BBU-RRH Association Optimization in Cloud-Radio Access Networks

Karen Boulos

université PARIS-SACLAY

UNIVERSITÉ PARIS SUD
Comprendre le monde, construire l'avenir

USJ 1875
Université Saint-Joseph de Beyrouth
جامعة القدّيس يوسف في بيروت

NNT : 2019SACLS209

Thèse de doctorat

BBU-RRH Association Optimization in Cloud-Radio Access Networks

Thèse de doctorat de l'Université Saint-Joseph de Beyrouth et de l'Université Paris-Saclay
préparée à l'École Supérieur d'Ingénieurs de Beyrouth et l'Université Paris-Sud

Ecole doctorale STIC
n°580
Spécialité de doctorat : Informatique

Thèse présentée et soutenue à Orsay, le 04 juillet 2019, par

KAREN BOULOS

Composition du Jury :

Sylvain Conchon Professeur des universités, Université Paris-Sud	Président du jury
Emilio Calvanese Strinati Directeur d'études, CEA	Rapporteur
Pascale Minet Chargée de recherche, INRIA	Rapporteur
Nadjib Ait Saadi Professeur, ESIEE Paris	Examinateur
Oriol Sallent Professeur, Université polytechnique de Catalogne	Examinateur
Samer Lahoud Professeur associé, USJ	Invité
Kinda Khawam Professeur associée, UVSQ	Co-encadrant de thèse
Marc Ibrahim Maître de conférences, USJ	Co-encadrant de thèse
Hadi Sawaya Professeur, USJ	Directeur de thèse, invité
Steven Martin Professeur des universités, Université Paris-Sud	Directeur de thèse

A ma grand-mère Antoinette.

A mon père Magdy, ma mère Nabila et mes deux frères Marc et Karl.

Trust in dreams, for in them is hidden the gate to eternity.

Gebran Khalil Gebran.

Remerciements

Avant de développer le travail, je profite en ces quelques lignes pour exprimer mes reconnaissances envers tous ceux qui ont contribué implicitement et explicitement à bien mener cette thèse.
Je remercie d'abord le Seigneur Jésus Christ pour être en vie et en bonne santé. Je remercie tous les membres de ma famille: Mon père Magdy et ma mère Nabila pour les êtres exceptionnels qu'ils sont. Sans leur support infini, cette thèse n'aurait jamais vu le jour. Je remercie mes deux frères Marc et Karl, qui par leur humour ont toujours apaisé mes soucis et rendu les longues journées de travail agréables. Une énorme pensée à cette si précieuse dame, ma grand-mère Antoinette, qui depuis toujours n'a cessé de m'encourager. Malgré son absence parmi nous, elle a toujours été mon refuge de consolation. C'est à elle que je dédie ce travail. Je tiens à exprimer toute ma reconnaissance envers Imad. Merci pour son support et son encouragement excessif.
Je remercie de même tous mes amis, collègues et compagnons. Surtout Catherine, Yara, Ola et Ghassan. Merci à chacun d'eux de m'avoir écouté aux moments de soucis, de m'apaiser en tristesse et d'être contents pour ma réussite. Je remercie vivement Juliana pour son aide et la sympathie qu'elle a montrée durant les préparatifs de la soutenance.
Un très grand merci à Gebran Al Ayan pour son support extrême durant cette thèse. Merci pour tout conseil qu'il m'a donné et pour cette confiance qu'il m'a accordée. J'exprime toute ma reconnaissance envers Yasser Karanouh. Merci pour ses conseils précieux et son encouragement excessif.
Merci de très fond du coeur à mon encadrante Kinda Khawam, qui fut pour moi bien plus qu'une encadrante visée. Merci pour sa disponibilité et son attitude réactive pour répondre à n'importe quelle question malgré les nombreuses charges qu'elle a. Merci à elle de m'avoir écouté plusieurs fois et d'être pour moi un très grand support technique et moral. Je tiens à exprimer toute ma reconnaissance envers mon encadrant Marc Ibrahim. Merci pour ses remarques et ses conseils techniques judicieux qui ont valorisé mon travail. Merci à lui pour ses interventions multiples surtout dans les procédures administratives. Très grand merci à mon encadrant Melhem El Helou pour ses conseils techniques et morals très constructifs. Merci pour ses remarques et sa rigueur scientifique qui ont toujours valorisé les articles scientifiques rédigés et les présentations orales délivrées. J'exprime de même toute reconnaissance envers Samer Lahoud, pour tout conseil scientifique et moral qu'il m'a donné. Merci à lui de m'écouter et conseiller plusieurs fois.
Je tiens à remercier mes deux directeurs de thèse Hadi Sawaya et Steven Martin. Merci pour leur confiance, disponibilité et support durant mon parcours. J'adresse mes remerciements chaleureuses à Pascale Minet, ma rapportrice, qui m'a donné des remarques très utiles pour la valorisation de mon manuscrit de thèse. Grand merci à Emilio Calvanese Strinati pour avoir accepté de rapporter ma thèse. Merci de même pour ses encouragements à plusieurs reprises. Merci à tous les membres du jury: Oriol Sallent, Sylvain Conchon et Nadjib Ait Saadi. Ce fut un grand plaisir de les avoir dans mon jury.
Enfin, J'adresse mes remerciements au conseil de la recherche de l'USJ, le CNRS-L et le service de l'AUF de m'avoir alloué une bourse pour le financement de ma thèse.

Résumé

De nos jours, la demande en trafic mobile a considérablement augmenté. Face à cette croissance, plusieurs propositions font l'objet d'étude pour remédier à un tel défi.

L'architecture des réseaux d'accès de type Cloud (C-RAN) est l'une des propositions pour faire face à cette demande croissante, et constitue une solution candidate potentielle pour les réseaux futurs 5G. L'architecture C-RAN dissocie deux éléments principaux de la station de base: La BBU ou "Baseband Unit", qui constitue une unité intelligente pour le traitement des données en bande de base, et le RRH ou "Remote Radio Head", constituant une antenne passive pour fournir l'accès aux utilisateurs (UEs). Grâce à l'architecture C-RAN, les BBUs sont centralement regroupées, alors que les RRHs sont distribués sur plusieurs sites. Plusieurs avantages sont ainsi dérivés, tels que le gain en multiplexage statistique, l'efficacité d'utilisation des ressources, et l'économie de puissance. Contrairement à l'architecture conventionnelle où chaque RRH est exclusivement associé à une BBU, dans l'architecture C-RAN plusieurs RRHs sont regroupés en une seule BBU lorsque les conditions de charge sont faibles. Ceci présente plusieurs avantages, tel que l'amélioration en efficacité énergétique et la minimisation de consommation de puissance. Dans cette thèse, nous adressons le problème d'optimisation des associations BBU-RRH. Nous nous intéressons à l'optimisation des regroupements des RRHs aux BBUs en tenant compte de critères multiples. Plusieurs contraintes sont ainsi envisagées, tel que la réduction de la consommation de puissance sous garantie de Qualité de Service (QoS) minimale. En outre, la prise en compte du changement du niveau d'interférence en activant/désactivant les BBUs est primordiale pour l'amélioration de l'efficacité spectrale. En plus, décider dynamiquement de la réassociation des RRHs aux BBUs sous des conditions de charges variables représente un défi, vu que les UEs connectés aux RRHs changeant leurs associations font face à des "handovers" (HOs).

Nous proposons d'abord une approche qui tient compte des interférences pour l'optimisation des associations BBU-RRH, et qui considère le comportement des RRHs lorsqu'ils sont associés à une seule BBU. Grâce à l'OFDMA (Orthogonol Frequency Division Multiple Access) appliqué au niveau d'une cellule, les RRHs partagent les ressources radio orthogonaux d'une BBU, et agissent comme une seule cellule équipée d'un système d'antennes distribués (DAS), entraînant ainsi une annulation des interférences "intra-cluster". Nous formulons d'abord le problème par une optimisation SPP (Set Partitioning Problem), dont le but revient à choisir les sous-ensembles des RRHs associés à une seule BBU, avec pour objectif principal la réduction de la consommation de puissance, sous contrainte de garantie minimale de QoS aux UEs. Le problème est représenté sous forme de programmation linéaire en nombre entiers et appartient à la catégorie des problèmes NP-complet. Par conséquent, nous proposons plusieurs heuristiques dont le but revient à fournir des solutions sous-optimales à complexité réduite. Nous appliquons ensuite ces heuristiques sur une topologie réelle et évaluons les résultats.

En outre, nous formulons le problème d'association BBU-RRH par une solution hybride. Particulièrement, nous avons recours à la théorie des jeux pour attribuer les RRHs aux BBUs, sous forme d'un jeu non-coopératif entre les RRHs en concurrence pour choisir leurs BBUs adéquates en fonction des conditions de charge, et du niveau d'interférence de chaque BBU. La convergence vers l'Equilibre de Nash Pur (PNE) est atteinte par l'algorithme de Meilleur Réponse et la Dynamique de Réplication. Le nombre de BBUs disponibles est décidé centralement en fonction de la QoS réalisée dans le réseau. L'objectif de notre proposition est de réduire les charges de signalisation imposées par les approches centralisées.

L'aspect dynamique du problème est traité de même. Particulièrement, sous des conditions de charge dynamiques, les associations BBU-RRH peuvent changer, entraînant ainsi des HOs aux UEs connectés aux RRHs subissant des réassociations. Notre but serait alors bi-objectif: La réduction de la consommation de puissance et la réduction du taux de réassociation des UEs desservies. Nous proposons ainsi un problème d'optimisation bi-objectif, ciblant les deux métriques, et formulé par SPP. De même, nous proposons une heuristique visant à trouver un bon compromis entre l'économie de puissance et la réduction du taux de réassociation. Nous proposons en outre, une approche hybride du problème dynamique, reposant sur la théorie des jeux, avec pour objectif de réaliser un bon compromis entre les deux métriques, tout en réduisant la charge de signalisation.

Une dernière contribution consiste à exploiter les avantages du C-RAN, avec le partage commun des ressources spectrales et des infrastructures réseaux entre plusieurs opérateurs mobiles coopérant ensemble. Nous proposons ainsi deux algorithmes différents de coopération dans les C-RANs, et évaluons les gains supplémentaires en économie de puissance résultant des associations BBU-RRH. Un premier algorithme repose sur une technique d'apprentissage par renforcement, le problème de Bandit Manchot (MAB), où les opérateurs utilisent mutuellement leurs bandes pour réduire les interférences. Un second, reposant sur l'agrégation des bandes (CA), qui constitue une spécification du standard LTE, version 10. Les deux propositions sont comparées par rapport au cas où les opérateurs ne partagent pas leurs ressources, dans le but d'évaluer les avantages de coopération dans l'architecture C-RAN.

Abstract

The demand on mobile traffic has been largely increasing nowadays. Facing such growth, several propositions are being studied to cope with this challenge.

Cloud-Radio Access Networks Architecture (C-RAN) is one of the proposed solution to address the increased demand, and is a potential candidate for future 5G networks. The C-RAN architecture dissociates two main elements composing the base station: The Baseband Unit (BBU), consisting in an intelligent element to perform baseband tasks functionalities, and the Remote Radio Head (RRH), that consists of a passive antenna element to provide access for serviced User Equipments (UEs). In C-RAN architecture, the BBUs migrate to a Cloud data center, while RRHs remain distributed across multiple sites. Several advantages are derived, such as statistical multiplexing gain, efficiency in resource utilization and power saving. Contrarily to conventional architecture, where each RRH is associated to one BBU, in C-RAN architecture, multiple RRHs can be embraced by one single BBU when network load conditions are low, bringing along several benefits, such as enhanced energy efficiency, and power consumption minimization. In this thesis, the BBU-RRH association optimization problem is addressed. Our aim is to optimize the BBU-RRH association schemes, taking into consideration several criteria. The problem presents many constraints: For example, achieving minimized power consumption while guaranteeing a minimum level of Quality of Service (QoS) is a challenging task. Further, taking into account the interference level variation while turning ON/OFF BBUs is paramount to achieve enhanced spectral efficiency. Moreover, deciding how to re-associate RRHs to BBUs under dynamic load conditions is also a challenge, since connected UEs face handovers (HOs) when RRHs change their associations.

In our work, we first propose an Interference-Aware approach for BBU-RRH association optimization, taking into consideration the RRHs behavior while assigned to one BBU. Due to Orthogonal Frequency Division Multiple Access (OFDMA) technique applied at a cell level, RRHs share orthogonal radio resources of one BBU, and act as a single cell equipped with a Distributed Antenna System (DAS), leading to intra-cluster interference cancellation. The problem is first formulated as a Set Partitioning Problem (SPP), to choose adequately the subsets of RRHs associated to one BBU, with main objective of reducing the power consumption and guaranteeing a minimum level of QoS for UEs. The problem is an Integer Linear Programing (ILP) and belongs to the NP-complete category. Consequently, we propose several heuristics, with the purpose of providing sub-optimal solutions, while reducing the computational requirements. We further, submit those heuristics to the stringency of a real network topology and evaluate the results.

We also formulate the BBU-RRH association problem as a hybrid solution. In particular, we resort to Game Theory to portray the RRHs assignment to BBUs, as a non-cooperative game among RRHs that compete on available BBUs in order to choose adequately their associations according to load and interference levels within each BBU. The convergence to Pure Nash Equilibrium (PNE) is attained via the Best Response dynamics and the Replicator Dynamics. The number of available BBUs is centrally decided according the realized QoS in the network. The purpose of our devised scheme is to reduce the amount of signaling load overhead that centralized approaches require.

Moreover, the dynamic aspect of the problem is investigated. In particular, under dynamic load variations, the BBU-RRH association schemes might change, causing HOs for serviced UEs connected to re-associated RRHs. Our purpose is bi-objective: The power consumption minimization, and the re-association rate reduction for serviced UEs. Thus, we present a bi-objective optimization problem targeting both metrics, and formulate it as a SPP. We also propose a heuristic, aiming to find a good compromise between power saving and re-association rate reduction. Further, we propose a hybrid formulation relying on Game Theory for the dynamic problem, targeting to realize good tradeoff between both metrics, while reducing signaling load overhead.

A last contribution consists in exploiting the benefits of C-RAN, along with common spectrum and active infrastructure sharing among multiple cooperating Mobile Network Operators (MNOs). In particular, we propose two different algorithms for MNOs cooperation in C-RAN, and assess the additional power saving derived from BBU-

RRH associations. A first algorithm relies on a reinforcement learning solution, the Multi-Armed Bandit problem (MAB), where MNOs can access each others' bands to mitigate interference. A second, consists of LTE release 10 specification, the Carrier Aggregation (CA). The two propositions are compared with the case where MNOs do not share their resources, to assess the benefits of cooperation in C-RAN architecture.

Acronymes

3GPP	Third Generation Partnership Project
5G	Fifth Generation
BBU	Baseband Unit
BFD	Best Fit Decreasing
BIL	Binary Integer Linear programing
CA	Carrier Aggregation
CAPEX	Capital Expenditures
CQI	Channel Quality Indicator
CSI	Channel State Information
CRS	Cell Specific Reference Signals
C-RAN	Cloud-Radio Access Networks
CPRI	Common Protocol Radio Interface
CoMP	Coordinated MultiPoint
CB	Coordinated Beamforming
CB/CS	Coordinated Beamforming and Coordinated Scheduling
CDD	Cyclic Delay Diversity
DAS	Distributed Antenna System
D-RAN	Distributed Radio Access Network
eX2	enhanced X2 interface
FFD	First Fit Decreasing
FBPA	Full Bin Packing Algorithm
GPS	Generalized Processor Sharing
HOs	Handovers
HARQ	Hybrid Automatic Retransmit reQuest
ILP	Integer Linear Programing
ICIC	Inter-Cell Interference Cancellation
ISI	Inter-Symbol Interference
JT	Joint Transmission technique
KPI	Key Performance Indicator
LSAS	Large Scale Antenna Systems
LTE	Long Term Evolution

M2M	Machine-to-Machine
MCN	Mobile Cloud Networking
MEC	Mobile Edge Computing
MNOs	Mobile Network Operators
MSC	Mobile Switching Center
MAB	Multi-Armed Bandit
MIMO	Multiple-Input Multiple-Output
NFD	Next Fit Decreasing
NGFI	Next Generation Fronthaul Interface
NGMN	Next Generation Mobile Networks
OFDMA	Orthogonal Frequency Division Multiple Access
OPEX	Operational Expenditures
PNE	Pure Nash Equilibrium
QoE	Quality of Experience
QoS	Quality of Service
RAU	Radio Aggregation Unit
RCC	Radio Cloud Center
RME	Radio Management Element
RRC	Radio Resource Control
RRM	Radio Resource Management
RRH	Remote Radio Head
RRS	Remote Radio System
RRU	Remote Radio Unit
RBs	Resource Blocks
SLA	Service Level Agreement
SPP	Set partitioning Problem
SINR	Signal-To-Interference Noise Ratio
SDR	Software Defined Radio
TCOs	Total Cost of Ownerships
TTI	Transmit Time Interval
TPS	Transmission Point Selection
WDM	Wavelength Division Multiplexing
WFD	Worst Fit Decreasing

Table des matières

Liste des tableaux

Table des figures

Chapitre 1

Introduction

The exponential growth of mobile traffic is a real hurdle for MNOs nowadays. New traffic patterns are introduced with different delay and bandwidth requirements. Facing such demand, MNOs resort to flexible and efficient solutions to cope with the continuous demand on traffic, without compromising their financial costs and return-on-investments. Cloud-Radio Access Network, known as Cloud-RAN or C-RAN, has been introduced to address the challenges that MNOs are facing, and to meet their requirements in terms of reducing the Capital Expenditures (CAPEX) and Operational Expenditures (OPEX). The concept of the C-RAN architecture relies on "Cloudification". In particular, the baseband processing elements (i.e, BBUs) are migrated to a cloud data-center, where centralized management of baseband tasks is introduced. Light-weighted radio elements (i.e, RRHs), remain distributed across many sites to provide access for UEs. This chapter presents the C-RAN architecture, and compares it with the conventional architecture, the Distributed RAN (D-RAN). Further, the motivations behind the C-RAN architecture are also presented, and a review on its different challenges is provided. The thesis objective and the summary of contributions are also detailed in this chapter.

1.1 Mobile Network Operators Challenges

Over the past few years, the demand for mobile traffic has been largely increasing. According to [3], the global mobile data will reach 77 exabytes per month by 2022. Recent reports show that more than 600 million (*i.e*, 648 million) mobile devices were added in 2017. By 2022, there will be 8.4 billion personal mobile-ready devices, and 3.9 billion Machine-to-Machine (M2M) connections.
In response to this growth, MNOs are obliged to increase the number of active base stations in their serving areas. However, this has many disadvantages :

- *The increased CAPEX and OPEX costs*: As new deployments must be done, MNOs are becoming aware about their Total Cost of Ownerships (TCOs). The TCOs mainly include the CAPEX, representing the cost of site construction and base station purchases, and the OPEX, representing the network maintenance and operation costs. It is reported that OPEX accounts for 60% of TCOs, whereas CAPEX accounts for 40% [1]. Thus, MNOs should consider carefully the OPEX while building their RAN, in a way to reduce their TCOs, without compromising the QoS of their subscribed UEs.
- *The increased amount of power consumption*: As the number of base stations increases, the associated power consumption gets more significant. High power consumption has several disadvantages and bad impacts on OPEX costs and environment.
- *The degradation of signal quality*: Due to hard spectrum reuse schemes, the signal quality is degraded, and the total level of interference is aggravated. Further, the deployment of small cells, for the purpose of capacity increase, imposes additional challenges regarding interference management and operational complexity [4].
- *Low base station utilization rate*: Because of mobile traffic dynamicity, some base stations may be overloaded, while others are under-used. As MNOs are designed to adequately serve UEs during day peak hours, base stations may often remain inefficiently exploited, noting that the processing power of idle mode base stations is wasted because they need the same level of energy as they do during busy hours [1].

All of the mentioned challenges are pushing MNOs to adopt a new cost-effective RAN architecture. In that context, the C-RAN architecture has been introduced, and has been motivated in many projects, such as the "Interworking and JOINt Design of an Open Access and Backhaul Network Architecture for Small Cells based on

Cloud Networks" [5], launched in 2012 by the European Commission, and the "Mobile Cloud Networking (MCN)" [6], launched in 2013 also by the European Commission. Further, many Asian-Pacific MNOs have been tempted by the advantages of C-RAN architecture, and have already started to plan its deployment. For example, Korean SK Telecom and NTT-DoCoMO have announced early trials of C-RAN in 2019. It will also be adopted in 2020 Olympic Games [7]. It has further gained interest among many local MNOs and vendors, such as AT&T, Vodafone, Verizon Communications, Orange, Intel, ZTE, Huawei and Nokia Bell Labs, which are still investing in its derived benefits, and defining its ecosystem.
The C-RAN consists of a new cloud architecture, aiming to face the need of TCOs reduction. This approach was conceived from the cloud computing concept. In particular, BBUs are migrated in a single data-center, where baseband resources are allocated on demand. Precisely, baseband elements are employed efficiently and follow the instantaneous load conditions in the network, instead of adopting the maximum traffic of individual base stations. Consequently, processing power is reduced and adapts to network instantaneous load.

1.2 D-RAN vs. C-RAN Architectures

1.2.1 D-RAN Architecture

In conventional architecture, base stations operate in a distributed fashion. In particular, each base station deployed on a site, consists of a BBU and a RRH. The BBU is an intelligent element used to perform the baseband tasks functionalities (*i.e*, physical and MAC layers functionalities). The RRH is a light-weighted radio element, acting as a passive antenna, and used to perform signal modulation and amplification. The BBU and RRH are linked through a fiber optic link using the Common Protocol Radio Interface (CPRI) [8]. Figure 1.1 illustrates the architecture of the base station in a D-RAN system.

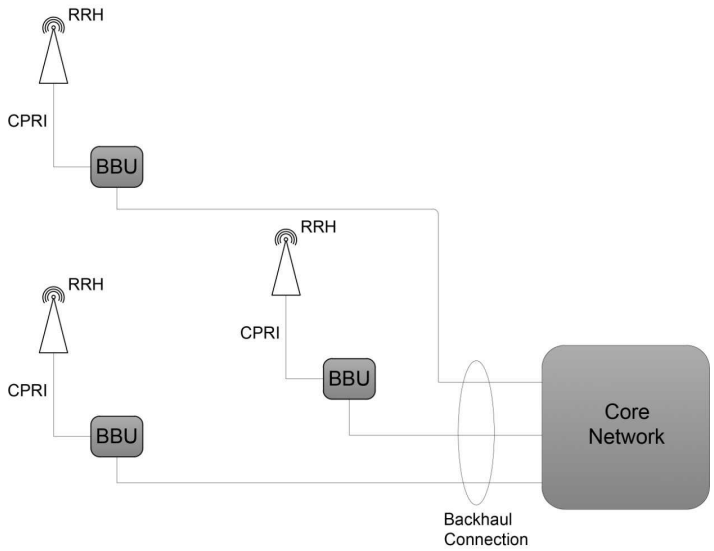


FIGURE 1.1: D-RAN Architecture

1.2.2 C-RAN Architecture

China Mobile Institute was the first to introduce the C-RAN concept in 2009. The C-RAN stands for Centralized, Collaborative, Cloud, and Clean RAN. In C-RAN, BBUs are migrated from sites, to be gathered in a single location. The latter consists of a central office or a super macro site used to aggregate BBUs. RRHs are connected to BBUs through high-bandwidth and low latency optical links. The Wavelength Division Multiplexing (WDM) is employed to interconnect RRHs to the BBU pool, using the CPRI protocol. This architecture is crucial for virtualization. In particular, it leverages the benefits of Software-Defined Radio (SDR) [9], allowing for baseband functionalities to be instantiated over software, without the need for a dedicated hardware, as in a D-RAN environment. Such feature enables the creation of the concept of RAN as a service (RANaaS), where network functions are executed over

the top of a virtualized environment controlled by a cloud Operating System. Figure 1.2 displays the architecture of C-RAN.

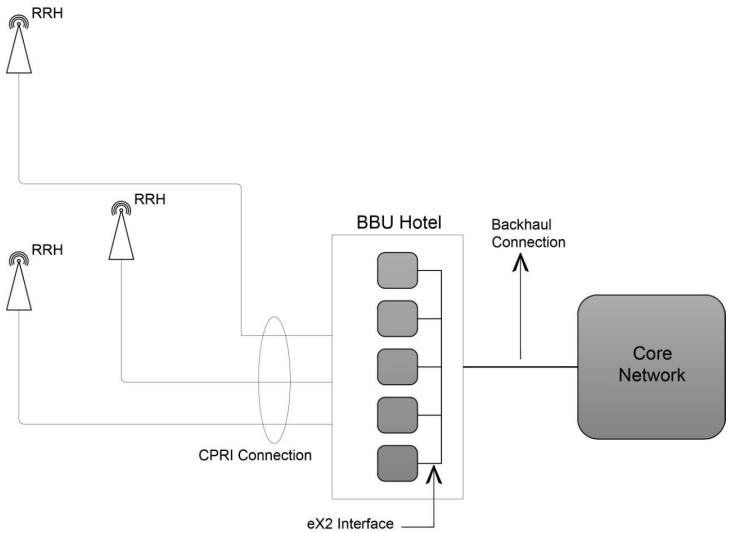


FIGURE 1.2: C-RAN Architecture

1.2.3 Motivations Behind C-RAN Architecture

The migration of BBUs to a central location allows to remove many hardware elements, such as routers, cooling and heating elements. Further, additional spaces and reduced costs are introduced, simplifying network management and providing deployment scalability and flexibility in terms of reconfigurability and extensibility. In particular, it facilitates the deployment of RAN and speeds up network construction in terms of site selection and civil works, since important equipments are gathered in a single location, reducing site visits for maintenance. Hence, the OPEX is significantly reduced, bringing along cost savings, which constitutes the most appealing property of C-RAN, motivating MNOs to adopt it.

Further, since multiple BBUs are co-located in single geographical location, they cooperate more easily. For example, the enhanced X2 interface (eX2) in the case of LTE would be more effective, resulting in improved performance and utilization efficiency. For example, facilitation in terms of network information exchange, such as signaling, and Channel State Information (CSI) of UEs, is introduced, leading to efficient utilization of baseband resources and interference management. Additionally, a load-balancing scheme among BBUs is also introduced, and adapts to the different traffic patterns that vary in time and spaces. The S1 interface is further enhanced, since one single router is used to aggregate traffic and to provide connection with the core network, instead of resorting to several small ones at each site, as in the case of D-RAN.

Besides its several benefits, the C-RAN is also a drive key toward future 5G networks [10], enabling several technologies, such as ultra-dense Multiple-Input Multiple-Output (MIMO) systems, Large Scale Antenna Systems (LSAS), and full-duplex systems [11].

Additionally, the C-RAN architecture allows for flexible BBU-RRH associations. For example, when network load conditions are low, several RRHs can be handled by one BBU, reducing the number of active BBUs and consequently the power consumption. Contrarily to the D-RAN architecture where one RRH is always mapped to one BBU. Further, when RRHs are associated to one BBU, they act as a DAS [12], and logically form a single cell. Thus, UEs' data are carried through different RBs in an orthogonal fashion. In particular, the data are multiplexed over orthogonal subcarriers, composing the bandwidth of one cell. UEs' data are received without any expected interference, and a cluster of RRHs is free of intra-cluster interference, which enhances UEs' radio conditions, leading to higher spectral efficiency and efficient resource utilization.

1.3 C-RAN Challenges

Besides its several advantages, the C-RAN architecture presents drawbacks which can limit its deployment in some cases :

- *Fronthaul Limitations*: In LTE systems, the transmission link with at least 10 Gbps and a maximum latency of 250 μs is paramount to transport the baseband data from BBUs to RRHs with respect to LTE specification. This leads to limit the distance between BBUs and RRHs, reaching a maximum of 20 km according to Next Generation Mobile Networks (NGMN) [13] specifications. As a consequence, fiber optic infrastructure is necessary at almost all locations, increasing the cost of CAPEX, especially in densified heterogeneous networks. Further, CPRI will encounter limitations in future 5G systems, where the capacity of one cell will reach 10 to 20 Gbps, requiring over 300 Gbps for CPRI which currently supports a maximum of 24 Gbps capacity. In fact, studies of functional baseband splits is highly motivated in the recent state-of-art [14] [15] [16]. The reason behind this is to reduce the data bandwidth and latency requirements between BBUs and RRHs. However, this may compromise the multiplexing gains introduced by the centralized architecture.
- *Coordinated Multipoint Technique implementation*: The C-RAN architecture facilitates for Coordinated Multipoint (CoMP) technique [17], where BBUs easily exchange CSI and signaling exchange for interference mitigation and enhanced spectral efficiency. However, this comes at the cost of increased complexity and overheads when the number of RRHs increases. A challenge in C-RAN architecture consists of CoMP implementation. In particular, when numerous cells with high data rate are handled, strict constraints on timing, imposed by LTE systems, are required. For instance, in D-RAN architecture, the required time of 3GPP LTE Hybrid Automatic Retransmit request (HARQ) protocol is 3 ms [18]. In C-RAN, additional fronthaul delays are introduced, requiring lower time processing on BBUs side than conventional schemes to respect the strict timing. When CoMP techniques are employed, more challenges and constraints on processing time are introduced, making its implementation a challenge.
- *Resource management*: In D-RAN architecture, various optimization algorithms are proposed for Radio Resource Management (RRM), such as scheduling algorithms, power control, interference mitigation, and admission control. The C-RAN introduces additional challenges in RRM, requiring new algorithms to be designed. Such algorithms not only focus on spectral efficiency enhancement or interference mitigation, but also allow optimized BBU-RRH associations. A proper design for BBU-RRH associations is a challenge, and should be performed carefully so as to adapt to traffic variations and respect a minimum level of QoS for serviced UEs. In this thesis, we will focus on the BBU-RRH association optimization, also known as RRH clustering, in the C-RAN architecture.

1.4 Thesis Contributions

Interference-Aware Approach for BBU-RRH Association

The BBU-RRH association problem has gained considerable interest among researchers. Most of them formulate the problem as a Bin Packing optimization, where BBUs are modeled as bins of fixed capacity and RRHs as objects of different volumes. The objective would be to find a suitable arrangement of objects into the bins in a way to reduce the number of used bins subject to the capacity of the latters. The BBU capacity is expressed in terms of the maximum number of supported UEs, the number of baseband tasks or the number of RBs. Although significant power savings are derived from this approach, the spectral efficiency improvement of a BBU while embracing multiple RRHs is not taken into account. In fact, RRHs share orthogonal radio resources of one BBU and act as a DAS, forming one logical cell. Intra-cluster interference is consequently canceled among RRHs assigned to one BBU. The latter might handle more UEs since radio conditions are improved. A first contribution of our work relies on taking into consideration the DAS behavior of RRHs while being associated to one BBU. Unlike existing studies, in our work we consider the interference level change in the network as a function of the chosen subset of RRHs clustered together, affecting the spectral efficiency of the BBU. An Interference-Aware approach is provided, and the problem is translated into a SPP [19], with main objective of reducing the power consumption subject to a minimum level of throughput requirement per UE. A comparison with the Bin Packing approach and the conventional architecture is provided, and results show that our Interference-Aware approach significantly enhances the power savings in comparison with the different schemes. Since the SPP belongs to the NP-complete category, we further propose a heuristic and compare it to the optimal solution. The results show that the proposed heuristic provides close performance to the optimal solution, while reducing the complexity. Additionally, we propose several heuristics

that we submit to the stringency of a real network topology and compare among them.

Hybrid Approach for BBU-RRH Association

Beside the Interference-Oblivious behavior of existing propositions, recent studies rely on centralized approaches for the problem. Those approaches require that RRHs frequently send their load and radio conditions to the cloud, where the metrics are assessed and decisions are made. In that case, the signaling load overhead of RRHs might impose constraints over the optical fronthaul connections between BBUs and RRHs. Particularly, when the number of RRHs is high. In that context, a second contribution of our work relies on formulating a hybrid scheme, with the purpose of reducing the amount of signaling load overhead between BBUs and RRHs. Two versions of the hybrid scheme are proposed : A first, adapting to a homogeneous UEs distribution in the network, where the number of UEs per cell is same. A second, adapting to a heterogeneous UEs distribution in the network, which is more realistic. The proposition relies on a game theoretic framework, where the association of RRHs to available BBUs is portrayed as an exact potential game among RRHs competing on available BBUs. RRHs association decisions are aware of the level of interference. In fact, each RRH seeks to reduce selfishly its own cost function. The latter is affected with the RRHs that choose concurrently the same BBUs (partaking the common radio resources) but also with those RRHs that joined other BBUs through their harmful interference impact. The number of available BBUs is centrally decided according to the realized throughput per BBU. The game among the RRHs is solved via the Best Response dynamics and the Replicator Dynamics consisting in a reinforcement learning technique. Results are compared with the global optimal solutions and show that the signaling load overhead generated by the hybrid proposition is significantly reduced, while performance metrics show closeness to the centralized approaches.

BBU-RRH Association with Re-association Optimization

Additionally, the dynamic aspect of the problem is treated in our work. Although few works [20] [21] [22] have considered the problem from a dynamic perspective, they did not take into consideration the re-association of RRHs to BBUs when network load conditions vary. In fact, under dynamic load variations, some UEs may arrive, while some others may quit the system, causing BBUs to be highly or lightly loaded. However, when a BBU is highly loaded, connected UEs might face a degradation in the level of QoS due to resource shortage, unless BBU-RRH association schemes change, so that RRHs re-associate to less loaded BBUs. On the other hand, when a BBU is lightly loaded, its assigned RRHs will re-associate to different BBUs in order to be turned OFF and to improve power savings. Frequent re-associations between BBUs and RRHs cause high rate of HOs between serviced UEs. In the present contribution, we tackle that problem by proposing a bi-objective optimization problem, formulated as a SPP, where the purpose is to jointly reduce the power consumption and the re-association rate of UEs, while guaranteeing a minimum level of throughput requirement per UE and being aware of the interference level change in the network. We also show a tradeoff between power savings and re-association rate of UEs. In fact, reducing the number of active BBUs may lead to resource shortage under any fluctuation in the load conditions, unless BBU-RRH association schemes change. However, this causes HOs for serviced UEs. On the other hand, activating many BBUs always guarantee the targeted QoS since enough resources would be available. Therefore, RRHs associations are not changed and UEs do not endure frequent HOs. However, this lowers the level of energy efficiency. Consequently, we propose a heuristic aiming to find a good tradeoff between both metrics, and whose performance outperform the Bin Packing problem, while reducing the complexity in comparison with the optimal solution. We lastly propose a hybrid approach for the problem, aiming to minimize the signaling load overhead in the network, and to find a good compromise between power savings and re-association rate reduction. The results of the hybrid proposition show close performance to the centralized scheme with less signaling load overhead. Further, its performance in terms of power savings and re-association rate reduction are enhanced in comparison with the Bin Packing approach.

BBU-RRH Association in a Multi-Operator Context

We lastly exploit the BBU-RRH association problem in a multi-operator context, which has not been treated from that perspective in the recent state-of-art. Our purpose is to examine the additional benefits in terms of energy and spectral efficiency when multiple MNOs put in common their resources in terms of infrastructure and spectrum within a C-RAN architecture. In fact, when centralized RAN management is allowed with MNOs cooperation, many benefits are derived, such as additional power savings resulted from BBU-RRH association optimization, and enhanced spectral and energy efficiency. Through two proposed algorithms, we show that despite the realized power economy

achieved in C-RAN architecture, additional power savings, and enhanced spectral and energy efficiency are realized when network cooperation is made possible. The first proposition relies on a reinforcement learning technique (the MAB problem) for common spectrum sharing, where MNOs can access each others' bands to mitigate interference within the shared RAN. The second relies on the CA technique, an existing feature of the release 10 LTE specification to increase the bandwidth. Results of the two propositions are compared with the case where MNOs do not share their resources, and show that additional power savings and enhanced spectral and energy efficiency are derived when MNOs cooperate in a C-RAN architecture.

1.5 Thesis Organization

The remaining of this thesis is organized as follows: BBU-RRH association problem is surveyed in chapter 2. We provide a wide review on the existing approaches, and come up with a key conclusion concerning the existing formulations. Chapter 3 provides the Interference-Aware proposition for the BBU-RRH association problem formulated as a SPP. Additionally, the heuristic is provided for the purpose of complexity reduction. We evaluate the results of our Interference-Aware proposition in comparison with the Bin Packing approach, and we consider different radio condition scenarios. The performance metrics of the heuristic are also evaluated in comparison with the optimal solution. In chapter 4, we propose several heuristics that we submit to the stringency of a real case topology [23], and compare among them.

Chapter 5, provides the hybrid formulation of the problem. The two versions of the proposition, adapting to uniform and non-uniform UEs distribution, are described. We first start by proposing the solution that adapts to a uniform UEs distribution in the network. We then enhance the proposed scheme to adapt to a heterogeneous UEs distribution, which is more realistic. A centralized Load- and Interference-Aware proposition is also presented in this chapter, and used as a benchmark to compare the results.

In chapter 6 the dynamic aspect of the problem is addressed. In particular, we present the bi-objective optimization formulation aiming to reduce the power consumption at one hand, and the re-association rate of UEs caused by the dynamic BBU-RRH re-associations, on the other hand. We further present the proposed heuristic, aiming to provide good compromise between power saving and re-association rate reduction. We also introduce the hybrid formulation of the problem, in order to reduce the amount of signaling load. In particular, the proposed algorithm in chapter 5 is adapted to reduce to re-association rate among UEs.

In chapter 7, we investigate the common spectrum and active infrastructure sharing among MNOs in C-RAN architecture. We present the two proposed schemes for MNOs cooperation, and assess the additional power savings derived from BBU-RRH association optimization, when cooperation is made possible in C-RAN. We compare both algorithms with the case where MNOs do not share their resources and assess the benefits of cooperation.

Chapter 8 concludes the thesis and provides the short and long term perspectives.

Chapitre 2

BBU-RRH Association Optimization

The key concept behind a centralized architecture is to enhance the efficiency in resource utilization, and thus limiting the growth of the MNOs TCOs, the CAPEX, and the OPEX. A flexible re-design of the BBU-RRH associations, introduced by the C-RAN architecture, is advantageous since RRHs assignment to BBUs adapts to the network load variations. As a consequence, the number of active BBUs decreases, leading to power consumption reduction and enhanced energy efficiency. In this chapter, we provide an exhaustive survey on the existing works, aiming to optimize the BBU-RRH association. We further come up with a key conclusion, concerning all existing schemes that do not take into account the interference level in the network when some BBUs are put into OFF state. Precisely, RRHs associated to one BBU share the bandwidth of the latter and act as a DAS. As a result, a cluster of RRHs is free of intra-cluster interference and more UEs are served. In this chapter, the RRHs behavior while being associated to one BBU is explained in details, and the need for a new approach to RRHs assignment is put forward.

2.1 Introduction

An efficient use of the radio resources is an important challenge for MNOs nowadays. The densification of base stations contributes to an imbalance in the resource consumption due to the disparity in mobile traffic across the network. Accordingly, some base stations may be overloaded, while others have radio resources to spare. Figure 2.1 illustrates an example of the traffic pattern during a day: within early hours of the day, we notice a huge load coming from the central office areas, while this amount of load decreases significantly by the end of the day. The opposite effect is shown in residential areas. The processing power in idle mode base stations is wasted because the same level of energy is needed as in busy hours, and because MNOs must provide 24x7 coverage [1]. In addition, base stations are dimensioned to handle a maximum number of active UEs during busy hours. This leads to an inefficient use of the total capacity, in particular during non-busy hours. A flexible assignment of resources is a must in order to cope with this challenge.

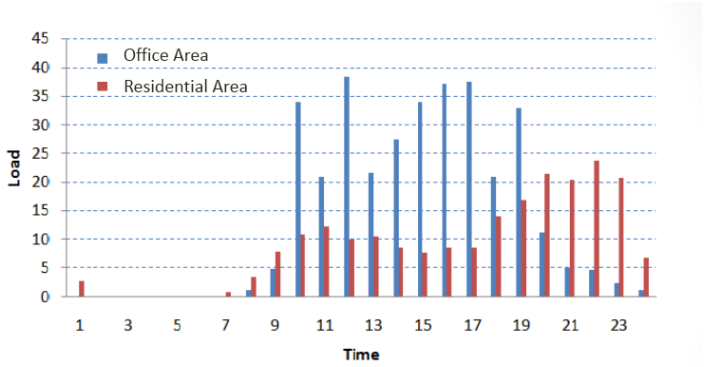


FIGURE 2.1: Traffic Load Variation [1]

2.2 BBU-RRH Flexible Association

In the conventional architecture, one RRH is associated with a single BBU (cf. Figure 2.2). When a BBU manages a single RRH, the connected UEs share the radio resources of the latter. Consequently, the average rate per UE is maximized. However, it is necessary to provide as many active BBUs as RRHs (*i.e*, radio sites). The concept of a flexible reassignment between BBUs and RRHs was firstly motivated in [24], for its various advantages:

— First, a dynamic change of the BBU-RRH associations can adapt to a non-uniform traffic and to the various fluctuations within the load conditions, leading to significant reduction in the number of active BBUs and thus, to an enhancement in the energy efficiency.

— Second, a reduction in the processing and signaling load overhead related to frequent HOs in cellular networks is introduced, as many RRHs are associated to one BBU, increasing the covered region of the latter, and allowing it to handle several RRHs without a degradation in the link quality of high mobility UEs.

— Third, a DAS system behavior [12], where multiple antenna elements are geographically scattered, is introduced, leading to diversity gain and to transmit power reduction.

Although several advantages are introduced from a flexible design, in the case of high load conditions, this might decrease the system capacity, which contributes to a degradation in the level of QoS, as many RRHs share the radio resources of one BBU. Therefore, it is crucial to dynamically decide the BBU-RRH association according to several criteria (cf. section 2.3). The design should be performed carefully so as to optimize the resource allocation.

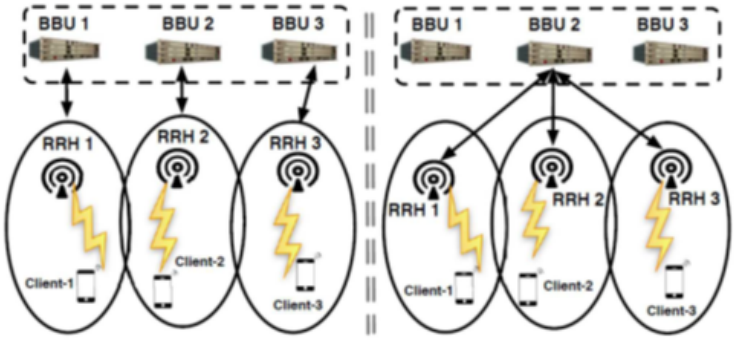


FIGURE 2.2: BBU-RRH Association [2]

System Architecture

Changing the associations between BBUs and RRHs implicates many elements. The most important ones are the Radio Management Element (RME) that controls the BBUs and orders the switching according to the collected traffic statistics per BBU, the optical switches, the splitters and the multiplexers. Figure 2.3 provides the functional blocks of the BBU-RRH switching procedure described as follows: Each RRH sends its traffic volume to the associated BBU periodically. The BBUs report an average amount of traffic volume per RRH to the RME. According to a traffic profile and to the existing connections between BBUs and RRHs, the switching plan is decided and sent to the BBUs. Afterwards, the decided scheme is executed by switching the BBU-RRH connections. A switching-end report is sent back to the RME to update the current connection schemes. The switching procedure is performed in the optical domain for its scalability and low-latency [2]. In particular, the baseband tasks between BBUs and RRHs are digitalized and carried over CPRI digital interface [8], enabling the use of the optical switches, splitters and multiplexers, along with the WDM technique.

2.3 BBU-RRH Association Criteria

Achieving an efficient scheme necessitates to take into consideration several criteria that are sketched in the following points:

— *The load balancing among BBUs*: A BBU charge cannot exceed a limited threshold, given normally as the BBU capacity. To obtain a good level of QoS, it is important to balance the BBUs charge, so that the system resists to any possible fluctuation in the load conditions.

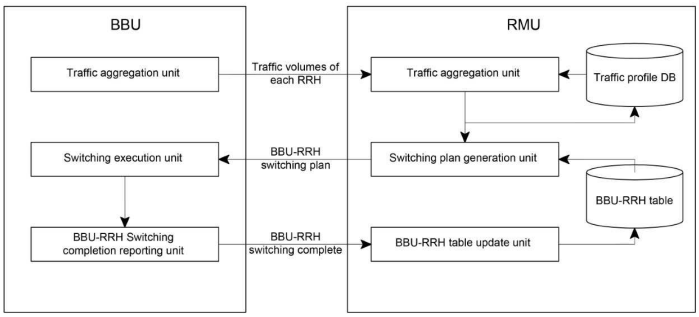


FIGURE 2.3: BBU-RRH Switching Procedure

— *The required QoS of UEs*: The key concept behind realizing a green network is finding a good compromise between QoS and power saving [25]. As previously stated, clustering several RRHs to a single BBU causes the degradation of the QoS level in the case of high load conditions. Contrarily, when one RRH is assigned to one BBU, the QoS would be guaranteed under any load condition. However, this increases the power consumption and lowers the energy efficiency of the MNOs.
— *The rate of HOs*: In 3GPP networks, HOs are handled by the Radio Resource Control (RRC) protocol [26]. This involves the roles of the BBUs and UEs. The RRHs remain transparent to that procedure, since they simply act as passive antenna elements. HOs occur when UEs change their serving BBUs. Contrarily to the conventional architecture, the HO factor in the C-RAN architecture introduces additional constraints on the BBU-RRH association design. On one hand,mapping neighboring RRHs to one BBU increases the coverage of the latter and contributes to HO rate reduction. On the other hand, re-associating a RRH to a different BBU increases the HO rate.
— *The interference level among RRHs*: Assigning many RRHs to one BBU lowers the level of interference in the network. In fact, the less is the number of active BBUs, the less the operating spectrum is reused. Finding a good compromise between the QoS necessitating to activate a high number of BBUs on one hand, and a low level of interference which requires the less use of the available bandwidth on the other hand, is crucial to realize an efficient design.

When multiple objectives, such as the QoS, the HO rate and the interference level are addressed, the BBU-RRH association problem remains a challenging task that will be targeted throughout the thesis.

2.4 BBU-RRH Association Approaches

The BBU-RRH association optimization has taken considerable interest among researchers. In this section, we review many relevant works and classify them into different approaches. Further, we conduct an important conclusion.

2.4.1 The Bin Packing Approach

The Bin Packing formulation [27] has been widely adopted in the recent state-of-art. The problem considers two finite sets of objects and bins. The objects are of different volumes, whereas the bins are of fixed capacity. The objective is to find a suitable arrangement of objects into the bins in a way to reduce the number of used bins. Within a C-RAN context, the BBUs are modeled as the bins of fixed capacity. The RRHs are the objects of different volumes, such as the number of required RBs [28] [20], the connected number of UEs [29] [30] and the required number of baseband tasks [31] [32].

Works in [20], [21], [33], [34], [31], [32], [35], [28] and [36] have considered a one-dimensional Bin Packing formulation. In [20], the authors propose a dynamic RRH assignment algorithm based on the First Fit Decreasing (FFD), a heuristic of the Bin Packing problem, combined with a dynamic re-assignment of RRHs mapped to crowded BBUs. They compare their solution to the classical FFD. Through conducted simulations, they prove that their proposition outperforms the classical FFD in terms of power saving and computational requirements. In the same line, [21] considers a “Full Bin Packing Algorithm (FBPA)” that aims to maximize the utilization of the available BBUs. In other terms, they resort to the Best Fit Decreasing (BFD) algorithm to associate the RRHs to available BBUs. They also

use a dynamic scheme with the objective of activating/deactivating BBUs when they are highly/lowly loaded respectively. A modified BFD is also adopted in [33], where the authors aim to load balance the baseband tasks among BBUs. Their proposed solution is compared to the classical BFD and shows a decrease in the number of available BBUs. The work in [34] also resorts to the BFD algorithm to solve the problem. In [31], the authors model the UEs baseband tasks as objects of different volumes that need to be packed into bins (*i.e*, BBUs). The FFD and the Next Fit Decreasing (NFD) are used to find acceptable solutions. One RRH is assumed able to be mapped to multiple BBUs, requiring inter-BBU communications. Yet, the resulting additional power consumption has been neglected. In [32], a heuristic simulated annealing method, that combines a Bin Packing algorithm with a simulated annealing scheme, has been introduced. A two-layered algorithm tries to map one or many RRHs to a single BBU within a first stage. A second stage is used to associate an unpacked RRH with multiple BBUs, and considers the additional power consumption required by inter-BBU communication. In [35], the authors propose a formulation by a Minimum Bin Slack algorithm [37], a heuristic of the Bin Packing, outperforming the FFD, with the purpose of load balancing the baseband tasks of different RRHs into virtual BBUs (considered as software applications running on servers). Furthermore, the BBU-RRH association problem is formulated as a Bin Packing problem in [28]. The objective is to reduce the number of active BBUs, leading to a decrease in network power consumption. In order to minimize the HO frequency, only neighboring RRHs are associated together. In [36], the authors consider a heterogeneous RAN context. In particular, they consider L-RRHs and H-RRHs with low and high power transmission respectively. They propose to cluster L-RRHs using fuzzy logic scheme [38] within a first stage. The H-RRH clustering is formulated as a Bin Packing optimization scheme and is performed within a second stage.

Several works consider the joint problem of resource optimization and BBU-RRH association in the C-RAN architecture. For example, [29] considers the joint problem of beamforming vector optimization [39] and BBU-RRH association. The authors decompose the joint problem into two subproblems. The first is solved via the weighted minimum mean square error, and derives what the authors call the “RRH-user equipments clusters”, defined as the subsets of RRHs serving jointly a set of UEs. The second subproblem is formulated as a Bin Packing optimization, where the “RRH-user equipments clusters” are packed into BBUs of limited capacity in terms of number of UEs. A QoS-aware formulation is considered in [40]. In particular, the authors target a UE association and a BBU-RRH association problems jointly. A first stage decides the UE association scheme by resorting to Lagrangien Relaxation [41]. The second stage assigns RRHs to BBUs, and formulates the problem via a Bin Packing optimization, solved by the BFD algorithm. The QoS is modeled in terms of waiting delay. A joint UE association and BBU-RRH association problem is also tackled in [42], where the authors consider a heterogeneous RAN context. First, they propose a FFD algorithm to map RRHs to BBUs. Then, they consider a greedy heuristic that aims to deactivate lightly loaded RRHs in a way to switch UEs to different cells, while guaranteeing a minimum level of QoS. Within the second stage, the authors first consider the small cell RRHs to be switched OFF. The authors in [43] consider a joint problem of resource allocation and BBU-RRH association. They also decompose their problem into two subproblems. To solve the resource allocation problem (*i.e*, the RBs assignment), they resort to the Binary Integer Linear programing (BIL) and they propose a heuristic to solve it. They model the second stage (*i.e*, the BBU-RRH association) as a Bin Packing approach and they resort to the FFD algorithm to solve it. The two subproblems are executed in a sequential manner. In their work, the authors do not take into consideration the interference level in the resource allocation stage (*i.e*, the first stage). A joint bandwidth/power allocation and BBU-RRH assignment problem is formulated in [44], where the authors resort to convex optimization [45] to solve the UE bandwidth and power allocation subproblem, and to a Bin Packing formulation to associate RRHs to BBUs. They consider BBUs of fixed capacities in terms of number of UEs and resort to the BFD algorithm to solve the subproblem.

2.4.2 The Knapsack Approach

Few works consider the formulation of the BBU-RRH association as a Knapsack problem [46]. A Knapsack problem, consists in packing the most valuable objects into a finite set of bins with fixed capacity. Thus, each object is represented by a weight reflecting its value.

The Knapsack formulation is adopted in [47], where the authors formulate a joint problem of resource/power allocation and BBU-RRH assignment. The number of bins (*i.e*, BBUs) is determined at the first stage of resource allocation, through dividing the total number of RBs that satisfies all UEs' demands over the available number of RBs in the system (*i.e*, the number of RBs composing the total bandwidth). A RRH's value would be its load demand. The second stage aims to associate the maximum number of highly loaded RRHs (*i.e*, the highest values) to the available set of active BBUs. In the same line, the authors in [48] propose a simulated annealing heuristic for the radio resource and power allocation subproblem, from which they predict the needed number of active BBUs, and resort to a Knapsack formulation to associate the most valuable RRHs (*i.e*, the most highly loaded RRHs) to

the decided set of active BBUs. A different context of the problem is investigated in [49], where the authors explore the benefits brought by the cooperation among multiple MNOs and further, propose to reduce the number of active BBUs by solving the Knapsack problem. The set of active BBUs is deduced as in [47] and [48], by dividing the total demand on RBs over the total available RBs in the system.

2.4.3 The Graph Based Approach

Few works consider a graph based formulation approach [50]. RRHs are modeled as vertices, and when two RRHs form one vertex, they are associated to the same BBU. Neighboring relationships between RRHs are modeled as edges. Thus, two neighboring RRHs are linked by an edge.
The work in [51] propose a formulation by a graph coloring method. Each RRH is represented as a vertex. Assigning a color to a vertex is attributing it a part of the bandwidth according to its traffic demand. Two adjacent vertices are not attributed the same color so as to mitigate the interference among them. The RRHs assigned to orthogonal parts of the bandwidth are associated to a single BBU. The results of the proposed scheme show improvement in terms of energy efficiency in comparison to the reuse-1 and to the FFR [52] schemes. The authors in [53], also adopts a graph based formulation, where RRHs are vertices that are connected to each other through edges, reflecting their neighborhood relationship. RRHs linked through edges are associated to the same BBU. The process is repeated until the resources of one BBU are entirely consumed. In [2], the authors propose to model the RRHs as vertices that are linked together through edges, whose weights represent the spare amount of radio resources of two linked vertices. The clustering algorithm chooses randomly a vertex and clusters it with the neighboring one whose edge shows the smallest weight. The algorithm is applied on the graph until all edges have infinite weights. In other terms, when the needed amount of resources to support each of the two vertices exceeds the available resources offered by one BBU. After clustering two vertices together, the graph is updated with its different vertices, weights and edges. In [54], the authors propose a clustering algorithm to reduce the HO rate of UEs between cells. Neighboring RRHs are connected to each other through edges, whose weights represent the amount of UEs performing HOs between different cells. The problem is translated to a community detection problem [55], where the objective is to find the cluster of nodes strongly related to each other through high weight edges.

2.4.4 The Evolutionary Algorithms Approach

Evolutionary algorithms are considered as meta-heuristics, aiming to solve optimization problems by running an iterative sequence of searches to exploit multiple subsets of solutions until convergence is reached. Candidate solutions are the ones that improve a fitness function that quantifies the value of a solution.
In [30], the authors suggest a QoS-aware mapping between BBUs and RRHs. Based on a Key Performance Indicator (KPI) they define, the authors propose an optimization problem aiming to maximize the defined metric. In other terms, the authors propose to reduce the number of blocked UEs, exceeding the BBU capacity expressed as the maximum number of blocked UEs. They resort to an evolutionary algorithm (*i.e*, the Genetic Algorithm [56]) to solve the problem. Again in [57] the authors propose a multi-objective optimization problem based on multiple KPIs, such as the number of blocked UEs and HO rates. The problem is subjected to the hard capacity of a BBU and to the neighborhood relationship between the RRHs. The algorithm is solved via two different evolutionary algorithms (*i.e*, the Genetic Algorithm and the Discrete Particle Swarm Optimization). A comparison between the two schemes is provided and shows that the Discrete Particle Swarm Optimization outperforms the Genetic Algorithm in the case of large network size. Dynamics schemes are proposed in [22] and [58], where the authors propose a multi-objective optimization problem, meeting many criteria based on fairness maximization, throughput enhancement and HO rate minimization. The authors resort to the Genetic Algorithm and the Discrete Particle Swarm Optimization to solve their problem.

2.4.5 A comparison between Existing Approaches

In fact, the existing approaches of the state-of-art present several advantages and disadvantages :

— **The Bin Packing Approach**: The Bin Packing approach remains the most efficient in reducing the number of active BBUs among all existing approaches, since its main objective would be to find the partitions of RRHs that reduce the number of used bins. Further, since the Bin Packing is constrained by a fixed capacity (*i.e*, the BBU capacity), the QoS is guaranteed, leading to enhanced energy efficiency. However, the Bin Packing approach is a combinatorial problem, and is of high computational requirement. Additionally, it is a centralized

approach, requiring high level of signaling overhead exchange, and is unaware of the level of interference in the network, which is reduced when several RRHs are assigned to one BBU, as will be explained in section 2.5. Also, the Bin Packing does not take into consideration the re-association of RRHs to BBUs when the dynamic aspect of the problem is treated, leading to high rate of HOs (cf. Chapter 6).

— **The Knapsack Approach**: The Knapsack problem provides a good estimation of the needed number of serving BBUs. However, its main objective remains to pack the most valuable objects (*i.e*, the highly loaded RRHs) into a predefined number of bins. In fact, this might violate the QoS for UEs connected to lowly loaded RRHs. Further, the Knapsack approach is a combinatorial problem, requiring high complexity. In addition, it relies on a centralized scheme, which increases the signaling load overhead requirement, and has an Interference-Oblivious behavior. Also, it does not take into account the re-association rate of UEs, increasing the HOs rate in the network.

— **The Graph Based Approach**: The graph based approach has the advantage of taking into consideration the neighborhood relationship among RRHs. In fact, when neighboring RRHs are associated to the same BBUs, the spectral efficiency is enhanced since the interference level decreases. Additionally, this reduces the HO rate, since mobile UEs will be connected to the same logical cell (*i.e*, BBU) while moving between the covered regions of neighboring RRHs. However, the neighborhood constraint might cause several disadvantages, such as violating the QoS. For example, neighboring RRHs might be clustered together in order to reduce the HO rate. However, this will cause the BBU to be overloaded and to degrade the throughput levels of UEs. Further, taking into consideration that constraint (*i.e*, neighborhood constraint), the number of active BBUs might not be necessarily reduced. Additionally, this approach is of high computational and signaling load requirement, and does not take into account the re-association rate of RRHs when dynamic load conditions occur.

— **The Evolutionary Algorithms Approach**: The Evolutionary algorithms remain the less in computational requirement since they are meta-heuristics. However, they do not necessarily provide optimal solutions and remain Interference-Oblivious approaches, while requiring huge amount of signaling load overhead and neglecting the re-association rate of UEs.

Table 2.1 provides a summary of the different advantages and disadvantages for each approach.
The Bin Packing approach is taken as a reference model to compare our devised schemes since its main objective is to reduce the number of active BBUs, while guaranteeing the QoS provided to UEs. In fact, as will be detailed in Chapter 3 our proposed approach relies on minimizing the power consumption, subject to a minimum throughput guarantee per UE, which mimics the behavior of the Bin Packing problem.

2.5 Incentives for a New Approach

A key drive to exploit the benefits of flexible associations between BBUs and RRHs, relies on designing efficient schemes. Thus, the close examination of the RRHs behavior, associated to one BBU, is crucial to attain the objective. The DAS is a behavior introduced by associating a number of RRHs to one BBU, allowing the change of cell boundaries within the network, since geographically distributed antenna elements (*i.e*, RRHs) are considered as a single cell. Changing the cell boundaries within the network contributes to reducing the inter-cell interference because the RRHs share the same pool of radio resources that of the BBU. However, this might limit the resources under high network load conditions.

2.5.1 RRHs Behavior While Associated to One BBU: A Distributed Antenna System Behavior

As previously mentioned, the C-RAN architecture breaks down the classical concept of a base station composed of one RRH and one BBU. The role of a BBU is to perform the MAC and physical layers functionalities. The data of the associated UEs is scheduled through a frame that is transmitted over the cluster of RRHs associated to one BBU. UEs' data are consequently carried through different RBs of one frame in an orthogonal fashion [59]. In particular, the data are multiplexed over orthogonal subcarriers, composing the bandwidth of one cell. Despite an overlap in the sidebands of each subcarrier, they can still be received without any interference. Figure 2.4 illustrates the subcarriers spacing while being multiplexed into a OFDMA symbol.

A cluster of RRHs, sharing the total bandwidth of the BBU, is free of intra-cluster interference and is considered as a single cell equipped with a DAS system. Hence, the spectral efficiency of a BBU is enhanced and more UEs are

	Advantages	Disadvantages
The Bin Packing Approach	— Reduces the number of active BBUs — Provides optimal solutions — Enhances the energy efficiency — Guarantees the QoS	— High computational requirement — Requires high amount of signaling load overhead — Interference-Oblivious approach — Does not take into consideration the re-associations of RRHs
The Knapsack Approach	— Provides good estimation for the number of active BBUs — enhances the energy efficiency	— High computational requirement — Might violate the QoS — Requires high amount of signaling load overhead — Interference-Oblivious approach — Does not take into consideration the re-association of RRHs
The Graph Based Approach	— Enhances the spectral efficiency — Takes into consideration the neighborhood relationship between RRHs — Reduces the HOs rate	— Does not always guarantee the QoS — High computational requirement — Does not necessarily reduce the number of active BBUs — Requires high amount of signaling load overhead — Does not take into consideration the re-association of RRHs

The Evolutionary Algorithms Approach	— Lower computational requirement	— Meta-heuristic and does not provide optimal solutions — Interference-Oblivious Approach — Requires high amount of signaling load overhead — Does not take into consideration the re-association of RRHs

TABLE 2.1: Advantages and Disadvantages of the Different Approaches

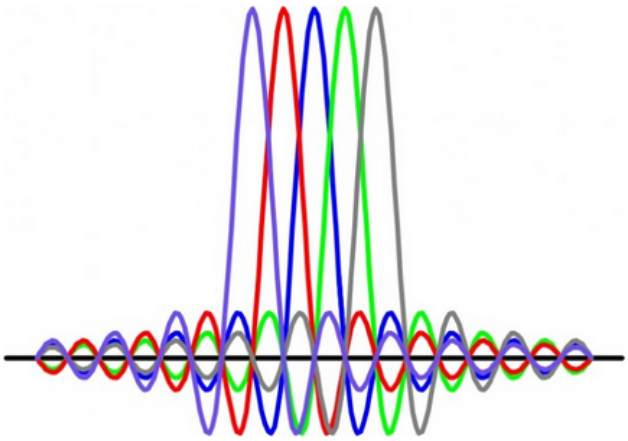
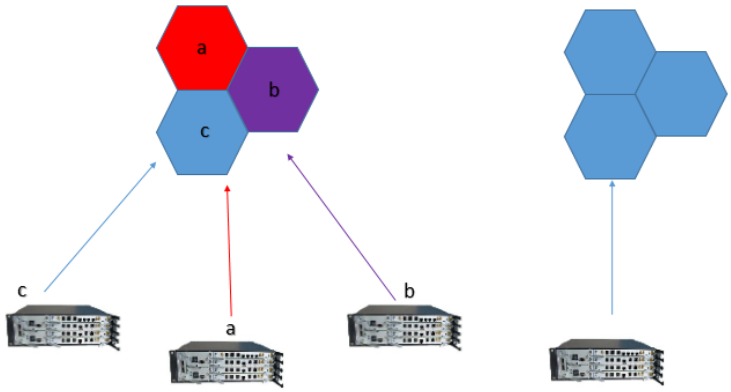


FIGURE 2.4: OFDMA Symbol

served. Figure 2.5 illustrates two examples of associations between BBUs and RRHs: 2.5(a) displays three RRHs, denoted as ‘a’, ‘b’ and ‘c’ respectively connected to BBUs ‘a’, ‘b’ and ‘c’. Three different frames are generated from the BBUs to serve UEs connected to RRHs. Assuming a reuse-1 scheme, where the operating bandwidth of one BBU is entirely reused within a different BBU, UEs connected to RRH ‘a’ experience interference from RRHs b and c. Same for UEs connected to RRHs b and c. Such inter-cell interferences degrade the level of the BBU spectral efficiency. Contrarily, Figure 2.5(b) displays three RRHs associated to one BBU: A single frame is generated from the BBU, and is transmitted over the cluster of RRHs. UEs connected to those RRHs do not endure the harmful intra-cluster interference. However, they share the limited resources of one BBU.

The DAS behavior of RRHs while being associated to the same BBU has not been investigated in the state-of-art, although many interesting formulations has been proposed. A common point among them is to consider a fixed capacity for the BBU. Few works, such as [60], [57] and [30], propose to associate neighboring RRHs to one BBU, so that interference mitigation techniques, such as CoMP, would be easier to implement among the cluster of RRHs. However, those works also consider a fixed BBU capacity, which is not the case when some BBUs are turned OFF and the interference level changes. As the Bin Packing approach is the most commonly used and is taken as a reference to compare our devised schemes, we review hereafter its formulation, followed by a conducted conclusion.



(a) Inter-Cell Interference (b) Intra-Cell Interference Cancellation

FIGURE 2.5

Bin Packing Optimization Formulation

Given a BBU B of capacity C, and a set of n RRHs $\{r_1, ..., r_n\}$, each with a demand $\{d_1, ..., d_n\}$ respectively, the problem consists in finding a number of BBUs S and an S-partition $\{Cl_1, ..., Cl_n\}$ of the set $\{1, ..., n\}$ such that $\sum_{i \in Cl_k} d_i \leq C$ for all $k = 1, ..., S$. A solution is optimal when we have a minimal S.

The ILP formulation of the problem is given as follows:

$$\textit{Minimize } S = \sum_{i=1}^{n} y_i \tag{2.1}$$

Subject to:

$$\sum_{i=1}^{n} d_j x_{ij} \leq C y_i, \forall i \in \{1, ..., n\}, \tag{2.2}$$

$$\sum_{i=1}^{n} x_{ij} = 1, \forall j \in \{1, ..., n\}, \tag{2.3}$$

$$y_i \in \{0, 1\}, \forall i \in \{1, ..., n\}, \tag{2.4}$$

$$x_{ij} \in \{0, 1\}, \forall i \in \{1, ..., n\}, \forall j \in \{1, ..., n\}, \tag{2.5}$$

where, $y_i = 1$ if BBU i is used and $x_{ij} = 1$ if r_j is associated to BBU i. The first inequality (2.2) expresses the BBU capacity constraints, (2.3) expresses that one RRH is associated to a single BBU.

Although such formulation has shown efficient results, it is clear that the BBU capacity is a fixed metric. Thus, this approach is *Interference-Oblivious*, and does not take into account the impact of the interference level variation that directly influences the BBU spectral efficiency.

A need for An Interference-Aware Approach

Enhancing the spectral efficiency of a BBU is being aware of the interference variation in the system. Taking into consideration the DAS behavior of the RRHs, a BBU capacity should not be expressed as a fixed metric such as the amount of supported UEs, the number of RBs or the baseband tasks. As this metric changes as a function of the interference level, a need for a relevant design must be targeted. This design should be performed carefully so as to optimize the resource allocation, taking into account the load factor of the served cells and the intra-cluster interference cancellation among RRHs associated with the same BBU. Further, setting the number of available BBUs is paramount. Increasing the number of BBUs hinders the energy efficiency and increases the interference level. Conversely, reducing the number of available BBUs leads to resource shortage.

2.5.2 Conclusion

In this chapter, we have presented the BBU-RRH association problem with its distinct formulations. Further, we have detailed the DAS behavior of the RRHs while being associated with one BBU, and which has not been taken into consideration in the existing works. Also, a need for a new approach of the problem is put forward. In the following chapter, an Interference-Aware formulation is provided. The Interference-Aware behavior of the algorithm is closely examined and compared to the Bin Packing approach.

Chapitre 3

An Interference-Aware Approach for BBU-RRH Association Optimization

In this chapter, the BBU-RRH association problem is addressed. Different from the previous works, our proposition has the originality to consider inter-cluster interference when forming RRH clusters. The BBU-RRH association problem is formulated as a SPP. The objective is to optimize network power consumption with minimum throughput requirements. Decisions are aware of the network interference level and spectral efficiency thus, leading to optimized BBUs assignment. The results are closely examined and the Interference-Aware behavior of the algorithm is put forward. In particular, to closely examine the behavior of our proposed scheme, we consider three different scenarios: The first one relies on a uniform UEs distribution per cell, leading to average radio conditions. The second one considers good radio conditions, where UEs are concentrated into zones close to the cell center. The third one considers bad radio conditions, where UEs are concentrated into zones close to the cell edge. Further, a comparison with the Bin Packing approach is provided to highlight the need of the Interference-Aware behavior. As the SPP belongs to the NP-complete category, a greedy heuristic is proposed and performance are evaluated through extensive simulations.

3.1 System Model

3.1.1 Network Topology

Our network topology consists of a cellular network composed of R RRHs. We define $\mathcal{R}$ as the set of RRHs, it is expressed as $\mathcal{R} = \{1, ..., r, ...R\}$. We also define $\mathcal{C}$ as the set of clusters, it is expressed as $\mathcal{C} = \{1, ..., c, ...C\}$. The set of clusters to choose should form a partition $\mathcal{P}$ of $\mathcal{R}$. $\mathcal{P}$ defines a partition of $\mathcal{R}$ if $c \cap c' = \emptyset \ \forall c, c' \in \mathcal{P}, c \neq c'$ and $\bigcup_{c \in \mathcal{P}} c = \mathcal{R}$. We also consider the binary variable $y_{r,c}$ defined as:

$$\begin{cases} y_{r,c} = 1 & \text{If RRH } r \text{ is associated to BBU } c \\ y_{r,c} = 0 & \text{Otherwise} \end{cases} \tag{3.1}$$

Measurement Consideration

Before considering the Interference-Aware proposition, it is worth defining the time scale during which the devised scheme is executed: The BBU-RRH association problem requires a time scale greater than milliseconds and even seconds. In fact, any change in the BBU-RRH associations requires a reconfiguration of the optical fronthaul connection which is very costly. In addition, the traffic pattern presents variations on the scale of minutes [61], [62]. Thus, average metrics are considered instead of instantaneous measurements. For that purpose, we decompose the network area into several discrete zones, and we consider the average number of UEs and radio conditions per zone.

Area Decomposition

Our system model relies on the classical meshing method [63]. In particular, the covered area is meshed into discrete zones. For illustration, Figure 3.1 displays a set of seven cells, quadratically meshed into multiple discrete zones.

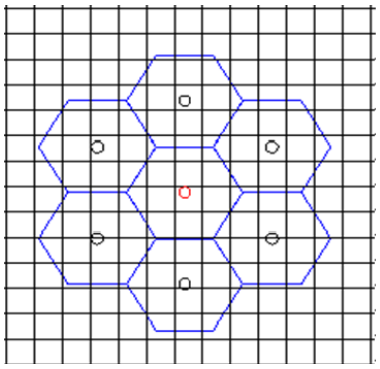


FIGURE 3.1: Meshed Area

Each zone is assimilated to a single test point from which the cumulative rate of UEs is generated. We denote by $n_{z,r}$ the average number of UEs located in a zone of index z, associated to a RRH of index r. We also define the average Signal-to-Interference-plus-Noise ratio of zone z, associated to RRH r, as $SINR_{z,r}$. To calculate the $SINR_{z,r}$ ratio, the average distances between the considered zone and all the cells within the network are calculated.
In paragraph 3.4.1, we evaluate many metrics as a function of different meshing steps values in order to assess its impact on the results.

Average Distance Calculation

Considering a quadratic meshing scheme, as in Figure 3.1, the average distance (D_{avr}) between a RRH of coordinates (x_0, y_0) and all points of coordinates (x, y) belonging to zone z is expressed as follows:

$$D_{avr} = \frac{1}{a^2} \iint_S \sqrt{(x - x_0)^2 + (y - y_0)^2} dxdy, \tag{3.2}$$

where a denotes the square side length (*i.e*, the meshing step) and S the square surface.

User Association

We assume that all UEs located in a zone z are associated to RRH r when the average distance D_{avr} between z and r is the smallest in comparison with the different distances between zone z and RRHs $r' \neq r$ in the network. In addition, D_{avr} must follow the following constraint: $D_{avr} \leq D_r$, where D_r denotes the cell radius of the RRH.

3.1.2 Employed CoMP Technique Within One BBU

As a BBU controls multiple antennas (*i.e*, RRHs), CoMP techniques are inevitable so that the BBU manages its UEs connected to its associated RRHs. Many CoMP techniques have been investigated in the state-of-art and are classified into three categories according to 3GPP [64] :

Joint Transmission

The Joint Transmission technique (JT) consists in transmitting the data of one UE simultaneously from a group of RRHs, so that the harmful interference will turn to be a constructive one. The transmission could be coherent or non-coherent. The coherent-JT necessitates a perfect synchronization between RRHs in order to realize the full advantage of it. It is based on spacial CSI updates, corresponding to a set of candidate RRHs that necessitate

additional constraint over the fronthaul connection between the BBUs and RRHs due to the limited bandwidth resources of the optical connections. Consequently, the coherent-JT requires a compromise between the realized gains and the additional overheads [65], along with the synchronization requirements. Non-coherent JT requires techniques such as Cyclic Delay Diversity (CDD) [66] to combat inter-symbol interference (ISI) due to multipath fading. Generally, JT techniques necessitate additional power requirements on the RRH side, since simultaneous transmissions are applied for UEs. This might be hefty, in particular in high load conditions, where the energy resources of one RRH is divided over many subcarriers to transmit UEs' data.

Coordinated Beamforming

By applying the Coordinated Beamforming (CB) technique, the interfering points (*i.e*, RRHs) select their beamforming weights, in order to shelter UEs from their harmful interference. This technique necessitates the exchange of CSI between a group of interfering RRHs, requiring additional constraint in terms of bandwidth over the optical connections between BBUs and RRHs. Further, recent studies, such as [67] and [68], consider the joint problem of Coordinated Beamforming and Coordinated Scheduling (CB/CS), in which scheduling is performed within each selected group of RRHs in order to determine which RRH transmits to which UE and in which time slot [17]. This technique significantly increases the complexity in terms of CSI exchange and overhead, although it increases the throughput and enhances UE SINR by mitigating interference.

Transmission Point Selection

The Transmission Point Selection (TPS) technique consists in transmitting the data of a UE by one of the selected RRHs, belonging to a set of cooperating ones. In the TPS technique, a UE sends the index of its preferred RRH (*i.e*, the RRH with the highest SINR) with its corresponding CSI. Only one RRH transmits at a time, so that the SINR of a UE is improved. The selected RRH might dynamically change from one subframe (*i.e*, 1 ms) to another. However, when CoMP techniques are employed among RRHs associated to the same cell (*i.e*, BBU), long term measurements are taken into account and help in determining the long term set of cooperating RRHs [17]. The TPS technique is the less to require the exchange of informations in terms of CSI, since one UE reports one single RRH. In addition, it does not require additional constraint in terms of power requirements on the RRH side, since a UE gets its data from one serving RRH at a time.

In our system model, the TPS technique is employed among a cluster of RRHs associated to a single BBU because of two advantages: Less power consumption on RRHs side at one hand ; and less CSI exchange between BBUs and RRHs, on the other hand. Resorting to TPS techniques frees our model from the limited bandwidth constraint of the optical fronthaul connections, since less CSI signaling is needed. Further, associating several RRHs to one BBU allows long term measurements because a BBU locally controls the RRHs and performs average metrics calculations, which reduces instantaneous CSI exchange.

3.1.3 Cluster Rate Calculation of One BBU

To calculate a cluster rate (*i.e*, BBU rate), we start by expressing the $SINR$ ratio of a given zone z, associated to RRH r. A RRH should be served by a BBU (*i.e*, cluster) of index c. Thus, the $SINR$ is denoted as $SINR_{z,r,c}$. It is expressed as follows:

$$SINR_{z,r,c} = \frac{P_r \cdot G_{z,r}}{N_0 + \sum\limits_{\substack{r' \neq r \\ r' \in c' \neq c}} P_{r'}.Gz,r'}, \tag{3.3}$$

where

P_r: The power emitted by the serving RRH r
$G_{z,r}$: The channel gain between z and r
N_0: The thermal noise power
$P_{r'}$: The power emitted by any RRH $r' \neq r$
$G_{z,r'}$: The channel gain between z and r'
c: The BBU that serves r
c': The BBU that serves r'

As previously detailed in chapter 2, the interference is canceled among RRHs served by the same BBU c, and act as a DAS system. This behavior is put forward within the $SINR$ expression. In particular, only RRHs r' associated to different clusters $c' \neq c$ inflect harmful interference upon a zone of index z, associated to RRH r that is mapped to BBU c. As the TPS is assumed to be implemented within each BBU, the useful signal is received from one single RRH r at a time.

We denote by $\chi_{z,r,c}$ the peak throughput per UE perceived by one UE located in zone z. It is expressed as follows:

$$\chi_{z,r,c} = BW \cdot \log_2\big(1 + SINR_{z,r,c}\big), \tag{3.4}$$

where BW is the total bandwidth of a BBU. We note that the throughput perceived by a UE is at its peak value when only one UE consumes the resources of a BBU.

Considering a fair resource sharing model among n_c UEs served by BBU c, and a full buffer traffic model, a UE located in zone z perceives an average throughput of $\frac{1}{n_c} \cdot \chi_{z,r,c}$. Consequently, when $n_{z,r}$ UEs (among n_c) are located in zone z, their cumulative throughput would be expressed as:

$$T\big(z,r,c\big) = \frac{n_{z,r}}{n_c} \cdot \chi_{z,r,c} \tag{3.5}$$

The cumulative throughput of RRH r, associated to BBU c, is expressed as:

$$T\big(r,c\big) = \sum_{z=1}^{Z} T\big(z,r,c\big), \tag{3.6}$$

where Z is the total number of discrete zones associated to RRH r.

The cumulative throughput of BBU c would be expressed as:

$$T\big(c\big) = \sum_{r=1}^{R} y_{r,c} \cdot T\big(r,c\big), \tag{3.7}$$

where R denotes the total number of RRHs within the network, and $y_{r,c}$ is the binary variable defined in (3.1).

3.1.4 BBU Power Consumption Model

The power consumption of a BBU is assumed to be a linear function of the cumulative rate of BBU c [69]. It is expressed as:

$$P\big(c\big) = A + B.T\big(c\big), \tag{3.8}$$

where

$P\big(c\big)$: The power consumed by an active BBU c
A: The minimum power consumption of an active BBU at 0 load
B: The coefficient variation of $P\big(c\big)$ as a function of $T\big(c\big)$
$T\big(c\big)$: The cumulative throughput of BBU c

3.2 Set Partitioning Problem Formulation

Our problem consists in choosing the subsets (*i.e*, clusters) of RRHs served by one BBU. The chosen subsets should define a partition $\mathcal{P}$ of the set $\mathcal{R}$ of RRHs at one hand, and should minimize a total cost function defined as the sum of costs of each chosen subset, on the other hand. A cost function of a subset c is defined as the power consumption of a BBU as in (3.8). Thus, the total cost S of a partition $\mathcal{P}$ would be expressed as follows:

$$S = \sum_{c \in \mathcal{P}} Cost\big(c\big) = \sum_{c \in \mathcal{P}} P\big(c\big) \tag{3.9}$$

We define a binary variable x_c, expressed as follows:

$$\begin{cases} x_c = 1 & \text{If the subset (\textit{i.e}, BBU) } c \text{ is chosen} \\ x_c = 0 & \text{Otherwise} \end{cases} \tag{3.10}$$

The optimization problem aiming to minimize the total cost function (3.9) is given as follows:

$$\textit{Minimize } S = \sum_{c \in \mathcal{C}} x_c.Cost(c)$$

Subject to:

$$\sum_{c \in \mathcal{C}} x_c.y_{r,c} = 1, \forall r \in \mathcal{R} \tag{3.11}$$

$$x_c.T(c) \geq n_c.D_{min}, \forall c \in \mathcal{P} \tag{3.12}$$

$$x_c \in \{0,1\}, \forall c \in \mathcal{C} \tag{3.13}$$

$$y_{r,c} \in \{0,1\}, \forall r \in \mathcal{R}, \forall c \in \mathcal{C} \tag{3.14}$$

Constraints (3.11) state that each RRH is mapped to a single BBU c. Constraints (3.12) state that each cluster c should at least realize a throughput per UE equal to D_{min}. Without constraints (3.12), the problem would be a classical SPP.

The variable to be optimized is $x_c \in \{0,1\}$. Thus, the problem belongs to the Integer Linear Programing (ILP) class. Further, since the problem is presented as a SPP, it belongs to the NP-complete category [19]. Thus, when the entry parameters, which in our case is the number of RRHs, significantly increase, the optimal solution would be practically intractable. For the sake of complexity reduction, we propose in the following an efficient greedy heuristic.

3.3 Greedy Heuristic

A greedy heuristic consists of several steps. Within each step, the algorithm chooses a local optimum. Such algorithms perform to reduce the complexity of NP-complete problems and the gap with optimal solutions.
In our case, the objective is to minimize the power consumption by clustering multiple RRHs into a single BBU. Our proposed heuristic consists in choosing, at each stage, the RRH that minimizes the power consumption of a cluster. In other words, at each stage, the local optimum in terms of power consumption is chosen.
The heuristic starts by initializing a set of candidate RRHs, denoted by $\mathcal{R}_{cand}$, to the set $\mathcal{R}$. $\mathcal{R}_{cand}$ represents the RRHs that are still unmapped to a given BBU. After the algorithm finishes its execution, $\mathcal{R}_{cand}$ is equal to $\emptyset$.

The heuristic consists of the following steps:

— Choose a random RRH r in the set $\mathcal{R}_{cand}$ and form the set $\mathcal{V}$ that is equal to $\mathcal{R}_{cand} - \{r\}$
— Choose a RRH $r' \in \mathcal{V}$ such that $Cost(\{r, r'\})$ is minimal and respects constraint (3.12), then update $\mathcal{V}$ to $\mathcal{V} - \{r'\}$
— Form the cluster $c = \{r, r', ...\}$ by adding more RRHs from $\mathcal{V}$ that minimize $Cost(c)$ and respect (3.12). Update $\mathcal{V}$ each time a RRH is added
— Stop when $\mathcal{V} = \emptyset$; in other terms, when there are no more unmapped RRHs that minimize $Cost(c)$ and respect (3.12)
— Update R_{cand} to $R_{cand} - c$ and choose randomly a RRH in R_{cand} to form a new cluster c'
— Stop when $R_{cand} = \emptyset$

Algorithm 1 describes the proposed heuristic.

Heuristic Complexity Analysis

In order to estimate the heuristic complexity, we examine its executed steps in extreme load scenarios (*i.e*, extremely low load and high load conditions).

```
Initialize R_cand = R;
while R_cand ≠ ∅ do
    Choose randomly r ∈ R_cand;
    Set c = {r};
    Set V = R_cand − {r};
    if V ≠ ∅ then
        Choose r' ∈ V such as Cost(c ∪ {r'}) is minimal;
        if (3.12) is satisfied then
            c ← c ∪ {r'};
            V ← V − {r'};
            Go to step 6;
            else
                V ← V − {r'};
                Go to step 6;
            end
        end
        else
            R_cand ← R_cand − c;
            Go to step 2;
        end
    end
end
```

Algorithm 1: Heuristic algorithm

— *Low load conditions*: At extremely low load conditions, associating all RRHs to one BBU would be the optimal mapping. The heuristic starts by randomly choosing a RRH r. It then searches among the $R-1$ RRHs for a RRH r', minimizing the total power consumption of cluster $c = \{r, r'\}$, and respecting a minimum level of QoS in c. In that stage, the total number of operations executed by the algorithm would be equal to $1+(R-1)$, since the random drop is executed within one operation, and the search for RRH r' is performed in $R-1$ operations. The heuristic continues the search among the $R-2$ RRHs to associate a RRH r'' to cluster c. When all RRHs in $\mathcal{R}_{cand}$ are mapped to c, the total number of operations executed by the heuristic would be equal to $1+(R-1)+(R-2)+...+1$, which is equal to $1+\frac{R\cdot(R-1)}{2}$. For high values of R, the complexity would be quadratic ($O(R^2)$).

— *High load conditions*: Under extremely high load conditions, the (1 :1) mapping scheme is the optimal association. In that case, the heuristic performs for each RRH a random drop, and tries to associate it with one of the unmapped RRHs in the set $\mathcal{V}$. Thus, the total number of executed operations is equal $1+(R-1)+1+(R-2)+...+1+(R-(R-1))$, which is equal to $(R-1)+\frac{R\cdot(R-1)}{2}$, equal to $\frac{(R-1)\cdot(R+2)}{2}$. Thus, for high values of R, the complexity would be quadratic, and the random drop of each RRH would be negligible.

3.4 Performance Evaluation

We evaluate the results of our proposed solutions according to different scenarios. In particular, the results are assessed according to different UEs distribution (*i.e*, different radio conditions). Different metrics are also evaluated as a function of the load conditions, and are listed below:

— The number of active BBUs
— The total power consumption derived in the system: $\sum_{c\in\mathcal{P}} P(c)$
— The realized power saving in the system, given by:

$$\frac{P(1:1)-\sum_{c\in\mathcal{P}} P(c)}{P(1:1)}, \tag{3.15}$$

where $P(1:1)$ reflects the power consumption in the conventional architecture (*i.e*, the (1 :1) association scheme)

— The energy efficiency, which is expressed as the total throughput in the system over the total power consumption, is given by:

$$\frac{\sum_{c\in\mathcal{P}} T(c)}{\sum_{c\in\mathcal{P}} P(c)} \tag{3.16}$$

— The mean throughput per UE, given by:

$$\frac{\sum_{c\in\mathcal{P}} T(c)}{\sum_{c\in\mathcal{P}} n_c}, \tag{3.17}$$

where $\sum_{c\in\mathcal{P}} n_c$ is the total number of UEs in the system

— The spectral efficiency per BBU, given by:

$$\frac{\sum_{c\in\mathcal{P}} T(c)}{BW \cdot N_{BBU}}, \tag{3.18}$$

where BW is the total bandwidth and N_{BBU} is the number of active BBUs

— The total derived throughput, given by:

$$\sum_{c\in\mathcal{P}} T(c) \tag{3.19}$$

3.4.1 Simulation Setup

We use matlab for simulation, and we display the results of a network topology consisting in seven cells in order to assess the performance of the optimal solution. We also evaluate the results for three different UEs distribution:

— *Uniform UE distribution*: In this case, the number of UEs per zone is the same.

— *Good radio conditions distribution*: In this case, UEs are concentrated into zones with good radio conditions (*i.e*, zones that are close to the cell center).

— *Bad radio conditions distribution*: UEs are concentrated into zones with bad radio conditions (*i.e*, zones that are close to the cell edge).

Meshing Step Evaluation We are interested in evaluating the total derived throughput and the total power consumption as a function of different values of meshing steps. A medium load condition is considered (*i.e*, the number of UEs per cell is 6), along with a uniform UE distribution in the network. Evaluations are listed in table 3.1. Further, we list the obtained number of active BBUs, at medium load conditions and uniform distribution of UEs for many meshing step values (cf. table 3.2).

TABLE 3.1: Meshing Step Evaluation: Total Throughput, Power Consumption

Meshing Step (m)	Total Throughput (Mb/s)	Power Consumption (W)
50	119.11	221.46
100	116.45	219.87
150	116.43	219.86
200	114.63	218.78
210	113.59	218.15
220	112.87	217.72
230	111.26	216.76
240	115.25	219.15
250	110.11	216.07

We notice that the range of difference between the maximum and minimum values of throughput is 7.55%, whereas it is equal to 2.43% in the case of power consumption. Further, the number of active BBUs remains 3 for all meshing step values, indicating that the considered range of values does not significantly impact the results of the optimization problem showing the same number of active BBUs for all values. In our assessments, a meshing step of 240 m will be considered to evaluate metrics. Additionally, the minimum throughput per UE threshold is set to 2 Mb/s (*i.e*, D_{min}), and a full buffer traffic scheme is considered. The propagation model employed in our case is the Cost Hata 231 for urban areas, and a frequency reuse-1 scheme is taken into account. The simulation parameters are given in table 3.3.

We first start by providing the results of our proposed approach in comparison with the Bin Packing scheme, the classical formulation of the state-of-art. Further, the results of the conventional scheme are also displayed.

TABLE 3.2: Meshing Step Evaluation: Number of Active BBUs

Meshing Step (m)	Number of Active BBUs
50	3
100	3
150	3
200	3
210	3
220	3
230	3
240	3
250	3

TABLE 3.3: Simulation Parameters

Parameter	Value
Traffic Model	Full Buffer
Scheduling Scheme	Fair Resource Sharing
Propagation Model	Cost Hata 231
Shadowing Standard Deviation	10 dB
Transmit Power of RRH	40 dBm
Thermal Noise Power	-174 dBm/Hz
BW	10 MHz
Cell Radius	500 m
A	50 W
B	0.6
D_{min}	2 Mb/s
Meshing Step (a)	240 m

3.4.2 Interference-Aware Algorithm vs. Bin Packing vs. No Clustering

Uniform UEs Distribution

In order to guarantee 2 Mb/s for all UEs, we assume that the system is performing Admission Control. The results show a maximum number of 12 UEs per BBU, (*i.e*, 84 UEs in the network) being able to pull 2 Mb/s when 7 BBUs are activated. Thus, the BBU capacity, in the Bin Packing case, is fixed to 12 UEs. The following results show the performance of the optimal solution of the problem (cf. section 3.2), in comparison with different schemes. A uniform UEs distribution is considered, leading to average radio conditions.

Figure 3.2 shows the number of active BBUs as a function of the load conditions (*i.e*, the number of UEs per cell). Our proposed algorithm is denoted as Interference-Aware Clustering Algorithm (IACA), the Bin Packing approach is denoted as Bin Packing and the conventional architecture as No Clustering scheme. For 1 UE per cell, the IACA scheme and the Bin Packing activate 1 BBU. When the number of UEs per cell is more than 1, the Bin Packing activates more BBUs, while the IACA maintains 1 activated BBU. The reason is that our proposed approach is considering the spectral efficiency of a BBU, which is significantly high when all RRHs are associated to a single BBU, and the cluster is free of intra-cluster interference. Contrarily, the Bin Packing is considering a fixed metric (*i.e*, the maximum number of UEs per BBU), which overlooks the radio conditions and the spectral efficiency. The IACA scheme keeps up with 1 active BBU until we reach 4 UEs per cell, while the Bin Packing algorithm activates 3. When the number of UEs per cell reaches 7, the Bin Packing activates 7 BBUs, while our scheme makes do with 3 BBUs. When 11 UEs per cell is reached, our algorithm activates all BBUs.

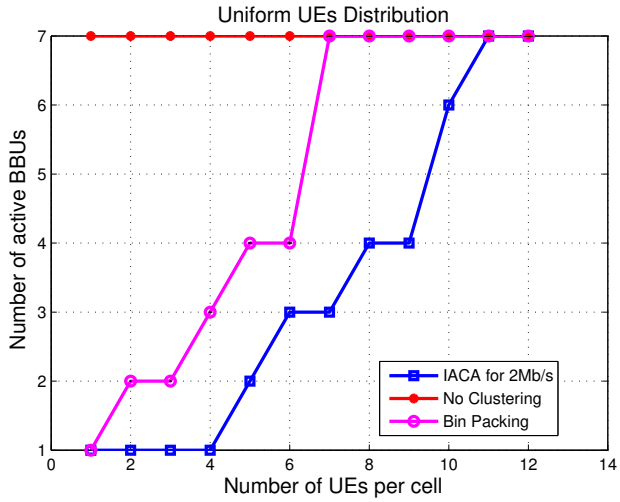


FIGURE 3.2: Number of Active BBUs: IACA vs. Bin Packing vs. No Clustering

Figure 3.3 displays the total power consumption in the system as a function of the number of UEs per cell. We notice that the lowest value is shown for 1 UE per cell, as the IACA and the Bin Packing schemes activate 1 BBU (cf. Figure 3.2). The power consumption increases when the number of active BBUs increases to satisfy a higher number of UEs. However, for the cases of 6 and 7 UEs per cell, where the IACA scheme activates 3 BBUs, we notice that the power consumption slightly increases. In fact, when more UEs are served, more throughput is derived to satisfy the QoS. Since the power consumption model varies linearly as a function of the derived rate, the value increases for the same number of active BBUs. For the Bin Packing case, we notice the same values for the same number of active BBUs, as the cases of 5 and 6 UEs per cell. In fact, the Bin Packing scheme chooses the partitions that minimize the number of active BBUs. In that case, many solutions exist for one single scenario. Consequently, we display the average result of provided solutions for a particular case (*i.e*, load condition). Thus, two scenarios showing the same number of active BBUs, display almost similar results, as the spaces of solutions might be the same for both cases. The power consumption attains its maximal value for the IACA scheme when the number of UEs per cell is 11, as all the BBUs are activated. We note that the power consumption of the conventional architecture shows a fixed value. The reason is that we consider fixed radio conditions and a full buffer traffic scheme. Thus, the total throughput value is the same as all resources are consumed.

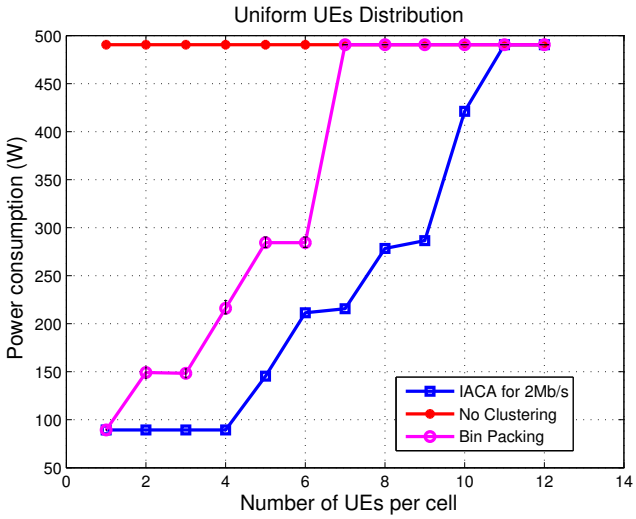


FIGURE 3.3: Power Consumption: IACA vs. Bin Packing vs. No Clustering

In Figure 3.4, the power saving metric is displayed as a function of the load conditions. The IACA scheme shows

the highest values of power savings since it activates the lowest number of BBUs. The Bin Packing realizes less power saving than the proposed approach since it activates more BBUs for different load conditions. The power saving realized by the conventional scheme is 0, since all BBUs are activated for all load conditions.

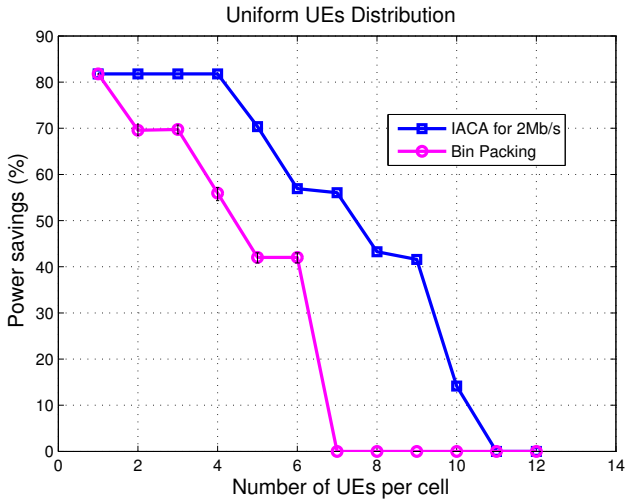


FIGURE 3.4: Power Savings: IACA vs. Bin Packing

Figure 3.5 shows the energy efficiency as a function of the number of UEs per cell. This metric underlines the Interference-Aware behavior of our proposed solution. For instance, for the cases of 6 and 7 UEs per cell, the same number of BBUs is activated as shown in Figure 3.2. However, the energy efficiency increases for 7 UEs per cell. The reason is that the algorithm chooses a different partition. In fact, for this case (*i.e*, 7 UEs per cell), the algorithm decides to increase the energy efficiency by choosing the partition that lowers the interference level between clusters. For instance, more adjacent RRHs are clustered together to reduce the interference level between one another, and thus, to enhance the efficiency in resource utilization instead of activating more BBUs as does the Bin Packing scheme. The same behavior is noticed for 8 and 9 UEs per cell. Our algorithm is more energy efficient than both the Bin Packing and the No Clustering schemes, except for the cases of 6 and 8 UEs per cell. In fact, the main objective is to reduce the power consumption, rather than to improve the energy efficiency. Within some partitions of RRHs, the energy efficiency could be lower, but the power consumption is reduced. We conclude that lowering the power consumption does not necessarily lead to enhanced energy efficiency. We note that the energy efficiency remains at a fixed level for the No Clustering scheme due to two reasons: first, all the BBUs are activated, leading to the consumption of all the radio resources since a full buffer traffic scheme is considered. Further, the radio conditions are fixed.

In Figure 3.6, the mean throughput per UE as a function of the number of UEs per cell is displayed. The mean throughput per UE decreases when the number of UEs per cell increases for the No Clustering scheme. In both IACA and Bin Packing schemes, the mean throughput per UE shows fluctuations while the number of UEs per cell increases. In fact, the higher the number of UEs, the higher the number of activated BBUs. However, activating more BBUs means introducing more interference. This explains the fluctuating behavior of both algorithms. In all cases, we notice that our algorithm tries to reduce the mean throughput per UE to 2 Mb/s in order to realize more power saving. Let us recall that the traffic model is full buffer, and thus, when we activate more BBUs, the resources are entirely consumed. However, the objective of the algorithm is to reduce the power consumption subject to minimum throughput requirements. We note that the Bin Packing scheme shows higher values of mean throughput per UE, as it activates more BBUs for the same load conditions.

Figure 3.7 shows the spectral efficiency per BBU as a function of the load conditions. We notice that the behavior of the spectral efficiency is the same as the energy efficiency (cf. Figure 3.5). In fact, both metrics show variations as a function of the interference level in the system. Thus, for some cases, the algorithm endeavors to increase the efficiency in resource utilization, rather than to increase the number of active BBUs, as previously explained. Increasing the resource utilization efficiency means enhancing both the spectral and energy efficiency. For more illustration, we display the chosen partitions of RRHs for four different cases (*i.e*, 6, 7, 8 and 9 UEs per cell). We note that for the cases of 6 and 7 UEs per cell, the number of active BBUs is 3. It is 4 for the cases of 8 and 9 UEs per cell. Cells of the same color are associated to the same BBU. In Figure 3.8(a), the proposed algorithm

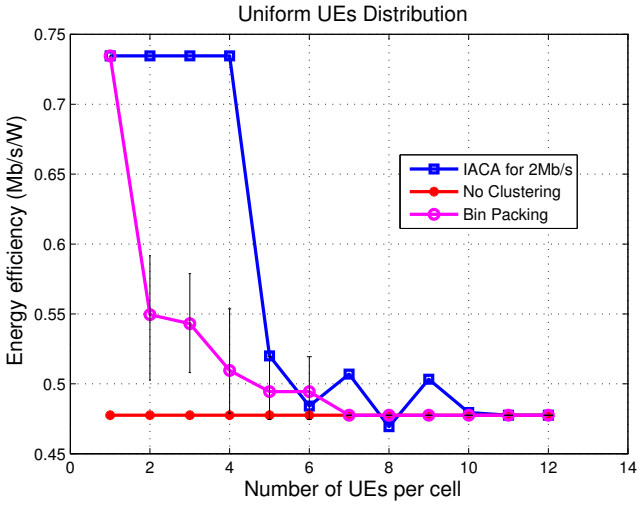


FIGURE 3.5: Energy Efficiency: IACA vs. Bin Packing vs. No Clustering

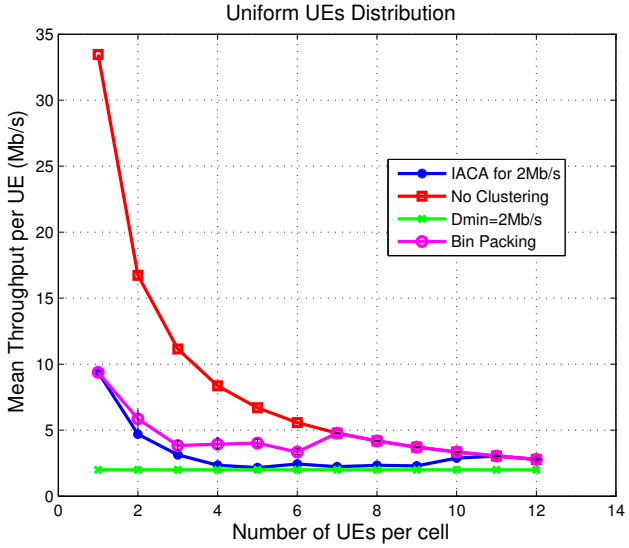


FIGURE 3.6: Mean Throughput per UE: IACA vs. Bin Packing vs. No Clustering

chooses a partition where it associates less adjacent RRHs to one BBU. Therefore, more cell edges exist, and the level of interference is higher than the case of 7 UEs per cell (Figure 3.8(b)), where we notice more adjacent RRHs clustered together. This explains how our algorithm manages the interference level to stay on the same number of active BBUs when the number of UEs increases. The same behavior is noticed for the cases of 8 and 9 UEs per cell in Figures 3.8(c) and 3.8(d).

Figure 3.9 displays the total throughput derived in the system as a function of load conditions. We notice that the throughput pattern follows the one of power consumption. In fact, this is normal since the power consumption is a linear function of the throughput. We also notice that the throughput increases for the same number of active BBUs, as the cases of 6 and 7 UEs per cell (cf. Figure 3.2). The latter highlights the Interference-Aware behavior of the proposed scheme since many partitions exist for the same number of active BBUs, affecting the cell boundaries in the network and changing the level of spectral efficiency. The throughput derived by the Bin Packing and the No Clustering schemes is higher than that obtained in IACA since more BBUs are activated.

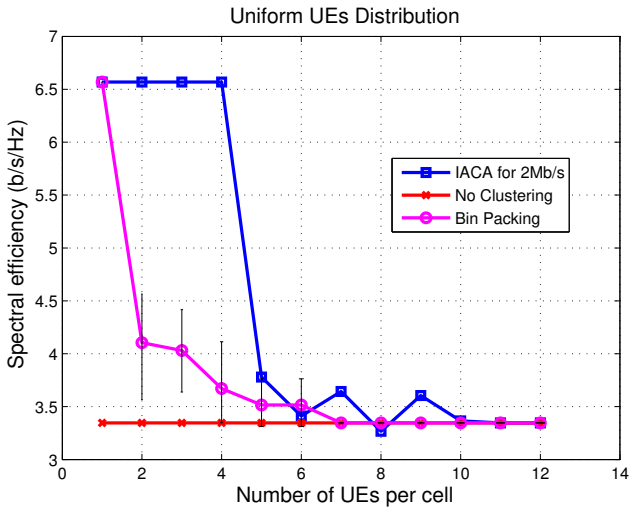
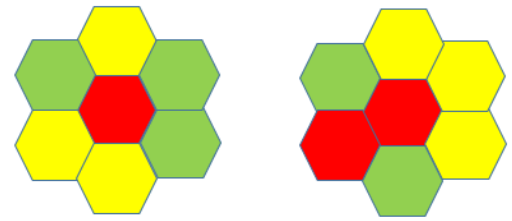
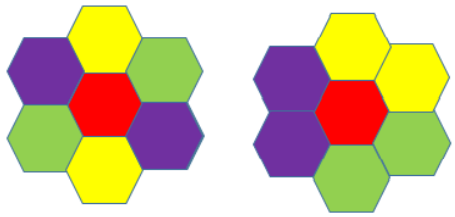


FIGURE 3.7: Spectral Efficiency: IACA vs. Bin Packing vs. No Clustering



(a) For 6 UEs per cell (b) For 7 UEs per cell



(c) For 8 UEs per cell (d) For 9 UEs per cell

FIGURE 3.8: Partitions of RRHs

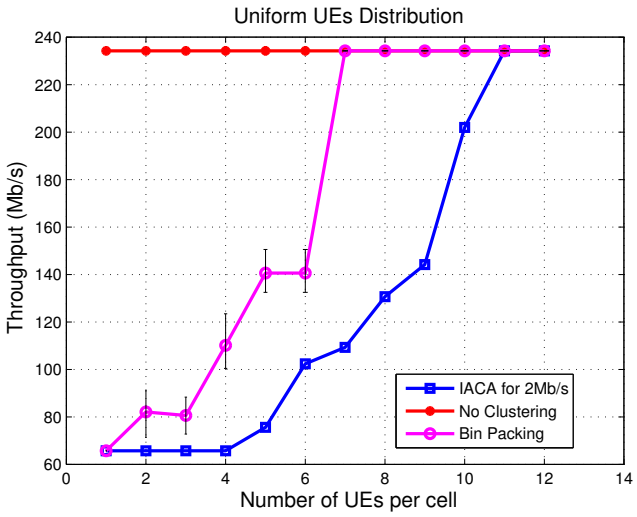


FIGURE 3.9: Total Throughput: IACA vs. Bin Packing vs. No Clustering

Good Radio Conditions Distribution

In the following, we display the results of the algorithm for UEs with good radio conditions. In that case, UEs are concentrated into zones that are close to the cell center. Thus, better radio conditions are experienced. For the Bin Packing scheme, we consider a BBU capacity of 16 UEs which is higher than the uniform UEs distribution case. In fact, the results have shown a maximum number of 16 UEs, being able to pull 2 Mb/s when 7 BBUs are activated.

In Figure 3.10, the number of active BBUs is displayed as a function of the number of UEs per cell. We notice that the IACA scheme makes do with one activated BBU, serving up to 5 UEs per cell (*i.e*, 35 UEs in the system) with a minimum mean throughput per UE of 2 Mb/s. For good radio conditions, the IACA scheme activates 2 BBUs to handle 6 UEs per cell (*i.e*, 42 UEs in the system), with a minimum mean throughput per UE of 2 Mb/s. The same number of UEs (*i.e*, 6 UEs per cell) is handled by 3 active BBUs when a uniform UEs distribution is considered. Lastly, for the case of 11 UEs per cell, the IACA scheme activates 4 BBUs when good radio conditions are considered, whereas it activates all the BBUs in the case of a uniform UEs distribution. We conclude that the BBU capacity is not the same for different radio conditions. This behavior again highlights the importance of Interference-Awareness, when designing the BBU-RRH association schemes. The Bin Packing scheme activates more BBUs for all load conditions, even when a higher BBU capacity is considered.

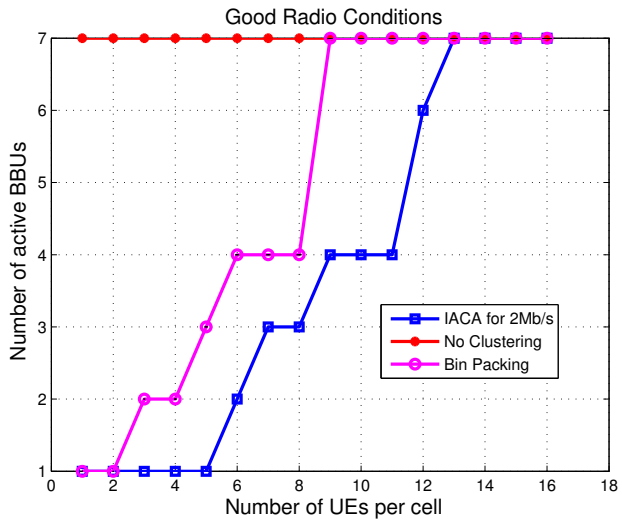


FIGURE 3.10: Number of Active BBUs: IACA vs. Bin Packing vs. No Clustering

Figure 3.11 displays the power consumption derived in the system as a function of the number of UEs per cell. We notice that for the same number of active BBUs, the power consumption is higher than the case of uniform UEs distribution (cf. Figure 3.3). For instance, for the case of 2 active BBUs, the derived power consumption is around 150 W for the good radio conditions case, whereas it is slightly less than 150 W when uniform radio conditions are considered. The same behavior is registered for the cases of 3, 4, 6 and 7 active BBUs (cf. Figures 3.10 and 3.2). The reason is that more throughput is derived when radio conditions are better. Thus, the power consumption is higher, since it is a linear function of throughput. However, even with higher power consumption, the energy efficiency is enhanced (cf. Figure 3.13) for the same number of active BBUs in comparison with the uniform UEs distribution. The Bin Packing algorithm shows higher power consumption than IACA whatever the radio conditions are, since it activates more BBUs for all cases.

In Figure 3.12, the power saving metric is globally enhanced in comparison with the uniform UEs distribution case. For example, for the case of 6 UEs per cell, the power saving is 70% when good radio conditions are considered, whereas it is slightly less than 60% for the case of uniform UEs distribution. This is normal since the needed number of active BBUs to serve a given number of UEs is higher for the case of uniform UEs distribution.

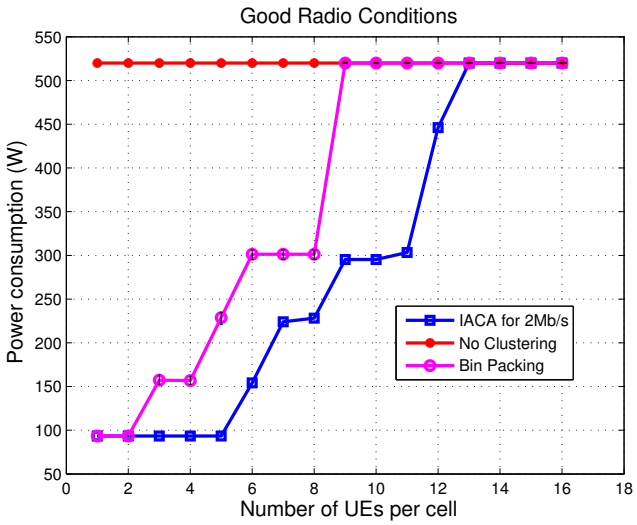


FIGURE 3.11: Power Consumption: IACA vs. Bin Packing vs. No Clustering

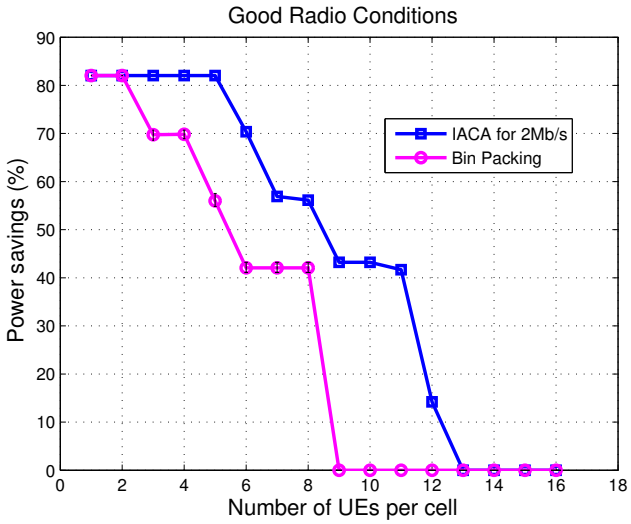


FIGURE 3.12: Power Savings: IACA vs. Bin Packing

The energy efficiency in Figure 3.13, is enhanced in comparison with the uniform radio conditions case (cf.

Figure 3.5). For instance, for the case of 2 active BBUs (*i.e*, 6 UEs per cell), the energy efficiency is slightly less than 0.6 Mb/s/W for good radio conditions, whereas it is less than 0.55 Mb/s/W for the case of a uniform UEs distribution, where we have 5 UEs per cell. The energy efficiency for good radio conditions has the same behavior as the uniform UEs distribution. In particular, for the same number of active BBUs (precisely, for the cases of 7 and 8 UEs, where 3 BBUs are activated, and 9, 10 and 11 UEs per cell, where 4 BBUs are activated), the energy efficiency is enhanced when the number of UEs per cell increases. The reason is that the RRHs partition changes in a way to lower the interference level and to increase the energy efficiency, instead of activating additional BBUs, as previously explained.

In Figure 3.14, the mean throughput per UE shows higher values for the No clustering scheme in comparison with the No Clustering scheme, where a uniform UEs distribution is considered (cf. Figure 3.6). For example, for the case of 1 UE per cell, the mean throughput per UE is slightly higher than 40 Mb/s where good radio conditions are considered. For the case of a uniform UEs distribution, it is less than 35 Mb/s. The IACA scheme shows a value around 10 Mb/s for good radio conditions, where 1 UE per cell is active, whereas it shows a value that is slightly less than 10 Mb/s for uniform UEs distribution. However, for other load conditions, we notice that the IACA scheme shows almost 2 Mb/s whatever the radio conditions are. The reason is that the IACA scheme reduces as much as possible the mean throughput per UE to gain more power saving for all radio conditions.

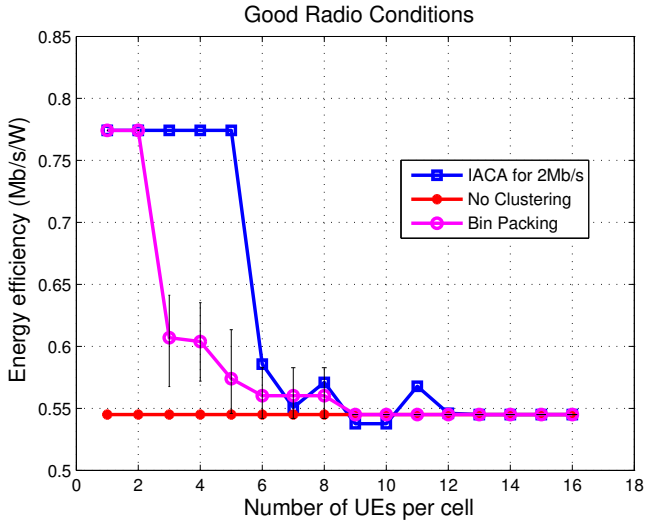


FIGURE 3.13: Energy Efficiency: IACA vs. Bin Packing vs. No Clustering

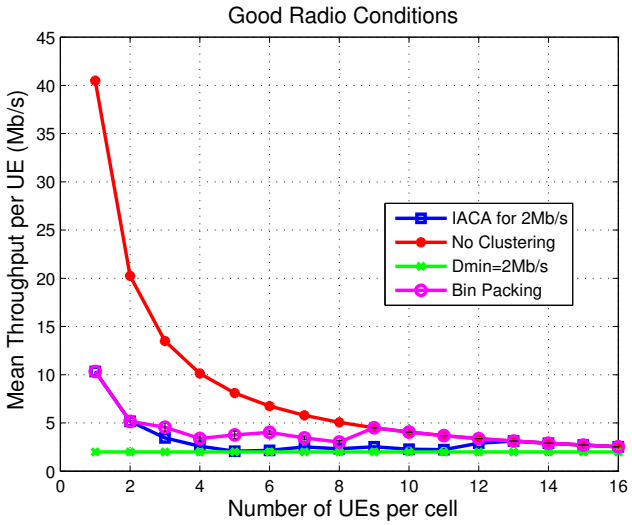


FIGURE 3.14: Mean Throughput per UE: IACA vs. Bin Packing vs. No Clustering

Figure 3.15 displays the spectral efficiency per BBU for good radio conditions. Same as the energy efficiency, we notice that the spectral efficiency is globally enhanced in comparison with the uniform UEs distribution case. For example, for the case of 1 UE per cell, where we have 1 active BBU in both cases of uniform and good radio conditions, the spectral efficiency is above 7 b/s/Hz for good radio conditions, whereas it is around 6.5 b/s/Hz for uniform UEs distribution (cf. Figure 3.7). Further, we notice that its pattern is the same as that of energy efficiency. Figures 3.16(a), 3.16(b), 3.16(c), 3.16(d) and 3.16(e) display the chosen partitions of RRHs for the cases of 7 and 8 UEs per cell, where we have 3 active BBUs and 9, 10 and 11 UEs per cell, where we have 4 active BBUs. As previously explained, we notice that the IACA scheme chooses a partition with less cell edges when the number of UEs per cell increases and the number of activated BBUs is the same. For instance, for the case of 7 UEs per cell, more cell edges exist, whereas for the case of 8 UEs per cell, we notice that the cell boundaries change and more adjacent RRHs are clustered together. For the cases of 9 and 10 UEs per cell the same partition of RRHs is chosen to serve UEs, and many cell edges are shown. For the case of 11 UEs per cell, the algorithm changes the partition to cluster more adjacent RRHs together.

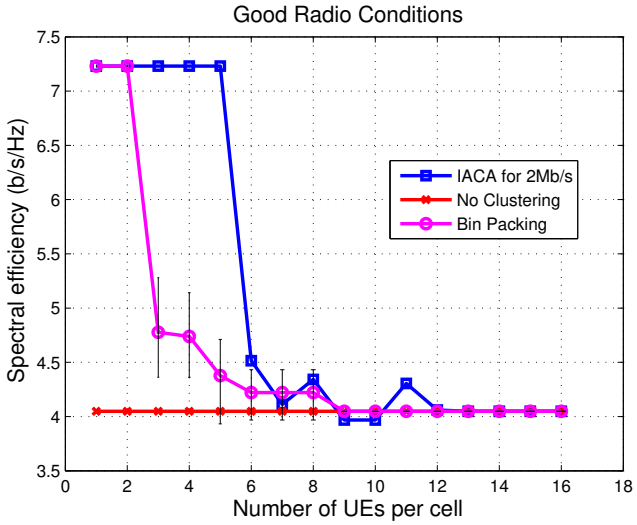
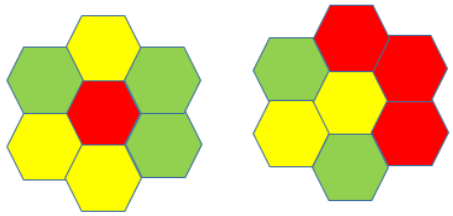
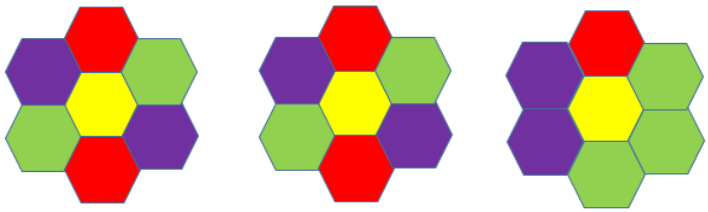


FIGURE 3.15: Spectral Efficiency: IACA vs. Bin Packing vs. No Clustering



(a) For 7 UEs per cell (b) For 8 UEs per cell



(c) For 9 UEs per cell (d) For 10 UEs per cell (e) For 11 UEs per cell

FIGURE 3.16: Partitions of RRHs

Figure 3.17 displays the total derived throughput as a function of load conditions. We notice that for good radio conditions, the total throughput is enhanced in comparison with the case of uniform UEs distribution. For instance, for the same number of active BBUs, the total derived throughput is higher than that of uniform UEs distribution case. The total throughput derived by IACA is less than the Bin Packing and the No Clustering schemes as they activate more BBUs.

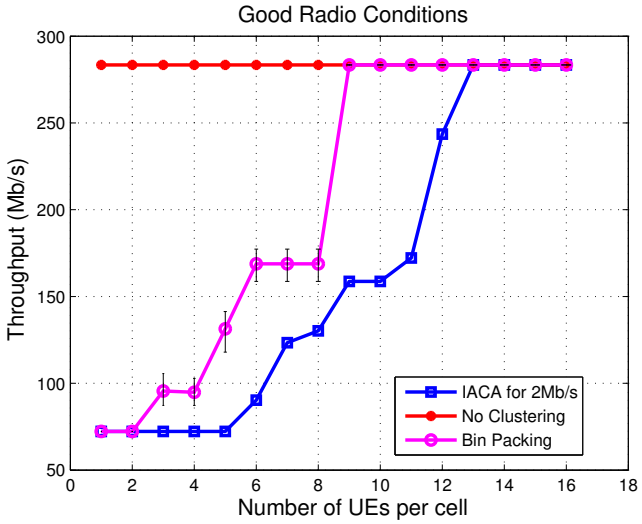


FIGURE 3.17: Total Throughput: IACA vs. Bin Packing vs. No Clustering

Bad Radio Conditions Distribution

In this subsection, the results are displayed for bad radio conditions, where UEs are concentrated into zones close to the cell edge. The BBU capacity in the Bin Packing case is fixed to 9 UEs since the results have shown a maximum number of 9 UEs per BBU, being able to pull 2 Mb/s when all BBUs are activated.

Figure 3.18 shows the number of active BBUs as a function of the number of UEs per cell. We notice that the IACA scheme significantly minimizes the number of active BBUs in comparison with the Bin Packing approach. For instance, for 4 UEs per cell (*i.e*, 28 UEs in the network), the IACA scheme makes do with 1 active BBU, whereas the Bin Packing activates 4 BBUs. The same behavior is noticed for 5 UEs per cell, where the IACA scheme activates 2 BBUs, whereas the Bin Packing activates 7 BBUs. In fact, the Interference-Aware behavior is more effective for the case of bad radio conditions than the cases of uniform and good radio conditions, since significant gap of difference is shown in comparison with the Bin Packing, which is an Interference-Oblivious approach.

Figure 3.19, displays the power consumption as a function of the load conditions. Similar to the number of active BBUs, the power consumption metric shows significant discrepancy between the IACA and the Bin Packing schemes. Again, being aware of the interference level for the bad conditions case is paramount to achieve the maximum power consumption minimization.

Figure 3.20 is showing the power saving metric as a function of the number of UEs per cell. Since the power consumption is significantly minimized in comparison with the Bin packing scheme, the power saving also shows the same behavior. We also notice that the power saving metric decreases in comparison with the cases of uniform and good radio conditions. For instance, for 7 UEs per cell, the realized power saving is around 40% for bad radio conditions, whereas it is around 55% for uniform radio conditions and 60% for good radio conditions (cf. Figures 3.4 and 3.12).

Figure 3.21 displays the energy efficiency as a function of the load conditions. We notice, for the case of bad radio conditions, the energy efficiency of the IACA scheme is enhanced in comparison with the cases of Bin Packing and No Clustering schemes, for all load conditions. We conclude that for the case of bad radio conditions, lowering the power consumption always enhances the energy efficiency.

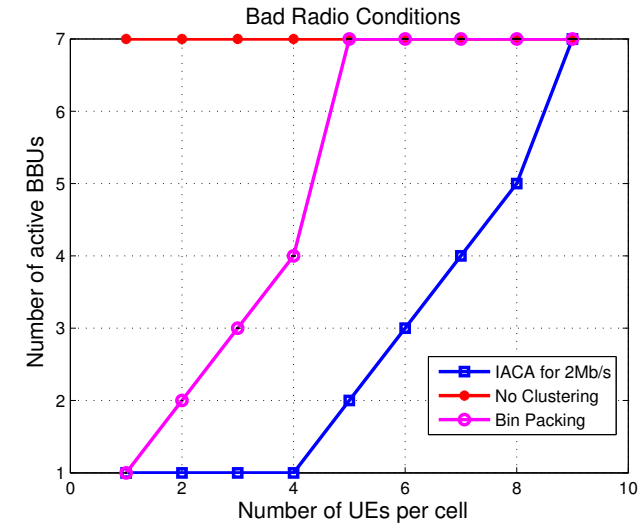


FIGURE 3.18: Number of Active BBUs: IACA vs. Bin Packing vs. No Clustering

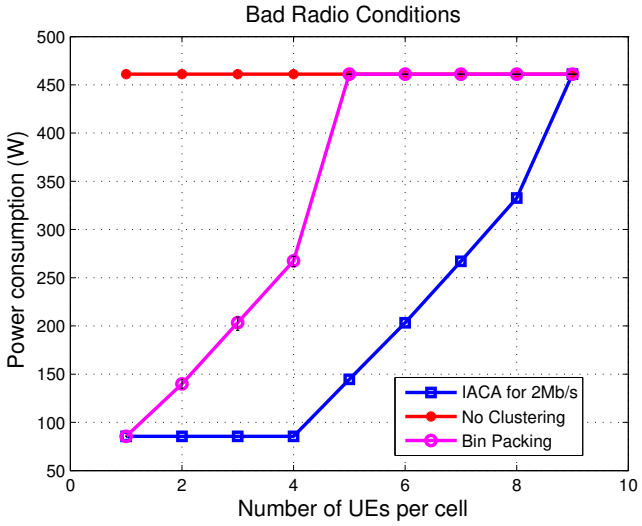


FIGURE 3.19: Power Consumption: IACA vs. Bin Packing vs. No Clustering

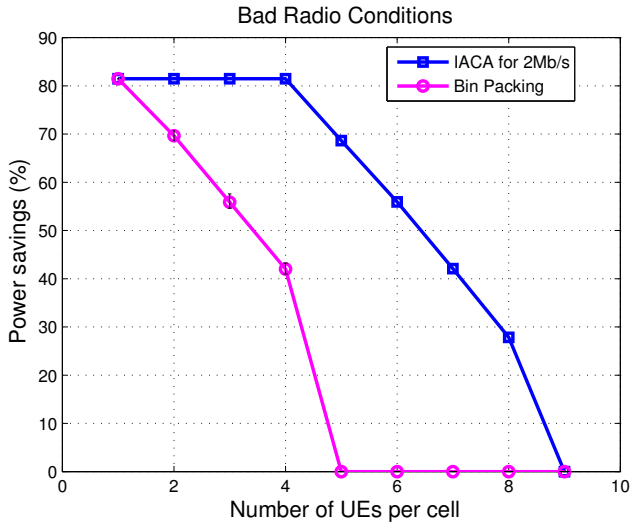


FIGURE 3.20: Power Savings: IACA vs. Bin Packing

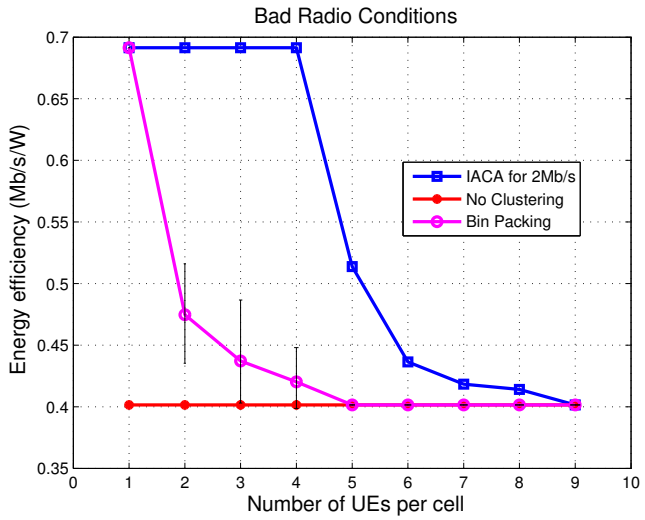


FIGURE 3.21: Energy Efficiency: IACA vs. Bin Packing vs. No Clustering

The mean throughput per UE is displayed in Figure 3.22 as a function of the number of UEs per cell. We notice that the mean throughput per UE shows lower values for the bad radio conditions case than the cases of uniform and good radio conditions (cf. Figures 3.6 and 3.14). For instance, for the case of 1 UE per cell, the mean throughput per UE shown by the IACA scheme is less than 10 Mb/s for the bad radio conditions case, whereas it is around 10 Mb/s for the uniform radio conditions case, and it is more than 10 Mb/s for the good radio conditions case. As the cases of uniform and good radio conditions, the IACA scheme reduces the mean throughput per UE to 2 Mb/s to gain more power saving. However, it does not show values lower than 2 Mb/s.

Figure 3.23 displays the spectral efficiency per BBU as a function of the load conditions. As the cases of uniform and bad radio conditions, the spectral efficiency follows the behavior of the energy efficiency. Further, it shows lower values in comparison with the cases of uniform and good radio conditions. For instance, for the cases of 1 UE per cell, where we have 1 active BBU for all cases, the spectral efficiency in the case of bad radio conditions shows a value that is less than 6 b/s/Hz, whereas it is more than 6.5 Mb/s/Hz for the case of uniform radio conditions, and more than 7 b/s/Hz for the case of good radio conditions (cf. Figures 3.7 and 3.15).

In Figure 3.24, the total derived throughput in the system is shown as a function of the load conditions. Similar to the power consumption, the total throughput is significantly reduced in comparison with the case of the Bin Packing scheme. However, the total derived throughput is enough to respect the QoS constraint, and guaranteeing a mean

throughput per UE of 2 Mb/s as shown in Figure 3.22.

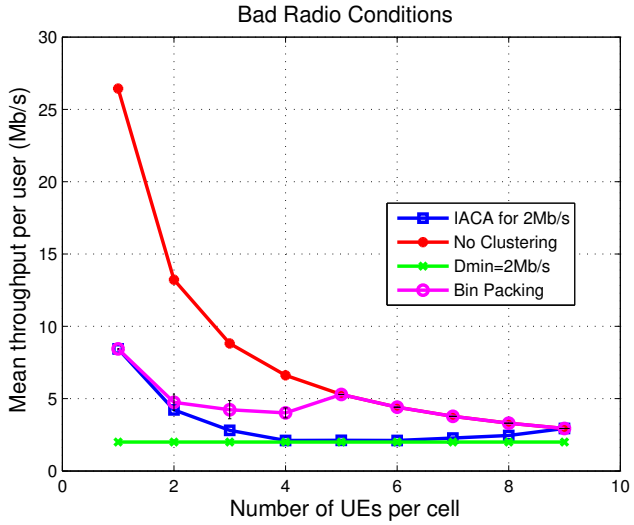


FIGURE 3.22: Mean Throughput per UE: IACA vs. Bin Packing vs. No Clustering

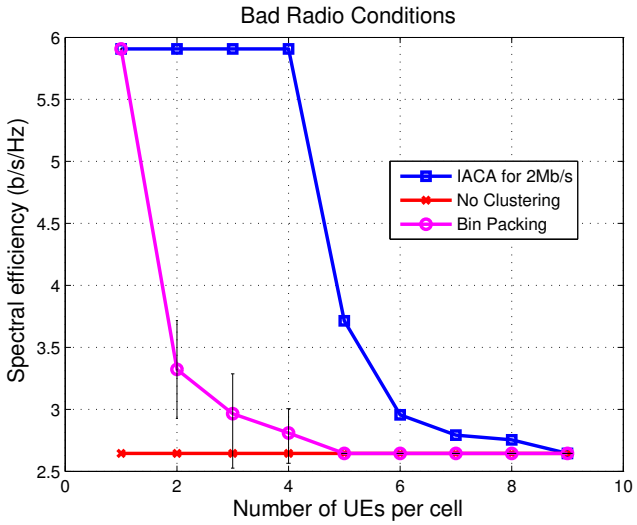


FIGURE 3.23: Spectral Efficiency: IACA vs. Bin Packing vs. No Clustering

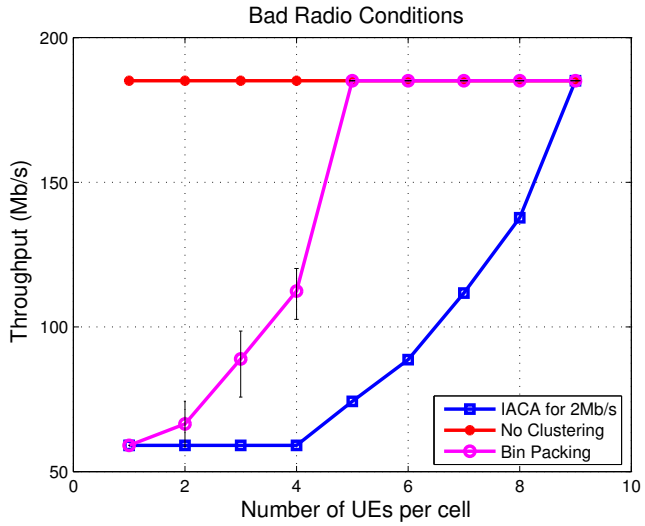


FIGURE 3.24: Total Throughput: IACA vs. Bin Packing vs. No Clustering

3.4.3 IACA vs. Heuristic vs. No Clustering

In this section, the results of the proposed heuristic are compared to the optimal solution of the Interference-Aware proposition and the No Clustering scheme. We note that the results are evaluated for a uniform UEs distribution. Further, the results of the heuristic are evaluated for 1000 iterations for each scenario (*i.e*, load condition), and average results are presented with 95% confidence interval.

We start by displaying the number of active BBUs as a function of the load conditions. In Figure 3.25, the number of active BBUs given by the heuristic is shown in comparison with the IACA and the No Clustering schemes. We notice that the heuristic activates the same number of BBUs as the IACA scheme, except for the cases of 7 and 9 UEs per cell. For the case of 7 UEs per cell, the heuristic shows an average value slightly higher than 3 active BBUs, which is the result of the optimal solution. For the case of 9 UEs per cell, the heuristic activates, on average, one additional BBU in comparison with the IACA scheme. The No Clustering scheme maintains 7 active BBUs.

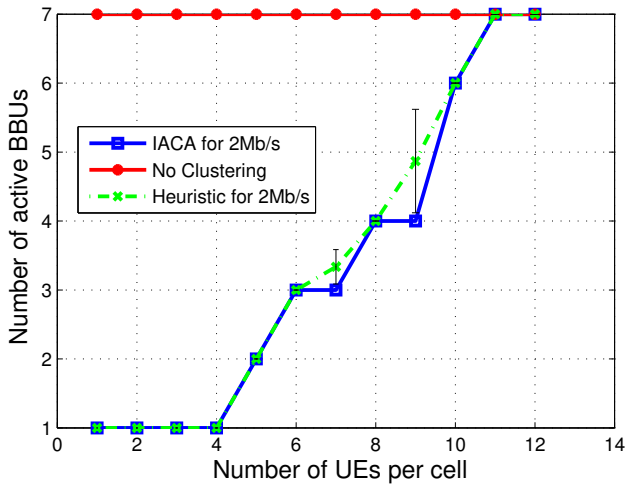


FIGURE 3.25: Number of Active BBUs: IACA vs. Heuristic vs. No Clustering

Figure 3.26 displays the power consumption of the heuristic as a function of the load conditions, in comparison with the IACA and to the No Clustering schemes. We notice that the power consumption given by the heuristic shows almost the same values of the IACA scheme, except for 7 and 9 UEs per cell because more BBUs are activated. In fact, this underlines the efficiency of the proposed heuristic since the main objective remains to decrease the

power consumption. The power consumption of the No Clustering scheme remains the highest since all BBUs are activated.

In Figure 3.27, the power saving metric of the heuristic is displayed in comparison with the IACA and the No Clustering schemes. Similar to the power consumption, the heuristic shows almost the same values as the IACA scheme, except for 7 and 9 UEs per cell since more BBUs are activated.

Figure 3.28 shows the energy efficiency of the heuristic, in comparison with the IACA and No Clustering schemes. The energy efficiency is higher for the heuristic solution, except for the cases of 7 and 9 UEs per cell. As previously explained, reducing the power consumption does not necessarily lead to an enhanced energy efficiency. In fact, the energy efficiency of the heuristic algorithm is lower than that of the IACA scheme when the number of active BBUs, shown by the heuristic, is higher (cf. figure 3.25). The reason is that less interference is the network is experienced in the case of IACA, since less BBUs are activated. The opposite behavior is noticed when the number of active BBUs is the same.

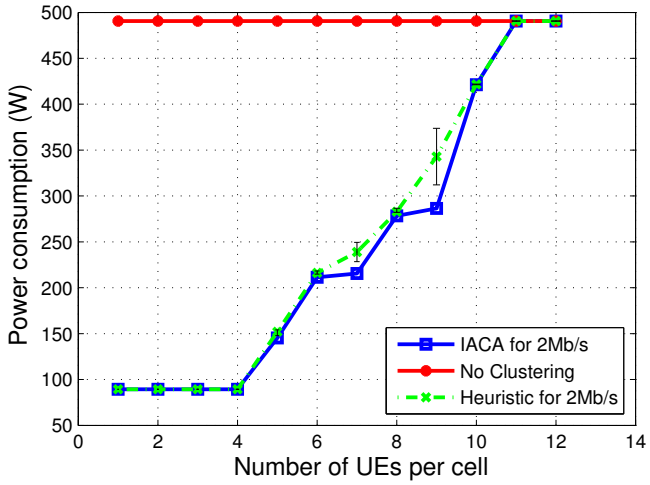


FIGURE 3.26: Power Consumption: IACA vs. Heuristic vs. No Clustering

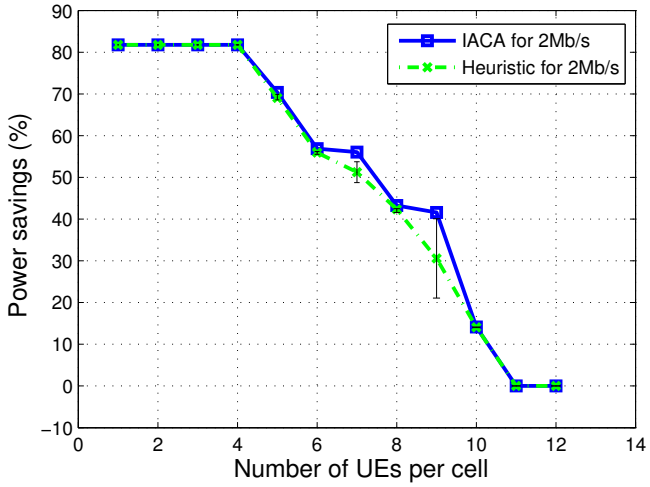


FIGURE 3.27: Power savings: IACA vs. Heuristic

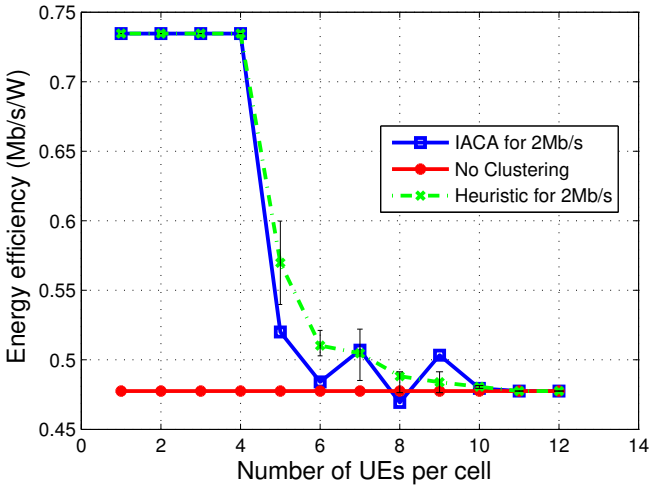


FIGURE 3.28: Energy Efficiency: IACA vs. Heuristic vs. No Clustering

The mean throughput per UE as a function of the load conditions is shown in Figure 3.29, in comparison with the IACA and the No Clustering schemes. Similar to the IACA scheme, the heuristic reduces the mean throughput per UE as much as possible to gain more power saving. However, we notice that the heuristic does not show values less than 2 Mb/s. The No Clustering scheme realizes the highest mean throughput per UE since it activates all BBUs and a full buffer traffic model is considered.

In Figure 3.30, the spectral efficiency per BBU for the heuristic is shown, in comparison with the IACA and the No Clustering schemes. Similar behavior to that of the energy efficiency is noticed. In particular, the spectral efficiency of the heuristic algorithm is higher than the IACA scheme, except for the cases where the IACA activates less BBUs than the heuristic. The reason, as previously mentioned, is related to the interference level which is lower when the IACA activates fewer BBUs.

In figure 3.31, the total derived throughput of the heuristic is displayed, in comparison with the IACA and the No Clustering schemes. We notice that the total throughput derived by the heuristic shows higher values than the IACA scheme. In fact, the heuristic enjoys higher spectral and energy efficiency. Thus, the total derived throughput is higher. Further, the IACA scheme main objective remains to reduce power consumption, at the cost of reduced throughput in comparison with other schemes.

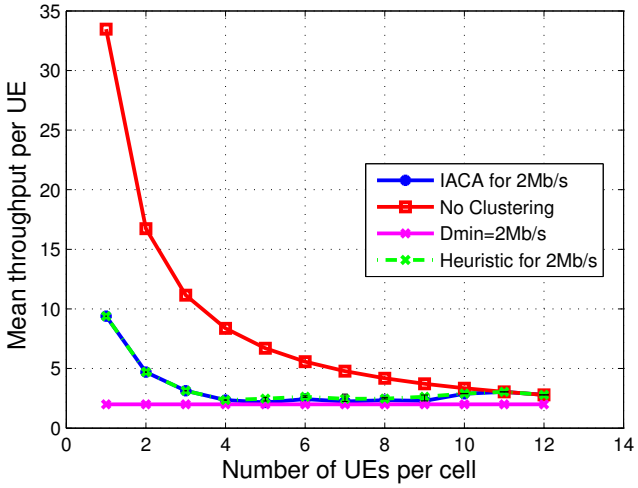


FIGURE 3.29: Mean Throughput per UE: IACA vs. Heuristic vs. No Clustering

Figure 3.32 displays the execution time of the IACA and the heuristic as a function of the load conditions. The results are executed over a hardware of 64-bit, Intel 7 core, 2.5 Ghz. We notice that the heuristic significantly reduces

the execution time in comparison with the IACA scheme, which underlines the efficiency of the proposed heuristic that shows close performance to the optimal solution with significant less computational requirements.

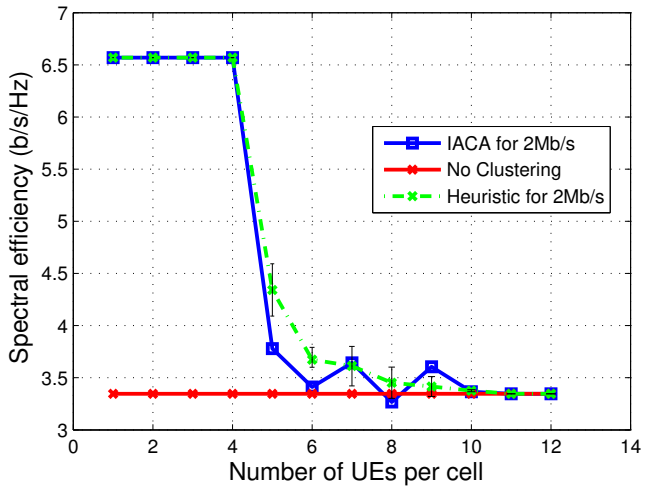


FIGURE 3.30: Spectral Efficiency: IACA vs. Heuristic vs. No Clustering

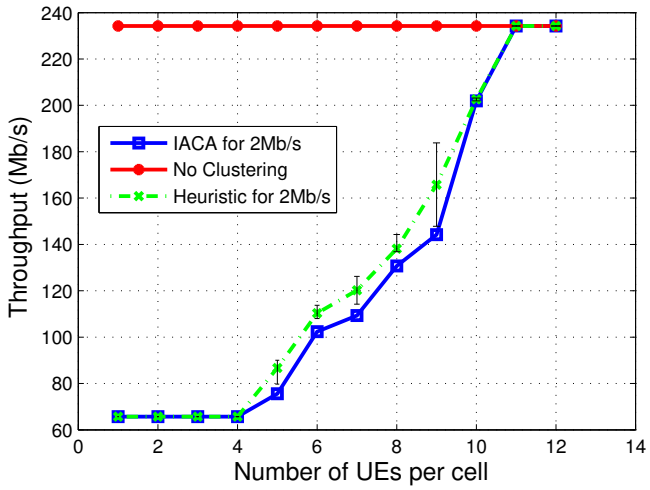


FIGURE 3.31: Total Throughput: IACA vs. Heuristic vs. No Clustering

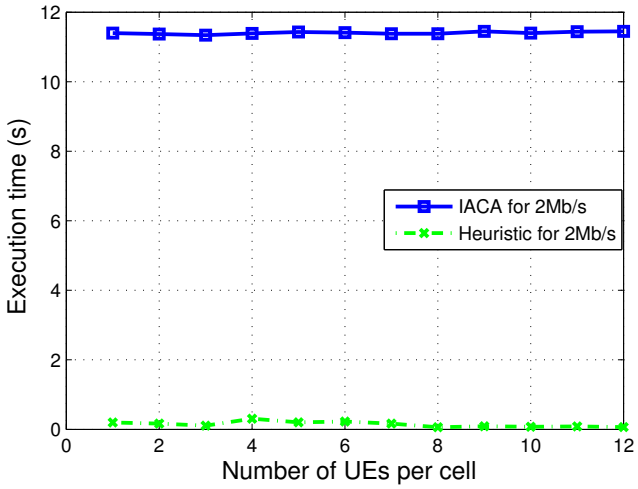


FIGURE 3.32: Execution Time: IACA vs. Heuristic

3.5 Concluding Remarks

In this chapter, an Interference-Aware algorithm has been proposed and evaluated in comparison with an Interference-Oblivious approach, the Bin Packing problem. Three scenarios has been taken into consideration to closely evaluate the results: The uniform UEs distribution, the good radio conditions and the bad radio conditions. Further, a greedy heuristic has been proposed and the results were evaluated and compared with the optimal solution of the Interference-Aware algorithm. The main conclusions are stated below:

- The Interference-Aware decisions are paramount to achieve maximum power saving. Particularly, the chosen partition of RRHs affects the level of interference, and thus, the spectral efficiency which is essential to decide the right number of active BBUs.
- The radio conditions affect the maximum number of UEs that a BBU can handle. For instance, when good radio conditions are considered, the number of active BBUs needed to handle a given load condition, is less than that needed for the cases of uniform or bad radio conditions, where more BBUs are activated to cope with the traffic demand.
- The power consumption minimization does not necessarily lead to energy efficiency enhancement. As shown from the results, some cases present the lowest values of power consumption. On the other hand, the energy efficiency does not strike the highest values. However, for bad radio conditions the power consumption minimization enhances the energy efficiency for all cases of load conditions.
- The proposed heuristic shows close performance to the optimal solutions with less computational requirements.

In the following chapter, we submit the proposed heuristic to the stringency of a real network topology. Further, we propose additional greedy heuristics for better comparative scoring.

Chapitre 4

Heuristics for BBU-RRH Association Optimization on a Real Case Topology

In chapter 3, we introduced an algorithm endowed an Interference-Aware behavior. Further, we proposed a greedy heuristic aiming to provide close performance to the optimal solution with less computational requirement. In this chapter, we submit the Interference-Aware behavior to the stringency of a real network topology consisting in 20 cells (i.e, the 4G network topology for the 14th district of Paris in France). Precisely, we propose several Interference-Aware heuristics, and we compare them. A first type of heuristics is inspired from efficient heuristics of the Bin Packing problem, such as the Best Fit Decreasing (BFD), the Worst Fit Decreasing (WFD) and the First Fit Decreasing (FFD). Those heuristics are adapted to compose with interference. A second type of heuristics is similar to the previously proposed heuristic of chapter 3. We first start by describing the classical BFD, WFD and FFD, then we introduce our proposed Interference-Aware Bin Packing heuristics. We further introduce the heuristics that are similar to the one described in chapter 3. We evaluate the results of the Interference-Aware BFD in comparison with the classical one. Lastly, we compare all proposed heuristics: The ones inspired from the Bin Packing scheme and the ones that are similar to the proposed scheme in chapter 3.

4.1 Best Fit Decreasing, Worst Fit Decreasing and First Fit Decreasing

— *Best Fit Decreasing:* The BFD algorithm is an efficient greedy heuristic, sorting the objects (*i.e*, RRHs) in a decreasing order in terms of volumes (*i.e*, number of UEs). The object to be packed is placed into the bin that embraces it the best. In other terms, the object is placed into the bin that has the best fitted empty place for the object volume. In order to apply the BFD in a C-RAN context, RRHs are sorted in terms of number of UEs. A RRH is associated to the BBU that suits it the best in terms of number of UEs. Precisely, the RRH is associated to the BBU that has the minimal empty place that suits it.

— *Worst Fit Decreasing:* The WFD algorithm differs from the BFD by placing the object into the bin that has the largest empty place and suits it. In a C-RAN context, a RRH is associated to the BBU having the maximal empty place and is able to embrace the RRH UEs.

— *First Fit Decreasing:* The FFD algorithm places the object into the bin that first fits it. In other terms, the RRH is mapped to the first BBU that can handle it.

4.2 Interference-Aware BFD, WFD and FFD

In Chapter 2 (cf. subsection 2.5.1), we have described and provided the formulation of the Bin Packing problem. Recall that the Bin Packing consists in finding a partition of objects in a way to minimize the number of used bins. The objects are packed into the bins in a way not to violate the maximum bin capacity. Within an Interference-Aware context, the constraint of maximum bin capacity is substituted with that of minimum throughput requirement per UE. In other terms, RRHs are associated to one BBU until a minimum level of throughput requirement per UE is violated. This constraint is expressed as follows:

$$T(c) \geq n_c \cdot D_{min}, \forall c \in \mathcal{P} \tag{4.1}$$

Let us recall that $T(c)$ expresses the total throughput derived in cluster c, n_c is the total number of UEs in cluster c, and D_{min} the minimum throughput per UE that should be available within each cluster. In fact, the mean throughput per UE shows a clear image of the BBU capacity that is affected by both the load conditions (*i.e*, number of UEs) and the interference impact of RRHs associated to different BBUs.

— *Interference-Aware BFD:* An Interference-Aware BFD consists in clustering the RRH within the BBU that has the lowest mean throughput per UE and respects (4.1), rather than clustering it into the BBU fitting it the best in terms of number of UEs and respecting the bin capacity. Similar to the classical BFD, the RRHs are sorted in decreasing order of their load (*i.e*, number of serviced UEs).

— *Interference-Aware WFD:* Similar to the BFD, the RRHs are sorted in decreasing order of their load. A RRH is associated to the BBU that shows the highest mean throughput per UE with respect to (4.1).

— *Interference-Aware FFD:* Again, the RRHs are sorted in decreasing order of their load, and one RRH is associated to the first BBU that respects constraint (4.1).

4.3 Greedy Heuristics

In this section, we present additional greedy heuristics that we propose, and are similar to the one described in Chapter 3. These heuristics are deemed "Random Load-Random Load", "Maximum Load-Minimum Power" and "Minimum Load-Minimum Power". They are described in the following:

— *Random Load-Random Load:* The "Random Load-Random Load" starts by choosing randomly two RRHs and associates them to one BBU in a way not to violate constraint (4.1). The heuristic keeps up by randomly choosing RRHs to be associated to the same cluster until no RRH respecting (4.1) is found. It then repeats the process until all RRHs are associated to BBUs.

— *Maximum Load-Minimum Power:* The "Maximum Load-Minimum Power" heuristic starts by choosing the RRH with the highest load, and associates it with the RRH that minimizes the total power consumption of the cluster with respect to constraint (4.1). It keeps up by associating RRHs that minimize the total power consumption of the cluster until no RRH respecting (4.1) is found. The process is repeated until all RRHs are clustered to BBUs.

— *Minimum Load-Minimum Power:* Contrarily to the "Maximum Load-Minimum Power", the "Minimum Load-Minimum Power" chooses the RRH with the lowest load, and associates it with the RRH that minimizes the total power consumption of the cluster without violating constraint (4.1).

We note that the proposed heuristic in chapter 3 will be referred to as "Random Load-Minimum Power", as it starts by randomly choosing a RRH and associates it with the one that minimizes the power consumption of the cluster.

4.4 Complexity Analysis

The classical heuristics of the Bin Packing problem (*i.e*, BFD, WFD and FFD) have a complexity of $R \cdot log(R)$ [70], where R is the total number of objects (*i.e*, RRHs). As the Interference-Aware Bin Packing heuristics behave similarly to the classical heuristics, they execute the same number of operations and have the same complexity. As for the greedy heuristics, their complexity remains quadratic for the previously explained reason in section 3.3 of chapter 3, knowing that the "Maximum Load-Minimum Power" and "Minimum Load-Minimum Power" heuristics, differ from the "Rand Load-Minimum Power" by choosing the maximum/minimum loaded RRHs while starting a stage. In that case, the complexity of the max and min load functions is equal to $O(log(n))$, where n is the number of unmapped RRHs within a given stage. However, for high number of RRHs the complexity of the logarithmic function would be negligible in comparison with the number of operations the heuristics perform to search for the RRH that minimizes the power consumption of the cluster.

4.5 Performance Evaluation

4.5.1 Simulation Setup

We use Matlab for simulations and we consider the 4G network topology for the 14th district of Paris in France [23]. Figure 4.1 displays the realistic positioning of the base stations (*i.e*, RRHs) in the 14th district of Paris. We note

that the system model remains the same as the previously described one in chapter 3 (cf. section 3.1).

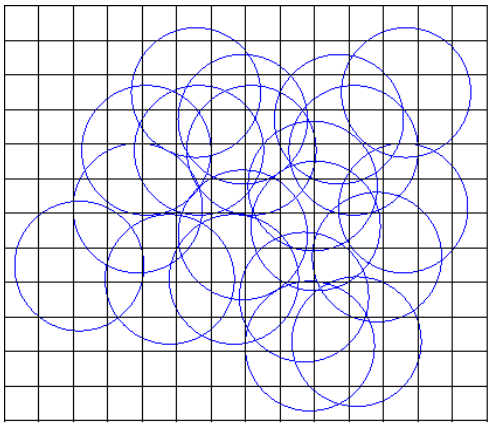


FIGURE 4.1: 4G Network Topology for the 14th District of Paris in France

We recall that a quadratic meshing of the covered area is considered. In this case, we resort to a uniform UEs distribution per cell (*i.e*, the number of UEs per zone is the same). The employed propagation model is the Cost Hata 231 for urban areas, and a frequency reuse-1 scheme is considered. We assume a cell radius of one base station to be 450 m since Paris is a dense region and one base station would have a shorter coverage range than base stations located in suburban or rural regions. Further, the value of D_{min} is set to 1 Mb/s. Same metrics of chapter 3 are used to compare the various schemes (cf. section 3.4). The simulation parameters are shown in table 4.1.

4.5.2 IA-BFD vs. BFD vs. No Clustering

In this subsection, we display the comparative results of the Interference-Aware BFD, denoted as IA-BFD, the classical BFD, denoted as BFD, and the conventional architecture denoted as No Clustering scheme. The results are displayed for three load conditions: Low, medium and high. For low load conditions, the number of UEs per cell varies between 1 and 5. For medium load conditions, the number of UEs per cell varies between 6 and 10. Lastly, for high load conditions, the number of UEs per cell is between 11 and 15. Simulations are repeated 2000 times, and results are averaged and shown with 95% confidence interval. Further, the bin capacity is set to 15 UEs because, through extensive simulations, we assessed the maximum number of UEs per BBU (set to 15) necessary to pull 1 Mb/s when 20 BBUs are activated.

Figure 4.2 displays the number of active BBUs as a function of the number of UEs per cell. At low load, the IA-BFD activates 2 BBUs, whereas the BFD scheme activates 4. This is due to the Interference-Aware behavior

TABLE 4.1: Simulation Parameters

Parameter	**Value**
Traffic Model	Full Buffer
Scheduling Scheme	Fair Resource Sharing
UEs Distribution	Uniform
Propagation Model	Cost Hata 231
Shadowing Standard Deviation	10 dB
Transmit Power of RRH	40 dBm
Thermal Noise Power	-174 dBm/Hz
BW	10 MHz
Cell Radius	450 m
A	50 W
B	0.6
D_{min}	1 Mb/s
Meshing Step	240 m

of the IA-BFD that takes into consideration the load and the interference conditions, contrarily to the BFD scheme that relies only on the number of UEs per cell for BBU-RRH association decisions. At medium load, the IA-BFD activates an average of 8 BBUs, whereas the BFD activates around 13 BBUs. At high load, the IA-BFD activates 14 BBUs, and the BFD activates all BBUs. In fact, under high load, the number of active BBUs increases, which in turn increases the level of interference since the same spectrum is reused. As a consequence, the IA-BFD does not activate all BBUs to lower the interference level that harmfully affects the QoS, and is contented with 14 activated BBUs, contrarily to the BFD and the No Clustering schemes that activate all BBUs.

Figures 4.3 and 4.4 show the power consumption and the realized power saving as a function of the number of UEs per cell. We notice that the power consumption shown by the IA-BFD is reduced in comparison with the BFD and the No Clustering schemes. Similarly, the realized power saving shown by the IA-BFD is increased in comparison with the classical BFD and the conventional architecture.

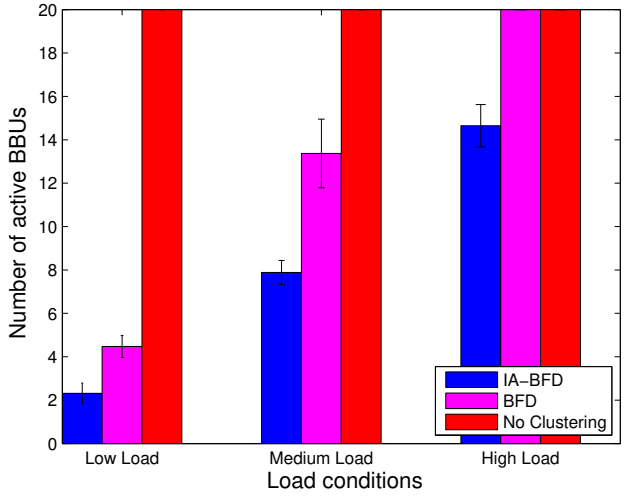


FIGURE 4.2: Number of Active BBUs: IA-BFD vs. BFD vs. No Clustering

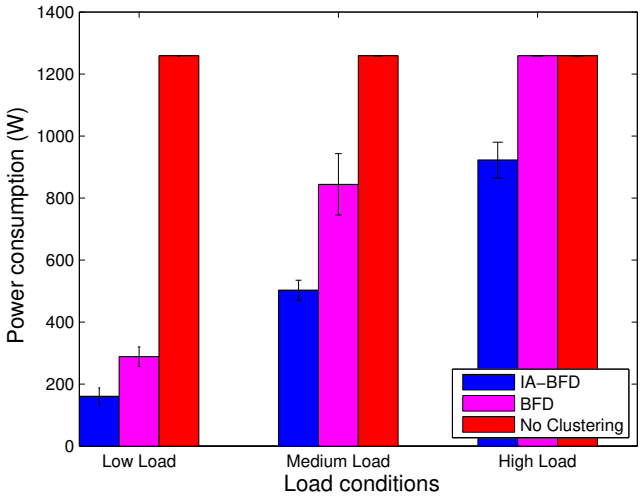


FIGURE 4.3: Power Consumption: IA-BFD vs. BFD vs. No Clustering

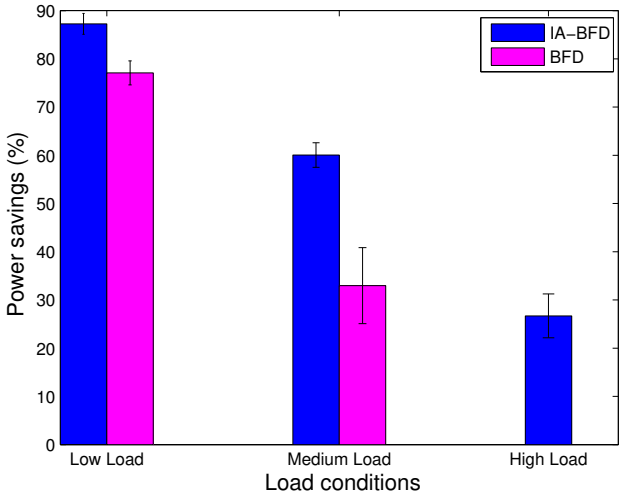


FIGURE 4.4: Power Savings: IA-BFD vs. BFD

Figure 4.5 shows the energy efficiency as a function of the load conditions. We notice that for low load, the energy efficiency significantly improves in comparison with the classical BFD and the conventional architecture. The reason is the low number of activated BBUs which leads to less interference in the network. However, we notice that for medium and high load conditions, the energy efficiency realized by the IA-BFD slightly improves in comparison with the classical BFD and the No Clustering schemes. The reason is due to the close positioning of the base stations that aggressively interfere each others as shown in Figure 4.1. Thus, when more BBUs are activated (as for the cases of medium and high load conditions), the interference level significantly increases, leading to a degradation in terms of energy and spectral efficiency (cf. Figure 4.7) until a stable level is reached, affecting slightly the energy and the spectral efficiency. However, the main objective of our algorithm is to reduce the power consumption, rather than to maximize the energy efficiency. Recall that reducing the power consumption does not necessarily lead to enhanced energy efficiency.

Figure 4.6 displays the mean throughput per UE as a function of the number of UEs per cell. We notice that the IA-BFD is showing the lowest level of mean throughput per UE, because it activates the lowest number of BBUs. Further, the IA-BFD shows a value close to 1 Mb/s for all load conditions, as the main objective is to realize the maximum power reduction with respect to a minimum level of throughput requirement per UE within each cluster, which is 1 Mb/s in that particular case.

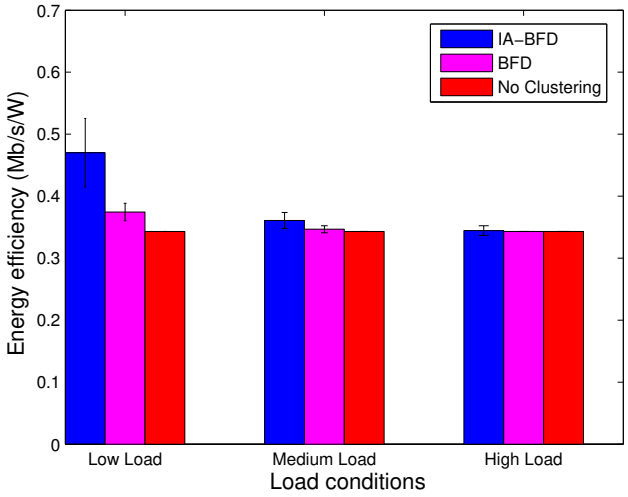


FIGURE 4.5: Energy Efficiency: IA-BFD vs. BFD vs. No Clustering

Figure 4.7, shows the spectral efficiency per BBU as a function of the number of UEs per cell. The spectral efficiency shown by the IA-BFD is significantly improved for low load conditions, in comparison with the BFD and the No Clustering schemes. Similar to the energy efficiency, for medium and high load conditions, the IA-BFD slightly improves the spectral efficiency as the number of BBUs is higher and the interference level reaches a stable level.

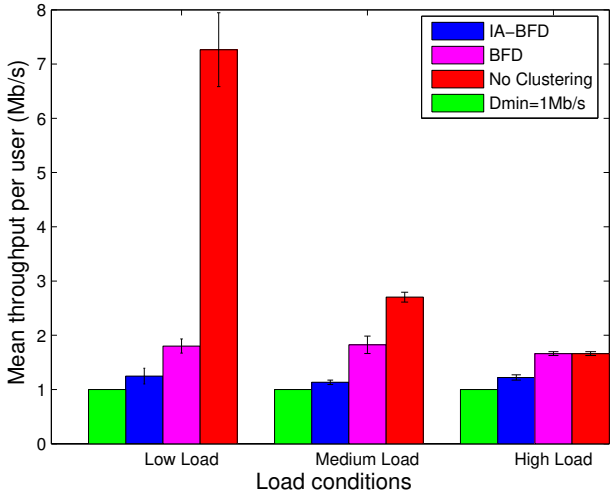


FIGURE 4.6: Mean Throughput per UE: IA-BFD vs. BFD vs. No Clustering

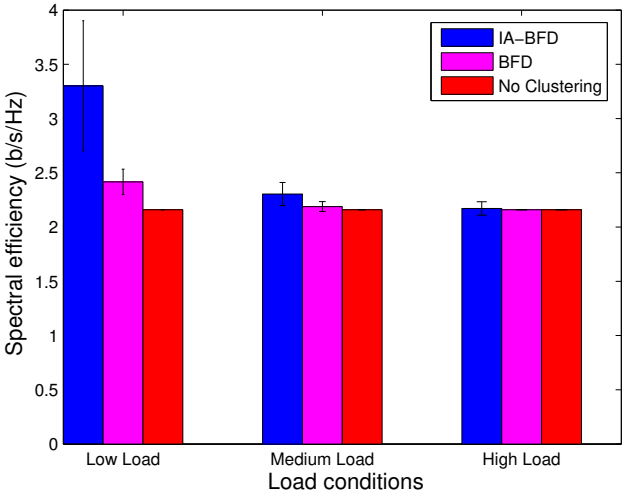


FIGURE 4.7: Spectral Efficiency: IA-BFD vs. BFD vs. No Clustering

In Figure 4.8, the total derived throughput in the system is displayed. We notice that the No Clustering scheme shows the highest total throughput as it activates all BBUs. The classical BFD also shows an improved total throughput as more BBUs are activated. However, the realized power saving of the BFD and the No Clustering schemes is significantly lower than the IA-BFD as shown in Figure 4.4.

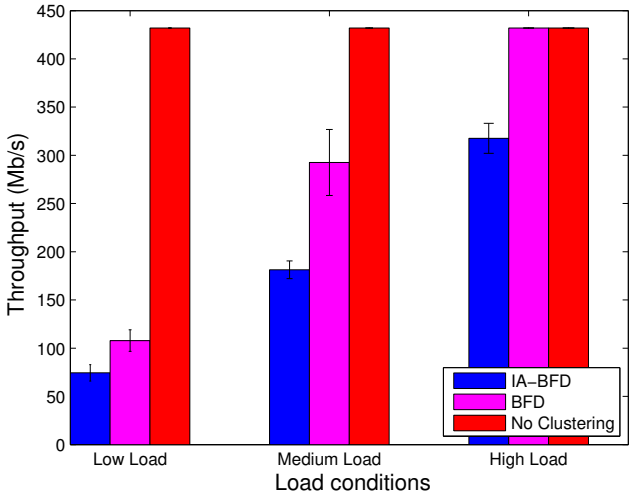


FIGURE 4.8: Total Throughput: IA-BFD vs. BFD vs. No Clustering

4.5.3 IA-BFD vs. IA-WFD vs. IA-FFD vs. Rand Load-Rand load vs. Rand-Min Pow vs. Max Load-Min Pow vs. Min Load-Min Pow

In this subsection, we compare the different heuristics inspired from the Bin Packing heuristics (cf. section 4.2), deemed IA-BFD, IA-WFD (Interference-Aware WFD), and IA-FFD (Interference-Aware FFD), and the greedy heuristics described in section 4.3. The greedy heuristics are denoted as “Rand Load-Rand Load” (Random Load-Random Load), “Rand Load-Min Power” (Random Load-Minimum Power), “Max Load-Min Power” (Maximum Load-Minimum Power) and “Min Load-Min Power” (Minimum Load-Minimum Power). The results are displayed for three load conditions, as stated in subsection 4.5.2, and metrics are averaged and shown with 95% confidence interval. Simulations are repeated for 2000 runs. The same metrics of the previous subsection are used to compare the results. Further, the execution time of all the heuristics is assessed.

Figure 4.9 shows the number of active BBUs as a function of the load conditions. We notice that for low load, the greedy heuristics activate 1 BBU, whereas the Interference-Aware Bin Packing heuristics activate 2 BBUs. In fact, within the IA-Bin Packing schemes, RRHs are packed in a way to verify some requirements, and decisions are taken one shot and do not consider the change of interference level due to the added RRHs. This is contrarily to the greedy heuristics that search among all the unmapped RRHs at each step, and choose the best suited one in terms of power consumption. For medium and high load conditions, all the heuristics show almost the same number of activated BBUs.

Figures 4.10 and 4.11 show the power consumption and the realized power saving of all heuristics as a function of the load conditions. For low load, we notice that the power consumption shown by the IA-Bin Packing heuristics is higher than that of the greedy heuristics since one additional BBU is activated. Similarly, the realized power saving metric is slightly less than 90% for the IA-Bin Packing scheme, whereas it slightly more than 90% for the greedy heuristics. The power consumption and the realized power saving at medium and high load show similar values since the number of activated BBUs for all heuristics is similar.

Figure 4.12 shows the energy efficiency realized by all the heuristics as a function of the load conditions. For low load, we notice that the greedy heuristics show significant enhanced energy efficiency in comparison with the IA-Bin Packing heuristics, although one additional BBU is activated in the IA-Bin Packing case. The reason is the absence of inter-cluster interference for the greedy heuristics, since all the RRHs are associated to one single BBU. For medium and high load conditions, the energy efficiency levels shown by all heuristics are similar. We recall that when more BBUs are activated, the interference level increases significantly until a stable level is reached. In that stage, when more BBUs are activated, the interference level would be slightly affected, leading to similar values of energy efficiency as shown in the cases of medium and high load conditions, where an average of 8 and 14 BBUs are activated respectively.

Figure 4.13 displays the mean throughput per UE realized by all schemes. We notice that the greedy heuristics show higher values of mean throughput per UE in comparison with the IA-Bin Packing, although one additional BBU is activated for the case of IA-Bin Packing. The reason is the absence of the inter-cluster interference in the greedy

heuristics, leading to enhanced spectral efficiency, that contributes to higher derived throughput within one BBU (cf. Figure 4.14). For medium and high load conditions, the heuristics show similar values of mean throughput per UE since the radio conditions are stable and the number of active BBUs is the same.

Figure 4.14, shows the spectral efficiency per BBU as a function of the load conditions. Similar to the energy efficiency, the spectral efficiency per BBU is enhanced at low load conditions for the greedy heuristics, in comparison with the IA-Bin Packing heuristics. The reason is the absence of inter-cluster interference as previously stated. For medium and high load conditions, the level of the spectral efficiency is the same for all heuristics.

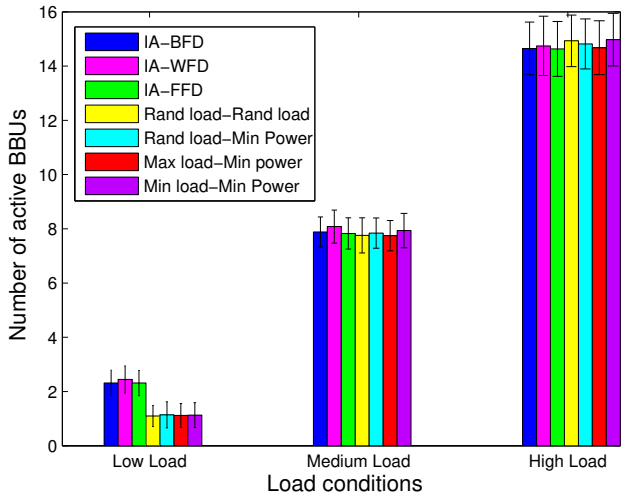


FIGURE 4.9: Number of Active BBUs: IA-BFD vs. IA-WFD vs. IA-FFD vs. Rand Load-Rand load vs. Rand-Min Power vs. Max Load-Min Power vs. Min Load-Min Power

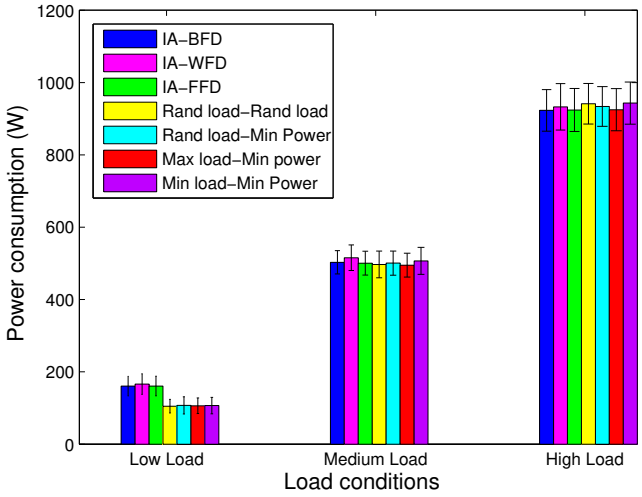


FIGURE 4.10: Power Consumption: IA-BFD vs. IA-WFD vs. IA-FFD vs. Rand Load-Rand load vs. Rand-Min Power vs. Max Load-Min Power vs. Min Load-Min Power

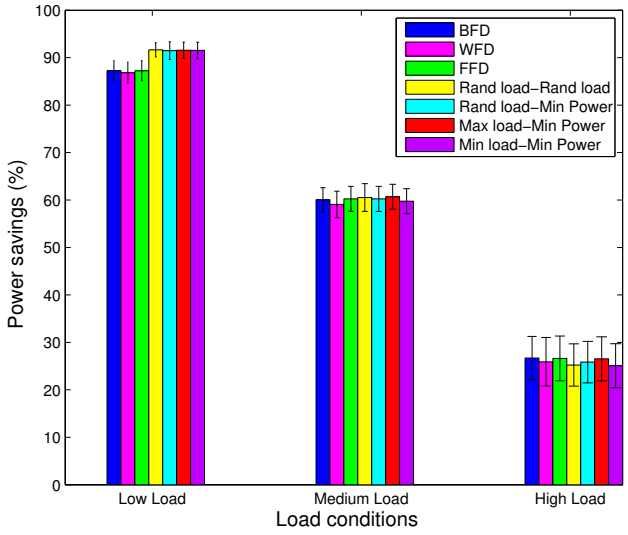


FIGURE 4.11: Power Savings: IA-BFD vs. IA-WFD vs. IA-FFD vs. Rand Load-Rand load vs. Rand-Min Power vs. Max Load-Min Power vs. Min Load-Min Power

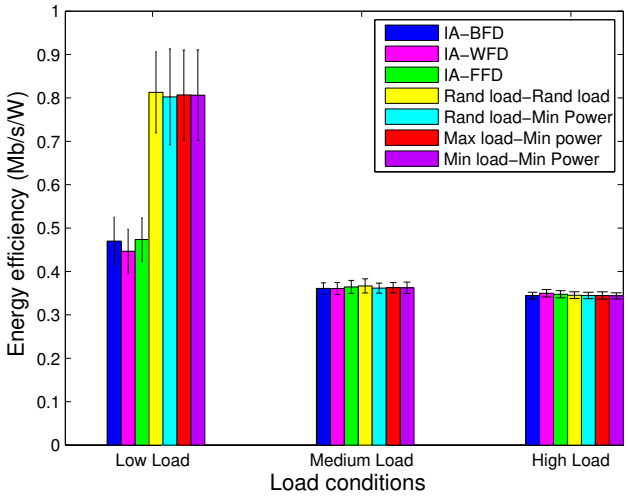


FIGURE 4.12: Energy Efficiency: IA-BFD vs. IA-WFD vs. IA-FFD vs. Rand Load-Rand load vs. Rand-Min Power vs. Max Load-Min Power vs. Min Load-Min Power

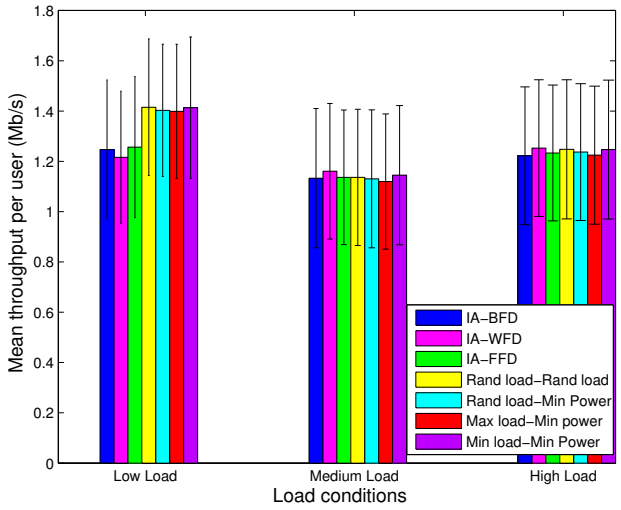


FIGURE 4.13: Mean Throughput per UE: IA-BFD vs. IA-WFD vs. IA-FFD vs. Rand Load-Rand load vs. Rand-Min Power vs. Max Load-Min Power vs. Min Load-Min Power

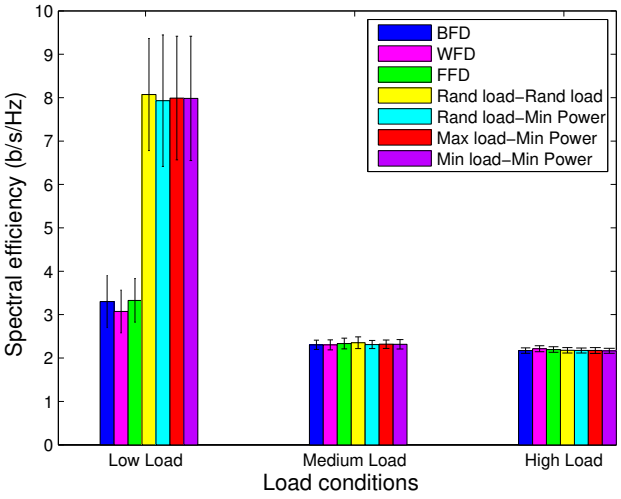


FIGURE 4.14: Spectral Efficiency: IA-BFD vs. IA-WFD vs. IA-FFD vs. Rand Load-Rand load vs. Rand-Min Power vs. Max Load-Min Power vs. Min Load-Min Power

The total derived throughput in the system is shown in Figure 4.15. For low load conditions, we notice that the greedy heuristics show higher total derived throughput than the IA-Bin Packing heuristics. The reason is due to the high level of spectral efficiency, as previously stated. For high and medium load conditions, the total derived throughput is the same for all heuristics since the number of active BBUs and the radio conditions are similar.

The execution time of all the algorithms is displayed in Figure 4.16 as a function of the load conditions. The algorithms are executed over a hardware of 64-bit, Intel 7 core, 2.5 Ghz. We notice that the execution time of the IA-Bin Packing heuristics is less than the greedy heuristics for all load conditions. As previously mentioned in section 4.4, the complexity of the IA-Bin Packing heuristic is $O(R \cdot \log(R))$, which is lower than that of the greedy heuristics with quadratic complexity. We further notice that at low load conditions, the greedy heuristics necessitate more execution time at the cases of medium and high load conditions. In fact, at low load condition, the heuristics need to frequently update the resulted clusters while adding RRHs, which takes more time than in the cases of medium and high load conditions, where the associations of many RRHs to one BBU is less frequent to preserve a minimum level of throughput requirement. Also, the execution time of the "Rand-load-Rand load" heuristic is slightly higher

than that of the “Rand-load-Min Power”, “Min-load-Min Power” and “Max-load-Min Power”. The reason is due to its random behavior. In other terms, in some cases, the “Rand-load-Rand load” performs many drops within the set of unmapped RRHs to find the one that respects the constraint of QoS, contrarily to the cases of “Rand-load-Min Power”, “Min-load-Min Power” and “Max-load-Min Power” that do not have this random behavior.

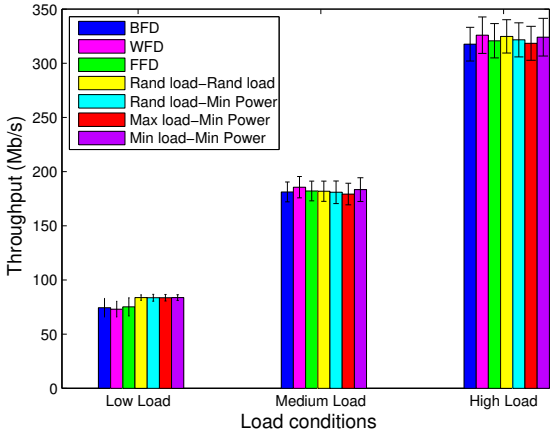


FIGURE 4.15: Total Throughput: IA-BFD vs. IA-WFD vs. IA-FFD vs. Rand Load-Rand load vs. Rand-Min Power vs. Max Load-Min Power vs. Min Load-Min Power

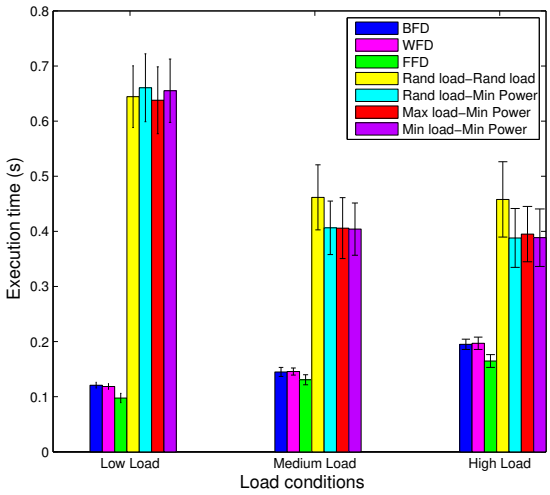


FIGURE 4.16: Execution Time: IA-BFD vs. IA-WFD vs. IA-FFD vs. Rand Load-Rand load vs. Rand-Min Power vs. Max Load-Min Power vs. Min Load-Min Power

4.6 Concluding Remarks

In this chapter, we described different Interference-Aware heuristics that we submitted to the stringency of a real network topology. A first type of heuristics are inspired from the classical heuristics of the Bin Packing Problem (*i.e*, BFD, WFD and FFD), and are adapted to behave in an Interference-Aware manner. A second type consists of greedy heuristics similar to the one described in chapter 3. Two important conclusions are conducted from the obtained results: The first one states that the greedy heuristics derive better power savings in comparison with the IA-Bin Packing heuristics at low load conditions. For medium and high load, the results shown by all heuristics are

similar. The second conducted conclusion concerns the computational complexity of the algorithms. Precisely, the greedy heuristics require more execution time than the Bin Packing heuristics.

Chapitre 5

A Hybrid Approach for BBU-RRH Association Optimization in C-RAN Architecture

In the previous chapters (3 and 4), the proposed schemes consisted in fully centralized approaches, which are usually known for their good results and efficient computational capabilities. However, those solutions require huge amount of signaling overhead since RRHs have to frequently send their load and radio conditions to the cloud, where the metrics are assessed and the decisions are made. The exchange of information might impose constraints over the optical fronthaul connections between BBUs and RRHs. Particularly, when the number of RRHs is high. In this chapter, a different approach of the problem is put forward: Since the assistance of the edge cloud in performing the baseband tasks has been highly motivated in the recent state-of-art, the BBU-RRH association problem is treated in that context. Precisely, we propose a hybrid solution relying on a Game Theoretic framework, where RRHs organize themselves locally into several clusters with the assistance of the edge cloud. We first propose a hybrid solution that adapts to a homogeneous distribution of UEs in the network. We then present an enhanced version of the proposed scheme that adapts to a heterogeneous distribution of UEs in the network, which is more realistic. The second version of our proposition relies on a Load-Aware scheme. In particular, RRHs organize themselves into several clusters, while being aware of the load condition in each BBU. Consequently, we propose a centralized Load- and Interference-Aware solution, that operates in two stages. The potential game among RRHs (i.e, the first stage) is solved via two different algorithms: The first relies on a Best Response dynamics. The second is based on reinforcement learning (i.e, the Replicator Dynamics). An analysis of the signaling load overhead generated by the hybrid scheme is provided in comparison with that generated in centralized solutions. Performance metrics of both versions of the hybrid proposition are also presented in comparison with centralized approaches. The results show that the proposed scheme provides close performance to the centralized approaches with reduced amount of signaling load. Further, a comparative study between the centralized Interference-Aware proposition of chapter 3, as well as the proposed two-stage Load- and Interference-Aware centralized algorithm, is provided to put forward the Load-Aware behavior. We lastly submit our hybrid proposition to the stringency of a real network topology and compare the results with the proposed heuristic in chapter 3.

5.1 Introduction

The BBU-RRH association problem has been formulated in a centralized fashion, requiring periodic exchange of signaling load overhead between BBUs and RRHs in order to make accurate decisions. However, for particular cases, such as massive MIMO and CoMP techniques, the optical fronthaul links might impose strict constraints. For example, to realize the full benefits of CoMP, the CSI exchange between UEs and BBUs is required. As the number of RRHs increases, the CSI signaling will significantly increase, causing a constrained fronthaul that must be taken into consideration [65]. In [71], the authors use an example to show that shifting all the baseband functionalities to the remote BBU pool necessitates approximately 1 Gbps of throughput on the fronthaul link, just to support a 75 Mbps UE data rate. Hence, several works, such as [15], [14], [72] and [73], propose different algorithms to optimally split the different parts of the radio stack between BBUs and RRHs in order to lower the load over the optical

fronthaul connections. Their purpose would be to find a good compromise between multiplexing gain and signaling load reduction. When a few parts of baseband tasks are executed at the edge side of the network, it would be interesting to investigate a BBU-RRH association proposition in a distributed fashion, with the purpose of reducing the burden of signaling load overhead required to perform the mapping decisions. By making use of the system architecture described in section 5.2, we propose that RRHs organize themselves at the edge side of the network, so that they will not periodically send their load and radio conditions to the centralized cloud.
Our proposition relies on two stages: A first stage is used to set the BBU-RRH association schemes via a non-cooperative game among the RRHs that compete over a predefined set of available resources (*i.e*, the BBUs), until the convergence to a Pure Nash Equilibrium (PNE) [74]. In particular, each RRH calculates its own cost function in order to decide to which cluster it should be associated. After organizing themselves into different clusters, the RRHs update their decisions to the centralized cloud, where the throughput per UE realized within each cluster is probed. Accordingly, a BBU is activated/deactivated.
In the rest of the chapter, we describe the system architecture in section 5.2. Section 5.3 introduces a first proposition relying on a hybrid Load-Balancing and Interference-Aware algorithm that is well suited for the case of homogeneous UEs distribution in the network. In section 5.4, an enhanced version of the solution, deemed Load- and Interference-Aware Clustering Algorithm, is introduced to adapt the devised scheme to a heterogeneous distribution of UEs in the network. The centralized Load- and Interference-Aware proposition is provided as well, as a benchmark. In section 5.5, we analyze the needed amount of signaling load overhead that needs to be accounted for in fully centralized and hybrid schemes. Section 5.6 provides firstly the performance evaluation of the Load-Balancing and Interference-Aware proposition, in comparison with the centralized proposition of chapter 3. The hybrid and centralized versions of the Load- and Interference-Aware propositions are secondly provided. We lastly extend the performance of the hybrid scheme to the stringency of a real network topology. Section 5.7 concludes the chapter.

5.2 System Architecture

Traditionally, the C-RAN architecture is based on CPRI to transport the baseband tasks from BBUs to RRHs. Such solution imposes very high bandwidth requirements on the optical fronthaul network. In that context, several working groups, such as the next generation fronthaul interface (NGFI) [75], has been formed to standarize the fronthaul interface for future 5G cellular networks.
According to [75], a BBU is redefined as the Radio Cloud Center (RCC). Further, a Remote Radio System (RRS) is introduced and is composed of the Remote Radio Unit (RRU), or in other terms RRH, and the Radio Aggregation Unit (RAU) (cf. Figure 5.1). The RAU system is composed of a RRU gateway as shown in 5.1. Further, it could embrace an edge BBU pool in some cases where the optical fronthaul is very constrained. Data from RRUs are multiplexed/demultiplexed at the RRU gateways (*i.e*, the edge cloud), which is used to aggregate the traffic from multiple cell sites, where several baseband functions, such as inter-site coordination functionalities, can be placed. The different parts of the baseband processing are performed at the respective endpoint depending on the functional split. NGFI whitepaper [75] describes several ways to perform the splitting between RRU and RRC (cf. Figure 5.2). The horizontal numbered lines represent the different splitting schemes : The functionalities below the lines will be executed at the RRS side, while those above the lines will be executed at the RCC side.

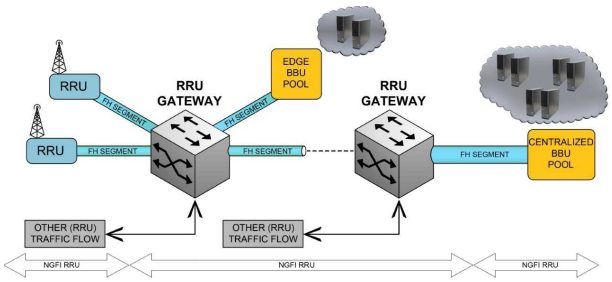


FIGURE 5.1: C-RAN Network Topology according to NGFI

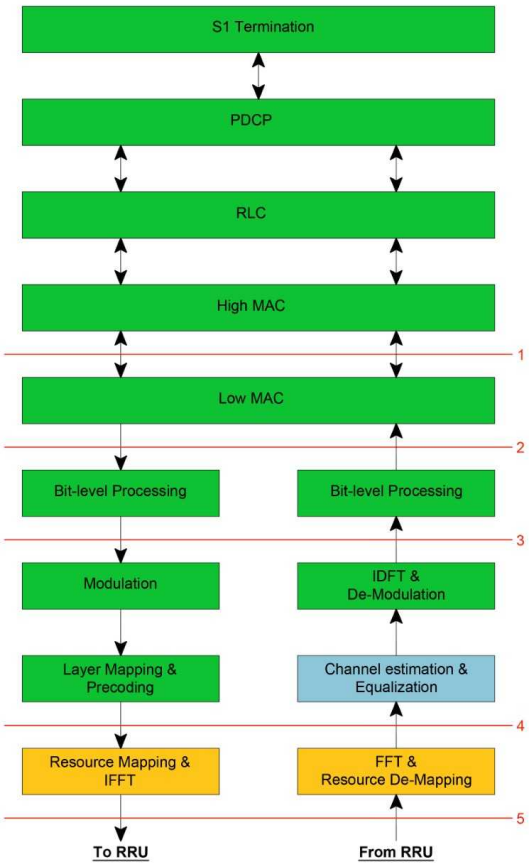


FIGURE 5.2: Different splits of RRS/RCC

Based on this architecture, we propose to make use of the edge cloud in order to collect the data that concern the load conditions of each RRH, helping them to locally organize themselves into a predefined set of clusters, in order to lower the signaling load overhead over the optical fronthaul links, connecting the RRS to BBUs.

5.3 A Load-Balancing and Interference-Aware Scheme for Homogeneous Distribution of UEs in the Network

This section describes a first proposition that adapts to a homogeneous distribution of UEs in the network. In other terms, the number of UEs within each cell is the same. The main objective of our devised scheme is to minimize the number of active BBUs, while guaranteeing a minimum level of throughput requirement per UE. Initially, the centralized cloud (*i.e*, RRC) sets the number of available BBUs.

— In the first stage, a potential game is played among the RRHs to associate to a given BBU. Each RRH reduces selfishly its own Load-Balancing and Interference-Aware cost function. The latter is affected with the RRHs that choose concurrently the same BBUs (partaking the common radio resources) but also, with those RRHs that joined other BBUs through their harmful interference impact.
— In the second stage, the centralized cloud decides whether to activate or to turn OFF a BBU according to the realized throughput per BBU. The two stages are iteratively executed until the number of active BBUs is decided.

Hybrid Load-Balancing and Interference-Aware Scheme Based on Game Theory

The two stages of our hybrid proposition are described within the following paragraphs. After attaining the PNE in the first stage, the second stage is unfolded, where the number of available BBUs can be modified to strike a compromise between reducing the power consumption and guaranteeing a minimal level of throughput for serviced UEs.

First Stage: Non-Cooperative Game for BBU-RRH Association

Non-cooperative game theory models the interactions between players competing for a common resource. Hence, it is well adapted to BBU-RRH association problem. We define a multi-player game $\mathcal{G}$ between R RRHs. The RRHs are assumed to make their decisions without knowing the decisions of each other.
We present the general framework of the game $\mathcal{G} = (\mathcal{R}, \mathcal{C}, Cost)$ that can be described as follows:

- The set $\mathcal{R}$ of RRHs as the set of players.
- The set $\mathcal{P}$ of available BBUs. An action of a RRH is selecting one of the available BBUs.
- The strategy of RRH r is denoted by the vector y_r, whose components are $y_{r,c}$ which define whether RRH r is associated to BBU c. Hence, $\mathbf{y} = (y_1, ..., y_R)$ is a pure strategy profile, and $\mathcal{C} = C_1 \times C_2 \times ... \times C_R$ is the space of all profiles.
- A set of cost functions $\{Cost_1, Cost_2, ..., Cost_R\}$ that quantify the players' preferences over the possible outcomes of the game. Outcomes are determined by a particular action chosen by RRH r, and the particular actions chosen by all other players r'.

Cost Function Each RRH seeks to minimize its own Load-Balancing and Interference-Aware cost function by choosing an adequate strategy (*i.e,* BBU c). The cost function of RRH r is given by what follows:

$$Cost_r(y_r, y_{-r}) = \sum_{c \in \mathcal{P}} y_{r,c} \cdot \big[\alpha \cdot \alpha' \cdot \underbrace{\sum_{r' \in \mathcal{R}} y_{r',c}}_{\text{Term 1}} + \beta \cdot \beta' \cdot \underbrace{\big(\sum_{r' \neq r} G_{r',r} \cdot (1 - y_{r',c}) + N_0\big)}_{\text{Term 2}}\big], \tag{5.1}$$

where $G_{r,r'}$ is the channel gain between RRHs r and r', and $y_{r,c}$ is the binary variable that is equal to one if BBU c is chosen.
The cost sustained by a given RRH r in any selected BBU depends upon the congestion impact inflected by RRHs r' associated to BBU c, sharing the same radio resources of the common BBU c. The latter is depicted by Term 1 in (5.1) and acts as a Load-Balancing function among active BBUs. Furthermore, the cost function should encompass the interference impact. Accordingly, Term 2 reflects the intra-cluster interference which is canceled between RRH r and other RRHs that have chosen the same strategy c.
The Load-Balancing cost is ponderated by α, and the interference cost is ponderated by β. We note that $\alpha + \beta = 1$. α' and β' are normalization factors.

The Pure Nash Equilibrium In a non-cooperative game, a solution is obtained when all players adhere to a NE. A NE is a profile of strategies in which no player will profit from deviating its strategy unilaterally. Hence, it is a strategy profile, where each player strategy is an optimal response to other players' strategies.
Our game $\mathcal{G}$ is a finite game and in general such games are not guaranteed to have a PNE. Nevertheless, they possess a mixed NE where each player has to continually change its BBU selection according to a distribution probability over the strategy set. However, for our game an exact potential function [76] exists, which means that the unilateral change of one RRH strategy y_r to y'_r results in a change of its cost function that is equal to the change of a so-called potential function $\phi : \mathcal{S}^R \to \mathbb{R}$.
An example of such a potential function is given by the following:

$$\phi(\mathbf{y}) = \sum_{c \in \mathcal{P}} \sum_{r \in \mathcal{R}} y_{r,c} \big(\frac{1}{2} \cdot \sum_{r' \neq r} y_{r',c} + y_{r,c} + \frac{1}{2} \cdot \sum_{r' \neq r} \big(2 - y_{r',c}\big) G_{r',r} + N_0\big), \tag{5.2}$$

Furthermore, such games (*i.e*, potential games) [77] have the appealing property of admitting at least PNE which is a desired property in this context for practical reasons.

Proof: We prove hereafter that the game at hand is an exact potential game. We simply need to show that if **y** and **y**′ are two profiles which only differ in the strategy of one RRH r, then:

$$Cost_r(y_r, y_{-r}) - Cost_r(y'_r, y_{-r}) = \phi(y_r, y_{-r}) - \phi(y'_r, y_{-r}) \tag{5.3}$$

We begin by computing $Cost_r(y_r, y_{-r}) - Cost_r(y'_r, y_{-r}) =$

$$\begin{aligned}
&\sum_c y_{r,c} \cdot \big[\alpha\alpha' \cdot \big(\sum_{r'\neq r} y_{r',c} + y_{r,c}\big) + \beta\beta' \cdot \big(\sum_{r'\neq r} G_{r',r}(1 - y_{r',c}) + N_0\big)\big] \\
&- \sum_c y'_{r,c} \cdot \big[\alpha\alpha' \cdot \big(\sum_{r'\neq r} y_{r',c} + y'_{r,c}\big) + \beta\beta' \cdot \big(\sum_{r'\neq r} G_{r',r}(1 - y_{r',c}) + N_0\big)\big] \\
&= \sum_c \big(y_{r,c} - y'_{r,c}\big) \cdot \big[\alpha\alpha' \cdot \big(\sum_{r'\neq r} y_{r',c} + (y_{r,c} + y'_{r,c})\big) + \beta\beta' \cdot \big(\sum_{r'\neq r} G_{r',r}\big(1 - y_{r',c}\big) + N_0\big)\big]
\end{aligned} \tag{5.4}$$

Now we rewrite $\phi(\mathbf{y})$ for clarity :

$$\begin{aligned}
\phi(\mathbf{y}) = &\sum_{r'\neq r} \big(\sum_c y_{r',c} \cdot \big(\alpha\alpha' \cdot \big(\frac{1}{2} \cdot \sum_{j\neq r, j\neq r'} y_{j,c} + \frac{1}{2} y_{r,c} + y_{r',c}\big) + \\
&\beta\beta' \cdot \big(\frac{1}{2} \cdot \sum_{j\neq r, j\neq r'} \big(2 - y_{j,c}\big) G_{j,r} + \frac{1}{2} \cdot \big(2 - y_{r,c}\big) G_{r,r'} + N_0\big)\big) \\
&+ \sum_c y_{r,c} \cdot \big(\alpha\alpha' \cdot \big(\frac{1}{2} \cdot \sum_{j\neq r} y_{j,c} + y_{r,c}\big) + \beta\beta' \cdot \big(\frac{1}{2} \cdot \sum_{j\neq r} (2 - y_{j,c}) G_{j,r} + N_0\big)\big)
\end{aligned} \tag{5.5}$$

We finally compute $\phi(y_r, y_{-r}) - \phi(y'_r, y_{-r}) =$

$$\begin{aligned}
&\sum_c \sum_{r'\neq r} y_{r',c} \big[\big(\alpha\alpha' \cdot \big(\frac{1}{2} \cdot \sum_{j\neq r, j\neq r'} y_{j,c} + \frac{1}{2} y_{r,c} + y_{r',c}\big) + \beta\beta' \cdot \big(\frac{1}{2} \cdot \sum_{j\neq r', j\neq r} \big(2 - y_{j,c}\big) G_{j,r} + \frac{1}{2} \cdot \big(2 - y_{r,c}\big) G_{r,r'} + N_0\big)\big)\big] \\
&- \sum_c \sum_{r'\neq r} y'_{r',c} \big[\big(\alpha\alpha' \cdot \big(\frac{1}{2} \cdot \sum_{j\neq r, j\neq r'} y'_{j,c} + \frac{1}{2} y'_{r,c} + y'_{r',c}\big) + \beta\beta' \cdot \big(\frac{1}{2} \cdot \sum_{j\neq r', j\neq r} \big(2 - y'_{j,c}\big) \cdot G_{j,r} + \frac{1}{2} \cdot \big(2 - y'_{r,c}\big) G_{r,r'} + N_0\big)\big)\big] \\
&+ \sum_c \big(y_{r,c} - y'_{r,c}\big) \cdot \big[\alpha\alpha' \cdot \big(\frac{1}{2} \cdot \sum_{j\neq r} y_{j,c} + (y_{r,c} + y'_{r,c})\big) + \beta\beta' \cdot \big(\frac{1}{2} \cdot \sum_{j\neq r} \big(2 - y_{j,c}\big) G_{j,r} + N_0\big)\big]
\end{aligned}$$

As $y_{j,c} = y'_{j,c}$ and $y_{r',c} = y'_{r',c}$ since **y** and **y**′ only differ in the strategy of RRH r, we have

$$\begin{aligned}
&= \sum_c \big(y_{r,c} - y'_{r,c}\big) \cdot \big(\alpha\alpha' \cdot \frac{1}{2} \cdot \sum_{r'\neq r} y_{r',c} - \frac{1}{2} \cdot \beta\beta' \cdot \sum_c \big(y_{r,c} - y'_{r,c}\big) \cdot \big(\sum_{r'\neq r} y_{r',c} \cdot G_{r,r'}\big)\big) \\
&+ \sum_c \big(y_{r,c} - y'_{r,c}\big) \cdot \big[\alpha\alpha' \cdot \big(\frac{1}{2} \cdot \sum_{r'\neq r} y_{r',c} + (y_{r,c} + y'_{r,c})\big) + \beta\beta' \cdot \big(\frac{1}{2} \cdot \sum_{r'\neq r} G_{r,r'}\big(2 - y_{r',c}\big) + N_0\big)\big] \\
&= \sum_c \big(y_{r,c} - y'_{r,c}\big) \cdot \big[\alpha\alpha' \cdot \big(\sum_{r'\neq r} y_{r',c} + (y_{r,c} + y'_{r,c})\big) + \beta\beta' \cdot \big(\sum_{r'\neq r} G_{r,r'}\big(1 - y_{r',c}\big) + N_0\big)\big]
\end{aligned} \tag{5.6}$$

As $G_{r,r'} = G_{r',r}$ (impact of RRH r on RRH r' is the same as impact of RRH r' on r), we have indeed:
$Cost_r(y_r, y_{-r}) - Cost_r(y'_r, y_{-r}) = \phi(y_r, y_{-r}) - \phi(y'_r, y_{-r})$.

Attaining the Pure Nash Equilibrium

In our work, we resort to two different algorithms to reach PNE. The first, namely the Best Response dynamics [78], guarantees the convergence to a PNE. The second is based on a reinforcement learning algorithm, namely the Replicator Dynamics [79], where the Nash equilibriums are learned through the game.

Best Response algorithm: A Best Response dynamics scheme consists of a sequence of iterations, where each player r chooses the best response to the other players' actions in the previous iteration. In particular, at each iteration t, a RRH chooses the cluster that provides it the lowest cost (5.1). Convergence is attained when all RRHs

choose the same strategies as in the previous iteration. In the first iteration, the choice of each player is the Best Response based on its arbitrary belief about what the other players will choose. We denote by $a_r(t)$ and $y_r(t)$ the action and the pure strategy of RRH r at iteration t respectively. The behavior of the Best Response is described in algorithm 2.

Initialize all vectors $y_r(0)$ of pure strategies;
repeat
  Each RRH selects a startegy c^* that respects $c^* = \text{argmin}_{c \in \mathcal{P}} Cost_r(t)$;
  Actions $a_{r \in \mathcal{R}}(t)$ of each RRH are updated;
until $y_r(t-1) = y_r(t) \forall r \in \mathcal{R}$;

Algorithm 2: Best Response Dynamics

In Best Response dynamics, players need to have a local view about other players choices. Consequently, at each iteration t, the edge cloud sends all vectors $y_{r \in \mathcal{R}}$ of pure strategies to all players (*i.e*, RRHs). According to different players' choices, each RRH makes its own choice. In that case, a RRH must be equipped with enough storage, in order to save all vectors $y_{r \in \mathcal{R}}$ of pure strategies.

Replicator Dynamics algorithm: We denote by p_r the vector of mixed strategy, corresponding to a probability distribution over the set of pure strategies, and whose components are $p_{r,c} \in [0,1]$, defining the probability that RRH r chooses BBU c, where $\sum_{c \in \mathcal{P}} p_{r,c} = 1$. Let K_r be the simplex of mixed strategies for RRH r. Any pure strategy y_r can be considered as a mixed strategy p_r, where vector $p_{r \in \mathcal{R}}$ denotes the unit probability vector with s^{th} component being a unity component, hence a corner of K_r. Let $\mathbb{K} = K_1 \times K_2 \times ... \times K_R$ be the space of all mixed strategies. A strategy $\mathbf{p} = (p_1, ..., p_R) \in \mathbb{K}$ specifies the (mixed or pure) strategies of all players. Following classical convention, we write $\mathbf{p} = (p_r, p_{-r})$, where p_{-r} denotes the vector of strategies played by all RRHs besides RRH r. The game mechanics work as follows: At $t = 0$, we begin by a uniform distribution of each vector $p_r(0)$, where each of its component $p_{r,c}(0)$ is set to $\frac{1}{|\mathcal{P}|}$, where $|\mathcal{P}|$ is the number of available strategies. At each iteration $t > 0$:

— Each RRH r chooses an action $a_r(t)$ according to probability distribution $p_r(t)$.
— Each RRH r learns the cost $Cost_r(t)$ resulting from the association of RRH r to cluster c, and from the set of other players' actions.
— Each RRH r updates the components $p_{r,c}(t+1)$ of its vector $p_r(t+1)$ in the following way:

$$p_{r,c}(t+1) = \begin{cases} p_{r,c}(t) + k(1 - \frac{Cost_r(t)}{Cost_{max}})(1 - p_{r,c}(t)) \\ \text{if} \quad a_r(t) = c \in \mathcal{C}, \\ p_{r,c}(t) - k(1 - \frac{Cost_r(t)}{Cost_{max}}) p_{r,c}(t) \\ \text{otherwise}, \end{cases} \tag{5.7}$$

where $0 < k < 1$ is a parameter and $Cost_{max}$ is the maximum cost. The behavior of the Replicator Dynamics is described in algorithm 3.

Theorem 1 *Let G be an instance of game $\mathcal{G}$. Let $\mathbb{K}^*$ be a set of mixed profiles, where at most one player plays a pure strategy. The learning algorithm, for any initial condition in $\mathbb{K} - \mathbb{K}^*$, always weakly converges to a Nash Equilibrium.*

Proof of convergence of Theorem 1 is given in the appendix A at the end of this thesis.

Algorithms of this form are fully distributed, and players do not need to know any information about each other. For example, vectors p_r are not sent by the edge cloud for RRHs. In fact, at any iteration t, RRH r only needs to know its own cost $Cost_r(t)$ and mixed strategy $p_r(t)$. To compute $Cost_r(t)$, the load and the radio condition of a chosen strategy need to be known by RRH r (cf. equation (5.1)). The load condition is sent by the edge cloud at each iteration t, as for the radio condition (Term 2), any RRH r makes use of the signaling information already present in the downlink of a system. For example, in the case of LTE, a UE assigned to a specific RB, measures its channel quality based on pilots (*i.e*, Cell Specific Reference Signals (CRS)) that are spread across the whole band independently of the individual UE allocation. Hence, Term 2 can be easily inferred from the CQI (Channel Quality Indicator) sent every TTI (Transmit Time Interval) by serviced UEs. In that case, less storage on RRHs side is needed, since each RRH only needs to store its own vector p_r of mixed strategy.

Initialize all vectors $p_r(0)$ of mixed strategies to uniform distribution;
Each RRH determines its own action $a_r(0)$;
repeat
    Each RRH selects a BBU c and receives a cost $Cost_r(t)$;
    Actions $a_r(t)$ of each RRH are updated;
    Vectors $p_r(t+1)$ of mixed strategies are updated and each component $p_{r,c}(t+1) \in p_r(t+1)$ is set to :

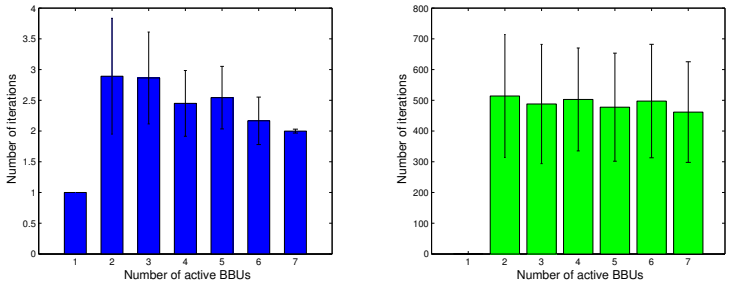
$$p_{r,c}(t+1) = \begin{cases} p_{r,c}(t) + k(1 - \frac{Cost_r(t)}{Cost_{max}})(1 - p_{r,c}(t)) \\ \text{if} \quad a_r(t) = c \in \mathcal{C}, \\ p_{r,c}(t) - k(1 - \frac{Cost_r(t)}{Cost_{max}})p_{r,c}(t) \\ \text{otherwise}, \end{cases} \tag{5.8}$$

until *Learning NE*;

Algorithm 3: Replicator Dynamics

Speed of Convergence We assessed through extensive simulations the convergence of algorithm 2 and algorithm 3 to PNE. We are interested in evaluating the speed of convergence of both algorithms for two different network topologies: The first consists of 7 RRHs. The second is based on a real network topology consisting in 20 RRHs [23]. The mean number of iterations to reach convergence is drawn as a function of the number of active BBUs (*i.e*, number of available strategies). The number of UEs per cell is fixed to 6 in the case of 7 RRHs and to 3 for the topology of the network of 20 RRHs. 1000 simulations are run for one scenario (*i.e*, fixed number of active BBUs).

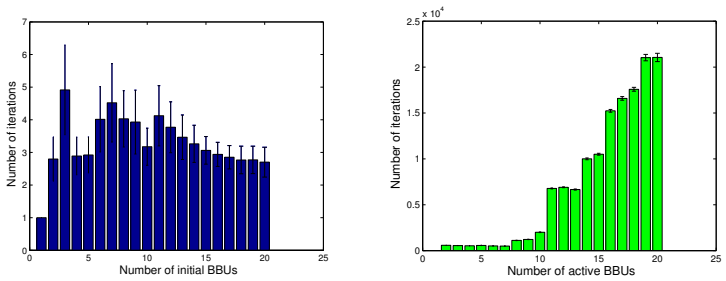
Figures 5.3(a) and 5.3(b) display the speed of convergence for the topology of 7 RRHs. We notice that the Best Response needs less number of iterations for convergence, and for all number of active BBUs in comparison with the Replicator Dynamics. For example, the Replicator Dynamics algorithm needs around 500 iterations for convergence in some cases, whereas the Best Response needs at most 3 iterations. Recall that the Replicator Dynamics is based on the reinforcement learning technique. Such algorithms are known to be slow in convergence. However, they behave independently and require less informations (and storage) than other algorithms (the Best Response algorithm).



(a) Best Response (b) Replicator Dynamics

FIGURE 5.3: Number of Iterations for 7 RRHs

Figures 5.4(a) and 5.4(b) display the speed of convergence for the topology of 20 RRHs. In this case, we notice that the Best Response algorithm needs significant less number of iterations than the Replicator Dynamics does. In particular, when the number of strategies increases. In fact, for high number of strategies, the convergence of the Replicator Dynamics is very slow. This is normal since it is based on reinforcement learning techniques, which are known to be slow in convergence. Nerveless, such solutions behave independently, and require less assistance. However, when the time for convergence significantly increases, semi-distributed approaches, such as the Best Response remain better solutions.



(a) Best Response (b) Replicator Dynamics

FIGURE 5.4: Number of Iterations for 20 RRHs

Second Stage: BBU Activation/Deactivation

In the second stage, the centralized cloud examines the lowest throughput per UE realized among BBUs, and verifies if constraint (5.9) is respected or not:

$$T(c) \geq n_c \cdot D_{min}, \forall c \in \mathcal{P} \tag{5.9}$$

Constraint (5.9) states that the throughput per UE should stay greater than a minimal threshold, denoted by D_{min}. Recall that $T(c)$ is the derived throughput in cluster c, and n_c the number of UEs in cluster c.
If constraint (5.9) is violated, the centralized cloud activates an additional BBU to lower the load per BBU. Accordingly, the second stage is launched, and the RRHs start competing again over the BBUs made available for them. Contrarily, if the lowest throughput per UE respects constraint (5.9), the centralized cloud decreases the number of active BBUs to reduce the power consumption, and relaunches the game again. The algorithm stops running and outputs a solution when it deactivates a BBU and activates it back again. More precisely, when the algorithm deactivates a BBU and realizes that constraint (5.9) is not respected, it falls back to the previous number of active BBUs (before the deactivation) and sets back the previous associations. We note that the centralized cloud sets the initial number of strategies to $\frac{R}{2}$.

5.4 A Load- and Interference-Aware Scheme for Heterogeneous UEs Distribution

The Load-Balancing and Interference-Aware proposition suits for a homogeneous UEs distribution in the network. In particular, this proposition aims to balance the number of RRHs among the activated BBUs, which is valid when the number of UEs connected to each RRH is similar. However, when a heterogeneous UEs distribution in the network is considered, a RRH needs to be aware of the load condition of its chosen strategy (*i.e*, BBU) in order to make an accurate decision. Consequently, in this section, we present an enhanced version of the cost function for RRHs, aiming to make a RRH aware of the load condition of its chosen strategy.
In order to assess the impact of the Load-Awareness in BBU-RRH association decisions, we propose a fully centralized version of the hybrid approach, where the BBU-RRH association decisions are set within the centralized cloud. The centralized approach consists of two stages: A first stage is used to decide the number of active BBUs. A set of candidate partitions, minimizing the total number of active BBUs and respecting a minimum level of QoS, is selected. A second stage consists in minimizing the total cost function of all RRHs among the preselected set of partitions. In section 5.6, a comparison between the centralized proposition of chapter 3, that only minimizes the total power consumption metric in the system, and the Load- and Interference-Aware centralized proposition is provided in paragraph 5.6.2, in order to evaluate the benefit brought by the Load-Awareness. We then compare between the centralized Load- and Interference-Aware proposition that results in a global optimal solution, and the hybrid Load- and Interference-Aware proposition.

5.4.1 Hybrid Load- and Interference-Aware Scheme Based on Game Theory

The algorithm framework of the hybrid Load- and Interference-Aware scheme is similar to the previously proposed algorithm in section 5.3. In particular, the algorithm consists of two stages: A first stage, which is a non-cooperative game played among the RRHs to associate to BBUs, and a second stage where the number of active BBUs is set.
In the first stage, the general framework of the game is the same as the previously described one in subsection 5.3. In particular, the set of players $\mathcal{R}$ and the space of all the pure strategies profiles $\mathcal{C}$ remain the same. An action of a RRH also consists in selecting one of the available BBUs in the set $\mathcal{P}$. As for the set of cost functions $\{Cost_1, Cost_2, ..., Cost_R\}$, a RRH cost is expressed as follows:

$$Cost_r(y_r, y_{-r}) = \sum_{c\in\mathcal{P}} y_{r,c}\big[\alpha\alpha'\big(\underbrace{\sum_{r'} n_{r'}y_{r',c}}_{\text{Term 1}} + \underbrace{n_r \sum_{r'\neq r} y_{r',c}}_{\text{Term 2}}\big) + \beta\beta'\underbrace{\big(\sum_{r'\neq r} G_{r',r}(1-y_{r',c}) + N_0\big)}_{\text{Term 3}}\big] \tag{5.10}$$

Terms 1 and 2 reflect the cost of RRH r, affected by the congestion impact of other RRHs that simultaneously share the radio resources of cluster c. In particular, term 1 is the total number of UEs served by RRHs r' associated to c, whereas term 2 works as a stressing factor to emphasize the relative load impact of RRH r onto other RRHs $r' \neq r$ when its load n_r is high on one hand, and when the number of RRHs associated to the same BBU is high, on the other hand. Term 3 expresses the inter-cluster interference which is cancelled between RRH r and other RRHs that have chosen the same BBU, as previoulsy explained.
The game among the RRHs remains an exact potential game since a potential function defined as in (5.11) exists, and its change is equal to the change of the cost function in (5.10) when the stategy of RRH r changes from y_r to y'_r :

$$\phi(\mathbf{y}) = \sum_{c\in\mathcal{P}}\sum_{r\in\mathcal{R}} y_{r,c}\big(\sum_{r'} y_{r',c}n_{r'} + \frac{1}{2}\sum_{r'\neq r}(2-y_{r',c})G_{r',r} + N_0\big) \tag{5.11}$$

Proof: If $\mathbf{y}$ and $\mathbf{y}'$ are two profiles which only differ in the strategy of one RRH r, then:

$$Cost_r(y_r, y_{-r}) - Cost_r(y'_r, y_{-r}) = \phi(y_r, y_{-r}) - \phi(y'_r, y_{-r}) \tag{5.12}$$

We begin by computing $Cost_r(y_r, y_{-r}) - Cost_r(y'_r, y_{-r}) =$

$$\begin{aligned}
&\sum_c y_{r,c}\cdot\big[\alpha\alpha'\big(\sum_{r'\neq r} n_{r'}\cdot y_{r',c} + n_r\cdot y_{r,c} + n_r\cdot\sum_{r'\neq r} y_{r',c}\big) + \beta\beta'\big(\sum_{r'\neq r} G_{r',r}(1-y_{r',c}) + N_0\big)\big] - \\
&\sum_c y'_{r,c}\cdot\big[\alpha\alpha'\cdot\big(\sum_{r'\neq r} n_{r'}\cdot y_{r',c} + n_r\cdot y'_{r,c} + n_r\cdot\sum_{r'\neq r} y_{r',c}\big) + \beta\beta'\big(\sum_{r'\neq r} G_{r',r}(1-y_{r',c}) + N_0\big)\big] \\
&= \sum_c \big(y_{r,c} - y'_{r,c}\big)\cdot\big[\alpha\alpha'\cdot\big(\sum_{r'\neq r}(n_r + n_{r'})\cdot y_{r',c} + n_r\cdot(y_{r,c} + y'_{r,c})\big) + \beta\beta'\big(\sum_{r'\neq r} G_{r',r}\big(1-y_{r',c}\big) + N_0\big)\big]
\end{aligned} \tag{5.13}$$

Now we rewrite $\phi(\mathbf{y})$ for clarity:

$$\begin{aligned}
\phi(\mathbf{y}) &= \sum_{r'\neq r}\sum_c y_{r',c}\cdot\alpha\alpha'\cdot\big(\sum_{j\neq r} y_{j,c}\cdot n_j + y_{r,c}\cdot n_r\big) + \frac{1}{2}\cdot\beta\beta'\cdot\big(\sum_{j\neq r, j\neq r'}\big(2-y_{j,c}\big)G_{j,r'} + \frac{1}{2}\cdot\big(2-y_{r,c}\big)G_{r,r'} + N_0\big) \\
&+ \sum_c y_{r,c}\cdot\alpha\alpha'\cdot\big(\sum_{j\neq r} y_{j,c}\cdot n_j + y_{r,c}\cdot n_r\big) + \frac{1}{2}\cdot\beta\beta'\cdot\big(\sum_{j\neq r}\big(2-y_{j,c}\big)G_{j,r} + N_0\big)
\end{aligned} \tag{5.14}$$

We finally compute $\phi(y_r, y_{-r}) - \phi(y'_r, y_{-r}) =$

$$\sum_c \sum_{r' \neq r} y_{r',c} \cdot \big[\alpha\alpha' \cdot \big(\sum_{r' \neq r} y_{r',c} \cdot n_{r'} + y_{r,c} \cdot n_r\big) + \frac{1}{2} \cdot \beta\beta' \cdot \big(\sum_{r' \neq r} \big(2 - y_{r',c}\big) G_{r,r'} + N_0\big)\big] -$$

$$\sum_c \sum_{r' \neq r} y'_{r',c} \cdot \big[\alpha\alpha' \big(\sum_{r' \neq r} y'_{r',c} \cdot n_{r'} + y'_{r,c} \cdot n_r\big) + \beta\beta' \cdot \frac{1}{2} \cdot \big(\sum_{r' \neq r} \big(2 - y'_{r',c}\big) \cdot G_{r,r'} + N_0\big)\big] +$$

$$\sum_c \alpha\alpha' \big(y_{r,c} - y'_{r,c}\big) \cdot \big[\big(\sum_{j \neq r} y_{j,c} \cdot n_j + n_r(y_{r,c} + y'_{r,c})\big) + \beta\beta' \cdot \big(\frac{1}{2} \cdot \sum_{j \neq r} \big(2 - y_{j,c}\big) \cdot G_{j,r} + N_0\big)\big]$$

As $y_{r',c} = y'_{r',c}$ since **y** and **y**′ only differ in the strategy of RRH r, we have (5.15)

$$= \sum_c \alpha\alpha' \big(y_{r,c} - y'_{r,c}\big) \cdot \big(\sum_{r' \neq r} y_{r',c} \cdot n_r\big) - \sum_c \beta\beta' \cdot \frac{1}{2} \cdot \big(y_{r,c} - y'_{r,c}\big) \cdot \big(\sum_{r' \neq r} y_{r',c} \cdot G_{r,r'}\big) +$$

$$\sum_c \big(y_{r,c} - y'_{r,c}\big) \cdot \big[\alpha\alpha' \cdot \big(\sum_{r' \neq r} y_{r',c} \cdot n_{r'} + n_r(y_{r,c} + y'_{r,c})\big) + \beta\beta' \cdot \big(\frac{1}{2} \cdot \sum_{r' \neq r} G_{r,r'} \big(2 - y_{r',c}\big) + N_0\big)\big]$$

$$= \sum_c \big(y_{r,c} - y'_{r,c}\big) \cdot \big[\alpha\alpha' \cdot \big(\sum_{r' \neq r} y_{r',c} \cdot \big(n_r + n_{r'}\big) + n_r(y_{r,c} + y'_{r,c}) + \beta\beta' \big(\sum_{r' \neq r} G_{r,r'} \big(1 - y_{r',c}\big) + N_0\big)\big]$$

As $G_{r,r'} = G_{r',r}$ we have:

$$Cost_r\big(y_r, y_{-r}\big) - Cost_r\big(y'_r, y_{-r}\big) = \phi\big(y_r, y_{-r}\big) - \phi\big(y'_r, y_{-r}\big) \tag{5.16}$$

The game is solved via the Best Response and the Replicator Dynamics as previously described. The second stage of the hybrid Load- and Interference-Aware scheme is the same as the one described in paragraph 7 of section 5.3.

5.4.2 Centralized Load- and Interference-Aware Scheme

The centralized approach is compounded of two stages executed sequentially. The first stage, formulated as a modified Bin Packing optimization, aims to reduce the number of active BBUs subject to a minimum level of throughput requirement. A set of candidate partitions of RRHs is therefore derived, and denoted as P_{cand}. The second stage consists in finding the partition that minimizes the total cost function of all RRHs defined as follows:

$$TC(\mathcal{P}) = \sum_{r \in \mathcal{R}} Cost(r) \tag{5.17}$$

First Stage: Bin Packing Optimization Formulation

In the Bin Packing problem, the algorithm chooses all the partitions $\mathcal{P}$ of objects that minimize the total number of used bins, subject to the bin capacity. We define S as the total number of used bins (*i.e*, active BBUs) and x_c as a binary variable that is equal to 1 if the BBU of index c is chosen and to 0 otherwise. The optimization problem is expressed as follows:

$$\textit{Minimize } S = \sum_{c=1}^{C} x_c$$

Subject to:

$$x_c \cdot T(c) \geq n_c \cdot D_{min}, \forall c \in \mathcal{C} \tag{5.18}$$

$$\sum_{c \in \mathcal{C}} x_c \cdot y_{r,c} = 1, \forall r \in \mathcal{R} \tag{5.19}$$

$$x_c \in \{0, 1\}, \forall c \in \mathcal{C} \tag{5.20}$$

$$y_{r,c} \in \{0, 1\}, \forall r \in \mathcal{R}, \forall c \in \mathcal{C} \tag{5.21}$$

Constraints (5.18) insure that a minimum level of throughput per UE (D_{min}) is respected within each cluster. Constraints (5.19) state that each RRH is associated to a single BBU.
The first stage is modeled as a modified Bin Packing optimization, where the constraint of bin capacity is substituted by 5.18. The reason behind this algorithm refinement is to be aware of the load conditions and the interference level, affecting the BBU capacity.

Second Stage: Load- and Interference-Aware Cost Minimization

The second stage of the algorithm chooses the partition $\mathcal{P}^*$ from the derived set $\mathcal{P}_{cand}$, that minimizes the Interference and Load-Aware cost function given as in (5.10). Therefore, $\mathcal{P}^*$ is expressed as:

$$\mathcal{P}^* = \text{argmin}_{\mathcal{P}\in\mathcal{P}_{cand}}\big(TC(\mathcal{P})\big) \tag{5.22}$$

The two stages are executed in a sequential manner until $\mathcal{P}^*$ is chosen.

5.5 Signaling Load Overhead Analysis

In this section, we analyze the signaling load overhead generated by both the centralized and the hybrid schemes.

5.5.1 Centralized Algorithms

In centralized schemes, the BBU-RRH association decisions are performed within the centralized cloud. Thus, the RRHs have to frequently report their load and radio conditions to the BBUs. In our system model, the covered area is meshed into several quadratic zones (cf. chapter 3, section 3.1). Consequently, informations about each zone are reported periodically. We denote by n_{load} and n_{radio} the number of bits needed to precode the load and the radio conditions of one zone respectively. The amount of information concerning the load and the radio conditions of one RRH is expressed as $\big(n_{load} + n_{radio}\big) \cdot Z$, where Z denotes the number of zones associated to one RRH. A RRH must also report its identity. Denoting by n_{idt} the number of bits used to precode one RRH identity, the total amount of bits a RRH should report is equal to $\big(n_{load} + n_{radio}\big) \cdot Z + n_{idt}$. Consequently, the total number of bits generated by all the RRHs is equal to $\big[\big(n_{load} + n_{radio}\big).Z + n_{idt}\big] \cdot R$, where R is the total number of RRHs in the system. Assuming that the clustering decisions are re-considered every period T, the signaling load overhead generated by all the RRHs in the network is expressed as follows:

$$\frac{\big[\big(n_{load} + n_{radio}\big).Z + n_{idt}\big] \cdot R}{T} \tag{5.23}$$

5.5.2 Hybrid Algorithms

Within the hybrid schemes, RRHs just need to send their final decisions to the centralized cloud. In other terms, each RRH sends its vector y_r of pure strategy, defining its chosen cluster. Thus, a matrix $\big(\boldsymbol{M}\big)_{R\times S}$ of dimensions $R \times S$ needs to be generated and sent to the centralized cloud. R and S denote the total number of RRHs and the available strategies respectively. In fact, the matrix $\big(\boldsymbol{M}\big)_{R\times S}$ has to be sent many times to the centralized cloud, since the latter decides to relaunch the game according to the realized throughput within each cluster (cf. paragraph 7). In that case, the generated signaling load overhead depends on the matrix dimensions, particularly, on the number of strategies S_{round} within a given round, and on the number of rounds relaunched by the centralized cloud to run the game among the RRHs. The total amount of signaling load overhead generated by hybrid schemes is expressed as follows:

$$\frac{\sum_{round} R \times S_{round}}{T} \tag{5.24}$$

We note that the number of rounds can be reduced by properly tuning the number of active BBUs according to the load conditions. However, in our simulations we start from a number of strategies equal to $\frac{R}{2}$, as previously mentioned.

5.5.3 Numerical Illustration

In order to compare the signaling load overhead generated by the centralized and the hybrid schemes, we use the parameters summarized in table 5.1.

Parameter	Value
n_{load}	4 bits
n_{radio}	12 bits
n_{idt}	16 bits
T	60 s

TABLE 5.1: Parameter Values

In the following, we display the results of two different networks: The first consists of 7 RRHs, the second is based on a real network topology [23] and consists of 20 RRHs. For the hybrid schemes, we evaluate the results of the Best Response and the Replicator Dynamics.

Figure 5.5 displays the signaling load overhead for the case of 7 RRHs as a function of the number of UEs per cell. In that case, the algorithm initiates the number of strategies to 3 BBUs for all load scenarios (*i.e*, number of UEs per cell). The signaling load overhead generated by the centralized scheme shows fixed values, whereas the hybrid schemes show variations as a function of the number of UEs per cell. In fact, the variations are due to two factors: The number of strategies (S_{round}), and the number of rounds. For example, for 6 and 7 UEs per cell, where the hybrid schemes activate 3 BBUs (cf. Figure 5.7), we notice the lowest signaling load overhead since the matrix $(\boldsymbol{M})_{R\times S}$ is sent two times only: A first time, a matrix of dimensions (7×3) is sent, and a second time a matrix of dimensions (7×2). Recall that the hybrid schemes decrement the number of strategies when the constraint of QoS is respected in each cluster, in order to realize power savings (cf. paragraph 7). When the centralized cloud realizes that the constraint of QoS is not respected, it sets back the previous number of strategies (*i.e*, 3 BBUs) and the previous BBU-RRH associations and stops the execution of the algorithm. As the number of UEs increases, the signaling load overhead increases, because the needed number of strategies S_{round} to compete on within each round is higher.

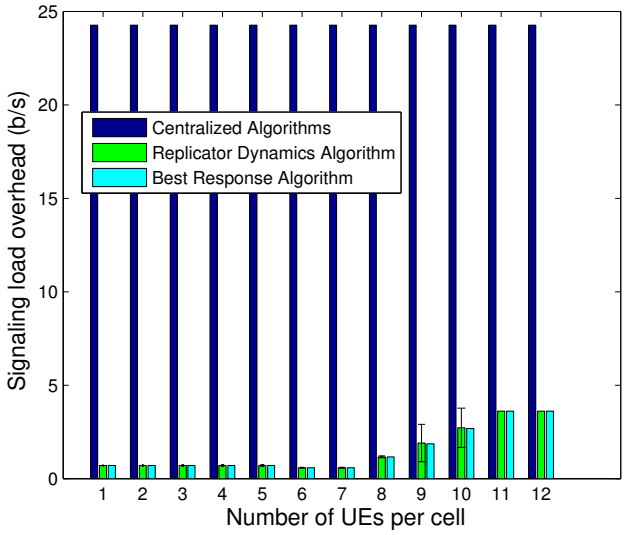


FIGURE 5.5: Signaling Load Overhead: 7 RRHs

Figure 5.6 displays the signaling load overhead for the real network topology. In that case, the algorithm starts from 10 strategies. Within the Replicator Dynamics, we notice that the lowest signaling load overhead is for the case of 4 UEs per cell. In that particular case, the needed number of active BBUs is 11, as shown in Figure 5.32, which is the closest to 10. The same applies for the Best Response, which shows a number of active BBUs slightly higher for 5 UEs per cell. Within the Replicator Dynamics (*i.e*, 4 UEs per cell), the centralized cloud launches the game among the RRHs twice: In a first run, the game consists of 10 strategies, and then in 11 strategies. The same applies for

the case of the Best Response (*i.e*, 5 UEs per cell). The highest amount of signaling load overhead remains for 6 and 7 UEs per cell, as we have the highest number of needed rounds and strategies to compete on. The centralized scheme always shows a fixed value for all load conditions, as all the RRHs send a fixed amount of informations at each clustering decision epoch. The displayed results start from $\frac{R}{2}$ number of strategies for all load conditions (*i.e*, number of UEs per cell). In fact, this case is not realistic when dynamic traffic variations are considered. For example, significant fluctuations within the load conditions might occur within a duration larger than the considered time scale T. By well adapting the period T, slight changes in the appropriate number of active BBUs is needed, which reduces the number of rounds launched by the centralized cloud to set the right number of active BBUs.

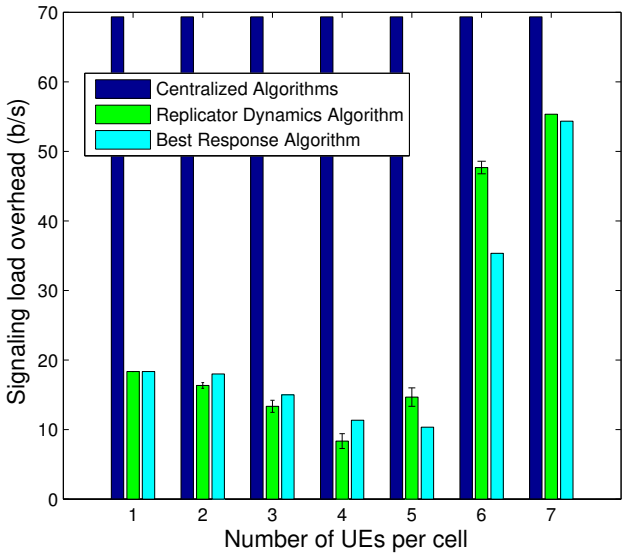


FIGURE 5.6: Signaling Load Overhead: 20 RRHs

5.6 Performance Evaluation

We start first by considering a small network size, consisting in seven cells. We note that the same system model as that described in chapter 3 is employed (cf. section 3.1). Thus, a quadratic meshing of the covered area is considered, with a uniform UEs distribution per cell. In other terms, the number of UEs per zone in each cell is the same. We set α and β, denoting the weights of the load and the interference costs respectively to 0.75 and 0.25 (cf. equation (5.1)). The reason to consider a high weight of load cost is to motivate the RRHs to associate to lightly loaded BBUs in order to balance the charges within each cell. In fact, we have noticed through simulations that high values of β might cause the RRHs to choose few BBUs from the available ones. In other terms, even for a high number of available strategies (*i.e*, BBUs), the RRHs choose to be associated to a low number of BBUs. Recall that the RRHs associated to one BBU will not face the intra-cluster interference among each other, since they act as a DAS system. Consequently, for high values of β, the RRHs are motivated to reduce their interference impact by being clustered together. We also consider a reuse-1 scheme and a full buffer traffic model. A minimum throughput per UE (D_{min}) is set to 2 Mb/s. The power consumption model used in our simulations is the same as in paragraph 3.1.4 of chapter 3. The simulation parameters are shown in table 5.2. We note that the performance metrics used to assess the results are calculated as in section 3.4 of chapter 3.

5.6.1 Load-Balancing and Interference-Aware Performance Evaluation: Uniform UEs Distribution

In this subsection, we evaluate the results of the hybrid Load-Balancing and Interference-Aware proposition that adapts to a uniform distribution of UEs per cell. We compare the results to the centralized Interference-Aware proposition described in chapter 3, and to the conventional architecture. Recall that the centralized proposition relies

TABLE 5.2: Simulation Parameters

Parameter	**Value**
Traffic Model	Full Buffer
Scheduling Scheme	Fair Resource Sharing
Users Distribution	Uniform
Propagation Model	Cost Hata 231
Shadowing Standard Deviation	10 dB
Transmit Power of RRH	40 dBm
Thermal Noise Power	-174 dBm/Hz
BW	10 MHz
Cell Radius	500 m
A	50 W
B	0.6
D_{min}	2 Mb/s
Meshing Step (a)	240 m

on a SPP formulation, where the objective is to choose the subsets of RRHs that form a partition of the initial set $\mathcal{R}$ of RRHs at one hand, and that minimizes a total cost function defined as the total power consumption in the system, on the other hand. We note that the hybrid approach seeks to reduce the power consumption through the second stage, by turning off lightly loaded BBUs, whereas the centralized scheme seeks that goal directly by minimizing the power consumption in the cost function. We refer by H-BR-IACA to the hybrid proposition, solved via the Best Response algorithm, and by H-DR-IACA to the hybrid proposition solved via the Replicator Dynamics. The centralized algorithm is referred to by C-IACA and the conventional scheme by No Clustering.

Figure 5.7 shows the number of active BBUs as a function of the number of UEs per cell. For 1 to 4 UEs per cell, the C-IACA and the hybrid schemes activate only one BBU. When the number of UEs per cell reaches 5, more BBUs are activated. All of the C-IACA, H-BR-IACA and H-DR-IACA provide the same number of active BBUs except for 9 and 10 UEs per cell, where the H-BR-IACA scheme activates an additional BBU in comparison with the centralized scheme for 9 UEs per cell. The H-DR-IACA scheme activates a slightly higher number of BBUs for 10 UEs per cell in comparison with the C-IACA and the H-BR-IACA schemes. The No Clustering scheme activates all available BBUs whatever the number of UEs per cell is. When 11 UEs per cell is reached, all schemes activate 7 BBUs.

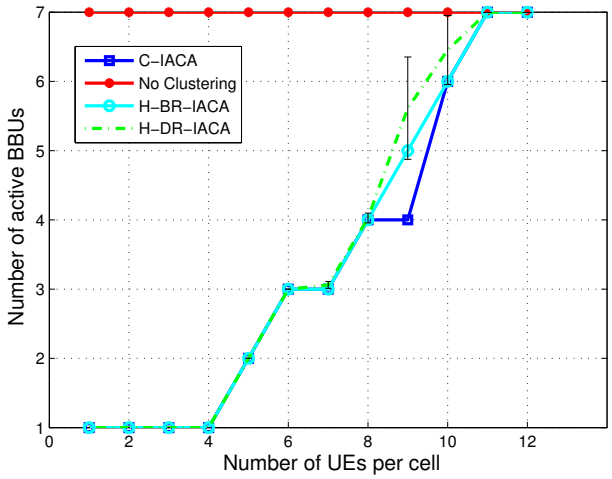


FIGURE 5.7: Number of Active BBUs: C-IACA vs. H-BR-IACA vs. H-DR-IACA vs. No Clustering

In Figure 5.8, the power consumption metric is displayed as a function of the load conditions. We notice that the hybrid schemes show very close values to the centralized scheme except for the cases of 9 and 10 UEs per cell, where the Best Response algorithm activates one additional BBU for the case of 9 UEs per cell, and the Replicator Dynamics activates a slightly higher number of active BBUs for the case of 10 UEs per cell in comparison with the

centralized and the Best Response schemes. The highest power consumption values remain for the No Clustering scheme, where all BBUs are activated.

Figure 5.9 displays the power saving metric as a function of the number of UEs per cell. Between 1 and 4 UEs per cell, we notice that the highest power savings are realized by the C-IACA and the hybrid schemes (slightly higher than 80%). Almost the same power economies are realized for all load conditions, except for 9 and 10 UEs per cell, where additional BBUs are activated by the hybrid schemes in comparison with the C-IACA as shown in Figure 5.7. Further, no power economy is achieved when the number of UEs per cell exceeds 11, as all BBUs are activated.

The energy efficiency is displayed in Figure 5.10 as a function of the load conditions. We notice that the energy efficiency values shown by the H-BR-IACA and H-DR-IACA are higher than those of the C-IACA scheme, except for the case of 9 UEs per cell. In fact, the main objective of the C-IACA is to reduce the power consumption. Again, reducing the power consumption does not necessary lead to enhanced energy efficiency, as stated before. The No Clustering scheme shows the same values of energy efficiency for all load conditions, because the radio conditions are fixed and all the resources are consumed by the traffic which is assumed to be full buffer.

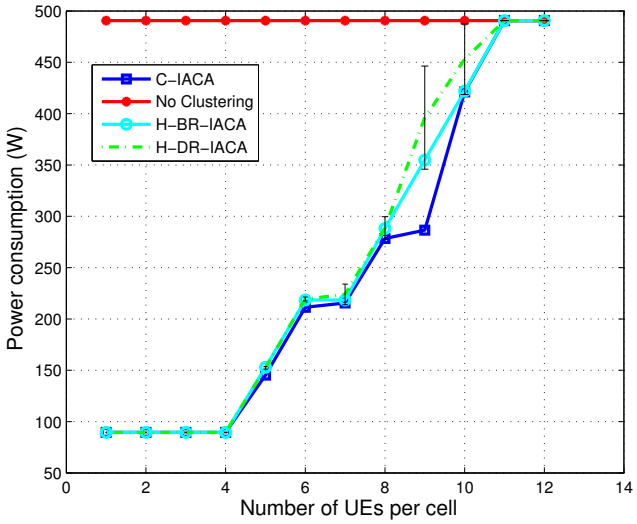


FIGURE 5.8: Power Consumption: C-IACA vs. H-BR-IACA vs. H-DR-IACA vs. No Clustering

Figure 5.11 shows the mean throughput per UE as a function of the number of UEs per cell. The highest values are attained through the No Clustering scheme, owing to the full buffer traffic model. In fact, when 7 BBUs are activated, all the resources are consumed, leading to high throughput. Particularly, when the number of UEs per cell is low. However, we notice that all of the centralized and hybrid schemes realize much more power economy, at the cost of lower mean throughput per UE in comparison with the No Clustering scheme. In addition, the C-IACA and the hybrid schemes guarantee the targeted QoS for UEs, and do not offer throughputs less than 2 Mb/s per UE. For high number of UEs per cell, the mean throughput per UE is reduced for all schemes (the centralized, hybrids and No Clustering), due to the scarcity in the radio resources.

Figure 5.12 displays the spectral efficiency per BBU as a function of the load conditions. The spectral efficiency behavior follows that of the energy efficiency, as both metrics reflect the load and radio conditions. In particular, the hybrid schemes show higher values of spectral efficiency than the centralized scheme, which remains the most efficient in reducing the power consumption.

Figure 5.13 shows the total derived throughput in the system as a function of the load conditions. The H-BR-IACA and the H-DR-IACA schemes show more throughputs than the case of C-IACA, as they consume more power.

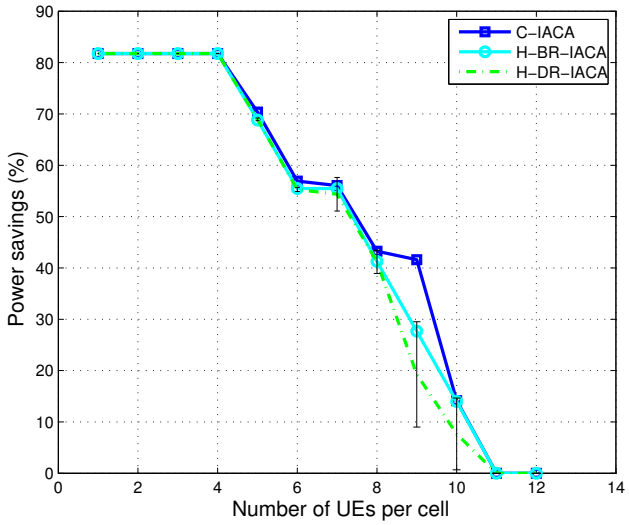


FIGURE 5.9: Power Savings: C-IACA vs. H-BR-IACA vs. H-DR-IACA

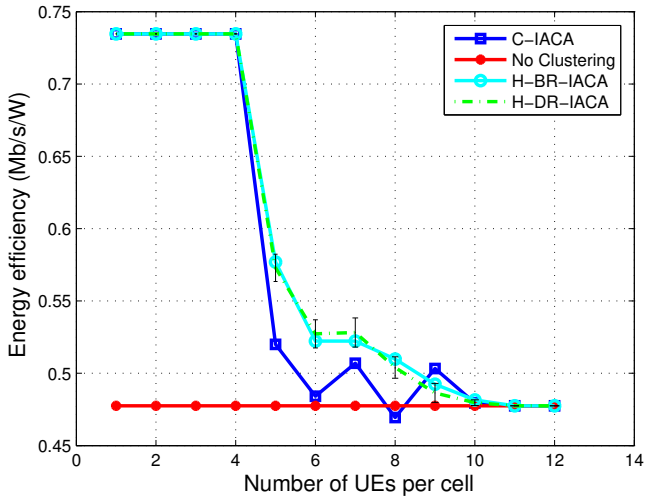


FIGURE 5.10: Energy Efficiency: C-IACA vs. H-BR-IACA vs. H-DR-IACA vs. No Clustering

Figure 5.14 represents the execution time of all algorithms as a function of the number of UEs per cell. We note that the algorithms are executed over a hardware of 64-bit, Intel 7 core, 2.5 Ghz. We notice that the H-BR-IACA scheme consumes less computational time for all load conditions in comparison with other schemes. The H-DR-IACA consumes less computational time than the C-IACA scheme for a number of UEs per cell that varies between 1 and 7. When the number of UEs per cell is greater than 7, the computational time of the H-DR-IACA is greater than all other schemes. In fact, when a higher number of strategies (*i.e*, number of BBUs) is available for the reinforcement learning algorithm (H-DR-IACA), the time convergence would be greater. However, reinforcement learning algorithms make their decisions independently, and require less storage than other schemes, since each player does not need to know any information on other players, except their attributed costs while making a choice.

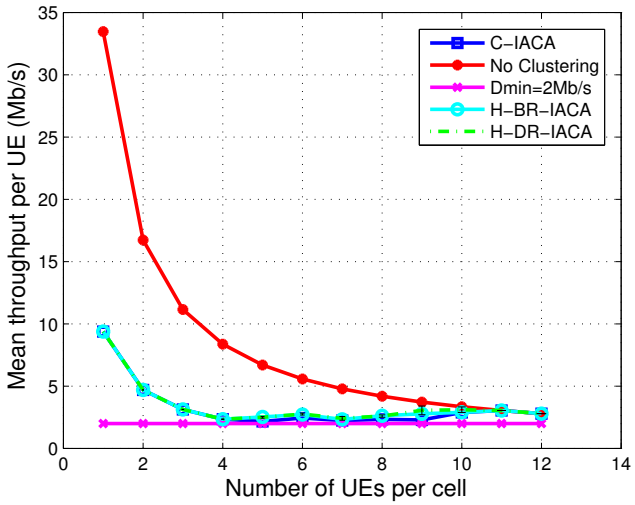


FIGURE 5.11: Mean Throughput per UE: C-IACA vs. H-BR-IACA vs. H-DR-IACA vs. No Clustering

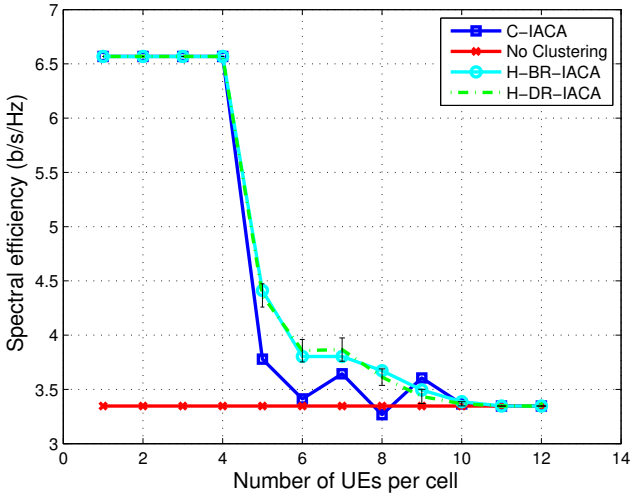


FIGURE 5.12: Spectral Efficiency: C-IACA vs. H-BR-IACA vs. H-DR-IACA vs. No Clustering

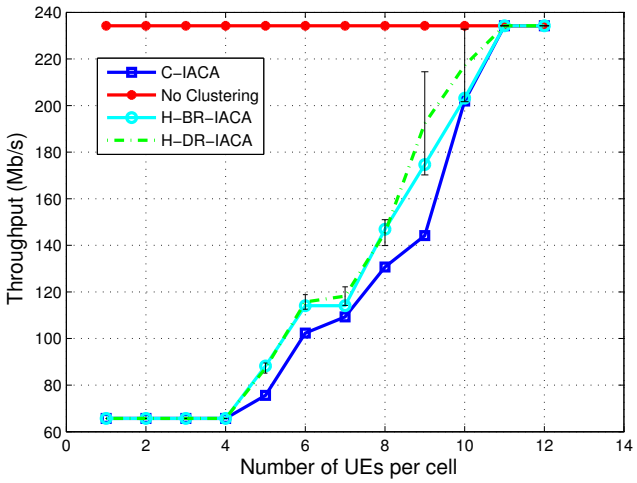


FIGURE 5.13: Total Throughput: C-IACA vs. H-BR-IACA vs. H-DR-IACA vs. No Clustering

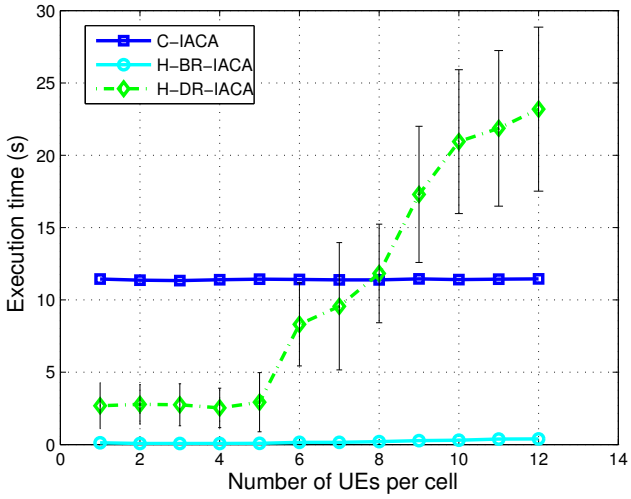


FIGURE 5.14: Execution Time: C-IACA vs. H-BR-IACA vs. H-DR-IACA

5.6.2 Load- and Interference-Aware Performance Evaluation: Heterogeneous UEs Distribution per cell

In this subsection, we provide a comparison between the centralized Load- and Interference-Aware algorithm (C-LIACA) presented in subsection 5.4.2, and the centralized Interference-Aware proposition of chapter 3 (cf. paragraph 5.6.2). The aim of this comparison is to evaluate the importance of the Load-Aware behavior, while performing the BBU-RRH association algorithm. Further, in paragraph 5.6.2, we compare the C-LIACA and the hybrid Load- and Interference-Aware (H-LIACA) that relies on Game Theory. The values of α and β remain equal to 0.75 and 0.25 for the same previously explained reason (cf. section 5.6). The simulations are performed according to different load conditions: low, medium and high. At low load conditions, the number of UEs per cell is between 1 and 4. For medium load conditions, the number of UEs per cell varies between 5 and 8. It is between 9 and 12 for high load conditions. Performance metrics are averaged and shown with 95% confidence interval. We adopt the same simulation parameters as in table 5.2.

C-IACA vs. C-LIACA vs. No Clustering

Figure 5.15 provides the number of active BBUs as a function of load conditions. We notice that both the C-IACA and C-LIACA schemes provide the same number of active BBUs for all load conditions. At low load, only one BBU is activated by the C-IACA and the C-LIACA. For medium load, almost 3 BBUs are activated by the centralized schemes. At high load, an average number of 6.5 active BBUs is obtained in the C-IACA and the C-LIACA schemes. The No Clustering scheme activates all the BBUs for all load conditions. We note that activating the same number of BBUs does not lead to the same realized energy efficiency, as the latter depends on the load factor and interference level (cf. Figure 5.18).

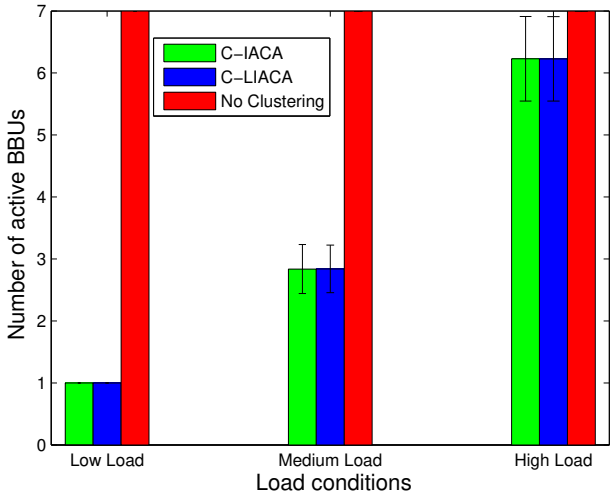


FIGURE 5.15: Number of Active BBUs: C-IACA vs. C-LIACA vs. No Clustering

Figure 5.16 displays the power consumption as a function of load conditions. We notice that for low and high load conditions, the power consumption derived in the C-IACA and the C-LIACA schemes is the same. As for medium load conditions, the C-LIACA shows more derived power consumption in comparison with the C-IACA. Recall that the C-LIACA is a Load- and Interference-Aware scheme. Thus, it improves the load distribution among the BBUs, leading to enhanced throughput. As the power consumption is a function of the total derived throughput, it shows higher value.

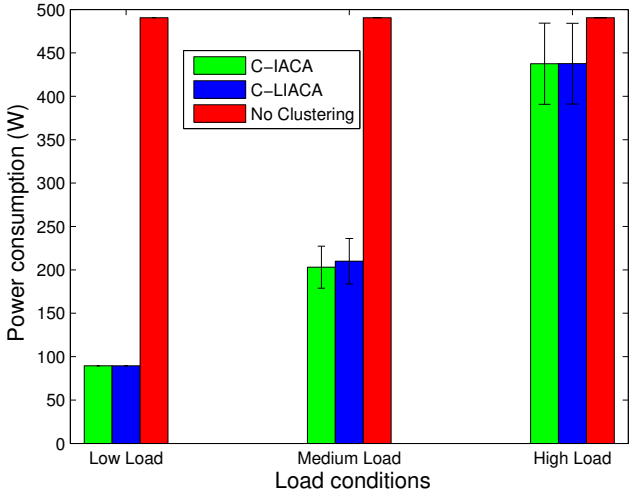


FIGURE 5.16: Power Consumption: C-IACA vs. C-LIACA vs. No Clustering

In Figure 5.17, the power saving metric is displayed as a function of the load conditions. We notice that for low

and high load conditions, both of the C-IACA and the C-LIACA schemes display the same power saving values because they both activate the same number of BBUs (cf. Figure 5.15). For medium load conditions, the power saving realized by the C-LIACA scheme is slightly less than the C-IACA scheme, as more power consumption is derived.

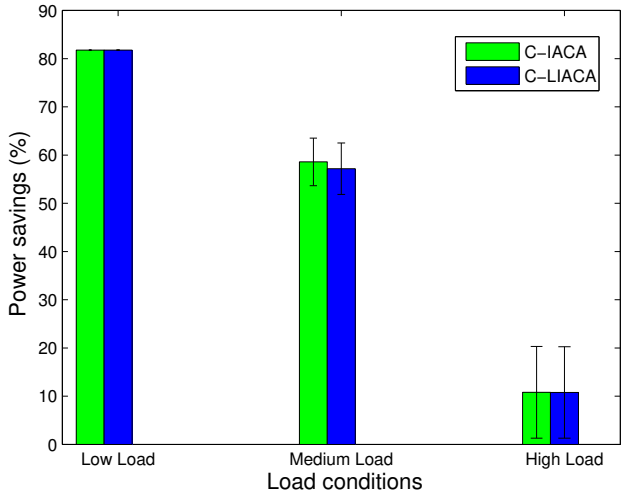


FIGURE 5.17: Power Savings: C-IACA vs. C-LIACA

Figure 5.18 shows the energy efficiency as a function of the different load conditions. We notice that for low and high load conditions, the realized energy efficiency by both the C-IACA and the C-LIACA schemes is the same. In fact, the lowest or highest load conditions are extreme cases that activate the lowest or highest number of BBUs, respectively. Consequently, the realized energy efficiency is the same in both algorithms, because the interference level is the same. For medium load, the C-LIACA scheme realizes higher energy efficiency than the C-IACA scheme. As previously stated, the C-LIACA improves the load distribution among the BBUs, leading to enhanced energy efficiency and throughput. Within a realistic scenario, the medium load condition is mostly encountered. Thus, it would be better to resort to the Load- and Interference aware approach. For low and high load conditions, the C-IACA scheme would be more efficient since it is a one-stage algorithm, focusing on power consumption minimization. We finally underline that the C-IACA and C-LIACA schemes realize more energy efficiency at low and medium load conditions, in comparison with the No Clustering scheme. For high load conditions, the energy efficiency is the same for all schemes.

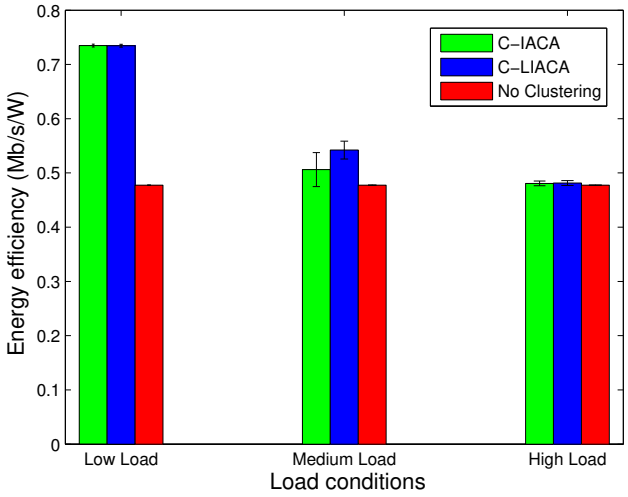


FIGURE 5.18: Energy Efficiency: C-IACA vs. C-LIACA vs. No Clustering

The mean throughput per UE is depicted in Figure 5.19 as a function of the load conditions. We notice that the No Clustering scheme shows the highest mean throughput per UE for all load conditions, owing to the full buffer traffic model. We also notice that all schemes respect a minimum throughput requirement and do not show values lower than 2 Mb/s. Additionally, the mean throughput per UE is the same for the C-IACA and the C-LIACA schemes at low and high load conditions. However, it is higher for the case of the C-LIACA in comparison with the C-IACA at medium load conditions, as the spectral efficiency is higher (cf. Figure 5.21).

Besides, Figure 5.20 shows the CDF function of the mean throughput per UE for the C-IACA and the C-LIACA schemes at medium load conditions. We particularly notice that the probability that a UE gets a throughput less than 2.5 Mb/s for the C-LIACA scheme is 50%, whereas it around 85% for the C-IACA scheme. However, both schemes guarantee a mean throughput per UE greater than 2 Mb/s. Such enhancement is particularly useful to cope with load fluctuations that might occur in the network, causing a degradation in QoS.

Figure 5.21 displays the spectral efficiency per BBU as a function of the different load conditions. Similar to the energy efficiency, the spectral efficiency for the C-LIACA is higher at medium load conditions than that of the C-IACA, as it is aware of the load conditions within each cluster. For low and high load conditions, the spectral efficiency is the same as we are in presence of extreme load conditions.

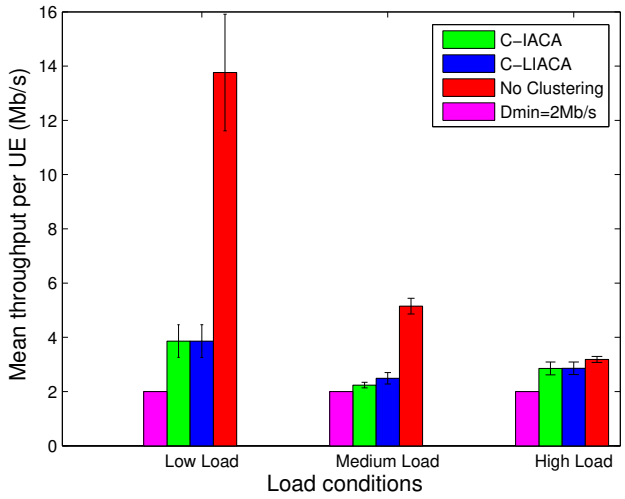


FIGURE 5.19: Mean Throughput per UE: C-IACA vs. C-LIACA vs. No Clustering

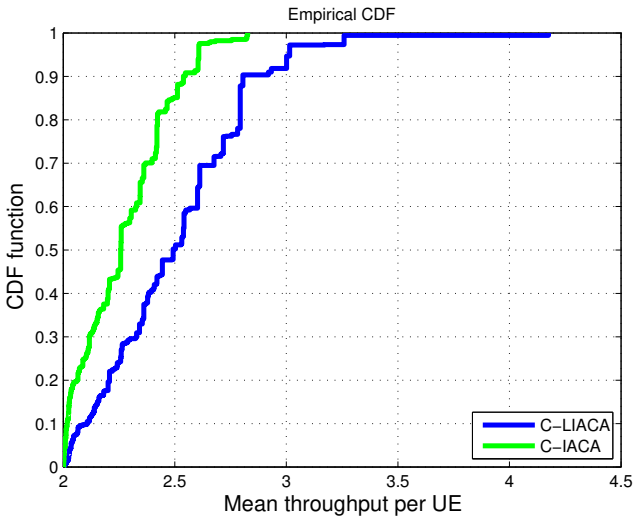


FIGURE 5.20: CDF of mean throughput per UE: C-IACA vs. C-LIACA vs. No Clustering

Figure 5.22 shows the total derived throughput as a function of the different load conditions. The C-LIACA shows more derived throughput at medium load conditions, as it a Load-Aware algorithm. For low and high load conditions, the total throughput is the same for both the C-IACA and the C-LIACA schemes. The No Clustering scheme shows the highest derived throughput for all load conditions. However, it does not realize any power economy.

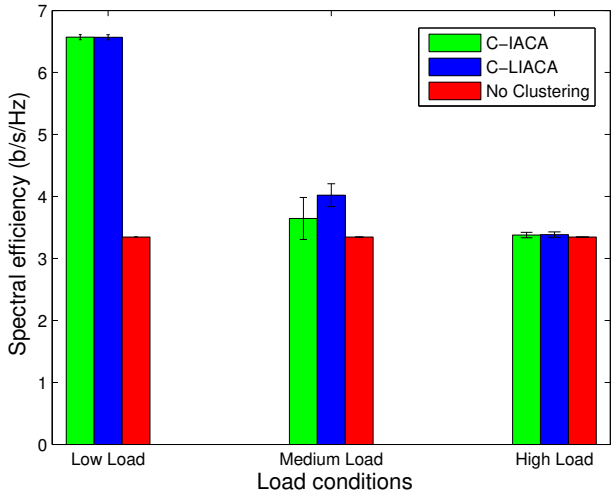


FIGURE 5.21: Spectral Efficiency: C-IACA vs. C-LIACA vs. No Clustering

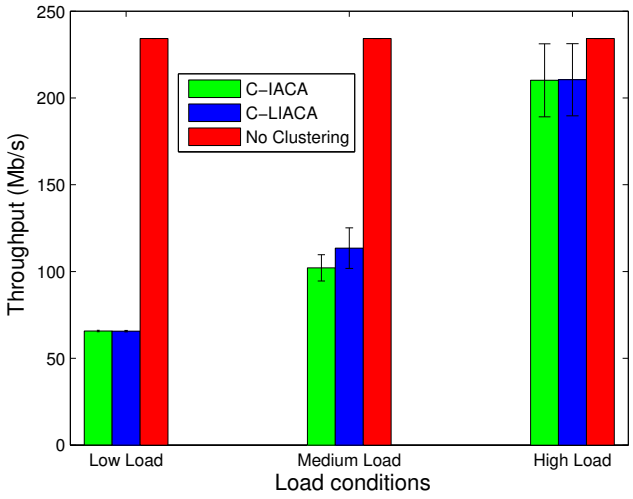


FIGURE 5.22: Total Throughput: C-IACA vs. C-LIACA vs. No Clustering

C-LIACA vs. H-BR-LIACA vs. H-DR-LIACA vs. No Clustering

Figure 5.23 represents the number of active BBUs as a function of the load conditions. At low load conditions, only one BBU is activated for all schemes (C-LIACA, H-BR-LIACA and H-DR-LIACA). When the network load conditions are moderate, the C-LIACA activates a number of BBUs that is slightly less than that of H-BR-LIACA and H-DR-LIACA. At high load conditions, the centralized and hybrid schemes turn on an average number of BBUs greater than 6. Here again, a slight economy is realized by all schemes, as they still refrain from activating all 7 BBUs. The No Clustering scheme activates all available BBUs whatever the load conditions are.

Figure 5.24 shows the power consumption as a function of the load conditions. For low load conditions, all of the centralized and hybrid schemes show the same power consumption, as they all activate one BBU. For medium load,

the H-BR-LIACA and H-DR-LIACA show slightly more power consumption than that of C-LIACA, as they activate slightly more BBUs. For high load conditions, the C-LIACA scheme always shows the lowest power consumption in comparison with the hybrid and the No Clustering schemes.

In Figure 5.25, the power saving metric is shown as a function of the load conditions. The highest power saving values are shown by the centralized scheme since it activates less BBUs. At medium and high load conditions, the hybrid schemes show slightly less power savings than the centralized scheme, as slightly more power is consumed (cf. Figure 5.24).

Figure 5.26 displays the energy efficiency as a function of the load conditions. When the network load conditions are low or high, almost the same energy efficiency is realized in the centralized and hybrid schemes. At medium load conditions, the centralized scheme realizes a slightly higher energy efficiency in comparison with the H-BR-LIACA and H-DR-LIACA schemes, as the number of active BBUs is lower, leading to less interference. The lowest energy efficiency remains for the No Clustering scheme, as more BBUs are activated and the interference level is higher.

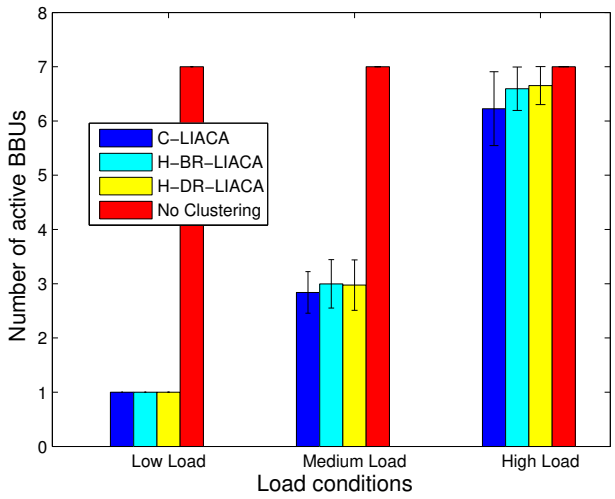


FIGURE 5.23: Number of Active BBUs: C-LIACA vs. H-BR-LIACA vs. H-DR-LIACA vs. No Clustering

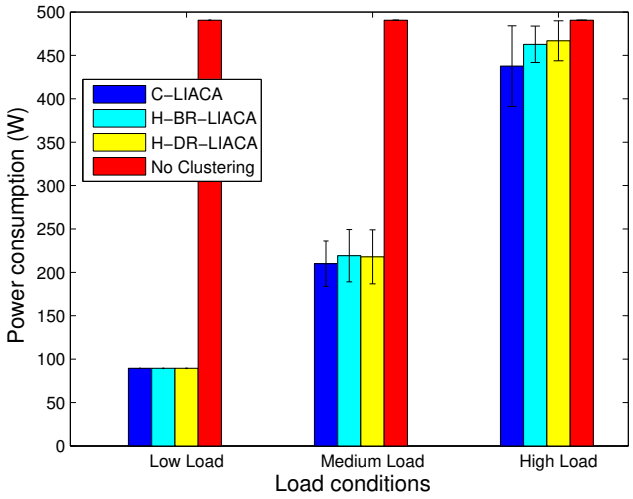


FIGURE 5.24: Power Consumption: C-LIACA vs. H-BR-LIACA vs. H-DR-LIACA vs. No Clustering

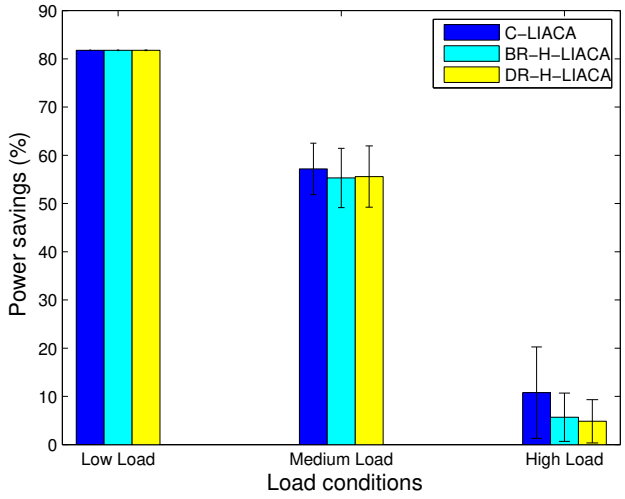


FIGURE 5.25: Power Savings: C-LIACA vs. H-BR-LIACA vs. H-DR-LIACA

Figure 5.27 shows the mean throughput per UE as a function of the load conditions. At low load conditions, the same mean throughput per UE is realized by the centralized and the hybrid schemes. The highest mean throughput per UE is attained through the No Clustering scheme. However, we notice that all of the C-LIACA, H-BR-LIACA and H-DR-LIACA schemes activate less BBUs at low and medium load conditions at the cost of lower mean throughput per UE. At high load conditions, the C-LIACA scheme shows a mean throughput per UE that is slightly less than those of the H-BR-LIACA and H-DR-LIACA schemes, as it activates a slightly lower number of BBUs. For all cases, we do not record values less than 2 Mb/s, which indicates that all schemes guarantee a minimum level of QoS.

Figure 5.28 shows the CDF function of the mean throughput per UE for all of the C-IACA, C-LIACA, H-BR-LIACA and H-DR-LIACA schemes at medium load conditions. We notice that the C-LIACA and the hybrid schemes guarantee higher values of mean throughput per UE than the C-IACA scheme, as they have a Load-Aware behavior, enhancing the load distribution among clusters and increasing the QoS. For values less than 2.6 Mb/s, the hybrid schemes guarantee higher values of mean throughput per UE than the case of C-LIACA. In fact, the QoS is not the sole target of the C-LIACA, that focuses on reducing the number of active BBUs within its first stage (cf. subsection 5.4.2). Thus, it is not necessary to always have the highest QoS (*i.e*, mean throughput per UE).

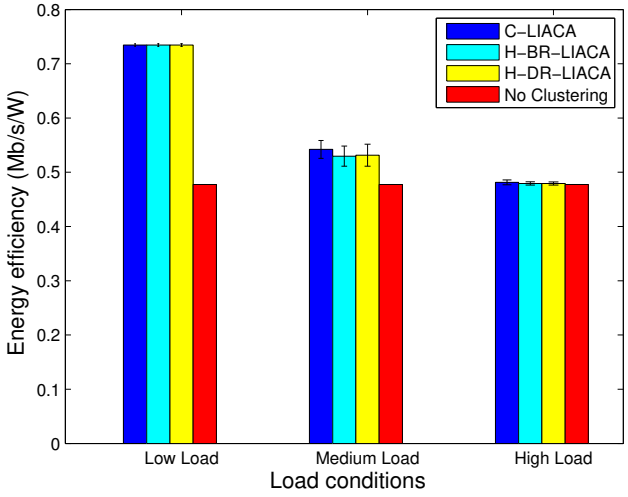


FIGURE 5.26: Energy Efficiency: C-LIACA vs. H-BR-LIACA vs. H-DR-LIACA vs. No Clustering

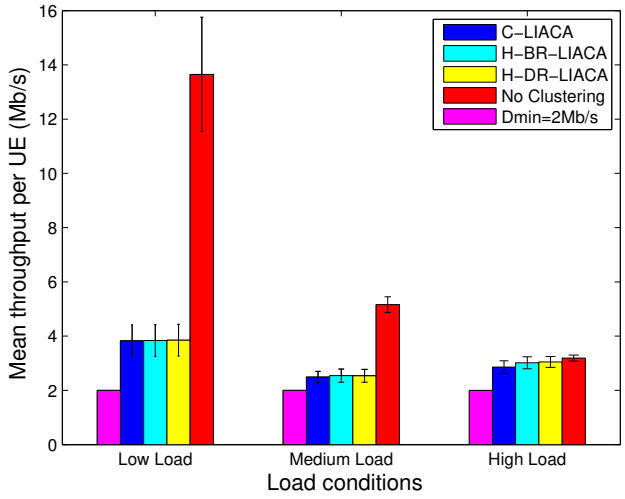


FIGURE 5.27: Mean Throughput per UE: C-LIACA vs. H-BR-LIACA vs. H-DR-LIACA vs. No Clustering

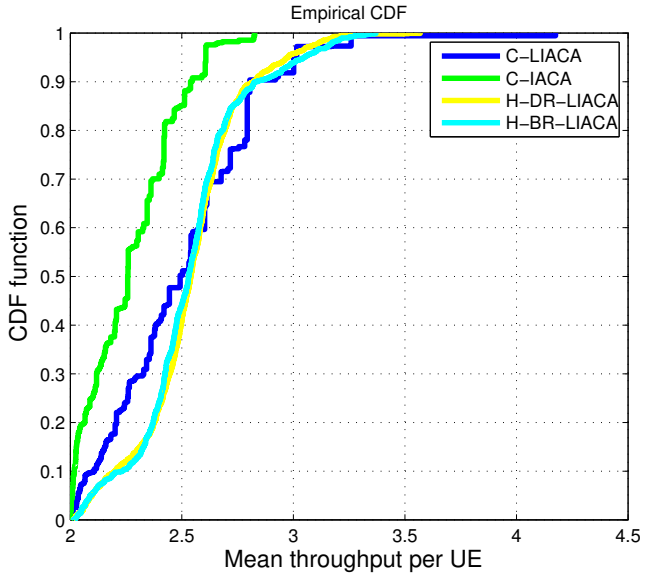


FIGURE 5.28: CDF of mean throughput per UE: C-IACA vs. C-LIACA vs. H-BR-IACA vs. H-DR-IACA

In Figure 5.29, the spectral efficiency per BBU is displayed as a function of load conditions. Similar to the energy efficiency, the spectral efficiency shows little discrepancy between the centralized and the hybrid schemes at low and high load conditions. For medium load conditions, the spectral efficiency realized by the centralized scheme remains the highest, as less BBUs are activated. The No Clustering scheme shows the lowest spectral efficiency, as the interference level is higher than all other schemes, since more BBUs are activated.

Figure 5.30 shows the total derived throughput as a function of load conditions. At low load conditions, the centralized and the hybrid schemes show the same values of throughput because only one BBU is activated. For high and medium load conditions, the C-LIACA scheme shows the lowest value since it realizes the highest power economy. The No Clustering scheme shows the highest values for all load conditions, as all the BBUs are activated and all the resources are consumed.

Figure 5.31 displays the execution time of all of the C-LIACA, H-BR-LIACA and H-DR-LIACA schemes, as a function of load conditions. The algorithms are executed over a hardware of 64-bit, Intel 7 core, 2.5 Ghz. We clearly

see how the H-BR-LIACA scheme consumes much less computational time for all load conditions in comparison with other schemes. The H-DR-LIACA scheme shows less computational time at low and medium load conditions in comparison with the C-LIACA scheme. Yet, when more BBUs are activated, or in other terms, a higher number of strategies is attributed to the reinforcement learning algorithm (H-DR-LIACA), the convergence time is greater than all other schemes, as in the case of high load conditions. However, reinforcement learning algorithms have the advantage of being fully distributed and require less assistance than other schemes.

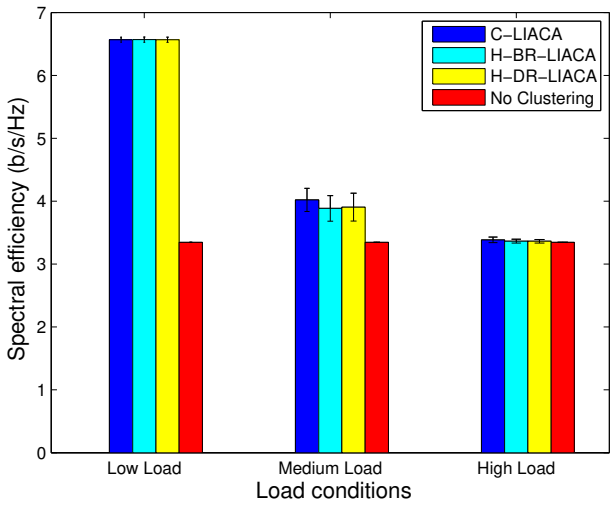


FIGURE 5.29: Spectral Efficiency: C-LIACA vs. H-BR-LIACA vs. H-DR-LIACA vs. No Clustering

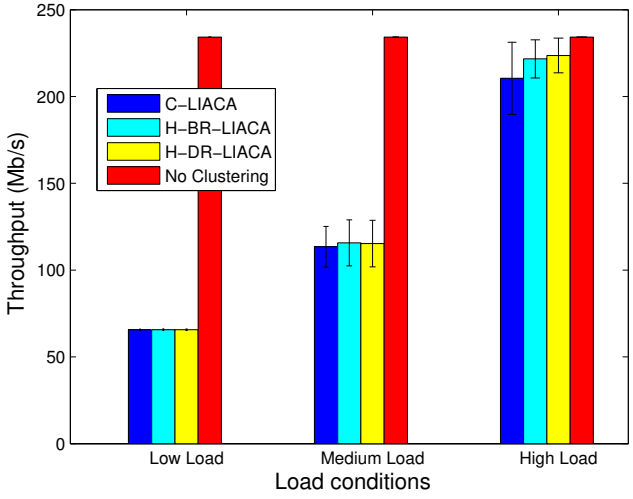


FIGURE 5.30: Total Throughput: C-LIACA vs. H-BR-LIACA vs. H-DR-LIACA vs. No Clustering

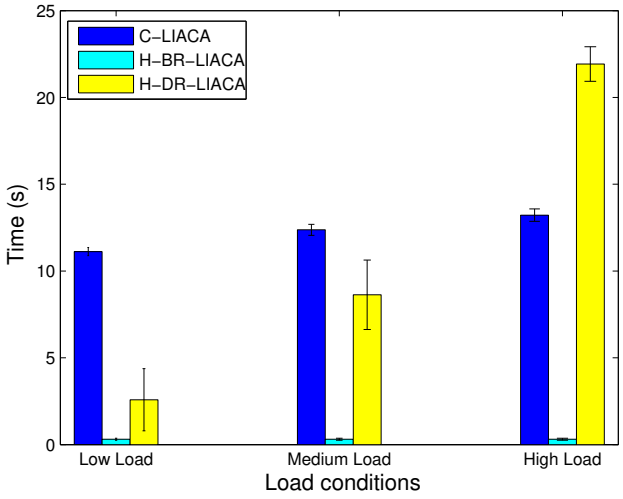


FIGURE 5.31: Execution Time: C-LIACA vs. H-BR-LIACA vs. H-DR-LIACA

The Hybrid Scheme in the Case of Real Case Topology

In this subsection, we evaluate the performance of the hybrid proposition in the case of the real network topology considered in chapter 4. In that case, we set the values of α and β to 0.9 and 0.1 respectively. In fact, for high values of β, the RRHs tend to be associated together in order to lower inter-cell interferences. In the case of the real network topology, the RRHs are highly motivated to be associated together, due to their closeness to each other. In that case, stressing the impact of the load would motivate them to organize themselves into different clusters (*i.e*, BBUs), since the interference impact is already high. The results are compared to the previously proposed heuristic in chapter 3 (cf. section 3.3). Recall that the heuristic consists in choosing randomly a RRH, and associates it with the one that minimizes the total power consumption within the cluster at one hand, and respects a minimum throughput requirement, on the other hand. The heuristic repeats the process until a minimum throughput requirement is violated. The simulation parameters used to assess the results are the same as in table 4.1 (cf. chapter 4 section 4.5).

Figure 5.32 displays the number of active BBUs as a function of the number of UEs per cell. For 1 UE per cell, all schemes show 1 active BBU. When the number of UEs per cell is 2, the H-BR-IACA scheme shows a number of BBUs slightly less than that of the centralized heuristic and the H-DR-IACA scheme. For higher number of UEs, the H-BR-IACA scheme shows higher values than that of the centralized heuristic. The H-DR-IACA activates more BBUs for all load conditions, as it is based on a reinforcement learning algorithm. We further notice that the centralized heuristic, the H-BR-IACA and the H-DR IACA activate a number of BBUs that is less than 20 when 7 UEs per cell is reached. This is due to the high level of interference when many BBUs are activated. In that case, all algorithms try to reduce the impact of interference by lowering the number of BBUs.

Figure 5.33 displays the power consumption metric as a function of the number of UEs per cell. For 1 and 2 UEs per cell, the power consumption derived by the centralized heuristic and the H-BR-IACA scheme is similar (slightly less for the case of the H-BR-IACA scheme). For higher number of UEs, the H-BR-IACA derives more power consumption than the centralized scheme, as it activates more BBUs. The H-DR-IACA activates more BBUs than both the H-BR-IACA and the centralized heuristic. However, all schemes show less power consumption than the conventional architecture that activates all BBUs.

The power saving metric is depicted in Figure 5.34 as a function of load conditions. For the cases of 1 and 2 UEs per cell, we clearly notice that the power saving metric is similar in the H-BR-IACA and the centralized heuristic schemes. The centralized heuristic realizes more power savings than the H-BR-IACA, as the number of UEs per cell increases. The lowest power savings remain for the H-DR-IACA since it activates more BBUs.

In Figure 5.35, we display the energy efficiency as a function of the number of UEs per cell. We notice that the H-BR-IACA shows the highest energy efficiency despite the fact that it realizes less power economy than the centralized heuristic for high load conditions. Again, realizing more power economy does not necessarily lead to enhanced energy efficiency. The H-DR-IACA shows less energy efficiency than the centralized heuristic and the H-BR-IACA, as it increases the number of active BBUs and thus, magnifies interference.

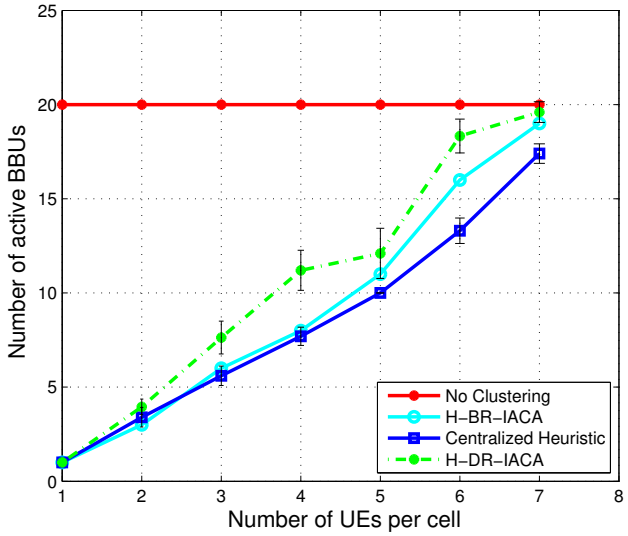


FIGURE 5.32: Number of Active BBUs: C-IACA vs. H-BR-IACA vs. H-DR-IACA vs. No Clustering

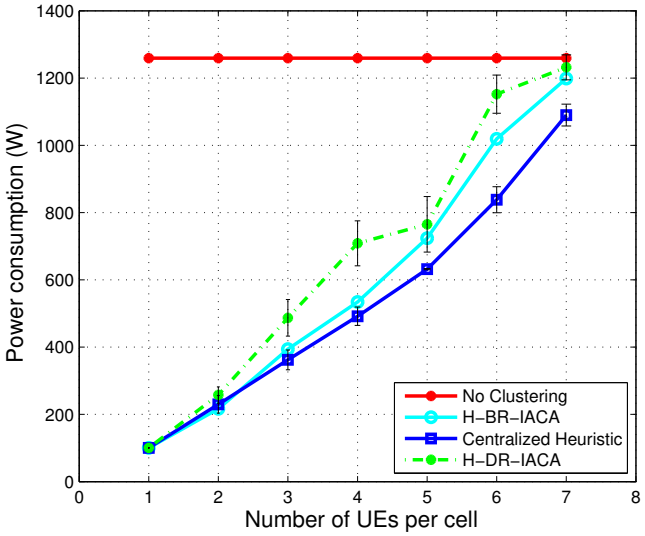


FIGURE 5.33: Power Consumption: C-IACA vs. H-BR-IACA vs. H-DR-IACA vs. No Clustering

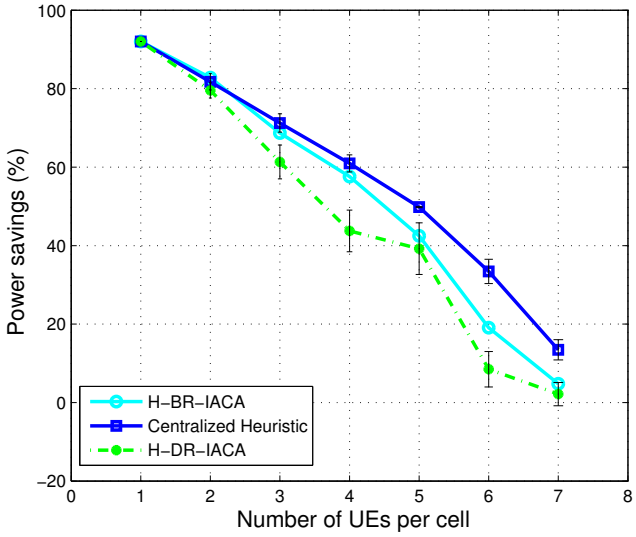


FIGURE 5.34: Power Savings: C-IACA vs. H-BR-IACA vs. H-DR-IACA

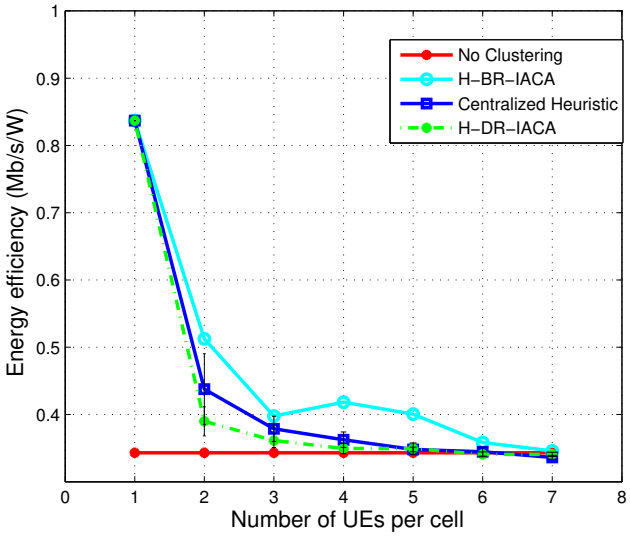


FIGURE 5.35: Energy Efficiency: C-IACA vs. H-BR-IACA vs. H-DR-IACA vs. No Clustering

Figure 5.36 shows the mean throughput per UE as a function of the number of UEs per cell. The No Clustering scheme displays the highest mean throughput per UE as it activates all BBUs while a full buffer traffic scheme is considered. The centralized heuristic reduces the mean throughput per UE to 2 Mb/s to realize more power savings for a number of UEs per cell greater than 1. The H-BR-IACA and the H-DR-IACA realize more mean throughput per UE than the case of the centralized heuristic, as they activate more BBUs. All of the centralized heuristic, H-BR-IACA and H-DR-IACA schemes do not show values less than 2 Mb/s, which underlines the respect of a minimal level of QoS.

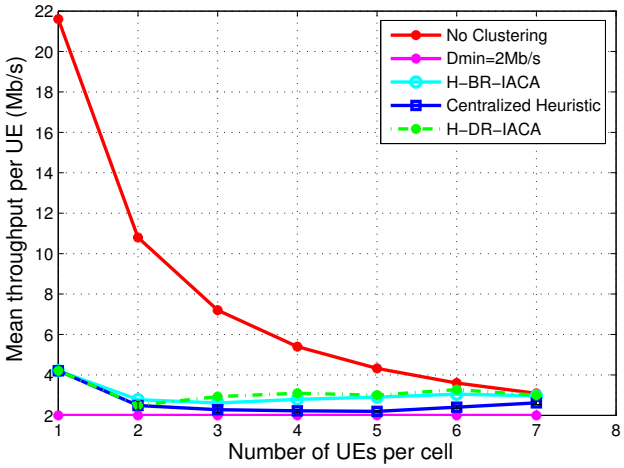


FIGURE 5.36: Mean Throughput per UE: C-IACA vs. H-BR-IACA vs. H-DR-IACA vs. No Clustering

Figure 5.37 shows the spectral efficiency per BBU as a function of the number of UEs per cell. Similar to the energy efficiency, the highest values of spectral efficiency are for the H-BR-IACA and the lowest for the No Clustering scheme, as the interference level is high.

The total derived throughput is depicted in Figure 5.38 as a function of load conditions. We notice that the highest values are realized for the No Clustering scheme, as all the BBUs are activated for all load conditions. We also notice that the centralized heuristic realizes the lowest throughputs, as it activates the lowest number of BBUs. Further, we notice that for the case of 5 UEs per cell, the H-BR-IACA scheme realizes more throughput than the H-DR-IACA scheme. For that particular case, the number of active BBUs in both cases is similar. However, the H-BR-IACA shows higher spectral efficiency, enhancing the throughput metric.

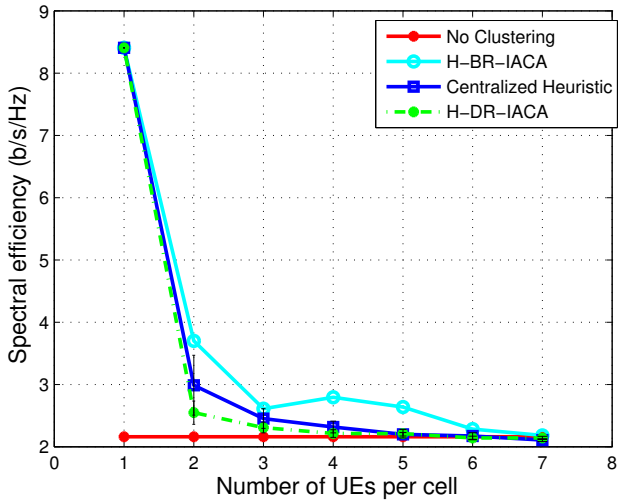


FIGURE 5.37: Spectral Efficiency: C-IACA vs. H-BR-IACA vs. H-DR-IACA vs. No Clustering

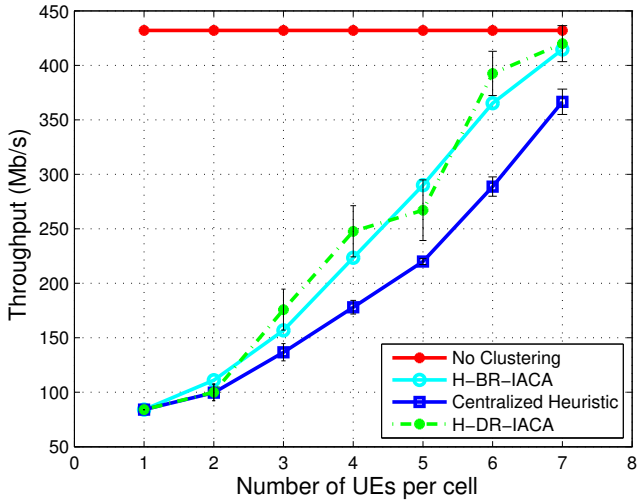


FIGURE 5.38: Total Throughput: C-IACA vs. H-BR-IACA vs. H-DR-IACA vs. No Clustering

Figure 5.39 displays the execution time of the centralized heuristic, the H-BR-IACA and the H-DR-IACA as a function of the load conditions. The results are executed over a hardware of 64-bit, Intel 7 core, 2.5 Ghz. We notice that the H-DR-IACA significantly increases the execution time in comparison with other schemes. In particular, for 5 UEs per cell and above. As the Replicator Dynamics is a reinforcement learning technique, it needs time for convergence, specifically when the number of strategies increases, as the case of 5 UEs per cell (cf. Figure 5.32). In fact, for high load conditions, the Replicator Dynamics will not be a practical solution, as the clustering decisions must be considered within a time scale of minutes. Thus, it would be better to resort to either semi-distributed schemes, such as the Best Response, or to centralized approaches in case of high bandwidth availability in the optical fronthaul connections.

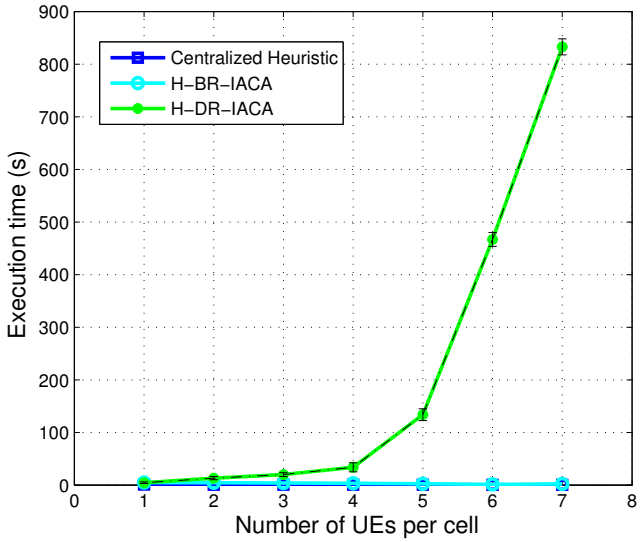


FIGURE 5.39: Execution time: C-IACA vs. H-BR-IACA vs. H-DR-IACA

5.7 Conclusion

In this chapter, we tackled the BBU-RRH association problem from a distributed perspective. The purpose is to reduce the amount of signaling load overhead between BBUs and RRHs. We have first proposed a solution for a homogeneous distribution of UEs in the network, then we enhanced our model to fit with a heterogeneous distribution of UEs, which is more realistic. For both solutions, we resorted to Game Theory, where the BBU-RRH associations were portrayed as a potential game among the RRHs. The NE of the game were reached via two algorithms: The Best Response dynamics and the Replicator Dynamics. We have further proposed a centralized algorithm relying on a Load- and Interference-Aware behavior, to compare the results. The main conclusions are summarized in the following points:

— The hybrid propositions show close performance to the optimal solutions with reduced amount of signaling load overhead.
— The signaling load overhead of the hybrid propositions is a function of the number of the initial set of strategies, and the number of strategies to compete on.
— The Load-Aware behavior in the BBU-RRH association decisions enhances the QoS (*i.e*, the mean throughput per UE), in particular at medium load conditions.
— The Replicator Dynamics is not practical for the case of high load conditions in a real case topology, as the number of strategies increases.

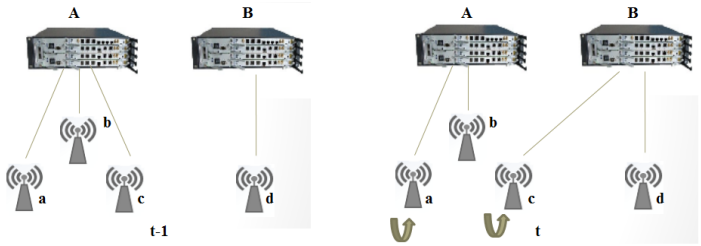
Chapitre 6

BBU-RRH Association Optimization with Re-association Consideration

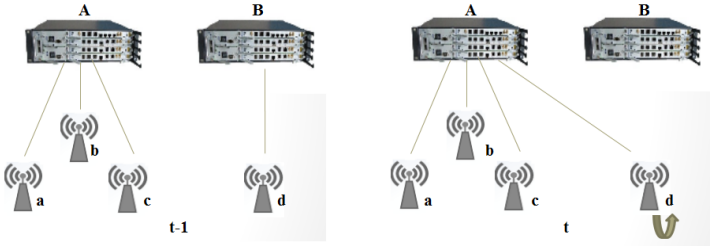
In this chapter, we investigate the BBU-RRH association problem under dynamic load variations. In particular, we aim to solve the frequent re-associations between BBUs and RRHs that occur when network load conditions vary. Precisely, under such variations some UEs may arrive, while some others may quit the system, causing the BBUs to be highly loaded or ligtly loaded. However, UEs connected to highly loaded BBUs might face a degradation in the level of QoS, unless the BBU-RRH association schemes change, so that some RRHs move to lowly loaded BBUs. On the other hand, UEs connected to lightly loaded BBUs might move to different ones in order to turn OFF lowly loaded BBUs. Frequent re-associations between BBUs and RRHs cause a serious degradation in the Quality of Experience (QoE) since connected UEs experience costly HOs to different BBUs. In this chapter, we tackle such a problem by proposing several algorithms. We first introduce a tunable bi-objective optimization problem formulated as a SPP. The purpose is to jointly minimize the power consumption and the re-association rate of UEs experiencing a BBU change. Through conducted simulations, we show a compromise between both metrics. We further, propose a heuristic approach that provides a good tradeoff between power saving and re-association rate reduction, and whose performance outperforms the Bin Packing case. Lastly, we propose a hybrid approach for the problem, aiming to minimize the signaling load overhead in the network and to find a good compromise between power saving and re-association rate reduction. The results of the hybrid proposition show close performance to the centralized scheme with less signaling load overhead. Further, its performance in terms of power savings and re-association rate reduction are enhanced in comparison with the Bin Packing approach.

6.1 Introduction

The HO mechanism is a essential in cellular networks. In particular, it defines the procedure through which a UE changes its serving cell (*i.e*, BBU). In the conventional architecture, HOs occur when UEs move between the covered regions of different RRHs. In fact, HO procedure is defined by the RRC protocol running on BBUs and UEs side. RRHs remain transparent from that procedure and simply act as passive antenna elements, delivering UEs' data. Unlike in conventional architecture, in C-RAN, BBU-RRH association schemes change dynamically. According to load conditions, RRHs might re-associate to different BBUs, causing the HO of UEs connected to re-associated RRHs. When HOs occur, several disadvantages are incurred. For example, 10% of video sessions are abandoned [80]. Web sessions are also abandoned [81]. In fact, IP-level measurements demonstrate that disconnections can reach tens of seconds in HO cases [82]. Based on those results, it is paramount to reduce the HOs within the network. In particular, reducing the re-association rate of UEs introduced by the flexible design of BBU-RRH association schemes is crucial to attain the objective.
Optimizing the BBU-RRH associations is paramount. Reducing the number of active BBUs may lead to resource shortage under any fluctuation in the load conditions, unless BBU-RRH association schemes change, so that RRHs move to less loaded BBUs. However, this causes HOs for serviced UEs. On the other hand, activating many BBUs always guarantee the targeted QoS since enough resources would be available. Therefore, RRHs associations are not changed and UEs do not endure frequent HOs. However, this lowers the level of energy efficiency. Finding a good tradeoff between power savings and re-association rate reduction will be targeted throughout the chapter.
The remainder of the chapter is organized as follows: Section 6.2 describes the system model. Section 6.3 pro-



(a) Scenario A



(b) Scenario B

FIGURE 6.1: BBU-RRH Re-association

vides a bi-objective formulation of the problem, targeting the joint power consumption and the re-association rate reduction. We also highlight the tradeoff between both metrics. Section 6.4 describes the proposed heuristic, aiming to find a good compromise between the two metrics. Section 6.5 formulates the problem in a hybrid fashion for the purpose of reducing the amount of signaling load overhead in the network. Section 6.6 provides the simulation results for the heuristic and the hybrid proposition. Section 6.7 concludes the chapter.

6.2 System Model

We use the reference model of section 3.1 presented in chapter 3. Further, in this chapter, dynamic UEs arrivals and departures are considered, and described within the following subsection.

Network Dynamics

We consider that UEs arrive and leave the network according to a random process with an average rate of λ and μ per time respectively. The number of active UEs is examined at predefined decision epochs, where clustering decisions are reconsidered.

Figures 6.1(a) and 6.1(b) illustrate two representative scenarios, where BBU-RRH re-associations might occur under dynamic load variations. In scenario A (cf. Figure 6.1(a)), UEs connect to any RRH associated to BBU 'A', causing the latter to be overloaded. At the decision epoch t, RRH clustering decisions are re-considered. Thus, a RRH connected to BBU 'A' (for example RRH 'c') might re-associate to BBU 'B' in order to lower the load of BBU 'A' and to guarantee a minimum level of QoS for the serviced UEs. However, connected UEs to RRH 'c' will face HOs to BBU 'B', causing a degradation in the QoE. On the other hand, in scenario B (cf. Figure 6.1(b)), UEs depart from the system. In particular, they disconnect from RRH 'd', causing BBU 'B' to be lightly loaded. At t, a clustering algorithm decides to turn OFF BBU 'B' in order to realize power economy. In that case, the remaining UEs connected to RRH 'd' face HOs to BBU 'A'.

As a matter of fact, the impact on a given RRH is directly related to the number of its serviced UEs. Thus, the cost of a RRH re-association within a cluster c is expressed as follows:

$$R(c) = \sum_{r=1}^{R} n_r \cdot |y_{r,c}^{t} - y_{r,c}^{t-1}|, \tag{6.1}$$

where n_r is the number of serviced UEs by RRH r at epoch t, $y_{r,c}^{t}$ and $y_{r,c}^{t-1}$ are binary variables reflecting the association between RRH r and cluster c at decision epochs t and $t-1$ respectively. When RRH r stays associated

to the same cluster at two consecutive epochs, the cost is zero. It is proportional to the number of serviced UEs by RRH r, when the latter moves to a new cluster at a new epoch t.

6.3 Joint Power Consumption and Re-association Rate Reduction

In this section, we introduce a tunable bi-objective optimization problem, aiming to minimize the joint power consumption and the re-association rate of UEs.
We start by introducing a cost function per cluster composed of two elements:

— The first element is the power consumption of cluster c (*i.e*, BBU c) within the current decision epoch t, defined as in section 3.1 of chapter 3.
— The second element reflects the number of UEs performing re-associations to cluster c between two consecutive decision epochs t and $t-1$. It is expressed as in (6.1)

The bi-objective cost function per cluster is expressed as follows:

$$Cost(c) = \alpha \cdot \alpha' \cdot P(c) + \frac{1}{2} \cdot \beta \cdot \beta' \cdot R(c), \tag{6.2}$$

where α is the weight attributed to the power consumption, β is the weight attributed to the re-association cost, α' and β' are the normalization factors. The term $\frac{1}{2}$ comes from the summation over all the clusters in the system, as we express within the following. We note that $\alpha + \beta = 1$.

The optimization problem aims to find the partition $\mathcal{P}^*$ of the set $\mathcal{R}$ of RRHs that minimizes the total cost function of the system, defined as the sum of each cluster cost (cf. equation (6.2)). Let x_c be the binary variable associated to the subset (*i.e*, cluster) c. It is equal to 1 if the subset c is chosen and 0 otherwise. The optimization problem is expressed as follows:

$$\textit{Minimize } S = \sum_{c \in \mathcal{C}} x_c \cdot Cost(c)$$

Subject to:

$$\sum_{c \in \mathcal{C}} x_c \cdot y_{r,c} = 1, \forall r \in \mathcal{R} \tag{6.3}$$

$$x_c \cdot T(c) \geq n_c \cdot D_{min}, \forall c \in \mathcal{P} \tag{6.4}$$

$$x_c \in \{0,1\}, \forall c \in \mathcal{C} \tag{6.5}$$

$$y_{r,c} \in \{0,1\}, \forall r \in \mathcal{R}, \forall c \in \mathcal{C} \tag{6.6}$$

Constraints (6.3) express that each RRH is associated to a single BBU c. Constraints (6.4) express the fact that a minimum throughput per UE (D_{min}) must be guaranteed in each cluster.
The optimization problem is a SPP, where an additional constraint is added to guarantee a minimum level of QoS per UE.

Tradeoff between Power Saving and Re-association Rate

In this subsection, we evaluate the behavior of our bi-objective optimization problem, and we show the tradeoff between the power saving and the re-association rate of UEs. The algorithm behavior is evaluated under two different load conditions: First, we consider moderate to high load conditions, where the number of UEs per RRH varies between 8 to 15. Then, we consider a low to moderate load conditions, where the number of UEs per RRH varies between 1 to 8 UEs per cell.
To examine the algorithm behavior, we use the following metrics: The number of active BBUs, the total power consumption, the realized power savings, the re-association rate of UEs, the mean throughput per UE and the total derived throughput in the system. Beside the re-association rate metric, all other metrics are calculated as in section 3.4 in chapter 3. The total re-association rate of UEs in the system is calculated as follows:

$$\left(\frac{\sum_{c \in \mathcal{P}} R(c)}{\sum_{c \in \mathcal{P}} n_c}\right) \times 100, \tag{6.7}$$

where $\frac{1}{2}\sum_{c\in\mathcal{P}} R(c)$ is the total number of re-associated UEs in the system and $\sum_{c\in\mathcal{P}} n_c$ is the total number of UEs in the network.
We consider a small network size that consists of seven cells and a uniform UEs distribution per cell (cf. section 3.1, chapter 3). We set λ and μ (the rates of UEs arrival and departure respectively) to 2 and 3 UEs/tu respectively. We consider a reuse-1 scheme and a full buffer traffic model. D_{min} is set to 2 Mb/s. The metrics are averaged and shown with 95% confidence interval. The simulation parameters are shown in table 6.1.

TABLE 6.1: Simulation parameters

Parameter	**Value**
Traffic Model	Full Buffer
Scheduling Scheme	Fair Resource Sharing
UEs Distribution	Uniform
λ	2 UEs/tu
μ	3 UEs/tu
Propagation Model	Cost Hata 231
Shadowing Standard Deviation	10 dB
Transmit Power of RRH	40 dBm
Thermal Noise Power	-174 dBm/Hz
BW	10 MHz
Cell Radius	500 m
A	50 W
B	0.6
D_{min}	2 Mb/s
Meshing Step (a)	240 m

Moderate to High Load Conditions

In this scenario, we consider moderate to high load conditions, where the number of UEs per RRH varies between 8 to 15 UEs.
In Figure 6.2, we display the number of active BBUs for different sets of α. In particular, we consider the following values of α: 0, 0.25, 0.5, 0.75 and 1. When α is 0, we record the highest number of active BBUs. As α increases, the number of active BBUs decreases. In fact, α reflects the importance attributed to the power consumption reduction. Thus, for low values of α, the system main objective would be to reduce the rate of re-associations that might increase when network load conditions vary. As a consequence, the algorithm activates a high number of BBUs, in order to guarantee a minimum level of QoS per UE without any need to re-activate additional BBUs and perform re-associations for the serviced UEs in case of load fluctuations. When α increases, the number of active BBUs decreases because more importance is attributed for power savings. On the other hand, this may cause re-activation for BBUs to guarantee a minimum level of QoS per UE, increasing the rate of re-associated UEs as displayed in Figure 6.5.

Figure 6.3 displays the power consumption for different sets of α. Similar to the number of active BBUs, the highest values of power consumption are obtained for the lowest values of α, since the system main objective is to decrease the re-association rate of UEs. As α increases, the power consumption decreases since the system main objective is to realize power savings.

Figure 6.4, shows the power saving metric for different sets of α. The lowest values of power savings are obtained for the lowest values of α (0 and 0.25), since the system main objective is to reduce the rate of re-associations. When α increases, more power savings are realized. However, this will increase the rate of re-associations (cf. Figure 6.5).

Figure 6.5 displays the re-association rate of UEs for different sets of α. When α has the lowest value, the re-association rate is at its lowest. For that particular case, the algorithm main objective is to reduce the re-association rate of UEs, in order to lower the HOs in the system. As α increases, β decreases and less importance is attributed for re-association rate reduction. As a consequence, the system main objective would be to realize more power savings. However, as previously explained, under any load fluctuation, the system activates additional BBUs to guarantee a minimum rate of throughput per UE, which increases the HOs in the system.

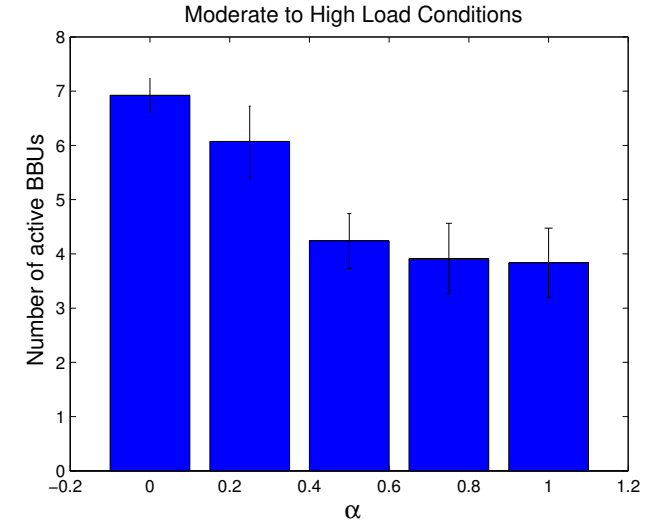


FIGURE 6.2: Number of Active BBUs for different sets α

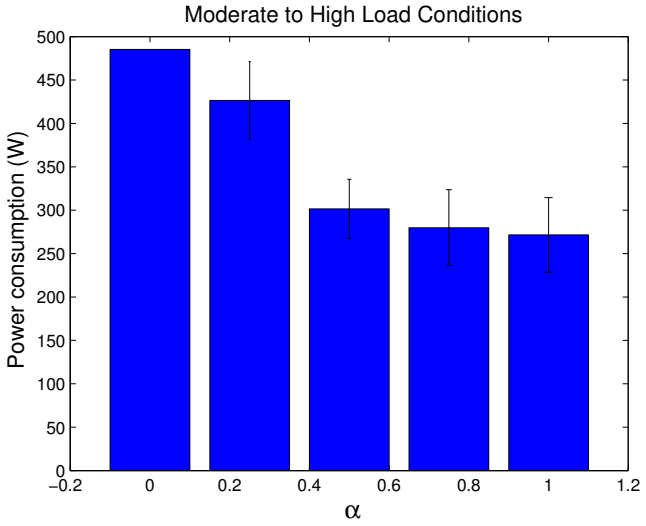


FIGURE 6.3: Power Consumption for different sets α

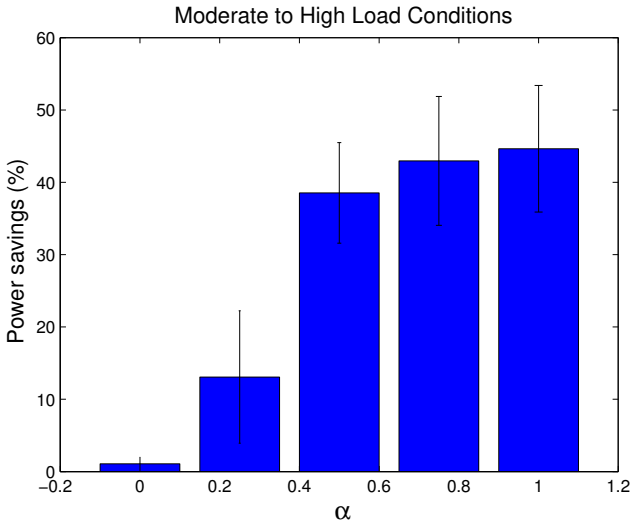


FIGURE 6.4: Power Savings for different sets α

Figure 6.6 shows the mean throughput per UE for different sets of α. The mean throughput per UE shows the highest values when power consumption reduction is forfeited (*i.e*, $\alpha = 0$). The mean throughput per UE is reduced when α increases, making room for power savings. In fact, for high values of α, less BBUs are activated to reduce the power consumption. Thus, lower throughputs are derived within BBUs. For the highest value of α, the mean throughput per UE is close to 2 Mb/s. The reason is that the algorithm is trying to reduce the throughput as much as possible to gain more power savings. However, the constraint of minimum throughput per UE (6.4) is guaranteed in all cases, and only values equal or higher than 2 Mb/s are realized.

Figure 6.7 displays the total derived throughput for different sets of α. Same as the mean throughput per UE, the highest values of total throughput are derived for the lowest values of α, because the number of active BBUs is high. As α increases, the total derived throughput decreases, making room for power savings. However, the re-association rate is higher.

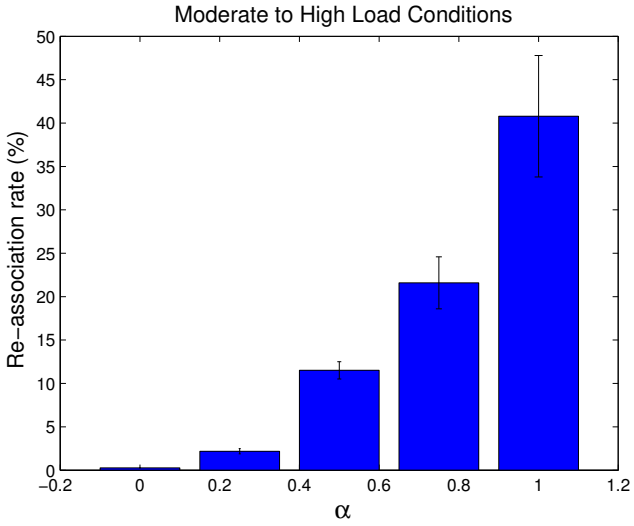


FIGURE 6.5: Re-association Rate of UEs for different sets α

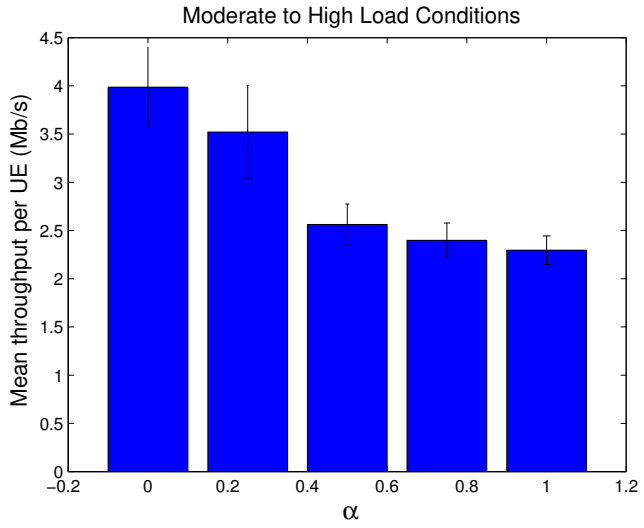


FIGURE 6.6: Mean Throughput per UE for different sets α

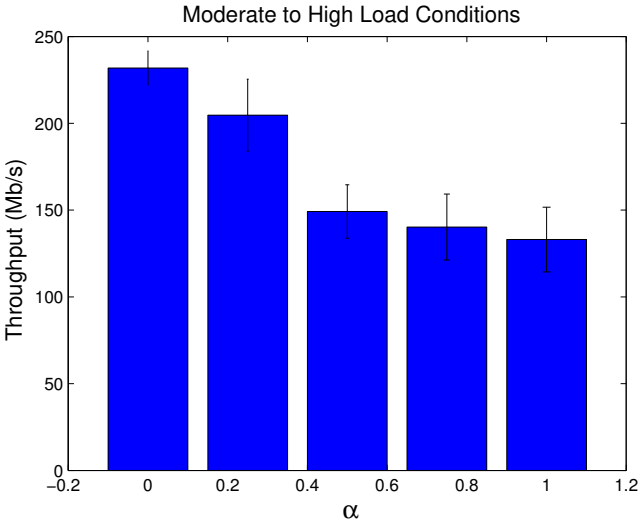


FIGURE 6.7: Total Throughput for different sets α

Low to Moderate Load Conditions

Within this scenario, we consider low to moderate load conditions, where the number of UEs per RRH varies between 1 to 8 UEs.
Figure 6.8 shows the number of active BBUs for different sets of α. We note that the same values of α are used to compare the results. We notice that the highest number of active BBUs is 4. Then, the values decrease as α increases. We note that for the case of low load conditions, the number of serviced UEs in the system is low. Thus, the re-association rate of UEs is low (cf. Figure 6.11). In that case, significant power savings are realized for almost all adopted values of α, even when more importance is devoted to re-association rate reduction, such as the case of $\alpha = 0.25$.

Figure 6.9 displays the power consumption metric for different sets of α. Similar to the number of active BBUs, the power consumption metric shows the highest value for $\alpha = 0$. However, for other values of α, the power consumption metric shows low values, because the re-association rate of UEs is low for all adopted cases.

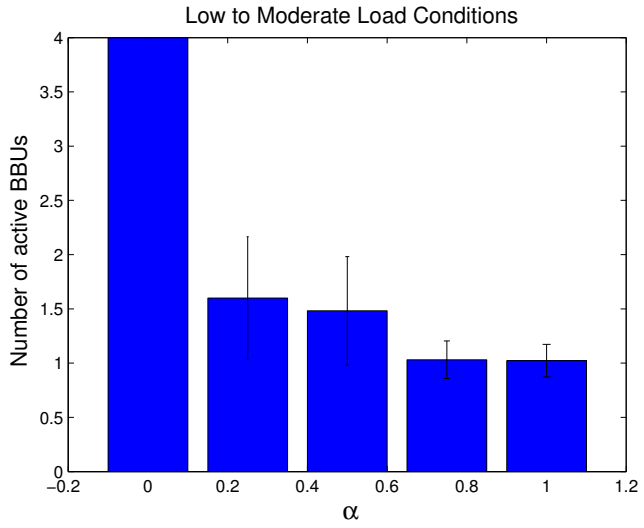


FIGURE 6.8: Number of Active BBUs for different sets α

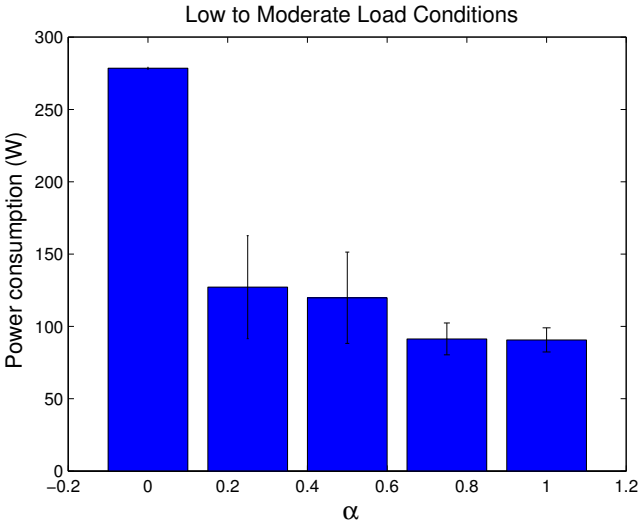


FIGURE 6.9: Power Consumption for different sets α

In Figure 6.10, we display the power saving metric for different sets of α. The lowest power saving value remains for $\alpha = 0$. For all other values of α, the realized power saving is above 70%, even for the cases where more importance is dedicated to re-association rate reduction, such as $\alpha = 0.25$.

Figure 6.11 shows the re-association rate of UEs for different sets of α. We clearly notice that the re-association rate of UEs is low for all the sets of α. As previously explained, the reason is that the number of UEs in the system is low, which lowers the rate of re-association. In the case of low load conditions, attributing more importance for power savings is more beneficial than stressing on re-association rate reduction when it comes to cost minimization.

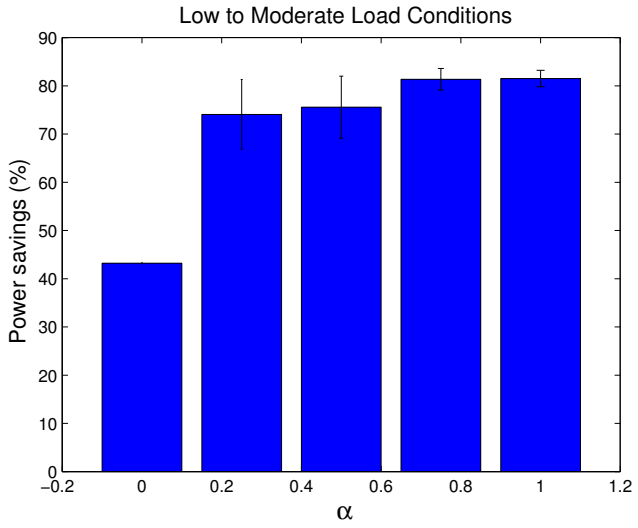


FIGURE 6.10: Power Savings for different sets α

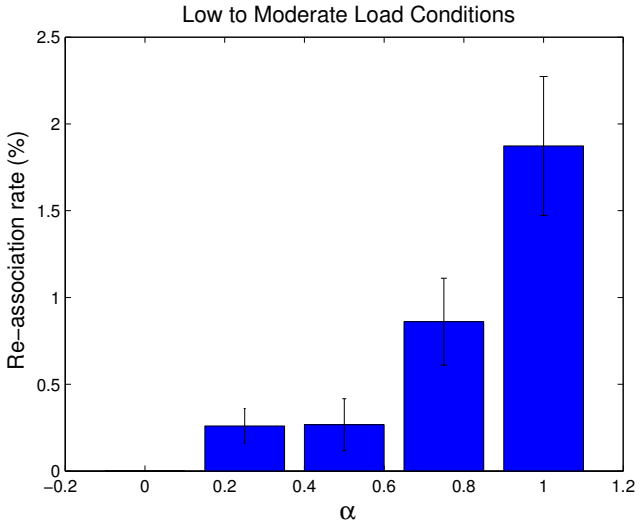


FIGURE 6.11: Re-association Rate of UEs for different sets α

Figure 6.12 shows the mean throughput per UE for different sets of α. We notice that the values are above 2 Mb/s for all cases. The reason is that the load conditions are low and the traffic model is assumed to be full buffer, consuming all the resources in the system. The highest value remains for the case of $\alpha = 0$ because the highest number of BBUs is activated.

In Figure 6.13, we display the total derived throughput in the system for different sets of α. Similar to the number of active BBUs and the power consumption metrics, the total derived throughput shows its highest value for the case of $\alpha = 0$. For all other sets of α, the total derived throughput shows lower values since more power saving is realized.

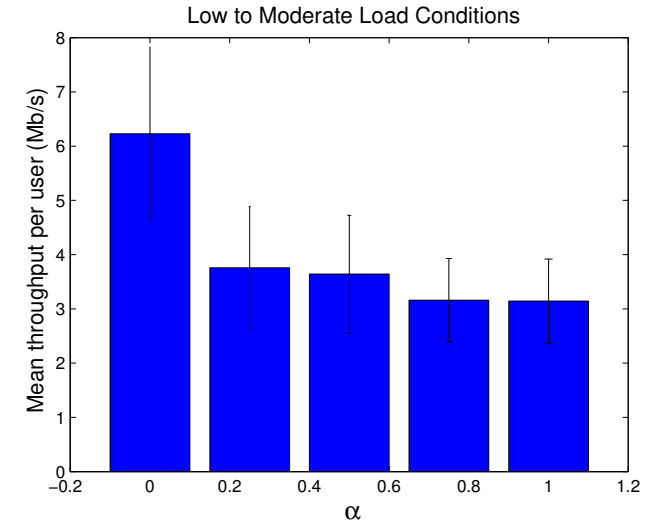


FIGURE 6.12: Mean Throughput per UE for different sets α

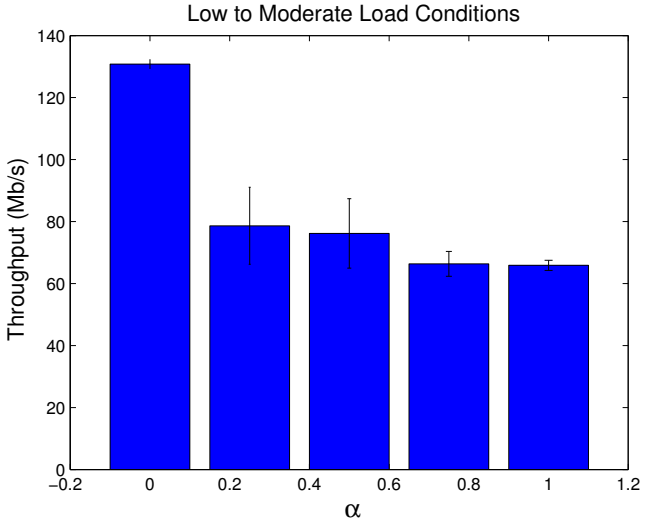


FIGURE 6.13: Total Throughput for different sets α

Concluding Remarks

According to the results, there is a clear tradeoff between the power saving metric and the re-association rate of UEs. However, for low to moderate load conditions, the re-association rate of UEs is low for all sets of α and β. In that case, stressing on high values of α (*i.e*, increasing the weight of power consumption reduction) would be more beneficial.

Figures 6.14 and 6.15 display the realized power savings versus the re-association rate of UEs for moderate to high and low to moderate load conditions respectively. In fact, the larger is the gap of difference between the re-association rate of UEs and the power saving, the better is the compromise between both metrics.

In figure 6.14, we notice that the largest gaps between both metrics remain for moderate values of α; for example, for $\alpha = 0.5$. The lowest gaps are for extreme cases such as, $\alpha = 1$ and $\beta = 0$ or vice versa. In Figure 6.15, the largest gaps are noticed for high values of α; for example, $\alpha = 1$. Whereas the lowest gap remains for $\alpha = 0$. This result again, proves that attributing more importance for power consumption reduction under low to moderate load conditions is more beneficial.

Within the following, we use the value of $\alpha = 0.5$ as a reference to compare the results under moderate to high load

conditions, and the value of $\alpha = 1$ as a reference to compare the results under low to moderate load conditions.

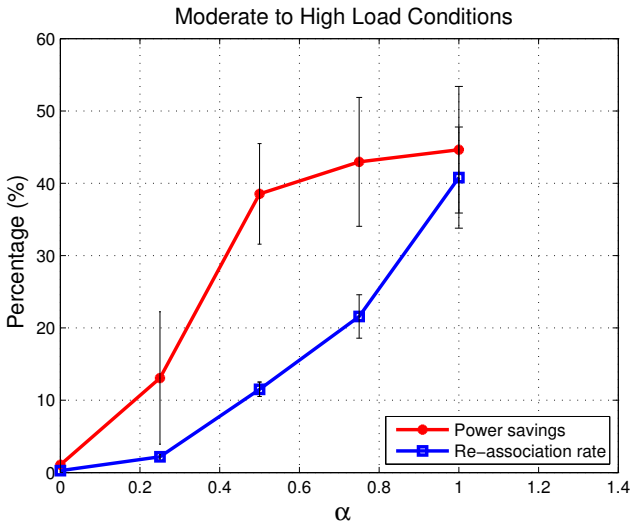


FIGURE 6.14: Power Savings vs. Re-association Rate

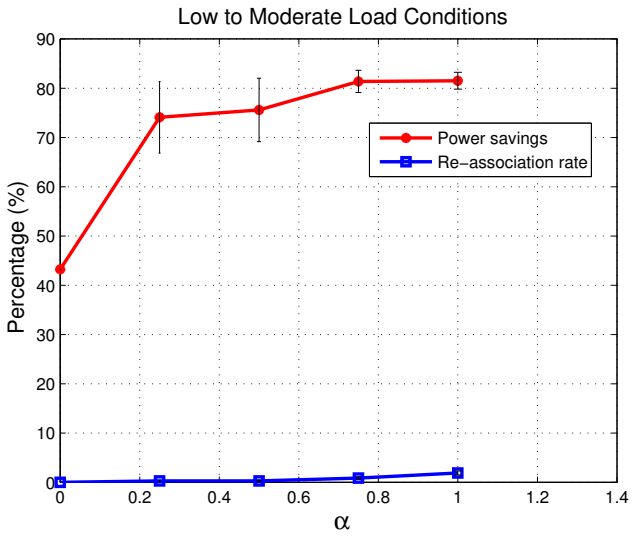


FIGURE 6.15: Power Savings vs. Re-association Rate

As the optimization problem is a SPP, it belongs to the NP-complete category [19]. It further belongs to the ILP category, since one variable (*i.e*, x_c) is optimized. In such case, when input parameters (*i.e*, the number of RRHs) are large, the problem becomes intractable. For the sake of complexity reduction, we introduce within the following section a greedy heuristic with less computational requirements, and which provides a good compromise between power savings and re-association rate reduction, as shown by the simulation results in subsection 6.6.1.

6.4 Heuristic Solution

Our proposed heuristic mainly consists of two stages described in algorithms 4 and 5. Algorithms 4 and 5 are executed sequentially.
We define $\mathcal{T} = \{t_{\mathcal{P}_1}, ..., t_{\mathcal{P}_c}, ...t_{\mathcal{P}_L}\}$ to be the set of all throughputs per UE derived in each cluster $c \in \mathcal{P}$, where $\mathcal{P}$ is the existing partition of RRHs in the system at the current decision epoch. L is the total number of active BBUs. We also define $\mathcal{U} = \{u_{c_1}, ..., u_{c_r}, ..., u_{c_N}\}$ to be the set of the number of UEs served by each RRH $r \in c$. N is the total

number of RRHs in cluster c. Finally, we define $\mathcal{B}$ to be the set of all BBUs having low to moderate load conditions. To obtain the set $\mathcal{B}$, we sort the clusters in an increasing order of throughput per UE, and we choose the second half of them.

Stage 1 describes the behavior of our heuristic when there are clusters that violate constraint (6.4). In other words, serviced UEs of such clusters are not guaranteed their minimal rate at the current decision epoch. At this stage, the algorithm starts by examining whether the lowest throughput per UE, given by a cluster $c \in \mathcal{P}$, obeys constraint (6.4). If not, it moves the RRH that serves the least number of UEs to the BBU that has the second lowest throughput per UE and respects constraint (6.4) (steps 5 to 19). The motivation behind this is to lower the amount of re-associated UEs. To do so, the heuristic examines the set $\mathcal{V}$ of available clusters that have not been tested (step 6, 7). If the heuristic does not find a cluster obeying constraint (6.4), to embrace the RRH in question, it activates an additional BBU (steps 22, 23). It continues on running until there are no clusters violating constraint (6.4).

Stage 2 describes how the heuristic tries to minimize the number of active BBUs by examining the maximum throughput per UE given by a cluster $c \in \mathcal{P}$. The heuristic tries to associate all the RRHs from cluster c to the BBU that has the most moderate throughput per UE (*i.e.,* the BBU that has the lowest throughput per UE in set $\mathcal{B}$) and respects constraint (6.4) (steps 6 to 16). If no such cluster is found, the algorithm stops the search (step 19).

1 Form the set $\mathcal{T}$;
2 $LessThrperUser \leftarrow min_{c\in\mathcal{P}}(t_{\mathcal{P}_c})$;
3 Form the set $\mathcal{U}$;
4 $\mathcal{V} \leftarrow \mathcal{P} - c$;
5 **while** $LessThrperUser < D_{min}$ **do**
6 **if** $\mathcal{V} \neq \emptyset$ **then** ;
7 Select $min_{c'\in\mathcal{V}}(t_{\mathcal{P}_{c'}})$;
8 $LessUser \leftarrow min_{r\in c}(\mathcal{U})$;
9 $c'' \leftarrow c' \cup \{r\}$;
10 **if** $t_{\mathcal{P}_{c''}} < D_{min}$ **then** ;
11 $\mathcal{V} \leftarrow \mathcal{V} - c'$;
12 Go to step 6;
13 **else**
14 $c' \leftarrow c''$;
15 Update $\mathcal{P}$ and $\mathcal{T}$;
16 $LessThrperUser \leftarrow min_{c\in\mathcal{P}}(t_{\mathcal{P}_c})$;
17 Update $\mathcal{U}$;
18 $\mathcal{V} \leftarrow \mathcal{P} - c$;
19 Go to step 5 ;
20 **end**
21 **else**
22 Add cluster $\{r\}$;
23 Update $\mathcal{P}$ and $\mathcal{T}$;
24 $LessThrperUser \leftarrow min_{c\in\mathcal{P}}(t_{\mathcal{P}_c})$;
25 Update $\mathcal{U}$;
26 $\mathcal{V} \leftarrow \mathcal{P} - c$;
27 Go to step 5 ;
28 **end**
29 **end**

Algorithm 4: Stage 1

Form the set $\mathcal{B}$;
$MaxThrperUser \leftarrow max_{c\in\mathcal{P}}(t_{\mathcal{P}_c})$;
$\mathcal{B} \leftarrow \mathcal{B} - c$;
$Stop = 0$;
while $Stop \neq 1$ **do**
    if $\mathcal{B} \neq \emptyset$ **then** ;
    Select $min_{c'\in\mathcal{B}}(t_{\mathcal{P}_{c'}})$;
    $c'' \leftarrow c' \cup \{c\}$;
    if $d_{\mathcal{P}_{c''}} < D_{min}$ **then** ;
    $\mathcal{B} \leftarrow \mathcal{B} - c'$;
    Go to step 6;
    else
        $c' \leftarrow c''$;
        Update $\mathcal{P}$, $\mathcal{B}$ and $\mathcal{T}$;
        $MaxThrperUser \leftarrow max_{c\in\mathcal{P}}(t_{\mathcal{P}_c})$;
        Go to step 6 ;
    end
    else
        $Stop = 1$;
    end
end

Algorithm 5: Stage 2

Heuristic Complexity

The heuristic proposition consists of two stages. Each stage has its own complexity. We consider a partition of RRHs that consists of k clusters, where $1 \leq k \leq R$. To study the heuristic complexity, we consider the cases of $k = 1$ and $k = R$, since they are extreme cases, and represent the highest complexities.

— When $k = 1$, the partition $\mathcal{P}$ consists of one active BBU. In the worst case, extreme high load conditions occur between two consecutive decision epochs, causing the re-association of each RRH to one BBU. The total number of steps the algorithm runs within the first stage is: $1 + 2 + ...R - 2$. As a matter of fact, the first stage runs until all the clusters respect a minimum level of throughput requirement per UE. In a first step, the algorithm re-associates the less loaded RRH of the highest loaded cluster to one BBU. In a second step, it associates the second less loaded RRH to one BBU, but verifies first whether it can be associated to the newly activated BBU. As the number of clusters increases, the number of verification steps increases, until all RRHs are re-associated, each to one BBU. In that case, stage 2 verifies within one step whether a lowly loaded cluster can be associated to the most moderately loaded one (cf. Algorithm 5). This would be unfeasible since the constraint of QoS would be violated. In that particular case, the first stage complexity would be equal to $1 + 2 + ...(R - 2) = \frac{(R-2)(R-1)}{2}$, and to 1 for the second stage. When the total number R of RRHs is high, the complexity is quadratic.

— When k is equal to R, the partition $\mathcal{P}$ of RRHs consists of the one-to-one association scheme. In extreme cases, a serious drop in the level of load conditions occurs between two consecutive decision epochs. Thus, one cluster (*i.e*, BBU) would be enough to accommodate the traffic demand. In that case, the first stage performs no action, since all clusters respect the constraint of QoS. Stage 2 verifies within a total number of $R - 1$ steps, whether all RRHs can be associated together. In that case, the complexity is linear.

Assuming a scenario where the first stage is executed with its highest complexity. In that case, the second stage is executed with its lowest complexity. In other terms, the two stages cannot be executed with their highest complexities simultaneously. We further note that the highest complexities of both algorithms are very rare cases, since the period of re-clustering decision is assumed to be well tuned in order to cope with the traffic variations.

6.5 Hybrid Approach of Joint Power Consumption and Re-association Reduction

In this section, we tackle the problem by proposing a hybrid version of the joint Power Consumption and Re-association rate Reduction for the purpose of reducing the signaling load overhead of the centralized schemes. In particular, our proposition relies on the previous proposition in chapter 5. Recall that our proposition is compounded of two stages executed in an iterative manner: A first stage (cf. paragraph 6.5.1) is used to set the BBU-RRH association schemes, and relies on a potential game among the RRHs to associate to a predefined set of clusters (*i.e*, BBUs). A second stage, is used to fix the number of active BBUs according to the realized throughput within each cluster.

6.5.1 Hybrid Load- and Interference-Aware Scheme with Re-association Consideration

Non-Cooperative Game for BBU-RRH Association

The general framework of the game $\mathcal{H} = (\mathcal{R}, \mathcal{C}, Cpu)$ is expressed as follows:
— The set $\mathcal{R}$ of players, defining the set of RRHs.
— The set $\mathcal{P}$ of available BBUs (*i.e,* strategies). An action of a RRH is selecting one of the available BBU.
— The vector y_r defines a strategy of RRH r. The binary variables $y_{r,c}$ defining the association between RRH r and cluster c. Therefore, $\mathbf{y} = (y_r)_{r\in\mathcal{R}} \in \mathcal{C}$ defines a pure strategy profile, and $\mathcal{C} = C_1 \times C_2 \times ... \times C_R$ would be the space of all profiles.
— The set of cost functions $\{Cpu_1, Cpu_2, ..., Cpu_R\}$ defining the players' preferences over the available strategies.

Cost Function The cost function that each RRH seeks to minimize is given as follows:

$$Cpu_r(y_r, y_{-r}) = \sum_{c\in\mathcal{P}} y^t_{r,c}\big[\alpha\alpha'\big(\underbrace{\sum_{r'} n_{r'} y^t_{r',c}}_{\text{Term 1}} + \underbrace{n_r \sum_{r'\neq r} y^t_{r',c}}_{\text{Term 2}}\big) + \beta\beta'\underbrace{\Big(\sum_{r'\neq r} G_{r',r}(1 - y^t_{r',c}) + N_0\Big)}_{\text{Term 3}} + \gamma\gamma'\big(\underbrace{n_r\big(1 - y^{t-1}_{r,c}\big)}_{\text{Term 4}}\big)\big], \tag{6.8}$$

where n_r and $n_{r'}$ define the number of UEs associated to RRHs r and r' respectively at the current decision epoch (t). The two binary variables $y^t_{r,c}$ and $y^{t-1}_{r,c}$ express the current and previous associations of RRH r to cluster c at decision epochs (t) and $(t-1)$ respectively. $G_{r,r'}$ is the channel gain between RRHs r and r'. α, β and γ are the corresponding weights of the load, interference and re-association costs respectively. α', β' and γ' are the normalization factors.
As explained in chapter 5, terms 1 and 2 reflect the load cost of RRH r, affected by the congestion impact of other RRHs that share the common radio resources of cluster c. In particular, term 1 is the total number of UEs associated to RRHs r', mapped to cluster c, whereas term 2 works as a stressing factor to emphasize the relative load impact of RRH r onto other RRHs $r' \neq r$ when its load n_r is high on one hand, and when the number of RRHs associated to the same BBU is high, on the other hand. Term 3 expresses the inter-cluster interference which is canceled between RRH r and other RRHs that have chosen the same BBU. Term 4 reflects the re-association cost of RRH r when the latter changes its cluster c within the current decision epoch (t). Precisely, if RRH r stays associated to the same cluster, at two consecutive decision epochs, the cost is zero. It is proportional to the number of connected UEs n_r when RRH r associates to a different cluster.

The Pure Nash Equilibrium The game at hand $\mathcal{H} = (\mathcal{R}, \mathcal{C}, Cpu)$ is a combination of two games: The first game $\mathcal{G} = (\mathcal{R}, \mathcal{C}, Cp)$ is the same as in section 5.4 in chapter 5, where each player cost function $Cp_r(y_r)$ is expressed as in (5.10). The second game $\mathcal{U} = (\mathcal{R}, \mathcal{C}, Cu)$ is a constant game, where each player's action only affects its own cost function $Cu_r(y_r)$, and not the costs of other players. The cost $Cu_r(y_r)$ is equal to term 4 in (6.8), reflecting the impact of re-association of RRH r, when the latter changes its BBU c between two consecutive decision epochs (t) and $(t-1)$.
Game $\mathcal{G}$ is an exact potential game as previously proved in chapter 5. Game $\mathcal{U}$ is a constant game, and thus, an

exact potential game. Game $\mathcal{H}$ is a convex combination of $\mathcal{G}$ and $\mathcal{U}$, again on the same action set, where the cost function for each player r is $(1-\gamma)\cdot Cp_r(y_r)+\gamma\cdot Cu_r(y_r)$. As the set of exact potential games is convex, our game $\mathcal{H}$ is an exact potential game [77].
In our work, we resort to the Best Response and the Replicator Dynamics to attain the PNE. The Best Response and Replicator Dynamics were previously described in paragraph 5.3, in chapter 5.

BBU Activation/Deactivation

The second stage behavior is similar to the one described in paragraph 7, in section 5.3. The only difference is within the deactivation process of a BBU. In particular, When a BBU is deactivated, the algorithm first checks the BBU with the lowest number of UEs, and then turns it OFF. The reason behind this is to lower the HO rate among connected UEs.

6.5.2 Signaling Load Analysis

In this subsection, we analyze the amount of signaling load overhead generated by the centralized and hybrid schemes.
In section 5.5, we have stated that the RRHs have to periodically report their load and radio conditions to the centralized cloud, where the decisions are taken. In this chapter, the problem is treated from a dynamic perspective, and the objective is to jointly optimize the power consumption and the re-association rate of UEs. Thus, at the current decision epoch t, the centralized cloud has to be aware of the previous BBU-RRH associations. In other terms, the binary variables $y_{r,c}^{t-1}$ of each RRH have to be known. In fact, such information does not need to be sent by RRHs, since it is already present in the cloud. Consequently, the total amount of signaling load overhead of centralized schemes is calculated as in (5.23), in section 5.5 of chapter 5.
As for the hybrid schemes, RRHs have to send their final decision matrix $(\boldsymbol{M})_{R\times S}$ to the centralized cloud, as stated in chapter 5. As previously explained, their radio conditions can be locally obtained through the CQI feedback of UEs. As for the clusters load conditions used by RRHs to calculate their own cost functions, it can be obtained from the edge cloud, without the centralized cloud assistance. Their previous associations $y_{r,c}^{t-1}$ can also be locally obtained, and used to calculate their own cost functions. The total amount of signaling load overhead is calculated as in section 5.5 of chapter 5.
Figure 6.16 shows the signaling load overhead, generated by the centralized and the hybrid schemes under two different load conditions (low to moderate and moderate to high). We note that the results are displayed for a set of seven cells, and the same set of parameter values are used as in table 5.1, in section 5.5, to assess the results. The centralized scheme shows a fixed value whatever the load conditions are, since the same amount of signaling overhead is sent at each decision epoch.

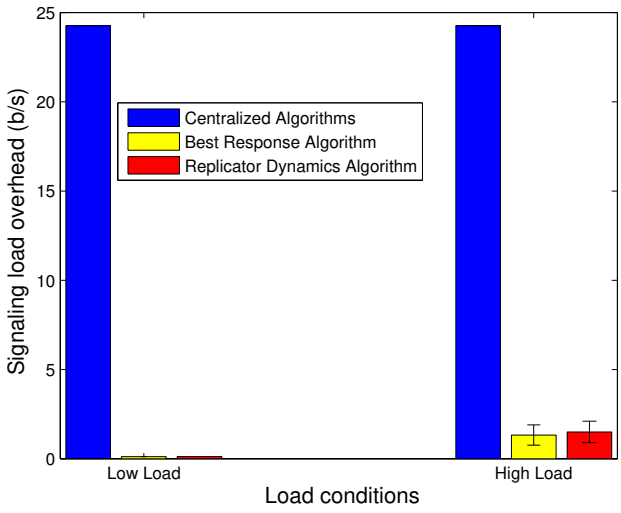


FIGURE 6.16: Signaling Load Overhead

The hybrid schemes show very low values of signaling load overhead for all load conditions. However, the signaling load overhead at low load conditions, is less than the case of high load conditions. As explained in chapter 5 (cf. subsection 5.5.3), the signaling load overhead for the hybrid schemes, is a function of the number of iterations executed by the second stage, and the number of strategies the RRHs compete on. As the number of strategies increases for high load conditions, the decision matrix dimensions ($R \times S$) would be higher, deriving more signaling load overhead. However, as a dynamic load scenario is considered, the needed number of active BBUs to accommodate the traffic demand will not significantly differ between two consecutive decision epochs, making the number of iterations launched by the second stage to attain the right number of active BBUs low (cf. section 7).

6.6 Performance Evaluation

In this section, the results of the heuristic (cf. subsection 6.4) and the hybrid schemes (cf. subsection 6.5) are provided. We note that the same simulation sets are used as in subsection 6.3, and a network topology consisting in seven cells is considered. Further, we use the same metrics as in subsection 6.3 to compare the results. Additionally, we examine the performance of the various algorithms in terms of computational requirement.

6.6.1 Optimal Solution vs. Heuristic vs. Bin Packing

The results of the heuristic are compared to the optimal solution and the Bin Packing schemes. As the Bin Packing model might give many partition solutions, minimizing the number of used bins for one particular case, we display the average result given by all the optimal partitions. We recall that the optimal solution at $\alpha = 1$ is used to compare the results under low load conditions, and $\alpha = 0.5$ is used under high load conditions for the previously explained reason (cf. paragraph 6.3).

Figure 6.17 displays the number of active BBUs as a function of two different load conditions: Low to moderate load conditions, denoted as low load, and moderate to high load conditions, denoted as high load. Recall that for the case of low load, the number of UEs per cell varies between 1 and 8. For the case of high load, the number of UEs varies between 8 and 15. We notice that for low load, the heuristic shows the same value as the optimal case (1 active BBU), whereas the Bin Packing scheme displays more than 2 active BBUs. Recall that the Bin Packing scheme is an Interference-Oblivious approach and is unaware of the level of spectral efficiency at the BBU level, which varies when the subset of RRHs varies. In case of high load, the highest value also remains for the Bin Packing scheme, whereas the heuristic and the optimal cases show similar values.

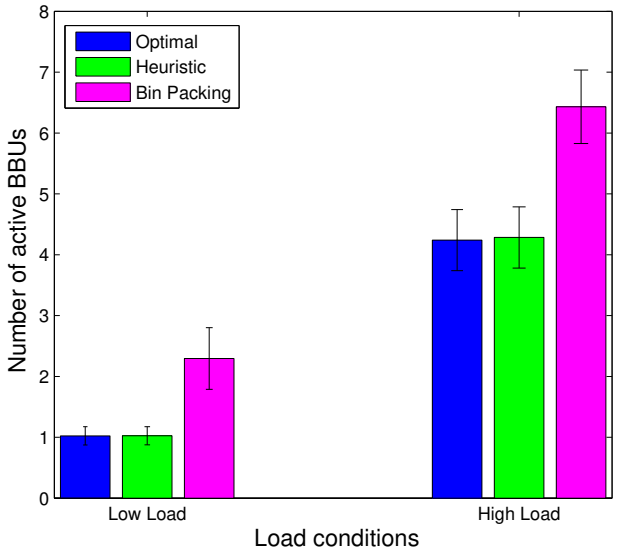


FIGURE 6.17: Number of Active BBUs: Optimal vs. Heuristic vs. Bin Packing

In Figure 6.18, the power consumption metric is displayed as a function of load conditions. Similar to the number

of active BBUs, the power consumption values shown by both the optimal solution and the heuristic scheme are the same for low load conditions. The Bin Packing scheme shows more power consumption as it activates more BBUs. For the case of high load conditions, the heuristic scheme shows slightly higher power consumption than the case of the optimal solution.

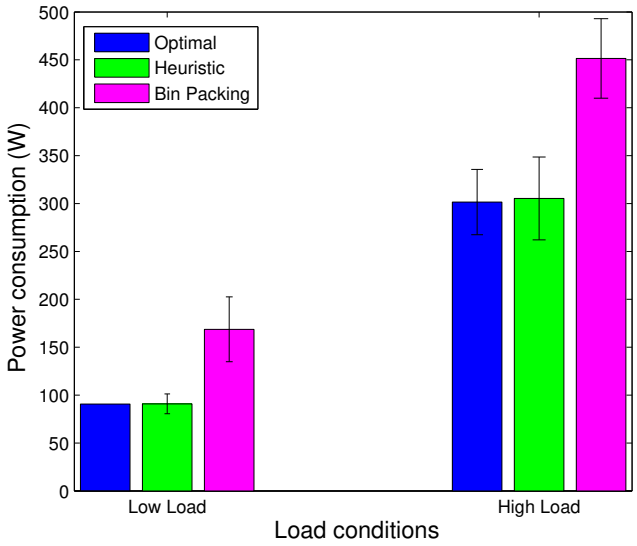


FIGURE 6.18: Power Consumption: Optimal vs. Heuristic vs. Bin Packing

Figure 6.19 displays the power saving metric as a function of load conditions. For all load conditions, we notice that the optimal solution and the heuristic show improved power saving in comparison with the Bin Packing, since they activate less BBUs.

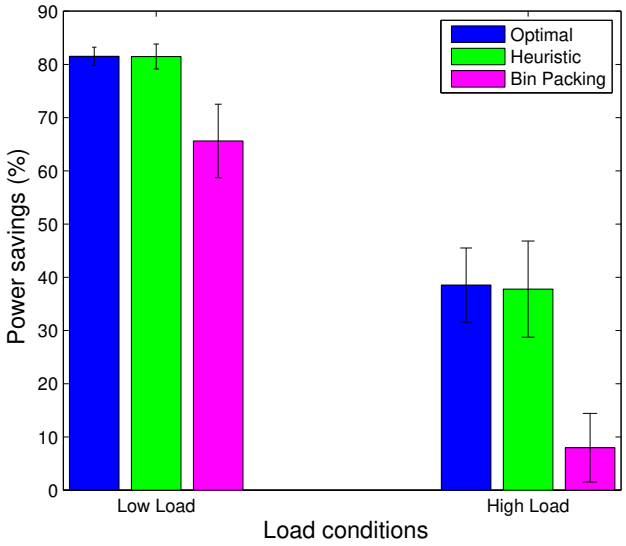


FIGURE 6.19: Power Savings: Optimal vs. Heuristic vs. Bin Packing

In Figure 6.20, the re-association rate of UEs is displayed as a function of load conditions. For low load conditions, the Bin Packing case shows significant higher re-association rate of UEs in comparison with the optimal and the heuristic schemes. In fact, the Bin Packing scheme does not take into consideration the re-association rate of UEs. Further, it activates more BBUs than other cases, which increases the rate of re-associations between two

different decision epochs, since RRHs might re-associate between many BBUs. For high load conditions, the heuristic shows slightly higher rate of re-association in comparison with the optimal case. The highest value remains for the Bin Packing scheme, since it activates more BBUs at one hand, and overlooks such metric, on the other hand. Our heuristic performs efficiently in terms of realized power saving and reduced re-association rate. For low load conditions, it acts similarly as the optimal case for $\alpha = 1$, and for high load conditions, its behavior, in terms of power savings and re-association rate, is close to the case of the optimal solution with $\alpha = 0.5$. This demonstrates its efficiency in striking a good compromise between both metrics.

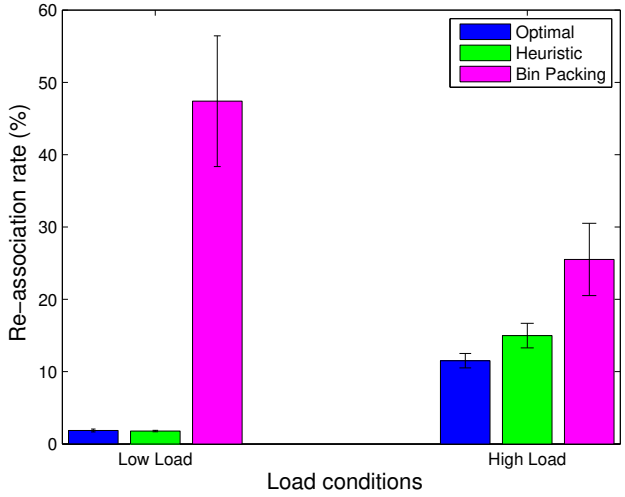


FIGURE 6.20: Re-association Rate of UEs: Optimal vs. Heuristic vs. Bin Packing

Figure 6.21 shows the mean throughput per UE as a function of load conditions. The highest mean throughput per UE values is realized by the Bin Packing scheme since it activates the highest number of BBUs, and a full buffer traffic scheme is considered, consuming all the resources. However, both cases of optimal solution and heuristic schemes do not show values less than 2 Mb/s and respect a minimum throughput requirement per UE.

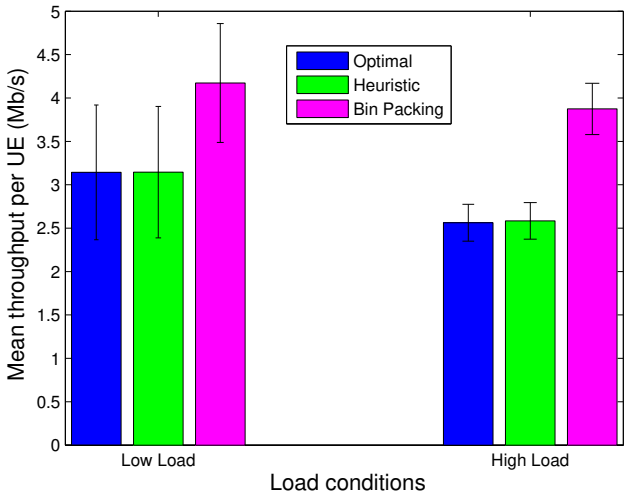


FIGURE 6.21: Mean Throughput per UE: Optimal vs. Heuristic vs. Bin Packing

Figure 6.22 displays the total derived throughput as a function of load conditions. Similar to the mean throughput per UE, the highest values are for the Bin Packing scheme, since more BBUs are activated. However, the Bin Packing scheme does not realize the same level of power economy as the optimal solution and the heuristic schemes. It further increases the rate of re-association.

The last given metric is the execution time of the heuristic in comparison with the optimal case as a function of load conditions. The results are executed over a hardware of 64-bit, Intel 5 core, 2.5 Ghz. Figure 6.23 shows that the execution time of the heuristic is much lower than that of the optimal case. The heuristic provides good performance with lower complexity, and is consequently efficient for implementation in practice.

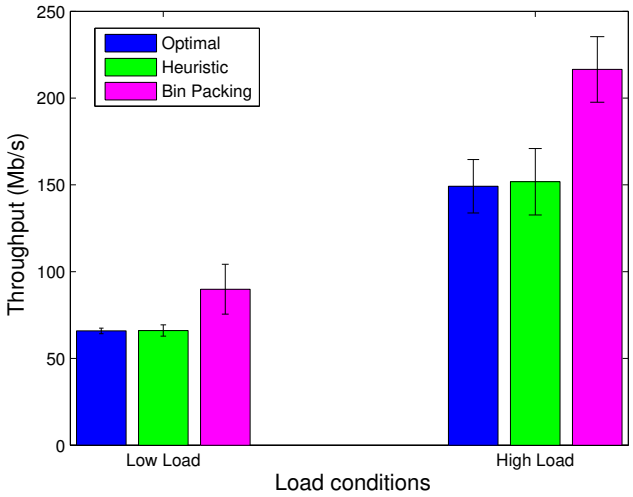


FIGURE 6.22: Total Throughput: Optimal vs. Heuristic vs. Bin Packing

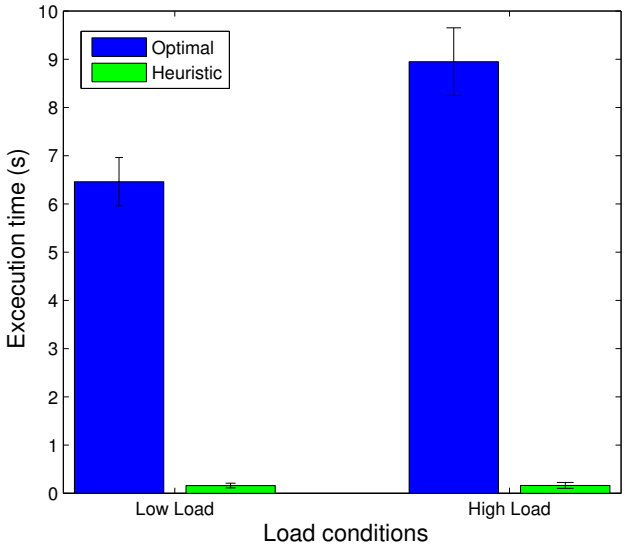


FIGURE 6.23: Execution Time: Optimal vs. Heuristic

6.6.2 Centralized Algorithm vs. Hybrid Algorithm with Re-association vs. Hybrid Algorithm without Re-association vs. Bin Packing

In this subsection, we provide the performance evaluation of the hybrid schemes in comparison with the centralized scheme and the Bin Packing approach. Similar to the heuristic, the performance are compared under two different load conditions (cf. subsection 6.6.1), and the same sets of α (in equation (6.2)) in the centralized scheme are employed to compare the results. We further note that α, β and γ (in equation 6.8) are set to 0.75, 0.125 and 0.125 respectively). The reason behind this is to motivate the RRHs to associate to the least loaded clusters in

case of high load conditions. In fact, for high values of β, RRHs tend to associate to fewer clusters to lower the interference level. For high values of γ, RRHs tend to not change their clusters between two consecutive decision epochs to lower the rate of re-association.

Centralized Algorithm vs. Best Response Algorithm with Re-association vs. Best Response Algorithm without Re-association vs. Bin Packing

Figure 6.24 displays the number of active BBUs as a function of load conditions. We notice that for low load conditions, the centralized and the Best Response algorithms (with and without re-association consideration) show 1 active BBU, whereas the Bin Packing scheme activates more BBUs since it is unaware of the interference level in the system. For high load conditions, the centralized scheme activates the lowest number of BBUs. The two hybrid schemes (with and without re-association) activates slightly more BBUs. The highest number of active BBUs remains for the Bin Packing scheme.

Figure 6.25 shows the power consumption metric as a function of load conditions. Similar to the number of active BBUs, at low load conditions, we notice the lowest values of power consumption for the Best Response and the centralized schemes, since only 1 BBU is activated. For high load conditions, the hybrid schemes show slightly higher power consumption than the case of the centralized scheme. The highest power consumption remains for the Bin Packing scheme, for all load conditions.

Figure 6.26 displays the power saving metric as a function of load conditions. The highest values of power saving are recorded for the Best Response and the centralized schemes, at low load conditions. For high load conditions, the highest power saving is fulfilled by the centralized approach, whereas the lowest is realized by the Bin Packing scheme.

Figure 6.27 shows the re-association rate of UEs as a function of load conditions. At low load, the re-association rate for the centralized algorithm and the Best Response algorithms are similar. In fact, at low load, few BBUs are activated. Thus, the re-association rates are low for all schemes. As for the Bin Packing approach, it shows a significant higher re-association rate in comparison with other schemes, since it activates more BBUs at one hand, and it is unaware of the re-association rate, on the other hand. At high load, the Best Response algorithm with re-association consideration shows a slightly higher rate of re-association in comparison with the centralized approach. It also shows a good improvement in terms of re-association reduction in comparison with the hybrid scheme without re-association. The Bin Packing scheme shows the highest value of re-association rate in comparison with other schemes.

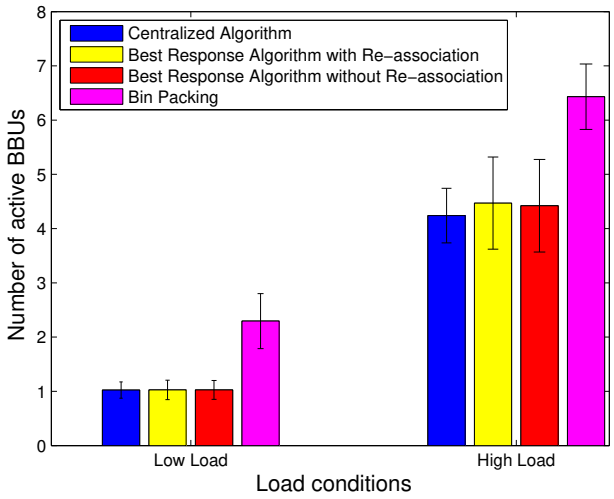


FIGURE 6.24: Number of Active BBUs: Centralized Algorithm vs. Best Response Algorithm with Re-association vs. Best Response Algorithm without Re-association vs. Bin Packing

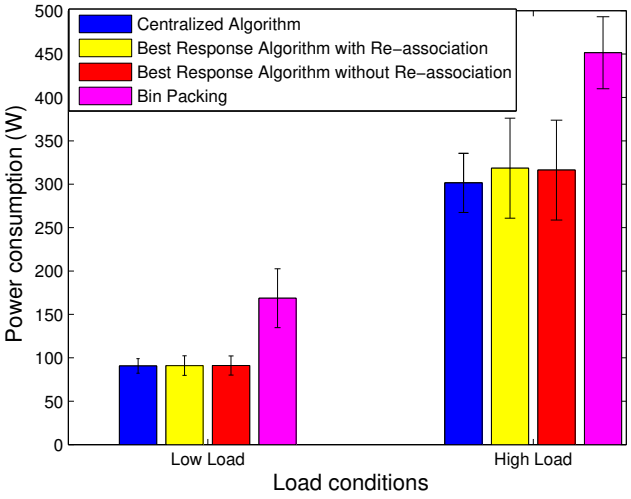


FIGURE 6.25: Power Consumption: Centralized Algorithm vs. Best Response Algorithm with Re-association vs. Best Response Algorithm without Re-association vs. Bin Packing

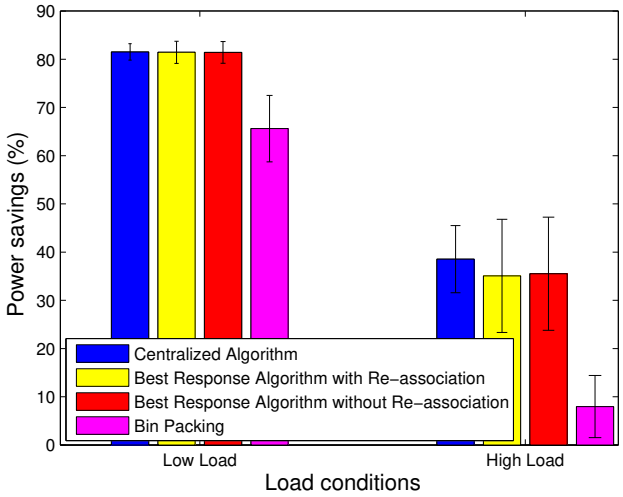


FIGURE 6.26: Power Savings: Centralized Algorithm vs. Best Response Algorithm with Re-association vs. Best Response Algorithm without Re-association vs. Bin Packing

Figure 6.28 displays the useful mean throughput per UE as a function of load conditions. At low load, the Bin Packing scheme shows higher mean throughput per UE in comparison with other schemes. As explained, when more BBUs are activated, all resources are consumed, since we have a full buffer traffic model. However, all schemes respect a minimum level of QoS and do not show values less than 2 Mb/s. At high load conditions, the centralized scheme shows slightly lower mean throughput per UE in comparison with the Best Response algorithm (with and without re-association). On the other hand, it shows slightly more power saving (cf. Figure 6.26). The Bin Packing scheme shows the highest mean throughput per UE for high load conditions, because more BBUs are activated. We note that the Best Response algorithm with re-association provides similar performance to the one without re-association, in terms of power saving and QoS (*i.e*, mean throughput per UE), while reducing the re-association rate of UEs.

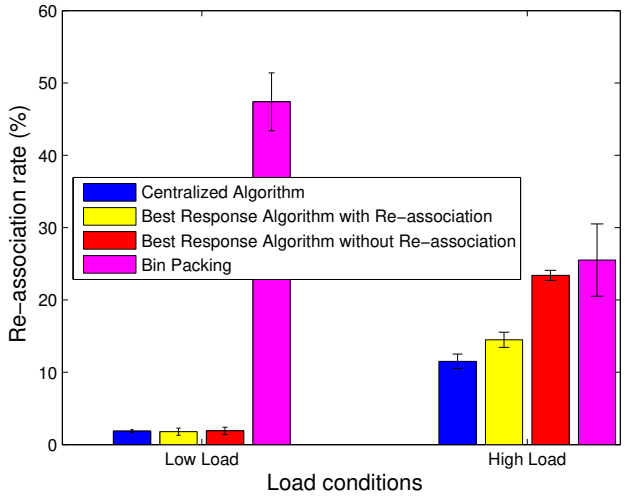


FIGURE 6.27: Re-association Rate of UEs: Centralized Algorithm vs. Best Response Algorithm with Re-association vs. Best Response Algorithm without Re-association vs. Bin Packing

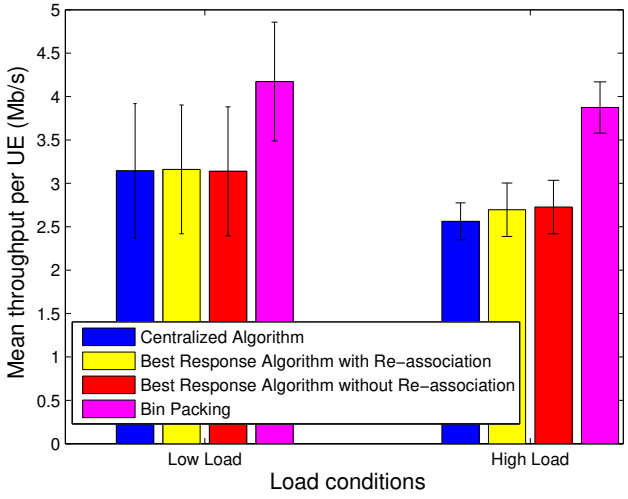


FIGURE 6.28: Mean Throughput per UE: Centralized Algorithm vs. Best Response Algorithm with Re-association vs. Best Response Algorithm without Re-association vs. Bin Packing

In Figure 6.29, the total derived throughput as a function of the load conditions is displayed. Similar to the mean throughput per UE, the highest value is shown by the Bin Packing scheme, for all load conditions. The lowest values are still for the centralized scheme, because it activates the lowest number of BBUs. On the other hand, it realizes the highest power savings.

Figure 6.30 displays the execution time of the centralized and the Best Response algorithms. We note that the algorithms are executed over a hardware of 64-bit, Intel 7 core, 2.5 Ghz. We notice that for all load conditions, the centralized scheme consumes more computational time in comparison with the Best Response algorithms with and without re-association. Obviously, under dynamic load conditions, taking into consideration the re-association rate of UEs in the hybrid proposition is very efficient, since it provides the same performance of the hybrid scheme without re-association, while reducing the re-association rate of UEs without affecting the computational time.

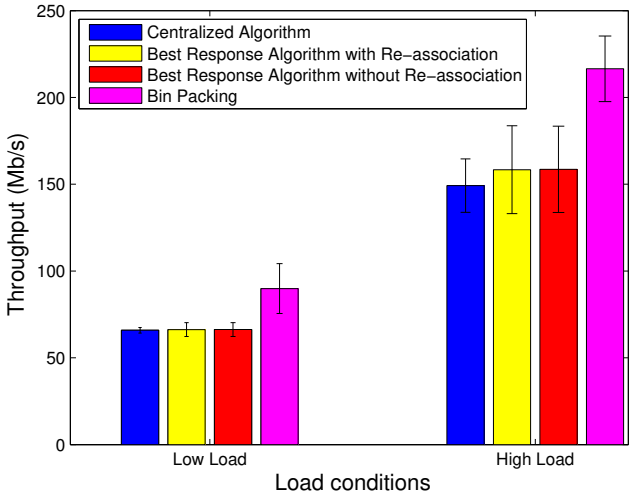


FIGURE 6.29: Total Throughput: Centralized Algorithm vs. Best Response Algorithm with Re-association vs. Best Response Algorithm without Re-association vs. Bin Packing

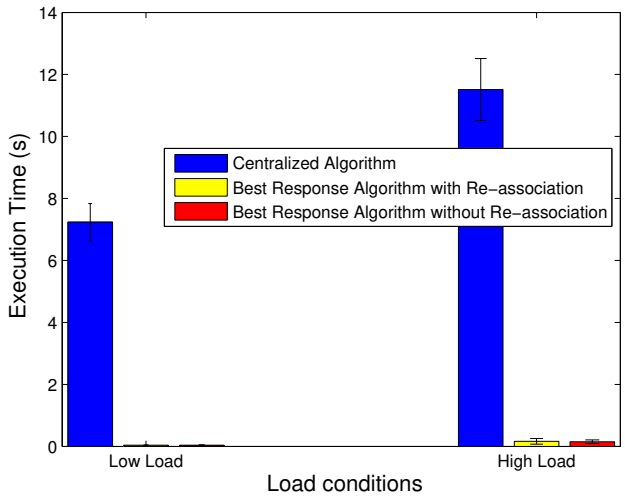


FIGURE 6.30: Execution Time: Centralized Algorithm vs. Best Response Algorithm with Re-association vs. Best Response Algorithm without Re-association

Centralized Algorithm vs. Replicator Dynamics Algorithm with Re-association vs. Replicator Dynamics Algorithm without Re-association vs. Bin Packing

Within this paragraph, we display the results of the Replicator Dynamics algorithm with and without re-association consideration in comparison with the centralized and the Bin Packing schemes.
Figure 6.31 displays the number of active BBUs as a function of load conditions. At low load conditions, the Replicator Dynamics algorithm (with and without re-association) provides the same number of active BBUs as the centralized scheme. The Bin Packing shows the highest number of BBUs because of its Interference-Oblivious behavior. At high load conditions, the Replicator Dynamics (with and without re-association) activates slightly more BBUs than the centralized scheme and the Best Response scheme (cf. Figure 6.24). Recall that the Replicator Dynamics is a reinforcement learning technique that is fully distributed. Thus, it shows a slightly degradation in performance in some cases. However, the Replicator Dynamics shows less activated BBUs in comparison with the Bin Packing approach, which does not take into consideration the interference level change in the network.

The power consumption metric as a function of load conditions is displayed in Figure 6.32. Similar to the number of active BBUs, the power consumption metric is the same at low load conditions for all of the centralized scheme and Replicator Dynamics algorithms, since they all activate 1 BBU. For high load conditions, the Replicator Dynamics algorithm (with and without re-association) shows higher power consumption values than the case of the centralized scheme and the Best Response scheme (cf. Figure 6.25), since they activate more BBUs. The highest power consumption values is still realized for the Bin Packing scheme, for all load conditions.

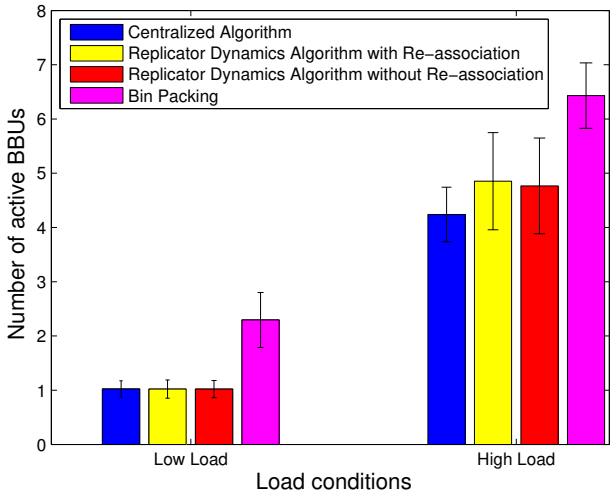


FIGURE 6.31: Number of Active BBUs: Centralized Algorithm vs. Replicator Dynamics Algorithm with Re-association vs. Replicator Dynamics Algorithm without Re-association vs. Bin Packing

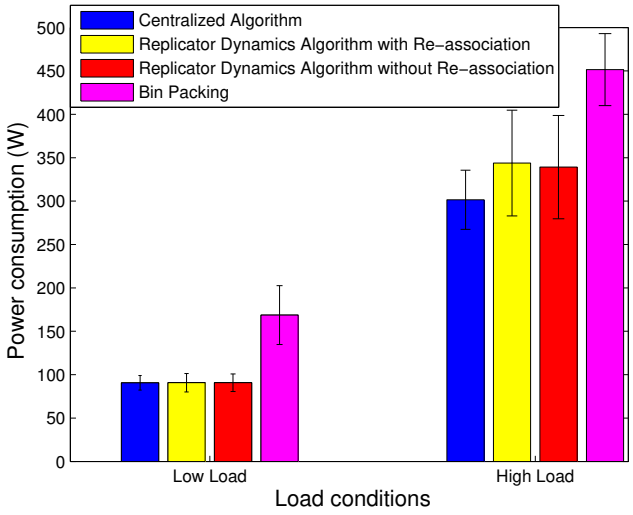


FIGURE 6.32: Power Consumption: Centralized Algorithm vs. Replicator Dynamics Algorithm with Re-association vs. Replicator Dynamics Algorithm without Re-association vs. Bin Packing

Figure 6.33 shows the power saving metric as a function of the load conditions. At high load conditions, the Replicator Dynamics algorithm (with and without re-association) shows less power saving than the centralized algorithm, as more BBUs are activated. However, more power savings are shown in comparison with the Bin Packing approach.

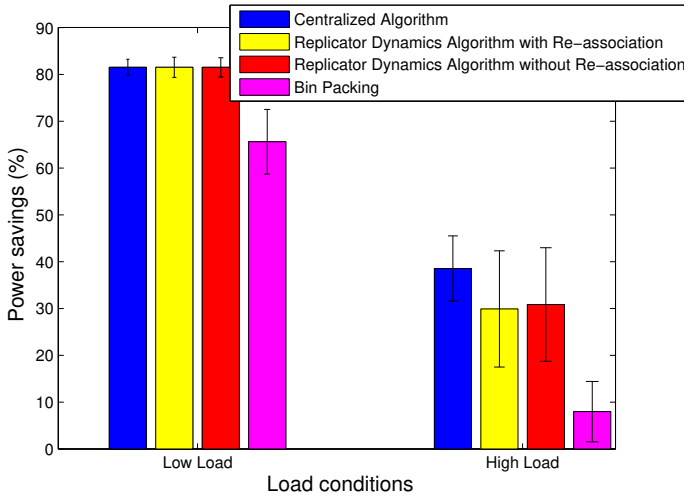


FIGURE 6.33: Power Savings: Centralized Algorithm vs. Replicator Dynamics Algorithm with Re-association vs. Replicator Dynamics Algorithm without Re-association vs. Bin Packing

Figure 6.34 displays the re-association rate of UEs as a function of load conditions. At low load conditions, the highest re-association rate is for the Bin Packing approach, as it activates more BBUs, and is unaware of the re-association factor. For high load conditions, the re-association rate of UEs shown by the Replicator Dynamics without re-association (*i.e*, $\gamma = 0$) is significantly high, and attains a level of 80%. Initially, the Replicator Dynamics is based on random drops on the set of available strategies (*i.e*, clusters). Thus, a RRH might re-associate to a new cluster with a high probability between two consecutive decision epochs, if the re-association factor is not considered. The re-association rate of UEs is significantly reduced by the Replicator Dynamics when it is taken into account (*i.e*, the Replicator Dynamics with Re-association), and attains a value of 14%. Taking into account such factor under dynamic load conditions is paramount, particularly, for the Replicator Dynamics which might degrade drastically the QoE, causing huge amount of HOs if the re-association factor is not carried out carefully.

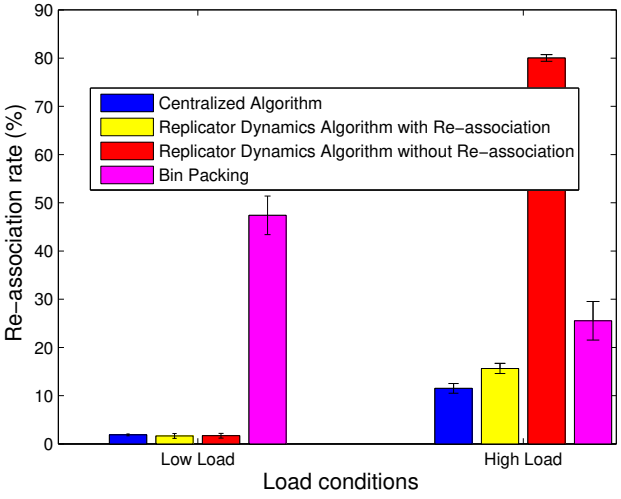


FIGURE 6.34: Re-association Rate of UEs: Centralized Algorithm vs. Replicator Dynamics Algorithm with Re-association vs. Replicator Dynamics Algorithm without Re-association vs. Bin Packing

The mean throughput per UE is displayed in Figure 6.35 as a function of load conditions. The highest values of mean throughput per UE are shown by the Bin Packing scheme, since it activates more BBUs, and a full buffer traffic scheme is considered. For high load conditions, we notice that the Replicator Dynamics with and without re-association shows the same values, which underlines the efficiency of the re-association factor that reduces

significantly the re-association rate of UEs without compromising the QoS.

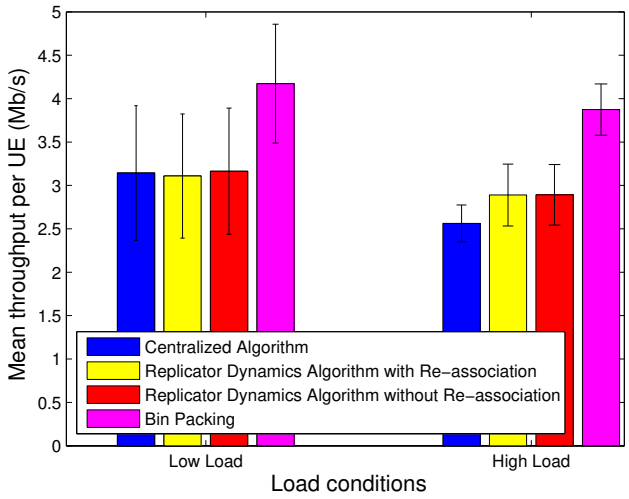


FIGURE 6.35: Mean Throughput per UE: Centralized Algorithm vs. Replicator Dynamics Algorithm with Re-association vs. Replicator Dynamics Algorithm without Re-association vs. Bin Packing

Figure 6.36 shows the total derived throughput as a function of load conditions. Again, we notice the highest values for the Bin Packing approach, since more BBUs are activated and more resources are consumed. We also notice that the realized throughput of the Replicator Dynamics with and without re-association is the same, which again highlights the efficiency of the re-association factor that does not reduce the QoS, while reducing the re-association rate of UEs.

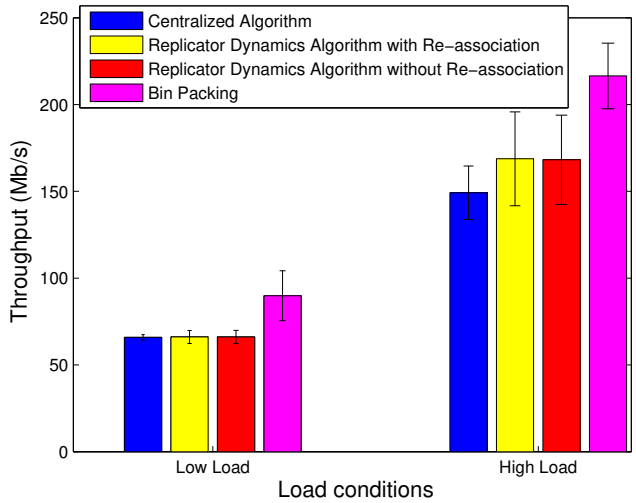


FIGURE 6.36: Total Throughput: Centralized Algorithm vs. Replicator Dynamics Algorithm with Re-association vs. Replicator Dynamics Algorithm without Re-association vs. Bin Packing

The last given metric is the execution time of the algorithms given by Figure 6.37, as a function of load conditions. As previously stated, the results are executed over a hardware of 64-bit, Intel 7 core, 2.5 Ghz. At low load conditions, the Replicator Dynamics algorithm shows less computational time than the centralized scheme. The reason is that the RRHs compete over a low number of strategies (*i.e*, BBUs). However, for high load conditions, the Replicator Dynamics algorithms show higher computational time with the increase in the number of strategies to compete on increases. We highlight the fact that the Replicator Dynamics operates independently while making its choice. To attain good results, the Replicator Dynamics needs time for convergence. However, it is necessary to carry out its

usage efficiently. For example, the Replicator Dynamics is efficient for the case of low number of strategies (*i.e*, the case of low load conditions). However, it could be hefty in terms of computational time when the number of strategies increases.

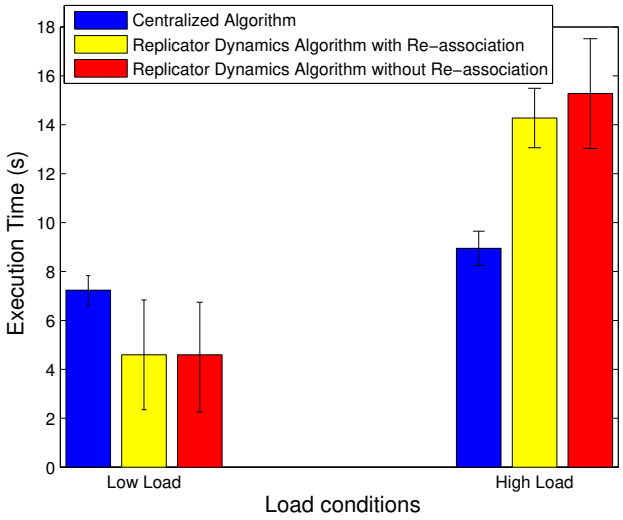


FIGURE 6.37: Execution Time: Centralized Algorithm vs. Replicator Dynamics Algorithm with Re-association vs. Replicator Dynamics Algorithm without Re-association vs. Bin Packing

6.7 Conclusion

In this chapter, we tackled the BBU-RRH association problem from a dynamic perspective. In particular, we have proposed many solutions to jointly optimize the power consumption and the re-association rate of UEs that occurs when the network load conditions vary. The main conclusions of the chapter are summarized within the following points:

- A tradeoff between the power saving and the re-association rate reduction exists: Activating all BBUs guarantee the QoS under any load condition, and avoids re-associating RRHs to BBUs. However, it hinders the energy efficiency. On the other hand, deactivating BBUs may lead to a resource shortage when fluctuations in the load conditions occur, which provokes the re-association of RRHs, in order to cope with the increase in traffic.
- Under low load conditions, prioritizing the power savings is more efficient, as the re-association rate of UEs is low in all cases. For high load conditions, attributing the same importance for power consumption and re-association rate reduction (*i.e*, $\alpha = 0.5$, $\beta = 0.5$) remains a good compromise between both metrics.
- The heuristic proposition achieves a good compromise between power saving and re-association rate reduction. In particular, for low load conditions, it shows close performance to the optimal solution for $\alpha = 1$, and close performance to the optimal case for $\alpha = 0.5$, under high load conditions.
- The hybrid proposition also shows good compromise in terms of power saving and re-association rate reduction. It further reduces the signaling load overhead generated by centralized approaches. Further, the hybrid scheme with re-association consideration shows similar performance to the hybrid proposition without re-association consideration, while reducing the re-association rate of UEs, especially in the case of Replicator Dynamics, where the re-association rate significantly increases if it is not taken into account.

Chapitre 7

On Common Spectrum and Active Infrastructure Sharing in C-RAN Architecture

Besides its derived benefits in terms of multiplexing gain, efficiency in resource utilization and power consumption reduction, the C-RAN architecture is a key enabler for cooperation among multiple MNOs. In particular, such architecture facilitates for common spectrum and active infrastructure sharing among MNOs that put in common their BBUs, RRHs and licensed spectrum for cooperation. Extended coverage and capacity are therefore derived, since MNOs rely on each others' resources in case of shortage. For example, UEs subscribed to one MNO may connect to a different one in case the latter provides better coverage in some regions. Seemingly, one MNO can access the operating spectrum of a different one to increase its capacity, deemed common spectrum network sharing. Further, BBUs are controlled in one single location, providing additional multiplexing gains and virtualization, which leads to efficient use of radio resources. In this chapter, we exploit the benefits of such cooperation within the C-RAN architecture. In fact, when centralized RAN management is allowed with MNOs cooperation, many benefits are derived, such as additional power savings resulted from BBU-RRH association optimization, and enhanced spectral and energy efficiency. In our work, we propose two different algorithms for MNOs cooperation in C-RAN: The first, deemed MAB algorithm, where common spectrum sharing is governed by a reinforcement learning solution (the MAB problem), and consists of an interference management technique, where MNOs can access each others' bands to mitigate interference within the shared RAN. The second, deemed CA algorithm, where the common spectrum sharing relies on the CA, which is a release 10 LTE specification technique to increase capacity. We compare both algorithms with the case where MNOs do not share their resources. The simulation results show that network performance are globally enhanced while MNOs cooperate together. Further, despite the realized power economy achieved in C-RAN architecture, additional power savings, and enhanced spectral and energy efficiency are realized when network cooperation is made possible.

7.1 Introduction

Network infrastructure sharing is being proposed as a promising solution. This paradigm involves a set of strategies, allowing MNOs to share their resources in order to attain the goal of power consumption reduction and enhanced resource utilization efficiency. Effectively, it is estimated about €2 billion of benefits from sharing part of a MNO cellular network [83]. The infrastructure sharing can be classified into three categories:

— The passive infrastructure sharing, consisting in sharing the sites, masts and building premises among MNOs.
— The active infrastructure sharing, consisting in sharing active network components such as antennas, switches and backhaul equipments.
— The roaming based sharing, when one MNO uses the coverage of a different MNO to serve its UEs at a certain region.

Early network sharing agreements involved the passive infrastructure sharing. Savings were achieved by sharing the basic resources, such as sites, accommodation, power and air conditioning. Because of the technical, commercial

and regulatory complexities, there are few recent examples of active infrastructure sharing, where MNOs share their active network components and radio resources. However, the need of additional savings and increased revenues is forcing the different regulators to overcome all of the stated complexities. For example, network infrastructure sharing is becoming an important part in 3GPP systems that define multiple scenarios which are stated below [84]:

1. *Multiple core networks sharing common radio access network scenario*: In this case, MNOs share their RAN but not their licensed spectrum. In particular, MNOs connect directly to their carrier within the shared RAN.
2. *Geographically split networks sharing scenario*: In this case, MNOs provide coverage for different parts of a country. However, together they provide coverage for the entire region. Many deployment scenarios are therefore distinguished:
   — Multiple MNOs employ national roaming for subscribed UEs. In this scenario, each RAN would be associated to one core network (cf. Figure 7.1).

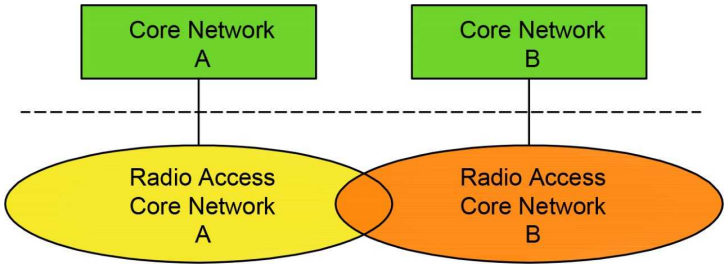


FIGURE 7.1: Geographically split network using national roaming

   — Each MNO has its individual core network connected to the shared RAN within the entire region. In that case, all MNOs have the possibility to use each others' allocated spectrum in the different parts of the entire region (cf. Figure 7.2).

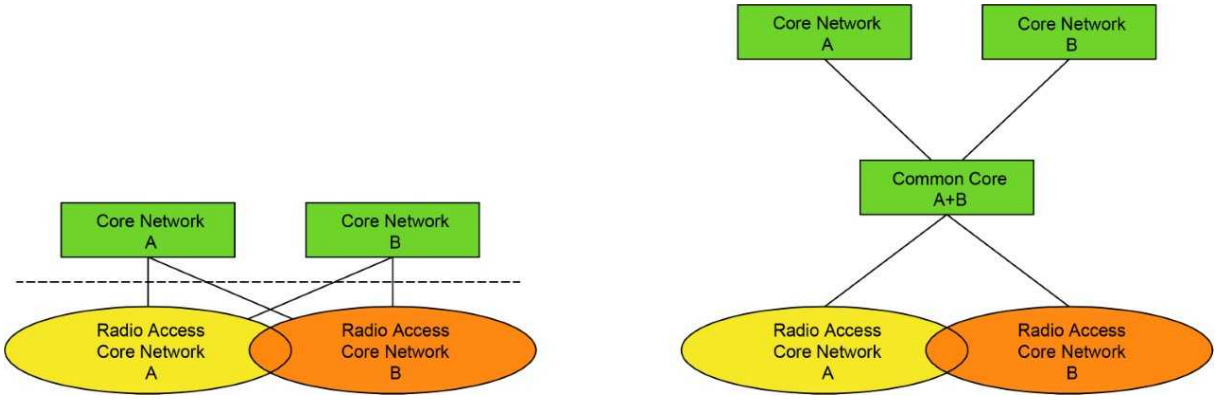


FIGURE 7.2: Geographically split network using dedicated or common core networks

3. *Common network sharing scenario*: In this case, multiple MNOs use the RAN of one MNO in a specific area. Outside the region, each MNO provides its own coverage.
4. *Common spectrum network sharing scenario*: In this case, MNOs jointly pool their allocated spectrum and share them.
5. *Multiple radio access networks sharing common core network scenario*: In this case, multiple RANs share a common network. The multiple RANs can belong to different MNOs.

In our work, we consider the geographically split networks, along with the common spectrum network sharing scenarios. In particular, we consider a realistic case positioning of antenna elements [23], where RRHs are split between two different MNOs, having each its own licensed spectrum. As a result, one UE might associate to its geographically closest RRH, that may belong to a MNO different from its hosting one. Further, cooperating MNOs jointly pool their BBUs located in a single data center, along with their spectral resources, so as to increase capacity and multiplexing gains. Therefore, we propose two different algorithms for MNOs cooperation in C-RAN (*i.e*, the MAB

algorithm and CA algorithm). Both algorithms aim to optimize the BBU-RRH association schemes. Additionally, they both manage the common spectrum sharing part by two different schemes:

— The first scheme is inspired from the reinforcement learning solutions targeted to resolve the MAB problem. In this scenario, MNOs access each others' bands so as to mitigate inter-cluster interferences. In particular, the goal would be to steer each BBU decision to choose autonomously the most appropriate operating band (*i.e*, the band of one MNO or another) with minimum interference level. As the selection of the least interfered band by BBUs is modeled by the adversarial MAB problem, we resort to the popular EXP3 algorithm [85], [86] to solve the problem.

— The second scheme relies on the CA technique, which is a feature in LTE-Advanced technology that aggregates the different spectrum so as to increase capacity.

The rest of the chapter is organized as follows: Section 7.2 relates few relevant works of network sharing. Section 7.3 describes the proposed algorithms for MNOs cooperation in C-RAN architecture (*i.e*, the MAB algorithm and CA algorithm). Further, in this section we detail the different types of the CA techniques, and its deployment scenarios with its RRM challenges. Section 7.4 describes the considered scenarios to assess the results, and provides a comparative study between the case, where two MNOs, adopting the C-RAN architecture, do not share their resources, and the proposed algorithms where MNOs cooperate and share their resources. Section 7.5 concludes the chapter.

7.2 Related Work

Many works have treated the operational aspects of network infrastructure sharing, such as scheduling techniques to efficiently exploit the shared pool of resources. In this context, the authors in [87] propose a scheduling algorithm, allowing the dynamic spectrum sharing between multiple MNOs. They propose to apply a Generalized Processor Sharing (GPS) strategy, where MNOs agree previously on their respective resource share, with a Bin Packing heuristic to share their spectral resources. In [88], the GPS strategy is further investigated. In their work, the authors demonstrate that respecting the sharing guarantees is not the ultimate way to maximize spectral efficiency. Thus, they aim to find a good compromise between satisfying MNOs predefined resource shares and spectral efficiency. In [89], the authors propose to abstract all the base stations of multiple MNOs, as a virtual big-base station controlled by a logically centralized control plane, making all decisions of HOs and interference management. An energy efficient scheme was introduced in [90]. In particular, the authors propose to make use of the coexistence of multiple MNOs in the same area to switch OFF base stations belonging to a particular MNO, and to roam its traffic to base stations of different MNOs. In the same line, the authors in [91] derive a closed form power consumption model as a function of inter-base stations distance. Their purpose is to decide whether a MNO should roam its UEs to a different MNO, whose base station is located at a specific distance from its own base station. In [92], the authors consider a single buyer MNO, sharing the network infrastructure of multiple sellers MNOs. They analyze the power consumption of the network and propose a scheme for optimizing the base station intensities of seller MNOs, in order to minimize the power consumption while achieving a minimum level of QoS for the buyer MNO. The authors in [93] exploit the network infrastructure sharing in a multi-tenant slicing context. The cooperation among multiple infrastructure providers (*i.e*, MNOs) is applied through two stages: A first stage is used to manage and orchestrate use cases and service models into network slices. A second stage is employed so as to provide optimal cooperation among multiple infrastructure providers according to the best match of each network slice to its required delay and QoS parameter. The authors in [94] propose a virtualization scheme for a cellular network sharing. In particular, they propose a two-level scheduler for multiple MNOs that share the same spectrum, achieving virtualization of radio resources.

Different from the previous propositions, few works propose techno-economical models for infrastructure sharing. For example, in [95] a techno-economical model is proposed, allowing UEs to select their targeted quality from different MNOs at one hand, and to maximize their profits, on the other hand. In their proposition, the authors translate the UEs' willingness to pay and their quality targets into time-varying resource requirements, which are used to determine the amount of network resources to be allocated by multiple MNOs. In the same line in [96], the authors propose a techno-economical model for dynamic short-term resource sharing. In particular, their proposition is built on short terms decisions, rather than long terms Service Level Agreements (SLAs) that do not cope with short term fluctuations in traffic.

Strategic cooperation among MNOs is investigated in [97] and [49]. In particular, authors aim to exploit the economical profits of coalitions among multiple cooperating MNOs. Based on fixed market shares and pre-allocated spectrum, authors aim to determine scenarios where MNOs have incentives to cooperate with each other. The work in [49] is the only one to exploit such cooperation within the C-RAN architecture. However, this work only focuses

on the strategic cooperation of MNOs in terms of monetary costs, and do not investigate the additional benefits of power savings, and enhanced spectral and energy efficiency brought by common spectrum and active infrastructure sharing in C-RAN. Further, this work proposes the CA technique for common spectrum sharing. However, CA has many RRM drawbacks and cannot be applied for all UEs types.

In fact, allowing the cooperation among MNOs in C-RAN brings along many benefits: MNOs do not only share their antenna elements (*i.e*, RRHs) or spectral resources, they also share their BBUs which are pooled together, providing additional multiplexing gains and virtualization of resources, leading to efficiency in radio resource utilization. As an example, the increased spectral efficiency derived from common spectrum sharing enhances the BBU capacity. Further, pooling of BBUs resources allows additional associations of RRHs to one single BBU. Thus, reduces the number of active BBUs. In fact, in previous works, the BBU-RRH association optimization has been treated in the case of one single MNO, and has not been investigated in the context of cooperating ones.
To the best of our knowledge, we are the first to exploit the advantages of common spectrum and active infrastructure sharing in the C-RAN architecture. Particularly, we are interested in evaluating the derived benefits in terms of additional power savings, and enhancement of spectral and energy efficiency. The contributions of the chapter are summarized in the following points:

— We propose to exploit the derived benefits of MNOs cooperation in terms of additional power savings, and enhanced spectral and energy efficiency in C-RAN architecture.
— Two different algorithms are proposed for such cooperation, where common spectrum sharing relies on two different schemes: A first, based on reinforcement learning solutions targeted to solve the MAB problem, where MNOs can access each others' bands to mitigate interference. The second is based on the CA technique which is a feature in LTE-Advanced technology that aggregates the spectrum so as to increase capacity.
— A realistic positioning of antenna elements is considered [23], where RRHs are split between two different MNOs, and results are assessed through many considered scenarios in terms of load conditions.

7.3 Cooperation Algorithms in C-RAN

In this section, we describe MNOs cooperating algorithms in C-RAN. We start by describing the MAB algorithm (cf. subsection 7.3.2). We then describe the CA algorithm (cf. subsection 7.3.4).

7.3.1 System Model

We use the reference model of section 3.1 presented in chapter 3. In this chapter, attributing the appropriate band to a formed cluster is treated in the MAB algorithm. In particular, each BBU has to choose a most adequate operating band to lower inter-cluster interferences. Thus, long term radio conditions are taken into account, so as to select the appropriate operating band. The network area in this case, remains decomposed into several discrete zones, where average number of UEs and radio conditions are taken into account.

7.3.2 MAB Algorithm

The different stages of both algorithms are similar to each other. We describe hereafter the different stages of the MAB algorithm.

— *Stage 1: UE Association* The UE association scheme consists in associating a given zone to its geographically closest RRH (cf. paragraph 3.1.1, chapter 3). When MNOs put in common their antenna elements, a UE subscribed to a given MNO might be located into a zone closer to a RRH belonging to a different MNO. Thus, within that stage, zones are associated to their geographically closest RRHs, belonging to all cooperating MNOs for a better received UEs signal level.
— *Stage 2: BBU-RRH Association* After applying the UE association scheme, a BBU-RRH association algorithm is executed. In fact, we apply the greedy heuristic described in chapter 3 to form clusters (cf. section 3.3). Within that stage, the clustering algorithm is applied over the set of RRHs operating on the same spectrum. In that case, the clustering algorithm is executed over each set of RRHs belonging to a given MNO.
— *Stage 3: Common Spectrum Sharing Scheme* This stage consists in managing the common pool of licensed spectrum of all cooperating MNOs. In that particular case, the EXP3 algorithm is applied so as to reduce the interference impact between the formed clusters of RRHs derived in the previous stage. Precisely, the

spectrum allocation among the different clusters changes in a way to reduce inter-cluster interferences. Therefore, the clusters (*i.e*, BBUs) are the players that compete for the MNOs spectrum. This stage is detailed in subsection 7.3.3.

— *Stage 4: BBU-RRH Association Re-execution* After applying the EXP3 algorithm, inter-cluster interferences are reduced, leading to enhanced BBU spectral efficiency. As a consequence, a BBU might support more UEs. In that stage, the clustering algorithm is re-executed over the set of RRHs operating on the same spectrum to realize more power savings.

7.3.3 Choosing the Operating Band

A cluster c needs to choose one band b among B available bands, where B is the number of licensed bands belonging to the different cooperating MNOs. We consider a vector space $\{0,1\}^B$ and number the available bands from 1 to B. The strategy space for any cluster c is the set $S_c \in \{0,1\}^B$ of size B, since each cluster chooses only one operating band b among B.
At a given iteration t, band b provides a reward $r_{c,b}(t)$ for cluster c, which is in interval $(0,1)$.

Reward of a Cluster

As previously mentioned, our goal is to steer the decision of each cluster to choose autonomously the most appropriate operating band with minimum interference level. For that purpose, we define the reward function as the sum of $SINR$ of all UEs served by the cluster. It is expressed as follows:

$$r_{c,b} = \sum_{r \in \mathcal{R}} \sum_{u \in \mathcal{U}} z_{u,r} \cdot y_{r,c} \cdot SINR_{u,r,c,b}, \tag{7.1}$$

where $z_{u,r}$ expresses the association between UE u belonging to the set $\mathcal{U}$ of all UEs, and RRH r. Variable $y_{r,c}$ expresses the association between RRH r and cluster c, as previously defined. $SINR_{u,r,c,b}$ expresses the $SINR$ ratio of UE u, connected to RRH r, which is associated to cluster c, operating on band b. $SINR_{u,r,c,b}$ is expressed as follows:

$$SINR_{u,r,c,b} = \frac{P_r \cdot G_{u,r}}{N_0 + \sum_{c' \neq c} \sum_{r' \neq r} y_{r',c'} \cdot s_{c',b} \cdot P_{r'} \cdot G_{u,r'}}, \tag{7.2}$$

where the binary variable $s_{c',b}$ is defined as follows:

$$\begin{cases} s_{c',b} = 1 & \text{If cluster } c' \text{ operates on band } b \\ s_{c',b} = 0 & \text{Otherwise} \end{cases} \tag{7.3}$$

We recall that intra-cluster interference between RRHs belonging to the same cluster is canceled, since the latters act as a single cell and behave in a DAS manner.

The EXP3 Algorithm

The cumulated reward of any cluster (*i.e*, BBU) c, given the horizon T, is expressed as:

$$G^c = \sum_{t=1}^{T} r_{c,b}(t), \tag{7.4}$$

where the strategy b is chosen randomly according to some distribution over S_c governed by EXP3 algorithm and described in algorithm 6. We want to compare this cumulated reward with the largest reward given by a fixed choice:

$$g = \max_b \sum_{t=1}^{T} r_{c,b}(t), \tag{7.5}$$

The purpose of any cluster c, given the horizon T, is to get a minimum regret at the end. The regret R_T is defined as the difference between the best fixed choice and the expected cumulated reward of any cluster c:

$$R_T^c = g - \mathbb{E}[G^c], \tag{7.6}$$

where the averaging is taken w.r.t. the probabilistic choices of cluster c.

Input Let k be a real in $(0, 1]$;
Initialization Set initial strategy weights $w_b(0) = 1$ for every strategy $b \in S_c$;
for $t = 1$ *to* T **do**
    Cluster c selects strategy s at random according to probability $p_b(t) = (1-k)\frac{w_s(t)}{\sum_{b=1}^{B} w_b(t)} + \frac{k}{B}$;
    Cluster c updates the weights as follows:

$$w_s(t+1) = \begin{cases} w_s(t)\text{exp}(\frac{kr_b(t)}{Bp_b(t)}) & \text{if } s = b \\ w_s(t) & \text{Otherwise} \end{cases}$$

end

Algorithm 6: EXP3 applied by any cluster c

Performance Evaluation of EXP3 Algorithm

In this paragraph, we evaluate the performance of the EXP3. In particular, we are interested in evaluating the cumulated reward given by the algorithm, given as in (7.4), in comparison with the case where MNOs do not access each others bands, and each operates on its own licensed band, deemed "Non-common spectrum sharing". Further, we are also interested in evaluating the convergence of the EXP3.

Figure 7.3 shows the cumulated reward given by the EXP3 algorithm in comparison with the non-common spectrum sharing scheme, as a function of the number of iterations. For the case of non-spectrum sharing, a fixed value is shown, since bands are distributed deterministically over the RRHs, which makes the reward function that varies as a function of inter-cluster interferences stable, contrarily to the EXP3 algorithm which has a probabilistic behavior. We also notice that the reward realized by the EXP3 algorithm is initially below that realized by the non-common spectrum sharing. As the number of iterations increases, the EXP3 algorithm converges to values around 0.35, greater than that of the non-spectrum sharing scheme, which is around 0.275.

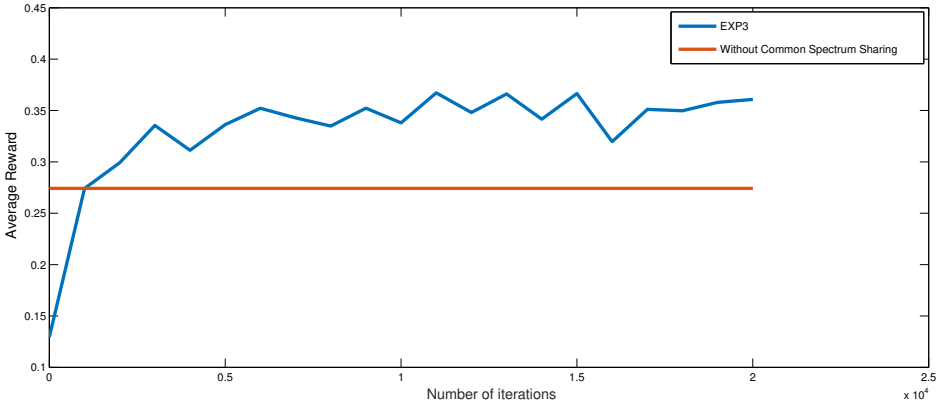


FIGURE 7.3: Cumulated Reward: EXP3 vs. Non-Common Spectrum Sharing

Figure 7.4 displays the probability of a strategy selection as a function of the number of iterations. In particular, we consider 4 players (*i.e*, 4 BBUs), denoted as 'BBU 1', 'BBU 2', 'BBU 3' and 'BBU 4'. We also consider 4 strategies (*i.e*, 4 bands), denoted as 'Band 1', 'Band 2', 'Band 3' and 'Band 4'. In fact, the results are evaluated for enough number of bands, since in realistic case scenarios the number of licensed bands attributed for MNOs are limited, making the number of strategies very low. We notice that each player chooses one band to operate on, which confirms that inter-cluster interferences are well managed by EXP3 algorithm. Further, we notice that convergence is attained after almost 3500 iterations for all players, which is considered a low value, since one iteration represents an average duration of 1.5 ms as shown in Figure 7.5.

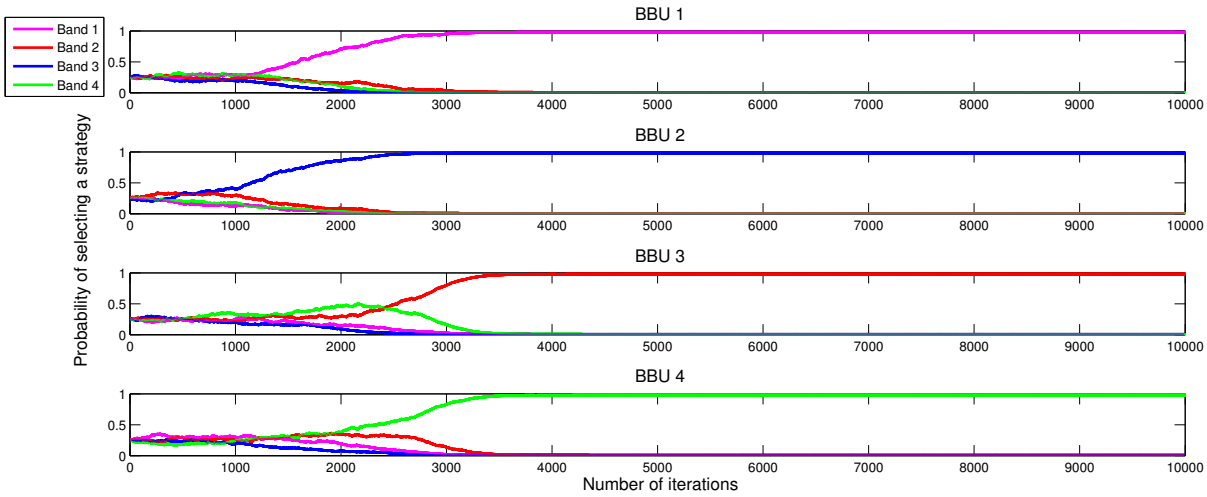


FIGURE 7.4: Probability of Selecting a Strategy

7.3.4 CA Algorithm

As previously mentioned in 7.3.2, the different stages of both algorithms (MAB and CA) are similar to each other. However, for the case of CA, stages 3 and 4 differ. In particular, the common spectrum sharing algorithm (stage 3) is managed via the CA technique. As for the fourth stage, the clustering algorithm is applied over all the RRHs in the system, since they all operate on the same aggregated bands. Contrarily to the MAB algorithm, which only applies the clustering scheme over the set of RRHs operating on the same band.
We review hereafter, the CA technique with its different types, deployment scenarios and RRM.

Carrier Aggregation Technique

Spectrum Scenarios Usually, three different types of spectrum scenarios are employed while applying the CA, and are illustrated in Figure 7.6:

- *The Intra-band contiguous CA*: This type of CA, acts as if it operates on a single band. In particular, the aggregated bands are contiguous to each other. The center frequencies are spaced from each other in a multiple of 300 kHz, in order to be compatible with LTE release 8/9 and to reserve the orthogonality of subcarriers with 15 kHz spacing [98]. This kind of scenario is the simplest to implement.
- *The Intra-band Non-contiguous CA*: This form is more sophisticated than the first case. Non-contiguous carriers are used within the same band. Thus, the multi-carrier signal cannot be treated as a single signal and consequently, many transceivers are required, adding more complexity, particularly for UEs that are constrained in power and cost [99].
- *The Inter-band non-contiguous CA*: This type of CA operates on different bands. It is possible also to use different fragmentation of bands, where some are of 10 MHz width only. One UE needs to use multiple transceivers, increasing the cost, performance and power. Further, there are also additional complexities that result from inter-modulation and cross modulation over the different transceivers. We note that this type of CA potentially contributes in improving the mobility robustness, as the radio propagation characteristics can be exploited over the different bands [100].

Deployment Scenarios Many deployment scenarios are also considered while applying the CA:

- *Deployment scenario 1*: The first case, depicted in Figure 7.7(a), is the most typical envisaged scenario. In this case, antennas have the same beam directions and patterns for each band. Within this scenario, both bands (F1 and F2) offer the same coverage, as shown in Figure 7.7(a).
- *Deployment scenario 2*: In this case, F1 and F2 cells are overladed and co-located in different bands (cf. Figure 7.7(b)). F2 has a smaller coverage than F1, due to its higher pathloss. F2 is used to offer more throughput and only F1 offers wider coverage.

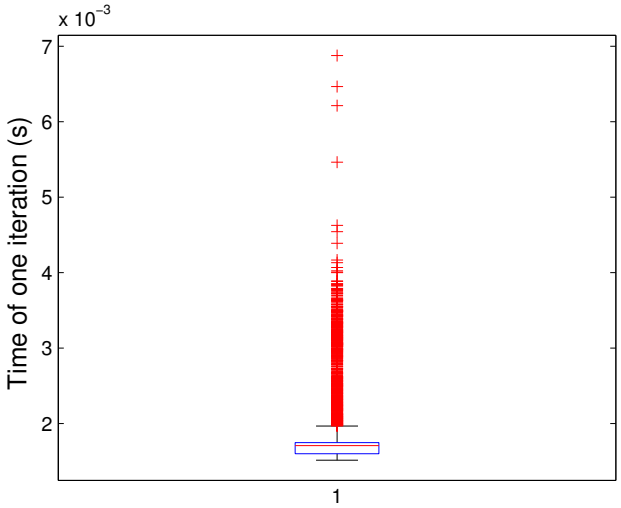


FIGURE 7.5: Time of Convergence of One Iteration

— *Deployment scenario 3*: A similar case to scenario 2, consists that F1 and F2 cells to be co-located. However, antennas for F2 are directed to the F1 cell edges so as to enhance the cell edge UEs of F1 cells, as shown in Figure 7.7(c)
— *Deployment scenario 4*: A fourth scenario consists in offering the macro coverage by F1, whereas RRHs operating on F2 are used to serve hotspots regions (cf. Figure 7.7(d)). RRHs are connected to the macro base station via fiber connections. Such deployment scenario is used to increase throughputs.

In our case, we consider that the carriers provide the same coverage. In other terms, they represent same beam directions and antenna patterns (*i.e*, deployment scenario 1)

RRM Challenges of CA As a UE can be scheduled over multiple carriers, having each different radio characteristics, new RRM challenges are introduced. In particular, the LTE release 10 must insure backward compatibility with LTE UEs:

The base stations configure a set of carriers in different bands for UEs, based on traffic load and QoS requirements. The multiplexing of multiple UEs in each carrier is carried out, and RBs scheduling is performed. Consequently, independent layer 1 transmission, containing HARQ and Link Adaptation per carrier is employed (cf. Figure 7.8) in order to ensure backward compatibility, so that LTE-Advanced UEs and LTE UEs can coexist.

Transmissions over different carriers is optimized by the use of independent link adaption per carrier according to experienced radio conditions. Therefore, resource allocation and adaptive adjustment of transmission parameters for each carrier must be treated jointly, provoking a large amount of signaling overhead on the upload direction, since UEs send their CSI informations and acknowledgment/non-acknowledgment per carrier.

The functionality of assigning multiple carriers for UEs is introduced by the RRM. In particular, in order to select a carrier for one UE, informations, such as terminal capability and QoS are gathered. Further, the measurements of traffic load per carrier and the CQI from UEs are also required to select the appropriate carriers. Additionally, a load-balancing scheme between different carriers is also favorable in order to obtain optimal system performance and to prevent the under-utilization of spectral resources [102], leading to increased complexity on base stations side.

The maximum uplink transmission power of one UE is also a constraint. In fact, for non-contiguous resource allocation, a UE transmits simultaneously over multiple carriers. Therefore, power back-off is needed in UE power amplifier, which increases the peak-to-average power ratio, resulting in a reduction in the maximum UE transmission power [103]. In fact, it is not always beneficial for power limited UEs to allocate multiple carriers, particularly, for cell edge UEs. Therefore, the carrier selection on the uplink level is crucial for the system performance optimization from a RRM perspective, due to the power constraint in uplink transmission.

CA introduces fairness issues in LTE systems. In fact, LTE-advanced UEs are assigned to multiple carriers, which increases their throughputs comparing to LTE UEs, causing the degradation of fairness. Additionally, carrier mobility

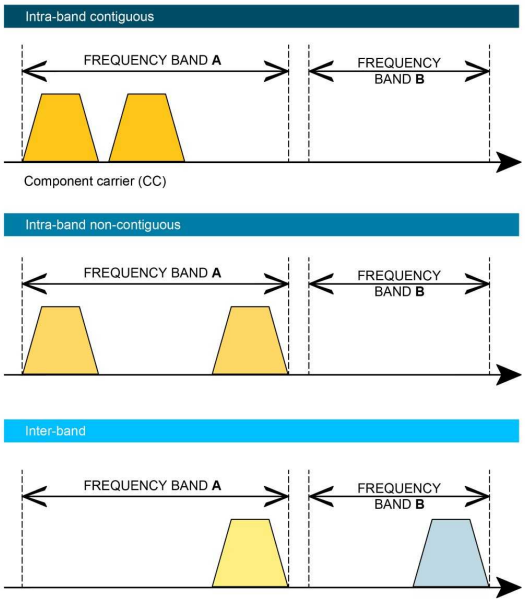


FIGURE 7.6: Different Types of CA

management could also impose a few constraints. For example, intra- and inter-frequency HOs are necessary, resulting in additional signaling load overhead related to the configurations of the RRC layers [104], since carriers change dynamically.

In the following we evaluate the performance of both algorithms in comparison with the case where two different MNOs do not share their resources, however, they both use the C-RAN architecture.

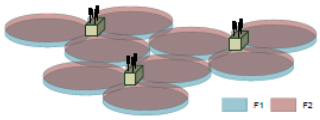
7.4 Performance Evaluation

To evaluate the results, we consider a realistic positioning of base stations [23], where RRHs are equally split between two different MNOs (1 and 2), as shown in Figure 7.9. In particular, each MNO has 10 RRHs, covering the 14th district of Paris.

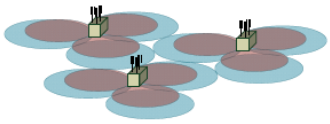
The results are evaluated according to many scenarios in terms of load conditions. In particular, we consider three load scenarios, defined as follows:

- *Scenario A*: In scenario A, we consider that MNO-1 is highly loaded, whereas MNO-2 is lightly loaded. Specifically, the number of UEs per cell in MNO-1 varies between 8 and 12. It is between 4 and 8 in MNO-2.
- *Scenario B*: In scenario B, both MNOs are highly loaded, and the number of UEs per cell varies between 8 and 12.
- *Scenario C*: Scenario C considers that both MNOs are lightly loaded, and the number of UEs per cell varies between 4 and 8.

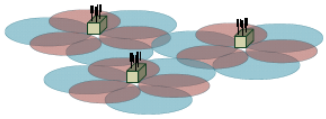
When MNOs are not cooperating, the number of active BBUs and the power consumption are calculated as previously detailed in chapter 3 (cf. section 3.4). However, when MNOs are cooperating, all resources are put together (*i.e*, BBUs, RRHs and spectrum). In that case, we consider that the power consumption of each MNO is proportional to the number of its subscribed UEs. In fact, this assumption has been adopted in [49], as it is considered a logical estimation of MNOs' load contribution in the shared network. For example, if we consider that MNO-1 has n_{MNO1} number of UEs in the network, and MNO-2 has n_{MNO1} UEs, the derived power consumptions of MNOs 1 and 2 would be $P_t \cdot \frac{n_{MNO1}}{n_{MNO1}+n_{MNO2}}$ and $P_t \cdot \frac{n_{MNO2}}{n_{MNO1}+n_{MNO2}}$ respectively, where P_t is the total power consumption in the system. The number of active BBUs is calculated in the same manner as the power consumption. We note that the power consumption model and parameters (cf. subsection 3.1.4, chapter 3) in the CA case remain the same. In fact, introducing the CA technique corresponds to handling a set of unaggregated multi-carriers which is already supported by a LTE base station supporting multiple frequency bands [105] [106].



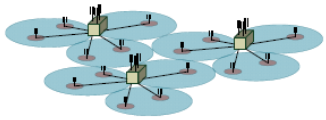
(a) Scenario 1 [101]



(b) Scenario 2 [101]



(c) Scenario 3 [101]



(d) Scenario 4 [101]

FIGURE 7.7: CA Deployment Scenarios

Same metrics, presented in chapter 3, are used to compare the results. We further introduce an additional one, given as follows:

$$R_{unsatisfied} = \frac{n_{unsatisfied}}{\sum_{c \in \mathcal{P}} n_c} \times 100, \tag{7.7}$$

$R_{unsatisfied}$ reflects the total number of unsatisfied UEs in the network. Precisely, this metric reflects the percentage of UEs that are not guaranteed a minimum level of throughput requirement (D_{min}). $n_{unsatisfied}$ is the total number of UEs within each cluster c where $T(c) < D_{min}$. Recall that $T(c)$ is the throughput realized within cluster c, and n_c is the total number of UEs served by cluster c. The simulation parameters are shown in table 7.1.

TABLE 7.1: Simulation Parameters

Parameter	**Value**
Traffic Model	Full Buffer
Scheduling Scheme	Fair Resource Sharing
Propagation Model	Cost Hata 231
Shadowing Standard Deviation	10 dB
Transmit Power of RRH	40 dBm
Thermal Noise Power	-174 dBm/Hz
BW	10 MHz
Cell Radius	450 m
A	50 W
B	0.6
D_{min}	2 Mb/s
Meshing Step (a)	240 m

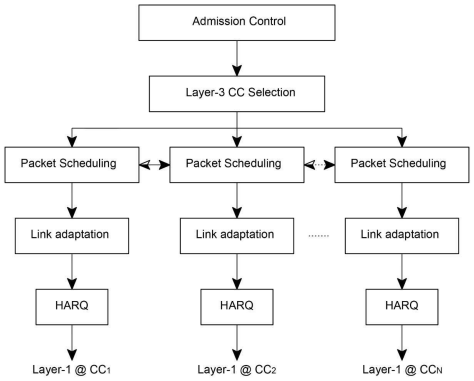


FIGURE 7.8: Structure of RRM for LTE advanced with Carrier Aggregation Support

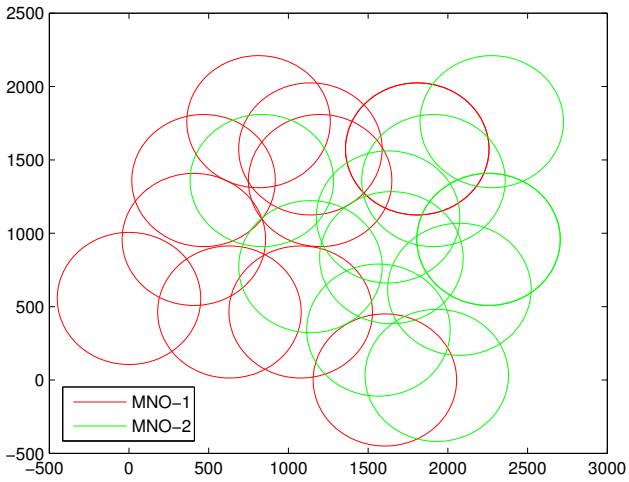


FIGURE 7.9: MNO 1 vs. MNO-2

7.4.1 Scenario A: Highly loaded MNO-Lightly loaded MNO

We start by assessing the different metrics of scenario A where MNO-1 is highly loaded, and MNO-2 is lightly loaded. The results are displayed for the following cases : First the case where MNOs do not share their resources, denoted as “Without cooperation”; the cooperative scenario that employs the MAB algorithm, denoted as “With Cooperation-MAB”; and the cooperative case that uses the CA algorithm, denoted as “With Cooperation-CA”. For the case where MNOs do not share their resources, the clustering algorithm is applied over each set of RRHs that belongs to one MNO. The results are averaged and displayed with 95% confidence interval.

Figure 7.10 shows the number of active BBUs for MNOs 1 and 2. When spectral and network resources are not shared, MNO-1 activates a high number of BBUs (around 9.5). After cooperation, resources of MNO-1 and MNO-2 are jointly pooled: A new UEs association scheme is applied for UEs and more spectrum are available, resulting in higher energy and spectral efficiencies. Thus, the number of active BBUs decreases. The MAB algorithm shows around 6 active BBUs, whereas the CA shows around 4.5 active BBUs. In fact, the spectrum handled by one BBU increases while applying the CA, which consequently increases the BBU capacity. Thus, fewer BBUs are activated in that case. However, the CA technique is not supported by LTE release 8 UEs. Further, it requires more complexity in terms of RRM, as previously stated. For the case of MAB, an interference management technique is applied among clusters, which also decreases the number of active BBUs. We note that for the MAB algorithm, the BBU capacity remains the same, as no aggregation for spectrum is applied, and a BBU is operating over one band. In the case of MNO-2, we notice that the number of active BBUs slightly increases when cooperative schemes are applied in comparison with the case where resources are not shared. In fact, UEs of MNO-1 use the resources of MNO-2,

increasing the number of active BBUs, and also the interference level. However, the total number of BBUs within the network globally decreases. We also notice that the MAB algorithm shows slightly higher number of active BBUs in comparison with the CA algorithm requiring more complexity for band aggregation.

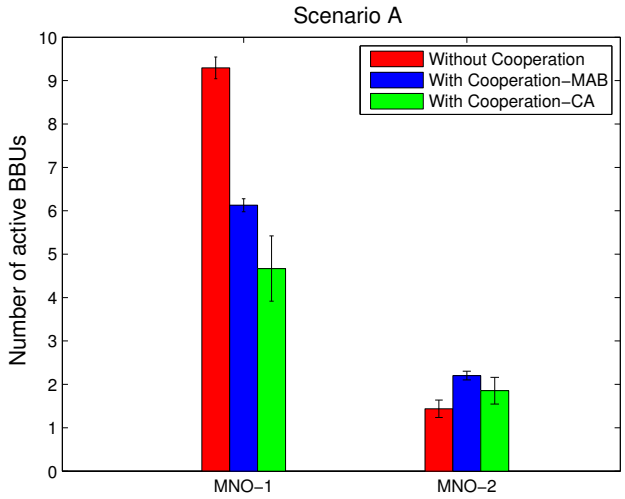


FIGURE 7.10: Scenario A: Number of Active BBUs

Figure 7.11 displays the power consumption of MNOs 1 and 2. Without cooperation, MNO-1 shows the highest power consumption as it is highly loaded. When cooperation algorithms are applied, the power consumption decreases. The power consumption of CA algorithm is further reduced in comparison with the MAB algorithm, as the CA increases the capacity of the BBU, decreasing the needed number of active BBUs. The power consumption of MNO-2 slightly increases for the cooperation algorithms in comparison with the case where resources are not shared, as MNO-1 consumes the resources of MNO-2. However, the power consumption in the system is globally decreased.

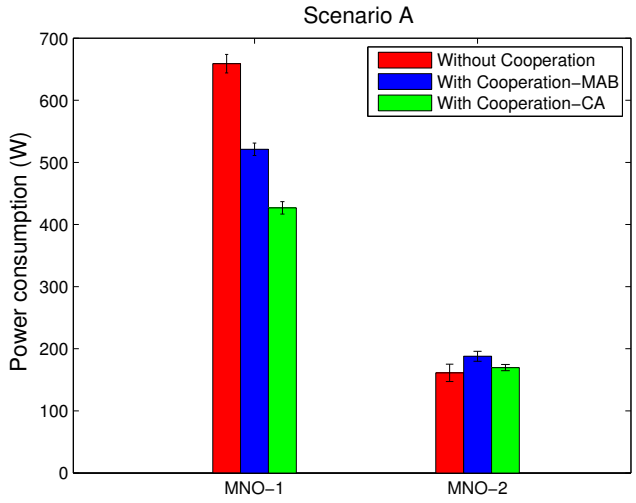


FIGURE 7.11: Scenario A: Power Consumption

Figure 7.12 shows the power saving realized by MNOs 1 and 2. For MNO-1, the power saving is significantly improved when cooperation schemes are applied, as less BBUs are activated. The highest power saving value remains for the CA algorithm, as the capacity of BBUs increases. For the case of MNO-2, the power saving values of the cooperation algorithms slightly decrease in comparison with the case where MNOs do not share their bands and infrastructures, as more resources are used, and more BBUs are activated to serve UEs of MNO-1.

In Figure 7.13, the energy efficiency is displayed for MNOs 1 and 2. In the case of MNO-1, the energy efficiency increases when MNOs cooperation, as less BBUs are activated and less interference is experienced. We notice that the energy efficiency of the CA technique is slightly increased in comparison with the MAB algorithm. In fact, for CA, less BBUs are activated, leading to higher energy efficiency. For MNO-2, we notice that the energy efficiency decreases for cooperation algorithms, comparing to the case where the resources are not shared, because more resources are consumed, leading to more power consumption. The MAB algorithm shows slightly lower value of energy efficiency than the CA algorithm as it consumes more power.

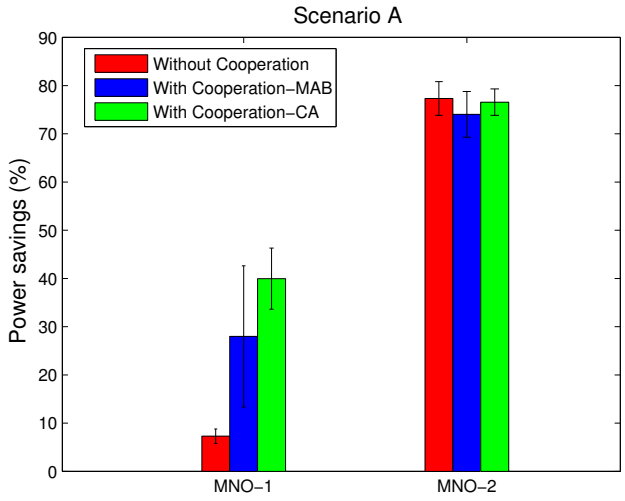


FIGURE 7.12: Scenario A: Power Savings

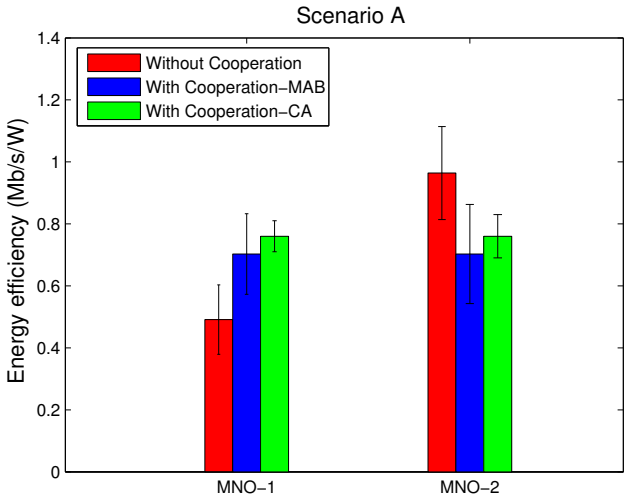


FIGURE 7.13: Scenario A: Energy Efficiency

Figure 7.14 displays the rate of unsatisfied UEs for MNOs 1 and 2. We recall that this metric reflects the number of UEs within each cluster that does not respect a minimum level of QoS requirement per UE. Without cooperation, MNO-1 shows a value of 25%. In fact, for scenario A, MNO-1 is highly loaded, which increases the rate of unsatisfied UEs. However, when cooperation is applied among MNOs, the rate of unsatisfied UEs significantly decreases for the MAB case, and is 0 for the CA case. For MNO-2, the MAB algorithm slightly increases the rate of unsatisfied UEs in order to cope with the demand of MNO-1 UEs that consume the resources of MNO-2, increasing both power consumption and interference level, as MNO-2 consumes the resources of MNO-1. However, the rate of UEs that are not guaranteed a minimum level of throughput requirement, globally decreases in the network.

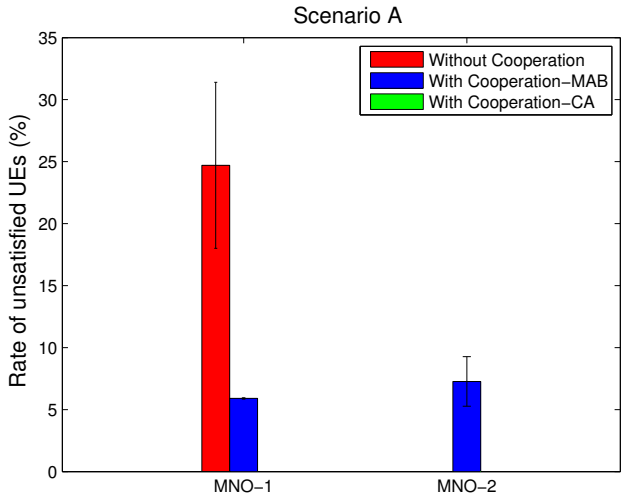


FIGURE 7.14: Scenario A: Rate of Unsatisfied Users

Figure 7.15 shows the mean throughput per UE for MNOs 1 and 2. For MNO-1, we notice the highest value for the MAB algorithm, though it displays less BBUs than the case where resources are not shared. In fact, this is due to the higher level of spectral efficiency (cf. Figure 7.16), enhanced in the case of MAB. Although the spectral efficiency of the CA algorithm is higher than that of MAB, the number of activated BBUs is lower, decreasing the mean throughput per UE in comparison with the MAB. For MNO-2, the mean throughput per UE is lower for cooperation algorithms. In fact, the spectral efficiency values shown by the MAB and the CA schemes are lower than that of without cooperation, although more power consumption is derived. In fact, the increased amount of power consumption in MNO-2 is invested to serve UEs of MNO-1.

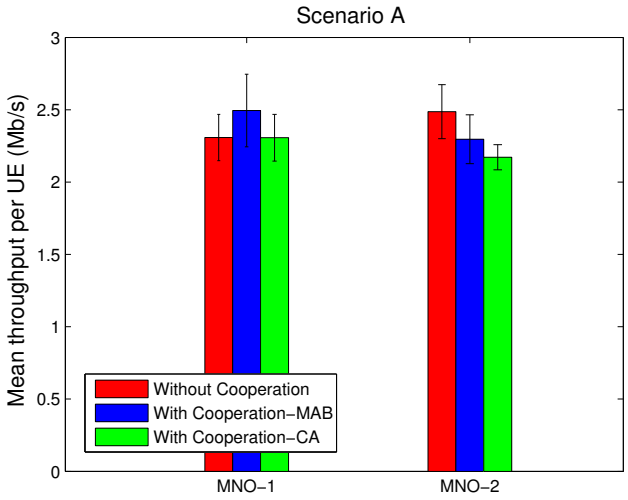


FIGURE 7.15: Scenario A: Mean Throughput per UE

Figure 7.16 displays the results of the spectral efficiency per BBU for MNOs 1 and 2. As the case of the energy efficiency, the spectral efficiency is enhanced for the cooperation algorithms in MNO-1. As for MNO-2, the spectral efficiency is decreased in presence of cooperation algorithms, since MNO-1 UEs consume the resources of MNO-2, increasing jointly the power consumption and the interference level in the network. The CA algorithm always shows higher spectral efficiency than the MAB algorithm. However, it is at the price of more complex RRM at BBUs side.

In Figure 7.17, the total derived throughput in the system is displayed for MNOs 1 and 2. For the case of MNO-1, the total throughput shows its highest value for the case of MAB algorithm, as it provides higher spectral efficiency

than the case where resources are not shared, at on hand, and it activates more BBUs than the CA algorithm, on the other hand. For MNO-2 case, the total derived throughput by the cooperation schemes is lower than that of without cooperation, as more resources are invested to cope with the high demand on traffic of MNO-1.

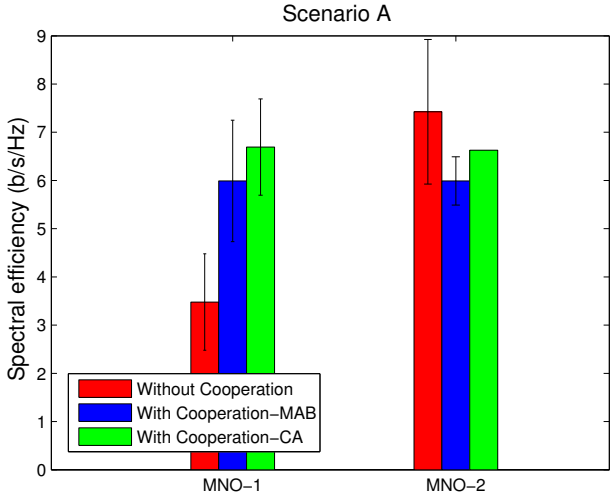


FIGURE 7.16: Scenario A: Spectral Efficiency

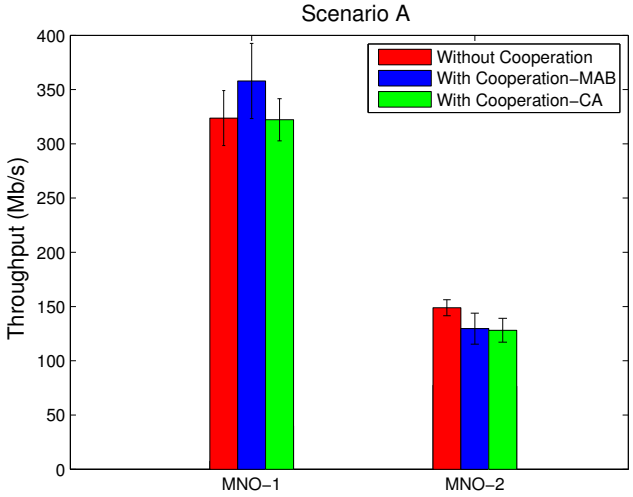


FIGURE 7.17: Scenario A: Total Throughput

7.4.2 Scenario B: Highly loaded MNOs

In scenario B, both MNOs are highly loaded (*i.e*, the number of UEs per cell varies between 8 and 12). Figure 7.18 shows the number of active BBUs for MNOs 1 and 2. We notice that both MNOs display a high number of activated BBUs when resources are not shared. After applying the cooperation algorithms, the interference level decreases in the MAB case, and the capacity increases in the CA case, leading to a significant decrease in the number of active BBUs, even when both MNOs are highly loaded. The CA slightly decreases the number of active BBUs in comparison with the MAB case, as the latter aggregates the bands at the BBU side, and augments the capacity.

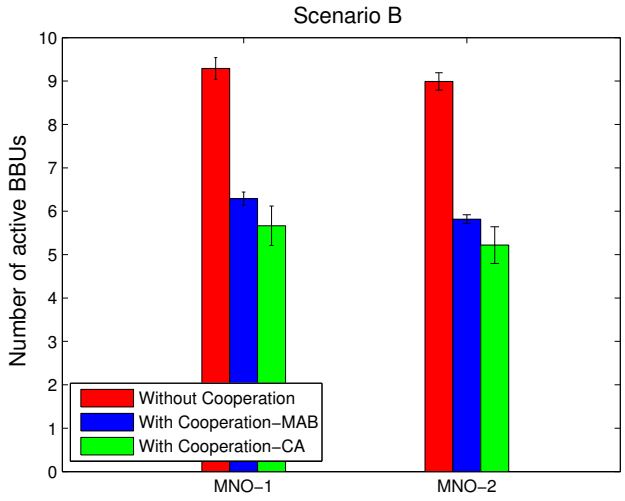


FIGURE 7.18: Scenario B: Number of Active BBUs

Figure 7.19 shows the total amount of power consumption for MNOs 1 and 2. Same as the number of active BBUs, the power consumption in both MNOs decreases when cooperation algorithms are applied. The lowest power consumption value remains for the CA algorithm, as less BBUs are activated due to the increased capacity of a BBU.

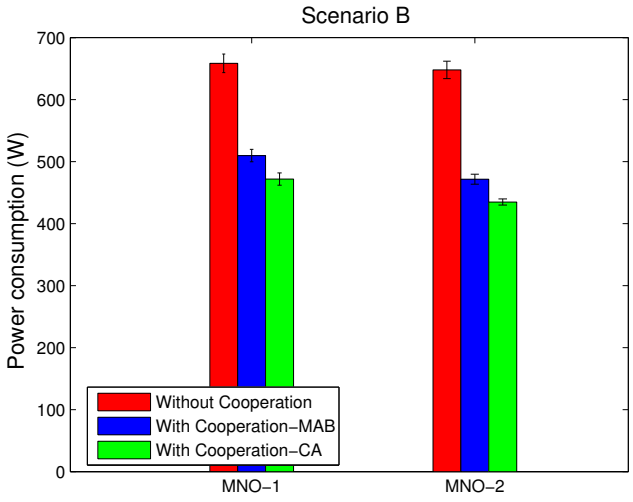


FIGURE 7.19: Scenario B: Power Consumption

In Figure 7.20, the power saving metric is displayed for MNO-1 and MNO-2. We notice that significant power saving is derived when cooperation algorithms are applied. The highest power economy remains for the CA case, due to the previously explained reason. Further, we notice that the power saving for MNO-2 is slightly higher than that of MNO-1, for all schemes, although the same load conditions are considered. In fact, this is due to the network topology which is not the same for both MNOs, affecting the interference level, and thus the different metrics.

In Figure 7.21, the energy efficiency is displayed. For both MNOs, the energy efficiency is increased when cooperation is applied. The highest energy efficiency is noticed for the CA algorithm, as it lowers the most the power consumption. Also, the MAB algorithm reduces the level of interference, increasing both energy and spectral efficiencies.

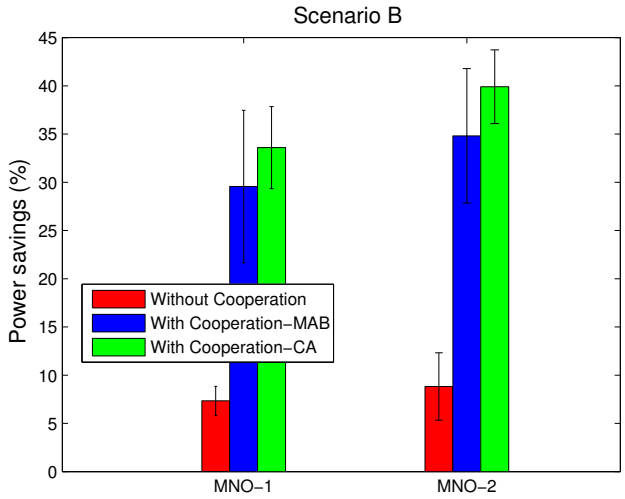


FIGURE 7.20: Scenario B: Power Savings

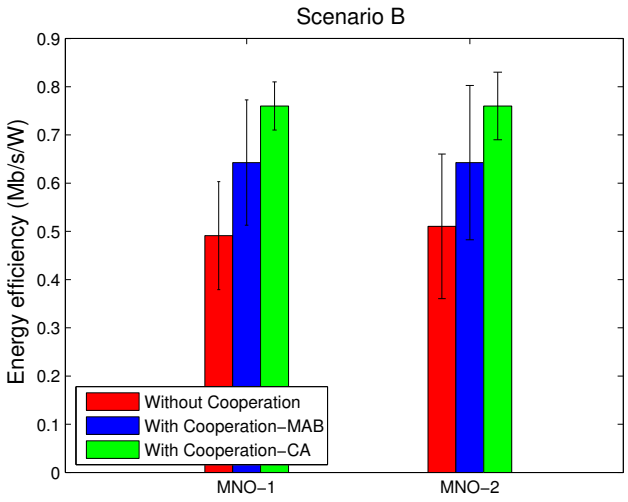


FIGURE 7.21: Scenario B: Energy Efficiency

Figure 7.22 displays the rate of unsatisfied UEs. For both MNOs, the rate is reduced when cooperation algorithms are applied. The MAB algorithm shows higher values than that of the CA. The reduction in the rate of unsatisfied UEs is due to the interference management performed by the MAB algorithm, and to the band aggregation in the case of CA algorithm. We also notice that the cooperation has more impact on MNO-1 than on MNO-2, as the rate of unsatisfied UEs is more reduced for MNO-1. This is due to the network topology, which is not the same for MNOs 1 and 2, and to the fact that MNO-2 derives more power savings than MNO-1, and thus limits the resources when cooperation algorithms are applied (cf. Figure 7.20).

The mean throughput per UE is displayed in Figure 7.23 for MNOs 1 and 2. For MNO-1, the mean throughput per UE is similar in all schemes. This is due to the increased level of spectral efficiency (cf. Figure 7.24) when cooperation schemes are applied, although less BBUs are activated. For MNO-2, the highest mean throughput per UE value remains for the case where resources are not shared, because more BBUs are activated. However, for both cooperation algorithms, the mean throughput per UE slightly decreases, despite the high amount of power saving. This is due to the increased level of spectral efficiency.

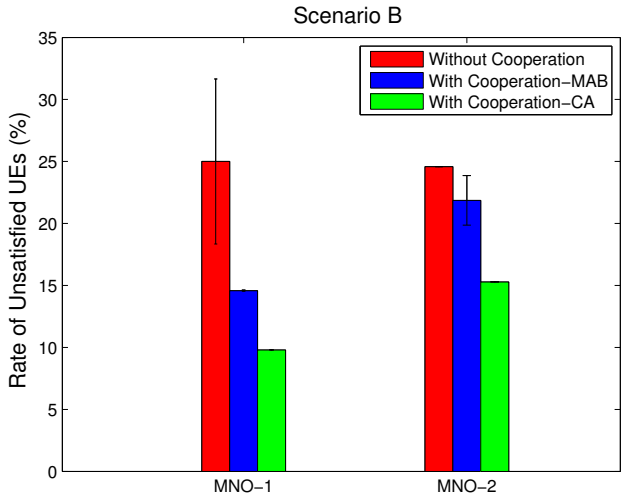


FIGURE 7.22: Scenario B: Rate of Unsatisfied UEs

In Figure 7.24, the spectral efficiency per BBU is displayed. We notice that the spectral efficiency is significantly enhanced for both MNOs when cooperation algorithms are applied (same as the case of energy efficiency). The highest values remain for the CA algorithm, as it deactivates more BBUs and lowers the interference.

Figure 7.25 displays the total derived throughput for all MNOs. Same as the mean throughput per UE, MNO-1 shows similar values for all schemes, whereas the metric slightly decreases for MNO-2 when cooperation is applied.

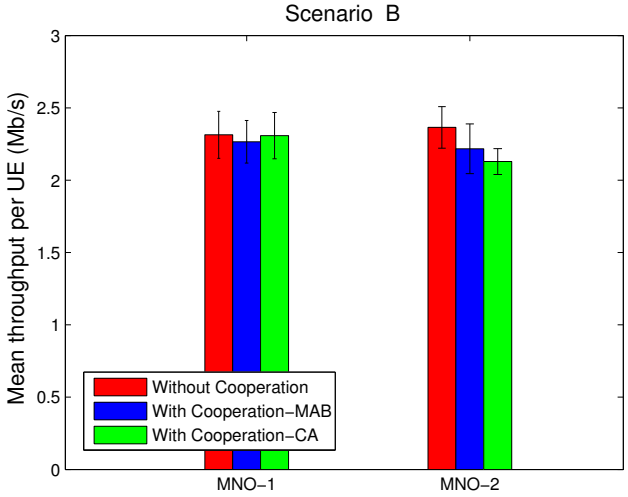


FIGURE 7.23: Scenario B: Mean Throughput per UE

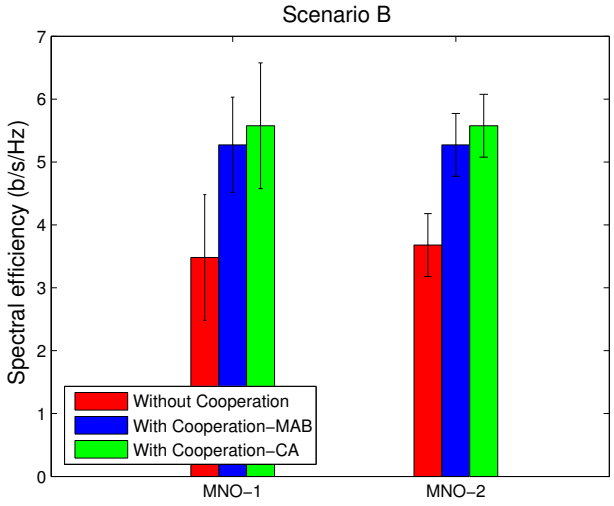


FIGURE 7.24: Scenario B: Spectral Efficiency

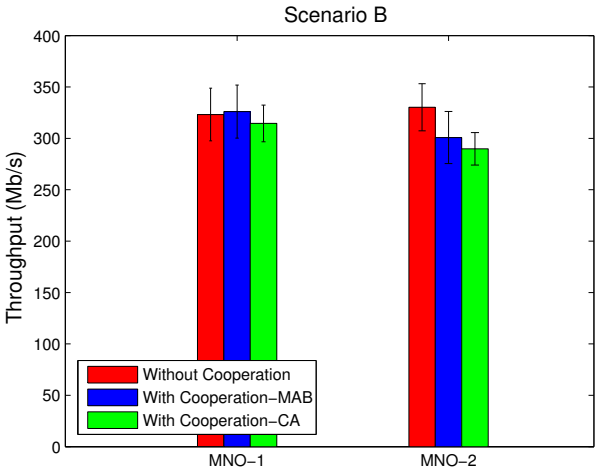


FIGURE 7.25: Scenario B: Total Throughput

7.4.3 Scenario C: Lightly loaded MNOs

In the case of scenario C, MNOs are lightly loaded (*i.e*, the number of UEs per cell varies between 4 and 8). Figure 7.26 displays the number of active BBUs for MNOs 1 and 2. Without cooperation, MNOs 1 and 2 show around 1.5 active BBUs. When cooperation algorithms are applied, the MAB shows around 1.2 active BBUs, whereas the CA shows around 0.9. Although low load conditions are considered, the number of active BBUs slightly decreases when cooperation algorithms are applied.

Figure 7.27 shows the power consumption metric. For MNO-1, we notice that the MAB algorithm derives almost same power consumption as the case where resources are not shared, although less BBUs are activated. In fact, the MAB algorithm enhances the spectral efficiency in comparison with the case where no resource sharing is considered (cf. Figure 7.32), which increases the total throughput in the system. As the power consumption is a linear function of the throughput, the MAB consumes more power. The CA algorithm decreases the power consumption, although the spectral efficiency is increased in comparison with the case where resources are not shared. However, the CA shows lower number of active BBUs in comparison with the MAB case, which also reduces the throughput and the power consumption. For the case of MNO-2, we do not notice the same values of that shown by MNO-1, though the same load conditions are considered. The reason is that the network topology is not the same.

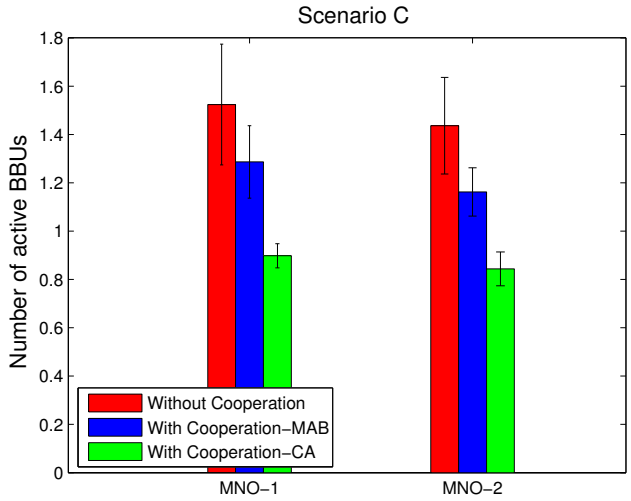


FIGURE 7.26: Scenario C: Number of Active BBUs

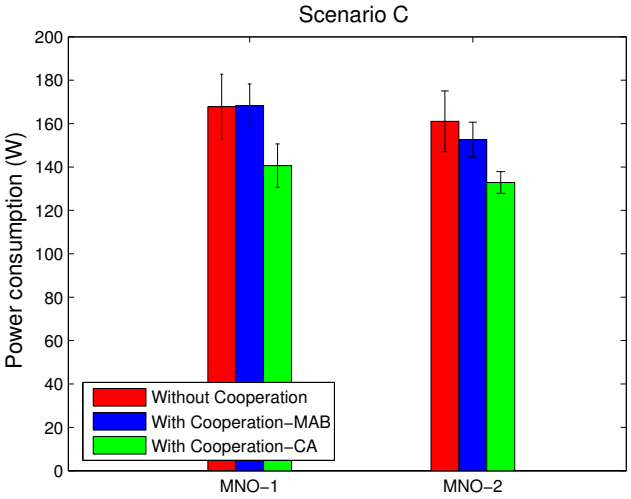


FIGURE 7.27: Scenario C: Power Consumption

The power saving metric is displayed in Figure 7.28. Similar to the power consumption metric, MNO-1 shows small discrepancy in power saving for the case where resources are not share, and the MAB case. The CA shows higher value than both cases of "Without-Cooperation" and the MAB algorithm, as less power consumption is derived. MNO-2 shows the highest value of power saving for the CA case, and the lowest for the case where resources are not shared.

Figure 7.29 displays the energy efficiency of MNOs 1 and 2. We notice that the energy efficiency is enhanced for both MNOs after applying the cooperation schemes. The highest values are for the CA algorithm, as it reduces the most the power consumption for both MNOs 1 and 2. Although it realizes more energy efficiency, the CA remains the most complicated scheme in terms of RRM, and is not supported by LTE release 8 UEs. The MAB algorithm also realizes high values of energy efficiency in comparison with the case where resource are not shared, as it performs interference management among clusters.

Figure 7.30 shows the rate of unsatisfied UEs for MNOs 1 and 2. We notice that the MAB algorithm increases the rate of unsatisfied UEs for both MNOs. However, it realizes more power savings than in the case where resources are not shared. We also notice that the values displayed in the MAB case, for both MNOs 1 and 2 are not the same, as the MAB realizes more power savings in MNO-2, limiting the resources. Further, network topologies for both MNOs are not the same.

In Figure 7.31, the mean throughput per UE is displayed for MNOs 1 and 2. For MNO-1, we notice the highest value for the MAB algorithm. As previously mentioned, the MAB enhances the spectral efficiency in comparison with the case where resources are not shared. Further, it activates more BBUs than the CA algorithm. The same behavior is noticed for MNO-2, as the mean throughput per UE shows the highest value for the case of MAB. The values shown by MNO-2 are not the same as MNO-1, since the network topologies are not the same. Although the mean throughput per UE is enhanced in the case of MAB, the rate of unsatisfied UEs is increased as shown in Figure 7.31. In fact, enhancing the throughput does not necessarily implicate a reduction in the number of unsatisfied UEs.

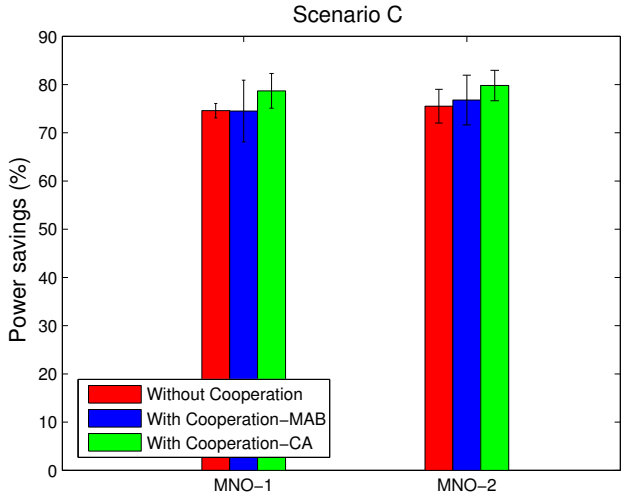


FIGURE 7.28: Scenario C: Power Savings

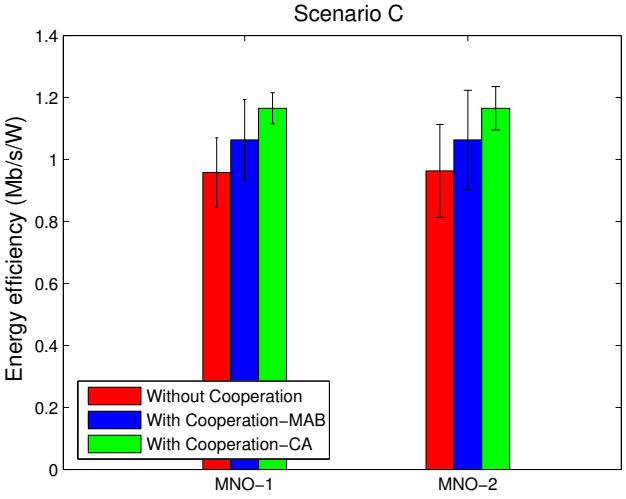


FIGURE 7.29: Scenario C: Energy Efficiency

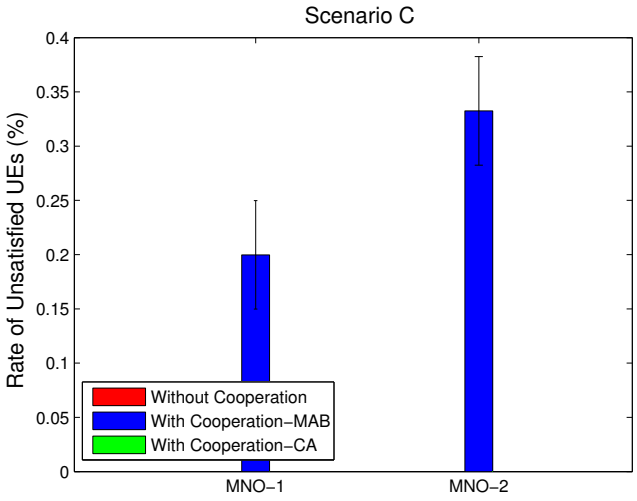


FIGURE 7.30: Scenario C: Rate of Unsatisfied UEs

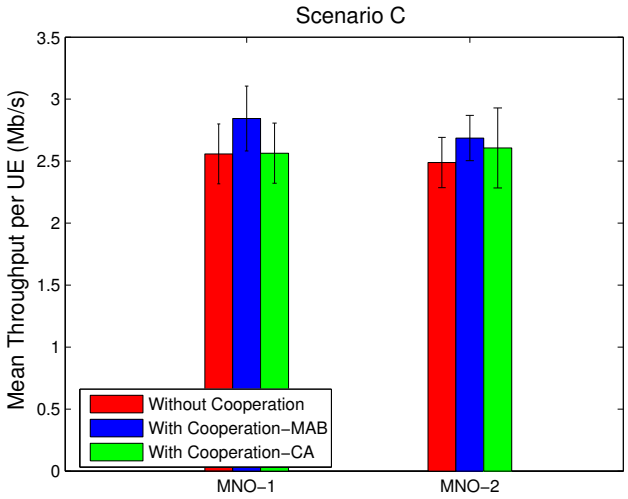


FIGURE 7.31: Scenario C: Mean Throughput per UE

Figure 7.32 displays the spectral efficiency per BBU for MNOs 1 and 2. Similar to the energy efficiency, both cooperative schemes enhance the spectral efficiency of MNOs 1 and 2. The highest values for the CA case, as it activates less BBUs than other cases.

Figure 7.33 shows the total derived throughput for MNOs 1 and 2. The total throughput shows its highest values for the MAB algorithm. As previously explained, the MAB enhances the spectral efficiency in comparison with the case where resources are not shared, and activates more BBUs than CA, which increases the throughput. The values of MNOs 1 and 2 are not the same, as network topologies are not the same.

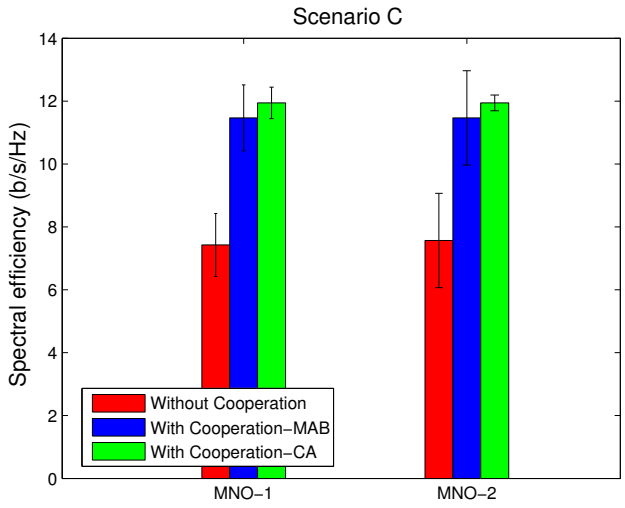


FIGURE 7.32: Scenario C: Spectral Efficiency

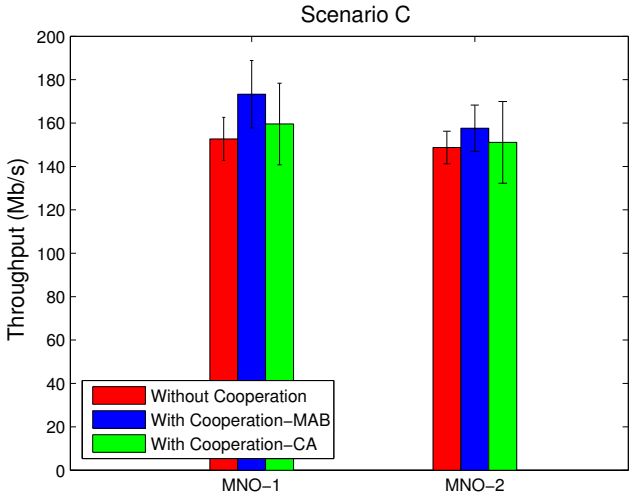


FIGURE 7.33: Scenario C: Total Throughput

7.5 Conclusion

In this chapter, we investigated the common spectrum and active infrastructure sharing among MNOs in C-RAN architecture. Two algorithms were proposed for MNOs cooperation: The first relies on an interference management technique to reduce inter-cluster interferences, and is inspired from reinforcement learning solutions targeted to resolve the MAB problem, in which each BBU chooses autonomously its most appropriate operating band. The second consists of the CA, which is a specification of LTE release 10, targeted to aggregate the bands so as to increase network capacity. A realistic positioning of antenna elements was considered, in which RRHs are split between two MNOs, providing different coverage. The results are compared to the case where resources are not shared, and according to different scenarios in terms of load conditions. Important conclusions are summarized in the following points:

— When one MNO is highly loaded and a second is lowly loaded, the first consumes the resources of the second, affecting its metrics which slightly degrade. However, the performance metrics are globally enhanced within the network.

— When both MNOs are highly loaded, performance metrics are enhanced for all MNOs when cooperation algorithms are applied.
— When MNOs are lowly loaded, additional power savings and increased energy and spectral efficiencies are derived.
— The CA algorithm shows more enhanced metrics than the MAB algorithm. However, it requires more RRM complexity. Further, it is not supported by all UE types, such as LTE release 8.
— Power savings, energy and spectral efficiency are globally enhanced in C-RAN architecture when cooperation among MNOs is made possible.

Chapitre 8

Conclusion

In this chapter, we summarize our contributions, and we provide the short and long term perspectives of the present work.

8.1 Summary of Contributions

This thesis addresses the BBU-RRH association optimization in the C-RAN architecture. We have proposed several algorithms for the problem, and evaluated the performance metrics through extensive simulations.
A first contribution in our work, relies on an Interference-Aware approach for BBU-RRH association optimization. The objective of the proposed approach is to reduce the total power consumption in the system, while being aware of intra-cluster interference canceled among RRHs associated to one BBU, owing to DAS behavior. Consequently, an ILP algorithm, relying on a SPP formulation is proposed, where the subsets of RRHs are chosen in a way to reduce a total cost function (*i.e*, the total power consumption), and to respect a minimum level of throughput requirement per UE. The simulation results show that our Interference-Aware approach outperforms the adopted proposition of the state-of-art, the Bin Packing formulation, in terms of power savings and energy efficiency. Due to the high computational requirements of NP-complete problems, such as the SPP, we have proposed a greedy heuristic with low complexity and close performance to the optimal solution. Further, we have adapted several Bin Packing heuristics to behave in an Interference-Aware manner, and applied them on a realistic antenna deployment scenario. Further many heuristics were proposed and compared among each other.
A second contribution in our work, relies on formulating a hybrid approach for BBU-RRH association optimization, based on Game Theory. Our approach relies on an exact potential game among RRHs, which is an appealing property as Best Response dynamics are guaranteed to converge to PNE. The purpose of our hybrid approach is to reduce the amount of signaling load overhead necessary in centralized schemes. In this work, we first propose an algorithm adapting to homogeneous UEs distribution in the network. We then enhance our proposition to suit a heterogeneous distribution of UEs, which is more realistic. The simulation results show that our propositions provide close performance to centralized schemes with reduced amount of signaling load overhead.
A third contribution addresses the dynamic aspect of BBU-RRH association problem. In particular, we aim to address the frequent re-associations between BBUs and RRHs, due to the dynamicity in network load conditions. In fact, when a RRH re-associates to a different BBU, connected UEs face HOs from one BBU to another, which causes a degradation in the level of QoE. To solve this issue, we have proposed a bi-objective optimization problem, addressing two targets: The power consumption reduction, and the UEs' re-association rate minimization. The problem is formulated as an ILP problem, and is presented as a SPP. Through simulations, we have shown a tradeoff between both metrics: On one hand, reducing the power consumption might change the BBU-RRH association schemes under any load fluctuation to guarantee a minimum level of QoS. This causes frequent HOs for serviced UEs. On the other hand, activating a high number of BBUs always guarantees the constraint of QoS, avoiding frequent re-associations. However, this hinders the energy efficiency. We have further proposed a heuristic, addressing both objectives, and showing a good compromise between both metrics. We have lastly, proposed a hybrid solution for the problem, with the purpose of reducing the signaling load overhead. The hybrid approach also shows good tradeoff between power consumption minimization, and re-association rate reduction, with reduced level of signaling load overhead.
A last contribution aims to investigate the common spectrum and active infrastructure sharing in the C-RAN ar-

chitecture, which has not been addressed in the recent state-of-art. In particular, we have proposed two different algorithms, where MNOs share their active network components and spectral resources in C-RAN. In that context, we have assessed the additional power savings derived from BBU-RRH association optimization. The first algorithm is based on an interference management technique, relying on the MAB problem, where each BBU chooses independently an operating band in a way to enhance its UEs' SINR. The second relies on a feature in LTE release 10 specification, the CA. We have evaluated the results under different scenarios in terms of load conditions, and compared them with the case where MNOs do not share their resources. The performance metrics are globally enhanced in terms of power saving, energy and spectral efficiency for both algorithms in comparison with the case of non-resource sharing.

To conclude, we summarize our contributions with their different application domains according to MNOs needs and preferences :

- **Interference-Aware Approach**: The Interference-Aware approach has the originality of being aware of the interference level in the network to reduce the power consumption. The results showed that it achieves more power savings than the Bin Packing scheme, which also seeks to reduce the power consumption by decreasing the number of active BBUs. Such approach suits well for MNOs whose power economy and OPEX costs are paramount. It also fits well for dense environments, where the operating spectrum is aggressively reused, since it reduces the interference level.
- **Hybrid Approach**: The hybrid approach has the originality of reducing the signaling load overhead between BBUs and RRHs. Thus, for MNOs limited by fronthaul resources, such solution would be adequate. It also presents a practical side, since some MNOs would prefer to connect RRHs to BBUs with microwave link connections which are known for their practicability in installation and low cost, however, they present huge limitation in terms of bandwidth. In that case, reducing the signaling load overhead would be paramount.
- **Clustering Approach with Re-association Consideration**: Considering the re-association of RRHs to BBUs in dynamic environments is paramount to avoid the degradation of the QoE for serviced UEs at one hand, and to achieve power economy on the other hand. In that context, our contributions have the appealing property of achieving a good tradeoff between both metrics (*i.e*, the power economy and the re-association rate of UEs), as shown by the results.
- **Clustering in Multi-operator Context**: The present solution has the originality of efficiently exploiting the resources. Particularly, it introduces a cooperative framework among MNOs within a C-RAN architecture, leading to higher power savings from BBU-RRH association, and to enhanced spectral efficiency either from interference management, or from increased capacities, since MNOs access each others' bands. Because of higher spectral and energy efficiency, OPEX and costs are reduced, which positively impacts MNOs of limited resources.

Table 8.1 summarizes all of the application domains of each approach.

8.2 Future Directions

We expose hereafter the future perspectives of the work in subsections 8.2.1, and 8.2.2, exposing short and long term perspectives respectively.

8.2.1 Short Term Perspectives

For short term perspectives, we first propose to exploit the BBU-RRH association problem with Inter-Cell Interference Cancellation (ICIC), and massive MIMO technique. In particular, in our system model we resorted to the reuse-1 scheme. By applying interference management techniques, such as ICIC, performance metrics could be improved, enhancing spectral and energy efficiencies. Further, massive MIMO technique introduces either diversity or spatial multiplexing gain, improving energy and spectral efficiencies, and deriving more power savings. Also, to accelerate the search of the Replicator Dynamics algorithm, RRHs could be divided into several subsets, where the algorithm would be applied in parallel onto each subset. In that case, faster solutions are obtained. Additionally, for the last contribution, we propose to derive decisions of cooperation as a function of inter-cell distances. In fact, taking into consideration the distances between cells, one MNO can accurately offload its traffic to a different one, which helps in deriving more power savings from cooperation, while guaranteeing a minimum level of QoS. We also propose to study the impact of the signaling load overhead derived from CA technique on a constrained optical fronthaul. In fact, the impact of signaling load overhead from CA in C-RAN architecture has not been investigated in the recent state-of-art. By proposing an algorithm that decides whether a UE should operate on a single carrier

Contribution	Application Domain
Interference-Aware Approach	— Very efficient when power economy is paramount for MNOs — Reduces OPEX — Works well for dense environment with high level of interference
Hybrid Approach	— Very efficient when the fronthaul connection between BBUs and RRHs is limited — Efficient for practical implementation
Clustering Approach with Re-association Consideration	— Very efficient for MNOs that prioritize the QoE for UEs — Realizes a good tradeoff between power savings and re-association rate
Clustering in Multi-operator Context	— Very efficient for MNOs with limited resources (*i.e*, limited in coverage and spectral resources) — Reduces MNOs OPEX

TABLE 8.1: Application Domains of the proposed Approaches

or multiple carriers, good tradeoff between signaling load reduction and enhanced throughputs is realized. The proposed algorithm must take into consideration several constraints, such as the traffic demand, a UE QoS, and the constraint on optical fronthaul capacity.

8.2.2 Long Term Perspectives

As for long term perspectives, we propose to extend our proposed algorithms to real testbed experiments. In fact, the C-RAN architecture can leverage the benefits of SDR. In particular, the baseband functionalities can be instantiated over software, without the need of dedicated hardware. In that case, resorting to open source softwares, such as OpenAirInterface [107] and Ettus Research's Universal Software Radio Platform (USRP) [108] would be an effective solution. Additionally, resorting to learning techniques, to dynamically decide clustering schemes, adapting to dynamic load variations would also be an interesting solution. Moreover, we propose to select BBUs in a more defined granularity, based on QoS requirements and BBUs resiliency [109] for different business models. In particular, by making use of network slicing concept, which is a promising feature in future 5G networks, sharing techniques are achieved between multiple MNOs, reducing CAPEX and OPEX, while respecting QoS requirements for different services and business models. Further, Mobile Edge Computing (MEC) [110] is a promising feature to reduce tight delay applications, requiring fast computational requirements. By exploiting edge computing, along with centralized RAN management, benefits in terms of multiplexing gains are realized, while respecting the different delay requirements of real time applications, such as M2M communication, smart vehicles, virtual reality etc.

Annexe A

The appendix is devoted to prove Theorem 1.

Let $\mathbf{p} = (p_1, p_2, \ldots, p_R)$ be a mixed profile of the game. $\mathbb{E}[Cost_r|\mathbf{p}]$ denotes the expected cost of RRH r with respect to a mixed profile $\mathbf{p}$.

According to [111], the Linear Reward-Inaction algorithm converges weakly towards a replication dynamic :

$$\frac{dp_{r,y}}{dt}(\mathbf{p}) = p_{r,y}\left(\mathbb{E}[Cost_r|\mathbf{p}] - E[Cost_r|p_{r,y} = 1, p_{-r}]\right) \tag{A.1}$$

This equation, called the (multi-population) replicator dynamics, is well-known to have its limit points related to Nash equilibriums (through the so-called Folk's theorem of evolutionary game theory [112]). More precisely, we have the following theorem :

Theorem 2 *The following are true for the solutions of Equation* (A.1) *:*
(i) All Nash equilibriums are stationary points.
(ii) All strict Nash equilibriums are asymptotically stable.
(iii) All stable stationary points are Nash equilibriums.

From [79], the limit for $k \to 0$ of the dynamics of stochastic algorithms is some Ordinary Differential Equations (ODE) whose stable limit points, when $t \to \infty$ (if they exist), can only be Nash equilibriums. Hence, if there is convergence for the ordinary differential equation, then one expects the replicator dynamic algorithm to reach an equilibrium. Moreover, in [113], the authors prove that such Nash equilibriums are pure.

Let us see if the continuous dynamic converges with stability arguments. Given a pure profile $\mathbf{y} = (y_1, y_2, \ldots, y_R)$, we denote by $\pi(\mathbf{y}|\mathbf{p})$ the probability that RRH r chooses the pure strategy y_r according to the mixed profile $\mathbf{p}$:

$$\pi(\mathbf{y}|\mathbf{p}) = \prod_{r=1}^{R} p_{r,y_r}. \tag{A.2}$$

To prove the convergence of the learning scheme Eq. (5.8), we define the following function $F : \mathbb{K} \to \mathbb{R}$:

$$F(\mathbf{p}) = \sum_{\mathbf{y}\in\mathcal{S}} \pi(\mathbf{y}|\mathbf{p})\phi(\mathbf{y}) \tag{A.3}$$

Let us study the evolution of function $F(\mathbf{p})$ over time. We focus on $\frac{dF}{dt}(\mathbf{p})$. By definition, we have

$$\frac{dF}{dt}(\mathbf{p}) = \sum_{r=1}^{R} \sum_{s\in S_r} \frac{\partial F}{\partial p_{r,s}} \frac{dp_{r,s}}{dt}(\mathbf{p})$$

We will compute first $\frac{\partial \pi(\mathbf{y}|\mathbf{p})}{\partial p_{r',y'_{r'}}}$ and then $\frac{\partial F}{\partial p_{r',y'_{r'}}}(\mathbf{p})$.

$$\frac{\partial \pi(\mathbf{y}|\mathbf{p})}{\partial p_{r',y'_{r'}}} = \begin{cases} \prod_{r=1, r\neq r'}^{R} p_{r,y_r} & \text{if} \quad y_{r'} = y'_{r'}, \\ 0 & \text{otherwise.} \end{cases} \tag{A.4}$$

This formula can be rewritten as

$$\frac{\partial \pi(\mathbf{y}|\mathbf{p})}{\partial p_{r',y'_{r'}}} = \pi(\mathbf{y}|p_{r',y'_{r'}} = 1, \mathbf{p}_{-r'}).$$

Hence, we get what follows :

$$\begin{aligned}\tfrac{\partial F}{\partial p_{r',y'_{r'}}}(\mathbf{p}) &= \sum_{\mathbf{y}\in\mathcal{S}} \frac{\partial \pi(\mathbf{y}|\mathbf{p})}{\partial p_{r',y'_{r'}}}\phi(\mathbf{y}) \\ &= \sum_{\mathbf{y}\in\mathcal{S}} \pi(\mathbf{y}|p_{r',y'_{r'}} = 1, \mathbf{p}_{-r'})\phi(\mathbf{y})\end{aligned}$$

From Equation (5.2), we can obtain :

$$\begin{aligned}&\tfrac{\partial F}{\partial p_{r',y'_{r'}}} - \tfrac{\partial F}{\partial p_{r',y_{r'}}} \\ &= \sum_{\mathbf{y}\in\mathcal{S}} \pi(\mathbf{y}|p_{r',y'_{r'}} = 1, \mathbf{p}_{-r'})(\phi(y'_{r'}, \mathbf{y}_{-r'}) - \phi(y_{r'}, \mathbf{y}_{-r'})) \\ &= \sum_{\mathbf{y}\in\mathcal{S}} \pi(\mathbf{y}|p_{r',y'_{r'}} = 1, \mathbf{p}_{-r'})(Cost'_r(y'_{r'}, \mathbf{y}_{-r'}) - Cost'_r(y_{r'}, \mathbf{y}_{-r'})) \\ &= \mathbb{E}[Cost_{r'}|p_{r',y'_{r'}} = 1, \mathbf{p}_{-r'}] - \mathbb{E}[Cost_{r'}|p_{r',y_{r'}} = 1, \mathbf{p}_{-r'}]\end{aligned}$$

So,

$$\frac{\partial F}{\partial p_{r',y'_{r'}}} - \frac{\partial F}{\partial p_{r',y_{r'}}} = \mathbb{E}[Cost_{r'}|p_{r',y'_{r'}} = 1, \mathbf{p}_{-r'}] - \mathbb{E}[Cost_{r'}|p_{r',y_{r'}} = 1, \mathbf{p}_{-r'}] \tag{A.5}$$

Hence, we have the following :

$$\begin{aligned}&\frac{dF}{dt}(\mathbf{p}) = \sum_{r\in\mathcal{R}}\sum_{y_r\in S_r} \frac{\partial F}{\partial p_{r,y_r}}\frac{dp_{r,y_r}}{dt}(\mathbf{p}) \\ &= \sum_{r\in\mathcal{R}}\sum_{y_r\in S_r} p_{r,y_r}\frac{\partial F}{\partial p_{r,y_r}}\left(\mathbb{E}[Cost_r|\mathbf{p}] - \mathbb{E}[Cost_r|p_{r,y_r} = 1, \mathbf{p}_{-r}]\right) \\ &= \sum_{r}\sum_{y_r\in S_r}\sum_{y'_r\in S_r} p_{r,y_r}p_{r,y'_r}\frac{\partial F}{\partial p_{r,y_r}}\cdot \\ &\left(\mathbb{E}[Cost_r|p_{r,y'_r} = 1, \mathbf{p}_{-r}] - \mathbb{E}[Cost_r|p_{r,y_r} = 1, \mathbf{p}_{-r}]\right) \\ &= \sum_{r}\sum_{y_r<y'_r} p_{r,y_r}p_{r,y'_r}\left(\frac{\partial F}{\partial p_{r,y_r}} - \frac{\partial F}{\partial p_{r,y'_r}}\right)\cdot \\ &\left(\mathbb{E}[Cost_r|p_{r,y'_r} = 1, \mathbf{p}_{-r}] - \mathbb{E}[Cost_r|p_{r,y_r} = 1, \mathbf{p}_{-r}]\right) \\ &= \textstyle\sum_r \sum_{y_r<y'_r} -p_{r,y_r}p_{r,y'_r}\cdot \\ &\left(\mathbb{E}[Cost_r|p_{r,y'_r} = 1, \mathbf{p}_{-r}] - \mathbb{E}[Cost_r|p_{r,y_r} = 1, \mathbf{p}_{-r}]\right)^2\end{aligned}$$

So, we can conclude that $\frac{dF}{dt}(\mathbf{p}) \leq 0$.

Thus F is non-decreasing along the trajectories of the replication dynamics. and, due to the nature of the learning algorithm, all solutions of the ODE (A.1) remain in the strategy space if initial conditions $\in [0, 1]$. From the previous computations, we know that $\frac{dF(\mathbf{p}^*)}{dt} = 0$ implies that $\forall r \in \mathcal{R}\forall y_r, y'_r \in S_r$:

$$p^*_{r,y_r} = 0, or\ \left(\mathbb{E}[Cost_r|p_{r,y_r} = 1, \mathbf{p}_{-r}] = \mathbb{E}[Cost_r|p_{r,y_r} = 1, \mathbf{p}_{-r}]\right)$$

Such a $\mathbf{p}^*$ is consequently a stationary point of the dynamics.

Since from Theorem 2, all stationary points that are not Nash equilibriums are unstable, Theorem 1 holds. Thus all solutions have to converge to some stationary point corresponding to Nash Equilibrium. We can deduce that the learning algorithm, for any initial condition in $\mathbb{K} - \mathbb{K}^*$, always converges to a Nash Equilibrium of instance G.

Publications

Conferences

[114] K. Boulos, K. Khawam, M. E. Helou, M. Ibrahim, H. Sawaya, and S. Martin, “An efficient scheme for BBU-RRH association in C-RAN architecture for joint power saving and re-association optimization,” in *7th IEEE International Conference on Cloud Networking, CloudNet 2018, Tokyo, Japan, October 22-24, 2018*, 2018

[115] K. Boulos, K. Khawam, M. E. Helou, M. Ibrahim, S. Martin, and H. Sawaya, “A hybrid approach for RRH clustering in cloud radio access networks based on game theory,” in *Proceedings of the 16th ACM International Symposium on Mobility Management and Wireless Access, MobiWac 2018, Montreal, QC, Canada, October 28 - November 02, 2018*, 2018

[116] K. Boulos, M. E. Helou, K. Khawam, M. Ibrahim, S. Martin, and H. Sawaya, “RRH clustering in cloud radio access networks with re-association consideration,” in *2018 IEEE Wireless Communications and Networking Conference, WCNC 2018, Barcelona, Spain, April 15-18, 2018*, 2018

[117] K. Boulos, M. E. Helou, M. Ibrahim, K. Khawam, H. Sawaya, and S. Martin, “Interference-aware clustering in cloud radio access networks,” in *2017 IEEE 6th International Conference on Cloud Networking (CloudNet)*, Sept 2017

Journals

[118] K. Boulos, K. Khawam, M. E. Helou, M. Ibrahim, S. Martin, H. Sawaya, “On Common Spectrum and Active Infrastructure Sharing in Cloud Radio Access Networks Architecture,” submitted to IEEE access 2019

Bibliographie

[1] C. Mobile, “C-RAN : the road towards green RAN,” *White Paper, ver*, 2011.

[2] K. Sundaresan, M. Y. Arslan, S. Singh, S. Rangarajan, and S. V. Krishnamurthy, “Fluidnet : A flexible cloud-based radio access network for small cells,” *IEEE/ACM Transactions on Networking*, April 2016.

[3] *Cisco Visual Networking Index : Global Mobile Data Traffic Forecast Update, 2016–2021*, Cisco, February 2017.

[4] “Mobile network operators face cost crunch, available online,” MarketingCharts, 2011, https://www.marketingcharts.com.

[5] *Interworking and JOINt Design of an Open Access and Backhaul Network Architecture for Small Cells based on Cloud networks*, IJOIN, 2012, https://cordis.europa.eu/project/rcn/105819/factsheet/fr.

[6] *Mobile Cloud Networking project*, MCN, 2013, http://mobile-cloud-networking.eu/site/.

[7] “Special articles on 5G technologies toward 2020 deployment,” *Technical Journal*, vol. 17, no. 4, 2016.

[8] “Common public radio interface (CPRI) : Interface specification,” CPRI, http://www.cpri.info.

[9] F. K. Jondral, “Software-defined Radio : Basics and evolution to cognitive radio,” *EURASIP J. Wirel. Commun. Netw.*, Aug. 2005.

[10] *Multi-layer and cloud-ready radio evolution towards 5G, white paper*, Nokia, 2016, https://gsacom.com/paper/multi-layer-cloud-ready-radio-evolution-towards-5g/.

[11] P. Rost, C. J. Bernardos, A. D. Domenico, M. D. Girolamo, M. Lalam, A. Maeder, D. Sabella, and D. Wübben, “Cloud technologies for flexible 5G radio access networks,” *IEEE Communications Magazine*, May 2014.

[12] D. Castanheira and A. Gameiro, “Distributed antenna system capacity scaling,” *Wireless Commun.*, 2010.

[13] *Next Generation Mobile Network*, NGMN, https://www.ngmn.org.

[14] G. Mountaser, M. L. Rosas, T. Mahmoodi, and M. Dohler, “On the feasibility of MAC and PHY split in Cloud RAN,” in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, March 2017.

[15] N. Mharsi, M. Hadji, D. Niyato, W. Diego, and R. Krishnaswamy, “Scalable and cost-efficient algorithms for baseband unit (BBU) function split placement,” in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, April 2018.

[16] Y. Huang, C. Lu, M. Berg, and P. Ödling, “Functional split of zero-forcing based massive MIMO for fronthaul load reduction,” *IEEE Access*, 2018.

[17] D. Lee, H. Seo, B. Clerckx, E. Hardouin, D. Mazzarese, S. Nagata, and K. Sayana, “Coordinated Multipoint transmission and reception in LTE-advanced : deployment scenarios and operational challenges,” *IEEE Communications Magazine*, February 2012.

[18] “Evolved Universal Terrestrial Radio Access (E-UTRA) ; Physical channels and modulation,” 3GPP, Tech. Rep. TS 36.211, Jul. 2008.

[19] M. Diaby, “Linear programming formulation of the set partitioning problem,” *International Journal of Operational Research*, 2010.

[20] D. Mishra, P. C. Amogh, A. Ramamurthy, A. A. Franklin, and B. R. Tamma, “Load-aware dynamic RRH assignment in cloud radio access networks,” in *2016 IEEE Wireless Communications and Networking Conference*, April 2016.

[21] T. Sigwele, A. S. Alam, P. Pillai, and Y. F. Hu, “Energy-efficient cloud radio access networks by cloud based workload consolidation for 5G,” *J. Netw. Comput. Appl.*, Jan. 2017.

[22] M. Khan, F. A. Sabir, and H. S. Al-Raweshidy, “Load balancing by dynamic BBU-RRH mapping in a self-optimised cloud radio access network,” in *2017 24th International Conference on Telecommunications (ICT)*, 2017.

[23] http://www.antennesmobiles.fr.

[24] S. Namba, T. Matsunaka, T. Warabino, S. Kaneko, and Y. Kishi, “Colony-RAN architecture for future cellular network,” in *2012 Future Network Mobile Summit (FutureNetw)*, 2012.

[25] F. Moety, “Joint minimization of power and delay in wireless access networks. (minimisation conjointe de la puissance et du délai dans les réseaux d’accès sans-fil),” Ph.D. dissertation, University of Rennes 1, France, 2014.

[26] “Evolved Universal Terrestrial Radio Access (E-UTRA) ; Radio Resource Control (RRC) ; Protocol specification,” 3GPP, Tech. Rep. 36.331, Jul. 2016.

[27] B. Korte and J. Vygen, *Combinatorial Optimization : Theory and Algorithms*. Springer Berlin Heidelberg, 2000.

[28] K. Boulos, M. E. Helou, and S. Lahoud, “RRH clustering in cloud radio access networks,” in *2015 International Conference on Applied Research in Computer Science and Engineering (ICAR)*, Oct 2015.

[29] K. Wang, W. Zhou, and S. Mao, “On joint BBU/RRH resource allocation in heterogeneous Cloud-RANs,” *IEEE Internet of Things Journal*, 2017.

[30] M. Khan, R. S. Alhumaima, and H. S. Al-Raweshidy, “Quality of service aware dynamic BBU-RRH mapping in cloud radio access network,” in *2015 International Conference on Emerging Technologies (ICET)*, Dec 2015.

[31] T. Sigwele, A. S. Alam, P. Pillai, and Y. F. Hu, “Evaluating energy-efficient cloud radio access networks for 5g,” in *2015 IEEE International Conference on Data Science and Data Intensive Systems*, Dec 2015.

[32] M. Qian, W. Hardjawana, J. Shi, and B. Vucetic, “Baseband processing units virtualization for cloud radio access networks,” *IEEE Wireless Communications Letters*, April 2015.

[33] S. R. Aldaeabool and M. F. Abbod, “Reducing power consumption by dynamic BBUs-RRHs allocation in C-RAN,” in *2017 25th Telecommunication Forum (TELFOR)*, Nov 2017.

[34] T. Sigwele, P. Pillai, and Y. F. Hu, “itree : Intelligent traffic and resource elastic energy scheme for Cloud-RAN,” in *2015 3rd International Conference on Future Internet of Things and Cloud*, Aug 2015.

[35] W. Al-Zubaedi and H. S. Al-Raweshidy, “A parameterized and optimized BBU pool virtualization power model for C-RAN architecture,” in *IEEE EUROCON 2017 -17th International Conference on Smart Technologies*, July 2017.

[36] O. Chabbouh, S. B. Rejeb, Z. Choukair, and N. Agoulmine, “A two-stage RRH clustering mechanism in 5G heterogeneous C-RAN,” in *5th International Workshop on ADVANCEs in ICT Infrastructures and Services (ADVANCE 2017)*, Jan. 2017.

[37] D. S. Johnson, “Fast algorithms for bin packing,” *J. Comput. Syst. Sci.*

[38] F. M. McNeill and E. Thro, *Fuzzy Logic : A Practical Approach*. Academic Press Professional, Inc., 1994.

[39] Y. Shi, J. Zhang, and K. B. Letaief, “Group sparse beamforming for green Cloud-RAN,” *IEEE Transactions on Wireless Communications*, May 2014.

[40] J. Yao and N. Ansari, “QoS-aware joint BBU-RRH mapping and user association in Cloud-RANs,” *IEEE Transactions on Green Communications and Networking*, Dec.

[41] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows : Theory, Algorithms, and Applications*, 1993.

[42] T. Sigwele, A. S. Alam, P. Pillai, and Y. F. Hu, “On the energy minimization of heterogeneous cloud radio access networks,” in *Wireless and Satellite Systems*, I. Otung, P. Pillai, G. Eleftherakis, and G. Giambene, Eds. Springer International Publishing, 2017.

[43] E. Aqeeli, A. Moubayed, and A. Shami, “Power-aware optimized RRH to BBU allocation in C-RAN,” *IEEE Transactions on Wireless Communications*, Feb 2018.

[44] A. Younis, T. X. Tran, and D. Pompili, “Bandwidth and energy-aware resource allocation for cloud radio access networks,” *IEEE Trans. Wireless Communications*, 2018.

[45] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[46] S. Martello and P. Toth, *Knapsack Problems : Algorithms and Computer Implementations*, 1990.

[47] M. Y. Lyazidi, N. Aitsaadi, and R. Langar, "Dynamic resource allocation for cloud-ran in LTE with real-time BBU/RRH assignment," in *2016 IEEE International Conference on Communications (ICC)*, May 2016.

[48] ——, "A dynamic resource allocation framework in LTE downlink for cloud-radio access network," *Computer Networks*, 2018.

[49] M. Vincenzi, A. Antonopoulos, E. Kartsakli, J. Vardakas, L. Alonso, and C. Verikoukis, "Cooperation incentives for multi-operator C-RAN energy efficient sharing," in *2017 IEEE International Conference on Communications (ICC)*, 2017.

[50] N. Wirth, *Algorithms &Amp ; Data Structures*, 1986.

[51] K. Wang, M. Zhao, and W. Zhou, "Traffic-aware graph-based dynamic frequency reuse for heterogeneous Cloud-RAN," in *2014 IEEE Global Communications Conference*, Dec 2014.

[52] M. Yassin, "Inter-cell interference coordination in wireless networks. (coordination des interférences intercellulaires dans les réseaux sans-fil)," Ph.D. dissertation, University of Rennes 1, France, 2015.

[53] T. Withawat, H. Takeshi, and O. Minoru, "Handover reduction using optical matrix switch for centralized radio access network," *IEICE Communications Express*, Aug 2016.

[54] D. Naboulsi, A. Mermouri, R. Stanica, H. Rivano, and M. Fiore, "On user mobility in dynamica cloud radio access networks," in *2018 IEEE Conference on Computer Communications, INFOCOM 2018, Honolulu, HI, USA, April 16-19, 2018*, 2018.

[55] S. Fortunato and A. Lancichinetti, "Community detection algorithms : a comparative analysis," in *4th International Conference on Performance Evaluation Methodologies and Tools, VALUETOOLS '09, Pisa, Italy, October 20-22, 2009*.

[56] C. A. C. Coello, G. B. Lamont, and D. A. V. Veldhuizen, *Evolutionary Algorithms for Solving Multi-Objective Problems (Genetic and Evolutionary Computation)*. Berlin, Heidelberg : Springer-Verlag, 2006.

[57] M. Khan, R. S. Alhumaima, and H. S. Al-Raweshidy, "QoS-aware dynamic RRH allocation in a self-optimized cloud radio access network with RRH proximity constraint," *IEEE Transactions on Network and Service Management*, 2017.

[58] M. Khan, Z. H. Fakhri, and H. S. Al-Raweshidy, "Semistatic cell differentiation and integration with dynamic BBU-RRH mapping in cloud radio access network," *IEEE Transactions on Network and Service Management*, March 2018.

[59] J. Li, X. Wu, and R. Laroia, *OFDMA Mobile Broadband Communications : A Systems Approach*. Cambridge University Press, 2013.

[60] S. Nanba, T. Warabino, and S. Kaneko, "BBU-RRH switching schemes for centralized RAN," in *7th International Conference on Communications and Networking in China, Kunming, Yunnan Province, China, August 8-10, 2012*, 2012.

[61] H. Wang, F. Xu, Y. Li, P. Zhang, and D. Jin, "Understanding mobile traffic patterns of large scale cellular towers in urban environment," in *Proceedings of the 2015 Internet Measurement Conference*, 2015.

[62] D. Lee, S. Zhou, X. Zhong, Z. Niu, X. Zhou, and H. Zhang, "Spatial modeling of the traffic density in cellular networks," *Wireless Communications, IEEE*, 02 2014.

[63] J. F. Thompson, Z. U. Warsi, and C. W. Mastin, *Numerical Grid Generation : Foundations and Applications*. New York, NY, USA : Elsevier North-Holland, Inc., 1985.

[64] T.-T. Tran, Y. Shin, and O.-S. Shin, "Overview of enabling technologies for 3GPP LTE-advanced," *EURASIP Journal on Wireless Communications and Networking*, Feb 2012.

[65] A. Liu and V. K. N. Lau, "Two-timescale user-centric RRH clustering and precoding optimization for Cloud RAN via local stochastic cutting plane," *IEEE Transactions on Signal Processing*, Jan 2018.

[66] A. Dammann, S. Plass, and S. Sand, "Cyclic delay diversity - a simple, flexible and effective multi-antenna technology for OFDM," in *2008 IEEE 10th International Symposium on Spread Spectrum Techniques and Applications*, Aug 2008.

[67] H. D. Thang, L. Boukhatem, M. KaneW, and S. Martin, "Performance-cost trade-off of joint beamforming and user clustering in cloud radio access networks," in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Oct 2017.

[68] H. D. Thang, L. Boukhatem, M. Kaneko, and S. Martin, “An advanced mobility-aware algorithm for joint beamforming and clustering in heterogeneous cloud radio access network,” in *Proceedings of the 21st ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, ser. MSWIM ’18, 2018.

[69] C. Liu, Y. Wan, L. Tian, Y. Zhou, and J. Shi, “Base station sleeping control with energy-stability tradeoff in centralized radio access networks,” in *2015 IEEE Global Communications Conference (GLOBECOM)*, Dec 2015.

[70] M. Maiza and M. S. Radjef, “Heuristics for solving the bin-packing problem with conflicts,” *Applied Mathematical Sciences (Ruse)*, 01 2011.

[71] C. Chang, R. Schiavi, N. Nikaein, T. Spyropoulos, and C. Bonnet, “Impact of packetization and functional split on C-RAN fronthaul performance,” in *2016 IEEE International Conference on Communications (ICC)*, May 2016.

[72] J. Liu, S. Zhou, J. Gong, Z. Niu, and S. Xu, “Graph-based framework for flexible baseband function splitting and placement in C-RAN,” in *2015 IEEE International Conference on Communications (ICC)*, June 2015.

[73] A. Checko, A. P. Avramova, M. S. Berger, and H. L. Christiansen, “Evaluating C-RAN fronthaul functional splits in terms of network level energy and cost savings,” *Journal of Communications and Networks*, April 2016.

[74] J. F. Nash, “Equilibrium points in n-person games,” *Proc. of the National Academy of Sciences*, 1950.

[75] Y. Z. et al., “White paper of next generation fronthaul interface v1.0,” China Mobile Research Institute, Tech. Rep., 2015.

[76] Q. D. L, Y. H. Chew, and B.-H. Soong, *Potential Game Theory : Applications in Radio Resource Allocation*, 1st ed. Springer Publishing Company, Incorporated, 2016.

[77] R. W. Rosenthal, “A class of games possessing pure-strategy nash equilibria,” *International Journal of Game Theory*, Dec 1973.

[78] M. Voorneveld, “Best-response potential games,” *Economics Letters*, 2000.

[79] O. Bournez and J. Cohen, “Learning equilibria in games by stochastic distributed algorithms,” in *Computer and Information Sciences III*, E. Gelenbe and R. Lent, Eds. Springer London, 2013.

[80] M. Z. Shafiq, J. Erman, L. Ji, A. X. Liu, J. Pang, and J. Wang, “Understanding the impact of network dynamics on mobile video user engagement,” in *The 2014 ACM International Conference on Measurement and Modeling of Computer Systems*. ACM, 2014.

[81] A. Balachandran, V. Aggarwal, E. Halepovic, J. Pang, S. Seshan, S. Venkataraman, and H. Yan, “Modeling web quality-of-experience on cellular networks,” in *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom ’14. ACM, 2014.

[82] C. Gomez, M. Catalan, X. Figueras, J. Paradells, and A. Calveras, “Impact of handover between umts and gprs on tcp/ip : An empirical approach,” in *IEEE Vehicular Technology Conference*, Sep. 2006.

[83] “Infrastructure sharing among MNOs,” Applied Value Telecom Series, 2014.

[84] “Service aspects and requirements for network sharing,” 3GPP, Tech. Rep., 2012, 3GPP TR 22.951 V11.0.0.

[85] P. Auer and R. Ortner, “Ucb revisited : Improved regret bounds for the stochastic multi-armed bandit problem,” *Periodica Mathematica Hungarica*, Sep 2010.

[86] P. Auer, N. Cesa-Bianchi, and P. Fischer, “Finite-time analysis of the multiarmed bandit problem,” *Machine Learning*, May 2002.

[87] S. Valentin, W. Jamil, and O. Aydin, “Extending generalized processor sharing for multi-operator scheduling in cellular networks,” in *2013 9th International Wireless Communications and Mobile Computing Conference (IWCMC)*, July 2013.

[88] I. Malanchini, S. Valentin, and O. Aydin, “Generalized resource sharing for multiple operators in cellular wireless networks,” in *2014 International Wireless Communications and Mobile Computing Conference (IWCMC)*, Aug 2014.

[89] A. Gudipati, L. E. Li, and S. Katti, in *Proceedings of the Third Workshop on Hot Topics in Software Defined Networking*, 2014.

[90] A. Antonopoulos, E. Kartsakli, A. Bousia, L. Alonso, and C. Verikoukis, “Energy-efficient infrastructure sharing in multi-operator mobile networks,” *IEEE Communications Magazine*, May 2015.

[91] M. A. Safwat, “Framework for multi-operator collaboration for green communication,” *IEEE Access*, 2018.

[92] T. Sanguanpuak, S. Guruacharya, E. Hossain, N. Rajatheva, and M. Latva-aho, “Infrastructure sharing for mobile network operators : Analysis of trade-offs and market,” *IEEE Transactions on Mobile Computing*, Dec 2018.

[93] W. Guan, X. Wen, L. Wang, and Z. Lu, “On-demand cooperation among multiple infrastructure networks for multi-tenant slicing : a complex network perspective,” *IEEE Access*, 2018.

[94] X. Costa-Perez, J. Swetina, T. Guo, R. Mahindra, and S. Rangarajan, “Radio access network virtualization for future mobile carrier networks,” *IEEE Communications Magazine*, July 2013.

[95] A. Lieto, I. Malanchini, V. Suryaprakash, and A. Capone, “Making the case for dynamic wireless infrastructure sharing : A techno-economic game,” in *2017 15th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, May 2017.

[96] Ö. U. Akgül, I. Malanchini, V. Suryaprakash, and A. Capone, “Dynamic resource allocation and pricing for shared radio access infrastructure,” in *2017 IEEE International Conference on Communications (ICC)*, May 2017.

[97] L. Cano, A. Capone, G. Carello, M. Cesana, and M. Passacantando, “Cooperative infrastructure and spectrum sharing in heterogeneous mobile networks,” *IEEE Journal on Selected Areas in Communications*, Oct 2016.

[98] E. Dahlman, S. Parkvall, and J. Skold, *4G : LTE/LTE-Advanced for Mobile Broadband*, 1st ed. Orlando, FL, USA : Academic Press, Inc., 2011.

[99] M. Dehghani, K. Arshad, and R. Mackenzie, “LTE-advanced radio access enhancements : A survey,” *Wirel. Pers. Commun.*

[100] L. Liu, M. Li, J. Zhou, X. She, L. Chen, Y. Sagae, and M. Iwamura, “Component carrier management for carrier aggregation in LTE-advanced system,” in *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring)*, May 2011.

[101] “Evolved universal terrestrial radio access (EUTRA) ; stage 2 description,” 3GPP TS 36.300, Tech. Rep., 2010.

[102] H. Wang, C. Rosa, and K. I. Pedersen, “Performance analysis of downlink inter-band carrier aggregation in LTE-advanced,” 09 2011.

[103] “LTE-A MC RF requirements for contiguous carriers,” 3GPP, Tech. Rep., 2009, r4-091910, 3GPP TSG-RAN WG4 51.

[104] K. Yagyu, T. Nakamori, H. Ishii, M. Iwamura, N. Miki, T. Asai, and J. Hagiwara, “Investigation on mobility management for carrier aggregation in LTE-advanced,” in *2011 IEEE Vehicular Technology Conference (VTC Fall)*, 2011.

[105] E. Dahlman, S. Parkvall, and J. Sköld, *4G : LTE/LTE-advanced for mobile broadband*, 03 2011.

[106] M. Deruyck, W. Joseph, B. Lannoo, D. Colle, and L. Martens, “Designing energy-efficient wireless access networks : LTE and LTE-advanced,” *IEEE Internet Computing*, Sep. 2013.

[107] *Openairinterface software alliance*, EURECOM, China Mobile, Orange, BUPT, IITH, OPNFV, 2016, http://www.openairinterface.org/.

[108] *Ettus research, a National Instrument company*, https://www.ettus.com.

[109] M. Y. Lyazidi, L. Giupponi, J. Mangues-Bafalluy, N. Aitsaadi, and R. Langar, “A novel optimization framework for C-RAN BBU selection based on resiliency and price,” in *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, Sept 2017.

[110] D. Sabella, A. Vaillant, P. Kuure, U. Rauschenbach, and F. Giust, “Mobile-edge computing architecture : The role of mec in the internet of things,” *IEEE Consumer Electronics Magazine*, 2016.

[111] P. S. Sastry, V. V. Phansalkar, and M. A. L. Thathachar, “Decentralized learning of nash equilibria in multi-person stochastic games with incomplete information,” *IEEE Transactions on Systems, Man, and Cybernetics*, May 1994.

[112] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics*. Cambridge University Press, 1998.

[113] P. Coucheney, C. Touati, and B. Gaujal, “Fair and efficient user-network association algorithm for multi-technology wireless networks,” in *IEEE INFOCOM 2009*, April 2009.

[114] K. Boulos, K. Khawam, M. E. Helou, M. Ibrahim, H. Sawaya, and S. Martin, “An efficient scheme for BBU-RRH association in C-RAN architecture for joint power saving and re-association optimization,” in *7th IEEE International Conference on Cloud Networking, CloudNet 2018, Tokyo, Japan, October 22-24, 2018*, 2018.

[115] K. Boulos, K. Khawam, M. E. Helou, M. Ibrahim, S. Martin, and H. Sawaya, “A hybrid approach for RRH clustering in cloud radio access networks based on game theory,” in *Proceedings of the 16th ACM International Symposium on Mobility Management and Wireless Access, MobiWac 2018, Montreal, QC, Canada, October 28 - November 02, 2018*, 2018.

[116] K. Boulos, M. E. Helou, K. Khawam, M. Ibrahim, S. Martin, and H. Sawaya, “RRH clustering in cloud radio access networks with re-association consideration,” in *2018 IEEE Wireless Communications and Networking Conference, WCNC 2018, Barcelona, Spain, April 15-18, 2018*, 2018.

[117] K. Boulos, M. E. Helou, M. Ibrahim, K. Khawam, H. Sawaya, and S. Martin, “Interference-aware clustering in cloud radio access networks,” in *2017 IEEE 6th International Conference on Cloud Networking (CloudNet)*, Sept 2017.

[118] K. Boulos, K. Khawam, M. E. Helou, M. Ibrahim, S. Martin, H. Sawaya, “On Common Spectrum and Active Infrastructure Sharing in Cloud Radio Access Networks Architecture,” submitted to IEEE access 2019.

université PARIS-SACLAY
ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Titre : Optimisation des associations BBU-RRH dans les réseaux Cloud-RAN

Mots clés : Cloud Radio Access Networks, Association BBU-RRH, Economie de puissance, Efficacité énergétique, Programmation linéaire en nombre entiers, Théorie des jeux

Résumé : De nos jours, la demande en trafic mobile a considérablement augmenté. Face à cette croissance, plusieurs propositions font l'objet d'étude pour remédier à un tel défi.
L'architecture des réseaux d'accès de type Cloud (C-RAN) est l'une des propositions pour faire face à cette demande croissante, et constitue une solution candidate potentielle pour les réseaux futurs 5G. L'architecture C-RAN dissocie deux éléments principaux de la station de base: La BBU ou "Baseband Unit", qui constitue une unité intelligente pour le traitement des données en bande de base, et le RRH ou "Remote Radio Head", constituant une antenne passive pour fournir l'accès aux utilisateurs (UEs). Grâce à l'architecture C-RAN, les BBUs sont centralement regroupées, alors que les RRHs sont distribués sur plusieurs sites. Plusieurs avantages sont ainsi dérivés, tels que le gain en multiplexage statistique, l'efficacité d'utilisation des ressources, et l'économie de puissance. Contrairement à l'architecture conventionnelle où chaque RRH est exclusivement associé à une BBU, dans l'architecture C-RAN plusieurs RRHs sont regroupés en une seule BBU lorsque les conditions de charge sont faibles. Ceci présente plusieurs avantages, tel que l'amélioration en efficacité énergétique et la minimisation de consommation de puissance. Dans cette thèse, nous adressons le problème d'optimisation des associations BBU-RRH. Nous nous intéressons à l'optimisation des regroupements des RRHs aux BBUs en tenant compte de critères multiples. Plusieurs contraintes sont ainsi envisagées, tel que la réduction de la consommation de puissance sous garantie de Qualité de Service (QoS) minimale. En outre, la prise en compte du changement du niveau d'interférence en activant/désactivant les BBUs est primordiale pour l'amélioration de l'efficacité spectrale. En plus, décider dynamiquement de la réassociation des RRHs aux BBUs sous des conditions de charges variables représente un défi, vu que les UEs connectés aux RRHs changeant leurs associations font face à des "handovers" (HOs).

Title : BBU-RRH Association Optimization in Cloud-Radio Access Networks

Keywords : Cloud Radio Access Networks, BBU-RRH Association, Power Saving, Energy Efficiency, Integer Linear Programing, Game Theory

Abstract : The demand on mobile traffic has been largely increasing nowadays. Facing such growth, several propositions are being studied to cope with this challenge.
Cloud-Radio Access Networks Architecture (C-RAN) is one of the proposed solution to address the increased demand, and is a potential candidate for future 5G networks. The C-RAN architecture dissociates two main elements composing the base station: The Baseband Unit (BBU), consisting in an intelligent element to perform baseband tasks functionalities, and the Remote Radio Head (RRH), that consists of a passive antenna element to provide access for serviced User Equipments (UEs). In C-RAN architecture, the BBUs migrate to a Cloud data center, while RRHs remain distributed across multiple sites. Several advantages are derived, such as statistical multiplexing gain, efficiency in resource utilization and power saving. Contrarily to conventional architecture, where each RRH is associated to one BBU, in C-RAN architecture, multiple RRHs can be embraced by one single BBU when network load conditions are low, bringing along several benefits, such as enhanced energy efficiency, and power consumption minimization. In this thesis, the BBU-RRH association optimization problem is addressed. Our aim is to optimize the BBU-RRH association schemes, taking into consideration several criteria. The problem presents many constraints: For example, achieving minimized power consumption while guaranteeing a minimum level of Quality of Service (QoS) is a challenging task. Further, taking into account the interference level variation while turning ON/OFF BBUs is paramount to achieve enhanced spectral efficiency. Moreover, deciding how to re-associate RRHs to BBUs under dynamic load conditions is also a challenge, since connected UEs face handovers (HOs) when RRHs change their associations.