ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

2018-ENST-0013

EDITE - ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité « Communication et Électronique »

présentée et soutenue publiquement par

Konstantinos ALEXANDRIS

le 9 mars 2018

Gestion de mobilité et de ressources pour des réseaux d'accès radio de cinquième génération

Directeur de thèse : **Navid NIKAEIN**

Jury

M. Raymond KNOPP, Professeur, EURECOM	Président du jury
M. Albert BANCHS, Professeur, Universidad Carlos III de Madrid	Rapporteur
M. Olav TIRKKONEN, Professeur, Aalto University	Rapporteur
M. Athanasios KORAKIS, Maître de Conférences, University of Thessaly	Examinateur
Mme Ingrid MOERMAN, Professeur, Ghent University	Examinateur
M. Thrasyvoulos SPYROPOULOS, Maître de Conférences, EURECOM	Examinateur

TELECOM ParisTech
école de l'Institut Télécom - membre de ParisTech

T H È S E

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

EURECOM
Sophia Antipolis

DISSERTATION

In Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
from Télécom ParisTech

Specialization

Communication and Electronics

École doctorale Informatique, Télécommunications et Électronique (Paris)

Presented by

Konstantinos ALEXANDRIS

Mobility Management and Resource Allocation towards 5G Radio Access Networks (RANs)

Defense scheduled on March 9th 2018

before a committee composed of:

Prof. Raymond Knopp	President of the Jury
Prof. Albert Banchs	Reporter
Prof. Olav Tirkkonen	Reporter
Prof. Athanasios Korakis	Examiner
Prof. Ingrid Moerman	Examiner
Prof. Thrasyvoulos Spyropoulos	Examiner
Prof. Navid Nikaein	Thesis Supervisor

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

EURECOM
Sophia Antipolis

THÈSE

présentée pour obtenir le grade de
Docteur de
Télécom ParisTech

Spécialité

Communication et Electronique

École doctorale Informatique, Télécommunications et Électronique (Paris)

Présentée par

Konstantinos ALEXANDRIS

Gestion de mobilité et de ressources pour des réseaux d'accès radio de cinquième génération

Soutenance de thèse prévue le 9 mars 2018

devant le jury composé de:

Prof. Raymond Knopp	Président du jury
Prof. Albert Banchs	Rapporteur
Prof. Olav Tirkkonen	Rapporteur
Prof. Athanasios Korakis	Examinateur
Prof. Ingrid Moerman	Examinateur
Prof. Thrasyvoulos Spyropoulos	Examinateur
Prof. Navid Nikaein	Directeur de thèse

Abstract

Mobility management and resource allocation are the key features of the current and next generation cellular systems that enable the users to change seamlessly their cell associations and maintain their service continuity with required quality-of-service (QoS). Software-defined networking (SDN) promises to overcome such adversities in network design by centrally controlling the behavior of the underlying network via an abstraction layer. These mechanisms are applied on those procedures in 5G mobile communications exploiting the holistic knowledge of the network.

In the first part of the thesis, we develop and examine the performance of the legacy 3GPP X2 handover (HO) using the OpenAirInterface (OAI) platform to understand not only the mobility management in 4G but also the interplay of system parameters on triggering the HO process towards designing a HO algorithm. As a next step, a load-aware HO decision algorithm for the next-generation heterogeneous networks (HetNets) is proposed considering the users service delay and the asymmetrical cells power. Results have shown that the proposed algorithm outperforms the conventional received signal strength (RSS) and distance-based ones. In addition, an SDN architecture is sketched to manage the system decisions resulting in a centralized network coordination.

In the second part of the thesis, the importance of users QoS in 5G networks turns our attention to radio resource management under the multi-connectivity setting. A utility-based resource allocation that considers the users QoS requirements is proposed and compared to proportional fairness schemes. Multi-connectivity use case is shown to outperform the legacy 3GPP single-connectivity in terms of user satisfaction and aggregated network throughput. In that setup, besides the air-interface limitations, we assume local routing among cells allowing to offload the backhaul/core network. Subsequently, we introduce backhaul capacity constraints leading to results that show multi-connectivity is still superior to the legacy one while retaining users QoS. Finally, we study the performance gain for a multi-connected user under spatio-temporal channel variability (large and small-scale fading) by exploiting the opportunistic scheduling for loaded and unloaded cell scenarios. Extracted results reveal the benefits of UPF utility function with opportunistic scheduling in terms of users QoS.

Acknowledgements

First of all, I would like to acknowledge my PhD thesis supervisor, Prof. Navid Nikaein, for giving me the chance to discover a new technology world, choose a challenging dissertation topic, be part of an exciting work environment and collaborate with esteemed industry experts all over the world. His guidance, support and patience, both in academic but also in personal level, made the completion of this work possible.

Also, I would like to express my sincere gratitude to Prof. Thrasyvoulos Spyropoulos. His constructive comments guided me towards the right direction to think creatively and out of the box. Moreover, I would like to express my strong appreciation to Prof. Raymond Knopp for acting as the president of my PhD committee as well as the committee members who graciously agreed to serve.

Special thanks go to Dr. Kostas Katsalis and Dr. Nikolaos Sapountzis for the smooth collaboration and helpful discussions. In particular, I am most grateful to my research partner Chia-Yu Chang. His precious help was more than crucial from a technical and non-technical perspective in order to achieve any goal within the limited time constraints.

A PhD is usually considered a solitary journey, where self-discipline needs to be strengthened. In this regard, I would like to take some time here and thank all my lifelong friends for encouraging and supporting me all these years as well as the ones that came to my life unexpectedly to change my mindset and make me a well-rounded person. Next, I am thankful to the people who created such a good atmosphere in the office and the lab.

Last but not least, I want to dedicate this thesis to my beloved family, without whom my entire studies would not have been possible. Through their deepest love and understanding, they are always present to stand next to me, help me deal better with hard times in life and keep my faith to carry on.

Contents

List of Figures

List of Tables

Acronyms

3GPP 3rd Generation Partnership Project

4G Fourth Generation

5G Fifth Generation

ACK Acknowledgment

ANRF Automatic Neighbor Relation Function

AP Access Point

API Application Programming Interface

AR Augmented Reality

AU Active User

BH Backhaul

BS Base Station

C-RAN Cloud-Radio Access Network

CA Carrier Aggregation

CAPEX Capital Expenses

CDF Cumulative Distribution Function

CLI Control Line Interface

CN Core Network

CoMP Coordinated MultiPoint

CONV Conventional

COTS Commercial-Off-The-Self

D2D Device-to-Device

DC Dual Connectivity

DIST Distance-based

DL Downlink

DRB Dedicated Radio Bearer

DU Disconnected User

E-UTRAN Evolved Universal Terrestrial Access

EESM Exponential Effective SINR Mapping

EMA Exponential Moving Average

eNB evolved Node B

EPC Evolved Packet Core

EPS Evolved Packet System

FDD Frequency Division Duplexing

GBR Guaranteed Bit Rate

GPRS General Packet Radio Service

GPS Global Positioning System

GTP GPRS Tunneling Protocol

HeNB Home eNB

HetNet Heterogeneous Network

HO Handover

HSPA High Speed Packet Access

HW Hardware

PHY Physical Layer

IoE Internet of Everything

IoT Internet of Things

KPI Key Performance Indicator

LA Load-aware

LTE Long Term Evolution

LTE-A Long Term Evolution-Advanced

M-MTC Massive Massive-Type Communication

M2M Machine-to-Machine

MAC Medium Access Control

MCS Modulation Coding Scheme

MCC Mobile Country Code

MCN Mobile Network Code

MEC Mobile Edge Computing

MME Mobile Management Entity

MIMO Multiple-Input Multiple-Output

MU Mobile User

mm-wave Milimeter Wave

NAS Non-Access Stratum

NGMN Next Generation Mobile Networks

OAI OpenAirInterface

OFDMA Orthogonal Frequency-Division Multiple Access

OPEX Operational Expenses

oaisim OAI emulator

P2P Point-to-Point

P-GW Packet Gateway

PCI Physical Cell Identity

PDCP Packet Data Convergence Protocol

PDF Probability Density Function

PRACH Physical Random Access Channel

PF Proportional Fairness

PRB Physical Resource Block

PHY Physical Layer

PS Processor Sharing

QoS Quality of Service

QCI QoS Class Identifier

RAB Radio Access Bearer

RACH Radio Access Channel

RAN Radio Access Network

RAT Radio Access Technology

REST Representational State Transfer

RLC Radio Link Control

RN Relay Nodes

RRC Radio Resource Control

RSRP Reference Signal Received Power

RSRQ Reference Signal Received Quality

RSS Received Signal Strength

S-GW Serving Gateway

SC Small Cell

SCTP Stream Control Transmission protocol

SDN Software-Defined Networking

SI Study Item

SIM Subscriber Identity Module

SINR Signal-to-Interference-plus-Noise Ratio

SFN System Frame Number

SN Sequence Number

SNR Signal-to-Noise Ratio

SOM Self-Organizing Map

SON Self-Organizing Network

SRB Signalling Radio Bearer

SU Static User

TA Tracking Area

TCP Transmission Control Protocol

TNL Transport Network Layer

TTT Time-to-Trigger

U2U User-to-User

U-MTC Ultra-Reliable Machine-Type Communication

UDN Ultra Dense Network

UDP User Datagram Protocol

UE User Equipment

UL Uplink

UMTS Universal Mobile Telecommunications System

UPF Utility Proportional Fairness

USRP Universal Software Radio Peripheral

V2X Vehicular-to-X

VNF Virtual Network Function

VoIP Voice over IP

VRB Virtual Resource Block

X2-AP X2 Application Protocol

Wi-Fi Wireless Fidelity

WiMAX Worldwide Interoperability for Microwave Access

Preface

Definitions, theorems, lemmas, propositions and examples share the same index within each chapter. The symbol □ stands for the end of proof of theorem, or lemma.

x	a variable
$\mathbf{x}$	a vector
$\mathbf{x}^T$	transpose of $\mathbf{x}$
$\mathbf{1}_N$	$N \times 1$ all-ones column vector
$\|\|\mathbf{x}\|\|$	the Euclidean distance of a vector $\mathbf{x}$
$\mathbb{N}$	the set of natural numbers
$\mathbb{R}$	the set of real numbers
$\mathbb{R}^+$	the set of positive real numbers
$\|A\|$	cardinality of a set A
$\mathbf{1}(\cdot)$	denotes the indicator function
$\min(a, b)$	the minimum value of a and b
$\mathcal{CN}(\mu, \sigma^2)$	denotes the complex Gaussian distribution of a random variable x with mean μ and variance σ^2
$\mathcal{N}(\mu, \sigma^2)$	denotes the Gaussian distribution of a random variable x with mean μ and variance σ^2
Log-$\mathcal{N}(\mu, \sigma^2)$	denotes the log-normal distribution of a random variable x with parameters μ and σ

Chapter 1

Introduction

1.1 5G Networks and Beyond

Societal changes throughout the world will have an impact on the way mobile and wireless communication systems are used in nowadays. A plethora of services as e-banking, e-health, augmented reality (AR), on-demand entertainment etc., begins to proliferate emerging industry and academia research consortia to define the 5G wireless access capabilities. The ubiquitous access to such services comes with the requirements of service continuity, seamless mobility, higher resource efficiency and low energy consumption. On the other hand, the avalanche effect of mobile and wireless traffic volume increment that is followed by the exponential growth in connected devices targeted not only to human-type communications that dominate the current networks but also to the machine-type ones. A total of 50 billion connected devices is predicted to be prevalent by 2020 [1, 2] with the upcoming technology of internet of things (IoT), massive machine-type communication (M-MTC) and ultra-reliable machine-type communication (U-MTC) promising to make our life and daily chores easier.

The future networks evolution has to support the coexistence of diverse applications with varied characteristics that impose a wide range of different requirements. To better reflect 5G requirements, we describe its disruptive capabilities based on the intended goals to achieve summarized by the following metrics: a) peak internet mobile (IM) terminal data rate ($\geq$ 10Gbps), b) connections density ($\geq$ 1M IM terminals/km^2), c) end-to-end latency ($\leq$ 5ms), d) guaranteed user data rate ($\geq$ 50Mbps) before channel coding, e) capability of human-oriented terminals ($\geq$ 20B), f) capability of IoT devices ($\geq$ 1T), g) mobility support at speed ($\geq$ 500km/h) for ground transportation, h) aggregated service reliability ($\geq$ 99.999%) and i) outdoor terminal location accuracy ($\leq$ 1m) [3]. By addressing the portraying sustainable development goals, several key performance indicators (KPIs) are presented by quantitative mathematical definitions to facilitate the technical solutions design and accomplish the new trend technologies demands [4]. In addition, non-quantitative criteria are related to the network management including software-based central architectures, simplified authentication, multi-tenancy and multi-RAT support, satellite communications and robust security and privacy to reduce the CAPEX/OPEX levels with current technologies.

The design principles to build the next-generation networks infrastructure host new types of services, new types of devices and different technologies spanning from the fronthaul to backhaul access radio access network (RAN). To this direction, 5G architecture will dramatically change

fostering to new radio paradigms. Recent advances towards pre-5G/LTE-Advanced Pro (LTE-A Pro) transition such small cells [5] lead to Ultra Dense Networks (UDNs) combined with massive multiple-input multiple-output (MIMO) and milimeter wave (mm-wave) radio mechanisms promise to increase capacity, reduce power emission and extend network coverage. Such environments are identified by a mixture of different entities, e.g., human-to-human, human-to-machine and machine-to-machine (M2M) communications. To conclude, operators can choose different strategies in network planning dominated by multiple technologies that enable new business models to support both legacy communication services as well as their future enhancements.

1.2 Challenges and Related work

Modern networks are characterized by different technologies interacting with varied type of entities. Such heterogeneity forms a new type of networks, the so-called heterogeneous networks (HetNets), including e.g., long term evolution (LTE) cellular base stations (BSs), wireless-fidelity (Wi-Fi) access points (APs), vehicular-to-x (V2X) and device-to-device (D2D) communications etc., as depicted in Fig. 1.1. The latter motivates to explore HetNets assymetrical characteristics aligned with the upcoming network architectures. To this direction, the thesis focuses on *mobility management* that is emerged to be revisited in LTE/LTE-A HetNets introducing new mechanisms beyond the current standardization schemes. Hence, this radical view of the network creates new radio interfaces ahead point-to-point (P2P) links schemes establishing simultaneously multiple connections in order to improve reliability. The new feature of introducing more than one radio interfaces leverages new ways of managing network resources. Tackling the problem of *resource management* under multi-connectivity, the thesis provides new solutions reviewing former techniques in the direction of next-generation mobile networks. To conclude, we turn our attention to such problems and discuss in detail the current system challenges as well as the design criteria beyond the state-of-the art proposed methodology.

1.2.1 Problem 1: Mobility management in LTE/LTE-A networks and beyond

Mobility is one of the key features of current and next generation cellular systems that enables the users to change seamlessly their point of attachment while using their data and voice services under dense small cell networks environment regimes. Critical issues are related to the deployment of new techniques based on spectrum availability, preliminary admission control and quality-of-service (QoS) criteria [6]. Future networks control mechanisms face to solve complex and multi-dimensional decision problems impacting users performance, e.g., throughput, service delay etc. during the handover procedure in mobility management.

Handover (HO) decision algorithms development for multiple-macrocell multiple-femtocell scenarios is even more prominent, since dense cell networks become an integral part of the 5G networks. Cell densification is emerged to cope with the tremendous growth of data coming from ubiquitous data hungry devices as mentioned before. Specifically, HetNets composed of different cells with different levels of power, i.e., different cell sizes, towards new network architectures adapted to different type of technologies. 3GPP Rel.9 introduces small cells (SC) as Home eNBs (HeNBs) to provide indoors coverage for closed subsriber groups with macro-cell coordination. Later, in 3GPP Rel.10, relay nodes (RNs) are presented as low-power SCs to enhance coverage, i.e., areas without fiber connections, as well as overcome bandwidth limitations that creates interference from the neighbor cells due to the reuse of the same bands. Current HetNets deploy

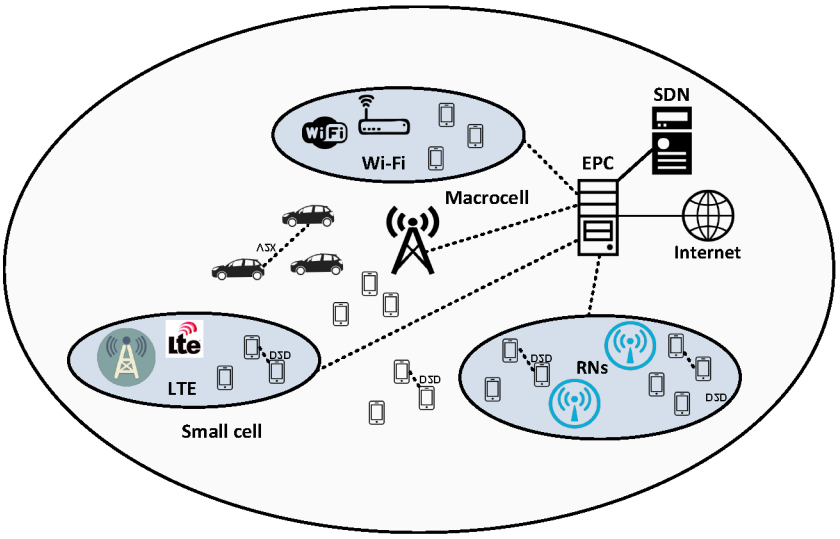


Figure 1.1: Heteregeneous networks towards 5G

a macro layer that supports high speed packet access (HSPA) with LTE overlay and SCs mainly to increase capacity and extend coverage. As a next step, future deployments based on such multi-layer and dense cell schemes shape the so-called UDNs. The latter will include macro setup upgrades with multi-carrier and carrier aggregation solutions while the hot zones will be served by SCs and WiFi APs [7]. Under those constraints, mobility management urges new types of HO decision algorithms that take into account the new architecture trends.

Many works have been done in HO mechanisms considering the overall network (i.e., network-based). A handover parameter self-optimization algorithm in long-term evolution (LTE) self-organising networks (SONs) [8] applies an operator policy based on different metrics (i.e., handover failure ratio, call dropping ratio and ping-pong HO ratio) for changing the hysteresis and the time-to-trigger (TTT) parameters and increases the system performance significantly. Furthermore, a parameter optimization for LTE handover using advanced self-organizing (feature) map (SOM) reduces the number of handover failures [9]. On the other hand, many works exist in HO decision algorithms performed by the user (i.e., user-based). The most commonly used are based on: i) received signal strength (RSS), ii) speed based, iii) interference-aware [10, 11]. Hence, these works do not consider hierarchical HetNets and user QoS metrics (e.g., user service delay). Software-defined networking (SDN) promises to overcome adversities in network design by centrally controlling network functions via an abstraction of the underlaying network (i.e., network virtualization). SDN architecture provides a central controller for the global network state that can affect the local control decisions of each individual radio element [12]. These mechanisms can also be applied on mobility management procedures in fifth generation (5G) mobile communications offering a holistic knowledge of the network.

1.2.2 Problem 2: Resource allocation under multi-connectivity in 5G networks

Multi-connectivity is one of the most promising techniques that can be used to achieve high performance improvements in 5G [13]. Moreover, with the arrival of Internet-of-Everything

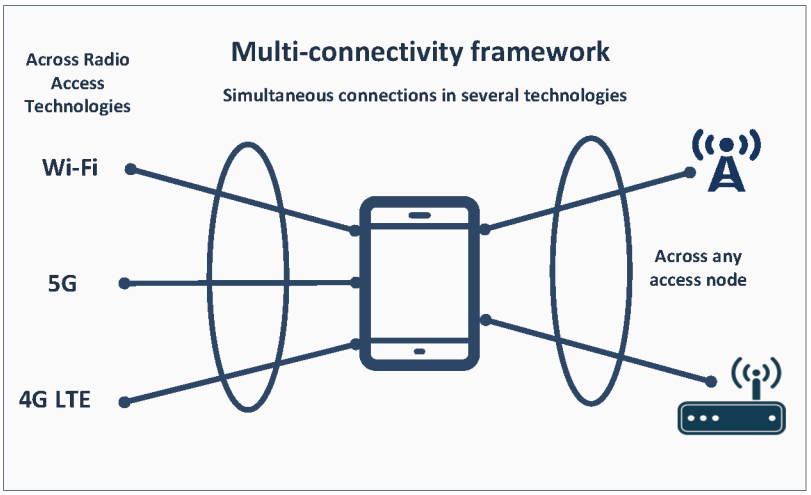


Figure 1.2: Multi-connectivity

(IoE), network traffic is tremendously increasing, thus the future networks needs not only to operate with multiple connections but also adhere to the devices QoS requirements. In 3GPP Rel.12 study item (SI) targeted to three main objectives: mobility robustness, signaling load reduction and per-user throughput enhancements [14]. In order to accomplish them, the idea of multi-connectivity was proposed where a user equipment (UE) is connected at the same time to multiple BSs as depicted in Fig. 1.2. This diversity of resources allows the UE to exploit in a smart way the multiple physical connections in order to attain the intended corresponding gains.

Dual-Connectivity (DC) in LTE: Multi-connectivity is not a new concept in mobile communication networks. LTE-advanced (LTE-A) Rel.12 introduced the so-called DC [14]. DC implies the UE connection to two base stations simultaneously and it can be seen as an extension of the carrier aggregation (CA) approach [15] where the transmission points are connected via a backhaul link.

Multi-Connectivity deployment in 5G: In terms of implementation, contrary the former network scheme, multi-connectivity in 5G can support a number of cells involved that is higher than two. The idea of multi-connectivity can incorporate connections across multiple heterogeneous RATs, e.g., LTE, 5G, WiFi etc., as well as across multiple radio interfaces of the same RAT.

From the network deployment perspective, multi-connectivity can run in many different deployment layers such as macro-cells as well as small cells. Three deployment scenarios are mainly taken into account: Scenario 1 (intra-freq macro and small cells are connected via non-ideal backhaul), Scenario 2 (inter-freq macro and small cells are connected via non-ideal backhaul) and Scenario 3 (intra/inter-freq small cells are connected via non-ideal backhaul) [16]. From the network architecture perspective, multi-connectivity offers balanced UL and DL operation and deploys independent 3GPP LTE protocal layer functions for different radio bearers [17]. The authors in [18] show that DC outperforms hard handover in terms of throughtput and latency reducing the signaling traffic per user. Finally, multi-connectivity not only provides simultaneous data transmission and reception, but it is also acting as a quick fail-over method when

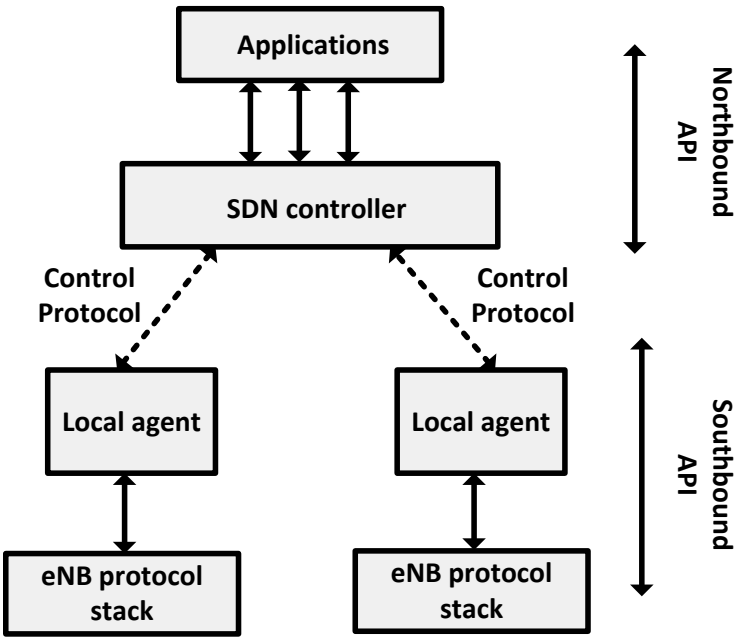


Figure 1.3: LTE with SDN control

connection is lost [19].

Hence, the importance of QoS turns our attention to network resource management towards 5G networks under the multi-connectivity regime. In general, many types of allocation have been proposed in literature related to former network technologies. The utility-based proportional fairness resource allocation criterion is proposed in [20] and achieves fairness in terms of the selected utility functions. Further, several works exploit utility functions as a measurement of the user application's QoS performance. In [21], different utility functions are used to formulate the utility proportional-fair optimization problem under good wireless channel conditions. Further, the authors in [22] deal with the power allocations of multi-class traffic in terms of four general types of utility functions: convex, concave, S-shape and inverse S-shaped. A time-variant utility function is introduced in [23] to minimize the delays for various foreground and background traffic demands. Recently, a few works appear to apply resource allocation with multi-connectivity support. The authors in [24] propose a multi-connectivity concept for a cloud radio access network and apply a simple proportional fair resource allocation showing performance results in terms of throughput and radio links failure. In [25], an approximation algorithm is presented to a user association that is optimal up-to an additive constant for the proportional fairness utility under the DC regime. Nevertheless, none of these works formally addresses utility-based proportional fairness schemes in a multi-connectivity wireless environment such as to retain users QoS.

To sum up, the thesis deals with the mobile and resource management towards 5G RANs by tackling the aforementioned problem challenges. If we refer to the 3GPP protocol stack [26], this type of management is related to the MAC and RRC layers for L2 network control

where the scheduling and handover processes mainly take place. To apply, such control can exploit the holistic knowledge of the network implying the existence of a network controller within a centralized architecture framework to acquire the information needed. The network controller can interact with the RAN by taking cross-layer decisions as depicted in Fig. 1.3. To this direction, SDN-based unified control-plane architecture retrieves and reconfigures real-time network control information for its data-plane use case and also consider Virtual Network Functions (VNFs) for the implementation of its components. The latter passes on a framework design that is open to high-layer application development, considers for easy associated services deployment and defines the relevant communication interfaces (northbound/southbound APIs). Specifically, a network application defines the requested policy to the SDN controller via a northbound API, e.g., REST API, the SDN controller communicates this policy to the local eNB agent via a southbound API, i.e., OpenFlow like control protocol (see FlexRAN control protocol [27]), and finally the policy is applied to each LTE protocol stack layer (in our case the MAC and RRC layer). Throughout this thesis, the process to be followed leads to centralized schemes implying the support of such systems for network management and monitoring of the underpin 5G networks by any operator.

1.3 Motivation and Contribution

Based on the previous section, the general requirements of 5G systems are progressively taking shape by raising many questions to the common study problems in the existing literature where there is room to improve various performance metrics and examine the design trade-offs.

1.3.1 Motivation

Most of the current studies on the problems presented and discussed in Section 1.2.1 (i) base on conventional assumptions not adapted to present networks complexity by nature, (ii) isolate network problems ignoring the holistic network view to offer a solution, (iii) neglect the practical aspects concerning design and implementation issues. In the following of this section, we point out the weaknesses of the state-of-the art works by emphasizing the need to come up with new sophisticated algorithms poised to evolve the cutting-edge algorithmic technologies that lead to the legacy communications systems modernization. Finally, Section 1.3.2 proposes new methods and techniques that can be applied to future standardization suggesting solutions that review older assumptions in network modelling and take into account the upcoming network architectures. In that sense, the goal of this thesis is to combine *theoretical and practical* insights so as to provide solutions and bring some answers to emerging network problems adopting a realistic approach. Specifically, the main motivation and incentives of this thesis coming from previous studies on mobility management and resource allocation are listed as follows:

Legacy handover performance study: Inspired by Problem 1 statement, as a first step, we focus on the legacy handover algorithm study for our better understanding of mobility management in LTE/LTE-A networks. In more detail, many works have been done comparing the S1 and X2 handover in terms of the EPC signaling load and the results prove that X2 handover can reduce EPC signaling load more than six times compared with S1 handover. X2 handover can be a sort of solution to decrease the load impact to the EPC and to increase the reliable inbound handover [28,29]. In addition, it reveals that the X2 handover triggering time is decreased with the increase on the eNB transmission power and vehicle speed using RSRP criterion on the

MATLAB platform [30]. Another work models the LTE handover scheme on an open source platform operated on the ns-3 platform; however, it does not compare the impact of different parameters on the handover delay [31]. Finally, this work uses the ns-3 platform to compare the measured RSRP and RSRQ level under different parameters: vehicle velocity, eNB transmission power and distance between UE and eNB. However, there is no comparison on the handover latency on different parameters [32]. An initial study on X2 HO latency can be found in [33] based on Helsinki's metropolitan area mobile network measurements. Nevertheless, changing system parameters in such cases is not feasible, since commercial systems are not open to experimenters community. Towards this direction, after implementing the X2 handover (HO) mechanism in OpenAirInterface (OAI) (open-source) platform, HO performance need to be examined in terms of network delay as well as the impact of system parameters on triggering the HO process.

Heterogeneity in next generation networks: Beyond legacy approach when heterogeneity comes to the picture, conventionalities related to handover decision algorithms need to be revised as discussed in the following. Denser deployments in HetNets experience high spatio-temporal load variations and thus require more advanced HO algorithms. In addition, power based algorithms (e.g., RSS-based) proposed in the literature cannot be applied, due to asymmetrical transmission powers among macrocells and SCs in the literature. Especially, in such environments, a UE usually remains connected to a macrocell that offers high transmission power (and is potentially overloaded due to its high number of concurrent users), while is placed close to one or more underloaded SCs. According to the latter, authors in [34] propose to scale down macrocell RSS with an appropriate factor to force connection to a SC. This factor is selected optimally based on the maximization of the SC assignment probability. Therefore, it was left as future work to examine policies that consider user throughput, delay or other users requirements etc. To this end, the problem of HO shall be revisited, in the context of future HetNets deployments, considering users QoS under transmission power asymmetry regimes beyond the conventional methods.

Software-defined networking (SDN): SDN promises to overcome network design adversities by centrally controlling network functions via an abstraction of the underlaying network. This is achieved by separating the mechanisms that are in charge of making the decisions about where the data are sent (control plane) from the mechanisms that transfer the data through the network (data plane). Software-defined radio access network (SoftRAN) proposes a centralized design, where a centralized controller is responsible for the global network state while each individual radio element handles local control decisions [12]. In [35], authors mention that existing expensive equipment and complex controller plane protocols in cellular wireless networks sparkle interest on research for future SDN architectures including new challenges in controller applications. Finally, Softcell architecture in [36] uses a centralized controller for achieving scalability and minimizing core network state via packet classification at the access edge and multi-dimensional aggregation respectively.

New envisioned SDN architectures claim to reduce energy UE consumption, signaling communication overhead and handover delays [37]. Further, distributed algorithmic approaches used in SONs do not scale well in large-scale networks with a large number of base stations and a centralized approach may meet better the global network needs. A centralized approach is referred in [38], where an SDN framework reduces real-time video freeze events via fast switching among the access points (APs) by using a system controller for handover management. Another SDN controller is used in OpenRoads [39] for lossless and seamless handover between WiFi-WiMAX in order to provide robustness of the channel via simultaneous transmission of packets over mul-

tiple interfaces. In that sense, SDN architecture can be considered as a solution to manage the handover decisions resulting in a centralized network coordination. The latter calls for proposing new SDN-based implementations that support HO functionalities combined with an algorithm's sophistication to better manage mobility in next generation networks. Emerging control frameworks such as FlexRAN [27] with support for both north- and south-bound protocols can be exploited for mobility management via RAN programmability.

Traffic differentiation: Based on Problem 2 statement, we focus on new resource allocation schemes under the multi-connectivity regime with different types of traffic. Most of the current works [40, 41] usually focus on one type of traffic while current technologies urge for traffic differentiation in next generation mobile networks. That implies joint UL/DL consideration for any link dependencies as well as different constraints arising from link parameters asymmetries adapted to multiple simultaneous connections traffic flows. The type of considered traffic can be generally categorized as follows:

- ***UL/DL*:** In that type of traffic, we consider uplink flows with direction from the UE to the BS and downlink flows with direction from the BS to the UE, i.e., user-to-network (UL) and network-to-user (DL) traffic. For instance, applications related to social networking (e.g., Facebook), augmented reality (e.g., Google Translate) and video-on-demand/video dissemination/content recommendation (e.g., Netflix) belong to this traffic category. Due to asymmetry in transmit powers of UEs and BSs, power control needs to be applied regarding the UL. Further, such traffic can be routed via the backhaul network infrastructure or locally in case a cache-assisted mobile edge computing (MEC) system [42, 43] is supported by intelligent BSs capabilities.

- ***U2U*:** In that type of traffic, we consider flows with direction from one UE to another UE, i.e., user-to-user (U2U) traffic. For instance, applications related to peer-to-peer (e.g., Hangout) and public safety (e.g., bSafe) communication networks belong to this type of traffic. In such scenarios, the traffic routing can be distinguished as: a) 1 hop: a UE sends traffic direct to another UE via D2D communication, b) 2 hop: a UE sends traffic to another UE via local routing by a common BS connection (no backhaul routing), c) 4 hop: a UE sends traffic to another UE via Evolved Packet Core (EPC)-based routing in the backhaul.

Consequently, new multi-connectivity scheduling schemes need to be created based on traffic differentiation across multiple connections of one or more RATs such that to improve user performance and network resource utilization towards 5G era.

Backhaul capacity limitations: As stressed earlier, multi-connectivity can greatly improve the system performance and fits well in various traffic flows whereas introduces several different challenges. Mobile networks appear to tremendously create a traffic data rate increment over the air-interface and it is argued that the major performance bottleneck will be on the *backhaul* networks infrastructure [44]. Consequently, some recent works focus on that direction and attempt to jointly consider radio access and backhaul capacity limitations. These are mostly taken into account in various schemes related to user association [45], energy efficiency resource allocation [46], power allocation [47], relay cooperative transmissions [48], caching-aware user association [49], load-balancing in HetNet [50], cell range extension [51], or multi-cell cooperative processing [52]. Compared to the aforementioned studies, potential backhaul network limitations among the end-users and the major network is crucial to be introduced, leading to the need

of more sophisticated resource allocation schemes under the multi-connectivity regime. Those schemes are significantly important since they directly impact both the overall network system performance and the received user QoS, especially when the backhaul capacity limitations come into the picture.

Spatio-temporal channel variability: In wireless mobile networks, the channel conditions are varying in time due to small-scale fading effects. In that sense, different users experience different channel conditions at a given time. Exploiting such diversity, scheduling mechanisms, the so-called *opportunistic*, benefit from the favorable channel conditions in assigning time and frequency slots to the mobile users. To this direction, opportunistic scheduling can be combined with multi-connectivity exploiting multiple connections among the users, i.e., different channel conditions based on cell diversity. Many opportunistic schemes have been proposed tackling the issue of scheduling from different aspects [53]. For instance, there are proposals that mainly focus on improving network throughput [54] while others aim to provide fairness under users QoS constraints [55]. Based on such schemes, new opportunistic scheduling techniques need to be examined under the multi-connectivity regime with network backhaul limitations.

1.3.2 Contributions and Outline

The goal of this thesis is to answer the questions raised by revising Problem 1 and 2 discussed in Section 1.2. In that sense, we provide realistic assumptions applied on the proposed system models to better reflect the network trends of novel architectures towards 5G deployed networks. The current manuscript can be divided into two parts: a) mobility management (Chapter 2 and Chapter 3) and b) resource allocation (Chapter 4, Chapter 5 and Chapter 6). Both parts cover the practical aspects presenting the implementation difficulties and lesson learnt to any practitioner as well as the theoretical approaches emphasizing the challenges of mathematical analysis to any theorist. Moreover, the analytical tools lie in queuing, probability, convex, non-convex and combinatorial optimization theory leading to results that offer the design insights and reveal the trade-offs involved. Specifically, the chapters of the thesis and the main contributions in each one of them, are organized as following:

Chapter 2-X2 Handover in LTE/LTE-A: In this chapter, we analyze the performance of the X2 handover from the user perspective. Furthermore, the impact of the different parameters on the handover decision algorithm is investigated. Presented results, obtained from the OpenAirInterface LTE/LTE-A emulation platform, demonstrate that main delay bottleneck resides in the uplink synchronization of the UE to the target eNB. The work related to this chapter is:

- K. Alexandris, N. Nikaein, R. Knopp, and C. Bonnet, "Analyzing X2 handover in LTE/LTE-A," in Wireless Networks: Measurements and Experimentation," WINMEE 2016, May 9, 2016, Arizona State University, Tempe, Arizona, USA.

Chapter 3-Load-aware Handover Decision Algorithm in Next-generation HetNets: In this chapter, we focus on new algorithms that take into account also the network perspective, e.g., cell load. Specifically, a load-aware algorithm is proposed considering the service delay that a user experiences from the network. In addition, an implementable framework based on Software Defined Networking (SDN) architecture is sketched to support the algorithm. The proposed algorithm is compared with the 3GPP legacy (RSS-based) one presented in Chapter 2 and a distance-based one. Extracted cell assignment probability and user service delay performance results show that the load-aware approach outperforms both of them. The work related to this chapter is:

- K. Alexandris, N. Sapountzis, N. Nikaein, and T. Spyropoulos, "Load-aware handover decision algorithm in next-generation HetNets," in IEEE WCNC 2016, 3-6 April 2016, Doha, Qatar.

Chapter 4-Utility-Based Resource Allocation under Multi-Connectivity in LTE: In this chapter, we examine a resource allocation problem under multi-connectivity in an evolved LTE network and propose a utility proportional fairness (UPF) resource allocation that supports QoS in terms of requested rates. We evaluate the proposed policy with the proportional fairness (PF) resource allocation through extensive simulations and characterize performance gain from both the user and network perspectives under different conditions. The work related to this chapter is:

- K. Alexandris, C.-Y. Chang, K. Katsalis, N. Nikaein, and T. Spyropoulos, "Utility-Based Resource Allocation under Multi-Connectivity in Evolved LTE," in IEEE VTC 2017, Sep 24-27, 2017, Toronto, Canada.

Chapter 5-Multi-Connectivity Resource Allocation with Limited Backhaul Capacity in Evolved LTE: In this chapter, we examine a UPF resource allocation problem under multi-connectivity in evolved LTE and propose a utility considering the users QoS with backhaul capacity limitations compared to Chapter 4 where only the air interface constraints appear. The proposed policy is compared with PF resource allocation through extensive simulations. Presented results show that multi-connectivity outperforms single connectivity in terms of network aggregated rate and users QoS satisfaction in different network case studies, i.e., empty and loaded cell scenarios with fixed and variable backhaul capacity. The work related to this chapter is:

- K. Alexandris, C.-Y. Chang, N. Nikaein, and T. Spyropoulos, "Multi-Connectivity Resource Allocation with Limited Backhaul Capacity in Evolved LTE," in IEEE WCNC 2018, April 15-18, 2018, Barcelona, Spain.

Chapter 6-Utility-based Opportunistic Scheduling under Multi-Connectivity with Limited Backhaul Capacity: In this chapter, we revisit the scenarios discussed in Chapter 5 exploiting the impact of spatiotemporal channel variability (large- and small-scale fading) via opportunistic scheduling. Specifically, we examine an opportunistic resource allocation problem under multi-connectivity and limited backhaul capacity in evolved LTE using the two aforementioned utility functions: PF and UPF. We then propose an efficient algorithm to deal with the formulated NP-hard problem. Such algorithm is guaranteed to produce a solution with an explicit bound toward the optimal solution. Finally, the simulation results justify that using UPF utility function with opportunistic scheduling in multi-connectivity can better satisfy QoS requirements. The work related to this chapter is:

- K. Alexandris, C.-Y. Chang, N. Nikaein, and T. Spyropoulos, "Utility-based Opportunistic Scheduling under Multi-Connectivity with Limited Backhaul Capacity," in IEEE Wireless Communications Letters, 2018, submitted, under review.

Chapter 2

X2 Handover in LTE/LTE-A

2.1 Introduction

Mobile data continuous growth emerges efficient technologies to satisfy the required quality of service (QoS) of the new services. Network mobile users force to change seamlessly their point of attachments during their service time. Handover in Long Term Evolution (LTE), as in previous generation of cellular systems, is a procedure to transfer a user equipment (UE) and its context from a source evolved NodeB (eNB) to a target eNB. It requires efficient handover decision algorithms in order to optimize both UE and network performance and quality. Handover is a "UE-assisted network-controlled" process in that the measurement is reported by UE, and the decision is made by the network, i.e. eNBs and/or Mobility Management Entity (MME).

As explicitly discussed in Chapter 1, X2 handover can be used to reduce EPC signaling load ($\sim$ six times) and ensure system's validity and reliability [28, 29]. In this chapter, we revisit X2 handover in order to analyze the impact of different system parameters in terms of latency contrary to other works that focus on measured RSRP/RSRQ level [32]. Finally, experiments to study the system performance are carried out in OpenAirInterface (OAI) [56] adding value to this work by using a real world LTE deployment compared to other works that are based on the MATLAB or ns3 platforms [30, 31]. In more detail, this chapter will focus on X2 handover in LTE/LTE-A that happens between eNBs [57]. In all the cases, both source and target eNBs are connected to the same MME and are located in the same tracking area (TA). The measurement cases cover the handover between two cells supporting the X2 interface between the eNBs. The chapter contributions can be summarized as follows:

- We discuss and sum up the X2 handover protocol as well as its own characteristics, parameters and further extensions towards 5G network technologies.
- We investigate the impact of the handover parameters such as frequency offsets and hysteresis that are commonly used in the handover decision algorithms criteria.
- We analyze and characterize the performance of X2 in terms of delay using the OAI emulation platform focusing on the Evolved Universal Terrestrial Access Network (E-UTRAN).
- We perform real world OAI X2 handover RF testbed measurements and we compare them with the OAI emulated results.

The remainder of the chapter is organized as follows. Section 2.2 introduces the system description and modeling approach. Section 2.3 presents the system implementation. Section 2.4 includes the emulated system evaluation. Section 2.5 shows real-world RF testbed measurements. Section 2.6 offers a brief discussion. Finally, Section 2.7 provides concluding remarks.

2.2 X2 Handover protocols

2.2.1 X2 Application Protocol

Handover architecture, deployment and implementation has entirely changed compared to the legacy 3GPP technologies. Universal Mobile Telecommunications System (UMTS) technology supported the Radio Network Controller (RNC), a network component that was in charge of handling any handover signaling capability. In LTE Evolved Packet System (EPS), RNC has been removed and the intelligence is kept in the eNB side that is responsible for handover. A connection has to be established among eNBs in order to signal with each others for handovering. This is managed through X2 interface, using X2 Application Protocol (X2-AP).

X2 interface can be established between one eNB and its neighbors in order to exchange the intended information. Hence, fully mesh topology is not mandated contrary to S1 interface where a star topology is used. Moreover, the protocol structure over X2 interface contains both the control and the data plane protocol stack that is the same as over the S1 interface as depicted in Fig. 2.1. The X2 topology as well as the X2-AP structure provide advantages related to the data forwarding operation as will be discussed later. In case X2 interface is not configured or the connection is blocked; handover can be performed via MME using S1 interface. The initialization of X2 interface starts with the neighbor identification, i.e., based on configuration or Automatic Neighbor Relation Function (ANRF) process. Subsequently, the Transport Network Layer (TNL) is set using the TNL address of the neighbor. Once the TNL is established, the X2 setup procedure is ready to run to exchange application level data needed for two eNBs in order to operate correctly via X2 interface. Specifically, the source eNB (i.e., the initiating eNB in which the UE is attached) sends the X2 Setup Request to the target eNodeB (i.e., the candidate eNB in which the UE intends to handover). The target eNB replies with the X2 Setup Response.

The X2 handover key features are [58]:

- The whole procedure is directly performed between the two eNBs.
- MME is involved only after the handover procedure is completed for the path switch procedure contrary to the S1 handover that is MME assisted decreasing the delay and the network signaling overhead.
- The release of source eNB resources is triggered via the target eNB at the end of the path switch procedure.

The X2 procedure can be described in five steps, as shown in Fig. 2.2:

1. **Before Handover:** UE is attached to the source eNB. The Dedicated Radio Bearers (DRBs) and Signalling Radio Bearers (SRBs) are established and UL/DL traffic is transmitted between the source eNB and the UE. The UE remains in the Radio Resource

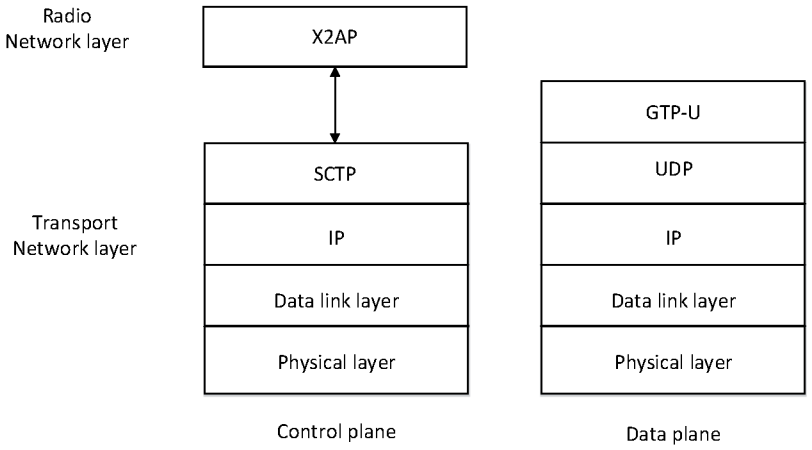


Figure 2.1: eNB X2 protocol stack for control-plane and data-plane

Control (RRC)-Connected, EMM-Registered, and ECM-connected states with respect to the source eNB, and keeps all the resources allocated by E-UTRAN and EPC.[1]

2. **Handover Preparation:** UE sends the periodical measurement report to the source eNB; this report contains information about the neighboring cells. The source eNB triggers the handover (i.e., eNB decides that the handover is necessary) based on the reported measurement results, i.e., A1-A5/B1,B2 event (see [59]) and chooses the best reported target cell by the UE. Then, the source eNB sends a X2 handover request to the target eNB. This message contains the information needed to perform the handover (e.g., UE context information, Radio Access Bearer (RAB) context, Target Cell ID). Considering the QoS in the RAB context, the target eNB performs call admission control and if it is able to provide the requested resources for the new UE, it sends a handover (HO) request acknowledgment (ACK) to the source through the X2 direct tunnel setup (i.e., handover is eNB accepted). The source eNB receives this message that includes the configuration of the GTP-U tunnels per radio access data radio bearer as well as the RRC Connection Reconfiguration message in a transparent container that the source eNB has to forward to the UE. In the RRC message, L1/L2 parameters are provided to the UE in order to be synchronized with the target eNB. Finally, the source eNB sends the HO command message that encloses the RRC Connection Reconfiguration message to the UE. If the target eNB cannot accept the HO request (due to load or the required setup), it responds to the source eNB with an X2 failure message. During this step, the UE states remain unchanged.

3. **Handover execution:** UE receives the RRC Connection Reconfiguration message and transits to the RRC idle state triggering the detachment from the source eNB. The source eNB sends the Sequence Number (SN) status transfer message that contains the Packet

[1]RRC represents the state of a UE with respect to the eNB. EMM, the EPS Mobility Management, represents the state of a UE with respect to the mobility management entity (MME). ECM, EPS Connection Management (ECM), represents a combination of RRC connection state between UE and eNB and S1 state between eNB and MME.

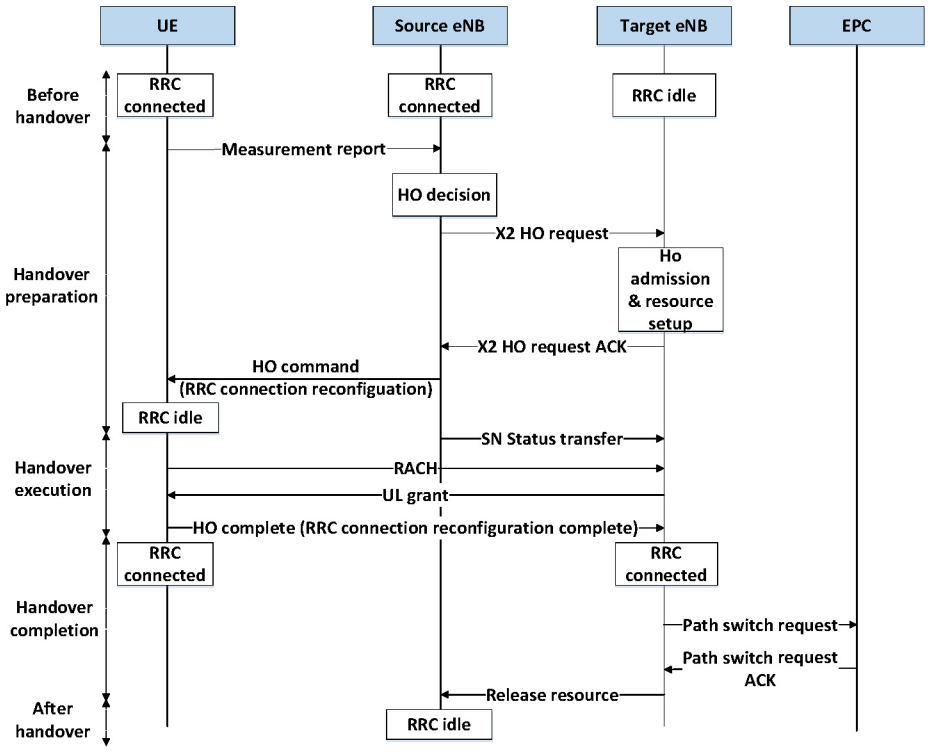


Figure 2.2: X2 handover procedure

Data Convergence Protocol (PDCP) sequence numbers to the target eNB through X2 interface. For UL the first missing data unit is included and for DL the next sequence number to be allocated. Then, UE is synchronized with the target based on the given parameters and send the HO Confirm message that encloses the RRC Connection Reconfiguration Complete to acknowledge the successful handover to the target eNB. As a result, the UE transits to the RRC connected state with respect to the target eNB. Concerning the UE synchronization, if a dedicated random access preamble has been received in the RRC Connection Reconfiguration message, the UE does not need to perform the random access procedure, i.e., contention free Random Access Channel (RACH) process. If this is not the case, the UE performs the normal random access procedure described in [60] (contention-based RACH).

4. **Handover Completion:** The target eNB receives the RRC Connection Reconfiguration Complete message and the path switch procedure is initiated between the target eNB and the MME/S-GW. The target eNB starts to forward all the packets received from the X2 interface to the UE before any new ones coming from the Serving Gateway (S-GW) (i.e., target eNB receives the end-marker from the old path switch and starts transmitting packets from the new path switch). Afterwards, the source eNB UE context is released via receiving UE release context message from the target eNB . Finally, the S1 bearer that was initially established between source eNB and UE is also released.

5. **After Handover:** UE is attached to the target eNB. The DRB and SRB are established

and UL/DL traffic is transmitted as in the initial step.

Mobility over X2 can be differentiated in four different modes according to the RAB QoS Class Identifier (QCI). The source eNB has to select based on the UE QoS requirements received (e.g., Guaranteed Bit Rate (GBR)/non-GBR traffic etc.). These modes are described as follows (see also Fig. 2.1):

- **Control plane:** Only Stream Control Transmission Protocol (SCTP) connection is established among the two eNBs for control plane messaging and no data forwarding via X2 interface is supported. In that case, all the packets that is intended to be transmitted through the S1 path or are PDCP processed (i.e., buffered locally, but not yet acknowledged by the UE).
- **DL data plane:** General Packet Radio Service (GPRS) Tunneling Protocol (GTP) tunnels will be established for downlink data forwarding on per radio access bearer. The X2 request message that is sent by the source eNB proposes the GTP tunnel establishment; then the tunnel endpoint is included in the X2 request ACK message if the establishment is accepted by the target eNB. Thus, the source eNB can start the packet forwarding process in parallel with the HO command transmission to the UE. This type of data forwarding includes packets arriving over the source S1 path and is known as "seamless handover". As an enhancement, packets that are PDCP processed can also be forwarded (PDCP SN is included in the GTP extension header). The aforementioned data forwarding is referred as "lossless handover", since there is no packet loss.
- **UL data plane:** Uplink forwarding can be similarly handled by taking into account the traffic coming from the UE side that is PDCP buffered, non-acknowledged by the source eNB and consequently non-forwarded through the S1 path. This mode is known as "selective retransmission", since the UE can be informed by the target eNB for not re-transmitting those packets accelerating the uplink re-transmission.
- **DL & UL data plane:** A combination of the above modes can be also performed decreasing the overall delay. Accompanied with the control plane messaging assures the overall packet transmission both for DL/UL retaining the handover procedure seamless to the UE side.

In general, X2 handover can be initiated by the eNB for several reasons:

1. **Quality-based handover:** The indicated QoS levels included in the measurement report by the UE are too low and the UE needs to switch to another eNB for enhancing its QoS metrics.
2. **Coverage-based handover:** UE is moving from the one cell to another. In general, it could be intra-LTE or from LTE to UMTS or Global System for Mobile communication (GSM), when the UE moves to an area that is not LTE-covered, i.e., inter-Radio Access Technology (inter-RAT).
3. **Load-based handover:** This is an optimization case concerning the load among different eNBs. The required information is transferred through the X2 load indication message. Based on the purpose served, the case falls into two categories:

3.1. **Load balancing:** This category handles the load imbalance management between two neighboring cells by taking into account the overall system capacity. The frequency exchange of load information is low (i.e., in the order of seconds).

3.2. **Interference coordination:** This category elaborates Radio Resource Management (RRM) processes optimization such as interference coordination. Using this information, the target eNB can decide its scheduling policy based on its interference sensitivity. The frequency exchange of load information is high (i.e., in the order of miliseconds).

2.2.2 Handover Criteria and Parameterization

In principle the LTE network setup considers for the deployment of eNBs in hexagonal topology. Let $\mathcal{B}$ denote the number of deployed eNBs and let $r_i^{\mathrm{dBm}}[k]$ denote the Reference Signal Received Power (RSRP) from each base station (BS) $i \in \mathcal{I} = \{1, \ldots, \mathcal{B}\}$ at time k [2] in dBm scale. Averaging is performed by an Exponential Moving Average (EMA) filter, i.e., low-pass filter, for smoothing any RSRP abrupt variations and is applied in the radio resource control (RRC) layer [59] (i.e., L3 filtering). High frequency fluctuations are filtered out and can be neglected. The filtered signal is expressed in dBm as follows:

$$\bar{r}_i^{\mathrm{dBm}}[k] \triangleq (1-\alpha)\bar{r}_i^{\mathrm{dBm}}[k-1] + \alpha r_i^{\mathrm{dBm}}[k], \tag{2.1}$$

where $\alpha \triangleq 2^{-q/4}$ and $q \in \mathcal{F}$ [3].

Handover is performed using a set of handover parameters. Here, we refer to their definitions as well as the fields they belong to in the corresponding RRC layer structures, i.e., ReportConfigEUTRA, MeasObjectEUTRA, QuantityConfigEUTRA, see also [59]. Specifically, the parameters that could be adjusted are:

- Time to trigger (ttt): Time during which specific criteria for the event needs to be met in order to trigger a measurement report (time-to-trigger as defined in ReportConfigEUTRA).
- Hysteresis (hys): the hysteresis parameter for this event (i.e., hysteresis as defined in ReportConfigEUTRA).
- OFN (ofn): the frequency specific offset of the neighbor cell frequency (i.e., offsetFreq as defined in MeasObjectEUTRA).
- OCN (ocn): the cell specific offset of the neighbor cell (i.e., cellIndividualOffset corresponding to the frequency of the neighbor cell as defined in MeasObjectEUTRA).
- OFS (ofs): the frequency specific offset of the serving cell frequency (i.e., offsetFreq as defined in MeasObjectEUTRA).
- OCS (ocs): the cell specific offset of the serving cell (i.e., cellIndividualOffset corresponding to the serving frequency as defined in MeasObjectEUTRA).

[2] The time k corresponds to the discretization of the continuous time t sampling at kT_s intervals, where T_s stands for the measurement sampling period.

[3] The set of integers $\mathcal{F}$ is defined in [59, 10.3.7.9]

- OFF (off): the offset parameter for this event (i.e., a3-offset as defined within ReportConfigEUTRA for this event).
- L3 Filtering coefficient RSRP/Reference Signal Received Quality (RSRQ) (q): Parameter for the EMA filter as defined in Eq. (2.1) (i.e., this parameter is defined within QuantityConfigEUTRA).

A well-known handover criterion, commonly used in conventional HO decision algorithms for mobile communication systems (also applied in 3GPP LTE), is based on RSRPs comparison method in which hysteresis and handover offsets are included. Specifically, we focus on A3 event and its condition that is used as a criterion for the cell selection. The criterion is expressed as follows:

$$\bar{r}_n^{\mathrm{dBm}}[k] + ofn + ocn > \bar{r}_s^{\mathrm{dBm}}[k] + ofs + ocs + hys + off, \tag{2.2}$$

where $s \in \mathcal{I}$ and denotes the serving cell, $n \in \mathcal{I} - s$ and denotes the neighbor cells. Finally, the handover parameters that are included in Eq. (2.2) are defined as described above.

The above inequality is interpreted as follows: when the RSRP of a neighbor cell (sum of the neighbor's RSRP and offsets, $\bar{r}_n^{\mathrm{dBm}}[k] + ofn + ocn$) becomes greater than that of the RSRP of the serving cell (sum of signal strength and offset, $\bar{r}_s^{\mathrm{dBm}}[k] + ofs + ocs$) and the difference is greater than the value of off (referred also as a3-offset), Event A3 is triggered and the UE reports the measurement results to the eNB. Hysteresis (hys) indicates the value of a handover margin between the source and the target cell. Finally, the inequality can be compressed as:

$$\bar{r}_n^{\mathrm{dBm}}[k] + S > \bar{r}_s^{\mathrm{dBm}}[k], \tag{2.3}$$

where $S = ofn + ocn - ofs - ocs - hys - off$. The S can be determined as a sum of the offsets including all the offset impacts in triggering the handover condition.

Other representative handover algorithms include not only Received Signal Strength (RSS) criteria; a brief description of the non-RSS HO algorithms is given as follows:

- **Interference-based:** Interference-aware handover decision algorithms enable the shifting to femtocell communication paradigm in HetNets, where co-tier and cross-tier interference is taking into account based on interference level at the cell sites or Received Signal Quality (RSQ) at the UEs [11].
- **Speed-based:** Speed handover decision algorithms typically compare the UE speed with specific thresholds to mitigate the HO probability for high speed users (i.e., fast handover case) decreasing the overall handover signaling cost. Such algorithms can be combined with load/traffic-type criteria that are discussed below [10].
- **Load-based:** To this direction, load-aware handover decision algorithms can be developed considering the service delay that a user experiences from the network. In addition, an implementable framework based on Software Defined Networking (SDN) architecture can be included to support the algorithm, as suggested to be a key enabler for the realization of 5G networks. This approach overcomes the shortcomings created by only considering Received Signal Strength (RSS) criteria in HO decision for HetNets [61] and is thoroughly analyzed in Chapter 3.

2.2.3 Handover Delay Analysis

Handover delay can be classified into two different main categories: the *protocol delay* that captures the processing time and handover signaling delay and the *transport delay* that captures the transmission time through the physical medium of the X2 link (wired or wireless). The average delay budget of a handover can be defined as:

$$Delay_{\mathrm{HO}} = T_{\mathrm{Before\_HO}} + T_{\mathrm{HO\_Preparation}} + \underbrace{T_{\mathrm{HO\_Execution}} + T_{\mathrm{HO\_Completion}} + T_{\mathrm{Margin}}}_{\text{detach time} \leq 65\text{ms}} \tag{2.4}$$

where $T_{\mathrm{Before\_Ho}}$ represents the time required to search and identify the unknown target cell identity. This is applicable only to the network-triggered handover (e.g., load-balancing), otherwise, it is 0. $T_{\mathrm{HO\_Preparation}}$ is the UE transition time from RRC connected state to RRC idle state where the RRC Connection Reconfiguration message is received from the source eNB (see Fig. 2.2–HO Preparation phase). This delay includes the X2-AP processing and transport. It is noted that the UE can still send and receive traffic before getting in RRC idle mode. In addition, 3GPP has set requirements for the length of the detach time observed by the UE [33]. $T_{\mathrm{HO\_Execution}}$ represents the time to acquire the random access (contention-free or contention-based) and receiving an uplink resource grant for sending the RRC Connection Reconfiguration complete message (see Fig. 2.2–HO execution phase) and it is set to 45 ms for contention-based random access with very small probability of collision. In case of contention-free random access, the C-RNTI generated by the target eNB is sent to the source eNB via the X2 HO request ACK and finally received by the UE as part of MobilityControlInfo field included in the RRCConnectionReconfiguration message. $T_{\mathrm{HO\_Completion}}$ delay is zero for the UE as the UE is already in the RRC connected state with respect to the target eNB (see Fig. 2.2–HO completion phase). On the target eNB side, this delay is non-zero and holds for the time interval till the reception of RRC Connection Reconfiguration complete message. T_{Margin} is the implementation-dependent margin time upper bounded to 20 ms. Thus, the length of the detach time is estimated to 65 ms [62].

The latter depends on the deployment scenario, for instance in a cloud-RAN (C-RAN) [63], eNBs may share a common memory space, which in turn simplifies the X2 messaging. Thus, the transport delay becomes negligible, since practically there is no physical link presence.

2.3 X2 Handover experimentation setup

OpenAirInterface (OAI) [56] is an open experimentation and prototyping platform initially developed by the Communication Systems Department at EURECOM in order to stress innovation targeting on cellular technologies such as LTE/LTE-A. OpenAirInterface comprises a highly optimized C implementation all of the elements of the 3GPP LTE Rel.8.6 and a subset of Rel.10 protocol stack for UE and eNB (PHY, MAC, RLC, RRC, PDCP, NAS driver) as well as new features beyond Rel.12 towards 5G as new radio (NR), masssive MIMO, C-RAN, SDN etc. To support real-world LTE RF experimentation, OAI provides software radio frontend based on the ExpressMIMO2 PCI Express (PCIe) board [4]. Besides ExpressMIMO2, OAI now supports the UHD interface on recent Commercial-Off-The-Shelf (COTS) USRP PC-hosted software radio platforms such as Ettus USRP B210/X300, Nuand BladeRF and Lime Microsystems LimeSDR.

[4] `http://openairinterface.eurecom.fr/expressmimo2`

Apart from real-time operation of the software modem on a hardware target, the full protocol stack can be run in emulation.

OpenAirInterface emulation environment (oaisim) [64] allows for virtualization of network nodes within physical machines and distributed deployment on wired Ethernet networks. Nodes in the network communicate via direct-memory transfer when they are part of the same physical machine and via multicast IP over Ethernet when they are in different machines. One way to emulate the wireless medium behavior is the full PHY mode that generates real I/Q samples after modulation. Specifically, it performs the convolution of these samples with a synthetic channel to simulate the influence of the RF chains and propagation channel on the signal, instead of sending them to a RF card. The resulting samples are given to the demodulator of the receiving node. Another way provided in OAI is the PHY abstraction mode that supports both Exponential effective SINR mapping (EESM) and Mutual Information based SINR mapping (MIESM) abstraction techniques [65]. PHY layer includes power and resource allocation to a specific UE, number of spatial layers, modulation (UL single FDD, DL OFDMA) and different coding schemes (MCS), TDD and FDD configuration (5, 10, 20 MHz), UL/DL LTE channels, transmission modes, HARQ support (UL/DL) and mainly channel chracteristics, i.e., path loss, shadowing and stochastic small scale fading (frequency flat/selective channels) parameters. The remainder of the protocol stack holds the same as in a real full-RF LTE implementation.

OpenAirInterface required parameters for large scale system simulations are highly dynamic and are provided by the specific OAI simulation tools. These tools include openair traffic generator, (e.g., TCP, UDP, VoIP traffic) and openair mobility generator (e.g., trace-based, random walk mobility patterns). The handover experiment is conducted based on the OpenAirInterface (OAI) built-in emulation platform [56, 66]. For the PHY layer we used the "full PHY mode" as defined above and the full-LTE protocol stack supporting the handover procedure as well as the mobility and traffic generator tools. Finally, a brief description of the network components and the network topology of our experiment is given below based on the OAI tools discussed in this section.

We conduct several experiments based on the OpenAirInterface (OAI) built-in emulation platform [56]. This platform implements standard compliant LTE UE and eNB protocol stacks spanning all the protocol layers. As mentioned before, we make use of the "full PHY mode" that generates real I/Q samples after encoding and modulation, but instead of sending them to a radio front-end to produce a RF signal as in a real system, samples are convolved with a synthetic channel to simulate the influence of the RF chains and propagation channel on the signal. The resulting samples are given to the demodulator of the receiving node.

2.3.1 Network Components

In order to perform handover experimentation, the minimal involved LTE network components are a single UE and two eNBs. The number of UE and eNB instances as well as their own characteristics can be provided as OAI input. Fig. 2.3 shows the considered network topology for the handover experiment and the corresponding coordinates for each network component, a brief description of the experiment's setup is given as follows:

- 2 eNBs: one source and one target single omnidirectional antenna cell
- 1 UE: single omnidirectional antenna user

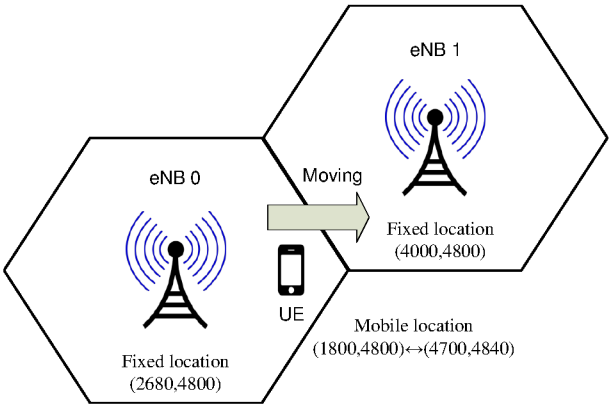


Figure 2.3: Network topology

- One radio chain (Tx/Rx) per UE/eNB wireless single-input single-output (SISO) link in full PHY mode without radio frontends.

All the network component parameters are fully re-configurable either via concrete configuration files provided as user input or via control line interface (CLI). The same holds for the handover parameters that are crucial for the successful handover operation.

2.3.2 Network topology and mobility

Using OAI, the network topology can be configured by the experimenter and is given as input to the OAI platform. The experimenter can choose if the network nodes will be in a fixed topology or they will be moving defining the specific mobility model (i.e., mobility traces as input, random way point, or random walk). The mobility is limited in a grid of 10000×10000. A brief description of the experiment's network topology is given as follows:

- UE is moving based on a specific mobility pattern defined by traces given as input.
- The cells are fixed, i.e., placed in a constant position as depicted in Fig. 2.3. One stands for the source eNB located in (2680,4800) and the other stands for the target one located in (4000,4800).

In our experimentation scenario a UE is attached to the source eNB moving towards the target eNB and back again to its initial position (i.e., ping-pong movement pattern from (1800,4800) to (4700,4840) as shown in Fig. 2.3. This pattern is designed to trigger multiple handovers and measure average handover delay.

2.4 System performance evaluation in OAI emulator (oaisim)

We perform a set of system-level emulations in oaisim to investigate and analyze the handover X2 protocol performance as well as the impact of the involved handover parameters. A summary of the emulation parameters is provided in Table 2.1.

Fig. 2.4 characterizes the performance of the successful X2 HO in the RAN segment in terms of delay. The first delay measurement includes the time that the UE is attached to the

Table 2.1: Emulation Parameters

Parameter	**Value**	**Parameter**	**Value**
Carrier Freq.	1.9 GHz	*Max. Tx Pwr (dBm)*	eNB 15 - UE 0
Bandwidth	5 MHz	*Max. MCS*	DL 26 - UL 16
Pathloss at 1km	-122dB	*RLC Mode*	UM
Pathloss Exp.	3.67	*Duplexing*	FDD
Noise Model	AWGN and Rayleigh fading	*Noise floor (dBm)*	-105.1
Trans. Mode	1	*Antenna*	Omni 0dBi

source eNB (i.e., UE RRC connected) and sends the Measurement report in order to connect to the target eNB, until the reception of the RRC Reconfiguration message (i.e., UE RRC idle). The second one includes the time interval between the RRC Reconfiguration message reception at the UE side and the RRC Reconfiguration complete message reception at the target eNB side after UE synchronization (i.e., UE RRC connected). Based on the experiments extracted statistics in Fig. 2.4, if we focus on the second delay measurement that represents the length of the detach time, it can be noticed that we are within the estimated time interval based on Eq. (2.4) ($\leq$ 65ms). The estimation takes into account the average case scenario handover execution delay, i.e., $\sim$ 45 ms as well as additional delays of the the custom implementation related to the contention-based random access, i.e., $\sim$ 20 ms. It is noted that the additional delay added by the contention-based procedure can be reduced introducing the contention-free mechanism described in Section 2.2.3 where the UE gets directly the C-RNTI via the RRC Connection Reconfiguration message. Subsequently, the C-RNTI is incorporated in the header of the RRC Connection Reconfiguration complete message that is sent later to the target eNB in the HO completion part such that the UE to be distinguished by the target eNB. The support of this feature in the implementation phase skips the transmission of PRACH/UL grant messages that is the major bottleneck in terms of handover execution delay, especially if the UE signal quality is low that may result in handover execution delay increment or even handover failure and connection re-establishment as well as ping-pong effects.

Fig. 2.5 and Fig. 2.6 depict the filtered RSRP measurements both for the source (serving cell) and the target eNB (target cell) in time during the handover procedure for different values of S offset as defined in Eq. (2.3). The two figures include the real signal measurements as well as the smoothed ones in order to suppress any abrupt variations of the real signal measurements performed. At the beginning, the UE is attached to the source eNB. As the UE is moving towards the target cell, the filtered RSRP of the target eNB, see Eq. (2.1), becomes S times less than the source eNB one and the handover criterion of Eq. (2.2) starts to be fulfilled. When the *ttt* time period passes since the handover condition is continuously satisfied, the UE sends the measurement report to the source eNB. The offset values play a crucial role in a successful handover process. If S is large, the handover process is triggered too early when the UE is too far away from the target to send the required signaling for synchronization and the handover process might fail. On the other hand, if the offset value S is small or negative then the handover is triggered later and it might happen the UE to be closer to the target and far away from the source eNB that renders the UE unable to send the required signaling (i.e., RRC Reconfiguration

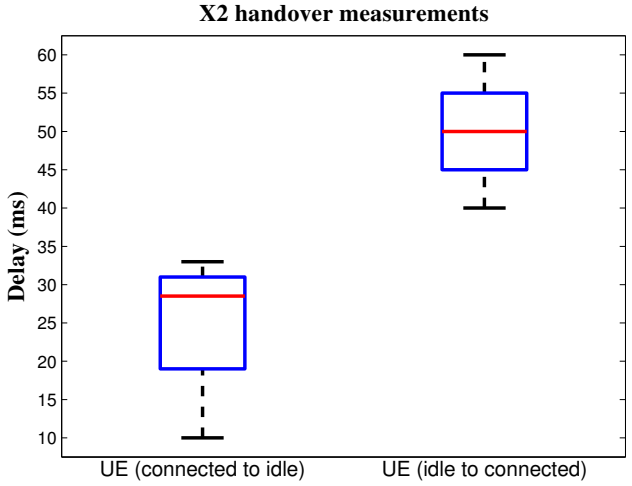


Figure 2.4: X2 handover measurements

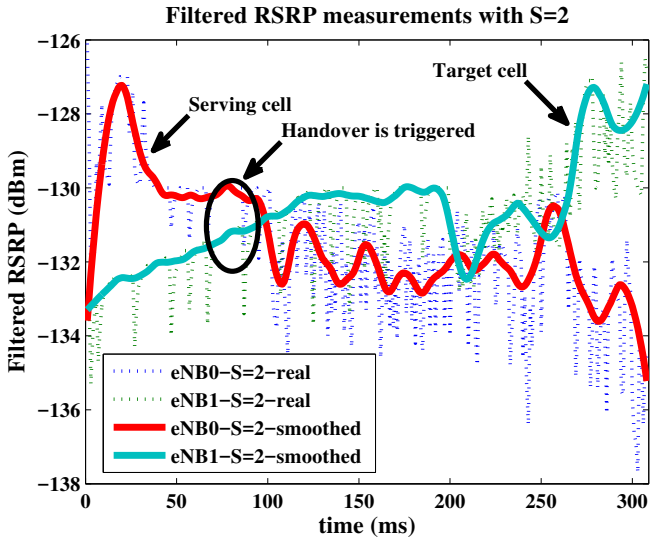


Figure 2.5: RSRP measurements both for source eNB (serving cell) and target eNB (target cell) for S=2

complete) to the source cell in order to detach from it. Both of the cases lead to additional delay and connection re-establishment procedure to the nearest cell need to be supported as discussed in the previous paragraph.

Finally, based on the OAI output logs (see OAI logger [66]) we captured the handover triggered regions across measurements in time. In both of the cases, the handover is triggered before the cross of the two curves, since $S > 0$. Another impact of the parameters, observing the OAI logger, is the earlier handover trigger when $S = 5$ compared to the $S = 2$ case as expected, since the criterion in Eq. (2.2) is satisfied faster. To this end, the selection of S parameter impacts the handover triggering and subsequently the handover execution delay as

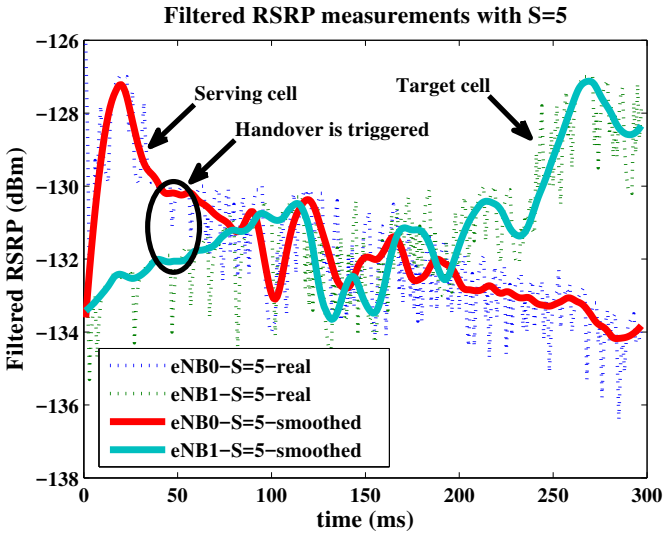


Figure 2.6: RSRP measurements both for source eNB (serving cell) and target eNB (target cell) for S=5

Table 2.2: Basic system Parameters

Parameter	**Value**	**Parameter**	**Value**
Carrier Freq.	2.6 GHz (Band 7)	*Max. Tx Pwr (dBm)*	eNB 10 - UE 0
Bandwidth	5 MHz	*Antenna*	Omni 0dBi
Trans. Mode	1	*Duplexing*	FDD
Max. MCS	DL 28 - UL 20	*RLC Mode*	UM

stressed earlier.

2.5 Experimental results with OAI X2 HO RF testbed

As an extension of X2 handover implementation in oaisim, we perform real-world RF measurements in the deployed OAI X2 HO experimental testbed presented in Appendix C. The objective is to test, study and analyze the OAI X2 handover system performance in a pragmatic testbed. To this direction, similarly to Section 2.4, we focus on the performance of the X2 handover in terms of handover delay that is the main implementation criticality. Finally, UE network information snapshots as well as RSRP measurements are collected using the Network Signal Guru Android application [67]. A summary of the system parameters is provided in Table 2.2.

Specifically, we perform 4 experiments as the Table 2.3 depicts to examine the X2 HO delay performance. The experiments are categorized in two basic groups: Experiment 1 & 2 correspond to the scenario of 4 dB difference of channel attenuation between target (Ch1) and source (Ch0) while Experiment 3 & 4 correspond to the scenario of -2 dB difference. In both groups, the

Table 2.3: Parameter notation

Experiment	Parameter	Value	Parameter	Value
#1	prach_config_index	0	Channels attenuation	Ch0: 12, Ch1:16
#2	prach_config_index	14	Channels attenuation	Ch0: 12, Ch1:16
#3	prach_config_index	0	Channels attenuation	Ch0: 12, Ch1:10
#4	prach_config_index	14	Channels attenuation	Ch0: 12, Ch1:10

PRACH Configuration Index is variable (0 or 14) [5].

Fig. 2.7 describes the performance of the X2 HO in the OAI RAN in terms of delay. The left-side bar columns delay measurements include the time that the source eNB receives the RRC Measurement report by the UE in order to connect to the target eNB (i.e., UE RRC connected) till the reception of the X2 Handover Request ACK message sent by the target eNB to the source eNB [6]. The right-side bar columns delay measurements include the time that the source eNB sends the RRC Connection Reconfiguration message to UE (i.e., UE RRC idle) till the reception at the target eNB side of the RRC Connection Reconfiguration Complete message reception after UE synchronization (i.e., UE RRC connected). Based on the experiments extracted results in Fig. 2.7, if we compare the left and right-side bar columns delay measurements, there is a large difference due to the main bottleneck of the handover process in terms of delay that is the transmission of PRACH/UL grant messages in the HO execution phase. Focusing on such bottleneck, Exp.2 & 4 apply PRACH Configuration Index 14 instead of 0 used in Exp.1 & 3 to reduce the delay as it is depicted. That implies more chances in Exp.2 & 4 for the target eNB to detect the PRACH message as the transmissions are performed in all sub frames of any SFN by using prach_config_index=14 compared to Exp.1 & 3. In addition, we examine the channel impact in PRACH reception by the target eNB; in Exp.3 & 4, we apply less attenuation to the target eNB (Ch1>Ch0) compared to Exp.1 & 2 (Ch0>Ch1) using octoBoxquadAtten's programmable attenuator (see Section C.2). Due to wireless environment variability, the handover triggering [7], RRC Connection Reconfiguration message detection by UE (more interference introduced by target eNB) and UE synchronization initiation are impacted. To this end, in Exp.3 the preamble is detected 1 frame later compared to Exp.1 (SFN needs to be even number-prach_config_index=0) and in Exp.4 the preamble is detected one sub frame later compared to Exp.2 based on our log files. The latter explains the additional delays, i,e., 10 ms for Exp.1 vs Exp.3 and 1 ms for Exp.2 vs Exp. 4 depicted in Fig. 2.7.

The current RF testbed implementation delay measurements exceed the estimated time interval based on Eq. (2.4) ($\leq$ 65ms) presented in Section 2.4 implying further considerations for real-time system optimization compared to the oaisim emulator results where delay measurements were lying within this interval. Finally, contention-free mechanism can be introduced to avoid PRACH/UL grant messages that represent the main handover bottleneck in terms of delay.

As follows, Fig. 2.8(a) and Fig. 2.8(b) depict network information as MCC, MCN, SIM

[5]PRACH Configuration Index determines the prach preamble timing and prach preamble type (see TS36.211-Table 5.7.1-2). PRACH Configuration Index 0 means that UE is allowed to transmit PRACH only at sub frame number 1 of every even system frame number (SFN) while PRACH Configuration Index 14 means that UE is allowed to transmit PRACH in all sub frame numbers of any SFN.

[6]The measurements are performed in the eNB side due to no access in the COTS UE signalling log files.

[7]The S parameter expressed in Eq. (2.3) is set to 10.

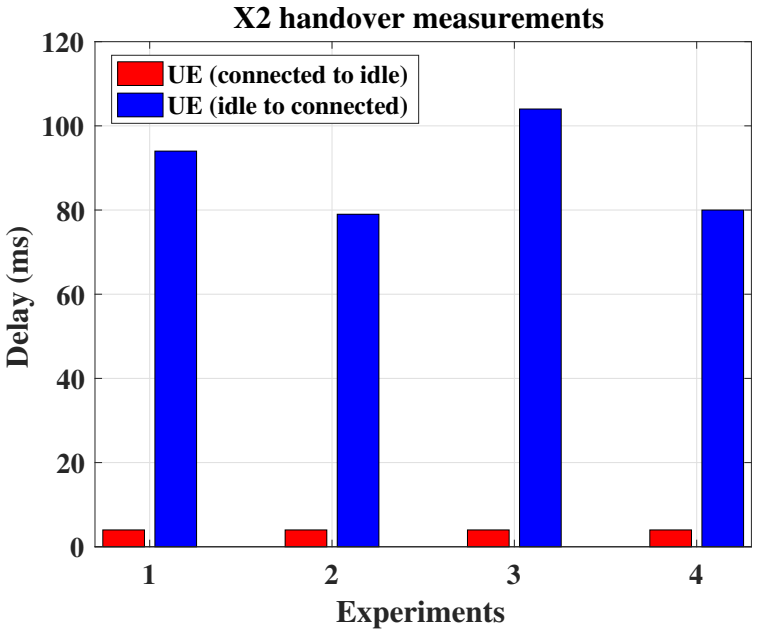


Figure 2.7: X2 handover measurements

state, TAC, RSRP, RSRQ etc. for both source/target eNB before and after HO. It is critical to state that the serving/neighbor physical cell id (PCI) is 0/1 before HO while PCI is 1/0 after HO showing that HO process was succeeded. Further, Fig. 2.9(a) and Fig. 2.9(b) offer the source/target RSRP before and after HO. It is observed that the serving cell is superior in terms of power before HO and the neighbor cell is superior to the source after HO, i.e., A3 event has been triggered since the neighbor cell becomes an offset better than the serving cell. With the increment of the source eNB and the decrement of the target eNB channel attenuator (Ch0/Ch1), we emulate the UE mobility in order to trigger the handover process. In Fig. 2.10, the serving/neighbor cell RSRP is presented across time signifying RSRP variation as described. Handover A3 event is triggered before the two curves cross, i.e., $S = 10 > 0$ in our case. Consequently, the selection of S parameter impacts the handover triggering and subsequently the handover execution delay as stressed earlier.

2.6 Discussion

Based on the current study, many contingent directions can be followed beyond the 3GPP standardization handover process. As mentioned earlier, LTE handover is UE assisted but network controlled. In more detail, handover event (e.g., A3 event) is triggered by UE and target eNB is reported to source eNB; later call admission control is performed by target eNB in order to accept or reject the UE handover. Subsequently, the centralized network control provided by SDN technologies can enable more sophisticated mobility management mechanisms that consider metrics as HO failures, dropped calls, ping-pong ratios, UE interference/speed/delay, cells load etc. To this direction, conventional handover decision rules as the one expressed in Eq. (2.3) can be changed introducing new offsets (i.e., S-like parameters) that are based on the aforementioned HO metrics and are optimized by taking into account the entire network

(a) Network information before HO (PCI =0) (b) Network information after HO (PCI =1)

Figure 2.8: Network information before/after HO

(a) Source/Target eNB RSRP before HO (PCI =0) (b) Source/Target eNB RSRP after HO (PCI =1)

Figure 2.9: Source/Target eNB RSRP before/after HO

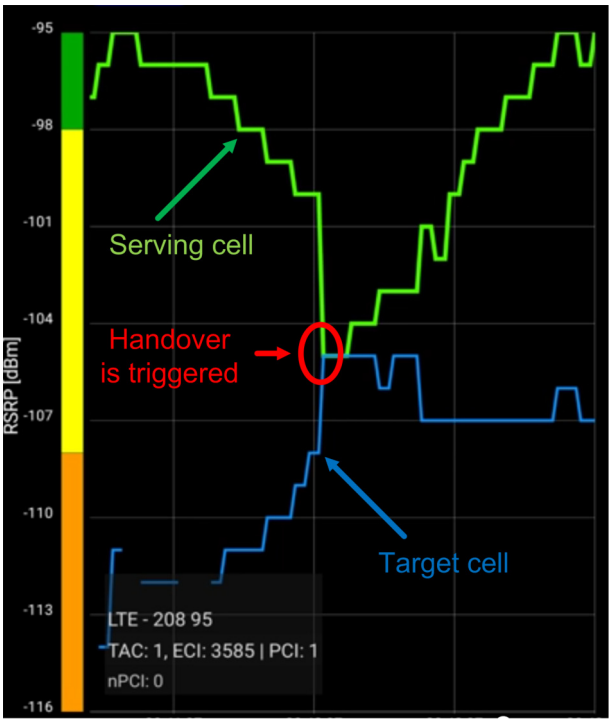


Figure 2.10: RSRP measurements both for source eNB (serving cell) and target eNB (target cell)

knowledge. In other words, handover triggering can be controlled and adjusted optimally by network operators relying on advanced handover decision schemes. Such schemes consider any network conditions and detect persistent bottlenecks in order to both reduce handover delay or failure and optimize network performance metrics, e.g., throughput, user QoS etc. To this direction, a load-aware HO decision algorithm is introduced and analyzed in Chapter 3 inspired by the concluding remarks of this chapter.

2.7 Conclusion

This chapter presents the details of the LTE/LTE-A X2 handover procedures and parameters. The performance of the X2 handover in terms of transition delay from RRC connected to idle state with respect to the source eNB, and from RRC idle to connected state for the target eNB are investigated. Further, a set of real world experiments were performed to assess the performance of X2 handover with real applications and radio channels and be compared with the emulated ones. Presented results show that the uplink synchronization procedure obtained through the Medium Access Control (MAC) RACH procedure contributes the most to the total latency, and that the contention-free preambles are preferred to avoid collision, especially in high load and mobility scenarios. Finally, the impact of the handover parameters on the decision algorithm is studied thoroughly.

Chapter 3

Load-aware Handover Decision Algorithm in Next-generation HetNets

3.1 Introduction

In Chapter 2, we analyzed the 3GPP legacy X2 HO studying the protocol and transport delay performance as well as the impact of the system's HO parameters. The aforementioned study shows that HO is a costly process in terms of time implying moderation in handover decisions, especially when HO is crucial to happen fast and seamlessly in the network.

HO decision algorithms in heterogeneous networks (HetNets) are emerging with the extreme densification and offloading of cellular systems as the key technologies to get 1000x data rate [68]. Many works have been done in HO decision algorithms considering the overall system performance from the *user* and *network perspective*, in the context of long-term evolution (LTE)/LTE–Advanced (LTE–A) networks [6]. Specifically, from the network perspective, advanced self-organizing map (SOM) is examined to suppress handovers in regions that coincide with many unnecessary HOs [9]; also the HO offset tuning is performed via network load-balancing for assuring handover to possible target cells and no return to their serving cell [69]. In addition, a HO algorithm that adapts the hysteresis and the time-to-trigger (TTT) based on certain network key-metrics (e.g., HO failure or ping-pong ratios) is studied in [8], along with the self-organising networks (SONs). The main focus so far has been to enhance the system performance from the user-perspective as discussed in Chapter 2. The main categories include: (i) received signal strength (RSS), (ii) user speed and (iii) interference-management [10, 11]. A thorough analysis for the selection of the hysteresis parameter based on the minimization of the outage probability and the probability of handover through dynamic optimization via a hybrid system is presented in [70].

However as mentioned in Chapter 1, denser deployments in HetNets face high variability in load emerging more advanced HO algorithms that *jointly* consider user and network perspectives. In addition, RSS-based legacy techniques cannot be applied, due to the assymetrical power characteristics among macrocells and picocells (encompassed by SCs) in HetNets. In that sense, UE doesn't consider underloaded SCs while is nearby and remains attached to the macrocell that is superior in terms of power. Authors in [34] suggest to scale down the macrocell RSS with an

appropriate factor and allow UEs to connect to other SCs. Hence, no user service requirements are considered such throughput, delay etc. The latter implies that review of HO proposed algorithms is needed considering the future HetNets deployment. The chapter contributions can be summarized as follows:
1) We construct a framework considering both user (i.e., RSS) and network (i.e., cell load) perspectives. Specifically, the *service delay* of a random flow is analytically derived, as a key QoS criterion for the HO decision algorithm under the assumption of uneven transmission power regimes. This approach overcomes the shortcomings created by only considering RSS criteria in HO decision for HetNets.
2) We propose a HO decision algorithm and sketch an implementable framework based on Software Defined Networking (SDN) architecture, as suggested to be a key enabler for the realization of 5G networks [71].
3) We investigate the related trade-offs involved, in different load variation scenarios. Further, the proposed algorithm is compared with the (i) one used in traditional LTE systems and (ii) a distance-based algorithm proposed in [34] and significantly outperforms them.

The remainder of this chapter is organized as follows. Section 3.2 provides the problem setup and system modeling approach. Section 3.3 presents the proposed load-aware HO decision algorithm, and Section 3.4 provides its applicability to the software-defined networking (SDN) framework. Simulation results are presented in Section 3.5. Section 3.6 offers a brief discussion. Finally, Section 3.7 provides concluding remarks.

3.2 System model and assumptions

We consider a network consisting of a macrocell and a group of picocells located at given distances D_j from it, where $j \in \{1, \ldots, \mathcal{P}\}$ and $\mathcal{P}$ is the total number of the picocells. The cells, i.e., macrocell/picocells, are denoted by m, p_j, respectively. In addition, the picocells are assumed circular on average, regardless of the fading effects presence. Each base station (BS) is considered to be located in the center of each circular cell $i \in \{m, p_j\}$ of a given radius R_i corresponding to the (x_i, y_i) coordinates. In the following, the system model is presented in more detail:

3.2.1 Air-interface model

A.1-Reference Signal Received Power (RSRP): Let $r_m^{\mathrm{dBm}}[k]$ and $r_{p_j}^{\mathrm{dBm}}[k]$ denote the RSRP from the macrocell BS and the picocell BS at time k [1] in dBm scale, respectively:

$$r_i^{\mathrm{dBm}}[k] \triangleq P_{T_x,i}^{\mathrm{dBm}} + P_{L,i}^{\mathrm{dB}}(d_i^k) + \psi_i^{\mathrm{dB}}[k],\ i \in \{m, p_j\}, \tag{3.1}$$

where $P_{T_x,i}^{\mathrm{dBm}}$ denotes the BS transmission power, $P_{L,i}^{\mathrm{dB}}(d_i^k)$ denotes the path loss in dB, d_i^k represents the distance from the BS to the user that is greater than a reference distance d_r and $\psi_i^{\mathrm{dB}}[k]$ stands for the shadowing fading in dB for each cell i. The pathloss $P_{L,i}^{\mathrm{dB}}(d_i^k) \triangleq 10\log_{10}(K) - 10\eta\log_{10}\left(d_i^k/d_r\right)$, where K is the path loss measured at reference distance d_r and η is the pathloss exponent. The K parameter is given by $K \triangleq [c_0/(4\pi f_c d_r)]^2$, where c_0 is the universal speed of light in vacuum and f_c is the carrier frequency. The shadow fading

[1] k corresponds to the discretization of the continuous time t sampling at kT_s intervals, where T_s stands for the measurement sampling period.

$\psi_i^{\mathrm{dB}}[k] \sim \mathcal{N}\left(0, \sigma_{\mathrm{dB},i}^2/\xi^2\right)$ and is assumed to be independent across i and k, where $\xi \triangleq 10/\ln(10)$. Averaging is performed by an exponential moving average (EMA) filter, i.e., low-pass filter, for smoothing any RSRP abrupt variations. High frequency fluctuations (i.e., Rayleigh fading) are filtered out and can be neglected. Consequently, the output filtered signal is:

$$\bar{r}_i^{\mathrm{dBm}}[k] \triangleq (1-\alpha)\bar{r}_i^{\mathrm{dBm}}[k-1] + \alpha r_i^{\mathrm{dBm}}[k], \tag{3.2}$$

where $\alpha \triangleq 2^{-q/2}$ and $q \in \mathbb{N}$.

3.2.2 Mobility and Traffic model

B.1-Network users: The users are distinguished in two different categories depending on their mobility status: a) **Static users** (SU): Users that are not moving on average and they don't intend to handover. The static users are divided in two subcategories based on their ongoing traffic: i) **Active users** (AU): Users that have already associated with a BS and generate new flows. ii) **Disconnected users** (DU): Users that are considered to be switched-off. This implies no association with a specific BS. b) **Mobile users** (MU): Users that are moving and intend to handover. The users mobility can be described with any mobility model.

The SUs are getting active with probability p_{r_i} in each cell i at each time k. On that account, the number of the AUs in each cell i at time k, $N_{\mathrm{AU},i}^k$, is distributed according to the binomial distribution with parameters $N_{\mathrm{SU},i}$ and p_{r_i}, where $N_{\mathrm{SU},i}$ stands for the total number of the static users in each cell i. The exact value of $N_{\mathrm{AU},i}^k$ is considered to be known to the BS at each time k. The MUs are always active and N_{MU} stands for their total number.

B.2-Traffic flows: New flows (all considered as best-effort) by active users are assumed to arrive according to a Poisson process with the total arrival rate in each cell i at time k [72]:

$$\lambda_i^k = \lambda\left(\tilde{N}_{\mathrm{MU},i}^k + N_{\mathrm{AU},i}^k\right), \tag{3.3}$$

where λ denotes the flow arrival rate per user. The flow size follows a general distribution with mean Y. We assume the Processor Sharing (PS) scheduling discipline and adopt the stationary M/G/1/PS system [72]. $\tilde{N}_{\mathrm{MU},i}^k \triangleq \mathbf{1}\left(u \notin i\right) + N_{\mathrm{MU},i}^k$, where $\mathbf{1}(\cdot)$ stands for the indicator function and becomes one when a mobile user u that intends to handover is not associated with the cell i (i.e., $u \notin i$) at time k; $N_{\mathrm{MU},i}^k$ denotes the total number of mobile user that are associated with the cell i at time k (i.e., a subset of N_{MU}). $N_{\mathrm{MU},i}^k$ is also acquainted by each cell i BS. If the mobile user u is already associated with the cell i, then $\mathbf{1}(\cdot)$ is zero and u is already included in $N_{\mathrm{MU},i}^k$. Thus, after computing $\tilde{N}_{\mathrm{MU},i}^k$ and plug-inning it to Eq. (3.3), we get an estimation of the total arrival rate after the HO procedure. This is also used for the prediction of the service delay, as it will be discussed in Section 3.3.

3.3 Handover algorithm

A well-known criterion, commonly used in conventional HO decision algorithms for mobile communication systems (e.g., 3GPP LTE), is based on RSRPs comparison method in which hysteresis and threshold are included [73]. Specifically, the criterion for the handover scenario is expressed as:

$$\bar{r}_{p_j}^{\mathrm{dBm}}[k] > \bar{r}_m^{\mathrm{dBm}}[k] + \Delta \ \wedge \ \bar{r}_m^{\mathrm{dBm}}[k] < r_{m,th}^{\mathrm{dBm}}, \tag{3.4}$$

Algorithm 1: Handover algorithm

Input : $\bar{\mathcal{D}}_m^k$ /* Predicted average delay of the macrocell m */
$\bar{\mathcal{D}}_{p_j}^k$ /* Predicted average delay of the picocell p_j */
$r_{p,th}$ /* Picocell RSRP threshold */

Output: User cell association

/* Select the picocell with the maximum received power */
$j^* \leftarrow \arg\max_j \bar{r}_{p_j}$;
/* User stays within the picocell area */
if $\left(\bar{r}_{p_{j^*}}[k] > r_{p,th}\right)$ **then**
  /* Handover decision rule */
  if $(\tilde{r}_{j^*}[k] > \bar{r}_m[k])$ **then**
    connect to picocell;
  else
    connect to macrocell;

where Δ denotes the hysteresis and $r_{m,th}$ is the minimum RSRP macrocell threshold value.

There are scenarios where a macrocell is overloaded and it becomes crucial for the user to connect to a picocell that is underloaded. Due to significant differences in transmission power among a macrocell and picocells, there is a huge gap between their received powers at the UEs (i.e., $\bar{r}_m^{\mathrm{dBm}}[k] \gg \bar{r}_{p_j}^{\mathrm{dBm}}[k]$). Hence, the above criterion does not often hold and the conventional HO algorithm cannot be triggered to maintain the required QoS (e.g., in terms of rate, delay and throughput). Thus, *load-aware* algorithms development are crucial for uneven received powers regimes found in HetNet scenarios.

The proposed load-aware (LA) policy is described in Algorithm 1. The algorithm is applied only to the mobile and static active users. The disconnected users are not taken into account, since they are assumed to be switched-off. $\bar{r}_i[k]$ stands for the RSRP from each cell i at time k in linear scale. The picocell that the user intends to handover is the one with the maximum received power, denoted as p_{j^*}. After the selection of the p_{j^*} picocell, the condition $\bar{r}_{p_{j^*}}[k] > r_{p,th}$, with $r_{p,th}$ standing for the picocell RSRP threshold, ensures that the user stays within the picocell coverage area and the received signal is retained adequately strong. Thus, the algorithm is running only inside the picocell. The latter avoids ping-pong effects due to picocell non-coverage, i.e., cell-edge users are not considered. Subsequently, we define $\tilde{r}_{j^*}[k] \triangleq \bar{r}_{p_{j^*}}[k] + \left[1 - f\left(\bar{\mathcal{D}}_m^k, \bar{\mathcal{D}}_{p_{j^*}}^k\right)\right] \bar{r}_m[k]$. $\bar{\mathcal{D}}_i^k$ denotes the *prediction* of the average (service) delay of a new flow of a cell i at time k that the user experiences when staying in the picocell. An analytical expression for $\bar{\mathcal{D}}_i^k$ is given later in Eq. (3.8). It is noted that the prediction of the average delay $\bar{\mathcal{D}}_i^k$ is performed at time k, but it represents an estimation for the entire time period that the user is staying within the picocell (i.e., user's delay assumed to be $\bar{\mathcal{D}}_i^k$ during this time period). Further, $f(\cdot)$ stands for a continuous function and is constructed properly to force the user to connect in the picocell p_{j^*}, iff $\bar{\mathcal{D}}_{p_{j^*}}^k \ll \bar{\mathcal{D}}_m^k$, despite the fact that $\bar{r}_m[k] \gg \bar{r}_{p_{j^*}}[k]$. This approach meets the user needs that cannot be supported by the congested macrocell and its association with a picocell is critical in terms of QoS. The condition that must hold to connect to the picocell

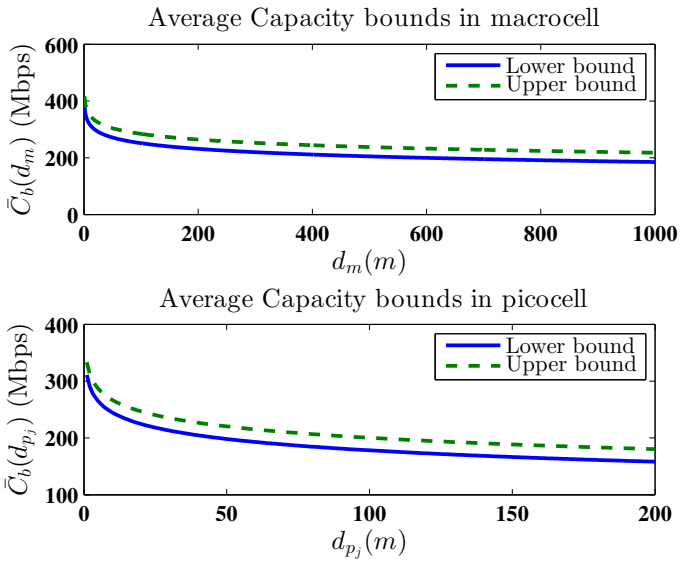


Figure 3.1: Upper and lower bounds of the average capacity for the macrocell and the picocell, as a function of the distance.

is given analytically by the following inequality that applies the above policy:

$$\tilde{r}_{j^*}[k] > \bar{r}_m[k] \Leftrightarrow \bar{r}_{p_{j^*}}[k] > f\left(\bar{\mathcal{D}}_m^k, \bar{\mathcal{D}}_{p_{j^*}}^k\right)\bar{r}_m[k], \tag{3.5}$$

otherwise the user is connected to the macrocell. Finally, we define the function

$$f\left(\bar{\mathcal{D}}_m^k, \bar{\mathcal{D}}_{p_j}^k\right) \triangleq \exp\left[-c\frac{\bar{\mathcal{D}}_m^k}{\bar{\mathcal{D}}_{p_j}^k}\right], \; c \geq 1 \tag{3.6}$$

with a co-domain in $[0, 1]$ that provides such a property as intended. Specifically, the exponential nature of the function provides quick convergence to zero when $\bar{\mathcal{D}}_{p_{j^*}}^k \ll \bar{\mathcal{D}}_m^k$. This manages to scale down the received powers asymmetry in order to hold Eq. (3.5) and associate the user with the underloaded picocell. Finally, the constant c stands for a tuning parameter. Next, we describe the computations needed for the prediction of $\bar{\mathcal{D}}_i^k$:

The $\tilde{C}(d_{l,i}) \triangleq \left[\bar{C}_L(d_{l,i}) + \bar{C}_U(d_{l,i})\right]/2$ is defined as the arithmetic mean of $\bar{C}_L(d_{l,i}), \bar{C}_U(d_{l,i})$ that represents the analytical lower and upper bounds of the averaged capacity as a function of the l^{th} static active user's distance, $d_{l,i}$, from each cell i (see Fig. 3.1). A detailed description of their analytical expression is given in Appendix A (see Eq. (A.2), (A.3)). The latter computations are consistent with the prediction of $\bar{\mathcal{D}}_i^k$ for the entire user's sojourn within the picocell, since $d_{l,i}$ remains invariant for the static users across time k. For the mobile users, there is no prior knowledge of their future positions, thus their corresponding lower and upper bounds of the averaged capacity for each cell i, $\bar{C}_L^i, \bar{C}_U^i$, are averaged over all the possible distances, assuming that the user moves uniformly through all the points within the picocell after multiple visits during its staying within it. The uniform distribution of the points as well as the corresponding capacity analysis are enclosed in Appendix A. Thus, the aforementioned computations are consistent also with the prediction of $\bar{\mathcal{D}}_i^k$ for the total user's sojourn within the picocell. Finally, their respective arithmetic mean is given by $\tilde{C}_i = \left(\bar{C}_L^i + \bar{C}_U^i\right)/2$.

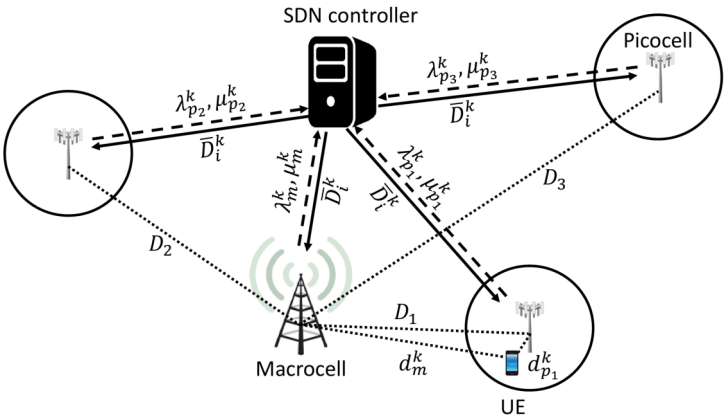


Figure 3.2: Network setup and SDN controller.

According to the above computations, the expected average rate [2] at time k for each cell i is determined by [74]:

$$\mathcal{R}_i^k = \frac{\tilde{N}^k_{\mathrm{MU},i}\tilde{C}_i + \sum_{l=1}^{N^k_{\mathrm{AU},i}} \tilde{C}\left(d_{l,i}\right)}{\tilde{N}^k_{\mathrm{MU},i} + N^k_{\mathrm{AU},i}}. \tag{3.7}$$

Consequently, based on the system assumptions (i.e., stationary M/G/1/PS), the prediction of $\bar{\mathcal{D}}_i^k$ is given by [72]:

$$\bar{\mathcal{D}}_i^k = \frac{1}{\mu_i^k - \lambda_i^k}, \tag{3.8}$$

where $\mu_i^k = \mathcal{R}_i^k / Y$ stands for the service rate at time k for each cell i.

3.4 SDN-based implementation

The proposed handover algorithm decides whether a mobile user should handover from its (currently) associated BS, to another nearby BS that might potentially promise enhanced QoS. To make such a decision, UE should be aware of the cell load or the offered user QoS for *both* BSs. To this end, a centralized entity determines the service delay of each individual cell based on the received measurements and communicates this with the underlying network. SDN architecture facilitates this procedure, since it offers a centralized programmable control for the underlying network [71]. Following the SDN outline, we consider three planes as illustrated in Fig. 3.2:
Controller tier: Each time k, the controller: a) receives the respective λ_i^k and μ_i^k from all cells, b) computes and sends the corresponding delays $\bar{\mathcal{D}}_i^k$ to all BSs needed (e.g., a certain BS should know not only his corresponding service delay, but also the one from its neighboring BSs).
Network tier: Each time k, BSs: a) send the respective λ_i^k and μ_i^k, b) receive the corresponding delays $\bar{\mathcal{D}}_i^k$, c) send the $\bar{\mathcal{D}}_i^k$ to the UE.
User tier: Each time k, the UE: a) receives the delays $\bar{\mathcal{D}}_i^k$, b) triggers the association procedure based on Algorithm 1.

[2] The rate considers a priori the user association, as it is used for the predicted delay. Thus, $\tilde{N}^k_{\mathrm{MU,i}}$ is plug-inned.

Table 3.1: Simulation setup parameters

parameter	value	parameter	value	parameter	value
R_m	1000 m	f_c	700 MHz	p_{r_m}	0.8
R_{p_j}	200 m	B	10 MHz	$p_{r_{p_j}}$	0.8
$P_{T_x,m}^{\mathrm{dBm}}$	43 dBm	d_r	1 m	Δ	3 dB
$P_{T_x,p_j}^{\mathrm{dBm}}$	21 dBm	q	4	Y	10 MBytes
N_0^{dBm}	-104.5 dBm	η	2	s	50
$r_{m,\mathrm{th}}^{\mathrm{dBm}}$	-90 dBm	$\sigma_{\mathrm{dB},m}$	8 dB	δ	10
$r_{p,\mathrm{th}}^{\mathrm{dBm}}$	-55 dBm	σ_{dB,p_j}	6 dB	c	1

3.5 Simulations

This section investigates the proposed LA HO policy performance compared to other HO policies. The entire framework for the simulation is built in MATLAB. The simulation parameters are provided in Table 3.1 [60] [3].

The MUs mobility model is based on a two-dimensional random walk (2D RW) limited in a finite space (i.e., circular macrocell) with step s or s/δ depending on where the user is moving inside (i.e., macrocell/picocell) and $\delta > 0$ denotes a degrading step factor [4]. This variable step simulates the scenario that the users are staying inside the picocell area for a sufficient amount of time compared to the macrocell one (e.g., in airports, malls, etc). As next, Fig. 3.3–3.4 present a scenario with three picocells placed in different distances from the macrocell ($D_1 = 250$ m, $D_2 = 350$ m, $D_3 = 450$ m) and one mobile user (i.e., $N_{\mathrm{MU}} = 1$) [5], as shown in Fig. 3.2.

Fig. 3.3 demonstrates the mobile user cell assignment probabilities, $P_r(u \in m),\ P_r(u \in p)$ for the macrocell and the picocell, respectively. These mobile user cell assignment probabilities are presented as a function of the flow arrival rate λ for different number of static users [6] $N_{\mathrm{SU},m} = 200, 400, 600$ and $N_{\mathrm{SU},p_j} = 10$. Their computation is performed via Monte-Carlo simulations within the picocells p_j. Especially, $P_r(u \in p)$ is also averaged out for all picocells p_j. Specifically, with the increment of λ, the total load in the macrocell augments, since the number of the macrocell static users is greater than the picocell ones, causing macrocell overload. Consequently, the difference in the corresponding delays increases (i.e., macrocell delay is greater than that of the picocells). Thus, the LA algorithm associates the user with the underloaded cell (i.e., picocell) or equivalently $P_r(u \in p)$ increases with the augmentation of λ. Also, as $N_{\mathrm{SU},m}$ increases the macrocell becomes overloaded with smaller values of λ, so $P_r(u \in p)$ tends to 1 in different $\lambda = 0.04, 0.06, 0.12$ for $N_{\mathrm{SU},m} = 200, 400, 600$, respectively. What is more, we include the RSS-based conventional (CONV) algorithm, described in Eq. (3.4) and compare it with the proposed LA one. In sharp contrast with the LA algorithm, the CONV algorithm always suggests the user to be associated with the macrocell. In case of congestion, this is suboptimal, since the

[3] N_0^{dBm} in Table 3.1 stands for the N_0 in dBm scale.

[4] The user coordinates $(x_i^k, y_i^k) \triangleq (x_i^{k-1} + s_i\cos(\phi), y_i^{k-1} + s_i\sin(\phi))$ in each cell i at each time k, where $s_i = \begin{cases} s & , i = m, \\ s/\delta & , i = p_j, \end{cases}$ and ϕ is uniformly distributed in $[0, 2\pi)$.

[5] The proposed algorithm as well as the performed analysis hold for the general type of cells (i.e., either overlapping or non-overlapping cells) with any radius R_i. Without loss of generality, we relax this assumption considering non-overlapping cells and the same radius for the picocells. The consideration of one mobile user is sufficient enough for studying the behavior of the algorithm without affecting its validity for the general case.

[6] At each time k, $N_{\mathrm{AU},i}^k$ is changing. On average, we have $\mathbb{E}[N_{\mathrm{AU},i}^k] = p_{r_i} N_{\mathrm{SU},i}$.

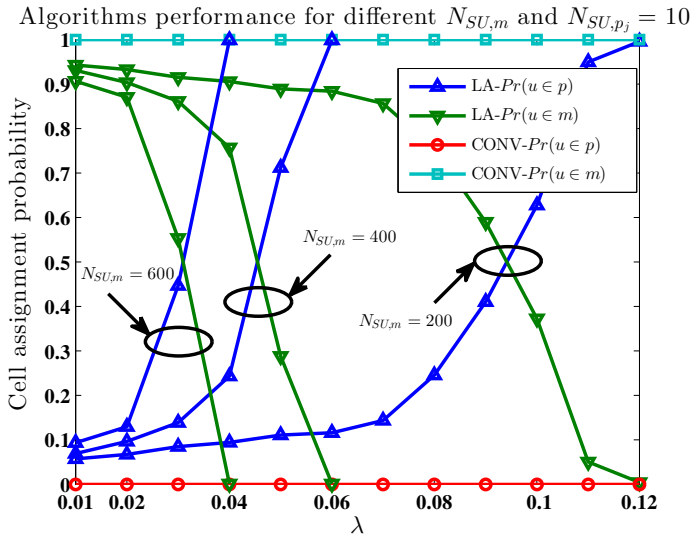


Figure 3.3: Cell assignment probability (CONV vs LA) for different number of static macrocell users $N_{\mathrm{SU},m}$ and number of static picocell users $N_{\mathrm{SU},p_j} = 10$.

user is placed close to one or more underloaded cells. $P_r(u \in m)$, is also plotted, showing the complementary behavior of the algorithm.

Fig. 3.4 shows the predicted delays, $\bar{\mathcal{D}}_m, \bar{\mathcal{D}}_p$ for the macrocell and the picocell, as a function of the flow arrival rate λ. Specifically, these are the $\bar{\mathcal{D}}_m^k, \bar{\mathcal{D}}_p^k$ that are also averaged out across time k. Their computation is performed via Monte-Carlo simulations for different number of static users $N_{\mathrm{SU},m} = 200, 400, 600$ and $N_{\mathrm{SU},p_j} = 10$ within the picocells p_j. Especially, $\bar{\mathcal{D}}_p$ is also averaged out for all picocells p_j. It is noticed that the average delay, $\bar{\mathcal{D}}_m$, increases *sharper* for larger values of $N_{\mathrm{SU},m}$ and smaller values of λ, according to Eq. (3.8). The latter implies that total arrival rate becomes faster equal to the system service rate. $N_{\mathrm{SU},m}$ and λ affect $\bar{\mathcal{D}}_m$ increment and consequently enlarge the difference between $\bar{\mathcal{D}}_m$, $\bar{\mathcal{D}}_p$. Hence, when their ratio becomes large ($\bar{\mathcal{D}}_m \gg \bar{\mathcal{D}}_p$), $f(\cdot)$ tends to zero and vanishes the gap between the received powers. Thus, the impact of $N_{\mathrm{SU},m}$, λ on $\bar{\mathcal{D}}_m$, $\bar{\mathcal{D}}_p$ ratio exceeds the impact of received powers asymmetry. This validates the LA algorithm functionality that associates the user with the underloaded picocells in the case of an overloaded macrocell ($P_r(u \in p)$ goes to one), regardless the gap in received powers. On the other hand, HO is not triggered if there is not much difference in $\bar{\mathcal{D}}_m$, $\bar{\mathcal{D}}_p$, i.e., the user QoS is satisfied ($P_r(u \in p)$ goes to zero). Consequently, RSS-based HO that maximizes the instantaneous rate of a user (i.e., the best modulation and coding scheme (MCS) is used), reflects user QoS only when the BS is lightly loaded. However, user performance, in terms of per flow delay, may be severely affected if the BS offering the best RSS is congested.

In Fig. 3.5, the picocell assignment probability, $P_r(u \in p)$, is depicted as a function of the user's distance from the macrocell for fixed number of $N_{\mathrm{SU},m} = 200$ and $N_{\mathrm{SU},p_j} = 10$ in the picocell. The proposed algorithm is compared with the CONV and a distance-based (DIST) one provided in [34], for different λ. The latter is applied in hierarchical macrocell/SCs network that is also our case. The DIST approach associates the user with its closest SC regardless the asymmetry in BSs powers. This is achieved using a scale down factor α based on the maximization of the SC assignment probability. Further, some of the system parameters are

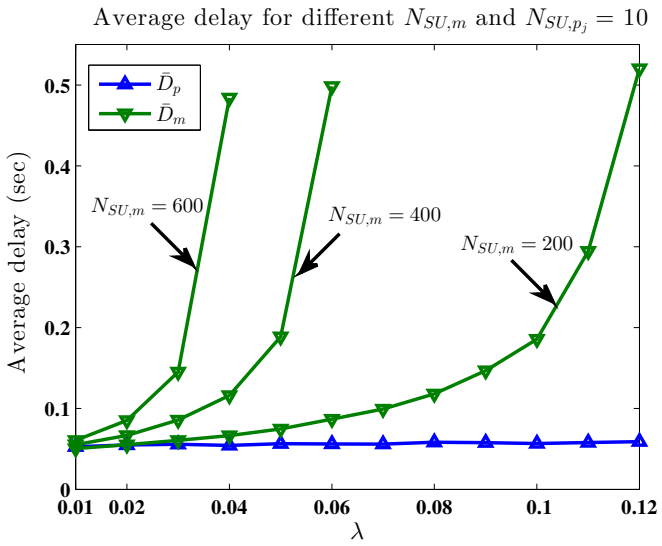


Figure 3.4: Average delay in the macrocell/picocell for different number of static macrocell users $N_{\mathrm{SU},m}$ and number of static picocell users $N_{\mathrm{SU},p_j} = 10$.

Table 3.2: Delay gain for $N_{\mathrm{SU},m} = 200$, $N_{\mathrm{SU},p_j} = 10$ and different λ

$\boldsymbol{\lambda}$	**0.05**	**0.08**	**0.1**	**0.11**	**0.12**
$\bar{\mathcal{D}}_{\mathrm{LA}}$ (sec)	0.0701	0.0992	0.1261	0.1125	0.0621
$\bar{\mathcal{D}}_{\mathrm{DIST}}$ (sec)	0.0621	0.0807	0.1109	0.1470	0.2808
$\bar{\mathcal{D}}_{\mathrm{CONV}}$ (sec)	0.0743	0.1160	0.1864	0.2708	0.5496
$\bar{\mathcal{D}}_{\mathrm{DIST}}/\bar{\mathcal{D}}_{\mathrm{LA}}$	0.8861	0.8141	0.8798	**1.3066**	**4.5235**
$\bar{\mathcal{D}}_{\mathrm{CONV}}/\bar{\mathcal{D}}_{\mathrm{LA}}$	**1.0596**	**1.1701**	**1.4790**	**2.4073**	**8.8547**

changed to adapt the simulated scenario in [34] [7]. For the DIST algorithm, $P_r(u \in p)$ is close to 1 after a specific user's distance from the picocell despite the fact that the macrocell is overloaded or not. However, with the LA approach, $P_r(u \in p)$ increases, depending on the macrocell load for different λ, no matter the user's proximity to the picocell. Thus, if the user is close to a picocell, its association with it may be unnecessary, since its QoS requirements may be already met. Finally, the CONV algorithm does not trigger the HO, due to BSs power assymetry.

The overall delay that the user experiences (i.e., $\bar{\mathcal{D}}_m$ from the macrocell and $\bar{\mathcal{D}}_p$ from the picocell depending on where it is associated) is presented in Table 3.2. The setup is considered the same as in Fig. 3.5. Using the CONV algorithm, it is noted that the overall delay, $\bar{\mathcal{D}}_{\mathrm{CONV}}$, is identical to the delay of the macrocell (see Fig. 3.4 for $N_{\mathrm{SU},m} = 200$ and $N_{\mathrm{SU},p_j} = 10$), since the user is always connected to the macrocell. The DIST algorithm retains the user connected to the macrocell up to a specific distance threshold and after that it attaches it to the picocell (see Fig. 3.5) with $P_r(u \in p)$ close to 1. Thus, the augmentation of the macrocell delay, when it becomes overloaded, compared to the picocell ones, increases the overall delay, $\bar{\mathcal{D}}_{\mathrm{Dist}}$. The

[7] Specifically, one macrocell and one picocell are considered in distance $D_1 = 200$ m, $a^* = 0.96$ as the optimal a for the specific D_1, $\Delta = 0$, $R_p = 150$ m and $r_{p,\mathrm{th}} = -51$ dBm. In addition, the mobility model is changed, since the authors in [34] assume that one mobile user moving across a line from the macrocell to the picocell coverage area.

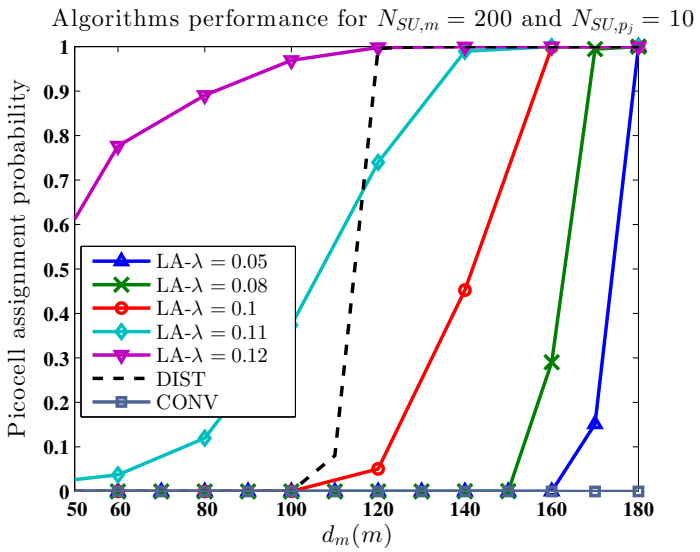


Figure 3.5: Picocell assignment probability (CONV vs DIST vs LA) for number of static macrocell users $N_{\mathrm{SU},m} = 200$ and number of static picocell users $N_{\mathrm{SU},p_j} = 10$ for different λ.

LA algorithm keeps a conservative policy when the load is low in the macrocell, i.e., the user remains associated with the macrocell. Hence, $P_r(u \in p)$ is low even if the user is close to the picocell. This policy gives slightly higher overall delays when $\lambda \leq 0.1$ compared to the DIST one, since the latter connects the user to the cell with the lower delay (i.e., picocell) earlier. Consequently, unnecessary handovers to the picocell and related signaling overhead are avoided with the LA approach. On the other hand, if the load is high in the macrocell, the LA algorithm associates the user with the underloaded picocell earlier than the DIST algorithm that is load-unaware giving significant performance to the overall delay, $\bar{\mathcal{D}}_{\mathrm{LA}}$. The respective ratios of the CONV/DIST algorithms over the LA algorithm (i.e., $\bar{\mathcal{D}}_{\mathrm{CONV}}/\bar{\mathcal{D}}_{\mathrm{LA}}$, $\bar{\mathcal{D}}_{\mathrm{DIST}}/\bar{\mathcal{D}}_{\mathrm{LA}}$) overall delays are demonstrated to show the corresponding gains in Table 3.2. The same trend holds for higher values of $N_{\mathrm{SU},m}$, but with different values of λ, as explained in Fig. 3.4. Finally, significant gains are provided in high load scenarios (i.e., $\simeq 4$ and $\simeq 8$ compared to the CONV and DIST approach).

3.6 Discussion

The possible directions beyond the proposed HO load-aware algorithm mainly focus on its applicability to other use cases. Such use cases urge for new handover decision rules corresponding to different functions $f(\cdot)$. For instance, to consider scenarios where picocells are overloaded compared to macrocell, i.e., $\bar{\mathcal{D}}_{p_{j^*}} \gg \bar{\mathcal{D}}_m$, a generalization of the current function in Eq. (3.6) is given as follows:

$$f\left(\bar{\mathcal{D}}_m^k, \bar{\mathcal{D}}_{p_j}^k\right) \triangleq \exp\left[c\left(-\frac{\bar{\mathcal{D}}_m^k}{\bar{\mathcal{D}}_{p_j}^k} + \frac{\bar{\mathcal{D}}_{p_j}^k}{\bar{\mathcal{D}}_m^k}\right)\right], \; c \geq 1. \tag{3.9}$$

The new generalized function is defined with a co-domain in $[0, +\infty]$. It is noted that the function scales down the power differences following the proposed HO rule, i.e., $\bar{r}_{p_{j^*}}[k] >$

$f\left(\bar{\mathcal{D}}_m^k, \bar{\mathcal{D}}_{p_{j^*}}^k\right)\bar{r}_m[k]$ and associates the user to the appropriate cell based on the predicted average delays [8]. In more detail, based on that rule, the user is associated to the picocell iff $\bar{\mathcal{D}}_m \gg \bar{\mathcal{D}}_{p_{j^*}}$, i.e., $f\left(\bar{\mathcal{D}}_m^k, \bar{\mathcal{D}}_{p_{j^*}}^k\right) \to 0^+$ while in the macrocell iff $\bar{\mathcal{D}}_m \ll \bar{\mathcal{D}}_{p_{j^*}}$, i.e., $f\left(\bar{\mathcal{D}}_m^k, \bar{\mathcal{D}}_{p_{j^*}}^k\right) \to +\infty$. Further, similar rules can be used in picocell-to-picocell HO scenarios without loss of generality. Towards this direction, more advanced HO rules can be applied by taking into account transient phenomena that are crucial in mobility scenarios, especially when the sojourn time in a cell (function of UE speed) is not sufficiently enough to provide system stationarity. Subsequently, the prediction of the service delay becomes complicated and switch to transient analysis tools is required. Finally, paving the path to 5G beyond the HO (RSS-based) legacy presented in Chapter 2, this chapter discusses an SDN-based programmable framework to centrally control the HO process. It can be noticed that UE receives the predicted delay for each cell by its serving cell in order to trigger the association procedure defined by the load-aware HO algorithm (see Fig. 3.2). Hence, that algorithm can directly run on the SDN controller side that is aware of the users RSRP such that to send only the cell association results to the interrelated users offloading the signaling overhead and bypassing any modifications in the UE protocol stack for the SDN protocol support.

3.7 Conclusion

This chapter focuses on user-centric HO decision algorithms for HetNets. It was shown that the proposed load-aware algorithm can significantly enhance the system performance by considering both user and network perspective in order to improve user QoS. To this end, it was compared with a conventional and a distance-based LTE-handover algorithm and we showed that it can outperform both of them according to the extracted cell assignment probability and user service delay performance results.

[8] Such load-aware scaling can be considered as a dynamic cell range extension showing a delay gain compared to the DIST/CONV algorithms.

Chapter 4

Utility-Based Resource Allocation under Multi-Connectivity in LTE

4.1 Introduction

In the previous chapters, we cover the topic of mobility management towards next-generation cellular networks. More precisely, we present the analytical insights about the involved trade-offs on handover decision algorithms for next-generation HetNets. The comparisons with the conventional methodologies reveal the new network trends that are met in heterogeneous environments. To this end, 5G communication systems are expected to fulfill the insatiable and heterogeneous service requirements in future deployments.

As stressed in Chapter 1, toward the development of 5G vision, it is expected that the mobile broadband service will be enhanced to provide a consistent user experience [75]. Considering the current cellular technologies, cell edge users and those experiencing high interference suffer from poor service, even when coordinated signal processing is applied. Along this direction, the coordination of multiple connections is expected to be a solution by re-designing radio/network functions [75]. To overcome such limitation, multi-connectivity was introduced in 5G as a mechanism of supporting multiple cell connections per end-user basis on several radio access technologies (RAT) or frequency bands [76, 77].

The multi-connectivity concept is characterized by effective resource utilization for seamless user experience [78], enhancement among capacity, coverage and mobility [79], reliable high-speed data delivery [80] and acting as a quick fail-over method [19]. In general, the multiple connections can be applied among heterogeneous inter-RAT or intra-RAT [81] or within a single RAT that is viewed as establishing multiple connections to different BSs. The dual-connectivity (DC) concept was introduced in LTE Rel.12 study item (3GPP TR36.842) as the baseline of multi-connectivity enabling two simultaneous connections across the end-users in an intra-RAT, i.e., transmit/receive data simultaneously from two distinct BSs in uplink (UL) and/or downlink (DL).

Despite its appealing, once of the significant challenge of multi-connectivity is presented in [24] related to the efficient radio resource utilization. Towards this direction, the current and the following chapters explore the new problem of resource allocation under multi-connectivity. The resource allocation in multi-connectivity is crucial to enhance the user performance and to deal with the increasing demand of traffic. In general, two types of application traffic exist, (i)

user-to-user such as content/video dissemination sharing, peer-to-peer gaming and public safety, and (ii) user-to-network and network-to-user such as social networking and video-on-demand. Nevertheless, another challenge is related to satisfy the quality of service (QoS) requirement of each traffic flow and optimize network throughput across multiple connections.

The work in [82] proposes an hierarchical architecture to control and coordinate multi-connectivity with resource allocation and interference management. In [83], the authors consider resource allocation at the macro node side based on DC enabled downlink (DL) connections supported by small cells. While these works focus on resource allocation schemes in multi-connectivity environment, users QoS-aware utility proportional fairness allocation schemes [20] are not proposed. In that sense, this chapter addresses the new problem of resource allocation under multi-connectivity case with user-to-user traffic and QoS requirement. The contributions of this chapter are summarized as follows:

1. We introduce the proportional fairness resource allocation problem that aims to maximize network aggregated throughput for multi-connectivity and compare with legacy single-connected case under different scenarios.

2. We propose a utility-based resource allocation that considers the QoS and analyze the performance gain with aforementioned proportional fairness resource allocation.

3. Finally, we investigate the impact of utility function on QoS satisfaction in terms of its shape.

The rest of the chapter is organized as follows. Section 4.2 provides the system model and assumptions used through this chapter. Section 4.3 formulates the optimization problem. Simulation results are discussed in Section 4.4. Section 4.5 offers a brief discussion. Finally, Section 4.6 presents the concluding remarks.

4.2 System model and assumptions

In this section, we describe the system model and the assumptions. We consider an area $\mathcal{L} \subset \mathbb{R}^2$ served by a set of BSs $\mathcal{B} = \{b_1, \cdots, b_{|\mathcal{B}|}\}$ and a set of UEs $\mathcal{U} = \{u_1, \cdots, u_{|\mathcal{U}|}\}$ is distributed in this area. For instance, in Fig. 4.1, an example is presented with $\mathcal{B} = \{b_1, b_2, b_3\}$ and $\mathcal{U} = \{u_1, u_2, u_3, u_4\}$. In following, we examine our model in more detail:

4.2.1 Air-interface model

A.1-Mobility: We assume user mobility, implying that all location related parameters are changing in time. For simplicity, we skip the time index t throughout the chapter.

A.2-Signal to interference plus noise ratio (SINR): The SINR of the received signal from the j-th BS (b_j) to the i-th UE (u_i) per Physical Resource Block (PRB) along the DL direction is denoted as:

$$\mathrm{SINR}^{\mathrm{D}}_{b_j,u_i} = \frac{\mathrm{RSRP}^{\mathrm{D}}_{b_j,u_i}}{\sum\limits_{b_k \neq b_j} p^{\mathrm{D}}_{b_k,u_i} \mathrm{RSRP}^{\mathrm{D}}_{b_k,u_i} + W^{\mathrm{D}}_{b_j} N_0}. \tag{4.1}$$

Respectively, the SINR from the i-th UE to the j-th BS in UL direction is denoted as $\mathrm{SINR}^{\mathrm{U}}_{u_i,b_j}$.

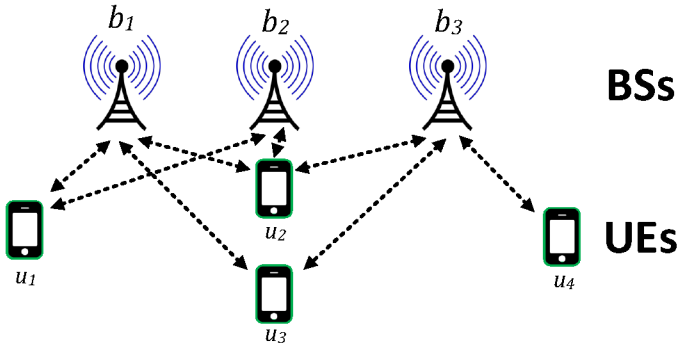


Figure 4.1: Multi-connectivity example

The Reference Signal Received Power (RSRP) is as $\text{RSRP}^{\text{D}}_{b_j,u_i} = L^{\text{D}}_{b_j,u_i} P^{\text{D}}_{b_j,u_i} G^{\text{D}}_{b_j,u_i}$ that includes the path loss and shadowing $L^{\text{D}}_{b_j,u_i}$ from the j-th BS to the i-th UE in DL ($L^{\text{U}}_{u_i,b_j}$ for UL), the transmitted power of the j-th BS to the i-th UE $P^{\text{D}}_{b_j,u_i}$ ($P^{\text{U}}_{u_i,b_j}$ for UL), and the combined antenna gain of the j-th BS and the i-th UE $G^{\text{D}}_{b_j,u_i}$ in the DL ($G^{\text{U}}_{u_i,b_j}$ for UL). The N_0^{dBm} [1] stands for the thermal noise density in dBm per Hz and $W^{\text{D}}_{b_j}$ is the j-th BS DL bandwidth per PRB in Hz, such that their product is the aggregated noise power per PRB for DL ($W^{\text{U}}_{b_j}$ for UL). In addition, high frequency fluctuations (i.e., Rayleigh fading) are assumed to be filtered and equalized.

Further, the RSRP from other BSs ($b_k \neq b_j$) in the denominator is assumed to be dependent on the *PRB overlapping probability* as $p^{\text{D}}_{b_k,u_i}$ for the i-th UE. In general, we assume the PRB allocation at each BS is uniformly distributed across all PRBs, so the *PRB overlapping probability* is defined as the summation of allocated PRB to all other UEs ($u_q \neq u_i$) in percentage ($p^{\text{U}}_{u_i,b_k}$ for UL):

$$p^{\text{D}}_{b_k,u_i} = \sum_{u_l \in \mathcal{U}} \sum_{u_q \neq u_i} x^{\star \text{D}}_{b_k,(u_l,u_q)}, \tag{4.2}$$

where $x^{\star \text{D}}_{b_k,(u_l,u_q)}$ is the percentage of allocated PRBs by the k-th BS to user-to-user traffic of user pair (u_l, u_q) along the DL that will be elaborated in **B.3** ($x^{\star \text{U}}_{b_k,(u_q,u_l)}$ for UL).

A.3-Physical data rate: The j-th BS can deliver a maximum physical data transmission rate $R^{\text{D}}_{b_j,u_i}$ to a UE. The physical data rate along the DL in bits per second (bps) is given in Eq. (4.3) based on the Shannon capacity formula:

$$R^{\text{D}}_{b_j,u_i} = B^{\text{D}}_{b_j} W^{\text{D}}_{b_j} \log_2 \left(1 + \text{SINR}^{\text{D}}_{b_j,u_i}\right), \tag{4.3}$$

where $B^{\text{D}}_{b_j}$ is the total number of DL PRBs of the j-th BS ($B^{\text{U}}_{b_j}$ for UL). In respect, the physical data rate from the i-th UE to the j-th BS in UL direction is as $R^{\text{U}}_{u_i,b_j}$.

A.4-Power Control: The open-loop power control is applied in UL and each UE compensates the path loss $L^{\text{U,dB}}_{u_i,b_j}$ and shadowing effects based on the power control parameters (i.e., α, P_0^{dBm}). Thus, the transmitted power of each PRB from the i-th UE to the j-th BS is:

$$P^{\text{U,dBm}}_{u_i,b_j} = \min\left(P^{\text{max,dBm}}_{u_i}, P_0^{\text{dBm}} + \alpha \cdot L^{\text{U,dB}}_{u_i,b_j}\right), \tag{4.4}$$

[1] The notation dB or dBm is skipped when we refer to the linear scale of any quantities.

where $P_{u_i}^{\max,\mathrm{dBm}}$ is the maximum transmitted power of the i-th UE. However, in the DL, no power control algorithm is applied and the transmitted power from each BS to all UEs is denoted as $P_{b_j,u_i}^{\mathrm{D,dBm}} = P_{b_j}^{\mathrm{D,dBm}}$. It is noted that the asymmetry in UL/DL transmitted powers due to the UL power control differentiates the UL/DL SINR values (see assumption **A.2**).

4.2.2 Connection and Traffic model

B.1-Multi-connectivity: Under multi-connectivity, users can be associated to and communicate with more than one BSs at the same time. We assume the multi-connectivity capability exists in both DL and UL for all UEs and a UE can be connected to a BS if both DL and UL SINR is above a pre-defined threshold, i.e., $\mathrm{SINR}_{\mathrm{th}}$. Hence, we define a set $\mathcal{E} \triangleq \{(u_i, b_j), (b_j, u_i) : \min\left(\mathrm{SINR}_{u_i,b_j}^{\mathrm{U}}, \mathrm{SINR}_{b_j,u_i}^{\mathrm{D}}\right) > \mathrm{SINR}_{\mathrm{th}}\}$ that represents all possible connections between UEs and BSs. Further, no connection can be established when the condition $\min\left(\mathrm{SINR}_{u_i,b_j}^{\mathrm{U}}, \mathrm{SINR}_{b_j,u_i}^{\mathrm{D}}\right) > \mathrm{SINR}_{\mathrm{th}}$ is not hold for the i-th UE and the j-th BS. Without loss of generality, asymmetric connection scenarios among UE and BSs, i.e., only UL or DL connection to one cell, can be allowed, but in such cases multi-cell connectivity can be limited, especially under bi-directional traffic presence.

B.2-Local-routing: In principle, routing in the backhaul network is necessary even for the user-to-user traffic served by the same BS. To offload the backhaul/core network, the concept of local-routing is applied [84] in which the user-to-user traffic is routed directly via intermediate BS, thus offloading the core and backhaul network, and reducing the number of hops taken by IP packets to reach the destination [85]. In this case, traffic is not routed via the core network, e.g., LTE evolved packet core (EPC), but only via the intermediate BS [86]. We focus on the user-to-user traffic flow that can be locally-routed, i.e., both user are connected to at least one common BS[2]. Such traffic flow is representative of the public safety (e.g., isolated BSs) and close community application (e.g., community-based video sharing) scenarios where content is shared locally among UEs. Thus, backhaul-routed case is not considered [3].

B.3-Active users pairs: Based on the minimum SINR requirement defined in **B.1**, a set that comprises all active user pairs served by the j-th BS is defined as $\mathcal{C}_{b_j} \triangleq \{(u_i, u_q) : (u_i, b_j), (b_j, u_q) \in \mathcal{E}, u_i \neq u_q\}$ and the user pair $(u_i, u_q) \in \mathcal{C}_{b_j}$ can have user-to-user traffic routed locally via BS b_j. Consequently, a set $\mathcal{C} \triangleq \bigcup_{b_j \in \mathcal{B}} \mathcal{C}_{b_j}$ is formed as the union of all active user pairs. Lastly, two sets are further defined for each user: $\mathcal{D}_{u_i} \triangleq \{u_q : (u_i, u_q) \in \mathcal{C}\}$ comprises all destined UEs from the i-th UE and $\mathcal{S}_{u_i} \triangleq \{u_q : (u_q, u_i) \in \mathcal{C}\}$ comprises all source UEs that can transport traffic to the i-th UE.

B.4-Traffic flow requested rate: It corresponds to the requested rate $\hat{R}_{u_i,u_q}$ determined as the total rate required by a group of U2U applications running on the top using an end-to-end established connection of user pair (u_i, u_q), that can go through any intermediate BS via local-routing[4].

4.3 Problem setup

Based on aforementioned system model, we formulate the optimization problem in this section.

[2]In Fig.4.1, the traffic from u_1 to u_2 can be local routed via b_1 or b_2.

[3]In Fig.4.1, the traffic from u_1 to u_4 is not considered.

[4]In Fig. 4.1, requested rate $\hat{R}_{u_1,u_2}$ and $\hat{R}_{u_2,u_1}$ can go through b_1 and b_2,

4.3.1 Utility function

Our objective here is to allocate optimally the resource to user-to-user traffic based on the applied utility function. We define $x^{\mathrm{U}}_{b_j,(u_i,u_q)}, x^{\mathrm{D}}_{b_j,(u_i,u_q)} \in \mathcal{X}$ as the percentage of allocated PRB to total PRBs in decimal form along UL/DL direction to transport user-to-user traffic of user pair (u_i, u_q) through BS b_j. It is noted that $x^{\mathrm{U}}_{b_j,(u_i,u_q)} = x^{\mathrm{D}}_{b_j,(u_i,u_q)} = 0$, if $(u_i, u_q) \notin \mathcal{C}_{b_j}$, i.e., no resource is allocated if such user pair can not be local routed through BS b_j. In the following, we introduce two different utility functions: one provides proportional fairness and the other one extends the proportional fairness by taking QoS into consideration.

4.3.1.1 Proportional Fairness (PF)

As the first utility function, we exploit a logarithmic utility function similar to the one in [87] that achieves "proportional fair" resource allocation. The "proportional fair" characteristic implies that if we select a data rate greater than the optimal one for a given user, then there will be at least one another data rate that will be decreased in a proportion larger than the proportion of the increased user data rate [88]. Such utility function is given as:

$$\Phi\left(y\right) = \log\left(y\right), \tag{4.5}$$

4.3.1.2 Utility Proportional Fairness (UPF)

In the former utility function, a significant drawback lies on the lack of QoS consideration. To incorporate the QoS into the utility function, we use the family of sigmoid functions as:

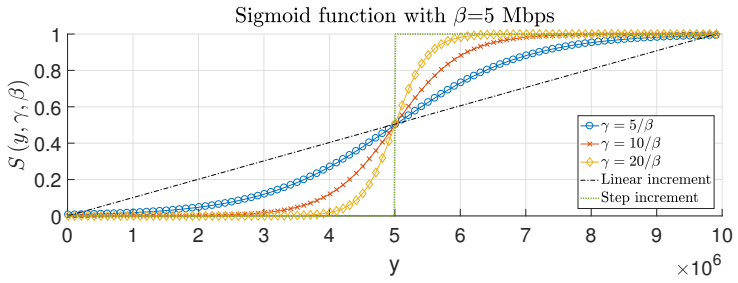
$$\Phi\left(y\right) = \log\left(S\left(y,\gamma,\beta\right)\right) = \log\left(\frac{1}{1+e^{-\gamma(y-\beta)}}\right), \tag{4.6}$$

$\gamma > 0$, $y \geq 0$, where β is the traffic requested rate defined as introduced in **B.3** of Section 4.2. The function $\Phi\left(y\right)$ is the natural logarithm of the sigmoid utility function, $S\left(y,\gamma,\beta\right)$, commonly used in literature [89] that preserves the traffic requested rate. In Fig. 4.2(a), the family of sigmoid function with different γ is compared with the linear increment and step increment function for $\beta = 5$ Mbps. If the allocated data rate y is less than the requested rate β, the sigmoid function family provides monotonic increment on its slope as shown in Fig. 4.2(b) whereas its slope is monotonic decreasing when the requested rate is achieved (i.e., $y > \beta$). Moreover, the value of γ impacts the shape of sigmoid function to be more like step or linear function that reflects the demand degree of requested rate. In other words, it echoes the trade-off between the network throughput maximization provided by the "linear" characteristic (better in the network perspective) and the QoS satisfaction by the "non-linear" characteristic (better in the user perspective).

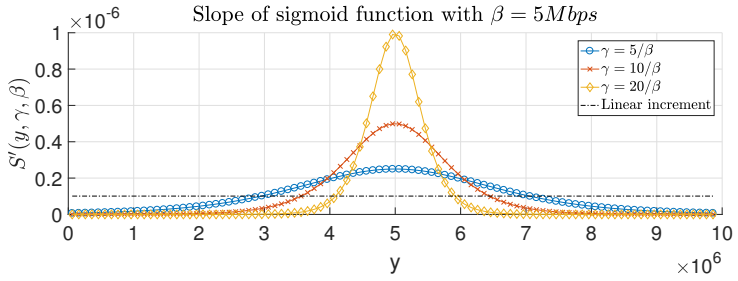
Under the multi-connectivity regime assumption (see **B.1** of Section 4.2), we can introduce our final objective function based on any utility function $\Phi\left(\cdot\right)$ presented in Eq. (4.5) or Eq. (4.6) as following:

$$U\left(\mathbf{x}^{\mathrm{U}}_{u_i,u_q}, \mathbf{x}^{\mathrm{D}}_{u_i,u_q}\right) \triangleq \Phi\left(\sum_{b_j \in \mathcal{B}} Q\left(x^{\mathrm{U}}_{b_j,(u_i,u_q)}, x^{\mathrm{D}}_{b_j,(u_i,u_q)}\right)\right), \tag{4.7}$$

where $\mathbf{x}^{\mathrm{U}}_{u_i,u_q} \triangleq \left[x^{\mathrm{U}}_{b_1,(u_i,u_q)}, \cdots, x^{\mathrm{U}}_{b_{|B|},(u_i,u_q)}\right]^T$, $\mathbf{x}^{\mathrm{D}}_{u_i,u_q} \triangleq \left[x^{\mathrm{D}}_{b_1,(u_i,u_q)}, \cdots, x^{\mathrm{D}}_{b_{|B|},(u_i,u_q)}\right]^T$.



(a) Sigmoid function



(b) Slope of sigmoid function

Figure 4.2: Shape of sigmoid function

The $Q(\cdot)$ function stands for the allocated rate of user pair (u_i, u_q) and is expressed as follows:

$$Q\left(x^{\mathrm{U}}_{b_j,(u_i,u_q)}, x^{\mathrm{D}}_{b_j,(u_i,u_q)}\right) \triangleq \min\left(x^{\mathrm{U}}_{b_j,(u_i,u_q)} R^{\mathrm{U}}_{u_i,b_j}, x^{\mathrm{D}}_{b_j,(u_i,u_q)} R^{\mathrm{D}}_{b_j,u_q}\right) \tag{4.8}$$

It is noted that the minimum operator guarantees the same allocated rate in both UL/DL directions for a single user-to-user traffic [5]. Specifically, each user cannot send or receive with an allocated rate greater than another user can send or receive. No mechanism support of this guarantee implies UL/DL buffering that renders the current model more complex introducing queuing theory analysis tools and is out of the scope of this study. Finally, the argument of $\Phi(\cdot)$ in Eq. (4.7) takes the aggregated allocated rate of user pair (u_i, u_q) over all common BSs (i.e., $\forall\ b_j \in \mathcal{B}$).

4.3.2 Problem formulation

Based on the proposed system model, assumptions and utility functions, we present our problem formulation. The problem falls into the category of the network utility maximization for resource allocation and is given in Eq. (4.9). A detailed explanation of the proposed optimization problem follows, presenting both the objective function and constraints.

Objective function: The objective is to allocate the resource $x^{\mathrm{U}}_{b_j,(u_i,u_q)}, x^{\mathrm{D}}_{b_j,(u_i,u_q)} \in \mathcal{X}$ to each user pair in order to maximize the aggregated network utility function over all user pairs that exchange traffic. The two aforementioned utility functions can be utilized.

[5] Neither the user-to-network traffic nor the network-to-user traffic applies such minimum operation.

$$
\begin{aligned}
\max_{\mathcal{X}} \quad & \sum_{(u_i,u_q)\in\mathcal{C}} U\left(\mathbf{x}^{\mathrm{U}}_{u_i,u_q}, \mathbf{x}^{\mathrm{D}}_{u_i,u_q}\right) \\
\text{s.t.} \quad & \sum_{(u_i,u_q)\in\mathcal{C}_{b_j}} x^{\mathrm{U}}_{b_j,(u_i,u_q)} \leq 1,\ \forall\ b_j, \\
& \sum_{(u_i,u_q)\in\mathcal{C}_{b_j}} x^{\mathrm{D}}_{b_j,(u_i,u_q)} \leq 1,\ \forall\ b_j, \\
& \sum_{b_j\in\mathcal{B}} \sum_{u_q\in\mathcal{D}_{u_i}} x^{\mathrm{U}}_{b_j,(u_i,u_q)} B^{\mathrm{U}}_{b_j} \leq B^{\mathrm{U}}_{u_i},\ \forall\ u_i, \\
& \sum_{b_j\in\mathcal{B}} \sum_{u_q\in\mathcal{S}_{u_i}} x^{\mathrm{D}}_{b_j,(u_q,u_i)} B^{\mathrm{D}}_{b_j} \leq B^{\mathrm{D}}_{u_i},\ \forall\ u_i, \\
& \sum_{b_j\in\mathcal{B}} \sum_{u_q\in\mathcal{D}_{u_i}} x^{\mathrm{U}}_{b_j,(u_i,u_q)} P^{\mathrm{U}}_{u_i,b_j} B^{\mathrm{U}}_{b_j} \leq P^{\max}_{u_i},\ \forall\ u_i.
\end{aligned} \tag{4.9}
$$

Constraints: Here, we present a detailed description for each one of them:

1. The first two constraints ensure that the number of allocated PRBs (expressed as the percentage in decimal form on total PRBs) at each BS b_j to all UEs will not exceed the total number of PRBs in both UL/DL.

2. The third and fourth constraints assure that the number of allocated PRBs to each UE among all connected BSs will not exceed the maximum number of allocated PRBs, $B^{\mathrm{U}}_{u_i}, B^{\mathrm{D}}_{u_i}$, in UL/DL of the i-th UE, respectively. Specifically, these constraints take into account all user-to-user traffic related to the i-th UE as $u_q \in \mathcal{D}_{u_i}$ of all destinate UEs in UL and $u_q \in \mathcal{S}_{u_i}$ of all source UEs in DL through any intermediate BS $(b_j \in \mathcal{B})$. Further, these constraints lie on the user capability defined as UE-category in 3GPP TS36.306.

3. Finally, the last constraint is related to the power control mechanism introduced in **A.4** of Section 4.2. It restricts the total transmitted power from the i-th user to all connected BSs to be within its power limitation as $P^{\max}_{u_i}$.

Finally, we note that the objective function contains the minimum operation that is concave but non-differentiable. Thus, we need to transform the objective function to a differentiable concave one in order to conclude to a unique tractable global optimal. The following paragraph describes such procedure.

4.3.3 Problem transformation

To deal with the minimum operation in Eq. (4.8), we replace $Q\left(x^{\mathrm{U}}_{b_j,(u_i,u_q)}, x^{\mathrm{D}}_{b_j,(u_i,u_q)}\right)$ with the auxiliary variable $z_{b_j,(u_i,u_q)} \in \mathcal{Z}$ and add two extra constraints $z_{b_j,(u_i,u_q)} \leq x^{\mathrm{U}}_{b_j,(u_i,u_q)} R_{u_i,b_j}$ and $z_{b_j,(u_i,u_q)} \leq x^{\mathrm{D}}_{b_j,(u_i,u_q)} R_{b_j,u_q}$. Then, we define $U\left(\mathbf{z}_{u_i,u_q}\right) \triangleq \Phi\left(\sum_{b_j\in\mathcal{B}} z_{b_j,(u_i,u_q)}\right)$, $\mathbf{z}_{u_i,u_q} \triangleq \left[z_{b_1,(u_i,u_q)}, \cdots, z_{b_{|\mathcal{B}|},(u_i,u_q)}\right]^T$ and the transformed problem in Eq. (4.10). The objective function in the transformed optimization problem is strictly concave as the sum of $U\left(\mathbf{z}_{u_i,u_q}\right)$ is strictly concave as proved in the Lemma 1.

$$
\begin{aligned}
\max_{\mathcal{X},\mathcal{Z}} \quad & \sum_{(u_i,u_q)\in\mathcal{C}} U\left(\mathbf{z}_{u_i,u_q}\right) \\
\text{s.t.} \quad & z_{b_j,(u_i,u_q)} \leq x^{\mathrm{U}}_{b_j,(u_i,u_q)} R^{\mathrm{U}}_{u_i,b_j},\ \forall\ b_j, \\
& z_{b_j,(u_i,u_q)} \leq x^{\mathrm{D}}_{b_j,(u_i,u_q)} R^{\mathrm{D}}_{b_j,u_q},\ \forall\ b_j, \\
& \sum_{(u_i,u_q)\in\mathcal{C}_{b_j}} x^{\mathrm{U}}_{b_j,(u_i,u_q)} \leq 1,\ \forall\ b_j, \\
& \sum_{(u_i,u_q)\in\mathcal{C}_{b_j}} x^{\mathrm{D}}_{b_j,(u_i,u_q)} \leq 1,\ \forall\ b_j, \\
& \sum_{b_j\in\mathcal{B}} \sum_{u_q\in\mathcal{D}_{u_i}} x^{\mathrm{U}}_{b_j,(u_i,u_q)} B^{\mathrm{U}}_{b_j} \leq B^{\mathrm{U}}_{u_i},\ \forall\ u_i, \\
& \sum_{b_j\in\mathcal{B}} \sum_{u_q\in\mathcal{S}_{u_i}} x^{\mathrm{D}}_{b_j,(u_q,u_i)} B^{\mathrm{D}}_{b_j} \leq B^{\mathrm{D}}_{u_i},\ \forall\ u_i, \\
& \sum_{b_j\in\mathcal{B}} \sum_{u_q\in\mathcal{D}_{u_i}} x^{\mathrm{U}}_{b_j,(u_i,u_q)} P^{\mathrm{U}}_{u_i,b_j} B^{\mathrm{U}}_{b_j} \leq P^{\max}_{u_i},\ \forall\ u_i
\end{aligned} \tag{4.10}
$$

Lemma 1. *The utility function $U\left(\mathbf{z}_{u_i,u_q}\right)$ in the transformed optimization problem is strictly concave.*

Proof. It is noticed that $U\left(\mathbf{z}_{u_i,u_q}\right)$ function can be written as:

$$
U(\mathbf{z}_{u_i,u_q}) = \Phi\left(\mathbf{1}^T_{|\mathcal{B}|}\mathbf{z}_{u_i,u_q}\right), \tag{4.11}
$$

and any utility function $\Phi(\cdot)$ is strictly concave. The first one is the logarithmic in Eq. (4.5) and the second one in Eq. (4.6) is proved to be strictly concave as the logarithm of the sigmoid function based on Lemma III.1 in [89]. It is known that any composition of a concave function with an affine function is concave [90]. Thus, as expressed in Eq. (4.11), we conclude that $U(\mathbf{z}_{u_i,u_q})$ is also concave. □

Combined with the linear constraints, the transformed optimization problem is a convex optimization problem with a unique tractable optimal solution.

4.4 Simulations

In this section, the performance evaluation results are presented for the optimization problem described in Section 4.3. Simulation parameters applied to UEs, BSs and network planning are mostly taken from 3GPP (TR36.814, TR36.942, TR25.942) and NGMN documents [91], and some important parameters are listed in Table 4.1 [6]. Moreover, all assumptions introduced in

[6] Without loss of generality, we assume the (maximum) BS Tx power that corresponds to a macrocell. Hence, the current work is fully applicable for small cells deployment. The same assumption holds for Chapter 5 and 6.

Table 4.1: Simulation parameters

Parameter	Value
LTE mode	FDD, SISO
Carrier frequency	DL: 2.19 GHz; UL: 2.0 GHz
Total PRBs of each BS	100 (20 MHz BW)
Maximum PRBs of each UE	100
Number of BSs	3
Initial UE distribution	Uniform of each BS
UE speed	Selected from $[3, 30, 120]$ km/h as [91]
UE direction	Uniform distributed in $[0, 360]$ degree
UE traffic model	Full buffer
SINR threshold	-10 dB
BS transmission power	46 dBm
UE maximum transmission power	23 dBm
Power control parameters	P_0^{dBm} = -58 dBm, $\alpha = 0.8$
Thermal noise density	-174 dBm/Hz

Table 4.2: Comparison of Single/Multi-connectivity

UE number in BS $b_1/b_2/b_3$	**Performance metric**	**Single-connected**	**Multi-connected**
Under-loaded case: 2/2/2	Connected BS	1	2.07
	Connected UE pairs	6	17.52
	Aggregated user rate	0.99 Mbps	20.04 Mbps
Uneven-loaded case: 6/2/2	Connected BS	1	1.34
	Connected UE pairs	34	49.95
	Aggregated user rate	11.68 Mbps	46.73 Mbps
Over-loaded case: 6/6/6	Connected BS	1	1.45
	Connected UE pairs	90	162.22
	Aggregated user rate	55.04 Mbps	57.01 Mbps

Section 4.2 are also held. The whole simulation framework is built in MATLAB [7] and we use the interior point algorithm to iteratively solve the linear-constrained convex optimization problem of both utility functions: PF and UPF.

4.4.1 Comparison: Single-connectivity and Multi-connectivity

Firstly, we compare the performance of legacy single-connectivity case with the multi-connectivity one using the same PF utility function. The considered legacy single-connected case associates each UE to only one BS and only allows each UE to communicate with the associated BS in both UL/DL directions in terms of the best received RSRP. In Table 4.2, we compare them in terms

[7] Our self-developed MATLAB LTE/LTE-A simulation platform is based on the 3GPP specifications and NGMN documents to provide a system level view of the whole mobile network. It can generate the network map via planning several BSs and mobile UEs, emulate the non-idealities of wireless channel model and provide a network-wide inspection on the user and network performance. The same simulator is used in Chapter 5 and 6.

Table 4.3: QoS metrics comparison of PF and UPF

Metric	Requested rate	PF problem	UPF problem
Satisfaction ratio	0.1Mbps	68.72%	91.39%
	0.5Mbps	42.07%	58.03%
	1Mbps	25.15%	35.33%
	5Mbps	6.42%	23.10%
	10Mbps	$< 1\%$	<1%
Unsatisfied normalized error	0.1Mbps	0.2122	0.0625
	0.5Mbps	0.4131	0.2317
	1Mbps	0.5483	0.3906
	5Mbps	0.7949	0.6607
	10Mbps	0.8833	0.8441

of the average number of connected BS, number of connected user pairs, and aggregated user rate. Three representative scenarios, namely under-loaded, uneven-loaded and over-loaded, are presented, where the $x/y/z$ notation represents the initially uniformly distributed of $x/y/z$ UEs over each BS. It is noted that the x, y, z variables variation creates the aforementioned scenarios related to the network load.

It can be seen from the table that the number of connected BS per UE increases with multi-connectivity especially for under-loaded scenario. This is because of the induced inter-cell interference is minor for under-loaded scenario. Moreover, we observe a higher number of connected user pair in multi-connectivity as each UE is able to transmit and receive traffic to a larger set of UEs through the local routing across different BSs. This advantage becomes significant of over-loaded scenario that allows more traffic diversity among users. Even the number of UEs increases, the core network signaling overhead is reduced benefiting from the local-routing adaption of the current network setup.

When comparing the aggregated user rate of all user pairs, we notice that the performance gain is significantly higher in under-loaded scenario followed by the uneven-loaded scenario. This gain is due to schedule UEs across all available BSs, that is indeed one of the expected merit of multi-connectivity. To sum up, the multi-connectivity not only has advantage in user perspective (i.e., more UEs can be reached through multiple BSs) but also in the network perspective (i.e., larger aggregated user rate) in different loading scenarios.

4.4.2 Performance analysis of PF & UPF in multi-connectivity

Then, we present the results of both utility functions with a fixed $\gamma = 10/\hat{R}$ in a scenario with 4 UEs over each BS (i.e., 4/4/4). Firstly, we define the final allocated rate of the user pair (u_i, u_q) among all intermediate BSs as

$$\zeta_{u_i,u_q} \triangleq \sum_{b_j \in \mathcal{B}} Q\left(x^{\star \mathrm{U}}_{b_j,(u_i,u_q)}, x^{\star \mathrm{D}}_{b_j,(u_i,u_q)}\right), \tag{4.12}$$

where $x^{\star \mathrm{U}}_{b_j,(u_i,u_q)}, x^{\star \mathrm{D}}_{b_j,(u_i,u_q)}$ are the optimization results of control variables $x^{\mathrm{U}}_{b_j,(u_i,u_q)}, x^{\mathrm{D}}_{b_j,(u_i,u_q)}$. Further, in a quantitative comparison on QoS, we define two different metrics: (i) Satisfaction ratio that represents the ratio of user pairs which are satisfied with the allocated rate represented

by the CDF [8] in (4.13), and (ii) Unsatisfied normalized error that shows the normalized Euclidean distance between the allocated rate ζ_{u_i,u_q} and the requested rate $\hat{R}_{u_i,u_q}$ when a user pair (u_i, u_q) is unsatisfied in (4.14).

$$M_{u_i,u_q} = \text{Prob}\left\{\zeta_{u_i,u_q} \geq \hat{R}_{u_i,u_q}\right\}. \tag{4.13}$$

$$E_{u_i,u_q} = \begin{cases} \left\| \frac{\zeta_{u_i,u_q} - \hat{R}_{u_i,u_q}}{\hat{R}_{u_i,u_q}} \right\|, & \text{if } \zeta_{u_i,u_q} < \hat{R}_{u_i,u_q}, \\ 0 \qquad\qquad\qquad\quad , & \text{o/w.} \end{cases} \tag{4.14}$$

Table 4.3 shows the results of the defined QoS metrics using both PF and UPF utility functions with five different requested rate $\hat{R} = \hat{R}_{u_i,u_q}$, $\forall\,(u_i, u_q) \in \mathcal{C}$. In terms of the satisfaction ratio, the UPF is much better than the PF one except in $\hat{R} = 10$ Mbps case in which both utility functions satisfy less than 1% of user pairs due to inter-cell interference impact. Further, UPF reduces the unsatisfied normalized error by allocating resources as close as possible to the requested rate. We can see that even the QoS requirement cannot be satisfied for some user pairs mostly in overloaded scenarios, the UPF still provides less error to the requested rate.

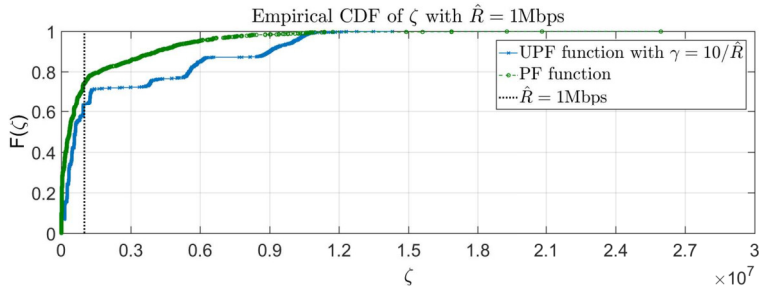
In more qualitative comparison, in Fig. 4.3, the CDF plot of the allocated rate to all user pairs with three different requested rates are shown. The ratio of satisfied user pair is higher for the UPF case in Fig. 4.3(a) and Fig. 4.3(b) and is almost the same of both UPF and PF in Fig. 4.3(c) that matches the satisfaction ratio in Table 4.3. For instance, the satisfaction ratio $M_{u_i,u_q} = 1 - 0.65 = 0.35$ of the UPF case in Fig. 4.3(a); however, $M_{u_i,u_q} = 0.25$ of the PF case (CDF is 0.75 at this point and it means 75% of user pairs are unsatisfied). We observe that the PF has the same CDF among different requested rate $\hat{R}$ (i.e., independent of $\hat{R}$) and possesses a longer tail due to the fact that it only maximizes the network throughput without considering any QoS requirements. In other words, UPF provides less network aggregated rate improving user pairs QoS, i.e., shorter tail considering $\hat{R}$, compared to PF that targets to maximize the network aggregated rate, i.e., longer tail and unconstrained of $\hat{R}$.

4.4.3 Impact of γ on UPF problem

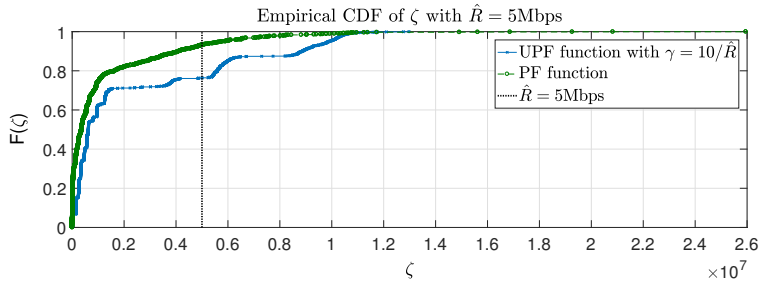
Fig. 4.4 compares the impact of γ on the UPF utility function in terms of the PDF plot of allocated rate ζ_{u_i,u_q} for all user pairs. For simplicity, we only provide the result with $\hat{R} = 1$ Mbps but the same phenomena can be observed for other requested rates. Firstly, it can be observed that the tail of PDF plot is longer with smaller γ, i.e., the tail is the longest of the three when $\gamma = 5/\hat{R}$. That is because the sigmoid function with smaller γ tends to be more linear and makes the allocated rate exceed the requested rate as shown in Fig. 4.2(a). The latter approach is better from the network perspective.

Moreover, we observe a significant amount of user pairs are with smaller allocated rate ($\zeta_{u_i,u_q} <$ 0.4 Mbps) when γ is large, $\gamma = 20/\hat{R}$, compared to the case when γ is small, $\gamma = 5/\hat{R}$. This is also due to the shape of the sigmoid function in Fig. 4.2(a) in which the function with larger γ is more like the step function and prefers to serve the user pair that is close to the QoS requirement. In that sense, some user pairs that are struggle to achieve the requested rate due to the poor SINR condition will be allocated with a smaller data rate or even be queued (e.g., $\zeta_{u_i,u_q} = 0$).

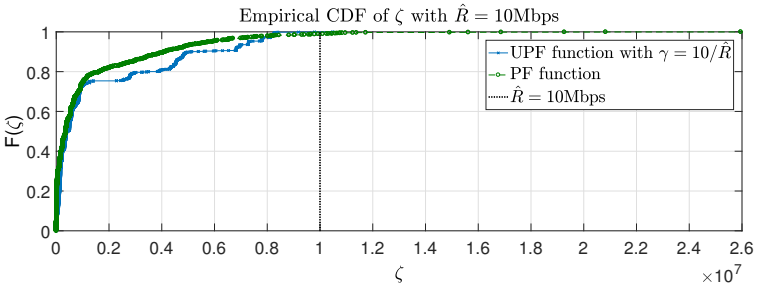
[8] The term ratio is corresponding to the $\frac{\zeta_{u_i,u_q}}{\hat{R}_{u_i,u_q}}$ related to the event $\frac{\zeta_{u_i,u_q}}{\hat{R}_{u_i,u_q}} \geq 1$ that implies satisfaction.



(a) $\hat{R}$ = 1 Mbps



(b) $\hat{R}$ = 5 Mbps



(c) $\hat{R}$ = 10 Mbps

Figure 4.3: CDF plot of PF/UPF utility functions with different $\hat{R}$

The latter approach is better from the user perspective. Consequently, the adjustment of sigmoid function shape based on the γ parameters can impact the allocation policy highlighting the trade-off of fairness from the user and network perspective.

4.5 Discussion

Based on the current study, extensions of the current multi-connectivity optimization framework can be extracted to improve future network deployments. As stressed earlier, the resource allocation policy can be adjusted via changing the value of γ of the UPF's sigmoid function. As it was shown with the increment of gamma, we have a transition from the linear to non-linear (step function) shape of the sigmoid, i.e., network (PF approach) to user perspective (UPF approach). Towards this direction, operators can impact the users network performance via UPF function shape adjustment based on traffic classification.

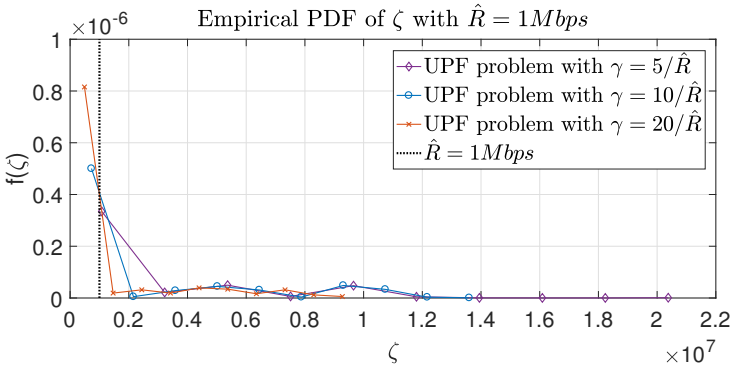


Figure 4.4: PDF plot of UPF utility function for several γ

Specifically, traffic classes can be prioritized based on the predefined QoS class identifier (QCI) such that to be optimized in a sequential way. This sequential optimization method is efficient even when the number of traffic class remains small; however, it needs to pre-define the priority and minimum quality of each class. Another approach is to perform the resource allocation for each traffic class in parallel. By doing so, the PRBs need to be abstracted as virtual resource block (VRB) with infinite resources, which means each utility maximization problem of each traffic can be executed in parallel. Then theses VRBs are mapped into physical PRBs after the optimization following the resource allocation policy and traffic class priority, and then transported from the logical to physical channels. To this end, MAC scheduler can apply different algorithms associated with different MAC layer functions as the ones proposed for resource allocation using a gamut of utility functions in order to fulfill different network demands. These demands can be related to the establishment of different radio bearers (e.g. dedicated bearer and default bearer in LTE) across different RATs that are converged at the higher layers (e.g. packet data convergent protocol (PDCP) or legacy IP).

Complementarily to this chapter, Chapter 5 introduces the user-to-network and network-to-user traffic, i.e., regular UL/DL traffic to/from remote server, formulating new schedulers mechanisms beyond the scope of U2U traffic. In more detail, even U2U traffic can be routed via isolated BSs based on local-routing, user-to-network and network-to-user traffic must be routed over the backhaul that appears to be an emergency performance bottleneck bringing out new limitations as the next chapter examines.

4.6 Conclusion

This chapter examines a resource allocation problem under multi-connectivity in an evolved LTE network. A utility proportional fairness resource allocation is proposed as an extension to the proportional fairness one, which takes into account the QoS requirement in terms of requested rates. Simulation results reveal that the multi-connectivity can boost the aggregated data rate of user-to-user traffic in under-loaded and uneven-loaded scenarios when compared with the single-connectivity case. In addition, UPF is able to fulfill the requested rate and increase the satisfaction ratio when there are available radio resources among multiple connections. Lastly, the shape of UPF function can be changed in accordance to either user or network perspectives.

Chapter 5

Multi-Connectivity Resource Allocation with Limited Backhaul Capacity in Evolved LTE

5.1 Introduction

In the previous chapter, we focus on the multi-connectivity resource allocation in the radio access layer assuming U2U traffic. In more detail, we study the new problem of utility-based resource allocation under the multi-connectivity regime and we compare the performance of the proposed scheme with the legacy single-connectivity use case. Hence, complementarily to the previous chapter, we focus on the regular UL/DL traffic to/from remote server [1]. In addition, while the proposed optimization problem formulated in Chapter 4 supports the constraints arise from a typical wireless environment under multi-connectivity regime, it does not take into account the core network (CN) restrictions appear with the regural UL/DL traffic routing, since U2U traffic can be routed via isolated BSs based on local-routing.

As highlighted in Chapter 1, the chance that the backhaul link capacities will not be sufficient to support the demand of the RAN, is relatively high. Consequently, as the date rate over the air-interface is increasing, the capacity over the backhaul network is emerged as a major performance bottleneck [45]. The main reasons are connected with (i) the evolution of access network functionalities, and (ii) the fact that multiple BS might have to share the capacity of a single backhaul link; impacting the backhaul capacities. Due to these reasons, more sophisticated, *backhaul-aware*, optimization algorimths need to be considered conforming with the new network demands, especially in dense heterogeneous deployment in 5G. To this end, in this chapter, we revisit the problem of resource allocation in multi-connectivity including backhaul constraints by taking into account the UL/DL users QoS. The contributions of this chapter are summarized as follows:

1. We introduce the system model and assumptions to be used to formulate the utility-based resource allocation problem for *multi-connectivity* in LTE that jointly considers the air-interface limitation and *backhaul* capacity constraints.

[1] In this chapter, we refer to the typical UL/DL users traffic, i.e., social networking etc. Without loss of generality, the current work can be extended to the U2U traffic scenarios.

2. Besides fairness among the users, we apply the utility-based resource allocation that takes into account the users *QoS* requirements for the UL and DL directions.

3. Simulation results show that multi-connectivity is superior to the legacy single-connectivity for different regimes, i.e., empty and loaded cell scenarios with fixed and variable backhaul capacity, in terms of network aggregated rate as well as users QoS satisfaction.

The rest of the chapter is organized as follows. The system model and applied assumptions are introduced in Section 5.2. In Section 5.3, we formulate the considered optimization problems. In Section 5.4, we provide the numerical results and give in-depth analysis. In Section 5.5, we offer a brief discussion. Finally, we conclude the chapter in Section 5.6.

5.2 System model and assumptions

This section describes the system model and the assumptions. We consider a set of BS $\mathcal{B} = \{b_1, \cdots, b_{|\mathcal{B}|}\}$ connected to the core network via backhaul network to serve a set of UE $\mathcal{U} = \{u_1, \cdots, u_{|\mathcal{U}|}\}$. An example is presented in Fig. 5.1 with $\mathcal{B} = \{b_1, b_2, b_3\}$ and $\mathcal{U} = \{u_1, u_2, u_3, u_4\}$. In Table 5.1, we summarize some applied notations used in this chapter.

5.2.1 Air-interface model

To better elucidate the assumptions of the current chapter in difference with the ones mentioned in Section 4.2, we redefine them in order to point out the new considered system setup:

A.1-Carrier frequency: Multi-connectivity deployments are categorized into two different types: (a) Intra-frequency or (b) Inter-frequency. The inter-frequency deployment stands for the case where a UE is multi-connected through different carrier frequencies, either from a single or multi-RAT. While the intra-frequency one refers to transmissions on the same frequency from a single-RAT, where Coordinated Multipoint (CoMP) schemes are required to mitigate interference [92]. In this chapter, we apply the inter-frequency single RAT deployment and denotes the DL carrier frequency as f_j^{D} for the j-th BS[2].

A.2-Physical data rate: Each BS b_j can deliver a maximum physical data transmission rate of $R_{j,i}^{\mathrm{D}}$ to a UE u_i (in absence of any other served UE). The physical data rate $R_{j,i}^{\mathrm{D}}$ for the DL in bps is given in Eq. (5.1) based on the Shannon capacity formula:

$$R_{j,i}^{\mathrm{D}} = B_j^{\mathrm{D}} W_j^{\mathrm{D}} \log_2 \left(1 + \mathrm{SINR}_{j,i}^{\mathrm{D}}\right), \tag{5.1}$$

where B_j^{D} is the j-th BS total number of DL PRBs, W_j^{D} is the j-th BS DL bandwidth per PRB in Hz and

$$\mathrm{SINR}_{j,i}^{\mathrm{D}} = \frac{\mathrm{RSRP}_{j,i}^{\mathrm{D}}}{\sum\limits_{b_k \neq b_j,\ f_k^{\mathrm{D}} = f_j^{\mathrm{D}}} \mathrm{RSRP}_{k,i}^{\mathrm{D}} + W_j^{\mathrm{D}} N_0}, \tag{5.2}$$

stands for the DL signal-to-interference-plus-noise-ratio (SINR) of the received signal from the j-th BS b_j to the i-th UE u_i. Respectively, the SINR for the UL direction is denoted as $\mathrm{SINR}_{i,j}^{\mathrm{U}}$.

[2]In the uplink (UL), notation becomes, e.g., f_j^{U}, while the same holds for the rest of the system parameters, and is partially skipped for brevity.

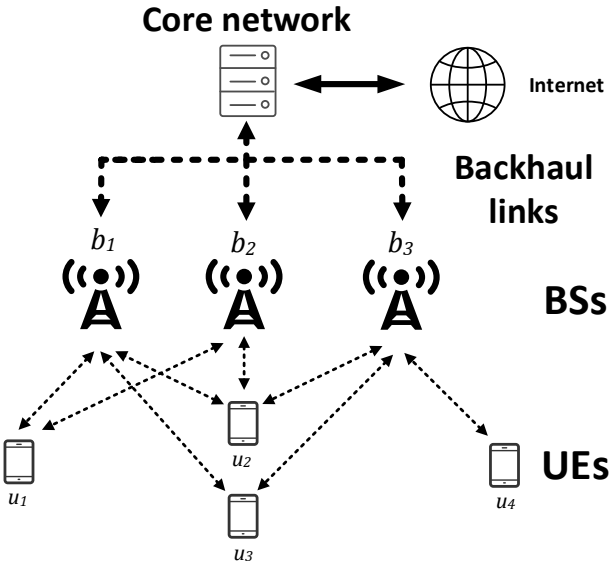


Figure 5.1: Multi-connectivity example in LTE network

The DL Reference Signal Received Power (RSRP) $\text{RSRP}^{\text{D}}_{j,i} = L^{\text{D}}_{j,i} P^{\text{D}}_{j,i} G^{\text{D}}_{j,i}$ includes the path loss and shadowing from the j-th BS to the i-th UE $L^{\text{D}}_{j,i}$, the transmission power of the j-th BS to the i-th UE $P^{\text{D}}_{j,i}$, and the combined antenna gain of the j-th BS and the i-th UE $G^{\text{D}}_{j,i}$. The N_0^{dBm} [3] stands for the thermal noise density in dBm per Hz, such that the product $W^{\text{D}}_j N_0$ is the DL aggregated noise power per PRB. Finally, fast fading effects are assumed to be filtered and equalized.

A.3-Uplink Power Control: The assumption related to the uplink power control is considered as the one mentioned in Chapter 4 (see assumption **A.4** of Section 4.2) and similarly we express the transmitted power of each PRB in UL as:

$$P^{\text{U,dBm}}_{i,j} = \min\left(P^{\text{max,dBm}}_{i}, P^{\text{dBm}}_{0} + \alpha \cdot L^{\text{U,dB}}_{i,j}\right). \tag{5.3}$$

However, in the DL, no power control algorithm is applied and the transmitted power from each BS to all UEs is denoted as $P^{\text{D,dBm}}_{j,i} = P^{\text{D,dBm}}_{j}$.

5.2.2 Connection and Traffic model

The multi-connectivity assumption as described in Chapter 4 is considered the same (see assumption **B.1** in Section 4.2), but new assumptions regarding the backhaul network and the users traffic need to be discussed:

B.1-Backhaul network: Backhaul connection can either directly go through wired solution (e.g., with star topology) or go through one or more wireless hub cells aggregation points (e.g., with tree topology) [93]. In our case, we assume that each BS is connected to the gateway of core network via a dedicated and direct connection (i.e., star topology), see Fig. 5.1. Finally, we characterize the provisioned capacity of the backhaul link from the j-th BS in both UL and DL as $C^{\text{U}}_{h,j}, C^{\text{D}}_{h,j}$ respectively.

[3] The notation dB or dBm is skipped when we refer to the linear scale of any quantities.

Table 5.1: Parameter notation

Parameter	Description	Direction
$\mathrm{RSRP}_{j,i}^{\mathrm{D\ or\ U}}$	Reference signal received power	DL/UL
$\mathrm{SINR}_{j,i}^{\mathrm{D\ or\ U}}$	Signal to interference plus noise ratio	DL/UL
$\mathrm{SINR}_{\mathrm{th}}$	Signal to interference plus noise ratio threshold	-
$B_j^{\mathrm{D\ or\ U}}$	Total PRB number of j-th BS	DL/UL
$M_i^{\mathrm{D\ or\ U}}$	Maximum PRB number of i-th UE	DL/UL
$W_j^{\mathrm{D\ or\ U}}$	Bandwidth per PRB in Hz	DL/UL
N_0	Thermal noise density	DL/UL
$R_{j,i}^{\mathrm{D\ or\ U}}$	Physical data rate	DL/UL
$\hat{R}_i^{\mathrm{D\ or\ U}}$	Traffic requested rate	DL/UL
$P_{j,i}^{\mathrm{D}}$	DL PRB transmission power of BS	DL
$P_{i,j}^{\mathrm{U}}$	UL PRB transmission power of UE	UL
α, P_0	Open loop power control parameters	UL
$P_i^{\max}$	Maximum transmission power of i-th UE	UL
$C_{h,j}^{\mathrm{D\ or\ U}}$	Backhaul link capacity in bps	DL/UL
$f_j^{\mathrm{D\ or\ U}}$	Carrier frequency in Hz	DL/UL

B.2-Traffic requested rate: The UEs are assumed to send and receive traffic from both UL and DL directions. The set of UEs that transmit in UL is defined as $\mathcal{S}$ and the set of UEs receive in DL is defined as $\mathcal{D}$. We denote the traffic requested rate per direction (UL/DL) as the total rate required by a group of user applications running on the top. This group of user applications can be mapped to bi-directional traffic flows corresponding to the UL and DL while the requested rate can reflect the intended user QoS. Specifically, a user u_i sends traffic in the UL by requesting data rate $\hat{R}_i^{\mathrm{U}}$ (UL traffic requested rate) and receives traffic in the DL by requesting rate $\hat{R}_i^{\mathrm{D}}$ (DL traffic requested rate).

5.3 Problem setup

5.3.1 Utility function

We define $x_{i,j}^{\mathrm{U}}, x_{j,i}^{\mathrm{D}} \in \mathcal{X}$ as the percentage in decimals on total PRBs allocated by BS b_j to a user u_i to send or receive traffic in UL and DL. Similarly to Chapter 4, we use two different utility functions; one provides proportional fairness (PF) and another one extends the proportional fairness by taking into account the users QoS requirements (UPF), expressed in Eq. (4.5)and (4.6).

Based on the system model introduced in Section 5.2, we firstly form 4 vectors as $\mathbf{x}_i^{\mathrm{U}} \triangleq \left[x_{i,1}^{\mathrm{U}}, \cdots, x_{i,|\mathcal{B}|}^{\mathrm{U}}\right]^T$, $\mathbf{x}_q^{\mathrm{D}} \triangleq \left[x_{1,q}^{\mathrm{D}}, \cdots, x_{|\mathcal{B}|,q}^{\mathrm{D}}\right]^T$, $\mathbf{R}_i^{\mathrm{U}} \triangleq \left[R_{i,1}^{\mathrm{U}}, \cdots, R_{i,|\mathcal{B}|}^{\mathrm{U}}\right]^T$ and $\mathbf{R}_q^{\mathrm{D}} \triangleq \left[R_{1,q}^{\mathrm{D}}, \cdots, R_{|\mathcal{B}|,q}^{\mathrm{D}}\right]^T$. Then, the per-user final objective utility functions of multi-connectivity proposed for $u_i \in \mathcal{S}$ and $u_q \in \mathcal{D}$ and based on any utility function $\Phi(\cdot)$ (PF/UPF) are

$$U_1\left(\mathbf{x}_i^{\mathrm{U}}\right) \triangleq \Phi\left(\left(\mathbf{x}_i^{\mathrm{U}}\right)^T \cdot \mathbf{R}_i^{\mathrm{U}}\right) = \Phi\left(\sum_{b_j \in \mathcal{B}} x_{i,j}^{\mathrm{U}} R_{i,j}^{\mathrm{U}}\right), \tag{5.4}$$

$$U_2\left(\mathbf{x}_q^{\mathrm{D}}\right) \triangleq \Phi\left(\left(\mathbf{x}_q^{\mathrm{D}}\right)^T \cdot \mathbf{R}_q^{\mathrm{D}}\right) = \Phi\left(\sum_{b_j \in \mathcal{B}} x_{j,q}^{\mathrm{D}} R_{j,q}^{\mathrm{D}}\right). \tag{5.5}$$

Note that $\left(\mathbf{x}_i^{\mathrm{U}}\right)^T \cdot \mathbf{R}_i^{\mathrm{U}}$ and $\left(\mathbf{x}_q^{\mathrm{D}}\right)^T \cdot \mathbf{R}_q^{\mathrm{D}}$ include the aggregated rate over all connected BSs in UL or DL direction for u_i and u_q. The value of $\hat{R}_i^{\mathrm{U}}$ or $\hat{R}_q^{\mathrm{D}}$ is assigned to β in Eq. (4.6) (sigmoid function) for UL or DL direction.

5.3.2 Problem formulation

Based on aforementioned system model, assumptions and utility functions, we then formulate the resource allocation problem. Such problem falls into the category of the network utility maximization for resource allocation and is given as:

$$\begin{aligned}
\max_{\mathcal{X}} \quad & \sum_{u_i \in \mathcal{S}} U_1\left(\mathbf{x}_i^{\mathrm{U}}\right) + \sum_{u_q \in \mathcal{D}} U_2\left(\mathbf{x}_q^{\mathrm{D}}\right) \\
\text{s.t. } \mathbf{C1:} \quad & \sum_{u_i \in \mathcal{S},\ (u_i, b_j) \in \mathcal{E}} x_{i,j}^{\mathrm{U}} \leq 1,\ \forall\ b_j \in \mathcal{B}, \\
\mathbf{C2:} \quad & \sum_{u_q \in \mathcal{D},\ (b_j, u_q) \in \mathcal{E}} x_{j,q}^{\mathrm{D}} \leq 1,\ \forall\ b_j \in \mathcal{B}, \\
\mathbf{C3:} \quad & \sum_{b_j \in \mathcal{B},\ (u_i, b_j) \in \mathcal{E}} x_{i,j}^{\mathrm{U}} B_j^{\mathrm{U}} \leq M_i^{\mathrm{U}},\ \forall\ u_i \in \mathcal{U}, \\
\mathbf{C4:} \quad & \sum_{b_j \in \mathcal{B},\ (b_j, u_q) \in \mathcal{E}} x_{j,q}^{\mathrm{D}} B_j^{\mathrm{D}} \leq M_q^{\mathrm{D}},\ \forall\ u_q \in \mathcal{U}, \\
\mathbf{C5:} \quad & \sum_{b_j \in \mathcal{B},\ (u_i, b_j) \in \mathcal{E}} x_{i,j}^{\mathrm{U}} B_j^{\mathrm{U}} P_{i,j}^{\mathrm{U}} \leq P_i^{\max},\ \forall\ u_i \in \mathcal{U}, \\
\mathbf{C6:} \quad & \sum_{u_i \in \mathcal{S},\ (u_i, b_j) \in \mathcal{E}} x_{i,j}^{\mathrm{U}} R_{i,j}^{\mathrm{U}} \leq C_{h,j}^{\mathrm{U}},\ \forall\ b_j \in \mathcal{B}, \\
\mathbf{C7:} \quad & \sum_{u_q \in \mathcal{D},\ (b_j, u_q) \in \mathcal{E}} x_{j,q}^{\mathrm{D}} R_{j,q}^{\mathrm{D}} \leq C_{h,j}^{\mathrm{D}},\ \forall\ b_j \in \mathcal{B}.
\end{aligned} \tag{5.6}$$

A detailed explanation of the proposed optimization problem is offered presenting the objective function as well as all seven constraints:

Objective function: The objective is to allocate optimally the resources $x_{i,j}^{\mathrm{U}}, x_{j,q}^{\mathrm{D}} \in \mathcal{X}$ for each user $u_i, u_q \in \mathcal{U}$ to maximize the sum of utility functions of all users via applying the per-user utility functions in Eq.(5.4) and Eq.(5.5).

Constraints: We elaborate all constraints as following:

- **C1-C2:** These constraints ensure that the number of allocated PRBs (expressed as the percentage in decimals on total PRBs) of each BS b_j to all users shall not exceed the total number of PRBs in both UL and DL directions.

- **C3-C4:** These two constraints assure that the total number of allocated PRBs to each user among all connected BSs will not exceed its maximum number of allocated PRBs, $M_i^{\mathrm{U}}, M_q^{\mathrm{D}}$, in UL and DL, respectively.

Table 5.2: Simulation parameters

Parameter	Value
LTE mode	FDD, SISO
Frequency band	Band 1 (3×20MHz BW)
Total PRBs of each BS	100 (20 MHz BW)
Maximum PRBs of each UE	100
Number of BSs	3
UE distribution	Uniform for each BS
UE traffic model	Full buffer
SINR threshold	-12 dB
BS transmission power	46 dBm
UE maximum transmission power	23 dBm
Power control parameters	P_0^{dBm} = -58 dBm, $\alpha = 0.8$
Thermal noise density	-174 dBm/Hz
Requested rate distribution	Uniform between 50 to 150 Mbps
γ of UPF utility function	$10/\beta$

- **C5:** Such constraint is related to the power control mechanism in the uplink and it restricts that the total transmission power of the i-th user to all connected BSs can not exceed its power limitations denoted as $P_i^{\max}$.
- **C6-C7:** These two constraints are related to the backhaul link and they assure that the aggregated rates at each BS b_j will not exceed the provisioned backhaul capacities both in the UL and DL directions.

Finally, the convexity of the proposed problem is proved in the following lemma:

Lemma 2. *The optimization problem proposed in Eq.* (5.6) *is convex* [4].

Proof. Any of the proposed functions $\Phi(\cdot)$ are strictly concave as the first one is the logarithmic and the second one is the logarithm of the sigmoid function [5]. It is known that any composition of a concave function with an affine function is concave [90]. Thus, we conclude that $U_1\left(\mathbf{x}_i^{\text{U}}\right)$ and $U_2\left(\mathbf{x}_q^{\text{D}}\right)$ are also concave. Consequently, the objective function is concave as a sum of concave functions $U_1\left(\mathbf{x}_i^{\text{U}}\right)$ and $U_2\left(\mathbf{x}_q^{\text{D}}\right)$. Combined with the linear constraints, the problem is convex with an unique tractable optimal solution. □

5.4 Simulations

In this section, the performance evaluation results are presented for the aforementioned optimization problem. Simulation parameters applied to UEs, BSs and network planning are mostly taken from 3GPP (TR36.814, TR36.942, TR25.942) and NGMN documents, and some important parameters are listed in Table 5.2. The whole simulation framework is built in MATLAB and any optimization results are obtained numerically using the interior point method.

[4] Convex solvers such as interior-point methods can be utilized to find the global optimal solutions with great efficiency [90].

[5] The natural logarithm of the sigmoid function is proved as strictly concave in [89].

Two main simulations scenarios are considered as follows:

- **Scenario A-Empty cell**: There is 1 BS that serves 0 UEs (i.e., 0% traffic load) while the remaining 2 BSs serve non-zero UEs. The empty cell traffic load is used in the evaluation of 3GPP report (TR 25.927) and characterizes some real-world measurement results. We use $0/z/z$ notation to represent this scenario where z is the number of non-zero served UEs per BS.
- **Scenario B-Loaded cell**: All cells serve non-zero UEs and we use $z/z/z$ notation to represent this scenario. The value of z is 2 or 4 in our simulations.

Further, we define two basic metrics: (a) the network aggregated rate as the sum of total allocated rate to all users for DL

$$\bar{R}^{\mathrm{D}} \triangleq \sum_{u_q \in \mathcal{D}} \sum_{b_j \in \mathcal{B}} \left(x_{j,q}^{\mathrm{D}}\right)^{\star} R_{j,q}^{\mathrm{D}} \tag{5.7}$$

where and $\left(x_{j,q}^{\mathrm{D}}\right)^{\star}$ are the optimization results in DL (same holds for UL), and (b) the user satisfaction ratio as the percentage of users that are satisfied with the allocated data rate

$$S^{\mathrm{D}} \triangleq \mathrm{Prob}\left\{ \sum_{b_j \in \mathcal{B}} \left(x_{j,q}^{\mathrm{D}}\right)^{\star} R_{j,q}^{\mathrm{D}} \geq \hat{R}_q^{\mathrm{D}} \right\} \tag{5.8}$$

(same holds for UL).

Finally, the empty and loaded cell scenarios are used to identify the corresponding lower and upper bounds of our performance metrics, i.e., best and worst case scenarios.

5.4.1 Single versus Multi-connectivity

Firstly, we compare the performance of single-connectivity and multi-connectivity under the case of infinite backhaul capacity (i.e., $C_{h,j}^{\mathrm{D}} = C_{h,j}^{\mathrm{U}} = \infty, \forall j$) [6]. The considered single-connectivity mode associates each UE to only one BS and only allows each UE to communicate with that associated BS in both UL and DL directions in terms of the best received RSRP. In following, we firstly provide the results of DL direction.

Fig. 5.2 depicts the aggregated rate of Scenario A under single and multi-connectivity for both PF and UPF case. It can be seen that multi-connectivity outperforms the single one, since it utilizes more resources of the empty cell to enhance the aggregated rate. Additionally, it is noted that PF outperforms UPF as PF aims to boost the aggregated rate while UPF targets to increase the user satisfaction. Further, the same trend is observed in both cases with 2 and 4 UEs per BS, since the applied resource allocation should give the approximately same results in terms of aggregated rate, but not per-user rate as shown in the following user satisfaction results. In Fig. 5.3, the user satisfaction ratio of Scenario A under single and multi-connectivity for both PF and UPF is shown. Multi-connectivity performs better than the single one as more resources are in use. Further, we notice that UPF is superior to the PF in both single and

[6] It is noted that the infinite backhaul scenario UL/DL traffic results are differentiated by the ones presented in the previous chapter for the U2U traffic, since joint UL/DL was considered. Here, the allocation of UL/DL directions is independent such that no restrictions come from the UL power control impact the DL allocation.

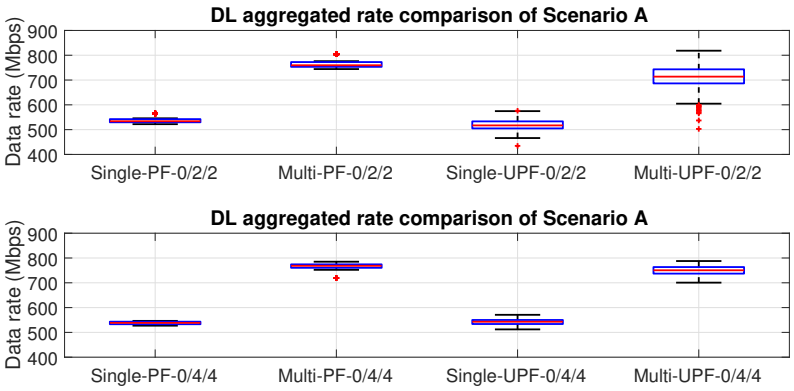


Figure 5.2: Aggregated rate of Scenario A

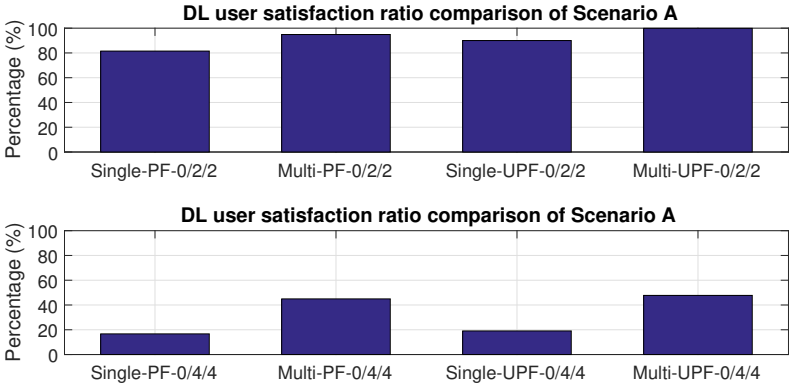


Figure 5.3: User satisfaction ratio of Scenario A

multi-connectivity (3%-9% gain) for 2 and 4 UEs per BS implying the trade-off between the network aggregated rate and the user satisfaction acquisition. Finally, as the number of users is increasing, the user satisfaction ratio is decreasing, since the per-user rate is not sufficient to satisfy the traffic requested rate.

In Fig. 5.4, we examine analogous results for Scenario B. It is noted that there are no significant differences between single and multi-connectivity as no empty cell resources are available, i.e., loaded cell case. It is noted that the aggregated rate is slightly higher compared to Scenario A due to more uniformly distributed users that are added while their SINRs variate. Further, in Fig. 5.5, we observe that UPF can still increase the user satisfaction when compared to PF in both single and multi-connectivity cases (4%-21% gain) for 2 and 4 UEs per BS. Lastly, the UL resource allocation is similar to the DL ones except for the uplink power control constraint in **C5** of Eq. (5.6). This constraint further restricts the number of allocated PRBs to each UE and increases the number of unallocated resources at each BS which can be further exploited in multi-connectivity. In that sense, the UL results of Scenario B[7] in Fig. 5.6 show a higher gain on the aggregate rate and user satisfaction ratio between single and multi-connectivity case when

[7]The UL results for Scenario A show the same behavior with the ones in DL and are omitted due to space limitations.

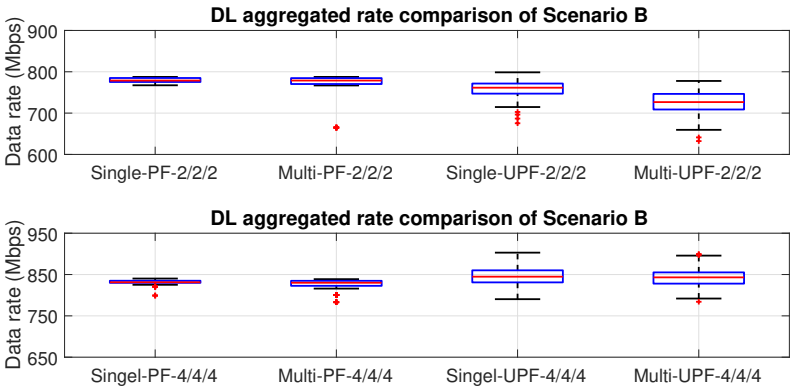


Figure 5.4: Aggregated rate of Scenario B

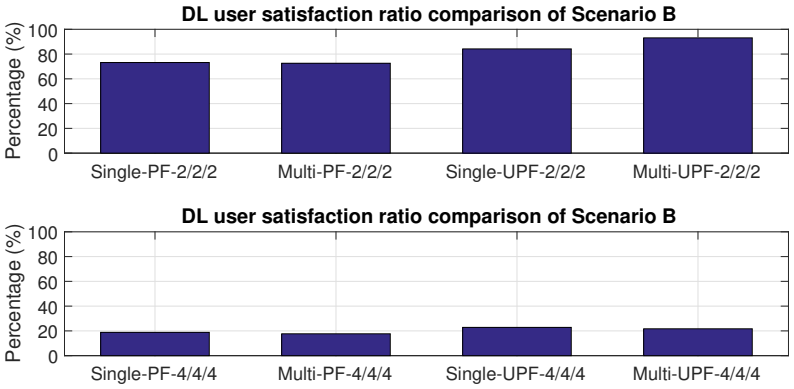


Figure 5.5: User satisfaction ratio of Scenario B

compared with the corresponding gain of DL results in Fig. 5.4 and Fig. 5.5.

5.4.2 Fixed backhaul capacity limitations

We demonstrate the results for DL with fixed backhaul capacity considering the same value is applied to all backhaul links. In Fig. 5.7, the average aggregated rate is shown for Scenario A and B with 2 UEs per BS under single and multi-connectivity as a function of the backhaul capacity for both UPF and PF. In both empty and loaded scenarios, the aggregated rate converges when backhaul capacity is provisioned to exceed 300 Mbps in both single and multi-connectivity cases. Regarding Scenario A (i.e., empty cell), it can be seen that multi-connectivity outperforms the single one even with insufficient backhaul capacity, i.e., less than 300 Mbps, via utilizing unallocated backhaul capacity of empty cell. For Scenario B, there are no significant differences between single and multi-connectivity, since in the loaded cell case multi-connectivity cannot utilize more resources than the single one compared to the empty cell case, where additional resources are available. Nevertheless, the multi-connectivity case can better reflect the trade-off between achieving higher data rate and attaining higher users satisfaction (i.e., PF versus UPF). In other words, more aggregated rate is allocated in multi-connectivity in order to better

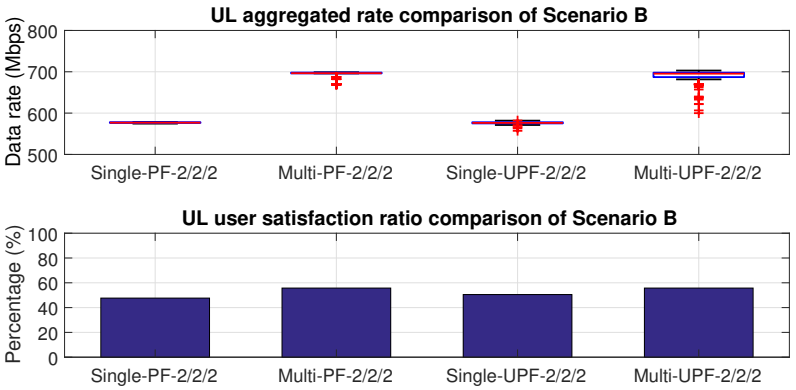
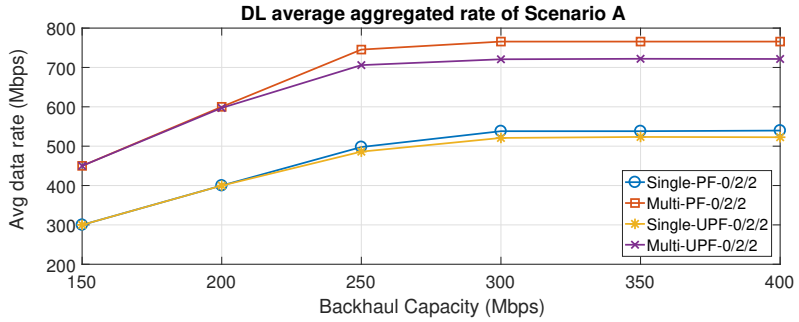
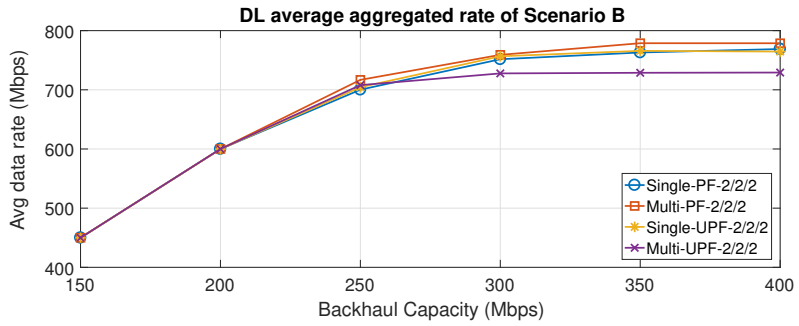


Figure 5.6: Uplink result of Scenario B with 2/2/2



(a) Scenario A: Empty cell



(b) Scenario B: Loaded cell

Figure 5.7: Backhaul capacity impact on average aggregate rate

satisfy the users requested rate when comparing with the single-connectivity case as shown in the following Table 5.3.

The user satisfaction ratio results is presented in Table 5.3 of both Scenario A and B and we observe that the satisfaction ratio can be increased through two manners: (a) apply multi-connectivity (from single to multi-connectivity) where with PF a 3%-31% average gain is observed and with UPF a 7%-34% average gain is offered, and (b) use the UPF utility function (from PF to UPF) where in single connectivity a 7%-10% average gain is inspected while in

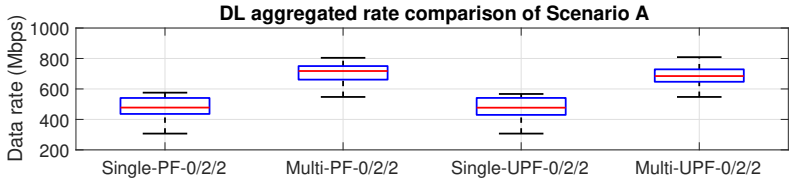
Table 5.3: User satisfaction ratio (%) of different BH capacity

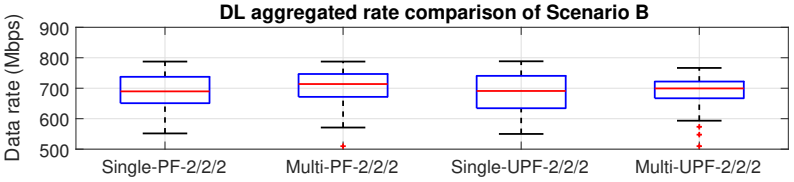
Backhaul Capacity (Mbps)	**Single PF**		**Multi PF**		**Single UPF**		**Multi UPF**	
	0/2/2	**2/2/2**	**0/2/2**	**2/2/2**	**0/2/2**	**2/2/2**	**0/2/2**	**2/2/2**
150	∼20	∼20	63	∼20	∼20	∼20	81	∼20
200	49	∼50	95	∼50	51	54	100	55
250	72	65	95	68	83	80	100	91
300	81	72	95	73	90	84	100	93
>350	81	72	95	73	90	84	100	93

multi connectivity a 8%-16% average gain is noticed. We highlight that multi-connectivity is beneficial in the empty cell case where it can boost the user satisfaction rate via exploiting more unallocated resources. In addition, UPF can be used to accelerate more the intended performance through redistributing the network resources based on the users requested rate especially in the loaded cell case when compared with the PF one.

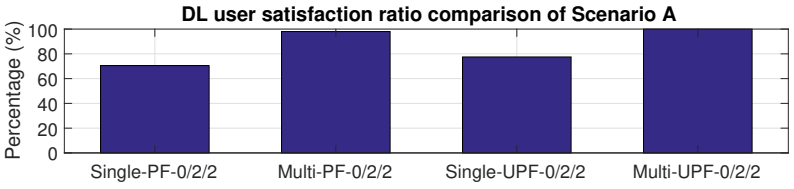
5.4.3 Dynamic backhaul capacity limitations

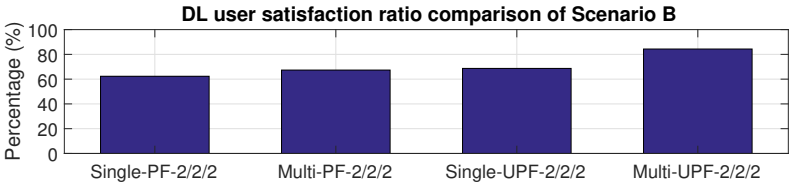
Based on the results of fixed backhaul capacity, we further investigate the time-varying backhaul scenarios (i.e., $C^{\mathrm{D}}_{h,j}(t)$ and $C^{\mathrm{U}}_{h,j}(t)$ are the backhaul capacity at time t). This scenario matches the wireless backhaul condition as introduced in Section 5.2 that could be half-duplex in 5G small cell deployment [94]. In Fig. 5.8(a), we use the time-varying uniformly distributed backhaul capacity between 150 Mbps and 350 Mbps to evaluate the aggregated rate and user satisfaction ratio for both empty and loaded scenarios. We observe that multi-connectivity provides higher aggregated rate for both scenarios (much higher for Scenario A and slightly higher for Scenario B) compared to single one as expected in empty (additional resources to use) and loaded (no additional resources to use) scenarios. Further, regarding the user satisfaction ratio results presented in Fig. 5.8(b), we compare the multi-connectivity to the single one where PF shows a gain of 28% in Scenario A and 5% in Scenario B while UPF offers a gain of 23% in Scenario A and 16% in Scenario B. Finally, respecting the case to compare PF with UPF, in single-connectivity we obtain a gain of 7% for both Scenario A and B while in multi-connectivity we see a gain of 2% for Scenario A and 16% for Scenario B. It can be noticed that the presented results closely lie in the range of gains presented with the finite backhaul capacity representing the anticipated system behavior even with backhaul capacity variability. Table 5.4 summarizes the resource utilization ratio in terms of (a) the percentage of allocated PRBs on air-interface, and (b) the percentage of allocated data rate on backhaul link. Multi-connectivity utilizes better the unallocated resources both in air-interface and backhaul when compared to the single one. It is noticed that the dynamic BH captures the scenario between two cases: a) Unlimited BH: the air-interface resources are fully used (100%) and the BH utilization ratio is < 100 %, b) Limited BH: the air-interface resources are not fully used (< 100 %) and the BH utilization ratio is 100 %. The latter observation clarifies the depicted results that are below 100 % in terms of utilization ratio.





(a) Aggregated rate





(b) User satisfaction ratio

Figure 5.8: Time-varying backhaul capacity of both scenarios

Table 5.4: Utilization ratio (%) of time-varying BH capacity

Utilization Ratio (%)	**Single PF**		**Multi PF**		**Single UPF**		**Multi UPF**	
	0/2/2	**2/2/2**	**0/2/2**	**2/2/2**	**0/2/2**	**2/2/2**	**0/2/2**	**2/2/2**
Air-interface	57	90	94	98	57	90	93	98
Backhaul	63	92	95	93	63	92	93	91

5.5 Discussion

The current study is focused on multi-connectivity framework for regular UL/DL traffic, i.e., remote server UL/DL traffic, by taking into account BH that is characterized as the major performance criticality between the RAN and the core network. Towards this direction, we reveal the air-interface and BH limitations in multi-connectivity MAC schedulers algorithms design. As future extensions of the current work, one can consider different BH topologies, e..g., tree/mesh topology etc., in order to examine wireless BH solutions existing in small cell infrastructures. Another focal point is to deal with the intra-frequency use case, where interference mitigation schemes need to be incorporated in the presented optimization framework. To this direction,

interference can be managed either with power allocation control variables or precoding-based interference alignment solutions under the multi-connectivity regime. That implies further considerations to encompass network interference, especially in load cell use cases, where users QoS needs to be retained.

Hence, the current multi-connectivity schedulers presented in Chapter 4 and 5 do not take into account the spatiotemporal channel variations, i.e., small-scale fading effects, that can ameliorate users QoS performance. Specifically, such technique is known as opportunistic scheduling and can take advantage of the channel fluctuations selecting properly the allocated resources corresponding to each UE in order to improve its data rate. Next chapter discusses and analyzes an opportunistic scheduling scheme in multi-connectivity with BH constraints to better exploit network resources under users QoS requirements.

5.6 Conclusion

To conclude, in this chapter, we propose a UPF resource allocation scheme under multi-connectivity considering backhaul constraints that targets to enhance users QoS. Presented results show that multi-connectivity outperforms single-connectivity in terms of: a) network aggregated rate and b) users QoS satisfaction exploiting better the air-interface and backhaul resources in both empty and loaded cell scenarios with fixed and variable backhaul capacity for the PF and UPF respective cases. The proposed scheme can be applied to other RATs beyond LTE with minor modifications on the legacy link schedulers.

Chapter 6

Utility-based Opportunistic Scheduling under Multi-Connectivity with Limited Backhaul Capacity

6.1 Introduction

In the former chapters (Chapter 4 and Chapter 5) we studied on the problem of multi-connectivity and examine different scenarios, i.e., loaded/unloaded, including air-interface and backhaul constraints. As stressed in Chapter 1, multi-connectivity targets for higher throughput, lower latency, and improving of capacity coverage and network robustness. The latter imposed in many works as formely mentioned using the simplest form of multi-connectivity, so-called dual connectivity (DC), supporting two simultaneous connections across end-users in LTE RAT. For instance, the work in [83] considers resource allocation at the macro base station (BS) assisted by DC-enabled downlink (DL) small BSs connections. While authors in [95] study the joint problem of user association and power allocation in uplink (UL) of a DC-enabled heterogeneous network.

In difference with the aforementioned works, the proposed optimization problems that are formulated in Chapter 4 and Chapter 5 retain QoS satisfaction via a utility-based proportional fairness scheme. Hence, the former proposed methodologies do not consider channel fluctuations, i.e., small-scale fading that are commonly used in opportunistic scheduling [53]. It is noted that the opportunistic scheduling can exploit channel variabilities and allow for better QoS-awareness as discussed in [96]. In that sense, the current chapter analyzes the impact of channel variabilities (hence the term opportunistic) under the multi-connectivity regime with backhaul capacity limitations. The chapter's contributions can be summarized as follows:

- We formulate an opportunistic resource allocation problem under multi-connectivity and limited backhaul capacity in evolved LTE using the two aforementioned utility functions: PF and UPF.

- We provide an efficient algorithm to deal with the formulated NP-hard problem. Such algorithm is guaranteed to produce a solution with an explicit bound toward the optimal solution.

- Simulation results justify that UPF utility function with opportunistic scheduling in multi-connectivity exploit channel variability to further satisfy users QoS requirements compared to the PF one.

The rest of the chapter is organized as follows. We introduce the considered system model in Section 6.2, elaborate the formulated problem and proposed algorithm in Section 6.3, and provide the simulation results in Section 6.4. We offer a brief discussion in Section 6.5. We conclude the chapter in Section 6.6.

6.2 System model and assumptions

We consider an area served by a set of BSs $\mathcal{B} = \{b_1, \cdots, b_{|\mathcal{B}|}\}$ connected to core network (CN) via the backhaul network and a set of user equipments (UEs) $\mathcal{U} = \{u_1, \cdots, u_{|\mathcal{U}|}\}$ is distributed in this area. The topology is the same with the one in Chapter 5 (see for instance Fig. 5.1).

6.2.1 Air-interface model

The assumption **A.1** related to the inter-frequency multi-connectivity deployment as described in Section 5.2 of Chapter 5 holds the same and some variations on the channel model (i.e., small-scale fading so as to consider channel flunctuations) and physical data rate definitions compared to Chapter 5 are presented in the following:

A.1-Channel model: In LTE RAT, the DL direction makes use of orthogonal frequency division multiple access with the minimum radio resource granularity termed as physical resource block (PRB) spanning one single time slot in time domain and one sub-channel (equals to several subcarriers) in frequency domain. Hence, a sub-channel set $\mathcal{K}$ can be allocated for each BS. Our considered channel model includes both large-scale and small-scale fading. The large-scale fading includes both path loss and shadowing as $L^{\mathrm{D}}_{j,i}$ from the j-th BS to the i-th UE, whereas the small-scale one comprises both multi-path effect and Doppler spread as $s^{\mathrm{D}}_{j,k,i}(t)$ from the j-th BS to the i-th UE over the k-th sub-channel at time t. For simplicity, we drop time index t in following paragraphs.

A.2-Physical data rate: The physical data rate that the i-th UE can get from the j-th BS of the k-th sub-channel in bps is $R^{\mathrm{D}}_{j,k,i} = W^{\mathrm{D}}_{j,k} \log_2\left(1 + \mathrm{SINR}^{\mathrm{D}}_{j,k,i}\right)$, where $W^{\mathrm{D}}_{j,k}$ is the bandwidth of the k-th sub-channel at the j-th BS in Hz and

$$\mathrm{SINR}^{\mathrm{D}}_{j,k,i} = \frac{\mathrm{RSRP}^{\mathrm{D}}_{j,k,i}}{\sum\limits_{b_m \neq b_j,\ f^{\mathrm{D}}_m = f^{\mathrm{D}}_j} \mathrm{RSRP}^{\mathrm{D}}_{m,k,i} + W^{\mathrm{D}}_{j,k} N_0} \tag{6.1}$$

is the DL signal-to-interference-plus-noise-ratio (SINR) of the received signal from the j-th BS to the i-th UE of the k-th sub-channel. The reference signal received power (RSRP) $\mathrm{RSRP}^{\mathrm{D}}_{j,k,i} = L^{\mathrm{D}}_{j,i} s^{\mathrm{D}}_{j,k,i} P^{\mathrm{D}}_{j,k,i} G^{\mathrm{D}}_{j,i}$ includes large-scale and small-scale fading, the transmission power of the j-th BS to the i-th UE at the k-th sub-channel as $P^{\mathrm{D}}_{j,k,i}$, and the combined antenna gain at both j-th BS and i-th UE as $G^{\mathrm{D}}_{j,i}$. Note N_0^{dBm} is the thermal noise density in dBm per Hz, such that the product $W^{\mathrm{D}}_{j,k} N_0$ represents the aggregated noise power.

6.2.2 Connection and Traffic model

The multi-connectivity regime holds as in Chapter 4 and 5, where the connectivity criterion described in assumption **B.1** of Section 4.2 holds the same, the i-th UE can be connected to the j-th BS if the average SINR $\overline{\text{SINR}}^{\text{D}}_{j,i}$ over all sub-channels is above a threshold, i.e., $\overline{\text{SINR}}_{\text{th}}$, where

$$\overline{\text{SINR}}^{\text{D}}_{j,i} = \frac{1}{|\mathcal{K}|} \sum_{k \in \mathcal{K}} \text{SINR}^{\text{D}}_{j,k,i}. \tag{6.2}$$

Limited capacity assumptions as the one in Section 5.2 (see assumption **B.1**) are taken into account. We denote the bottleneck backhaul capacity from the CN toward the j-th BS as $C^{\text{D}}_{h,j}$ that can potentially limit the achievable user data rate. In addition, the assumption **B.2** presented in Section 5.2 related to the users traffic requested rate holds the same.

6.3 Problem setup

6.3.1 Utility function

We define $x^{\text{D}}_{j,k,i} \in \{0,1\}$ as the binary variables representing the allocation or not of the k-th sub-channel to the i-th UE from the j-th BS. Hence, $y^{\text{D}}_{j,k,i} = x^{\text{D}}_{j,k,i} \cdot R^{\text{D}}_{j,k,i}$ is the corresponding physical data rate.

Based on the system model introduced in Section 5.2, we firstly form 2 vectors as $\mathbf{x}^{\text{D}}_i = \left[x^{\text{D}}_{1,1,i},...,x^{\text{D}}_{1,|\mathcal{K}|,i}, x^{\text{D}}_{2,1,i},...,x^{\text{D}}_{2,|\mathcal{K}|,i},...,x^{\text{D}}_{|\mathcal{B}|,1,i},...,x^{\text{D}}_{|\mathcal{B}|,|\mathcal{K}|,i}\right]^T$ is the vector form of binary allocation variables and $\mathbf{R}^{\text{D}}_i = \left[R^{\text{D}}_{1,1,i},...,R^{\text{D}}_{1,|\mathcal{K}|,i}, R^{\text{D}}_{2,1,i},...,R^{\text{D}}_{2,|\mathcal{K}|,i},...,R^{\text{D}}_{|\mathcal{B}|,1,i},...,R^{\text{D}}_{|\mathcal{B}|,|\mathcal{K}|,i}\right]^T$ represents the vectorized physical data rate. Then, the network utility function of the i-th UE is $U\left(\mathbf{x}^{\text{D}}_i\right) \triangleq \Phi\left(\left(\mathbf{R}^{\text{D}}_i\right)^T \cdot \mathbf{x}^{\text{D}}_i\right)$, where $\Phi(\cdot)$ can be any of the two utility functions presented in Chapter 4 (see Eq. (4.5)and (4.6). Further, the arguments of both PF and UPF utility functions include the aggregated data rate overall all connected BSs and allocated sub-channels. Finally, UPF utility function assigns $\hat{R}^{\text{D}}_i$ to its β argument (see Eq. (4.6)).

6.3.2 Problem formulation

Optimization Problem: The problem falls into the category of the network utility maximization for resource allocation and is given as follows:

$$\max_{\mathbf{x}^{\text{D}}_i \in \{0,1\}^{|\mathcal{B}|\times|\mathcal{K}|}} \sum_{u_i \in \mathcal{U}} U\left(\mathbf{x}^{\text{D}}_i\right) \tag{6.3}$$

$$\text{s.t.} \quad \textbf{C1:} \sum_{u_i \in \mathcal{U}} \sum_{k \in \mathcal{K}} x^{\text{D}}_{j,k,i} \leq B^{\text{D}}_j, \forall\ b_j \in \mathcal{B},$$

$$\textbf{C2:} \sum_{b_j \in \mathcal{B}} \sum_{k \in \mathcal{K}} x^{\text{D}}_{j,k,i} \leq M^{\text{D}}_i, \forall\ u_i \in \mathcal{U},$$

$$\textbf{C3:} \sum_{u_i \in \mathcal{U}} x^{\text{D}}_{j,k,i} \leq 1, \forall\ b_j \in \mathcal{B}, k \in \mathcal{K}. \tag{6.4}$$

Table 6.1: Simulation parameters

Parameter	Value
LTE mode	FDD, SISO, Band 1 (3×20 MHz bandwidth)
Total PRBs of each BS	100 (20 MHz bandwidth)
Maximum allocated PRBs of UE	100
Number of BSs	3
UE traffic model	Full buffer
SINR threshold	-12 dB
BS transmission power	46 dBm
Thermal noise density	-174 dBm/Hz
Requested rate distribution	Uniform between 10 to 100 Mbps

Objective function: The objective is to allocate optimally the resources $x^{\mathrm{D}}_{j,k,i} \in \{0,1\}$ for each UE $u_i \in \mathcal{U}$ to maximize the sum of utility functions of all UEs. Both PF/UPF utility functions described in Section 8.5.2 can be applied.

Constraints: In following, we present a detailed description on each constraint; **C1:** It ensures the number of allocated sub-channels of BS b_j to all UEs will not exceed the total number of available sub-channels at b_j, B^{D}_j; **C2:** It assures the total number of allocated sub-channels to UE u_i among all connected BSs will not exceed the maximum allocatable sub-channels toward u_i, M^{D}_i. It relies on the user capability defined as the UE-category in 3GPP TS36.306; **C3:** It ensures each single sub-channel will not be assigned to more than one UE implying sub-channel exclusivity.

6.3.3 Proposed algorithm

The formulated optimization problem is NP-hard. Then, we prove it has a submodular and monotone objective function and three matroid constraints in Lemma 3.

Lemma 3. *The formulated optimization problem is with submodular and monotone objective function (Eq.* (6.3)*) and three matroid constraints (Eq.* (6.4)*).*

Proof. The proof is given in Appendix B. □

Hence, we propose a greedy algorithm in Algorithm 2 with $f(\cdot)$ as the objective function and $\mathcal{I}_1/\mathcal{I}_2/\mathcal{I}_3$ as three constraints detailed in Eq. (B.3) and Eq. (B.4) of Appendix B. Such algorithm can efficiently solve the problem with polynomial time complexity. Moreover, we show the outcome provided by the proposed algorithm is with an explicit bound toward the optimal solution in following theorem [1].

Theorem 3.1. *Let OPT be the optimal solution of the formulated problem and $S^\star$ be the output of Algorithm 2. Then, it holds that $f(S^\star) \geq \frac{1}{4}$ OPT.*

Proof. Define $\mathcal{I} = \bigcap_{l=1}^{L} \mathcal{I}_l$, where $\mathcal{I}_l$ is matroid constraints. Even $\mathcal{I}$ is not a matroid constraint, the output of greedy algorithm of Algorithm 2 is guaranteed with $f(S^\star) \geq \frac{1}{L+1}$ OPT shown

[1] The given problem falls in the NP-hard class and a greedy algorithm with an explicit bound on the optimization problem solution is provided as follows. Such methodology argues sub-modularity with matroid constraints and solves the problem in the discrete domain coming with guarantees contrarily to common continuous relaxation approach.

Algorithm 2: Proposed greedy algorithm with complexity: $\mathcal{O}\left(\left(|\mathcal{B}|\cdot|\mathcal{K}|\right)^2\cdot|\mathcal{U}|\right)$

Input : $R^{\mathrm{D}}_{j,k,i}$, $\hat{R}^{\mathrm{D}}_i$, B^{D}_j, M^{D}_i, $\forall\ b_j\in\mathcal{B},\ k\in\mathcal{K},\ u_i\in\mathcal{U}$
Output: $S^\star$
$S_0\leftarrow\emptyset$; $t\leftarrow 0$; /* Parameter initialization */
$\mathcal{I}\leftarrow\mathcal{I}_1\bigcap\mathcal{I}_2\bigcap\mathcal{I}_3$; /* Get intersection of three constraints */
while $t<\sum_{b_j\in\mathcal{B}}B^D_j$ **do**
  $t\leftarrow t+1$;
  $v\leftarrow\arg\max_{l\in\mathcal{I}\setminus S_{t-1}} f\left(S_{t-1}\cup\{l\}\right)$; /*v gets largest gain*/
  if $v\neq\emptyset$ **then**
    $S_t\leftarrow S_{t-1}\bigcup v$; /*Include v in set S_t*/
  else
    break;
$S^\star\leftarrow S_t$;

in [97]. As $L=3$ (i.e., three matroid constraints) in the formulated problem, we prove Theorem 3.1. □

Last but not least, to satisfy the backhaul capacity limitation, $C^{\mathrm{D}}_{h,j}$, we further scale the allocated data rate to the i-th UE as

$$R^{\mathrm{D}}_i\triangleq\sum_{b_j\in\mathcal{B}}\sum_{k\in\mathcal{K}}\left(x^{\mathrm{D}}_{j,k,i}\right)^\star R^{\mathrm{D}}_{j,k,i}\cdot\min\left(\frac{C^{\mathrm{D}}_{h,j}}{\sum_{u_i\in\mathcal{U}}\sum_{k\in\mathcal{K}}\left(x^{\mathrm{D}}_{j,k,i}\right)^\star R^{\mathrm{D}}_{j,k,i}},1\right),\tag{6.5}$$

where $\left(x^{\mathrm{D}}_{j,k,i}\right)^\star$ are the outcomes of proposed algorithm. In reality, such scaling can be done via adjusting the modulation and coding scheme but is out of scope of this study.

6.4 Simulations

Two major scenarios are simulated[2] using self-developed MATLAB system-level simulator with proposed algorithm: **Scenario A-Empty cell**: There is 1 BS that serves 0 UEs, i.e, 0% traffic load, whereas the rest 2 BSs serve non-zero UEs. Such empty cell traffic load is used in the evaluation of 3GPP report (TR 25.927) and characterizes some real-world measurements. The $0/z/z$ notation represents such scenario where z is the number of non-zero served UEs per BS. **Scenario B-Loaded cell**: All cells serve non-zero UEs and we use $z/z/z$ notation to represent this scenario. The value of z is 2 in our simulations[3]. Then, our considered performance metrics include (i) user satisfaction ratio as the percentage of users that are satisfied with the allocated data rate, i.e., $P_{sat}\triangleq\mathrm{Prob}\{R^{\mathrm{D}}_i\geq\hat{R}^{\mathrm{D}}_i\}$ and (ii) unsatisfied normalized error as the normalized

[2]Simulation parameters applied to UEs, BSs and network planning are mostly taken from 3GPP (TR36.814, TR36.942, TR25.942) and NGMN documents, and some important parameters are listed in Table 6.1

[3]Without loss of generality, we focus on the impact of zero and non-zero cell users and any variation of z can cover this case study.

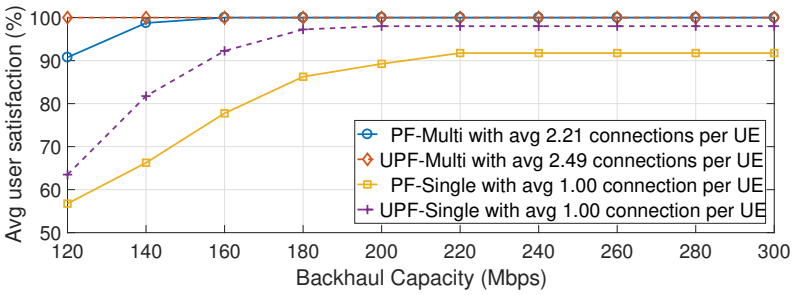


Figure 6.1: User satisfaction ratio versus backhaul capacity $(0/z/z)$

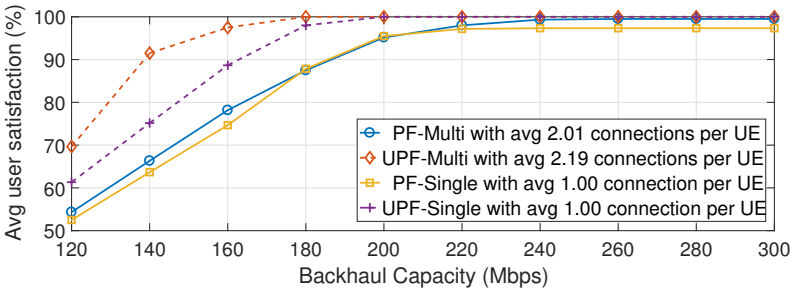


Figure 6.2: User satisfaction ratio versus backhaul capacity $(z/z/z)$

Euclidean distance from the allocated rate to the requested rate formulated as

$$E_i^{\mathrm{D}} = \begin{cases} \left\| \frac{R_i^{\mathrm{D}} - \hat{R}_i^{\mathrm{D}}}{\hat{R}_i^{\mathrm{D}}} \right\|, & \text{if } R_i^{\mathrm{D}} < \hat{R}_i^{\mathrm{D}}, \\ 0 \qquad\quad , & \text{o/w.} \end{cases} \tag{6.6}$$

In Fig. 6.1, empty cell scenario results are shown with several combinations of PF/UPF with single/multi-connectivity. We can see that using UPF utility function can boost the satisfaction ratio even with single-connectivity; however, some QoS requirements can never be fulfilled even we continuously increase the backhaul capacity. It is due to the lack of utilizing unallocated resources at empty cell and thus multi-connectivity shows its advantage. Further, using UPF with multi-connectivity can better fulfill QoS requirements even with fewer backhaul capacity. Then, the results of loaded cell scenario are shown in Fig. 6.2. Using multiple connections is not enough to fulfill QoS requirements if only PF is adopted. Such observation is due to there is no unallocated resources from empty cell; hence, we need to efficiently reshuffle allocated resources via using UPF.

Finally, Fig. 6.3 shows the opportunistic scheduling benefits in loaded cell scenario with UPF and multi-connectivity compared to the non-opportunistic one. Note the non-opportunistic scheduling can optimally rely on the average SINR (i.e., $\overline{\mathrm{SINR}}_{j,i}^{\mathrm{D}}$); however, its worst case is based on the minimum SINR among all sub-channels (i.e., $\min_{k \in \mathcal{K}} \left(\mathrm{SINR}_{j,k,i}^{\mathrm{D}} \right)$). The opportunistic one can utilize more connections and provides a smaller error no matter to the best or worst case of non-opportunistic scheduling.

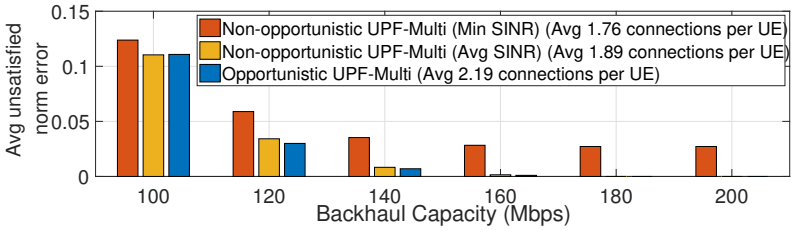


Figure 6.3: Unsatisfied error versus backhaul capacity $(z/z/z)$

6.5 Discussion

The possible directions beyond the proposed opportunistic scheme extend the current scheme to evaluate system performance by taking into account spatio-temporal channel fluctuations in more complex scenarios aligned with 5G technologies. As stressed earlier, UPF favores users QoS exploiting the time and frequency variant channel conditions even in load cell scenarios as well as utilizing more connections. The latter calls for further considerations of new opportunistic schemes beyond the current study under the multi-connectivity regime seeking for novel applications and new challenges.

In more detail, the use of opportunistic scheduling in more complex scenarios introduces multi-objective QoS schedulers, where different users QoS criteria need to be fulfilled while fairness is imposed. Further, different clusters among mobile users can be formed based on channel variability and traffic requirements applied in cooperative schemes, e.g., D2D communications. Towards this direction, a member from each cluster is chosen to convey information to neighboring users in case of emergency, e.g., disaster management communications, applying a best-channel scheduling policy as the one presented in this chapter. Hence, flow-level dynamics based on queuing theory modeling provide a better overview approaching more realistic scenarios for different types of network traffic. To this end, combined schemes as U2U and UL/DL traffic can be jointly scheduled focusing on diverse traffic patterns requirements under multi-connectivity and network BH limitations.

6.6 Conclusion

In this chapter, we formulate the utility-based opportunistic resource allocation problem under multi-connectivity and limited backhaul capacity. The proposed greedy algorithm can efficiently solve the problem with an explicit bound toward the optimal solution. The simulation results justify the benefits of UPF utility function with opportunistic scheduling in terms of QoS satisfaction ratio and unsatisfied normalized error.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

Mobility is an important aspect of the current and next generation cellular systems considering the aggressive densification of network deployments and the spatio-temporal load variations often met in HetNets. Nowdays, users intend to change seamlessly their point of attachment under the umbrella of a reliable and consistent mobility management service. SDN promises to overcome such adversities in network design by centrally controlling network functions via an abstraction of the underlaying network. These mechanisms are applied on mobility management procedures in 5G mobile communications exploiting the holistic knowledge of the network.

In parallel to mobility management, resource allocation plays a crucial role to meet the insatiable demands of users services, especially in networks deteriorated by data hungry devices towards 5G. The current cellular technologies are usually only providing single cell connectivity which limits cell edge users performance in terms of QoS when the network resource management comes to the picture. In single cell connectivity, the end user is only allowed to transmit and receive data with only one BS. To this end, multi-connectivity is considered as an efficient approach supporting simultaneous connections to several technologies or bands providing more resources to the end-users and exploiting cell diversity.

The overall study aims to suggest new radio functionalities beyond the simplistic approaches found in conventional networks as well as to pose questions related to the problems arised so as to provide solutions by better understanding the future network trends and needs. Throughout the thesis chapters, we consider various models and we perform the required analysis by using tools lie in queuing, probability, convex, non-convex and combinatorial optimization theory. The derived expressions and algorithms are aligned with the 3GPP standardization and can be exploited by any operator for efficient mobility and resource management in future 5G deployments. Our contributions are summarized as following:

- In Chapter 2, to better understand mobility, we implement the X2 HO in OAI platform and examine its performance in terms of network delay as well as the impact of system parameters on triggering the HO process. The analysis creates a detailed view of the HO process both in emulated and real world RF testbed environment that shows the main delay bottleneck resides in the UE's uplink synchronization through the MAC RACH procedure implying moderation in handover decisions.

- In Chapter 3, motivated by mobility management challenges in dense network environments as part of future networks, we focus on new handover algorithms beyond the conventional legacy ones presented in Chapter 2 considering both the network and user perspective. Specifically, a load-aware algorithm is proposed and an SDN architecture framework is sketched to support the HO functionality. The algorithm intends to alleviate the assymetrical power regime met in HetNets and take into account user QoS criteria as service delay. The extracted cell assignment probability and the user service delay performance results shows the load-aware algorithm is superior compared to the legacy RSS and distance-based schemes.

- In Chapter 4, inspired by multi-connectivity new added feature towards 5G, we examine new policies in resource allocation problem under multi-connectivity regime in an evolved LTE network. A UPF resource allocation that takes into account the users requested rates as QoS criterion is proposed and is compared to the PF allocation policy throughout simulations. The gain of multiple connections compared to the single-connectivity case is characterized in under-loaded and uneven-loaded scenarios. Finally, UPF shows to outperform PF in terms of users satisfaction, i.e., requested rate is achieved even in high loaded scenarios. It is noted that the shape of UPF can be changed by the corresponding operator in accordance to either user or network perspective.

- In Chapter 5, based on the previous multi-connectivity study, we investigate the impact of backhaul capacity limitations while in Chapter 4 only the air interface constraints are taken into account. The UPF policy is compared with the PF common one via extensive simulations showing that multi-connectivity outperforms single-connectivity in different network use cases, i.e., empty and loaded cell scenarios under fixed and variable backhaul capacity constraints. The presented results reveal the gain of multiple connections compared to the single one even the network aggregated rate is degraded due to the backhaul and the user satisfaction is impacted.

- In Chapter 6, we deepen into the spatiotemporal channel variability (large- and small-scale fading) by exploiting the opportunistic scheduling for loaded and unloaded cell scenarios under backhaul constraints similar to the ones presented in Chapter 5. In more detail, a utility-based opportunistic resource allocation problem under multi-connectivity and limited backhaul capacity is formulated. A greedy algorithm is proposed to solve an NP-hard problem and an explicit bound toward the optimal solution is derived. The offered results show the benefits of UPF utility function with opportunistic scheduling in terms of QoS satisfaction ratio and unsatisfied normalized error.

7.2 Future work

Towards 5G networks, various problems have been proposed and studied both in radio access and core network leaving room for many open issues. Consequently, the maximum performance improvements, the corresponfing trade-offs as well as the arising dependencies between the access-backhaul networks in the context of next generation communications systems are not clear yet. Having proposed concrete mobile management and resource allocation techniques within specific frameworks, the future research directions we are planning to pursue implying our vision beyond 5G are the following:

- *Joint resource allocation and user association in multi-connectivity:* Toward 5G era, developed frameworks related to user association in HetNets can be applied to multi-connectivity scenarios and jointly combined with our resource allocation ones. A tractable and analytical framework for joint access and backhaul network optimization on user association presented in [45]. To this direction, such a framework can be extended to support multiple associations under backhaul limitations so as to provide seamless handovers to the intended users and take into account resource allocation policies based on a generalized utility function aligned with an SDN-based architecture. Such methodologies can offer a centralized optimization framework focusing on users QoS enhancements via a transparent mobility scheme.

- *FlexRAN multi-connectivity mobile and resource management API:* Based on the above multi-connectivity framework, the current FlexRAN API (see Appendix C) can be further extended to support the corresponding radio and CN functionalities based on such algorithmic scheme decisions. To this direction, new gamut of VNFs related to SDN-aided mobility and resource management control, i.e., FlexRAN local agent and centralized controller APIs, need to be developed as well as the RAN and CN OAI multi-connectivity features need to be supported. This kind of sophistication can leverage new mechanisms wrapped up as a 5G prototype to pave the path for next-generation network deployments.

Other future work steps, include the consideration of combined U2U and UL/DL traffic in the context of co-existence of public safety networks and 5G network deployments.

- *Public safety networks under multi-connectivity:* 5G is expected to be the next RAT for Public Safety (PS) communications and specifications items have emerged to this regard [85]. Beyond 3GPP Rel.14, scenarios include multi-hop communications as part of self-backhauled LTE mesh networks [98] promise to improve users QoS in autonomous networking environments. Being aligned with such scenarios, we can further extend our multi-connectivity resource allocation framework to support mixed U2U (2-hop) and UL/DL (4-hop) traffic via local or backhaul routing depending on the network mesh topology. To this end, an adaptable SDN framework that offloads the backhaul can be provided while in parallel can apply a holistic optimization approach that retain users QoS in complex multi-RAT environments.

Chapter 8

Résumé en français

8.1 Réseaux vision 5G

Les changements sociétaux à travers le monde auront un impact sur la façon dont les systèmes de communication mobiles et sans fil sont utilisés de nos jours. Une pléthore de services tels que la banque en ligne, l'e-santé, la réalité augmentée, le divertissement à la demande, etc., commence à proliférer dans les consortiums de recherche émergents de l'industrie et des universités à définir les capacités d'accès sans fil 5G. L'accès permanent à ces services répond aux exigences de continuité de service, de mobilité sans faille, d'efficacité accrue des ressources et de faible consommation d'énergie. D'autre part, l'effet d'avalanche de l'augmentation de volume de trafic mobile et sans fil qui est suivie par la croissance exponentielle des dispositifs connectés ciblait non seulement les communications de type humain qui dominent les réseaux actuels mais aussi ceux de type machine. Un total de 50 milliards d'appareils connectés devrait exister d'ici 2020 [1, 2] avec la prochaine technologie de l'Internet des objets (IoT), la communication massive de type machine (M-MTC) et l'ultra-communication fiable de type machine (U-MTC) promettant de faciliter notre quotidien.

L'évolution future des réseaux doit supporter la coexistence de diverses applications aux caractéristiques variées qui imposent un large éventail d'exigences différentes. Afin de mieux refléter les exigences de la 5G, nous décrivons ses capacités perturbatrices en fonction des objectifs visés, résumés par les mesures suivantes: a) débit Internet maximal au terminal ($\geq$ 10Gbps), b) densité de connexions ($\geq$ 1M terminaux/km^2), c) latence de bout en bout ($\leq$ 5ms), d) débit de données utilisateur garanti ($\geq$ 50Mbps) avant le codage des canaux, e) capacité des terminaux à orientation humaine ($\geq$ 20B), f) capacité des périphériques IoT ($\geq$ 1T), g) support de mobilité en déplacement terrestre ($\geq$ 500km/h), h) la fiabilité du service agrégé ($\geq$ 99.999%) et i) la précision de l'emplacement du terminal extérieur ($\leq$ 1m) [3]. En visant les objectifs de développement durable, plusieurs indicateurs clés de performance (KPI) sont présentés par des définitions mathématiques quantitatives afin de faciliter la conception de solutions techniques et d'accomplir les nouvelles demandes de technologies de tendance [4]. En outre, des critères non quantitatifs sont liés à la gestion du réseau, y compris les architectures centrales logicielles, l'authentification simplifiée, le support multiclient et multi-RAT, les communications par satellite et la sécurité et la confidentialité pour réduire les niveaux CAPEX/OPEX avec les technologies actuelles.

Les principes de conception pour construire l'infrastructure de réseaux de nouvelle génération

hébergent de nouveaux types de services, de nouveaux types de dispositifs et différentes technologies allant du réseau d'accès radio de liaison de bout en bout à l'accès de backhaul (RAN). Dans ce sens, l'architecture 5G changera radicalement en faveur de nouveaux paradigmes radio. Les récentes avancées vers la transition Pro-5G/LTE-Advanced Pro (LTE-A Pro) telles que les petites cellules [5] conduisent à des réseaux Ultra Dense (UDN) combinés à de multiples sorties multi-entrées (MIMO) et des mécanismes de transmission micro-ondes (mm-wave) promettent d'augmenter la capacité, de réduire les émissions de puissance et d'étendre la couverture du réseau. De tels environnements sont identifiés par un mélange de différentes entités, par exemple, des communications d'homme à homme, d'homme à machine et de machine à machine (M2M). En conclusion, les opérateurs peuvent choisir différentes stratégies dans la planification de réseau dominée par de multiples technologies qui permettent aux nouveaux modèles d'entreprise de prendre en charge à la fois les services de communication existants et leurs améliorations futures.

8.2 Motivation et contribution de la thèse

Les réseaux modernes sont caractérisés par différentes technologies interagissant avec différents types d'entités. Une telle hétérogénéité forme un nouveau type de réseaux, les réseaux dits hétérogènes (HetNets), incluant par exemple les stations de base (BS) LTE, les points d'accès sans fil (Wi-Fi), communications Car-to-x (V2X) et de périphérique à périphérique (D2D) etc., comme représenté sur la figure 8.1. Ce dernier motive à explorer les caractéristiques asymétriques de HetNets alignées avec les architectures de réseau à venir. Dans cette direction, la thèse se concentre sur *gestion de mobilité* qui est apparue pour être revisitée dans les HetNets LTE/LTE-A introduisant de nouveaux mécanismes au-delà des schémas de standardisation actuels. Ainsi, cette vision radicale du réseau crée de nouvelles interfaces radio en amont des schémas de liens point à point (P2P) établissant simultanément plusieurs connexions afin d'améliorer la fiabilité. La nouvelle fonctionnalité de l'introduction de plus d'une interface radio exploite de nouvelles façons de gérer les ressources du réseau. S'attaquant au problème de *gestion de ressources* en multi-connectivité, la thèse fournit de nouvelles solutions en passant en revue les anciennes techniques dans le sens des réseaux mobiles de la prochaine génération. En conclusion, nous portons notre attention sur de tels problèmes et discutons en détail les défis du système actuel, ainsi que les critères de conception au-delà de la méthodologie de l'état de l'art proposée.

8.2.1 Problème 1: Gestion de mobilité dans les réseaux LTE/LTE-A et au-delà

La mobilité est l'une des principales caractéristiques des systèmes cellulaires actuels et de la prochaine génération qui permet aux utilisateurs de modifier de façon transparente leur point de rattachement tout en utilisant leurs services voix et données dans des environnements denses à réseaux de petites cellules. Les problèmes critiques sont liés au déploiement de nouvelles techniques basées sur la disponibilité du spectre, le contrôle d'admission préliminaire et les critères de qualité de service (QoS) [6]. Les futurs mécanismes de contrôle des réseaux sont destinés à résoudre des problèmes de décision complexes et multidimensionnels ayant une incidence sur le fonctionnement de l'équipement usager (UE), par exemple le débit, le délai de service, etc. pendant la procédure de HO dans la gestion de mobilité.

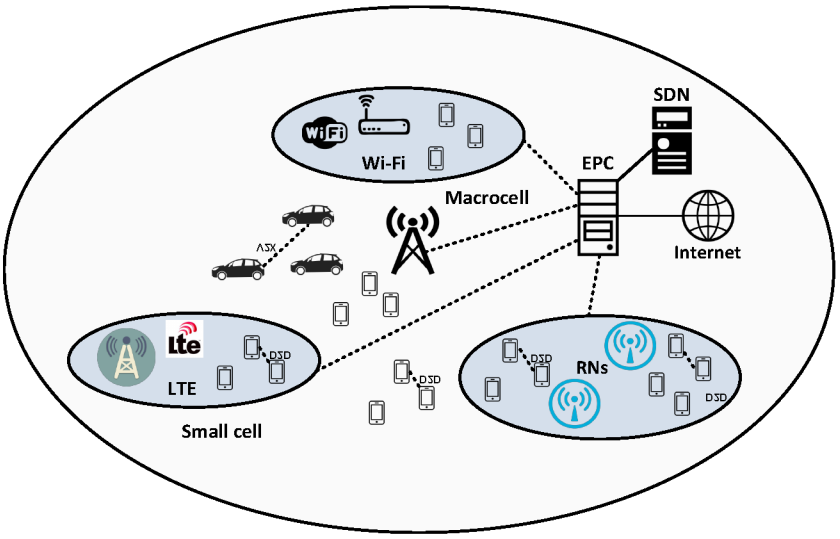


Figure 8.1: Réseaux héterogènes vision 5G

De nombreux travaux ont été réalisés dans les mécanismes HO en considérant le réseau global (c.-à-d., basés sur le réseau). Un algorithme d'auto-optimisation des paramètres de HO dans les réseaux auto-organisés (SON) LTE [8] applique une politique d'opérateur basée sur différentes métriques (c.-à-d., ping-pong HO ratio) pour modifier les paramètres d'hystérésis et de temps de déclenchement (TTT) et augmente considérablement les performances du système. En outre, une optimisation des paramètres pour le HO LTE à l'aide d'une carte auto-organisée avancée (SOM) réduit le nombre d'échecs de HO intercellulaire [9]. D'autre part, de nombreux travaux existent dans les algorithmes de décision HO effectués par l'utilisateur (c.-à-d., basés sur l'utilisateur). Les plus communément utilisés sont basés sur: i) l'intensité du signal reçu (RSS), ii) la vitesse, iii) le mode d'interférence [10, 11]. Par conséquent, ces travaux ne prennent pas en compte les HetNets hiérarchiques et les métriques QoS de l'utilisateur (par exemple, le délai de service de l'utilisateur). Le réseau à définition logicielle (SDN) promet de surmonter les adversités dans la conception de réseau en contrôlant centralement les fonctions de réseau via une abstraction du réseau sous-jacent (c.-à-d., la virtualisation de réseau). L'architecture SDN fournit un contrôleur central pour l'état du réseau global qui peut affecter les décisions de contrôle local de chaque élément radio individuel [12]. Ces mécanismes peuvent également être appliqués aux procédures de gestion de mobilité dans les communications mobiles de cinquième génération (5G) offrant une connaissance holistique du réseau.

8.2.2 Problème 2: Gestion de ressources en multi-connectivité dans les réseaux 5G

La multi-connectivité est l'une des techniques les plus prometteuses qui peuvent être utilisées pour améliorer les performances de 5G [13]. De plus, avec l'arrivée d'Internet-of-Everything (IoE), le trafic réseau augmente énormément, les futurs réseaux doivent non seulement fonctionner avec plusieurs connexions, mais aussi respecter les exigences de QoS des périphériques. Dans 3GPP Rel.12, l'objet d'étude (SI) cible sur trois objectifs principaux: la robustesse de la mobilité, la réduction de la charge de signalisation et des améliorations de débit par utilisateur [14].

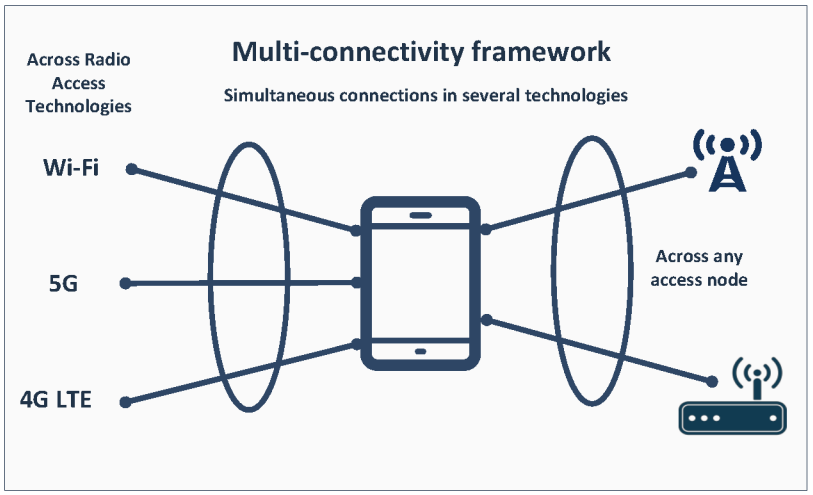


Figure 8.2: Multi-connectivité

Afin de les accomplir, l'idée de multi-connectivité a été proposée où un UE est connecté en même temps à plusieurs BS comme représenté sur la figure 8.2. Cette diversité de ressources permet à l'UE d'exploiter intelligemment les multiples connexions physiques afin d'atteindre les gains correspondants prévus.

Dual-Connectivité (DC) en LTE: La multi-connectivité n'est pas un nouveau concept dans les réseaux de communication mobiles. LTE-advanced (LTE-A) Rel.12 introduit le DC [14]. DC implique la connexion UE à deux stations de base simultanément et il peut être vu comme une extension de l'approche d'agrégation de porteuse (CA) [15] où les points de transmission sont connectés via une liaison backhaul.

Déploiement multi-connectivité en 5G: En termes de mise en œuvre, contrairement à l'ancien schéma de réseau, la multi-connectivité en 5G peut supporter un nombre de cellules impliquées supérieur à deux. L'idée de multi-connectivité peut incorporer des connexions à travers plusieurs RAT hétérogènes, par exemple, LTE, 5G, WiFi, etc., ainsi que sur plusieurs interfaces radio du même RAT.

Par conséquent, l'importance de la QoS attire notre attention sur la gestion de ressources réseau vers les réseaux 5G dans le cadre du régime de multi-connectivité. En général, de nombreux types d'allocation ont été proposés dans la littérature relative aux anciennes technologies de réseau. Le critère de gestion de ressources par ordonnancement proportionnels et équitables fondé sur les services publics est proposé dans [20] et assure l'équité en termes de fonctions d'utilité sélectionnées. En outre, plusieurs travaux exploitent les fonctions d'utilité en tant que mesure des performances QoS de l'application utilisateur. Dans [21], différentes fonctions d'utilité sont employées pour formuler le problème d'optimisation proportionnelle à l'utilité dans de bonnes conditions de canal sans fil. De plus, les auteurs de [22] traitent des allocations de puissance du trafic multi-classes en fonction de quatre types généraux de fonctions d'utilité: convexe, concave, en forme de S et en S inversé. Une fonction d'utilité variant dans le temps est introduite dans [23] afin de minimiser les délais pour diverses demandes de trafic de premier et arrière plans. Récemment, quelques travaux semblent appliquer l'allocation de ressources avec un support multi-connectivité. Les auteurs de [24] proposent un concept de connectivité mul-

tiple pour un réseau de C-RAN et appliquent une allocation proportionnelle juste et équitable des ressources montrant les résultats de performance en termes de débit et de défaillance des liaisons radio. Dans [25], un algorithme d'approximation est présenté à une association d'usager qui est optimale jusqu'à une constante additive pour l'utilité d'équité proportionnelle sous le régime DC. Néanmoins, aucun de ces travaux n'aborde pas officiellement les schémas d'équité proportionnelle basés sur les fonctions d'utilité dans un environnement sans fil à connectivité multiple, par exemple pour conserver la qualité de service des utilisateurs.

Le but de cette thèse est de répondre aux questions soulevées en révisant les problèmes 1 et 2 abordés dans la section 1.2. Dans ce sens, nous fournissons des hypothèses réalistes appliquées aux modèles de système proposés afin de mieux refléter les tendances de réseau des architectures nouvelles vers les réseaux déployés 5G. Le manuscrit actuel peut être divisé en deux parties: a) gestion de mobilité (chapitre 2 et chapitre 3) et b) gestion de ressources (chapitre 4, chapitre 5 et chapitre 6). Les deux parties couvrent les aspects pratiques présentant les difficultés de mise en œuvre et les leçons apprises à tout praticien ainsi que les approches théoriques soulignant les défis de l'analyse mathématique à tout théoricien. De plus, les outils d'analyse se situent dans la théorie des fils d'attente, des probabilités, de l'optimisation convexe, non-convexe et combinatoire conduisant à des résultats qui offrent des aperçus de conception et révèlent les compromis impliqués. Plus précisément, les chapitres de la thèse et les principales contributions dans chacun d'entre eux, sont organisés comme suit:

8.3 Chapitre 2 - "X2 Handover in LTE/LTE-A"

Dans ce chapitre, nous analysons les performances du handover (HO) X2 du point de vue de l'utilisateur. De plus, l'impact des différents paramètres sur l'algorithme de décision de HO est étudié. Les résultats présentés, obtenus à partir de la plateforme d'émulation OpenAirInterface (OAI) LTE/LTE-A, démontrent que le principal goulot d'étranglement réside dans la synchronisation de l'UE en montée vers l'eNB cible.

8.3.1 Critère de HO 3GPP

Un critère de HO bien connu, couramment utilisé dans les algorithmes de décision HO traditionnels pour les systèmes de communication mobiles (également appliqué dans le LTE 3GPP), est basé sur la méthode de comparaison RSRPs dans laquelle les décalages d'hystérésis et de HO sont inclus. Plus précisément, nous nous concentrons sur l'événement A3 et son état qui est utilisé comme critère pour la sélection des cellules. Le critère est exprimé comme suit:

$$\bar{r}_n^{\mathrm{dBm}}[k] + ofn + ocn > \bar{r}_s^{\mathrm{dBm}}[k] + ofs + ocs + hys + off, \tag{8.1}$$

où $s \in \mathcal{I}$ et indique la cellule de service, $n \in \mathcal{I} - s$ et indique les cellules voisines. Enfin, les paramètres de HO qui sont inclus dans l'équation (2.2) sont définis comme décrit ci-dessus.

L'inégalité ci-dessus est interprétée comme suit: lorsque le RSRP d'une cellule voisine (somme des RSRP et offsets du voisin, $\bar{r}_n^{\mathrm{dBm}}[k] + ofn + ocn$) devient plus grande que celle de la RSRP de la cellule de desserte (somme de l'intensité et de l'offset du signal, $\bar{r}_s^{\mathrm{dBm}}[k] + ofs + ocs$) et la différence est supérieur à la valeur de off (également appelé décalage a3), l'événement A3 est déclenché et l'UE signale les résultats de mesure à l'eNB. Hystérésis (hys) indique la valeur d'une marge de HO entre la source et la cellule cible.

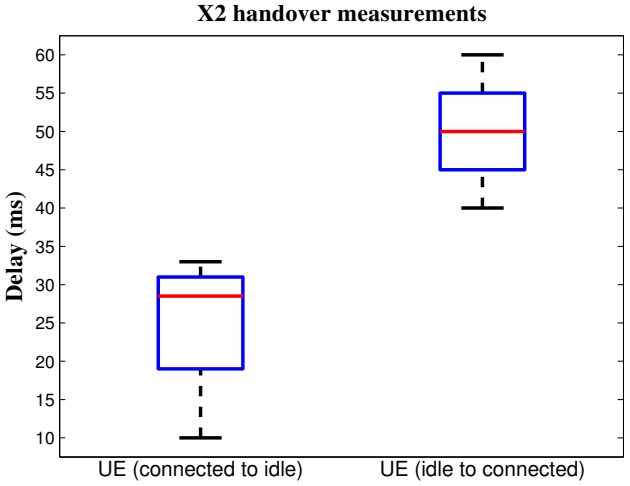


Figure 8.3: Mesures Handover X2

8.3.2 Résultats d'émulation

8.3.2.1 Évaluation de la performance du système

La figure 8.3 caractérise les performances de l'accomplissement de HO X2 dans le segment RAN en termes de délai. La première mesure de délai comprend le temps pendant lequel l'UE est connecté à l'eNB source (c.-à-d., UE RRC connecté) et envoie le rapport de mesure pour se connecter à l'eNB cible, jusqu'à la réception du message RRC Reconfiguration (c.-à-d., UE RRC inactif). La seconde comprend l'intervalle de temps entre la réception du message de reconfiguration RRC du côté UE et la réception du message complet de reconfiguration RRC du côté eNB cible après la synchronisation UE (c.-à-d., UE RRC connecté). Sur la base des expériences extraites des statistiques de la figure 8.3, si nous nous concentrons sur la deuxième mesure de délai qui représente la durée du temps de détachement, on peut remarquer que nous sommes dans l'intervalle de temps estimé basé sur l'équation (2.4) ($\leq$ 65 ms). L'estimation prend en compte le délai d'exécution du scénario HO de cas moyen, à savoir $\sim$ 45 ms, ainsi que les délais supplémentaires de l'implémentation personnalisée liée à l'accès aléatoire basé sur la contention, à savoir $\sim$ 20 ms. On notera que le délai supplémentaire ajouté par la procédure basée sur la contention peut être réduit en introduisant le mécanisme sans contention décrit dans la section 2.2.3 où l'UE obtient directement le C-RNTI via le message de reconfiguration de connexion RRC. Par la suite, le C-RNTI est incorporé dans l'en-tête du message complet de reconfiguration de connexion RRC qui est envoyé plus tard à l'eNB cible dans la partie d'achèvement HO de telle sorte que l'UE soit distingué par l'eNB cible. La prise en charge de cette fonctionnalité dans la phase d'implémentation ignore la transmission des messages d'accord PRACH/montée qui constituent le principal goulot d'étranglement en termes de délai d'exécution, surtout si la qualité du signal UE est faible.

8.3.2.2 Résultats expérimentaux avec OAI X2 RF testbed

Dans le prolongement de la mise en œuvre de HO intercellulaire X2 dans l'émulateur oaisim [64], nous effectuons des mesures RF dans le monde réel dans le banc d'essai expérimental OAI HO X2 déployé présenté dans l'annexe C. L'objectif est de tester, étudier et analyser les performances du système de HO OAI X2 dans un banc d'essai réel. Dans ce sens, comme dans la section 2.4, nous nous intéressons à la performance de HO X2 en termes de délai de HO qui est le critère principal de l'implémentation. Enfin, les instantanés d'informations sur le réseau UE ainsi que les mesures RSRP sont collectés à l'aide de l'application Android Network Guru Android [67]. Un résumé des paramètres du système est fourni dans le tableau 2.2.

Spécifiquement, nous effectuons 4 expériences comme le décrit le tableau 2.3 pour examiner les performances du délai HO X2. Les expériences sont classées en deux groupes de base: Les expériences 1 & 2 correspondent au scénario de 4 dB de différence d'atténuation entre la cible (Ch1) et la source (Ch0) tandis que les expériences 3 & 4 correspondent au scénario de -2 dB . Dans les deux groupes, l'index de configuration PRACH est variable (0 ou 14).

La figure 8.4 décrit les performances du HO X2 dans le OAI RAN en termes de délai. Les mesures de délai des colonnes à barres latérales comprennent le temps pendant lequel l'eNB source reçoit le rapport de mesure RRC par l'UE afin de se connecter au eNB cible (à savoir UE RRC connecté) jusqu'à la réception du message ACK de demande de HO X2 envoyé par l'eNB cible à la source eNB [1]. Les mesures actuelles de délai d'implémentation du banc d'essai RF dépassent l'intervalle de temps estimé basé sur l'équation (2.4) ($\leq$ 65 ms) présentée dans la section 2.4 impliquant d'autres considérations pour l'optimisation du système en temps réel comparé aux résultats de l'émulateur oaisim où les mesures de délai se situaient dans cet intervalle. Enfin, un mécanisme sans contention peut être introduit pour éviter les messages d'accord PRACH/montée qui représentent le goulot d'étranglement de HO principal en termes de délai.

Le travail lié à ce chapitre est:

- K. Alexandris, N. Nikaein, R. Knopp, and C. Bonnet, "Analyzing X2 handover in LTE/LTE-A," in Wireless Networks: Measurements and Experimentation," WINMEE 2016, May 9, 2016, Arizona State University, Tempe, Arizona, USA.

8.4 Chapitre 3 - "Load-aware Handover Decision Algorithm in Next-generation HetNets"

Dans ce chapitre, nous nous intéressons aux nouveaux algorithmes qui prennent également en compte la perspective réseau, par exemple la charge de cellules. Plus précisément, un algorithme prenant en charge la charge est proposé en tenant compte du délai de service qu'un utilisateur confronte sur le réseau. En outre, un cadre implémentable basé sur l'architecture SDN est esquissé pour prendre en charge l'algorithme. L'algorithme proposé est comparé à l'héritage 3GPP (basé sur RSS) présenté dans le chapitre 2 et basé sur la distance. La probabilité d'attribution de cellules extraites et les résultats de performance de délai de service utilisateur montrent que l'approche de charge sensible surpasse les deux.

[1] Les mesures sont effectuées du côté eNB en raison de l'absence d'accès aux journaux de signalisation COTS UE.

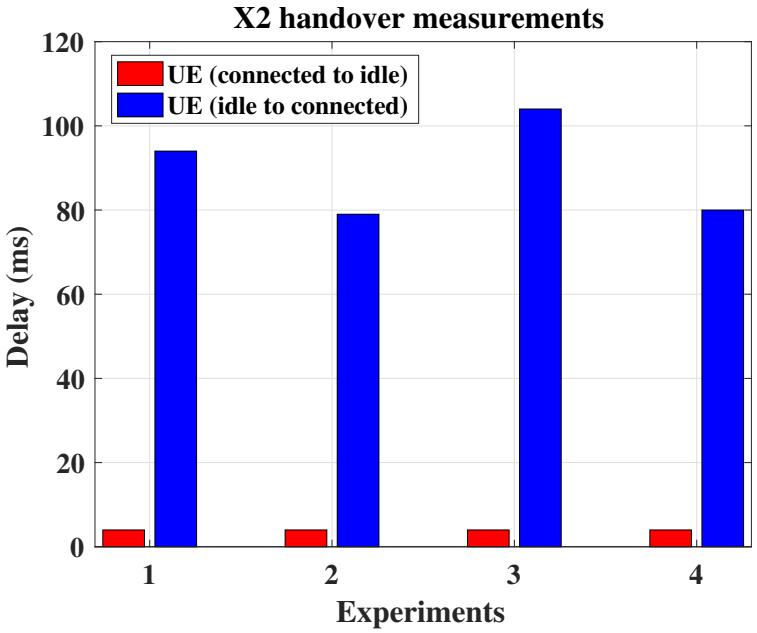


Figure 8.4: Mesures Handover X2

8.4.1 Algorithme de HO sensible à la charge

La politique de load-aware (LA) proposée est décrite in Algorithm 3. L'algorithme est appliqué uniquement aux utilisateurs actifs mobiles et statiques. Les utilisateurs déconnectés ne sont pas pris en compte, car ils sont supposés être désactivés. $\bar{r}_i[k]$ représente le RSRP de chaque cellule i à l'instant k en échelle linéaire. La picocellule que l'utilisateur a l'intention de transférer est celle avec la puissance maximale reçue, notée p_{j^*}. Après la sélection de picocellule p_{j^*}, la condition $\bar{r}_{p_{j^*}}[k] > r_{p,th}$, avec $r_{p,th}$, le seuil de RSRP de picocellule garantit que l'utilisateur reste dans la couverture de la picocellule et que le signal reçu reste suffisamment fort. Ainsi, l'algorithme fonctionne uniquement à l'intérieur de la picocellule. Ce dernier évite les effets de ping-pong dus à la non-couverture des picocellules, c.-à-d., que les utilisateurs de bord de cellule ne sont pas pris en compte. Par la suite, nous définissons $\tilde{r}_{j^*}[k] \triangleq \bar{r}_{p_{j^*}}[k] + \left[1 - f\left(\bar{\mathcal{D}}_m^k, \bar{\mathcal{D}}_{p_{j^*}}^k\right)\right] \bar{r}_m[k]$. $\bar{\mathcal{D}}_i^k$ indique la *prédiction* du délai moyen (de service) d'un nouveau flux d'une cellule i à l'instant k que l'utilisateur rencontre quand rester in le picocellule. Une expression analytique pour $\bar{\mathcal{D}}_i^k$ est donnée par l'équation (3.8). Il est à noter que la prédiction du délai moyen $\bar{\mathcal{D}}_i^k$ est effectuée à l'instant k, mais elle représente une estimation pour toute la période que l'utilisateur reste dans la picocellule (c.-à-d., le délai de l'utilisateur supposé être $\bar{\mathcal{D}}_i^k$ pendant cette période). De plus, $f(\cdot)$ représente une fonction continue et est construit correctement pour forcer l'utilisateur à se connecter dans la picocellule p_{j^*}, ssi $\bar{\mathcal{D}}_{p_{j^*}}^k \ll \bar{\mathcal{D}}_m^k$, malgré le fait que $\bar{r}_m[k] \gg \bar{r}_{p_{j^*}}[k]$. Cette approche répond aux besoins des utilisateurs qui ne peuvent pas être pris en charge par la macrocellule congestionnée et son association avec une picocellule est critique en termes de QoS. La condition qui doit tenir pour se connecter à la picocellule est donnée analytiquement par l'inégalité suivante qui applique la politique ci-dessus:

$$\tilde{r}_{j^*}[k] > \bar{r}_m[k] \Leftrightarrow \bar{r}_{p_{j^*}}[k] > f\left(\bar{\mathcal{D}}_m^k, \bar{\mathcal{D}}_{p_{j^*}}^k\right) \bar{r}_m[k], \tag{8.2}$$

Algorithm 3: Algorithme de HO

Entrée: $\bar{\mathcal{D}}_m^k$ /* Délai moyen prévue de la macrocellule m */
$\bar{\mathcal{D}}_{p_j}^k$ /* Délai moyen prévue de la picocellule p_j */
$r_{p,th}$ /* Seuil RSRP de la picocellule RSRP */
Sortie : Association d'utilisateur
/* Sélectionner la picocellule à la puissance reçue maximale */
$j^* \leftarrow \arg\max_j \bar{r}_{p_j}$;
/* L'utilisateur reste dans la picocellule */
if $\left(\bar{r}_{p_{j^*}}[k] > r_{p,th}\right)$ **then**
  /* Règle de décision HO */
  if $(\tilde{r}_{j^*}[k] > \bar{r}_m[k])$ **then**
    connectée à la picocellule;
  else
    connectée à la macrocellule;

sinon, l'utilisateur est connecté à la macrocellule. Enfin, nous définissons la fonction

$$f\left(\bar{\mathcal{D}}_m^k, \bar{\mathcal{D}}_{p_j}^k\right) \triangleq \exp\left[-c\frac{\bar{\mathcal{D}}_m^k}{\bar{\mathcal{D}}_{p_j}^k}\right],\ c \geq 1 \tag{8.3}$$

avec un co-domaine in $[0, 1]$ qui fournit une telle propriété comme prévu. Plus précisément, la nature exponentielle de la fonction fournit une convergence rapide à zéro lorsque $\bar{\mathcal{D}}_{p_{j^*}}^k \ll \bar{\mathcal{D}}_m^k$. Cela permet de réduire l'asymétrie des puissances reçues afin de conserver l'équation (8.2) et d'associer l'utilisateur à la picocellule sous-chargée. Enfin, la constante c représente un paramètre d'optimisation.

8.4.2 Résultats de la simulation

Dans la figure 8.5, la probabilité d'association de picocellule, $P_r(u \in p)$, est représentée en fonction de la distance de l'utilisateur de la macrocellule pour un nombre fixe de $N_{\mathrm{SU},m} = 200$ et $N_{\mathrm{SU},p_j} = 10$ dans la picocellule. L'algorithme proposé est comparé à la CONV et à celui basé sur la distance (DIST) fourni dans [34], pour λ différent. Ce dernier est appliqué dans le réseau macrocellule/SCs hiérarchique qui est aussi notre cas. L'approche DIST associe l'utilisateur à son SC le plus proche, sans tenir compte de l'asymétrie des puissances BS. Ceci est obtenu en utilisant un facteur de réduction d'échelle α basé sur la maximisation de la probabilité d'association SC. Pour l'algorithme DIST, $P_r(u \in p)$ est proche de 1, après une distance spécifique d'utilisateur de la picocellule, malgré que la macrocellule est surchargée ou non. Cependant avec l'approche LA, $P_r(u \in p)$ augmente, suivant la charge de la macrocellule pour λ différent, par rapport la distance de l'utilisateur à la picocellule. Par conséquent, si l'utilisateur est proche de la picocellule, son association avec elle peut-être inutile, car ses exigences de qualité sont déjà satisfaits. Enfin l'algorithme CONV ne déclenche pas le HO, à cause de la puissance asymétrique des stations de base.

Le délai total que l'utilisateur confronte (c.-à-d., $\bar{\mathcal{D}}_m$ de la macrocellule et $\bar{\mathcal{D}}_p$ de la picocellule selon auquelle il est associé) est présenté dans 8.1. La configuration est considérée

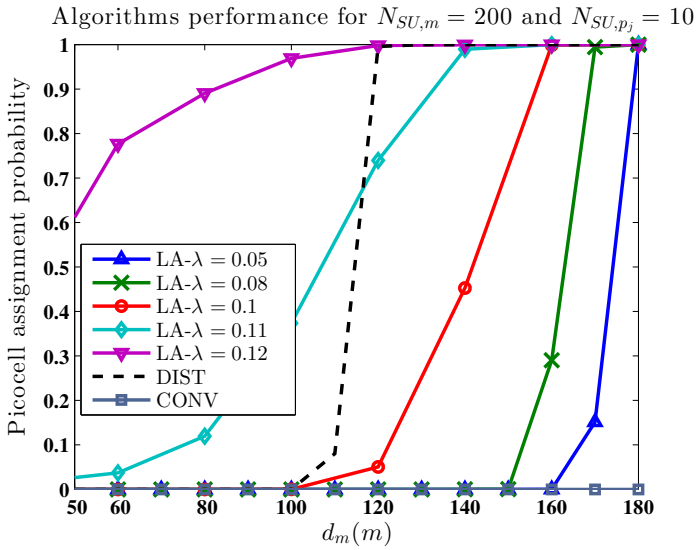


Figure 8.5: Probabilité d'association par picocellule (CONV vs DIST vs LA) pour un nombre $N_{\mathrm{SU},m} = 200$ d'utilisateurs statiques en macrocellules et un nombre $N_{\mathrm{SU},p_j} = 10$ d'utilisateurs statiques en picocellules pour différents λ.

Table 8.1: Gain en délai pour $N_{\mathrm{SU},m} = 200$, $N_{\mathrm{SU},p_j} = 10$ et différents λ

$\boldsymbol{\lambda}$	**0.05**	**0.08**	**0.1**	**0.11**	**0.12**
$\bar{\mathcal{D}}_{\mathrm{LA}}$ (sec)	0.0701	0.0992	0.1261	0.1125	0.0621
$\bar{\mathcal{D}}_{\mathrm{DIST}}$ (sec)	0.0621	0.0807	0.1109	0.1470	0.2808
$\bar{\mathcal{D}}_{\mathrm{CONV}}$ (sec)	0.0743	0.1160	0.1864	0.2708	0.5496
$\bar{\mathcal{D}}_{\mathrm{DIST}}/\bar{\mathcal{D}}_{\mathrm{LA}}$	0.8861	0.8141	0.8798	**1.3066**	**4.5235**
$\bar{\mathcal{D}}_{\mathrm{CONV}}/\bar{\mathcal{D}}_{\mathrm{LA}}$	**1.0596**	**1.1701**	**1.4790**	**2.4073**	**8.8547**

comme la même que sur la figure 8.5. En utilisant l'algorithme CONV, on note que le délai global, $\bar{\mathcal{D}}_{\mathrm{CONV}}$, est identique au délai de la macrocellule (voir la figure 3.4 pour $N_{\mathrm{SU},m} = 200$ et $N_{\mathrm{SU},p_j} = 10$), puisque l'utilisateur est toujours connecté à la macrocellule. L'algorithme DIST retient l'utilisateur connecté à la macrocellule jusqu'à un seuil de distance spécifique et ensuite il l'attache à la picocellule (voir la figure 8.5) avec $P_r(u \in p)$ proche de 1. Ainsi, l'augmentation du délai de la macrocellule, quand elle est surchargée, comparée à celle des picocellules, augmente le délai global, $\bar{\mathcal{D}}_{\mathrm{Dist}}$. L'algorithme LA conserve une politique conservatrice lorsque la charge est faible dans la macrocellule, c.-à-d., que l'utilisateur reste associé à la macrocellule. Partant, $P_r(u \in p)$ est bas même si l'utilisateur est proche de la picocellule. Cette politique donne des délais globaux légèrement plus élevés lorsque $\lambda \leq 0,1$ par rapport à la DIST, puisque cette dernière connecte l'utilisateur à la cellule avec le délai le plus bas (c.-à-d., picocellule) plus tôt. Par conséquent, de HO inutiles vers la picocellule et les surdébits de signalisation associés sont évités avec l'approche LA.

D'un autre côté, si la charge est élevée dans la macrocellule, l'algorithme LA associe l'utilisateur à la picocellule sous-chargée plus tôt que l'algorithme DIST qui ne supporte pas la charge, donnant des performances significatives au délai global, $\bar{\mathcal{D}}_{\mathrm{LA}}$. Les rapports respectifs des algorithmes CONV/DIST sur l'algorithme LA (c.-à-d., $\bar{\mathcal{D}}_{\mathrm{CONV}}/\bar{\mathcal{D}}_{\mathrm{LA}}$, $\bar{\mathcal{D}}_{\mathrm{DIST}}/\bar{\mathcal{D}}_{\mathrm{LA}}$) les délais globaux sont

Table 8.2: Notation des paramètres

Paramètre	Decription	Direction
$\mathrm{RSRP}_{j,i}^{\mathrm{D\ or\ U}}$	Puissance reçue de signal de référence	descente/montée
$\mathrm{SINR}_{j,i}^{\mathrm{D\ or\ U}}$	Signal sur interférence et bruit	descente/montée
$\mathrm{SINR}_{\mathrm{th}}$	Seuil de signal sur interférence et bruit	-
$B_j^{\mathrm{D\ or\ U}}$	Nombre total de PRB pour la $j^{\text{ème}}$ BS	descente/montée
$W_j^{\mathrm{D\ or\ U}}$	Bande passante par PRB en Hz	descente/montée
N_0	Densité de bruit thermique	descente/montée
$R_{j,i}^{\mathrm{D\ or\ U}}$	Débit du canal physique	descente/montée
$\hat{R}_i^{\mathrm{D\ or\ U}}$	Débit de trafic demandé	descente/montée
$P_{j,i}^{\mathrm{D}}$	Puissance de transmission de la BS de PRB en descente	descente
$P_{i,j}^{\mathrm{U}}$	Puissance de transmission de la BS de PRB en montée	montée
α, P_0	Contrôle de paramètres de puissance en boucle ouverte	montée
$P_i^{\max}$	Puissance maximale de transmission de l' $i^{\text{ème}}$ UE	montée

présentés pour montrer le gain correspondant dans le tableau 8.1. La même tendance vaut pour les valeurs supérieures de $N_{\mathrm{SU},m}$, mais avec des valeurs différentes de λ, comme expliqué dans la figure 3.4. Enfin, des gains significatifs sont fournis dans des scénarios de forte charge (c.-à-d., $\simeq 4$ et $\simeq 8$ par rapport à l'approche CONV et DIST).

Le travail lié à ce chapitre est:

- K. Alexandris, N. Sapountzis, N. Nikaein, and T. Spyropoulos, "Load-aware handover decision algorithm in next-generation HetNets," in IEEE WCNC 2016, 3-6 April 2016, Doha, Qatar.

8.5 Chapitre 4 - "Utility-Based Resource Allocation under Multi-Connectivity in LTE"

Dans ce chapitre, nous examinons un problème d'allocation de ressources sous multi-connectivité dans un réseau LTE et proposons une allocation de ressource UPF (utilité par proportions équitables) qui prend en charge la QoS en termes de débits demandés. Nous évaluons la politique proposée avec l'allocation des ressources de l'équité proportionnelle (PF) grâce à des simulations approfondies et caractérisons le gain de performance tant du point de vue de l'utilisateur que du réseau dans différentes conditions.

8.5.1 Modèle de système et hypothèses de problème

Nous considérons une zone $\mathcal{L} \subset \mathbb{R}^2$ servie par un ensemble de BS $\mathcal{B} = \{b_1, \cdots, b_{|\mathcal{B}|}\}$ et un ensemble d'UE $\mathcal{U} = \{u_1, \cdots, u_{|\mathcal{U}|}\}$ est distribué dans cette zone. Dans la suite, nous examinons notre modèle plus en détail: Sur la base du modèle de système susmentionné, nous formulons le problème d'optimisation dans cette section.

A.1-Mobility: Nous assumons la mobilité de l'utilisateur, ce qui implique que tous les paramètres liés à l'emplacement changent dans le temps.

A.2-Débit de données physique: La $j^{\text{ème}}$ BS peut délivrer un débit de transmission de données physique maximum $R^{\mathrm{D}}_{b_j,u_i}$ à un UE. Le débit de données physique de la descente en bits par seconde (bps) est donné dans l'équation (8.4) basée sur la formule de capacité de Shannon:

$$R^{\mathrm{D}}_{b_j,u_i} = B^{\mathrm{D}}_{b_j} W^{\mathrm{D}}_{b_j} \log_2 \left(1 + \ \mathrm{SINR}^{\mathrm{D}}_{b_j,u_i}\right), \tag{8.4}$$

où $B^{\mathrm{D}}_{b_j}$ est le nombre total de PRB en descente de la $j^{\text{ème}}$ BS ($B^{\mathrm{U}}_{b_j}$ pour montée). En ce qui concerne le débit de données physique de l' $i^{\text{ème}}$ UE à la $j^{\text{ème}}$ BS en montée est comme $R^{\mathrm{U}}_{u_i,b_j}$.

A.3-Contrôle de puissance: Le contrôle de puissance en boucle ouverte est appliqué en montée et chaque UE compense la perte de chemin $L^{\mathrm{U,dB}}_{u_i,b_j}$ et effets d'ombrage basés sur les paramètres de contrôle de puissance (c.-à-d., $\alpha, P_0^{\mathrm{dBm}}$). Ainsi, la puissance transmise de chaque PRB de la $i^{\text{ème}}$ UE à la $j^{\text{ème}}$ BS est:

$$P^{\mathrm{U,dBm}}_{u_i,b_j} = \min\left(P^{\mathrm{max,dBm}}_{u_i}, P_0^{\mathrm{dBm}} + \alpha \cdot L^{\mathrm{U,dB}}_{u_i,b_j}\right), \tag{8.5}$$

où $P^{\mathrm{max,dBm}}_{u_i}$ est la puissance maximale transmise de l' $i^{\text{ème}}$ UE. Cependant, en descente, aucun algorithme de contrôle de puissance n'est pas appliqué et la puissance transmise de chaque BS à tous les UE est notée $P^{\mathrm{D,dBm}}_{b_j,u_i} = P^{\mathrm{D,dBm}}_{b_j}$. On notera que l'asymétrie des puissances transmises en montée/descente au contrôle de puissance en montée différencie les valeurs en montée/descente de SINR.

A.4-Multi-connectivité: Dans le cadre de la connectivité multiple, les utilisateurs peuvent être associés et communiquer avec plusieurs BS en même temps. Nous supposons que la capacité de connectivité multiple existe en montée/descente pour tous les UE et qu'un UE peut être connecté à une BS si en montée et en descente SINR dépassent un seuil prédéfini, c.-à-d., $\mathrm{SINR}_{\mathrm{th}}$. Par conséquent, nous définissons un ensemble $\mathcal{E} \triangleq \{(u_i, b_j), (b_j, u_i) : \min\left(\mathrm{SINR}^{\mathrm{U}}_{u_i,b_j}, \mathrm{SINR}^{\mathrm{D}}_{b_j,u_i}\right) > \mathrm{SINR}_{\mathrm{th}}\}$ qui représente toutes les connexions possibles entre UE et BS.

A.5-Routage-local: En principe, le routage dans le réseau backhaul est nécessaire même pour le trafic d'utilisateur à utilisateur servi par la même BS. Dans ce cas, le trafic n'est pas routé via le réseau central, par exemple le cœur de reseau (EPC) LTE, mais seulement via l'intermédiaire BS [86]. Nous nous concentrons sur le flux de trafic utilisateur-utilisateur qui peut être routé localement, c.-à-d., que les deux utilisateurs sont connectés à au moins un BS.

A.6-Paires d'utilisateurs actifs: Basé sur l'exigence SINR minimale définie dans **B.1**, un ensemble qui comprend toutes les paires d'utilisateurs actives desservies par la $j^{\text{ème}}$ BS est défini comme $\mathcal{C}_{b_j} \triangleq \{(u_i, u_q) : (u_i, b_j), (b_j, u_q) \in \mathcal{E}, u_i \neq u_q\}$ et la paire d'utilisateurs $(u_i, u_q) \in \mathcal{C}_{b_j}$ peut avoir un trafic d'utilisateur à utilisateur acheminé localement via BS b_j. Par conséquent, un ensemble $\mathcal{C} \triangleq \bigcup_{b_j \in \mathcal{B}} \mathcal{C}_{b_j}$ est formé comme l'union de toutes les paires d'utilisateurs actifs. Enfin, deux ensembles sont définis pour chaque utilisateur: $\mathcal{D}_{u_i} \triangleq \{u_q : (u_i, u_q) \in \mathcal{C}\}$ comprend tous les UEs destinés à partir du $i^{\text{ème}}$ UE et $\mathcal{S}_{u_i} \triangleq \{u_q : (u_q, u_i) \in \mathcal{C}\}$ comprend tous les UE source qui peuvent transport du trafic vers l' $i^{\text{ème}}$ UE.

A.7-Débit demandé pour le flux de trafic: Il correspond au débit demandé $\hat{R}_{u_i,u_q}$ déterminé comme le débit total requis par un groupe d'applications U2U s'exécutant en haut en utilisant une connexion établie de bout en bout de la paire d'utilisateurs (u_i, u_q), qui peut passer par n'importe quelle BS intermédiaire via local-routing.

8.5.2 Fonction d'utilité

Notre objectif ici est d'allouer de manière optimale la ressource au trafic utilisateur-utilisateur en fonction de la fonction d'utilité appliquée. Nous définissons $x^{\mathrm{U}}_{b_j,(u_i,u_q)}, x^{\mathrm{D}}_{b_j,(u_i,u_q)} \in \mathcal{X}$ comme pourcentage du PRB alloué au total des PRB sous forme décimale selon la direction en montée/descente pour transporter le trafic utilisateur-utilisateur de la paire d'utilisateurs (u_i, u_q) à BS b_j. Dans ce qui suit, nous introduisons deux fonctions d'utilité différentes: l'une fournit une équité proportionnelle et l'autre étend l'équité proportionnelle en tenant compte de la QoS.

8.5.2.1 Équité proportionnelle (PF)

En tant que première fonction d'utilité, nous exploitons une fonction d'utilité logarithmique similaire à celle de [87] qui réalise une allocation de ressources "proportionnellement équitable". Cette fonction d'utilité est donnée comme:

$$\Phi\left(y\right) = \log\left(y\right), \tag{8.6}$$

8.5.2.2 Utilité par proportions équitables (UPF)

Dans l'ancienne fonction d'utilité, un inconvénient important réside dans l'absence de prise en compte de la qualité de service. Pour incorporer la QoS dans la fonction d'utilité, nous utilisons la famille des fonctions sigmoïdes comme:

$$\Phi\left(y\right) = \log\left(S\left(y,\gamma,\beta\right)\right) = \log\left(\frac{1}{1+e^{-\gamma(y-\beta)}}\right), \tag{8.7}$$

$\gamma > 0$, $y \geq 0$, où β est le débit demandé défini en **A.7** de la section 8.5.1. La fonction $\Phi\left(y\right)$ est le logarithme naturel de la fonction d'utilité sigmoïde, $S\left(y,\gamma,\beta\right)$, couramment utilisé dans la littérature (voir [89]) qui préserve le débit demandé.

Sous l'hypothèse du régime de multi-connectivité (voir **A.4** de la section 8.5.1), nous pouvons introduire notre fonction objective finale basée sur toute fonction d'utilité $\Phi\left(\cdot\right)$ présenté à l' équation (8.6) ou à l' équation (8.7) comme suit:

$$U\left(\mathbf{x}^{\mathrm{U}}_{u_i,u_q}, \mathbf{x}^{\mathrm{D}}_{u_i,u_q}\right) \triangleq \Phi\left(\sum_{b_j\in\mathcal{B}} Q\left(x^{\mathrm{U}}_{b_j,(u_i,u_q)}, x^{\mathrm{D}}_{b_j,(u_i,u_q)}\right)\right), \tag{8.8}$$

où where $\mathbf{x}^{\mathrm{U}}_{u_i,u_q} \triangleq \left[x^{\mathrm{U}}_{b_1,(u_i,u_q)}, \cdots, x^{\mathrm{U}}_{b_{|B|},(u_i,u_q)}\right]^T$, $\mathbf{x}^{\mathrm{D}}_{u_i,u_q} \triangleq \left[x^{\mathrm{D}}_{b_1,(u_i,u_q)}, \cdots, x^{\mathrm{D}}_{b_{|B|},(u_i,u_q)}\right]^T$.

La fonction $Q\left(\cdot\right)$ représente le débit alloué de la paire d'utilisateurs (u_i, u_q) et s'exprime comme suit:

$$Q\left(x^{\mathrm{U}}_{b_j,(u_i,u_q)}, x^{\mathrm{D}}_{b_j,(u_i,u_q)}\right) \triangleq \min\left(x^{\mathrm{U}}_{b_j,(u_i,u_q)} R^{\mathrm{U}}_{u_i,b_j}, x^{\mathrm{D}}_{b_j,(u_i,u_q)} R^{\mathrm{D}}_{b_j,u_q}\right) \tag{8.9}$$

Il est à noter que l'opérateur minimum garantit le même débit alloué en montée et en descente pour un trafic utilisateur à utilisateur.

8.5.3 Formulation du problème

Basé sur le modèle de système proposé, les hypothèses et les fonctions d'utilité, nous présentons notre formulation du problème. Le problème tombe dans la catégorie de la maximisation de l'utilité du réseau pour l'allocation des ressources et est donné dans l'équation (8.10).

Fonction objective: L'objectif est d'allouer la ressource $x^{\mathrm{U}}_{b_j,(u_i,u_q)}, x^{\mathrm{D}}_{b_j,(u_i,u_q)} \in \mathcal{X}$ à chaque paire d'utilisateurs afin de maximiser la fonction d'utilité réseau agrégée sur toutes les paires d'utilisateurs qui échangent du trafic. Les deux fonctions d'utilité susmentionnées peuvent être utilisées.

$$
\begin{aligned}
\max_{\mathcal{X}} \quad & \sum_{(u_i,u_q)\in\mathcal{C}} U\left(\mathbf{x}^{\mathrm{U}}_{u_i,u_q}, \mathbf{x}^{\mathrm{D}}_{u_i,u_q}\right) \\
\text{s.c.} \quad & \sum_{(u_i,u_q)\in\mathcal{C}_{b_j}} x^{\mathrm{U}}_{b_j,(u_i,u_q)} \leq 1, \ \forall\ b_j, \\
& \sum_{(u_i,u_q)\in\mathcal{C}_{b_j}} x^{\mathrm{D}}_{b_j,(u_i,u_q)} \leq 1, \ \forall\ b_j, \\
& \sum_{b_j\in\mathcal{B}} \sum_{u_q\in\mathcal{D}_{u_i}} x^{\mathrm{U}}_{b_j,(u_i,u_q)} B^{\mathrm{U}}_{b_j} \leq B^{\mathrm{U}}_{u_i}, \ \forall\ u_i, \\
& \sum_{b_j\in\mathcal{B}} \sum_{u_q\in\mathcal{S}_{u_i}} x^{\mathrm{D}}_{b_j,(u_q,u_i)} B^{\mathrm{D}}_{b_j} \leq B^{\mathrm{D}}_{u_i}, \ \forall\ u_i, \\
& \sum_{b_j\in\mathcal{B}} \sum_{u_q\in\mathcal{D}_{u_i}} x^{\mathrm{U}}_{b_j,(u_i,u_q)} P^{\mathrm{U}}_{u_i,b_j} B^{\mathrm{U}}_{b_j} \leq P^{\max}_{u_i}, \ \forall\ u_i.
\end{aligned} \tag{8.10}
$$

Contraintes: Ici, nous présentons une description détaillée pour chacun d'entre eux:

1. Les deux premières contraintes garantissent que le nombre de PRB alloués (exprimé sous forme de pourcentage décimal sur le total des PRB) à chaque BS b_j pour tous les UE ne dépassera pas le nombre total de PRB en montée et en descente.

2. Les troisième et quatrième contraintes assurent que le nombre de PRB alloués à chaque UE parmi toutes les BS connectées ne dépassera pas le nombre maximum de PRB alloués, $B^{\mathrm{U}}_{u_i}, B^{\mathrm{D}}_{u_i}$, du $i^{\text{ème}}$ UE, respectivement. Plus précisément, ces contraintes prennent en compte tout le trafic utilisateur-utilisateur lié au $i^{\text{ème}}$ UE comme $u_q \in \mathcal{D}_{u_i}$ de tous les UE destinataires dans montée et $u_q \in \mathcal{S}_{u_i}$ de tous les UE source en descente par n'importe quelle BS ($b_j \in \mathcal{B}$). En outre, ces contraintes reposent sur la capacité de l'utilisateur définie comme catégorie UE dans 3GPP TS36.306.

3. Enfin, la dernière contrainte est liée au mécanisme de contrôle de puissance introduit dans **A.3** de la section 8.5.1. Il limite la puissance totale transmise de l' $i^{\text{ème}}$ UE à toutes les BS connectées pour être dans sa limite de puissance comme $P^{\max}_{u_i}$.

Le problème proposé dans l'équation (8.10) est prouvé être convexe.

8.5.4 Résultats de la simulation

Les principaux résultats de la simulation sont présentés comme suit:

Table 8.3: Comparaison: connectivité unique et multi-connectivité

Nombre d'UE dans le BS $b_1/b_2/b_3$	**Métrique de performance**	**Connectivité unique**	**Multi-connectivité**
Cas de sous-chargé: 2/2/2	BS connecté	1	2.07
	Pairs d'UE connectés	6	17.52
	Débit utilisateur agrégé	0.99 Mbps	20.04 Mbps
Cas de charge moyen: 6/2/2	BS connecté	1	1.34
	Pairs d'UE connectés	34	49.95
	Débit utilisateur agrégé	11.68 Mbps	46.73 Mbps
Cas de surchargé: 6/6/6	BS connecté	1	1.45
	Pairs d'UE connectés	90	162.22
	Débit utilisateur agrégé	55.04 Mbps	57.01 Mbps

8.5.4.1 Comparaison: connectivité unique et multi-connectivité

Premièrement, nous comparons les performances d'un cas de connectivité unique hérité avec celle d'une connectivité multiple en utilisant la même fonction d'utilité PF. Le cas unique à connexion unique considéré associe chaque UE à une seule BS et permet seulement à chaque UE de communiquer avec la BS associée en montée et en descente en termes de RSRP le mieux reçu. On peut voir dans le tableau que le nombre de BS connectées par UE augmente avec la multi-connectivité, en particulier pour les scénarios sous-chargés. Ceci est dû au fait que l'interférence inter-cellulaire induite est mineure pour un scénario sous-chargé. De plus, nous observons un plus grand nombre de paires d'utilisateurs connectés dans la multi-connectivité car chaque UE est capable d'émettre et de recevoir du trafic vers un plus grand ensemble d'UE via le routage local à travers différentes BS. Cet avantage devient important du scénario surchargé qui permet une plus grande diversité de trafic parmi les utilisateurs. Même le nombre d'UE augmente, le surdébit de signalisation du réseau central est réduit en bénéficiant de l'adaptation du routage local de la configuration réseau actuelle. En comparant le débit d'utilisateurs agrégé de toutes les paires d'utilisateurs, nous remarquons que le gain de performance est significativement plus élevé dans le scénario sous-chargé suivi du scénario à charge inégale. Ce gain est dû à la présence d'UE dans toutes les BS disponibles, ce qui est effectivement l'un des mérites attendus de la multi-connectivité.

8.5.4.2 Impact de γ sur le problème UPF

La figure 8.6 compare l'impact de γ sur la fonction d'utilité UPF en termes de la représentation de la fonction de densité de probabilité (PDF) du débit alloué agrégé ζ_{u_i,u_q} pour toutes les paires d'utilisateurs. Pour simplifier, nous fournissons seulement le résultat avec $\hat{R} = 1$ Mbps mais les mêmes phénomènes peuvent être observés pour d'autres débits demandés. Tout d'abord, on peut observer que la queue du tracé PDF est plus longue avec un plus petit γ, c.-à-d., que la queue est la plus longue des trois quand $\gamma = 5/\hat{R}$. C'est parce que la fonction sigmoïde avec un plus petit γ tend à être plus linéaire et fait que le débit alloué agrégé dépasse le débit demandé comme le montre la figure 4.2(a). Cette dernière approche est meilleure du point de vue du réseau.

De plus, nous observons qu'un nombre significatif de paires d'utilisateurs ont un débit alloué

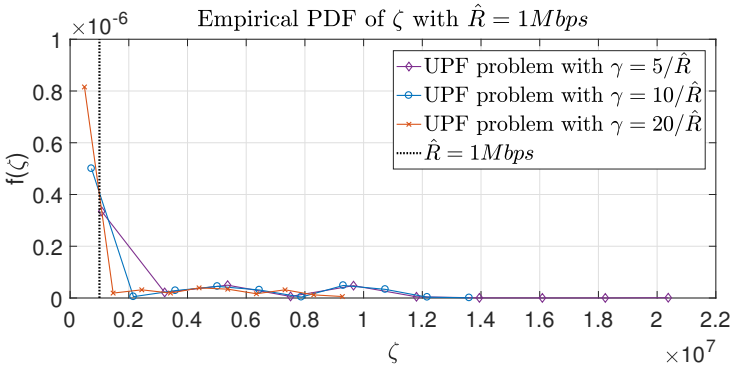


Figure 8.6: Tracé PDF d'UPF pour différents γ

agrégé plus faible ($\zeta_{u_i,u_q} < 0,4$ Mbps) quand γ est grand, $\gamma = 20/\hat{R}$, comparé au cas où γ est petit, $\gamma = 5/\hat{R}$. Ceci est aussi dû à la forme de la fonction sigmoïde dans la figure 4.2(a) dans laquelle la fonction avec un plus grand γ est plus proche de la fonction step et préfère servir la paire d'utilisateurs proche du QoS exigence. En ce sens, certaines paires d'utilisateurs qui ont du mal à atteindre le débit demandé en raison de la mauvaise condition SINR seront allouées avec un débit de données agrégé plus faible ou même en file d'attente (par exemple, $\zeta_{u_i,u_q} = 0$). Cette dernière approche est meilleure du point de vue de l'utilisateur. Par conséquent, l'ajustement de la forme de la fonction sigmoïde sur la base des paramètres γ peut avoir un impact sur la politique d'allocation mettant en évidence le compromis entre l'équité et la perspective de l'utilisateur et du réseau.

Le travail lié à ce chapitre est:

- K. Alexandris, C.-Y. Chang, K. Katsalis, N. Nikaein, and T. Spyropoulos, "Utility-Based Resource Allocation under Multi-Connectivity in Evolved LTE," in IEEE VTC 2017, Sep 24-27, 2017, Toronto, Canada.

8.6 Chapitre 5 - "Multi-Connectivity Resource Allocation with Limited Backhaul Capacity in Evolved LTE"

Dans ce chapitre, nous examinons un problème d'allocation de ressource UPF sous multi-connectivité en LTE et proposons une utilité considérant la QoS des utilisateurs avec des limitations de capacité backhaul par rapport au chapitre 4 où seules les contraintes radio apparaissent. La politique proposée est comparée à l'allocation des ressources PF par le biais de simulations approfondies. Les résultats présentés montrent que la connectivité multiple surpasse la connectivité unique en termes de débit agrégé du réseau et de QoS des utilisateurs la satisfaction dans différentes études de cas de réseau, c.-à-d., des scénarios de cellules vides et chargées avec une capacité de liaison fixe et variable.

8.6.1 Modèle de système et hypothèses de problème

Nous considérons un ensemble de BS $\mathcal{B} = \{b_1, \cdots, b_{|\mathcal{B}|}\}$ connecté au réseau central via un réseau backhaul pour servir un ensemble de UE $\mathcal{U} = \{u_1, \cdots, u_{|\mathcal{U}|}\}$. Pour mieux élucider les

Table 8.4: Notations additionnelles des paramètres

Paramètres	Description	Direction
$M_i^{\text{D or U}}$	Nombre de PRB maximal de l' $i^{ème}$ UE	descente/montée
$\hat{R}_i^{\text{D or U}}$	Debit de trafic demandé	descente/montée
$C_{h,j}^{\text{D or U}}$	Capacité du lien backhaul en bps	descente/montée
$f_j^{\text{D or U}}$	Fréquence de porteuse en Hz	descente/montée

hypothèses du chapitre actuel en différence avec celles mentionnées dans la section 8.5.1, nous les redéfinissons afin de souligner la nouvelle configuration système considérée:

A.1-Fréquence de porteuse: Dans ce chapitre, nous appliquons le déploiement RAT inter-fréquence unique et dénote la fréquence porteuse en descente comme f_j^{D} pour la $j^{\text{ème}}$ BS [2].

A.2-Réseau Backhaul: Dans notre cas, nous supposons que chaque BS est connectée à la passerelle du réseau central via une connexion dédiée et directe (c.-à-d., topologie en étoile), voir la figure 5.1. Enfin, nous caractérisons la capacité provisionnée de la liaison backhaul à partir de la $j^{\text{ème}}$ BS en montée et en descente comme $C_{h,j}^{\text{U}}, C_{h,j}^{\text{D}}$ respectivement.

A.3-Débit de trafic demandé: Les UE sont supposés envoyer et recevoir du trafic provenant des directions en montée et en descente. L'ensemble des UE qui transmettent en montée est défini comme $\mathcal{S}$ et l'ensemble des UE reçus en descente est défini comme $\mathcal{D}$. Plus précisément, un utilisateur u_i envoie du trafic en montée en demandant un débit de données $\hat{R}_i^{\text{U}}$ (débit de trafic demandé en montée) et reçoit du trafic en descente en demandant un débit $\hat{R}_i^{\text{D}}$ (débit de trafic demandé en descente).

8.6.2 Fonction d'utilité

Nous définissons $x_{i,j}^{\text{U}}, x_{j,i}^{\text{D}} \in \mathcal{X}$ comme les nombres en décimaux sur le total des PRB alloué par BS b_j à un utilisateur u_i pour envoyer ou recevoir du trafic en montée et en descente.

$$U_1\left(\mathbf{x}_i^{\text{U}}\right) \triangleq \Phi\left(\left(\mathbf{x}_i^{\text{U}}\right)^T \cdot \mathbf{R}_i^{\text{U}}\right) = \Phi\left(\sum_{b_j \in \mathcal{B}} x_{i,j}^{\text{U}} R_{i,j}^{\text{U}}\right), \tag{8.11}$$

$$U_2\left(\mathbf{x}_q^{\text{D}}\right) \triangleq \Phi\left(\left(\mathbf{x}_q^{\text{D}}\right)^T \cdot \mathbf{R}_q^{\text{D}}\right) = \Phi\left(\sum_{b_j \in \mathcal{B}} x_{j,q}^{\text{D}} R_{j,q}^{\text{D}}\right). \tag{8.12}$$

Notez que $\left(\mathbf{x}_i^{\text{U}}\right)^T \cdot \mathbf{R}_i^{\text{U}}$ et $\left(\mathbf{x}_q^{\text{D}}\right)^T \cdot \mathbf{R}_q^{\text{D}}$ incluent le débit agrégé sur toutes les BS connectées en montée ou en descente pour u_i et u_q. La valeur de $\hat{R}_i^{\text{U}}$ ou $\hat{R}_q^{\text{D}}$ est assignée à β dans l'équation (8.7) (fonction sigmoïde) en montée et en descente.

[2] En montée, la notation devient, par exemple, f_j^{U} et la même règle est appliquée pour le reste des paramètres du système.

8.6.3 Formulation du problème

Sur la base du modèle de système susmentionné, des hypothèses et des fonctions d'utilité, nous formulons ensuite le problème d'allocation des ressources. Un tel problème tombe dans la catégorie de la maximisation de l'utilité du réseau pour l'allocation des ressources et est donné comme suit:

$$
\begin{aligned}
\max_{\mathcal{X}} \quad & \sum_{u_i \in \mathcal{S}} U_1\left(\mathbf{x}_i^{\mathrm{U}}\right) + \sum_{u_q \in \mathcal{D}} U_2\left(\mathbf{x}_q^{\mathrm{D}}\right) \\
\text{s.c. } \mathbf{C1:} \quad & \sum_{u_i \in \mathcal{S},\ (u_i, b_j) \in \mathcal{E}} x_{i,j}^{\mathrm{U}} \leq 1,\ \forall\ b_j \in \mathcal{B}, \\
\mathbf{C2:} \quad & \sum_{u_q \in \mathcal{D},\ (b_j, u_q) \in \mathcal{E}} x_{j,q}^{\mathrm{D}} \leq 1,\ \forall\ b_j \in \mathcal{B}, \\
\mathbf{C3:} \quad & \sum_{b_j \in \mathcal{B},\ (u_i, b_j) \in \mathcal{E}} x_{i,j}^{\mathrm{U}} B_j^{\mathrm{U}} \leq M_i^{\mathrm{U}},\ \forall\ u_i \in \mathcal{U}, \\
\mathbf{C4:} \quad & \sum_{b_j \in \mathcal{B},\ (b_j, u_q) \in \mathcal{E}} x_{j,q}^{\mathrm{D}} B_j^{\mathrm{D}} \leq M_q^{\mathrm{D}},\ \forall\ u_q \in \mathcal{U}, \\
\mathbf{C5:} \quad & \sum_{b_j \in \mathcal{B},\ (u_i, b_j) \in \mathcal{E}} x_{i,j}^{\mathrm{U}} B_j^{\mathrm{U}} P_{i,j}^{\mathrm{U}} \leq P_i^{\max},\ \forall\ u_i \in \mathcal{U}, \\
\mathbf{C6:} \quad & \sum_{u_i \in \mathcal{S},\ (u_i, b_j) \in \mathcal{E}} x_{i,j}^{\mathrm{U}} R_{i,j}^{\mathrm{U}} \leq C_{h,j}^{\mathrm{U}},\ \forall\ b_j \in \mathcal{B}, \\
\mathbf{C7:} \quad & \sum_{u_q \in \mathcal{D},\ (b_j, u_q) \in \mathcal{E}} x_{j,q}^{\mathrm{D}} R_{j,q}^{\mathrm{D}} \leq C_{h,j}^{\mathrm{D}},\ \forall\ b_j \in \mathcal{B}.
\end{aligned} \tag{8.13}
$$

Une explication détaillée du problème d'optimisation proposé est donnée présentant la fonction objectif ainsi que les sept contraintes:

Fonction objective: L'objectif est d'allouer de manière optimale les ressources $x_{i,j}^{\mathrm{U}}, x_{j,q}^{\mathrm{D}} \in \mathcal{X}$ pour chaque utilisateur $u_i, u_q \in \mathcal{U}$ pour maximiser la somme des fonctions d'utilité de tous les utilisateurs en appliquant les fonctions d'utilité par utilisateur (voir equation (8.11) et equation (8.12)).

Contraintes: Nous élaborons toutes les contraintes comme suit:

- **C1-C2:** Ces contraintes garantissent que le nombre de PRB alloués (exprimés en pourcentage en décimales sur le total des PRB) de chaque BS b_j à tous les utilisateurs ne dépasse pas le nombre total de PRB en montée et en descente.
- **C3-C4:** Ces deux contraintes assurent que le nombre total de PRB alloués à chaque utilisateur parmi toutes les BS connectées ne dépassera pas son nombre maximum de PRB alloués, $M_i^{\mathrm{U}}, M_q^{\mathrm{D}}$, en montée et en descente, respectivement.
- **C5:** Une telle contrainte est liée au mécanisme de contrôle de puissance en montée et elle restreint que la puissance de transmission totale de l' $i^{ème}$ UE à toutes les BS connectées ne peut pas dépasser ses limites de puissance $P_i^{\max}$.
- **C6-C7:** Ces deux contraintes sont liées à la liaison backhaul et elles assurent que les débits agrégés à chaque BS b_j ne dépasseront pas les capacités backhaul provisionnées dans les directions en montée et en descente.

Le problème proposé dans l'équation (8.13) est prouvé d'être convexe.

8.6.4 Résultats de la simulation

Deux scénarios de simulation principaux sont considérés comme suit:

- **Scénario A-Cellule vide**: Il y a 1 BS qui ne sert pas d'UE (c.-à-d., charge de trafic 0%) tandis que les 2 BS restantes servent un nombre d'UE non-nulle. La charge de trafic cellulaire vide est utilisée dans l'évaluation du rapport 3GPP (TR 25.927) et caractérise certains résultats de mesure dans le monde réel. Nous utilisons la notation $0/z/z$ pour représenter ce scénario où z est le nombre d'UE desservis par BS non différent de zéro.
- **Scénario B-Cellule chargée**: Toutes les cellules servent un nombre d'UE non-nulle et nous utilisons la notation $z/z/z$ pour représenter ce scénario. La valeur de z est 2 ou 4 dans nos simulations.

En outre, nous définissons deux paramètres de base: (a) le débit agrégé du réseau en tant que somme du débit total alloué à tous les utilisateurs en descente

$$\bar{R}^{\mathrm{D}} \triangleq \sum_{u_q \in \mathcal{D}} \sum_{b_j \in \mathcal{B}} \left(x_{j,q}^{\mathrm{D}}\right)^{\star} R_{j,q}^{\mathrm{D}} \tag{8.14}$$

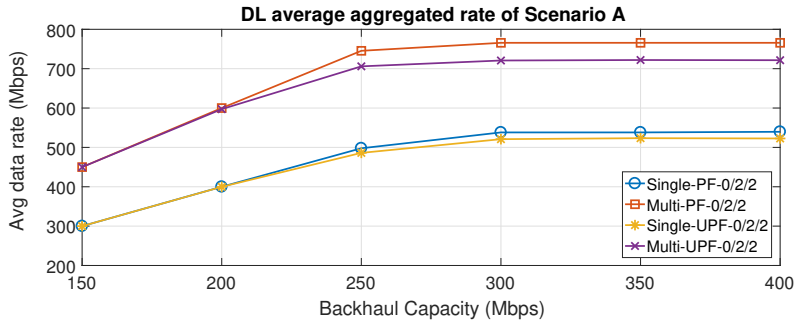
où et $\left(x_{j,q}^{\mathrm{D}}\right)^{\star}$ sont les résultats d'optimisation en descente (même chose pour montée), et (b) le ratio de satisfaction d'utilisateurs en pourcentage d'utilisateurs satisfaits du débit alloué

$$S^{\mathrm{D}} \triangleq \mathrm{Prob}\left\{\sum_{b_j \in \mathcal{B}} \left(x_{j,q}^{\mathrm{D}}\right)^{\star} R_{j,q}^{\mathrm{D}} \geq \hat{R}_q^{\mathrm{D}}\right\} \tag{8.15}$$

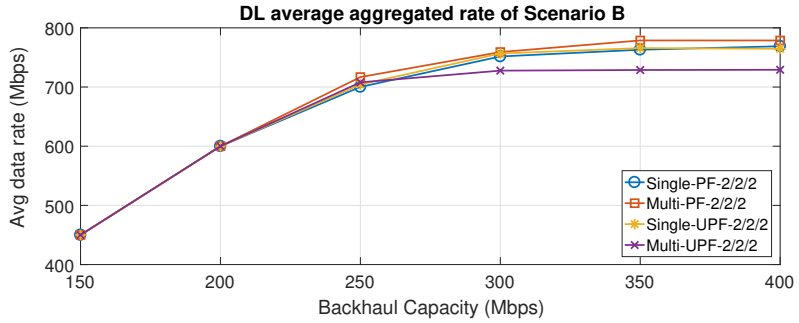
(même chose en montée).

Nous démontrons les résultats en descente avec une capacité de backhaul fixe en considérant la même valeur est appliquée à tous les liens backhaul. Sur la figure 8.7, le débit agrégé moyen est représenté pour les scénarios A et B avec 2 UE par BS sous connectivité unique et multi-connectivité en fonction de la capacité backhaul à la fois pour UPF et PF. Dans les scénarios vides et chargés, le débit agrégé converge lorsque la capacité backhaul est provisionnée pour dépasser 300 Mbit/s dans les cas de connectivité unique et multi-connectivité. En ce qui concerne le scénario A (c.-à-d., cellule vide), on peut voir que la connectivité multiple surpasse la seule même avec une capacité de liaison insuffisante, c.-à-d., moins de 300 Mbit/s, en utilisant la capacité non allouée de la cellule vide. Pour le scénario B, il n'y a pas de différences significatives entre la connectivité unique et la connectivité multiple, car dans la cellule chargée, la connectivité multiple ne peut pas utiliser plus de ressources que la seule par rapport au cas vide, où des ressources supplémentaires sont disponibles. Néanmoins, le cas de multi-connectivité peut mieux refléter le compromis entre atteindre un débit de données plus élevé et atteindre une plus grande satisfaction des utilisateurs (c.-à-d., PF par rapport à UPF). En d'autres termes, un débit plus agrégé est alloué en multi-connectivité afin de mieux satisfaire le débit demandé par les utilisateurs lors de la comparaison avec le cas de connectivité unique, comme indiqué dans le tableau suivant 8.5.

Les résultats du ratio de satisfaction d' utilisateurs sont présentés dans le tableau 8.5 du scénario A et B et nous observons que le ratio de satisfaction peut être augmenté de deux



(a) Scenario A: Cellule vide



(b) Scenario B: Cellule chargée

Figure 8.7: Influence de la capacité backhaul sur débit moyen agrégé

Table 8.5: Ratio de satisfaction d'utilisateurs (%) pour différentes capacités backhaul

Backhaul Capacity (Mbps)	**Single PF**		**Multi PF**		**Single UPF**		**Multi UPF**	
	0/2/2	**2/2/2**	**0/2/2**	**2/2/2**	**0/2/2**	**2/2/2**	**0/2/2**	**2/2/2**
150	∼20	∼20	63	∼20	∼20	∼20	81	∼20
200	49	∼50	95	∼50	51	54	100	55
250	72	65	95	68	83	80	100	91
300	81	72	95	73	90	84	100	93
>350	81	72	95	73	90	84	100	93

manières: (a) appliquer la multi-connectivité (du simple au multi- connectivité) avec PF avec un gain moyen de 3% - 31% et avec un UPF un gain moyen de 7% - 34%, et (b) d'utiliser la fonction d'utilité UPF (de PF vers UPF) connectivité unique un gain moyen de 7% - 10% est inspecté alors qu'en multi connectivité, un gain moyen de 8% - 16% est remarqué. Nous soulignons que la multi-connectivité est bénéfique dans le cas des cellules vides, où elle peut augmenter le ratio de satisfaction d' utilisateurs en exploitant davantage de ressources non allouées. En outre, la fonction UPF peut être utilisée pour accélérer les performances prévues en redistribuant les ressources réseau en fonction du débit demandé par les utilisateurs, en particulier dans le cas de cellule chargée par rapport à celui de PF.

Enfin, les scénarios de cellule vide et chargée sont utilisés pour identifier les limites inférieure et supérieure correspondantes de nos métriques de performance, c.-à-d., les meilleurs et les pires scénarios.

Le travail lié à ce chapitre est:

- K. Alexandris, C.-Y. Chang, N. Nikaein, and T. Spyropoulos, "Multi-Connectivity Resource Allocation with Limited Backhaul Capacity in Evolved LTE," in IEEE WCNC 2018, April 15-18, 2018, Barcelona, Spain.

8.7 Chapitre 6 - "Utility-based Opportunistic Scheduling under Multi-Connectivity with Limited Backhaul Capacity"

Dans ce chapitre, nous revenons sur les scénarios discutés dans le chapitre 5 exploitant l'impact de la variabilité du canal spatio-temporel (évanouissements à grande et petite échelle) via l'ordonnancement opportuniste. Plus précisément, nous examinons un problème d'allocation de ressource opportuniste sous multi-connectivité et capacité de liaison limitée dans LTE évolué en utilisant les deux fonctions d'utilité susmentionnées: PF et UPF. Nous proposons ensuite un algorithme efficace pour traiter le problème NP-difficile formulé. Un tel algorithme est garanti pour produire une solution avec une liaison explicite vers la solution optimale. Enfin, les résultats de la simulation justifient que l'utilisation de la fonction d'utilité UPF avec une planification opportuniste dans la multi-connectivité puisse mieux répondre aux exigences de QoS.

8.7.1 Modèle de système et hypothèses de problème

Nous considérons une zone desservie par un ensemble de BS $\mathcal{B} = \{b_1, \cdots, b_{|\mathcal{B}|}\}$ connecté au cœur de réseau (CN) via le réseau de backhaul et un ensemble d'UE $\mathcal{U} = \{u_1, \cdots, u_{|\mathcal{U}|}\}$ est distribué dans cette zone. La topologie est la même avec celle du chapitre 5 et nous nous concentrons uniquement en descente.

L'hypothèse **A.1** relative au déploiement multi-connectivité inter-fréquence comme décrit dans la section 8.6.1 contient la même chose et quelques variations sur le modèle de canal (c.-à-d., l'évanouissement à petite échelle afin de prendre en compte les erreurs de canal) et les définitions de débit physique par rapport au chapitre 5 sont présentés comme suit:

A.1-Modèle de canal: Dans le RAT LTE, la direction descente utilise l'accès multiple par répartition orthogonale de la fréquence avec la granularité minimale des ressources radio appelée PRB (Physical Resource Block) couvrant une seule tranche de temps dans le domaine temporel et un sous-canal (égal à plusieurs sous-porteuses) domaine. Par conséquent, un ensemble de sous-canaux $\mathcal{K}$ peut être alloué pour chaque BS. Notre modèle de canal considéré inclut à la fois les évanouissements à grande échelle et à petite échelle. L'évanouissement à grande échelle inclut à la fois la perte de chemin et l'ombrage comme $L^{\mathrm{D}}_{j,i}$ de la $j^{ème}$ BS au $i^{\text{ème}}$ UE , alors que la petite échelle comprend l'effet multi-chemin et la propagation Doppler comme $s^{\mathrm{D}}_{j,k,i}(t)$ de la $j^{ème}$ BS au $i^{\text{ème}}$ UE sur le $k^{\text{ème}}$ sous-canal au temps t. Pour plus de simplicité, nous laissons tomber l'index de temps t dans les paragraphes suivants.

A.2-Modèle de connexion et de trafic: Le régime de multi-connectivité est comme dans le chapitre 4 et 5, où le critère de connectivité décrit dans l'hypothèse **A.4** de la section 8.5.1 contient la même chose, l' $i^{\text{ème}}$ UE peut être connecté à la $j^{ème}$ BS si le SINR moyen $\overline{\mathrm{SINR}}^{\mathrm{D}}_{j,i}$

sur tous les sous-canaux est au-dessus d'un seuil, soit $\overline{\text{SINR}}_{\text{th}}$, où

$$\overline{\text{SINR}}^{\text{D}}_{j,i} = \frac{1}{|\mathcal{K}|} \sum_{k \in \mathcal{K}} \text{SINR}^{\text{D}}_{j,k,i}. \tag{8.16}$$

Les hypothèses de capacité limitée comme celles de la section 8.6.1 (voir l'hypothèse **A.2**) sont prises en compte. Nous notons la capacité de backhaul vers la $j^{ème}$ BS comme $C^{\text{D}}_{h,j}$ qui peut potentiellement limiter le débit de données utilisateur réalisable. En outre, l'hypothèse **A.3** présentée dans la section 8.6.1 liée au trafic demandé par les utilisateurs est la même.

8.7.2 Fonction d'utilité

Nous définissons $x^{\text{D}}_{j,k,i} \in \{0,1\}$ comme les variables binaires représentant l'allocation ou non du $k^{ème}$ sous-canal au $i^{ème}$ UE à partir de la $j^{ème}$ BS. Par conséquent, $y^{\text{D}}_{j,k,i} = x^{\text{D}}_{j,k,i} \cdot R^{\text{D}}_{j,k,i}$ est le débit physique correspondant.

Basé sur le modèle de système introduit dans la section 8.6.1, nous formons d'abord 2 vecteurs $\mathbf{x}^{\text{D}}_i = \left[x^{\text{D}}_{1,1,i},...,x^{\text{D}}_{1,|\mathcal{K}|,i}, x^{\text{D}}_{2,1,i},...,x^{\text{D}}_{2,|\mathcal{K}|,i},...,x^{\text{D}}_{|\mathcal{B}|,1,i},...,x^{\text{D}}_{|\mathcal{B}|,|\mathcal{K}|,i}\right]^T$ est la forme vectorielle des variables d'allocation binaire et $\mathbf{R}^{\text{D}}_i = \left[R^{\text{D}}_{1,1,i},...,R^{\text{D}}_{1,|\mathcal{K}|,i}, R^{\text{D}}_{2,1,i},...,R^{\text{D}}_{2,|\mathcal{K}|,i},...,R^{\text{D}}_{|\mathcal{B}|,1,i},...,R^{\text{D}}_{|\mathcal{B}|,|\mathcal{K}|,i}\right]^T$ représente le débit physique vectorisé. Ensuite, la fonction d'utilité du réseau de l' $i^{ème}$ UE est $U\left(\mathbf{x}^{\text{D}}_i\right) \triangleq \Phi\left(\left(\mathbf{R}^{\text{D}}_i\right)^T \cdot \mathbf{x}^{\text{D}}_i\right)$, où $\Phi(\cdot)$ peut être l'une des deux fonctions d'utilité présentées aux équations (8.6) et (8.7). En outre, les arguments des fonctions d'utilité PF et UPF incluent le débit de données agrégé globalement toutes les BS connectées et les sous-canaux alloués. Enfin, la fonction d'utilité UPF assigne $\hat{R}^{\text{D}}_i$ à son argument β (voir l'équation (8.7)).

8.7.3 Formulation du problème

Problème d'optimisation: Le problème tombe dans la catégorie de la maximisation de l'utilité réseau pour l'allocation des ressources et est donné comme suit:

$$\max_{\mathbf{x}^{\text{D}}_i \in \{0,1\}^{|\mathcal{B}|\times|\mathcal{K}|}} \sum_{u_i \in \mathcal{U}} U\left(\mathbf{x}^{\text{D}}_i\right) \tag{8.17}$$

$$\text{s.c.} \quad \textbf{C1:} \sum_{u_i \in \mathcal{U}} \sum_{k \in \mathcal{K}} x^{\text{D}}_{j,k,i} \leq B^{\text{D}}_j, \forall\ b_j \in \mathcal{B},$$

$$\textbf{C2:} \sum_{b_j \in \mathcal{B}} \sum_{k \in \mathcal{K}} x^{\text{D}}_{j,k,i} \leq M^{\text{D}}_i, \forall\ u_i \in \mathcal{U},$$

$$\textbf{C3:} \sum_{u_i \in \mathcal{U}} x^{\text{D}}_{j,k,i} \leq 1, \forall\ b_j \in \mathcal{B}, k \in \mathcal{K}. \tag{8.18}$$

Fonction objective: L'objectif est d'allouer de façon optimale les ressources $x^{\text{D}}_{j,k,i} \in \{0,1\}$ pour chaque UE $u_i \in \mathcal{U}$ pour maximiser la sum des fonctions d'utilité de tous les UE. Les deux fonctions d'utilité PF / UPF décrites dans Section 8.5.2 peuvent être appliquées.

Contraintes: Dans la suite, nous présentons une description détaillée de chaque contrainte; **C1:** Il garantit que le nombre de sous-canaux alloués de BS b_j à tous les UE ne dépassera pas le nombre total de sous-canaux disponibles à b_j, B^{D}_j ; **C2:** Il assure que le nombre total de sous-canaux alloués à UE u_i parmi toutes les BS connectées ne dépassera pas les sous-canaux

allouables maximum vers u_i, M_i^{D} . Il repose sur la capacité utilisateur définie comme la catégorie UE dans 3GPP TS36.306; **C3:** Il garantit que chaque sous-canal unique ne sera pas assigné à plus d'un UE impliquant l'exclusivité du sous-canal.

Le problème d'optimisation formulé est NP-difficile et se révèle être avec une fonction objective submodulaire et monotone et trois contraintes matrioïdes. Enfin, un algorithme glouton est proposé qui peut résoudre efficacement le problème avec une liaison explicite vers la solution optimale.

Dernier point, mais non des moindres, pour satisfaire la limitation de la capacité de backhaul, $C_{h,j}^{\mathrm{D}}$, nous augmentons encore le débit de données alloué au $i^{\text{ème}}$ UE comme:

$$R_i^{\mathrm{D}} \triangleq \sum_{b_j \in \mathcal{B}} \sum_{k \in \mathcal{K}} \left(x_{j,k,i}^{\mathrm{D}}\right)^{\star} R_{j,k,i}^{\mathrm{D}} \cdot \min\left(\frac{C_{h,j}^{\mathrm{D}}}{\sum_{u_i \in \mathcal{U}} \sum_{k \in \mathcal{K}} \left(x_{j,k,i}^{\mathrm{D}}\right)^{\star} R_{j,k,i}^{\mathrm{D}}}, 1\right), \tag{8.19}$$

où $\left(x_{j,k,i}^{\mathrm{D}}\right)^{\star}$ sont les résultats de l'algorithme proposé. En réalité, une telle mise à l'échelle peut être effectuée en ajustant le schéma de modulation et de codage, mais elle sort du cadre de cette étude.

8.7.4 Résultats de la simulation

Nous nous concentrons sur le scénario B - Cellule chargée: Toutes les cellules servent des UE non-nulles et nous utilisons la notation $z/z/z$ pour représenter ce scénario. Ensuite, notre métrique de performance considérée est:

$$E_i^{\mathrm{D}} = \begin{cases} \left\| \frac{R_i^{\mathrm{D}} - \hat{R}_i^{\mathrm{D}}}{\hat{R}_i^{\mathrm{D}}} \right\|, & \text{if } R_i^{\mathrm{D}} < \hat{R}_i^{\mathrm{D}}, \\ 0 \qquad\quad , & \text{o/w.} \end{cases} \tag{8.20}$$

Dans la figure 8.8, les résultats du scénario de cellule vide sont affichés avec plusieurs combinaisons de PF/UPF avec une connectivité unique/multiple. Nous pouvons voir que l'utilisation de la fonction d'utilité UPF peut augmenter le ratio de satisfaction même avec une connectivité unique. Cependant, certaines exigences QoS ne peuvent jamais être remplies, même si nous augmentons continuellement la capacité de backhaul. Cela est dû à l'absence d'utilisation de ressources non allouées à la cellule vide et donc la multi-connectivité montre son avantage. En outre, l'utilisation de la fonction UPF avec une connectivité multiple permet de mieux répondre aux exigences de qualité de service, même avec une capacité de transport moins importante. Ensuite, les résultats du scénario de cellule chargée sont montrés sur la figure 8.8. L'utilisation de plusieurs connexions n'est pas suffisante pour répondre aux exigences de QoS si seulement PF est adopté. Une telle observation est due à l'absence de ressources non allouées provenant de cellules vides; par conséquent, nous devons réorganiser efficacement les ressources allouées via l'utilisation de la fonction UPF. Enfin, la figure 8.8 montre les avantages de l'ordonnancement opportuniste dans le scénario de cellule chargée avec UPF et multi-connectivité par rapport à la non-opportuniste. Notez que l'ordonnancement non opportuniste peut s'appuyer de façon optimale sur le SINR moyen (c.-à-d., $\overline{\mathrm{SINR}}_{j,i}^{\mathrm{D}}$); cependant, son pire cas est basé sur le SINR minimum parmi tous les sous-canaux (c.-à-d., $\min_{k \in \mathcal{K}} \left(\mathrm{SINR}_{j,k,i}^{\mathrm{D}}\right)$). L'opportuniste peut utiliser plus de connexions et fournit une erreur plus petite, quel que soit le meilleur ou le pire des cas d'ordonnancement non opportuniste.

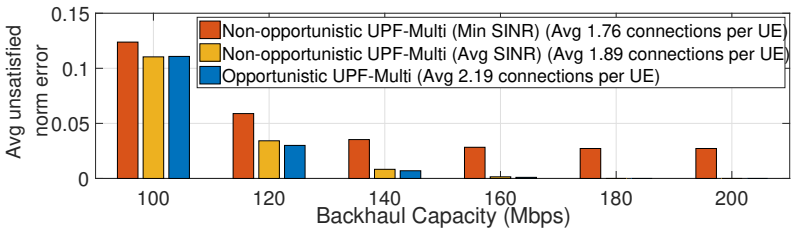


Figure 8.8: Erreur insatisfait par rapport à capacite backhaul $(z/z/z)$

Le travail lié à ce chapitre est:

- K. Alexandris, C.-Y. Chang, N. Nikaein, and T. Spyropoulos, "Utility-based Opportunistic Scheduling under Multi-Connectivity with Limited Backhaul Capacity," in IEEE Wireless Communications Letters, 2018, submitted, under review.

Appendix A

A.1 Capacity analysis-I

The average capacity, assuming shadowing and Rayleigh flat fading [1], is expressed as:

$$\bar{C}\left(d_i\right) = \mathbb{E}_{\psi_i,|h|^2}\left[B\log_2\left(1+\rho_i K\left(d_i/d_r\right)^{-\eta}\psi_i\left|h\right|^2\right)\right], \tag{A.1}$$

where $\rho_i \triangleq P_{T_x,i}/N_0$ denotes the transmit signal-to-noise ratio (SNR) [2], $P_{T_x,i}$ is the transmission power in linear scale, d_i represents the distance from the BS cell i to the user that is greater than a reference distance d_r, B stands for the available bandwidth, N_0 represents the aggregated noise power over B in linear scale, $\psi_i \sim \text{Log-}\mathcal{N}\left(0, \sigma^2_{\text{dB},i}/\xi^2\right)$ represents the shadowing fading in linear scale and $h \sim \mathcal{CN}\left(0,1\right)$ defines the Rayleigh flat fading channel coefficient. To the best of our knowledge, an analytical expression for the averaged capacity does not exist. Therefore, its lower and upper bound, $\bar{C}_b\left(d_i\right)$, $b \in \{L, U\}$, can be easily computed, as follows.

A.1.1 Lower bound

Exploiting the concavity of the logarithmic function, an analytical lower bound is given by:

$$\begin{aligned}
&\bar{C}\left(d_i\right) = \mathbb{E}_{\psi_i,|h|^2}\left[B\log_2\left(1+\rho_i K\left(d_i/d_r\right)^{-\eta}\psi_i\left|h\right|^2\right)\right] \Leftrightarrow \\
&\bar{C}\left(d_i\right) = \mathbb{E}_{\psi_i}\left[\mathbb{E}_{|h|^2}\left[B\log_2\left(1+\rho_i K\left(d_i/d_r\right)^{-\eta}\psi_i\left|h\right|^2\right)\right]\right] \Leftrightarrow \\
&\bar{C}\left(d_i\right) \geq \mathbb{E}_{\psi_i}\left[\mathbb{E}_{|h|^2}\left[B\log_2\left(\rho_i K\left(d_i/d_r\right)^{-\eta}\psi_i\left|h\right|^2\right)\right]\right] \Leftrightarrow \\
&\bar{C}\left(d_i\right) \geq \underbrace{B\left\{\log_2\left(\rho_i K\left(d_i/d_r\right)^{-\eta}\right) - \gamma/\ln\left(2\right)\right\}}_{\bar{C}_L(d_i)},
\end{aligned} \tag{A.2}$$

where γ denotes the Euler-Mascheroni constant.

[1] It is noted that the capacity is based on the received signal before the filtering process, thus Rayleigh fading effect is taken into account.

[2] SINR could also be introduced, but would make the analysis complex.

A.1.2 Upper bound

Using Jensen's inequality for concave functions, the upper bound is computed analytically as follows:

$$\begin{aligned}
&\bar{C}\left(d_i\right) = \mathbb{E}_{\psi_i,|h|^2}\left[B\log_2\left(1+\rho_i K\left(d_i/d_r\right)^{-\eta}\psi_i\left|h\right|^2\right)\right] \Leftrightarrow \\
&\bar{C}\left(d_i\right) = \mathbb{E}_{\psi_i}\left[\mathbb{E}_{|h|^2}\left[B\log_2\left(1+\rho_i K\left(d_i/d_r\right)^{-\eta}\psi_i\left|h\right|^2\right)\right]\right] \Leftrightarrow \\
&\bar{C}\left(d_i\right) \leq \underbrace{B\log_2\left(1+\rho_i K\left(d_i/d_r\right)^{-\eta}\mu_{\psi_i}\right)}_{\bar{C}_U(d_i)}, \qquad \text{(A.3)}
\end{aligned}$$

where $\mu_{\psi_i} = \mathbb{E}\left[\psi_i\right] = \exp\left[\sigma^2_{\mathrm{dB},i}/2\xi^2\right]$.

A.2 Capacity analysis-II

The picocell points x, y are distributed uniformly over a circle according to the following distribution:

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{\pi\left(R^2_{p_j}-d^2_r\right)} & ,\ d_r \leq r \leq R_{p_j}, \\ 0 & ,\ o/w, \end{cases} \qquad \text{(A.4)}$$

where $r = \sqrt{(x-x_{p_j})^2+(y-y_{p_j})^2}$. Average capacity bounds are obtained by averaging over all the possible distances (i.e., over all the possible points) within the picocell on the respective lower and upper capacity bound for each cell i. This can be analytically described as:

$$\bar{C}^i_b = \int_Y\int_X \bar{C}_b\left(d_i\left(x,y\right)\right) f_{X,Y}\left(x,y\right)dxdy,\ \ b\in\{L,U\}. \qquad \text{(A.5)}$$

A.2.1 Lower bound

The average lower bound of the average capacity by applying cartesian to polar coordinates transformation in Eq. (A.5) is computed as:

$$\begin{aligned}
\bar{C}^i_L &= \frac{1}{\pi\left(R^2_{p_j}-d^2_r\right)}\int_0^{2\pi}\int_{d_r}^{R_{p_j}} r\bar{C}_L\left(\tilde{d}_i\left(r,\theta\right)\right)drd\theta = \\
&= B\left[\log_2\left(\rho_i\right) + \frac{1}{\pi\left(R^2_{p_j}-d^2_r\right)}\underbrace{\int_0^{2\pi}\int_{d_r}^{R_{p_j}} r\log_2\left(K\left[\tilde{d}_i\left(r,\theta\right)/d_r\right]^{-\eta}\right)drd\theta}_{J_1} - \gamma/\ln\left(2\right)\right]. \qquad \text{(A.6)}
\end{aligned}$$

A.2.2 Upper bound

Applying the same methodology used for the lower bound, the corresponding average upper bound is given by:

$$\begin{aligned}\bar{C}_U^i &= \frac{1}{\pi\left(R_{p_j}^2 - d_r^2\right)} \int_0^{2\pi} \int_{d_r}^{R_{p_j}} r\bar{C}_U\left(\tilde{d}_i\left(r,\theta\right)\right) dr d\theta \\ &= \frac{B}{\pi\left(R_{p_j}^2 - d_r^2\right)} \times \underbrace{\int_0^{2\pi} \int_{d_r}^{R_{p_j}} r\log_2\left(1 + \rho_i K\left[\tilde{d}_i\left(r,\theta\right)/d_r\right]^{-\eta}\mu_{\psi_i}\right) dr d\theta}_{J_2}. \end{aligned} \tag{A.7}$$

The function $\tilde{d}_i\left(r,\theta\right)$ is given by:

$$\tilde{d}_i\left(r,\theta\right) = \begin{cases} \sqrt{\left[r\cos\left(\theta\right) - x^*\right]^2 + \left[r\sin\left(\theta\right) - y^*\right]^2} & , \ i = m, \\ r & , \ i = p_j, \end{cases} \tag{A.8}$$

where $x^* = x_m - x_{p_j}$ and $y^* = y_m - y_{p_j}$.

This function represents the user distance from the macrocell and the picocell BS in polar coordinates, when the user is located within the picocell. Finally, the J_1 and J_2 integrals are numerically computed.

Appendix B

B.1 Proof of Lemma 3

We define the function $h(\mathbf{x}) : \{0,1\}^{|\mathcal{N}|} \to \mathbb{R}$ expressed as

$$h(\mathbf{x}) = g\left(\mathbf{a}^T \cdot \mathbf{x}\right) = g\left(\sum_{n=1}^{|\mathcal{N}|} a_n x_n\right), \tag{B.1}$$

where $g(\cdot)$ is monotone and concave, $\mathbf{a} = \left[a_1, \cdots, a_{|\mathcal{N}|}\right]^{\mathrm{T}}$ and $a_n \in \mathbb{R}^+$. The above function is equivalent to the set function $h(\mathcal{S}) : 2^{\mathcal{N}} \to \mathbb{R}$, where $h(\mathcal{S}) = g\left(\sum_{n\in\mathcal{S}} a_n\right)$ and $\mathcal{S} = \{n \in \mathcal{N} : x_n = 1\}$.

Definition 1. *A set function is characterized as monotone iff $h(\mathcal{Q}) \geq h(\mathcal{P})$ for every $\mathcal{P} \subseteq \mathcal{Q} \subseteq \mathcal{N}$.*

Following holds with inequality for every $\mathcal{P} \subseteq \mathcal{Q} \subseteq \mathcal{N}$:

$$\sum_{n\in\mathcal{Q}} a_n \geq \sum_{n\in\mathcal{P}} a_n \Leftrightarrow g\left(\sum_{n\in\mathcal{Q}} a_n\right) \geq g\left(\sum_{n\in\mathcal{P}} a_n\right), \tag{B.2}$$

since g is monotone proving the monotonicity of h.

Proposition 1. *If $a_n \in \mathbb{R}^+$ and $g(\cdot)$ is concave, then the function $h(\mathcal{S}) = g\left(\sum_{n\in\mathcal{S}} a_n\right)$ is submodular.*

Proof. The proof is based on part (a) of Lemma 2.6.2. in [99]. □

Applying Proposition 1, we prove the submodularity of $h(\cdot)$.

The objective function in Eq. (6.3) can be rewritten as the set function $f(\mathcal{V}) : 2^{\mathcal{V}} \to \mathbb{R}$ that is expressed as

$$f(\mathcal{V}) = \sum_{u_i\in\mathcal{U}} \Phi\left(\sum_{b_j\in\mathcal{B}} \sum_{k\in\mathcal{K}} R^{\mathrm{D}}_{j,k,i} \mathbf{1}_{\mathcal{W}}\left(x^{\mathrm{D}}_{j,k,i}\right)\right), \tag{B.3}$$

where $\mathcal{V} = \mathcal{B} \times \mathcal{K} \times \mathcal{U}$, $\mathbf{1}_{\mathcal{W}}(\cdot)$ is the indicator function on the set $\mathcal{W} = \{(j,k,i) \in \mathcal{V} : x^{\mathrm{D}}_{j,k,i} = 1\}$.

Proposition 2. $\Phi(\cdot)$ *utility functions are monotone and strictly concave.*

Proof. The PF utility function is the logarithmic function and thus is monotone and strictly concave. While the UPF takes the natural logarithm of sigmoid function proven as monotone and strictly concave following Lemma III.1. in [89]. □

As each term in the summation of objective function $f(\cdot)$ lying on the family of $h(\cdot)$ in Eq. (B.1), where $\Phi(\cdot) \equiv g(\cdot)$ is monotone and concave, $\mathcal{N} \equiv \mathcal{B} \times \mathcal{K}$, $a_n \equiv R^{\mathrm{D}}_{j,k,i}$ and $x_n \equiv x^{\mathrm{D}}_{j,k,i}$. Thus, we conclude that the objective function $f(\cdot)$ is monotone and submodular as it is the summation of monotone and submodular functions.

Finally, we define three constraints in Eq. (6.4) equivalently as $\mathcal{I}_l,\ l \in \{1,2,3\}$ in Eq. (B.4) as supersets:

$$\mathcal{I}_1 = \{S \subseteq 2^{\mathcal{V}} : |S \cap 2^{\{j,\mathcal{K},\mathcal{U}\}}| \leq B_j, \forall\ b_j \in \mathcal{B}\}, \tag{B.4a}$$

$$\mathcal{I}_2 = \{S \subseteq 2^{\mathcal{V}} : |S \cap 2^{\{\mathcal{B},\mathcal{K},i\}}| \leq M_i, \forall\ u_i \in \mathcal{U}\}, \tag{B.4b}$$

$$\mathcal{I}_3 = \{S \subseteq 2^{\mathcal{V}} : |S \cap 2^{\{j,k,\mathcal{U}\}}| \leq 1, \forall\ b_j \in \mathcal{B},\ k \in \mathcal{K}\}. \tag{B.4c}$$

To prove these constraints are matroid (see definition in [97]) with $\mathcal{V}$ as ground set, two properties shall be fulfilled:

(a) For all sets $\mathcal{P}$ and $\mathcal{Q}$ with $\mathcal{P} \subseteq \mathcal{Q} \subseteq \mathcal{V}$, if $\mathcal{Q} \in \mathcal{I}_l$, i.e., the sub-channel allocation defined by $\mathcal{Q}$ does not violate the constraint $\mathcal{I}_l$, then $\mathcal{P} \in \mathcal{I}_l$ is held, i.e., the constraint $\mathcal{I}_l$ is not violated by sub-channel allocation in $\mathcal{P}$, since $\mathcal{P} \subseteq \mathcal{Q}$. All three aforementioned constraints possess such property.

(b) For all sets $\mathcal{P}, \mathcal{Q} \in \mathcal{I}_l$ and $|\mathcal{Q}| > |\mathcal{P}|$, i.e., more sub-channels are allocated in total defined by $\mathcal{Q}$, $\exists\ e \in \mathcal{Q} \backslash \mathcal{P}$ such that $\mathcal{P} \cup \{e\} \in \mathcal{I}_l$. Since not all sub-channels are allocated defined by $\mathcal{P}$ (otherwise, $\mathcal{Q}$ will either violate the constraint $\mathcal{I}_l$, i.e., $\mathcal{Q} \notin \mathcal{I}_l$, or has no more allocated sub-channels as $\mathcal{P}$, i.e., $|\mathcal{Q}| \leq |\mathcal{P}|$), there is at least one sub-channel that can be further allocated and will not violate the constraint $\mathcal{I}_l$. Hence, such property is held for all three constraints.

Thus, we conclude that $\mathcal{I}_l,\ \forall\ l \in \{1,2,3\}$ are matroid.

Appendix C

Towards open-source 4G/5G experimental platforms [100], we are advancing to the OAI testbed implementation for local or remote access experimentation. The deployed OAI testbed was used to perform network measurements in order to characterize and analyze the X2 handover system behavior described in Chapter 2. The supported OAI testbed's hardware (HW) equipment was further exploited to extend the X2 handover implementation based on the OAI emulator (oaisim) and proceed to its RF-based release for real-time communications prototyping.

C.1 OAI testbed for X2 handover experimentation

In that section, we describe the OAI testbed including the supported nodes, the provided web portal and the offered services. The whole testbed infrastructure is a solution for accessing, managing and monitoring the OAI network resources in a real-time running environment.

C.1.1 OAI testbed description

The remote and local access to the OAI testbed facilities is given via the OAI testbed web portal. The NITOS Scheduler [101], a tool that is responsible for managing testbed resources, i.e., including nodes description, reservation and verification per time slot for a given experiment, was adapted to support the OAI testbed LTE network resources. A detailed description of the OAI testbed portal can be found at `https://oailab.eurecom.fr/oai-testbed/`, as well as analytical guidelines on how to access and register to it. The web portal's user interface is depicted in Fig. C.1 and the provided OAI testbed LTE nodes with their supported features are shown in Fig. C.2.

The list of nodes is provided as follows:

Node 1: OAI Juju charms, COTS UE (Samsung Galaxy S6), **Node 2:** OAI eNB (USRP B210), **Node 3:** LTErf service, **Node 4:** OAI eNB (USRP B210), **Node 5:** COTS UE (Huawei e392), OAI HSS, Docker service, **Node 6:** COTS UE (Huawei e3276), OAI HSS, Docker service, **Node 7:** OAI EPC (MME/SPGW), **Node 8:** OAI EPC (MME/SPGW).

In more detail, except for the OAI compatible hardware (2 USRP B210) and software (OAI eNB/EPC), we provide a gamut of services, e.g., OAI Juju charms [102], LTErf and Docker. Moreover, the OAI testbed is fully equipped with 3 COTS UEs (Huawei e392/ Huawei e3276/Sam-

Figure C.1: OAI testbed portal

sung Galaxy S6) and suitable RF frontend equipment, i.e., antennae, filters, etc. Tutorials for the experimenters can be found in the following links:

- OAI testbed guidelines:
  `https://oailab.eurecom.fr/oai-testbed/index.php/oai-nodes/lte-testbed`
- LTErf guidelines for OAI:
  `https://oailab.eurecom.fr/oai-testbed/index.php/oai-nodes/lterf-guidelines`

The provided links include a step by step analytical guide on how an experimenter can access the OAI testbed after the registration process. As a second step, we provide the commands that can be used to manage and run the OAI LTE network equipment. Last but not least, detailed instructions for the use of the LTErf service via REST API commands are incorporated to aid the OAI testbed's LTE network parameterization during the phase of the experiments.

C.2 OAI X2 experimental setup

The initial implementation of X2 handover was based on the oaisim without including the RF support using compatible software/hardware equipment (OAI RF lte-softmodem) as discussed in Chapter 2. Currently, the X2 HO implementation is extended to support the RF functionality [103] as well as it is integrated with the OAI testbed.

The E-UTRAN supported messages and functionalities for the X2 HO procedure on the eNB side are described as follows:

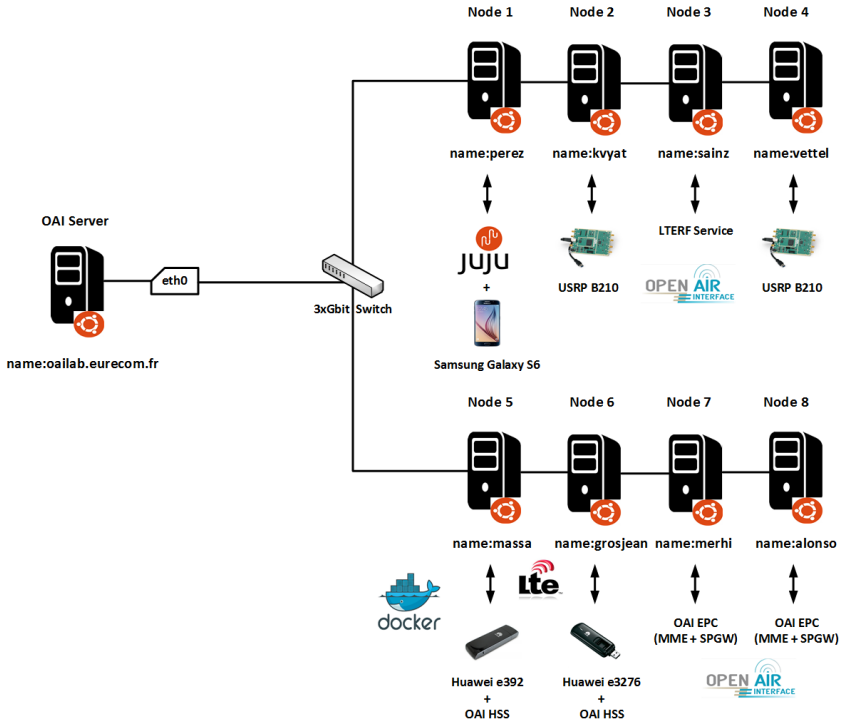


Figure C.2: OAI testbed nodes setup

PHY/MAC layer API: RAR, **RRC layer API:** RRC Measurement report, RRC Connection Reconfiguration, RRC Connection Reconfiguration Complete, **X2 API:** X2 Setup Request, X2 Setup Response, X2 Handover Request, X2 Handover Request ACK.

The experimentation setup is depicted in Fig. C.3 and includes 2 OAI eNBs (2 USRP B210-source/target eNB) synchronised via GPS. In addition, octoBoxquadAtten programmable modules are used as attenuators on both the two eNBs to model UE mobility, e.g., emulate fading effects etc. At that instant, UE can send its E-UTRAN RRC Measurement report for triggering the HO in the eNB side. Specifically, the selected attenuators offer stable attenuation accuracy via wide choice of attenuation (63 dB range), frequency ranges (DC to 6000 MHz) and step sizes (0.5 step). In addition, the module can be controlled remotely through Ethernet interface via an attenuation control unit. For the core network side, a COTS EPC, i.e., Nokia Bell-labs, is used to support the X2 HO functionalities for EPC, e.g., Path switch/Modify Bearer signaling, not currently supported by OAI; this part is included in the HO Execution process (see Fig. 2.2 in Chapter 2).

C.3 OAI X2 handover demonstration with OAI emulator (oaisim)

A brief description of the X2 HO intra-frequency oaisim deployment presented in Chapter 2 and supported in OAI X2 experimental testbed is discussed as follows. The demonstration includes 2 eNBs (one source and one target single omnidirectional antenna cell) and 1 UE (single omnidirectional antenna user). In addition, one radio chain (Tx/Rx) per UE/eNB wireless single-input single-output (SISO) link in full PHY emulation mode (without radio frontends)

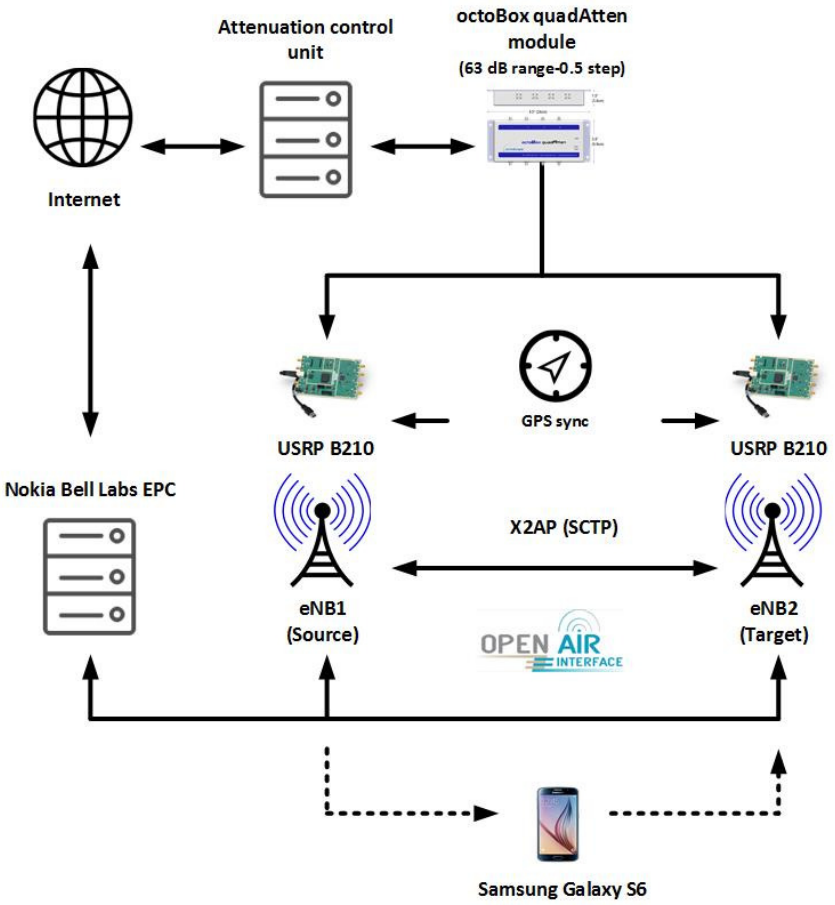


Figure C.3: X2 HO experimental setup schematic

is implemented. UE is moving based on a specific mobility pattern (i.e., ping-pong movement pattern) defined by traces that are given as input. The cells are in fixed topology, i.e., placed in a constant position. Before handover, UDP traffic [1] is exchanged between the UE and the source eNB (eNB0). Handover procedure is triggered and after handover execution and completion, UDP traffic is exchanged between the UE and the target eNB (eNB1). A youtube video link of the demonstration can be found at: `https://www.youtube.com/watch?v=PiQxLgY3Gsk`.

C.4 OAI X2 handover experimental RF testbed demonstration

As an extension of X2 handover implementation in oaisim, we perform real-world RF measurements in the deployed OAI X2 HO experimental testbed. The handover testbed provides small-cell to small-cell X2 intra-frequency handover functionalities, as described in [57]. The demonstration includes 2 OAI eNBs (i.e., 2x USRP B210) and 1 COTS UE (Samsung Galaxy S6). The two eNBs are synchronised via a GPS module as mentioned before and are connected via Ethernet or locally, where SCTP is enabled for communication, i.e., exchanging the required X2

[1] Additional measurements can be performed using TCP traffic to figure out the impact on HO delay increment due to packets retransmission.

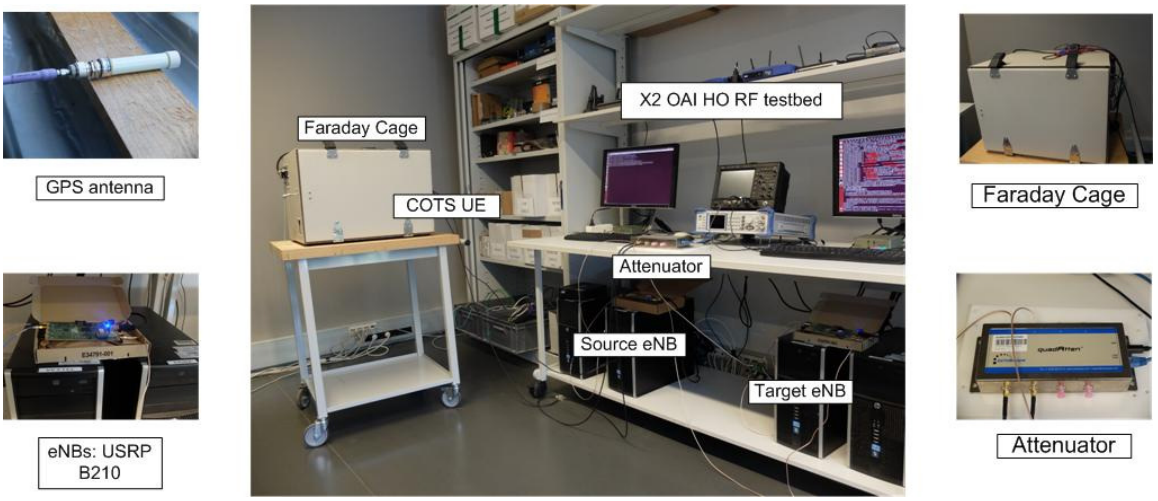


Figure C.4: X2 HO experimental setup in the lab

signalling, and GTP-U for the user data-plane. This is an initial step for testing the E-UTRAN part including the HO preparation and HO execution steps of X2 HO. For the EPC part; an OAI compatible (Nokia Bell Labs EPC) is used as descripted in the previous sections. A youtube video link of the demonstration can be found at: `https://www.youtube.com/watch?v=KlwampmhFxw`.

C.5 SDN-based handover control

This section describes a Software Defined Network (SDN)-based control and coordination framework used for the handover parameters control that it will be released soon for future experimentation. In more detail, the supported handover parameters that can be adjusted using the proposed framework are mentioned in Section 2.2.2.

FlexRAN controller/agent: Towards 5G, programmable RAN underlay is essential in order to facilitate the control and managements procedures. In the LTE domain, SDN control can be used to facilitate network agility over the physical or virtualized core network and the base stations. On the top of the current testbed RAN deployment for X2 handover experimentation, an SDN controller is built to provide centrally handover programmability. For the programmability of the eNB, emerging control frameworks such as FlexRAN with support for both northbound and southbound protocols can be used, whose design and implementation can be found in [27]. A high level representation of the proposed architecture that allows to change the above parameters on-the-fly is depicted in Fig. C.5. Additional information about FlexRAN programmability and setup can be found at: `https://gitlab.eurecom.fr/FlexRAN/FlexRAN/wikis/FlexRANwithCOTSUE`.

Specifically, FlexRAN is composed from two basic entities: a) the FlexRAN controller and b) the FlexRAN agent. The boundary between the network controller and agent is traversed by the RAN-specific southbound application programming interface (API), the so-called FlexRAN agent API. The boundary between the applications and the FlexRAN controller layer is traversed by the northbound API, the so-called FlexRAN application API.

- **FlexRAN agent:** FlexRAN separates the control from the data plane by detaching the

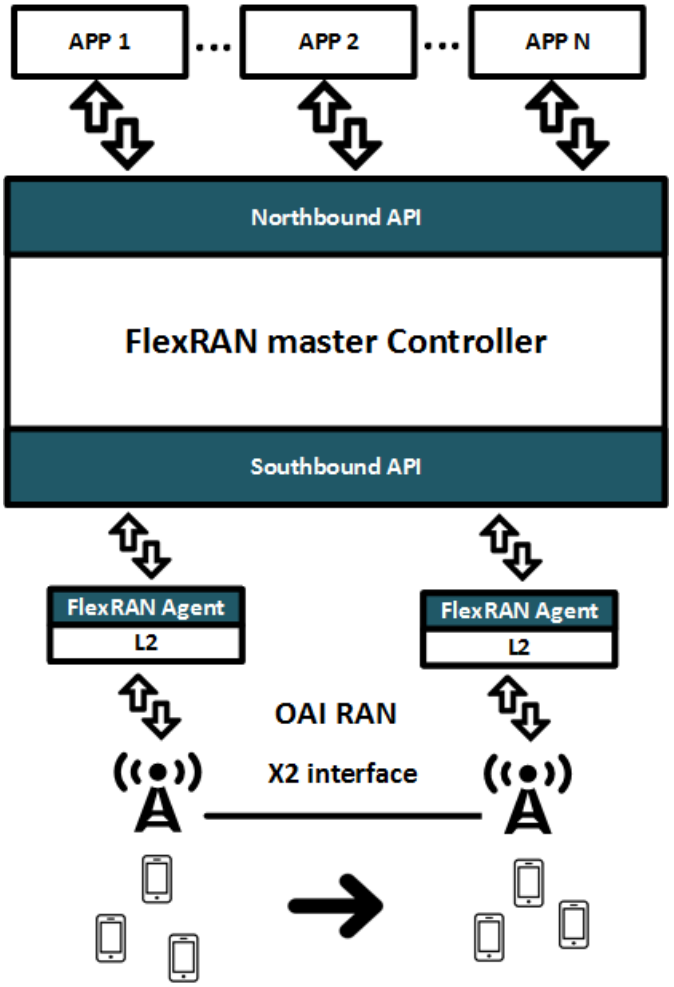


Figure C.5: FlexRAN scenario

L2 protocol logic from the relevant actions and by consolidating all the related control operations in a logically centralized controller, which in FlexRAN comprises of the FlexRAN master controller and the agents interacting via the FlexRAN protocol. In other words, the FlexRAN Agent API constitutes the southbound API. The protocol messages are implemented using Google protobuf messages; the current implementation supports the RLC/MAC protocol logic and RRC logic is ongoing development related to the handover parameters control.

- **FlexRAN controller:** The master controller constitutes the brain of the FlexRAN control plane as it manages the operation of the FlexRAN agent. In more detail, the applications running on top of the FlexRAN master controller make control decisions for the various functions of the L2. These decisions are manifested in the form of FlexRAN protocol messages that are sent from the master controller to the corresponding FlexRAN agent(s). The agents provide a protocol message handler and dispatcher module, that is responsible to process these messages and to forward them to the L2 control module

Table C.1: ho.yaml file content

```
  rrc:
  -x2_ho:
  parameters:
         ttt_ms:     :!!int 100
            hys:     :!!int 0
            ofn:     :!!int 0
            ocn:     :!!int 0
            ofs:     :!!int 0
            ocs:     :!!int 0
            off:     :!!int 0
filter_coeff_rsrp:  :!!float 0.5
filter_coeff_rsrq:  :!!float 0.5
```

by invoking the proper function of the L2 API. As northbound API, we use the "curl" command-line tool to send by an application via URLs the intended parameter configuration.

If an experimenter wants to set/get the handover parameters in order to see their impact in the network can send a curl command using the northbound API (application-controller) and invoke the corresponding L2 (RRC layer) getters/setters using the southbound API (controller-agent). An example of the corresponding commands is given in the following:

Set handover parameters:

```
$ curl -X POST http://localhost:9999/ho_manager/ho.yaml
```

Get handover parameters:

```
$ curl -X GET http://localhost:9999/stats_manager/rrc
```

```
$ curl -X GET http://localhost:9999/stats_manager/all
```

The above example shows how the experimenter can set or get the handover parameters values via curl commands invoking the corresponding setters/getters function in L2 layer. Specifically, the experimenter sets the values in the corresponding .yaml files (like the one depicted in Table C.1) or gets the values using the stats_manager (returns rrc or all L2 layers parameters).

The centralized network view offered by FlexRAN could enable more sophisticated mobility management mechanisms that consider additional factors, e.g., the load of cells or use-driven resource requirements of mobile devices, leading to an optimization of the device-network associations [27]. That kind of sophistication can be devised by the running applications on the top described in the following. The application layer consists of all the controlling and business applications running on top of the FlexRAN controller. One of these applications can be considered mobility management based on different handover criteria. A well-known handover criterion supported by OAI RAN platform, commonly used in conventional HO decision algorithms (3GPP LTE), is based on RSRPs comparison method in which hysteresis and handover offsets are included as described in Chapter 2. As an extension, other representative handover algorithms consider not only RSS criteria as the one described in Chapter 3 are planned for future development.

Bibliography

[1] Cisco. Cisco visual networking index forecast projects 18-fold growth in global mobile internet data traffic from 2011 to 2016. [Online]. Available: http://tinyurl.com/7gn9x9s

[2] Ericsson. 5G radio access: Research and vision. [Online]. Available: http://www.ericsson.com/res/docs/whitepapers/wp-5g.pdf

[3] 5G-PPP–5G Vision. [Online]. Available: https://5g-ppp.eu/wp-content/uploads/2015/02/5G-Vision-Brochure-v1.pdf

[4] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. D. Schotten, T. Hidekazu, H. M. Tullberg, M. A. Uusitalo, B. Timus, and M. Fallgren, "Scenarios for 5g mobile and wireless communications: the vision of the METIS project," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 26–35, 2014.

[5] Small Cell Forum–Release 10.0. [Online]. Available: http://scf.io/en/documents/new_documents.php

[6] D. Xenakis, N. Passas, L. Merakos, and C. Verikoukis, "Mobility management for femtocells in LTE-advanced: Key aspects and survey of handover decision algorithms," *IEEE Communications Surveys Tutorials*, 2014.

[7] Nokia, "Ultra Dense Network (udn)," White Paper, 2015.

[8] T. Jansen, I. Balan, J. Turk, I. Moerman, and T. Kürner, "Handover parameter optimization in LTE self-organizing networks," in *IEEE 72nd Vehicular Technology Conference Fall (VTC 2010-Fall)*, 2010.

[9] N. Sinclair, D. Harle, I. Glover, J. Irvine, and R. Atkinson, "An advanced som algorithm applied to handover management within LTE," *IEEE Transactions on Vehicular Technology*, 2013.

[10] H. Zhang, W. Ma, W. Li, W. Zheng, X. Wen, and C. Jiang, "Signalling cost evaluation of handover management schemes in LTE-advanced femtocell," in *IEEE 73rd Vehicular Technology Conference (VTC Spring)*, 2011.

[11] D. Lopez-Perez, A. Ladanyi, A. Jüttner, and J. Zhang, "Ofdma femtocells: Intracell handover for interference and handover mitigation in two-tier networks," in *2010 IEEE Wireless Communications and Networking Conference (WCNC)*, 2010.

[12] A. Gudipati, D. Perry, L. E. Li, and S. Katti, "Softran: Software defined radio access network," in *Proceedings of the Second ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking*, 2013, pp. 25–30.

[13] 5G PPP Architecture Working Group, "View on 5G architecture," 2016.

[14] 3GPP, "Study on Small Cell enhancements for E-UTRA and E-UTRAN; Higher layer aspects," 3rd Generation Partnership Project (3GPP), TR 36.842, Dec 2013.

[15] ——, "Evolved Universal Terrestrial Radio Access Network (eutran);carrier aggregation; base station (bs) radio transmission and reception (release 10)," 3rd Generation Partnership Project (3GPP), TR 36.808, Jul 2013.

[16] D. S. Michalopoulos, I. Viering, and L. Du, "User-plane multi-connectivity aspects in 5G," in *23rd International Conference on Telecommunications (ICT)*, May 2016, pp. 1–5.

[17] C. Rosa, K. Pedersen, H. Wang, P. H. Michaelsen, S. Barbera, E. Malkamaki, T. Henttonen, and B. Sebire, "Dual connectivity for lte small cell evolution: functionality and performance aspects," *IEEE Communications Magazine*, vol. 54, no. 6, pp. 137–143, June 2016.

[18] M. Polese, M. Mezzavilla, and M. Zorzi, "Performance comparison of dual connectivity and hard handover for lte-5g tight integration," in *Proceedings of the 9th EAI International Conference on Simulation Tools and Techniques*, ser. SIMUTOOLS'16, 2016.

[19] Ericsson, "5G radio access," White Paper, 2014.

[20] W. H. Wang, M. Palaniswami, and S. H. Low, "Application-oriented flow control: fundamentals, algorithms and fairness," *IEEE/ACM Transactions on Networking*, vol. 14, no. 6, pp. 1282–1291, 2006.

[21] G. Tychogiorgos, A. Gkelias, and K. K. Leung, "Utility-proportional fairness in wireless networks," in *IEEE PIMRC 2012*, pp. 839–844.

[22] J.-W. Lee and J.-A. Kwon, "Utility-based power allocation for multiclass wireless systems," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 7, pp. 3813–3819, 2009.

[23] M. Proebster, M. Kaschub, and S. Valentin, "Context-aware resource allocation to improve the quality of service of heterogeneous traffic," in *2011 IEEE International Conference on Communications (ICC)*, 2011, pp. 1–6.

[24] F. B. Tesema, A. Awada, I. Viering, M. Simsek, and G. P. Fettweis, "Mobility modeling and performance evaluation of multi-connectivity in 5G intra-frequency networks," in *IEEE Globecom Workshops*, 2015.

[25] N. Prasad and S. Rangarajan, "Exploiting dual connectivity in heterogeneous cellular networks," in *2017 15th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, May 2017, pp. 1–8.

[26] F. Cardoso, F. Pereira de Figueiredo, R. Vilela, and J. Paulo Miranda, "A case study on protocol stack integration for 3gpp lte evolved node b," Nov 2014, pp. 1–6.

[27] X. Foukas, N. Nikaein, M. M. Kassem, M. K. Marina, and K. Kontovasilis, "Flexran: A flexible and programmable platform for software-defined radio access networks," in *Proceedings of the 12th International on Conference on Emerging Networking EXperiments and Technologies*, ser. CoNEXT '16, 2016, pp. 427–441.

[28] S. Oh, B. Ryu, and Y. Shin, "Epc signaling load impact over s1 and x2 handover on lte-advanced system," in *2013 Third World Congress on Information and Communication Technologies (WICT)*, Dec 2013, pp. 183–188.

[29] S. Oh, H. Kim, B. Ryu, and N. Park, "Inbound mobility management on lte-advanced femtocell topology using x2 interface," in *Proceedings of 20th International Conference on Computer Communications and Networks (ICCCN)*, July 2011, pp. 1–5.

[30] E. A. Ibrahim, M. R. M. Rizk, and E. F. Badran, "Study of lte-r x2 handover based on a3 event algorithm using matlab," in *2015 International Conference on Information and Communication Technology Convergence (ICTC)*, Oct 2015, pp. 1155–1159.

[31] N. Baldo, M. Requena-Esteso, M. Miozzo, and R. Kwan, "An open source model for the simulation of lte handover scenarios and algorithms in ns-3," in *Proceedings of the 16th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, 2013, pp. 289–298.

[32] M. Assyadzily, A. Suhartomo, and A. Silitonga, "Evaluation of x2-handover performance based on rsrp measurement with friis path loss using network simulator version 3 (ns-3)," in *Information and Communication Technology (ICoICT), 2014 2nd International Conference on*, May 2014, pp. 436–441.

[33] Atte Helenius, "Performance of handover in long term evolution," Master Thesis, Aalto University, 2008.

[34] J.-M. Moon and D.-H. Cho, "Novel handoff decision algorithm in hierarchical macro/femtocell networks," in *2010 IEEE Wireless Communications and Networking Conference (WCNC)*, 2010.

[35] L. Li, Z. Mao, and J. Rexford, "Toward software-defined cellular networks," in *2012 European Workshop on Software Defined Networking (EWSDN)*, Oct 2012, pp. 7–12.

[36] X. Jin, L. E. Li, L. Vanbever, and J. Rexford, "Softcell: Scalable and flexible cellular core network architecture," in *Proceedings of the 9th ACM Conference on Emerging Networking Experiments and Technologies*, 2013, pp. 163–174.

[37] T. Mahmoodi and S. Seetharaman, "On using a sdn-based control plane in 5G mobile networks," *IEEE Vehicular Technology Magazine*, vol. 9, no. 3, 2014.

[38] P. Dely, A. Kassler, L. Chow, N. Bambos, N. Bayer, H. Einsiedler, C. Peylo, D. Mellado, and M. Sanchez, "A software-defined networking approach for handover management with real-time video in wlans," *Journal of Modern Transportation*, vol. 21, no. 1, pp. 58–65, 2013.

[39] K.-K. Yap, T.-Y. Huang, M. Kobayashi, M. Chan, R. Sherwood, G. Parulkar, and N. McKeown, "Lossless handover with n-casting between wifi-wimax on openroads," Mobicom '09, Sep. 2009.

[40] F. Capozzi, G. Piro, L. Grieco, G. Boggia, and P. Camarda, "Downlink packet scheduling in lte cellular networks: Key design issues and a survey," *IEEE Communication Surveys and Tutorials*, 2013.

[41] H. S. Dhillon, M. Kountouris, and J. G. Andrews, "Downlink MIMO hetnets: Modeling, ordering results and performance analysis," *IEEE Transactions on Wireless Communications*, 2013.

[42] A. Huang, N. Nikaein, T. Stenbock, A. Ksentini, and C. Bonnet, "Low latency MEC framework for SDN-based LTE/LTE-A networks," in *IEEE International Conference on Communications (ICC)*, May 2017.

[43] C.-Y. Chang, K. Alexandris, N. Nikaein, K. Katsalis, and T. Spyropoulos, "Mec architectural implications for lte/lte-a networks," in *Proceedings of the Workshop on Mobility in the Evolving Internet Architecture*, ser. MobiArch '16, 2016, pp. 13–18.

[44] S. Tombaz, P. Monti, F. Farias, M. Fiorani, L. Wosinska, and J. Zander, "Is backhaul becoming a bottleneck for green wireless access networks?" in *IEEE International Conference on Communications (ICC)*, 2014, pp. 4029–4035.

[45] N. Sapountzis, T. Spyropoulos, N. Nikaein, and U. Salim, "Optimal downlink and uplink user association in backhaul-limited hetnets," in *the 35th Annual IEEE International Conference on Computer Communications (INFOCOM)*, April 2016, pp. 1–9.

[46] D. W. K. Ng, E. S. Lo, and R. Schober, "Energy-efficient resource allocation in multi-cell ofdma systems with limited backhaul capacity," *IEEE Transactions on Wireless Communications*, vol. 11, no. 10, pp. 3618–3631, Oct 2012.

[47] S. A. Ahmad and D. Datla, "Distributed power allocations in heterogeneous networks with dual connectivity using backhaul state information," *IEEE Trans. on Wireless Communications*, vol. 14, no. 8, pp. 4574–4581, Aug 2015.

[48] O. Somekh, O. Simeone, A. Sanderovich, B. M. Zaidel, and S. Shamai, "On the impact of limited-capacity backhaul and inter-users links in cooperative multicell networks," in *42nd Annual Conference on Information Sciences and Systems*, March 2008, pp. 776–780.

[49] F. Pantisano, M. Bennis, W. Saad, and M. Debbah, "Cache-aware user association in backhaul-constrained small cell networks," in *12th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, 2014, pp. 37–42.

[50] H. Beyranvand, W. Lim, M. Maier, C. Verikoukis, and J. A. Salehi, "Backhaul-aware user association in fiwi enhanced lte-a heterogeneous networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 2992–3003, 2015.

[51] M. Jaber, M. Imran, R. Tafazolli, and A. Tukmanov, "An adaptive backhaul-aware cell range extension approach," in *IEEE International Conference on Communication Workshop (ICCW)*, 2015, pp. 74–79.

[52] A. Chowdhery, W. Yu, and J. M. Cioffi, "Cooperative wireless multicell ofdma network with backhaul capacity constraints," in *IEEE International Conference on Communications (ICC)*, 2011, pp. 1–6.

[53] A. Asadi and V. Mancuso, "A survey on opportunistic scheduling in wireless communications," *IEEE Communications Surveys Tutorials*, vol. 15, no. 4, pp. 1671–1688, April 2013.

[54] M. Andrews and L. Zhang, "Scheduling algorithms for multicarrier wireless data systems," *IEEE/ACM Transactions on Networking*, vol. 19, no. 2, pp. 447–455, April 2011.

[55] S. Shakkottai, "Effective capacity and qos for wireless scheduling," *IEEE Transactions on Automatic Control*, vol. 53, no. 3, pp. 749–761, April 2008.

[56] Openairinterface software alliance. [Online]. Available: http://www.openairinterface.org/

[57] 3GPP. 3rd Generation Partnership Project; Evolved Universal Terrestrial Radio; X2 Application Protocol (X2AP) specification (Release 10); Technical Specification TS 36.321 v10.7.0. [Online]. Available: http://www.3gpp.org

[58] S. Sesia, I. Toufik, and M. Baker, *LTE, The UMTS Long Term Evolution: From Theory to Practice*. Wiley Publishing, 2009.

[59] 3GPP. 3rd Generation Partnership Project; Technical Specification Group Radio Acess Network; Radio Resource Control (RRC); Protocol specification (Release 10); Technical Specification TS 36.331 v10.6.0. [Online]. Available: http://www.3gpp.org

[60] ——. 3rd Generation Partnership Project; Technical Specification Group Radio Acess Network; Medium Acess Control (MAC); Protocol specification (Release 10); Technical Specification TS 36.321 v10.6.0. [Online]. Available: http://www.3gpp.org

[61] K. Alexandris, N. Sapountzis, N. Nikaein, and T. Spyropoulos, "Load-aware handover decision algorithm in next-generation HetNets," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2016.

[62] 3GPP TR 25.912, "Feasibility Study for Evolved Universal Terrestrial Radio Access (UTRA) and Universal Terrestrial Radio Access Network (UTRAN), Version 12.0.0," Tech. Rep., 2014.

[63] C.-Y. Chang, R. Schiavi, N. Nikaein, T. Spyropoulos, and C. Bonnet, "Impact of packetization and functional split on C-RAN fronthaul performance," in *IEEE International Conference on Communications (ICC)*, May 2016.

[64] OAI X2 API oaisim. [Online]. Available: https://gitlab.eurecom.fr/oai/openairinterface5g/tree/feature-39-X2-handover

[65] F. Kaltenberger, I. Latif, and R. Knopp, "On scalability, robustness and accuracy of physical layer abstraction for large-scale system-level evaluations of LTE networks," in *Asilomar Conference on Signals, Systems, and Computers, Pacific Grove (ASILOMAR)*, 2013.

[66] N. Nikaein, R. Knopp, F. Kaltenberger, L. Gauthier, C. Bonnet, D. Nussbaum, and R. Ghaddab, "Demo: Openairinterface: An open lte network in a pc," in *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*, 2014, pp. 305–308.

[67] Network signal guru. [Online]. Available: https://play.google.com/store/apps/details?id=com.qtrun.QuickTest&hl=fr

[68] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, and J. Zhang, "What will 5g be?" *IEEE Journal on Selected Areas in Communications*, 2014.

[69] A. Lobinger, S. Stefanski, T. Jansen, and I. Balan, "Load balancing in downlink LTE self-optimizing networks," in *IEEE 71st Vehicular Technology Conference (VTC 2010-Spring)*, 2010.

[70] C. Fischione, G. Athanasiou, and F. Santucci, "Dynamic optimization of generalized least squares handover algorithms," *IEEE Transactions on Wireless Communications*, 2014.

[71] P. Rost, C. Bernardos, A. Domenico, M. Girolamo, M. Lalam, A. Maeder, D. Sabella, and D. Wübben, "Cloud technologies for flexible 5g radio access networks," *IEEE Communications Magazine*, 2014.

[72] M. Harchol-Balter, *Performance Modeling and Design of Computer Systems.* Cambridge University Press, 2010.

[73] G. Pollini, "Trends in handover design," *IEEE Communications Magazine*, 1996.

[74] N. Sapountzis, S. Sarantidis, T. Spyropoulos, N. Nikaein, and U. Salim, "Reducing the energy consumption of small cell networks subject to QoE constraints," in *Proceedings of the Global Communications Conference, GLOBECOM 2014, Texas, USA*, 2014.

[75] NGMN, "NGMN 5G white paper," 2015.

[76] E. Dahlman, G. Mildh, S. Parkvall, J. Peisa, J. Sachs, Y. Selén, and J. Sköld, "5G wireless access: requirements and realization," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 42–47, 2014.

[77] A. Prasad, F. S. Moya, M. Ericson, R. Fantini, and O. Bulakci, "Enabling ran moderation and dynamic traffic steering in 5G," *IEEE VTC-Fall*, 2016.

[78] C. L. I, S. Han, Z. Xu, S. Wang, Q. Sun, and Y. Chen, "New paradigm of 5G wireless Internet," *IEEE Journal on Selected Areas in Communications*, 2016.

[79] Nokia, "5G Masterplan - Five keys to create the new communications era," White Paper, 2016.

[80] Huawei, "5G network architecture a high-level view," White Paper, 2016.

[81] S. Chandrashekar, A. Maeder, C. Sartori, T. Höhne, B. Vejlgaard, and D. Chandramouli, "5G multi-RAT multi-connectivity architecture," in *IEEE ICC workshops*, 2016.

[82] J. Deng, O. Tirkkonen, R. Freij-Hollanti, T. Chen, and N. Nikaein, "Resource allocation and interference management for opportunistic relaying in integrated mmwave/sub-6 GHz 5G networks," *IEEE Communications Magazine*, vol. 55, no. 6, pp. 94–101, 2017.

[83] S. Singh, M. Geraseminko, S. p. Yeh, N. Himayat, and S. Talwar, "Proportional fair traffic splitting and aggregation in heterogeneous wireless networks," *IEEE Communications Letters*, vol. 20, no. 5, pp. 1010–1013, May 2016.

[84] 3GPP, "Feasibility study for proximity services," TR 22.803, 2013.

[85] R. Favraud, A. Apostolaras, N. Nikaein, and T. Korakis, "Towards moving public safety networks," *IEEE Communications Magazine*, vol. 54, no. 3, March 2016.

[86] A. Laya, K. Wang, A. A. Widaa, J. Alonso-Zarate, J. Markendahl, and L. Alonso, "Device-to-device communications and small cells: enabling spectrum reuse for dense networks," *IEEE Wireless Communications*, vol. 21, no. 4, pp. 98–105, 2014.

[87] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2706–2716, 2013.

[88] S. Shakkottai and R. Srikant, "Network optimization and control," *Foundations and Trends in Networking*, 2008.

[89] A. Abdel-Hadi and C. Clancy, "A utility proportional fairness approach for resource allocation in 4G-LTE," in *IEEE ICNC*, 2014.

[90] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[91] NGMN, "Next generation mobile networks radio access performance evaluation methodology," Tech. Rep., 2008.

[92] M. Shariat, D. M. G. Estévez, A. Vijay, K. Safjan, P. Rugeland, I. da Silva, J. Lorca, J. Widmer, M. Fresia, Y. Li, and I. Siaud, "5G radio access above 6 GHz," *Trans. Emerging Telecommunications Technologies*, vol. 27, no. 9, pp. 1160–1167, 2016.

[93] A. D. L. Oliva, X. C. Perez, A. Azcorra, A. D. Giglio, F. Cavaliere, D. Tiegelbekkers, J. Lessmann, T. Haustein, A. Mourad, and P. Iovanna, "Xhaul: toward an integrated fronthaul/backhaul architecture in 5G networks," *IEEE Wireless Communications*, vol. 22, no. 5, pp. 32–40, 2015.

[94] U. Siddique, H. Tabassum, E. Hossain, and D. I. Kim, "Wireless backhauling of 5G small cells: Challenges and solution approaches," *IEEE Wireless Communications*, vol. 22, no. 5, pp. 22–31, 2015.

[95] Y. Liu, M. Derakhshani, and S. Lambotharan, "Dual Connectivity in Backhaul-limited Massive-MIMO HetNets: User Association and Power Allocation," in *IEEE Globecom 2017*.

[96] O. Grøndalen, A. Zanella, K. Mahmood, M. Carpin, J. Rasool, and O. N. Østerbø, "Scheduling Policies in Time and Frequency Domains for LTE Downlink Channel: A Performance Comparison," *IEEE Trans. on Vehicular Technology*, vol. 66, no. 4, pp. 3345–3360, April 2017.

[97] A. Krause and D. Golovin, *Submodular Function Maximization.* Cambridge University Press, 2014, pp. 71–104.

[98] R. Favraud, N. Nikaein, and C.-Y. Chang, "QoS guarantee in self-backhauled LTE mesh networks," in *IEEE Global Communications Conference (GLOBECOM)*, Dec 2017.

[99] D. Topkis, *Supermodularity and Complementarity*, ser. Frontiers of Economic Research. Princeton University Press, 2011.

[100] K. Pechlivanidou, K. Katsalis, I. Igoumenos, D. Katsaros, T. Korakis, and L. Tassiulas, "NITOS testbed: A cloud based wireless experimentation facility," in *26th International Teletraffic Congress (ITC)*, Sep 2014, pp. 1–6.

[101] NITOS Scheduler. [Online]. Available: https://github.com/NitLab/NITOS-Scheduler

[102] OAI Juju charms. [Online]. Available: https://jujucharms.com/u/navid-nikaein/

[103] OAI X2 API RF. [Online]. Available: https://gitlab.eurecom.fr/oai/openairinterface5g/tree/x2-ka-test