Multihoming in heterogeneous wireless networks

Ghina Dandachi

A JOINT PHD THESIS
Specialty: Telecommunications

Ecole doctorale Informatique, Télécommunications et Electronique de Paris
Ecole doctorale des Sciences et Technologie au Liban

Presented by

DANDACHI Ghina

To obtain the degree of
DOCTOR FROM TELECOM SUDPARIS

Multihoming in Heterogeneous Wireless Networks

Defended on July 21, 2017

Thesis Defense Committee:

M. Mohamad ASSAAD, Professor, CentraleSupelec	Referee
M. Zaher DAWY, Professor, American University of Beirut	Referee
M. Nazim AGLOUMINE, Professor, Université d'Evry Val d'Essone	Examiner
M. Guy PUJOLLE, Professor, Université de Pierre et Marie Curie	Examiner
M.YOUMNI ZIADE, Assistant Professor, Beirut Arab University	Examiner
Ms. Nada CHENDEB, Associate Professor, Lebanese University	Supervisor
M. Salaheddine ELAYOUBI, Senior radio expert, Orange Labs	Supervisor
M. Ziad FAWAL, Professor, Lebanese University	Thesis co-director
M. Tijani CHAHED, Professor, SAMOVAR, Telecom SudParis	Thesis director

N° NNT : 2017TELE0014

THÈSE EN COTUTELLE
Spécialité : Télécommunications

Ecole doctorale Informatique, Télécommunications et Electronique de Paris
Ecole doctorale des Sciences et Technologie au Liban

Présentée par

DANDACHI Ghina

Pour obtenir le grade de
DOCTEUR DE TELECOM SUDPARIS

Le Multihoming dans les Réseaux Sans Fil Hétérogènes

Soutenue le 21 Juillet, 2017

Devant le jury composé de :

M. Mohamad ASSAAD, Professeur, CentraleSupelec	Rapporteur
M. Zaher DAWY, Professeur, Université américaine de Beyrouth	Rapporteur
M. Nazim AGLOUMINE, Professeur, Université d'Evry Val d'Essone	Examinateur
M. Guy PUJOLLE, Professeur, Université de Pierre et Marie Curie	Examinateur
M.YOUMNI ZIADE, Maître de conférences, Université Arabe de Beyrouth	Examinateur
Mme Nada CHENDEB, Professeur associée, Université Libanaise	Encadrante
M. Salaheddine ELAYOUBI, Ingénieur R&D, Orange Labs	Encadrant
M. Ziad FAWAL, Professeur, Université Libanaise	Co-directeur de Thèse
M. Tijani CHAHED, Professeur, SAMOVAR, Telecom SudParis	Directeur de Thèse

N° NNT : <>

Abstract

Fifth generation mobile networks (5G) are being designed to introduce new services that require extreme broadband data rates and utlra-reliable latency. 5G will be a paradigm shift that includes heterogeneous networks with densification, virtualized radio access networks, mm-wave carrier frequencies, and very high device densities. However, unlike the previous generations, it will be a holistic network, tying any new 5G air interface and spectrum with the currently existing LTE and WiFi.

In this context, we focus on new resource allocation strategies that are able to take advantage of multihoming in dual access settings. We model such algorithms at the flow level and analyze their performance in terms of flow throughput, system stability and fairness between different classes of users.

We first focus on multihoming in LTE/WiFi heterogeneous networks. We consider network centric allocations where a central scheduler performs local and global proportional fairness (PF) allocations for different classes of users, single-homed and multihomed users. By comparison with a reference model without multihoming, we show that both strategies improve system performance and stability, at the expense of more complexity for the global PF. We also investigate user centric allocation strategies where multihomed users decide the split of a file using either peak rate maximization or network assisted strategy. We show that the latter strategy maximizes the average throughput in the whole network. We also show that network centric strategies achieve higher data rates than the user centric ones.

Then, we focus on Virtual Radio Access Networks (V-RAN) and particularly on multi-resource allocation therein. We investigate the feasibility of virtualization without decreasing neither users performance, nor system's stability. We consider a 5G heterogeneous network composed of LTE and mm-wave cells in order to study how high frequency networks can increase system's capacity. We show that network virtualization is feasible without performance loss when using the dominant resource fairness strategy (DRF). We propose a two-phase allocation (TPA) strategy which achieves a higher fairness index than DRF and a higher system stability than PF. We also show significant gains brought by mm-wave instead of WiFi.

Eventually, we consider energy efficiency and compare DRF and TPA

strategies with a Dinklebach based energy efficient strategy. Our results show that the energy efficient strategy slightly outperforms DRF and TPA at low to medium load in terms of higher average throughput with comparable power consumption, while it outperforms them at high load in terms of power consumption. In this case of high load, DRF outperforms TPA and the energy efficient strategy in terms of average throughput. As for Jain's fairness index, TPA achieves the highest one.

Keywords- 5G networks, Heterogeneous Networks, Virtual Radio Access Networks, Millimeter wave, LTE, Multihoming, Flow-level modeling, Resource allocation, Multi-resource allocation, Power consumption.

Résumé

Les réseaux mobiles de la cinquième génération (5G) sont conçus pour introduire de nouveaux services nécessitant des débits de données extrêmement hauts et une faible latence. 5G sera un changement de paradigme qui comprend des réseaux hétérogènes densifiés, des réseaux d'accès radio virtualisés, des fréquences porteuses à ondes millimétrées et des densités de périphériques très élevées. Cependant, contrairement aux générations précédentes, 5G sera un réseau holistique, intégrant n'importe quelle nouvelle technologie radio avec les technologies LTE et WiFi existant.

Dans ce contexte, on se concentre sur de nouvelles stratégies d'allocation de ressources capables de bénéficier du multihoming dans le cas d'accès double au réseau. On modélise ces algorithmes au niveau du flux et analyse leurs performances en termes de débit, de stabilité du système et d'équité entre différentes catégories d'utilisateurs.

On se concentre tout d'abord sur le multihoming dans les réseaux hétérogènes LTE/WiFi. On considère les allocations centrées sur le réseau où un planificateur central effectue des allocations d'équité proportionnelle (PF) locale et globale pour différentes classes d'utilisateurs, utilisateurs individuels (single-homed) et multi-domiciliés (multihomed). Par rapport à un modèle de référence sans multihoming, on montre que les deux stratégies améliorent la performance et la stabilité du système, au détriment d'une plus grande complexité pour la stratégie PF globale. On étudie également les stratégies d'allocation centrées sur l'utilisateur, dans lesquelles les utilisateurs multihomed décident la partition de la demande d'un fichier en utilisant soit la maximisation du débit crête, soit la stratégie assistée par réseau. On montre que cette dernière stratégie maximise le débit moyen dans l'ensemble du réseau. On montre également que les stratégies centrées sur le réseau permettent d'obtenir des débits de données plus élevés que ceux centrés sur l'utilisateur.

Ensuite, on se concentre sur les réseaux d'accès radio virtuels (V-RAN) et en particulier sur l'allocation de multi-ressources. On étudie la faisabilité de la virtualisation sans diminuer ni la performance des utilisateurs, ni la stabilité du système. On considère un réseau hétérogène 5G composé de cellules LTE et mm-wave afin d'étudier comment les réseaux haute fréquence peuvent augmenter la capacité du système. On montre que la

virtualisation du réseau est réalisable sans perte de performance lors de l'utilisation de la stratégie "dominant resource fairness" (DRF). On propose une stratégie d'allocation en deux phases (TPA) qui montre un indice d'équité plus élevé que DRF et une stabilité du système plus élevée que PF. On montre également des gains importants apportés par l'adoption des fréquences mm-wave au lieu de WiFi.

Finalement, on considère l'efficacité énergétique et compare les stratégies DRF et TPA avec une stratégie éconergétique basée sur l'algorithme de Dinklebach. Les résultats montrent que la stratégie éconergétique dépasse légèrement DRF et TPA à charge faible ou moyenne en termes de débit moyen plus élevé avec une consommation d'énergie comparable, alors qu'elle les surpasse à une charge élevée en termes de consommation d'énergie moins élevée. Dans ce cas de charge élevée, DRF surpasse TPA et la stratégie éconergétique en termes de débit moyen. En ce qui concerne l'indice d'équité de Jain, TPA réalise l'indice d'équité le plus élevé parmi d'autres stratégies.

Mots clés- Réseaux 5G, Réseaux hétérogènes, Réseaux d'accès radio virtuel, Ondes millimétriques, LTE, Multihoming, Modélisation du flux, Allocation de ressources, Allocation multi-ressources, Consommation d'énergie.

To my parents, my brother and sisters
Thank you for all of your support along the way
To Ahmad

Acknowledgements

I owe my gratitude to all the people who have made this thesis possible with their help, support and contributions.

First and foremost, I would like to thank my director, Prof. Tijani CHAHED, who has given me an invaluable opportunity to do research and work on challenging and extremely interesting subjects over the past three years, my supervisor, Prof. Salaheddine ELAYOUBI, for his special theoretical ideas and mathematical expertise, and my supervisor in Lebanon, Dr. Nada CHENDEB, for her pursuance. They have been great mentors throughout my Ph.D. by helping me establish a direction of research and providing valuable guidance and advice. I will never forget the beautiful moments and the dialogues we had on various subjects.

I am also grateful to honorable Dr. Mohamad ASSAAD and Dr. Zaher DAWY for accepting to referee my thesis manuscript and for providing their valuable comments. Moreover, my thanks go to Prof. Nazim AGLOUMINE, Prof. Guy PUJOLLE and Dr. Youmni ZIADE who has generously accepted to be part of my thesis evaluation committee.

I would like to thank everybody at the Lebanese University, especially Prof. Fawaz EL-OMAR, Prof. Mohamad KHALIL and Prof. Ziad FAWAL for making this joint supervision PhD study possible. My sincere thanks also goes to Prof. Bertrand GRANADO, Ms. Marilyn GALOPIN and Ms. Catherine BOUDEAU from the doctoral school "EDITE"- Paris, M. Christophe DIGNE, Ms. Francoise ABAD and Ms. Veronique GUY from TELECOM SudParis , as well as to Mrs. Jana ELHAJJ and Mrs. Zeinab IBRAHIM from the doctoral school "EDST"-Lebanon for their help especially in the administrative procedure for my Thesis defense.

I would like to thank my friends Noujoude, Aya, Nour, Ranime, Asmaa, and Shohreh for their friendship. I cherish every single moment we have shared in Paris or in Lebanon. You were really, a beautiful support along the way. Noujoude and Aya, I never forget our craziest times.

Most importantly, I would like to thank my beautiful fiancé Ahmad for his love, support, and positive energy. He has always been by my side, especially during the hardest moments of my Ph.D.

Last but not least, I owe my deepest thanks to my wonderful family, my parents, my sisters and my brother, who are always there for me even when

were thousand miles apart. I express my gratitude to my parents for having guided me through life, and supported and encouraged me to move France to pursue my Ph.D. studies. You enlightened my life with knowledge.

Contents

List of Figures

List of Tables

List of Abbreviations

ANDSF Access Network Discovery and Selection Function

BBF Bottleneck-Based Fairness

BBU Baseband Unit

BMF Bottleneck Maximum Fairness

C-RAN Cloud Radio Access Network

CA Carrier Aggregation

CAPEX Capital Expenditures

CoMP Coordinated Multipoint

CP Cyclic Prefix

CPRI Common Public Radio Interface

CPU Central Processing Unit

CQI Channel Quality Indicator

CSMA/CA Carrier Sensing Multiple Access/Collision Avoidance

D2D Device-to-Device

DC Dual Connectivity

DIDA Data Identification in ANDSF

DRF Dominant Resource Fair

DRFQ Dominant Resource Fair Queuing

E3F EARTH Energy Efficiency Evaluation Framework

EARTH Energy Aware Radio and Network Technologies

EDF Earliest Deadline First

EE Energy Efficiency

eICIC Enhanced Inter-Cell Interference Coordination

ETSI European Telecommunications Standards Institute

FCC Federal Communications Commission

FFT/IFFT Fast/Inverse Fast Fourier Transform

FIFO First In First Out

GOPS Giga Operations Per Second

GSM Global Systems for Mobile Communications

GTP GPRS Tunneling Protocol

HARQ Hybrid Automatic Repeat Request

HetNets Heterogeneous Networks

HIP Host Identity Protocol

I/Q In-Phase/Quadrature

ICT Information and Communications Technologies

IFOM IP Flow Mobility

IoT Internet of Things

ISMP Inter-System Mobility Policies

ISRP Inter-System Routing Policy

L-GW Local Gateway

LAA License Assisted Access

LAN Local Area Network

LIPA Local IP Access

LTE Long Term Evolution Network

LTE Long Term Evolution

LWA LTE/WLAN Aggregation

M2M Machine-to-Machine

MAC Medium Access Control

MIMO Multiple Input Multiple Output

mm-wave millimeter wave

MPTCP Multi-Path Transport Control Protocol

NLOS Non-Line-of-Sight

O&M Operations and Maintenance

OFDM Orthogonal Frequency Division Multiplexing

OPEX Operational Expenditures

OPIIS Operator Policies for IP Interface Selection

PA Power Amplifier

PDCP Packet Data Convergence Protocol

PDU Packet Data Unit

PF Proportional Fairness

PFS Proportionally Fair Scheduling

PMIP Proxy Mobile IP

QoE Quality-of-Experience

QoS Quality of Service

RAM Random Access Memory

RF Radio Frequency

RR Round Robin

RRH Remote Radio Head

S-GW Serving Gateway

SaMOG S2a-Based Mobility on GTP

SCBS Small Cell Base Stations

SCTP Stream Control Transport Protocol

SDR Software Defined Radio

SNR Signal-to-Noise Ratio

TDMA Time Division Multiple Access

TPA Two-Phase Allocation

TRX Transceiver

TTI Transmission Time Interval

UE User Equipment

UMTS Universal Mobile Telecommunications System

V-RAN Virtual Radio Access Networks

Chapter 1

Introduction

A continuous revolution of wireless networks was predicted by Nicola Tesla, the inventor of wireless telegraph, in 1915 [1]:

> It's all a wonderful thing. Wireless is coming to mankind in its full meaning like a hurricane some of these days. Some day there will be, say, six great wireless telephone stations in a world system connecting all the inhabitants on this earth to one another, not only by voice, but by sight. It's surely coming.
>
> Nikola Telsa Sees a Wireless Vision, NYT, 1915

The mobile cellular era started in the early 1980s. Since then, four cellular generations were implemented offering several services as detailed in Fig. 1.1. The first generation used analog transmission for speech services. Then, second-generation (2G) mobile systems, such as the Global Systems for Mobile Communications (GSM), were introduced in the end of 1991 offering low bit data rate and digital speech service. Third generation (3G) mobile systems were then born in 2001 as a network design independent of the platform technology offering different services, including high quality audio, video calls, and broadband wireless data. Mobile Internet connectivity gained wide spread popularity with the fourth generation Long Term Evolution (LTE) mobile communication system introduced in 2009. LTE eliminated circuit switching, and employed an all-IP network with packet switching that supports low to high mobility applications and a wide range of data rates.

By 2021, the fifth generation mobile network is expected to appear as a holistic network which will ensure user experience continuity for more challenging services with huge capacity, real-time constraints and massive objects connection. Among these services we note: virtual reality/augmented reality applications, the Internet-of-Things (IoT), ultra-reliability for Device-to-Device (D2D) communication, video streaming, etc. The wide adoption

of advanced multimedia applications increases mobile and Wi-Fi traffic and requires faster, higher bandwidth and more intelligent networks.

The objective of the cellular generation evolution was always to provide higher data rates with better user experience. With the emergence of new services and communication types (M2M, D2D, IoT, etc.), network densification and collaboration between different access technologies becomes necessary in order to take advantage of all available bandwidth. It is important to note that users' traffic activity takes place mainly within their homes or offices. In this case, users may have access concurrently to cellular broadband and Wi-Fi access points during long periods of the day. In this context, 3GPP recently standardized "Dual connectivity" for simultaneous connections on two access networks. A general definition of this simultaneous connection feature with sending and receiving data on more than one access network is called "multihoming". Multihoming, considered as an evolution of traffic offloading, increases resource usage as well as multihoming capable users' data rates.

Exploiting resources to their maximum, we are still limited by the used spectrum. As spectrum is scarce in currently explored bands, higher frequency bands, including millimeter waves, are needed. The adoption of millimeter wave frequency bands (mm-wave) in 5G networks offers a higher spectrum allocation up to $1.7GHz$ which serves as a solution for the huge capacity requirements. The use of mm-wave is now standardized, IEEE 802.11ad is one such example known as WiGig. The advantages of mm-wave include also smaller size of antennas which allows to implement more than 64 antennas with directional beams on few centimeters. However, mm-wave networks require more processing than conventional sub-GHz networks.

This network densification with mm-waves and all of these enhancements increases the capital and operational expenditures (CAPEX and OPEX) which contradicts with the green feature of 5G networks. One possible solution for decreasing CAPEX and OPEX is the Radio Access Network (RAN) centralization or virtualization known as C-RAN/V-RAN enabled by the network function virtualization (NFV) techniques. The main concept behind C-RAN is to split the Baseband Unit (BBU) and the Remote Radio Head (RRH) located at the cell site. The connection between the BBU and the RRH is ensured by a low latency fronthaul. Moreover, BBUs are pooled and centralized or virtualized into a BBU pool. The adoption of C-RAN/V-RAN in 5G networks reduces the CAPEX by deploying only low cost and small footprint RRH instead of deploying a whole base station. It allows also dynamic many-to-one BBU-to-RRH allocation depending on the required processing resources which in turn decreases the OPEX. C-RAN's energy efficiency is also increased by the negligible power consumption needed for RRH cooling and by performing coordination functions between access networks such as coordinated multipoint (CoMP) and enhanced inter-cell interference coordination (eICIC).

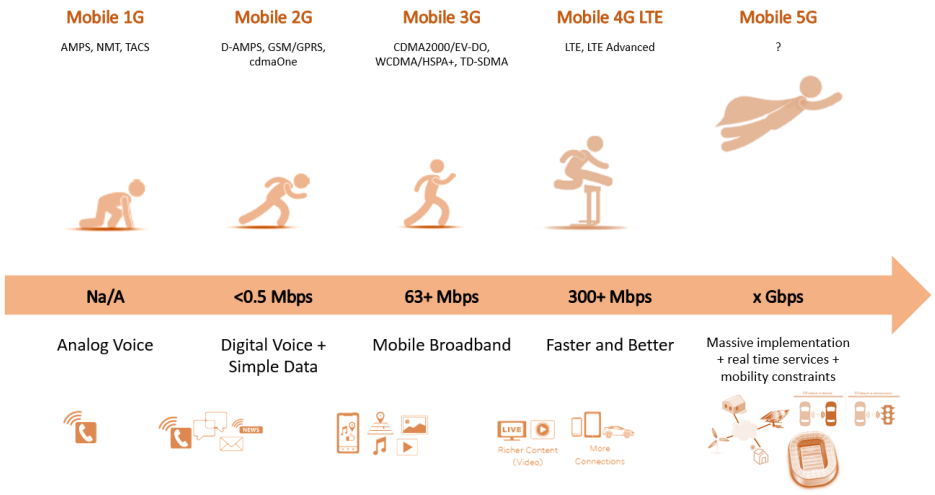


Figure 1.1: Approximate timeline of the evolution of the mobile communications standards.

1.1 Scope and contributions

Motivated by the need of 5G networks to deliver new services with higher data rates, we focus in this thesis on performance of heterogeneous networks composed of macro and small cells, in the presence of multihoming capability.

We first consider two categories of resource allocation strategies: network centric and user centric. The network centric strategies refer in our case to local Proportional Fairness (PF) and global PF. The user centric ones are: peak rate maximization and network assisted strategy. These strategies are modeled at the flow level and solved both analytically and numerically in order to evaluate the achievable data rate for each class of users as well as system stability.

Second, we investigate the V-RAN context. In this context, we adopt multi-resource allocation for jointly allocating radio and processing resources, and investigate whether it is possible to virtualize a radio access network while maintaining users performance and how this virtualization will affect users performance and system stability. For this aim, we considered three strategies: PF, dominant resource fairness (DRF), and a two-phase allocation (TPA) which we proposed and which achieves a trade-off between system stability and users' allocation fairness.

Eventually, we investigate the energy efficiency of resource allocation in the V-RAN and study energy efficiency of DRF and TPA. We compare the joint resource allocation strategies with energy efficient allocation using Din-

klebach's algorithm, and the baseline network model without virtualization. These strategies were modeled and evaluated numerically with considering real power models from the literature.

1.2 Organization of the thesis

The remainder of this dissertation is organized as follows.

Chapter 2 provides an extensive introduction to 5G networks, outlining its fundamental architecture including HetNets, C-RAN/V-RAN, mm-wave networks, and multihoming and interworking techniques. In addition, 5G research directions are outlined with a detailed analysis of related research studies and findings.

In Chapter 3, we evaluate the performance of multihoming in HetNets. Two main approaches are covered: network centric versus user centric. We evaluate the achievable throughput and system stability obtained by different allocation strategies for both approaches.

In Chapter 4, we focus on V-RAN and evaluate multi-resource allocation strategies. Emphasis is placed on users' performance to meet the same results obtained in the case of HetNets without V-RAN. As for reducing the operational cost of 5G networks, we evaluate the power consumption and savings coming from network virtualization.

Finally, Chapter 5 concludes the dissertation and presents some future research directions.

1.3 Publications

[1] G. Dandachi, S. E. Elayoubi, T. Chahed, N. Chendeb and H. Jebalia, "Comparing Resource Allocation Schemes in Multi-Homed LTE/WiFi Access Networks," *2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)*, Boston, MA, 2015.

[2] G. Dandachi, S. E. Elayoubi, T. Chahed and N. C. Taher, "Performance evaluation of user centric multihoming strategies in LTE/WiFi networks," *2016 IEEE Wireless Communications and Networking Conference*, Doha, 2016.

[3] G. Dandachi; S. Elayoubi; T. Chahed; N. Chendeb, "Network centric versus user centric multihoming strategies in LTE/WiFi networks," in *IEEE Transactions on Vehicular Technology* , vol.PP, no.99, 2016.

[4] G. Dandachi, T. Chahed, S. E. Elayoubi, N. C. Taher and Z. Fawal, "Joint allocation strategies for radio and processing resources in Virtual-Radio Access Network (V-RAN)," submitted.

[5] G. Dandachi, T. Chahed, S. E. Elayoubi and N. C. Taher, "Joint Resource Allocation in C-RAN/V-RAN," submitted.

[6] G. Dandachi, T. Chahed, S. E. Elayoubi and N. C. Taher, "On green multihoming-capable virtualized networks," under preparation.

Chapter 2

Resource Orchestration in 5G: Overview and Litterature Review

2.1 Introduction

Fifth generation (5G) is not as previous generations, an evolution of the existing, but it is rather considered as a cellular network revolution that builds on the evolution of existing technologies. These technologies are complemented by new radio concepts that are designed to meet the new and challenging requirements of some use cases today's radio access networks cannot support [2] [3].

This revolution is necessary to offer new services to 5G users with good quality of service (QoS). These services include:

- Good service even in very crowded places.
- Similar user experience for end-users on the move as for static users.
- The Internet of Things (IoT). Basically, anything that profits from being connected will be connected.
- Machine-to-machine (M2M) or device-to-device (D2D) communication with real-time constraints, enabling new functionalities for traffic safety, traffic efficiency, smart grid, and e-health.
- Huge capacity increase that could be achieved by having more spectrum, better spectrum efficiency and a large number of small cells.

In parallel to the data starving services, several technological concepts that were not supported in previous cellular generations are now potential

5G scenarios to answer users demands. We mainly note: D2D communications, ultra-reliable communications, massive machine communications, IoT, Cloud computing, and hybrid networks.

However, the full image of 5G is not clear until now, and research projects are being conducted in order to fit all puzzle pieces and figure out 5G's unified big picture by 2020. Among these projects, we note 5G-PPP [4], NGMN [5], METIS [6], COHERENT [7], and mmMAGIC [8].

On the other hand, ultra high data rates, extremely low latency, anywhere anytime coverage, huge energy saving – most of the promises made by 5G are associated with their respective challenges. Among these challenges we address in this thesis *network densification* in the form of heterogeneous networks (HetNets). Heterogeneous architecture is an underlining feature of 5G, however deployment and management of HetNets in 5G scenarios is yet to be explored. Given the need to satisfy overwhelming capacity demands in 5G, *mm-wave spectrum* (3-300 GHz) is expected to offer a very compelling long term solution by providing additional spectrum to 5G networks. Hence, the challenge is the integration of mm-wave in heterogeneous and dense networks as well as the backward compatibility and integration with legacy 4G/3G networks. Furthermore, Cloud radio access networks (C-RAN) contribution to 5G is considered as a cost effective and energy efficient solution for dense 5G deployment. From an energy point of view, cost and energy consumption are major considerations for 5G. C-RAN and energy efficiency techniques could help in performance improvements.

Although HetNets were introduced in 4G networks, their complexity has increased in 5G networks. In this chapter, we will try to build a clear image of HetNets in 5G cellular networks. We consider different technologies with a special focus on mm-wave networks given its important role in 5G networks. We then address the available standards in HetNets that allow interworking and multihoming between different radio access technologies. Afterwards, we consider the virtualization of 5G HetNets and its benefits. Different resource allocation strategies in the literature are also presented for single-resource as well as for multi-resources. Finally, we give an overview of existing works addressing energy efficiency strategies in 5G networks.

2.2 Heterogeneous networks/Multi-RAT

Today's 3G and 4G networks are designed primarily with a focus on peak rate and spectral efficiency improvements. In the 5G era, we will see a shift towards network efficiency with 5G systems based on dense heterogeneous networks architectures. HetNets are among the most promising low-cost approaches to meet the industry's capacity growth needs and deliver a uniform connectivity experience. A HetNet comprises a group of small cells that support aggressive spectrum spatial reuse coexisting within macro cells

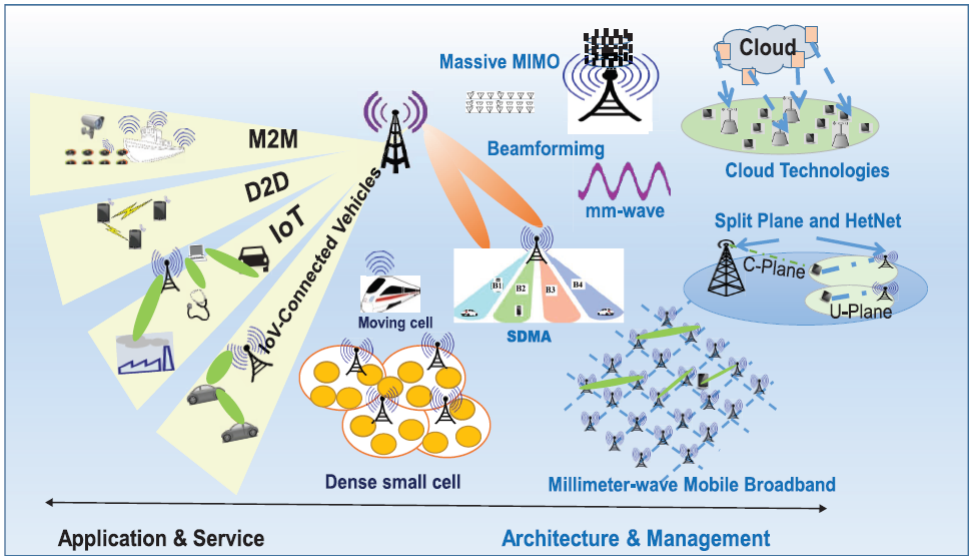


Figure 2.1: Next Generation 5G Wireless Networks (Source: [3]).

as shown in Fig. 2.2. However, HetNets will be architected to incorporate an increasingly diverse set of frequency bands within a range of network topologies, including macro cells in licensed bands (e.g., long term evolution network or LTE) and small cells in licensed or unlicensed bands (e.g., WiFi). New higher frequency spectrum (e.g., millimeter-wave or mm-wave) may also be deployed in small cells to enable ultra-high-data-rate services.

2.2.1 Architecture

HetNets are formed of macro cells and small cells. A macro cell is generally divided into several sectors in order to increase the spatial frequency reuse which increases the network capacity. Typically, a macro cell is implemented as a tri-sectorial base station (BS) with each sector of 120°. However, different definitions are considered for choosing the cell type, it can consider the radius of the cell, the number of connected users and the deployment options.

As their name indicates, small cells provide a smaller coverage area than a macro cell. As shown in Fig. 2.2 , a macro cell overlaps several small cells. There are several types of small cells such as micro, pico, femto and relay cells, ordered in decreasing order of coverage and transmission power. These small cells can be managed by the same operator as a macro cell or by a different operator and require a lower installation cost. In addition, it is worth to note that small cells are mainly deployed in order to support the increasing rates of data services but can also support voice services.

In 5G HetNets, macro and small cells may be connected to each other

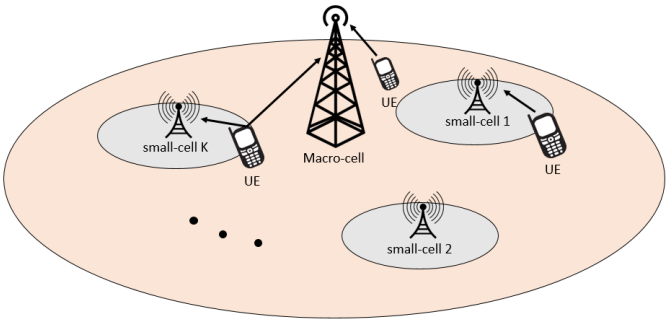


Figure 2.2: Heterogeneous network model.

via ideal or non-ideal backhaul, resulting in different levels of coordination across the network for mobility and interference management. Increasing degrees of network cooperation, from loose network node coordination to completely centralized control (i.e. tight coupling), will provide increasing levels of network capacity.

2.2.2 WiFi small cells

Widely deployed WiFi systems are playing an increasingly important role in offloading data traffic from the heavily loaded cellular network, especially in indoor traffic hotspots and in poor cellular coverage areas. Very recently, the Federal Communications Commission (FCC) voted to make 100 MHz of spectrum in the 5 GHz band available for unlicensed WiFi use based on the IEEE 802.11ac standard [9], giving carriers and operators more opportunities to push data traffic to WiFi. WiFi access points have even been regarded as a distinct tier of small cells in heterogeneous cellular networks.

Wireless local access networks (WLAN) technology evolution is mainly carried out within the WLAN IEEE 802.11 working group which released multiple set of standards for various operating frequencies and ranges specification. The first release was IEEE 802.11 original standard that was defined in 1997 and clarified in 1999 with a data rate up to 11 Mbps [10]. In this thesis, we focus on the last two standards IEEE 802.11n [11] and IEEE 802.11ac [12] as the newest sub-6 GHz standards. WiFi networks implemented in most home networks are IEEE 802.11n based. The latter operates at both 2.4 and 5 GHz frequencies and employs orthogonal frequency division multiplexing (OFDM) modulation technique. The antenna technology used with the IEEE 802.11n standard is known as Multiple Input, Multiple Output (MIMO) which allows the coordination with similar technologies and offers data rates up to 300 Mbps. IEEE standards evolved by introducing IEEE 802.11ac operating at 5 GHz with higher channel bandwidth up to 160 MHz and a data rate up to 866 Mbps. The IEEE 802.11ac

standard uses a wider channel and an improved modulation scheme that also supports more clients called multi-user MIMO. The mm-wave spectrum in the IEEE 802.11ad standard [13, 14] will be tackled later in this chapter.

In 3GPP, the LTE/WiFi interworking became possible by implementing a modem-level aggregation for superior performance leveraging dual connectivity standardized in Release 12 (R12) [15]. A new standard is being studied: the LTE/WLAN Aggregation (LWA) for mobile operators leveraging existing carrier WiFi deployments.

European Telecommunications Standards Institute (ETSI) defined in [16] two different ways of integrating heterogeneous wireless networks: loose coupling and tight coupling interworking. In a loosely coupled system, shown in Fig. 2.3, the heterogeneous wireless networks are not connected directly. Instead, they are connected to the same IP network (i.e., the internet). Loose coupling uses the subscriber databases without the need for a user plane interface. To use the WiFi network, the User Equipment (UE) first needs to scan for available WiFi APs. It needs to authenticate on the selected AP and then sends or receives data. Even if mechanisms such as the Access Network Discovery and Selection Function (ANDSF) and Hotspot 2.0 [17] aim to accelerate the process, the UE still needs time before it can use the WiFi network and in practice, it has to stay for a while in the AP coverage to start offloading.

In contrast, a tightly coupled system, as shown in Fig. 2.4 consists of a common packet scheduler for cellular and WLAN systems, connecting the latter to the mobile core network and achieving the integration between both systems at the lower layers. The UE still needs to use WiFi security mechanisms, which are time consuming. This was standardized by 3GPP on Release 10 [18].

A very tight solution was proposed in [20] between WiFi access points and LTE eNodeBs. The main idea is to connect WiFi access points that are covered by an eNodeB to this eNodeB. Such very tight coupling is made possible by putting security functions and the layer 3 (L3) protocol stack of the gateway in the network. In other words, residential gateways as well as access points specifically deployed by the operator are virtualized. The device deployed in the customer premise or in the hot spot is called a virtual residential gateway and the device that hosts security and L3 functions is called a gateway hotel as it is possible to implement gateway functions of several customers in the same equipment (see figure 2.5). The main principle of the proposal is to keep all control functions (security, mobility, session management) in the LTE network and to use WiFi only to transmit data. Very tight coupling between LTE and WiFi makes possible to help the terminal to very quickly set up Layer 2 connection with WiFi access point. The objective is to allow terminals that are covered by a residential virtual gateway for a short period (10 seconds to 1 minute) to use the WiFi network to offload the cellular network.

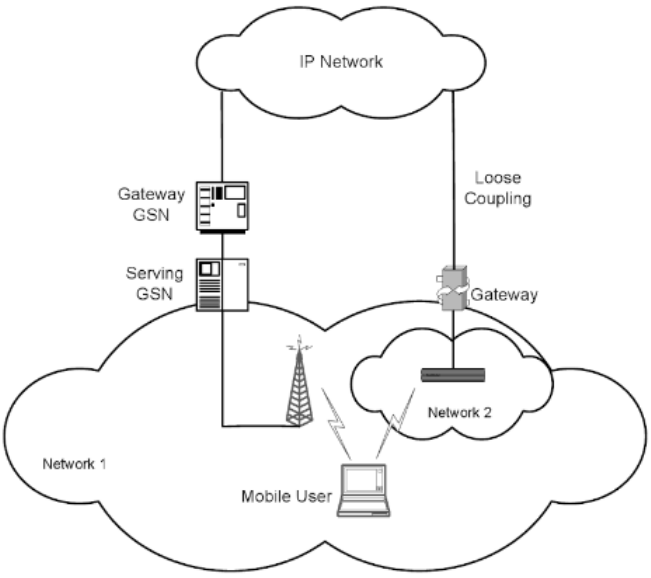


Figure 2.3: HetNet architecture with loose coupling (Source: [19]).

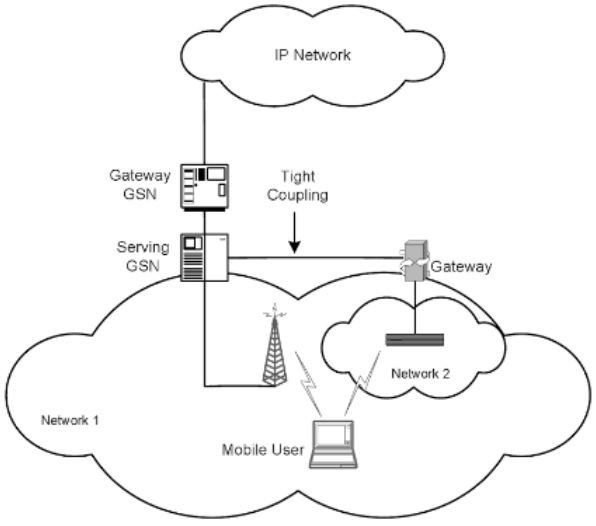


Figure 2.4: HetNet architecture with tight coupling (Source: [19]).

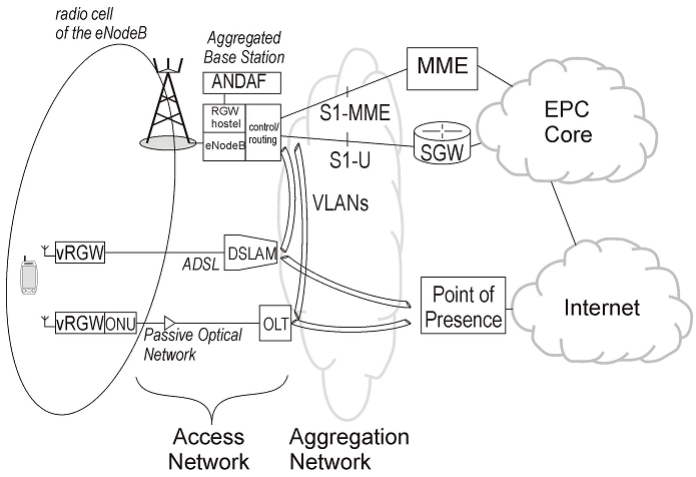


Figure 2.5: HetNet architecture with very tight coupling (Source: [19]).

2.2.3 LTE small cells

LTE small cell networks are highly dense networks constituting of home eNodeBs, indoor enterprise eNodeBs as well as outdoor deployed eNodeBs. Some of the major challenges of the LTE small cell networks are: 1) maintaining the desired QoS with respect to downlink and uplink packet data transmission 2) efficient handover 3) interference co-ordination with neighbors. Especially in the uplink direction, i.e., from UE to small cell eNodeB, the task of delivering a wide variety of application layer packets is complicated due to limited transmission power of the UE, limited battery resources at UE and time-varying nature of wireless channels. Thus in order to satisfy the wide variety of applications in the uplink direction, an efficient QoS aware uplink scheduler in the eNodeB is required for guarantying uplink QoS for all the packet data transmissions.

From an architectural point of view, two deployment scenarios were identified in [21], namely small cells co-existing with macro cells,known as *Hot Spot*, and small cells without macro cells known as *Not spot*. In such areas, only basic network coverage is needed, which can be adequately supported by lower cost small cells rather than more expensive resource from the macro site. Not-spot small cells are perfect for network coverage extension to reach the rural areas, both indoors and outdoors. The Not-spot scenario may potentially suffer however from high volume of handover signaling load, which may impact the users Quality-of-Experience (QoE).

Hot spots [22] on the other hand enable the operator to provide additional capacity where needed. A new architecture with split control and user plane has been proposed in 3GPP Release 12 [23]. In this architecture, the control plane will be handled by a macro cell and user plane will be handled

by small cells. Since small cells are deployed within the radio coverage of an existing macro cell network, necessary techniques should be put in place in order to enable small cells to work autonomously upon the failure of the corresponding macro cell.

2.2.4 Mm-wave small cells

Capacity for wireless communication depends on spectral efficiency and bandwidth. It is also related to cell size. Cell sizes are becoming small and physical layer technology is already at the boundary of Shannon capacity [24]. It is the system bandwidth that remains unexplored. Presently, almost all wireless communications use spectrum in 300 MHz to 3 GHz band, often termed as "sweet spot" or "beachfront spectrum" [25]. In order to increase capacity, wireless communications cannot help facing the new challenges of high frequency bandwidth. The key essence of next generation 5G wireless networks lies in exploring this unused, high frequency mm-wave band, ranging from 3 $\sim$ 300 GHz. Even a small fraction of available mm-wave spectrum can support hundreds of times of more data rate and capacity over the current cellular spectrum [26]. Thus, the availability of a big chunk of mm-wave spectrum is opening up a new horizon for spectrum constrained future wireless communications [26].

Indeed, the usage of mm-wave in cellular networks is a promising solution because of the huge channel bandwidth offered by this technology. The essential component of mm-wave systems is the directional beamforming that provides array gains that can be used to overcome the high path loss and achieve sufficient link margins. Adaptive beamforming using large arrays for array gain distinguishes mm-wave and microwave wireless systems. Hence, modeling beamforming in mm-wave networks is critical for precise characterization of the network behavior and accurate evaluation of its performance.

An example of mm-wave access networks is the IEEE 802.11ad standard [14] so called WiGig. IEEE 802.11ad specifies the physical and MAC layers in the 60GHz band to support multi-gigabit wireless applications including instant wireless synchronization, wireless display of high definition (HD) streams, cordless computing, and internet access. In the physical layer, two operating modes are defined, the OFDM mode for high performance applications (e.g., high data rate), and the single carrier (SC) mode for low power and low complexity implementation. In addition, a hybrid multiple access of carrier sensing multiple access/collision avoidance (CSMA/CA) and time division multiple access (TDMA) is adopted for transmissions among devices. CSMA/CA is more suitable for bursty traffic such as web browsing to reduce latency, while TDMA is more suitable for traffic such as video transmission to support better QoS.

A. Beamforming in mm-wave

The main objective of adaptive beamforming is to shape the beam patterns (e.g., by beamsteering) so that the received signal-to-noise ratio (SNR) is maximized. Full control of beam pattern shaping requires changing both the amplitude and phase of transmitted signals. The need for low-cost and low-power hardware, however, has pushed mm-wave towards a simpler analog architecture that contains only digitally controlled constant modulus phase shifters. Hybrid precoding proposed in [27] divides the required precoding processing between the analog and digital domains, and hence allows better control of the beam shape.

B. Mm-wave mobile boradband frame structure

As in 4G systems, mm-wave uses also OFDM and single-carrier waveform as multiplexing schemes. We show in Fig. 2.6 a mm-wave frame structure as described in [28]. The basic transmission time interval (TTI) is a solt of $62.5\mu s$ duration. Subframe, frame and superframe's duration are chosen equal to those in LTE systems (1 ms, 10 ms and 40 ms, respectively) in order to facilitate the interworking between both technologies. The cyclic prefix (CP) is chosen to be 520 ns, which gives sufficient margin to accommodate the longest path, different deployment scenarios, and the potential increase of delay spread in the case of small antenna arrays or wider beams. The subcarrier spacing is chosen to be 480 kHz, small enough to stay within the coherent bandwidth of most multipath channels expected in mm-wave. The corresponding OFDM symbol length (without CP) is 2.08 us, resulting in 20 percent CP overhead. The subcarrier spacing is also wide enough to keep the size of fast/inverse fast Fourier transform (FFT/IFFT) small (2048 points for 1 GHz system bandwidth) and accommodate inaccuracies of low-cost clocks.

C. Interworking between mm-wave and LTE

A hybrid LTE/mm-wave system can improve coverage and ensure seamless user experience in mobile applications. In a hybrid LTE/mm-wave system, system information, control channel, and feedback are transmitted in the LTE system, making the entire millimeter-wave spectrum available for data communication. Compared with millimeter waves, the radio waves at < 3 GHz frequencies can better penetrate obstacles and are less sensitive to non-line-of-sight (NLOS) communication link or other impairments such as absorption by foliage, rain, and other particles in the air. Therefore, it is advantageous to transmit important control channels and signals via cellular radio frequencies, while utilizing the millimeter waves for high data rate communication.

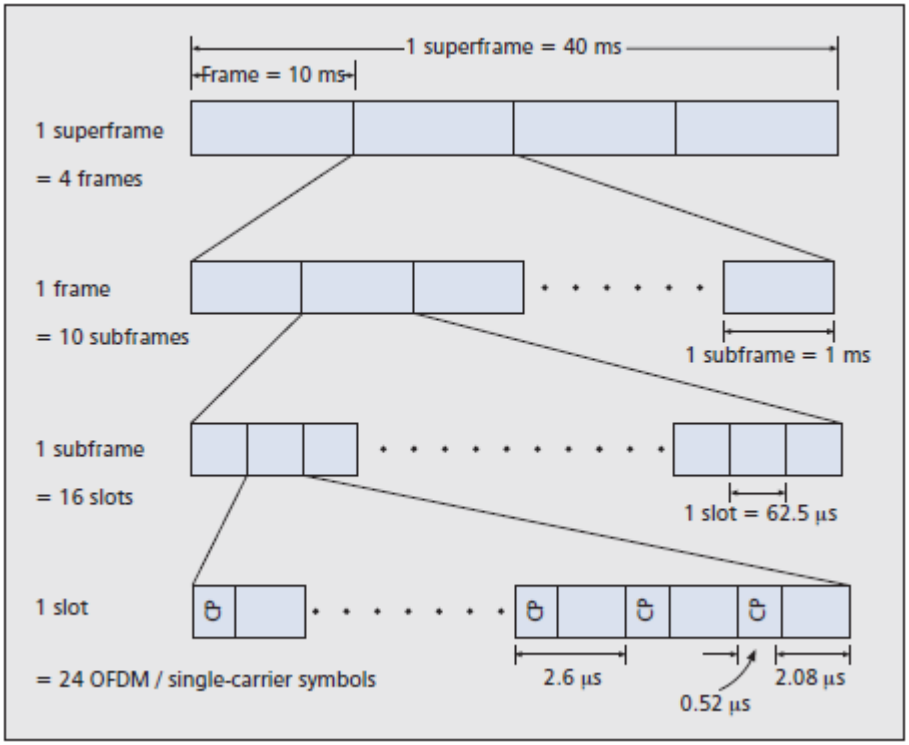


Figure 2.6: Mm-Wave frame structure [28].

Authors in [29] propose a RAN-level tight coupling solution for LTE/WiGig interworking. A control/user-plane (C/U) splitting mechanism has been considered for LTE and WiGig in mm-wave HetNet based on the proposition in 3GPP Release 12 for LTE-Advanced systems [30]. By using the proposed C/U splitting, user data traffic is offloaded to mm-wave small cells while user mobility management is centrally controlled by LTE macro cells. The proposed architecture enables RAN-level LTE/WiGig interworking and makes full use of the novel technologies specified for WiGig.

2.3 Multihoming

HetNets were designed such that traffic can be offloaded between available access networks. However, concurrent multiple access to more than one network in wireless networks has recently been standardized in Release 12 under the name of "Dual Connectivity" [15]. In this section, we introduce the aspects and standards enabling multihoming's concept implementation with focus on the *dual connectivity* standard. We also present a literature overview on interworking and network selection strategies in this context.

2.3.1 Multihoming aspects

Multihoming was first proposed as a redundancy solution for wired networks. Recently, the coexistence of different wireless access network technologies has renewed this concept and became an attractive topic for study during the past years. Wireless networks multihoming concept started with offloading [31, 32, 33], passing by load balancing [34], optimal distribution [35] [36] [37], as well as concurrent multiple access [38, 39, 40].

In data offloading, users' data is redirected intelligently in the network in order to avoid overload as well as improve the end user experience. Data offloading had limited success in 3GPP Release 8 because of WiFi/LTE integration complexities. However, the introduction of LTE-Advanced in 3GPP Release 10 made data offloading more relevant. Three key data offloading areas are proposed: Local IP Access (LIPA), Selected IP Traffic Offload (SIPTO) and IP Flow Mobility (IFOM). We will be looking at the details of each of these methods later on in this section.

Non-seamless offloading methods were proposed in [31] in order to enhance the service discontinuity when a user chooses to turn on the WiFi interface. A cross-system traffic steering framework is proposed in [32] as a cost-effective integration of WiFi and LTE small cells and showed 200% gains on traditional schedulers. Authors in [33] proposed a network assisted offloading strategy in a heterogeneous network where users apply traffic steering proportions transmitted by the network.

Load balancing concept was introduced in wired networks [41]. Such load balancing system must determine the available bandwidth through an

access link, assign incoming and outgoing traffic, and detect access links failure. For this aim, a reliable routing protocol must be considered [42]. Similarly, load balancing management could be obtained in heterogeneous wireless networks by dynamically optimizing the packets' split ratio between multiple access networks as shown in [34]. Such strategy might be based on the load information and channel quality information at each access network. A dynamic load balancing management scheme was proposed in [43] for WiFi/LTE networks in order to balance the network load. Moreover, several user assignment studies in HetNets use the optimization of a utility function on which the network operator splits traffic between different access networks. Such optimization-based load balancing was adopted in [35], [36] and [37] but did not consider the possibility of simultaneous multiple access.

Simultaneous multiple access is the newest concept in HetNets. For simplicity, we use the term "multihoming" instead of simultaneous multiple access. A resource allocation model was proposed for multihomed access networks with traffic flows optimal allocation in [38]. The adoption of multihoming has been proven to enhance the quality of streaming services in wired networks as shown in [44]. Multihoming can also improve streaming services in wireless networks as shown in [39, 40].

2.3.2 Multihoming technology enablers

Throughout the past years, 3GPP and IETF worked hard in order to standardize different HetNets interwoking schemes. Their main interest was to standardize the users mobility between accesses, the transport layer support of multihoming, and frequency resource scheduling known as "Dual Connectivity" (DC).

A. Mobility in Heterogneous Networks

Non-seamless offloading between LTE and WiFi is disturbing, especially for real-time applications that require the continuity of service (e.g., VoIP, Video Conference, HTTP page). It is highly desirable that mobile operators provide seamless service continuity between cellular and WiFi accesses with involving both user plane routing and control plane functions. This seamless continuity can be first supported by ensuring a service layer continuity even when the IP address has changed which is not supported in TCP/IP. In this section, we present some of the seamless continuity standardized technologies. Several mechanisms are proposed by 3GPP describing the offload management in 3GPP networks. I-WLAN is the first approach allowing local area network access to the 3GPP core.

Next we present two approaches in Release 10 that act on different axes by choosing either offloading or load balancing. While LIPA and SIPTO allow offloading of the traffic into a local area network in order to reduce

the load on cellular networks, IFOM allows the usage of dual radio connections through WLAN and 3GPP systems. More precisely, LIPA consists in offloading the traffic seamlessly at a local gateway (L-GW) breakout point into a local area network (LAN) when needed. SIPTO presents a selective offloading from the serving gateway (S-GW) at a Femtocell or macrocell scale towards a (L-GW) connected to the internet in order to decrease the load on the cellular network in a crowded region [45]. Moreover, the WLAN offload mechanism IFOM describes the mechanism adopted when a UE has several data sessions over different types of access simultaneously: 3GPP and WiFi networks. Until now, the user selects which traffic to offload on WiFi and which to keep on the 3GPP access (for example LTE). However, the previously described offloading techniques require mobility management mechanisms at the network layer for roaming and offloading of low-latency and higher data rates packet services. For this reason, 3GPP and IETF introduced respectively two mobility protocols: GTP[1] and PMIP[2]. Both GTP and PMIP based mobility are explained in [46] from a technical and economical points of view in such a way so as to optimize the interworking between WiFi and other networks. In general, a session continuity can be maintained: (i) between WiFi and 3GPP networks by using GTP, and (ii) between WiFi and non-3GPP networks by using PMIP.

These offloading mechanisms and mobility solutions are complemented by the 3GPP network selection strategy, the ANDSF, where the ANDSF element is a server located in the operator's network that distributes the network selection information and policies using a standardized interface (S14). ANDSF was first standardized in Release 8 where it introduced the Inter-System Mobility Policies (ISMP), it was then enhanced in Release 9 and 10 in which it introduced the Inter-System Routing Policy (ISRP) for routing IP traffic simultaneously over multiple radio interfaces. In Release 11, data identification in ANDSF (DIDA) was introduced, and finally in Release 12, it introduced the Operator Policies for IP Interface Selection (OPIIS).

Authors in [47] studied the non-seamless offload between cellular and non-trusted WiFi aiming to achieve an enhanced offloading from WiFi to LTE with guaranteed session continuity using IP encapsulation technique. Several works addressed host mobility in the current IP architecture using the Host Identity Protocol (HIP) proposed by IETF. Authors in [48] provide an in-depth look at HIP, discussing its architecture, design, benefits and potential drawbacks.

In order to support real-time services and QoS over trusted WiFi, operators deploy WiFi APs that are considered trusted and enable mobile services

[1]GTP: GPRS Tunneling Protocol, A group of protocols used for GPRS packet traffic and control in mobile core networks

[2] PMIP: Proxy Mobile IP, the network-based mobility mechanism supported by 3GPP2 and WiMAX

and features over these WiFi APs such as real-time services, and generally, end-to-end QoS over trusted WiFi, this solution is called SaMOG[3]. This integration is described as a use of Small Cell Base Stations (SCBS) within LTE cell coverage. According to [32] these SCBSs have the ability to transmit over both licensed and unlicensed spectrum. Traffic is steered in such a way that delay-sensitive applications are routed using LTE and delay-tolerant ones through WiFi. Some studies aim to minimize the SCBS intervention, while others aim to maximize it. Once SCBS is chosen, UEs are scheduled according to their QoS requirements using a scheduling mechanism (earliest deadline first (EDF), proportionally fair scheduling (PFS), Max-timely, and Min-Resources). This steering approach allows to reach 5× more gain than that obtained in a random distribution ([32]).

B. Multihoming at transport layer

In addition to the mobility described above and maintaining the IP connection of a user when offloading, static multihoming of a user connected simultaneously to multiple access networks has multiple IP addresses. However, regular TCP can support only one flow which mean only one IP address. For this reason, several transport protocols were proposed, we will present here an overview of multihoming-capable protocols.

Transport layer multihoming started with node multihoming which is an old concept defined as a device having more than one wired access interface. Two main standards were proposed: Stream Control Transport Protocol (SCTP) in 2000 [49, 50] and Multi-Path Transport Control Protocol (MPTCP) in 2010 [51]. SCTP uses only one path for transfer and keeps the other available paths for packet retransmission or for backup in case of handover or link failure. SCTP suffered however from the middleboxes blocking problem for SCTP packets. In this context, MPTCP was designed to answer the need for a concurrent multipath transfer function and a wide scale deployment by ensuring compatibility with lower and upper layers and by bypassing the middleboxes that may act at the transport level and may affect the TCP and MPTCP traffic. In addition, it targets better throughput and resilience by supporting the concurrent use of multiple paths seamlessly.

C. Frequency resources aggregation

Since the operator's first choice is to add more capacity on licensed spectrum, carrier aggregation (CA) technology [52] has been standardized in Long Term Evolution (LTE) Releases 10–12. CA was first proposed to aggregate

[3]SaMOG: S2a-Based Mobility on GTP, where GTP is a mobility solution proposed for 3GPP network using the S2a interface and adopting the Evolved Packet Core (EPC) of LTE architecture in the mobility process.

multiple small band segments into maximum 100 MHz virtual bandwidth to achieve higher data rate in LTE small cells.

Frequency multi-connection is also being standardized by 3GPP. LTE dual connectivity is introduced in Release 12 [15] as a realization of different spectrum allocation between a macro cell and a small cell. Several work items in Release 13 differentiated between dual connectivity in LTE/LTE-A HetNets, the License Assisted Access (LAA), and in LTE/WLAN HetNets, the LTE/WLAN Aggregation (LWA).

In LTE Dual Connectivity (DC), a user equipment (UE) maintains two downlink radio links, one to the macro cell and one to the small cell with control signaling sent only to the macro eNB. In other words, the UE can move under the coverage of the LTE macro cell without incurring any handover events. On the other hand, the uplink user plane of the UE is sent on either the macro cell link or the small cell link, whilst the downlink user plane has the additional option of being split and using both links (link aggregation). The downlink user plane bearer splitting occurs at the Packet Data Convergence Protocol (PDCP) protocol layer such that PDCP PDUs are sent either from the macro cell or forwarded over the X2 interface to the small cell. The small cell's eNB queues the PDCP PDUs and determines when to schedule their transmissions. Since PDCP PDUs may arrive out-of-sequence at the UE, the PDCP layer includes reordering functionality.

Several studies have been made in this field from different aspects in order to boost the standardization of DC mechanism in 5G networks. Information sharing and reporting issues were investigated in [53] with a focus on data split requirements. A novel method addressing the MeNBs-SeNBs pairing and the UEs grouping problem was proposed in [54]. Assignment problem with limited available reference signals was studied in [55], while flow control was analyzed in [56] and [57].

2.3.3 Interworking types

Several heterogeneous network types were considered in the literature. Heterogeneity in wired networks mainly consisted in accessing a server using more than one ISP, which means different routes. Generally, wired networks mutilhoming is considered as redundancy in case of failure. Few works tackled multihoming in such networks, we note [44] in which the authors conducted a study on multihoming streaming in a residential context using a DSL and a cable connection. This study showed significant QoS improvement for connection splitting and migration in case of congestion.

Conversely, wireless networks interworking gained a huge reputation. Several HetNet models were proposed along with performance evaluation and interworking technologies standardization. Next, we present two main categories for wireless networks interworking: (i) interworking between access networks with the same technology, mainly 3GPP, and (ii) interworking

between different wireless technologies with a focus on the interworing between 3GPP and WLAN networks.

A. Inter-3GPP interworking

Network densification using LTE small cells has been an important evolution direction in 3GPP, since LTE Release 10, to provide the necessary means to accommodate the anticipated huge traffic growth. Moreover, LTE small cells can be deployed both with macro coverage and standalone, indoor or outdoor, and can also be deployed sparsely or densely based on each case requirements. LTE interference coordination in such HetNets is widely studied and several radio coordination features are proposed. For example, we note downlink joint transmission, dynamic point blanking known as coordinated scheduling and enhanced inter-cell interference coordination (eICIC).

Joint access control and spectrum resource allocation is studied in [58] in a multi-access network composed of an LTE macro cell and several femto cells. Authors in [59] studied the eICIC in HetNets with co-channel deployment of small cells sharing the same licensed spectrum with macro cells and showed that they provide high-speed localized services.

Recently, different scenarios for LTE small cells deployment is studied in [60] under the dual connectivity feature in which detailed system-level simulations demonstrated how dual connectivity can improve end-user throughput and mobility performance. Authors in [53] explored dual connectivity technical challenges between LTE macro and small cells such as buffer status report, power headroom calculation and reporting, user power saving operations such as discontinuous reception, and increased device complexity to support bearer split in dual connectivity. They also showed that there are benefits of uplink bearer split in terms of increased per-user throughput between MeNB and SeNB at the expense of higher complexity in UE functionality. Authors in [61] compared load balancing and bearer splitting for the LTE DC architecture for indoor and outdoor scenarios and showed that both strategies achieve comparable throughput at the expense of additional complexity for bearer split.

B. Heterogeneous Interworking

The ability to exploit different access network technologies while providing a seamless subscriber experience has a clear appeal for all service providers and network operators. This is why interworking between HetNets was adopted. Several combinations of access networks were studied including, but not limited to, UMTS/WiMAX [62], WiFi/UMTS [63], WiFi/WiMAX [64] WiFi/HSDPA [38], WiFi/LTE [32, 36, 65, 66], and recently in 2017 mmWave/LTE [67]. However, not too many studies considered simultaneous multihoming. In the following, we present an overview of research works

concerning different cases of heterogeneous interworking.

A seamless handover approach between UMTS and WiMAX was proposed in [62] under a tight coupling architecture. This seamless handover was achieved by adding a sublayer on top of Layer 2 PDCP in UMTS and of MAC in WiMAX. The interworking between WiFi and UMTS was studied in [63] from the user equipment point of view in which they proposed a network selection scheme based on the battery power level in both normal and power saving modes. The interworking between WiFi and WiMAX was presented in [64]. The model consisted in WiFi accesses backhauled by WiMAX networks, the users are supposed to connect to the WiFi AP only and the main focus was on the interference management between WiFi APs. Multihoming was first studied in [38] between WiFi and HSDPA by using a strategy that finds the best packet splitting ratio between both accesses and showed service rate improvement.

Many studies addressed the interworking between WiFi and LTE. A tight integration between WiFi and LTE studied the co-location between WiFi and LTE small cells in [32], and proposed a cross-system traffic steering framework function of the traffic load and QoS requirements. Optimized cell-association and RAT assignment in such networks was studied in [66] and showed edge users throughput improvement by 1.8×. As an extension, [68] proposed a QoS-aware scheduling algorithm that optimizes the on-time throughput metric for tightly coupled LTE-WiFi small cells and showed a 3× improvement in the number of satisfied users. Authors in [36] considered a centralized radio resources scheduler that can communicate with both BS and APs offering two main functions: network selection and resource allocation, assuming that a user can access only one network at a time. From an architectural point of view, authors in [65] provided an overview on the LTE/WLAN interworking architecture, network selection, and security and mobility procedures. In addition, a quantitative study on the performance of indoor WiFi IEEE 802.11ac deployment is presented in [69] by using the offloading capabilities controlled by the user equipment itself. An analytical model of user throughput in IEEE 802.11ac is provided. A traffic steering method from LTE to IEEE 802.11ac is proposed based on SINR and throughput metrics aiming to ensure that users offloaded to WiFi acquire their minimum required data rate. The performance study showed that a minimum of 10 Mbps could be guaranteed in an interworking between IEEE 802.11ac and LTE. Very recently [67], mmwave offloading was proposed for interworking with LTE, which we assume in this thesis to be an important key technology in 5G networks. However, authors did not consider the case of multihoming and its benefit in such interworking.

2.3.4 Network selection decision

The network selection strategy in HetNets in the literature can be classified into three approaches: network centric, user centric, and hybrid decisions. We present here an overview for different research works in this domain and their contributions for network selection decision.

Network centric strategies generally propose a central scheduler managed by the operator. This central scheduler takes into consideration resource allocation between cell users. Several works addressed the interworking between HetNets using network centric scheduler, we note [66, 38, 36, 70]. Alternatively, user centric strategies delegate the traffic splitting or offloading to the users. For example, the user equipment might decide based on the battery power level combined with the consumption on each access network with preferring to offload on WiFi networks in the battery saving mode [63, 71]. Similarly, authors in [72] proposed an autonomous interface selection architecture for mobility management. The main idea was to choose an interface according to performance gain instead of throughput and to avoid frequent handover. Multihoming was considered from a user centric point of view in [73] where the user requests bandwidth share from available network without taking into consideration the system load information, which might be misleading sometimes.

A hybrid scheme for radio resource management consists in assisting wireless users decisions by broadcasting aggregate information about the network state. The operators' broadcasted policies aim to provide a better user experience as described in ANDSF [74]. In [75], authors proposed an association scheme that combines both centralized and decentralized approaches in a hybrid network composed of HSDPA and 3G LTE cells. The operator controls the UE decision through the load information broadcasted to users about each cell. An extension of this work is provided in [76] based on the interworking between WLAN and 3G LTE. The optimized distribution of users and their utilities is achieved through a strategy based on a Bayes-Nash equilibrium, where the operator can influence this equilibrium by broadcasting the channel quality indicator (CQI). Authors in [77] proposed a network selection algorithm based on the media independent handover concept (IEEE 802.21). This algorithm is split in two coordinating parts at the user equipment and the core network in order to decide the most suitable access network during a call establishment or a handover. A network-assisted user-centric WiFi offloading model was proposed in [33] by maximizing the per-user throughput. The heterogeneous network collects information of all users, calculates the optimized split ratio and broadcasts it to the users to use it when offloading.

2.4 RAN cloudification

Aiming to fill the blanks in the 5G's complete image, we introduce in this section the virtual radio access network (V-RAN). The rationale behind V-RANs starts with the emergence of cloud computing such as Amazon Web Services, Microsoft Azure and Google App Engine. In parallel, the rapid growth in mobile media applications and platforms was limited by energy and computational resources which imposed restrictions on the advancement of multimedia applications. That's why cloud computing was proposed as a support for mobile platforms by leveraging the heavy-computational services by executing them on the cloud. The mobile cloud computing [78] was considered as the intersection between mobile computing and cloud computing.

Cloud radio access networks (Cloud-RAN or C-RAN) architecture is considered as an innovation in HetNets. C-RAN allows scaling the mobile data network effectively under recent network challenges. C-RAN reduces both expenditures of mobile networks that are facing exponentially increasing data traffic demand [79] [80]. A logical evolution of C-RAN architecture is a V-RAN, a programmable architecture that is software definable and tuneable.

2.4.1 Macro cell

An LTE eNodeB is composed of one baseband unit (BBU) and up to three remote radio heads (RRHs) that can be connected. To connect the BBU and each RRH, an optical interface compliant with the common public radio interface (CPRI) specification, which is standard, is required (see Fig. 2.7).

The BBU is responsible for digital baseband signal processing. IP packets received from the core network are modulated into digital baseband signals and transmitted to the RRH. The digital baseband signals received from the RRH are demodulated and IP packets are transmitted to the core network. As for RRH, an RRH transmits and receives wireless signals. An RRH converts the digital baseband signals from BBU that are subject to protocol-specific processing into radio frequency signals and power amplifies them to transmit them to UE. On the contrary, the RF signals received from UE are amplified and converted into digital baseband signals for transmission to the BBU.

2.4.2 C-RAN/V-RAN

In C-RAN, the RRHs are located at the cell site and the BBU is implemented separately and performs centralized signal processing for the RAN. The decentralized BBU enables agility, faster delivery, cost savings and improved coordination of radio capabilities across a set of RRHs. A number of BBUs can be aggregated to form a pool of baseband units (BBU pool).

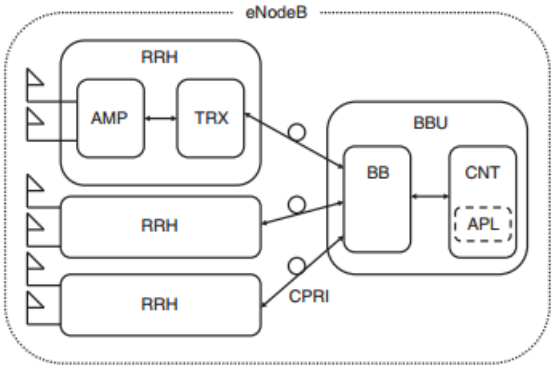


Figure 2.7: eNodeB hardware architecture (Source: [81]).

The V-RAN architecture virtualizes the BBU functionality and services in a centralized BBU pool (V-BBU) in the Central Office (CO) that can effectively manage on-demand resource allocation, mobility, and interference control for a large number of interfaces using programmable software layers. Complex RAN functions become easier also, we note: precoding, energy efficient allocation, enhanced inter-cell interference coordination [82], coordinated multi-point transmission [83] [84] and dual connectivity. V-RAN architecture enjoys software-defined capacity and scaling limits. It enables selective content caching, which helps to further reduce CAPEX and OPEX as well as improve user experience based on its cloud infrastructure.

In other words, V-RAN will open the door for many new applications in 5G. For example, it offers the possibility of using signal processing software dedicated to a special purpose based on the actual service. However, the realization of these benefits requires suitable strategies for an efficient usage of computing resources [85] [86], energy efficient resource allocation [87], sufficient fronthaul capacity [88] and effective BBU placement [89]. Authors in [90] studied heterogeneous C-RAN resource sharing at spectrum, infrastructure, and network levels.

A. Functional splitting

The C-RAN architecture can be divided into two types, based on the RRH and BBU functionalities: Full Centralization and Partial Centralization.

In full centralization, the functionalities of Layer 1, Layer 2, Layer 3 and signaling as well as operations and maintenance (O&M) are concentrated in the BBU, while RRH has only the radio functionalities as shown in Fig. 2.8. This provides optimum architecture for implementing network optimization techniques, however, it requires a large bandwidth and very low latency link to BBU hotel, to carry the baseband in-phase/quadrature (I/Q) signals.

Partial centralization's baseband processing functions (Layer 1) are lo-

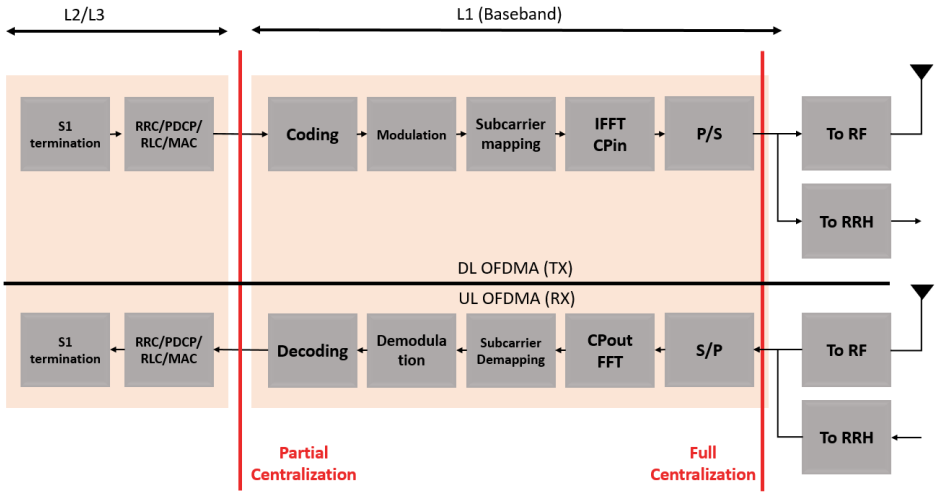


Figure 2.8: Functional splitting of full and partial centralization.

cated in the RRH along with radio functions (see Fig. 2.8). This configuration greatly reduces front-haul bandwidth requirements as compared to full centralization. In return, bringing baseband processing in the RRH level makes the upgrade and multi-cell collaborative signal processing less convenient [80].

B. Dynamic BBU pool virtualization

C-RAN architecture physically decouples the base station processing and radio units from each other. But to achieve the true benefits, the functional decoupling in C-RAN is of great importance. As the network load varies over time, this make the customary one-to-one mapping highly sub-optimal in some situations. The relationship between processing units and radio units should be dynamically optimized according to the network conditions which we call BBU pooling.

BBU pooling is the second evolution step. It brings the capability of load balancing between BBUs and cost effective redundancy for protection against BBU failure. BBU resources are then dimensioned from aggregated requirements of the cells covered by the pool. The main benefits of pooling are on CAPEX, enhanced scalability and failover mechanisms. In addition, BBU pooling includes LTE-A features for enhanced spectral efficiency and coverage fairness such as Coordinated Multipoint (CoMP) and eICIC.

In [91] the authors developed a load based adaptive architecture for small cell networks, in which small cells/RRH are dynamically connected to the BBU according to the load on each small cell and the overall system. This type of mapping not only optimizes RAN performance and saves the

computing resources but also provides an opportunity for energy saving as we will show later in this thesis.

2.5 Resource allocation strategies

Resource allocation and scheduling is defined as the act of assigning resources to a set of tasks. A set of constraints must be met by any scheduler such as deadline and minimum resource allocation. The decision and the scheduling problems address the feasibility of the scheduling. Resource scheduling started with the periodic scheduling in 1973 [92] by assigning zero or one resources at a time. Then another version allows sharing a resource or assigning more than one resource at a time. Among the proposed single-resource scheduling algorithms we note First In First Out (FIFO), Earliest Deadline algorithm (EDF) [92], Round Robin (RR), fair queuing (max-min fair scheduling), proportionally fair scheduling, and Scheduling optimization problems. However, none of these strategies extended directly to multiple resources. Radio resource allocation strategies focused either on rate-adaptive algorithms [93, 94] on fairness algorithms [95, 96] that try to avoid the starvation of some users.

In this thesis, we consider different variants of the 5G architecture described previously and evaluate users performance in each of these architectures. For this aim, we use different resource allocation strategies to compare system's stability and achievable users throughput. We also focus in this comparison on the fairness between different classes of users. In this section, we explore single and multi-resource allocation strategies we found in the literature.

2.5.1 Case of single type of resources

A fundamental step in the understanding of resource allocation mechanisms in the Internet has been the formulation by Kelly et al. [97] of congestion control in terms of network utility maximization. In a scenario of a fixed number of connections across different routes in a network, this approach characterizes an equilibrium and leads to the formulation of dynamic, distributed methods to achieve it. A unifying mathematical formulation to fair throughput assignment (which is called the "α-fairness") has been proposed in [98]; the degree of fairness is expressed by a parameter α defined on the whole half line $[0, \infty)$; it controls the trade-off between efficiency (total throughput maximization) and fairness. In particular, the case $\alpha \to \infty$ corresponds to max-min fairness (that can be considered to be the most fair allocation), the case $\alpha = 2$ corresponds to delay minimization, the case $\alpha \to 1$ corresponds to proportional fair assignment and the case $\alpha = 0$ corresponds to throughput maximization (that can be considered to be the most

efficient). α-fairness notion is commonly used to describe various network protocols (e.g., [99], [100]).

The user utility function is defined as $U(x)$ where x is the capacity share the user gets and the fair capacity sharing according to the utility criterion is defined as the solution of the maximization problem that considers the total utility of all users with $U(x) = \frac{x^{1-\alpha}}{1-\alpha}$. In this thesis, we focus on the case of proportional fairness (PF) allocation where $\alpha = 1$ and $U(x)$ is defined in the literature [97] by:

$$U(x) = logx \tag{2.1}$$

The goal of PF is to keep the proportionality among all service classes of users through resources allocation. PF also presents a trade-off between sum rate, i.e. efficiency, and link fairness in a heterogeneous network.

2.5.2 Case of multiple types of resources

Unlike conventional works on resource allocation which focus on single resource allocation [97, 101, 102], C-RAN introduces the need to study multi-resource allocation strategies, both at the RRH and BBU.

A study on multi-resource allocation in cloud computing systems is presented in [103]. The authors proposed so-called Dominant Resource Fair (DRF) allocation as a generalization of the max-min fairness to multiple resource types. DRF has been considered in [104] in the context of simultaneous fair allocation of multiple, continuously-divisible resources called Bottleneck-Based Fairness (BBF). The latter identifies the bottleneck resources and allocates resources such that each user receives all his required resources or gets at least his entitlement on some bottleneck resource. The work in [105] introduces Bottleneck Maximum Fairness (BMF), which is a simplified definition of BBF that ensures fairness on a bottleneck resource. A comparison between Proportional Fairness (PF), BMF and DRF showed the superiority of PF and BMF over DRF in offering better efficiency-fairness trade-off.

A Dominant Resource Fair Queuing (DRFQ) allocation was proposed in [106]. It generalizes the concept of virtual time from classical fair queuing to multiple resources that are consumed at different rates over time. Authors in [107] proposed an enhanced sharing of data center resource types such as Central Processing Unit (CPU), Random Access Memory (RAM) and disk storage. Dominant resource fairness was also studied in [108] in the presence of heterogeneous servers in a cloud computing system sharing computational resources such as processing, memory and storage. The proposed multi-resource allocation mechanism called DRF for Heterogeneous servers (DRFH) generalizes the allocation from single server in DRF to multiple heterogeneous servers.

However, when applied to our virtualized heterogeneous model these multi-resource allocation strategies do not offer same results of PF superiority over DRF because of multi-resources heterogeneity. To realize the ideal DRF allocation proposed by [103], authors in [105] stated that one can employ a water-filling algorithm. Water-filling consists on increasing the resource allocation at the same rate for all users until some resource is fully used, with repeating this process until all resources are fully used. In addition, following Ghodsi et al. [103] multi-resource allocations, including DRF, should satisfy the following properties: (i) sharing incentive, in that any allocation is assured at least a fraction of $1/n$ of all dominant resource, (ii) strategy-proofness, in which a user cannot improve her allocation by lying, (iii) envy-freeness, in which a user should not prefer the allocation of another user, and (iv) pareto efficiency, in that no user allocation can be increased without decreasing the allocation of at least another user.

2.6 5G and energy issues

The Information and Communications Technologies (ICT) account for a considerable portion of the total energy consumption. Statistics of 2012 tell that the annual average power consumption by ICT industries was over 200 GW, where telecommunication infrastructure and devices accounted for 25% [109]. Moreover, it is expected that in 5G era, millions more base stations with higher functionality and billions more devices with ever higher data rates will be connected [110]. Therefore, dramatic improvements of Energy Efficiency (EE) are required to ensure sustainable energy consumption in ICT [111].

Various efforts are done to cut down the energy consumption of telecommunication networks. The Energy Aware Radio and Network Technologies (EARTH) project sponsored by EU, has built a framework to support the EE evaluation over the large scale and long term, which is named the EARTH Energy Efficiency Evaluation Framework (E3F) [112]. E3F offers the power consumption breakdown for eNodeB components of LTE wireless system. Meanwhile, a flexible power model is built to support the E3F evaluation, which considers differentiation of BSs types. Furthermore, each type of BS is divided into a group of hardware components. The power of each hardware component is affected by several scaling factors, including bandwidth, antenna, modulation, coding rate, and load as presented in [86].

2.6.1 Energy consumption

A. Energy consumption in cellular networks

Energy consumption in cellular networks could be evaluated generally by considering the power consumed by all the components as well as the dy-

namic radio power used for transmission function of the load, or particularly by considering the power consumed by each allocated resources.

The consumed power at the base station follows the model provided by EARTH in [113] generalized to all BS types, including macro, micro, pico and femto BSs. Different transceiver (TRX) parts power consumption is analyzed:

- Antenna interface: The influence of the antenna type on the power efficiency is modeled by a certain amount of loss mainly at the feeder.
- Power amplifier (PA): The power consumption in PA suffers from non-linear effects which rises the poor power efficiency η_{PA}.
- Radio Frequency RF: The RF power consumption depends of the required bandwidth, the allowable signal-to-noise-and-distortion ratio, and the resolution of the analog-to-digital conversion.
- Baseband unit (BB): The BB unit power consumption includes the power consumed by functions such as filtering, modulation/demodulation, digital pre-distortion, signal detection, and channel coding/decoding.
- Power supply and cooling: The power supply and active cooling consumption is presented as a loss that scales linearly with the power consumption of other components.

B. Energy consumption in WiFi

The energy consumption in WiFi is less costly than cellular networks because of the reduced coverage and the lower number of users. The power consumption in this case depends of the AP's two states: Idle or Dynamic [114]. In a WiFi AP, the power consumption of PA, RF, BB, and power supply and cooling components are reduced or neglected.

C. Energy consumption in mm-wave

In a mm-wave small cell, the power consumption includes the baseband functions, the RF chains and the phase shifters. The other power consuming-part is the power amplifier (PA) which is the most power consuming part in a mm-wave access network. The power consumption in a mm-wave small cell depends of AP's state: Idle or Dynamic [115].

2.6.2 Energy efficiency maximization

EE and sustainability of 5G networks have recently received significant attention from mobile operators, vendors and research projects.

A large amount of work has been reported on EE resource allocation in mobile networks. An energy efficient analysis was provided for LTE HetNets

in [116] using realistic power models defined in the EARTH project. Mainly, energy saving techniques such as sleep mode were proposed for idle femto cells. In the same way, authors in [117, 118] proposed small cells activation for the offloading from macro cells to small cells as a strategy to increase power savings.

As for HetNets with multihoming, authors in [119] and [120] developed an uplink and downlink energy efficient allocation model for bandwidth and power resources in a heterogeneous wireless network. In the downlink case, they adopted a win-win strategy that achieves cooperation between different operators. Similar works on network resource allocation with multihoming are presented in [121, 122] with power consumption minimization.

To our knowledge, few works tackled the energy efficiency in V-RANs architecture with multihoming. Authors in [123] focused on the baseband unit role to decrease the fixed power consumption. They proposed a dynamic allocation of BBUs to RRH based on traffic conditions by switching them between ON and OFF. In addition, authors in [124] tried to decrease the backhaul power consumption by using caching as a solution. Authors in [125] proposed an EE maximization under average minimum data rate, maximum fronthaul capacity and maximum transmission power of BSs in a C-RAN model with mm-wave backhaul. However, this work did not target the multihoming problem, neither the heterogeneity of access networks. Considering a similar case with micro-wave backhaul, authors in [126] modeled the general power consumption in such networks.

Several works studied also the mm-wave energy efficiency. A comparison between mm-wave and 2 GHz system was provided in [127]. This study showed that mm-wave is more efficient than 2GHz system for high SINR value while for low SINR value 2GHz systems outperforms the mm-wave. Precoding energy efficient strategies were the main focus point for other works: authors in [128] studied the hybrid precoding subconnected architecture, and authors in [129] compared the fully-connected and the array-of-subarray architectures and found that the array-of-subarray has less power consumption.

2.7 Conclusion

We presented in this chapter a general overview of HetNets in 5G cellular networks. HetNets emerged as a promising low-cost approach for network densification. The interworking schemes range from load balancing, to offoading and multihoming; the latter being the focus of the present thesis. We described multihoming aspects and technology enablers available in 3GPP releases and those proposed by IETF. These technologies mainly include mobility protocols, transport layer's protocols, and dual connectivity mechanism in 5G. We reported on works on heterogeneous networks interworking,

highlighting different network selection strategies.

We also described V-RAN's architecture and defined BBU and RRH entities based on the different functional splitting types. We showed that BBU virtualization offers new efficiency and coverage enhancements by means of CoMP and eICIC. We reported on resource allocation works for both single type and multiple types of resources. We focused on proportional fairness and dominant resource fairness strategies for single resource and multi-resource allocations, respectively. We finally presented energy consumption aspects in different wireless networks, described power consuming parts and reported different energy efficiency works in the literature, for HetNets, C-RAN and multihoming.

In the next chapter, we present our work on the performance evaluation of multihoming in HetNets and compare network versus user centric resource allocation strategies.

Chapter 3

Network Centric versus User Centric Multihoming

3.1 Introduction

Facing the important increase of data traffic and the need for higher data rates, the fifth generation networks will be based on low cost, dense HetNets. HetNets pushed the operators to look for the best interworking techniques, standards and advanced features that aim to improve the overall system performance. In particular, network access in HetNets was previously proposed for offloading traffic or load balancing. Yet, simultaneous multiple access attracted operators with the increasing demands called "multihoming". Multihoming was standardized in Release 12 [15] under the name of "Dual Connectivity".

In this chapter, we model macro cell/small cell multihoming for both network centric and user centric approaches. Although several studies tackled network and user centric allocations, none of them considered the multihoming key feature. A central scheduler allocates resources in a network centric strategy as presented in [36, 38, 66, 70]. While in the user centric approach, the users decide by themselves the traffic splitting ratio or offloading to an access or another [63, 71, 72, 73]. None of these works however studied resource allocation in the presence of multihoming, and this is the object of our first contribution.

We specifically develop, in the following, analytical models to evaluate the performance of several resource allocation strategies while considering multihoming and compare them numerically versus a baseline strategy without multihoming. We consider both user centric and network centric allocation strategies. For the network centric strategies, we apply Proportional Fairness (PF) on both global and local levels to the system: global PF considers both access networks as a whole system whereas local PF considers each of the access networks individually, independently from the other one.

Although other works considered PF resource allocation ([130, 131, 132, 133] to state just a few), we model here the case of multihomed users in a heterogeneous system which was not considered in other works. We also extend our network centric study to evaluate the impact of opportunistic scheduling on the system's performance. Opportunistic scheduling in cellular networks was studied in many works (for instance [134], [135]) as well as opportunistic beamforming in [136], we consider it in this work in the context of a multihomed system.

User centric policies are performed at the application layer. We study two strategies: a simple one in which the multihomed user is aware of the peak rates of each network interface and splits its traffic proportionally between them, and a network assisted one in which the operator broadcasts information about the traffic intensities and the capacities of each system. The multihomed user uses this information so as to maximize its throughput.

3.2 System description

We consider the downlink of a multi-access wireless network composed of one LTE macro cell having in its coverage K WiFi small cells, as shown in Figure 4.1. We suppose that this system serves two types of users depending on their mobile equipment: single-homed and multihomed users. Single-homed users can have a single connection at a time depending on the quality of the received signal (LTE or WiFi) whereas multihomed users can benefit from multihoming when they are covered by both access networks by activating simultaneous connection mode.

As mentioned above, resources can be managed in a network centric or user centric manner.

In network centric, the allocation is performed jointly for both layers' resources: LTE and WiFi, as follows:

- A low bandwidth backhaul links the central scheduling node to the WiFi AP and the LTE eNodeB. The central scheduler operates at a large time scale - order of the flow dynamics - and determines the amount of resources to be allocated to each flow on each radio interface.

- Second level system schedulers, located within the WiFi AP and the LTE eNB, receive the first scheduler allocation output and use it to allocate effectively radio resources to users. We assume that these schedulers operate on a lower time scale and take into account the instantaneous fast fading variations in LTE.

In the user centric resource allocation, we study two strategies for traffic splitting:

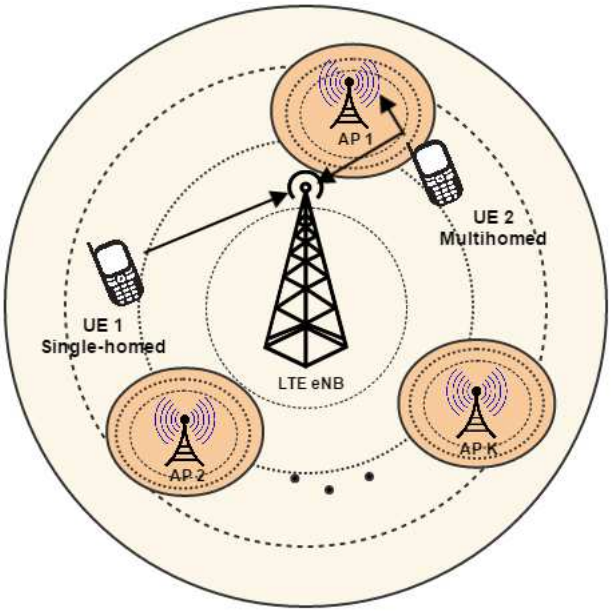


Figure 3.1: System model.

- Peak rate maximization: users do not receive any information about the system load. They split their traffic proportionally to the peak rate offered by each interface.
- Network assisted strategy: users use the load information broadcasted by the network in order to maximize their throughput performance.

3.3 Network centric resource allocations

We assume that flows are elastic, with exponentially distributed file size with mean σ, and that they arrive to the system following a Poisson process with global traffic intensity $\lambda = \lambda^M + \sum_k (\lambda_k^S + \lambda_k^{MH})$ where λ^M, λ_k^S and λ_k^{MH} are the mean traffic intensities for single-homed LTE macro cell users, single-homed small cell users and multihomed users, respectively, $k = 1, \cdots, K$.

Let the capacity shares resulting from different resource allocation strategies be denoted as follows:

- $x^M(\mathbf{n})$ and $x_k^{MH_M}(\mathbf{n})$ are resource capacity shares in the LTE macro cell for single-homed LTE users and multihomed users, having a connection with the k^{th} WiFi small cell, respectively.
- $x_k^S(\mathbf{n})$ and $x_k^{MH_S}(\mathbf{n})$ are resource capacity shares in the WiFi small cell numbered k for WiFi-only single-homed users and multihomed ones, respectively.

where $\mathbf{n} = (n^M, n_1^S, n_1^{MH}, \cdots, n_k^S, n_k^{MH})$ and where n^M, n_k^S and n_k^{MH}, with $k = 1, \cdots, K$, denote the number of users of each class: LTE only macro cell users, WiFi only small cell users, and multihomed users (both macro and small cells), respectively.

These shares depend on the allocation strategy, as we will show next for the two strategies: local PF wherein PF is applied to each access network separately from the other one, and global PF wherein it is applied to all accesses as a whole.

For simplicity, we first study the system with homogeneous radio conditions over the cells, with a cell capacity denoted by C^M for LTE macro cell and C^S for WiFi small cells. We then generalize our model to cover the case of multiple radio conditions.

3.3.1 Local Proportional Fairness

Generally, a user utility is defined as $U(x)$ where x is the capacity share the user gets. One example of utility is the so-called α-fair one, given by: $U(x) = \frac{x^{1-\alpha}}{1-\alpha}$. For the case of proportional fairness allocation, $\alpha = 1$ and $U(x)$ is defined in the literature [97] by $U(x) = \log x$.

In our case, local PF is obtained by maximizing the utility function for the macro cell and each small cell independently. For the macro cell, the utility function is defined as the sum of the logarithms of the flow rates obtained by single-homed and multihomed users served in the macro cell as follows:

maximize:

$$U_M = n^M \log(x^M C^M) + \sum_{k=1}^{K} n_k^{MH} \log(x_k^{MH_M} C^M) \tag{3.1}$$

subject to:

$$n^M x^M + \sum_{k=1}^{K} n_k^{MH} x_k^{MH_M} = 1 \tag{3.2}$$

$$x^M, x_k^{MH_M} \in [0, 1] \tag{3.3}$$

with $k = 1, \cdots, K$. Eq. (3.2) is the capacity constraint of the macro cell access network which ensures that the sum of capacity shares obtained by the users served in macro cell is equal to 1.

For the k^{th} small cell, the utility function is defined as the sum of the logarithms of the flow rates obtained by single-homed and multihomed users served by small cell k as follows:

maximize:

$$U_S^k = n_k^S \log(x_k^S C^S) + n_k^{MH} \log(x_k^{MH_S} C^S) \tag{3.4}$$

subject to:

$$n_k^S x_k^S + n_k^{MH} x_k^{MH_S} = 1 \tag{3.5}$$

$$x_k^S, x_k^{MH_S} \in [0, 1] \tag{3.6}$$

with $k = 1, \cdots, K$. Eq. (3.5) is the constraint that limits the sum of the users' capacity shares on small cell to 1.

We next formulate the Lagrangian function for each access network, composed of the utility function and its corresponding constraints, and obtain the capacity shares for each access network. For instance, the Lagrangian of the problem shown in Eq. (3.1) is given by:

$$L_M(x^M, x_1^{MH_M}, \cdots, x_K^{MH_M}, \nu_M) = n^M \log(x^M C^M) + \sum_k n_k^{MH} \log(x_k^{MH_M} C^M) - \nu_M(n^M x^M + \sum_k n_k^{MH} x_k^{MH_M} - 1) \tag{3.7}$$

with ν_M the Lagrangian multiplier. Then

$$\frac{\delta L_M}{\delta x^M} = \frac{n^M}{x^M} - n^M \nu_M = 0 \tag{3.8}$$

$$\frac{\delta L_M}{\delta x_k^{MH_M}} = \frac{n_k^{MH}}{x_k^{MH_M}} - n_k^{MH} \nu_M = 0 \tag{3.9}$$

By replacing x^M and $x_k^{MH_M}$ in Eq. (3.2), we find $\nu_M = n^M + \sum_k n_k^{MH}$ and so the unique solution to the primal problem is given by:

$$x^M = x_k^{MH_M} = 1/(n^M + \sum_{k=1}^{K} n_k^{MH}) \tag{3.10}$$

Similarly, by solving Eq. (3.4) under the constraint (3.5) we find:

$$x_k^S = x_k^{MH_S} = 1/(n_k^S + n_k^{MH}) \tag{3.11}$$

Note that, as fast fading is not taken into account, local PF is equivalent to round robin (RR) resource sharing: the round robin scheduling strategy serves users in a cyclic manner. In addition, this model for WiFi small cell throughput ignores the inefficiency of WiFi resource allocation where collisions lead to a decrease in the capacity when several users are served simultaneously. This gives an upper bound of the performance. We will consider this inefficiency in the case of heterogeneous radio conditions in section 3.5.

3.3.2 Global Proportional Fairness

Global PF is obtained by jointly maximizing the capacity shares ($x^M(\mathbf{n})$,$x_k^S(\mathbf{n})$, $x_k^{MH_M}(\mathbf{n})$, $x_k^{MH_S}(\mathbf{n})$) of macro cell users, k^{th} small cell users, k^{th} multihomed users on macro cell access and k^{th} multihomed users on small cell access, respectively.

This is obtained by jointly maximizing the utility function defined in Eq. (3.12) as the sum of the logarithms of the flow rates obtained by single-homed and multihomed users on each access network.
maximize:

$$\begin{aligned} U =& n^M \log(x^M C^M) + \sum_{k=1}^{K} n_k^{MH} \log(x_k^{MH_M} C^M + x_k^{MH_S} C^S) \\ &+ \sum_{k=1}^{K} n_k^S \log(x_k^S C^S) \end{aligned} \tag{3.12}$$

subject to:

$$n^M x^M + \sum_{k=1}^{K} n_k^{MH} x_k^{MH_M} = 1 \tag{3.13}$$

$$n_k^S x_k^S + n_k^{MH} x_k^{MH_S} = 1 \tag{3.14}$$

$$x^M, x_k^{MH_M} \in [0,1] \tag{3.15}$$

$$x_k^S, x_k^{MH_S} \in [0,1] \tag{3.16}$$

with $k = 1, \cdots, K$. Eqs. (3.13) and (3.14) are the constraints which ensure that the capacity shares on LTE and WiFi access networks, respectively, equal to 1.

The maximization problem is solved numerically. A closed-form expression for the capacity shares can however be obtained when the system is simplified to one macro cell and one small cell by deriving the Lagrangian function of the problem described in Eq. (3.12) under the constraints in Eqs. (3.13) and (3.14).

$$\begin{aligned} L =& n^M \log(x^M C^M) + n^{MH} \log(x^{MH_M} C^M + x^{MH_S} C^S) + n^S \log(x^S C^S) \\ &- \nu_M (n^M x^M + n^{MH} x^{MH_M} - 1) - \nu_S (n^S x^S + n^{MH} x^{MH_S} - 1) \end{aligned} \tag{3.17}$$

with ν_M and ν_S the Lagrangian multipliers. Then:

$$\frac{\delta L}{\delta x^M} = \frac{n^M}{x^M} - n^M \nu_M \tag{3.18}$$

$$\frac{\delta L}{\delta x^S} = \frac{n^S}{x^S} - n_S \nu^S \tag{3.19}$$

$$\frac{\delta L}{\delta x^{MH_M}} = \frac{n^{MH} C^M}{x^{MH_M} C^M + x^{MH_S} C^S} - n^{MH} \nu_M \tag{3.20}$$

$$\frac{\delta L}{\delta x^{MH_S}} = \frac{n^{MH} C^S}{x^{MH_M} C^M + x^{MH_S} C^S} - n^{MH} \nu_S \tag{3.21}$$

By replacing Eqs. (3.18) and (3.19) in Eqs. (3.13) and (3.14) for $K = 1$, we find x^{MH_M} and x^{MH_S} function of ν_M and ν_S, respectively. Then, we find the Lagrange multipliers by substituting x^{MH_M} and x^{MH_S} in Eq. (3.20)

$$\nu_M = \frac{C^M (n^M + n^S + n^{MH})}{C^M + C^S} \tag{3.22}$$

$$\nu_S = \frac{C^S (n^M + n^S + n^{MH})}{C^M + C^S} \tag{3.23}$$

The unique solution to the problem is given by:

$$x^M = \frac{C^M + C^S}{C^M (n^M + n^S + n^{MH})} \tag{3.24}$$

$$x^S = \frac{C^M + C^S}{C^S (n^M + n^S + n^{MH})} \tag{3.25}$$

$$x^{MH_M} = \frac{C^M n^S - C^S n^M + C^M n^{MH}}{C^M n^{MH} (n^M + n^S + n^{MH})} \tag{3.26}$$

$$x^{MH_S} = \frac{C^S n^M - C^M n^S + C^S n^{MH}}{C^S n^{MH} (n^M + n^S + n^{MH})} \tag{3.27}$$

where n^M, n^S and n^{MH} denote the number of LTE, WiFi and multihomed users, respectively.

3.3.3 Performance metrics

Since the operator decides the distribution of the resources instantaneously, we model the network centric system by a Markov Chain with state $\mathbf{n}$ defined, as described above, by: $\mathbf{n} = (n^M, n_1^S, n_1^{MH}, \cdots, n_k^S, n_k^{MH})$, where, again, n^M, n_k^S and n_k^{MH} denote macro cell single-homed users, small cell single-homed users and multihomed users (both macro and small cells), re-

spectively, and with transition rates equal to:

$$q(\mathbf{n}, \mathbf{n} + \mathbf{e}^M) = \lambda^M \tag{3.28}$$

$$q(\mathbf{n}, \mathbf{n} + \mathbf{e}_k^S) = \lambda_k^S \tag{3.29}$$

$$q(\mathbf{n}, \mathbf{n} + \mathbf{e}_k^{MH}) = \lambda_k^{MH} \tag{3.30}$$

$$q(\mathbf{n}, \mathbf{n} - \mathbf{e}^M) = n^M x^M \frac{C^M}{\sigma} \tag{3.31}$$

$$q(\mathbf{n}, \mathbf{n} - \mathbf{e}_k^S) = n_k^S x_k^S \frac{C^S}{\sigma} \tag{3.32}$$

$$q(\mathbf{n}, \mathbf{n} - \mathbf{e}_k^{MH}) = n_k^{MH} \frac{x_k^{MH_M} C_l + x_k^{MH_S} C^S}{\sigma} \tag{3.33}$$

where $\mathbf{e}^M$, $\mathbf{e}_k^S$ and $\mathbf{e}_k^{MH}$ denote the vector with one at the corresponding entry and zero elsewhere, for macro cell, small cell and multihomed users between macro cell and small cell k, respectively, with $k = 1, \cdots, K$.

The diagonal element is:

$$q(\mathbf{n}, \mathbf{n}) = -\sum_{\mathbf{n}_i \neq \mathbf{n}} q(\mathbf{n}, \mathbf{n}_i) \tag{3.34}$$

with $\mathbf{n}_i$ the row elements of the transition matrix except $\mathbf{n}$.

The steady state distribution is then obtained by solving:

$$\begin{cases} \Pi(\mathbf{n}).Q(\mathbf{n}) = 0 \\ \Pi(\mathbf{n}).\mathbf{e} = 1 \end{cases} \tag{3.35}$$

$\Pi(\mathbf{n})$ being the vector of the steady-state probabilities and $\mathbf{e}$ a vector of ones.

Once the vector $\Pi(\mathbf{n})$ is obtained, the global performance parameters can be calculated. The average number of users in each class of users is obtained by the formula:

$$N^M = \sum_{n^M \in \mathbf{n}} n^M \Pi(\mathbf{n}) \tag{3.36}$$

$$N_k^S = \sum_{n_k^S \in \mathbf{n}} n_k^S \Pi(\mathbf{n}) \tag{3.37}$$

$$N_k^{MH} = \sum_{n_k^{MH} \in \mathbf{n}} n_k^{MH} \Pi(\mathbf{n}) \tag{3.38}$$

A. Mean delay and average throughput

Using Little's formula, we deduce the mean delay of users of each class:

$$\delta^M = \frac{N^M}{\lambda^M}; \quad \delta_k^S = \frac{N_k^S}{\lambda_k^S}; \quad \delta_k^{MH} = \frac{N_k^{MH}}{\lambda_k^{MH}} \tag{3.39}$$

And finally, we deduce the average throughput obtained by users of each class:

$$D^M = \frac{\lambda^M \sigma}{N^M}; \quad D_k^S = \frac{\lambda_k^S \sigma}{N_k^S}; \quad D_k^{MH} = \frac{\lambda_k^{MH} \sigma}{N_k^{MH}} \tag{3.40}$$

with $k = 1, \cdots, K$.

B. System stability

We now turn to the system stability limits. For simplicity, we assume that all small cells have the same traffic arrival intensity for WiFi only small cell users and multihomed ones: i.e., $\lambda_k^S = \lambda^S$ and $\lambda_k^{MH} = \lambda^{MH}$ for all k. Given that $p^M = \frac{\lambda^M}{\lambda}$, $p^S = \frac{\lambda^S}{\lambda}$ and $p^{MH} = \frac{\lambda^{MH}}{\lambda}$, the sum $p^M + K(p^S + p^{MH}) = 1$.

At the stability limit, there are different cases for macro and small cells queues.

Macro cell queue A macro cell queue may serve only macro cell users with load ρ_{M1}:

$$\rho_{M1} = \frac{p^M \lambda \sigma}{C^M} \tag{3.41}$$

Nevertheless, a macro cell queue may also serve both macro cell only users and multihomed users at stability limits with load ρ_{M2} as follows:

$$\rho_{M2} = \frac{(p^M + K p^{MH}) \lambda \sigma}{C^M} \tag{3.42}$$

where ρ_{M1} and ρ_{M2} are the lower and upper bounds respectively on the macro cell queue load.

Small cell queue Likewise the macro cell queue, at stability limits, the small cell queue has two extreme values, it may serve only small cell users with load ρ_{S1} equal to:

$$\rho_{S1} = \frac{p^S \lambda \sigma}{C^S} \tag{3.43}$$

or may serve both small cell only users and multihomed users with load ρ_{S2}:

$$\rho_{S2} = \frac{(p^S + p^{MH}) \lambda \sigma}{C^S} \tag{3.44}$$

We take the realistic case when the macro cell queue is more loaded than the small cell queue without taking into consideration the multihomed users, i.e., $\rho_{M1} > \rho_{S1}$. Hence, two cases are possible:

1. For $\rho_{M1} > \rho_{S2}$, all multihomed users are served by the small cell access. Thus, the system stability corresponds to the maximum capacity for which all queues are stable , i.e., the most loaded $\rho_{M1} \leq 1$, which gives:
$$\lambda \leq \frac{C^M}{p^M \sigma}$$
.

2. For $\rho_{M1} < \rho_{S2}$, multihomed users cannot be served completely by small cell access because, at stability limits, it becomes much loaded than the macro cell with $\frac{p^{MH}}{C^S} > \frac{p^M}{C^M} - \frac{p^S}{C^S} > 0$. In this case, at stability limits, multihomed users are served by both queues with equalizing their load. A proportion p^{MH_S} of multihomed users traffic will be served by the small cell and the remaining $(1 - p^{MH_S})$ will be served by the macro cell , where p^{MH_S} is the solution of:
$$\frac{p^M + Kp^{MH}(1 - p^{MH_S})}{C^M} = \frac{p^S + p^{MH}p^{MH_S}}{C^S}$$
Based on this equalization, the stability limits correspond to $\rho_{M1} \leq 1$ or $\rho_{S2} \leq 1$ as follows:
$$\lambda \leq \frac{C^S}{(p^S + p^{MH}p^{MH_S})\sigma}$$

3.4 User centric resources allocation

We now turn to the user centric approach and evaluate the performance of two strategies: peak rate maximization and network assisted strategies. Our aim is to find the optimal value of the proportion β of a file a multihomed user receives on the LTE macro cell; the remaining file proportion, $(1 - \beta)$, is received on the WiFi small cell. Obviously, β depends on the applied strategy.

For simplicity, we assume that all WiFi small cells have the same traffic arrival intensity for WiFi only users and multihomed ones: i.e., $\lambda_k^S = \lambda^S$ and $\lambda_k^{MH} = \lambda^{MH}$ for all k.

3.4.1 Peak rate maximization

As already mentioned, peak rate maximization strategy supposes that multihomed users have only information on the throughput of each system: C^M and C_k^S and use this information to find β, the file split ratio between LTE and WiFi:
$$\beta = \frac{C^M}{C^M + C^S}$$

3.4.2 Network assisted policy

In the network assisted policy, the network operator broadcasts detailed information about the offered traffic in each class, for example the arrival rates λ. Multihomed users aim to maximize their throughput given by:

$$D_k^{MH}(\beta) = \min(\frac{D^M(\beta)}{\beta}, \frac{D_k^S(\beta)}{1-\beta}) \tag{3.45}$$

where $D^M(\beta)$ and $D_k^S(\beta)$ are the average user throughputs for single homed macro and small cell users, respectively, and are given by:

$$D^M(\beta) = C^M - \lambda^M\sigma - \sum_{k=1}^{K}\lambda_k^{MH}\sigma\beta \tag{3.46}$$

$$D_k^S(\beta) = C_k^S - \lambda_k^S\sigma - \lambda_k^{MH}\sigma(1-\beta) \tag{3.47}$$

We define the overall average user throughput by taking into account all classes of users as:

$$D(\beta) = \frac{\lambda^M}{\lambda}D^M(\beta) + \sum_{i=1}^{K}(\frac{\lambda_k^S}{\lambda}D_k^S(\beta) + \frac{\lambda_k^{MH}}{\lambda}D_k^{MH}(\beta)) \tag{3.48}$$

The multihomed users' objective is to find the optimal traffic split that maximizes their throughput given by Eqn. (3.45).

We get the following solution for $k > 1$:

$$\beta^* = \frac{\delta - \sqrt{\delta^2 - 4(K-1)\lambda^{MH}\sigma(C^M - \lambda^M\sigma)}}{2(K-1)\lambda^{MH}\sigma} \tag{3.49}$$

with:

$$\delta = C^M - \lambda^M\sigma + C^S - \lambda^S\sigma + (K-1)\lambda^{MH}\sigma \tag{3.50}$$

For the case of one LTE macro cell and one WiFi small cell (i.e. $K = 1$):

$$\beta^* = \arg\max\min(\frac{C^M - \lambda^M\sigma - \lambda^{MH}\sigma\beta}{\beta}, \frac{C^S - \lambda^S\sigma - \lambda^{MH}\sigma(1-\beta)}{1-\beta}) \tag{3.51}$$

If $\frac{C^M - \lambda^M\sigma - \lambda^{MH}\sigma\beta}{\beta} < \frac{C^S - \lambda^S\sigma - \lambda^{MH}\sigma(1-\beta)}{1-\beta}$: $\beta \in \min(\beta_0, 1)$ with

$$\beta_0 = \frac{C^M - \lambda^M\sigma}{C^M - \lambda^M\sigma + C^S - \lambda^S\sigma} \tag{3.52}$$

which gives that $\beta = \beta_0$.

If $\frac{C^M-\lambda^M\sigma-\lambda^{MH}\sigma\beta}{\beta} > \frac{C^S-\lambda^S\sigma-\lambda^{MH}\sigma(1-\beta)}{1-\beta}$: $\beta \in \max(0,\beta_0)$. In this case, the solution is $\beta = \beta_0$.

This yields for both cases the optimal split: $\beta^* = \beta_0$.

$$\beta^* = \frac{C^M - \lambda^M\sigma}{C^M - \lambda^M\sigma + C^S - \lambda^S\sigma} \tag{3.53}$$

Moreover, we have the following result:

Theorem 1. *The selfish policy where multihomed users maximize their own throughput corresponds to a global optimum for the average throughput of all users over the cell, i.e., maximizes D in equation (3.48).*

Proof. It is sufficient for this to verify that β^* in equation (3.49) optimizes equation (3.48). See Appendix A for the proof. □

This correspondence between selfish optimum and global optimum can be interpreted as follows: when the network operator maximizes multihomed users' throughput, it offers multihomed users the opportunity to leave the system faster, and thus free resources for single-homed users.

3.4.3 Performance metrics

In the user centric approach, the traffic split decision is independent of the instantaneous state of the system which can be modeled by a set of $K+1$ independent queues. These queues are one macro cell and K small cell queues, all Processor Sharing (PS).

For peak rate assisted strategy, the stability limits correspond to the total traffic arrival rate that keeps both queues stable. Given that the split ratio is $\beta = \frac{C^M}{C^M+C^S}$, the upper bound for λ is:

$$\lambda < \min\{\frac{C^M}{(p^M + \sum_k p_k^{MH}\beta)\sigma}, \frac{C^S}{(p_k^S + p_k^{MH}(1-\beta))\sigma}\}, \qquad \forall k \in [1,K]$$

For the network assisted strategy, we assume that the macro cell LTE queue becomes unstable before the other K small cell WiFi queues do. Given that $\lambda^M \ggg \lambda_k^{MH}$ at stability limits of the LTE queue, the traffic split of multihomed users on the LTE access $\beta \to 0$ and the load of the LTE macro cell is:

$$\rho^M = \frac{\lambda^M\sigma}{C^M} \tag{3.54}$$

After the saturation of LTE queue, all multihomed users join completely the WiFi queue and the k^{th} WiFi small cell stability limits is equal to:

$$\rho_k^S = \frac{\lambda_k^S\sigma}{C^S} + \frac{\lambda_k^{MH}\sigma}{C^S} \tag{3.55}$$

The system stability corresponds to the stability of both LTE and WiFi systems by finding the traffic arrival that maintains $\rho^M \leq 1$ and $\rho_k^S \leq 1$ determined by the following inequality:

$$\lambda < \min\{\frac{C^M}{p^M\sigma}, \frac{C^S}{(p^S + p^{MH})\sigma}\}, \qquad \forall k \in [1, K]$$

3.5 Heterogeneous radio conditions

Finally, we consider the realistic case where users at different positions in the cell receive different signal strengths from the network. The objective here is to find the optimal multihoming policies when different radio conditions are experienced by multihomed users. A typical example is one with two radio conditions for multihomed users: indoor, with a better WiFi signal than the LTE one, and outdoor, with a better LTE signal than the WiFi one.

We consider the following radio conditions' distribution:

- N radio conditions for the LTE macro cell users, each with weight p_n^M computed as the number of flow arrivals from this class over the total LTE macro cell flow arrivals. Among these radio conditions, $1 \rightarrow KJ$ correspond to multihomed users, and $KJ+1 \rightarrow N$ correspond to LTE users. We denote by C_n^M the throughput observed in the LTE macro cell by a user of class $n \in [1, N]$.

- I radio conditions in each WiFi small cell, each with weight p_i^S. We also rearrange the radio conditions for WiFi so that the first J radio conditions $1 \rightarrow J$ correspond to multihomed users, and $J + 1 \rightarrow I$ correspond to WiFi users. We denote by C_i^S the throughput observed in the WiFi small cell by a user of class $i \in [1, I]$.

Without loss of generality, we focus on the global PF strategy for the network centric approach and on the network assisted strategy for the user centric approach.

3.5.1 Network centric approach

We start with the heterogeneous radio conditions applied to the global PF strategy. Let $\hat{n}^M$ and $\hat{n}^S$ be the number of users in LTE macro cell and WiFi small cell queues, respectively, including multihomed users. For ease of expression, we denote the number of multihomed users by $\hat{n}^{MH}$:

$$\hat{n}^{MH} = \hat{n}^M p_j^M = \hat{n}^S p_j^S \tag{3.56}$$

for $j = 1, \cdots, J$.

The optimization problem to solve becomes:
maximize:

$$\begin{aligned} U =& \hat{n}^M \sum_{n=KJ+1}^{N} p_n^M \log(x_n^M C_n^M) + \\ & K\hat{n}^S \sum_{i=J+1}^{I} p_i^S \log(x_i^S C_i^S) + \\ & K(\sum_{j=1}^{J} n^{\hat{M}H} \log(x_j^{MH_M} C_j^M + x_j^{MH_S} C_j^S) \end{aligned} \tag{3.57}$$

subject to:

$$\hat{n}^M \sum_{n=KJ+1}^{N} p_n^M x_n^M + K\hat{n}^M \sum_{j=1}^{J} p_j^M x_j^{MH_M} = 1 \tag{3.58}$$

$$\hat{n}^S \sum_{i=J+1}^{I} p_i^S x_i^S + \hat{n}^S \sum_{j=1}^{J} p_j^S x_j^{MH_S} = 1 \tag{3.59}$$

The capacity share values are obtained by solving numerically the problem in equation (3.57), subject to constraints (3.58) and (3.59).

3.5.2 User centric approach

In this section, we study the heterogeneous radio conditions influence on users' performance when applying the network assisted strategy.

Let $\hat{\lambda}^M$ and $\hat{\lambda}^S$ be the flow arrival intensities in LTE and WiFi queues, respectively, including multihomed users. When these multihomed users choose a policy expressed by $\beta = (\beta_1, ..., \beta_J)$, the cell loads become equal to:

$$\rho^M(\beta) = \hat{\lambda}^M \sigma\big[K \sum_{j=1}^{J} \frac{p_j^M \beta_j}{C_j^M} + \sum_{n=KJ+1}^{N} \frac{p_n^M}{C_n^M}\big] \tag{3.60}$$

$$\rho^S(\beta) = \hat{\lambda}^S \sigma\big[\sum_{j=1}^{J} \frac{p_j^S (1-\beta_j)}{C_j^S} + \sum_{i=J+1}^{I} \frac{p_i^S}{C_i^S}\big] \tag{3.61}$$

In this case, the system is modeled as $K+1$ independent PS queues with cell capacities given by:

$$\hat{C}^M(\beta) = \big[K \sum_{j=1}^{J} \frac{p_j^M \beta_j}{C_j^M} + \sum_{n=KJ+1}^{N} \frac{p_n^M}{C_n^M}\big]^{-1} \tag{3.62}$$

$$\hat{C}^S(\beta) = \big[\sum_{j=1}^{J} \frac{p_j^S (1-\beta_j)}{C_j^S} + \sum_{i=J+1}^{I} \frac{p_i^S}{C_i^S}\big]^{-1} \tag{3.63}$$

In this context, the average throughput of a multihomed user that belongs to class $h \in [1, J]$ is given by:

$$D_h^{MH}(\beta) = \min\{\frac{C_h^M}{\beta_h}(1 - \hat{\lambda}^M \sigma(K \sum_{j \neq h} \frac{p_j^M \beta_j}{C_j^M} + \sum_{n > KJ} \frac{p_n^M}{C_n^M})) - K\hat{\lambda}^M \sigma p_h^M; \frac{C_h^S}{1-\beta_h}(1 - \hat{\lambda}^S \sigma(\sum_{j \neq h} \frac{p_j^S(1-\beta_j)}{C_j^S} + \sum_{i > J} \frac{p_i^S}{C_i^S})) - \hat{\lambda}^S \sigma p_h^S\} \quad (3.64)$$

Let β_h^* be the optimal policy that maximizes the throughput of multihomed users of class h, given the policy of other classes $j \neq i$:

$$\beta_h^* = \arg\max_{\beta_h} \{D_h^{MH}(\beta)\} \quad (3.65)$$

and let $\beta/\mathrm{h} = (\beta_1, ..., \beta_{h-1}, \beta_{h+1}, ..., \beta_J)$ be the policies chosen by all multihomed users classes except h. The solution of the maximization in equation (3.65) is given in Eq. (3.66). For more details see Appendix B.

$$\beta_h^*(\beta/\mathrm{h}) = \frac{\delta - \sqrt{\delta^2 - 4\lambda_h^{MH} \sigma C_h^M (1 - \rho^M(\beta/\mathrm{h}))}}{2\lambda_h^{MH} \sigma} \quad (3.66)$$

with:

$$\delta = C_h^M(1 - \rho^M(\beta/\mathrm{h})) + C_h^S(1 - \rho^S(\beta/\mathrm{h})) + \lambda_h^{MH}\sigma \quad (3.67)$$

$$\rho^M(\beta/\mathrm{h}) = \hat{\lambda}^M \sigma(K \sum_{j \neq h} \frac{p_j^M \beta_j}{C_j^M} + \sum_{n > KJ} \frac{p_n^M}{C_n^M}) \quad (3.68)$$

$$\rho^S(\beta/\mathrm{h}) = \hat{\lambda}^S \sigma(\sum_{j \neq h} \frac{p_j^S(1-\beta_j)}{C_j^S} + \sum_{i > J} \frac{p_i^S}{C_i^S}) \quad (3.69)$$

and

$$\lambda_h^{MH} = K\hat{\lambda}^M p_h^M - \hat{\lambda}^S p_h^S \quad (3.70)$$

Consequently, if all users apply this described strategy, the solution of the set of the J fixed point Eq. (3.66) yields the resulting policy under the sufficient stability condition. This is expressed by replacing β/h by $\beta^*/\mathrm{h} = (\beta_1^*, ..., \beta_{h-1}^*, \beta_{h+1}^*, ..., \beta_J^*)$.

Definition 1. *The sufficient stability condition is the traffic region that maintains the stability of both systems (macro and small cells) for all multihoming policies:*

$$\hat{\lambda}^M \leq \frac{1}{\sigma(K\sum_{j=1}^{J}\frac{p_j^M}{C_j^M}+\sum_{n>KJ}\frac{p_n^M}{C_n^M})} \tag{3.71}$$

$$\hat{\lambda}^S \leq \frac{1}{\sigma(\sum_{j=1}^{J}\frac{p_j^S}{C_j^S}+\sum_{i>J}\frac{p_i^S}{C_i^S})} \tag{3.72}$$

Theorem 2. *The fixed point solution of (3.66) exists and corresponds to a Nash equilibrium under the sufficient stability condition.*

Theorem 2 is proven in Appendix C.

In other words, the Nash equilibrium exists for traffic loads before the stability point. Near this latter, some policies may cause system instability, preventing thus from proper convergence.

3.6 Simulation and numerical results

In this section, we present simulation parameters used to evaluate the performance of the proposed resource allocation strategies. For this aim, we consider a network where each LTE eNodeB coexists with $K = 3$ WiFi APs (IEEE 802.11n). For more realistic radio conditions, we consider a dense LTE network deployed in a large European city with an average radius of 350 meters and an LTE frequency band equal to 1800 MHz. Based on the Channel Quality Indicator (CQI) measurements and the resource block (RB) throughputs, authors in [137] found the average LTE macro cell capacity to be equal to $C^M = 30.5$ Mbps .

In addition, we consider a typical WiFi small cell deployment of 70 meters radius and 15 Mbps average capacity. Assuming a homogeneous distribution of users over the covered system area, we define the arrival rate of users for each access network as shown in Table 4.1. We consider that 50% of users equipment are multihoming capable, and that they can connect simultaneously to LTE and WiFi whenever possible which implies that only multihoming capable users that are covered by WiFi and LTE benefit the multihoming, which equals 15% of multihomed users on each WiFi access network.

Our results are plotted as a function of the global offered traffic $\lambda\sigma$ in Mbps, with $\sigma = 5$ Mbits the average flow size. More details on output performance parameters are given in Table 3.2.

User centric performances are evaluated by using the closed form equations of throughput. However, network centric ones are evaluated by a Monte Carlo simulation of the Markov Chain in MATLAB with maximizing the utility function of capacity shares between users flows.

Table 3.1:
Simulation parameters

Parameter	**Value**
LTE macro cell radius r_{MC}	350 m
LTE average rate	30.5 Mbps
Number of WiFi small cells (k)	3
WiFi small cell radius r_{SC}	70 m
WiFi average rate	15 Mbps
Average flow size	5 Mbits
Global traffic intensity λ	$0 \cdots 35[1/s]$
LTE users traffic intensity	$55\%\lambda$
k^{th} AP WiFi users traffic intensity	$7.5\%\lambda$
Multihomed users connected to LTE macro cell andk^{th} WiFi small cell traffic intensity	$7.5\%\lambda$

Table 3.2: Output parameters

Output Parameter	**Definition**
D^M	Throughput of macro cell users
D_k^S	Throughput of k^{th} small cell users
D_k^{MH}	Throughput of multihomed users on macro cell and k^{th} small cell
D_{ql}	Throughput of macro cell access network including macro cell only users and multihomed users
D_{qw}	Throughput of k^{th} small cell access network including small cell only and multihomed users served on this access
x axis	Offered traffic equal to the total arrival rate in the whole system multihplied by the average flow size σ

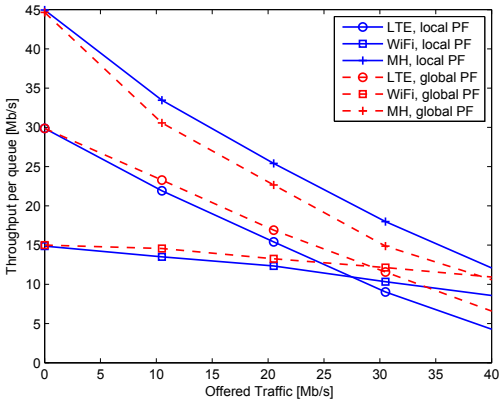


Figure 3.2: Impact of network centric scheduling strategies on users' performance.

3.6.1 Network centric approach

A. Achievable throughput

We present in Figure 3.2 the achievable throughput for all classes of users as a function of the offered traffic, both for local and global PF network centric resource allocation strategies. With the chosen parameters, multihomed users will be able to achieve higher throughput with local PF compared to that achieved by applying global PF. Whereas single-homed users achieve a better performance with the global PF strategy. This can be explained by the fact that local PF applied on a wireless interface ignores other interfaces' allocation and thus do an over-provisioning for multihomed users contrary to global PF that takes into consideration the resource allocation on all interfaces and achieves a better fairness between single-homed and multihomed users. This increase of fairness between users when applying global PF strategy is explained by the decrease in the difference between user's throughput.

We also notice that, at low loads, multihomed users achieve higher throughput than single-home ones, and which is equal to the aggregation of both systems' capacities, then their throughput decreases when traffic increases.

It is worth to note that based on the chosen parameters of traffic intensity on each access network, LTE access network receives higher proportion of user arrivals than WiFi. This difference is explained by the fast decrease of achievable throughput for LTE users versus a slower decrease of WiFi users' throughput. With the increase of traffic intensity, the achievable throughput by LTE becomes less than that achievable by WiFi.

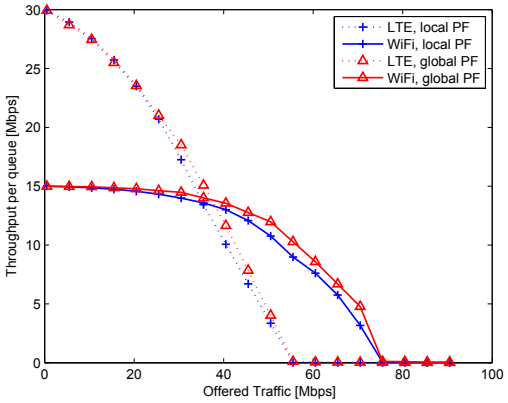


Figure 3.3: User performance and system stability for network centric strategies.

B. System stability

Under the assumption of an infinite queue with no blocking, we observe that both strategies local PF and global PF achieve the same stability point 55 Mbps of offered traffic (which is equal to the traffic intensity when the average achievable throughput reaches zero) as shown in Figure 3.3. This figure shows the stability region for each system (LTE and Wi-Fi) by plotting their average throughput as a function of the offered traffic, both for local and global PF strategies. However, the difference of stability between both queues is noticeable: the LTE system becomes unstable at 55 Mbps before the WiFi system does at 75 Mbps. This can be explained by the fact that multihomed users distribution corresponds to users distribution in the system as we assume that 50% of users are multihoming capable users. This gives that only 15% of multihomed traffic can be served on a WiFi AP due to its small coverage. Note that the stability point is when both LTE and WiFi systems are stable. In order to find this stability point, we intentionally ignore the inefficiency of WiFi resource allocation and consider only the upper bound of users' performance.

Furthermore, we note that the sum of the throughputs per queue can be higher or lower than the offered traffic. Keeping in mind that this achievable throughput per queue is inversely proportional to the traffic arrival intensity, i.e., the requested offered traffic, a low traffic intensity allows the users to receive higher throughput than their target because they do not share (or they share with a small number of users) the network resources and can leave the system very fast. However, a high traffic intensity causes the high number of users to share the network resources. This share is shown in Figure 3.3 by a lower throughput than the target which causes the users to spend more time before leaving the system. If the traffic intensity (i.e.,

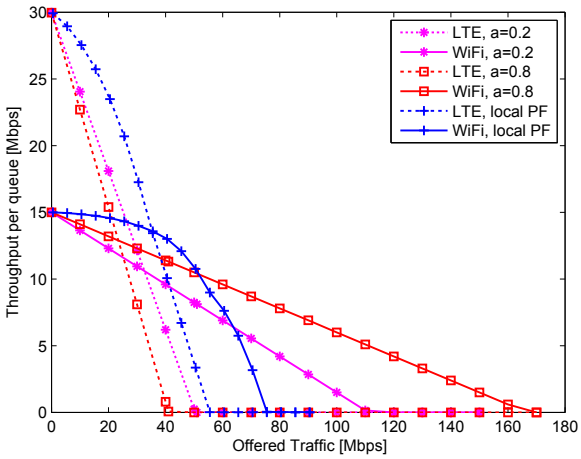


Figure 3.4: Throughput variation for each wireless access network as a function of offered traffic, comparison between local PF and reference model for $a = 0.2$ and 0.8.

offered traffic) increases, the system reaches the saturation point.

C. Benefit of multihoming

In order to emphasize the benefit of multihoming, we compare the local PF strategy with a baseline one without multihoming. In this baseline strategy, the multihoming capable users (50% of devices in our numerical examples) are served either by LTE or WiFi because multihoming is not considered. We denote by a the proportion of multihoming capable users served by LTE instead of the proportion of a file a multihomed user receives on LTE. Consequently, $(1-a)$ denotes the proportion of those users served by WiFi. Based on this pre-known distribution, the stability conditions become:

$$\rho^M + \sum_{k=1}^{K} \rho_k^{MH_M} = \frac{\lambda^M \sigma}{C^M} + \sum_{k=1}^{K} \frac{a\lambda_k^{MH}\sigma}{C^M} < 1$$

$$\rho_k^S + \rho_k^{MH_S} = \frac{\lambda_k^S \sigma}{C^S} + \frac{(1-a)\lambda_k^{MH}\sigma}{C^S} < 1$$

Figure 3.4 shows a comparison between the network with multihoming and the baseline case without multihoming. For the baseline case, we consider two values of a; this choice of a has a large impact on the network performance.

We notice that for the LTE system, local PF achieves higher throughput and higher stability. Whereas for WiFi, local PF brings higher throughput but lower stability. This is due to the fact that local PF achieves a proportional distribution of resources between both access types by changing the

multihomed users capacity share on each access. However, since the baseline case does not allow multihoming and WiFi is less loaded than LTE in local PF, we obtain this difference of stability between both strategies. It is worth noting that global stability is enhanced with multihoming.

D. Impact of fast fading

We now consider the second level scheduler for the LTE macro cell. A channel aware PF scheduler operates at the fast fading time scale, as assumed in [135]. This fast fading is modeled as a gain G added to the LTE system as a function of the number of flows in progress: when the LTE macro cell has n^M flows, and an optimized capacity share x^M, the obtained throughput by LTE flows is equal to $x^M C^M G(n^M)$, where $G(n^M)$ corresponds to the opportunistic scheduling gain that is a function of several parameters such as the number of flows in the channel, the channel model, the receiver type and the Multiple Input Multiple Output (MIMO) scheme [134].

The system can still be modeled as a Markov Chain with state dependent departure rates due to the opportunistic scheduling gain. The transition rates described in Eqs. (3.31) and (3.33) for the LTE macro cell users in the Markov chain become:

$$q(\mathbf{n}, \mathbf{n} - \mathbf{e}^M) = n^M x^M \frac{C^M G(n^M)}{\sigma} \tag{3.73}$$

$$q(\mathbf{n}, \mathbf{n} - \mathbf{e}_k^{MH}) = n_k^{MH} \frac{x_k^{MH_M} C^M G(n^M) + x_k^{MH_S} C^S}{\sigma} \tag{3.74}$$

This gain is applied to LTE users only, including multihomed ones. The achievable throughput for a WiFi user remains as described in Eq. (3.32). We consider the scheduling gain calculated in [135] for a MIMO 2×2 LTE system and an Additive White Gaussian Noise (AWGN) channel and that converges to $G(\infty) = 1.7$ starting from a number of active users in the LTE cell equal to 19.

The numerical simulations in Figure 3.5 show that the overall performance of the system is improved when opportunistic scheduling is considered. We see clearly that opportunistic scheduling alleviates the load of LTE and extends the benefit of multihoming to higher traffic loads.

3.6.2 User centric approach

Now, we compare the performance of user centric resource allocation strategies: peak rate maximization and network assisted.

As we can see in Figure 3.6 the network assisted strategy enables multihomed and LTE-only users to achieve higher throughput, especially at higher traffic loads. Yet, WiFi-only users have a slightly lower throughput with the network assisted strategy than that obtained with the peak rate

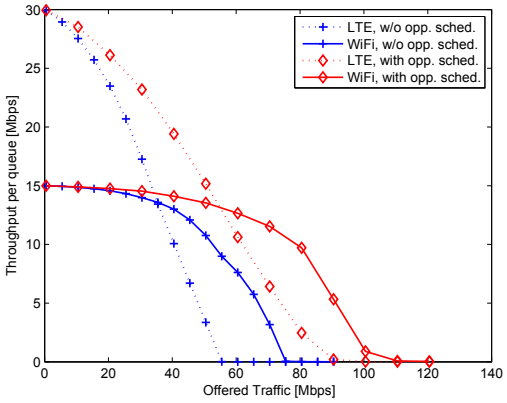


Figure 3.5: Impact of opportunistic scheduling on performance.

maximization. This difference of performance relates to the resource allocation strategy: while peak rate maximization uses throughput perceived by the users for each access type, network assisted strategy is more robust and takes into consideration the traffic intensity in each system. Consequently, we observe that multihomed users achieve higher throughput with the network assisted strategy when compared with single-homed users (LTE-only and WiFi-only). This is also the case for the peak rate maximization strategy but not at high traffic loads because this strategy does not take into consideration system's traffic intensity.

Furthermore, we show in Figure 3.7 that the LTE macro cell reaches instability before WiFi small cell does, and this is because of the difference between coverage and hence the fact that LTE serves more users than WiFi. In fact, the network assisted strategy brings larger stability for LTE, but it does the opposite for WiFi as more multihomed users are served on WiFi. This is the optimal distribution of load from a user point of view. This strategy improves the whole stability of the system by balancing multihomed users between the two sub-systems.

3.6.3 Comparison with network centric allocation strategy

Another important evaluation is the comparison between user centric (network assisted) and network centric (global PF) strategies shown in Figure 3.8. We observe that LTE and WiFi queues achieve a better performance with global PF strategy whereas WiFi-only users' performance becomes slightly lower than that achieved with the network assisted strategy. The difference of performance relates to the precision of the allocation strategy: global PF performs instantaneous resource allocation, while network assisted uses average values for resource allocation decision. In contrast, the advantage of network assisted strategy is the computational simplicity

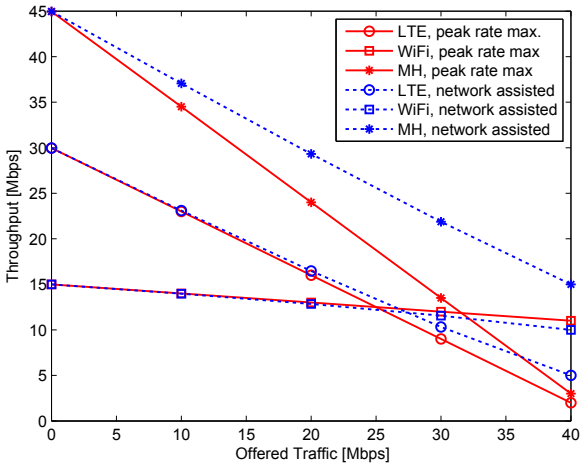


Figure 3.6: Impact of user centric scheduling strategies on users' performance.

when compared with the complexity of global PF. However, both strategies achieve the same stability points for LTE and WiFi queues because the stability point is independent of the resource allocation strategy.

3.6.4 Case of heterogeneous radio conditions

In order to evaluate the impact of the heterogeneous radio conditions on multihomed users' achievable throughput, we apply now the network assisted strategy. We consider two types of multihomed users' radio conditions: indoor and outdoor. Recall that indoor users receive a high WiFi signal and a low LTE signal, whereas outdoor users receive a low WiFi signal and a high LTE one. In our experiment, we consider a path loss of 6 dB for WiFi, and consider that 50% of multihomed users are indoor. We show in Figure 3.9 that indoor users achieve a higher throughput than outdoor ones. This is explained by the fact that indoor users receive a good WiFi signal.

3.7 Conclusion

We studied in this chapter different types of multihoming resource allocation strategies in multi-access LTE/WiFi networks, and evaluated their impact on single-homed and multihomed users as well as on the global system, in terms of achievable throughput and stability. We considered two approaches: network centric and user centric. In the former, the network decides how to distribute resources between all classes of users whereas in the latter, the split of file request on both access interfaces is decided by the multihomed user itself.

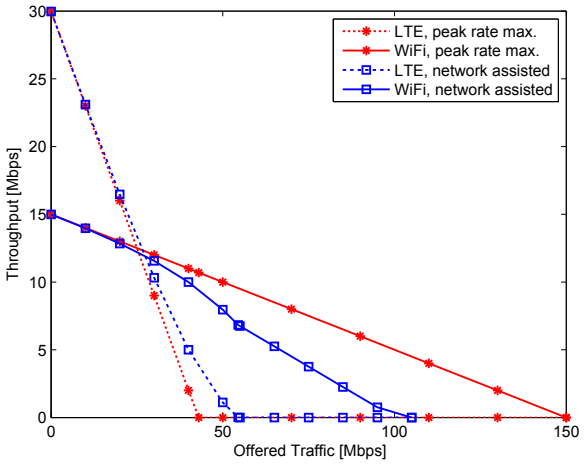


Figure 3.7: LTE and WiFi queues performance and system stability.

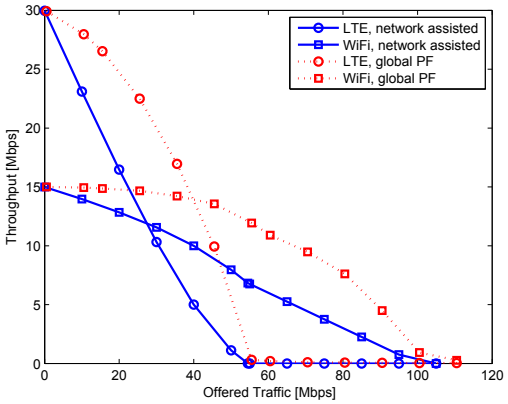


Figure 3.8: Comparison of user centric and network centric strategies.

In each approach, we focused on two strategies: local versus global PF for the network centric strategy and peak rate maximization versus network assisted for the user centric one. We proved also that the network assisted strategy, wherein the user benefits from traffic intensity information on each radio access in addition to their maximal rates, corresponds to the network centric global optimum that maximizes the average throughput in the whole system.

In addition, we compared network centric global PF to the user centric network assisted strategy. We showed that global PF achieves a better performance in terms of throughput. However, network assisted strategy requires less computational resources.

Finally, we evaluated the achievable rates of multihomed users in heterogeneous radio conditions: indoor and outdoor and showed that the former

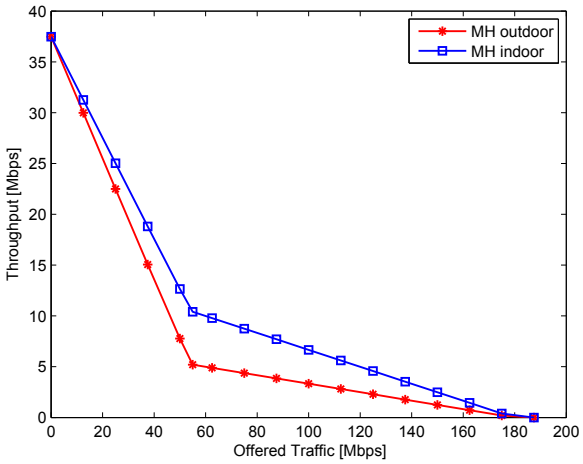


Figure 3.9: Multihomed achievable throughput for indoor and outdoor users.

makes them take more advantage of WiFi and hence achieve higher rates.

Chapter 4

Joint Radio/Processing Resource Allocation in V-RAN

4.1 Introduction

We introduce in this chapter V-RAN integration in a 5G network by adopting the previous model of macro cells/small cells interworking with multihoming. V-RAN was proposed as a solution to limit the capital and operating expenditure (CAPEX and OPEX).

We introduced V-RAN main components in section 2.4: the base- band unit (BBU) responsible for the baseband functions and the remote radio head (RRH) which integrates the radio functions. A BBU pool in the V-RAN ensures the simplicity of any future cell deployment because in this case the deployment will be handled at the BBU level with only deploying the RRH site which is less costly than the BBU unit deployment. Yet, the V-RAN increases user experience by applying enhanced inter-cell interference techniques [82] and coordinating multi-point [138, 84] at the BBU pool.

We address in this chapter the problem of joint allocation of multiple types of resources, mainly frequency and processing resources, to different classes of users with heterogeneous demands. Frequency resources are limited by the spectrum allocated to each access network and processing resources are limited by the BBU capacity offered by a given network operator. We also consider the case of multihomed users and the heterogeneity of access types each requiring different processing resources.

Although resource allocation strategies have been widely studied for single resources in [139, 101] and multiple types of resources [140, 141, 103, 106], these strategies cannot be applied directly to the V-RAN because they do not consider the case of independent resources. For this aim, we modify existing strategies and consider two cases for the V-RAN: with sufficient and

with limited processing resources.

Furthermore, we focus on the trade-off between power consumption and the achievable rate in the V-RAN with multihoming. We build on the existing state-of-the art solutions in order to find an energy efficient joint allocation strategy without compromising the achievable throughput in a multihoming capable V-RAN.

4.2 V-RAN for heterogeneous networks

4.2.1 V-RAN architectural considerations

HetNets are the result of large macro cells combination with small cells. These small cells increase capacity in loaded areas, or fill in the area with low coverage [142]. Small cells are depolyed in the macro eNB by adding low-power base stations as eNBs, Home eNBs, Relay Nodes, or Remote Radio Heads (RRH). In particular, RRH based small cell deployment can have distinct architectures [143]:

- Distributed baseband architecture consisting of having BBU and RRH units in the small cell.
- Common baseband architecture including a BBU connected to one or more RRH where RRH can be geographically separated from the BBU.

As mentioned in section 2.4, RAN architecture has evolved from all-in-one architecture to distributed base stations with RRH separated from the BBU [144]. First, BBU and RRH were separated by fiber links that carry the baseband wireless signal [145]. In this architecture, RRH's role is receiving the wireless signal and power amplification; while BBU implements the antenna and other layer 2 and 3 functionalities. Second, BBUs were unified in a BBU system platform with unified processing board hardware platform and support Software Defined Radio (SDR) upgrade. However, this second architecture does not support real-time dynamic resource allocation. Third, BBUs were virtualized in data centers that enable dynamic service provisioning forming thus a BBU pool to different virtualized base stations and different air interface standards. Moreover, this virtualization does not include small cells only, but also traditional RAN architectures such as BS in 2G networks, NodeB in 3G networks, and eNodeB in 4G networks.

4.2.2 System description

We consider a heterogeneous V-RAN composed of a macro cell (MC) co-existing with K small cells (SCs). Each access network is composed of an RRH as shown in Fig. 4.1. These RRH are connected to their corresponding BBU forming thus a BBU pool.

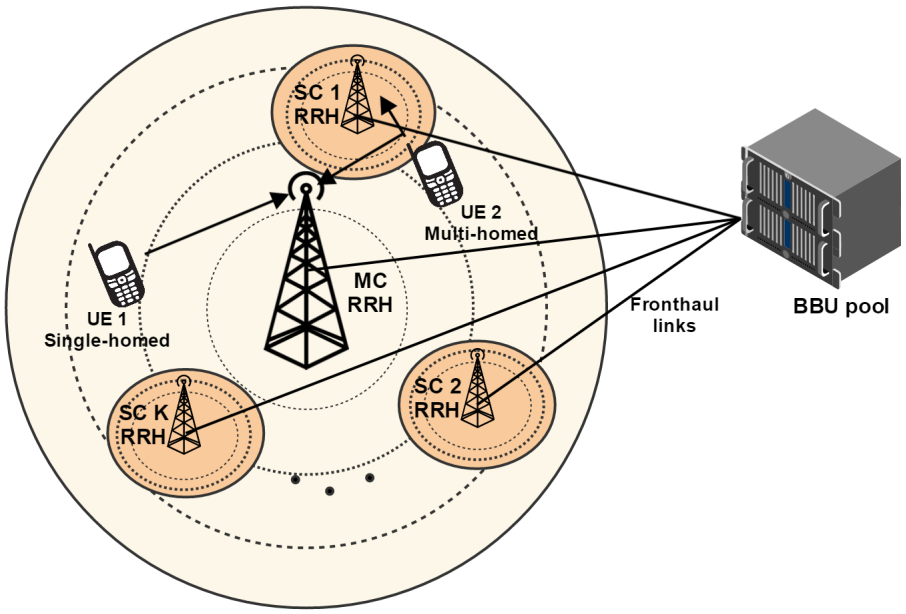


Figure 4.1: V-RAN general model

In such centralized networks, the RRH handles the analog processing as well as the digital to analog conversion. The BBU pool is responsible for the digital baseband processing functions. Optical fiber cables are used to provide data transmission between the RRHs and the BBU pool and ensure an ideal fronthaul that supports in-phase/quadrature samples exchange between RRH and BBU.

Multihomed users achieve dual connectivity on lower layers. The medium access control (MAC) layer manages resource mapping between physical layers of macro and small cells, their data modulation, hybrid automatic repeat request (HARQ) as well as channel coding on the corresponding carrier component. In addition, this unique MAC layer supports cross carrier scheduling.

In this chapter, we keep similar notations for global traffic intensity and users classes as previously. We also define the following parameters for the wireless access and the V-RAN:

For the wireless access We define $x_i^M(\mathbf{n})$ and $x_{ki}^{MH_M}(\mathbf{n})$ as the bandwidth shares in the macro cell for single-homed macro cell users and multihomed users having a connection with the k^{th} small cell, respectively, with $k = 1, \cdots, K$ and i denotes the Channel Quality indicator, $i = \text{CQI}$. We also define $x_{ki}^S(\mathbf{n})$ and $x_{ki}^{MH_S}(\mathbf{n})$ as the bandwidth shares in the SC number k for small cell-only single-homed users and multihomed ones, respectively.

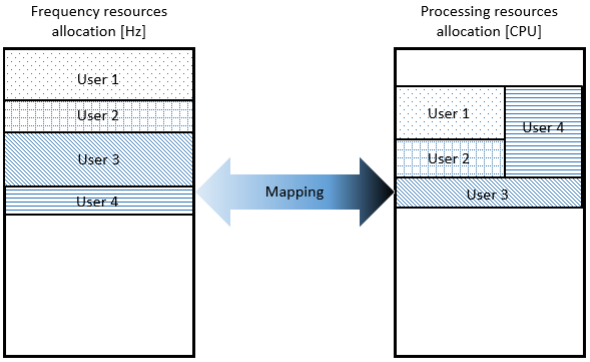


Figure 4.2: Equivalence between radio and processing resource allocation in V-RAN.

For the V-RAN We define $y_i^M(\mathbf{n})$ and $y_{ki}^{MH_M}(\mathbf{n})$ as the CPU shares in the V-RAN for single-homed macro cell users and multihomed users, having a connection with the k^{th} small cell, respectively. We also define $y_{ki}^S(\mathbf{n})$ and $y_{ki}^{MH_S}(\mathbf{n})$ as the CPU shares in the SC number k for SC-only single-homed users and multihomed ones, respectively.

A multi-resource allocation must be performed jointly on the radio and processing resources. Fig. 4.2 shows that an allocated amount of frequency resources requires a corresponding amount of processing resources in the BBU so that they yield together the same throughput perceived by the end user, taking into account different efficiency parameters in each of them. This correspondence depends of the spectral efficiency (SE_i) in the wireless access, with i = CQI, and the V-RAN efficiency given by the processing efficiency $\gamma = \{\gamma_i^M, \gamma_i^S\}$ which is function of the users' radio conditions and their access network technology. We shall explicit more this statement in the next section.

For the performance evaluation, we consider the achievable data rate by each class of users taking into consideration the heterogeneity of users' radio conditions in the macro and small cells. The achievable data rate R, in both access networks, is function of the spectral efficiency and the allocated spectral resources given by:

$$R = x \times B \times SE \tag{4.1}$$

where x is the bandwidth share, B is the channel bandwidth in [Hz] and SE is the spectral efficiency which is function of the CQI and the access technology in [bps/Hz].

4.3 Case without multihoming

We first focus on multi-resource allocation strategies without multihoming in which we consider both 1) the spectral resource share at the macro or the small cells and 2) the processing resource share at the BBU pool.

In this chapter, we adapt the strategies proposed in [103, 105, 108] to our case where radio resources of macro and small cells are independent and all users share the same processing resources from the BBU pool.

We start by evaluating a baseline network model where no V-RAN is considered. We then present the resource allocation adapted to our system model in case of V-RAN. We start by the Proportional Fairness (PF), then the Dominant Resource Fairness (DRF) and finally our proposed strategy called Two-Phase Allocation (TPA).

4.3.1 Baseline network model without V-RAN

The baseline radio access network consists of a HetNet composed of a macro cell and a set of small cells. We consider a proportional fairness resource allocation that finds bandwidth capacity shares allocated for each class of users. These capacity shares are obtained by maximizing the utility function equal to the sum of logarithms of users' throughput as defined in [97].

In this case, the achievable rate by macro cell and small cells users are independent. The capacity shares of macro cell users are equal to the solution of Eq. (4.2) subject to conditions given in Eq. (4.3).
Maximize:

$$U_M = \sum_i n_i^M \log(x_i^M SE_i B^M) \tag{4.2}$$

subject to:

$$n_i^M x_i^M = 1, \tag{4.3a}$$

$$x_i^M \in [0, 1] \tag{4.3b}$$

with $i = \text{CQI}$. Eq. (4.3a) is the capacity constraint of the macro cell access network which ensures that the sum of capacity shares obtained by the users served by the macro cell equals to 1.

For the k^{th} small cell, the utility function is defined as the sum of the logarithms of the flow rates obtained by the k^{th} small cell users as follows:
Maximize:

$$U_S^k = \sum_i n_{ki}^S \log(x_{ki}^S SE_i B^S) \tag{4.4}$$

subject to:

$$n_{ki}^S x_{ki}^S = 1, \tag{4.5a}$$

$$x_{ki}^S \in [0, 1] \tag{4.5b}$$

with $k = 1, \cdots, K$ and $i = \text{CQI}$. Eq. (4.5a) is the constraint that limits the sum of the users' capacity shares on the SC k to 1.

The solutions of problems in Eq. (4.2) and (4.4) are given by:

$$x_i^M = 1/(\sum_i n_i^M) \tag{4.6a}$$

$$x_{ki}^S = 1/(\sum_i n_{ki}^S) \tag{4.6b}$$

4.3.2 Proportional fairness with V-RAN

We now model the proportional fair resource allocation while considering the V-RAN. We do so based on our model provided in chapter 3 by introducing a new condition for the processing resource allocation.

We maximize the following utility function for the system:

$$U = \sum_{i \in \mathrm{I}} n_i^M \log\left(x_i^M B^M SE_i\right) + \sum_k \sum_{i \in \mathrm{I}} n_{ki}^S \log\left(x_{ki}^S B^S SE_i\right) \tag{4.7}$$

subject to:

$$\sum_{i \in \mathrm{I}} n_i^M x_i^M \leq 1, \tag{4.8a}$$

$$\sum_{i \in \mathrm{I}} n_{ki}^S x_{ki}^S \leq 1, \tag{4.8b}$$

$$\sum_{i \in \mathrm{I}} n_i^M y_i^M + \sum_k \sum_{i \in \mathrm{I}} n_{ki}^S y_{ki}^S \leq 1 \tag{4.8c}$$

where

$$x_i^M B^M SE_i = y_i^M V \gamma_i^M \tag{4.9}$$

$$x_{ki}^S B^S SE_i = y_{ki}^S V \gamma_i^S \tag{4.10}$$

The above maximization problem consists in finding the capacity shares for all the users by maximizing the achievable throughput in a proportional fair allocation, taking into consideration the V-RAN capacity in condition (4.8c). In addition, we consider the difference of the processing efficiency as a function of the technology and the adopted modulation; this processing efficiency is denoted by γ_i^M and γ_i^S for the macro and small cell, respectively. The mapping between allocated spectral resources and allocated processing resources yielding the same throughput is shown in Eqns. (4.9) and (4.10). In the following, we replace y by x according to these mapping equations.

The Lagrangian of this system is given by:

$$\begin{aligned} L = & \sum_{i\in \mathrm{I}} n_i^M \log\left(x_i^M B^M SE_i\right) + \sum_k \sum_{i\in \mathrm{I}} n_{ki}^S \log\left(x_{ki}^S B^S SE_i\right) \\ & - \mu_1 (\sum_{i\in \mathrm{I}} n_i^M x_i^M - 1) - \sum_k \mu_2^k (\sum_{i\in \mathrm{I}} n_{ki}^S x_{ki}^S - 1) \\ & - \mu_3 (\sum_{i\in \mathrm{I}} n_i^M \frac{B^M SE_i}{V\gamma_i^M} x_i^M + \sum_k \sum_{i\in \mathrm{I}} n_{ki}^S \frac{B^S SE_i}{V\gamma_i^S} x_{ki}^S - 1) \end{aligned} \tag{4.11}$$

The Karush Kuhn Tucker (KKT) conditions for problem (4.11) are:

$$\nabla L = 0, \tag{4.12a}$$

$$\sum_{i\in \mathrm{I}} n_i^M x_i^M - 1 \leq 0, \tag{4.12b}$$

$$\sum_{i\in \mathrm{I}} n_{ki}^S x_{ki}^S - 1 \leq 0, \tag{4.12c}$$

$$\sum_{i\in \mathrm{I}} n_i^M \frac{B^M SE_i}{V\gamma_i^M} x_i^M + \sum_k \sum_{i\in \mathrm{I}} n_{ki}^S \frac{B^S SE_i}{V\gamma_i^S} x_{ki}^S - 1 \leq 0, \tag{4.12d}$$

$$\mu_1 \geq 0, \quad \mu_2^k \geq 0, \quad \mu_3 \geq 0 \tag{4.12e}$$

and the complementary slackness:

$$\mu_1 (\sum_{i\in \mathrm{I}} n_i^M x_i^M - 1) = 0, \tag{4.13a}$$

$$\mu_2^k (\sum_{i\in \mathrm{I}} n_{ki}^S x_{ki}^S - 1) = 0, \qquad \forall k = 1, \cdots, K \tag{4.13b}$$

$$\mu_3 (\sum_{i\in \mathrm{I}} n_i^M \frac{B^M SE_i}{V\gamma_i^M} x_i^M + \sum_k \sum_{i\in \mathrm{I}} n_{ki}^S \frac{B^S SE_i}{V\gamma_i^S} x_{ki}^S - 1) = 0 \tag{4.13c}$$

In order to solve this problem, we first assume that $\mu_1 = 0$, $\mu_2^k = 0$ and $\mu_3 \neq 0$ and we solve the problem. Then we solve the problem by assuming that $\mu_1 \neq 0$, $\mu_2^k \neq 0$ and $\mu_3 = 0$. The solution is as follows:

$$x_i^M = \min\left(\frac{V\gamma_i^M}{n_t B^M SE_i}, \frac{1}{\sum_{i\in \mathrm{I}} n_i^M}\right) \tag{4.14a}$$

$$x_{ki}^S = \min\left(\frac{V\gamma_i^S}{n_t B^S SE_i}, \frac{1}{\sum_{i\in \mathrm{I}} n_{ki}^S}\right) \tag{4.14b}$$

with $n_t = \sum_{i\in \mathrm{I}} n_i^M + \sum_k \sum_{i\in \mathrm{I}} n_{ki}^S$.

It is worth to note that the advantage of PF allocation is the fairness between all users. However, this fairness could decrease the system's efficiency when applied to our case because by ensuring fairness between all

users, some resources might be left unused which violates the *Pareto efficiency* condition. To overcome this, we focus next on two resource allocation strategies that achieve a compromise between efficiency and fairness: DRF and a new proposal which we call TPA.

4.3.3 Dominant resource fairness with V-RAN

The adaptation of DRF to our context lies in its extension to take into consideration the correspondence between V-RAN's processing resources and the radio ones. To realize the ideal DRF allocation proposed by [103], authors in [105] stated that one can employ a water-filling algorithm. This is possible by increasing the resource allocation at the same rate until some resource is fully used. In this case, users using that resource are frozen while other users allocations are increased until all the resources are frozen. We show in Algorithm (1) the steps of DRF resource allocation applied to our case. The advantage of this strategy is that it is *Pareto efficient* because it ensures that no fraction of any resource is left needlessly idle.

Algorithm 1 Water-filling algorithm

Parameters:
$dx \leftarrow$ increment step
$x_i^M \leftarrow$ macro cell class i user share
$x_{ki}^S \leftarrow$ small cell k, class i user share
Algorithm:
do
  for each class $i \in macrocell$ **do**
    if *macrocell* stable **then**
      $x_i^M \leftarrow x_i^M + dx.$
    else
      $stop(1, i) = 1.$
  for each $k \in K$ **do**
    for each class $i \in smallcell\ k$ **do**
      if SC stable **then**
        $x_{ki}^S \leftarrow x_{ki}^S + dx.$
      else
        $stop(k + 1, i) = 1.$
while V-RAN is stable and $\sum\sum(stop) \leq$ nb of classes

4.3.4 Two-phase allocation with V-RAN

In this section, we propose a two-phase resource allocation strategy which achieves the advantages of the two above-mentioned strategies: it specifically achieves the fairness of PF and the Pareto efficiency of water-filling.

The resource allocation shares are first based on those obtained by Eq. (4.7). Then a feedback step allocates the unused resources remaining in the network. This strategy is shown in Algorithm (2) where f (resp. f') is the mapping (resp. de-mapping) function between radio resources and V-RAN's (resp. V-RAN resources and radio ones) as follows:

$$y = f(x) = \frac{x\ B\ SE}{V\ \gamma} \tag{4.15a}$$

$$x = f'(y) = \frac{y\ V\ \gamma}{B\ SE} \tag{4.15b}$$

with the parameters $B = \{B^M, B^S\}$, $SE = \{SE_i\}$ and $\gamma = \{\gamma_i^M, \gamma_i^S\}$ are chosen according to the user's class.

Algorithm 2 Two-phase allocation

Parameters:
$x_i^M \leftarrow$ From Eq. (4.14a)
$x_{ki}^S \leftarrow$ From Eq. (4.14b)
$feedback \leftarrow$ logical index of accesses that can increase
their users' shares
Algorithm:
if $\sum_i n_i^M x_i^M \leq 1$ **then**
$feedback(MC) = 1$
for $k = 1 : K$ **do**
if $\sum_i n_{ki}^S x_{ki}^S \leq 1$ **then**
$feedback(SC\ k) = 1$
Update the allocation for the accesses where feedback is 1
$x(feedback) = f'(\frac{1-n(\overline{feedback})f(x(\overline{feedback}))}{n(feedback)})$

4.4 Case with multihoming

In this section, we consider the network model provided in section 4.3.2 with multihomed users. We model the multi-resource allocation strategies proposed in section 4.3 with adding the corresponding parameters for multihomed users.

We first present the baseline network model without V-RAN. We then provide resource allocation with PF, DRF and TPA taking V-RAN into consideration.

4.4.1 Baseline network model without V-RAN

The baseline network with multihoming is a heterogeneous radio access network where users can be single-homed or multihomed according to the user

equipment capabilities as well as to the received signals from the macro cell and the small cells. We consider PF resource allocation. The problem in Eq. (4.16) consists in maximizing the sum of the logarithms of the flow rates obtained by single-homed and multihomed users on each access network. The solution is the capacity share values $x_i^M(\mathbf{n})$,$x_{ki}^S(\mathbf{n})$, $x_{ki}^{MH_M}(\mathbf{n})$, $x_{ki}^{MH_S}(\mathbf{n})$ of macro cell users, k^{th} small cell users, k^{th} multihomed users on macro cell access and k^{th} multihomed users on small cell, respectively.
Maximize:

$$\begin{aligned} U = &\sum_i n_i^M \log(x_i^M SE_i B^M) \\ &+ \sum_i \sum_{k=1}^{K} n_{ki}^S \log(x_{ki}^S SE_i B^S) \\ &+ \sum_i \sum_{k=1}^{K} n_{ki}^{MH} \log(x_{ki}^{MH_M} SE_i B^M + x_{ki}^{MH_S} SE_i B^S) \end{aligned} \tag{4.16}$$

subject to:

$$n_i^M x_i^M + \sum_{k=1}^{K} n_{ki}^{MH} x_{ki}^{MH_M} = 1, \tag{4.17}$$

$$n_{ki}^S x_{ki}^S + n_{ki}^{MH} x_{ki}^{MH_S} = 1, \tag{4.18}$$

$$x_i^M, x_{ki}^{MH_M} \in [0,1], \tag{4.19}$$

$$x_{ki}^S, x_{ki}^{MH_S} \in [0,1] \tag{4.20}$$

with $k = 1, \cdots, K$. Eqs. (4.17) and (4.18) are the constraints which ensure that the sum of capacity shares on the macro cell and the small cell access networks, respectively, is equal to 1.

For K small cells, this maximization problem is solved numerically in MATLAB. However, if we assume $K = 1$ with one macro cell and only one small cell, a closed form expression for the capacity shares is obtained by deriving the Lagrangian function of the problem described in Eq. (4.16) under the corresponding constraints for $K = 1$ in Eqs. (4.17) and (4.18).

$$\begin{aligned} L = &n^M \log(x^M SE\ B^M) + n^S \log(x^S SE\ B^S) \\ &+ n^{MH} \log(x^{MH_M} SE\ B^M + x^{MH_S} SE\ B^S) \\ &- \nu_M(n^M x^M + n^{MH} x^{MH_M} - 1) \\ &- \nu_S(n^S x^S + n^{MH} x^{MH_S} - 1) \end{aligned} \tag{4.21}$$

with ν_M and ν_S the Lagrange multipliers. Then:

$$\frac{\delta L}{\delta x^M} = \frac{n^M}{x^M} - n^M \nu_M, \tag{4.22}$$

$$\frac{\delta L}{\delta x^S} = \frac{n^S}{x^S} - n^S \nu_S, \tag{4.23}$$

$$\frac{\delta L}{\delta x^{MH_M}} = \frac{n^{MH} SE\ B^M}{x^{MH_M} SE\ B^M + x^{MH_S} SE\ B^S} - n^{MH} \nu_M, \tag{4.24}$$

$$\frac{\delta L}{\delta x^{MH_S}} = \frac{n^{MH} SE\ B^S}{x^{MH_M} SE\ B^M + x^{MH_S} SE\ B^S} - n^{MH} \nu_S \tag{4.25}$$

By replacing Eqs. (4.22) and (4.23) in Eqs. (4.17) and (4.17) for $K = 1$, we find x^M and x^S as a function of ν_M and ν_S, respectively. Then we find the Lagrange multipliers by substituting x^{MH_M} and x^{MH_S} in Eq. (4.24)

$$\nu_M = \frac{B^M(n^M + n^S + n^{MH})}{B^M + B^S} \tag{4.26}$$

$$\nu_S = \frac{B^S(n^M + n^S + n^{MH})}{B^M + B^S} \tag{4.27}$$

The unique solution to the problem is given by:

$$x^M = \frac{B^M + B^S}{B^M(n^M + n^S + n^{MH})} \tag{4.28}$$

$$x^S = \frac{B^M + B^S}{B^S(n^M + n^S + n^{MH})} \tag{4.29}$$

$$x_{MH_M} = \frac{B^M n^S - B^S n^M + B^M n^{MH}}{B^M n^{MH}(n^M + n^S + n^{MH})} \tag{4.30}$$

$$x_{MH_S} = \frac{B^S n^M - B^M n^S + B^S n^{MH}}{B^S n^{MH}(n^M + n^S + n^{MH})} \tag{4.31}$$

where n^M, n^S and n^{MH} denote the number of macro cell, small cell and multihomed users, respectively.

4.4.2 Proportional fairness with V-RAN and multihoming

We extend the model provided in section 4.4.1 by introducing processing resources condition at the BBU pool as well as resource mapping expressions that maintain the correspondence between frequency and processing resources for each class of single-homed and multihomed users.

The capacity shares in this case are the maximization solution of the problem in Eq. (4.32) under radio and processing resource stability conditions.

Maximize:

$$\begin{aligned} U = & \sum_{i \in \mathrm{I}} n_i^M \log\left(x_i^M B^M SE_i\right) \\ & + \sum_k \sum_{i \in \mathrm{I}} n_{ki}^S \log\left(x_{ki}^S B^S SE_i\right) \\ & + \sum_k \sum_{i \in \mathrm{I}} n_{ki}^{MH} \log\left(x_{ki}^{MH_M} B^M SE_i + x_{ki}^{MH_S} B^S SE_i\right) \end{aligned} \tag{4.32}$$

subject to following conditions (Eq. (4.33)) that maintain stability in all access networks as well as in the BBU pool:

$$\sum_{i \in \mathrm{I}} n_i^M x_i^M B^M + \sum_k \sum_{i \in \mathrm{I}} n_{ki}^{MH} x_{ki}^{MH_M} B^M \leq B^M, \tag{4.33a}$$

$$\sum_{i \in \mathrm{I}} n_{ki}^S x_{ki}^S B^S + \sum_{i \in \mathrm{I}} n_{ki}^{MH} x_{ki}^{MH_S} B^S \leq B^S, \tag{4.33b}$$

$$\begin{aligned} & \sum_{i \in \mathrm{I}} n_i^M y_i^M + \sum_k \sum_{i \in \mathrm{I}} n_{ki}^S y_{ki}^S + \\ & \sum_k \sum_{i \in \mathrm{I}} n_{ki}^{MH} (y_{ki}^{MH_M} + y_{ki}^{MH_S}) \leq 1 \end{aligned} \tag{4.33c}$$

taking into consideration the mapping functions between spectral resources and processing resources:

$$x_i^M B^M SE_i = y_i^M V \gamma_i^M, \tag{4.34a}$$

$$x_{ki}^S B^S SE_i = y_{ki}^S V \gamma_i^S, \tag{4.34b}$$

$$x_{ki}^{MH_M} B^M SE_i = y_{ki}^{MH_M} V \gamma_i^M, \tag{4.34c}$$

$$x_{ki}^{MH_S} B^S SE_i = y_{ki}^{MH_S} V \gamma_i^S \tag{4.34d}$$

This allocation strategy considers the difference of processing efficiency as a function of the technology and the modulation adopted, this processing efficiency is denoted by γ_i^M, γ_i^S. The mapping between spectral efficiency and processing efficiency is presented in Eqns. (4.34).

Here also, applying proportional fairness to our network model with multihoming does not allow us to benefit from all available resources in the network. For this reason, we consider next DRF as well as TPA allocation strategies.

4.4.3 Dominant resource fairness with V-RAN and multihoming

In this section, we present the dominant resource fairness described in section 4.3.3 while considering the multihomed users in the network. For this reason,

Algortihm 3 is the extension of Algorithm 1; it allocates resources to single-homed and multihomed users in each access network. It is important to note that the water-filling with multihoming satisfies DRF's properties defined in [103].

Algorithm 3 is a sharing incentive in that no user is better off if resources are equally partitioned among them. Second, a user cannot increase his allocation by lying about his requirements which maintains the strategy proofness. Third, water-filling with multihoming is envy-free in that no user would trade his allocation with another user. Eventually, it is Pareto efficient since all resources are used, to increase a user allocation we must decrease the allocation of another one.

Algorithm 3 Water-filling algorithm

Parameters:
$dx \leftarrow$ increment step
$x_i^M, x_{ki}^{MH_M} \leftarrow$ macro cell class i single-homed and
multihomed users share resp.
$x_{ki}^S, x_{ki}^{MH_S} \leftarrow$ small cell k, class i single-homed and
multihomed users share resp.
Algorithm:
do
  for each user class $i \in$ *macrocell* **do**
    if *macrocell* stable **then**
      $x_i^M \leftarrow x_i^M + dx.$
      $x_{ki}^{MH_M} \leftarrow x_{ki}^{MH_M} + dx.$
    else
      $stop(1, i) = 1.$
  for each $k \in K$ **do**
    for each user class $i \in$ *smallcell* k **do**
      if *smallcell* stable **then**
        $x_{ki}^S \leftarrow x_{ki}^S + dx.$
        $x_{ki}^{MH_S} \leftarrow x_{ki}^{MH_S} + dx.$
      else
        $stop(k + 1, i) = 1.$
while V-RAN is stable and $\sum\sum(stop) \leq$ nb of classes

4.4.4 Two-phase allocation with V-RAN and multihoming

The two-phase heuristic allocation with multihomed users is defined as the extension of section 4.4.2. In this case, a feedback step after finding the capacity shares must be added: we calculate the unused resources and re-distribute them among the users while maintaining small cells, macro cell

and V-RAN's stability. f and f' are calculated using Eq. (4.15). Recall that f (resp. f') calculates the mapping (resp. de-mapping) function between radio and V-RAN (resp. V-RAN and radio). This strategy allows to increase the allocation efficiency and maintains the fairness between all users.

Algorithm 4 Two-phase allocation

Parameters:
$x_i^M,\ x_{ki}^S,\ x_{ki}^{MH_M},\ x_{ki}^{MH_S} \leftarrow$ solution of problem (4.32)
$feedback \leftarrow$ logical index of accesses that can increase their users' shares
Algorithm:
if $\sum_i n_i^M x_i^M + \sum_k \sum_i n_{ki}^{MH} x_{ki}^{MH_M} \leq 1$ **then**
    $feedback(MC) = 1$
for $k = 1 : K$ **do**
    if $\sum_i n_{ki}^S x_{ki}^S + \sum_i n_{ki}^{MH} x_{ki}^{MH_S} \leq 1$ **then**
        $feedback(SC\ k) = 1$
Update the allocation for the accesses where feedback is 1
$x(feedback) = f'(\frac{1-n(\overline{feedback})f(x(\overline{feedback}))}{n(feedback)})$

4.5 Accounting for power consumption in V-RAN

In this section, we elaborate on the energy efficiency of resource allocation strategies applied on the V-RAN architecture. The V-RAN network architecture is an essential part of energy efficient 5G networks since the baseband processing virtualization leads to more energy-efficient cooling, better coordination (CoMP) and dynamic cell reconfiguration. Many recent works have surveyed the energy efficient gains obtained with C-RAN [146, 147, 148, 126]. However, few researchers addressed the energy efficient resource allocation in C-RAN/V-RAN. Dynamic resource allocation in C-RAN was studied in [123] and showed a 70% power consumption reduction. As for multihoming, an energy efficient study was proposed in [121].

We focus in this section on the trade-off between power consumption and the achievable rate in the V-RAN with multihoming and compare the performance of multi-resource allocation strategies from an energy efficient point of view.

For the power models, we consider the holistic energy efficiency evaluation framework (E^3F) that has been developed within the Energy Aware Radio and neTwork tecHnologies (EARTH) project [113, 112]. E^3F presents a quantitative evaluation for radio access network's operation in a cellular network with a detailed power model of a base station.

We consider the V-RAN presented in Fig. 4.1, with heterogeneous access

networks composed of an LTE macro cell and K mm-wave small cells. In fact, network densification combined with the use of cm-wave and mm-wave frequency bands above 10GHz increases the available bandwidth for short range communications in dense areas. However, reliable mm-wave communication consist of power-hungry transceiver chains as well as complex processing techniques [128, 149]. We first consider the network without V-RAN. Then, we evaluate the power consumption of DRF and TPA strategies. We also evaluate the case of assigning multihomed users to a corresponding access network using previously mentioned parameter a, defined as the rate of multihomed users served by the macro cell. Finally, we consider an energy efficient allocation by using the Dinklebach's algorithm [150].

4.5.1 Modeling power consumption in V-RAN

In the following, we provide power models for both access types: macro and small cells, in a V-RAN with BBU pooling.

The total power consumption is equal to the sum of RRH's and BBU's power consumption:

$$P_{tot} = P_{MC}^{RRH} + \sum_{k} P_{SC_k}^{RRH} + s * P^{BBU} \tag{4.35}$$

where P_{MC}^{RRH}, $P_{SC_k}^{RRH}$, and P^{BBU} denote the power consumption at the MC's RRH, the power consumption at the k^{th} SC and the power consumption of a BBU server respectively. s denotes the number of BBUs active in the BBU pool.

An LTE RRH power consumption is provided by EARTH in [113] as:

$$P_{MC}^{RRH} = N_{TRX}(P_0 + \Delta_p P_{out}) \tag{4.36}$$

where N_{TRX} denotes the number of transmission chains, P_0 is the linear model parameter to represent the power consumption at the zero RF output power without the baseband power consumption and Δ_p is the slope of the load dependent power consumption. This fixed power consumption P_0 is independent from the access network load and is called the tax of the coverage, it is consumed by the broadcast channels that are continuously emitting even at 0 load. P_{out} is the dynamic RF power equal to $\rho_{MC} P_{max}$ where ρ_{MC} is the load in the macro cell, and P_{max} is the maximum transmit power. All these parameters are summarized in Table 4.3.

As for mm-wave small cells, the RRH power consumption $P_{SC_k}^{RRH}$ is assumed to be a constant value regardless of the traffic load [115].

Finally, the total BBU pool power consumption is defined as:

$$P_{BBU}^{tot} = s \times P_{BBU} \tag{4.37}$$

where s is the total number of allocated BBUs in the pool, and P_{BBU} is the power consumption in a BBU.

4.5.2 Energy efficiency of resource allocation schemes

In this section, we present resources allocation strategies considered in the previous section in order to study the trade-off between the achievable data rate and the power consumption. We focus on multihomed users' role in increasing or decreasing the total power consumption in the network. The baseline model without V-RAN, used as a reference to which we compare our results, is described in section 4.3.1.

A. Case of DRF with multihoming

In this section, we adopt the dominant resource fairness algorithm, Algorithm 3, that allocates resources to single-homed and multihomed users in each access network in the V-RAN.

In this strategy, the BBUs assignment is dynamic, which means that the central scheduler at the operator 1) increases the number of BBUs turned ON as a function of the system load 2) assigns a BBU to serve a set of small cells and macro cells depending on their load.

Let Eq. (4.38) be the rule the central scheduler follows in order to choose the number s of BBU for a given network.

$$s = \left\lceil \frac{C^M \rho^M + \sum_k C_k^S \rho_k^S}{C_{cap}} \right\rceil \tag{4.38}$$

with C^M, C_k^S and C_{cap} denoting the processing capacity in GOPS[1] required for the macro cell at full load, for the small cells at full load, and the processor maximum capacity. ρ^M and ρ_k^S denote the load at both layers.

Based on the load in each access layer, we calculate the total power consumption in the network with V-RAN by applying the DRF allocation strategy.

B. Case of DRF in V-RAN without multihoming

In this case, we consider that multihoming capable users are connected either to the macro cell or to the small cell. A proportion a of multihoming capable users are connected to the macro cell while a proportion $(1-a)$ receive data from the small cells. The aim is to study how the power consumption varies with varying a. We also consider a dynamic BBU assignment with total number of BBU equal to s given in Eq. (4.38).

C. Case of TPA with multihoming

For the two-phase allocation algorithm, Algorithm 4, the BBUs assignment is dynamic and follows Eq. (4.38). The total power consumption is equal to the sum of powers consumed in all access networks.

[1]GOPS: Giga Operations Per Second

It is important to note that TPA is composed of two phases, and does not take into consideration energy efficiency when allocating resources.

4.5.3 Energy efficient allocation for V-RAN

Now, we propose an energy efficient allocation that maximizes the energy efficiency in a V-RAN with multihoming.

Let $P(\mathbf{x})$ be the power consumption function for a given access network, $R(\mathbf{x})$ be the achievable throughput function of the access network.

$$P(\mathbf{x}) = N_{TRX}(P_0' + \rho_{MC}(x)\Delta_p P_{max}) + \sum_{k=1}^{K} P_{SC_k}^{RRH} + sP_{BBU} \tag{4.39a}$$

$$R(\mathbf{x}) = \sum_i n_i^M x_i^M B^M SE_i + \sum_k \sum_i n_{ki}^S x_{ki}^S B^S SE_i + \sum_k \sum_i n_{ki}^{MH}(x_{ki}^{MH_M} B^M SE_i + x_{ki}^{MH_S} B^S SE_i) \tag{4.39b}$$

As energy efficiency increases with decreasing energy-per-bit, the optimization problem that maximizes the energy efficiency is equivalent to that of minimizing the energy-per-bit. In this case, we formulate the optimization problem as follows:

P 1.

$$\textit{minimize} \quad \frac{P(\boldsymbol{x})}{R(\boldsymbol{x})} \tag{4.40}$$

$$\textit{subject to} \quad \sum_i n_i^M x_i^M + \sum_k \sum_i n_{ki}^{MH} x_{ki}^{MH_M} \leq 1 \tag{4.41}$$

$$\textit{and} \quad \sum_i n_{ki}^S x_{ki}^S + \sum_i n_{ki}^{MH} x_{ki}^{MH_S} \leq 1 \tag{4.42}$$

$$\textit{and} \quad \sum_{i\in \mathrm{I}} n_i^M y_i^M V + \sum_k \sum_{i\in \mathrm{I}} n_{ki}^S y_{ki}^S V + \sum_k \sum_{i\in \mathrm{I}} n_{ki}^{MH}(y_{ki}^{MH_M} + y_{ki}^{MH_S})V \leq V \tag{4.43}$$

taking into consideration the mapping functions between spectral efficiency and processing efficiency:

$$x_i^M B^M SE_i = y_i^M V \gamma_i^M \tag{4.44a}$$

$$x_{ki}^S B^S SE_i = y_{ki}^S V \gamma_i^S \tag{4.44b}$$

$$x_{ki}^{MH_M} B^M SE_i = y_{ki}^{MH_M} V \gamma_i^M \tag{4.44c}$$

$$x_{ki}^{MH_S} B^S SE_i = y_{ki}^{MH_S} V \gamma_i^S \tag{4.44d}$$

Note that solving the fractional optimization problem P1 directly is very challenging. For this reason, we derive a parametric optimization problem out of the fractional programming by introducing a parameter θ [150] as follows:

P 2.

$$\textit{minimize} \quad z(\theta) = P(\boldsymbol{x}) - \theta\ R(\boldsymbol{x}) \tag{4.45}$$

$$\textit{subject to} \quad \sum_i n_i^M x_i^M + \sum_k \sum_i n_{ki}^{MH} x_{ki}^{MH_M} \leq 1 \tag{4.46}$$

$$\textit{and:} \quad \sum_i n_{ki}^S x_{ki}^S + \sum_i n_{ki}^{MH} x_{ki}^{MH_S} \leq 1 \tag{4.47}$$

$$\sum_{i\in \mathrm{I}} n_i^M y_i^M V + \sum_k \sum_{i\in \mathrm{I}} n_{ki}^S y_{ki}^S V + \sum_k \sum_{i\in \mathrm{I}} n_{ki}^{MH} (y_{ki}^{MH_M} + y_{ki}^{MH_S}) V \leq V \tag{4.48}$$

where the parametric problem P2 is a convex optimization for a given θ solved by applying Dinkelbach's method defined in Algorithm 5.

Algorithm 5 Dinkelbach's method

1: Initialize θ;
2: **do**
3: Determine $Z(\theta)$ and $\mathbf{x}^*$;
4: $\theta^* = \frac{P(\mathbf{x}^*)}{R(\mathbf{x}^*)}$;
5: **while** $z(\theta^*) > \epsilon$

In order to find $\mathbf{x}^*$ for a given θ, we solve the linear minimization in Problem 2.

This allocation strategy considers the difference of processing efficiencies γ_i^M, γ_i^S as a function of the technology and the adopted modulation. The mapping between spectral efficiency and processing efficiency is presented in Eqns. (4.44).

4.6 Simulation and numerical results

4.6.1 Simulation parameters

We suppose that our network is composed of a heterogeneous virtual radio access network composed of an LTE macro cell and mm-wave small cells, with different radio resources and processing requirements. We consider the inter-site multi-connectivity architecture proposed in [8]. By choosing the mm-wave technology, we target high data rates allowed by this technology that can reach several Gbps.

Table 4.1: Parameters affecting baseband power consumption. Default values chosen in [86] and network scenarios.

Description	Default	LTE	mm-wave
Bandwidth [MHz]	10	20	1750
Number of Antennas	2	4	8
Modulation	6	2, 4, 6	2, 4, 6
Coding rate	1	0.5, 0.5, 3/4	0.5, 0.5, 3/4
CQI	15	5, 8, 13	15, 18, 23

Table 4.2: Processing efficiency as a function of MCS in LTE and mm-wave in [Mbps/CPU].

Class	LTE (TC)	mm-wave (LDPC)
1/2 QPSK	2.32	0.607
1/2 16QAM	4.6885	1.098
3/4 64 QAM	9.27	2.025

In these simulations, we consider the LTE macro cell described previously in chapter 3, and three mm-wave small cells of $r_{SC} = 70$ m radius that operate at 60 GHz frequency band with 1.7 GHz bandwidth and deploy an 8×8 antenna array [151]. Other parameters adopted in Chapter 3 remain the same. Using the mapping information between CQI, MCS, data rate and distance for a 60 GHz link [152, 153] and for an LTE access network in [154], the users' distribution is calculated for an LTE and a mm-wave access network.

Let λ denote the overall arrival rate. It is equal to the sum of all arrival rates. The traffic intensity in this case is equal to $\lambda \times \sigma$ in [Mbps], where σ Mbits denotes the mean elastic flow size with exponential distribution.

The processing efficiency γ_i^M and γ_i^S are important parameters in that they differentiate between computational requirements in each access network. Table 4.1 presents the different classes of users in LTE and mm-wave. We estimate LTE and mm-wave processing efficiency using information provided in [86]. We show in Table 4.2 the processing efficiency γ in [Mbps/CPU]. We calculate it for LTE and mm-wave respectively for the three CQIs corresponding to modulations $(2, 4, 6)$ under coding rates $(1/2, 1/2, 3/4)$, respectively. It is important to note that multihomed users processing efficiency on the macro and the small cells are independent and follow the same parameters shown in Table 4.2.

4.6.2 Case without mulihoming

In this section, we compare the achievable throughput obtained by each class of users for each of the allocation strategies under two cases: sufficient

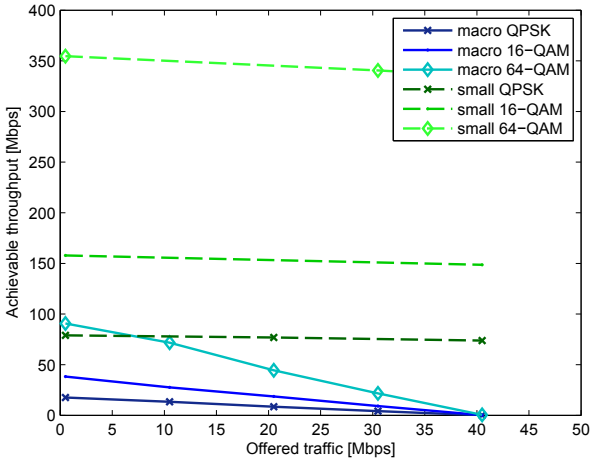


Figure 4.3: Performance evaluation without V-RAN and with single-homed users only.

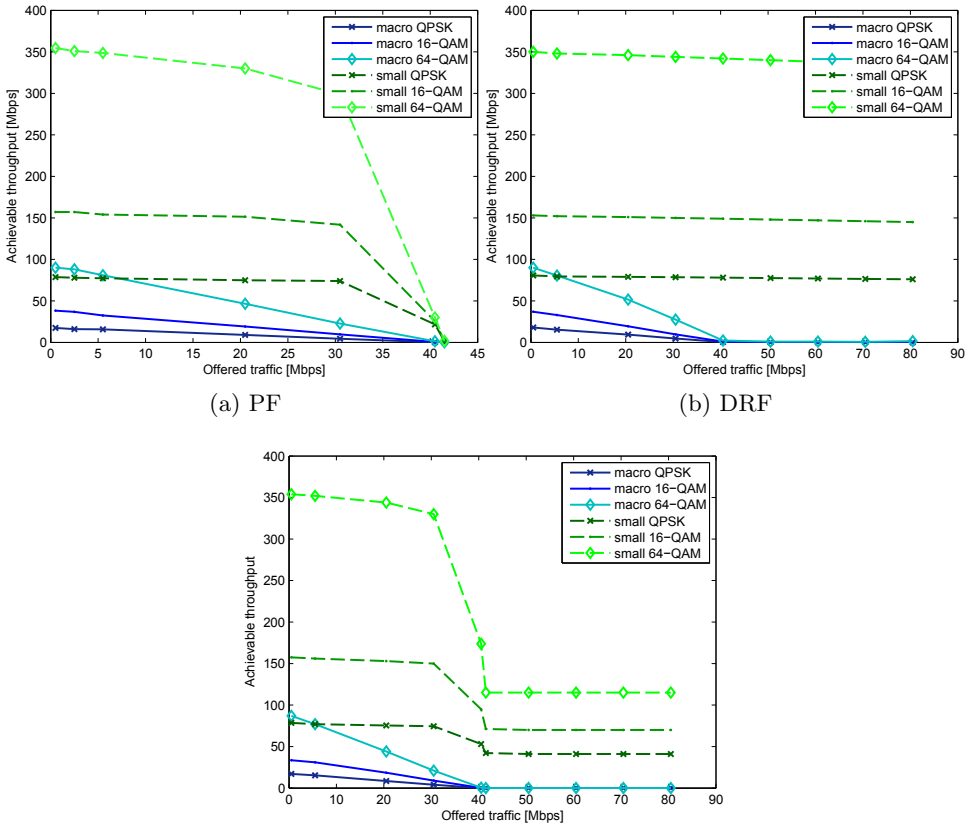
processing resources or limited processing resources.

A. V-RAN with sufficient resources

We first consider that processing resources in the V-RAN are sufficient. We evaluate the achievable throughput in the baseline HetNet (without V-RAN) as a function of the offered traffic as shown in Fig. 4.3 based on the proportional fairness resource allocation. The obtained results show that macro and small cell users data rates are independent since each access network allocates its frequency resources independently. We observe that as the offered traffic in the system increases, the load in the macro cell increases while small cells load increases at a lower rate. This limitation of the macro cell is related to the simulation scenario we considered: macro cell receives higher percentage of offered traffic than small cells.

For a V-RAN with sufficient CPU processing resources (1200 CPU units), we show in Fig. 4.4 the achievable throughput as a function of the offered traffic for each of the proposed strategies. In Fig. 4.4a we see that the proportional fairness strategy achieves a throughput equivalent to that obtained for the baseline network model for low load. Macro cell's stability point is limited by the spectral resources saturation. We denote by stability point the maximum offered traffic that can be transported by the access network before it gets saturated, i.e., the point at which, the achievable throughput becomes equal to zero. On the other hand, we see that small cells stability point is limited to 42 Mbps by the allocation strategy.

It is interesting to see in Fig. 4.4b that the DRF allocation strategy offers better stability than PF and allows to achieve higher data rates for small cell users similar to those obtained by the baseline strategy. This



(a) PF

(b) DRF

(c) TPA

Figure 4.4: Comparison of: (4.4a) proportional fairness (PF), (4.4b) dominant resource fairness (DRF) and (4.4c) two-phase allocation (TPA) strategies' achievable throughput for different classes of users when V-RAN has sufficient processing resources.

allocation strategy allows to virtualize our network without loosing on the performance and stability.

As for the TPA allocation results depicted in Fig. 4.4c, we observe an improvement over PF in terms of achievable throughput and especially in terms of system stability. We also observe an increased fairness between different classes of users illustrated by closer throughput plots.

Fig. 4.5 presents a comparison of the average achievable throughput between the baseline network model and the different allocation strategies. Obviously, the DRF strategy, i.e., waterfilling, achieves the highest performance with an average throughput almost equal to the baseline one when no virtualization was considered. PF appears to have the worst performance with network stability limited to 42 Mbps of offered traffic. In between, the TPA strategy achieves better fairness between different classes of users.

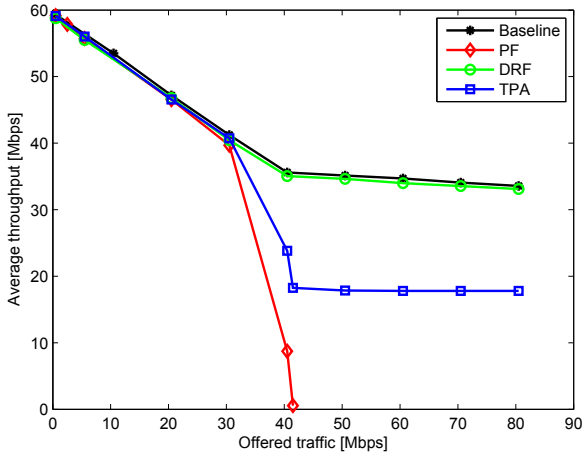


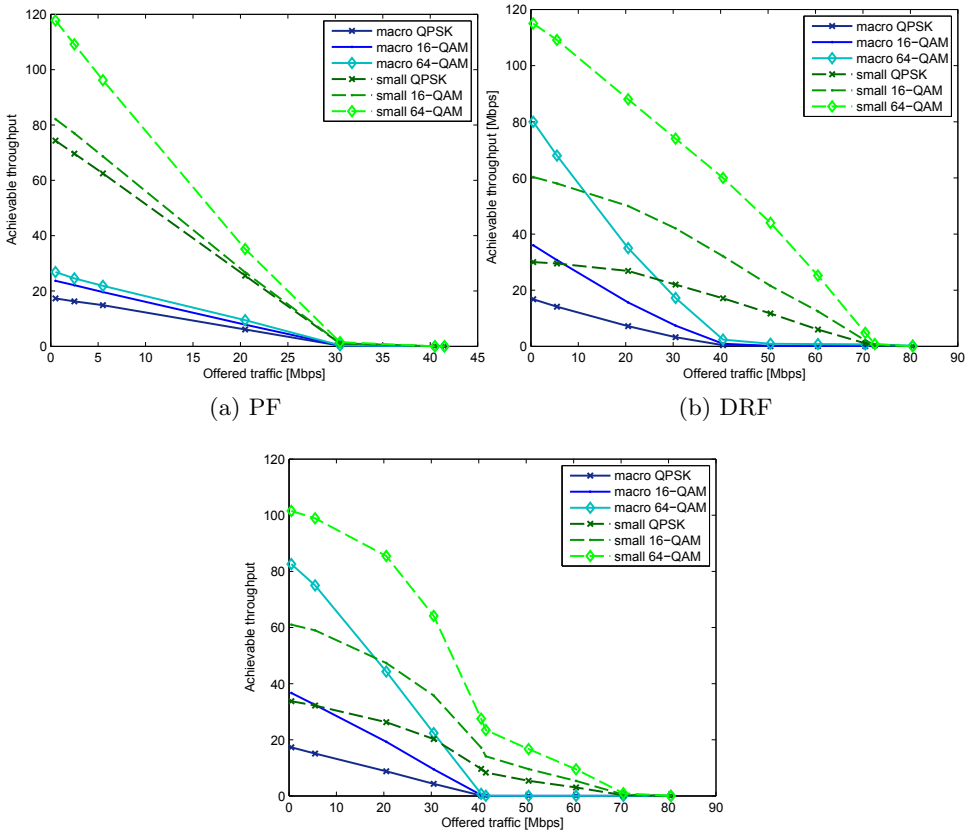
Figure 4.5: Comparing average throughput of different strategies when V-RAN has sufficient resources.

B. V-RAN with limited resources

We assume in this section that V-RAN's processing resources are limiting and compare the different allocation strategies as shown in Figure 4.6. The plotted results show the variation of achievable throughput as a function of the offered traffic for each class of users. For 60 CPU units of limited processing capacity, we observe in Fig. 4.6a that the achievable throughput obtained by all the users decreases for all access networks and reaches the limits of the system's stability at 30 Mbps for both small cells and macro cells. We also find that the maximum achievable throughout at low load decreases for all classes of users.

When comparing PF to the water-filling strategy (i.e., DRF), we observe in Fig. 4.6b that DRF achieves a higher stability for both macro and small cells as well as increased achievable throughput by macro cell users and decreased achievable throughput for small cell users at low load. This variation of allocation is related to the consideration of heterogeneous requirements for each class of users in PF, while disregarding this information in DRF. This stability improvement is explained by the waste of resources in PF that had proven to be not Pareto efficient. While DRF does not stop before all possible spectral and processing resources are allocated, PF chooses between two possible allocations based on the available spectral and processing resources, including unused resources in the network.

Fig. 4.6c shows the two-phase allocation simulation results. We observe that at limited processing capacity, TPA combines the advantages of both PF and DRF. It increases the fairness between different classes of users achieving throughputs that are closer to each other. It also increases the network stability point for both access networks to a point equal to that



(a) PF

(b) DRF

(c) TPA

Figure 4.6: Comparison of: (4.6a) proportional fairness (PF), (4.6b) dominant resource fairness (DRF) and (4.6c) two-phase allocation strategies' achievable throughput of different classes of users when V-RAN has restrictive processing resources.

achieved by DRF.

By comparing the average throughput of the above mentioned allocation strategies, we see in Fig. 4.7 that waterfilling and TPA have comparable results. At low load, PF outperforms water-filling (DRF) and TPA. However, at medium to high load, we observe that water-filling and TPA outperform PF especially from a stability point of view.

C. Jain Fairness Index comparison

In Fig. 4.8 , we evaluate the Jain's fairness index [155] in order to quantify the fairness of each allocation strategy as a function of the offered traffic when no multihoming is considered. For sufficient processing resources at the V-RAN, we see in Fig. 4.8a that all allocation strategies achieve the

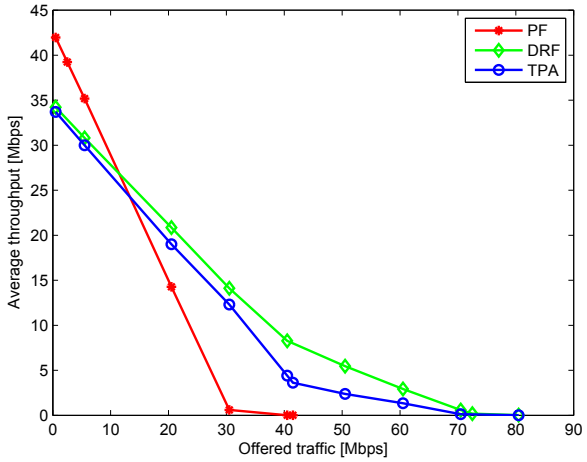
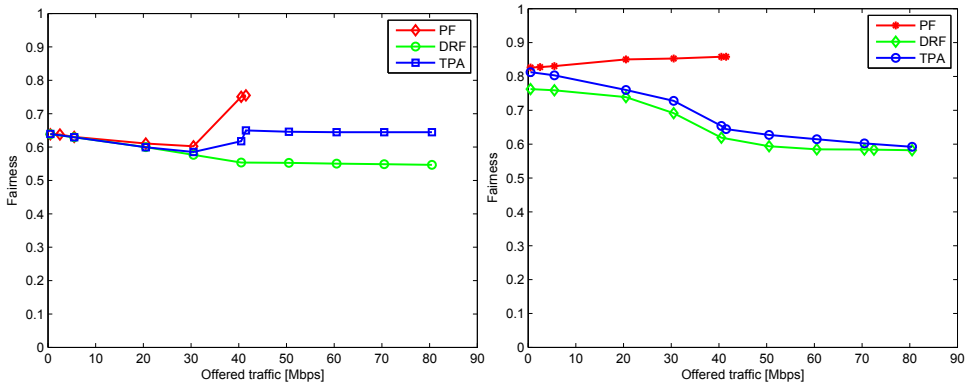


Figure 4.7: Comparing average throughput of different strategies when V-RAN is restrictive.



(a) Fairness index vs. the offered traffic with sufficient V-RAN resources.

(b) Fairness index vs. the offered traffic with limiting V-RAN resources.

Figure 4.8: Fairness index.

same fairness at low load. At higher load, PF shows the highest fairness, water-filling shows the lowest fairness index, and TPA appears to offer a higher fairness than water-filling and better stability than PF. For V-RAN with limiting resources, we observe in Fig. 4.8b that PF, water-filling and TPA achieve higher fairness, with clearer fairness difference. PF achieves always the highest fairness, and water-filling offers the lowest fairness index.

4.6.3 Case with mulihoming

In this section, we show the simulation results for the V-RAN with multi-homing. We specifically investigate how the presence of multihomed users

influences resource allocation. As above, we first consider the network when V-RAN has sufficient processing resources and then present the simulation results for the case when the V-RAN has limiting processing resources at the BBU pool.

A. V-RAN with sufficient resources

The baseline simulation results using PF allocation (without V-RAN) in Fig. 4.9 are considered as a reference to evaluate multi-resource allocation techniques as our aim is to find the allocation strategy that allows us to pool BBU resources in the network without compromising the system's performance.

Fig. 4.9 shows that multihomed users achieve higher throughput than single-homed ones; it becomes equal to the latter when the macro cell is fully loaded. Given the difference between macro and small cells loads, we see that the macro cell reaches the stability limit earlier than the small cells which have lower load and larger spectrum to be allocated.

Simulation results in Fig. 4.10 show the achievable throughput for each class of users as a function of the offered traffic for PF, DRF and TPA strategies. With sufficient processing resources of 1200 CPU unit, PF allocation strategy achieves results almost comparable to those obtained in the baseline model. However, at high load, all users' classes throughputs in all access networks decrease to zero at 40 Mbps. Given that spectral and processing resources are sufficient, we deduce here that the allocation strategy itself is limiting and does not allow to benefit from all the available resources. Multihomed users appear to follow the same strategy: they combine the achievable data rate in the macro cell and the connected small cell, and when one of these networks is saturated multihomed users will be served completely by the other access network. This concurrent access is the privilege of multihomed users over single-homed users.

Dominant resource fairness results implemented as a water-filling algorithm show in Fig. 4.10b the comparable results with the baseline network model. This proves again that DRF achieves an efficient pooling of BBU resources even in the presence of multihomed users. Similarly to the results without multihoming shown in Fig. 4.4b, we find here that DRF improves the system stability with achieving higher data rates for small cell users at high load after the macro cell gets saturated.

Simulation results plotted in Fig. 4.10c show that TPA achieves comparable data rates with the baseline strategy at low to medium loads. At high load, when the macro cell reaches the stability limits, multihomed users join small cells and receive the same data rate as single-homed users.

As a comparison, we show in Fig. 4.11 the average achievable throughput in all multi-resource allocation strategies as well as the baseline strategy when no BBU pooling is considered. The water-filling strategy appears to

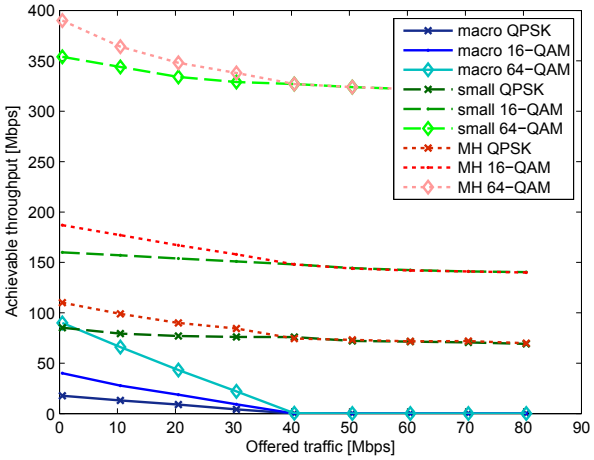


Figure 4.9: Performance evaluation without V-RAN architecture and with multihoming.

achieve comparable results with the baseline strategy; it also outperforms the two-phase allocation strategy. The latter achieves an acceptable average data rate as well as a high system stability. Clearly, even after considering the multi-resource allocation for the proportional fairness the results are not satisfactory which means that proportional fairness is not well suited to our multi-resource problem. PF has two main problems: 1) wasted resources and 2) limited system stability because of the wasted resources.

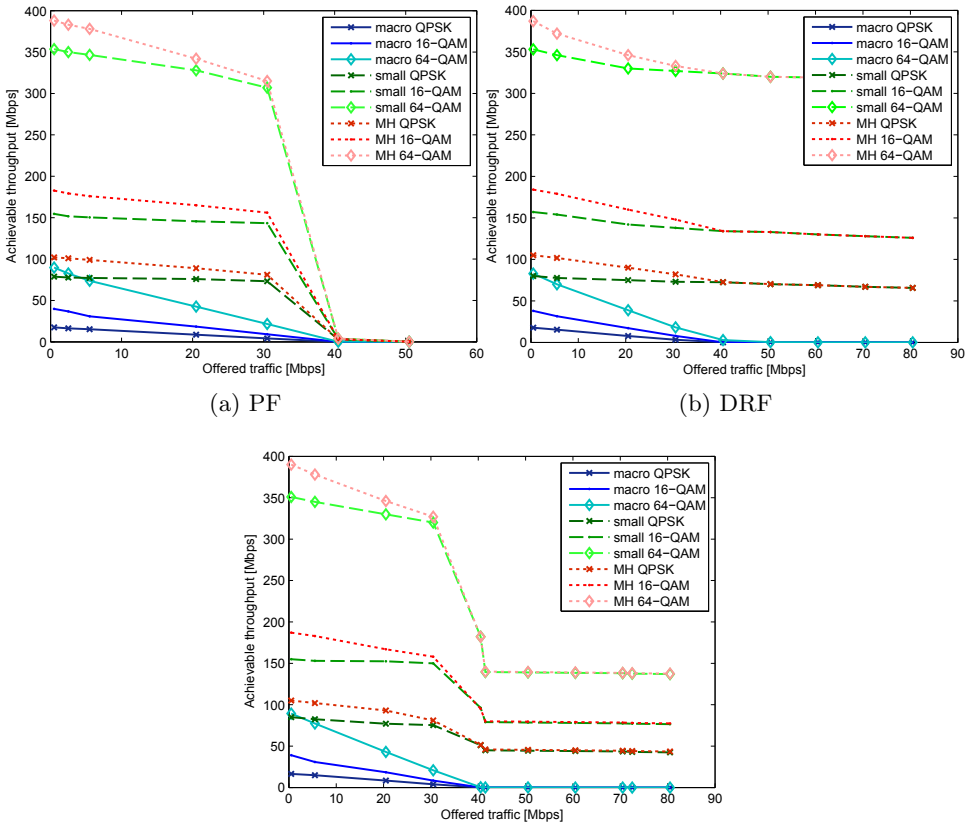
B. V-RAN with limiting resources

In this section, we study the case of limited processing resources in a V-RAN with multihomed users. Here, we present a comparison of the average data rate obtained by different multi-resource allocation strategies.

Fig. 4.12 shows the average achievable throughput by the three allocation strategies. The obtained results show that for low load, PF outperforms water-filling (DRF) and TPA, while at medium to high loads, DRF and TPA achieve comparable results. In this case of restrictive V-RAN, DRF outperforms both PF and TPA.

C. Jain Fairness Index comparison

Fig. 4.13 shows the Jain's fairness index variation of each allocation strategy as a function of the offered traffic when multihoming is considered. For sufficient processing resources at the V-RAN, we see in Fig. 4.13a that all allocation strategies achieve the same fairness at low load. At higher loads, PF offers the highest fairness index, while water-filling offers the lowest one. TPA outperforms PF with highest system stability and outperforms



(a) PF

(b) DRF

(c) TPA

Figure 4.10: Comparison of: (4.10a) proportional fairness (PF), (4.10b) dominant resource fairness (DRF) and (4.10c) two-phase allocation strategies' achievable throughput of different classes of users when V-RAN has sufficient processing resources in case of multihoming.

water-filling with highest fairness index. For V-RAN with limiting resources, the difference between the fairness values becomes clearer in Fig. 4.13b. Whereas PF always offers the highest fairness and water-filling the lowest one. An operator can choose PF, TPA, or water-filling by prioritizing the achievable rate, fairness or stability.

4.6.4 Power consumption evaluation

In the following simulations, we consider the case of one macro cell and $K = 20$ mm-wave small cells with sufficient processing resources. We consider $K = 20$ small cells instead of 3 that was adopted previously in order to focus on the energy efficiency in a very dense HetNet deployment. We also disregard the heterogeneity between users classes in the network in order to

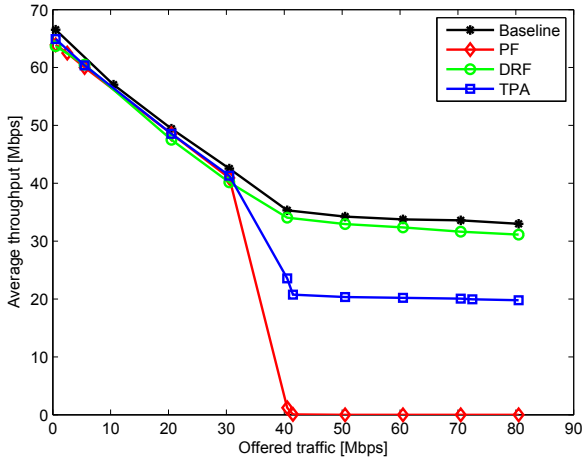


Figure 4.11: Comparing average throughput of different strategies when V-RAN has sufficient resources and with multihoming.

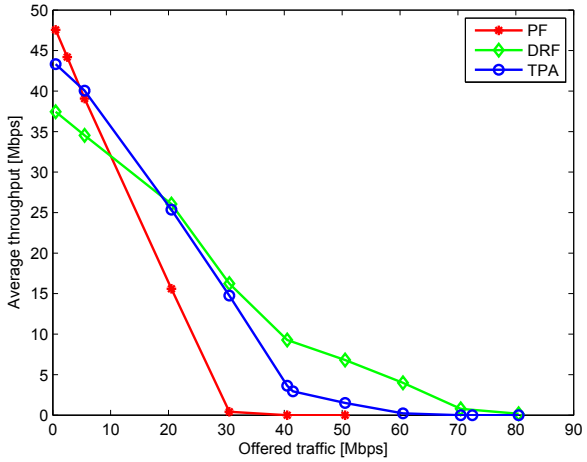
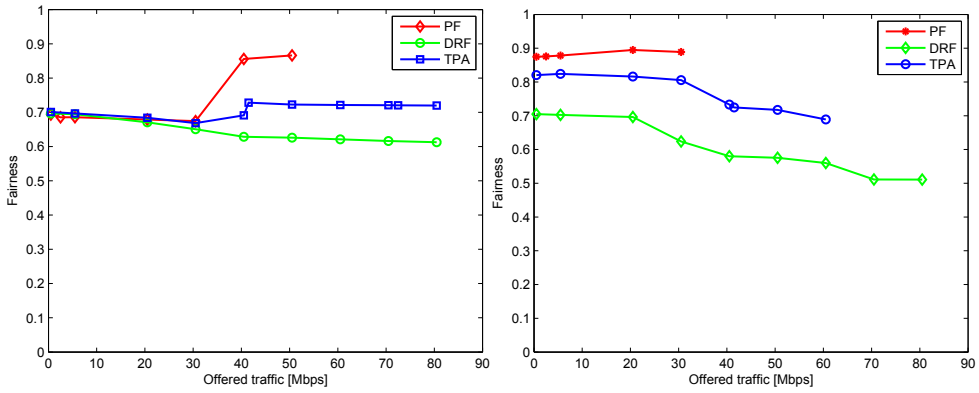


Figure 4.12: Comparing average throughput of different strategies when V-RAN is restrictive and with multihoming.



(a) Fairness index vs. the offered traffic with sufficient V-RAN resources.

(b) Fairness index vs. the offered traffic with limiting V-RAN resources.

Figure 4.13: Fairness index, system with multihoming.

reduce system's complexity.

The previously adopted parameters such as macro and small cell radii, frequency bands, bandwidths, processing efficiencies remain the same in this section. We also keep the parameters presented in Tables 4.1 and 4.2.

Given that we assume in this section only one radio condition for the LTE macro cell users (CQI= 13), and only one radio condition for the mm-wave small cell users (CQI= 23), the flow arrivals are modified such that

$$\lambda = \hat{p}^{M}\lambda + \sum_{k} \hat{p}^{S}\lambda + \sum_{k} \hat{p}^{MH}\lambda \tag{4.49}$$

where small cell users and multihomed users arrival rates are obtained as a function of the small cell and the macro cell radii, r_{SC} and r_{MC} respectively, and $\hat{p}^{M}$,$\hat{p}^{S}$ and $\hat{p}^{MH}$ are given by

$$\hat{p}^{S} = \hat{p}^{MH} = (r_{SC}/r_{MC})^{2} \tag{4.50}$$

$$\hat{p}^{M} = 1 - 2K\hat{p}^{S} \tag{4.51}$$

As for the power consumption parameters, we present in table 4.3 the macro cell's power consumption parameters.

Table 4.3: Power model parameters

Power (Watts)	N_{TRX}	P_{max} [W]	P_0[W]	Δ_p	P_{sleep}
RRH	6	20	84	2.8	56

According to [115], the small cell RRH power consumption is constant and equal to $P_{SC}^{k} = 60W$.

And a BBU consumes $P_{BBU} = 120W$ and is able to process 324 GOPS.

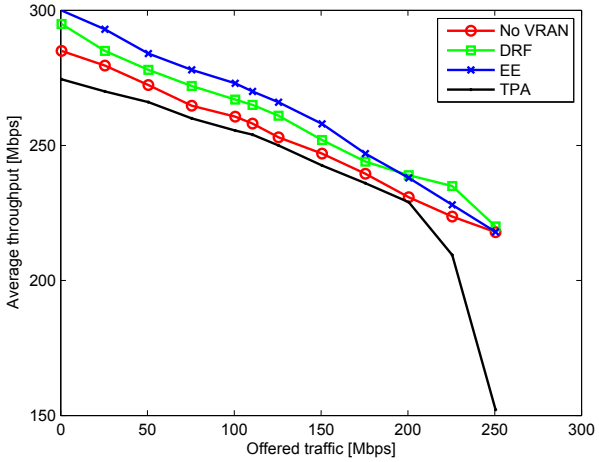


Figure 4.14: Comparing average throughput variation under different allocation strategies.

A. Comparing average data rates

Figure 4.14 presents the average data rate as a function of the offered traffic for each allocation strategy.

When no V-RAN is considered, we apply the local PF strategy, the average throughput in this case serves as a baseline to which we compare other strategies. We observe that at low to medium load, the energy efficient strategy (EE) achieves the highest average throughput with 4.26% higher than the baseline, while DRF achieves an average of 2.35% higher than the baseline. While at high load, DRF achieves the highest performance while TPA decreases rapidly to 30% lower average throughput than the baseline without DRF.

B. Comparing power consumptions

In Fig. 4.15, we compare the total power consumption as a function of the offered traffic for the above-mentioned allocation strategies. We observe that the system virtualization diminishes significantly the total power consumption due to the variable assignment of BBUs to the access networks. Depending on the network load and on the adopted strategy, the total power consumption in the network is reduced between 40% to 93% of the total power consumed in the system without virtualization.

We found that the EE strategy achieves similar power consumption with DRF at low load, higher consumption at medium load ($100-200$ Mbps), and lower consumption at high load ($200-250$ Mbps). While TPA achieves lower consumption at low to medium loads, then requires higher consumption than

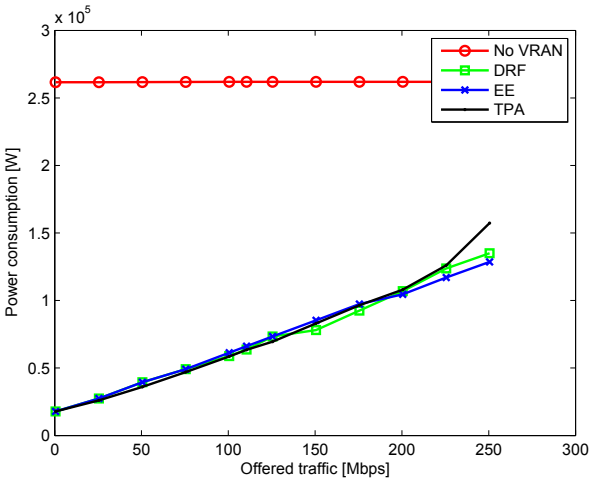


Figure 4.15: Comparing power consumption variation under different allocation strategies.

both EE and DRF at higher load.

Now if we combine both data rate and total power variations, we find that the energy efficient strategy achieves the best trade-off between the energy consumption and the average data rate at low to medium load. While DRF achieves the better trade-off at higher load.

C. Comparing Dynamic DRF and Energy Efficient allocation

In this section, we aim to figure out the reason behind the obtained performance for DRF and EE strategies. Generally, energy efficient maximization is achieved either by increasing the users achievable throughput or by reducing the power consumption. When comparing the achievable data rate for each class of users in Fig. 4.16, we find that the energy efficient strategy offers a higher throughput for small cell users, but at the expense of lower throughput for macro cell users. Even though small cell users require higher processing, the energy efficient strategy assigns more resources to small cell users whenever possible trying to reduce power consumption by reducing the number of users. This is possible because this allocation strategy takes both energy and throughput into consideration, contrary to DRF which focuses on the throughput only. However, at high load, the macro cell becomes unstable when using the EE strategy, which results in moving multihomed users load on the small cells, and reducing thus small cell users' throughput. This also was translated by an increased power consumption as seen in figure 4.15.

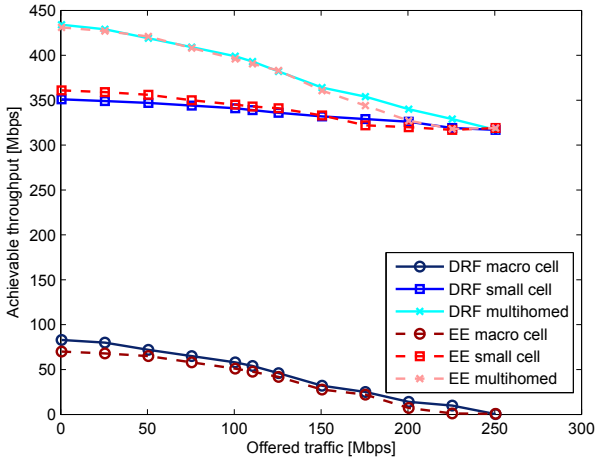


Figure 4.16: Comparing achievable data rate by each class of users for both DRF and energy efficient allocation strategies.

D. Jain fairness index comparison

Despite the higher average throughput offered by EE, we show in Fig. 4.17 that it offers a lower fairness than DRF and TPA. In addition, it is worth to note that TPA achieves comparable results at low to medium load with DRF and EE in terms of average throughput and power consumption, but it surely achieves a higher fairness between the users.

4.7 Conclusion

We presented in this chapter different resource allocation strategies under different virtual radio access networks scenarios. Three allocation strategies were considered: proportional fairness (PF), dominant resource fairness (DRF), and a newly proposed two-phase heuristic allocation (TPA).

We started with a V-RAN model without dual connectivity, i.e., multi-homing, and then considered dual connectivity in the network. We jointly allocated radio resources as well as processing ones. For both cases when V-RAN has limiting and sufficient processing resources, we found that DRF outperforms other strategies in terms of achievable throughput, PF offers the highest fairness, and TPA offers a trade-off between fairness and achievable throughput and maintains system's stability. We also found that when using PF, V-RAN limits small cells performance while DRF maintains a good performance and stability for these small cells.

By comparing these strategies to the baseline network performance when no virtualization is considered, we found that DRF achieves comparable throughput and stability in the system allowing thus to virtualize the HetNet

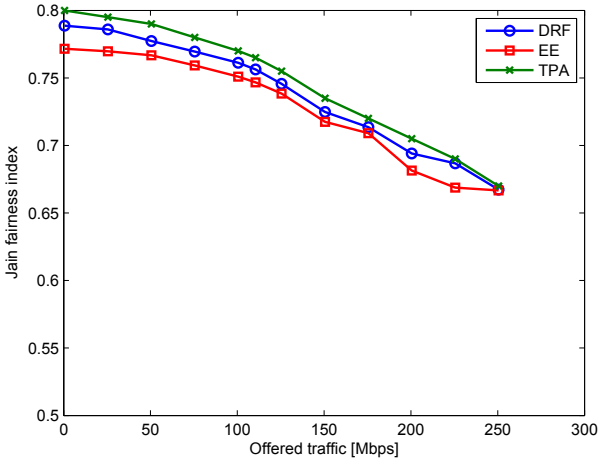


Figure 4.17: Jain fairness index vs. the offered traffic.

without performance loss.

Knowing that 5G is being designed to be a green network, we studied V-RAN's energy efficiency with multihoming, and this by adopting multi-resource allocation strategies taking into consideration two parameters instead of only one: the achievable throughput and the total power consumption in the network. We compared local PF applied to a baseline network model without V-RAN to the network model with V-RAN using DRF, TPA, and energy efficient allocation using Dinklebach's algorithm. We considered only sufficient processing resources case by allocating BBUs dynamically to the RRHs.

The energy efficient strategy has been shown to outperform all other strategies at low to average load both in terms of increased average throughput and reduced power consumption, which means reduced network OPEX. However, DRF outperformed the other strategies at high load. When comparing the jain fairness index, TPA outperformed other strategies in terms of fairness, and the energy efficient strategy has shown the lowest fairness index.

Chapter 5

Conclusion and Perspectives

5.1 Thesis summary

Since the introduction of mobile broadband in 3G cellular networks in 2001, users are continuously asking for more services with increasing capacity and latency requirements. Nowadays, the first 5G network will see the light by 2020 offering a large number of services with huge capacity and ultra-reliable communication. Until then, a rigorous work is needed from researchers to address all possible topics in 5G networks.

The increasing number of connected devices requires a larger number of small cells deployment accompanied with wireless LAN proliferation in houses, offices, malls and almost any indoor place. This coexistence of several access technologies could be exploited in order to achieve higher data rates than that could be achieved by single access networks. By adopting the multihoming capability in a network, multihomed users reach higher data rates than single-homed ones.

The first contribution of this thesis focused on evaluating a HetNet with multihoming capability. This work modeled resource allocation strategies by applying: (1) network centric approaches: local and global proportional fairness and (2) user centric approaches: peak rate maximization and network assisted strategies. Our results showed that the network assisted strategy offers a global network optimum. We also found that global PF achieves a better performance than user centric strategies in terms of throughput at the expense of higher computational requirements.

However, the usage of dense HetNets increases significantly cells implementation and upgrade costs, i.e., CAPEX. This cost can be reduced by considering a 5G network with C-RAN/V-RAN. In this case, the base station functionalities are split between cell locations (RRH) and the centralized pool (BBU pool). The challenge here is to choose a joint resource allocation strategy that considers both radio and processing resource requirements. For this aim, we considered three joint-allocation strategies: proportional

fairness (PF), dominant resource fairness (DRF), and proposed a new two-phase heuristic allocation (TPA). We notably showed that DRF outperforms other strategies in terms of achievable throughput when V-RAN's processing resources are limiting or sufficient while maintaining good performance and stability for small cells contrary to PF. We also showed that TPA achieves a trade-off between PF and DRF by offering higher throughput than PF and ensuring higher fairness than DRF. Furthermore, we showed that DRF is a good choice for RAN virtualization because it offers comparable results in terms of throughput and stability with the baseline network model (without virtualization).

Finally, and because V-RAN is intended to be a green solution, we extended our investigation to an energy efficiency study of power consumption in 5G V-RAN aiming to achieve a lower OPEX. We considered the previous V-RAN resource allocation strategies, DRF and TPA, as well as an energy efficient allocation strategy using Dinklebach's algorithm and the baseline case where no-virtualization is applied. We added a power component to the joint allocation strategies, and compared both throughput and power consumption metrics. At low to medium load, the energy efficient strategy using Dinklebach's algorithm outperformed all remaining strategies in terms of higher average throughput while having comparable total power consumption. However, at high system load, DRF outperformed TPA and the energy efficient strategy in terms of higher throughput, but the energy efficient strategy achieves the lowest power consumption. We also showed that TPA outperforms both strategies in terms of users' allocation fairness.

5.2 Future research perspectives

We explored in this work the resource allocation strategies applied to different 5G architecture scenarios where the users' classes are split between multihomed and single-homed. While we considered in this work only elastic traffic, 5G services consist of several types of traffic: real-time, circuit, elastic, and multi-media streaming. A direct extension would be the investigation of real-time traffic in 5G where resource allocation strategies will require additional real-time constraints.

In addition to the multihoming and pooling design considerations in 5G V-RAN, caching is yet another feature that must be considered. Caching can improve users' quality of experience by bringing storage functionality to the network edges instead of centralized servers reducing thus latency.

Moreover, it would be very interesting to evaluate the studied network scenarios and technologies by the means of testbeds in order to validate our results. An experimental framework for a HetNet with mm-wave small cells as well as a virtualized network can be used to study multihoming concepts in 5G systems.

Finally, we believe that 5G infrastructure should not be limited to the case of one operator only. It would be interesting to explore the coexistence of different operators either cooperative or in competition. We aim in our future works to explore different sharing configurations and pricing models for the services offered to the users.

We detail in the following the previously introduced research perspectives.

5.2.1 Real-time traffic in 5G

Deploying 5G mobile networks requires a highly enhanced mobile broadband experience. The range of applications and services that can be deployed include ultra-reliable communication services with enhanced latency [3]. In addition to the achieved higher throughput, we aim to add a latency constraint, and evaluate resource allocation in 5G networks with both constraints.

With the adoption of mm-wave in 5G networks [156], a round trip latency of 1 ms is required [157]. However, there is very little work explaining ways of achieving this stringent requirement. Our aim consists mainly on modeling the real-time traffic [158] in 5G networks at the flow level and finding the capacity shares that ensure a good user experience.

5.2.2 Caching in V-RAN

Another strategy to decrease the latency in delay-sensitive content retrieval is using content caching in the V-RAN. In online social networks, users tend to choose contents that are recommended by friends, or currently trending. This content can be cached beforehand [159, 160].

In [161], edge caching has been developed for small cell networks. However, this work considered the case of single-homed access, while multihomed access is possible. Consequently, we aim to study the impact of multihoming on V-RAN edge caching. We specifically focus on a heterogeneous network where mm-wave small cells coexist with LTE macro cell.

5.2.3 V-RAN testbed

To further examine the feasibility of V-RAN 5G system, we find it interesting to conduct a series of experiments on a realistic RAN system [162]. This could be possible by virtualizing an eNodeB into RRH and BBU, with evaluating the joint allocation strategies in order to verify the obtained results, and observe how a real system would perform under different load conditions with mm-wave and multihoming. In order to demonstrate the efficiency of V-RANs in practice, a large-size testbed with corresponding trial tests for V-RANs must be considered.

5.2.4 Economical aspects

Although the focus was on 5G architecture performance evaluation with multihoming applied to different scenarios of macro and small cells, including LTE, WiFi and mm-wave, the follow-up to this work could focus on the economical aspects as well [163]. We aim to consider the case of a multi-tenant cooperative network, where several operators coexist [164], and where the profit sharing between different operators can for instance be modeled function of the subscription revenues, the infrastructure and operation costs using economical strategies stemming from coalition games [165] and Shapely value [166].

Bibliography

[1] Nikola Tesla. Nikola tesla sees a wireless vision. `http://www.tfcbooks.com/tesla/1915-10-03.htm`. [Accessed: Apr. 26, 2017].

[2] Ericsson White Paper. 5g radio access, capabilities and technologies. http://www.ericsson.com/res/docs/whitepapers/wp-5g.pdf, Apr. 2016. [Accessed: Feb. 17, 2017].

[3] M. Agiwal, A. Roy, and N. Saxena. Next generation 5g wireless networks: A comprehensive survey. *IEEE Communications Surveys Tutorials*, 18(3):1617–1655, thirdquarter 2016.

[4] 5G PPP Architecture Working Group. View on 5g architecture. White Paper, Jul. 2016.

[5] NGMN Alliance. 5G White Paper – Final Deliverable. Technical report, White Paper, Feb. 2015.

[6] Jose F. Monserrat, Genevieve Mange, Volker Braun, Hugo Tullberg, Gerd Zimmermann, and Ömer Bulakci. Metis research advances towards the 5g mobile and wireless system definition. *EURASIP Journal on Wireless Communications and Networking*, 2015(1):53, 2015.

[7] A. Kostopoulos, G. Agapiou, F. C. Kuo, K. Pentikousis, A. Cipriano, D. Panaitopol, D. Marandin, K. Kowalik, K. Alexandris, C. Y. Chang, N. Nikaein, M. Goldhamer, A. Kliks, R. Steinert, A. Mämmelä, and T. Chen. Scenarios for 5g networks: The coherent approach. In *2016 23rd International Conference on Telecommunications (ICT)*, pages 1–6, May 2016.

[8] Qualcomm. Initial concepts on 5G architecture and integration. *Deliverable D3.1*, 2016.

[9] FCC. FCC Increases 5GHz Spectrum for Wi-Fi, Other Unlicensed Uses. `https://www.fcc.gov/document/fcc-increases-5ghz-spectrum-wi-fi-other-unlicensed-uses`, Mar. 2014. [Accessed: Feb. 25, 2017].

[10] Ieee standard for information technology - telecommunications and information exchange between systems - local and metropolitan networks - specific requirements - part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications: Higher speed physical layer (phy) extension in the 2.4 ghz band. *IEEE Std 802.11b-1999*, pages 1–96, Jan. 2000.

[11] S. G. Sankaran, B. J. Zargari, L. Y. Nathawad, H. Samavati, S. S. Mehta, A. Kheirkhahi, P. Chen, K. Gong, B. Vakili-Amini, J. A. Hwang, S. W. M. Chen, M. Terrovitis, B. J. Kaczynski, S. Limotyrakis, M. P. Mack, H. Gan, M. Lee, R. T. Chang, H. Dogan, S. Abdollahi-Alibeik, B. Baytekin, K. Onodera, S. Mendis, A. Chang, Y. Rajavi, S. H. M. Jen, D. K. Su, and B. A. Wooley. Design and implementation of a cmo 802.11n soc. *IEEE Communications Magazine*, 47(4):134–143, Apr. 2009.

[12] Iso/iec/ieee international standard - information technology – telecommunications and information exchange between systems – local and metropolitan area networks – specific requirements – part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications amendment 4. *ISO/IEC/IEEE 8802-11:2012/Amd.4:2015(E) (Adoption of IEEE Std 802.11ac-2013)*, pages 1–430, Aug. 2015.

[13] Iso/iec/ieee international standard for information technology–telecommunications and information exchange between systems–local and metropolitan area networks–specific requirements-part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications amendment 3: Enhancements for very high throughput in the 60 ghz band (adoption of ieee std 802.11ad-2012). *ISO/IEC/IEEE 8802-11:2012/Amd.3:2014(E)*, pages 1–634, Mar. 2014.

[14] E. Perahia, C. Cordeiro, M. Park, and L. L. Yang. Ieee 802.11ad: Defining the next generation multi-gbps wi-fi. In *2010 7th IEEE Consumer Communications and Networking Conference*, pages 1–5, Jan. 2010.

[15] 3GPP. (E-UTRAN), Overall description, Stage 2 (Release 12). *TS36.300, v12.6.0.*

[16] ETSI TR 101-957. Requirements and architectures for interworking between hiperlan/2 and 3G cellular systems. Technical report, ETSI, 2001. Online; accessed 24-Jan-2016.

[17] Alcatel-Lucent. Wifi roaming-building on andsf and hotspot 2.0, 2012.

[18] 3GPP. Architecture enhancements for non-3gpp accesses. Technical specification TS 23.402, 2012. Release 10.

[19] Victor C. M. Leung. *Multihomed Communication with SCTP (Stream Control Transmission Protocol)*. Auerbach Publications, Boston, MA, USA, 2013.

[20] X. Lagrange. Very tight coupling between lte and wi-fi for advanced offloading procedures. In *2014 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, pages 82–86, Apr. 2014.

[21] 3GPP TS 36.932. Scenarios and requirements for small cell enhancements for eutra and eutran. Release 12, Version 12.1.0, Oct. 2014.

[22] J. Robson. A white paper by the ngmn alliance: Small cell backhaul requirements. Next Generation Mobile Networks, Jun. 2012.

[23] 3GPP. LTE; Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description, Stage 2. TS 36.300, v 12.7.0, Oct. 2015.

[24] B. Bangerter, S. Talwar, R. Arefi, and K. Stewart. Networks and devices for the 5g era. *IEEE Communications Magazine*, 52(2):90–96, Feb. 2014.

[25] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang. What will 5g be? *IEEE Journal on Selected Areas in Communications*, 32(6):1065–1082, Jun. 2014.

[26] F. Khan, Z. Pi, and S. Rajagopal. Millimeter-wave mobile broadband with large scale spatial processing for 5g mobile communication. In *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1517–1523, Oct. 2012.

[27] O. El Ayach, S. Rajagopal, S. Abu-Surra, Zhouyue Pi, and R.W. Heath. Spatially sparse precoding in millimeter wave mimo systems. *Wireless Communications, IEEE Transactions on*, pages 1499–1513, Mar. 2014.

[28] Z. Pi and F. Khan. An introduction to millimeter-wave mobile broadband systems. *IEEE Communications Magazine*, 49(6):101–107, Jun. 2011.

[29] H. Peng, T. Yamamoto, and Y. Suegara. Extended user/control plane architectures for tightly coupled lte/wigig interworking in millimeter-wave heterogeneous networks. In *2015 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1548–1553, Mar. 2015.

[30] 3GPP TS 36.300. E-UTRA and E-UTRAN, overall description. v12.1.0.

[31] Wonyong Yoon and Beakcheol Jang. Enhanced non-seamless offload for lte and wlan networks. *Communications Letters, IEEE*, 17(10):1960–1963, Oct. 2013.

[32] M. Simsek, M. Bennis, M. Debbah, and A. Czylwik. Rethinking offload: How to intelligently combine wifi and small cells? In *2013 IEEE International Conference on Communications (ICC)*, pages 5204–5208, Jun. 2013.

[33] Byoung Hoon Jung, Nah-Oak Song, and Dan Keun Sung. A network-assisted user-centric wifi-offloading model for maximizing per-user throughput in a heterogeneous network. *Vehicular Technology, IEEE Transactions on*, 63, May 2014.

[34] Ruiming Yang, Yongyu Chang, Jia Sun, and Dacheng Yang. Traffic split scheme based on common radio resource management in an integrated lte and hsdpa networks. In *Vehicular Technology Conference (VTC Fall), 2012 IEEE*, pages 1–5, Sept. 2012.

[35] G. Aristomenopoulos, T. Kastrinogiannis, and S. Papavassiliou. Multiaccess multi-cell distributed resource management framework in heterogeneous wireless networks. *Vehicular Technology, IEEE Transactions on*, 61(6):2636–2650, Jul. 2012.

[36] Peng Xue, Peng Gong, Jae Hyun Park, Daeyoung Park, and Duk Kyung Kim. Radio resource management with proportional rate constraint in the heterogeneous networks. *Wireless Communications, IEEE Transactions on*, 11(3):1066–1075, Mar. 2012.

[37] M.C. Lucas-Estañ and J. Gozalvez. On the real-time hardware implementation feasibility of joint radio resource management policies for heterogeneous wireless networks. *Mobile Computing, IEEE Transactions on*, 12(2):193–205, Feb. 2013.

[38] N. Capela and S. Sargento. Optimizing network performance with multihoming and network coding. In *Globecom Workshops (GC Wkshps), 2012 IEEE*, pages 210–215, Dec. 2012.

[39] D. Jurca and P. Frossard. Media flow rate allocation in multipath networks. *Multimedia, IEEE Transactions on*, 9(6):1227–1240, Oct. 2007.

[40] Loss tolerant bandwidth aggregation for multihomed video streaming over heterogeneous wireless networks. *Wireless Personal Communications*, 75(2):1265–1282, 2014.

[41] Fanglu Guo, Jiawu Chen, Wei Li, and Tzi cker Chiueh. Experiences in building a multihoming load balancing system. In *IEEE INFOCOM 2004*, volume 2, pages 1241–1251 vol.2, Mar. 2004.

[42] H. Fujinoki. Improving reliability for multi-home inbound traffic: Mhlb/i packet-level inter-domain load-balancing. In *2009 International Conference on Availability, Reliability and Security*, pages 248–256, Mar. 2009.

[43] Zunli Yang, Qinghai Yang, Fenglin Fu, and Kyung Sup Kwak. A novel load balancing scheme in lte and wifi coexisted network for ofdma system. In *2013 International Conference on Wireless Communications and Signal Processing*, pages 1–5, Oct. 2013.

[44] A. Habib and J. Chuang. Multihoming media streaming. In *Performance, Computing, and Communications Conference, 2005. IPCCC 2005. 24th IEEE International*, pages 499–504, Apr. 2005.

[45] 3GPP TR 23.829 v10.0.0. Local ip access and selected ip traffic offload (lipa-sipto) (release 10).

[46] Ruckus. Interworking wi-fi and mobile networks. *Ruckus Wireless Incorporation, U.S.*, 2013.

[47] W. Yoon and B. Jang. Enhanced non-seamless offload for lte and wlan networks. *IEEE Communications Letters*, 17(10):1960–1963, Oct. 2013.

[48] P. Nikander, A. Gurtov, and T. R. Henderson. Host identity protocol (hip): Connectivity, mobility, multi-homing, security, and privacy over ipv4 and ipv6 networks. *IEEE Communications Surveys Tutorials*, 12(2):186–204, Second 2010.

[49] R. Stewart and C. Metz. Sctp: new transport protocol for tcp/ip. *IEEE Internet Computing*, 5(6):64–69, Nov. 2001.

[50] R. Stewart. Stream control transmission protocol. In *RFC*.

[51] M. Handley S. Barre A. Ford, C. Raiciu and J. Iyengar. Architectural guidelines for multipath tcp development. *IETF*, 2011.

[52] R. Zhang, Z. Zheng, M. Wang, X. Shen, and L. L. Xie. Equivalent capacity in carrier aggregation-based lte-a systems: A probabilistic analysis. *IEEE Transactions on Wireless Communications*, 13(11):6444–6460, Nov. 2014.

[53] S. C. Jha, K. Sivanesan, R. Vannithamby, and A. T. Koc. Dual connectivity in lte small cell networks. In *2014 IEEE Globecom Workshops (GC Wkshps)*, pages 1205–1210, Dec. 2014.

[54] A. Mukherjee. Macro-small cell grouping in dual connectivity lte-b networks with non-ideal backhaul. In *2014 IEEE International Conference on Communications (ICC)*, pages 2520–2525, Jun. 2014.

[55] A. Zakrzewska, D. López-Pérez, S. Kucera, and H. Claussen. Dual connectivity in lte hetnets with split control- and user-plane. In *2013 IEEE Globecom Workshops (GC Wkshps)*, pages 391–396, Dec. 2013.

[56] H. Wang, C. Rosa, and K. I. Pedersen. Inter-enb flow control for heterogeneous networks with dual connectivity. In *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, pages 1–5, May 2015.

[57] A. Mukherjee. Optimal flow bifurcation in networks with dual base station connectivity and non-ideal backhaul. In *2014 48th Asilomar Conference on Signals, Systems and Computers*, pages 521–524, Nov. 2014.

[58] Zhefeng Jiang and S. Mao. Access strategy and dynamic downlink resource allocation for femtocell networks. In *2013 IEEE Global Communications Conference (GLOBECOM)*, pages 3528–3533, Dec. 2013.

[59] D. Lopez-Perez, X. Chu, A. V. Vasilakos, and H. Claussen. Power minimization based resource allocation for interference mitigation in ofdma femtocell networks. *IEEE Journal on Selected Areas in Communications*, 32(2):333–344, Feb. 2014.

[60] 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*, 54(6):137–143, Jun. 2016.

[61] P. Legg, P. Fotiadis, and P. Soldati. Load balancing and aggregation algorithms for lte dual connectivity. In *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, pages 1–5, May 2016.

[62] B. Liu, N. Boukhatem, P. Martins, and P. Bertin. Multihoming at layer-2 for inter-rat handover. In *21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, pages 1173–1178, Sept. 2010.

[63] M. Z. Chowdhury, Y. M. Jang, Choong Sub Ji, Sunwoong Choi, Hongseok Jeon, Junghoon Jee, and C. Park. Interface selection for power management in umts/wlan overlaying network. In *2009 11th International Conference on Advanced Communication Technology*, volume 01, pages 795–799, Feb. 2009.

[64] S. E. Elayoubi and M. Francisco. Capacity of hierarchical wifi/wimax networks. In *2009 IEEE International Conference on Communications*, pages 1–5, Jun. 2009.

[65] G. I. Naik. Lte wlan interworking for wi-fi hotspots. In *2010 Second International Conference on COMmunication Systems and NETworks (COMSNETS 2010)*, pages 1–2, Jan. 2010.

[66] A. Y. Panah, S. p. Yeh, N. Himayat, and S. Talwar. Utility-based radio link assignment in multi-radio heterogeneous networks. In *2012 IEEE Globecom Workshops*, pages 618–623, Dec. 2012.

[67] Ehab Mahmoud Mohamed. Cloud cooperated heterogeneous cellular networks for delayed offloading using millimeter wave gates. *International Journal of Electronics and Telecommunications*, 63(1):51–64, 2017.

[68] S. p. Yeh, A. Y. Panah, N. Himayat, and S. Talwar. Qos aware scheduling and cross-radio coordination in multi-radio heterogeneous networks. In *2013 IEEE 78th Vehicular Technology Conference (VTC Fall)*, pages 1–6, Sept. 2013.

[69] L. Hu, L. L. Sanchez, M. Maternia, I. Z. Kovacs, B. Vejlgaard, P. Mogensen, and H. Taoka. Modeling of wi-fi ieee 802.11ac offloading performance for 1000x capacity expansion of lte-advanced. In *2013 IEEE 78th Vehicular Technology Conference (VTC Fall)*, pages 1–6, Sept. 2013.

[70] M. Ismail and W. Zhuang. A distributed multi-service resource allocation algorithm in heterogeneous wireless access medium. *IEEE Journal on Selected Areas in Communications*, 30(2):425–432, Feb. 2012.

[71] D. H. Bui, K. Lee, S. Oh, I. Shin, H. Shin, H. Woo, and D. Ban. Greenbag: Energy-efficient bandwidth aggregation for real-time streaming in heterogeneous mobile wireless networks. In *2013 IEEE 34th Real-Time Systems Symposium*, pages 57–67, Dec. 2013.

[72] F. Juanjuan, C. Shanzhi, H. Bo, and S. Yan. An autonomic interface selection method for multi-interfaces mobile terminal in heterogeneous environment. In *2009 WRI World Congress on Computer Science and Information Engineering*, volume 5, pages 107–111, Mar. 2009.

[73] M. Ismail, A. Abdrabou, and W. Zhuang. Cooperative decentralized resource allocation in heterogeneous wireless access medium. *IEEE Transactions on Wireless Communications*, 12(2):714–724, Feb. 2013.

[74] 3GPP TS 24.312. Access Network Discovery and Selection Function (ANDSF) Management Object (MO). (Release 8).

[75] S. E. Elayoubi, E. Altman, M. Haddad, and Z. Altman. A hybrid decision approach for the association problem in heterogeneous networks. In *2010 Proceedings IEEE INFOCOM*, pages 1–5, Mar. 2010.

[76] E. Altman, P. Wiecek, and M. Haddad. The association problem with misleading partial channel state information. In *2012 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 2242–2246, Apr. 2012.

[77] Alexandros Kaloxylos, Ioannis Modeas, Fotos Georgiadis, and Nikos Passas. Network selection algorithm for heterogeneous wireless networks: From design to implementation. *Network Protocols and Algorithms*, 1(2):27–47, 2010.

[78] Niroshinie Fernando, Seng W. Loke, and Wenny Rahayu. Mobile cloud computing: A survey. *Future Generation Computer Systems*, 29(1):84 – 106, 2013. Including

Special section: AIRCC-NetCoM 2009 and Special section: Clouds and Service-Oriented Architectures.

[79] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann. Cloud RAN for Mobile Networks – A Technology Overview. *IEEE Communications Surveys Tutorials*, 17(1):405–426, Firstquarter 2015.

[80] C-ran the road towards green ran. Technical report, China Mobile Research Institute, 2011.

[81] Kimio Watanabe and Mamoru Machida. Outdoor lte infrastructure equipment (enodeb). *Fujitsu Sci. Tech. J*, 48(1):27–32, 2012.

[82] S. Deb, P. Monogioudis, J. Miernik, and J. P. Seymour. Algorithms for enhanced inter-cell interference coordination (eicic) in lte hetnets. *IEEE/ACM Transactions on Networking*, 22(1):137–150, Feb. 2014.

[83] M. Sawahashi, Y. Kishiyama, A. Morimoto, D. Nishikawa, and M. Tanno. Coordinated multipoint transmission/reception techniques for lte-advanced [coordinated and distributed mimo]. *IEEE Wireless Communications*, 17(3):26–34, Jun. 2010.

[84] V. N. Ha, L. B. Le, and N. D. Đào. Coordinated multipoint transmission design for cloud-rans with limited fronthaul capacity constraints. *IEEE Transactions on Vehicular Technology*, 65(9):7432–7447, Sept. 2016.

[85] Navid Nikaein. Processing radio access network functions in the cloud: Critical issues and modeling. In *Proceedings of the 6th International Workshop on Mobile Cloud Computing and Services*, MCS '15, pages 36–43, New York, NY, USA, 2015. ACM.

[86] C. Desset, B. Debaillie, V. Giannini, A. Fehske, G. Auer, H. Holtkamp, W. Wajda, D. Sabella, F. Richter, M. J. Gonzalez, H. Klessig, I. Gódor, M. Olsson, M. A. Imran, A. Ambrosy, and O. Blume. Flexible power modeling of lte base stations. In *2012 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 2858–2862, Apr. 2012.

[87] K. Guo, M. Sheng, J. Tang, T. Q. S. Quek, X. Wang, and Z. Qiu. Cooperative transmission meets computation provisioning in downlink c-ran. In *2016 IEEE International Conference on Communications (ICC)*, pages 1–6, May 2016.

[88] D. Wubben, P. Rost, J. S. Bartelt, M. Lalam, V. Savin, M. Gorgoglione, A. Dekorsy, and G. Fettweis. Benefits and impact of cloud computing on 5g signal processing: Flexible centralization through cloud-ran. *IEEE Signal Processing Magazine*, 31(6):35–44, Nov. 2014.

[89] S. Xu and S. Wang. Efficient algorithm for baseband unit pool planning in cloud radio access networks. In *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, pages 1–5, May 2016.

[90] M. A. Marotta, N. Kaminski, I. Gomez-Miguelez, L. Z. Granville, J. Rochol, L. DaSilva, and C. B. Both. Resource sharing in heterogeneous cloud radio access networks. *IEEE Wireless Communications*, 22(3):74–82, Jun. 2015.

[91] W. Ni and I. B. Collings. A new adaptive small-cell architecture. *IEEE Journal on Selected Areas in Communications*, 31(5):829–839, May 2013.

[92] C. L. Liu and James W. Layland. Scheduling algorithms for multiprogramming in a hard-real-time environment. *J. ACM*, 20(1):46–61, Jan. 1973.

[93] H. Yin and H. Liu. An efficient multiuser loading algorithm for ofdm-based broadband wireless systems. In *Global Telecommunications Conference, 2000. GLOBECOM '00. IEEE*, volume 1, pages 103–107 vol.1, 2000.

[94] I. C. Wong and B. L. Evans. Optimal resource allocation in the ofdma downlink with imperfect channel knowledge. *IEEE Transactions on Communications*, 57(1):232–241, Jan. 2009.

[95] W. Rhee and J. M. Cioffi. Increase in capacity of multiuser ofdm system using dynamic subchannel allocation. In *VTC2000-Spring. 2000 IEEE 51st Vehicular Technology Conference Proceedings (Cat. No.00CH37026)*, volume 2, pages 1085–1089 vol.2, 2000.

[96] Ellen L. Hahne and Robert G. Gallager. Round robin scheduling for fair flow control in data communication networks. In *ICC*, pages 103–107, 1986.

[97] D. K. H. Tan F. P. Kelly, A. K. Maulloo. Rate control for communication networks: Shadow prices, proportional fairness and stability. *The Journal of the Operational Research Society*, 49(3):237–252, 1998.

[98] J. Mo and J. Walrand. Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on Networking*, 8(5):556–567, Oct. 2000.

[99] M. Allman, V. Paxson, and W. Stevens. TCP Congestion Control RFC 2581. 1999.

[100] D. X. Wei, C. Jin, S. H. Low, and S. Hegde. Fast tcp: Motivation, architecture, algorithms, performance. *IEEE/ACM Transactions on Networking*, 14(6):1246–1259, Dec. 2006.

[101] S. K. Baruah, N. K. Cohen, C. G. Plaxton, and D. A. Varvel. Proportionate progress: A notion of fairness in resource allocation. In *Proceedings of the Twenty-fifth Annual ACM Symposium on Theory of Computing*, STOC '93, pages 345–354, New York, NY, USA, 1993. ACM.

[102] Tian Lan, David Kao, Mung Chiang, and Ashutosh Sabharwal. An axiomatic theory of fairness in network resource allocation. In *Proceedings of the 29th Conference on Information Communications*, INFOCOM'10, pages 1343–1351, Piscataway, NJ, USA, 2010. IEEE Press.

[103] Ali Ghodsi, Matei Zaharia, Benjamin Hindman, Andy Konwinski, Scott Shenker, and Ion Stoica. Dominant Resource Fairness: Fair Allocation of Multiple Resource Types. In *Proceedings of the 8th USENIX Conference on Networked Systems Design and Implementation*, NSDI'11, pages 323–336, Berkeley, CA, USA, 2011. USENIX Association.

[104] Danny Dolev, Dror G. Feitelson, Joseph Y. Halpern, Raz Kupferman, and Nathan Linial. No justified complaints: On fair sharing of multiple resources. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, ITCS '12, pages 68–75, New York, NY, USA, 2012. ACM.

[105] Thomas Bonald and James Roberts. Multi-resource fairness: Objectives, algorithms and performance. *SIGMETRICS Perform. Eval. Rev.*, 43(1):31–42, Jun. 2015.

[106] Ali Ghodsi, Vyas Sekar, Matei Zaharia, and Ion Stoica. Multi-resource fair queueing for packet processing. In *Proceedings of the ACM SIGCOMM 2012 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, SIGCOMM '12, pages 1–12, New York, NY, USA, 2012. ACM.

[107] Thomas Bonald and James Roberts. Enhanced cluster computing performance through proportional fairness. *Performance Evaluation*, 79:134 – 145, 2014. Special Issue: Performance 2014.

[108] W. Wang, B. Liang, and B. Li. Multi-resource fair allocation in heterogeneous cloud computing systems. *IEEE Transactions on Parallel and Distributed Systems*, 26(10):2822–2835, Oct. 2015.

[109] A. Fehske, G. Fettweis, J. Malmodin, and G. Biczok. The global footprint of mobile communications: The ecological and economic perspective. *IEEE Communications Magazine*, 49(8):55–62, Aug. 2011.

[110] Z. Hasan, H. Boostanimehr, and V. K. Bhargava. Green cellular networks: A survey, some research issues and challenges. *IEEE Communications Surveys Tutorials*, 13(4):524–540, Fourth 2011.

[111] C. L. I, C. Rowell, S. Han, Z. Xu, G. Li, and Z. Pan. Toward green and soft: a 5g perspective. *IEEE Communications Magazine*, 52(2):66–73, Feb. 2014.

[112] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume, and A. Fehske. How much energy is needed to run a wireless network? *IEEE Wireless Communications*, 18(5):40–49, Oct. 2011.

[113] Gunther Auer, Oliver Blume, Vito Giannini, Istvan Godor, M Imran, Ylva Jading, Efstathios Katranaras, Magnus Olsson, Dario Sabella, Per Skillermark, et al. D2. 3: Energy efficiency analysis of the reference systems, areas of improvements and target breakdown. *EARTH*, pages 1–69, 2010.

[114] Salvatore Chiaravalloti, Filip Idzikowski, Łukasz Budzisz, and Ing Adam Wolisz. Power consumption of wlan network elements. 2011.

[115] Gia Khanh Tran, Hidekazu Shimodaira, Roya Ebrahim Rezagah, Kei Sakaguchi, and Kiyomichi Araki. Practical evaluation of on-demand smallcell on/off based on traffic model for 5g cellular networks. In *Wireless Communications and Networking Conference (WCNC), 2016 IEEE*, pages 1–7. IEEE, 2016.

[116] Arsalan Saeed, Amir Akbari, Mehrdad Dianati, and Muhammad Ali Imran. Energy efficiency analysis for lte macro-femto hetnets. In *Wireless Conference (EW), Proceedings of the 2013 19th European*, pages 1–5, Apr. 2013.

[117] A. Prasad, A. Maeder, and C. Ng. Energy efficient small cell activation mechanism for heterogeneous networks. In *2013 IEEE Globecom Workshops (GC Wkshps)*, pages 754–759, Dec. 2013.

[118] I. Ashraf, F. Boccardi, and L. Ho. Sleep mode techniques for small cell deployments. *IEEE Communications Magazine*, 49(8):72–79, Aug. 2011.

[119] M. Ismail and W. Zhuang. Green radio communications in a heterogeneous wireless medium. *IEEE Wireless Communications*, 21(3):128–135, Jun. 2014.

[120] M. Ismail, E. Serpedin, and K. Qaraqe. A win-win cooperative downlink resource allocation for green communications in a heterogeneous wireless medium. In *2014 IEEE Globecom Workshops (GC Wkshps)*, pages 1115–1119, Dec. 2014.

[121] Q. D. Vu, L. N. Tran, M. Juntti, and E. K. Hong. Energy-efficient bandwidth and power allocation for multi-homing networks. *IEEE Transactions on Signal Processing*, 63(7):1684–1699, Apr. 2015.

[122] S. Kim, B. G. Lee, and D. Park. Energy-per-bit minimized radio resource allocation in heterogeneous networks. *IEEE Transactions on Wireless Communications*, 13(4):1862–1873, Apr. 2014.

[123] M. A. Khan, S. Leng, Wang Xiang, and Kun Yang. Architecture of heterogeneous wireless access networks: A short survey. In *TENCON 2015 - 2015 IEEE Region 10 Conference*, pages 1–6, Nov. 2015.

[124] Z. Zhao, M. Peng, Z. Ding, W. Wang, and H. V. Poor. Cluster content caching: An energy-efficient approach to improve quality of service in cloud radio access networks. *IEEE Journal on Selected Areas in Communications*, 34(5):1207–1221, May 2016.

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

[126] D. Sabella, A. de Domenico, E. Katranaras, M. A. Imran, M. di Girolamo, U. Salim, M. Lalam, K. Samdanis, and A. Maeder. Energy efficiency benefits of ran-as-a-service concept for a cloud-based 5g mobile network infrastructure. *IEEE Access*, 2:1586–1597, 2014.

[127] M. O. Al Kalaa, A. Imran, and H. H. Refai. mmwave based vs 2 ghz networks: What is more energy efficient? In *2016 International Wireless Communications and Mobile Computing Conference (IWCMC)*, pages 67–71, Sept. 2016.

[128] X. Gao, L. Dai, S. Han, C. L. I, and R. W. Heath. Energy-efficient hybrid analog and digital precoding for mmwave mimo systems with large antenna arrays. *IEEE Journal on Selected Areas in Communications*, 34(4):998–1009, Apr. 2016.

[129] C. Lin and G. Y. Li. Energy-efficient design of indoor mmwave and sub-thz systems with antenna arrays. *IEEE Transactions on Wireless Communications*, 15(7):4660–4672, Jul. 2016.

[130] Michal Pioro, Gabor Malicsko, and Gabor Fodor. *NETWORKING 2002: Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; Mobile and Wireless Communications*, chapter Optimal link capacity dimensioning in proportionally fair networks, pages 277–288. SPRINGER, 2002.

[131] L. Massoulie and J. Roberts. Bandwidth sharing: objectives and algorithms. *IEEE/ACM Transactions on Networking*, 10(3):320–328, Jun 2002.

[132] Se-Young Yun and Alexandre Proutiere. Distributed proportional fair load balancing in heterogenous systems. In *Proceedings of the 2015 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, SIGMETRICS '15, pages 17–30, New York, NY, USA, 2015. ACM.

[133] I. H. Hou and P. Gupta. Proportionally fair distributed resource allocation in multiband wireless systems. *IEEE/ACM Transactions on Networking*, 22(6):1819–1830, Dec 2014.

[134] Richard Combes, Salah-Eddine Elayoubi, and Zwi Altman. Cross-layer analysis of scheduling gains: Application to lmmse receivers in frequency-selective rayleigh-fading channels. In *Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), 2011 International Symposium on*, pages 133–139. IEEE, 2011.

[135] Yuanye Wang. *System Level Analysis of LTE-Advanced: with Emphasis on Multi-Component Carrier Management*. PhD thesis, Department of Electronic Systems, Aalborg University, 2010.

[136] P. Viswanath, D. N. C. Tse, and R. Laroia. Opportunistic beamforming using dumb antennas. 48(6):1277–1294, Jun. 2002.

[137] Motorola. Realistic lte performance – from peak to subscriber experience. Technical report, Motorola, 2009.

[138] M. Sawahashi, Y. Kishiyama, A. Morimoto, D. Nishikawa, and M. Tanno. Coordinated multipoint transmission/reception techniques for lte-advanced [coordinated and distributed mimo]. *IEEE Wireless Communications*, 17(3):26–34, Jun. 2010.

[139] Frank P Kelly, Aman K Maulloo, and David KH Tan. Rate control for communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research society*, 49(3):237–252, 1998.

[140] S. K. Baruah, J. E. Gehrke, and C. G. Plaxton. Fast scheduling of periodic tasks on multiple resources. In *Proceedings of 9th International Parallel Processing Symposium*, pages 280–288, Apr. 1995.

[141] Michael Isard, Vijayan Prabhakaran, Jon Currey, Udi Wieder, Kunal Talwar, and Andrew Goldberg. Quincy: Fair scheduling for distributed computing clusters. In *Proceedings of the ACM SIGOPS 22Nd Symposium on Operating Systems Principles*, SOSP '09, pages 261–276, New York, NY, USA, 2009. ACM.

[142] J. Wannstrom and K. Mallinson. Hetnet/small cells. `http://www.3gpp.org/hetnet`, 2014.

[143] D. Bladsjo, M. Hogan, and S. Ruffini. Synchronization aspects in lte small cells. *Communications Magazine, IEEE*, 51(9):70–77, Sept. 2013.

[144] Bo Han, V. Gopalakrishnan, Lusheng Ji, and Seungjoon Lee. Network function virtualization: Challenges and opportunities for innovations. *Communications Magazine, IEEE*, 53(2):90–97, Feb. 2015.

[145] C-ran: The road towards green ran. CCMC, Nov. 2011.

[146] M. Peng, Y. Li, J. Jiang, J. Li, and C. Wang. Heterogeneous cloud radio access networks: a new perspective for enhancing spectral and energy efficiencies. *IEEE Wireless Communications*, 21(6):126–135, Dec. 2014.

[147] D. Pompili, A. Hajisami, and T. X. Tran. Elastic resource utilization framework for high capacity and energy efficiency in cloud ran. *IEEE Communications Magazine*, 54(1):26–32, Jan. 2016.

[148] C. L. I, C. Rowell, S. Han, Z. Xu, G. Li, and Z. Pan. Toward green and soft: a 5g perspective. *IEEE Communications Magazine*, 52(2):66–73, Feb. 2014.

[149] S. Buzzi, C. L. I, T. E. Klein, H. V. Poor, C. Yang, and A. Zappone. A survey of energy-efficient techniques for 5g networks and challenges ahead. *IEEE Journal on Selected Areas in Communications*, 34(4):697–709, Apr. 2016.

[150] Werner Dinkelbach. On nonlinear fractional programming. *Management science*, 13(7):492–498, 1967.

[151] S. Rangan, T. S. Rappaport, and E. Erkip. Millimeter-wave cellular wireless networks: Potentials and challenges. *Proceedings of the IEEE*, 102(3):366–385, Mar. 2014.

[152] Mehmet R. Yuce, editor. *Ultra-Wideband and 60 GHz Communications for Biomedical Applications*. Springer, 2014.

[153] Fei Hu, editor. *Opportunities in 5G Networks: A Research and Development Perspective*. CRC Press, 2016.

[154] Ying Wang and Ahmed Abdelhadi. Optimal power allocation for LTE users with different modulations. *CoRR*, abs/1507.07159, 2015.

[155] Raj Jain, Dah-Ming Chiu, and W. Hawe. A quantitative measure of fairness and discrimination for resource allocation in shared computer systems. *CoRR*, cs.NI/9809099, 1998.

[156] Z. Pi and F. Khan. System design and network architecture for a millimeter-wave mobile broadband (mmb) system. In *34th IEEE Sarnoff Symposium*, pages 1–6, May 2011.

[157] GSMA Intelligence. Understanding 5g: Perspectives on future technological advancements in mobile. White paper, 2014.

[158] Y. P. Li, B. J. Hu, H. Zhu, Z. H. Wei, and W. Gao. A delay priority scheduling algorithm for downlink real-time traffic in lte networks. In *2016 IEEE Information Technology, Networking, Electronic and Automation Control Conference*, pages 706–709, May 2016.

[159] E. Bastug, M. Bennis, and M. Debbah. Living on the edge: The role of proactive caching in 5g wireless networks. *IEEE Communications Magazine*, 52(8):82–89, Aug. 2014.

[160] S. Ye, Y. Tao, Z. Zhang, and P. Zhang. Fronthaul-constrained resource allocation for downlink cached c-ran. In *2016 First IEEE International Conference on Computer Communication and the Internet (ICCCI)*, pages 305–309, Oct. 2016.

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

[162] M. Peng, Y. Sun, X. Li, Z. Mao, and C. Wang. Recent advances in cloud radio access networks: System architectures, key techniques, and open issues. *IEEE Communications Surveys Tutorials*, 18(3):2282–2308, thirdquarter 2016.

[163] G. Smail and J. Weijia. Techno-economic analysis and prediction for the deployment of 5g mobile network. In *2017 20th Conference on Innovations in Clouds, Internet and Networks (ICIN)*, pages 9–16, Mar. 2017.

[164] L. Cano, A. Capone, G. Carello, M. Cesana, and M. Passacantando. On optimal infrastructure sharing strategies in mobile radio networks. *IEEE Transactions on Wireless Communications*, 16(5):3003–3016, May 2017.

[165] Y. Chen, L. Duan, J. Huang, and Q. Zhang. Balancing income and user utility in spectrum allocation. *IEEE Transactions on Mobile Computing*, 14(12):2460–2473, Dec. 2015.

[166] R. Shi, Lanlan Rui, H. Huang, X. Qiu, H. Guo, and P. Zhang. A shapley value-based forwarding strategy in information-centric networking. In *2016 18th Asia-Pacific Network Operations and Management Symposium (APNOMS)*, pages 1–4, Oct. 2016.

[167] Shizuo Kakutani et al. A generalization of brouwer's fixed point theorem. 1941.

Appendix A

Proof of Theorem 1

Our objective is to prove that the selfish policy of multihomed users corresponds to the global optimum of the average throughput by verifying that the optimal value of β^* is the same.

A.1 Selfish optimum

Let β^*_{MH} be the maximum of Eq. (3.45):

$$\beta^*_{MH} = \arg\max \min \left(\underbrace{\frac{C^M - \lambda^M\sigma - \sum_{k=1}^{K} \lambda_k^{MH}\sigma\beta}{\beta}}_{①}, \underbrace{\frac{C_k^S - \lambda_k^S\sigma - \lambda_k^{MH}\sigma(1-\beta)}{1-\beta}}_{②} \right) \tag{A.1}$$

If ① < ②: We find that $\beta \in \min([0,1] \cap [\beta_1, \beta_2])$ with $0 < \beta_1 < 1 < \beta_2$ for a stable system. With β_1 and β_2 the solution of ① < ②.

$$\beta_{1,2} = \frac{\delta \mp \sqrt{\delta^2 - 4(K-1)\lambda^{MH}\sigma(C^M - \lambda^M\sigma)}}{2(K-1)\lambda^{MH}\sigma} \tag{A.2}$$

for δ as defined in Eq. (3.67). In this case, the solution is $\beta = \min([\beta_1, 1]) = \beta_1$.

If ① > ②: We find that $\beta \in \max([0,1] \cap (]-\infty, \beta_1] \cap [\beta_2, \infty[))$ with $0 < \beta_1 < 1 < \beta_2$ for a stable system. In this case, the solution is $\beta = \max([0, \beta_1]) = \beta_1$. Which makes $\beta^*_{MH} = \beta_1$.

A.2 Global optimum

Now, let β_g^* be the maximum of Eq. (3.48):

$$\beta_g^* = \arg\max \Bigg(\min \Bigg(\underbrace{\frac{C^M - \lambda^M\sigma - \sum_{k=1}^{K} \lambda_k^{MH}\sigma\beta}{\beta}}_{①}, \underbrace{\frac{C_k^S - \lambda_k^S\sigma - \lambda_k^{MH}\sigma(1-\beta)}{1-\beta}}_{②} \Bigg) - \lambda^S\sigma(1-\beta) - \lambda^M\sigma\beta \Bigg) \quad \text{(A.3)}$$

If ① < ②: From the min operator, we get that $\beta \in [\beta_1, 1]$ with f the function to maximize decreasing over $[\beta_1, 1]$. In this case, the maximum is achieved for $\beta = \beta_1$.

If ① > ②: $\beta \in [0, \beta_1]$ with the obtained function to maximize increasing over this domain. In this case, $\beta = \beta_1$. Which makes $\beta_g^* = \beta_1$.

By this, we prove that the multihomed users selfish policy corresponds to the global optimum of the system for $\beta_{MH}^* = \beta_g^* = \beta_1$.

Appendix B

Maximization of Eq. (3.66)

In order to prove that Eq. (3.66) corresponds to the maximum of Eq. (3.65), we combine both Eqs. (3.64) and (3.65) as:

$$\beta_h^* = \arg\max_{\beta_h} \left\{ \min \left(\textcircled{1}, \textcircled{2} \right) \right\} \tag{B.1}$$

with:

$$\textcircled{1} = \frac{C_h^M}{\beta_h}(1 - \hat{\lambda}^M \sigma (K \sum_{j \neq h} \frac{p_j^M \beta_j}{C_j^M} + \sum_{n > KJ} \frac{p_n^M}{C_n^M})) - K \hat{\lambda}^M \sigma p_h^M \tag{B.2}$$

$$\textcircled{2} = \frac{C_{min}^S}{1 - \beta_h}(1 - \hat{\lambda}^S \sigma (\sum_{j \neq h} \frac{p_j^S (1 - \beta_j)}{C_{min}^S} + \sum_{i > J} \frac{p_i^S}{C_{min}^S})) - \hat{\lambda}^S \sigma p_h^S \tag{B.3}$$

If $\textcircled{1} < \textcircled{2}$: We find that $\beta \in \min([0,1] \cap [\beta_1, \beta_2])$ with β_1 and β_2 the solution of $\textcircled{1} < \textcircled{2}$ defined as an equation of second degree with $0 < \beta_1 < 1 < \beta_2$.

$$\beta_{1,2} = \frac{\delta \mp \sqrt{\delta^2 - 4 \lambda_h^{MH} \sigma C_h^M (1 - \rho^M(\beta/\mathrm{h}))}}{2 \lambda_h^{MH} \sigma} \tag{B.4}$$

with δ and $\rho^M(\beta/\mathrm{h})$ as defined in Eqs. (3.67) and (3.68) respectively. In this case, the maximum of D_h^{MH} is obtained for $\beta = \min([\beta_1, 1]) = \beta_1$.

If $\textcircled{1} > \textcircled{2}$: We find that $\beta \in \max([0,1] \cap (] - \infty, \beta_1] \cap [\beta_2, \infty[) = \max([0, \beta_1]) = \beta_1$.

The solution of multihomed users selfish policy under heterogeneous radio conditions is thus $\beta_h^*(\beta/\mathrm{h}) = \frac{\delta - \sqrt{\delta^2 - 4 \lambda_h^{MH} \sigma C_h^M (1 - \rho^M(\beta/\mathrm{h}))}}{2 \lambda_h^{MH} \sigma}$ for both cases $\textcircled{1} < \textcircled{2}$ and $\textcircled{1} > \textcircled{2}$.

Appendix C

Proof of Theorem 2

Our objective is to prove first that the fixed point solution of Eq. (3.66) exists. We then prove that it corresponds to a Nash equilibrium.

Let A be a compact in $\mathbb{R}^J$ defined by $x_j \in [0,1]$, $j = 1, ..., J$ (a hypercube). According to Brouwer's fixed point theorem [167], function $f(.)$ defined on A by Eq. (3.66) admits a fixed point if it is continuous and has values in A. The continuity of f is obvious, we need to prove that $\forall(\beta/\mathrm{h}) \in [0,1]^{J-1}$ the solution of Eq. (3.66) gives a value $\beta_h \in [0,1]$.

Under the sufficient stability condition in Eqs. (3.71) and (3.72), we have:

$$\rho^M(\beta/\mathrm{h}) \leq 1$$
$$\rho^S(\beta/\mathrm{h}) \leq 1$$

Knowing that:

$$\delta = C_h^M(1 - \rho^M(\beta/\mathrm{h})) + C_{min}^S(1 - \rho^S(\beta/\mathrm{h})) + \lambda_h^{MH}\sigma \geq 0 \tag{C.1}$$

it is easy to show the lower bound :

$$\frac{\delta - \sqrt{\delta^2 - 4\lambda_h^{MH}\sigma C_h^M(1 - \rho^M(\beta/\mathrm{h}))}}{2\lambda_h^{MH}\sigma} \geq 0 \tag{C.2}$$

As of the upper bound, we use the fact that:

$$\delta \geq C_h^M(1 - \rho^M(\beta/\mathrm{h})) + \lambda_h^{MH}\sigma \tag{C.3}$$

If we multiply by $-4\lambda_h^{MH}\sigma$, add δ^2, and factorize, we get:

$$(\delta - 2\lambda_h^{MH}\sigma)^2 \leq \delta^2 - 4\lambda_h^{MH}\sigma C_h^M(1 - \rho^M(\beta/\mathrm{h}))$$

Since $0 \leq \delta - 2\lambda_h^{MH}\sigma$, we get:

$$\delta - 2\lambda_h^{MH}\sigma \leq \sqrt{\delta^2 - 4\lambda_h^{MH}\sigma C_h^M(1 - \rho^M(\beta/\mathrm{h}))}$$

By rearranging appropriately the terms, we prove the existence of the upper bound:

$$\frac{\delta - \sqrt{\delta^2 - 4\lambda_h^{MH}\sigma C_h^M(1-\rho^M(\beta/\mathrm{h}))}}{2\lambda_h^{MH}\sigma} \leq 1 \tag{C.4}$$

By this, we prove that $\forall(\beta/\mathrm{h}) \in [0,1]^{J-1}$, and for the sufficient stability condition of the system, the β_h calculated by Eq. (3.66) $\in [0,1]$:

$$0 \leq \frac{\delta - \sqrt{\delta^2 - 4\lambda_h^{MH}\sigma C_h^M(1-\rho^M(\beta/\mathrm{h}))}}{2\lambda_h^{MH}\sigma} \leq 1 \tag{C.5}$$

Now consider the game where the strategy of each class of multihomed users is its traffic split β_j. Let users of classes $j \neq h$ apply the policies β_j^* obtained from Eq. (3.66). If class h users apply a policy β_h different from β_h^* in Eq. (3.66), they will degrade their throughput as β_h^* is by definition the traffic split that maximizes the multihomed throughput. This proves the existence of a Nash equilibrium in the system.