Universitat Politechnica de Catalunya

Ph.D. Thesis

Traffic Offloading in Future Heterogeneous Mobile Networks

Author:
Panagiotis Trakas

Supervisors:
Dr Ferran Adelantado i Freixer
Associate Professor
Universitat Oberta de Catalunya (UOC)
Dr Christos Verikoukis
Fellow Researcher
Telecommunications Technological Center of Catalonia (CTTC)

Tutor:
Dr Jordi Perez Romero
Associate Professor
Universitat Politecnica de Catalunya (UPC)

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

in the

Department of Signal Theory and Communications
Universitat Politechnica de Catalunya (UPC)

Barcelona, September 2018

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Traffic offloading in future heterogeneous mobile networks

Panagiotis Trakas

ADVERTIMENT La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del repositori institucional UPCommons (http://upcommons.upc.edu/tesis) i el repositori cooperatiu TDX (http://www.tdx.cat/) ha estat autoritzada pels titulars dels drets de propietat intel·lectual **únicament per a usos privats** emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei UPCommons o TDX. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a UPCommons (*framing*). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del repositorio institucional UPCommons (http://upcommons.upc.edu/tesis) y el repositorio cooperativo TDR (http://www.tdx.cat/?locale-attribute=es) ha sido autorizada por los titulares de los derechos de propiedad intelectual **únicamente para usos privados enmarcados** en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio UPCommons No se autoriza la presentación de su contenido en una ventana o marco ajeno a UPCommons (*framing)*. Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the institutional repository UPCommons (http://upcommons.upc.edu/tesis) and the cooperative repository TDX (http://www.tdx.cat/?locale-attribute=en) has been authorized by the titular of the intellectual property rights **only for private uses** placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized neither its spreading nor availability from a site foreign to the UPCommons service. Introducing its content in a window or frame foreign to the UPCommons service is not authorized (framing). These rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

Abstract

The rise of third-party content providers and the introduction of numerous applications has been driving the growth of mobile data traffic in the past few years. In order to tackle this challenge, Mobile Network Operators (MNOs) aim to increase their networks' capacity by expanding their infrastructure, deploying more Base Stations (BSs). Particularly, the creation of Heterogeneous Networks (HetNets) and the application of traffic offloading through the dense deployment of low-power BSs, the small cells (SCs), is one promising solution to address the aforementioned explosive data traffic increase.

Due to their financial implementation requirements, which could not be met by the MNOs, the emergence of third parties that deploy small cell networks creates new business opportunities. Thus, the investigation of frameworks that facilitate the implementation of outsourced traffic offloading, the collaboration and the transactions among MNOs and third-party small cell owners, as well as the provision of participation incentives for all stakeholders is essential for the deployment of the necessary new infrastructure and capacity expansion.

The aforementioned emergence of third-party content providers and their applications not only drives the increase in mobile data traffic, but also create new Quality of Service (QoS) as well as Quality of Experience (QoE) requirements that the MNOs need to guarantee for the satisfaction of their subscribers. Moreover, even though the MNOs accommodate this traffic, they do not get any monetary compensation or subsidization for the required capacity expansion. On the contrary, their revenues reduce continuously. To that end, it is necessary to research and design network and economic functionalities adapted to the new requirements, such as QoE-aware Radio Resource Management and Dynamic Pricing (DP) strategies, which both guarantee the subscriber satisfaction and maximization the MNO profit (to compensate the diminished MNOs' revenues and the increasing deployment investment).

Following a thorough investigation of the state-of-the-art, a set of research directions were identified. This dissertation consists of contributions on network sharing and outsourced traffic offloading for the capacity enhancement of MNO networks, and the design

of network and economic functions for the sustainable deployment and use of the densely constructed HetNets. The contributions of this thesis are divided into two main parts, as described in the following.

The first part of the thesis introduces an innovative approach on outsourced traffic offloading, where we present a framework for the Multi-Operator Radio Access Network (MORAN) sharing. The proposed framework is based on an auction scheme used by a monopolistic Small Cell Operator (SCO), through which it leases its SC infrastructure to MNOs. As the lack of information on the future offered load and the auction strategies creates uncertainty for the MNOs, we designed a learning mechanism that assists the MNOs in their bid-placing decisions. Our simulations show that our proposal almost maximizes the social welfare, satisfying the involved stakeholders and providing them with participation incentives.

The second part of the thesis researches the use of network and economic functions for MNO profit maximization, while guaranteeing the users' satisfaction. Particularly, we designed a model that accommodates a plethora of services with various QoS and QoE requirements, as well as diverse pricing, that is, various service prices and different charging schemes. In this model, we proposed QoE-aware user association, resource allocation and joint resource allocation and dynamic pricing algorithms, which exploit the QoE-awareness and the network's economic aspects, such as the profit. Our simulations have shown that our proposals gain substantial more profit compared to traditional and state-of-the-art solutions, while providing a similar or even better network performance.

Resumen

El aumento de los proveedores de contenido de terceros y la introducción de numerosas aplicaciones ha impulsado el crecimiento del tráfico de datos en redes móviles en los últimos años. Para hacer frente a este desafío, los operadores de redes móviles (Mobile Network Operators, MNOs) apuntan a aumentar la capacidad de sus redes mediante la expansión de su infraestructura y el despliegue de más estaciones base (BS). Particularmente, la creación de Redes Heterogéneas (Heterogenous Networks, HetNets) y la aplicación de descarga de tráfico a través del despliegue denso de BSs de baja potencia, las células pequeñas (small cells, SCs), es una solución prometedora para abordar el aumento del tráfico de datos explosivos antes mencionado.

Debido a sus requisitos de implementación financiera, que los MNO no pudieron cumplir, la aparición de terceros que implementan redes de células pequeñas crea nuevas oportunidades comerciales. Por lo tanto, la investigación de marcos que faciliten la implementación de la descarga tercerizada de tráfico, la colaboración y las transacciones entre MNOs y terceros propietarios de células pequeñas, así como la provisión de incentivos de participación para todas las partes interesadas esencial para el despliegue de la nueva infraestructura necesaria y la expansión de la capacidad.

La aparición antes mencionada de proveedores de contenido de terceros y sus aplicaciones no solo impulsa el aumento del tráfico de datos móviles, sino también crea nuevos requisitos de calidad de servicio (Quality of Service, QoS) y calidad de la experiencia (Quality of Experience, QoE) que los operadores de redes móviles deben garantizar para la satisfacción de sus suscriptores. Además, a pesar de que los operadores de redes móviles adaptan este tráfico, no obtienen ninguna compensación monetaria o subsidio por la expansión de capacidad requerida. Por el contrario, sus ingresos se reducen continuamente. Para ello, es necesario investigar y diseñar funcionalidades económicas y de red adaptadas a los nuevos requisitos, tales como las estrategias QoE-conscientes de gestión de recursos de radio y de precios dinámicos (Dynamic Pricing, DP), que garantizan la satisfacción del abonado y la maximización de la ganancia de operador

móvil (para compensar los ingresos de los MNOs disminuidos y la creciente inversión de implementación).

Después de una investigación exhaustiva del estado del arte, se identificaron un conjunto de direcciones de investigación. Esta disertación consiste en contribuciones sobre el uso compartido de redes y la descarga tercerizada de tráfico para la mejora de la capacidad de redes MNO, y el diseño de funciones económicas y de red para el despliegue y uso sostenible de las HetNets densamente construidas. Las contribuciones de esta tesis se dividen en dos partes principales, como se describe a continuación.

La primera parte de la tesis presenta un enfoque innovador sobre la descarga subcontratada de tráfico, en el que presentamos un marco para el uso compartido de la red de acceso de radio de múltiples operadores (Multi-Operator RAN, MORAN). El marco propuesto se basa en un esquema de subasta utilizado por un operador monopólico de celda pequeña (Small Cell Operator, SCO), a través del cual arrienda su infraestructura SC a MNOs. Como la falta de información sobre la futura carga de red y las estrategias de subasta creaban incertidumbre para los MNO, diseñamos un mecanismo de aprendizaje que asiste a los MNO en sus decisiones de colocación de pujas. Nuestras simulaciones muestran que nuestra propuesta casi maximiza el bienestar social, satisfaciendo a las partes interesadas involucradas y proporcionándoles incentivos de participación.

La segunda parte de la tesis investiga el uso de las funciones económicas y de red para la maximización de los beneficios de los MNOs, al tiempo que garantiza la satisfacción de los usuarios. Particularmente, diseñamos un modelo que acomoda una gran cantidad de servicios con diversos requisitos de QoS y QoE, tanto como diversos precios, es decir, varios precios de servicio y diferentes esquemas de cobro. En este modelo, propusimos algoritmos QoE-conscientes para asociación de usuarios, asignación de recursos y conjunta asignación de recursos y de fijación dinámica de precios, que explotan la conciencia de QoE y los aspectos económicos de la red, como la ganancia. Nuestras simulaciones han demostrado que nuestras propuestas obtienen un beneficio sustancial en comparación con las soluciones tradicionales y del estado del arte, a la vez que proporcionan un rendimiento de red similar o incluso mejor.

Acknowledgements

By writing these lines, one of the most important chapters of my life comes to an end. The submission of this Ph.D dissertation was made possible thanks to the help and support of many people. Herein, I would like to express my gratitude to all of them.

Firstly, I would like to express my sincere gratitude to my advisors Dr. Ferran Adelantado and Dr. Christos Verikoukis for the continuous support of my doctoral studies and related research, for their patience, motivation, and immense knowledge. Their guidance assisted me with conducting my research and and the writing of this thesis. They gave me the opportunity to work within the framework of the CROSSFIRE European research project, which provided me with the means towards high-end research, attention to project meetings and international conferences.

I thank my fellow project team-mates in UOC and Iquadrat, Georgia Tseliou and Georgios Kollias, for the help they provided me at the beginning of my studies and their continuous support, for the brainstorming sessions, for sharing the anxiousness before deadlines, and for all the good times we had in Barcelona.

Special thanks go to all my friends, to the old ones that followed me faithfully through this journey and to the new ones, especially Alexandra Mpousia and Maria Oikonomakou with whom I shared experiences and worries. Other good friends that belong here and have my gratitude are Petya Dacheva, Mireia Ribas, David Mudroncik and Juanjo Hernandez.

I would also like to thank my close friends, who, since the time they knew I was going to Barcelona to do research and pursue the Ph.D degree at the Polytechnic University of Catalonia, they have always been supporting me.

As in every step forward in my life, I cannot forget that this achievement could not have been possible without the support of my family: my parents, Thanasis and Chrysa, and my brother, Vasilis. They supported me endlessly throughout this challenging period of my life.

I would like to complete these acknowledgements by extending my gratitude to everyone else who, directly or not, consciously or not, has contributed to this thesis.

Contents

List of Figures

List of Tables

Abbreviations

2G	Second Generation
3G	Third Generation
3GPP	Third Generation Partnership Project
4G	Fourth Generation
5G	Fifth Generation
ACP	ACcess Permission
AGO	Auction game-based offloading mechanism
AP	Access Point
APO	Access point owner
BS	Base Station
C-RAN	Cloud RAN
CACA	Cost-constrained Association Control Algorithm
CAPEX	Capital Expenditure
CDF	Cumulative distribution function
CGO	Congestion game-based offloading mechanism
CSP	Cellular Service Provider
DOFF	Data offloading game
DP	Dynamic Pricing
eNB	evolved Node B
FPA	First-Price sealed bid auction
FSP	Femtocell Service Provider
GWCN	Gateway Core Network
H-CoMP	Hybrid coordinated multipoint transmission
HetNet	Heterogeneous Network
HW	Holt-Winters

IoT	Internet of Things
IQX	Exponential interdependency of quality of experience and quality of service
ISD	Inter-Site Distance
ISP	Internet Service Provider
ITU-T	International Telecommunication Union's Telecommunication Standardization Sector
KPI	Key Performance Indicators
LDP	Location Dependent Pricing
LOS	Line of sight
LTE-A	Long Term Evolution Advanced
LTP	Linear Threshold Policy
MCS	Modulation and Coding Scheme
MIMO	Multiple-Input multiple-Output
MINLP	Mixed Integer Non-Linear Problem
mmWave	Millimetre-wave
MNO	Mobile Network Operator
MOCN	Multiple Operator Core Network
MORAN	Multiple Operator Radio Access Network
MOS	Mean Opinion Score
MTP	Multiple Threshold Policy
NLOS	Non-Line of sight
NP	Non deterministic polynomial time
NTP	Network Termination Point
NV	Network Virtualization
OFDMA	Orthogonal frequency division multiple access
OPEX	Operational Expenditure
OS	Overall user satisfaction
OSM	Overall user satisfaction maximization
OTT	Over-the-top
PCC	Policy and Charging Control
PF	Proportional Fair
PFEE	Proportional fair energy efficiency

PLMN	Public land mobile network
PM	Profit Maximizing resource allocation algorithm
PUCCH	Physical Uplink Control Channel
PUSCH	Physical Uplink Shared Channel
Q-AAA	QoE-aware Association Algorithm
QADP	Quality-aware dynamic price
QoE	Quality of Experience
QoS	Quality of Service
RA	Resource Allocation
RAN	Radio Access Network
ROI	Return on Investment
RRM	Radio Resource Management
SC	Small Cell
SCaaS	Small Cell as a Service
SCO	Small Cell Operator
SH	Small cell Holder
SIC	Successive Interference Cancellation
SINR	Signal-to-Interference-plus-Noise-Ratio
SLA	Service Level Agreement
SP	Service Profile
SSP	Small cell Service Provider
SWM	Social Welfare Maximization
TDAP	Time dependent adaptive pricing
TRX/RF	Transceiver/Radio Frequency
UE	User Equipment
VCG	Vickrey-Clarke-Groves
VM	Virtual Machine
VU	Vehicular User
μWave	Microwave

For my parents, Thanasis and Chrysa, and my brother, Vasilis.

Chapter 1

Introduction

This chapter presents the context of the addressed problems, the motivation behind our proposals, and the structure of the thesis. This chapter is organized as follows. Section 1.1 presents the motivation of the thesis, whereas Section 1.2 describes the structure of the thesis. Finally, we present our contributions in Section 1.3.

1.1 Motivation

Since the launch of the second generation mobile networks (2G), the telecommunications industry has been developing with rapid rates. 2G popularized the use of mobile phones, and widened the mobile telephony market from a limited market in the business sector to the general market and the typical consumer. With the introduction of the third generation (3G), the mobile users became accustomed to the use of data services. Moreover, 3G along with the first appearance of smartphones (i.e. iOS and android devices) allowed the extensive use of mobile applications, paving the way for the fourth generation (4G), with the general consumer using services for every aspect of her daily life (e.g. business, recreation, socialization etc.). The emergence of numerous mobile applications along with their wide use have created an increasing trend in the mobile data demand. This growth in traffic load has been following an exponential increase over the last few years, which is expected to continue in the future. Particularly, the average monthly data consumption is expected to reach 49 exabytes by 2021 according to Cisco [2]. As a result, the telecommunications industry along with the academia have been researching ways to address this challenge, and increase the network capacity.

One of the most promising solutions is the densification of the existing networks with the deployment of numerous low-power base stations (BSs), also known as small cells

(SCs). The small cells can be deployed in areas where the typical high-power, macrocell BS cannot serve the entire offered load, creating Heterogeneous Networks (HetNets). In this case, the traffic is offloaded from the macrocell to the small cells, allowing the dense reuse of the spectrum, and hence increasing the network capacity. However, despite its potential, the ubiquitous deployment of small cells requires high Capital Expenditure (CAPEX). Therefore, a single Mobile Network Operator (MNO) may not be able to make such investments, which also pose high financial risks.

The answer to the financial issues of traffic offloading came through a different trend of telecommunications research, Network Virtualization (NV). NV abstracts the physical infrastructure and radio resources and isolates them to virtual resources, allowing for their splitting among multiple MNOs. Therefore, NV facilitates the sharing of network infrastructure, as well as spectrum by multiple MNOs [3]. As a result, a number of different Radio Access Network (RAN), spectrum sharing scenarios and use cases have been introduced by telecommunication standardization bodies such as the third generation partnership project (3GPP), regulatory organizations (e.g. ITU-T), as well as third-party organizations (e.g. Small Cell Forum). This in turn enabled the MNOs to cooperate and form agreements for sharing each others' infrastructure, or form joint ventures for the joint infrastructure deployment and operation.

The possibilities provided by NV helped independent third parties to emerge and enter the telecommunications market as neutral hosts that lease network infrastructure. These neutral hosts, also known as Small Cell Operators (SCOs), can be authorities of highly-populated metropolitan areas, or owners of large venues that deploy small cell infrastructure and lease it to MNOs [4, 5], treating it as an additional revenue source. This outsourcing of small cell deployment and traffic offloading helps the MNOs to distribute the necessary CAPEX among several SCOs (i.e. different SCOs in different locations). It is true that this CAPEX reduction is translated into an increase in the MNOs' Operational Expenditure (OPEX) through the price paid to lease the small cell infrastructure. However, this model poses lower financial risk compared to the ubiquitous small cell deployment, and at the same time allows the MNOs to increase their network capacity in a dynamic way, that is, at the locations and the time periods where and when it is needed.

Despite the capacity and economic benefits that outsourced traffic offloading offers, there are new challenges that need to be addressed, since this business model was recently created and adopted.

- There is still a lack of frameworks that cover the whole range of traffic offloading

markets. Particularly, traffic offloading agreements may vary from simple arrangements between a single small cell owner and its MNO, to detailed, long-term agreements among multiple MNOs and neutral hosts, who own hundreds of small cells as mentioned above. In such competitive markets, there is always the possibility that the stakeholders show undesirable market behaviour. Therefore, the proposed frameworks should also take into account the competition among MNOs or SCOs, to provide fairness to all participants in the offloading market, and prevent unethical and illegal practices (e.g. collusion in auctions, market manipulation etc.), while promoting healthy competition that benefits both the stakeholders and the consumers.

- Economic incentives must be provided for both the MNOs and the SCOs. As the SCOs undertake the deployment of the small cell networks and its financing, they need appropriate economic incentives. This in turn requires leasing strategies (e.g. leasing schemes, duration of leasing periods, pricing schemes etc.) that guarantee a significant Return on Investment (ROI[1]). Regarding the MNOs, even though traffic offloading can boost substantially their networks' performance and improve the user experience, it should not be expensive. That is, an MNO would be interested in investing in traffic offloading only if it provided him with either a low-cost alternative for serving its traffic or an increase in profit due to the growth in the served load. Thus, the proposed traffic offloading frameworks should also guarantee long-term sustainability of the traffic offloading markets, and provide the stakeholders with economic incentives.

- The MNOs need guidelines and strategies for the wide adoption of outsourced traffic offloading. In the near future, traffic offloading will be required ubiquitously. This means that an MNO must have access to numerous small cell networks in different sites. And as mentioned earlier, even though the MNO does not pay the small cell CAPEX, it covers the small cell OPEX, which on a large scale can prove to be a significant expense. Therefore, an MNO must plan cautiously where, when and how much to pay for traffic offloading outsourcing, taking into account the expected traffic demand, the competition, technical limitations of the small cell network (e.g. maximum number of supported tenants, backhaul capacity, number of small cells per site etc.), the offloading cost, and finally the potential profit. Therefore, it is imperative to propose mechanisms that assist the MNOs in adopting beneficial offloading strategies, while taking into account the aforementioned parameters and conditions.

[1] ROI is defined as the benefit of an investment relative to the investment's cost[6].

Taking the above challenges into consideration, we can formulate the traffic offloading problem. The main issue is how traffic offloading frameworks should be designed in order to guarantee legality and fairness, and provide incentives for its adoption by stakeholders. Then, within such a framework we need to answer where, when and how much an MNO should invest on traffic offloading, in order to have both technical and financial gains. Similarly, an SCO needs to know how extensive its small cell deployment should be, and which pricing schemes and price levels are best to incentivize MNOs lease its infrastructure, in order to recoup its investment.

As it will be shown in Chapter 2.3.1, there is a multitude of works that study variations of the traffic offloading problem, which propose their own frameworks for the traffic offloading market, as well as MNO and/or SCO strategies on leasing small cell infrastructure. In most of these works we encounter the assumption that the SCOs own not only small cell infrastructure, but also licensed spectrum. Hence, in these scenarios the MNOs solely need to come to an agreement with the SCO for the offloaded traffic volume, the offloading duration and the corresponding price. However, as licensed spectrum can be acquired only through national auctions, their license fees can raise to billions of euros [7]. Since the typical, incumbent MNO offers nation-wide services, it gains enough revenue to recoup its infrastructure CAPEX and OPEX, as well as the licence fee during the license's duration (e.g. 20 years for UK's 4G auction [8]). Conversely, a typical SCO's service is offered in small geographical areas, which in turn limits its maximum revenue. To that end, it is not safe to assume that a local SCO can acquire a spectrum license, and get a significant Return on Investment through a small cell network.

In the first contribution of this thesis, we study and propose a solution for the traffic offloading problem. In our case, we focus our study on a sharing scheme different from the literature's trend, with regard to the spectrum used in the shared RAN. Specifically, we examine the Multi-Operator RAN (MORAN) sharing configuration [1], whose business model is known as Small Cell as a Service (SCaaS) [9]. SCaaS concerns the provision of solely small cell infrastructure to a group of MNOs, which is usually deployed by an SCO. The main difference with the sharing models in the literature is in the fact that with MORAN the transceiver and radio frequency (TRX/RF) elements (i.e. schedulers, licensed bandwidth etc.) are not shared. Hence, MORAN allows each MNO to apply its own policy on the Quality of Service (QoS) the users receive. This means that even though the MNOs' spectrum is used at a shared RAN, they still retain exclusive rights on its usage. Therefore, in a MORAN sharing arrangement there is a need for a mutual Service Level Agreement (SLA) between the MNO and the SCO for the use of the spectrum, for instance, the spectrum manager leasing arrangement or the de facto leasing arrangement [10]. In this scenario, we propose a framework for the organization of the small cell resource allocation among the leasing MNOs, and the corresponding

transactions. Furthermore, we propose a novel mechanism that assists the MNOs in their decision making, in order to achieve their network and economic objectives within a competitive market, and under the uncertainty that the dynamic variations of the traffic load, along with the opponent MNOs' actions incur.

After solving the traffic offloading problem, an MNO has at its disposal additional small cell infrastructure, which can be used to provide better service to its subscribers and improve its own economic gains. Consequently, the solution of the traffic offloading problem leads us to a new area of study, which examines how an MNO can use efficiently and profitably its HetNet. The MNO network's performance and financial gains depend highly on how network and economic functionalities are designed and applied. Therefore, in order to identify the appropriate network and economic functions, and design them so that they guarantee subscriber satisfaction and raise the MNO profit, we first need to identify the challenges that need to be addressed.

- The aforementioned raise in mobile data demand has been greatly impacted by the rise of over-the-top (OTT) content providers, and the dissemination of their content through numerous applications for mobile devices. These applications are described by various QoS requirements, which along with the use of multiple devices per user [11] increase the heterogeneity of the traffic demand. The burden on the networks is further increased by the introduction of Quality of Experience (QoE) requirements for the aforementioned services and applications. The provision of seamless connectivity and QoE-aware services to their users is essential for the MNOs, as QoE is one of the key elements that has been attracting the interest of telecommunication stakeholders the past few years [12], and a key design factor for the future 5G networks [13].

- This emergence of OTT applications and the corresponding traffic increase have caused a drop in the MNOs' revenues [14–16]. This revenue decrease finds its roots in the gradual replacement of the MNOs' voice and messaging services (i.e. their main source of revenue from the middle 1990s until the late 2000s) by their OTT counterparts (e.g. Skype, Whatsapp etc.). Furthermore, the MNOs' data service prices (i.e. price per down/uploaded GB) have been decreasing over the years due to the introduction of new technologies (i.e. migration from 2G to 3G and lately to 4G), along with the market competition. Additionally, the content providers only reap the benefits of using the MNO infrastructure without cost, as they do not subsidize either the infrastructure deployment or its operation.

Therefore, the MNOs face a two-fold challenge: meet a multitude of different QoS and QoE requirements due to the existence of numerous, distinctive services and applications, and increase their revenues in an effort to maximize their profit. As previously mentioned, the initial and fundamental step for improving the service quality is the capacity growth of the MNO networks through densification with small cells. However, given a particular HetNet deployment it is important to recognize the parameters that provide satisfactory service to the subscribers, and in turn high profit to the MNO. With regard to the user experience, it has been proven that the relation between QoS and QoE has a non-linear nature [17]. This means that small degradations in the received QoS can impact significantly the users' perception of QoE. Moreover, the users' perception is influenced by other factors, which can be either technical such as the device characteristics [18], or non-technical such as the service price [19]. Under these circumstances, and in order to address the MNOs' two-fold challenge, it is essential to design both network and economic functionalities adapted to these new requirements, as mentioned previously. For instance, functions such as QoE-aware Radio Resource Management (RRM) and user association strategies, as well as smart and dynamic pricing schemes can be used to affect both technical and economic parameters in a network. By doing so, an MNO can offer high service quality to its subscribers, while increasing its profit.

As it will be shown in Chapter 2.3.2, the majority of works on RRM and cell selection place the focus on the provision of high QoS/QoE and other network related parameters (e.g. power allocation, spectral and/or energy efficiency, fairness etc.). However, they do not take into account the impact of their proposals on the MNOs' financial parameters. With regard to the economic functions an MNO can use to affect both its network and economic aspects, the literature has placed its focus on smart, dynamic pricing. In particular, the proposals on pricing schemes have been used in order to influence the users and steer the traffic demand from the peak to the off-peak traffic hours and locations. Nevertheless, even though these proposals result in reducing congestion, they cannot affect the network performance during unavoidable congestion periods.

In the second contribution of this thesis, we focus on RAN (i.e. user association and resource allocation) and economic (i.e. pricing) functions. Contrary to the literature, our proposals on these network and economic functions take into account a multitude of parameters that increase the traffic heterogeneity, as well as diverse pricing, and always aim to benefit both the user satisfaction and the MNO financial gains.

1.2 Thesis Structure

The scope of this Ph.D dissertation is the design of novel schemes for improving the performance of current and future HetNets. At the same time, we address the financial challenges from an operator's point of view, given the competition in the telecommunications market along with the ever increasing data traffic demands, and requirements in service quality. The proposed solutions enable the sustainable densification of HetNets, as well as their profitable operation. Particularly, this thesis is based on two research directions, as it can be seen in Fig. 1.1: 1) the study of the telecommunications stakeholders's strategies when they compete or collaborate in order to strike beneficial arrangements in small cell infrastructure sharing scenarios for traffic offloading; 2) the study of network and economic functions (i.e. user association, resource allocation, and pricing) in HetNets aiming for a service provision that satisfies the subscribers and maximizes the MNO profit.

The remaining part of the thesis includes four chapters. Chapter 2 consists of two main parts that provide background information on network sharing, and a literature review respectively. Specifically, the information on network sharing concerns a variety of scenarios and use cases defined by 3GPP, and subsequently network architecture configurations for their implementation. With regard to the literature review, it is divided in three parts. Initially, we present works that solve the traffic offloading problem, that is, scenarios where multiple MNOs or SCOs compete to maximize their utility (e.g. cost reduction, revenue maximization etc.) by buying or selling small cell resources. Subsequently, we survey works on user association and resource allocation,where the objectives vary between improvement of the network performance, and increase of the MNOs' economic gains. In the same vein, we finalize the literature review with a presentation of various works on dynamic, smart pricing, which is used as a means to address network congestion, better the service provision, and guarantee financial gains for the MNOs.

The contributions of this thesis are presented in the next two chapters. Specifically, both study areas of our contributions examine the techno-economical aspects of traffic offloading, however from a different perspective. Initially, we examine the impact of competition among MNOs, when they share small cell network infrastructure. Subsequently, we analyse ways to improve the MNO profit as well as the user satisfaction through intelligent management and pricing of the acquired small cell resources.

Particularly, Chapter 3 investigates the traffic offloading problem under the SCaaS approach. The main contributions in Chapter 3 are the proposal of an analytical, realistic model for SCaaS that includes and combines the technological constraints and economic

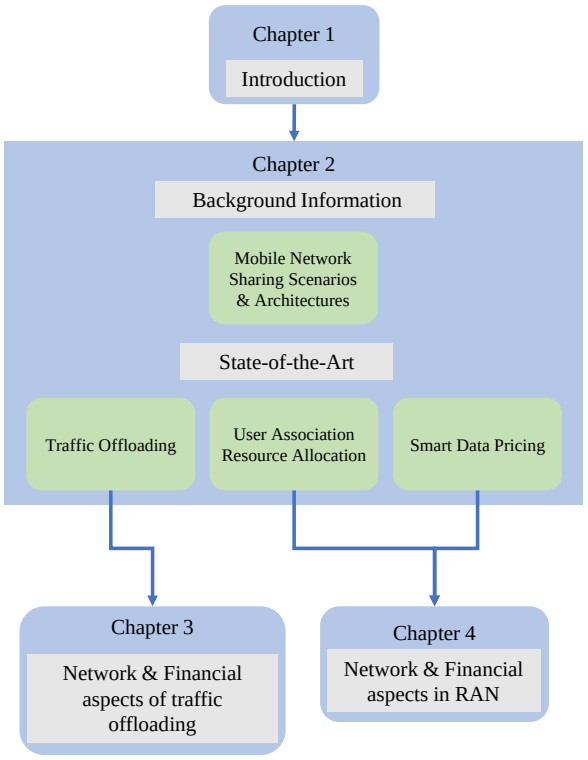


FIGURE 1.1: Thesis Structure

objectives of each stakeholder. We further propose an auction scheme for the efficient allocation of the small cell resources and the optimization of the stakeholders' economic gains. To that end, we provide a learning mechanism that assists the MNOs with their auction strategies, in realistic conditions with lack of information on future traffic load, and the strategies of the competing MNOs.

Chapter 4 is dedicated to the study of network (i.e. user association, resource allocation) and economic (i.e. pricing) functions, and the proposal of the corresponding QoE-aware algorithms for the maximization of the MNO profit, while guaranteeing high user satisfaction. Chapter 4's main contributions are three QoE-aware, greedy, heuristic algorithms in a scenario with various services and numerous QoS/QoE requirements. We further take into account diverse pricing, that is, a range of service prices for different pricing schemes. In this scenario, we propose a user association algorithm for dual-band 5G HetNets (i.e. microwave macrocells overlaid with millimetre wave small cells), aiming to maximize the MNO profit. Subsequently, we propose a resource allocation and a joint resource allocation and dynamic pricing algorithm that maximize the MNO profit within a HetNet, while taking into account constraints on the fairness, and the overall user satisfaction. Finally, Chapter 5 provides conclusions for the works presented in the previous chapters, and discusses potential issues for future research.

1.3 Research Contributions

The presented in this dissertation ideas have been published in various scientific journals and conferences. In the following, we list our publications according to their respective contribution, and the corresponding chapter in this thesis.

The publication list consists of:

[**J**]: 2 journal papers.

[**C**]: 3 conference papers.

Chapter 3 is based on 1 conference and 1 journal paper.

[**C1**] P. Trakas, F. Adelantado, and C. Verikoukis, *"A novel learning mechanism for traffic offloading with small cell as a service,"* in IEEE International Conference on Communications (ICC), pp. 6893-6898, June 2015.

URL: *https://ieeexplore.ieee.org/document/7249424*

[**J1**] P. Trakas, F. Adelantado, and C. Verikoukis, *"Network and Financial Aspects of Traffic Offloading with Small Cell as a Service,"* IEEE Transactions on Wireless Communications, vol. 17, no. 11, pp. 7744-7758, Nov. 2018.

URL: *https://ieeexplore.ieee.org/document/8470260*

Chapter 4 is based on 2 conference and 1 journal paper.

[**C2**] P. Trakas, F. Adelantado, N. Zorba, and C. Verikoukis, *"A quality of experience-aware association algorithm for 5G heterogeneous networks,"* in IEEE International Conference on Communications (ICC), May 2017.

URL: *https://ieeexplore.ieee.org/document/7996869*

[**C3**] P. Trakas, F. Adelantado, N. Zorba, and C. Verikoukis, *"A QoE-aware joint resource allocation and dynamic pricing algorithm for heterogeneous networks,"* in IEEE Global Communications Conference (GLOBECOM), Dec. 2017

URL: *https://ieeexplore.ieee.org/document/8254131*

[**J2**] P. Trakas, F. Adelantado, and C. Verikoukis, *"QoE-aware resource allocation for profit maximization under user satisfaction guarantees in HetNets with differentiated services,"* to appear in IEEE Systems Journal.

URL: *https://ieeexplore.ieee.org/document/8521660*

Chapter 2

Background and State-of-the-Art

2.1 Introduction

The continuous demand for new services and the increasing mobile data consumption has been driving the research carried out by both the industry and the academia for advancing telecommunications technologies. This research has resulted in the worldwide adoption of LTE-A, and 5G in the near future, always focusing on network throughput, delay and fairness, aiming for better QoS and QoE. However, in order to adopt new technologies, fund further advancements, and offer satisfactory service to the users, it is important to assure the long-term sustainability of the deployed networks.

To this end, research for sustainable capacity growth has been carried out based on small cell network densification, RAN sharing, and the corresponding transactions between telecommunication stakeholders. Similarly, little attention has been also given to the impact of network functions (e.g. resource allocation) on the networks' financial aspects. Finally, economic functions such as pricing have been researched for influencing the traffic demand, and hence improving the service provision. This chapter presents information on RAN sharing scenarios, use cases, and architectures as well as a review of state-of-the-art works on small cell sharing. Subsequently, we complete our literature review with works on RAN functions (i.e. user association and resource allocation), and dynamic, smart pricing in order to show the effect they can have on network performance and MNO financial parameters.

The chapter is organized as follows. Section 2.2 lists the scenarios, uses cases and architectures of RAN sharing proposed by 3GPP and other organizations such as the small cell forum. Section 2.3 provides the literature review and is divided in two parts, according to our contributions. The first part is dedicated on works that examine variations of

the traffic offloading problem and the interactions among stakeholders. The second part presents works on the efficient use of HetNet resources, and particularly focuses on user association, resource allocation, and dynamic, smart pricing. Section 2.4 describes open issues and challenges that have not yet been addressed by the research works presented in the literature review. Finally, Section 2.5 concludes the chapter.

2.2 Mobile Network Architectures and Network Sharing

In this section, we present network sharing scenarios and use cases as defined by 3GPP [20], along with the corresponding architectures proposed by 3GPP [21], as well as other organizations such as the small cell forum [1]. Initially, let us introduce the actors and their roles in network sharing, as defined by 3GPP [20].

In a network scenario, 3GPP defines two main types of actors: the Hosting RAN Provider, and the Participating Operator. The Hosting RAN Provider owns network infrastructure, which it shares with or leases to one or more Participating Operators. The Hosting RAN Provider has primary operational access to licensed spectrum, which is part of the sharing agreement. However, the Hosting RAN Provider does not necessarily need to own licensed spectrum. Therefore, it can be solely an infrastructure provider, who manages the shared network utilizing spectrum provided by the Participating Operators (as agreed upon an SLA [10]). It should be noted that the Hosting RAN Provider is a notion that can be implemented as outsourcing, joint ventures, or leasing agreements for operating, owning the RAN infrastructure or managing the sharing agreements. Thus, when we consider scenarios of small cell network sharing for traffic offloading, the neutral host or SCO (as introduced in Chapter 1) has the role of the Hosting RAN Provider.

Regarding the Participating Operator, it is an entity that makes an agreement with the Hosting RAN Provider in order to serve its users with the shared network, which can be used concurrently by more Participating Operators. Under the sharing agreement, the Participating Operator can use a portion of the aforementioned licensed spectrum (or its own spectrum) in order to serve its users, and specifically under its own control in order to set its own service policies (e.g. the resource allocation and scheduling for QoS/QoE provision).

2.2.1 Network sharing scenarios and use cases

In this section, we present various network sharing scenarios and use cases defined by 3GPP [20].

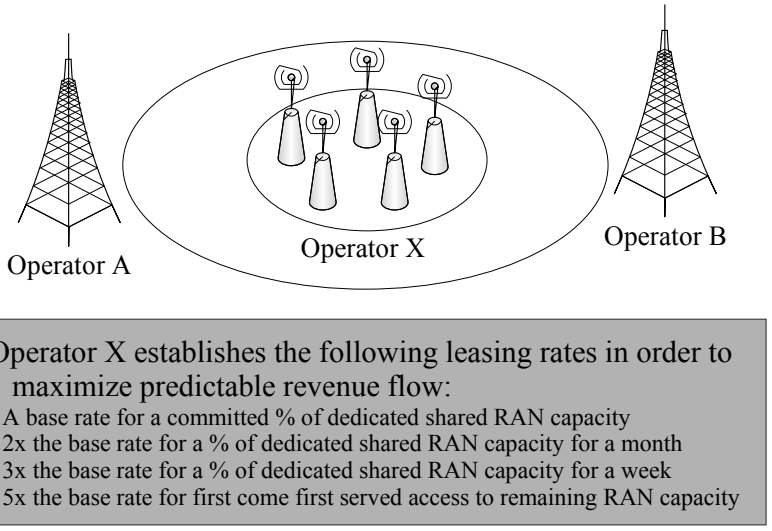


FIGURE 2.1: Example of leasing rates for Operator X's (i.e. the Hosting RAN Provider) Shared RAN in a Maximizing RAN sharing revenue scenario

Maximizing RAN sharing revenue

The efficient use of the network for the subscriber satisfaction, and the revenue maximization are two goals that all operators share. Nevertheless, in a network sharing scenario a trade-off may arise between these two objectives. The trade-off depends on parameters such as the offered load, and others regarding the sharing agreements, such as the number of the participating operators, the details of their demands (e.g. whether the demand is static or dynamic, the duration and the volume of resources etc.), and the sharing's economic parameters (e.g. additional service costs, and sharing prices/revenue). To this end, this use case proposes a variety of rates imposed by the Hosting RAN Provider on the Participating Operators, which depend on the period of time described by the sharing agreement. Hence, a basic rate can be charged for a permanent portion of dedicated shared RAN capacity, whereas the rate can be increased as the agreement's duration decreases (e.g. to months, weeks, on-demand agreements etc.), as illustrated in the example in Fig. 2.1. By doing so, the Hosting RAN Provider offers a flexibility in the sharing agreements, accommodating the Participating Operators different needs, while guaranteeing high revenues.

Asymmetric RAN Resource Allocation

This use case refers to Joint Ventures, which are created by two Participating Operators, with the purpose to build, operate and maintain the hosting RAN. The "assymetry" in this use case refers to the difference in the investment level the Participating Operators have in the shared RAN (e.g. 60% of investment is done by one operator, and 60% of

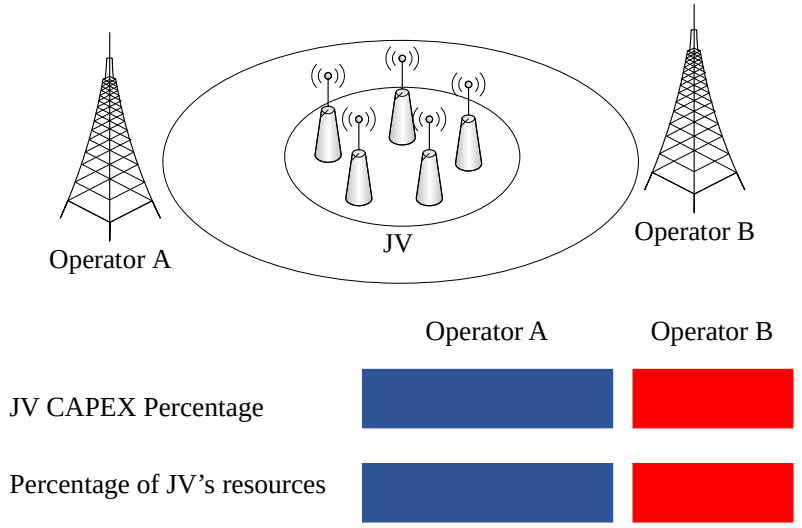


FIGURE 2.2: Asymmetric RAN resource allocation in the Joint Venture SC network of two operators

resources are dedicated for him), as shown in Fig. 2.2. Thus, during a peak-traffic hour the allocation and scheduling of the shared resources for the two operators is proportional to their respectful investment level. It should be noted that the sharing in this use case considers only the RAN and not the core network infrastructure.

Dynamic RAN Sharing Enhancements & On-demand Automated Capacity Brokering

In these two use cases, the Participating Operators' capacity requirements may vary during different time periods of the day or the week. Therefore, the Participating Operator demands a set of capacity allocations in a shared RAN for different time periods in the near future according to the expected traffic load, or on an on-demand basis for short-term additional capacity. These use cases mainly refer to Participating Operators, who already own RAN infrastructure but need additional capacity during high traffic time periods to serve their subscribers. With such flexibility, the Hosting RAN provider can manage its shared RAN in order to optimize the network performance and the financial gains. On the other hand, the Participating operators can optimize their subscribers' service by leasing the shared infrastructure according to the patterns of their traffic load, or guarantee the service quality during sudden traffic surges.

Typical examples for the application of both use cases involve the concentration of large crowds in small geographical areas, such as sports stadiums or open-air events (e.g. music concerts etc.), which are organized well in advance, and the operators have accurate estimates on the additional capacity they will need to serve properly their users. Such real-life example is the deployment of small cell infrastructure at the Mestalla stadium

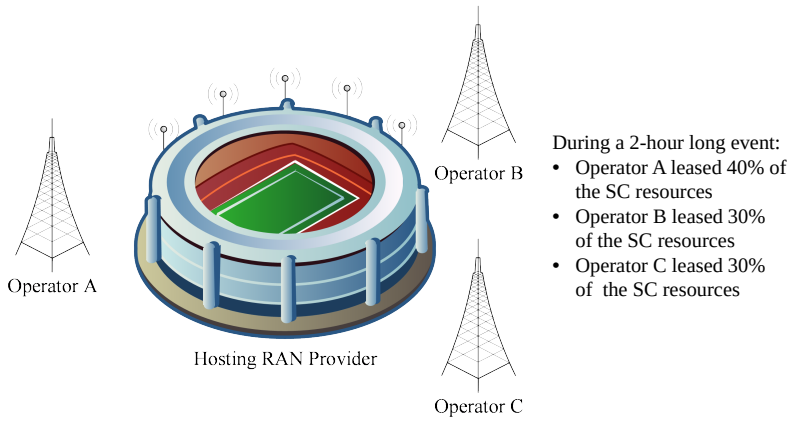


FIGURE 2.3: Leasing of small cell resources for an event in a sports stadium

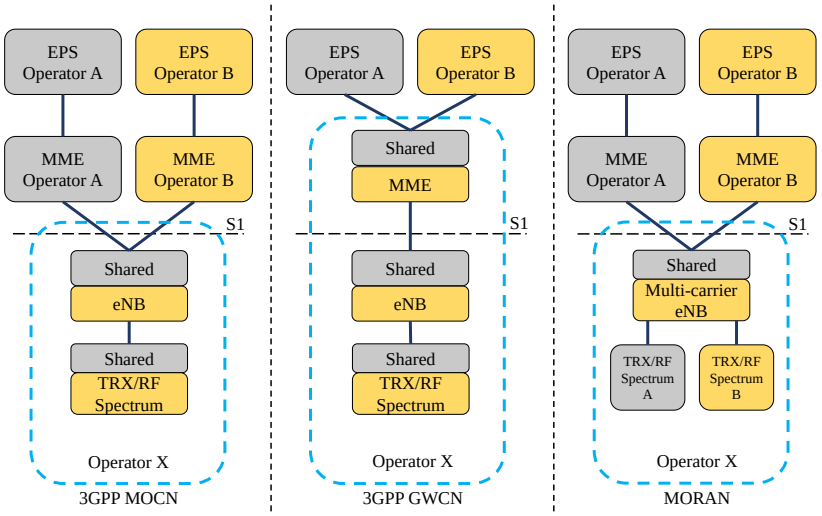


FIGURE 2.4: Three main approaches to network sharing (designed according to Fig. 4-2 in [1])

in Valencia by the collaboration of Telefonica and Nokia [5] (visual representation in Fig. 2.3). Thanks to the multi-tenancy capabilities of the installed infrastructure, the small cell capacity can be shared among multiple Participating Operators.

2.2.2 Network sharing architecture configurations

In this section we present the network sharing architecture configurations that can realize the network sharing use cases mentioned in the previous section. Three main approaches have won the attention of the telecommunications industry for application in LTE-A networks[1]. 3GPP has proposed the Multiple Operator Core Network (MOCN), and Gateway Core Network (GWCN) [21], whereas the Multiple Operator RAN (MORAN) is a non-3GPP configuration [1]. All three approaches are illustrated in Fig. 2.4 to

[1]These configurations can be also used as guidelines for future 5G network RAN sharing scenarios.

facilitate their comprehension and highlight their differences. We consider the example where the Hosting RAN provider is labelled as 'Operator X', and the Participating Operators as 'Operator A' and 'Operator B'. As it can observed, in the case of MOCN the sharing operators can share all the architecture elements along with the spectrum, excluding the core networks. With GWCN, the operators also share elements at the core network level.

With regard to the spectrum shared in MOCN and GWCN, it must be provided by a typical MNO, owner of licensed spectrum. This MNO would be one of the Participating Operators, who also acts as the Hosting RAN provider (operator 'X'). This means that MOCN and GWCN entail the ownership of spectrum by the Hosting RAN provider, which however is not required in the definition provided by 3GPP, as noted previously. Additionally, such a sharing configuration requires not only an infrastructure, but also a spectrum sharing agreement, which may not be always desirable by the MNOs (e.g. due to MNO competition). Moreover, a requirement for shared spectrum limits the possibilities for third-party, independent infrastructure owners from acting as Hosting RAN providers, due to the high spectrum license fees as explained in Chapter 1 [7]. In turn, this limits the general deployment of small cell networks, which is imperative for the increase of the network capacity through traffic offloading, and one of the requirements for future 5G networks.

These issues can be overcome with the use of the MORAN configuration. In MORAN, the MNOs share the backhaul connection and the BS hardware, except for the TRX/RF aspects, and their spectrum [1]. This means that MORAN is a configuration more flexible than the 3GPP approaches, as it enables the MNOs to have a higher level of control over the cell-specific radio parameters and radio resource management. Furthermore, they are free to employ their own service policies, such as service differentiation, and the provision of particular QoS/QoE levels. Moreover, the Hosting RAN provider does not need to acquire licensed spectrum, since the spectrum used for the operation of the shared RAN must be provided by the MNOs themselves. As a result, MORAN allows the emergence of third-party infrastructure owners, and the wide adoption of the SCaaS business model.

In order to better show the applicability of MORAN, in the following we provide two real-world examples that can be described as SCaaS (technically MORAN) use cases. The first example is the small cell network deployed in Los Angeles (USA) [4], where Phillips and the city Council have installed 4G LTE small cells in street light poles. The infrastructure, with Ericsson's technology, has been initially designed to provide IoT services, but it will be upgraded in the future to allow the MNOs to lease the infrastructure to deliver broadband connectivity.

The second example has been deployed in Mestalla Stadium, located in Valencia (Spain) [5]. In this case, the owner of the stadium deployed an SC network in cooperation with Telefonica, the largest MNO in Spain. The goal is to improve the visitors' mobile experience. Telefonica installed Flexi Zone small cells, nodes produced by Nokia and implementing MORAN [22], which allow not only Telefonica to deliver the service, but other MNOs to share the RAN. This deployment does not only address the broadband delivery but it also becomes a new source of revenue for the stadium owner. In both of these use cases, the small cell network owners can use as guidelines network sharing use cases described in Section 2.2.1 or in [20].

Furthermore, the third-party MORAN business model has been attracting new companies, who want to establish themselves as neutral hosts in the cellular network market. For instance, back in October 2014, Analysis Mason pointed out in [23] that towercos (independent tower companies, owners of mobile network passive infrastructure) were showing interest to enter the active RAN sharing market as neutral hosts. Such a towerco is Cellnex, which has been acquiring both macrocell and small cell sites, as well as infrastructure in fast rhythms over the years. Recent examples have been the acquisition of CommsCon, which manages telephony and data coverage at emblematic sites in Italy with the use of small cells [24], as well as the most recent agreements being the acquisition of 3000 new sites in France by Bouygues Telecom [25], and the incorporation of 2239 sites in Switzerland through a consortium acquisition of Swiss Towers AG [26].

2.3 State-of-the-Art

This section showcases a variety of approaches over a multitude of scenarios and use cases of the traffic offloading problem, as well as the use of network and economic functions for the improvement of the network performance and the MNO's economic benefits. Specifically, Section 2.3.1 describes the state-of-the-art regarding the solution of the traffic offloading problem, whereas Section 2.3.2 shows works on user association and resource allocation. Finally, Section 2.3.3 includes works on smart dynamic pricing.

2.3.1 Traffic Offloading

In the context of traffic offloading, several proposals have appeared in the literature. Furthermore, most of these studies have addressed the traffic offloading problem with a game theoretical approach, mainly modelling it with auction schemes [27–36]. This widely adopted game theoretical approach can explained by the fact that game theory enables the modelling of strategic interaction between two or more actors in situations

with strict rules and results, as in the traffic offloading scenarios described in the following.

A combinatorial auction scheme for the solution of the traffic offloading problem with WiFi Access Points (APs) is proposed in [27]. The authors design a reverse auction, where the MNO conducts the auction, and the AP owners bid to sell their unused WiFi capacity. In order to ascertain the AP owners' individual rationality and truthfulness, the authors introduce a novel payment rule, which is based on the well-known Vickrey-Clarke-Groves (VCG) scheme, and moreover accounts for the trade-off between the offloading cost and the gains in customer service. In order to tackle the high complexity of the optimal solution of the problem, the authors propose a low-complexity greedy algorithm, capable to solve the offloading problem in polynomial time even for large network use cases.

A double auction scheme among multiple MNOs and multiple third-party WiFi or femtocell access point owners, managed by an independent broker, is proposed in [28]. The MNOs place their bids in order to show their capacity demands, whereas the AP owners bid to set their capacity pricing. The broker guarantees truthfulness in bidding, hence maximizing the market's efficiency while maintaining its profit. The authors solve the double auction's Social Welfare Maximization (SWM) problem, and propose a low complexity, iterative, double auction mechanism, which manages to converge to the optimal solution. Moreover, the mechanism incentivizes participation, guarantees truthfulness and offers confidentiality, as the stakeholders do not need to disclose their utility functions to the broker.

An auction-based incentive framework for leasing on-demand bandwidth resources (referred to as iDEAL) is presented in [29]. In iDEAL, the offloading problem is formulated with the use of a reverse auction. In this auction, the MNO acts as the auctioneer and buyer. The bidders are third-party owners of WiFi hotspots, who bid to sell their unused capacity to the MNO. iDEAL takes into account the MNO's traffic spatial variations, conducting auctions in multiple geographical areas, and creating competition among the AP owners. By doing so, the MNO achieves substantial savings, by minimizing its service cost. Furthermore, iDEAL generates incentives to the third parties to sell the capacity, guarantees truthfulness in their bidding, and guards against collusion among them. Finally, iDEAL's performance and efficiency is evaluated with the use of simulations on real cellular network traces.

A multiple reserve price based auction mechanism, named EasyBid, is proposed in [30]. As in [29], the MNO is the auctioneer, and the femtocell owners are the sellers and bidders. The authors assume that the femtocell owners make imprecise estimations of their own valuations, and introduce the perceived valuation concept. In this scenario,

Easybid's objective is to find a multi-reserve price based solution that maximizes the MNO's utility. Finally, heuristic algorithms are used in order to conduct truthful auctions between an MNO and femtocell owners, taking into account the fact that the sellers have knowledge of their perceived valuations, which however could differ from the actual ones.

In a similar way, a VCG auction-based incentive framework for accessing selfish femtocells is studied in [31]. In it, the authors create a framework to incentivize femtocell owners to provide service to macrocell users, subscribers of a particular MNO. Particularly, the framework aims to maximize the system efficiency, and it is formulated with two auction schemes. A multi-unit reverse auction is conducted for a single macrocell user scenario, whereas a double auction scheme is conducted for multiple macrocell users. In both auction schemes, the MNO is the auctioneer, the femtocell owners are the bidders, and truthfulness is guaranteed for the placed bids.

An auction-based marketplace that assists MNOs to rent unused capacity from residential users, owners of WiFi or femtocell access points, is proposed in [32]. In this work, the authors design the marketplace as reverse, combinatorial auction, where the MNO is the auctioneer and the residential users are the bidders and sellers. The use of this auction scheme forces the sellers to bid truthfully, hence avoiding market manipulation. The MNO needs to solve the allocation of the resources available in the auction, which the authors formulate as an MNO cost minimization problem. However, due to the high complexity of the problem's solution (i.e. NP-hard), the authors further propose a greedy algorithm, which solves the allocation problem in polynomial time, even for large network instances, while guaranteeing the bidders' truthfulness.

In [33], the authors study the design of incentive mechanisms for third-party data offloading, taking into account a budget constraint for the Cellular Service Provider (CSP). In this paper, a CSP owns macrocellular infrastructure over a geographical area, which is divided into several smaller regions. The CSP aims to maximize the achieved capacity gain through offloading traffic to third party resource owners, while taking into consideration fairness in terms of QoS in the different regions. In order to achieve both objectives, the CSP runs a single-price[2], reverse auction with third-party resource owners (i.e. the bidders/sellers) in order to increase its network's capacity by leasing out the bidders' small cells. The authors propose two efficient competitive auction mechanisms, which do not depend on the knowledge or estimates of the future traffic, while taking

[2]Single or uniform-price refers to auctions where all the winning bidders pay the same price per unit of the auctioned good[37].

into account the CSP's limited budget. Both mechanisms are proved to be non-budget-deficit, individually rational, incentive compatible, and have a theoretical lower bound regarding their capacity gain.

Two game-theoretical proposals on the opportunistic offloading of Vehicular Users' (VU) data traffic through carrier WiFi (operated by the MNO) are proposed in [34]. The two proposals take into consideration the VUs' satisfaction, the offloading performance, as well as the MNO's revenue. The first proposal is an auction game-based offloading mechanism (AGO), whereas the second proposal is congestion game-based offloading mechanism (CGO). When AGO is employed, the MNO conducts auctions for selling WiFi access opportunities to VUs (i.e. the bidders). The authors designed the AGO mechanism according to the single-object ascending clock auction, which is proved to be truthful and individually rational. The AGO's use leads to social welfare's maximization, improvements in the offloading performance and MNO revenue. On the other hand it results in low VU utility along with fairness issues among VUs, due to the high offloading prices, and the possibility to WiFi access only by winning bidders. When CGO is used, all the VUs can access the carrier WiFi. Moreover, a VU makes offloading decisions after considering other VUs' actions, and its own utility maximization. The employment of the CGO mechanism leads to improvements in the offloading performance, and better VU utility and VU fairness compared to the AGO mechanism. For both mechanisms, the VUs are provided with a method for predicting the potential capacity gains from WiFi offloading, and its cost, in order to improve their decisions.

A trading marketplace for the leasing of WiFi capacity by a single MNO is proposed in [35]. The marketplace is modelled as a combinatorial reverse auction conducted by the MNO, and where third-party, WiFi AP owners bid their available capacity and its corresponding price. The allocation problem of the traffic to be offloaded is formulated as a combinatorial reverse auction, where the MNO's subscribers are covered partially. The authors propose a VCG-based payment rule, which secures the bidders' individual rationality and truthfulness, in order to avoid the manipulation of the market. The problem is solved both optimally and with the use of three greedy algorithms. The greedy algorithms are shown to preserve the individual rationality and truthfulness properties, along with their capability of solving the problem in polynomial time, even for large network instances.

A combinatorial auction framework for joint traffic offloading and BS switching off is proposed in [36]. The authors use traffic offloading as a means for the MNOs to decrease both the operation cost and the energy consumption, by switching off their macrocell BSs and serving their traffic through third party small cells. The proposed scheme consists of three steps. Initially, the MNOs estimate the future load and place multiple bids per

small cell in order to lease enough capacity to serve a portion of the expected maximum offered load. The rationale behind this bidding strategy is to avoid leasing unnecessary capacity, with the risk of not serving all of the MNO traffic. Subsequently, the third party decides on the capacity allocation and the pricing of the small cell resources. Finally, the MNOs decide on whether they will switch off their BSs according the auction result. The framework is modelled with a multi-objective optimization function, where the objectives are the maximization of the third party's profit, the maximization of the MNOs' added profits, and the minimization of the energy consumption.

A cost-sharing framework between two MNOs for the lease of an SCO's small cell infrastructure is proposed in [38]. In [38], traffic offloading outsourcing is used by MNOs to avoid the investment of small cell infrastructure, and the raise in total energy consumption of their networks. The authors propose a novel cost-sharing policy, which divides fairly the different expenses between the MNOs, by taking into account accurate estimates of the small cell leasing cost. In order to guarantee the accuracy of their cost assessments, the authors provide a realistic network-cost model. This model's variable expenses depend highly on the network's energy consumption, whose precise estimation is ensured by accurate forecasts of the expected traffic load.

Traffic offloading through a third party, with WiFi or femtocell APs, using non-cooperative game theory is considered in [39]. The authors use a two-stage multi-leader multi-follower game, called data offloading game (DOFF), where the BSs offer prices to the APs, and the latter decide the offloading volume. The authors solve the social welfare maximization problem, and examine the market behaviour with a game theoretical analysis. Particularly, they characterize the Nash Equilibrium for two typical market scenarios: a *perfect competition market*, and a *monopoly market*, and examine the impact they have on traffic volume and the corresponding leasing price.

A distributed market pricing framework, where mobile data flows are offloaded to APs, is proposed in [40]. The problem is formulated with a multi-leader multi-follower Stackelberg game, where the leaders are service providers and the followers are consumers within the aforementioned framework. With regard to the APs, the authors consider the cases where their offloading capacity is limited and unlimited. Their analysis leads to the establishment of the existences and uniqueness of the Stackelberg equilibrium in both cases, and a closed form is provided for the unlimited capacity case. Moreover, a distributed pricing algorithm is proposed for the convergence to an equilibrium, which is shown to be approximating the social optimum.

A refunding framework for offloading MNO macrocell traffic to small cell holders' (SHs) networks with limited capacity is proposed in [41]. The problem is formulated with a two-stage refunding-admission game, where the MNO is the leader, and the SHs are the

followers. In the first stage, the MNO maximizes its revenue, whereas in the second stage the SHs make decisions between offloading roaming macrocell users and receiving a refund, and satisfying their own requirements on QoS. The MNO's and SHs' decisions take into account constraints on the transmission power, interference, the Signal-to-Interference-plus-Noise-Ratio (SINR), and the small cells' backhaul capacity. The authors propose two refunding schemes: access-based refunding when the service provides guaranteed QoS, and usage-based refunding when the SH offers best-effort service.

An economic framework for traffic offloading to privately owned femtocells is described in [42]. The femtocells can be accessed by an MNO's public users through a hybrid access mode, and profit sharing is used to motivate the femtocell owners to offload traffic. A two-stage sequential game is modeled for revenue distribution, resource allocation, and service selection. Particularly, in the first stage the MNO decides on the revenue ratio that will be distributed among the participating femtocells. Next, in the second stage the femtocells determine the volume of resources that will be reserved for their subscribers' and the MNO's users. At the same time, the users (both fetmocell subscribers and public users) decide on their service, that is, whether to connect to the macrocell or a femtocell.

Similarly, in [43] the authors propose an ACcess Permission (ACP) transaction framework, which enables an MNO to buy ACP from multiple Femtocell Service Providers (FSPs). This framework provides also incentives to the FSPs in order to lease their available resources. The FSPs determine if and at which locations they will lease their resources, and competition among them arises when their infrastructure covers the same geographical area. Given the MNO's traffic uncertainty, and in order for the FSPs to yield the best strategies, the authors propose an adaptive strategy updating algorithm, based on an online learning process, known as the non-stochastic multi-armed bandit problem. It is further proven that the algorithm strategies' and the optimal strategies that provide the maximum payoff, yield a particular maximum payoff gap.

The data offloading problem is studied for the first time with the use of an one-to-many bargaining model in [44]. The scenario is modeled as a monopoly, where one MNO bargains with third party access point owners (APOs) for offloading its traffic. The bargaining is designed and analysed both in a sequential and a concurrent manner, and the Nash Bargaining Solution is provided for both bargaining protocols. The authors discover the APO benefits from bargaining in earlier steps (sequential bargaining), and the APO losses when the bargaining is conducted concurrently. Finally, the authors investigate the impact that APO grouping has on the bargaining solution.

In [45], although the scenario is similar, the situation differs, since all of the service providers' subscribers also play a major role in the dynamic selection of their service

provider. A hierarchical dynamic game framework models the interactive decisions, and an evolutionary game describes the subscriber's service selection. As for the transactions between the MNO and the SSP, they are modeled with a Stackelberg differential game.

The problem of traffic offloading through a third-party WiFi AP from the MNO's perspective is examined in [46]. The MNO aims to maximize its revenue when Successive Interference Cancellation (SIC) is applied to the system. The problem is solved with both a centralized approach and a threshold-based distributed offloading scheme, showing the SIC benefits on the MNO's revenue.

After reviewing the works presented in this section, we observe that in all works the SCOs provide not only infrastructure, but also spectrum resources. This means that the SCOs are actual service providers, who can have subscribers' of their own, and their business model can be completely independent from the typical MNOs. Thus, out of the three network sharing architecture configurations described in Section 2.2.2, the literature review considers only the MOCN or the GWCN configurations. As a result, the outsourced traffic offloading problem has not been addressed for the MORAN configuration, or the SCaaS business model, which we have described in Section 1.1.

Except for the lack of works on MORAN, we have also observed that the literature is mostly focused on the economic aspects of traffic offloading. Particularly, these works solely focus on the way the stakeholders interact with each other (e.g. auctions, bargaining etc.) in order to fulfil a particular objective (e.g. cost minimization, social welfare maximization etc.) and how it impacts their economic parameters. However, these works fail to address the technological aspects of traffic offloading and their connection with the financial aspects. For instance, we have not seen how the spectrum use in different tiers (i.e. orthogonal or co-channel), the limitations in backhaul or the small cell cluster density may impact the system throughput and the end users, and how the stakeholders act in different use cases. As the stakeholders' actions determine the network performance and with the economic gains or losses, it is interesting to examine the strategies they will adopt for different use cases described by technological parameters.

2.3.2 User Association and Resource Allocation

In this Section we outline proposals for the use of the user association and resource allocation network functions for the improvement of network performance. Particularly, we have not only included works that use these network functions as the tool to improve the network performance. As our research focuses on the use of network functions for both the provision of satisfactory service and the economic sustainability of the network (i.e. MNO profit maximization), we further include the included relevant works that

introduce an economic objective or constraint, which show us that this approach has not been thoroughly examined by the research community.

User Association

A proactive method for the solution of the user-cell association problem in small cell (SC) networks is proposed in [47]. In [47], the authors use the users' context regarding their data demands on specific applications in order to improve the decisions on cell selection. Particularly, they make use of past information on the service of users in order to make predictions for the users' QoE -in the Mean Opinion Score (MOS) scale- through collaborative filtering techniques. A cost function is defined for each user as the relative error between the actual and the predicted QoE. Then, the user association problem is formulated as the minimization of the sum of all the cost functions in the system. In order to tackle the high complexity of the problem (NP-hard), the authors design a matching game with externalities, and solve it with a low complexity decentralized algorithm, which always converges to stable matching, and achieves lower bandwidth utilization compared to conventional cell selection techniques.

The authors in [48] propose a UE context-aware approach for associating users to small cell base stations, aiming to improve the users' QoS. Particularly, the UE context taken into consideration refers to the application that is active at a specific time, and its QoS requirements (i.e. data rate, delay and packet error rate), as well as the UE's hardware type (i.e. smartphone, tablet and laptop). Specifically, the hardware type is used to prioritize one application over the others (e.g. a video streaming application has the highest priority for a laptop, but not for a smartphone). Subsequently, the above QoS metrics and the UE's hardware priority are used to define each user's utility function. Then, the user association problem is formulated as the maximization of the sum of the users' utilities with the use of a matching game with externalities. In order to avoid the high complexity of the problem's optimal solution, the authors propose a decentralised algorithm where each UE and BS create a preference list for each other, leading to stable matchings in low number of iterations, and better QoS performance compared to traditional user association algorithms.

The user association problem in an SC network operating in the 60-GHz band is studied in [49]. The authors' objective is the minimization of the maximum resource utilization in the system. The formulated resource minimization problem is a combinatorial, non-convex problem and NP-hard. The problem is reformulated based on the Lagrangian duality theory and solved with a distributed association algorithm, with a projected sub-gradient method. The proposed algorithm is time efficient and converges asymptotically

to the optimal values, utilizing substantially less resources compared to conventional association algorithms.

The problem of re-associating users within a network of WiFi access points under cost-migration[3] constraints is studied in [50]. In [50], the authors aim to improve the bandwidth utilization and the user experience in the system, by migrating some users' service to new APs after an initial AP selection. The problem is formulated as the maximization of the minimum user throughput with the constraint of maintaining the migration cost under a particular value. The problem is solved by a dual-stage approximation algorithm named CACA (Cost-constrained Association Control Algorithm). In the first stage, the algorithm solves the sub-problem of user removal from the initial user association, whereas in the second stage it solves the problem of associating these users to new APs.

The problem of user-cell association and the service scheduling in a two-tier network is studied in [51]. The paper's objective is the minimization of the load served among all the BSs in the system, which is formulated as the minimization of the service time of all the users' requests. This problem is reformulated and solved optimally by a sequential fixing algorithm, which however has a high complexity. To this end, the authors propose three near-optimal approximation algorithms; a rounding, a greedy and a randomized approximation algorithm, for which they calculate the complexity and the upper bounds for the maximum expected service time. Finally, the service scheduling is handled by a greedy algorithm, which is shown to minimize the expected waiting time.

The joint cell selection and resource allocation problem in a video transmitting wireless heterogeneous network is studied in [52]. The authors in [52] design a multi-objective optimization framework, aiming to maximize the users' QoE, and minimize the power and bandwidth consumption. The authors solve optimally the multi-objective problem and find the Pareto optimal solutions by using a weighted Tchebycheff approach, and a dual decomposition technique, revealing the tradeoffs between the design objectives.

So far, the reviewed works have network-related objectives (e.g. improvement of resource utilization). In the following, we present studies which not only focus on network performance improvement, but also on economic objectives. Particularly, the majority of these works follow a *greener-network* approach, aiming at cost reduction through proposals on energy efficiency.

The user association problem for energy cost minimization in heterogeneous networks with wireless backhauls is studied in [53]. In this system, the macrocell BS has an optical fiber backhaul to the core network, whereas the small cell BSs employ a wireless

[3] In [50], the migration cost is defined as the number of migrated users.

backhaul solution through the macrocell, using multi-hop in a tree topology. Both BS types receive power both from the main power grid and from renewable resources. The authors formulate the joint user association and green energy allocation as an energy cost minimization problem. In order to avoid the optimal's solution high complexity they propose a four stage solution, with the use of four low complexity algorithms; an algorithm to estimate the necessary energy consumption, a green energy allocation algorithm, a user association algorithm, and a green energy reallocation algorithm for increasing the efficiency of the green energy consumption.

The joint user association and sub-channel power allocation problem is studied in [54], for the cost-effective interference coordination in dense HetNets. In [54], the authors also take into consideration the user generated traffic, which may result in the switching-off of small cell BSs with low offered loads for a greater cost-efficiency. The aforementioned problem is formulated as the maximization of the difference between the aggregate utilities (functions of the BS total rate) of all the BSs in the system minus the total BS cost. The initial convex optimization problem is reformulated to its Lagrangian dual problem, which is solved by the proposed subgradient search algorithm.

The joint switching-off and user association problem in HetNets consisting of cellular macrocells and WiFi APs is studied in [55]. The authors aim to minimize the HetNet's total cost, and formulate the problem as the minimization of a function that balances the HetNet's energy consumption and the revenue of its cellular network assets. The problem then is divided into a user association and a BS switching-off sub-problems. For the user association problem, the authors provide an optimal association policy, which depends on the energy efficiency and revenue. Since the BS switching-off problem cannot be solved in polynomial time, they provide two low-complexity heuristic algorithms; one switches off the BS that generates the highest cost gain when turned off, whereas the other algorithm switches off the BS with the most users served by other APs, within its coverage area.

The authors in [56] study jointly the cell selection and antenna allocation problem in 5G massive Multiple Input Multiple Output (MIMO) networks. In this work, the users have different QoS requirements, and compete among themselves in order to be associated to the BS that maximizes their rate-based utility. Similarly, the BSs can modify the allocation of their antennas to different users in order to maximize their revenue. The problems of cell selection and antenna allocation are formulated as subgames of a hierarchical evolutionary game framework. In order to achieve an equilibrium, the authors propose two algorithms; one for the deterministic version of the evolutionary game, which is based on replicator dynamics, and one for the stochastic version, which is based on a Markov chain.

Resource Allocation

The joint subcarrier and power allocation problem for multiuser orthogonal frequency division multiple access (OFDMA) systems is studied in [57]. The authors follow a user-centric approach for improving the quality perceived by the users, by taking into consideration application-layer parameters and subjective human perception (i.e. MOS) in the resource management. The problem is formulated as the maximization of the minimum MOS (max-min MOS). For its solution, two user-oriented algorithms are proposed. The first algorithm solves suboptimally the max-min MOS problem, guaranteeing that all users share the same QoE. On the contrary, the second algorithm uses a QoE-aware approach and provides the trade-off between the appropriate QoE level and the system spectral efficiency.

Similarly, the joint resource and power allocation problem is studied in [58] for the optimization of the QoE in wireless networks. Particularly, the authors aim to maximize the minimum MOS in the system, with a constraint on the minimum number of users that should receive this satisfaction level. In order to solve this problem, the authors propose a heuristic QoE-aware resource and power allocation algorithm that satisfies the aforementioned requirement, improving the fairness in the system compared to algorithms that aim to straightforward QoE maximization.

The problem of proportional fair scheduling in multi-cell OFDMA networks for QoE maximization is addressed in [59]. A concave utility function for QoE provision is used to achieve global optimality. This utility function is based on the MOS model in order to consider application-specific characteristics. The authors formulate the problem as the maximization of the sum user utility, which they solve with the use of opportunistic gradient scheduling.

The multiple-service, class-based bearer-level resource scheduling problem in congested LTE networks is studied in [60] and [61]. In both works, the authors formulate the scheduling of LTE resources to service bearers as a multi-objective optimization problem, aiming at the maximization of the throughput, fairness provision among the users, and QoS guarantee by minimizing the loss and delay. In [60], the optimization problem is mapped to a Proportional Fair Knapsack problem with minimal manipulation. Due to its NP-hardness, the authors reformulate the problem by simplifying the objective functions with Gaussian weights, which is then solved by the proposed heuristic algorithm. In [61], the problem is solved with the use of a two-level Fair-QoS Broker algorithm. The algorithm's higher level provides per-class fairness for all service classes with the

application of a game theoretic model. The lower level of the algorithm allocates resources optimally for improving the QoS and the throughput, using a greedy-knapsack algorithm.

The issue of QoE-based coordinated resource allocation in the downlink of SC clusters with transport network constraints is addressed in [62]. The authors aim to maximize the aggregate user QoE, which they formulate as a convex maximization problem. The optimal solution of the problem is provided by forming the Lagrangian dual and employing the subgradient projection method. In order for the authors to provide a practical solution to the problem (i.e. discrete number of physical resource blocks to be assigned per user, instead of the real number provided by the optimal solution), they design a low-complexity heuristic algorithm, which uses the marginal QoE as the metric for the allocation decisions.

The problem of joint QoE and energy aware load management, power allocation, and channel allocation in small cell networks is addressed in [63]. The paper's objective is the maximization of the aggregate utility of the users in the system, which is a function of their perceived QoE minus part of the operational and energy cost of the small cell serving them. In order to solve this NP-hard, combinatorial problem, the authors divide it into two sub-problems addressed with an iterative manner. Specifically, one problem for the optimization of the power allocation and load management, while the other game optimizes the channel allocation. Furthermore, they propose a two-dimensional-action extended weakly acyclic game in order to solve the two problems optimally and in a distributed way. The solution of the problems is achieved through the use of two types of best response algorithms, which are considered to be executed with the assistance of virtual agents acting in the cloud.

The problem of achieving proportional fair energy efficiency (PFEE) with resource management in energy harvesting-based wireless networks is studied in [64]. Due to the complexity of the maximization of the PFEE problem, the authors convert it to a tractable form through Lagrangian relaxation. For the solution of the converted version, they propose a sub-optimal, PFEE, resource allocation algorithm based on a block coordinate descent method, which determines the subchannel and power allocation.

The issue of optimizing the the spectral and energy efficiency, as well as delay performances using dynamic resource allocation for delay-sensitive traffic in cloud-RANs (C-RANs) is addressed in [65]. In detail, the authors design a hybrid coordinated multipoint transmission (H-CoMP) scheme for the downlink in C-RANs, and analyse the relation between the fronthaul consumption and cooperation gain. The formulation of the queue-aware power and resource allocation of the delay sensitive traffic is done with

an infinite horizon Markov decision process. A low complexity stochastic gradient algorithm is proposed for the solution of the problem, which manages to address future traffic uncertainty due to imperfect channel state information.

The joint optimization of QoE and power allocation in multi-cell mobile networks is researched in [66]. The problem is formulated as the maximization of the aggregate BS utility, which is defined as the product of the users' QoE (associated to a BS), divided by the BS's power consumption. The use of this novel utility function aims at the fair maximization of the user QoE, while minimizing the power consumption. Due to the non-convexity and the integer programming nature of the problem, the authors solve the two subproblems of subchannel and power allocation separately. The two problems are solved with the proposed two-step iterative optimization method, which is based on Lagrangian functions.

As mentioned previously, the majority of works on resource allocation and scheduling focus solely on the provision of high QoS/QoE or other technical aspects (e.g. power allocation, fairness etc.), as the studies shown up to this point. In the following, we review works on resource allocation or scheduling, which taking into account the impact of their proposals on economic aspects.

The problem of packet scheduling for fair QoS provisioning, while guaranteeing high revenues for the operator of a wireless network is studied in [67]. To this end, the authors propose a packet scheduling scheme that maximizes the utility of users who are described by different services, priority classes for each service, and various QoS requirements, while minimizing the network operator's revenue loss.

The problem of fair resource allocation for the maximization of the MNO revenue is addressed in [68]. In [68], the users may renegotiate with the MNO regarding their service level in terms of QoS requirements. Moreover, when the available bandwidth does not suffice for the satisfaction of all the user demands, the users compete among them for the limited bandwidth resources in an auction scheme conducted by the MNO. This auction-based resource allocation scheme is formulated as a multi-objective optimization problem, which maximizes the Jain's fairness index, the bandwidth resource utilization, and the MNO's revenue.

A method for the simultaneous resource allocation in both licensed and unlicensed bands in the small cells of a HetNet is proposed in [69]. The concurrent resource allocation in both bands for the small cell users is designed as an optimization problem, which has constraints on the interference imposed on macrocell-served users by the small cell-served users. In order to achieve fairness among the small cell-served users, the authors set a constraint that demands the satisfaction of the small cell users' minimum rate

requirements. The optimization problem is solved twice; the authors first maximize the users' sum rate, and then the MNO revenue. Both solutions present similar results, as the MNO revenue maximization for usage-based pricing services demands the maximization of the users' rate.

The problem of profit maximization for a broadcasting MNO through resource allocation is studied in [70]. Particularly, the MNO broadcasts a set of video contents (e.g. tv channels) to different users groups. Each user has her own utility function, which is defined as the perceived QoE minus the charge for the service. In this scenario, the authors aim to maximize the MNO profit, with QoE constraints for the users. In order to do so, they introduce a marginal-based principle for the maximization of the profit with resource allocation. This marginal profit principle is then used in a resource allocation algorithm, which determines the rates of the broadcast contents, by making profit maximizing decisions.

The trade-off between service provider profit and user utility maximization in OFDM resource allocation wireless systems with multiple service classes is researched in [71]. In [71], the authors aim to strike a balance between the two conflicting objectives for both the time and frequency division multiplexing. In order to achieve their objective, they solve the multi-objective problem using Lagrangian relaxation, and find the Pareto-efficient policies that satisfy best both the service provider and the users.

The problem of cost-effective resource allocation in mobile networks that employ cloud-RAN and mobile cloud computing is addressed in [72]. In this scenario, the mobile cloud's role is the computation of the users' tasks, which are then forwarded to the users through the C-RAN. The problem is formulated as the minimization of the MNO's cost, which is a function of the user rate, and the computation capability of the virtual machine serving the corresponding user, while satisfying the user, mobile cloud, and fronthaul constraints. This non-convex problem is converted to its equivalent weighted minimum mean square error problem, which is then solved by an iterative algorithm.

As we can observe in this section, the majority of works on user association, resource allocation and scheduling in the context of 4G and 5G networks focus mainly on the provision of high QoS/QoE and other network aspects (e.g. power allocation, fairness etc.). However, these works fail to take into account the impact of their proposals on the financial aspects of the MNOs (e.g. cost, revenue etc.). As we are entering into the 5G era, and the MNOs need to serve satisfactorily their subscribers and generate large revenues (for the compensation of their infrastructure investments), it is important to analyse the connection between QoS/QoE provision policies and MNO economic parameters, and contribute with proposals that address both the network and economic objectives.

2.3.3 Dynamic, Smart Pricing

In this section, we present works that make use of dynamic smart pricing as a mean to improve not only the MNOs' revenues, but also the subscriber satisfaction. As shown next, most dynamic pricing schemes in the literature are based on time or location dependent pricing, which aim to steer the traffic demand from peak to off-peak traffic hours and areas [73–78]. In [74], the authors explore charging schemes to be used in time-dependent pricing for the maximization of an Internet Service Provider (ISP) in a monopoly market. In this market, the users decide whether they will use a particular wireless data service for a specific time period, depending on their valuation of the service, and the service's price, which is determined dynamically by the ISP. The charging schemes under study are the usage-based, the flat-rate, and the cap with meter[4] schemes. In order to examine the profitability of each scheme, the authors design a two-stage Stackelberg game (the ISP is the leader) for all of the charging schemes, and obtain each game's equilibrium with the use of backward induction.

The problem of traffic steering utilizing time dependent adaptive pricing (TDAP) with threshold policies is addressed in [75]. In [75], an ISP offers mobile internet services, charging its subscribers with a K-class cap with meter scheme. This means that the ISP offers various applications with K distinct classes, each of them defined by a particular monthly data capacity threshold, user rate, and price. In order to apply TDAP, the ISP divides a 24-hour day in timeslots, and counts the volume of traffic served within each timeslot. Subsequently, the ISP determines next day's charging rate of each timeslot in order to incentivize the users use their applications during off-peak traffic hours. The charging is determined according to two policies; the Multiple Threshold Policy (MTP) and the Linear Threshold Policy (LTP). In both cases, the charging increases with the usage. In MTP there is a discrete and limited number of thresholds, whereas in LTP, the charging is determined by a linear function, hence there can be infinite charging values.

The authors in [76] demonstrate the importance of location for improving the efficiency of time dependent pricing. Particularly, after conducting a large-scale trace-driven analysis they discovered that a single TDP scheme cannot be applied in every urban area, since they are described by different traffic patterns (e.g. business and residential districts). In order to address the spatial heterogeneity of the data traffic, they propose applying TDP separately at each BS, based on the corresponding traffic pattern, and name this scheme Location Dependent Pricing (LDP).

[4]The cap with meter scheme refers to the typical mobile data plans, where a user enjoys a flat rate charging for an upper limit of data volume. When this data limit is surpassed, the user is charged according to their data consumption.

The authors in [77] steer the traffic demand through a novel TDP scheme in order to achieve greener mobile networks and subscriber satisfaction. In order to achieve their goal, they propose the integration of all the parameters that constitute a telecommunications bill, such as QoS, application, and time dependent dynamic pricing factors into a single charging metric, the eBit. The eBit has a single monetary price per unit, however different applications and QoS levels correspond to a different number of eBits per data units. By employing eBit with usage-based pricing, the authors expect that the users will perceive a more satisfactory service, since their responsible service consumption will eliminate congestion, increase the QoS, and the charging fees will correspond to their actual consumption. For the successful adoption of eBit, the authors propose the regular information of the subscribers, so that they can understand their current charging, as well as their current bill at any particular point of consumption.

The increase of MNO revenue, along with improvements in the resource utilization efficiency, and the user satisfaction through time-dependent pricing is addressed in [78]. In [78], the time-dependent pricing scheme is implemented with reverse pricing. This means that the users disclose their willingness to pay for a particular service by reporting through bidding the desired volume of data for a particular price. If a placed bid is above a specific threshold, which is set secretly by the MNO, the users can be served. In order to achieve the aforementioned objectives, and avoid the manipulation of prices through false reporting, the authors formulate the reverse pricing scheme with a four-stage Stackelberg game. In the first stage, the MNO sets a revenue maximizing, initial unit data price, based on predictions for the traffic demand. In the following stage, the users announce their data demands according to the initial unit price, and indirectly their willingness to pay. Subsequently, the MNO determines the resource recommendation rule, and the hidden bid-acceptance threshold. Finally, the users that can be served decide whether or not they will receive the service for the price they had placed a bid.

The problem of revenue maximization in time-varying multi-hop wireless networks is addressed in [79]. In the system described in [79], multiple flows associated to users share the resources of the network. Each flow has a particular rate requirement, whereas the delay depends on the service level requested by the user, which determines the flow's priority (in terms of delay), and its price. Moreover, each user has a utility, which depends on both the flow's rate and delay, but also on the flow's price. This means that even though a flow's rate and priority may be high, the corresponding high price may affect the user's experience, and indirectly her service requests. The authors formulate a revenue maximization problem with constraints on the network stability, and the QoS requirements of the flows. For the solution of the problem, a quality-aware dynamic price (QADP) algorithm is proposed. QADP is an online policy, which dynamically

adapts weights on each flow, sets a price on them, and finally determines the scheduling that maximizes the revenue under the aforementioned requirements.

Up to this point, we have seen that dynamic pricing is mainly used as a tool to motivate the users change their data consumption habits, in order to avoid network congestion. However, such approaches do not provide solutions for high customer satisfaction during inevitable congestion periods. In the following, we present works where the service provider adopts a dynamic pricing scheme, which is used to address the trade-off between user satisfaction and revenue maximization during congestion.

The issue of a single MNO's revenue maximization through utility-based dynamic pricing is studied in [80]. The initial objective of this paper is the maximization of both the system utility (i.e. the aggregate user utility), and the MNO's revenue. However, as there is a trade-off between the two objectives, the authors propose a scheme that adapts the resource allocation and pricing to the point that it maintains a near-optimal MNO revenue, while inflicting an acceptable low degradation of the system utility compared to its maximum possible.

The authors in [81] study the revenue maximization problem of a cloud provider through a joint virtual machine (VM) scheduling and pricing approach. The authors introduce the notion of partial utility, which is defined as a client's willingness to pay when receiving a specific degradation in the service quality. By using partial utility, a cloud provider can redistribute VM resources from a low-class, low-charge client request to a different client request, which will maximize the provider's revenue. However, this service degradation does not affect the former client, as long as an appropriate discount is offered, which is determined based on her willingness to pay.

Taking into consideration the presented works, we observe that they consider dynamic pricing as the means to steer the traffic demand and direct it from peak to off-peak traffic hours and locations, in an effort to minimize congestion periods. That is, dynamic pricing is used as a tool to motivate the users to change their data consumption habits, thus avoiding network congestion. However, there is a lack of works that examine real-time dynamic pricing schemes that are applied during inevitable congestion periods in mobile networks in order to improve the subscriber satisfaction, in a manner similar to the proposal in [81] for cloud computing services.

2.4 Open Issues and Challenges

As shown in the previous sections, extensive research has been conducted over the years in outsourced traffic offloading, and a plethora of schemes for network and economic

functions have been proposed. Even though the contribution is considerable, there are still issues that have not been addressed and approaches that have not yet been implemented.

Regarding the literature of outsourced traffic offloading, we have already mentioned that there is a lack of works on the MORAN sharing approach, and the corresponding SCaaS business model. As it is a business model that has already been applied [4, 5, 24–26], and is one of the cornerstones for the densification of networks in the 5G era, it is imperative that research should be conducted, contributing with a plethora of approaches for its efficient implementation in real-life scenarios. Our contribution on outsourced traffic offloading for the SCaaS business model is provided in Chapter 3.

As for the research done on user association and resource allocation, the majority of studies is done towards the improvement of specific network parameters (e.g. QoE, fairness, energy efficiency etc.). Only a small part of the research community has studied the economic impact of their proposals on network functions, and the connection between network and economic parameters has not been thoroughly investigated. We present our contributions on user association and resource allocation in Chapter 4.

Similarly, the majority of works on smart dynamic pricing focus on the approach of traffic steering for the avoidance of congestion. On the other hand, the connection of pricing to the customer satisfaction and the possibility of exploiting this connection through pricing schemes during congestion with real-time schemes has not been properly addressed. Our contribution on dynamic pricing is presented in Chapter 4.

Taking the above into consideration, we understand that there is a wide area to be researched regarding the connection between the technological and financial aspects of mobile networks. From the strategies MNOs use for traffic offloading transactions, to the use of network and economic functions for both the customer satisfaction, and the financial sustainability of the network.

2.5 Concluding Remarks

In this chapter we have presented background information and the literature review, which are the basis of our contributions described in the following chapters. Initially, we present various network sharing scenarios and use cases, as well as network sharing architecture configurations. This background information are the foundation for the study of all the traffic offloading scenarios and use cases presented in this chapter's literature review, as well as our own contribution on outsourced traffic offloading with SCaaS.

Subsequently, we provide the state-of-the-art on outsourced traffic offloading, which even though has been studied by a plethora of researchers, its main focus has been placed on the financial aspects of the subject. To that end, our contribution on traffic offloading deals both with its financial and technological aspects, and especially the connections between them.

The second part of the literature review presents the use of network (i.e. user association and resource allocation) and economic (i.e. dynamic pricing) functions for improved network performance and financial output. As the majority of proposals on network functions examine solely their impact on the network performance, our contributions on both user association and resource allocation address both the network and economic objectives of an MNO, as well as analyse their relations. Regarding dynamic pricing, the state-of-the-art mainly treats it as a traffic steering tool in order to avoid network congestion and thus guarantee satisfying service provision. Due to the lack of works that use dynamic pricing in real-time, we propose a scheme that applies it during congestion periods and improves the user satisfaction. Finally, we summarize the open issues and challenges regarding traffic offloading, user association, resource allocation and dynamic pricing.

Chapter 3

Network and Financial aspects of traffic offloading

3.1 Introduction

The expected exponential increase in mobile data traffic during the next few years [2] burdens the MNOs with maintaining a sustainable capacity growth for meeting these new demands. This need for capacity growth can be satisfied with the deployment of new infrastructure, and particularly with the densification of the RAN with small cells. Particularly, the use of small cell infrastructure can be used for mobile data offloading, and is considered as one of the most promising solutions to the aforementioned problem.

However, the ubiquitous deployment of small cell infrastructure by a single MNO can pose a high financial risk, as explained in the previous chapters. To that end, the MNOs tend to outsource their traffic offloading needs to independent third parties. These third parties, also known as Small Cell Operators (SCOs), are owners of small cell infrastructure, which they lease to MNOs when and where their network needs a capacity boost.

In order for outsourced traffic offloading to be established as a common practice in the telecommunications market, both sides (MNOs and SCOs) need to be provided with appropriate incentives. SCOs need to price their small cell infrastructure accordingly so that they can recoup their investment, and make a profit. On the other hand, the MNOs need to offload their traffic with a reasonable cost, which will maintain their profitability and hence justify outsourced traffic offloading over own deployment of small cell networks. Therefore, the traffic offloading problem is solved when the following

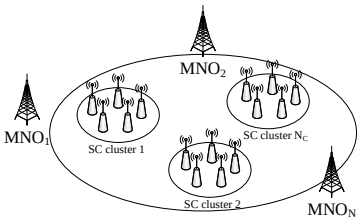
questions are answered. When, where, for how long and for how much money should an MNO request a traffic offloading service.

Apart from the description and formulation of the traffic offloading problem, there are also implementation issues. As shown in Section 2.3.1, the literature has focused on use cases where the SCO not only owns infrastructure, but also spectrum with which it serves the MNOs' users. This means that these use cases consider either the MOCN or GWCN network sharing architecture configurations, which we have described in Section 2.2.2. These two configurations may be plausible in scenarios where the Hosting RAN provider is an MNO. However, it is quite improbable for them to be adopted by an SCO due to the high cost the acquirement of licensed spectrum imposes, as explained in Section 1.1. Consequently, this requirement would impede the wide adoption of traffic offloading, as third parties would not be able to enter this market and alleviate the small cell deployment cost from the MNOs.

In order to provide a more realistic approach for the traffic offloading problem, we have studied the business model of the MORAN sharing configuration, also known as small cell as a service (SCaaS). With MORAN, the Hosting RAN provider (i.e. the SCO), offers access only to the backhaul and BS hardware, except for the TRX/RF aspects and their spectrum. Hence, the MNOs can plan the use of their spectrum resources at the leased infrastructure, and employ their own Radio Resource Management policies. Additionally, an SCO can offer SCaaS by simply deploying small cell and backhaul infrastructure, without the need for own spectrum license and core network.

The use of the MNOs' spectrum by an SCO may seem radical, considering the importance and value of this asset. As mentioned previously, MORAN allows MNOs to choose the service policy at the SCO infrastructure and also broadcast their public land mobile network (PLMN) identities [1]. That is, an SCaaS agreement makes the leased small cells to appear as part of the MNOs' networks, with the difference of being operated by a neutral host. Moreover, the above are guaranteed with legal means, that is, with the use of binding contracts in the form of Service Level Agreements [10].

This chapter examines the case where a monopolistic SCO offers SCaaS to multiple MNOs, hence addressing the scenarios with a single SCO in the area under study (e.g. stadiums, airports, shopping malls etc.) [82, 83]. Particularly, this chapter investigates the interaction of the capacity needs and the economic constraints of the stakeholders, and aims to gain insight into techno-economic implications of the SCaaS paradigm. We analyse the MNOs' auction strategies, and their impact on the system, and then propose a novel learning mechanism for improving the MNOs' bidding strategies. The main contributions of this chapter are summarized in the following:



(a)

SCO spectrum allocation	Player	Spectrum availability	Spectral Efficiency
Dedicated	MNO_i	$(1-X_i^D)B_i$	SE_i^D
	SCO	$\boldsymbol{x_i^D} B_i$	SE_{sc}^D
Co-channel	MNO_i	$X_i^C B_i$	SE_i^C
		$(1-X_i^C)B_i$	SE_i^D
	SCO	$\boldsymbol{x_i} B_i$	SE_{sc}^C

(b)

FIGURE 3.1: (a) System Model and (b) Spectrum Allocation for the two channel deployments

- We propose a realistic analytical model for traffic offloading under the SCaaS approach, considering multiple small cell clusters, while including both the technological constraints (e.g. bandwidth availability, backhaul capacity) and the financial goals of each stakeholder.

- The SCaaS model can be implemented under two spectrum deployment use cases: i) the dedicated spectrum deployment, and ii) the co-channel deployment. This work shows the profit-capacity trade-off for each use case and proposes the mathematical framework to define the capacity limits of SCaaS in each use case.

- As the stakeholders' profit and the total system capacity depend highly on i) the deployment density (i.e. density of deployed eNBs and small cells), ii) the competition level among MNOs (i.e. the capacity needs of each MNO), iii) the SCO backhaul capacity, iv) the reuse or not of the spectrum bands, and v) each actor's cost function, the presented results offer useful insights to select the adequate values of the parameters involved in the SCaaS approach.

- The proposed auction scheme relies on the perfect knowledge of the MNOs' future loads. As each MNO is not aware of its future load and the future load of the contending MNOs (due to future uncertainty, as well as privacy of sensitive information), we provide a mathematical framework based on a new learning mechanism to overcome the lack of reliable information.

The rest of the chapter is organized as follows. Section 3.2 describes the system model, and Section 3.3 states the stakeholders' objectives. Section 3.4 describes the auction scheme, the Social Welfare Maximization problem, and the learning mechanism devised to estimate the key variables in the absence of relevant knowledge. Numerical results and analysis are presented in Section 3.5. Section 3.6 discusses the applicability of our model in real-world scenarios and section 3.7 concludes the chapter.

3.2 System Model

The aim of this chapter is to study traffic offloading for the SCaaS business model, and gain insight into its techno-economic trade-offs. The scenario is characterized by multiple non-overlapping hotspot areas served by multiple MNOs and where a third party has deployed a number of SC clusters. As the hotspot areas are geographically limited, each MNO covers the hotspot areas with a single eNB[1], as shown in Fig. 3.1a. Therefore, the system model is constituted by one eNB per MNO and N_C SCO clusters (one in each hotspot area). The set of eNBs is denoted by $\mathcal{N} = \{1, 2, \ldots, N\}$ and the MNO owner of eNB $i \in \mathcal{N}$ is denoted by MNO_i. The set of SCO clusters is denoted by $\mathcal{N}_\mathcal{C} = \{1, 2, \ldots, N_C\}$, where an SC cluster $l \in \mathcal{N}_\mathcal{C}$ consists of N_{sc_l} small cells, and is connected to the internet or to the core network through a backhaul network with a capacity C_{BH_l} (in Mbps). C_{BH_l} is therefore the capacity of the link between the SC cluster l and the core network/internet, and it is upper bounded by the backhaul technology, such as a wireless millimetre-wave link or an optical fibre connection [84]. We assume a frequency reuse factor 1 for the SC tier. Although in Fig. 3.1a eNBs from different MNOs are not co-sited, the subsequent analysis can be also used for RAN sharing scenarios (i.e. when multiple MNOs share a single eNB), as it only assumes that the average spectral efficiency of the diverse eNBs is the same. The notation used henceforth is summarized in Table 3.1.

Each BS is characterized by the spectral efficiency, defined as the transmission rate achievable per bandwidth unit, and the available bandwidth. For a given MNO_i, the available licensed bandwidth is denoted by B_i (in MHz) and the spectral efficiency SE_i (in Mbps/Hz) can be approximated by $SE_i = \mathbb{E}[\log_2(1 + SINR_{ik})]$, where $\mathbb{E}[\cdot]$ is the mathematical expectation and $SINR_{ik}$ is the Signal-to-Interfernece-plus-Noise ratio of a user k served by MNO_i. For an SC cluster l, the spectral efficiency is defined analogously and denoted by SE_{sc_l}. In turn, the maximum bandwidth supported by each SC cluster is denoted by B_{sc} and it is limited either by the technology's specifications or by the deployed hardware's capabilities. Based on these definitions, the capacity of MNO_i, defined as the maximum throughput that can be served by MNO_i, is given by $C_i = B_i SE_i$. As for the capacity of the SC cluster l, it is given by $N_{sc_l} B_{sc} SE_{sc_l}$. Note that, whereas the traffic served by MNO_i is limited by its capacity C_i, the traffic served by the SC cluster l is limited either by C_{BH_l} or by the SC cluster capacity, i.e. $\min\{C_{BH_l}, N_{sc_l} B_{sc} SE_{sc_l}\}$.

[1]This study is focused on a single macrocell sector, hence all hotspot areas are considered to be located in the geographical area of a macrocell sector. However, the results can be extrapolated to a larger area that consists of multiple macrocell sectors.

Table 3.1: MNO and SCO Notation

Notation	Description
$\mathcal{N}$	Set of eNBs
$\mathcal{N_C}$	Set of SC clusters
N_{sc_l}	Number of small cells in SC cluster l
C_{BH_l}	Backhaul capacity in SC cluster l
B_i	eNB i's available bandwidth
SE_i	eNB i's expected spectral efficiency
C_i	eNB i's maximum capacity without offloading
B_{sc}	Maximum bandwidth supported by all SC clusters
SE_{sc_l}	SC cluster l's expected spectral efficiency
L_i	MNO_i's offered load
L_{h_i}	MNO_i's offered load in all SC clusters
$L_{h_{il}}$	MNO_i's offered load in SC cluster l
L_{n_i}	MNO_i's offered load outside of SC clusters
$\mathcal{T}$	Set of a 24-hour day equal-period timeframes
$\boldsymbol{x_i}$	MNO_i's transferred bandwidth vector
x_{il}	MNO_i's transferred bandwidth at SC cluster l
X_i	MNO_i's total transferred bandwidth percentage
L_i^{mc}	Load served by eNB i
L_i^{sc}	MNO_i's load served by all SC clusters
L_{il}^{sc}	MNO_i's load served by SC cluster l
L_i^T	MNO_i's total served load or throughput
D, C	Dedicated spectrum, Co-channel deployment superindexes
Θ_i	Throughput difference between co-channel and dedicated spectrum deployment
P_i	MNO_i's profit
$\boldsymbol{b_i}$	MNO_i's bid vector
b_{il}	MNO_i's bid at SC cluster l
R_i	MNO_i's revenue
CL_i	MNO_i's load cost
a_i, d_i	MNO_i's cost shaping parameters
P_{sc}	SCO's total profit
P_{sc_l}	SCO's profit at SC cluster l
CL_{sc_l}	SCO's cost at SC cluster l
a_{sc_l}, d_{sc_l}	SCO's cost shaping parameters
$P_{sc_l}^{min}$	SCO's minimum profit for SC cluster l
z	SC capacity pricing factor
b_{il}^{min}	MNO_i's reserve price for SC cluster l
$\overline{a}$	Forecast value of any parameter a

Let us define the offered load of MNO_i as L_i (in Mbps). This offered load can be divided into two components: the offered load generated within each hotspot ($L_{h_{il}}$ with $L_{h_i} = \sum_{l \in \mathcal{N_C}} L_{h_{il}}$) and the offered load generated elsewhere (L_{n_i}), i.e. $L_i = L_{h_i} + L_{n_i}$. Each of these components can vary in time and space. During time periods where L_{h_i} is low, the need for capacity provided by the SCO declines. Conversely, high hotspot loads result in a raising interest for the usage of the SCO infrastructure. Hereafter, the day is divided into a set $\mathcal{T} = \{1, 2, \ldots, T\}$ of equal timeframes, during which the

load is considered constant. It is nonetheless worth noting that, even though the load in a timeframe $t \in \mathcal{T}$ is not necessarily the same every day, it follows a daily pattern. Even though the real-time offered load and user SE do not remain constant during a timeframe, their instantaneous variations from their average (i.e. L_i, SE_i and SE_{sc_l}) can be disregarded when considering a timeframe's offloading decision [29].

As it will be shown in Section 3.4, an MNO_i needs to estimate both L_{h_i} and the corresponding bandwidth resources transferred to the SC clusters (henceforth denoted as $\boldsymbol{x_i}$), for serving its load as well as its profit maximizing objective. To that end, and since it is not possible to know beforehand the actual spectral efficiency of each user roaming the system, we use SE_i and SE_{sc_l}. However, this assumption introduces some limitations. Particularly, as a user roams a small cell, her spectral efficiency will be at times either higher or lower than SE_{sc_l}, leading to lower or higher requirement of $\boldsymbol{x_i}$ respectively. In case of higher $\boldsymbol{x_i}$ requirements, it is possible that part of L_{h_i} will not be served. This in turn can further lead to MNO revenue and profit loss as it will be shown in Section 3.3.2.

Regarding the provided SC capacity, it is true that it should reflect the MNOs' demand. However, increasing the capacity requires the further instalment of SC and backhaul infrastructure. Such deployments and the corresponding SC capacity pricing for the recoup of their investment are always part of a long-term business plan. Hence, since our system model examines SCaaS on a day-to-day basis and short timeframes according to the traffic's trends (e.g. hours), the study of additional SC infrastructure and its corresponding pricing are out of the scope of this contribution.

3.3 Stakeholders' Network and Financial Objectives

MNOs have a two-fold objective. First, they must guarantee the users' QoS. Second, the network must be managed so as to maximize their economic profit. On the other hand, the SCO aims to repay the deployment investment and generate profit. In the following, the analyses of the MNO throughput, and the stakeholders' profit are detailed.

3.3.1 MNO Throughput

In this work, we consider that the QoS of an MNO_i's users is guaranteed by serving the offered load. In other words, the users receive satisfactory service when $L_i^T = L_i$, where L_i^T (in Mbps) is the total load served[2] by MNO_i (through the own eNB and the SCO

[2]Henceforth, we will use the terms total served load and throughput interchangeably.

leased capacity). L_i^T is divided into the load served through the MNO_i infrastructure, L_i^{mc}, and the load served through the SCO, $L_i^{sc} = \sum_{l \in \mathcal{N}_\mathcal{C}} L_{il}^{sc}$. Therefore, $L_i^T = L_i^{mc} + L_i^{sc}$. Since the SCO uses the MNOs' bandwidth for the RAN operation of the SC tier, we define the *transferred bandwidth percentage* vector for MNO_i to all N_C SC clusters as $\boldsymbol{x_i} \triangleq (x_{il} : \forall l \in \mathcal{N}_\mathcal{C})$, $x_{il} \in [0, 1]$, and the total transferred bandwidth percentage of MNO_i is $X_i = \max_{l \in \mathcal{N}_\mathcal{C}}(x_{il})$. The transferred bandwidth percentages of all MNOs are given by matrix $\boldsymbol{x} = (\boldsymbol{x_i} : \forall i \in \mathcal{N})$. Hence, an MNO_i provides $x_{il}B_i$ of its bandwidth at each SC cluster l. Thus, the MNO_i will be allowed to offload a maximum load through the SCO equal to $L_i^{sc} = \sum_{l \in \mathcal{N}_\mathcal{C}} N_{sc_l} x_{il} B_i SE_{sc_l}$ (in Mbps).

Since the SCaaS approach requires the transfer[3] of licensed bandwidth from the MNO to the SCO, the way in which the transfer is conducted will impact the system performance. Specifically, the deployment of HetNets presents two use cases regarding the SC tier spectrum band deployment: the dedicated spectrum deployment and the co-channel deployment. The former (dedicated spectrum) is characterized by the orthogonal use of spectrum in the SCO and the MNO. Thus, an MNO_i that transfers a band $X_i B_i$ to the SCO only uses the non-transferred spectrum band (i.e. $(1 - X_i)B_i$) at the eNB. Conversely, in the latter (co-channel deployment) the spectrum is partially reused by the two tiers, and the eNB makes use of the whole band B_i regardless of the transferred spectrum to the SC tier, $X_i B_i$. These differences are outlined in Fig. 3.1b. In the sequel, superindexes D and C differentiate the dedicated spectrum deployment and the co-channel deployment parameters, respectively.

According to interesting studies, the achieved throughput depends on the use case (co-channel and dedicated spectrum), and the key parameter that differentiates them is the spectral efficiency of the SCO infrastructure, SE_{sc}, which is higher for the dedicated spectrum use case than for the co-channel use case ($SE_{sc}^D > SE_{sc}^C$) [85]. Thus, for a given offloaded load L_{il}^{sc}, the bandwidth required by the SCO in the dedicated spectrum case is smaller than the bandwidth required in the co-channel case, $x_{il}^D B_i < x_{il}^C B_i$. This fact causes differences in the maximum total served load.

Proposition 3.1 (Dedicated spectrum deployment throughput)**.** *Given an MNO_i that transfers $\boldsymbol{x_i^D} B_i$ MHz to the SCO, the throughput in a dedicated spectrum deployment is expressed as*

$$\begin{aligned} L_i^{DT}(\boldsymbol{x_i^D}) &= L_i^{Dmc}(X_i^D) + L_i^{Dsc}(\boldsymbol{x_i^D}) \\ &= (1 - X_i^D) B_i SE_i^D + \sum_{l \in \mathcal{N}_\mathcal{C}} N_{sc_l} x_{il}^D B_i SE_{sc_l}^D \end{aligned} \tag{3.1}$$

[3] As detailed in Section 1.1, the transfer of bandwidth is defined as a Service Level Agreement (SLA) between the legal licensee (the MNO) and the lessee (the SCO) by which the latter is allowed to use part of the spectrum of the former.

where SE_i^D is the average spectral efficiency of MNO_i in dedicated spectrum bands

$$x_{il}^D \leq \frac{L_{h_{il}}}{N_{sc_l} B_i SE_{sc_l}^D} = x_{il}^{Dmax}, \ \forall l \in \mathcal{N}_\mathcal{C} \tag{3.2}$$

Proof. L_{h_i} can be served completely by the SCO or jointly served between the SCO and MNO_i. If the offered load in the hotspot is completely served by the SCO, the transferred bandwidth reaches its maximum and it holds that $L_{h_{il}} = N_{sc_l} x_{il}^D B_i SE_{sc_l}^D$. Therefore, $x_{il}^D \leq \frac{L_{h_{il}}}{N_{sc_l} B_i SE_{sc_l}^D}$. □

Proposition 3.2 (Co-channel deployment throughput)**.** *Given an MNO_i that transfers $\boldsymbol{x_i^C} B_i$ MHz to the SCO, the total load served in a co-channel deployment is expressed as*

$$\begin{aligned} L_i^{CT}(\boldsymbol{x_i^C}) &= L_i^{Cmc}(X_i^C) + L_i^{Csc}(\boldsymbol{x_i^C}) \\ &= [X_i^C SE_i^C + (1 - X_i^C) SE_i^D] B_i + \sum_{l \in \mathcal{N}_\mathcal{C}} N_{sc_l} x_{il}^C B_i SE_{sc_l}^C \end{aligned} \tag{3.3}$$

where SE_i^C is the average spectral efficiency of MNO_i in co-channel bands, and

$$x_{il}^C \leq \frac{L_{h_{il}}}{N_{sc_l} B_i SE_{sc_l}^C} = x_{il}^{Cmax}, \ \forall l \in \mathcal{N}_\mathcal{C} \tag{3.4}$$

Proof. In a co-channel deployment, the bandwidth transferred from MNO_i to the SCO is simultaneously shared by MNO_i with average spectral efficiency SE_i^C and by the SCO with spectral efficiency SE_{sc}^C. The rest of the bandwidth (i.e. $(1 - X_i^C)B_i$) is exclusively used by MNO_i with spectral efficiency SE_i^D. The proof of limits of x_{il}^C is analogous to the one of Proposition 3.1. □

As already observed, the spectral efficiency determines the capacity of the MNO and the offloading capacity. In turn, it depends a great deal on aspects such as the number of small cells, location or bandwidth allocation, thereby impacting on the MNO's served traffic, by limiting or increasing the achievable maximum throughput. Therefore, the maximum served load is achieved with the co-channel deployment in some scenarios, and with dedicated spectrum deployment in some others. The selection of each deployment depends on the ratio between spectral efficiencies.

Proposition 3.3 (Spectrum deployment selection)**.** *Given an MNO_i with L_i and an SCO, if $\exists \boldsymbol{x_i^C}, \boldsymbol{x_i^D} \in (0, 1]$ such that $L_i^{Csc} = L_i^{Dsc}$, then it holds $L_i^{CT} > L_i^{DT}$ if $(1-\delta_i)X_i^C < X_i^D$, where $\delta_i = \frac{SE_i^C}{SE_i^D}$, with $\delta_i \in (0, 1)$. It holds $L_i^{DT} > L_i^{CT}$ otherwise.*

Proof. Let us define the difference between the load served in a co-channel scenario (L_i^{CT}) and the load served in a dedicated spectrum deployment (L_i^{DT}) as $\Theta_i = L_i^{CT} - L_i^{DT}$.

The co-channel deployment will achieve higher served load if $\Theta_i > 0$. By using (3.1) and (3.3),

$$\Theta_i(\boldsymbol{x_i^C}, \boldsymbol{x_i^D}) = [(X_i^D - (1-\delta_i)X_i^C)SE_i^D + \sum_{l \in \mathcal{N}_C} SE_{sc_l}^D N_{sc_l}(\delta_{sc_l} x_{il}^C - x_{il}^D)]B_i \tag{3.5}$$

where $\delta_{sc_l} = \frac{SE_{sc_l}^C}{SE_{sc_l}^D} \in (0,1)$. Expression (3.5) can be easily written as $\Theta_i = \Theta_i^{mc} + \Theta_i^{sc}$, where $\Theta_i^{mc} = L_i^{Cmc} - L_i^{Dmc}$ and $\Theta_i^{sc} = L_i^{Csc} - L_i^{Dsc}$. Since Proposition 3.3 assumes $\Theta_i^{sc} = 0$, then $\Theta_i > 0$ is equivalent to $\Theta_i^{mc} > 0$. Therefore, $(1-\delta_i)X_i^C < X_i^D$ is obtained by arranging $\Theta_i^{mc} > 0$. □

According to the results in [85], SE_i is expected to be slightly higher for the dedicated spectrum deployment (i.e. $\delta_i < 1$ but $\delta_i \to 1$). Hence, when traffic is offloaded, $(1-\delta_i)X_i^C < X_i^D$ is valid for a wide range of x_{il}^C and x_{il}^D. Results on these dependencies are shown in Section 3.5.

3.3.2 MNO Profit

As mentioned earlier in this section, the MNOs' objectives are twofold, since they embrace both financial and capacity aspects. Specifically, MNO_i aims at maximizing the profit (P_i), defined as the revenue minus the expenses, while guaranteeing the QoS. It should be noted that in order to offload its traffic, each MNO_i partakes in auctions conducted by the SCO for each SC cluster, at the beginning of each timeframe. In the auctions, each MNO_i bids an amount of money $\boldsymbol{b_i} \triangleq (b_{il} : \forall l \in \mathcal{N}_C), b_{il} \in \mathbb{R}_+$ to use the SCO infrastructure. Then, the SCO distributes the SC capacity to the MNOs according to the placed bids. Hence, the profit can be written as

$$P_i(\boldsymbol{x_i}, \boldsymbol{b_i}) = R_i L_i^T(\boldsymbol{x_i}) - CL_i(X_i) - \sum_{l \in \mathcal{N}_C} b_{il} \ [€], \tag{3.6}$$

where R_i is the revenue per throughput unit (€/Mbps), and CL_i is the load cost. CL_i includes variable costs such as energy cost, maintenance cost etc. As shown in (3.6), MNO_i's profit (P_i) is not only tightly coupled with L_i^T but also with the load cost (CL_i). The load cost has been widely modeled in the literature with the use of convex functions [28, 29, 39, 44, 86]. The use of a convex function is suitable for describing the network congestion cost as well as the effect of subscriber churn. Particularly, during peak traffic periods the load cost increases rapidly, depicting the economic consequences

of congestion. Hence, let us assume that CL_i is also convex, and can be expressed as

$$CL_i(X_i) = \frac{a_i(L_i^{mc}(X_i))^2}{C_i + d_i - L_i^{mc}(X_i)}, \quad L_i^{mc} \in [0, C_i], \tag{3.7}$$

where C_i (in Mbps) is the maximum capacity of MNO_i without offloading, factor a_i defines the rate with which the cost increases (in €/Mbps), and d_i (in Mbps) moves the asymptotic discontinuity of the cost function to $L_i = C_i + d_i$. Note that when the MNO_i network operates at its maximum capacity ($C_i = B_i SE_i^D$), the load cost is high but not infinite. Therefore, $d_i > 0$. These parameters are used to adjust the cost function of the economic entity.

3.3.3 SCO Profit

The SCO aims to recoup both its investment and operational cost, and generate profit P_{sc}, by leasing out large volumes of SC capacity at high prices. The profit is defined as the sum of the profits of each SC cluster l, P_{sc_l}, which in turn is given by the difference of the MNOs' bids deducting the load cost, denoted as CL_{sc_l}. Hence, the profit can be written as

$$P_{sc} = \sum_{l \in \mathcal{N}_C} P_{sc_l} = \sum_{l \in \mathcal{N}_C} \left(\sum_{i \in \mathcal{N}} b_{il} - CL_{sc_l} \right) \quad [€]. \tag{3.8}$$

Due to the differences in the structure of the MNO network and the SCO network, as well as their operation and maintenance, each financial entity is characterized by a different OPEX, and therefore a different cost. Despite the difference of the magnitude of MNO and SCO OPEX, these two economic entities offer the same type of service. In this sense, network congestion impacts in a similar manner the SCO's finances. Therefore, the SC cluster l's load cost can be also described as a convex function

$$CL_{sc_l} = \frac{a_{sc_l} L_{sc_l}^2}{C_{BH_l} + d_{sc_l} - L_{sc_l}}, \; L_{sc_l} \in [0, C_{BH_l}]. \tag{3.9}$$

where $L_{sc_l} = \sum_i L_{il}^{sc}$ is the total load served by SC cluster l, and a_{sc_l} and d_{sc_l} are the characteristic parameters of CL_{sc_l}. To ensure its objective, the SCO sets a minimum profit for each SC cluster l, $P_{sc_l}^{min}$ (in €), below which the SCO will not accept the bids and therefore there will not be offloading. The minimum profit is defined as a portion of the SC cluster's cost and determined by a profit factor $z > 1$, as shown in Proposition 3.4.

Proposition 3.4 (Minimmum SCO revenue and profit)**.** *An SCO that receives the bids from a set $\mathcal{N}$ of MNOs and imposes a minimum profit margin over SC cluster l's total cost, $\sum_{i \in \mathcal{N}} b_{il} \geq z(CL_{sc_l})$ with $z > 1$, must receive a minimum revenue for SC cluster l*

equal to

$$\sum_{i \in \mathcal{N}} b_{il} \geq z \frac{a_{sc_l} L_{sc_l}^2}{C_{BH_l} + d_{sc_l} - L_{sc_l}}. \tag{3.10}$$

or alternatively, a minimum profit expressed as $P_{sc_l}^{min} = (z-1)CL_{sc_l}$.

Proof. $P_{sc_l}^{min}$ is found by substituting (3.9) and (3.10) in $P_{sc_l} = \sum_{i \in \mathcal{N}} b_{il} - CL_{sc_l}$. □

Hence, the SCO determines the SC clusters' pricing and the corresponding minimum profit by adjusting the value of the profit factor z.

3.4 The Auction

The distribution of the existing resources, i.e. the available spectrum and the cost of using the SCO infrastructure, are the response to two main objectives: firstly, the QoS guaranty (the capacity objective), and secondly high profit generation (the economic objectives of both the MSPs and the SCO). In this scenario, the auction is the mechanism by which the interaction of these competing objectives results in incentives for all the involved parties.

3.4.1 The Auction Mechanism

An auction is a trading process for goods or services, where potential buyers bid money in order to decide the buyer(s), the price of the good, and its distribution. In any auction there are three aspects that must be defined: the auctioneer (i.e. the auction-conducting party), the bidders (i.e. the buying parties) and the good to be auctioned. In the scenario under study, the SC clusters' capacity is the auctioned good, the SCO plays the role of the auctioneer, and the MNOs are the bidders. At the beginning of each timeframe t, the SCO conducts an individual auction for each SC cluster $l \in \mathcal{N}_\mathcal{C}$, during which each MNO_i places a bid b_{il}. Given the set of bids for an SC cluster l, the SCO distributes L_{sc_l} in a proportionally fair manner [87], and informs the MNOs about the actual volume of leased capacity $L_{il}^{sc}, i \in \mathcal{N}$. The proportional fair allocation rule charges all bidders with the same price per SC capacity ratio, and hence there is no possibility for overcharging. Hence, the MSPs have a fair chance to serve more traffic (i.e. network incentive) and gain profit (i.e. economic incentive). Furthermore, we assume that the auction is a sealed-bid auction. This type of auctions are characterized by the simultaneous submission of bids, so that none of the bidders know the bids of the other bidders.

Definition 3.5 (Proportional fair capacity allocation)**.** If a given set of MNOs, $\mathcal{N}$, place a set of bids, b_{jl} with $j \in \mathcal{N}, l \in \mathcal{N_C}$, for the total load served by SC cluster l (L_{sc_l}), the maximum load that a specific MNO_i, with $i \in \mathcal{N}$, can offload to the SC cluster is given by

$$L_{il}^{sc} = \frac{b_{il}}{b_{il} + \sum_{j \in \mathcal{N} \setminus \{i\}} b_{jl}} L_{sc_l}, \tag{3.11}$$

where L_{sc_l} is found by solving $\sum_{i \in \mathcal{N}} b_{il} = zCL_{sc_l}(L_{sc_l})$.

3.4.2 Conducting the auction

3.4.2.1 Social Welfare

In this section we address the traffic offloading problem as a Social Welfare Maximization (SWM) problem. The SW is defined as the aggregate payoff of all the MNOs and the SCO. The maximization takes into account the availability of MNOs' resources, as well as technical constraints imposed by the SC cluster deployment. Hence, by replacing (3.1) and (3.7) in (3.6), and (3.9) in (3.8) the SWM problem for the dedicated spectrum deployment[4] is formulated as

$$\begin{aligned}
\max_{\boldsymbol{x}} \quad & P(\boldsymbol{x}) = \sum_{i \in \mathcal{N}} P_i(\boldsymbol{x_i}) + P_{sc} \qquad (3.12)\\
& = \sum_{i \in \mathcal{N}} [R_i \left((1 - X_i^D) B_i SE_i^D + \sum_{l \in \mathcal{N_C}} N_{sc_l} x_{il}^D B_i SE_{sc_l}^D \right) \\
& - \frac{a_i [(1 - X_i^D) B_i SE_i^D]^2}{C_i + d_i - (1 - X_i^D) B_i SE_i^D}] - \sum_{l \in \mathcal{N_C}} \left(\frac{a_{sc_l} L_{sc_l}^2}{C_{BH_l} + d_{sc_l} - L_{sc_l}} \right) \\
\text{s.t.} \quad & 0 \leq x_{il} \leq x_{il}^{max} \leq 1, \ \ i \in \mathcal{N}, l \in \mathcal{N_C}, \qquad (3.12a)\\
& B_{sc} \geq \sum_{i \in \mathcal{N}} x_{il} B_i, \ i \in \mathcal{N}, l \in \mathcal{N_C}, \qquad (3.12b)\\
& C_{BH_l} \geq \sum_{i \in \mathcal{N}} L_{il}^{sc}, \ i \in \mathcal{N}, l \in \mathcal{N_C}. \qquad (3.12c)
\end{aligned}$$

The optimization problem in (3.12) maximizes the aggregate profit of all the stakeholders constrained by the availability of resources to be transferred from $MNO_{i \in \mathcal{N}}$ to the SCO (3.12a), the technological limits of the small cells (3.12b), and the backhaul capacity (3.12c). Constraints (3.12b) and (3.12c) limit the capacity of the SCO infrastructure, and consequently results in a competition among MNOs. Likewise, it is worth noting that bids are cancelled out in (3.12), since $\boldsymbol{b_{i \in \mathcal{N}}}$ is a cost for MNO_i but in turn it is a revenue for the SCO. This is the reason why bids are not explicitly included in (3.12),

[4]The SWM problem formulation for the co-channel deployment is derived in the same manner by replacing (3.3) in (3.6).

although they are implicitly included in P_{sc}, since there is a minimum SCO profit (or a minimum SCO revenue) to create an incentive for offloading according to Proposition 3.4. Moreover, the bids and the auction are connected with the SWM problem through the allocation of the SC capacity L_{sc_l} (as described in **Definition** 3.5). L_{sc_l} and the bids determine the SC capacity allocation, and in turn the MNO throughput. That is, the solution of (3.12) determines the bids and the *transferred bandwidth percentage* matrix $\boldsymbol{x}$. However, as it can be observed in (3.12) this solution depends on both network (i.e. $SE_i, B_i, SE_{sc_l}, N_{sc_l}, C_{BH_l}$) and economic (i.e. R_i, a_i, d_i, z etc.) parameters of the stakeholders.

Next, we define the *reserve price* by combining (3.9), (3.10) and (3.11) in (3.8).

Definition 3.6 (Reserve price)**.** The reserve price is the minimum price b_{il} that MNO_i has to pay in order to lease L_{il}^{sc} SC capacity of cluster l, and is written as

$$b_{il}^{min} = z\frac{a_{sc_l} L_{il}^{sc} L_{sc_l}}{C_{BH_l} + d_{sc_l} - L_{sc_l}}. \tag{3.13}$$

All MNOs are interested in bidding the reserve price to maximize their profit, which can be calculated with (3.13) if all information is available to all MNOs. That is, if we consider that the MNOs have complete information over their own future load, the auction, the SCO and the competing MNOs' parameters, then they will be able to place bids equal to the reserve price, thereby avoiding increasing their bids and ending up with the same or less or even more but unnecessary leased SC capacity.

3.4.2.2 Learning mechanism

In realistic competitive SCaaS scenarios, MNOs cannot have the information to bid the reserve prices ($b_{il}, l \in \mathcal{N_C}$). Since the auctions are conducted at the beginning of the timeframe, MNO_i cannot know its own load. In addition, the bids and the load of the competing MNOs are not public (i.e., b_j and L_j, for $j \in \mathcal{N}\backslash\{i\}$, are unknown to MNO_i). Given that, MNO_i must estimate the load in the hotspots ($L_{h_{il}}$), the load outside the hotspots (L_{n_i}) and the aggregate bids of the opponents ($\sum_{j\in\mathcal{N}\backslash\{i\}} b_{jl}, l \in \mathcal{N_C}$, henceforth denoted by $b_{jl\neq i}$ for simplicity) to properly select each b_{il}. It should be noted that each MNO needs to estimate the aggregate bid and not each competing MNO's bid separately, thanks to the definition of the proportional fair capacity allocation mechanism in (3.11). Hence, the more accurate the estimations are, the better the auction results are. In the following, the estimates for MNO_i's offered load and the opponents' bids are denoted as $\overline{L_i}$ and $\overline{b_{jl\neq i}}$, respectively. Analogously to the SWM problem stated in (3.12), the maximization of the profit remains as the objective. However, in a competitive environment

with limited available information, each MNO_i formulates the problem as

$$\max_{\boldsymbol{x_i}} \quad P_i(\boldsymbol{x_i}|\overline{b_{jl\neq i}}) = R_i L_i^T(\boldsymbol{x_i}) - CL_i(X_i) - \sum_{l\in\mathcal{N}_\mathcal{C}} b_{il}(x_{il}|\overline{b_{jl\neq i}}) \tag{3.14}$$

$$\text{s.t.} \quad 0 \leq x_{il} \leq x_{il}^{max} \leq 1,\ l \in \mathcal{N}_\mathcal{C}, \tag{3.14a}$$

$$B_{sc_l} \geq \sum_{i\in\mathcal{N}} x_{il} B_i,\ l \in \mathcal{N}_\mathcal{C}, \tag{3.14b}$$

$$C_{BH_l} \geq \sum_{i\in\mathcal{N}} L_{il}^{sc},\ l \in \mathcal{N}_\mathcal{C}. \tag{3.14c}$$

where $b_{il}(x_{il}|\overline{b_{jl\neq i}})$ denotes the bid selection of MNO_i after estimating $b_{jl\neq i}$. The solution to (3.14) is also the reserve price, though important parameters such as L_{sc} have to be derived previously.

Proposition 3.7 (Minimum bid selection)**.** *Given an MNO_i with a targeted offloaded load $L_{il}^{sc}, l \in \mathcal{N}_\mathcal{C}$, and assuming that both the SCO capacity allocation policy (3.11) and the structure of the SCO cost and profit functions (3.10) or (3.13) are known, the minimum winning bid for SC cluster l is achieved by solving $\lambda_l b_{il}^3 + \phi_l b_{il}^2 + \xi_l b_{il} + \omega_l = 0$, where $\lambda_l = L_{il}^{sc} - C_{BH_l} + d_{sc_l}$; $\xi_l = b_{jl\neq i}^2 L_{il}^{sc}$; $\omega_l = z a_{sc_l} (L_{il}^{sc})^2 b_{jl\neq i}^2$; $\phi_l = (z a_{sc_l} L_{il}^{sc} + 2b_{jl\neq i}) L_{il}^{sc} - (C_{BH_l} + d_{sc_l}) b_{jl\neq i}$.*

Proof. The expression is obtained by combining (3.11) and (3.13) and rearranging the result. □

The solution to Proposition 3.7 provides the minimum bids b_{il} for all SC clusters to offload L_{il}^{sc} (i.e. b_{il}^{min}). Since estimates are used instead of actual values, the solution stated in Proposition 3.7 is $b_{il}(x_{il}|\overline{b_{jl\neq i}}), l \in \mathcal{N}_\mathcal{C}$. For accurate estimations, $b_{il}(x_{il}|\overline{b_{jl\neq i}}) \approx b_{il}^{min}$.

The learning mechanism described in the sequel is designed to accurately estimate the load and $b_{jl\neq i}$ at an SC cluster l. The bid selection is run at every timeframe by each MNO. Our proposed learning mechanism consists of three components: a forecasting method to predict the offered load (L_i and L_{h_i}), and two methods to estimate $b_{jl\neq i}$ (a reinforcement learning algorithm along with an adaptive search range scheme).

Traffic Forecasting

In order to forecast the MNOs' traffic loads L_i and L_{h_i}, the well-known and computationally efficient Holt-Winters (HW) method [88] is used. Also known as Triple Exponential Smoothing, it takes into account the level, trend and seasonal changes in the observed dataset. There are two HW models according to the type of the seasonality (described

as the periodic repetition in time-series data), known as multiplicative and additive seasonal models. The former refers to a proportional change in the values of the time series from season to season, whereas the latter refers to a particular absolute change. In our case, the multiplicative model is used to capture the random, small, seasonal variations of the traffic pattern.

Definition 3.8 (Holt-Winters method [88])**.** At timeframe $t \in \mathcal{T}$, let us define the forecast offered load for timeframe $(t+m) \in \mathcal{T}$, with $m \in \mathbb{N}_{>0}$, as $\overline{L}_{t+m}$. According to HW method, $\overline{L}_{t+m}$ is calculated as $\overline{L}_{t+m} = (S_t + mv_t)I_{t-T+m}$, where S_t is a (smoothed) estimation of the level (i.e. a local average of the dataset), v_t is an estimation of the linear trend (slope) of the time series, I_t denotes the seasonal component (i.e. the expectation for a specific timeframe based on its past season values), and $T = |\mathcal{T}|$ is the number of daily timeframes (season length).

$$\begin{aligned} S_t &= \alpha\frac{L_t}{I_{t-T}} + (1-\alpha)(S_{t-1} + v_{t-1}) \\ v_t &= \beta(S_t - S_{t-1}) + (1-\beta)v_{t-1} \\ I_t &= \gamma\frac{L_t}{S_t} + (1-\gamma)I_{t-T} \end{aligned}$$

with three smoothing factors $\alpha, \beta, \gamma \in [0,1]$ that show the dependence on the past values of the time series. Initialization values are $S_t = \sum_{t=1}^{T} \frac{L_t}{T}$, $v_t = 0$, $I_t = \frac{L_t}{S_t}$, $t = 1...T$.

As the auction is carried out at every timeframe, $m = 1$.

Opponents' bids estimation

Let us define a continuous set containing the actual value of the aggregate bids of the contending MNOs for SC cluster l, $\mathcal{A}_{b_{jl\neq i}} = [b^{min}_{jl\neq i}, b^{max}_{jl\neq i}]$, with $b^{min}_{jl\neq i} \leq b_{jl\neq i} \leq b^{max}_{jl\neq i}$. The learning algorithm devised to estimate the opponents' bids must be able to find out $b_{jl\neq i}$ in $\mathcal{A}_{b_{jl\neq i}}$. Nevertheless, using a continuous set is shown inefficient in these scenarios, since the convergence requires thousands of iterations [89]. Let us then define the uniform discretization of $\mathcal{A}_{b_{jl\neq i}}$ as the discrete set of K elements $\mathcal{A}'_{b_{jl\neq i}} = \{\frac{1}{K-1}(K-k)b^{min}_{jl\neq i} + (k-1)b^{max}_{jl\neq i}\}_{k=1...K}$, and the probability of each element of $\mathcal{A}'_{b_{jl\neq i}}$ at timeframe t of day n as $p_{kl}(t,n)$, for $k = 1 \dots K$. Note that the opponents' bids are linked with their capacity needs, which in turn are related to the previous timeframes' needs, and follow a daily pattern. Therefore, $p_{kl}(t,n)$ is correlated both with $p_{kl}(t-1,n)$ and $p_{kl}(t,n-1)$. In the following, a modification of the reinforcement learning algorithm Exp3 (described in [90]) is presented in Algorithm 1. The application of Exp3 and the modification shown hereafter to the discrete set $\mathcal{A}'_{b_{jl\neq i}}$ turns the selection of each $\overline{b_{jl\neq i}}$ into a non-stochastic multi-armed bandit problem, where each $\overline{b_{jl\neq i}} \in \mathcal{A}'_{b_{jl\neq i}}$ is regarded as one arm.

Algorithm 1 is designed to continuously update $p_{kl}(t,n)$, which is initialized for each SC cluster l at day n_0 as $p_{kl}(t,n_0) = \frac{1}{K}$ for $k = 1 \dots K$ and $t = 1 \dots T$. Later, for every $t \in \mathcal{T}$ and $n > n_0$, $p_{kl}(t,n)$ is updated based on the accuracy of the estimation. Although Algorithm 1 is aimed to estimate $b_{jl\neq i}$, the actual value of $b_{jl\neq i}$ is not available to MNO_i even after the auction. Hence, the accuracy of $\overline{b_{jl\neq i}}$ must be controlled indirectly through an alternative variable, such as $\overline{L_{il}^{sc}}$. We define the relative difference of estimated and actual L_{il}^{sc} at timeframe t of day n as

$$\Delta L_{il}^{sc}(t,n) = \frac{\overline{L_{il}^{sc}}(t,n) - L_{il}^{sc}(t,n)}{L_{il}^{sc}(t,n)} \tag{3.15}$$

where $\overline{L_{il}^{sc}}(t,n)$ and $L_{il}^{sc}(t,n)$ are the estimated and the actual offloaded load at timeframe t of day n, respectively. If $\overline{b_{jl\neq i}}$ is accurate, then $\Delta L_{il}^{sc}(t,n) \to 0$; otherwise, $\Delta L_{il}^{sc}(t,n) \nrightarrow 0$. Based on the accuracy of the estimation, $p_{kl}(t,n)$ will be updated through a reward defined as follows: for a given estimate $\overline{b_{jl\neq i}} = \frac{1}{K-1}(K-k)b_{jl\neq i}^{min} + (k-1)b_{jl\neq i}^{max}$ (i.e. equal to the kth element of $\mathcal{A}'_{b_{jl\neq i}}$) the reward for an accurate estimation at timeframe t of day n, referred to as $r_{kl}(t,n) \in [0,1]$, must be close to 1. Conversely, if $\overline{b_{jl\neq i}}$ is not accurate, then $r_{kl}(t,n)$ should be smaller than 1. In the same vein, if $\Delta L_{il}^{sc}(t,n)$ is similar to $\Delta L_{il}^{sc}(t,n-\nu)$, with $\nu \in \mathbb{N}_{>0}$, the reward should be high since load and bids present a strong daily pattern. Thus, we define the reward as

$$r_{kl}(t,n) = \begin{cases} \left(1 - \left|\frac{\Delta L_{il}^{sc}}{\mu_{\Delta L_{il}^{sc}}}\right|\right)_+ & \text{if } \mu_{\Delta L_{il}^{sc}} < \mu_{HW} \text{ and } \sigma_{\Delta L_{il}^{sc}} < \sigma_{HW} \\ \left(1 - 2\left|\Delta L_{il}^{sc}\right|\right)_+ & \text{otherwise} \end{cases} \tag{3.16}$$

where $\mu_{\Delta L_{il}^{sc}}$ and $\sigma_{\Delta L_{il}^{sc}}$ are the mean and the standard deviation of ΔL_{il}^{sc}, μ_{HW} and σ_{HW} are the respective thresholds, and $(x)_+ = \max(0,x)$. Accordingly, given an estimate $\overline{b_{jl\neq i}}$ equal to the k^*th element of $\mathcal{A}'_{b_{jl\neq i}}$ at timeframe t of day n, the updated set of probabilities is given by

$$p_{kl}(t,n+1) = \begin{cases} \frac{p_{kl}(t,n)\exp\left(\eta_l \frac{r_{kl}(t,n)}{K p_{kl}(t,n)}\right)}{p_{kl}(t,n)\exp\left(\eta_l \frac{r_{kl}(t,n)}{K p_{kl}(t,n)}\right) + \sum_{wl\neq kl} p_{wl}(t,n)} & \text{if } k = k^* \\ \frac{p_{kl}(t,n)}{p_{k^*l}(t,n)\exp\left(\eta_l \frac{r_{k^*l}(t,n)}{K p_{k^*l}(t,n)}\right) + \sum_{wl\neq k^*l} p_{wl}(t,n)} & \text{otherwise} \end{cases} \tag{3.17}$$

where $\eta_l \in (0,1]$ is the learning speed. Note that each timeframe is evaluated once per day. That is, one day corresponds to one iteration of the learning mechanism.

Search Range Scheme

In order for Algorithm 1 to work, $\mathcal{A}'_{b_{jl\neq i}}$ must contain $b_{jl\neq i}$. Hence, if the actual aggregate bid does not lie in the initial estimate of the action set, the algorithm will not be able

Algorithm 1: Estimation of opponent's auction strategy at SC cluster $l \in \mathcal{N}_C$

For a given timeframe $t \in \mathcal{T}$
Initialization:

if $(n = n_0)$ **then**
  $p_{kl}(t,n) = \frac{1}{K}$, for $k = 1 \dots K$
  Determine $\mathcal{A}_{b_{jl\neq i}}$ and $\mathcal{A}'_{b_{jl\neq i}}$
end if
Estimation:
Forecast L_{il}^{sc} according to the traffic forecasting algorithm (Section 3.4.2.2)
Select randomly k^* based on $p_{kl}(t,n)$, for $k = 1 \dots K$
$\overline{b_{jl\neq i}} = \frac{1}{K-1}(K-k^*)b_{jl\neq i}^{min} + (k^*-1)b_{jl\neq i}^{max}$
Auction:
Run the auction
Update of probabilities:
Calculate ΔL_{il}^{sc} and $r_{kl}(t,n)$, for $k = 1 \dots K$, according to (3.15) and (3.16)
Calculate $p_{kl}(t,n+1)$ based on (3.17)

to predict it. Then again, even if the initial estimate is correct, the algorithm will not be able to follow any changes of $b_{jl\neq i}$ outside the predefined range. Therefore, an extension of the variable parameter space scheme described in [89] is used. The aim of this scheme is to center the probability distribution obtained by Algorithm 1 by expanding and reducing $\mathcal{A}_{b_{jl\neq i}}$ (and consequently, the associated discretized $\mathcal{A}'_{b_{jl\neq i}}$). The distribution is centered in order to assure that $b_{jl\neq i}$ falls within $\mathcal{A}'_{b_{jl\neq i}}$, and to improve the mechanism's convergence speed. There are two pairs of important parameters in the search range scheme, which are the expansion (ϵ_{r_l}) and the reduction (ρ_{r_l}) rates, defined as the speed with which the range is expanded/reduced, and the expansion (ϵ_{cl}) and reduction (ρ_{c_l}) coefficients. A low expansion/reduction rate/coefficient (i.e. low ϵ_{r_l}, ρ_{r_l}, ϵ_{c_l} and ρ_{c_l} values) could enable the search range scheme to follow rapid $b_{jl\neq i}$ variations. Conversely, big expansion/reduction rates/coefficients (i.e. high ϵ_{r_l}, ρ_{r_l}, ϵ_{c_l} and ρ_{c_l} values) might impede the convergence of the algorithm. Let us first define the conditions used by the search range scheme:

$[C_1]$: $b_{jl\neq i}^{25} < b_{jl\neq i}^{25-\epsilon_{c_l}}$ and $b_{jl\neq i}^{75} > b_{jl\neq i}^{75+\epsilon_{c_l}}$

$[C_2]$: $b_{jl\neq i}^{25} > (b_{jl\neq i}^{max} + b_{jl\neq i}^{min})/2$

$[C_3]$: $b_{jl\neq i}^{25} = b_{jl\neq i}^{75}$

$[C_4]$: $p_{kl}(t,n) = 1/K$ for $k = 1 \dots K$

$[C_5]$: $\mu_{\Delta L_{il}^{sc}} > \mu_{SR_l}$ and $\Delta L_{il}^{sc}(t,n) < \Delta L_{SR_l}$

$[C_6]$: $b_{jl\neq i}^{25} > b_{jl\neq i}^{25+\rho_{c_l}}$ and $b_{jl\neq i}^{75} < b_{jl\neq i}^{75-\rho_{c_l}}$

where μ_{SR_l} and ΔL_{SR_l} are minimum and maximum thresholds, and $b^p_{jl\neq i}$ is the pth percentile of $b_{jl\neq i}$. Initially, the algorithm presented in [89] only defines conditions C_1 and C_6. However, conditions C_2-C_5 have been proposed to cope with the described scenario. Condition C_2 is used when the distribution's mode is at $b^{max}_{jl\neq i}$. The same applies to C_3, but for the lower endpoint $b^{min}_{jl\neq i}$. Condition C_4 is used when the algorithm is in a stalemate due to the actual bid being far from $A_{b_{jl\neq i}}$, so that $r_{kl}(t,n)=0$. Finally, C_5 is examined only when none of the rest is satisfied. It is applied when the mechanism is in a stalemate and does not place $A_{b_{jl\neq i}}$ close to the real bid. Based on conditions C_1-C_6, the action set $\mathcal{A}_{b_{jl\neq i}}$ is updated as,

$$b^{min}_{jl\neq i} = \begin{cases} b^{min}_{jl\neq i} + (-1)^n \chi_n (\psi^{min}_n - b^{min}_{jl\neq i}) & \text{if } C_n \text{ with } n=\{1,2,3,4,6\} \\ \psi^{min}_n \overline{b_{jl\neq i}} & \text{if } C_n \text{ with } n=5 \\ b^{min}_{jl\neq i} & \text{otherwise} \end{cases} \tag{3.18}$$

$$b^{max}_{jl\neq i} = \begin{cases} b^{max}_{jl\neq i} + \chi_n (b^{max}_{jl\neq i} - \psi^{max}_n) & \text{if } C_n \text{ with } n=\{1,2,4\} \\ b^{max}_{jl\neq i} - \chi_n (b^{max}_{jl\neq i} - \psi^{max}_n) & \text{if } C_n \text{ with } n=\{3,6\} \\ \psi^{max}_n \overline{b_{jl\neq i}} & \text{if } C_n \text{ with } n=5 \\ b^{max}_{jl\neq i} & \text{otherwise} \end{cases} \tag{3.19}$$

where $\chi_n = \epsilon_{r_l}$ for $n=[1,4]$; $\chi_6 = \rho_{r_l}$; $\psi^{min}_n = b^{25}_{jl\neq i}$ for $n=\{1,2,6\}$; $\psi^{min}_n = b^{max}_{jl\neq i}$ for $n=\{3,4\}$; $\psi^{min}_5 = 0.9$; $\psi^{max}_n = b^{75}_{jl\neq i}$ for $n=\{1,2,6\}$; $\psi^{max}_n = b^{min}_{jl\neq i}$ for $n=\{3,4\}$; $\psi^{max}_5 = 1.1$. Given that probabilities $p_{kl}(t,n)$ are only calculated for the discrete set $\mathcal{A}'_{b_{jl\neq i}}$, these probabilities are updated through interpolation/extrapolation after the reduction/expansion of $\mathcal{A}'_{b_{jl\neq i}}$. The proposed algorithms, both the opponents' bids estimation and range search, may adapt three main parameters: the learning speed, the expansion rate and the reduction rate. In particular, if $\mu_{\Delta L^{sc}_{il}}$ is high, the algorithm should be able to accurately estimate $b_{jl\neq i}$ faster. Accordingly, two additional conditions are defined:

[C_7]: $\mu_{\Delta L^{sc}_{il}} > \mu_{H\&S_1}$

[C_8]: $\mu_{\Delta L^{sc}_{il}} < \mu_{H\&S_2}$ and $\sigma_{\Delta L^{sc}_{il}} < \sigma_{H\&S}$

where $\mu_{H\&S_1}$, $\mu_{H\&S_2}$ and $\sigma_{H\&S}$ are mean and deviation thresholds. An update mechanism is proposed based on C_7 and C_8 as $\eta_l = \eta_l + 0.1$, $\epsilon_{r_l} = 1.2\epsilon_{r_l}$, $\rho_{r_l} = 1.2\rho_{r_l}$ for C_7, and $\eta_l = \eta_l - 0.1$, $\epsilon_{r_l} = 0.5\epsilon_{r_l}$, $\rho_{r_l} = 0.9\rho_{r_l}$ for C_8. The specific values of the update mechanism have been obtained through simulations. The complexity of the learning algorithm is the number of elements in $A'_{b_{jl\neq i}}$ (i.e. K) multiplied by the number of timeframes T. Hence, the computational complexity of the mechanism is $O(d_{cnv}TK)$, where d_{cnv} denotes the number of days until convergence.

TABLE 3.2: MNO Network-Financial Parameters

MNO	
Bandwidth (B_i)	20 MHz
Sector Spect. Eff. (SE_i^D,SE_i^C)	see Table 3.4
Revenue per 1-h timeframe (R_i)	3.375 €/Mbps
Cost shaping factor (a_i)	0.189 €/Mbps [91]
Cost shaping constant (d_i)	10 Mbps [91]
Hotspot offered load (L_{h_i})	0.6 L_i

TABLE 3.3: SCO Network-Financial Parameters

SCO	
Number of SC clusters (N_C)	2
Number of SCs per cluster (N_{sc_l})	{ (4,4),(4,5),(5,5),(5,6),(6,6) }
Max. Bandwidth (B_{sc})	20 MHz
Spectral Efficiency (SE_{sc}^D,SE_{sc}^C)	see Table 3.4
Backhaul Capacity (C_{BH})	[100,1000] Mbps
Cost shaping factor (a_{sc})	0.072 €/Mbps [91]
Cost shaping constant (d_{sc})	8.9 Mbps [91]
Profit factor (z)	1.2

TABLE 3.4: Average spectral efficiency for different SC deployments

Spectral Efficiency (Mbps/MHz)				
N_{sc}	Dedicated channel		Co-channel	
	Macrocell	SC clusters	Macrocell	SC clusters
(4,4)	3.0703	(4.8946,4.8946)	2.8041	(2.0601,2.0601)
(4,5)	3.0703	(4.8616,4.1812)	2.7869	(2.0381,1.9889)
(5,5)	3.0703	(4.181,4.181)	2.7668	(1.9868,1.9868)
(5,6)	3.0703	(4.1676,4.1449)	2.7435	(1.9843,1.9736)
(6,6)	3.0703	(4.1327,4.1327)	2.7121	(1.972,1.972)

3.5 Performance Evaluation

3.5.1 Scenario description and parameters

The scenario used for the numerical analysis in this section consists of $N = 2$ MNOs and an SCO with two SC clusters of co-channel small cells in the MNOs' overlapping macrocell coverage areas. In order to increase their profits and throughput, the MNOs participate in the auctions conducted by the SCO, and decide on the bid prices according to the bandwidth requirements (i.e. $\boldsymbol{x_i}$), and the auction competition. According to its business strategy, the SCO selects offline the value of z, which remains static for periods significantly longer than the auction timeframes. The values of the main parameters used in our simulations are obtained from Small Cell Forum's reports [84] and listed in Tables 3.2 and 3.3. In the baseline scenario, 60% of the offered load is generated

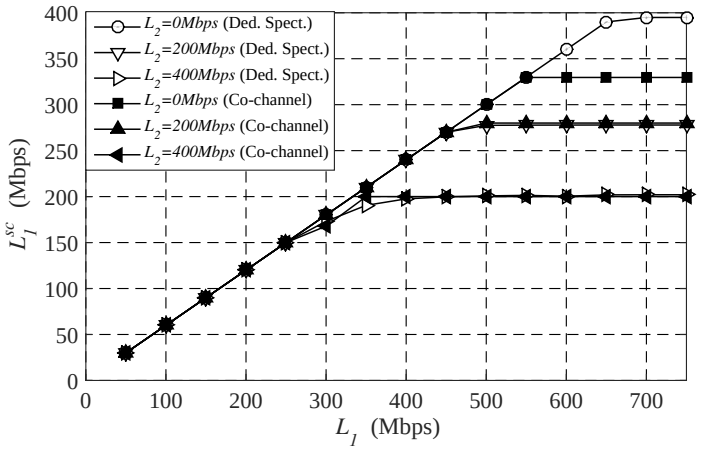
in the hotspots, and the remaining 40% outside the hotspots (i.e. $L_{h_i} = 0.6L_i$ and $L_{n_i} = 0.4L_i$). In general, an SC cluster is comprised of a set of street-level small cells connected to a Network Termination Point (NTP) through a multi-hop millimetre-wave wireless network [84]. A custom-made MATLAB®-based simulator has been developed to calculate the spectral efficiency values shown in Table 3.4. The simulator is compliant with the guidelines provided by 3GPP in [92–94].

3.5.2 Network Throughput

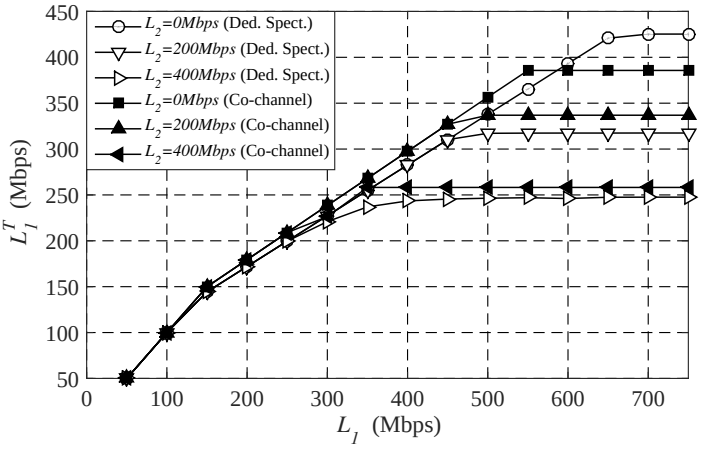
The impact of the competition among MNOs and the differences arisen with dedicated spectrum and co-channel deployments are analysed in the described scenario with two clusters of $N_{sc_1} = N_{sc_2} = 4$ small cells and a backhaul capacity $C_{BH_1} = C_{BH_2} = 200$ Mbps. Moreover, in order to shed light on the impact of spatial traffic variations, we consider two hotspot traffic allocations: in the first allocation, $L_{h_{il}} = 0.3 \cdot L_i, \forall i \in \mathcal{N}, \forall l \in \mathcal{N_C}$ (Fig. 3.2), and in the second one $(L_{h_{11}}, L_{h_{12}}) = (0.4, 0.2) \cdot L_1$, $(L_{h_{21}}, L_{h_{22}}) = (0.2, 0.4) \cdot L_2$ (Fig. 3.3). In this scenario, we assume that the MNOs have no a priori knowledge on their load nor their opponents' bids. Hence, we conduct simulations with the use of the proposed learning mechanism to acquire the following results.

Fig. 3.2a and Fig. 3.2b depict the offloaded traffic (L_1^{sc}) and the total served load of MNO_1 (L_1^T) as a function of the MNO_1 offered load (L_1) for different MNO_2 offered load levels (L_2), and for load distribution $L_{h_{il}} = 0.3 \cdot L_i, \forall i \in \mathcal{N}, l \in \mathcal{N_C}$. The same experiments have been carried out for the dedicated spectrum and co-channel deployments. As both MNOs have the same cost functions (i.e. the values of all parameters in (3.7) are equal), results for MNO_1 and MNO_2 are symmetrical, and only MNO_1 figures are included. Furthermore, since both MNOs' hotspot traffic is equally divided between the SC clusters (i.e. $L_{h_{il}} = 0.3 \cdot L_i, \forall i \in \mathcal{N}, l \in \mathcal{N_C}$), during the auctions none of the clusters is prioritized and therefore $L_{i1}^{sc} = L_{i2}^{sc}, \forall i \in \mathcal{N}$.

It can be observed in Fig. 3.2a that L_{h_1} is completely served by the SCO until it reaches a maximum level, after which the SCO is unable to serve more traffic. This maximum L_1^{sc} differs depending on the competition among MNOs (or, in other words, the offered load of MNO_2) and the spectrum deployment use case. Focusing on the results of MNO_1 in a scenario without competition (i.e. $L_2 = 0Mbps$), Fig. 3.2a shows that the maximum offloaded traffic (L_1^{sc}) is higher for the dedicated spectrum option than for the co-channel use case. This is explained by the access link capacity of the SCO in each case. In that sense, if no cost or backhaul capacity restrictions were considered, the maximum offloaded traffic could be found from its definition as $L_i^{sc} = \sum_{l \in \mathcal{N_C}} N_{sc_l} x_{il} B_i SE_{sc_l}$. Based on this assumption, the maximum offloaded traffic is calculated in the hypothetical case



(a)



(b)

FIGURE 3.2: Traffic served by (a) the SC clusters and (b) MNO_1's throughput versus the total offered load for MNO_1, with $L_{h_{il}} = 0.3 \cdot L_i, \forall i \in \mathcal{N}, l \in \mathcal{N}_\mathcal{C}$

where all bandwidth is transferred from MNO_1 to the SCO (i.e. $X_i = 1$). Specifically, if $X_i = 1$, the maximum offloaded traffic is $L_1^{Dsc} = 783Mbps$ and $L_1^{Csc} = 329Mbps$. Accordingly, Fig. 3.2a shows that, for $L_2 = 0Mbps$, L_1^{sc} is limited by the backhaul capacity ($\sum_{l\in\mathcal{N}_\mathcal{C}} C_{BH_l} = 400Mbps$) in the dedicated spectrum case, and by the access link capacity (namely the spectral efficiency of the small cells) in the co-channel case. As expected, L_1^{sc} falls as L_2 increases, since the backhaul capacity is shared out among the two MNOs.

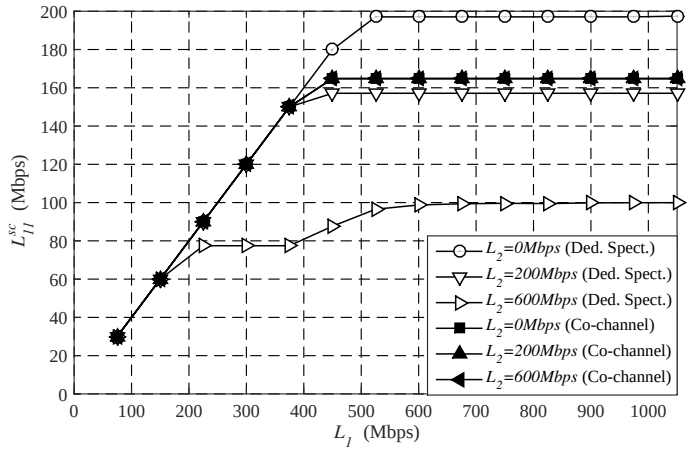
As for $L_2 = 400Mbps$ in Fig. 3.2a, it can be seen that $L_1^{sc} \cong \sum_{l\in\mathcal{N}_\mathcal{C}} C_{BH_l}/2$ for $L_1 \geq 350Mbps$. The same occurs to MNO_2 due to symmetry, and so an important outcome can be stated. In general, for a set of N MNOs with equal cost functions, and $L_{h_{il}} \in (C_{BH_l}/N, N_{sc_l} SE_{sc_l} B_i]$, $\forall i \in \mathcal{N}, l \in \mathcal{N}_\mathcal{C}$ (L_{il}^{sc} is not limited by the access link capacity, but by C_{BH_l}), the SC capacity is equally divided among the set of MNOs, $L_i^{sc} \cong \sum_{l\in\mathcal{N}_\mathcal{C}} C_{BH_l}/N$, whether or not $L_{h_{il}}$ and $L_{h_{jl}}$ are equal $\forall i, j \in \mathcal{N}, l \in \mathcal{N}_\mathcal{C}$ and

$i \neq j$. In other words, under high competition scenarios, C_{BH_l} is equally shared among MNOs. This is due to the fact that the cost incurred by an MNO to offload a unit of load increases with the competition, but it is upper bounded. In particular, the bid will never be higher than the value that makes the cost of offloading be above the cost of not offloading. If all MNOs have the same cost functions, they all bid the same amount, when L_{h_i} is high, and therefore they offload the same amount of traffic (according to Definition 3.5).

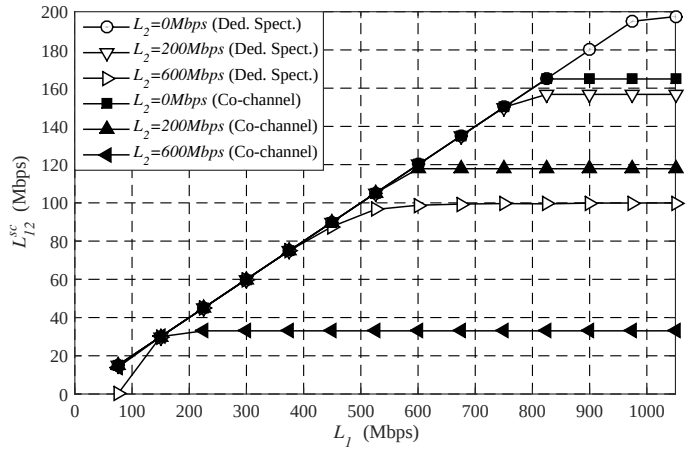
Fig. 3.2b depicts MNO_1's throughput (L_1^T). Although the trend of L_1^T is similar to the trend of L_1^{sc}, we point out two important results. First, the MNO always prioritizes the traffic in the hotspot since it is more profitable. This can be observed in the decrease of the gradient of L_1^T for an offered load L_1 higher than 150 Mbps (see Fig. 3.2b). Second, Fig. 3.2b shows that $L_1^{CT} > L_1^{DT}$ as long as the hotspot traffic is completely served (this fact is proved in Proposition 3.3). Conversely, when the hotspot load is not completely served, the spectrum deployment that provides the best results in terms of total served load will depend on the spectral efficiency in each case.

As in Fig. 3.2, Fig 3.3a, 3.3b, and 3.3c depict MNO_1's offloaded traffic at the SC clusters (L_{11}^{sc} and L_{12}^{sc}), and the total served load (L_1^T), respectively, as a function of the offered load (L_1). The only difference in this use case is the uneven hotspot traffic allocation, that is, $(L_{h_{11}}, L_{h_{12}}) = (0.4, 0.2)\cdot L_1$ and $(L_{h_{21}}, L_{h_{22}}) = (0.2, 0.4)\cdot L_2$. The main difference observed between the two use cases is in the way traffic is offloaded at the SC clusters. As mentioned above, it is $L_{11}^{sc} = L_{12}^{sc}$ for any (L_1, L_2), when the traffic is evenly distributed among the SC clusters. Conversely, this does not occur for uneven traffic allocations. We observe that given an MNO_2 level of competition, MNO_1's offloading traffic can be the same (i.e. $L_{11}^{Dsc} = L_{12}^{Dsc}$) only for high L_1 values (e.g. $L_{12}^{sc} = 200$ Mbps for $L_1 = 1000$ Mbps). This can be explained by the fact that the MNOs may prioritize the traffic among hotspots. This is especially observed for the co-channel deployment, where MNO_1 prefers to offload at SC cluster 1. This prioritization occurs because it is cheaper for MNO_1 to offload $L_{h_{11}}$, due to the lower competition in SC cluster 1 ($L_{h_{21}} = 0.2 \cdot L_2$), and its lower offloading demand at SC cluster 2 ($L_{h_{11}} > L_{h_{12}}$).

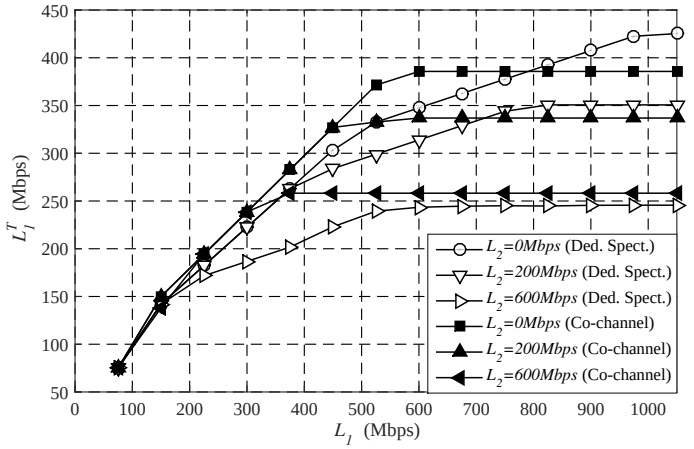
Despite this prioritization, the maximum L_i^{sc} of each MNO_i is the same as in the even traffic distribution case, for both spectrum deployments. Consequently, each MNO_i achieves the same maximum L_i^T in both cases, for both spectrum deployments. The difference lies in the fact that this maximum L_i^T is achieved for different (L_1, L_2) values in each use case. For instance, given $L_2 = 0$ the maximum $L_1^{DT} \cong 425$ Mbps is achieved for $L_1 = 700$ Mbps when $L_{h_{11}} = L_{h_{12}}$, whereas for $L_1 = 1000$ Mbps when $L_{h_{11}} \neq L_{h_{12}}$. Similarly, when C_{BH} is equally divided between the two MNOs, the maximum L_i^{DT} of each MNO_i is the same in both use cases, however it is achieved for different levels



(a)

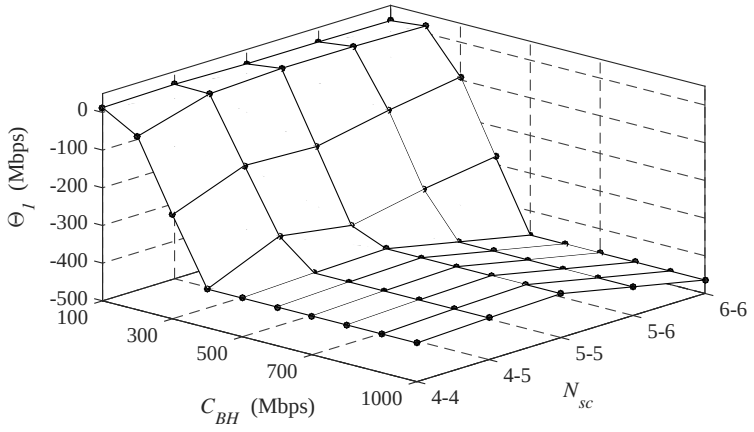


(b)

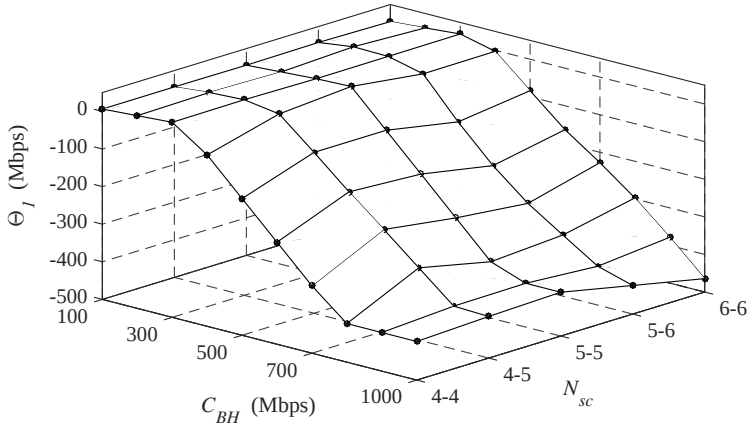


(c)

FIGURE 3.3: Traffic served by (a-b) the SC clusters and (c) MNO_1's throughput versus the total offered load for MNO_1, with $(L_{h_{11}}, L_{h_{12}}) = (0.4, 0.2) \cdot L_1$, $(L_{h_{21}}, L_{h_{22}}) = (0.2, 0.4) \cdot L_2$



(a) L_2=0 Mbps



(b) L_2=1600 Mbps

FIGURE 3.4: Served load difference of the dedicated spectrum and co-channel deployments, Θ_1

of competition (i.e. for $(L_1, L_2) = (400, 400)$ Mbps when $L_{h_{11}} = L_{h_{12}}$, whereas for $(L_1, L_2) = (600, 600)$ Mbps when $L_{h_{11}} \neq L_{h_{12}}$). As mentioned above, this occurs due to the prioritization of traffic in the SC clusters. Each MNO_i will prioritize the traffic at the SC cluster with the highest load distribution and the least competition, since it is cheaper and generates higher revenue. As the offered load increases in the cluster with the lower load distribution, the revenue from offloading this traffic justifies the corresponding increase in bid by MNO_i.

Impact of Backhaul Capacity: According to the results obtained previously, Lsc_1 can be limited either by C_{BH_l} or by the SCO access link capacity ($N_{sc_l} B_1 x_1 SE_{sc_l}$). It was also noted that the dedicated spectrum deployment presents higher access link capacity than the co-channel deployment, and therefore it is more likely to be limited by C_{BH_l}. We consider the same scenario as previously, but the backhaul capacity values vary within the range $C_{BH_l} \in [100, 1000]$ Mbps. In order to analyse the impact of C_{BH_l},

the metric defined in (3.5) as $\Theta_i = L_i^{CT} - L_i^{DT}$ is plotted in Fig. 3.4 when $L_1 = 1700$ Mbps, $L_2 = 0$ Mbps or $L_2 = 1600$ Mbps, and $L_{h_{il}} = 0.3 \cdot L_i, \forall i \in \mathcal{N}, l \in \mathcal{N}_\mathcal{C}$.

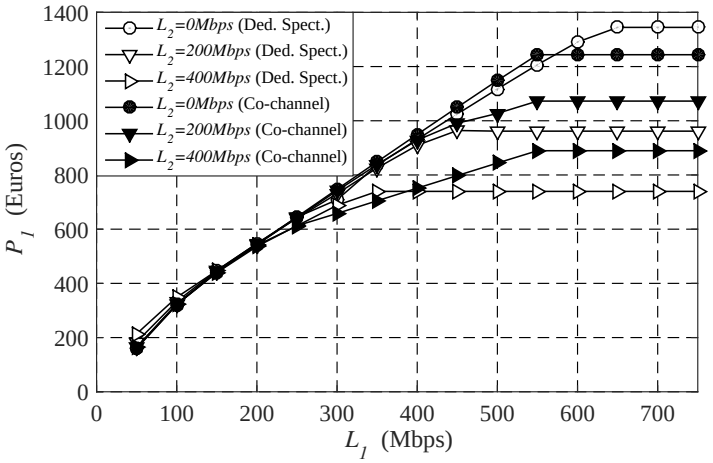
Fig. 3.4a shows that, regardless N_{sc}, both use cases (dedicated spectrum and co-channel) provide similar results when C_{BH_l} is small, thereby obtaining $\Theta_i \cong 0$. As observed in Fig. 3.2, when L_i^{sc} is not constrained by C_{BH_l}, the maximum L_1^{Dsc} is higher than the maximum L_1^{Csc}. However, if C_{BH_l} is smaller than these maximum offloaded traffic values, both use cases perform in similar terms. This is what can be seen in Fig. 3.4a for small C_{BH_l} values, where $\Theta_i \cong 0$ Mbps. However, as C_{BH_l} is increased, the co-channel deployment offloading gets limited by the access link capacity (due to the lower spectral efficiency of the co-channel use case) while the dedicated deployment remains constrained by the backhaul capacity. Therefore, whereas L_i^{CT} remains constant despite the increase of C_{BH_l}, L_i^{DT} grows with C_{BH_l}. Consequently, Θ_i falls. Finally, Θ_i stabilizes when the dedicated spectrum deployment also becomes limited by the access link capacity.

In Fig. 3.4a, Θ_1 falls slower for $(N_{sc_1}, N_{sc_2}) = (6, 6)$ than for $(N_{sc_1}, N_{sc_2}) = (4, 4)$ (e.g. Θ_1 is -224Mbps and -92Mbps respectively for $C_{BH_l} = 300$Mbps). Hence, the maximum access link capacity depends not only on the spectral efficiency, but also on the SC density. Based on this, it is clear in Fig. 3.4a that the increase of the access link capacity achieved in a denser SC cluster (i.e. high N_{sc}) is translated into a slower fall of Θ_i. Fig. 3.4b plots the same results when L_2 is high and, therefore, C_{BH_l} is shared between MNO_1 and MNO_2. The explanation is exactly the same as for Fig. 3.4a, though in this case, and according to the outcome discussed previously for Fig. 3.2a, MNO_1 can lease SC capacity equal to $C_{BH_l}/2$ instead of C_{BH_l} (MNO_2 leases the rest of C_{BH_l}). This is the reason why we can observe the same trend for Θ_1 in Fig. 3.4b, however for a wider range of C_{BH_l} values.

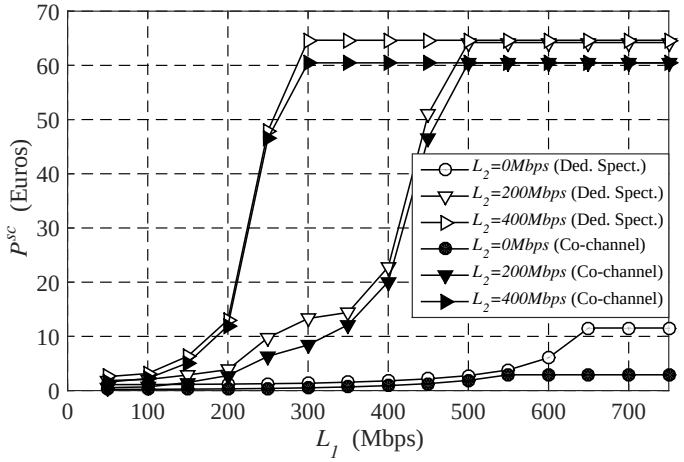
3.5.3 Stakeholders' Profits

There is a tight relation between the auction results depicted in Fig. 3.2 and the profit obtained by the MNO and the SCO. In order to highlight it, Fig. 3.5a and 3.5b show MNO_1's and SCO's profit respectively, for the same scenario and use case as in Fig. 3.2. Regarding P_1, we observe that it follows the same trend as L_1^T. This behaviour is explained with (3.6). Particularly, by offloading volumes of traffic significantly larger than what its own eNB can serve, MNO_1 generates a large revenue that covers both its own OPEX (i.e. CL_1) and the offloading cost (i.e. the bids).

Regarding the SCO's profit, we observe in Fig. 3.5b that P_{sc} is convex. This is explained by the fact that the MNOs can place bids close to the reserve price for low to medium total demand with the assistance of the learning mechanism (as will be explained in



(a)



(b)

FIGURE 3.5: Profit gained by (a) MNO_1 and (b) SCO versus the total offered load for MNO_1, with $L_{h_{il}} = 0.3 \cdot L_i, \forall i \in \mathcal{N}, l \in \mathcal{N}_\mathcal{C}$

Section 3.5.5). Hence, P_{sc_l} is slightly higher than $P_{sc_l}^{min} = (z-1)CL_{sc_l}$, which is convex. We further observe the impact of the MNO competition on the SCO profit. Particularly, we see that as the competition increases (i.e. $L_2 = \{200, 400\}$Mbps), P_{sc} becomes saturated for lower L_1 values. This is explained by the fact that the higher the competition, the higher the need for SC capacity. Thus, P_{sc} will be high when all of the contending MNOs have high offloading demands. Finally, it can be seen that P_{sc} is higher with the dedicated spectrum deployment. This occurs because with the dedicated spectrum deployment the MNOs offload slightly more traffic than with the co-channel deployment, as observed in Fig. 3.2a.

3.5.4 Auction Scheme Comparison

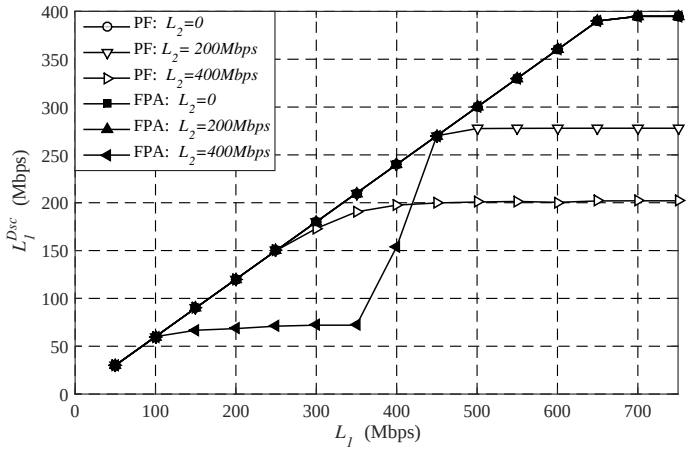
In this section, we conduct a comparison between our proposed Proportional Fair Auction scheme (PF) and the well-known First-Price sealed bid auction scheme (FPA) [37]. Particularly, we solve the SWM problem in (3.12), and compare the performance of each auction in terms of offloaded traffic (L_i^{sc}), and total system throughput ($\sum_{i\in\mathcal{N}} L_i^T$).

In FPA, the highest bidder wins the auction, and pays the bid it had submitted. In order to adjust FPA in our system model, we consider the following. We assume that two MNOs place their bids b_{1l} and b_{2l} in order to lease L_{1l}^{sc} and L_{2l}^{sc}, respectively, from a single SC cluster l. Since we solve the SWM problem, the MNOs will bid at least the reserve price (i.e. $b_{il} \geq b_{il}^{min}, i \in \mathcal{N}, l \in \mathcal{N}_\mathcal{C}$), so that the MNOs can maximize their profit. If the submitted bids $b_{1l} + b_{2l}$ produce the minimum profit $P_{sc_l}^{min}$ for $L_{sc_l} \in (0, C_{BH_l}]$ (i.e. $b_{1l} + b_{2l} = zCL_{sc_l}(L_{sc_l})$), L_{sc_l} will be allocated among the MNOs according to the following rule:

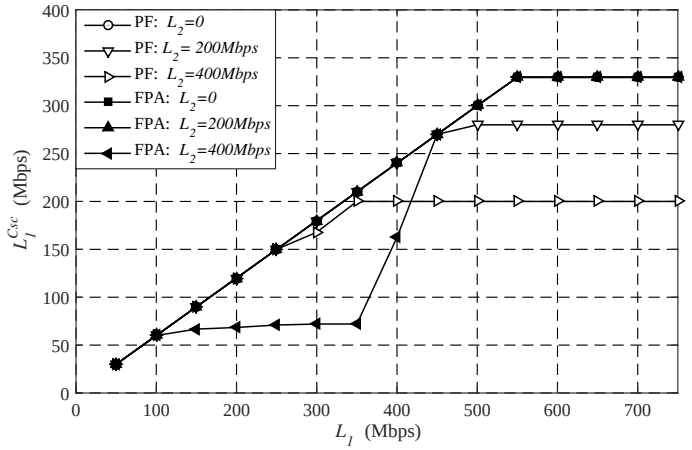
The winning bid of MNO_i b_{il}, $i \in \{1,2\}, l \in \mathcal{N}_\mathcal{C}$ is the one that maximizes the SC cluster l's profit, that is, the winning bid is given by $b_{il} = \arg\max_{b_{jl}, j\in\mathcal{N}} (P_{sc_l} = b_{jl} - CL_{sc_l}(L_{jl}^{sc}))$. *Hence, MNO_i will lease the requested SC capacity L_{il}^{sc} by paying b_{il}. Then, MNO_j will lease the remaining SC capacity $L_{jl}^{sc} = L_{sc_l} - L_{il}^{sc}$ by paying its placed bid, b_{jl}.*

For the comparison of the two auction schemes we use the same scenario and use case as in subsection 3.5.2 (i.e. $N_{sc_1} = N_{sc_2} = 4$, $C_{BH_1} = C_{BH_2} = 200$Mbps, and $L_{h_{il}} = 0.3 \cdot L_i, \forall i \in \mathcal{N}, \forall l \in \mathcal{N}_\mathcal{C}$) for 2 MNOs. Fig. 3.6a and 3.6b depict L_1^{Dsc} and L_1^{Csc}, respectively, as a function of L_1 for different L_2 values. The same experiments have been carried out for the proposed PF scheme and the compared FPA scheme. As expected, when there is no competition (i.e. $L_2 = 0$) MNO_1 will bid the reserve price in either case, and offload as much traffic as possible (as explained in Section 3.5.2), hence achieving the same L_1^{sc}. MNO_1's strategy changes when the competition increases (i.e. $L_2 > 0$). As described in Section 3.5.2, with the PF scheme the SC cluster capacity is divided proportionally among the MNOs according to their offered load, which shapes their offloading demands. Furthermore, for offered loads beyond a particular value (in this case $L_i > 350$ Mbps, $i \in \mathcal{N}$), each MNO_i offloads traffic approximately equal to $\sum_{l\in\mathcal{N}_\mathcal{C}} C_{BH_l}/2$.

Conversely, we observe for both spectrum deployments that with the FPA scheme most of the SC capacity is allocated to the MNO with the highest offered load. Moreover, the opponent MNO does not offload the remaining SC capacity. This occurs because the allocation rule overcharges the loser of the auction. We notice that for L_2=400 Mbps and $L_1 \leq$400 Mbps, where MNO_2 wins the auctions, MNO_1 does not lease the remaining C_{BH_l}. Conversely, MNO_1 wins the auctions for L_1>400 Mbps and offloads as much traffic as possible ($L_1^{sc} = \sum_{l\in\mathcal{N}_\mathcal{C}} L_{h_{il}}$ or $L_1^{sc} \cong \sum_{l\in\mathcal{N}_\mathcal{C}} C_{BH_l}$).



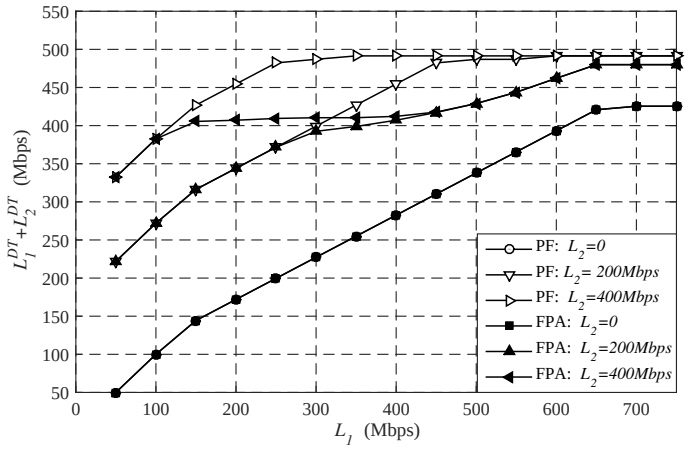
(a)



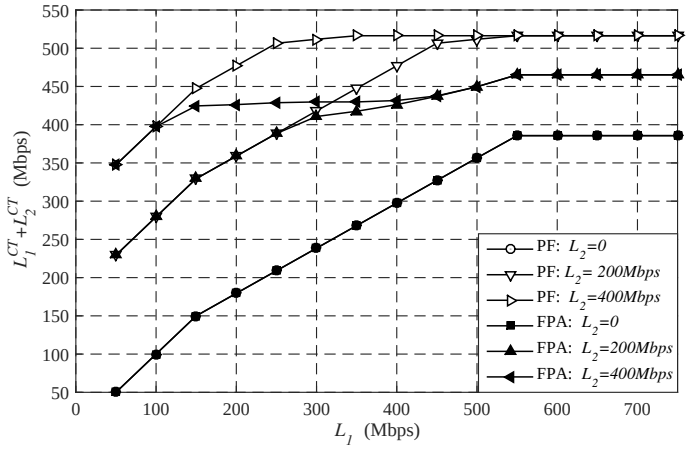
(b)

FIGURE 3.6: Leased SC capacity for the PF and FPA schemes for (a) dedicated spectrum and (b) co-channel deployment versus L_1 for MNO_1, with $L_{h_{il}} = 0.3 \cdot L_i, \forall i \in \mathcal{N}, l \in \mathcal{N}_\mathcal{C}$

The results of the above allocation rules are shown in Fig. 3.7a and 3.7b. These figures depict $\sum_{i\in\mathcal{N}} L_i^{DT}$ and $\sum_{i\in\mathcal{N}} L_i^{CT}$ respectively, as a function of L_1 for different L_2 values. It can be observed in both figures that the proposed PF scheme either outperforms or performs equally with the compared FPA scheme. The sum throughput is the same for both auctions schemes when there is no competition (i.e. $L_2 = 0$), since the bidding behaviour is the same as explained previously. When the competition is higher (i.e. L_2>0), PF achieves higher throughput than FPA. This occurs due to the allocation rule of FPA, which favours the winner, but disincentivizes the loser with expensive SC capacity. Thus, we observe that when the competition is high (i.e. $L_2 = [200, 400]$ Mbps), the sum throughput's gradient decreases for medium loads (i.e. $L_1 \in [150, 350]$ Mbps). When L_1 increases beyond the value where $L_{h_{1l}} = C_{BH_l}, l \in \mathcal{N}_\mathcal{C}$ (i.e. $L_i > 333$ Mbps), the gradient increases up to the point that the maximum sum throughput is achieved.



(a)

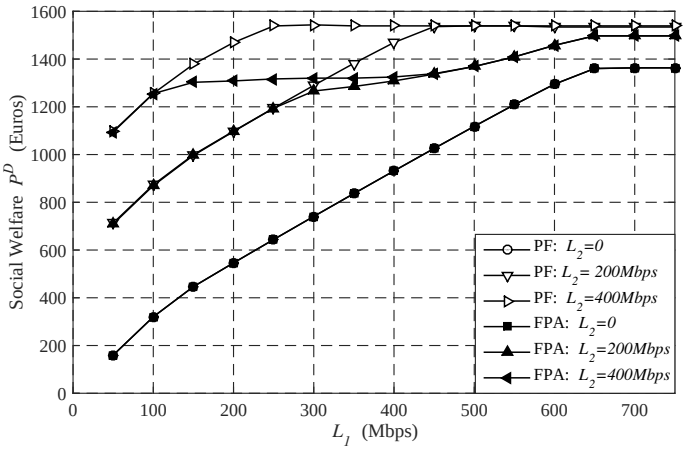


(b)

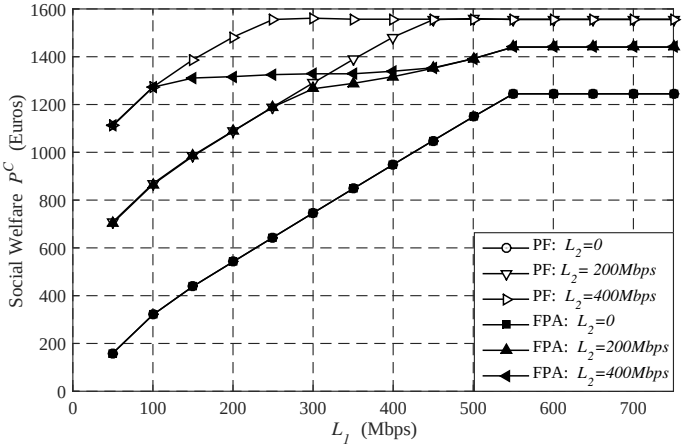
FIGURE 3.7: System sum throughput $\sum_{i\in\mathcal{N}} L_i^T$ of the PF and FPA schemes for (a) dedicated spectrum and (b) co-channel deployment versus L_1 for MNO_1, with $L_{h_{il}} = 0.3 \cdot L_i, \forall i \in \mathcal{N}, l \in \mathcal{N}_\mathcal{C}$

Fig. 3.8a and 3.8b depict the Social Welfare ($P = \sum_{i\in\mathcal{N}} P_i + P_{sc}$) for the dedicated spectrum and co-channel deployments as a function of MNO_1's offered load (L_1) for different MNO_2 offered load levels (L_2). Due to the tight relation between the MNO throughput and profit (the high MNO revenues cover their costs as also observed in Fig. 3.5a), we notice that the SW shows a trend similar to that of the sum throughput. Hence, it can be observed in Fig. 3.8a and 3.8b that PF either outperforms or performs equally with FPA.

From the above we conclude that the PF scheme results in a more balanced allocation of the SC capacity, which then leads to higher throughput, and higher profits for the auction stakeholders.



(a)



(b)

FIGURE 3.8: Social Welfare P of the PF and FPA schemes for (a) dedicated spectrum and (b) co-channel deployment versus L_1 for MNO_1, with $L_{h_{il}} = 0.3 \cdot L_i, \forall i \in \mathcal{N}, l \in \mathcal{N}_\mathcal{C}$

3.5.5 Learning Mechanism Performance

The performance of the learning mechanism described in Section 3.4.2.2 is evaluated in a scenario where future loads and opponents' bids are unknown. We use the same scenario and use case as in subsection 3.5.2 (i.e. $N_{sc_1} = N_{sc_2}$=4, $C_{BH_1} = C_{BH_2}$ = 200Mbps, and $L_{h_{il}} = 0.3 \cdot L_i, \forall i \in \mathcal{N}, \forall l \in \mathcal{N}_\mathcal{C}$) for 2 MNOs. Regarding L_i, the day is divided into $T = 24$ timeframes, and the traffic pattern for both MNOs is modelled as a bimodal distribution with two peaks, the first one at timeframe $t = 12$ and the second one at $t = 16$. In the simulation, each L_i instance is randomly generated according to this pattern. Table 3.5 shows the values of the parameters for the three algorithms that compose the learning mechanism, which have been obtained through simulations.

Since L_{h_i} is distributed evenly and $L^{sc}_{i1} = L^{sc}_{i2}$, we provide results on the convergence

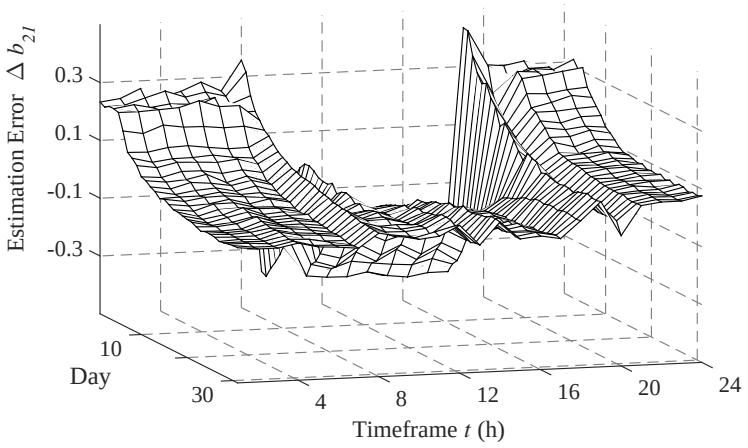
TABLE 3.5: Learning Mechanism parameters

Traffic forecasting		**Adaptive Search range**	
α	0.2	Expansion coefficient (ϵ_c)	1
β	0.1	Reduction coefficient (ρ_c)	17
γ	0.1	Expansion rate (ϵ_r)	0.3
Average threshold (μ_{HW})	0.15	Reduction rate (ρ_r)	0.5
Deviation threshold (σ_{HW})	0.2	Average threshold (μ_{SR})	0.2
Learning Algorithm		Estimated-Real difference threshold (ΔL_{SR})	0.15
Size of $\mathcal{A}'_{b_{j\neq i}}$ (K)	3		
Learning speed (η)	0.3		

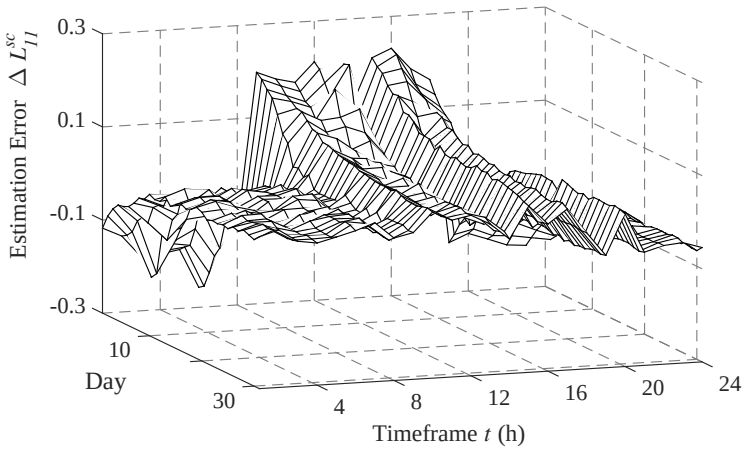
speed and the performance of the learning mechanism only for SC cluster 1. The error in the estimation of L_{11}^{sc} and b_{21} (in this case, as only two MNOs are considered, $b_{j1\neq 1} = b_{21}$) are hereafter denoted by ΔL_{11}^{sc}, defined in (3.15), and $\Delta b_{21} = \frac{\overline{b_{21}} - b_{21}}{b_{21}}$, where $\overline{b_{21}}$ is the estimate of b_{21}. Fig. 3.9a and 3.9b show the convergence of the estimation errors Δb_{21} and ΔL_{11}^{sc}, respectively, within a range of 31 days (note that the estimation of each timeframe is carried out once per day, and thus 31 days are equivalent to 31 iterations). It can be observed that despite the initial poor estimates of b_{21}, the mechanism accomplishes estimation errors of the order [-10,10]% within the first 8-14 iterations in all timeframes, and small errors upon convergence (e.g. $|\Delta L_{11}^{sc}|$<6% and $|\Delta b_{21}|$<7.5%).

It is observed that the highest absolute estimation errors of b_{21} occur at t={7,19,21,24}, far from the two peak hours defined at t=12 and t=16. In fact, the absolute error $|\overline{b_{21}} - b_{21}|$ is higher for high loaded timeframes, but the significant differences between b_{21} in high and low loaded timeframes makes Δb_{21} be higher for low loaded timeframes. In order to see the impact of these errors, Fig. 3.10 shows the throughput, the profit and the bid of MNO_1 when all the information is available to all stakeholders (i.e. the SWM problem defined in (3.12) is solved and $\Delta L_{11}^{sc} = \Delta b_{21} = 0$), and when information is not available and the learning mechanism is used (i.e. with ΔL_{11}^{sc} and Δb_{21} shown in Fig. 3.9). It can be observed in Fig. 3.10a that the served load (or throughput) achieved with the learning mechanism matches almost perfectly the results obtained when all information is known. This good performance is owing to: i) good estimations plotted in Fig. 3.9, and ii) less accurate estimations occur when capacity needs are less stringent.

Similar results are obtained in Fig. 3.10b for the profit of MNO_1, P_1. However, it is worth noting that the profit decreases with the learning mechanism at timeframe t=12 (the peak hour). This reduction of the profit is the result of a high bid at t=12 (depicted in Fig. 3.10c). If the offered bid b_1 is higher than the reserve price (the bid when all information is available), the profit drops. Generally, the high bid is a consequence of the overestimation of b_{2l} and/or L_{1l}^{sc}. In this particular case, Fig. 3.9 shows at t=12 that



(a)



(b)

FIGURE 3.9: Convergence of error of the estimated (a) b_{21} and (b) L_{11}^{sc} made by MNO_1

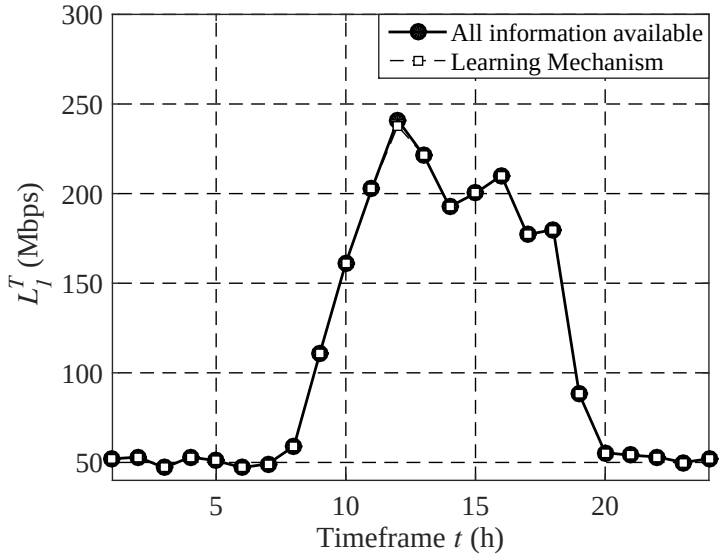
$\Delta L_{11}^{sc} > 0$ and $\Delta b_{21} < 0$; the former overestimates L_{11}^{sc} and the latter underestimates b_{21}. Therefore, the overestimation of L_{11}^{sc} causes the decrease of the profit.

Yet, despite the slight decrease in the profit experienced during the peak hour, the learning mechanism presents a very good performance in the simulated scenario.

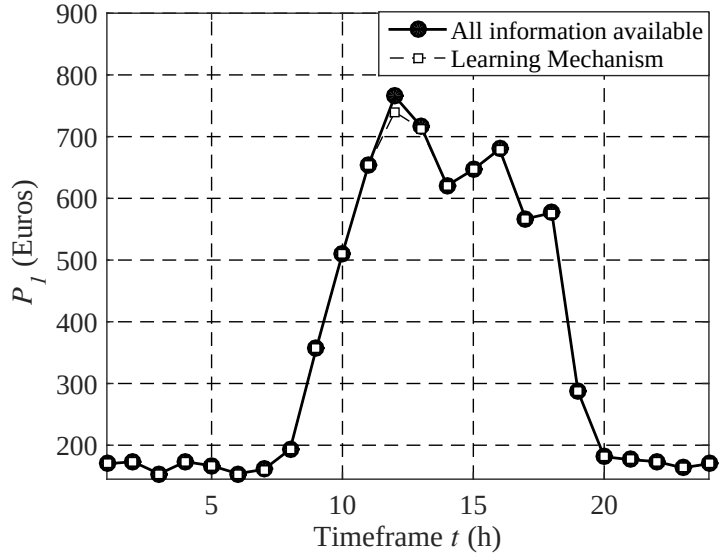
Summary of Results

The main key findings of Section 3.5 can be summarized in the following:

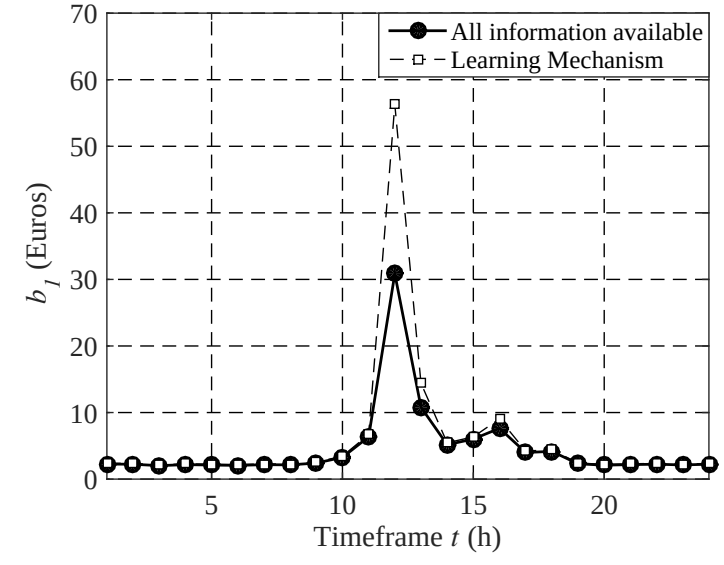
- MNOs achieve higher throughput when the competition is low with the dedicated spectrum allocation, whereas when the competition is high the co-channel spectrum allocation achieves higher performance. This occurs independent of the traffic distribution in the SC clusters.



(a)



(b)



(c)

FIGURE 3.10: Comparison between (a) throughput, L_1^T; (b) profit, P_1; and (c) bid, b_1, of MNO_1

- The MNO throughput can be limited either by the backhaul capacity or the access link capacity. Due to its higher spectral efficiency, the dedicated spectrum allocation throughput is limited by the access link capacity for lower backhaul capacity values than the co-channel allocation throughput.
- The MNO profit follows the same trend as the MNO throughput. Regarding the SCO profit, it follows the same trend as the SCO cost and it increases with MNO offered load, according to the definition of the minimum profit P_{sc}^{min} in Proposition 3.4.
- The proportional fair auction scheme results in a more balanced allocation of the SC capacity compared to the first price auction, which also results in higher throughput, and higher profits for the auction stakeholders.
- Our proposed learning mechanism converges to its solution in a short number of iterations.
- Our proposed learning mechanism provides results similar to the optimal solution of the social welfare problem in (3.12).

3.6 Applicability

Regarding the applicability of our model (and generally the models found in the literature) in real-world scenarios, we have encountered companies in the mobile network infrastructure market, which have adopted similar market frameworks in their business models. In the past few years, we have observed that the continuous growth of the telecommunication industry has brought forth the need for expansion of the existing 4G and the deployment of new 5G networks. To that end, MNOs intend to expand their network, and are looking for new cell sites. This demand has created a market of valuable telecommunication assets, which already counts billions of dollars only in the United States. These assets include cellular and telecommunication towers, antennas, license agreements, which are either leased or sold to MNOs or other third-party investors.

Various companies have taken advantage of this situation by offering a marketplace, mediating the transactions between sellers and buyers, and some of them even offering network management services. Two examples of companies that offer this kind of brokering services are Cell At Auction, LLC [95] and SteepSteel, LLC [96]. Both these companies have created their own marketplace, where owners of cellular towers or cell sites can either lease or sell their property. Moreover, each item to be sold or leased

is auctioned, in an effort to maximize the seller's profit. Depending on the availability of cell towers in a geographical area, there can be either a monopoly (i.e. a single owner-lessor/seller as the SCO in our scenario) or a competition among owners.

Discovering a multitude of such companies, we notice an expanding trend in the market of mobile network infrastructure, which will play an important role in the deployment of 5G networks, as it facilitates the process of infrastructure acquirement and capacity augmentation for MNOs. Moreover, we have seen that the corresponding marketplaces are not based on a single model, and hence we also encounter brokers that offer marketplaces based on auction schemes. Such practices prove that auction schemes can be used in realistic scenarios as frameworks for the transaction and resource allocation of mobile network infrastructure.

Autonomous bidding or trading software enables the adoption of such schemes thanks to their capacity of functioning in the high frequencies that auctions or bargaining actions are conducted, as described in [97]. These *software agents* are designed with machine learning or other artificial intelligence tools in order to act in the interest of their owner and improve their decisions based to previous outcomes, similar to our approach in this contribution.

3.7 Concluding Remarks

In this chapter, we have presented our contribution on traffic offloading under the SCaaS approach, where a small cell operator owns the SC infrastructure, and the mobile network operators transfer spectrum resources to serve their users. The problem of the efficient capacity distribution, both from a network and a financial perspective, has been modelled with a proportionally fair auction scheme. We further show that the uncertainty about future traffic load poses the necessity to develop learning mechanisms to assist the auction. Modified versions of the Holt-Winters method and the Exp3 reinforcement learning algorithm have been proposed to deal with the load forecasting and the opponents' bid estimation, respectively. Extensive simulations with the proposed mechanism were used to study the MNOs' auction strategies, for two spectrum deployments, and different SC densities. The results show how the competition level among the MNOs impacts their profit and capacity, and analyse their trade-off. Moreover, they reveal that the capacity is limited by either technical (i.e. backhaul or spectral efficiency) or economic causes, and explain how they are connected to the MNOs' competition. Finally, we show that the proposed forecasting, learning and auction scheme copes with the uncertainty efficiently, and provides results comparable with scenarios without uncertainty.

Chapter 4

Network and Financial aspects in RAN

4.1 Introduction

The exponential growth of mobile data traffic experienced over the last years is expected to continue in the future. This traffic growth is mainly the result of the surge in demand of multimedia and video content (usually offered by independent content providers) and the explosion in the number of devices and broadband connections. Therefore, mobile traffic will be heterogeneous from a dual perspective, since there will be a wide range of possible QoS requirements (e.g. video on-demand, online gaming and messaging are very different in terms of QoS) and it will be originated/received by diverse devices (e.g. tablets, laptops or smartphones) [11]. In this context, it has been shown that the quality perceived by the users can not be fully captured with QoS metrics, thus making QoE one of the most important Key Performance Indicators (KPI) [12]. Therefore, in the future MNOs will have to be able to meet not only the envisioned boost of the traffic demand, but also its heterogeneity in terms of QoS/QoE requirements.

As mentioned in the previous chapters, RAN densification is one of the main strategies to catch up with the future intense and diverse demand [98]. Mobile industry must invest large amounts of capital in the deployment of dense HetNets or the leasing of small cell infrastructure (as described in our first contribution) to be able to provide seamless connectivity and high QoS/QoE. Therefore, the densification strategy to cope with the increase of the demand has a significant impact on the balance sheet of the different stakeholders, and particularly for MNOs (e.g. increase of the deployment or infrastructure leasing cost). These financial aspects/constraints are exacerbated by the so-called *traffic and revenue paradox/challenge* [15]; specifically, although it may seem

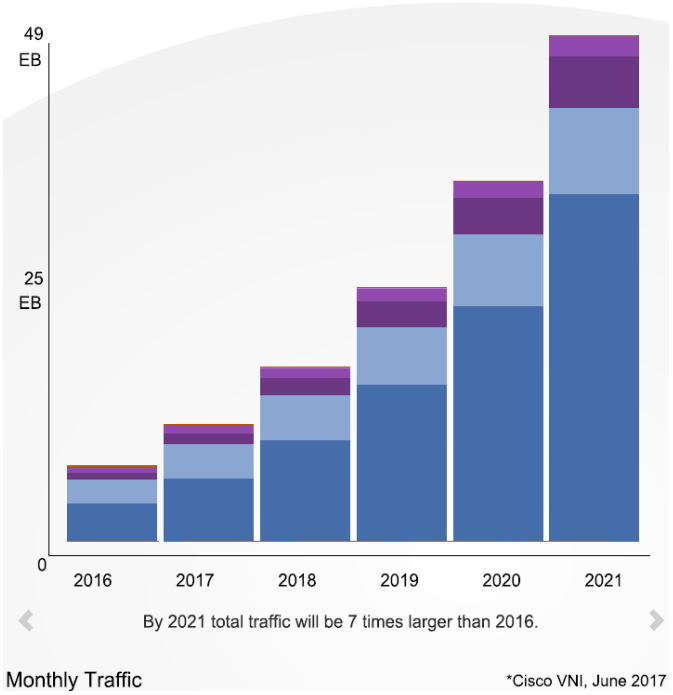


FIGURE 4.1: Forecast of monthly internet traffic by Cisco

contradictory, the described traffic boost shown in Fig. 4.1 has increased the content providers' profits while, simultaneously, has diminished the MNOs' revenues. This occurs because the MNO's basic services (voice and messaging) have been gradually replaced by their third-party counterparts. Moreover, the MNO's data service prices have been decreasing over the years, due to the market competition.

Therefore, MNOs face a two-fold challenge: meet the QoE requirements and maximize the profit. It has been proven that the relation between QoS and QoE has a non-linear nature [17]. This means that small degradations in the received QoS can significantly impact on the perceived QoE level. Yet, QoE is influenced by other factors such as pricing or device characteristics [19]. In this context, it is necessary to design network and economic functionalities adapted to the new requirements, such as QoE-aware Radio Resource Management strategies and smart dynamic pricing, and always trying to maximize the profit (to compensate the diminished MNOs' revenues and the increasing deployment investment).

As we have shown in Section 2.3.2, the majority of works on User Association, Resource Allocation (RA) and scheduling in the context of 4G and 5G networks focus mainly on the provision of high QoS/QoE and other network aspects (e.g. power allocation, fairness etc.), however without taking into account the impact of their proposals on the financial aspects of the MNOs (e.g. profit). Regarding smart dynamic pricing, we have shown in Section 2.3.3 that the majority of the works in the literature use dynamic pricing in order to steer the traffic demand and direct it from peak to off-peak traffic hours

and locations, in an effort to minimize congestion periods. That is, dynamic pricing is used as a tool to motivate the users change their data consumption habits, thus avoiding network congestion. However, such approaches fail to provide high customer satisfaction during inevitable congestion periods.

In this chapter, we study the user association, resource allocation and dynamic pricing problems aiming to maximize the MNO profit, while offering high QoE to the users, under fairness and overall user satisfaction (OS) constraints. We consider HetNets composed of macrocell and SC base stations (BSs), with dynamic traffic described by numerous QoS/QoE demands. In contrast to the literature, we further consider diverse pricing, that is, various service prices and different pricing schemes. In order to address the challenges of traffic heterogeneity and high network profitability, our solutions exploit the QoE-awareness and the network's economic aspects (i.e. the MNO profit). The main contributions of the chapter are summarized in the following:

- In Section 4.5.1, we propose a greedy QoE-aware user association, heuristic algorithm to maximize the MNO profit under fairness constraints in HetNets composed of macrocell BSs operating in the sub-6GHz microwave (μWave) band and small cells operating in the millimetre-wave (mmWave) band.
- In Section 4.5.2, we propose a QoE-aware resource allocation algorithm for profit maximization under overall user satisfaction and fairness constraints, analyse the connection among the individual and overall user satisfaction, fairness, pricing and profit, as well as how they impact each other. Particularly:
  - We propose a heuristic, greedy, low-complexity, QoE-aware resource allocation algorithm that maximizes the MNO profit while imposing constraints on the minimum overall users satisfaction and on the fairness among users. Our simulation results show that the proposed algorithm outperforms state-of-the-art algorithms.
  - We shed light on the trade-off between users' satisfaction, fairness and MNO profit. We show that, given the non-linear relation between QoS and QoE, there is room for profit maximization resources allocation solutions without penalizing the quality perceived by the users.
  - We show that the use of diverse pricing schemes (e.g. data-based and time-based pricing) can lead to the prioritization of some services over the others in *pure* profit maximizing algorithms. Moreover, as the QoE is affected by the price level, the adjustment of the service prices by the MNO can affect both the profit and overall satisfaction. In line with this, we provide an

accurate set of simulation results that shows the sensitivity of profit and overall satisfaction with respect to pricing schemes and price level.

- Going beyond the proposal of a dedicated resource allocation scheme included in Section 4.5.2, in Section 4.5.3 we propose a greedy joint resource allocation and dynamic pricing algorithm for MNO profit maximization under overall satisfaction constraints. Particularly:
  - Instead of impacting the MNO profit and system performance by changing equally the price of all users (as we did in Section 4.5.2), in Section 4.5.3 we propose a heuristic algorithm that maximizes the MNO profit in a real-time scale, during congestion, by changing the charging for specific users.
  - Our proposal on dynamic pricing provides real-time results and can be applied on the resource allocation scheme and pricing type an MNO already uses.

The rest of the chapter is organized as follows. Section 4.2 describes the system model, and Section 4.3 states the MNO's objectives. In Section 4.4, we formulate the profit optimization problems, and in Section 4.5 we propose the QoE-aware profit maximizing algorithms. We validate our algorithms in Section 4.6, and conclude the chapter in Section 4.7.

4.2 System Model

The considered network is composed of a set of macrocells and a set of small cells, all of them deployed by a single MNO. We denote this set of BSs, both macrocells and small cells, as $\mathcal{B} = \{1, 2, \ldots, N_B\}$, where N_B is the total number of BSs. The bandwidth allocated to each BS $i \in \mathcal{B}$ is hereafter referred to as b_i (in Hz). The notation used henceforth is summarized in Table 4.1.

It should be noted that in Section 4.5.1 we addressed the user association problem for a 5G HetNet setup. To that end, the considered network is composed of a set of macrocells operating in the μWave band and a set of small cells operating in the mmWave band, all of them deployed by a single MNO as in Section 4.5.3 and Section 4.5.2. The mmWave small cell deployment has been extensively addressed in the literature and proposed as a pivotal solution in 5G for two main reasons. First, the bandwidth availability in mmWave bands is higher than in the μWave bands, thereby alleviating the spectrum scarcity problem; second, thanks to the limited interference realized by the use of highly directional antennas, very dense deployments are feasible, thus enhancing the network spectral efficiency [99].

TABLE 4.1: Notation

Notation	Description
$\mathcal{B}$	Set of BSs
$\mathcal{U}$	Set of users
$\mathcal{U}_i$	Set of users in BS i
$\mathcal{U}_i^t$	Set of time-based charged users in BS i
$\mathcal{U}_i^B$	Set of data-based charged users in BS i
$\mathcal{S}$	Set of services
$\mathcal{Q}$	Set of QoE levels
$\mathcal{D}$	Set of devices
b_i	BS i's available bandwidth
π_k	Service Profile (SP) k
s_k, q_k, p_k	Service, QoE class, service price of SP k
θ_k^B	Per data unit charging of service k
θ_k^t	Per time unit charging of service k
r_j	User j's data rate
r_{kd}	Target data rate for SP k, device d
w_{ij}	Resources allocated to user j by BS i
w_i	$\sum_{j\in\mathcal{U}_i} w_{ij}$
ε_{ij}	Spectral efficiency between user j and BS i
Q_j^{kd}	QoE level of user j with SP k, device d
Q_k^{tg}	Target QoE level for SP k
Q_k^{drop}	Service dropping QoE level for SP k
$\widehat{Q}_j^{kd}$	QoS-based component for QoE mapping
Q_p	Price-based component for QoE mapping
$\alpha_{kd}, \gamma_{kd}, \beta_{kd}$	SP-dependent constants for $\widehat{Q}_j^{kd}$
v_k	User-dependent constant for Q_p
σ_{ij}	Satisfaction of user $j \in \mathcal{U}_i$
J_i	Jain's fairness index in BS i
OS_i	Overall User Satisfaction (OS) in BS i
OS_i^{max}	Maximum possible OS in BS i
ϕ_i	Relative OS in BS i, OS_i/OS_i^{max}
ϕ^{min}	Relative OS threshold
P_i	BS i profit
R_i	BS i revenue
CB_i	BS i bandwidth utilization cost
c_i, h_i	Cost adjusting factors
λ_{ij}	Price percentage a user j pays

The MNO serves a set of users $\mathcal{U} = \{1, 2, \ldots, N_U\}$, where N_U is the total number of users. It is assumed that users are not served by more than a single BS simultaneously, and therefore we define the set of users served by BS $i \in \mathcal{B}$ as $\mathcal{U}_i$, where $\mathcal{U} = \cup_{i\in\mathcal{B}}\mathcal{U}_i$ and $\cap_{i\in\mathcal{B}}\mathcal{U}_i = \emptyset$. MNOs have put the focus on the QoS and QoE as the target KPI in the design of networks [12]. Accordingly, in our model each user has a contract with the MNO that specifies a desired QoE for each service, denoted in the sequel as *Service Profile* (SP). If we define the set of services as $\mathcal{S} = \{s : s = 1 \ldots S\}$ and the set of QoE

classes as $\mathcal{Q} = \{q : q = 1 \dots Q\}$ ($\mathcal{Q}$ is assumed to be a discrete and finite set), a generic SP can be defined as $\pi_k = (s_k, q_k, p_k)$, where p_k is the price of the service (in €), $s_k \in \mathcal{S}$ and $q_k \in \mathcal{Q}$. Focusing on p_k, it is worth noting that its definition depends on the service s_k. Thus, some services are charged based on the amount of transmitted/received data and some others are based on the connection time. Let us define the price for a data-based charged service as θ_k^B (in €/MB) and for a time-based charged service as θ_k^t (in €/sec). Moreover, we denote by $\mathcal{U}_i^B, \mathcal{U}_i^t \subseteq \mathcal{U}_i$ the sets of data-based and time-based charged users served by BS i, respectively (i.e. $\mathcal{U}_i^B \cup \mathcal{U}_i^t = \mathcal{U}_i$ and $\mathcal{U}_i^B \cap \mathcal{U}_i^t = \emptyset$). For the cases that our proposal on real-time dynamic pricing is applied, the users' service price must be reduced during short time periods T, as will be explained in detail in Section 4.5.3. To that end, a BS i determines the percentage of the price p_k a user $j \in \mathcal{U}_i$ will pay at a time period T, which we denote by $\lambda_{ij} \in [0, 1]$. The general expression of p_k for a time period T can be expressed as

$$p_k = \begin{cases} T\lambda_{ij}\theta_k^t & \text{If user } j \in \mathcal{U}_i^t \\ \frac{T \cdot r}{8}\lambda_{ij}\theta_k^B & \text{If user } j \in \mathcal{U}_i^B, \end{cases} \tag{4.1}$$

where r (in Mbps) is the user transmission rate[1]. As for the perceived QoE, in general any user with a service profile π_k has a target QoE level, Q_k^{tg}, and a minimum QoE level below which the session is dropped, Q_k^{drop} (in the Mean Opinion Score (MOS) scale [17]). We assume that both values are established in the contract between the user (as a customer) and the MNO. Regarding the price percentage λ_{ij}, it will be henceforth considered to be equal to 1 with the exception of the sections dedicated to dynamic pricing.

Although the perceived QoE is influenced by multiple factors, as it will be detailed in Section 4.3, we now focus on the impact of the user device. Nowadays, a single user can get connected to the network with different devices (tablet, laptop, smartphone, etc), each one with specific characteristics. These characteristics of the device, such as the screen quality or screen size, are relevant since they may improve or worsen the perceived QoE. For instance, to perceive similar QoE levels, lower image resolution and hence lower transmission bit rate (i.e. lower QoS) is required for a user using a video service in a small-sized screen smartphone than for the same user with a large screen tablet [18]. Therefore, the characteristics of the device must be also taken into account to design efficient radio resources management algorithms. We define the set of devices as $\mathcal{D} = \{d : d = 1 \dots D\}$, and the mapping function that links the device-SP pair with the required transmission rate, r_{kd}, as $f : (\pi_k, d) \rightarrow r_{kd}$. According to the definitions, the QoE perceived by a user $j \in \mathcal{U}$ with an SP π_k and using a device $d \in \mathcal{D}$, namely

[1]The user rate in (4.1) is divided by 8, so that it is expressed in MB/sec, and hence the price p_k in monetary units (i.e. €).

Q_j^{kd}, will be higher than the target QoE Q_k^{tg} if the transmission rate from the serving BS to the user j is higher than r_{kd}. In other words, the target QoE is met at time period t if $r_j(t) = w_{ij}(t)\varepsilon_{ij}(t)b_i \geq r_{kd} = f(\pi_k, d)$, where $r_j(t)$ is the actual transmission rate of user $j \in \mathcal{U}_i$ (in Mbps), $w_{ij}(t) \in [0, 1]$ is the portion of BS $i \in \mathcal{B}$ radio resources allocated to user j, and $\varepsilon_{ij}(t)$ is the spectral efficiency of the link between user j and BS i (in bps/Hz), which can be approximated as $\varepsilon_{ij}(t) = \log_2(1 + SINR_{ij})$, where $SINR_{ij}$ is the Signal to Interference and Noise Ratio received by user j, when served by BS i.

Note that, for a given device d and service s_k, each QoE class is translated into an equivalent QoS level (e.g. different r_{kd} values are required for streaming SD and HD video). This means that the user can choose among Q QoE classes on a contract basis for each service, depending on the personal preferences (e.g. preference for high browsing speed but SD video), and this choice impacts on the minimum QoS requirements. In turn, for a given service s_k and a QoE class q_k, the user can receive the service through a diversity of devices. For each device, the required transmission rate can also differ.

Based on the definitions stated above, it is clear that the satisfaction of users is tightly coupled with the perceived QoE. Specifically, if the satisfaction of user j served by BS i, namely $\sigma_{ij}(t)$, is defined within the interval [0,1], when $Q_j^{kd}(t) \leq Q_k^{drop}$, the session is dropped and the satisfaction is equal to 0. Conversely, when $Q_j^{kd}(t) \geq Q_k^{tg}$, the satisfaction is equal to 1. Thus, according to [19], the satisfaction is defined as

$$\sigma_{ij}(t) = \begin{cases} 0 & \text{if } Q_j^{kd}(t) \leq Q_k^{drop} \\ \frac{Q_j^{kd}(t) - Q_k^{drop}}{Q_k^{tg} - Q_k^{drop}} & \text{if } Q_j^{kd}(t) \in (Q_k^{drop}, Q_k^{tg}) \\ 1 & \text{otherwise.} \end{cases} \tag{4.2}$$

4.3 MNO's objectives

In order to propose RRM and pricing schemes based on network and economic functions, we first need to identify and analyse the two MNO's objectives. First, they must offer to the users the QoE agreed in the SP, and guarantee fairness when the system is congested. Second, the network must be managed so as to maximize their economic profit. In the following, the analyses of the QoE, fairness, overall user satisfaction and the profit are detailed.

4.3.1 Network Performance Metrics

Based on the analysis described in [19], the perceived QoE $Q_j^{kd}(t)$ can be divided into two components: the QoS-based component ($\widehat{Q}_j^{kd}(t)$) and the price-based component

$(Q_p(p_k))$.

$$Q_j^{kd}(t) = \widehat{Q}_j^{kd}(t) \cdot Q_p(p_k). \tag{4.3}$$

The QoS-based component, $\widehat{Q}_j^{kd}(t) \in [1, 5]$ (in the MOS scale), shows the effect of QoS level on QoE. In the literature, the QoE is usually modelled to have an exponential interdependency with the QoS, also known as the IQX hypothesis [17]. Using the transmission rate $r_j(t)$ as the reference QoS metric, and according to the IQX hypothesis, we can express $\widehat{Q}_j^{kd}(t)$ as

$$\widehat{Q}_j^{kd}(t) = \alpha_{k_j d_j} e^{-\beta_{k_j d_j} \Delta r_j(t)} + \gamma_{k_j d_j}, \tag{4.4}$$

where $\Delta r_j(t) = r_{kd} - r_j(t)$, $\alpha_{k_j d_j}, \gamma_{k_j d_j} > 0$ (both in the MOS scale) and $\beta_{k_j d_j} > 0$ (in sec/bit) are SP-device dependent constants. Regarding the price-based component, it captures how the perception of the quality improves (worsens) as the price falls (rises). As in [19], $Q_p(p_k)$ is modelled as

$$Q_p(p_k) = 1 - v_k p_k, \tag{4.5}$$

where $v_k > 0$ is an adjusting factor measured in $€^{-1}$. Particularly, v_k is the factor that determines the sensitivity of the perceived QoE to price variations for a user j and a service s_k. We assume that the value of v_k and hence $Q_p(p_k)$ can be different for each user j, in order to capture the effect of p_k on each user individually. As it can be observed in (4.5), if the user does not pay for the service (i.e. $p_k = 0$), the price-based component reaches the maximum value, $Q_p(0) = 1$, thereby increasing the perceived QoE in (4.3). That is, the more a user pays for a service, the higher the expectations on the received quality are.

It should be pointed out that $Q_{k_j}^{tg}$ corresponds to the QoE level a user j wants to perceive, it is constant, and does not depend on the service price. Hence, the MNO's objective is to offer user j a service with $Q_j^{kd} = Q_{k_j}^{tg}$ by providing the required QoS and corresponding price combination.

Fairness

MNOs aim to offer fairness among users both when the available resources suffice to provide them all with the target QoE (i.e. $Q_j^{kd} = Q_k^{tg}, \forall j \in \mathcal{U}$) and when not all of them can be appropriately served (i.e. $Q_j^{kd} < Q_k^{tg}$ for some users). In this contribution, we adopt the well-known Jain's fairness index [100] of the users' satisfaction level $\sigma_{ij}(t)$ as the QoE-fairness metric. Hence, at a specific time period t a BS i's Jain's index can be

expressed as

$$J_i(t) = \frac{\left(\sum_{j\in\mathcal{U}_i} \sigma_{ij}(t)\right)^2}{|\mathcal{U}_i| \sum_{j\in\mathcal{U}_i} \sigma_{ij}^2(t)} \in [0,1], \tag{4.6}$$

where $|\mathcal{U}_i|$ denotes the cardinality of $\mathcal{U}_i$ (i.e. the number of users served by BS i).

It should be noted that in Section 4.5.1 we imposed a fairness constraint that demands equal satisfaction for all the served users (i.e. $J_i(t) = 1$). In detail, we considered that for the set of users served by BS i, QoE fairness is achieved if $\sigma_{ij}(t) = \sigma_{in}(t)$ for any $j, n \in \mathcal{U}_i$ with service profiles $\pi_{k_j} = (s_{k_j}, q_{k_j}, p_{k_j})$ and $\pi_{k_n} = (s_{k_n}, q_{k_n}, p_{k_n})$, and devices $d_j, d_n \in \mathcal{D}$, respectively. When b_i is not enough to offer $\sigma_{ij}(t) = 1, \forall j \in \mathcal{U}_i$, all $\sigma_{ij}(t)$ are decreased (by reducing $w_{ij}(t)$) until $\sigma_{ij}(t) = \sigma_{in}(t)\ \forall j, n \in \mathcal{U}_i$. If $\sigma_{ij}(t) = \sigma_{in}(t)$ is only true for the trivial solution (i.e. $\sigma_{ij}(t) = 0$), users with $\sigma_{ij}(t) = 0$ are dropped (i.e. $w_{ij}(t) = 0$).

Overall User Satisfaction

We observe in (4.6) that $J_i(t)$ depends on the standard deviation of the users' satisfaction in BS i; the lower the standard deviation, the higher the $J_i(t)$. Hence, a high $J_i(t)$ can be achieved even when both the average and the standard deviation of the users' satisfactions ($\sigma_{ij}(t)$) are low. This means that high fairness does not guarantee high QoE for the users. To this end, we define the Overall User Satisfaction (OS) in a BS i as the sum of of the satisfaction of all the users connected to i, $OS_i(t) = \sum_{j\in\mathcal{U}_i} \sigma_{ij}(t)$. Similarly, the total OS in the system is defined as the aggregate satisfaction of all the users, $OS(t) = \sum_{i\in\mathcal{B}} OS_i(t)$. Thus, the objective of the MNO is to achieve high overall satisfaction (OS_i) and high Jain's index (J_i) values $\forall i \in \mathcal{B}$.

Since the number of users connected to a BS and the satisfaction of each user depends on their spectral efficiency and on the allocation of resources, the maximum overall user satisfaction varies along time. Let us define, for a given time interval t, the maximum achievable Overall User Satisfaction at BS i as $OS_i^{max}(t) = \max_{w_{ij}(t)} \{OS_i(t)\}$. Based on the definition, the Overall User Satisfaction achieved with a specific resource allocation can be expressed as a fraction of the maximum value. Therefore, we define the relative overall user satisfaction $\phi_i(t) = OS_i(t)/OS_i^{max}(t) \in [0,1]$ as a QoE-aware performance metric. The objective of the MNO is then given by

$$\phi_i(t) \geq \phi^{min}, \forall i \in \mathcal{B}, \tag{4.7}$$

where ϕ^{min} is a minimum threshold defined by the MNO.

4.3.2 MNO Economic Profit

The objective of the MNO is the maximization of the profit while satisfying the QoE required by the users. Specifically, the total profit $P(t)$ of the MNO is the sum of the individual profits of each BS $P_i(t)$, i.e. $P(t) = \sum_{i\in\mathcal{B}} P_i(t)$. In [28], $P_i(t)$ is expressed as the revenue obtained from the traffic served at time t, $R_i(t)$, minus the cost incurred when serving the traffic, which depicts the bandwidth utilization cost, $CB_i(t)$. Therefore,

$$P(t) = \sum_{i\in\mathcal{B}} P_i(t) = \sum_{i\in\mathcal{B}} (R_i(t) - CB_i(t)),\ [€]. \tag{4.8}$$

The revenue of BS i, $R_i(t)$, is usually the price of the services paid by the users in $\mathcal{U}_i$. That is, $R_i(t) = \sum_{j\in\mathcal{U}_i} R_{ij}(t)$, where $R_{ij}(t) = p_{k_j}$ is the revenue paid by user j with an SP π_{k_j}, when connected to BS i at time period t for a duration of T seconds. With regard to $CB_i(t)$, it is a convex and increasing exponential function of the total resources used by BS i, $w_i(t) = \sum_{j\in\mathcal{U}_i} w_{ij}(t)$ [28], and for a duration of T seconds it can be written as

$$CB_i(t) = c_i e^{h_i w_i(t) b_i} T, \tag{4.9}$$

where c_i (in $€/sec$) and h_i (in MHz^{-1}) are adjusting factors that capture the differences in the operational cost of the different BSs (e.g. macrocells and small cells have different transmit power, maintenance cost, site rent, etc). Substituting (4.1) and (4.9) into (4.8), and denoting the SP of a generic user j as π_{k_j}, the profit of BS i at time period t with a duration of T seconds when $Q_j^{k_j d_j}(t) \in (Q_{k_j}^{drop}, Q_{k_j}^{tg}]$ is given by

$$\begin{aligned} P_i(t) &= \sum_{j\in\mathcal{U}_i} p_{k_j} - c_i e^{h_i \sum_{j\in\mathcal{U}_i} w_{ij}(t) b_i} T \\ &= T \left(\sum_{j\in\mathcal{U}_i^B} \frac{\varepsilon_{ij}(t) w_{ij}(t) b_i}{8} \theta_{k_j}^B \mathbb{1}(\sigma_{ij}(t) > 0) \right. \\ &\quad \left. + \sum_{j\in\mathcal{U}_i^t} \theta_{k_j}^t \mathbb{1}(\sigma_{ij}(t) > 0) - c_i e^{h_i \sum_{j\in\mathcal{U}_i} w_{ij}(t) b_i} \right), \end{aligned} \tag{4.10}$$

where $\mathbb{1}(\cdot)$ is the binary indicator function, which is equal to 1 if the condition is true and 0 otherwise. We use the binary indicator function in order to emphasize that a BS i will not receive revenue if the allocated resources to user j, $w_{ij}(t)$, do not suffice for a satisfactory service with $\sigma_{ij}(t) > 0$.

It can be seen in (4.10) that the profit is influenced by multifarious factors, such as the perceived QoE (which in turn depends on multiple factors), the cost, the radio resources usage, etc. The maximization of the profit involves all these factors.

4.4 Problem Formulation

As explained in the previous section, the MNO aims to maximize its profit, while satisfying the required QoE of all users. However, when not all users can be served with the required QoE due to network congestion, the MNO must ensure fairness among them and the highest possible QoE level. In the following, we present the formulation of the profit maximization problem for our contributions.

4.4.1 QoE-Aware User Association

It should be noted that for our contribution in user association we adopted a QoE-based charging policy where the price (and revenue) is reduced when satisfaction is below 1, as proposed in [19]. In other words, the service price is reduced when $\sigma_{ij}(t) < 1$. Thus, based on [19], for a user j served by BS i and with a SP π_{k_j}, the BS i revenue is given by

$$R_{ij}(t) = \sigma_{ij}(t) \cdot p_{k_j} \tag{4.11}$$

Let us define the association of user j to BS i at time period t as $x_{ij}(t)$, where $x_{ij}(t) = 1$ if user j is served by BS i and $x_{ij}(t) = 0$ otherwise. In order to capture the impact of dynamic traffic demand and channel conditions, we assume that the users are mobile within the area under study and maximize the profit for a period of N_S subframes of T seconds duration. Taking the above into consideration, the user association problem for profit maximization is formulated based on (4.10) and (4.11) as

$$\max_{x_{ij}, w_{ij}, i\in\mathcal{B}, j\in\mathcal{U}} \sum_{t=1}^{N_S} P(t) = \sum_{t=1}^{N_S}\sum_{i\in\mathcal{B}}\sum_{j\in\mathcal{U}} x_{ij}(t)\sigma_{ij}(t)p_{k_j}T - \sum_{t=1}^{N_S}\sum_{i\in\mathcal{B}} c_i e^{h_i b_i \sum_{j\in\mathcal{U}} x_{ij}(t)w_{ij}(t)}T, \tag{4.12}$$

$$s.t. \quad \sum_{i\in\mathcal{B}} x_{ij} \leq 1,\ \forall i \in \mathcal{B},\ \forall j \in \mathcal{U}, \tag{4.12a}$$

$$w_i \in [0,1],\ \forall i \in \mathcal{B}, \tag{4.12b}$$

$$\sigma_{ij} = \sigma_{in},\ \forall i \in \mathcal{B},\ \forall j, n \in \mathcal{U}_i, \tag{4.12c}$$

In the optimization problem, users cannot be connected to more than a single BS (4.12a), the maximum bandwidth allocated by BS i is b_i, that is $\sum_{j\in\mathcal{U}_i} w_{ij}(t) = w_i(t) \leq 1$ (4.12b), and QoE fairness must be guaranteed (4.12c). Due to the binary nature of the association variable $x_{ij}(t)$, this maximization problem is a Mixed Integer Non-Linear Problem (MINLP) and cannot be solved in polynomial time (i.e. it is NP-hard).

4.4.2 QoE-Aware Resource Allocation and Dynamic Pricing

Given a particular association of users in the system's BSs, the maximization of the total MNO profit is equivalent to the maximization of each BS i's profit individually through resource allocation (and dynamic pricing when applied). Hence, the BS i's resource allocation and dynamic pricing problem for profit maximization at time t is formulated based on (4.10) as

$$\max_{w_{ij},\lambda_{ij},j\in\mathcal{U}_i} P_i(t) = T\left(\sum_{j\in\mathcal{U}_i^B}\frac{\varepsilon_{ij}(t)w_{ij}(t)b_i}{8}\lambda_{ij}\theta_{k_j}^B\mathbb{1}(\sigma_{ij}(t)>0)\right. \tag{4.13}$$

$$\left. + \sum_{j\in\mathcal{U}_i^t}\lambda_{ij}\theta_{k_j}^t\mathbb{1}(\sigma_{ij}(t)>0) - c_i e^{h_i\sum_{j\in\mathcal{U}_i} w_{ij}(t)b_i}\right),$$

$$s.t. \quad w_i \in [0,1],\ \forall i \in \mathcal{B}, \tag{4.13a}$$

$$J_i(t) \geq J^{min},\ \forall i \in \mathcal{B}, \tag{4.13b}$$

$$\phi_i(t) \geq \phi^{min}. \tag{4.13c}$$

In the optimization problem, the maximum bandwidth allocated by BS i is b_i, that is $\sum_{\forall j\in\mathcal{U}_i} w_{ij}(t) = w_i(t) \leq 1$ (4.13a), and QoE fairness must be guaranteed for a minimum Jain's index value, J^{min} (4.13b). Finally, the relative overall user satisfaction must be higher than the minimum threshold ϕ^{min} (4.13c). Since we use the binary indicator function in $P_i(t)$, the optimization problem in (4.13) is a MINLP problem, whose computational complexity is NP-hard [101].

It should be noted that for our contribution dedicated solely to the resource allocation problem wE do not consider dynamic pricing. Hence, for the solution of this problem the pricing factor λ_{ij} is stable and always equal to 1. Regarding the formulation of the joint resource allocation and dynamic pricing problem we did not impose a fairness condition. Thus, the constraint in (4.13b) is not considered for the solution of the problem in (4.13).

4.5 QoE-Aware Algorithms

As it was shown in the previous section, the formulated problems are described by NP-hardness. In order to address the high complexity of their solution, in this section we present low-complexity, greedy, heuristic algorithms, which we proposed in our contributions.

Algorithm 2: QoE-Aware user Association Algorithm

Create $\mathcal{A}(t)$ as the set of users with $\sigma_{ij}(t-1) < 1$ or a NLOS channel with serving BS.
if $\mathcal{A}(t) \neq \emptyset$ **then**
  for $j \in \mathcal{U} \setminus \mathcal{A}(t)$ **do**
    $x_{ij}(t) = x_{ij}(t-1), \forall i \in \mathcal{B}$
  end
  Initialize $\overline{w_i}(t) = w_i(t-1)$; $\overline{P_i}(t) = P_i(t-1)$; $\overline{R_{ij}}(t) = R_{ij}(t-1)$; $\forall i \in \mathcal{B}$, $\forall j \in \mathcal{U}_i$
  while $\mathcal{A}(t) \neq \emptyset$ **do**
    Select user j randomly from $\mathcal{A}(t)$
    $m = \arg\max_{i \in \mathcal{B}} x_{ij}(t-1)$
    $\overline{P_m}(t) = \overline{P_m}(t) - \overline{R_{mj}}(t) + + c_m e^{h_m \overline{w_m}(t) b_m} \left(1 - e^{-h_m w_{mj}(t-1) b_m}\right)$
    $\overline{w_m}(t) = \overline{w_m}(t) - w_{mj}(t-1)$
    for $i \in \mathcal{B}$ **do**
      $\widehat{w_{ij}}(t) = \min\left(\frac{r_{kd}}{\varepsilon_{ij}(t) b_i}, 1 - \overline{w_i}(t)\right)$
      Calculate $\widehat{\sigma_{ij}}(t), \widehat{R_{ij}}(t)$ according to (4.2),(4.11) for $\widehat{w_{ij}}(t)$
      $\widehat{P_i}(t) = \overline{P_i}(t) + \widehat{R_{ij}}(t) - - c_i e^{h_i \overline{w_i}(t) b_i} (e^{h_i \widehat{w_{ij}}(t) b_i} - 1)$
      $\widehat{w_i}(t) = \overline{w_i}(t) + \widehat{w_{ij}}(t)$
    end
    if $\exists i \in \mathcal{B}$ *such that* $\widehat{\sigma_{ij}}(t) = 1$ **then**
      $m = \arg\max_{i \in \mathcal{B}: \widehat{\sigma_{ij}}(t) = 1} \{\widehat{P_i}(t) + \sum_{v \in \mathcal{B} \setminus \{i\}} \overline{P_v}(t)\}$
      $x_{mj}(t) = 1$; $x_{vj}(t) = 0, \forall v \in \mathcal{B} \setminus m$
      $\overline{w_m}(t) = \widehat{w_i}(t)$; $\overline{R_{mj}}(t) = \widehat{R_{mj}}(t)$; $\overline{P_m}(t) = \widehat{P_m}(t)$
    else if $\exists i \in \mathcal{B}$ *such that* $\widehat{\sigma_{ij}}(t) \in (0,1)$ **then**
      $m = \arg\max_{i \in \mathcal{B}: \widehat{\sigma_{ij}}(t) \in (0,1)} \{\widehat{P_i}(t) + \sum_{v \in \mathcal{B} \setminus \{i\}} \overline{P_v}(t)\}$
      $x_{mj}(t) = 1$; $x_{vj}(t) = 0, \forall v \in \mathcal{B} \setminus m$
      $\overline{w_m}(t) = \widehat{w_i}(t)$; $\overline{R_{mj}}(t) = \widehat{R_{mj}}(t)$; $\overline{P_m}(t) = \widehat{P_m}(t)$
    else
      $x_{vj}(t) = 0, \forall v \in \mathcal{B}$
    end
    $\mathcal{A}(t) = \mathcal{A}(t) \setminus \{j\}$
  end
end

4.5.1 QoE-Aware User Association

For the solution of the profit maximization problem in (4.12), we proposed a greedy, low complexity algorithm, $O(n^2)$, which is presented in Algorithm 2. This algorithm takes as input the user's SP-device pair (π_{k_j}, d_j), as well as the BSs' state in the previous subframe (i.e. $w_{ij}(t-1), R_{ij}(t-1)$, $\forall j \in \mathcal{U}_i, \forall i \in \mathcal{B}$), to maximize the MNO profit through the user association. At the beginning of the subframe t, the algorithm creates the set $\mathcal{A}(t)$ with all the users that, being served by a BS, had a satisfaction below 1 at

time $t-1$, i.e. $\sigma_{ij}(t-1) < 1$, and the users that consumed a lot of resources due to a Non-Line-Of-Sight (NLOS) connection[2] (line 1). If the set is not empty, all users with $\sigma_{ij}(t-1) = 1$ are associated to the same BS they were associated with at time $t-1$ (line 4). For the association of the rest of the users, the algorithm makes use of the estimates of $w_i(t)$, $P_i(t)$ and $R_{ij}(t)$, denoted as $\overline{w_i}(t)$, $\overline{P_i}(t)$ and $\overline{R_{ij}}(t)$.

Users in $\mathcal{A}(t)$ are selected randomly, one by one, in order to avoid the prioritization of users during the association procedure, and hence guarantee fairness . Each user's contribution to the estimated BS profit, where the user was connected to at time $t-1$ is subtracted (lines 8-11). The algorithm estimates the resources needed/available in each BS, $\widehat{w_{ij}}(t)$, the expected revenue, $\widehat{R_{ij}}(t)$, and the expected satisfaction, $\widehat{\sigma_{ij}}(t)$ (line 13). Then, for all BSs, if there are BSs that provide satisfaction equal to 1, the user will be associated -in a greedy manner- to the BS that maximizes the MNO profit while $\widehat{\sigma_{ij}}(t) = 1$ (lines 18-21). If there are not BSs that could provide $\widehat{\sigma_{ij}}(t) = 1$, but $\widehat{\sigma_{ij}}(t) > 0$, the user will be associated with the BS that maximizes the MNO profit with $\widehat{\sigma_{ij}}(t) > 0$ (lines 22-25). The rest of users are not associated to any BS, since there are not enough resources in the neighbouring BSs or the channels between the user and the BSs are in outage (line 27). The procedure is repeated for all users in $\mathcal{A}(t)$.

4.5.2 QoE-Aware Resource Allocation

As it can be observed in (4.13), the resource allocation for profit optimization is constrained by minimum fairness and overall user satisfaction values. In this section we analyse the interaction between satisfaction, fairness and profit, and propose a greedy, heuristic, QoE-aware resource allocation algorithm for profit maximization.

Based on (4.13), the MNO profit can be maximized by reducing the cost and/or increasing the revenue. However, it is noteworthy that for a given association of users to BS i, $\mathcal{U}_i$, the bandwidth utilization cost depends on the total amount of resources allocated to users, regardless of how they are distributed among the users. Therefore, the utilization cost ($CB_i(t)$) is fixed for a given number of total resources ($w_i(t)$).

In turn, (4.1) shows that the revenue presents a differentiated behaviour for time-based charged services and data-based charged services. Whereas the revenue generated by a user with a time-based charged service remains constant as long as the connection is not dropped, the revenue generated by users with a data-based charged service increases with the amount of transferred data. In other words, after providing time-based charged

[2]In a mmWave link, an NLOS connection demands significantly more spectrum resources than an LOS connection in order to achieve the same QoS/QoE [99]. Hence, it is reasonable to re-associate a user with an NLOS connection to a different BS an LOS links in order to achieve more efficient spectrum resource usage.

users with the minimum amount of resources required to guarantee that the perceived QoE is above the minimum QoE level, the revenue can be increased by allocating the rest of resources to data-based charged users. Note that in terms of QoE the aforementioned resources allocation strategy is translated into low satisfaction of time-based charged users (their $\sigma_{ij}(t)$ is low but above 0) and higher satisfaction of data-based charged users. These differences between the two types of services lead to resources allocation unfairness (i.e. low Jain's index values, $J_i(t)$) and could result in low relative overall user satisfaction ($\phi_i(t)$). This is the reason why (4.13b) and (4.13c) impose minimum fairness and minimum relative overall user satisfaction levels, respectively, and this is also the reason why the heuristic algorithm presented in the sequel takes both aspects into account. This analysis shows that there is a trade-off between the MNO profit and the network performance, as the optimization of one of the two objectives comes at the expense of the other.

In order to overcome the complexity of the maximization problem stated in (4.13), we propose a low-complexity resource allocation algorithm, $O(n^2)$, presented in Algorithm 3, that maximizes the MNO profit for a minimum fairness and OS level (i.e. J^{min} and ϕ^{min}). This greedy, Profit Maximizing resource allocation algorithm (referred to as PM) takes as input the users associated to a BS at period t, and each user's SP-device pair (p_k, d).

PM is divided into three parts. In the first part, PM determines the resource allocation that maximizes BS i's Overall Satisfaction $OS_i(t)$ (steps 1-15). Subsequently, using as input the resource allocation obtained from the first part of the algorithm, PM determines the resource allocation that maximizes $P_i(t)$, while satisfying $\phi_i(t) \geq \phi^{min}$ (steps 16-29). Finally, the profit maximizing resources allocation determined in the previous step is iteratively modified until the minimum overall user satisfaction and the minimum fairness constraints are satisfied (steps 30-32).

The first part of Algorithm 3 (steps 1-15) is a greedy algorithm that aims to determine the resources allocation (i.e. the value of $w_{ij}(t)\ \forall j \in \mathcal{U}_i$) that maximizes the OS_i. The algorithm initially computes the resources required to meet the maximum user satisfaction for each user. By substituting (4.2) in (4.3), $w_{ij}(t)$ can be expressed as

$$\begin{aligned} w_{ij}(t) &= \frac{1}{\varepsilon_{ij}(t) b_i \beta_{k_j d_j}} \Big[r_{kd} \beta_{k_j d_j} \\ &+ \log\left(\frac{\sigma_{ij}(t)(Q_{k_j}^{tg} - Q_{k_j}^{drop}) + Q_{k_j}^{drop}}{\alpha_{k_j d_j} Q_p(p_{k_j})} - \frac{\gamma_{k_j d_j}}{\alpha_{k_j d_j}} \right) \Big] \end{aligned} \tag{4.14}$$

Subsequently, the algorithm assigns resources iteratively to the user $j \in \mathcal{U}_i$ with the least resource requirements. Particularly, each user is allocated the resources calculated

Algorithm 3: Profit Maximizing RA Algorithm (PM)

1 Set $\mathcal{U}_i' = \mathcal{U}_i$
2 Compute $w_{ij}(t) \leq 1$ to maximize $\sigma_{ij}(t)$ for each $j \in \mathcal{U}_i$ using (4.14)
3 $w_i(t) = 0$
4 **while** $\mathcal{U}_i' \neq \emptyset$ **do**
5 Find user $j \in \mathcal{U}_i'$ with $\min(w_{ij}(t))$
6 **if** $w_{ij}(t) \leq 1 - w_i(t)$ **then**
7 $w_i(t) = w_i(t) + w_{ij}(t)$
8 **else if** $\sigma_{ij}(t) \geq \sigma_{ij}^{min}$ *for* $w_{ij}(t) = 1 - w_i(t)$ **then**
9 $w_{ij}(t) = 1 - w_i(t)$ and $w_i(t) = 1$
10 **else**
11 $w_{ij}(t) = 0$
12 **end**
13 $\mathcal{U}_i' \leftarrow \mathcal{U}_i' - \{j\}$
14 **end**
15 Set $OS_i^{max}(t) = OS_i(t)$, $\phi_i(t) = 1$ and $\sigma_{step} = \sigma_{step}^{max}$
16 **while** $\phi_i(t) \geq \phi^{min}$ **and** $\sigma_{step} \geq \sigma_{step}^{min}$ **do**
17 **forall** $j \in \mathcal{U}_i$ **do**
18 $\sigma_{ij}^{-}(t) = \max\left(\sigma_{ij}(t) - \sigma_{step}, 0\right)$
19 $\sigma_{ij}^{+}(t) = \min\left(\sigma_{ij}(t) + \sigma_{step}, 1\right)$
20 Calculate the total profit $P_i(t)$ and $\phi_i(t)$ with $\sigma_{ij}(t) = \sigma_{ij}^{-}(t)$ and $\sigma_{ij}(t) = \sigma_{ij}^{+}(t)$
21 Store the maximum profit $P_i(t)$ s.t. $\phi_i(t) \geq \phi^{min}$ and $w_i(t) \leq 1$ in $P_{ij}'(t)$ and the corresponding $\phi_i(t)$ and $\sigma_{ij}(t)$ in $\phi_{ij}'(t)$ and $\sigma_{ij}'(t)$
22 **end**
23 $j^* = \arg\max_j \left(P_{ij}'(t)\right)$
24 **if** $P_{ij^*}'(t) > P_i(t)$ **then**
25 $P_i(t) = P_{ij^*}'(t)$, $\phi_i(t) = \phi_{ij^*}'(t)$, $\sigma_{ij^*}(t) = \sigma_{ij^*}'(t)$ and the corresponding $w_{ij}(t)$ values are updated
26 **else**
27 Reduce σ_{step}
28 **end**
29 **end**
30 Calculate Jain's index $J_i(t)$ and set $\sigma_{step} = \sigma_{step}^{max}$
31 Repeat steps 16-29 while $J_i(t) < J^{min}$
32 Select the maximum $P_{ij}'(t)$ in 23 that reduces $|J_i(t) - J^{min}|$

previously until the total available resources are depleted. In case of resources depletion, the resources allocated to the remaining users are reduced as long as the user satisfaction is above the minimum acceptable satisfaction value σ_{ij}^{min} or are set to 0 otherwise. With this, the overall user satisfaction is maximized, but fairness is not taken into account.

The second part of PM (steps 16-29) is a greedy, profit maximization algorithm on the basis of the resources allocation resulted from the first part of the algorithm (steps 1-15). Specifically, for each user the user satisfaction is decreased (step 18) and increased

(step 19) with σ_{step} (the resources allocation $w_{ij}(t)$ is calculated from (4.14)), and then the profit is calculated for both satisfaction values. Only $\sigma_{ij}(t)$ values that increase the profit are considered as feasible results. This procedure is repeated iteratively for each user and for different values of σ_{step} as long as the relative overall user satisfaction is above the minimum threshold. The resources allocation is updated with the distribution of resources that provides the maximum profit for a relative user satisfaction level above ϕ^{min}, that is, the optimal solution in a single iteration. Therefore, the resulting resources allocation of the second part of the PM algorithm converges to a local optimal solution of the profit maximization problem for a given minimum satisfaction, which at times can be a global maximum [102].

The last part of the PM algorithm (steps 30-32) introduces the fairness. Thus, starting from the resources allocation obtained in the second part of the algorithm, PM executes the same greedy process run in the second part (steps 16-29) but this time only solutions that converge to J^{min} are selected (i.e. with declining $|J_i(t) - J^{min}|$ values). This last part, and hence PM, is terminated when there are no alternative resource allocations that satisfy the relative overall satisfaction threshold and the convergence of $J_i(t)$ to J^{min}.

It should be noted that PM consists of three greedy iterative algorithms, which in each of their iterations make the optimal decision for a subproblem (e.g. maximize the OS_i or P_i by allocating resources to a single user at a time). However, it is probable that it converges to locally optimal solutions instead of the global optimum [102]. Thus, PM may not always perform optimally, which we will examine in Section 4.6.2.

4.5.3 Dynamic Pricing

As explained in Section 4.3.1, the service price is one of the key factors that affects the users' QoE perception. Particularly, by observing expressions (4.3), (4.4) and (4.5), we notice that when p_k is lowered (increased), the QoE's price-based component ($Q_p(p_k)$) increases (decreases). For instance, when the charging is lowered to $p'_k < p_k$, Q_p increases (i.e. $Q_p(p'_k) > Q_p(p_k)$). Thus, a particular target QoE level Q_k^{tg} can be achieved by offering the same service with lower price and rate $r'_j < r_{kd}$, that is, $Q_k^{tg} = \widehat{Q}_j^{kd}(r_{kd})Q_p(p_k) = \widehat{Q}_j^{kd}(r'_j)Q_p(p'_k)$. Conversely, when p_k is raised, it is not possible to achieve the highest QoE levels, even when the QoS reaches its peak (i.e. $r_j = r_{kd}$).

Therefore, reducing p_k (i.e. $\lambda_{ij} \in [0, 1)$) allows for the reduction of a user's rate $r_j(t)$ without lowering her satisfaction, which in turn reduces her resource utilization $w_{ij}(t)$. Thus, our proposal on dynamic pricing is the use of price reduction in order to release spectrum resources from specific users. These released resources can be then used on

other users to improve the OS during congestion. Moreover, our proposal on dynamic pricing can be used to increase $P_i(t)$. Particularly, when this scheme is used, we decrease both $R_{ij}(t)$ and $CB_i(t)$. Hence, when the cost reduction is higher than the revenue loss, $P_i(t)$ becomes higher. Finally, this form of dynamic pricing can be applied on the existing pricing schemes an MNO uses along with the employed resource allocation scheme.

In order to apply our proposal on dynamic pricing while solving the profit maximization problem in (4.13), we propose a greedy, low complexity algorithm $O(n^3)$, presented in Algorithm 4. This algorithm takes as input the BS i's users' SP-device pair (p_{k_j}, d_j), to maximize the MNO profit for a minimum relative OS level ϕ^{min} through resource allocation and dynamic pricing.

Similar to Algorithm 3, the rationale behind Algorithm 4 is to initially determine the resource allocation that maximizes BS i's $OS_i(t)$. Subsequently, using as input this allocation, the algorithm determines the resource allocation and pricing (i.e. λ_{ij}) that maximize $P_i(t)$ for the required OS constraint (i.e. $\phi_i(t) \geq \phi^{min}$).

Initially, Algorithm 4 calculates the $w_{ij}(t)$ needed to serve each user with the maximum $\sigma_{ij}(t)$ (step 1). Subsequently, it allocates resources starting from the user with the least resource requirements towards the user with the highest (steps 3-11). Finally, either all users are satisfied (i.e. $OS_i = N_U$) or the resources are depleted. Next, the algorithm sets $OS_i^{max}(t) = OS_i(t)$, $\phi_i(t) = 1$, and the discrete set $\Lambda = \{0, 0.1, \ldots, 1\}$ of λ_{ij} values that BS i can use for pricing.

In the following iterative procedure (steps 13-28), each user's satisfaction is decreased (step 15) and increased (step 16) with σ_{step} (the resource allocation $w_{ij}(t)$ is calculated from (4.14)). For both satisfaction values, we reduce p_k for the λ_{ij} values in Λ, and calculate the corresponding profit (steps 14-21). Only $\sigma_{ij}(t)$ values that increase the profit are considered as feasible results. This procedure is repeated iteratively for each user and for different values of σ_{step} as long as the relative overall user satisfaction is above the minimum threshold. The resource allocation and pricing are updated with the distribution of resources and the λ_{ij} values that provide the maximum profit for a relative user satisfaction level above ϕ^{min}.

It should be noted that for our contribution in dynamic pricing we did not consider a fairness constraint. As in Algorithm 3, when Algorithm 4 is applied for $\phi^{min} = 1$, the corresponding Jain's fairness index for the served users is always the maximum possible (as it will be shown in Section 4.6.2). This occurs due to the fact that when $\phi^{min} = 1$ both algorithms skip their second (and third) part, since the overall user satisfaction achieved in the first part cannot be reduced. Thus, all of the served users receive a satisfaction $\sigma_{ij}(t) = 1$, with the possible exception of a single user receiving $\sigma_{ij}(t) \in (0, 1)$

Algorithm 4: Profit Maximizing joint RA-DP Algorithm

Calculate $w_{ij}(t) \leq 1$ for maximum $\sigma_{ij}(t), \forall j \in \mathcal{U}_i$
$w_i(t) = 0$
forall $j \in \mathcal{U}_i$ *in ascending order of* $w_{ij}(t)$ **do**
  if $w_{ij}(t) \leq 1 - w_i(t)$ **then**
    $w_i(t) = w_i(t) + w_{ij}(t)$
  else if $\sigma_{ij}(t) \geq \sigma^{min}$ *for* $w_{ij}(t) = 1 - w_i(t)$ **then**
    $w_{ij}(t) = 1 - w_i(t)$ and $w_i(t) = 1$
  else
    $w_{ij}(t) = 0$
  end
end
Set $OS_i^{max}(t) = OS_i(t)$, $\phi_i(t) = 1$, $\sigma_{step} = \sigma_{step}^{max}$ and $\Lambda = \{0, 0.1, \ldots, 1\}$
while $\phi_i(t) \geq \phi^{min}$ **and** $\sigma_{step} \geq \sigma_{step}^{min}$ **do**
  for $j \in \mathcal{U}_i$ **do**
    $\sigma_{ij}^{-}(t) = \max\left(\sigma_{ij}(t) - \sigma_{step}, 0\right)$
    $\sigma_{ij}^{+}(t) = \min\left(\sigma_{ij}(t) + \sigma_{step}, 1\right)$
    for $\lambda_{ij} \in \Lambda$ **do**
      Calculate the total profit $P_i(t)$ and $\phi_i(t)$ with $\sigma_{ij}(t) = \sigma_{ij}^{-}(t)$ and $\sigma_{ij}(t) = \sigma_{ij}^{+}(t)$
      Store the maximum profit $P_i(t)$ s.t. $\phi_i(t) \geq \phi^{min}$ and $w_i(t) \leq 1$ in $P'_{ij}(t, \lambda_{ij})$ and the corresponding $\phi_i(t)$ and $\sigma_{ij}(t)$ in $\phi'_{ij}(t, \lambda_{ij})$ and $\sigma'_{ij}(t, \lambda_{ij})$
    end
  end
  $(j^*, \lambda_{ij}^*) = \arg\max_{j, \lambda_{ij}} \left(P'_{ij}(t, \lambda_{ij})\right)$
  if $P'_{ij^*}(t, \lambda_{ij}^*) > P_i(t)$ **then**
    $P_i(t) = P'_{ij^*}(t, \lambda_{ij}^*)$, $\phi_i(t) = \phi'_{ij^*}(t, \lambda_{ij}^*)$, $\sigma_{ij^*}(t) = \sigma'_{ij^*}(t, \lambda_{ij}^*)$ and the corresponding $w_{ij}(t), \lambda_{ij}(t)$ values are updated
  else
    Reduce σ_{step}
  end
end

(hence $J_i(t) = 1$ or $J_i(t) \cong 1$). Therefore, the use of a fairness algorithm component as in Algorithm 3 was deemed unnecessary, since this contribution's objective is the dynamic pricing and not the fairness-OS-profit trade-off, which we analyse thoroughly in our contribution dedicated solely to resource allocation.

4.5.4 Feasibility

The feasibility of the above algorithms depends on the availability of $w_{ij}(t-1)$, $R_{ij}(t-1)$ and the information needed to calculate $\widehat{\sigma_{ij}}(t)$ and $\widehat{R_{ij}}(t)$ (for Algorithm 2), the users' SINR (for the calculation of their spectral efficiency $\varepsilon_{ij}(t)$), as well as monitoring the

actual user rate $r_j(t)$. In LTE-A, the SINR is calculated by the device and sent to the BS over the Physical Uplink Control Channel (PUCCH) or the Physical Uplink Shared Channel (PUSCH) [93]. In LTE-A, real-time monitoring of the User Equipment (UE) application layer data throughput performance is used for measuring the provided QoS [103]. The necessary information for $\widehat{\sigma_{ij}}(t)$ and $\widehat{R_{ij}}(t)$ can be obtained in real time with a module such as the Policy and Charging Control (PCC) in LTE-A, which can control the QoS on a per service data flow, apply different charging models, as well as control usage monitoring to make dynamic policy decisions [104]. Moreover, such a module would be responsible and able to apply our diverse pricing scheme as well as our proposal on dynamic pricing.

4.6 Performance Evaluation

In this section, we evaluate the performance of each algorithm (in dedicated subsections) presented in Section 4.5, which we proposed for the solution of the profit maximization problems in Section 4.4.

4.6.1 QoE-Aware User Association

4.6.1.1 Scenario description and parameters

The scenario used for the performance evaluation consists of a cluster with 4 mmWave small cells deployed in the coverage area of a macrocell eNB sector. The cluster is square shaped and centred at location $c = (x_c, 0)$, as shown in the layout depicted in Fig. 4.2. Along simulations, x_c is randomly selected according to a uniform distribution with $x_c \in [100, 190]m$. The mmWave channel is modelled as a three-states channel[99], with LOS, NLOS and outage states. Although high directivity of antennas compensates partially the path loss, the probability of LOS communications falls rapidly as the distance between transmitter and receiver increases. In the scenario we define R_{sc} as the distance at which the probability of having a LOS communication is 0.55. According to [99], $\mathbb{P}_{LOS}(R_{sc}) = 0.55$ holds for R_{sc}=40m in the 28GHz band. In the sequel the Inter-Site Distance (ISD) between small cells, R_{ISD}, is expressed as a multiple of R_{sc}, that is, $R_{ISD} = nR_{sc}$, with $n \in \mathbb{N}$, as we aim to shed light on the impact of a varying small cell ISD.

Users are uniformly distributed within the cluster and move with a speed of 5 km/h. The SP of each user is selected with equal probability among the SPs defined in Table 4.2. As it can be observed in Table 4.2, three services are considered, each one with two QoE

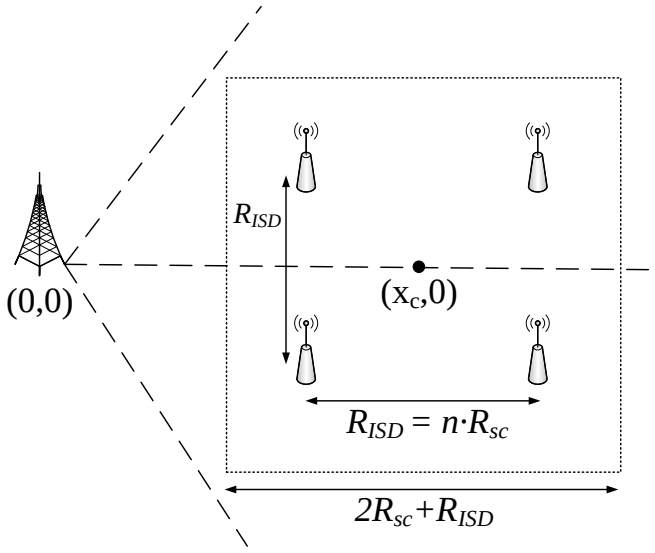


FIGURE 4.2: Simulation scenario topology

TABLE 4.2: Service Profiles' parameters

Service	**QoE class**	$\{\mathbf{r_{k1}}, \mathbf{r_{k2}}, \mathbf{r_{k3}}\}$ (Mbps)	$\theta_{\mathbf{k}}^{\mathbf{t}}$ **or** $\theta_{\mathbf{k}}^{\mathbf{B}}$
Service 1	Basic	80	$3 \cdot 10^{-4}$€/MB
(Data Based)	Premium	100	$4 \cdot 10^{-4}$€/MB
Service 2	Basic	$\{50, 65, 80\}$	3.5€/h
(Time Based)	Premium	$\{60, 75, 90\}$	5€/h
Service 3	Basic	$\{70, 85, 90\}$	3.5€/h
(Time Based)	Premium	$\{80, 100, 120\}$	5€/h

classes $\mathcal{Q} = \{\text{Basic, Premium}\}$: Service 1 is a data-based charged service, and Services 2 and 3 are time-based charged services. Likewise, 3 different devices are considered, and the corresponding transmission rates associated to each SP, r_{kd}, are also included in Table 4.2. Note that for each SP, r_{kd} is the transmission rate required to perceive a QoE equal to Q_k^{tg}. In the simulations, the transmission rate that results in a perceived QoE equal to Q_k^{drop} is set to $r_{kd}^{drop} = 0.7 r_{kd}$ for all SPs. Moreover, v_k is selected so as to have $Q_p(p_k) = 0.9$ in (4.5), and

$$\begin{cases} \alpha_{kd} = \frac{Q_k^{tg}}{Q_p(p_k)} - \gamma_{kd} & \text{(4.15a)} \\ \beta_{kd} = -\frac{1}{\Delta r_j^{drop}} \ln\left(\frac{Q_k^{drop} - \gamma_{kd} Q_p(p_k)}{Q_k^{tg} - \gamma_{kd} Q_p(p_k)}\right), & \text{(4.15b)} \end{cases}$$

where $\Delta r_j^{drop} = r_{kd} - r_{kd}^{drop}$, and $\gamma_{kd} = 1$, for all π_k, $d \in \mathcal{D}$ and $(Q_k^{tg}, Q_k^{drop}) = (3.5, 2.5)$ for the Basic QoE class of all services and $(Q_k^{tg}, Q_k^{drop}) = (4.5, 3.5)$ for the Premium QoE class of all services .

Parameters used for the BSs, both eNBs and small cells, are listed in Table 4.3. For the bandwidth allocation in the two tiers, we adopted the 5G configuration proposed by

TABLE 4.3: BS parameters

Parameter	**Macrocell**	**Small cell**
c_i (€)	$5 \cdot 10^{-6}$	$5 \cdot 10^{-6}$
h_i (MHz^{-1})	$39 \cdot 10^{-3}$	$16 \cdot 10^{-3}$
b_i (MHz)	200	500
Transmission Power (dBm)	43	37

a leading telecommunications vendor [105]. The 3GPP LTE-A channel model used for macrocells is described in [93] and for small cells (in the 28GHz band) in [99]. Antenna gains are set to 0 dB and, due to high antenna directivity in the mmWave band, small cells communications are assumed to be noise-limited. In conventional cellular networks, the cell selection is based on schemes that connect the users to the BS with the strongest signal [106]. Hence, in the following we compare the proposed QoE-aware Association Algorithm (denoted as Q-AAA) with a SINR-based cell selection algorithm (referred to as SINR-Alg). This means that when we apply the SINR-Alg the users are served by the BS with the highest SINR. It should be noted that the resource allocation in both algorithms satisfies the condition for QoE fairness (i.e. $\sigma_{ij}(t) = \sigma_{in}(t)$ for any $j, n \in \mathcal{U}_i$).

4.6.1.2 Comparison with SINR Algorithm

Fig. 4.3 shows the expected total utilization of the spectrum, which is defined as $W = \mathbb{E}\left[W(t) = \frac{\sum_{i\in\mathcal{B}} w_i(t) b_i}{\sum_{i\in\mathcal{B}} b_i}\right]$. As it can be observed, the proposed algorithm (Q-AAA) presents higher bandwidth utilization than the cell association algorithm based on the SINR (SINR-Alg) for both $R_{ISD} = 2R_{sc}$ and $R_{ISD} = 3R_{sc}$. As expected, SINR-Alg should always present the best spectrum efficiency, and consequently, the lowest bandwidth utilization, since users tend to use the most efficient Modulation and Coding Scheme (MCS). Interesting enough, a different behaviour can be noticed between the two algorithms regarding N_U and R_{ISD}. As it can be seen, for Q-AAA utilization increases with N_U, and is comparable regarding R_{ISD}. As R_{ISD} rises, the average distance between BSs and users grows as well. Thus, the probability of serving users with NLOS links raises, increasing W. The broad difference in W between the two R_{ISD} values observed for SINR-Alg can be explained with the help of Fig. 4.4.

Fig. 4.4 shows the percentage of time during which the users had a satisfaction above 0, i.e. $\sigma_{ij}>0$. For both algorithms, the users perceive $\sigma_{ij}>0$ for less time as N_U and R_{ISD} increase. As the system's bandwidth demands increase with both N_U and R_{ISD}, the system can accommodate a lower percentage of users. We further observe that Q-AAA achieves gains up to 25% and 81% for R_{ISD}=$\{2,3\}R_{sc}$ respectively. This can be explained by the fact that Q-AAA prioritizes users that can obtain an appropriate

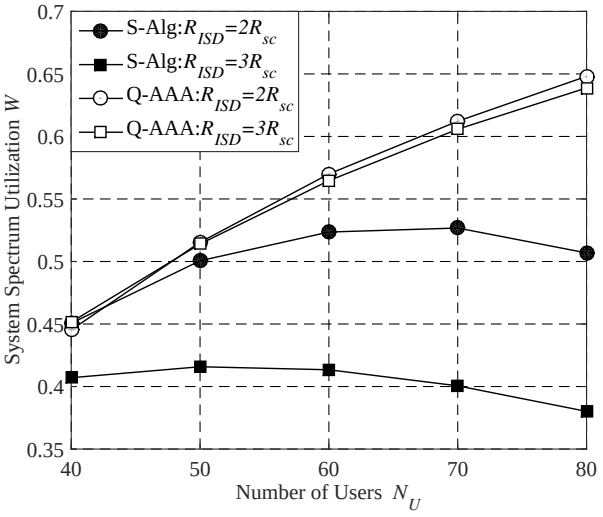


FIGURE 4.3: Bandwidth utilization (W)

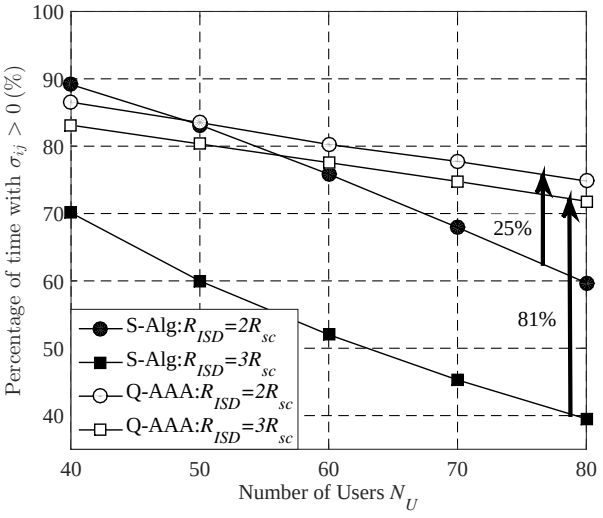


FIGURE 4.4: Percentage of time with a satisfaction above 0 ($\sigma_{ij} > 0$)

QoE over users that cannot, whereas SINR-Alg connects users without considering the available BS resources. Thus, SINR-Alg congests the BSs more frequently, dropping users to satisfy the QoE fairness condition.

Fig. 4.5 presents the empirical CDFs of σ_{ij}, for $N_U = \{40, 70\}$ and $R_{ISD} = \{2, 3\}R_{sc}$. As it can be observed, σ_{ij} diminishes as N_U and R_{ISD} increase and the system gets congested more frequently. Hence, the QoE is degraded more regularly, while the service time with $\sigma_{ij} > 0$ decreases as well (see Fig. 4.4). It can be deduced from Fig. 4.3-4.5 that Q-AAA results in less frequent congestion of the BSs, and provides higher QoE for longer time periods. The above explains why P increases with Q-AAA, whereas it decreases for the SINR-Alg not only with R_{ISD}, but also with N_U, as shown in Fig. 4.6.

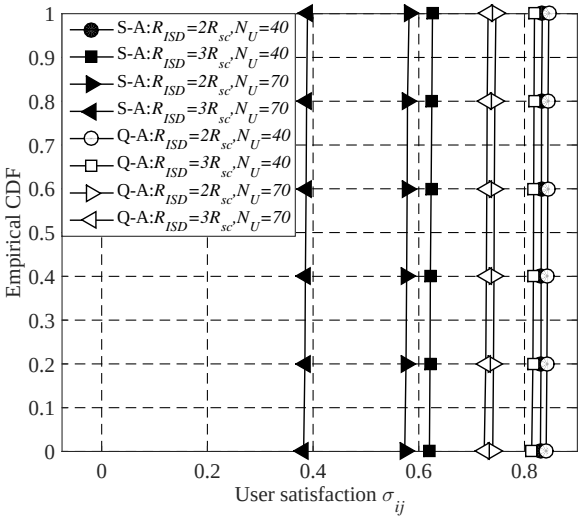


FIGURE 4.5: CDF of the user satisfaction

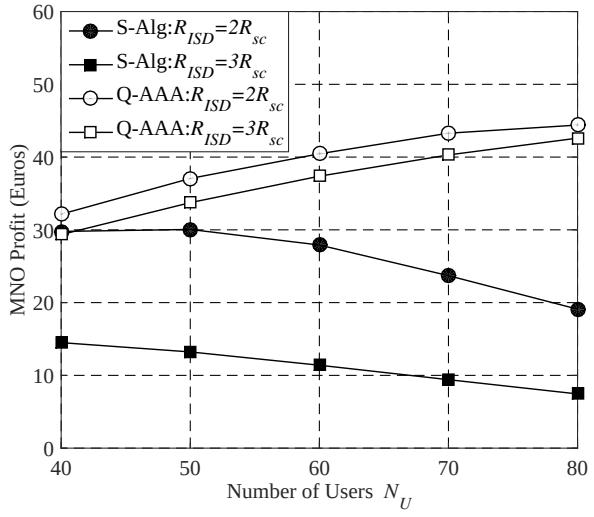


FIGURE 4.6: Total MNO profit (P)

This happens because as Q-AAA offers higher user satisfaction for longer time period, the MNO generates larger revenue, compensating the higher bandwidth utilization.

Our proposed algorithm, Q-AAA, manages to offer higher QoE to the users and profit to the MNO compared to the reference algorithm, because it bases its decisions on both the technological (i.e. QoS/QoE requirements) and economic (i.e. pricing and profit) context of the network. This scheme can guarantee the sustainability of a network, providing the incentives for adoption in future 5G deployments.

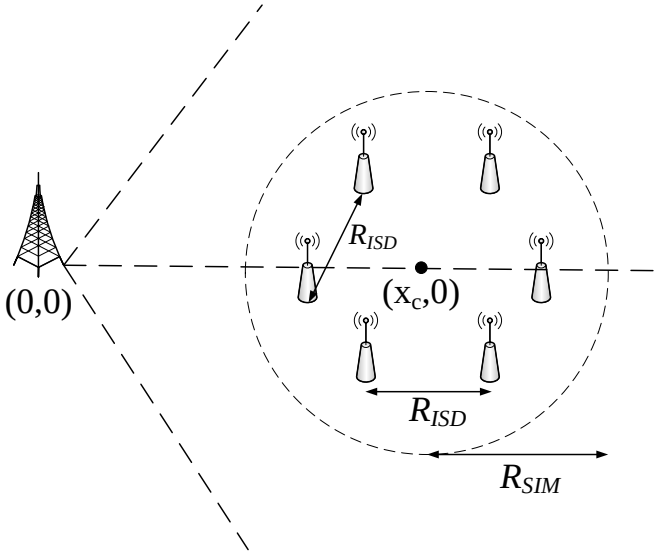


FIGURE 4.7: Simulation scenario topology

Summary of results

- Q-AAA requires a higher system bandwidth utilization than SINR-Alg, which however leads to lower congestion of BSs along with users served with higher satisfaction and for more time.
- Q-AAA gains substantially higher profit than SINR-Alg, thanks to the better network performance.
- The increase in inter-site distance among the small cells deteriorates the network performance and profit gains in both algorithms, however this loss is significantly lower for Q-AAA than SINR-Alg.

4.6.2 Resource Allocation

4.6.2.1 Scenario description and parameters

The scenario used for the performance evaluation consists of a cluster with 6 small cells deployed in the coverage area of a macrocell sector. The cluster is circular shaped and centred at location $c = (x_c, 0)$, as shown in the layout depicted in Fig. 4.7. Along simulations, x_c is randomly selected according to a uniform distribution with $x_c \in [100, 190]m$. The ISD between two small cells equals $R_{ISD} = 50m$.

Users are uniformly distributed within a radius of $R_{SIM} = 75m$ from c, and the SP of each user is selected with equal probability among the SPs defined in Table 4.4. As it can be observed in Table 4.4, three services are considered, each one with two QoE

TABLE 4.4: Service Profiles' parameters

Service	**QoE class**	$\{\mathbf{r_{k1}}, \mathbf{r_{k2}}, \mathbf{r_{k3}}\}$ (Mbps)	$\theta_{\mathbf{k}}^{\mathbf{t}}$ **or** $\theta_{\mathbf{k}}^{\mathbf{B}}$
Service 1	Basic	5.5	1.5€/GB
(Data Based)	Premium	7	2€/GB
Service 2	Basic	$\{3.5, 4, 5\}$	4€/h
(Time Based)	Premium	$\{4, 4.5, 5.5\}$	7€/h
Service 3	Basic	$\{4.5, 5.5, 6\}$	4€/h
(Time Based)	Premium	$\{5, 6, 7\}$	7€/h

TABLE 4.5: BS parameters

Parameter	**Macrocell**	**Small cell**
c_i (€/sec)	$5 \cdot 10^{-5}$	$5 \cdot 10^{-5}$
h_i (MHz^{-1})	0.28	0.275
b_i (MHz)	20	20
Transmission Power (dBm)	43	30

classes $\mathcal{Q} = \{\text{Basic, Premium}\}$: Service 1 is a data-based charged service, and Services 2 and 3 are time-based charged services. Likewise, 3 different devices are considered, and the corresponding transmission rates associated to each SP, r_{kd}, are also included in Table 4.4. Note that for each SP, r_{kd} is the transmission rate required to perceive the service's target QoE level, Q_k^{tg}. In the simulations, the transmission rate that results in a perceived QoE equal to Q_k^{drop} is set to $r_{kd}^{drop} = 0.7 r_{kd}$ for all SPs. Therefore, when a user is served with a data rate below or equal to $0.7 r_{kd}$, the connection is dropped. Moreover, v_k is selected randomly so as to have a price-based QoE component $Q_p(p_k) \in [0.8, 0.9]$ in (4.5), α_{kd} and β_{kd} are calculated according to (4.15a) and (4.15b), and $\gamma_{kd} = 1$, for all SPs π_k, devices $d \in \mathcal{D}$ and target and drop QoE levels $(Q_k^{tg}, Q_k^{drop}) = (3.5, 2.5)$ for Basic QoE class and $(Q_k^{tg}, Q_k^{drop}) = (4.5, 3.5)$ for Premium QoE class of all services.

Parameters used for the BSs, both eNBs and small cells, are listed in Table 4.5. For the carrier bandwidth allocated to each tier, we adopted 3GPP LTE-A's channel models described in [93], and the antenna gains are set to 0 dB. For the cell selection, we associate the users to the BS with the highest SINR, as it is common practice in mobile networks [106]. For PM, the values for the change in user satisfaction are $\sigma_{step} = \{0.01, 0.05\}$. Moreover, the minimum acceptable satisfaction level for all algorithms is $\sigma_{ij}^{min} = 0.01$ (note that with null user satisfaction, i.e. $\sigma_{ij}(t) = 0$, the session is dropped). The results shown in the following subsections were acquired through Monte-Carlo simulations, where each simulation iteration examines a single network instance of a $T = 1$ *sec* duration. It should be noted that we assume perfect channel estimation, and hence we do not inspect the network and economic impact of imperfect channel estimations or rate fluctuations during consecutive instances.

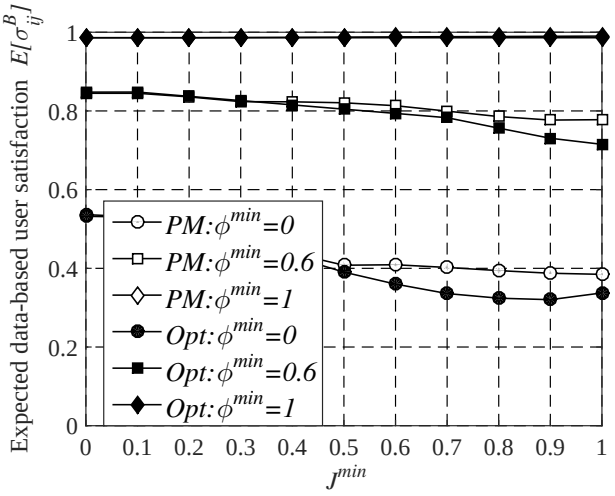


FIGURE 4.8: Expected user satisfaction for data-based users ($E[\sigma_{ij}^B]$) versus target minimum Jain's index (J^{min}) for $N_U = 80$ users

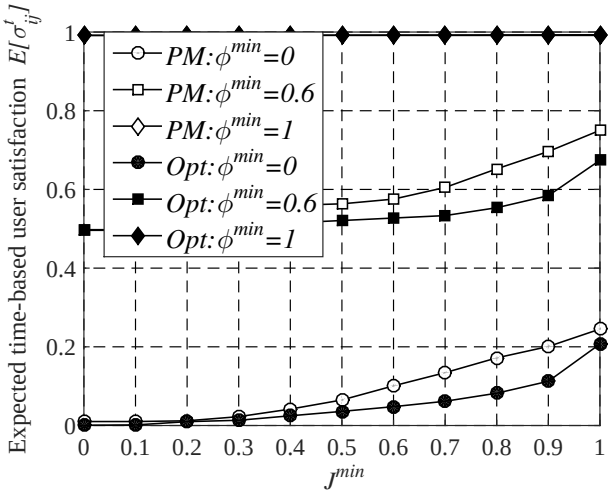


FIGURE 4.9: Expected user satisfaction for time-based users ($E[\sigma_{ij}^t]$) versus target minimum Jain's index (J^{min}) for $N_U = 80$ users

4.6.2.2 Impact of fairness and Satisfaction constraints

Initially, we present how the fairness and relative overall satisfaction objectives (i.e. J^{min} and ϕ^{min}) affect the user satisfaction, the network performance and the MNO profit, when we apply PM on a system with $N_U = 80$ users. We also present the results generated by the optimal solution of the problem in (4.13) (henceforth labelled as Opt).

In Fig. 4.8 and 4.9 we show the expected user satisfaction of the data-based and time-based charged users, denoted by $E[\sigma_{ij}^B]$ and $E[\sigma_{ij}^t]$, respectively. Unserved users are not considered in the calculation of $E[\sigma_{ij}^t]$ and $E[\sigma_{ij}^B]$, since when users are not served the satisfaction of the user is $\sigma_{ij}(t) = 0$. We observe that for low to medium relative overall user satisfaction levels ϕ^{min} and no fairness constraint ($J^{min} = 0$) there is a substantial

difference between the user satisfaction of data-based and time-based charged users (i.e. 521% and 69% higher mean satisfaction of data-based charged users ($E[\sigma_{ij}^B]$) than of time-based charged users ($E[\sigma_{ij}^t]$), when relative overall satisfaction is $\phi^{min} = 0$ and $\phi^{min} = 0.6$ respectively). As explained in section 4.5.2, a BS i can maximize its profit by serving time-based charged users with the minimum acceptable satisfaction (thus with the minimum allocation of resources $w_{ij}(t)$). At the same time, the data-based charged users require to be served with high rates (and satisfaction) in order for BS i to gain high revenue. However, as the fairness requirement increases ($J^{min} \uparrow$), the difference between the mean satisfaction of data-based and time-based charged users ($E[\sigma_{ij}^B]$ and $E[\sigma_{ij}^t]$) decreases. In order to increase fairness ($J_i(t)$), the standard deviation of the satisfaction among the users in BS i must be decreased (as explained in Section 4.5.2). This can only be achieved by increasing the resources allocated to time-based charged users at the expense of reducing the resources allocated to data-based charged users. Finally, we observe that for high relative overall satisfaction constraints (i.e. $\phi^{min} = 1$, or in other words, the overall user satisfaction must be always the maximum one, OS_i^{max}), the PM algorithm skips the second and the third parts of Algorithm 3, since the overall user satisfaction achieved in the first part of PM (step 15) can not be reduced. Therefore, users are either served with maximum satisfaction ($\sigma_{ij}(t) = 1$) or dropped[3]. Regarding the optimal results, we observe that they show the same trend as PM, and small differences for medium to high J^{min} values.

TABLE 4.6: Percentage of Served Users for $N_U = 80$

PM	**Data-Charged Users**		**Time-Charged Users**	
ϕ^{min} / J^{min}	0	1	0	1
0	52.69%	54.13%	67.12%	68.74%
1	52.68%	52.68%	65.05%	65.05%

Table 4.6 shows the percentage of users served according to their service's pricing scheme, for target minimum Jain's index and relative overall users satisfaction (J^{min} and ϕ^{min}) equal to 0 and 1 (i.e. minimum and maximum levels) for $N_U = 80$ users, according to the PM results[4]. Note that J^{min} is the objective, but the actual value of the Jain's index $J_i(t)$ could not reach J^{min} if there are users with very low SINR levels. We observe that for a given number of users N_U, fairness and relative overall satisfaction objectives (J^{min} and ϕ^{min}) have little or no effect on the percentage of served users. However, we observe that more time-based than data-based charged users are served in all cases (i.e. $[23, 27]$% more time-based charged users). This occurs because in general time-based charged services have lower requirements in terms of transmission rate than data-based

[3]As resources are allocated until their depletion, the last user served could receive resources that result in $0 < \sigma_{ij}(t) < 1$ (steps 8-9).

[4]The corresponding optimal results are approximately the same with absolute differences below 1%, and hence they are omitted.

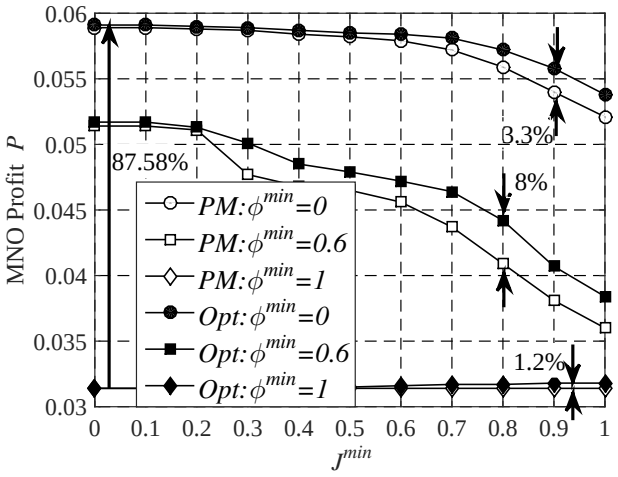


FIGURE 4.10: Profit (P) versus target minimum Jain's index (J^{min}) for $N_U = 80$ users

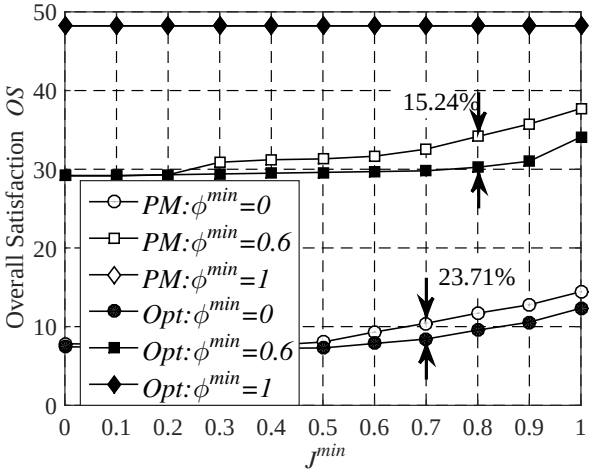


FIGURE 4.11: Overall satisfaction (OS) versus target minimum Jain's index (J^{min}) for $N_U = 80$ users

charged services. Therefore, although more time-based charged users are served (see Table 4.6), they are served with lower satisfaction, for low and medium relative overall satisfaction objectives (see Fig. 4.8 and 4.9).

Fig. 4.10 and Fig. 4.11 depict the MNO profit P and the total overall satisfaction OS for $N_U = 80$ users, respectively. In these two figures, we can observe the trade-off among the MNO profit P, the Overall Satisfaction (OS), and the objective minimum fairness (J^{min}). Particularly, profit is maximized when there are no satisfaction and fairness constraints, which also leads to the lowest Overall Satisfaction. This occurs because the lack of fairness and satisfaction objectives turns the PM algorithm into a *pure* profit maximization solution. Conversely, when the objective relative overall satisfaction (ϕ^{min}) is high, the profit is substantially lower than the maximum (i.e. $[65, 87]\%$ profit gain for $\phi^{min} = 0$ over $\phi^{min} = 1$). This decrease of the profit is caused

by the fact that the proposed algorithm initially maximizes the satisfaction; then, it modifies the solution to increase the profit as long as the minimum relative satisfaction objective (ϕ^{min}) is satisfied. Therefore, the higher the value of ϕ^{min}, the closer to a *pure* satisfaction maximization algorithm PM is. Similar to Fig. 4.8 and 4.9, the optimal results follow the same trend as the PM results, with differences appearing for medium to high J^{min} values. Regarding the MNO profit, we see that PM performs close to the optimum in most cases. The highest differences are observed for the low and medium OS requirements (i.e. 3.3% and 8% difference for $\phi^{min} = 0$ and $\phi^{min} = 0.6$ respectively), whereas there are minuscule differences when the OS must be maximized (i.e. $\phi^{min} = 1$).

From the results in Fig. 4.8-4.11, it can be concluded that when there are two pricing schemes (data-based and time-based charging), the profit can be maximized only when data-based charged users are prioritized over the time-based charged users in terms of average user satisfaction (which leads to low fairness). At the same time, serving more time-based charged users allows the MNO to gain revenue with a low cost. Finally, the trade-off between Profit and Overall Satisfaction shows that the increase of one of them implies the decrease of the other, as specified in the analysis in Section 4.5.2.

4.6.2.3 Comparison with SoA algorithms and impact of pricing

In the following (i.e. Fig. 4.12-4.16 and Table 4.7), we compare PM with two algorithms referred to as Alg-5 [58] and Alg-6 [70], and the optimal solution of (4.13) (i.e. Opt).

Alg-5[58] is an iterative resources allocation algorithm that maximizes the user QoE. In each iteration, Alg-5 allocates enough resources to satisfy a single user, starting from the user with the highest spectral efficiency towards the user with the lowest spectral efficiency.

Alg-6 is an iterative resources allocation algorithm proposed in [70]. In the scenario proposed in [70], an MNO offers wireless video broadcasting services, and aims to maximize its profit. Particularly, the MNO broadcasts a set of video contents (e.g. tv channels) to different users groups. Each user has her own utility function, which is defined as the perceived QoE minus the charge for the service. In order to maximize the profit, Alg-6 increases the rate of a single content until either all contents are served with their *"ideal rate"*[5] or the available bandwidth is fully utilized. The content whose rate will be increased is the one that maximizes the marginal MNO profit, provided that it is non-zero.

[5]In [70], *"ideal rate"* is defined as the rate that satisfies perfectly all the users that share a particular content.

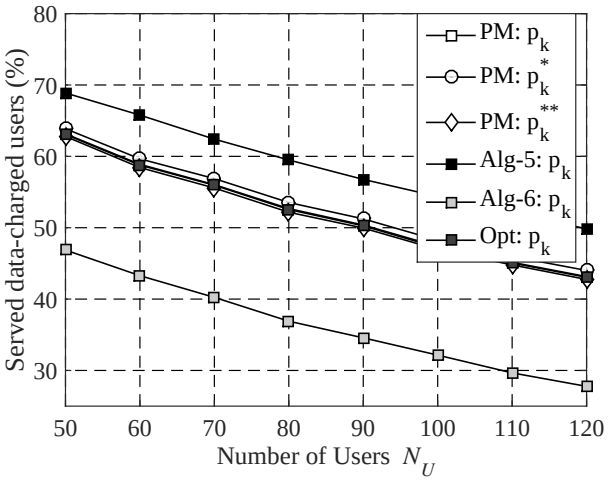


FIGURE 4.12: Percentage of served Data-based charged users

In order to compare Alg-6 with PM, we assume that each user demands a content, which is unique to herself. The *"ideal rate"* of a broadcast content described in Alg-6 corresponds to the user's SP-device pair required transmission rate r_{kd}, in our system model. Therefore, Alg-6 has to broadcast N_U unique contents, whose rate requirements are defined by each user's SP-device pair. Hence, the broadcasting resources allocation problem is transformed into the typical resources allocation problem, where each user's rate (i.e. QoE) is decided separately.

It should be noted that we do not concur with pure profit maximizing policies that result in low quality service provision, as presented previously. To that end, the results provided in the following for PM and Opt are obtained with strict fairness and satisfaction objectives, that is $(J^{min}, \phi^{min}) = (1, 1)$, which maximize the OS. In order to study the variations of the price on PM, two additional price values have been defined with respect to the price stated in Table 4.4: $p_k^* = 0.75p_k$ and $p_k^{**} = 1.5p_k$. However, initially we only compare PM, Alg-5, Alg-6 and the optimal results with price equal to p_k.

***Comparison of PM, Alg-5, Alg-6 and Opt*:** Fig. 4.12 and 4.13 show the percentage of served data-based and time-based charged users respectively, whereas Fig. 4.14 and Table 4.7 show the expected satisfaction for data-based and time-based charged users ($E[\sigma_{ij}^B]$ and $E[\sigma_{ij}^t]$ respectively[6]). Initially, only the price p_k is considered. As expected, we observe in Fig. 4.12 and 4.13 that the percentage of served users decreases as the total number of users N_U is increased, for both data-based and time-based users. This fact is caused by the network congestion and, consequently, by the lack of resources to serve all the users.

[6]The results in Table 4.7 are included as a table because they are constant for $N_U \in [50, 120]$ users.

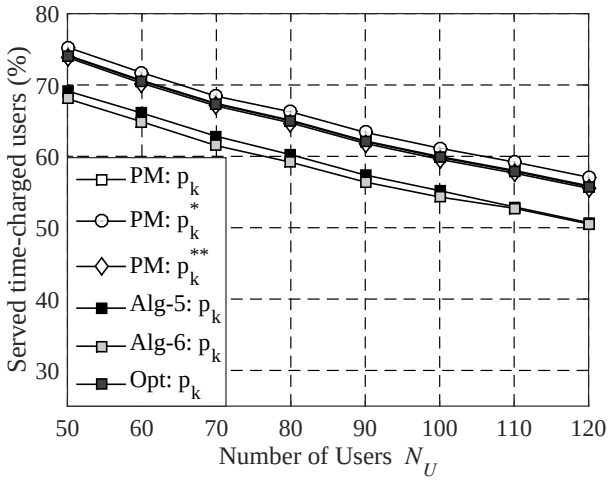


FIGURE 4.13: Percentage of served Time-based charged users

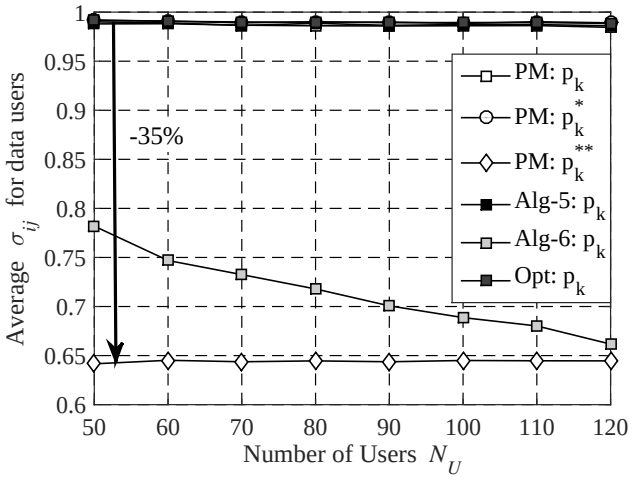


FIGURE 4.14: Expected satisfaction for data-based charged users

With Alg-5, both the percentage of served users (Fig. 4.12 and 4.13) and the expected user satisfaction for data-based charged users ($E[\sigma_{ij}^B]$ in Fig. 4.14) and time-based charged users (first row of Table 4.7) is almost the same. This occurs because resources allocation algorithms that are exclusively aimed to QoE maximization (like Alg-5) do not prioritize services based on their pricing scheme. Conversely, with Alg-6 substantially more time-based charged users are served ($[45, 81]\%$ more time-based than data-based charged users are served when comparing Fig. 4.12 and Fig. 4.13), but with significantly lower expected user satisfaction (in line with PM results when $\phi^{min} = 0$ - see Fig. 4.8 and 4.9). This is explained by the fact that Alg-6 is aimed to maximize the profit. If the BS allocates to time-based users the minimum amount of resources to avoid their connection being dropped: i) resources are better distributed among users; ii) dropping is reduced; and iii) the number of served time-based users and the profit are increased. Similarly to Alg-6, it can be observed in Fig. 4.12 and Fig. 4.13 that PM always

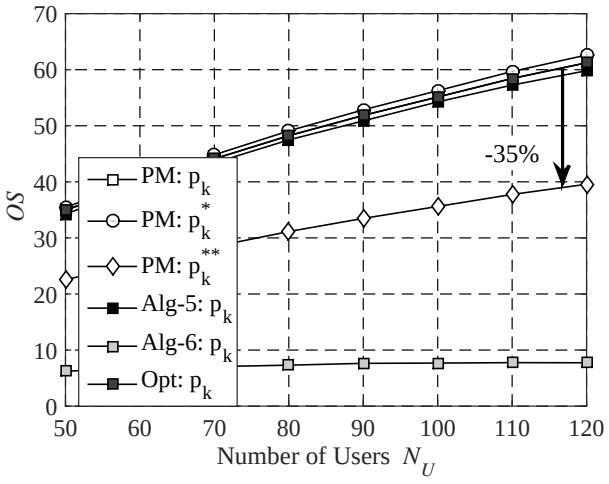


FIGURE 4.15: Overall Satisfaction (OS)

serves a larger percentage of time-based charged users, and this difference increases with the total number of users N_U ($[17, 29]\%$ more time than data-based charged users are served). This can be explained by the fact that the time-based charged users have lower rate requirements on average than the data-based charged users due to the use of different devices, as it can be seen in Table 4.4. Therefore, time-based charged users with low rate requirements are prioritized over data-based charged users in order to achieve the high user satisfaction performance observed in Table 4.7.

TABLE 4.7: Expected satisfaction for time-based charged users ($E[\sigma_{ij}^t]$)

Time-Charged Users				
Price level	**PM**	**Alg-5**	**Alg-6**	**Opt**
p_k	0.9909	0.9901	0.01	0.9929
p_k^*	0.9903	-	-	-
p_k^{**}	0.6427	-	-	-

Summing up, given a particular relative overall satisfaction constraint (ϕ^{min}) and two charging schemes (i.e. data and time-based charging), the profit will be maximized by serving as many time-based charged users as possible. Additionally, if the objective of the MNO is the maximization of the profit without any fairness and satisfaction constraints (i.e. PM with $(J^{min}, \phi^{min}) = (0, 0)$ and Alg-6), the data-based charged users will have higher expected satisfaction than the time-based charged users.

Fig. 4.15 shows the comparison of the system's overall user satisfaction OS. As expected, Alg-5 and PM offer the highest OS, whereas for Alg-6 OS is kept low and slightly increasing with the number of users N_U. As mentioned earlier, Alg-5 sorts the users according to their spectral efficiency, and then allocates the resources until they are exhausted. This means that Alg-5 will first serve the users with the highest spectral efficiency regardless of their service's requirements. Conversely, PM finds iteratively the

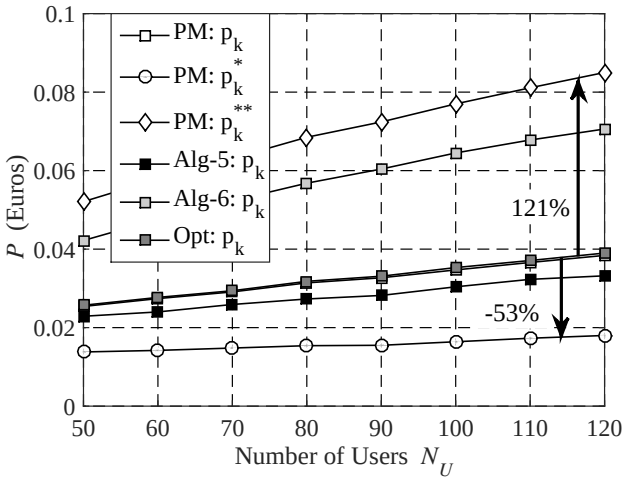


FIGURE 4.16: MNO Profit P

user with least resource requirements $w_{ij}(t)$ and serves her with the maximum satisfaction. In Fig. 4.15, the difference between PM and Alg-5 in terms of Overall Satisfaction (OS) is within the range $[1.14, 2.36]\%$. Alg-5 performs well when there is a single service with a single rate requirement. However, in a scenario with heterogeneous traffic as well as diverse pricing, a more elaborate algorithm such as PM is required in order to serve the users with even higher satisfaction, while gaining large MNO profit.

Fig. 4.16 presents the comparison of the MNO profit for the PM, Alg-5, Alg-6 and Opt. As expected, we observe that Alg-6 gains the highest profit, which increases with the number of users N_U. Due to the use of strict OS and fairness constraints, PM achieves lower profit than Alg-6. Nevertheless, it outperforms notably Alg-5 (i.e. $[10.92, 15.96]\%$), even though they share an almost equal OS performance. Regarding the results from the optimal solution of (4.13), we observe that in all of the Fig. 4.12-4.16 and Table 4.7 PM and Opt have the same performance. A small difference is visible only in Fig. 4.16, where we can see that Opt generates more profit than PM by a small margin (i.e. $[1.18, 1.73]\%$ profit gain for Opt over PM). This gain in profit corresponds to an even smaller loss in OS performance (i.e. $< 1\%$ smaller OS than PM).

In light of the results presented above, PM is a profit maximizing resource allocation algorithm, which guarantees similar QoE performance results to the ones achieved with Alg-5 (a QoE maximizing algorithm), thanks to the satisfaction and fairness constraints (in Fig. 4.12-4.16 $J^{min} = \phi^{min} = 1$). However, the good performance in terms of QoE is not translated in a decrease of the profit. Specifically, PM achieves higher profit than Alg-5 for all scenarios. Finally, when strict satisfaction and fairness constraints are applied, PM provides approximately optimal results for the solution of problem (4.13).

***Impact of price level*:** Results of Fig. 4.12-4.16 have been so far analysed with the price level p_k stated in Table 4.4. Now, we examine in the same figures how the price levels affect the PM algorithm in terms of users satisfaction, the percentage of served users, the MNO profit and the Overall Satisfaction. In order to do this, the price values p_k, $p_k^* = 0.75p_k$ and $p_k^{**} = 1.5p_k$ are considered. Particularly, in order to examine the price change on the users, we generate the users' parameters for the original prices p_k, and then execute our algorithm for the decreased and increased values.

Taking the above into consideration as well as the corresponding results, it can be observed that the decrease of the price ($p_k \rightarrow p_k^*$) results in an increase of the percentage of served users (see Fig. 4.12 and 4.13) and the mean users' satisfaction (see Fig. 4.14 and Table 4.7). Note that the reduction of the price improves the perception of the user thanks to the price-based component of the QoE (as shown in expression (4.5)). The opposite behaviour is observed when the price is increased ($p_k \rightarrow p_k^{**}$). Consequently, as we observe in Fig. 4.15 and 4.16 the higher MNO profit gained by the higher price p_k^{**} (a gain in the range of 105-121%), comes at the expense of lower expected user satisfaction $E[\sigma_{ij}^B]$, $E[\sigma_{ij}^t]$ (with a decrease around 35%), and lower Overall Satisfaction OS (around 35%). Conversely, when p_k is decreased to $p_k^* = 0.75p_k$, there is a slight increase in the percentage of served data-based and time-based charged users (in the range of 1.1-2% for data-based charged users and 1.4-2.1% for time-based charged users), and in the Overall Satisfaction ($1.4 - 2.2\%$ gain). However, the improvement in the percentage of served users and in the overall satisfaction is achieved at the cost of a significant profit loss (around $45 - 53\%$ drop of the profit). Thus, a decrease of the price improves the perception of the users (i.e. higher satisfaction) while it reduces the MNO profit, and vice versa.

Summary of results

- The trade-off between Profit and Overall Satisfaction shows that the increase of one of them implies the decrease of the other.
- Given a particular minimum relative overall satisfaction (ϕ^{min}) and two charging schemes (i.e. data and time-based charging), profit (P) will be maximized by serving as many time-based charged users as possible. Additionally, if the MNO maximizes the profit without any fairness and satisfaction constraints, the data-based charged users will have higher expected satisfaction than the time-based charged users.

- PM guarantees similar QoE performance results to the ones achieved with Alg-2 (a QoE maximizing algorithm), thanks to the satisfaction and fairness constraints, while achieving higher profit than Alg-2 for all scenarios.
- When fairness and relative overall satisfaction are forced to be maximum ($J^{min} = \phi^{min} = 1$), PM provides approximately optimal results for the solution of problem (4.13).
- Decreasing the price improves the users' perception while it reduces the MNO profit, and vice versa.

4.6.3 Dynamic Pricing

As it was shown in Sections 4.5.2 and 4.5.3, the proposed algorithms for the solution of the particular profit maximization problems share common parts. As we presented in the previous section, our contribution in pure resource allocation revealed the impact that price changes can have on the system performance and MNO profit. To that end, we researched ways to apply a price adjustment scheme in a real-time manner, which lead to our proposal on dynamic pricing presented in Section 4.5.3. Therefore, the main objective in this section is to demonstrate the performance of our dynamic pricing scheme, and its applicability to different types of resource allocation algorithms.

4.6.3.1 Scenario description and parameters

The scenario used for the performance evaluation of our proposal on joint resource allocation and dynamic pricing is the same as the scenario we use for the evaluation of our contribution in Section 4.6.2. The only differences between the two setups are found in some of the SP and BS parameter values, which we present in Tables 4.8 and 4.9 respectively. Furthermore, we assume dedicated spectrum allocation per tier, adopted 3GPP LTE-A's channel models described in [93], and set the Antenna gains to 0 dB. For the cell selection, we associate the users to the BS with the highest SINR, as it is common practice in mobile networks [106].

4.6.3.2 Comparison with SoA algorithm and impact of dynamic pricing

The following results were acquired through Monte-Carlo simulations. We compare our proposed algorithm (referred to as PM) with a QoE maximizing algorithm, referred to as Overall user Satisfaction Maximization (OSM) [58] (also used for the performance

TABLE 4.8: Service Profiles' parameters

Service	**QoE class**	$\{\mathbf{r_{k1}}, \mathbf{r_{k2}}, \mathbf{r_{k3}}\}$ (Mbps)	$\theta_{\mathbf{k}}^{\mathbf{t}}$ **or** $\theta_{\mathbf{k}}^{\mathbf{B}}$
Service 1	Basic	7	1.5€/GB
(Data Based)	Premium	9	2€/GB
Service 2	Basic	$\{4.5, 5, 6.5\}$	4€/h
(Time Based)	Premium	$\{5, 6, 7\}$	7€/h
Service 3	Basic	$\{6, 7, 7.5\}$	4€/h
(Time Based)	Premium	$\{6.5, 7.5, 8.5\}$	7€/h

TABLE 4.9: BS parameters

Parameter	**Macrocell**	**Small cell**
c_i (€)	$5 \cdot 10^{-5}$	$5 \cdot 10^{-5}$
h_i (MHz^{-1})	0.3	0.295
b_i (MHz)	20	20
Transmission Power (dBm)	43	30

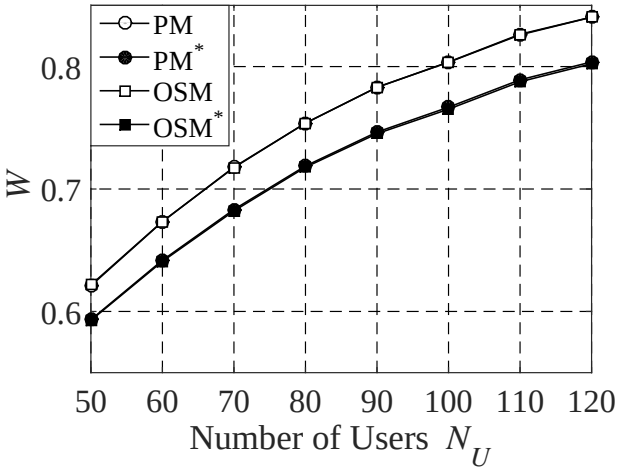


FIGURE 4.17: System bandwidth utilization (W)

evaluation in the previous section). We remind that OSM is a resource allocation algorithm that maximizes the user QoE through an iterative procedure, which in each iteration allocates enough resources to satisfy a single user, starting from the user with the highest spectral efficiency towards the user with the lowest. In order to show the benefits of our dynamic pricing scheme, we provide results for both algorithms with and without applying dynamic pricing (the symbol * refers to the algorithms with dynamic pricing applied). For PM, when dynamic pricing is not applied, it is $\Lambda = \{1\}$. For PM, the values for the change in user satisfaction are $\sigma_{step} = \{0.01, 0.05\}$. Moreover, the minimum acceptable satisfaction level for all algorithms is $\sigma_{ij}^{min} = 0.01$. It should be noted that we provide results for PM with $\phi^{min} = 1$, aiming to offer the users the service agreed in their SPs.

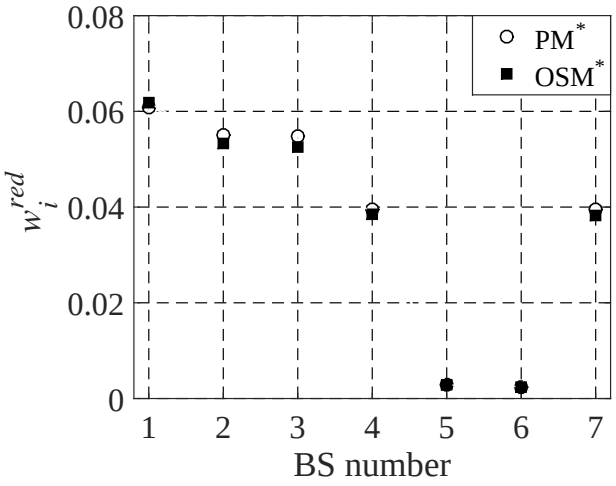


FIGURE 4.18: BS bandwidth utilization reduction (w_i^{red})

Fig. 4.17 shows the expected total utilization of the spectrum, which is defined as $W = \mathbb{E}\left[W(t) = \frac{\sum_{i\in\mathcal{B}} w_i(t)b_i}{\sum_{i\in\mathcal{B}} b_i}\right]$ versus the number of users N_U. We observe that PM and OSM consume the same portion of the bandwidth whether dynamic pricing is applied or not. Regarding the gain from dynamic pricing, we see that both PM* and OSM* use $[4.46, 4.92]\%$ less of their total resources compared to PM and OSM respectively.

Fig. 4.18 depicts the bandwidth utilization reduction of each BS i (w_i^{red}), when dynamic pricing is used, where $i = 1$ denotes the macrocell BS. In order to produce the presented results, we averaged w_i^{red} of each BS i for all simulated N_U values. We observe that both algorithms share similar W and w_i^{red} gains, when dynamic pricing is applied. This occurs because the reduction in $w_{ij}(t)$ depends on the users' individual parameters, that is, her current spectral efficiency $\varepsilon_{ij}(t)$, her rate requirement r_{kd}, and the impact that the service price has on her QoE perception Q_p. We further notice that there is a high deviation in w_i^{red} among the BSs. Particularly, w_i^{red} at BSs 5 and 6 is almost zero. This is a result of low load in these two BSs. Due to the use of the SINR-based cell selection scheme, the small cells closer to the eNB are associated with a small number of users. Hence, they have low spectrum requirements, and if dynamic pricing is applied, the revenue loss will be higher than the cost reduction.

Fig. 4.19 and Fig. 4.20 shows the algorithms' performance on the MNO profit P and the overall user performance OS, respectively. We see that our proposal outperforms OSM in terms of profit, and shows a slight gain over OS as well ($[1.65, 2.83]\%$ gain). As mentioned earlier, OSM sorts the users according to their spectral efficiency $\varepsilon_{ij}(t)$, and then allocates the resources until they are exhausted. This means that OSM will first serve the users with the highest $\varepsilon_{ij}(t)$ regardless of their service's requirements. Conversely, PM sorts the users according to their resource requirements $w_{ij}(t)$ for serving them with their maximum $\sigma_{ij}(t)$. OSM performs well when there is a single rate

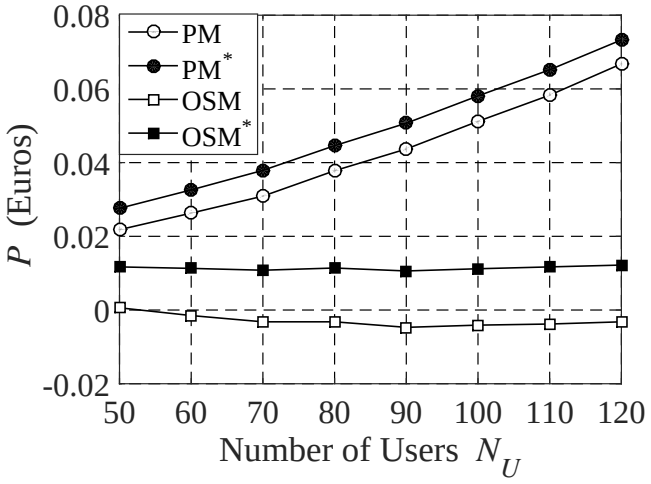


FIGURE 4.19: MNO Profit (P)

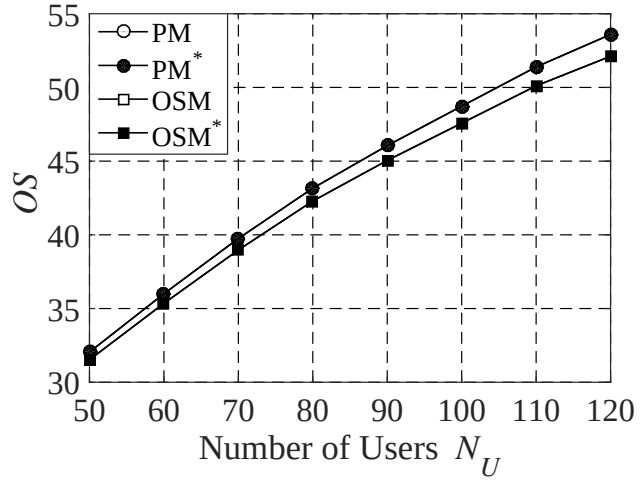


FIGURE 4.20: Overall User Satisfaction (OS)

requirement. However, in a scenario with heterogeneous traffic as well as diverse pricing, a more elaborate algorithm such as PM is required in order to serve the users with even higher satisfaction, while gaining large MNO profit.

Regarding the effect of dynamic pricing, we observe that both PM* and OSM* perform the same as PM and OSM respectively, in terms of OS. Conversely, PM* and OSM* provide substantial gains in the MNO profit P ($[9.7, 26.6]\%$ gain for PM* and $[325, 1835]\%$ gain for OSM*[7]). For PM, this is explained by the fact that the decisions on resource allocation and dynamic pricing are made in order to maximize the BS profit while achieving a minimum OS performance (refer to Alg. 4's steps 23-27). Therefore, PM* provides the same OS as PM, however for lower W and cost. As for OSM, we

[7]The high gains observed for OSM* are explained by the fact that OSM's profit is significantly low and close to 0.

applied our dynamic pricing scheme on the resource allocation determined by the original algorithm. Hence, we obtain the same OS, but for a higher profit owing to the cost reduction. If dynamic pricing was applied within the original OSM algorithm, the BS revenue would be significantly low (even zero), as the algorithm would always reduce the service price (i.e. low λ_{ij}) in order to maximize OS.

Our proposed algorithm manages to offer significantly higher profit for a similar network performance compared to the reference algorithm, because it bases its decisions on both technological (i.e. QoS/QoE requirements) and economic (i.e. pricing and profit) context of the network. Moreover, our proposal on dynamic pricing has been proven to complement resource allocation schemes in order to increase substantially the MNO profit. Additionally, in a different application of dynamic pricing the released resources can be used to increase the QoE of the users, while maintaining high MNO profits.

Summary of results

- PM outperforms slightly OSM in terms of overall user satisfaction (independent of the use of dynamic pricing), for the same system bandwidth utilization.
- PM gains substantially higher profit than OSM, while guaranteeing higher QoE performance.
- Dynamic pricing should be used at BS with high resource requirements and bandwidth utilization cost, so that the cost reduction (over)compensates the revenue loss.
- The use of dynamic pricing increases significantly the profit gained for both PM and OSM, while maintaining the same network performance.

4.7 Concluding Remarks

In this chapter, we presented our contributions on user association, resource allocation and joint resource allocation and dynamic pricing in a single MNO's HetNet described by traffic heterogeneity and diverse pricing. The objective of these works was the maximization of the MNO's profit while providing high QoE to the users. With this objective, we proposed low-complexity heuristic algorithms, each of them with particular constraints that guarantee fairness and high overall users satisfaction (a QoE performance metric). Particularly:

- We proposed a greedy, heuristic user association algorithm, which bases its decisions on the capability of the BSs to offer an acceptable QoE level to the user, while at the same time maximizing the MNO profit. We evaluated the performance of the proposed algorithm by comparing it with a SINR-based algorithm. The simulation results show the adaptability of the proposed algorithm to traffic heterogeneity by achieving substantially higher profit and QoE, in contrast to the SINR-based one. This proposal has led to the publication of [107].

- We proposed a low complexity, greedy, heuristic resource allocation algorithm, namely PM, which maximizes the profit of the MNO as long as fairness and overall users satisfaction (a QoE performance metric) levels are kept above specific thresholds. We highlighted the overall satisfaction-fairness-profit trade-off and showed that the proposed resources allocation algorithm achieves similar results in terms of users' satisfaction when compared to QoE maximization algorithms (e.g. Alg-5), while outperforming them in terms of profit gains. Moreover, we demonstrated that our algorithm approximates the optimal solution of the profit maximizing problem. With the analysis of two pricing schemes (one for data-based charged users and another for time-based charged users), we further showed that *pure* profit maximizing algorithms prioritize the satisfaction of data-based charged users over time-based charged users, and we proved that PM with strict fairness and satisfaction constraints smooths this effect. Finally, we shed light on how the changes in price levels can improve or worsen the user experience, and the corresponding effects of these changes on the network performance and the MNO profit. Our contribution on resource allocation resulted in the publication of [108].

- We proposed a greedy, heuristic, joint resource allocation and dynamic pricing algorithm, which bases its decisions on profit maximization, while satisfying a constraint on overall user satisfaction. We evaluated the performance of the proposed algorithm with and without applying dynamic pricing by comparing it with a state of the art resource allocation algorithm. Our results verify the adaptability of the proposed algorithm to traffic heterogeneity, by providing higher OS and profit than the algorithm in comparison. Finally, we show that our proposal on dynamic pricing can be applied on different algorithms, allowing them to improve either the MNO profit or the OS performance. Our proposal on dynamic pricing produced our work in [109].

Chapter 5

Conclusions and Future Work

This chapter concludes the dissertation with the summary of the main contributions, and the presentation of potential subjects for future research. Specifically, Section 5.1 outlines the conclusions derived from each contribution part, and Section 5.2 lists research lines towards the extension of our contributions.

5.1 Conclusions

The ever increasing data traffic demand in mobile networks since the introduction of the data-oriented 3G has been driving the telecommunications research for faster, more reliable and high-capacity networks, which has resulted in various proposals for their enhancement. One of the most promising solutions for increasing the network capacity is the dense deployment of small cells, and their use for mobile data offloading. However, even though a ubiquitous small cell deployment can solve the traffic demand challenge, it poses high financial risks due to the corresponding high capital requirements. To that end, Mobile Network Operators (MNOs) started outsourcing mobile data offloading to independent third parties also known as Small Cell Operators (SCOs), owners of small cell infrastructure.

One approach that can enable the access of third parties in the telecommunication infrastructure market and therefore in outsourced traffic offloading is the business model of the Multiple Operator Radio Access Network (MORAN) sharing, also known as Small Cell as a Service (SCaaS). In SCaaS, the SCO acts solely as an infrastructure provider, and hence the MNOs need to provide their own licensed spectrum resources for the RAN operation. Moreover, with MORAN the MNOs are the ones that determine how their spectrum will be used by the small cell infrastructure, applying their own service policies as in their own network.

As the deployment of small cell infrastructure in high traffic areas creates a market of small cell capacity, it attracts the interest of multiple MNOs, creating a competition among them for the leasing of the limited small cell resources. As a result, there is a need for frameworks that determine the transactions and the small cell resource allocation in this market, and provide fair participation opportunities for all the interested MNOs.

Taking the above into consideration, our first contribution (presented in Chapter 3) was focused on outsourced traffic offloading, where we examined the case where a monopolistic SCO offers SCaaS to multiple MNOs, addressing the scenarios with a single SCO in the area under study (e.g. stadiums, airports etc.) [82, 83]. Particularly, we investigated the interaction of the capacity needs and the economic constraints of the stakeholders, and gained insight into techno-economic implications of the SCaaS paradigm. We analysed the MNOs' auction strategies, and their impact on the system, and then proposed a novel learning mechanism for improving the MNOs' bidding strategies. The main contributions of this work are summarized in the following:

- We proposed a realistic analytical model for traffic offloading under the SCaaS approach, considering multiple small cell clusters, while including both the technological constraints (e.g. bandwidth availability, backhaul capacity) and the financial goals of each stakeholder.
- The SCaaS model can be implemented under two spectrum deployment use cases: i) the dedicated spectrum deployment, and ii) the co-channel deployment. This work showed the profit-capacity trade-off for each use case and proposed the mathematical framework to define the capacity limits of SCaaS in each use case.
- As the stakeholders' profit and the total system capacity depend highly on i) the deployment density (i.e. density of deployed eNBs and small cells), ii) the competition level among MNOs (i.e. the capacity needs of each MNO), iii) the SCO backhaul capacity, iv) the reuse or not of the spectrum bands, and v) each actor's cost function, the presented results offer useful insights to select the adequate values of the parameters involved in the SCaaS approach.
- The proposed auction scheme relies on the perfect knowledge of the MNOs' future loads. As each MNO is not aware of its future load and the future load of the contending MNOs (due to future uncertainty, as well as privacy of sensitive information), we provided a mathematical framework based on a novel learning mechanism to overcome the lack of reliable information.

The aforementioned raise in mobile data demand has been the result of content dissemination by third-party content providers and the plethora of applications for mobile

devices. These applications are described by various Quality of Service (QoS) and Quality of Experience (QoE) requirements, which along with their use on multiple devices per user increase the heterogeneity of the traffic demand. This ever-increasing, heterogeneous traffic demand has created not only technical challenges for the MNOs, but also financial. Even though the MNOs accommodate this plethora of services and applications with their networks, they do not receive any monetary compensation from the content providers. Moreover, the service prices have been declining over the years due to the MNO competition.

Under these circumstances, the MNOs need to guarantee the provision of seamless connectivity and high satisfaction to their subscribers, while generating high revenues to recoup their HetNet investments and make a profit. Hence, after solving the traffic offloading problem, an MNO has at its disposal additional small cell infrastructure, which can be used to provide better service to its subscribers and improve its own economic gains. Consequently, the solution of the traffic offloading problem led us to a new area of study, which examines how an MNO can use efficiently and profitably its HetNet.

After solving the traffic offloading problem, an MNO has at its disposal additional small cell infrastructure, which can be used to provide better service to its subscribers and improve its own economic gains. Consequently, the solution of the traffic offloading problem led us to a new area of study, which examines how an MNO can use efficiently and profitably its HetNet. The MNO network's performance and financial gains depend highly on how network and economic functionalities are designed and applied. Therefore, in order to identify the appropriate network and economic functions, and design them so that they guarantee subscriber satisfaction and raise the MNO profit, we first need to identify the challenges that need to be addressed. Therefore, in this thesis' second contribution part we presented our proposals on network and economic functionalities in order to address these challenges.

Specifically, we presented our proposals on user association, resource allocation and joint resource allocation and dynamic pricing in a single MNO's HetNet described by traffic heterogeneity and diverse pricing. The objective of these works was the maximization of the MNO's profit while providing high QoE to the users. With this objective, we proposed low-complexity heuristic algorithms, each of them with particular constraints that guarantee fairness and high overall users satisfaction (a QoE performance metric). Particularly:

- We proposed a greedy, heuristic user association algorithm, which bases its decisions on the capability of the BSs to offer an acceptable QoE level to the user,

while at the same time maximizing the MNO profit. We evaluated the performance of the proposed algorithm by comparing it with a SINR-based algorithm. The simulation results show the adaptability of the proposed algorithm to traffic heterogeneity by achieving substantially higher profit and QoE, in contrast to the SINR-based one. This proposal has led to the publication of [107].

- We proposed a greedy, heuristic resource allocation algorithm, namely PM, which maximizes the profit of the MNO as long as fairness and overall users satisfaction (a QoE performance metric) levels are kept above specific thresholds. We highlighted the overall satisfaction-fairness-profit trade-off and showed that the proposed resources allocation algorithm achieves similar results in terms of users' satisfaction when compared to QoE maximization algorithms (e.g. Alg-5), while outperforming them in terms of profit gains. Moreover, we demonstrated that our algorithm approximates the optimal solution of the profit maximizing problem. With the analysis of two pricing schemes (one for data-based charged users and another for time-based charged users), we further showed that *pure* profit maximizing algorithms prioritize the satisfaction of data-based charged users over time-based charged users, and we proved that PM with strict fairness and satisfaction constraints smooths this effect. Finally, we shed light on how the changes in price levels can improve or worsen the user experience, and the corresponding effects of these changes on the network performance and the MNO profit. Our contribution on resource allocation resulted in the publication of [108].

- We proposed a greedy, heuristic, joint resource allocation and dynamic pricing algorithm, which bases its decisions on profit maximization, while satisfying a constraint on overall user satisfaction. We evaluated the performance of the proposed algorithm with and without applying dynamic pricing by comparing it with a state of the art resource allocation algorithm. Our results verify the adaptability of the proposed algorithm to traffic heterogeneity, by providing higher OS and profit than the algorithm in comparison. Finally, we show that our proposal on dynamic pricing can be applied on different algorithms, allowing them to improve either the MNO profit or the OS performance. Our proposal on dynamic pricing produced our work in [109].

5.2 Future Work

The profitable provision of satisfactory service in Heterogeneous Networks was the main objective of the contributions in this dissertation. Nevertheless, there are still open

issues and interesting research subjects that neither this dissertation nor the state-of-the-art has addressed. To that end, in this section we provide a list of new lines for future investigation with respect to the contributions in this dissertation.

The open issues regarding the first contribution part of the thesis on the outsourcing of traffic offloading and generally leasing of telecommunications infrastructure can be outlined as follows:

- In this work, we consider the case of a single SCO offering SCaaS to multiple MNOs. Hence, as the SCO is the sole SCaaS provider in the system, it can act as a monopoly. This case is more suited to scenarios where there can be a single SCO in the area under study (e.g. the owner of a stadium or an airport etc.). A future direction and interesting case study considers multiple SCOs in the same area. In this case, not only the MNOs, but also the SCOs compete among them. The SCOs need to set their prices properly in order to lease their capacity and maximize their profit. On the other hand, the MNOs need to adapt their strategies not only according to the competition among them, but also according to the available choices in small cell capacity and the varying prices.

- Taking into consideration that outsourced traffic offloading belongs to the general case of telecommunication infrastructure sharing, this contribution could be extended in order to be applied to different telecommunication infrastructure and applications. As we are approaching the launch of the first commercial 5G network, various new technologies such as the Internet of Things (IoT), Machine to Machine communications, vehicular or even drone networks will generate an abundance of new data, and introduce new services in the market. Such networks will consist of numerous devices, which will not be deployed by a small number of actors (e.g. MNOs) as in the case of small cell densification. Therefore, infrastructure sharing will be the only solution, when an MNO needs a particular type of infrastructure in an area for providing a specific service. Moreover, as such demands will be generating ubiquitously and continuously, intelligent autonomous schemes such our proposal will be essential for the automatic and profitable undertaking of the necessary actions.

With respect to the second contribution part of the thesis, which is dedicated to radio resource management (RRM) and dynamic pricing schemes, there are open issues that can be summarized in the following:

- Another interesting research subject is the design of a price adjustment scheme either for profit maximization or network performance improvement, based on our

contributions in Chapter 4. The idea for this extension of the proposed schemes is to propose strategies on the degree (i.e. how much should the service prices be decreased and for how many users in the area under study) an MNO should apply dynamic pricing, according to a 24-hour traffic pattern. The MNO would set particular objectives regarding the overall user satisfaction, fairness and profit, and then adjust the charging based on a learning mechanism in order to achieve these objectives, in a manner similar to our first contribution in [82] and [83].

- During the last few years, a number of schemes have emerged, aiming to enhance the network performance such as Device to Device communications, clustering and caching. However, similar to the network functions addressed in our contributions (i.e. user association and resource allocation), these schemes have not been thoroughly examined for their economic impact on the applying MNO, and how the balance between profit maximization and user satisfaction can be achieved. Thus, it would be interesting to conduct research on these schemes, following our techno-economic approach.

- Due to the exponential increase in data traffic and the devices generating and handling it (e.g. new users, network densification, IoT etc.) there is a challenge regarding the abundance of information that needs to be processed in order to provide high levels of satisfactory service provision. In such conditions, the exploitation of machine learning and other artificial intelligence tools pose among the most promising solutions for the aforementioned challenges. As they can handle enormous volumes of data, it is interesting to research schemes that enable dynamic, near-optimal network management, while taking taking into consideration the cost structure and the economic objectives of the MNO.

Concluding, this dissertation has contributed to the state-of-the-art by proposing initially a novel online learning mechanism for the profitable leasing of third-party small cell infrastructure by MNOs, and finally, by presenting QoE-aware user association, resource allocation and dynamic pricing profit maximization algorithms in HetNets with traffic heterogeneity and diverse pricing. These two contributions have provided insight on the sustainable MNO network enhancement with leased small cell infrastructure, and the profitable operation of such HetNets with RRM and smart pricing schemes. Finally, the outlined ideas on future research lines shows that there is still room to further advance the presented subjects, and even apply our techno-economic investigation approach on other telecommunication research areas.

Bibliography

[1] Small Cell Forum, 191.08.02, "Multi-operator and neutral host small cells," Dec 2016.

[2] Cisco, San Jose, CA, USA, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021," Feb 2017.

[3] C. Liang and F. R. Yu, "Wireless network virtualization: A survey, some research issues and challenges," *IEEE Communications Surveys Tutorials*, vol. 17, pp. 358–380, Firstquarter 2015.

[4] Ericsson Press Backgrounder, "City of Los Angeles and Philips shine a light on 4G LTE app coverage with Ericsson," Nov 2015.

[5] Nokia Solutions and Networks, "Nokia's Centralized RAN technology deployed by Telefonica to improve the mobile experience at Valencia's Mestalla Stadium," Jun 2016.

[6] P. Farris, N. Bendle, P. Pfeifer, and D. Reibstein, *Marketing Metrics: The Manager's Guide to Measuring Marketing Performance*. Pearson Education, 2015.

[7] Ofcom, "Award of the 800 MHz and 2.6 GHz spectrum bands – Publication of final results of auction under regulation 111 of the Wireless Telegraphy (Licence Award) Regulations 2012," Mar 2013.

[8] Ofcom, "The award of 800 MHz and 2.6 GHz spectrum-Information Memorandum," Jul 2012.

[9] Real Wireless, Report for Virgin Media, "An assessment of the value of small cell services to operators," Oct 2012.

[10] R. Kelly, A. LaFrance, and S. Sanders, "Spectrum trading in the eu and the us - shifting ends and means," *Telecommunication Laws and Regulations 2012, Global Legal Group Ltd.*, Aug 2011.

[11] Appier Research Report, "Cross-Screen User Behavior Report, Asia 1H 2016," 2016.

[12] Ofcom, "Measuring mobile voice and data quality of experience," 2013.

[13] 5G-PPP, "The 5G Infrastructure Public Private Partnership: the next generation of communication networks and services," 2014.

[14] Mobile Squared, "From Resistance to Partnership Operators shift into monetising OTT," Nov 2014.

[15] Mobile Squared, "Annual Market Review 2014 and PRS market outlook 2015," Jul 2015.

[16] Mobile Squared, "Annual Market Review 2015-16 PRS market outlook 2016-17," Jul 2016.

[17] M. Fiedler, T. Hossfeld, and P. Tran-Gia, "A generic quantitative relationship between quality of experience and quality of service," *IEEE Network*, vol. 24, pp. 36–41, March 2010.

[18] A. Rehman, K. Zeng, and Z. Wang, "Display device-adapted video quality-of-experience assessment," 2015.

[19] P. Reichl, P. Maillé, P. Zwickl, and A. Sackl, "On the fixpoint problem of qoe-based charging," in *6th International ICST Conference on Performance Evaluation Methodologies and Tools*, pp. 235–242, Oct 2012.

[20] 3GPP TR 22.852 V13.1.0., "Study on radio access network (ran) sharing enhancements," Sep 2014.

[21] 3GPP TR 23.251 V14.0.0., "Network sharing; architecture and functional description," Mar 2017.

[22] Nokia Solutions and Networks, "Small cell deployments: you don't have to learn the hard way," 2016.

[23] Analysis Mason, "Active RAN sharing business models can bring benefits to towercos as well as operators," Oct 2014.

[24] Cellnex Telecom, "Cellnex Telecom acquires CommsCon," Jun 2016.

[25] Cellnex Telecom, "Cellnex to add 3000 new sites to its portfolio in France," Feb 2017.

[26] Cellnex Telecom, "Cellnex to incorporate 2,239 sites in Switzerland," May 2017.

[27] S. Paris, F. Martisnon, I. Filippini, and L. Clien, "A bandwidth trading marketplace for mobile data offloading," in *2013 Proceedings IEEE INFOCOM*, pp. 430–434, April 2013.

[28] G. Iosifidis, L. Gao, J. Huang, and L. Tassiulas, "A double-auction mechanism for mobile data-offloading markets," *IEEE/ACM Transactions on Networking*, vol. 23, pp. 1634–1647, Oct 2015.

[29] W. Dong, S. Rallapalli, R. Jana, L. Qiu, K. K. Ramakrishnan, L. Razoumov, Y. Zhang, and T. W. Cho, "ideal: Incentivized dynamic cellular offloading via auctions," *IEEE/ACM Transactions on Networking*, vol. 22, pp. 1271–1284, Aug 2014.

[30] Z. Lu, P. Sinha, and R. Srikant, "Easybid: Enabling cellular offloading via small players," in *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, pp. 691–699, April 2014.

[31] S. Hua, X. Zhuo, and S. Panwar, "A truthful auction based incentive framework for femtocell access," in *2013 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 2271–2276, April 2013.

[32] S. Paris, F. Martignon, I. Filippini, and A. Capone, "A truthful auction for access point selection in heterogeneous mobile networks," in *2012 IEEE International Conference on Communications (ICC)*, pp. 3200–3205, June 2012.

[33] Y. Zhang, S. Tang, T. Chen, and S. Zhong, "Competitive auctions for cost-aware cellular traffic offloading with optimized capacity gain," in *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, pp. 1–9, April 2016.

[34] N. Cheng, N. Lu, N. Zhang, X. Zhang, X. S. Shen, and J. W. Mark, "Opportunistic wifi offloading in vehicular environment: A game-theory approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, pp. 1944–1955, July 2016.

[35] S. Paris, F. Martignon, I. Filippini, and L. Chen, "An efficient auction-based mechanism for mobile data offloading," *IEEE Transactions on Mobile Computing*, vol. 14, pp. 1573–1586, Aug 2015.

[36] A. Bousia, E. Kartsakli, A. Antonopoulos, L. Alonso, and C. Verikoukis, "Multi-objective auction-based switching-off scheme in heterogeneous networks: To bid or not to bid?," *IEEE Transactions on Vehicular Technology*, vol. 65, pp. 9168–9180, Nov 2016.

[37] D. J. Salant, *A Primer on Auction Design, Management, and Strategy*. MIT Press, 2014.

[38] A. Bousia, E. Kartsakli, A. Antonopoulos, L. Alonso, and C. Verikoukis, "Sharing the small cells for energy efficient networking: How much does it cost?," in *2014 IEEE Global Communications Conference*, pp. 2649–2654, Dec 2014.

[39] L. Gao, G. Iosifidis, J. Huang, and L. Tassiulas, "Economics of mobile data offloading," in *2013 Proceedings IEEE INFOCOM*, pp. 3303–3308, April 2013.

[40] K. Wang, F. C. M. Lau, L. Chen, and R. Schober, "Pricing mobile data offloading: A distributed market framework," *IEEE Transactions on Wireless Communications*, vol. 15, pp. 913–927, Feb 2016.

[41] Y. Yang, T. Q. S. Quek, and L. Duan, "Backhaul-constrained small cell networks: Refunding and qos provisioning," *IEEE Transactions on Wireless Communications*, vol. 13, pp. 5148–5161, Sept 2014.

[42] C. H. Chai, Y. Y. Shih, and A. C. Pang, "A spectrum-sharing rewarding framework for co-channel hybrid access femtocell networks," in *2013 Proceedings IEEE INFOCOM*, pp. 565–569, April 2013.

[43] Y. Chen, J. Zhang, and Q. Zhang, "Incentive mechanism for hybrid access in femtocell network with traffic uncertainty," in *2013 IEEE International Conference on Communications (ICC)*, pp. 6333–6337, June 2013.

[44] L. Gao, G. Iosifidis, J. Huang, L. Tassiulas, and D. Li, "Bargaining-based mobile data offloading," *IEEE Journal on Selected Areas in Communications*, vol. 32, pp. 1114–1125, June 2014.

[45] K. Zhu, E. Hossain, and D. Niyato, "Pricing, spectrum sharing, and service selection in two-tier small cell networks: A hierarchical dynamic game approach," *IEEE Transactions on Mobile Computing*, vol. 13, pp. 1843–1856, Aug 2014.

[46] X. Kang, Y. K. Chia, S. Sun, and H. F. Chong, "Mobile data offloading through a third-party wifi access point: An operator's perspective," *IEEE Transactions on Wireless Communications*, vol. 13, pp. 5340–5351, Oct 2014.

[47] F. Pantisano, M. Bennis, W. Saad, S. Valentin, M. Debbah, and A. Zappone, "Proactive user association in wireless small cell networks via collaborative filtering," in *2013 Asilomar Conference on Signals, Systems and Computers*, pp. 1601–1605, Nov 2013.

[48] F. Pantisano, M. Bennis, W. Saad, S. Valentin, and M. Debbah, "Matching with externalities for context-aware user-cell association in small cell networks," in *2013 IEEE Global Communications Conference (GLOBECOM)*, pp. 4483–4488, Dec 2013.

[49] G. Athanasiou, P. C. Weeraddana, C. Fischione, and L. Tassiulas, "Optimizing client association for load balancing and fairness in millimeter-wave wireless networks," *IEEE/ACM Transactions on Networking*, vol. 23, pp. 836–850, June 2015.

[50] W. Wong, A. Thakur, and S. H. G. Chan, "An approximation algorithm for AP association under user migration cost constraint," in *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, pp. 1–9, April 2016.

[51] H. Zhou, S. Mao, and P. Agrawal, "Approximation algorithms for cell association and scheduling in femtocell networks," *IEEE Transactions on Emerging Topics in Computing*, vol. 3, pp. 432–443, Sept 2015.

[52] Y. Xu, R. Q. Hu, Y. Qian, and T. Znati, "Tradeoffs in video transmission over wireless heterogeneous networks: Energy, bandwidth and qoe," in *2015 IEEE International Conference on Communications (ICC)*, pp. 3483–3489, June 2015.

[53] Q. Yang, X. Deng, and B. Wang, "Energy cost minimization in green heterogeneous cellular networks with wireless backhauls," in *2016 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, pp. 698–704, Dec 2016.

[54] X. Huang, L. Shi, L. Chen, and Q. Chen, "Cost-effective interference coordination scheme in high dense small cell heterogeneous network," in *2015 10th International Conference on Communications and Networking in China (ChinaCom)*, pp. 27–31, Aug 2015.

[55] S. Kim, S. Choi, and B. G. Lee, "A Joint Algorithm for Base Station Operation and User Association in Heterogeneous Networks," *IEEE Communications Letters*, vol. 17, pp. 1552–1555, August 2013.

[56] D. Niyato, F. Adachi, P. Wang, and D. I. Kim, "Competitive cell association and antenna allocation in 5G massive MIMO networks," in *2015 IEEE International Conference on Communications (ICC)*, pp. 3867–3872, June 2015.

[57] M. Rugelj, U. Sedlar, M. Volk, J. Sterle, M. Hajdinjak, and A. Kos, "Novel Cross-Layer QoE-Aware Radio Resource Allocation Algorithms in Multiuser OFDMA Systems," *IEEE Transactions on Communications*, vol. 62, pp. 3196–3208, Sept 2014.

[58] V. F. Monteiro, D. A. Sousa, T. F. Maciel, F. R. M. Lima, E. B. Rodrigues, and F. R. P. Cavalcanti, "Radio resource allocation framework for quality of experience optimization in wireless networks," *IEEE Network*, vol. 29, pp. 33–39, Nov 2015.

[59] Y. H. Cho, H. Kim, S. H. Lee, and H. S. Lee, "A QoE-Aware Proportional Fair Resource Allocation for Multi-Cell OFDMA Networks," *IEEE Communications Letters*, vol. 19, pp. 82–85, Jan 2015.

[60] N. Ferdosian, M. Othman, B. M. Ali, and K. Y. Lun, "Multi-Targeted Downlink Scheduling for Overload-States in LTE Networks: Proportional Fractional Knapsack Algorithm With Gaussian Weights," *IEEE Access*, vol. 5, pp. 3016–3027, 2017.

[61] N. Ferdosian, M. Othman, B. M. Ali, and K. Y. Lun, "Fair-QoS Broker Algorithm for Overload-State Downlink Resource Scheduling in LTE Networks," *IEEE Systems Journal*, pp. 1–12, 2018.

[62] M. Li, P. N. Tran, H. K. Tütüncüoğlu, and A. Timm-Giel, "Coordinated radio resource allocation in LTE femtocell cluster considering transport limitations," in *2015 IEEE International Conference on Communications (ICC)*, pp. 3113–3118, June 2015.

[63] D. Wu, Q. Wu, Y. Xu, and Y. C. Liang, "QoE and Energy Aware Resource Allocation in Small Cell Networks with Power Selection, Load Management and Channel Allocation," *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, pp. 1–1, 2017.

[64] B. C. Chung, K. Lee, and D. H. Cho, "Proportional Fair Energy-Efficient Resource Allocation in Energy-Harvesting-Based Wireless Networks," *IEEE Systems Journal*, vol. PP, no. 99, pp. 1–11, 2017.

[65] J. Li, M. Peng, A. Cheng, Y. Yu, and C. Wang, "Resource allocation optimization for delay-sensitive traffic in fronthaul constrained cloud radio access networks," *IEEE Systems Journal*, vol. 11, pp. 2267–2278, Dec 2017.

[66] H. Shao, W. Jing, X. Wen, Z. Lu, H. Zhang, Y. Chen, and D. Ling, "Joint Optimization of Quality of Experience and Power Consumption in OFDMA Multicell Networks," *IEEE Communications Letters*, vol. 20, pp. 380–383, Feb 2016.

[67] B. Al-Manthari, H. Hassanein, N. A. Ali, and N. Nasser, "Fair Class-Based Downlink Scheduling with Revenue Considerations in Next Generation Broadband Wireless Access Systems," *IEEE Transactions on Mobile Computing*, vol. 8, pp. 721–734, June 2009.

[68] T. Taleb, N. Nasser, and M. P. Anastasopoulos, "An auction-based pareto-optimal strategy for dynamic and fair allotment of resources in wireless mobile networks," *IEEE Transactions on Vehicular Technology*, vol. 60, pp. 4587–4597, Nov 2011.

[69] A. R. Elsherif, W. P. Chen, A. Ito, and Z. Ding, "Resource allocation and inter-cell interference management for dual-access small cells," *IEEE Journal on Selected Areas in Communications*, vol. 33, pp. 1082–1096, June 2015.

[70] W. Ji, B. W. Chen, Y. Chen, and S. Y. Kung, "Profit improvement in wireless video broadcasting system: A marginal principle approach," *IEEE Transactions on Mobile Computing*, vol. 14, pp. 1659–1671, Aug 2015.

[71] K. D. Lee and T. S. P. Yum, "On pareto-efficiency between profit and utility in ofdm resource allocation," *IEEE Transactions on Communications*, vol. 58, pp. 3277–3285, November 2010.

[72] K. Wang, K. Yang, X. Wang, and C. S. Magurawalage, "Cost-effective resource allocation in c-ran with mobile cloud," in *2016 IEEE International Conference on Communications (ICC)*, pp. 1–6, May 2016.

[73] S. Sen, C. Joe-Wong, S. Ha, and M. Chiang, "Incentivizing time-shifting of data: a survey of time-dependent pricing for internet access," *IEEE Communications Magazine*, vol. 50, pp. 91–99, November 2012.

[74] L. Zhang, W. Wu, and D. Wang, "Time dependent pricing in wireless data networks: Flat-rate vs. usage-based schemes," in *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, pp. 700–708, April 2014.

[75] C. H. Chang, P. Lin, J. Zhang, and J. Y. Jeng, "Time dependent adaptive pricing for mobile internet access," in *2015 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 540–545, April 2015.

[76] Y. Li and F. Xu, "Trace-driven analysis for location-dependent pricing in mobile cellular networks," *IEEE Network*, vol. 30, pp. 40–45, March 2016.

[77] R. Schoenen, H. U. Sokun, and H. Yanikomeroglu, "Green Cellular Demand Control with User-in-the-Loop Enabled by Smart Data Pricing Using an Effective Quantum (eBit) Tariff," in *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, pp. 1–7, Sept 2016.

[78] S. Y. Jung and S. L. Kim, "Resource allocation with reverse pricing for communication networks," in *2016 IEEE International Conference on Communications (ICC)*, pp. 1–6, May 2016.

[79] Y. Song, C. Zhang, Y. Fang, and P. Lin, "Revenue maximization in time-varying multi-hop wireless networks: A dynamic pricing approach," *IEEE Journal on Selected Areas in Communications*, vol. 30, pp. 1237–1245, August 2012.

[80] B. Zeng, X. Fang, and L. Qing, "Utility-based dynamic revenue pricing scheme for wireless operators," in *2013 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 4612–4617, April 2013.

[81] J. Simão and L. Veiga, "Partial utility-driven scheduling for flexible sla and pricing arbitration in clouds," *IEEE Transactions on Cloud Computing*, vol. 4, pp. 467–480, Oct 2016.

[82] P. Trakas, F. Adelantado, and C. Verikoukis, "A novel learning mechanism for traffic offloading with small cell as a service," in *2015 IEEE International Conference on Communications (ICC)*, pp. 6893–6898, June 2015.

[83] P. Trakas, F. Adelantado, and C. Verikoukis, "Network and Financial Aspects of Traffic Offloading With Small Cell as a Service," *IEEE Transactions on Wireless Communications*, vol. 17, pp. 7744–7758, Nov 2018.

[84] Real Wireless, "The business case for urban small cells," Feb 2014.

[85] X. Lin, J. G. Andrews, and A. Ghosh, "Modeling, analysis and design for carrier aggregation in heterogeneous cellular networks," *IEEE Transactions on Communications*, vol. 61, pp. 4002–4015, September 2013.

[86] Y. Zhao, S. Wang, S. Xu, X. Wang, X. Gao, and C. Qiao, "Load balance vs energy efficiency in traffic engineering: A game theoretical perspective," in *2013 Proceedings IEEE INFOCOM*, pp. 530–534, April 2013.

[87] R. T. Maheswaran and T. Başar, "Nash equilibrium and decentralized negotiation in auctioning divisible resources," *Group Decision and Negotiation*, vol. 12, no. 5, pp. 361–395, 2003.

[88] P. R. Winters, "Forecasting sales by exponentially weighted moving averages," *Manage. Sci.*, vol. 6, pp. 324–342, Apr. 1960.

[89] M. N. Howell and T. J. Gordon, "Continuous action reinforcement learning automata and their application to adaptive digital filter design," *J. Eng. Applicat. Automat. Contr*, p. 254, 2001.

[90] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire, "The nonstochastic multiarmed bandit problem," *SIAM J. Comput.*, vol. 32, pp. 48–77, Jan. 2003.

[91] C. Joe-Wong, S. Sen, and S. Ha, "Offering supplementary network technologies: Adoption behavior and offloading benefits," *IEEE/ACM Transactions on Networking*, vol. 23, pp. 355–368, April 2015.

[92] 3GPP TR 36.872 V12.1.0., "Small cell enhancements for e-utra and e-utran - physical layer aspects," Dec 2013.

[93] 3GPP TR 36.814 V9.0.0., "Evolved universal terrestrial radio access (e-utra); further advancements for e-utra physical layer aspects," Mar 2010.

[94] A. El-Nashar, M. El-saidny, and M. Sherif, *Design, Deployment and Performance of 4G-LTE Networks: A Practical Approach.* Wiley Publishing, 1st ed., 2014.

[95] Cell at Auction, "Who we are," 2013.

[96] SteepSteel, "About us & Why?," 2016.

[97] M. Wellman, A. Greenwald, and P. Stone, *Autonomous Bidding Agents: Strategies and Lessons from the Trading Agent Competition.* Intelligent Robotics and Auton, MIT Press, 2007.

[98] A. Gotsis, S. Stefanatos, and A. Alexiou, "Ultradense networks: The new wireless frontier for enabling 5g access," *IEEE Vehicular Technology Magazine*, vol. 11, pp. 71–78, June 2016.

[99] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 32, pp. 1164–1179, June 2014.

[100] R. Jain, D. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," *CoRR*, vol. cs.NI/9809099, 1998.

[101] O. Günlük and J. Linderoth, "Perspective reformulation and applications," in *Mixed Integer Nonlinear Programming* (J. Lee and S. Leyffer, eds.), (New York, NY), pp. 61–89, Springer New York, 2012.

[102] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms, 3rd Edition.* MIT Press, 2009.

[103] 3GPP TR 37.901 V15.0.0., "User Equipment (UE) application layer data throughput performance," Mar 2018.

[104] 3GPP TR 23.203 V14.0.0., "Policy and charging control architecture," Jun 2016.

[105] V. Held, "White paper: Ten key rules of 5g deployment," *Nokia Solutions and Networks.*

[106] J. G. Andrews, "Seven ways that hetnets are a cellular paradigm shift," *IEEE Communications Magazine*, vol. 51, pp. 136–144, March 2013.

[107] P. Trakas, F. Adelantado, N. Zorba, and C. Verikoukis, "A quality of experience-aware association algorithm for 5G heterogeneous networks," in *2017 IEEE International Conference on Communications (ICC)*, pp. 1–6, May 2017.

[108] P. Trakas, F. Adelantado, and C. Verikoukis, "QoE-aware resource allocation for profit maximization under user satisfaction guarantees in HetNets with differentiated services," *to appear in IEEE Systems Journal*.

[109] P. Trakas, F. Adelantado, N. Zorba, and C. Verikoukis, "A QoE-Aware Joint Resource Allocation and Dynamic Pricing Algorithm for Heterogeneous Networks," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, pp. 1–6, Dec 2017.