Architecture and protocols for public safety users in the 5G cellular networks

Sarkis Moussa

A dissertation submitted in partial fulfillment for the degree of Doctor of Philosophy In Computer Science

Doctoral School 536
"Agrosciences et Sciences"
Laboratoire Informatique d'Avignon (LIA)

Presented by:
Sarkis Moussa

Architecture and Protocols for Public Safety Users in the 5G Cellular Networks

Defended on 06/30/2023, before the committee members:

Nader Mbarek	Professor	Université de Bourgogne	Reviewer
Tayeb Lemlouma	Associate Professor	Université de Rennes	Reviewer
Fen Zhou	Associate Professor	Avignon Université	Examiner
Yezekael Hayel	Professor	Avignon Université	Examiner
Abderrahim Benslimane	Professor	Avignon Université	Thesis supervisor

Abstract

Public Safety Networks (PSNs) are wireless communication systems designed to meet the needs of emergency responders, including firefighters, police, and many other Public Safety (PS) agencies. These networks are used to prevent or respond to incidents that pose a threat to people or property. Traditionally, these PSNs were supported by reliable, but low-rate radio technologies that provide limited services such as voice communication among Public Safety Users (PSUs). Consequently, their capability to take advantage of recent developments in wireless networks and broadband applications was restricted. At the forefront of wireless communication technologies, 5th Generation (5G) and beyond Cellular Networks (CNs), are ideal for this purpose due to their advanced infrastructure and tailored techniques developed for broadband services. Their capacity for high data transmission, low latency in data exchange, and ability to support a significant number of connected devices make them perfectly suited to overcome the limitations associated with PSNs.

Integrating PSNs into 5G can significantly improve the performance of PSUs. It enables PS agencies to respond more effectively to emergencies, improve communication among first responders, and access critical information in real-time. In this regard, the objective of this thesis is to develop models and architecture that guarantee effective communication among PSUs through the use of cellular resources. We consider different scenarios, including situations where resources are solely dedicated to PSUs and where resources are shared with primary users. In these scenarios, Non-Orthogonal Multiple Access (NOMA) technique and Device-to-Device (D2D) communication ensure an efficient allocation of limited resources for a larger number of PSUs. Additionally, in this thesis, we explore scenarios where cellular resources are not available (e.g. when the Base Station (BS) is not accessible). We develop strategies to maintain the continuous functioning of PSUs in such situations using the Multi-access Edge Computing (MEC) system.

To accomplish these objectives, we first focus on the formulation of a resource allocation problem in the in-band overlay D2D communication. The rationale for using the overlay mode is to guarantee the availability of Resource Blocks (RBs) that are dedicated exclusively to PSUs, minimizing therefore the interference between Cellular Users (CUs) and PSUs. Furthermore, we consider the NOMA technique for radio access, which allows multiple PSUs to share the same RBs, improving thereby the system performance in terms of spectral efficiency, achieved throughput, and number of PSUs accessing the network. This is achieved by implementing a heuristic that groups the PSUs that will share the same resources, so that the total amount of

bandwidth used is minimized. We then allocate sufficient power to each PSU using the Particle Swarm Optimization (PSO) algorithm.

As a second approach, we propose a novel scheme for the underlay D2D communication scenario. This scheme also relies on the NOMA technique and is based on a mixed integer nonlinear programming problem for sum throughput maximization. It takes into account the power budget, the users required rates, and the Successive Interference Cancellation (SIC) constraints. Since the maximization problem is computationally challenging, we design a heuristic algorithm that selects the appropriate CUs to share their resources with the PS clusters. Then, given this selection, we compute the optimal power allocation in each PS cluster using the Lagrange multiplier method.

During a disaster, the infrastructure might get damaged, in particular, BSs, resulting in the disruption of PSUs' access to the core network. To address this critical issue and ensure the continuous availability of PS services under any circumstances, we investigate the use of the Proximity-based Services (ProSe) standard which plays a crucial role in enabling D2D communication in licensed and unlicensed spectrum. Accordingly, we propose and validate a new architecture for PSNs using the Simu5G network simulator. This architecture uses the NOMA technique and places both, the ProSe function and the ProSe application server, in the MEC system that is installed in a relay station. This approach guarantees network availability for as many PSUs as possible, and enables them to access the required information with minimal latency in the licensed spectrum, while they continue to operate in the unlicensed spectrum securely and efficiently.

Keywords: *Public safety network, device-to-device communications, non-orthogonal multiple access, optimization, multi-access edge computing, proximity-based services, Simu5G.*

Résumé

Les réseaux de sécurité publique (PSNs) sont des systèmes de communication sans fil conçus pour satisfaire les besoins des intervenants d'urgence, notamment les pompiers, la police et de multiples autres organismes de sécurité publique (PS). Ces réseaux sont employés pour prévenir les incidents qui constituent une menace pour les personnes ou les biens, ou pour y répondre. Ils sont traditionnellement basés sur des technologies radio fiables, mais à faible débit, fournissant ainsi des services limités tels que la communication vocale entre les utilisateurs de la sécurité publique (PSUs). De ce fait, leur capacité de tirer parti des développements récents des réseaux sans fil et des applications à large bande était limitée. Les réseaux cellulaires de la cinquième génération (5G) et au-delà, qui sont à la pointe des technologies de communication sans fil, sont idéaux à cette fin grâce à leur infrastructure avancée et aux techniques adaptées mises en place pour offrir des services à large bande. Leur capacité de transmission de données à haut débit, leur faible latence lors de l'échange de données et leur capacité à prendre en charge un nombre important de dispositifs connectés les rendent parfaitement adaptés pour surmonter les limitations associées aux PSNs.

L'intégration des PSNs dans la 5G peut améliorer considérablement les performances des PSUs. Elle permet aux agences de PS de répondre plus efficacement aux situations d'urgence et d'accéder aux informations essentielles en temps réel. Elle améliore également la communication entre les premiers intervenants. À cet égard, l'objectif de cette thèse est de développer des modèles et une architecture qui garantissent une communication efficace entre les PSUs en utilisant les ressources cellulaires. Nous envisageons différents scénarios, y compris des situations où les ressources sont uniquement dédiées aux PSUs et des situations où les ressources sont partagées avec les utilisateurs primaires. Dans ces scénarios, la technique d'accès multiple non orthogonal (NOMA) et la communication de dispositif à dispositif (D2D) garantissent une allocation efficace des ressources limitées pour un plus grand nombre de PSUs. En outre, dans cette thèse, nous explorons des scénarios dans lesquels les ressources cellulaires ne sont pas disponibles (par exemple, lorsque la station de base n'est pas accessible). Nous développons des stratégies pour maintenir le fonctionnement continu des PSUs dans de telles situations, tout en employant le système de calcul de périphérie multi-accès (MEC).

Afin de réaliser ces objectifs, nous nous intéressons d'abord à la formulation d'un problème d'attribution de ressources dans la communication D2D superposée en bande sous licence. Ce mode superposé permet de garantir la disponibilité de blocs de ressources (RBs) dédiés exclusivement aux PSUs, minimisant ainsi l'interférence

entre les utilisateurs cellulaires (CUs) et les PSUs. En outre, nous considérons la technique NOMA pour l'accès radio, qui permet à plusieurs PSUs de partager les mêmes RB, améliorant ainsi la performance du système en termes d'efficacité spectrale, de débit atteint et de nombre de PSUs accédant au réseau. À cette fin, nous appliquons une heuristique qui regroupe les PSUs qui partageront les mêmes ressources, de sorte que la consommation de la bande passante soit réduite au minimum. Nous attribuons ensuite une puissance suffisante à chaque PSU à l'aide de l'algorithme d'optimisation par essaims de particules (PSO).

Dans une deuxième approche, nous proposons un nouveau schéma pour le scénario de communication D2D sous-couche. Ce schéma repose également sur la technique NOMA et est basé sur un problème de programmation mixte non linéaire à nombre entier pour la maximisation du débit total. Ce problème prend en compte les contraintes liées au budget de puissance, aux débits requis par les utilisateurs et à l'annulation successives des interférences (SIC). Comme il s'agit d'un problème de maximisation difficile à résoudre, nous concevons un algorithme heuristique qui sélectionne les CUs appropriées pour partager leurs ressources avec les clusters de PS. Ensuite, compte tenu de cette sélection, nous calculons l'allocation optimale de puissance dans chaque groupe de PS en utilisant la méthode du multiplicateur de Lagrange.

Lors d'une catastrophe, l'infrastructure peut être endommagée, en particulier les stations de base, ce qui perturbe l'accès des PSUs au réseau central. Pour résoudre ce problème majeur et garantir la disponibilité continue des services PS en toutes circonstances, nous examinons l'utilisation de la norme des services basés sur la proximité (ProSe). Cette norme joue un rôle crucial en permettant aux communications D2D de se dérouler dans les spectres avec et sans licence. En conséquence, nous proposons et validons une nouvelle architecture pour les PSNs à l'aide du simulateur de réseau Simu5G. Cette architecture utilise la technique NOMA et place la fonction ProSe et le serveur d'application ProSe dans le système MEC installé dans une station relais. Cette approche garantit la disponibilité du réseau pour le plus grand nombre possible de PSUs et leur permet d'accéder aux informations requises avec une latence minimale dans le spectre sous licence, tout en assurant le fonctionnement dans le spectre sans licence en toute sécurité et de manière efficace.

Mots-clés: *Réseaux de sécurité publique, communications de dispositif à dispositif, accès multiple non orthogonal, optimisation, calcul de périphérie multi-accès, services basés sur la proximité, Simu5G.*

Acknowledgements

I would like to express my sincere gratitude to all those who have contributed to the completion of this thesis. Their support, guidance, and motivation have been invaluable along the way.

In particular, I would like to extend my heartfelt appreciation to my lovely parents and siblings for their unconditional love, support, and encouragement, which have been the cornerstone of my progress in completing this thesis. I am truly lucky to have such a loving and supportive family. I cannot express enough gratitude for the trust you have placed in me. This thesis is as much yours as it is mine, and I dedicate it to each and every one of you. Thank you for being my rock and my source of inspiration.

Contents

List of Figures

List of Tables

List of Acronyms

3GPP Third Generation Partnership Project 29

5G 5th Generation 22

AF Application Function 35

AI Artificial Intelligence 103

AMF Access and Mobility management Function 34

API Application Programming Interfaces 32

AR Augmented Reality 103

AUSF AUthentication Server Function 35

AWGN Additive White Gaussian Noise 39

BER Bit Error Ratio 92

BS Base Station 22

CDMA Code Division Multiple Access 29

CH Cluster Head 39

CN Cellular Networks 21

CPU Central Processing Unit 35

CU Cellular Users 21

D2D Device-to-Device 22

DMR Digital Mobile Radio 28

DN Data Network 35

DNS Domain Name System 35

EASDF Edge Application Server Discovery Function 35

eNodeB evolved NodeB 30

EPC Evolved Packet Core 30

ETSI European Telecommunications Standards Institute 35

E-UTRAN Evolved-UMTS Terrestrial Radio Access Network 30

eMBB enhanced Mobile BroadBand 22

FDMA Frequency Division Multiple Access 29

FirstNet First responder Network authority 29

FPA Fixed Power Allocation 43

FTPA Fractional Transmit Power Allocation 43

gNodeB next generation NodeB 88

GSM Global System for Mobile communication 29

HSS Home Subscriber Server 30

IoT Internet of Things 33

IP Internet Protocol 30

ISM Industrial, Scientific and Medical band 32

ITU International Telecommunications Union 22

KKT Karush-Kuhn-Tucker 23

LMR Land Mobile Radio 28

LTE Long-Term Evolution 28

LTE-A Long-Term Evolution-Advanced 51

MEC Multi-access Edge Computing 22

MME Mobile Management Entity 30

mMTC massive Machine-Type Communications 22

mmWave millimeter Wave 33

NAS Non-Access Stratum 34

NEF Network Exposure Function 35

NF Network Functions 34

NFV Network Functions Virtualization 33

NOMA Non-Orthogonal Multiple Access 22

NR New Radio 91

NRF Network Repository Function 35

NrNic NR Network interface card 91

NrUe New radio User equipment 91

NS Network Slicing 33

NSACF Network Slice Admission Control Function 35

NSSAAF Network Slice-specific and SNPN Authentication and Authorization Function 35

NSSF Network Slice Selection Function 35

OFDMA Orthogonal Frequency Division Multiple Access 23

OMA Orthogonal Multiple Access 36

PCF Policy Control Function 35

PCRF Policy and Charging Rules Function 30

PDN Packet Data Networks 30

P-GW Packet data network GateWay 30

P25 Project 25 28

PMR Private Mobile Radio 28

PPDR Public Protection and Disaster Relief 28

ProSe Proximity-based Services 23

PS Public Safety 21

PSN Public Safety Networks 21

PSO Particle Swarm Optimization 23

PSU Public Safety Users 21

QoS Quality of Service 21

RAM Random-Access Memory 35

RB Resource Blocks 24

SBA Service-Based Architecture 34

SC Superposition Coding 38

SDN Software-Defined Networking 34

S-GW Serving GateWay 30

SIC Successive Interference Cancellation 23

SINR Signal-to-Interference-plus-Noise Ratio 52

SMF Session Management Function 34

TDMA Time Division Multiple Access 29

TETRA TErrestrial Trunked RAdio 28

TETRAPOL TErrestrial Trunked RAdio POLice 28

UAV Unmanned Aerial Vehicles 103

UDM Unified Data Management 35

UE User Equipment 29

UPF User Plane Functions 34

uRLLC ultra-Reliable Low-Latency Communications 22

V2V Vehicle-to-Vehicle 64

Chapter 1

Introduction

Contents

1.1 Thesis Context

Public Safety Networks (PSN) consist of dedicated communication systems that are specifically designed to cater to the needs of Public Safety (PS) agencies, such as police, firefighters, etc. These networks have traditionally been designed to provide voice messages and calls that can be reliably transmitted during emergencies or disasters. However, with advances in communication technologies, there is a growing necessity to expand the capabilities of PSNs to improve their Quality of Service (QoS) by adapting them to new technologies and services. These improvements include supporting high-quality real-time video, as well as accessing and sharing accurate and timely information in critical situations. To achieve this, sufficient resources should be allocated to Public Safety Users (PSU).

The allocation of resources to PSUs can be accomplished either through dedicated PSNs or through the sharing of Cellular Networks (CN) resources with Cellular Users (CU). The establishment of dedicated PSNs poses challenges that must be addressed. These include the deployment of new infrastructure, the limited spectrum availability, and the guarantee of the cost-effectiveness and sustainability of the network [1]. One promising solution to overcome these challenges is to enable PSUs to share CNs with CUs. This approach allows for addressing the aforementioned issues by leveraging the existing infrastructure already in place, thus reducing the need for extensive deployment of new infrastructure. Additionally, the CNs provide a pool of available spectrum, which can be efficiently utilized to meet the QoS requirements of the PSUs.

As such, the integration of PSNs into 5th Generation (5G) CNs is justified as these networks are expected to accommodate a wide range of applications and use cases. The International Telecommunications Union (ITU) has classified 5G networks into three categories, namely enhanced Mobile BroadBand (eMBB), which offers high-speed downloads and increased data transmission capacity. ultra-Reliable Low-Latency Communications (uRLLC), which provides nearly instantaneous responses and minimizes connection latency. And finally, massive Machine-Type Communications (mMTC), which can accommodate a significant number of connected devices. PS use cases, known as mission-critical applications, fall under the uRLLC and mMTC categories [2]. This integration leverages the capabilities of 5G networks to provide reliable and efficient communication services for PSUs. However, sharing of CNs' resources also introduces certain considerations that need to be addressed. One key challenge is interference management, ensuring that the interference will not cause service degradation for both PSUs and CUs [3]. In addition, it becomes crucial to ensure the availability of resources in various PS scenarios [4]. Thus, effective interference management techniques must be conceived to allocate resources, manage interference, and ensure resource availability according to the specific requirements of each PS use case.

1.2 Motivations and Contributions

According to a recent prediction [5], global internet traffic is expected to increase dramatically and reach an estimated 150.7 exabytes per month by 2023. This is a substantial amount of data that will be transmitted over the internet and is expected to grow at a rate of 24% up to 2026. However, this increased demand for internet connectivity and the proliferation of connected devices come with many challenges, including interference management and the shortage of available spectrum, which affects not only commercial applications but also PSNs' communication [6]. This brings up problems such as congestion, slower data speeds, dropped calls, and increased latency, which can hinder emergency responders' ability to communicate effectively during critical events. Hence, to respond to this ever-growing traffic and its requirements, it is essential to adopt measures that ensure more efficient use of the available spectrum. This could involve optimizing the existing wireless infrastructure, developing new technologies to enable more efficient use of the available spectrum, etc. Without these measures, the continued growth of internet traffic could lead to significant challenges in providing reliable and high-quality internet services to both commercial and PS users.

Applying appropriate technologies that facilitate the sharing of CNs resources between CUs and PSUs is crucial to fulfill the needs of PSUs. This includes implementing Device-to-Device (D2D) communication, adopting Non-Orthogonal Multiple Access (NOMA) technique, and deploying Multi-access Edge Computing (MEC) system. These technologies play a key role in optimizing spectrum usage, reducing latency, and providing the required QoS for PSUs. D2D communication enables direct communication between nearby devices without the need for a Base Station (BS), thus reducing the load on the network and providing faster and more reliable

connections. The NOMA technique enables multiple users to share the same frequency band simultaneously, thereby improving spectral efficiency and maximizing the use of available resources. MEC system allows for the offloading of computationally intensive tasks to the edge of the network, which reduces latency and improves performance. The integration of these technologies into the existing network infrastructure will enhance the capacity of PSNs and meet the increasing demands for reliable and efficient communication of PSUs.

In this thesis, we study a wireless system containing CUs, known as primary users, and PSUs organized in clusters, known as secondary users. These PSUs aim to access the network by exploiting the unused resources of the CUs through the overlay D2D communication, thus eliminating the interferences experienced by CUs. Furthermore, we employ the NOMA technique to enhance PSUs' performance and spectral efficiency. We formulate and solve an optimization problem that involves radio resource allocation and power distribution for PS clusters. When clusters experience a heavy load, PSUs can be partitioned into smaller groups. These groups will be able to share the same resources. We use a heuristic algorithm to form these groups and determine the number of radio resources to be allocated to each cluster. Then, we apply the Particle Swarm Optimization (PSO) algorithm to assign power to each NOMA group, with the aim of maximizing the sum-throughput. The simulation results confirm that our proposed approach outperforms the Orthogonal Frequency Division Multiple Access (OFDMA) technique.

Secondly, we explore a resource allocation method for PSUs that takes into account both underlay D2D communication and the NOMA technique, while also being aware of the interference that will occur at CUs level. Our primary contribution is conceiving a mixed integer nonlinear programming model under power, minimum rate, and Successive Interference Cancellation (SIC) constraints. As a solution to the formulated problem, we first propose a heuristic algorithm for the appropriate matching between CUs and PS clusters, taking into account the interference threshold of the CUs. Then, we apply the Karush-Kuhn-Tucker (KKT) conditions, for an optimal power allocation for the PSUs inside each cluster. We conduct extensive simulations to demonstrate the effectiveness of our approach compared to OFDMA and other state-of-the-art approaches.

Finally, we present a novel architecture that provides uninterrupted PS services, even in the event of infrastructure damage during a disaster. The architecture we propose utilizes the NOMA technique and integrates the Proximity-based Services (ProSe) standard into the MEC system. By incorporating ProSe into the MEC system, we can ensure that the needs of PSUs are met in both the in-band and out-of-band scenarios while also minimizing latency. To evaluate our proposed architecture, we use the Simu5G simulator. Our findings demonstrate that NOMA increases the spectral efficiency, allowing more PSUs to be served with fewer resources. Moreover, by integrating ProSe into the MEC system, PSUs can obtain the necessary information with minimal latency and high reliability in both in-band and out-of-band scenarios.

1.3 Organization

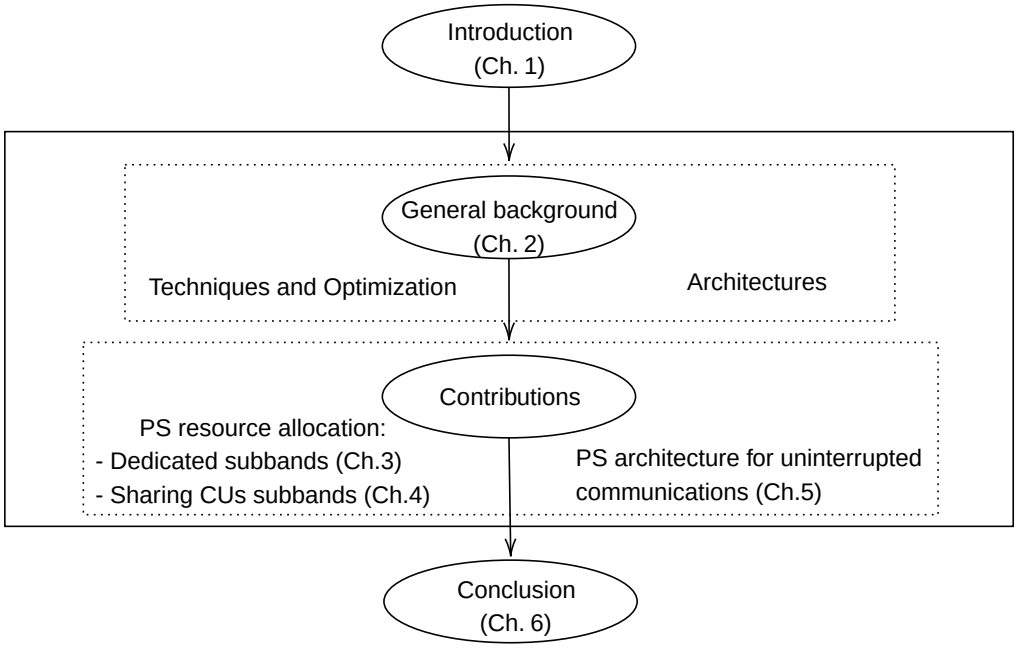

Figure 1.1: Thesis organization

The thesis is organized into six chapters. Figure 1.1 summarizes the organization of the manuscript, which is detailed as follows:

• Chapter 2 provides a general overview of the topics that are relevant to our research. The first part of the chapter offers an overview of PSNs, as well as different CNs architectures. The second part is dedicated to discussing the different multiple access techniques and optimization algorithms that form the basis of this thesis.

• In Chapter 3, our focus is on allocating subbands and power to meet the rate requirements of PSUs. These subbands consist of cellular Resource Blocks (RB) that are unoccupied by accus and can be utilized by PSUs without causing interference to the primary users. Our proposed approach leverages NOMA and in-band frequencies to allocate the resources to PSUs. The primary objective is to maximize the sum-throughput while ensuring that each PSU achieves its required rate.

• Chapter 4 focuses on how to allocate the necessary resources to PSUs while simultaneously sharing them with CUs. Since both PSUs and CUs use the same resources, the aim is to ensure that the resulting interference from PSUs is carefully managed to avoid disrupting the functioning of primary users. Subsequently, we ensure that the remaining resources meet the requirements of the PSUs. Various metrics are evaluated, including total and average user throughput, fairness, and outage probability.

• We present our proposed PS architecture in Chapter 5. It takes into account the NOMA system and employs the ProSe standard in the MEC system. This

architecture aims to ensure spectral efficiency and continuity of PSUs operation, especially in a disaster situation where connection to the BS may be lost. To assess the performance of the proposed architecture, we use the Simu5G simulator to evaluate the spectrum efficiency and latency.

• Finally, we summarize the work presented in the previous chapters and outline directions for future work in Chapter 6.

1.4 Publications

Journal Publication

• S. Moussa, A. Benslimane, R. Darazi et C. Jiang, "Power allocation based NOMA and underlay D2D communication for Public Safety Users in the 5G Cellular Network," IEEE Systems Journal, 2023.

Conference Publications

• S. Moussa and A. Benslimane, "Design of MEC-based Architecture Integrating ProSe and NOMA for Public Safety Users in 5G and Beyond Cellular Networks, " IEEE IWCMC, June 2023, Marrakesh, Morocco.

• S. Moussa, A. Benslimane, R. Darazi and C. Jiang, "Resource allocation for Public Safety Users in the 5G Cellular Network," IEEE Globecom, 2021, Madrid, Spain.

Chapter 2

A background study providing the context for this thesis

Contents

In this chapter, we discuss relevant topics to the work presented in this thesis. First, we provide an overview of the PSNs and the CNs, namely: Long-Term Evolution (LTE), 5G, and beyond 5G. We present different architectures and functions that have been used in the CNs. We also provide a general background on D2D communication and the MEC system, as well as the difference between in-band and out-of-band D2D communication. Next, we compare the OFDMA technique, which has been adopted in LTE and 5G CNs, and NOMA technique, which is a promising technique for future wireless communication networks and is the main topic of this thesis. The difference in frame structure between LTE and 5G is also discussed in this chapter. Finally, we present previous studies that have explored optimization theory in the context of user grouping and power allocation within NOMA. Additionally, a further explanation of PSO and Lagrange multipliers is provided, as these methods are employed throughout this thesis.

2.1 Public Safety and Cellular Networks

2.1.1 Public Safety Networks

PSNs are communication networks designed specifically to support PS and emergency response operations. Generally, public agencies administer these networks. Their primary goal is to establish reliable and secure communication channels among first responders, such as police, firefighters, and others. PSNs offer a wide range of services designed to support PS operations. These services include voice and data communication, location tracking, interoperability among different agencies, etc. Despite the interchangeability between the terms PS communication, Public Protection and Disaster Relief (PPDR) [7] and mission-critical communication [8], they are slightly different. PS communication systems provide day-to-day protection and support for the general public, while PPDR and mission-critical systems are designed for specific scenarios. Mission-critical communication refers to communication in high-risk emergencies, while PPDR focuses on communication systems and technologies employed during disaster relief and recovery missions.

Traditionally, PSNs have relied on radio technologies such as Private Mobile Radio (PMR) and Land Mobile Radio (LMR), which include systems like TErrestrial Trunked RAdio (TETRA), Digital Mobile Radio (DMR), Project 25 (P25) and TErrestrial Trunked RAdio POLice (TETRAPOL). These digital communication systems offer more advanced features, such as encryption and data transmission, compared to early analog systems. Although these technologies have been successful in providing reliable voice communication, their utilization of low data rates limits their capacity to take advantage of the most recent developments in wireless networks and broadband applications. For instance, the P25 and TETRAPOL technologies can achieve a maximum data rate of around 10 Kbps, while TETRA Release 1 can achieve a maximum of around 30 Kbps. For more information, Jarwan et al. [8] present the technical characteristics of the above systems, containing information on frequency bands, channel bandwidth, access method, modulation technique, peak data rates, and supported applications.

It is crucial to provide adequate resources to meet the demands of PSUs to access and share real-time video and accurate information about events. There are ongoing efforts to meet these demands [4]. In 2012, the First responder Network authority (FirstNet) aimed to provide emergency responders with the first nationwide broadband network dedicated to PS [9]. It was granted a license to use 700 MHz broadband spectrum throughout the United States. The Third Generation Partnership Project (3GPP) community subsequently developed a set of standards for mission-critical functions in broadband networks, including LTE and 5G. Releases 12 [10] and 13 [11] addressed the specific PS requirements, including group calling, ProSe, and push-to-talk. ProSe has recently regained attention in release 17 [12] for its implementation in 5G systems by focusing on three main functions: 5G ProSe direct discovery, 5G ProSe direct communication, and 5G ProSe User Equipment (UE)-to-network relay. CNs are an ideal fit for PSNs, as their infrastructure is already adapted for broadband services, providing reliable communication, low latency for data exchange, and the ability to adapt dynamically to changes.

2.1.2 Cellular Networks

CNs are wireless communication networks consisting of a series of cells, each of which is covered by at least one fixed BS. The latter provides cell coverage, enabling wireless transmission of voice, data, and other content. CNs have evolved through a succession of technology generations, each offering new capabilities. The evolution of these generations is described very briefly as follows: 1G technology, which began in 1980, only allowed analog voice calls using the Frequency Division Multiple Access (FDMA) technique [13]. Then in the early 1990s, with the Global System for Mobile communication (GSM) standard, 2G technology introduced digital voice services and internet connectivity with the Time Division Multiple Access (TDMA) technique [14]. Later on, mobile TV, video telephony, and video conference were provided by 3G technology, using the Code Division Multiple Access (CDMA) technique. Further evolution occurred with the advent of 4G technology, known as LTE technology. It has led to additional enhancements by enabling higher data rate applications such as high-definition television, cloud computing, and video gaming. It also introduced the concepts of OFDMA technique. Finally, 5G technology has emerged, it is intended for applications that require very high data rates, low latency, and a massive number of connected devices, using the same OFDMA technique as 4G technology [15].

Given that D2D communication is the focus of this thesis, and considering that this technology was introduced with the advent of LTE, in the following we shall discuss the LTE architecture, after which we cover D2D communication. Moreover, we will also approach the 5G technology to cover the concept of edge computing. We will build on it to develop our architecture.

2.1.2.1 Long-Term Evolution architecture

LTE technology has offered higher data rates, greater coverage area, lower latency, and better spectral efficiency, compared to previous generations. These improvements were achieved through a new architecture and radio interface. In ad-

dition, this technology uses OFDMA modulation and supports flexible bandwidths that take values in {1.4, 3, 5, 10, 15 and 20} MHz [16]. Figure 2.1 shows the LTE network architecture [8], which is organized as follows: The UE domain, the Evolved-UMTS Terrestrial Radio Access Network (E-UTRAN) domain, the Evolved Packet Core (EPC) domain and finally the service domain. Each function in these domains has its role in establishing and maintaining communication between two UEs in the LTE network.

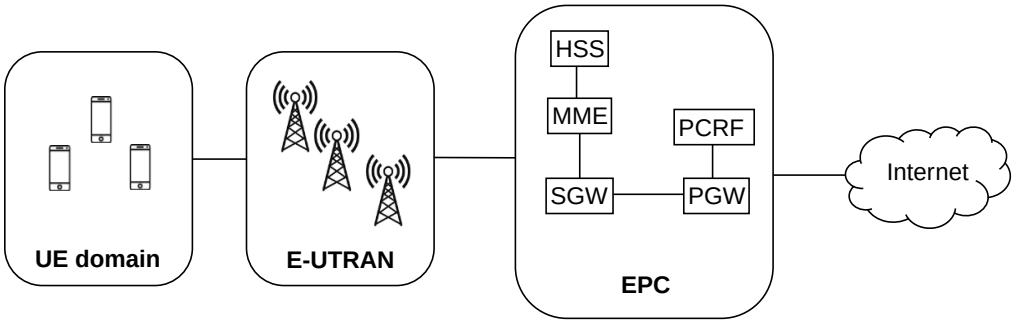

Figure 2.1: LTE network architecture

The E-UTRAN is the radio access part of the network and consists of several BSs known as evolved NodeB (eNodeB). It is in charge of reliable radio frequency transmission between the LTE network and the UEs. The core network is composed of several functions, including the following [17]: the Serving GateWay (S-GW), which is the main part of the EPC and the link point between the E-UTRAN and the EPC. The Mobile Management Entity (MME) is the control node that manages the signaling between the UE and the EPC. It also handles session and mobility management at the control plane level. Another component of the LTE network is the Packet data network GateWay (P-GW). It performs the assignment of Internet Protocol (IP) addresses to UEs and connects the network to other Packet Data Networks (PDN). It also ensures QoS satisfaction via the Policy and Charging Rules Function (PCRF). The Home Subscriber Server (HSS) has the duty of maintaining user information, such as subscriptions and PDNs information, while also offering authentication services and supporting mobility management [18, 19].

2.1.2.2 Device-to-Device communication

D2D communication in CNs allows nearby UEs to communicate over a direct link using CN frequencies, rather than passing radio signals through the BS and the core network [20]. To meet the PS requirements, 3GPP has developed a set of standards for critical functions in broadband networks (LTE and 5G). These standards help PSUs that rely on legacy technologies and communications to benefit from the latest technologies and better increase their bandwidth while reducing their communication latency. ProSe is one such standard deployed in the LTE architecture to enable D2D communication for PSUs in CNs [10]. The main roles of ProSe can be summarized as follows: First, it assists in user discovery and network-assisted communication when the communications are within LTE network coverage.

Second, if two devices are in close proximity and one of them is out of LTE network coverage, ProSe assists in applying the device to the network relay. Finally, when mobile devices have no LTE network coverage, direct communication can be applied via the LTE Direct.

The ProSe function and the ProSe application server were integrated into the existing LTE architecture to initiate the D2D communication concept for PSUs in CNs. Figure 2.2 shows the updated LTE architecture following the introduction of ProSe. LTE-Uu, S1 and S6a are LTE network reference points used for interaction between UEs, E-UTRAN and some core network functions. While the PC reference points are added to provide connectivity for the ProSe standard.

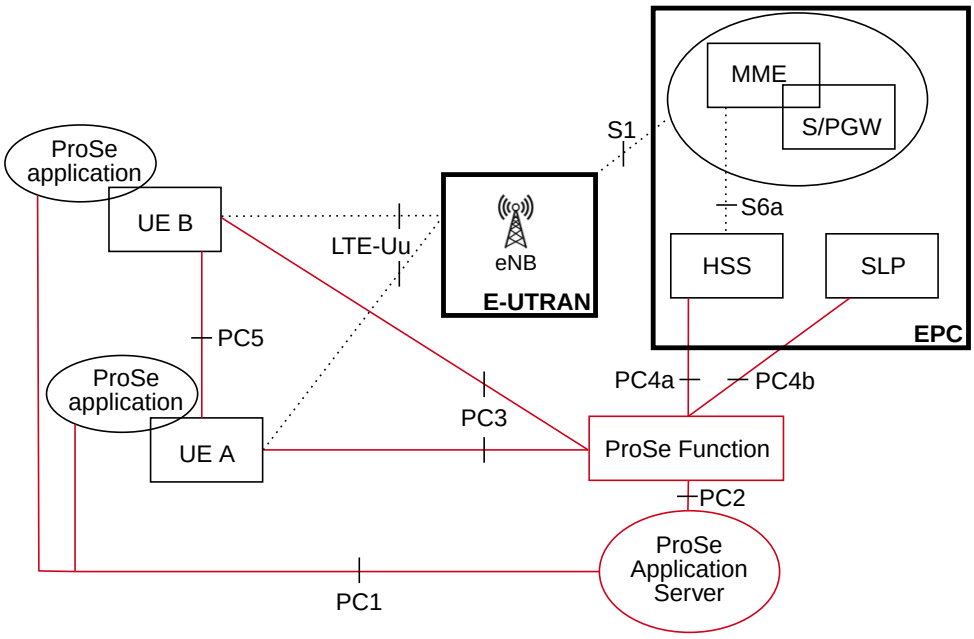

Figure 2.2: LTE-ProSe Architecture

The ProSe Function consists of three main sub-functions that cooperate to ensure continuous and secure utilization of ProSe for nearby users [21]. They are defined as follows:

- Direct Provisioning Function: It is intended to provide the UE with the essential parameters for D2D discovery and communication, whether the UE is served by the E-UTRAN or not. In summary, the direct provisioning function ensures the proper configuration of UEs to support ProSe services.
- Direct Discovery Name Management Function: It authorizes and secures discovery requests of ProSe subscribers using their relative data stored in the HSS.
- EPC-level Discovery ProSe Function: provides network-related functionalities, such as authorization, charging, and subscriber information management.

App Server: It is the entity from which the user downloads ProSe-related applications that offer services based on the corresponding Application Programming Interfaces (API). It interacts with both the UE and the ProSe function and stores the user profile for the applications based on the ProSe service.

The establishment of direct communication links can be classified into two categories [22]: in-band, where both cellular and D2D traffic make use of the same licensed cellular spectrum. This category can be further subdivided into underlay D2D communication, where cellular and D2D users share the available spectrum simultaneously, and overlay D2D communication, where part of the available spectrum of the CUs is dedicated to D2D. The second category is out-of-band, where D2D traffic leverages unlicensed spectrum (i.e., Industrial, Scientific and Medical band (ISM) spectrum) to establish direct communication. It can be divided into controlled and autonomous communication [23]. Under controlled mode, the D2D communication is monitored by the CNs, whereas in the case of autonomous communication, the users themselves are in charge of controlling the communication without involving the BS. Figure 2.3 depicts the representation of this classification [24].

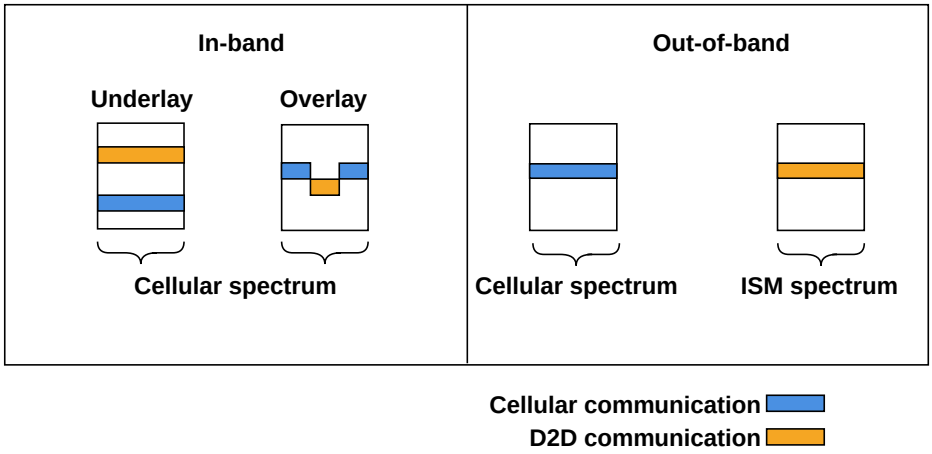

Figure 2.3: D2D scenarios

In the " autonomous" communication case, the concept of D2D communication is similar to that of the mobile Ad-hoc network, while in the other cases, this communication is controlled by the eNodeB. Each case of communication has its benefits and drawbacks; in "underlay" communication, the cellular spectrum unused by CUs is reused by D2D communication, which improves the spectrum efficiency, however, the interference becomes difficult to manage. In "overlay" communication, the interference problem is almost solved by dedicating part of the cellular spectrum to D2D communication, but this can result in decreasing the spectrum efficiency.

2.1.2.3 5th Generation architecture

5G radio networks are the evolution of LTE-Advanced networks. They transform the way of communication by using a wide range of wireless frequencies. 5G allows

for operation in three main frequency bands: low, medium, and high bands [25]. Over the low bands, 5G can provide better coverage with lower data rates, it can handle increased network capacity and support billions of devices connecting to the internet every day. These bands are ideal for voice applications and for improving coverage in rural areas, for example. While across the medium bands, 5G offers a balance of coverage and speed, making it a good choice for a variety of applications and use cases such as broadband connectivity that enhances the user experience of mobile devices, smart factories, etc. Whereas over the high bands, known as millimeter Wave (mmWave), 5G offers extremely high speed with limited coverage, making it the right option for emerging technologies that require increased throughput and fast download speeds, such as augmented and virtual reality, smart cities, and more.

Figure 2.4 shows some 5G associated applications and their use cases that can be classified as follows [26]: eMBB, which offers faster download speeds and throughput, uRLLC, which enables near-instantaneous responses and limits latency among connections, and mMTC, which supports the billions of devices that make up the Internet of Things (IoT) [27].

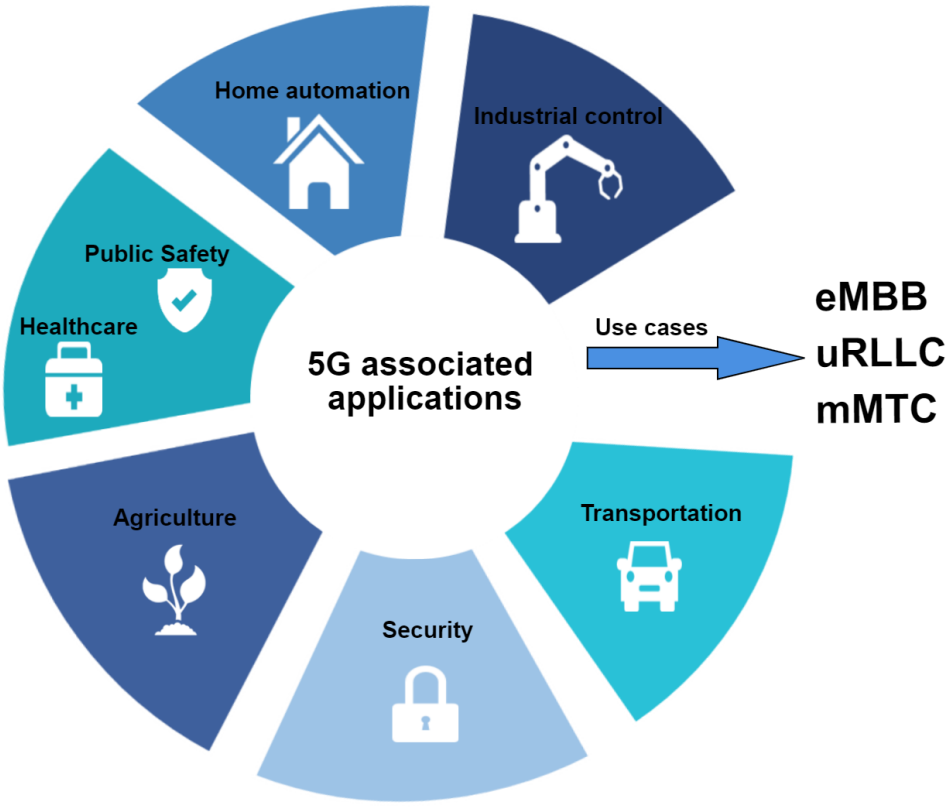

Figure 2.4: 5G Use cases and applications

In contrast to previous generations of CNs, which were solely applicable to cellular devices, 5G is a fundamentally different concept that has evolved towards enabling the IoT. Numerous technologies have emerged from this perspective [28], such as MEC, Network Slicing (NS), Network Functions Virtualization (NFV) [29],

Software-Defined Networking (SDN), and many others, enabling many applications, including mission-critical communication, connected vehicles, traffic management, smart grids, environmental monitoring, smart homes and buildings, and smart cities.

Figure 2.5 illustrates the 5G Service-Based Architecture (SBA), in which many functions are the same as in the LTE architecture under a different name, and many new functions have been added [30]. The remarkable feature of the SBA is the decoupling of the data plane from the control plane. The Network Functions (NF) related to the control plane are connected via service-based interfaces (referred to as Network followed by the name of the function, e.g. Namf, Nsmf, etc.), and their services can be invoked via API. Whereas the point-to-point connection is being deployed in the data plane. One of the main advantages of this configuration is the distributed behavior of the data plane functions that can be located closer to the users, thus reducing latency and offloading the network.

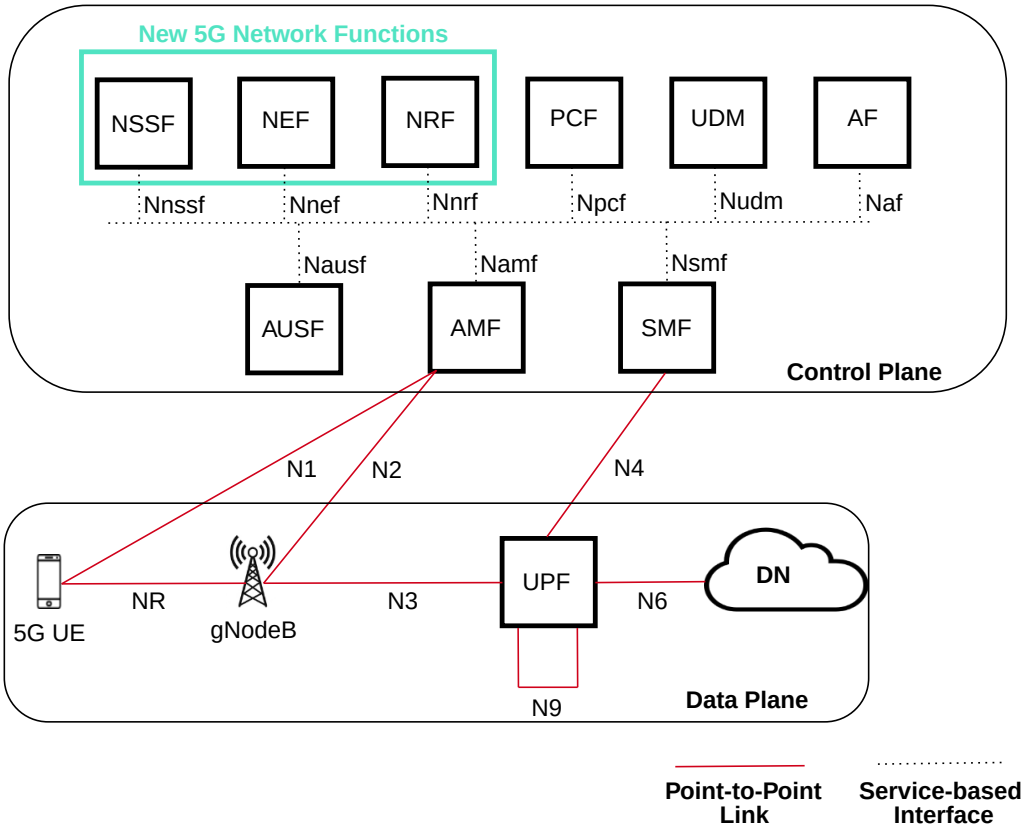

Figure 2.5: 5G Service-Based Architecture

In this paragraph, we briefly describe the 5G functions and their functionalities and show their corresponding functions in the LTE architecture. The Access and Mobility management Function (AMF) secures UE data, maintains subscriber mobility, and controls Non-Access Stratum (NAS) communication between the core network and the UEs. The Session Management Function (SMF) manages sessions and assigns IP addresses to UEs in addition to selecting and controlling User Plane Functions (UPF). These two functions have the same functionality as MME, S-GW,

and P-GW control planes in LTE. Processing of authentication is the only focus of the AUthentication Server Function (AUSF). Policy Control Function (PCF) and Unified Data Management (UDM) are analogous to PCRF and HSS in LTE networks, respectively. The Network Exposure Function (NEF) serves as an intermediary between services and NFs and helps structure the exposure of network data and services privately and securely. Network Repository Function (NRF) acts as a directory service and provides the necessary information when it receives a NF discovery request from other NF instances. Network Slice Selection Function (NSSF) assists in selecting the appropriate network slices for UEs and assigning the necessary AMF. UPF provides services suitable for user plane processing, such as routing, packet forwarding, and QoS management, etc., similarly to S-GW and P-GW user planes in LTE. It plays a key role in the deployment of MEC in the 5G networks. Application Function (AF) provides top-level services to users by interacting with other NFs and with application servers, ensuring that the appropriate services are supplied to UEs. Data Network (DN) represents the external data network or the internet. Network Slice-specific and SNPN Authentication and Authorization Function (NSSAAF) [31] and Network Slice Admission Control Function (NSACF) [32] are two newly functions added to support network slicing. NSSAAF is in charge of service authentication and authorization, while NSACF monitors and controls the number of registered UEs and the number of protocol data unit sessions per network slice. Additionally, the Edge Application Server Discovery Function (EASDF) [33] has direct UPF connectivity and handles Domain Name System (DNS) messages according to SMF instructions.

2.1.2.4 Multi-Access Edge Computing

The concept of MEC involves shifting computational processes from a central location to the edge of the network close to the end user, thus reducing the amount of data to be transmitted to the cloud or data center. This paves the way for new types of applications by offering extremely low latency and improved bandwidth [34]. The 5G SBA interactions between NFs allow the mapping of MEC onto AF, then based on the configured policies, this AF can use the services and information offered by other NFs. As stated by European Telecommunications Standards Institute (ETSI), a MEC system is made up of two layers, the MEC system level, and the MEC host level [35]. Figure 2.6 represents an example of integrating the MEC system in 5G architecture.

The MEC host holds the virtualization infrastructure where the MEC applications and MEC services are executed, it is most often deployed in the DN. The MEC services (e.g., location services and radio network information service) are consumed by MEC applications using standard MEC APIs [36]. These APIs allow MEC services to be explored by MEC applications, similar to how the SBA explores NFs and their services. Two entities, the MEC platform manager and virtualization infrastructure manager, monitor and manage the status of the MEC host. They provide information such as available computing resources (e.g., Random-Access Memory (RAM), Central Processing Unit (CPU) and disk), available MEC services (e.g., location service and radio network information service), etc. to the MEC orchestrator

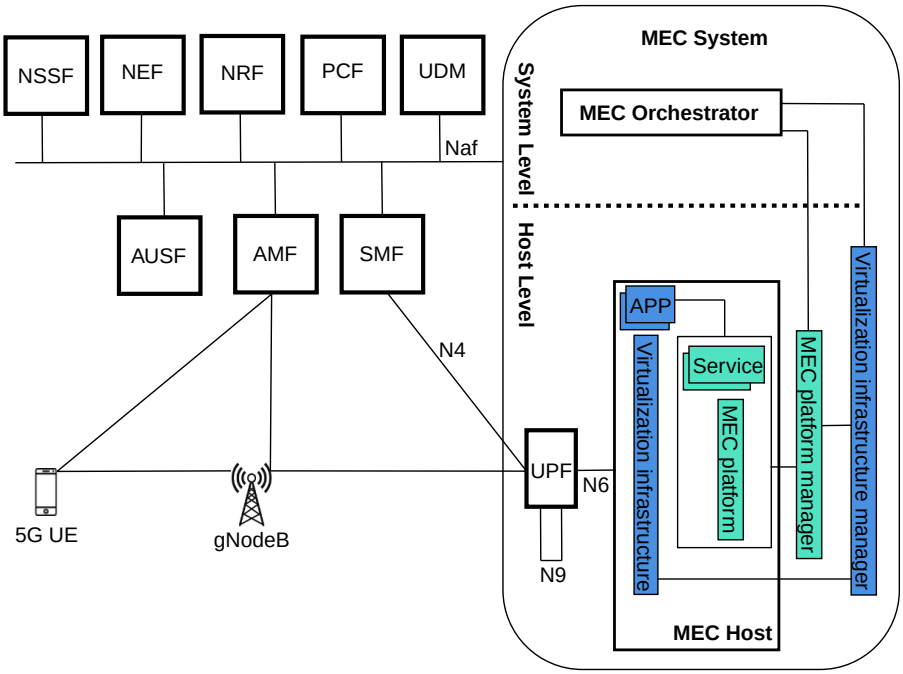

Figure 2.6: MEC deployment in 5G

during the operation of MEC applications [37].

A key component of the MEC system level is the MEC orchestrator, which is considered as a management entity that provides an overview of the entire MEC system. To meet UEs latency requirements, the MEC orchestrator selects the best MEC hosts based on available resources and services [38]. It also triggers the instantiation, termination, and relocation of applications as needed. It interacts with the core network functions, either directly if trusted by 3GPP, or via the NEF if not.

2.2 Multiple access techniques in LTE, 5G and beyond 5G

With the rapid development of IoT, 5G networks are faced with a massive amount of connected devices and need to meet their latency requirements and different types of services. As a result, spectrum efficiency becomes one of the main challenges of this connectivity growth. Many technologies have been proposed for this purpose, one of them being the NOMA technique. As opposed to the currently used Orthogonal Multiple Access (OMA) technique, NOMA is a very promising technique to solve the spectral efficiency by allowing users to non-orthogonally share the same resources [39].

2.2.1 Orthogonal Multiple Access technique

Prior to the arrival of orthogonality in access technology, the bandwidth was divided into different time slots TDMA or different frequency bands FDMA, etc., and users could use these resources separately. However, none of these techniques can meet the high demands of today's radio access systems.

In the LTE and 5G networks, the OFDMA technique has been considered. Under this technique, the available bandwidth is divided into several orthogonal subcarriers, grouped into frames. This allows users to share the same frequency band using different subcarriers. Figure 2.7 shows an example of LTE frame consisting of 10 subframes of one ms [40], each subframe is split into slots called RBs. The RB is the smallest unit of resources that can be allocated to a user, its size is fixed to 180 kHz in 0.5 ms. Moreover, the RB is composed of 84 resource elements of 15 kHz and 0.0714 ms each.

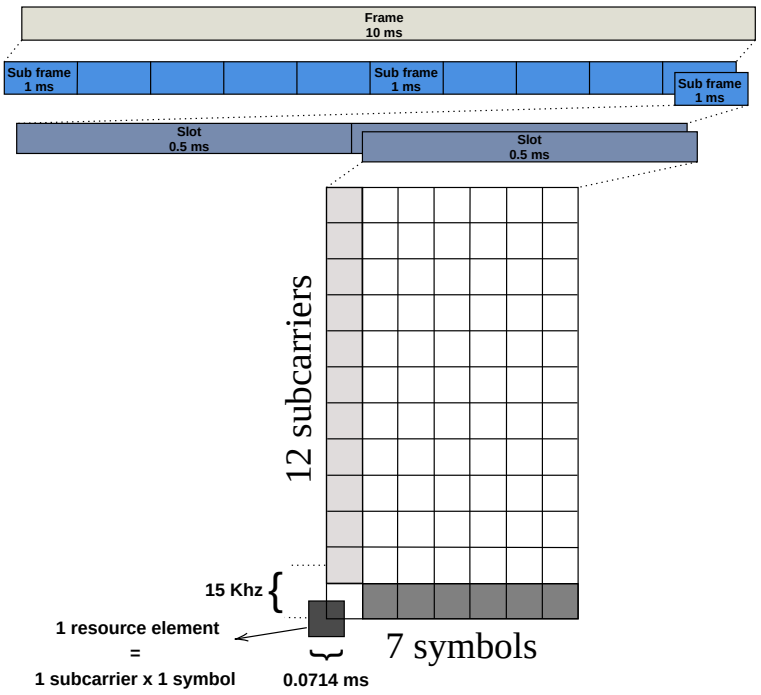

Figure 2.7: LTE Frame

Similarly to the LTE frame, under 5G networks, a frame consists of 10 subframes of one ms, and each subframe is divided into RBs. However, the RB size varies according to an introduced numerology index μ [41], so the RB size is no longer fixed as in LTE networks, and is calculated by using the following formula: 12 subcarriers of 15 x 2^{μ} kHz and 14 symbols in $2^{-\mu}$ ms; For $\mu \in \{0,\ldots,4\}$, which allows for greater flexibility in resource utilization [42, 43]. Figure 2.8 shows an example of a 5G frame when μ is equal to three. In this example, the size of the RB is 1200 kHz in 0.125 ms. Moreover, the RB is composed of 168 resource elements of 120 kHz and 0.0089 ms each.

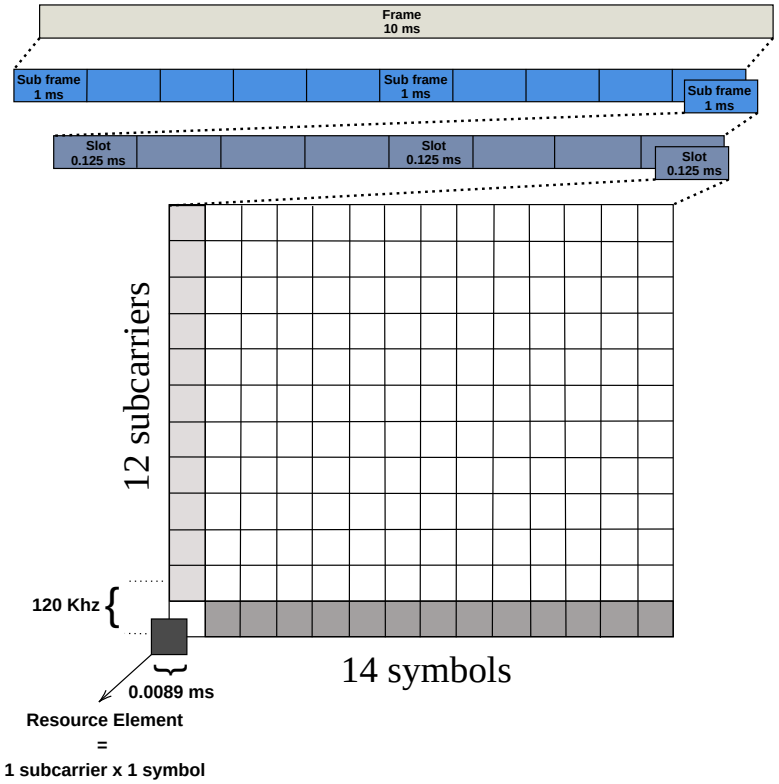

Figure 2.8: 5G Frame example when μ is equal to three

2.2.2 Non-Orthogonal Multiple Access technique

NOMA is a very promising technique that is not yet standardized to be used in CNs. It is proposed as a complement to the OMA technique, as it assists to share orthogonal resources in a non-orthogonal way, thus allowing for more users to access the network. In this thesis, we focus on the use of NOMA for PSUs for its ability to support data transmission by a group of users who simultaneously use limited resources.

The concept of using NOMA in CNs implies that multiple users must be multiplexed simultaneously on the same subband (group of RBs). This concept is based on Superposition Coding (SC) [44] on the transmitter side, and SIC [45] on the receiver side. Many categories of NOMA, such as power domain NOMA [46–50] (where signals are separated using different power levels), code domain NOMA [51,52] (where signals are separated using different signature codes), pattern domain [53] NOMA (where signals are separated using different patterns), etc., have been considered in the literature to separate the signal of each user. In our research, we concentrated on the power domain NOMA.

A power domain example of three users using either NOMA or OFDMA technique is shown in Figure 2.9. In NOMA, the signals to be sent to the three users are multiplexed in the same subband and are differentiated by different power levels [54], whereas in OFDMA, the signals sent to the users have been assigned different frequencies [55].

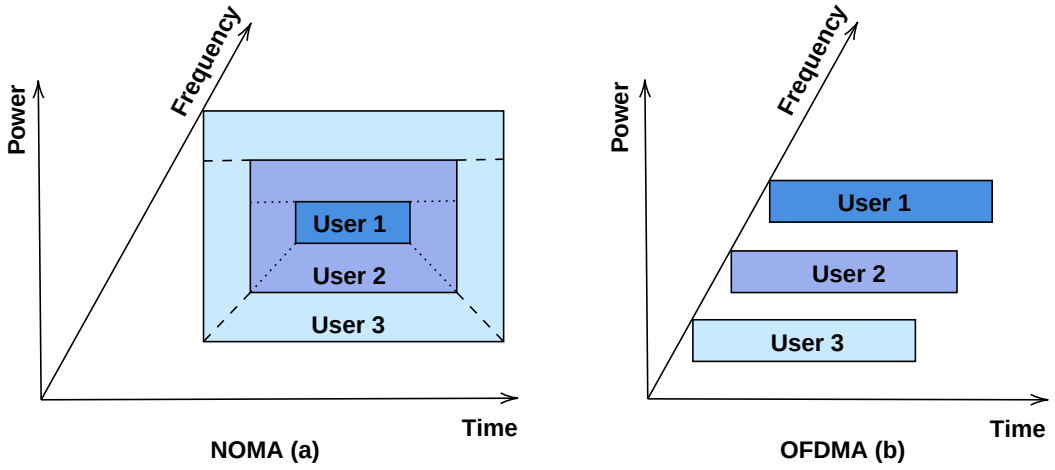

Figure 2.9: NOMA vs OFDMA

2.2.2.1 Downlink NOMA technique

In downlink NOMA [46, 56], multiplexed users over the same subband receive all signals, including interfering and required signals. Therefore, it is necessary for the transmitter, such as a BS or a Cluster Head (CH), to apply SC. Accordingly, the signals transmitted to the receivers use different power coefficients (α_i), which are numbers between zero and one, such that the sum of all these coefficients is equal to one. The allocation of the power coefficient will depend on the channel condition of the UE. Those with a strong channel condition will be assigned a low power coefficient, while those with a weak channel condition will be given a higher power coefficient. Hence, the users having weaker channel conditions will result in stronger interference in the network. Upon reception of the signals, SIC will then be applied at the receiver end. The process is demonstrated in Figure 2.10, which displays three users receiving data in the downlink. User 1 processes a stronger channel condition compared to users 2 and 3, whereas user 2 has a stronger channel condition than user 3. h, x, and n denote the channel gain, desired message, and Additive White Gaussian Noise (AWGN) of users, respectively.

By applying SIC, each user can decode the messages of the users who have a lower channel gain than its own, while treating the messages of those with a stronger channel gain as interference. As a result, user 1, which has the highest channel gain, can retrieve its message by eliminating all the interference after decoding the messages of the other users. User 2 can only decode the message of user 3 and treats that of user 1 as interference. User 3, who has the lowest channel gain, cannot decode messages from other users, thus these messages are considered as interference.

To evaluate user performance, many utilities can be considered, such as throughput achieved, fairness, outage probability, etc. The common metric of these utilities is the data rate achieved by each user. (2.1) shows the achieved date rate by user i, where $i \in \{1, 2, ..., I\}$, who are sharing the same resources in the downlink NOMA. Without loss of generality, we assume that the user channel gain between the transmitter (T) and the receiver i (R_i) follows this order: $h_{T,R_1} > h_{T,R_2} > ... > h_{T,R_I}$.

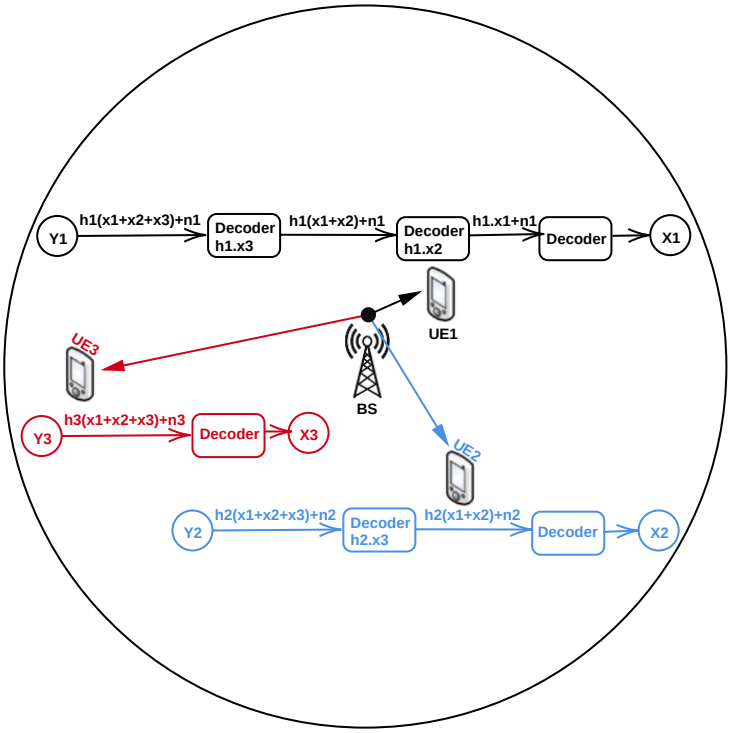

Figure 2.10: NOMA Downlink

$$R_i = b \,.\, log_2(1 + \frac{\alpha_i P \,.\, h_{T,R_i}}{h_{T,R_i} \,.\, (\sum_{j=1}^{i-1} \alpha_j) P + v.b}) \tag{2.1}$$

Where b represents the bandwidth assigned to the communication in NOMA and $\alpha_i P$ is the portion of the power allocated to each UE i over b.

Referring to the same example in Figure 2.10, and after performing SIC by the users, the data rate that can be achieved by each user is computed respectively as follows:

$$R_1 = b \,.\, log_2(1 + \frac{\alpha_1 P \,.\, h_{BS,UE_1}}{v.b}) \tag{2.2}$$

$$R_2 = b \,.\, log_2(1 + \frac{\alpha_2 P \,.\, h_{BS,UE_2}}{h_{BS,UE_2} \,.\, \alpha_1 P + v.b}) \tag{2.3}$$

$$R_3 = b \,.\, log_2(1 + \frac{\alpha_3 P \,.\, h_{BS,UE_3}}{h_{BS,UE_3} \,.\, (\alpha_1 + \alpha_2) P + v.b}) \tag{2.4}$$

2.2.2.2 Uplink NOMA technique

At the beginning of the uplink process, a receiver (e.g. a BS) sends a control message to the transmitting users. This message includes a power allocation coefficient for each user, allowing the BS to receive multiple signals over the same subband. Although these signals may interfere with each other, they are all desired signals.

(2.5) shows the achieved date rate by user i, where $i \in \{1, 2, ..., I\}$, who are sharing the same subband in the uplink NOMA [46, 56].

$$R_i = b \; . \; log_2(1 + \frac{\alpha_i P \, . \, h_{T_i,R}}{(\sum_{j=i+1}^{I} \alpha_j P \, . \, h_{T_j,R}) + v.b}) \tag{2.5}$$

After receiving the signal, the BS applies SIC and begins decoding the users' messages. It starts with the highest channel gain user, the one closest to it. This user experiences interference from all other users sharing the same resources. Once the BS has decoded that user's message, it repeats the process for the user with the second highest channel gain, and then for the third until it reaches the user with the lowest channel gain, who does not experience any interference from other users.

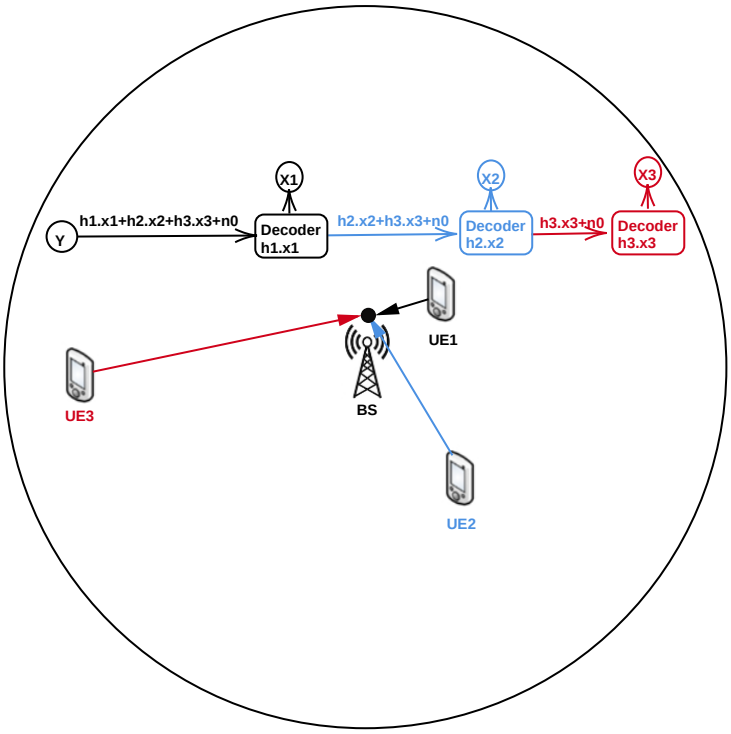

Figure 2.11: NOMA Uplink

Figure 2.11 shows an example of three users sharing the same subband in the NOMA uplink. After the SIC is applied by the BS, the data rates of the three users are as follows:

$$R_1 = b \; . \; log_2(1 + \frac{\alpha_1 P \, . \, h_{UE_1,BS}}{\alpha_2 P \, . \, h_{UE_2,BS} + \alpha_3 P \, . \, h_{UE_3,BS} + v.b}) \tag{2.6}$$

$$R_2 = b \; . \; log_2(1 + \frac{\alpha_2 P \, . \, h_{UE_2,BS}}{(\alpha_3 P \, . \, h_{UE_3,BS} + v.b}) \tag{2.7}$$

$$R_3 = b \; . \; log_2(1 + \frac{\alpha_3 P \, . \, h_{UE_3,BS}}{v.b}) \tag{2.8}$$

(2.1) and (2.5) demonstrate that the data rate achieved by the users in the NOMA technique is closely related to the number of users sharing the same subband, and to the power allocated to each user. A thorough study of the allocation of these resources is therefore essential to take full advantage of the NOMA technique. The effective implementation of NOMA and the achievement of better performance than OFDMA depend heavily on the optimization of resource allocation, and thus on the improved use of scarce resources in wireless communication systems.

2.3 Optimization theory

Optimization theory aims to develop mathematical models and algorithms to maximize or minimize functions that correspond to a problem, subject to certain constraints. These functions are known as objective functions. Each mathematical model is composed of three basic elements: the objective function, the constraints, and the decision variables. In this thesis, we are concerned with integer nonlinear programming [57], because of the nonlinear behavior that characterizes the NOMA technique. The following mathematical model is a generic model that can be adapted to any nonlinear programming problem.

$$min \;\; f(x_1, x_2, ..., x_n) \tag{2.9}$$

$s.t.$

$$g_i(x_1, x_2, ..., x_n) \leq b_i \quad \forall i = 1, ..., m \tag{2.10}$$

$$x_j \in \{0, 1\} \quad \forall j = 1, ..., n \tag{2.11}$$

The problem represented by (2.9) is a minimization problem and becomes nonlinear if either the objective function $f(X)$ or any of the constraints are nonlinear functions of X. It should be mentioned that the problem can also be a maximization problem by using $max \;\; f(X)$ or the $min - f(X)$. (2.10) and (2.11) refer to the constraints and the decision variables respectively, with m constraints and n variables. The constraints can have various forms such as equal to ($=$), greater than or equal to ($\geq$), etc. while decision variables can take different ranges such as values between zero and one ($x_j \in [0, 1]$), or real numbers ($x_j \in \mathbb{R}$), etc. [58]. Solving integer nonlinear programs is known to be a very challenging task, due to the combination of the complexity of both nonlinear and integer variables.

2.3.1 Different optimization methods for resource allocation

In the OFDMA technique, each user is given a distinct and non-overlapping subband to transmit and receive data during their designated time slot. This technique provides users with exclusive access to their assigned subband, preventing any interference from other users. In contrast, the NOMA technique allows multiple users to share the same subband simultaneously. This introduces some challenges, such as

determining the appropriate distribution (user grouping) for an allocated subband, and the power allocation for each user within the shared subband. The problem of resource allocation in NOMA becomes very complex if these challenges are considered simultaneously. However, by splitting them into several problems, the overall problem becomes solvable and can be addressed effectively.

2.3.1.1 User grouping problem

Grouping users in NOMA involves determining how to distribute users to improve system performance. Various algorithms and techniques have been employed to achieve this objective. One such method is the exhaustive search, which explores all possible combinations of user groups to find the best solution. However, its computational complexity grows exponentially with the number of users, making it impractical for real-time systems [59, 60]. Another approach is the channel gain difference strategy [46, 61, 62], which groups users based on the disparity in their channel gain. Although simple and effective, it does not always deliver optimal performance. The climbing pairing [60] and the Hungarian algorithm [63] are also utilized in this context, providing a better solution compared to the channel gain difference method. However, they require iterative calculations, making them complex for large-scale systems. Metaheuristic algorithms such as simulating annealing [60,64] and PSO [65–67] have been employed as well, yielding satisfactory results. Yet, these algorithms may necessitate parameter tuning and increased computing resources depending on the scenarios. The artificial neural network approach [68] has also been considered, where neural networks learn the optimal grouping of users using training data. The network is trained on a large data set, and its parameters are adjusted to optimize a specific objective function. This method requires significant computing resources and time. Considering that resource allocation in wireless communication systems must be performed within milliseconds, the user grouping methods need to be of low complexity while ensuring system performance.

2.3.1.2 Power allocation problem

Power allocation in NOMA is the process of distributing power among users within each group to optimize the overall system performance. Various algorithms and approaches have been developed to tackle this task. The full space algorithm [69] exhaustively explores the entire power allocation space to find the best possible solution. However, due to its considerably high complexity, it is not suitable for systems that require real-time operation. Another approach is the Fixed Power Allocation (FPA) algorithm [69, 70], which assigns a fixed power level to each user within a group. Although FPA has low complexity, it often results in sub-optimal performance since it does not adapt to the varying channel conditions experienced by different users. On the other hand, the Fractional Transmit Power Allocation (FTPA) algorithm [69, 71, 72] allows for flexible power distribution by assigning fractions of the available power to users based on their specific channel conditions. FTPA is a local optimization algorithm that improves performance compared to FPA while still maintaining a manageable level of complexity. Similar to user grouping, metaheuristic algorithms such as simulated annealing [64] and PSO [73] can also be applied

to power allocation in NOMA. However, it is important to note that these algorithms may find local optimal solutions and face challenges in effectively handling constraints. Additionally, the KKT conditions [46], which is a mathematical theory, offers a closed-form solution for power allocation that can achieve the optimal global solution. Furthermore, machine learning methods such as deep learning [74] and reinforcement learning [75] have been also explored for power allocation. These approaches have the potential to learn complex patterns and optimize system performance. However, they often require large data sets, which can be time-consuming and resource-intensive, especially during the initial stages of implementation.

Various approaches have been explored to address the resource allocation problem in the NOMA technique, including exact methods, heuristic methods, and decomposition methods. Exact methods guarantee that a global optimal solution is obtained, but their applicability is often limited by factors such as computational complexity, time constraints, and practical considerations. On the other hand, heuristic methods and decomposition methods aim to provide a solution that may not be globally optimal but can be obtained within a reasonable time. Heuristic methods employ efficient algorithms to find sub-optimal solutions, while decomposition methods divide the original problem into smaller, more manageable sub-problems for easier resolution. These alternative categories of methods offer practical approaches to tackle the resource allocation problem in NOMA, balancing the trade-off between optimality and computational feasibility.

2.3.2 Overview on resource allocation methods used in this thesis

As we mentioned above, the resource allocation for PSUs consists of two main phases. The first phase is user grouping, which is performed using the channel gain difference strategy in this thesis. The second phase is the power allocation, for which we investigate two methods. We initially explore the PSO algorithm, which yields satisfactory results. However, due to the necessity to adjust the parameters for different scenarios, we subsequently examine the KKT conditions. These conditions provide a closed-form solution and ensure that an optimal global solution is obtained. We find it worth providing a concise overview of these two problem-solving approaches.

2.3.2.1 Particle Swarm Optimization method

PSO algorithm was developed by James Kennedy and Russell Eberhart in 1995 [76]. This algorithm is based on the movement of swarms; it was intended for social behavior (e.g. the movement of a flock of birds), and then it became one of the most popular nature-inspired metaheuristic optimization algorithms for solving various optimization problems in science and engineering [77–79]. The principle of PSO lies on a population (swarm) of candidate solutions (particles) that move in the search space with the aim that these particles find the optimum. To enable the movement of these particles, a simple mathematical formula is applied to their position and

speed. Furthermore, to control the behavior and efficiency of the algorithm, different acceleration factors, random numbers, and inertia factor are used. The movement of each particle is influenced by its best-known local position (Pbest), and at the same time is influenced by the best-known position (Gbest), which is updated according to the best positions found by the other particles. As a result, the swarm is expected to move toward the best solution.

Consider that we have a population size of M and several particles (variables) of N. For each iteration, the velocity and position of each m and n are updated as follows:

Velocity :

$$V_{m,n}^{k+1} = W.V_{m,n}^{k} + C_1.R_1.(Pbest_{m,n}^{k} - X_{m,n}^{k}) + C_2.R_2.(Gbest_{n}^{k} - X_{m,n}^{k}) \tag{2.12}$$

Position :

$$X_{m,n}^{k+1} = X_{m,n}^{k} + V_{m,n}^{k+1} \tag{2.13}$$

The parameters W, C_i, and R_i $(i = 1, 2)$ represent the inertial weight, the accelerations factors, and the random numbers. They play an important role in determining the movement of particles towards the optimal solution. The primary function of W is to control the global-local optimum search, whereas C_i, and R_i are mainly responsible for controlling the directions and velocity of particle movement. A higher inertia weight helps the algorithm perform a global optimality search, while a lower weight favors the local search. Therefore, a higher weight is needed at the beginning of the search to determine the general position of the optimal solution, and later, a lower weight can be used to help the algorithm converge to the optimal solution. Generally, the commonly employed values for these parameters are 0.9 to 0.4 for W, [0,1] for C_i, and 2 to 2.05 for R_i [80].

Once the above calculations are performed and the right parameters are determined, the fitness function is computed, which represents the objective function to be optimized. The value of $X_{m,n}^{k+1}$ that results in the highest fitness function result will be considered as the best particle positions, thus indicating the optimal values of the objective function variables.

2.3.2.2 Lagrange multiplier method

The Lagrange multiplier method is an efficient way of solving optimization problems; it allows critical points to be obtained, whether they are local or global optima, under equality constraints. For problems with inequality constraints, the Lagrange multipliers method can be generalized to the KKT approach, which is useful for nonlinear constrained programs [81]. The KKT approach states that an optimal point must be either a critical point in the interior of the feasible set or on its boundary. Once we know on which boundary of the feasible region it lays, the KKT conditions become active along that boundary and determine the set of critical points [82].

The subsequent example is a general maximization problem, on which we will give a general idea of the application of the Lagrangian and the KKT conditions.

$$\max_{x \in \mathbb{R}^n} f(x) \tag{2.14}$$

s.t

$$g_i(x) \leq b_i \quad \forall i = 1, ..., m. \tag{2.15}$$

After assigning a reward of λ_i unit to promote feasibility in the inequality-constrained maximization problem, the hard constraints listed in (2.15) can be added into the objective function as $\lambda_i[b_i - g_i(x)]$. Then, the Lagrangian can be formed as follows:

$$L(x|\lambda) = f(x) + \sum_{i=1}^{m} \lambda_i[b_i - g_i(x)] \tag{2.16}$$

The KKT approach outlines four sets of conditions for a point $\bar{x}$ to be considered a local optimum. These conditions are described as follows:

- **Stationarity**

$\nabla\{f(\bar{x}) + \sum_{i=1}^{m} \lambda_i[b_i - g_i(\bar{x})]\} = 0$

- **Primal feasibility** :

$g_i(\bar{x}) \leq b_i$ for all $i = 1, ..., m$

- **Dual feasibility** :

$\lambda_i \geq 0$ for all $i = 1, ..., m$

- **Complementary slackness** :

$\lambda_i[b_i - g_i(\bar{x})] = 0$ for all $i = 1, ..., m$

The stationarity conditions ensure that the gradient is within all the gradients of constraints, therefore there is no way to keep improving feasibly. The primal feasibility conditions are used to ensure that $\bar{x}$ is going to satisfy all the original constraints. The dual feasibility conditions ensure that all the multipliers are greater or equal to zero. Finally, the complementary slackness conditions demonstrate that only the constraints binding at $\bar{x}$ matter, if we have a constraint that does not bind, λ_i would be zero and we do not care about the gradient of that constraint because it is not binding.

2.4 Summary

The main objective of this chapter is to provide an overview of the various topics that are closely related to our thesis work. We aim to enhance PS services by leveraging cutting-edge technologies offered by CNs. PS services rely on D2D communication, which is enabled in CNs through the ProSe standard. Therefore, in this chapter, we provided a review of LTE and 5G CNs and explored the deployment of D2D communication in these networks. Additionally, PS services are crucial in critical scenarios like emergencies and disasters, where establishing a connection to BS may be challenging or even impossible. To address this issue, we propose implementing ProSe functionalities in the MEC. In Section 2.1.2.4, we introduced the MEC and its integration in 5G CNs. Effective management of radio resources is essential for integrating PSNs in CNs. Consequently, we investigated the RBs structure in both LTE and 5G networks. We also examined the NOMA access technique, which plays a key role in enabling a maximum number of PSUs to access the network. Finally, we presented various optimization methods which we will employ to solve optimization problems related to effective radio resource allocation for PSUs. By covering these topics, our aim was to provide a general understanding of the concepts and principles underlying our research on PS.

Chapter 3

Efficient power allocation scheme for PSUs with dedicated subbands

Contents

3.1 Introduction

Ensuring communication with PSUs in all events, and specifically in disaster situations, is one of the main challenges in CNs. Various types of communications, such as in-band and out-band communications, can be used to integrate the PSNs into the CNs. In this chapter, we focus on the in-band overlay D2D communication, which effectively reduces the interference caused by CUs, and ensures the availability of resources for PSUs at any time and for different events. Furthermore, by considering the NOMA-based system, the user's throughput and the resource wastage

problem are improved. Our goal is to provide the necessary resources to PSUs and, at the same time, to maximize the use of these resources.

Our proposed approach involves using a heuristic algorithm to assign appropriate subbands to the PS clusters, followed by the use of the PSO algorithm to distribute the available power that can be used in each subband by the PSUs. According to the simulation results, our proposed approach proves to be more efficient than the OFDMA technique in terms of user sum-throughput. In addition, we study how the sum-throughput is affected by the number of PSUs of the same group. Finally, the simulations show that the relation between throughput and fairness is a requirement-dependent trade-off, such that achieving optimal fairness requires a decrease in the sum-throughput.

The remainder of this chapter is arranged as follows: Section 3.2 presents a brief summary of some literature related to this chapter. In Section 3.3, we provide an overview of the network structure of our proposed model and describe the formulation of the resource allocation problem. Section 3.4 delves into the specifics of our proposed solution which is based on both heuristic and PSO algorithms. We then present the results of our simulations and compare our approach with the OFDMA-based method in Section 3.5. Lastly, Section 3.6 provides the concluding remarks.

3.2 Related work

To deal with the emergency communication scenarios, a clustering method was proposed in [83], which reduces the system's energy consumption and realizes the continuous communication, based on D2D multicasting technology, leading to optimizing the allocation of the network resources. The concept of "D2D group" communication using NOMA, was introduced in [84], to enable one D2D transmitter to communicate with multiple D2D receivers simultaneously. The authors of [46, 85] show the advantage of using the clustering strategy in NOMA, and how the difference in channel coefficient influences the performance and results.

In [86], the authors proposed a selective operation of overlay D2D communication in areas with high call traffic density, to overcome the problem of spectrum wastage. A novel sub-grant scheme in the overlay D2D communication, where the symbol basis method is used to grant the allocated and unused resources to other users in the vicinity, was studied in [87]. The PSO algorithm was employed in [66, 67] to assign power to individual subbands, and subsequently, this power was allocated to different multiplexed users based on their respective channel gains. The authors suggested the use of this algorithm due to its ability to efficiently find near-optimal solutions to complex and constrained problems.

Inspired by the aforementioned works, we have developed a resource allocation problem aimed at maximizing power allocation for clusters of PSUs in the context of overlay D2D communication, while utilizing the NOMA technique.

3.3 Problem formulation

3.3.1 System model

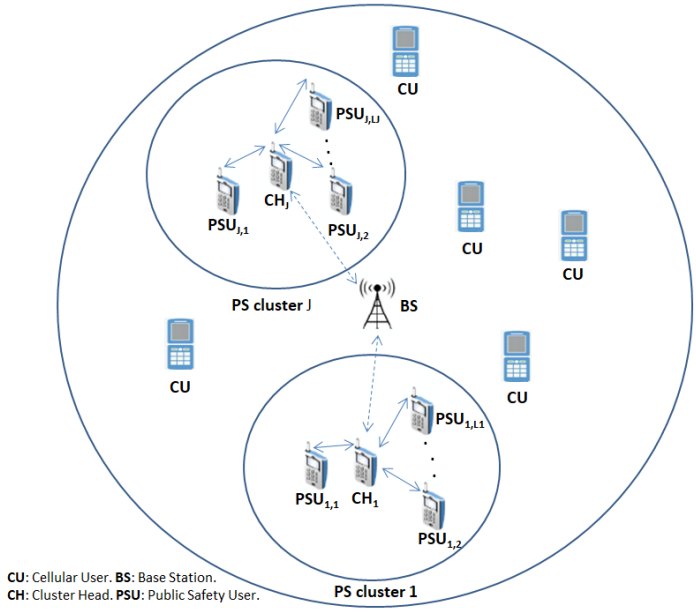

Figure 3.1: Illustration of overlay D2D communication for PSUs with a NOMA-based CN

The considered network is made up of J clusters of critical D2D users referred to as PS clusters. These clusters denoted by $\{PS$ cluster 1, ..., PS cluster $J\}$, are made up of one CH, i.e., $\{CH_j\}$ where $j \in \{1, ..., J\}$, and L_j PSU, i.e., $\{PSU_{j,1}, \ldots, PSU_{j,L_j}\}$. The clusters coexist with a macro cell where a macro BS is serving I CUs, i.e., $\{CU_1, CU_2, ..., CU_I\}$ that are uniformly distributed throughout the cell.

Each cluster is formed around a particular node, known as the CH. This node is an extremely powerful mobile device installed by the safety organization in a designated location. It has access to the BS and serves as a relay for all PSUs located within its coverage area. The CH processes and manages the allocation of subbands for clusters, ensuring that PSUs can communicate with each other and have access to the BS. PSUs are uniformly distributed within the cluster, which has a maximum radius of d_{max}. Within the PS cluster, NOMA-based D2D communication is employed, where only PSUs of the same cluster can share the same resources simultaneously. This results in only one type of interference: the intra-cluster interference among PSUs of the same cluster. It is important to note that there is no interference resulting from the CUs, due to the dedication of resources for PSUs in the overlay D2D communication. Users are deployed as illustrated in Figure 3.1.

OFDMA has been considered in Long-Term Evolution-Advanced (LTE-A) and 5G, however in a PS scenario, where there is a dense number of users with limited resources, NOMA is a more promising system due to its ability to support many

users to share the same subband. This results in greater spectrum efficiency by giving more PSUs the opportunity to access the network.

3.3.2 Preliminary

A signal received without sharing the bandwidth with other users is composed of the transmitted power (P), the channel gain (h), the transmitted signal (x), and the AWGN (n), as shown in (3.1).

$$y = \sqrt{P}.h.x + n \tag{3.1}$$

h is made up of the small-scale fading and the path loss between the transmitter and the receiver. Based on (3.1), the received Signal-to-Interference-plus-Noise Ratio (SINR) can be calculated as follows:

$$SINR = \frac{P.h}{N_0.b} \tag{3.2}$$

Where b represents the bandwidth allocated to the data transmission between the transmitter and the receiver, and N_0 corresponds to the noise power spectral density. The user's data rate (R) is calculated using the Shannon capacity formula:

$$R = b \ . \ log_2(1 + SINR) \tag{3.3}$$

Given the example in Figure 3.1, the overall system bandwidth (B) is split into M RBs, i.e., $\{RB_1, RB_2, ..., RB_M\}$. The CUs use the orthogonal spectrum resources, and D2D pairs do not multiplex them, so the interference among these CUs is negligible. Pc_i is defined as the power assigned to a subband W_i in the downlink communication between the BS and the CU_i, and h_{BS,CU_i} represents the channel gain between them. Therefore, the data rate of each CU_i can be calculated according to the following equation:

$$R_{CU_i} = W_i \ . \ log_2(1 + \frac{Pc_i \ . \ h_{BS,CU_i}}{N_0 \ . \ W_i}) \tag{3.4}$$

By considering the overlay D2D communication within the PS clusters, unused cellular resources can be allocated to these clusters without interfering with the CUs. A set of RBs, denoted as W_j, will be allocated to each cluster j. The intra-cluster interference is therefore the only interference to be taken into account, and the $SINR$ of each PSU within the cluster must be processed above a threshold level ($SINR_{thr}$) to ensure high-quality communication for these users. Ps_j represents the maximum amount of power that can be accommodated in subband W_j.

To control the communication within the PS clusters, the NOMA technique is employed. This involves assigning power coefficients ($\alpha_{j,l}$), which range from zero to one, to the signal transmitted to PSUs in cluster j, such that $\sum_{\forall l}(\alpha_{j,l}) = 1$. The PSU with a strong channel condition is assigned a low power coefficient and the one

with a weak channel condition is assigned a higher power coefficient. Without loss of generality, it is assumed that all the channels in the cluster j follow this order: $h_{CH_j,PSU_{j,1}} > h_{CH_j,PSU_{j,2}} > ... > h_{CH_j,PSU_{j,L_j}}$. Upon receiving the signal, and after performing SIC, each user can decode the message of users having a weaker channel gain than its own and treats the message of those with stronger channel gain as interference. Consequently, the PSUs rate can be calculated as follows:

$$R_{PSU_{j,l}} = W_j \,.\, log_2(1 + \frac{\alpha_{j,l} Ps_j \,.\, h_{CH_j,PSU_{j,l}}}{h_{CH_j,PSU_{j,l}} \,.\, (\sum_{t=1}^{l-1} \alpha_{j,t}) Ps_j + N_0 \,.\, W_j}) \tag{3.5}$$

Based on the preceding information, the overall rate of the downlink system can be expressed as:

$$R_{sys} = \sum_{j=1}^{J} \sum_{l=1}^{L_j} R_{PSU_{j,l}} + \sum_{i=1}^{I} R_{CU_i} \tag{3.6}$$

This overall rate is composed of two parts: the first is the rate attained by PSUs in all clusters, and the second is the rate accomplished by CUs. It is obvious that the amount of power allocated to PSUs affects the overall rate. Therefore, our main concern is to find the best allocation of these powers.

3.3.3 Problem statement

To achieve the highest overall rate, it is necessary to maximize the rate achieved by PS clusters. The performance of CUs is not affected by the presence of these clusters, so there is no need to consider the rate of CUs. Therefore, the objective function can be written as follows:

$$max(R_{sys}) \tag{3.7}$$

s.t.

$$SINR_{PSU_{j,l}} \geq SINR_{thr}, \quad j \in \{1, ..., J\}, l \in \{1, ..., L_j\} \tag{3.7a}$$

$$\sum_{l=1}^{L_j} \alpha_{j,l} = 1, \quad j \in \{1, ..., J\} \tag{3.7b}$$

$$\frac{Ps_j}{2^{(L_j - l + 1)}} < \alpha_{j,l} Ps_j \leq \frac{Ps_j}{2^{(L_j - l)}}, \quad j \in \{1, ..., J\}, l \in \{1, ..., L_j\} \tag{3.7c}$$

$$\sum_{i=1}^{I} Pc_i + \sum_{j=1}^{J} Ps_j \leq Pmax \tag{3.7d}$$

Wherein (3.7a) ensures that the $SINR$ of PSUs must be maintained above the threshold level. Constraint (3.7b) guarantees that the sum of the allocated power for communication with PSUs in cluster j does not exceed the assigned power budget (Ps_j) for that cluster. For an efficient SIC in the downlink NOMA cluster, (3.7c) specifies the interval within which the power of each PSU takes a value (A detailed derivation of this constraint is given in the **Appendix**). Finally, constraint (3.7d) ensures that the sum of the allocated powers to all the users in the network does not exceed the total power budget $Pmax$ of the BS.

3.4 Resource allocation algorithms

As stated in the optimization theory Section, various optimization algorithms can be utilized depending on the type of problem being optimized. Our objective function, which involves selecting subbands for clusters and determining the power allocation for each PSU within those clusters, is classified as a mixed integer non-linear programming problem. Therefore, we have applied two algorithms to solve the problem we have formulated. The initial algorithm uses a heuristic approach that assigns the appropriate subbands to each cluster by performing grouping within clusters when necessary. After grouping the PSUs of the same cluster, we apply the PSO algorithm that will search to optimally allocate the available power of each subband among the PSUs of the same group.

3.4.1 Heuristic algorithm for subband allocation

Initially, each cluster is considered as a single group, and all of its PSUs are multiplexed together to utilize the same subband. Later on, the cluster might be subdivided into several groups depending on the number of its PSUs and their rate threshold. If Ps_j is insufficient to satisfy all of the PSUs in cluster j, this cluster will be divided into two groups. Each group will consist of $\frac{L_j}{2}$ PSUs if there is an even number of PSUs in the cluster. Otherwise, the first group will have the floor value of $\frac{L_j}{2}$ and the second group will have the ceiling value of it. A unique subband will then be assigned to each group, and this process will be repeated until an appropriate grouping of the cluster is determined.

Figure 3.2 illustrates an example of various options for grouping 12 PSUs belonging to the same cluster. The first option considers that all PSUs form a single group, while other options follow the user distribution pattern used in [46]. The authors have shown that it is advantageous to distribute users with high channel gain into different groups and combine them with users with low channel gain. With this strategy, we form new groups by selecting the users with the highest and lowest channel gain from the previous groups. This process is repeated iteratively until all users are considered.

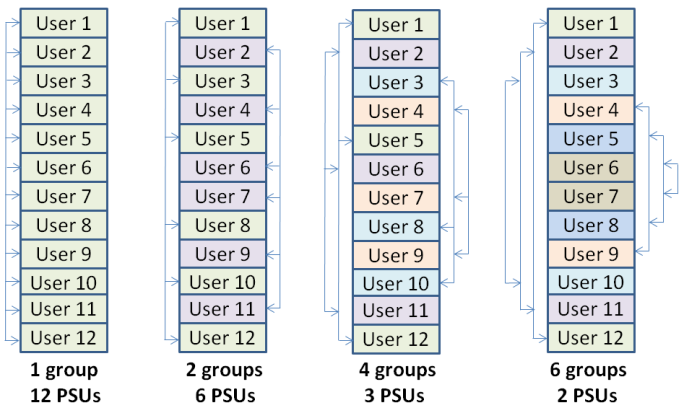

Figure 3.2: Various options for grouping 12 PSUs

The method of assigning the subbands to clusters is outlined in **Algorithm 1**. The input parameters are the number of clusters and the number of PSUs in each cluster. N_j represents the number of subbands needed to fulfill the rate threshold of all the PSUs in the clusters. Starting with the first cluster, and based on the number of its PSUs and their threshold rate, the CH can decide if a single subband is sufficient to satisfy their need. If it is the case, only one subband will be assigned to this cluster (steps 4-6). Otherwise, we proceed to group this cluster into two groups (step 9). We then evaluate if a single subband per group is sufficient; if yes, we assign one subband per group (steps 11-13). In the contrary case, we repeat the grouping procedure for the group that still has unsatisfied PSUs (step 15). This process is repeated for all PS clusters in the network.

Algorithm 1 Heuristic algorithm for subbands allocation

Input: J: PS clusters and L_j: PSUs in cluster j.
Output: V_j: A vector indicating the number of subbands for each cluster j.

System Initialization
1: $g = 0$; //Nb of subbands
2: $j = 0$;
Main
3: **for** j=1 to J **do**
4: **if** 1 subband is sufficient for all PSUs in the cluster j **then**
5: g=g+1; //Increment g
6: V_j=g; //Assign 1 subband for cluster j
7: **end if**
8: **else**
9: Grouping into 2 groups;
10: **for** each group **do**
11: **if** 1 subband is sufficient for all PSUs in the group **then**
12: g=g+1;
13: V_j=g; //Assign g subbands for cluster j
14: **else**
15: Repeat 9
16: **end if**
17: **end for**
18: **end for**
19: Return V_j //Return the number of subbands for cluster j

3.4.2 Particle Swarm Optimization (PSO) algorithm for power allocation to PSUs

PSO is a well-known optimization algorithm in the literature, which is classified as a metaheuristic method. It is particularly suited to address nonlinear problems, especially when the search space is complex and high-dimensional. This makes it appropriate for our problem of optimizing power allocation for PSUs of the same group. In our case, particles represent the various power coefficients to be allocated to PSUs in the group, which will be evaluated by the fitness function. PSO uses the

fitness function, which serves as the objective function, to guide the search for an optimal solution. In response to its Pbest and Gbest positions, each particle adjusts its position and its velocity within the search space. The particle that results in the highest fitness value will be then selected as the best solution for power allocation.

Algorithm 2 shows the PSO algorithm we employed to assign the power dedicated to each group. This algorithm takes various inputs such as the number of PSUs in the group, their channel gain, the acceleration coefficients, and the inertial weight. The output Δ_g^j includes the best powers that will be assigned to the PSUs in each group g of cluster j. The initialization process is outlined in steps (3-5), which involves setting the initial positions of both the particles and the swarm based on random values. Steps (9-12) explain how each iteration updates the position of the particles based on three factors: its current velocity, its best solution found so far, and the best solution found by its neighboring particles. Next, as described in steps (13-15), we assess our objective function, which consists of the sum-throughput in each group. After each iteration, we update the best position of the particles and the swarm, thus indicating the best power allocation in each group.

Algorithm 2 Public Safety User's power allocation using PSO algorithm

Input: The number of PSUs in each group, their channel gain, the acceleration coefficients (C_1 and C_2), and the inertial weight W.

Output: Δ_g^j.

System Initialization

1: **for** each group **do**
2: **for** each particle $m = 1$ **to** M **do**
3: Initialize the particle's random position $\mathrm{X}_{g,m}$
4: Set the particle's best known position ($pbest$) to the initial position
5: Set the swarm's best known position ($gbest$) to the initial position
6: **end for**

Main

7: **while** the end condition is not satisfied **do**
8: **for** each particle $m = 1$ **to** M **do**
9: Update the particle's velocity:
10: $V_{g,m}^{k+1} = W.V_{g,m}^{k} + C_1.R_1.(P_{g,m}^{k} - X_{g,m}^{k}) + C_2.R_2.(G_g^{k} - X_{g,m}^{k})$
11: Update the particle's position:
12: $X_{g,m}^{k+1} = X_{g,m}^{k} + V_{g,m}^{k+1}$
13: Evaluate the fitness function f as:
14: $R_{PSU_g^j} = \sum_{l=1}^{L_g}[W_j \ . \ log_2(1 + \frac{\alpha_{g,l} Ps_{\mathrm{j}} \, . \, h_{CH_{\mathrm{j}},PSU_{g,l}}}{h_{CH_{\mathrm{j}},PSU_{g,l}} . (\sum_{t=1}^{l-1} \alpha_{g,\mathrm{t}}) Ps_{\mathrm{j}} \ + \ N_0 \, . \, W_j})]$
15: Update the particle's best known position ($pbest_{g,m}$) and the swarm's best known position ($gbest_g$) based on the fitness function
16: **end for**
17: **end while**
18: **end for**
19: Return Δ_g^j //Return the best allocation of powers to the PSUs in each group g of cluster j

3.5 Numerical Evaluation

3.5.1 Simulation setup

The simulation is carried out with MATLAB. We considered one cluster with a radius of 100 m, containing a different number of PSUs. We calculate the power allocated to each PSU using the PSO algorithm, and we evaluate our proposed method against the conventional OFDMA technique.

The parameters used in the simulations are summarized in Table 3.1. d_{max} is the maximum distance between the CH and a PSU. We chose a SIC threshold of -10 dBm to distinguish signals of users within the same group. This chosen value of SIC allows more users to be multiplexed to use the same resources, regardless of the demodulation complexity of the SIC receiver. For the PSO algorithm, we utilize a population size of 100 particles, the inertia weight started by 0.75 and is decreased by a random number, while the acceleration factors are set to 2. The final results were derived by taking the maximum values from running 100 simulations of 1000 iterations. The main loop of the PSO algorithm terminates when the algorithm reaches the number of iterations or if the gbest value is not optimized further than a tolerance factor of 10^{-12}.

Table 3.1: Default values of the simulation parameters.

Simulation parameters	Default values
d_{max}	100 m
Number of deployed PSUs	2-12
Subband allocated power	46 dBm
Bandwidth of a RB	180 kHz
SIC threshold	-10 dBm
PSO Population Size	100
PSO Maximum Iterations	1000
PSO Inertia Weight	0.75 decreased by Rand
PSO Acceleration Factor	2
Maximum runs	100

In this chapter, we have considered a normalized channel gain in our simulation; therefore, (3.5) becomes as follows:

$$R_{PSU^j_{g,l}} = W_j \,.\, log_2(1 + \frac{\alpha_{g,l} Ps_j \,.\, \delta^j_{g,l}}{\delta^j_{g,l}.(\sum_{t=1}^{l-1} \alpha_{g,\mathrm{t}}) Ps_j + w}) \tag{3.8}$$

Where w is the number of RBs in each subband, $\delta^j_{g,l} = h/(N_0 \,.\, W_j/w)$ is the normalized channel gain, with h being the channel gain, N_0 is the noise power spectral

density and W_j/w is the bandwidth of one RB. The normalized channel gain we used for simulations is: [60 55 50 45 40 35 30 25 20 15 10 5].

3.5.2 Simulation results

Figure 3.3 illustrates a comparison of user sum-throughput between the NOMA and OFDMA techniques, with varying numbers of PSUs in one group for the NOMA technique. The PSUs throughput threshold is set to 180 *kbps*. For two PSUs, we took the smallest and the largest values of the normalized channel gain vector, for three PSUs we added the second smallest value, for four we added the second largest value, etc. As the number of PSUs increases, the sum-throughput increases for both methods. Moreover, our method, which utilizes the PSO algorithm to optimally distribute total power among PSUs, demonstrates better performance than the OFDMA technique.

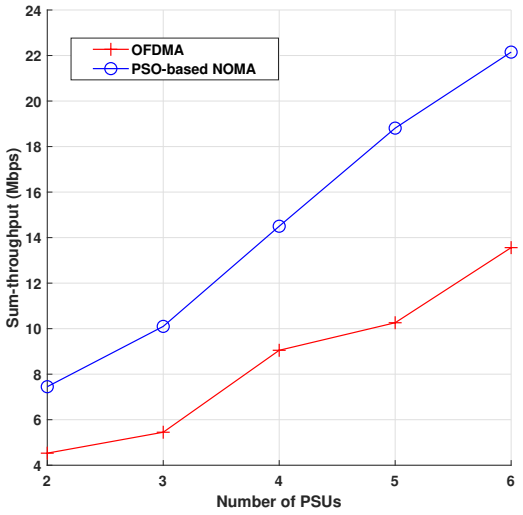

Figure 3.3: Sum-throughput when the number of PS users ranges from 2 to 6 users in one PS cluster

By increasing the SIC threshold, fewer PSUs will be able to use the same resources, thus reducing the demodulation complexity. Figure 3.4 shows different possibilities for grouping the PSUs of the same cluster. These PSUs utilize the normalized channel gain vector values. As the number of multiplexed PSUs in the same group decreases, the sum-throughput decreases. This means that, based on the SIC threshold and demodulation complexity, the more PSUs multiplexed in a group, the higher the sum-throughput will be.

The impact of the number of RBs in one subband on the sum-throughput of PSUs is depicted in Figure 3.5. As the number of RBs and PSUs in a group increases, the sum-throughput of PSUs also increases. Notably, when the number of PSUs is low with a small number of RBs, the sum-throughput changes slowly, but when the number of RBs is high, the sum-throughput increases rapidly. Therefore, it

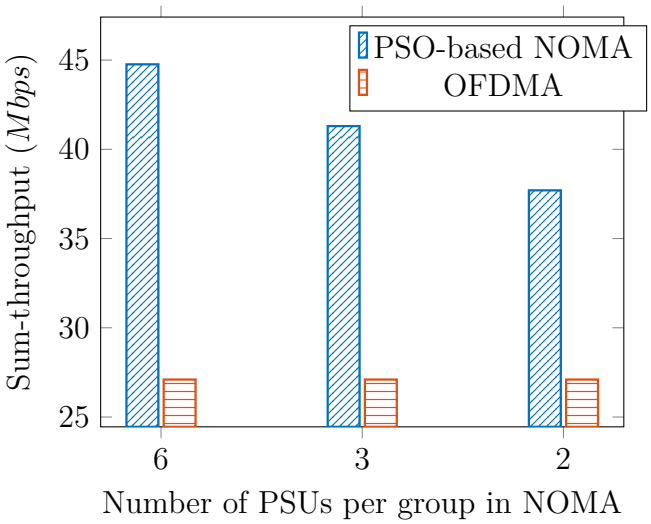

Figure 3.4: Variation of sum-throughput according to the variation of the number of PSUs of the same group

is suitable to have a high number of PSUs in the same group when using a large number of RBs.

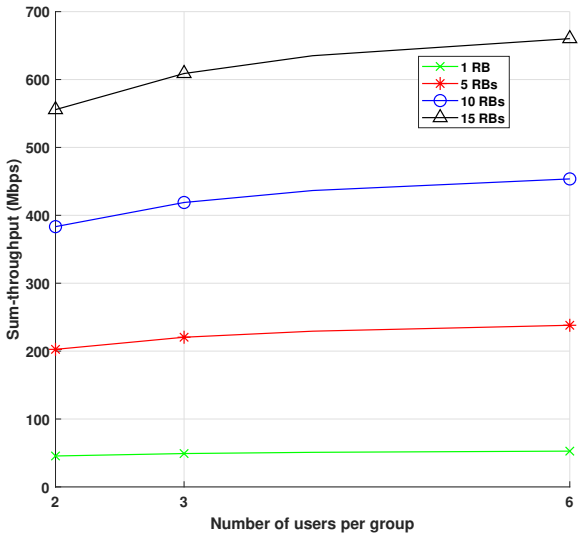

Figure 3.5: The impact of the number of RBs and users per group on the sum-throughput of users

We evaluate the users' throughput fairness by using Jain's fairness index in (3.9) [88]. The result ranges from $1/L_j$ (worst case) to 1 (best case, where all the PSUs achieve the same throughput).

$$\text{Jain's fairness index} = \frac{(\sum_{l=1}^{L_j} R_l)^2}{L_j \sum_{l=1}^{L_j} R_l^2} \tag{3.9}$$

Where R_l is the rate achieved by each user l.

The simulation is performed for two scenarios: the first one concerns the maximum sum-throughput achieved by using the PSO algorithm (we refer to this scheme by the term $\texttt{Max} - \texttt{throughput}$), and the second one concerns the maximum fairness (referred as $\texttt{Max} - \texttt{fairness}$) using $R_{PSU^j_{g,1}} = R_{PSU^j_{g,2}} = ... = R_{PSU^j_{g,L_g}}$, which can be written and solved as follows:

$$W_j \ . \ log_2(1 + \frac{\alpha_{g,1} Ps_j \ . \ \delta^j_{g,1}}{w}) = W_j \ . \ log_2(1 + \frac{\alpha_{g,2} Ps_j \ . \ \delta^j_{g,2}}{\delta^j_{g,2} . \alpha_{g,1} Ps_j + w}) \tag{3.10}$$

$$... = W_j \ . \ log_2(1 + \frac{\alpha_{g,L_g} Ps_j \ . \ \delta^j_{g,L_g}}{\delta^j_{g,L_g} . (\sum_{t=1}^{L_g - 1} \alpha_{g,t}) Ps_j + w})$$

According to the results presented in Figure 3.6, the $\texttt{Max} - \texttt{throughput}$ approach yielded an inverse relation between users' throughput fairness and the number of multiplexed PSUs. This implies that when more PSUs are grouped together, their throughput fairness decreases. Whereas with OFDMA technique, the fairness depends on the number of RBs used, such that an increase in the number of RBs leads to an increase in the fairness of the PSUs throughput. In contrast, if our goal is to achieve maximum fairness, we can reduce the power allocated to the PSU with the highest channel gain. This would result in a decrease in the throughput achieved by that user, but would allow the other PSUs to achieve that same throughput. Hence, maximum fairness among all PSUs is achieved.

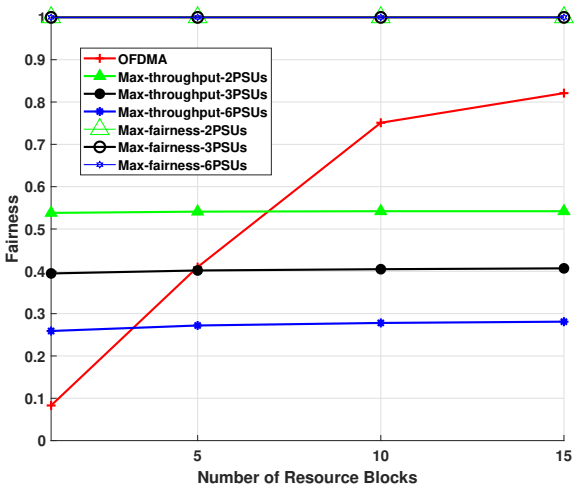

Figure 3.6: Users fairness index based on the number of available RB

3.6 Summary

The chapter focuses on presenting a novel framework for bandwidth and power allocation in a NOMA-based system that utilizes overlay D2D communication for PSUs. The goal is to maximize the sum-throughput in each group of PSUs while satisfying the minimum data rate threshold for each PSU. To achieve this objective, the chapter proposes a two-stage approach. The first stage involves a heuristic algorithm for selecting PSUs to form groups, and the second stage employs the PSO algorithm to optimally allocate power in each group.

The proposed framework is compared to an OFDMA-based one in terms of sum-throughput for varying numbers of PSUs in a group. The results show that the proposed framework outperforms the OFDMA-based one. Additionally, the chapter investigates the impact of grouping on sum-throughput and throughput fairness. The study reveals that the fewer the number of PSUs grouped together, the higher the fairness.

The next chapter will explore the scenario of underlay D2D communication, where interference to and from CUs must be taken into account. This scenario ensures enhanced spectral efficiency while meeting the requirements of both PSUs and CUs. Therefore, the upcoming chapter will focus on studying bandwidth and power allocation in this context.

Chapter 4

Resource distribution through collaborative resource sharing for PSUs

Contents

4.1 Introduction

The need for efficient data traffic handling in future wireless communication systems has become a pressing issue with the increasing demand for wireless con-

nectivity. The introduction of advanced technologies such as NOMA and D2D communication has been identified as a viable solution to address this challenge. These technologies allow for the efficient use of spectral and power resources, which are critical factors in meeting the demands of data traffic in 5G and beyond wireless communication systems. To leverage these technologies, we propose a novel scheme that integrates PSNs in CNs using the underlay D2D communication. This integration presents an opportunity to improve the performance of PSUs that rely on low-rate technologies and enhance spectral efficiency. With this scheme, PSNs can be integrated into CNs as an underlay network, which will coexist with the primary CUs with a controlled level of interference.

We formulate a mixed integer nonlinear programming problem for sum throughput maximization, to optimize the channel allocation and achieve the required rate of PSUs. This optimization takes into account the power budget, the users required rates, and the SIC constraints. Since the maximization problem is computationally challenging, we design a heuristic algorithm that selects the appropriate CUs to share their resources with PS clusters. Then, given this selection, we compute the optimal power allocation in each PS cluster (or group) using the KKT conditions in the Lagrange multiplier method. Evaluation results demonstrate that our approach improves spectral efficiency and provides higher sum throughput compared to other works in the literature. We also conduct extensive simulations to compare our throughput maximization approach with the fairness maximization approach. Finally, we compute the outage probability, to verify the impact of the power allocation on the throughput achieved by the PSUs in both approaches.

The rest of this chapter proceeds as follows: A detailed literature review of radio resource allocation techniques is given in Section 4.2. Section 4.3 describes the system model and presents the problem formulation. Then, we introduce the proposed optimization method in Section 4.4. Experimental results are shown in Section 4.5, along with a discussion. Finally, concluding remarks are provided in a comprehensive conclusion.

4.2 Related work

The underlay RB sharing between one CU and one D2D pair was considered in [89], the objective is to maximize the number of D2D pairs for which the QoS requirements are satisfied. An NP-hard problem was formulated, and then a heuristic algorithm was used to find a sub-optimal solution. Based on D2D multicasting, authors in [83] proposed a clustering method to deal with emergency communication scenarios. They formulated a resource allocation problem by multiplexing, over the same channel, a cluster composed of a transmitter and many receivers together with a CU. Then they used bipartite transformation and the Hungarian algorithm to solve this formulated problem. A review of recent advances in D2D communications regarding public safety applications was presented in [90]. In particular, the authors consider the topics of device discovery under critical conditions, how to establish a D2D cluster, relaying mechanisms using mobile devices, and Vehicle-to-Vehicle (V2V) communications for road safety. In all the above-mentioned works, commu-

nications rely on the OFDMA technique, which provides good communication and interference management among users, and which continues to prove its effectiveness until today. Recently, the NOMA technique has emerged, enabling more efficient resource allocation when compared to OFDMA, and this is achieved by allowing multiple users to share the same RB, thus supporting massive connectivity, reducing latency, and improving spectral efficiency, attracting thereby the attention of most current researches.

The authors of [91] studied a multi-carrier NOMA cognitive radio system, in which each licensed band dedicated to primary users, can be opportunistically used to simultaneously serve a group of secondary users. The main objective of this paper is to achieve energy efficiency. To this end, a non-convex fractional optimization problem has been formulated and solved by using Dinkelbach's algorithm to transfer the problem to a parameterized optimization problem. The authors then used an iterative optimization approach to obtain the solution for the energy efficiency maximization problem. In [92], both the full-duplex and half-duplex D2D communications were studied in the underlay CNs. NOMA is used to manage the interference between devices and CUs using mutual SIC and to boost the performance of D2D underlay systems. These works examined the power levels in NOMA using the D2D communication. However, they are limited due to the fact that resources are shared with a maximum of one D2D pair, which hinders the effective utilization of the spectrum.

The concept of NOMA-based "D2D group" communication, was introduced in [84,93], the idea is to enable one D2D transmitter to communicate with multiple D2D receivers simultaneously. The authors formulated an NP-hard problem to maximize the overall rate, and they gave the sub-optimal solution for this problem using the many-to-one two-sided matching theory. The authors argue that multiple groups can use the same RB, which is not consistent with the concept of grouping in NOMA, where devices of the same group share the same RBs. The authors of [94] presented a D2D communication network using the NOMA technique. Once the performance of the CUs degrades due to the interference caused by D2D users, these D2D users use the available resource of the unlicensed band through the duty cycle technique, helping to mitigate the interference caused to the CU and improving the data rate of the users. They formulated a nonlinear mixed integer programming problem and then used a collective intelligence approach, namely the whale optimization algorithm, to solve this problem. They did not mention how groups are formed, nor the criteria for selecting CUs and how to transit between licensed and unlicensed bands. Amer et al. in [95] conducted a literature review of the following technologies: NOMA, D2D, full-duplex, and cooperative networks. They proposed a NOMA-based system with multiple CUs and D2D transmitters multiplexed on the same sub-channel, and based on Pareto improvement, they used a two-sided stable many-to-one matching algorithm, for sub-channel assignment.

By using the NOMA technique, the authors of [46] considered a single BS and many directly connected users to it. They proposed a heuristic algorithm for grouping these users and solved the optimization problem using the Lagrange multiplier method. Moreover, they demonstrated the difference in the working principles of the

downlink and uplink NOMA technique and the influence of normalized channel gain on the performance and results. In [85], Chen et al. considered the same scenario as in [46], they added an inter-cluster dynamic programming algorithm to achieve the overall maximum energy efficiency. Authors in [61], proposed a new clustering scheme and showed that it can increase the system throughput by about eight percent compared to the same solution proposed in [46]. In [96], the authors addressed the problem of many CUs and one D2D pair sharing the same sub-channel. They derived optimal conditions for power control of CUs on each sub-channel. Then, they proposed a dual-based iterative algorithm to solve the problem. This body of work is complementary to our focus on the NOMA technique for PSUs. In fact, we employ a D2D underlay system in our prototype with multiple PSUs in the same cluster (or group) to highlight the efficiency of spectrum usage.

4.3 Problem Formulation

4.3.1 System model

In contrast to the previous chapter, which focused on using dedicated resources for PSUs in the overlay D2D communication, this chapter and as illustrated in Figure 4.1, explores the underlay D2D communication, where PS clusters share the resources of CUs.

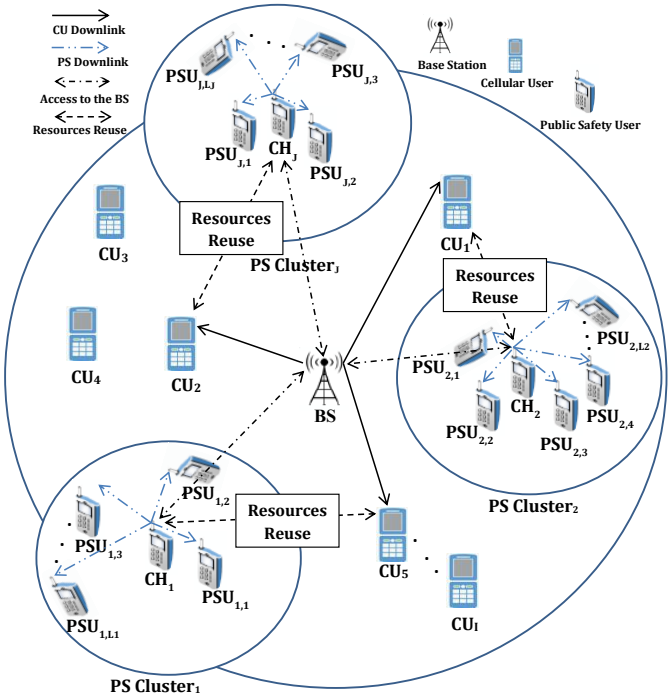

Figure 4.1: System Architecture

The downlink transmission herein adopts the OFDMA technique, therefore the interference among the CUs is negligible. Hence, by considering the underlay D2D communication in the network, the PSUs can access the CNs. Accordingly, within

Table 4.1: List of notations

Notation	Description
I	Number of CUs and subbands
J	Number of PS clusters
L_j	Number of PSUs in cluster$_j$
$h_{x,y}$	Channel gain between transmitter x and receiver y
N_0	Noise power spectral density
W_i	Bandwidth allocated to subband (i)
Pc_i	Portion of power allocated to CU_i in W_i
Ps_j^i	Remained power to PS cluster$_j$ from CU_i
$\alpha_{j,l}Ps_j^i$	Portion of power allocated to each PSU_l in cluster$_j$
$\eta_{j,i}$	Binary variable indicating whether cluster$_j$ shares CU_i resources or not
$I_{j,l}^{In}$	Interference caused to PSU_l by other PSUs within the cluster$_j$.
$I_{i,l}^{Cel}$	Interference caused to PSU_l by CU_i.
Υ	Sum of the CU interference and the noise power spectral density in the same cluster

the same subband, PSUs considered as secondary users will be multiplexed with a CU that is the primary user of that subband, and will use its remaining resources, provided that the interference caused to that CU is below a given threshold. We summarize the major notations used throughout this chapter in Table 4.1.

4.3.2 Preliminary

Considering the underlay D2D communication in the network, the allocated and unused CU resources can be shared with the PS clusters. To maintain a good level of communication in the network, intra-cluster interference and CU interference must be taken into account. The SINR of CUs and PSUs must be maintained above a predefined threshold.

We denote the power allocated to CU_i in subband W_i by Pc_i, and let Ps_j^i be the remaining power for cluster$_j$ in W_i. After performing the resource sharing, the SINR at the receiver side of the cellular communication is computed as follows:

$$SINR_{CU_i} = \frac{Pc_i \,.\, h_{BS,CU_i}}{\eta_{j,i}(Ps_j^i \,.\, h_{CH_j,CU_i}) + N_0 \,.\, W_i} \tag{4.1}$$

Where $\eta_{j,i}$ is a binary variable indicating whether cluster$_j$ shares CU_i resources ($\eta = 1$) or not ($\eta = 0$). The SINR at PS receiver side can be written as:

$$SINR_{PSU_{j,l}} = \frac{\alpha_{j,l}Ps_j^i \,.\, h_{CH_j,PSU_{j,l}}}{I_{j,l}^{In} + I_{i,l}^{Cel} + N_0 \,.\, W_i} \tag{4.2}$$

$I_{j,l}^{In}$ and $I_{i,l}^{Cel}$ represent the interference caused to PSU_l in cluster$_j$ by PSUs with higher channel gain (i.e., PSU_1, ... ,PSU_{l-1}), and the interference caused by CU_i respectively. They can be computed as follows:

$$I_{j,l}^{In} = h_{CH_j,PSU_{j,l}} . (\sum_{k=1}^{l-1} \alpha_{j,k}) Ps_j^i \tag{4.3}$$

$$I_{i,l}^{Cel} = Pc_i \ . \ h_{BS,PSU_{j,l}} \tag{4.4}$$

The rate achieved by the receiver PSU_l in cluster$_j$ can be given by:

$$R_{\mathrm{PSU}_{j,l}} = W_i \ . \ log_2(1 + SINR_{PSU_{j,l}}) \tag{4.5}$$

The total system rate is calculated as follows:

$$R_{sys} = \sum_{j=1}^{J} \sum_{l=1}^{L_j} R_{\mathrm{PSU}_{j,l}} + \sum_{i=1}^{I} R_{\mathrm{CU}_i} \tag{4.6}$$

4.3.3 Problem statement

In this Section, we formulate the sum rate maximization problem based on two parameters: $\eta_{j,i}$ and αPs_j^i. The former indicates the CUs chosen for each cluster of PS, whereas the latter denotes the power allocation for each PSU in the cluster. This optimization problem can be expressed in the following manner:

$$\max_{\eta_{j,i}, \alpha Ps_j^i} (Rsys) \tag{4.7}$$

s.t.

$$SINR_{CU_i} \geq SINR_{thr_i}, i \in \{1, ..., I\} \tag{4.7a}$$

$$SINR_{PSU_{j,l}} \geq SINR_{thr_{j,l}}, j \in \{1, ..., J\}, l \in \{1, ..., L_j\} \tag{4.7b}$$

$$\sum_{l=1}^{L_j} \alpha_{j,l} = 1, j \in \{1, ..., J\} \tag{4.7c}$$

$$\alpha_{j,l} \in [0, 1], j \in \{1, ..., J\}, l \in \{1, ..., L_j\} \tag{4.7d}$$

$$\frac{Ps_j^i}{2^{(L_j - l + 1)}} < \alpha_{j,l} Ps_j^i \leq \frac{Ps_j^i}{2^{(L_j - l)}}, j \in \{1, ..., J\}, l \in \{1, ..., L_j\} \tag{4.7e}$$

$$\sum_{i=1}^{I} Pc_i + \sum_{j=1}^{J} Ps_j^i \leq Pmax \tag{4.7f}$$

$$\eta_{j,i} \in \{0,1\}, i \in \{1,...,I\}, j \in \{1,...,J\} \tag{4.7g}$$

$$\sum_{j=1}^{J} \eta_{j,i} = 1, i \in \{1,...,I\} \tag{4.7h}$$

Wherein (4.7a) and (4.7b) ensure that the SINR of all CUs and PSUs must be maintained above the threshold level. Constraints (4.7c) and (4.7d) are related to NOMA power allocation in the PS clusters, the first constraint guarantees that the sum of the power allocated for all the PSUs in cluster$_j$ does not exceed the remaining power budget (Ps_j^i) to the CH_j, while the second ensures that the power allocation factor is positive. For an efficient SIC in the downlink NOMA cluster, (4.7e) indicates the upper and lower power bands that can be assigned to each PSU. Whereas (4.7f) ensures that the sum of the assigned powers to each subband in the network does not exceed the total power budget $Pmax$ assigned by the BS. (4.7g) shows that $\eta_{j,i}$ is a binary variable indicating whether cluster$_j$ shares CU_i resources or not. Finally, (4.7h) ensures that the resources of CU_i can only be used by a single PS cluster.

The formulated problem turns out to be infeasible when the remaining power budget cannot support the minimum SINR requirements of constraint (4.7b). Therefore, it is important to perform a feasibility check of (4.7) before solving it.

4.4 Proposed Optimization Method

4.4.1 Cellular users selection

The optimization of the problem (4.7) can be divided into two parts: first, we need to find the CUs with the maximum resources that can be shared with PS cluster$_j$, to do this we propose a heuristic algorithm (**Algorithm 3**) to multiplex the CUs and PSUs into the same resources, then we distribute the power obtained by the clusters among their PSUs, to this end, we use the Lagrange multiplier method as explained next in Section 4.4.2. Based on the number of PSUs in the cluster, their minimum rate requirements, and the SIC threshold, the CH of each cluster decides the minimum power required (P_{Req}^j). Therefore, determining the optimal transmit power conditions of CUs is of great importance. By setting the rate of each CU to its threshold, we have:

$$R_{CU_i} = R_{th_i} \tag{4.8}$$

$$W_i \,.\, log2(1 + \frac{Pc_i^{'}.h_{BS,CU_i}}{\eta_{j,i}(Ps_j^i.h_{CH_j,CU_i}) + N_0 \,.\, W_i}) = R_{th_i} \tag{4.9}$$

After performing some algebraic operations, (4.9) can be expressed as follows:

$$(2^{\frac{R_{th_i}}{W_i}} - 1) = \frac{Pc_i^{'}.h_{BS,CU_i}}{\eta_{j,i}(Ps_j^i.h_{CH_j,CU_i}) + N_0 \,.\, W_i} \tag{4.10}$$

$$Pc_i^{'} = (2^{\frac{R_{thr_i}}{W_i}} - 1)(\frac{\eta_{j,i}(Ps_j^i \,.\, h_{CH_j,CU_i}) + N_0 \,.\, W_i}{h_{BS,CU_i}}) \tag{4.11}$$

Our goal is to find the CUs with the maximum remaining resources to be used by the PS clusters. Without loss of generality, we assume that the PS clusters follow this order: $P_{Req}^1 > P_{Req}^2 ... > P_{Req}^J$.

Algorithm 3 Heuristic algorithm for choosing the best CUs

Input: I: The number of RBs and the number of CUs, J: The number of clusters.
Output: η: $(I$x$J)$ binary matrix. It indicates whether the cluster$_j$ shares the RB_i of CU_i or not.

System Initialization

$\eta_{(j,i)} = 0; \forall j \in \{1, ..., J\}$ and $\forall i \in \{1, ..., I\}$
$Temp_i^j = 0; \forall j \in \{1, ..., J\}$ and $\forall i \in \{1, ..., I\}$

Main

$j = 1$; //Starting by the first cluster
for $i = 1$ to I **do**
  if $\eta_{(:,i)} \neq 1$ // (If there is no CU already assigned) **then**
    Compute $Pc_i^{'}$ using (4.11)
    if $Pc_i - Pc_i^{'} > 0$ // (If there is a remaining power that can be shared) **then**
      $Temp_i^j = Pc_i - Pc_i^{'}$;
    else
      $Temp_i^j = 0$
    end if
  else
    $Temp_i^j = 0$
  end if
end for
$idx = \arg\max_i (Temp_i^j)$ // Choose CU i with the highest power to share
if $Pc_{idx} - Pc_{idx}^{'} \geq P_{Req}^j$ **then**
  # Single-cluster NOMA // No grouping
  $\eta_{(j,idx)} = 1$
else
  # Multi-clusters NOMA
  # Grouping PSUs into many groups
  # Choose 1 CU for each group
  $\eta_{(j,idx)} = 1$
end if
$j = j + 1$
if $j \leq J$ **then**
  go to 4
else
  Return η // The matrix η indicates shared subbands i with PS cluster j
end if

Algorithm 3 is a heuristic that chooses the best CUs in terms of the remaining power for each PS cluster. Starting with PS cluster1 and all CUs (steps 3 and 4), if the subband (i) of CU_i is not already assigned to any PS cluster (step 5), we compute $Pc_i^{'}$ using (4.11), to ensure that we have guaranteed the required threshold for CU_i after the sharing. Next, we select the CU_i with the maximum remaining power to share with the PS cluster$_j$ (step 16). Subsequently, we check whether the remaining resources are greater than the power requested by the CH_j (step 17). If it is the case, we use single-cluster NOMA and consider all the PSUs of the cluster$_j$ as a single group. If not, we group these PSUs and choose one CU for each group. In step 27, we repeat the previous iteration for all the clusters. Finally, matrix η is the output of the algorithm, where j, i^{th} element indicates the subband (i) of CU_i that is shared with the PS cluster$_j$.

In what follows, we detail the grouping method we used in case the remaining power is not sufficient for single-cluster NOMA.

The grouping method

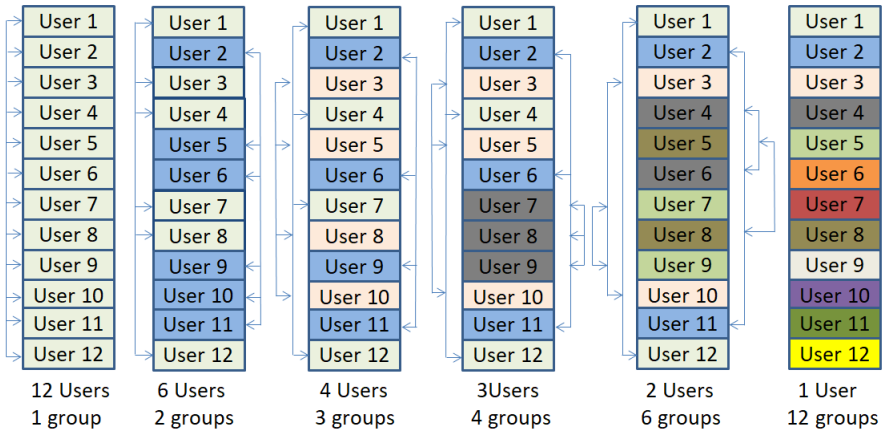

Figure 4.2: Different grouping possibilities. Users connected by the arrows (same color) are grouped into the same group.

The clusters can be divided into groups depending on the number of PSUs in each cluster and their required rate. For a small number of PSUs in the cluster, all these PSUs can be considered as a single group (single-cluster NOMA) and can share the resources with a single CU. However, when the number of PSUs increases and their required rate is no longer met, grouping the PSUs of the same cluster (multi-cluster NOMA) becomes a necessity. This grouping method ensures the satisfaction of the PSUs rate and the robustness of our proposed solution.

By improving the throughput of NOMA intermediate users, the authors of [61] proposed an improved grouping scheme over the hybrid grouping scheme widely used in the literature. Based on this work, in our scenario, when the number of PSUs increases, the grouping is performed as follows: The CH sorts its PSUs according to their channel gain. Every two PSUs will form a group. The first and last groups form two large groups by adding all the PSUs from the other groups according to their group number parity. Each of these two groups is formed to share the resources

of one CU. In case the required rate of PSUs is not met, the same process is repeated for each group. Figure 4.2 is an example of different groups of 12 PSUs in the same cluster.

4.4.2 Optimal Power Allocation in the NOMA cluster

Here we consider the case of single-cluster NOMA. In case we have multi-clusters NOMA, a similar procedure of single-cluster NOMA can be followed for each group. The total rate in cluster$_j$ is calculated as follows:

$$R_j = W_i \,.\, \sum_{l=1}^{L_j} log_2(1 + \frac{\alpha_{j,l} Ps_j^i \,.\, h_{CH_j,PSU_{j,l}}}{I_{j,l}^{In} + I_{i,l}^{Cel} + N_0 \,.\, W_i}) \tag{4.12}$$

The optimal power allocation problem for all the PSUs in each PS cluster is formulated as follows:

$$\max_{\alpha Ps_j^i} R_j \tag{4.13}$$

s.t.

$$R_{PSU_{j,l}} \geq R_{thr_{j,l}}, l \in \{1, ..., L_j\} \tag{4.13a}$$

$$\eta_{j,i} Ps_j^i \leq Pc_i - Pc_i^{'}, i \in \{1, ..., I\} \tag{4.13b}$$

$$\alpha_{j,l} Ps_j^i - \sum_{k=1}^{l-1} \alpha_{j,k} Ps_j^i \geq Pmin_l, \forall l \in \{1, ..., L_j\} \tag{4.13c}$$

Where (4.13a) ensures that the rate of each PSU must be maintained above its threshold level, (4.13b) indicates the power allocated to each cluster, and finally (4.13c) denotes the SIC constraints where $Pmin_l$ is equal to $\frac{Pmin}{h_{CH_j,PSU_{j,l-1}}}$, and $Pmin$ is the minimum power required to decode the desired signal from the entire message signal.

In the above optimization problem, the objective function is concave, constraints are affine, and state conditions are satisfied, thus problem has a unique optimal solution. Now, we consider the Lagrange of the optimization problem (4.13):

$$L(\alpha_{j,l} Ps_j, \lambda, \mu, \phi) = W_i \,.\, \sum_{l=1}^{L_j} log_2(1 + \frac{\alpha_{j,l} Ps_j^i . h_{CH_j,PSU_{j,l}}}{I_{j,l}^{In} + I_{i,l}^{Cel} + N_0 \,.\, W_i}) + \tag{4.14}$$

$$\sum_{l=1}^{L_j} \lambda_l (R_{PSU_{j,l}} - R_{thr_{j,l}}) + \mu(Pc_i - Pc_i^{'} - \eta_{j,i} Ps_j^i) +$$

$$\sum_{l=1}^{L_j} \phi_l (\alpha_{j,l} Ps_j^i - \sum_{k=1}^{l-1} \alpha_{j,k} Ps_j^i - Pmin_l)$$

By applying the KKT conditions, a solution $\alpha_{j,l}Ps_j^i$ is a local maximum if only exist λ_l, μ, and ϕ_l, called KKT multipliers, such that the following four groups of conditions hold:

• Stationarity:

$$\frac{\delta_L}{\delta_{\alpha_{j,l}Ps_j^i}} = \frac{W_i \,.\, h}{\sum\limits_{k=1}^{l} \alpha_{j,l}Ps_j^i \,.\, h + \Upsilon_j^i} + (\lambda_l - \sum_{k=l+1}^{L_j} \lambda_k.(2^{R_{\mathrm{j,k}}} - 1)) - \mu + \tag{4.15}$$

$$(\phi_l - \sum_{k=l+1}^{L_j} \phi_k) - W_i \,.\, h \,.\, \frac{\sum\limits_{k=l+1}^{L_j} \alpha_{j,k}Ps_j^i \,.\, h}{(\sum\limits_{t=1}^{k-1} \alpha_{j,t}Ps_j^i \,.\, h + \Upsilon_j^i)(\sum\limits_{t=1}^{k} \alpha_{j,t}Ps_j^i \,.\, h + \Upsilon_j^i)} = 0$$

$$\frac{\delta_L}{\delta_{\lambda_l}} = \alpha_{j,l}Ps_j^i \,.\, h - (2^{R_{j,k}} - 1)(h \,.\, \sum_{k=1}^{l-1} \alpha_{j,k}Ps_j^i + \Upsilon_j^i) = 0 \tag{4.16}$$

$$\frac{\delta_L}{\delta_{\mu}} = Pc_i - Pc_i^{'} - \eta_{j,i}Ps_j^i = 0 \tag{4.17}$$

$$\frac{\delta_L}{\delta_{\phi_l}} = \alpha_{j,l}Ps_j^i - \sum_{k=1}^{l-1} \alpha_{j,k}Ps_j^i - Pmin_l = 0 \tag{4.18}$$

Where h, Υ and $2^{R_{j,k}}$ are the abbreviations of $h_{CH_j,PSU_{j,l}}$, $I_{i,l}^{Cel} + N_0 \,.\, W_i$, and $2^{\frac{R_{thr_{j,k}}}{W_i}}$ respectively, they were used to simplify the equation, otherwise it will be very long and difficult to understand.

• Primal Feasibility (PF) :

$$R_{PSU_{j,l}} \geq R_{thr_{j,l}} \tag{PF1}$$

$$\eta_{j,i}Ps_j^i \leq Pc_i - Pc_i^{'} \tag{PF2}$$

$$\alpha_{j,l}Ps_j^i - \sum_{k=1}^{l-1} \alpha_{j,k}Ps_j^i \geq Pmin_l \tag{PF3}$$

• Dual Feasibility (DF) :

$$\lambda_l \geq 0 \tag{DF1}$$

$$\mu \geq 0 \tag{DF2}$$

$$\phi_l \geq 0 \tag{DF3}$$

• Complementary Slackness (CS) :

$$\sum_{l=1}^{L_j} \lambda_l (R_{PSU_{j,l}} - R_{thr_{j,l}}) = 0 \tag{CS1}$$

$$\mu(Pc_i - Pc_i^{'} - \eta_{j,i} Ps_j^i) = 0 \tag{CS2}$$

$$\sum_{l=2}^{L_j} \phi_l (\alpha_{j,l} Ps_j^i - \sum_{k=1}^{l-1} \alpha_{j,k} Ps_j^i - Pmin_l) = 0 \tag{CS3}$$

The stationarity conditions ensure that the gradient is within all the gradients of constraints, therefore there is no way to keep improving feasibly. The PF conditions are used to ensure that $\alpha_{j,l} Ps_j^i$ is going to satisfy all the original constraints. The DF conditions ensure that all the multipliers are greater or equal to zero. Finally, the CS conditions demonstrate that only the constraints binding at $\alpha_{j,l} Ps_j^i$ matter, if we have a constraint that does not bind, λ and ϕ would be zero and we do not care about the gradient of that constraint because it is not binding.

The Lagrange multipliers belong to three sets of constraints: $A = \mu$, $B = \{\lambda_1, \lambda_2, ..., \lambda_{L_j}\}$ and $C = \{\phi_2, \phi_3, ..., \phi_{L_j}\}$. These sets are the power allocation constraint for each cluster, the minimum data rate constraints of each PSU, and finally, the power allocation constraints for each PSU in the cluster, taking into consideration the SIC order. Based on [46], the combinations of Lagrange multiplier satisfying the KKT conditions can be computed as $2^{L_j - 1}$. For $L_j = 2$, the two combinations are: $\mu\phi_2$ and $\mu\lambda_2$. For $L_j = 3$, the four combinations are: $\mu\phi_2\phi_3$, $\mu\phi_2\lambda_3$, $\mu\lambda_2\phi_3$ and $\mu\lambda_2\lambda_3$. For $L_j = 4$, the eight combinations are: $\mu\phi_2\phi_3\phi_4$, $\mu\phi_2\phi_3\lambda_4$, $\mu\phi_2\lambda_3\phi_4$, $\mu\phi_2\lambda_3\lambda_4$, $\mu\lambda_2\phi_3\phi_4$, $\mu\lambda_2\phi_3\lambda_4$, $\mu\lambda_2\lambda_3\phi_4$ and $\mu\lambda_2\lambda_3\lambda_4$. Therefore, based on the number of PSUs inside the cluster, the solution set can be written as follows: $S = \{\mu, \lambda_2 or \phi_2, \lambda_3 or \phi_3, ..., \lambda_{L_j} or \phi_{L_j}\}$.

After stating the number of combinations of Lagrange multipliers satisfying the KKT conditions, Table 4.2 shows the optimal powers allocation and the necessary conditions for $L_j = 2, 3$ and 4 PSUs.

Two additional sets are needed to generalize the solution for the optimal power allocation within each cluster, $B' = S - B$ and $C' = S - C$. The following equation shows the generalized form of the solution, where $\alpha_{j,1} Ps_j^i$ denotes the optimal power allocation for the PSU with the highest channel gain in each cluster, while $\alpha_{j,l} Ps_j^i$ are the optimal power allocations for all remaining PSUs in the cluster.

$$\alpha_{j,1}Ps_j^i = \frac{Pc_i - Pc_i^{'}}{\prod\limits_{\substack{k=2\\k\notin B'}}^{L_j} 2^{R_{j,k}}.\prod\limits_{\substack{k=2\\k\in B'}}^{L_j} 2} - (\frac{\Upsilon}{h}).\frac{(\prod\limits_{\substack{k=2\\k\notin B'}}^{L_j} 2^{R_{j,k}} - 1)}{\prod\limits_{\substack{k=2\\k\notin B'}}^{L_j} 2^{R_{j,k}}.\prod\limits_{1}^{k'-1-x} 2} - \sum_{\substack{k=2\\k\notin C'}}^{L_j} \frac{Pmin_k}{\prod\limits_{\substack{t=2\\t\notin B'}}^{k-1} 2^{R_{j,t}}.\prod\limits_{1}^{k-1-y} 2} \quad (4.19)$$

$$\alpha_{j,l}Ps_j^i = \begin{cases} (2^{R_{j,l}} - 1).(\sum_{k=1}^{l-1} \alpha_{j,k}Ps_j^i + \frac{\Upsilon}{h}), & \text{if } l \notin B' \\ \\ \sum_{k=1}^{l-1} \alpha_{j,k}Ps_j^i + Pmin_l, & \text{if } l \in B' \end{cases}$$

k', x and y are defined respectively as: the greatest $k \notin B'$, $\sum\limits_{k\notin B'} k$ and $\sum\limits_{\substack{t<k\\t\notin B'}} t$.

Table 4.2: Powers allocation and conditions

L_j PSUs	**Optimal powers allocation**	**Necessary conditions**
$L_j = 2$	$\alpha_{j,1}Ps_j^i = \frac{Pc_i - Pc_i^{'}}{2} - \frac{Pmin_2}{2}$ $\alpha_{j,2}Ps_j^i = \frac{Pc_i - Pc_i^{'}}{2} + \frac{Pmin_2}{2} = \alpha_{j,1}Ps_j^i + Pmin_2$	$R_{PSU_{j,l}} - R_{thr_{j,l}} > 0, \forall l = 1,2$
	$\alpha_{j,1}Ps_j^i = \frac{Pc_i - Pc_i^{'}}{2^{(R_{j,2})}} - (\frac{\Upsilon}{h})(\frac{2^{(R_{j,2})}-1}{2^{(R_{j,2})}})$ $\alpha_{j,2}Ps_j^i = (2^{(R_{j,2})} - 1)(\alpha_{j,1}Ps_j^i + \frac{\Upsilon}{h})$	$R_{PSU_{j,l}} - R_{thr_{j,l}} > 0, \forall l = 1$ $\alpha_{j,l}Ps_j^i - \sum_{k=1}^{l-1} \alpha_{j,k}Ps_j^i - Pmin_l > 0, \forall l = 2$
$L_j = 3$	$\alpha_{j,1}Ps_j^i = \frac{Pc_i - Pc_i^{'}}{2.2} - \frac{Pmin_2}{2} - \frac{Pmin_3}{2.2}$ $\alpha_{j,2}Ps_j^i = \alpha_{j,1}Ps_j^i + Pmin_2$ $\alpha_{j,3}Ps_j^i = \alpha_{j,1}Ps_j^i + \alpha_{j,2}Ps_j^i + Pmin_3$	$R_{PSU_{j,l}} - R_{thr_{j,l}} > 0, \forall l = 1,2,3$
	$\alpha_{j,1}Ps_j^i = \frac{Pc_i - Pc_i^{'}}{2.2^{(R_{j,3})}} - (\frac{\Upsilon}{h})(\frac{2^{(R_{j,3})}-1}{2.2^{(R_{j,3})}}) - \frac{Pmin_2}{2}$ $\alpha_{j,2}Ps_j^i = \alpha_{j,1}Ps_j^i + Pmin_2$ $\alpha_{j,3}Ps_j^i = (2^{(R_{j,3})} - 1)(\alpha_{j,1}Ps_j^i + \alpha_{j,2}Ps_j^i + \frac{\Upsilon}{h})$	$R_{PSU_{j,l}} - R_{thr_{j,l}} > 0, \forall l = 1,2$ $\alpha_{j,l}Ps_j^i - \sum_{k=1}^{l-1} \alpha_{j,k}Ps_j^i - Pmin_l > 0, \forall l = 3$
	$\alpha_{j,1}Ps_j^i = \frac{Pc_i - Pc_i^{'}}{2.2^{(R_{j,2})}} - (\frac{\Upsilon}{h})(\frac{2^{(R_{j,2})}-1}{2^{(R_{j,2})}}) - \frac{Pmin_3}{2.2^{(R_{j,2})}}$ $\alpha_{j,2}Ps_j^i = (2^{(R_{j,2})} - 1)(\alpha_{j,1}Ps_j^i + \frac{\Upsilon}{h})$ $\alpha_{j,3}Ps_j^i = \alpha_{j,1}Ps_j^i + \alpha_{j,2}Ps_j^i + Pmin_3$	$R_{PSU_{j,l}} - R_{thr_{j,l}} > 0, \forall l = 1,3$ $\alpha_{j,l}Ps_j^i - \sum_{k=1}^{l-1} \alpha_{j,k}Ps_j^i - Pmin_l > 0, \forall l = 2$
	$\alpha_{j,1}Ps_j^i = \frac{Pc_i - Pc_i^{'}}{(2^{(R_{j,2})})(2^{(R_{j,3})})} - (\frac{\Upsilon}{h})(\frac{(2^{(R_{j,2})})(2^{(R_{j,3})})-1}{(2^{(R_{j,2})})(2^{(R_{j,3})})})$ $\alpha_{j,2}Ps_j^i = (2^{(R_{j,2})} - 1)(\alpha_{j,1}Ps_j^i + \frac{\Upsilon}{h})$ $\alpha_{j,3}Ps_j^i = (2^{(R_{j,3})} - 1)(\alpha_{j,1}Ps_j^i + \alpha_{j,2}Ps_j^i + \frac{\Upsilon}{h})$	$R_{PSU_{j,l}} - R_{thr_{j,l}} > 0, \forall l = 1$ $\alpha_{j,l}Ps_j^i - \sum_{k=1}^{l-1} \alpha_{j,k}Ps_j^i - Pmin_l > 0, \forall l = 2,3$
	$\alpha_{j,1}Ps_j^i = \frac{Pc_i - Pc_i^{'}}{2.2.2} - \frac{Pmin_2}{2} - \frac{Pmin_3}{2.2} - \frac{Pmin_4}{2.2.2}$	

	$\alpha_{j,2}Ps_j^i = \alpha_{j,1}Ps_j^i + Pmin_2$	$R_{PSU_{j,l}}$ - $R_{thr_{j,l}} > 0, \forall l = 1,2,3,4$
	$\alpha_{j,3}Ps_j^i = \alpha_{j,1}Ps_j^i + \alpha_{j,2}Ps_j^i + Pmin_3$	
	$\alpha_{j,4}Ps_j^i = \alpha_{j,1}Ps_j^i + \alpha_{j,2}Ps_j^i + \alpha_{j,3}Ps_j^i + Pmin_4$	
	$\alpha_{j,1}Ps_j^i = \frac{Pc_i - Pc_i^{'}}{2.2.2^{(\mathbf{R}_{j,4})}} - (\frac{\Upsilon}{h})(\frac{2^{(\mathbf{R}_{j,4})}-1}{2.2.2^{(\mathbf{R}_{j,4})}}) - \frac{Pmin_2}{2} - \frac{Pmin_3}{2.2}$	
	$\alpha_{j,2}Ps_j^i = \alpha_{j,1}Ps_j^i + Pmin_2$	$R_{PSU_{j,l}}$ - $R_{thr_{j,l}} > 0, \forall l = 1,2,3$
	$\alpha_{j,3}Ps_j^i = \alpha_{j,1}Ps_j^i + \alpha_{j,2}Ps_j^i + Pmin_3$	$\alpha_{j,l}Ps_j^i$ - $\sum_{k=1}^{l-1}\alpha_{j,k}Ps_j^i$ - $Pmin_l > 0, \forall l = 4$
	$\alpha_{j,4}Ps_j^i = (2^{(\boldsymbol{R}_{j,4})}-1)(\alpha_{j,1}Ps_j^i + \alpha_{j,2}Ps_j^i + \alpha_{j,3}Ps_j^i + \frac{\Upsilon}{h})$	
	$\alpha_{j,1}Ps_j^i = \frac{Pc_i - Pc_i^{'}}{2.2.2^{(\boldsymbol{R}_{j,3})}} - (\frac{\Upsilon}{h})(\frac{2^{(\mathbf{R}_{j,3})}-1}{2.2^{(\mathbf{R}_{j,3})}}) - \frac{Pmin_2}{2} - \frac{Pmin_4}{2.2.2^{(\mathbf{R}_{j,3})}}$	
	$\alpha_{j,2}Ps_j^i = \alpha_{j,1}Ps_j^i + Pmin_2$	$R_{PSU_{j,l}}$ - $R_{thr_{j,l}} > 0, \forall l = 1,2,4$
	$\alpha_{j,3}Ps_j^i = (2^{(\boldsymbol{R}_{j,3})}-1)(\alpha_{j,1}Ps_j^i + \alpha_{j,2}Ps_j^i + \frac{\Upsilon}{h})$	$\alpha_{j,l}Ps_j^i$ - $\sum_{k=1}^{l-1}\alpha_{j,k}Ps_j^i$ - $Pmin_l > 0, \forall l = 3$
	$\alpha_{j,4}Ps_j^i = \alpha_{j,1}Ps_j^i + \alpha_{j,2}Ps_j^i + \alpha_{j,3}Ps_j^i + Pmin_4$	
	$\alpha_{j,1}Ps_j^i = \frac{Pc_i - Pc_i^{'}}{2.(2^{(\boldsymbol{R}_{j,3})})(2^{(\boldsymbol{R}_{j,4})})} - (\frac{\Upsilon}{h})(\frac{(2^{(\boldsymbol{R}_{j,3})})(2^{(\boldsymbol{R}_{j,4})})-1}{2.(2^{(\boldsymbol{R}_{j,3})})(2^{(\boldsymbol{R}_{j,4})})}) - \frac{Pmin_2}{2}$	
	$\alpha_{j,2}Ps_j^i = \alpha_{j,1}Ps_j^i + Pmin_2$	$R_{PSU_{j,l}}$ - $R_{thr_{j,l}} > 0, \forall l = 1,2$
	$\alpha_{j,3}Ps_j^i = (2^{(\boldsymbol{R}_{j,3})}-1)(\alpha_{j,1}Ps_j^i + \alpha_{j,2}Ps_j^i + \frac{\Upsilon}{h})$	$\alpha_{j,l}Ps_j^i$ - $\sum_{k=1}^{l-1}\alpha_{j,k}Ps_j^i$ - $Pmin_l > 0, \forall l = 3,4$
$L_j = 4$	$\alpha_{j,4}Ps_j^i = (2^{(\boldsymbol{R}_{j,4})}-1)(\alpha_{j,1}Ps_j^i + \alpha_{j,2}Ps_j^i + \alpha_{j,3}Ps_j^i + \frac{\Upsilon}{h})$	
	$\alpha_{j,1}Ps_j^i = \frac{Pc_i - Pc_i^{'}}{2.2.2^{(\boldsymbol{R}_{j,2})}} - (\frac{\Upsilon}{h})(\frac{2^{(\mathbf{R}_{j,2})}-1}{2^{(\mathbf{R}_{j,2})}}) - \frac{Pmin_3}{2.2^{(\mathbf{R}_{j,2})}} - \frac{Pmin_4}{2.2.2^{(\mathbf{R}_{j,2})}}$	
	$\alpha_{j,2}Ps_j^i = (2^{(\boldsymbol{R}_{j,2})}-1)(\alpha_{j,1}Ps_j^i + \frac{\Upsilon}{h})$	$R_{PSU_{j,l}}$ - $R_{thr_{j,l}} > 0, \forall l = 1,3,4$
	$\alpha_{j,3}Ps_j^i = \alpha_{j,1}Ps_j^i + \alpha_{j,2}Ps_j^i + Pmin_3$	$\alpha_{j,l}Ps_j^i$ - $\sum_{k=1}^{l-1}\alpha_{j,k}Ps_j^i$ - $Pmin_l > 0, \forall l = 2$
	$\alpha_{j,4}Ps_j^i = \alpha_{j,1}Ps_j^i + \alpha_{j,2}Ps_j^i + \alpha_{j,3}Ps_j^i + Pmin_4$	
	$\alpha_{j,1}Ps_j^i = \frac{Pc_i - Pc_i^{'}}{2.(2^{(\boldsymbol{R}_{j,2})}).(2^{(\boldsymbol{R}_{j,4})})} - (\frac{\Upsilon}{h})(\frac{(2^{(\boldsymbol{R}_{j,2})}).(2^{(\boldsymbol{R}_{j,4})})-1}{(2^{(\boldsymbol{R}_{j,2})}).(2^{(\boldsymbol{R}_{j,4})})}) - \frac{Pmin_3}{2.2^{(\boldsymbol{R}_{j,2})}}$	
	$\alpha_{j,2}Ps_j^i = (2^{(\boldsymbol{R}_{j,2})}-1)(\alpha_{j,1}Ps_j^i + \frac{\Upsilon}{h})$	$R_{PSU_{j,l}}$ - $R_{thr_{j,l}} > 0, \forall l = 1,3$
	$\alpha_{j,3}Ps_j^i = \alpha_{j,1}Ps_j^i + \alpha_{j,2}Ps_j^i + Pmin_3$	$\alpha_{j,l}Ps_j^i$ - $\sum_{k=1}^{l-1}\alpha_{j,k}Ps_j^i$ - $Pmin_l > 0, \forall l = 2,4$
	$\alpha_{j,4}Ps_j^i = (2^{(\boldsymbol{R}_{j,4})}-1)(\alpha_{j,1}Ps_j^i + \alpha_{j,2}Ps_j^i + \alpha_{j,3}Ps_j^i + \frac{\Upsilon}{h})$	
	$\alpha_{j,1}Ps_j^i = \frac{Pc_i - Pc_i^{'}}{2.(2^{(\boldsymbol{R}_{j,2})}).(2^{(\boldsymbol{R}_{j,3})})} - (\frac{\Upsilon}{h})(\frac{(2^{(\boldsymbol{R}_{j,2})}).(2^{(\boldsymbol{R}_{j,3})})-1}{(2^{(\boldsymbol{R}_{j,2})}).(2^{(\boldsymbol{R}_{j,3})})}) - \frac{Pmin_4}{2.(2^{(\boldsymbol{R}_{j,2})}).(2^{(\boldsymbol{R}_{j,3})})}$	
	$\alpha_{j,2}Ps_j^i = (2^{(\boldsymbol{R}_{j,2})}-1)(\alpha_{j,1}Ps_j^i + \frac{\Upsilon}{h})$	$R_{PSU_{j,l}}$ - $R_{thr_{j,l}} > 0, \forall l = 1,4$
	$\alpha_{j,3}Ps_j^i = (2^{(\boldsymbol{R}_{j,3})}-1)(\alpha_{j,1}Ps_j^i + \alpha_{j,2}Ps_j^i + \frac{\Upsilon}{h})$	$\alpha_{j,l}Ps_j^i$ - $\sum_{k=1}^{l-1}\alpha_{j,k}Ps_j^i$ - $Pmin_l > 0, \forall l = 2,3$
	$\alpha_{j,4}Ps_j^i = \alpha_{j,1}Ps_j^i + \alpha_{j,2}Ps_j^i + \alpha_{j,3}Ps_j^i + Pmin_4$	

	$\alpha_{j,1}Ps_j^i = \frac{Pc_i - Pc_i'}{(2^{(R_{j,2})})(2^{(R_{j,3})})(2^{(R_{j,4})})} - (\frac{\Upsilon}{h})(\frac{(2^{(R_{j,2})})(2^{(R_{j,3})})(2^{(R_{j,4})})-1}{(2^{(R_{j,2})})(2^{(R_{j,3})})(2^{(R_{j,4})})})$	
	$\alpha_{j,2}Ps_j^i = (2^{(R_{j,2})} - 1)(\alpha_{j,1}Ps_j^i + \frac{\Upsilon}{h})$	$R_{PSU_{j,l}} - R_{thr_{j,l}} > 0, \forall l = 1$
	$\alpha_{j,3}Ps_j^i = (2^{(R_{j,3})} - 1)(\alpha_{j,1}Ps_j^i + \alpha_{j,2}Ps_j^i + \frac{\Upsilon}{h})$	$\alpha_{j,l}Ps_j^i - \sum_{k=1}^{l-1}\alpha_{j,k}Ps_j^i - Pmin_l > 0, \forall l = 2,3,4$
	$\alpha_{j,4}Ps_j^i = (2^{(R_{j,4})} - 1)(\alpha_{j,1}Ps_j^i + \alpha_{j,2}Ps_j^i + \alpha_{j,3}Ps_j^i + \frac{\Upsilon}{h})$	

4.5 Simulation Results

4.5.1 Simulation setup

Table 4.3: Default values of the simulation parameters.

Simulation parameters	Default values
d_{min}	1 m
d_{max}	100 m
Power per subband	26 dBm
Noise power spectral density (N_0)	-100 dBm
SIC threshold	-10 dBm
Bandwidth of a RB	180 Khz
Carrier frequency	2 Ghz
Shadow fading	7 dB
L_{urban}	2.3 dB
Maximum runs	1000

In this Section, we evaluate the performance of our solution. Extensive simulations have been conducted using our simulation framework implemented in MATLAB. The scenario is a BS located at the center of a square area of 1000 m x 1000 m, and many PSUs located in a cluster with a radius of 100 m.

The channel gain $h_{\mathrm{x,y}}$ between the transmitter x and the receiver y including the path-loss and the small-scale fading, is computed as follows: $h_{\mathrm{x,y}} = L + L_{\mathrm{fad}}$. The path-loss of all our simulated PSUs is generated randomly between the minimum distance (d_{min}) in line-of-sight (LoS) and the maximum distance (d_{max}) in non-line-of-sight ($NLoS$), using the free-space path loss model [83, 97]. (4.20) and (4.21) show the two path-loss equations in LoS and $NLoS$ used for our simulations.

$$L_{LoS}(d) = 32.45dB + 20log_{10}(f_{UE}) + 20log_{10}(d_{min}/1000) \tag{4.20}$$

$$L_{NLoS}(d) = 9.5dB + 45log_{10}(f_{UE}) + 40log_{10}(d_{max}/1000) + L_{urban} \tag{4.21}$$

d_{min} and d_{max} are the minimum and maximum distance between the CH and a PSU and are equal to 1 m and 100 m respectively. The noise power density at each PSU is chosen to be $N_0 = -100$ dBm. The power allocated for each RB in the downlink communication is equal to 26 dBm [98]. Table 4.3 shows the simulation parameters used in our experiments.

4.5.2 Simulation results

The simulations are made up of three parts: We first benchmark our NOMA-based resource allocation method with the OFDMA-based method, and compare it with the work done in [91, 92]. Then, we use Jain's fairness index [88] to assess whether users are getting a fair share of resources. Finally, we study the outage probability in both methods, to see whether PSUs reach the required rate or are out of service.

4.5.2.1 Our `Max − throughput` method versus state-of-the-art works

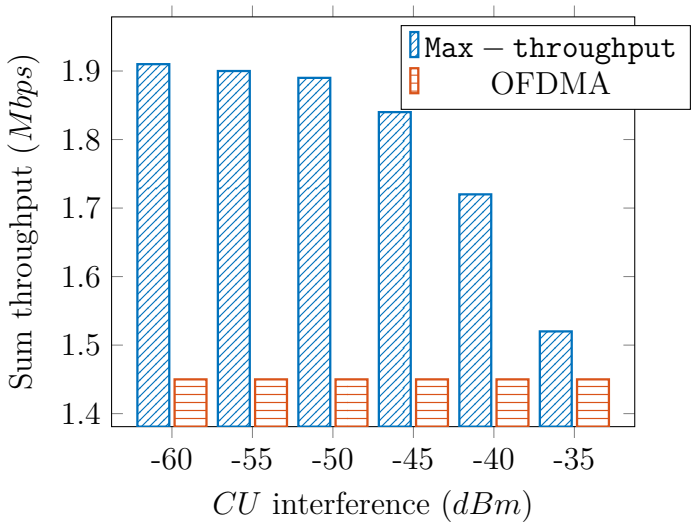

Figure 4.3: Variation of sum throughput with different values of CU interference

We investigate the impact of CU interference on the sum throughput. We consider four PSUs sharing the same RB with a CU. We set the power dedicated to the cluster to 23 dBm. We plot the sum throughput obtained by both, our NOMA-based and the OFDMA-based allocation, as a function of CU interference (see Figure 4.3). Our method yields higher sum throughput values when compared with OFDMA-based allocation. Depending on the interference caused by the CU, these values vary from 1.9 to 1.52 $Mbps$ in our method, whereas due to the absence of interference

between users in the second method, a fixed value of 1.45 $Mbps$ is obtained. However, if the CU interference continues to increase, the sum throughput decreases to reach lower values than those achieved by OFDMA allocation.

The Figure 4.4 shows the comparison of the average user throughput achieved by the two approaches. By setting the interference caused by the CU to -40 dBm and varying the number of PSUs in the cluster, we observe a decrease in the average throughput from 0.835 to 0.289 $Mbps$ in our proposed method when the number of PSUs increases from two to six. Whereas, in the second method, it decreases from 0.823 to 0.222 $Mbps$. Notably, our allocation method consistently outperforms the OFDMA-based method in terms of average user throughput for the same number of PSUs.

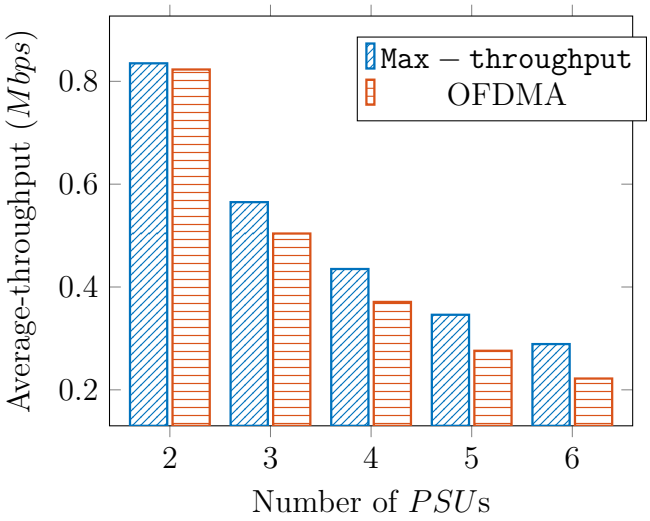

Figure 4.4: Variation of Average-throughput with the number of PSUs

When the number of PSUs increases in the cluster, it is necessary to use the multi-clusters NOMA and select more CUs to share their resources with this cluster. To make this selection, the CH divides all its connected PSUs into groups. Each group will separately use the resources of one CU. Figure 4.2 details the grouping method we used when the number of PSUs in the cluster increases. Figure 4.5 shows the comparison between the total number of PSUs and the number of allocated subbands. We considered 20, 40, 60, 80, 100, and 120 PSUs with three different throughput thresholds, namely 250, 400, and 800 $Kbps$. The 800 $Kbps$ threshold can be reached by all users when up to two PSUs share the same subband, while the 400 $Kbps$ threshold can be reached when up to four PSUs share the same subband, and finally, the 250 $Kbps$ threshold is reached when six PSUs share the same subband. As the number of PSUs increases, the total number of subbands required increases to meet the interference constraints. Furthermore, with the increase in the throughput threshold, the total number of subbands required in turn increases, due to the decrease in the number of PSUs grouped together. Therefore, depending on the throughput threshold of the PSUs, the more PSUs are in the

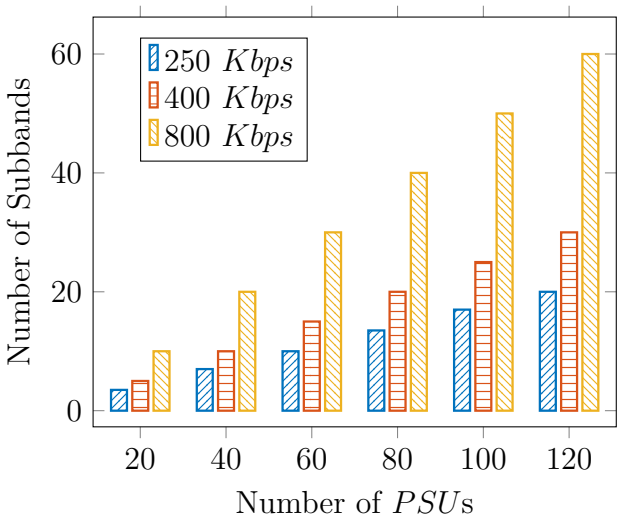

Figure 4.5: Variation of the number of RBs with the number of PSUs

same group to share the same subband, the fewer subbands are required to satisfy that throughput threshold. This results in better spectral efficiency in our method compared to the work in [91, 92], in which a maximum of two PSUs can share the same subband regardless of the required rate.

For example, if we consider the scenario of 60 users and the cases of two and three PSUs using the same resources. Calculating the spectral efficiency as a fraction of the throughput divided by the bandwidth, we get 0.311 $bits/sec/Hz$ for two PSUs and 0.4722 $bits/sec/Hz$ for three PSUs, representing a spectral efficiency of 52%.

4.5.2.2 Rate fairness

Figure 4.6 (a) depicts the throughput obtained in both scenarios, while (b) shows the fairness obtained. As the number of PSUs increases from two to six, the sum throughput increases for both the $\texttt{max} - \texttt{throughput}$ and $\texttt{max} - \texttt{fairness}$ methods (a). However, it increases at a higher ratio when it comes to the $\texttt{max} - \texttt{throughput}$ (a, blue line (+)). This is because the method does not address the fairness issue among the users, and hence the high throughput is achieved by the user with the highest channel gain. On the other hand, the fairness in terms of throughput achieved by these users is low and varies between 0.5 and 0.62 (b, blue line (+)). To maximize fairness, we decrease the power allocated to the user with the highest channel gain. As a result, the throughput achieved by this user will decrease, allowing other users to achieve the same throughput, and resulting in maximum fairness among users (b, black line (*)), with a sum throughput slightly below the optimal value (a, black line (*)).

Table 4.4 shows an example of the throughput achieved by four users in both the $\texttt{max} - \texttt{throughput}$ and $\texttt{max} - \texttt{fairness}$ methods, using random channel gains.

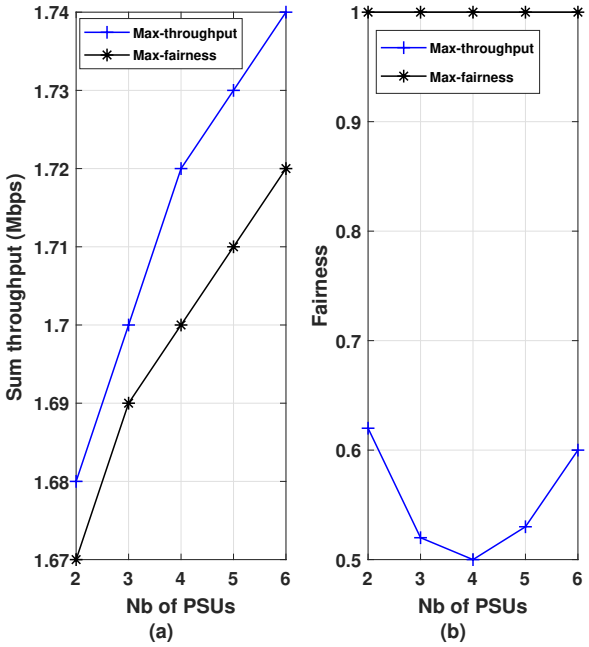

Figure 4.6: Throughput and fairness for different numbers of PSUs

Table 4.4: The comparison of the throughput achieved in both Maximum throughput and Maximum fairness methods

User	1	2	3	4
Channel gain	0.154	0.108	0.079	0.043
$\texttt{Max}-\texttt{throughput}$ (Mbps)	1.195	0.1816	0.1812	0.1805
$\texttt{Max}-\texttt{fairness}$ (Mbps)	0.42824	0.42824	0.42824	0.42824

Figure 4.7 illustrates the comparison of the sum throughput of 2 PSUs as a function of different power levels. The simulation is performed for both NOMA ($\texttt{Max}-\texttt{throughput}$ and $\texttt{Max}-\texttt{fairness}$) and OFDMA-based methods. We observe that the total throughput increases with the increase of power allocated to the cluster. In addition, the sum throughput obtained with the two NOMA-based methods shows better performance than that achieved with the OFDMA-based method. More importantly, the $\texttt{Max}-\texttt{throughput}$ method always presents better results than the $\texttt{Max}-\texttt{fairness}$ method. This is due to the maximum power allocated to the user with the best channel gain.

4.5.2.3 Outage probability

Figure 4.8 plots the outage probability for the $\texttt{Max}-\texttt{throughput}$ method for two PSUs in the same group. Different power and throughput levels are considered. Similarly, Figure 4.9 presents the same for the $\texttt{Max}-\texttt{fairness}$ method. The power varies between 0.004 and 0.4 W, the throughput varies between 100 $Kbps$ and 1.26

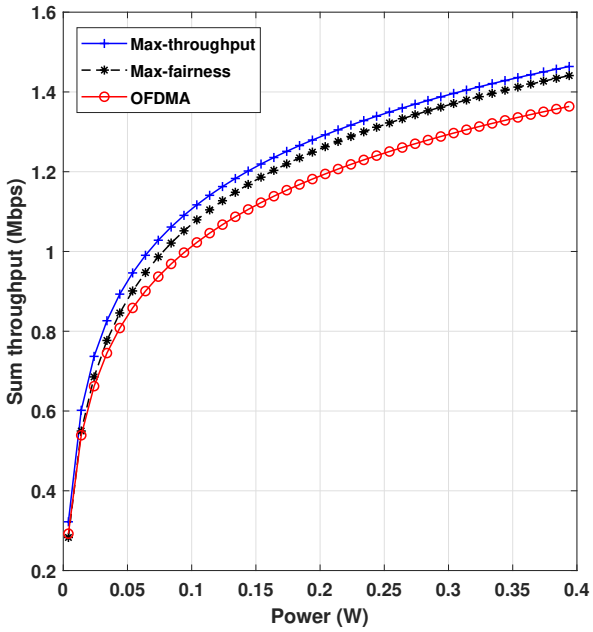

Figure 4.7: Sum throughput variation according to the power allocated to the PSUs

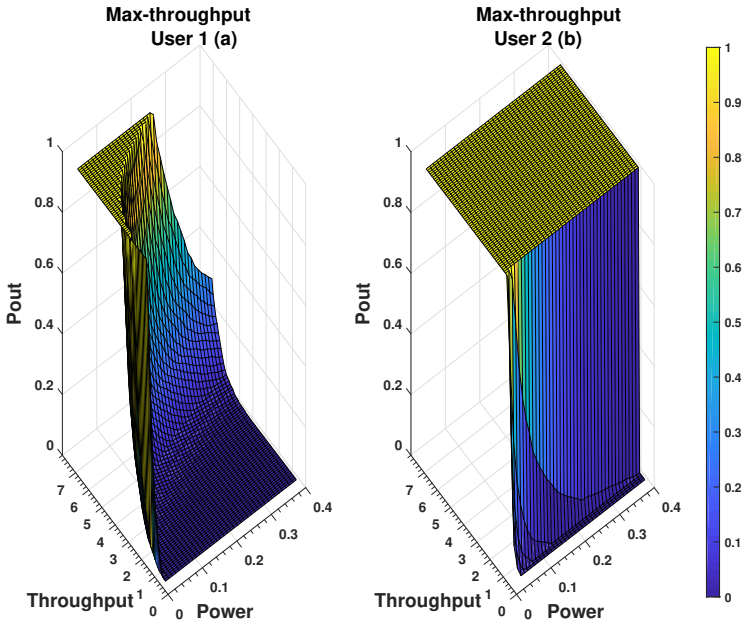

Figure 4.8: Outage probability of 2 PSUs in the `Max − throughput` method versus the remaining CU power.

$Mbps$ (which is plotted between 0.55 and seven), and the outage probability varies between zero and one. When the power increases and the throughput decreases, the outage probability tends to zero, which corresponds to the satisfaction of the

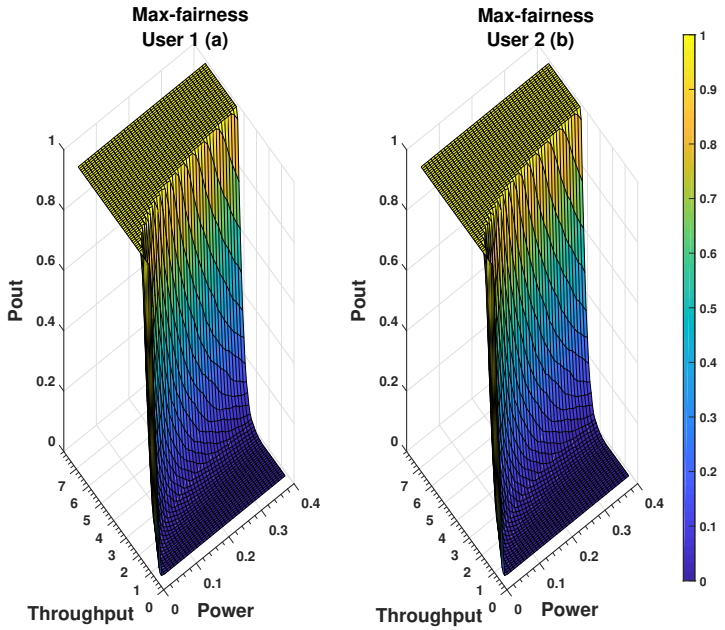

Figure 4.9: Outage probability of 2 PSUs in the $\texttt{Max}-\texttt{fairness}$ method versus the remaining CU power.

achieved throughput by the PSUs. Figure 4.8 demonstrates that there is a significant variation in the outage probability between the two PSUs. For the user with better channel gain (a), when the power increases (more than 0.25 W), the highest throughput of seven can be reached with a very low outage probability, while for the user with lower channel gain (b), it reaches a maximum of one with negligible outage for a power greater than 0.1 W. Whereas, Figure 4.9 clearly illustrates that the same outage probability is obtained by both PSUs (a and b). A throughput between 3.5 and 4 is attained for both users, with a very low outage probability, when the power is greater than 0.25 W.

4.6 Summary

This chapter discusses the application of the NOMA technique and the underlay D2D communication for clusters of PSUs. The aim is to allocate power to these PSUs in a way that maximizes the sum-throughput. To achieve this, a heuristic algorithm is used to select the appropriate CUs for the clusters. Then, a convex optimization problem is formulated under power, rate, and SIC constraints to optimally allocate the remaining power of the selected CUs to the PSUs in the clusters. The problem is then solved using the generalized Lagrange multiplier method (The KKT approach). The simulation reveals interesting results which show that our proposed solution outperforms the one using OFDMA, and also has better spectrum efficiency compared to existing works. Furthermore, it shows the comparison between $\texttt{Max}-\texttt{throughput}$ and $\texttt{Max}-\texttt{fairness}$ methods in terms of sum throughput, fairness, and outage probability.

In some PS scenarios, it may not be possible for the CH to connect to the BS. This situation prevents communication with the PSUs in the network. Finding an alternative and securing this communication is of utmost importance as many people living in danger may depend on it. One way to alleviate this issue is through the MEC system, which brings certain functions closer to the end-users, resulting in reduced latency and improved user experience. The next chapter presents a new PS architecture that incorporates both the NOMA technique and the MEC system to improve communication among PSUs and ensure its occurrence in both licensed and unlicensed bands.

Chapter 5

NOMA and MEC integration in a proposed PS architecture: enhancing efficiency and ensuring connectivity

Contents

5.1 Introduction

The existence of CNs is crucial to provide broadband services for PS purposes. However, during a disaster, damage to the infrastructure such as BSs can disrupt PSUs trying to access the core network, even if a relay station is utilized. To ensure uninterrupted PS services in all circumstances and to tackle the aforementioned issue, ProSe standard plays a crucial role. It enables direct communication between devices, known as D2D communication, to take place in licensed and unlicensed frequency bands. The present chapter introduces a novel architecture for PS that

employs NOMA technique and incorporates the ProSe function and ProSe application server into the MEC system. The primary goal of this approach is to provide network access to the highest possible number of PSUs using NOMA technique. Additionally, the MEC system helps to acquire the information required by PSUs with minimal latency in the licensed spectrum, while providing secure and managed operation for these users in the unlicensed spectrum.

The effectiveness of the proposed architecture is evaluated using the Simu5G network simulator. The analysis revealed that the incorporation of the NOMA technique led to a significant improvement in spectral efficiency of up to 28.8%. Furthermore, the incorporation of MEC system in the proposed architecture is highly beneficial in reducing the latency and improving reliability. In this regard, we analyze the impact of increasing the number of requested applications and numerology index on the task response time, which is a critical metric for PSUs.

The remaining part of this chapter is structured as follows: In Section 5.2, a brief overview of related work and 5G network simulator is provided. Section 5.3 elaborates on the proposed architecture and simulation environment. The performance evaluation and validation of the architecture are presented in Section 5.4. Finally, Section 5.5 concludes the chapter.

5.2 Related Work

The idea of integrating PSNs into CNs is a recent concept, and efforts are ongoing to propose technologies and architectures to achieve this goal. The 3GPP community has developed a set of standards for mission-critical functions in broadband networks, such as LTE and 5G. 3GPP TS 23.303 [11] was the first to address the specific needs of PS. Group calling, ProSe, and push to talk were addressed in this release, and then critical video and interworking with existing PMR systems (i.e., narrowband/broadband interworking) were addressed later in 3GPP TS 36.101 [99] and 3GPP TR 21.915 [100]. Recently, ProSe has regained attention for its implementation in 5G systems with release 17 focusing on three main functions: 5G ProSe Direct Discovery, 5G ProSe Direct Communication, and 5G ProSe UE-to-Network Relay [12].

A review of recent advances in D2D communications regarding PS applications such as search and rescue missions, coverage extension, and road safety, is presented in [90]. Many topics related to emergency deployment, spectrum management, and radio resource management schemes in LTE-based PSNs are discussed in [8], the authors also study the architecture of LTE-based PSNs and provide different deployment and migration solutions. In [4], different possible PS scenarios in 5G networks are presented, as well as the standardization process of integrating existing PSNs with commercial CNs, and their challenges. The various PS communication infrastructures are described in [101], as well as the key features of the PSN and the role of adding cognitive radio technology in enhancing the first responder experience by improving their communication capabilities. In [102], the authors review the various possible use cases for PS, the status of 3gpp standards, and future challenges. They discuss the need to support mobile backhauling in LTE-based moving-cell scenarios.

Based on Release 12, authors of [103] have thoroughly reviewed D2D operations in the beyond 4G network. Their comprehensive insights, covering from the system architecture to the radio interface, boost the knowledge to practice D2D communications over cellular networks. Chochliouros et al. present a specific use case for PSNs in [104], which is based on the architectural approach proposed by the 5G ESSENCE project and focuses on MEC in 5G. This use case intends to offer a critical push-to-talk service, along with chat and location services. In [105], the authors present a non-standalone 5G ETSI MEC-based architecture for mission-critical push-to-talk (MCPTT) services. They also examine a scenario where the E-UTRAN operates in isolation, which would occur when the connection to eNodeBs is lost. In [106], the focus is on exploring how critical communications can be accomplished using a 5G network, two different use cases were explored: 1) enabling priority communications on a commercial mobile network, and 2) establishing rapidly deployable networks for emergency and tactical operations. The authors further describe the key enabling technologies for these communications.

To realize the emergency communication for PSNs most efficiently, in this chapter we propose to improve the above studies by using MEC system and NOMA technique in 5G and beyond networks. Additionally, since PS personnel are considered as communication devices with limited battery capacity, choosing one of them as CH can severely impact their battery lifetime and thus leave many first responders without any means of communication. Therefore, we prioritize the use of a dedicated CH with the appropriate communication equipment and power supply to ensure uninterrupted communication during emergencies.

After developing a PS architecture, we validate it by evaluating the effect of the location of the MEC system in the network, and by extending some functionalities of the Simu5G simulator, notably by integrating the NOMA technique. We refer to Chapter 4, and we use the optimal power calculated to be assigned to each transmitted signal in the NOMA technique, taking into account that the interference from cellular users is equal to zero.

Simu5G is a popular 5G simulation tool that provides a 3GPP compliant 5G simulation model library based on the OMNeT++ framework. It allows simulating the complete protocol stack using models for each layer of the stack, making it easy to extend the simulator's functionality and validate new network architecture. Despite the availability of other simulation tools, such as Vienna 5G SL, 5G-LENA, 5G-air simulator, etc. we decided to use Simu5G because it implements the MEC system, which is compliant with the ETSI standard. Additionally, the other simulators have some limitations. For instance, Vienna 5G SL only models the physical and MAC layers, 5G-LENA is limited to time division duplex (TDD) mode, and 5G-air simulator does not support network-controlled D2D communications and dual connectivity scenarios, just like 5G-LENA [107].

5.3 Network architecture

In this Section, we present our proposed architecture for supporting PSUs in 5G and beyond CNs. We also describe the simulation design used to test and evaluate

the architecture.

5.3.1 Proposed architecture

To ensure that PS services are always accessible, we are proposing a PS architecture in 5G and beyond networks that incorporates MEC-integrated ProSe and NOMA. The proposed architecture, illustrated in Figure 5.1, consists of two parts: the edge cloud tier and the core cloud tier. The former includes the MEC host, which is a compact core placed close to the PSUs to meet their specific demands. While the latter comprises the necessary network and application control functions required to connect and communicate among PSUs, as well as the edge cloud management elements, known as the MEC orchestrator. The MEC host is in charge of running MEC applications and MEC services through its virtualization infrastructure. Standard MEC APIs are used by MEC applications to access MEC services. Two entities, known as MEC platform manager and virtualization infrastructure manager, monitor and manage the status of the MEC host. They provide information such as available computing resources such as RAM, CPU, and disk, as well as available MEC services like location and radio network information services. This information is provided to the MEC orchestrator during MEC application operation.

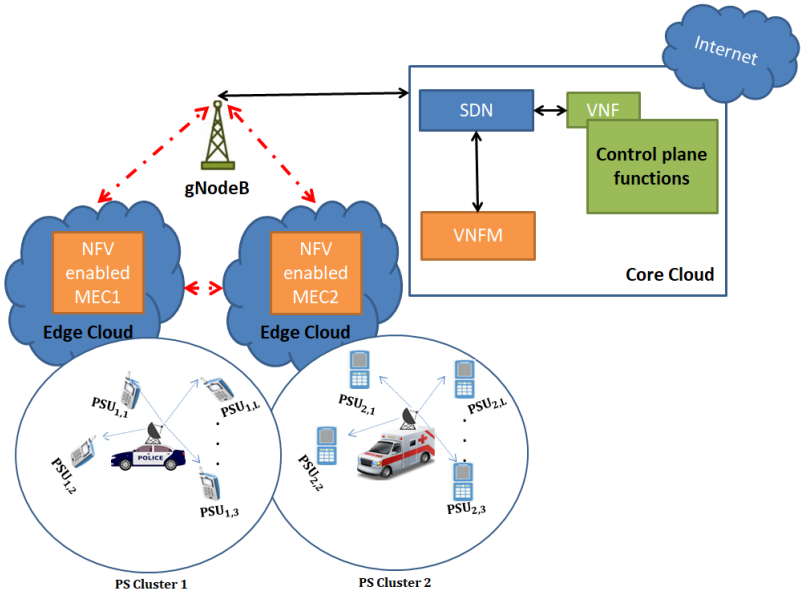

Figure 5.1: Proposed architecture

We apply NOMA within PS clusters, where we only consider the intra-cluster interference. The cluster is established around a central node referred to as the CH. The latter is a movable device deployed by a security agency at a specific location, serving as a resource allocation and management tool. It is used to manage communication among PSUs that are experiencing connectivity difficulties. Additionally, it includes a local core server that stores information about all PSUs connected to it and sends this information regularly to the core network through a next generation NodeB (gNodeB).

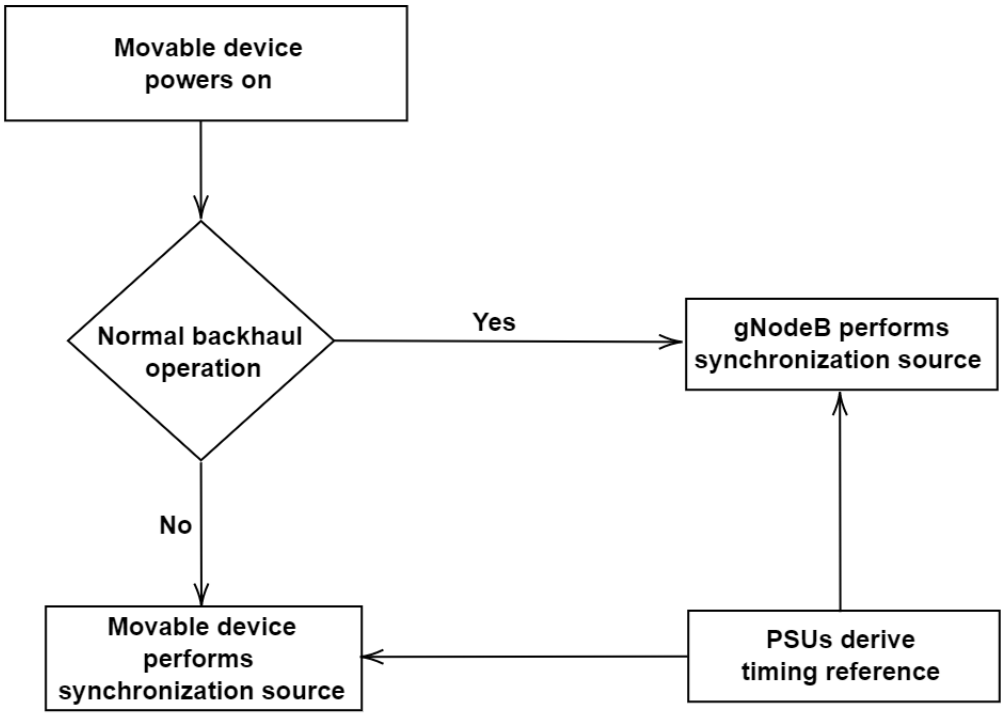

Figure 5.2: PSUs synchronization procedure

PSUs are considered to be in partial coverage or out-of-coverage. When the movable device powers on, if a synchronization signal from a gNodeB can be detected, all PSUs synchronize to the gNodeB. Otherwise, PSUs are in out-of-coverage. Figure 5.2 depicts a diagram containing the PSUs synchronization procedure. When the movable device detects a disruption in its connection to the gNodeB, it becomes a synchronization source. Thus, the timing reference of PSUs is derived from the movable device. Synchronization signals are the first step to be performed when a PSU turns on, and contain information such as the frame number for D2D communications, system bandwidth, synchronization source ID, synchronization source type, etc. [103].

The integration of ProSe into LTE networks has helped PSNs that previously relied on outdated technologies to upgrade and improve their bandwidth and reduce communication latency [36]. In this chapter, we suggest positioning the PS function and the PS application server in the MEC host, which are referred to as the ProSe function and ProSe app server by the 3GPP TS [10–12]. This approach ensures the continuity of PS services in all scenarios and allows communication to occur in either the licensed or unlicensed spectrum.

Our proposed 5G service-based architecture is depicted in Figure 5.3. The movable device contains the PS MEC host, which is connected to both the core network and the MEC orchestrator through a UPF. This UPF is crucial for the integration of MEC into the 5G network, it performs the same routing and packet forwarding functions as the S-GW and P-GW user plane in LTE networks. The PS function provides the necessary infrastructure (user and control plane functions) for secure di-

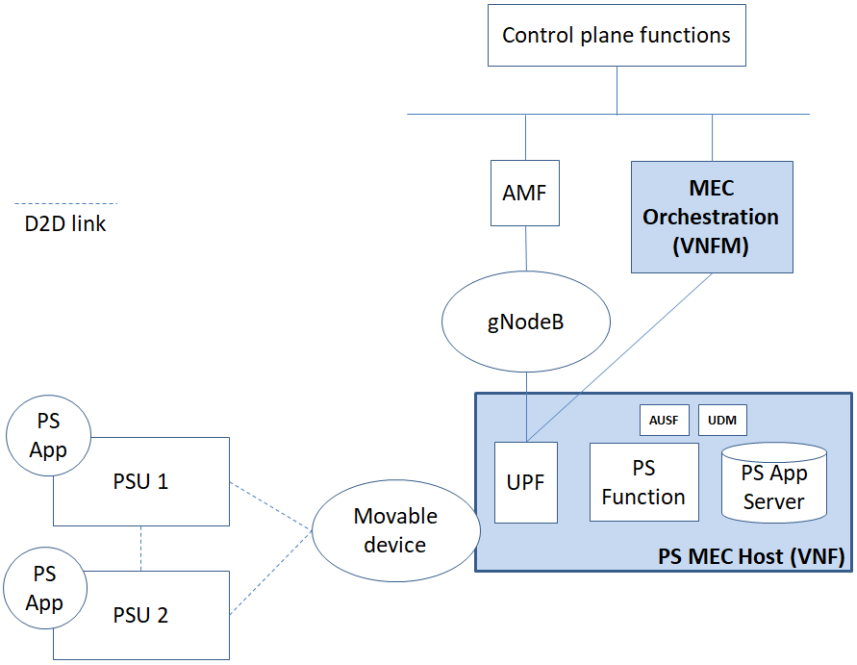

Figure 5.3: 5G service-based architecture

rect communication among PSUs; it provisions, stores, and maintains a unique PS identifier for each connected PSU, it also ensures voice communication, real-time video analysis, location-based services for emergency response, etc. This guarantees that private mission-critical push-to-talk (MCPTT) calls and communication within default groups, based on organizational membership, can still take place in the event of a disaster. The PS application server manages PS connections at the application layer and provides support for the PS function. The local EPC server contains a UDM function and an AUSF that store PSUs information and regularly update and send them to the core network.

5.3.2 Simu5G and our simulation design

In this Section, we briefly introduce Simu5G, the new OMNET++ based model library for simulating 5G networks. We also discuss the specifics and simulation design of our proposed architecture.

Simu5G is a 5G new radio network simulator; it is the evolution of the well-known 4G network simulator "SimuLTE". Following the decoupling of the user and control planes in the 5G network, Simu5G only models the user plane functions. Regarding the control plane functions, a Binder module is implemented in Simu5G. This module maintains the network data structures, and contains network information that is accessed by the network nodes via direct method calls, thus enabling the control plane functions without the need to set up complex protocol state machines. As shown in Figure 5.4, Simu5G consists of modules that communicate via messages in the form of frames or packets. These modules can be simple or compound modules, having gates that represent the input and output interfaces of the module from where

messages are sent and received, and are linked by arrows called connections [107].

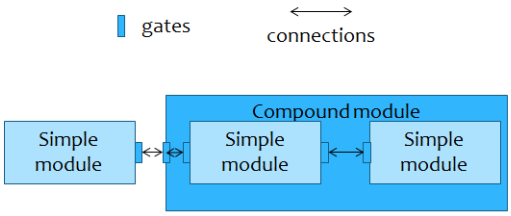

Figure 5.4: OMNET++ module connection

Table 5.1: NrNic protocol stack and their main functions

Protocol	**Main functions**
PDCP	Header compression Assigns/creates a connection identifier (CID) for packets
RLC	Specifying the buffering mode (transparent mode (TM), unacknowledged mode (UM), acknowledged mode (AM))
MAC	Time slots (TTI) Numerology Scheduling
PHY	Transmitted and received power Channel model Signal-to-interference-plus-noise ratio (SINR) Block error rate (BLER)

The compound modules New radio User equipment (NrUe) and gNodeB are the main components of the Simu5G library, providing New Radio (NR) capabilities. The NrUe module includes all protocol layers (from physical to application), while the gNodeB module includes only protocols up to layer 3. The physical layer of each module contains an NR Network interface card (NrNic) implementing NR functionality. In addition, the gNodeB module contains an additional point-to-point (P2P) interface for wired connectivity to the core network. The NrNic is made up of a protocol stack including packet data convergence protocol (PDCP), radio link control (RLC), medium access control (MAC), and physical protocol (PHY); their main functions are summarized in Table 5.1. It is important to note that the NrNic UE module has a dual stack of these protocols (one for LTE and another for NR) to allow dual connectivity with both [107].

Figure 5.5 shows the Simu5G-based simulation network of our proposed architecture. We consider a movable device connected to both a gNodeB via the x2 interface, and the MEC system via UPFs. The MEC system contains a MEC orchestrator, a user application lifecycle management proxy (UALCMP), and three

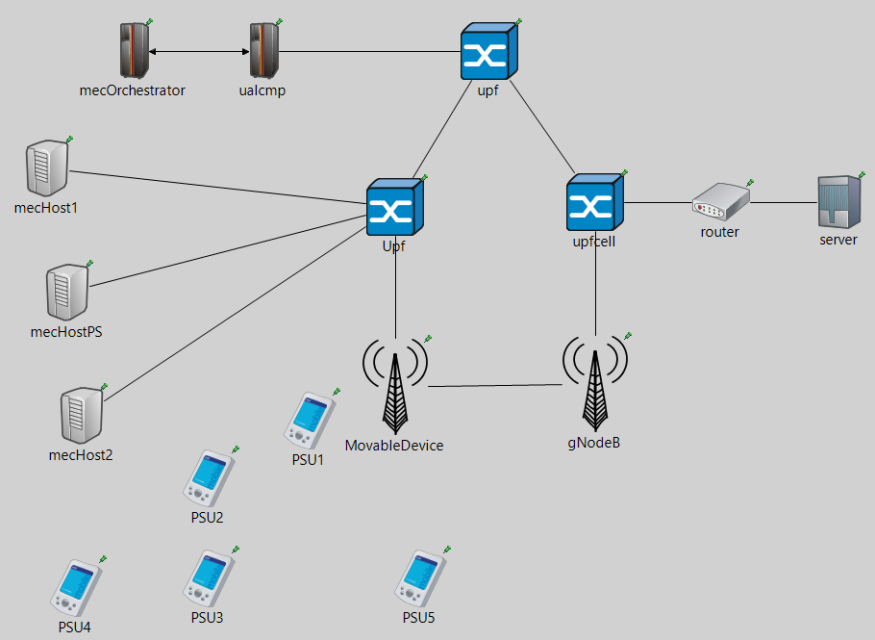

Figure 5.5: Simu5G-based simulation network

MEC hosts, namely mecHost1, mecHost2, and mecHostPS that are associated with the movable device. The PSUs are randomly deployed around this device, and run the device application that requests the instantiation of the MEC application from the UALCMP. Two scenarios are considered: the first is when PSUs receive data from the movable device in the downlink, and the second is when PSUs run a PS application to request content from the network. This content is supposed to be on-boarded in the mecHostPS during network initialization [108].

The evaluation conducted in Section 5.4 is performed based on the above two scenarios. In the first scenario, we consider the NOMA technique and show its impact on the spectral efficiency, while in the second scenario, we study the impact of MEC host location and the numerology index on the latency of accessing the required information. The numerology index determines the symbol duration, the subcarrier spacing, and the cyclic prefix length, which are essential parameters in communication systems.

5.4 Evaluation

The evaluation of the proposed architecture is made up of two parts. In the first part (Section 5.4.1), PSUs receive 800 *kbps* user datagram protocol (UDP) packets, in the downlink traffic, from the movable device randomly located at (511 m, 415 m) from the top left origin. Initially, for the evaluation of NOMA, we consider the numerology index equal to zero. Then, in the second part (Section 5.4.2) and by using Matlab, we analyze the Bit Error Ratio (BER) of the SINR obtained in Section 5.4.1. Finally, in the third part (Section 5.4.3), we consider PSUs that use different PS applications requesting some content from the network. The simulation parameters are shown in Table 5.2.

Table 5.2: Default values of the simulation parameters.

Simulation parameters	Default values
Bandwidth	10 MHz (50 Resource Blocks)
Carrier Frequency	2 GHz
Numerology	0 - 3 (15 - 120 KHz subcarrier spacing)
movable device Tx power	46 dBm
PSUs Tx power	23 dBm
Simulation time (A)	10 s
Simulation time (B)	50 s
SIC threshold	10 dBm
CPU speed	330000 instructions per second

5.4.1 Spectral efficiency evaluation for both NOMA and OFDMA techniques

With a potential rise in the number of PSUs in disaster scenarios and the crucial nature of their communication, it is imperative to ensure this communication with the limited resources allocated to them, as lives depend on it and every moment counts. NOMA is a very promising technique for 5G and beyond, it allows multiplexing many users with the same resources. To evaluate the impact of using the NOMA technique on the performance of the architecture, we consider the spectral efficiency as a metric, which is calculated according to (5.1).

$$Spectral\ efficiency = \frac{Data\ rate\ (bps)}{Channel\ Bandwidth\ (Hz)} \tag{5.1}$$

We started the experiments with two PSUs (namely PSU1 and PSU2) randomly located at (409 m, 439 m) and (329 m, 498 m), approximately 105 m and 200 m away from the movable device. As shown in Table 5.3, we first benchmark our NOMA-based method with the OFDMA-based method in terms of spectral efficiency. Our method yields better performance results than the OFDMA-based method, with a spectral efficiency increase of 7%.

Next, we investigate the impact of the propagation model (path loss, shadowing, and fading) on spectral efficiency. We plot the spectral efficiency obtained by our NOMA-based method and the OFDMA-based method (see Figure 5.6), given linear user mobility. We made both PSUs travel a distance of 50 m, moving them closer to the movable device and further away from it. The movement of PSUs consists of a horizontal translation along the x-axis with a translation angle of 0°and 180°, respectively towards and away from the vehicle. Figure 5.6 is generated by recording the spectral efficiency at each position reached by the PSUs, in both the NOMA and OFDMA-based methods. We observe that our NOMA-based method always

Table 5.3: Spectral efficiency of NOMA versus OFDMA for two PSUs

	OFDMA		**NOMA**	
User	1	2	1	2
SINR	47.62	44.33	28.81	1.058
log2(1+SINR)	5.6035	5.5024	4.8977	1.0412
Spectral efficiency	**5.5530**		**5.9389**	

results in better spectral efficiency compared to the OFDMA-based method. This is due to the reduced resource consumption when the users are multiplexed through NOMA. Furthermore, when both PSUs are headed towards the vehicle and attain 50 m, the spectral efficiency increases by 10% compared to only 4% when PSUs attain 50 m away from the vehicle. This is because the distance between the PSUs and the vehicle has a significant impact on the propagation model.

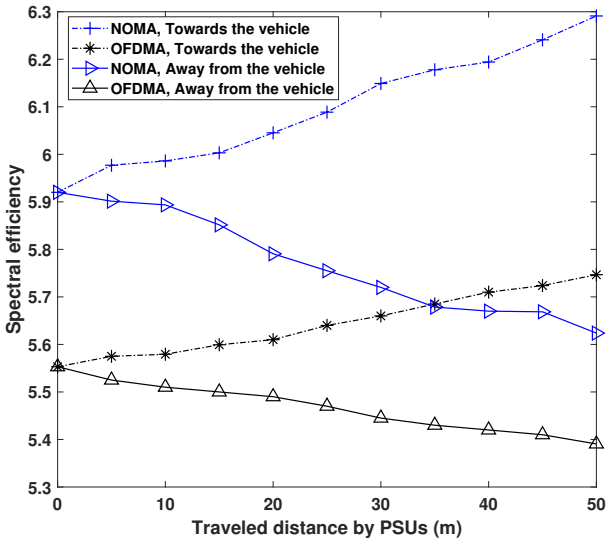

Figure 5.6: Spectral efficiency versus the PSUs traveled distance

We extend the assessment of the proposed architecture, by increasing the number of PSUs connected to the network and attempting to share the same resources using the NOMA technique. Table 5.4 shows the comparison of the spectral efficiency obtained by our NOMA-based approach and the one based on OFDMA, for three and four users (by adding PSU3 and PSU4 in Figure 5.5, respectively). The results show that our proposed approach achieves 12.3% and 28.8% better spectral efficiency for three and four users, respectively, compared to the OFDMA approach. These findings highlight the superior performance of our NOMA-based approach, particularly when more PSUs are connected to the network. This improvement is due to the ability of the NOMA technique to allow multiple users to share the same resources and confirms that the more PSUs are multiplexed together, the more efficient the spectrum usage.

Table 5.4: Spectral efficiency of NOMA versus OFDMA for three and four PSUs

Nb	**SINR**		**log2(1+SINR)**		**Spectral efficiency**	
	OFDMA	NOMA	OFDMA	NOMA	OFDMA	NOMA
3 PSUs						
1	54.13	17.72	5.7848	4.2265		
2	49.9	1.058	5.6696	1.0412	**5.6183**	**6.3089**
3	41.24	1.058	5.4005	1.0412		
4 PSUs						
1	45.21	12.15	5.5301	3.717		
2	39.27	1.058	5.3316	1.0412	**5.3094**	**6.8406**
3	38.19	1.058	5.2924	1.0412		
4	32.91	1.058	5.0836	1.0412		

When the number of PSUs in the same NOMA group continues to increase, the power allocated to the group will no longer meet the minimum SINR required by these users. It will therefore be necessary to use the multi-cluster NOMA, which consists of grouping the PSUs and then allocating resources to each group separately. We have not considered the multi-clusters NOMA in this chapter, our main concern is to implement NOMA in Simu5G and to show the impact of the channel model, which is 3GPP compliant, on this implementation.

5.4.2 Bit error ratio analysis

We then examine the performance of the OFDMA and NOMA techniques in terms of BER. Our analysis considers the theoretical equation (5.2) which provides an approximation for the AWGN model and for a rectangular constellation of M-QAM, where M = 2^k and k is an even number [109].

$$P_B \approx \frac{2(1-L^{-1})}{log_2 L} Q[\sqrt{(\frac{3log_2 L}{L^2-1})\frac{2E_b}{N_0}}] \tag{5.2}$$

Figure 5.7 illustrates the BER results as a function of different SINR levels. The simulation is conducted for one user in OFDMA and two users in NOMA techniques. Two scenarios are analyzed: the first scenario involves adjusting the power coefficient for NOMA-based users (left vs. right sides), while in the second scenario, we investigate different modulation orders (upper vs. lower sides). As can be noticed, the NOMA power coefficients have a significant impact on the BER of NOMA-based users. When more power is assigned to user 2, the BER of both users becomes closer.

The values $\alpha 1 = 0.95$ and $\alpha 2 = 0.05$ indicate that user 1 is receiving the maximum power in its assigned power range. This is in accordance with our simulations in Section 5.4.1, which are based on throughput maximization. The results indicate that OFDMA users can achieve an SINR of about 45 dB, with a BER of less than 10^{-6}. The same is for user 1 in NOMA, which achieves an SINR of about 28 dB.

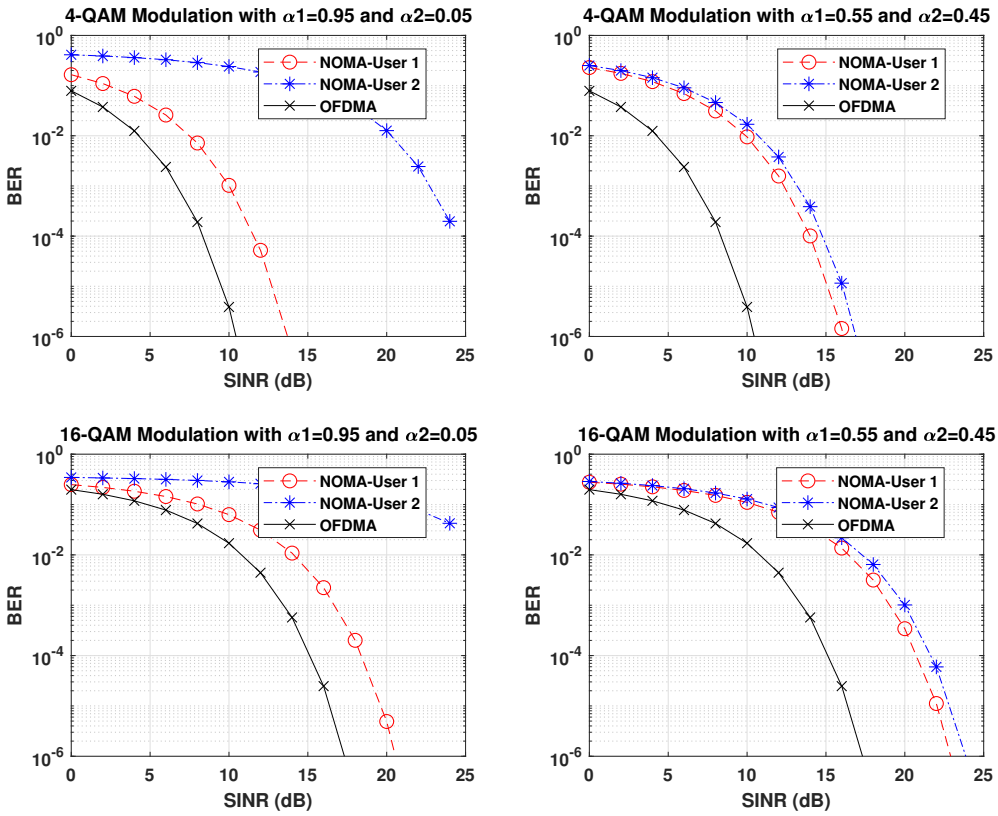

Figure 5.7: BER signal performance for OFDMA and 2-users NOMA using 4 and 16-QAM modulations.

However, user 2, who experiences interference from user 1 and has a lower channel gain, has a significantly poor BER of about 10^{-1}. The performance of user 2 in terms of BER improves as the power allocated to it increases and the modulation order decreases.

Figure 5.8 depicts the BER results for three NOMA-based users. When aiming for maximum throughput, users 2 and 3 in the NOMA technique have inferior BER performance compared to user 1 and the OFDMA user. Achieving a good BER becomes more complex as the number of users in a NOMA group increases. Increasing the modulation order or the number of users in a NOMA group leads to an increase in the BER. To address this issue, users with lower channel gain should receive a higher power allocation, or lower modulation orders should be used for these users. When the maximum available power is reached and the BER remains high, alternative modulation techniques, such as frequency-based modulation, should be considered.

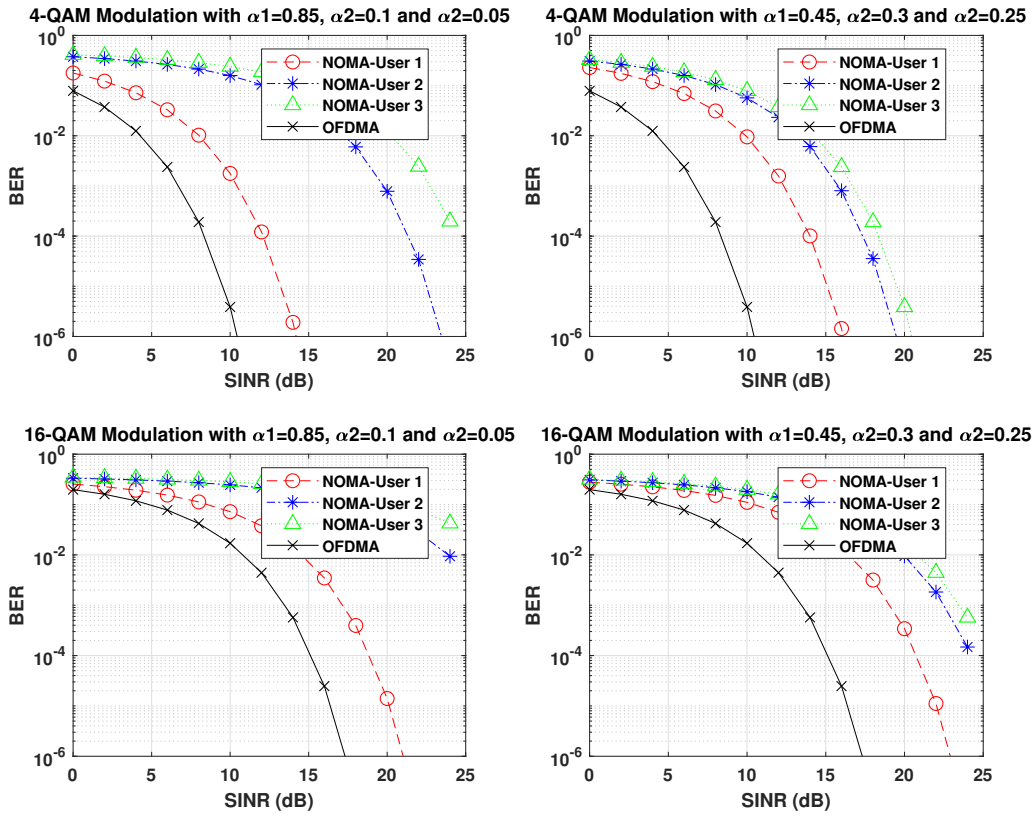

Figure 5.8: BER signal performance for OFDMA and 3-users NOMA using 4 and 16-QAM modulations.

5.4.3 Latency measurement when integrating ProSe into the MEC

The architecture is also being evaluated for critical PS applications that require a very low response time to complete some of their tasks. PSU5 uses applications that request certain contents from the network, including the radio network information service (RNIS) that collects radio access network-level information, and the location-based service. The MEC hosts have the same computational capacity, but only mecHostPS has these services running on its MEC platform, hence it will be chosen by the MEC orchestrator to deploy the MEC application. After the reception and computation of the requests, mecHostPS sends back the response. The task response time consists of several phases, including the uplink and downlink delays that occur between the PSU and the MEC application. Additionally, the processing time that depends on the load of the request and the MEC host's CPU instructions. Lastly, the service response time, comprising the service response delay as well as the network delay required to access the MEC host responsible for providing the service.

The response time of the different phases, including uplink (Up) and downlink (Down) times, processing time (Proc), and service response time (Serv), is shown in Figure 5.9. As the number of applications increases from 15 to 90, the overall

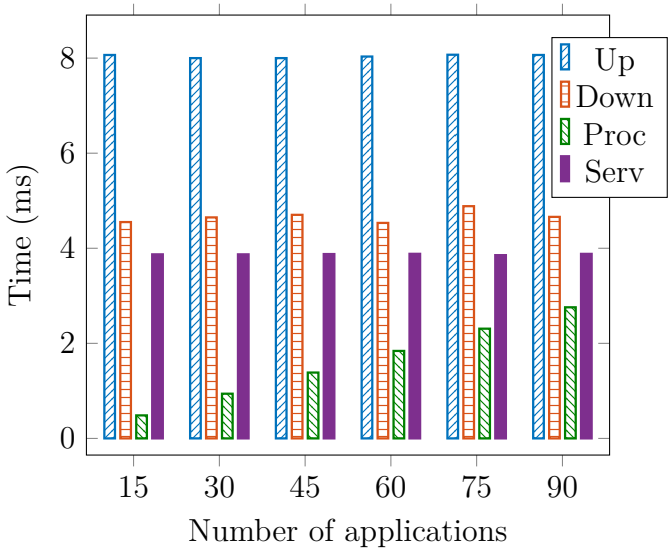

Figure 5.9: Average response time of each phase of the task request for a variable number of applications with a numerology equal to zero

response time, which is the sum of all the phases, increases from 17.461 ms to 19.366 ms. This increase is due to the time required to reach the MEC host and to process the request. Since in our simulation, the MEChostPS is directly connected to the movable vehicle through a UPF, the only delay affecting the response time is the processing delay. This delay is considered the optimal delay to reach the services. However, if the services are not located in the MEC host directly connected to the vehicle, the network time to reach the MEC host increases, resulting in an additional latency of 30 ms with an S-GW and P-GW located in a centralized MEC site, 10 ms with the X2 interface, or 200 μs with a virtual network function (VNF) [108], depending on the scenario.

Figure 5.10 compares the average response time of each phase based on different numerologies. The numerologies used in the comparison are 0, 1, 2, and 3, which correspond to subcarrier spacing of 15 kHz, 30 kHz, 60 kHz, and 120 kHz respectively. The results show high reliability with no packet loss and low latencies across all numerologies. Furthermore, as the numerology increases from zero to three, the overall delay decreases considerably from 17.33 ms to 5.974 ms with no packet loss. The major contributors to the delay in this scenario are the uplink and downlink delays, while the other phases remain relatively constant.

The results indicate that the proposed PS architecture enhances the communication capabilities of PSUs by improving spectral efficiency and achieving ultra-high reliability and very low latencies, under different numerologies and numbers of applications requested from the MEC host.

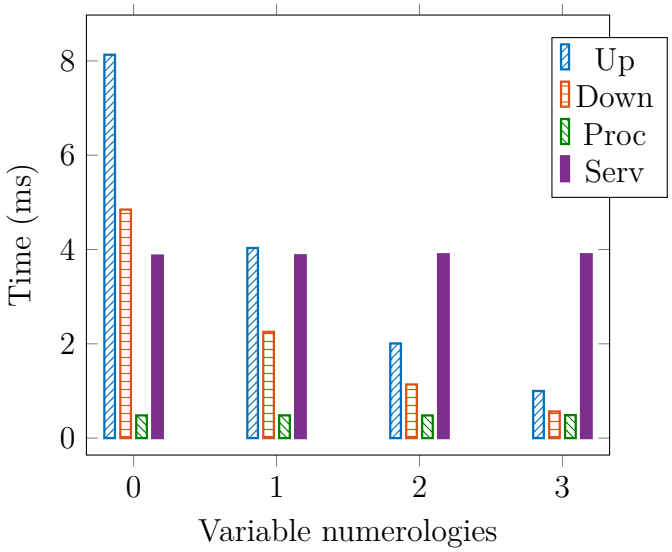

Figure 5.10: Average response time of each phase of the task request for variable numerology with a number of applications equal to 15

5.5 Summary

This chapter delves into the development of a MEC-based architecture designed for PSUs in 5G and beyond networks. The main objective is to improve the communication performance for PSUs by providing communication for a maximum number of PSUs while minimizing latency. To achieve this goal, we explore the integration of the NOMA technique and the ProSe standard in the architecture. We evaluate the effectiveness of this approach using the Simu5G simulator. The results of this study indicate that the integration of the NOMA technique has significant advantages in terms of spectral efficiency, which means that more PSUs can be served with fewer resources. Additionally, the integration of ProSe into the MEC system ensures that PSUs can receive the necessary information with very low latency and high reliability, both in the in-band and out-of-band, which is essential for real-time communication. Furthermore, we analyze the impact of different numerologies and the number of applications requested on the latency. This analysis is crucial to evaluate the performance of the proposed MEC architecture in various usage scenarios.

Chapter 6

Conclusions and Future Work

Contents

In this thesis, we addressed two main problems of resource allocation for PSUs in CNs. First, the resource allocation for PSUs by using the remaining resources of CUs considering the overlay D2D communication. Second, the resource allocation by simultaneously sharing with the CUs their resources in the underlay D2D communication. Furthermore, we proposed an architecture that incorporates ProSe standard into the MEC system to guarantee the connectivity of PSUs in both the in-band and out-of-band.

6.1 Summary

To tackle the first problem, we propose a novel bandwidth and power allocation framework in a NOMA-based system that incorporates overlay D2D communication for PSUs. The primary objective of this framework is to maximize the sum-throughput within each PSUs group, while also ensuring that the minimum data rate threshold for each PSU is met. To achieve this, our framework is divided into two distinct stages. The first stage involves the implementation of a heuristic algorithm designed to select the appropriate PSUs and form groups based on their channel gain. In the second stage, we took advantage of the PSO algorithm to allocate the powers for PSUs in each group. Through this choice, we could maximize the system's overall performance and ensure that the required data rate threshold for each PSU is achieved. The simulation results revealed an improved sum-throughput compared to the OFDMA technique. Additionally, we investigated the impact of the number of grouped users, which is more beneficial for throughput than for fairness. Finally, we shed light on a key element: the difference in the fairness of user throughput between the $\texttt{Max} - \texttt{throughput}$ and $\texttt{Max} - \texttt{fairness}$ approaches, showing that is a trade-off such that achieving optimal fairness requires a decrease in the sum-throughput.

Subsequently, we addressed the second problem by introducing an approach for the sharing of CUs' resources among PSUs in the context of underlay D2D communication. This approach involves the simultaneous utilization of CUs' resources by PSUs. A major concern we took into account at this level was the interference that could arise between CUs and PSUs. Indeed, following the formulation of a mixed-integer nonlinear problem, we developed a heuristic algorithm to select CUs while preventing interference that could disrupt their operation. After that, we leveraged the Lagrange and KKT conditions to achieve an optimal allocation of the available power among the PSUs within each formulated group. To evaluate the effectiveness of our proposed approach, we conducted extensive simulations. These simulations involved comparing our approach with an OFDMA-based method, as well as other relevant works in the literature. The simulations presented interesting results, highlighting the superiority of our proposed solution over the OFDMA-based approach. Additionally, our approach exhibited superior spectrum efficiency when compared to existing works that restrict the maximum number of users in each group to two. Finally, we conducted comprehensive simulations to assess the trade-offs in terms of throughput, fairness, and outage probability. These simulations provided valuable insights into the performance characteristics of our approach.

Moving on to the more global perspective, we proposed a MEC-based architecture for PSUs in 5G networks and beyond. The objective of this architecture was to maximize the number of PSUs that can access the network while ensuring their communication in both the in-band and out-of-band. To achieve this, we considered the NOMA technique and proposed placing the ProSe standard in the MEC host located in a movable device. This device is connected to the BS and helps the PSUs to communicate by assigning the necessary resources. When the connection to the BS is established, this device will be synchronized with the BS, and the PSUs will operate in the in-band. However, if the connection with the BS is interrupted, the PSUs, with the help and management of ProSe, will continue to operate in the out-of-band. We assessed the effectiveness of our proposed architecture by conducting simulations using the Simu5G simulator. The results demonstrated that the incorporation of the NOMA technique led to a significant enhancement in spectral efficiency when compared to the OFDMA-based approach. This allowed for a greater number of PSUs to access the network. Furthermore, we analyzed the impact of maximizing the throughput on the BER experienced by each user. The results show that PSUs with weak channel conditions experienced a bad BER compared to the user with best channel conditions. Finally, the integration of ProSe into the MEC system ensured that PSUs could receive the required information with minimal latency and high reliability. This functionality is applicable in both the in-band and out-of-band, which is crucial for maintaining uninterrupted PSU operations. We examined the impact of varying the numerology and the number of requested applications on latency. We observed that increasing the numerology index led to a decrease in latency due to reduced uplink and downlink times. Conversely, an increase in the number of requested applications resulted in higher latency due to the additional processing time required.

6.2 Future Work

Throughout this thesis, we have shown the significant advantages of using D2D communication, NOMA technique, and MEC system for PSUs communications. Yet, several other interesting directions remain unexplored and deserve further investigation. In the following, we introduce some of these promising directions.

• **Considering both throughput and fairness metrics:** In the present work, we have focused on separately optimizing the metrics of throughput and fairness, maximizing each without considering the other. However, combining these metrics into a unified problem, while introducing a controllable factor that takes both aspects into account, can provide a significant trade-off between both metrics. This trade-off involves balancing the objective of maximizing throughput while simultaneously ensuring fairness among the PSUs. Moreover, this can yield an additional advantage of enhancing the BER experienced by users with weak channel conditions, leading to improved overall system performance.

• **Improving Simu5G simulations:** To further improve the simulations carried out in Chapter 5, it would be useful to increase the number of PSUs and to consider the use of a multi-cluster NOMA. In addition, it would be beneficial to implement not only the localization of ProSe in Simu5G but also its functionality. Improving the simulation capabilities by considering more PSUs would not only improve the overall system performance, but also enable a better representation of real-world scenarios, thus improving the accuracy of the results. Moreover, the exploration of ProSe functionality presents an opportunity to integrate new features and functions into the simulation, thus improving the effectiveness and efficiency of the system.

• **Improving coverage for disaster situations:** To improve coverage in disaster situations, it may be useful to consider scenarios other than movable devices, such as Unmanned Aerial Vehicles (UAV). By using UAVs, the system can cover more areas that would otherwise be inaccessible or too difficult for movable devices to reach. For example, following a natural disaster, certain regions may be cut off from the rest of the world, so drones can be deployed to deliver essential supplies to these remote areas.

• **Integrating Artificial Intelligence (AI) and Augmented Reality (AR) technologies for PSUs:** Both technologies, AI and AR, are continuously gaining attention in various fields due to their capacity to collect, analyze and enhance the quality of results. Recognition applications and real-time data analysis in AI and wearable devices in AR can be considered crucial requirements for PS. For instance, the integration of these two technologies can significantly improve PSNs by providing remote assistance to PSUs. Remote assistance is a vital aspect of PS, especially when PSUs require immediate support from experts with specialized knowledge or experience. With AI and AR, remote experts can gain visibility in real-time into the same environment as the PSUs. Consequently, this enables them to guide PSUs on which direction to go, which obstacles to avoid, and which area to prioritize in search and rescue missions.

• **Taking advantage of the network slicing in PS applications:** Network slicing and its ability to divide a physical network into several virtual networks, each with its own set of resources, priorities, and QoS requirements, can provide different types of services and applications for PSs. Through network slicing, PSNs can dedicate specific slices to particular users, ensuring that their critical applications and services have the resources and priorities they need to operate efficiently. For instance, during a disaster or other emergency situation, network slicing can be used to allocate dedicated network slices to PSUs. This ensures that their communication channels are not impacted by the high network traffic, which could potentially delay or prevent emergency response operations.

Appendix

To ensure proper downlink connectivity under NOMA, the signal power transmitted from the cluster head to the devices must be set correctly. This is because the cluster head enables the devices to operate over the same spectrum. Therefore, their signals must be transmitted at different power levels to perform SIC. More precisely, the signal power of any device is required to be higher than the sum of the signal powers of all devices with relatively higher channel gain [46]. We assume there are L devices connected to their cluster head, then the constraints on the signals powers are as follows:

$$P_2 > P_1 \tag{A.1}$$

$$P_3 > P_2 + P_1 \tag{A.2}$$

$$\cdots$$

$$P_i > P_{i-1} + ... + P_1 \tag{A.i}$$

$$\cdots$$

$$P_L > P_{L-1} + ... + P_1 \tag{A.L}$$

To satisfy the above conditions, we first want to maximize the transmission power of device 1 (which is the device with the highest channel gain). This is done by using the result of Lemma 1 in paper [46].

Hence $\alpha_{\mathrm{j},1} P s_j^i \leq \frac{P s_j^i}{2^{(L_j - 1)}}$.

Next, to perform SIC between device 1 and device 2 which has the second highest channel gain, we need P2 to be greater than P1 and take a value outside its range, thus $\frac{P s_j^i}{2^{(L_j - 1)}} < \alpha_{\mathrm{j},2} P s_j^i$.

P3 should be greater than P1, P2, and P1 + P2 to perform SIC. In the extreme case $\alpha_{\mathrm{j},1} P s_j^i$ takes the highest value $(\frac{P s_j^i}{2^{(L_j - 1)}})$ and $\alpha_{\mathrm{j},2} P s_j^i$ takes the smallest value $((\frac{P s_j^i}{2^{(L_j - 1)}}) + \delta)$.

Thus $\alpha_{\mathrm{j},1}Ps_j^i + \alpha_{\mathrm{j},2}Ps_j^i = (2((\frac{Ps_j^i}{2^{(L_j-1)}}) + \delta)) = ((\frac{Ps_j^i}{2^{(L_j-2)}}) + \delta)$, this means $\alpha_{\mathrm{j},3}Ps_j^i > (\frac{Ps_j^i}{2^{(L_j-2)}})$ and obviously SIC requires $\alpha_{\mathrm{j},2}Ps_j^i \leq (\frac{Ps_j^i}{2^{(L_j-2)}})$

Hence $(\frac{Ps_j^i}{2^{(L_j-1)}}) < \alpha_{\mathrm{j},2}Ps_j^i \leq (\frac{Ps_j^i}{2^{(L_j-2)}})$.

The same reasoning applies to the rest of the devices and we have

$\frac{Ps_j^i}{2^{(L_j-l+1)}} < \alpha_{\mathrm{j},l}Ps_j^i \leq \frac{Ps_j^i}{2^{(L_j-l)}}$

Bibliography

[1] A. Chaudhry and R. Hafez, "LMR and LTE for Public Safety in 700 MHz Spectrum," *Wireless Communications and Mobile Computing*, vol. 2019, 06 2019.

[2] Q. Qiu, S. Liu, S. Xu, and S. Yu, "Study on Security and Privacy in 5G-Enabled Applications," *Wireless Communications and Mobile Computing*, 12 2020.

[3] J. Li, X. Lin, K. K. Nagalapur, Z. Qi, A. Lahuerta-Lavieja, T. Chapman, S. Agneessens, H. Sahlin, D. Guldbrand, and J. Åkesson, "Toward Providing Connectivity When and Where It Counts: An Overview of Deployable 5G Networks," *IEEE Communications Standards Magazine*, vol. 6, no. 4, pp. 56–64, 2022.

[4] R. Favraud, A. Apostolaras, N. Nikaein, and T. Korakis, "Toward moving public safety networks," *IEEE Communications Magazine*, vol. 54, no. 3, pp. 14–20, 2016.

[5] Gitnux, "Internet Traffic Statistics And Trends in 2023," March 24 2023. Accessed on April 5, 2023.

[6] C. Sudhamani, M. Roslee, J. J. Tiang, and A. U. Rehman, "A Survey on 5G Coverage Improvement Techniques: Issues and Future Challenges," *Sensors*, vol. 23, no. 4, 2023.

[7] G. Baldini, S. Karanasios, D. Allen, and F. Vergari, "Survey of Wireless Communication Technologies for Public Safety," *IEEE Communications Surveys Tutorials*, vol. 16, no. 2, pp. 619–641, 2014.

[8] A. Jarwan, A. Sabbah, M. Ibnkahla, and O. Issa, "LTE-Based Public Safety Networks: A Survey," *IEEE Communications Surveys Tutorials*, vol. 21, no. 2, pp. 1165–1187, 2019.

[9] M. M. Sohul, M. Yao, X. Ma, E. Y. Imana, V. Marojevic, and J. H. Reed, "Next generation public safety networks: A spectrum sharing approach," *IEEE Communications Magazine*, vol. 54, no. 3, pp. 30–36, 2016.

[10] 3GPP, "Universal Mobile Telecommunications System (UMTS); LTE; Proximity-based services (ProSe)," Technical Report (TR) 23.303, 3rd Generation Partnership Project (3GPP), 09 2014. Version 12.2.0.

[11] 3GPP, "Universal Mobile Telecommunications System (UMTS); LTE; Proximity-based services (ProSe)," Technical Report (TR) 23.303, 3rd Generation Partnership Project (3GPP), 07 2016. Version 13.4.0.

[12] 3GPP, "5G; Proximity based Services (ProSe) in the 5G System (5GS)," Technical Specification (TS) 23.304, 3rd Generation Partnership Project (3GPP), 07 2022. Version 17.3.0.

[13] K. S and D. Vali, "A Survey on Evolution of Mobile Communication from 1G to 7G," in *3rd National Conference on Image Processing, Computing, Communication, Networking and Data Analytics*, pp. 391–397, 06 2018.

[14] N. Bhandari, S. Devra, and K. Singh, "Evolution of Cellular Network: From 1G to 5G," in *International Journal of Engineering and Techniques*, vol. 3, pp. 98–105, 2017.

[15] A. F. M. Shahen Shah, "A Survey From 1G to 5G Including the Advent of 6G: Architectures, Multiple Access Techniques, and Emerging Technologies," in *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 1117–1123, 2022.

[16] N. A. Mohammed, A. M. Mansoor, and R. B. Ahmad, "Mission-Critical Machine-Type Communication: An Overview and Perspectives Towards 5G," *IEEE Access*, vol. 7, pp. 127198–127216, 2019.

[17] U. Singh, A. Dua, S. Tanwar, N. Kumar, and M. Alazab, "A Survey on LTE/LTE-A Radio Resource Allocation Techniques for Machine-to-Machine Communication for B5G Networks," *IEEE Access*, vol. 9, pp. 107976–107997, 2021.

[18] D. Salunkhe, P. Kende, H. Takale, and J. Joshi, "A Review On Architecture Technology and Challenges of 4G LTE," 03 2016.

[19] B. Raghothaman, E. Deng, R. Pragada, G. Sternberg, T. Deng, and K. Vanganuru, "Architecture and protocols for LTE-based device to device communication," *2013 International Conference on Computing, Networking and Communications (ICNC)*, pp. 895–899, 2013.

[20] M. Wang and Z. Yan, "A Survey on Security in D2D Communications," *Mobile Netw Appl 22*, pp. 195–208, 2017.

[21] P. Mach, Z. Becvar, and T. Vanek, "In-Band Device-to-Device Communication in OFDMA Cellular Networks: A Survey and Challenges," *IEEE Communications Surveys Tutorials*, vol. 17, no. 4, pp. 1885–1922, 2015.

[22] R. I. Ansari, C. Chrysostomou, S. A. Hassan, M. Guizani, S. Mumtaz, J. Rodriguez, and J. J. P. C. Rodrigues, "5G D2D Networks: Techniques, Challenges, and Future Prospects," *IEEE Systems Journal*, vol. 12, no. 4, pp. 3970–3984, 2018.

[23] U. N. Kar and D. K. Sanyal, "An overview of device-to-device communication in cellular networks," *ICT Express*, vol. 4, pp. 203–208, 2018.

[24] A. Arash, J. Peter, and M. Vincenzo, "Modeling Multi-mode D2D Communications in LTE," *arXiv.org, Computer Science, Networking and Internet Architecture*, 2014.

[25] S. B. Biradar and G. G. Hallur, "Economic implication of spectrum bands used in 5g: A multicountry study of spectrum allocation," in *2022 International Conference on Decision Aid Sciences and Applications (DASA)*, pp. 584–590, 2022.

[26] D. Mourtzis, J. Angelopoulos, and N. Panopoulos, "Smart manufacturing and tactile internet based on 5g in industry 4.0: Challenges, applications and new trends," *Electronics*, vol. 10, 12 2021.

[27] Q. Qiu, S. Liu, S. Xu, and S. Yu, "Study on Security and Privacy in 5G-Enabled Applications," *Wireless Communications and Mobile Computing*, vol. 2020, pp. 1–15, 12 2020.

[28] J. M. G. Aranda, E. J. S. Cabrera, D. Haro-Mendoza, and F. A. Salinas, "5G networks: A review from the perspectives of architecture, business models, cybersecurity, and research developments," *Novasinergia*, vol. 4, no. 1, pp. 6–41, Jun. 2021.

[29] C. Leite, P. S. Barreto, M. F. Caetano, and R. Amaral, "A Framework for Performance Evaluation of Network Function Virtualisation in 5G Networks," *2020 XLVI Latin American Computing Conference (CLEI)*, pp. 314–321, 2020.

[30] 3GPP, "5G; System Architecture for the 5G System," Technical Specification (TS) 23.501, 3rd Generation Partnership Project (3GPP), 09 2018. Version 15.3.0.

[31] 3GPP, "5G; System architecture for the 5G System (5GS) ," Technical Specification (TS) 23.501, 3rd Generation Partnership Project (3GPP), 10 2020. Version 16.6.0.

[32] 3GPP, "5G; System Architecture for the 5G System (5GS)," Technical Specification (TS) 23.501, 3rd Generation Partnership Project (3GPP), 07 2022. Version 17.5.0.

[33] 3GPP, "5G; 5G System Enhancements for Edge Computing; Stage 2," Technical Specification (TS) 23.548, 3rd Generation Partnership Project (3GPP), 05 2022. Version 17.2.0.

[34] A. Virdis, G. Nardini, G. Stea, and D. Sabella, "End-to-End Performance Evaluation of MEC Deployments in 5G Scenarios," *J. Sens. Actuator Networks*, vol. 9, p. 57, 2020.

[35] *ETSI White Paper, "MEC in 5G networks"*, June 2018.

[36] *ETSI GS MEC 009 V2.1.1, "Multi-access Edge Computing (MEC); General principles for MEC Service APIs"*, (2019-01).

[37] N. Alessandro, N. Giovanni, S. Giovanni, and V. Antonio, "Rapid prototyping and performance evaluation of ETSI MEC-based applications," *Simulation Modelling Practice and Theory*, vol. 123, p. 102700, 2023.

[38] A. Noferi, G. Nardini, G. Stea, and A. Virdis, "Deployment and configuration of MEC apps with Simu5G," *arXiv preprint arXiv:2109.12048*, 2021.

[39] N. A. Mohammed, A. M. Mansoor, and R. B. Ahmad, "Mission-Critical Machine-Type Communication: An Overview and Perspectives Towards 5G," *IEEE Access*, vol. 7, pp. 127198–127216, 2019.

[40] S. Chen, "A novel td-lte frame structure for heavy uplink traffic in smart grid," in *2014 IEEE Innovative Smart Grid Technologies - Asia (ISGT ASIA)*, pp. 158–163, 2014.

[41] R.-J. Wang, C.-H. Wang, G.-S. Lee, D.-N. Yang, W.-T. Chen, and J.-P. Sheu, "Resource Allocation in 5G with NOMA-based Mixed Numerology Systems," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, pp. 1–6, 2020.

[42] A. A. Zaidi, R. Baldemair, V. Moles-Cases, N. He, K. Werner, and A. Cedergren, "Ofdm numerology design for 5g new radio to support iot, embb, and mbsfn," *IEEE Communications Standards Magazine*, vol. 2, no. 2, pp. 78–83, 2018.

[43] P. K. Korrai, E. Lagunas, A. Bandi, S. K. Sharma, and S. Chatzinotas, "Joint Power and Resource Block Allocation for Mixed-Numerology-Based 5G Downlink Under Imperfect CSI," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 1583–1601, 2020.

[44] S. Vanka, S. Srinivasa, Z. Gong, P. Vizi, K. Stamatiou, and M. Haenggi, "Superposition Coding Strategies: Design and Experimental Evaluation," *IEEE Transactions on Wireless Communications*, vol. 11, no. 7, pp. 2628–2639, 2012.

[45] J. Andrews, "Interference cancellation for cellular systems: a contemporary overview," *IEEE Wireless Communications*, vol. 12, no. 2, pp. 19–29, 2005.

[46] M. S. Ali, H. Tabassum, and E. Hossain, "Dynamic User Clustering and Power Allocation for Uplink and Downlink Non-Orthogonal Multiple Access (NOMA) Systems," *IEEE Access*, vol. 4, pp. 6325–6343, 2016.

[47] J. Li, T. Gao, B. He, W. Zheng, and F. Lin, "Power Allocation and User Grouping for NOMA Downlink Systems," *Applied Sciences*, vol. 13, no. 4, 2023.

[48] S. M.-B. Fakhrildin, A. G. Wadday, and A. H. Al-Nakkash, "A Comparative Study of Perfect and Imperfect SIC in Downlink PD-NOMA based on Sharing Bandwidth," *IOP Conference Series: Materials Science and Engineering*, vol. 1105, 2021.

[49] J. Li, T. Gao, B. He, W. Zheng, and F. Lin, "Power Allocation and User Grouping for NOMA Downlink Systems," *Applied Sciences*, vol. 13, p. 2452, 02 2023.

[50] S. Moussa, A. Benslimane, R. Darazi, and C. Jiang, "Resource allocation for Public Safety Users in the 5G Cellular Network," in *2021 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, 2021.

[51] X. Fu, M. Pischella, and D. Le Ruyet, "On gaussian approximation algorithms for scma," in *2019 16th International Symposium on Wireless Communication Systems (ISWCS)*, pp. 155–160, 2019.

[52] H. Nikopour and H. Baligh, "Sparse code multiple access," in *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 332–336, 2013.

[53] S. Chen, B. Ren, Q. Gao, S. Kang, S. Sun, and K. Niu, "Pattern Division Multiple Access—A Novel Nonorthogonal Multiple Access for Fifth-Generation Radio Networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 4, pp. 3185–3196, 2017.

[54] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal Multiple Access for 5G and Beyond," *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2347–2381, 2017.

[55] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-Orthogonal Multiple Access (NOMA) for Cellular Future Radio Access," *2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*, pp. 1–5, 2013.

[56] I. Chebbi, "Study and performance analysis of two-way cooperative NOMA networks," *hal*, 2021.

[57] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*. John Wiley & Sons, 2006.

[58] Y. Collette and P. Siarry, *Optimisation Multiobjectif*. Groupe Eyrolles, 2002.

[59] J. Li, T. Gao, B. He, W. Zheng, and F. Lin, "Power allocation and user grouping for noma downlink systems," *Applied Sciences*, vol. 13, no. 4, 2023.

[60] X. Zhang, J. Wang, J. Wang, and J. Song, "A novel user pairing in downlink non-orthogonal multiple access," in *2018 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, pp. 1–5, 2018.

[61] L. Li, Z. Feng, Y. Tang, Z. Peng, L. Wang, and W. S. and, "User clustering scheme for downlink of noma system," *KSII Transactions on Internet and Information Systems*, vol. 14, pp. 1363–1376, March 2020.

[62] A. Benjebbour, A. Li, Y. Kishiyama, H. Jiang, and T. Nakamura, "System-level performance of downlink noma combined with su-mimo for future lte enhancements," in *2014 IEEE Globecom Workshops (GC Wkshps)*, pp. 706–710, 2014.

[63] Z. J. Ali, N. K. Noordin, A. Sali, F. Hashim, and M. Balfaqih, "An efficient method for resource allocation and user pairing in downlink non-orthogonal multiple access system," in *2019 IEEE 14th Malaysia International Conference on Communication (MICC)*, pp. 124–129, 2019.

[64] O. Abuajwa, M. Roslee, and Z. Yusoff, "Simulated annealing for resource allocation in downlink noma systems in 5g networks," *Applied Sciences*, vol. 11, p. 4592, 05 2021.

[65] S. P. Kumaresan, C. K. Tan, C. K. Lee, and Y. H. Ng, "Low-complexity particle swarm optimisation-based adaptive user clustering for downlink non-orthogonal multiple access deployed for 5g systems," *World Review of Science, Technology and Sustainable Development*, vol. 18, no. 1, pp. 7–19, 2022.

[66] D. Pliatsios and P. Sarigiannidis, "Power Allocation in Downlink Non-orthogonal Multiple Access IoT-enabled Systems: A Particle Swarm Optimization Approach," in *2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, pp. 416–422, 2019.

[67] D. Pliatsios and P. Sarigiannidis, "Resource Allocation Combining Heuristic Matching and Particle Swarm Optimization Approaches: The Case of Downlink Non-Orthogonal Multiple Access," in *Information 2019*, vol. 10, 2019.

[68] S. P. Kumaresan, C. K. Tan, and Y. H. Ng, "Efficient user clustering using a low-complexity artificial neural network (ann) for 5g noma systems," *IEEE Access*, vol. 8, pp. 179307–179316, 2020.

[69] N. Otao, Y. Kishiyama, and K. Higuchi, "Performance of non-orthogonal access with sic in cellular downlink using proportional fair-based resource allocation," in *2012 International Symposium on Wireless Communication Systems (ISWCS)*, pp. 476–480, 2012.

[70] J. Li, H.-H. Chen, and Q. Guo, "On the performance of noma systems with different user grouping strategies," *IEEE Wireless Communications*, pp. 1–14, 2022.

[71] M.-R. Hojeij, C. A. Nour, J. Farah, and C. Douillard, "Advanced resource allocation technique for a fair downlink non-orthogonal multiple access system," in *2018 25th International Conference on Telecommunications (ICT)*, pp. 439–444, 2018.

[72] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, "System-level performance evaluation of downlink non-orthogonal multiple access (noma)," in *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 611–615, 2013.

[73] Y. Cui, P. Liu, Y. Zhou, and W. Duan, "Energy-efficient resource allocation for downlink non-orthogonal multiple access systems," *Applied Sciences*, vol. 12, no. 19, 2022.

[74] W. Saetan and S. Thipchaksurat, "Application of deep learning to energy-efficient power allocation scheme for 5g sc-noma system with imperfect sic," in *2019 16th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, pp. 661–664, 2019.

[75] S. Zhang, L. Li, J. Yin, W. Liang, X. Li, W. Chen, and Z. Han, "A dynamic power allocation scheme in power-domain noma using actor-critic reinforcement learning," in *2018 IEEE/CIC International Conference on Communications in China (ICCC)*, pp. 719–723, 2018.

[76] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95 - International Conference on Neural Networks*, vol. 4, pp. 1942–1948 vol.4, 1995.

[77] A. Masaracchia, D. B. Da Costa, T. Q. Duong, M.-N. Nguyen, and M. T. Nguyen, "A PSO-Based Approach for User-Pairing Schemes in NOMA Systems: Theory and Applications," *IEEE Access*, vol. 7, pp. 90550–90564, 2019.

[78] Y.-X. Guo and H. Li, "A power allocation method based on particle swarm algorithm for NOMA downlink networks," *Journal of Physics: Conference Series*, vol. 1087, p. 022033, 09 2018.

[79] H. Xiao, Y. Wang, Q. Cheng, and Y. Wang, "An Improved PSO-Based Power Allocation Algorithm for the Optimal EE and SE Tradeoff in Downlink NOMA Systems," in *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 1–5, 2018.

[80] M. ALAM, "Particle Swarm Optimization: Algorithm and its Codes in MATLAB," *ResearchGate*, 03 2016.

[81] B. Ghojogh, A. Ghodsi, F. Karray, and M. Crowley, " KKT Conditions, First-Order and Second-Order Optimization, and Distributed Optimization: Tutorial and Survey," *arXiv:2110.01858*, 10 2021.

[82] L.-C. Kung, "Operations Research (3): Theory," *Online course, fr.coursera.org/learn/operations-research-theory.*

[83] M. Yin, W. Li, L. Feng, P. Yu, and X. Qiu, "Emergency Communications Based on Throughput-Aware D2D Multicasting in 5G Public Safety Networks," *Sensors (Basel)*, vol. 20, no. 7, p. 1901, 2020.

[84] J. Zhao, Y. Liu, K. K. Chai, Y. Chen, M. Elkashlan, and J. Alonso-Zarate, "NOMA-Based D2D Communications: Towards 5G," in *2016 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, 2016.

[85] R. Chen, F. Shu, J. Wang, and L. Zhang, "User Clustering and Power Allocation for Energy Efficiency Maximization in Downlink Non-Orthogonal Multiple Access Systems," *Appl. Sci.*, vol. 11, no. 2, 2021.

[86] G. D. Swetha and G. R. Murthy, "Selective overlay mode operation for D2D communication in dense 5G cellular networks," in *2017 IEEE Symposium on Computers and Communications (ISCC)*, pp. 704–709, 2017.

[87] D. M. Soleymani, J. Mueckenheimi, M. Harounabadi, A. M. Waswa, Z. Shaik, and A. Mitschele-Thiel, "Implementation aspects of hierarchical radio resource management scheme for overlay D2D," in *2017 9th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, pp. 154–161, 2017.

[88] R. Jain, D.-M. Chiu, and W. Hawe, "A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems," *ACM Transaction on Computer Systems*, 1984.

[89] C. Liu and J. Zheng, "A QoS-Aware Resource Allocation Algorithm for Device-to-Device Communication Underlaying Cellular Networks," in *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, pp. 1–5, 2017.

[90] A. Alnoman and A. Anpalagan, "On D2D communications for public safety applications," in *2017 IEEE Canada International Humanitarian Technology Conference (IHTC)*, pp. 124–127, 2017.

[91] H. Al-Obiedollah, H. Bany Salameh, S. Abdel-Razeq, A. Hayajneh, K. Cumanan, and Y. Jararwehe, "Energy-efficient opportunistic multi-carrier NOMA-based resource allocation for beyond 5G (B5G) networks," *Simulation Modelling Practice and Theory*, vol. 116, p. 102452, 2022.

[92] A. Kilzi, J. Farah, C. A. Nour, and C. Douillard, "Inband Full-Duplex D2D Communications Underlaying Uplink Networks with Mutual SIC NOMA," in *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1–7, 2020.

[93] J. Zhao, Y. Liu, K. K. Chai, Y. Chen, and M. Elkashlan, "Joint Subchannel and Power Allocation for NOMA Enhanced D2D Communications," *IEEE Transactions on Communications*, vol. 65, no. 11, pp. 5081–5094, 2017.

[94] M. Le, Q.-V. Pham, H.-C. Kim, and W.-J. Hwang, "Enhanced Resource Allocation in D2D Communications With NOMA and Unlicensed Spectrum," *IEEE Systems Journal*, vol. 16, no. 2, pp. 2856–2866, 2022.

[95] A. Amer, A.-M. Ahmad, and S. Hoteit, "Resource Allocation for Downlink Full-Duplex Cooperative NOMA-Based Cellular System with Imperfect SI Cancellation and Underlaying D2D Communications," *Sensors*, vol. 21, 2021.

[96] Y. Pan, C. Pan, Z. Yang, and M. Chen, "Resource Allocation for D2D Communications Underlaying a NOMA-Based Cellular Network," *IEEE Wireless Communications Letters*, vol. 7, no. 1, pp. 130–133, 2018.

[97] ITU-R, "Propagation data and prediction methods for the planning of short-range outdoor radiocommunication systems and radio local area networks in the frequency range 300 MHz to 100 GHz," *ITU-R Recommendation P.1411-10*, 2019. Accessed: Mars 17, 2022.

[98] E. Driouch, W. Ajib, and C. Assi, "Power control and clustering in heterogeneous cellular networks," *Wireless Networks*, vol. 23, pp. 2509–2520, 2017.

[99] 3GPP, "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception," Technical Report (TR) 36.101, 3GPP, 11 2017. version 14.5.0.

[100] 3GPP, "Digital cellular telecommunications system (Phase 2+) (GSM); Universal Mobile Telecommunications System (UMTS); LTE; 5G; Release description," Technical Report (TR) 21.915, 3GPP, 10 2019. version 15.0.0.

[101] A. Abu Alkheir and H. T. Mouftah, "Cognitive radio for public safety communications," *Wireless Public Safety Networks 2. San Diego, CA, USA: Elsevier*, 2016.

[102] M. A. Imran, Y. A. Sambo, and Q. H. Abbasi, "Integrating Public Safety Networks to 5G: Applications and Standards," *Wiley-IEEE Press*, pp. 233–251, 2019.

[103] S.-Y. Lien, C.-C. Chien, F.-M. Tseng, and T.-C. Ho, "3GPP device-to-device communications for beyond 4G cellular networks," *IEEE Communications Magazine*, vol. 54, no. 3, pp. 29–35, 2016.

[104] I. P. Chochliouros, A. S. Spiliopoulou, P. l. Lazaridis, Z. D. Zaharis, M.-R. Spada, J. Pérez-Romero, B. Blanco, H. Khalife, E. E. Khaleghi, and M.-A. Kourtis, "5G for the Support of Public Safety Services," *Wireless Pers Commun 120*, pp. 2321–2348, 2021.

[105] R. Solozabal, A. Sanchoyerto, E. Atxutegi, B. Blanco, J. O. Fajardo, and F. Liberal, "Exploitation of Mobile Edge Computing in 5G Distributed Mission-Critical Push-to-Talk Service Deployment," *IEEE Access*, vol. 6, pp. 37665–37675, 2018.

[106] M. Höyhtyä, K. Lähetkangas, J. Suomalainen, M. Hoppari, K. Kujanpää, K. Trung Ngo, T. Kippola, M. Heikkilä, H. Posti, J. Mäki, T. Savunen, A. Hulkkonen, and H. Kokkinen, "Critical Communications Over Mobile Operators' Networks: 5G Use Cases Enabled by Licensed Spectrum Sharing, Network Slicing and QoS Control," *IEEE Access*, vol. 6, pp. 73572–73582, 2018.

[107] G. Nardini, D. Sabella, G. Stea, P. Thakkar, and A. Virdis, "Simu5G–An OMNeT++ Library for End-to-End Performance Evaluation of 5G Networks," *IEEE Access*, vol. 8, pp. 181176–181191, 2020.

[108] G. Nardini, G. Stea, A. Virdis, and D. Sabella, "Simu5G: A System-level Simulator for 5G Networks," *10th International Conference on Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH 2020)*, pp. 68–80, 07 2020.

[109] B. SKLAR, *Digital Communications: Fundamentals and Applications (2nd ed.)*. Prentice Hall., Jan. 2001.