KADİR HAS UNIVERSITY
SCHOOL OF GRADUATE STUDIES
DEPARTMENT OF ENGINEERING AND NATURAL SCIENCES

NOMA-BASED RADIO RESOURCE ALLOCATION FOR MACHINE TYPE COMMUNICATION IN 5G AND BEYOND CELLULAR NETWORK

SÜMEYRA ALDEMİR

ASST. PROF. YALÇIN ŞADİ

MASTER OF SCIENCE THESIS

ISTANBUL, OCTOBER, 2021

Sümeyra Aldemir | Master of Science Thesis | 2021

NOMA-BASED RADIO RESOURCE ALLOCATION FOR MACHINE TYPE COMMUNICATION IN 5G AND BEYOND CELLULAR NETWORK

BY

SÜMEYRA ALDEMİR

A THESIS SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES OF KADIR HAS UNIVERSITY IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN ELECTRONICS ENGINEERING

ISTANBUL, OCTOBER, 2021

DECLARATION OF RESEARCH ETHICS /
METHODS OF DISSEMINATION

I, SÜMEYRA ALDEMİR, hereby declare that;

- this master's thesis is my own original work and that due references have been appropriately provided on all supporting literature and resources;
- this master's thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- I have followed *Kadir Has University Academic Ethics Principles prepared in accordance with The Council of Higher Education's Ethical Conduct Principles.*

In addition, I understand that any false claim in respect of this work will result in disciplinary action in accordance with University regulations.

Furthermore, both printed and electronic copies of my work will be kept in Kadir Has Information Center under the following condition as indicated below: The full content of my thesis will be accessible from everywhere by all means.

SÜMEYRA ALDEMİR
07.OCTOBER.2021

KADİR HAS UNIVERSITY

SCHOOL OF GRADUATE STUDIES

ACCEPTANCE AND APPROVAL

This work entitled NOMA-based Radio Resource Allocation for Machine Type Communications in 5G and Beyond Cellular Networks prepared by SÜMEYRA ALDEMİR has been judged to be successful at the defense exam on 07.OCTOBER.2021 and accepted by our jury as master's thesis.

APPROVED BY:

Asst. Prof. Dr. Yalçın Şadi (Advisor) ....
Kadir Has University

Assoc. Prof. Dr. Eylem Erdoğan ....
Istanbul Medeniyet University

Assoc. Prof. Dr. Atilla Özmen ....
Kadir Has University

I certify that the above signatures belong to the faculty members named above.

....
Prof. Dr. M. Timur Aydemir
Dean of School of Graduate Studies
DATE OF APPROVAL: 07.OCTOBER.2021

TABLE OF CONTENTS

NOMA-based Radio Resource Allocation for Machine Type Communications in 5G and Beyond Cellular Networks

ABSTRACT

The rapid increase of machine-to-machine (M2M) communications brings challenges in the design of cellular networks. The adversity of meeting the Quality-of-Service (QoS) requirements of a massive number of machine-type communications (MTC) devices with insufficient radio resources has emerged. The main goal of this thesis is to investigate the minimum bandwidth resource allocation problem for non-orthogonal multiple access (NOMA) based M2M communications in 5G and beyond cellular networks. A polynomial-time persistent resource allocation algorithm considering NOMA and the periodicity of the MTC traffic is proposed to solve the problem fast and efficiently. The proposed algorithm consists of two phases. In the first phase, M2M clusters are grouped into NOMA sub-clusters using a technique that minimizes the number of NOMA sub-clusters for a set of devices. NOMA sub-clusters are then allocated to resource blocks (RB) in the second phase, considering their QoS requirements while achieving minimum bandwidth reservation. Through simulations, the performance of the proposed algorithm is presented in comparison to the previously proposed access grant time interval (AGTI) based radio resource allocation algorithms. It is illustrated that the proposed algorithm improves the spectrum efficiency significantly addressing the spectrum-scarcity issue.

Keywords: Machine-to-Machine (M2M) Communications, Non-orthogonal Multiple Access (NOMA), 5G and Beyond Cellular Networks, Resource Allocation.

5G ve Ötesi Hücresel Ağlarda Makine Tipi İletişim için NOMA Tabanlı Radyo Kaynak Atama

ÖZET

Makineden makineye iletişimdeki hızlı artış, hücresel ağların tasarımında zorluklar getiriyor. Yetersiz radyo kaynaklarına sahip çok sayıda makine tipi iletişim cihazının hizmet kalitesi gereksinimlerini karşılamanın zorluğu ortaya çıktı. Bu tezin temel amacı, 5G ve hücresel ağların ötesinde, dikgen olmayan çoklu erişim tabanlı makineler arası iletişimleri için minimum bant genişliği kaynak tahsisi problemini araştırmaktır. Problemi hızlı ve verimli bir şekilde çözmek için NOMA ve MTC trafiğinin periyodikliğini dikkate alan bir polinom zamanlı kalıcı kaynak tahsis algoritması önerilmiştir. Önerilen algoritma, iki aşamadan oluşmaktadır. İlk aşamada, M2M kümeleri, bir dizi cihaz için NOMA alt kümelerinin sayısını en aza indiren bir teknik kullanılarak NOMA alt kümelerine ayrılmaktadır. İkinci aşamada ise NOMA alt kümeleri, kaynak bloklarına servis kalitesi gereklilikleri saglanırken minimum bant genişliği kullanılacak şekilde atanmaktadır. Simülasyonlar ile, önerilen algoritmanın performansı, daha önce önerilen erişim izni zaman aralığına dayalı radyo kaynak atama algoritmalarıyla karşılaştırmalı olarak sunulmuştur. Önerilen algoritmanın spektrum etkinligini önemli ölçüde artırdığı gösterilmiştir.

Anahtar Sözcükler: Makineler Arası (M2M) İletişim, Dikgen Olmayan Çoklu Erişim (NOMA), 5G ve Ötesi Hücresel Ağlar, Kaynak Atama

ACKNOWLEDGEMENTS

I want to state my sincere appreciation to my advisor, Asst. Prof. Yalçın Şadi, for supporting me throughout my graduate studies, preparing this thesis and every decision I have made. I also want to express my gratitute to members of my thesis committee Assoc. Prof. Atilla Özmen and Assoc. Prof. Eylem Erdoğan, for their constructive comments and participation on my defense committee.

I gratefully acknowledge Kadir Has University, Department of Electrical - Electronics Engineering funding, and Asst. Prof. Yalçın Şadi's research grant from the TUBITAK 3501 Career Development Program (Career Program), "NOMA-based Radio Resource Allocation for Machine-Type Communications in 5G and Beyond Cellular Networks", Project No: 118E920 for granting me a scholarship during my graduate studies. I have served as a teaching assistant in the Department of Electrical - Electronics Engineering at Kadir Has University and a research assistant in Asst. Prof. Şadi's project during my master's degree.

More importantly, I thank to my parents - Afiyet & Refika Aldemir, my brothers - Ahmet & Abdullah Aldemir and my beloved cousin - Elif Seda Topuzoğlu. Thanks to their presence and endless support, I was able to finish this thesis. Next, I thank my best friend F. Batuhan Okumuş, who has shared my best moments since 2015. Lastly, I thank my office friend Gülsüm Yiğit for her help and friendship. Especially on these coronavirus days, I could not finish my graduate studies without their psychological support.

To my Parents

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS/ABBREVIATIONS

z_k	Indicator for the use of band k
y_{jk}	Assigning MTC Cluster j to band k
x_{ij}	Indication of if MTC device i is in MTC cluster j
C_i	Cluster i
C_{ik}	Sub-cluster k grouped from cluster i
p_j	Common Packet Generation Period of the Devices in the Sub-cluster j
d_j	Maximum Allowable Jitter
$\mathbf{B}_n$	Indication of how many NOMA sub-clusters from each MTC cluster are assigned to band n
$\mathbf{B}_n(i)$	Number of sub-cluster assigned from the set i in band n
K_i	The number of unallocated sub-cluster from cluster i
P	Transmit Power in Uplink Communication
P_Δ	Detection Threshold at SIC Receiver
N	Number of sub-bands
δ_i^*	Delay Bound
τ	Duration of a Resource Block
τ_i	Transmiison Time of Device i
Θ	The number of possible combinations for otimal user clustering
γ^*	The Highest Normalized Channel Gain
γ_i	Normalized Channel Gain of i
Δ_{rem}	The difference between the delay tolerance and and experienced delay due to higher priority sub-clusters
Δ_i	The additional delay the device experiences due to devices in higher priority sub-clusters

5G	Fifth Generation
M2M	Machine-to-Machine Communication
H2H	Human-to-Human
H2M	Human-to-Machine
GSM	Global System for Mobile Communication
RF	Radio Frequency
OFDM	Orthogonal Frequency-Division Multiple Access
TDM	Time Division Multiplexing
FDM	Frequency Division Multiplexing
OMA	Orthogonal Multiple Access
NOMA	Non-Orthogonal Multiple Access
MTD	Machine Type Devices
MTC	Machine Type Communication
QoS	Quality of Service
AGTI	Access Grant Time Interval
RA	Random Access
SIC	Successive Interference Cancellation
IP	Integer Programming
CDMA	Code-Division Multiple Access
IDMA	Interleave-Division Multiple Access
LDS-CDMA	Low Density Spreading aided CDMA
MUSA	MultiUser Shared Access
RSMA	Resource Spread Multiple Access
SCMA	Sparse Code Multiple Access
BS	Base Station
SNR	Signal-to-Noise Ratio
RB	Resource Block
NR	New Radio
FB	Frequency Band
TTI	Transmission Time Interval
RRS	Random Resource Schedule

CRS	Channel-aware Resource Schedule
CSI	Channel State Information
CA	Clustering Algorithm
DUC	Dynamic User Clustering

1. INTRODUCTION

1.1 Machine-to-Machine Communication

The machine-to-machine concept is based on the technology that machines communicate with each other without any human intervention in the communication cycle. Even though it seems like a new technology, M2M communication has been used for many years. M2M technologies are used most commonly in the vehicular telematics, security, and medical field. The communication system architecture consists of time-triggered and event-triggered devices, human-to-human (H2H) devices, as shown in Fig. 1.1. We have been using and editing tools for H2H or human-to-machine (H2M) communications. Yet, there is considerable potential for M2M devices in applications of smart homes, smart grids, smart cities, and the industrial environment without any human effort. M2M communications use sensors and RFIDs to collect data and transmit the collected data traditionally through Wi-fi, Ethernet, GSM, or Radio Frequency (RF). M2M servers and M2M devices in diverse network domains require a wide coverage field to enable communication. Therefore, usage of the cellular technologies will increase to gather data. Different types of multiple access technologies are used for M2M applications in cellular networks. For example, 4G systems are supported by Orthogonal Frequency-Division Multiple Access (OFDMA). The resources can be assigned to users by using either frequency division multiplexing (FDM) or time-division multiplexing (TDM) techniques for OFDMA. In FDM, the total bandwidth available is divided into frequency bands while TDM divides time frames into slots keeping all transmits over the same frequency band. However, Orthogonal Multiple Access (OMA) techniques are insufficient to keep up with the increase in the number of connected machines in MTC. Therefore, Non-Orthogonal Multiple Access (NOMA), where several machine type

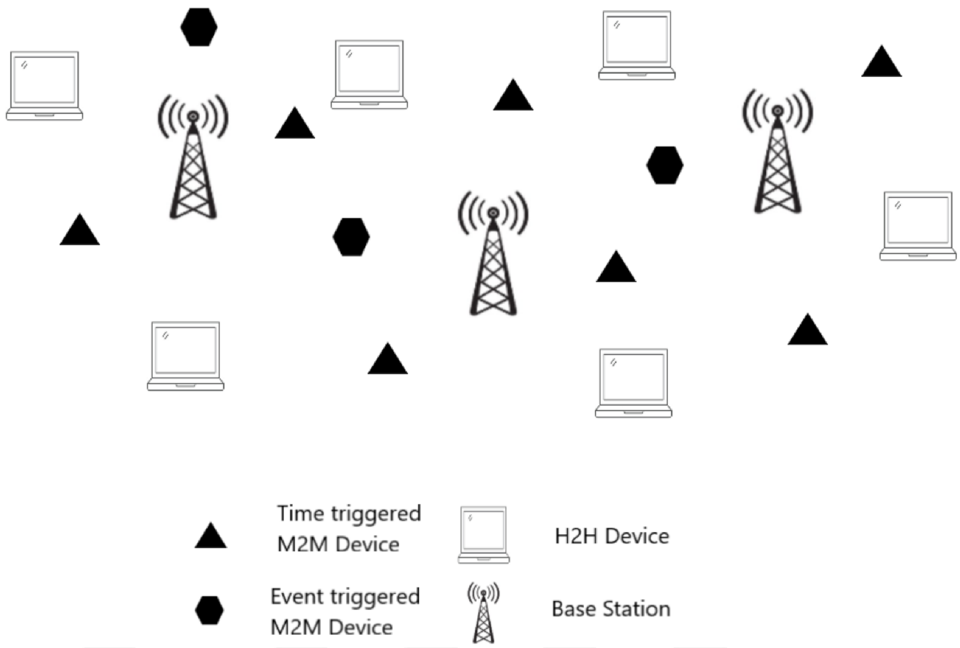

Figure 1.1 System architecture

devices (MTDs) are allowed to participate in the same orthogonal channel, is one of the solutions to support massive machine type connectivity in 5G. There are different service requirements for M2M applications that must be considered in design of access techniques. For example, some applications, such as security systems, are delay-sensitive but need to deliver small data, while applications such as intelligent grids are delay-tolerant but need to deliver large data. Therefore, the QoS requirements of M2M devices need to be met effectively to avoid radio resource wastage and service outage.

1.2 Non-Orthogonal Multiple Access (NOMA)

Radio access technologies (RATs) are effectively designed to increase system spectral capability and connectivity to meet the enhancing high data rate and connectivity requirements in next-generation 5G wireless networks. NOMA has attracted research attention in this area as an encouraging technique for 5G and beyond cellular networks. NOMA has the potential to improve system spectral capability and connectivity limit. When we compare NOMA with the current standard OMA techniques, NOMA introduces many advantages, such as massive connectivity, reduced

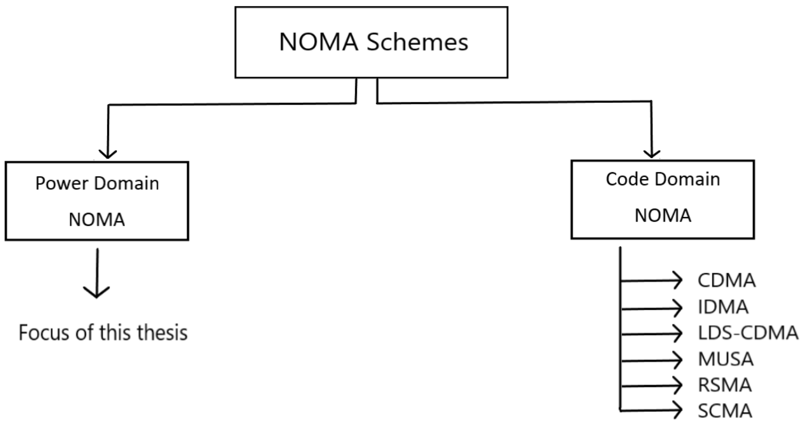

Figure 1.2 Classification of NOMA schemes

latency with high reliability, enhanced spectrum efficiency, etc. NOMA systems, using non-orthogonal resource allocation, accommodate more users than orthogonal multiple access (OMA). In future radio access systems, OMA will not be sufficient to meet the big resource demand.

The main idea behind NOMA is to serve multiple users over the same radio resource in terms of frequency, time, and space, simultaneously. While achieving this, it aims to minimize inter-user interference. NOMA serves individual users with higher adequate bandwidth, and at the same time, it permits scheduling more users than the available resources. NOMA can be categorized into two as power-domain and code-domain NOMA. Code-domain NOMA separates many users with the help of some generalized "codewords." Moreover, classic Code-Division Multiple Access (CDMA), covering Low Density Spreading aided CDMA (LDS-CDMA), Interleave-Division Multiple Access (IDMA), Resource Spread Multiple Access (RSMA), MultiUser Shared Access (MUSA), and Sparse Code Multiple Access (SCMA) constitute the code-domain. Since the focus of this thesis is on power-domain NOMA, we do not go further into these schemes. On the other hand, power-domain NOMA permits multiple users at the same frequency, time, and space but with distinct power grades, as summarized in Fig. 1.2. NOMA differs from OMA by superposing message signals from multiple users in the power domain with the benefit of individual channel gain differences. Then, successive interference cancellation (SIC) is applied for multi-user detection and decoding of the signals at the receiver side of NOMA.

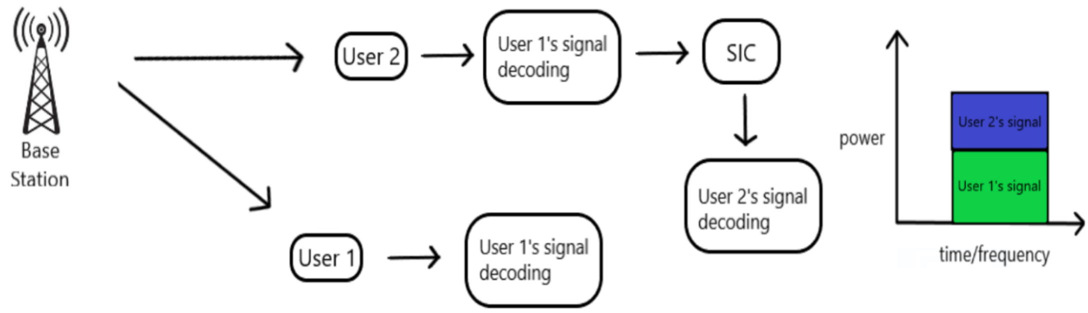

Figure 1.3 Downlink NOMA setup for 2 users

Decoding is performed per user while other users are assumed as noise.

NOMA can be further considered as downlink NOMA and uplink NOMA based on the direction of information transfer. In downlink NOMA, devices get signals over the same radio resources from the base station (BS). Downlink NOMA uses a power allocation technique where considerable transmission powers are used for weak channel conditions and lower transmit powers for users with better channel conditions. Therefore, considering a user in the NOMA cluster, most of the interference is caused by high-power message signals of weak-channel users. To obtain the desired signal, the strong interferences are canceled for each user by decoding, then remodulating and removing them from the received signal. All intra-cluster interferences are canceled by the user with the highest channel gain, while the user with the lowest channel gains faces interferences from all users within the cluster. In Fig. 1.3, there is an illustration of downlink NOMA for two users. The BS allocates diverse transmission powers for both users' signals. The signal with the most considerable transmit power, user 1, is decoded first at both users. The decoded signal utilizing SIC is removed by user 2, and it keeps decoding its signal. In uplink NOMA, devices send their message signals over the same radio resources to BS. In an uplink scenario, users transmit their individual signals with different transmit powers. The received signal at the BS is a superposed signal with a random white Gaussian noise added. The power of each device is limited to the lifetime of the battery. Unlike downlink NOMA, the battery powers of all users can be completely used on the condition that the users' channel gains are different enough from each other. Suppose that the channel gains are very close to each other. In that case, the

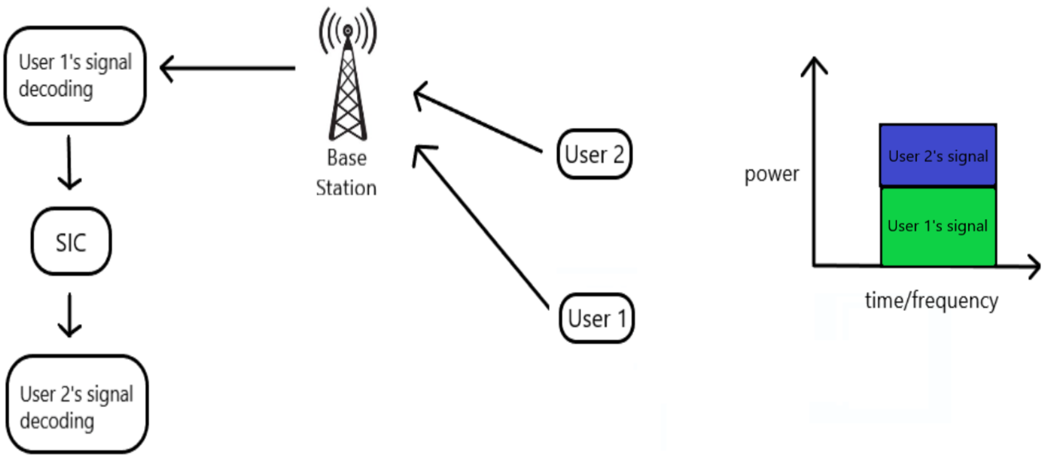

Figure 1.4 Uplink NOMA setup for 2 users

performance of the user with better channel gain can be improved by using power control while keeping the performance of the users with weaker channel gains at a specific level. In order to implement SIC in BS and decode the signals, it is essential to keep the divergence of the various superposed message signals within the received signal. Each message signal faces different channel gains because different users have different channels in an uplink scenario. So, the received signal power, which corresponds to the user with the strongest channel, is probably the strongest at the BS. Therefore, the strongest signal is firstly decoded at the base station and faces interference from all other users in the cluster. As a consequence, the user with the highest channel gain faces interference from all users in the NOMA cluster, where the user with the lowest channel gain experiences no interference from any user in the cluster. Fig.1.4 visualizes the basic functioning of the uplink NOMA setup for two users. Since the signal arriving with the largest receive power is user 1, it is decoded first. Then, the BS extracts its contribution utilizing SIC and decodes the second user's signal. Even though there are advantages of NOMA, it is still a challenging field to work including many different aspects such as dynamic user pairing, resource allocation, outage probability, and user fairness. User pairing is essential in order to obtain maximum efficiency due to co-channel interference in NOMA systems. Another challenging issue in NOMA is to manage proper communication in 5G with a high data rate, low delay, and secure way. For any wireless system, outage analysis is essential to understand the performance of the system. In

NOMA, less outage occurs for users randomly allocated in a cell compared to OMA. While NOMA pays attention to inter-cell interference, it should also pay attention to the outage examination to understand the outage pattern of edge users. Other challenges include the impact of transmission distortion, NOMA with multiple antennas, carrier aggregation, and NOMA with antenna selection. They all need to be investigated and improved to facilitate NOMA at high efficiency.

1.3 Related Work

The number and importance of machine-type communication devices in our lives have increased with the deployment of e-health systems, smart grid, systems used to monitor and protect assets and natural life, and smart houses, to name a few. Unfortunately, sufficient radio resources are not reserved for machine-type communication devices. H2H and H2M communications are given priority due to prevent performance degradation. When radio resources are reserved for M2M communications, H2H applications such as internet browsing, online games, video streaming, etc., experience resource scarcity, and as a result they experience excessive delays. Therefore, many M2M devices use restricted radio resources [1]. The network has to sparingly use the scarce radio resources to provide fair QoS because of the essential change in the number of users [2]. Machine-type devices are used for different purposes in various places having very diverse QoS requirements. Smoke detectors may have high latency, and lower throughput compared to a connected goggle providing augmented reality [3]. Also, while security systems have to transmit small data in a very limited time, systems such as smart meters can send larger data with a delay. These diverse QoS requirements of the systems should be taken into account while designing M2M communications. In [4, 5], the authors recommend clustering MTC devices with common QoS features without considering bandwidth efficiency. In [6, 7], effective bandwidth usage has been taken into account by taking advantage of the periodicity of M2M traffic, and efficient resource use is aimed at satisfying the QoS requirements of MTC devices. MTC devices provide access to the BS using the random access (RA) procedure. Even though MTC devices need to send a small data payload, the detection and use of RA procedure in M2M communication bring problems such as high latency, low transmission success rate, and high energy consumption. Certain methods have been proposed to overcome these problems [10]. Innovative methods are presented in [12] and [13] for MTDs to communicate without the need for the RA procedure. However, we do not focus on the RA procedure. Some stages of the RA procedure were assumed to be managed in advance. NOMA has been considered as a strategy that aids with efficient utilization of spectrum resources for 5G and beyond 5G [14]. Analysis of NOMA and how to meet requirements for 5G is discussed in [15, 16]. The main approach of NOMA is to assist all users with the same radio resources at the same time. NOMA technique consists of power domain and code domain. In code domain NOMA [17–19], the users at the transmitter side assign different codes, and the receiver separates the signals by looking at the codes. On the other hand, in power domain NOMA [24–27], the message signals of multiple users are superposed with various power levels. As a result, the receivers implement successive interference cancellations (SIC) to decode the received signals. The fundamentals of NOMA was exploited in [14] for downlink transmissions, and uplink transmissions [20]. Although NOMA allows serving multiple MTC devices simultaneously, it is still challenging to deal with massive connectivity and ensure spectrum efficiency. Therefore, different user pairing and user clustering methods are investigated as a solution to massive connectivity problems [26–31]. The massive growth of M2M traffic and the number of connected devices in cellular networks naturally bring some challenges in supporting this traffic. The challenge of meeting the traffic and QoS requirements of many MTC devices with limited radio resources have pushed researchers to look for a solution in 5G and beyond networks. As a candidate, the periodic characteristics of most M2M traffic can be used with a persistent resource allocation scheme in which radio resources are allocated periodically without any extra control signaling for long durations. NOMA is one of the effective methods used to increase the spectral efficiency of mobile communication networks for a vast number of MTC devices. In light of this idea, a NOMA-based resource allocation algorithm for uplink is proposed to minimize the bandwidth used while meeting the periodic traffic and quality of service requirements of M2M devices.

1.4 Original Contributions

It is hard to meet the QoS requirements of a massive number of MTC devices with insufficient radio resources in cellular networks. Therefore, this thesis aims to determine the user clustering and resource allocation algorithms for uplink NOMA systems that reduce the usage of frequency bands while meeting the requirements of the packet generation period and the maximum allowed jitter requirements of MTC devices. Thesis contributions can be listed as follows:

1. We examine the fundamental principles of uplink and downlink NOMA.
2. We propose a novel NOMA clustering technique that maximizes the number of M2M devices in a cluster. The proposed scheme utilizes the channel gain differences among the users in a NOMA cluster.
3. We propose a polynomial-time persistent resource allocation algorithm considering NOMA and the periodicity of the machine type communications traffic.
4. Through simulations, the performance of the proposed algorithm is presented in comparison to the previously proposed access grant time interval-based radio resource allocation algorithms. It is shown that the proposed algorithm enhances the spectrum efficiency significantly.

1.5 Thesis Outline

The remainder of the thesis is organized as follows. Chapter 2 explains the system model and the assumptions used throughout the thesis. Then, in Chapter 3, the minimum bandwidth resource allocation problem is formulated. Next, Chapter 4 presents the NOMA clustering and resource allocation algorithms. Simulations are presented in Chapter 5. Finally, concluding remarks are given in Chapter 6.

2. SYSTEM MODEL AND ASSUMPTIONS

System model and assumptions used throughout this thesis are detailed in this chapter.

2.1 NOMA

The requirement to access and analyze information from anywhere and at any time has led to the development of M2M communication in many ways. The main difference between M2M and H2H applications is that the number of M2M applications is much higher. Because of that, effective use of the bandwidth allocated for M2M is one of the main goals. However, there are different QoS requirements for MTC devices; most M2M applications need to transmit small data periodically. Using these features of MTC devices, we can perform spectral efficiency for massive connectivity by clustering M2M devices and allocate them to radio resources. NOMA system has been regarded as an assuring multiple access technology for the 5G wireless communication systems because it allows energy and spectral efficient communication while meeting diverse QoS requirements. NOMA is categorized into code-domain NOMA and power domain NOMA. Power domain NOMA supports multiple users at the same frequency, time, and space but with different power levels. In the power domain NOMA, the channel diversity can be effectively exploited over SIC. We consider K MTC devices randomly distributed within a circular area and communicating with a single BS at the center of that area using uplink power-domain NOMA. The BS and each device operate a single antenna configuration. A NOMA cluster is created with devices that are non-orthogonally allocated on the same resource block. On the other hand, each NOMA cluster is allocated on a frequency resource block that is orthogonal to other frequency resource blocks. Consider a 4 MTC device uplink

NOMA cluster with channel gains h_1, h_2, h_3 and h_4 where $h_1 > h_2 > h_3 > h_4$. The transmit power affects the decoding arrangement and the channels' fading effects for the uplink because channel coefficients to the BS are not identical at the same time. In the before-mentioned circumstances, the signal arriving with the enormous receive power is decoded first, the BS subtracts its additive utilizing SIC and continues decoding the residual signal. When the normalized channel gains of each device are γ_1, γ_2, γ_3 and γ_4 and P is the transmission power. Then the next states must be provided for efficient SIC at the base station. For example,

$$P\gamma_1 - P\gamma_2 - P\gamma_3 - P\gamma_4 \geq P_\Delta, \tag{2.1}$$

$$P\gamma_2 - P\gamma_3 - P\gamma_4 \geq P_\Delta, \tag{2.2}$$

$$P\gamma_3 - P\gamma_4 \geq P_\Delta, \tag{2.3}$$

Equations (2.1), (2.2), and (2.3) shows the essential conditions for effective decoding of the first, second, and third device, respectively, in order to prioritize the fourth device. According to the above conditions, the essential power constraints for effective SIC in a K MTC device uplink NOMA cluster can be stated as follows:

$$P\gamma_i - \sum_{j=i+1}^{K} P\gamma_j \geq P_\Delta, i = 1, 2, ..., (K-1). \tag{2.4}$$

2.2 QoS Characterization

QoS provisioning allows controlling traffic and providing sufficiently good performance for critical applications. It allows prioritization of high-performance applications and relieves the overall network traffic. MTC devices consist of event-triggered and time-triggered devices. There are different QoS characterizations for each case. Packet delay tolerance, aggregate maximum bit rate, required minimum bit rates, and acceptable packet loss rates can be considered as QoS requirements in event-triggered cases. In contrast, bandwidth efficiency, delay tolerance, packet loss, and jitter can be considered for time-triggered devices. This thesis considers the packet generation period (p_i) and the maximum allowed jitter value (d_i) to define QoS requirements of time-triggered devices while ensuring spectral efficiency. The time difference between two successful packet generation and transmission is defined as jitter. Jitter is considered as the major QoS metric because it completely expresses the time performance of periodic data transfer. Meeting jitter requirements is crucial for many MTC applications such as critical security processes, real-time control systems, health applications, and navigation data communications. Considering the QoS constraint of time-triggered MTC devices, the device delay must be provided within delay tolerance in the system. In another way, jitter requirements must be provided. Even in the worst delay case, the delay must be less than the suitable delay tolerance (jitter). If a device tries to transmit a packet during the transmission of higher priority devices simultaneously, the worst case may occur. Denote the number of time-triggered devices on the same band as N, the jitter bound as δ_i^* and transmission time of device i as τ_i. MTC devices are sorted in increasing order of packet generation periods. Then, the QoS constraint is formulated as follows:

$$\delta_i^* = \tau_i + \sum_{l=1}^{i-1} [p_i/p_l]\tau_l \leq d_i \quad for \quad i \in [1, N] \tag{2.5}$$

where p_i is the packet generation period of device i.

2.3 M2M Communication Scenario

We consider a cellular network architecture in which a base station supports multiple MTC devices with different QoS requirements. The cell edge users' lowest received signal-to-noise ratio (SNR) in cellular mobile communications generally limits the system performance. In OMA transmission, if the system design aims to maximize the system capability, those users located at the cell edge will seldom be scheduled. For this reason, an unfair resource allocation or even user failures occurs. Therefore, the power-based NOMA technique is used in the uplink transmission of MTC traffic, and the SNR of the arriving signals is taken into account when applying SIC at the base station. We consider K uniformly distributed devices within a circular area for uplink NOMA. γ_i responsible for distance-based path loss and shadowing expresses the normalized channel gain between device i and the BS. The normalized channel gains of the devices are sorted according to descending sequence of $\gamma_1 > \gamma_2 > \gamma_3 ... > \gamma_K$. In most M2M communication applications, time-triggered devices that generate periodical data are used. Industrial supply systems, e-health, and intelligent transportation systems can be given as an example of time-triggered devices. The solution presented in this thesis addresses NOMA-based radio resource allocation of time-triggered MTC devices in 5G and beyond the cellular network.

2.4 Clustering Techniques

The most fundamental design problems for uplink and downlink NOMA systems are well-designed user clustering and power allocation between users. As a solution to deal with massive connectivity problems, [26] uses a stochastic geometry procedure to create an intensive wireless system on both uplink and downlink NOMA. Performance results are investigated according to outage and achievable rate, either in pseudo-closed forms or short closed forms. Additionally, a massive grant-free NOMA scheme where devices have firm latency requirements and no retransmission opportunities exist has been proposed in [27]. Moreover, the effect of user grouping on the achievement of NOMA systems is examined in [28]. It is shown that the performance gain of NOMA compared to OMA can be additionally increased by choosing users whose channel requirements are more distinguishing. Different types of clustering techniques have been investigated in the literature. Specifically, authors in [29] introduce and investigate a hybrid OMA-NOMA designs for mMTC uplink scenarios by the scheduling designs random resource schedule (RRS) and channel-aware resource schedule (CRS). Under the RRS design, MTC devices are allocated to aggregation channels independently and randomly. At the same time, processing of RRS scheme, channel state information (CSI) is not used for the scheduling methods. Contrary to RRS, CSI is used for resource scheduling in CRS, and MTC devices are initially allocated from the one with better fading gain. Meanwhile, users are clustered singly if the user verifies NOMA relevance constraint with the interference from the user already existing in the cluster in [30]. Another clustering scheme, dynamic user clustering (DUC) [31], takes advantage of differences in channel gain between users in a NOMA cluster and groups them into an individual cluster or multiple clusters. In [31], it is tried to maximize the sum-throughput of the system without considering the spectral efficiency required for resource assignment for MTC.

This thesis proposes a low complexity user clustering scheme for uplink NOMA systems for machine-to-machine communications. The proposed scheme consists of two stages. First, time-triggered periodic devices with common period and jitter values are grouped into QoS clusters. "n" represents the number of devices per QoS cluster, varying from $2 < n < K$. Provided the range of n, the number of clusters can alter between 1 and K/2. Now, let specify the variable $x_{i,j}$ as follows:

$$x_{i,j} = \begin{cases} 1, & if\ a\ device\ i\ is\ grouped\ into\ cluster\ j \\ 0, & otherwise \end{cases} \tag{2.6}$$

where j=1,2, $\cdots$, K/2 and i=1,2, $\cdots$, K. Then, it exploits the channel gain differences among devices and creates NOMA sub-clusters from the QoS clusters. This scheme relies on 3 dB channel gain differences between each device before user clustering in sub-clusters. By this procedure, it aims to maximize the number of devices in a NOMA sub-cluster and thus to minimize the number of resource blocks required for allocation.

2.5 Resource Allocation

NOMA ensures greater flexibility in resource utilization than traditional OMA systems. A resource block is the smallest radio resource unit assigned to a user, as shown in Fig. 2.1. An RB is a time-frequency unit with a bandwidth of 180 kHz and a duration of 0.5 ms. In 5G cellular networks, the new radio (NR) is used to support massive connections by providing the flexibility referred to as *multiple numerologies* [32]. Subcarrier spacings and cyclic prefixes are used as a flexible frame

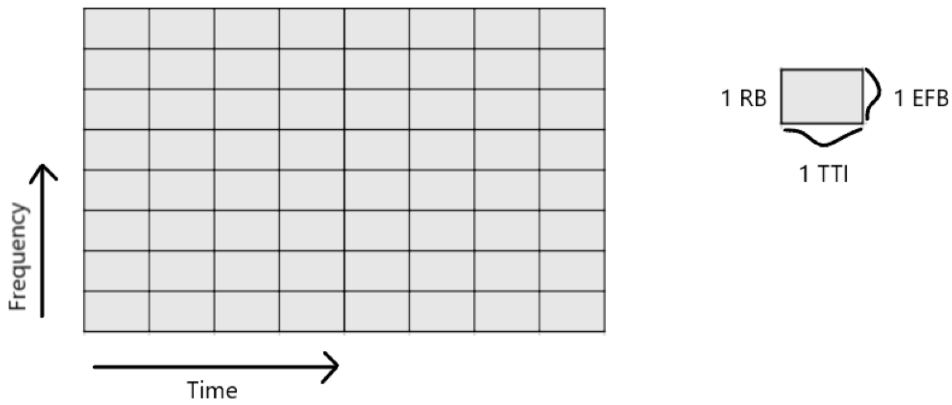

Figure 2.1 Structure of time-frequency resource blocks

structure. The space among the centers of two sequential subcarriers is called the subcarrier spacing, and the subcarrier spacing values are given 15, 30, 60, 120, and 240 kHz. In this thesis, a fixed numerology is used over the entire bandwidth part reserved for MTC communications. Resource block granularity is supported in NR as in LTE, although NR features enable to change RB architecture [33]. The frequency band of 1 RB width is called the element frequency band (EFB), and the time band of 1 RB width is called transmission time interval (TTI), as demonstrated in Fig. 2.1.

Each MTC device should be assigned to one EFB. Time-triggered M2M devices are prioritized; i.e., After time-triggered M2M devices are allocated to EFBs, then event-triggered M2M and H2H devices can be allocated if resources are present. For time-triggered MTC devices, only 1 RB is assigned periodically because of the small data transmission characteristic of MTC traffic. Packet generation periods are assumed to be integer multiples of 1 RB duration. The time-triggered M2M devices are prioritized and allocated in increasing order of packet generation periods.

3. MINIMUM BANDWIDTH RESOURCE ALLOCATION PROBLEM

A large number of M2M devices use limited radio resources that are constrained from H2H devices. Therefore, it is essential to minimize the number of frequency bands used by MTC devices considering fixed numerology. The minimum bandwidth resource allocation goal is to provide spectral efficiency while satisfying the packet generation period, maximum allowable jitter value, and SNR constraint of time-triggered M2M devices. MTC devices with a common packet generation period and maximum allowable jitter value are first grouped into QoS clusters. Then, these clusters are further clustered into NOMA sub-clusters by the proposed NOMA technique. The constraints of the optimization problem include periodic data generation and jitter tolerance requirements of MTC devices, and the SNR constraint for NOMA. The detailed information about QoS requirements is given in Section 2.3.

While meeting the requirements of periodic traffic and QoS requirements of a certain number of MTC devices, NOMA-based resource allocation, which aims to minimize the bandwidth usage, is formulated below as an optimization problem.

$$
\begin{aligned}
&min && \sum_k z_k && (3.1a)\\
&s.t. && \sum_k y_{jk} = 1,\ \forall j && (3.1b)\\
& && \sum_j y_{jk} \leq N z_k,\ \forall k && (3.1c)\\
& && \sum_j x_{ij} = 1,\ \forall i && (3.1d)\\
& && P\gamma_i x_{ij} - \sum_l x_{lj} P\gamma_l \geq P_\Delta,\ \forall i, j && (3.1e)\\
& && \tau + \sum_{l=1}^{j-1} \frac{p_j}{p_l} \tau y_{lk} \leq d_j + (1 - y_{jk})D,\ \forall j && (3.1f)\\
&variables && z_k \in \{0,1\},\ y_{jk} \in \{0,1\},\ x_{ij} \in \{0,1\} && (3.1g)
\end{aligned}
$$

The independent variables of the problem are z_k, y_{jk} and x_{ij} (3.1g). z_k is used to indicate whether band k is used; y_{jk} indicates whether NOMA sub-cluster j is assigned to band k, The variable x_{ij} indicates whether the MTC device i is clustered into the NOMA sub-cluster j. z_k, y_{jk} and x_{ij} are binary variables.

The objective function of the optimization problem is to minimize the frequency band usage. (3.1a). MTC devices with the same period and jitter values are divided into clusters, and then these clusters are ordered in ascending order. Each MTC cluster is then divided into sub-clusters using the NOMA technique. Each sub-cluster is assigned to an RB. It aims to reduce the number of frequency bands used as a result of assigning MTC devices of all sub-clusters to frequency bands so that traffic and QoS requirements are met. Equation (3.1b) states that a frequency band must be assigned for a NOMA sub-cluster j. Inequality (3.1c) indicates that at least one NOMA sub-cluster should be assigned to band k indicating that band k is used in resource allocation. N is a large number that will cause inequality when z_k takes a value of 1, regardless of the number of NOMA sub-clusters. Equation (3.1d) states that an MTC device can be assigned to only one NOMA sub-cluster. Inequality (3.1e) shows the SNR constraint required for a successful resolution at the base station for a NOMA sub-cluster j. Here, γ_i represents the normalized channel gain for MTC device i, P represents the power used in the uplink communication, and P_Δ represents the detection threshold at the SIC receiver. Finally, inequality (3.1f) ensures that the QoS requirements are met for NOMA sub-clusters ($y_{jk} = 1$) assigned to the same band. When $y_{jk} = 0$, the inequality for the value $D = \tau + \sum_{l=1}^{N} \frac{p_j}{p_l}\tau - d_j$ is satisfied in all conditions. Here, τ represents an RB duration; p_j is the common packet generation period of the devices in the sub-cluster j; d_j represents the maximum allowable jitter value.

4. NOMA-BASED MINIMUM BANDWIDTH RESOURCE ALLOCATION ALGORITHM

The formulated optimization problem is the previous section is and Integer Programming (IP) problem which is combinatorial in nature and difficult to solve in general. Mainly, for bandwidth minimization, the optimum user clustering solution requires exhaustive search to form a NOMA sub-cluster. One needs to consider all possible combinations of MTC devices to group them optimally. That is why the optimal resource allocation solution can not be reached practically for a large number of devices. For instance; in an uplink NOMA system with K users, the number of possible combinations for optimum user clustering can be stated as:

$$\Theta = \sum_{i=2}^{K} \binom{K}{i} \tag{4.1}$$

In this case, we propose a two-step solution for the problem. First, we group the MTC devices with a common packet generation period and maximum allowable jitter value into the same QoS cluster to decrease the overall complexity of resource allocation and then further group these clusters into NOMA sub-clusters using normalized channel gains of MTC devices. NOMA clustering solution is detailed in Sections 4.1 and 4.2. Then, given the NOMA clustering, we propose a resource allocation algorithm as detailed in Sections 4.3 and 4.4.

4.1 NOMA Clustering Algorithm

The NOMA clustering algorithm given in Algorithm 1 groups each of the M QoS clusters that consist of MTC devices with a common packet generation period and maximum jitter value into NOMA sub-clusters. Sub-clustering is done by using normalized channel gain values of MTC devices. This process continues until all

Algorithm 1 NOMA Clustering Algorithm

1: **input:** M clusters $\{C_1, C_2, ..., C_M\}$
2: **output:** NOMA sub-clusters $\{C_{i1}, C_{i2}, ..., C_{iK_i}\}$ for all C_i, $i \in [1, M]$;
3: **for** $i = 1 : M$ **do**
4:   $k \leftarrow 1$,
5:   **while** $C_i \neq \emptyset$ **do**
6:     $\gamma^* \leftarrow \gamma(C_i[1])$
7:     $C_{ik} \leftarrow C_{ik} + C_i[1]$
8:     $C_i \leftarrow C_i - C_i[1]$
9:     **for** $j \in C_i$ **do**
10:       **if** $\gamma_j \leq \gamma^*/2$ **then**
11:         $C_{ik} \leftarrow C_{ik} + j$
12:         $C_i \leftarrow C_i - j$
13:         $\gamma^* \leftarrow \gamma_j$
14:       **end if**
15:     **end for**
16:     $k \leftarrow k + 1$,
17:   **end while**
18: **end for**

Figure 4.1 NOMA Clustering Algorithm

QoS clusters are divided into NOMA sub-clusters, and each device is assigned to a sub-cluster.

The NOMA Clustering Algorithm is explained in detail as follows. The input is M QoS clusters, while the output is K_i sub-clusters of NOMA generated from each QoS cluster C_i (Lines 1 and 2). The algorithm creates sub-clusters from each set C_i for all $i \in [1, M]$ (Line 3). k indicates the number of sub-clusters created and is initially set to 1 (Line 4). The algorithm continues to generate sub-clusters for each cluster C_i as long as there are devices in C_i that have not yet been allocated to a NOMA sub-cluster(Line 5). The algorithm first selects the device with the highest normalized channel gain γ^* for each NOMA sub-cluster C_{ik} (Lines 6 and 7). Then, among the remaining devices in C_i, the first device whose normalized channel gain is equal to or less than half of the previously selected device's normalized channel gain is selected (Lines 9-15). This process ends when no device that meets this requirement is found. Devices added to the existing NOMA sub-cluster are removed from C_i. The algorithm terminates when all QoS clusters are divided into NOMA sub-clusters, and each device is assigned to a NOMA sub-cluster.

4.2 NOMA Clustering Algorithm Illustration Through An Example

The working principle of our clustering algorithm is explained through an example in Fig. 4.2. We consider one QoS cluster C_1 of 12 devices and the normalized channel gain of devices in the cluster are $\gamma_1 = 30$, $\gamma_2 = 29.5$, $\gamma_3 = 29$, $\gamma_4 = 28.5$, $\gamma_5 = 28$, $\gamma_6 = 27.5$, $\gamma_7 = 27$, $\gamma_8 = 26.5$, $\gamma_9 = 26$, $\gamma_{10} = 25.5$, $\gamma_{11} = 25$, and $\gamma_{12} = 24.5$, respectively. We start from device with the highest normalized channel gain $\gamma_1 = 30$ and compare with the second device $\gamma_2 = 29.5$. Since the difference between that two devices, γ_1 - γ_2 = 30 - 29.5 = 0.5, does not satisfy half the maximum power which is equal to the 3 dB, we move on to the next device, $\gamma_3 = 29$. Again, the normalized channel gain difference between the devices, γ_1 - γ_3 = 30 - 29 = 1, does not satisfy 3 dB condition. Similarly, we check normalized channel gain difference between first device and fourth, fifth, sixth devices, respectively. Results are γ_1 - γ_4 = 30 - 28.5 = 1.5, γ_1 - γ_5 = 30 - 28 = 2, γ_1 - γ_6 = 30 - 27.5 = 2.5. However, there is still no 3 dB difference. Thus, we can not create a sub-cluster with any of that devices. Then, we proceed to next device, $\gamma_7 = 27$. The normalized channel gain difference with the first device is γ_1 - γ_7 = 30 - 27 = 3. Since it satisfies the 3 dB difference condition, we can create a sub-cluster called C_{11} that contains first and seventh devices. This time, we continue to check the normalized channel gain difference with seventh device and eighth device. The result is γ_7 - γ_8 = 27 - 26.5 = 0.5. It does not satisfy the condition. Therefore, we move on the next device, $\gamma_9 = 26$. The normalized channel gain difference is γ_7 - γ_9 = 27 - 26 = 1. Since the difference is not 3 dB, we can not add ninth device to the sub-cluster C_{11}. For tenth, eleventh and twelfth devices, the normalized channel gain difference does not meet the condition. Therefore, the sub-cluster C_{11} includes only first and seventh devices. Then, we pick the device with the second highest normalized channel gain that is not in any sub-cluster, $\gamma_2 = 29.5$. We compare the normalized channel gain of second device with the normalized channel gain of other devices. There is no need to look for devices that are already in sub-cluster. Since the channel gain difference of second and eighth devices is γ_2 - γ_8 = 29.5 - 26.5 = 3, we create another sub-cluster called C_{12} that includes second and eighth devices. Unfortunately, the channel gain difference between eighth and next devices does not satisfy the condition. With this procedure, we allocate all devices to sub-clusters as shown in the figure in detail. The QoS cluster is grouped into 6 NOMA sub-clusters each containing 2 MTC devices satisying the SNR constraint given in Equation 3.1e.

4.3 Resource Allocation Algorithm

The proposed Resource Allocation Algorithm given in Algorithm 2 performs the assignment of RBs to K_i NOMA sub-clusters according to the priorities in order of increasing packet generation periods. The proposed algorithm assigns one RB to each NOMA sub-cluster periodically without violating the QoS requirements of the devices in NOMA sub-cluster. Resource Allocation Algorithm is explained in detail as follows.

K_i NOMA sub-clusters generated from each of the M QoS clusters obtained using NOMA Clustering Algorithm is sorted by ascending packet generation period. Each element of the vector $\mathbf{K}$, $\mathbf{K}_i$, specifies the number of NOMA sub-clusters of the MTC set i that have not yet been assigned an RB (Line 3). The algorithm updates the number of unassigned sub-clusters at each iteration and terminates when all sub-clusters are allocated RBs (Line 4). $\mathbf{B}_n$ is an M-dimensional vector showing how many NOMA sub-clusters from each MTC cluster are assigned to band n. Therefore, $\mathbf{B}_n(i)$ is the number of sub-clusters assigned from the cluster i in band n. For each band n, $\mathbf{B}_n$ is initialized to zero (Line 5). The algorithm assigns NOMA sub-clusters in descending order of priority in each band. For each MTC cluster i assigned to a band, Δ_i specifies the maximum waiting time due to higher priority MTC clusters

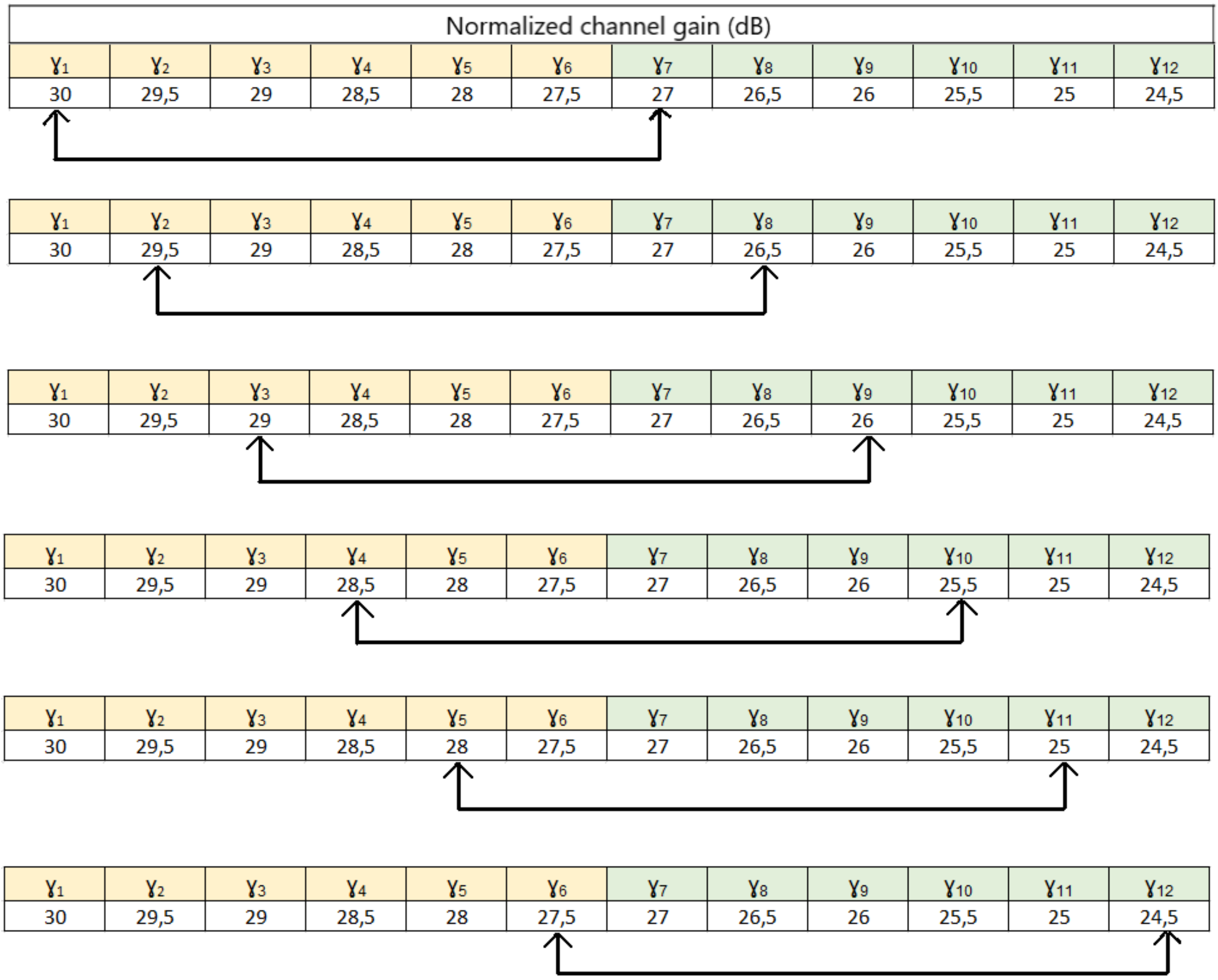

Figure 4.2 The illustration of NOMA clustering algorithm for uplink transmission of 12 devices

Algorithm 2 **Resource Allocation Algorithm**

input: RB length τ, p_i, d_i, K_i for all $i \in [1, M]$
output: Number of sub-bands N, for $n \in [1, N]$ resource allocation B_n;
$n \leftarrow 1, \mathbf{K} \leftarrow [K_1, K_2, ..., K_M]$;
while $\mathbf{K} \neq zeros(1, M)$ **do**
  $\mathbf{B}_n \leftarrow zeros(1, M)$;
  for $i = 1 : M$ **do**
    $\Delta_i \leftarrow 0$
    for $j = 1 : i$ **do**
      $\Delta_i \leftarrow \Delta_i + \mathbf{B}_n(j) * \lceil \frac{p_i}{p_j} \rceil * \tau$
    end for
    $\Delta_{rem} \leftarrow d_i - \Delta_i$
    if $\Delta_{rem} > 0$ **then**
      $\mathbf{B}_n(i) = \min\{K(i), \lfloor \frac{\Delta_{rem}}{\tau} \rfloor\}$
    end if
  end for
  $K \leftarrow K - B_n$
  $n \leftarrow n + 1$
end while
$N \leftarrow n - 1$

Figure 4.3 Resource Allocation Algorithm previously assigned to the same band (Lines 7-10). Each NOMA sub-cluster must be assigned to an RB of τ time before its next packet is created without violating the maximum allowable jitter requirement. Therefore, the appropriate assignment time Δ_{rem} for each MTC cluster is determined accordingly (Line 11). The number of NOMA sub-clusters to be assigned from the considered MTC set to the current band is determined to be $\lfloor \frac{\Delta_{rem}}{\tau} \rfloor$ at most, since each NOMA sub-cluster needs an RB of duration τ (Lines 12-14). The algorithm maximizes the utilization of each band by assigning the maximum number of NOMA sub-clusters from each QoS cluster. After NOMA sub-clusters have been assigned to band n without violating QoS requirements from all clusters, the number of remaining sub-clusters from each MTC cluster is updated (Line 16). The algorithm continues with the assignment of the remaining sub-clusters to the next band (Line 17).

4.4 Resource Allocation Algorithm Illustration Through An Example

The working principle of resource allocation algorithm is described through an example in Fig. 4.4. Consider 4 QoS clusters with $d_1 = 2$, $p_1 = 2$, $\tau_1 = 1$, $K_1 = 3$, $d_2 = 3$, $p_2 = 4$, $\tau_2 = 1$, $K_2 = 2$, $d_3 = 3$, $p_3 = 5$, $\tau_3 = 1$, $K_3 = 2$, $d_4 = 3$, $p_4 = 6$, $\tau_4 = 1$, $K_4 = 3$ as the cluster parameters. The clusters are denoted by C_1, C_2, C_3, C_4, respectively. The allocation of RBs to NOMA sub-clusters starts with cluster C_1. Since the highest priority cluster is C_1, there is no delay imposed by other clusters on C_1. Therefore, we set the Δ_i to 0. The Δ_{rem} is $d_1 - \Delta_1 = 2 - 0 = 2$. So, $\Delta_{rem}/\tau_1 = 2/1 = 2$ NOMA sub-clusters from cluster C_1 can be allocated to the first band $\mathbf{B}$. Here, $\mathbf{B}_1(1)$ can be updated to 2. Now, the cluster C_2 is considered for the first band $\mathbf{B}$. The delay applied on cluster C_2 by cluster C_1 on the initial band B is calculated as, $\mathbf{B}_1(1) * \lceil \frac{p_i}{p_j} \rceil * \tau_1 = 2 * \lceil \frac{4}{2} \rceil * 1 = 4$. However, since $4 > d_2 = 3$ the delay applied on cluster C_2 is greater than its delay tolerance. Therefore, any NOMA sub-cluster from QoS cluster C_2 cannot be allocated to the initial band $\mathbf{B}$. Similar to C_2, the delay applied from cluster C_1 on clusters C_3 and C_4 are $2 * \lceil \frac{5}{2} \rceil * 1 = 6$ and $2 * \lceil \frac{6}{2} \rceil * 1 = 6$, respectively. Thus, any NOMA sub-clusters from QoS clusters C_3 and C_4 cannot be allocated to the initial band. Next, the second band $\mathbf{B}$ is considered.

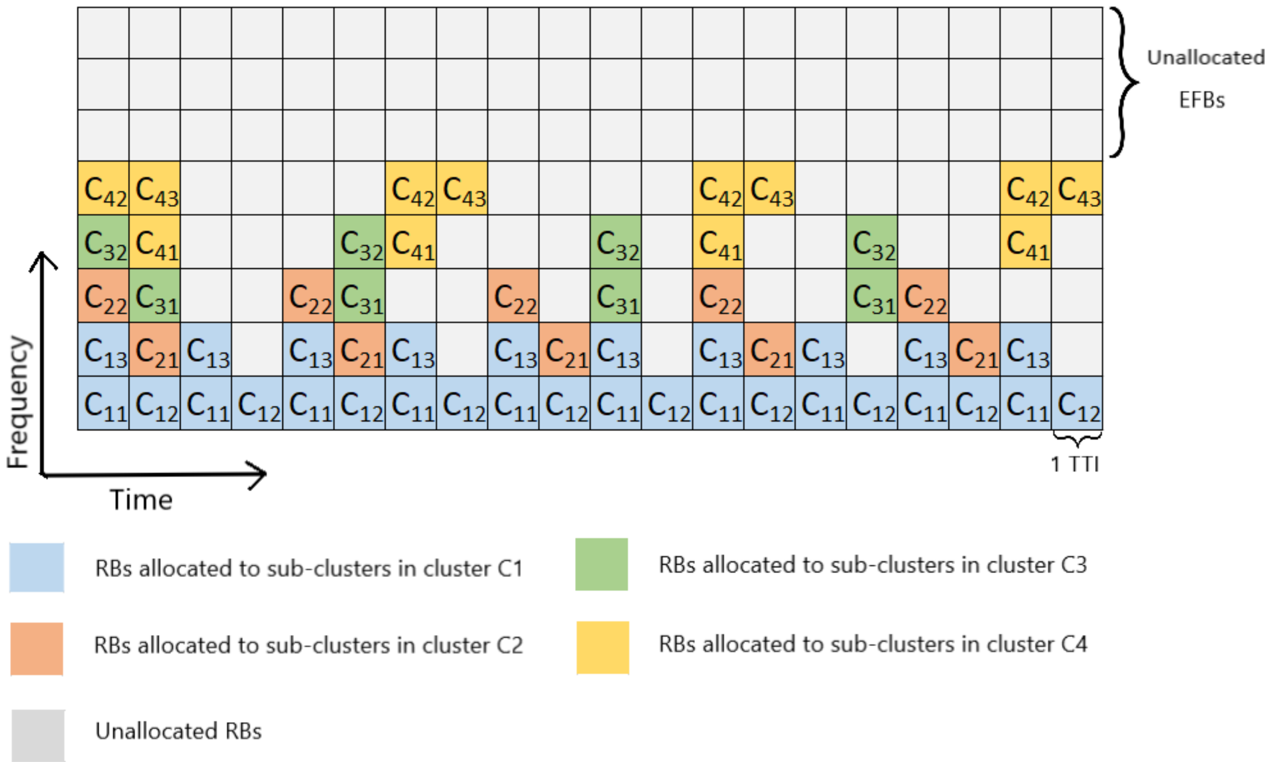

Figure 4.4 The illustration of resource allocation algorithm example

Here, the third NOMA sub-cluster from cluster C_1 is still unallocated. Therefore, first this sub-cluster is allocated on the second band **B**. Then, we proceed with the second highest priority cluster, C_2. The maximum wait time, Δ_2 is $\lceil \frac{3}{2} \rceil * 1 = 2$. The Δ_{rem} is $d_2 - \Delta_2 = 3 - 2 = 1$. According to the formula $\Delta_{rem}/\tau_2 = 1/1 = 1$ sub-cluster from cluster C_2 can be allocated to the second band. When the clusters C_3 and C_4 are considered, the sub-clusters on the second band would apply $\lceil \frac{5}{2} \rceil + \lceil \frac{5}{4} \rceil = 5$ and $\lceil \frac{6}{2} \rceil + \lceil \frac{6}{4} \rceil = 5$ delays, respectively. Since $5 > d_3 = 3$ and $d_4 = 3$, we cannot allocate any sub-clusters from cluster C_3 and cluster C_4. Following that, we move on to the third band. There is still one sub-cluster from cluster C_2. Therefore, we allocate this sub-cluster first on the third **B**. Then, we proceed with the next highest priority cluster, C_3. The maximum wait time, Δ_3 is $\lceil \frac{p_3}{p_2} \rceil * \tau_2 = \lceil \frac{5}{4} \rceil * 1 = 2$. The Δ_{rem} is $d_3 - \Delta_3 = 3 - 2 = 1$. So, we can allocate $\Delta_{rem}/\tau_3 = 1/1 = 1$ sub-cluster from cluster C_3 on the third **B**. The delay applied to cluster C_4 which is caused by cluster C_3 on the third band is, $\mathbf{B}_3(4) = \lceil \frac{6}{4} \rceil + \lceil \frac{6}{5} \rceil = 4 > d_4 = 3$. Then, we cannot allocate any sub-cluster from cluster C_4 on the third band. We allocate the remaining sub-clusters on the fourth and fifth bands following the same procedure.

5. SIMULATIONS & PERFORMANCE EVALUATION

This chapter aims at analyzing the performance of the proposed NOMA clustering algorithm and resource allocation algorithm to the previously proposed algorithms through simulations.

Table 5.1 summarizes the values of the parameters used in the simulations. It is assumed that the bandwidth reserved for M2M communication is 18 MHz, and it is divided into 100 bands, each corresponding to one RB-wide frequency band. Each QoS cluster consists of a random number of MTC devices distributed in the [10,100] range, and 12 clusters are used. Packet generation period and maximum jitter tolerance values of QoS clusters are $p_i \in \{10, 20, 20, 40, 100, 100, 200, 250, 500, 500, 10^3, 10^5\}(ms)$ and $\delta_i \in \{2, 4, 6, 12, 50, 60, 80, 100, 150, 200, 500, 10^4\}(ms)$, $\forall i \in \{1, ..., 12\}$, respectively. Performance results are obtained in MATLAB by averaging 1000 independent runs for each simulation scenario. In the simulations, the persistent resource allocation algorithm given in Algorithm 2 is used as the resource allocation algorithm.

In addition to the proposed Algorithm 1 as the NOMA clustering algorithm, different clustering algorithms suggested in the literature for comparison purposes random resource scheduling (RRS), channel-dependent resource scheduling (CRS) [29], dynamic user clustering (DUC) [31] are used. For the random resource schedule algorithm, the number of devices to be assigned and the available number of channels resources to be assigned are determined. If the number of channels is more than the number of MTC devices assigned, a random and independent assignment is performed as one device for each channel. However, if the number of MTC devices assigned is more than the number of channels, devices are randomly assigned to all

Parameters	Values
Number of MTC devices in each QoS cluster	[10,100]
Number of QoS clusters	12
Packet generation period, p_i	10,20,20,40,100,100,200, 250,500,500,10^3,10^5 (ms)
Maximum jitter tolerance, δ_i	2,4,6,12,50,60,80,100, 150,200,500,10^4 (ms)
Maximum NOMA sub-cluster size	4
Number of runs	1000

Table 5.1 Simulation Parameters

channels one by one first. The remaining devices are then assigned to channels to be used in common with previously assigned devices. This assignment is executed without using any channel state information. On the other hand, the channel-dependent resource schedule aims to use better channel resources by using the channel state information. The channel gains are ordered in descending sequence in CRS, where the aggregators know the fading gain. Then, the allocation of devices is performed in the same way as in the RRS algorithm. This time, however, the placement of devices starts with the one with the best channel gain. In the DUC algorithm, clustering is done by using the channel gain differences of the devices. Normalized channel gains are ordered in descending sequence first. Then, when there is a channel gain difference of at least ten times between the device and the next device, that number is defined as alpha. If the alpha is smaller than six, the cluster size equals to alpha value. If the alpha is greater or equal to six, then the cluster size equals six. The values of the parameters used for uplink NOMA is given in Table 5.2.

In Figure 5.1, the proposed clustering algorithm (CA) is compared with other clustering algorithms; RRS, CRS, DUC, and OMA. These NOMA clustering algorithms have been used by integrating them into the resource allocation algorithm proposed in Algorithm 2. A NOMA sub-cluster can take the maximum size of 1 for the OMA scenario, while it is 4 for other clustering algorithms. Normalized bandwidth is investigated for different number of QoS clusters. Normalized bandwidth is the ratio

Parameters	**Values**
System effective bandwidth	18 MHz
Bandwidth of a resource block	180 kHz
Duration of a resource block	0.5 ms
Number of available bands	100
Transmit power, P	24 dBm
Detection threshold at SIC receiver, P_{Δ}	10 dBm

Table 5.2 Simulation Parameters for Uplink NOMA

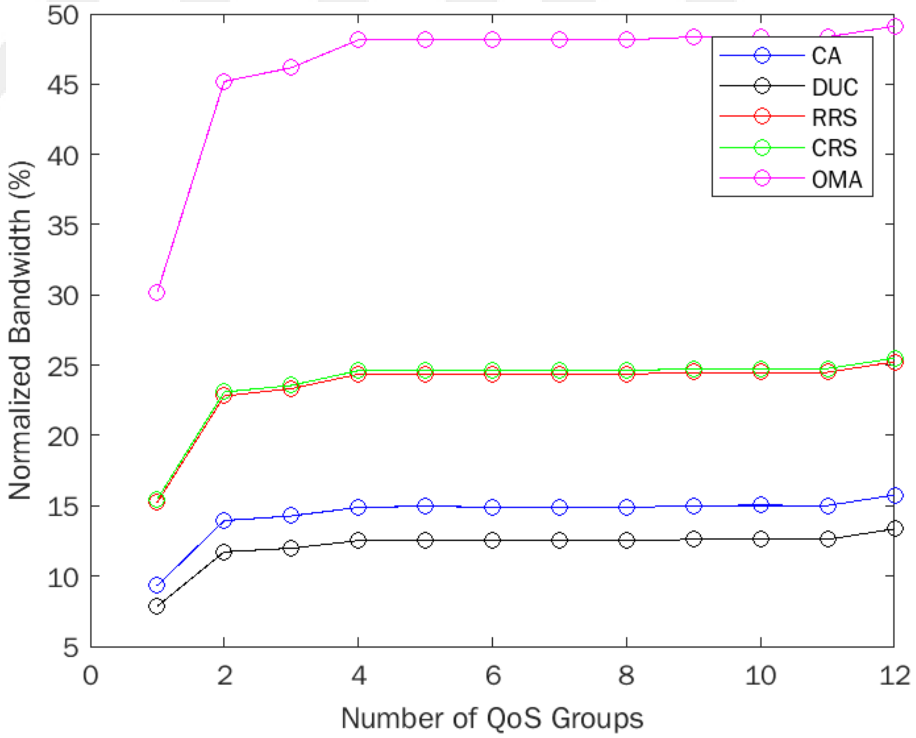

Figure 5.1 Bandwidth performance of the resource allocation algorithm for CA, DUC, RRS, CRS and OMA clustering solutions by number of QoS groups.

of the bandwidth required by the proposed solution to the bandwidth needed for the previously proposed (AGTI-based) algorithm. The proposed NOMA clustering algorithm gives better results than CRS and RRS clustering algorithms in terms of spectral efficiency, and it shows a lower performance than the DUC clustering algorithm. In DUC algorithm, which aims to maximize the sum-throughput of the system, spectral efficiency is not considered. DUC may seem to give a better result, but it consumes more power because it needs to dynamically adjust user clustering to ensure user fairness. Since we fixed the normalized channel gain difference between users as 3 dB to ensure successful SIC at the receiver, the CA algorithm provides a less complex algorithm using less power. RRS gives a slightly better result than CRS because it has less restrictions. Therefore, the use of CSI did not help to minimize the bandwidth usage. However, while each user is assigned individually as long as there are enough channels, like OMA, it gives worse results compared to the CA algorithm. In contrast, it provides a better result than OMA because it assigns devices jointly in the case of limited channels. For OMA, the proposed resource allocation solution (Algorithm 2) reduces the required bandwidth by at least 50% for all scenarios. Bandwidth performance remains stable as the number of QoS clusters increases. This shows the resilience of the resource allocation algorithm against QoS diversity. When used together with the proposed NOMA clustering algorithm, the proposed resource allocation algorithm reduces frequency band usage by 85% for the maximum NOMA cluster size of 4.

Similarly, Fig. 5.2 illustrates the normalized bandwidth for scheduling algorithms CA, RRS, CRS, DUC, and OMA for different number of QoS groups for implicit deadline case. In the implicit deadline case, each device has its maximum jitter tolerance equal to its period. Performance results are similar to Fig. 5.1. The proposed algorithms perform better for the implicit deadline case since it is less restrictive in terms of QoS requirements.

Fig. 5.3 and Fig 5.4 illustrate the normalized bandwidth for resource allocation algorithms CA, RRS, CRS, DUC, and OMA for different number of devices for the

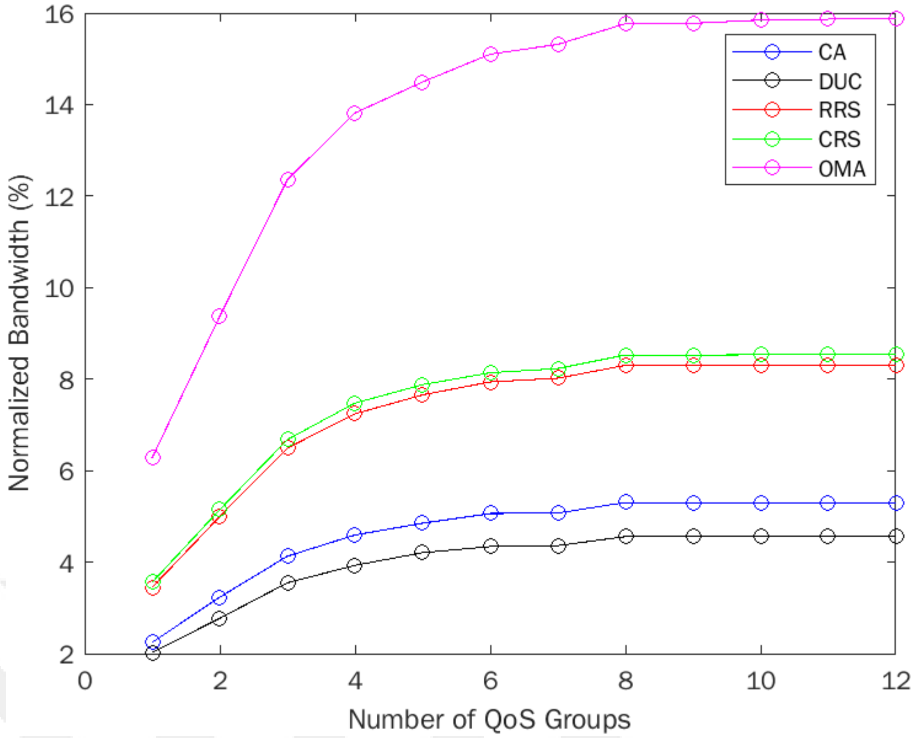

Figure 5.2 Bandwidth performance of the resource allocation algorithm for CA, DUC, RRS, CRS and OMA clustering solutions by number of QoS groups for implicit deadline case.

uniform and the implicit deadline cases, respectively. The number of QoS clusters is 12, and a different number of devices are uniformly shared between those QoS clusters. The proposed algorithm, CA, gives better performance than the other clustering algorithms such as RRS and CRS, while it performs very close to clustering algorithm DUC. Here again, the proposed algorithms perform better for implicit deadline case than the uniform deadline case.

In Fig. 5.5, the effect of maximum NOMA cluster size on the performance of the proposed algorithm is investigated. It is seen that the bandwidth efficiency increases significantly as the cluster size increases. On the other hand, the performance increase decreases as the maximum cluster size value increases. For example, increasing the cluster size from 2 to 3 reduces the normalized bandwidth by 10%, while increasing the cluster size from 3 to 4 decreases the bandwidth by 5%. Considering the complexity of the increase of NOMA cluster size, 3 can be regarded as a practical and spectral-efficient NOMA cluster size.

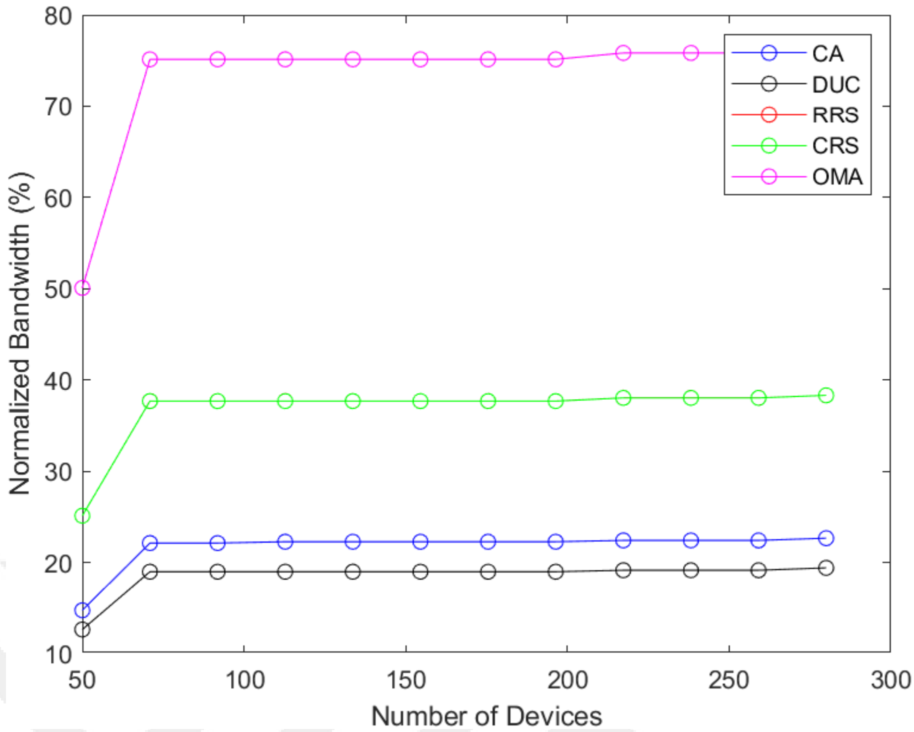

Figure 5.3 Bandwidth performance of the resource allocation algorithm for CA, DUC, RRS, CRS and OMA clustering solutions by number of devices for uniform deadline case.

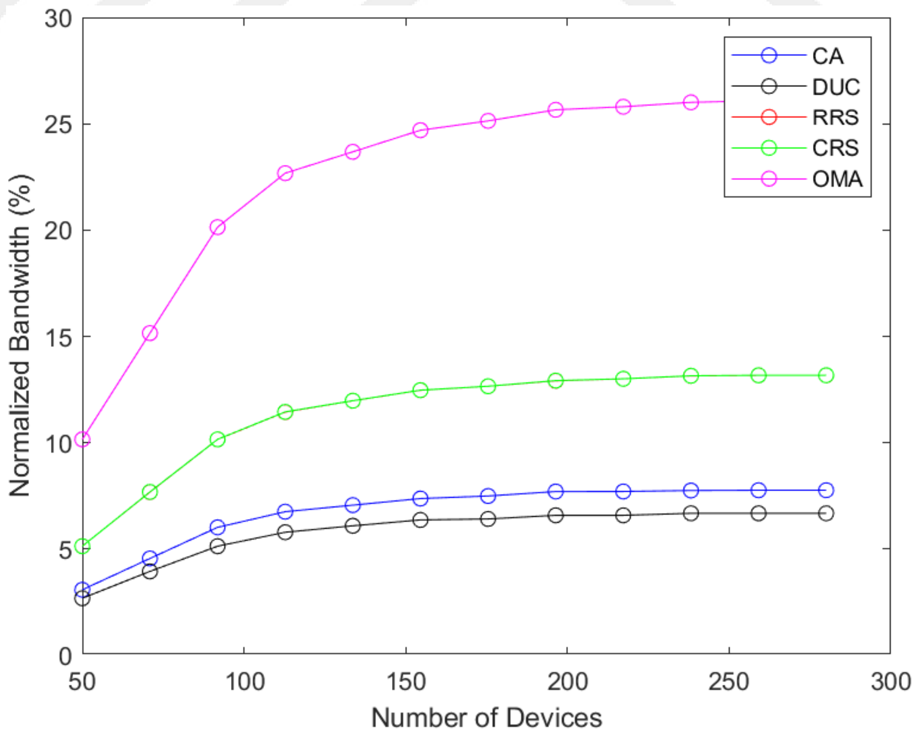

Figure 5.4 Bandwidth performance of the resource allocation algorithm for CA, DUC, RRS, CRS and OMA clustering solutions by number of devices for implicit deadline case

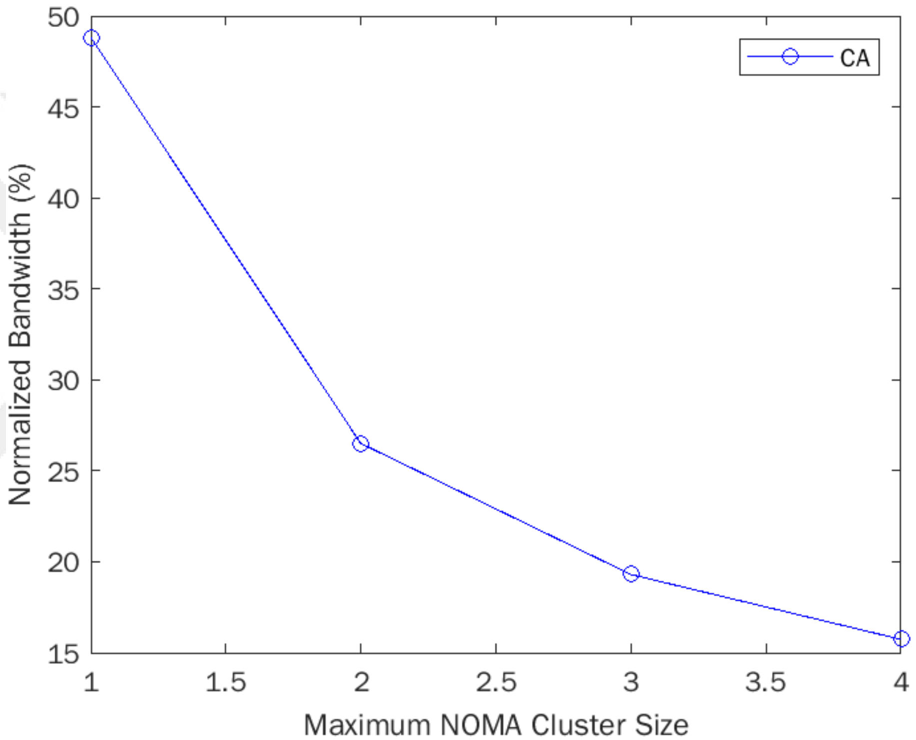

Figure 5.5 Bandwidth performance of the resource allocation algorithm relative to the maximum NOMA cluster size.

6. CONCLUSION

In this thesis, the limited resource problem caused by the rapid growth of M2M communication has been evaluated. Although there are many M2M devices, because of the machines' small data transmission and periodicity feature, we can offer a solution to the massive connection by clustering the devices and assigning them to an RB. Effective user clustering and allocation are the most fundamental design problems for the NOMA system for all these reasons. As a conclusion, a NOMA-based resource allocation study is investigated for the native support of massive M2M communications in 5G and beyond cellular networks. Considering the QoS requirements of M2M devices, a NOMA-based optimization framework is established, and a radio resource allocation algorithm employing a NOMA clustering algorithm is proposed within this framework. The proposed two-phase algorithm first groups QoS clusters into NOMA sub-clusters, minimizing the number of NOMA sub-clusters. The second phase assigns NOMA sub-clusters to resource blocks to reduce bandwidth, considering the QoS requirements characterized by jitter and periodicity. The effect of the number of QoS groups on the normalized bandwidth is observed for CA, DUC, RRS, CRS, and OMA, and the bandwidth performance of the proposed resource allocation algorithm has emerged for the uniform deadline case as a result of this observation. When the proposed NOMA clustering algorithm and the proposed resource allocation algorithm are used together, it has been observed that an 85% efficiency is achieved in the frequency band usage. Again, the effect of QoS on the normalized bandwidth is investigated for the implicit deadline case. The results of the implicit deadline case are similar to the uniform deadline case, but the performance of the proposed algorithms is better for the implicit deadline case because, considering the QoS requirements, the implicit deadline case has a less restrictive attitude. Then, the effect of the number of devices on the resource allocation algorithm was observed, and the relationship between the change in the number of devices and the normalized bandwidth has emerged for the uniform deadline case for the given clustering solutions. Considering the number of QoS clusters fixed to 12, the effects of the change in the number of devices allocated to the clusters were observed. Likewise, the bandwidth performance of the algorithm is examined for the implicit deadline case, and it is seen that the implicit deadline case gives better results than the uniform deadline case. Lastly, the connection between the bandwidth performance of the resource allocation algorithm and the cluster size of NOMA is observed for the proposed clustering solution. It is concluded that bandwidth efficiency is directly proportional to cluster size. In addition, it is observed that the efficiency increase rate decreases as the cluster size increases. Overall, performance results show that the proposed solution significantly reduces the bandwidth used compared to OMA and previously submitted clustering techniques.

As a future study, the optimization framework and the proposed solution strategy will be extended to incorporate multiple numerology based physical layer in New Radio as specified in 5G.

REFERENCES

1. V. Misic and J. Misic, Machine-to-Machine Communications: Architectures, Technology, Standards, and Applications. Boca Raton, FL, USA: CRC Press, 2014.
2. A. Ali, W. Hamouda and M. Uysal, "Next generation M2M cellular networks: challenges and practical considerations," in IEEE Communications Magazine, vol. 53, no. 9, pp. 18-24, September 2015, doi: 10.1109/MCOM.2015.7263368.
3. Z. Dawy, W. Saad, A. Ghosh, J. G. Andrews and E. Yaacoub, "Toward Massive Machine Type Cellular Communications," in IEEE Wireless Communications, vol. 24, no. 1, pp. 120-128, February 2017, doi: 10.1109/MWC.2016.1500284WC.
4. Globecom Workshops, Anaheim, CA, 2012, pp. 1620-1625, doi: 10.1109/GLOCOMW.2012.6477828.A. Kumar, A. Abdelhadi and C. Clancy, "A Delay Optimal Multiclass Packet Scheduler for General M2M Uplink," in IEEE Systems Journal, vol. 13, no. 4, pp. 3815-3826, Dec. 2019, doi: 10.1109/JSYST.2019.2901001.
5. A. Hassebo, A. Rezk and M. A. Ali, "A Hybrid Uplink Scheduling Approach for Supporting Mission-Critical Smart Grid applications in Commercial 4G Cellular Networks," 2018 9th IEEE Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), 2018, pp.303-308, doi:10.1109/UEMCON.2018.8796758.
6. Z. H. Hussien and Y. Sadi, "Flexible radio resource allocation for machine type communications in 5G cellular networks," 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, 2018, pp. 1-4.
7. G. Karadag, R. Gul, Y. Sadi and S. Coleri Ergen, "QoS-Constrained Semi-Persistent Scheduling of Machine-Type Communications in Cellular Networks," in IEEE Transactions on Wireless Communications, vol. 18, no. 5, pp. 2737-2750, May 2019.
8. M. Hasan, E. Hossain and D. Niyato, "Random access for machine-to-machine communication in LTE-advanced networks: issues and approaches," in IEEE Communications Magazine, vol. 51, no. 6, pp. 86-93, June 2013.

9. K. Lee, S. Kim and B. Yi, "Throughput comparison of random access methods for M2M service over LTE networks," 2011 IEEE GLOBECOM Workshops (GC Wkshps), Houston, TX, 2011, pp. 373-377.
10. M. S. Ali, E. Hossain and D. I. Kim, "LTE/LTE-A Random Access for Massive Machine-Type Communications in Smart Cities," in IEEE Communications Magazine, vol. 55, no. 1, pp. 76-83, January 2017.
11. P. Popovski et al., "Wireless Access for Ultra-Reliable Low-Latency Communication: Principles and Building Blocks," in IEEE Network, vol. 32, no. 2, pp. 16-23, March-April 2018, doi: 10.1109/MNET.2018.1700258.
12. C. Wang, Y. Chen, Y. Wu and L. Zhang, "Performance Evaluation of Grant-Free Transmission for Uplink URLLC Services," 2017 IEEE 85th Vehicular Technology Conference (VTC Spring), Sydney, NSW, 2017, pp. 1-6, doi: 10.1109/VTC-Spring.2017.8108593.
13. S. Ali, N. Rajatheva and W. Saad, "Fast Uplink Grant for Machine Type Communications: Challenges and Opportunities," in IEEE Communications Magazine, vol. 57, no. 3, pp. 97-103, March 2019, doi: 10.1109/MCOM.2019.1800475.
14. 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), 2013, pp. 1-5, doi: 10.1109/VTCSpring.2013.6692652.
15. Islam, S. M. Riazul and Zeng, Ming and Dobre, Octavia Kwak, Kyung. (2019). Non-Orthogonal Multiple Access (NOMA): How It Meets 5G and Beyond.
16. Y. Wang, B. Ren, S. Sun, S. Kang and X. Yue, "Analysis of non-orthogonal multiple access for 5G," in China Communications, vol. 13, no. Supplement2, pp. 52-66, 2016, doi: 10.1109/CC.2016.7833460.
17. L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-lin and Z. Wang, "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends," in IEEE Communications Magazine, vol. 53, no. 9, pp. 74-81, September 2015, doi: 10.1109/MCOM.2015.7263349.
18. 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," in IEEE Transactions on Vehicular Technology, vol. 66, no. 4, pp. 3185-3196, April 2017, doi: 10.1109/TVT.2016.2596438.

19. Z. Zhang, H. Sun and R. Q. Hu, "Downlink and Uplink Non-Orthogonal Multiple Access in a Dense Wireless Network," in IEEE Journal on Selected Areas in Communications, vol. 35, no. 12, pp. 2771-2784, Dec. 2017, doi: 10.1109/JSAC.2017.2724646.
20. Y. Endo, Y. Kishiyama and K. Higuchi, "Uplink non-orthogonal access with MMSE-SIC in the presence of inter-cell interference," 2012 International Symposium on Wireless Communication Systems (ISWCS), 2012, pp. 261-265, doi: 10.1109/ISWCS.2012.6328370.
21. S. Lien and K. Chen, "Massive Access Management for QoS Guarantees in 3GPP Machine-to-Machine Communications," in IEEE Communications Letters, vol. 15, no. 3, pp. 311-313, March 2011, doi: 10.1109/LCOMM.2011.011811.101798.
22. S. Lien, K. Chen and Y. Lin, "Toward ubiquitous massive accesses in 3GPP machine-to-machine communications," in IEEE Communications Magazine, vol. 49, no. 4, pp. 66-74, April 2011, doi: 10.1109/MCOM.2011.5741148.
23. A. G. Gotsis, A. S. Lioumpas and A. Alexiou, "Evolution of packet scheduling for Machine-Type communications over LTE: Algorithmic design and performance analysis," 2012 IEEE
24. M. Shirvanimoghaddam, M. Dohler and S. J. Johnson, "Massive Non-Orthogonal Multiple Access for Cellular IoT: Potentials and Limitations," in IEEE Communications Magazine, vol. 55, no. 9, pp. 55-61, Sept. 2017, doi: 10.1109/MCOM.2017.1600618.
25. M. Shirvanimoghaddam, M. Condoluci, M. Dohler and S. J. Johnson, "On the Fundamental Limits of Random Non-Orthogonal Multiple Access in Cellular Massive IoT," in IEEE Journal on Selected Areas in Communications, vol. 35, no. 10, pp. 2238-2252, Oct. 2017, doi: 10.1109/JSAC.2017.2724442.
26. Z. Zhang, H. Sun and R. Q. Hu, "Downlink and Uplink Non-Orthogonal Multiple Access in a Dense Wireless Network," in IEEE Journal on Selected Areas in Communications, vol. 35, no. 12, pp. 2771-2784, Dec. 2017, doi: 10.1109/JSAC.2017.2724646.

27. R. Abbas, M. Shirvanimoghaddam, Y. Li and B. Vucetic, "A Novel Analytical Framework for Massive Grant-Free NOMA," in IEEE Transactions on Communications, vol. 67, no. 3, pp. 2436-2449, March 2019, doi: 10.1109/TCOMM.2018.2881120.
28. Z. Ding, P. Fan and H. V. Poor, "Impact of User Pairing on 5G Nonorthogonal Multiple-Access Downlink Transmissions," in IEEE Transactions on Vehicular Technology, vol. 65, no. 8, pp. 6010-6023, Aug. 2016, doi: 10.1109/TVT.2015.2480766.
29. LÓPEZ, Onel L. Alcaraz, et al. Hybrid resource scheduling for aggregation in massive machine-type communication networks. Ad Hoc Networks, 2019, 94: 101932.
30. M. Pischella and D. Le Ruyet, "NOMA-Relevant Clustering and Resource Allocation for Proportional Fair Uplink Communications," in IEEE Wireless Communications Letters, vol. 8, no. 3, pp. 873-876, June 2019, doi: 10.1109/LWC.2019.2897796.
31. M. S. Ali, H. Tabassum and E. Hossain, "Dynamic User Clustering and Power Allocation for Uplink and Downlink Non-Orthogonal Multiple Access (NOMA) Systems," in IEEE Access, vol. 4, pp. 6325-6343, 2016, doi: 10.1109/ACCESS.2016.2604821.
32. Release 15, document TR 21.915 V15.0.0, 3GPP. Sep 2019.
33. Y. Sadi, S. Erkucuk and E. Panayirci, "Flexible Physical Layer based Resource Allocation for Machine Type Communications Towards 6G," 2020 2nd 6G Wireless Summit (6G SUMMIT), 2020, pp. 1-5, doi: 10.1109/6GSUMMIT49458.2020.9083921.

CURRICULUM VITAE

Personal Information
Name and Surname : Sümeyra Aldemir

Academic Background
Bachelor's Degree Education :Kadir Has University (GPA: 3.86)
Post Graduate Education: Kadir Has University (GPA: 3.68)
Foreign Languages: English (C1)
Korean (A1)

Work Experience
Karabük Kardemir Demir-çelik Fabrikaları A.Ş. 07.08.2017 – 06.09.2017
Aselsan A.Ş. 18.06.2018 – 13.07.2018
Kadir Has University, Assistant of Computer Lab. 01.10.2018 – 01.06.2019
Dione Local Development Center 01.07.2019 – 30.08.2019
Tübitak, Project Assistant 11.09.2019 – 07.10.2021