RESOURCE ALLOCATION FOR MASSIVE MULTIPLE INPUT MULTIPLE OUTPUT BASED COMMUNICATIONS SYSTEMS

A Thesis Submitted to
the Graduate School of Engineering and Sciences of
İzmir Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in Electronics and Communication Engineering

by
Saadet Simay YILMAZ

June 2023
İZMİR

We approve the thesis of **Saadet Simay YILMAZ**

Examining Committee Members:

Prof. Dr. Berna ÖZBEK
Department of Electrical and Electronics Engineering, Izmir Institute of Technology

---

Prof. Dr. Mustafa A. ALTINKAYA
Department of Electrical and Electronics Engineering, Izmir Institute of Technology

---

Prof. Dr. Radosveta İVANOVA SOKULLU
Department of Electrical and Electronics Engineering, Ege University

---

Prof. Dr. Ali Özgür YILMAZ
Department of Electrical and Electronics Engineering, Middle East Technical University

---

Prof. Dr. Barış ATAKAN
Department of Electrical and Electronics Engineering, Izmir Institute of Technology

9 June 2023

Prof. Dr. Berna ÖZBEK
Supervisor, Department of Electrical and Electronics Engineering
Izmir Institute of Technology

Prof. Dr. Mustafa A. ALTINKAYA
Head of the Department of
Electrical and Electronics Engineering

Prof. Dr. Mehtap EANES
Dean of the Graduate School of
Engineering and Sciences

To the memory of my beloved father, Süleyman YILMAZ.

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to my advisor, Prof. Dr. Berna Özbek, for directing me toward various exciting research areas and for her expert guidance, mentorship, and continuous support throughout this research journey. Her vast knowledge, valuable insights, invaluable advice, and constructive feedback have been instrumental in shaping the direction and quality of this thesis. It has been an honor for me to have the chance to work with her. I would like to thank her from the bottom of my heart for her dedication, patience, and encouragement, which have been a constant source of motivation for me since my ten-year academic journey.

I would also like to express my gratitude to my thesis progress committee members, Prof. Dr. Mustafa Aziz Altınkaya and Prof. Dr. Radosveta İvanova Sokullu, for their valuable feedback, contributions, suggestions, and critical reviews of my research study. I owe particular thanks to each member of the thesis committee, Prof. Dr. Ali Özgür Yılmaz and Prof. Dr. Barış Atakan, for their valuable contributions to revising the thesis. Their expertise and insights have been invaluable in improving the quality and rigor of this thesis.

I would like to express my deepest gratitude to my colleagues and close friends who have supported and contributed to completing my Ph.D. thesis. I owe particular thanks to Enes Ataç for his companionship, motivation talks, endless support and help during my Ph.D. journey, my roommate İrem Cumalı for her support and understanding and Caner Göztepe for many valuable discussions and sharing of his experiences. This thesis and the study would not have been possible without their support, guidance, and encouragement.

Last but definitely not least, I would like to extend my deepest gratitude and love to my beloved family, Saliha and Süleyman Yılmaz, Simge Yılmaz Kavcar, Yücel Kavcar for their years of love and endless unconditional support. Very special thanks to my nephew, Kerem Kavcar, for bringing me so much happiness. This accomplishment would not have been possible without them. Thank you my family both for giving me strength and for always being with me. And my father, Süleyman Yılmaz, you are still with me in my smile, tears, and success!

ABSTRACT

RESOURCE ALLOCATION FOR MASSIVE MULTIPLE INPUT MULTIPLE OUTPUT BASED COMMUNICATIONS SYSTEMS

In this thesis, we examine resource allocation strategies for massive multiple-input and multiple-output (MIMO) based wireless communications systems to increase system performance, considering computation-intensive applications with low-latency communication. Firstly, we propose user selection algorithms for non-orthogonal multiple-access (NOMA)-based massive MIMO systems in densely deployed scenarios to increase the sum data rate. Then, we investigate mobile edge computing (MEC) as a solution to enable computation-intensive and delay-critical applications. We propose resource allocation algorithms considering the downlink and uplink transmit powers, the task offloading decision factor and the computing resources to reduce both transmission and computing delays for the massive MIMO-NOMA-assisted MEC system. Finally, we consider a cooperative MEC system where helpers assist in the execution of cell-edge users' computation-intensive tasks with low latency. On the other hand, the task offloading in MEC can introduce security concerns as the offloaded data may be intercepted and overheard by eavesdroppers. Since ensuring a secure task offloading scheme in MEC is important, we formulate the optimization problem to minimize both offloading and computing delays while satisfying security constraints for a massive MIMO-based cooperative MEC. We provide performance results based on sum data rate, delay and total offloading data for the proposed schemes in massive MIMO based wireless communication systems.

ÖZET

YOĞUN ÇOK GİRİŞLİ ÇOK ÇIKIŞLI HABERLEŞME SİSTEMLERİ İÇİN KAYNAK TAHSİSİ

Bu tezde, düşük gecikmeli haberleşme ile yoğun hesaplama gerektiren uygulamalar için, sistem performansını artırmak amacıyla yoğun çok girişli ve çok çıkışlı (MIMO) kablosuz iletişim sistemlerinde kaynak tahsisi stratejileri incelenmektedir. Öncelikle, toplam veri hızını artırmak amacıyla yoğun olarak konuşlandırılmış senaryolarda dikgen olmayan çoklu erişim (NOMA) tabanlı yoğun MIMO sistemleri için kullanıcı seçim algoritmaları önerilmektedir. Ardından, yoğun hesaplama gerektiren ve gecikme açısından kritik olan uygulamaları gerçekleştirmek için bir çözüm olarak mobil uç hesaplama (MEC) sistemi incelenmektedir. Yoğun MIMO-NOMA destekli MEC sistemi için, hem iletim hem de hesaplama gecikmelerini azaltmak amacıyla aşağı bağlantı ve yukarı bağlantı iletim güçlerini, görev iletim karar faktörünü ve hesaplama kaynaklarını dikkate alan kaynak tahsis algoritmaları önerilmektedir. Son olarak, yardımcıların düşük gecikme süresiyle hücre ucundaki kullanıcıların yoğun hesaplama gerektiren görevlerini yürütmesine yardımcı olan işbirlikçi bir MEC sistemi dikkate alınmaktadır. Öte yandan, MEC'de görev iletimi sırasında, verilerin gizli dinleyiciler tarafından engellenebilir ya da duyulabilir olması güvenlik sorunu yaratabilir. MEC'de güvenli bir şekilde görev iletimini sağlamak önemli olduğundan, yoğun MIMO tabanlı işbirlikçi MEC sisteminde güvenlik kısıtlamaları göz önünde bulundurularak, hem iletim hem de hesaplama gecikmelerini en aza indirmek için bir optimizasyon problemi formüle edilmektedir. Yoğun MIMO tabanlı kablosuz iletişim sistemlerinde, önerilen şemalar için toplam veri hızı, gecikme ve toplam iletilen veriye dayalı performans sonuçları sağlanmaktadır.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

3D . . . . . Three-Dimension
6G . . . . . Sixth Generation
APs . . . . . Access Points
AR . . . . . Augmented Reality
ASD . . . . . Angular Standard Deviation
AWGN . . . . . Additive White Gaussian Noise
B5G . . . . . Beyond Fifth Generation
BS . . . . . Base Station
CCI . . . . . Co-channel Interference
CIR . . . . . Channel Impulse Response
CPU . . . . . Central Processing Unit
CSI . . . . . Channel State Information
CSIT . . . . . Channel State Information at the Transmitter
DFT . . . . . Discrete Fourier Transform
DL . . . . . Downlink
EE . . . . . Energy Efficiency
FDD . . . . . Frequency Division Duplex
i.i.d. . . . . Independent and identically distributed
IoT . . . . . Internet of Things
MEC . . . . . Mobile Edge Computing
MIMO . . . . . Multiple-Input Multiple-Output
massive MIMO . . . . . Massive Multiple-Input Multiple-Output
MMSE . . . . . Minimum Mean Squared Error
MSE . . . . . Mean Squared Error
NLoS . . . . . Non-Line Of Sight
NOMA . . . . . Non-Orthogonal Multiple Access
OMA . . . . . Orthogonal Multiple Access
PLS . . . . . Physical Layer Security
RAN . . . . . Radio Access Network
SIC . . . . . Successive Interference Cancellation
SINR . . . . . Signal-to-interference-plus-noise-ratio
SNR . . . . . Signal-to-Noise Ratio
SUS . . . . . Semiorthogonal User Selection

TDD . . . . . Time Division Duplex
TDMA . . . . . Time Division Multiple Access
UAVs . . . . . Unmanned Aerial Vehicles
UE . . . . . User Equipment
UL . . . . . Uplink
ULA . . . . . Uniform Linear Array
VR . . . . . Virtual Reality
ZF . . . . . Zero-forcing

CHAPTER 1

INTRODUCTION

1.1. Background and Motivation

In recent years, there has been an exponential increase in demand for wireless data traffic and the number of connected devices. In addition to this demand, mobile users request seamless service at higher data rates to perform data-intensive applications. To cope with this growing request, it is necessary to introduce new technologies that can enhance spectral efficiency while enabling spatial reuse of the available spectrum. This can be achieved by utilizing multiple antennas at the base station (BS) and/or user equipment (UE), which increases the number of degrees of freedom available through the propagation channel. Therefore, multiple-input and multiple-output (MIMO) has been presented to significantly improve the sum data rate without increasing the bandwidth or the transmit power.

With the increased demand for higher spectral efficiency and reduced delay, massive MIMO technology at the BS with a large number of antennas has been provided as a promising solution for these requirements. By using a large number of antennas, the BS can accommodate multiple co-channel users simultaneously. In addition to the high spectral efficiency, these systems offer several other advantages, such as diversity gain, improved reliability, and increased energy efficiency. Furthermore, accurately estimating the channel response between each transmit-receive link is critical for effectively leveraging the benefits of massive MIMO. Several studies have demonstrated that as the number of antennas in a system grows, the user channel tends to be sparse (Lahbib et al., 2019). The massive MIMO channel responses exhibit sparsity because a few number of paths contain the majority of the received energy. Thus, instead of relying on a large training sequence to estimate the channel matrix, it is recommended to leverage the sparsity of the channel response. Estimating the channel response in sparse uplink channels of massive MIMO systems can be accomplished using compressive sensing (CS) to recover the original sparse signal with only a small number of measurements. Numerous greedy algorithms have been developed in the literature to address the CS problem. These

algorithms include matching pursuit (MP) (Mallat and Zhang, 1993), orthogonal matching pursuit (OMP) (Davis et al., 1997), regularized orthogonal matching pursuit (ROMP) (Needell and Vershynin, 2010), stagewise orthogonal matching pursuit (StOMP) (Donoho et al., 2012), orthogonal complementary matching pursuit (OCMP) (Rath and Guillemot, 2009), compressive sampling matching pursuit (CoSaMP) (Needell and Tropp, 2009), and gradient pursuit (GP) (Blumensath and Davies, 2008).

Although the next-generation wireless systems provide several techniques, such as MIMO and massive MIMO, it is essential to design multiple access schemes. Non-orthogonal multiple-access (NOMA) has become an important technology due to its higher spectral efficiency and massive connectivity compared to orthogonal multiple access (OMA) schemes, such as time-division multiple access (TDMA), frequency-division multiple access (FDMA), and code-division multiple access (CDMA). In NOMA, the same resource block of time, frequency, or code is shared by multiple users; superposition coding and successive interference cancellation (SIC) can be used at the transceivers to manage multiple-access interference.

In densely deployed scenarios that accommodate a large number of users, user selection becomes essential since the BS cannot serve all of them simultaneously. Because there can be more users to be supported than the number of transmit antennas available at the BS. Thus, user selection is an important technique to improve system performance while reducing inter-user interference.

On the other hand, increasing mobile computation-intensive applications and the finite computation capacities of devices have recently generated new challenges in next-generation systems. Since the applications have strict latency requirements and heavy computation needs, the computing capability of the devices, e.g., local central processing unit (CPU) frequency and memory, may be insufficient. Thus, the limited computation capability can result in a degraded quality of user experience, such as excessive delay and power consumption (Chen et al., 2022). Moreover, to fulfill the demands of real-time tasks, it is essential to transmit a massive amount of data with very low latency. Towards this end, mobile edge computing (MEC) is one of the most promising solutions for computation-intensive and delay-sensitive applications such as virtual reality (VR), augmented reality (AR), autonomous driving, interactive gaming, remote healthcare systems, unmanned aerial vehicles (UAVs) and the Internet of Things (IoT). In MEC systems, computing is implemented at the network edge, such as the BS, to significantly reduce transmission distances compared to the cloud computing. MEC servers can be connected to the BSs either through a backhaul link or directly integrated into the BSs

using generic computing platforms (Zhao et al., 2019). This provides cloud-computing services in close proximity to mobile users. Compared with cloud computing, in MEC systems, users offload computation-intensive tasks to the powerful MEC servers for execution, which reduces latency. Thus, MEC can considerably decrease computation latency and significantly reduce traffic loads on the backhaul networks (Zhao et al., 2020a).

Besides the advantages of the MEC system, it can be challenging to offload computation tasks to the MEC when mobile devices are located far away from the network edge having weak channel gain. To overcome these issues, cooperative edge computing leveraging user cooperation is an effective solution (Liu, 2019). This approach involves nearby mobile devices in close proximity to the BS sharing their computational and radio resources to assist those far away mobile devices in executing computation tasks. During task offloading, part or all of the computationally intensive tasks on mobile devices are transferred to the MEC server for remote computation. Thus, the computation capacity is improved (Pan et al., 2021) and, the delay resulting from task offloading and task execution at the server can be significantly reduced.

The performance of MEC systems can be improved by the multiple access strategy since multiple users may need to access the same server for task offloading (Wen et al., 2020). In this regard, NOMA can be implemented in MEC systems to meet the demands of extremely high data rates (Li et al., 2020). By combining MEC with NOMA, delay performance can be improved compared to MEC with OMA systems. On the other hand, utilizing massive MIMO technology in MEC systems can enhance both spectral efficiency and energy efficiency (Feng et al., 2020). Specifically, the coexistence of MEC and massive MIMO is necessary to enable massive wireless connectivity with high data rates, low-latency, and large computing capabilities. Therefore, the integration of NOMA based massive MIMO into MEC systems can further enhance computing capability, increase spectral efficiency, and reduce task delay (Pham et al., 2020).

On the other hand, there can be security challenges during the offloading process in MEC. Since the offloading data may be intercepted and overheard by eavesdroppers, a secure task offloading scheme in MEC systems is essential (Wang et al., 2020). In this context, the physical layer security (PLS) technique has provided as a promising solution to address the security issues for task offloading in MEC systems (Qian et al., 2021a).

In this thesis, we study resource allocation for massive MIMO based communications systems. Firstly, the proposed user selection algorithms are employed for NOMA-MIMO based systems and massive MIMO-NOMA based systems to improve the sum data rate. For massive MIMO systems, we consider reconstructing the sparse channel of

massive MIMO with the OMP to reduce the feedback load. In addition, we investigate the effect of the user selection algorithm on massive MIMO-NOMA based MEC systems to reduce overall computing and transmission delay significantly.

It is important to investigate resource allocation for MEC systems in densely deployed scenarios. This thesis also focuses on the integration of massive MIMO and NOMA technologies into the MEC system to facilitate the offloading and further improve the performance of the MEC system in terms of the amount of offloaded data and overall computing and transmission delay. Furthermore, in the system, there may be a cell-edge user with a computation-intensive and latency-critical task. The cell-edge user may suffer from a low signal-to-noise ratio (SNR) or there may not be a strong direct transmission link to the BS. Considering these users, we study the cooperative computation offloading and resource allocation strategies in the MEC system. Thus, a multi-helper cooperative MEC system based on NOMA has been proposed to maximize the total offloading data under latency and power constraints. In addition, we investigate the massive MIMO based cooperative MEC system to minimize the offloading and computing delay. Since secure offloading in MEC systems is of critical importance, we also study secure offloading for the massive MIMO based cooperative MEC system.

1.2. Outline of the Thesis

The thesis is organized as follows:

- In Chapter 2, the background information related to MIMO, massive MIMO and NOMA technologies is described. Next, we briefly review the MIMO and massive MIMO channel models. Since user selection plays a critical role in improving system performance, we explain the traditional user selection algorithms. Moreover, we introduce the CS approach, including OMP, to reconstruct the compressed signal accurately and reduce the feedback load. Accordingly, in Section 2.5, the compressive sensing based low complexity user selection is proposed for massive MIMO systems by reconstructing the sparse channel of massive MIMO with the OMP to reduce the feedback load (Yılmaz and Özbek, 2020). Section 2.6 introduces a user selection algorithm for NOMA based MIMO systems (Yılmaz et al., 2022). Moreover, the proposed user selection algorithm is also applied to NOMA based massive MIMO systems in Section 2.7.

- In Chapter 3, an overview of the MEC system is introduced. Section 3.2 addresses the task offloading delay minimization problem for the massive MIMO-NOMA based MEC system (Yılmaz and Özbek, 2023). The proposed framework enables both cell-center and cell-edge users to offload their tasks to the MEC server by applying an efficient user pairing, offloading and computation scheme. In Section 3.3, the user selection algorithm is applied to the massive MIMO-NOMA based MEC system to reduce the task offloading delay significantly.

- In Chapter 4, cooperative MEC systems are represented. Section 4.1 presents a multi-helper cooperative MEC system based on NOMA (Yılmaz and Özbek, 2022). We present the resource allocation problem to maximize the total offloading data under the latency and power constraints for the proposed framework. Section 4.2 proposes a novel framework for a cooperative MEC system with massive MIMO and NOMA technologies to minimize the overall delay, including security aspects (Yılmaz et al., 2023). The problem of minimizing the overall computing and transmission delay over the massive MIMO-NOMA assisted MEC system is formulated under computing capability and transmit power constraints.

- Chapter 5 concludes the thesis and discusses potential future research directions.

CHAPTER 2

USER SELECTION FOR MIMO BASED SYSTEMS

In this chapter, background information about multiple-input multiple-output (MIMO), massive MIMO, non-orthogonal multiple access (NOMA) based systems are provided. Then, user selection algorithms are investigated. Following this, we propose user selection algorithms for massive MIMO system, NOMA based MIMO system and NOMA based massive MIMO systems.

2.1. Multiple-Input Multiple-Output (MIMO) Systems

MIMO systems have emerged as one of the most promising technologies for wireless communications to provide a reliable communication link through spatial diversity and a higher data rate by using multiplexing techniques without requiring additional power or bandwidth. Specifically, MIMO systems can achieve multiplexing gain by transmitting different signals through different antennas, which results in a higher data rate. These systems can achieve diversity gain by transmitting the same signals through different paths, resulting in a higher probability of detection and higher reliability. Moreover, MIMO technology achieves higher spectral efficiency since multiple antennas at the BS can simultaneously serve multiple users while separating them in the spatial domain.

2.1.1. Channel Model for MIMO Systems

Let N be the number of antennas at the BS and K be the number of users with a single antenna, where the MIMO channel vector $\mathbf{h}_k$ is defined as;

$$\mathbf{h}_k = \sqrt{L_k}\,\mathbf{g}_k \tag{2.1}$$

where $\mathbf{h}_k$ is the $N \times 1$ channel vector between the BS and the k^{th} user. The path-loss coefficient for the k^{th} user is represented by L_k, which is the large-scale fading component.

The small-scale Rayleigh fading component of the channel with size $N \times 1$ is represented by $\mathbf{g}_k$, which each element of $\mathbf{g}_k$ is modeled as independent and identically distributed (i.i.d.) and Gaussian random variables denoted by $\mathcal{CN}(0, 1)$.

2.2. Massive MIMO Systems

Massive MIMO systems have been the most outstanding technology in terms of supporting high spectral efficiency and energy efficiency for the wireless communications. Adopting the multi-user concept and providing extra degrees of freedom in the spatial domain can significantly improve the multiplexing gain and reliability of such systems, allowing many tens of users to be served more effectively in the same time-frequency resource in massive MIMO systems.

Current MIMO systems generally employ up to eight antennas at the BS, whereas in massive MIMO systems, the BS is equipped with hundreds of antennas. The key idea is based on using a large number of transmit antennas to serve multiple users simultaneously where the number of antennas at the BS is much higher than the number of served users. The fading effects are completely eliminated as the number of antennas at the BS approaches infinity (Marzetta, 2010). In particular, using a high number of antennas, the small-scale fading effect can be eliminated through channel hardening, resulting in a deterministic scalar channel model (Hochwald et al., 2004). More specifically, when increasing the number of antennas at the BS, the channel variations decrease and the channel hardening effect appears in which the variations of the channel gain in time and frequency decrease. Moreover, massive MIMO systems can achieve very high sum data rate under favorable propagation conditions by utilizing simple linear processing (Rusek et al., 2013). These advantages make massive MIMO very attractive for wireless communication systems.

2.2.1. Channel Model for Massive MIMO Systems

In this subsection, the correlated Rayleigh fading with the local scattering channel model is considered (Björnson et al., 2017). We define $\mathbf{h}_k \in \mathbb{C}^{Nx1}$ is the channel vector between the k^{th} user and the BS with N antennas. For two-dimensional (2D) channel models, the channel vector due to the effect of clusters of scatterers is written as follows (Ngo et al., 2013) (Björnson et al., 2017) (Zhang et al., 2022);

$$\mathbf{h}_k = \sqrt{\frac{L_k}{N_p}} \sum_{i=1}^{N_p} \mathbf{a}\left(\phi_{k,i}\right) c_{ki} \tag{2.2}$$

where N_p is the number of Non-Line Of Sight (NLoS) paths from the BS to the k^{th} user. c_{ki} is the channel coefficient between the k^{th} user and the BS associated with each path which is modeled as Rayleigh fading by $\mathbb{CN}(0,1)$. Here, $\mathbf{a}\left(\phi_{k,i}\right) \in \mathbb{C}^{Nx1}$ is the steering array vector at the BS which denotes the antenna array response of the i^{th} path of the k^{th} user in the direction of $\phi_{k,i}$. For uniform linear array (ULA) antenna models, the array response (Hawej and Shayan, 2018) (Dai et al., 2019) (Hawej and Shayan, 2019) is defined as;

$$\mathbf{a}\left(\phi_{k,i}\right) = \left[1,\ e^{-j2\pi\frac{D}{\lambda}cos(\phi_{k,i})}, \ldots, e^{-j2\pi\frac{D}{\lambda}(N-1)cos(\phi_{k,i})}\right]^T \tag{2.3}$$

where $\phi_{k,i}$ is the angle of azimuth of the i^{th} path for the k^{th} user and $D/\lambda = 0.5$, where λ is the wavelength of the carrier signal.

The azimuth angle can be expressed as $\phi_{k,i} \triangleq \phi_k + \delta_i$ with a nominal angle $\phi_k \in [-90°, 90°]$ and a random deviation modeled as a uniform variate which lies within, $\delta_i \sim U\left[-\sqrt{3}\sigma_\phi, \sqrt{3}\sigma_\phi\right]$, from the nominal angle with the angular standard deviation (ASD), σ_ϕ (Björnson et al., 2017).

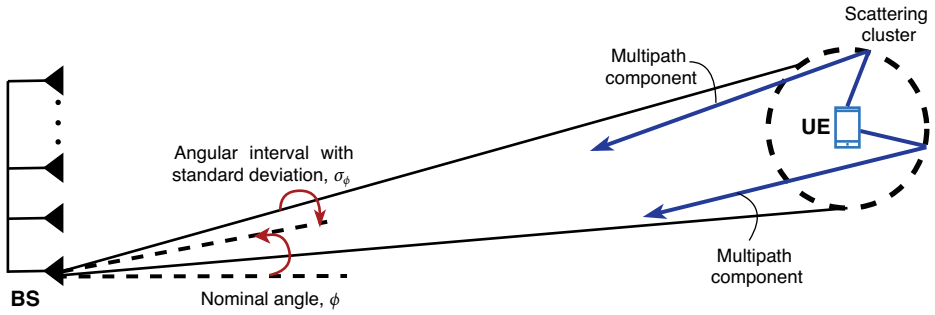


Figure 2.1. Illustration of NLoS propagation under the local scattering model (Source: Björnson et al. 2017).

Fig. 2.1 shows that the scattering is localized around the user equipment (UE). The nominal angle ϕ and ASD σ_ϕ of the multi-path components are key parameters to model the spatial correlation matrix.

On the other hand, the channel model given in (2.1) can also be used as massive MIMO channel (Khansefid and Minn, 2014) (Zheng et al., 2014) (Zeng et al., 2020a) (Chaves et al., 2022), namely uncorrelated massive MIMO channel.

2.3. Non-Orthogonal Multiple Access (NOMA)

NOMA is investigated as one of the possible enabling multiple access technologies for wireless communications systems to meet the requirements such as ultra-low latency and massive connectivity while enhancing the overall spectral efficiency. The main concept of NOMA is to simultaneously serve multiple users at the same time, same frequency, or same code (Shi et al., 2020). In literature, there are mainly two types of NOMA technologies: Power-domain NOMA (Islam et al., 2017) and code-domain NOMA (Shental et al., 2017). NOMA uses superposition coding at the transmitter so that the SIC receiver can separate the users both in the uplink and downlink channels. We consider power domain NOMA in which the power difference of the users is exploited for multi-user multiplexing and apply SIC to mitigate the inter-user interference (Wei et al., 2020).

Since NOMA employs more than one user at the same resource block, it improves the sum data rate compared to its traditional counterpart (Qian et al., 2021b), orthogonal multiple access (OMA), such as time division multiple access (TDMA), frequency division multiple access (FDMA) and code division multiple access (CDMA) (Dai et al., 2015) (Islam et al., 2017) (Wang et al., 2018) (Wu et al., 2018). On the other hand, the latency performance of power-domain NOMA versus OMA has also been examined in recent literature. Specifically, the study presented in (Yu et al., 2018) has investigated the link-layer capacity, namely, the effective capacity, of NOMA under probabilistic delay constraints and shown that NOMA outperforms OMA at high signal-to-noise ratios. Further investigations have investigated the performance of NOMA systems in terms of ultra-reliable low-latency communications requirements (Amjad and Musavian, 2018).

2.3.1. Downlink NOMA

Fig. 2.2 shows the downlink (DL) NOMA for the case of one BS with N antenna and two user equipments (UEs) with single-antennas.

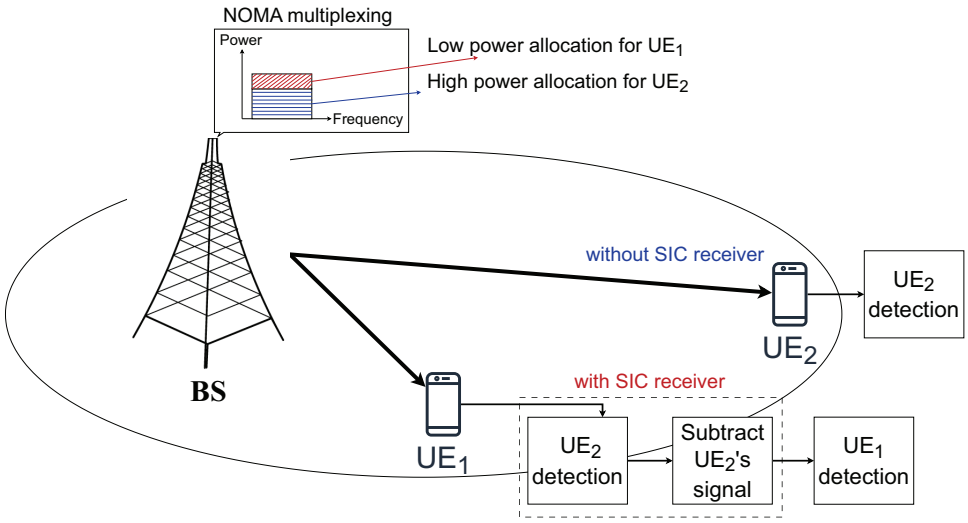


Figure 2.2. DL NOMA system with two UEs.

The overall system transmission bandwidth is set to B. The BS transmits the signal s_k for each k user, where $k = 1, 2$ and $\mathbb{E}\left[|s_k|^2\right] = 1$ with transmit power P_k, where $\sum_{k=1}^{2} P_k = P$. In downlink NOMA, the signals of the two UEs, s_1 and s_2, are superposed as follows (Luo and Zhang, 2016);

$$s = \sqrt{P_1}s_1 + \sqrt{P_2}s_2 \tag{2.4}$$

The received signal at kth user is given as;

$$\mathbf{y}_k = \mathbf{h}_k^H \mathbf{w}^d s + \mathbf{n}_k \tag{2.5}$$

where $\mathbf{h}_k$ is the complex channel vector coefficient of size $N \times 1$ between kth user and the BS, $\mathbf{w}^d$ is the normalized precoder vector of size $N \times 1$ for the kth user and is applied to mitigate inter-user interference. Here, $\mathbf{w}^d = \frac{\mathbf{h}_k}{\|\mathbf{h}_k\|}$ is maximum ratio transmission (MRT) precoding technique and can be computed through either the strong or weak user channels. $\mathbf{n}_k$ is additive white Gaussian noise (AWGN) whose elements are modeled by $\mathcal{CN}(0, \sigma_n^2)$, where σ_n^2 is the variance of AWGN.

Multi-user signal separation needs to be implemented at the UE side so that each UE can retrieve its signal and decode its own data. This can be achieved using the SIC

technique. For the case of SIC, the optimal order for decoding is in the order of the decreasing channel gain. Based on this order, any user can correctly decode the signals of other users whose decoding order comes before the corresponding user. Thus, in a two-UE case, assuming that the channel gain of the UE_1 is higher than the UE_2 as given in Fig. 2.2, UE_2 does not perform interference cancellation since it comes first in the decoding order. UE_1 first decodes s_2 and subtracts its component from received signal y_1; then it decodes s_1 without interference from s_2 (Luo and Zhang, 2016). Thus, the data rates of the users are defined as;

$$\begin{aligned} R_1 &= B\log_2\left(1+\frac{P_1\left|\mathbf{h}_1^H\mathbf{w}^d\right|^2}{\sigma_n^2}\right) \\ R_2 &= B\log_2\left(1+\frac{P_2\left|\mathbf{h}_2^H\mathbf{w}^d\right|^2}{P_1\left|\mathbf{h}_2^H\mathbf{w}^d\right|^2+\sigma_n^2}\right) \end{aligned} \tag{2.6}$$

On the other hand, in general, consider N antenna at the transmitter and K single antenna users. The total number of clusters is denoted by M and each cluster accommodates U users. The SIC technique is used between intra-cluster users to detect the signals. We define the channel vector of size $N\times 1$ between the transmitter and the u^{th} user at the m^{th} cluster by $\boldsymbol{h}_{u,m}$. For each cluster, the relation $\|\boldsymbol{h}_{1,m}\| > \|\boldsymbol{h}_{2,m}\| > \ldots > \|\boldsymbol{h}_{U,m}\|$ is maintained. Then, the normalized precoder vector $\mathbf{w}_m^d$ for each cluster is applied. Thus, the DL signal-to-interference-plus-noise ratio (SINR) of u^{th} user in m^{th} cluster is calculated by;

$$\gamma_{u,m}^d = \frac{P_{u,m}\left|\mathbf{h}_{u,m}^H\mathbf{w}_m^d\right|^2}{\underbrace{\left|\mathbf{h}_{u,m}^H\mathbf{w}_m^d\right|^2\sum_{j=1}^{u-1}P_{j,m}}_{\text{Inter-user interference}}+\underbrace{\sum_{i=1,i\neq m}^{M}\left(\left|\mathbf{h}_{u,m}^H\mathbf{w}_i^d\right|^2\sum_{j=1}^{U}P_{j,i}\right)}_{\text{Inter-cluster interference}}+\sigma_n^2} \tag{2.7}$$

The average downlink sum data rate of the system with SIC can be expressed as follows;

$$R_{sum}^d = B\sum_{m=1}^{M}\sum_{u=1}^{U}\mathbb{E}\left\{\log_2\left(1+\gamma_{u,m}^d\right)\right\} \tag{2.8}$$

2.3.2. Uplink NOMA

Fig. 2.3 shows uplink (UL) NOMA with two UEs transmitting signals to the BS on the same frequency resource and at the same time, and SIC conducted at the BS for UE multi-user signal separation.

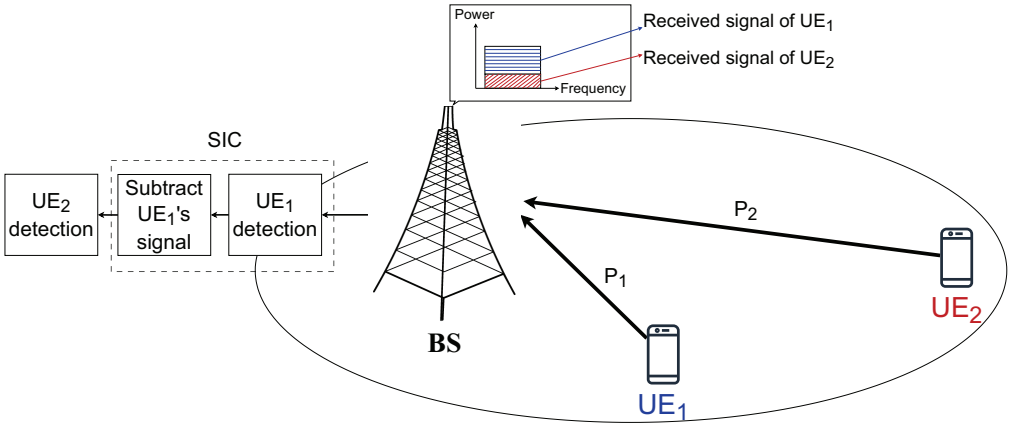


Figure 2.3. UL NOMA system with two UEs.

In uplink NOMA, the signals s_k for each k user, where $k = 1, 2$ are transmitted simultaneously to the BS through the user's channel. The received signal at the BS, $\mathbf{y} \in \mathbb{C}^{N\times 1}$ is a superposed signal of s_1 and s_2 as follows (Luo and Zhang, 2016);

$$\mathbf{y} = \mathbf{h}_1\sqrt{P_1}s_1 + \mathbf{h}_2\sqrt{P_2}s_2 + \mathbf{n} \tag{2.9}$$

where $\mathbf{n} \in \mathbb{C}^{N\times 1}$ is the AWGN vector at the BS.

We assume UE_1 is the cell-center user and UE_2 is the cell-edge user, and the BS conducts SIC according to the descending order of channel gains. The received signals of the first and second UE at the BS after post-coding with the combiner vector $\mathbf{w} \in \mathbb{C}^{N\times 1}$ are written, respectively, by (Thet et al., 2020);

$$\begin{aligned} r_1 &= \mathbf{w}^H\mathbf{h}_1\sqrt{P_1}s_1 + \mathbf{w}^H\mathbf{h}_2\sqrt{P_2}s_2 + \mathbf{w}^H\mathbf{n}, \\ r_2 &= \mathbf{w}^H\mathbf{h}_2\sqrt{P_2}s_2 + \mathbf{w}^H\mathbf{n}. \end{aligned} \tag{2.10}$$

For two UEs case, since the received signal at the UE_1 is decoded without employing SIC due to its highest received power, the interference from the UE_2 cannot be removed, while SIC is applied at the UE_2 successively to remove the interference from the UE_1. Thus, the data rates of the users are defined as;

$$R_1 = B\log_2\left(1+\frac{P_1\left|\mathbf{w}^H\mathbf{h}_1\right|^2}{P_2\left|\mathbf{w}^H\mathbf{h}_2\right|^2+\sigma_n^2}\right)$$
$$R_2 = B\log_2\left(1+\frac{P_2\left|\mathbf{w}^H\mathbf{h}_2\right|^2}{\sigma_n^2}\right) \tag{2.11}$$

In general, when we consider uplink NOMA cluster, the SINR of u^{th} user in m^{th} cluster is calculated by;

$$\gamma_{u,m} = \frac{P_{u,m}\left|\mathbf{w}_m^H\mathbf{h}_{u,m}\right|^2}{\underbrace{\sum_{j=u+1}^{U}P_{j,m}\left|\mathbf{w}_m^H\mathbf{h}_{j,m}\right|^2}_{\text{Inter-user interference}}+\underbrace{\sum_{i=1,i\neq m}^{M}\left(\sum_{j=1}^{U}P_{j,i}\left|\mathbf{w}_m^H\mathbf{h}_{j,i}\right|^2\right)}_{\text{Inter-cluster interference}}+\sigma_n^2} \tag{2.12}$$

The average uplink sum data rate of the system with SIC can be expressed as follows;

$$R_{sum} = B\sum_{m=1}^{M}\sum_{u=1}^{U}\mathbb{E}\left\{\log_2\left(1+\gamma_{u,m}\right)\right\} \tag{2.13}$$

In this thesis, we study a synchronous UL-NOMA, which assumes perfectly time-synchronized users' signals are received at the BS.

2.4. User Selection Algorithms

The performance of multi-user MIMO systems depends on the user selection approach. In a dense system that requires to transmit a very large number of users, user

selection is necessary since the number of users that can be simultaneously supported can be limited by the number of transmit antennas at the BS. The instantaneous channels among the users can be non-orthogonal, which results in mutual inter-user interference. Thus, BS can increase sum data rate by selecting the best set of users to establish communication.

In the literature, user selection has been widely investigated. There are two main user selection algorithms, as the capacity-based greedy user selection and the semi orthogonal user group (SUS) selection given in (Yoo and Goldsmith, 2006), where iteratively selects the user considering the channel norm and the correlation coefficient. To ease the problems of high computational complexity and high feedback, the greedy user selection algorithm has been given in (Razi et al., 2010) based on the rate allocation in vector perturbation precoding systems, which reduces the computational complexity by removing the insignificant users from the candidate user set. Similarly, the authors in (Tran et al., 2012) have presented a greedy low-complexity scheduling algorithm for multi-user MIMO downlink channels, in which the product of squared row norms of the effective channels is used as the selection metric. In (Ko and Lee, 2012), a low-complexity scheduling algorithm using chordal distance as an orthogonality measure has been presented with the block diagonalization (BD) scheme to maximize the total throughput for a multi-user MIMO downlink system.

In (Xu et al., 2014), the joint antenna selection and user selection problem has been solved in distributed massive MIMO systems under the backhaul capacity constraint. In (Benmimoune et al., 2015), the joint strategy has been examined, which performs antenna selection and schedules the users to maximize the sum data rate. In (Li et al., 2018), the user selection scheme for a hybrid architecture based on discrete Fourier transform (DFT) processing has been examined by considering the achievable rate of the system and guaranteeing the fairness of selection for a massive MIMO multi-user system.

2.5. Low Complexity User Selection for Massive MIMO Systems

In this section, a user selection algorithm is proposed with the reconstruction of the sparse massive MIMO channel using the CS algorithm (Yılmaz and Özbek, 2020) by eliminating the users based on channel correlation and employing the CS algorithm to reduce the feedback load in the system.

To fully harvest the benefit of excessive BS antennas in massive MIMO systems, knowledge of channel state information at the transmitter (CSIT) is essential. However, it is challenging to obtain an accurate CSIT. Since the training overhead for CSIT acquisition grows proportionally with the number of BS antennas, it can be huge in such systems. The previous studies avoid this challenge by adopting time-division duplexing (TDD), where the CSIT can be obtained by exploiting channel reciprocity and the uplink pilot-aided training overhead is proportional to the number of users. However, channel reciprocity does not hold for massive MIMO systems with frequency-division duplexing (FDD). Pilot-based channel estimation and uplink channel feedback are required, which consume spectrum resources. Therefore, it is also essential to consider CSIT acquisition for FDD systems.

Many studies have shown that the effective dimension of a massive MIMO channel is actually much less than its original dimension because of the limited local scattering effect in the propagation environment (Dai et al., 2018). Specifically, the massive MIMO channel has an approximately sparse representation under the DFT basis when the BS is equipped with a ULA. As a consequence, CS algorithm, which exploits the hidden sparsity under the DFT basis, has been examined for downlink channel estimation and feedback (Lu et al., 2019). In addition to that, the performance of massive MIMO systems depends on the user selection approach. In this section, we present a user selection for massive MIMO systems with reduced feedback load. The computational complexity of the conventional user selection schemes is very high to be implemented in massive MIMO systems. Our objective is to improve the sum data rate and reduce the feedback load in massive MIMO systems through a user selection. In this algorithm, we eliminate the users based on the channel correlation, and OMP algorithm is performed to reduce the feedback load.

2.5.1. System Model

We consider the Massive MISO system model with N antennas at the BS and K single-antenna users, under the assumption of $N >> K >> 1$ as shown in Fig. 2.4. In a ULA antenna model, neighboring antennas are spaced by $D = \lambda/2$.

The received signal at the k^{th} UE is given by;

$$y_k = \mathbf{h}_k^H \mathbf{w}_k s_k + \sum_{i=1, i\neq k}^{K} \mathbf{h}_k^H \mathbf{w}_i s_i + n_k \tag{2.14}$$

where $\mathbf{h}_k \in \mathbb{C}^{Nx1}$ is the channel vector, $\mathbf{w}_k \in \mathbb{C}^{Nx1}$ is the k^{th} UE precoder, the transmitted symbol vector as $\mathbf{s} = [s_1, \ldots, s_K]^T \in \mathbb{C}^{Kx1}$, and n_k is the AWGN with zero mean and σ_n^2 variance, $\mathcal{CN}(0, \sigma_n^2)$.

The first term of the right side of (2.14) contains the desired signal for the k^{th} user, the second term represents the interference caused by the other users, and the last term is the noise.

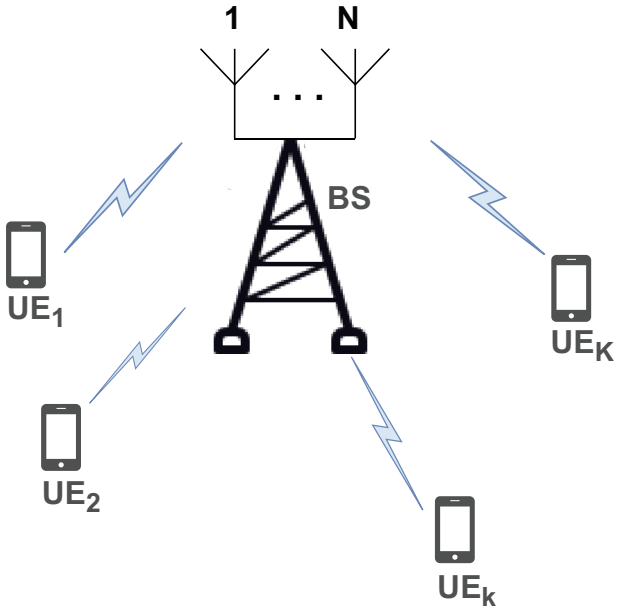


Figure 2.4. Massive MISO system model.

For 2D channel models, the channel vector $\mathbf{h}_k \in \mathbb{C}^{Nx1}$ for the k^{th} user due to the effect of clusters of scatterers is defined as given in (2.2). The total transmit power is equally shared among all users.

The received signal for all UEs is expressed as;

$$\mathbf{y} = \mathbf{H}^H \mathbf{W}\mathbf{s} + \mathbf{n}, \tag{2.15}$$

where the channel matrix is $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_K] \in \mathbb{C}^{NxK}$, $\mathbf{y} = [\mathbf{y}_1 \ldots \mathbf{y}_K]^T \in \mathbb{C}^{Kx1}$, and the precoder matrix is $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_K] \in \mathbb{C}^{NxK}$, which is determined with Zero-forcing (ZF) precoding as follows;

$$\mathbf{W} = \eta\, \mathbf{H} \left(\mathbf{H}^H \mathbf{H}\right)^{-1} \tag{2.16}$$

In order to keep the short-term power constant, the factor η is calculated as;

$$\eta = \frac{1}{\sqrt{tr\left(\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\right)}}. \tag{2.17}$$

Then, the SINR for k^{th} user is;

$$\gamma_k = \frac{\left|\mathbf{h}_k^H\mathbf{w}_k\right|^2}{\sum\limits_{j=1,j\neq k}^{K}\left|\mathbf{h}_k^H\mathbf{w}_j\right|^2 + (1/\rho)}, \forall k \in K \tag{2.18}$$

where ρ is the average signal-to-noise ratio (SNR).

The average sum data rate is calculated as;

$$R_{\text{M-MISO}} = B\sum_{k=1}^{K}\mathbb{E}\left\{\log_2\left(1+\gamma_k\right)\right\} \tag{2.19}$$

2.5.2. Proposed Algorithm

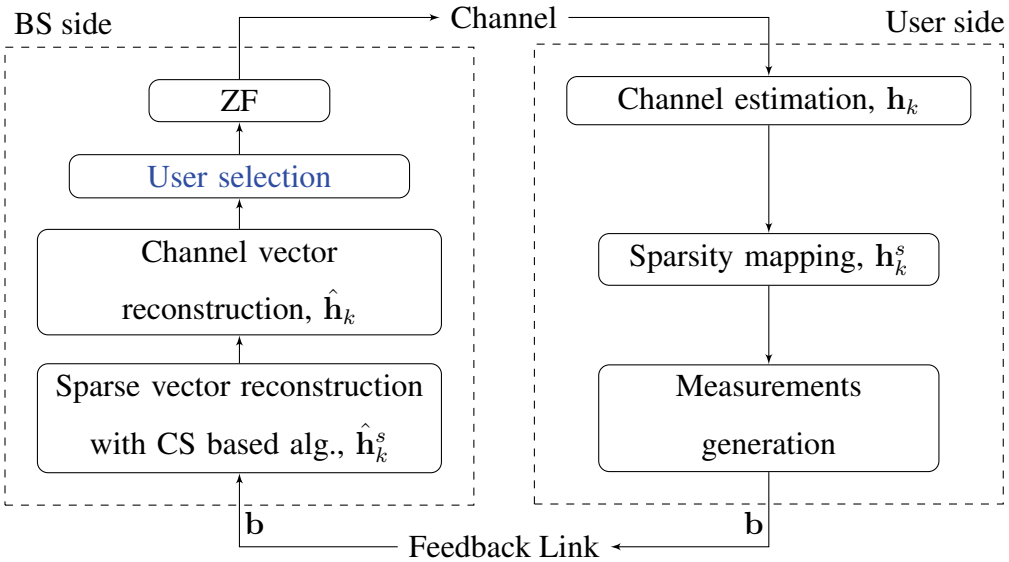


Figure 2.5. The proposed system model.

Fig. 2.5 presents the general system structure for the proposed system. The proposed user selection algorithm eliminates the users based on channel correlation by employing the CS algorithm, which reduces the feedback load in the system. The steps of the proposed user selection algorithm are explained in the following three subsections, as sparsity mapping, channel vector reconstruction and the proposed user selection method.

Sparsity Mapping: The basic procedure of the sparsity mapping at the user side is illustrated in Fig. 2.6. We assume that each user has perfect the downlink channel vector $\mathbf{h_k}$.

Due to the antenna correlation at the BS and limited local scattering effects, most of the multi-path energy for each user tends to be concentrated in a relatively small region within the virtual angular domain. Therefore, the channel is expected to have a sparse representation in the virtual angular domain, so only a small fraction of components is significant and the others are zero.

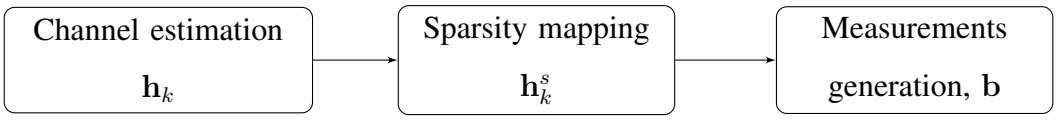


Figure 2.6. Sparsity mapping at the user side.

The channel vector $\mathbf{h_k}$ is represented using virtual channel representation with a proper basis;

$$\mathbf{h}_k = \mathbf{U}^H \mathbf{h}_k^s \tag{2.20}$$

where $\mathbf{h}_k^s \in \mathbb{C}^{Nx1}$ is the channel representation in the virtual angular domain, and $\mathbf{U}$ is DFT matrix with $\mathbb{C}^{NxN}$.

The n^{th} column of $\mathbf{U}$ is given by;

$$\mathbf{u}_n \triangleq \frac{1}{\sqrt{N}} \left[1 \dots e^{-j\frac{2\pi(m-1)n}{N}} \dots e^{-j\frac{2\pi(N-1)n}{N}}\right]^T \tag{2.21}$$

for $m, n = 1, 2, \dots, N$.

After mapping $\mathbf{h}_k$ to a sparse channel vector $\mathbf{h}_k^s$, the measurement matrix is used to reduce the number of samples as the channel vector. For $\mathbf{h}_k^s$ and m feedback measurements (FM), the random measurement vector $\mathbf{b} \in \mathbb{C}^{mx1}$ is generated as follows;

$$\mathbf{b} = \mathbf{A}^T \mathbf{h}_k^s, \tag{2.22}$$

where $\mathbf{A} \in \mathbb{C}^{Nxm}$ is a measurement matrix, which is generated off-line and known at both the user and the BS sides. $\mathbf{A}$ is sampled from i.i.d. Gaussian distributed entries with zero mean and $1/m$ variance.

After the measurement vector, $\mathbf{b}$, is generated, it is fed back to the BS perfectly to perform downlink precoding.

Channel Vector Reconstruction: The basic procedures of the applying CS algorithm and the proposed user selection at the BS side are illustrated in Fig. 2.7.

To fully utilize the spatial multiplexing and the array gain of massive MIMO, the CSIT is essential. However, it is inefficient to estimate the entire CSI using long pilot training symbols at the BS. We should exploit the hidden sparsity in the CSIT estimation and feedback process, where CS based algorithm is the efficient reconstruction of a sparse signal from a few samples.

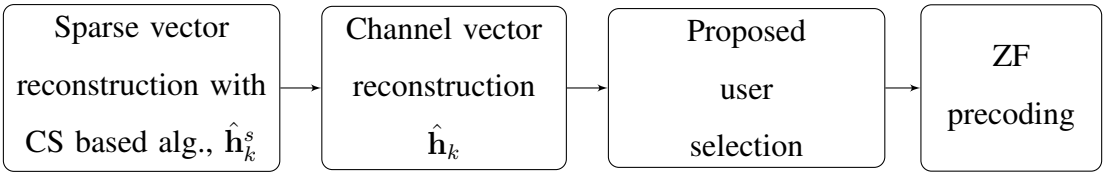


Figure 2.7. Channel vector reconstruction at the BS.

At the BS, as the CS-based algorithm, OMP is used for reconstructing the sparse channel vector $\hat{\mathbf{h}}_k^s$. In the OMP, since all columns of $\mathbf{A}$ are correlated with the $\mathbf{b}$, the sparse signal is reconstructed iteratively. In each iteration, the algorithm finds the column of $\mathbf{A}$ which is most correlated to the $\mathbf{b}$ and adds its index. The stopping criterion is based on the sparsity level (SL) value.

The reconstructed channel state information (CSI), $\hat{\mathbf{h}}_k$, is obtained by mapping back the reconstructed sparse channel vector $\hat{\mathbf{h}}_k^s$ based on the same basis used at the user side. After that, the proposed user selection algorithm is applied to the reconstructed CSI. The detail of the proposed algorithm is given in Algorithm 1.

Finally, the BS performs the ZF precoding to eliminate inter-user interference and achieve the sum data rate based on the reconstructed channel matrix $\hat{\mathbf{H}}$ instead of $\mathbf{H}$ in (2.16).

Proposed User Selection Method: The main objective of the proposed algorithm is to improve the sum data rate and reduce the complexity. The steps of the proposed user selection algorithm are given in Algorithm 1. Its complexity is determined to be upper-bounded by $C_{ip}\frac{K(K-1)}{2}$, where C_{ip} is the complexity of the normalized inner-product for each UE pair in Step 2.

Algorithm 1 Proposed User Selection Method

Step 1: Initialization:

$$S_0 = \{1, \ldots, K\}, \quad \mathcal{S} = \emptyset \tag{2.23}$$

Step 2: Determine the correlation coefficient $\beta_{k,j}$ between all UE pairs $j \neq k$:

$$\beta_{k,j} = \frac{\left|\hat{\mathbf{h}}_k^H \hat{\mathbf{h}}_\mathbf{j}\right|}{\left\|\hat{\mathbf{h}}_k\right\| \left\|\hat{\mathbf{h}}_j\right\|} \tag{2.24}$$

Note that $\beta_{k,j} = \beta_{j,k}$.

Step 3: Select the UE to be eliminated according to the degree of orthogonality.

- For k^{th} UE, define $\mathcal{US}_k$ set that holds the users whose the correlation coefficient higher than β.

$$\mathcal{US}_k = \{j \in S_0 : \beta_{k,j} > \beta, \forall k \in S_0\} \tag{2.25}$$

where β is a small positive constant value. It characterizes the allowed degree of orthogonality between two channel vectors.

Step 4: For k^{th} UE, if the number of elements in $\mathcal{US}_k$ is 0, add this k^{th} UE to the selected user set $\mathcal{S}$ is:

$$\mathcal{S} = \mathcal{S} \cup \{k\}, \text{ if Card } (US_k) = 0. \tag{2.26}$$

- The precoding matrix and the sum data rate are calculated based on the set of selected UEs, $\mathcal{S}$.

2.5.3. Performance Evaluations

In the considered Massive MISO system, there is one BS employed between N=48 and N=96 antennas, and K=16 users with single-antenna. For higher number of antennas, the channel hardening occurs, and scheduling all UEs is optimal, which means that more advanced scheduling does not provide gains (Dierks et al., 2015) (Björnson et al., 2017).

There are only NLOS components in the system, where the standard deviation, ASD, is determined as $\sigma_\phi = 10^\circ$ (Dierks and Juenger, 2016). The SNR value, ρ, is defined as 15 dB, where the path loss coefficients for all users are assumed to be the same (Hawej and Shayan, 2018). The optimal β is determined through the numerical simulations. We consider different feedback measurement values FM, and sparsity levels, SL, while providing the performance results.

We examine the sum data rate versus the number of antennas at the BS and then compare the performance of the proposed algorithm with the pair-wise SUS algorithm in (Dierks and Juenger, 2016), labeled as "*pair-wise SUS*", and all UEs selected case. Specifically, the pair-wise SUS algorithm first schedules all users. Then, at each iteration, the scheduling algorithm finds the user pair with the largest correlation coefficient. From this pair, the user with the smaller channel gain is removed. This continues until the correlation coefficient between the remaining users are small enough. Moreover, the complexity of the pair-wise SUS algorithm is determined to be upper-bounded by $C_{ip}\frac{K(K-1)}{2}+C_{vn}K$, where C_{vn} is the complexity of one vector 2-norm.

The channel reconstruction performance is measured in terms of the mean square error (MSE) based on the reconstructed channel matrix $\hat{\mathbf{H}}$ and the actual channel matrix $\mathbf{H}$ as follows;

$$MSE = \frac{\mathbb{E}\left\{\left\|\hat{\mathbf{H}}-\mathbf{H}\right\|_F^2\right\}}{N\,K} \tag{2.27}$$

where $\|.\|_F$ represents the Frobenius norm.

Fig. 2.8 and Fig. 2.9 show the MSE performance of the proposed algorithm to indicate the effect of different FM and SL values on the quality of the channel reconstruction.

Fig. 2.8 shows the MSE performance of the proposed algorithm for SL=N$/4$ and different values of FM. The numerical results show that when the FM is increased, the

MSE decreases. The reason is the increasing the number of measurements that send to the BS side via the feedback link.

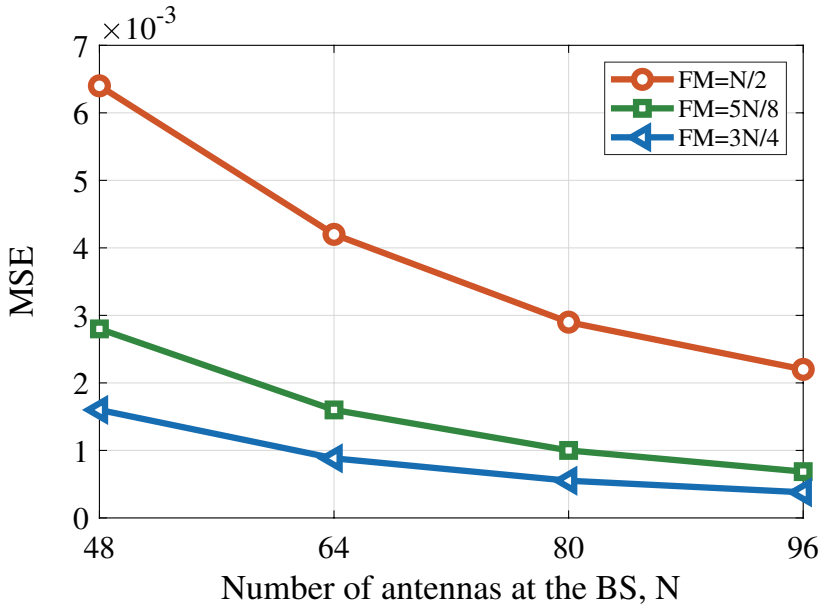


Figure 2.8. MSE results for different FM in the proposed algorithm for K=16 and SL=N/4.

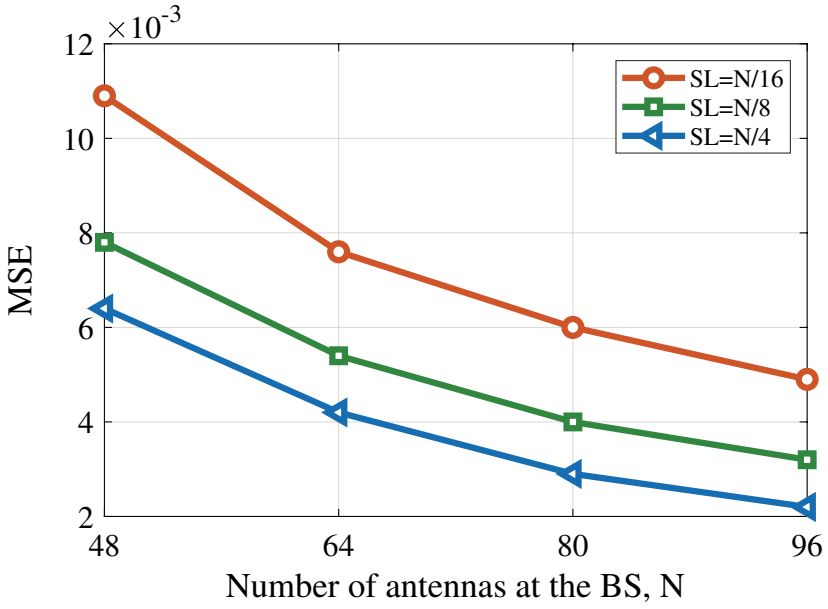


Figure 2.9. MSE results for different SL in the proposed algorithm for K=16 and FM=N/2.

Fig. 2.9 shows the MSE performance of the proposed algorithm for FM=N/2 and different values of SL. It is seen that higher sparsity level decreases the MSE performance. This is because the CS algorithm provides better performance to reconstruct the sparse signal with higher sparsity levels which indicate the number of the non-zero coefficients of the original signal.

From Fig. 2.8 and Fig. 2.9, we conclude that FM=3N/4 and SL=N/4 give the minimum MSE value, thus improving the channel reconstruction in the proposed system.

Fig. 2.10 presents the sum data rate performance of the proposed algorithm as a function of different β values for FM=3N/4 and SL=N/4. For any number of antennas, the results show that $\beta = 0.7$ has the highest sum data rate since it allows to eliminate correlated UEs. Thus, we fix $\beta = 0.7$ for the rest of the results.

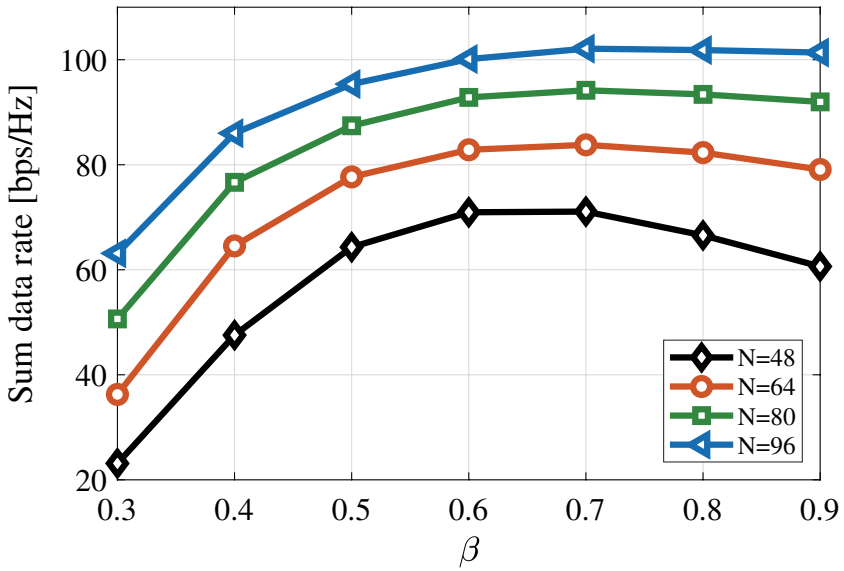


Figure 2.10. Sum data rate results for different β in the proposed algorithm for K=16, FM=3N/4 and SL=N/4.

Fig. 2.11 illustrates the average data rate per user of all user selection algorithms when FM=3N/4 and SL=N/4, and $\beta = 0.7$ for the pair-wise SUS algorithm and the proposed algorithm to compare fairly. It is worth noting that the proposed algorithm provides the best performance in terms of average data rate per user compared to the algorithms of all UEs selected and pair-wise SUS. It is illustrated that applying user selection is more advantageous for the low number of BS antennas since the variations of the channel gain are high. When the number of antennas increases, the average data rate per user

of the pair-wise SUS algorithm and the proposed algorithm close to each other due to the decreasing degrees of orthogonality. Specifically, the proposed algorithm achieves a 9% and 2% higher sum data rate compared to the pair-wise SUS algorithm for N=48 and N=96, respectively. For N=48, the proposed algorithm schedules 13 users while the pair-wise SUS algorithm schedules 14 users. This indicates that the proposed algorithm eliminates more users than the pair-wise SUS algorithm. Furthermore, the proposed algorithm reduces complexity compared to the pair-wise SUS algorithm since the pair-wise SUS algorithm has the factor of $C_{vn}K$, which depends linearly on the number of users in the system.

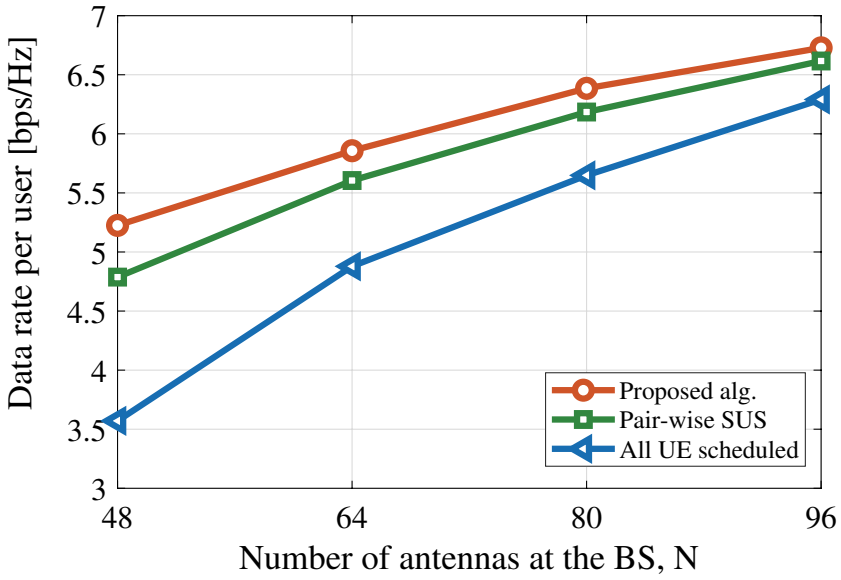


Figure 2.11. Comparison of the user selection algorithms for K=16, β=0.7, FM=3N/4 and SL=N/4.

2.6. User Selection for NOMA based MIMO Systems

In this section, we propose a user-set selection algorithm in a densely deployed environment for an uplink NOMA based MIMO system to improve the sum data rate (Yılmaz et al., 2022). The simulation results verify the advantage of the proposed user-set selection with power allocation over conventional OMA systems.

2.6.1. System Model

We consider an uplink MIMO-NOMA communication in which a BS has N antennas and is serving K devices with a single-antenna, where K≥2N as shown in Fig. 2.12. These devices can be smartphones, laptops, tablets and IoT devices.

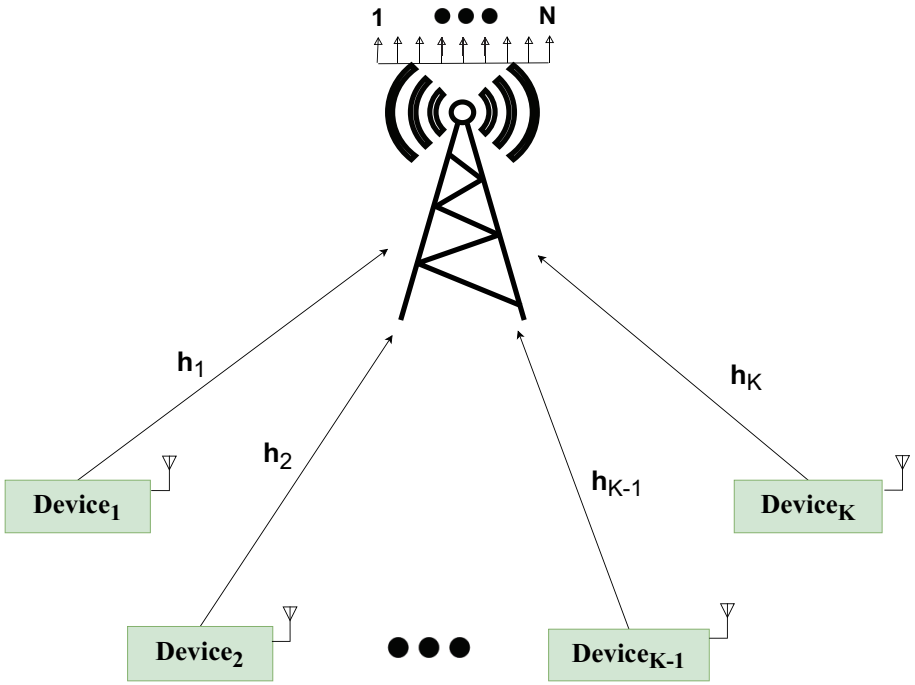


Figure 2.12. Uplink NOMA based MIMO system model.

The BS with N antennas can serve up to N devices with traditional OMA systems. On the other hand, in the uplink NOMA system, the BS simultaneously supports two user sets. As a result, the total number of devices supported by the BS is equal to 2N (Kim et al., 2015). Thus, the selection of 2N devices among K total devices is crucial and is proposed in Section 2.6.2.

There are two user sets as strong and weak where each of them includes N devices. We consider the two-users NOMA with a total of M clusters, including one strong and one weak user for each cluster, where the number of clusters is equal to the number of antennas at the BS, M=N and m $\in \{1, \ldots, \mathrm{M}\}$. The strong users set represents the devices that have high channel gains, whereas the devices in the weak users set have relatively low channel gains. Accordingly, the uplink channel vectors belonging to each set are denoted

as $\mathbf{h}_{m,j}$ between the BS and the device in the corresponding set of $j = 1, 2$ for each m^{th} cluster and determined as in (2.1). The channel matrices belonging to the strong users set, $\mathbf{H}_1$, and weak users set, $\mathbf{H}_2$, in all clusters are given by;

$$\mathbf{H}_1 = [\mathbf{h}_{1,1} \dots \mathbf{h}_{m,1} \dots \mathbf{h}_{M,1}] \tag{2.28}$$

$$\mathbf{H}_2 = [\mathbf{h}_{1,2} \dots \mathbf{h}_{m,2} \dots \mathbf{h}_{M,2}] \tag{2.29}$$

where the overall $N \times M$ channel matrices $\mathbf{H}_j$, whose element $\mathbf{h}_{m,j}$ with size $N \times 1$ uplink channel vector, can be given as $\mathbf{H}_j = \mathbf{G}_j \, \mathbf{D}_j^{1/2}$ (Khansefid and Minn, 2014) (Zeng et al., 2020a). Here, $\mathbf{D}_j = \text{diag}\{L_{1,j}, L_{2,j}, \dots, L_{M,j}\} \in \mathbb{R}^{M \times M}$ is a diagonal matrix and represents the large-scale fading component including path loss. $\mathbf{G}_j$ is a $N \times M$ matrix including $N \times 1$ vector of $\mathbf{g}_{m,j}$ belonging to each device in the j^{th} set and m^{th} cluster. Thus, we have $\mathbf{h}_{m,j} = \sqrt{L_{m,j}}\, \mathbf{g}_{m,j}$ between the BS and a device of $m = 1, 2, ..., M$ for the set of $j = 1, 2$.

The received signal at the BS is the summation of the signals coming from these two sets as;

$$\mathbf{y} = \sum_{m=1}^{M} \sum_{j=1}^{2} \mathbf{h}_{m,j} \sqrt{\alpha_{m,j}} s_{m,j} + \mathbf{n}, \tag{2.30}$$

where $\alpha_{m,j}$ is the power allocation factor of the device in the j^{th} set and m^{th} cluster within the range $0 < \alpha_{m,j} \leq 1$. The uplink symbol of the device in the j^{th} set and m^{th} cluster is given by $s_{m,j}$ having $\mathbb{E}[|s_{m,j}|^2] \leq P$, where P is the maximum transmit power per device. The AWGN vector of size $N \times 1$ is given by $\mathbf{n}$ and each element of the vector is modeled by a Gaussian distribution random variable with zero mean and σ_n^2 variance.

We can re-write (2.30) as;

$$\mathbf{y} = (\mathbf{H}_1 \mathbf{s}_1 + \mathbf{H}_2 \mathbf{s}_2) + \mathbf{n}, \tag{2.31}$$

where $M \times 1$ transmitted signal vectors for the strong, $\mathbf{s}_1$, and weak, $\mathbf{s}_2$, users set are given respectively as;

$$\mathbf{s}_1 = \left[\sqrt{\alpha_{1,1}}s_{1,1} \dots \sqrt{\alpha_{m,1}}s_{m,1} \dots \sqrt{\alpha_{M,1}}s_{M,1}\right]^T,$$

$$\mathbf{s}_2 = \left[\sqrt{\alpha_{1,2}}s_{1,2} \dots \sqrt{\alpha_{m,2}}s_{m,2} \dots \sqrt{\alpha_{M,2}}s_{M,2}\right]^T.$$

According to the received signal in (2.31), both symbols belonging to the strong and weak users are extracted at the BS with SIC decoding. Firstly, the signals of the devices in the strong users set are decoded under the inter-set interference caused by the devices in the weak users set. Then, the SIC is applied to decode the signals in the weak users set by subtracting the signals of strong users from the received signal. Thus, the signals in the weak users set are decoded without the inter-set interference.

In order to cancel the intra-set interference, the ZF postcoding technique is employed at the BS. It is assumed that the BS can have the perfect CSI belonging to all devices. The ZF postcoding matrix $\mathbf{W}_j^{'}$ with $j = 1, 2$ is determined through overall channel matrix by;

$$\mathbf{W}_j^{'} = \mathbf{H}_j^H \left(\mathbf{H}_j \mathbf{H}_j^H\right)^{-1}, \tag{2.32}$$

where the normalized ZF postcoding matrix is given by;

$$\mathbf{W}_j = \left[{\mathbf{w}_{1,j}}^T \dots {\mathbf{w}_{m,j}}^T \dots {\mathbf{w}_{M,j}}^T\right]^T, \tag{2.33}$$

with $\mathbf{w}_{m,j}$ is the ZF postcoder vector with the length of $1 \times N$ for the device in the j^{th} set and m^{th} cluster. It is determined by; $\mathbf{w}_{m,j} = \frac{\mathbf{w}_{m,j}^{'}}{\left\|\mathbf{w}_{m,j}^{'}\right\|}$ where $\mathbf{w}_{m,j}^{'}$ is the m^{th} row of $\mathbf{W}_j^{'}$.

Firstly, the signals of the strong devices are decoded under inter-set interference by using the postcoding matrix $\mathbf{W}_1$. Accordingly, the received signal vector of the strong set, $\mathbf{r}_1 = [r_{1,1} \ \dots r_{m,1} \dots r_{M,1}]^T$, can be expressed as;

$$\mathbf{r}_1 = \mathbf{W}_1\mathbf{y} = \mathbf{W}_1\mathbf{H}_1\mathbf{s}_1 + \mathbf{W}_1\mathbf{H}_2\mathbf{s}_2 + \mathbf{W}_1\mathbf{n}. \tag{2.34}$$

From (2.34), the received signal of the strong user in the m^{th} cluster is given by;

$$r_{m,1} = \mathbf{w}_{m,1}\mathbf{h}_{m,1}\sqrt{\alpha_{m,1}}s_{m,1} + \sum_{i=1}^{M} \mathbf{w}_{m,1}\mathbf{h}_{i,2}\sqrt{\alpha_{i,2}}s_{i,2} + \sum_{j=1, j\neq m}^{M} \mathbf{w}_{m,1}\mathbf{h}_{j,1}\sqrt{\alpha_{j,1}}s_{j,1} + \mathbf{w}_{m,1}\mathbf{n}, \tag{2.35}$$

where the first term represents the desired signals of the strong users set and the second term is the inter-set interference caused by the weak users set.

For the strong user in the m^{th} cluster, the instantaneous SINR is given by;

$$\text{SINR}_{m,1} = \frac{\alpha_{m,1} P \left|\mathbf{w}_{m,1}\mathbf{h}_{m,1}\right|^2}{\sum\limits_{i=1}^{M} \alpha_{i,2} P \left|\mathbf{w}_{m,1}\mathbf{h}_{i,2}\right|^2 + \sum\limits_{j=1, j\neq m}^{M} \alpha_{j,1} P \left|\mathbf{w}_{m,1}\mathbf{h}_{j,1}\right|^2 + \left\|\mathbf{w}_{m,1}\right\|^2 \sigma_n^2}. \tag{2.36}$$

For decoding of the signals of weak users set, SIC is applied and, in this case, there will be no inter-set interference. Thus, after utilizing ZF postcoding matrix $\mathbf{W}_2$, the received signal vector of the weak set, $\mathbf{r}_2 = \left[r_{1,2} \ \ldots r_{m,2} \ldots r_{M,2}\right]^T$, is expressed as;

$$\mathbf{r}_2 = \mathbf{W}_2\mathbf{H}_2\mathbf{s}_2 + \mathbf{W}_2\mathbf{n}. \tag{2.37}$$

From (2.37), the received signal of the weak user in the m^{th} cluster is given by;

$$r_{m,2} = \mathbf{w}_{m,2}\mathbf{h}_{m,2}\sqrt{\alpha_{m,2}}s_{m,2} + \left(\sum_{j=1, j\neq m}^{M} \mathbf{w}_{m,2}\mathbf{h}_{j,2}\sqrt{\alpha_{j,2}}s_{j,2}\right) + \mathbf{w}_{m,2}\mathbf{n}. \tag{2.38}$$

For the weak user in the m^{th} cluster, the instantaneous SINR is defined by;

$$\text{SINR}_{m,2} = \frac{\alpha_{m,2} P \left|\mathbf{w}_{m,2}\mathbf{h}_{m,2}\right|^2}{\sum\limits_{j=1, j\neq m}^{M} \alpha_{j,2} P \left|\mathbf{w}_{m,2}\mathbf{h}_{j,2}\right|^2 + \left\|\mathbf{w}_{m,2}\right\|^2 \sigma_n^2}. \tag{2.39}$$

The average data rate of the user in the j^{th} set and the m^{th} cluster is given by;

$$R_{m,j} = B\ \mathbb{E}\left\{\log_2\left(1 + \text{SINR}_{m,j}\right)\right\}. \tag{2.40}$$

The sum data rate of M devices for each user set, j=1,2, is given by;

$$R_{j,\text{NOMA}} = \sum_{m=1}^{M} R_{m,j}. \tag{2.41}$$

Thus, the overall sum data rate in NOMA system is determined by;

$$R_{\text{sum}}^{\text{NOMA}} = \sum_{j=1}^{2} R_{j,\text{NOMA}}. \tag{2.42}$$

For the purpose of comparing, the average data rates of devices are given in the corresponding sets for conventional OMA system as;

$$R_{j,\text{OMA}} = \frac{B}{2} \sum_{m=1}^{M} \mathbb{E}\left\{ \log_2 \left(1 + \frac{P \left| \mathbf{w}_{m,j} \mathbf{h}_{m,j} \right|^2}{\left\| \mathbf{w}_{m,j} \right\|^2 \sigma_n^{\,2}} \right) \right\}, \tag{2.43}$$

where $\frac{1}{2}$ is added since two time slots are required to support $2M$ devices in the OMA system with N antennas, whereas NOMA can support $2M$ devices during a single time slot. Therefore, the overall sum data rate for OMA is given by;

$$R_{\text{sum}}^{\text{OMA}} = \sum_{j=1}^{2} R_{j,\text{OMA}}. \tag{2.44}$$

Here, the power allocation factors of the strong user set are selected as equal to each other;

$$\alpha_{1,1} = \ldots = \alpha_{m,1} = \ldots = \alpha_{M,1} = \alpha_1 \tag{2.45}$$

Similarly, the power allocation factors of the weak user set are selected as equal to each other;

$$\alpha_{1,2} = \ldots = \alpha_{m,2} = \ldots = \alpha_{M,2} = \alpha_2 \tag{2.46}$$

2.6.2. User-Set Selection for NOMA based MIMO Systems

In this section, we propose a user-set selection algorithm to improve the performance of the uplink MIMO-NOMA system.

As illustrated in Fig. 2.13, the proposed user-set selection algorithm selects totally $2M$ devices without affecting the transmission of strong users. We firstly determine the strong and weak user sets considering their channel gains. Then, we select M devices from each set based on both the channel gain and the correlation properties to mitigate inter-set interference. After that, we determine the optimal power values α_1 and α_2 for both strong and weak user sets, respectively.

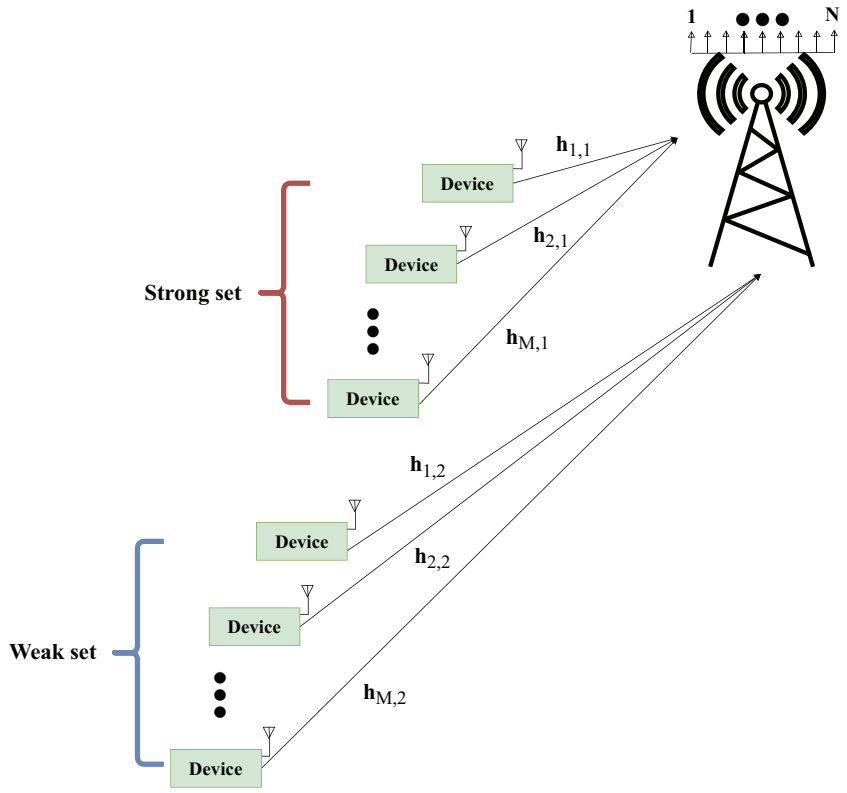


Figure 2.13. Proposed user set selection for uplink NOMA based MIMO.

A total of M devices having higher channel gains in the strong users set are selected. The devices in the weak users set cause inter-set interference to devices in the strong users set. Therefore, the proposed user-set selection algorithm determines weak users through an elimination procedure based on the channel correlation between strong and weak user sets. In order to select M devices for weak users set, the BS compares their orthogonality with the selected devices in strong users set. Details of the proposed algorithm are explained in Algorithm 2.

Algorithm 2 Proposed User-Set Selection Method

Step 1: Channel gains of K devices are sorted in descending order:

$$\mathcal{S}_{\text{ord}} = \{\|\mathbf{h}_k\| \mid \|\mathbf{h}_k\| > \|\mathbf{h}_{k+1}\|; k \in \{1, 2, \ldots, K\}\}. \tag{2.47}$$

Step 2: Utilizing $\mathcal{S}_{\text{ord}}$, BS classifies the K devices into two groups. The set of devices with the higher channel gains is represented by $\mathcal{S}_{\text{s}}$, whereas $\mathcal{S}_{\text{w}}$ includes the set of devices with the lower channel gains.

$$\mathcal{S}_s = \left\{\|\mathbf{h}_1\|, \ldots, \left\|\mathbf{h}_{\lfloor K/2 \rfloor}\right\|\right\}, \tag{2.48}$$

in here $\lfloor . \rfloor$ indicates the floor function.

$$\mathcal{S}_w = \left\{\left\|\mathbf{h}_{\lfloor K/2 \rfloor + 1}\right\|, \ldots, \|\mathbf{h}_K\|\right\}. \tag{2.49}$$

Step 3: From $\mathcal{S}_{\text{s}}$, strong users set is formed by selecting the first M devices,

$$\mathbf{H}_1 = \{\mathbf{h}_m \mid \mathbf{h}_m \in \mathcal{S}_s, m = 1, \ldots, M\}. \tag{2.50}$$

Step 4: BS determines the correlation coefficient $\beta_{m,i}$ between device m in strong users set and the device i in set of $\mathcal{S}_{\text{w}}$:

$$\beta_{m,i} = \frac{\left|\mathbf{h}_m^H \mathbf{h}_i\right|}{\|\mathbf{h}_m\| \, \|\mathbf{h}_i\|}, \tag{2.51}$$

where $m = 1, \ldots, M$ and $i = \lfloor K/2 \rfloor + 1, \ldots, K$. When $\beta_{m,i}$ is equal to zero, two channel vectors are orthogonal to each other. When this value is closer to 1, two channel vectors are in similar directions and then they are highly correlated.

Step 5: As in (Yılmaz and Özbek, 2020), a set of $\mathcal{US}_i$ that satisfy the following criterion for i^{th} device in $\mathcal{S}_w$ is constructed:

$$\mathcal{US}_i = \{i \in \mathcal{S}_w \mid \beta_{m,i} > \beta; \forall m\}, \tag{2.52}$$

where β is a fixed threshold holding a value between 0 and 1. As a result, $\mathcal{US}_i$ includes i^{th} weak device that is correlated to the selected strong users.

Step 6: For the i^{th} device, if the set of $\mathcal{US}_i$ is not empty, i^{th} device is removed from the set of $\mathcal{S}_{\text{w}}$ by,

$$\mathcal{S}_w = \mathcal{S}_w \setminus \{\mathbf{h}_i\}\,, \text{ if Card } (\mathcal{US}_i) \neq 0. \tag{2.53}$$

- $\mathcal{S}_w$ includes the set of weak devices that are orthogonal to the selected strong users.

Step 7a: If the cardinality of $\mathcal{S}_w$ is equal or higher than M, the proposed user set selection algorithm selects the devices in the weak users set as follows:

$$\mathbf{H}_2 = \{\mathbf{h}_k \mid \mathbf{h}_k \in \mathcal{S}_w, k = 1, \ldots, M\}\,. \tag{2.54}$$

- The NOMA based MIMO transmission is performed for the selected devices in (2.50) and (2.54).

Step 7b: If the cardinality of $\mathcal{S}_w$ is less than M, the devices in the weak users set are selected from the set of $\mathcal{S}_{\text{ord}}$ by,

$$\mathbf{H}_2 = \left\{\mathbf{h}_i \mid \mathbf{h}_i \in \mathcal{S}_{\text{ord}}, i = \left\lfloor \frac{K}{2} \right\rfloor + 1, \ldots, \left\lfloor \frac{K}{2} \right\rfloor + M \right\}. \tag{2.55}$$

- The OMA-based MIMO transmission is performed for the selected devices in (2.50) and (2.55).

Power Allocation Scheme

We examine an optimal power allocation to find the coefficients of α_1 and α_2. Our objective function is to maximize the sum data rate of devices in NOMA based MIMO system, defined by;

$$f(\boldsymbol{\alpha}) = R_{1,\text{NOMA}} + R_{2,\text{NOMA}} \tag{2.56}$$

where $\boldsymbol{\alpha} = [\alpha_1, \alpha_2]$.

The sum data rate maximization problem subject to the data rate constraint based on OMA-based MIMO system is given by;

$$\max_{\boldsymbol{\alpha}} \quad f(\boldsymbol{\alpha}) \tag{2.57}$$

$$\text{s.t.} \quad R_{1,\text{NOMA}} \geq R_{1,\text{OMA}}, \tag{2.57a}$$

$$R_{2,\text{NOMA}} \geq R_{2,\text{OMA}}, \tag{2.57b}$$

$$0 < \alpha_1, \alpha_2 \leq 1. \tag{2.57c}$$

Constraints (2.57a) and (2.57b) give the minimal data rate constraints in which the achievable data rate in uplink NOMA should be no less than a conventional OMA system. Constraint (2.57c) gives the interval of the α_1 and α_2.

The maximization function in (2.57) is equivalent to minimizing -f(.), thus the corresponding optimization problem can be solved efficiently with the standard nonlinear programming optimization tools (Grace, 1990). The minimum of a constrained nonlinear multivariate function can be found using the interior-point method. As a result of the interior-point method, the optimum output values are obtained as α_1^* and α_2^* to provide the maximum sum data rate under the given constraints.

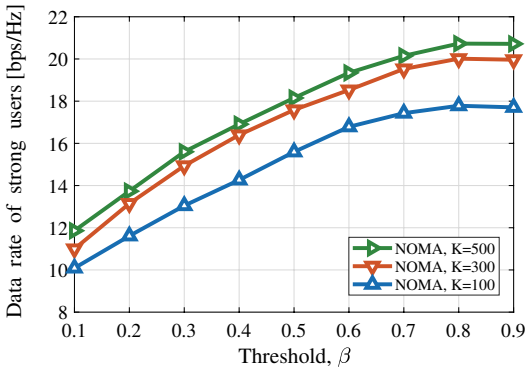
2.6.3. Performance Evaluations

This subsection presents the simulation results to evaluate the performances of NOMA based MIMO system. The devices are uniformly placed in the cell area which is placed between 30 m and 250 m from the BS. The noise power spectrum density is -159 dBm/Hz, and the maximum transmit power and the bandwidth of each device are set to 11 dBm and 200 kHz, respectively (Abozariba et al., 2019). The path loss is determined by $L = 128.1 + 37.6 \log_{10}(d_{m,j}\,(km))$ (Dai and Lyu, 2020) where $d_{m,j}$ is the distance between the device m in the set j and the BS. In the OMA-based multi-user MIMO, the strong devices are selected as in (2.50) while the weak devices are selected as either in (2.54) or (2.55). After that, these selected devices establish transmission through two time slots in OMA based systems whereas only one time slot is allocated for the NOMA based multi-user MIMO.

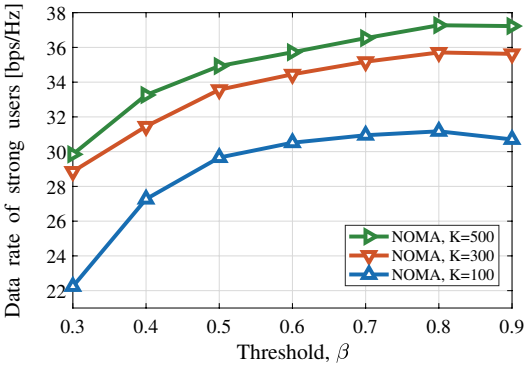
We compare the performance of the proposed user set selection algorithm for multi-user MIMO systems based on NOMA and OMA. The data rate of the strong users is given for the proposed user set selection with NOMA based system and the sorting based user set selection with NOMA. The effect of the number of devices, K, and the effect of the threshold, β, on the sum data rate and on the data rate of the strong users are provided. For all schemes, we determine the value of β up to 0.9 in order to implement the NOMA system within the proper restriction.

In Fig. 2.14, the effect of the orthogonality threshold is investigated for N=2, N=4 and N=8 regarding to the uplink NOMA based MIMO system. To find the optimum threshold value, β, we consider all N values individually, since when we increase the N value, OMA-based MIMO transmission is performed for the lower value of β according

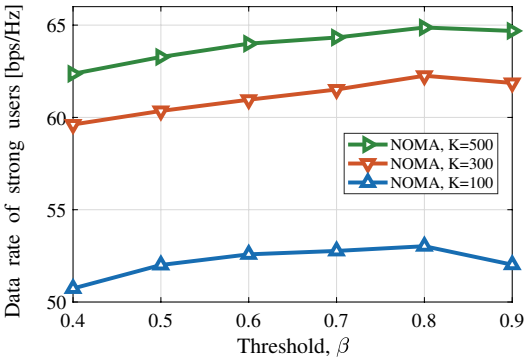
to the proposed user-set selection method given in Algorithm 2. The data rate of the strong users for different threshold values is examined since the threshold value directly affects the performance of the strong users set. It is observed that $\beta = 0.8$ provides the highest data rate of the strong users for all the case N and different K values, and the weak devices are selected as much as orthogonal to the strong devices. Then, the value of β continues to increase, the data rate of the strong users decreases due to the increasing inter-set interference with the selection of nearly non-orthogonal weak users. The optimized β value, which is 0.8, is kept for each N for the remaining part of the simulation results. Moreover, it is observed that selecting $2M$ users among a higher number of users, K, improves the data rate.



(a) N = 2.



(b) N = 4.



(c) N = 8.

Figure 2.14. Data rate of strong users for the proposed NOMA systems versus the threshold β, for different K and N.

Fig. 2.15 shows the sum data rate of the proposed user set selection through NOMA and OMA for all N with $\beta = 0.8$ in the densely deployed devices. For the different number of devices and antennas at BS, it is observed that the NOMA based MIMO system has superior performance than OMA based MIMO systems through the proposed user set selection algorithm. Specifically, for N=2 and K=100, the proposed user set selection with NOMA achieves approximately 23 bps/Hz, whereas the proposed user set selection with OMA provides approximately 18 bps/Hz. For the case N=4 and K=100, the proposed user set selection with NOMA achieves approximately 42 bps/Hz, while the proposed user set selection with OMA attains approximately 34 bps/Hz. Similarly, for the case N=8 and K=100, the proposed user set selection with NOMA achieves approximately 74 bps/Hz, while the proposed user set selection with OMA attains approximately 64 bps/Hz. When the K value is increased from 100 to 500, the proposed user set selection through NOMA achieves a higher sum data rate for both the systems with N=4 and N=8 as 15%, and 13% for the system with N=2. Furthermore, the performance results show that increasing number of BS antennas improves the sum data rate for both NOMA and OMA based MIMO systems.

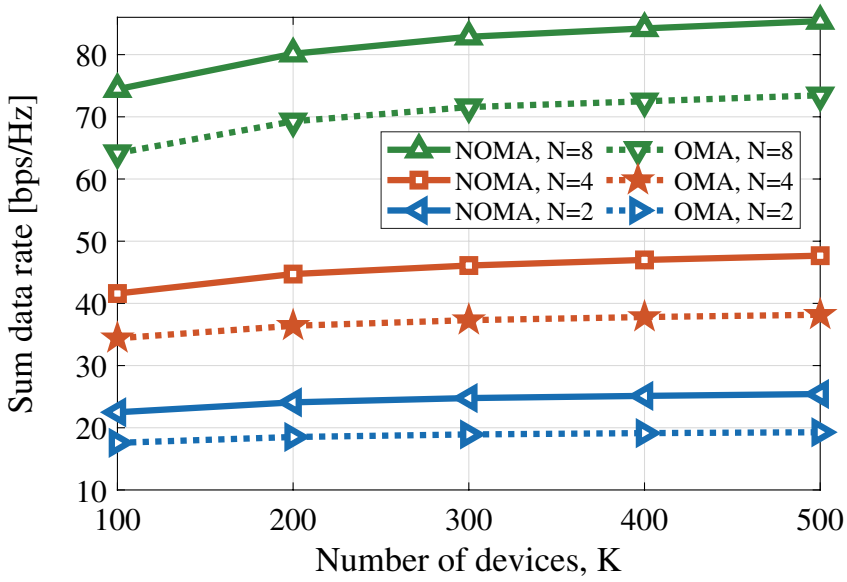
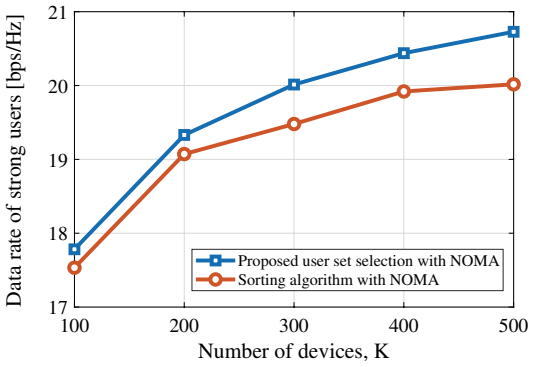


Figure 2.15. The sum data rate of NOMA and OMA through the proposed user selection algorithm for different N with $\beta = 0.8$.

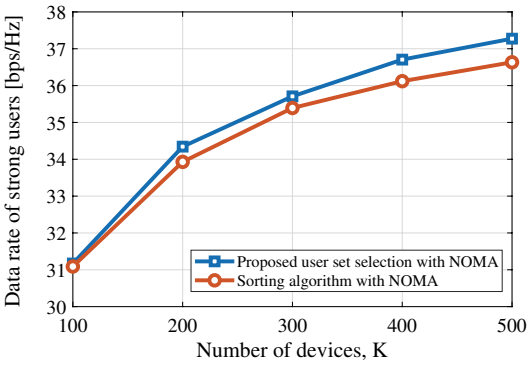
For different number of BS antennas, the proposed user set selection algorithm is compared to the sorting algorithm after determining the optimal power allocation factors

in (2.57). The sorting algorithm determines the strong and weak users sets by considering only their channel gains. Firstly, K devices are sorted based on the channel gains in descending order and classified into two groups as in (2.48) and (2.49). Using (2.48), the first M devices in $\mathcal{S}_s$ which have the highest channel gains are assigned to the strong users set, while using (2.49), the first M devices in $\mathcal{S}_w$ which have the highest channel gains are assigned to the weak users set.

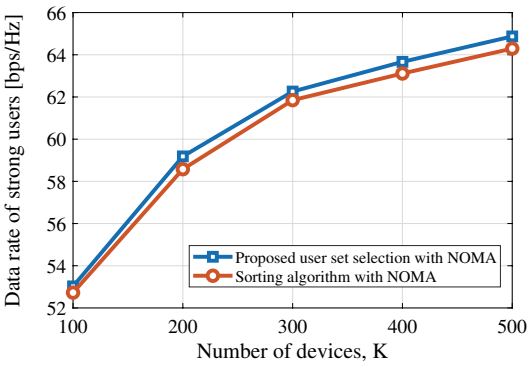
As given in Fig. 2.16, the proposed user set selection algorithm for all N values with $\beta = 0.8$ provides a better performance, in terms of the data rate of the strong users than the sorting algorithm. The reason is that the proposed user set selection algorithm considers the orthogonality between the weak users and the strong users, which reduces inter-set interference. Specifically, as the N value increases, the performance gain of the proposed user set selection algorithm decreases due to decreasing the variation of channel gains.



(a) N = 2.



(b) N = 4.



(c) N = 8.

Figure 2.16. The comparison of user-set selection algorithms in NOMA based MIMO systems for different N with $\beta = 0.8$.

2.7. User Selection for NOMA based Massive MIMO Systems

In this section, we consider the user selection and power allocation scheme for an uplink NOMA based massive MIMO system for a densely devices scenario. We perform the massive MIMO implementation instead of MIMO as given in the uplink NOMA based MIMO system model in Section 2.6. For the NOMA based massive MIMO system, the 2D channel model is used as given in (2.2).

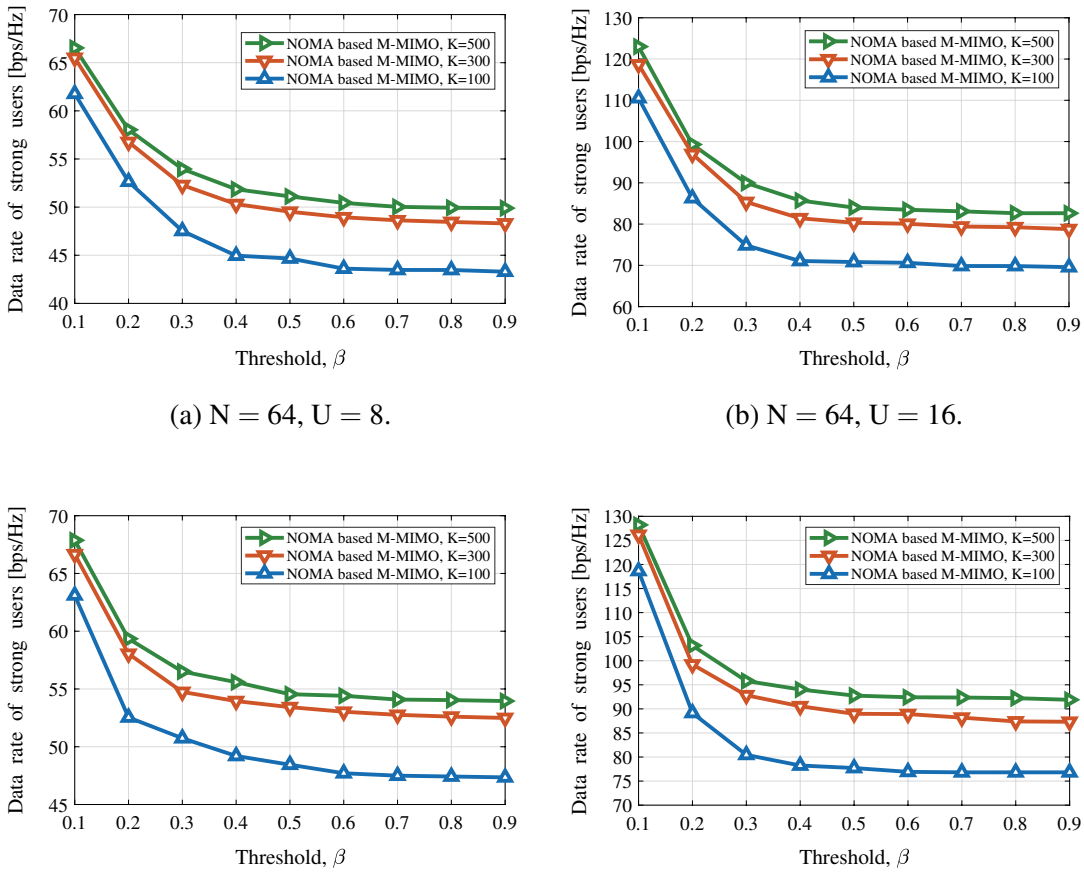
The NOMA based massive MIMO system is composed of N-antenna BS and K single-antenna devices. Among all of these K devices, U devices are selected to be simultaneously served. Thus, we assume that $N >> U >> 1$. We select U users among K users through Algorithm 2, thus there are $U/2$ users for the strong set and $U/2$ users for the weak set. This implies that M=U$/2$ in the Algorithm 2.

2.7.1. Performance Evaluations

The simulation results are provided to compare the proposed user-set selection algorithm for multi-user massive MIMO systems based on NOMA and OMA. The data rate of the strong users is given for the proposed user-set selection with NOMA and the sorting-based user-set selection with NOMA systems. The effect of the number of devices K and the threshold β for the different number of antennas at the BS on the sum data rate and the data rate of the strong users are investigated. The same simulation parameters are used as given in Section 2.6.3. In figures, massive MIMO is abbreviated as "*M-MIMO*".

Fig. 2.17 investigates the optimum β values for different U at N=64 and N=128. The data rate of strong users versus the threshold values, β, is given among different K users. It is shown that $\beta = 0.1$ gives the highest data rate of strong users for different K users at fixed N and U. Thus, $\beta = 0.1$ is selected as the optimum threshold value. For each case, there is an SNR difference of about 23 dB between the strong and weak users to perform NOMA effectively.

In Fig. 2.18, the sum data rate comparison of NOMA and OMA based massive MIMO systems with the proposed algorithm is given, employing $\beta = 0.1$. The NOMA based massive MIMO outperforms the OMA counterpart for Fig. 2.18(a) and Fig. 2.18(b). Moreover, N=128 has a higher sum data rate than N=64 for each U among K users in both NOMA and OMA systems. Specifically, the NOMA based massive MIMO system with N=128 has increased the sum data rate by 73% compared to the OMA based mas-



(a) N = 64, U = 8.

(b) N = 64, U = 16.

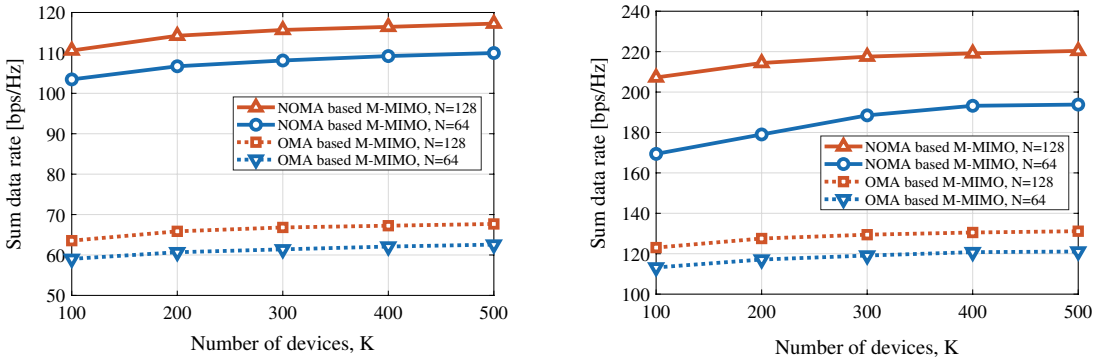
(c) N = 128, U = 8.

(d) N = 128, U = 16.

Figure 2.17. Data rate of strong users for the proposed NOMA based massive MIMO systems versus the threshold β for different N and K.

sive MIMO system with N=128 while serving U = 8 among K users as in Fig. 2.18(a). Similarly, the NOMA based massive MIMO system with N=64 has increased the sum data rate by 76% compared to its OMA counterpart while serving U = 8 among K users. When we increase the serving users to U = 16 as in Fig. 2.18(b), the sum data rates for both NOMA and OMA increase compared to U = 8 for the same number of antennas at BS. Compared to the OMA case, the NOMA based massive MIMO system improves the sum data rate performance with a percentage of 56% and 68% while serving U = 16 among K users for N=64 and N=128, respectively.

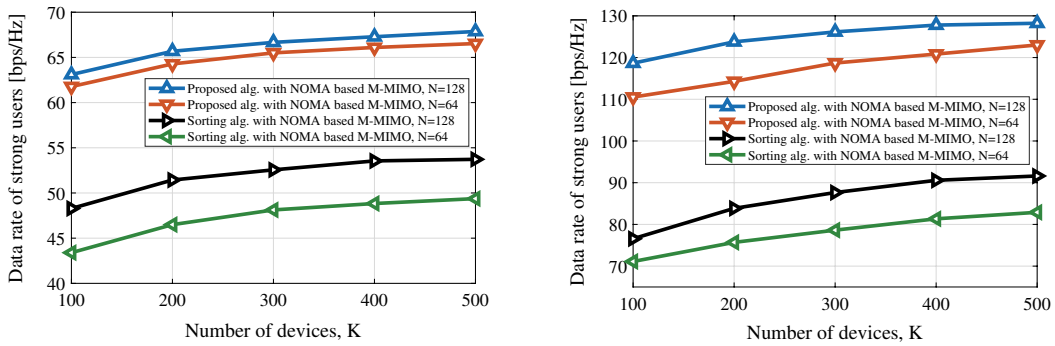
Moreover, the NOMA based massive MIMO with N=128 has increased the sum data rate by around 7 bps/Hz compared to N=64 for U = 8 among K users as in Fig. 2.18(a). In addition to that, when we increase the serving users to U = 16 among K users as in Fig. 2.18(b), the NOMA based massive MIMO with N=128 has increased the sum data rate by around 31 bps/Hz compared to N=64.



(a) U = 8. (b) U = 16.

Figure 2.18. The sum data rate comparison of NOMA and OMA based massive MIMO systems with the proposed algorithm for different U and N at $\beta = 0.1$.

Fig. 2.19 compares the proposed user-set selection and sorting algorithm with NOMA based massive MIMO in terms of the data rate of strong users for the different number of antennas at BS, N, and selected U users among K total users. Fig. 2.19(a) and Fig. 2.19(b) show that the proposed user-set selection algorithm outperforms the sorting algorithm for any N and K values. Specifically, the proposed algorithm provides 37% and 28% higher data rate of strong users than the sorting algorithm for N=64 and N=128, respectively while serving U = 8 among K users in Fig. 2.19(a). On the other hand, Fig. 2.19(b) demonstrates that the proposed algorithm achieves 51% and 46% higher data rate of strong users compared to the sorting algorithm for N=64 and N=128, respectively while serving U = 16 among K users.



(a) U = 8. (b) U = 16.

Figure 2.19. The comparison of user-set selection algorithms for different U and N at $\beta = 0.1$.

2.8. Conclusion

In this chapter, we have presented background information about MIMO, massive MIMO and NOMA systems. Then, we have proposed the user selection algorithm by reconstructing the sparse channel of massive MIMO with the OMP to reduce the feedback load. The proposed algorithm eliminates users according to the channel correlation to improve the sum data rate performance. Also, the effects of feedback measurement and sparsity level parameters have been provided through the simulations. Since the feedback measurement and the sparsity level severely impact the channel reconstruction, the choice of these parameters is important. The simulation results show that the proposed algorithm outperforms the traditional approaches in terms of data rate, reduced feedback load and low complexity.

Moreover, for the uplink NOMA based MIMO system, a user-set selection algorithm that determines the devices incurring the least inter-set interference has been proposed to improve the sum data rate. For the uplink NOMA based MIMO system, simulation results have verified the advantage of the proposed user-set selection with power allocation over conventional OMA systems. Moreover, in a densely deployed environment, the same user-set selection and power allocation scheme methods have been extended to the uplink NOMA based massive MIMO systems.

CHAPTER 3

MASSIVE MIMO-NOMA BASED MEC SYSTEMS

In this chapter, an overview of mobile edge computing (MEC) system is given in detail. Then, we present a framework for a MEC system integrated with massive MIMO and NOMA technology to demonstrate the advantages of massive connectivity, higher spectral efficiency and lower delay for delay-sensitive applications (Yılmaz and Özbek, 2023). We aim to minimize the overall computing and transmission delay under the users' transmit power and MEC computing capability. Thus, multiple users in the system, including the user at the cell-edge, can offload their tasks to the MEC server under an overall delay constraint through the pairing scheme for massive MIMO-NOMA. Performance results are provided regarding to the sum data rate and overall system delay compared with the OMA-massive MIMO and massive MIMO based MEC systems. Moreover, since the system performance of massive MIMO-NOMA based MEC system depends on the user selection approach, we apply a user selection for a densely deployed system to improve the performance. Therefore, the overall transmission and computing delay in MEC can be decreased significantly.

3.1. Mobile Edge Computing (MEC)

Recent advances in the next-generation wireless technologies have motivated several computationally intensive and latency-critical applications such as virtual reality (VR), augmented reality (AR), mobile online gaming, face recognition, autonomous driving, telesurgery, unmanned aerial vehicles (UAVs) and Internet of Things (IoT) (Liao et al., 2020). These applications require ultra-low-latency communication, computation and control among many wireless devices (Zeng et al., 2020a). Since the devices have small physical sizes, limited computation capacities and limited power sources, it is challenging to handle intensive computation load at the user side. Moreover, this reduces the quality of the user experience, resulting in excessive delay and power consumption.

To overcome these limitations, cloud computing offers one possible solution by offloading computation-intensive tasks from users to the cloud. However, because of the data propagation through wide area networks, the long propagation distances between the

devices and the centralized cloud, cloud computing can cause excessive latency computation and heavy traffic loads at the backhaul networks (Pham et al., 2 019). Therefore, cloud computing may not support latency-critical applications.

To mitigate the limitations and concerns of conventional cloud computing, MEC has been a promising solution to enable computation-intensive and latency-critical applications. MEC utilizes powerful cloud-computing capabilities, such as a MEC server integrated into the BS within the Radio Access Network (RAN) and eliminates the requirement for traffic to be routed through the core network. Compared with cloud computing, in MEC systems, users offload computation-intensive tasks to the powerful MEC servers in proximity to BSs for execution, which avoids data delivery over the backhaul networks and reduces latency (Zhang et al., 2019). MEC can significantly reduce computation latency and traffic loads at backhaul networks (Wang et al., 2019). One of the other benefits of the MEC system is that the energy consumption of the devices can be significantly reduced by offloading computation-intensive tasks to a MEC server for execution (Ren et al., 2018). Besides, since the tasks can be computed at the adjacent BS instead of the remote cloud center, the congestion in the core network can be effectively relieved. As a result, by moving the computing and storage features to the edge, MEC also provides a distributed and decentralized service environment characterized by low latency and high-rate access (Malik and Vu, 2020).

3.1.1. Offloading Scheme

In MEC systems, there are three different offloading schemes as shown in Fig. 3.1:

- Remote computing only,
- Binary offloading and,
- Partial offloading.

In remote computing only, all users offload their computing tasks to the MEC server. In binary offloading, users' tasks are inseparable, and users can either offload their computational tasks to the MEC server or process them locally. This practically corresponds to highly integrated or relatively simple tasks, such as speech recognition and natural language translation (Cao et al., 2019). On the other hand, in partial offloading, the computational tasks of users are partitioned into two parts, with one processed locally and the other offloaded for edge execution. This corresponds to applications with multiple fine-grained procedures/components, e.g., AR applications. For binary offloading, users

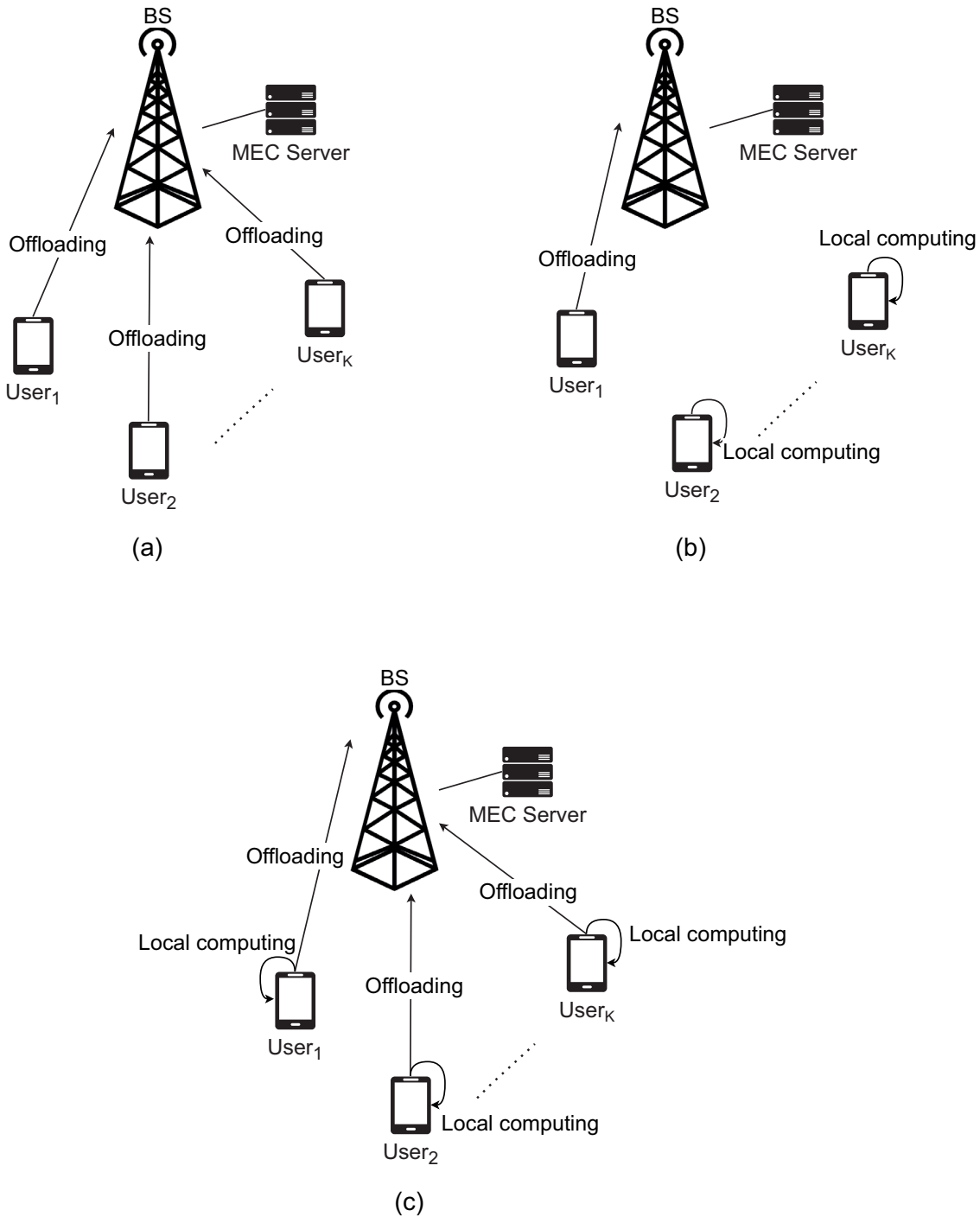


Figure 3.1. Offloading strategies in MEC: (a) remote computing only (b) binary offloading (c) partial offloading.

prefer to offload when their connection to the MEC server is good, and the MEC server's computational resource is sufficient. For the case of partial offloading, the distribution of the computational tasks is determined by a number of parameters, including the characteristics of the computational tasks, the computing capabilities of the users and MEC server and the wireless channel conditions (Zeng et al., 2020a).

A taxonomy of several existing MEC schemes is provided in Fig. 3.2 in terms of offloading strategy, system objective and approach for the single antenna, MIMO and massive MIMO systems.

***Single antenna-based MEC systems*:** Many studies have analyzed single antenna-based MEC systems with different multiplexing access techniques, such as TDMA, FDMA, CDMA, NOMA and OFDMA. The papers of (Qiu et al., 2022) (Irum et al., 2022) have performed the remote computing only. The binary offloading scheme has been used in (Chen et al., 2016) (Ren et al., 2018) (Hmimz et al., 2019) (Wang et al., 2019) (Liang et al., 2020), whereas the partial offloading scheme has been considered in (Ren et al., 2018) (Sun et al., 2019) (Wang et al., 2019) (Tiwari et al., 2020) (Fang et al., 2020) (Baidas, 2020) (Wang et al., 2021) (Chen et al., 2022). As another offloading scheme, the cooperative computing scheme has been presented in (Huang and Liu, 2018) (Liu, 2019) (Yao et al., 2019) (Cao et al., 2019) (Wen et al., 2020) (Tan et al., 2021) (Li et al., 2021) (Pan et al., 2021) (Yılmaz and Özbek, 2022).

In terms of system objective, the energy consumption minimization has been investigated in (Huang and Liu, 2018) (Cao et al., 2019) (Yao et al., 2019) (Wang et al., 2019) (Hmimz et al., 2019) (Wen et al., 2020) (Tan et al., 2021) (Li et al., 2021) (Wang et al., 2021) (Qiu et al., 2022). While the papers of (Huang and Liu, 2018), (Wen et al., 2020) and (Li et al., 2021) have formulated the energy consumption minimization for a cooperative NOMA based MEC system, the papers of (Wang et al., 2019) (Qiu et al., 2022) and (Wang et al., 2021) have provided the energy consumption minimization problem for NOMA-assisted MEC systems. The authors in (Cao et al., 2019) have considered the user cooperation approach based on a four-slot cooperation protocol. However, in (Yao et al., 2019), a computing architecture for cooperative computation offloading with multi-source, multi-relay, and a single edge server has been proposed in an OFDMA wireless network with four computing modes. Similarly, the authors in (Tan et al., 2021) have investigated a joint offloading decision, collaboration decision, computing resource allocation and communication resource allocation problem in the multi-user collaborative mobile edge computing network (C-MEC) based on OFDMA to minimize the total energy consumption under the delay constraint.

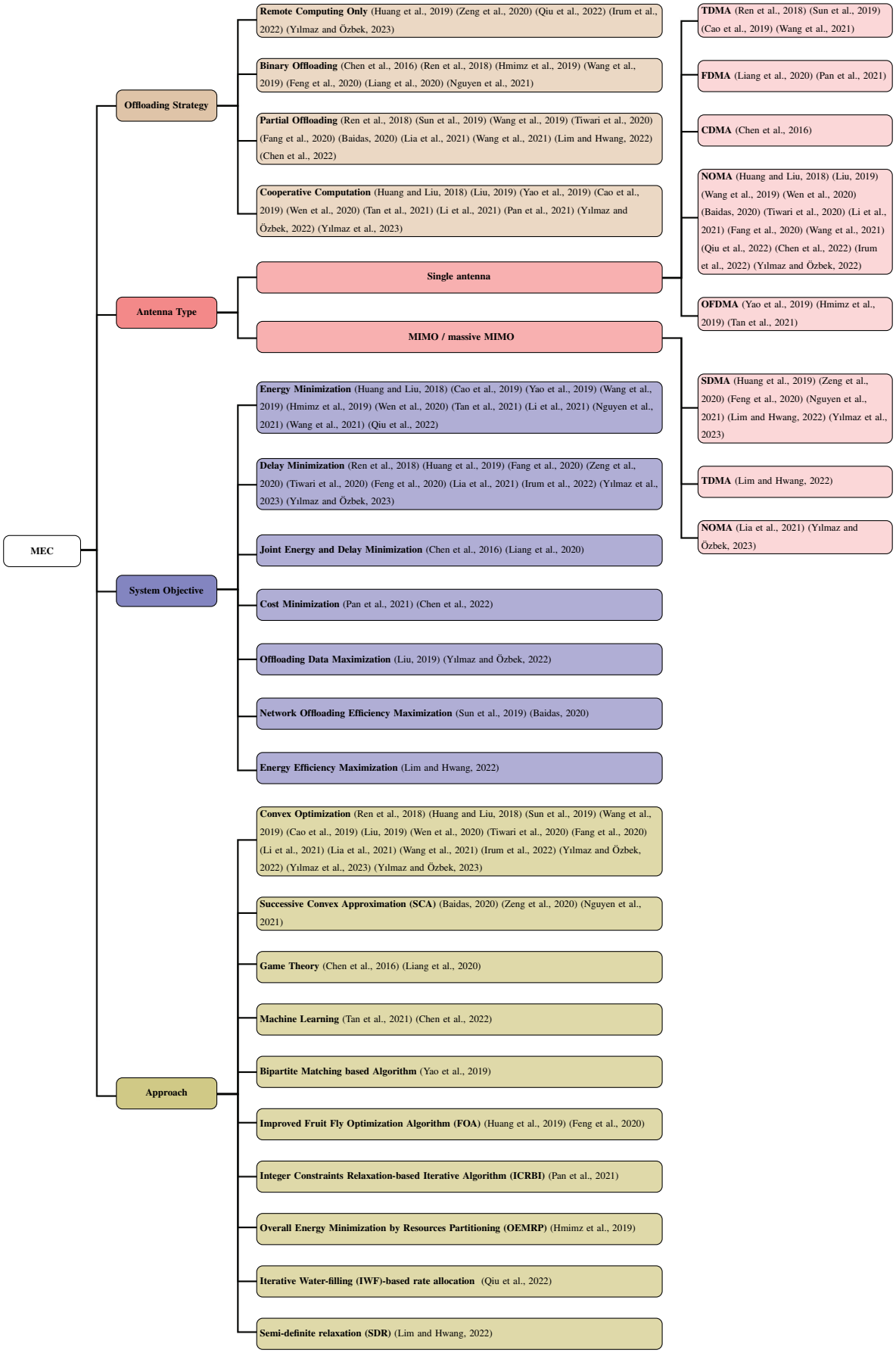


Figure 3.2. A taxonomy of existing MEC schemes.

On the other hand, the delay minimization problem has been formulated in (Ren et al., 2018) (Fang et al., 2020) (Tiwari et al., 2020) and (Irum et al., 2022). In particular, the task delay minimization problem has been investigated for NOMA-enabled multi-user MEC systems in (Fang et al., 2020), (Tiwari et al., 2020) and (Irum et al., 2022). However, the authors in (Ren et al., 2018) have investigated the joint communication and computation resource allocation for a TDMA-based multi-user mobile edge computation offloading (MECO) system to improve the quality of experience (QoE) for users by minimizing the weighted-sum delay of all devices.

The papers of (Chen et al., 2016) and (Liang et al., 2020) have considered the joint energy and delay minimization. In (Chen et al., 2016), a game theoretic approach has been presented for the computation offloading decision-making problem among multiple mobile device users for mobile-edge cloud computing with the CDMA technique. In (Liang et al., 2020), the problem of interference-aware single-cell multi-user computation offloading has been studied in the MEC. They have applied game theory and reinforcement learning methods to achieve effective computation offloading.

The cost minimization problem has considered in (Pan et al., 2021) and (Chen et al., 2022). In (Pan et al., 2021), a cooperative task computation framework has been presented to exploit the computation resource in UEs to accomplish more tasks while minimizing the power consumption of UEs. In the considered framework, the offloading decision, the computational frequency and the offloading power for each UE have been optimized jointly to minimize the system total cost, which consists of the cost charged for the UE's power consumption and the penalty caused by the unaccomplished tasks. In (Chen et al., 2022), decentralized computation offloading in a NOMA based multi-user MEC system has been investigated, where long-term average network computation cost is minimized in terms of power consumption and buffering delay.

The offloading data maximization problem has been investigated in (Liu, 2019) and (Yılmaz and Özbek, 2022). Specifically, cooperative edge computing has been studied in (Liu, 2019) for a basic three-node model, which exploits the parallel transmission structure of NOMA over the whole resource block to maximize the sum offloading data subject to the latency constraints. In (Yılmaz and Özbek, 2022), a cooperative MEC has been presented that exploits the combination of NOMA and multiple helpers to maximize the total offloading data subject to the latency constraints.

The network offloading efficiency maximization problem has been introduced in (Sun et al., 2019) and (Baidas, 2020). The TDMA based MEC offloading is performed for (Sun et al., 2019), whereas the problem is formulated in (Baidas, 2020) for a clustered

NOMA-enabled MEC system.

***MIMO and massive MIMO based MEC systems*:** Besides the single antenna-based MEC systems, the integration of MIMO/massive MIMO technologies into the MEC system has been considered with SDMA, TDMA and NOMA techniques, as in Fig. 3.2. As an offloading strategy, the papers of (Huang et al., 2019) and (Zeng et al., 2020) consider remote computing only. The binary offloading has been performed in (Feng et al., 2020) (Nguyen et al., 2021), whereas (Lim and Hwang, 2022) and (Lia et al., 2021) consider the partial offloading.

As system objective, the delay minimization problem has been formulated in (Huang et al., 2019) (Zeng et al., 2020) (Feng et al., 2020) and (Lia et al., 2021). Specifically, the authors in (Huang et al., 2019) (Zeng et al., 2020) and (Feng et al., 2020) have studied joint communication and computation resource allocation problem for a single cell massive MIMO based MEC system to minimize the maximum offloading delay over multiple users. However, the authors in (Lia et al., 2021) have applied massive MIMO-NOMA technology to the MEC system.

On the other hand, the authors in (Nguyen et al., 2021) have studied the optimization of computation task offloading and resource allocation in MIMO wireless systems considering perfect and imperfect CSI estimation to minimize the maximum weighted energy consumption. The paper of (Lim and Hwang, 2022) has considered the energy efficiency (EE) maximization problem. In particular, energy-efficient beamforming and resource allocation have been studied for multi-access edge computing systems consisting of multi-antenna access points (APs) and single-antenna users.

3.2. Task Offloading for Delay Minimization in MEC

In MEC systems, several users may access the same server for task offloading, which requires massive connectivity and a multiple-access strategy. Therefore, efficient and stable wireless communication is needed to satisfy the seamless task offloading to provide a transmission framework for improving system throughput and reducing overall delay. The overall delay includes transmission and computing delays. Due to the real-time processing requirements of MEC systems, it is promising to integrate massive MIMO and NOMA with MEC.

The aforementioned studies in Fig. 3.2, such as (Baidas, 2020) (Fang et al., 2020) (Tiwari et al., 2020) (Wang et al., 2021) (Qiu et al., 2022) (Chen et al., 2022) and

(Irum et al., 2022) have analyzed single antenna-based MEC systems with NOMA technology to minimize energy consumption, network computation cost, and task delay or maximize the network offloading efficiency. Nonetheless, all these studies consider single antenna-based offloading with the NOMA.

On the other hand, the work in (Wang et al., 2019) has considered computation offloading via multi-antenna NOMA to improve multi-user MEC systems' performance, in which the BS has four antennas. The weighted sum-energy minimization problem has been formulated for partial offloading and binary offloading.

The following works consider MIMO or massive MIMO assisted MEC system. The authors of (Nguyen et al., 2021) have considered a single-cell massive MIMO system to minimize the maximum weighted energy consumption, while the paper of (Lim and Hwang, 2022) has considered maximizing the EE of the MIMO based MEC system. The following studies (Huang et al., 2019) (Zeng et al., 2020) and (Feng et al., 2020) have considered the delay minimization for massive MIMO based MEC systems. Although these studies improve MEC system performance through massive MIMO, they do not consider the NOMA system, which improves spectral efficiency in massive MIMO based MEC system.

The previous studies focused on only the NOMA or massive MIMO based MEC systems, while we propose a massive MIMO and NOMA assisted MEC system for delay-sensitive applications to minimize the overall computing and transmission delay for remote computing in which all users in each cluster offload the computing tasks to the MEC server, considering both offloading and computing phases (Yılmaz and Özbek, 2023). The main motivation is to construct an efficient MEC mechanism where not only the users with higher channel gain and computational-intensive tasks but also cell-edge users with lower channel gain can offload their tasks through user pairing, offloading and computation scheme. The study of (Lia et al., 2021) has provided the massive MIMO and NOMA based MEC system to reduce the overall delay among all users without considering the cluster concept, while we apply in the same transmission delay for the users in the same cluster in NOMA (Yılmaz and Özbek, 2023). In addition to that, a partial offloading scheme has been adopted in (Lia et al., 2021), while we present an offloading for remote computing scheme.

To be specific, in this section, we present a massive MIMO-NOMA assisted MEC system for remote computing. Specifically, the users with relatively higher channel gains are called as strong and the users at the cell-edge are called as weak. While pairing the weak users with the strong ones in each cluster, the weak users can offload a portion

of their data to the MEC. In each cluster, the strong user determines the transmission delay to offload its data, while a portion of the weak user's data is also offloaded during this transmission. We formulate the problem of minimizing the overall computing and transmission delay over the massive MIMO based NOMA system under the computing capability and transmit power constraints. Then, the optimization problem is transformed into a linear problem and solved with a convex optimization tool, i.e., the interior-point method.

3.2.1. System Model

As shown in Fig. 3.3, we consider an uplink massive MIMO-NOMA assisted MEC system for remote computing in which a BS is equipped with an antenna array of N elements and serves K single-antenna users, under the assumption that $K << N$. The sets of users and antennas are denoted as $\mathcal{K} = \{1, \ldots, K\}$ and $\mathcal{N} = \{1, \ldots, N\}$, respectively. We consider the case of two users in each cluster for NOMA having totally M clusters with $M \leq N$ and $m \in \{1, \ldots, M\}$. Then, the number of clusters, M, is determined as $K/2$. Accordingly, in the proposed system, an efficient user clustering is applied.

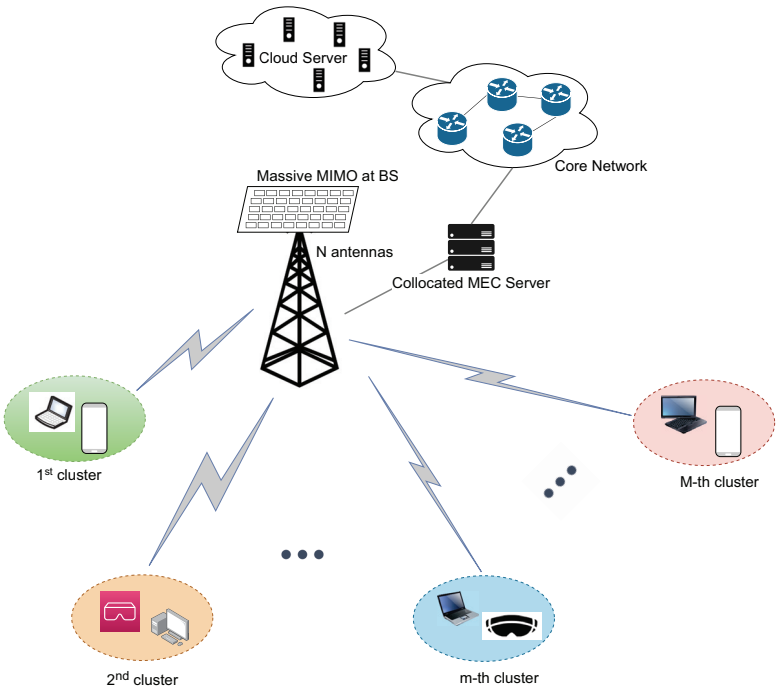


Figure 3.3. The massive MIMO-NOMA assisted MEC system.

We apply the High-High channel gain user pairing strategy for the proposed framework to pair these K users (Rauniyar et al., 2020). Specifically, the channel gains of K users are sorted in descending order. Then, these K users are divided into two sets, each including M users. The set of the first M users with the higher channel gains is defined by the strong users set, while the weak users set includes the set of remaining M users with lower channel gains from $M+1$ to K. Then, the two-user cluster is formed by pairing the users from each set. Accordingly, the first cluster is formed by pairing the first user of the strong users set with the first user of the weak users set, and so on through the High-High channel gain users pairing strategy. Specifically, the first cluster includes the 1st user and $M+1$th user, and the Mth cluster contains the Mth user and Kth user.

The uplink channel vectors, $\mathbf{h}_{m,j}$, belonging to j^{th} set and m^{th} cluster are given as (2.1), in Subsection 2.6.1. The channel matrices belonging to the strong users set, $\mathbf{H}_1$, and weak users set, $\mathbf{H}_2$, in all clusters are given in (2.28) and (2.29), respectively.

In the proposed framework, the delay mainly includes transmission time and computing time in a remote computing scheme. We focus on the uplink transmission and do not consider the required time to transmit the computing data from the MEC server to the user in the downlink phase.

Transmission Scheme: The proposed framework includes both offloading phase and computing phase. Firstly, we provide the offloading phase where the users transmit their tasks to the MEC server.

The received signal at the BS is given by (2.30) and (2.31). Here, we assume that the power allocation factor of the strong user in each m^{th} cluster is different from each other. Similarly, the power allocation factor of the weak user in each m^{th} cluster is different from each other. Accordingly, the power allocation factors of the strong users in the m^{th} cluster are given as $[\alpha_{1,1} \dots \alpha_{m,1} \dots \alpha_{M,1}]$ and the power allocation factors of the weak users in the m^{th} cluster are given by $[\alpha_{1,2} \dots \alpha_{m,2} \dots \alpha_{M,2}]$ with $m \in \{1, \dots, M\}$.

The average data rate of the user in the j^{th} set and the m^{th} cluster, $R_{m,j}$, is defined in (2.40). Then, the sum data rate in the system is given by $R_{\text{sum}}^{\text{NOMA}}$ as in (2.42).

3.2.2. Problem Formulation and Proposed Solution

In this section, we propose the computation scheme for the massive MIMO-NOMA assisted MEC system. Then, we present the optimization problem and give the solution based on the interior-point algorithm to tackle the problem of minimizing overall delay.

Computation Scheme: The main objective of the proposed system is to execute the data belonging to strong users under delay constraints while offloading a portion of weak users' data to MEC at the same transmission time. Accordingly, the strong user determines the transmission delay in each cluster, m, to offload its data to the MEC. The total task size for the strong user is initially defined as $D_{m,1}$. The transmission time to offload $D_{m,1}$ for the strong user in the m^{th} cluster is given by;

$$T_{m,1}^{t} = \frac{D_{m,1}}{R_{m,1}} \tag{3.1}$$

Since the transmission delay in each cluster is determined by the strong user, we calculate the offloaded data by the weak user as follows;

$$T_{m,1}^{t}\ R_{m,2} = D_{m,2} \tag{3.2}$$

When the users' data is received, the MEC server allocates its computing resources to the tasks and the computing is performed for each cluster. The computing time at the MEC server belonging to m^{th} cluster is expressed as;

$$T_{m}^{c} = \frac{(D_{m,1} + D_{m,2})\, C_{mec}}{f_{m}^{mec}} \tag{3.3}$$

where C_{mec} denotes the number of CPU cycles required to calculate one bit in the MEC server, which is also named the computation intensity. f_{m}^{mec} denotes the CPU frequency allocated to the m^{th} cluster by MEC.

Finally, the total time to perform the task in the m^{th} cluster is expressed as;

$$T_{m} = T_{m,1}^{t} + T_{m}^{c} \tag{3.4}$$

Our aim is to minimize the overall computing and transmission delay in all clusters by jointly optimizing the users' transmit power and MEC computing capacity. Then, we define the optimization problem as follows;

$$\min_{\boldsymbol{\alpha},\mathbf{F}} \quad \max_{m} \ (T_m) \tag{3.5}$$

$$\text{s.t. } 0 < \alpha_{m,j} \leq 1,\ j = \{1,2\},\ m \in \{1,\ldots,M\} \tag{3.5a}$$

$$\sum_{m=1}^{M} f_m^{mec} \leq F_{max} \tag{3.5b}$$

$$\hat{R}_{m,2} < R_{m,2}, \quad m \in \{1,\ldots,M\}. \tag{3.5c}$$

where $\boldsymbol{\alpha} = [\alpha_{1,1},\ldots,\alpha_{m,1},\ldots,\alpha_{M,1},\ \alpha_{1,2},\ldots,\alpha_{m,2},\ldots,\alpha_{M,2}]$, $\mathbf{F} = [f_1^{mec},\ldots,f_M^{mec}]$ and the available computing resource at the MEC server defined by the maximum CPU operating frequency is F_{max}. $\hat{R}_{m,2}$ is the data rate of the m^{th} weak user for OMA system and is expressed as;

$$\hat{R}_{m,2} = \frac{B}{2}\mathbb{E}\left\{\log_2\left(1 + \frac{P\left|\mathbf{w}_{m,2}\mathbf{h}_{m,2}\right|^2}{\left\|\mathbf{w}_{m,2}\right\|^2 {\sigma_n}^2}\right)\right\}. \tag{3.6}$$

The constraint (3.5a) shows the range of power allocation factors for each user in the j^{th} set and m^{th} cluster. The constraint (3.5b) gives the total computing resources. The constraint (3.5c) gives the minimum data rate requirement, in which the data rate of the weak user in the m^{th} cluster should be higher than the OMA data rate.

In order to reduce the computing time T_m^c, we can use all available CPU computing resources and then the constraint (3.5b) is re-defined as;

$$\sum_{m=1}^{M} f_m^{mec} = F_{max} \tag{3.7}$$

Accordingly, we share the computing resources among the clusters equally as follows;

$$f_m^{mec} = \frac{F_{max}}{M} \tag{3.8}$$

To solve the optimization problem (3.5), we transform the min-max problem into a minimization problem by introducing an auxiliary variable ν as follows;

$$\min_{\boldsymbol{\alpha}} \nu \tag{3.9}$$

s.t. (3.5a) (3.5c) and (3.8),

$$T_m \leq \nu, \quad m \in \{1, \ldots, M\}\,. \tag{3.9a}$$

Thus, we have transformed the non-convex problem (3.5) into a convex problem (3.9) using the auxiliary variable. On this basis, the solution to the problem (3.9) is found by applying the interior-point method subject to the nonlinear inequalities constraints.

In the interior-point method, a log-barrier term is used for the inequality constraints, and the problem with inequality constraints can be reduced to having only equality constraints (Waltz et al., 2006). Barrier functions are usually a logarithmic function and can be used to transform a constrained problem into a sequence of unconstrained problems. These functions avoid the iterates from leaving the feasible region by acting as a barrier.

In this way, we reformulate the objective function in (3.9) as an auxiliary function with a barrier parameter μ by;

$$S_\mu(\mathbf{x}) = \nu + \mu\ P(\mathbf{x}) \tag{3.10}$$

where P (.) is an interior penalty function as;

$$P(\mathbf{x}) = -\sum_{i=1}^{K} \log\left[-G_i(\mathbf{x})\right],\ \text{for } G_i(\mathbf{x}) < 0 \tag{3.11}$$

where the nonlinear inequality constraints in (3.5c) and (3.9a) are rearranged so that $\hat{R}_{m,2} - R_{m,2} < 0$ and $T_m - \nu \leq 0$ for $m \in \{1, \ldots, M\}$, respectively. Then, each of these nonlinear inequalities is represented by $\mathbf{G} = [G_1, G_2, \ldots, G_K]$.

Algorithm 3 outlines the detailed steps of the interior-point method. In the minimization problem, $\mathbf{x}$ is determined as a vector of the following components; $\mathbf{x} = [\alpha_{m,j}, \nu]$, $\forall m, j$. A feasible solution to Problem (3.9) is the vector $\mathbf{x}$ satisfying all the constraints. The initial values, $\mathbf{x}^1$, are decided by determining the lower and upper bounds range for each component $\alpha_{m,j}, \nu$ in $\mathbf{x}$. For this method, the initial values of power allocation factors, $\alpha_{m,j}$, and the overall delay, ν, are defined to satisfy the constraints.

Algorithm 3 Minimization of the Overall Delay Through Interior-Point Algorithm

Input: $\mathbf{x} = [\alpha_{m,j}, \nu]$, f_m^{mec}, $D_{m,1}$, $\mathbf{h}_{m,j}$ for $j = \{1, 2\}$ and $m \in \{1, \ldots, M\}$.

• Initialization Step:

Select a growth parameter, $\zeta > 1$.

Select a stopping parameter, $\varepsilon > 0$.

Set u=1.

Select an initial value of the barrier parameter $\mu_u > 0$.

Rearrange the inequality constraints in (3.5c) and (3.9a) as $\mathbf{G} = [G_1, G_2, \ldots, G_K]$, $G_i(.) \leq 0$, $i = 1, 2, \ldots, K$.

Choose initial feasible points $\mathbf{x}^u$ with $\mathbf{G}(\mathbf{x}^u) < 0$.

Reformulate the objective function as an auxiliary function by,

$$S_{\mu_u}(\mathbf{x}) = \nu + \mu_u P(\mathbf{x})$$

where

$$P(\mathbf{x}) = -\sum_{i=1}^{K} \log\left[-G_i(\mathbf{x})\right]$$

• Iteration Step:

Starting from $\mathbf{x}^u$, use an unconstrained search technique such as an iterative descent method applicable to unconstrained problems including steepest descent or Newton's method to find the point that minimizes $S_{\mu_u}(\mathbf{x})$ and call it as the new starting point, $\mathbf{x}^{u+1}$

• Stopping Criterion Step:

if $\|\mathbf{x}^{u+1} - \mathbf{x}^u\| < \varepsilon$, stop **then**

    $\mathbf{x}^{u+1}$ is an estimate of the optimal solution

else

    $\mu_{u+1} = \zeta\mu_u$

    Reformulate the $S_{\mu_{u+1}}(\mathbf{x})$ with $u = u + 1$

    Go to iteration step

end if

Output: $\boldsymbol{\alpha}^*$and ν^*

After finding the optimized solutions through Algorithm 3, $\boldsymbol{\alpha}^*$and ν^*, we calculate the total data of strong users, D_1, and the total data of weak users, D_2, over all clusters as given below, respectively.

$$D_1 = \sum_{m=1}^{M} D_{m,1} \tag{3.12}$$

$$D_2 = \sum_{m=1}^{M} D_{m,2} \tag{3.13}$$

Thus, the total executed data in the MEC system is given by;

$$D = D_1 + D_2. \tag{3.14}$$

The complexity of the interior-point method in Algorithm 3 can be determined as $\mathcal{O}\left(\sqrt{n}\,\frac{1}{\varepsilon}\right)$ iterations, where n is the number of variables in the problem, depending on mainly the number of users, K, in the system (Lesaja, 2009). Thus, the number of users and the choice of the convergence tolerance, ε, affect the complexity. In the algorithm, convergence tolerance, ε, is selected as 10^{-6}.

OMA-massive MIMO based MEC scheme: The system performance is compared with the OMA-massive MIMO based MEC scheme (Kim et al., 2015) as in Fig. 3.4. We serve the same number of users within two-time slots and equally share computing resources among the users.

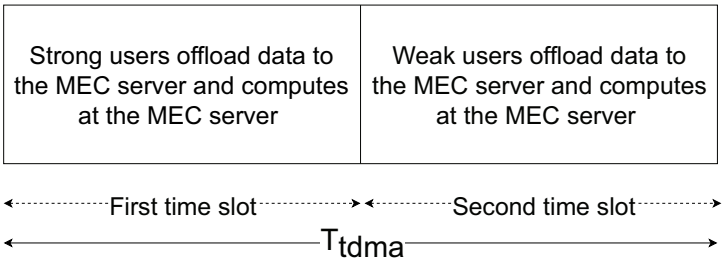


Figure 3.4. Time slot allocation for OMA-massive MIMO based MEC scheme.

The average data rate of the m^{th} weak user, $\hat{R}_{m,2}$, is given in (3.6). Then, the average data rate of the m^{th} strong user, $\hat{R}_{m,1}$, is given by;

$$\hat{R}_{m,1} = \frac{B}{2}\mathbb{E}\left\{\log_2\left(1 + \frac{P\left|\mathbf{w}_{m,1}\mathbf{h}_{m,1}\right|^2}{\left\|\mathbf{w}_{m,1}\right\|^2 {\sigma_n}^2}\right)\right\}. \tag{3.15}$$

Thus, the sum data rate for OMA-massive MIMO based MEC scheme is given by;

$$R_{\text{sum}}^{\text{OMA}} = \sum_{m=1}^{M}\sum_{j=1}^{2}\hat{R}_{m,j}. \tag{3.16}$$

Under these circumstances, the transmission time and the computing time are calculated and rearranged. The transmission time to offload $D_{m,1}$ for the m^{th} strong user is given by;

$$\hat{T}_{m,1}^{t} = \frac{D_{m,1}}{\hat{R}_{m,1}} \tag{3.17}$$

The computing time at the MEC server to compute data belonging to m^{th} strong user is expressed as;

$$\hat{T}_{m,1}^{c} = \frac{D_{m,1}\,C_{mec}}{f^{mec}} \tag{3.18}$$

where $f^{mec} = \frac{F_{max}}{K}$.

Then, the total time for m^{th} strong user in the OMA-massive MIMO based MEC scheme is given as $\hat{T}_{m,1} = \hat{T}_{m,1}^{t} + \hat{T}_{m,1}^{c}$.

Similarly, the transmission time to offload $D_{m,2}$ for the m^{th} weak user is given by;

$$\hat{T}_{m,2}^{t} = \frac{D_{m,2}}{\hat{R}_{m,2}} \tag{3.19}$$

The computing time at the MEC server to compute data belonging to m^{th} weak user is expressed as;

$$\hat{T}_{m,2}^{c} = \frac{D_{m,2}\,C_{mec}}{f^{mec}} \tag{3.20}$$

The total time for the m^{th} weak user in the OMA-massive MIMO based MEC scheme is given as $\hat{T}_{m,2} = \hat{T}^t_{m,2} + \hat{T}^c_{m,2}$.

The overall delay for the OMA-massive MIMO based MEC scheme is given;

$$T_{\text{tdma}} = \max_{\forall m} \left\{ \hat{T}_{m,1} \right\} + \max_{\forall m} \left\{ \hat{T}_{m,2} \right\}. \tag{3.21}$$

Massive MIMO based MEC system: The system performance is also compared with the massive MIMO based MEC system. The instantaneous SINR is defined by;

$$\widetilde{\text{SINR}}_k = \frac{\alpha_k P \left| \mathbf{v}_k \, \mathbf{h}_k \right|^2}{\sum\limits_{i=1, i \neq k}^{K} \alpha_i P \left| \mathbf{v}_k \mathbf{h}_i \right|^2 + \left\| \mathbf{v}_k \right\|^2 {\sigma_n}^2}, \forall k \in \mathcal{K}. \tag{3.22}$$

where $\mathbf{v}_k$ is the normalized ZF postcoder vector with the size $1 \times N$ and the ZF postcoding matrix is $\mathbf{V} = \mathbf{H}^H (\mathbf{H}\mathbf{H}^H)^{-1}$. The normalized ZF postcoding vector is given as $\mathbf{v}_k = \frac{\mathbf{V}_k}{\|\mathbf{V}_k\|}$, where $\mathbf{V}_k$ is the k^{th} row of $\mathbf{V}$. Here, the channel matrix is $\mathbf{H} = [\mathbf{H}_1 \, \mathbf{H}_2]$ including $\mathbf{h}_k$ that is the k^{th} column of $\mathbf{H}$ with size $N \times 1$ represents either strong or weak user channel vector. α_k is the power allocation factor, which represents either strong or weak user power allocation factor.

Then, the average data rate of the k^{th} user is calculated as follows;

$$\widetilde{R}_k = B \, \mathbb{E} \left\{ \log_2 \left(1 + \widetilde{\text{SINR}}_k \right) \right\}, \forall k \in \mathcal{K}. \tag{3.23}$$

Thus, the sum data rate for massive MIMO based MEC system is expressed as;

$$R_{\text{m-MIMO}} = \sum_{k=1}^{K} \widetilde{R}_k \tag{3.24}$$

The transmission time to offload $\widetilde{D}_k$ for the k^{th} user is given by;

$$\widetilde{T}^t_k = \frac{\widetilde{D}_k}{\widetilde{R}_k} \tag{3.25}$$

where $\widetilde{D}_k$ is the total task size which represents either strong or weak user task size.

The computing time at the MEC server to compute data belonging to k^{th} user is expressed as;

$$\widetilde{T}_k^c = \frac{\widetilde{D}_k \, C_{mec}}{f^{mec}} \tag{3.26}$$

Then, the total time for k^{th} user in the massive MIMO based MEC scheme is given as; $\widetilde{T}_k = \widetilde{T}_k^t + \widetilde{T}_k^c$.

Thus, the overall delay the massive MIMO based MEC scheme is expressed as;

$$T_{\text{m-MIMO}} = \max_{\forall k \in \mathcal{K}} \widetilde{T}_k. \tag{3.27}$$

3.2.3. Performance Evaluations

This section provides the simulation results to illustrate the performance of the proposed massive MIMO-NOMA assisted based MEC framework compared with the massive MIMO based MEC and the OMA-massive MIMO based MEC systems.

Table 3.1. Simulation parameters.

Parameter	Value
K	16
B	1 MHz
P	10 dBm
F_{max}	20 GHz
C_{mec}	100 cycles/bit
$D_{m,1}$	1 Mbits
ν_{max}	5 sec

For the massive MIMO based MEC system, the same transmit powers as in the proposed scheme are used, while for the OMA-massive MIMO based MEC the power of all users is chosen as P, as in (Kim et al., 2015). Consequently, the OMA-massive MIMO

based MEC uses approximately 63% higher average transmit power than the proposed massive MIMO-NOMA based and massive MIMO based MEC schemes. In figures, massive MIMO is abbreviated as "*M-MIMO*".

The system parameters are given in Table 3.1. The users are uniformly located in the considered area within a radius of 300 m. The noise power spectral density is -174 dBm/Hz. The path loss is determined by $L_{m,j} = 30.6 + 36.7 \log_{10}(d_{m,j}\,(m))$ where $d_{m,j}$ corresponds to the distance between the user and the BS (Zeng et al., 2020).

Fig. 3.5 provides the sum data rate of MEC systems for the different numbers of antennas at the BS for K=16. As shown in the figure, the proposed massive MIMO-NOMA based MEC achieves a higher sum data rate than the massive MIMO based MEC and OMA-massive MIMO based MEC for all N values. Specifically, for N=128, the proposed massive MIMO-NOMA based MEC achieves 65 Mbps, while the massive MIMO based MEC provides 62 Mbps and the OMA-massive MIMO based scheme enables 26 Mbps. Similarly, for N=512, the proposed MEC achieves around 8 Mbps and 47 Mbps higher sum data rates than its counterparts with the massive MIMO based MEC and the OMA-massive MIMO based MEC, respectively. The simulation results show the advantages of NOMA in the MEC system based on achievable data rates.

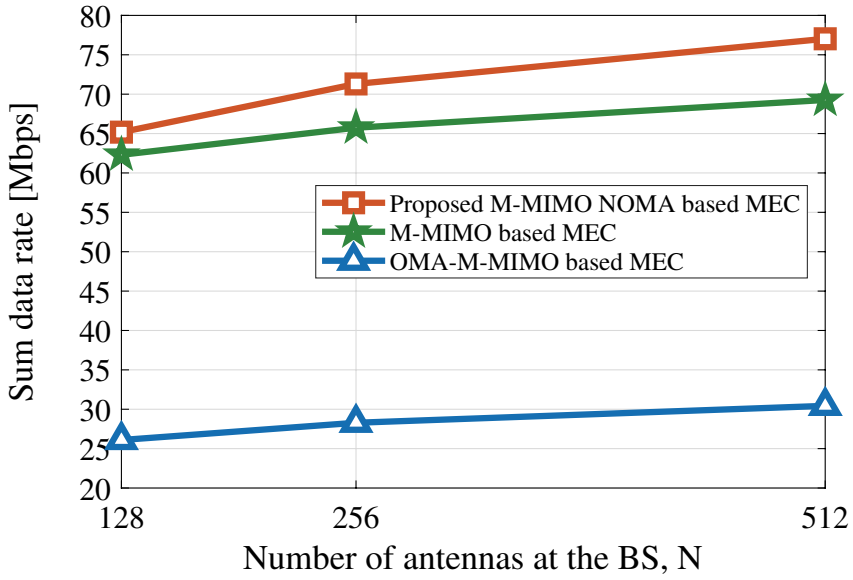


Figure 3.5. The sum data rate versus the number of antennas, N, for K=16, M=8 and P=10 dBm.

Fig. 3.6 presents the overall delay versus the number of antennas at the BS for K=16. The overall delay decreases as the number of antennas increases since the data rates of strong users are increased in the massive MIMO system, which reduces the transmission delay. Specifically, the proposed MEC with N=512 reduces the overall delay by 16% compared with the case of N=128 by serving multiple users simultaneously. Moreover, the proposed massive MIMO-NOMA based MEC outperforms the massive MIMO based MEC and the OMA-massive MIMO based MEC in terms of the overall delay for all N values, e.g., the proposed framework reduces the overall delay by 263 ms and 227 ms compared with the OMA-massive MIMO based MEC for N=128 and N=512, respectively. Accordingly, the proposed massive MIMO-NOMA based MEC reduces the overall delay by 5% and 10% compared with the massive MIMO based MEC for N=128 and N=512, respectively. These performance results confirm the benefits of the proposed joint massive MIMO and NOMA based MEC system in terms of the overall delay.

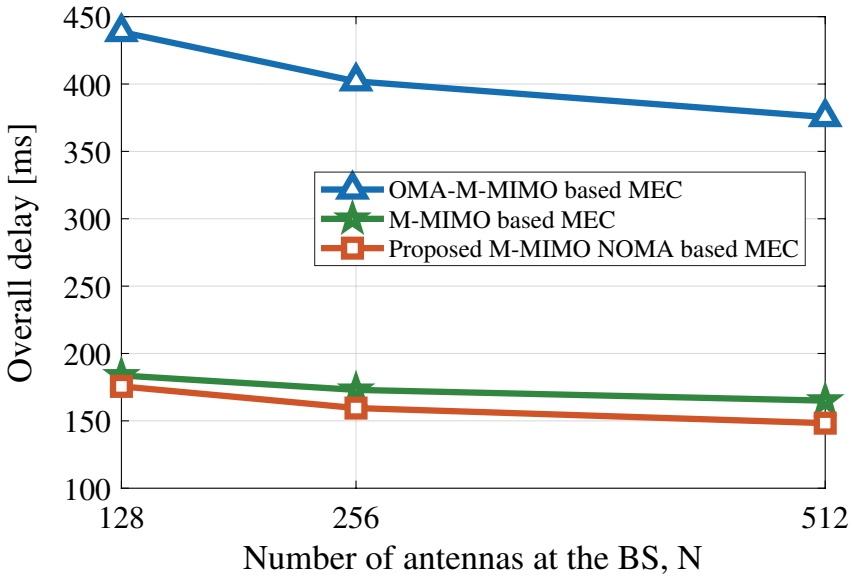


Figure 3.6. The overall delay versus the number of antennas, N, for K=16, M=8 and P=10 dBm.

In Fig. 3.7, the overall delay versus the maximum transmit power, P, is shown for N=256 and K=16. When the maximum transmit power is increased, the data rate of strong users significantly increases, resulting in a reduction in transmission delay and, thus overall delay. The proposed MEC framework reduces the overall delay by 9% at P=0 dBm and 8% at P=10 dBm compared with the massive MIMO based MEC scheme.

Similarly, the overall delay of the proposed system is reduced by 62% at P=0 dBm and 60% at P=10 dBm compared with the OMA-massive MIMO based MEC.

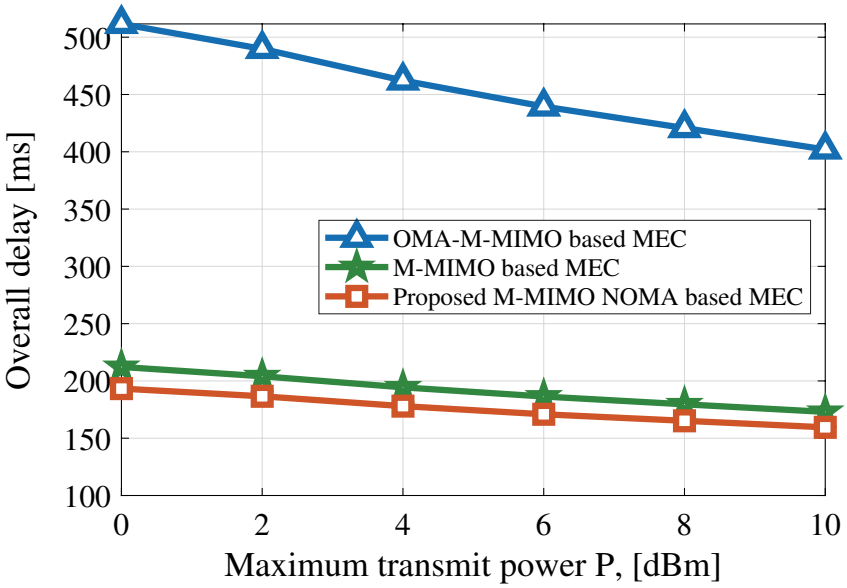


Figure 3.7. The overall delay versus the maximum transmit power, P for N=256, K=16, M=8.

Table 3.2 gives the average transmit power per strong user, P_s, and per weak user, P_w, for the different number of antennas at P=10 dBm. It is observed that the average transmit power of strong user is higher than those of weak user. Furthermore, the average transmit power belonging to strong users increases when the number of antennas is increased, resulting in an increased data rate of the strong users and, thus reducing transmission delay.

Table 3.2. The average transmit power per strong user and per weak user for N and K=16, M=8, P=10 dBm.

N	P_s [dBm]	P_w [dBm]
128	8.48	-3.14
256	8.50	-4.63
512	8.56	-6.16

Fig. 3.8 investigates the effect of the number of users, K, on the overall delay for N=256. With the increasing number of users, the CPU frequency allocated by MEC to the m^{th} cluster for the proposed massive MIMO-NOMA based MEC and k^{th} user for the OMA-massive MIMO based MEC and massive MIMO based MEC systems decreases. This causes an increasing computing delay at the MEC server. Thus, it results in a higher overall delay. In particular, the proposed framework reduces the overall delay by 207 ms and 308 ms compared with the OMA-massive MIMO based MEC when K=8 and K=32, respectively. Accordingly, the proposed massive MIMO-NOMA based MEC reduces the overall delay by 6% and 10% compared with the massive MIMO based MEC when K=8 and K=32, respectively.

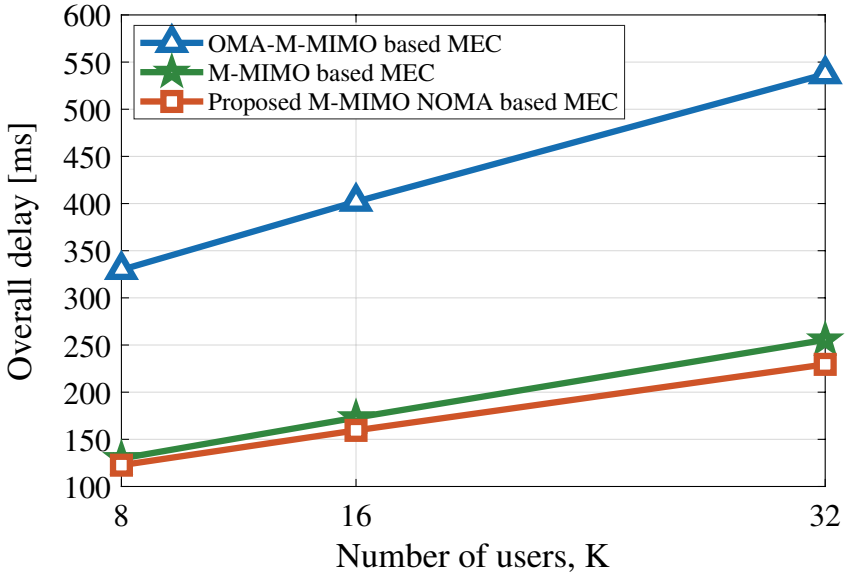


Figure 3.8. The overall delay versus the number of users, K for N=256 and P=10 dBm.

3.3. User Selection for Delay Minimization in MEC

Apart from the previous section, in this section, we perform a user selection approach to reduce the overall delay significantly and implement 2D channel model as given in (2.2). Our aim is to demonstrate the effect of the user selection method on the delay minimization for the massive MIMO-NOMA based MEC system.

We consider the uplink massive MIMO-NOMA assisted MEC system for remote computing in which a BS is equipped with an antenna array of N elements and K single-antenna users, as given in Fig. 3.3. As the proposed user-set selection method, we apply Algorithm 2. Here, the only difference in the algorithm is that the NOMA based massive MIMO transmission is performed in Step 7b for the selected users instead of the OMA-based MIMO transmission. Thus, the selection of U users among K total users is determined under the assumption that $U << N$, and these U users form two sets, including $U/2$ users in each. After we select the U users, we consider the case of two users in each cluster for NOMA having totally M clusters with $m \in \{1, \ldots, M\}$. Then, the number of clusters, M, is determined as $U/2$. Accordingly, in the proposed system, the High-High channel gain user pairing strategy is applied to pair these U users.

3.3.1. Performance Evaluations

This section provides the simulation results to show the performance of the proposed massive MIMO-NOMA based MEC framework compared with the massive MIMO based MEC and the OMA-massive MIMO based MEC systems. The same system parameters are implemented as in Section 3.2.3. We perform the user selection algorithm among K=300 users.

Fig. 3.9 explores the effects of the different threshold β on the sum data rate for the proposed massive MIMO-NOMA based MEC system versus the different number of antennas at the BS. It is shown that β =0.1 has the highest sum data rate for all N values. Besides, it is observed that utilizing higher number of antennas increases the sum data rate by achieving antenna gain.

Fig. 3.10 compares the average sum data rate of the proposed massive MIMO-NOMA based MEC, the massive MIMO based MEC and the OMA-massive MIMO based MEC systems for β =0.1 and U=16. The proposed massive MIMO-NOMA based MEC significantly outperforms the conventional schemes. For example, at N=128, the proposed system achieves about 90 Mbps, while the massive MIMO based MEC and the OMA-massive MIMO based MEC systems only achieve 71 Mbps and 35 Mbps, respectively. In addition, the proposed algorithm provides a 20 Mbps and 58 Mbps higher sum data rate than the massive MIMO based MEC and the OMA-massive MIMO based MEC, respectively, when N=512.

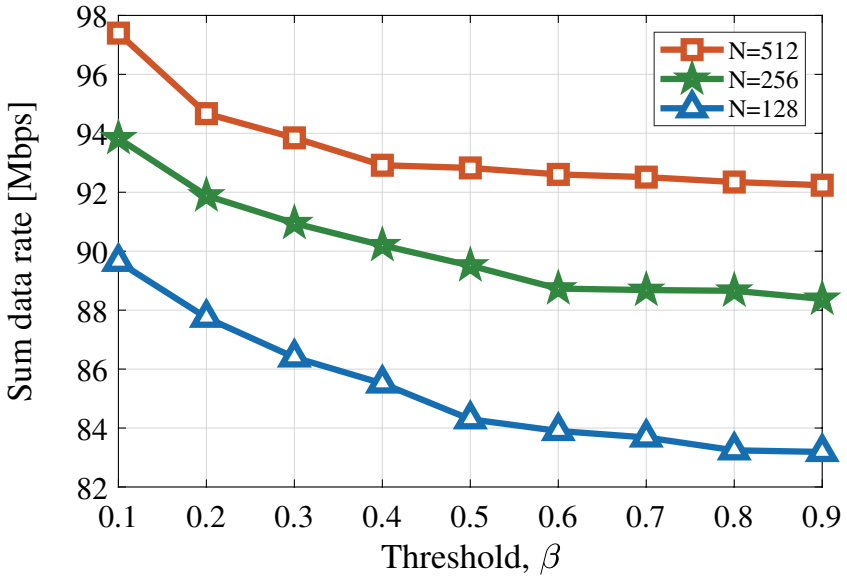


Figure 3.9. The sum data rate for the proposed massive MIMO-NOMA based MEC system versus the threshold β for different N, U=16, K=300, P=10 dBm.

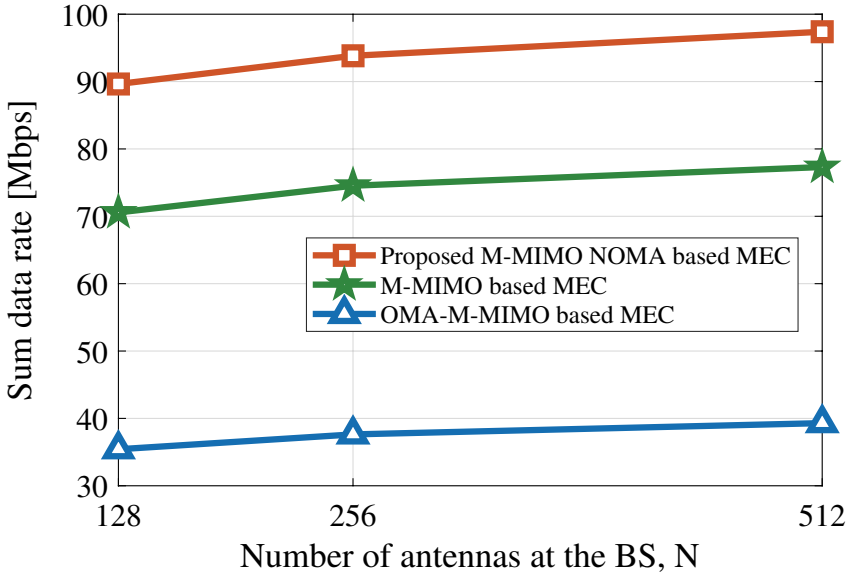


Figure 3.10. The sum data rate versus the number of antennas, N, for U=16, K=300, β =0.1 and P=10 dBm.

In Fig. 3.11, we investigate the impact of the user selection strategy on the overall delay in the proposed massive MIMO-NOMA based MEC system for the different numbers of BS antennas, N. We observe that for different N, the proposed MEC system with user selection significantly reduces overall delay compared to the case U=16, K=16. For example, when N=128, the overall delay with the user selection decreases by 99 ms; when N=512, that reduces by 61 ms compared to the case U=16, K=16. Thus, it is observed that applying the user selection algorithm improves the MEC performance in terms of the overall delay.

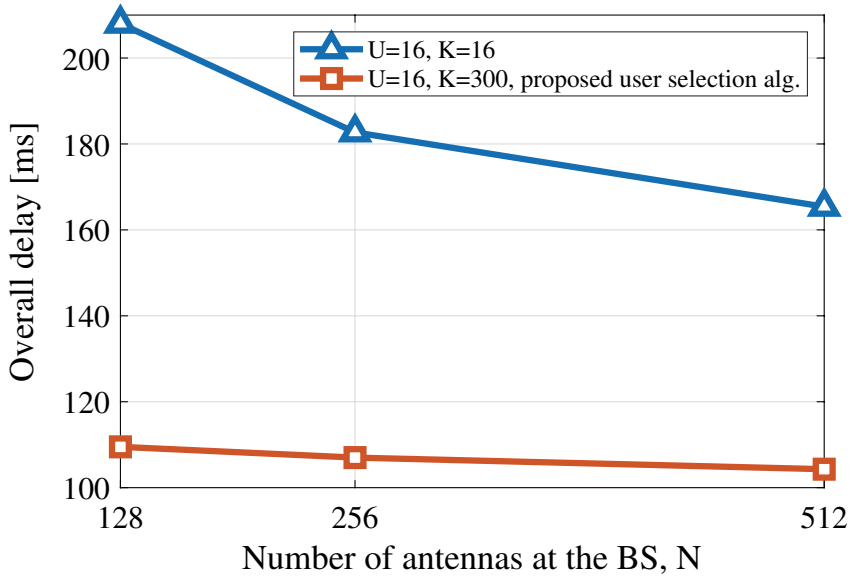


Figure 3.11. The effect of the user selection on the overall delay for the proposed massive MIMO-NOMA based MEC with different N at P=10 dBm.

Fig. 3.12 demonstrates the overall delay versus the number of antennas at the BS for the proposed massive MIMO-NOMA based MEC, the massive MIMO based MEC and the OMA-massive MIMO based MEC systems with the selected U=16 users among K=300. The overall delay decreases as the number of antennas increases since the data rates of strong users are improved, which reduces the transmission delay. Specifically, the proposed MEC with N=512 reduces the overall delay by 5% compared to N=128 by simultaneously serving the same number of users. Furthermore, the proposed massive MIMO-NOMA based MEC outperforms the massive MIMO based MEC and the OMA-massive MIMO based MEC in terms of the overall delay for all N values. In particular, the proposed framework reduces the overall delay by 168 ms and 30 ms compared with

the OMA-massive MIMO based MEC and the massive MIMO based MEC for N=128, respectively. When the number of antennas is increased from 128 to 512, the overall delay in the proposed MEC system decreases by 154 ms and 27 ms compared with the OMA-massive MIMO based MEC and the massive MIMO based MEC, respectively.

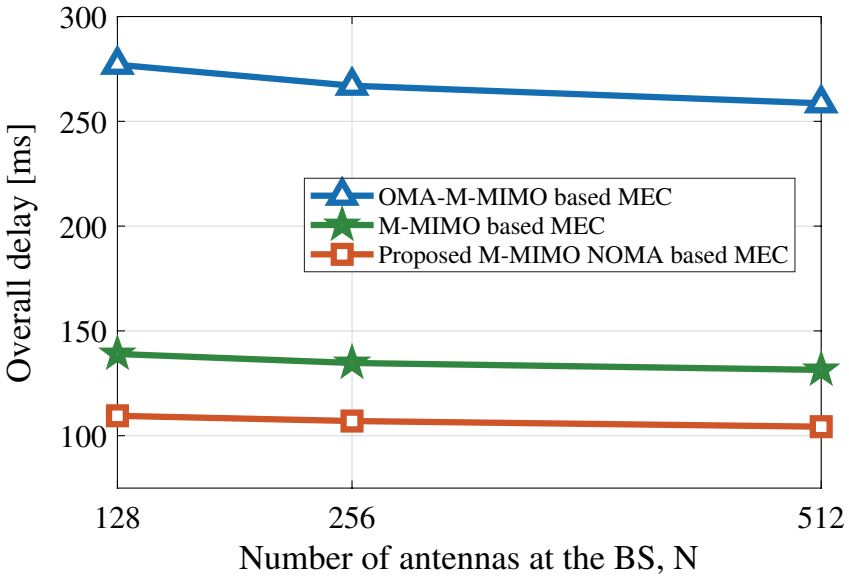


Figure 3.12. The overall delay versus the number of antennas at the BS, N for U=16, K=300 and P=10 dBm.

Table 3.3 presents the average transmit power per strong user, P_s, and per weak user, P_w, considering the different number of antennas for U=16, K=300 and P=10 dBm. The average transmit power of strong user is higher than those of weak user. Moreover, the average transmit power of strong users increases when the number of antennas is incremented from N=128 to N=512. Thus, it results in an increased data rate for strong users and reduces transmission delay.

Table 3.3. The average transmit power per strong user and per weak user for N and U=16, K=300, P=10 dBm.

N	P_s [dBm]	P_w [dBm]
128	8.64	-2.49
256	8.65	-5.05
512	8.67	-6.74

In Fig. 3.13, we compare the performance of the overall delay when the number of the selected users, U, is varied at K=300 and N=256. The figure shows that the overall delay increases with the number of selected users among K=300. The reason is due to the increasing computing delay at the MEC server. Thus, the larger number of selected users results in a higher overall delay. Compared with the schemes in the massive MIMO based MEC and the OMA-massive MIMO based MEC, the proposed scheme achieves a lower overall delay for the different numbers of selected users. Specifically, for U=8, the proposed MEC reduces the overall delay by 14 ms and 124 ms compared to the massive MIMO based MEC and the OMA-massive MIMO based MEC, respectively. When U=32, the proposed MEC reduces the overall delay by 33 ms and 240 ms compared with the massive MIMO based MEC and the OMA-massive MIMO based MEC, respectively.

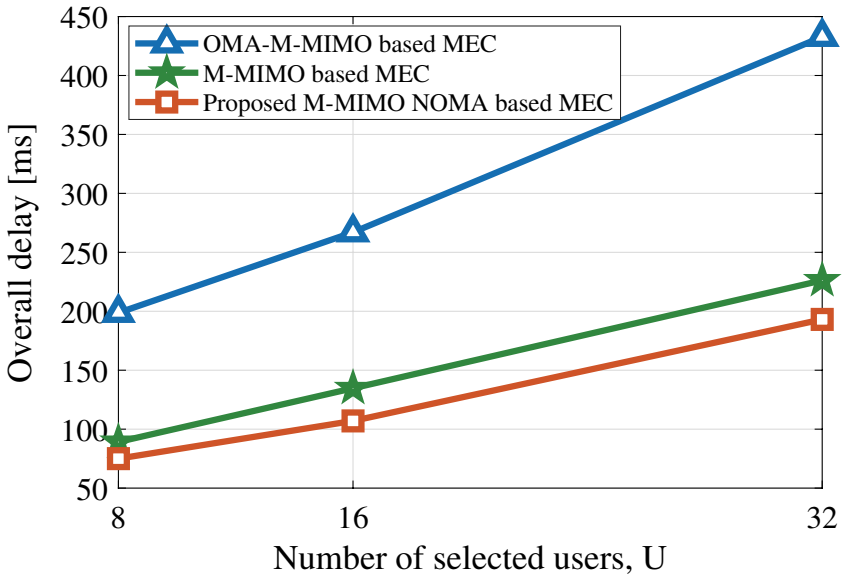


Figure 3.13. The overall delay versus the number of selected users, U for K=300, N=256 and P=10 dBm.

3.4. Conclusion

In this chapter, the joint NOMA and massive MIMO assisted MEC system with a remote computing scheme has been proposed for delay-sensitive applications. We have shown that the combination of MEC with NOMA improves the system performance by

simultaneously serving K users with N antennas. Moreover, by combining massive MIMO and MEC technologies, more users can offload computational-intensive tasks simultaneously to the MEC while reducing the overall delay. We have formulated the overall computing and transmission delay minimization problem for massive MIMO NOMA assisted MEC systems. As a result, the proposed framework enables both cell-center and cell-edge users to offload their tasks to the MEC server by applying an efficient user pairing, offloading and computation scheme. The simulation results verify the benefits of the proposed joint massive MIMO and NOMA with the MEC system. Furthermore, the proposed algorithm has been extended to the densely deployed scenario for the NOMA and massive MIMO assisted MEC system, and a user selection algorithm has been applied. Thus, the overall delay has been reduced through the user selection approach.

CHAPTER 4

COOPERATIVE MEC SYSTEMS

In this section, cooperative MEC systems are introduced for the cell-edge users far from the MEC server to improve the offloading performance of these users. The cell-center users in the system ensure cooperation by offloading cell-edge users' tasks to the MEC server and computing some parts of them locally. Following this idea, cooperative NOMA-MEC systems are proposed with a single antenna to maximize the total offloading data (Yılmaz and Özbek, 2022) and with massive MIMO to minimize the overall delay while ensuring security (Yılmaz et al., 2023).

4.1. Multi-helper NOMA for Cooperative MEC

The next-generation wireless systems are expected to support a number of computation intensive and delay-sensitive applications. Since many devices are computation and power limited, MEC has been deemed as a promising way to enhance computation service. However, there are some main cases that need to be overcome. Firstly, the computation resource at the BS cannot be always sufficient to support all devices in densely deployed scenario. Secondly, there could be no strong direct transmission link to the BS.

Considering the above issues, in this section, we propose a novel cooperative MEC that exploits the combination of NOMA and multiple helpers (Yılmaz and Özbek, 2022). The motivation of employing the NOMA technology is to reduce the offloading latency and improve the performance of the MEC based systems. Therefore, these two communication techniques, MEC and NOMA, can be combined to provide gains in terms of the total offloading data and the latency performance.

In the proposed system including a user, multiple helpers and a BS, the user can simultaneously offload its computation-intensive tasks to the helpers using NOMA when there is no strong direct transmission link between the user and the BS. Then, the helpers can compute and offload these tasks through NOMA. Thus, in the proposed scheme, the computation and offloading modes at the helpers are determined with respect to the optimized task offloading decision factor. The simulation results show that the proposed NOMA based cooperative MEC significantly increases the total offloading data under the

latency constraints compared to the benchmark schemes including one helper with strong direct transmission link.

State-of-the-art on Computation Offloading in MEC:

Recently, several works have been published addressing the issue of latency minimization, task offloading, resource management and energy consumption in MEC-enabled systems.

Motivated by the performance gains of applying NOMA over OMA, both NOMA uplink and downlink transmissions have been applied in (Ding et al., 2019a) to reduce the latency and energy consumption considering NOMA based MEC offloading. The offloading delay and energy minimization have been respectively optimized in (Ding et al., 2018) and in (Ding et al., 2019b) for the different users with different computation requirements in NOMA based MEC systems. In (Ye et al., 2019), a problem has been formulated to maximize the probability of successful computation by jointly optimizing the offloading time consumption, the power allocation and the offloading ratios for NOMA based MEC network, where two users may offload their computation tasks to a MEC server. However, in (Ding et al., 2018) (Ding et al., 2019a) (Ding et al., 2019b) (Ye et al., 2019), only one cluster of two users has been considered and the resource allocation among the different clusters of users forming NOMA has been ignored.

Since each resource is suggested to be multiplexed by a small number of users due to decoding complexity and error propagation, the importance of resource allocation among the different clusters of users forming NOMA has been stated in (Zafar et al., 2013) (Wan et al., 2019). Thus, the completion time and the total energy consumption minimization problem have been studied in (Yang et al., 2019a) for an uplink NOMA based MEC system by considering resource allocation for the different clusters with two users to perform NOMA. In (Pan et al., 2019), the total energy consumption has been minimized considering the NOMA based transmission in both task uploading and downloading. In addition, the authors in (Mao et al., 2019) have presented a total energy consumption minimization problem while achieving the computation latency constraint for a NOMA-assisted MEC system.

Although the above studies have demonstrated the effectiveness of MEC through NOMA in enhancing the computation performance of wireless networks, the computation resource at the BS cannot be always sufficient to support all devices in a densely deployed scenario, there may be a cell-edge user with computationally intensive tasks suffering from a low SNR, or there may not be a strong direct transmission link to the BS. To deal with this issue, a potential solution is to offload part of the computationally

intensive tasks to helpers via NOMA. In the light of this solution, in (Cao et al., 2019), a cooperative edge computing in both computation and communication has been given for a basic three-node MEC system consisting of one user node, one helper node and one AP node with a MEC server integrated. The main objective is to minimize the total energy consumption while satisfying the user's computation latency constraint based on the TDMA transmission protocol. In (Liu, 2019), a basic three-node MEC with NOMA based cooperative edge computing has been presented to maximize sum offloading data subject to the latency constraints. Similarly, the authors of (Li et al., 2021) have represented a NOMA-aided user cooperation scheme in a three-node MEC wireless power transfer (WPT) system based on the energy consumption minimization problem. The only difference between the system models in (Liu, 2019) and (Li et al., 2021) is the available computation tasks at the helper. The helper does not have computation tasks in (Li et al., 2021). However, these studies (Cao et al., 2019) (Liu, 2019) (Li et al., 2021) are based on only one helper scenario. In addition to that, the user can directly offload its own tasks to the BS in these three-node MEC systems. On the other hand, different from the basic three-node MEC system, a cooperative task computation framework has been considered in (Pan et al., 2021) with the purpose of maximizing the number of accomplished tasks and minimizing the power consumption of users. In this cooperative task computation system between the user and the BS, one user can help the other user for task computation via D2D transmissions.

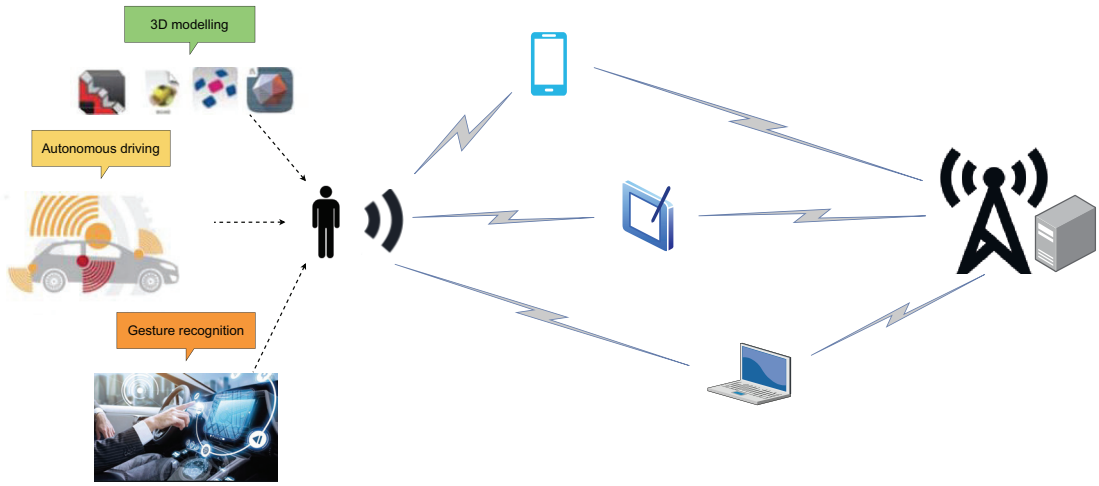


Figure 4.1. The scenario of the cooperative MEC based on NOMA with multiple helpers.

Motivated by the above discussions, as shown in Fig. 4.1, the user can have computation-intensive mobile applications, such as augmented reality, virtual reality, autonomous driving, gesture recognition and three-dimension (3D) modeling. Also, these tasks can also be executed under latency constraints. Therefore, it is a challenging task for the devices to handle these intensive computation loads with the latency requirements. If there were a strong direct transmission link between the user and the MEC server namely the three-node MEC system, the user would offload its computation-intensive and delay-sensitive tasks to a nearby MEC server for remote task execution. However, for the user which is at the cell edge, there is no strong direct transmission link between the user and the MEC server. Therefore, the cell-edge user chooses to offload its computationally intensive and latency-critical tasks to the server through the aid of helpers.

In this section, by considering the state-of-the-art work listed above, we propose a new paradigm of cooperative MEC with multiple helpers based on NOMA. In the proposed framework, a user simultaneously offloads its tasks using NOMA to many helpers at the first slot. Then, the helpers can both compute and offload the user's tasks at the second slot. NOMA is adopted for offloading in both time slots to increase the total offloading data. Furthermore, since the distances between the user and the helpers affect the maximization of the total offloading data, adjusting the distances is important while achieving latency constraints of the user's applications. As a result, the communication resource optimization is performed to improve the user computation experience through emerging MEC systems.

To be specific, we propose a multi-helper cooperative MEC system based on NOMA to maximize the total offloading data under the latency and power constraints while in the literature (Cao et al., 2019), (Liu, 2019) and (Li et al., 2021), only one helper case for the MEC system has been considered and the NOMA scheme is only applied for the link between the user and the helper while it is not used for the link between the helper and the BS. Afterward, we analyze the proposed framework under the optimum distances of the multiple helpers to maximize the total offloading data. We provide simulation results to show the superiority of the proposed framework in terms of the total offloading data compared with the benchmarking solutions in (Cao et al., 2019), (Liu, 2019) and (Li et al., 2021).

4.1.1. System Model

We consider a NOMA based cooperative MEC system which consists of a BS integrated with a MEC server, one user and K helpers, as illustrated in Fig. 4.2. Let $\mathcal{K} = \{1, 2, \ldots, K\}$ denote the set of helpers. In the system, all helpers have the same hardware specifications. These helper nodes can be a laptop, or a tablet, which have certain computation and communication resources.

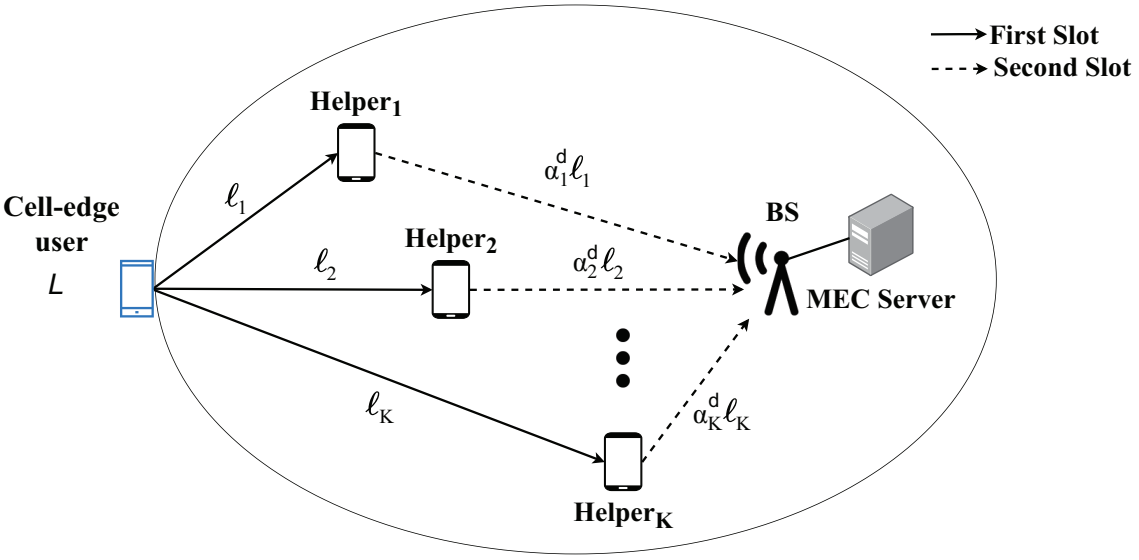


Figure 4.2. The proposed cooperative MEC based on NOMA with K helpers.

We assume that the user has individual computation tasks with data size, L, to complete successfully under a common latency constraint, but the helpers, Helper$_k$ $\forall k \in \mathcal{K}$, do not have computation tasks. The BS is integrated with the MEC server to execute the computation-intensive tasks offloaded by the helpers. It is assumed that there is no strong direct transmission link between the user and the MEC server since the user is at the cell edge. The user sends a certain part of its tasks to the MEC server through the helpers. Besides, these helpers are at the cell-center.

We consider the partial computation offloading operation mode that assumes the computational tasks can be divided into two independent parts. One part is executed locally, while the other part is offloaded to the helpers. In Fig. 4.2, the user simultaneously offloads ℓ_k part of its own data L, to kth helper and locally computes the rest of the data.

Each helper computes and offloads the tasks received from the user. The Helper$_k$ uses task offloading decision factor α_k^d to decide the fraction of the offloading and the computing modes. That is, for a particular time instant, the Helper$_k$ can have $0 < \alpha_k^d < 1$ to act the double modes. In the system, α_k^d is obtained through the optimization algorithm and cannot be 0 or 1, which leads to the helpers working in cooperative mode while executing a certain portion of the tasks and establishing transmission to offload the remaining portion of the tasks.

For a cooperative edge computing system, the cooperation between the user and the helpers is essential. The main motivation of the proposed cooperative MEC system is that, for a common latency constraint T, the user offloads its own tasks to the MEC server with the aid of K helpers. Latency constraint, T, is divided into two time slots such as t_u and t_h as shown in Fig. 4.3. In the first time slot, t_u, the user offloads the tasks to the helpers. Then, the helpers offload and compute a certain part of the user's tasks in the second time slot, t_h.

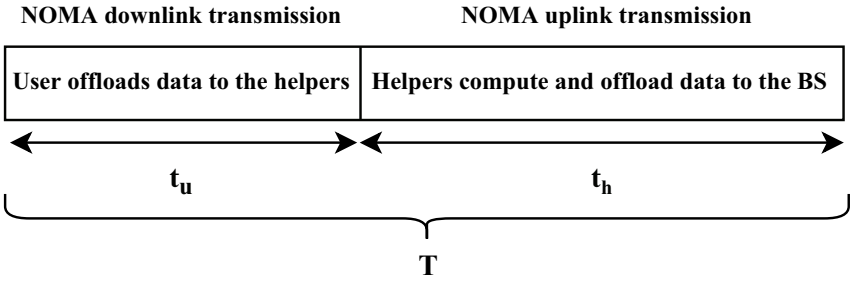


Figure 4.3. Latency constraint scheme.

In the first time slot which performs NOMA downlink transmission, without loss of generality, K helpers are sorted as $g_1 \geq g_2 \geq \ldots \geq g_K$ where g_k is the channel gain between the user and the Helper$_k$, $\forall k \in \mathcal{K}$. The channel amplitude is modeled by using Rayleigh distribution with a variance of distance-dependent path loss coefficient. The corresponding CSI is assumed to be available at each receiving node.

In the first time slot t_u, the offloaded data from the user to the Helper$_k$ is calculated as $\ell_k = t_u R_k^D$, and the average data rate R_k^D is given by (Manglayev et al., 2016) (Huang and Liu, 2018) (Liu, 2019) (Zhu et al., 2019) (Glei and Belgacem Chibani, 2019) (Huang et al., 2020) (Zhu et al., 2020) and defined as follows;

$$R_k^D = B\ \mathbb{E}\left\{\log_2\left(1+\frac{P_k^D\, g_k}{g_k \sum\limits_{j=1}^{k-1} P_j^D + \sigma_d^2}\right)\right\} \tag{4.1}$$

where P_k^D denotes the downlink transmit power for the Helper$_k$, and σ_d^2 is a variance of zero-mean complex AWGN, where $\sigma_d^2 = N_0\, B$ with N_0 is the noise power spectral density.

In the second time slot which performs NOMA uplink transmission, depending on α_k^d values, all helpers simultaneously offload their data, $\ell_{k,o} = \alpha_k^d\,(\ell_k)$, to the BS relying on the uplink NOMA scheme. Under this scheme, the helpers are sorted based on their channel gains, namely $g_{K,o} \geq g_{K-1,o} \geq \ldots \geq g_{1,o}$ where $g_{k,o}$ is the channel gain between the Helper$_k$ and the BS. Then, the BS utilizes the SIC technique to decode the data coming from the helpers. According to the principle of the SIC, the BS first decodes the information from the helper with the larger channel gain and then removes it from the other helpers' signals. Therefore, the average data rate of the Helper$_k$ is given by (Zuo and Tao, 2017) (Yang et al., 2018) (Mao et al., 2019) (Azam et al., 2019);

$$R_k^U = B\ \mathbb{E}\left\{\log_2\left(1+\frac{P_k^U\, g_{k,o}}{\sum\limits_{j=1}^{k-1} P_j^U\, g_{j,o} + \sigma_u^2}\right)\right\} \tag{4.2}$$

where P_k^U is the uplink transmit power of the Helper$_k$. σ_u^2 is a variance of zero-mean complex AWGN, where $\sigma_u^2 = N_0\, B$.

4.1.2. Proposed Framework and Problem Formulation

In this section, a total offloading data (TOD) maximization problem with given constraints is formulated for the proposed cooperative MEC based on NOMA with the K helpers. In the proposed framework, since there is no strong direct transmission link between the cell-edge user and the MEC server, the user sends a certain part of its tasks to the MEC server through the cell-center helpers. The helpers do not act as a pure relay since they can compute some portions of the offloaded tasks from the cell-edge user and

also offload some portions of these computation tasks to a MEC server. In the proposed approach, we only focus on the offloading part at the user side rather than computing since our aim is to maximize TOD.

The objective function is TOD that is sum of the user's and helpers' offloading data and defined as function $w(.)$ given below

$$w(\mathbf{t}, \mathbf{P}) = \sum_{k=1}^{K} \left(t_u R_k^D + t_h R_k^U \right) \tag{4.3}$$

Since the data $\alpha_k^d \left(t_u R_k^D \right)$ will be offloaded from the Helper$_k$ to the BS in the uplink transmission, the function $w(.)$ can be rewritten as the following function f(.);

$$f(\mathbf{t}, \mathbf{P}, \boldsymbol{\alpha}^d) = \sum_{k=1}^{K} \left(t_u R_k^D + \alpha_k^d (t_u R_k^D) \right) \tag{4.4}$$

Accordingly, TOD maximization problem subject to the latency constraints and power allocation factors is given by;

$$\max_{\mathbf{t}, \mathbf{P}, \boldsymbol{\alpha}^d} \quad f(\mathbf{t}, \mathbf{P}, \boldsymbol{\alpha}^d) \tag{4.5}$$

$$\text{s.t.} \quad R_k^D \geq R_{th,k}^D, \quad \forall k \in \mathcal{K} \tag{4.5a}$$

$$\sum_{k=1}^{K} P_k^D \leq P_u \tag{4.5b}$$

$$P_k^U \leq P_{max}^U, \quad \forall k \in \mathcal{K} \tag{4.5c}$$

$$t_u + t_h \leq T \tag{4.5d}$$

$$\frac{\left(1 - \alpha_k^d\right) t_u \, R_k^D \, \theta}{f_k} \leq T - t_u, \quad \forall k \in \mathcal{K} \tag{4.5e}$$

$$0 < \alpha_k^d < 1, \quad \forall k \in \mathcal{K}. \tag{4.5f}$$

where $\mathbf{t} = [t_u, t_h]$, $\mathbf{P} = \left[P_1^D, \ldots, P_K^D, P_1^U, \ldots, P_K^U\right]$ and $\boldsymbol{\alpha}^d = \left[\alpha_1^d, \alpha_2^d, \ldots, \alpha_K^d\right]$, P_u is the maximum total transmit power of the user, and P_{max}^U is the maximum transmit power of each helper. θ is the density/complexity of the task computing (i.e., the number of CPU cycles/bit) and represents the amount of CPU cycles needed for computing one bit.

f_k represents the local computation capability of each Helper$_k$, also known as the CPU frequency.

Constraint (4.5a) gives the minimum data rate constraints in downlink that should not be less than a given threshold rate $\mathrm{R}_{th,k}^{D}$. Constraints (4.5b) and (4.5c) represent the power constraints in NOMA downlink and NOMA uplink transmissions, respectively. Constraint (4.5d) shows the total offloading time constraint for both the user and the helpers. Constraint (4.5e) represents the local computing time constraint of each helper. Constraint (4.5f) indicates the range of the task offloading decision factor for each helper.

In the proposed approach, we firstly investigate the optimum distances of the helpers which maximize TOD assuming there is no energy restriction on the helpers. After determining the optimum distances, we select the helpers having these distances. Then, the optimization algorithm is applied to find the optimal time, power allocation factors and the task offloading decision factor.

4.1.3. Problem Solution

In this section, we derive the solution to Problem (4.5) for the proposed system with the K helpers NOMA based cooperative MEC. Firstly, we give the solution for the case of K=2 and then generalize to the K helpers. Since the Problem (4.5) is a constrained nonlinear multivariable problem, we apply the interior point method. In this method, violation of inequality constraints is prevented by adding a barrier term to the objective function that ensures the optimal unconstrained values to be in the feasible space.

NOMA downlink transmission: Since Problem (4.5) is an increasing function with respect to P_{k}^{D}, the constraint associated with the downlink transmit power can be written as $P_1^D + P_2^D = P_u$ to maximize the objective function. Thus, in the downlink transmission, $\xi \in (0, 1)$ is the power allocation factor and becomes one of the optimization parameters in the Problem (4.5). Then, the allocated downlink transmit power, P_{1}^{D}, for the Helper$_1$ is determined by ξP_u while the allocated downlink transmit power, P_{2}^{D}, for the Helper$_2$ is calculated as $(1 - \xi)\, P_u$.

The average achievable rate of the Helper$_k$ in an OMA system (Zhu et al., 2019) is given by;

$$R_k^{OMA} = \frac{B}{2}\, \mathbb{E}\left\{\log_2\left(1 + \frac{P_u\, g_k}{\sigma_d^2}\right)\right\} \tag{4.6}$$

where the factor $\frac{1}{2}$ is due to the fact that conventional OMA results in a multiplexing loss of $\frac{1}{2}$.

In the downlink transmission, the achievable rate in the NOMA system should be no less than in the OMA system (Zhu et al., 2019). In this case, $\mathrm{R}_{th,k}^{D}$ is set as R_{k}^{OMA}. Then, the range of ξ can be obtained from the constraint (4.5a) directly as follows;

$$
\begin{aligned}
{R_1}^{D} &\geq {R_1}^{OMA} \\
\log_2\left(1+\frac{\xi P_u\, g_1}{\sigma_d^2}\right) &\geq \frac{1}{2}\log_2\left(1+\frac{P_u\, g_1}{\sigma_d^2}\right) \\
\xi &\geq \frac{\left(\sqrt{1+\frac{P_u g_1}{\sigma_d^2}}-1\right)\sigma_d^2}{P_u\, g_1}
\end{aligned}
\tag{4.7}
$$

$$
\begin{aligned}
{R_2}^{D} &\geq {R_2}^{OMA} \\
\log_2\left(1+\frac{(1-\xi)\, P_u\, g_2}{\xi P_u\, g_2+\sigma_d^2}\right) &\geq \frac{1}{2}\log_2\left(1+\frac{P_u\, g_2}{\sigma_d^2}\right) \\
\xi &\leq \frac{\left(\sqrt{1+\frac{P_u g_2}{\sigma_d^2}}-1\right)\sigma_d^2}{P_u\, g_2}
\end{aligned}
\tag{4.8}
$$

Thus, the range of ξ is given by;

$$
\frac{\left(\sqrt{1+\frac{P_u g_1}{\sigma_d^2}}-1\right)\sigma_d^2}{P_u\, g_1} \leq \xi \leq \frac{\left(\sqrt{1+\frac{P_u g_2}{\sigma_d^2}}-1\right)\sigma_d^2}{P_u\, g_2} \tag{4.9}
$$

Then, we can extend the downlink power allocation factors to the K helpers. Assume that n, k represent the index of each helper, which provides cooperation between the user and the MEC server. The superscript of $\xi_{*}^{n,k}$ denotes the indices of the pairing

helpers and the upper bound of $\xi_k^{n,k}$ is given as z_k (Zhu et al., 2019). Thus, it is written as;

$$\begin{aligned}
\xi_2^{1,2} =& \frac{\left(\sqrt{1+\frac{P_u g_2}{\sigma_d^2}}-1\right)\sigma_d^2}{P_u\, g_2} \triangleq z_2 \\
\xi_3^{1,3} =& \xi_3^{2,3} = \frac{\left(\sqrt{1+\frac{P_u g_3}{\sigma_d^2}}-1\right)\sigma_d^2}{P_u\, g_3} \triangleq z_3 \\
& \cdot \\
& \cdot \\
& \cdot \\
\xi_K^{1,K} =& \xi_K^{2,K} = \ldots = \xi_K^{K-1,K} = \frac{\left(\sqrt{1+\frac{P_u g_K}{\sigma_d^2}}-1\right)\sigma_d^2}{P_u\, g_K} \triangleq z_K
\end{aligned} \tag{4.10}$$

Considering the magnitude of the channel gain in the downlink transmission, the order of z_k in (4.10) for $\forall k \in \mathcal{K}$ is given as;

$$z_1 \leq z_2 \leq \ldots \leq z_K \tag{4.11}$$

where $z_1 = \frac{\left(\sqrt{1+\frac{P_u g_1}{\sigma_d^2}}-1\right)\sigma_d^2}{P_u\, g_1}$.

In this way, the range of the power allocation factors, ξ_k, of each helper is determined as;

$$z_{k-1} \leq \xi_k \leq z_{k+1} \tag{4.12}$$

where $1 \leq k \leq K-1,\ K \geq 3$ and z_0 denotes 0.

Hence, the downlink transmit power for each helper is written as $\xi_k P_u$, while the downlink transmit power for the Helper$_K$ is found from the expression as $P_u - \left(\sum_{k=1}^{K-1} \xi_k P_u\right)$.

It can be noticed that Problem (4.5) is non-decreasing with respect to t_u. In order to maximize the total offloading data, the total latency constraint, T, is the sum of t_u and t_h since the range of t_u is given in $0 < t_u < T$. Thus, the constraint (4.5d) becomes;

$$t_u + t_h = T \tag{4.13}$$

Accordingly, the constraint (4.5e) can be rewritten as

$$\frac{\left(1-\alpha_k^d\right) t_u \, R_k^D \, \theta}{f_k} \leq t_h, \quad \forall k \in \mathcal{K} \tag{4.14}$$

NOMA uplink transmission: In the second time slot t_h since the data $\alpha_k^d \left(t_u R_k^D\right)$ will be offloaded from the Helper$_k$ to the BS with data rate R_k^U, we can obtain the following equations;

$$\alpha_1^d \left(t_u R_1^D\right) = (T - t_u) \, B \, \log_2 \left(1 + \frac{P_1^U g_{1,o}}{\sigma_u^2}\right) \tag{4.15}$$

$$\alpha_2^d \left(t_u R_2^D\right) = (T - t_u) \, B \, \log_2 \left(1 + \frac{P_2^U g_{2,o}}{P_1^U g_{1,o} + \sigma_u^2}\right) \tag{4.16}$$

On the other hand, the uplink transmission power must be lower than or equal to P_{max}^U in the uplink as shown in constraint (4.5c). Thus, we can derive the nonlinear constraints related to the optimization Problem (4.5) for the uplink transmission powers, P_1^U and P_2^U, using (4.15) and (4.16) as follows;

$$\frac{\left(2^{\frac{\alpha_1^d \left(t_u R_1^D\right)}{t_h}} - 1\right) \sigma_u^2}{g_{1,o}} \leq P_{max}^U \tag{4.17}$$

$$\frac{\left(2^{\frac{\alpha_2^d \left(t_u R_2^D\right)}{t_h}} - 1\right) \left(2^{\frac{\alpha_1^d \left(t_u R_1^D\right)}{t_h}}\right) \sigma_u^2}{g_{2,o}} \leq P_{max}^U \tag{4.18}$$

Then, we can extend the uplink transmission powers given in constraint (4.5c) to the K helpers, $\forall k \in \mathcal{K}$ as;

$$\frac{\left(2^{\frac{\alpha_k^d \left(t_u R_k^D\right)}{t_h}} - 1\right) \prod_{i=1}^{k-1} \left(2^{\frac{\alpha_i^d \left(t_u R_i^D\right)}{t_h}}\right) \sigma_u^2}{g_{k,o}} \leq P_{max}^U \tag{4.19}$$

Furthermore, the range of the task offloading decision factor for each helper is determined to provide better cooperation through the helpers in the system. Thus, constraint (4.5f) is redefined as follows;

$$0.3 \leq \alpha_k^d \leq 0.7, \quad \forall k \in \mathcal{K}. \tag{4.20}$$

Accordingly, the optimization Problem (4.5) can be rewritten as follows;

$$\max_{\mathbf{t},\mathbf{P},\boldsymbol{\alpha}^d} \quad f(\mathbf{t}, \mathbf{P}, \boldsymbol{\alpha}^d) \tag{4.21}$$

$$\text{s.t.} \quad (4.12), (4.14), (4.19), (4.20). \tag{4.22}$$

It is also equivalent to minimizing -f(.), thus the corresponding optimization problem can be efficiently solved by using some standard nonlinear programming optimization tools (Grace, 1990). The minimum of a constrained nonlinear multivariate function can be determined using the interior-point method.

The details of the interior-point algorithm are summarized in Algorithm 4 for the multi-helper scenario. In the minimization problem, $\mathbf{x}$ is defined as a vector of the components; $\mathbf{x} = \left[t_u, \xi_k, \alpha_k^d\right], \forall k \in \mathcal{K}$. The vector $\mathbf{x}$ satisfying all the constraints is called a feasible solution to the Problem (4.21). The initial values, $\mathbf{x}^1$, are determined by defining lower and upper bounds range for each component t_u, ξ_k and α_k^d in $\mathbf{x}$.

Then, the auxiliary function with a barrier parameter μ is given as;

$$S_\mu(\mathbf{x}) = -f(\mathbf{x}) + \mu\, P(\mathbf{x}) \tag{4.23}$$

where P(.) is given by;

$$P(\mathbf{x}) = -\sum_{i=1}^{2K} \log\left[-G_i(\mathbf{x})\right], \text{ for } G_i(\mathbf{x}) < 0 \tag{4.24}$$

where $\mathbf{G} = [G_1, G_2, \ldots, G_{2K}]$ represents the nonlinear inequality constraints in (4.14) and (4.19).

Algorithm 4 Optimizing TOD Based on Interior-Point Algorithm for K Helpers

Input: $g_k, g_{k,o}$, $\mathbf{x} = \left[t_u, \xi_k, \alpha_k^d\right]; \forall k \in \mathcal{K}$

1: Rearrange the inequality constraints in (4.14) and (4.19) as $\mathbf{G} = [G_1, G_2, \ldots, G_{2K}]$, $G_i(.) \leq 0,\ i = 1, 2, \ldots, 2K$.

2: Reformulate the objective function as an auxiliary function by,

$$S_{\mu_j}(\mathbf{x}) = -f(\mathbf{x}) + \mu_j P(\mathbf{x})$$

where $P(\mathbf{x})$ is given (4.24).

3: Call ***Algorithm 3***

Output: $t_u^*, \xi_k^*, \alpha_k^{d*};\ \ \forall k \in \mathcal{K}$

As a result of Algorithm 4, the optimum output values as $t_u^*, \xi_k^*, \alpha_k^{d*}, \forall k \in \mathcal{K}$ are obtained in order to provide maximum total offloading data under the constraints.

4.1.4. Performance Evaluation

In this section, we present the simulation results to evaluate the performance of the proposed system. The simulation parameters are listed in Table 4.1.

Table 4.1. Simulation parameters.

Parameter	Value
Carrier frequency, f_c	3.5 GHz
Transmission bandwidth, B	1 MHz
Noise variance, σ_d^2, σ_u^2	-159 dBm
CPU frequency, f_k	1 GHz
Required CPU cycles of tasks, θ	1000 cycles/bit (Li et al., 2021)
Maximum total transmit power of the user, P_u	0.5 W (Liu, 2019)
Maximum transmit power of the helper, P_{max}^U	0.8 W (Liu, 2019)

For both the user and helpers, the channel is modeled by using Rayleigh fading components with distance-dependent path loss, whose parameters depend on whether the receiver is the BS or the helper.

When the receiver is the BS, the path loss model conforms to "3GPP TR 36.814, Table B.1.2.1-1, B.1.2.1-2, UMi (3GPP, 2010)" for the distance between the k^{th} helper and the BS, $d_{k,BS}[m]$.

$$L(d_{k,BS})[dB] = 36.7\log_{10}(d_{k,BS}) + 22.7 + 26\log_{10}(f_c) \tag{4.25}$$

When the receiver is the helper, the path loss model conforms to "3GPP TR 36.843, A.2.1.2, UMi (3rd Generation Partnership Project, 2014)" for the distance between the user and the corresponding Helper$_k$, $d_{u,k}[m]$.

$$L(d_{u,k})[dB] = \max(PL(d_{u,k}), PL\_B(d_{u,k})) \tag{4.26}$$

where

$$PL = 20\log_{10}(d_{u,k}) + 46.4 + 20\log_{10}(f_c/5) \tag{4.27}$$

$$PL\_B = (44.9 - 6.55\log_{10}(h_{MS}))\log_{10}(d_{u,k}) + 5.83\log_{10}(h_{MS}) + 14.78 + 34.97\log_{10}(f_c) \tag{4.28}$$

where h_{MS} is the device antenna height (Kyösti et al., 2008) (Meinilä et al., 2010).

The performance results of the proposed scheme are compared with the one helper based MEC systems in (Cao et al., 2019) and (Li et al., 2021). In Figures 4.6 and 4.7, the proposed system is labeled as "*two helpers with NOMA*", whereas the benchmark systems studied in (Cao et al., 2019) is labeled as "*one helper with TDMA*" and (Li et al., 2021) is labeled as "*one helper with direct link*". Moreover, TDMA based cooperative offloading scheme is also performed for the two helpers case, namely, "*two helpers with TDMA*". In the TDMA scheme, four time slots are needed to offload the user's tasks to the BS through the two helpers.

The distance between the user and the BS, d, is selected at 750 meters. Accordingly, the cell-center helpers are deployed between $\frac{d}{3}$ and $\frac{2d}{3}$. It is assumed that the helpers cannot be close to the user and the MEC server from a distance of 50 meters. In the proposed system, the distances between the user and the helpers are determined according to TOD results. On the other hand, in the one helper systems in (Cao et al., 2019) and (Li et al., 2021), the helper is located in the middle of the user and BS, which is 375 meters.

TOD is obtained for various distances using Algorithm 4 by dividing d into 50 meters intervals. In this way, we decide where the helpers should be located to maximize TOD. The total offloading data is shown in Fig. 4.4 as a function of different choices of $d_{u,1}$ and $d_{u,2}$ for T$= 50$ ms. In the figure, $d_{u,1}$ is the distance between the user and the Helper$_1$, whereas $d_{u,2}$ denotes the distance between the user and the Helper$_2$. According to Fig. 4.4, the possible distance pairs giving the highest total offloading data are the case of fixed $d_{u,1} = 250$ m or $d_{u,2} = 500$ m which are plotted in detail in Fig. 4.5.

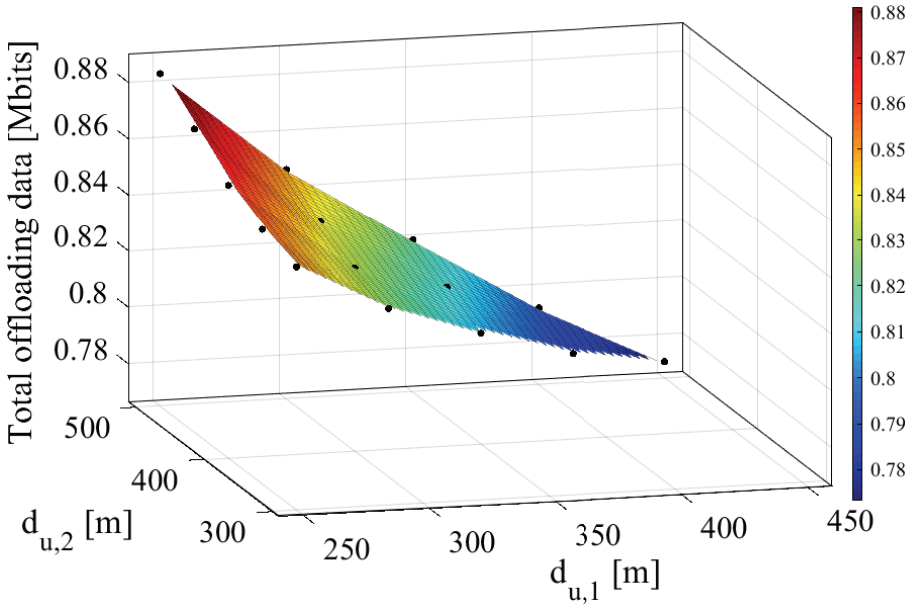


Figure 4.4. Average total offloading data versus $d_{u,1}$ and $d_{u,2}$, T$= 50$ ms.

Fig. 4.5 illustrates the total offloading data versus the different $d_{u,1}$ and $d_{u,2}$ distance pairs for the proposed system for T$= 50$ ms. The increasing curve in Fig. 4.5(a) represents $d_{u,1} = 250$ m is fixed and $d_{u,2}$ is changing through the x-axis. On the other hand, the descending curve in Fig. 4.5(b) represents $d_{u,2} = 500$ m is fixed and $d_{u,1}$ is changing through the x-axis. Fig. 4.5(a) indicates that the closer the Helper$_2$ is to the BS, the higher the total offloading data. Fig. 4.5(b) shows that the closer the Helper$_1$ is to the user, the higher the total offloading data. In other words, when the channel gain between the user and the Helper$_1$ becomes stronger in the downlink transmission and the channel gain between the Helper$_2$ and the BS becomes stronger in the uplink transmission, we obtain the maximum total offloading data.

By considering TOD results given in Fig. 4.5, the optimum distance pair, $[d_{u,1}, d_{u,2}]$, is chosen to have the maximum total offloading data as $[250, 500]$. For the remaining part of the simulation results, the optimum distance pair $[250, 500]$ is utilized.

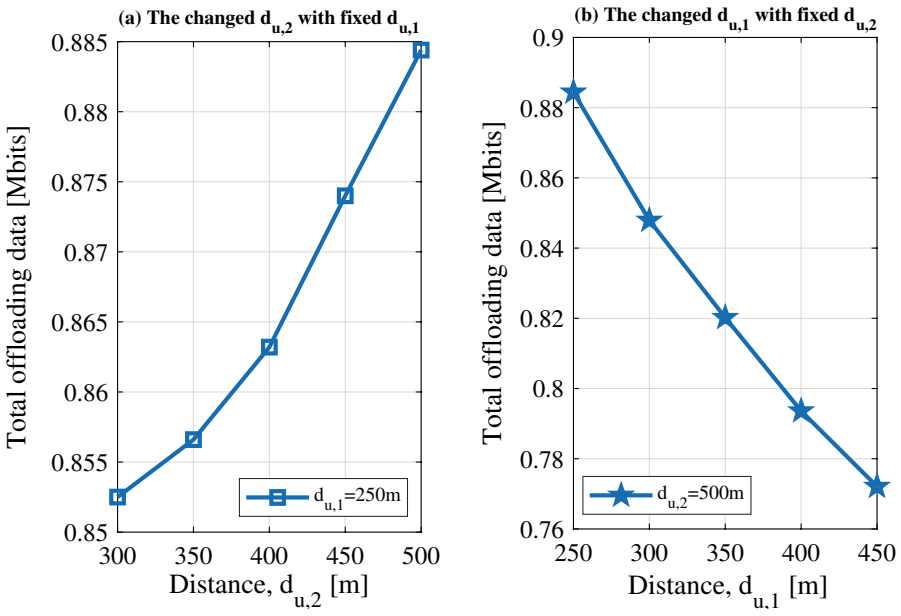


Figure 4.5. The average total offloading data versus fixed $d_{u,1}$ and $d_{u,2}$ distance pairs, T= 50 ms.

In Fig. 4.6, we provide the total offloading data results considering the latency constraints. The latency, T, is a scalable value depending on applications, in this system, it is taken as between 50 ms and 80 ms. Increasing T stretches the latency constraint since the tasks of the user and the helpers can be executed more flexibly. As shown in Fig. 4.6, the system performance is improved in all systems when T increases since more user tasks are offloaded to the helpers. It is observed that two helpers case with NOMA or TDMA schemes outperforms the one helper systems. Furthermore, TOD results demonstrate that the proposed two helpers with NOMA system provides the best performance. This is due to the fact that the user's computation-intensive tasks are distributed between two helpers instead of offloading to the BS directly and two helpers can offload more tasks to the BS using NOMA. The total offloading data in the proposed system increases dramatically comparing with the one helper with direct link in (Li et al., 2021) system as 0.14 Mbits at T= 50 ms and 0.23 Mbits at T= 80 ms. Besides, the proposed two helpers with NOMA system achieves higher total offloading data performance on around 17% compared to the two helpers with TDMA scheme. The proposed two helpers with NOMA system provides a 0.29 Mbits higher TOD at T= 50 ms and 0.47 Mbits higher TOD at T= 80 ms compared to the one helper with TDMA in (Cao et al., 2019).

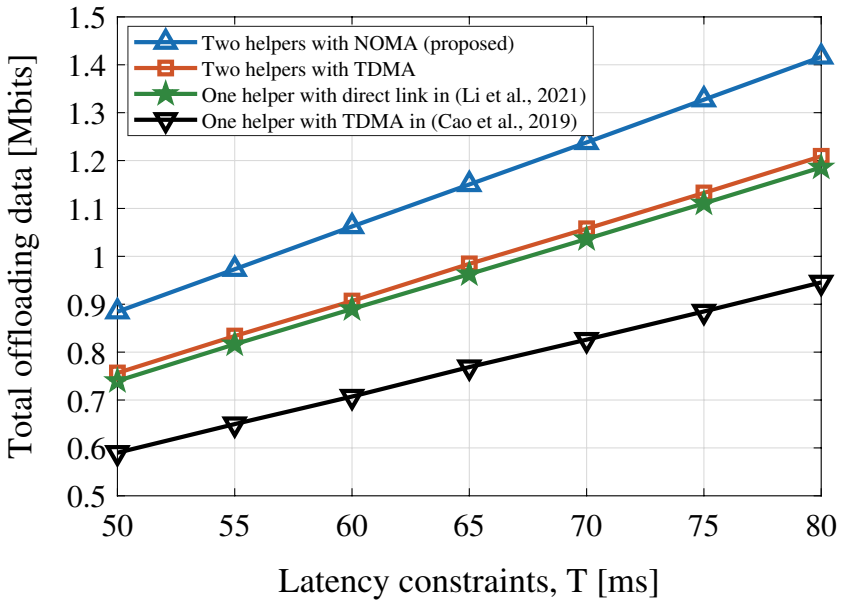


Figure 4.6. The average total offloading data versus latency.

We also investigate the effect of the maximum transmit power of the user, P_u, on the total offloading data and the downlink power allocation factor, ξ, for T$=50$ ms. Fig. 4.7 shows that when P_u increases, the total offloading data increases for all schemes due to increasing in the downlink transmission rates, R_m^D. Specifically, the proposed two helpers with NOMA system achieves 24% and 18% improvement in TOD over the one helper with direct link in (Li et al., 2021) system by using the same transmit power of the user, $P_u = 0.1$ W and $P_u = 0.8$ W, respectively. Besides, the proposed algorithm with NOMA outperforms its TDMA counterpart by 24% and 16% in TOD for $P_u = 0.1$ W and $P_u = 0.8$ W, respectively. It is shown that the proposed two helpers with NOMA system can provide 60% and 47.3% higher TOD than the one helper with TDMA in (Cao et al., 2019) for $P_u = 0.1$ W and $P_u = 0.8$ W, respectively.

Fig. 4.8 illustrates that larger P_u results in a smaller ξ. The reason is that increasing P_u leads to allocate more power to the Helper$_2$. Then, the offloading data, ℓ_1 and ℓ_2 are allocated fairly in the proposed system.

The performance results based on uplink transmit powers, downlink transmit powers, the amount of the offloading data at the user and the helpers side and TOD are given in Table 4.2 for the different distance pairs for T$=50$ ms and $P_u = 0.5$ W. Table 4.2 implies that there is an optimization between time constraints; t_h is lower than t_u so that the user can offload more data to the helpers. Furthermore, the uplink transmit powers, P_k^U, are allocated proportionally to the channel gain between the Helper$_k$ and the BS.

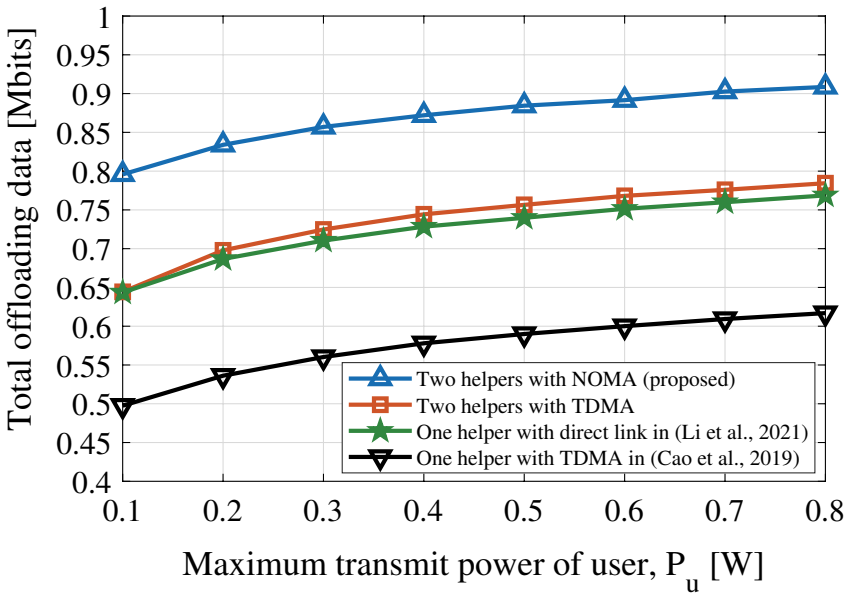


Figure 4.7. The average total offloading data versus maximum transmit power of user, T= 50 ms.

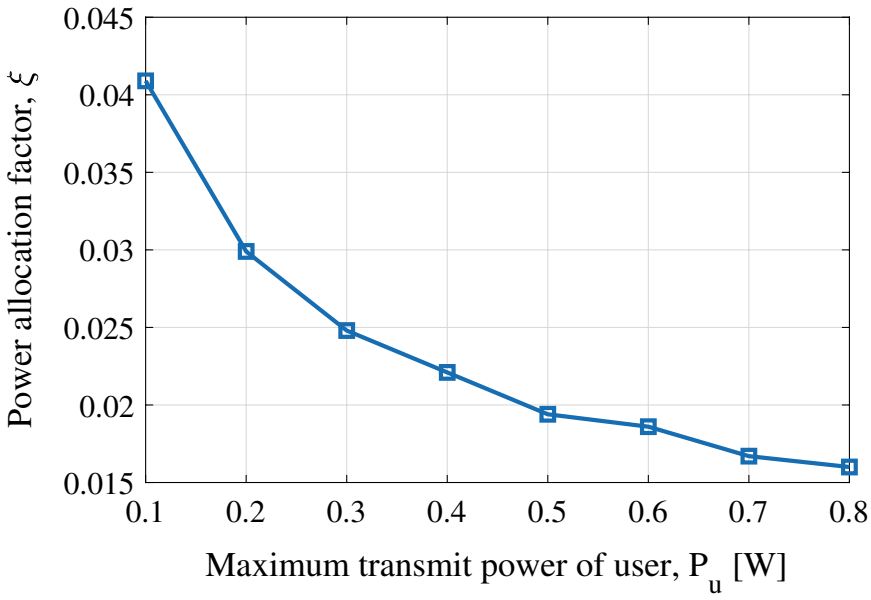


Figure 4.8. The power allocation factor, ξ, versus the maximum transmit power of user, $d_{u,1} = 250$ m, $d_{u,2} = 500$ m, and T= 50 ms.

In Table 4.2, the distance pair of $[250, 300]$ provides more user data offloaded to the helpers, while less data is offloaded to the BS due to longer distance between the helpers and the BS. This inference is easily comprehended from the optimized task offloading decision factor, $\boldsymbol{\alpha}^{d*}$. Thus, it results in less TOD compared to the optimum distance pair of $[250, 500]$. Furthermore, although the optimized task offloading decision factor, $\boldsymbol{\alpha}^{d*}$ takes the maximum value, TOD is lower for the distance pair of $[450, 500]$. The reason is that since the user has to complete the computation-intensive tasks under the latency constraint, when the helpers are far away from the user, less data will be offloaded to the helpers to meet the latency constraint.

Table 4.2. Performance results for different $d_{u,1}$ and $d_{u,2}$ distance pairs, for T$=50$ ms, $\mathrm{P}_u = 0.5$ W.

$\mathrm{d}_{u,1}$ [m]	$\mathrm{d}_{u,2}$ [m]	t_u^* [ms]	α_1^{d*}	α_2^{d*}	$\mathrm{P}_1^{\mathrm{D}}$ [dBm]	$\mathrm{P}_2^{\mathrm{D}}$ [dBm]	$\mathrm{P}_1^{\mathrm{U}}$ [dBm]	$\mathrm{P}_2^{\mathrm{U}}$ [dBm]	ℓ_1 [Mbits]	ℓ_2 [Mbits]	$\ell_{1,\mathrm{o}}$ [Mbits]	$\ell_{2,\mathrm{o}}$ [Mbits]	TOD [Mbits]
250	300	33.5	0.53	0.51	6.3	27	16.5	29	0.33	0.24	0.17	0.12	0.85
250	400	32.4	0.63	0.6	8.3	27	19.4	29	0.33	0.21	0.20	0.13	0.86
250	**500**	32.5	0.67	0.66	9.9	26.9	22.6	29	0.34	0.19	0.23	0.13	**0.88**
300	350	32.4	0.64	0.62	7.8	27	15.7	29	0.29	0.22	0.18	0.13	0.83
350	400	32.4	0.68	0.67	9	27	14.7	29	0.28	0.20	0.19	0.14	0.81
350	500	33.3	0.69	0.69	10.3	26.9	18.3	29	0.29	0.19	0.20	0.13	0.82
450	500	34.2	0.7	0.7	11.1	26.9	14.8	29	0.26	0.19	0.18	0.13	0.77

We also discuss the total computing data since it is important to execute the amount of the user's task. This variable indicates that how much of the user's data is cooperatively executed at both the BS and helpers side under the latency constraint. For the fixed optimum distance pair $[250, 500]$ and T$=50$ ms, the proposed two helpers with cooperation scheme achieves higher total computing data performance on around 18% compared to the without cooperation case where the optimized $\boldsymbol{\alpha}^{d^*}$ values equal to 1 and the helpers act as a relay. This result indicates that the cooperation becomes more important to execute more user's data.

The performance of TOD is compared through multiple helpers for the cases K$=$2 and K$=$3. The optimum distances for K$=$3 are determined as $\mathrm{d}_{u,1} = 250$ m, $\mathrm{d}_{u,3} = 300$ m and $\mathrm{d}_{u,2} = 500$ m to maximize TOD. Thus, we show that K$=$3 provides a gain of about 15 kbits on TOD compared to K$=$2 case for T$=50$ ms. The contribution of adding one helper to the proposed framework is affected by power constraints.

4.2. Massive MIMO based Cooperative MEC with Secure Offloading

In the MEC system, the achievement of edge computing operations depends on transmission data rates. One of the key technologies for wireless systems to achieve high data rates is massive MIMO, which is being increasingly adopted in different frameworks. The massive MIMO based MEC strategy has greatly assisted the offloading in the MEC system due to the significant gains in both spectral and energy efficiencies (Zeng et al., 2020a). The massive MIMO can yield higher transmission rates for offloading in MEC. Besides, since massive MIMO can simultaneously support a larger number of users in offloading, the wireless data transmission delay, especially in the uplink, is reduced.

Apart from massive MIMO and MEC technology, the combination of NOMA and massive MIMO yields great potential in MEC systems, such as higher spectral and energy efficiencies, massive connectivity and lower delay.

While the advantages of MEC technology are in decreasing computation latency and traffic loads on the backhaul networks, secure offloading in MEC systems is of critical importance. Due to the broadcast nature of wireless communications, the computation tasks offloaded from users to the MEC server may be overheard by nearby eavesdroppers, leading to security threats to users (Mao et al., 2022). Without proper security mechanisms, the advantages of MEC technology will be diminished by the damage caused by eavesdroppers (Elgendy et al., 2020). Therefore, a secure task offloading scheme is essential for successfully completing computing tasks (Wang et al., 2020). To address this, the physical layer security (PLS) technique has been a promising solution to ensure the security of task offloading in MEC systems. Specifically, PLS exploits the nature of wireless channels to achieve secure information transmission with affordable complexity (Xu et al., 2021).

4.2.1. Related Works

In this subsection, we present the literature review on massive MIMO-assisted MEC systems and the secure task offloading scheme in MEC systems.

Massive MIMO assisted MEC Systems:

Many works have focused on energy consumption and energy efficiency for massive MIMO based MEC systems. The authors of (Hao et al., 2019) have considered an energy minimization problem for a massive MIMO enabled heterogeneous network (HetNet) with MEC to show that the energy consumption can be reduced by employing

massive MIMO with a maximum-ratio combing detector. In (Malik and Vu, 2019) and (Malik and Vu, 2020), the authors have formulated an energy optimization problem at both the users and the MEC server for a delay constrained massive MIMO based MEC system. Furthermore, the authors of (Mukherjee and Lee, 2020) have explored an edge computing-enabled, cell-free multicell massive MIMO system. They have analyzed the impact of the successful computation probability on the total energy consumption using queuing theory and stochastic geometry. In the paper of (Zhao et al., 2020b), the minimization of the user's total energy consumption has been considered by jointly optimizing the user's offloading data, transmission power and offloading rate for a massive MIMO based MEC system.

The delay minimization problem has been investigated in the following studies for massive MIMO based MEC systems. As given in Fig. 3.2, the authors of (Huang et al., 2019) have investigated a delay minimization problem for a single-cell, massive MIMO assisted MEC system. It is shown that the delay can be reduced by employing massive MIMO with a maximum-ratio combing detector. In (Zeng et al., 2020), the overall delay minimization among all users has been studied for a massive MIMO assisted MEC system with a joint allocation of wireless and computational resources considering the perfect and imperfect CSI cases. The authors of (Feng et al., 2020) have designed a single-cell multi-user massive MIMO MEC system based on the joint resource allocation to minimize the maximum delay consisting of pilot transmission delay, data transmission delay and server computation delay.

Secure MEC Systems:

Since the offloading data may be intercepted and overheard by eavesdroppers due to the broadcast nature of wireless communications, the MEC system brings security challenges. Thus, some studies have focused on designing a secure task offloading scheme to avoid information leakage. In (Lin et al., 2019), the computation efficiency maximization problem has been studied in a multi-user NOMA-enabled MEC system with PLS. The secure computation efficiency problem has been formulated by jointly optimizing the transmission power and the CPU frequency of local computing. The authors in (Yang et al., 2019b) and (Wang et al., 2020) have considered a multi-user uplink offloading scenario with one eavesdropper. The authors have examined a joint optimization of the computing task allocation, local CPU frequency, offloading power and time slots to minimize the total energy consumption. In (Xiao et al., 2020), the authors have examined a deep reinforcement learning-based mobile offloading scheme for edge computing against jamming attacks and interference. A safe reinforcement learning has been used to avoid

choosing the risky offloading policy that fails to meet the computational latency requirements of the tasks. Moreover, in (Zhang et al., 2021), an optimization problem has been introduced to minimize the weighted sum of the execution latency and energy consumption subject to communication and computation resource constraints for a MEC system consisting of one MEC server, multiple mobile devices and one eavesdropper. In (Xu et al., 2021), secure computation offloading has been studied for multi-user multi-server MEC-enabled IoT. The joint optimization of communication and computation resource allocation, a partial offloading ratio have been performed to maximize the total secrecy offloading data considering offloading latency and secrecy constraints. In (Wu et al., 2022), a NOMA-assisted secure computation offloading has been investigated under the eavesdropping attack, in which a wireless user forms a NOMA pair with an edge-computing user to provide cooperative jamming to the eavesdropper while gaining the opportunity of sending its data.

The following studies have investigated the latency minimization problem for secure offloading MEC system. Specifically, in the study of (Wang et al., 2019a), an uplink NOMA based MEC system has been considered with one eavesdropper. The task completion time minimization problem has been studied subject to the worst-case secrecy rate, the transmit power and secrecy outage probability constraints. A secure and low-latency offloading MEC system with one eavesdropper has been presented in (Zhou et al., 2020). The minimization of total latency has been formulated by jointly optimizing the users' transmit power, computing capacity allocation and user association subject to security and computing resource constraints. Furthermore, in (Sun et al., 2020), a power allocation algorithm has been provided to achieve an optimal secure data rate and reduce the whole task latency of both communication and computation. The authors in (Wang et al., 2022) have investigated PLS in a NOMA based MEC system with hybrid SIC decoding. A latency minimization problem has been formulated by jointly designing computational resource allocation, task assignment and power allocation.

Motivated by the above background, we present an overall delay minimization problem in the massive MIMO-NOMA based cooperative MEC system for the scenario as shown in Fig. 4.9 where there is a cell-edge user with a computation-intensive and latency-critical task (i.e., AR, VR, real-time online gaming, or remote healthcare applications) (Yılmaz et al., 2023). The cell-edge user may experience a low SINR or there may not be a strong direct transmission link to the BS. Hence, it is needed to execute this cell-edge user's computation-intensive task with low latency. In this case, all cell-center users, termed helpers, assist in executing this task using a cooperative communication strategy

Figure 4.9. The scenario of massive MIMO based cooperative MEC system model.

based on massive MIMO and NOMA technologies. In this framework, the overall delay is determined by taking into account both downlink and uplink transmissions. NOMA is performed in the downlink transmission between the cell-edge user and helpers. On the other hand, a massive MIMO channel is employed for the uplink transmission, including data transmission between helpers and the MEC server. In addition to that, we consider the computation delay at the MEC server by executing the offloaded tasks. Furthermore, these computation-intensive and latency-critical tasks may include private, financial and identity information such as medical records or payments. Thus, secure offloading in the MEC system is investigated in uplink transmission, where the eavesdropper might intend to overhear the offloaded tasks from the helpers. Different from the existing works, we present a cooperative MEC system including secure offloading. In the proposed framework, we aim to minimize both offloading and computing delay by providing cooperation through massive MIMO and NOMA while satisfying security constraints.

In this section, massive MIMO and NOMA are introduced into a cooperative MEC system to minimize the overall delay. In particular, we investigate a secure massive MIMO based cooperative MEC by considering the overall delay, including both offloading and computing. In the overall system, NOMA and massive MIMO communication are performed in the downlink and uplink transmission, respectively. We propose efficient algorithms for both the downlink and uplink transmission in the proposed framework. The overall delay minimization problem is formulated in the secure MEC system considering both the secrecy rate and the uplink delay. The performance results show that the proposed massive MIMO-NOMA based cooperative MEC system significantly reduces the overall delay through multiple helpers. Furthermore, massive MIMO and NOMA technologies facilitate secure offloading in a cooperative MEC framework.

4.2.2. System Model

We consider a massive MIMO based cooperative MEC system model consisting of one BS with N antennas, K single-antenna helpers, and one user with a single-antenna, under the assumption of $N >> K >> 1$, as illustrated in Fig. 4.10. NOMA is applied in the downlink transmission between the user and the helpers. On the other hand, in the uplink transmission between the helpers and the BS, a massive MIMO is employed.

It is assumed that the perfect CSI between all nodes is available at the BS. The BS is connected to a MEC server by an optical fiber link to provide computing service for

the users in its coverage. In massive MIMO, a ULA antenna model where neighboring antennas are spaced by $D = \lambda/2$.

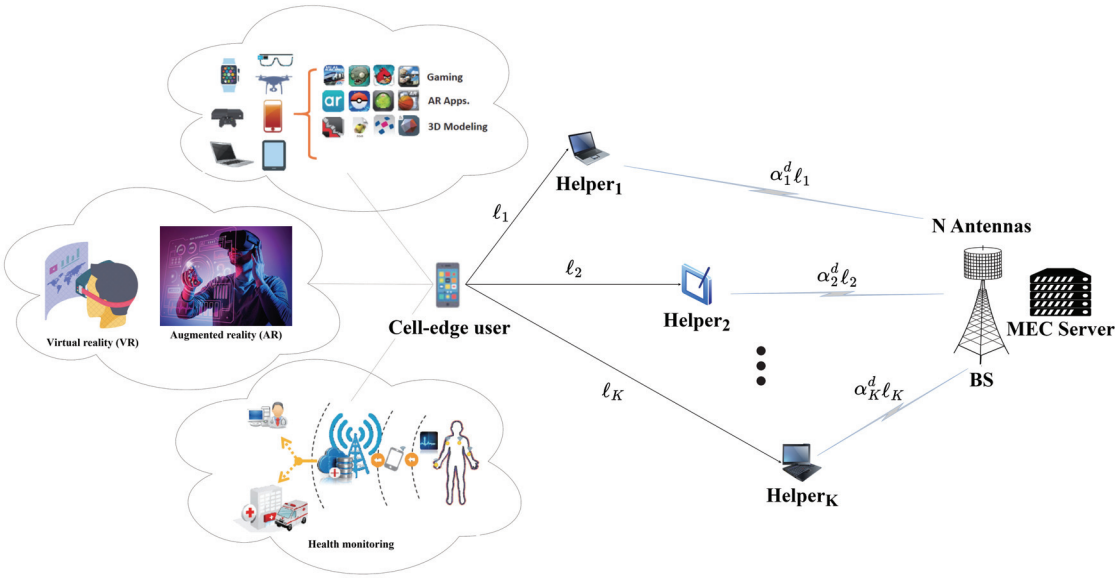


Figure 4.10. The proposed massive MIMO based cooperative MEC system model with K helpers.

The user has computation tasks with a data size of L. It is assumed that there is no strong direct transmission link between the user and the MEC server since the user is at the cell edge. Thus, for the offloading phase, the user sends a certain part of its tasks to the MEC server through the helpers. It is assumed that these helpers are at the cell-center as shown in Fig. 4.10. The BS is integrated with the MEC server to execute the computation-intensive tasks that are offloaded by the helpers. Upon receiving the data, the MEC server applies computing to these tasks.

We consider partial offloading on the helper side, where helpers' resources are partitioned into two parts: one part is processed locally and the remaining is offloaded to the MEC server. In the proposed framework, these computational tasks, which depend on various parameters, are partitioned based on offloading decision factor, α_k^d. After the user simultaneously offloads a certain part of its own data, L, to the helpers, the Helper$_k$ uses α_k^d factor to decide the portion of data offloading and data computing.

In downlink transmission, NOMA based scheme is used to offload the cell-edge user's tasks to the K helpers simultaneously. Then, the average data rate R_k^D is given as in (4.1).

In the uplink transmission, in the massive MIMO system, K helpers simultaneously transmit their symbols to the BS. Then, the received data vector $\mathbf{y} \in \mathbb{C}^{N\times 1}$ at the BS is given by;

$$\mathbf{y} = \sum_{k=1}^{K} \sqrt{P_k^U}\, \mathbf{h_k} s_k + \mathbf{z} \tag{4.29}$$

where $\mathbf{s} = [s_1, \ldots, s_k, \ldots s_K]^T \in \mathbb{C}^{Kx1}$ is the transmitted symbol vector, $\forall k \in \mathcal{K}$. P_k^U is the uplink transmit power of the Helper$_k$, $\mathbf{h_k} \in \mathbb{C}^{N\times 1}$ is the channel vector between the k^{th} helper and the BS, and the 2D massive MIMO channel model is implemented as in (2.2). $\mathbf{z}$ is the AWGN vector with zero mean and σ_u^2 variance, $\mathbb{C}\mathcal{N}(0, \sigma_u^2)$.

The average data rate R_k for Helper$_k$ is defined as;

$$R_k = B\ \mathbb{E}\left\{\log_2\left(1+\gamma_k\right)\right\} \tag{4.30}$$

where γ_k is the received SINR for the k^{th} helper after applying linear detector.

The uplink sum data rate of K helpers is computed as;

$$R = \sum_{k=1}^{K} R_k \tag{4.31}$$

The minimum mean-squared error (MMSE) detection scheme is used at the BS to detect the uplink data. To compute γ_k for each helper, a linear MMSE detector scheme is employed by;

$$\mathbf{V}_{\text{MMSE}} = \left(\mathbf{H}^H\mathbf{H} + \sigma_u^2\mathbf{I}_K\right)^{-1}\mathbf{H}^H \tag{4.32}$$

where $\mathbf{V}_{\text{MMSE}} \triangleq [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_K] \in \mathbb{C}^{K\times N}$ is the MMSE matrix and the channel matrix is given by $\mathbf{H} \triangleq [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_K] \in \mathbb{C}^{N\times K}$.

We detect the data symbol for the k^{th} helper by;

$$s_k = \mathbf{v}_k\mathbf{y} \tag{4.33}$$

where $\mathbf{v}_k$ is the k^{th} row vector of $\mathbf{V}_{\text{MMSE}}$.

We obtain the received symbol belonging to the k^{th} helper as;

$$s_k = \sqrt{P_k^U}\mathbf{v}_k\mathbf{h}_k s_k + \sum_{i=1,i\neq k}^{K} \sqrt{P_i^U}\mathbf{v}_k\mathbf{h}_i s_i + \mathbf{v}_k\mathbf{z} \tag{4.34}$$

where the first term represents the received data symbol of the k^{th} helper, while the second and the third terms represent the interference from other helpers and noise, respectively.

Thus, the SINR of the k^{th} helper at the MMSE detector output can be calculated as;

$$\gamma_k = \frac{P_k^U \left|\mathbf{v}_k\mathbf{h}_k\right|^2}{\sum\limits_{i=1,i\neq k}^{K} P_i^U \left|\mathbf{v}_k\mathbf{h}_i\right|^2 + \left\|\mathbf{v}_k\right\|^2 \sigma_u^2} \tag{4.35}$$

4.2.3. Proposed Framework and Problem Formulation

In this section, we first introduce the offloading and computing scheme within the scope of the proposed framework and then formulate the corresponding optimization problem. Then, we investigate the secure offloading in the proposed MEC system. In the proposed framework, the cell-center helpers assist the cell-edge user in offloading the user's tasks to the MEC server since there is no strong direct transmission link between the cell-edge user and the MEC server. We cannot consider the helpers as pure relays since they can compute some parts of the offloaded tasks from the cell-edge user and at the same time, they can offload the remaining parts of these computation tasks to the MEC server. Moreover, we assume that there is no energy restriction on the helpers.

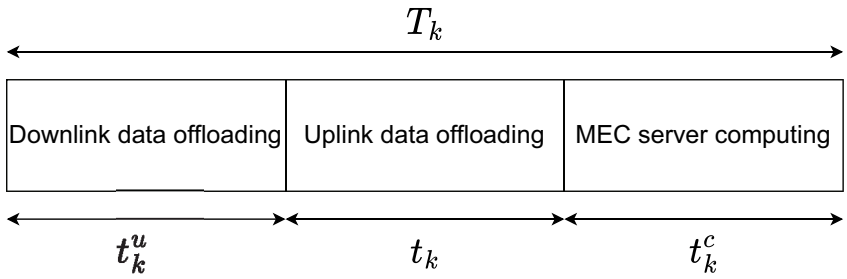


Figure 4.11. Overall delay scheme for offloading and computing.

Fig. 4.11 shows the overall delay scheme for offloading and computing in the system. The overall delay, T_k, consists of the downlink data offloading, t_k^u, uplink data offloading, t_k, and computing at the MEC server, t_k^c, for each helper.

Accordingly, the overall delay for each helper can be expressed as follows;

$$T_k = t_k^u + t_k + t_k^c \tag{4.36}$$

The downlink transmission delay for the user to offload a certain part of its data, L, to the K helpers can be expressed as;

$$t_k^u = \frac{\ell_k}{R_k^D} \tag{4.37}$$

where $\ell_k = \frac{L}{K}$, $\forall k \in \mathcal{K}$.

The required uplink transmission delay for offloading data from each helper to the MEC server is given by;

$$t_k = \frac{\alpha_k^d \, \ell_k}{R_k} \tag{4.38}$$

For a given offload task $\alpha_k^d \, \ell_k$ by the k^{th} helper, the delay of computing this task at the MEC server is given by;

$$t_k^c = \frac{\alpha_k^d \, \ell_k \, \theta}{f_{k,s}} \tag{4.39}$$

where the computing resource assigned to k^{th} helper is $f_{k,s}$ by the MEC server.

Since the helpers have to wait for all downlink data to perform computation and offloading, the uplink transmission is performed after the downlink transmission is completed. The solution to minimization of the downlink transmission delay, t_u^*, is obtained through Algorithm 5, which is given in the following section.

The aim of optimization in a massive MIMO based cooperative MEC system is to minimize the overall delay for both offloading and computing. Towards this end, we jointly optimize the offloading decision factors, transmit powers and the MEC server's computing resources to minimize the overall delay.

The overall delay of offloading and computing for each helper can be expressed by $w_k(.)$ as a function of $\mathbf{P}^D, \mathbf{P}^U, \mathbf{f}, \boldsymbol{\alpha}^d$ as follows;

$$w_k(\mathbf{P}^D, \mathbf{P}^U, \mathbf{f}, \boldsymbol{\alpha}^d) = T_k \tag{4.40}$$

where $\mathbf{P}^D = \left[P_1^D, \ldots, P_K^D\right], \mathbf{P}^U = \left[P_1^U, \ldots, P_K^U\right], \mathbf{f} = [f_{1,s}, \ldots, f_{K,s}]$ and $\boldsymbol{\alpha}^d = \left[\alpha_1^d, \ldots, \alpha_K^d\right]$.

Hence, our optimization problem can be formulated as the minimization of the overall delay subject to the given constraints as;

$$\min_{\mathbf{P}^D, \mathbf{P}^U, \mathbf{f}, \boldsymbol{\alpha}^d} \quad \max_k \quad w_k(\mathbf{P}^D, \mathbf{P}^U, \mathbf{f}, \boldsymbol{\alpha}^d) \tag{4.41}$$

$$\text{s.t.} \quad R_k^D \geq R_{th,k}^D, \quad \forall k \in \mathcal{K} \tag{4.41a}$$

$$\sum_{k=1}^{K} P_k^D \leq P_u, \tag{4.41b}$$

$$\sum_{k=1}^{K} f_{k,s} \leq F_{max}, \tag{4.41c}$$

$$\frac{\left(1 - \alpha_k^d\right) \ell_k \, \theta}{f_k} \leq T_k - t_k^u, \quad \forall k \in \mathcal{K}, \tag{4.41d}$$

$$0 < P_k^U \leq P_{max}^U, \quad \forall k \in \mathcal{K}, \tag{4.41e}$$

$$0 < \alpha_k^d < 1, \quad \forall k \in \mathcal{K}. \tag{4.41f}$$

where P_u is the maximum total transmit power of the cell-edge user. At the MEC server, its computational capacity, F_{max}, is shared among all helpers.

Constraint (4.41a) gives the minimum data rate constraint that the achievable rate in downlink transmission should be equal to or higher than a given threshold rate $\mathrm{R}_{th,k}^D$. Constraint (4.41b) represents the power constraints in the NOMA downlink transmission. (4.41c) is the computing resource constraint of the MEC server. (4.41d) denotes the required time for the local computation of the remaining tasks at Helper_k. (4.41e) represents the maximum transmit power of Helper_k. (4.41f) shows the range of α_k^d, which gives the ratio of tasks offloaded to the MEC server and computed locally at each helper.

4.2.3.1. Problem Solution

In this section, we derive the solution to Problem (4.41) for the proposed framework with K helpers massive MIMO based cooperative MEC.

The constraint (4.41c) can be rearranged as the sum of the computing resources allocated to helpers is equal to the maximum CPU operating frequency of the MEC server as;

$$\sum_{k=1}^{K} f_{k,s} = F_{max}, \tag{4.42}$$

Accordingly, the computing resources allocated to each helper are shared equally, such as $f_{k,s} = \frac{F_{max}}{K}, \ \forall k \in \mathcal{K}$.

Furthermore, α_k^d shows the ratio of tasks offloaded to the MEC server and the tasks executed locally at each helper. For the case of α_k^d is 0 or 1, the tasks can be executed either at the helpers or at the MEC server without any cooperation. In order to guarantee cooperation in the proposed MEC framework, the value of α_k^d in (4.41f) is determined between 0.3 and 0.7. In this way, the helpers execute some tasks regarding their computing resources while offloading the remaining tasks to the MEC server. Then, we re-write the constraint (4.41f) as;

$$0.3 \leq \alpha_k^d \leq 0.7, \quad \forall k \in \mathcal{K}. \tag{4.43}$$

Accordingly, in order to solve the optimization problem (4.41), the auxiliary variable τ is introduced. In this way, it is considered that the overall delay T_k for each helper is equal to each other and denoted by τ. Thus, the optimization Problem (4.41) becomes the minimization of τ and can be rearranged as follows;

$$\min_{\mathbf{P}^D, \mathbf{P}^U, \boldsymbol{\alpha}^d} \ \tau \tag{4.44}$$

$$\begin{aligned} \text{s.t.} \quad & (4.41a), (4.41b), (4.41e), (4.42), (4.43), \\ & \frac{\left(1-\alpha_k^d\right) \ell_k \, \theta}{f_k} \leq \tau - t_k^u, \quad \forall k \in \mathcal{K}, && (4.44a) \\ & \{t_k^u + t_k + t_k^c\} \leq \tau, \quad \forall k \in \mathcal{K}. && (4.44b) \end{aligned}$$

Constraint (4.44b) shows that the overall delay for each helper should be equal to or less than a constant, τ.

Firstly, we give the solution for the case of K=2 and then generalize it to the K helpers.

In this way, firstly, we focus on the solution for the downlink transmission part. The constraint (4.41b) associated with the downlink transmit power for helpers can be written as $P_1^D + P_2^D = P_u$. Thus, $\xi \in (0, 1)$ is determined as the power allocation factor and becomes one of the optimization parameters. Then, the allocated downlink transmit power, P_1^D, for the Helper_1 is determined by ξP_u while the allocated downlink transmit power, P_2^D, for the Helper_2 is calculated as $(1 - \xi) P_u$.

The achievable rate of the Helper_k in an OMA system is given by R_k^{OMA} as in (4.6). We define $\mathrm{R}_{th,k}^D$ as $\mathrm{R}_k^{\mathrm{OMA}}$. Then, the range of ξ can be obtained directly from the constraint related to the new optimization problem (4.44), (4.41a), as given previously in (4.9).

Then, we can extend the downlink power allocation factors to the K helpers as given in (4.10), (4.11) and (4.12).

In order to minimize the overall delay in (4.44), firstly, the maximum downlink transmission delay, t_k^u, is minimized for a given range of the power allocation factors, $\boldsymbol{\xi}$;

$$\min_{\boldsymbol{\xi}} \max_{k} \quad t_k^u \tag{4.45}$$

$$\text{s.t.} \qquad (4.12)$$

where $\boldsymbol{\xi} = \{\xi_1, \ldots, \xi_{K-1}\}$.

To solve the Problem (4.45), we convert min-max problem to a minimum problem by defining the auxiliary variable, t_u. Accordingly, this implies that the downlink transmission delay t_k^u for each helper is equal to each other and denoted by t_u.

$$\min_{\boldsymbol{\xi}} \; t_u \tag{4.46}$$

$$\text{s.t.} \qquad (4.12),$$

$$t_k^u \leq t_u, \quad \forall k \in \mathcal{K}. \tag{4.46a}$$

Constraint (4.46a) indicates that the downlink transmission delay for each helper should be equal to or less than a constant, t_u.

Thus, the minimum of a constrained nonlinear multivariate function of (4.46) can be obtained using the interior-point method. Here, $\mathbf{x}$ is defined as a vector of the

components; $\mathbf{x} = [t_u, \boldsymbol{\xi}]$. The vector $\mathbf{x}$ satisfying all the constraints is called a feasible solution for the Problem (4.46). The initial values, $\mathbf{x}^1$, are determined through lower and upper bounds for each component in $\mathbf{x}$.

Then, the auxiliary function with a barrier parameter μ is expressed as;

$$S_\mu(\mathbf{x}) = t_u + \mu\, P(\mathbf{x}) \tag{4.47}$$

where $P(\mathbf{x})$ is given in (3.11) with $\mathbf{G} = [G_1, G_2, \ldots, G_K]$ representing the nonlinear inequality constraints in (4.46a) such as $t_k^u - t_u \leq 0; \forall k \in \mathcal{K}$.

Once the $\boldsymbol{\xi}$ values have been obtained, we find the downlink transmit power for helper k, $P_k^{D^*}$. The details of the solution for the downlink transmission delay, t_u^* for K helpers are summarized in Algorithm 5.

Algorithm 5 Solution of the Downlink Transmission Delay based on Interior-Point Algorithm for K Helpers

---

Input: $\mathbf{x} = [t_u, \boldsymbol{\xi}]$, g_k, ℓ_k; $\forall k \in \mathcal{K}$

1: Rearrange the inequality constraints in (4.46a) as $\mathbf{G} = [G_1, G_2, \ldots, G_K]$, $G_i(.) \leq 0$, $i = 1, 2, \ldots, K$.

2: Reformulate the objective function as an auxiliary function by,

$$S_{\mu_j}(\mathbf{x}) = t_u + \mu_j P(\mathbf{x})$$

where $P(.)$ is expressed as in (3.11).

3: Call ***Algorithm 3***

Output: t_u^*, $P_k^{D^*}$

After the downlink transmission delay, t_u^*, is determined through Algorithm 5, we focus on the uplink transmission to solve the corresponding optimization Problem (4.44) efficiently by using standard nonlinear programming optimization tools (Grace, 1990). Accordingly, the constraint (4.44a) is rewritten as;

$$\frac{\left(1 - \alpha_k^d\right) \ell_k\, \theta}{f_k} \leq \tau - t_u^*, \quad \forall k \in \mathcal{K}, \tag{4.48}$$

Similarly, we can also write the overall delay constraint (4.44b) for each helper;

$$\{t_u^* + t_k + t_k^c\} \leq \tau, \quad \forall k \in \mathcal{K} \tag{4.49}$$

Then, we reformulate the Problem (4.44) for the uplink transmission part as the minimization of τ under $\mathbf{P}^U, \boldsymbol{\alpha}^d$ with the constraints (4.41e), (4.42), (4.43), (4.48) and (4.49). Thus, the minimum of a constrained nonlinear multivariate function can be obtained using the interior-point method.

The interior-point method for the solution of the overall delay is given in Algorithm 6. In this minimization problem, $\mathbf{x}$ is defined as a vector of the components; $\mathbf{x} = \left[\tau, \alpha_k^d, P_k^U\right], \forall k \in \mathcal{K}$. The vector $\mathbf{x}$ satisfying all the constraints is called a feasible solution for the Problem (4.44). The initial values, $\mathbf{x}^1$, are defined through lower and upper bounds for each component τ, α_k^d and P_k^U in $\mathbf{x}$. As a result of Algorithm 6, the output values are obtained as $\alpha_k^{d*}, P_k^{U*}, \forall k \in \mathcal{K}$ to obtain minimum overall delay, τ^*, under the given constraints.

Algorithm 6 Solution of the Overall Delay based on Interior-Point Algorithm for K Helpers

---

Input: $\mathbf{x} = \left[\tau, \alpha_k^d, P_k^U\right], g_k, \mathbf{h_k}$; $\forall k \in \mathcal{K}$ and t_u^*

1: Rearrange the inequality constraints in (4.48) and (4.49) as $\mathbf{G} = [G_1, G_2, \ldots, G_{2K}], G_i(.) \leq 0,\ i = 1, 2, \ldots, 2K$.

2: Reformulate the objective function as an auxiliary function by,

$$S_{\mu_j}(\mathbf{x}) = \tau + \mu_j P(\mathbf{x})$$

where $P(.)$ is given in (4.24).

3: Call ***Algorithm 3***

Output: $\tau^*, P_k^{U*}, \alpha_k^{d*}$; $\forall k \in \mathcal{K}$

Thus, the uplink delay, t_h, is calculated as the summation of uplink transmission delay, t^*, and computing delay at the MEC server, t_c^* as;

$$t_h = t^* + t_c^* \tag{4.50}$$

where the uplink transmission delay and computing delay at the MEC server are given, respectively;

$$t^* = \max_{\forall k \in \mathcal{K}} t_k, \tag{4.51}$$

$$t_c^* = \max_{\forall k \in \mathcal{K}} t_k^c. \tag{4.52}$$

4.2.4. Secure Offloading in MEC System

As shown in Fig. 4.12, we consider a secure MEC offloading scenario where an eavesdropper with a single antenna near the BS can overhear the messages transmitting from helpers to the BS (Sun et al., 2020). Specifically, this eavesdropper is passive and never transmits signal and attempts to intercept communications between helpers and the MEC server. Thus, the eavesdropper passively listens to uplink communications. The aim of the helpers is to offload their computation tasks to the MEC server partially while satisfying secrecy constraints. Therefore, we consider PLS technology to ensure that the computing tasks are securely offloaded to the MEC server.

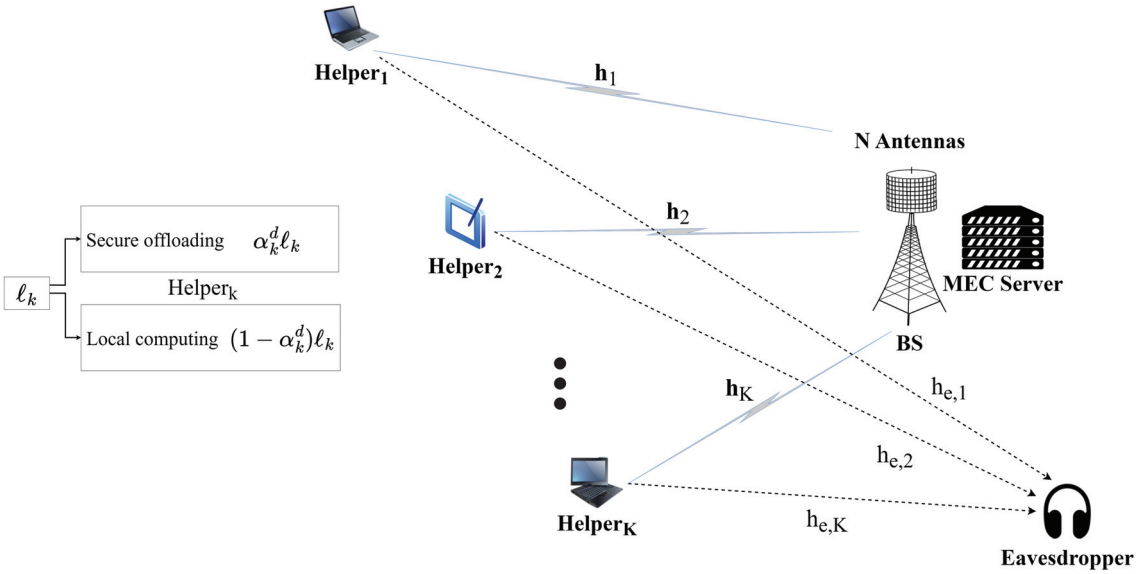


Figure 4.12. The proposed massive MIMO based cooperative MEC system model with secure offloading.

With the partial offloading, $\alpha_k^d\ \ell_k$ bits of computing tasks are securely offloaded to the MEC server and the Helper$_k$ can compute the remaining $(1-\alpha_k^d)\ \ell_k$ bits locally. During the uplink transmission, where the helpers send their data to the BS, the received signal at the eavesdropper is given by;

$$y_e = \sum_{k=1}^{K} \sqrt{P_k^U}\ h_{e,k}\, s_k + n \tag{4.53}$$

where $h_{e,k}$ is the channel between the k^{th} helper and the eavesdropper, which is modeled by the Rayleigh distribution with a variance of distance-dependent path loss coefficient and n is AWGN with zero mean and $\sigma_e^2 = N_0\, B$ variance.

The achievable secrecy rate of each helper, $R_{k,s}$, is given by;

$$R_{k,s} = \max\left\{R_k - R_{k,e}, 0\right\}, \quad \forall k \in \mathcal{K} \tag{4.54}$$

where R_k is defined in (4.30) and $R_{k,e}$ denotes the data rate belonging to k^{th} helper at the eavesdropper;

$$R_{k,e} = B\ \ \mathbb{E}\left\{\log_2\left(1 + \frac{P_k^U\ |h_{e,k}|^2}{\sigma_e^2}\right)\right\} \tag{4.55}$$

Then, the sum secrecy rate is given by;

$$R_s = \sum_{k=1}^{K} R_{k,s} \tag{4.56}$$

In order to avoid information leakage to the eavesdropper, the PLS technique is adopted in the offloading process. The achievable secrecy rate of any helper should be non-negative; otherwise, this helper would stop offloading tasks to the MEC server. When the channel gain of the helper is higher than the channel gain of the eavesdropper, secure transmission is guaranteed. Otherwise, the helper does not offload its computation task to the MEC server since the channel gain of the helper is lower than the channel gain of the eavesdropper. Specifically, if $R_k \leq R_{k,e}$, we cannot ensure secure transmission, and (4.54) results in the value of 0. Thus, we assume that the channel gain of the helper is higher than the channel gain of the eavesdropper.

While designing a secure offloading mechanism in the proposed MEC framework, we define a constraint based on secrecy rate to diminish the offloading information leakage to the eavesdropper. According to (Yang et al., 2019b) (Sun et al., 2020) (Wang et al., 2020) and (Zhang et al., 2021), the transmission delay by considering the security constraint is expressed as follows;

$$t_k = \frac{\alpha_k^d \, \ell_k}{R_{k,s}}, \quad \forall k \in \mathcal{K}. \tag{4.57}$$

Thus, we change the uplink transmission delay for offloading, t_k, in (4.38) into (4.57) as uplink secure transmission delay. The uplink delay, t_h, is calculated as in (4.50) where t_k is given in (4.57).

As a result, we reformulate the MEC optimization problem in (4.44) by minimizing the overall delay based on secure transmission as follows;

$$\min_{\mathbf{P}^U, \boldsymbol{\alpha}^d} \ \tau \tag{4.58}$$

$$\text{s.t.} \quad \text{(4.41e), (4.42), (4.43), (4.48), (4.49)}$$

$$R_{k,e} < R_k, \quad \forall k \in \mathcal{K}. \tag{4.58a}$$

For the Problem in (4.58) with its given constraints, we perform Algorithm 6 to determine the uplink transmission delay under secrecy constraints.

4.2.5. Performance Evaluation

In this section, the performance of the proposed framework is evaluated through the simulation parameters listed in Table 4.3. The distance between the cell-edge user and the BS, d, is set at 750 meters. The cell-center helpers are located at $d_{u,1} = \frac{d}{3}$ and $d_{u,2} = \frac{2d}{3}$ for the K=2 case. We assume that helpers are at least 50 meters away from the cell-edge user and the MEC server. For the case of K=3 and K=4 helpers, their distances are distributed between $d_{u,1} = \frac{d}{3}$ and $d_{u,2} = \frac{2d}{3}$. The channel is modeled using Rayleigh fading components with distance-dependent path loss, whose parameters depend on whether the receiver is the BS as given in (4.25) or the helper as given in (4.26).

As performance metrics, the overall delay performance for different number of helpers, *K*, different number of antennas, *N*, and different amount of user's offloaded data, *L* are provided. In addition, the uplink sum data rate is obtained for different parameters.

Table 4.3. Simulation parameters.

Parameter	Value
N	32,64,128,256
K	2,3,4
Carrier frequency, f_c	2 GHz
B	1 MHz (Zeng et al., 2020a)
Noise power spectral density, N_0	-174 dBm/Hz
P_u	0.5 W
P_{max}^U	0.8 W
θ	1000 cycles/bit (Cao et al., 2018)
τ_{max}	1000 ms
F_{max}	20 GHz (Zeng et al., 2020)
f_k	2 GHz
h_{MS}	1.5 m
Path number	35
ε	10^{-6}
ζ	2

Performance Results of Cooperative MEC:

The uplink delay performance for the different numbers of antennas and the number of helpers at the fixed data amount of L=1.4 Mbits is shown in Fig. 4.13. As the number of BS antennas, N, increases, the offloading delay decreases. The reason is that since the data rate of helpers increases with the number of antennas, the uplink transmission delay for offloading reduces. It results in a lower uplink delay since the decrease in uplink transmission delay is more dominant than the computing delay at the MEC server with the increasing number of antennas at the BS. Moreover, the higher number of helpers, K=4, reduces the uplink delay by an average of 32% and 14% compared to K=2 and K=3, respectively.

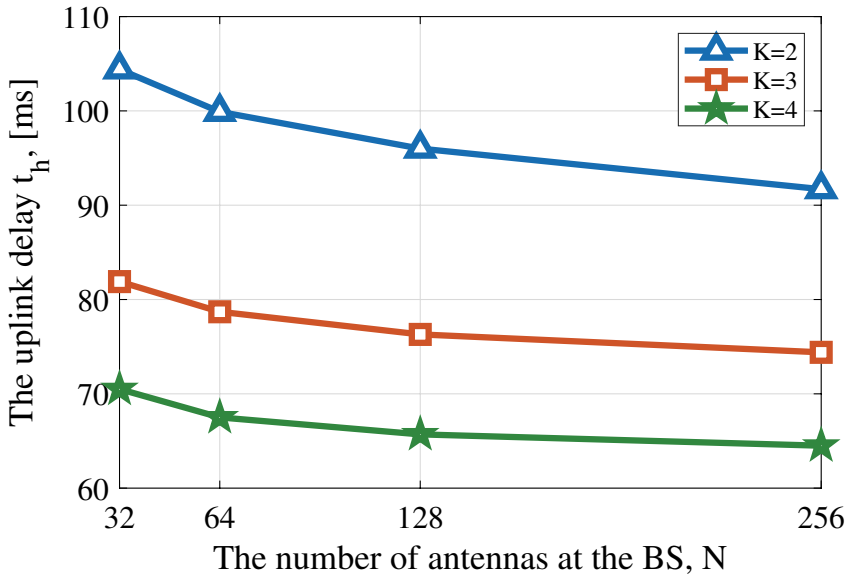


Figure 4.13. The uplink delay versus the number of antennas and helpers for L=1.4 Mbits.

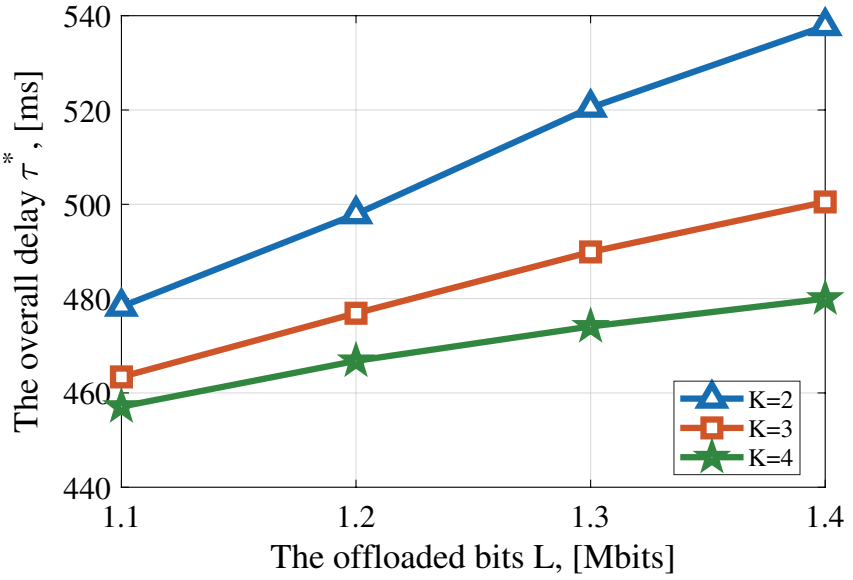


Figure 4.14. The overall delay versus the offloaded data for different number of helpers for N=32.

Fig. 4.14 illustrates the overall delay performance versus the user's offloaded data, L, which is equally distributed to each helper for the different number of helpers and N=32. The overall delay increases with an increase in the amount of the user's offloaded data, L. The reason is that the higher amount of offloaded data leads to an increased transmission delay. In addition, the computing delay increases due to the higher amount of data that needs to be processed by the same server computing resources. The cooperative MEC with K=4 reduces the overall delay by 21.2 ms and 6.3 ms compared to the cases of K=2 and K=3, respectively for L=1.1 Mbits. On the other hand, for L=1.4 Mbits, this difference increases and the cooperative MEC with K=4 reduces the overall delay by 57.8 ms and 20.5 ms compared to the cases of K=2 and K=3, respectively.

Fig. 4.15 provides the uplink sum data rate performance of the different number of helpers and antennas for L=1.4 Mbits. When we increase the number of antennas, the sum data rate of all schemes increases. Besides, the cooperative MEC with K=4 has the highest sum data rate for all cases. Specifically, the cooperative MEC with K=4 achieves 86% and 97% higher sum data rates compared to the case of K=2 at N=32 and N=256, respectively.

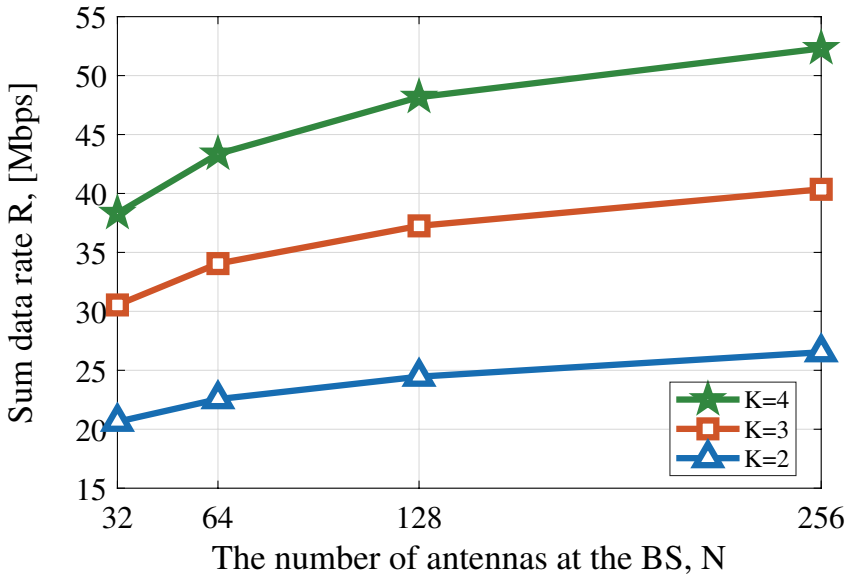


Figure 4.15. The average sum data rate versus the number of antennas and helpers for L=1.4 Mbits.

To further investigate the downlink and uplink delays in detail, Table 4.4 gives the overall delay, τ^*, downlink transmission delay, t_u^*, and uplink delay, t_h, as a result of the proposed algorithm for different K and N values at a fixed L=1.4 Mbits. It is shown that downlink transmission results in a higher delay than uplink transmission due to the downlink power constraints and the higher amount of offloaded data. In addition, the computing delay at the MEC server, t_c^*, decreases with the increased number of helpers for fixed N. The reason is that the offloading decision factor, α_k^{d*}, and offloaded data ℓ_k reduce with K. When we increase the number of antennas at the BS, N, the computing delay at the MEC server, t_c^*, increases slightly because of the increment in α_k^{d*}. It is observed that the cooperative MEC with K=4 has a lower overall delay and when the number of antennas is increased from N=32 to N=256, the uplink delay decreases due to the increasing sum data rate.

Table 4.4. The delay performance for different number of antennas and helpers for L=1.4 Mbits.

K	N	t_u^* [ms]	t^* [ms]	t_c^* [ms]	t_h [ms]	τ^* [ms]
2	32	427.4	55.9	48.5	104.4	537.8
3	32	419	34.9	47	81.9	500.5
4	32	408.9	26.2	44.3	70.5	480
2	256	427.4	42.8	48.9	91.7	533.1
3	256	419	26.3	48.1	74.4	494.4
4	256	408.9	18.7	45.8	64.5	473

Table 4.5. The comparison of with respect to α_k^{d*} and $\mathbf{P}$ values for the different number of helpers at L=1.4 Mbits and N=32.

K	α_1^{d*}	α_2^{d*}	α_3^{d*}	α_4^{d*}	P_1^{D*}	P_2^{D*}	P_3^{D*}	P_4^{D*}	P_1^{U*}	P_2^{U*}	P_3^{U*}	P_4^{U*}
2	0.69	0.69	-	-	18.5	26.3	-	-	27.6	26.2	-	-
3	0.66	0.67	0.67	-	14.4	20.6	25.5	-	28.6	27.1	26.3	-
4	0.61	0.62	0.62	0.63	11.4	18.8	20.7	24.7	27.9	27.4	26.9	26.3

Table 4.5 shows the outputs of the proposed algorithms including α_k^{d*}, $P_k^{D*}[dBm]$, $P_k^{U*}[dBm]$, $\forall k \in \mathcal{K}$ for N=32 and L=1.4 Mbits. It is observed that the value of α_k^{d*} is slightly decreased when the number of helpers are increased. This indicates that K=4 tends to compute more data locally at each helper than K=2 case due to the decreasing amount of ℓ_k. Also, the downlink and uplink transmit powers of helpers are allocated inversely proportional to their distances to the users and the MEC server, respectively.

Performance Results for Secure Offloading in MEC System:

The performance results of the secure offloading MEC system are obtained for the case of K=2 helpers. The eavesdropper is positioned at two different distances from the helpers. Table 4.6 shows the distances between the helpers and the eavesdropper, in which $d_{e,1}$ is the distance to the Helper$_1$ and $d_{e,2}$ denotes the distance to the Helper$_2$.

Table 4.6. The distance between the helpers and eavesdropper.

Case #	$d_{e,1}$ [m]	$d_{e,2}$ [m]
1	550	300
2	750	500

We provide simulation results to evaluate the uplink delay for the proposed framework in various locations of the eavesdropper. In addition to that, we provide comparison results with the *Secure full offloading* where all helpers offload all their tasks to the MEC server for computing. In the system model, it corresponds to the case of $\alpha_k^d = 1,\ \forall k \in \mathcal{K}$. In order to provide fair comparison results, the same uplink transmit power of each helper is used in both the proposed secure MEC and secure full offloading systems. Then, the average sum secrecy rate for both the partial and the full offloading MEC systems will be the same.

In Table 4.7, the uplink delay performance versus the number of antennas is depicted for Case 1 and Case 2 with L=1.4 Mbits. It is shown that we provide secure task offloading at the expense of increasing uplink delay. As the distance between the helpers and the eavesdropper is increased for Case 2, the uplink delay reduces accordingly. The proposed cooperative MEC outperforms the secure full offloading for any number of antennas, which shows the efficiency of the partial offloading in secrecy.

Fig. 4.16 demonstrates the average sum secrecy rate versus the number of antennas for the proposed secure MEC scheme at L=1.4 Mbits. It is shown that Case 2 has a higher sum secrecy rate compared to Case 1 since the wireless channel deteriorates when the distance between the helpers and the eavesdropper is increased.

Table 4.7. The uplink delay performance for the proposed secure MEC and the secure full offloading, L=1.4 Mbits with the different number of antennas.

Comparing schemes	**Case 1**				**Case 2**			
	The uplink delay t_h, [ms]				The uplink delay t_h, [ms]			
	N=32	N=64	N=128	N=256	N=32	N=64	N=128	N=256
Secure full offloading	168.4	154.1	144.1	136	154.9	146.1	137.4	132.1
Proposed secure MEC	117.3	107.8	100.8	95.2	107.9	102.2	96.1	92.5
Proposed MEC	103.8	99	94.4	91.3	103.8	99	94.4	91.3

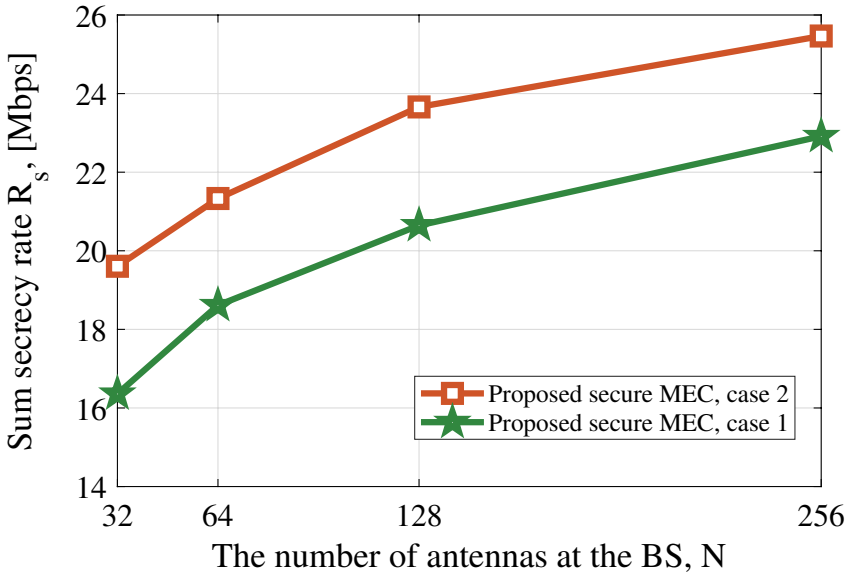


Figure 4.16. The average sum secrecy rate versus the number of antennas for the proposed secure MEC, L=1.4 Mbits.

4.3. Conclusion

In this chapter, we first focus on cooperative MEC systems to improve the offloading performance of cell-edge users through helpers. Accordingly, we have proposed the multi-helper NOMA based cooperative MEC system, where the helpers compute and offload the user's computational tasks. Specifically, we have developed the efficient framework to maximize the total offloading data subject to the latency constraints. Furthermore, the optimum distances of the helpers have been determined to obtain the maximum total offloading data. The importance of cooperation has been discussed through the total computing data which indicates the amount of user's executed data in the system. The simulation results have demonstrated that the proposed system has better performance compared to the one helper systems in terms of the total offloading data. Through the proposed framework, we have shown that both the total offloading and computing data under latency constraints are improved by employing more than one helper in the NOMA based cooperative MEC system.

On the other hand, to investigate the delay performance of the cooperative MEC systems, we have proposed a massive MIMO based cooperative secure MEC system where cooperation is established through the cell-center helpers. The overall system has been investigated as downlink transmission by applying NOMA technology and the uplink transmission by performing massive MIMO communication. We have formulated the overall delay minimization for the proposed framework under computing capability and transmit power constraints. In addition, we have investigated the proposed algorithm for a secure offloading MEC system. The simulation results have indicated that the overall delay is reduced when the number of antennas and helpers is increased in the proposed framework, which also achieves secure offloading. Moreover, the proposed secure MEC system decreases the uplink delay compared to the secure full offloading scheme while having the same secrecy rate, which shows the superiority of cooperative schemes. As a result, we have demonstrated that massive MIMO and NOMA technologies facilitate secure offloading in cooperative MEC systems.

CHAPTER 5

CONCLUSION

In this thesis, we have focused on the resource allocation algorithms for massive MIMO systems to improve the system performance of wireless communications.

In the first part of the thesis, we have introduced MIMO, massive MIMO and NOMA based communications systems. Since user selection strategies are important to maximize spectral efficiency in these systems, we have proposed user selection algorithms for densely deployed scenarios. The proposed algorithm eliminates users based on channel correlation by employing the CS algorithm, which reduces the feedback load in the system. The simulation results have shown that the proposed algorithm outperforms traditional user selection schemes in terms of data rate per user and computational complexity. Furthermore, we have proposed a user-set selection for the densely deployed scenario to improve the sum data rate of the NOMA based MIMO systems. The proposed algorithm has been designed to determine the devices incurring the least inter-set interference. For the uplink NOMA based MIMO system, the simulation results have shown that the proposed user-set selection with power allocation outperforms conventional OMA systems. Furthermore, we have performed the proposed user-set selection algorithm for an uplink NOMA based massive MIMO system to achieve a higher sum data rate than OMA based massive MIMO systems.

In the second part of the thesis, we have examined MEC systems to allow mobile devices to offload delay-sensitive and computation-intensive tasks to the nearby edge servers. We have investigated the integration of NOMA and massive MIMO into MEC systems to facilitate offloading and improve the delay performance of MEC. We have proposed the NOMA and massive MIMO assisted MEC system for delay-sensitive applications to minimize the overall computing and transmission delay under the users' transmit power and MEC computing capability. Through the pairing scheme for massive MIMO-NOMA, the users with the higher channel gain can offload all their data, while the users with the lower channel gain can offload a portion of their data to the MEC. The simulation results have shown that the sum data rate and overall system delay of the proposed system outperform the OMA-massive MIMO based and massive MIMO based MEC systems. The proposed system enables more users to offload computationally intensive tasks to the MEC while reducing the overall delay. Besides, we have examined the effect of

the user selection algorithm for the massive MIMO-NOMA based MEC system in terms of the overall delay. We have shown that user selection improves system performance by reducing the overall delay.

To further improve the MEC performance, we have presented the cooperative MEC system through helpers. In the proposed NOMA based MEC, which includes a cell-edge user and multiple helpers, the user can simultaneously offload its computation-intensive tasks to the helpers since there is no strong direct transmission link between the user and the MEC. Then, the helpers can both compute and offload these tasks. In the proposed scheme, the computation and offloading modes at the helpers are determined with respect to the optimized task offloading decision factor. The simulation results have shown that the proposed NOMA based cooperative MEC significantly increases the total offloading data under the latency constraints compared to the benchmark schemes with one helper with a strong direct transmission link. Furthermore, we have investigated the delay performance of the cooperative MEC system. Towards this end, we have presented a novel framework for a cooperative MEC system by employing massive MIMO and NOMA technologies, including security aspects. The proposed algorithm has minimized the overall delay in downlink and uplink transmission while satisfying security constraints under computing capability and transmit power constraints. The simulation results have demonstrated that massive MIMO based NOMA improves the performance of the secure MEC system by employing more than one helper.

As future works, the potential integrations of the MEC system could be addressed to further improve the system performance. The MEC system with the assistance of digital twins (DT) could be investigated to enable intelligent resource allocation and network management while creating real-time digital representation of the physical equipments. The reconfigurable intelligent surface (RIS)-assisted MEC systems could be considered to further reduce the overall delay. The low-complexity resource allocation algorithms based on Artificial Intelligence (AI)/Machine Learning (ML) could be employed in DT enabled MEC systems. Furthermore, secure offloading transmission schemes under the existence of multiple eavesdroppers could be considered in massively deployed MEC system scenarios. In addition to that, practical constraints such as asynchronous UL-NOMA and channel estimation errors in offloading schemes could be addressed in massive MIMO-MEC systems.

REFERENCES

3GPP (2010). Evolved universal terrestrial radio access (E-UTRA); further advancements for E-UTRA physical layer aspects. *TR 36.814*.

3rd Generation Partnership Project (2014). Study on LTE device to device proximity services; radio aspects. *TR 36.843*.

Abozariba, R., M. K. Naeem, M. Patwary, M. Seyedebrahimi, P. Bull, and A. Aneiba (2019). NOMA-based resource allocation and mobility enhancement framework for IoT in next generation cellular networks. *IEEE Access 7*, 29158–29172.

Amjad, M. and L. Musavian (2018). Performance analysis of NOMA for ultra-reliable and low-latency communications. In *IEEE Globecom Workshops (GC Wkshps)*, Abu Dhabi, United Arab Emirates, pp. 1–5.

Azam, I., M. B. Shahab, and S. Y. Shin (2019). User pairing and power allocation for capacity maximization in uplink NOMA. In *42nd International Conference on Telecommunications and Signal Processing (TSP)*, Budapest, Hungary, pp. 690–694.

Baidas, M. W. (2020). Offloading-efficiency maximization for mobile edge computing in clustered NOMA networks. In *IEEE 11th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, Vancouver, BC, Canada, pp. 0101–0107.

Benmimoune, M., E. Driouch, W. Ajib, and D. Massicotte (2015). Joint transmit antenna selection and user scheduling for massive MIMO systems. In *IEEE Wireless Communications and Networking Conference (WCNC)*, New Orleans, LA, USA, pp. 381–386. IEEE.

Björnson, E., J. Hoydis, and L. Sanguinetti (2017). *Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency*. Now Foundations and Trends.

Blumensath, T. and M. E. Davies (2008). Gradient pursuits. *IEEE Transactions on Signal Processing 56*(6), 2370–2382.

Cao, X., F. Wang, J. Xu, R. Zhang, and S. Cui (2018). Joint computation and communication cooperation for mobile edge computing. In *2018 16th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, pp. 1–6.

Cao, X., F. Wang, J. Xu, R. Zhang, and S. Cui (2019). Joint computation and communication cooperation for energy-efficient mobile edge computing. *IEEE Internet of Things Journal 6*(3), 4188–4200.

Chaves, R. S., M. V. S. Lima, E. Cetin, and W. A. Martins (2022). User selection for massive MIMO under line-of-sight propagation. *IEEE Open Journal of the Communications Society 3*, 867–887.

Chen, X., L. Jiao, W. Li, and X. Fu (2016). Efficient multi-user computation offloading for mobile-edge cloud computing. *IEEE/ACM Transactions on Networking 24*(5), 2795–2808.

Chen, Z., L. Zhang, Y. Pei, C. Jiang, and L. Yin (2022). NOMA-based multi-user mobile edge computation offloading via cooperative multi-agent deep reinforcement learning. *IEEE Transactions on Cognitive Communications and Networking 8*(1), 350–364.

Dai, J., A. Liu, and V. K. N. Lau (2018). FDD Massive MIMO channel estimation with arbitrary 2d-array geometry. *IEEE Transactions on Signal Processing 66*(10), 2584–2599.

Dai, J., A. Liu, and V. K. N. Lau (2019). Joint channel estimation and user grouping for massive MIMO systems. *IEEE Transactions on Signal Processing 67*(3), 622–637.

Dai, L., B. Wang, Y. Yuan, S. Han, I. Chih-lin, and Z. Wang (2015). Non-orthogonal multiple access for 5g: Solutions, challenges, opportunities, and future research trends. *IEEE Communications Magazine 53*(9), 74–81.

Dai, Y. and L. Lyu (2020). NOMA-enabled CoMP clustering and power control for green internet of things networks. *IEEE Access 8*, 90109–90117.

Davis, G. M., S. Mallat, and M. Avellaneda (1997). Adaptive greedy approximations. *Constructive Approximation 13*, 57–98.

Dierks, S. and N. Juenger (2016). Scheduling for massive MIMO with few excess antennas. In *20th International ITG Workshop on Smart Antennas (WSA)*, Munich, Germany, pp. 1–5. VDE.

Dierks, S., W. Zirwas, M. Jäger, B. Panzner, and G. Kramer (2015). MIMO and massive MIMO - Analysis for a local area scenario. In *23rd European Signal Processing Conference (EUSIPCO)*, Nice, France, pp. 2451–2455.

Ding, Z., P. Fan, and H. V. Poor (2019a). Impact of non-orthogonal multiple access on the offloading of mobile edge computing. *IEEE Transactions on Communications 67*(1), 375–390.

Ding, Z., D. W. K. Ng, R. Schober, and H. V. Poor (2018). Delay minimization for NOMA-MEC offloading. *IEEE Signal Processing Letters 25*(12), 1875–1879.

Ding, Z., J. Xu, O. A. Dobre, and H. V. Poor (2019b). Joint power and time allocation for NOMA–MEC offloading. *IEEE Transactions on Vehicular Technology 68*(6), 6207–6211.

Donoho, D. L., Y. Tsaig, I. Drori, and J.-L. Starck (2012). Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit. *IEEE Transactions on Information Theory 58*(2), 1094–1121.

Elgendy, I. A., W.-Z. Zhang, Y. Zeng, H. He, Y.-C. Tian, and Y. Yang (2020). Efficient and secure multi-user multi-task computation offloading for mobile-edge computing in mobile IoT networks. *IEEE Transactions on Network and Service Management 17*(4), 2410–2422.

Fang, F., Y. Xu, Z. Ding, C. Shen, M. Peng, and G. K. Karagiannidis (2020). Optimal resource allocation for delay minimization in NOMA-MEC networks. *IEEE Transactions on Communications 68*(12), 7867–7881.

Feng, W., J. Zheng, and W. Jiang (2020). Joint pilot and data transmission power control and computing resource allocation algorithm for massive MIMO-MEC

networks. *IEEE Access 8*, 80801–80811.

Glei, N. and R. Belgacem Chibani (2019). Power allocation for energy-efficient downlink NOMA systems. In *19th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA)*, Sousse, Tunisia, pp. 611–613.

Grace, A. (1990). *Optimization Toolbox: For Use with MATLAB: User's Guide*. Math Works Inc.

Hao, Y., Q. Ni, H. Li, and S. Hou (2019). Energy-efficient multi-user mobile-edge computation offloading in massive MIMO enabled Hetnets. In *IEEE International Conference on Communications (ICC)*, Shanghai, China, pp. 1–6.

Hawej, M. and Y. R. Shayan (2018). Iterative weighted nuclear norm minimization-based channel estimation for massive multi-user MIMO systems. In *IEEE 88th Vehicular Technology Conference (VTC-Fall)*, Chicago, IL, USA, pp. 1–5.

Hawej, M. and Y. R. Shayan (2019). Evaluation of massive MU-MIMO channel estimation based on uplink achievable-sum rate criteria. In *IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, Edmonton, AB, Canada, pp. 1–5.

Hmimz, Y., M. El Ghmary, T. Chanyour, and M. O. Cherkaoui Malki (2019). Computation offloading to a mobile edge computing server with delay and energy constraints. In *International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS)*, Fez, Morocco, pp. 1–6.

Hochwald, B., T. Marzetta, and V. Tarokh (2004). Multiple-antenna channel hardening and its implications for rate feedback and scheduling. *IEEE Transactions on Information Theory 50*(9), 1893–1909.

Huang, T., Y. Zhang, H. Wu, W. Jiang, C. Yao, M. Xu, and J. Feng (2019). Joint pilot and data transmission power control and computing resource allocation for the massive MIMO based MEC network. In *IEEE 19th International Conference on*

Communication Technology (ICCT), Xi'an, China, pp. 860–865.

Huang, Y. and Y. Liu (2018). User cooperation for NOMA-based mobile edge computing. In *IEEE International Conference on Communication Systems (ICCS)*, Chengdu, China, pp. 395–400.

Huang, Y., Y. Liu, and F. Chen (2020). NOMA-aided mobile edge computing via user cooperation. *IEEE Transactions on Communications 68*(4), 2221–2235.

Irum, T., M. Usman Ejaz, and M. Elkashlan (2022). Minimizing task offloading delay in NOMA-MEC wireless systems. In *4th Global Power, Energy and Communication Conference (GPECOM)*, Nevsehir, Turkey, pp. 632–637.

Islam, S. M. R., N. Avazov, O. A. Dobre, and K.-s. Kwak (2017). Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges. *IEEE Communications Surveys & Tutorials 19*(2), 721–742.

Khansefid, A. and H. Minn (2014). Performance bounds for massive MIMO uplink. In *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Atlanta, GA, USA, pp. 632–636.

Kim, B., W. Chung, S. Lim, S. Suh, J. Kwun, S. Choi, and D. Hong (2015). Uplink NOMA with multi-antenna. In *IEEE 81st Vehicular Technology Conference (VTC Spring)*, Glasgow, UK, pp. 1–5.

Ko, K. and J. Lee (2012). Multiuser MIMO user selection based on chordal distance. *IEEE Transactions on Communications 60*(3), 649–654.

Kyösti, P., J. Meinilä, L. Hentila, X. Zhao, T. Jämsä, C. Schneider, M. Narandzic, M. Milojević, A. Hong, J. Ylitalo, V.-M. Holappa, M. Alatossava, R. Bultitude, Y. Jong, and T. Rautiainen (2008, 02). IST-4-027756 WINNER II D1.1.2 v1.2 WINNER II channel models. *EBITG, TUI, UOULU, CU/CRC, NOKIA, Tech. Rep 11*.

Lahbib, N. D., M. Cherif, M. Hizem, and R. Bouallegue (2019). Massive MIMO uplink channel estimation using compressive sensing. In *International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, Split,

Croatia, pp. 1–6.

Lesaja, G. (2009). Introducing interior-point methods for introductory operations research courses and/or linear programming courses. *The Open Operational Research Journal 3*(1), 1–12.

Li, B., F. Si, W. Zhao, and H. Zhang (2021). Wireless powered mobile edge computing with NOMA and user cooperation. *IEEE Transactions on Vehicular Technology 70*(2), 1957–1961.

Li, H., F. Fang, and Z. Ding (2020). Joint resource allocation for hybrid NOMA-assisted MEC in 6G networks. *Digital Communications and Networks 6*(3), 241–252.

Li, J., Y. Han, S. Jin, and F.-C. Zheng (2018). User scheduling schemes of DFT-based hybrid beamforming multiuser systems. In *IEEE/CIC International Conference on Communications in China (ICCC)*, Beijing, China, pp. 200–204. IEEE.

Lia, D., N. Qin, B. Li, X. Jing, C. Du, and C. Wan (2021). Resource allocation method based on massive MIMO NOMA MEC on distribution communication network. *IOP Conference Series. Earth and Environmental Science 634*(1), 012069.

Liang, S., H. Wan, T. Qin, J. Li, and W. Chen (2020). Multi-user computation offloading for mobile edge computing: A deep reinforcement learning and game theory approach. In *IEEE 20th International Conference on Communication Technology (ICCT)*, Nanning, China, pp. 1534–1539.

Liao, H., Z. Zhou, X. Zhao, L. Zhang, S. Mumtaz, A. Jolfaei, S. H. Ahmed, and A. K. Bashir (2020). Learning-based context-aware resource allocation for edge computing-empowered industrial IoT. *IEEE Internet of Things Journal 7*(5), 4260–4277.

Lim, H. and T. Hwang (2022). Energy-efficient beamforming and resource allocation for multi-antenna MEC systems. *IEEE Access 10*, 18008–18022.

Lin, H., Y. Cao, Y. Zhong, and P. Liu (2019). Secure computation efficiency maximization in NOMA-enabled mobile edge computing networks. *IEEE*

Access 7, 87504–87512.

Liu, Y. (2019). Exploiting NOMA for cooperative edge computing. *IEEE Wireless Communications 26*(5), 99–103.

Lu, W., Y. Wang, X. Wen, X. Hua, S. Peng, and L. Zhong (2019). Compressive downlink channel estimation for FDD massive MIMO using weighted l_p minimization. *IEEE Access 7*, 86964–86978.

Luo, F.-L. and C. Zhang (2016). *Non-Orthogonal Multiple Access (NOMA): Concept and Design*, *in Signal Processing for 5G: Algorithms and Implementations*, IEEE, 2016, 143-168.

Malik, R. and M. Vu (2019). Multi-access edge computation offloading using massive MIMO. In *IEEE Global Communications Conference (GLOBECOM)*, Waikoloa, HI, USA, pp. 1–6.

Malik, R. and M. Vu (2020). Energy-efficient computation offloading in delay-constrained massive MIMO enabled edge network using data partitioning. *IEEE Transactions on Wireless Communications 19*(10), 6977–6991.

Mallat, S. and Z. Zhang (1993). Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing 41*(12), 3397–3415.

Manglayev, T., R. C. Kizilirmak, and Y. H. Kho (2016). Optimum power allocation for non-orthogonal multiple access (NOMA). In *IEEE 10th International Conference on Application of Information and Communication Technologies (AICT)*, Baku, Azerbaijan, pp. 1–4.

Mao, S., S. Leng, and Y. Zhang (2019). Joint communication and computation resource optimization for NOMA-assisted mobile edge computing. In *IEEE International Conference on Communications (ICC)*, Shanghai, China, pp. 1–6.

Mao, S., L. Liu, N. Zhang, M. Dong, J. Zhao, J. Wu, and V. C. M. Leung (2022). Reconfigurable intelligent surface-assisted secure mobile edge computing networks. *IEEE Transactions on Vehicular Technology 71*(6), 6647–6660.

Marzetta, T. L. (2010). Noncooperative cellular wireless with unlimited numbers of

base station antennas. *IEEE Transactions on Wireless Communications 9*(11), 3590–3600.

Meinilä, J., P. Kyösti, L. Hentilä, T. Jämsä, E. Suikkanen, E. Kunnari, and M. Narandzic (2010). D5.3: WINNER+ final channel models. *Wireless World Initiative New Radio WINNER*, 119–172.

Mukherjee, S. and J. Lee (2020). Edge computing-enabled cell-free massive MIMO systems. *IEEE Transactions on Wireless Communications 19*(4), 2884–2899.

Needell, D. and J. Tropp (2009). CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis 26*(3), 301–321.

Needell, D. and R. Vershynin (2010). Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit. *IEEE Journal of Selected Topics in Signal Processing 4*(2), 310–316.

Ngo, H. Q., E. G. Larsson, and T. L. Marzetta (2013). The multicell multiuser MIMO uplink with very large antenna arrays and a finite-dimensional channel. *IEEE Transactions on Communications 61*(6), 2350–2361.

Nguyen, T. T., L. B. Le, and Q. Le-Trung (2021). Computation offloading in MIMO based mobile edge computing systems under perfect and imperfect CSI estimation. *IEEE Transactions on Services Computing 14*(6), 2011–2025.

Pan, Y., M. Chen, Z. Yang, N. Huang, and M. Shikh-Bahaei (2019). Energy-efficient NOMA-based mobile edge computing offloading. *IEEE Communications Letters 23*(2), 310–313.

Pan, Y., C. Pan, K. Wang, H. Zhu, and J. Wang (2021). Cost minimization for cooperative computation framework in MEC networks. *IEEE Transactions on Wireless Communications 20*(6), 3670–3684.

Pham, Q.-V., F. Fang, V. N. Ha, M. J. Piran, M. Le, L. B. Le, W.-J. Hwang, and Z. Ding (2020). A survey of multi-access edge computing in 5g and beyond: Fundamentals, technology integration, and state-of-the-art. *IEEE Access 8*,

116974–117017.

Pham, Q.-V., L. B. Le, S.-H. Chung, and W.-J. Hwang (2019). Mobile edge computing with wireless backhaul: Joint task offloading and resource allocation. *IEEE Access 7*, 16444–16459.

Qian, L., W. Wu, W. Lu, Y. Wu, B. Lin, and T. Q. S. Quek (2021a). Secrecy-based energy-efficient mobile edge computing via cooperative non-orthogonal multiple access transmission. *IEEE Transactions on Communications 69*(7), 4659–4677.

Qian, L., Y. Wu, F. Jiang, N. Yu, W. Lu, and B. Lin (2021b). NOMA assisted multi-task multi-access mobile edge computing via deep reinforcement learning for industrial internet of things. *IEEE Transactions on Industrial Informatics 17*(8), 5688–5698.

Qiu, H., S. Gao, Y. Chen, and G. Tu (2022). Energy-efficient rate allocation for NOMA-MEC offloading under outage constraints. *IEEE Communications Letters 26*(11), 2710–2714.

Rath, G. and C. Guillemot (2009). Sparse approximation with an orthogonal complementary matching pursuit algorithm. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, Taipei, Taiwan, pp. 3325–3328.

Rauniyar, A., P. Engelstad, and O. N. Østerbø (2020). An adaptive user pairing strategy for uplink non-orthogonal multiple access. In *IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1–7.

Razi, A., D. J. Ryan, I. B. Collings, and J. Yuan (2010). Sum rates, rate allocation, and user scheduling for multi-user MIMO vector perturbation precoding. *IEEE Transactions on Wireless Communications 9*(1), 356–365.

Ren, J., G. Yu, Y. Cai, and Y. He (2018). Latency optimization for resource allocation in mobile-edge computation offloading. *IEEE Transactions on Wireless Communications 17*(8), 5506–5519.

Rusek, F., D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and

F. Tufvesson (2013). Scaling up MIMO: Opportunities and challenges with very large arrays. *IEEE Signal Processing Magazine 30*(1), 40–60.

Shental, O., B. M. Zaidel, and S. S. Shitz (2017). Low-density code-domain NOMA: Better be regular. In *IEEE International Symposium on Information Theory (ISIT)*, Aachen, Germany, pp. 2628–2632.

Shi, Z., H. Wang, Y. Fu, G. Yang, S. Ma, F. Hou, and T. A. Tsiftsis (2020). Zero-forcing-based downlink virtual MIMO-NOMA communications in iot networks. *IEEE Internet of Things Journal 7*(4), 2716–2737.

Sun, H., F. Zhou, and R. Q. Hu (2019). Joint offloading and computation energy efficiency maximization in a mobile edge computing system. *IEEE Transactions on Vehicular Technology 68*(3), 3052–3056.

Sun, Y., Z. Su, Y. Zhao, D. Deng, F. Zhu, and J. Xia (2020). Mobile cooperative sensing based secure communication strategy of edge computational networks for smart cities. *IEEE Access 8*, 150750–150758.

Tan, L., Z. Kuang, L. Zhao, and A. Liu (2021). Energy-efficient joint task offloading and resource allocation in ofdma-based collaborative edge computing. *IEEE Transactions on Wireless Communications*, 1–1.

Thet, N. W. M., S. Khan, E. Arvas, and M. K. Özdemir (2020). Impact of mutual coupling on power-domain non-orthogonal multiple access (NOMA). *IEEE Access 8*, 188401–188414.

Tiwari, A., T. Goyal, and S. Gurugopinath (2020). Latency minimization in uplink non-orthogonal multiple access-based mobile edge computing. In *IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, Bangalore, India, pp. 1–5.

Tran, L.-N., M. Bengtsson, and B. Ottersten (2012). Iterative precoder design and user scheduling for block-diagonalized systems. *IEEE Transactions on Signal Processing 60*(7), 3726–3739.

Waltz, R. A., J. L. Morales, J. Nocedal, and D. Orban (2006, Jul.). An interior

algorithm for nonlinear optimization that combines line search and trust region steps. *Mathematical Programming 107*(3), 391–408.

Wan, D., M. Wen, X. Cheng, S. Mumtaz, and M. Guizani (2019). A promising non-orthogonal multiple access based networking architecture: Motivation, conception, and evolution. *IEEE Wireless Communications 26*(5), 152–159.

Wang, F., J. Xu, and Z. Ding (2019). Multi-antenna NOMA for computation offloading in multiuser mobile edge computing systems. *IEEE Transactions on Communications 67*(3), 2450–2463.

Wang, J.-B., H. Yang, M. Cheng, J.-Y. Wang, M. Lin, and J. Wang (2020). Joint optimization of offloading and resources allocation in secure mobile edge computing systems. *IEEE Transactions on Vehicular Technology 69*(8), 8843–8854.

Wang, K., F. Fang, D. B. d. Costa, and Z. Ding (2021). Sub-channel scheduling, task assignment, and power allocation for OMA-based and NOMA-based MEC systems. *IEEE Transactions on Communications 69*(4), 2692–2708.

Wang, K., H. Li, Z. Ding, and P. Xiao (2022). Reinforcement learning based latency minimization in secure NOMA-MEC systems with hybrid SIC. *IEEE Transactions on Wireless Communications*, 1–1.

Wang, X., J. Wang, L. He, and J. Song (2018). Outage analysis for downlink NOMA with statistical channel state information. *IEEE Wireless Communications Letters 7*(2), 142–145.

Wang, X., W. Wu, B. Lyu, and H. Wang (2019a). Delay minimization for secure NOMA mobile-edge computing. In *IEEE 19th International Conference on Communication Technology (ICCT)*, Xi'an, China, pp. 1529–1534.

Wei, Z., L. Yang, D. W. K. Ng, J. Yuan, and L. Hanzo (2020). On the performance gain of NOMA over OMA in uplink communication systems. *IEEE Transactions on Communications 68*(1), 536–568.

Wen, Y., X. Zhou, F. Fang, H. Zhang, and D. Yuan (2020). Joint time and power

allocation for cooperative NOMA based MEC system. In *IEEE 92nd Vehicular Technology Conference (VTC2020-Fall)*, Victoria, BC, Canada, pp. 1–5.

Wu, Q., W. Chen, D. W. K. Ng, and R. Schober (2018). Spectral and energy-efficient wireless powered IoT networks: NOMA or TDMA? *IEEE Transactions on Vehicular Technology 67*(7), 6663–6667.

Wu, Y., G. Ji, T. Wang, L. Qian, B. Lin, and X. Shen (2022). Non-orthogonal multiple access assisted secure computation offloading via cooperative jamming. *IEEE Transactions on Vehicular Technology 71*(7), 7751–7768.

Xiao, L., X. Lu, T. Xu, X. Wan, W. Ji, and Y. Zhang (2020). Reinforcement learning-based mobile offloading for edge computing against jamming and interference. *IEEE Transactions on Communications 68*(10), 6114–6126.

Xu, G., A. Liu, W. Jiang, H. Xiang, and W. Luo (2014). Joint user scheduling and antenna selection in distributed massive MIMO systems with limited backhaul capacity. *China Communications 11*(5), 17–30.

Xu, J., P. Zhu, J. Li, and X. You (2021). Secure computation offloading for multi-user multi-server MEC-enabled IoT. In *IEEE International Conference on Communications (ICC)*, Montreal, QC, Canada, pp. 1–6.

Yang, H., J.-B. Wang, M. Cheng, C. Chang, J.-Y. Wang, M. Lin, and M. Chen (2019b). Secure resource allocation in mobile edge computing systems. In *IEEE Global Communications Conference (GLOBECOM)*, Waikoloa, HI, USA, pp. 1–6.

Yang, Z., J. Hou, and M. Shikh-Bahaei (2018). Energy efficient resource allocation for mobile-edge computation networks with NOMA. In *IEEE Globecom Workshops (GC Wkshps)*, Abu Dhabi, United Arab Emirates, pp. 1–7.

Yang, Z., C. Pan, J. Hou, and M. Shikh-Bahaei (2019a). Efficient resource allocation for mobile-edge computing networks with NOMA: Completion time and energy minimization. *IEEE Transactions on Communications 67*(11), 7771–7784.

Yao, M., L. Chen, T. Liu, and J. Wu (2019). Energy efficient cooperative edge computing with multi-source multi-relay devices. In *IEEE 21st International*

Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), Zhangjiajie, China, pp. 865–870.

Ye, Y., G. Lu, R. Q. Hu, and L. Shi (2019). On the performance and optimization for MEC networks using uplink NOMA. In *IEEE International Conference on Communications Workshops (ICC Wkshps)*, Shanghai, China, pp. 1–6.

Yılmaz, S. S. and B. Özbek (2020). Compressive sensing based low complexity user selection for massive MIMO systems. In *IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, Antwerp, Belgium, pp. 1–5.

Yılmaz, S. S. and B. Özbek (2022). Multi-helper NOMA for cooperative mobile edge computing. *IEEE Transactions on Intelligent Transportation Systems 23*(7), 9819–9828.

Yılmaz, S. S. and B. Özbek (2023). Massive MIMO-NOMA based MEC in task offloading for delay minimization. *IEEE Access 11*, 162–170.

Yılmaz, S. S., B. Özbek, M. İlgüy, B. Okyere, L. Musavian, and J. Gonzalez (2022). User selection for NOMA-based MIMO with physical-layer network coding in Internet of Things applications. *IEEE Internet of Things Journal 9*(16), 14998–15006.

Yılmaz, S. S., B. Özbek, and R. Mumtaz (2023). Delay minimization for massive MIMO based cooperative mobile edge computing system with secure offloading: (Invited paper). *IEEE Open Journal of Vehicular Technology 4*, 149–161.

Yoo, T. and A. Goldsmith (2006). On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming. *IEEE Journal on Selected Areas in Communications 24*(3), 528–541.

Yu, W., L. Musavian, and Q. Ni (2018). Link-layer capacity of NOMA under statistical delay QoS guarantees. *IEEE Transactions on Communications 66*(10), 4907–4922.

Zafar, A., M. Shaqfeh, M.-S. Alouini, and H. Alnuweiri (2013). On multiple users scheduling using superposition coding over Rayleigh fading channels. *IEEE Communications Letters 17*(4), 733–736.

Zeng, M., W. Hao, O. A. Dobre, Z. Ding, and H. V. Poor (2020a). Massive MIMO-assisted mobile edge computing: Exciting possibilities for computation offloading. *IEEE Vehicular Technology Magazine 15*(2), 31–38.

Zeng, M., W. Hao, O. A. Dobre, and H. V. Poor (2020). Delay minimization for massive MIMO assisted mobile edge computing. *IEEE Transactions on Vehicular Technology 69*(6), 6788–6792.

Zhang, L., J. Wu, S. Mumtaz, J. Li, H. Gacanin, and J. J. P. C. Rodrigues (2019). Edge-to-edge cooperative artificial intelligence in smart cities with on-demand learning offloading. In *IEEE Global Communications Conference (GLOBECOM)*, Waikoloa, HI, USA, pp. 1–6.

Zhang, Y., J. Zhang, J. Zhang, G. Liu, Y. Zhang, and Y. Yao (2022). A theoretical analysis of favorable propagation on massive MIMO channel with generalized angle distributions. *Electronics 11*(14), 2150.

Zhang, Z., Y. Fu, G. Cheng, X. Lan, and Q. Chen (2021). Secure offloading design in multi-user mobile-edge computing systems. In *IEEE 6th International Conference on Computer and Communication Systems (ICCCS)*, Chengdu, China, pp. 695–703.

Zhao, C., Y. Cai, A. Liu, M. Zhao, and L. Hanzo (2020a). Mobile edge computing meets mmwave communications: Joint beamforming and resource allocation for system delay minimization. *IEEE Transactions on Wireless Communications 19*(4), 2382–2396.

Zhao, C., Y. Cai, M. Zhao, and Q. Shi (2019). Joint hybrid beamforming and offloading for mmWave mobile edge computing systems. In *IEEE Wireless Communications and Networking Conference (WCNC)*, Marrakesh, Morocco, pp. 1–6.

Zhao, W., B. Wang, H. Bao, and B. Li (2020b). Secure energy-saving resource allocation on massive MIMO-MEC system. *IEEE Access 8*, 137244–137253.

Zheng, K., S. Ou, , and X. Yin (2014). Massive MIMO channel models: A survey. *International Journal of Antennas and Propagation 2014*(11), 1–10.

Zhou, Y., P. L. Yeoh, C. Pan, K. Wang, M. Elkashlan, Z. Wang, B. Vucetic, and Y. Li (2020). Offloading optimization for low-latency secure mobile edge computing systems. *IEEE Wireless Communications Letters 9*(4), 480–484.

Zhu, F., Y. Huang, Y. Liu, and X. Zhang (2020). Cooperative computation offloading in NOMA-based edge computing. In *IEEE/CIC International Conference on Communications in China (ICCC)*, Chongqing, China, pp. 208–213.

Zhu, L., J. Zhang, Z. Xiao, X. Cao, and D. O. Wu (2019). Optimal user pairing for downlink non-orthogonal multiple access (NOMA). *IEEE Wireless Communications Letters 8*(2), 328–331.

Zuo, H. and X. Tao (2017). Power allocation optimization for uplink non-orthogonal multiple access systems. In *9th International Conference on Wireless Communications and Signal Processing (WCSP)*, Nanjing, China, pp. 1–5.

VITA

Saadet Simay Yılmaz received the B.Sc. degree from the Department of Electrical and Electronics, the minor degree from the Department of Chemistry, Izmir Institute of Technology, in 2014, and the M.Sc. degree from the Department of Electrical and Electronics, Izmir Institute of Technology, in 2017.

During the M.Sc. program, she worked on a project supported by the Republic of Turkey Ministry of Science, Industry and Technology under SAN-TEZ 0686.STZ.2014 Programme.

Since 2016, she has been a Research Assistant with the Department of Electrical and Electronics Engineering, Izmir Institute of Technology. She is also a Researcher of an International Project supported by European Union under Horizon 2020-MSCA-RISE Programme. In the framework of this project, she visited GreenSphere Unipessoal-LDA, at Aveiro, Porto, Portugal, to study mobile edge computing systems and YL-Verkot Oy at Tampere, Finland, to study Hardware-in-the-loop testbed architecture.

Her current research interests include wireless communications, 5G, beyond 5G, massive MIMO systems, mobile edge computing and digital twin.