Performance Evaluation of Cluster-based Concurrent Uplink Transmissions in MIMO-NOMA Networks

Abhishek Kumar*, Frank Y. Li*, and Jorge Martinez-Bauset†
*Dept. of Information and Communication Technology, University of Agder (UiA), N-4898 Grimstad, Norway
†Departamento de Comunicaciones, Universitat Politècnica de València (UPV), València 46022, Spain
Email: {abhishek.kumar, frank.li}@uia.no; jmartinez@upv.es

***Abstract*—To obtain benefits for non-orthogonal multiple access (NOMA) based concurrent transmissions for uplink Internet of things (IoT) traffic in multi-antenna enabled networks, device clustering has been a challenging task. Most previous studies focused on grouping devices into clusters that are served by different beams in multiple input multiple output (MIMO)-NOMA networks. In this paper, we perform an exploratory study on the impact of both intra- and inter-cluster interference on the performance of uplink concurrent transmissions considering that multiple clusters are served by a single beam. For performance assessment, we define two metrics, cluster throughput and transmission latency, and evaluate network performance with various network configurations. The study provides insight on how to configure device clusters and perform access control in order to maximize performance and improve fairness. As devices closer to the base station experience less path attenuation, we introduce distinct access control mechanisms to improve transmissions fairness for concurrent transmissions from difference clusters.**

***Index Terms*—Multi-antenna, non-orthogonal multiple access, uplink traffic, concurrent transmissions, performance assessment.**

I. Introduction

Beyond fifth generation (B5G) communication systems are expected to handle a huge number of simultaneously connected users, facilitating diverse categories of applications. In addition to supporting traditional enhanced mobile broadband (eMBB) services for human-type communications (HTC), use cases for massive machine-type communications (mMTC) or massive Internet of things (mIoT), and ultra reliable low latency communications (URLLC) will bring many different types of connected users into B5G systems. According to [1], the estimated device density for B5G mMTC/mIoT applications will be 1$\sim$10 million devices per square kilometer. Under such circumstances, it is imperative to design effective mechanisms for radio resource allocation and network management.

Multiple access is a vital procedure for wireless connectivity in all generations of mobile networks. A primary task for any multiple access scheme is to efficiently and fairly distribute radio resources, in both frequency and time domains, among a large number of devices, such as HTC user equipment (UE), tablets, sensors, IoT devices. Traditionally, orthogonal multiple access (OMA) based schemes where each device occupies dedicated radio resources have been a dominant solution until 5G. For B5G system, non-orthogonal multiple access (NOMA) schemes which allow concurrent transmissions of multiple users on the same (time and frequency) resource have attracted lots of attentions in recent years. By performing successive interference cancellation (SIC), a receiver is able to retrieve the signals from multiple devices even if they transmit concurrently using the same radio resource. Thanks to the capability of resource sharing, NOMA systems achieve higher cumulative capacity than orthogonal multiple access systems, appearing as a strong potential candidate for multiple access in B5G systems.

Although NOMA outperforms OMA in many cases and offers a number of advantages for lifting accumulative rate, the complexity of user pairing algorithms increases exponentially with the number of users and NOMA protocols' scalability performance appears shaky. On the other hand, employing multiple antennas at a base station (BS), which takes the advantage of spatial diversity with beamforming, is another strategy to increase system capacity [2]. Furthermore, combining NOMA with beamforming is a powerful technique for increasing capacity efficiency in B5G networks by exploiting the benefits from both the power and the space domains. Through beamforming, a beam of signals is employed to accommodate a group of users with the same or a similar angle towards the BS. Within that group, the channels are divided into multiple single input single output channels. Different beams are assigned to different groups of devices so that the number of devices covered by each beam is much reduced.

A major issue when implementing NOMA in a multi-antenna network is user clustering. It is a challenging task to decide how to associate users to clusters, and how to serve these clusters by different beams, in order to exploit the benefits of NOMA. Numerous studies investigated the circumstances where adding NOMA to spatial multiplexing might be beneficial. For uplink transmissions in a multi-antenna network, a strong-weak user pairing strategy that is determined by the amount of channel path loss was presented in [3]. The authors in [4] proposed a mmWave-NOMA system and analyzed the rate maximization problem by decomposing it into a user clustering and ordering problem, followed by power allocation. Furthermore, [5] proposed a scheme optimizing user clustering, power, and resource (time slot or bandwidth) allocation for downlink transmissions in NOMA-OMA hybrid systems. In a multiple input multiple output (MIMO)-NOMA scenario, [6] proposed both resource conscious user clustering and learning-assisted user clustering. Moreover, the effect of mobility on NOMA interference was addressed in [7].

However, the majority of existing studies concentrate on user grouping under distinct beams in uplink MIMO-NOMA

networks. So far, user clustering under a single beam for uplink transmissions in multi-antenna NOMA systems has received little attention. In this paper, we study multi-antenna based NOMA systems for uplink IoT traffic, focusing on devices associated into different clusters which are served by a single beam. We evaluate the performance of uplink data transmissions considering both inter- and intra-cluster interference, as well as channel conditions. We measure the performance by two metrics, cluster throughput and transmission delay. We evaluate network performance under multiple scenarios through extensive simulations. Our results shed light on how to configure device clusters in order to maximize performance in such a network. As we observe that devices closer to the BS receive better service, two access control schemes are investigated to improve performance fairness.

The remainder of the paper is structured as follows. Following a description of the network scenario in Sec. II, we present the developed access scheme for uplink traffic in Sec. III. Afterwards, simulations are conducted with different network configurations and numerical outcomes are discussed in Sec. IV. Finally, the paper is concluded in Sec. V.

II. Network Scenario

Consider a single cell in a wireless network with multiple antennas mounted at the BS. We focus on uplink traffic from multiple devices based on the NOMA principle. All devices are simple battery-powered mMTC/mIoT devices and are stationary, each equipped with a single antenna. It is assumed that each device uses the same amount of transmit power. Given the simplicity of the devices, automatic power control is not considered [8]. Although multiple beams may exist, we focus on one beam towards a certain direction covering a set of devices. Devices disposed at different geographical locations are grouped into clusters, where each cluster is typically composed by multiple devices. In this network, devices in all clusters are served by the same antenna beam and concurrent transmissions are enabled.

Fig. 1 illustrates the scenario under study, where multiple devices grouped into clusters contend to transmit uplink traffic to a BS equipped with M antennas following the NOMA transmission principle. The antenna is situated at a height of $\mathbf{g}$ meters above the ground. For the sake of simplicity, we consider only two clusters in this paper, C_j ($j = 1, 2$), each containing N devices, and the center of C_1 and C_2 is located d_1 and d_2 meters away from the BS, respectively. Devices belonging to the same cluster are uniformly spread within a certain radius from the cluster center. Furthermore, we assume that cluster C_1 is geographically located nearer to the BS than cluster C_2. Thus, the total received signal y at the BS including the transmissions from both clusters is given by

$$y = \sum_{j=1}^{2} \sum_{i=1}^{N} \mathbf{H}_i^j x_i^j + n \tag{1}$$

where x_i^j, $\mathbf{H}_i^j$, and n stand for the transmitted signal by device D_i^j, complex channel gain vector between the BS and D_i^j, and additive noise, respectively. The signal transmitted by device D_i^j, denoted by x_i^j, is expressed as follows:

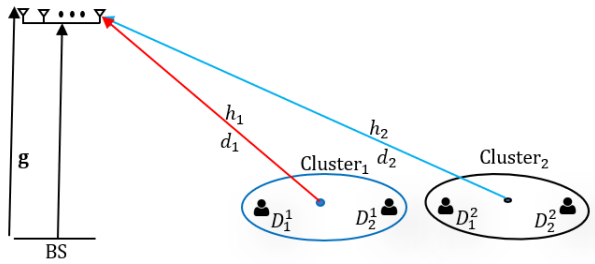

Fig. 1. Network scenario for cluster-based uplink NOMA transmissions.

$$x_i^j = \sqrt{P} s_i^j \ , \tag{2}$$

where P refers to the maximum amount of transmit power available at each device, and s_i^j is the transmitted signal of each device with unit variance. The additive noise present in the channel is assumed to have a Gaussian distribution with zero mean and variance σ^2, i.e., $n \sim \mathcal{CN}\left(0, \sigma^2\right)$. Additionally, we assume that the complex channel gain vector, denoted by $\mathbf{H}_i^j \sim \mathcal{CN}\left(\mathbf{0}_M, \mathbf{I}_M\right)$, follows Rayleigh fading with zero-mean complex Gaussian distribution.

The time is slotted and devices from both clusters transmit their signals using the same radio resource. Whether the transmission of a device is successful or not depends on the interference level generated by other concurrent transmission(s), as well as the channel condition. The BS may be able to recover the original signal(s) via SIC, considering the specific ordering of the received signal power levels from the other transmissions. The decoding order of the uplink received signals is determined by the channel gain, normalized by the noise power, in a descending order. When the BS is able to decode the strongest signal, given that its signal-to-inference ratio (SINR) is sufficiently large, it can subtract the decoded signal from the received signal and then try to decode the next strongest signal. The decoding process ends when the SINR of a signal is insufficient [9]. When only one device transmits in a timeslot, whether the signal can be decoded or not depends on channel condition. The transmission will be successful if the channel condition is good enough; otherwise the transmission failed.

III. Data Transmissions and Performance Metrics

In this section we present the principle of concurrent uplink transmissions in the envisaged network and define two metrics that will be used later for performance evaluation.

A. Transmission Principle

1) Transmit power and path loss: For each transmission, all devices draw the same amount of transmit power and occupy the same radio resource. Within each timeslot, one or more transmissions may occur, initiated from devices in the same or different clusters.

The path loss between a device and the BS depends on the exact location of each device and channel gains may vary from

device to device. There are also variations in SINR at the BS, since devices inside the same cluster are located nearby, but devices belonging to different clusters may be far away from each other. The path loss is calculated as 128 + 37.6log$_{10}d$, where d is the distance between a device and the BS expressed in kilometers.

2) Device behavior and message transmission: The BS has information about device identity and cluster location. A device is regarded as active if it has one packet in its buffer. Let $n_j \in [0, 1, 2]$ indicate the number of active devices in cluster C_j where $j = 1, 2$. A two dimensional system state with n_1 active devices in cluster C_1 and n_2 active devices in cluster C_2 is denoted by $\mathbf{s} = (n_1, n_2)$. For example, at state $\mathbf{s} = (0, 1)$, there is only a single active device at C_2 and no active devices at C_1, i.e., no concurrent transmissions occur. At state $\mathbf{s} = (2, 2)$, there are four active devices in this network, and a total of four concurrent transmissions will occur in a timeslot. At each timeslot, the same radio resources are used by all active devices to transmit their packets to the BS [11].

3) Medium access and message detection: A multi-user detection mechanism based on the SIC principle is utilized at the BS to detect the messages transmitted by devices. It is worth highlighting that to decode successfully two or more messages sent by different devices using SIC, the received signal strength differences at the BS from various devices have to be sufficiently large. If this requirement is met, the BS will be able to differentiate the messages sent by distinct devices[1]. As a result, the strongest received signal, which corresponds to the transmission with the highest channel gain is decoded first, by treating the signals sent from other devices as interference. Following the SIC principle, the other signals may also be decoded by subtracting the already decoded strongest signal(s).

4) Access control schemes: To achieve optimal network performance, a suitable access control mechanism may be required for mMTC/mIoT traffic. To this end, we introduce the concept of access probability, $\theta \in [0, 1]$, which is multicasted by the BS to devices in all clusters before they start to transmit. Whether an active device will transmit a message in the imminent timeslot or not depends on the value of θ. A message that is currently in the buffer will be sent in the beginning of the current timeslot if θ is lower than a specific threshold. Otherwise, the transmission attempt is postponed. In this paper, we have investigated two different mechanisms for access control as presented later in Subsec. IV-C.

B. Performance Parameters

Two performance parameters have been defined in this study, as cluster throughput and transmission latency.

1) Cluster throughput: It is defined as the average number of messages that are successfully transmitted per cluster per timeslot. The cluster throughput of cluster C_j $(j = 1, 2)$ is denoted by $Th_j(\boldsymbol{s})$, when the system is in state $\boldsymbol{s}$.

[1]According to real-life experiments performed in [10], this SINR difference threshold, denoted by α, is $\alpha \geq 10$ dB.

TABLE I
NETWORK PARAMETERS CONFIGURED IN OUR SIMULATIONS [9]

Parameter	Value
System type	Single cell
Number of clusters	2
Cluster radius	25 m
Antenna height (**g**)	30 m
Maximum active devices in a cluster	2
Path loss model	128 + 37.6log$_{10}$ d
Standard deviation for shadow fading	8 dB
Total transmit power (P)	23 dBm

2) Transmission latency: It is defined as the average number of timeslots required to transmit a message successfully per device. The transmission latency experienced by a message in C_j $(j = 1, 2)$ is denoted by $L_j(\boldsymbol{s})$, when the system is in state $\boldsymbol{s}$. Note that the system behavior in a time slot is independent of its behavior in any other timeslot.

IV. SIMULATIONS AND NUMERICAL RESULTS

Consider a single-cell multi-antenna system that is illustrated in Fig. 1, where one BS serves two clusters within a single beam with two devices in each cluster. To assess network performance under various system configurations, we have performed extensive Monte Carlo simulations based on MATLAB. Except the results presented in Subsec. IV-C, we configure θ=1.

To obtain $Th_j(\boldsymbol{s})$, a single simulation run evaluates the outcome of each timeslot when the system is in state $\boldsymbol{s} = (n_1, n_2)$. That is, the number of successfully decoded messages at the BS in a timeslot when the system is in state $\boldsymbol{s}$. To obtain transmission latency, a simulation run might be composed of multiple timeslots. At each system state $\boldsymbol{s}$, we keep track on the number of retransmissions a failed message experienced in successive independent timeslots until all packets have been successfully received. The $Th_j(\boldsymbol{s})$ and $L_j(\boldsymbol{s})$ results presented hereafter are the average values obtained from 10^6 simulation runs.

A. Simulation Scenarios and Network Configurations

The height of the antenna at the BS is configured to be $\mathbf{g} = 30$ meters. The radius of both clusters is 25 meters, and the center of each cluster is located d_1 and d_2, where $d_1 < d_2$, meters away from the BS for C_1 and C_2, respectively. Furthermore, the devices in each cluster are uniformly distributed, and their distances to the BS vary slightly depending on their exact location within the cluster. The transmit power for all the devices is identical, and configured to P = 23 dBm. A list of the key network parameters configured in our evaluation scenarios is provided in Tab. I.

B. Performance Evaluation: Cluster Throughput and Latency

To assess network performance in terms of the two metrics defined above, we configure the studied network with a variety of system parameters including distances and number of

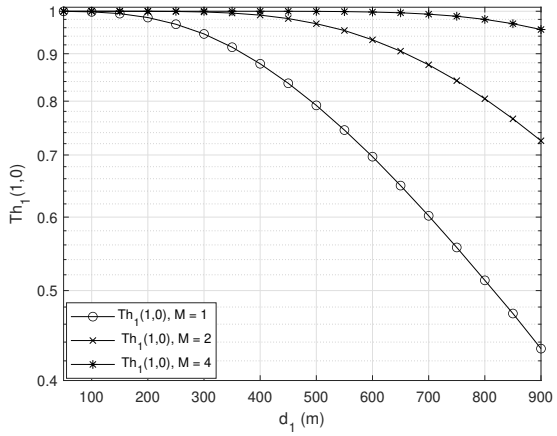

Fig. 2. Throughput $Th_1\,(1,0)$ as distance d_1 and antenna number M vary.

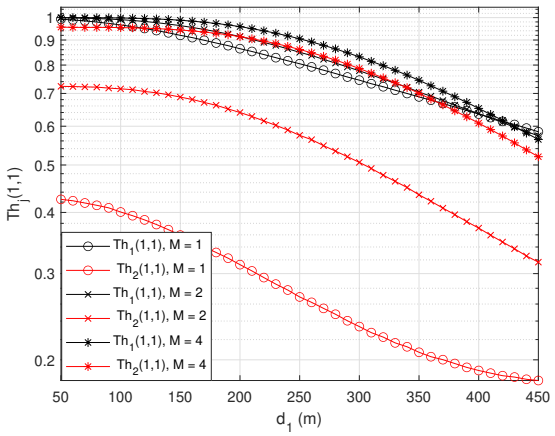

Fig. 3. Throughput $Th_j\,(1,1)$ with d_1 and M for a fixed $d_2 = 900$ m.

TABLE II
TRANSMISSION LATENCY FOR STATE $\boldsymbol{s} = (1,0)$

d (m)	$M=1$	$M=2$	$M=4$
900	2.8700	1.6530	1.0905
500	1.4725	1.0617	1.0006

antennas at the BS, which is denoted by M. The results based on five representative states are reported in this subsection.

1) Performance of state $\boldsymbol{s} = (1,0)$: The curves in Fig. 2 show the evolution of $Th_1\,(1,0)$ with d_1 and M, when there is only one active device in the network. Clearly, the obtained cluster throughput improves as more antennas are available but it deteriorates as the distance to the BS becomes larger. This is because as the number of antennas at the BS increases, so does the received SINR at the BS. As a result, the probability of successfully decoding a packet increases, leading to higher cluster throughput. On the other hand, a larger distance between transmitter and receiver causes more serious path loss and correspondingly a weaker SINR at the BS, leading to lower cluster throughput. This behavior is also observed in the measured transmission latency which is shown in Tab. II. That is, more antennas will help reducing latency, whereas larger distances will lead to longer latency.

2) Performance of state $\boldsymbol{s} = (1,1)$: The curves in Fig. 3 reveal the evolution of $Th_j\,(1,1)$ for both C_1 and C_2 as d_1 and M vary, where the distance from the BS to C_2 is set to a fixed value of $d_2 = 900$ meters. In this system state, each cluster only has one active device and each device has an equal opportunity to occupy the same radio resource whenever it has a packet to transmit. Concurrent transmissions, one from each cluster, occur in this state.

As observed, for a fixed number of antennas at the BS, $Th_j\,(1,1)$ decreases as d_1 becomes larger. As C_1 is closer to the BS than C_2 is, we observe that $Th_1\,(1,1) > Th_2\,(1,1)$. Furthermore, for a given distance of d_1, the variation of $Th_1\,(1,1)$ is less significant with more antennas. However, for C_2 which is located farther away from the BS than C_1 is, $Th_2\,(1,1)$ improves significantly as M increases. This is because the inter-cluster interference increases as d_1 increases, given that d_2 and M are kept fixed. At the maximum distance of $d_1 = 450$ m for C_1, the SINR of C_1 signal is still high enough and the fraction of successful signal decoding is hardly affected. Then, increasing the number of antennas does not have significant impact on throughput performance. On the other hand, for the farthest located cluster C_2, $Th_2\,(1,1)$ significant improves as M increases, thanks to the resulted channel gain improvement when more antennas are deployed.

TABLE III
TRANSMISSION LATENCY FOR $\boldsymbol{s} = (1,1)$, $d_1 = d_2/2$

d_2 (m)	M	$L_1\,(\boldsymbol{s})$	$L_2\,(\boldsymbol{s})$
900	1	2.1242	6.3310
-	2	2.1698	3.8459
-	4	2.2089	2.4029
500	1	2.2788	3.3260
-	2	2.2765	2.4108
-	4	2.188	2.188

The average transmission latency for both clusters at different distances d_1 and d_2, and number of antennas M, is shown in Tab. III. The values in this table reveal that the transmission latency decreases noticeably as M increases, and the latency increases with a longer distance.

3) Performance of state $\boldsymbol{s} = (1,2)$: Let us now assess network performance when there are 1 and 2 devices in C_1 and C_2 respectively, where $d_2 = 900$ m. As can be observed in Fig. 4, when d_1 is small, the throughput of the cluster closer to the BS, $Th_1\,(1,2)$, increases with more antennas. However, as d_1 approaches 450 m, increasing M might have a negligible or even a negative effect on $Th_1\,(1,2)$. This is due to the fact that inter-cluster interference becomes more serious when d_1 increases given that d_2 is fixed. When d_1 is sufficiently long, inter-cluster interference dominates and adding more antennas does not help increasing sufficiently the SINR of the C_1 signal received at the BS for decoding. For C_2, which is the cluster located farther away from the BS, the observed behavior is slightly different. For a fixed d_1 value, $Th_2\,(1,2)$ initially increases and then begins to decline when the number of antennas increases. This is because the impact of intra-device interference is not relevant when M is small. As more antennas are available, intra-device interference becomes more significant, leading to reduced $Th_2\,(1,2)$. The same behavior trend applies also to transmission latency, as shown in Tab. IV.

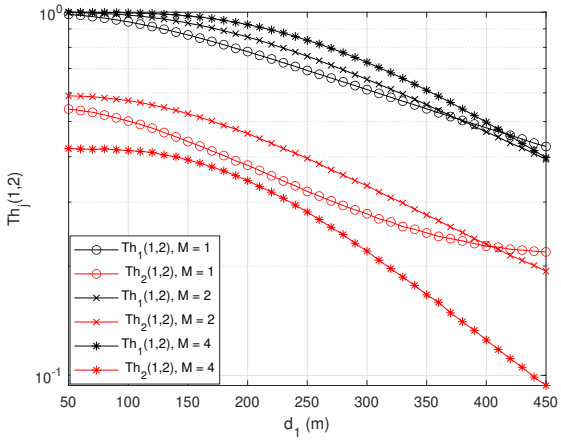

Fig. 4. Throughput $Th_j\,(1,2)$ with d_1 and M for a fixed $d_2 = 900$ m.

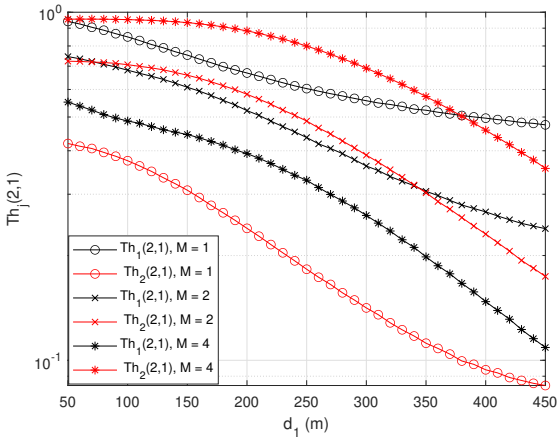

Fig. 5. Throughput $Th_j\,(2,1)$ with d_1 and M for a fixed $d_2 = 900$ m.

TABLE IV
TRANSMISSION LATENCY FOR $\boldsymbol{s_1} = (1,2)$, $\boldsymbol{s_2} = (2,1)$, $d_1 = d_2/2$

d_2 (m)	M	$L_1\,(\boldsymbol{s_1})$	$L_2\,(\boldsymbol{s_1})$	$L_1\,(\boldsymbol{s_2})$	$L_2\,(\boldsymbol{s_2})$
900	1	2.9116	10.0388	4.9813	12.7722
-	2	3.1511	11.2604	9.2570	6.5631
-	4	3.1148	22.2910	19.4156	3.4467
500	1	3.2944	9.8232	7.4926	5.3821
-	2	3.2377	14.1996	12.7828	3.4870
-	4	3.1396	23.5182	19.652	3.1349

4) Performance of state $\boldsymbol{s} = (2,1)$: This state represents another network configuration with two active devices in C_1 and one active device in C_2 where d_2 is still fixed as $d_2 = 900$ m. As shown in Fig. 5, the throughput of the cluster that is the farthest away from the BS, $Th_2(2,1)$, increases with more antennas for small values of d_1. However, the behavior is inverse for the cluster that is closer to the BS. A shorter inter-cluster distance and more antennas will result in lower throughput for cluster C_1, $Th_1(2,1)$. This is due to the fact that deploying more antennas for a given distance leads to deteriorated SINR for decoding the signals received from C_1 but improved SINR for the signals received from C_2. Furthermore, with a longer distance of d_1, the cluster throughput for both clusters falls. This is because when distance d_1 grows, the SINR for cluster C_1 falls and a shorter inter-cluster distance causes stronger inter-cluster interference. For transmission latency, the same behavior trend has been observed, as evident from Tab. IV.

5) Performance of state $\boldsymbol{s}$ = (2,2): Lastly, we examine the network performance when both clusters have two active devices, i.e., state $\mathbf{s}$ = (2,2) with varying d_1 and M, and fixed d_2 = 900 m. As shown in Fig. 6, the cluster throughput of both clusters descends with a shorter inter-cluster distance (i.e., a larger d_1) for a given M. With more antennas, the throughput difference between the two clusters becomes smaller as the inter-cluster distance reduces. When the inter-cluster distance is sufficiently low (about $d_2 - d_1 = 650$ m), this difference becomes insignificant. An intriguing finding regarding this behavior is that *with more antennas and shorter inter-cluster distance, the cluster throughput intends to be nearly identical for all clusters.* The same trend is also reflected in the latency performance as shown in Tab. V.

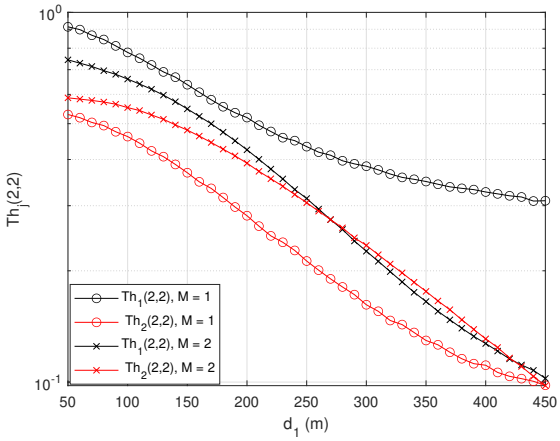

Fig. 6. Throughput $Th_j\,(2,2)$ with d_1 and M, for a fixed $d_2 = 900$ m.

C. Performance Evaluation: Access Control Mechanisms

To further assess the impact of access control, we investigate cluster throughput by introducing the following two access control mechanisms (ACMs) following the same principle.

- ACM-1: The same access probability applies to both clusters. Devices in all clusters will select a random value uniformly between 0 and 1. If the selected value by a device is less than θ which is broadcasted by the BS, it will transmit in the current timeslot. Otherwise, the transmission is postponed.
- ACM-2: Distinct access probabilities for different clusters. Since the cluster closer to the BS experiences less path loss, the BS may assign a higher access probability to the cluster that is farther away for the purpose of achieving fairer access among devices from different clusters. The value of θ is re-configurable for each timeslot.

TABLE V
TRANSMISSION LATENCY FOR $\boldsymbol{s} = (2,2)$, $d_1 = d_2/2$

d_2 (m)	M	$L_1(\boldsymbol{s})$	$L_2(\boldsymbol{s})$
900	1	7.3336	21.2576
-	2	20.4300	21.3219
-	4	46.2076	47.4393
500	1	16.0301	19.6514
-	2	29.8769	33.4475
-	4	46.5151	63.6770

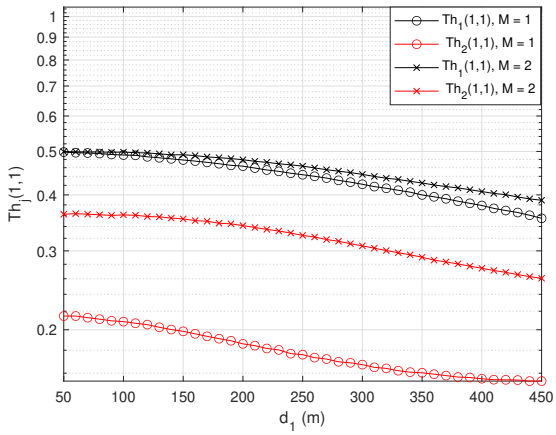

Fig. 7. $Th_j\ (1,1)$ for ACM-1: The same θ for both clusters.

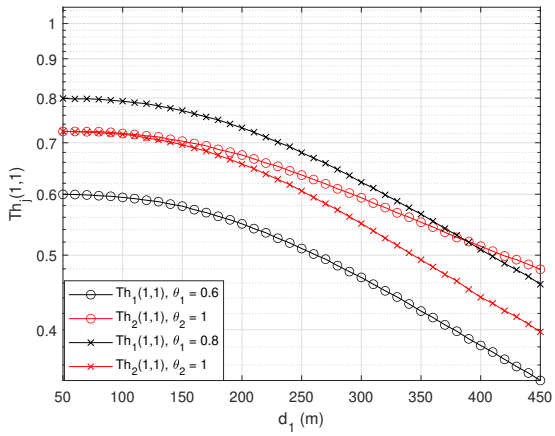

Fig. 8. $Th_j\ (1,1)$ for ACM-2: Different θ values for clusters C_1 and C_2.

For fairness performance comparison, we intend to find out suitable θ values which result in more balanced cluster throughput for different clusters. To do so, we select the system state of **s** = (1,1) meaning that there is one active device in each cluster. Depending on the employed ACM, devices may have different opportunities to access the same radio resource for their transmissions.

The obtained cluster throughput results for ACM-1 and ACM-2 are illustrated in Figs. 7 and 8 respectively. For ACM-1, we configure $\theta = 0.5$ for both clusters. For ACM-2, we configure $\theta = 0.8$ or 0.6 for cluster C1 while always keeping $\theta = 1$ for cluster C2, and $M = 2$. In both figures, the gap between the two curves that are marked with the same symbol (circle or asterisk) represents the throughput difference between two clusters. Clearly, ACM-2 achieves better access fairness since we give more opportunities to C2 in ACM-2. When employing ACM-2, a comparatively high θ for C1 will lead to both high throughput and fair access for both clusters.

V. Conclusions

In this paper, we evaluate the performance of uplink concurrent transmissions in MIMO-NOMA networks where IoT devices are grouped into multiple clusters under the coverage of a single beam. By introducing two performance parameters *cluster throughput* and *transmission latency*, we reveal the potential benefits by enabling concurrent transmissions for uplink traffic from grouped devices with different network configurations. In addition, we assess the effect of access control for the purpose of achieving fair access from different clusters. The main findings of this work are: 1) *Inter- and intra-cluster interference* plays a substantial role for the performance of cluster-based uplink concurrent transmissions in MIMO-NOMA networks; 2) Compared with transmissions without access control, the performance of transmissions (in terms of fairness) based on access control with proper parameter configurations will be significantly better; and 3) The performance of data transmissions in a MIMO-NOMA network is heavily affected by multiple factors for instance inter- and intra-device distance as well as the number of antennas at the BS. Therefore, scenario-oriented network configurations are highly recommended.

Acknowledgment

The research leading to these results has received funding from the NO Grants 2014-2021, under project contract no. 42/2021, RO-NO-2019-0499 - "A Massive MIMO Enabled IoT Platform with Networking Slicing for Beyond 5G IoV/V2X and Maritime Services (SOLID-B5G)". The work of Jorge Martinez-Bauset was supported in part by Grants PID2021-123168NB-I00 and TED2021-131387B-I00 funded by MCIN/AEI/10.13039/501100011033 and ERDF *"A way of making Europe"*, and by Grant AICO/2021/138 funded by Generalitat Valenciana.

References

[1] X. Chen, D. W. K. Ng, W. Yu, E. G. Larsson, N. Al-Dhahir, and R. Schober, "Massive access for 5G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 615–637, Mar. 2021.

[2] Z. Ding, F. Adachi, and H. V. Poor, "The application of MIMO to nonorthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 537–552, Jan. 2016.

[3] W. A. Al-Hussaibi and F. H. Ali, "Efficient user clustering, receive antenna selection, and power allocation algorithms for massive MIMO NOMA systems," *IEEE Access*, vol. 7, pp. 31865–31882, Feb. 2019.

[4] A. S. Rajasekaran, O. Maraqa, H. U. Sokun, H. Yanikomeroglu, and S. Al-Ahmadi, "User clustering in mmWave-NOMA systems with user decoding capability constraints for B5G networks," *IEEE Access*, vol. 8, pp. 209949–209963, Nov. 2020.

[5] A. Mahmoudi, B. Abolhassani, S. M. Razavizadeh, and H. H. Nguyen, "User clustering and resource allocation in hybrid NOMA-OMA systems under Nakagami-m fading," *IEEE Access*, vol. 10, pp. 38709–38728, Apr. 2022.

[6] M. Shahjalal, M. H. Rahman, M. O. Ali, B. Chung, and Y. M. Jang, "User clustering techniques for massive MIMO-NOMA enabled mmWave/THz communications in 6G," in *Proc. Int. Conf. Ubiquitous and Future Netw. (ICUFN)*, 2021, pp. 379–383.

[7] M. Alzard, S. Althunibat, and N. Zorba, "On the performance of non-orthogonal multiple access considering random waypoint mobility model," in *Proc. IEEE GLOBECOM*, Dec. 2022.

[8] Z. Ding, R. Schober, and H. V. Poor, "Unveiling the importance of SIC in NOMA systems – Part 1: State of the art and recent findings," *IEEE Commun. Lett.*, vol. 24, no. 11, pp. 2373–2377, Nov. 2020.

[9] A. Kumar, F. Y. Li, and J. Martinez-Bauset, "Revealing the benefits of rate-splitting multiple access for uplink IoT traffic," in *Proc. IEEE GLOBECOM*, Dec. 2022.

[10] Y. Qi, X. Zhang, and M. Vaezi, "Over-the-air implementation of NOMA: New experiments and future directions," *IEEE Access*, vol. 9, pp. 135828–135844, Sep. 2021.

[11] T. N. Weerasinghe, V. Casares-Giner, I. A. M. Balapuwaduge, and F. Y. Li, "Priority enabled grant-free access with dynamic slot allocation for heterogeneous mMTC traffic in 5G NR networks," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3192 - 3206, May 2021.