LATENCY MINIMIZATION FOR MEC-ENABLED MASSIVE MIMO SYSTEMS: A JOINT BIPARTITE GRAPH AND SUCCESSIVE CONVEX APPROXIMATION METHODS

Hieu V. Nguyen*, Mai T. P. Le, Vien Nguyen-Duy-Nhat, Tien V. Thai, Nguyen Van Phong

The University of Danang - University of Science and Technology, Vietnam

*Corresponding author: nvhieu@dut.udn.vn

(Received: April 25, 2025; Revised: June 08, 2025; Accepted: June 12, 2025)

DOI: 10.31130/ud-jst.2025.23(9D).567E

Abstract - This paper addresses the challenge of minimizing latency in massive multiple-input multiple-output (mMIMO) networks integrated with mobile edge computing (MEC) and non-orthogonal multiple access (NOMA). To reduce the latency for MEC-enabled mMIMO, we propose a novel NOMA-based pairing strategy, named linear bottleneck assignment problem (LBAP), that leverages channel correlation to facilitate the successive interference cancellation (SIC) technique. A two-step optimization framework is proposed: (i) a user pairing algorithm based on LBAP is executed, and then (ii) a resource allocation method utilizing the successive convex approximation (SCA) technique is applied to solve the resulting non-convex problem. Simulation results demonstrate that the proposed framework significantly outperforms not only the traditional mMIMO system but also the greedy NOMA pairing in terms of computational latency for MEC, making it a practical candidate for next-generation low-latency wireless systems.

Key words - Massive MIMO; MEC; NOMA pairing; latency optimization; successive convex approximation.

1. Introduction

The evolution toward fifth-generation (5G) and upcoming sixth-generation (6G) wireless systems has intensified the demand for ultra-low latency, high throughput, and massive device connectivity [1], [2]. Emerging applications such as autonomous driving, augmented reality, and remote surgery impose strict latency constraints, often in the order of a few milliseconds.

To meet these stringent requirements, massive MIMO (mMIMO) has been considered as a 5G architecture that overcomes limitations of conventional cellular deployments [3]. Furthermore, mobile edge computing (MEC) enables low-latency processing by relocating computation tasks closer to user equipment (UE) [4]. The integration of MEC with mMIMO allows UEs to offload computationally intensive tasks to edge nodes, effectively reducing response time and energy consumption in a wide range of application [5]–[8].

Further performance gains can be realized by incorporating non-orthogonal multiple access (NOMA) into mMIMO systems [9]. NOMA enables multiple users to share the same time-frequency resources via power-domain multiplexing, significantly improving spectral efficiency. However, system performance is heavily influenced by the user pairing strategy, particularly in large-scale deployments [10], [11].

Despite the individual benefits of mMIMO, MEC, and NOMA, their joint optimization - especially for worst-case latency minimization - remains underexplored. Existing works have primarily focused on energy-latency trade-offs or resource allocation without addressing maximum latency concerns [12], [13].

In this context, we propose a joint solution for worst-case latency minimization in MEC-enabled mMIMO networks employing linear bottleneck assignment problem (LBAP) for NOMA user pairing. Our contributions are summarized as follows:

- A LBAP NOMA pairing algorithm based on channel correlation is introduced to replace traditional random pairing, enhancing SIC effectiveness.

- A latency-aware optimization model is formulated, capturing both transmission and computation delays under power and offloading constraints.

- A two-step solution is proposed, combining NOMA pairing with a low-complexity SCA-based algorithm to solve the non-convex latency minimization problem.

- Simulation results validate the proposed method's efficiency in reducing system latency and achieving rapid convergence compared to conventional approaches.

2. Preliminaries

2.1. MEC-Enabled massive MIMO

We consider an uplink MEC-based mMIMO system, where the single-antenna EUEs are enabled to offload their computation tasks to the MEC server located at the BS. The BS equipped with M antennas is responsible for handling all the messages and executing the task offloading sent from EUEs. In this work, we propose to adopt a NOMA technique where multiple UEs share the same time-frequency resources through power-domain multiplexing. Thus, this model allows us to enhance spectral efficiency, as well as reducing interference and latency.

Without loss of generality, we let $2K$ denote the number of EUEs in a cellular mMIMO system. One EUE is paired with the other to employ the SIC technique as the NOMA principle. The UEs are randomly distributed across the coverage area and paired with the others, sharing the same resource block.

2.2. Channel Model

The wireless channel between the BS and EUE k_x, $k \in \mathcal{K} \triangleq \{1,2,\dots,K\}, x \in \{1,2\}$, is modeled as:

$$\mathbf{h}_{k_x} = \sqrt{\beta_{k_x}}\mathbf{g}_{k_x}, \tag{1}$$

where the entries of $\mathbf{g}_{k_x} \sim \mathcal{CN}(0,1)$ represents small-scale Rayleigh fading, while the large-scale fading β_{k_x} is defined as:

$$\beta_{k_x} = d_{k_x}^{-\alpha}, \tag{2}$$

where d_{k_x} is the distance between the BS and EUE k_x, α is the path loss exponent.

2.3. Notations

Table 1 summarizes the main notations used in the paper, including network parameters, decision variables, and several operators.

***Table 1.** Notations and descriptions*

Notation	Description
M	The number of antennas at the BS
K	The number of EUE pairs
$\mathbf{h}_{k_x}$	Channel vector between BS and EUE k_x
β_{k_x}	Path loss (large-scale fading) between the BS and EUE k_x
$\mathbf{g}_{k_x}$	Fading Rayleigh, $g_{m,k} \sim \mathcal{CN}(0,1)$
d_{k_x}	Distance between BS m and EUE k_x
α	Path-loss exponent
P_{k_1}, P_{k_2}	Transmit power of EUE k_1 and k_2 in pair k
$P_{k_1}^{max}, P_{k_2}^{max}$	Maximum transmit power of EUE k_1 and k_2
D_{k_x}	Task size (in bits) for EUE k_x
C_{k_x}	The number of CPU cycles per bit for EUE k_x
ρ_k	Offloading ratio of EUE k's task
f_{serv}	CPU frequency of the MEC server
R_{k_x}	Data rate achieved by EUE k_x
L_{k_x}	Total delay of EUE k_x
$L_{\max}$	Maximum tolerable delay

2.4. NOMA Pairing and Data Transmission

By applying NOMA principle, all EUEs reuse the same time-frequency resource. SIC technique is applied to the pairs of EUEs.

Consider the pair k of EUE k_1 and k_2, the message of EUE k_1 with the strong channel is decoded in priority, and then it is removed from the received signal before decoding the message of EUE k_2. Power coefficients allocated to EUE k_1 and k_2 are P_{k_1} and P_{k_2}, respectively. By applying SIC, received signals used for decoding the messages of EUE k_1 and k_2 are respectively determined as follows.

$$\begin{aligned}\mathbf{y}_{k_1} &= \mathbf{h}_{k_1}\sqrt{P_{k_1}}x_{k_1} + \mathbf{h}_{k_2}\sqrt{P_{k_2}}x_{k_2} \\ &+ \textstyle\sum_{k'=1,k'\neq k}^{K} \mathbf{h}_{k'_1}\sqrt{P_{k'_1}}x_{k'_1} + \mathbf{h}_{k'_2}\sqrt{P_{k'_2}}x_{k'_2}, \\ &+ \mathbf{n}_{\text{BS}},\end{aligned} \tag{3}$$

$$\begin{aligned}\mathbf{y}_{k_2} &= \mathbf{h}_{k_2}\sqrt{P_{k_2}}x_{k_2} + \sum_{k'=1,k'\neq k}^{K} \mathbf{h}_{k'_1}\sqrt{P_{k'_1}}x_{k'_1} \\ &+ \mathbf{h}_{k'_2}\sqrt{P_{k'_2}}x_{k'_2} + \mathbf{n}_{\text{BS}},\end{aligned} \tag{4}$$

where x_{k_1} and x_{k_2} are the desired messages transmitted from EUEs k_1 and k_2, respectively. $\mathbf{n}_{\text{BS}}$ is the additive white Gaussian noise (AWGN) vector of which each entry follows $\sim \mathcal{CN}(0,\sigma^2)$.

The Signal-to-Interference-Plus-Noise Ratios (SINRs) for decoding EUE k_1 and k_2 can be respectively formulated as

$$\text{SINR}_{k_1} = \frac{P_{k_1}|\mathbf{a}_{k_1}\mathbf{h}_{k_1}|^2}{P_{k_2}|\mathbf{a}_{k_1}\mathbf{h}_{k_2}|^2 + \text{IN}_1 + \sigma^2}, \tag{5}$$

$$\text{SINR}_{k_2} = \frac{P_{k_2}|\mathbf{a}_{k_2}\mathbf{h}_{k_2}|^2}{\text{IN}_2 + \sigma^2}, \tag{6}$$

where $\mathbf{a}_{k_1} = \mathbf{h}_{k_1}^H$ and $\mathbf{a}_{k_2} = \mathbf{h}_{k_2}^H$ are the receiver vectors, and the interference IN_1 and IN_2 are respectively given as

$$\text{IN}_1 \triangleq \textstyle\sum_{k'\neq k}^{K} \left(P_{k'_1}\left|\mathbf{a}_{k_1}\mathbf{h}_{k'_1}\right|^2 + P_{k'_2}\left|\mathbf{a}_{k_1}\mathbf{h}_{k'_2}\right|^2\right), \tag{7}$$

$$\text{IN}_1 \triangleq \textstyle\sum_{k'\neq k}^{K} \left(P_{k'_1}\left|\mathbf{a}_{k_2}\mathbf{h}_{k'_1}\right|^2 + P_{k'_2}\left|\mathbf{a}_{k_2}\mathbf{h}_{k'_2}\right|^2\right). \tag{8}$$

3. Latency Optimization Problem

3.1. Data Rate and Latency

From (5) and (6), the achievable rate of EUE k_x is computed by

$$R_{k_x} = \text{B}\log_2\left(1 + \text{SINR}_{k_x}\right), \tag{9}$$

where B is the system bandwidth. Then, the latency of EUE k_x includes transmission delay, computing delay, and queue delay. Since the queue delay is considered as a given value, we omit this term without any effect on the problem. Therefore, the total latency of EUE pair k can be expressed as

$$\begin{aligned}L_k &= \max_x\left\{L_{\text{trans},k_x} + L_{\text{comp},k_x}\right\} \\ &= \max_x\left\{\frac{\rho_{k_x}D_{k_x}}{R_{k_x}} + \frac{\rho_{k_x}D_{k_x}C_{k_x}}{f_{\text{serv}}}\right\},\end{aligned} \tag{10}$$

where L_{trans,k_x} and L_{comp,k_x} are the transmission delay and computing delay, respectively. The variable $\rho_{k_x} \in [0,1]$ is the offloading ratio. D_{k_x} is the size (in bits) of the offloading task, while C_{k_x} and f_{serv} represent the resource and frequency to process the data at the MEC server.

3.2. Latency Minimization Problem Formulation

We formulate the latency optimization problem for the mMIMO system with MEC and NOMA. Our objective is to minimize the maximum latency experienced by any user in the network, focusing on the worst-case scenario. Mathematically, the problem can be formulated as:

$$\min_{\{P_{k_x},\rho_{k_x}\}} \quad \max_{k=1,\dots,K} L_k \tag{11a}$$

$$\text{s.t.} \qquad 0 \leq P_{k_1} \leq P_{k_1}^{max}, \forall k \tag{11b}$$

$$0 \le P_{k_2} \le P_{k_2}^{max}, \forall k \quad (11c)$$

$$\mathbf{0} \le \boldsymbol{\rho} \le \mathbf{1} \quad (11d)$$

Constraints (11b) and (11c) ensure that the transmit power of each EUE does not exceed its maximum allowable limit. Constraint (11d) enforces that the offloading ratio stays within the valid range of 0 to 1. The optimization problem presented above is challenging due to its non-convex nature. The function R_k in the latency expression depends non-linearly on the transmit power P_k and is affected by cross-interference between EUEs in the NOMA system. Consequently, the term $\frac{\rho_{k_x} D_{k_x}}{R_{k_x}}$ becomes non-convex with respect to the decision variables (P_{k_x}, ρ_{k_x}). Furthermore, the min-max structure of the objective function adds another layer of complexity. To address the min-max structure, we introduce an auxiliary variable τ and transform problem (11) into:

$$\min_{\{P_k, \rho_k\}, \tau} \quad \tau \quad (12a)$$

$$\text{s.t.} \quad L_k \le \tau, \forall k, \quad (12b)$$

$$(11b) - (11d). \quad (12c)$$

However, the constraint $L_k \le \tau$ remains non-convex due to the non-convex function L_k. It is further challenging that the strategy of EUE pairing has a large effect on the latency minimization problem. Therefore, we propose an efficient pairing scheme before solving problem (12).

4. Two-Step Algorithm for NOMA pairing and resource allocation

To improve the NOMA-based resource allocation for MEC-enabled mMIMO network, we propose the two-step algorithm: (1) we propose the linear bottleneck assignment problem (LBAP) for NOMA pairing; and then (2) we propose a fast transformation to apply SCA framework for the resource allocation to achieve the minimal latency as solving problem (12).

4.1. Proposed Linear Bottleneck Assignment Problem (LBAP) for NOMA Pairing Algorithm

We propose an LBAP-based NOMA pairing for which SIC is applied efficiently. In mMIMO systems, the BS typically employs the maximum ratio combining (MRC) scheme to reduce the complexity for decoding the messages from EUEs. However, this approach may lead to substantial interference between UEs with highly correlated channels. To address this issue, we first define two sets of K EUEs: K EUEs with the better channel conditions in the first set $\mathcal{V}_1$, and the other K EUEs in the second set $\mathcal{V}_2$. Moreover, we define a new set containing all the channel correlation values between two EUEs with one EUE from $\mathcal{V}_1$ and the other from $\mathcal{V}_2$. The channel correlation between UEs $i \in \mathcal{V}_1$ and $j \in \mathcal{V}_2$ is calculated as

$$C_{i,j} = \frac{|\mathbf{h}_i^H \mathbf{h}_j|^2}{\|\mathbf{h}_i\|^2 \|\mathbf{h}_j\|^2}. \quad (13)$$

This metric ranges from 0 (orthogonal channels) to 1 (perfectly correlated channels), providing a measure of spatial correlation between channel vectors. By letting $\mathcal{E} = \{C_{i,j}\}$, the LBAP-based NOMA pairing algorithm is executed using the combination of the binary search and maximum matching on the bipartite graph $\mathcal{G} = (\mathcal{V}_1 \cup \mathcal{V}_2, \mathcal{E})$.

4.2. Latency Optimization using SCA

Following the identification of NOMA pairs, we obtain problem (9). To tackle the non-convexity, we apply the successive convex approximation (SCA) method. First, we expand the constraint $L_k \le \tau$ as:

$$\frac{\rho_{k_x} D_{k_x}}{R_{k_x}} + \frac{\rho_{k_x} D_{k_x} C_{k_x}}{f_{\text{serv}}} \le \tau. \forall k \in \mathcal{K}, x \in \{1,2\}. \quad (14)$$

By introducing the variable substitution $\bar{\tau}_{k_x} = \frac{1}{R_{k_x}}$, we rewrite this as:

$$D_{k_x} \rho_{k_x} \bar{\tau}_{k_x} + \frac{\rho_{k_x} D_{k_x} C_{k_x}}{f_{\text{serv}}} \le \tau, \forall k, x. \quad (15)$$

The product $\rho_{k_x} \bar{\tau}_{k_x}$ remains non-convex. To address this, we introduce the upper bound of the product function as

$$\rho_{k_x} \bar{\tau}_{k_x} \le \frac{\rho_{k_x}^{(n)}}{2\bar{\tau}_{k_x}^{(n)}} \bar{\tau}_{k_x}^2 + \frac{\bar{\tau}_{k_x}^{(n)}}{2\rho_{k_x}^{(n)}} \rho_{k_x}^2 := F^{(n)}(\rho_{k_x} \bar{\tau}_{k_x}), \forall k, x, \quad (16)$$

where $\bar{\rho}_{k_x}^{(n)}$ represents the value of $\bar{\rho}_{k_x}$ at the n-th iteration.

The transformed convex problem becomes:

$$\min_{P_k, \rho_k, \tau, \bar{\tau}} \quad \tau \quad (17a)$$

$$D_{k_x} F^{(n)}(\rho_{k_x} \bar{\tau}_{k_x}) + \frac{\rho_{k_x} D_{k_x} C_{k_x}}{f_{\text{serv}}} \le \tau, \forall k, x \quad (17b)$$

$$(11b) - (11d). \quad (17c)$$

It can be seen that problem (17) is convex and can be efficiently solved using standard convex optimization tools.

The procedure of SCA framework for solving (11) is summarized in **Algorithm 1**.

Algorithm 1: SCA for Latency Minimization in (11)

1: **Input:** EUE pairs and system parameters, $\epsilon = 10^{-3}$.

2: **Output:** Optimal power allocation $\boldsymbol{P}^*$ and offloading ratios $\boldsymbol{\rho}^*$.

Initialization:

3: Set iteration counter $n \leftarrow 0$.

4: Generate feasible starting point: $(P^{(0)}, \rho^{(0)}, \tau^{(0)}, \bar{\tau}^{(0)})$.

5: **Repeat**

6: Linearize non-convex constraints around current point to form convex subproblem.

7: Solve convex subproblem (17) to obtain the solution $(P^{sol}, \rho^{sol}, \tau^{sol}, \bar{\tau}^{sol})$.

8: Update the solution: $(P^{(n+1)}, \rho^{(n+1)}, \tau^{(n+1)}, \bar{\tau}^{(n+1)}) \leftarrow (P^{sol}, \rho^{sol}, \tau^{sol}, \bar{\tau}^{sol})$.

9: Compute new objective value $\tau^{(n+1)}$.

10: $n \leftarrow n + 1$.

11: **Until** $|\tau^{(n)} - \tau^{(n-1)}| < \epsilon$

12: **Return** $(P^*, \rho^*) = (P^{(n)}, \rho^{(n)})$.

4.3. Complexity and Convergence Analysis

The LBAP-based NOMA pairing algorithm has a complexity of $O(K^3 logK^2)$, as it requires calculating a channel correlation matrix of size $2K \times 2K$ and finding the maximum matching.

For the SCA algorithm, each iteration involves solving a convex problem of size K, which typically costs $O\left((v^2c^{2.5} + c^{3.5})\log(1/\epsilon)\right)$, where $v = 4K + 1$ và $c = 6K$ are the number of variables and constraints, respectively, and ϵ is the convergence threshold.

5. Performance Analysis

5.1. Simulation setup

The performance of the proposed solution is evaluated via Monte Carlo simulations. The system parameters are configured in accordance with practical 5G/6G network deployments, as summarized in Table 2. To validate the effectiveness of the proposed scheme, we compare the proposed LBAP-based user pairing NOMA mMIMO architecture with other two benchmark strategies:

- Greedy-based NOMA pairing mMIMO where user pairing is computed by the famous greedy scheme (the highest channel correlation is paired at first), and

- Traditional mMIMO without NOMA (the MRC receiver is applied to directly decode the messages).

Table 2. *Simulation Parameter Configuration*

Parameter	Value
Simulation area	1 km × 1 km
Number of BSs (M)	64
Number of EUEs ($2K$)	[10, 20, 30]
Bandwidth	20 MHz
Noise power spectral density	−174 dBm/Hz
Maximum transmit power ($P_{k_x}^{max}$)	10 dBm
MEC server CPU frequency(f_{serv})	5 GHz
Path-loss exponent (α)	3.7
Task size, D_{k_x}	1Kb

5.2. Convergence analysis of SCA algorithm

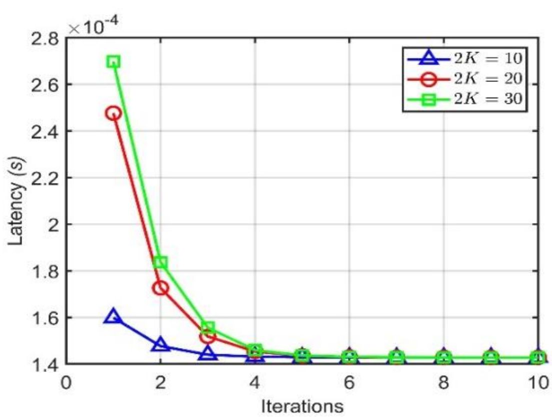

Figure 1. *Convergence speed comparison of the SCA algorithm under three different number of users*

Figure 1 illustrates the convergence behavior of the proposed SCA-based optimization algorithm under three different different number of users. At $2K = 10$, the algorithm starts with a delay of approximately $2.6 \times 10^{-4} s$ and converges to $1.43 \times 10^{-4} s$ after 6 iterations. Even when the initial delays can be different with $2K = \{20, 30\}$ decrease to $1.99 \times 10^{-4} s$ and $1.3 \times 10^{-4} s$, respectively, and the final values still converge at around $1.43 \times 10^{-4} s$. Notably, all configurations exhibit fast convergence, requiring only 5–6 iterations to reach a steady state, demonstrating the effectiveness of the proposed algorithm.

5.3. System latency assessment

5.3.1. Delay distribution analysis

The cumulative distribution function (CDF) of the system delay illustrated in Figure 2 demonstrates the superior performance of the proposed LBAP NOMA-based mMIMO architecture compared to other schemes. At the 80th percentile user threshold, the proposed approach achieves a delay of $1.56 \times 10^{-4} s$, which is lower than that of the Greedy NOMA pairing and traditional mMIMO. These results validate the effectiveness of the LBAP NOMA-based pairing algorithm in improving system delay performance.

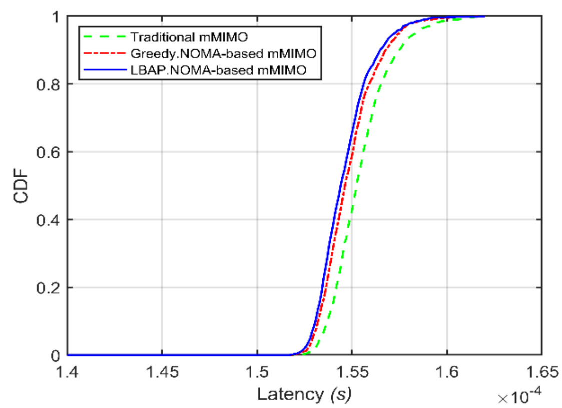

Figure 2. *Comparison of delay CDF between network architectures*

6. Conclusion

This paper has presented a novel joint optimization framework that integrates LBAP NOMA-based user pairing with the successive convex approximation (SCA) method to minimize the latency in mMIMO networks with MEC support. By leveraging channel correlation for user pairing and adopting an efficient two-step optimization strategy, the proposed solution achieves notable improvements in both latency reduction and scalability compared to traditional mMIMO and greedy pairing approaches.

Simulation results confirm that the proposed scheme is particularly effective at high MEC processing frequencies, achieving a minimum latency of 1.43×10^{-4} seconds at 5 GHz. Even as the user population increases from 10 to 30, the system maintains performance robustness, highlighting its scalability potential.

Nonetheless, the effectiveness of the proposed approach heavily depends on the availability of edge processing resources, highlighting the importance of robust and scalable MEC infrastructure. Moreover, practical deployment may encounter challenges such as multi-carrier environments and strong channel correlations among users, which could impact performance.

Future work will investigate advanced user clustering strategies, including learning-based approaches for dynamic and context-aware optimization. Additional research directions include developing adaptive mechanisms to effectively handle user mobility and extending the framework to jointly optimize multiple quality-of-service (QoS) requirements. These efforts aim to enhance the practicality and robustness of the proposed scheme for real-world 5G and next-generation 6G wireless networks.

Acknowledgment: This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.04-2023.40.

REFERENCES

[1] W. Saad, M. Bennis, and M. Chen, 'A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems', *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, 2020.

[2] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, 'The Road Towards 6G: A Comprehensive Survey', *IEEE Open J. Commun. Soc.*, vol. 2, pp. 334–366, 2021.

[3] S. Willhammar, H. Iimori, J. Vieira, L. Sundström, F. Tufvesson and E. G. Larsson, "Achieving Distributed MIMO Performance with Repeater-Assisted Cellular Massive MIMO", in *IEEE Communications Magazine*, vol. 63, no. 3, pp. 114-119, March 2025.

[4] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, 'A Survey on Mobile Edge Computing: The Communication Perspective', *IEEE Commun. Surv. Tutor.*, vol. 19, no. 4, pp. 2322–2358, 2017.

[5] S. S. Yılmaz, B. Özbek and R. Mumtaz, "Delay Minimization for Massive MIMO Based Cooperative Mobile Edge Computing System With Secure Offloading", in *IEEE Open Journal of Vehicular Technology*, vol. 4, pp. 149-161, 2023.

[6] M. Zeng, W. Hao, O. A. Dobre and H. V. Poor, "Delay Minimization for Massive MIMO Assisted Mobile Edge Computing", in *IEEE Transactions on Vehicular Technology*, vol. 69, no. 6, pp. 6788-6792, June 2020.

[7] E. T. Michailidis, N. I. Miridakis, A. Michalas, E. Skondras, D. J. Vergados and D. D. Vergados, "Energy Optimization in Massive MIMO UAV-Aided MEC-Enabled Vehicular Networks", in *IEEE Access*, vol. 9, pp. 117388-117403, 2021.

[8] R. Malik and M. Vu, "Energy-Efficient Computation Offloading in Delay-Constrained Massive MIMO Enabled Edge Network Using Data Partitioning", in *IEEE Transactions on Wireless Communications*, vol. 19, no. 10, pp. 6977-6991, Oct. 2020.

[9] A. Akbar, S. Jangsher, and F. A. Bhatti, 'NOMA and 5G emerging technologies: A survey on issues and solution techniques', *Comput. Netw.*, vol. 190, p. 107950, 2021.

[10] X.-T. Dang, M. T. P. Le, H. V Nguyen, and O.-S. Shin, 'Optimal User Pairing for NOMA-assisted Cell-Free Massive MIMO System', in *2022 IEEE Int. Conf. Commun. Electron. (ICCE)*, 2022, pp. 7–12.

[11] X.-T. Dang, M. T. P. Le, H. V Nguyen, S. Chatzinotas, and O.-S. Shin, 'Optimal User Pairing Approach for NOMA-based Cell-Free Massive MIMO Systems', *IEEE Trans. Veh. Technol.*, vol. 72, no. 4, pp. 4751–4765, 2023.

[12] H. V Nguyen, M. T. P. Le, T. D. Ho, P. V. Tuan, and H. Nguyen-Le, 'Joint Latency Minimization and Power Allocation for MEC-Enabled MU-MISO Networks', in *2024 10th Int. Conf. Commun. Electron. (ICCE)*, 2024, pp. 753–757.

[13] T. V. Thai, M. T. P. Le, H. V. Nguyen, and O.-S. Shin, 'NOMA-Aided Cell-Free Massive MIMO with MEC: A Trade-Off Between Latency and Energy Consumption', in *2024 IEEE Int. Conf. Consum. Electron.-Asia (ICCE-Asia)*, Danang: IEEE, Nov. 2024, pp. 1–5.