IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 20 June 2023, accepted 18 August 2023, date of publication 29 August 2023, date of current version 6 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3309628

RESEARCH ARTICLE

Joint Task Assignment, Power Allocation and Node Grouping for Cooperative Computing in NOMA-mmWave Mobile Edge Computing

AZADEH KHAZALI[1], ARASH BOZORGCHENANI[2], (Member, IEEE), DANIELE TARCHI[3], (Senior Member, IEEE), MAHROKH G. SHAYESTEH[1,4], (Senior Member, IEEE), AND HASHEM KALBKHANI[5]

[1]Department of Electrical and Computer Engineering, Urmia University, Urmia 57179-44514, Iran
[2]School of Computing, University of Leeds, LS2 9JT Leeds, U.K.
[3]Department of Electrical, Electronic and Information Engineering "Guglielmo Marconi," University of Bologna, 40126 Bologna, Italy
[4]Wireless Research Laboratory, ACRI, Electrical Engineering Department, Sharif University of Technology, Tehran 14588-89694, Iran
[5]Faculty of Electrical Engineering, Urmia University of Technology, Urmia 5716617165, Iran

Corresponding author: Daniele Tarchi (daniele.tarchi@unibo.it)

The work of Daniele Tarchi was supported in part by the ECOSISTER Project funded under the National Recovery and Resilience Plan (NRRP), Mission 04 Component 2 Investment 1.5—NextGenerationEU, Call for tender n. 3277 (30/12/2021) under Award 0001052 (23/06/2022).

ABSTRACT In this paper, we investigate the cooperation of idle computation resources of nearby mobile devices in mobile edge computing (MEC) systems, in which each mobile device jointly offloads computation tasks to a MEC node and a nearby mobile device by employing non-orthogonal multiple access (NOMA) in a millimeter-wave (mmWave) heterogeneous network. In this setup, the nearby device acts as a helper by performing local computation and offloading data simultaneously to the MEC system. We formulate an optimization problem for joint taSk assignmenT, poweR allOcation and Node Grouping (STRONG) aiming to minimize the energy consumption of devices (i.e., user and helper devices). To tackle this problem, we present a two-step solution. First, we adopt a low complexity search-based algorithm for both helper and MEC server selection. Next, considering the non-convex nature of the energy minimization problem, we develop algorithms that provide sub-optimal solutions for power allocation to the helper and MEC server, as well as the offloading task ratios between them. Numerical results are provided to validate the effectiveness of our proposed algorithms. The results not only validate the efficiency of our approach but also demonstrate the superiority of our cooperative NOMA-based MEC scenario compared to methods without cooperation and other cooperation-based scenarios.

INDEX TERMS Mobile edge computing, distributed computing, NOMA, millimeter wave communication, energy efficiency.

I. INTRODUCTION

In recent years, with the explosive emergence of mobile computing services, e.g., virtual reality (VR) and augmented reality (AR), limitations on battery and computation capacities bring new challenges for mobile devices (MDs) towards sixth generation (6G) networks. One solution to overcome these challenges is mobile-edge computing, or as it has been renamed by ETSI, multi-access edge computing (MEC). Among several techniques enabled through MEC, computation offloading offers an ultra-low latency environment with real-time access to network resources, and enables MDs to extend their battery lifetime, by offloading various computational tasks to edge servers deployed at the small-cells, macro-cells, and Wi-Fi access points (APs) [1]. Therefore, MEC offloading enhances both

The associate editor coordinating the review of this manuscript and approving it for publication was Adamu Murtala Zungeru.

quality-of-experience (QoE) and quality-of-service (QoS) of MDs [2].

However, the capacity of resources of MEC servers is not infinite and, as a result, MEC-based systems may face challenges dealing with a high volume of requests from MDs. Moreover, the interaction between MDs and the MEC server may be negatively affected by long distances, leading to a decrease in channel gain during data exchange. One solution for these challenges is the implementation of cooperative MEC. By leveraging nearby users as helpers, the computational and radio resources of these neighboring MDs can be utilized to assist far MDs in completing their computation tasks. Thus, such cooperative MEC scenario not only overcomes the limitations imposed by the MEC server capacity and long-distance interactions, but also balances the resources and workloads of MDs [3].

As a consequence of the development of sensor networks and Internet of Things (IoT), machine-to-machine (M2M) communications are expected to grow to billions in the next few decades. Moreover, the data traffic demand of MDs is expected to grow to 10 times. Thus, to support these demands, the capacity and data rate of current wireless communication networks should be enhanced [4]. To this end, millimeter-wave (mmWave) communications and non-orthogonal multiple access (NOMA) attracted considerable attention both in industry and academia [4].

The integration of mmWave communications and MEC provides high capacity for nearby MDs and low latency services, particularly those with high data-consuming applications [5]. To further increase data rates and provide short-range communications, network densification through mmWave small cells (SCs) have gained great attention. SCs benefit from the unique features of mmWave communications including directionality and large bandwidth, resulting in a considerable reduction of co-channel interference [6].

On the other hand, NOMA technology can effectively enhance spectral efficiency (SE), connectivity, and reduce access latency [7], [8]. In the context of MEC, a NOMA-based MEC network has been proposed by using the superposition coding in power domain to support multiple MDs within the same resource block [9]. Therefore, leveraging the advantages of high bandwidth and data rate, both mmWave SCs and NOMA techniques are efficient methods for MEC systems [10], [11]. Moreover, the rapid battery life depletion remains a significant challenge in modern networks. To enhance the computation performance and reduce energy consumption of mobile users, it is of great importance to investigate power allocation strategies for MEC networks [12], [13].

Most of the current works focus on simple networks, where their MEC models do not consider cooperative approaches, and usually part of task computation is executed locally at MDs and the rest is offloaded to the MEC servers [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25]. Under the large number of users that are supported by 5G and its beyond, the possibility of collaboratively exploiting idle-state users with unused computation resources acting as helpers can be considered to improve the performance of MEC. Inspired by this fact, the helper-assisted cooperative NOMA-MEC has been investigated. In [3] and [26], a basic three-node system consisting of one MD, one helper, and one AP is proposed, where the helper acts as a relay for offloading a part of MD's task to the AP. In [27], a novel multi-helper cooperative NOMA for MEC was proposed and the total offloading data was optimized.

However, when a large number of helpers and MEC servers in ultra-dense networks participate in cooperative NOMA-MEC, the performance of task offloading may degrade by increasing the complexity and error propagation of successive interference cancellation (SIC)-based detection. Therefore, to further enhance the performance of NOMA-MEC systems by utilizing the spatial diversity offered by cooperative communication, the development of a node grouping algorithm becomes essential. The focus of this paper is to propose a node grouping algorithm and address the energy minimization challenges in a cooperative mmWave-NOMA MEC network. The contribution of this research is discussed in Section A.

A. CONTRIBUTIONS

This paper addresses the cooperative edge computation offloading in mmWave-NOMA heterogeneous network. In this network, the helper can offload part of the task to the MEC server while processing the remaining portion locally. Although cooperation reduces the burden on the MEC server, it leads to higher energy consumption of helper. Thus, we formulate energy consumption minimization of the network subject to various constraints including 1) the sum of the offloading task ratios, 2) the offloading times for both user and the helper, 3) the delays in completing the task locally at the user, helper and the MEC server, 4) the total energy consumed by the MDs for offloading and locally computing the task, and 5) the total energy consumed by the helper for offloading and local computing. Further, taSk assignmenT, poweR allOcation and Node Grouping (STRONG) are jointly optimized. As mentioned in Table 1, the proposed methods are different from those developed in [3], [14], [15], [16], [17], [18], [19], [26], [27], and [28], in the aspect of system model, multiple access method, frequency type or investigated objectives. The main contributions of this paper are summarized as follows:

1. We investigate a MEC scheme that enables cooperative NOMA-mmWave transmissions, in which the helper can offload part of the task to the MEC server while processing the remaining portion locally.
2. We propose a grouping scheme that facilitates the selection of the helper and MEC server for cooperative NOMA-mmWave transmission.
3. Task assignment and power allocation of nodes are jointly designed aiming at improving the energy minimization within the MEC scheme. This problem is

TABLE 1. Comparison between the proposed cooperative mmWave-NOMA heterogeneous network and the most related works.

References	MEC system model	Multiple access methods	Frequency type	Investigated Objectives
Our work	cooperative	NOMA	mmWave	Node grouping, Energy minimization, successful computation probability
[14-19]	Non-cooperative	NOMA	microwave	Energy minimization
[26]	cooperative	OMA	microwave	Energy minimization
[3]	cooperative	NOMA	microwave	energy consumption minimization and offloading data maximization
[27]	cooperative	NOMA	microwave	Offloading data
[28]	cooperative	NOMA	microwave	successful computation probability

proven to be non-convex and, later, is converted to a convex problem for which the sub-optimal task assignment and minimum power allocation values are obtained.

4. We derive a closed-form expression for the successful computation probability. We also discover that the locations of MEC server and helper significantly impact the successful computation probability. Motivated by this, we derive the successful computation probability of the proposed network.
5. Finally, through simulation results, we demonstrate that the proposed system displays superior performance in comparison to other benchmark schemes.

The rest of this paper is organized as follows. In Section II, we review the state of the art related to works considering NOMA and mmWave in MEC system. In Section III, we describe the system model. In Section IV, we derive the successful computation probability. Section V introduces the proposed solution for energy consumption minimization. In Section VI, we present performance evaluation results and, finally, Section VI concludes the paper.

II. RELATED WORKS

The fundamental concept of computation offloading in MEC system involves transferring the latency-critical and computation-intensive tasks of MDs to the MEC server [29]. The task offloading is operated in either partial or binary modes [30]. In the binary offloading mode, the computation tasks are entirely offloaded to the MEC server, whereas in the partial offloading mode, the computation tasks can be shared among several MEC servers and the offloader itself [31].

The deployment of MEC servers near the APs provides cloud-like computing services for the MDs [32]. However, supporting a massive number of requests from MDs becomes complex for APs due to their limited computational capacities. Therefore, to alleviate the workload of APs, cooperative computing paradigms can be considered by exploiting the computation capability of several neighboring MEC servers. For example, in [33], an algorithm was proposed to encourage the MDs to offer their unused resources. In [34], the authors considered a co-computing scenario in which an MD offloads the computational data to a helper. The work in [35] introduced a task offloading scheme based on D2D collaboration. It was shown in [36] that collaborative computation offloading among MDs in an energy harvesting scenario can prolong the network lifetime. Machine learning approaches have also been extensively proposed in the literature for computation offloading in MEC. An auction-based approach was presented in [37] to incentivize MEC servers to participate in resource sharing with MDs where the allocation of computation resources to MDs is performed with a pair of deep neural networks. In [38], both on-line and off-policy solutions based on Bandit theory were proposed for allocating the computation resources to the MDs in a highly dynamic vehicular environment.

Recently, several studies have been devoted to investigating the benefits of NOMA in MEC systems, especially for enhancing energy efficiency (EE). In [14], authors introduced an optimization framework that jointly optimizes the transmit power, user clustering, and computing for minimizing energy consumption of a NOMA-based MEC system. A similar problem was investigated in [15] by considering the offloading tasks and power levels of each user. The work in [16] considered NOMA-assisted MEC system, where time allocation and power levels were optimized to minimize the energy consumption of computation offloading. A joint computational and radio resource allocation problem for NOMA-assisted MEC system in heterogeneous networks was investigated in [17]. The authors in [18] considered NOMA for task offloading and result downloading in the MEC system, where transmit power, transmission time allocation, and task offloading partitions were optimized to minimize the total energy consumption. An energy-efficient multitask multi-access scheme in NOMA-based MEC system was studied in [19].

Besides [14], [15], [16], [17], [18], [19], several other works [20], [21], [22], [23], [24], [25] focused on the latency minimization in NOMA-based MEC system. The minimization of overall delay for completing the computation requirements of MDs by jointly optimizing users' offloaded workloads and NOMA transmission time was considered in [20], while in [21] the authors aimed at minimizing the sum of downloading, offloading, and execution times. The study in [22] proposed the minimization of maximum task execution time across all devices for uplink NOMA-based MEC system by jointly considering SIC ordering and computation resource allocation. In [23], an optimal offloading policy was presented to minimize the mean latency. The authors in [24] proposed to maximize the probability of guaranteeing

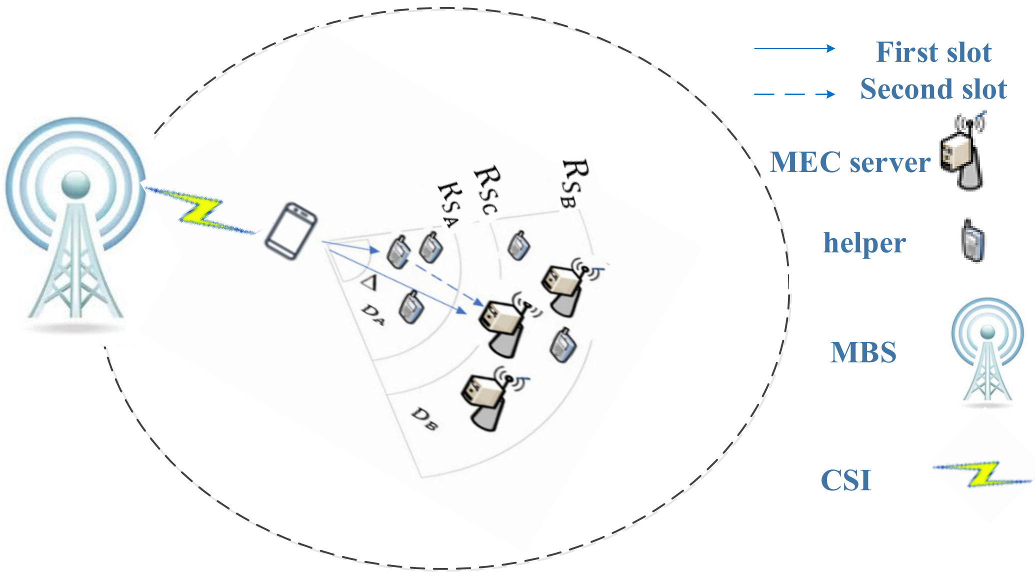


FIGURE 1. Proposed NOMA-mmWave based MEC heterogeneous network.

the latency requirements. In [25], the trade-off between the communication performance (i.e., network coverage) and latency was presented. Energy consumption and task processing delay can sometimes be seen as two competing objectives; hence, there exists a trade-off between them [1]. Some works, such as [39], addressed this trade-off and studied energy consumption optimization and task completion delay simultaneously.

Unlike the existing works, [14], [15], [16], [17], [18], [19], the promising gains of mmWave have motivated us on integrating MEC and mmWave techniques. Given the propagation characteristics of mmWaves, cooperative NOMA communication emerges as an efficient approach to increase the SE of the network. Although the authors in [3] and [26] combined MEC with cooperative communication, the system models considered in these works differ from ours. For instance, [26] focuses on OMA-based system model, while [3] introduces cooperative computing within a NOMA-based MEC system, which consists of a user, a helper, and a MEC server deployed at an AP. However, there are notable limitations in these prior works. In [3], although the helper can receive and compute the user's task in addition to acting as a relay, the selection of helper was not discussed, and the energy consumption and delay of the MEC server were not considered. Also, in [3], mmWave band, beamforming, and MIMO antennas were not incorporated. Our approach addresses these limitations by integrating MEC and cooperative NOMA communication in the context of mmWave networks. By considering helper selection, energy efficiency, delay constraints, and leveraging mmWave features, we aim to provide an involved solution compared to the previous works. Also, the proposed techniques are distinctively different from those developed in [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], and [25], as summarized in Table 1. To the best of our knowledge, the systematic design of cooperative mmWave-NOMA MEC systems has not been yet investigated. Moreover, node grouping as a key aspect of achieving performance in NOMA, has not been considered in the previous studies [3], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25]. Although [28] explores helper selection, unlike our scenario, they assume a fixed MEC server. while the MEC server has a more powerful computing capacity and influences the successful computation probability. Also, the helper selection method in [28] does not optimize the successful computation probability and relies on the conventional channel gain method.

III. SYSTEM MODEL

We model partial offloading in a NOMA-mmWave based MEC heterogeneous network, consisting of multiple MEC servers, one macrocell base station (MBS), and multiple MDs, where one is considered as the reference user and the others are helpers. We suppose that the reference user generates a random beamforming vector, as shown in Fig. 1; only the helpers that fall into the region D_A with radius R_{S_A} and a central angle 2Δ are considered. In addition, only the MEC servers located within the region D_B with a maximum radius of R_{S_B}, a minimum radius of R_{S_c}, and a central angle of2Δ, can be scheduled. These assumptions ensure the maximal angle difference Δ between the selected helper or MEC server and their associated beams. As depicted in Fig. 1, $R_{S_A} < R_{S_c} < R_{S_B}$. It is assumed that all nodes operate in half-duplex mode and are equipped with multiple antennas. Moreover, the MBS utilizes channel state information (CSI) from all nodes to determine parameters such as transmit powers and offloading portions. Once determined, these parameters are communicated to the user, helpers and the MEC server. As depicted in Fig. 2, we assume that the user and helpers'

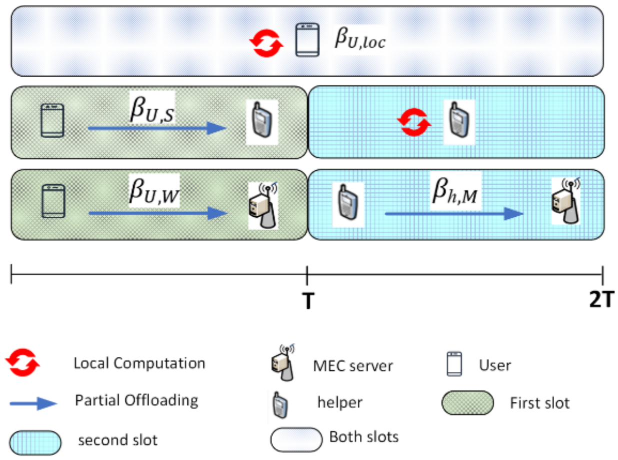


FIGURE 2. **Execution time for the proposed NOMA-mmWave based partial offloading protocol.**

computation tasks should be executed within a given time $2T$ and T, respectively, while local computing and offloading can be executed in parallel.

The two time slots used to perform the cooperative NOMA-based partial offloading are shown in Fig. 2. Denote $L_U\beta_{U,h}$ and $L_U\beta_{U,W}$ as the portions of the user's task, with a size of L_U bits, offloaded to the selected helper and MEC server, respectively. We define $\beta_{U,h} = \beta_{U,S} + \beta_{h,M}$ where $\beta_{U,S}$ is the portion of the user's task computed by the selected helper and $\beta_{h,M}$ is the portion of the user's task offloaded from helper to MEC server.

In the first time slot, having duration T, the user's task is offloaded in the uplink direction to both the selected helper and MEC server, i.e., $L_U(\beta_{U,S} + \beta_{h,M})$ bits, to the helper and the other part, i.e., $L_U\beta_{U,W}$ bits, to the MEC server for remote executions. The remaining portion of the task $\left(1 - \beta_{U,S} - \beta_{U,W} - \beta_{h,M}\right)L_U$, denoted as $\beta_{U,loc}$, is locally executed within duration $2T$. In the second time slot, the selected helper partially offloads the task $(L_U\beta_{h,M})$ to the MEC system, while simultaneously executing the remaining portion $L_U(\beta_{U,S})$ locally. We assume that the local computation of user can be done in both time slots.

A. OFFLOADING AWARE MMWAVE-NOMA TRANSMISSION

During the task offloading phase, at the first time slot, the user transmits its task with L_U bits exploiting linear superposition of data to the selected helper and MEC server simultaneously as:

$$x = \sqrt{P_S}x_S + \sqrt{P_W}x_W \tag{1}$$

where x_S and x_W are the unit power message signals of the strong (i.e., highest channel gain node) and weak nodes (i.e., lowest channel gain node), corresponding to the selected helper and MEC server, respectively, while p_S and p_W denote the transmission power levels allocated to the strong and weak nodes, respectively. Hence, the received signals at the kth strong and weak node can be written as:

$$y_{k,A} = g_{k,A}x + n_{k,A} \tag{2}$$

where $A \in \{S, W\}$, S and W represent the strong and weak nodes, respectively, and $n_{k,A}$ is the zero-mean additive Gaussian noise with the variance N_0. When $y_{k,A}$ contributes to the strong nodes (i.e., $A = S$), k defines the index of helpers, otherwise (i.e., $A = W$), it corresponds to the index of MEC servers. By applying SIC at the selected helper and MEC server, the desired signal of each of them is detected. Moreover, $g_{k,A}$ denotes the mmWave channel gain from the user A to the helper or MEC server, k, which consists of one line-of-sight (LOS) path and several non-line-of-sight (NLOS) paths. To this end, the mmWave channel gain can be expressed as [40]:

$$g_k = \sqrt{N}\frac{\rho_{k,L}w\left(\theta_{k,L}\right)}{\sqrt{1+d_k^{\omega L}}} + \sqrt{N}\sum\nolimits_{nl=1}^{NL}\frac{\rho_{nl,NL}w\left(\theta_{nl,NL}\right)}{\sqrt{1+d_k^{\omega NL}}} \tag{3}$$

where $\theta_{k,L}$ and $\theta_{nl,NL}$ denote the normalized directions of receiver for LOS and NLOS paths, respectively, $\rho_{k,L}$ and $\rho_{nl,NL}$ represent the complex channel gain of receiver for LOS and NLOS paths, respectively, and ω_L and ω_{NL} denote the path loss exponents for LOS and NLOS paths, respectively. Also, d_k shows the distance between transceivers, NL is the number of NLOS paths and N denotes the number of transmit antennas of user, while we assume an array steering vector $\omega(\theta)$ shown as [40]:

$$\omega(\theta) = \frac{1}{\sqrt{N}}\left[1, e^{-j\pi\theta}, \ldots, e^{-j\pi(N-1)\theta}\right]^T \tag{4}$$

As described in [40] and [41], the effect of LOS path is dominant in mmWave communication. Thus, the mmWave channel model can be simplified as:

$$g_k = \sqrt{N}\frac{\rho_{k,L}w\left(\theta_{k,L}\right)}{\sqrt{1+d_k^{\omega L}}} \tag{5}$$

One solution for node scheduling is to ask helpers and MEC servers nearby an MD to provide feedback on their effective channel gains to the MBS, and then the MBS considers the nodes with the highest channel gains. However, if there are many MDs in the network, it will incur in a considerable system overhead to the MBS. Therefore, we consider random beamforming to reduce system overhead [42]. Random beamforming is exploited to serve NOMA nodes via one randomly generated beam as follows [40]:

$$P = \omega(\bar{\theta}) \tag{6}$$

where $\bar{\theta}$ is uniformly distributed in the range [-1,1]. Then, the effective channel gain of node k is defined as [40]:

$$\left|g_k^H P\right|^2 = \frac{N\left|\rho_k\right|^2\left|\omega\left(\theta_k\right)^H P\right|^2}{1+d_k^{\omega}} = \frac{\left|\rho_k\right|^2\left|\sum_{n=0}^{N-1} e^{j\pi n(\bar{\theta}-\theta_k)}\right|^2}{N\left(1+d_k^{\omega}\right)}$$

$$= \frac{|\rho_k|^2 \, sin^2\left(\frac{\pi N(\bar{\theta}-\theta_k)}{2}\right)}{N\left(1+d_k^{\omega}\right) sin^2\left(\frac{\pi(\bar{\theta}-\theta_k)}{2}\right)}$$

$$= \frac{|\rho_k|^2}{(1+d_k^{\omega})} F_M(\bar{\theta}-\theta_k) \tag{7}$$

where $F_M(\cdot)$ denotes the LV et al. [40] as follows:

$$F_M(V) = \frac{sin^2(0.5\pi VN)}{N\, sin^2(0.5\pi V)} \tag{8}$$

B. FORMULATION OF SPECTRAL EFFICIENCY

In the proposed scenario of NOMA transmission, one helper and one MEC server are selected. The SINR of MEC server at the helper (γ_S^{xw}) is calculated as [43]:

$$\gamma_S^{xw} = \frac{P_W |g_S|^2}{P_S |g_S|^2 + N_0} \tag{9}$$

where g_S is the channel gain coefficient from the user to the helper. The helper first decode the message of MEC server if $\gamma_S^{xw} > \gamma_W^{th}$, where γ_W^{th} is the required SINR for successful detection of the message of MEC server [43]. Then, the helper detects its desired signal after removing the message of MEC server known as SIC, with the following SINR (γ_S^{xs}):

$$\gamma_S^{xs} = \frac{P_S |g_S|^2}{N_0} \tag{10}$$

The received SINR at MEC server for detecting x_W (γ_W^{xw}) is calculated as [43]:

$$\gamma_W^{xw} = \frac{P_W |g_W|^2}{P_S |g_W|^2 + N_0} \tag{11}$$

where g_W denotes the channel gain coefficient from user to MEC server. Then, by denoting the bandwidth of this system as B Hz, the SE of helper (η_{SE}^S) and MEC server (η_{SE}^W) are computed as:

$$\eta_{SE}^S = B\, log_2\left(1+\gamma_S^{xs}\right) \tag{12}$$

$$\eta_{SE}^W = B\, log_2\left(1+\gamma_W^{xw}\right) \tag{13}$$

The helper offloads $L_U\beta_{h,M}$ bits with transmit power p_M to the MEC server at the second time slot. Similarly, the SNR ($\gamma_M^{h\to M}$) and SE ($\eta_{SE}^{h\to M}$) from helper to MEC server are defined as:

$$\gamma_M^{h\to M} = \frac{P_M |g_M|^2}{N_0} \tag{14}$$

$$\eta_{SE}^{h\to M} = B\, log_2\left(1+\gamma_M^{h\to M}\right) \tag{15}$$

where g_M denotes the channel coefficient from helper to MEC server.

C. OFFLOADING PHASE TIME AND ENERGY CONSUMPTION

By utilizing the SE in (12)-(13) and (15), the task offloading time from user to the helper and MEC server and from helper to MEC server can be written as [11]:

$$T_{U,S} = \frac{L_U\left(\beta_{U,h}\right)}{\eta_{SE}^S} \tag{16}$$

$$T_{U,W} = \frac{L_U\beta_{U,W}}{\eta_{SE}^W} \tag{17}$$

$$T_{h,M} = \frac{L_U\beta_{h,M}}{\eta_{SE}^{h\to M}} \tag{18}$$

The total offloading energy consumption of the system is obtained as:

$$E_{off} = T_{U,S}P_s + T_{U,W}P_W + T_{h,M}P_M \tag{19}$$

D. PROCESSING TIME AND ENERGY CONSUMPTION

Let us define the same CPU frequency at MD and helper as f_{loc} (in cycles per second) and the number of CPU cycles to compute one bit of their task as C. The local computation time of user (T_U^{loc}) and helper (T_h^{loc}) are written as:

$$T_U^{loc} = \frac{\beta_{U,loc}L_U C}{f_{loc}} \tag{20}$$

$$T_h^{loc} = \frac{\beta_{U,S}L_U C}{f_{loc}} \tag{21}$$

Thus, the total energy consumption of local computing can be given by:

$$E_{loc} = k_u\beta_{U,loc}L_U C(f_{loc})^2 + k_u\beta_{U,S}L_U C(f_{loc})^2 \tag{22}$$

where k_u is the effective capacitance coefficient of user or helper for each CPU cycle. Unlike the previous works [3], [14], [15], [16], [17], we consider execution latency (T_{MEC}) and energy consumption of MEC server (E_{MEC}) and define them as

$$T_{MEC} = \frac{(\beta_{U,W}+\beta_{h,M})L_U C}{f_{MEC}} \tag{23}$$

$$E_{MEC} = k_m(\beta_{U,W}+\beta_{h,M})L_U C(f_{MEC})^2 \tag{24}$$

where f_{MEC} denotes CPU frequency at MEC server. As the processed tasks produce small-size results, time, and power consumptions for downloading the computational results can be ignored [3]. Finally, the total energy consumption is given by:

$$E_T = E_{loc} + E_{off} + E_{MEC} \tag{25}$$

IV. SUCCESSFUL COMPUTATION PROBABILITY

The successful computation probability in the proposed scenario is defined as the probability of completing the task of the MD which includes 1) local computation by the MD and the helper 2) all offloading phases, 3) and completion of the offloaded tasks. Accordingly, successful computation probability can be written as:

$$P_{sc} = P\left(T_{U,S} < T, T_{U,W} < T, T_{h,M} < T, T_U^{loc} \leq 2T, T_h^{loc} \leq T, T_{MEC} \leq t_d\right) \tag{26}$$

Proposition 1: By utilizing (7) and any given parameters, i.e., $\beta_{U,S}$, $\beta_{U,W}$, $\beta_{h,M}$, T, P_s, P_W and P_M, the closed-form expression of (26) can be obtained as shown in (27), at the bottom of the page.

Proof: The proof is given in Appendix A.

Successful computation probability of the proposed network can be characterized through the closed-form expression of Proposition 1, which provides insight into understanding the proposed network. Moreover, through the closed-form expression, we will verify the results obtained in the simulation.

V. PROPOSED STRONG ALGORITHM

In this section, the node grouping, joint task assignment and power allocation problems to minimize the total energy consumption of NOMA-based MEC system is formulated as:

$$\left[G^* P^* \beta^*\right] = \underset{G,P,\beta}{\text{argmin}}\ E_T \tag{28}$$

subject to:

$$\beta_{U,S} + \beta_{U,W} + \beta_{h,M} + \beta_{U,loc} = 1 \tag{29}$$

$$\frac{L_U \beta_{U,S}}{\eta_{SE}^{S}} < T, \frac{L_U \beta_{U,W}}{\eta_{SE}^{W}} < T, \frac{L_U \beta_{h,M}}{\eta_{SE}^{h\to M}} < T \tag{30}$$

$$\frac{\beta_{U,S} L_U C}{f_{loc}} < T, \frac{\beta_{U,loc} L_U C}{f_{loc}} < 2T, \frac{(\beta_{U,W} + \beta_{h,\mathrm{M}}) L_U C}{f_{MEC}} < t_d \tag{31}$$

$$k_u \beta_{U,loc} L_U C (f_{loc})^2 + T_{U,S} P_s + T_{U,W} P_W < E_{max,u} \tag{32}$$

$$T_{h,M} P_M + k_u \beta_{U,A} L_U C (f_{loc})^2 < E_{max,h} \tag{33}$$

where G^*, P^* and β^* are the optimal grouping of helper and MEC server, allocated power to helper and MEC server and offloading task ratios of helper and MEC server, respectively. Constraint (29) ensures that the sum of offloading task ratios is equal to 1. Constraint (30) states that the offloading times of user and helper are restricted to T. Constraint (31) ensures that the delay for completing task locally in the user, helper and MEC server cannot exceed delay limits $2T$, T, and t_d, respectively, where it is assumed that t_d is lower than T. Constraint (32) indicates that the sum of MD's energies for offloading and locally computing the task cannot exceed its maximum energy shown by $E_{\max,u}$. Finally, constraint (33) ensures that the total energy of helper for offloading and locally computing cannot exceed its maximum energy denoted by $E_{\max,h}$.

Due to the high computational complexity of optimization problem in (28), we divide it into two independent sub-problems. In the first sub-problem, we present a sub-optimal solution for independently grouping MEC server with helper, while in the second sub-problem, power allocation and task assignment are jointly formulated aiming at minimizing the energy consumption of the proposed cooperative NOMA-based MEC system.

A. NODE GROUPING

We consider random beamforming for the designed cooperative NOMA-mmWave based MEC system, where a single random beam is generated by the MD, thus the knowledge of channel vectors of all nodes is not required. This can significantly reduce system overhead in an ultra-dense network. Also, since mmWave transmission is highly directional, in the proposed NOMA grouping algorithm, we avoid scheduling those helpers and MEC servers that may have low signal levels, which also reduces the search space; thus, reducing system overhead. We assume that helper selection procedure, which is performed by MBS, does not incur any significant latency due to its short execution time [30], [44], [45].

The optimal grouping of helper and MEC server for implementation of NOMA is a discrete problem. It can be solved by searching over all existing pairs of MEC servers and helpers. However, this approach is computationally expensive. To reduce the computational complexity and increase the successful computation probability, we propose an intuitive algorithm for MEC server selection. First, we select the highest channel gain MEC server and then search over the helpers among the users and select the helper with the minimum energy consumption. The selected helper node and the MEC server are then used for NOMA implementation. This scheme is executed at the MBS side and is described in more detail in Algorithm 1. It should be noted that the following algorithm has a very low execution time[1]; hence, the impact of this latency on the system is negligible.

[1]The execution takes 1 ms in a modest hardware.

$$P_{sc} = \begin{cases} exp\left(-\dfrac{N_0}{\left(\frac{P_W}{2^{\frac{L_U \beta_{U,W}}{BT}}-1} - P_S\right)} \dfrac{\left(1+d_{U,W}^{\alpha}\right)}{F_M\left(v-\theta_{U,W}\right)}\right) \cdot exp\left(-\left(2^{\frac{L_U\left(\beta_{U,S}+\beta_{U,W}\right)}{BT}} - 1\right) \dfrac{N_0}{P_S} \dfrac{\left(1+d_{U,S}^{\alpha}\right)}{F_M\left(v-\theta_{U,S}\right)}\right) \cdot \\ exp\left(-\left(2^{\frac{L_U \beta_{h,M}}{BT}} - 1\right) \dfrac{N_0}{p_M} \dfrac{\left(1+d_{h,M}^{\alpha}\right)}{F_M\left(v-\theta_{h,M}\right)}\right), \text{if } T_U^{loc} \leq 2T, T_h^{loc} \leq T, T_{MEC} \leq t_d \\ 0, \quad otherwise \end{cases} \tag{27}$$

Algorithm 1 Node Grouping Scheme

Input: The number of MEC servers as L and helpers as K
Channel gains: $|g_l|^2$ for $l = 1, 2, \ldots, L$
Output: Selected users, as l^*, k^*
1. Select MEC server
Calculate $g = \left[|g_1|^2, |g_2|^2, \ldots, |g_L|^2\right]$
$[\sim, order] = (sort(g), 'descend')$ $l^* = order(1)$;
2. Select helper
Calculate $E_T = \left[E_T^{l^*,1}, E_T^{l^*,2}, \ldots, E_T^{l^*,K}\right]$
$[:, order] = \min(E_T)$
$k^* = order$
3. Return l^* and k^*.

B. JOINT POWER ALLOCATION AND TASK ASSIGNMENT

After solving the first sub-problem, i.e., node grouping, the second sub-problem which minimizes the total energy consumption by allocating power and task portions, is formulated as

$$[P^* \beta^*] = \underset{P,\beta}{argmin}\ E_T \tag{34}$$

subject to (29)-(33).

The non-convexity of the objective function in (34) makes the optimization problem hard to solve. Hence, we present some mathematical solutions to transform (34) into a convex problem as follows.

Proposition 2: The optimal offloading time for the energy minimization problem (34) satisfies the following condition:

$$T_{U,S} = T_{U,W} = T_{h,M} = T \tag{35}$$

Proof: The proof is given in Appendix B.

Then, by utilizing (35), transmission powers can be equivalently transformed to:

$$P_M = \left(\left(2^{\frac{\beta_{h,M}L_U}{TB}} - 1\right)N_0\right) \Big/ \left(|g_M|^2\right) \tag{36}$$

$$P_S = \left(\left(2^{\frac{(\beta_{U,S}+\beta_{h,M})L_U}{TB}} - 1\right)N_0\right) \Big/ \left(|g_S|^2\right) \tag{37}$$

$$P_W = \left(\left(2^{\frac{\beta_{U,W}L_U}{TB}} - 1\right)\left(P_S \cdot |g_W|^2 + N_0\right)\right) \Big/ \left(|g_W|^2\right) \tag{38}$$

Proposition 2 provides transmission power of nodes, which enables us to introduce the following proposition.

Proposition 3: Utilizing the optimal values of transmission powers (36)-(38), The objective function given by (34) is convex.

Proof: The proof is given in Appendix C.

Then, due to the convexity of (34), the KKT conditions are applied to solve the maximization problem. Therefore, the Lagrangian of (34) considering the constraints (29)-(33) can be derived as:

$$\begin{aligned}
&L\left(\beta_{U,W}, \beta_{U,S}, \beta_{h,M}, \beta_{U,loc}, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5\right) \\
&= \Big(TP_W + TP_S + TP_M + k_u\beta_{U,loc}L_U C(f_{loc})^2 \\
&\quad + k_u\beta_{U,S}L_U C(f_{loc})^2 + k_m(\beta_{U,W} + \beta_{h,\mathrm{M}})L_U C(f_{MEC})^2\Big) \\
&\quad + \lambda_1\left(\frac{\beta_{U,S}L_U C}{f_{loc}} - T\right) + \lambda_2\left(\frac{\beta_{U,loc}L_U C}{f_{loc}} - 2T\right) \\
&\quad + \lambda_3\left(k_u\beta_{U,loc}L_U C(f_{loc})^2 + TP_S + TP_W - E_{max,u}\right) \\
&\quad + \lambda_4\left(k_u\beta_{U,S}L_U C(f_{loc})^2 + TP_M - E_{max,h}\right) \\
&\quad + \lambda_5\left(\beta_{U,S} + \beta_{U,W} + \beta_{h,M} + \beta_{U,loc} - 1\right) \\
&\quad + \lambda_6\left(\frac{(\beta_{U,W} + \beta_{h,\mathrm{M}})L_U C}{f_{MEC}} - t_d\right)
\end{aligned} \tag{39}$$

where λ_1, λ_2, λ_3, λ_4, λ_5, and λ_6 are the Lagrange multipliers of the constraints on delay for completing task locally in MD, and delay for computation on helper, sum of user's energies for offloading and locally computing the task, total energy of helper for offloading and locally computing, sum of offloading task portions, and processing time of the MEC server, respectively.

We can obtain the solution of (39) by a dual decomposition method as follows:

$$\begin{aligned}
&\underset{\lambda_1,\lambda_2,\lambda_3,\lambda_4,\lambda_5,\lambda_6}{max}\ \underset{\beta_{U,W},\beta_{U,S},\beta_{U,l\ oc}}{min} \\
&L\left(\beta_{U,W}, \beta_{U,S}, \beta_{h,M}, \beta_{U,l\ oc}, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6\right)
\end{aligned} \tag{40}$$

Partial derivative of (39) with respect to $\beta_{U,W}$, $\beta_{U,S}$, $\beta_{h,M}$, and $\beta_{U,loc}$ can be obtained as follows:

$$\begin{aligned}
\frac{\partial L}{\partial \beta_{U,W}} &= \frac{ln\ (2)\, L_U}{B} 2^{\frac{\beta_{U,W}L_U}{TB}}\left(P_S + \frac{N_0}{|g_W|^2}\right)(1+\lambda_3) + \lambda_5 \\
&\quad + k_m L_U C(f_{MEC})^2 + \lambda_6\frac{L_U C}{f_{MEC}}
\end{aligned} \tag{41}$$

$$\begin{aligned}
\frac{\partial L}{\partial \beta_{U,S}} &= ln\ (2)(1+\lambda_3)\left(\frac{L_U N_0 2^{\frac{(\beta_{U,S}+\beta_{h,M})L_U}{TB}}\left(2^{\frac{\beta_{U,W}L_U}{TB}}\right)}{B\,|g_S|^2}\right) \\
&\quad + k_u L_U C(f_{loc})^2(1+\lambda_4) + \lambda_1\frac{L_U C}{f_{loc}} + \lambda_5
\end{aligned} \tag{42}$$

$$\begin{aligned}
\frac{\partial L}{\partial \beta_{h,M}} &= (1+\lambda_3)\left(\frac{ln\ (2)\, L_U N_0 2^{\frac{(\beta_{U,S}+\beta_{h,M})L_U}{TB}}\left(2^{\frac{\beta_{U,W}L_U}{TB}}\right)}{B\,|g_S|^2}\right) \\
&\quad + \frac{ln\ (2)\, 2^{\frac{\beta_{h,M}L_U}{TB}} L_U N_0}{B\,|g_M|^2}(1+\lambda_4) \\
&\quad + \lambda_5 + k_m L_U C(f_{MEC})^2 + \lambda_6\frac{L_U C}{f_{MEC}}
\end{aligned} \tag{43}$$

$$\frac{\partial L}{\partial \beta_{U,loc}} = (1+\lambda_3)k_u L_U C(f_{loc})^2 + \lambda_2\frac{L_U C}{f_{loc}} + \lambda_5 \tag{44}$$

The values of $\beta_{U,W}\beta_{U,S}$, $\beta_{h,M}$ and $\beta_{U,loc}$ are obtained by setting (41)-(43) to zero as in (45)–(48), shown at the bottom of the next page, where $[x]^+$ denotes $(x, 0)$. Also, setting (44) to zero yields:

$$\lambda_5 = -(1+\lambda_3)k_u L_U C(f_{loc})^2 - \lambda_2\frac{L_U C}{f_{loc}} \tag{49}$$

To update Lagrange multipliers, gradient method is adopted as follows:

$$\lambda_1(i+1) = \left[\lambda_1(i) + \xi(i)\left(\frac{\beta_{U,S}(i)L_U C}{f_{loc}} - T\right)\right]^+ \tag{50}$$

$$\lambda_2(i+1) = \left[\lambda_2(i) + \xi(i)\left(\frac{\beta_{U,loc}(i)L_U C}{f_{loc}} - 2T\right)\right]^+ \tag{51}$$

$$\begin{aligned}\lambda_3(i+1) = \Bigg[\lambda_3(i) + \xi(i)\Bigg(&k_u\beta_{U,loc}(i)L_U C(f_{loc})^2 \\ &+TP_S 2^{\frac{\beta_{U,W}(i)L_U}{TB}} + \frac{T\left(2^{\frac{\beta_{U,W}(i)L_U}{TB}} - 1\right)N_0}{|g_W|^2} \\ &-E_{max,u}\Bigg)\Bigg]^+\end{aligned} \tag{52}$$

$$\begin{aligned}\lambda_4(i+1) = \Big[\lambda_4(i) + \xi(i)(&k_u\beta_{U,S}(i)L_U C(f_{loc})^2 \\ &+TP_M - E_{max,h})\Big]^+\end{aligned} \tag{53}$$

$$\lambda_6(i+1) = \left[\lambda_6(i) + \xi(i)\left(\frac{(\beta_{U,W} + \beta_{h,M})L_U C}{f_{MEC}} - t_d\right)\right]^+ \tag{54}$$

where $\xi(i)$ is a dynamically positive value step size updated as $\xi(i+1) = \xi(i)/i$. Algorithm 2 shows the iterative method for solving (39). The iterative algorithm continues until it converges to the minimum energy consumption, which occurs when the condition $\left((E_{loc} + E_{off} + E_{MEC}) - (E_{loc}(l) + E_{off}(l) + E_{MEC}(l))\right) \leq \varepsilon$ for a maximum tolerance ε is satisfied or the maximum number of L_{max} is reached.

By utilizing Algorithms 1 and 2, the problem of node grouping and joint power allocation and task assignment is solved. The computational complexity of the proposed method consists of the complexities of node grouping and task assignment scheme. The complexity of proposed node grouping algorithm is $O(KL)$. Also, the complexity of proposed task assignment scheme is of the order $O(I_{max}L_{max})$ [46].Thus, the overall complexity of node grouping, and task assignment scheme is $O(I_{max}L_{max} \cdot KL)$. This complexity is lower than some the grid search algorithm, where node grouping and task assignment has a complexity of $O\left((K+L)^{K+L} \cdot V_S^K \cdot V_W^L \cdot V_h^L\right)$ where V_S^K, V_W^L and V_h^L are the number of possible values of task assignment values of helper, MEC and helper to MEC, respectively. Moreover, as shown in the next section, the performance of proposed suboptimum algorithms is close to the grid search method with much lower complexity than the grid search method.

TABLE 2. Simulation parameters.

PARAMETER	VALUE
Radius R_{S_A}	10 m
Radius R_{S_C}	20 m
Radius R_{S_B}	30 m
$E_{max,u}$,	20 dBm
$E_{max,h}$	0 dBm
Noise power	-174 dBm/Hz
f_{loc}	0.5 GHz
C	1000 cycle/bit
L_U	10 Kbits
T	10 ms
M	64
k_u	10^{-27}
k_m	10^{-29}
f_{MEC}	1 GHz

VI. PERFORMANCE EVALUATION

In this section, numerical results obtained through computer simulations are provided to verify the derived analytical results and investigate the efficiency of the proposed solutions for STRONG problem for cooperative offloading in NOMA-mmWave MEC system. The general network parameters are summarized in Table 2.

A. P_{sc} WITH FIXED HELPER

In this sub-section, we evaluate the impacts of various parameters on the successful computation probability for the proposed node grouping scenario, where a helper with a fixed location is selected to group with MEC server.

Fig. 3(a) shows the successful computation probability versus the distance between user and MEC server for different lengths of task. It is observed that analytical and simulation results match very well, which verifies (27). Meanwhile, we can see that the successful computation probability decreases with the increase of distance. The reason

$$\beta^*_{U,S} = \left[\frac{TB}{L_U}log_2\left(\frac{-\left(k_u L_U C(f_{loc})^2(1+\lambda_4) + \lambda_1\frac{L_U C}{f_{loc}} + \lambda_5\right)B|g_S|^2}{L_U N_0 2^{\frac{(\beta^*_{U,W}+\beta^*_{h,M})L_U}{TB}} ln\ (2)(1+\lambda_3)}\right)\right]^+ \tag{45}$$

$$\beta^*_{U,W} = \left[\frac{TB}{L_U}log_2\left(\frac{B\left(k_u L_U C(f_{loc})^2(1+\lambda_4) + \lambda_1\frac{L_U C}{f_{loc}} - \left(k_m L_U C(f_{MEC})^2 + \lambda_6\frac{L_U C}{f_{MEC}}\right)\right)}{\left(\frac{1}{|g_W|^2} - \frac{1}{|g_S|^2}\right) ln\ (2)(1+\lambda_3)N_0 L_U}\right)\right]^+ \tag{46}$$

$$\beta^*_{h,M} = \left[\frac{TB}{L_U}log_2\left(\frac{B|g_M|^2(k_u L_U C(f_{loc})^2(1+\lambda_4) + \lambda_1\frac{L_U C}{f_{loc}} - \left(k_m L_U C(f_{MEC})^2 + \lambda_6\frac{L_U C}{f_{MEC}}\right)}{ln\ (2) L_U N_0(1+\lambda_4)}\right)\right]^+ \tag{47}$$

$$\beta^*_{U,loc} = \left[1 - \beta^*_{U,S} - \beta^*_{U,W} - \beta^*_{h,M}\right]^+ \tag{48}$$

Algorithm 2 Task Assignment Scheme

1. Initialize I_{max}, ε, L_{max}
2. For $l = 1 : L_{max}$, do
3. Initialize $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6$
4. For $i = 1 : I_{max}$, do
5. $\hat{\beta}_{U,S}(i)$ $$\longleftarrow \left[\frac{TB}{L_U} log_2 \left(\frac{-\left(k_u L_U C(f_{loc})^2(1+\lambda_4)+\lambda_1 \frac{L_U C}{f_{loc}}+\lambda_5\right)B|g_S|^2}{L_U N_0 2^{\frac{(\beta^*_{U,W}+\beta^*_{h,M})L_U}{TB}}(1+\lambda_3)} \right) \right]^+$$
6. $\hat{\beta}_{U,W}(i)$ $$\longleftarrow \left[\frac{TB}{L_U} log_2 \left(\frac{B\left(k_u L_U C(f_{loc})^2(1+\lambda_4)+\lambda_1 \frac{L_U C}{f_{loc}}-\left(k_m L_U C(f_{MEC})^2+\lambda_6 \frac{L_U C}{f_{MEC}}\right)\right)}{\left(\frac{1}{|g_W|^2}-\frac{1}{|g_S|^2}\right)\ln(2)(1+\lambda_3)N_0 L_U} \right) \right]^+$$
7. $\hat{\beta}_{h,M}(i)$ $$\longleftarrow \left[\frac{TB}{L_U} log_2 \left(\frac{B|g_M|^2(k_u L_U C(f_{loc})^2(1+\lambda_4)+\lambda_1 \frac{L_U C}{f_{loc}}-\left(k_m L_U C(f_{MEC})^2+\lambda_6 \frac{L_U C}{f_{MEC}}\right)}{\ln(2)L_U N_0(1+\lambda_4)} \right) \right]^+$$
8. $\hat{\beta}_{U,loc}(i) \longleftarrow \left[1 - \hat{\beta}_{U,S}(i)-\hat{\beta}_{U,W}(i)-\hat{\beta}_{h,M}(i)\right]^+$
9. Calculate $E_{off}\left(\hat{\beta}_{U,S}(i), \hat{\beta}_{U,W}(i), \hat{\beta}_{h,M}(i)\right)$, $E_{loc}\left(\hat{\beta}_{U,loc}(i), \hat{\beta}_{U,S}(i)\right)$ and $E_{MEC}(\hat{\beta}_{U,W}(i), \hat{\beta}_{h,M}(i))$
10. $\lambda_1(i+1) \longleftarrow \left[\lambda_1(i) + \xi(i)(\frac{\hat{\beta}_{U,S}(i)L_U C}{f_{loc}} - T)\right]^+$
11. $\lambda_2(i+1) \longleftarrow \left[\lambda_2(i) + \xi(i)(\frac{\hat{\beta}_{U,loc}(i)L_U C}{f_{loc}} - 2T)\right]^+$
12. $$\lambda_3(i+1) \longleftarrow \left[\lambda_3(i) + \xi(i)(k_u\hat{\beta}_{U,loc}(i)L_U C(f_{loc})^2 + T\frac{(2^{\frac{(\hat{\beta}_{U,S}(i)+\hat{\beta}_{h,M}(i))L_U}{TB}}-1)2^{\frac{\hat{\beta}_{U,W}(i)L_U}{TB}}N_0}{|g_S|^2} + \frac{T(2^{\frac{\hat{\beta}_{U,W}(i)L_U}{TB}}-1)N_0}{|g_W|^2} - E_{max,u})\right]^+$$
13. $$\lambda_4(i+1) \longleftarrow \left[\lambda_4(i) + \xi(i)(k_u\hat{\beta}_{U,S}(i)L_U C(f_{loc})^2 + T\frac{(2^{\frac{\hat{\beta}_{h,M}(i)L_U}{TB}}-1)N_0}{|g_M|^2} - E_{max,h})\right]^+$$
14. $\lambda_5(i+1) \longleftarrow -(1+\lambda_3(i))k_u L_U C(f_{loc})^2 - \lambda_2(i)\frac{L_U C}{f_{loc}}$
15. $\lambda_6(i+1) \longleftarrow \left[\lambda_6(i) + \xi(i)(\frac{(\hat{\beta}_{U,W}+\hat{\beta}_{h,\mathrm{M}})L_U C}{f_{MEC}} - t_d)\right]^+$
16. $\xi(i+1) \longleftarrow \xi(i)/i$
17. end for
18. if $\left((E_{loc} + E_{off} + E_{MEC}) - (E_{loc}(l) + E_{off}(l) + E_{MEC}(l))\right) \leq \varepsilon$, then
19. $\beta^*_{U,S} \longleftarrow \hat{\beta}_{U,S}(i)$
20. $\beta^*_{U,W} \longleftarrow \hat{\beta}_{U,W}(i)$
21. $\beta^*_{h,M} \longleftarrow \hat{\beta}_{h,M}(i)$
22. $\beta^*_{U,loc} \longleftarrow \hat{\beta}_{U,loc}(i)$
23. Break
24. Else
25. $\left(E_{off}(l+1) + E_{loc}(l+1) + E_{MEC}(l+1)\right) \longleftarrow \left(E_{off}(i) + E_{loc}(i) + E_{MEC}(i)\right)$
26. $i \longleftarrow i + 1$
27. end if
28. end for
29. return $\beta^*_{U,S}, \beta^*_{U,W}, \beta^*_{h,M}$ and $\beta^*_{U,loc}$

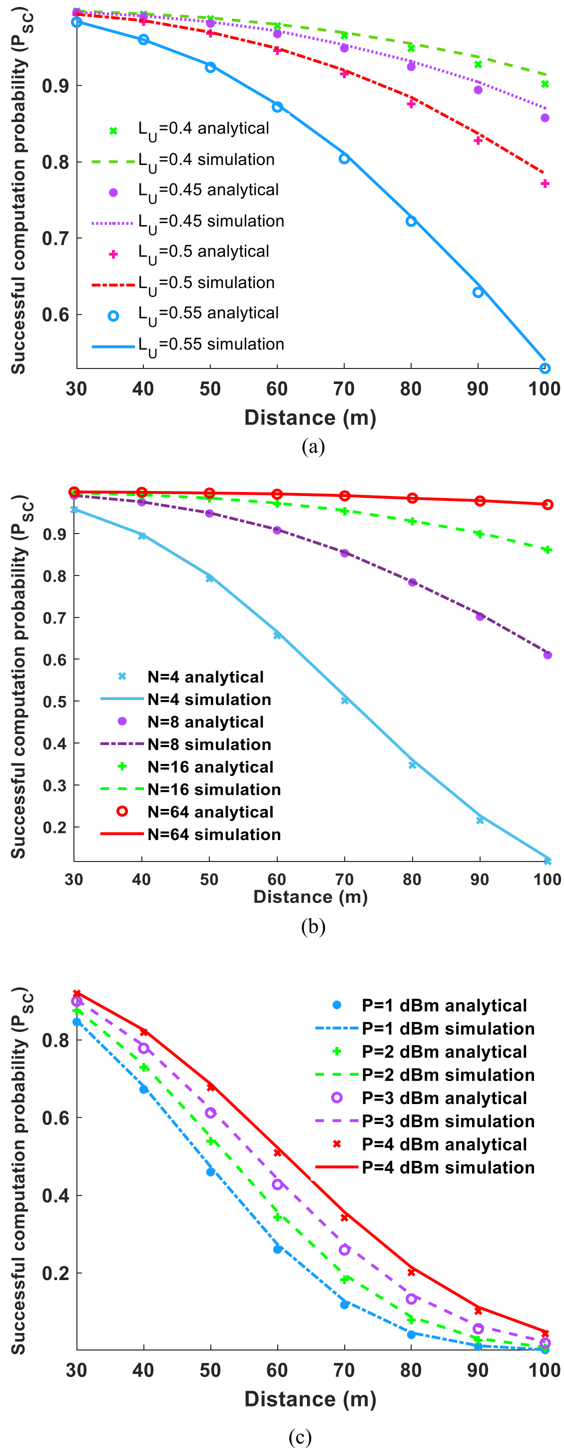


FIGURE 3. P_{sc} **versus distance between user and MEC server (a) for different** L_U**, (b) for different transmit antenna numbers** (N)**, and (c) for different transmit powers** (P)**.**

is that MEC server is assumed to be resource-rich, and by increasing the distance, the performance of the proposed cooperation scheme degrades due to the poor channel gain for performing cooperative computing. It can also be seen that the performance decreases by increasing the task length L_U, which is mainly due to the increasing of task offloading time.

In Fig 3(b), the impact of different number of transmit antennas (N) on the successful computation probability versus the distance between user and MEC server is depicted. As expected, increasing the number of transmit antennas results in higher successful computation probability since transmission links between user, helper, and MEC server are improved by providing higher directive gain resulting in less task offloading time.

Fig. 3(c) shows P_{sc} versus the distance between user and MEC server for different transmit powers (P). As expected, a higher transmit power leads to higher successful probability since task offloading time decreases.

B. PERFORMANCE COMPARISON

To show the efficiency of the proposed solution for STRONG problem, we provide a comparison with other benchmark solutions as following. Please note that in the rest of the paper we refer to our solutions to the joint problems as STRONG.

- *Reference [3]:* In this approach, a NOMA-aided cooperative computing scheme with three-node MEC system consisting of a helper, a user, and a MEC server is considered, in which user simultaneously offloads data to helper and MEC server by employing NOMA. Then, the helper can locally compute data or offload data to MEC server simultaneously, where the selection of helper is not considered, and the network is a traditional cellular network without utilization of mmWave band, beamforming, and MIMO antennas. Moreover, the energy and delay of MEC server are ignored in this work.
- *Proposed approach without cooperation (PWC)*: In this scheme, the user offloads data to helper and MEC server simultaneously and the helper can only compute data locally.
- *Proposed approach without node grouping (PWNG)*: In this scheme, the proposed system is considered with constant helper and MEC system without any node grouping scheme.

In Fig. 4, energy consumption as a function of input data of user L_U is demonstrated. It can be observed that the STRONG scheme outperforms the other benchmarks. Energy consumptions of all methods increase with the increment of input data (L_U). The difference between PWNG and STRONG methods indicates the efficiency of proposed node grouping scheme. Also, PWC scheme consumes more energy than STRONG and the PWNG methods. The reason is that in PWC scheme, the helper cannot offload data to MEC server although it is near to MEC server, and offloading data to MEC server is more efficient. Thus, the benchmarks with cooperation phase such as STRONG and PWNG approach outperform PWC. Moreover, although [3] and STRONG scheme have some similarities, in [3], mmWave band, beamforming, and MIMO antennas were not implemented. Furthermore, no node grouping algorithm was considered in [3]. In addition, in [3], the authors do not have any constraints on the available energy of user and helper for offloading

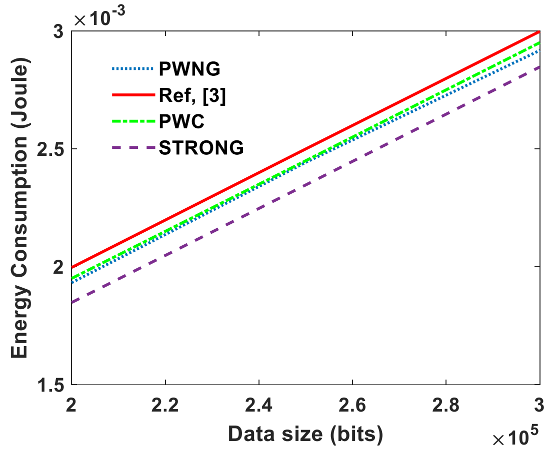


FIGURE 4. Average energy consumption versus data size (L_U).

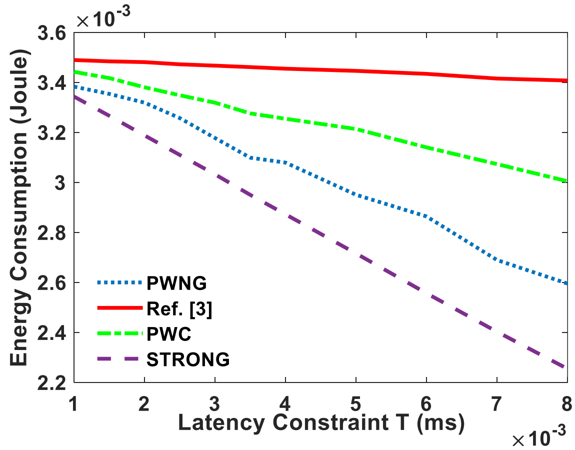


FIGURE 5. Average energy consumption versus T.

and local computation in the energy consumption minimization problem. It should be noticed that the results for all schemes in Figs. 4-7, are obtained by using the same parameter values used in our proposed solution for the sake of fairness.

Fig. 5, depicts energy consumption as a function of latency constraint T. As shown, all schemes experience reduction of energy consumption by increasing the latency constraints, since, for small T, user and helper need to offload input bits at a high rate to satisfy latency constraint, which leads to high energy consumption. It is seen that [3] has the gentlest slope. This is due to the fact that mmWave band, beamforming, MIMO antennas, and node grouping were not implemented in [3] and nodes experience lower channel gains. Therefore, for offloading, higher power is required; so local computation is more preferred by system. Whereas, in other schemes offloading leads to lower energy consumption.

Fig. 6 represents the results of offloading data which is defined as $(1-\beta_{U,loc})L_U$ versus the latency constraint. As observed, STRONG scheme achieves the best results

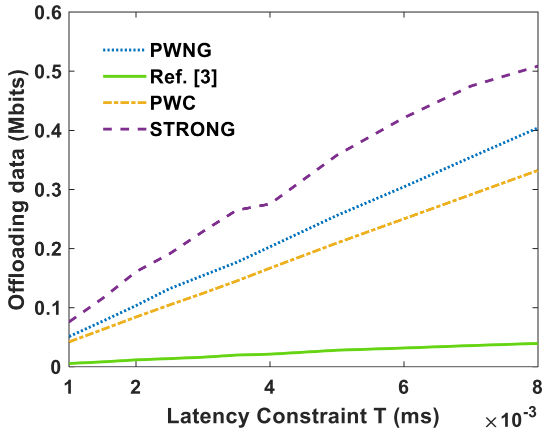


FIGURE 6. Average offloading data versus latency constraint T.

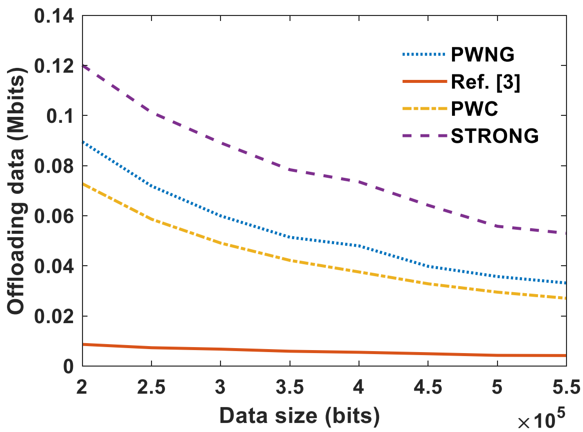


FIGURE 7. Average offloading data versus data size (L_U).

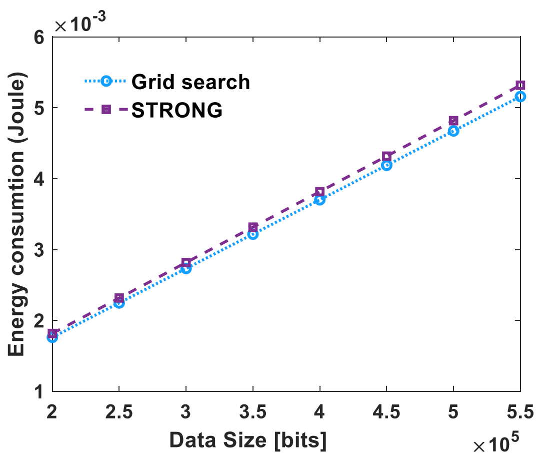


FIGURE 8. Comparison between the performance of optimum and proposed algorithm versus data size (L_U).

in comparison with other schemes. Also, in all methods, offloading data increases with the increment of latency constraint. This fact can be explained from the relation between offloading parts of data (45)-(48) and time constraint which is approximately linear.

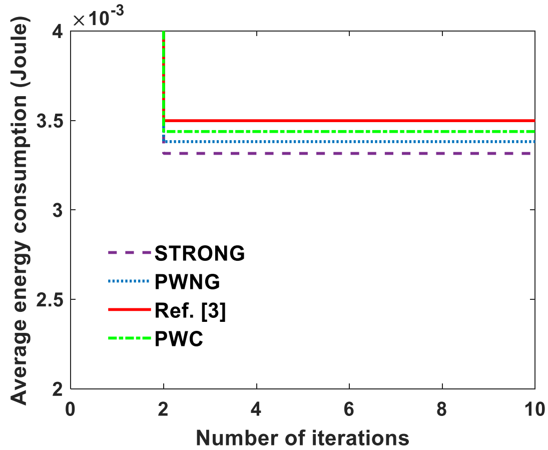


FIGURE 9. Convergence speed of the proposed solution.

In Fig. 7, the impact of number of input computation bits (L_U) on the offloading data is shown. We can observe the effectiveness of STRONG scheme which provides a gap with other benchmarks. The offloading data in all schemes decrease as L_U increases. This result can be explained from (45)-(48) where larger L_U leads to smaller offloading portion, and local computation becomes more energy efficient than offloading.

In Fig. 8, the small gap between the STRONG and grid search shows the efficiency of proposed algorithm with low complexity order. Moreover, by averaging the result of 10^5 simulations in Fig. 9, it can be observed that the proposed solution converged after small iteration. Therefore, the proposed algorithm leads to near optimal solution with low complexity and small number of iterations.

VII. CONCLUSION

In this paper, we proposed joint task assignment, power allocation, and node grouping for cooperative computing in NOMA-mmWave MEC system, where each user offloads computation tasks to MEC and an idle nearby MD. We studied energy consumption minimization problem under the constraints on the energies, delays, and sum of offloading task ratios. Despite the non-convexity of the formulated problem, we presented a two-step solution that decreased the computation complexity. We proposed efficient algorithms to compute sub-optimal solutions for powers allocated to helper and MEC server, offloading task ratios of helper and MEC, and node grouping. Also, we derived the expression for the successful computation probability which provides insight the proposed network. An extensive performance evaluation was presented to validate our proposed algorithms. The advantage of the proposed cooperative NOMA-assisted offloading in terms of reducing energy consumption compared to the existing works was shown. As a future work, we plan to extend our single-user scenario to multi-user cooperative NOMA-mmWave MEC system and incorporate ML solutions into our approach.

APPENDIX A
PROOF OF PROPOSITION 1

Proposition 1 is driven through law of total probability and considering that the variables $|g_S|^2$, $|g_W|^2$, and $|g_M|^2$ are not present in the $T_U^{loc} \geq 2T, T_h^{loc} \geq T, T_{MEC} \geq t_d$ and $T_U^{loc} \leq 2T, T_h^{loc} \leq T, T_{MEC} \leq t_d$ we conclude that

$$\begin{cases} P(T_U^{loc} \geq 2T, T_h^{loc} \geq T, T_{MEC} \geq t_d) = 1 \\ P(T_U^{loc} \leq 2T, T_h^{loc} \leq T, T_{MEC} \leq t_d) = 1 \end{cases} \tag{55}$$

By utilizing the law of total probability and (55), the successful computation probability in $P(T_U^{loc} \geq 2T, T_h^{loc} \geq T, T_{MEC} \geq t_d) = 1$ equals zero. Thus, the successful computation probability for $P(T_U^{loc} \leq 2T, T_h^{loc} \leq T, T_{MEC} \leq t_d) = 1$ can be rewritten as (56), shown at the bottom of the page.

APPENDIX B
PROOF OF PROPOSITION 2

In problem (34), the energy consumption for offloading the task to helper can be expressed as:

$$E_{S,off} = \frac{P_S L_U(\beta_{U,S}+\beta_{h,M})}{B\, log_2\left(1 + \frac{p_S|g_S|^2}{N_0}\right)} \tag{57}$$

The partial derivative of the above function with respect to P_S and task coefficient $\beta_{U,S}$ and $\beta_{h,M}$ is obtained as (58), shown at the bottom of the page. The above equation is more than 0. Thus, $E_{S,off}$ increases with the increase of transmit power. Also, the offloading time can be expressed as:

$$T_{U,S} = \frac{L_U(\beta_{U,S}+\beta_{h,M})}{\eta_{SE}^S} \tag{59}$$

By reducing the transmit power P_S, offloading time will be increased. Therefore, for an energy minimization problem, the maximum possible transmission time T should be utilized for each node. Also, in NOMA transmission, the transmission times of both nodes should be the same. Hence, the offloading time of both nodes is equal to T. Similarly, the offloading time of helper in the second phase is equal to T too. Thus, the Proposition 2 is proved.

APPENDIX C
PROOF OF PROPOSITION 3

The convexity of (34) is proved through positive semi-definite Hessian matrix [47]. The Hessian matrix of the objective function (34) can be shown as in (60), at the top of the next page, where $\Omega = \frac{N_0 2^{\frac{(\beta_{U,S}+\beta_{h,M}+\beta_{U,W})L_U}{TB}}}{T|g_S|^2}$. Then, we define a non-zero column vector $V = [V_1 V_2 V_3]^T$ to obtain:

$$\begin{aligned} &V^T H\,(EE)\,V \\ &= \left(\frac{L_U ln\ (2)}{B}\right)^2 \left(\frac{N_0 2^{\frac{(\beta_{U,S}+\beta_{h,M}+\beta_{U,W})L_U}{TB}}}{T\,|g_S|^2}\right) \\ &\quad \times (V_1+V_2+V_3)^2 \\ &\quad + \left(\frac{L_U ln\ (2)}{B}\right)^2 2^{\frac{\beta_{U,W}L_U}{TB}} N_0 V_1^2 \left(\frac{1}{T\,|g_W|^2} - \frac{1}{T\,|g_S|^2}\right) \\ &\quad + \left(\frac{L_U ln\ (2)}{B}\right)^2 \frac{\left(2^{\frac{\beta_{h,M}L_U}{TB}}\right) N_0}{T\,|g_M|^2} V_3^2 \end{aligned}$$

$$\begin{aligned} P_{sc} &= P\Bigg(N_0 \Big/ \left(\frac{P_W}{2^{\frac{L_U\beta_{U,W}}{BT}}-1} - P_S\right) < |g_W|^2, \left(2^{\frac{L_U(\beta_{U,S}+\beta_{h,M})}{BT}}-1\right)\frac{N_0}{P_S} < |g_S|^2, \left(2^{\frac{L_U\beta_{h,M}}{BT}}-1\right)\frac{N_0}{p_M} \\ &\quad < |g_M|^2 \,|(T_U^{loc} \leq 2T, T_h^{loc} \leq T, T_{MEC} \leq t_d)\Bigg) \\ &\quad \cdot P\left(T_U^{loc} \leq 2T, T_h^{loc} \leq T, T_{MEC} \leq t_d\right) = \frac{\left(1+d_{U,W}^{\alpha}\right)}{F_M\left(v-\theta_{U,W}\right)} \frac{\left(1+d_{U,S}^{\alpha}\right)}{F_M\left(v-\theta_{U,S}\right)} \frac{\left(1+d_{h,M}^{\alpha}\right)}{F_M\left(v-\theta_{h,M}\right)} \\ &\quad \cdot \int_{\left(2^{\frac{L_U\beta_{h,M}}{BT}}-1\right)\frac{N_0}{p_M}}^{\infty} \int_{\left(2^{\frac{L_U(\beta_{U,S}+\beta_{h,M})}{BT}}-1\right)\frac{N_0}{P_S}}^{\infty} \int_{\frac{N_0}{\left(\frac{P_W}{2^{\frac{L_U\beta_{U,W}}{BT}}-1}-P_S\right)}}^{\infty} e^{-x\frac{\left(1+d_{U,W}^{\alpha}\right)}{F_M\left(v-\theta_{U,W}\right)}} e^{-x\frac{\left(1+d_{U,S}^{\alpha}\right)}{F_M\left(v-\theta_{U,S}\right)}} e^{-x\frac{\left(1+d_{h,M}^{\alpha}\right)}{F_M\left(v-\theta_{h,M}\right)}} dx \\ &= e^{-\frac{N_0}{\left(\frac{P_W}{2^{\frac{L_U\beta_{U,W}}{BT}}-1}-P_S\right)}\frac{(1+d_{U,W}^{\alpha})}{F_M(v-\theta_{U,W})}} e^{-\left(2^{\frac{L_U(\beta_{U,S}+\beta_{h,M})}{BT}}-1\right)\frac{N_0}{P_S}\frac{(1+d_{U,S}^{\alpha})}{F_M(v-\theta_{U,S})}} e^{-\left(2^{\frac{L_U\beta_{h,M}}{BT}}-1\right)\frac{N_0}{p_M}\frac{(1+d_{h,M}^{\alpha})}{F_M(v-\theta_{h,M})}} \end{aligned} \tag{56}$$

$$\frac{\partial E_{S,off}}{\partial P_S} = \frac{L_U(\beta_{U,S}+\beta_{h,M}) ln\ (2) \left(ln\ \left(1+\frac{p_S|g_S|^2}{N_0}\right)\left(1+\frac{p_S|g_S|^2}{N_0}\right) - \frac{p_S|g_S|^2}{N_0}\right)}{B\, ln^2\left(1+\frac{p_S|g_S|^2}{N_0}\right)\left(1+\frac{p_S|g_S|^2}{N_0}\right)} \tag{58}$$

$$H(EE) = \left(\frac{L_U ln\ (2)}{B}\right)^2 \cdot \begin{bmatrix} 2^{\frac{\beta_{U,W}L_U}{TB}}\left(\frac{\left(2^{\frac{(\beta_{U,S}+\beta_{h,M})L_U}{TB}}-1\right)N_0}{T|g_S|^2}+\frac{N_0}{T|g_W|^2}\right) & \Omega & \Omega \\ \Omega & \Omega & \Omega \\ \Omega & \Omega & \left(\Omega+\frac{2^{\frac{\beta_{h,M}L_U}{TB}}N_0}{T|g_M|^2}\right) \end{bmatrix} \quad (60)$$

Since $|g_S|^2 > |g_W|^2$, V_1, V_2 and V_3 are non-zero, and all the other parameters are positive, $V^T H(EE) V > 0$; consequently, the Hessian matrix of the objective function (34) is positive semi-definite and the objective function (34) is convex.

REFERENCES

[1] A. Bozorgchenani, F. Mashhadi, D. Tarchi, and S. A. S. Monroy, "Multi-objective computation sharing in energy and delay constrained mobile edge computing environments," *IEEE Trans. Mobile Comput.*, vol. 20, no. 10, pp. 2992–3005, Oct. 2021.

[2] P. Mach and Z. Becvar, "Mobile edge computing: A survey on architecture and computation offloading," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1628–1656, 3rd Quart., 2017.

[3] Y. Huang, Y. Liu, and F. Chen, "NOMA-aided mobile edge computing via user cooperation," *IEEE Trans. Commun.*, vol. 68, no. 4, pp. 2221–2235, Apr. 2020.

[4] T. N. Nguyen and T. Jeon, "A conditional clustering scheme for hybrid NOMA in millimeter wave communication system," *Int. J. Adv. Smart Converg.*, vol. 8, no. 4, pp. 34–39, 2019.

[5] M. Merluzzi, P. Di Lorenzo, S. Barbarossa, and V. Frascolla, "Joint resource allocation for latency-constrained dynamic computation offloading with MEC," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshop (WCNCW)*, Apr. 2019, pp. 1–6.

[6] A. Ghosh, T. A. Thomas, M. C. Cudak, R. Ratasuk, P. Moorut, F. W. Vook, T. S. Rappaport, G. R. MacCartney, S. Sun, and S. Nie, "Millimeter-wave enhanced local area systems: A high-data-rate approach for future wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1152–1163, Jun. 2014.

[7] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2294–2323, 3rd Quart., 2018.

[8] F. Zhou, Y. Wu, R. Q. Hu, Y. Wang, and K. K. Wong, "Energy-efficient NOMA enabled heterogeneous cloud radio access networks," *IEEE Netw.*, vol. 32, no. 2, pp. 152–160, Mar. 2018.

[9] L. Zhu, Z. Xiao, X.-G. Xia, and D. Oliver Wu, "Millimeter-wave communications with non-orthogonal multiple access for B5G/6G," *IEEE Access*, vol. 7, pp. 116123–116132, 2019.

[10] C. Zhao, Y. Cai, A. Liu, M. Zhao, and L. Hanzo, "Mobile edge computing meets mmWave communications: Joint beamforming and resource allocation for system delay minimization," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2382–2396, Apr. 2020.

[11] F. Fang, Y. Xu, Z. Ding, C. Shen, M. Peng, and G. K. Karagiannidis, "Optimal task assignment and power allocation for NOMA mobile-edge computing networks," 2019, *arXiv:1904.12389*.

[12] C. You, K. Huang, H. Chae, and B.-H. Kim, "Energy-efficient resource allocation for mobile-edge computation offloading," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1397–1411, Mar. 2017.

[13] W. Zhang, Y. Wen, K. Guan, D. Kilper, H. Luo, and D. O. Wu, "Energy-optimal mobile cloud computing under stochastic wireless channel," *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4569–4581, Sep. 2013.

[14] A. Kiani and N. Ansari, "Edge computing aware NOMA for 5G networks," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 1299–1306, Apr. 2018.

[15] F. Wang, J. Xu, and Z. Ding, "Multi-antenna NOMA for computation offloading in multiuser mobile edge computing systems," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2450–2463, Mar. 2019.

[16] Z. Ding, J. Xu, O. A. Dobre, and H. V. Poor, "Joint power and time allocation for NOMA–MEC offloading," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 6207–6211, Jun. 2019.

[17] Z. Song, Y. Liu, and X. Sun, "Joint radio and computational resource allocation for NOMA-based mobile edge computing in heterogeneous networks," *IEEE Commun. Lett.*, vol. 22, no. 12, pp. 2559–2562, Dec. 2018.

[18] Y. Pan, M. Chen, Z. Yang, N. Huang, and M. Shikh-Bahaei, "Energy-efficient NOMA-based mobile edge computing offloading," *IEEE Commun. Lett.*, vol. 23, no. 2, pp. 310–313, Feb. 2019.

[19] Y. Wu, B. Shi, L. P. Qian, F. Hou, J. Cai, and X. S. Shen, "Energy-efficient multi-task multi-access computation offloading via NOMA transmission for IoTs," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4811–4822, Jul. 2020.

[20] Y. Wu, K. Ni, C. Zhang, L. P. Qian, and D. H. K. Tsang, "NOMA-assisted multi-access mobile edge computing: A joint optimization of computation offloading and time allocation," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 12244–12258, Dec. 2018.

[21] Y. Wu, L. P. Qian, K. Ni, C. Zhang, and X. Shen, "Delay-minimization nonorthogonal multiple access enabled multi-user mobile edge computation offloading," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 3, pp. 392–407, Jun. 2019.

[22] L. Lv, J. Chen, Q. Ni, Z. Ding, and H. Jiang, "Cognitive non-orthogonal multiple access with cooperative relaying: A new wireless frontier for 5G spectrum sharing," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 188–195, Apr. 2018.

[23] J. Liu, Y. Mao, J. Zhang, and K. B. Letaief, "Delay-optimal computation task scheduling for mobile-edge computing systems," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 1451–1455.

[24] T. Zhao, S. Zhou, X. Guo, and Z. Niu, "Tasks scheduling and resource allocation in heterogeneous cloud for delay-bounded mobile edge computing," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–7.

[25] Y. Zhong, X. Ge, T. Han, Q. Li, and J. Zhang, "Tradeoff between delay and physical layer security in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 7, pp. 1635–1647, Jul. 2018.

[26] X. Cao, F. Wang, J. Xu, R. Zhang, and S. Cui, "Joint computation and communication cooperation for energy-efficient mobile edge computing," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4188–4200, Jun. 2019.

[27] S. S. Yilmaz and B. Özbek, "Multi-helper NOMA for cooperative mobile edge computing," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 9819–9828, Jul. 2022.

[28] M. Ren, J. Chen, B. He, K. Wu, Y. Zhou, X. Xue, and L. Yang, "Cooperative NOMA-MEC with helper scheduling," *IEEE Commun. Lett.*, vol. 26, no. 7, pp. 1693–1697, Jul. 2022.

[29] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, "On multi-access edge computing: A survey of the emerging 5G network edge cloud architecture and orchestration," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1657–1681, 3rd Quart., 2017.

[30] Y. Ye, R. Q. Hu, G. Lu, and L. Shi, "Enhance latency-constrained computation in MEC networks using uplink NOMA," *IEEE Trans. Commun.*, vol. 68, no. 4, pp. 2409–2425, Apr. 2020.

[31] H. Sun, F. Zhou, and R. Q. Hu, "Joint offloading and computation energy efficiency maximization in a mobile edge computing system," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 3052–3056, Mar. 2019.

[32] Z. Liang, Y. Liu, T.-M. Lok, and K. Huang, "Multiuser computation offloading and downloading for edge computing with virtualization," *IEEE Trans. Wireless Commun.*, vol. 18, no. 9, pp. 4298–4311, Sep. 2019.

[33] X. Wang, X. Chen, W. Wu, N. An, and L. Wang, "Cooperative application execution in mobile cloud computing: A Stackelberg game approach," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 946–949, May 2016.

[34] C. You and K. Huang, "Exploiting non-causal CPU-state information for energy-efficient mobile cooperative computing," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 4104–4117, Jun. 2018.

[35] L. Pu, X. Chen, J. Xu, and X. Fu, "D2D fogging: An energy-efficient and incentive-aware task offloading framework via network-assisted D2D collaboration," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3887–3901, Dec. 2016.

[36] A. Bozorgchenani, S. Disabato, D. Tarchi, and M. Roveri, "An energy harvesting solution for computation offloading in fog computing networks," *Comput. Commun.*, vol. 160, pp. 577–587, Jul. 2020.

[37] F. Mashhadi, S. A. S. Monroy, A. Bozorgchenani, and D. Tarchi, "Optimal auction for delay and energy constrained task offloading in mobile edge computing," *Comput. Netw.*, vol. 183, Dec. 2020, Art. no. 107527.

[38] A. Bozorgchenani, S. Maghsudi, D. Tarchi, and E. Hossain, "Computation offloading in heterogeneous vehicular edge networks: On-line and off-policy bandit solutions," *IEEE Trans. Mobile Comput.*, vol. 21, no. 12, pp. 4233–4248, Dec. 2022.

[39] X. Lan, L. Cai, and Q. Chen, "Execution latency and energy consumption tradeoff in mobile-edge computing systems," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Aug. 2019, pp. 123–128.

[40] T. Lv, Y. Ma, J. Zeng, and P. T. Mathiopoulos, "Millimeter-wave NOMA transmission in cellular M2M communications for Internet of Things," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1989–2000, Jun. 2018.

[41] G. Lee, Y. Sung, and J. Seo, "Randomly-directional beamforming in millimeter-wave multiuser MISO downlink," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1086–1100, Feb. 2016.

[42] A. Khazali, D. Tarchi, M. G. Shayesteh, H. Kalbkhani, and A. Bozorgchenani, "Energy efficient uplink transmission in cooperative mmWave NOMA networks with wireless power transfer," *IEEE Trans. Veh. Technol.*, vol. 71, no. 1, pp. 391–405, Jan. 2022.

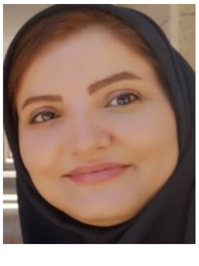
[43] Y. Liu, Z. Ding, M. Elkashlan, and H. V. Poor, "Cooperative non-orthogonal multiple access with simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 938–953, Apr. 2016.

[44] X. Hu, K.-K. Wong, and K. Yang, "Wireless powered cooperation-assisted mobile edge computing," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2375–2388, Apr. 2018.

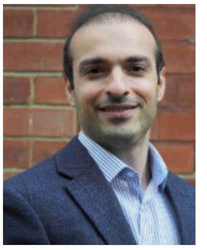
[45] Z. Yang, C. Pan, J. Hou, and M. Shikh-Bahaei, "Efficient resource allocation for mobile-edge computing networks with NOMA: Completion time and energy minimization," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 7771–7784, Nov. 2019.

[46] S. Sobhi-Givi, M. G. Shayesteh, and H. Kalbkhani, "Energy-efficient power allocation and user selection for mmWave-NOMA transmission in M2M communications underlaying cellular heterogeneous networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 9, pp. 9866–9881, Sep. 2020.

[47] K. Wang, F. Fang, and Z. Ding, "Joint optimization of task assignment and power allocation for NOMA-aided MEC systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Waikoloa, HI, USA, Dec. 2019, pp. 1–6.



AZADEH KHAZALI received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from Urmia University, Iran. She spent a 6-month research activity with the University of Bologna, Italy, as a Visiting Researcher. Her research interests include cellular networks, optimization technique, 5G, and 6G.



ARASH BOZORGCHENANI (Member, IEEE) received the B.Sc. and M.Sc. degrees in IT, in 2013 and 2016, respectively, Iran, and the Ph.D. degree in telecommunications and IT from the University of Bologna, Italy, in 2019. Later, he held a postdoctoral positions with the University of Bologna, and with Lancaster University, U.K., from 2019 to 2023. He was a Visiting Researcher with the University of Manitoba, Canada, in 2019. He is currently an Assistant Professor (Lecturer) with the University of Leeds, U.K. He has been involved in both national Gaucho (PRIN 2015, Italy) and European (H2020 SANCUS) projects. His research interests include resource allocation, optimization and machine learning techniques in wireless communications, and future networked systems (B5G/6G). He was a TPC member, the Session Chair, and an organizing member of IEEE flagship conferences.

DANIELE TARCHI (Senior Member, IEEE) was born in Florence, Italy, in 1975. He received the M.Sc. degree in telecommunications engineering and the Ph.D. degree in informatics and telecommunications engineering from the University of Florence, Florence, in 2000 and 2004, respectively.

From 2004 to 2010, he was a Research Associate with the University of Florence. From 2010 to 2019, he was an Assistant Professor with the University of Bologna, Bologna, Italy. He is currently an Associate Professor with the University of Bologna. He is the author of around 150 published articles in international journals and conference proceedings. He has been involved in several national and international research projects and collaborates with several foreign research institutes. His research interests include wireless communications and networks, satellite communications and networks, edge computing, fog computing, smart cities, and optimization techniques.

Prof. Tarchi has been the Symposium Co-Chair of IEEE WCNC 2011, IEEE Globecom 2014, IEEE Globecom 2018, and IEEE ICC 2020; and a workshop Co-Chair of IEEE ICC 2015. He is an Editorial Board Member for IEEE Transactions on Vehicular Technology, IEEE Open Journal of the Communication Society, and *IET Communications*.

MAHROKH G. SHAYESTEH (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from the University of Tehran, Tehran, Iran, the M.Sc. degree in electrical engineering from the Khajeh Nasir University of Technology, Tehran, and the Ph.D. degree in electrical engineering from the Amir Kabir University of Technology, Tehran. She is currently a Professor with the Department of Electrical Engineering, Urmia University, Urmia, Iran. She is also with the Wireless Research Laboratory, Advanced Communication Research Institute (ACRI), Department of Electrical Engineering, Sharif University of Technology, Tehran. Her research interests include wireless communications, signal, and image processing.

HASHEM KALBKHANI received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from Urmia University, Iran. He is currently an Assistant Professor with the Faculty of Electrical Engineering, Urmia University of Technology, Urmia, Iran. His research interests include wireless networks, machine learning, and signal processing.

Open Access funding provided by 'Alma Mater Studiorum - Università di Bologna' within the CRUI CARE Agreement