

Hybrid NOMA Offloading in Multi-User MEC Networks

Zhiguo Ding, *Fellow, IEEE*, Dongfang Xu, *Student Member, IEEE*, Robert Schober, *Fellow, IEEE*, and H. Vincent Poor, *Life Fellow, IEEE*

Abstract

Non-orthogonal multiple access (NOMA) assisted mobile edge computing (MEC) has recently attracted significant attention due to its superior capability to reduce the energy consumption and the latency of MEC offloading. In this paper, a general hybrid NOMA-MEC offloading strategy is proposed, which includes conventional orthogonal multiple access (OMA) and pure NOMA based offloading as special cases. A multi-objective optimization problem is formulated to minimize the energy consumption for MEC offloading, and a low-complexity resource allocation solution is derived and shown to be Pareto-optimal. Furthermore, by analyzing the properties of the obtained resource allocation solution, important insights regarding NOMA-MEC offloading are obtained. For example, it is proved that pure NOMA-MEC offloading cannot outperform hybrid NOMA-MEC. In addition, a precise condition under which NOMA-MEC outperforms OMA-MEC is established, and shown to match the one previously developed for the two-user special case. Furthermore, the developed analytical results also establish an interesting analogy between the proposed hybrid NOMA-MEC power allocation scheme and the well-known water-filling strategy.

Index Terms

Non-orthogonal multiple access (NOMA), mobile edge computing (MEC), multi-objective optimization, Pareto optimality.

I. INTRODUCTION

Mobile edge computing (MEC) has been recognized as an important enabling technique for the next generation of wireless networks [1]–[3]. The key idea behind MEC is to employ the infrastructure close to the edge of mobile networks, such as base stations, as computing servers, such that the mobile users can offload their computationally-intensive delay-critical tasks to the

Z. Ding and H. V. Poor are with the Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544, USA. Z. Ding is also with the School of Electrical and Electronic Engineering, the University of Manchester, Manchester, UK (email: zhiguo.ding@manchester.ac.uk, poor@princeton.edu). D. Xu and R. Schober are with the Institute for Digital Communications, Friedrich-Alexander-University Erlangen-Nurnberg (FAU), Germany (email: dongfang.xu@fau.de, robert.schober@fau.de).

servers. Because of the superior computing capabilities of the MEC servers, these tasks can be computed faster than when they are computed locally. From the energy consumption perspective, MEC offers mobile users the benefit of prolonging their battery lifetimes, since they do not have to spend energy for local computing. The concept of MEC is particularly important for the application of machine learning in the context of the wireless Internet of Things (IoT), where devices often are energy constrained and have limited computing capabilities [4]–[6]. Instead of relying on the IoT devices to locally compute machine learning tasks, MEC offers an energy efficient and low-latency alternative for task computation.

Energy and spectrally efficient offloading is key to the success of MEC networks, which motivates the application of non-orthogonal multiple access (NOMA) to MEC offloading [7], [8]. In particular, in [9], analytical results for the probability of successful MEC offloading were developed to demonstrate that the use of NOMA transmission can significantly reduce the delay and the energy consumption of MEC offloading. In [10]–[12], NOMA-MEC networks employing orthogonal frequency division multiplexing (OFDM) were studied, where sophisticated approaches for joint power and subcarrier allocation were developed. In [13]–[15], the use of unmanned aerial vehicles (UAVs) in NOMA-MEC networks was considered, and various designs for joint MEC resource allocation and UAV trajectory planning were developed. In [16] and [17], security provisioning for NOMA-MEC networks was investigated, where the presence of passive eavesdroppers was assumed and different approaches for security enhancement compared to orthogonal multiple access (OMA) based MEC were developed. To facilitate energy-constrained IoT, the application of wireless power transfer was considered in [18] and [19], where sophisticated algorithms for joint task allocation and delay minimization were proposed.

Unlike the aforementioned works, which require multiple users to complete their MEC offloading simultaneously, this paper considers a general hybrid NOMA scheme for MEC offloading, i.e., a user is also allowed to solely occupy a time slot and use OMA transmission for its offloading. The use of hybrid NOMA-MEC offloading was investigated for the special case of two users in [20], [21], where several insightful properties of MEC offloading were revealed. For example, the authors of [20], [21] showed that the use of hybrid NOMA offloading always yields an energy consumption no more than pure NOMA offloading, where all the users have to finish offloading concurrently. Furthermore, the authors of [20] provided a precise condition under which NOMA-MEC outperforms OMA-MEC. In particular, if one user's task deadline is

less than two times the other user's deadline, NOMA-MEC can outperform OMA-MEC. This condition is intuitive as explained in the following. If a user's deadline is not demanding, the user can afford to wait until the other user finishes offloading, and then use OMA for its own offloading, which avoids transmission in the presence of the strong interference caused by the other user. However, these conclusions were developed for the special case of two users, and it is not clear whether they are applicable for more general multi-user networks, which provides the motivation for this paper.

In particular, this paper focuses on a general multi-user MEC network, and the contributions of this paper are three-fold, as listed in the following:

- A general multi-user hybrid NOMA-MEC offloading strategy is first developed, where the offloading strategy designed in [20], [21] for two users can be viewed as a special case. In particular, for the proposed NOMA-MEC strategy, the users' offloading is scheduled according to the urgency of their tasks. Unlike for OMA-MEC, when a user with a more urgent task offloads, other users with less urgent tasks can still carry out offloading. Furthermore, unlike pure NOMA-MEC, which forces the users to start and finish their offloading simultaneously, the proposed strategy offers the users the opportunity to offload their tasks more flexibly, e.g., a user can wait until all the other users with more urgent tasks have finished their offloading. As a result, both pure NOMA-MEC and OMA-MEC are special cases of the proposed general hybrid offloading strategy.
- A multi-objective optimization problem is formulated in order to minimize the users' energy consumption for MEC offloading while meeting delay and transmit power constraints. We note that it is challenging to solve the formulated optimization problem because of its multi-objective and non-convex nature. Nevertheless, a low-complexity successive optimization algorithm is proposed by exploiting an important feature of the considered NOMA-MEC network, namely that the transmission strategy of a user with a less demanding task deadline has no impact on the data rate of a user with a more urgent task. By applying the proposed successive optimization algorithm, a closed-form expressions for the users' offloading power and time allocation parameters can be obtained.
- The properties of the proposed successive optimization algorithm are analyzed, which provides greater insight into NOMA-MEC. For example, the developed analytical results show that pure NOMA cannot outperform hybrid NOMA, which is consistent with the insight obtained for the two-user special case in [20], [21]. A precise condition for NOMA-

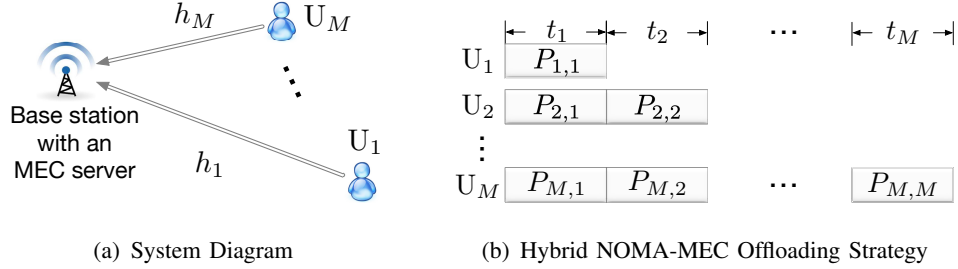


Fig. 1. Illustration of the considered NOMA-MEC system.

MEC to outperform OMA-MEC is also established, which reveals that the number of bits to be offloaded and the existing co-channel interference play important roles when deciding which of the two modes, namely OMA-MEC and NOMA-MEC, is preferable. This result is different from the conclusion reached for the two-user case in [20], [21]. However, it is worth noting that this new condition can be degraded to the one developed for the two-user case [20]. Another important insight obtained in this paper is that each user adopts a power allocation strategy similar to the well-known water-filling strategy [22]. Furthermore, it is proved that the proposed successive optimization algorithm realizes one of the Pareto optimal solutions for the considered multi-objective optimization problem, and the provided simulation results confirm the Pareto-optimality of the obtained solution.

II. SYSTEM MODEL

Consider a multi-user MEC network with M users, denoted by U_m , $1 \leq m \leq M$, and a base station equipped with an MEC server, as shown in Fig. 1(a). Each node is assumed to be equipped with a single antenna. Each user needs to compute a computationally-intensive and latency-critical task which is to be offloaded to the base station. Denote U_m 's deadline by D_m . Without loss of generality, assume that the users' deadlines are ordered as follows: $D_1 \leq \dots \leq D_M$.

A general hybrid NOMA-MEC offloading strategy is proposed in this paper, where the users are scheduled to complete their offloading according to the urgency of their tasks, as shown in Fig. 1(b). In particular, during the first t_1 seconds, U_1 is asked to complete its offloading, because its deadline is the most demanding one. In addition to U_1 , the other users are also allowed to carry out offloading during the first t_1 seconds. During the next t_2 seconds, U_2 is asked to complete its offloading, where the other users, U_m , $2 \leq m \leq M$, can continue offloading their tasks simultaneously. This offloading strategy continues in a successive manner according to the users' delay requirements, as shown in Fig. 1(b). During the last t_M seconds, only U_M is

served, because all other users should have already completed their offloading by then. Denote the users' transmit powers during t_n by $P_{m,n}$, $1 \leq n \leq m$, $1 \leq m \leq M$.

We note that both pure OMA and pure NOMA offloading are special cases of this hybrid NOMA offloading scheme. For example, by setting $t_m = 0$ for $m \geq 2$ (or $P_{m,n} = 0$ for $n \geq 2$ and $m \geq 2$), all users need to complete their offloading within the first t_1 seconds, which corresponds to a pure NOMA scheme. If $P_{m,n} = 0$ for $n < m$, i.e., only a single user is allowed to transmit at a time, the hybrid NOMA scheme is degraded to the pure OMA scheme. As shown in [20], for OMA offloading, for minimization of the offloading energy consumption, the optimal delay is given by $t_m^{\text{OMA}} = D_m - D_{m-1}$, and the optimal power allocation coefficient, denoted by P_m^{OMA} , is obtained by ensuring $(D_m - D_{m-1}) \log(1 + P_m^{\text{OMA}}|h_m|^2) = N$, where the users' channel coefficients are denoted by h_m and it is assumed that all users' tasks contain the same number of nats, denoted by N . Impairment by additive white Gaussian noise with normalized variance is assumed at all receivers.

By applying the proposed hybrid NOMA-MEC scheme, at t_n , the base station receives the following signal:

$$y_n = \sum_{m=n}^M \sqrt{P_{m,n}} h_m s_{m,n} + w_n, \quad (1)$$

where $s_{m,n}$ denotes the signal sent by U_m during t_n , and w_n denotes the white Gaussian noise. In order to ensure that U_n experiences the same performance as with OMA during t_n , the following successive interference cancellation (SIC) decoding order is used. In particular, at t_n , the base station decodes U_m 's signal first before decoding U_{m-1} 's signal, $m > n$, i.e., U_m 's signal is decoded in the $(M - m + 1)$ -th SIC step with the following offloading data rate:

$$R_{m,n} = \log \left(1 + \frac{P_{m,n}|h_m|^2}{\sum_{j=n}^{m-1} P_{j,n}|h_j|^2 + 1} \right). \quad (2)$$

As a result, at t_n , U_n 's signal is decoded last with the offloading data rate of $R_{n,n} = \log(1 + P_{n,n}|h_n|^2)$, which means that U_n experiences interference-free transmission as in OMA.

MEC offloading imposes two constraints. One is to ensure that U_m can offload all its N nats by the end of t_m , i.e.,

$$\sum_{n=1}^m t_n R_{m,n} \geq N, \quad (3)$$

and the other constraint is to meet the deadline of MEC offloading, i.e.,

$$\sum_{n=1}^m t_n \leq D_m. \quad (4)$$

Define $E_m \triangleq \sum_{n=1}^m P_{m,n} t_n$, which denotes U_m 's overall energy consumption. Furthermore, define $\mathbf{p}_m = [p_{m,1} \ \cdots \ p_{m,m}]^T$, which collects U_m 's power allocation coefficients, and $\mathbf{x} \triangleq [t_2 \ \cdots \ t_M \ \mathbf{p}_2^T \ \cdots \ \mathbf{p}_M^T]^T$, which collects the M users' time and power allocation coefficients, where \mathbf{A}^T denotes the transpose of \mathbf{A} . In this paper, the following multi-objective energy minimization problem is considered:

$$\min_{\mathbf{x}} \quad \mathbf{E}_M \triangleq [E_2 \ \cdots \ E_M]^T \quad (\text{P1a})$$

$$\text{s.t.} \quad \sum_{n=1}^m t_n R_{m,n} \geq N, \quad 2 \leq m \leq M \quad (\text{P1b})$$

$$\sum_{n=1}^m t_n \leq D_m, \quad t_m \geq 0, \quad 2 \leq m \leq M \quad (\text{P1c})$$

$$0 \leq P_{m,n} \leq P_m^{\text{OMA}}, \quad 2 \leq m \leq M, 1 \leq n \leq m, \quad (\text{P1d})$$

where P_m^{OMA} , a user's transmit power in OMA, is used in (P1d) in order to facilitate the performance comparison between OMA-MEC and NOMA-MEC. Another benefit to use P_m^{OMA} as the transmit power budget is that the optimal choices for U_1 's transmit parameters, t_1 and $P_{1,1}$, can be straightforwardly obtained as follows: $t_1^* = D_1$ and $P_{1,1}^* = P_1^{\text{OMA}}$, which is the reason why only U_m 's parameters, $m > 1$, are optimized in problem P1.

Remark 1: Problem P1 is a typical multi-objective optimization problem, where there are potential conflicts between the multiple users' objectives. For example, one user's choices for its power allocation coefficients may reduce its own offloading energy consumption, but could increase other users' energy consumption. For a multi-objective optimization problem, it is of interest to find Pareto optimal solutions which are defined as follows. Define \mathcal{E}_M as a set collecting all feasible \mathbf{E}_M . Denote a feasible solution by $\bar{\mathbf{x}}$, and the corresponding objective value by $\bar{\mathbf{E}}_M$. If $\bar{\mathbf{x}}$ is Pareto optimal, there is no other element in \mathcal{E}_M that dominates $\bar{\mathbf{E}}_M$, i.e., there is no $\mathbf{z} \in \mathcal{E}_M$ and $\mathbf{z} \neq \bar{\mathbf{E}}_M$ satisfying $\mathbf{z} \prec \bar{\mathbf{E}}_M$, where \prec denotes an element-wise inequality. In the other words, $\bar{\mathbf{E}}_M$ is a minimal element of \mathcal{E}_M [23].

Remark 2: One approach to find the Pareto optimal solutions of problem P1 is to apply the scalarization method and convert the multi-objective optimization problem to a single-objective optimization problem, e.g., using a weighted sum, $\sum_{m=1}^M w_m E_m$, as the objective function, where w_m denotes the weights [23]. Different Pareto-optimal solutions can be obtained by varying the weights, w_m . We note that if scalarization is employed, problem P1 is still challenging to solve due to the following two difficulties. First, $P_{m,n}$ and t_m are coupled in the objective function as

well as in the constraints. Second, $R_{m,n}$ in P1b is not a concave function of the power allocation coefficients. To avoid these obstacles, in the following section, a low-complexity algorithm for solving problem P1 will be introduced, and its Pareto-optimality will be analyzed.

III. A LOW-COMPLEXITY SUCCESSIVE RESOURCE ALLOCATION ALGORITHM

In this section, a low-complexity successive algorithm is first proposed to solve Problem P1, and then the properties of the solution obtained with the algorithm are revealed.

A. Description of Successive Resource Allocation

An important observation for the considered NOMA-MEC problem is that due to the use of SIC, U_n 's choices for t_n and $P_{n,i}$ have no impact on U_m 's data rate, $m < n$. An extreme example is U_1 's data rate which is $R_{1,1} = t_1 \log(1 + P_{1,1}|h_1|^2)$ and depends on t_1 and $P_{1,1}$ only. This motivates the use of a successive optimization strategy, where U_m 's transmission parameters are optimized after U_{m-1} 's. As a result, E_m can be minimized by focusing on optimizing U_m 's parameters only, because U_n 's parameters, $1 \leq n \leq m-1$, have already been optimized and U_i 's parameters, $i > m$, have no impact on U_m 's data rate.

Based on the above discussions, the proposed low-complexity algorithm decomposes Problem P1 into a series of subproblems and solves the subproblems in a successive manner, as shown in Algorithm 1, where $\mathbf{p}_m^* = [p_{m,1}^* \ \cdots \ p_{m,m}^*]^T$, and problem P2 is defined as follows:

$$\min_{\mathbf{p}_m, t_m} E_m \quad (\text{P2a})$$

$$\text{s.t.} \quad \sum_{n=1}^m t_n R_{m,n} \geq N \quad (\text{P2b})$$

$$0 \leq t_m \leq D_m - \sum_{n=1}^{m-1} t_n \quad (\text{P2c})$$

$$0 \leq P_{m,n} \leq P_m^{\text{OMA}}, \quad 2 \leq n \leq m. \quad (\text{P2d})$$

Because Algorithm 1 solves Problem P1 by decomposing it into a series of the subproblems shown in (P2), the solution obtained with Algorithm 1 is expected to be a suboptimal solution of Problem P1. However, surprisingly Algorithm 1 realizes one of the Pareto optimal solutions of Problem P1, as shown at the end of this section.

As discussed before, the successive feature of Algorithm 1 lies in the fact that problem P2 is optimized by assuming that U_n 's parameters, $1 \leq n \leq m$, have already been optimized. This

Algorithm 1 Low-Complexity Successive Optimization

- 1: Set $P_{1,1}^* = P_1^{\text{OMA}}, t_1^* = D_1$
 - 2: $m = 1$
 - 3: **while** $m < M$ **do**
 - 4: $m = m + 1$.
 - 5: Find the optimal solutions for $t_m^*, P_{m,n}^*, 1 \leq n \leq m$,
by solving problem P2
 - 6: **end**
 - 7: The outcome of the algorithm is given by

$$\mathbf{x}^* \triangleq [t_2^* \ \cdots \ t_M^* \ \mathbf{p}_2^{*T} \ \cdots \ \mathbf{p}_M^{*T}]^T$$
-

implies that t_n and $P_{n,i}$, $1 \leq n < m$, shown in problem P2 are constant and fixed¹. We note that Algorithm 1 is initialized with the choices, $P_{1,1}^* = P_1^{\text{OMA}}, t_1^* = D_1$, because P_1^{OMA} and D_1 are the optimal solutions for U_1 's transmission parameters.

We note that problem P2 is not convex due to the fact that t_m and $P_{m,n}$ are coupled in the objective and constraint functions. In order to reduce the number of the Lagrange multipliers to be used in the Karush–Kuhn–Tucker (KKT) conditions, the upper bound on $P_{m,n}$ is first removed, which means that the optimization problem considered in (P2) can be recast as follows:

$$\min_{\mathbf{p}_m, t_m} P_{m,m} t_m + \sum_{n=1}^{m-1} P_{m,n} t_n \quad (\text{P3a})$$

$$\text{s.t.} \quad (\text{P2b}), (\text{P2c})$$

$$0 \leq P_{m,n}, 1 \leq n \leq m. \quad (\text{P3b})$$

We note that problem P3 is still equivalent to problem P2, which will be formally proved after the optimal solution for problem P3 is obtained.

We further note that for the first $m - 1$ time slots, i.e., t_i , $1 \leq i \leq m - 1$, it is possible that U_m does not want to use all of them for offloading. Define \mathcal{S}_m as the set collecting the indices of the time slots, in which U_m decides to transmit, e.g., $\mathcal{S}_m = \{1, \dots, m - 1\}$ means that U_m transmits continuously during the first $m - 1$ time slots, and $\mathcal{S}_m = \emptyset$ means that U_m relies on t_m only. Define also $a_{m,n} = \frac{1}{\sum_{j=n}^{m-1} P_{j,n} |h_j|^2 + 1}$. Then, a closed-form solution of problem P3 is provided in the following lemma.

Lemma 1. *For the optimization problem in (P3), if $\min\{a_{m,1}, \dots, a_{m,m-1}\} \geq e^{-\frac{N - \sum_{k=1, k \in \mathcal{S}_m}^{m-1} t_k \log a_{m,k}}{t_m + \sum_{k=1, k \in \mathcal{S}_m}^{m-1} t_k}}$, the optimal solution for the power allocation is one of the following*

¹More rigorously, the notations, t_n^* and $P_{n,i}^*$, $1 \leq n < m$, should be used in problem P2, but the superscript, *, is omitted for notational simplicity.

two solutions. The first solution is based on pure OMA, i.e., $P_{m,n}^{O*} = 0$, for $1 \leq n \leq m-1$, and $P_{m,m}^{O*} = \frac{e^{\frac{N}{t_m}} - 1}{|h_m|^2}$. The second solution is based hybrid NOMA as shown in the following:

$$P_{m,n}^* = \begin{cases} \frac{e^{\frac{N - \sum_{k=1, k \in \mathcal{S}_m}^{m-1} t_k \log a_{m,k}}{t_m + \sum_{k=1, k \in \mathcal{S}_m}^{m-1} t_k} - 1}}{|h_m|^2}, & \text{if } n = m \\ \frac{a_{m,n} e^{\frac{N - \sum_{k=1, k \in \mathcal{S}_m}^{m-1} t_k \log a_{m,k}}{t_m + \sum_{k=1, k \in \mathcal{S}_m}^{m-1} t_k} - 1}}{a_{m,n} |h_m|^2}, & \text{if } n < m \text{ \& } n \in \mathcal{S}_m \\ 0, & \text{otherwise} \end{cases}. \quad (5)$$

Otherwise, the pure OMA power allocation solution is optimal. For both OMA and hybrid NOMA, the optimal choice of t_m is the same and given by $t_m^* = D_m - D_{m-1}$.

Proof. See Appendix A. □

Remark 3: The implementation of Lemma 1 requires the *a priori* knowledge of \mathcal{S}_m . Therefore, a straightforward way to use Lemma 1 is to find the potential choices of \mathcal{S}_m and compare their corresponding energy consumption by using the lemma. There is a more computationally efficient alternative, as explained in the following. Lemma 1 shows that the optimal solution for t_m is always $D_m - D_{m-1}$ and does not depend on $P_{m,n}$. By substituting t_m^* into problem P3, problem P3 becomes a convex problem with regard to $P_{m,n}$, and hence off-the-shelf convex optimization solvers can be straightforwardly used to find $P_{m,n}$ without *a priori* knowledge of \mathcal{S}_m .

Remark 4: Recall that the considered offloading scheme provides a general framework, and includes three possible power allocation strategies, namely pure NOMA, pure OMA, and hybrid NOMA. The proof of Lemma 1 shows that the pure NOMA strategy cannot outperform hybrid NOMA, and hence can be ignored. This observation is consistent with the results obtained for the two-user special case [20], [21].

Remark 5: While Lemma 1 provides closed-form expressions for the power allocation solutions for pure OMA and hybrid NOMA, it is not clear which of the two yields a better performance. For the two-user special case, it was shown in [20] that hybrid NOMA outperforms OMA if one user's task deadline is less than twice the other user's deadline. A similar condition for a more general setting will be established in the next subsection.

Remark 6: For the two-user case considered in [20], [21], the hybrid NOMA strategy is simple, i.e., one user chooses to transmit in one time slot and the other user transmits in two time slots. The hybrid NOMA strategy provided in Lemma 1 is more complicated. For example, U_M might choose to transmit at t_1 and t_M , and remain silent during the other time slots, i.e.,

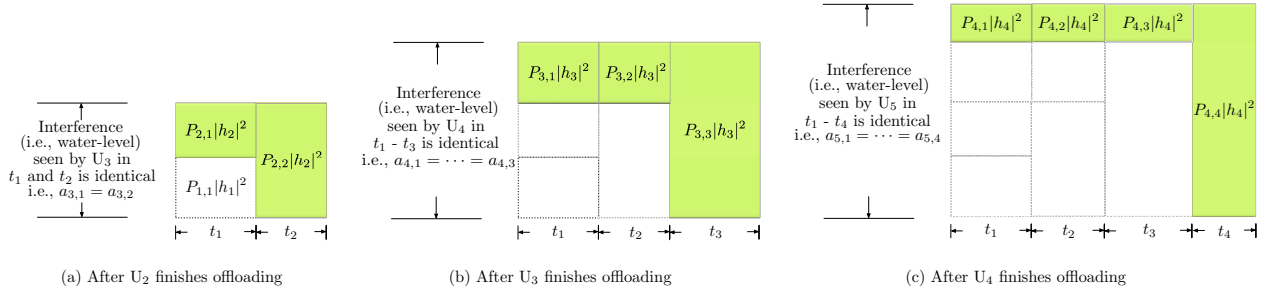


Fig. 2. Illustration of the water-filling like power allocation strategy indicated by Lemma 2.

$\mathcal{S}_M = \{1, M\}$. In the next subsection, Lemma 1 will be further simplified by proving that among all the hybrid NOMA strategies with different \mathcal{S}_m , the continuous transmission strategy, where $\mathcal{S}_m = \{1, \dots, m-1\}$, results in the lowest energy consumption.

B. Properties of Successive Resource Allocation

In this section, the properties of successive resource allocation are analyzed by focusing on the important case when $D_n - D_{n-1} \leq -\frac{N}{\log a_{n,1}}$, for $2 \leq n \leq M$, which is the feasibility condition for the hybrid NOMA solution with $\mathcal{S}_m = \{1, \dots, m-1\}$ as will be explained later. One reason to focus on this particular case is that it corresponds to the important time-critical situation, i.e., the duration left for U_n 's offloading in OMA, $D_n - D_{n-1}$, is small. The other reason is that more concise conclusions about NOMA-MEC offloading can be obtained.

Recall that $a_{m,n} = \frac{1}{\sum_{j=n}^{m-1} P_{j,n}|h_j|^2 + 1}$, which means that $a_{m,n}$ is an indicator for how much interference U_m suffers at t_n . The following lemma shows an important property of $a_{m,n}$.

Lemma 2. Assume that $D_n - D_{n-1} \leq -\frac{N}{\log a_{n,1}}$, for $2 \leq n \leq M$, and assume that the hybrid NOMA solution with $\mathcal{S}_m = \{1, \dots, m-1\}$ is adopted in each step of Algorithm 1. Then, the following equality holds

$$a_{m,i} = a_{m,l}, \quad (6)$$

for $i \neq l$, $1 \leq i \leq m-1$, and $1 \leq l \leq m-1$.

Proof. See Appendix B. □

Remark 7: Lemma 2 indicates that the hybrid NOMA power allocation solution is similar to the water-filling power allocation strategy [22]. In Fig. 2, a four-user example is used to illustrate this interesting property. U_2 's transmit powers in t_1 and t_2 are allocated carefully; such that the interference levels (or water levels) seen by U_3 at t_1 and t_2 are the same, i.e., $a_{3,1} = a_{3,2}$. Furthermore, compared to t_2 , U_2 experiences less interference in t_1 , and hence uses more power

in t_2 . This is also analogous to the principle of water-filling, i.e., more water (i.e., transmit power) is poured to a hole which is deeper (i.e., t_2). Similarly, U_3 chooses its transmit powers to ensure that the interference experienced by U_4 in t_1 , t_2 , and t_3 reaches the same levels, i.e., $a_{4,1} = a_{4,2} = a_{4,3}$.

Intuitively, if U_4 suffers the same amount of interference in each of the first three time slots, it should use the same transmit power during these three time slots. Indeed, by combining Lemmas 1 and 2, this intuition regarding the users' transmit powers can be straightforwardly proved as shown in the following corollary.

Corollary 1. *Assuming that $D_n - D_{n-1} \leq -\frac{N}{\log a_{n,1}}$, for $2 \leq n \leq M$, and assuming that the hybrid NOMA solution with $\mathcal{S}_m = \{1, \dots, m-1\}$ is adopted in each step of Algorithm 1, U_m 's transmit powers are identical for all t_i , $1 \leq i \leq m-1$, i.e., $P_{m,i} = P_{m,l}$, for $i \neq l$, $1 \leq i \leq m-1$, and $1 \leq l \leq m-1$.*

Remark 8: Lemma 2 can also be used to simplify the feasibility condition of hybrid NOMA shown in Lemma 1, i.e., $\min\{a_{m,1}, \dots, a_{m,m-1}\} \geq e^{-\frac{N - \sum_{k=1, k \in \mathcal{S}_m}^{m-1} t_k \log a_{m,k}}{t_m + \sum_{k=1, k \in \mathcal{S}_m}^{m-1} t_k}}$ can be simplified to $D_m - D_{m-1} \leq -\frac{N}{\log a_{m,1}}$, where $t_m^* = D_m - D_{m-1}$ is used. Furthermore, Lemma 2 can be used to simplify the optimal solution of problem P3 by establishing the following two facts. The first compares the OMA and hybrid NOMA solutions, as stated in the following lemma.

Lemma 3. *Assume that $D_n - D_{n-1} \leq -\frac{N}{\log a_{n,1}}$, for $2 \leq n \leq M$, and U_n , $2 \leq n \leq m-1$, chooses the hybrid NOMA solution with $\mathcal{S}_n = \{1, \dots, n-1\}$. U_m 's energy consumption in hybrid NOMA with $\mathcal{S}_m = \{1, \dots, m-1\}$ is no more than that of the pure OMA solution.*

Proof. See Appendix C. □

The second fact is that among all possible hybrid NOMA strategies, U_m can reduce its energy consumption by choosing to transmit continuously, i.e., by adopting the hybrid NOMA solution with $\mathcal{S}_m = \{1, \dots, m-1\}$, as shown in the following lemma.

Lemma 4. *Assume that $D_n - D_{n-1} \leq -\frac{N}{\log a_{n,1}}$, for $2 \leq n \leq M$, and assume that U_n , $2 \leq n \leq m-1$, chooses the hybrid NOMA solution with $\mathcal{S}_n = \{1, \dots, n-1\}$. Thus, among all of U_m 's hybrid NOMA power allocation solutions shown in (5), the one with $\mathcal{S}_m = \{1, \dots, m\}$ yields the lowest energy consumption.*

Proof. See Appendix D □

Lemmas 3 and 4 indicate that U_m prefers the hybrid NOMA solution with $\mathcal{S}_m = \{1, \dots, m-1\}$ if U_n , $n < m$, also chooses the same solution. By using a simple proof by mathematical induction and also the fact that U_2 chooses the hybrid NOMA solution whenever it is feasible which was established in [20], a more concise conclusion for the optimal solution of problem P3 can be straightforwardly obtained, as shown in the following corollary.

Corollary 2. Assume that $D_m - D_{m-1} \leq -\frac{N}{\log a_{m,1}}$, $2 \leq m \leq M$. For the optimization problem shown in (P3), the optimal power allocation strategy is given by

$$P_{m,n}^* = \begin{cases} \frac{e^{\frac{N-D_{m-1} \log a_{m,1}}{D_m}} - 1}{|h_m|^2}, & \text{if } n = m \\ \frac{a_{m,n} e^{\frac{N-D_{m-1} \log a_{m,1}}{D_m}} - 1}{a_{m,1} |h_m|^2}, & \text{if } n < m \end{cases}. \quad (7)$$

The optimal choice of t_m is given by $t_m^* = D_m - D_{m-1}$.

Remark 9: Compared to Lemma 1, Corollary 2 is much more concise and insightful. For example, Corollary 2 shows that there is a single optimal solution for problem P3, and the feasibility condition for this hybrid NOMA based optimal solution is simply $D_m - D_{m-1} \leq -\frac{N}{\log a_{m,1}}$. For the two-user case, it was shown in [20] that the condition to switch from hybrid NOMA to OMA is $D_2 \geq 2D_1$. We note that the condition shown in Corollary 2 is consistent with that established for the two-user case. In particular, for $M = 2$, $D_m - D_{m-1} \leq -\frac{N}{\log a_{m,1}}$ is equivalent to $D_2 \geq D_1 + \frac{N}{\log(1+P_{1,1}|h_1|^2)} = 2D_1$ because $D_1 \log(1+P_{1,1}|h_1|^2) = N$.

Furthermore, by using Lemma 2, the equivalence between problem P3 and problem P2 can be established, as shown in the following lemma.

Lemma 5. Assuming that $D_m - D_{m-1} \leq -\frac{N}{\log a_{m,1}}$, $2 \leq m \leq M$, the solution shown in (7) is the optimal solution of problem P2.

Proof. See Appendix E. □

Finally, the Pareto optimality of the obtained solution is shown in the following.

Lemma 6. Assuming that $D_m - D_{m-1} \leq -\frac{N}{\log a_{m,1}}$, $2 \leq m \leq M$, the solution shown in (7) is a Pareto optimal solution of the multi-objective optimization problem shown in (P1).

Proof. See Appendix F. □

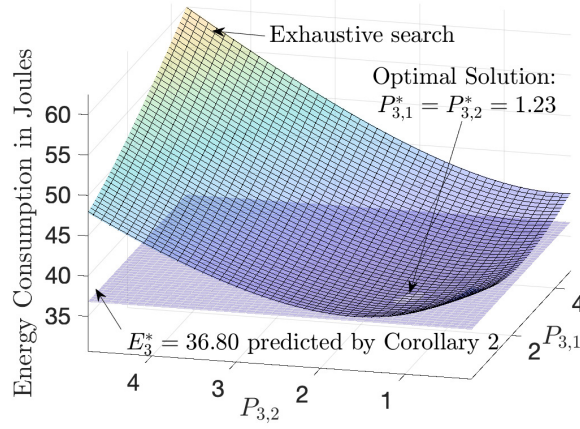


Fig. 3. Verification of the optimal solution shown in Corollary 2. $M = 3$ and $N = 10$, and the users' task deadlines are $D_1 = 8$, $D_2 = 12$, and $D_3 = 16$, respectively. For the exhaustive search, $P_{3,3}$ is obtained by using $P_{3,1}$ and $P_{3,2}$ to solve the following equation, $t_1 R_{3,1} + t_2 R_{3,2} + t_3 R_{3,3} = N$ and then checking whether the obtained $P_{3,3}$ satisfies the transmit power constraint in (P1d).

Remark 10: As shown in the proof for Lemma 6, the energy consumption vector realized by the proposed algorithm, denoted by $\mathbf{E}_M^* \triangleq [E_2^* \ \dots \ E_M^*]$, is a minimal element of all feasible \mathbf{E}_M [23]. In other words, there is no feasible \mathbf{E}_M which dominates \mathbf{E}_M^* , i.e., $\mathbf{E}_M \prec \mathbf{E}_M^*$. The Pareto optimality of the obtained solution will be investigated in detail in the next section by using the scalarization method.

IV. NUMERICAL STUDIES

In this section, the performance of the proposed NOMA-MEC offloading scheme is studied by using computer simulations.

A. Optimality of the Proposed Successive Optimization Algorithm

In Fig. 3, the optimal solution of problem P2 shown in Corollary 2 is verified by focusing on the three-user case and using an exhaustive search as a benchmark scheme. In particular, for the exhaustive search, U_2 's transmission parameters are fixed by using the outcome of the first step of Algorithm 1, and t_3 is also fixed as $t_3 = D_3 - D_2$. For illustration purposes, all users' channels are assumed to be normalized. As can be seen from Fig. 3, the minimum energy consumption predicted by Corollary 2 matches perfectly the exhaustive search result, which verifies the accuracy of Corollary 2.

TABLE I
WEIGHTED ENERGY CONSUMPTION $\sum_{m=1}^M w_m E_m$ IN JOULES

	$N = 8$	$N = 10$	$N = 12$	$N = 14$	$N = 16$
E_{ES}^w with \mathbf{w}_1	18.57	29.44	45.27	68.29	101.79
E_{SO}^w with \mathbf{w}_1	18.57	29.44	45.27	68.29	101.79
$E_{ES}^w - E_{SO}^w$	1.07×10^{-4}	1.55×10^{-4}	1.03×10^{-5}	2.63×10^{-10}	1.64×10^{-6}
E_{ES}^w with \mathbf{w}_2	19.66	31.59	49.19	75.11	113.21
E_{SO}^w with \mathbf{w}_2	19.98	32.26	50.48	77.47	117.38
$E_{ES}^w - E_{SO}^w$	-0.32	-0.67	-1.29	-2.36	-4.17

The verification for the Pareto optimality of the obtained solution is based on the scalarization method which solves the following optimization problem:

$$(P4): \min_{\mathbf{x}} \sum_{m=1}^M w_m E_m \quad \text{s.t.} \quad (P1b), (P1c), (P1d).$$

Recall that different Pareto optimal solutions can be obtained by solving problem (P4) for different weights [23]. Therefore, in order to show that the solution obtained by the proposed algorithm is Pareto optimal, it is sufficient to show that the proposed solution can realize the same performance as the scalarization method with a particular set of weights. For Table I, two sets of the weights are used, namely $\mathbf{w}_1 = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$ and $\mathbf{w}_2 = \frac{1}{6} \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$, where E_{SO}^w and E_{ES}^w denote the weighted energy consumption achieved by the proposed algorithm and the scalarization method, respectively. Because problem (P4) is non-convex, an exhaustive search is used to solve this problem, where a step size of 0.1 is used. The simulation parameters used to generate Table I are the same as the ones used for Fig. 3. As can be seen from Table I, the proposed algorithm results in the same performance as the scalarization method with \mathbf{w}_1 , which confirms the Pareto-optimality of the proposed algorithm provided in Lemma 5. The fact that the proposed algorithm can realize a Pareto-optimal solution is surprising, but might be due to the following reason. The proposed algorithm is similar to a greedy algorithm, where each user's transmission strategy is obtained separately by solving problem P2. As a result, the outcome of the proposed algorithm might be one of the stationary points of the considered problem, where no user has an incentive to deviate from its chosen strategy. The fact that $E_{ES}^w - E_{SO}^w$ is a small positive number instead of zero is explained in the following. For an optimization problem with continuous variables, the exhaustive search based scheme can only approximate the optimal solution. In other words, increasing the resolution of an exhaustive search reduces the gap between the outcome of the search and the optimal performance, but this gap cannot

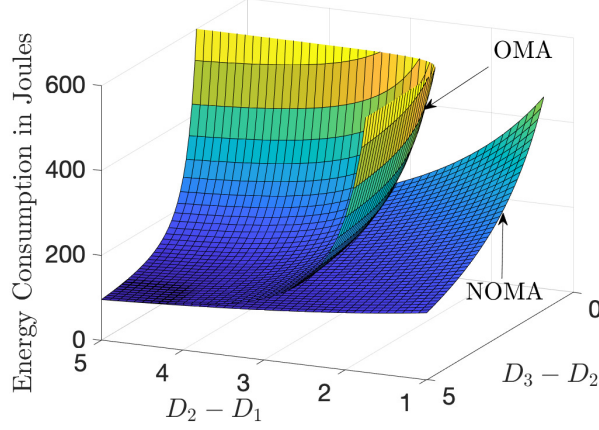


Fig. 4. Impact of the users' task deadlines on the energy consumed by MEC offloading with $M = 3$, $D_1 = 5$, and $N = 10$.

go to zero. Another important observation from Table I is that, for w_2 , the solution obtained by the proposed algorithm results in a higher energy consumption than the exhaustive search. This observation is also expected since the proposed algorithm finds only one of the many Pareto optimal solutions.

B. Performance of the Proposed Successive Optimization Algorithm

In this section, the performance of the proposed algorithm is studied by using OMA-MEC as a benchmark scheme, where the impact of different choices of the system parameters on MEC offloading is investigated. In Fig. 4, the overall energy consumption is shown as a function of the users' task deadlines, where $D_2 - D_1$ and $D_3 - D_2$ are used as the $x - y$ coordinates because they are the optimal choices for t_2 and t_3 , respectively. As can be observed from Fig. 4, the use of hybrid NOMA-MEC can result in a significant reduction in energy consumption, compared to OMA-MEC. Furthermore, an interesting observation from Fig. 4 is that hybrid NOMA-MEC and OMA-MEC achieve the same performance for large $D_m - D_{m-1}$, which confirms Corollary 2, as for large $D_m - D_{m-1}$ the NOMA feasibility condition cannot be satisfied and hence hybrid NOMA-MEC degrades to OMA-MEC.

In Fig. 5, the impact of N on the energy consumption of MEC offloading is studied. As can be observed from the figure, the energy consumed by MEC offloading increases with N , which is due to the fact that a larger N means that more bits need to be offloaded and hence more energy needs to be consumed. For the considered range of N , the use of hybrid NOMA-MEC offers a reduction in energy consumption compared to OMA-MEC. Furthermore, Fig. 5 shows that the performance gain of NOMA-MEC over OMA-MEC is significantly larger for larger N , i.e., when the users' tasks are large, it is more beneficial to use hybrid NOMA-MEC.

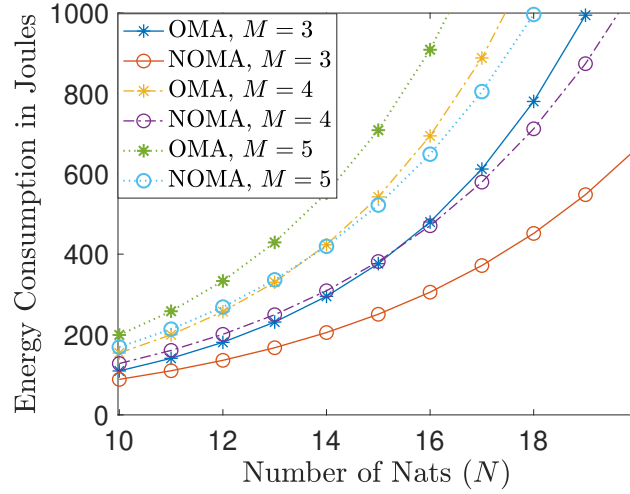


Fig. 5. Impact of the size of the tasks on the energy consumed by MEC offloading. $D_1 = 8$, and $D_m - D_{m-1} = \frac{D_1}{2}$.

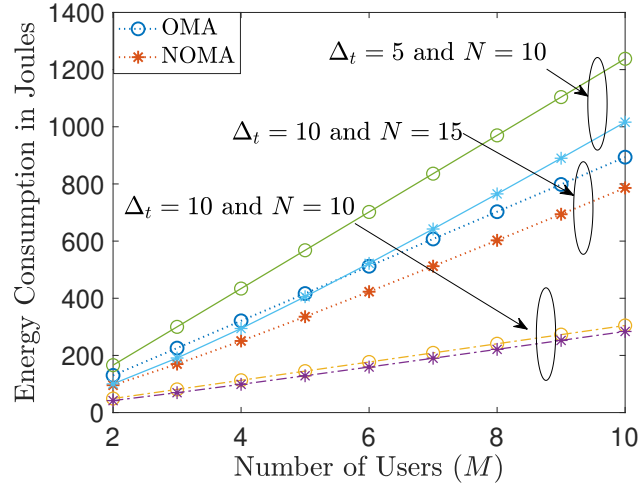


Fig. 6. Impact of the number of users on the energy consumed by MEC offloading. $D_1 = \Delta_t$, and $D_m - D_{m-1} = \frac{\Delta_t}{2}$.

In Fig. 6, the performance of MEC offloading is shown as a function of the number of users. In particular, the figure shows that increasing the number of users increases the overall energy consumption for both hybrid NOMA-MEC and OMA-MEC. Fig. 6 also shows that the two schemes require a higher energy consumption for MEC offloading, if a smaller $D_m - D_{m-1}$ is used, which can be explained by using OMA-MEC as an example. Recall that for OMA-MEC, U_m solely relies on t_m for offloading, and the optimal choice of t_m is $D_m - D_{m-1}$, which means that a smaller $D_m - D_{m-1}$ leads to less time available for offloading and hence more offloading energy is required. Fig. 6 shows that the performance gain of NOMA-MEC over OMA-MEC is larger if $D_m - D_{m-1}$ is smaller. This is expected since for OMA-MEC, U_m relies on $D_m - D_{m-1}$ (or t_m) for offloading, and a reduction of $D_m - D_{m-1}$ can cause a significant increase of energy consumption. For hybrid NOMA-MEC, U_m uses not only t_m but also t_n , $n < m$, for offloading,

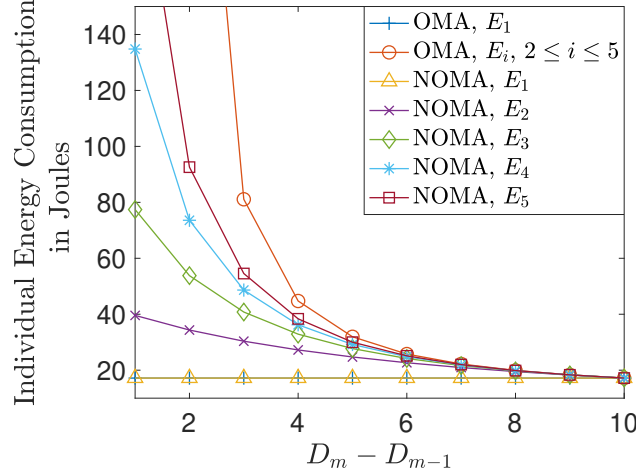


Fig. 7. Illustration of the condition under which hybrid NOMA-MEC and OMA-MEC achieve the same performance. $M = 5$, $N = 10$, and $D_1 = 5$. The differences between D_{m-1} and D_m are identical for $2 \leq m \leq M$.

TABLE II
THE VALUES OF $-\frac{N}{\log a_{m,1}}$ FOR THE CURVES SHOWN IN FIG. 7

$D_m - D_{m-1}$	1	2	3	4	5	6	7	8	9	10
$m = 2$	5.50	6.00	6.50	7.00	7.50	8.00	8.50	9.00	9.50	10.00
$m = 3$	4.00	4.67	5.33	6.00	6.67	7.33	8.00	8.67	9.33	10.00
$m = 4$	3.25	4.00	4.75	5.50	6.25	7.00	7.75	8.50	9.25	10.00
$m = 5$	2.80	3.60	4.40	5.20	6.00	6.80	7.60	8.40	9.20	10.00

which makes hybrid NOMA-MEC less sensitive to the choice of $D_m - D_{m-1}$.

Finally, the conditions under which hybrid NOMA-MEC and OMA-MEC yield the same performance are studied in Fig. 7, where the individual energy consumptions, E_m , are shown as functions of $D_m - D_{m-1}$. Recall that there are two possible optimal solutions for problem P2, one based on hybrid NOMA and one based on pure OMA. Remark 8 and Corollary 2 show that the feasibility of hybrid NOMA depends on the value of $D_m - D_{m-1}$, i.e., $D_n - D_{n-1} \leq -\frac{N}{\log a_{n,1}}$. Therefore, a large value of $D_m - D_{m-1}$ can cause the situation, where the hybrid NOMA solution is infeasible and the OMA solution is used. Fig. 7 and Table II confirm this conclusion, where increasing $D_m - D_{m-1}$ eventually leads to the situation, where hybrid NOMA-MEC is degraded to OMA-MEC. U_1 's energy consumptions in the OMA and NOMA modes are the same, and E_1 is not a function of $D_m - D_{m-1}$ since $P_{1,1}^*$ and t_1^* are not functions of the other users' parameters. Among the other four users, U_i 's energy consumption is less than U_j 's, for $i < j$, since U_j suffers from more interference as illustrated in Fig. 2.

V. CONCLUSIONS

In this paper, a general hybrid NOMA-MEC offloading strategy has been proposed, and a multi-objective optimization problem has been formulated for minimization of the users' energy

consumption for MEC offloading, where a low-complexity Pareto-optimal resource allocation solution has been obtained. Furthermore, by analyzing the properties of the obtained resource allocation solution, important insights regarding NOMA-MEC offloading have been obtained.

In this paper, an exhaustive search has been used for the scalarization method to solve the weighted-sum optimization problem, whereas an important direction for future research is to develop a low complexity alternative by applying advanced optimization tools, such as monotonic optimization. In addition, the SIC decoding order used in this paper was determined based on the users' task deadlines, but more sophisticated criteria for the SIC decoding order can be used to further improve the performance of MEC offloading.

APPENDIX A

PROOF FOR LEMMA 1

The lemma can be proved based on the following two steps. The first step is to fix t_m and find the closed-form expressions for the optimal power allocation solution. The second step is to prove that $t_m^* = D_m - D_{m-1}$ for all possible power allocations.

A. Finding the Power Allocations

The Lagrangian of problem P3 can be expressed as follows:

$$L = P_{m,m}t_m + \sum_{n=1}^{m-1} P_{m,n}t_n - \sum_{n=1}^m \lambda_n P_{m,n} + \lambda_0 \left(N - t_m R_{m,m} - \sum_{n=1}^{m-1} t_n R_{m,n} \right), \quad (8)$$

where the λ_i 's, $0 \leq i \leq m$, denote the Lagrange multipliers.

For a fixed t_m , it is straightforward to show that problem P3 is convex, and hence the optimal power allocation solutions can be obtained from solving the following KKT conditions:

$$\begin{cases} t_n - \lambda_0 t_n \frac{a_{m,n}|h_m|^2}{1+a_{m,n}|h_m|^2 P_{m,n}} - \lambda_n = 0, & 1 \leq n \leq m-1 \\ t_m - \lambda_0 t_m \frac{|h_m|^2}{1+|h_m|^2 P_{m,m}} - \lambda_m = 0 \\ \lambda_0 (t_m R_{m,m} + \sum_{n=1}^{m-1} t_n R_{m,n} - N) = 0 \\ \lambda_n P_{m,n} = 0, & 1 \leq n \leq m, \quad (\text{P2b}), (\text{P2c}), \& (\text{P3b}) \end{cases},$$

where the first two equations are obtained by using the Lagrangian shown in (8). The three considered types of transmission modes, namely pure OMA, pure NOMA, and hybrid NOMA, can be obtained as follows.

1) *Pure OMA*: Assume that $\lambda_n \neq 0$, $1 \leq n \leq m-1$, which means that $P_{m,n} = 0$, $1 \leq n \leq m-1$, and $\lambda_0 \neq 0$. Hence, $t_m R_{m,m} = N$, which yields the OMA solution shown in the lemma.

2) *Hybrid NOMA*: Recall that \mathcal{S}_m includes the time slots during which U_m transmits. Therefore, for a hybrid NOMA solution, given \mathcal{S}_m , $\lambda_n = 0$ for $n \in \mathcal{S}_m$. Because $\lambda_n = 0$ for $n \in \mathcal{S}_m$, $\lambda_0 \neq 0$, which means that U_m definitely transmits at t_m . Because $\lambda_0 \neq 0$, $t_m R_{m,m} + \sum_{n=1, n \in \mathcal{S}_m}^{m-1} t_n R_{m,n} - N = 0$. Therefore, the KKT conditions can be simplified as follows:

$$\begin{cases} 1 - \lambda_0 \frac{a_{m,n}|h_m|^2}{1+a_{m,n}|h_m|^2 P_{m,n}} = 0, & n \in \mathcal{S}_m \\ 1 - \lambda_0 \frac{|h_m|^2}{1+|h_m|^2 P_{m,m}} = 0 \\ t_m \log(1 + P_{m,m}|h_m|^2) + \sum_{n=1, n \in \mathcal{S}_m}^{m-1} t_n \log(1 + a_{m,n} P_{m,n} |h_m|^2) = N \end{cases}, \quad (9)$$

which can be further simplified as follows:

$$\begin{cases} \frac{a_{m,n}}{1+a_{m,n}|h_m|^2 P_{m,n}} = \frac{1}{1+|h_m|^2 P_{m,m}}, & n \in \mathcal{S}_m \\ t_m \log(1 + P_{m,m}|h_m|^2) + \sum_{n=1, n \in \mathcal{S}_m}^{m-1} t_n \log(1 + a_{m,n} P_{m,n} |h_m|^2) = N \end{cases}. \quad (10)$$

Defining $y_n = \log(1 + a_{m,n}|h_m|^2 P_{m,n})$ for $n \in \mathcal{S}_m$ and $y_m = \log(1 + |h_m|^2 P_{m,m})$, the KKT conditions can be simplified as follows:

$$\begin{cases} \log a_{m,n} + y_m = y_n, & n \in \mathcal{S}_m \\ t_m y_m + \sum_{n=1, n \in \mathcal{S}_m}^{m-1} t_n y_n = N \end{cases},$$

which yields the following solutions:

$$y_m = \frac{N - \sum_{k=1, k \in \mathcal{S}_m}^{m-1} t_k \log a_{m,k}}{t_m + \sum_{k=1, k \in \mathcal{S}_m}^{m-1} t_k},$$

and

$$y_n = \log a_{m,n} + \frac{N - \sum_{k=1, k \in \mathcal{S}_m}^{m-1} t_k \log a_{m,k}}{t_m + \sum_{k=1, k \in \mathcal{S}_m}^{m-1} t_k},$$

for $n \in \mathcal{S}_m$. By using y_m and y_n , $P_{m,m}^*$ and $P_{m,n}^*$ can be obtained as shown in the lemma, where the condition shown in the lemma, $a_{m,n} \geq e^{-\frac{N - \sum_{k=1, k \in \mathcal{S}_m}^{m-1} t_k \log a_{m,k}}{t_m + \sum_{k=1, k \in \mathcal{S}_m}^{m-1} t_k}}$, is required to ensure that $P_{m,n}^* \geq 0$, $1 \leq n \leq m-1$. We note that $P_{m,n}^* \geq 0$ is a sufficient condition to ensure $P_{m,m}^* \geq 0$ because $a_{m,1} \leq 1$.

3) *Pure NOMA*: The pure NOMA solution, i.e., $P_{m,m} = 0$, can be discarded by treating it as a special case of hybrid NOMA and showing that it results in higher energy consumption than hybrid NOMA. Without loss of generality, we focus on the hybrid NOMA solution with $\mathcal{S}_m = \{1, \dots, m-1\}$. The corresponding pure NOMA power allocation strategy can be obtained by letting $t_m \rightarrow 0$.

By using the hybrid NOMA solution shown in the lemma, the total energy consumption is given by

$$E_H = t_m \frac{e^{\frac{N - \sum_{k=1}^{m-1} t_k \log a_{m,k}}{t_m + \sum_{k=1}^{m-1} t_k}} - 1}{|h_m|^2} + \sum_{n=1}^{m-1} t_n \frac{a_{m,n} e^{\frac{N - \sum_{k=1}^{m-1} t_k \log a_{m,k}}{t_m + \sum_{k=1}^{m-1} t_k}} - 1}{a_{m,n} |h_m|^2}, \quad (11)$$

where the energy consumption of pure NOMA is also a special case of E_H by letting $t_m \rightarrow 0$.

By treating E_H as a function of t_m , i.e., $E_H(t_m)$, the superiority of hybrid NOMA over pure NOMA can be proved by showing $E_H(t_m) \leq E_H(0)$, as done in the following. The first order derivative of E_H with respect to t_m is given by

$$\begin{aligned} E'_H = & \frac{e^{\frac{N - \sum_{k=1}^{m-1} t_k \log a_{m,k}}{t_m + \sum_{k=1}^{m-1} t_k}} - 1}{|h_m|^2} - \frac{t_m}{|h_m|^2} e^{\frac{N - \sum_{k=1}^{m-1} t_k \log a_{m,k}}{t_m + \sum_{k=1}^{m-1} t_k}} \frac{N - \sum_{k=1}^{m-1} t_k \log a_{m,k}}{(t_m + \sum_{k=1}^{m-1} t_k)^2} \\ & - \sum_{n=1}^{m-1} \frac{a_{m,n} t_n}{a_{m,n} |h_m|^2} e^{\frac{N - \sum_{k=1}^{m-1} t_k \log a_{m,k}}{t_m + \sum_{k=1}^{m-1} t_k}} \frac{N - \sum_{k=1}^{m-1} t_k \log a_{m,k}}{(t_m + \sum_{k=1}^{m-1} t_k)^2}. \end{aligned} \quad (12)$$

By defining $\theta = N - \sum_{k=1}^{m-1} t_k \log a_{m,k}$ and $\tau_m = \sum_{k=1}^m t_k$, the expression for the energy consumption can be simplified as follows:

$$\begin{aligned} |h_m|^2 E'_H = & e^{\frac{\theta}{t_m + \tau_{m-1}}} - 1 - t_m e^{\frac{\theta}{t_m + \tau_{m-1}}} \frac{\theta}{(t_m + \tau_{m-1})^2} - \sum_{n=1}^{m-1} t_n e^{\frac{\theta}{t_m + \tau_{m-1}}} \frac{\theta}{(t_m + \tau_{m-1})^2} \\ = & e^{\frac{\theta}{t_m + \tau_{m-1}}} - 1 - e^{\frac{\theta}{t_m + \tau_{m-1}}} \frac{\theta}{t_m + \tau_{m-1}} = g\left(\frac{\theta}{t_m + \tau_{m-1}}\right), \end{aligned}$$

where $g(x) \triangleq e^x - 1 - xe^x$. In [20], it has been proved that $g(x)$ is a monotonically non-increasing function of x for $x \geq 0$, which means that $|h_m|^2 E'_H \leq g(0) = 0$. Therefore, the energy consumed by hybrid NOMA is a monotonically non-increasing function of t_m , i.e., $E_H(t_m) \leq E_H(0)$. Therefore, pure NOMA requires energy consumption no less than hybrid NOMA, and hence, the pure NOMA solution can be discarded.

B. Finding the Optimal Value for t_m

By using the closed-form expressions for the power allocation, problem P3 can be recasted as follows:

$$\min_{t_m} \quad P_{m,m}^* t_m + \sum_{n=1}^{m-1} P_{m,n}^* t_n \quad (P5a)$$

$$\text{s.t.} \quad 0 \leq t_m \leq D_m - \sum_{n=1}^{m-1} t_n. \quad (P5b)$$

In problem P5, either the power allocation strategy for pure OMA or that for hybrid NOMA can be used. For the case of hybrid NOMA, in the previous section, it was proved that the energy consumption for hybrid NOMA, E_H , is a monotonically non-increasing function of t_m , which means that the optimal choice of t_m is $t_m^* = D_m - \sum_{n=1}^{m-1} t_n$.

For the OMA case, the total energy consumption is given by

$$E_O = t_m \frac{e^{\frac{N}{t_m}} - 1}{|h_m|^2}. \quad (13)$$

The first order derivative of E_O with respect to t_m is given by

$$E'_O = \frac{e^{\frac{N}{t_m}} - 1 - \frac{N}{t_m} e^{\frac{N}{t_m}}}{|h_m|^2} = \frac{g\left(\frac{N}{t_m}\right)}{|h_m|^2}. \quad (14)$$

Exploiting again the fact that $g(x)$ is a monotonically non-increasing function of x , the energy consumed by pure OMA is a monotonically non-increasing function of t_m , and hence $t_m^* = D_m - \sum_{n=1}^{m-1} t_n$, which is the same as for hybrid NOMA.

Finally, because of the successive nature of Algorithm 1, i.e., $t_n^* = D_n - D_{n-1}$, for $2 \leq n < m$, and hence $t_m^* = D_m - D_{m-1}$. The proof of the lemma is complete.

APPENDIX B

PROOF FOR LEMMA 2

Recall that $a_{m,i}$ is defined as follows:

$$a_{m,i} = \frac{1}{\sum_{j=i}^{m-1} P_{j,i} |h_j|^2 + 1}. \quad (15)$$

By using the hybrid NOMA power allocation solutions, $a_{m,i}$ can be first expressed as follows:

$$\begin{aligned} \frac{1}{a_{m,i}} &= \sum_{j=i}^{m-1} P_{j,i} |h_j|^2 + 1 = P_{i,i} |h_i|^2 + \sum_{j=i+1}^{m-1} P_{j,i} |h_j|^2 + 1 \\ &= e^{\frac{N - \sum_{k=1}^{i-1} t_k \log a_{i,k}}{t_i + \sum_{k=1}^{i-1} t_k}} + \sum_{j=i+1}^{m-1} \frac{a_{j,i} e^{\frac{N - \sum_{k=1}^{j-1} t_k \log a_{j,k}}{t_j + \sum_{k=1}^{j-1} t_k}} - 1}{a_{j,i}}, \end{aligned} \quad (16)$$

for $1 \leq i \leq m-2$, and

$$\frac{1}{a_{m,m-1}} = e^{\frac{N - \sum_{k=1}^{m-2} t_k \log a_{m-1,k}}{t_{m-1} + \sum_{k=1}^{m-2} t_k}}. \quad (17)$$

The lemma can be proved by mathematical induction. The smallest value for m in the lemma is $m = 3$. Therefore, the proof is divided into two parts. The first part is about the base case, where $\frac{1}{a_{3,2}} - 1 = \frac{1}{a_{3,1}} - 1$ is proved. The second part is for the inductive step, where, by assuming that the lemma holds for $m-1$, the lemma for the case of m is proved.

A. The Base Case $m = 3$

In this part, we aim to prove the equivalence of $a_{3,1}$ and $a_{3,2}$. For the case of $m = 3$, (17) can be expressed as follows:

$$\frac{1}{a_{3,2}} = e^{\frac{N-t_1 \log a_{2,1}}{t_2+t_1}} \quad (18)$$

On the other hand, $a_{3,1}$ can be written as follows:

$$\begin{aligned} \frac{1}{a_{3,1}} &= e^{\frac{N}{t_1}} + \frac{a_{2,1} e^{\frac{N-t_1 \log a_{2,1}}{t_2+t_1}} - 1}{a_{2,1}} \\ &= e^{\frac{N}{t_1}} + e^{\frac{N-t_1 \log a_{2,1}}{t_2+t_1}} - \frac{1}{a_{2,1}}. \end{aligned} \quad (19)$$

Both the expressions for $a_{3,1}$ and $a_{3,2}$ are related to $a_{2,1}$. Recall that $a_{2,1} = \frac{1}{P_{1,1}|h_1|^2+1}$, where $P_{1,1}$ is the transmit power used by U_1 at t_1 . Because of the constraint $t_1 \log(1 + P_{1,1}|h_1|^2) = N$, the following equality holds:

$$e^{\frac{N}{t_1}} = \frac{1}{a_{2,1}}, \quad (20)$$

which means that (19) can be simplified as follows:

$$\frac{1}{a_{3,1}} = e^{\frac{N-t_1 \log a_{2,1}}{t_2+t_1}}. \quad (21)$$

Combining (18) and (21), one can conclude that the lemma holds for the case of $m = 3$, i.e.,

$$a_{3,1} = a_{3,2}. \quad (22)$$

B. Inductive Step

Assume now that the lemma holds for the cases of i , $3 \leq i \leq m-1$, i.e.,

$$a_{i,1} = \dots = a_{i,i-1}, \quad 3 \leq i \leq m-1. \quad (23)$$

The aim of this part is to show that the lemma also holds for the case of m , i.e.,

$$a_{m,1} = \dots = a_{m,m-1}. \quad (24)$$

First rewrite $a_{m,i}$, $1 \leq i \leq m-2$, shown in (16) as follows:

$$\frac{1}{a_{m,i}} = e^{\frac{N-\sum_{k=1}^{m-2} t_k \log a_{m-1,k}}{t_{m-1}+\sum_{k=1}^{m-2} t_k}} + \sum_{j=i}^{m-2} e^{\frac{N-\sum_{k=1}^{j-1} t_k \log a_{j,k}}{t_j+\sum_{k=1}^{j-1} t_k}} - \sum_{j=i+1}^{m-1} \frac{1}{a_{j,i}}. \quad (25)$$

Comparing (25) to (17), one can observe that proving (24) is equivalent to proving the following:

$$\check{\Delta} \triangleq \sum_{j=i}^{m-2} e^{\frac{N - \sum_{k=1}^{j-1} t_k \log a_{j,k}}{t_j + \sum_{k=1}^{j-1} t_k}} - \sum_{j=i+1}^{m-1} \frac{1}{a_{j,i}} = 0, \quad (26)$$

for any $1 \leq i \leq m-2$. $\check{\Delta}$ can be further rewritten as follows:

$$\begin{aligned} \check{\Delta} &= \sum_{j=i}^{m-2} \left(e^{\frac{N - \sum_{k=1}^{j-1} t_k \log a_{j,k}}{t_j + \sum_{k=1}^{j-1} t_k}} - \frac{1}{a_{j+1,i}} \right) \\ &= \sum_{j=i}^{m-2} \left(e^{\frac{N - \left(\sum_{k=1}^{j-1} t_k \right) \log a_{j,1}}{t_j + \sum_{k=1}^{j-1} t_k}} - \frac{1}{a_{j+1,1}} \right), \end{aligned} \quad (27)$$

where the last step follows from the assumption in (23).

Define the difference between the two terms in the bracket of (27) as Δ , which can be expressed as follows:

$$\begin{aligned} \Delta &= e^{\frac{N - \left(\sum_{k=1}^{j-1} t_k \right) \log a_{j,1}}{t_j + \sum_{k=1}^{j-1} t_k}} - \frac{1}{a_{j+1,1}} \\ &= e^{\frac{N - \left(\sum_{k=1}^{j-1} t_k \right) \log a_{j,1}}{t_j + \sum_{k=1}^{j-1} t_k}} - e^{-\log a_{j+1,1}}, \end{aligned} \quad (28)$$

which means that $\Delta = 0$ if the difference of the exponents of the two exponential functions in (28), denoted by $\bar{\Delta}$, is zero. We note that $\bar{\Delta}$ can be expressed as follows:

$$\begin{aligned} \bar{\Delta} &= \frac{N - \left(\sum_{k=1}^{j-1} t_k \right) \log a_{j,1}}{t_j + \sum_{k=1}^{j-1} t_k} + \log a_{j+1,1} \\ &= \frac{N - \left(\sum_{k=1}^{j-1} t_k \right) \log a_{j,1} + \left(t_j + \sum_{k=1}^{j-1} t_k \right) \log a_{j+1,1}}{t_j + \sum_{k=1}^{j-1} t_k}. \end{aligned} \quad (29)$$

Recall that $a_{m,n}$ can also be related to the users' data rates. Therefore, $\bar{\Delta}$ can be further expressed as follows:

$$\begin{aligned} \bar{\Delta} &= \frac{N + t_j \log a_{j+1,1} + \sum_{k=1}^{j-1} t_k (\log a_{j+1,1} - \log a_{j,1})}{t_j + \sum_{k=1}^{j-1} t_k} \\ &= \frac{N + t_j \log a_{j+1,j} + \sum_{k=1}^{j-1} t_k (\log a_{j+1,k} - \log a_{j,k})}{t_j + \sum_{k=1}^{j-1} t_k}, \end{aligned}$$

where the last step follows from the assumption in (23).

On the one hand, $\log a_{j+1,j}$ can be related to U_j 's offloading data rate at t_j as follows:

$$\log a_{j+1,j} = -\log (P_{j,j} |h_j|^2 + 1) = -R_{j,j}. \quad (30)$$

On the other hand, $\log a_{j+1,k} - \log a_{j,k}$ can be related to U_j 's offloading data rate at t_k as follows:

$$\begin{aligned} \log a_{j+1,k} - \log a_{j,k} &= -\log \frac{\sum_{p=k}^j P_{p,k} |h_p|^2 + 1}{\sum_{p=k}^{j-1} P_{p,k} |h_p|^2 + 1} \\ &= -\log \left(1 + \frac{P_{j,k} |h_j|^2}{\sum_{p=k}^{j-1} P_{p,k} |h_p|^2 + 1} \right) = -R_{j,k}. \end{aligned} \quad (31)$$

So the difference $\bar{\Delta}$ can be finally expressed as follows:

$$\bar{\Delta} = \frac{N - R_{j,j} - \sum_{k=1}^{j-1} R_{j,k}}{t_i + \sum_{k=1}^{j-1} t_k} = 0, \quad (32)$$

where the last step follows from the fact that, with the used hybrid NOMA power allocation, U_j can successfully finish its offloading. Because $\bar{\Delta} = 0$, $\check{\Delta} = 0$, and hence, $a_{m,1} = \dots = a_{m,m-1}$, i.e., the lemma holds for the case of m .

By combining the conclusions from the base case and the inductive step, the lemma is proved.

APPENDIX C

PROOF FOR LEMMA 3

Recall that the OMA power allocation solution is $P_{m,m}^{O*} = \frac{e^{\frac{N}{t_m}} - 1}{|h_m|^2}$, and hence the energy consumed by OMA offloading is given by $E_O = t_m \frac{e^{\frac{N}{t_m}} - 1}{|h_m|^2}$. The difference between the energy consumed by hybrid NOMA and OMA, denoted by $\tilde{\Delta} \triangleq |h_m|^2(E_H - E_O)$, is given by

$$\tilde{\Delta} = t_m e^{\frac{N - \sum_{k=1}^{m-1} t_k \log a_{m,k}}{t_m + \sum_{k=1}^{m-1} t_k}} + \sum_{n=1}^{m-1} t_n \frac{a_{m,n} e^{\frac{N - \sum_{k=1}^{m-1} t_k \log a_{m,k}}{t_m + \sum_{k=1}^{m-1} t_k}} - 1}{a_{m,n}} - t_m e^{\frac{N}{t_m}}. \quad (33)$$

The lemma is proved by showing that $\tilde{\Delta} \leq 0$ for $a_{m,n} e^{\frac{N - \sum_{k=1}^{m-1} t_k \log a_{m,k}}{t_m + \sum_{k=1}^{m-1} t_k}} - 1 \geq 0$.

The difference $\tilde{\Delta}$ can be further rewritten as follows:

$$\begin{aligned} \tilde{\Delta} &= t_m e^{\frac{N - \sum_{k=1}^{m-1} t_k \log a_{m,k}}{t_m + \sum_{k=1}^{m-1} t_k}} + \sum_{n=1}^{m-1} t_n e^{\frac{N - \sum_{k=1}^{m-1} t_k \log a_{m,k}}{t_m + \sum_{k=1}^{m-1} t_k}} - \sum_{n=1}^{m-1} t_n \frac{1}{a_{m,n}} - t_m e^{\frac{N}{t_m}} \\ &= \sum_{n=1}^m t_n e^{\frac{N - \sum_{k=1}^{m-1} t_k \log a_{m,k}}{t_m + \sum_{k=1}^{m-1} t_k}} - \sum_{n=1}^{m-1} t_n \frac{1}{a_{m,n}} - t_m e^{\frac{N}{t_m}}. \end{aligned} \quad (34)$$

By applying Lemma 2, the difference $\tilde{\Delta}$ can be simplified as follows:

$$\tilde{\Delta} = e^{\frac{N - (\sum_{k=1}^{m-1} t_k) \log a_{m,1}}{t_m + \sum_{k=1}^{m-1} t_k}} \sum_{n=1}^m t_n - \frac{1}{a_{m,1}} \sum_{n=1}^{m-1} t_n - t_m e^{\frac{N}{t_m}}. \quad (35)$$

Recall that $\tau_m = \sum_{k=1}^m t_k$, which can be used to simplify $\tilde{\Delta}$ as follows:

$$\tilde{\Delta} = \tau_m e^{\frac{N - \tau_{m-1} \log a_{m,1}}{\tau_m}} - \tau_{m-1} \frac{1}{a_{m,1}} - t_m e^{\frac{N}{t_m}}. \quad (36)$$

We note that by using Lemma 2, the feasibility condition of hybrid NOMA, $\min\{a_{m,1}, \dots, a_{m,m-1}\} \geq e^{-\frac{N - \sum_{k=1}^{m-1} t_k \log a_{m,k}}{t_m + \sum_{k=1}^{m-1} t_k}}$, can be simplified as follows:

$$a_{m,1}^{-1} \leq e^{\frac{N}{t_m}}. \quad (37)$$

Because the feasibility condition is a lower bound on $a_{m,1}$, $\tilde{\Delta}$ in (36) can be expressed as an explicit function of $a_{m,1}$ as follows:

$$\tilde{\Delta} = \tau_m e^{\frac{N}{\tau_m}} (a_{m,1}^{-1})^{\frac{\tau_{m-1}}{\tau_m}} - \tau_{m-1} (a_{m,1}^{-1}) - t_m e^{\frac{N}{t_m}} = f(a_{m,1}^{-1}), \quad (38)$$

where $f(x) \triangleq \tau_m e^{\frac{N}{\tau_m}} x^{\frac{\tau_{m-1}}{\tau_m}} - \tau_{m-1} x - t_m e^{\frac{N}{t_m}}$.

The lemma can be proved by showing $f(x) \leq 0$ for $x \leq e^{\frac{N}{t_m}}$. The first order derivative of $f(x)$ is given by

$$\begin{aligned} f'(x) &= \tau_m e^{\frac{N}{\tau_m}} \left(\frac{\tau_{m-1}}{\tau_m} \right) x^{\frac{\tau_{m-1}}{\tau_m} - 1} - \tau_{m-1} \\ &= \tau_{m-1} e^{\frac{N}{\tau_m}} x^{-\frac{t_m}{\tau_m}} - \tau_{m-1}, \end{aligned} \quad (39)$$

which is a monotonically decreasing function of x . Therefore, for $x \leq e^{\frac{N}{t_m}}$, $f'(x)$ is lower bounded as follows:

$$f'(x) \geq \tau_{m-1} e^{\frac{N}{\tau_m}} \left(e^{\frac{N}{t_m}} \right)^{-\frac{t_m}{\tau_m}} - \tau_{m-1} \geq 0, \quad (40)$$

which means that $f(x)$ is a monotonically increasing function of x . Therefore, for $x \leq e^{\frac{N}{t_m}}$, $f(x)$ is upper bounded as follows:

$$\begin{aligned} f(x) &\leq f\left(e^{\frac{N}{t_m}}\right) = \tau_m e^{\frac{N}{\tau_m}} e^{\frac{N \tau_{m-1}}{t_m \tau_m}} - \tau_{m-1} e^{\frac{N}{t_m}} - t_m e^{\frac{N}{t_m}} \\ &= \tau_m \left(e^{\frac{N}{\tau_m} + \frac{N \tau_{m-1}}{t_m \tau_m}} - e^{\frac{N}{t_m}} \right). \end{aligned} \quad (41)$$

Define the difference of the exponents of the two exponential functions in (41) as $\check{\Delta}$, which can be evaluated as follows:

$$\check{\Delta} \triangleq \frac{N}{\tau_m} + \frac{N \tau_{m-1}}{t_m \tau_m} - \frac{N}{t_m} = \frac{N \tau_m - N \tau_m}{t_m \tau_m} = 0.$$

This means that

$$f(x) \leq f\left(e^{\frac{N}{t_m}}\right) = 0.$$

Therefore, $\tilde{\Delta} \leq 0$, i.e., the energy consumption of OMA offloading is no less than that of hybrid NOMA. The proof of the lemma is completed.

APPENDIX D

PROOF FOR LEMMA 4

The lemma can be proved by using the hybrid NOMA solution with $\tilde{\mathcal{S}}_m = \{2, \dots, m-1\}$ as an example and showing that it does not consume less energy than the hybrid NOMA solution with $\mathcal{S}_m = \{1, \dots, m-1\}$. Recall that the energy consumed by the hybrid NOMA solution with $\mathcal{S}_m = \{1, \dots, m-1\}$ is given by

$$E_H = t_m \frac{e^{\frac{N - \sum_{k=1}^{m-1} t_k \log a_{m,k}}{t_m + \sum_{k=1}^{m-1} t_k}} - 1}{|h_m|^2} + \sum_{n=1}^{m-1} t_n \frac{a_{m,n} e^{\frac{N - \sum_{k=1}^{m-1} t_k \log a_{m,k}}{t_m + \sum_{k=1}^{m-1} t_k}} - 1}{a_{m,n} |h_m|^2}, \quad (42)$$

where the energy consumed by the solution with $\tilde{\mathcal{S}}_m = \{2, \dots, m-1\}$, denoted by \tilde{E}_H , is a special case of E_H by letting $t_1 \rightarrow 0$. Therefore, in order to prove $\tilde{E}_H \geq E_H$, it is sufficient to show that E_H is a monotonically non-increasing function of t_1 . The first order derivative of E_H with respect to t_1 is given by

$$\begin{aligned} E'_H = & \sum_{n=1}^m \frac{t_n}{|h_m|^2} e^{\frac{N - \sum_{k=1}^{m-1} t_k \log a_{m,k}}{t_m + \sum_{k=1}^{m-1} t_k}} \left(\frac{-\log a_{m,1}}{t_m + \sum_{k=1}^{m-1} t_k} - \frac{N - \sum_{k=1}^{m-1} t_k \log a_{m,k}}{(t_m + \sum_{k=1}^{m-1} t_k)^2} \right) \\ & + \frac{a_{m,1} e^{\frac{N - \sum_{k=1}^{m-1} t_k \log a_{m,k}}{t_m + \sum_{k=1}^{m-1} t_k}} - 1}{a_{m,1} |h_m|^2}. \end{aligned} \quad (43)$$

Because E_H is achieved by the hybrid NOMA solution with $\mathcal{S}_m = \{1, \dots, m-1\}$, Lemma 2 can be applied to simplify E'_H as follows:

$$\begin{aligned} E'_H &= \frac{1}{|h_m|^2} e^{\frac{N - \tau_{m-1} \log a_{m,1}}{\tau_m}} \left(-\frac{N + t_m \log a_{m,1}}{\tau_m} \right) + \frac{a_{m,1} e^{\frac{N - \tau_{m-1} \log a_{m,1}}{\tau_m}} - 1}{a_{m,1} |h_m|^2} \\ &= \frac{1}{|h_m|^2} e^{\frac{N - \tau_{m-1} \log a_{m,1}}{\tau_m}} \left(\frac{\tau_m - N - t_m \log a_{m,1}}{\tau_m} \right) - \frac{1}{|h_m|^2} e^{-\log a_{m,1}} \\ &= \frac{1}{|h_m|^2} p(-\log a_{m,1}), \end{aligned}$$

where $p(x) \triangleq e^{\frac{N + \tau_{m-1} x}{\tau_m}} \left(\frac{\tau_m - N + t_m x}{\tau_m} \right) - e^x$.

Recall that the feasibility condition for hybrid NOMA has been provided in (37) and it can be further rewritten as follows:

$$-\log a_{m,1} \leq \frac{N}{t_m}. \quad (44)$$

Therefore, the fact that $E'_H \leq 0$ can be proved by proving that $p(x) \leq 0$ for $x \leq \frac{N}{t_m}$, as shown in the following. Rewrite $p(x)$ as follows:

$$p(x) = e^{\frac{N + \tau_{m-1} x}{\tau_m}} e^{\log\left(\frac{\tau_m - N + t_m x}{\tau_m}\right)} - e^x. \quad (45)$$

Define $\tilde{p}(x) \triangleq \frac{N+\tau_{m-1}x}{\tau_m} + \log\left(\frac{\tau_m-N+t_mx}{\tau_m}\right) - x$. We note that $p(x) \leq 0$ is equivalent to $\tilde{p}(x) \leq 0$. The first order derivative of $\tilde{p}(x)$ is given by

$$\tilde{p}'(x) = \frac{\tau_{m-1}}{\tau_m} + \frac{t_m}{\tau_m} \frac{\tau_m}{\tau_m - N + t_mx} - 1. \quad (46)$$

We note that $\tilde{p}'(x)$ is a monotonically non-increasing function of x , which means

$$\tilde{p}'(x) \geq \tilde{p}'\left(\frac{N}{t_m}\right) = \frac{-t_m}{\tau_m} + \frac{t_m}{\tau_m} \frac{\tau_m}{\tau_m - N + N} = 0, \quad (47)$$

since $x \leq \frac{N}{t_m}$. Therefore, $\tilde{p}(x)$ is a monotonically non-decreasing function of x , which means that the use of the upper bound on x , $x \leq \frac{N}{t_m}$, yields the following upper bound on $p(x)$:

$$p(x) \leq p\left(\frac{N}{t_m}\right) = e^{\frac{N+\tau_{m-1}\frac{N}{t_m}}{\tau_m}} e^{\log\left(\frac{\tau_m-N+N}{\tau_m}\right)} - e^{\frac{N}{t_m}} = 0. \quad (48)$$

Therefore, $E'_H \leq 0$, and hence E_H is a monotonically non-increasing function of t_1 , which means that the hybrid NOMA solution without using t_1 results in no less energy consumption than the solution with $\mathcal{S}_m = \{1, \dots, m-1\}$. The proof is complete.

APPENDIX E

PROOF FOR LEMMA 5

The lemma can be proved by showing that the power allocation solutions of problem P3 do not violate the transmit power constraint of problem P2. Since P_m^{OMA} is the transmit power which is sufficient for U_m to complete OMA offloading by using t_m only, one can conclude that $P_{m,m}^* \leq P_m^{\text{OMA}}$, which means that the lemma can be proved by showing that $P_{m,n}^* \leq P_m^{\text{OMA}}$, $n < m$.

Recall that $P_{m,n}^*$, $n < m$, can be expressed as follows:

$$P_{m,n}^* = \frac{a_{m,n} e^{\frac{N - \sum_{k=1}^{m-1} t_k \log a_{m,k}}{t_m + \sum_{k=1}^{m-1} t_k}} - 1}{a_{m,n} |h_m|^2} = \frac{a_{m,1} e^{\frac{N - \tau_{m-1} \log a_{m,1}}{\tau_m}} - 1}{a_{m,1} |h_m|^2}, \quad (49)$$

where the last step follows from Lemma 2. Therefore, the difference between $P_{m,n}^*$ and the maximal power P_m^{OMA} is given by

$$\begin{aligned} \Delta_m &\triangleq P_{m,n}^* - P_m^{\text{OMA}} = \frac{a_{m,1} e^{\frac{N - \tau_{m-1} \log a_{m,1}}{\tau_m}} - 1}{a_{m,1} |h_m|^2} - P_m^{\text{OMA}} \\ &= \frac{a_{m,1} e^{\frac{N - \tau_{m-1} \log a_{m,1}}{\tau_m}} - 1 - P_m^{\text{OMA}} a_{m,1} |h_m|^2}{a_{m,1} |h_m|^2}, \end{aligned}$$

such that the lemma can be proved by showing that $\Delta_m \leq 0$.

Recall that, in OMA, P_m^{OMA} is used to deliver N nats by using t_m only, i.e., $t_m \log(1 + P_m^{\text{OMA}}|h_m|^2) = N$, which can be used to rewrite the difference as follows:

$$\Delta_m = \frac{a_{m,1} e^{\frac{N - \tau_{m-1} \log a_{m,1}}{\tau_m}} - 1 - a_{m,1} \left(e^{\frac{N}{t_m}} - 1 \right)}{a_{m,1} |h_m|^2}. \quad (50)$$

Since $a_{m,1} \leq 1$, $\Delta_m \leq 0$ can be proved by showing the following inequality

$$\bar{\Delta}_m \triangleq \frac{N - \tau_{m-1} \log a_{m,1}}{\tau_m} - \frac{N}{t_m} \leq 0. \quad (51)$$

$\bar{\Delta}_m$ can be further rewritten as follows:

$$\begin{aligned} \bar{\Delta}_m &= \frac{t_m N - t_m \tau_{m-1} \log a_{m,1} - \tau_m N}{\tau_m t_m} \\ &= \frac{-t_m \tau_{m-1} \log a_{m,1} - \tau_{m-1} N}{\tau_m t_m}, \end{aligned} \quad (52)$$

where it is important to point out that $-t_m \tau_{m-1} \log a_{m,1}$ is non-negative since $a_{m,1} \leq 1$.

We recall that the feasibility condition for the hybrid NOMA solution is $t_m \leq -\frac{N}{\log a_{m,1}}$, which means that the difference can be upper bounded as follows:

$$\bar{\Delta}_m \leq \frac{\frac{N}{\log a_{m,1}} \tau_{m-1} \log a_{m,1} - \tau_{m-1} N}{\tau_m t_m} = 0, \quad (53)$$

which means that $P_{m,n}^* \leq P_m^{\text{OMA}}$, $n < m$. Therefore, the proof of the lemma is complete.

APPENDIX F

PROOF FOR LEMMA 6

Define \mathcal{E}_M as a set collecting all feasible \mathbf{E}_M . The lemma is equivalent to the statement that there is no vector in \mathcal{E}_M dominating \mathbf{E}_M^* . The lemma can be proved by using again mathematical induction.

For the base case, $M = 2$, a conclusion stronger than the lemma can be proved. In particular, we will show that \mathbf{E}_2^* is the minimum element of \mathcal{E}_2 , i.e., $\mathbf{E}_2^* < \mathbf{E}_2$ or equivalently $E_2^* < \mathbf{E}_2(1)$, for any $\mathbf{E}_2 \in \mathcal{E}_2$, where $\mathbf{E}_m(i)$ denotes the i -th element of the $(m-1)$ vector, \mathbf{E}_m . Recall that E_2^* is obtained by solving the following optimization problem

$$\min_{t_2, P_{2,n}} E_2 \quad (\text{P6a})$$

$$\text{s.t. } t_1^* R_{2,1} + t_2 R_{2,2} \geq N, \quad t_2 \leq D_2 - D_1, \quad t_2 \geq 0, \quad (\text{P6b})$$

$$0 \leq P_{2,n} \leq P_2^{\text{OMA}}, \quad 1 \leq n \leq 2. \quad (\text{P6c})$$

Because problem P6 is obtained from problem P1 by discarding the constraints related to U_i 's parameters, $i > 2$, the optimal value of problem P6 is an achievable lower bound on the optimal value of problem P1. In other words, $E_2^* \leq E_2(1)$, for any $E_2 \in \mathcal{E}_2$. Therefore, the lemma holds for the base case.

For the inductive step, assume that the lemma holds for the case with $(m-1)$ users, i.e., E_{m-1}^* is Pareto optimal. In order to prove that E_m^* is also Pareto optimal, a proof by contradiction is used. In particular, assume that the lemma does not hold for the case with m users, i.e., there exists a vector in \mathcal{E}_m which is denoted by \bar{E}_m and satisfies $\bar{E}_m \prec E_m^*$. Because of the assumption that E_{m-1}^* is Pareto optimal, $\bar{E}_m \prec E_m^*$ means that $\bar{E}_m(m) < E_m^*$, and $\bar{E}_m(i) = E_i^*$ for $1 \leq i \leq m-1$.

Since E_m^* is the optimal value of problem P2, $\bar{E}_m(m) < E_m^*$ is possible only if the following event happens. In particular, one or multiple users choose their transmission parameters different from $P_{m,n}^*$ and t_n^* , for $1 \leq n \leq m-1$, which changes the feasibility set of problem P2. Among these users, denote by $U_{\bar{n}}$, $2 \leq \bar{n} \leq n$, the user whose signal is decoded first, which leads to the following two conclusions. The first is that $\bar{E}_m(\bar{n}) = E_{\bar{n}}^*$, and the second is that the users, U_i , $1 \leq i \leq \bar{n}-1$, choose t_i^* and $P_{i,j}^*$, $1 \leq j \leq i$, for their transmission. The statement that $U_{\bar{n}}$ did not choose $t_{\bar{n}}^*$ and $P_{\bar{n},j}^*$, $1 \leq j \leq \bar{n}$, but still realizes $E_{\bar{n}}^*$ is equivalent to the statement that, for the following optimization problem, there are two different optimal solutions realizing the same optimal value:

$$\min_{\mathbf{P}_{\bar{n}}, t_{\bar{n}}} E_{\bar{n}} \quad (\text{P7a})$$

$$\text{s.t.} \quad \sum_{j=1}^{\bar{n}} t_j^* R_{\bar{n},j} + t_{\bar{n}} R_{\bar{n},\bar{n}} \geq N, \quad t_{\bar{n}} \leq D_{\bar{n}} - \sum_{j=1}^{\bar{n}-1} t_j, \quad t_{\bar{n}} \geq 0, \quad (\text{P7b})$$

$$0 \leq P_{\bar{n},j} \leq P_{\bar{n}}^{\text{OMA}}, \quad 1 \leq j \leq \bar{n}. \quad (\text{P7c})$$

However, Corollary 2 shows that there is a single optimal solution for problem P7. Therefore, there cannot exist an \bar{E}_m in \mathcal{E}_m which satisfies $\bar{E}_m \prec E_m^*$, which completes the inductive step.

By combining the conclusions from the base case and the inductive step, the lemma is proved.

REFERENCES

- [1] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2322–2358, Dec. 2017.
- [2] Q.-V. Pham *et al.*, "A survey of multi-access edge computing in 5G and beyond: Fundamentals, technology integration, and state-of-the-art," *IEEE Access*, vol. 8, pp. 116 974–117 017, 2020.

- [3] J. Santos, T. Wauters, B. Volckaert, and F. De Turck, "Towards low-latency service delivery in a continuum of virtual resources: State-of-the-art and research directions," *IEEE Commun. Surveys Tuts.*, to appear in 2021.
- [4] H. Li, K. Ota, and M. Dong, "Learning IoT in edge: Deep learning for the internet of things with edge computing," *IEEE Netw.*, vol. 32, no. 1, pp. 96–101, Feb. 2018.
- [5] W. Y. B. Lim *et al.*, "Federated learning in mobile edge networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 2031–2063, Apr. 2020.
- [6] Z. Ding, R. Schober, and H. Vincent Poor, "No-pain no-gain: DRL assisted optimization in energy-constrained CR-NOMA networks," *IEEE Trans. Commun.*, to appear in 2021.
- [7] M. Elbayoumi, M. Kamel, W. Hamouda, and A. Youssef, "NOMA-assisted machine-type communications in UDN: State-of-the-art and challenges," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 2, pp. 1276–1304, 2020.
- [8] O. Maraqa, A. S. Rajasekaran, S. Al-Ahmadi, H. Yanikomeroglu, and S. M. Sait, "A survey of rate-optimal power domain NOMA with enabling technologies of future wireless networks," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 4, pp. 2192–2235, Fourth quarter 2020.
- [9] Z. Ding, P. Fan, and H. V. Poor, "Impact of non-orthogonal multiple access on the offloading of mobile edge computing," *IEEE Transactions on Communications*, vol. 67, no. 1, pp. 375–390, Jan. 2019.
- [10] Q.-V. Pham, H. T. Nguyen, Z. Han, and W.-J. Hwang, "Coalitional games for computation offloading in NOMA-enabled multi-access edge computing," *IEEE Trans. Veh. Tech.*, vol. 69, no. 2, pp. 1982–1993, Feb. 2020.
- [11] A. Kiani and N. Ansari, "Edge computing aware NOMA for 5G networks," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 1299–1306, Aug. 2018.
- [12] M. Sheng, Y. Dai, J. Liu, N. Cheng, X. Shen, and Q. Yang, "Delay-aware computation offloading in NOMA MEC under differentiated uploading delay," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2813–2826, 2020.
- [13] W. Feng, J. Tang, N. Zhao, X. Zhang, X. Wang, K.-K. Wong, and J. Chambers, "Hybrid beamforming design and resource allocation for UAV-aided wireless-powered mobile edge computing networks with NOMA," *IEEE J. Sel. Areas Commun.*, pp. 1–1, to appear in 2021.
- [14] I. Budhiraja, N. Kumar, S. Tyagi, and S. Tanwar, "Energy consumption minimization scheme for NOMA-based mobile edge computation networks underlying UAV," *IEEE Systems Journal*, pp. 1–10, to appear in 2021.
- [15] Y. Xu, T. Zhang, D. Yang, Y. Liu, and M. Tao, "Joint resource and trajectory optimization for security in UAV-assisted MEC systems," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 573–588, Jan. 2021.
- [16] L. Qian, W. Wu, W. Lu, Y. Wu, B. Lin, and T. Q. S. Quek, "Secrecy-based energy-efficient mobile edge computing via cooperative non-orthogonal multiple access transmission," *IEEE Trans. Commun.*, vol. 69, no. 7, pp. 4659–4677, Jul. 2021.
- [17] B. Li, W. Wu, W. Zhao, and H. Zhang, "Security enhancement with a hybrid cooperative NOMA scheme for MEC system," *IEEE Trans. Veh. Tech.*, vol. 70, no. 3, pp. 2635–2648, Mar. 2021.
- [18] B. Li, F. Si, W. Zhao, and H. Zhang, "Wireless powered mobile edge computing with NOMA and user cooperation," *IEEE Trans. Veh. Tech.*, vol. 70, no. 2, pp. 1957–1961, Feb. 2021.
- [19] F. Zhou and R. Q. Hu, "Computation efficiency maximization in wireless-powered mobile edge computing networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3170–3184, May 2020.
- [20] Z. Ding, J. Xu, O. A. Dobre, and H. V. Poor, "Joint power and time allocation for NOMA-MEC offloading," *IEEE Wireless Commun. Lett.*, vol. 68, no. 6, pp. 6207–6211, Jun. 2019.
- [21] Z. Ding, D. W. K. Ng, R. Schober, and H. V. Poor, "Delay minimization for NOMA-MEC offloading," *IEEE Signal Process. Lett.*, vol. 25, no. 12, pp. 1875–1879, Dec. 2018.
- [22] T. Cover and J. Thomas, *Elements of Information Theory*, 6th ed. Wiley and Sons, New York, 1991.
- [23] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, Cambridge, UK, 2003.