Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 5143640, 13 pages
https://doi.org/10.1155/2022/5143640

Research Article

Joint Task Partition and Resource Allocation for Multiuser Cooperative Mobile Edge Computing

Gang Xie,[1] **Zhenzhen Wang**,[1] **and Yuanan Liu**[2]

[1]*School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China*
[2]*School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing, China*

Correspondence should be addressed to Gang Xie; xiegang@bupt.edu.cn

Received 11 February 2022; Revised 27 May 2022; Accepted 22 June 2022; Published 29 July 2022

Academic Editor: Jun Cai

Exploiting the idle computation resources distributed at wireless devices (WDs) can enhance the mobile edge computing (MEC) computation performance. This paper studies a multiuser cooperative computing system consisting of one local user and multiple helpers, in which the user solicits multiple nearby WDs acting as helpers for cooperative computing. We design an efficient orthogonal frequency-division multiple access- (OFDMA-) aided three-phase transmission protocol, under which the user's computation-intensive tasks can be executed in parallel by local computing and offloading. Under this setup, we study the energy consumption minimization problem by optimizing the user's task partition, jointly with the communication and computation resources allocation for task offloading and results downloading, subject to the user's computation latency constraint. For the nonconvex problem, we first transform the original problem into a convex one and then use the Lagrange duality method to obtain the globally optimal solution. Compared with other benchmark schemes, numerical results validate the effectiveness of the proposed joint task partition and resource allocation (JTPRA) scheme.

1. Introduction

The real-time communication and computation of massive wireless devices (WDs) (e.g., smart wearable devices and laptops) promote the rapid growth of emerging applications (e.g., face recognition, smart grid, and autonomous driving) [1]. In fact, these applications or tasks may be computation-intensive and latency-critical, but WDs are generally of small size and only have the finite battery power. Hence, how to enhance their computation capabilities and reduce the computation latency is one crucial but challenging task to be handled. To deal with such limitations, mobile edge computing (MEC) has been proposed as a promising technology by providing cloud-like computing at the network edge (e.g., base stations (BSs) and access points (APs)) [2].

Various efforts have been devoted to handling technical challenges against different computation task models. Two extensively adopted task models in the current research works are partial and binary offloading, respectively. Note that in partial offloading, the mutual dependency of the computing tasks significantly affects the computation offloading process [3]. That means partial offloading can be classified as the task-call graph and the data-partition model. Also, there exists different types of MEC system architectures, such as single user single-server [4–7], multiuser single-server [8–14], and single/multiuser multiserver [15–18].

Due to the increasing number of WDs, resource contention may occur on MEC servers. Under this circumstance, cooperative computing provides a viable solution by utilizing abundant idle computation resources distributed at WDs [19, 20]. In a basic two-user device-to-device (D2D) cooperative computing system, [19] jointly optimized both users' local computing and task offloading decisions over time, in order to minimize their weighted sum-energy consumption. Under a single-user single-helper single-server setup, [21, 22] jointly optimized the communication and computation resources allocation at both the user and helper based on time-division multiple access (TDMA) and nonorthogonal multiple access (NOMA), respectively. In a cellular D2D MEC system, [23] proposed a joint task management architecture to achieve efficient information interaction and task management. Also, by integrating D2D into the MEC system, [24] jointly optimized

D2D pairing, task split, and the communication and computation resource allocation, in order to improve the system computation capacity.

In the above research works, the user is mostly assumed to cooperate with single helper at the same time. In practical design, multiple helpers can simultaneously share their own computation and communication resources to help the user [20, 25–27]. In the D2D MEC system, [25] jointly optimized helpers' selection and the communication and computation resource allocation for minimizing the energy consumption. In [26], a multihelper MEC with NOMA-based cooperative edge computing has been presented to maximize the total offloading data subject to the latency constraints. However, the system model in [25, 26] ignores results downloading, which may be not applicable for practical design. Thus, [20, 27] focus on the joint task offloading and results downloading. In the D2D-enabled multihelper MEC system, [27] jointly optimized the time and rate for task offloading and results downloading, as well as the computation frequency for task execution, in order to minimize the computation latency. Unlike the binary offloading model in [20, 27], it investigated a multiuser computational offloading scheme, in which the controlling user partially distributes its computing tasks to multiple trusted helpers. Also, [20] ignores computation resource of the local user and dynamic management of computation frequency.

Despite the recent research progress, cooperative computing still faces some technical challenges. First, the previous research works mostly consider cooperating with the MEC server or single helper. When multiple helpers share unused resources to help the user, how to effectively coordinate the cooperation between the user and multiple helpers for achieving computing diversity remains challenging, especially when the helper number becomes large. Second, the previous research works generally ignore potential performance improvement brought by dynamic management of computation frequency as well as results downloading. When the MEC system considers these options, how to solve such a complex problem is also challenging.

Motivated by this, we consider a multiuser cooperative MEC system consisting of one user and multiple nearby WDs serving as helpers. The user has individual computation tasks to be executed within a given time block. To implement the cooperation between the user and the helpers, the time block is divided into three phases. In the first phase, the user simultaneously offloads the computing tasks to multiple nearby helpers. In the second phase, the helpers execute their assigned computation bits. In the three phase, the helpers send the computation results back to the user. Under this setup, this paper develops an energy-efficient multiuser cooperative MEC design by optimizing the user's task partition, jointly with the communication and computation resource allocation for task offloading and results downloading. The main contributions of this paper are summarized as follows.

(1) We propose an MEC framework for multiuser cooperative computing, in which the user can simultaneously offload the computing tasks to multiple nearby helpers

(2) We design an OFDMA-aided three-phase transmission protocol involving results downloading, which efficiently coordinates the cooperation between the user and multiple nearby helpers

(3) For the energy consumption minimization problem, we optimize the user's task partition, jointly with the communication and computation resource allocation for task offloading and results downloading. Due to nonconvexity of this problem, we first transform it into a convex one and then use the Lagrange duality method to obtain the globally optimal solution

The rest of this paper is organized as follows. Section 2 introduces the system model. The proposed joint task partition and resource allocation problem is formulated in Section 3. The joint task partition and resource allocation algorithm is presented in Sections 4. Section 5 provides numerical results, followed by the conclusion in Section 6.

Notation is as follows: we employ uppercase boldface letters and lowercase boldface ones for matrices and vectors, respectively. Δ is represented by "denoted by" $[x]_a^b$. And $[x]^+$ is denoted by $\{b, \min\{a, x\}\}$ and $\max\{0, x\}$, respectively. A continuous random variable z uniformly distributed over $[a, b]$ is denoted by $z \sim U\ [a, b]$. $|\mathbf{A}|$ denotes the determinant of a matrix $\mathbf{A}$. Moreover, $l_u + \sum_{k=1}^{K} l_k = L_u$ and R^+ stand for the sets of nonnegative real vectors of dimension K and positive real numbers, respectively.

2. System Model

As shown in Figure 1, we consider a multiuser cooperative MEC system, which consists of one user and a set.

$K =^{\Delta} \{1, \cdots, K\}$ of nearby helpers all equipped with single antenna. We focus on a time block with length T, where the user should execute the computing tasks with data-size L_u (in bits) within this block. Here, T is no larger than the channel coherence time [21]. Suppose that there is a central controller that is responsible for collecting the network information, such as the global channel state information (CSI), accordingly, the central controller can send the optimized strategies to the user and helpers to take actions [21, 22]. For easy implementation, it is further assumed that the task offloading and result downloading channel reciprocity are leveraged in this paper [20].

Specifically, the L_u bits generally can be divided into $K + 1$ independent parts for local computing and offloading to the helpers, respectively. Let $l_u \geq 0$ and $l_1 \geq 0, \cdots, l_K \geq 0$ denote the numbers of bits for local computing at the user and offloading to K helpers, respectively. Then, we have

$$l_u + \sum_{k=1}^{K} l_k = L_u. \tag{1}$$

2.1. Local Computing. The l_u bits are executed locally with the optimal central process unit (CPU) frequency given as [21]

$$f_u = \frac{c_u l_u}{T}, \tag{2}$$

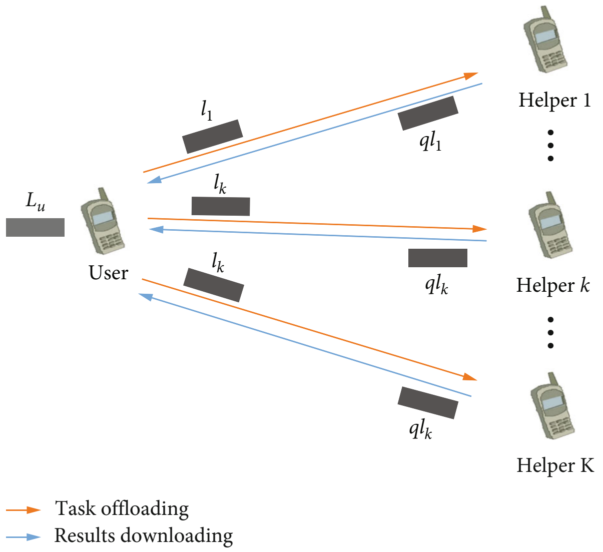

Figure 1: System model of multiuser cooperative MEC.

where c_u denotes the number of CPU cycles for computing 1-bit input-data at the user. Note that f_u is subject to the maximum frequency constraint, that is,

$$f_u \leq f_u^{\max}. \tag{3}$$

Accordingly, the user's energy consumption for local computing is given by

$$E_u^{\text{comp}} = \gamma_u c_u l_u f_u^2. \tag{4}$$

where γ_u denotes a constant related to the user's hardware architecture [21]. Replacing f_u in (4) with (2), E_u^{comp} can thus be reexpressed as

$$E_u^{\text{comp}} = \gamma_u \frac{(c_u l_u)^3}{T^2}. \tag{5}$$

2.2. Remote Computing at Helpers. The OFDMA-aided three-phase transmission protocol is shown in Figure 2. At first, the user first offloads l_k bits to the k -th helper with duration $t^{\text{off}}{}_k$ via OFDMA in the task offloading phase, $k \in K$. Then, the k -th helper executes its assigned bits with duration $t^{\text{comp}}{}_k$ in the task execution phase. At last, in the result downloading phase, the k -th helper sends the computation results back to the user with duration $t^{\text{dl}}{}_k$ via OFDMA. Note that the cooperation between the user and K helpers does not affect each other. To meet the user's latency requirement, we have the following time constraint:

$$t_k^{\text{off}} + t_k^{\text{comp}} + t_k^{\text{dl}} \leq T, \forall k \in \mathcal{K}. \tag{6}$$

In the following, we describe the OFDMA-aided three-phase transmission protocol in detail.

2.2.1. Phase I (Task Offloading). Let $h^{\text{off}}{}_k$ denote the channel power gain from the user to the k -th helper, $k \in K$. The achievable offloading rate at the k -th helper is given by

$$l_k = t_k^{\text{off}} r_k^{\text{off}} \left(p_k^{\text{off}} \right), \tag{7}$$

where W in Hz denotes one frequency resource block, $p^{\text{off}}{}_k$ is the transmit power for offloading data to the k -th helper, and $\sigma_k{}^2$ is the power of additive white Gaussian noise (AWGN) at the k -th helper. Hence, we have the offloaded bits l_k from the user to the k -th helper as

$$l_k = t_k^{\text{off}} r_k^{\text{off}} \left(p_k^{\text{off}} \right). \tag{8}$$

Accordingly, the total energy consumption for task offloading consumed by the user is expressed as

$$E_u^{off} = \sum_{k=1}^{K} t_k^{off} p_k^{off}. \tag{9}$$

2.2.2. Phase II (Task Execution). After receiving l_k bits, the k -th helper executes with the optimal CPU frequency given as

$$f_k = \frac{l_k c_k}{T - t_k^{\text{off}} - t_k^{\text{dl}}}, \tag{10}$$

where c_k denotes the number of CPU cycles for computing 1-bit input-data at the k -th helper. Similarly as in (3), f_k is also subject to the maximum frequency constraint, that is,

$$f_k \leq f_k^{\max}, \forall k \in \mathcal{K}. \tag{11}$$

Consequently, the energy consumption for cooperative computation at the k -th helper is expressed as

$$E_k^{\text{comp}} = \gamma_k \frac{(l_k c_k)^3}{\left(T - t_k^{\text{off}} - t_k^{\text{dl}}\right)^2}, \tag{12}$$

where γ_k denotes a constant related to the k -th helper's hardware architecture [21].

2.2.3. Phase III (Result Downloading). After executing the user's assigned bits, the k -th helper begins sending the computation results back to the user via OFDMA. Let h_k^{dl} denote the channel power gain from the k -th helper to the user. The achievable downloading rate from the k -th helper is given by

$$r_k^{\text{dl}} \left(p_k^{\text{dl}} \right) = W \log_2 \left(1 + \frac{p_k^{\text{dl}} h_k^{\text{dl}}}{\sigma_0^2} \right), \tag{13}$$

where p_k^{dl} is the transmit power of the k -th helper, and σ_0^2 is the power of AWGN at the user. The corresponding computation results are thus given by

$$ql_k = t_k^{\text{dl}} r_k^{\text{dl}} \left(p_k^{\text{dl}} \right), \tag{14}$$

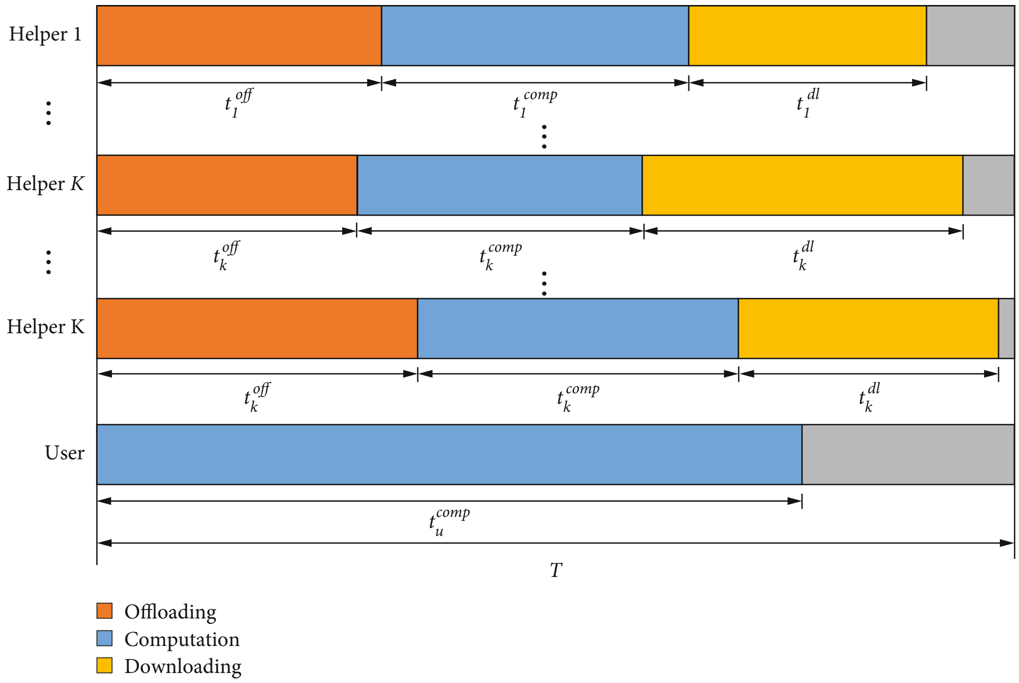

Figure 2: An illustration of the OFDMA-aided three-phase transmission protocol.

where $q \in R^+$ denotes the normalized ratio between the size of computation results and the size of computing tasks [20]. The energy consumption for results downloading consumed by the k -th helper is expressed as

$$E_k^{\mathrm{dl}} = t_k^{\mathrm{dl}} p_k^{\mathrm{dl}}. \tag{15}$$

3. Problem Formulation

In this paper, we aim to minimize the total energy consumption of the multiuser cooperative MEC system (i.e., $E_u^{\mathrm{off}} + E_u^{\mathrm{comp}} + \sum_{k=1}^K (E_k^{\mathrm{dl}} + E_k^{\mathrm{comp}})$) by jointly optimizing the user's task partition, the task offloading time, the result downloading time, and the transmit power of the user and helpers, subject to the user's computation latency constraint T. Specifically, the energy consumption minimization problem is formulated as

$$(\mathrm{P1}): \min_{l, t_{\mathrm{off}}, t_{\mathrm{dl}}, p_{\mathrm{off}}, p_{\mathrm{dl}}} E_u^{\mathrm{off}} + \sum_{k=1}^K \left(E_k^{\mathrm{dl}} + E_k^{\mathrm{comp}}\right) + E_u^{\mathrm{comp}}\, s.t. l_k \le t_k^{\mathrm{off}} r_k^{\mathrm{off}} \left(p_k^{\mathrm{off}}\right), \forall k \in \mathscr{K}, \tag{16a}$$

$$\mathrm{q} l_k \le t_k^{\mathrm{dl}} r_k^{\mathrm{dl}} \left(p_k^{\mathrm{dl}}\right), \forall k \in \mathscr{K}, \tag{16b}$$

$$\sum_{k=1}^K t_k^{\mathrm{off}} p_k^{\mathrm{off}} \le E_{\max}^{\mathrm{off}}, \tag{16c}$$

$$t_k^{\mathrm{dl}} p_k^{\mathrm{dl}} \le E_{\max}^{\mathrm{dl}}, \forall k \in \mathscr{K}, \tag{16d}$$

$$t_k^{\mathrm{dl}} \le T, t_k^{\mathrm{off}} \le T, \forall k \in \mathscr{K}, \tag{16e}$$

$$\mathbf{l} \in \mathbb{R}_{\ge 0}^{\mathrm{K}+1}, \mathbf{t_{off}} \in \mathbb{R}_{\ge 0}^{\mathrm{K}}, \mathbf{t_{dl}} \in \mathbb{R}_{\ge 0}^{\mathrm{K}}, \begin{matrix} \mathbf{p_{off}} \in \mathbb{R}_{\ge 0}^{\mathrm{K}}, \mathbf{p_{dl}} \in \mathbb{R}_{\ge 0}^{\mathrm{K}} \\ (1), (3), (11) \end{matrix}, \tag{16f}$$

where $\mathbf{l} \triangleq \{l_u, l_1, \cdots, l_K\} \in \mathbb{R}_{\ge 0}^{K+1}$, $\mathbf{t}_{\mathrm{off}} \triangleq \{t_1^{\mathrm{off}}, \cdots, t_K^{\mathrm{off}}\} \in \mathbb{R}_{\ge 0}^K$, $\mathbf{t}_{\mathrm{dl}} \triangleq \{t_1^{\mathrm{dl}}, \cdots, t_K^{\mathrm{dl}}\} \in \mathbb{R}_{\ge 0}^K$, $\mathbf{p}_{off} \triangleq \{p_1^{\mathrm{off}}, \cdots, p_K^{\mathrm{off}}\} \in \mathbb{R}_{\ge 0}^K$, and $\mathbf{p}_{\mathrm{dl}} \triangleq \{p_1^{\mathrm{dl}}, \cdots, p_K^{\mathrm{dl}}\} \in \mathbb{R}_{\ge 0}^K$. (1) denotes the user's task partition constraint, (3) and (11) denote the maximum CPU frequency constraints at the user and helpers, respectively, (16a) and (16b) denote the constraints for data transmission between the user and the helpers, and (16c) and (16d) denote the transmission energy consumption constraints at the user and helpers, respectively. Note that in problem (P1), we replace the equality in (8) and (14) as the inequality constraints (16a) and (16b), respectively. (16a) and (16b) should be met with strict equality at optimality of problem (P1). This is consistent with intuition. Because of the coupling of t_k^{off} and p_k^{off} and $t^{\mathrm{dl}}{}_k$ and $p^{\mathrm{dl}}{}_k$ in the objective function and the constraints (16a) and (16b), problem (P1) is nonconvex.

3.1. Feasibility of Problem (P1). Before solving problem (P1), we need to guarantee its feasibility so that the multiuser cooperative MEC system can support the latency-constrained task execution. Let $L_{\max}$ denote the maximum data size in bits supported by the proposed MEC system within duration T. There is no doubt that only when $L_{\max} \ge L_u$, problem (P1) is feasible, or problem (P1) is infeasible. Hence, we check the feasibility of problem (P1) by determining $L_{\max}$. Intuitively, $L_{\max}$ is obtained when the user and helpers make full use of the communication and computation resources in the proposed MEC system. This corresponds to letting the constraints (3) and (11) be met with strict equality in problem (P1). Then, the data maximization problem is formulated as

$$\text{(P2)}: L_{\max} \triangleq \max_{\mathbf{t^{off}},\mathbf{t^{dl}},\mathbf{p^{off}},\mathbf{p^{dl}},\mathbf{l}} \frac{Tf_u^{\max}}{c_u} + \sum_{k=1}^{K} \frac{\left(T - t_k^{\mathrm{off}} - t_k^{\mathrm{dl}}\right) f_k^{\max}}{c_k} \; s.t. \frac{\left(T - t_k^{\mathrm{off}} - t_k^{\mathrm{dl}}\right) f_k^{\max}}{c_k} \le t_k^{\mathrm{off}} r_k^{\mathrm{off}} \left(p_k^{\mathrm{off}}\right), \forall k \in \mathcal{K},$$

$$q\frac{\left(T - t_k^{\mathrm{off}} - t_k^{\mathrm{dl}}\right) f_k^{\max}}{c_k} \le t_k^{\mathrm{dl}} r_k^{\mathrm{dl}}\left(p_k^{\mathrm{dl}}\right), \forall k \in \mathcal{K}, \sum_{k=1}^{K} t_k^{\mathrm{off}} p_k^{\mathrm{off}} \le E_{\max}^{\mathrm{off}},$$

$$t_k^{\mathrm{dl}} p_k^{\mathrm{dl}} \le E_{\max}^{\mathrm{dl}}, \forall k \in \mathcal{K} \quad . \quad (16e), (16f) \tag{17a}$$

Due to the similarity of between problems (P1) and (P2), problem (P2) can be solved like problem (P1). By comparing $L_{\max}$ and L_u, we finally check the feasibility of problem (P1).

4. Optimal Solution

In this section, we first transform problem (P1) into a convex one and then present an efficient algorithm to obtain the globally optimal solution.

To accomplish this target, we introduce two auxiliary variable vectors $\mathbf{y_{off}} \triangleq [y_1^{off}, \cdots, y_K^{off}]$ and $\mathbf{y_{dl}} \triangleq [y_1^{dl}, \cdots, y_K^{dl}]$ with $y_k^{\mathrm{off}} = t_k^{\mathrm{off}} p_k^{\mathrm{off}}$ and $y_k^{\mathrm{dl}} = t_k^{\mathrm{dl}} p_k^{\mathrm{dl}}$, $\forall k \in K$. Then, it holds that $p^{\mathrm{off}}{}_k = y_k{}^{\mathrm{off}}/t^{\mathrm{off}}{}_k$ if $t^{\mathrm{off}}{}_k > 0$, and $p_k^{\mathrm{off}} = 0$ if either $y_k^{\mathrm{off}} = 0$ or $t_k^{\mathrm{off}} = 0$. Similarly, this also applies to $p^{\mathrm{dl}}{}_k = y_k{}^{\mathrm{dl}}/t^{\mathrm{dl}}{}_k$. By substituting $p_k^{\mathrm{off}} = y_k^{\mathrm{off}}/t_k^{\mathrm{off}}$ and $p_k^{dl} = y_k^{dl}/t_k^{dl}$, problem (P1) can be reformulated as

$$\text{(P1.1)}: \min_{l, t_{\mathrm{off}}, t_{\mathrm{dl}}, y_{\mathrm{off}}, y_{\mathrm{dl}}} \sum_{k=1}^{K} \left(y_k^{\mathrm{off}} + y_k^{dl} + \frac{\gamma_k (l_k c_k)^3}{\left(T - t_k^{\mathrm{off}} - t_k^{\mathrm{dl}}\right)^2} \right) + \frac{\gamma_u (l_u c_u)^3}{T^2},$$

$$s.t\, l_k \le t_k^{\mathrm{off}} r_k^{\mathrm{off}} \left(\frac{y_k^{\mathrm{off}}}{t_k^{\mathrm{off}}} \right), \forall k \in \mathcal{K}, \tag{18a}$$

$$q l_k \le t_k^{dl} r_k^{dl} \left(\frac{y_k^{dl}}{t_k^{dl}} \right), \forall k \in \mathcal{K}, \tag{18b}$$

$$\sum_{k=1}^{K} y_k^{\mathrm{off}} \le E_{\max}^{\mathrm{off}}, \tag{18c}$$

$$y_k^{dl} \le E_{\max}^{dl}, \forall k \in \mathcal{K}, \tag{18d}$$

$$t_k^{dl} \le T, t_k^{\mathrm{off}} \le T, \forall k \in \mathcal{K}, \tag{18e}$$

$$\mathbf{l} \in \mathbb{R}_{\ge 0}^{K+1}, \mathbf{t}_{\mathrm{dl}} \in \mathbb{R}_{\ge 0}^{K}, \mathbf{t}_{\mathrm{off}} \in \mathbb{R}_{\ge 0}^{K}, \mathbf{y}_{\mathrm{dl}} \in \mathbb{R}_{\ge 0}^{K}, \mathbf{y}_{\mathrm{off}} \in \mathbb{R}_{\ge 0}^{K}. \quad (1), (3), (11) \tag{18f}$$

Lemma 1. *Problem (P1.1) is a convex problem.*

Proof. It is obvious that the function $r_k{}^{\mathrm{off}}(x)$ is a concave function with respect to $x \ge 0$. As the perspective operation maintains convexity, $x r_k^{\mathrm{off}}(y/x)$ is jointly concave with respect to $x \ge 0$ and $y \ge 0$ [28]. Similarly, this also applies to $x r_k^{dl}(y/x)$. Therefore, the set defined by the constraints (18a)–(18d) is convex. The Hessian of $l_k^3/(T - t_k^{\mathrm{off}} - t_k^{dl})^2$ is

$$\mathbf{H} = \begin{bmatrix} \frac{6l_k}{\left(T - t_k^{\mathrm{off}} - t_k^{dl}\right)^2} & \frac{6l_k^2}{\left(T - t_k^{\mathrm{off}} - t_k^{dl}\right)^3} & \frac{6l_k^2}{\left(T - t_k^{\mathrm{off}} - t_k^{dl}\right)^3} \\ \frac{6l_k^2}{\left(T - t_k^{\mathrm{off}} - t_k^{dl}\right)^3} & \frac{6l_k^3}{\left(T - t_k^{\mathrm{off}} - t_k^{dl}\right)^4} & \frac{6l_k^3}{\left(T - t_k^{\mathrm{off}} - t_k^{dl}\right)^4} \\ \frac{6l_k^2}{\left(T - t_k^{\mathrm{off}} - t_k^{dl}\right)^3} & \frac{6l_k^3}{\left(T - t_k^{\mathrm{off}} - t_k^{dl}\right)^4} & \frac{6l_k^3}{\left(T - t_k^{\mathrm{off}} - t_k^{dl}\right)^4} \end{bmatrix}. \tag{19}$$

The leading principal mirrors of $\mathbf{H}$ are given by

$$|\mathbf{\Delta_1}| = \frac{6l_k}{\left(T - t_k^{\mathrm{off}} - t_k^{\mathrm{dl}}\right)^2} \ge 0,$$

$$|\mathbf{\Delta_2}| = \begin{vmatrix} \frac{6l_k}{\left(T - t_k^{\mathrm{off}} - t_k^{dl}\right)^2} & \frac{6l_k^2}{\left(T - t_k^{\mathrm{off}} - t_k^{dl}\right)^3} \\ \frac{6l_k^2}{\left(T - t_k^{\mathrm{off}} - t_k^{dl}\right)^3} & \frac{6l_k^3}{\left(T - t_k^{\mathrm{off}} - t_k^{dl}\right)^4} \end{vmatrix} = 0. \tag{20}$$

□

From the above analysis, we can validate that $l_k^3/(T - t_k^{\mathrm{off}} - t_k^{dl})^2$ is convex and so is l_u^3/T^2. Hence, problem (P1.1) is convex.

In view of Lemma 1, to gain engineering insights, we next leverage the Lagrange duality method to solve problem (P1.1).

[28]

Let $\lambda_1 \in R^K{}_{\ge 0}$ and $\lambda_2 \in R^K{}_{\ge 0}$ indicate the dual variables related to the constraints in (18a) and (18b), respectively, and let $\mu_1 \in \mathrm{R} \ge 0$, $\mu_2 \in \mathrm{R}$, and $\mu_3 \in R^K{}_{\ge 0}$ be the dual variables related to the constraints in (18c), (1), and (18d), respectively.

Define $\lambda \triangleq [\lambda_1, \lambda_2]$, $\mu \triangleq [\mu_1, \mu_2, \mu_3]$, $\lambda_1 \triangleq [\lambda_{1,1}, \cdots, \lambda_{1,\mathrm{K}}]$, $\lambda_2 \triangleq [\lambda_{2,1}, \cdots, \lambda_{2,\mathrm{K}}]$, and $\mu_3 \triangleq [\mu_{3,1}, \cdots, \mu_{3,\mathrm{K}}]$. The partial

Lagrangian of problem (P1.1) is given by

$$\begin{aligned}\mathscr{L}(\mathbf{t}_{\mathbf{off}}, \mathbf{t}_{\mathbf{dl}}, \mathbf{y}_{\mathbf{off}}, \mathbf{y}_{\mathbf{dl}}, \mathbf{l}, \lambda, \mu) = \sum_{k=1}^{K}\Bigg(&(1+\mu_1)y_k^{off} + (\mu_{3,k}+1)y_k^{dl} \\ &- \mu_{3,k}E_{\max}^{\mathrm{dl}} + \frac{\gamma_k(l_k c_k)^3}{\left(T - t_k^{\mathrm{off}} - t_k^{dl}\right)^2} \\ &+ (\lambda_{1,k} - \mu_2 + \lambda_{2,k}q)l_k \\ &- \lambda_{1,k}t_k^{\mathrm{off}} r_k^{\mathrm{off}}\left(\frac{y_k^{\mathrm{off}}}{t_k^{\mathrm{off}}}\right) - \lambda_{2,k}t_k^{dl} r_k^{dl}\left(\frac{y_k^{dl}}{t_k^{dl}}\right)\Bigg) \\ &+ \frac{\gamma_u(l_u c_u)^3}{T^2} - \mu_2 l_u + \mu_2 L_u - \mu_1 E_{\max^{\mathrm{off}}}.\end{aligned} \tag{21}$$

The dual function of problem (P1.1) is expressed as

$$\begin{aligned}g(\lambda,\mu) = &\min_{\mathbf{t}_{\mathrm{off}},\mathbf{t}_{\mathrm{dl}},\mathbf{y}_{\mathrm{off}},\mathbf{y}_{\mathrm{dl}},\mathbf{l}} \mathscr{L}(\mathbf{t}_{\mathrm{off}}, \mathbf{t}_{\mathrm{dl}}, \mathbf{y}_{\mathrm{off}}, \mathbf{y}_{\mathrm{dl}}, l, \lambda, \mu) \\ &\text{s.t.} \quad (3), (11), (18e), (18f).\end{aligned} \tag{22}$$

As a result, the dual problem of problem (P1.1) is given by

$$\begin{aligned}&(\text{P1.1-dual}):\ \max_{\lambda,\mu} g(\lambda,\mu) \\ &\text{s.t.}\lambda \in \mathbb{R}_{\geq 0}^{2\mathrm{K}}, \mu_1 \geq 0, \mu_2 \in \mathbb{R}, \mu_3 \in \mathbb{R}_{\geq 0}^{\mathrm{K}}.\end{aligned} \tag{23}$$

Denote Ψ and λ^{opt} and μ^{opt} as the feasible set and the optimal dual variables for problem (P1.1-dual), respectively.

Since problem (P1.1) is convex and satisfies Slater's condition, there is zero duality gap between problems (P1.1) and (P1.1-dual) [28]. Next, we first find the dual function $g\,(\lambda,\mu)$ by solving problem (22) under any given $(\lambda,\mu) \in \Psi$ and then obtain λ^{opt} and μ^{opt} to maximize $g\,(\lambda,\mu)$.s.

4.1. Derivation of Dual Function $g\,(\lambda,\mu)$. Denote $(\mathbf{t}_{\mathrm{off}}^*, \mathbf{t}_{\mathrm{dl}}^*, \mathbf{y}_{\mathrm{off}}^*, \mathbf{y}_{\mathrm{dl}}^*, \mathbf{l}^*)$ as the optimal solution for problem (22) under any given $(\lambda,\mu) \in \Psi$, $(\mathbf{t}_{\mathrm{off}}^{\mathrm{opt}}, \mathbf{t}_{\mathrm{dl}}^{\mathrm{opt}}, \mathbf{y}_{\mathrm{off}}^{\mathrm{opt}}, \mathbf{y}_{\mathrm{dl}}^{\mathrm{opt}}, \mathbf{l}^{\mathrm{opt}})$ as the optimal primal solution for problem (P1.1), respectively. In the following, we find the dual function $g\,(\lambda,\mu)$ by solving problem (22) under any given $(\lambda,\mu) \in \Psi$. Equivalently, we decompose (22) into $K+1$ subproblems as follows:

$$\begin{aligned}(\text{P1.1-sub1}):\ &\min_{t_1^{\mathrm{off}},t_1^{\mathrm{dl}},y_1^{\mathrm{off}},y_1^{\mathrm{dl}},l_1} (1+\mu_1)y_1^{\mathrm{off}} \\ &+ (\mu_{3,1}+1)y_1^{dl} - \mu_{3,1}E_{\max^{dl}} + \frac{\gamma_1(l_1c_1)^3}{\left(T - t_1^{\mathrm{off}} - t_1^{\mathrm{dl}}\right)^2} \\ &+ (\lambda_{1,1} - \mu_2 + \lambda_{2,1}q)l_1 - \lambda_{1,1}t_1^{\mathrm{off}} r_1^{\mathrm{off}}\left(\frac{y_1^{\mathrm{off}}}{t_1^{\mathrm{off}}}\right) \\ &- \lambda_{2,1}t_1^{dl} r_1^{dl}\left(\frac{y_1^{dl}}{t_1^{dl}}\right)\end{aligned}$$

$$\text{s.t.} \quad (11), 0 \leq t_1^{\mathrm{off}} \leq T, 0 \leq t_1^{dl} \leq T, 0 \leq l_1, 0 \leq y_1^{\mathrm{off}}, 0 \leq y_1^{dl}, \vdots,$$

$$\begin{aligned}(\text{P1.1-sub}K):\ &\min_{t_K^{\mathrm{off}},t_K^{\mathrm{dl}},y_K^{\mathrm{off}},y_K^{\mathrm{dl}},l_K} (1+\mu_1)y_K^{\mathrm{off}} + (\mu_{3,K}+1)y_K^{\mathrm{dl}} \\ &- \mu_{3,K}E_{\max}^{\mathrm{dl}} + \frac{\gamma_K(l_K c_K)^3}{\left(T - t_K^{\mathrm{off}} - t_K^{dl}\right)^2} + (\lambda_{1,K} - \mu_2 + \lambda_{2,K}q)l_K \\ &- \lambda_{1,K}t_K^{\mathrm{off}} r_K^{\mathrm{off}}\left(\frac{y_K^{\mathrm{off}}}{t_K^{\mathrm{off}}}\right) - \lambda_{2,K}t_K^{dl} r_K^{dl}\left(\frac{y_K^{dl}}{t_K^{dl}}\right), \\ &s.t. \quad (11), 0 \leq t_K^{\mathrm{off}} \leq T, 0 \leq t_K^{dl} \leq T, \\ &0 \leq l_K, 0 \leq y_K^{\mathrm{off}}, 0 \leq y_K^{dl},\end{aligned}$$

$$\begin{aligned}(\text{P1.1-sub}K+1):\ &\min_{l_u} \frac{\gamma_u(l_u c_u)^3}{T^2} - \mu_2 l_u \\ &s.t. \quad (3).\end{aligned} \tag{24}$$

As these subproblems are independent of each other, they can be parallelly solved. Also, the optimal solutions for problems (P1.1-sub1)-(P1.1-sub$K$+1) are presented in Lemmas 2_1,⋯, 2_K, and 3, respectively. Note that we only show the proof of Lemma 2 since Lemmas 2_1,⋯, 2_{k-1}, 2_{k+1},⋯, 2_K, and 3 can be similarly proved via Karush-Kuhn-Tucker (KKT) conditions.

Lemma 2. *Under given $(\lambda,\mu) \in \Psi$, the optimal solution ($(y_k^{off})^*, (y_k^{dl})^*, (t_k^{off})^*, (t_k^{dl})^*, l_k^*)$ to problem (P1.1subk) satisfies*

$$\left(y_k^{off}\right)^* = \left(t_k^{off}\right)^*\left(p_k^{off}\right)^*, \tag{25}$$

$$\left(y_k^{dl}\right)^* = \left(t_k^{dl}\right)^*\left(p_k^{dl}\right)^*, \tag{26}$$

$$l_k^* = (M_k)^*\left(T - \left(t_k^{off}\right)^* - \left(t_k^{dl}\right)^*\right), \tag{27}$$

$$\left(t_k^{off}\right)^* = \begin{cases} T, & \rho_{k,1} > 0, \\ [0, T], & \rho_{k,1} = 0, \\ 0, & \rho_{k,1} < 0, \end{cases} \tag{28}$$

$$\left(t_k^{dl}\right)^* = \begin{cases} T, & \rho_{k,2} > 0, \\ [0, T], & \rho_{k,2} = 0, \\ 0, & \rho_{k,2} < 0, \end{cases} \tag{29}$$

where $(p_k^{\mathrm{off}})^* = [\lambda_{1,k}W/\ln 2(1+\mu_1) - \sigma_k^2/h_k^{\mathrm{off}}]^+$, $(p_k^{\mathrm{dl}})^* = [\lambda_{2,k}W/\ln 2(1+\mu_{3,k}) - \sigma_0^2/h_k^{\mathrm{dl}}]^+$, and

$$M_k^* = \begin{cases} \left[\sqrt{\dfrac{\mu_2 - \lambda_{1,k} - \lambda_{2,k}q}{3\gamma_k c_k^3}}\right]_0^{f_k^{\max}/c_k}, & \mu_2 - \lambda_{1,k} - \lambda_{2,k}q \geq 0, \\ 0, & \mu_2 - \lambda_{1,k} - \lambda_{2,k}q < 0, \end{cases} \tag{30}$$

$$\rho_{k,1} = \lambda_{1,k} r_k^{\mathrm{off}}\left(\left(p_k^{\mathrm{off}}\right)^*\right) - \alpha_1 \frac{f_k^{\max}}{c_k} - 2\gamma_k (c_k M_k^*)^3 - \frac{\lambda_{1,k} W\left(h_k^{\mathrm{off}}/\sigma_k^2\right)\left(p_k^{\mathrm{off}}\right)^*}{\ln 2\left(1 + \left(h_k^{\mathrm{off}}/\sigma_k^2\right)\left(p_k^{\mathrm{off}}\right)^*\right)}, \tag{31}$$

$$\rho_{k,2} = \lambda_{2,k} r_k^{\mathrm{dl}}\left(\left(p_k^{\mathrm{dl}}\right)^*\right) - \alpha_1 \frac{f_k^{\max}}{c_k} - 2\gamma_k (c_k M_k^*)^3 - \frac{\lambda_{2,k} W\left(h_k^{\mathrm{dl}}/\sigma_0^2\right)\left(p_k^{\mathrm{dl}}\right)^*}{\ln 2\left(1 + \left(h_k^{\mathrm{dl}}/\sigma_0^2\right)\left(p_k^{\mathrm{dl}}\right)^*\right)}, \tag{32}$$

$$\alpha_1 = \begin{cases} 0, & M_k^* < \dfrac{f_k^{\max}}{c_k}, \\ \mu_2 - \lambda_{1,k} - \lambda_{2,k} q - 3\gamma_k c_k^3 (M_k^*)^2, & M_k^* = \dfrac{f_k^{\max}}{c_k}. \end{cases} \tag{33}$$

Proof. Since problem (P1.1-subk) is convex and satisfies Slater's condition, there is zero duality gap between problems (P1.1-subk) and its dual problem [28]. Hence, we use the KKT conditions to solve problem (P1.1-subk). The Lagrangian function of problem (P1.1-subk) is given by

$$\begin{aligned} \mathscr{L}_k = {} & (1 + \mu_1) y_k^{\mathrm{off}} + \left(\mu_{3,k} + 1\right) y_k^{\mathrm{dl}} + \frac{\gamma_k (l_k c_k)^3}{\left(T - t_k^{\mathrm{off}} - t_k^{\mathrm{dl}}\right)^2} \\ & - \mu_2 l_k + \lambda_{1,k}\left(l_k - t_k^{\mathrm{off}} r_k^{\mathrm{off}}\left(\frac{y_k^{\mathrm{off}}}{t_k^{\mathrm{off}}}\right)\right) + \lambda_{2,k}\left(q l_k - t_k^{dl} r_k^{dl}\left(\frac{y_k^{dl}}{t_k^{dl}}\right)\right) \\ & + \alpha_1\left(l_k - \frac{\left(T - t_k^{off} - t_k^{dl}\right) f_k^{\max}}{c_k}\right) + a_1\left(t_k^{\mathrm{off}} - T\right) \\ & - a_2 t_k^{\mathrm{off}} - b_1 l_k - \beta_1 y_k^{\mathrm{off}} - \eta_1 y_k^{\mathrm{dl}} + d_1\left(t_k^{\mathrm{dl}} - T\right) - d_2 t_k^{\mathrm{dl}}, \end{aligned} \tag{34}$$

where a_1, a_2, b_1, α_1, β_1, η_1, d_1, and d_2 are the nonnegative Lagrange multipliers associated with $t_k^{\mathrm{off}} \le T0 \le t_k^{\mathrm{dl}}$, $0 \le l_k$, $l_k \le (T - t_k^{\mathrm{off}} - t_{\mathrm{k}}^{\mathrm{dl}}) f_k^{\max}/c_k$, $0 \le y_k^{\mathrm{off}}$, $0 \le y_k^{\mathrm{dl}}$, $t_k^{\mathrm{dl}} \le T$, and $0 \le t^{\mathrm{dl}}{}_k$, respectively.

According to the KKT conditions, it follows that

$$a_1\left(t_k^{\mathrm{off}} - T\right) = 0, \tag{35}$$

$$a_2 t_k^{\mathrm{off}} = 0, \tag{36}$$

$$\mathrm{b}_1 \mathrm{l}_k = 0, \tag{37}$$

$$\beta_1 y_k^{\mathrm{off}} = 0, \tag{38}$$

$$\eta_1 y_k^{\mathrm{dl}} = 0, \tag{39}$$

$$d_1\left(t_k^{\mathrm{dl}} - T\right) = 0, \tag{40}$$

$$d_2 t_k^{\mathrm{dl}} = 0, \tag{41}$$

$$a_1\left(l_k - \frac{\left(T - t_k^{\mathrm{off}} - t_k^{\mathrm{dl}}\right) f_k^{\max}}{c_k}\right) = 0, \tag{42}$$

$$\begin{aligned} \frac{\partial \mathscr{L}_k}{\partial t_k^{\mathrm{off}}} = {} & \frac{2\gamma_k (l_k c_k)^3}{\left(T - y_k^{\mathrm{off}} - y_k^{\mathrm{dl}}\right)^3} - \lambda_{1,k} W \log 2\left(1 + \frac{y_k^{\mathrm{off}}}{t_k^{\mathrm{off}}} \frac{h_k^{\mathrm{off}}}{\sigma_k^2}\right) \\ & + \frac{\lambda_{1,k} W\left(y_k^{\mathrm{off}}/t_k^{\mathrm{off}}\right)\left(h_k^{\mathrm{off}}/\sigma_k^2\right)}{\ln 2\left(1 + \left(y_k^{\mathrm{off}}/t_k^{\mathrm{off}}\right)\left(h_k^{\mathrm{off}}/\sigma_k^2\right)\right)} \\ & + \alpha_1 \frac{f_k^{\max}}{c_k} + \mathrm{a}_1 - \mathrm{a}_2 = 0, \end{aligned} \tag{43}$$

$$\begin{aligned} \frac{\partial \mathscr{L}_k}{\partial t_k^{\mathrm{dl}}} = {} & \frac{2\gamma_k (l_k c_k)^3}{\left(T - t_k^{\mathrm{off}} - t_k^{\mathrm{dl}}\right)^3} - \lambda_{2,k} W \log 2\left(1 + \frac{y_k^{\mathrm{dl}}}{t_k^{\mathrm{dl}}} \frac{h_k^{\mathrm{dl}}}{\sigma_0^2}\right) \\ & + \frac{\lambda_{2,k} W\left(y_k^{\mathrm{dl}}/t_k^{\mathrm{dl}}\right)\left(h_k^{\mathrm{dl}}/\sigma_0^2\right)}{\ln 2\left(1 + \left(y_k^{\mathrm{dl}}/t_k^{\mathrm{dl}}\right)\left(h_k^{\mathrm{dl}}/\sigma_0^2\right)\right)} + \alpha_1 \frac{f_k^{\max}}{c_k} \\ & + d_1 - d_2 = 0, \end{aligned} \tag{44}$$

$$\frac{\partial \mathscr{L}_k}{\partial y_k^{\mathrm{off}}} = (1 + \mu_1) - \frac{\lambda_{1,k} W\left(h_k^{\mathrm{off}}/\sigma_k^2\right)}{\ln 2\left(1 + \left(y_k^{\mathrm{off}}/t_k^{\mathrm{off}}\right)\left(h_k^{\mathrm{off}}/\sigma_k^2\right)\right)} - \beta_1 = 0, \tag{45}$$

$$\frac{\partial \mathscr{L}_k}{\partial y_k^{\mathrm{dl}}} = \left(1 + \mu_{3,k}\right) - \frac{\lambda_{2,k} W\left(h_k^{\mathrm{dl}}/\sigma_0^2\right)}{\ln 2\left(1 + \left(y_k^{\mathrm{dl}}/t_k^{\mathrm{dl}}\right)\left(h_k^{\mathrm{dl}}/\sigma_0^2\right)\right)} - \eta_1 = 0, \tag{46}$$

$$\frac{\partial \mathscr{L}_k}{\partial l_k} = \frac{3\gamma_k c_k^3 l_k^2}{\left(T - t_k^{\mathrm{off}} - t_k^{\mathrm{dl}}\right)^2} - \mu_2 + \lambda_{1,k} + \lambda_{2,k} q + \alpha_1 - b_1 = 0, \tag{47}$$

□

where (35)–(42) denote the complementary slackness condition, (43)–(47) are the first-order derivative conditions of L_k with respect to $t^{\mathrm{off}}{}_k$, $t^{\mathrm{dl}}{}_k$, $y_k{}^{\mathrm{off}}$, $y_k{}^{\mathrm{dl}}$, and l_k, respectively. Hence, we have (27) and (28) based on (45) and (46), respectively, and (29) holds due to (47). In addition, we have (29) based on (47) and some manipulations.

Also, by substituting (25) and (27) into (43) and (26) and (27) into (44) and assuming $\rho_{k,1} = a_1 - a_2$ and $\rho_{k,2} = d_1 - d_2$, we have $\rho_{k,1}$ and $\rho_{k,2}$ in (31) and (28), respectively. Consequently, the optimal $\left(t_k^{\mathrm{off}}\right)^*$ and $\left(t_k^{\mathrm{dl}}\right)^*$ are given in (28) and (29), respectively.

Lemma 3. *Under given $(\lambda, \mu) \in \Psi$, the optimal solution l_u^* to problem (P1.1-subK+1) is*

$$l_u^* = \left[T\sqrt{\frac{\mu_2}{3\gamma_u c_u^3}} \right]_0^{Tf_u^{\max}/c_u}. \tag{48}$$

Remark 4. Note that in (28) and (29), if $\rho_{k,i} = 0$ (for any $i \in \{1, 2\}$), the optimal solution $(t_k^{\text{off}})^*$ or $(t_k^{\text{dl}})^*$ is generally nonunique. In this case, we choose $(t_k^{\text{off}})^* = 0$ or $(t_k^{\text{dl}})^* = 0$ for evaluating $g(\lambda, \mu)$. Since such choice may not be feasible or optimal for problem (P1.1), we add an additional step to find the primal optimal $(t_k^{\text{off}})^{\text{opt}}$ and $(t_k^{\text{dl}})^{\text{opt}}$, as will be shown in Section 4.3.

By combining Lemmas $2_1,\cdots, 2_K$ with 3, $g(\lambda,\mu)$ is evaluated for any given $(\lambda, \mu) \in \Psi$.

4.2. Obtaining λ^{opt} and μ^{opt} to Maximize $g(\lambda, \mu)$. With ($\mathbf{t}_{\text{off}}^*, \mathbf{t}_{\text{dl}}^*, \mathbf{y}_{\text{off}}^*, \mathbf{y}_{\text{dl}}^*, \mathbf{l}^*$) obtained, we then solve problem (P1.1-dual) to maximize $g(\lambda, \mu)$. Due to the property of $g(\lambda, \mu)$, the ellipsoid method is utilized to obtain $(\lambda^{\text{opt}}, \mu^{\text{opt}})$ [29]. For the objective function in (22), one subgradient is given by [29]

$$\begin{aligned}\mathbf{e} = \Bigg[& l_1^* - \left(t_1^{\text{off}}\right)^* r_1^{\text{off}}\left(\frac{\left(y_1^{\text{off}}\right)^*}{\left(t_1^{\text{off}}\right)^*}\right),\cdots,l_K^* - \left(t_K^{\text{off}}\right)^* r_K^{\text{off}}\left(\frac{\left(y_K^{\text{off}}\right)^*}{\left(t_K^{\text{off}}\right)^*}\right), ql_1^* \\ & - \left(t_1^{\text{dl}}\right)^* r_1^{\text{dl}}\left(\frac{\left(y_1^{\text{dl}}\right)^*}{\left(t_1^{\text{dl}}\right)^*}\right),\cdots,ql_K^* - \left(t_K^{\text{dl}}\right)^* r_K^{\text{dl}}\left(\frac{\left(y_K^{\text{dl}}\right)^*}{\left(t_K^{\text{dl}}\right)^*}\right), \sum_{k=1}^{K}\left(y_k^{\text{off}}\right)^* \\ & - E_{\max}^{\text{off}}, L_u - l_u^* - \sum_{k=1}^{K} l_k^*\left(y_1^{\text{dl}}\right)^* - E_{\max}^{\text{dl}},\cdots,\left(y_K^{\text{dl}}\right)^* - E_{\max}^{\text{dl}} \Bigg].\end{aligned} \tag{49}$$

For the constraints $\lambda \in \mathbb{R}_{\geq 0}^{2K}, \mu_1 \geq 0, \mu_2 \in \mathbb{R}$ and $\mu_3{}^{\in} R^K{}_{\geq 0}$, the subgradients are $e_1,\cdots,e_{2K}$, e_{2K+1}, e_{2K+2}, and $e_{2K+3},\cdots,e_{3K+2}$, respectively. Note that e_i is of all zero entries except for the i -th entry being one.

4.3. Finding the Optimal Primal Solution to (P1). Having obtained λ^{opt},μ^{opt}, we still need to further solve problem (P1.1). By replacing (λ,μ) with λ^{opt},μ^{opt} in Lemmas $2_1,\cdots, 2_K$, and 3, we obtain the corresponding $\mathbf{p}_{\text{off}}^{\text{opt}} = [(\mathbf{p}_1^{\text{off}})^{\text{opt}}, L, (\mathbf{p}_K^{\text{off}})^{\text{opt}}]$, $\mathbf{p}_{\text{dl}}^{\text{opt}} = [(\mathbf{p}_1^{\text{dl}})^{\text{opt}},\cdots,(\mathbf{p}_K^{\text{dl}})^{\text{opt}}]$, $M^{\text{opt}} = [M_1^{\text{opt}},\cdots,M_K^{\text{opt}}]$, and $l_u{}^{\text{opt}}$, respectively. However, due to the nonuniqueness of $(t_k^{\text{off}})^*$ and $(t_k^{\text{dl}})^*, k \in \mathscr{K}$, we implement an extra procedure to obtain the optimal solution of other variables for problem (P1). With $\mathbf{p}_{\text{off}}^{\text{opt}}$, $\mathbf{p}_{\text{dl}}^{\text{opt}}$, $\mathbf{M}^{\text{opt}}$, and l_u^{opt}, the optimal solution must satisfy $l_k = M_k^{\text{opt}}(T - t_k^{\text{off}} - t_k^{\text{dl}})$, $y_k^{\text{off}} = t_k^{\text{off}} (p_k^{\text{off}})^{\text{opt}}$, and $y_k^{\text{dl}} = t_k^{\text{dl}}(p_k^{\text{dl}})^{\text{opt}}$. By substituting them in (P1.1), we have the following linear program (LP) to obtain $\mathbf{t}_{\text{off}}^{\text{opt}}$ and $\mathbf{t}_{\text{dl}}^{\text{opt}}$:

$$\begin{gathered}\min_{t_{\text{off}},t_{\text{dl}}} \sum_{k=1}^{K}\left(t_k^{\text{off}}\left(p_k^{\text{off}}\right)^{\text{opt}} + t_k^{\text{dl}}\left(p_k^{\text{dl}}\right)^{\text{opt}} + \gamma_k\left(M_k^{\text{opt}}c_u\right)^3\left(T - t_k^{\text{off}} - t_k^{\text{dl}}\right)\right), \\ \text{s.t.} M_k^{\text{opt}}\left(T - t_k^{\text{off}} - t_k^{\text{dl}}\right) \leq t_k^{\text{off}} r_k^{\text{off}}\left(\left(p_k^{\text{off}}\right)^{\text{opt}}\right), \forall k \in \mathscr{K}, \\ qM_k^{\text{opt}}\left(T - t_k^{\text{off}} - t_k^{\text{dl}}\right) \leq t_k^{\text{dl}} r_k^{\text{dl}}\left(\left(p_k^{\text{dl}}\right)^{\text{opt}}\right), \forall k \in \mathscr{K}, \\ \sum_{k=1}^{K} t_k^{\text{off}}\left(p_k^{\text{off}}\right)^{\text{opt}} \leq E_{\max}^{\text{off}}, \\ t_k^{\text{dl}}\left(p_k^{\text{dl}}\right)^{\text{opt}} \leq E_{\max}^{\text{dl}}, \forall k \in \mathscr{K}, \\ \sum_{k=1}^{K} M_k^{\text{opt}}\left(T - t_k^{\text{off}} - t_k^{\text{dl}}\right) + l_u^{\text{opt}} = L_u, \\ t_k^{\text{off}} + t_k^{\text{dl}} \leq T, \forall k \in \mathscr{K}, \\ 0 \leq t_k^{\text{off}}, 0 \leq t_k^{\text{dl}}, \forall k \in \mathscr{K}.\end{gathered} \tag{50}$$

Since problem (50) is an instance of LP, it can be solved by the interior-point method [28]. Finally, we obtain the globally optimal solution for problem (P1). The proposed joint task partition and resource allocation (JTPRA) scheme is thus summarized in Algorithm 1.

Remark 5. With Lemmas $2_1,\cdots, 2_K$, and 3, the following insights can be obtained as follows:

(1) As for local computing, it is observed from Lemma 3 that $l_u{}^{\text{opt}}$ generally increases as T becomes large. This indicates that the user prefers executing more tasks when the user's computation latency constraint becomes loose

(2) As for cooperative computing, it is evident that, based on Lemma 2, the offloading power $(p_k^{\text{off}})^{\text{opt}}$ increases as the channel power gain $h^{\text{off}}{}_k$ becomes stronger. That is, the user prefers offloading more tasks to the closer helper, in order to reduce the marginal energy consumption for offloading. Similarly, this also applies to $(p_k^{\text{dl}})^{\text{opt}}$

4.4. Complexity. The complexity of the ellipsoid method is $O(N^2)$, where N is the number of dual variables and $N = 3K + 2$ in (23) [29]. Moreover, the complexity of the interior-point method is $O(M^{3.5} \log (1/\varepsilon))$ where M is the number of optimal variables, $\log (1/\varepsilon)$ is the iteration complexity order, and $M = 2K$ is in (50) [28]. Hence, the total complexity of Algorithm 1 is $O(K^{3.5} \log (1/1\varepsilon - \varepsilon))$.

5. Simulation Results

We provide simulation results for verifying the effectiveness of the proposed joint task partition and resource allocation

```
Initialization: Given an ellipsoid $\varepsilon((\boldsymbol{\lambda},\boldsymbol{\mu}),\mathbf{A})$ containing $(\boldsymbol{\lambda}^{\mathbf{opt}},\boldsymbol{\mu}^{\mathbf{opt}})$, where $(\boldsymbol{\lambda},\boldsymbol{\mu})$ is the center point of $\varepsilon$ and $\mathbf{A} \succ 0$ characterizes the volume of $\varepsilon$.
repeat
    Obtain $\mathbf{p}^*_{\mathbf{off}}$, $\mathbf{p}^*_{\mathbf{dl}}$, $\mathbf{y}^*_{\mathbf{off}}$, $\mathbf{y}^*_{\mathbf{dl}}$, $\mathbf{t}^*_{\mathbf{off}}$, and $\mathbf{t}^*_{\mathbf{dl}}$ by Lemmas $2_1, \cdots, 2_K$ and 3 under given $(\boldsymbol{\lambda},\boldsymbol{\mu}) \in \Psi$, respectively;
    Compute the subgradients of $g\,(\lambda,\mu)$, then update $(\boldsymbol{\lambda},\boldsymbol{\mu})$ using the ellipsoid method [29];
until $(\boldsymbol{\lambda},\boldsymbol{\mu})$ converge to a specified accuracy.
Set$(\lambda^{\mathbf{opt}}, \mu^{\mathbf{opt}}) \longleftarrow (\lambda, \mu)$.
Output: Obtain $\mathbf{p}^{\mathbf{opt}}_{\mathbf{off}}$,$\mathbf{p}^{\mathbf{opt}}_{\mathbf{dl}}$, and $l^{opt}_u$ based on $2_1$,⋯,$2_K$ and 3 by replacing $(\boldsymbol{\lambda},\boldsymbol{\mu})$ with $\lambda^{opt}$,$\mu^{\mathbf{opt}}$, and then compute $\mathbf{t}^{\mathbf{opt}}_{\mathbf{off}}$, $\mathbf{t}^{\mathbf{opt}}_{\mathbf{dl}}$, and $\mathbf{l}^{\mathbf{opt}}$ by solving the LP in (50).
```

Algorithm 1

(JTPRA) scheme in Section 4, as compared against the following five benchmark schemes:

(1) Local Computing with Optimal Frequency (LCOF): the computation tasks are executed locally with the optimal CPU frequency, and thus the optimal energy consumption for local computing is $E^{\text{opt}}_{\text{local}} = \gamma_u (L_u c_u)^3 / T^2$

(2) Local Computing with Fixed Frequency (LCFF): the computation tasks are executed locally with the maximum CPU frequency, and thus the energy consumption for local computing is $E^{\text{fixed}}_{\text{local}} = \gamma_u L_u c_u (f^{\max}_u)^2$

(3) Full Offloading (FO): the computation tasks are partitioned into K parts for offloading to nearby helpers, which corresponds to solving problem (P1) by setting $l_u = 0$

(4) Joint Offloading Ratio and CPU Frequency (JORCF) [7]: in this scheme, the user adjusts both the offloading ratio and CPU frequency to cooperate with the MEC server

(5) Fixed Frequency (FF) [20]: let the constraints (3) and (11) be met with strict equality. This corresponds to solving problem (P1) by setting $f_u = f^{\max}_u$ and $f_k = {f_k}^{\max}$, $k \in K$

In the simulation, the distance between the user and helpers is $d \sim U\ [d_{\min}, d_{\max}]$ meters, where $d_{\max} = 30$ meters and $d_{\min} = 1$ meters. The path loss between any two nodes is modeled as $bd^{-\varphi}$, where $b = 10^{-3}$ corresponds to the path loss at a reference distance of 1 meter, d denotes the distance from the user to a helper, and $\varphi = 3$ is the path loss exponent [21]. Also, the helpers' maximum CPU frequencies are assumed to be uniformly chosen from the set {1.6, 2.4, 3}GHz. The other parameters are set as shown in Table 1 unless otherwise specified.

Table 1: Simulation parameters.

Frequency resource block W	1 MHz
Effective capacitance coefficient $\gamma_u = \gamma_1 = \cdots = \gamma_K$	3 × 10 − 27
Noise power $\sigma_0^2 = \sigma_1^2 = \cdots = \sigma_K^2$	-120 dBm [22]
Computation intensities $c_u = c_1 = \cdots = c_K$	10^3 cycle/bit [21]
User's maximum CPU frequency ${f_u}^{\max}$	2 GHz [21]
Available energy for data transmission $E^{\text{off}}_{\max} = E^{\text{dl}}_{\max}$	0.5 joule
Normalized ratio between the size of computation results and the size of computing tasks q	0.2
Number of channel realizations	500

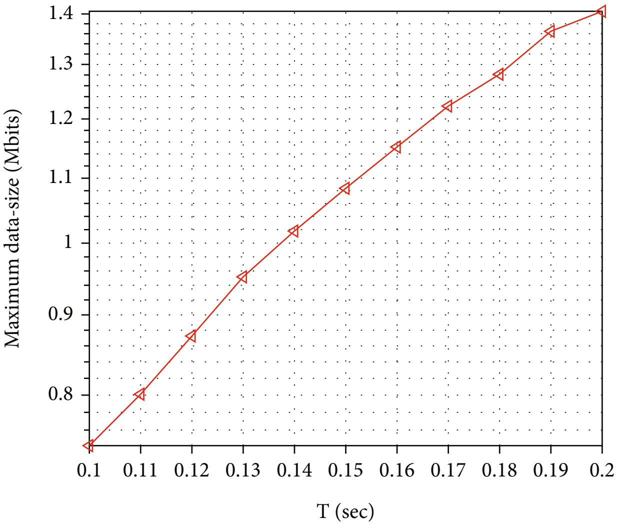

Figure 3: The maximum data-size versus the block length T.

Figure 3 shows the maximum data-size versus the block length T where $L_u = 0.2$ Mb and $K = 3$. In the following simulation, under given T, we set L_u smaller than $L_{\max}$ to guarantee the feasibility of problem (P1).

Figure 4 shows the average energy consumption versus data size L_u where $T = 0.15$ sec and $K = 3$. It is observed that our proposed JTPRA achieves the minimum energy consumption than other schemes. Moreover, we have some observations as follows.

(1) The average energy consumption by all the schemes increases as L_u increases. JTPRA achieves significant performance gain over FF. This indicates the benefit of computation frequency optimization in energy saving for MEC

(2) For schemes with unchanged CPU frequency, FF achieves a lower energy consumption than LCFF. This is because the user prefers offloading the computing tasks to the helpers whose maximum CPU frequencies are below that of the user, compared with local computing

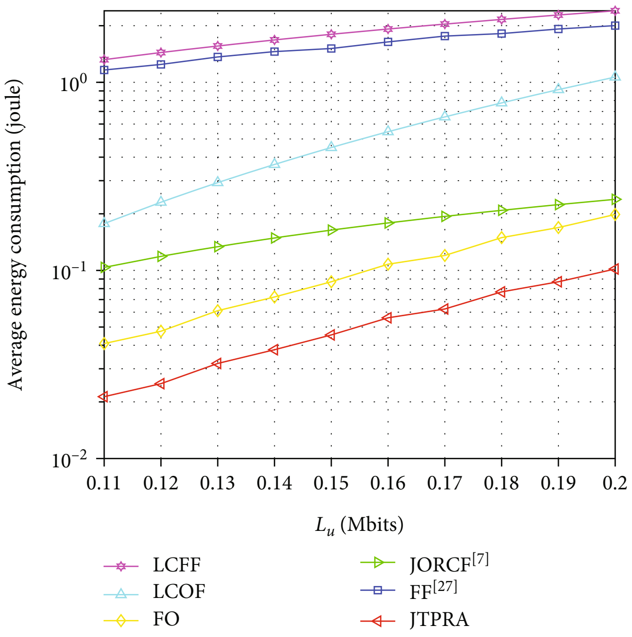

FIGURE 4: The average energy consumption versus data size L_u.

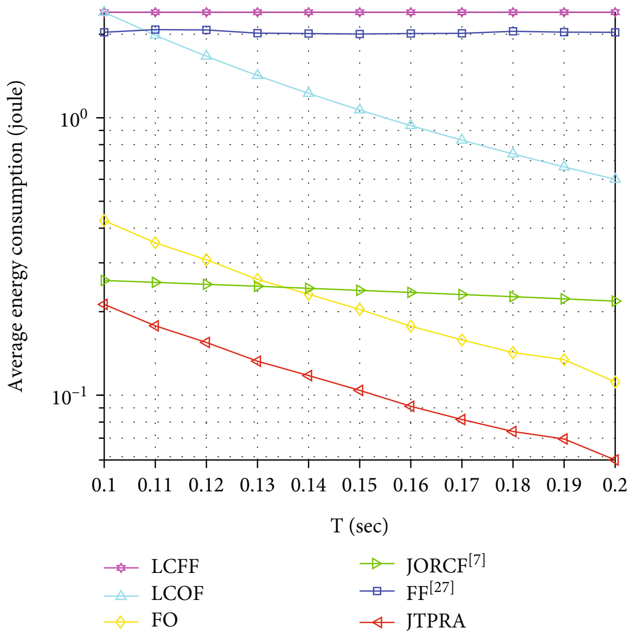

FIGURE 5: The average energy consumption versus the block length T.

(3) For schemes involving optimizing CPU frequency, FO achieves significant energy reduction than LCOF, which is because the helpers' optimal CPU frequencies are far below than that of the user. Also, by comparison with JORCF, JTPRA has about 69.9% energy consumption reduction on average. This indicates the performance gain brought by proximity

Figure 5 shows the average energy consumption versus the block length T where $L_u = 0.2$ Mb and $K = 3$. We have generally similar observations in Figure 5 as in Figure 4. Specifically, it is observed that our proposed JTPRA has about 52.4% energy consumption reduction on average, compared with JORCF. Moreover, we have some observations as follows. (1) For schemes with unchanged CPU frequency, LCFF remains unchanged as T increases, while FF keeps almost unchanged. This indicates there is no need for the user and helpers to increase the transmission rate when the latency requirement is loose.

(2) For schemes involving optimizing CPU frequency, the average energy consumption by all the schemes decreases as T increases, which is because as T increases, the user and helpers can lower down their optimal CPU frequency for consuming less energy. However, JORCF decreases with T very slowly. This indicates fixing the transmission rate is not conducive to improve the MEC performance.

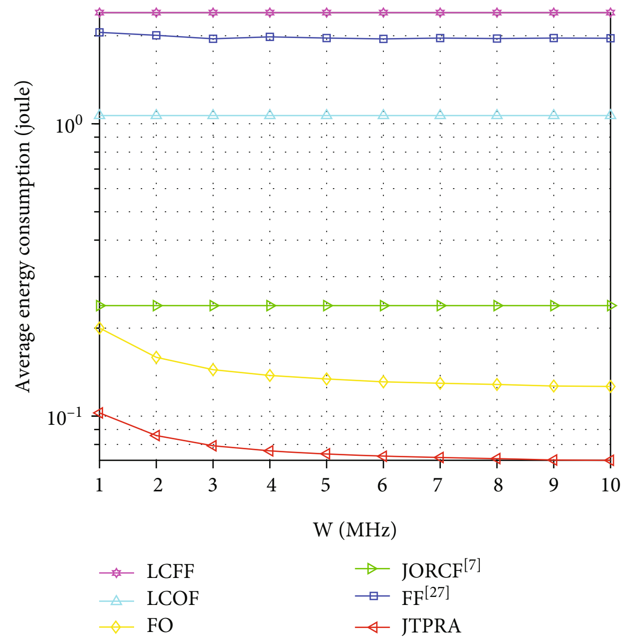

FIGURE 6: The average energy consumption versus the frequency resource block W.

Figure 6 shows the average energy consumption versus the frequency resource block W where $T = 0.15$ sec, $L_u = 0.2$ Mb, and $K = 3$. We have generally similar observations in Figure 6 as in Figure 5. Specifically, it is observed that our proposed JTPRA has about 67.5% energy consumption reduction on average, compared with JORCF. Moreover, for schemes involving optimizing CPU frequency other than LCOF and JORCF, the average energy consumption first steadily decreases and then keeps almost unchanged as W increases. This is because a large W not only signifies a high transmission rate but incurs decreased transmission energy consumption between the user and the helpers.

Figure 7 shows the average energy consumption versus the helper number K where $L_u = 0.2$ Mb and $T = 0.15$ sec.

We have generally similar observations in Figure 7 as in Figure 6. Specifically, it is observed that our proposed JTPRA has about 65.1% energy consumption reduction on average,

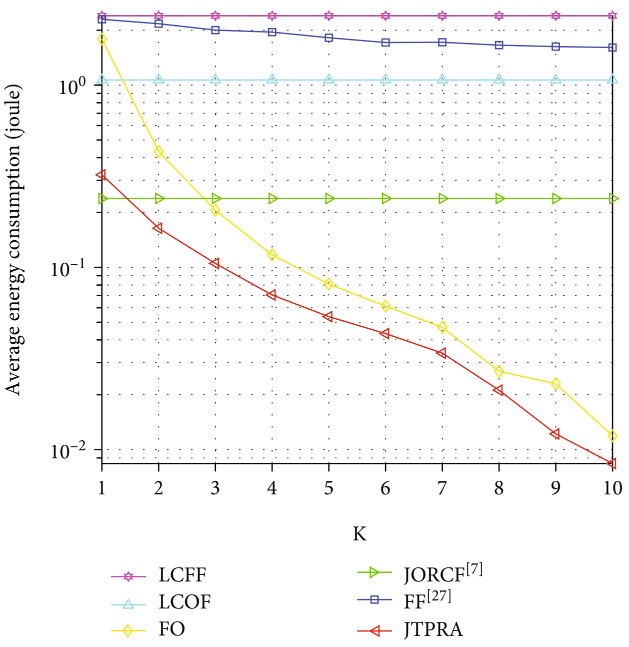

Figure 7: The average energy consumption versus the helper number K.

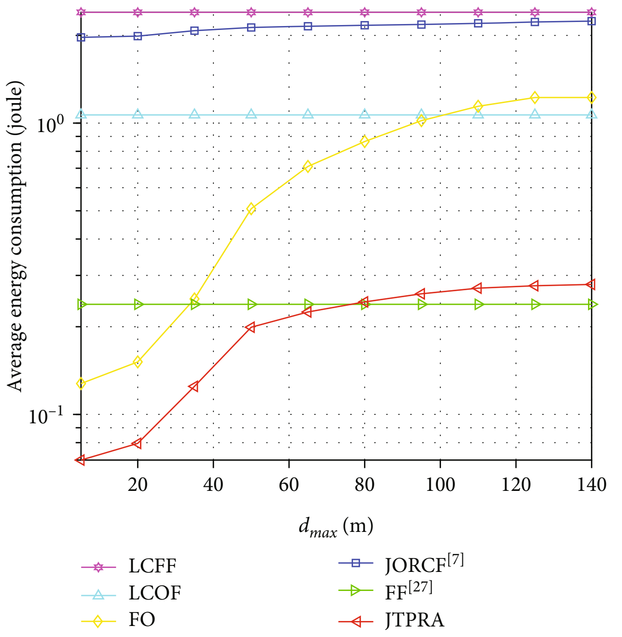

Figure 8: The average energy consumption versus the maximum distance $d_{\max}$ from the user to the helpers.

compared with JORCF. Obviously, for schemes with unchanged CPU frequency, FF decreases as K becomes large. In addition, for schemes involving optimizing CPU frequency, JTPRA achieves a lower energy consumption than both LCOF and FO, while when $K < 2$, JORCF achieves a lower energy consumption than JTPRA, while the reverse is true when Kbecomes large. This is because the more helpers whose optimal CPU frequencies are below that of the user are helpful for achieving more significant energy reduction.

Figure 8 shows the average energy consumption versus the maximum distance $d_{\max}$ from the user to the helpers where $L_u = 0.2$ Mb, $T = 0.15$ sec, and $K = 3$. We have generally similar observations in Figure 8 as in Figure 7. Specifically, it is observed that our proposed JTPRA has about 15.1% energy consumption reduction on average, compared with JORCF. For schemes involving optimizing CPU frequency, FO first grows rapidly as $d_{\max}$ increases and then becomes even worse than both JORCF and LCOF, while JTPRA steadily increases as $d_{\max}$ increases. This is because as $d_{\max}$ increases, the channel gain between the user and the helpers becomes smaller, which leads to increased transmission energy consumption between the user and the helpers, and thus local computing is more beneficial than computation offloading at large $d_{\max}$ values.

6. Conclusion

In this paper, we have proposed a novel joint task partition and resource allocation (JTPRA) scheme, in which nearby helpers share their own communication and computation resources to help the user. By considering an efficient OFDMA-aided three-phase transmission protocol, we proposed an energy-efficient design framework by jointly optimizing the user's task partition, and the communication and computation resources allocation for task offloading and results downloading, subject to the user's computation latency constraint. Based on convex optimization methods, we presented an efficient algorithm to obtain the globally optimal solution. Extensive numerical results demonstrated the merits of the proposed JTPRA scheme over alternative benchmark schemes.

Due to space limitation, there are some other challenging problems to be handled in this paper, which are investigated as follows to inspire future work.

(1) Although this paper considered single-user multihelper model, our results are extendable to more general ones with multiuser multihelper. In this case, we can design helper selection policy to pair each user with one or multiple helpers, such that the helpers can use the proposed JTPRA scheme to help the computation of the paired user. However, how to efficiently handle the joint optimization problem of helper selection and resource allocation is a quite challenging problem worthy of further study

(2) Due to easy implementation of OFDMA, we designed the proposed protocol based on it in this paper. To further improve the system performance, we can next exploit other orthogonal multiple access schemes, e.g., NOMA schemes [22] and sparse code multiple access (SCMA) [30]

(3) In terms of energy saving, we achieved the expected goal. But for MEC standardization, how to improve the propose scheme's implementation like D2D, e.g., symbol synchronization and signaling interaction, is a difficult problem worth pursuing in the future [31]

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest in the work.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under grant No. 61531007.

References

[1] M. Chiang and T. Zhang, "Fog and IoT: an overview of research opportunities," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 854–864, 2016.

[2] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, and V. Young, "Mobile edge computing: a key technology towards 5G," *ETSI White Paper*, vol. 11, no. 11, pp. 1–16, 2015.

[3] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: the communication perspective," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2322–2358, 2017.

[4] Y. Mao, J. Zhang, and K. B. Letaief, "Joint task offloading scheduling and transmit power allocation for mobile-edge computing systems," in *2017 IEEE wireless communications and networking conference (WCNC)*, pp. 1–6, San Francisco, CA, USA, March 2017.

[5] D. Han, W. Chen, and Y. Fang, "Joint channel and queue aware scheduling for latency sensitive mobile edge computing with power constraints," *IEEE Transactions on Wireless Communications*, vol. 19, no. 6, pp. 3938–3951, 2020.

[6] Y. Mao, J. Zhang, and K. B. Letaief, "Dynamic computation offloading for mobile-edge computing with energy harvesting devices," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3590–3605, 2016.

[7] W. Yoo, W. Yang, and J. -M. Chung, "Energy consumption minimization of smart devices for delay-constrained task processing with edge computing," in *2020 IEEE International Conference on Consumer Electronics (ICCE)*, pp. 1–3, Las Vegas, NV, USA, January 2020.

[8] M. Liu and Y. Liu, "Price-based distributed offloading for mobile-edge computing with computation capacity constraints," *IEEE Wireless Communications Letters*, vol. 7, no. 3, pp. 420–423, 2018.

[9] F. Wang, J. Xu, X. Wang, and S. Cui, "Joint offloading and computing optimization in wireless powered mobile-edge computing systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 3, pp. 1784–1797, 2018.

[10] S. Bi and Y. J. Zhang, "Computation rate maximization for wireless powered mobile-edge computing with binary computation offloading," *IEEE Transactions on Wireless Communications*, vol. 17, no. 6, pp. 4177–4190, 2018.

[11] H. Sun, F. Zhou, and R. Q. Hu, "Joint offloading and computation energy efficiency maximization in a mobile edge computing system," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 3052–3056, 2019.

[12] X. Huang, S. Zeng, D. Li, P. Zhang, S. Yan, and X. Wang, "Fair computation efficiency scheduling in NOMA-aided mobile edge computing," *IEEE Wireless Communications Letters*, vol. 9, no. 11, pp. 1812–1816, 2020.

[13] J. Xu and J. Yao, "Exploiting physical-layer security for multiuser multicarrier computation offloading," *IEEE Wireless Communications Letters*, vol. 8, no. 1, pp. 9–12, 2019.

[14] F. Wang, J. Xu, and Z. Ding, "Multi-Antenna NOMA for computation offloading in multiuser mobile edge computing systems," *IEEE Transactions on Communications*, vol. 67, no. 3, pp. 2450–2463, 2019.

[15] T. X. Tran and D. Pompili, "Joint task offloading and resource allocation for multi-server mobile-edge computing networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 856–868, 2019.

[16] Y. Wang, Y. Zhang, M. Sheng, and K. Guo, "On the interaction of video caching and retrieving in multi-server mobile-edge computing systems," *IEEE Wireless Communications Letters*, vol. 8, no. 5, pp. 1444–1447, 2019.

[17] C. Guo, W. He, and G. Y. Li, "Optimal fairness-aware resource supply and demand management for mobile edge computing," *IEEE Wireless Communications Letters*, vol. 10, no. 3, pp. 678–682, 2021.

[18] K. Li, M. Tao, and Z. Chen, "Exploiting computation replication for mobile edge computing: a fundamental computation-communication tradeoff study," *IEEE Transactions on Wireless Communications*, vol. 19, no. 7, pp. 4563–4578, 2020.

[19] Q. Lin, F. Wang, and J. Xu, "Optimal task offloading scheduling for energy efficient d2d cooperative computing," *IEEE Wireless Communications Letters*, vol. 23, no. 10, pp. 1816–1820, 2019.

[20] L. Wang, M. Guan, Y. Ai, Y. Chen, B. Jiao, and L. Hanzo, "Beamforming-Aided NOMA expedites collaborative multiuser computational offloading," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 10, pp. 10027–10032, 2018.

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

[22] Y. Huang, Y. Liu, and F. Chen, "NOMA-Aided Mobile edge computing via user cooperation," *IEEE Transactions on Wireless Communications*, vol. 68, no. 4, pp. 2221–2235, 2020.

[23] R. Chai, J. Lin, M. Chen, and Q. Chen, "Task execution cost minimization-based joint computation offloading and resource allocation for cellular D2D MEC systems," *IEEE Systems Journal*, vol. 13, no. 4, pp. 4110–4121, 2019.

[24] Y. He, J. Ren, G. Yu, and Y. Cai, "D2D communications meet mobile edge computing for enhanced computation capacity in cellular networks," *IEEE Transactions on Wireless Communications*, vol. 18, no. 3, pp. 1750–1763, 2019.

[25] Y. Li, G. Xu, J. Ge, P. Liu, X. Fu, and Z. Jin, "Jointly optimizing helpers selection and resource allocation in D2D mobile edge computing," in *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, Seoul, Korea (South), May 2020.

[26] S. S. Yilmaz and B. Ozbek, "Multi-helper NOMA for cooperative¨ mobile edge computing," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, 2021.

[27] H. Xing, L. Liu, J. Xu, and A. Nallanathan, "Joint task assignment and resource allocation for d2d-enabled mobile-edge computing," *IEEE Transactions on Communications*, vol. 67, no. 6, pp. 4193–4207, 2019.

[28] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge Univ. Press, Cambridge, U.K, 2004.

[29] S. Boyd, *Ellipsoid Method*, Stanford Univ, Sacramento, CA, USA, 2008.

[30] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: research challenges and future trends," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 10, pp. 2181–2195, 2017.

[31] K. Doppler, M. Rinne, C. Wijting, C. B. Ribeiro, and K. Hugl, "Device-to-device communication as an underlay to lte-advanced networks," *IEEE Communications Magazine*, vol. 47, no. 12, pp. 42–49, 2009.