Green Task Offloading with Bandit Learning in Fog Computing Systems

Xin Gao, Xi Huang, Ziyu Shao
School of Information Science and Technology, ShanghaiTech University
Email: {gaoxin, huangxi, shaozy}@shanghaitech.edu.cn

***Abstract*—In fog computing systems, to improve the quality-of-service (QoS) of computation-intensive IoT applications on resource-limited IoT devices, tasks are offloaded to nearby fog nodes to achieve lower latency. However, it is still challenging to make timely and energy-efficient offloading decision about *whether* to offload and *where* to offload for each incoming task. The challenges come down to: how to manage the tradeoff between task latency and energy consumption; how to conduct effective learning in face of environment uncertainties such as processing capacities and transmission rates; and how to integrate the learning procedure with online control so that the incurred performance loss, *a.k.a.*, the *regret*, can be minimized. In this paper, by employing Lyapunov optimization techniques and bandit learning methods, we propose *LAGO* (Learning-Aided Green Offloading), an efficient online algorithm which aims to minimize the average task latency under long-term energy budget constraints. Our theoretical analysis and simulation results show that LAGO effectively reduces the time-average task latency with an $O(\alpha_1/V + \alpha_2\sqrt{\log T/T})$ regret bound over time horizon T while ensuring the long-term energy constraints, where α_1 and α_2 are positive constants and V is a tunable parameter.**

I. Introduction

In recent years, the rapid development of Internet of Things (IoT) technology has spawned a wealth of applications that require intensive computation in real-time [1]. Such applications constantly generate tasks to be processed on resource-constrained IoT devices. To improve the system performance and QoS in terms of energy efficiency and task latency, the IoT applications often resort to the assistance of external platforms such as recently proposed fog computing systems [2], by *offloading* tasks from IoT devices, *a.k.a.*, *user nodes*, to their nearby processing nodes with enhanced computing and storage capacity, *a.k.a. fog nodes*, through wireless connections.

However, it is challenging to develop an effective online offloading scheme, by which each user node must decide *which* task to be offloaded to *which* of its accessible fog node. The challenges lie in three aspects. The *first* is to handle the *tradeoff* between task latency and energy consumption. The advantage of offloading comes from the powerful processing capacity of fog nodes, thereby resulting in shorter task latency. Nonetheless, the potential benefit of task offloading can be offset by induced considerable energy consumption of task transmission. In face of constantly generated tasks with heterogeneous sizes and computation demands, such a tradeoff must be handled even more carefully. The *second* is about minimizing the impact of various uncertainties of system dynamics on the overall system performance. For example, the processing capacity and the wireless channel state can vary in a wide range among fog nodes and user nodes. Such exact information is instructive for making offloading decisions but usually remains unknown until revealed by feedback after tasks' actual completion. Therefore, the control procedure of task offloading must proceed with the aid of particular learning procedure that effectively learns the statistics about such uncertainties from collected feedback. In the process of learning, the well-known *exploration-exploitation* tradeoff must be carefully treated, since 1) over-exploitation, *i.e.*, user nodes stick to offloading tasks to particular fog nodes, may prevent user nodes from collecting informative feedback from other potentially better fog nodes; 2) over-exploration, *i.e.*, user nodes blindly switch among different fog nodes, may induce excessive delay and energy consumption. The *third* is about properly dealing with the interplay between the control and learning procedure, because ineffective learning can produce inaccurate estimates and misguide the control procedure, while control procedure, if improperly conducted, can lead to noisy feedback that misinforms the learning procedure.

So far, existing works basically follow two lines to develop task offloading schemes. One line of works [3] [4] typically formulate the task offloading problem as a stochastic optimization problem and adopt Lyapunov optimization techniques [5] to transform it into a sequence of subproblems, then develop online algorithms to solve them on a time-slot basis. Works in this line often implicitly assume the instant system dynamics to be readily given at the beginning of each time slot, which could be invalid in practice as aforementioned. The other line of works [6] consider the cases where some system dynamics such as the node processing capacities are unknown; then they apply bandit methods to learn the statistics of the uncertainties to make offloading decisions. However, such methods are only able to handle the total energy budget constraint, but fail to conduct fine-grained control to maintain the long-term energy constraint on each of the user and fog nodes.

In this paper, we combine the ideas of the above two lines of works to address the aforementioned challenges. Particularly, we focus on the task offloading problem between user nodes, *i.e.*, IoT devices, and fog nodes with uncertainties of system dynamics. Our contributions and key results are as follows:

• **Modeling and Problem Formulation:** We develop an energy-constrained fog computing system model with unknown node processing capacities and link transmission rates. Based on the model, we formulate a task offloading prob-

lem to minimize the average task latency under long-term energy constraints. The problem consists of two parts, one is a stochastic control problem which ensures the energy constraints, the other is an extended stochastic combinatorial multi-armed bandit (CMAB) problem.

• **Algorithm Design:** By adopting Lyapunov optimization technique [5] and UCB (upper confidence bound) method [7] to handle the tradeoff between online control and learning, we propose *LAGO*, an energy-efficient online task offloading scheme that makes effective task offloading decisions under long-term energy constraints.

• **Theoretical Analysis:** Our theoretical analysis shows that LAGO achieves a tunable latency-energy tradeoff while guaranteeing the energy constraints, and has a regret bound of the order $O(\alpha_1/V + \alpha_2\sqrt{\log T/T})$.

• **Numerical Evaluation:** We conduct extensive simulations to evaluate the latency-energy tradeoff under LAGO, and compare the performance of LAGO with its variants that have different choices of UCB methods. Simulation results verify the effectiveness of LAGO in task latency reduction and energy efficiency guarantee.

The rest of this paper is organized as follows. We present our system model in Section II and problem formulation in Section III. Next we show the algorithm design in Section IV and performance analysis in Section V. Section VI shows our simulation results, while Section VII concludes the paper.

II. System Model

A. Basic Model

We consider a time-slotted fog computing system and focus on the interplay between one given IoT user node and N fog nodes. [1] We denote by set $\mathcal{N} = \{0, 1, \ldots, N\}$ as the set of all nodes, including the user node which is indexed by 0, and the fog nodes indexed by other non-zero indices from 1 to N. During each time slot t, due to wireless channel state variation, the user node has access to only a subset of the fog nodes. For each time slot $t \in \{0, 1, \ldots\}$, we denote the set of nodes accessible from the user node (including itself) by $\mathcal{N}(t) \subseteq \mathcal{N}$.

The system proceeds as follows. At the beginning of each time slot t, the user node generates a number of tasks to be processed, denoted by set $\mathcal{A}(t)$. We assume that the task generation process is *i.i.d.* over time slots and satisfies that $|\mathcal{A}(t)| \leq a_{\max}$ for some constant $a_{\max}$. Each task $i \in \mathcal{A}(t)$ has a size of $L_i(t)$ (in the unit of bits) and a computation demand of $W_i(t)$ CPU cycles, which are assumed known upon the task's arrival and upper bounded by constant $l_{\max}$ and $w_{\max}$, respectively. Besides, each task has a deadline $\tau_{\max}$; if a task's total latency exceeds the deadline, it will be considered failed.

Then the user node should decide, for each of such tasks, whether to offload it and which node it is offloaded to. For ease of analysis, we assume that each task can be either processed locally or uploaded to one of the fog nodes, although our model can be directly extended to the scenarios with splittable tasks. We denote the offloading decision for task $i \in \mathcal{A}(t)$ by $I_i(t) \in \mathcal{N}(t)$. Particularly, $I_i(t) = 0$ indicates that task i will not be processed locally on the user node; otherwise, task i will be offloaded to fog node $I_i(t)$. According to the offloading decisions, tasks are scheduled and processed with results fed back to the user node at the end of time slot t.

[1] Our framework can be extended to the case of multiple users.

B. Optimization Objectives

Task Latency: For latency-sensitive IoT applications [2], task latency remains one of the key measurements of QoS. A task's latency mainly consists of the *transmission latency* (if it is offloaded to one of the fog nodes) and its *processing latency*. In the following, we present related notations to such latencies for each task $i \in \mathcal{A}(t)$ in detail.

On one hand, we denote the transmission latency of task i by $D^{(tr)}_{i,I_i(t)}(t)$. If $I_i(t) = 0$, *i.e.*, the task is to be processed locally, then no transmission latency will be induced ($D^{(tr)}_{i,I_i(t)}(t) = 0$). Otherwise, if task i is to be offloaded to fog node $I_i(t)$, then by defining $R_{i,n}(t)$ as the transmission rate allocated to task i when it is offloaded from the user node to fog node $n \in \mathcal{N}(t) \setminus \{0\}$, we write its transmission latency as

$$D^{(tr)}_{i,I_i(t)}(t) = L_i(t)\mathbb{1}\{I_i(t) > 0\}/R_{i,I_i(t)}(t), \tag{1}$$

where $D^{(tr)}_{i,I_i(t)}(t)$ is a function of the offloading decision $I_i(t)$. On the other hand, we denote the processing latency of task i by $D^{(pr)}_{i,I_i(t)}(t)$. Recall that the computation demand of task i is quantified by $W_i(t)$, its required number of CPU cycles. Considering the heterogeneity of processing capacities among fog nodes and the user node, we define $F_{i,I_i(t)}(t)$ as the CPU cycle frequency assigned to task i when it is allocated to node $I_i(t)$. The processing latency of task i can be written as

$$D^{(pr)}_{i,I_i(t)}(t) = W_i(t)/F_{i,I_i(t)}(t). \tag{2}$$

By definitions in (1) and (2), the total latency of task i is

$$\begin{aligned} D_{i,I_i(t)}(t) &= D^{(tr)}_{i,I_i(t)}(t) + D^{(pr)}_{i,I_i(t)}(t) \\ &= L_i(t)\mathbb{1}\{I_i(t) > 0\}/R_{i,I_i(t)}(t) + W_i(t)/F_{i,I_i(t)}(t). \end{aligned} \tag{3}$$

Recall that if the total latency of task i is greater than its deadline $\tau_{\max}$, then it will be considered failed. For ease of analysis, we set the latency for such failure tasks as $\tau_{\max}$.

In practice, however, due to wireless channel variations, $R_{i,n}(t)$ is often unknown before task i's transmission. Likewise, the exact value of $F_{i,n}(t)$ is revealed only after task i's completion. Therefore, we view both $R_{i,n}(t)$ and $F_{i,n}(t)$ as random variables with unknown distributions and means. Further, we assume both $R_{i,n}(t)$ and $F_{i,n}(t)$ to be *i.i.d.* for each task i, lower bounded by constant $r_{\min}$ and $f_{\min}$, respectively. In addition, we assume the existence of the mean of their reciprocals, denoted by $\rho_n \triangleq \mathbb{E}[1/R_{i,n}(t)]$ and $\phi_n \triangleq \mathbb{E}[1/F_{i,n}(t)]$, respectively.

Energy Consumption: Because of the resource limit on the user node and fog nodes, energy efficiency is another objective worth considering. Typically, for the user node, its energy consumption includes the transmission energy and the local processing energy, while for each fog node, the energy

consumption is mainly made up by CPU processing energy. For each time slot t, we denote by $\eta_n(t)$ the transmission energy of sending one bit from the user node to fog node $n \in \mathcal{N}(t) \setminus \{0\}$, and by $\kappa_n(t)$ the energy consumption for processing tasks with one CPU cycle on node $n \in \mathcal{N}(t)$. We assume that $\eta_n(t)$ and $\kappa_n(t)$ are readily attainable at the beginning of each time slot t, while upper bounded by some constants $\eta_{\max}$ and $\kappa_{\max}$, respectively. Then by defining the energy consumption of each node $n \in \mathcal{N}$ in time slot t as $E_n(t)$, we can write it as

$$\begin{aligned} E_0(t) = &\sum_{i \in \mathcal{A}(t)} \kappa_0(t) W_i(t) \mathbb{1}\{I_i(t) = 0\} \\ &+ \sum_{i \in \mathcal{A}(t)} \sum_{n=1}^{N} \eta_n(t) L_i(t) \mathbb{1}\{I_i(t) = n\}. \end{aligned} \tag{4}$$

and the energy consumption of each fog node n as

$$E_n(t) = \sum_{i \in \mathcal{A}(t)} \kappa_n(t) W_i(t) \mathbb{1}\{I_i(t) = n\}. \tag{5}$$

When it comes to energy efficiency, for each node $n \in \mathcal{N}$, we assume that it is quantified by an energy consumption budget b_n, such that the long-term time-average energy consumption should be capped by b_n, *i.e.*,

$$\limsup_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\left[E_n(t)\right] \leq b_n, \forall n \in \mathcal{N}. \tag{6}$$

III. Problem Formulation

We formulate the task offloading problem as the following stochastic network optimization problem, which aims to minimize the expected total task latency over finite time horizon T with energy efficiency constraints.

$$\begin{aligned} \underset{\{I_i(t)\}_{i,t}}{\text{minimize}} \quad & \sum_{t=0}^{T-1} \sum_{i \in \mathcal{A}(t)} \mathbb{E}\left[D_{i,I_i(t)}(t)\right] \\ \text{subject to} \quad & (6), I_i(t) \in \mathcal{N}(t), \forall i \in \mathcal{A}(t), \forall t. \end{aligned} \tag{7}$$

Provided that the exact information of allocated CPU cycle frequencies $\{F_{i,n}(t)\}_{i,n}$ and transmission rates $\{R_{i,n}(t)\}_{i,n}$ is given at the beginning of each time slot, Problem (7) can be solved by applying Lyapunov optimization method [5]. However, such assumptions usually do not hold in practice, thus the uncertainties must be learned.

To solve problem (7) with such uncertainties, we reformulate our problem as a stochastic combinatorial multi-armed bandit (CMAB) problem [7]. In the basic stochastic CMAB problem, a player chooses a subset of arms, *a.k.a.* super arm, from a fixed arm set to play in each round $t \in \{0, \ldots, T-1\}$. Each arm being played will reveal a reward to the player. The reward of each arm is assumed to be a random variable with unknown distribution and mean, and i.i.d. across different rounds. The goal of the player is to maximize the cumulative reward over finite time horizon of T rounds.

To fit our problem into such a setting, we regard the user node as the player and regard each offloading choice as an arm. Note there are $N+1$ arms in total because the choice of local processing is also regarded as an arm. In each time slot t, the player chooses an arm $I_i(t)$ from subset $\mathcal{N}(t)$ for each task $i \in \mathcal{A}(t)$, then the super arm be chosen in time slot t is the arm set $\{I_i(t)\}_{i \in \mathcal{A}(t)}$. If the player chooses arm n for task i, she will receive a random reward of

$$X_{i,I_i(t)}(t) \triangleq \tau_{\max} - D_{i,I_i(t)}(t). \tag{8}$$

The player's goal is to maximize the expected time-average reward of T time slots, *i.e.* $\frac{1}{T}\sum_{t=0}^{T-1}\sum_{i \in \mathcal{A}(t)} \mathbb{E}[X_{i,I_i(t)}(t)]$, under energy constraints shown in (6).

Our CMAB model is different from the basic CMAB model in the following ways: At first, an arm can be played for more than one time in one time slot since two different tasks can be allocated to the same node. Second, the available arm set is only a subset of all arms and it is varying over time slots. Third, we are faced with the energy constraints.

We denote X^* as the maximal expected time-average reward that can be achieved by the optimal policy for T time slots, then the regret of T time slots under offloading decision $\{I_i(t)\}_{i,t}$ is defined as

$$R(T) \triangleq X^* - \frac{1}{T} \sum_{t=0}^{T-1} \sum_{i \in \mathcal{A}(t)} \mathbb{E}\left[X_{i,I_i(t)}(t)\right]. \tag{9}$$

Since to maximize the reward is the same as minimizing the regret, we equivalently solve the following problem to find an optimal policy for the player:

$$\begin{aligned} \underset{\{I_i(t)\}_{i,t}}{\text{minimize}} \quad & R(T) \\ \text{subject to} \quad & (6), I_i(t) \in \mathcal{N}(t), \forall i \in \mathcal{A}(t), \forall t. \end{aligned} \tag{10}$$

Note that (10) is equivalent to the original problem (7). By the definition of regret $R(T)$ in (9) and the definition of task reward in (8), we have

$$R(T) = X^* - \frac{1}{T} \sum_{t=0}^{T-1} \sum_{i \in \mathcal{A}(t)} \left(\tau_{\max} - \mathbb{E}\left[D_{i,I_i(t)}(t)\right]\right). \tag{11}$$

Substitute the task latency expression (3) into (11), and by $\mathbb{E}\left[1/F_{i,n}(t)\right] = \phi_n$ and $\mathbb{E}\left[1/R_{i,n}(t)\right] = \rho_n$, we obtain

$$\begin{aligned} R(T) = X^* - \frac{1}{T} \sum_{t=0}^{T-1} \sum_{i \in \mathcal{A}(t)} &\big(\tau_{\max} \\ &- \mathbb{E}\left[\rho_{I_i(t)} L_i(t) \mathbb{1}\left\{I_i(t) > 0\right\} + \phi_{I_i(t)} W_i(t)\right]\big). \end{aligned} \tag{12}$$

Note that $R(T)$ is related to unknown parameters $\{\phi_n\}_{n \in \mathcal{N}}$ and $\{\rho_n\}_{n \in \mathcal{N} \setminus \{0\}}$. To develop an efficient algorithm for problem (10), we have to face with two challenges. One of them is to deal with the unknown parameters that are strongly related to the problem objective, the other is to take the energy constraints into consideration.

IV. Algorithm Design

Motivated by the idea of integrating bandit learning and Lyapunov optimization in recent works [7] [8], we propose Learning-Aided Green Offloading (LAGO) algorithm to solve problem (10), as shown in Algorithm 1. In the following subsections, we specify the design of LAGO in detail.

Algorithm 1 Learning-Aided Green Offloading (LAGO)

Initialize $\bar{\rho}_n(0) = \hat{\rho}_n(0) = 0$ for each fog node $n \in \mathcal{N} - \{0\}$, and $\bar{\phi}_n(0) = \hat{\phi}_n(0) = h_n(0) = 0$ for each node $n \in \mathcal{N}$. Initialize $V = 1$. In each time slot $t \in \{0, 1, \dots\}$:
%%Learn the UCB estimates.
for each node $n \in \mathcal{N}(t)$ **do**
  if $h_n(t) > 0$ **then**
    Update $\hat{\phi}_n(t)$ according to (14).
    Update $\hat{\rho}_n(t)$ according to (13) if n is a fog node.
  end if
end for
%%Control and optimization.
for each task $i \in \mathcal{A}(t)$ **do**
  Set $I_i(t) \leftarrow \arg\min_{n \in \mathcal{N}(t)} v_{i,n}(t, \hat{\rho}_n(t), \hat{\phi}_n(t))$ and assign task i to node $I_i(t)$.
end for
Update virtual queues $\{Q_n(t)\}_{n \in \mathcal{N}(t)}$ according to (23).
%%Update counters and empirical means.
for each node $n \in \mathcal{N}(t)$ **do**
  Update $h_n(t)$ and $\bar{\phi}_n(t)$ according to (18) and (20).
  Update $\bar{\rho}_n(t)$ according to (19) if n is a fog node.
end for

A. Learning with UCB Method

We obtain the unknown system parameters by incorporating learning methods into our algorithm. To learn efficiently, we need to address the balance between maximizing expected reward (exploitation) and acquiring new knowledge (exploration), *i.e.*, exploitation-exploration tradeoff. In our CMAB formulation, parameters ρ_n (if $n > 0$) and ϕ_n are learned based on the received feedback after each play. The more times arm n is played, the more reliable estimates of ρ_n (if $n > 0$) and ϕ_n are. But this will give rise to higher regret when arm n is suboptimal. In our LAGO algorithm, we use UCB method to address the tradeoff. Specifically, we use UCB1 to estimate the values of ρ_n and ϕ_n as follows in each time slot t:

$$\hat{\rho}_n(t) = [\bar{\rho}_n(t) - \rho_{\max}\sqrt{3\log t/(2h_n(t))}]^+, \forall n \in \mathcal{N} \setminus \{0\}, \quad (13)$$

$$\hat{\phi}_n(t) = [\bar{\phi}_n(t) - \phi_{\max}\sqrt{3\log t/(2h_n(t))}]^+, \forall n \in \mathcal{N}, \quad (14)$$

in which $[\cdot]^+ \triangleq \max\{\cdot, 0\}$, $\rho_{\max} \triangleq 1/r_{\min}$, $\phi_{\max} \triangleq 1/f_{\min}$, $h_n(t)$ is the total number of times that arm n is chosen during the first t time slots, and $\bar{\rho}_n(t)$ and $\bar{\phi}_n(t)$ are the empirical averages of $1/R_{i,n}(t)$ and $1/F_{i,n}(t)$ respectively. The definitions of $h_n(t)$, $\bar{\rho}_n(t)$, and $\bar{\phi}_n(t)$ are

$$h_n(t) \triangleq \sum_{\tau=0}^{t-1} \sum_{i \in \mathcal{A}(\tau)} \mathbb{1}\{I_i(\tau) = n\}, \forall n \in \mathcal{N}, \quad (15)$$

$$\bar{\rho}_n(t) \triangleq \frac{1}{h_n(t)} \sum_{\tau=0}^{t-1} \sum_{i \in \mathcal{A}(\tau)} \frac{\mathbb{1}\{I_i(\tau) = n\}}{R_{i,n}(\tau)}, \forall n \in \mathcal{N} \setminus \{0\}, \quad (16)$$

$$\bar{\phi}_n(t) \triangleq \frac{1}{h_n(t)} \sum_{\tau=0}^{t-1} \sum_{i \in \mathcal{A}(\tau)} \frac{\mathbb{1}\{I_i(\tau) = n\}}{F_{i,n}(\tau)}, \forall n \in \mathcal{N}. \quad (17)$$

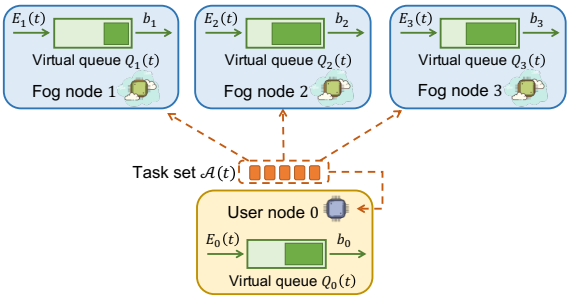


Fig. 1. Illustration of virtual queues in a fog computing system with one user node and three fog nodes. Each node $n \in \{0, 1, 2, 3\}$ keeps a virtual queue $Q_n(t)$ to handle its long-term energy constraint. The virtual queue $Q_n(t)$ has an input of $E_n(t)$ and a constant output of b_n, and the energy constraint of node n is satisfied if $Q_n(t)$ is strongly stable.

Note that $h_n(t)$, $\bar{\rho}_n(t)$ and $\bar{\phi}_n(t)$ can be updated iteratively in time slot t as follows:

$$h_n(t) = h_n(t-1) + \sum_{i \in \mathcal{A}(t)} \mathbb{1}\{I_i(t) = n\}, \forall n \in \mathcal{N}, \quad (18)$$

$$\bar{\rho}_n(t) = \bar{\rho}_n(t-1)h_n(t-1)/h_n(t) + \sum_{i \in \mathcal{A}(t)} \mathbb{1}\{I_i(t) = n\}/(R_{i,n}(t)h_n(t)), \forall n \in \mathcal{N} \setminus \{0\}, \quad (19)$$

$$\bar{\phi}_n(t) = \bar{\phi}_n(t-1)h_n(t-1)/h_n(t) + \sum_{i \in \mathcal{A}(t)} \mathbb{1}\{I_i(t) = n\}/(F_{i,n}(t)h_n(t)), \forall n \in \mathcal{N}. \quad (20)$$

Note that updating $\bar{\rho}_n(t)$ and $\bar{\phi}_n(t)$ requires the values of $\{R_{i,I_i(t)}(t)\}_{i \in \mathcal{A}(t), I_i(t) > 0}$ and $\{F_{i,I_i(t)}(t)\}_{i \in \mathcal{A}(t)}$. These values can be obtained from the received feedback as follows:

$$R_{i,I_i(t)}(t) = L_i(t)/d_i^{(tr)}(t), \forall i \in \mathcal{A}(t), I_i(t) > 0, \quad (21)$$

$$F_{i,I_i(t)}(t) = W_i(t)/d_i^{(pr)}(t), \forall i \in \mathcal{A}(t), \quad (22)$$

where $d_i^{(tr)}$ and $d_i^{(pr)}$ are transmission latency and processing latency of task i that are received from the feedback information. Other learning methods such as UCB variants and ϵ-greedy algorithm can also be used in our framework.

B. Control and Optimization with Lyapunov Method

In this subsection, we take energy constraints into consideration. Applying the idea of transforming inequality constraints into queue stability problems [5], we introduce a virtual queue $Q_n(t)$ for each node $n \in \mathcal{N}$ with $Q_n(0) = 0$ to handle the energy constraints. At each time slot t, the output of queue $Q_n(t)$ is b_n, *i.e.* the energy upper bound of node n, and the input of queue $Q_n(t)$ is the energy consumption on node n during time slot t. The update function of queue $Q_n(t)$ is:

$$Q_n(t+1) = [Q_n(t) - b_n]^+ + E_n(t). \quad (23)$$

The energy constraint on node n is satisfied when $Q_n(t)$ is strongly stable [5]. Figure 1 shows an example of such virtual queues. Applying Lyapunov optimization method, our algorithm LAGO can guarantee the stability of virtual queues and minimize the regret at the same time.

In every time slot t, we make offloading decision based on virtual queue backlogs, task informations, and UCB estimates introduced in (13) and (14). We define a set of value functions $\{v_{i,n}(t,\rho,\phi)\}_{i\in\mathcal{N}}$ for each task $i\in\mathcal{A}(t)$ as follows:

$$\begin{aligned} v_{i,n}(t,\rho,\phi) \triangleq\ & Q_n(t)\kappa_n(t)W_i(t) \\ & + Q_0(t)\eta_n(t)L_i(t)\mathbb{1}\{n>0\} \\ & + V(\phi W_i(t)+\rho L_i(t)\mathbb{1}\{n>0\}), \forall n\in\mathcal{N}. \end{aligned} \tag{24}$$

Then algorithm LAGO makes offloading decision for each task $i\in\mathcal{A}(t)$ as follows:

$$I_i(t)=\arg\min_{n\in\mathcal{N}(t)} v_{i,n}(t,\hat{\rho}_n(t),\hat{\phi}_n(t)), \tag{25}$$

that is, task i will be allocated to node $n\in\mathcal{N}(t)$ with minimal value $v_{i,n}(t,\hat{\rho}_n(t),\hat{\phi}_n(t))$.

Note that as V increases, LAGO is more willing to choose node with small $\phi_n W_i(t)+\rho_n L_i(t)\mathbb{1}\{n>0\}$ according to (24) and (25). ϕ_n can be viewed as the average latency of processing with one CPU cycle on node n, and ρ_n can be viewed as the average latency of transmitting one bit of task to node n. Usually, the processing latency of one CPU cycle on the user node, *i.e.* ϕ_0, is much larger than that of fog nodes, thus the user node usually tends to offload tasks to fog nodes in the case of big V under LAGO.

On the other hand, the user node is more willing to process tasks locally when the virtual queue backlog $Q_0(t)$ of the user node is small while the virtual queue backlogs $\{Q_n(t)\}_{n\in\mathcal{N}(t)\setminus\{0\}}$ of fog nodes are large. This is due to that LAGO should guarantee the stabilities of all virtual queues to satisfy the energy constraints.

As shown in Algorithm 1, LAGO is composed of three parts in every time slot. First, it learns the UCB estimates of unknown ϕ_n and ρ_n. Then it makes offloading decision for every task based on the UCB estimates to optimize the average task latency while guaranteeing energy constraints. Finally, LAGO updates the arm (action) hitting times and the empirical means based on feedback of every task in current time slot, and these latest information will be used to update the UCB estimates in the next time slot.

V. Theoretical Analysis

In this section, we show that LAGO guarantees the energy constraints and provide a theoretical regret bound for it.

A. Energy Consumption Bound

For any energy bound vector $\boldsymbol{b}=(b_0,\ldots,b_N)$, it is said to be *feasible* if there exists an offloading policy such that all energy constraints are satisfied. We define the set of all feasible energy bound vectors as the *maximal feasibility region*. Then we have the following theorem for the feasibility of LAGO and the proof is shown in Appendix A.

Theorem 1: Assume that the energy bound vector $\boldsymbol{b}$ lies in the interior of the maximal feasibility region, then the energy constraints in (7) are satisfied under LAGO algorithm. Moreover, the virtual queues defined in (23) are strongly stable and there exists some constant $\epsilon>0$ such that

$$\limsup_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}\sum_{n\in\mathcal{N}}\mathbb{E}[Q_n(t)]\le\frac{B+V(\theta_1+\theta_2)}{\epsilon}, \tag{26}$$

where $B\triangleq\frac{1}{2}\sum_{n\in\mathcal{N}}b_n^2+\frac{1}{2}a_{\max}\max\{\kappa_{\max}l_{\max},\eta_{\max}w_{\max}\}+\frac{1}{2}Na_{\max}\kappa_{\max}l_{\max}$, $\theta_1\triangleq 2w_{\max}\phi_{\max}a_{\max}$, $\theta_2\triangleq 2l_{\max}\rho_{\max}a_{\max}$.

Remark 1: Theorem 1 shows that LAGO is feasible to problem (7) when $\boldsymbol{b}$ is interior to the maximal feasibility region. Moreover, the total virtual queue backlog increases linearly with the increase of tradeoff parameter V. This implies that the average node energy consumptions are approaching the energy bound vector $\boldsymbol{b}$ as parameter V increases.

B. Regret Bound

We provide an upper bound for the regret $R(T)$ under LAGO algorithm as follows and the proof is shown in Appendix B:

Theorem 2: Under algorithm LAGO, the regret of T time slots defined in (9) has an upper bound shown as follows:

$$\begin{aligned} R(T)\le & \frac{B}{V}+\left(\frac{3}{T}+\sqrt{\frac{3a_{\max}(N+1)\log T}{2T}}\right)\frac{\theta_1}{2} \\ & +\left(\frac{3}{T}+\sqrt{\frac{3a_{\max}N\log T}{2T}}\right)\frac{\theta_2}{2}. \end{aligned} \tag{27}$$

where $\theta_1/2=w_{\max}\phi_{\max}a_{\max}$ and $\theta_2/2=l_{\max}\rho_{\max}a_{\max}$ are the maximal possible task processing latency and transmission latency respectively.

Remark 2: The first term B/V in (27) is brought by the control part of the algorithm, the second term is brought by the learning of processing latency, and the third term is brought by the learning of transmission latency. As we can see from Theorem 2, the upper bound of regret depends on the time horizon length T and the tradeoff parameter V. As T increases to infinite, the regret bound will decrease to B/V since the last two terms in the regret bound is of the order $O(\sqrt{\log T/T})$ when T is large. On the other hand, the first term B/V will decrease to 0 as V increases to infinite. Therefore, to get a smaller regret we can increase both V and T.

VI. Simulation Results

A. Simulation Settings

We conduct extensive simulation in a fog computing system with $N=20$ fog nodes and one user node, and the simulation is run over $T=5\times10^5$ time slots, based on commonly adopted settings in fog computing systems [9], [10], which are specified as follows.

• *Task arrivals:* The number of task arrivals in each time slot is generated by Poisson distribution with mean 10 and then bounded into region $[1,20]$. The size of each task is sampled from distribution $\mathrm{Unif}(10^6,10^7)$ bits. The computation complexity, *i.e.* the number of CPU cycles needed for processing one bit of task, is set to 1000 cycles/bit for every task.

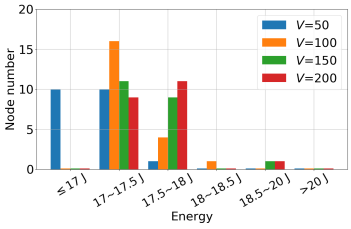


Fig. 2. Effect of V on node energy consumption.

• *Transmission rate:* The transmission rate to fog node in every time slot is generated by distribution $\mathrm{Unif}(r_{n,\min}, r_{n,\max})$, where $r_{n,\min}$ and $r_{n,\max}$ are generated by distribution $\mathrm{Unif}(5\times 10^6, 1.5\times 10^7)$ bits/s and $\mathrm{Unif}(5\times 10^7, 1.5\times 10^8)$ bits/s respectively at the beginning of simulation.

• *Processing rate:* In every time slot, the task processing rate on user node is generated by distribution $\mathrm{Unif}(10^9, 10^{10})$, and the task processing rate on fog node n is generated by distribution $\mathrm{Unif}(f_{n,\min}, f_{n,\max})$, where $f_{n,\min}$ and $f_{n,\max}$ are generated by distribution $\mathrm{Unif}(5\times 10^9, 1.5\times 10^{10})$ cycles/s and $\mathrm{Unif}(1.5\times 10^{10}, 2.5\times 10^{10})$ cycles/s respectively at the beginning of the simulation.

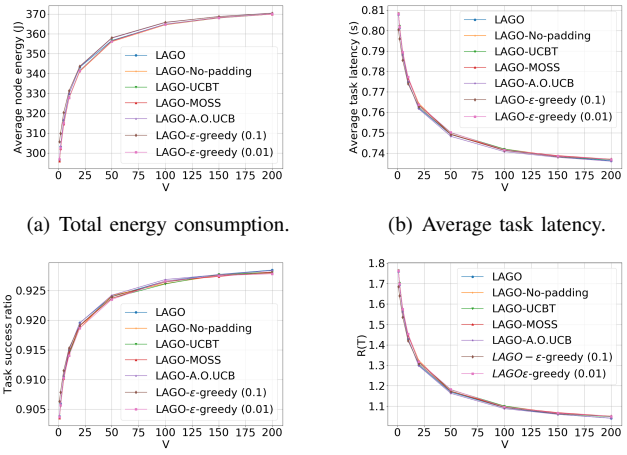
• *Energy consumption:* In every time slot, the unit processing energy on user node is sampled from $\mathrm{Unif}(10^{-10}, 5\times 10^{-10})$ J/cycle, the unit processing energy on each fog node is sampled from $\mathrm{Unif}(5\times 10^{-9}, 1.5\times 10^{-8})$ J/cycle, and the unit transmission energy from user node to each fog node n is sampled from $\mathrm{Unif}(10^{-7}, 10^{-6})$ J/bit. The energy upper bound is set to 20 J for every node.

B. Energy Consumptions vs. V under LAGO

We investigate the performance of LAGO under different values of V. Figure 2 illustrates the histogram of time-average node energy consumptions under LAGO. For example, the leftmost blue bar indicates that there are 10 nodes whose average energy consumption is less than or equal to 17 J when $V = 50$. From the figure we observe that when $V = 50$, all nodes but only one consumes less than 17.5 J energy. But when $V = 200$, more than half of the 21 nodes consume more than 17.5 J energy. The observation implies when V increases, the energy consumption on every node tends to increase. However, when V increases to 200, the energy budgets are still satisfied on all node as none of them consumes more than 20 J. The simulation result implies that when V increases, the user node is more willing to offload. On the other hand, our algorithm is feasible under different values of V.

C. LAGO vs. LAGO's Variants

As we mentioned before, we can substitute UCB1 with other learning methods in LAGO. In our simulation, we investigate the performance under the following learning methods: *UCB-tuned* (*UCBT*) [11], *Moss* [12], *asymptotically optimal UCB* (*A.O.UCB*) [13], and ϵ-*greedy* algorithm [14]. We refer to the LAGO variant which is equipped with learning method X as LAGO-X algorithm. We further consider the case when we choose offloading decision



(a) Total energy consumption. (b) Average task latency.

(c) Cumulative Success Ratio. (d) Regret.

Fig. 3. Effect of V on performance of LAGO-Xs.

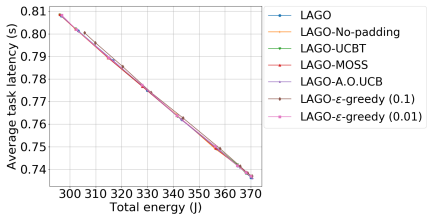


Fig. 4. Latency-Energy tradeoff of LAGO and its variants.

$I_i(t) = \arg\min_{n\in\mathcal{N}(t)} v_{i,n}(t, \bar{\rho}_n(t), \bar{\phi}_n(t))$ for each task i in time slot t, and refer to this variant as LAGO-*No-padding* algorithm. Note that in LAGO-*No-padding* algorithm, we drop the exploration of learning process by setting the padding function of UCB1 to zero.

The performance of LAGO and its variants are illustrated in Figure 3, which shows that LAGO and its variants perform very similarly. Subfigures 3-(a) and 3-(b) show that when V increases, the total energy consumption increases, but the average task latency decreases. The result is reasonable since the user node is more willing to offload to achieve low latency when V is large, but the total energy consumption increases because of the increasing processing energy consumption on fog nodes and the growing transmitting energy consumption on the user node. The result implies that we can adjust the tradeoff between latency-energy tradeoff by tuning the value of V. We also investigate the cumulative task success ratio in subfigure 3-(c), from which we observe that the success ratio increases as V increases. Subfigure 3-(d) shows that the regret of algorithm decreases with the increase of V, which verifies our theoretical analysis in Section V.

To compare the performance of LAGO and its variants, we illustrate the latency-energy tradeoff curves of these algorithms in Figure 4. The figure shows that all algorithms except for LAGO-ϵ-*greedy* with $\epsilon = 0.1$ achieve the similar latency-energy tradeoff, though they achieve different total energy consumption and average task latency under the same V. An interesting observation is that though there is no exploration

in the learning part of LAGO-*No-padding*, it performs almost the same as LAGO and other variants that have exploration in learning. On the other hand, LAGO-ϵ-*greedy* with $\epsilon = 0.1$ performs even worse than LAGO-*No-padding* as it achieves higher average task latency under the same total energy consumption. The reason for this result is that with Lyapunov control we avoid sampling a node too frequently by controlling the stability of its virtual queue. In other words, even without the exploration in learning part of the algorithm, we will still explore other nodes when the virtual queue of the empirically best node is too large, *i.e.*, the control part of these algorithms bring opportunity for exploration. As for LAGO-ϵ-*greedy*, its performance is degraded because of over-exploration.

VII. Conclusion

In this paper, we studied the task offloading problem in a fog computing system with unknown node processing capacities and link transmission rates under long-term energy constraints. By integrating Lyapunov optimization techniques with bandit learning methods, we proposed LAGO, an online task offloading scheme which endeavors to achieve minimal task latency while satisfying the energy constraints. Our theoretical results show that LAGO achieves a tunable latency-energy tradeoff under the energy constraints, and the regret bound of LAGO is of the order $O(\alpha_1/V + \alpha_2\sqrt{\log T/T})$ over time horizon T. By conducting extensive simulations, we verified our theoretical results and investigated the performance of LAGO's variants by replacing the learning block with variants of bandit algorithms.

Appendix A
Proof of Theorem 1

First, we define a Lyapunov function as

$$L\left(\boldsymbol{Q}\left(t\right)\right) \triangleq \frac{1}{2}\sum_{n\in\mathcal{N}}\left(Q_n\left(t\right)\right)^2, \tag{28}$$

in which $\boldsymbol{Q}(t)$ is the vector of all virtual queue backlogs in time slot t. Then we define the corresponding Lyapunov drift as

$$\Delta L\left(\boldsymbol{Q}\left(t\right)\right) \triangleq \mathbb{E}\left[L\left(\boldsymbol{Q}\left(t+1\right)\right) - L\left(\boldsymbol{Q}\left(t\right)\right)|\boldsymbol{Q}\left(t\right)\right]. \tag{29}$$

To develop an upper bound for $\Delta L\left(\boldsymbol{Q}\left(t\right)\right)$, we consider the difference of the Lyapunov function $L(\boldsymbol{Q}(t))$:

$$\begin{aligned}
&L\left(\boldsymbol{Q}\left(t+1\right)\right) - L\left(\boldsymbol{Q}\left(t\right)\right)\\
&\leq \frac{1}{2}\sum_{n\in\mathcal{N}}[(Q_n(t)-\rho_n)^2 + (E_n(t))^2 + 2Q_n(t)E_n(t)]\\
&\quad - \frac{1}{2}\sum_{n\in\mathcal{N}}(Q_n(t))^2\\
&= \frac{1}{2}\sum_{n\in\mathcal{N}}(b_n^2 + (E_n(t))^2) + \sum_{n\in\mathcal{N}}Q_n(t)(E_n(t)-\rho_n).
\end{aligned} \tag{30}$$

According to (4), (5) and the boundedness assumptions of the model, we have $E_0\left(t\right) \leq a_{\max}\max\left\{\kappa_{\max}l_{\max}, \eta_{\max}w_{\max}\right\}$ and $E_n\left(t\right) \leq a_{\max}\kappa_{\max}l_{\max}$, thus

$$\begin{aligned}
&\frac{1}{2}\sum_{n\in\mathcal{N}}\left(b_n^2 + \left(E_n\left(t\right)\right)^2\right)\\
&\leq \frac{1}{2}a_{\max}\max\left\{\kappa_{\max}l_{\max}, \eta_{\max}w_{\max}\right\}\\
&\quad + \frac{1}{2}Na_{\max}\kappa_{\max}L_{\max} + \frac{1}{2}\sum_{n\in\mathcal{N}}b_n^2 = B.
\end{aligned} \tag{31}$$

It follows by (30) that

$$L(\boldsymbol{Q}(t+1)) - L(\boldsymbol{Q}(t)) \leq B + \sum_{n\in\mathcal{N}}Q_n(t)(E_n(t)-\rho_n). \tag{32}$$

Then we obtain the upper bound of Lyapunov drift as

$$\Delta L(\boldsymbol{Q}(t)) \leq B + \mathbb{E}\left[\sum_{n\in\mathcal{N}}Q_n(t)(E_n(t)-\rho_n)|\boldsymbol{Q}(t)\right]. \tag{33}$$

Suppose the optimal policy makes offloading decision $\{I_i^*(t)\}_{i\in\mathcal{A}(t)}$ in each time slot t, then we have

$$X^* = \frac{1}{T}\sum_{t=0}^{T-1}\sum_{i\in\mathcal{A}(t)}\mathbb{E}[X_{i,I_i^*(t)}(t)]. \tag{34}$$

By (34), we can rewrite the regret $R(T)$ under policy $\{I_i(t)\}_{i,t}$ as

$$R(T) = \frac{1}{T}\sum_{t=0}^{T-1}\sum_{i\in\mathcal{A}(t)}(\mathbb{E}[X_{i,I_i^*(t)}(t)] - \mathbb{E}[X_{i,I_i(t)}(t)]). \tag{35}$$

Then by $X_{i,n}(t) = \tau_{\max} - D_{i,n}(t)$, it follows that

$$R(T) = \frac{1}{T}\sum_{t=0}^{T-1}\sum_{i\in\mathcal{A}(t)}(\mathbb{E}[D_{i,I_i(t)}(t)] - \mathbb{E}[D_{i,I_i^*(t)}(t)]). \tag{36}$$

Substitute (3) into above equation, we have

$$\begin{aligned}
&R\left(T\right) =\\
&\frac{1}{T}\sum_{t=0}^{T-1}\sum_{i\in\mathcal{A}(t)}\left(\mathbb{E}\left[\frac{L_i\left(t\right)\mathbb{1}\left\{I_i\left(t\right)>0\right\}}{R_{i,I_i(t)}\left(t\right)} + \frac{W_i\left(t\right)}{F_{i,I_i(t)}\left(t\right)}\right]\right.\\
&\qquad \left. - \mathbb{E}\left[\frac{L_i\left(t\right)\mathbb{1}\left\{I_i^*\left(t\right)>0\right\}}{R_{i,I_i^*(t)}\left(t\right)} + \frac{W_i\left(t\right)}{F_{i,I_i^*(t)}\left(t\right)}\right]\right).
\end{aligned} \tag{37}$$

Since $\mathbb{E}\left[1/R_{i,n}\left(t\right)\right] = \rho_n$ and $\mathbb{E}\left[1/F_{i,n}\left(t\right)\right] = \phi_n$, given offloading decision $\{I_i(t)\}_i$, we have

$$\begin{aligned}
&R\left(T\right) =\\
&\frac{1}{T}\sum_{t=0}^{T-1}\sum_{i\in\mathcal{A}(t)}(\mathbb{E}[\rho_{I_i(t)}L_i(t)\mathbb{1}\{I_i(t)>0\} + \phi_{I_i(t)}W_i(t)]\\
&\qquad - \mathbb{E}[\rho_{I_i^*(t)}L_i(t)\mathbb{1}\{I_i^*(t)>0\} + \phi_{I_i^*(t)}W_i(t)]).
\end{aligned} \tag{38}$$

We define the one-slot regret as

$$\begin{aligned}
\Delta R(t) \triangleq &\sum_{i\in\mathcal{A}(t)}(\rho_{I_i(t)}L_i(t)\mathbb{1}\{I_i(t)>0\} + \phi_{I_i(t)}W_i(t))\\
&- \sum_{i\in\mathcal{A}(t)}(\rho_{I_i^*(t)}L_i(t)\mathbb{1}\{I_i^*(t)>0\} + \phi_{I_i^*(t)}W_i(t)),
\end{aligned} \tag{39}$$

For simplicity of expression, we let

$$D^*(t) \triangleq \sum_{i\in\mathcal{A}(t)} \rho_{I_i^*(t)} L_i(t) \mathbb{1}\{I_i^*(t)>0\} + \phi_{I_i^*(t)} W_i(t), \quad (40)$$

then we have

$$\Delta R(t) z \triangleq \sum_{i\in\mathcal{A}(t)} \left(\rho_{I_i(t)} L_i(t) \mathbb{1}\{I_i(t)>0\} + \phi_{I_i(t)} W_i(t)\right) - D^*(t). \quad (41)$$

Regret $R(T)$ can be expressed as

$$R(T) = \frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}\left[\Delta R(t)\right]. \quad (42)$$

Next, define the drift-plus-regret as

$$\Delta_V(\boldsymbol{Q}(t)) = \Delta(\boldsymbol{Q}(t)) + V\mathbb{E}\left[\Delta R(t)\,|\boldsymbol{Q}(t)\right], \quad (43)$$

which is the weighted sum of Lyapunov drift and expected one-slot regret. By (33), we get an upper bound for the drift-plus-regret:

$$\Delta_V(\boldsymbol{Q}(t)) \le B + \mathbb{E}\left[\sum_{n\in\mathcal{N}} Q_n(t)\left(E_n(t)-b_n\right) + V\Delta R(t)\,|\boldsymbol{Q}(t)\right]. \quad (44)$$

The right-hand side of (44) can be rewritten as follows according to (4), (5), and (41):

$$\begin{aligned}
&\Delta_V(\boldsymbol{Q}(t)) \le B - V\mathbb{E}\left[D^*(t)\,|\boldsymbol{Q}(t)\right] - \sum_{n\in\mathcal{N}} b_n Q_n(t) \\
&+ \mathbb{E}\left[\sum_{i\in\mathcal{A}(t)} \left(Q_0(t)\kappa_0(t) W_i(t) + V\phi_0 W_i(t)\right) \cdot \mathbb{1}\left\{I_i(t)=0\right\}|\boldsymbol{Q}(t)\right] \\
&+ \mathbb{E}\left[\sum_{i\in\mathcal{A}(t)}\sum_{n=1}^{N} (Q_0(t)\eta_n(t)L_i(t) + Q_n(t)\kappa_n(t)W_i(t) \right. \\
&\qquad \left. + V\rho_n L_i(t) + V\phi_n W_i(t))\mathbb{1}\left\{I_i(t)=n\right\}|\boldsymbol{Q}(t)\right] \\
&= B - V\mathbb{E}\left[D^*(t)\,|\boldsymbol{Q}(t)\right] - \sum_{n\in\mathcal{N}} b_n Q_n(t) \\
&+ \mathbb{E}\left[\sum_{i\in\mathcal{A}(t)}\sum_{n\in\mathcal{N}} v_{i,n}(t,\rho_n,\phi_n)\,\mathbb{1}\left\{I_i(t)=n\right\}|\boldsymbol{Q}(t)\right] \\
&= B - V\mathbb{E}\left[D^*(t)\,|\boldsymbol{Q}(t)\right] - \sum_{n\in\mathcal{N}} b_n Q_n(t) \\
&+ \mathbb{E}\left[\sum_{i\in\mathcal{A}(t)}\sum_{n\in\mathcal{N}} v_{i,n}(t,\bar{\rho}_n,\bar{\phi}_n)\mathbb{1}\{I_i(t)=n\}|\boldsymbol{Q}(t)\right] \\
&+ \mathbb{E}\left[\sum_{i\in\mathcal{A}(t)} V\left((\rho_{I_i(t)} - \bar{\rho}_{I_i(t)}(t))L_i(t) \right.\right. \\
&\qquad \left.\left. + (\phi_{I_i(t)} - \bar{\phi}_{I_i(t)}(t))W_i(t)\mathbb{1}\{I_i(t)>0\}\right)|\boldsymbol{Q}(t)\right].
\end{aligned} \quad (45)$$

Since $\phi_n \le \phi_{\max}$, $L_i(t) \le l_{\max}$, $\rho_n \le \rho_{\max}$, $W_i(t) \le w_{\max}$, and $A(t) \le a_{\max}$, it follows that

$$\begin{aligned}
&\Delta_V(\boldsymbol{Q}(t)) \le B - V\mathbb{E}\left[D^*(t)\,|\boldsymbol{Q}(t)\right] - \sum_{n\in\mathcal{N}} b_n Q_n(t) \\
&+ \mathbb{E}\left[\sum_{i\in\mathcal{A}(t)}\sum_{n\in\mathcal{N}} v_{i,n}(t,\bar{\rho}_n,\bar{\phi}_n)\mathbb{1}\{I_i(t)=n\}|\boldsymbol{Q}(t)\right] \\
&+ Va_{\max}\left(\rho_{\max} l_{\max} + \phi_{\max} w_{\max}\right).
\end{aligned} \quad (46)$$

We refer to policies that makes i.i.d. offloading decision in each time slot t as a function of observable system state $(\eta_n(t), \kappa_n(t), L_i(t), W_i(t))_{n,i}$ as *S-only* policies. By the assumption in Theorem 1, $\boldsymbol{b}$ is an interior point of the maximal feasibility region. Then there must exist some $\epsilon > 0$ such that $\boldsymbol{b} - \epsilon\mathbf{1}$ is also an interior point of the maximal feasibility region. Then there exists an S-only policy which makes offloading decision $\{I_i^\epsilon(t)\}_i$ in each time slot t such that

$$\mathbb{E}\left[E_n^\epsilon(t)\right] + \epsilon \le b_n, \ \forall n \in \mathcal{N} \quad (47)$$

holds for all time slots, where $E_n^\epsilon(t)$ is the energy consumption of node n under offloading decision $\{I_i^\epsilon(t)\}_i$.

Since our policy $\{I_i(t)\}_i$ minimizes the right-hand side of (46), we have

$$\begin{aligned}
&\Delta_V(\boldsymbol{Q}(t)) \le B - V\mathbb{E}\left[D^*(t)\,|\boldsymbol{Q}(t)\right] - \sum_{n\in\mathcal{N}} b_n Q_n(t) \\
&+ \mathbb{E}\left[\sum_{i\in\mathcal{A}(t)}\sum_{n\in\mathcal{N}} v_{i,n}(t,\bar{\rho}_n,\bar{\phi}_n)\mathbb{1}\{I_i^\epsilon(t)=n\}\right] \\
&+ Va_{\max}\left(\rho_{\max} l_{\max} + \phi_{\max} w_{\max}\right) \\
&\le B - \sum_{n\in\mathcal{N}} b_n Q_n(t) + Va_{\max}\left(\rho_{\max} l_{\max} + \phi_{\max} w_{\max}\right) \\
&+ \mathbb{E}\left[\sum_{i\in\mathcal{A}(t)}\sum_{n\in\mathcal{N}} v_{i,n}(t,\bar{\rho}_n,\bar{\phi}_n)\mathbb{1}\{I_i^\epsilon(t)=n\}\right]
\end{aligned} \quad (48)$$

Since $I_i^\epsilon(t)$ is independent of $\boldsymbol{Q}(t)$. By the definition of $v_{i,n}(t,\rho,\phi)$, it follows that

$$\begin{aligned}
&\Delta_V(\boldsymbol{Q}(t)) \le B - \sum_{n\in\mathcal{N}} b_n Q_n(t) \\
&+ Va_{\max}(\rho_{\max} l_{\max} + \phi_{\max} w_{\max}) \\
&+ \sum_{n\in\mathcal{N}} Q_n(t)\mathbb{E}\left[E_n^\epsilon(t)\right] + V\sum_{i\in\mathcal{A}(t)} \mathbb{E}\left[\left(\bar{\phi}_{I_i^\epsilon(t)}(t) W_i(t) \right.\right. \\
&\qquad \left.\left. + \bar{\rho}_{I_i^\epsilon(t)}(t) L_i(t)\,1\{n>0\}\right)\right] \\
&\le B + V(\theta_1+\theta_2) + \sum_{n\in\mathcal{N}} Q_n(t)\left(\mathbb{E}\left[E_n^\epsilon(t)\right] - b_n\right)
\end{aligned} \quad (49)$$

where $E^\epsilon(t)$ and $\Delta R^\epsilon(t)$ are the energy consumption and one-slot regret under policy $\{I_i^\epsilon(t)\}_i$, $\theta_1 \triangleq 2w_{\max}\phi_{\max}a_{\max}$, and $\theta_2 \triangleq 2l_{\max}\rho_{\max}a_{\max}$. It follows by (47) that

$$\Delta_V(\boldsymbol{Q}(t)) \le B + V(\theta_1+\theta_2) - \epsilon\sum_{n\in\mathcal{N}} Q_n(t). \quad (50)$$

By substituting definition (29) into above inequality yields

$$L\left(\boldsymbol{Q}\left(t+1\right)\right)-L\left(\boldsymbol{Q}\left(t\right)\right)+V\mathbb{E}\left[\Delta R\left(t\right)|\boldsymbol{Q}\left(t\right)\right] \leq B+V\left(\theta_1+\theta_2\right)-\epsilon\sum_{n\in\mathcal{N}}Q_n\left(t\right), \quad (51)$$

and it follows that

$$L\left(\boldsymbol{Q}\left(t+1\right)\right)-L\left(\boldsymbol{Q}\left(t\right)\right) \leq B+V\left(\theta_1+\theta_2\right)-\epsilon\sum_{n\in\mathcal{N}}Q_n\left(t\right). \quad (52)$$

Taking expectation of both sides of the inequality and summing over T time slots, we have

$$\mathbb{E}\left[L\left(\boldsymbol{Q}\left(T\right)\right)\right]-\mathbb{E}\left[L\left(\boldsymbol{Q}\left(0\right)\right)\right] \leq\left(B+V\left(\theta_1+\theta_2\right)\right)T-\epsilon\sum_{t=0}^{T-1}\sum_{n\in\mathcal{N}}\mathbb{E}\left[Q_n\left(t\right)\right]. \quad (53)$$

Dividing both side by $T\epsilon$ and rearrange the items, we have

$$\frac{1}{T}\sum_{t=0}^{T-1}\sum_{n\in\mathcal{N}}\mathbb{E}\left[Q_n\left(t\right)\right]+\frac{1}{T}\mathbb{E}\left[L\left(\boldsymbol{Q}\left(T\right)\right)\right] \leq\frac{B+V\left(\theta_1+\theta_2\right)}{\epsilon}+\frac{\mathbb{E}\left[L\left(\boldsymbol{Q}\left(0\right)\right)\right]}{T}. \quad (54)$$

Since $L(\boldsymbol{Q}(0))=0$ and $L(\boldsymbol{Q}(t))\geq 0$ for all t, we have

$$\frac{1}{T}\sum_{t=0}^{T-1}\sum_{n\in\mathcal{N}}\mathbb{E}\left[Q_n\left(t\right)\right]\leq\frac{B+V\left(\theta_1+\theta_2\right)}{\epsilon}. \quad (55)$$

Let $T\to\infty$ we obtain

$$\limsup_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}\sum_{n\in\mathcal{N}}\mathbb{E}\left[Q_n(t)\right]\leq\frac{B+V(\theta_1+\theta_2)}{\epsilon}<\infty, \quad (56)$$

which implies that

$$\limsup_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[Q_n\left(t\right)\right]<\infty,\ \forall n\in\mathcal{N}. \quad (57)$$

Appendix B
Proof of Theorem 2

We choose an optimal S-only policy which make offloading decisions $\{I_i^*(t)\}_{i\in\mathcal{A}(t)}$ in every time slot t. According to (32) and (39), we have

$$\begin{aligned}&L\left(\boldsymbol{Q}\left(t+1\right)\right)-L\left(\boldsymbol{Q}\left(t\right)\right)+V\Delta R\left(t\right)\\&\leq B+\sum_{n\in\mathcal{N}}Q_n\left(t\right)\left(E_n\left(t\right)-b_n\right)\\&+V\sum_{i\in\mathcal{A}(t)}\left(\rho_{I_i(t)}L_i(t)\mathbb{1}\left\{I_i(t)>0\right\}+\phi_{I_i(t)}W_i(t)\right)\\&-V\sum_{i\in\mathcal{A}(t)}\left(\rho_{I_i^*(t)}L_i(t)\mathbb{1}\left\{I_i^*(t)>0\right\}+\phi_{I_i^*(t)}W_i(t)\right).\end{aligned} \quad (58)$$

By expression (4) and (5), it follows that

$$\begin{aligned}&L\left(\boldsymbol{Q}\left(t+1\right)\right)-L\left(\boldsymbol{Q}\left(t\right)\right)+V\Delta R\left(t\right)\leq B\\&+\sum_{i\in\mathcal{A}(t)}\left(Q_0\left(t\right)\kappa_0\left(t\right)W_i\left(t\right)+V\phi_0W_i\left(t\right)\right)\\&\qquad\cdot\left(\mathbb{1}\left\{I_i\left(t\right)=0\right\}-\mathbb{1}\left\{I_i^*\left(t\right)=0\right\}\right)\\&+\sum_{i\in\mathcal{A}(t)}\sum_{n=1}^{N}\left(Q_n(t)\kappa_n(t)W_i(t)+Q_0(t)\eta_n(t)L_i(t)\right.\\&\qquad\left.+V\phi_nW_i\left(t\right)+V\rho_nL_i\left(t\right)\right)\\&\qquad\cdot\left(\mathbb{1}\left\{I_i\left(t\right)=n\right\}-\mathbb{1}\left\{I_i^*\left(t\right)=n\right\}\right)\\&+Q_0\left(\sum_{i\in\mathcal{A}(t)}\kappa_0\left(t\right)W_i\left(t\right)\mathbb{1}\left\{I_i^*\left(t\right)=0\right\}\right.\\&\qquad\left.+\sum_{i\in\mathcal{A}(t)}\sum_{n=1}^{N}\eta_n(t)L_i(t)\mathbb{1}\{I_i^*(t)=n\}-b_0\right)\\&+\sum_{n=1}^{N}Q_n(t)\left(\sum_{i\in\mathcal{A}(t)}\kappa_n(t)W_i(t)\mathbb{1}\{I_i^*(t)=n\}-b_n\right).\end{aligned} \quad (59)$$

Since the optimal policy $\{I_i^*(t)\}_{i,t}$ is feasible and i.i.d. over time slots, we have

$$\mathbb{E}\left[\sum_{i\in\mathcal{A}(t)}\kappa_0\left(t\right)W_i\left(t\right)\mathbb{1}\left\{I_i^*\left(t\right)=0\right\}+\sum_{i\in\mathcal{A}(t)}\sum_{n=1}^{N}\eta_n\left(t\right)L_i\left(t\right)\mathbb{1}\left\{I_i^*\left(t\right)=n\right\}\right]\leq b_0 \quad (60)$$

and

$$\mathbb{E}\left[\sum_{i\in\mathcal{A}(t)}\kappa_n\left(t\right)W_i\left(t\right)\mathbb{1}\left\{I_i^*\left(t\right)=n\right\}\right]\leq b_n,\quad\forall n\in\{1,\ldots,N\}, \quad (61)$$

i.e., the average energy consumption on each node $n\in\mathcal{N}$ in every time slot is no larger than b_n. By (60), (61) and (59), we have

$$\begin{aligned}&\mathbb{E}\left[L\left(\boldsymbol{Q}\left(t+1\right)\right)-L\left(\boldsymbol{Q}\left(t\right)\right)+V\Delta R\left(t\right)\right]\leq B\\&+\mathbb{E}\left[\sum_{i\in\mathcal{A}(t)}\left(Q_0\left(t\right)\kappa_0\left(t\right)W_i\left(t\right)+V\phi_0W_i\left(t\right)\right)\right.\\&\qquad\left.\cdot\left(\mathbb{1}\left\{I_i\left(t\right)=0\right\}-\mathbb{1}\left\{I_i^*\left(t\right)=0\right\}\right)\right]\\&+\mathbb{E}\left[\sum_{i\in\mathcal{A}(t)}\sum_{n=1}^{N}\left(Q_n\left(t\right)\kappa_n\left(t\right)W_i\left(t\right)\right.\right.\\&\qquad\left.\left.+Q_0\left(t\right)\eta_n\left(t\right)L_i\left(t\right)+V\phi_nW_i\left(t\right)+V\rho_nL_i\left(t\right)\right)\right.\\&\qquad\left.\cdot\left(\mathbb{1}\left\{I_i\left(t\right)=n\right\}-\mathbb{1}\left\{I_i^*\left(t\right)=n\right\}\right)\right].\end{aligned} \quad (62)$$

By the definition in (24), we can rewrite the above inequality as

$$\mathbb{E}\left[L\left(\boldsymbol{Q}\left(t+1\right)\right)-L\left(\boldsymbol{Q}\left(t\right)\right)+V\Delta R\left(t\right)\right]\leq B + \mathbb{E}\left[\sum_{i\in\mathcal{A}(t)}\sum_{n\in\mathcal{N}} v_{i,n}\left(t,r_n,f_n\right) \cdot\left(\mathbb{1}\left\{I_i\left(t\right)=n\right\}-\mathbb{1}\left\{I_i^*\left(t\right)=n\right\}\right)\right]. \tag{63}$$

Define

$$C_1(t)\triangleq\sum_{i\in\mathcal{A}(t)}\sum_{n\in\mathcal{N}} v_{i,n}\left(t,r_n,f_n\right) \cdot\left(\mathbb{1}\left\{I_i\left(t\right)=n\right\}-\mathbb{1}\left\{I_i^*\left(t\right)=n\right\}\right), \tag{64}$$

then

$$\mathbb{E}\left[L\left(\boldsymbol{Q}\left(t+1\right)\right)-L\left(\boldsymbol{Q}\left(t\right)\right)+V\Delta R\left(t\right)\right] \leq B+\mathbb{E}\left[C_1\left(t\right)\right]. \tag{65}$$

Summing (65) over t and dividing both sides by TV, we obtain

$$\mathbb{E}\left[L\left(\boldsymbol{Q}\left(T\right)\right)\right]-\mathbb{E}\left[L\left(\boldsymbol{Q}\left(0\right)\right)\right]+\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\Delta R\left(t\right)\right] \leq \frac{B}{V}+\frac{1}{TV}\sum_{t=0}^{T-1}\mathbb{E}\left[C_1\left(t\right)\right]. \tag{66}$$

Since $L\left(\boldsymbol{Q}\left(T\right)\right)\geq 0$ and $L\left(\boldsymbol{Q}\left(0\right)\right)=0$, we have

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\Delta R\left(t\right)\right]\leq\frac{B}{V}+\frac{1}{TV}\sum_{t=0}^{T-1}\mathbb{E}\left[C_1\left(t\right)\right]. \tag{67}$$

A. Upper Bound of $C_1(t)$

Next, we want to find the upper bound of $\mathbb{E}[C_1(t)]$. Consider a policy which makes offloading decision $\{I_i'(t)\}_{i\in\mathcal{A}(t)}$ in each time slot t such that

$$I_i'\left(t\right)=\arg\min_{i\in\mathcal{N}(t)} v_{i,n}\left(t,\rho_n,\phi_n\right),\ \forall i\in\mathcal{A}(t). \tag{68}$$

Then we have

$$\begin{aligned}&\sum_{n\in\mathcal{N}} v_{i,n}(t,\hat{\rho}_n(t),\hat{\phi}_n(t))\mathbb{1}\{I_i'(t)=n\}\\ &\leq\sum_{n\in\mathcal{N}} v_{i,n}(t,\hat{\rho}_n(t),\hat{\phi}_n(t))\mathbb{1}\{I_i^*(t)=n\},\ \forall i\in\mathcal{A}(t).\end{aligned} \tag{69}$$

It follows that our algorithm $\{I_i(t)\}_{i,t}$ satisfies

$$\begin{aligned}C_1(t)&=\sum_{i\in\mathcal{A}(t)}\sum_{n\in\mathcal{N}} v_{i,n}\left(t,\rho_n,\phi_n\right)\\ &\qquad\cdot\left(\mathbb{1}\left\{I_i\left(t\right)=n\right\}-\mathbb{1}\left\{I_i^*\left(t\right)=n\right\}\right)\\ &\leq\sum_{i\in\mathcal{A}(t)}\sum_{n\in\mathcal{N}} v_{i,n}\left(t,\rho_n,\phi_n\right)\\ &\qquad\cdot\left(\mathbb{1}\left\{I_i\left(t\right)=n\right\}-\mathbb{1}\left\{I_i'\left(t\right)=n\right\}\right)\\ &\leq\sum_{i\in\mathcal{A}(t)}\sum_{n\in\mathcal{N}} v_{i,n}\left(t,\rho_n,\phi_n\right)\\ &\qquad\cdot\left(\mathbb{1}\left\{I_i\left(t\right)=n\right\}-\mathbb{1}\left\{I_i'\left(t\right)=n\right\}\right)\\ &+\sum_{i\in\mathcal{A}(t)}\sum_{n\in\mathcal{N}} v_{i,n}(t,\hat{\rho}_n(t),\hat{\phi}_n(t))\\ &\qquad\cdot\left(\mathbb{1}\left\{I_i'\left(t\right)=n\right\}-\mathbb{1}\left\{I_i\left(t\right)=n\right\}\right),\end{aligned} \tag{70}$$

where the last equality is by (25). Rearrange the right-hand side of (70), we obtain

$$\begin{aligned}C_1(t)\leq&\sum_{i\in\mathcal{A}(t)}\sum_{n\in\mathcal{N}}(v_{i,n}(t,\rho_n,\phi_n)-v_{i,n}(t,\hat{\rho}_n(t),\hat{\phi}_n(t)))\\ &\qquad\cdot\mathbb{1}\left\{I_i\left(t\right)=n\right\}\\ &+\sum_{i\in\mathcal{A}(t)}\sum_{n\in\mathcal{N}}(v_{i,n}(t,\hat{\rho}_n(t),\hat{\phi}_n(t))-v_{i,n}(t,\rho_n,\phi_n))\\ &\qquad\cdot\mathbb{1}\left\{I_i'(t)=n\right\}.\end{aligned} \tag{71}$$

By the definition of $v_{i,n}(t,\rho,\phi)$ in (24), we have

$$\begin{aligned}&v_{i,n}(t,\rho_n,\phi_n)-v_{i,n}(t,\hat{\rho}_n(t),\hat{\phi}_n(t))\\ &=V\left(\phi_n-\hat{\phi}_n\left(t\right)\right)W_i\left(t\right)\\ &+V(\rho_n-\hat{\rho}_n(t))L_i(t)\mathbb{1}\{n>0\},\forall i\in\mathcal{A}(t),n\in\mathcal{N}.\end{aligned} \tag{72}$$

Substitute into (71), we obtain

$$\begin{aligned}C_1(t)\leq&V\sum_{i\in\mathcal{A}(t)}\sum_{n\in\mathcal{N}}\Big[(\phi_n-\hat{\phi}_n(t))W_i(t)\\ &+(\rho_n-\hat{\rho}_n(t))L_i(t)\mathbb{1}\{n>0\}\Big]\mathbb{1}\{I_i(t)=n\}\\ &+V\sum_{i\in\mathcal{A}(t)}\sum_{n\in\mathcal{N}}\Big[(\hat{\phi}_n(t)-\phi_n)W_i(t)\\ &+(\hat{\rho}_n(t)-\rho_n)L_i(t)\mathbb{1}\{n>0\}\Big]\mathbb{1}\left\{I_i'(t)=n\right\}.\end{aligned} \tag{73}$$

Define

$$\begin{aligned}C_2\left(t\right)=&\sum_{i\in\mathcal{A}(t)}\sum_{n\in\mathcal{N}}\Big[\left(\phi_n-\hat{\phi}_n\left(t\right)\right)W_i\left(t\right)\\ &+\left(\rho_n-\hat{\rho}_n\left(t\right)\right)L_i\left(t\right)\mathbb{1}\left\{n>0\right\}\Big]\mathbb{1}\left\{I_i\left(t\right)=n\right\}\end{aligned} \tag{74}$$

and

$$\begin{aligned}C_3\left(t\right)=&\sum_{i\in\mathcal{A}(t)}\sum_{n\in\mathcal{N}}\Big[\left(\hat{\phi}_n\left(t\right)-\phi_n\right)W_i\left(t\right)\\ &+\left(\hat{\rho}_n\left(t\right)-\rho_n\right)L_i\left(t\right)\mathbb{1}\left\{n>0\right\}\Big]\mathbb{1}\left\{I_i'\left(t\right)=n\right\},\end{aligned} \tag{75}$$

then the upper bound of $C_1(t)$ can be written as follows:

$$C_1\left(t\right)\leq V\left(C_2\left(t\right)+C_3\left(t\right)\right). \tag{76}$$

And it follows that

$$\sum_{t=0}^{T-1}\mathbb{E}[C_1(t)] \le V\left(\sum_{t=0}^{T-1}\mathbb{E}[C_2(t)] + \sum_{t=0}^{T-1}\mathbb{E}[C_3(t)]\right). \quad (77)$$

B. Upper Bound of $C_2(t)$

a) : Define event $G_n(t) \triangleq \left\{\phi_n < \hat{\phi}_n(t)\right\}$ for each $n \in \mathcal{N}$ and define event $J_n(t) \triangleq \{\rho_n < \hat{\rho}_n(t)\}$ for each $n \in \mathcal{N} - \{0\}$, then we have

$$\begin{aligned}
\mathbb{E}[C_2(t)] &= \sum_{n\in\mathcal{N}}\sum_{i\in\mathcal{A}(t)}\mathbb{E}\Big[(\phi_n - \hat{\phi}_n(t))W_i(t) \\
&\quad\cdot(\mathbb{1}\{G_n(t)\} + \mathbb{1}\{G_n^c(t)\})]\,\mathbb{1}\{I_i(t)=n\} \\
&+ \sum_{n\in\mathcal{N}}\sum_{i\in\mathcal{A}(t)}\mathbb{E}\left[(\rho_n - \hat{\rho}_n(t))L_i(t)\,\mathbb{1}\{n>0\}\right. \\
&\quad\left.\cdot(\mathbb{1}\{J_n(t)\} + \mathbb{1}\{J_n^c(t)\})\right]\mathbb{1}\{I_i(t)=n\} \\
&\overset{(a)}{\le} \sum_{n\in\mathcal{N}}\sum_{i\in\mathcal{A}(t)}\mathbb{E}\left[(\phi_n - \hat{\phi}_n(t))W_i(t)\mathbb{1}\{G_n^c(t)\}\right] \\
&\quad\cdot\mathbb{1}\{I_i(t)=n\} \\
&+ \sum_{n\in\mathcal{N}}\sum_{i\in\mathcal{A}(t)}\mathbb{E}[(\rho_n - \hat{\rho}_n(t))L_i(t)\mathbb{1}\{n>0\}\mathbb{1}\{J_n^c(t)\}] \\
&\quad\cdot\mathbb{1}\{I_i(t)=n\} \\
&\le w_{\max}\sum_{n\in\mathcal{N}}\sum_{i\in\mathcal{A}(t)}\mathbb{E}[(\phi_n - \hat{\phi}_n(t))\mathbb{1}\{G_n^c(t)\}] \\
&\quad\cdot\mathbb{1}\{I_i(t)=n\} \\
&+ l_{\max}\sum_{n\in\mathcal{N}}\sum_{i\in\mathcal{A}(t)}\mathbb{E}\left[(\rho_n - \hat{\rho}_n(t))\mathbb{1}\{n>0\}\mathbb{1}\{J_n^c(t)\}\right] \\
&\quad\cdot\mathbb{1}\{I_i(t)=n\}
\end{aligned} \quad (78)$$

where inequality (a) is because $(\phi_n - \hat{\phi}_n(t))\mathbb{1}\{G_n(t)\} \le 0$ and $(\rho_n - \hat{\rho}_n(t))\,\mathbb{1}\{J_n(t)\} \le 0$. Define

$$C_{4,n}(t) \triangleq \sum_{i\in\mathcal{A}(t)}\mathbb{E}[(\phi_n - \hat{\phi}_n(t))\mathbb{1}\{G_n^c(t)\}]\mathbb{1}\{I_i(t)=n\} \quad (79)$$

for each node $n \in \mathcal{N}$ and

$$C_{5,n}(t) \triangleq \sum_{i\in\mathcal{A}(t)}\mathbb{E}\left[(\rho_n - \hat{\rho}_n(t))\,\mathbb{1}\{n>0\}\,\mathbb{1}\{J_n^c(t)\}\right] \cdot\mathbb{1}\{I_i(t)=n\} \quad (80)$$

for each fog node $n \in \mathcal{N} - \{0\}$, then the upper bound of $\mathbb{E}[C_2(t)]$ can be written as

$$\begin{aligned}
\mathbb{E}[C_2(t)] &\le w_{\max}\sum_{n\in\mathcal{N}}\mathbb{E}[C_{4,n}(t)] \\
&\quad + l_{\max}\sum_{n\in\mathcal{N}-\{0\}}\mathbb{E}[C_{5,n}(t)].
\end{aligned} \quad (81)$$

Let the first time when node n is chosen be time slot t_n. Define event $K_n(t) \triangleq \left\{\phi_n - \bar{\phi}_n(t) \le \phi_{\max}\sqrt{\frac{3\log t}{2h_n(t)}}\right\}$ for all $n \in \mathcal{N}$. Summing $C_{4,n}(t)$ over time gives

$$\begin{aligned}
\sum_{t=0}^{T-1}C_{4,n}(t) &= \sum_{t=0}^{T-1}\sum_{i\in\mathcal{A}(t)}(\phi_n - \hat{\phi}_n(t))\mathbb{1}\{G_n^c(t)\} \\
&\quad\cdot\mathbb{1}\{I_i(t)=n\} \\
&\overset{(a)}{=} \sum_{t=t_n+1}^{T-1}\sum_{i\in\mathcal{A}(t)}(\phi_n - \hat{\phi}_n(t))\mathbb{1}\{G_n^c(t)\}\mathbb{1}\{I_i(t)=n\} \\
&= \sum_{t=t_n+1}^{T-1}\sum_{i\in\mathcal{A}(t)}(\phi_n - \hat{\phi}_n(t))\mathbb{1}\{G_n^c(t)\}\,(\mathbb{1}\{K_n(t)\} \\
&\quad+\mathbb{1}\{K_n^c(t)\})\,\mathbb{1}\{I_i(t)=n\} \\
&= \sum_{t=t_n+1}^{T-1}\sum_{i\in\mathcal{A}(t)}(\phi_n - \hat{\phi}_n(t))\mathbb{1}\{G_n^c(t)\cap K_n(t)\} \\
&\quad\cdot\mathbb{1}\{I_i(t)=n\} \\
&+ \sum_{t=t_n+1}^{T-1}\sum_{i\in\mathcal{A}(t)}(\phi_n - \hat{\phi}_n(t))\mathbb{1}\{K_n^c(t)\}\mathbb{1}\{I_i(t)=n\} \\
&\overset{(b)}{\le} \sum_{t=t_n+1}^{T-1}\sum_{i\in\mathcal{A}(t)}(\phi_n - \hat{\phi}_n(t))\mathbb{1}\{G_n^c(t)\cap K_n(t)\} \\
&\quad\cdot\mathbb{1}\{I_i(t)=n\} \\
&+ \phi_{\max}\sum_{t=t_n+1}^{T-1}\sum_{i\in\mathcal{A}(t)}\mathbb{1}\{K_n^c(t)\}\,\mathbb{1}\{I_i(t)=n\}
\end{aligned} \quad (82)$$

where equality (a) is because $\mathbb{1}\{I_i(t)=n\} = 0$ when $t \le t_n$, and inequality (b) is because of $\phi_n \le \phi_{\max}$ and $\hat{\phi}_n(t) \ge 0$. Define

$$U_{1,n}(t) \triangleq \sum_{i\in\mathcal{A}(t)}(\phi_n - \hat{\phi}_n(t))\mathbb{1}\{G_n^c(t)\cap K_n(t)\} \cdot\mathbb{1}\{I_i(t)=n\} \quad (83)$$

and

$$U_{2,n}(t) \triangleq \sum_{i\in\mathcal{A}(t)}\mathbb{1}\{K_n^c(t)\}\,\mathbb{1}\{I_i(t)=n\}, \quad (84)$$

then by (82) we have

$$\begin{aligned}
&\sum_{t=0}^{T-1}\sum_{n\in\mathcal{N}}\mathbb{E}[C_{4,n}(t)] \\
&\quad\le \sum_{t=t_n+1}^{T-1}\sum_{n\in\mathcal{N}}\left(\mathbb{E}[U_{1,n}(t)] + \phi_{\max}\mathbb{E}[U_{2,n}(t)]\right).
\end{aligned} \quad (85)$$

When event $K_n(t)$ happens, $f_n - \bar{f}_n(t) \leq f_{\max}\sqrt{\frac{3\log t}{2h_n(t)}}$. Since $\hat{f}_n(t) = \max\{\bar{f}_n(t) - f_{\max}\sqrt{\frac{3\log t}{2h_n(t)}}, 0\}$, we have

$$\phi_n - \hat{\phi}_n(t) = \left(\phi_n - \bar{\phi}_n(t)\right) + \left(\bar{\phi}_n(t) - \hat{\phi}_n(t)\right) \leq 2\phi_{\max}\sqrt{\frac{3\log t}{2h_n(t)}}. \quad (86)$$

It follows that

$$\begin{aligned}
\sum_{t=t_n+1}^{T-1} U_{1,n}(t) &= \sum_{t=t_n+1}^{T-1}\sum_{i\in\mathcal{A}(t)}\left(\phi_n - \hat{\phi}_n(t)\right) \\
&\quad \cdot \mathbb{1}\{G_n^c(t)\cap K_n(t)\}\mathbb{1}\{I_i(t)=n\} \\
&\leq \sum_{t=t_n+1}^{T-1}\sum_{i\in\mathcal{A}(t)} 2\phi_{\max}\sqrt{\frac{3\log t}{2h_n(t)}}\mathbb{1}\{G_n^c(t)\cap K_n(t)\} \\
&\quad \cdot \mathbb{1}\{I_i(t)=n\} \\
&\leq \sum_{t=t_n+1}^{T-1}\sum_{i\in\mathcal{A}(t)} 2\phi_{\max}\sqrt{\frac{3\log t}{2h_n(t)}}\mathbb{1}\{I_i(t)=n\} \\
&\leq \sum_{t=t_n+1}^{T-1}\sum_{i\in\mathcal{A}(t)} 2\phi_{\max}\sqrt{\frac{3\log T}{2h_n(t)}}\mathbb{1}\{I_i(t)=n\} \\
&= \phi_{\max}\sqrt{6\log T}\sum_{t=t_n+1}^{T-1}\frac{1}{\sqrt{h_n(t)}}\sum_{i\in\mathcal{A}(t)}\mathbb{1}\{I_i(t)=n\}
\end{aligned} \quad (87)$$

Let $a_n(t)$ be the number of times that node n is chosen in time slot t under our policy, *i.e.* $a_n(t) = \sum_{i\in\mathcal{A}(t)} I_i(t)$, then it follows that

$$\sum_{t=t_n+1}^{T-1} U_{1,n}(t) \leq \phi_{\max}\sqrt{6\log T}\sum_{t=t_n+1}^{T-1}\frac{a_n(t)}{\sqrt{h_n(t)}}. \quad (88)$$

Let $t_{n,m}$ be the mth time slot when node n is chosen, and let $M_n(T-1)$ be the time slot when node n is lastly chosen before time slot T. Then we have

$$\begin{aligned}
\sum_{t=t_n+1}^{T-1}\frac{a_n(t)}{\sqrt{h_n(t)}} &= \sum_{m=1}^{M_n(T-1)}\frac{a_n(t_{n,m})}{\sqrt{h_n(t_{n,m})}} \\
&\overset{(a)}{\leq} \sum_{m=1}^{M_n(T-1)}\frac{a_{\max}}{\sqrt{m}} \overset{(b)}{\leq} a_{\max}\left(1 + \int_1^{M_n(T-1)}\frac{1}{\sqrt{m}}\right) \\
&= 2a_{\max}\sqrt{M_n(T-1)} \leq 2a_{\max}\sqrt{h_n(T-1)}.
\end{aligned} \quad (89)$$

Inequality (a) is because $a_n(t) \leq a_{\max}$ and $h_n(t_{n,m}) \geq m$. Inequality (b) is because of the basic relationship between summation and integral. Then it follows that

$$\sum_{t=t_n+1}^{T-1} U_{1,n}(t) \leq 2a_{\max}\phi_{\max}\sqrt{6h_n(T-1)\log T}. \quad (90)$$

By Jensen's inequality, we have

$$\begin{aligned}
\frac{1}{N+1}\sum_{n\in\mathcal{N}}\sqrt{h_n(T-1)} &\leq \sqrt{\frac{1}{N+1}\sum_{n\in\mathcal{N}} h_n(T-1)} \\
&\leq \sqrt{\frac{1}{N+1}(a_{\max}T)}.
\end{aligned} \quad (91)$$

Thus it follows that

$$\begin{aligned}
&\sum_{t=t_n+1}^{T-1}\sum_{n\in\mathcal{N}}\mathbb{E}[U_{1,n}(t)] \\
&\qquad \leq 2a_{\max}\phi_{\max}\sqrt{a_{\max}(N+1)T\log T}.
\end{aligned} \quad (92)$$

On the other hand, by using the Chernoff-Hoeffding bound we have

$$\begin{aligned}
&\sum_{t=t_n+1}^{T-1}\sum_{n\in\mathcal{N}}\mathbb{E}[U_{2,n}(t)] \\
&= \sum_{t=t_n+1}^{T-1}\sum_{n\in\mathcal{N}}\sum_{i\in\mathcal{A}(t)}\mathbb{E}[\mathbb{1}\{K_n^c(t)\}]\mathbb{1}\{I_i(t)=n\} \\
&= \sum_{t=t_n+1}^{T-1}\sum_{n\in\mathcal{N}} a_n(t)\,\mathbb{E}[\mathbb{1}\{K_n^c(t)\}] \\
&= \sum_{t=t_n+1}^{T-1}\sum_{n\in\mathcal{N}} a_n(t)\Pr\{K_n^c(t)\} \\
&= \sum_{t=t_n+1}^{T-1}\sum_{n\in\mathcal{N}} a_n(t)\Pr\left\{\phi_n - \bar{\phi}_n(t) > \phi_{\max}\sqrt{\frac{2\log t}{h_n(t)}}\right\} \\
&\leq \sum_{t=t_n+1}^{T-1}\sum_{n\in\mathcal{N}} a_n(t)\exp\left(-\frac{2(h_n(t))^2}{h_i(t)\phi_{\max}^2}\cdot\phi_{\max}^2\frac{3\log t}{2h_n(t)}\right) \\
&= \sum_{t=t_n+1}^{T-1}\sum_{n\in\mathcal{N}} a_n(t)\exp(-3\log t) \\
&\leq \sum_{t=1}^{T-1}\sum_{n\in\mathcal{N}} a_n(t)\,t^{-3} \\
&\leq a_{\max}\sum_{t=1}^{\infty} t^{-3} = a_{\max}\left(1 + \sum_{t=2}^{\infty} t^{-3}\right) \\
&\leq a_{\max}\left(1 + \int_1^{\infty} t^{-3}dt\right) = \frac{3}{2}a_{\max}.
\end{aligned} \quad (93)$$

Taking expectation of both sides of (82) and by (90) (93) we obtain

$$\begin{aligned}
&\sum_{t=0}^{T-1}\sum_{n\in\mathcal{N}}\mathbb{E}[C_{4,n}(t)] \\
&\quad \leq 2a_{\max}\phi_{\max}\sqrt{a_{\max}(N+1)T\log T} + \frac{3}{2}a_{\max}\phi_{\max}.
\end{aligned} \quad (94)$$

Similarly, we have

$$\sum_{t=0}^{T-1}\sum_{n\in\mathcal{N}}\mathbb{E}\left[C_{5,n}\left(t\right)\right] \leq 2a_{\max}\rho_{\max}\sqrt{a_{\max}NT\log T}+\frac{3}{2}a_{\max}\rho_{\max}. \quad (95)$$

Summing $\mathbb{E}(C_2(t))$ over time slots and by (81) we have

$$\sum_{t=0}^{T-1}\mathbb{E}\left[C_2\left(t\right)\right] \leq w_{\max}\sum_{t=0}^{T-1}\sum_{n\in\mathcal{N}}\mathbb{E}\left[C_{4,n}\left(t\right)\right] + l_{\max}\sum_{t=0}^{T-1}\sum_{n=1}^{N}\mathbb{E}\left[C_{5,n}\left(t\right)\right]. \quad (96)$$

Plugging (94) and (95) into it yield

$$\begin{aligned}&\sum_{t=0}^{T-1}\mathbb{E}\left[C_2\left(t\right)\right]\\ &\leq w_{\max}a_{\max}\phi_{\max}\left(2\sqrt{6a_{\max}\left(N+1\right)T\log T}+\frac{3}{2}\right) \quad (97)\\ &+ l_{\max}a_{\max}\rho_{\max}\left(2\sqrt{6a_{\max}NT\log T}+\frac{3}{2}\right).\end{aligned}$$

C. Upper Bound of $C_3(t)$

Recall that

$$C_3\left(t\right)=\sum_{i\in\mathcal{A}(t)}\sum_{n\in\mathcal{N}}\left[\left(\hat{\phi}_n\left(t\right)-\phi_n\right)W_i\left(t\right) + \left(\hat{\rho}_n\left(t\right)-\rho_n\right)L_i\left(t\right)\mathbb{1}\left\{n>0\right\}\right]\mathbb{1}\left\{I_i'\left(t\right)=n\right\}. \quad (98)$$

We bound the expectation of $C_3(t)$ as

$$\begin{aligned}\mathbb{E}\left[C_3\left(t\right)\right]&=\mathbb{E}\left[\sum_{i\in\mathcal{A}(t)}\sum_{n\in\mathcal{N}}\left[\left(\hat{\phi}_n\left(t\right)-\phi_n\right)W_i\left(t\right)\right.\right.\\ &\left.\left.\cdot\left(\mathbb{1}\left\{G_n\left(t\right)\right\}+\mathbb{1}\left\{G_n^c\left(t\right)\right\}\right)\right]\mathbb{1}\left\{I_i'\left(t\right)=n\right\}\right]\\ &+\mathbb{E}\left[\sum_{i\in\mathcal{A}(t)}\sum_{n\in\mathcal{N}}\left[\left(\hat{\rho}_n\left(t\right)-\rho_n\right)L_i\left(t\right)\mathbb{1}\left\{n>0\right\}\right.\right.\\ &\left.\left.\cdot\left(\mathbb{1}\left\{J_n\left(t\right)\right\}+\mathbb{1}\left\{J_n^c\left(t\right)\right\}\right)\right]\mathbb{1}\left\{I_i'\left(t\right)=n\right\}\right]\\ &\leq\mathbb{E}\left[\sum_{i\in\mathcal{A}(t)}\sum_{n\in\mathcal{N}}\left[\left(\hat{\phi}_n(t)-\phi_n\right)W_i(t)\mathbb{1}\left\{G_n\left(t\right)\right\}\right]\right.\\ &\left.\cdot\mathbb{1}\left\{I_i'\left(t\right)=n\right\}\right]\\ &+\mathbb{E}\left[\sum_{i\in\mathcal{A}(t)}\sum_{n\in\mathcal{N}}\left[\left(\hat{\rho}_n(t)-\rho_n\right)L_i(t)\mathbb{1}\left\{J_n(t)\right\}\right]\right.\\ &\left.\cdot\mathbb{1}\left\{n>0\right\}\mathbb{1}\left\{I_i'\left(t\right)=n\right\}\right]\\ &\leq w_{\max}\mathbb{E}\left[\sum_{i\in\mathcal{A}(t)}\sum_{n\in\mathcal{N}}\left[\left(\hat{\phi}_n(t)-\phi_n\right)\mathbb{1}\left\{G_n(t)\right\}\right]\right.\\ &\left.\cdot\mathbb{1}\left\{I_i'\left(t\right)=n\right\}\right]\\ &+l_{\max}\mathbb{E}\left[\sum_{i\in\mathcal{A}(t)}\sum_{n\in\mathcal{N}}\left[\left(\hat{\rho}_n(t)-\rho_n\right)\mathbb{1}\left\{n>0\right\}\mathbb{1}\left\{J_n(t)\right\}\right]\right.\\ &\left.\cdot\mathbb{1}\left\{I_i'(t)=n\right\}\right].\end{aligned} \quad (99)$$

We define

$$C_{6,n}(t)\triangleq\sum_{i\in\mathcal{A}(t)}\left[\left(\hat{\phi}_n(t)-\phi_n\right)\mathbb{1}\{G_n(t)\}\right]\mathbb{1}\{I_i'(t)=n\} \quad (100)$$

for each $n\in\mathcal{N}$, and define

$$C_{7,n}(t)\triangleq\sum_{i\in\mathcal{A}(t)}\left[\left(\hat{\rho}_n(t)-\rho_n\right)\mathbb{1}\{n>0\}\mathbb{1}\left\{J_n\left(t\right)\right\}\right]\cdot\mathbb{1}\left\{I_i'\left(t\right)=n\right\} \quad (101)$$

for each $n\in\mathcal{N}-\{0\}$.

We consider the case when $t\leq t_n$ and $t\geq t_n+1$ separately. When $t\leq t_n$, $\hat{f}_n(t)=0$ and the event $G_n\left(t\right)=\left\{\phi_n<\hat{\phi}_n\left(t\right)\right\}$ will not happen. Thus $C_{6,n}(t)=0$ when $t\leq t_n$.

When $t\geq t_n+1$, suppose event $G_n(t)$ happens. Then we have $\hat{\phi}_n(t)\geq\phi_n\geq 0$, which implies that $\hat{\phi}_n=\bar{\phi}_n\left(t\right)-\phi_{\max}\sqrt{\frac{3\log t}{2h_n(t)}}$ and it follows that $\phi_n\leq\bar{\phi}_n\left(t\right)-\phi_{\max}\sqrt{\frac{3\log t}{2h_n(t)}}$.

Thus we can bound $\mathbb{E}[C_{6,n}(t)]$ as follows:

$$\begin{aligned}
&\mathbb{E}\left[C_{6,n}\left(t\right)\right] \\
&= \sum_{i\in\mathcal{A}(t)} \mathbb{E}\left[\left(\hat{\phi}_n\left(t\right)-\phi_n\right)\mathbb{1}\left\{G_n\left(t\right)\right\}\right]\mathbb{1}\{I_i'(t)=n\} \\
&\leq \phi_{\max}\sum_{i\in\mathcal{A}(t)} \mathbb{E}\left[\mathbb{1}\left\{G_n\left(t\right)\right\}\right]\mathbb{1}\left\{I_i'\left(t\right)=n\right\} \\
&\leq \phi_{\max}\sum_{i\in\mathcal{A}(t)} \Pr\left\{G_n\left(t\right)\right\}\mathbb{1}\left\{I_i'\left(t\right)=n\right\} \\
&= \phi_{\max}\sum_{i\in\mathcal{A}(t)} \Pr\left\{\phi_n \leq \bar{\phi}_n(t)-\phi_{\max}\sqrt{\frac{3\log t}{2h_n\left(t\right)}}\right\} \\
&\qquad\qquad \cdot\mathbb{1}\left\{I_i'\left(t\right)=n\right\}.
\end{aligned} \tag{102}$$

By Chernoff-Hoeffding bound,

$$\begin{aligned}
&\Pr\left\{\phi_n \leq \bar{\phi}_n\left(t\right)-\phi_{\max}\sqrt{\frac{3\log t}{2h_n\left(t\right)}}\right\} \\
&= \Pr\left\{\bar{\phi}_n\left(t\right)\geq \phi_n+\phi_{\max}\sqrt{\frac{3\log t}{2h_n\left(t\right)}}\right\} \\
&\leq \exp\left(-\frac{2\left(h_n\left(t\right)\right)^2}{h_i\left(t\right)\phi_{\max}^2}\cdot\phi_{\max}^2\frac{3\log t}{2h_n\left(t\right)}\right) \\
&= \exp\left(-3\log t\right)=t^{-3}.
\end{aligned} \tag{103}$$

Thus we have

$$\mathbb{E}\left[C_{6,n}\left(t\right)\right]\leq \phi_{\max}a_n'\left(t\right)t^{-3}, \tag{104}$$

where $a_n'(t)$ is the number of times that node n is chosen in time slot t by policy $\{I_i'(t)\}$, *i.e.*, $a_n'\left(t\right)=\sum_{i\in\mathcal{A}(t)}I_i'\left(t\right)$.

Summing $\mathbb{E}[C_{6,n}(t)]$ over time slots $\{0,\ldots,T-1\}$ and nodes $n\in\mathcal{N}$, and applying (104), we obtain

$$\begin{aligned}
&\sum_{t=0}^{T-1}\sum_{n\in\mathcal{N}}\mathbb{E}\left[C_{6,n}\left(t\right)\right]\leq \phi_{\max}\sum_{n\in\mathcal{N}}\sum_{t=t_n+1}^{T-1}a_n'\left(t\right)t^{-3} \\
&\leq \phi_{\max}\sum_{n\in\mathcal{N}}\sum_{t=1}^{T-1}a_n'\left(t\right)t^{-3} \\
&= \phi_{\max}\sum_{t=1}^{T-1}A\left(t\right)t^{-3}\leq \phi_{\max}a_{\max}\sum_{t=1}^{T-1}t^{-3} \\
&\leq \phi_{\max}a_{\max}\sum_{t=1}^{\infty}t^{-3}=\phi_{\max}a_{\max}\left(1+\sum_{t=2}^{\infty}t^{-3}\right) \\
&\leq \phi_{\max}a_{\max}\left(1+\int_{t=1}^{\infty}\frac{1}{t^3}dt\right)=\frac{3}{2}\phi_{\max}a_{\max}.
\end{aligned} \tag{105}$$

Similarly, we have

$$\sum_{t=0}^{T-1}\sum_{n=1}^{N}\mathbb{E}\left[C_{7,n}\left(t\right)\right]\leq\frac{3}{2}\rho_{\max}a_{\max} \tag{106}$$

for every fog node $n\in\mathcal{N}-\{0\}$.

Summing $\mathbb{E}[C_3(t)]$ over time slots and by (99) (100), and (101) we have

$$\begin{aligned}
\sum_{t=0}^{T-1}\mathbb{E}\left[C_3\left(t\right)\right]&\leq w_{\max}\sum_{t=0}^{T-1}\sum_{n\in\mathcal{N}}\mathbb{E}\left[C_{6,n}\left(t\right)\right] \\
&\quad + l_{\max}\sum_{t=0}^{T-1}\sum_{n=1}^{N}\mathbb{E}\left[C_{7,n}\left(t\right)\right].
\end{aligned} \tag{107}$$

Plugging (105) and (106) into it yield

$$\sum_{t=0}^{T-1}\mathbb{E}\left[C_3\left(t\right)\right]\leq\frac{3}{2}w_{\max}\phi_{\max}a_{\max}+\frac{3}{2}l_{\max}\rho_{\max}a_{\max}. \tag{108}$$

D. Upper Bound of Regret

Plugging (77), (97) and (108) into (67), we obtain

$$\begin{aligned}
&\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\Delta R\left(t\right)\right] \\
&\leq \frac{B}{V}+\left(\frac{3}{T}+\sqrt{\frac{3a_{\max}\left(N+1\right)\log T}{2T}}\right)\frac{\theta_1}{2} \\
&\quad +\left(\frac{3}{T}+\sqrt{\frac{3a_{\max}N\log T}{2T}}\right)\frac{\theta_2}{2}
\end{aligned} \tag{109}$$

where $B=\frac{1}{2}a_{\max}\max\left\{\kappa_{\max}l_{\max},\eta_{\max}w_{\max}\right\}+\frac{1}{2}Na_{\max}\kappa_{\max}l_{\max}+\frac{1}{2}\sum_{n\in\mathcal{N}}b_n^2$, $\theta_1=2w_{\max}\phi_{\max}a_{\max}$, $\theta_2=2l_{\max}\rho_{\max}a_{\max}$.

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] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 637–646, 2016.

[3] 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.

[4] L. Pu, X. Chen, J. Xu, and X. Fu, "D2d fogging: An energy-efficient and incentive-aware task offloading framework via network-assisted d2d collaboration," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3887–3901, 2016.

[5] M. J. Neely, "Stochastic network optimization with application to communication and queueing systems," *Synthesis Lectures on Communication Networks*, vol. 3, no. 1, pp. 1–211, 2010.

[6] S. Ghoorchian and S. Maghsudi, "Multi-armed bandit for energy-efficient and delay-sensitive edge computing in dynamic networks with uncertainty," *arXiv preprint arXiv:1904.06258*, 2019.

[7] F. Li, J. Liu, and B. Ji, "Combinatorial sleeping bandits with fairness constraints," *arXiv preprint arXiv:1901.04891*, 2019.

[8] Y. Sun, S. Zhou, and J. Xu, "Emm: Energy-aware mobility management for mobile edge computing in ultra dense networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 11, pp. 2637–2646, 2017.

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

[10] G. Zhang, F. Shen, Z. Liu, Y. Yang, K. Wang, and M.-T. Zhou, "Femto: Fair and energy-minimized task offloading for fog-enabled iot networks," *IEEE Internet of Things Journal (Early Access)*, 2018. doi:10.1109/JIOT.2018.2887229.

[11] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine Learning*, vol. 47, no. 2-3, pp. 235–256, 2002.

[12] J.-Y. Audibert and S. Bubeck, "Minimax policies for adversarial and stochastic bandits," in *Proceedings of COLT*, 2009.
[13] T. Lattimore and C. Szepesvári, "Bandit algorithms," *preprint*, 2018.
[14] J. Vermorel and M. Mohri, "Multi-armed bandit algorithms and empirical evaluation," in *Proceedings of ECML*, 2005.