IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 13, 2020, accepted March 10, 2020, date of publication March 20, 2020, date of current version March 31, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2982356

Latency and Energy Optimization for MEC Enhanced SAT-IoT Networks

GAOFENG CUI[1,2], (Member, IEEE), XIAOYAO LI[1], LEXI XU[3], (Member, IEEE), AND WEIDONG WANG[1,2], (Member, IEEE)

[1]School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2]Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China
[3]Network Technology Research Institute, China United Network Communications Corporation, Beijing 100048, China

Corresponding authors: Gaofeng Cui (cuigaofeng@bupt.edu.cn) and Lexi Xu (davidlexi@hotmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61971054 and Grant 61601045.

ABSTRACT Mobile edge computing (MEC) enhanced satellite based internet of things (SAT-IoT) is an important complement for terrestrial networks based IoT, especially for the remote and depopulated areas. For MEC enhanced SAT-IoT networks with multiple satellites and multiple satellite gateways, the coupled user association, offloading decision, computing and communication resource allocation should be jointly optimized to minimize the latency and energy cost. In this paper, the latency and energy optimization for MEC enhanced SAT-IoT networks are formulated as a dynamic mixed-integer programming problem, which is hard to obtain the optimal solutions. To tackle this problem, we decompose the complex problem into two sub-problems. The first one is computing and communication resource allocation with fixed user association and offloading decision, and the second one is joint user association and offloading with optimal resource allocation. For the sub-problem of resource allocation, the optimal solution is proven to be obtained based on Lagrange multiplier method. And then, the second sub-problem is further formulated as a Markov decision process (MDP), and a joint user association and offloading decision with optimal resource allocation (JUAOD-ORA) is proposed based on deep reinforcement learning (DRL). Simulation results show that the proposed approach can achieve better long-term reward in terms of latency and energy cost.

INDEX TERMS Latency and energy optimization, MEC, SAT-IoT, deep reinforcement learning.

I. INTRODUCTION

Internet of things (IoT) plays an important role in future intelligent networks. To provide reliable connection and high-quality service for massive devices in IoT, the fifth generation (5G) wireless networks treat massive connection as an indispensable component and devote many efforts to make it satisfy the requirements of tremendous emerging services [1]. For the conceiving sixth generation (6G) wireless networks, techniques utilized to support IoT services will still be highlighted [2]. However, most of existing techniques for IoT are based on the terrestrial networks, such as long range (LoRa), narrow band IoT (NB-IoT), etc. These terrestrial networks can work well with complementary telecommunication infrastructures. However, they may not work effectively for the remote areas, such as sea and depopulated areas.

To provide seamless coverage and continuous services, satellite has become an important component for networks beyond 5G (B5G) or 6G [3]. Specially, satellites can support massive IoT devices geo-distributed widely, and they can also provide high-efficient backhaul links for terrestrial network based IoT [4]. In [5], Sanctis et al. describe satellite based internet of things (SAT-IoT) and further discuss several issues for SAT-IoT, such as quality of service (QoS) management, network interoperability etc. In [6], heterogenous space and terrestrial integrated networks for IoT is given and analyzed with several research challenges. The typical IoT applications and services in space information networks are given in [7], and four types of traffic are classified according to the delay tolerance. In [8], the internet of space things is introduced with software defined networking (SDN) and network function virtualization (NFV). Low earth orbit (LEO) satellite constellation is proposed for IoT in [9], and the comparison between SAT-IoT and terrestrial networks based IoT is also provided. In [10], small satellites are combined to form a freely-drifting swarm for IoT in the Arctic areas.

With the development and deployment of SAT-IoT networks, data processing and resource management for

The associate editor coordinating the review of this manuscript and approving it for publication was Zhenyu Zhou.

SAT-IoT has become an important and challenging issue that needs to be paid more attentions. Meanwhile, traditional satellite networks with few gateways cannot cope with the situation of massive data transmission and processing effectively. Mobile edge computing (MEC) enhanced satellite networks are emerging to improve QoS of high-speed satellite-terrestrial networks [11]. In MEC enhanced satellite networks, caching and computing resource can work collaboratively with communication resource to reduce latency and save energy, as well as provide reliable service with improved users experience. Since IoT devices are usually low power equipments with limited communication and computing resource, offloading the data generated by IoT devices to the satellites or gateways for further processing is a promising choice and a widely adopted solution for on-orbit or planning satellite systems [12]. However, the MEC enhanced SAT-IoT will make the resource management more complicate, and several issues such as user association, offloading decision, computing and communication resource allocation should be considered cooperatively to improve the network energy efficiency and reduce latency.

A. RELATED WORK

Radio resource allocation for the forward links of multibeam satellite networks is analyzed in [13] with flexible satellite payload. In [14], integrated satellite-ground industrial IoT is proposed and beam power is optimized with non-orthogonal multiple access (NOMA) to match the required transmission rate. Joint beamforming design and resource allocation for terrestrial-satellite cooperative systems is investigated in [15]. Deep learning based long-term power allocation is analyzed in [16] for NOMA downlink in SAT-IoT networks. However, these literatures mainly focus on the management of communication resource, whilst the joint computing and communication resource management, which will affect the latency and QoS in SAT-IoT networks, is not taken into consideration.

Joint computing and communication resource management has been mainly studied for terrestrial networks. MEC is investigated in detail from the communication perspective in [17]. Hassan *et al.* analyze the role of edge computing in IoT [18]. Liao *et al.* propose a learning-based resource allocation for edge computing enhanced industrial IoT [19]. Cui *et al.* study the tradeoff between the energy consumption and latency for MEC based IoT networks in [20], and a constrained multi-objective optimization problem is formulated and solved by the non-dominated sorting genetic algorithm. Cao *et al.* consider a three-node MEC system with partial and binary offloading models [21], and a joint computing and communication scheme is proposed to improve the energy efficiency of the nodes. Zhou *et al.* analyze the vehicular fog computing in [22] with information asymmetry and information uncertainty. Ning *et al.* construct a three-layer offloading framework for intelligent internet of vehicles to minimize the overall energy consumption via deep reinforcement learning (DRL) [23]. However, all of the above works do not consider the effects of user association, while the number of IoT devices is usually larger than that of access points. The admission control, computational resource allocation and power control are jointly optimized for MEC enhanced IoT networks in [24], the joint computation offloading and user association for multi-tasks MEC systems is investigated in [25] to minimize overall energy consumption. However, all of the above literatures mainly focus on terrestrial networks based IoT, and the backhaul links from the base stations or access points to the central units are assumed to be ideal with fiber connection. While for the SAT-IoT networks, the fast-moving LEO satellites will make the channel change quickly. Moreover, the power available for the satellites is very limited, and the links from the IoT devices to the satellites and satellites to the gateways are both wireless links, which are quite different from the terrestrial networks based IoT.

The mobility of caching, computing and communication resource is analyzed in [26], and an optimal resource mobility utilization strategy is devised. Huang *et al.* analyze the problem of collecting data from IoT gateways through LEO satellite under time-varying uplinks in an energy-efficient way [27], and the energy consumption is minimized based on Lyapunov optimization. However, the offloading decision and computing resource allocation are not taken into consideration. The delay and power consumption for terrestrial-satellite systems are modeled and analyzed in [28] with joint computing and communication allocation, and the joint optimization problem is solved by using dual decomposition method. Wang *et al.* conduct a computation offloading game framework for MEC enhanced satellite networks with considering the intermittent communication caused by satellite orbiting, and the response time and energy consumption of the tasks are optimized with the Nash equilibrium [29]. In [30], Cheng *et al.* design a space-air-ground integrated network (SAGIN) architecture, where flying unmanned aerial vehicles (UAVs) act as edge servers and can connect to the cloud servers via satellites. A policy gradient-based actor-critic learning algorithm, which can achieve near-optimal performance with low complexity, is proposed to make the offloading decision, and a heuristic algorithm is adopted to allocate computation resources for tasks. However, the existing works ignore the joint optimization of user association, offloading decision, computing and communication resource allocation for MEC enhanced SAT-IoT networks.

B. AIMS AND SCOPE

In this paper, we focus on the MEC enhanced SAT-IoT networks with multiple satellites and multiple satellite gateways. More specifically, satellite gateways serve as distributed clouds, which can provide abundant computing resources. Satellites act as edge computing nodes and provide access to satellite gateways for IoT devices. Since the satellites are power limited, and the on-board computing and communication resource are scarce, joint user association, offloading decision, computing and communication resource allocation are investigated to reduce the service latency and power

consumption of satellites. Obviously, the joint optimization problem will be affected by several factors coupled with each others, and it is hard to obtain the optimal solutions with existing methods or standalone optimization. Therefore, we decompose the complex problem into two sub-problems, which are computing and communication resource allocation with fixed user association and offloading decision, as well as joint user association and offloading with optimal resource allocation. To achieve long-term reward in terms of latency and power consumption, Lagrange multiplier method and deep Q-Network (DQN) are utilized collaboratively to tackle the two sub-problems. The main contributions of this paper can be summarized as follows:

- We present a framework for latency and energy optimization in MEC enhanced SAT-IoT networks with multiple satellites and multiple satellite gateways. Unlike the existing methods, user association, offloading decision, computing and communication resource allocation are jointly investigated by decomposing the complex problem into two related sub-problems.
- For the computing and communication resource allocation with fixed user association and offloading decision, the optimal solution can be obtained by using Lagrange multiplier method. While for the joint user association and offloading decision sub-problem, we formulate it as a Markov decision process (MDP) with large state and action space, and DQN is adopted to maximize the long-term reward in terms of latency and power consumption.
- The performance of the proposed DRL-based joint user association and offloading decision with optimal resource allocation (JUAOD-ORA) approach is evaluated through extensive simulation and compared with reference schemes. Simulation results illustrate the effectiveness of the proposed approach.

The rest of this paper is organized as follows. In Section II, we present the latency and energy cost model and formulate the optimization problem. In Section III, the optimal resource allocation with fixed user association and offloading decision is shown in subsection A, and then the joint user association and offloading decision with optimal resource allocation is presented with DRL-based method in the following subsection. Section IV presents the simulation results and evaluates the proposed approach. Section V concludes this paper.

II. SYSTEM MODEL AND PROBLEM FORMULATION

Fig. 1 shows a typical scenario of MEC enhanced SAT-IoT networks with multiple satellites and multiple gateways. It is assumed that each IoT terminal can access one satellite at most, but each satellite can set up multiple links between one satellite and gateways for the large onboard antennas and geographical separated gateways. IoT terminals in this paper are assumed to be sink nodes that can aggregate data from their surrounding sensors. Since the sink nodes can only have local information, they need to forward the aggregated

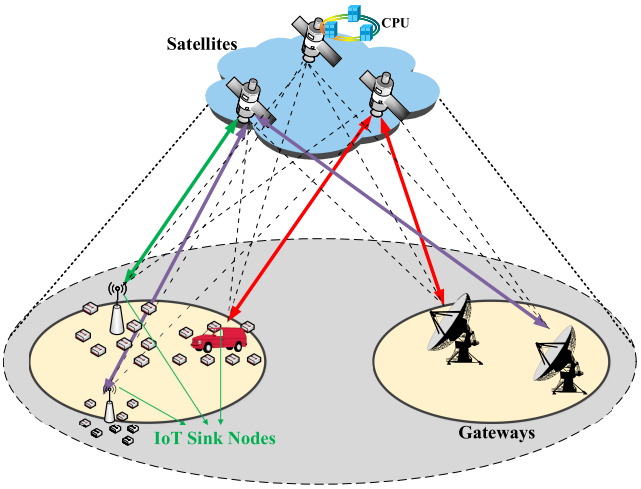

FIGURE 1. MEC enhanced SAT-IoT networks.

data to satellites or satellite gateways for further information processing or extraction. Thus, the aggregated data collected and post-processed by each IoT terminal is packed as a task that waits to be transmitted and handled. Every IoT terminal has one task that needs to be handled, and the total number of tasks is assumed to be K. Moreover, every task is assumed to be offloaded to the satellites or the satellite gateways without local processing, and the processing results will be sent back to IoT terminals. Although some IoT terminals may have the ability to handle tasks locally, we focus on the aggregated services [4] that need to be handled remotely, and our method can be extended to the situations with local processing easily. If the task is offloaded to the gateways with high computation capacity, the processing latency will be low, but it suffers from large propagation and transmission latency. While for the case of task processing at the satellites, the scarce computation and energy resources will bring large processing latency.

Moreover, the tasks will be scheduled slot by slot. At each time slot, several tasks will be selected to be offloaded to satellites or gateways with orthogonal resource. Once a task k being scheduled, its allocated resource may last for several time slots till the IoT terminals receive the feedbacks from satellites or gateways. This resource allocation mechanism, also named as demand assigned multiple access (DAMA) is commonly used in SAT-IoT networks to guarantee the efficiency and reliability of communication links [5] [30]. Therefore, there may be no or extremely few resources available for the unserved tasks at some time slots. To minimize the weighted-sum latency of all of tasks and energy cost, every task needs to be handled by jointly optimizing user association, offloading decision, computing and communication resource allocation. The system models adopted and problem formulation are listed in the following subsections.

A. LATENCY MODEL

For every task k, the latency T_k is composed of four components, which are waiting latency, transmission latency,

propagation latency and processing latency. Since the task can be handled at the satellites or the gateways, T_k will be analyzed with two possible cases.

If task k is handled at the satellite s, the latency T_k^S can be expressed as,

$$T_k^S = \rho(l-1) + \frac{N_k}{C_{k,s}^l} + \frac{N_k}{z_{k,s}^l} + P_{k,s}^l, \tag{1}$$

where ρ is the length of one time slot, and $\rho(l-1)$ denotes the waiting latency of task k which is scheduled at lth time slot. N_k is the number of bits in task k. $C_{k,s}^l$ denotes the allocated communication capacity of the link from terminal to the satellite for the task k at time slot l. The allocated capacity can be achieved with several ways, such as orthogonal sub-bands or time-frequency blocks. $z_{k,s}^l$ is the computing capacity allocated to the task k by the satellite s. $P_{k,s}^l = \frac{2d_{k,s}}{c}$ is the propagation latency for task k served by satellite s, where $d_{k,s}$ is the distance between IoT terminal and satellite, and c is the speed of light. Please note that the transmission latency from satellite to the terminal is omitted, because there is usually very small number of bits in the return link for the IoT services, such as results of object identification and acknowledgment information. However, the propagation latency from the satellite to the terminals cannot be omitted with $2d_{k,s}$ which is included in $P_{k,s}^l$.

If task k is handled at the gateway g, it will be transferred via the satellite s without data processing. Thus, the latency T_k^G can be expressed as,

$$T_k^G = \rho(l-1) + \frac{N_k}{C_{k,s}^l} + \frac{N_k}{D_{k,g}^l} + \frac{N_k}{q_{k,g}^l} + P_{k,g}^l, \tag{2}$$

where $D_{k,g}^l$ is the allocated communication capacity for the link of task k from the satellite to the gateway g. $q_{k,g}^l$ is the computing capacity allocated to the task k by gateway g. $P_{k,g}^l = \frac{2d_{k,s}+2d_{s,g}}{c}$, and $d_{s,g}$ is the distance between satellite s and gateway g. The transmission latency from the gateway to satellite and satellite to terminal are also omitted as analyzed in (1).

With (1) and (2), the latency of task k can be written as,

$$T_k = \begin{cases} T_k^S, & \alpha_k = 1; \\ T_k^G, & \beta_k = 1. \end{cases} \tag{3}$$

In (3), α_k and β_k are the offloading indicators for task k. If $\alpha_k = 1$, task k will be handled at the satellite. Otherwise, the task k will be handled at the gateway with $\beta_k = 1$. In addition, α_k and β_k satisfy $\alpha_k + \beta_k = 1$.

With the latency of every task k, the system latency can be defined as,

$$T_{eff} = \sum_l \sum_s \sum_{k \in \overline{\Omega}_{s,l}} \omega_k T_k. \tag{4}$$

In (4), system latency is defined as the weighted-sum latency of all of tasks. $\overline{\Omega}_{s,l}$ denotes the set of new tasks which are scheduled to be associated with the satellite s at time slot l, and ω_k denotes the weights of each task. According to (1)-(3), the T_{eff} is related to the user association, offloading decision, available resource and resource allocation at each time slot.

B. ENERGY COST MODEL

In SAT-IoT networks, IoT terminals and satellites are both energy-limited nodes. In this paper, IoT terminals are assumed to offload the tasks directly to the satellites or gateways without local task handling for simplicity, and every IoT terminal transmits with a constant power level. Thus, we only need to analyze the energy cost at the satellites.

Suppose the computation capability of the onboard central processing unit (CPU) as ξ cycle/bit, and ξ is mainly determined by the architecture of the CPU. The operating frequency of the CPU is denoted as f_s cycle/s. Thus, the computing capacity of satellite can be denoted as $Z_s = f_s/\xi$ with the unit of bit/s. Meanwhile, the energy cost of CPU is affected by the operating frequency, and can be denoted as κf_s^2 for each circle. Therefore, the energy required for the satellite s at time slot l can be expressed as,

$$E_{s,l} = \sum_{k \in \overline{\Psi}_{s,l}} N_k \xi \kappa f_s^2. \tag{5}$$

In (5), $\overline{\Psi}_{s,l}$ denotes the new tasks scheduled to be handled by the satellite s at time slot l. The energy cost for satellite s is related to the number of bits of all tasks in $\overline{\Psi}_{s,l}$.

C. PROBLEM FORMULATION

According to the latency and energy cost model, tasks handled at the satellite are beneficial to decrease the system latency, but the energy cost of satellite will increase unexpectedly. In this paper, we intend to minimize the weighted-sum of system latency and energy cost. Thus, the problem can be formulated as,

$$\begin{aligned} \min_{\substack{\overline{\Omega}_{s,l}, \overline{\Psi}_{s,l}, \overline{\Phi}_{g,l}, \\ C_{k,s}^l, D_{k,g}^l, z_{k,s}^l, q_{k,g}^l}} & \left(\eta T_{eff} + (1-\eta) \sum_l \sum_s E_{s,l} \right) \\ s.t. \quad & \sum_{k \in \Omega_{s,l}} C_{k,s}^l \le X_s, \\ & \sum_{k \in \Phi_{g,l}} D_{k,g}^l \le Y_g, \\ & \sum_{k \in \Psi_{s,l}} z_{k,s}^l \le Z_s, \\ & \sum_{k \in \Phi_{g,l}} q_{k,g}^l \le Q_g, \end{aligned} \tag{6}$$

where $\Omega_{s,l}$ denotes the tasks associated with satellite s at time slot l, and it includes the set of new associated tasks $\overline{\Omega}_{s,l}$ and ongoing tasks associated with satellite s at time slot l. $\Phi_{g,l}$ denotes the tasks associated with gateway g at time slot l and $\bigcup_g \Phi_{g,l} \subset \bigcup_s \Omega_{s,l}$, because the tasks need to be transferred to the gateways via the satellites. $\overline{\Phi}_{g,l}$ is the set of new tasks associated with gateway g at time slot l. $\Psi_{s,l}$ is the set of new tasks that will be handled by satellite s at

time slot l, and $\Psi_{s,l} \subset \Omega_{s,l}$ because some tasks will only be transferred via satellite s without being handled. X_s and Y_g is the maximum communication capacity available for the links from IoT terminals to the sth satellite and the links from satellites to the gth gateway respectively. Similarly, Z_s and Q_g is the maximum computing capacity for the satellite s and gateway g respectively. Please note that the resources occupied by the ongoing tasks $k \in \Omega_{s,l}/\overline{\Omega}_{s,l}$ cannot be used by the new scheduled tasks at time slot l. η and 1-η denote the weights of latency and energy cost respectively, and $0 \leq \eta \leq 1$.

It can be seen from (6) that the weighted-sum system latency and energy cost is affected by the user association, offloading decision and resource allocation at each time slot. Moreover, the user association, offloading decision and resource allocation at time slot l will affect the states of time slot $l+1$. For example, if all the tasks served at time slot l cannot be finished at the current time slot, there will be no available resources for the unserved tasks at time slot $l+1$. Thus, the problem formulated in (6) can be seen as a dynamic programming problem based on joint user association, offloading and resource allocation. Since there are a series of variables, it is difficult to solve the problem with traditional methods. We will propose a DRL based latency and energy optimization method by decomposing the complex problem into two sub-problems.

III. LATENCY AND ENERGY OPTIMIZATION BASED ON DEEP REINFORCEMENT LEARNING

According to Section II, user association and offloading decision indicators are discrete, while the resource allocation variables are continuous. Thus, the problem formulated in (6) is a dynamic mixed-integer problem, which is a non-convex problem and hard to find the optimal solutions. To tackle this problem, we decompose the problem into two sub-problems to reduce the its complexity. The first sub-problem is optimal computing and communication resource allocation with fixed user association and offloading decision, which will be solved with Lagrange multiplier method. The second sub-problem is the joint user association and offloading decision with optimal resource allocation, which will be solved with DRL based algorithm. The two sub-problems are analyzed in the following subsection A and subsection B, respectively..

A. OPTIMAL RESOURCE ALLOCATION WITH FIXED USER ASSOCIATION AND OFFLOADING DECISION

With fixed user association and offloading decision, the weighted-sum of latency and energy cost at the lth time slot can be defined as,

$$W_l = \sum_s \left(\sum_{k \in \overline{\Omega}_{s,l}} \eta \omega_k T_k + (1-\eta) E_{s,l} \right), \tag{7}$$

where W_l can be seen as the objective function for the lth time slot, and the value of W_l will be affected by the computing and communication resource allocation cooperatively. Let $\widetilde{X}_{s,l}$ be the sum of allocated communication resource for the ongoing tasks associated with sth satellite at lth time slot, let $\widetilde{Y}_{g,l}$ be the sum of allocated communication resource for the ongoing tasks associated with gth gateway at lth time slot. Similarly, $\widetilde{Z}_{s,l}$ and $\widetilde{Q}_{g,l}$ denotes the sum of computing resource allocated to the ongoing tasks handled at the sth satellite and gth gateway respectively. Thus, the minimization of weighted function W_l can be rewritten as,

$$\begin{aligned} \min_{C^l_{k,s}, D^l_{k,g}, z^l_{k,s}, q^l_{k,g}} & \sum_s \left(\sum_{k \in \overline{\Omega}_{s,l}} \eta \omega_k T_k + (1-\eta) E_{s,l} \right) \\ s.t. & \sum_{k \in \overline{\Omega}_{s,l}} C^l_{k,s} \leq X_s - \widetilde{X}_{s,l}, \\ & \sum_{k \in \overline{\Phi}_{g,l}} D^l_{k,g} \leq Y_g - \widetilde{Y}_{g,l}, \\ & \sum_{k \in \overline{\Psi}_{s,l}} z^l_{k,s} \leq Z_s - \widetilde{Z}_{s,l}, \\ & \sum_{k \in \overline{\Phi}_{g,l}} q^l_{k,g} \leq Q_g - \widetilde{Q}_{g,l}. \end{aligned} \tag{8}$$

In (8), the user association and offloading indicators expressed in $\overline{\Omega}_{s,l}$, $\overline{\Phi}_{g,l}$ and $\overline{\Psi}_{s,l}$ are assumed to be known, and the optimization of the problem listed in (8) is only related to the computing and communication resource allocation. With (1)-(3), W_l can be further rewritten as,

$$\begin{aligned} W_l &= \sum_s \left(\sum_{k \in \overline{\Omega}_{s,l}} \eta \omega_k T_k + (1-\eta) E_{s,l} \right) \\ &= \sum_s \sum_{k \in \overline{\Omega}_{s,l}} \overline{\eta} \rho (l-1) + \sum_s \sum_{k \in \overline{\Omega}_{s,l}} \overline{\eta} \frac{N_k}{C^l_{k,s}} \\ &\quad + \sum_s \sum_{k \in \overline{\Psi}_{s,l}} \overline{\eta} \frac{N_k}{z^l_{k,s}} + \sum_s \sum_{k \in \overline{\Phi}_{g,l}} \overline{\eta} \left(\frac{N_k}{D^l_{k,g}} + \frac{N_k}{q^l_{k,g}} \right) \\ &\quad + \sum_s \sum_{k \in \overline{\Psi}_{s,l}} \overline{\eta} P^l_{k,s} + \sum_s \sum_{k \in \overline{\Phi}_{g,l}} \overline{\eta} P^l_{k,g} \\ &\quad + \sum_s (1-\eta) E_{s,l}. \end{aligned} \tag{9}$$

In (9), $\overline{\eta} = \eta \omega_k$. The waiting latency and propagation latency in (9) will not be affected by the computing and communication resource allocation. Besides, $E_{s,l}$ is also determined if $\overline{\Psi}_{s,l}$ is known. Thus, the equivalent W_l with the factors that affect the computing and communication resource in (9) can be redefined as,

$$\widetilde{W}_l = \sum_s \left(\sum_{k \in \overline{\Omega}_{s,l}} \frac{\overline{\eta} N_k}{C^l_{k,s}} + \sum_{k \in \overline{\Psi}_{s,l}} \frac{\overline{\eta} N_k}{z^l_{k,s}} + \sum_{k \in \overline{\Phi}_{s,l}} \left(\frac{\overline{\eta} N_k}{D^l_{k,g}} + \frac{\overline{\eta} N_k}{q^l_{k,g}} \right) \right) \tag{10}$$

With (10), the problem defined in (8) can be rewritten as,

$$\begin{aligned}\min_{C^l_{k,s},D^l_{k,g},z^l_{k,s},q^l_{k,g}} \quad & \widetilde{W}_l \\ s.t. \quad & \sum_{k\in\overline{\Omega}_{s,l}} C^l_{k,s} \le X_s - \widetilde{X}_{s,l}, \\ & \sum_{k\in\overline{\Phi}_{g,l}} D^l_{k,g} \le Y_g - \widetilde{Y}_{g,l}, \\ & \sum_{k\in\overline{\Psi}_{s,l}} z^l_{k,s} \le Z_s - \widetilde{Z}_{s,l}, \\ & \sum_{k\in\overline{\Phi}_{g,l}} q^l_{k,g} \le Q_g - \widetilde{Q}_{g,l}. \end{aligned} \tag{11}$$

The objective function $\widetilde{W}_l$ and constraints in (11) are all convex, so the problem defined in (11) is a convex problem. According to (10) and (11), the optimal resource allocation utilized to minimize $\widetilde{W}_l$ can be achieved via Theorem 1.

Theorem 1: With (10) and (11), the optimal $C^l_{k,s}$, $D^l_{k,g}$, $z^l_{k,s}$ and $q^l_{k,g}$ at time slot l for $k \in \overline{\Omega}_{s,l}$ can be achieved via Karush-Kuhn-Tucker (KKT) conditions as,

$$\begin{aligned} C^l_{k,s} &= \frac{\overline{X}_{s,l}\sqrt{\overline{\eta}N_k}}{\sum_{k\in\overline{\Omega}_{s,l}}\sqrt{\overline{\eta}N_k}}, \\ D^l_{k,g} &= \frac{\overline{Y}_{g,l}\sqrt{\overline{\eta}N_k}}{\sum_{k\in\overline{\Phi}_{g,l}}\sqrt{\overline{\eta}N_k}}, \\ z^l_{k,s} &= \frac{\overline{Z}_{s,l}\sqrt{\overline{\eta}N_k}}{\sum_{k\in\overline{\Psi}_{s,l}}\sqrt{\overline{\eta}N_k}}, \\ q^l_{k,g} &= \frac{\overline{Q}_{g,l}\sqrt{\overline{\eta}N_k}}{\sum_{k\in\overline{\Phi}_{g,l}}\sqrt{\overline{\eta}N_k}}. \end{aligned} \tag{12}$$

Proof: The proof can be found in Appendix.

According to (12), the optimal computing and communication resource allocation can be obtained with fixed user association and offloading decision for each time slot. With the optimal solution for resource allocation, we can optimize the user association and offloading decision further in the next subsection.

B. JOINT USER ASSOCIATION AND OFFLOADING DECISION WITH OPTIMAL RESOURCE ALLOCATION

Although we can obtain the optimal computing and communication resource allocation for every time slot, the optimal resource allocation for several sequential time slots is coupled with each other. Moreover, the joint user association and offloading decision is still a non-convex problem with integer programming problem. Therefore, traditional methods based on optimization theory cannot be applied to tackle this problem directly. To address the problem with affordable complexity, we model it as an MDP problem, and DRL based method is utilized to achieve long-term reward in terms of latency and energy cost.

In general, MDP can be denoted as a tuple ($\mathbb{H}$, $\mathbb{A}$, $\mathbb{P}$, $\mathbb{R}$). $\mathbb{H}$ denotes the state space of the system, $\mathbb{A}$ is the action space, $\mathbb{P}$ is space of the transition probability from state h_l to h_{l+1} with action a_l, and $\mathbb{R}$ is the reward/cost with state h_l and action a_l. Thus, the MDP corresponding to the problem defined in (6) can be expressed as,

(1) **State**($\mathbb{H}$): Since the user association, offloading decision and resource allocation for tasks are made slot by slot, the states are defined for every time slot. Thus, the state at time slot l can be defined as $h_l = \{\widetilde{\Omega}_l, \widetilde{\Psi}_l, \widetilde{\Phi}_l, \widetilde{\mathbf{U}}_l, \widetilde{\mathbf{X}}_l, \widetilde{\mathbf{Y}}_l, \widetilde{\mathbf{Q}}_l, \widetilde{\mathbf{Z}}_l, \mathbf{PL}_l\}$. $\widetilde{\Omega}_l$ is defined as $\{\widetilde{\Omega}_{1,l}, \widetilde{\Omega}_{2,l}, \cdots, \widetilde{\Omega}_{\overline{S},l}\}$, and $\widetilde{\Omega}_{s,l}$ is the set of ongoing tasks associated with satellite s at time slot l. $\widetilde{\Psi}_l$ can be expressed as $\{\widetilde{\Psi}_{1,l}, \widetilde{\Psi}_{2,l}, \cdots, \widetilde{\Psi}_{\overline{S},l}\}$, and $\widetilde{\Psi}_{s,l}$ is the set of ongoing tasks handled by satellite s. $\widetilde{\Phi}_l$ is defined as $\{\widetilde{\Phi}_{1,l}, \widetilde{\Phi}_{2,l}, \cdots, \widetilde{\Phi}_{\overline{G},l}\}$, and $\widetilde{\Phi}_{g,l}$ is the set of ongoing tasks served by gateway g at time slot l. $\widetilde{\mathbf{U}}_l$ is the set of unserved tasks at the time slot l. $\widetilde{\mathbf{X}}_l$ is defined as $\{\widetilde{X}_{1,l}, \widetilde{X}_{2,l}, \cdots, \widetilde{X}_{\overline{S},l}\}$, and it is the communication resources occupied by the ongoing tasks served by satellite s at time slot l. Similarly, $\widetilde{\mathbf{Y}}_l = \{\widetilde{Y}_{1,l}, \widetilde{Y}_{2,l}, \cdots, \widetilde{Y}_{\overline{G},l}\}$, denotes the communication resources occupied by the ongoing task served by the gateways at time slot l. $\widetilde{\mathbf{Q}}_l = \{\widetilde{Q}_{1,l}, \widetilde{Q}_{2,l}, \cdots, \widetilde{Q}_{\overline{G},l}\}$ and $\widetilde{\mathbf{Z}}_l = \{\widetilde{Z}_{1,l}, \widetilde{Z}_{2,l}, \cdots, \widetilde{Z}_{\overline{S},l}\}$ is the computing resource occupied by the ongoing tasks served by satellites and gateways at time slot l respectively. $\mathbf{PL}_l$ denotes the matrix that includes the locations of all IoT terminals, satellites and gateways.

(2) **Action**($\mathbb{A}$): For each time slot l, the unserved tasks $k \in \widetilde{\mathbf{U}}_l$ need to be associated with a satellite or gateway to be handled. Moreover, the computing and communication resource needs to be allocated with the corresponding user association and offloading decision. Define $\mathbf{C}^l = \{C^l_{1,s}, C^l_{2,s}, \cdots, C^l_{K,s}\}$ and $\mathbf{D}^l = \{D^l_{1,s}, D^l_{2,s}, \cdots, D^l_{K,s}\}$ be the sets of communication resources allocated to the tasks scheduled at time slot l. $\mathbf{Z}^l = \{z^l_{1,s}, z^l_{2,s}, \cdots, z^l_{K,s}\}$ and $\mathbf{Q}^l = \{q^l_{1,s}, q^l_{2,s}, \cdots, q^l_{K,s}\}$ are the computing resources allocated to the tasks scheduled at time slot l. Since the resource allocation can be obtained with (12), we only need to define the action space for user association and offloading decision with lower dimensions. Thus, the action at time slot l can be defined as $a_l = \{\overline{A}_{1,l}, \overline{A}_{2,l}, \cdots, \overline{A}_{K,l}\}$. $\overline{A}_{k,l} = \{A_1, A_2, A_3, A_4\}$, in which $A_1 \in \{0, 1\}$ denotes that whether the task k will be scheduled at lth time slot or not, $A_2 \in \{1, 2, \cdots, S\}$ denotes the satellite associated with the task k, $A_3 \in \{0, 1\}$ denotes whether the task k will be handled by its associated satellite or by the gateway, and $A_4 \in \{1, 2, \cdots, G\}$ denotes the gateway utilized to handle the task k. With a specific action a_l, the user association and offloading indicators $\overline{\Omega}_l$, $\overline{\Psi}_l$, and $\overline{\Phi}_l$ can be obtained correspondingly.

(3) **Transition Probability**($\mathbb{P}$): For an MDP, the transition probability from one state to another should be provided with an action a_l. However, it is difficult to get the accurate probability from one state to another for the scenario investigated in this paper. Because some states are continuous variables, and both state space and action space are very large,

it is impossible to get transition probability for all of states h_l and actions a_l. Hence a model-free DRL framework is considered.

(4) **Reward**($\mathbb{R}$): To minimize the weighted-sum of system latency and energy cost, the reward $R(h_l, a_l)$ at time slot l with state h_l and action a_l is defined as,

$$R(h_l, a_l) = \sum_s \left(\sum_{k \in \overline{\Omega}_{s,l}} \left[J - (\eta \omega_k T_k + (1-\eta) E_{s,l}) \right] + \sum_{k \in \widetilde{\mathbf{U}}_l / \left(\sum_s \overline{\Omega}_{s,l}\right)} - (\eta \omega_k T_k) \right) \quad (13)$$

In (13), $R(h_l, a_l)$ will be affected by the state h_l and action a_l. For each time slot l, new scheduled tasks $k \in \overline{\Omega}_{s,l}$ will feedback a reward $J - (\eta \omega_k T_k + (1-\eta) E_{s,l})$, which is negatively correlated with the weighted-sum of latency and energy cost. J is a constant value that makes the reward positive. And the tasks $k \in \widetilde{\mathbf{U}}_l / \left(\sum_s \overline{\Omega}_{s,l}\right)$ will feedback $-(\eta \omega_k T_k)$ as a reward, in which T_k only consists of waiting latency and $E_{s,l}$ is zero since the tasks have not been scheduled.

Suppose π to be the policy of the action, value function $V(h|\pi)$ is defined to investigate the long-term performance of the policy π. The value function is defined as,

$$V(h|\pi) = \mathbb{E}\left[\sum_l \gamma^l R(h_l, a_l) | h_0 = h, \pi \right], \quad (14)$$

where γ denotes the discount factor, and the value function can be seen as a modified expectation of weighted-sum of system latency and energy cost defined in (6) with $\gamma = 1$. Therefore, the problem defined in (6) is equivalent to find an optimal policy π^* to maximize the value function of each state as,

$$\pi^*(h) = \arg\max_a \left[R(h, a) + \sum_{\bar{h}} \gamma V(\bar{h}|\pi^*) \right]. \quad (15)$$

In (15), state $\bar{h}$ can be achieved with action a and state h. To find the optimal policy π^*, Q-learning can be utilized [31]. However, in the scenario presented in this paper, the state space and action space will increase exponentially with the number of terminals. Hence it is extremely difficult for Q-learning to establish the Q-table. Function approximation method is able to handle the curse of dimensionality [32]. By combining Q-learning and neural network, which turns Q-learning's Q-table into Q-Network, DQN can avoid the capacity limitation of Q-table [33].

DQN introduces the target network on the basis of the main network, which has the same structure as the main network. The main network is updated every iteration, while the target work is updated at regular intervals by copying the parameters of the main network. The target network is utilized to obtain the target Q-value $Q^*(h, a)$ which is defined as

$$Q^*(h, a) = R(h, a) + \gamma \max_{\bar{a}} Q^*(\bar{h}, \bar{a}). \quad (16)$$

$Q^*(h, a)$ is the maximum expected reward utilized to evaluate the value of the action selected in a specific state. And $Q^*(h, a)$ is the target that approximated Q-function $Q(h, a; \theta)$ will approach through training.

The Loss-Function is defined as,

$$L(\theta) = \mathbb{E}\left[(Q^*(h, a) - Q(h, a; \theta))^2 \right], \quad (17)$$

in which θ is the weight of network and will be trained to convergence with the goal of minimizing $L(\theta)$. Meanwhile, Q-function $Q(h, a; \theta)$ gradually progresses to $Q^*(h, a)$. Besides of function approximation method, experience replay (ER) is also used in DQN to overcome the problem of correlation and non-stationary distribution of empirical data for training [34].

By adopting the DQN, the proposed JUAOD-ORA for MEC enhanced SAT-IoT is shown in Algorithm 1, where G denotes the maximum of training episode, and ζ denotes the experience replay buffer.

Algorithm 1 Joint User Association and Offloading Decision With Optimal Resource Allocation Based on DQN

Require:
IoT terminal information: N_k, $\widetilde{\mathbf{U}}_l$
Satellite information: $\widetilde{\Omega}_l$, $\widetilde{\Psi}_l$, $\widetilde{\mathbf{X}}_l$, $\widetilde{\mathbf{Q}}_l$
Gateway information: $\widetilde{\Phi}_l$, $\widetilde{\mathbf{Y}}_l$, $\widetilde{\mathbf{Z}}_l$
Nodes location: $\mathbf{PL}_l$
Ensure:
Offloading decision: $\overline{\Omega}_l$, $\overline{\Psi}_l$, $\overline{\Phi}_l$
Resource allocation: $\mathbf{C}^l$, $\mathbf{D}^l$, $\mathbf{Z}^l$, $\mathbf{Q}^l$
Initialize network with γ, ε and ζ.
Observe.
while *episode* $< G$ **do**
  reset state h_l and time slot l.
  while $\widetilde{\mathbf{U}}_l \neq \varnothing$ **do**
    Select an action a_l according to ε-greedy policy.
    Allocate resource according to eq.(12).
    Calculate reward $R(h_l, a_l)$.
    Update next state h_{l+1}.
    Save $(h_l, a_l, R(h_l, a_l), h_{l+1})$, and update ζ.
    Update θ.
    l ++.
  end while
  episode ++.
end while

ε-greedy policy is utilized to balance the exploration and utilization of models [33]. At each time slot l, an action will be selected randomly with probability ε, otherwise the action will be selected according to Q-function with probability 1-ε, and ε will decay from $\varepsilon_{initial}$ to ε_{final} through P steps.

Before training, agent will observe for O steps to accumulate experience. At each time slot l, agent takes system state h_l as input to the neural network, and selects an action according to ε-greedy policy. Once we get the offloading decision through the selected action a_l, computing and communication

TABLE 1. Simulation parameters.

Parameter	Value
Scenario parameters	
Task size N_k	8×10^4-1.2×10^5 bit
The weights of each task ω_k	1
Height of LEO satellites	1000 km
Coverage radius of LEO satellites	1000 km
CPU compress cycle ξ	1000 cycles/bit
Energy cost of CPU κf_s^2	3×10^{-10} J/cycle
Satellite communication capacity X_s	10 Mbps
Gateway communication capacity Y_g	50 Mbps
Satellite computing capacity Z_s	10^{10} cycles/s
Gateway computing capacity Q_g	5×10^{10} cycles/s
DQN parameters	
Maximum training episode G	50000
Size of replay buffer ζ	20000
Observation size O	5000
Discount factor γ	0.95
Completion reward J	20
Learning rate	0.001
$\varepsilon_{initial}$	1.0
ε_{final}	0.001
P	20000

resource can be allocated according to (12). With the offloading decision and resource allocation results, the latency and energy cost of the system at this slot can be calculated, and reward $R(h_l, a_l)$ can be obtained according to (13). Then the state will be updated to h_{l+1} and stored into the replay buffer with h_l, a_l and $R(h_l, a_l)$. Partial samples of the replay buffer will be randomly extracted for training at each step to disrupt the correlation of data, and old samples will be overwritten with new ones when the buffer is full. State will be reset if all tasks are served after several steps, and training will be stopped when *episode* reaches the maximum episode G.

IV. SIMULATION RESULTS

In this section, we evaluate the performance of the proposed JUAOD-ORA. Simulation parameters are listed in Table. 1. In the simulation, MEC enhanced SAT-IoT system with two LEO satellites and two gateways is considered. Each terminal and gateway is randomly distributed within the coverage of all satellites. The height and coverage radius of satellites are set to 1000 km. Referring to [35] and [36], we set the satellite communication capacity X_s and gateway communication capacity Y_g to 100 Mbps and 500 Mbps respectively. As for CPU configurations, we set the compress cycle ξ of CPU to 1000 cycles/bit [37], and the energy cost of CPU κf_s^2 is set to 3×10^{-10} J/cycle [38].The satellite computing capacity Z_s and gateway computing capacity Q_g are respectively set to 10^{10} cycles/s and 5×10^{10} cycles/s, scilicet 10^7 bits/s and 5×10^7 bits/s. In addition, DQN parameters are shown in the Table. 1.

A. CONVERGENCE OF JUAOD-ORA

In this subsection, we present the training results of the proposed JUAOD-ORA based on DQN. Fig. 2 shows the convergence process of the $L(\theta)$ defined in (17) in the training process. It can be seen that $L(\theta)$ converges close to 0 after about 1.1×10^4 steps, which indicates that the approximated Q-function $Q(h, a; \theta)$ has almost approached the target action-value function $Q^*(h, a)$. Please note that the action space in this paper is very large, so the number of steps required for convergence is large. However, the complexity of the proposed method remains low, because the large amount of steps are only needed for training phases, and the optimal offloading decision and resource allocation can be achieved based on the trained Q-network without iterations. The convergence process of total cost, as discussed in (7) and calculated by the weighted-sum of latency and energy, is showed in Fig. 3. The total cost converges almost synchronously with $L(\theta)$ at about 0.8×10^4-th episode. In general, the convergence of the proposed algorithm is fast, which enables the algorithm to cope with the dynamic environment.

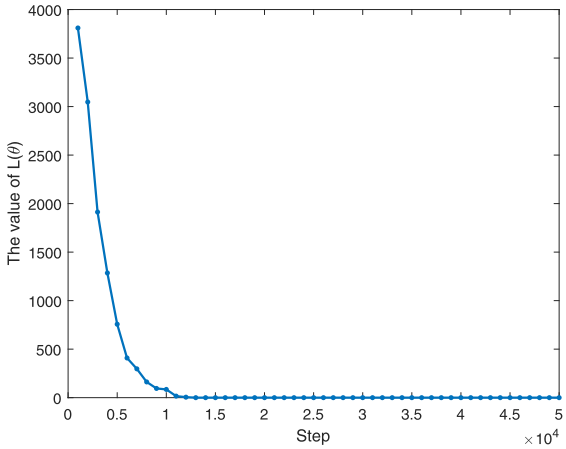

FIGURE 2. Convergence process of loss.

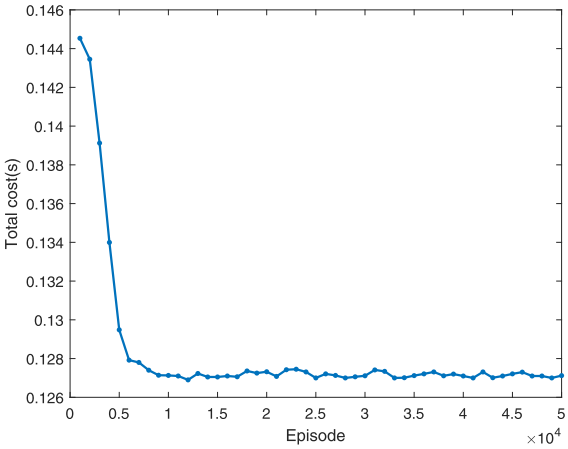

FIGURE 3. Convergence process of total cost.

B. PERFORMANCE ANALYSIS

In this subsection, we compare the proposed JUAOD-ORA algorithm with 'Random' method and heuristic method based on simulated annealing(SA), which is proposed in [39] to solve task offloading problem in MEC system, and we

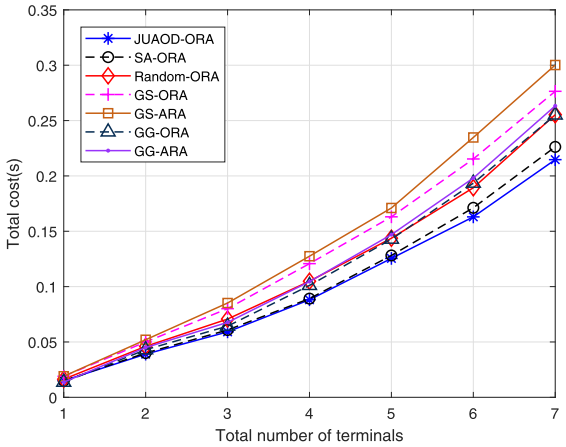

FIGURE 4. Total cost v.s. total number of terminals.

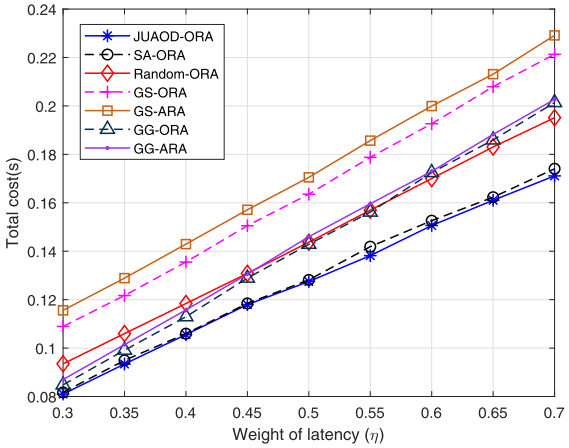

FIGURE 5. Total cost v.s. η.

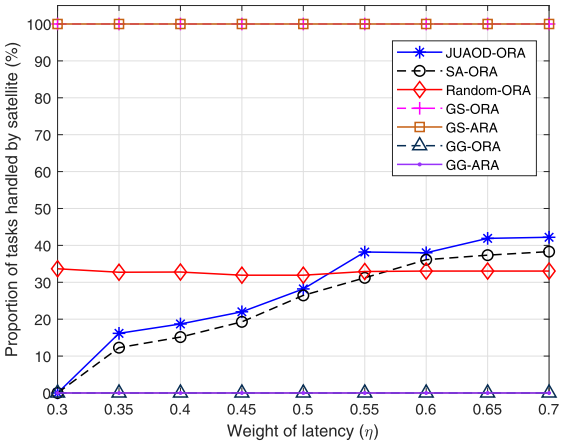

FIGURE 6. Proportion of tasks handled by satellites v.s. η.

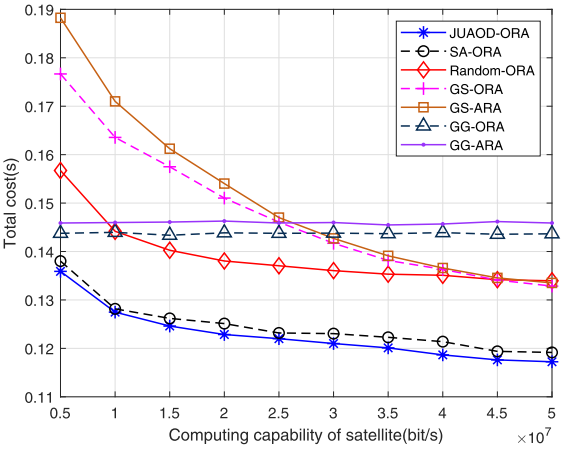

FIGURE 7. Total cost v.s. satellite computing capability.

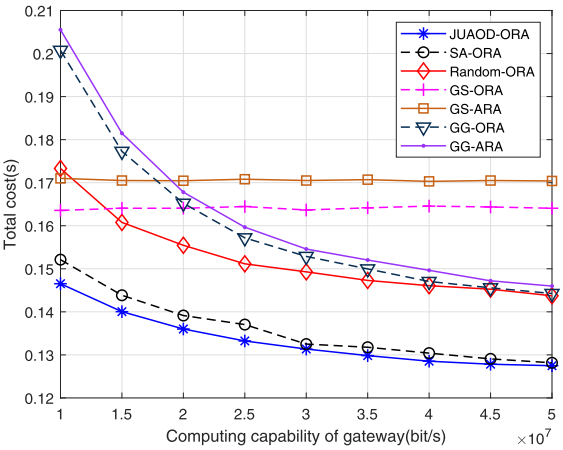

FIGURE 8. Total cost v.s. gateway computing capability.

simplify the SA algorithm to make it applicable to our MEC enhanced SAT-IoT scenario for comparison. To properly evaluate the performance of the JUAOD-ORA, Theorem 1 will be utilized to allocate resources when the offloading decision are made by reference schemes. The detailed description is as follows:

1) Random: Tasks will be randomly offloaded to satellite or gateway, and the resources will be allocated according to optimal resource allocation (ORA) proposed in Theorem 1. In Fig. 4-Fig.8, this method is labeled as Random-ORA.
2) SA: The offloading decision will be obtained through SA algorithm designed in [39]. Meanwhile, the resources will be allocated by ORA proposed in Theorem 1. In Fig. 4-Fig.8, this method is labeled as SA-ORA.

In addition, to evaluate the efficiency of the ORA, we design two other 'Greedy' offloading algorithms called 'greedy on satellite'(GS) and 'greedy on gateway'(GG), which will offload all tasks to satellite and gateway, respectively. ORA and average resource allocation (ARA), which will allocate communication resource and computing resource evenly, will be utilized respectively with two 'Greedy' methods. Every point in the simulation results is obtained by taking the average value over 5000 tests.

Fig. 4 shows the total cost performance of the proposed JUAOD-ORA as the number of terminals increases. We can find that both GS and GG with the ORA proposed in

Theorem 1 perform better than the ARA. When the number of terminals is 1, the total cost of the ORA and the ARA are equal, because the resource allocated by different methods is the same with each other for this case. Nevertheless, as the number of terminals increases, the performance of the ORA will gradually outperform the ARA, which indicates the higher efficiency of our proposed ORA.

In addition, we can see that the increment of number of terminals leads to the increase of total cost, because the limited resources are shared by more tasks. SA algorithm performs close to JUAOD-ORA algorithm when the number of terminals is small. However, when the number of terminals is large, the performance of SA algorithm will gradually deteriorate, because the action set grows exponentially with the number of terminals, and SA algorithm is more likely to fall into a local optimal solution. Moreover, parameters of SA algorithm (e.g., start temperature, the number of iterations, annealing rate) will directly affect the performance of the algorithm, and should be updated when the environment parameters change.

Fig. 5 and Fig. 6 depict the influence of weight η on the total cost and the proportion of tasks handled by satellites, respectively. For 'Greedy' methods, including GG and GS, the performance of the ORA is better than the ARA. The simulation curves are increasing linearly in Fig. 5, because η and total cost has a linear relationship. In Fig. 6, the proportion of tasks handled by satellite is calculated by dividing the number of tasks processed on the satellite by the total number of tasks. For 'Random' method and 'Greedy' methods, weight η has no effect on task offloading decision, so the proportion of tasks handled by satellite basically unchanged as η increases. However, for SA algorithm and the proposed JUAOD-ORA, η will affect the scheduling decision. As η increases, the weight of energy cost decreases, and it is better to offload more tasks to satellites with large computing capabilities. As shown in Fig. 6, the proportion of tasks handled by satellite with JUAOD-ORA increases as η becomes larger, which demonstrates that it can make offloading decision effectively.

In Fig. 7, satellite computing capability is adopted as variable to investigate the performance of the proposed algorithm. Obviously, the performance of the ORA is better than the ARA as analyzed above, and the curves of GS-ORA and GS-ARA are getting closer as the satellite computing capability increases. This is because the computing resource of satellite is beneficial to the reduction of computing and propagation latency, and the latency gap between different methods will also become smaller. Employing both GG-ORA and GG-ARA, the total cost is stable with little change, since all tasks are offloaded to gateway and their performance are not affected by the satellite computing capability. In addition, Fig. 7 also clearly shows the performance of the proposed JUAOD-ORA is better than SA algorithm.

Similarly, total cost with respect to gateway computing capability is shown in Fig. 8. For 'Greedy' methods, the ORA also performs better than the ARA. The curves of GS-ORA and GS-ARA are basically unchanged, since the increase of the gateway computing capability has no effect on GS methods. For other methods, the increasing of gateway computing capability leads to the decrease of total cost, because tasks can be assigned with more resources. Simulation results illustrate that the propose JUAOD-ORA algorithm can also achieve better performance than the reference algorithms.

To summarize, the proposed ORA method performs better than the ARA. Compared with SA algorithm, the proposed JUAOD-ORA algorithm performs better with lower complexity, and it can adapt to dynamic environments of SAT-IoT networks.

V. CONCLUSION

In this paper, we present an MEC enhanced SAT-IoT networks for IoT services in remote areas. To jointly optimize the weighted latency and energy cost of the system, we formulate the optimization problem and decompose it into two subproblems. Then, an optimal algorithm based on Lagrange multiplier method is proposed to solve the resource allocation sub-problem with fixed user association and offloading decision. Furthermore, a joint user association and offloading decision with optimal resource allocation algorithm based on deep reinforcement learning is proposed to solve the offloading decision sub-problem. Simulation results illustrate that the proposed algorithm can reduce the latency and energy cost effectively.

APPENDIX
PROOF OF THE THEOREM 1

With (10) and (11), the Lagrangian function $\mathcal{L}$ of $\widetilde{W}_l$ can be expressed as,

$$\mathcal{L} = \widetilde{W}_l + \mu\left(\sum_{k\in\overline{\Omega}_{s,l}} C^l_{k,s} - \overline{X}_{s,l}\right) + \theta\left(\sum_{k\in\overline{\Phi}_{g,l}} D^l_{k,g} - \overline{Y}_{g,l}\right) + \nu\left(\sum_{k\in\overline{\Psi}_{s,l}} z^l_{k,s} - \overline{Z}_{s,l}\right) + \sigma\left(\sum_{k\in\overline{\Phi}_{g,l}} q^l_{k,g} - \overline{Q}_{g,l}\right), \tag{18}$$

where $\mu \geq 0$, $\theta \geq 0$, $\nu \geq 0$, and $\sigma \geq 0$ are Lagrange multipliers. $\overline{X}_{s,l}$, $\overline{Y}_{g,l}$, $\overline{Z}_{s,l}$ and $\overline{Q}_{g,l}$ are the available resources for the new scheduled tasks to be handled at time slot l, and they can be expressed as follows,

$$\begin{aligned} \overline{X}_{s,l} &= X_s - \widetilde{X}_{s,l}, \\ \overline{Y}_{g,l} &= Y_g - \widetilde{Y}_{g,l}, \\ \overline{Z}_{s,l} &= Z_s - \widetilde{Z}_{s,l}, \\ \overline{Q}_{g,l} &= Q_s - \widetilde{Q}_{s,l}. \end{aligned} \tag{19}$$

For simplicity, $C^l_{k,s}$ is taken as an example, same method can be used to get $D^l_{k,g}$, $z^l_{k,s}$ and $q^l_{k,g}$. Thus, the partial

differential of $\frac{\partial \mathcal{L}}{\partial C_{k,s}^{l}}$ can be written as,

$$\frac{\partial \mathcal{L}}{\partial C_{k,s}^{l}} = -\frac{\overline{\eta} N_k}{\left(C_{k,s}^{l}\right)^2} + \mu. \tag{20}$$

According to KKT conditions, we can get the $C_{k,s}^{l}$ with $\frac{\partial \mathcal{L}}{\partial C_{k,s}^{l}} = 0$ as,

$$C_{k,s}^{l} = \sqrt{\frac{\overline{\eta} N_k}{\mu}}. \tag{21}$$

Moreover, μ can be achieved with the KKT conditions $\mu\left(\sum_{k\in\overline{\Omega}_{s,l}} C_{k,s}^{l} - \overline{X}_{s,l}\right) = 0$. Thus, the $C_{k,s}^{l}$ can be rewritten as,

$$C_{k,s}^{l} = \frac{\overline{X}_{s,l}\sqrt{\overline{\eta} N_k}}{\sum_{k\in\overline{\Omega}_{s,l}} \sqrt{\overline{\eta} N_k}}. \tag{22}$$

Therefore, *Theorem 1* can be proven.

REFERENCES

[1] L. Chettri and R. Bera, "A comprehensive survey on Internet of Things (IoT) toward 5G wireless systems," *IEEE Internet Things J.*, vol. 7, no. 1, pp. 16–32, Jan. 2020.

[2] Z. Zhou, J. Feng, L. Tan, Y. He, and J. Gong, "An air-ground integration approach for mobile edge computing in IoT," *IEEE Commun. Mag.*, vol. 56, no. 8, pp. 40–47, Aug. 2018.

[3] L. Matti and L. Kari, "Key drivers and research challenges for 6G ubiquitous wireless intelligence," 6G Flagship, Oulu, Finland, White Paper, Sep. 2019. [Online]. Available: http://jultika.oulu.fi/Record/isbn978-952-62-2354-4

[4] S. Cioni, R. De Gaudenzi, O. Del Rio Herrero, and N. Girault, "On the satellite role in the era of 5G massive machine type communications," *IEEE Netw.*, vol. 32, no. 5, pp. 54–61, Sep. 2018.

[5] M. De Sanctis, E. Cianca, G. Araniti, I. Bisio, and R. Prasad, "Satellite communications supporting Internet of remote things," *IEEE Internet Things J.*, vol. 3, no. 1, pp. 113–123, Feb. 2016.

[6] W.-C. Chien, C.-F. Lai, M. S. Hossain, and G. Muhammad, "Heterogeneous space and terrestrial integrated networks for IoT: Architecture and challenges," *IEEE Netw.*, vol. 33, no. 1, pp. 15–21, Jan. 2019.

[7] M. Bacco, L. Boero, P. Cassara, M. Colucci, A. Gotta, M. Marchese, and F. Patrone, "IoT applications and services in space information networks," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 31–37, Apr. 2019.

[8] I. F. Akyildiz and A. Kak, "The Internet of space things/CubeSats," *IEEE Netw.*, vol. 33, no. 5, pp. 212–218, Sep. 2019.

[9] Z. Qu, G. Zhang, H. Cao, and J. Xie, "LEO satellite constellation for Internet of Things," *IEEE Access*, vol. 5, pp. 18391–18401, 2017.

[10] D. Palma and R. Birkeland, "Enabling the Internet of arctic things with freely-drifting small-satellite swarms," *IEEE Access*, vol. 6, pp. 71435–71443, 2018.

[11] Z. Zhang, W. Zhang, and F.-H. Tseng, "Satellite mobile edge computing: Improving QoS of high-speed satellite-terrestrial networks using edge computing techniques," *IEEE Netw.*, vol. 33, no. 1, pp. 70–76, Jan. 2019.

[12] B. Deng, C. Jiang, H. Yao, S. Guo, and S. Zhao, "The next generation heterogeneous satellite communication networks: Integration of resource management and deep reinforcement learning," *IEEE Wireless Commun.*, early access, Nov. 22, 2019, doi: 10.1109/MWC.001.1900178.

[13] G. Cocco, T. de Cola, M. Angelone, Z. Katona, and S. Erl, "Radio resource management optimization of flexible satellite payloads for DVB-S2 systems," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 266–280, Jun. 2018.

[14] X. Liu, X. Zhai, W. Lu, and C. Wu, "QoS-guarantee resource allocation for multibeam satellite industrial Internet of Things with NOMA," *IEEE Trans Ind. Informat.*, early access, Nov. 26, 2019, doi: 10.1109/TII.2019.2951728.

[15] Y. Zhang, L. Yin, C. Jiang, and Y. Qian, "Joint beamforming design and resource allocation for terrestrial-satellite cooperation system," *IEEE Trans. Commun.*, vol. 68, no. 2, pp. 778–791, Feb. 2020.

[16] Y. Sun, Y. Wang, J. Jiao, S. Wu, and Q. Zhang, "Deep learning-based long-term power allocation scheme for NOMA downlink system in S-IoT," *IEEE Access*, vol. 7, pp. 86288–86296, 2019.

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

[18] N. Hassan, S. Gillani, E. Ahmed, I. Yaqoob, and M. Imran, "The role of edge computing in Internet of Things," *IEEE Commun. Mag.*, vol. 56, no. 11, pp. 110–115, Nov. 2018.

[19] H. Liao, Z. Zhou, X. Zhao, L. Zhang, S. Mumtaz, A. Jolfaei, S. H. Ahmed, and A. K. Bashir, "Learning-based context-aware resource allocation for edge computing-empowered industrial IoT," *IEEE Internet Things J.*, early access, Dec. 31, 2019, doi: 10.1109/JIOT.2019.2963371.

[20] L. Cui, C. Xu, S. Yang, J. Z. Huang, J. Li, X. Wang, Z. Ming, and N. Lu, "Joint optimization of energy consumption and latency in mobile edge computing for Internet of Things," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4791–4803, Jun. 2019.

[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 Things J.*, vol. 6, no. 3, pp. 4188–4200, Jun. 2019.

[22] Z. Zhou, H. Liao, X. Zhao, B. Ai, and M. Guizani, "Reliable task offloading for vehicular fog computing under information asymmetry and information uncertainty," *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 8322–8335, Sep. 2019.

[23] Z. Ning, P. Dong, X. Wang, L. Guo, J. J. P. C. Rodrigues, X. Kong, J. Huang, and R. Y. K. Kwok, "Deep reinforcement learning for intelligent Internet of vehicles: An energy-efficient computational offloading scheme," *IEEE Trans. Cognit. Commun. Netw.*, vol. 5, no. 4, pp. 1060–1072, Dec. 2019.

[24] S. Li, N. Zhang, S. Lin, L. Kong, A. Katangur, M. K. Khan, M. Ni, and G. Zhu, "Joint admission control and resource allocation in edge computing for Internet of Things," *IEEE Netw.*, vol. 32, no. 1, pp. 72–79, Jan. 2018.

[25] Y. Dai, D. Xu, S. Maharjan, and Y. Zhang, "Joint computation offloading and user association in multi-task mobile edge computing," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 12313–12325, Dec. 2018.

[26] M. Sheng, D. Zhou, R. Liu, Y. Wang, and J. Li, "Resource mobility in space information networks: Opportunities, challenges, and approaches," *IEEE Netw.*, vol. 33, no. 1, pp. 128–135, Jan. 2019.

[27] H. Huang, S. Guo, W. Liang, K. Wang, and A. Y. Zomaya, "Green data-collection from geo-distributed IoT networks through low-earth-orbit satellites," *IEEE Trans. Green Commun. Netw.*, vol. 3, no. 3, pp. 806–816, Sep. 2019.

[28] S. Fu, J. Gao, and L. Zhao, "Integrated resource management for terrestrial-satellite systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 3256–3266, Mar. 2020.

[29] Y. Wang, J. Yang, X. Guo, and Z. Qu, "A game-theoretic approach to computation offloading in satellite edge computing," *IEEE Access*, vol. 8, pp. 12510–12520, 2020.

[30] X. Cheng, F. Lyu, W. Quan, C. Zhou, H. He, W. Shi, and X. Shen, "Space/aerial-assisted computing offloading for IoT applications: A learning-based approach," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 5, pp. 1117–1129, May 2019.

[31] H. van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-learning," in *Proc. 30th AAAI Conf. Artif. Intell.*, Phoenix, AZ, USA, 2016, pp. 2094–2100.

[32] I. Osband, C. Blundell, A. Pritzel, and B. Van Roy, "Deep exploration via bootstrapped DQN," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 4026–4034.

[33] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015.

[34] L. Lv, S. Zhang, D. Ding, and Y. Wang, "Path planning via an improved DQN-based learning policy," *IEEE Access*, vol. 7, pp. 67319–67330, 2019.

[35] Y. Su, Y. Liu, Y. Zhou, J. Yuan, H. Cao, and J. Shi, "Broadband LEO satellite communications: Architectures and key technologies," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 55–61, Apr. 2019.

[36] C. Gavrila, M. Alexandru, V. Popescu, C. Sacchi, and D. Giusto, "Satellite SDR gateway for M2M and IoT applications," in *Proc. IEEE Aerosp. Conf.*, Big Sky, MT, USA, Mar. 2019, pp. 1–9.

[37] Y. Wang, J. Zhang, X. Zhang, P. Wang, and L. Liu, "A computation offloading strategy in satellite terrestrial networks with double edge computing," in *Proc. IEEE Int. Conf. Commun. Syst. (ICCS)*, Chengdu, China, Dec. 2018, pp. 450–455.
[38] E. Benkhelifa, T. Welsh, L. Tawalbeh, Y. Jararweh, and A. Basalamah, "User profiling for energy optimisation in mobile cloud computing," *Procedia Comput. Sci.*, vol. 52, pp. 1159–1165, Jan. 2015.
[39] W. Ni, H. Tian, S. Fan, and B. Liu, "Revenue-maximized offloading decision and fine-grained resource allocation in edge network," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Marrakesh, Morocco, Apr. 2019, pp. 1–6.

GAOFENG CUI (Member, IEEE) received the Ph.D. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2013. From 2019 to 2020, he was a Visiting Scholar at the University of California at Riverside. He is currently an Associate Professor at the Information and Electronics Technology Laboratory, School of Electronic Engineering, Beijing University of Posts and Telecommunications. His research interests include satellite communication, 6G mobile communication, radio resource management in telecommunication systems, the Internet of Things, and the Internet of Vehicles.

XIAOYAO LI received the B.E. degree in electronic engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2018, where he is currently pursuing the M.S. degree with the Information and Electronics Technology Laboratory. His research focuses on satellite communication.

LEXI XU (Member, IEEE) received the M.S. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2009, and the Ph.D. degree from the Queen Mary University of London, London, U.K., in 2013. He is currently a Senior Engineer at the China Unicom Network Technology Research Institute. He is also a China Unicom delegate in ITU, ETSI, and CCSA. His research interests include big data, self-organizing networks, satellite networks, and radio resource management in telecommunication systems.

WEIDONG WANG (Member, IEEE) received the Ph.D. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2002. He is currently a Professor at the School of Electronic Engineering, Beijing University of Posts and Telecommunications. His researches focus on satellite communication, radio resource management in telecommunication systems, the Internet of Things, intelligent transportation, cognitive radio, and mobile terminals. He is also a Senior Member of the China Association of Communication.