



Federated deep reinforcement learning-based computation offloading in a LEO satellite edge computing system*

Min JIA[‡], Jian WU, Xinyu WANG, Qing GUO

School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150006, China

E-mail: jiamin@hit.edu.cn; 20B905014@stu.hit.edu.cn; wang_xinyu@hit.edu.cn; qguo@hit.edu.cn

Received May 28, 2024; Revision accepted Oct. 15, 2024; Crosschecked

Abstract: Recent studies have shown that system capacity is very important for cellular networks. In this paper, we consider maximizing the weighted sum-rate of the cellular network downlink and uplink, where each cell consists of a full-duplex (FD) base station (BS) and half-duplex (HD) users. Federated learning (FL) can train models in the absence of centralized data, which can achieve privacy protection of user data. A low earth orbit (LEO) satellite edge computing system (LSECS) can be formed by placing the mobile edge computing (MEC) servers on LEO satellites, which greatly increases the processing capacities of the satellites. Therefore, this paper considers a combination of FL and MEC and proposes an FL-based computation offloading algorithm to maximize the weighted sum-rate while ensuring the security of user data. We consider solving the sub-channel assignment and power allocation problems using deep reinforcement learning (DRL) algorithms with excellent global search capabilities. The simulation results show that our proposed algorithm achieves the maximum weighted sum-rate compared with the baseline algorithms and excellent convergence.

Key words: Federated learning; Low earth orbit satellite; Mobile edge computing; Deep reinforcement learning; Computation offloading

<https://doi.org/10.1631/FITEE.2400448>

CLC number:

1 Introduction

In wireless communication, wireless full-duplex (FD) can significantly improve spectral efficiency, which has attracted great attention from academia and industry (He et al., 2023; Sun et al., 2023). Teklu et al. (2024) analyzed the FD-assisted multi-user multiple-input, multiple-output (MIMO) system and proposed a signal-to-leakage-plus-noise ratio precoding scheme to combat interference and significantly improve energy efficiency. To address the problem of self-interference (SI) in integrated access and backhaul networks based on FD communication operating in the millimeter wave band, Yu et al. (2023)

proposed SI cancellation techniques for spatial, radio frequency (RF), and digital domains, which significantly improve cell throughput gain. To solve the resource allocation problems in multi-cell FD networks, Fawaz et al. (2023) proposed an optimal queue-aware joint scheduling and power allocation algorithm, which can significantly improve user device throughput. To support in-band FD, Alkhrijah et al. (2023) designed a control frame exchange protocol and proposed a site-selection method that effectively improves spectral efficiency and significantly reduces packet delay. However, these researchers have not considered the problems of user data privacy and secure transmission.

Federated learning (FL) has emerged as a distributed machine learning paradigm that can effectively ensure the security of data transmission (Kamal et al., 2023) by sending local model parameters

[‡] Corresponding author

* Project supported by the National Natural Science Foundation of China (No. 62231012) and the Fundamental Research Funds for the Central Universities (No. 2022FRFK060008)

ORCID: Jian WU, <http://orcid.org/0009-0003-8460-1064>

Zhejiang University Press 2024

to FL servers for global aggregation.

In the traditional cloud computing paradigm, FL servers are placed in remote cloud computing centers that receive massive amounts of data for centralized processing, which applies heavy pressure on the center (Lv and Xiu, 2020) and results in significant communication delay and user privacy problems.

To ensure the privacy security of each base station (BS), we do not place servers for global aggregation on the BSs. At the same time, to ensure that BSs in different scenarios can obtain the allocation strategies of other BSs, we consider applying the idea of mobile edge computing (MEC) to the low earth orbit (LEO) satellite network (Chen XM et al., 2023) with excellent system robustness to form a LEO satellite edge computing system (LSECS) (Jia et al., 2023; Gao et al., 2024; Wang Q et al., 2024) and provide global aggregation for BSs. Faced with scenarios where FL needs to be performed across multiple geographically separated remote clusters or devices in remote areas that lack communication infrastructure (e.g., rural regions, maritime areas), existing FL techniques primarily use ground networks, and devices in the above scenarios cannot aggregate local model parameters without the help of non-terrestrial networks. Therefore, the above problems can be solved by using the advantages of LEO satellites in FL (Han et al., 2024). In addition, LEO satellites typically have low orbital altitudes and can reach speeds of up to 7.8 km/s, providing fast communication (e.g., the propagation delay from the ground BSs to the LEO satellites is about 5ms). This feature also enables LEO satellites to provide highly flexible and fast FL services.

Recently, FL has been widely applied to LEO satellite networks (Chen H et al., 2022). To ensure data against malicious intrusions, Uddin and Kumar (2023) proposed a distributed approach based on FL in a software-defined networking (SDN) environment, and designed an SDN backbone network equipped with a traffic regulator to ensure secure data transmission between devices in the satellite-IoT framework. To address the problem that massive satellites collaboratively train machine learning models without sharing local datasets in some scenarios, Razmi et al. (2022) proposed an FL algorithm based on FedAvg, which significantly improves the test accuracy. To achieve robust and reliable connectivity in satellite communications (Satcoms),

Kang et al. (2024) proposed a Satcom-based FL model framework to continuously collect and aggregate FL model parameters transmitted by terrestrial mobile devices, optimizing system energy consumption and performance. Aiming at cyberattacks on Satcom systems, Salim et al. (2024) proposed a comprehensive threat detection model based on deep FL (DFL), which uses a decentralized data-level preprocessing (DLP) mechanism to significantly improve accuracy. For the space-air-ground integration network (SAGIN), Xu HT et al. (2023) proposed a suitable collaborative FL architecture and proposed an abnormal traffic detection method to solve the problem of artificial feature engineering. In view of the highly dynamic SAGIN heterogeneous structure, to adapt to unbalanced resources and varying environments, Tang et al. (2023) proposed a traffic offloading method based on federated reinforcement learning (RL), which significantly reduces the packet drop rate. User privacy and data security problems need to be considered, but the efficient management of resources is also an important problem to be solved.

For resource allocation problems that require high-dimensional state and action spaces, traditional optimization algorithms are difficult to solve efficiently. As a data-driven intelligent algorithm with powerful global search ability, deep reinforcement learning (DRL) has attracted wide attention (Dai et al., 2022; Jia et al., 2024; Wu et al., 2024) by letting the agents interact with the environment iteratively to learn the optimal policy. Aiming at the problems of multi-slot and multi-user resource allocation in downlink cellular networks, Zhao et al. (2024) considered the dynamics of the environment and co-channel interference and proposed a transformer-based DRL algorithm that significantly improves the spectrum efficiency and user fairness. To meet the quality of service (QoS) requirements such as reliability, delay, and transmission rate of the internet of controllable things, Xiao et al. (2023) proposed a resource allocation algorithm based on a decentralized Markov chain, which significantly improves the energy efficiency. Aiming at the problem of maximizing sum-rate and fairness under the resource and power constraints of BSs and device-to-device (D2D) pairs in cellular networks, Vishnoi et al. (2023) proposed a resource allocation algorithm based on centralized DRL, which significantly alleviates co-channel interference and improves the sum-rate and fairness.

Aiming at the energy efficiency requirements of cellular networks, Tran et al. (2023) considered jointly the sub-channel selection and power allocation problems and proposed a low-complexity resource allocation algorithm based on multi-agent DRL, which significantly saves signaling overhead and improves energy efficiency.

Therefore, aiming at the sub-channel allocation and power allocation problems of uplinks and downlinks in cellular networks, which require high-dimensional state and action spaces, this paper considers the use of a DRL algorithm to solve these problems and maximize the weighted sum-rate of downlinks and uplinks. At the same time, to ensure that BSs in different scenarios can obtain the allocation strategies of other BSs and ensure their own privacy security, this paper considers using FL to let each BS maintain a DRL algorithm, and upload the gradient information of all BSs to the LSECS for global aggregation at intervals. After global aggregation, the LEO satellite sends the new gradient information to each BS so that it can update its own DRL algorithm. To summarize, the main contributions of this paper are as follows:

1. A secure transmission method based on FL for LSECS is proposed, and secure data transmission is realized.
2. To solve the problems of sub-channel selection and power allocation, a computation offloading algorithm based on a deep Q-network (DQN) is proposed to achieve efficient resource management in cellular networks under the conditions of BS SI and user co-channel interference.
3. The experimental results show that the proposed algorithm greatly enhances the weighted sum-rate and achieves excellent convergence.

2 Preliminaries

2.1 Notations

We summarize the main notations used in this paper as shown in Table 1.

2.2 DQN

DRL can train the agents to execute actions by interacting with the environment to obtain optimal rewards. As an improvement of classical Q-learning, the DQN algorithm (Liao et al., 2023) combines the

Table 1 Table of main notations

Notations	Description
I	Number of base stations
K	Number of downlink half-duplex users
J	Number of uplink half-duplex users
S	Number of sub-channels
E	Number of episode
T	Number of slots
i_K	Number of downlink half-duplex users
i_J	Number of uplink half-duplex users
$\eta_{i,k}$	Weight of the downlink of user k
$\lambda_{i,j}$	Weight of the uplink of user j
$p_{i,k}^d(n)$	Transmission power from base station to user
$p_{i,j}^u(n)$	Transmission power from user to base station
$N_{i,k}$	Gaussian noise variance of user k
$N_{i,0}$	Gaussian noise variance of base station
$S_{i,k}^d$	Set of sub-channels in the downlink of user k
$S_{i,j}^u$	Set of sub-channels in the uplink of user j
β_i	Self-interference cancellation coefficient
$h_{i,k}(n)$	Channel gain between base station and user k
$h_{i,j}(n)$	Channel gain between user j and base station
$h_{i,k,j}(n)$	Channel gain between users j and k
$P_{i,j}$	Maximum transmission power of user j
$P_{i,0}$	Maximum transmission power of base station

advantages of Q-learning and neural networks (NNs) and retains memories for learning past experience. The process is described below.

At time step t , agent i observes the environment, obtains observations $s_i(t)$, and then executes the action $a_i(t)$ based on the state $s_i(t)$. The environment is affected by the executed action $a_i(t)$ and subsequently transfers to the next state $s_i(t+1)$. Therefore, the reward $r_i(t)$ is calculated for the agent to adjust the policy to evaluate the effect of the action. A deep neural network (DNN) is used as a function approximator. The input of the DNN is the state $s_i(t)$ and the output is the action value $Q_i(s_i(t), a; \theta_i)$. The action $a_i(t)$ can be obtained by $a_i(t) = \max_a Q_i(s_i(t), a; \theta_i)$. At time step t , the interaction of agent i with the environment forms the experience that is described by the tuple $(s_i(t), a_i(t), r_i(t), s_i(t+1))$, which is stored in the experience replay.

Each agent uses two DNN structures—a online network and target network, respectively to enhance the stability and convergence of learning. Agent i randomly samples mini-batch $(s_i(l), a_i(l), r_i(s_i(l), a_i(j)), s_i(l+1))$ from experience replay for training and updating the network parameters θ_i and θ_i^- . Agent i uses the online network to calculate the evaluation Q value

$Q_i(s_i(l+1), a; \theta_i)$, using the target network to calculate the target Q value $y_i(l) = r_i(l) + \gamma \max_{a'} \hat{Q}_i(s_i(l+1), a'; \theta_i^-)$. Agent i performs gradient descent using the mean squared error (MSE) (Wang ZJ et al., 2024) between the evaluated Q values and the target Q value to update the online network. Then, the target network is updated according to $\theta_i^- = \theta_i$ every Z steps.

3 System model

3.1 LSECS model

The LSECS model proposed in this paper is shown in Fig. 1. On the ground side, we consider a cellular network consisting of I cells, each of which consists of an FD BS, J uplink half-duplex (UHD) users, and K downlink half-duplex (DHD) users. We consider cells with different scenarios, where users in each cell have different rate levels. Each BS maintains a DRL model and uploads local model parameters to a LEO satellite at regular intervals.

On the LEO satellite side, we consider placing MEC servers on LEO satellites to form the LSECS to provide global aggregation of local model parameters uploaded by the ground BSs. After global aggregation, each BS downloads the global model parameters from the LEO satellite, which is used to update the local model parameters.

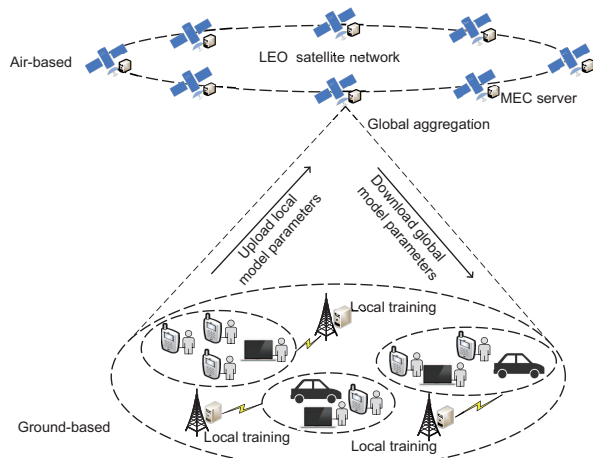


Fig. 1 LEO satellite edge computing system (LSECS) model

3.2 Channel model

In this paper, we consider the COST-231 Hata model as the radio wave propagation model:

$$\text{PL} = 46.3 + 33.9 \lg f - 13.82 \lg h_b - \alpha(h_u) + (44.9 - 6.55 \lg h_b) \lg d + C, \quad (1)$$

where PL is the path loss (dB), f is the carrier frequency (MHz), h_b is the effective antenna height of the BS (m), h_u is the effective antenna height of the user (m), d is the distance between the BS and the users (km), $\alpha(h_u)$ is the effective height correction factor (dB) of the user antenna, and C is the correction factor, which is 3 dB for urban and 0 dB for rural or suburban areas. $\alpha(h_u)$ in the small city scenario considered in this paper can be calculated as follows:

$$\alpha(h_u) = (1.1 \lg f - 0.7) h_u - 1.56 \lg f + 0.8. \quad (2)$$

3.3 Ground cell model

The communication and interference links in each ground cell are shown in Fig. 2. A single cell consists of a BS containing imperfect FD and multiple half-duplex (HD) users. An HD user can be a UHD user or a DHD user. Due to the FD characteristic of the network, the BS suffers from its own interference, and the UHD users and DHD users on the same sub-channels also interfere with each other.

In each ground cell, we consider that the channels are reciprocal and the uplink and downlink channels' gain are the same. This paper considers that the BS knows all the channel gain, noise power, SI cancellation factor, and downlink and uplink weights assigned to all users. In each ground cell, we assume an orthogonal frequency division multiple access (OFDMA) system with S sub-channels. We consider that all sub-carriers are perfectly synchronized (Fu et al., 2023) and the BS operates in FD mode. In each time slot, the BS allocates the sub-channel to the downlink or uplink of the user and determines the relevant transmission power. We consider the BS as an imperfect FD node with SI (Sultan and Shamseldeen, 2024). We define an SI cancellation coefficient of the i^{th} BS, expressed by $0 \leq \beta_i \leq 1$, where $\beta_i = 0$ means that the SI is completely cancelled out and $\beta_i = 1$ means that there is no SI cancellation.

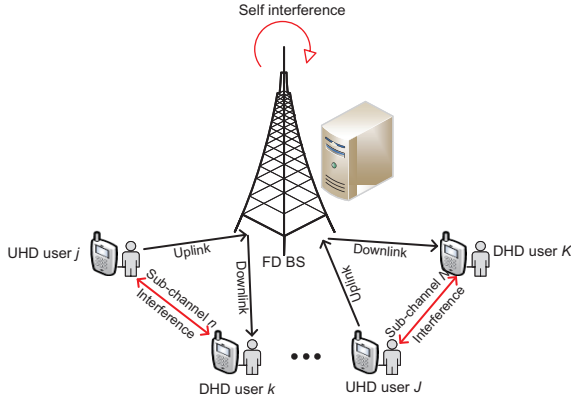


Fig. 2 Ground cell model

4 Problem statements

The optimization objective of this paper is to maximize the weighted sum-rate of the DHD users and UHD users under the constraints of the total power of the BS and the transmission power of each user. Let $\eta_{i,k}$ and $\lambda_{i,j}$ represent the system preference of the system for the downlink achievable rate of user k and the preference for the uplink achievable rate of user j in the i th cell, respectively. The system can dynamically adjust parameters $\eta_{i,k}$ and $\lambda_{i,j}$ according to the users' requirements on the uplink and downlink achievable rates.

We define the downlink weighted sum-rate in the i th cell as:

$$R_i^d = \sum_{k=1}^{i_K} \sum_{n \in S_{i,k}^d} \eta_{i,k} \log_2 \left(1 + \frac{h_{i,k}(n) p_{i,k}^d(n)}{N_{i,k} + h_{i,k,j_n}(n) p_{i,j_n}^u(n)} \right), \quad (3)$$

where j_n represents the UHD user who selects sub-channel n for uplink data transmission.

We define the uplink weighted sum-rate in the i th cell as:

$$R_i^u = \sum_{j=1}^{i_J} \sum_{n \in S_{i,j}^u} \lambda_{i,j} \log_2 \left(1 + \frac{h_{i,j}(n) p_{i,j}^u(n)}{N_{i,0} + \beta_i p_{i,k_n}^d(n)} \right), \quad (4)$$

where k_n represents the DHD user who selects sub-channel n for downlink data transmission.

Let $P_{i,0}$ and $P_{i,j}$ be the maximum available transmission power of the BS and user j in the i th cell, respectively. The proposed optimization problem $P1$ can be expressed as:

$$P1: \quad \text{maximize} \quad R_i^d + R_i^u, \quad (5)$$

$$\text{st:} \quad \sum_{k=1}^{i_K} \sum_{n \in S_{i,k}^d} p_{i,k}^d(n) \leq P_{i,0} \quad \forall i, \quad (6)$$

$$\sum_{n \in S_{i,j}^u} p_{i,j}^u(n) \leq P_{i,j} \quad \forall i, j, \quad (7)$$

$$p_{i,j}^u(n), p_{i,k}^d(n) \geq 0 \quad \forall i, j, k, n, \quad (8)$$

$$\cup_{j=1}^J S_{i,j}^u, \cup_{k=1}^K S_{i,k}^d \subseteq \{1, 2, \dots, S\} \quad \forall i, \quad (9)$$

$$S_{i,v}^d \cap S_{i,x}^d = \phi, S_{i,v}^u \cap S_{i,x}^u = \phi \quad \forall i, \forall v \neq x, \quad (10)$$

$$S_{i,j}^u \cap S_{i,j}^d = \phi, S_{i,k}^u \cap S_{i,k}^d = \phi \quad \forall i, j, k. \quad (11)$$

Constraints (7) and (8) represent, respectively, the power constraints of the BS and users. Constraint (9) shows that the power is positive. Constraint (10) indicates that the system does not exceed S sub-channels. Constraint (11) shows that one sub-channel cannot be assigned to two different users at the same time. Constraint (12) indicates the HD feature of HD users.

5 Proposed method

The specific process of the proposed method is shown in Algorithm 1. In this paper, the state, action, and reward are defined as follows:

1. State: To better describe the main characteristics of cellular network, we define the state space of the BS agent i as follows:

$$s_i = \{h_{i,k}(n), h_{i,j}(n), h_{i,k,j}(n)\} \quad \forall j, k, n. \quad (12)$$

2. Action: BS agent i performs actions, including allocating BS and user transmission power and setting of allocated sub-channels, as follows:

$$a_i = \{p_{i,k}^d(n), p_{i,j}^u(n), S_{i,k}^d, S_{i,j}^u\} \quad \forall j, k, n. \quad (13)$$

Algorithm 1 Federated DQN-based computation offloading algorithm

```

1: Set the federated global aggregation period  $A_g$ , the
   total number of training episodes  $E$  and exploration
   parameter  $\epsilon$ , the network parameters  $\theta = \{\theta_i, \theta_i^-\}$ 
2: Initialize the experience replay
3:  $\forall i \in I$ , initiate the online network  $\theta_i$  and the target
   network  $\theta_i^- = \theta_i$ 
4: for each episode  $e = 1, 2, \dots, E$  do
5:   Get initial observation  $s_0$ 
6:   for each slot  $t = 1, 2, \dots, T$  do
7:     for each BS  $i = 1, 2, \dots, I$  do
8:       Choose a random probability  $p$ 
9:       if  $p \leq \epsilon$  then
10:        Randomly select an action  $a_i(t)$ 
11:       else
12:         $a_i(t) = \max_a Q_i(s_i(t), a; \theta_i)$ 
13:       end if
14:       Perform the selected action  $a_i(t)$ , calculate
       the reward  $r_i(t)$ , and obtain the next state
        $s_i(t+1)$ 
15:       Store  $(s_i(t), a_i(t), r_i(t), s_i(t+1))$  into the
       experience replay
16:       Randomly sample mini-batch
        $(s_i(l), a_i(l), r_i(s_i(l), a_i(l)), s_i(l+1))$ 
       from the experience replay
17:       Set target Q value  $y_i(l) = r_i(l) +$ 
        $\gamma \max_{a'} \hat{Q}_i(s_i(l+1), a'; \theta_i^-)$ .
18:       Perform gradient descent on
       the loss function  $F(\theta) =$ 
        $\frac{1}{|L|} \sum_{l \in L} (Q_i(s_i(l), a_i(l); \theta_i) - y_i(l))^2$  to
       update the online network.
19:       Every  $Z$  steps, the target network is updated
       according to  $\theta_i^- = \theta_i$ 
20:     end for
21:   end for
22:   if  $e \bmod A_g = 0$  then
23:     The BSs upload the local model parameters to
     the LEO satellite for global aggregation
24:     The LEO satellite sends the global model pa-
     rameters to each BS for training and updating
25:   end if
26: end for

```

3. Reward: To maximize the weighted sum-rate of the i^{th} cell, we define the reward as:

$$r_i = \{R_i^d + R_i^u\}. \quad (14)$$

The workflow of the proposed federated DQN-based computation offloading algorithm is shown in Fig. 3. The LEO satellite maintains a global DRL model and each BS builds its own local DRL model

using the same network structure. The structure of the DQN model is shown in Fig. 3. Each BS agent has two NNs, namely the online network and the target network. The specific DQN process was introduced in Section 2.2.

The local update of the i^{th} agent is implemented by minimizing the above loss function $F(\theta)$. The global model is obtained by weighted averaging of the parameters of all the local models. The minimization of the global loss function $F_i(\theta)$ follows:

$$\min_{\theta} F(\theta) = \sum_{i=1}^I \frac{\rho_i}{\rho} F_i(\theta), \quad (15)$$

where θ represents network parameters $\theta = \{\theta_i, \theta_i^-\}$, ρ_i is the number of samples at BS i , and ρ is the total number of samples at all BSs.

Because the observed state in each cell cannot fully characterize the entire cellular network environment, FL provides an effective way of improving model performance by using decentralized local DRL models (El Houda et al., 2024). In the e^{th} iteration, each BS agent interacts with a LEO satellite acting as the model aggregator as follows:

1. Local update: Agent i first receives the latest global model parameters $\theta(e-1) = \{\theta_i(e-1), \theta_i^-(e-1)\}$ from the LEO satellite to obtain the local model parameters. Agent i then calculates the gradient based on historical experience and updates the local model parameters using Eq. (16):

$$\theta_i(e) = \theta(e-1) - \eta \cdot \nabla F_i(\theta(e)). \quad (16)$$

2. Upload: After the local update is completed, the i^{th} BS sends $\theta_i(e)$ to the LEO satellite.

3. Aggregation and feedback: The LEO satellite receives all the uploaded local model parameters and aggregates the models to obtain the updated global model:

$$\theta(e) = \sum_{i=1}^I \frac{n_i}{n} \theta_i(e). \quad (17)$$

We think the process of global aggregation can be completed in one time slot. Although the mobility of the LEO satellites may impact the system performance, by observing the formula of free space loss $L_p = 32.4 + 20 \lg(d) + 20 \lg(f)$, where d is the distance between the LEO satellite and the ground cells, and f is the communication frequency, it can be seen that the logarithm base 10 should be taken for the

distance between the LEO satellite and the ground cells. Therefore, changes in d caused by the mobility of the LEO satellite in one time slot has little influence on the free space loss. Therefore, changes in d caused by the mobility of the LEO satellite in one time slot has little influence on the channel gain of the LEO satellite-ground link. Therefore, changes in d caused by the mobility of the LEO satellite in one time slot has little influence on the system performance. Moreover, we focus on maximizing the weighted sum-rate of the DHD and UHD users under the constraints of the total power of the BS and the transmission power of each user. It can be seen from Eqs. (3) and (4) that the weighted sum-rate of the system is independent of the channel gain of the LEO satellite-ground link. Therefore, we neglect the mobility of LEO satellites in our work.

6 Experimental results and analysis

In this section, we evaluate the proposed federated DQN-based computation offloading algorithm in the cellular network with imperfect FD BSs and HD users. We summarize the simulation parameters used in this paper as shown in Table 2.

The channel gain remains constant in each time slot and varies independently between time slots. The users in different cells have different mobility levels. Different users in the same cell have different mobility speeds.

In this paper, all models were implemented in PyTorch and trained using the Adam optimizer. For the NNs used in this paper, the number of inputs is the state dimension, the number of outputs is the action dimension, the network has two hidden layers with 128 and 64 neurons, respectively, and the activation function is the ReLU activation function.

Fig. 4 depicts the loss function for each agent as the number of training steps increases. As shown in Fig. 4, the loss function of each agent decreases quickly and stabilizes toward 0, and can achieve convergence in very few training steps.

Fig. 5 depicts the weighted sum-rate of each agent in the cellular network for different numbers of episodes. As can be seen from Fig. 5, the weighted sum-rate of each agent increases as the number of episodes increases and finally stabilizes, which indicates that each agent has learned the effective computation offloading strategy during the interaction

Table 2 Table of simulation parameters

Parameter	Value
I	3
K	10
J	10
S	64
E	24
T	500
$P_{i,0}$	43 dBm
$P_{i,j}$	23 dBm
$\eta_{i,k}$	0.5
$\lambda_{i,j}$	0.5
β_i	10^{-11}
f	2 GHz
h_b	30 m
h_u	1.5 m
ϵ	0.9
Batch size	500
Cell radius	1 km
Total bandwidth	10 MHz
Bandwidth of each sub-channel	150 KHz
Noise density	-170 dBm/Hz

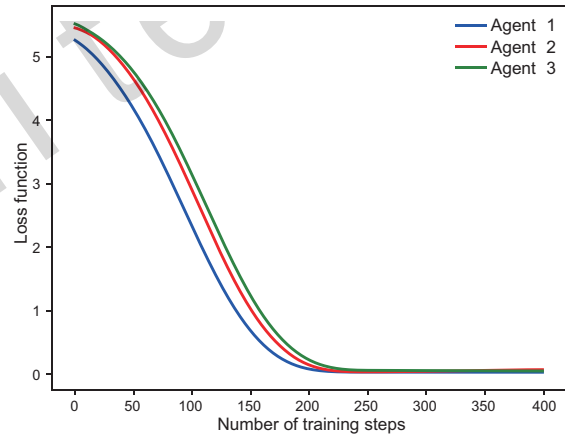


Fig. 4 Evaluation loss function for each agent over training steps

with the environment. From Algorithm 1, it can be seen that all agents train the NNs to learn the offloading strategies in each slot, and an episode includes multiple slots. Therefore, each agent can achieve convergence in very few training episodes.

We define the sum of the weighted sum-rate of all agents as the total weighted sum-rate. Fig. 6 depicts the total weighted sum-rate under different aggregation frequencies as the number of episodes increases. Fig. 6 shows that our proposed algorithm achieves excellent convergence performance under different aggregation frequencies. It can be seen from Fig. 6 that the obtained total weighted sum-rate is optimal when the aggregation frequency is 2 in the

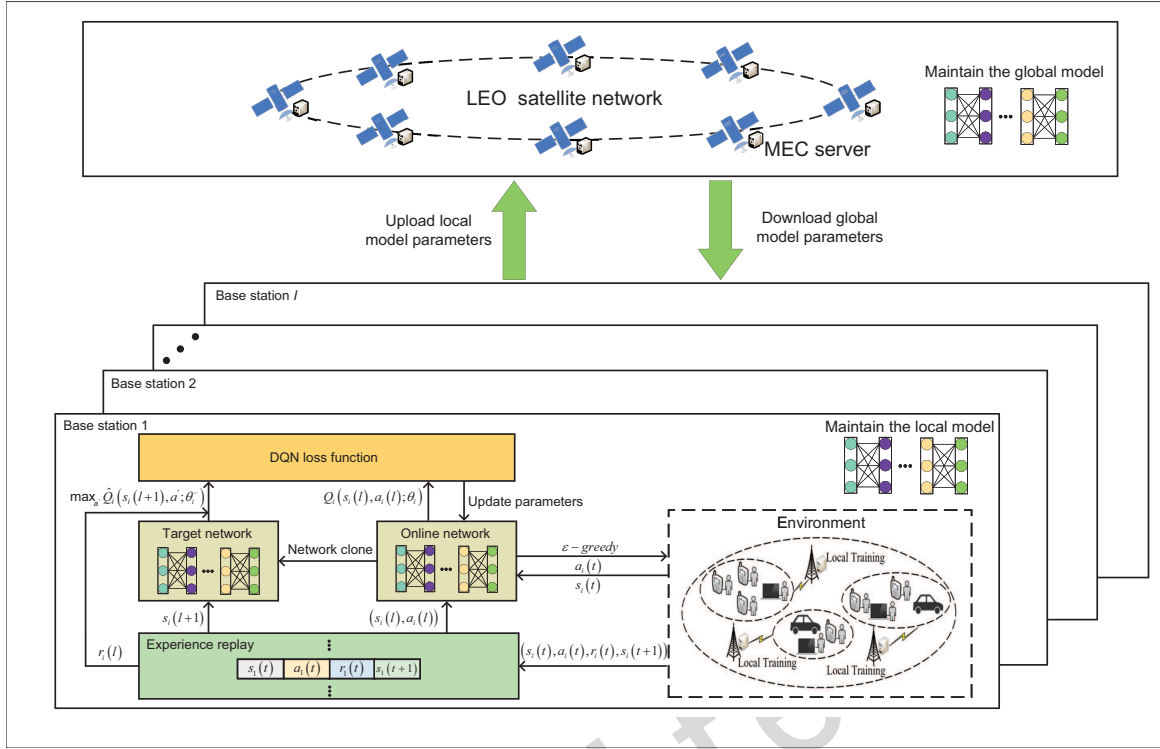


Fig. 3 Proposed federated deep Q-network (DQN)-based computation offloading architecture

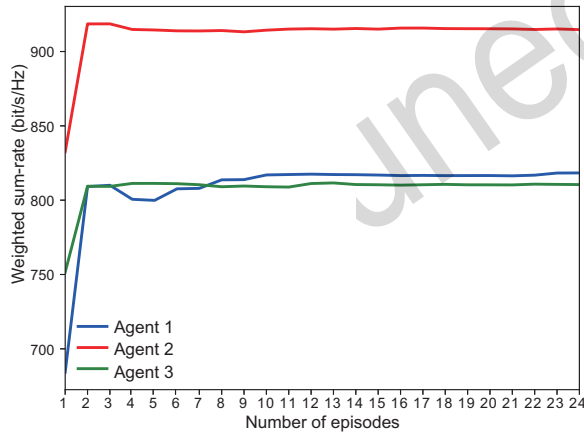


Fig. 5 Evaluation weighted sum-rate for each agent over episodes

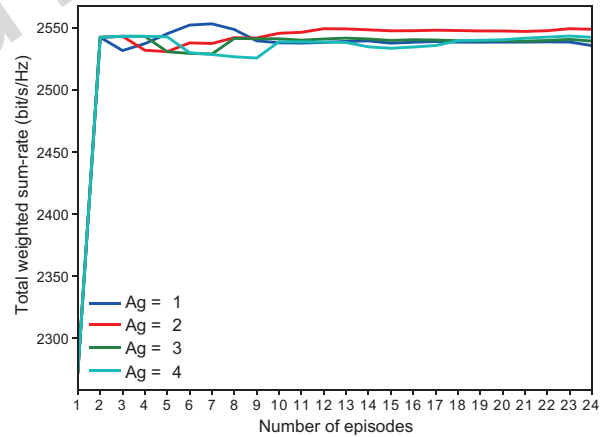


Fig. 6 Evaluation total weighted sum-rate for different aggregation frequencies over episodes

scenario of this paper under different contrasting aggregation frequencies.

Fig. 7 depicts the total weighted sum-rate at different learning rates as the number of episodes increases. Under different learning rates, our proposed algorithm achieves fast convergence. When the learning rate is set to e^{-1} , our proposed algorithm obtains the optimal total weighted sum-rate under different contrasting learning rates. As seen in Fig. 7, when the learning rate is set to e^{-4} , al-

though the final performance level is achieved in the first episode, the system does not converge at that time. This is because a lower learning rate will cause the updating of NN parameters to be too conservative and slow to jump out of the local optimum. Therefore, when the learning rate is set to e^{-4} , the system shows poor performance but follows a different learning trend from the other three lines.

To verify the efficiency of the proposed algorithm, we compared the following algorithms by sim-

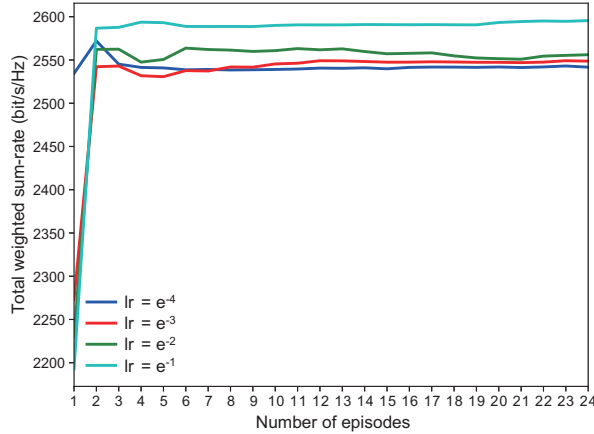


Fig. 7 Evaluation total weighted sum-rate for different learning rates over episodes

ulation: the federated multi-agent RL (FMARL) (Xu X et al., 2024), distributed multi-agent deep Q-Learning (DMADQL) algorithm (Lim and Vu, 2023), and random algorithms. In the FMARL algorithm, each agent learns independently through the FL algorithm during the training phase and operates autonomously through the stochastic gradient descent (SGD) algorithm during the execution phase. In the DMADQL algorithm, each agent uses a private DQN that exploits only local information without information exchange with BSs and/or between agents. The random algorithm randomly generates sub-channel selection strategies and power allocation schemes.

Fig. 8 depicts the total weighted sum-rate of different algorithms for different numbers of users. As shown in Fig. 8, the greater the number of users, the higher the total weighted sum-rate, because the total weighted sum-rate is proportional to the number of users. Among the four algorithms, the proposed algorithm obtains the optimal total weighted sum-rate, which shows that the proposed algorithm can achieve efficient resource management with different numbers of users in the system.

Fig. 9 depicts the total weighted sum-rate of different algorithms for different numbers of sub-carriers per user. It can be seen from Fig. 9 that a higher number of sub-carriers per user leads to a higher total weighted sum-rate, which is obviously reasonable. The total weighted sum-rate is proportional to the number of total sub-carriers. Among the four algorithms, our proposed algorithm obtains the optimal total weighted sum-rate, indicating that the proposed algorithm learns excellent resource allocation strategies in the process of interacting with

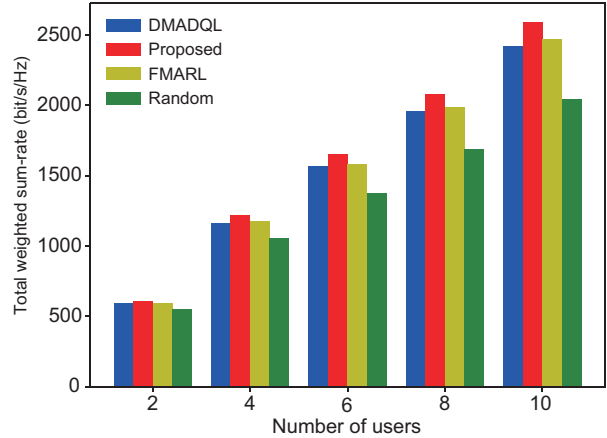


Fig. 8 Evaluation total weighted sum-rate of different algorithms for different numbers of users

the environment, and can achieve efficient computation offloading.

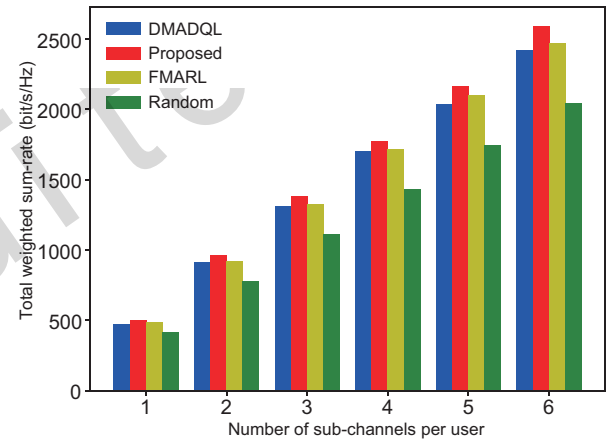


Fig. 9 Evaluation total weighted sum-rate of different algorithms for different numbers of sub-carriers per user

By comparing the results of different algorithms, it can be seen that our proposed algorithm significantly improves the total weighted sum-rate. Compared with the FMARL algorithm, the proposed federated DQN-based computation offloading algorithm is more suitable for processing the problem of discrete user sub-carrier selection. Therefore, the total weighted sum-rate of the FMARL algorithm is lower than our proposed algorithm. The DMADQL algorithm relies only on the data inside each cell for training, and cannot obtain the parameters of other BS agents, so the total weighted sum-rate is lower than our proposed algorithm. The random algorithm randomly generates the resource allocation schemes and cannot interact with the environment to learn

the corresponding strategies, so the total weighted sum-rate is lower than other algorithms.

To illustrate the scalability and credibility of the simulation results, we depict the total weighted sum-rate for different numbers of agents as the number of episodes increases in Fig. 10. As can be seen from Fig. 10, the proposed algorithm has good convergence performance and significantly improves the total weighted sum-rate with different numbers of agents, indicating that cellular networks of different sizes can learn efficient computation offloading strategies.

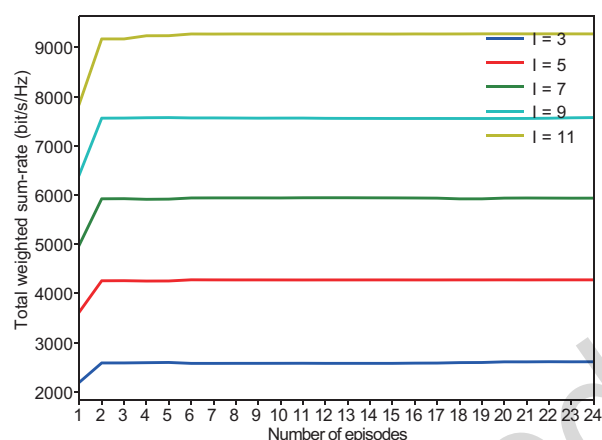


Fig. 10 Evaluation total weighted sum-rate for different numbers of agents over episodes

7 Conclusions

In this paper, the LSECS model is established, and then the weighted sum-rate maximization problem is modeled. Aiming at the problems of sub-channel allocation and power allocation, this paper proposes a computation offloading algorithm based on federated DQN. The simulation results show that the proposed algorithm can significantly improve the weighted sum-rate with excellent convergence.

Contributors

Min JIA, Jian WU, and Qing GUO designed the research. Min JIA, Jian WU, and Xinyu WANG processed the data. Jian WU drafted the paper. Min JIA helped organize the paper. Jian WU revised and finalized the paper.

Conflict of interest

All the authors declare that they have no conflict of interest.

Data availability

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

References

- Alkhrijah Y, Camp J, Rajan D, 2023. Multi-band full duplex MAC protocol (MB-FDMAC). *IEEE J Sel Areas Commun*, 41(9):2864-2878. <https://doi.org/10.1109/JSAC.2023.3287546>
- Chen H, Xiao M, Pang ZB, 2022. Satellite-based computing networks with federated learning. *IEEE Wirel Commun*, 29(1):78-84. <https://doi.org/10.1109/MWC.008.00353>
- Chen XM, Xu ZB, Shang L, 2023. Satellite internet of things: challenges, solutions, and development trends. *Front Inform Technol Electron Eng*, 24(7):935-944. <https://doi.org/10.1631/FITEE.2200648>
- Dai XY, Zhao C, Wang X, et al., 2022. Image-based traffic signal control via world models. *Front Inform Technol Electron Eng*, 23(12):1795-1813. <https://doi.org/10.1631/FITEE.2200323>
- El Houda ZA, Moudoud H, Brik B, 2024. Federated deep reinforcement learning for efficient jamming attack mitigation in O-RAN. *IEEE Trans Veh Technol*, 73(7):9334-9343. <https://doi.org/10.1109/TVT.2024.3359998>
- Fawaz H, Lahoud S, Helou ME, et al., 2023. Queue-aware resource allocation in full-duplex multi-cellular wireless networks. *IEEE J Sel Areas Commun*, 41(9):2852-2863. <https://doi.org/10.1109/JSAC.2023.3287541>
- Fu H, Si WJ, Kim IM, 2023. Deep learning-based joint pilot design and channel estimation for OFDM systems. *IEEE Trans Commun*, 71(8):4577-4590. <https://doi.org/10.1109/TCOMM.2023.3280937>
- Gao YF, Ji Z, Zhao KL, et al., 2024. Game-based computation offloading and power allocation for LEO constellation networks in distributed and dynamic environment. *IEEE Int Things J*, 11(4):7040-7058. <https://doi.org/10.1109/JIOT.2023.3314650>
- Han DJ, Hosseinalipour S, Love DJ, et al., 2024. Cooperative federated learning over ground-to-satellite integrated networks: joint local computation and data offloading. *IEEE J Sel Areas Commun*, 42(5):1080-1096. <https://doi.org/10.1109/JSAC.2024.3365901>
- He ZY, Xu W, Shen H, et al., 2023. Full-duplex communication for ISAC: joint beamforming and power optimization. *IEEE J Sel Areas Commun*, 41(9):2920-2936. <https://doi.org/10.1109/JSAC.2023.3287540>
- Jia M, Wu J, Zhang L, et al., 2023. Joint optimization communication and computing resource for LEO satellites with edge computing. *Chin J Electron*, 32(5):1011-1021. <https://doi.org/10.23919/cje.2022.00.314>
- Jia M, Wu J, Guo Q, et al., 2024. Service-oriented SA-GIN with pervasive intelligence for resource-constrained users. *IEEE Netw*, 38(2):79-86. <https://doi.org/10.1109/MNET.2024.3353414>
- Kamal M, Rashid I, Iqbal W, et al., 2023. Privacy and security federated reference architecture for internet of things. *Front Inform Technol Electron Eng*, 24(4):481-508. <https://doi.org/10.1631/FITEE.2200368>

- Kang YH, Zhu YF, Wang D, et al., 2024. Joint server selection and handover design for satellite-based federated learning using mean-field evolutionary approach. *IEEE Trans Netw Sci Eng*, 11(2):1655-1667. <https://doi.org/10.1109/TNSE.2023.3328776>
- Liao Y, Yang ZJ, Yin ZS, et al., 2023. DQN-based adaptive MCS and SDM FOR 5G massive MIMO-OFDM downlink. *IEEE Commun Lett*, 27(1):185-189. <https://doi.org/10.1109/LCOMM.2022.3210928>
- Lim B, Vu M, 2023. Distributed multi-agent deep q-learning for load balancing user association in dense networks. *IEEE Wirel Commun Lett*, 12(7):1120-1124. <https://doi.org/10.1109/LWC.2023.3250492>
- Lv ZH, Xiu WQ, 2020. Interaction of edge-cloud computing based on SDN and NFV for next generation IoT. *IEEE Int Things J*, 7(7):5706-5712. <https://doi.org/10.1109/JIOT.2019.2942719>
- Razmi N, Matthiesen B, Dekorsy A, et al., 2022. Ground-assisted federated learning in LEO satellite constellations. *IEEE Wirel Commun Lett*, 11(4):717-721. <https://doi.org/10.1109/LWC.2022.3141120>
- Salim S, Moustafa N, Hassanian M, et al., 2024. Deep-federated-learning-based threat detection model for extreme satellite communications. *IEEE Int Things J*, 11(3):3853-3867. <https://doi.org/10.1109/JIOT.2023.3301626>
- Sultan R, Shamseldeen A, 2024. Uplink-downlink cochannel interference cancellation in RIS-aided full-duplex networks. *IEEE Syst J*, 18(2):1220-1223. <https://doi.org/10.1109/JSYST.2024.3379438>
- Sun YW, Duan BY, Su X, et al., 2023. Performance analysis on reconfigurable intelligent surface and network-controlled repeater in 3GPP release-18. *Front Inform Technol Electron Eng*, 24(12):1815-1828. <https://doi.org/10.1631/FITEE.2300321>
- Tang FX, Wen C, Chen XH, et al., 2023. Federated learning for intelligent transmission with space-air-ground integrated network toward 6G. *IEEE Netw*, 37(2):198-204. <https://doi.org/10.1109/MNET.104.2100615>
- Teklu MB, Choi DY, Meng WX, 2024. Resource efficient full-duplex mode of transmissions under imperfect CSI. *IEEE Trans Broadcast*, 70(1):87-98. <https://doi.org/10.1109/TBC.2023.3323929>
- Tran DD, Sharma SK, Ha VN, et al., 2023. Multi-agent DRL approach for energy-efficient resource allocation in URLLC-enabled grant-free NOMA systems. *IEEE Open J Commun Soc*, 4:1470-1486. <https://doi.org/10.1109/OJCOMS.2023.3291689>
- Uddin R, Kumar SAP, 2023. SDN-based federated learning approach for satellite-IoT framework to enhance data security and privacy in space communication. *IEEE J Radio Freq Identif*, 7:424-440. <https://doi.org/10.1109/JRFID.2023.3279329>
- Vishnoi V, Budhiraja I, Gupta S, et al., 2023. A deep reinforcement learning scheme for sum rate and fairness maximization among D2D pairs underlying cellular network with NOMA. *IEEE Trans Veh Technol*, 72(10):13506-13522. <https://doi.org/10.1109/TVT.2023.3276647>
- Wang Q, Chen XM, Qi Q, 2024. Energy-efficient design of satellite-terrestrial computing in 6G wireless networks. *IEEE Trans Commun*, 72(3):1759-1772. <https://doi.org/10.1109/TCOMM.2023.3334813>
- Wang ZJ, Gao WF, Li GH, et al., 2024. Path planning for unmanned aerial vehicle via off-policy reinforcement learning with enhanced exploration. *IEEE Trans Emerg Top Comput Intell*, 8(3):2625-2639. <https://doi.org/10.1109/TETCI.2024.3369485>
- Wu J, Jia M, Zhang NT, et al., 2024. Multi-agent deep reinforcement learning-based computation offloading in LEO satellite edge computing system. *IEEE Commun Lett*, 28(10):2352-2356. <https://doi.org/10.1109/LCOMM.2024.3440489>
- Xiao Y, Song YQ, Liu J, 2023. Multi-agent deep reinforcement learning based resource allocation for ultra-reliable low-latency internet of controllable things. *IEEE Trans Wirel Commun*, 22(8):5414-5430. <https://doi.org/10.1109/TWC.2022.3233853>
- Xu HT, Han SY, Li XH, et al., 2023. Anomaly traffic detection based on communication-efficient federated learning in space-air-ground integration network. *IEEE Trans Wirel Commun*, 22(12):9346-9360. <https://doi.org/10.1109/TWC.2023.3270179>
- Xu X, Li RP, Zhao ZF, et al., 2024. The gradient convergence bound of federated multi-agent reinforcement learning with efficient communication. *IEEE Trans Wirel Commun*, 23(1):507-528. <https://doi.org/10.1109/TWC.2023.3279268>
- Yu B, Qian C, Lee J, et al., 2023. Realizing high power full duplex in millimeter wave system: design, prototype and results. *IEEE J Sel Areas Commun*, 41(9):2893-2906. <https://doi.org/10.1109/JSAC.2023.3287609>
- Zhao D, Zheng Z, Qi PF, et al., 2024. Resource allocation in multi-user cellular networks: a transformer-based deep reinforcement learning approach. *China Commun*, 21(5):77-96. <https://doi.org/10.23919/JCC.ea.2021-0665.202401>