aerospace

MDPI

Article

A DRL-Based Satellite Service Allocation Method in LEO Satellite Networks

Yafei Zhao, Jiaen Zhou *, Zhenrui Chen and Xinyang Wang

State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China; zhaoyafei@bupt.edu.cn (Y.Z.); czr2018@bupt.edu.cn (Z.C.); wangxinyang@bupt.edu.cn (X.W.)
* Correspondence: circularje@bupt.edu.cn

Abstract: Satellite computing represents a recent computational paradigm in the development of low Earth orbit (LEO) satellites. It aims to augment the capabilities of LEO satellites beyond their current transparent relay functions by enabling real-time processing, thereby providing low-latency computational services to end users. In LEO constellations, a significant deployment of computationally capable satellites is orchestrated to offer enhanced computational resources. Challenges arise in the optimal allocation of terminal services to the most suitable satellite due to overlapping coverage among neighboring satellites, compounded by constraints on satellite energy and computational resources. The satellite service allocation (SSA) problem is recognized as NP-hard, yet assessing allocation methods through results allows for the application of deep reinforcement learning (DRL) to obtain improved solutions, partially addressing the SSA challenge. In this paper, we introduce a satellite computing capability model to quantify satellite computational resources. A DRL model is proposed to address service demands, computational resources, and resolve service allocation conflicts, strategically placing each service on appropriate servers. Through simulation experiments, numerical results demonstrate the superiority of our proposed method over baseline approaches in service allocation and satellite resource utilization, showcasing advancements in this field.

Keywords: LEO satellite network; satellite computing; satellite service allocation; deep reinforcement learning

Citation: Zhao, Y.; Zhou, J.; Chen, Z.; Wang, X. A DRL-Based Satellite Service Allocation Method in LEO Satellite Networks. *Aerospace* **2024**, *11*, 386. https://doi.org/10.3390/aerospace11050386

Academic Editor: Mikhail Ovchinnikov

Received: 13 March 2024
Revised: 4 May 2024
Accepted: 8 May 2024
Published: 13 May 2024

1. Introduction

In recent years, with the continuous reduction in the size of computing hardware, the enhancement of computational capabilities, and the decrease in power consumption, LEO satellites equipped with computing cards have been able to provide a certain level of edge computing capability [1]. Satellite edge computing has emerged as a novel computing paradigm, allowing the processing of tasks on LEO satellites [2–4]. This not only reduces user service latency but also conserves valuable bandwidth resources from satellites to ground-based cloud centers [5–7]. This paradigm is advantageous for LEO satellite missions such as border surveillance and unmanned area monitoring [6–12]. Due to the typically limited power availability of a few hundred watts for LEO satellites, with only a portion allocated for computing purposes, the computational resources on a single satellite are constrained [13,14]. Additionally, these resources usually cover only a limited geographical area. Therefore, the efficient utilization of satellite computing resources in integrated satellite terrestrial networks (ISTNs) is crucial for the quality of LEO satellite services [15–17]. In an ISTN, satellite networks are positioned at the edge and connected to ground networks through ground stations [18]. However, the data link from users to satellites and then to the cloud center is too long. For delay-sensitive services, users typically access services directly at satellite nodes to meet their service requirements. This imposes higher demands on satellite resource capacity. Moreover, when multiple satellites have the ability to serve the same users simultaneously, selecting the appropriate satellites to achieve higher resource utilization becomes a research-worthy problem.

Many scholars have conducted research on this topic. For single satellite coverage scenarios, Wang et al. proposed the JCORA strategy [19], which considers energy consumption and latency as system costs to determine whether tasks should be offloaded to the access satellite. The work in [20] takes into account the energy and computational constraints of the satellite and presents a dynamic offloading strategy based on Lyapunov theory. Wang et al. proposed a double-edge offloading mechanism in [21]. When the offloading capacity of ground edge servers is insufficient, the access satellite selects different cost-matching modes based on the size or type of the tasks to schedule edge servers in adjacent satellite nodes. For multi-satellite coverage scenario, the work by [22–24] proposes the joint optimization problem of task offloading and computing resource allocation for ground users and then decoupling it, while in [25], the collaborative computing offloading method in satellite edge computing is described, in which edge satellite nodes are used to perform tasks for the source satellite to optimize multidimensional resources. In [22], Jia et al. adopted game theory methods to obtain offloading decisions in Nash equilibrium solutions. After proving the resource optimization problem is convex, they solved it by using the Lagrange multiplier method. In [23], the goal was to minimize the energy consumption of the satellite while meeting task latency requirements, and in [24], the focus is more on user service quality, with the weighted sum of user task latency and energy consumption being used as indicators. The above works have studied various schemes to provide computing offloading services to users but have not focused on the access choices of users in multi-satellite coverage scenarios. Some multi-satellite computing offloading schemes only schedule tasks and allocate resources using the access satellite's decision.

Some authors have considered the association between users and satellites. In [26], a joint client selection and resource allocation strategy are proposed for the multi-task joint learning system, which selects satellite nodes based on wireless communication factors and local task data training results. In [27–29], a correlation strategy of a three-layer architecture that contains the satellite, base station, and user is considered. Another proposal [27] aims to reduce end-to-end delay by developing an iterative algorithm based on approximation and relaxation methods, which solves the optimization problem of association and resource allocation under load balancing and demand constraints of user device. A different solution [28] focuses on the user, searching for base stations with a maximum rate and minimum load and connecting them to the satellite through the best channel quality. In [29], Li et al. adopted a multi-agent DRL algorithm which sets users, base stations, and satellites as agents. The SI-R based utility function is considered when selecting the access base stations or satellites for users. The sample is be sent to the MADDPG model for training. In [30], the authors propose a satellite–gateway association strategy using the graph-theory approach, where a bipartite graph is established with the weight set from the channel gain and the coverage of satellites. Accordingly, the association solution is obtained by finding the maximum weighted matching. The satellite selection problem is considered as a potential game in [31], where ground user equipment competes for satellite and channel resources. In [32], the authors proposed a multi-objective satellite selection strategy for multiple users based on reinforcement learning. The work by [31,32] both considers factors such as remaining visible time, terminal-satellite elevation angle, and available channel numbers to assist users in selecting satellites for decision-making. In [33,34], the GS algorithm is used for initial matching in the user-satellite association/computation offloading subproblem to obtain the user's satellite selection decision results. In the design of the preference list, communication conditions are still the main factor. Although the above papers examine the access satellite selection of users under multi-satellite coverage, most of them only refer to the impact of the communication environment and rarely consider the resource utilization and service satisfaction of users in satellite systems, which is not conducive to the load balancing of the entire satellite–terrestrial integrated network.

Some studies have considered the association between users and satellites. The work by [26] proposes a joint client selection and resource allocation strategy for the multi-task joint learning system, which selects satellite nodes based on wireless communication fac-

tors and local task data training results. In [27–29], the authors consider the correlation strategy of the three-layer architecture which contains the satellite, base station, and user. In [27], the authors aim to reduce end-to-end delay by developing an iterative algorithm based on approximation and relaxation methods, which solves the optimization problem of association and resource allocation under load balancing and demand constraints of the user device. The solution described in [28] focuses on the user, searching for base stations which have maximum rate and minimum load and connecting them to the satellite through the best channel quality. In [29], Li et al. adopted a multi-agent DRL algorithm which sets users, base stations, and satellites as agents. The SINR-based utility function is considered when selecting the access base stations or satellites for users. The sample is sent to the MADDPG model for training. In [30], the authors proposed a satellite–gateway association strategy using the graph-theory approach, where a bipartite graph is established with the weight set from the channel gain and the coverage of satellites. Accordingly, the association solution is obtained by finding the maximum weighted matching. The satellite selection problem is considered as a potential game in [31] ,where ground user equipment competes for satellite and channel resources. In [32], the authors proposed a multi-objective satellite selection strategy for multiple users based on reinforcement learning. In [31,32], factors such as remaining visible time, terminal–satellite elevation angle, and available channel numbers are considered to assist users in selecting satellites for decision-making. Other works [33,34] have used the GS algorithm for initial matching in the user–satellite association/computation offloading subproblem to obtain the user's satellite selection decision results. In the design of the preference list, communication conditions are still the main factor. Although the above papers above have studied the access satellite selection of users under multi-satellite coverage, most of them only refer to the impact of the communication environment and rarely consider the resource utilization and service satisfaction of users in satellite systems, which is not conducive to load balancing of the entire satellite–terrestrial integrated network.

In the literature, a major issue arises when there is an abundance of residual communication resources on a satellite but insufficient computational or storage resources. Algorithms that solely consider communication resources may erroneously allocate users to such satellites, leading to service interruptions or unavailability. Similarly, algorithms solely focusing on computational resources are constrained by communication, storage, and cache resources. Thus, only by comprehensively considering all resources required for user service can more effective services be provided to users and various resources on satellites utilized more fully.

Unlike the background work mentioned above, we innovatively consider five types of resources including communication, computation (CPU, GPU), storage, and cache. To comprehensively considering both user service demands and the characteristics of resource capacity on satellites, we utilize DRL to devise an efficient method for allocating user services. Based on the above considerations, this work conducts a detailed modeling of satellite resources and adopts DRL methods to achieve optimal access satellite selection decisions for users in multi-satellite coverage scenarios with the goal of maximizing the user service rate and system resource utilization rate. Our main contributions can be summarized as follows:

- In accordance with the actual computational resource conditions of LEO satellites, we have modeled the computational resources of LEO satellites. Additionally, we have formulated the SSA problem as a Markov decision process (MDP).
- We designed an adaptive reinforcement learning model capable of adjusting service allocation for varying numbers of users. This model outputs optimal service allocation schemes based on the specific user demands.
- Based on the proposed model, we generated several datasets and conducted model evaluation experiments as well as algorithm assessment experiments. These evaluations were aimed at assessing the performance of our proposed model and algorithm. The results indicate that our approach outperforms baseline methods.

The paper is organized as follows. Section 2 describes the system model of ISTN and introduces the satellite computing resource model. Section 3 formalizes the SSA problem as an MDP. Section 4 models our proposed DRL algorithm. Section 5 is the experimental part, which evaluates the performance of our work. Finally, Section 6 concludes this paper and looks ahead to future work.

2. System Model

2.1. Network Model

As illustrated in Figure 1, this paper considers a satellite–ground collaborative network model based on an LEO constellation. Satellites in this model allocate corresponding computational resources based on the terminal's business requirements and provide services to the terminals. The satellite–ground collaborative system under consideration comprises N LEO satellites and K users, where $\mathcal{N}$ and $\mathcal{K}$ represent the sets of LEO satellites and users, respectively. In this integrated system, in addition to providing traditional satellite communication (SATCOM) radio access services to UEs, the LEO satellites also provide edge computing services to users for processing edge data. For convenience, we assume that each satellite has a beam capable of serving users within its beam coverage area and that each user can only connect to one LEO satellite.

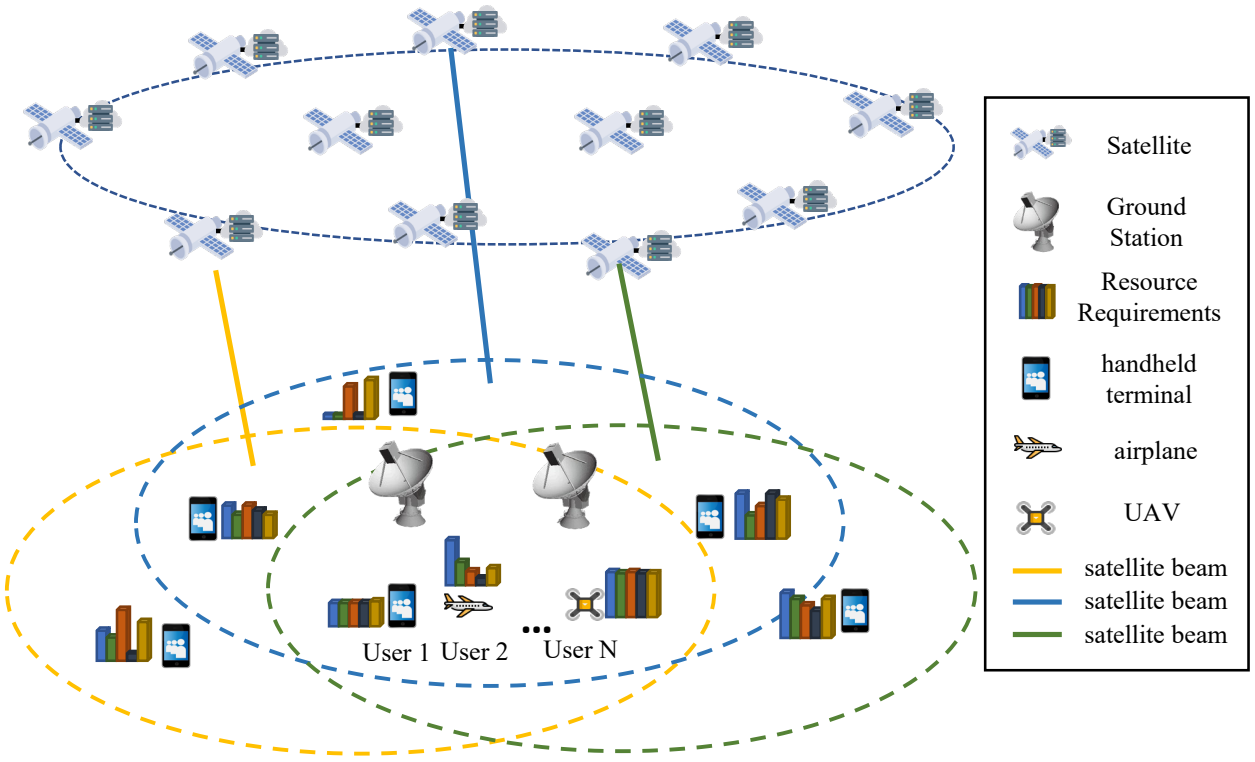


Figure 1. Network model.

We employ the homogeneous binomial point process (BPP) satellite distribution model as provided in [35] to configure the spatial distribution of the LEO satellites. As depicted in Figure 2, the BPP is considered a more suitable alternative to satellite networks compared to the Poisson point process (PPP) since it is better suited for modeling a finite number of points within a finite area. The subsequent proposition outlines the distribution of the homogeneous BPP on a sphere.

Proposition 1. *For a point in homogeneous BPP, the azimuth angle is uniformly distributed between* 0 *and* 2π*, i.e.,* $\phi_{BPP} \sim U[0, 2\pi]$ *and the cumulative distribution function (CDF) of each point's polar angle(of the spherical coordinate)* θ_{BPP} *as follows:*

$$F_{\theta_{BPP}}(\theta) = \frac{1 - \cos\theta}{2}, 0 \leq \theta_{BPP} \leq \pi, \tag{1}$$

θ_{BPP} *can be generated as follows:*

$$\theta_{BPP} = \arccos(1 - 2U[0,1]), 0 \leq \theta_{BPP} \leq \pi. \tag{2}$$

Note that the BPP given in subsequent parts of this paper refers to homogeneous BPP unless otherwise stated. The distribution of user terminals is modeled randomly.

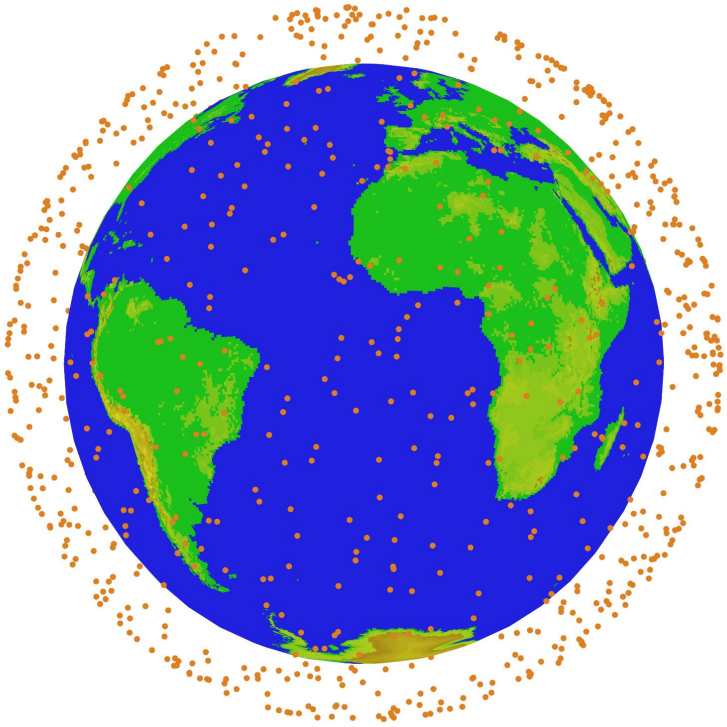


Figure 2. LEO constellation architecture.

In this work, we consider that each user can only run one service at the same time and that each service can only be provided by a single satellite. Therefore, each user can only access one satellite at the same time while the same satellite can serve multiple users simultaneously.

2.2. Resource Model

The transmission of operational data between the terminal and the satellite is conducted through a satellite-to-ground link, where each operation utilizes a fixed communication resource in the LEO. Introducing the variable $\delta(t) \triangleq [\delta_{n,k}(t)]_{\forall(n,k)}$ denotes the link connectivity status between the satellite and the user.

$$\delta_{n,k}(t) = \begin{cases} 1, UE_k \text{ is connected with } LEO_n. \\ 0, otherwise. \end{cases} \tag{3}$$

Due to the constraint that each user can only access a single satellite, the following restrictions are established:

$$\sum_{\forall n \in \mathcal{N}} \delta_{n,k}(t) \leq 1, \forall k \in \mathcal{K}. \tag{4}$$

Due to the increasing complexity of current service, the conventional approach of measuring node resources solely based on bandwidth and CPU frequency has become insufficient. In order to address this challenge, as shown in Figure 3, we have developed a five-dimensional resource model, considering the capabilities of communication bandwidth, CPU general computation, GPU mathematical computation, storage, and caching. A unified measurement model has been constructed to account for these dimensions. We denote communication resource quantity as bandwidth (BW), and for the standardization of diverse resources, we map the communication resource quantity to the interval $[0, \alpha]$ and represent it as B.

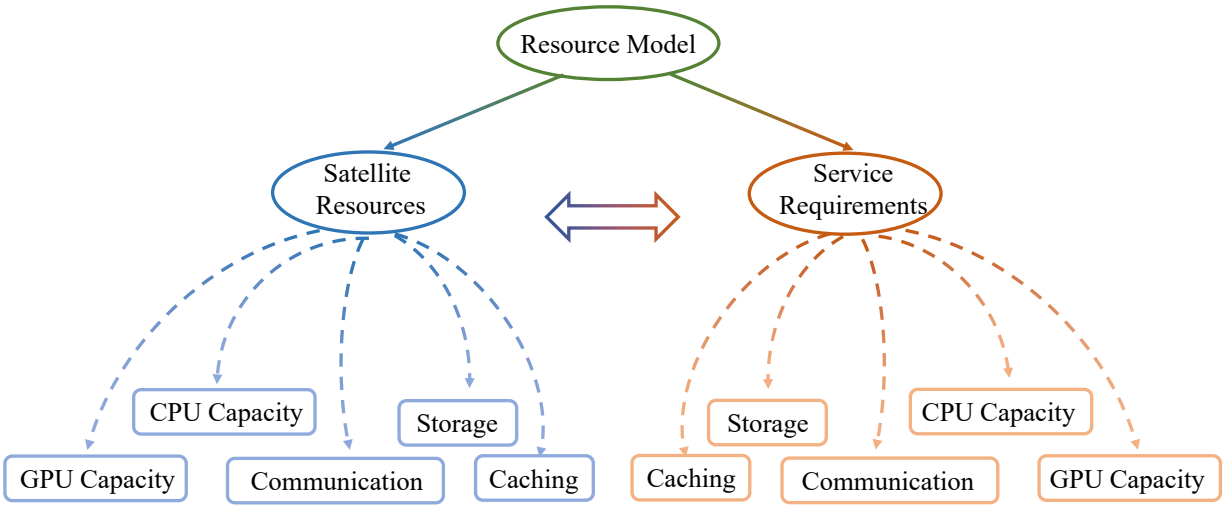


Figure 3. Resource and requirements model.

For convenience, we define the n-th LEO satellite and k-th user as Sat_n and UE_k, respectively. The channel gain between the LEO satellite and the user can be expressed as follows:

$$h_{n,k} = \frac{G^{Sat} G^{UE} \psi(\mu_{n,k})}{PL_{n,k}}, \tag{5}$$

where $PL_{n,k} = (4\pi f_c d_{n,k}/c)^2$ and $d_{n,k}$ represent the free-space path loss between the n-th LEO satellite and the k-th user; G^{Sat} and G^{UE} are the maximum antenna gains for the LEO satellite and user, respectively; and $\mu_{n,k}$ is the maximum boresight angle from Sat_n to UE_k. The $\psi(\mu)$ in the channel gain formula represents the beamforming pattern function and is expressed as follows:

$$\psi(\mu) = \begin{cases} 1, \mu = 0, \\ 4\left|\frac{J_1(wa \sin\mu)}{wa \sin\mu}\right|, \mu \neq 0, \end{cases} \tag{6}$$

where $w = 2\pi f_c/c$, a, f_c, and c are the antenna aperture radius, operation frequency, and light speed, respectively; and $J_1(\cdot)$ denotes the first-order Bessel function. Consequently, the available communication resources for a single satellite can be expressed as follows:

$$B_n^{Sat} = \frac{W_n^{Sat}}{W_{\max}^{Sat}} = \frac{W_n^{Sat}}{\max\{W_1^{Sat}, W_1^{Sat}, ..., W_N^{Sat}\}}. \tag{7}$$

The communication resource demand for a user is expressed as follows:

$$B_{n,k}^{UE} = \alpha \frac{R_{n,k}^{UE}}{W_{\max}^{Sat}}, \tag{8}$$

where

$$R_{n,k}^{UE} = W_k^{UE} \log_2 \left(1 + \frac{P_{n,k} h_{n,k}}{\sigma_n W_k^{UE}} \right), \tag{9}$$

where σ_n represents the noise power density per Hz. Therefore, the following can be obtained:

$$\sum_{\forall k \in \mathcal{K}} \delta_{n,k}(t) B_{n,k}^{UE} \leq B_n^{Sat}, \forall (n,t). \tag{10}$$

We employ a methodology similar to that used in constructing communication resource models to establish mathematical representations for the remaining four types of resources. For the sake of brevity, we refrain from elaborating on their actual mapping processes into mathematical forms. Ultimately, the mapping patterns for the representations of the other four resources also follow the method of referencing the maximum capacity of a single satellite resource, with all resources being mapped into specific ranges.

$$G_{n,k}^{UE} = \alpha \frac{G_{n,k}^{UE}}{G_{\max}^{Sat}}, \tag{11}$$

$$C_{n,k}^{UE} = \alpha \frac{C_{n,k}^{UE}}{C_{\max}^{Sat}}, \tag{12}$$

$$O_{n,k}^{UE} = \alpha \frac{O_{n,k}^{UE}}{O_{\max}^{Sat}}, \tag{13}$$

$$F_{n,k}^{UE} = \alpha \frac{F_{n,k}^{UE}}{F_{\max}^{Sat}}. \tag{14}$$

Therefore, the resource demand for a user can be represented as follows:

$$\mathcal{D} = \left\{ B^{UE}, G^{UE}, C^{UE}, O^{UE}, F^{UE} \right\}. \tag{15}$$

The resource capacity of a satellite can be expressed as follows:

$$\mathcal{V} = \left\{ B^{Sat}, G^{Sat}, C^{Sat}, O^{Sat}, F^{Sat} \right\}. \tag{16}$$

Similar to communication resources, these four types of resources also need to satisfy the constraint that the resource requirements of a single satellite user are less than the remaining available resources on the satellite. Therefore, it is necessary to satisfy the overall constraint:

$$\sum_{\forall k \in \mathcal{K}} \delta_{n,k}(t) \mathcal{D}_{n,k}^{UE} \prec \mathcal{V}_n^{Sat}, \forall (n,t). \tag{17}$$

3. Problem Statement

Due to the diverse requirements of edge services for various resources and the significant differences in the availability of resources on different satellites, there is an uneven distribution of resources in space, which also exhibit tidal-like variations over time, and an NP-hard problem arises in achieving the optimal match between services and resources. The SSA problem aims to assign each service to the most suitable satellite, thereby enhancing global resource utilization efficiency. This section analyzes the SSA problem, models its characteristics, and transforms it into a MDP, laying the groundwork for the introduction of a reinforcement learning solution. The definitions of relevant symbols are listed in the Table 1.

Table 1. Variable description.

Variable	Description
$\mathcal{N} = \{1, 2, ..., n, ...N\}$	Set of LEO satellites with available resources
$\mathcal{K} = \{1, 2, ..., k, ...K\}$	Set of users with resource requirements
$\mathcal{D} = \{B^{UE}, G^{UE}, C^{UE}, O^{UE}, F^{UE}\}$	Multi-dimensional resource requirements for user services
$\mathcal{V} = \{B^{Sat}, G^{Sat}, C^{Sat}, O^{Sat}, F^{Sat}\}$	Remaining available resources on LEO satellites
$Loc_k^{UE} = (x_k^{UE}, y_k^{UE}, z_k^{UE})$	The three-dimensional coordinates of users with service requests
$Loc_n^{Sat} = (x_n^{Sat}, y_n^{Sat}, y_n^{Sat})$	The three-dimensional coordinates of LEO satellites
d	Orbital altitude of LEO satellites
$\alpha_{n,k}$	The minimum communication elevation angle between UE_k and Sat_n
φ	Indicator for the existence of services on the satellite
$l_n^t =< Loc_n^{Sat}, V_n, \varphi >$	A struct used to describe the state of a single satellite
$e_k =< Loc_k^{UE}, D_n >$	Mathematical model of a UE
π_t	The satellite assigned to the UE_t in timeslot t; if unassigned, it is indicated as 0

For the SSA problem, we primarily focus on two constraints. The first one is that UE should be able to establish an effective communication link with LEO satellites:

$$\frac{\overrightarrow{Loc_{\pi_t}^{Sat}} \cdot \overrightarrow{Loc_k^{UE}}}{|Loc_{\pi_t}^{Sat}||Loc_k^{UE}|} \geqslant \alpha_{\pi_t,k}\ , \forall t\ and\ \pi_t \neq 0. \tag{18}$$

The second constraint is the resource constraint mentioned in formula [n], as follows:

$$\sum_{\forall k \in \mathcal{K}} \delta_{\pi_t,k}(t) \mathcal{D}_{\pi_t,k}^{UE} \prec \mathcal{V}_{\pi_t}^{Sat}, \forall t\ and\ \pi_t \neq 0. \tag{19}$$

Due to the presence of constrained multidimensional resources in the ISTN and varying resource utilization patterns across different services, we assume that to achieve desired service effects, each service has specific resource requirements. Satellite resources exceeding these requirements do not contribute to higher service quality for UE. Given the limited resources of satellites, there is competition among different services for the utilization of the same resources. The deployment of each service can be independently determined based on the current satellite's remaining resources. With a sufficiently fine-grained time granularity, in each timeslot t, only one service is allocated to an appropriate satellite, and the satellite resource status is updated. This process continues in the next timeslot, with suitable resources being allocated for the service. Inspired by [36–38], we propose the DRL-based approach in this paper.

In this scenario, akin to the edge user allocation (EUA) problem mentioned in [39], we can model the SSA problem as an MDP as shown in Figure 4. This includes the satellite network state S, satellite service allocation actions A, and the effectiveness of service allocation R. Therefore, the length of the Markov chain equals the number of users requesting services in each timeslot. S represents the satellite network state in each timeslot.

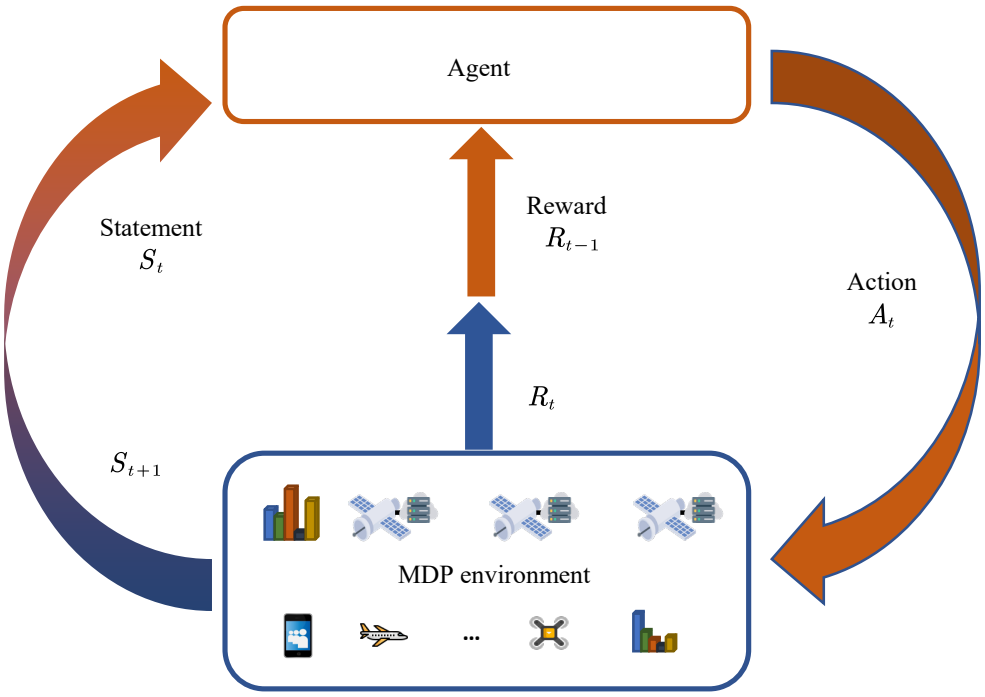


Figure 4. MDP schematic diagram.

$$S = \{s_1, s_2, ..., s_t, ..., s_N\}, \tag{20}$$

$$s_t =< e_t, l_{\mathcal{N}} > . \tag{21}$$

Each network state s_t can be jointly determined by the network state from the previous moment and the chosen action A. In our study, we set A as the satellite number n to which the UE is currently connected in the current state.

$$A = \{\pi_1, \pi_2, ..., \pi_t, ...\pi_n\}. \tag{22}$$

After the execution of action π_t in each timeslot t, the state S is updated based on the amount of resources occupied by the service. This involves updating the resource allocation of the satellite to which the service has been assigned. When a service cannot be allocated to any satellite, we consider the service as incomplete. In such cases, it will move on to the next timeslot, and there is no need to update the satellite resource status.

When the MDP concludes and when all services have been allocated, the reward calculation is based on the overall network resource allocation results. Typically, in the SSA problem, evaluation metrics for resource allocation results include resource utilization, the service allocation rate, and others. Our ultimate goal is to allocate suitable resources for as many services as possible while achieving a high level of resource utilization. Therefore, we define the reward as follows:

$$R = \lambda \times R^{UE}(A) + (1 - \lambda)R^{Sat}(A). \tag{23}$$

where R^{UE} represents the UE service allocation rate, and R^{Sat} represents the satellite resource utilization rate. λ is a predefined constant used to control the weighting coefficients of the two metrics. The calculation methods for R^{UE} and R^{Sat} can be expressed as follows:

$$R^{UE}(A) = \frac{1}{K}\sum_{t=1}^{K}\pi_t, \tag{24}$$

$$R^{Sat}(A) = avg\left\{ M \in \mathcal{D}, Z \in \mathcal{V} | \frac{\sum_{k \in \{ k | \pi_k \neq 0 \}} M_k^{UE}}{\sum_{n \in \mathcal{N}} Z_n^{Sat}} \right\}. \tag{25}$$

where M and Z refer to any element within the sets $\mathcal{D} = \{B^{UE}, G^{UE}, C^{UE}, O^{UE}, F^{UE}\}$ and $\mathcal{V} = \{B^{Sat}, G^{Sat}, C^{Sat}, O^{Sat}, F^{Sat}\}$, respectively. A higher R^{UE} value indicates that a greater number of user service requests have been satisfied. Furthermore, a higher R^{Sat} value signifies that satellite resources are being utilized more effectively.

Therefore, the final optimization problem can be formulated as follows:

$$\max avg\left\{ M \in \mathcal{D}, Z \in \mathcal{V} | \frac{\sum_{k \in \{ k | \pi_k \neq 0 \}} M_k^{UE}}{\sum_{n \in \mathcal{N}} Z_n^{Sat}} \right\} \tag{26}$$

$$s.t.C1 : \frac{\overrightarrow{Loc_{\pi_t}^{Sat}} \cdot \overrightarrow{Loc_k^{UE}}}{\left| Loc_{\pi_t}^{Sat} \right| \left| Loc_k^{UE} \right|} \geqslant \alpha_{\pi_t,k} \ , \forall t \ and \ \pi_t \neq 0 \tag{27}$$

$$s.t.C2 : \sum_{\forall k \in \mathcal{K}} \delta_{\pi_t,k}(t) D_{\pi_t,k}^{UE} \prec V_{\pi_t}^{Sat}, \forall t \ and \ \pi_t \neq 0 \tag{28}$$

4. Proposed Method

4.1. DRL-Based Method

The SSA problem is established in the context of the ISTN, where the number of satellites in the network tends to be relatively fixed, but the fluctuation in user numbers occurs on a smaller time scale. An effective solution should be capable of providing allocation schemes for varying user counts. In addressing this issue, we experimented with various traditional methods in this work, but their efficiency was relatively low. Therefore, considering the dynamic nature of user numbers, we explored the application of reinforcement learning to tackle this problem. For the MDP process constructed earlier, we devised a reinforcement learning approach based on small-sample training. This method can allocate satellite resources for users in networks of different scales, even larger than those used in training.

We use $\Omega(a|s) = P(A_t = a|S_t = s)$ to denote the policy of the agent, i.e., the probability of taking action a given the state s. $P_t^{\pi}(s)$ represents the probability of the agent being in state s at time t under the policy Ω. Therefore, the state visitation distribution for a policy is defined as follows:

$$v^{\Omega}(s) = (1 - \gamma) \sum_{t=0}^{\infty} \gamma^t P_t^{\pi}(s), \tag{29}$$

where $(1 - \gamma)$ serves as a normalization factor to ensure the total probability sums to 1. Due to the presence of actions, we define the action-value function as follows:

$$Q^{\pi}(s, a) = E_{\pi}[G_t | S_t = s, A_t = a]. \tag{30}$$

Furthermore, we define the state-value function, which is the sum of the product of all action probabilities and their respective values under the policy, as follows:

$$V^{\pi}(s) = \sum_{a \in A} \Omega(a|s) Q^{\pi}(s, a). \tag{31}$$

Therefore, when a fixed policy is employed, the value function for taking an action in a specific state is given by the following:

$$Q^{\pi}(s, a) = r(s, a) + \gamma \sum_{s' \in S} P\left(s'|s, a\right) V^{\pi}\left(s'\right). \tag{32}$$

We optimize L by gradient descent using the double DQN algorithm:

$$L = r + \gamma Q_{\omega^-}\left(s', arg \max_{a'} Q_{\omega}(s', a')\right). \tag{33}$$

The procedural flowchart of the methodology we employed is illustrated in Figure 5.

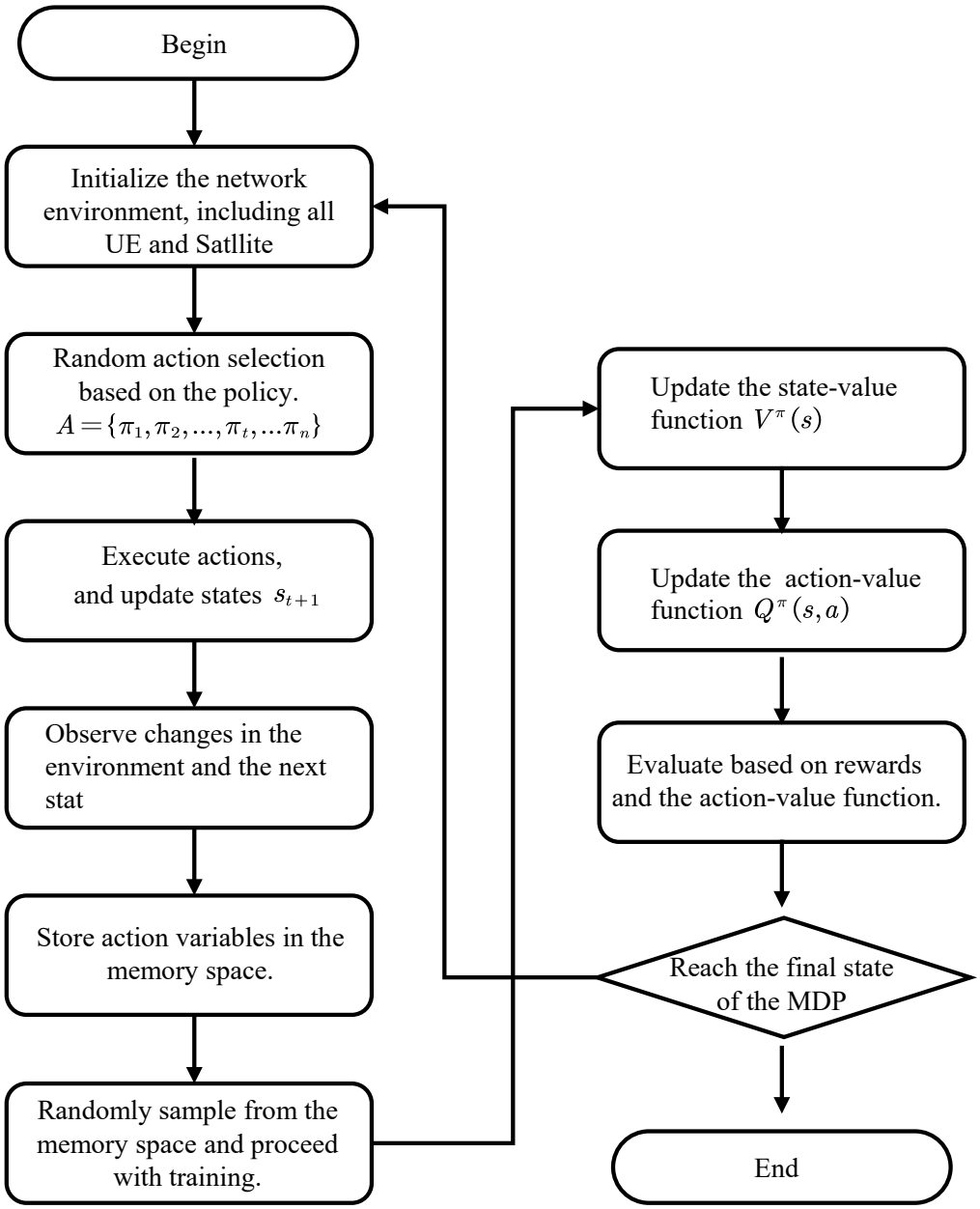


Figure 5. Proposed algorithm flowchart.

4.2. MCF-Based Method

As a benchmark for evaluating our method, we selected an advanced algorithm in the field of resource allocation, which is summarized in the flowchart below. This algorithm represents the state of the art in edge user service allocation algorithms known to date. In contrast to the algorithms described in the Introduction, this algorithm considers four types of resources and achieves high-level efficiency in user service allocation at a relatively low algorithmic complexity. The maximum capacity first (MCF) algorithm assesses users' computational demands, prioritizes them based on the assessment, and strives to allocate

users to servers with higher resource utilization [40]. This heuristic algorithm, based on a greedy approach, has demonstrated outstanding performance in ground networks, making it a suitable choice for comparison in our work.The scheme's concrete steps are depicted in Figure 6.

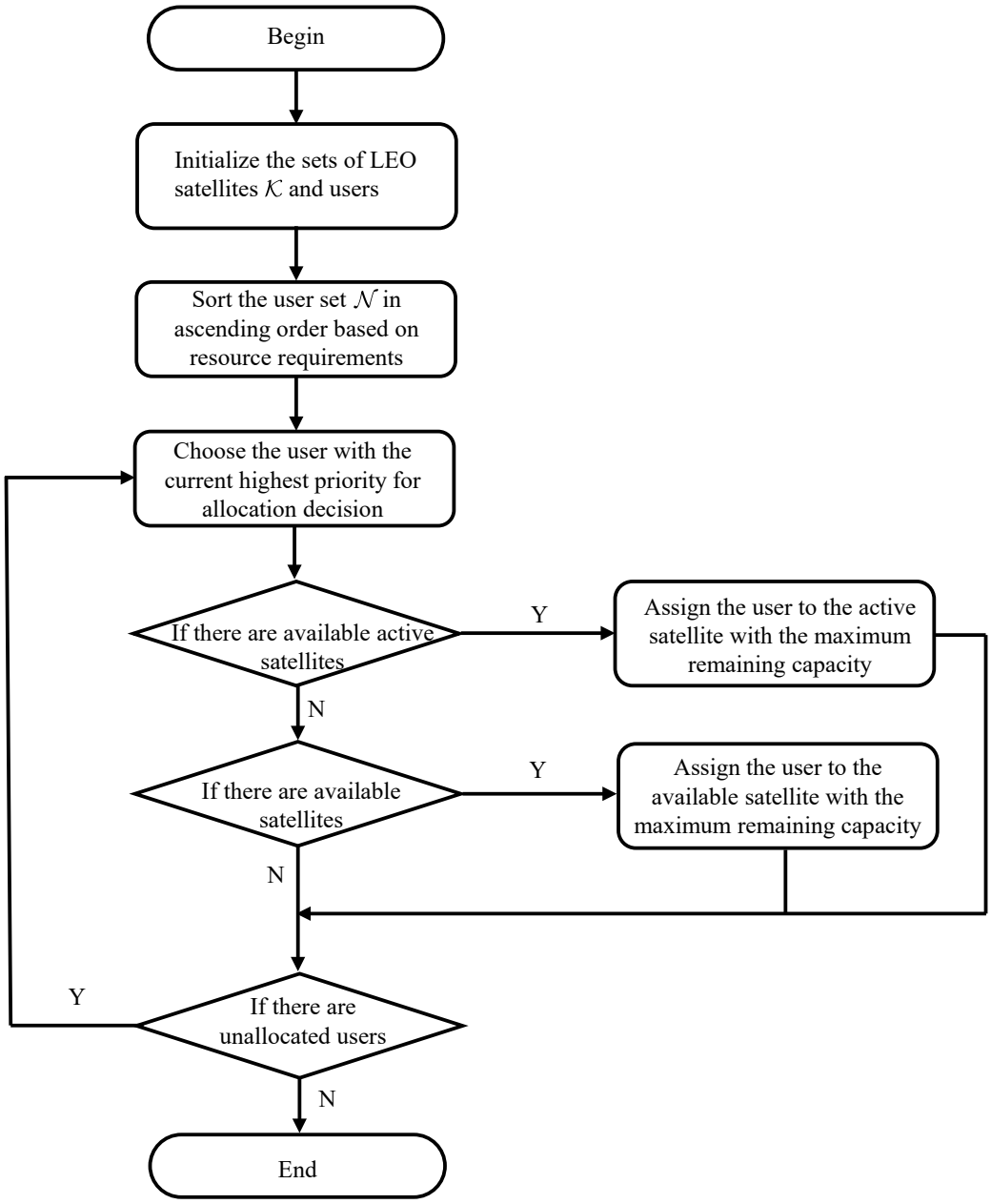


Figure 6. MCF flowchart.

5. Results

In this section, we describe the simulation experiments and evaluate the proposed algorithm based on numerical results. The parameters used during the simulation are summarized in Table 2. For the simulation, we established a three-dimensional coordinate system with the Earth's center as the origin, based on the structure depicted in Figure 2. We considered a LEO satellite constellation system, where the orbital altitude of the LEO satellites was set at 1000 km above the Earth's surface. The constellation comprised 66 LEO satellites, and each user in the system was guaranteed to be within the coverage range of at least one satellite.

To initially validate the effectiveness of the satellite resource model we constructed, we designed a common pre-experiment for simulation. Two approaches were employed, one utilizing the five-dimensional resource representation proposed in this paper and the other using the traditional representation considering only communication and computation dimensions. The simulation followed the service allocation pattern of traditional cloud data centers, where server resource amounts are fixed, and the total resource demand for all services equaled the total server resources. The maximum clique first (MCF) algorithm was used for simulation, and the results are shown in Figure 7.

Table 2. Variable Description.

Variable	Description
LEO satellite bandwidth, W_n^{Sat}	500 MHz
LEO satellite altitude, d	1000 km
LEO satellite antenna gain, G^{Sat}	40 dBi
UE antenna gain, G^{UE}	30 dBi
Number of UEs, K	100–1000
Number of sats, N	66
Mapping range, α	50
Minimum communication elevation angle β	60°

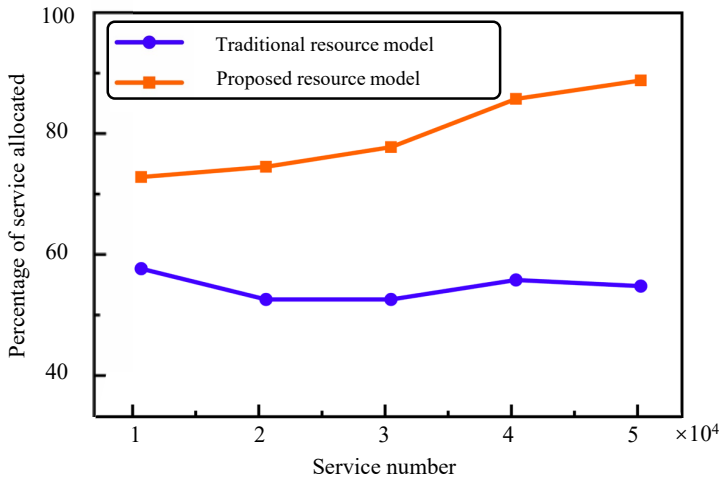


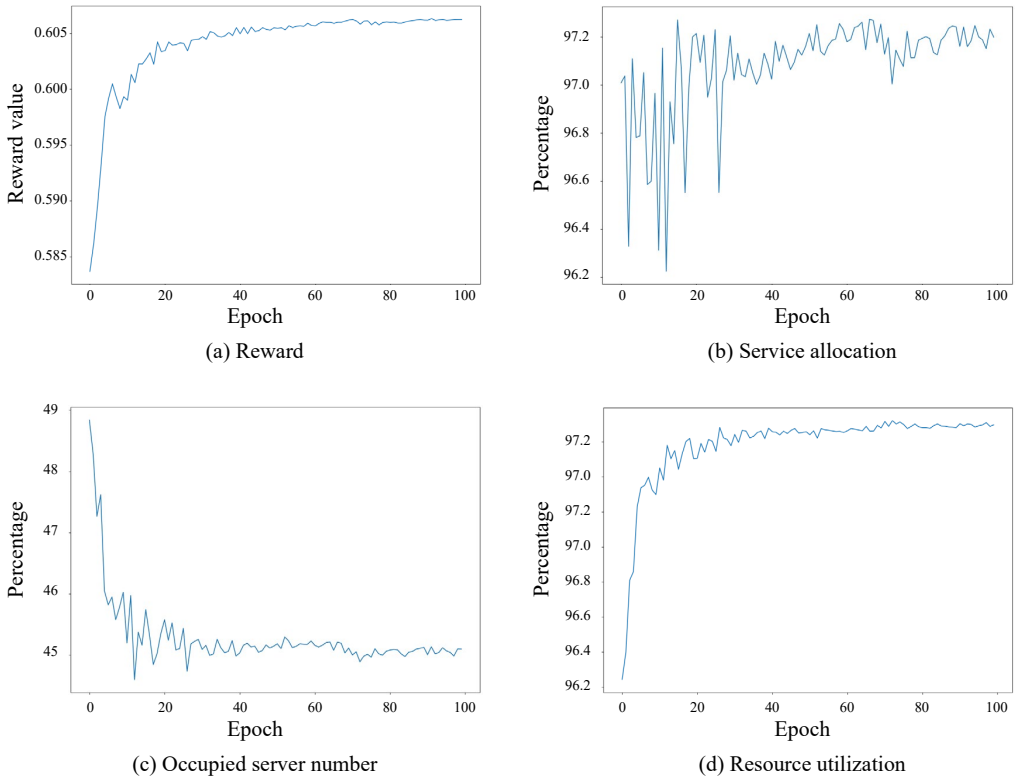
Figure 7. Service and resource match result.

The numerical results indicate that our proposed five-dimensional resource model outperforms the traditional two-dimensional resource representation in resource allocation service, and it exhibits superior performance in larger-scale networks.

Subsequently, we conducted simulations for the SSA problem based on the method proposed in this paper. We generated 10,000 sets of satellite and user data for the training the reinforcement learning model and 1000 sets for testing and validating the proposed method and the comparison algorithm. The data during the training process are illustrated in Figure 8, where "epoch" denotes the number of rounds during the algorithm training process. To enhance the algorithm's performance, we set the number of training epochs for the DRL algorithm to 100 epochs for each training.

From Figure 8, it can be observed that during the training process, the reward consistently increases and reaches convergence after 50 epochs. This indicates the reasonability of the reward setting and the successful construction of the model. Regarding the three metrics—user allocation rate, server occupancy, and resource utilization rate—a clear convergence can also be observed in the graph, demonstrating the effectiveness of the proposed method in addressing the SSA problem.

Subsequently, we conducted comparative experiments. In these experiments, we used a dataset with 100 users for training the model and datasets with user counts ranging from 100 to 1000 for validation.For a convincing simulation, we ensured that each user would only generate one service demand at any given moment. Additionally, we conducted simulation tests under scenarios where the volume of service demands continued to increase.



(a) Reward (b) Service allocation

(c) Occupied server number (d) Resource utilization

Figure 8. DRL model training process.

From Figure 9a, it can be observed that as the number of services demand increases, the proportion of successfully established services gradually decreases. In the range of 100 to 300 services, referred to as the "stable state", where satellite resources are relatively sufficient, the three methods exhibit similar success rates. As the number of services further increases, between 300 and 700, termed the "saturation state", the proposed method demonstrates significant superiority, with the ability to allocate an additional 5% of services. When the number of services exceeds 700, considered the "overload state" due to severe resource shortages, the three methods show similar performance. Overall, the proposed method consistently provides a higher service allocation rate relative to the comparison algorithms.

Figure 9b shows that in the stable state, the performance of the MCF algorithm is similar to the proposed algorithm, with both significantly outperforming the random algorithm. After entering the saturation state, the proposed algorithm gradually exhibits higher superiority, achieving approximately a 10% improvement compared to the MCF algorithm.

From Figure 9c, it can be observed that in both the stable and saturation states, the proposed algorithm and the MCF algorithm exhibit significant optimization in terms of server resource utilization. Upon entering the overload state, due to resource scarcity, none of the three algorithms can further increase resource utilization. Although the difference with the MCF algorithm is not substantial, the proposed algorithm consistently demonstrates performance superior to the MCF algorithm in this scenario.

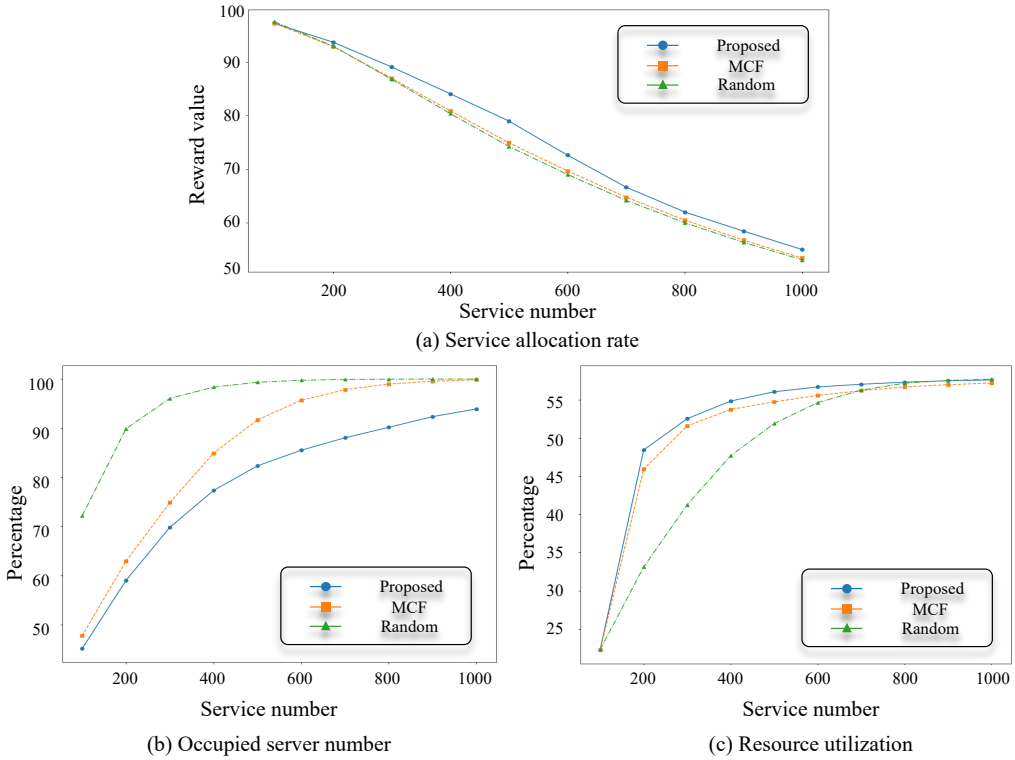


Figure 9. Experimental results for each algorithm.

6. Conclusions

In this paper, we discuss the SSA problem in the context of ISTN. In this scenario, LEO satellites provide services to UEs with diverse resource requirements, allocating resources based on service demands. Subsequently, we have formulated the optimal service allocation problem under the constraint of limited satellite resources, considering user allocation rate and satellite resource utilization rate. We model this problem as a variable-length MDP. To address this problem effectively, we propose an optimization algorithm based on reinforcement learning. Following the development of the algorithm, we conducted simulation experiments, using random allocation and the MCF algorithm as experimental benchmarks. According to the results, our proposed algorithm demonstrates rapid convergence starting from 50 epochs with outstanding performance. Moreover, compared to the baseline algorithm, our method exhibits significant advantages in several aspects including service allocation rate, occupied server number, and resource utilization, highlighting the superiority and effectiveness of the proposed approach.

Author Contributions: Conceptualization, Y.Z. and J.Z.; methodology, Y.Z. and J.Z.; software, Y.Z. and J.Z.; validation, Y.Z., J.Z. and Z.C.; formal analysis, Z.C. and X.W.; investigation, Z.C. and X.W.; resources, Z.C. and X.W.; data curation, Z.C. and X.W.; writing—original draft preparation, Y.Z. and J.Z.; writing—review and editing, Z.C. and X.W.; visualization, Y.Z. and J.Z.; supervision, Y.Z. and J.Z.; project administration, Y.Z.; funding acquisition, J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the BUPT Excellent Ph.D. Students Foundation under grant no. CX2023152 and BUPT innovation and entrepreneurship support program under grant no. 2024-YC-T001.

Data Availability Statement: The data presented in this study are contained within the article.

Acknowledgments: The authors acknowledge the contributions of School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, for supporting this work with the research facilities and resources.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

LEO	low Earth orbit
SSA	satellite service allocation
DRL	deep reinforcement learning
MDP	Markov decision process
SATCOM	satellite communication
BPP	binomial point process
PPP	Poisson point process
CDF	Cumulative distribution function
BW	bandwidth
EUA	edge user allocation
ISTN	integrated satellite–terrestrial network

References

1. Mahmood, N.H.; Alves, H.; López, O.A.; Shehab, M.; Moya Osorio, D.P.; Latva-Aho, M. Six Key Features of Machine Type Communication in 6G. In Proceedings of the 2020 2nd 6G Wireless Summit (6G SUMMIT), Levi, Finland, 17–20 March 2020; pp. 1–5.
2. Tang, J.; Bian, D.; Li, G.; Hu, J.; Cheng, J. Resource Allocation for LEO Beam-Hopping Satellites in a Spectrum Sharing Scenario. *IEEE Access* **2021**, *9*, 56468–56478. [CrossRef]
3. Xiong, X.; Zheng, K.; Lei, L.; Hou, L. Resource Allocation Based on Deep Reinforcement Learning in IoT Edge Computing. *IEEE J. Sel. Areas Commun.* **2020**, *38*, 1133–1146. [CrossRef]
4. Kodheli, O.; Maturo, N.; Chatzinotas, S.; Andrenacci, S.; Zimmer, F. NB-IoT via LEO Satellites: An Efficient Resource Allocation Strategy for Uplink Data Transmission. *IEEE Int. Things J.* **2022**, *9*, 5094–5107. [CrossRef]
5. Ivanov, A.; Stoliarenko, M.; Kruglik, S.; Novichkov, S.; Savinov, A. Dynamic Resource Allocation in LEO Satellite. In Proceedings of the 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC), Tangier, Morocco, 24–28 June 2019; pp. 930–935.
6. Zhou, J.; Sun, Z.; Zhang, R.; Lin, G.; Zhang, S.; Zhao, Y. A Cloud-Edge Collaboration CNN-Based Routing Method for ISAC in LEO Satellite Networks. In Proceedings of the 2nd Workshop on Integrated Sensing and Communications for Metaverse, ISACom '23, Helsinki, Finland, 18 June 2023; Association for Computing Machinery: New York, NY, USA, 2023; pp. 25–29.
7. Huang, L.; Feng, X.; Zhang, C.; Qian, L.; Wu, Y. Deep Reinforcement Learning-Based Joint Task Offloading and Bandwidth Allocation for Multi-User Mobile Edge Computing. *Digit. Commun. Netw.* **2019**, *5*, 10–17. [CrossRef]
8. Yuan, S.; Sun, Y.; Peng, M. Joint Beam Direction Control and Radio Resource Allocation in Dynamic Multi-beam LEO Satellite Networks. *IEEE Trans. Veh. Technol.* **2024**, 1–15. [CrossRef]
9. Sun, Y.; Chen, S.; Wang, Z.; Mao, S. A Joint Learning and Game-Theoretic Approach to Multi-Dimensional Resource Management in Fog Radio Access Networks. *IEEE Trans. Veh. Technol.* 2023, **72**, 2550–2563. [CrossRef]
10. Zhu, J.; Sun, Y.; Peng, M. Timing Advance Estimation in Low Earth Orbit Satellite Networks. *IEEE Trans. Veh. Technol.* **2024**, *73*, 4366–4382. [CrossRef]
11. Yuan, S.; Sun, Y.; Peng, M. Joint Network Function Placement and Routing Optimization in Dynamic Software-defined Satellite-Terrestrial Integrated Networks. *IEEE Trans. Wirel. Commun.* **2024**, *23*, 5172–5186. [CrossRef]
12. Xv, H.; Sun, Y.; Zhao, Y.; Peng, M.; Zhang, S. Joint Beam Scheduling and Beamforming Design for Cooperative Positioning in Multi-beam LEO Satellite Networks. *IEEE Trans. Veh. Technol.* **2023**, *73*, 5276–5287. [CrossRef]
13. Leng, T.; Li, X.; Hu, D.; Cui, G.; Wang, W.; Wen, M. Collaborative Computing and Resource Allocation for LEO Satellite-Assisted Internet of Things. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 4212548.
14. Zhao, J.; Chen, S.; Jin, C. Data scheduling and resource allocation in LEO satellite networks for IoT task offloading. *Wirel. Netw.* **2023**. [CrossRef]
15. He, Y.; Wang, Y.; Qiu, C.; Lin, Q.; Li, J.; Ming, Z. Blockchain-Based Edge Computing Resource Allocation in IoT: A Deep Reinforcement Learning Approach. *IEEE Int. Things J.* **2021**, *8*, 2226–2237. [CrossRef]
16. Chen, Z.; Lin, G.; Zhou, J.; Zhao, Y. Research on Satellite Routing Method Based on Q-Learning in Failure Scenarios. In Proceedings of the 2023 Chinese Intelligent Systems Conference, Ningbo, China, 14–15 October 2023; Lecture Notes in Electrical Engineering; Jia, Y., Zhang, W., Fu, Y., Wang, J., Eds.; Springer Nature: Singapore, 2023; pp. 433–445.
17. Huang, J.; Yang, Y.; Lee, J.; He, D.; Li, Y. Deep Reinforcement Learning Based Resource Allocation for RSMA in LEO Satellite-Terrestrial Networks. *IEEE Trans. Commun.* **2023**, *72*, 1341–1354. [CrossRef]
18. Baeza, V.M.; Ortiz, F.; Lagunas, E.; Abdu, T.S.; Chatzinotas, S. Gateway Station Geographical Planning for Emerging Non-Geostationary Satellites Constellations. *IEEE Netw.* **2023**, 1–1. [CrossRef]

19. Cheng, L.; Feng, G.; Sun, Y.; Liu, M.; Qin, S. Dynamic Computation Offloading in Satellite Edge Computing. In Proceedings of the ICC 2022—IEEE International Conference on Communications, Seoul, Republic of Korea, 16–20 May 2022; pp. 4721–4726.
20. Dai, C.-Q.; Luo, J.; Fu, S.; Wu, J.; Chen, Q. Dynamic User Association for Resilient Backhauling in Satellite–Terrestrial Integrated Networks. *IEEE Syst. J.* **2020**, *14*, 5025–5036. [CrossRef]
21. Feng, L.; Liu, Y.; Wu, L.; Zhang, Z.; Dang, J. A Satellite Handover Strategy Based on MIMO Technology in LEO Satellite Networks. *IEEE Commun. Lett.* **2020**, *24*, 1505–1509. [CrossRef]
22. Jia, M.; Zhang, L.; Wu, J.; Guo, Q.; Gu, X. Joint computing and communication resource allocation for edge computing towards Huge LEO networks. *China Commun.* **2022**, *19*, 73–84. [CrossRef]
23. Li, X.; Zhang, H.; Zhou, H.; Wang, N.; Long, K.; Al-Rubaye, S.; Karagiannidis, G.K. Multi-Agent DRL for Resource Allocation and Cache Design in Terrestrial-Satellite Networks. *IEEE Trans. Wirel. Commun.* **2023**, *22*, 5031–5042. [CrossRef]
24. Nguyen-Kha, H.; Ha, V.N.; Lagunas, E.; Chatzinotas, S.; Grotz, J. Two-Tier User Association and Resource Allocation Design for Integrated Satellite-Terrestrial Networks. In Proceedings of the 2023 IEEE International Conference on Communications Workshops (ICC Workshops), Rome, Italy, 28 May–1 June 2023; pp. 1234–1239.
25. Qiu, J.; Zhang, H.; Zhou, L.; Hu, P.; Wang, J. A Reinforcement Learning Based Resource Access Strategy for Satellite-Terrestrial Integrated Networks. In *Machine Learning and Intelligent Communication*; Jiang, X., Ed.; Springer Nature: Cham, Switzerland, 2023; pp. 97–107.
26. Song, Y.; Li, X.; Ji, H.; Zhang, H. Energy-Aware Task Offloading and Resource Allocation in the Intelligent LEO Satellite Network. In Proceedings of the 2022 IEEE 33rd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Kyoto, Japan, 12–15 September 2022; pp. 481–486.
27. Sun, W.; Cao, B. Efficient Transmission of Multi Satellites-Multi Terrestrial Nodes Under Large-Scale Deployment of LEO. In *Wireless Sensor Networks*; Hao, Z., Dang, X., Chen, H., Li, F., Eds.; Springer: Singapore, 2020; pp. 140–154.
28. Wang, B.; Feng, T.; Huang, D. A Joint Computation Offloading and Resource Allocation Strategy for LEO Satellite Edge Computing System. In Proceedings of the 2020 IEEE 20th International Conference on Communication Technology (ICCT), Nanning, China, 28–31 October 2020; pp. 649–655.
29. Wang, B.; Xie, J.; Huang, D.; Xie, X. A Computation Offloading Strategy for LEO Satellite Mobile Edge Computing System. In Proceedings of the 2022 14th International Conference on Communication Software and Networks (ICCSN), Chongqing, China, 10–12 June 2022; pp. 75–80.
30. Wang, R.; Zhu, W.; Liu, G.; Ma, R.; Zhang, D.; Mumtaz, S.; Cherkaoui, S. Collaborative Computation Offloading and Resource Allocation in Satellite Edge Computing. In Proceedings of the GLOBECOM 2022—2022 IEEE Global Communications Conference, Rio de Janeiro, Brazil, 4–8 December 2022; pp. 5625–5630.
31. Wang, Y.; Zhang, J.; Zhang, X.; Wang, P.; Liu, L. A Computation Offloading Strategy in Satellite Terrestrial Networks with Double Edge Computing. In Proceedings of the 2018 IEEE International Conference on Communication Systems (ICCS), Chengdu, China, 19–21 December 2018; pp. 450–455.
32. Wei, K.; Tang, Q.; Guo, J.; Zeng, M.; Fei, Z.; Cui, Q. Resource Scheduling and Offloading Strategy Based on LEO Satellite Edge Computing. In Proceedings of the 2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall), Norman, OK, USA, 27–30 September 2021; pp. 1–6.
33. Wu, Y.; Hu, G.; Jin, F.; Zu, J. A Satellite Handover Strategy Based on the Potential Game in LEO Satellite Networks. *IEEE Access* **2019**, *7*, 133641–133652. [CrossRef]
34. Zhang, M.; Wu, X.; Zhang, Z.; Liu, D.; Yin, F. User Selection and Resource Allocation for Satellite-Based Multi-Task Federated Learning System. In *Artificial Intelligence in China*; Liang, Q., Wang, W., Mu, J., Liu, X., Na, Z., Eds.; Springer Nature: Singapore, 2023; pp. 375–382.
35. Wang, R.; Kishk, M.A.; Alouini, M.-S. Evaluating the Accuracy of Stochastic Geometry Based Models for LEO Satellite Networks Analysis. *IEEE Commun. Lett.* **2022**, *26*, 2440–2444. [CrossRef]
36. Nguyen-Kha, H.; Ha, V.N.; Lagunas, E.; Chatzinotas, S.; Grotz, J. LEO-to-User Assignment and Resource Allocation for Uplink Transmit Power Minimization. In Proceedings of the WSA & SCC 2023: 26th International ITG Workshop on Smart Antennas and 13th Conference on Systems, Communications, and Coding, Braunschweig, Germany, 27 February 2023; pp. 1–6.
37. Van Chien, T.; Lagunas, E.; Ta, T.H.; Chatzinotas, S.; Ottersten, B. User Scheduling and Power Allocation for Precoded Multi-Beam High Throughput Satellite Systems with Individual Quality of Service Constraints. *arXiv* **2021**, arXiv:2110.02525.
38. Zhang, S.; Yan, S.; Wang, D.; Liu, X.; Peng, M. Multi-Service Oriented Multi-Dimensional Resource Requirement Conflicts Coordination in Radio Access Networks. In Proceedings of the ICC 2023—IEEE International Conference on Communications, Rome, Italy, 28 May–1 June 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 4810–4815.
39. Chang, J.; Wang, J.; Li, B.; Zhao, Y.; Li, D. Attention-Based Deep Reinforcement Learning for Edge User Allocation. *IEEE Trans. Netw. Serv. Manag.* **2023**, *21*, 590–604. [CrossRef]
40. Lai, P.; He, Q.; Grundy, J.; Chen, F.; Abdelrazek, M.; Hosking, J.; Yang, Y. Cost-Effective App User Allocation in an Edge Computing Environment. *IEEE Trans. Cloud Comput.* **2022**, *3*, 1701–1713. [CrossRef]