Hindawi
Mobile Information Systems
Volume 2020, Article ID 8868888, 10 pages
https://doi.org/10.1155/2020/8868888

Research Article

LEO Satellite Channel Allocation Scheme Based on Reinforcement Learning

Fei Zheng,[1,2] Zhao Pi ,[1] Zou Zhou ,[1] and Kaixuan Wang[3]

[1]*Ministry of Education Key Laboratory of Cognitive Radio and Information Processing, Guilin University of Electronic Technology, Guilin 541004, China*
[2]*State key Laboratory of Networking and Switching Technology (Beijing University of Posts and Telecommunications), Beijing 100876, China*
[3]*Department of Information, Shanxi University of Finance and Economics, Taiyuan 030006, China*

Correspondence should be addressed to Zou Zhou; zhouzou@guet.edu.cn

Received 15 August 2020; Revised 29 October 2020; Accepted 30 November 2020; Published 12 December 2020

Academic Editor: Chi-Hua Chen

Delay, cost, and loss are low in Low Earth Orbit (LEO) satellite networks, which play a pivotal role in channel allocation in global mobile communication system. Due to nonuniform distribution of users, the existing channel allocation schemes cannot adapt to load differences between beams. On the basis of the satellite resource pool, this paper proposes a network architecture of LEO satellite that utilizes a centralized resource pool and designs a combination allocation of fixed channel preallocation and dynamic channel scheduling. The dynamic channel scheduling can allocate or recycle free channels according to service requirements. The Q-Learning algorithm in reinforcement learning meets channel requirements between beams. Furthermore, the exponential gradient descent and information intensity updating accelerate the convergence speed of the Q-Learning algorithm. The simulation results show that the proposed scheme improves the system supply-demand ratio by 14%, compared with the fixed channel allocation (FCA) scheme and by 18%, compared with the Lagrange algorithm channel allocation (LACA) scheme. The results also demonstrate that our allocation scheme can exploit channel resources effectively.

1. Introduction

In recent years, with the development of wireless communication technology, the terrestrial cellular network is facing the explosive growth of data traffic. Although the terrestrial cellular network has the advantages of short delay and large bandwidth, it still has some limitations [1]. Due to the limit of geographical environment and economy, it is difficult for cellular networks to cover special areas such as oceans, deserts, forests, and islands. Ocean navigation, geological exploration, environmental emergency rescue, and other scenarios rescue require an all-weather, wide-coverage, highly reliable communication mode. Satellite communication can solve the above problems well by virtue of wide coverage, small geographic limitation, and large system capacity.

The satellite communication system has experienced the development of global beams, regional beams, and spot beams. Flexible resource allocation between spot beams can further improve system performance. Among various types of satellites, LEO satellites have the characteristics of low path loss, short communication delay, and flexible orbital position [2]. LEO constellations can achieve seamless coverage of global regions [3, 4]. With the development of satellite communication technology, intersatellite link (ISL) and on-board processing (OBP) can support satellite routing and data processing independently from the terrestrial network. The maturity and reliability of IP technology also make the application of IP technology in satellite networks become the trend in the future [5]. In addition to communications payloads, the satellites carry automated directed surveillance by broadcast (ADS-B) payloads, which are primarily used for aircraft flight surveillance and tracking aircraft position reports, as well as navigation augmentation payloads. Ground-based gateway stations are also capable of communicating with multiple satellites simultaneously, synthesizing data streams from different satellites [6]. The cost reduction of satellite manufacture and launch has also promoted the rapid development of the LEO satellite Internet industry. LEO satellite network is becoming an important part of the future global mobile communication system.

Satellite communication systems are typical resource-constrained systems. Available spectrum, power, time slot, and other resources are extremely scarce and precious [7]. An efficient network resource allocation scheme is urgently needed to solve the above problem in the satellite communication systems. Due to the dynamic coverage change caused by satellite movement and nonuniform distribution of ground users, the traffic load is changing all the time, while satellite on-board resources are solidified at the factory setting. Traditional fixed channel allocation (FCA) scheme is difficult to adapt to rapidly changing business requests. Dynamic channel allocation (DCA) can realize resource cross-beam scheduling and has a higher resource utilization rate than FCA [8]. The business request is a discrete dynamic process in communication networks, and the allocation results at the current time will affect the decision at a subsequent time. The existing dynamic channel allocation algorithms focus on the instantaneous performance of the LEO satellite system and ignore the time-domain relevance problem in the channel allocation process [7].

Reinforcement Learning (RL), as an emerging technology, provides a new solution to solve complex decision-making problems [9]. Under the background of rapidly growing data and complex system structure, RL can better adapt to complex decision-making problems, which are difficult for traditional algorithms. By combining satellite resource allocation with RL, the decision-making ability of the satellite system can be well enhanced [7].

This paper considers the difference in service distribution and the time correlation of channel allocation in satellite communication systems. The Q-Learning algorithm is used for dynamic channel allocation in a LEO satellite. The main contributions are as follows:

(i) The on-board resource pool is introduced to manage channel resources in the LEO satellite network. The resource pool integrates information processing, resource allocation and resource acquisition, enabling cross-beam scheduling of channels. So that the system can better adapt to business differences between beams.

(ii) A two-step allocation scheme combining fixed channel preallocation and dynamic channel scheduling is proposed to schedule the channel. The system preallocates some fixed channels for each beam cell before services arriving; dynamic channel allocation schedules channel according to the services request.

(iii) RL improves the decision-making ability of the system on resource allocation. The problem is described as a Markov decision process with defining state space, action space, and reward function. The system trains the optimal channel allocation strategy through a Q-learning algorithm for channel resource allocation.

(iv) Exponential gradient descent and information intensity updating accelerate the convergence of the algorithm and improve the decision-making speed of the LEO satellite system.

The rest of this paper is organized as follows. In Section 2, we give related works. In Section 3, we describe the architecture of LEO satellite network based on on-board resource pool and establish the channel allocation model and problem optimization strategy. We give the specific content of the algorithm and distribution process in Section 4. In Section 5, we present and discuss simulation results. Finally, conclusions are presented in Section 6.

2. Related Work

In this section, we introduce some related works about LEO satellite networks and satellite resource allocation.

2.1. LEO Satellite Communication System. The size of the LEO satellite constellation is becoming larger and larger due to the advantages of technology and cost. A large-scale constellation can better achieve global coverage and greatly expand the system capacity [10]. In highly complex and frequently changing systems, it is critical to consider the load on the underlying network components due to user behavior. The massive traffic loads also challenge the quality of service (QoS) of LEO Satellite communication systems [11]. The satellite system is different from the terrestrial network, so researchers adopt some special frames and protocols according to the particularity of satellite systems, including data relay satellite (DRS) system, delay-tolerant network (DTN), and performance enhancement system (PES). However, these satellite communication protocols based on TCP/IP have poor mobility, high overhead, and high complexity [4]. Further, most of the existing satellite network protocols are only applicable to medium Earth orbit (MEO) geosynchronous Earth orbit (GEO) satellites. Therefore, network architecture and resource management system are particularly important for LEO satellites.

2.2. LEO Satellite Network Architecture. Recently, the construction of commercial LEO satellite systems is active all over the world, but it is hard to avoid some challenges in the network architecture and resource management. The architecture of the O3b system in the MEO satellite network and the OneWeb system adopts a transparent forwarding mechanism. These two systems have no interstar networking, outing, and switching function, and the system resource utilization is low when business is highly dynamic [12]. The architecture of Iridium and SpaceX relies on ISL to achieve intersatellite networking, but their networking technologies are relatively backward, the control plane and forwarding plane are highly coupled, and the resource scheduling mechanism requires more human intervention, which all reduce the resources utilization efficiency [13]. To solve the above problems, researchers have made a lot of efforts on LEO satellite network architecture and corresponding resource allocation scheme.

As a resource management unit that is widely used in the terrestrial wireless network, a resource pool can realize resource sharing and dynamic scheduling according to service requirements and improve spectrum efficiency. However, current works mainly focus on the resource pool architecture design of earth-gate-station (EGS) or the centralized management of satellite network virtualization, rather than satellite resource pool. Reference [14] proposes a design scheme of EGS based on resource pool architecture. By integrating digitizing, the resource pool can achieve signal processing and baseband processing functions, the utilization of high-speed data communication resources can be effectively improved in satellite networks. In view of the problems existing in the "chimney" architecture of EGS, [15, 16] propose architectures based on resource pool to solve the instability of EGS systems. The researchers compare the two architectures with and without resource pooling and found that the resource pooled system architecture is more reliable while improving the efficiency and flexibility of device resource use. Reference [17] presently analyses the contradiction between resource constraint and business demand in satellite networks and proposes the concept of "on-board resource virtualization". Further, researchers construct a mission-oriented satellite network resource management model and conduct on-board resource allocation by means of resource sharing and collaborative management. At the present stage, satellite communications are creating suitable operational control systems for different functions and different series of satellites in order to achieve efficient utilization of resources [18].

2.3. LEO Satellite Resource Allocation Scheme. The satellite resource allocation scheme will directly affect the user's QoS and system performance. Reference [19] considers the trade-off between the maximum total system capacity and interbeam fairness to obtain the optimal allocation scheme by a subgradient algorithm. Reference [20] optimizes the allocation strategy by calculating and comparing user transmission rates under different transmission modes and strong interference. Reference [21] explains the physical layer structure of a multibeam satellite system, simplifies the three-dimensional coordinate system of the ground user to the two-dimensional coordinate system in the equatorial plane. Further, researchers calculate the maximum channel capacity according to the satellite beam coverage area and transmission power. Reference [22] proposes a beam-hopping algorithm, which adjusts the beam size according to the business distribution. Reference [23] uses a heuristic algorithm to achieve frequency band selection and beam allocation and adopts Lagrangian dual algorithm and water-filling-assisted Lagrangian dual algorithm to achieve power allocation. Reference [24] proposes a channel allocation scheme of mixed random access and on-demand access, which reduces system delay within the throughput threshold. This scheme provides effective solutions for services with different delay sensitivities. The above satellite resource allocation scheme improves the system performance in some aspects. However, they only focus on the instantaneous performance of the system and ignore the time correlation in the resource allocation process. The allocation result of the previous time will indirectly lead to the subsequent allocation effect, which will undoubtedly affect the system resource utilization.

The satellite channel allocation can be regarded as a sequential decision problem, and a decision is made on the arriving user request within each interval T. RL is a good way to adapt to this decision-making problem. References [25, 26] uses augmentation learning to solve channel allocation and congestion control in satellite Internet of things (SIoT). Compared with traditional algorithms, RL can improve performance in terms of energy consumption and blocking rate. Reference [27] extends single-agent deep reinforcement learning (DRL) to multiagents and propose a collaborative multiagent DRL method so as to improve transmission efficiency and achieve the desired goal with lower complexity. Reference [28] discusses a scheme of combining RL and resource allocation in different heterogeneous satellites and multiple service requirements and demonstrates the application effect of DRL in heterogeneous satellite networks (HSN). However, there are few researches on LEO satellite resource allocation. Most of the research has focused on MEO and GEO satellites. Therefore, this paper applies RL to the LEO satellite resource allocation. We adopt emerging technologies to solve LEO satellite channel allocation challenges in a different way.

3. System Model

In this section, we propose a LEO satellite network architecture based on an on-board resource pool and explain the centralized resource allocation in detail. Further, we establish an optimization model based on the user supply-demand ratio.

3.1. Framework of LEO Satellite Network. Figure 1 shows a LEO satellite network architecture. In the network layer, adjacent satellites transmit data through ISLs, and multiple satellites cooperate to complete the global coverage. The centralized resource pool can manage the channel, computing, caching, and other resources. In the link layer, the network control center (NCC) provides services for users by uploading data from satellites. Edge cloud computing devices are connected through multiple satellite relays. Idle computing resources on the constellation network can also be used as edge cloud devices. LEO constellation is composed of numerous LEO satellites, which can provide services for users in cities, suburbs, and oceans in the global region.

Figure 2 shows the structure of a centralized resource pool in a LEO satellite. Each centralized resource pool is the

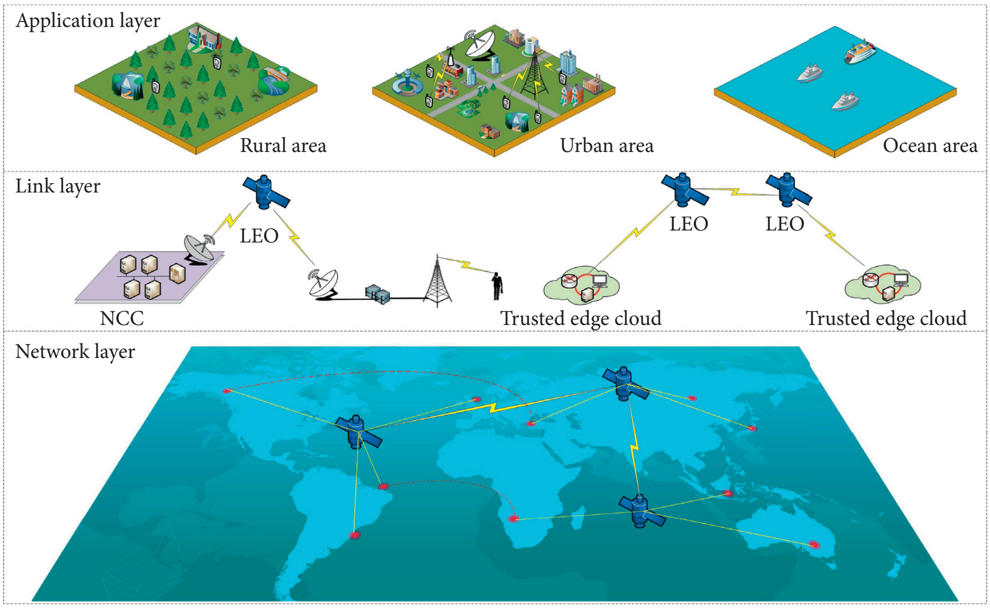

Figure 1: LEO satellite network architecture.

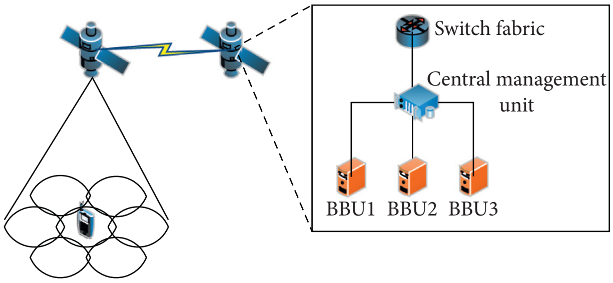

Figure 2: Structure of on-board resource pool.

core of the whole system, which integrates information processing, resource allocation, and resource collection. Resources between the satellites are connected through switch fabrics and resources are allocated in real time. A central management unit centrally manages BBU under the switching structure. For a single satellite, a centralized resource pool composed of high-performance processors can process services of all beams within its coverage, as shown in Figure 3. Compared with traditional dynamic resource allocation, the satellite with centralized resource pool can achieve resource allocation cross beams. The centralized resource pool not only processes and allocates resources for user's requests, but also schedules resource according to the utilization of resources in each beam to adapt to the business differences.

3.2. Channel Allocation Modelling. A LEO satellite has N beams on the ground through phased array antennas, represented by a set $X = \{x_n | n = 1, 2, \ldots, N\}$. The system available channels are represented by a set $Y = \{y_m | m = 1, 2, \ldots, M\}$, and system total bandwidth is B_{tot}. Users in beam x_n can be represented by a set $U = \left\{u_{n,k} | n = 1, 2, \ldots, N,\ k = 1, 2, \ldots, K\right\}$.

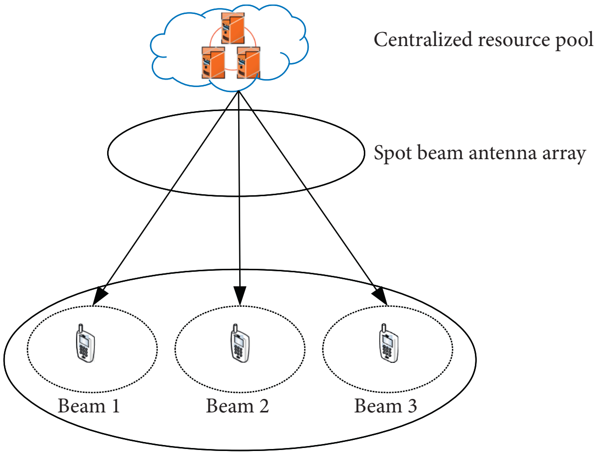

Figure 3: Channel allocation mapping under a single satellite.

The system allocates channels by frequency multiplexing between beams. Furthermore, channel and power allocation matrix are defined as follows:

$$
\begin{aligned}
V &= \left[V_1,\ V_2,\ \ldots\ V_N\right]^T = \begin{bmatrix} v_{1,1} & \cdots & v_{1,M} \\ \vdots & v_{n,m} & \vdots \\ v_{N,1} & \cdots & v_{N,M} \end{bmatrix}, \\
P &= \left[P_1,\ P_2,\ \ldots\ P_N\right]^T = \begin{bmatrix} p_{1,1} & \cdots & p_{1,M} \\ \vdots & p_{n,m} & \vdots \\ p_{N,1} & \cdots & p_{N,M} \end{bmatrix},
\end{aligned}
\tag{1}
$$

where $v_{n,m} \in \{0, 1\}$ in matrix V, $v_{n,m} = 1$ represents the channel y_m is used in the beam x_n, otherwise is not. The maximum transmits power of a beam and a system are P_c and P_{tot}, respectively. The channel gain of each beam can be expressed by a gain matrix

$$H = \left[H_1,\ H_2,\ \dots\ H_N \right]^T = \begin{bmatrix} h_{1,1} & \cdots & h_{1,M} \\ \vdots & h_{n,m} & \vdots \\ h_{N,1} & \cdots & h_{N,M} \end{bmatrix}. \tag{2}$$

For a user $u_{n,k}$ in the beam x_n, the useful signal and cofrequency interference received in the channel y_m are as follows:

$$S_{n,m}^{k} = p_{n,m} h_{n,m}, \tag{3}$$

$$R_{n,m}^{k} = \sum_{i=1, i \neq n}^{N} p_{i,m} h_{i,m}. \tag{4}$$

The SINR of $u_{n,k}$ can be calculated by equations (3) and (4); further, the channel rate of $u_{n,k}$ in the channel y_m can be calculated by the following equation:

$$C_{n,m}^{k} = B \log_2\left(1 + \frac{S}{N}\right) = B_{n,k} \log_2\left(1 + \frac{p_{n,m} h_{n,m}}{\sum_{i=1, i \neq n}^{N} p_{i,m} h_{i,m} + n_0 B_{n,k}}\right), \tag{5}$$

where n_0 is the noise power spectral density, $B_{n,k}$ is the bandwidth allocated to the user $u_{n,k}$. To evaluate system performance, a user supply-demand ratio is defined as follows:

$$\eta_n^k = \frac{C_{n,m}^{k}}{C_{n,m}^{k'}}. \tag{6}$$

In equation (6), $C_{n,m}^{k'}$ is user's request rate. Satellite channel allocation can be seen as a sequence decision-making problem in an interval T. Our optimization goal is to maximize the user supply-demand ratio under limited channel resources. Therefore, channel allocation is expressed as the following optimization:

$$\max \sum_{n=1}^{N} \sum_{k=1}^{K} \eta_n^k, \text{ s.t. } \begin{cases} \sum_{n=1}^{N} \sum_{k=1}^{K} C_{n,m}^{k} \leq C_{\text{tot}}, \\ \sum_{n=1}^{N} \sum_{m=1}^{M} p_{n,m} \leq P_{\text{tot}}, \\ \sum_{m=1}^{M} p_{n,m} \leq P_c. \end{cases} \tag{7}$$

The optimization objective in (7) is to maximize the supply-demand ratio in the system. The constraints indicate that the sum of user service rate must not exceed system capacity, the sum of channel transmit power must not exceed total transmit power limit, and the sum of channel transmit power within a single beam must not exceed the power limit of a single beam.

4. Channel Allocation Scheme

The purpose of RL is to improve the decision-making ability of the LEO satellite system in the process of channel allocation so as to improve resource utilization further. In this section, we define the state space, the action space, and the reward function of the Q-learning algorithm and adopt Q-learning algorithm to train the optimal channel allocation strategy.

Figure 4 shows the interaction process between a satellite system and the environment. The environment is the collection of terrestrial users in the satellite system, and the state is the channel allocation state of the system user. Furtherly, the action is the system assigning channels to users. We model the channel allocation of a satellite system as a Markov decision process (MDP). MDP is a set of sequential decision processes with Markov attributes. MDP contains a set of state $s_t \in s$, action $a_t \in A(s)$, reward $r_t \in R$, and state transition probability $p(s_{t+1}|s_t, a_t)$. The state transition probability $p(s_{t+1}|s_t, a_t)$ refers to the probability of environment transition to a new state s_{t+1} after performing an action a_t under state s_t. The goal of MDP is to specify a policy that maximizes the agent's reward from the environment. We use a model-free method in this paper, which does not need to model the state transition probability. According to the established optimization problem, we define the states, actions, and reward.

4.1. State Definition. The state matrix is constructed according to the channel assignment of users in each beam.

$$W = \begin{bmatrix} w_{1,1} & \cdots & w_{1,K} \\ \vdots & w_{n,k} & \vdots \\ w_{N,1} & \cdots & w_{N,K} \end{bmatrix}, \tag{8}$$

where $w_{n,k} \in \{-1, 0, 1\}$, $w_{n,k} = 0$ represents no user, $w_{n,k} = -1$ represents a user allocated no channel, and $w_{n,k} = 1$ represents a user allocated channels. The number of matrix columns is the maximum number of users in all beams, and the rows are system beams. When all requesting users have been allocated channels or the system has no available channels, the training process reaches the termination status and the allocation process ends.

4.2. Action Definition. The system selects suitable channels from the action set $A(s)$ and allocates these channels to users according to the current state. Channel assignment is defined as action a_t:

$$a_t = \{m | m \in A(s), A(s) \subseteq Y\}. \tag{9}$$

The agent randomly selects actions from the action set $A(s)$ with probability ε. Also the agent selects the action with maximum Q value with probability $1 - \varepsilon$. When the training steps are enough, the action value of each state in Q table will converge to the optimal value.

4.3. Reward Definition. Reward is the feedback from environments to agent after agent acts according to the current state, and it can be used to measure the performance of actions. An appropriate reward setting can guide an agent to train the optimal strategy better. The goal in optimization (7) is to maximize the system supply-demand ratio. Thus we set

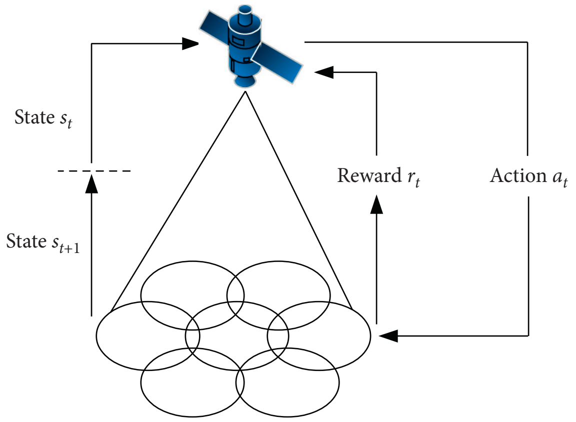

Figure 4: Dynamic channel allocation model based on RL.

the reward function as a function that is positively related to the supply-demand ratio.

$$r_t = 100 \times \sum_{n,k} \eta_n^k. \tag{10}$$

4.4. Algorithm Optimization. In order to accelerate the convergence of the *Q*-Learning algorithm, we make two improvements based on the original *Q*-Learning algorithm: the exponential gradient descent and the information intensity updating strategy.

The exponential gradient descent is that the random exploration probability ε decreases exponentially with the increase of training steps in the course of action selection, as shown in the following equation:

$$\varepsilon = \varepsilon_0 \cdot e^{-\frac{l}{l_0}}, \tag{11}$$

where l_0 is the maximum number of training steps. Exploring with a greater probability can ensure the diversification of action selection in the early training stage, and thus can avoid falling into local optimum; with the training step increasing, the exploration probability begins to decrease, and selecting optimal action with a larger greedy probability can accelerate the convergence of the algorithm.

The information intensity updating strategy is to define the information intensity to express the quality of the action and to update the *Q* table by information intensity. The information intensity $J(s_t, a)$ is defined in equation (12). It reflects the quality of action in the current state, where Δ is 1 by default. The *Q* table will update only when the reward is greater than the maximum reward in the current state. Further, *Q* table updates as shown in the heuristic function are defined in the following equation:

$$J(s_t, a) = \begin{cases} J(s_{t-1}, a)\dfrac{r_{\max}}{r_t}, & a \neq a_t, \\ \Delta, & \text{else}, \end{cases} \tag{12}$$

$$H(s_t, a) = \begin{cases} \max\limits_a [Q(s_t, a) - Q(s_t, a_t)] + \dfrac{J(s_t, a_t)}{\sum_i J(s_t, a_i)}, & a_t \text{ is the best action}, \\ 0, & \text{else}. \end{cases} \tag{13}$$

Under the guidance of information intensity, the heuristic function updates the optimal behavior. Through iterative accumulation, the agent will train the state-action decision plan with the largest reward.

4.5. Trade-Off Analysis. Firstly, in order to simplify the allocation process, we assume that the transmit power and SNR of each channel are the same. Then, the training time of the *Q*-DCA is highly dependent on the number of states and actions. The number of states and actions largely determines the quality of the final allocation scheme. Due to the strict latency requirements of satellite communication systems, we reduce the number of states and actions appropriately to shorten the training time of the algorithm.

4.6. Allocation Process. The allocation scheme has two steps: fixed channel preallocation and dynamic channel allocation on demands. Before each service request arrives, the system first preallocates some fixed channels for each beam cell; after fixed preallocation is finished, if channel resource cannot meet user's demands in some beams, the resource pool will perform dynamic channel allocation. Table 1 shows the process of satellite system channel allocation.

5. Simulation Results and Discussions

In order to verify the performance of the proposed dynamic channel allocation scheme, we carry out simulation experiments on the MATLAB platform and compare the proposed scheme with the FCA scheme and LACA scheme.

The system receives the user's request in each beam at each service interval (the service arrival model system is subject to the Poisson distribution with parameter λ, the service duration is subject to the negative exponential distribution with parameter μ_1, and the bandwidth request is subject to the normal distribution with parameters μ_2, σ^2). After centralized statistics of requests, the system allocates

Table 1: Allocation process.

Initialize system parameters	
1	Preallocation: Assign M channel to each beam
2	**for** Business request time $t = 1 : T$
3	**if** Resource is rich; recycle surplus resources
4	**else** resource is poor:Dynamic allocation
5	Allocate resources from resource pool
6	initialize parameter, learning rate α discount factor γ, initial explore probability ε_0, Q table
7	Reconstruct state based on business request $s = I$
8	**for** Episode = 1:max_episode
9	**while** (s_{t+1} is terminal state)
10	Confirm initial state s_t
11	Update explore probability ε
12	Choose best a_t^* or Choose randomly a_t
13	Execute action, get reward r_t
14	Update Q table
15	Jump to next state s_{t+1}
16	**End**
17	End of training, output Q table
18	Choose best strategy according to Q table π^*
19	Channel allocation
20	**End**
21	**End**

channel resources to each user, counting supply-demand ratios and blocking rates. Table 2 shows the specific parameters of the satellite system.

Compared with the proposed algorithm in simulation, the FCA scheme adopts the average allocation. The bandwidth resources are evenly distributed to all users. The LACA scheme adopts the minimum variance of supply and demand (MDSV), comparing the performance of three schemes in different scenarios.

5.1. The System Performance in Beam Number Variation Scenario. In this scenario, all beam traffic distribution parameters are the same. The number of beams increases from 10 to 50, simulating that the numbers of accessing users increase gradually and the available resources transition from rich to scarce. Figures 5 and 6 show the system performance of three schemes in the scenario of a gradual increase in the number of beams.

As shown in Figure 6, with the increase in the number of beams, the system blocking rate also increases. The reason is that with the expansion of the beam range, more users are connected to the current satellite communication system, and the bandwidth resources allocated to each user are also reduced. When the number of beams is increased to 16, the system starts to overload and block; meanwhile, the proposed Q-DCA scheme can further improve the system supply-demand ratio compared with the FCA scheme and LACA scheme. For example, when the number of beams reaches 20, the system supply-demand ratio of the three schemes are 0.725, 0.645, and 0.615 respectively, which means that the performance of the proposed Q-DCA algorithm is 12% and 18% better than FCA scheme and LACA scheme.

We analyze the differences in calculation time between the three allocation schemes, as shown in Figure 7.

Table 2: System simulation parameter.

Simulation parameter	Value
Satellite height	500 km
Downlink frequency	10.7–12.7 GHz
Maximum beams	40
Number of channels	16
Maximum transmission rate	1000 Mbps
Service rate threshold	100 kbps
Maximum transmitting power	23 dBW
Maximum power of beam	20 dBW
Antenna angle of beam	1°
Learning rate	0.1
Discount factor	0.9
Initial explore probability	0.9
Maximum step	10000
Service arrival rate	[10, 40] times/hour
Business duration	[3, 6] minutes

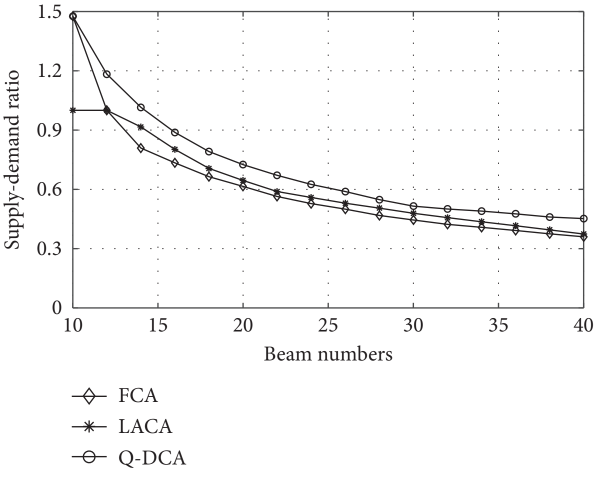

Figure 5: Supply-demand Ration under varying beam number.

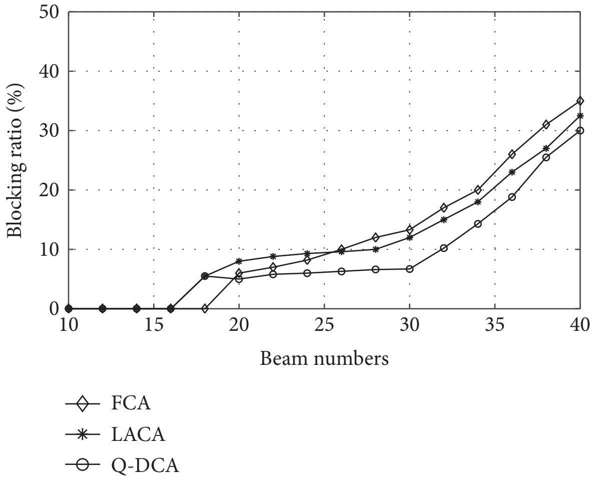

Figure 6: Blocking Ration under varying beam number.

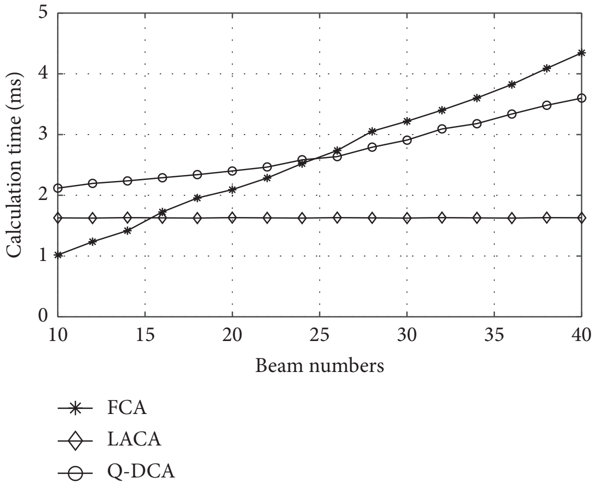

Figure 7: Calculation time of three schemes.

Because the FCA scheme adopts a uniform allocation principle, the number of calculations is relatively small, and therefore its calculation time is minimal. For the LACA scheme, as the beam increases, it takes longer to calculate the function extremes. And in Q-DCA, we use the trained strategy for channel allocation. Under each service request, only the Q table needs to be updated each time to get the optimal allocation scheme. Although the FCA scheme takes the least amount of time, it has the highest blocking rate when resources are tight. And the time complexity of the Q-DCA scheme is lower than the LACA scheme.

5.2. The System Performance in Beam Number Fixed Scenario. In this scenario, the number of satellite beam is fixed as 10, while the business request in beam increases from 900 Mbps to 1700 Mbps, simulating the scene in which the user changes from sparse to dense. Figures 8 and 9 show the system performance of the three schemes in the scene in which the number of beams is fixed.

When the number of beams is fixed at 10, the system supply-demand ratio decreases with the increase of the total system traffic. It can be seen that when the total numbers of business requests exceed 1000 Mbps, the system starts to block. At this time, the system business requests have exceeded the system payload. When the business request is 1500 Mbps, the system supply-demand ratio of the three schemes is 0.589, 0.542, and 0.475, respectively. At the same time, when the blocking rate is 30%, the system traffic volumes of the three schemes are 1620 Mbps, 1500 Mbps, and 1430 Mbps, respectively. In other words, the proposed Q-DCA scheme can further improve the system business processing capacity compared with the previous two algorithms while ensuring the same system blocking rate.

5.3. Spectrum Utilization and Algorithm Convergence Performance. In this scenario, the number of satellite beams is 10, and the total business volume of the system is 1000 Mbps. Comparing the convergence speed of the original Q-learning algorithm and the improved Q-learning algorithm when the system resources are exactly exhausted. Figure 10 shows the comparison of the convergence performance of the two algorithms.

As shown in Figure 10, the original Q-Learning algorithm starts to converge after about 4000 steps, while the improved Q-Learning algorithm starts to converge after about 2000 steps. Reflected in the actual application scenario, the improved Q-Learning algorithm can already shorten the system processing time by one time, thereby shortening the on-board processing delay.

Figure 11 analyzes the channel utilization of the original Q-Learning algorithm and the improved the Q-Learning algorithm when the system resources are abundant and scarce. It can be seen that the channel utilization of the two algorithms is almost the same whether the system resources are abundant or scarce, except for the convergence speed of the algorithm. Therefore, the improved algorithm will not change its utilization rate of the system resources.

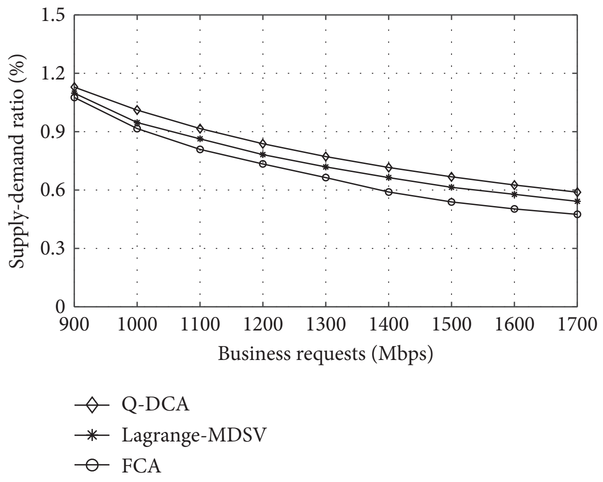

Figure 8: Supply-demand Ration under fixed beam number.

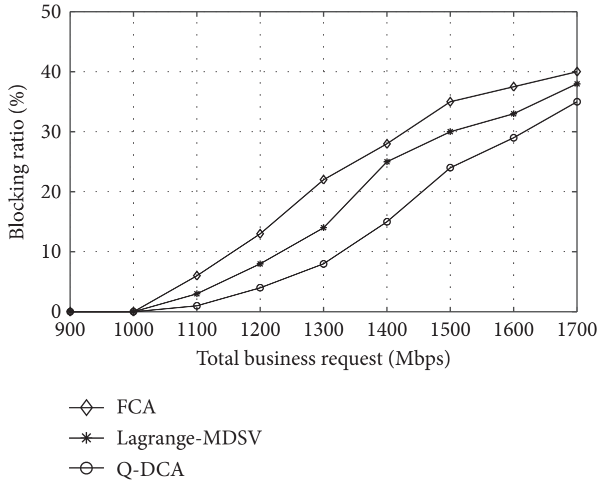

Figure 9: Blocking Ration under fixed beam number.

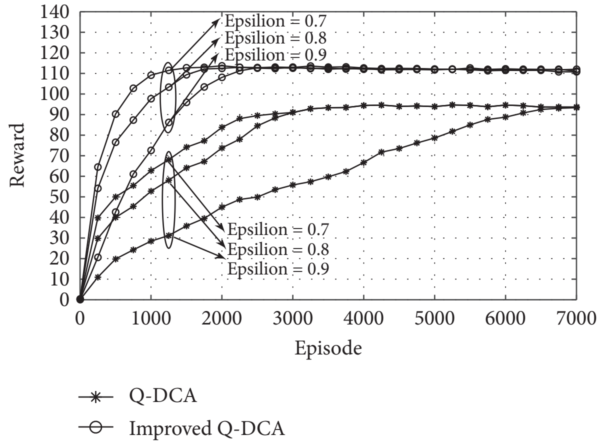

Figure 10: Algorithm convergence rate comparison.

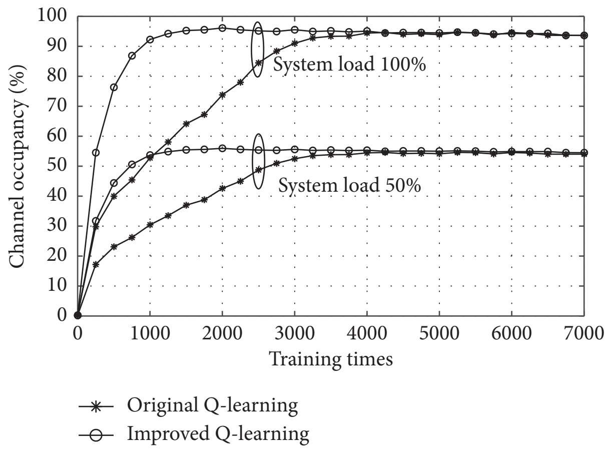

Figure 11: Spectral utilization efficiency comparison.

6. Conclusion

This paper proposes a LEO satellite network architecture based on a satellite resource pool. The system manages channel resources through a centralized resource pool to adapt to the traffic difference between beams. We adopt the *Q*-learning algorithm in RL for dynamic channel allocation. The simulation section analyzes the system performance and time complexity of FCA, LACA, and Q-DCA schemes in different scenarios. Analysis shows better performance of the proposed scheme in terms of channel allocation. Furthermore, we analyze the convergence of the *Q*-Learning algorithm and its impact on channel utilization. Simulation results show the effectiveness and the convergence performance of our proposed scheme.

Data Availability

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

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The work presented in this paper has been supported by Dean Project of Key Laboratory of Cognitive Radio and Information Processing, Ministry of Education (CRKL180104), Open Foundation of the State Key Laboratory of Networking and Switching Technology (Beijing University of Posts and Telecommunications) (SKLNST-2020-1-08), Project of Promoting Basic Scientific Research Ability of Young and Middle-aged Teachers in colleges and universities in Guangxi (2019KY0255), National Science Foundation of Shanxi Province (201801D121116), and Special Program of Guangxi Science and Technology Base and Talents (2018AD19048).

References

[1] H. Liang, *Research of Wireless Network Resource Management under CRAN Architecture*, Beijing University of Posts and Telecommunication, Beijing, China, 2016.

[2] L. Liu, *Research on Location Management and User Access Technology of LEO Satellite Communication*, University of Electronic Science and Technology of China, Chengdu, China, 2010.

[3] C. Qi, *Architecture Research for Low Earth Orbit Satellite Internet of Things*, Nanjing University of Posts and Telecommunications, Nanjing, China, 2019.

[4] C. Qiu, H. Yao, F. R. Yu, F. Xu, and C. Zhao, "Deep Q-learning aided networking, caching, and computing resources allocation in software-defined satellite-terrestrial networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 6, pp. 5871–5883, 2019.

[5] D. J. He, P. You, and S. W. Yong, "Mobility management in LEO satellite communication networks," *China Space Science and Technology*, vol. 36, no. 3, pp. 1–14, 2016.

[6] F. Fang and M. G. Wu, "Research on the development of global LEO satellite constellation," *Aerodynamic Missile Journal*, vol. 5, no. 6, pp. 88–92, 2020.

[7] S. J. Liu, *The Research on Dynamic Resource Management Techniques for Satellite Communication System*, Beijing University of Posts and Telecommunication, Beijing, China, 2018.

[8] Z. Liu, "Research on satellite communication resource allocation algorithm based on Reinforcement Learning," *Mobile Communication*, vol. 43, no. 5, pp. 27–32, 2019.

[9] S. Richard and B. Andrew, *Reinforcement Learning: An Introduction*, pp. 1–7, University of Electronic Science and Technology of China, Chengdu, China, 2017.

[10] Y. L. Liu and L. D. Zhu, "A suboptimal routing algorithm for massive LEO satellite networks," in *Proceedings of the International Symposium on Networks, Computers and Communications*, pp. 1–5, ISNCC, Rome, Italy, 2018.

[11] S. D. Feng, H. P. Zhu, and G. X. Li, "Dynamic modeling and simulation for LEO satellite networks," in *Proceedings of the 11th IEEE International Conference on Communication Technology*, pp. 37–40, ICCT, Hangzhou, China, 2018.

[12] H. J. Liu, P. Qin, N. W. Wang, Z. Lu, and B. Zhou, "Research on architecture design and resource allocation algorithm of LEO constellation," *Journal of China Academy of Electronic Sciences*, vol. 13, no. 06, pp. 631–635, 2018.

[13] X. Qi and J. Y. Sun, "Advances and challenges for software-defined LEO small satellite networks," in *Proceedings of the 16th Annual Meeting of Satellite Communication*, pp. 89–93, Beijing, China, 2020.

[14] K. W. Wang and S. Wang, "Design of satellite ground station based on resource pool architecture," *Communications World*, vol. 26, no. 8, pp. 16-17, 2019.

[15] M. M. Zhang, "Application analysis of resource pool architecture in satellite communications ground station network," *Information Technology and Information Technology*, vol. 7, pp. 101-102, 2019.

[16] J. H. Chang and A. J. Liu, "System reliability analysis of satellite communication center station under resource pooling architecture," *Communications Technology*, vol. 53, no. 2, pp. 375–381, 2020.

[17] W. T. Zhai, *Research on Virtualization Resource Management Technology of Satellite Network*, XiDian University, Xian, China, 2019.

[18] Y. Q. Xu, "The influence of resource pool architecture on the," *Construction of Satellite Communication Earth Station Network*, vol. 37, no. 11, pp. 160–162, 2020.

[19] H. Wang, A. J. Liu, X. F. Pan, and L. L. Jia, "Optimal bandwidth allocation for multi-spot-beam satellite communication systems," in *Proceedings of the 2013 International Conference on Mechatronic Sciences, Electric Engineering and Computer (MEC)*, Shengyang, China, 2013.

[20] G. Colavolpe, A. Modenini, A. Piemontese, and A. Ugolini, "On the application of multiuser detection in multibeam satellite systems," in *Proceedings of the IEEE International Conference on Communications*, vol. ICC, pp. 898–902, London, UK, 2015.

[21] A. Ivanov, M. Stoliarenko, S. Kruglik, S. Novichkov, and A. Savinov, "Dynamic resource allocation in LEO satellite," in *Proceedings of the International Wireless Communications and Mobile Computing Conference 2019, IWCMC*, pp. 930–935, Beiijing, China, 2019.

[22] T. Zhang, L. Zhang, and D. Shi, "Resource allocation in beam hopping communication system," in *Proceedings of the IEEE/AIAA 37th Digital Avionics Systems Conference 2018, DASC*, pp. 1–5, London, UK, 2018.

[23] P. Zuo, T. Peng, W. Linghu, and W. Wang, "Resource allocation for cognitive satellite communications downlink," *IEEE Access*, vol. 6, pp. 75192–75205, 2018.

[24] R. Chang, Y. He, G. Cui et al., "An allocation scheme between random access and DAMA channels for satellite networks," in *Proceedings of the IEEE International Conference*, pp. 1–6, Shenzhen, China, 2016.

[25] B. Zhao, J. Liu, Z. Wei, and I. You, "A deep reinforcement learning based approach for energy-efficient channel allocation in satellite Internet of things," *IEEE Access*, vol. 8, pp. 62197–62206, 2020.

[26] Z. Wang, J. X. Zhang, X. Zhang, and W. B. Wang, "Reinforcement learning based congestion control in satellite Internet of things," in *Proceedings of the 11th International Conference on Wireless Communications and Signal Processing 2019, WCSP*, pp. 1–6, Xi'an, China, 2019.

[27] X. Hu, "Multi-agent deep reinforcement learning-based flexible satellite payload for mobile terminals," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 9, pp. 9849–9865, 2020.

[28] B. Deng, C. Jiang, H. Yao, S. Guo, and S. Zhao, "The next generation heterogeneous satellite communication networks: integration of resource management and deep reinforcement learning," *IEEE Wireless Communications*, vol. 27, no. 2, pp. 105–111, 2020.