Learning-Based Admission Control for Low-Earth-Orbit Satellite Communication Networks

CHENG Lei, QIN Shuang, FENG Gang

(University of Electronic Science and Technology of China, Chengdu 611731, China)

DOI: 10.12142/ZTECOM.202303008

https://link.cnki.net/urlid/34.1294.TN.20230830.1858.005, published online August 31, 2023

Manuscript received: 2022-07-22

Abstract: Satellite communications has been regarded as an indispensable technology for future mobile networks to provide extremely high data rates, ultra-reliability, and ubiquitous coverage. However, the high dynamics caused by the fast movement of low-earth-orbit (LEO) satellites bring huge challenges in designing and optimizing satellite communication systems. Especially, admission control, deciding which users with diversified service requirements are allowed to access the network with limited resources, is of paramount importance to improve network resource utilization and meet the service quality requirements of users. In this paper, we propose a dynamic channel reservation strategy based on the Actor-Critic algorithm (AC-DCRS) to perform intelligent admission control in satellite networks. By carefully designing the long-term reward function and dynamically adjusting the reserved channel threshold, AC-DCRS reaches a long-run optimal access policy for both new calls and handover calls with different service priorities. Numerical results show that our proposed AC-DCRS outperforms traditional channel reservation strategies in terms of overall access failure probability, the average call success rate, and channel utilization under various dynamic traffic conditions.

Keywords: satellite communications; admission control; dynamic channel reservation; actor-critic

Citation (Format 1): CHENG L, QIN S, FENG G. Learning-based admission control for low-earth-orbit satellite communication networks [J]. ZTE Communications, 2023, 21(3): 54 - 62. DOI: 10.12142/ZTECOM.202303008
Citation (Format 2): L. Cheng, S. Qin, and G. Feng, "Learning-based admission control for low-earth-orbit satellite communication networks," *ZTE Communications*, vol. 21, no. 3, pp. 54 - 62, Sept. 2023. doi: 10.12142/ZTECOM.202303008.

1 Introduction

With the increasing number of users and service types in terrestrial wireless communication networks, it is impractical to provide wireless communication services anytime and anywhere alone[1]. The satellite communication network has the prominent advantages of long-distance communications, ubiquitous coverage, large capacity and high reliability, and can be a complement to terrestrial networks. It is not restricted by complex geographic conditions and harsh environments and can provide broadband multimedia services to user terminals (UT) in any area, even where terrestrial network resources are insufficient[2].

Due to the advantages of shorter propagation delay and lower operational expenditure, low-earth-orbit (LEO) satellite communication systems have been commonly used to provide user terminals with full coverage and real-time wireless communication services[3]. Usually, LEO communication systems can exploit multi-beam technology to irradiate lots of blocks of cellular networks in their coverage area, which are called beam cells. When a UT establishes a communication connection with an LEO satellite, one challenge faced is the frequent handover from one beam cell to another, due to the fast movement of LEO satellites. If there are insufficient channel resources in the targeted beam cell, the connection would be interrupted. Frequent handover failure and new call blocking would severely degrade the network performance and/or quality of service (QoS) of users. Moreover, with the rapid development of multimedia applications, diversified service requirements pose a great challenge to the network[4 - 6]. On the other hand, the satellite channel resources are limited, which usually cannot satisfy the requirements of all services. Considering the diversified service requirements with multi-priority services, admission control is particularly critical, as it decides which services are allowed to be admitted.

A typical solution for admission control of multi-priority ser-

This work was supported by the ZTE Industry-University-Institute Cooperation Funds.

vices in satellite systems is to allocate channel resources of beam cells by a priority-based channel reservation strategy, which has been intensively investigated in satellite communication systems[4 – 10]. The basic idea is to reserve a certain number of channels for handover calls and new calls with different service priorities, to guarantee the priority of handover calls and delay-sensitive services to ensure the continuity of calls for moving UTs.

Existing channel reservation strategies are mainly classified into two categories, fixed channel reservation (FCR) and dynamic channel reservation (DCR). A guaranteed handover FCR strategy was proposed in Ref. [7], which reserves a portion of channel resources dedicated to handover calls. Some improvements have been made later, such as the channel status-based reservation strategy (CSRS) [8] and time-based channel reservation algorithm (TCRA)[9], which set the number of reserved channels based on the information including the status of the cell and/or the remaining time. However, fixed reserved channels cannot adapt to the dynamic environment and multi-service requests, causing a high blocking rate for new calls. In Refs. [10 – 15], the authors proposed adaptive DCR strategies based on different prior information to dynamically change the number of channels reserved. The authors of Ref. [10] proposed to adjust the number of reserved channels, according to the current number of ongoing calls (voice or video traffic) and the localization of users. In Ref. [11], a grey model was used to decide whether the calls need to handover and then dynamically adjust the channel reservation number based on the counter. The authors of Ref. [12] leveraged the number of mobile stations in neighbor locations and the average handover call arrival rate to reserve channels. Ref. [13] considered the varying characteristics of the wireless channel to allocate resources, aiming at maximizing spectral efficiency. The authors of Ref. [15] proposed an adaptive probability-based reservation strategy (APRS) based on mobile users' location information and the handover probability, to improve the utilization of reserved channels in reservation time. Due to the imbalance between the new call blocking rate and handover call failure rate, the system performance is not satisfactory. Some researchers have used heuristic algorithms to adjust the thresholds, and the authors of Ref. [4] proposed a probability-based channel reservation strategy for improving the quality of service. The authors of Ref. [5] proposed a threshold-based DCR scheme to set optimal thresholds for different-priority services by the genetic algorithm.

However, all aforementioned schemes cannot respond quickly to dynamic changes and uneven distribution of service requirements, since they only consider finding the optimum in the current state, while a long-term optimization is needed to improve the system performance. With this regard, some researchers resort to exploiting machine learning algorithms for designing intelligent channel reservation schemes to achieve long-term performance improvement for complex satellite networks. The authors of Ref. [15] proposed a dynamic channel allocation algorithm based on deep reinforcement learning (DRL), which uses convolutional neural networks to extract useful features to make accurate admission decisions. It can effectively reduce the blocking rate and improve system throughput. But this work focused on processing the connection relationship of the UT in the beam and considered only a single service type. The authors of Ref. [6] proposed a multi-service DCR strategy based on the deep Q network to improve the overall service quality of the system, by examining the impact of current channel reservation results on the future environment. They mainly considered how to reserve channels for new calls, while ignoring the impact of handover calls. Unfortunately, all the aforementioned works lack consideration of multi-priority services and frequent handovers in highly dynamic LEO satellite networks. Therefore, it is imperative to develop an intelligent admission control scheme to maximize long-term system performance by performing appropriate channel resource allocation for LEO satellite networks.

In this paper, we propose an intelligent DCR strategy based on the Actor-Critic algorithm (AC-DCRS), which dynamically adjusts the reserved channel thresholds for multi-service calls. While traditional solutions only obtain the optimal solution of the current state in a memoryless system, our proposed AC-DCRS based on reinforcement learning can consistently approach the long-term optimal solution by considering the Markov property of the channel reservation problem. Specifically, the Actor-Critic algorithm is leveraged to deal with continuous state space and high-dimensional action space. Through interactions with the network environment, AC-DRCS can well balance the admission of handover calls and new calls of multiple priorities by setting corresponding thresholds under the current traffic state.

The rest of the paper is structured as follows. Section 2 presents the system model and problem formulation. Section 3 elaborates on the proposed AC-based DCR strategy. We evaluate the performance of the proposed strategy in Section 4 and finally conclude the paper in Section 5.

2 System Model and Problem Formulation

We consider a typical LEO communication network shown in Fig. 1. The coverage area of a single moving LEO satellite consists of multiple adjacent beam cells. A UT in the beam cell establishes a connection to the LEO satellite directly or through the base station in the beam cell. Handover call requests will arrive during the movement of the UT and satellites. At the same time, new call requests may also arrive requiring channel resources from the connected beam cell.

2.1 LEO Mobility Model

Without loss of generality, we consider a one-dimensional square continuous beam cell model, which can be readily extended to high-dimensional models. Since the moving speed of

the UT is much slower than that of the LEO satellites, it can be regarded as relatively static. Hypothetically, a LEO satellite moving horizontally to the left at a speed v_{sat} relative to UT is equivalent to UT moving to the right at the same speed relative to beam cells. We further adopt a cyclic mechanism to simplify the periodicity of satellite movement and the simplified satellite movement model is shown in Fig. 2. When the UT leaves the rightmost cell N as it moves, it will enter the leftmost cell 1. Handover call arrivals, new call arrivals, and call termination may constantly occur during the movement.

2.2 Channel Reservation Model for Multi-Priority Service

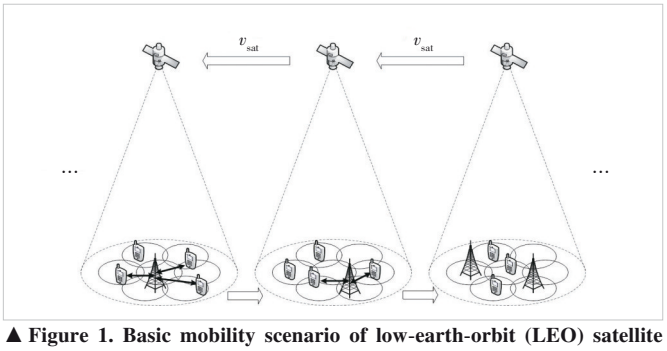
We assume that there are s types of services, and the number of new calls of the i-th service follows a Poisson distribution with an average arrival rate λ_n^i. Then the total arrival rate of the i-th service is given as the sum of the arrival rates of new calls and handover calls as follows:

$$\lambda^i = \lambda_n^i + \lambda_h^i , \tag{1}$$

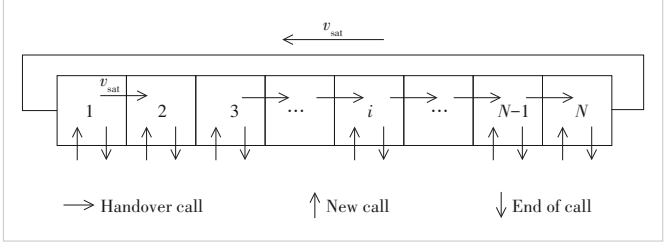
where λ_h^i is the arrival rate of the i-th service for handover calls. Obviously, the handover calls of the i-th service come from the previous beam cell and its arrival rate λ_h^i can be derived as follows[5]:

$$\lambda_h^i = \lambda_n^i \frac{(1 - P_{af}^i)P_{h1}}{1 - (1 - P_{hf}^i)P_{h2}} , \tag{2}$$

where P_{h1}, P_{h2}, P_{af} and P_{hf} are the handover success probability of the source cell, that of the target cell, the new call blocking rate of the i-th service, and the handover call failure rate of the i-th service, respectively.



▲ **Figure 1. Basic mobility scenario of low-earth-orbit (LEO) satellite communication system**



▲ **Figure 2. Simplified mobility model of low-earth-orbit (LEO) satellite communication system**

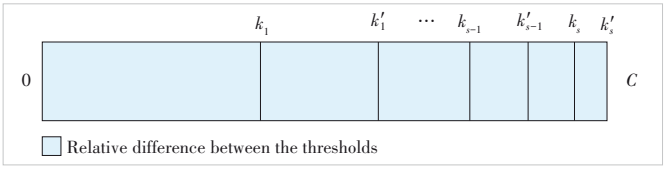
Besides, we denote the proportion of new calls of i-th service as p_i, and then the arrival rate of the i-th service can be expressed as:

$$\lambda_n^i = \lambda_n \times p_i , \tag{3}$$

where λ_n is the average arrival rate of total new calls and p_i satisfies $p_1 + p_2 + p_3 + \cdots + p_s = 1$. We further assume that the duration of all calls obeys an exponential distribution of a parameter u, so the average duration of the call is $1/u$ s.

We adopt a threshold-based channel reservation strategy to realize the admission control of satellite beam cells. We consider that the available bandwidth in a cell is equally allocated to all the channels, and each channel can be assigned to a call. Then, each admitted call will be assigned a channel with enough power to guarantee the quality of service. In this work, we focus on developing an intelligent admission control mechanism for LEO satellite communications. Thus, for simplifying the analysis, we just assume that there is enough power in a cell to guarantee the service quality of each admitted call. For a new call or a handover call, if the number of occupied channels in the current beam cell is less than the corresponding threshold, the call will be admitted successfully and assigned a channel with power and bandwidth resources, and the number of occupied channels is updated. Otherwise, the call will be blocked. After each decision period, the threshold will be updated according to our adjustment algorithm. We set a threshold k_i' for handover calls of the i-th service, while a threshold k_i is set for new calls of the i-th service. As handover calls are prior to new calls[16], we have $k_i \leqslant k_i'$. We also consider that the s types of services have certain priorities, in the way that the type with larger index numbers will be reserved for more channels than those with smaller index numbers. Therefore, the relationship between thresholds of all services satisfies $0 \leqslant k_1 \leqslant k_1' \leqslant \cdots \leqslant k_s \leqslant k_s' \leqslant C$, where C is the total number of beam cell channels. To maximize utilization of the channels, $k_s' = C$ is assumed in our model. Fig. 3 illustrates the proposed multi-priority service threshold-based channel reservation strategy.

The set of all thresholds is denoted by $K = \{k_1, k_1', k_2, k_2', \cdots, k_s, k_s'\}$. Intuitively, dynamically adjusting the thresholds K to control call admission can effectively improve the overall system performance. On the one hand, if the low-priority service threshold is set too low, low-priority service calls will be hard to get admission, even if there are no high-



▲ **Figure 3. Illustration of multi-priority service threshold-based channel reservation strategy**

priority service calls. Some channel resources may be wasted, resulting in low system channel utilization. On the other hand, if the low-priority service threshold is set too high, excessive low-priority service calls may be admitted to occupy too many channels. This will decrease the admission success rate of high-priority services in the future. Therefore, in this paper, we focus on designing a channel reservation strategy to dynamically adjust the thresholds of multi-priority calls, to realize intelligent admission control.

2.3 Problem Formulation

The overall access failure probability at time t, denoted by $O(t)$, as a system performance metric, is defined as:

$$O(t) = \alpha_0 P_{af}(t) + \alpha_1 P_{hf}(t) = \alpha_0 \sum_{i=1}^{s} \beta_i P_{af}^i(t) + \alpha_1 \sum_{i=1}^{s} \beta_i P_{hf}^i(t), \tag{4}$$

where α_0 and α_1 are the balance factors of new calls and handover calls respectively, which are used to measure the different impacts of new call blockage and handover call failure. And β_i is the balance factor of the i-th priority service, which is used to judge the significance of multi-priority services. Meanwhile, we modify $P_{af}^i(t)$ and $P_{hf}^i(t)$ as long-term metrics of the i-th service as follows:

$$P_{af}^i(t) = N_{af}^i / N_a^i, \tag{5}$$

$$P_{hf}^i(t) = N_{hf}^i / N_h^i, \tag{6}$$

where N_{af}^i, N_a^i, N_{hf}^i and N_h^i represent the number of new calls blocked for the i-th service, the total number of new calls for the i-th service, the number of handover calls failed for the i-th service, and the total number of handover call for the i-th service, respectively.

To improve the overall system performance, we minimize $O(t)$, i.e., minimize $P_{af}^i(t)$ and $P_{hf}^i(t)$ in the long term. As mentioned above, the setting of K directly affects the failure rate of new calls and handover calls. When the state space of each beam cell is modeled as a continuous-time M/M/C/C Markov chain, the closed-form relationship of the new call blocking rate, handover call failure rate and K can also be proved[5]. Therefore, we formulate our optimization problem as follows:

$$\begin{aligned} &\max \quad -O(t), \\ &\text{s.t.} \quad 0 < k_1 < k_1' < \cdots < k_s < k_s' \leqslant C \qquad (7.1), \\ &\qquad k_i, k_i' \in Z, \quad i = 1, 2, \cdots, s \qquad (7.2), \end{aligned} \tag{7}$$

where each threshold is limited to an integer for the convenience of adjustment in the dynamic channel reservation strategy. As the number of system channels C and that of services s can be large in real environments, using the brute force method to calculate the optimal thresholds in the current state will cause an exponential increase in time and space complexity. In addition, due to the rapid changes in the environment, optimal thresholds should be derived in real time. Thus, using static optimization to solve Problem (7) is infeasible and thus we resort to a learning-based solution.

3 Intelligent Admission Control Based on Dynamic Channel Reservation Strategy

In this section, we control service call admission by adjusting the reserved channel thresholds and model the problem of dynamically adjusting reserved channel thresholds as a Markov decision process (MDP). First, we slot the time as decision periods with the slot length T_Δ. In each time slot, multiple calls arrive according to the Poisson distribution. The system will adjust the reservation thresholds at the end of each decision period. The maximum number of calls in a decision period is set to N, and even if the decision period is not over, the decision will be made immediately.

3.1 MDP Model

An MDP model consists of a five-tuple $< S, A, P, R, \pi >$, where S, A, P, R and π represent state space, action space, transition probability between states, reward function, and policy for selecting actions based on the state, respectively, which are defined as follows:

1) State(S): we assume that the channel resources of the beam cell remain unchanged. The state is defined as:

$$s(t) \in \{ c ; \lambda ; K \} \qquad t = nT_\Delta, \ n = 0, 1, 2, \cdots, \tag{8}$$

where c is the normalized number of channels that have been occupied in the considered beam cell and satisfies $c \leqslant C$; $\lambda = \{ \lambda_n^1, \lambda_h^1, \cdots, \lambda_n^s, \lambda_h^s \}$ is the set of the call arrival rates of new calls and handover calls that satisfies Eq. (2); $K = \{ k_1, k_1', \cdots, k_s, k_s' \}$ is the set of the normalized reserved channel thresholds of new calls and handover calls that satisfies Eq. (7.1).

2) Action(A): we define the action as current normalized reserved channel thresholds, which can be expressed as:

$$a(t) = K^T = \{ k_1, k_1', \cdots, k_s, k_s' \}^T. \tag{9}$$

In each decision period, action will be taken based on the current state, which will control the admission by setting reserved channel thresholds.

3) Transition Probability(P): generally, the state transition function of the Markov decision process is a certain function $P: S \times A \times S \to [0, 1]$, which represents the probability of the transition to the state s' given state s after taking action a. Since state transition depends on not only the last action but also the traffic changes caused by the movement of satellites and UTs and call termination in the channel, it cannot be explicitly expressed in our problem.

4) Reward(R): for a single service call within a decision pe-

riod, the reward function is defined as:

$$r=\begin{cases}0 & \text{access successfully}\\ -\alpha_0\beta_i \text{ or } -\alpha_1\beta_i & \text{access failed}\\ -L & (7.1)\text{ not meet}\end{cases}. \tag{10}$$

As is defined in Eq. (10), when a new call or a handover call is blocked, a negative value $-\alpha_0\beta_i$ / $-\alpha_1\beta_i$ will be given as a punishment. L is a very large constant as a penalty for dissatisfying the constraint (7.1). Thus, the reward function of a decision period can be defined as:

$$r_\Delta=\sum_{\alpha,\beta,i} r/N^i_{\alpha,\beta}=-\alpha_0\sum_{i=1}^{s}\frac{\beta_i N^i_{af}}{N^i_a}-\alpha_1\sum_{i=1}^{s}\frac{\beta_i N^i_{hf}}{N^i_h}. \tag{11}$$

Substituting Eqs. (4) – (6) can be further expressed as:

$$r_\Delta=-\left(\alpha_0\sum_{i=1}^{s}\beta_i P^\Delta_{af_i}+\alpha_1\sum_{i=1}^{s}\beta_i P^\Delta_{hf_i}\right)=-O_\Delta(t). \tag{12}$$

Then our optimization problem of maximizing $-O(t)$ can be approximately solved by maximizing the accumulated reward $\sum_T r_\Delta$ in the long run.

5) Policy(π): we use a random policy $\pi(a|s)\rightarrow[0,1]$ to represent the probability of selecting the action a given the current state s.

In our MDP model, we use the state-value function to evaluate the value of state s, which can be expressed as:

$$V_\pi(s)=E^\pi\left[\sum_{k=0}^{\infty}\gamma^k r^{t+k}(s^{t+k},a^{t+k})|s_t\right], \tag{13}$$

where γ is the discount factor representing the discount contribution of the future states to the current state. Besides, the action-value function is used to evaluate the selected action a in the current state s, and can be expressed as:

$$Q^\pi(s,a)=E^\pi\left[\sum_{k=0}^{\infty}\gamma^k r^{t+k}(s^{t+k},a^{t+k})|s_t,a_t\right]. \tag{14}$$

Assuming that the MDP starts from the state $s^t\in S$, it experiences a trajectory as:

$$\kappa\sim\{s^t,a^t,s^{t+1},a^{t+1},\cdots,s^{t+T},a^{t+T}\}. \tag{15}$$

Since the policy is stochastic, the trajectory κ is uncertain. Denote the probability of trajectory κ as $\pi_\xi(\kappa)$, and the cumulative reward of trajectory κ is $R(\kappa)=\sum_{k=0}^{T}\gamma^k r^{t+k}$. As a result, the objective function can be rewritten as:

$$\max\ -O(t)\approx U(\pi_\xi)=E_{\kappa\sim\pi_\xi(\kappa)}[R(\kappa)]=\int_{\kappa\sim\pi_\xi(\kappa)}R(\kappa)d\kappa. \tag{16}$$

3.2 Actor-Critic-Based Dynamic Channel Reservation Strategy

This MDP problem can be solved by using the reinforcement learning (RL) algorithm. Specifically, we use the Actor-Critic framework[17] to model high-dimensional discrete action space, which is a combination of the Actor and the Critic. The Critic uses a neural network to approximate the state-value function and to judge the actions by temporal difference (TD) errors. The Actor uses another neural network to approximate the optimal policy and then selects the action while interacting with the environment.

1) Actor: The Actor will constantly improve the policy by TD errors. In our MDP problem, the policy π_ξ is modeled as a conditional probability distribution parameterized by ξ. Thus the process of modifying the policy is equivalent to the process of updating the parameter ξ. Through the back-propagation algorithm, ξ is updated as follows:

$$\xi_{\text{new}}=\xi_{\text{old}}+\alpha_{\text{actor}}\nabla_\xi U(\pi_\xi), \tag{17}$$

where α_{actor} is the learning rate of the Actor, and the gradient $\nabla_\xi U(\pi_\xi)$ is as follows:

$$\nabla_\xi U(\pi_\xi)=\nabla_\xi\log\pi_\xi(a|s)\times A^\pi(s,a), \tag{18}$$

where $A^\pi(s,a)$ is the advantage function.

In this problem, we use Gaussian probability distribution to formulate the policy, which can be expressed as:

$$\pi_\xi(a|s)=\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{(a-\mu(s))^2}{2\sigma^2}\right), \tag{19}$$

where $\mu(s)$ is the expectation and σ is the standard deviation of the selected action. Meanwhile, $\mu(s)$ is the action with the highest probability at the state s and σ represents the extent of exploration over all actions. Exploration and exploitation can be well balanced by exploiting the Gaussian distribution. Thus the policy can be modified through the process of updating $\mu(s)$.

To update $\mu(s)$, we extract a feature vector $\phi(s)$ from the current state as the input of the Actor neural network, which is expressed as:

$$\phi(s)=(c;\lambda;K)^T. \tag{20}$$

The neural network will then output the normalized average of reserved channel thresholds, which is denoted by $\mu(s)=(u_1,u_1',\cdots,u_s,u_s')^T$. Thus the policy can be further derived as a $2s$-dimensional Gaussian probability distribution:

$$\pi_\xi(a|s)=\frac{1}{(\sqrt{2\pi})^s(|Cov|)^{\frac{1}{2}}}e^{\left(-\frac{1}{2}(a-u(s))^T Cov^{-1}(a-u(s))\right)}, \tag{21}$$

where Cov is the covariance matrix with diagonal σ^2. Based on the policy, an action vector will be generated and the system will transit to a new state after a decision period T_Δ.

2) Critic: The Critic is used to approximate the state-value function $V_\pi(s)$. Traditional RL uses Q-value tables to record state values, which will face the problem of dimensional explosion under the scenario of large state space. To cope with this problem, a neural network parameterized by θ is utilized to approximate the state-value function $V_\theta(s)$. In the critic process, $V_\theta(s)$ will be updated by updating parameters θ.

To evaluate the gap between the actual value and the approximated value of the state-value function in the state s, the definition of TD-error is given as follows:

$$\delta_t = V_\pi(s^{t+1}) - V_\theta(s^t), \tag{22}$$

where $V_\pi(s^{t+1}) = r^{t+1} + \gamma V_\theta(s^{t+1})$ according to the bootstrapping method in the RL framework. To guide the updating of parameters and improve the performance, the objective of the critic process is designed to minimize the TD-error δ_t and can be re-expressed as:

$$\min\ 1/2\ (\delta_t)^2\ . \tag{23}$$

θ is updated by the gradient descent method as follows:

$$\theta_{\text{new}} = \theta_{\text{old}} - \alpha_{\text{critic}} |\delta_t| \nabla_{\theta_{\text{old}}} V_{\theta_{\text{old}}}(s^t), \tag{24}$$

where α_{critic} is the learning rate of the Critic.

3) Actor-Critic: The Actor updates the policy based on the state-value estimated by the Critic, while the Critic updates the state-value function according to the actions selected by the Actor and the state transitions generated by interactions with the environment. Besides, the performance of the Actor can be improved by replacing δ_t with an advantage function. Then the parameter update process in Actor can be rewritten as:

$$\nabla_\xi U(\pi_\xi) = \nabla_\xi \log \pi_\xi(a^t|s^t)\delta_t, \tag{25}$$

$$\xi_{\text{new}} = \xi_{\text{old}} + \alpha_{\text{actor}} \nabla_{\xi_{\text{old}}} \log \pi_{\xi_{\text{old}}}(a^t|s^t)\delta_t. \tag{26}$$

In summary, the proposed AC-DCRS is summarized as follows:

Algorithm 1. AC-DCRS

Input: $N, M, T, T_\Delta, \sigma, \alpha_{\text{actor}}, \alpha_{\text{critic}}, \lambda, \gamma$.
Output: Optimal dynamic adjustment policy π_ξ.
Initialize: $t = 0, n = 1, \xi = \xi_0, \theta = \theta_0, a = a_0, s = s_0, \Phi(s_0)$;
Repeat:

1. Action selection:
   Input $\Phi(s^t)$ into the Actor network to get $\mu(s^t)$ and select action a, i.e., adjust the threshold once.
2. According to the threshold adjusted by action, control incoming calls, while $(t \leqslant nT_\Delta)$:
   1) if a service call arrives, judge its service type and priority
   2) determine whether the call access is successful by the corresponding threshold:
      a) if c is less than the corresponding threshold, the call is admitted successfully and can be allocated a channel resource $c \leftarrow c + 1$;
      b) else the call is blocked.
   3) record the result of this call
   4) $t \leftarrow t + 1$.
3. State transition and reward feedback:
   1) obtain the access result in this T_Δ
   2) transition into the new state s^{t+1}, get a reward r^{t+1}
   3) update state feature vector $\Phi(s^{t+1})$
   4) calculate the state-value function $V_\theta(s^{t+1})$, $V_\theta(s^t)$.
4. Update policy:
   1) Critic network calculates and outputs TD-error $\delta_t = r^{t+1} + \gamma V_\theta(s^{t+1}) - V_\theta(s^t)$
   2) update Critic network parameters $\theta \leftarrow \theta - \alpha_{\text{critic}} |\delta_t| \nabla_\theta V_\theta(s^t)$
   3) update Actor network parameters $\xi \leftarrow \xi + \alpha_{actor} \nabla_\xi \log \pi_\xi(a^t|s^t)\delta_t$.
5. $n \leftarrow n + 1$, $s^t \leftarrow s^{t+1}$.

Until: $t \geqslant T$.

3.3 Complexity Analysis

In this subsection, we analyze the computing and space complexity of AC-DCRS and compare it with three baseline algorithms, i.e., FCR, handover priority fixed channel reservation strategy (HPFCR), and DCR.

In FCR and HPFCR, the thresholds are fixed. In HPFCR, the handover calls are given higher priority, and the thresholds for new calls are set to the same value. After the initial setting, the thresholds of DCR will dynamically change according to the proportion of the number of calls of various services after the decision period T_Δ has passed. Its normalized threshold can be expressed as:

$$K_{\text{DCR}} = c' + \frac{1 - c'}{n} [n_1, n_1 + n_1', \cdots, n_1 + \cdots + n_s'], \tag{27}$$

where c' represents the normalized number of shared channels, which can be used by all calls, n represents the total number of calls arrived, and n_i and n_i' represent the number of new calls and handover calls of the i-th service respectively. As both FCR and HPFCR use a fixed threshold, the computing complexity is $O(1)$and the space complexity is $O(s)$. On the other hand, DCR will dynamically change the threshold according to Eq. (27), and thus the computing complexity and space complexity are both equal to $O(s)$.

In our AC-DCRS, the neural networks are introduced to fit the policy function and the threshold is obtained according to the output feature vector. Specifically, we use a fully-

connected neural network including two dense hidden layers. Suppose the number of neurons of the two layers is N_{L_1} and N_{L_2}. The dimension of the input feature vector is $4s + 1$ and the dimension of the output feature vector is $2s$. Therefore, the computing complexity is $O(s^2 \times N_{L_1} \times N_{L_2})$. We need to store the weights and bias of the middle layer and the values of thresholds, which results in a space complexity of $O(N_{L_1} + N_{L_2} + s)$. By using additional space and computing resources, our AC-DCRS can intelligently adjust the threshold and achieve a better performance.

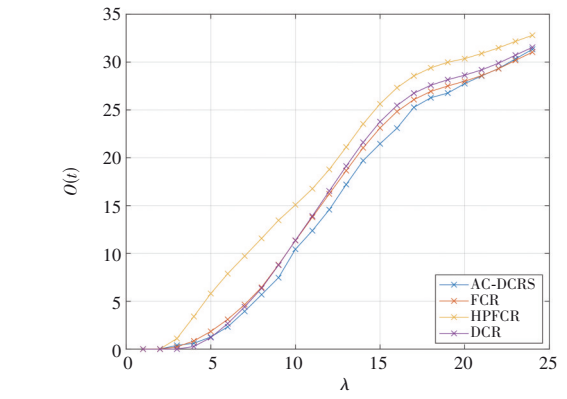
4 Performance Evaluation

4.1 Simulation Setting

We use a typical satellite system mobility model and basic assumptions, the parameters for the satellite communication network and the AC algorithm are shown in Tables 1 and 2 respectively.

4.2 Numerical Results

First, we examine the relationship between overall access failure probability $O(t)$, which is approximately equal to the negative long-term accumulated reward, with the varying total call arrival rate λ. As shown in Fig. 4, since our optimization objective is to minimize $O(t)$, the greater $O(t)$, the worse overall system performance. We can see that the proposed AC-DCRS algorithm can learn better admission control strategies and achieve better overall system performance, compared with FCR, HPFCR, and DCR in most traffic scenarios. However, there is no obvious advantage in the scenarios of very low and high traffic loads. This is because AC-DCRS needs some trial-and-error interactions with the environment. Reducing the thresholds of low-priority services causes some call failures in low-traffic scenarios, and multiple next-highest priority services get admission to quickly filling up the channel, which affects the overall access failure probability in high-traffic load scenarios.



AC-DCRS: DCR strategy based on Actor-Critic algorithm
DCR: dynamic channel reservation
FCR: fixed channel reservation
HPFCR: handover priority fixed channel reservation strategy

▲**Figure 4. $O(t)$ with varying λ**

Next, we explore the relationship between the channel utilization and the average call success rate with the varying total call arrival rate λ. From Figs. 5 and 6, we can find that the channel utilization increases with the traffic load, and the average call success rate decreases with the traffic load. We can

▼**Table 1. Simulation parameters**

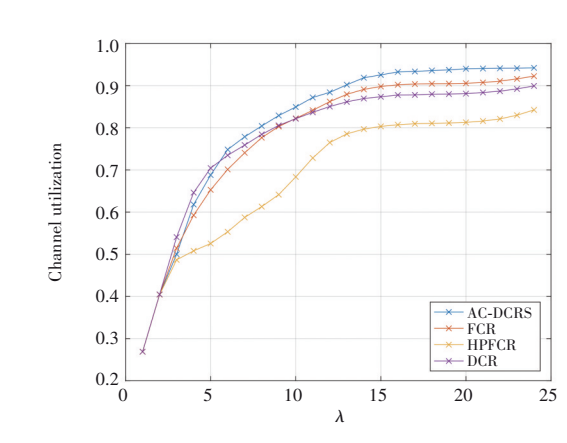
Parameter	Value
Number of services s	3
Number of beam cell channels	100
Average call duration parameter u	1/30 s
Call arrival rate ratio p_1: p_2: p_3	0.2:0.3:0.5
Decision period T_Δ	5 s
Maximum number of calls N	50
Balance factor α_0, α_1	0.4, 0.6
Balance factor β_1, β_2, β_3	0.2, 0.3, 0.5
FCR normalized fixed threshold setting	[0.73, 0.75, 0.82, 0.85, 0.86, 1]
HPFCR normalized fixed threshold setting	[0.73, 0.75, 0.73, 0.85, 0.73, 1]
DCR normalized initial threshold setting	[0.73, 0.75, 0.82, 0.85, 0.86, 1]
Number of periods played	20 000T_Δ
Total call arrival rate λ	2 – 25 calls/s

DCR: dynamic channel reservation
FCR: fixed channel reservation
HPFCR: handover priority fixed channel reservation

▼**Table 2. AC algorithm parameters**

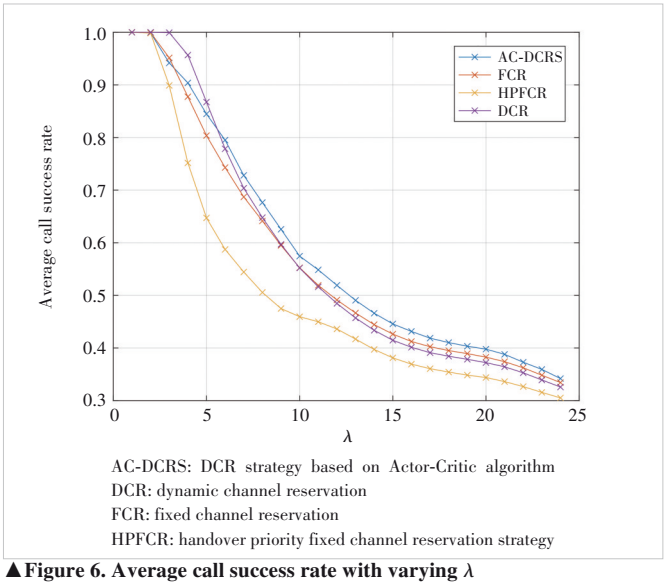
Parameter	Value
Discount factor γ	0.99
Learning rate of policy α_{actor}	0.002
Learning rate of value function α_{critic}	0.005
Action selection variance σ	0.05

AC: Actor-Critic



AC-DCRS: DCR strategy based on Actor-Critic algorithm
DCR: dynamic channel reservation
FCR: fixed channel reservation
HPFCR: handover priority fixed channel reservation strategy

▲**Figure 5. Channel utilization with varying λ**

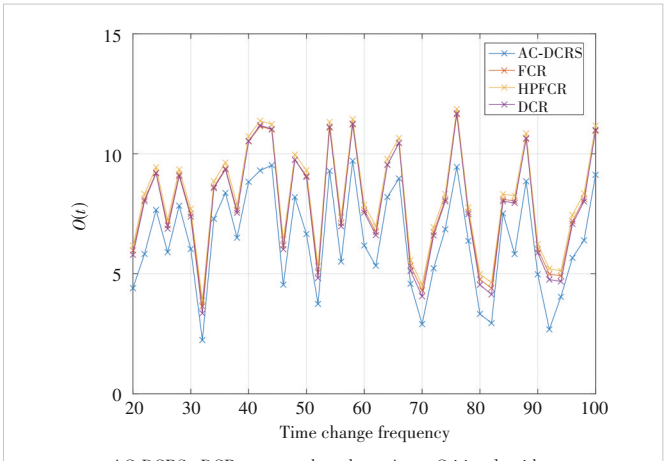


AC-DCRS: DCR strategy based on Actor-Critic algorithm
DCR: dynamic channel reservation
FCR: fixed channel reservation
HPFCR: handover priority fixed channel reservation strategy

▲**Figure 6. Average call success rate with varying λ**

also find that AC-DCRS outperforms FCR, HPFCR, and DCR in these two aspects. This is because AC-DCRS can well balance the call admission of all services from the level of the entire system. Ensuring the admission of high-priority service calls makes as many calls of multiple services as possible get admission.

We assume that the total call arrival rate changes dynamically at a certain time frequency. The initial total call arrival rate is 8 calls/s, and the range of change is $[\lambda_t - 2, \lambda_t + 2]$, where λ_t is the current total call arrival rate. As shown in Fig. 7, the AC-DCRS can achieve better system performance in different dynamic scenarios, compared with comparison algorithms. This is because our AC-DCRS can learn the optimal admission control strategy under the current traffic and can adjust the threshold in real time.



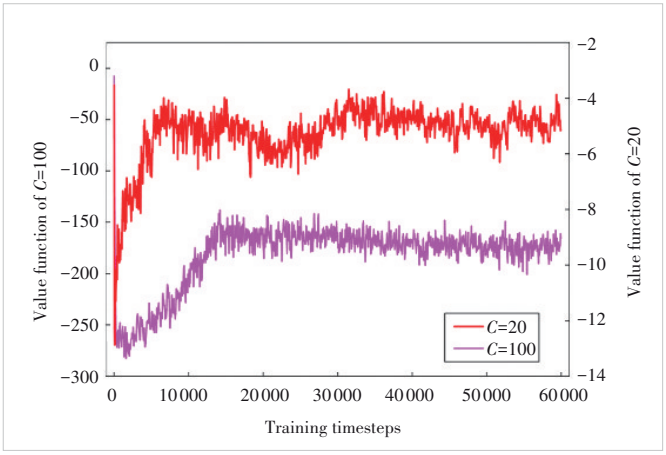
AC-DCRS: DCR strategy based on Actor-Critic algorithm
DCR: dynamic channel reservation
FCR: fixed channel reservation
HPFCR: handover priority fixed channel reservation strategy

▲**Figure 7. $O(t)$ with change frequency varying**

Finally, we show the convergence of the value function in the AC-DCRS algorithm. We separately consider the convergence in the small state space (C = 20) and big state space (C= 100) cases. We observe the dynamic change of the state-value function at a certain state as the Critic evolves. As shown in Fig. 8, we can find that after certain training steps, the value function converges. The convergence speed varies with the sizes of state space, for the reason that it requires more iterations to traverse a larger state space to reach optimal strategy. In addition, the obtained strategy through training can be applied to similar scenarios in different satellite beam cells. The training data of different satellite beam cells in similar scenarios can be shared for migration training, which will accelerate the convergence to the optimal strategy.

5 Conclusions

In this paper, we have proposed a dynamic channel reservation strategy AC-DCRS based on the Actor-Critic algorithm to realize intelligent admission control in a satellite network. AC-DCRS can learn an optimal admission policy for both new calls and handover calls with different service priorities, which will improve the performance of both the user side and the network. Numerical results show that our proposed AC-DCRS algorithm achieves better long-term overall system performance, average call success rate, and channel utilization under different traffic conditions and dynamic scenarios compared with traditional channel reservation strategies.



▲**Figure 8. Value function with time varying**

References

[1] LIU J J, SHI Y P, FADLULLAH Z M, et al. Space-air-ground integrated network: a survey [J]. IEEE communications surveys & tutorials, 2018, 20(4): 2714 – 2741. DOI: 10.1109/COMST.2018.2841996

[2] DING R, CHEN T T, LIU L, et al. 5G integrated satellite communication systems: architectures, air interface, and standardization [C]//International Conference on Wireless Communications and Signal Processing (WCSP). IEEE, 2020: 702 – 707. DOI: 10.1109/WCSP49889.2020.9299757

[3] DENG R Q, DI B Y, ZHANG H L, et al. Ultra-dense LEO satellite constellation design for global coverage in terrestrial-satellite networks [C]// Global Communications Conference. IEEE, 2021: 1 – 6. DOI: 10.1109/

GLOBECOM42002.2020.9322362
[4] DUAN C F, DUAN R Q, FENG J, et al. A novel channel allocation strategy in low earth orbit satellite networks [C]//6th International Conference on Computer and Communications (ICCC). IEEE, 2021: 8 – 13. DOI: 10.1109/ICCC51575.2020.9345173
[5] ZHOU J, YE X G, PAN Y, et al. Dynamic channel reservation scheme based on priorities in LEO satellite systems [J]. Journal of systems engineering and electronics, 2015, 26(1): 1 – 9. DOI: 10.1109/JSEE.2015.00001
[6] LI Z W, XIE Z C, LIANG X W. Dynamic channel reservation strategy based on DQN algorithm for multi-service LEO satellite communication system [J]. IEEE wireless communications letters, 2021, 10(4): 770 – 774. DOI: 10.1109/LWC.2020.3043073
[7] MARAL G, RESTREPO J, DEL RE E, et al. Performance analysis for a guaranteed handover service in an LEO constellation with a "satellite-fixed cell" system [J]. IEEE transactions on vehicular technology, 1998, 47(4): 1200 – 1214. DOI: 10.1109/25.728509
[8] WANG X L, WANG X X. The research of channel reservation strategy in LEO satellite network [C]//11th International Conference on Dependable, Autonomic and Secure Computing. IEEE, 2014: 590 – 594. DOI: 10.1109/DASC.2013.131
[9] BOUKHATEM L, GAITI D, PUJOLLE G. A channel reservation algorithm for handover issues in LEO satellite systems based on a satellite-fixed cell coverage [C]//IEEE VTS 53rd Vehicular Technology Conference. IEEE, 2001: 2975 – 2979. DOI: 10.1109/VETECS.2001.944147
[10] BEYLOT A L, BOUMERDASSI S. Adaptive channel reservation schemes in multitraffic LEO satellite systems [C]//IEEE Global Telecommunications Conference. IEEE, 2002: 2740 – 2743. DOI: 10.1109/GLOCOM.2001.966272
[11] ZOU Q Y, ZHU L D. Dynamic channel allocation strategy of satellite communication systems based on grey prediction [C]//International Symposium on Networks, Computers and Communications (ISNCC). IEEE, 2019: 1 – 5. DOI: 10.1109/ISNCC.2019.8909122
[12] CHATTERJEE S, SAHA J, BANERJEE S, et al. Neighbour Location Based Channel Reservation scheme for LEO Satellite communication [C]//International Conference on Communications, Devices and Intelligent Systems (CODIS). IEEE, 2012: 73 – 76. DOI: 10.1109/CODIS.2012.6422139
[13] RAHMAN M, WALINGO T, TAKAWIRA F. Adaptive handover scheme for LEO satellite communication system [C]//Proceedings of AFRICON. IEEE, 2015: 1 – 5. DOI: 10.1109/AFRCON.2015.7332051
[14] CHEN L M, GUO Q, WANG H Y. A handover management scheme based on adaptive probabilistic resource reservation for multimedia LEO satellite networks [C]//WASE International Conference on Information Engineering. IEEE, 2010: 255 – 259. DOI: 10.1109/ICIE.2010.67
[15] LIU S J, HU X, WANG W D. Deep reinforcement learning based dynamic channel allocation algorithm in multibeam satellite systems [J]. IEEE access, 2018, 6: 15733 – 15742. DOI: 10.1109/ACCESS.2018.2809581
[16] CHOWDHURY P K, ATIQUZZAMAN M, IVANCIC W. Handover schemes in satellite networks: state-of-the-art and future research directions [J]. IEEE communications surveys & tutorials, 2006, 8(4): 2 – 14. DOI: 10.1109/COMST.2006.283818
[17] BHATNAGAR S, SUTTON R S, GHAVAMZADEH M, et al. Natural actor-critic algorithms [J]. Automatica, 2009, 45(11): 2471 – 2482. DOI: 10.1016/j.automatica.2009.07.008

Biographies

CHENG Lei received her BS degree in communication engineering from University of Electronic Science and Technology of China (UESTC) in 2019. She now is pursuing her PhD degree at the National Key Laboratory of Wireless Communications, UESTC. Her research interests include resource management and network control in space-air-ground/satellite-terrestrial integrated networks by using optimization theory and machine learning techniques.

QIN Shuang (blueqs@uestc.edu.cn) received his BS degree in electronic information science and technology and PhD degree in communication and information system from University of Electronic Science and Technology of China (UESTC) in 2006 and 2012, respectively. He is currently a professor with the National Key Laboratory of Wireless Communications, UESTC. His research interests include wireless and mobile networks.

FENG Gang received his BE and ME degrees in electronic engineering from University of Electronic Science and Technology of China (UESTC) in 1986 and 1989, respectively, and PhD degree in information engineering from The Chinese University of Hong Kong, China in 1998. He joined the School of Electric and Electronic Engineering, Nanyang Technological University, Singapore in December 2000 as an assistant professor and became an associate professor in October 2005. He is currently a professor with the National Laboratory of Wireless Communications, UESTC. He has extensive research experience and has published widely on wireless networking. His research interests include next generation mobile networks, mobile cloud computing, and AI-enabled wireless networking. He has received the IEEE ComSoc TAOS Best Paper Award and the ICC Best Paper Award in 2019. A number of his papers have been highly cited.