Satellite Staring Beam Scheduling Strategy Based on Multi-agent Reinforcement Learning

Hongtao Zhu(✉), Zhenyong Wang, Dezhi Li, and Qing Guo

School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China
{zhuhongtao,ZYWang,lidezhi,QGuo}@hit.edu.cn

Abstract. Low Earth Orbit (LEO) satellites are an important part of Space-Air-Ground Integrated Networks (SAGIN), which play an irreplaceable role in providing global communication and emergency communication. With the development of phased array technology, many satellites begin to try to use staring beam technology, which can make the beam serve a hot spot on the ground as long as possible by adjusting its phased array parameters, so as to reduce the impact of fast switching on the service performance of LEO satellites. In the satellite service time, how to balance the load of each satellite and meet the communication needs of hot spots is an important problem to be considered. Excellent beam allocation strategy can reduce the network handover rate and signaling overhead. In this paper, the satellite staring beam scheduling problem is transformed into a two-dimensional model, and we propose a novel satellite beam scheduling strategy based on multi-agent reinforcement learning that aims to maximize system performance. Each satellite is regarded as an individual agent, and the decision is to provide communication beam for the current hot spot area. Compared with the beam allocation algorithm based on KM, simulation results show that the proposed strategy can effectively reduce the handoff rate of hot spots when the coverage is satisfied.

Keywords: Low orbit satellite · Multi-agent reinforcement learning · Staring beam scheduling

1 Introduction

Satellite communication can provide seamless wireless signal coverage to support and expand ground communication, which has become an important research direction of 5G and future 6G [1–3]. With the large-scale deployment of OneWeb, StarLink, TeleSat and other mega constellations, LEO satellite shows its advantages in reducing communication delay, providing wide area coverage, and not affected by the ground environment. However, due to the high mobility of LEO satellite, terminals need to switch frequently to maintain the connection of communication links. This will increase the signaling overhead and drop call rate

Q. Guo et al. (Eds.): WiSATS 2021, LNICST 410, pp. 23–34, 2022.
https://doi.org/10.1007/978-3-030-93398-2_3

of the system, and seriously affect the system throughput. Therefore, more and more scholars have studied the handoff problem of LEO satellite, hoping to reduce the impact of high-speed mobility.

According to whether the direction angle of satellite beam in LEO system is adjustable, LEO system is usually divided into two types: one is satellite fixed cell system (SFCS), the other is earth fixed cell system (EFCS) [4]. EFCS is also called staring beam satellite system. In [5], a new satellite handoff strategy based on the potential game of mobile terminal in LEO satellite communication network is proposed. In the software defined satellite network (SDSN) architecture, the author regards satellite handoff as a bipartite graph and proposes a terminal random-access algorithm based on the target of user space maximization. In [6], a fixed beam LEO satellite model is introduced, and the author proposes a method to analyze the throughput of fixed beam according to its coverage time. In [7], the authors propose a performance comparison of fixed and dynamic channel allocation techniques in a LEO satellite system, and they study the case of earth-fixed cell systems with different kinds of fixed and mobile users. In [8], the author presented a comprehensive literature review on applications of deep reinforcement learning (DRL) in communications and networking. In [9], a dynamic channel reservation (DCR) strategy based on deep Q network is proposed for multi-service LEO satellite communication system, which can improve the overall quality of service (QoS) of the system. Inspired by this, we will try to solve the staring beam scheduling problem by reinforcement learning.

The rest of the paper is organized as follows. In Sect. 2, we introduce the LEO satellite beam scheduling system model and an optimal problem is proposed under the constraints of the number of satellite beams and satellite capacity. In Sect. 3, we solve the problem by using a multi-agent DQN learning algorithm. Simulation results are analyzed in Sect. 4 and conclusions are drawn in Sect. 5.

2 System Model

We consider the problem of LEO satellite beam scheduling during satellite operation, as shown in Fig. 1. The set of satellites is denoted by $\mathcal{M} = \{1, 2, ..., M\}$. In this paper, the satellite is equipped with a multi beam phased array antenna system, so it can gaze at one or more hot spots by adjusting the parameters of the antenna array. Each satellite can form up to K beams, which is denoted by $\mathcal{K} = \{1, 2, ..., K\}$. The earth's surface is divided into a fixed number of hot spots according to the degree of user service demand, and the set of communication hot spots is denoted by $\mathcal{N} = \{1, 2, ..., N\}$. Figure 1 shows the coverage of two orbiting satellites to the hot spot area on the ground. The yellow dotted line is the satellite orbit. From t1 to t2, area A is within the coverage of Leo1, so Leo1 can adjust the antenna to maintain the connection to area A. At t3, leo1 exceeds the visible range of area A and starts to serve area F. At t2, because both leo1 and leo2 are within the visible range of area A, that is, a region may have multiple satellites covering at the same time. Also when the service satellite of a region leaves, it is necessary for this region to access another satellite to

Fig. 1. LEO satellite system model of staring beam.

ensure that the communication will not be interrupted. Therefore, staring beam satellite system mainly involves global satellite beam scheduling problem.

In order to simplify the model, this paper uses a two-dimensional plane model to model the staring beam scheduling problem, as shown in Fig. 2. A series of hot spots (red spots) are evenly distributed in the plane area. The satellite (blue dot) moves along the fixed orbit (yellow dotted line) at the same speed. When the satellite moves beyond the plane area, it will appear from the other side of the map and continue to move along the track. Each satellite can provide staring services for one or more regions at the same time, which is constrained by the maximum number of beams and the elevation angle between the satellite and the staring region. And each region can also accept the service of multiple satellites.

When the satellite serves a hot spot area, this paper assumes that the satellite can completely cover this area to avoid the discussion of incomplete beam coverage. When dealing with staring beam scheduling, there are three main factors considered in this paper:

1) Satellites should provide services to the nearest hot spots as far as possible to improve the service quality;
2) When the satellite moves out of a hot spot area, other satellites should continue to cover the area to reduce the drop rate of the area;
3) The satellite load capacity and the capacity of hot spot area should be considered in beam scheduling. If the satellite capacity is not enough to fully serve the current region, other nearby satellites should participate in the service to ensure the service quality.

Assume that the capacity of the satellite is $U = \{u_1, u_2, \ldots, u_N\}$, the remaining beam of each satellite is $B = \{b_1, b_2, \ldots, b_M\}$, and the capacity requirement

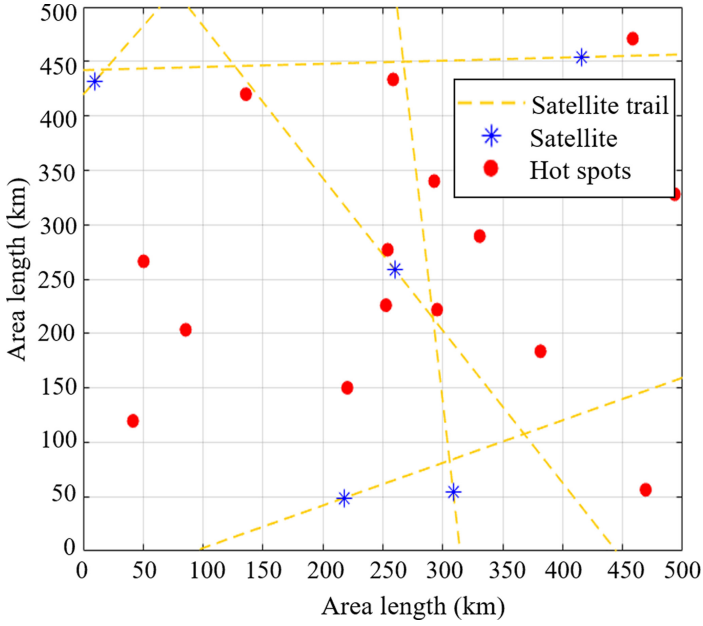

Fig. 2. Two dimensional model of staring beam satellite system.

of hot spot area is $L = \{l_1, l_2, \ldots, l_N\}$. When satellite i is connected with hot spot j, satellite i will provide services for hot spot j as much as possible. At this time, the remaining capacity of satellite i is:

$$s_i' = s_i - \min(u_i, l_j) \tag{1}$$

the remaining capacity requirement of hot spot j is:

$$u_j' = u_j - \min(u_i, l_j) \tag{2}$$

and the number of remaining beams of satellite i is:

$$b_i' = b_i - 1 \tag{3}$$

Considering the two-dimensional plane model we built, the elevation relationship between the hot spot area and the satellite will be transformed into the Euclidean distance between them. If the satellite i coordinate is $(x_{i,1}, x_{i,2})$ and the hot spot area j coordinate is $(y_{j,1}, y_{j,2})$, then the Euclidean distance between the satellite and the hot spot area can generate the weight matrix of M rows and N columns:

$$\boldsymbol{W} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1N} \\ w_{21} & w_{22} & \cdots & w_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ w_{M1} & w_{M2} & \cdots & w_{MN} \end{bmatrix} \tag{4}$$

where $w_{ij} = \sqrt{\sum_{k=1}^{2} (x_{ik} - y_{jk})^2}, x_i \in X, y_j \in Y$. Then we introduce the beam allocation matrix F of satellite and hot spot area as follows:

$$\boldsymbol{F} = \begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1N} \\ f_{21} & f_{22} & \cdots & f_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ f_{M1} & f_{M2} & \cdots & f_{MN} \end{bmatrix} \tag{5}$$

where $f_{m,n} \in \{0, 1\}$, $f_{m,n} = 1$ means that satellite m allocates a beam to hot spot area n. The service radius of the satellite is R. And the matrix F can be obtained by:

$$f_{m,n} = \begin{cases} 0, \ w_{m,n} > R, or \ u_m = 0, or \ l_j = 0 \\ 1, \ others \end{cases} \tag{6}$$

We introduce $c_{i,j}$ to represent the capacity of satellite i allocated to hot spot area j, then the capacity allocation matrix C can be expressed as:

$$\boldsymbol{C} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1N} \\ c_{21} & c_{22} & \cdots & c_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ c_{M1} & c_{M2} & \cdots & c_{MN} \end{bmatrix} \tag{7}$$

In this paper, incomplete service rate, handover rate and insufficient capacity rate are used to measure the performance of beam allocation. Incomplete service rate P_b refers to the proportion of hot spot area whose business requirements can't be met. It can be given by:

$$P_b = \frac{\sum\limits_{j \in \mathcal{N}} g_j}{N} \tag{8}$$

where g_j is:

$$g_j = \begin{cases} 0, \ \sum\limits_{i \in \mathcal{M}} C_{ij} < l_j \\ 1, \ \text{others} \end{cases} \tag{9}$$

Handover rate P_h is used to measure the frequency of satellite switching hot spots in the service process, which is given by:

$$P_h = \frac{\sum\limits_{n \in \mathcal{N}} \sum\limits_{m \in \mathcal{M}} h_{m,n}^t}{\sum\limits_{n \in \mathcal{N}} \sum\limits_{m \in \mathcal{M}} f_{m,n}^t} \tag{10}$$

where $h_{m,n}^t$ is:

$$h_{m,n}^t = \begin{cases} 0, \ f_{m,n}^t = f_{m,n}^{t+1} \\ 1, \ \text{others} \end{cases} \tag{11}$$

P_c is the insufficient capacity rate of hot spot area demand, which is a supplement to P_b. It can be expressed as:

$$P_c = \frac{\sum\limits_{i\in\mathcal{M}}\sum\limits_{j\in\mathcal{N}} c_{i,j}}{\sum\limits_{j\in\mathcal{N}} l_j} \tag{12}$$

Then the staring beam scheduling problem is formulated as follows:

$$\min_{c_{m,n}} \quad P = \alpha_1 P_b + \alpha_2 P_h + \alpha_3 P_c \tag{13}$$

$$\text{s.t.} \quad \sum_{n\in\mathcal{N}} f_{m,n} \le K, \quad \forall m \in \mathcal{M} \tag{13a}$$

$$\sum_{n\in\mathcal{N}} c_{m,n} \le u_m, \quad \forall m \in \mathcal{M} \tag{13b}$$

$$f_{m,n} \in \{0,1\}, \quad \forall m \in \mathcal{M},\ \forall n \in \mathcal{N} \tag{13c}$$

$$c_{m,n} \in [0, u_m], \quad \forall m \in \mathcal{M},\ \forall n \in \mathcal{N} \tag{13d}$$

where α_1, α_2, and α_3 are positive parameters.

3 Multi-agent Deep Q-Learning Algorithm

Deep Q-Learning (DQN) algorithm is a classic and effective reinforcement learning algorithm, which can solve complex problems in many communication scenarios. We extend the deep Q-learning to the multi-agent cases to solve the problem of staring beam scheduling. In the multi-agent DQN model, the learning and decision of each agent are realized by DQN algorithm. $s_{e,t}^{m,k}, a_{e,t}^{m,k}, r_{e,t}^{m,k}$, and $s_{e,t+1}^{m,k}$ represent the state, action, reward and the next state of agent (satellite) i at time t of the e-th training round. The online Q function fitted by neural network and the objective Q function are randomly initialized. With the continuous interaction between the agent and the environment, the generated action sequence is stored in the experience pool. In order to minimize the error function $L(\theta^m)$, a batch of sequences are randomly selected from the experience playback pool every certain interval. $L(\theta^m)$ is given by

$$L(\theta^m) = \left(r_j^m + \gamma \max_{A^i} Q^m\left(s_{j+1}^m, a^m; \theta^{m-}\right) - Q^m\left(s_j^m, a_j^m; \theta^m\right)\right)^2 \tag{14}$$

where γ is the discount factor, θ^{m-} is the parameters of target value network, and θ^m is the parameters of online value network.

The multi-agent deep Q-learning algorithm is shown in Algorithm 1. State $s_{e,t}^m = \left[W_{e,t}^i, a_{e,t-1}^m, a_{e,t-2}^m, C_{e,t}^m\right]$ represents the allocation state of the k-th beam of satellite m of the e-th round during training at the time t. The reward for satellite m performing action $a_{e,t}^{m,k}$ under the k-th beam at the time t in the e-th round is $r_{e,t}^{m,k}$. and it can be expressed as

$$r_{e,t}^{m,k} = -r\_l_{e,t}^{m,k} - a \times r\_b_{e,t}^{m,k} + b \times r\_c_{e,t}^{m,k} \tag{15}$$

Algorithm 1. Multi-Agent Deep Q-Learning Algorithm

Initilization:

- Satellite capacity set S, hot spot area capacity demand set U, satellite beam number set B, satellite coverage radius R, satellite geographic coordinates $(x_{i,1}, x_{i,2})$, $i \in \mathcal{M}$,hot spots area geographic coordinate $(y_{j,1}, y_{j,2})$, $j \in \mathcal{N}$;
- Initialize the status of each satellite $s_{1,1}^{m,k}, m \in \mathcal{M}, k \in \mathcal{K}$;
- Initialize the action value function of each agent with random parameters $Q^i\left(s^i, a^i; \theta^i\right), i \in \mathcal{M}$;
- $epoch = \{1, 2, ..., E\}$, $time = \{1, 2, ..., T\}$;

for $e \in epoch$ **do**
  for $t \in time$ **do**
    for $k \in \mathcal{K}$ **do**
      for $m \in \mathcal{M}$ **do**
        Using ε-greedy based exploration strategy $\pi^{\varepsilon}\left(s_{e,t}^{m,k}\right)$ to get action $a_{e,t}^{m,k}$, the reward $r_{e,t}^{m,k}$, and the transition state $s_{e,t+1}^{m,k}$, store sequence $\left(s_{e,t}^{m,k}, a_{e,t}^{m,k}, r_{e,t}^{m,k}, s_{e,t+1}^{m,k}\right)$ in D^m, update U, L, F, and C;
      end for
      if $t > 200$ and $t\%5 == 0$ **then**
        for $m \in \mathcal{M}$ **do**
          Select a batch of sequence $\left(s_j^{m,k}, a_j^{m,k}, r_j^{m,k}, s_{j+1}^{m,k}\right)$ from D^m randomly;
          Set $y_j^m = \begin{cases} r_j^m \\ r_j^m + \gamma \max_{A^m} Q^m\left(s_{j+1}^m, a^m; \theta^{m-}\right) \end{cases}$;
          $L\left(\theta^m\right) = \left(\mathrm{y}_j - Q^m\left(s_j^m, a_j^m; \theta^m\right)\right)^2$;
          Gradient descent update θ^m from $L\left(\theta^m\right)$;
          update $\theta^{m-} = \theta^m$;
        end for
      end if
    end for
  end for
end for

where a is the penalty coefficient of disconnection, and b is the reward coefficient of connection.

Here $r\_l_{e,t}^{m,k}$ is the distance penalty, which aims to make the satellite serve the nearest area as far as possible. $r\_b_{e,t}^{m,k}$ is the disconnection penalty and $r\_c_{e,t}^{m,k}$ is the connection reward. In this way, the handover rate can be reduced. And the three parameters is given by

$$r\_l_{e,t}^{m,k} = \sqrt{\sum_{k=1}^{2} \left(x_{ik} - y_{jk}\right)^2}, x_i \in X, y_j \in Y \tag{16}$$

$$r\_b_{e,t}^{m,k} = \begin{cases} -La & c_{m,n}^{t-1} = 1 \; and \; c_{m,n}^{t} \neq 1 \\ 0 & \text{else} \end{cases} \tag{17}$$

$$r\_c_{e,t}^{m,k} = \begin{cases} \sqrt{\sum_{j=1}^{2} \left(x_{m,j} - y_{n,k}\right)^2}, x_m \in X, y_{a_n} \in Y & c_{m,n}^{t-1} = c_{m,n}^{t} = 1 \\ 0 & \text{else} \end{cases} \tag{18}$$

By synthesizing the three kinds of rewards and adjusting their coefficients, the agent is expected to learn the corresponding strategies to meet the performance requirements.

4 Simulation Results

In this section, we present the simulation results of proposed strategy what we called multi-agent DQN and compare it with Beam scheduling algorithm based on KM. In this paper, we consider that a reasonable beam scheduling scheme should minimize the sum of the connection distances between the satellite and the hot spot region, that is, to find the minimum value of the edge weight of the bipartite graph $G = (V, E)$. So the problem is transformed into the optimal matching of bipartite graph, which can be solved by the classical KM algorithm.

The simulation parameters are summarized in Table 1. We consider a square area with 500 km side length, where hot spot areas are randomly distributed within the area. When the number and capacity of satellites are fixed, we focus on the impact of the maximum number of beams and the number of hot spots on the algorithm performance. In terms of satellite capacity, refer to StarLink single satellite capacity (17 Gbps) and Iridium system single satellite capacity (7.5 Gbps), the satellite capacity is set to 3 Gbps and the maximum capacity demand of hot spot area is 1 Gbps.

Table 1. Simulation parameters

Parameter	Value
Number of satellites M	5
Number of satellite beams K	3
Maximum service range of satellite R	250 km
Maximum capacity of satellite s_{max}	3 Gbps
Number of hot spots N	15
Maximum capacity demand of hot spots u_{max}	1 Gbps
Area size La^2	$250 \times 250\,\text{km}^2$
Satellite velocity v	10 km/s
Simulation duration T	300 s

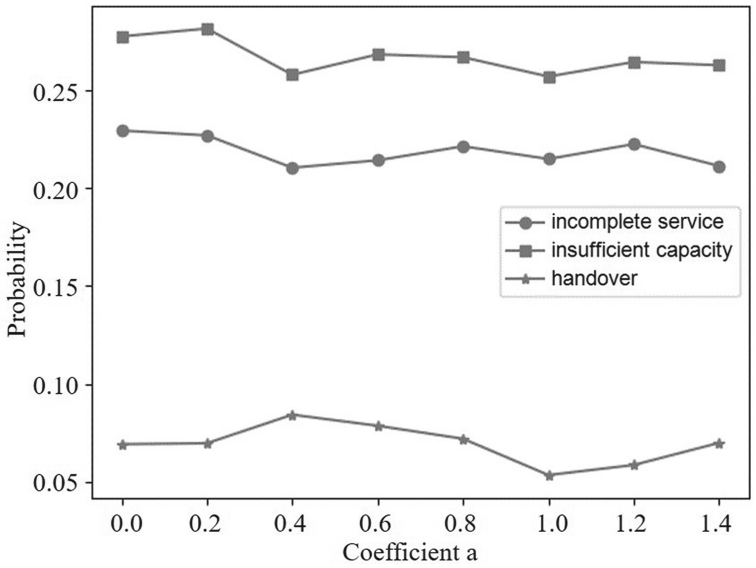

Fig. 3. The performance of the system varies with the penalty coefficient a of disconnection.

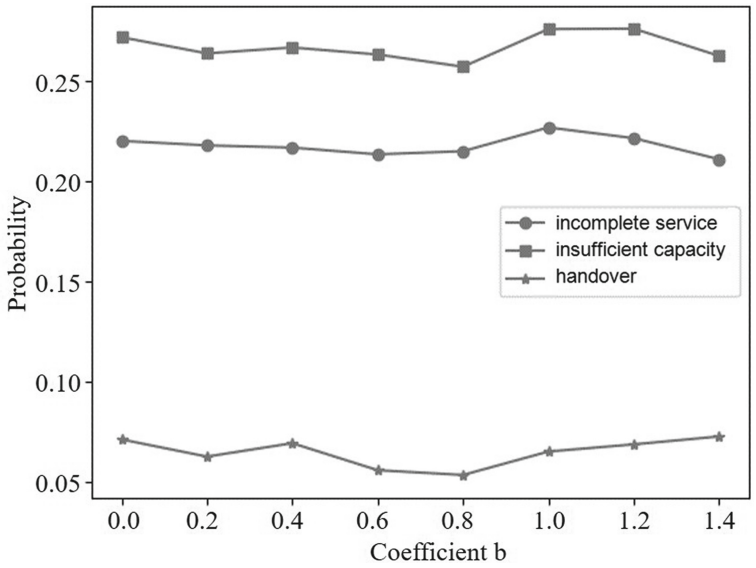

Fig. 4. The performance of the system varies with the reward coefficient b of connection.

In Fig. 3 and Fig. 4, we compared the effect of the penalty coefficient a of disconnection and the reward coefficient b of connection on the beam scheduling performance. It can be found that when $a = 1.0$ and $b = 0.75$, the handover rate of the system reaches the lowest, and the rate of incomplete service and insufficient capacity rate are also at a low value. So we set $a = 1.0$ and $b = 0.75$ in the following simulation.

Figure 5 shows the influence of the number of beams of satellite on the handover rate, incomplete service rate and insufficient capacity rate. With the increase of the maximum number of beams, the performance of KM algorithm and multi-agent DQN algorithm is improved. When the number of beams is small, the performance of KM algorithm is better than that of multi-agent DQN algorithm. However, with the increase of the number of beams, their coverage performance is almost the same, but the multi-agent DQN is significantly better than KM algorithm in the handover rate. This is because the punishment of satellite switching is strengthened in the training, so the satellite connection strategy is adjusted.

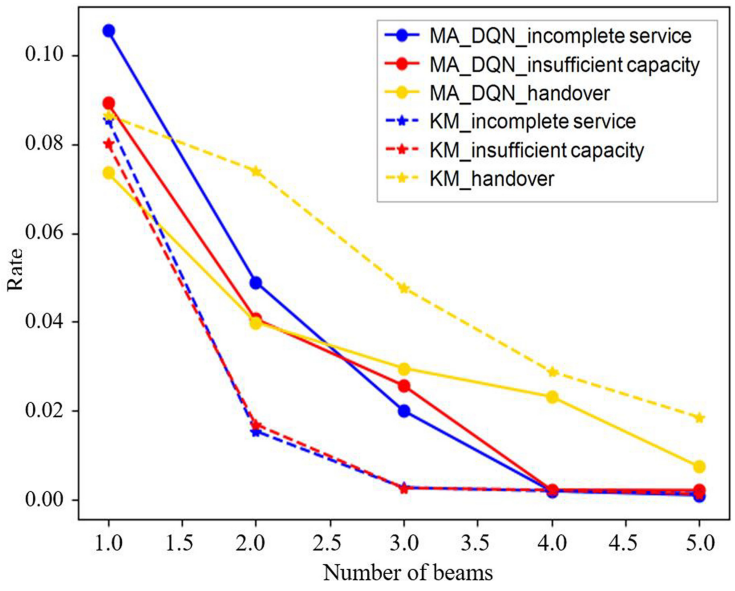

Fig. 5. The performance of the system varies with the maximum number of beams.

Figure 6 shows the influence of the number of hot spot area on the handover rate, incomplete service rate and insufficient capacity rate. As the number of hot spots increases, the performance of both algorithms decreases. The handover rate of multi-agent DQN algorithm is always better than that of KM algorithm. But its coverage performance is not as good as KM. In general, by adjusting the reward setting, the reinforcement learning algorithm achieves better performance

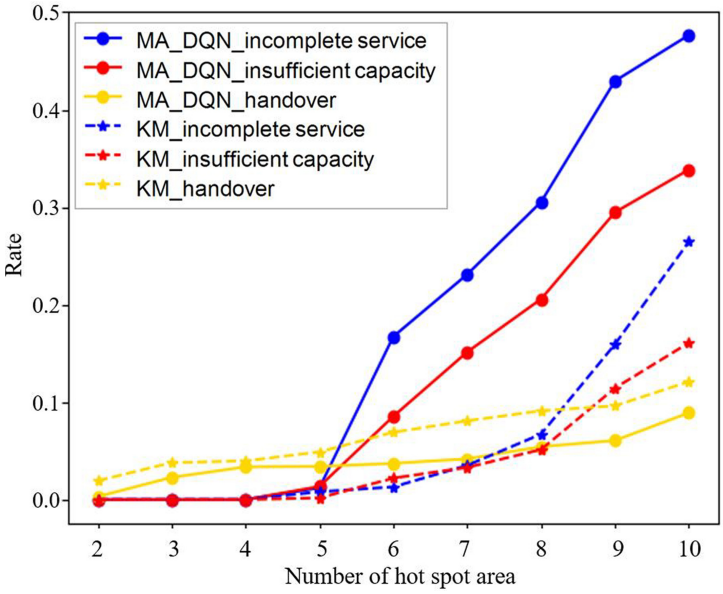

Fig. 6. The performance of the system varies with the number of hot spot area.

in the handover rate. In the future, when the number of satellites is increasing, the coverage performance will not be the main consideration, while reducing the handover rate can greatly reduce the signaling overhead of the system, which indicates that multi-agent DQN algorithm is a desirable staring beam scheduling algorithm.

5 Conclusion

In this paper, we have investigated the staring beam scheduling problem in LEO satellite network. By establishing a two-dimensional model, an optimal problem is proposed under the constraints of the number of satellite beams and satellite capacity. We have solved the problem by multi-agent DQN algorithm. Compared with the beam scheduling algorithm based on KM, multi-agent DQN algorithm can adjust the weight to change the scheduling strategy to meet the optimization requirements. Simulation results have shown that the proposed algorithm can reduce the handover rate, which is of great significance to reduce the network signaling overhead.

References

1. Chen, S., Sun, S., Kang, S.: System integration of terrestrial mobile communication and satellite communication - the trends, challenges and key technologies in B5G and 6G. China Commun. **17**(12), 156–171 (2020). https://doi.org/10.23919/JCC.2020.12.011

2. Rinaldi, F., et al.: Non-terrestrial networks in 5G beyond: a survey. IEEE Access **8**, 165178–165200 (2020). https://doi.org/10.1109/ACCESS.2020.3022981
3. Giordani, M., Zorzi, M.: Non-terrestrial networks in the 6G era: challenges and opportunities. IEEE Netw. **35**(2), 244–251 (2021). https://doi.org/10.1109/MNET.011.2000493
4. Restrepo, J., Maral, G.: Cellular geometry for world-wide coverage by non-geo satellites using "earth-fixed cell" technique. In: Proceedings of GLOBECOM 1996. 1996 IEEE Global Telecommunications Conference, vol. 3, pp. 2133–2137 (1996). https://doi.org/10.1109/GLOCOM.1996.592010
5. Wu, Y., Hu, G., Jin, F., Zu, J.: A satellite handover strategy based on the potential game in LEO satellite networks. IEEE Access **7**, 133641–133652 (2019). https://doi.org/10.1109/ACCESS.2019.2941217
6. Baik, J.S., Kim, J.H.: Analysis of the earth fixed beam duration in the LEO. In: 2021 International Conference on Information Networking (ICOIN), pp. 477–479 (2021). https://doi.org/10.1109/ICOIN50884.2021.9333981
7. Boukhatem, L., Beylot, A., Gaiti, D., Pujolle, G.: Performance analysis of dynamic and fixed channel allocation techniques in a LEO constellation with an "earth-fixed cell" system. In: Globecom 2000 - IEEE. Global Telecommunications Conference. Conference Record (Cat. No.00CH37137), vol. 2, pp. 1145–1149 (2000). https://doi.org/10.1109/GLOCOM.2000.891316
8. Luong, N.C., et al.: Applications of deep reinforcement learning in communications and networking: a survey. IEEE Commun. Surv. Tutor. **21**(4), 3133–3174 (2019). https://doi.org/10.1109/COMST.2019.2916583
9. Li, Z., Xie, Z., Liang, X.: Dynamic channel reservation strategy based on DQN algorithm for multi-service LEO satellite communication system. IEEE Wirel. Commun. Lett. **10**(4), 770–774 (2021). https://doi.org/10.1109/LWC.2020.3043073