

심층 강화 학습을 사용한 저궤도 위성 네트워크 핸드오버

Thanh Phung Truong, 이정화, Arooj Masood, 조성래

중앙대학교 컴퓨터공학과

{tptruong, jeonghwa, arooj}@uclab.re.kr, srcho@cau.ac.kr

Handover in LEO Satellite Networks using Deep Reinforcement Learning

Thanh Phung Truong, Jeonghwa Lee, Arooj Masood, Sungrae Cho

Department of Computer Science and Engineering, Chung-Ang Univ.

Abstract

Low Earth orbit (LEO) satellites have an important role to the future communication networks, where the ground users can be covered by multiple satellites. Since the LEO satellite has its speed and moving around on the orbit, the handover problem need to be considered to cover the quality of the service. Thus, in this paper, we consider a LEO satellite network with handover problem and apply a deep reinforcement learning framework to solve the optimization problem.

I. Introduction

Satellite communications can provide seamless wireless signal coverage so as to supplement and extend terrestrial communications [1]. The LEO satellites are in constant motion that the LEO satellite can cover end users for a limited period of time [2]. Due to the moving characteristic of the LEO satellite, the quality of the LEO satellite – ground users may not be satisfied. Thus, a handover problem in LEO satellite is an important problem when considering a satellite network.

In this work, we investigate an LEO satellite system model that the satellites cover and serve the ground users (UEs), in which, the handover optimization problem is proposed. Then, we apply a deep reinforcement learning framework to the model to solve the problem to warrant the quality of the system.

In section II, we form the system model and established the problem and solve the problem by applying deep reinforcement learning in section III. Finally, we conclude this work in section IV.

II. System Model and Problem Formulation

The scenario of the satellite handover is illustrated in Figure 1, which includes M LEO satellites $\{S_1, S_2, \dots, S_M\}$ are serving N ground users (UEs) $\{U_1, U_2, \dots, U_N\}$ on the Earth surface. We denote $\theta_{m,n}$ is the elevation angle between user n and satellite m . To guarantee the link quality of the satellite – user communication, the elevation angle must satisfy the serving elevation angle constraint as

$$\theta_{m,n} \geq \theta_0, \forall m \in M, \forall n \in N \quad (1)$$

where θ_0 is the minimum elevation angle.

We consider the handover problem in a specific time period T , it can be divided into E coverage sections as $\{t_1, t_2, \dots, t_E\}$, each coverage section represents the state changing times of the satellite coverage.

To perform the coverage of the satellite, we denote $c_{n,m}^e$ is the coverage indicator between satellite m and user n during timing section t_e as

$$c_{n,m}^e = \begin{cases} 1, & \text{user } n \text{ is covered by satellite } m \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

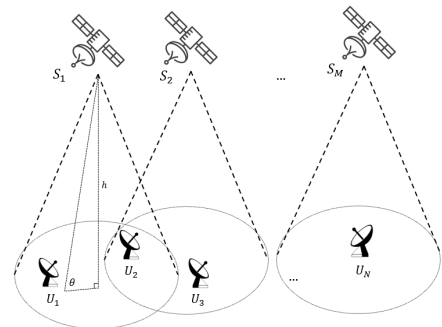


Figure 1. The Satellite Handover System Model

The set of satellites that cover user n during time section t_e is given by

$$S_n^e = \{m \mid c_{n,m}^e = 1, m \in M\}, \forall n \in N \quad (3)$$

and usage indicator in the section t_e is introduced as

$$x_{n,m}^e = \begin{cases} 1, & \text{user } n \text{ is served by satellite } m \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

The bandwidth of each satellite is divided into C channels with the same bandwidth, and each user can only use one channel of a satellite during a time section, then, the constraints of the usage bandwidth are

$$\sum_{n \in N} x_{n,m}^e \leq C, \forall m \in M \quad (5)$$

$$\sum_{m \in M} x_{n,m}^e \leq 1, \forall n \in N \quad (6)$$

We aim to maximize the total number of UEs that are served by the satellite while satisfying the serving elevation angle constraint and the usage bandwidth constraints. Then, the optimization problem is defined as

$$\max_{x_{n,m}^e} \sum_{n \in N, m \in M} x_{n,m}^e \quad (7)$$

$$\text{s.t.} \quad \theta_{m,n} \geq \theta_0, \forall m \in M, \forall n \in N \quad (7a)$$

$$x_{n,m}^e = \{0,1\} \quad (7b)$$

$$\sum_{n \in N} x_{n,m}^e \leq C, \forall m \in M \quad (7c)$$

$$\sum_{m \in M} x_{n,m}^e \leq 1, \forall n \in N \quad (7d)$$

which (7a) represents the serving elevation angle constraint, (7b) indicates two possible cases of the usage indicator and (7c), (7d) are the constraints of the usage bandwidth.

III. Deep Reinforcement Learning for Satellite Handover

To show the solution, we use a multi-agent Q-Learning algorithm to train the users to choose the best action. The reinforcement learning is formulated as

- *State*: The state of the system as the time t_e includes the coverage indicator of all satellites for all UEs $c_{n,m}^e$ and the number of available channels of the satellite m in this time section L_m^e . Thus, the total number of states is $(M \times N) + M = M(N + 1)$. Then, the state is formulated as

$$s = \{c_{n,m}^e, L_m^e\}, \forall n \in N, \forall m \in M \quad (8)$$

- *Action*: The action of each agent k in the time section t_e is defined as the usage indicator, which selects the serving satellite. Thus, the number of actions of each agent is $M \times N$. It is formulated as

$$a = x_{n,m}^e, \forall n \in N, \forall m \in M \quad (9)$$

- *Reward*: Since the objective is to maximize the total number of UEs that are served by the satellite, the decision reward function is

$$r_d = \sum_{n \in N, m \in M} x_{n,m}^e \quad (10)$$

To satisfy the constraints, we set the penalty value as 10 and define some penalty indicators as

$$p_c = \begin{cases} 1, & \text{if } x_{n,m}^e = 1 \text{ and } c_{n,m}^e = 0 \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

$$p_u = \begin{cases} 1, & \sum_{n \in N} x_{n,m}^e > C, \forall m \in M \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

Then, the total reward function of the model at the time section t_e is formulated as

$$r = r_d - 10 \times (p_c + p_u) \quad (13)$$

Based on the action-value ($Q(s_t, a_t | \theta)$), the agents will select the best action (or the offloading decision) with the biggest action-value to act to the environment. The neural network is updated by minimizing the error between the action-value of each action and the target value. The target value of each agent is calculated as

$$y = r + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1} | \theta) \quad (14)$$

The training algorithm of multi-agent Q learning is as follow

Algorithm 1: Deep Reinforcement Learning for LEO Satellite Handover

```

1 Set up the environment
2 Initialize state  $s_n^0$  and state-action value  $Q_n(s_n^0, a^0)$ 
3 while  $t^e < T$  do
4   for agent  $n$  in  $N$  do
5     Observe  $s_n^e = c_{n,m}^e, L_m^e$ 
6     Choose action  $a_n^e = x_{n,m}^e$ 
7     Obtain action of all agent  $a^e = a_1^e, a_2^e, \dots$ 
8     Receive the reward  $r_n^e(s_n^e, a^e)$  and the next state
9     update the Q-values
10   $t^e = t^e + 1$ 

```

IV. Conclusion

In this paper, a deep reinforcement learning approach is used to optimize a handover problem in the LEO satellite networks. We proposed a LEO satellite scenario and formulated a handover optimization problem, then, we applied a multi-agent Q learning framework to solve the problem. In the future, based on this work, we will consider a more practical scenario, such as power, complex usage channels, timing, etc., then we will perform simulation on the specific satellite constellation to estimate the performance of multi-agent Q learning when applying to handover satellite problem.

ACKNOWLEDGMENT

본 연구는 2019년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임

[NRF-2019R1A2C1090447]

References

- [1] S. Zhang, A. Liu and X. Liang, "A Multi-objective Satellite Handover Strategy Based on Entropy in LEO Satellite Communications," 2020 IEEE 6th International Conference on Computer and Communications (ICCC), 2020, pp. 723-728, doi: 10.1109/ICCC51575.2020.9344916.
- [2] Z. Wu, F. Jin, J. Luo, Y. Fu, J. Shan and G. Hu, "A Graph-Based Satellite Handover Framework for LEO Satellite Communication Networks," in IEEE Communications Letters, vol. 20, no. 8, pp. 1547-1550, Aug. 2016, doi: 10.1109/LCOMM.2016.2569099.