Low Earth Orbit Satellite Scheduling Optimization based on Deep Reinforcement Learning

Junyoung Kim[1], Soyi Jung[2]
[1]*Dept. Artificial Intelligence Convergence Network, Ajou University, Suwon, 16499, South Korea*
[2]*Dept. Electrical and Computer Engineering, Ajou University, Suwon, 16499, South Korea*
{junzero0615, sjung}@ajou.ac.kr

***Abstract*—In next-generation 6G scenarios, non-terrestrial networks employing low Earth orbit (LEO) satellites will be pivotal in achieving ultra-wide coverage, ultra-connectivity, and ultra-precision. Although LEO satellites provide comprehensive global coverage, their rapid mobility introduces frequent handovers, requiring sophisticated scheduling to maintain uninterrupted service. This paper proposes a deep reinforcement learning-based scheduling algorithm in order to improve service rate and continuity for terrestrial users in multi-LEO environments.**

***Index Terms*—non-terrestrial networks, LEO satellites, handover, deep reinforcement learning, scheduling optimization**

I. Introduction

The 3rd generation partnership project (3GPP) has proposed a conditional handover procedure for NTN to address challenges associated with the rapid mobility of LEO satellites and the movement of UE. This procedure is triggered when specific conditions are met, and various handover triggering events have been defined [1]. Among these, the A3 event, based on signal strength, is widely used. Recent research has also explored handover techniques that consider the distance to the cell center or satellite [2]. However, simple event-based designs and traditional mathematical optimization methods have limitations in performing handovers at the optimal time in multi-satellite environments, which is crucial for maximizing UE throughput. Accordingly, this paper proposes a deep reinforcement learning-based LEO scheduling technique to enhance UE service continuity and received signal strength by enabling terrestrial UEs to perform handovers at the optimal time. The performance of the proposed method is analyzed [3] [4].

II. Handover mechanism for Deep Reinforcement Learning Implementation

A. Handover Mechanism

In NTN, it is essential for terrestrial UEs to perform handovers at the appropriate time, considering the rapid mobility of LEO satellites, to maintain continuous service. The A3 event, which is commonly used for handovers, is defined by equation (1) as the condition where a handover is triggered when the received signal strength of the target cell (u_t) exceeds the sum of the received signal strength of the currently serving cell (u_s) and a predefined threshold.

$$u_t > u_s + threshold, \qquad (1)$$

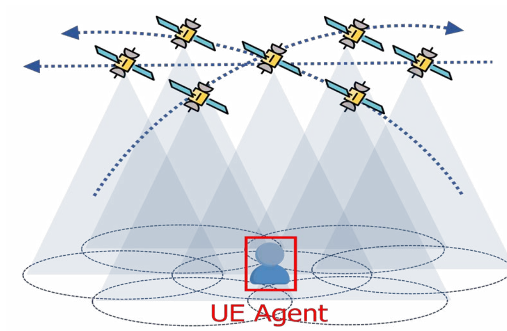

Fig. 1: System model

B. Proposed Deep Reinforcement Learning Algorithm

This paper aims to optimize the scheduling problem in which a terrestrial UE agent performs handovers at the optimal time to maximize received signal strength, based on the handover mechanism defined in Section II-A, using a deep-q-network (DQN) reinforcement learning algorithm. In the proposed DQN approach, which is designed to maximize the received signal strength as LEO satellites move, the Markov decision process (MDP) state (s_t) is defined as a combination of the distance between the UE and each LEO (s_d), the elevation angle between the UE and LEO (s_e), the UE's received signal strength (s_g), and the UE's signal-to-interference-plus-noise ratio (SINR) value (s_r), such that $s_t = [s_d, s_e, s_g, s_r]$. Notably, in the process of selecting a LEO, the UE classifies a LEO as connectable only if the value of s_e meets the minimum requirement of 30°. From a scheduling perspective, where the UE selects the LEO satellite for service, the UE's action (a_t) is defined as follows: if the signal strength of the currently serving LEO (u_s) is sufficient, the action is to maintain the current connection. If a candidate satellite satisfies the condition in equation (1), the action is to perform a handover. The action to maintain the connection is represented as 0, and the action to select a candidate satellite is represented as 1, expressed as $a_t = [0, 1]$.

The objective of the DQN-based LEO scheduling method proposed in this paper is to enhance system throughput by maximizing the UE's received signal strength through timely handovers. To achieve this, the reward (r_t) assigned to the UE agent is decomposed into the following components: received signal strength (r_g), UE data rate (r_a), handover cost (r_c), and

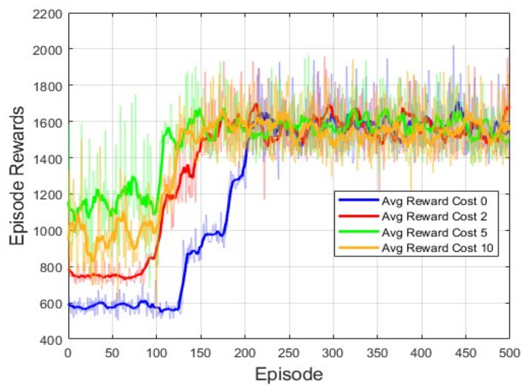

Fig. 2: Training performance of each handover costs

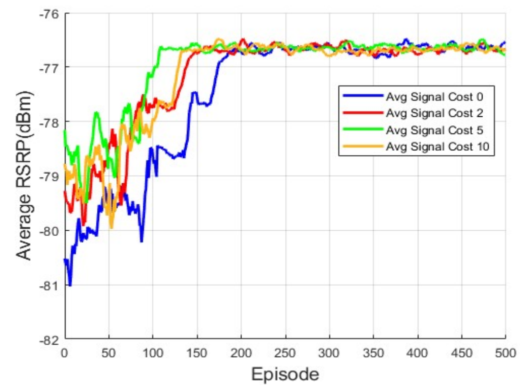

Fig. 3: Agent received signal strength of each handover cost

TABLE I: RL simulation parameters

Parameter	**Value**
Discount factor γ	0.98
Learning rate	0.001
Batch size, Buffer size	32, 50000

TABLE II: LEO environment parameters

Parameter	**Value**
LEO altitude, Number of LEO	600, 9
Frequency band	Ka-band(20GHz)
UE characteristics	VSAT
LoS probability scenario	Suburban & Rural

service duration time (r_s), with the total reward defined as $r_t = r_g + r_a + r_c + r_s$. Specifically, r_a is defined by the relationship among the satellite channel bandwidth (B), the number of UE (N_t), and s_r, as shown in (2).

$$r_a = \left(\frac{B}{N_t}\right) \times \log_2\left(1 + s_r\right), \tag{2}$$

Additionally, to prevent frequent handovers, a penalty term for handover cost, denoted as r_c, is introduced to minimize the number of handovers and maximize r_s.

A key feature of DQN is its ability to enhance learning stability by using experience replay samples stored in replay memory as input to both the Q-network and the target network. This setup allows the agent to select actions that minimize the loss function (L) by periodically duplicating the Q-network to create a target network, enabling faster learning. The target value, which serves as the reference for forming the target network, is defined in equation (3), while the loss value of the DQN based on the agent's actions is defined in equation (4).

$$Y_t = r_t + \gamma \max Q(s_{t+1}, a'; \theta), \tag{3}$$

$$L = (Y_t - Q(s_t, a_t; \theta))^2. \tag{4}$$

III. Performance Evaluation

Experiments were conducted in a Python environment to develop a deep reinforcement learning-based LEO scheduling model, utilizing the training parameters listed in Table I. To assess the performance of the scheduling technique, models were created with varying handover costs as part of the reward structure, with the outcomes shown in Fig. 2. As illustrated in Fig. 2, setting the handover cost to zero results in frequent handovers, leading to slower convergence. Conversely, applying a moderate handover cost leads to the fastest convergence compared to other cost values. Furthermore, Fig. 3 depicts the received signal strength of the terrestrial UE agent in relation to the handover cost. The training scenarios were aligned with 3GPP standards, as detailed in Table II. The findings show that the received signal strength of the agent converges most rapidly when a moderate handover cost is applied.

IV. Conclusion

This paper explores a DQN-based scheduling technique based on the handover mechanism for terrestrial UEs, considering the rapid mobility of LEO satellites. The proposed approach sets the groundwork for future studies on predicting optimal handover timing in scenarios with mega-constellations of satellites and numerous terrestrial users, with possible extensions utilizing multi-agent deep reinforcement learning methods.

Acknowledgment

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT)(RS-2024-00359330).

References

[1] 3GPP TR 38.821 v16.1.0, "Solutions for NR to support non-terrestrial networks (NTN) (Release 16)," 3rd Generation Partnership Project (3GPP), Technical Report 38.821, May 2021.

[2] E. Juan, M. Lauridsen, J. Wigard, and P. Mogensen, "Handover solutions for 5G low-earth orbit satellite networks," *IEEE Access*, vol. 10, pp. 93 309–93 325, August 2022.

[3] Y. Cao, S.-Y. Lien, and Y.-C. Liang, "Deep reinforcement learning for multi-user access control in non-terrestrial networks," *IEEE Transactions on Communications*, vol. 69, no. 3, pp. 1605–1619, March 2021.

[4] S. He, T. Wang, and S. Wang, "Load-aware satellite handover strategy based on multi-agent reinforcement learning," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, 2020, pp. 1–6.