

# Mobility-Aware Multi-Agent Learning for Efficient Handover in Multi-Layer Satellite Network

Junyoung Kim<sup>†</sup>, Huiyeon Jang<sup>†</sup>, and Soyi Jung<sup>‡</sup>

<sup>†</sup>Dept. Artificial Intelligence Convergence Network, Ajou University, Suwon, 16499, South Korea

<sup>‡</sup>Dept. Electrical and Computer Engineering, Ajou University, Suwon, 16499, South Korea

{junzero0615, timd0801, sjung}@ajou.ac.kr

**Abstract**—In 6G networks, providing ultra-space, ultra-connectivity, and ultra-precision services is a key objective, driving active research and standardization efforts. Among these, non-terrestrial networks (NTNs), particularly those using low Earth orbit (LEO) satellites, have gained significant attention. However, the high mobility of satellites and the harsh space environment pose challenges to seamless service for terrestrial user equipment (UE), especially mobile UE (MUE). This paper proposes a reinforcement learning-based handover optimization method for multi-layer satellite networks. The scheme utilizes long short-term memory (LSTM) networks to predict the time-series mobility of satellites and MUEs, and applies a multi-agent deep Q-network (MADQN) to optimize handover decisions. Simulation results show that the proposed method improves the quality of service (QoS) for MUEs and enhances overall system throughput.

**Index Terms**—satellite communication, handover, multi-layer satellite scenario, long-short term memory prediction, multi-agent deep reinforcement learning.

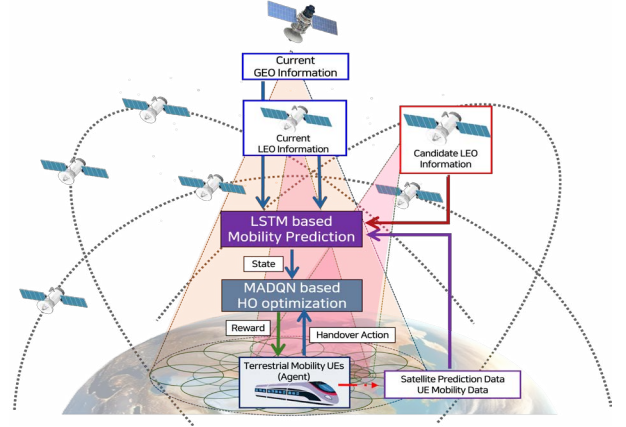


Fig. 1: GEO-LEO Multi-layer satellite networks.

## I. INTRODUCTION

In next-generation 6G networks, extensive efforts are underway to establish non-terrestrial networks (NTNs) for ultra-space communication coverage. In particular, low Earth orbit (LEO) satellite communications are regarded as a key NTN platform, with active standardization in the 3rd generation partnership project (3GPP) [1], [2]. However, the high mobility of satellites poses challenges to seamless service for terrestrial user equipment (UE), especially mobile UE (MUE), where handover (HO) decisions must account for both satellite and terminal mobility [3], [4].

This paper proposes CLM-HO, a handover optimization scheme for a multi-layer satellite network (GLMLN) with coexisting GEO and LEO satellites. The scheme employs a convolutional neural network (CNN) to analyze satellite and MUE mobility patterns, followed by a long short-term memory (LSTM) network to predict MUE mobility [5]. A multi-agent deep Q-network (MADQN) is then used to optimize handover (HO) decisions, ensuring MUE quality of service (QoS). Simulation results show that CLM-HO significantly improves key performance indicators (KPIs) compared to conventional MADQN-based methods [6].

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT)(RS-2024-00359330) (Corresponding authors: Soyi Jung)

## II. SYSTEM MODEL

In this study, we propose a mobility prediction and reinforcement learning-based handover optimization scheme for a GLMLN, as illustrated in Fig. 1. The LEO satellite set as  $\mathbf{L} = \{l | l = 1, 2, \dots, N_l\}$  and the GEO satellite set is denoted as  $\mathbf{G} = \{g | g = 1, 2, \dots, N_g\}$ . The MUEs are traveling at altitude  $h_u$  and velocity  $v_u$ , with the set of all MUEs expressed as  $\mathbf{M} = \{u | u = 1, 2, \dots, N_u\}$ .

The proposed CLM-HO scheme consists of two main procedures: 1) *MUE mobility prediction and data generation using CNN-LSTM* and 2) *handover optimization using MADQN based on the predicted data*. In the first stage, the time-series orbital data of satellites are utilized to extract features and predict MUE mobility through a long short-term memory (LSTM) network, which is well-suited for sequential data analysis. To address the limitations of LSTM in modeling long sequences, a convolutional neural network (CNN) is incorporated to capture spatial relationships within the input data and prevent overfitting.

The predicted mobility data from the CNN-LSTM model are then incorporated as the state variable,  $s_{pd}^u(t)$ , in the second stage, *handover optimization using CLM-HO*. The reinforcement learning (RL) agent, representing the MUE, selects actions  $a_{pd}^u(t)$  based on  $s_{pd}^u(t)$  to maximize the cumulative reward  $r_{pd}^u(t)$ . This approach enables optimal handover de-

TABLE I: Simulation parameters.

| Definition                   | Value             |
|------------------------------|-------------------|
| Operating frequency (DL)     | Ka-band (20 GHz)  |
| Satellite altitude(GEO, LEO) | 35,786 km, 600 km |
| $N_g, N_l, N_u$              | 1, 20, 4          |
| Batch size                   | 256               |
| Learning rate                | 0.0001            |
| Replay buffer                | 50,000            |

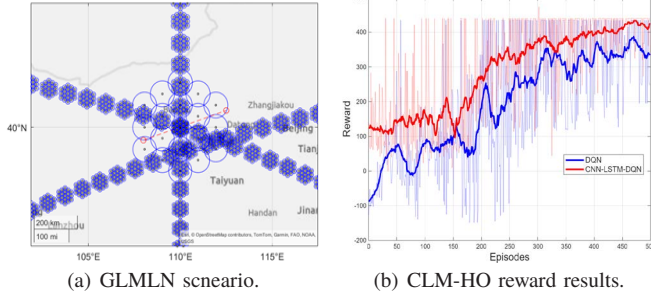


Fig. 2: GLMLN simulation environments and performance.

cisions, improving both service continuity and overall system throughput. The RL process follows a Markov decision process (MDP) defined by the state  $s_{pd}^u(t)$ , action  $a_{pd}^u(t)$ , and reward  $r_{pd}^u(t)$ , as detailed below:

- $s_{pd}^u(t) = \{\mathbf{S}_v(t), \mathbf{S}_d(t), \mathbf{M}_u(t)\}$ : The observation space available to the MUE agent in the GLMLN consists of three components: serving satellite information,  $\mathbf{S}_v(t)$ ; candidate satellite information,  $\mathbf{S}_d(t)$ ; and current MUE information,  $\mathbf{M}_u(t)$ .
- $a_{pd}^u(t) = \{a_u^m(t), a_u^h(t)\}$ : The MUE agent determines its action based on the current state  $s_{pd}^u(t)$ , where, upon satisfying the A3 HO triggering condition, it selects either to maintain the connection with the current serving satellite,  $a_u^m(t)$ , or to perform a handover to a candidate satellite,  $a_u^h(t)$  [7].
- $r_{pd}^u(t) = \omega_{pd}^u(t) + \eta \phi_{pd}^u(t) - C$ : To ensure the QoS of MUEs, the reward function  $r_{pd}^u(t)$  is formulated based on the relationship between the data rate,  $\omega_{pd}^u(t)$ , and the time of stay,  $\phi_{pd}^u(t)$ . In addition, a HO cost,  $C$ , is introduced to penalize frequent handovers, thereby encouraging handover decisions at appropriate times.

### III. PERFORMANCE EVALUATION

For the performance evaluation of the proposed CLM-HO scheme in the GLMLN, a simulation scenario with coexisting GEO and LEO satellites was designed, as depicted in Fig. 2(a). Each satellite is configured to generate 19 beams, and all channel models and parameters are adopted from the 3GPP standard specifications [1], [2]. Simulations were executed on a system equipped with an AMD Ryzen 7900X CPU and an NVIDIA GeForce RTX 4070 Super GPU.

The performance of the reinforcement learning model is evaluated based on reward convergence, as shown in Fig. 2(b). The proposed CLM-HO scheme achieves faster convergence,

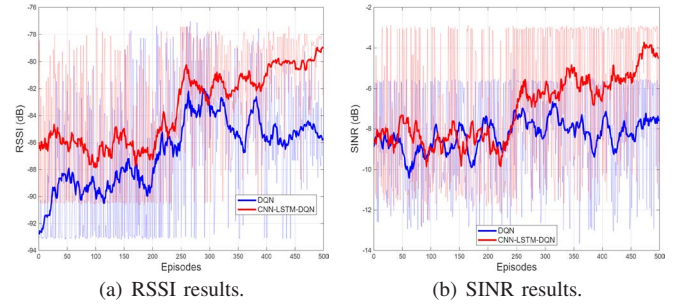


Fig. 3: GLCN simulation environments.

starting around 200 episodes, and demonstrates improved reward performance compared to the baseline MADQN method, which does not incorporate mobility prediction data. Additionally, KPIs including reference signal strength (RSSI) and signal-to-interference-plus-noise ratio (SINR), are analyzed, as illustrated in Fig. 2. As shown in Fig. 3(a), the proposed CLM-HO achieves superior KPI performance compared to MADQN, with further improvements observed in SINR, as depicted in Fig. 3(b). These results confirm that incorporating mobility prediction data into the CLM-HO framework enhances KPI performance relative to conventional reinforcement learning methods.

### IV. CONCLUSION

This paper proposed CLM-HO, a handover optimization scheme based on reinforcement learning incorporating mobility prediction data. The proposed method demonstrated improvements across all KPIs compared to conventional reinforcement learning approaches. As future work, we plan to extend this study by integrating reinforcement learning for human feedback (RLHF) to optimize parameters and weighting factors, enabling handover strategies that better accommodate diverse UE.

### REFERENCES

- [1] 3GPP TR 38.811 v16.1.0, "Study on new radio (NR) to support non-terrestrial networks (Release 15)," 3rd Generation Partnership Project (3GPP), Technical Report 38.811, September 2020.
- [2] 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.
- [3] J. Wu, S. Su, X. Wang, J. Zhang, and Y. Gao, "Accelerating handover in mobile satellite network," in *Proc. IEEE Conference on Computer Communications (INFOCOM)*, Vancouver, Canada, May 2024, pp. 531–540.
- [4] G. S. Kim, Y. Cho, S. Park, S. Jung, and J. Kim, "Quantum multiagent reinforcement learning for joint cube satellites and high-altitude long-endurance aerial vehicles in SAGIN," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 61, no. 4, pp. 9490–9510, August 2025.
- [5] 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, November 2021.
- [6] S. He, T. Wang, and S. Wang, "Load-aware satellite handover strategy based on multi-agent reinforcement learning," in *Proc. GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, 2020, pp. 1–6.
- [7] 3GPP TS 38.331 v18.3.0, "Radio resource control (RRC) protocol specification (Release 18)," 3rd Generation Partnership Project (3GPP), Technical Specification 38.331, October 2024.