Optimized Handover Management for Reliable Connectivity in GEO-LEO Satellite Networks via Predictive Reinforcement Learning

Huiyeon Jang[1], Junyoung Kim[1], Minsu Shin[2], In-Sop Cho[2], Soyi Jung[3]
[1]*Dept. Artificial Intelligence Convergence Network, Ajou University, Suwon, 16499, South Korea*
[2]*Electronics and Telecommunications Research Institute, Daejeon, 34129, South Korea*
[3]*Dept. Electrical and Computer Engineering, Ajou University, Suwon, 16499, South Korea*
{timd0801, junzero0615}@ajou.ac.kr, {msshin, lookatstar}@etri.re.kr, sjung@ajou.ac.kr

***Abstract*—This paper presents a novel methodology to optimize the handover decision of mobile terminals (MTs) within a cooperative geostationary Earth orbit (GEO) and low Earth orbit (LEO) satellite network. Unlike prior research, which primarily focuses on inter-LEO satellite handovers for stationary terminals, this work addresses the handover challenges associated with MTs in a hybrid GEO-LEO satellite architecture. The proposed framework employs a convolutional neural network (CNN)-long-short-term memory (LSTM) encoder-decoder model to accurately predict the reference signal received power (RSRP) of potential handover target satellites, enabling more informed decision-making. A multi-agent double deep Q-network (MADDQN) algorithm is implemented to determine the optimal handover target, leveraging predicted RSRP data while considering factors such as network load and satellite connection duration. The proposed approach minimizes unnecessary handovers, enhances data throughput for MTs, and ensures stable connectivity and quality of service under diverse network conditions. The proposed method contributes to handover optimization in cooperative GEO-LEO satellite networks, offering valuable insights for next-generation satellite communication systems.**

***Index Terms*—Satellite communication networks, handover, multi-agent reinforcement learning**

I. Introduction

With the rapid advancement of mobile communication technologies, the number of communication users and the demand for higher-quality services have significantly increased [1], [2]. To address these demands, satellite networks have emerged as a key technology to complement terrestrial networks by providing extensive wireless coverage and enabling real-time communication capabilities [3]–[5]. In particular, low Earth orbit (LEO) satellite networks have gained considerable attention due to their lower latency, reduced power consumption, and enhanced scalability compared to geostationary Earth orbit (GEO) satellite systems. These characteristics make LEO satellites well-suited for high-capacity, low-latency applications. However, the inherent mobility of LEO satellites requires frequent handovers to maintain seamless connectivity and ensure uninterrupted service [6]–[8].

To address these challenges, a cooperative GEO-LEO satellite network architecture has been proposed, leveraging the fixed positioning of GEO satellites for reliable connectivity and the low-latency performance of LEO satellites for enhanced communication efficiency [9], [10]. This hybrid architecture facilitates inter-satellite collaboration to optimize service quality. However, the heterogeneous and cross-domain nature of this cooperative network introduces significant complexity, making handover optimization a critical research challenge.

Extensive studies have explored efficient handover mechanisms for LEO satellite networks. Notably, deep reinforcement learning (DRL) techniques have been applied to train user equipment (UE) to predict successful handovers based on predefined satellite orbital patterns. These methods have improved handover timing, reduced access delays, and minimized collision rates while maintaining high success rates [11]. Further advancements introduced multi-agent deep Q-learning (MADQL) to optimize handover decisions by considering satellite load and remaining visibility time, thereby minimizing unnecessary handovers in dynamic traffic environments. A multi-agent double deep Q-network (MADDQN) framework was later developed to enhance data transmission efficiency and reduce handover frequency by leveraging local state information, including satellite visibility time, active connections, estimated transmission rates, packet queue status, and neighboring UEs' handover decisions [13]. Additionally, a MADQL-based load-balancing strategy was proposed to dynamically adjust handover decisions based on satellite coverage, load conditions, and visibility time, resulting in reduced handover frequency and network congestion prevention [14].

Despite these advancements, most research has primarily focused on optimizing handovers for stationary UEs within LEO satellite networks. The optimization of handovers for mobile terminals (MTs) in a cooperative GEO-LEO satellite network remains a relatively unexplored area. To address this gap, this study proposes a novel handover optimization framework that minimizes handover frequency while maximizing data rate for MTs operating within a hybrid GEO-LEO satellite network. The proposed approach employs a convolutional neural network (CNN)-long short-term memory (LSTM) encoder-decoder model to predict the reference signal received power (RSRP) of potential handover target satellites. Using these predictions,

ISBN 978-3-903176-73-7

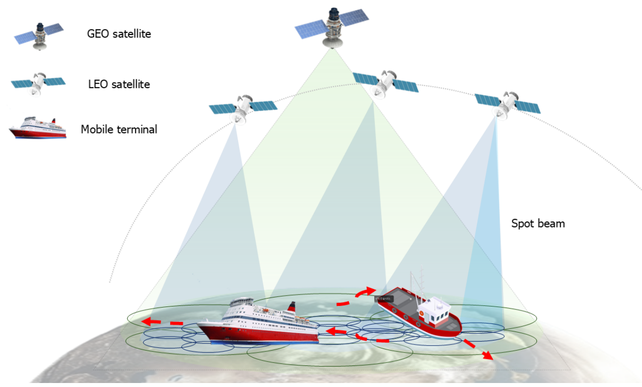

Fig. 1: Architecture of the GEO-LEO satellite network.

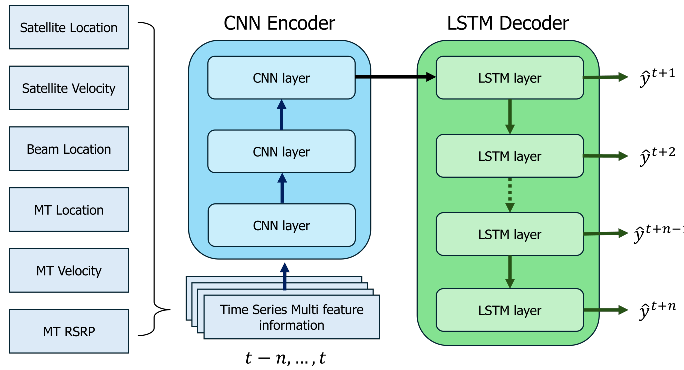

Fig. 2: CNN-LSTM encoder decoder model.

a MADDQN framework is implemented to determine the optimal target satellite, considering network load and satellite connection duration. By integrating deep learning-based prediction with RL-driven decision-making, the proposed method enhances handover efficiency, improves data throughput, and ensures stable connectivity, offering valuable insights for next-generation satellite communication systems.

II. System Model

A. Satellite Network Model

As illustrated in Fig. 1, we consider a multi-beam satellite communication system consisting of N LEO satellites operating in conjunction with a single GEO satellite. This network model focuses on a specialized scenario where MTs are deployed in environments where terrestrial base stations are unavailable. Each satellite functions as a regenerative node and is equipped with inter-satellite communication links that utilize the Ka-band frequency spectrum [8].

MTs are assumed to be located within the coverage area of the GEO satellite. In accordance with ITU regulations [15], LEO satellites are required to operate without causing significant interference to GEO satellites. To ensure compliance with these regulations, the LEO and GEO satellites are assumed to operate on distinct frequency bands within the same spectrum, thereby mitigating inter-satellite interference between them. MTs are capable of establishing communication with both LEO and GEO satellites; however, at any given time, each MT is connected to only one satellite. Furthermore, all satellites are assumed to allocate bandwidth resources uniformly among the MTs they serve. Based on these assumptions, the downlink transmission rate $R_{i,j}(t)$ of an MT m_i from beam b_j at time step t can be expressed as follows:

$$R_{i,j}(t) = \frac{W}{M_j(t)} \log_2\left(1 + \mathrm{SINR}_{i,j}(t)\right), \tag{1}$$

where W denotes the bandwidth of b_j, $M_j(t)$ represents the number of MTs connected to b_j at the time t and $SINR_{i,j}(t)$ is the signal-to-interference-plus-ratio ($SINR$) between MT i and satellite beam b_j, and it is given by:

$$\mathrm{SINR}_{i,j} = \frac{P_{rx}}{\sum_i P_{I,i} + P_N}, \tag{2}$$

where P_{rx} denotes the received signal power, $P_{I,i}$ represents the interference power from the i-th interferer, and P_N is the noise power. The received signal power P_{rx} is determined by transmit power P_{tx}, receiver antenna gain G_{rx}, transmitter antenna gain G_{tx}, and path loss (PL), and is expressed as:

$$P_{rx} = P_{tx} + G_{tx}(\theta) + G_{rx} - PL. \tag{3}$$

The transmitter antenna gain $G_{tx}(\theta)$ is computed as:

$$G_{tx}(\theta) = \begin{cases} 1, & \text{if } \theta = 0, \\ \frac{1}{4}\left|\frac{J_1(ka\sin\theta)}{ka\sin\theta}\right|^2, & \text{if } 0 < \theta \le 90^\circ, \end{cases} \tag{4}$$

where the θ is the angle between the beam boresight and the direction from the satellite toward the MT. The PL is determined as:

$$PL = P_{\mathrm{LoS}}(FSPL + SF) + P_{\mathrm{NLoS}}(FSPL) + SF + CL, \tag{5}$$

which incorporates the probability of line-of-sight (LoS) propagation, modeled as a function of the elevation angle, along with the free-space path loss ($FSPL$), shadow fading (SF), and cluster loss (CL). The shadow fading accounts for attenuation due to shading effects and is characterized by a log-normal distribution. $FSPL$ is expressed as:

$$FSPL = 32.45 + 20\log_{10}(f_c) + 20\log_{10}(d), \tag{6}$$

where f_c is the center frequency in GHz and d is the distance in meters between the MT and the satellite [16].

B. CNN-LSTM Encoder-Decoder based Deep Learning Model for RSRP Prediction

To effectively predict variations in downlink RSRP induced by the mobility of satellites and MTs, we propose a multi-feature CNN-LSTM encoder-decoder model. As illustrated in Fig. 2 [17], the proposed framework integrates a CNN module in the encoder and an LSTM module in the decoder, leveraging the complementary strengths of both architectures. The CNN module efficiently captures local dependencies and inter-signal correlations, while the LSTM module is well-suited for learning temporal dependencies in sequential data. However, a standalone LSTM structure may struggle to extract meaningful spatial features from the input data, which are crucial for understanding variations in signal strength. To address this limitation, the proposed model incorporates a CNN module in the encoder to first extract spatial correlations among multiple signals before passing the processed features to the LSTM module for temporal modeling [18]. This integration enhances the model's ability to learn complex dependencies in both spatial and temporal domains, thereby improving predictive accuracy in dynamic satellite communication environments.

The proposed model utilizes multi-feature input data to capture RSRP fluctuations caused by satellite and MT mobility. It includes the Earth-centered Earth-fixed (ECEF) coordinates of the satellite, beam, and MT to represent their positions, along with their velocity vectors in ECEF coordinates to account for movement patterns. Additionally, the observed RSRP at the current time step is incorporated to learn signal strength variations. By integrating spatial positioning and motion dynamics, the model effectively accounts for key factors influencing RSRP changes. The target output is a predicted sequence of future RSRP values, learned from the given input data. This enables the model to capture complex signal fluctuations driven by satellite and MT mobility, ensuring reliable RSRP forecasting in real-time communication scenarios.

The CNN Encoder extracts spatial correlations using one-dimensional convolution layers, followed by max-pooling to reduce sequence length while retaining essential features. The LSTM Decoder processes these features to predict future RSRP values, leveraging temporal dependencies through iterative forecasting. By combining CNNs for spatial extraction and LSTMs for temporal modeling, the model effectively learns RSRP variations, ensuring accurate and robust predictions in dynamic satellite networks.

III. Distributed MADDRQN-based Predictive Handover Decision

A. Problem Formulation

This section formulates the beam-level handover problem in a cooperative GEO-LEO satellite network, considering transitions between LEO satellites and between LEO and GEO satellites. The primary objective is to optimize the trade-off between minimizing handover frequency and maximizing data rates for MTs while ensuring efficient resource utilization in the network.

To achieve this balance, the handover decision-making process is optimized to distribute network load efficiently and maintain stable connectivity across satellite beams. At each time step t, each MT selects a target beam for handover from the top B beams with the highest *RSRP*, excluding its current serving beam. The total set of candidate beams for MT m_i at time t is denoted as $B_i(t) = \{b_0, b_1, b_2, \dots, b_B\}$, where b_0 represents the current serving beam. Let $x_{i,j}(t) \in \{0,1\}$ be the binary handover decision variable, where $x_{i,j} = 1$ indicates that MT i transitions to beam j at time t, and $x_{i,j} = 0$ otherwise. The data rate $R_{i,j}(t)$ achieved by MT i when connected to beam j at time t is defined as:

$$R_i(t) = \sum_{j \in B_i(t)} x_{i,j}(t) R_{i,j}(t). \tag{7}$$

A handover event occurs when an MT transitions from its current serving beam to a new beam. The total handover count $H_i(t)$ is defined as:

$$H_i(t) = \sum_{j \in B_i(t), j \neq b_0} x_{i,j}(t). \tag{8}$$

The optimization problem is formulated to minimize handover frequency while maximizing data rates:

$$\min_{\mathbf{x}(t)} \Big\{ \omega_1 \sum_{i=1}^{N} \sum_{t=1}^{T} H_i(t) - \omega_2 \sum_{i=1}^{N} \sum_{t=1}^{T} R_i(t) \Big\}. \tag{9}$$

Here, $\mathbf{x}(t) = \{\mathbf{x}_1(t), \mathbf{x}_2(t), \dots, \mathbf{x}_N(t)\}$ represents the handover decisions of all MTs, ω_1 is the weight for minimizing handover frequency, and ω_2 is the weight for maximizing data rate. The objective function maintains a balance between reducing unnecessary handovers and ensuring optimal throughput. Minimizing handover frequency enhances connection stability by reducing service interruptions, ensuring prolonged and uninterrupted connectivity. The optimization problem is thus designed to simultaneously achieve minimal handover frequency and maximal data rates for continuous service quality. The *continuity time* of an MT at time t is defined as $C_i(t)$, which represents the number of consecutive time steps that MT i remains connected to the same serving beam up to time t. The $C_i(t)$ is expressed as:

$$C_i(t) = (C_i(t-1)+1)(1-H_i(t)) + 1 \cdot H_i(t). \tag{10}$$

To facilitate long-term performance evaluation and enhance system stability, a normalization function for service continuity time, denoted $T_i(t)$, is introduced as:

$$T_i(t) = \frac{\sqrt{C_i(t)}}{\sqrt{T_{opt}}}, \tag{11}$$

where T_{opt} represents the total duration considered for long-term optimization. This normalization ensures that increasing service continuity enhances system stability while maintaining a balanced optimization framework. Since $T_i(t)$ is normalized between 0 and 1, it is necessary to also normalize the data rate $R_i(t)$ to a similar range. Therefore, $R_i(t)$ is scaled by its maximum value R_{max} to ensure both $T_i(t)$ and $R_i(t)$ are within the same normalized range for balanced optimization. The corresponding objective function is expressed as:

$$\max_{\mathbf{x}(t)} \Big\{ \omega_1 \sum_{i=1}^{N} \sum_{t=1}^{T_{opt}} T_i(t) + \omega_2 \sum_{i=1}^{N} \sum_{t=1}^{T_{opt}} \frac{R_i(t)}{R_{\max}} \Big\}, \tag{12}$$

$$\text{s.t.} \sum_{j=0}^{B+1} x_{i,j}(t) = 1, \forall i \in \mathcal{N}, \forall t \in \mathcal{T}, \tag{13}$$

$$x_{i,j}(t) \in \{0,1\}, j \in \{0,1,\dots,B\}, \forall i \in \mathcal{N}, \tag{14}$$

$$R_i(t) > 0, \forall i \in \mathcal{N}, \forall t \in \mathcal{T}, \tag{15}$$

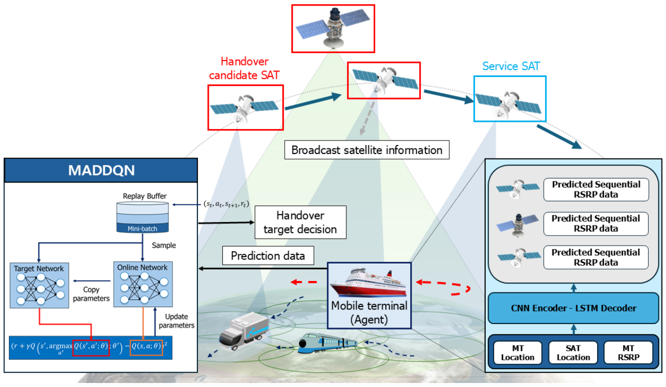

Fig. 3: Proposed distributed MADDQN-RL-based predictive handover decision system architecture.

$$\omega_1 + \omega_2 = 1,\ \omega_1,\ \omega_2 \geq 0. \tag{16}$$

Each MT is required to select either its current beam or one of the B candidate beams at time t. This decision-making process is modeled as a binary selection problem, ensuring that each MT maintains a positive data rate at all times. The system prioritizes balancing handover minimization and data rate maximization, controlled by the weight parameters ω_1 and ω_2, which can be adjusted based on network conditions.

B. Proposed Distributed DRL for Predictive Handover Optimization

The handover decision problem in GEO-LEO satellite networks is a sequential decision-making process, where each beam selection affects future connection stability and data performance. The highly dynamic environment, influenced by satellite and MT mobility, makes traditional rule-based optimization ineffective. The unpredictability of satellite beam coverage, fluctuating link quality, and user movement necessitate a learning-based approach that dynamically optimizes handover decisions while maintaining network stability and high throughput. To address this, we propose a multi-agent distributed deep reinforcement learning (MADDRQN) framework, enabling decentralized handover decision-making for multiple MTs in the network. Unlike traditional single-agent RL, which optimizes handovers from an individual perspective, MADDRQN allows collaborative learning among MTs, sharing state and reward information to enhance overall network efficiency. This ensures that handover decisions not only optimize individual MT performance but also contribute to network load balancing, reducing congestion in heavily utilized beams.

RL in high-dimensional state spaces is challenging, particularly in satellite networks with dynamic beam selection. To improve learning efficiency and policy generalization, we adopt deep Q-networks (DQN), which integrate Q-learning with artificial neural networks (ANNs). The Q-value function is:

$$Q(s,a;\theta) = r + \gamma Q\big(s', \arg\max_{a'} Q(s',a';\theta);\theta\big), \tag{17}$$

where s, a are the current state and action, and s', a' the next state and action, r the reward, and $\gamma \in [0,1]$ the discount factor. The parameter θ represents the ANN weights approximating the Q-value function.

Conventional DQN updates both the predicted Q-value $Q(s,a)$ and the target Q-value $Q(s',a')$ using the same ANN, leading to training instability. To mitigate this, we employ double deep Q-network (DDQN), introducing a target network θ^- that stabilizes training. The updated Q-value equation in DDQN is:

$$Q(s,a;\theta) = r + \gamma Q\big(s', \arg\max_{a'} Q(s',a';\theta);\theta'\big). \tag{18}$$

In the DDQN training process, only the primary network θ is updated in each iteration, while the target network θ^- is periodically synchronized by copying all parameters from θ. This strategy improves the stability of DRL training and enhances policy convergence. All agents share a common state, action, and reward structure, ensuring consistency in learning and decision-making. [19], [20].

Building upon the advantages of DDQN, this study introduces a distributed MADDRQN-based beam-level predictive handover optimization algorithm designed for GEO-LEO satellite cooperative networks. The state, action, and reward structures for each MT in the proposed MADDRQN-based handover decision mechanism are defined as follows.

- ***State Space:*** The state vector s_i^t for each agent i at time t is expressed as $s_i^t = [B_i^t, B_{c_0}^t, \ldots, B_{c_{n-1}}^t]$, where B_i^t represents the current serving beam information and beam information for n candidate beams. The beam information vector B is defined as $B_i^t = [i_s, i_b, t, n, b, R]$, where i_s represents the index of the satellite, i_b is the beam index, t is the time index, n is the number of users in the beam, b is the frequency band information, and $R = [r_0, \ldots, r_n]$ is the vector of predicted $RSRP$ values. The state information is shared among agents with overlapping beams in the beam information vector B.
- ***Action Space:*** The action space for each agent is defined as a discrete set of K candidate beams, indexed sequentially from 0 to $K-1$. At any given time t, the action a_i^t represents the beam index selected by the agent i. Formally, $a_i^t \in \{0, 1, 2, \ldots, K-1\}$. When $a_i^t = 0$, the agent maintains its connection to the current beam, thereby avoiding a handover. Conversely, for $a_i^t = k$ where $1 \leq k \leq K-1$, the agent initiates a handover to the k-th beam.
- ***Reward Design:*** The reward function is designed to optimize both data rate and connection stability while minimizing unnecessary handovers. It is formulated as:

  $$r_i(t) = R(s_i(t), a_i(t)) = \omega_1 T_i(t) + \omega_2 R_i(t), \tag{19}$$

  where $T_i(t)$ denotes the continuity time, and $R_i(t)$ represents the achieved data rate. The weighting parameters ω_1 and ω_2 balance the trade-off between stable connectivity and throughput performance. This formulation encourages prolonged service continuity while penalizing frequent handovers, ensuring optimized handover decisions.

The proposed system architecture is shown in Fig. 3, where each MT has a prediction model for RSRP forecasting and

Algorithm 1 Distributed predictive-MADDQN handover decision

1: **Initialize:** $\forall i \in \{1,\dots,N\}$
2: $\quad Q(s,a;\theta_i), \quad Q(s,a;\theta_i')$
3: $\quad \theta_i' \leftarrow \theta_i$
4: $\quad D_i \leftarrow \text{InitReplayBuffer}(|D_i|_{\max})$
5: $\quad \varepsilon \leftarrow \varepsilon_{\text{start}}$
6: **Training:**
7: **for** episode = 1 **to** MaxEpisode **do**
8: $\quad \varepsilon \leftarrow \max(\varepsilon_{\min}, \varepsilon \times \varepsilon_{\text{decay}})$
9: $\quad$ **for** time step = 1 **to** T **do**
10: $\quad\quad$ **for** each agent $i \in \{1,\dots,N\}$ **do**
11: $\quad\quad\quad$ Compute $\{Q_{\theta_i}(s_i,a)\}_{a=0}^{K}$
12: $\quad\quad\quad$ Select a_i via ε-greedy:

$$a_i = \begin{cases} \arg\max_{a\in\{0,\dots,K\}} Q_{\theta_i}(s_i,a), & \text{with prob } (1-\varepsilon), \\ \text{random in } \{0,\dots,K\}, & \text{with prob } \varepsilon \end{cases}$$

13: $\quad\quad$ **end for**
14: $\quad\quad$ Execute joint action $(a_1,\dots,a_N)$ and observe $(\{s_i'\},\{r_i\},\text{done})$
15: $\quad$ **end for**
16: $\quad$ **for** each agent i **do**
17: $\quad\quad$ Store (s_i,a_i,r_i,s_i') in D_i
18: $\quad$ **end for**
19: $\quad$ **if** $|D_i| \geq B$ **for some** i **then**
20: $\quad\quad$ **for** each agent i **do**
21: $\quad\quad\quad$ Sample minibatch $\{(s_{b,i},a_{b,i},r_{b,i},s_{b,i}',d_{b,i})\}_{b=1}^{B}$ from D_i
22: $\quad\quad\quad$ Compute target:

$$y_b = r_{b,i} + (1-d_{b,i})\,\gamma \max_a Q_{\theta_i'}(s_{b,i}',a)$$

23: $\quad\quad\quad$ **Update** θ_i by minimizing:

$$L(\theta_i) = \frac{1}{B}\sum_{b=1}^{B}\Big[Q_{\theta_i}(s_{b,i},a_{b,i}) - y_b\Big]^2$$

24: $\quad\quad$ **end for**
25: $\quad\quad$ **Target network update**:

$$\theta_i' \leftarrow \tau\,\theta_i + (1-\tau)\,\theta_i' \quad \forall i \in \{1,\dots,N\}$$

26: $\quad$ **end if**
27: **end for**

a handover decision model. All MTs share a common state, action, and reward structure while maintaining individual neural networks and replay buffers. The handover mechanism consists of handover candidate selection, RSRP prediction, and MADDRQN-based decision-making. The handover mechanism follows three key phases:

1) ***Candidate Beam Selection:*** To determine potential handover targets, satellites meeting the minimum elevation angle requirement are identified, excluding the currently serving satellite. The RSRP values of the candidate beams are denoted as r_k for $k \in \{1,2,\dots,K\}$, are ranked in descending order, and the top n beams are selected as candidates. This results in a candidate beam set $\mathscr{B}_{\rfloor} = \{k | r_k \in \{1,2,\dots,n\}\}$, with the final handover candidate beam set represented as $\mathscr{B}_{final} = \{\mathscr{B}_c \cup b_s\}$.
2) ***Future RSRP Prediction:*** To enable proactive handover decisions, a CNN-LSTM encoder-decoder model predicts future RSRP values for each MT. Given an input sequence $x = \{x_1,x_2,\dots,x_T\}$, the model generates a predicted RSRP sequence $\hat{y} = \{\hat{y}_{t+1},\dots,\hat{y}_{t+T_{\text{future}}}\}$. By capturing both spatial and temporal signal variations, this approach enhances decision accuracy and reduces unnecessary handovers.
3) ***Optimal Beam Selection using MADDQN:*** The state information s_i^t is processed by the DDQN model, which computes the Q-value $Q_\theta(s_i,k)$ for each candidate beam k, representing the expected long-term reward. The beam with the highest Q-value is selected as the optimal handover target, ensuring a balance between connection stability and data rate optimization.

This structured approach ensures efficient and adaptive handover decision-making, minimizing handover frequency while maintaining seamless connectivity and enhanced network performance in GEO-LEO satellite networks. The proposed MADDRQN-based handover decision process follows the structured learning procedure outlined in Algorithm 1. First, for each agent $i \in \{1,\dots,N\}$, initialize the action-value function $Q(s,a;\theta_i)$ and the corresponding target network $Q(s,a;\theta_i')$. Then, set the initial condition $\theta_i' \leftarrow \theta_i$ to ensure that the main network and the target network are identical at the initial stage. Additionally, create a replay buffer D_i for each agent and initialize the exploration rate ε as $\varepsilon \leftarrow \varepsilon_{\text{start}}$ (lines 1–5). Then, within each episode, up to T time steps are executed. At each time step, agent i computes the values of $Q_{\theta_i}(s_i,a)$, where $a \in \{0,1,\dots,K\}$. Based on the ε-greedy policy, the action a_i is selected. After all agents execute their selected actions in the environment, they observe the next state s_i', the reward r_i, and the episode termination indicator *done* (lines 10–16). Each agent i stores the tuple (s_i,a_i,r_i,s_i') in its replay buffer D_i. This enables offline learning by accumulating past state-action-reward-next state experiences (lines 17–18). If the size of the replay buffer $|D_i|$ is at least the mini-batch size B, a mini-batch is randomly sampled. Each experience in the mini-batch is indexed by b, and the target value y_b is computed as described on line 22, where $d_{b,i}$ indicates whether the episode has terminated, and γ is the discount factor. The loss function is defined as in line 23, where θ_i is updated to minimize this loss. Finally, the target network parameters θ_i' are updated using the soft update method in line 25.

IV. Performance evaluation

A. Simulation Environment

The performance of the proposed predictive handover scheme is evaluated through simulations conducted using the MATLAB Satellite Communication Toolbox. The simulation scenario, illustrated in Fig. 4, consists of a single GEO satellite and multiple LEO satellites distributed across three distinct orbital

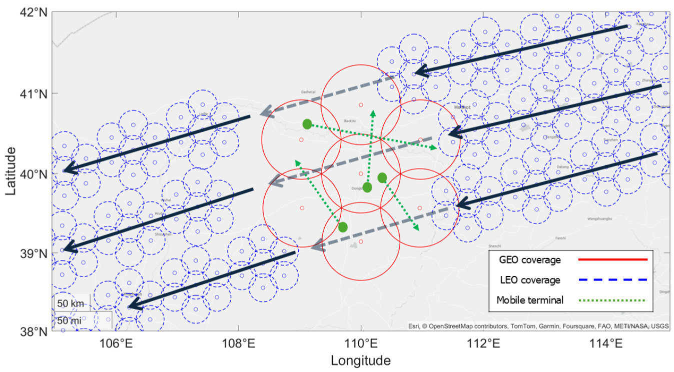

Fig. 4: Simulation environment.

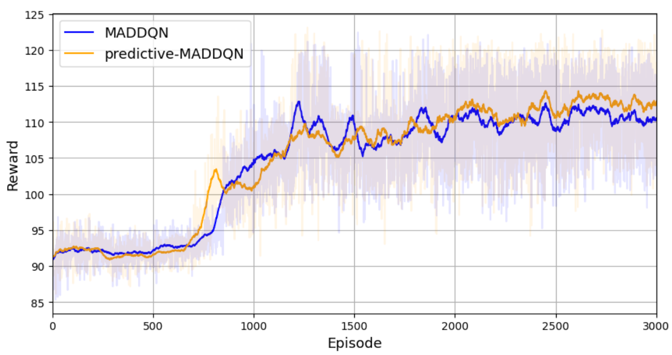

Fig. 5: Average cumulative reward as a function of the number of episodes.

TABLE I: System level simulation parameters

Parameter	**Value (LEO)**	**Value (GEO)**
Carrier frequency	20 [GHz] (Ka-band)	20.4 [GHz] (Ka-band)
Bandwidth	19.8–20.2 [GHz]	20.2–20.6 [GHz]
Altitude	1,200 [Km]	35,786 [Km]
TX antenna gain G_T	38.5 [dBi]	58.5 [dBi]
Beam diameter	40 [Km]	110 [Km]
Antenna aperture	0.5 [m]	5 [m]
EIRP density	10 [dBW/MHz]	40 [dBW/MHz]
Antenna temperature	150 [K]	
Boltzmann constant, k	-228.6 [dBW/K/Hz]	
Rx antenna gain G_R	39.7 [dBi]	
Simulation time	180 [s]	
Time step	1 [s]	

TABLE II: CNN-LSTM encoder-decoder learning parameters

Parameter	**Value**
Input feature	16
Hidden parameter	250
Number of layers	6
Batch size	32
Dropout rate	0.2
Time step (second)	1
Input sequence length	5
Target sequence length	3

TABLE III: MADDQN-RL learning parameters

Parameter	**Value**
Batch size	64
Learning rate	0.00005
Replay buffer size	10,000
Discount factor	0.95
Number of agents	4
Episode	3,000
Hidden size	128
Update frequency	1

planes. Under the assumption of a spherical Earth model, the velocity of each LEO satellite is determined based on Kepler's laws as a function of altitude. The LEO satellites are evenly spaced within their respective orbital planes, and each satellite is equipped with seven circular spot beams.

A minimum elevation angle of 10 degrees is required for an MT to establish a connection with a satellite. MTs travel at a constant velocity along a randomly generated straight-line trajectory within the GEO satellite's coverage area. The system-level simulation parameters are listed in Table I, while the CNN-LSTM encoder-decoder learning parameters are detailed in Table I, while the CNN-LSTM encoder-decoder learning parameters are detailed in Table II. Each MT utilizes a pre-trained CNN-LSTM model for RSRP prediction, and the MADDRQN training parameters are provided in Table III.

To comprehensively assess the performance of the proposed approach, the following handover strategies are considered for comparison:

- ***Measurement-based handover (MHO):*** The conventional handover approach selects the beam based on the measured RSRP of the MT. The MT evaluates the RSRP of potential handover target satellites and performs a handover when a beam is detected with a signal strength exceeding the current beam's strength by a predefined offset.
- ***Predictive signal average handover (PSAHO):*** This strategy selects the beam with the highest average predicted RSRP over a specified time window. By leveraging predicted RSRP trends, PSAHO accounts for signal fluctuations and selects the optimal handover target based on its expected performance over time.
- ***MADDQN-based handover (MADDQN):*** This RL-based method does not utilize predictive RSRP data. Each MT independently optimizes its handover decisions through partial state observability, selecting the optimal beam without sharing state information with other agents. Multiple MTs learn simultaneously, refining their policies through RL.

B. Simulation Results

To validate the performance improvements enabled by predictive information, we compare the learning performance of predictive-MADDRQN (i.e., proposed algorithm) and standard MADDRQN during training. Additionally, we evaluate the pro-

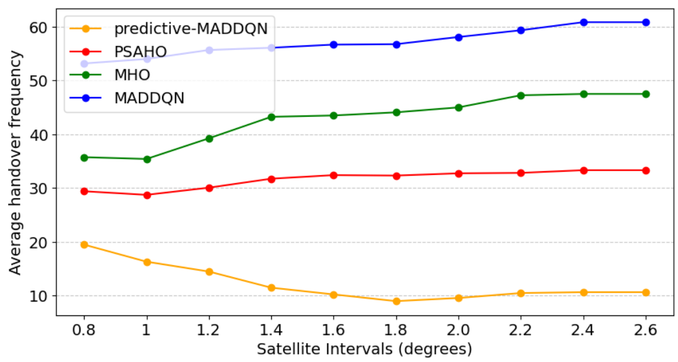

Fig. 6: Average handover frequency for various satellite intervals.

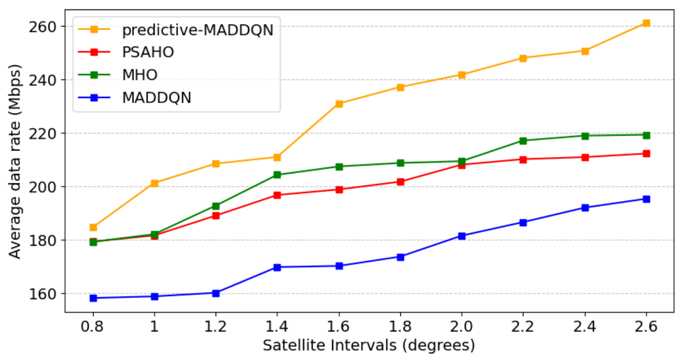

Fig. 7: Average data rate for various satellite intervals.

posed handover scheme against three baseline methods—MHO, PSAHO, and MADDRQN—to assess handover frequency, connection stability, and data throughput under varying conditions.

Fig. 5 illustrates the learning convergence of predictive-MADDRQN and MADDQN in the simulation environment, where MT speed is set to 100 km/h, LEO satellite spacing is 0.8 degrees, and each MT observes an average of 241 satellites simultaneously. The results show that both algorithms generally converge around episode 1500; however, predictive-MADDRQN converges to a higher reward value, demonstrating its superior learning efficiency and effectiveness in leveraging predictive data. This confirms that incorporating predictive RSRP information enhances decision-making and improves overall system performance. The effect of LEO satellite density on handover frequency is analyzed in Fig. 6 and Fig. 7, where x-axis represents the satellites intervals between LEO satellites. MTs are assumed to move at a constant speed of 100 km/h. Fig. 6 present the results of the analysis on the effect of satellite intervals between LEO satellites on handover frequency and data throughput. As the satellite interval between LEO satellites increases, most schemes exhibit a gradual increase in handover frequency. This is due to the reduced connection duration with each LEO satellite, leading to a higher likelihood of a gap before connecting to the next LEO satellite. To mitigate this gap, the handover frequency with GEO satellites increases, and this phenomenon becomes more pronounced as the satellite

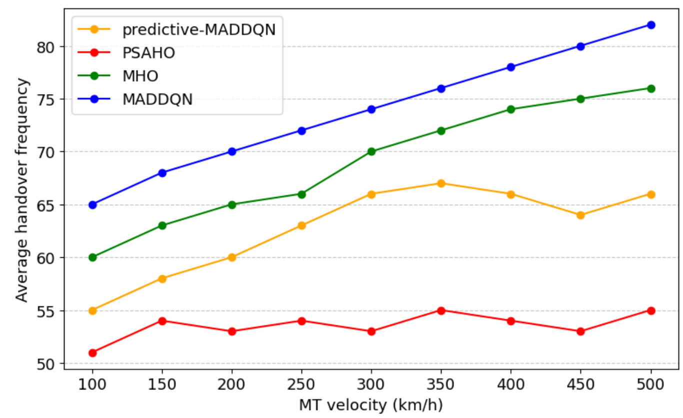

Fig. 8: Average handover frequency for various MT velocity.

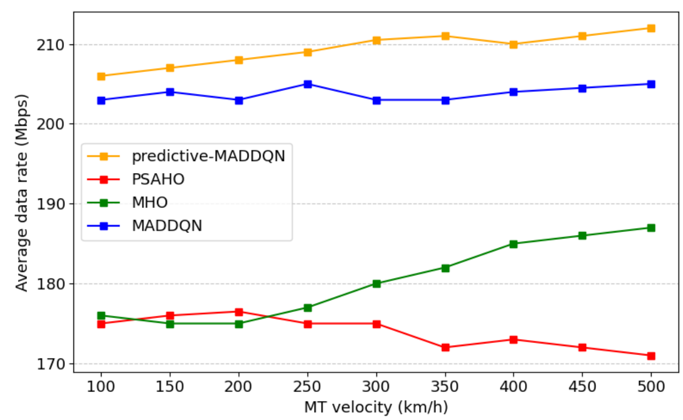

Fig. 9: Average data rate for various MT velocity.

interval widens. In particular, predictive-MADDRQN maintains a relatively lower handover frequency compared to other methods and demonstrates stable performance even as the satellite interval increases. This suggests that the predictive information utilized by proposed handover scheme enables it to determine the optimal handover timing, thereby reducing unnecessary handovers and maintaining more stable network connectivity. On the other hand, Fig. 7 shows a trend of increasing data throughput as the satellite interval increases. This is attributed to the reduction in interference between satellites. proposed predictive-MADDRQN consistently achieves the highest data throughput across all intervals, indicating that it is effective in optimizing network performance by minimizing handover frequency while maximizing data rate.

The effect of MT velocity on handover frequency and data throughput is presented in Fig. 8 and Fig. 9, where the x-axis represents the MT speed. In Fig. 8, predictive-MADDRQN and PSAHO exhibit significantly lower average handover frequencies compared to MHO and MADDRQN. Specifically, predictive-MADDRQN shows a tendency for the handover frequency to increase as the MT increase, but beyond 300 km/h, the frequency levels off and remains stable. This indicate that predictive-MADDRQN uses predictive information to suppress unnecessary handovers and maintain a more stable network connection at higher speeds. In contrast, MHO and MADDQN show a continuous increase in handover frequency with speed, leading to higher network overhead and reduced efficiency. Fig. 9 shows that the proposed handover scheme maintains the highest data transmission rate across all velocities. By minimizing unnecessary handovers and maintaining a high data rate, predictive-MADDQN ensures more stable connections, demonstrating the effectiveness of predictive-based handover optimization.

V. Conclusion

This paper addresses the handover optimization problem in cooperative GEO-LEO satellite networks supporting mobile terminals. Traditional RSRP-based methods struggle with frequent handovers and network overhead due to LEO satellite mobility. To overcome these challenges, we propose a novel predictive handover optimization scheme that integrates a CNN-LSTM encoder-decoder model for RSRP forecasting and a multi-agent deep RL framework for optimal decision-making. By leveraging predictive RSRP information, the proposed approach minimizes unnecessary handovers, maximizes per-MT data throughput, and enhances network stability even in high-mobility satellite environments. Simulation results confirm that predictive-MADDRQN outperforms traditional methods, achieving lower handover frequency and higher data rates. The proposed scheme ensures robust connectivity and optimized resource utilization, making it an effective solution for next-generation GEO-LEO satellite networks.

Acknowledgment

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2021-0-01715, Development of Communications Payload for GEOKOMPSAT-3)

References

[1] J. Chen, Z. Wei, Y. Wang, L. Sang, and D. Yang, "A service-adaptive multi-criteria vertical handoff algorithm in heterogeneous wireless networks," in *Proc. IEEE Int. Symp. Personal Indoor Mobile Radio Commun. (PIMRC)*, Sydney, NSW, Australia, Sep. 2012, pp. 899–904.

[2] H. Tataria, M. Shafi, A. F. Molisch, M. Dohler, H. Sjöland, and F. Tufvesson, "6G wireless systems: Vision, requirements, challenges, insights, and opportunities," *IEEE*, vol. 109, no. 7, pp. 1166–1199, Jul. 2021.

[3] F. Wang, D. Jiang, Z. Wang, J. Chen, and T. Q. S. Quek, "Seamless handover in LEO based non-terrestrial networks: service continuity and optimization," *IEEE Trans. Commun.*, vol. 71, no. 2, pp. 1008–1023, Feb. 2023.

[4] Starlink, "Starlink Internet services," Sep. 2021. [Online]. Available: https://www.starlink.com

[5] L. Yang, X. Yang, and Z. Bu, "A conditional handover strategy based on trajectory prediction for high-speed terminals in LEO satellite networks," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Denver, Co, USA, Jun. 2024, pp. 1671-1701.

[6] C. Han, L. Huo, X. Tong, H. Wang, and X. Liu, "Spatial anti-jamming scheme for Internet of satellites based on the deep reinforcement learning and Stackelberg game," *IEEE Trans. Veh. Technol.*, vol. 69, no. 5, pp. 5331–5342, May 2020.

[7] S. Zhang, A. Liu, C. Han, X. Ding, and X. Liang, "A network-flows-based satellite handover strategy for LEO satellite networks," *IEEE Wireless Commun. Lett.*, vol. 10, no. 12, pp. 2669–2673, Dec. 2021.

[8] H. Zhang, Z. Wang, S. Zhang, Q. Meng, and H. Luo, "Optimizing link-identified forwarding framework in LEO satellite networks," in *Proc. 21st Int. Symp. Model. Optim. in Mobile, Ad Hoc, Wireless Netw. (WiOpt)*, Singapore, Singapore, Aug. 2023, pp. 41–48.

[9] H. Wang, W. Qi, X. Xia, Y. Liu, Y. Xing, and C. Mei, "GEO and LEO collaborative architecture evolution in future integrated satellite-terrestrial network," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. and Broadcasting (BMSB)*, Beijing, China, Jun. 2023, pp. 1–6.

[10] J. Lee, C. Park, S. Park, and A. Molisch, "Handover protocol learning for LEO satellite networks: Access delay and collision minimization," *IEEE Trans. Wireless Commun.*, vol. 23, no. 7, pp. 7624–7637, Jul. 2024.

[11] N. Badini, M. Jaber, M. Marchese, and F. Patrone, "User centric satellite handover for multiple traffic profiles using deep Q-learning," *IEEE Trans. Aerosp. Electron. Syst.,*, vol. 60, no. 6, pp. 8591–8604, Dec. 2024.

[12] F. Yang, W. Wu, Y. Gao, Y. Sun, T. Sun, and P. Si, "Multi-agent fingerprints-enhanced distributed intelligent handover algorithm in LEO satellite networks," *IEEE Trans. Veh. Technol.*, vol. 73, no. 10, pp. 15255–15269, Oct. 2024.

[13] N. Badini, M. Jaber, M. Marchese, and F. Patrone, "Reinforcement learning-based load balancing satellite handover using NS-3," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Rome, Italy, May/Jun. 2023, pp. 2595–2600.

[14] D.-H. Jung, H.-J. Nam, and J. Choi, "Coverage analysis of GEO satellite networks," in *Proc. 22nd Int. Symp. Model. Optim. Mobile, Ad Hoc, Wireless Netw. (WiOpt)*, Seoul, Republic of Korea, Oct. 2024, pp. 70–75

[15] International Telecommunication Union (ITU), *Radio Regulations. Chapter VI: Provision of Services and Stations*, Geneva, Switzerland, 2016. [Online]. Available: https://life.itu.int/radioclub/rr/art22.pdf.

[16] 3rd Generation Partnership Project (3GPP), "Study on New Radio (NR) to support non-terrestrial networks (Release 15), " Technical Report (TR) 38.811 v15.4.0, Sep. 2020.

[17] C. Nguyen, T. M. Hoang, and A. A. Cheema, "Channel estimation using CNN-LSTM in RIS-NOMA assisted 6G network," *IEEE Trans. Mach. Learn. Commun, Netw.*, vol. 1, pp. 43–60, Jun. 2023.

[18] Z. Wang, X. Su, and Z. Ding, "Long-term traffic prediction based on LSTM encoder-decoder architecture," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 10, pp. 6561–6571, Oct. 2021.

[19] R. S. Sutton, and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.

[20] F. G. Ortiz-Gomez, D. Tarchi, R. Martínez, A. Vanelli-Coralli, M. A. Salas-Natera, and S. Landeros-Ayala, "Cooperative multi-agent deep reinforcement learning for resource management in full flexible VHTS systems," *IEEE Trans. Cogn. Commun. Netw.*, vol. 8, no. 1, pp. 335–349, Mar. 2022.