Hierarchical Reinforcement Learning Based Resource Scheduling and Handover Control in Integrated Satellite-Ground Network

Huiyeon Jang[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
{timd0801, sjung}@ajou.ac.kr

***Abstract*—Satellite-terrestrial integrated networks (STINs) are essential for sixth-generation wireless technology (6G) global connectivity, but the low Earth orbit (LEO) satellites' scale and mobility create resource management challenges with frequent handovers and limited spectrum. This yields a complex control problem where handover decisions and resource scheduling are tightly coupled, exceeding conventional optimization and standard reinforcement learning (RL) capabilities. We propose a hierarchical reinforcement learning (HRL) framework that decouples this problem through coordinated high-level handover control and low-level resource scheduling policies. Simulations show our HRL approach surpasses non-hierarchical RL and conventional methods, significantly reducing handovers and increasing throughput in dynamic STINs.**

***Index Terms*—satellite communication, handover, resource allocation, hierarchical reinforcement learning.**

I. Introduction

The emerging sixth-generation wireless technology (6G) is driving satellite-terrestrial network integration for global connectivity [1]. While satellite–terrestrial integrated networks (STINs) with low Earth orbit (LEO) satellites provide low latency and high throughput, satellite mobility and network scale create resource management challenges: frequent user handovers and limited spectrum/power sharing [2]. Since handover decisions depend on resource availability and scheduling affects handover-induced quality loss, this coupled high-dimensional control problem exceeds traditional and flat reinforcement learning (RL) capabilities. We propose a hierarchical reinforcement learning (HRL) framework with coordinated high-level handover control and low-level resource scheduling policies. Simulations demonstrate superior performance over non-hierarchical RL and conventional methods, reducing handovers while increasing throughput in dynamic STINs.

II. Resource Allocation and Handover Control for Hierarchical Reinforcement Learning

A. Satellite-Terrestrial Coexistence Scenario

This paper considers the downlink communication scenario in a network where the terrestrial network and the satellite network coexist, as shown in Fig. 1. In this scenario, the LEO satellite network and the terrestrial base station (GBS) have

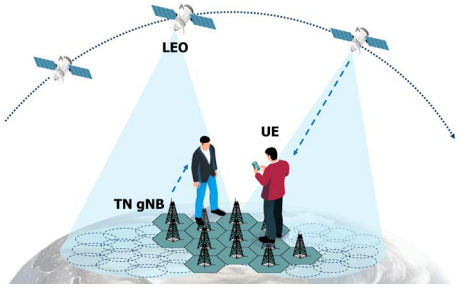

Fig. 1: System model.

partially overlapping bandwidths. Therefore, efficient utilization of frequency resources and interference management are crucial.

B. Proposed Hierarchical Reinforcement Learning Algorithm

This paper presents a HRL framework for joint handover and resource allocation optimization in wireless networks, targeting maximum throughput with minimal handover frequency through intelligent resource block allocation. The framework employs a two-tier architecture: a high-level agent selecting handover targets and a low-level agent managing bandwidth allocation. Both levels are formulated as Markov decision processes (MDPs) with the following specifications:

1) **State**: The high-level and low-level agents share a common state representation. The state at time t is defined as $\mathbf{s}_t = [\mathbf{c}_t, \mathbf{s}_t, \mathbf{d}_t, \mathbf{t}_t, \mathbf{v}_t]$, where $\mathbf{c}_t \in \{0,1\}^N$ denotes the connection status vector for N users, $\mathbf{s}_t \in \mathbb{R}^{N\times 3}$ represents the received signal reference power (RSRP) measurements, $\mathbf{d}_t \in \mathbb{R}^N$ indicates the throughput vector, $\mathbf{t}_t \in \mathbb{R}^N$ represents the connection duration vector, and $\mathbf{v}_t \in \mathbb{R}^N$ denotes the relative velocity vector.
2) **Action**: The high-level agent is responsible for mobility management and determines the handover policy by selecting a discrete action $a_i(t) \in \{0,1\}$, where $a_i(t) = 1$ triggers a handover to the highest RSRP cell, while $a_i(t) = 0$ indicates maintaining the current connection. Following this decision, the low-level agent produces a continuous vector $\mathbf{u}(t) \in \mathbb{R}^8$, which is assigned to the

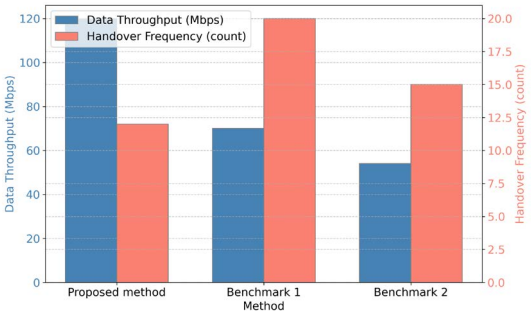

Fig. 2: Comparison of throughput performance and handover frequency.

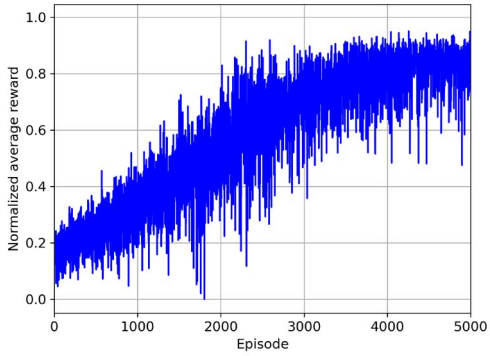

Fig. 3: Normalized average reward per training episode.

fractions of bandwidth per cell and the placements of the subband for the users of each beam.

3) **Reward**: The high-level and low-level agents utilize a common reward function that reflects a trade-off between two conflicting objectives: connection duration and data throughput. The reward function is formulated as follows,

$$R(t) = \omega_1 T(t) + \omega_2 D(t), \tag{1}$$

where $T(t)$ and $D(t)$ denotes the average connection duration time and average data throughput, respectively. The weighting factors ω_1 and ω_2 are design parameters that determine the relative importance of the two objectives and can be tuned according to specific system requirements or performance priorities.

III. Performance Evaluation

Table I summarizes the simulation parameters for the performance evaluation of the proposed method, while the hyperparameters used for training are presented in Table II. The simulations were conducted on a system equipped with an AMD Ryzen 7900X CPU and an NVIDIA GeForce RTX 4070 Super GPU.

To evaluate the performance of the proposed scheme, we define the two benchmarks: 1) a conventional measurement-based handover scheme with random resource division and 2) an RL-based handover scheme with random resource division. The convergence of the proposed scheme is first demonstrated in Fig. 3, which illustrates the cumulative reward over training episodes. This shows that the proposed scheme converges to an optimal policy after 4000 episodes. Based on this stable learning, Fig. 2 shows the average data throughput and the handover frequency during an episode. The results indicate that our proposed HRL scheme provides the highest data throughput with the lowest handover frequency. It can be observed that the proposed HRL scheme guarantees the best data throughput by dynamically allocating frequency resources while minimizing the effects of interference between the terrestrial and satellite networks.

TABLE I: Simulation environment parameters

Parameter	Value
Altitude of the LEO satellite	600 km
The frequency bandwidth of LEO	19.9 $\sim$ 20.2 GHz
The beam radius of LEO	25 km
The frequency bandwidth of GBS	20.0 $\sim$ 20.2 GHz
The inter-site distance of GBS	7500 m

TABLE II: HRL simulation parameters

Parameter	Value
Discount factor γ	0.99
Learning rate	0.0001
Batch size	64 episode
Buffer size	1000 episode

IV. Conclusions

In this paper, we proposed a HRL framework to solve the coupled resource scheduling and handover control problem in STINs. By decomposing the task into high-level handover and low-level scheduling policies, our approach was shown via simulation to significantly reduce handover frequency while increasing user throughput. The results validate our HRL framework as a robust and effective solution for resource management in next-generation integrated networks, outperforming both conventional and non-hierarchical RL methods.

Acknowledgment

This work was supported in by the Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korea government (MSIT) (RS-2024-00396992, Development of Cube Satellite based on Core Technologies in Low Earth Orbit Satellite Communications).

References

[1] J. Jang, J. Kim, J. Kim, and S. Jung, "Joint interference approximation and guard-band management for spectrum-efficient integrated NTN–TN networks," *IEEE Internet of Things Journal*, vol. 12, no. 15, pp. 32 220–32 236, Aug. 2025.

[2] X. Zhu and C. Jiang, "Integrated satellite-terrestrial networks toward 6G: Architectures, applications, and challenges," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 437–461, Jan. 2022.