A Hierarchical Learning Approach for Optimal Resource Allocation Strategy for NTN-TN Coexistence Scenarios

Junyoung Kim[†], Jiseok Jang[†], and Soyi Jung[‡]
[†]*Dept. Artificial Intelligence Convergence Network, Ajou University, Suwon, 16499, South Korea*
[‡]*Dept. Electrical and Computer Engineering, Ajou University, Suwon, 16499, South Korea*
{junzero0615, star12191254, sjung}@ajou.ac.kr

***Abstract*—In sixth-generation (6G) networks, non-terrestrial networks (NTNs) using low Earth orbit (LEO) satellites are expected to be pivotal in extending coverage and service diversity. Most existing studies on LEO-based communications focus on downlink (DL) services within the Long-Term Evolution (LTE) framework. This paper addresses an NTN–terrestrial network (TN) coexistence scenario based on the fifth-generation (5G) orthogonal frequency division multiple access (OFDMA) architecture and proposes a hierarchical reinforcement learning-based resource allocation and scheduling scheme (HRL-RAS). The HRL-RAS optimizes resource distribution to meet heterogeneous user demands and improve key performance indicators (KPIs) such as data rate for future integrated NTN–TN systems.**

***Index Terms*—satellite communication, coexistence scenario, hierarchical reinforcement learning**

I. Introduction

With the rapid evolution of communication technologies, low Earth orbit (LEO) satellite communications have become a key enabler for sixth-generation (6G) use cases. However, their high mobility and wide coverage pose challenges such as ensuring continuous user equipment (UE) service, managing mobility, and addressing resource allocation and frequency overlap in terrestrial network (TN) and non-terrestrial network (NTN) coexistence [1]. To mitigate these issues, the 3rd Generation Partnership Project (3GPP) has defined NTN–TN coexistence scenarios in major standardization documents, with ongoing efforts to refine the specifications [2].

In current 3GPP specifications, all NTN–TN coexistence scenarios are defined using the Long-Term Evolution (LTE) architecture and parameters as the baseline. To accommodate diverse user equipment (UE) requirements, this paper proposes a hierarchical reinforcement learning (HRL)-based resource allocation and scheduling scheme (HRL-RAS) for optimizing resource allocation in orthogonal frequency division multiple access (OFDMA) systems [3], [4]. The proposed HRL-RAS is evaluated in an NTN downlink (DL)–TN DL coexistence scenario, as illustrated in Fig. 1. By adaptively adjusting and assigning resource blocks (RBs) according to individual UE demands, the HRL-RAS enhances key performance indicators (KPIs) such as data rate.

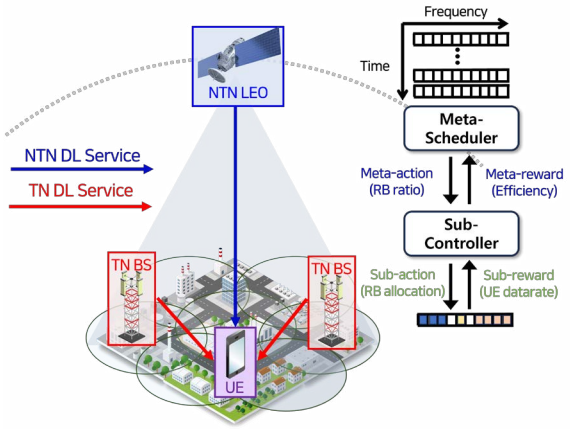


Fig. 1: HRL-RAS system model.

II. System Model

As illustrated in Fig. 1, this paper introduces a hierarchical reinforcement learning (HRL) framework to optimize RBs scheduling for UE requesting DL services over a 5G OFDMA architecture [3]. Each UE is classified into one of three service categories: enhanced mobile broadband (eMBB), ultra-reliable and low-latency communications (uRLLC), or massive machine-type communications (mMTC). In the proposed HRL-RAS scheme, the meta-scheduler (m) determines the total RBs assigned to each service type, while the sub-controller (s) allocates RBs to subframes according to the policy determined by m. The UE set is defined as $\mathbf{K} = \{k | k = 1, 2, \ldots, N_k\}$, where m corresponds to a 10 ms frame and s to a 1 ms subframe in the OFDMA structure. The Markov decision processes (MDPs) for each HRL layer are described as follows:

1) Meta-State ($S^m(t) = \{k_e^m, k_u^m, k_c^m\}$): In the upper layer, the meta-scheduler m periodically (every 10 ms) collects DL service demand information for all UEs classified as eMBB (k_e^m), uRLLC (k_u^m), or mMTC (k_c^m) under the NTN–TN coexistence scenario. Based on this information, m determines the RB allocation for each service category.

TABLE I: Simulation parameters.

Definition	Value
Operating frequency (DL)	S-band (2 GHz)
Satellite altitude (LEO)	600 km
Bandwdith, Subcarrier Spacing	20 Mhz, 15 KHz
N_k	1000
$SINR_\tau$	1 dB
Training Epoch, Step	1000, 500
Learning rate	0.0001
Discount factor (γ)	0.99

2) Meta-Action ($A^m(t) = \{a_e^m, a_u^m, a_c^m\}$): Every 10 ms, m determines the RB allocation ratios for the current frame—a_e^m for eMBB, a_u^m for uRLLC, and a_c^m for mMTC—based on the collected service demand data.

3) Meta-Reward ($R^m(t)$): The meta-reward is defined as the proportion of UEs whose signal-to-interference-plus-noise ratio (SINR) exceeds a predefined threshold $SINR_\tau$, as given in Eq. (1). The goal is to design an RB allocation and scheduling policy that maximizes UE throughput in the OFDMA structure while minimizing interference.

$$R^m(t) = \frac{\sum_{k=1}^{k=N_k, k\in\mathbf{K}} \mathbf{1}(SINR_k \geq SINR_\tau)}{\sum_{k=1}^{k=N_k, k\in\mathbf{K}} \mathbf{1}}. \quad (1)$$

4) Sub-State ($S^s(t) = \{S^m(t), k_e^s, k_c^s, k_u^s\}$): In the lower layer, the sub-controller s operates at the subframe level (1,ms) within the OFDMA structure. Using the allocation policy determined by $S^m(t)$, s collects information on the UEs—k_e^s (eMBB), k_c^s (uRLLC), and k_u^s (mMTC)—to which each DL RB will be assigned at the current time step.

5) Sub-Action ($A^s(t) = \{a_e^s, a_c^s, a_u^s\}$): Based on $S^s(t)$, the sub-controller allocates RBs within its bandwidth to the corresponding service requests: a_e^s for eMBB, a_c^s for uRLLC, and a_u^s for mMTC.

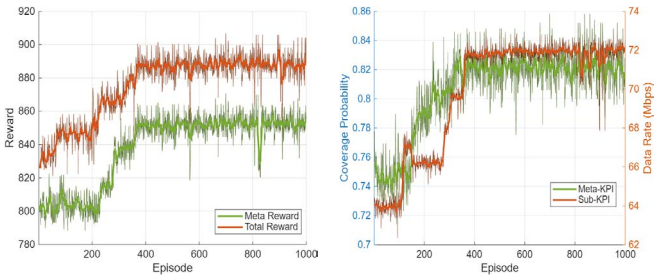
6) Sub-Reward ($R^s(t)$): The sub-reward is defined as the data rate achieved by each UE, as expressed in Eq. (2). This reward formulation is designed to optimize RB scheduling, reduce interference within the subframe, and enhance KPIs.

$$R^s(t) = \frac{B}{N_{k,s}} \log_2(1 + SINR_k). \quad (2)$$

In summary, this work proposes HRL-RAS, a hierarchical MDP-based framework for optimized resource allocation in NTN–TN coexistence scenarios within the 5G architecture, employing a hierarchically structured actor–critic algorithm for fine-grained RB allocation policy adjustment.

III. Performance Evaluation

As illustrated in Fig. 1, the performance of the proposed HRS-RAS model in an NTN–TN coexistence scenario is evaluated through simulations conducted in a Python environment, with reference to the 3GPP simulation parameters and TableI [2]. The total simulation time is set to 50 ms, where, at every 10 ms interval, m determines an allocation strategy based on the service-specific resource requirements and transmits $A^m(t)$ to each s (subframe). Each s subsequently executes the resource allocation action $A^s(t)$ every 1,ms to satisfy $A^m(t)$, while learning to maximize the reward.



(a) HRL-RAS reward convergence. (b) KPIs results.

Fig. 2: HRL-RAS simulation results.

To assess the performance metrics of the proposed HRL-RAS model, the convergence behavior of the reward function was analyzed, as illustrated in Fig. 2(a). The training results indicate that convergence commenced at approximately the 420th episode, and the similarity between the convergence trends of m and the overall reward curve confirms the effectiveness of the model. Furthermore, as shown in Fig. 2(b), an episode-wise analysis of the KPI metrics for the upper and lower layers demonstrates that both $R^m(t)$ and $R^s(t)$ exhibit clear convergence trends. Specifically, $R^m(t)$ aims to enhance the SINR of UEs allocated to the OFDMA bandwidth over a 50,ms duration, thereby enabling the formulation of an optimal RB allocation policy. In parallel, $R^s(t)$ optimizes the number of RBs assigned per subframe at 1,ms intervals, based on the upper-level resource allocation policy, to improve UE data rates. The observed improvement in both $R^m(t)$ and $R^s(t)$ KPIs over successive episodes confirms the successful derivation of the HRL-RAS strategy.

IV. Conclusion

This paper proposes HRL-RAS, an optimal resource allocation scheme based on an HRL architecture for NTN–TN coexistence in the 5G OFDMA framework. HRL-RAS improves overall resource allocation efficiency and enhances DL service KPIs for individual UEs. Future work will extend this approach using human-feedback reinforcement learning to address dynamically varying service demands.

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," pp. 32 220–32 236, August 2025.

[2] 3GPP TR 38.863 v16.1.0, "Solutions for NR to support non-terrestrial networks (NTN): NTN related RF and co-existence aspects (Release 18)," 3rd Generation Partnership Project (3GPP), Technical Report 38.863, June 2024.

[3] T. D. Kulkarni, K. R. Narasimhan, A. Saeedi, and J. B. Tenenbaum, "Hierarchical deep reinforcement learning:Intergrating temporal abstraction and intrinsic motivation," in *Proc. 30th International Conference on Neural Information Processing Systems (NIPS 2016)*, Barcelona, Spain, December 2016.

[4] J. Zhu, Y. Shi, Y. Zhou, C. Jiang, and L. Kuang, "Hierarchical learning and computing over space-ground integrated networks," *IEEE Transactions on Mobile Computing*, pp. 1–17, 2025 (Early Access).