Self-Directed Spectrum Allocation Framework for Integrated TN-NTN 6G Networks

Vaskar Chakma and Wooyeol Choi
School of Computer Science and Engineering, Chung-Ang University, Seoul, Republic of Korea
{vaskar, wchoi}@cau.ac.kr

***Abstract*—This paper proposes a self-adaptive channel assignment framework based on Q-learning, where agents learn optimal policies by observing network load, interference conditions, and temporal traffic dynamics within a Markov decision process (MDP). A multi-objective reward function is designed to jointly optimize system throughput, user fairness, and interference mitigation, while an ϵ-greedy strategy is employed to facilitate effective exploration. Simulation results demonstrate stable convergence, achieving an average reward of 37.5 and an average throughput of 28.5 Mbps. Moreover, the proposed approach achieves a Jain's fairness index of 0.75 and reduces interference by 26.3% compared to random allocation by adaptively responding to dynamic traffic patterns.**

***Index Terms*—TN-NTN integration, spectrum allocation, Q-learning, 6G networks.**

I. Introduction

As wireless networks transition to the sixth-generation (6G) systems, they need to support connectivity beyond ground-based networks. Integrated terrestrial and non-terrestrial networks (TN-NTN) merge ground base stations with LEO satellites, high-altitude platforms (HAPs), and UAVs, forming a multi-layered architecture enabling worldwide connectivity [1], [2]. LEO satellites move at speeds of 7 to 8 km/s, which causes rapid changes in network topology and requires regular handovers [3]. The unequal interference between terrestrial and satellite systems, along with different QoS needs, makes conventional optimization approaches impractical [4].

Artificial intelligence, especially reinforcement learning (RL), enables autonomous network management by learning from interactions [5]. Q-learning is a model-free RL method that learns policies without requiring an explicit environment model. Recent research on Q-learning for spectrum allocation [6] and deep reinforcement learning for wireless resources [7] shows promise. This paper proposes a Q-learning framework for TN-NTN spectrum allocation, incorporating state space encoding of network loads, interference, and temporal patterns. We chose Q-learning for the TN-NTN integration problem, as illustrated in Fig. 1, because network loads, interference levels, and temporal patterns can be discretised to form the state space.

II. Related Work

Research has focused on integrating terrestrial and non-terrestrial networks, especially in 6G networks. Mahboob and Liu [5] comprehensively review AI-powered satellite-based

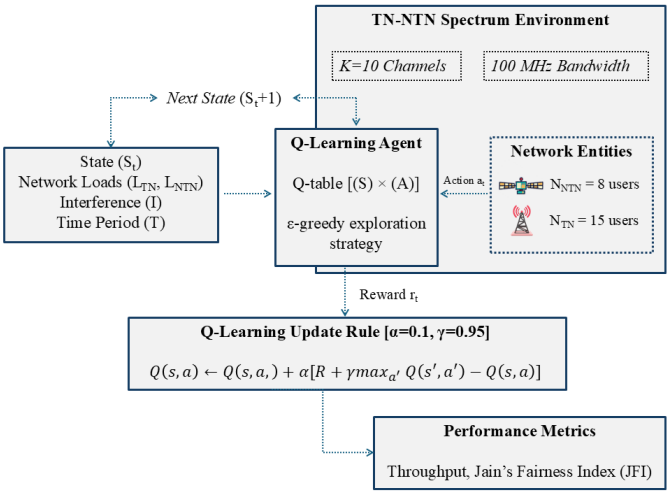

Fig. 1. System architecture of the Q-learning-based spectrum allocation framework.

NTNs, emphasizing reinforcement learning for dynamic resource management in LEO satellite networks. Reinforcement learning for spectrum allocation has shown promise across various wireless network scenarios. Adebayo et al. [8] suggested a DQN-based deep reinforcement learning approach for dynamic spectrum sharing between LTE and NB-IoT, enhancing throughput while upholding Jain's fairness index. Ukpong et al. [9] achieved 96.34% interference avoidance in TV whitespace cognitive radio networks using DRL models for increased dynamic spectrum access.

Several studies have tackled resource allocation challenges specific to satellite networks. A multi-agent DRL framework for interference-aware channel allocation in LEO satellite uplink situations was proposed by Cho et al. [10], utilizing sequential agent training to address non-stationarity. Coordinated spectrum allocation solutions help manage NTN-terrestrial service interference, according to their research. Deep reinforcement learning methods like DQN, Double DQN, and Dueling DQN excel in big state spaces [11], but tabular Q-learning excels in intermediate-sized problems for interpretability and efficiency. Elhachmi [6] showed that distributed Q-learning can optimise dynamic spectrum allocation in cognitive radio-based IoT networks, especially with careful state discretisation. Unlike prior Q-learning approaches applied to single-network scenarios, our framework jointly encodes TN-NTN

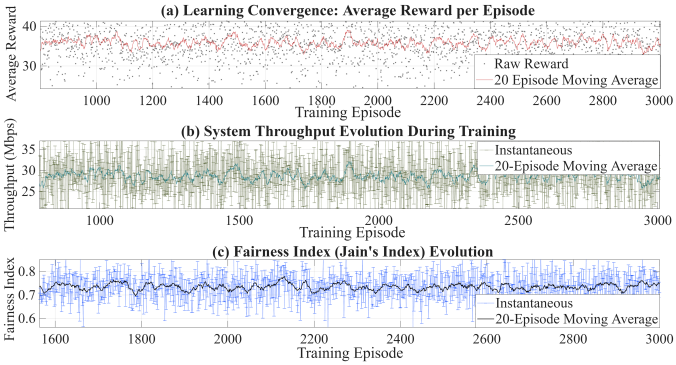

Fig. 2. Learning convergence after approximately 2,000 training episodes: (a) average reward per episode, (b) system throughput, and (c) Jain's fairness index with 20-episode moving averages.

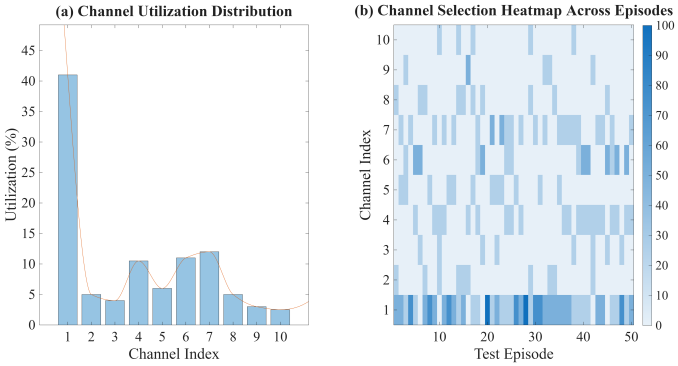

Fig. 3. Channel utilisation patterns: (a) distribution across 10 channels and (b) selection heatmap across test episodes.

co-channel interference and temporal traffic asymmetry into the state space.

III. System Model

A. Network models

We consider an integrated system with $N_{\text{TN}} = 15$ terrestrial users, $N_{\text{NTN}} = 8$ non-terrestrial users, and $K = 10$ frequency channels sharing 100 MHz bandwidth. LEO satellites operate at an altitude of $h = 550$ km with a velocity of $v = 7.5$ km/s. Dynamic traffic follows sinusoidal patterns $D_i(t) = \mu_i(1 + 0.5\sin(2\pi t/T)) + \mathcal{N}(0, \sigma_i^2)$, where $i \in \{\text{TN}, \text{NTN}\}$ represents network type, $\mu_{\text{TN}} = 50$ Mbps, $\mu_{\text{NTN}} = 30$ Mbps, and $\mathcal{N}(0, \sigma_i^2)$ denotes Gaussian noise. Co-channel interference between TN and NTN is $I = \frac{1}{K}\sum_{k=1}^{K} |L_{\text{TN}}^{(k)} - L_{\text{NTN}}^{(k)}|$, where L denotes the percentage load on channel k.

B. Q-learning formulation

The spectrum allocation problem is formulated as MDP $\langle \mathcal{S}, \mathcal{A}, \mathcal{R}, \gamma \rangle$ where the state space $\mathcal{S}$ discretises average TN load (10 bins), average NTN load (10 bins), interference level (5 bins), and time period (4 bins) for a total of $|\mathcal{S}| = 2000$ states. The action space $\mathcal{A} = \{1, \ldots, K\}$ represents channel selection. Each action selects a single channel shared across all users, with load-based interference modeling capturing multi-user contention effects. The multi-objective reward function can be expressed as

$$\begin{aligned} R(s,a) =& \text{Throughput}(s,a) - 0.3 \cdot \text{LoadImbalance}(s) \\ & - 0.2 \cdot I(s) + 10 \cdot \mathcal{F}(s), \end{aligned} \tag{1}$$

where $\mathcal{F} = (\sum_i L_i)^2/(2K\sum_i L_i^2)$ is Jain's fairness index. The coefficients of the multi-objective reward function were selected empirically based on preliminary experiments to achieve a balanced trade-off. The Q-learning update rule is $Q(s,a) \leftarrow Q(s,a) + \alpha[R + \gamma \max_{a'} Q(s',a') - Q(s,a)]$ with the learning rate $\alpha = 0.1$ and discount factor $\gamma = 0.95$. An ϵ-greedy exploration strategy decays from 1.0 to 0.01 over training episodes.

IV. Simulation Results

A. Training Convergence and Policy Behavior Analysis

Simulations were carried out in MATLAB R2025b [1] using 5,000 training episodes of 100 steps each. The Q-learning agent was compared to a random allocation baseline, which selects channels at random. Figs. 2, 3, and 4 together show how the proposed Q-learning-based TN–NTN spectrum allocation framework learns, how it uses the spectrum, and how its policies work.

Fig. 2 shows that the learning process converges steadily over 5,000 training episodes. The 20-episode moving average reduces the significant variance of instantaneous rewards and shows that convergence happens after about 2,000 episodes with an average reward of approximately 37.5 per episode. The throughput settles at approximately 28.5 Mbps, while Jain's fairness index stays at 0.75 ± 0.08, showing that fairness between terrestrial and non-terrestrial users was maintained during the training.

Fig. 3 shows how the learned policy affects spectrum usage. Channel 1 is the most used resource, with 39% of the total usage. The other channels have balanced usage, with 4% to 12% of the total usage, which demonstrates effective load distribution. The channel selection heatmap in Fig. 3(b) shows that the test episodes were able to adjust over time. The repeated selection of Channels 1, 8, and 10 shows that the learned policy adapts to temporal network dynamics.

Finally, Fig. 4 shows how the learned policy structure works. The Q-values for the states shown in Fig. 4(a) demonstrate a distinct difference between actions. In general, the values go up as the channel index goes up, which means that the agent has learned to prefer certain actions based on the properties of the channel. The distribution of maximum Q-values across all states in Fig. 4(b) shows that policies are coming together. The distribution is not symmetrical; most states have low maximum Q-values, while a few have much higher values. This suggests that certain state-action pairs yield higher expected returns.

[1] Results from a single representative run. The simulation code is available here for reproduction: https://doi.org/10.5281/zenodo.18468333.

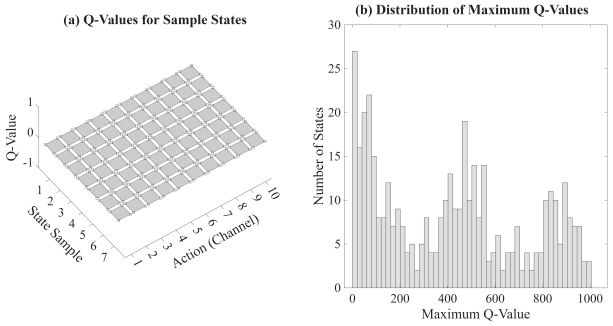

Fig. 4. Q-value analysis: (a) Q-values for sample states across action channels and (b) histogram of maximum Q-values per state.

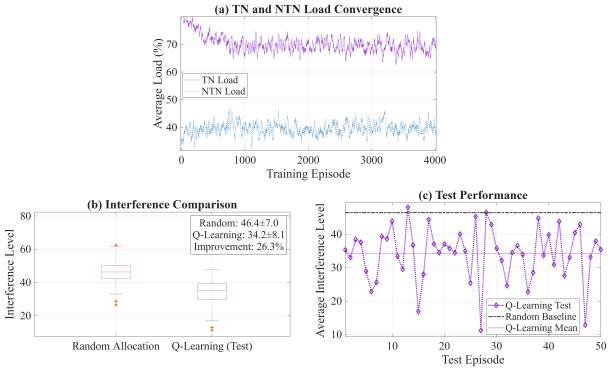

Fig. 5. Performance evaluation: (a) TN and NTN load convergence, (b) interference level comparison between random allocation and Q-learning, and (c) test-phase interference performance.

B. Interference Mitigation and Load Balancing

We evaluate the trained policy's ability to reduce interference and distribute load to see if the Q-learning approach is really working. Fig. 5(a) shows the load convergence behaviour, which shows how the Q-learning agent balances spectrum allocation between terrestrial and non-terrestrial networks during training. Initially, the TN load exceeds 80% but stabilises around 70% within the first few hundred episodes. The NTN load, on the other hand, stays consistent at around 40%. This indicates that the agent learns to maintain distinct but stable operating points for each type of network to avoid excessive contention over shared spectrum resources.

The statistical comparison in Fig. 5(b) shows strong evidence of less interference. The boxplot makes it evident that random allocation results in a median interference level of about 46.4 with a lot of variation. On the other hand, the Q-learning approach gets a median of about 34.2 during test episodes, which is much lower. This is a drop of more than a quarter in average interference, or 26.3%. This indicates improved signal quality and there is less packet loss in real-world use.

Fig. 5(c) illustrates test performance across 50 independent episodes. This shows how the implemented policy acted without any exploration noise. The Q-learning agent always keeps interference levels far lower than the random baseline. Most test episodes had interference levels between 25 and 45. The Q-learning mean, represented by the horizontal red line, stays around 34.2, while the random baseline is around 46.4. This is because traffic patterns and network circumstances change all the time. Some episodes had very little interference, going below 20. This shows that the policy can take advantage of good network states when they happen.

V. Conclusion

This research presented a Q-learning methodology for intelligent spectrum allocation in integrated TN-NTN systems. Simulations spanning 5,000 episodes exhibited consistent convergence, achieving a throughput of approximately 28.5 Mbps and a fairness index of 0.75. The learned policy shows that it can choose less congested channels while maintaining fair resource distribution among users. While both approaches handle balanced traffic, Q-learning provides significant advantages during dynamic traffic variations through fairness assurances via multi-objective optimisation and comprehensible policies. Future work will explore deep Q-networks for bigger state spaces, multi-agent RL for making decisions across satellites, comparisons with greedy and DQN baselines, and digital twin-based policy validation.

References

[1] C. T. Nguyen *et al.*, "Emerging Technologies for 6G Non-Terrestrial-Networks: From Academia to Industrial Applications," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 3852–3885, 2024. DOI: 10.1109/OJCOMS.2024.3418574

[2] M. Giordani and M. Zorzi, "Non-Terrestrial Networks in the 6G Era: Challenges and Opportunities," *IEEE Network*, vol. 35, no. 2, pp. 244–251, 2021. DOI: 10.1109/MNET.011.2000493

[3] A. Guidotti *et al.*, "Role and Evolution of Non-Terrestrial Networks Toward 6G Systems," *IEEE Access*, vol. 12, pp. 55945–55963, 2024. DOI: 10.1109/ACCESS.2024.3389459

[4] F. Fourati and M.-S. Alouini, "Artificial intelligence for satellite communication: A review," *Intelligent and Converged Networks*, vol. 2, no. 3, pp. 213–243, 2021. DOI: 10.23919/ICN.2021.0015

[5] S. Mahboob and L. Liu, "Revolutionizing Future Connectivity: A Contemporary Survey on AI-Empowered Satellite-Based Non-Terrestrial Networks in 6G," *IEEE Communications Surveys & Tutorials*, vol. 26, no. 2, pp. 1279–1321, 2024. DOI: 10.1109/COMST.2023.3347145

[6] Elhachmi, J., "Distributed reinforcement learning for dynamic spectrum allocation in cognitive radio-based internet of things," *IET Networks*, vol. 11, no. 6, pp. 207–220, 2022. DOI: 10.1049/ntw2.12051

[7] N. Naderializadeh, J. J. Sydir, M. Simsek, and H. Nikopour, "Resource Management in Wireless Networks via Multi-Agent Deep Reinforcement Learning," *IEEE Transactions on Wireless Communications*, vol. 20, no. 6, pp. 3507–3523, 2021. DOI: 10.1109/TWC.2021.3051163

[8] S. O. Adebayo, A. Barnawi, T. Sheltami, and M. Felemban, "Dynamic spectrum sharing in heterogeneous wireless networks using deep reinforcement learning," *Internet of Things*, vol. 32, p. 101635, 2025. DOI: 10.1016/j.iot.2025.101635

[9] U. C. Ukpong *et al.*, "Deep reinforcement learning agents for dynamic spectrum access in television whitespace cognitive radio networks," *Scientific African*, vol. 27, p. e02523, Mar. 2025. DOI: 10.1016/j.sciaf.2024.e02523

[10] Y. Cho, W. Yang, D. Oh, and H.-S. Jo, "Multi-Agent Deep Reinforcement Learning for Interference-Aware Channel Allocation in Non-Terrestrial Networks," *IEEE Communications Letters*, vol. 27, no. 3, pp. 936–940, Mar. 2023. DOI: 10.1109/LCOMM.2023.3237207

[11] M. Wang, X. Liu, F. Wang, Y. Liu, T. Qiu, and M. Jin, "Spectrum-efficient user grouping and resource allocation based on deep reinforcement learning for mmWave massive MIMO-NOMA systems," *Scientific Reports*, vol. 14, no. 1, p. 8884, Apr. 2024. DOI: 10.1038/s41598-024-59241-x