

Deep Reinforcement Learning for User Reconnection in Software-Defined Large-Scale LEO Satellite Networks

Yiming Fu, Wei Su

School of Electronic and Information Engineering
Beijing Jiaotong University
Beijing, China
Email: {25110090, wsu}@bitu.edu.cn

Bin Yang

School of Computer and Information Engineering
Chuzhou University
Chuzhou, China
Email: yangbinchi@gmail.com

Alex Hou

Tsinghua International School
Beijing, China
Email: 17800211460@163.com

Abstract— The high dynamic characteristic of Low Earth Orbit (LEO) satellites makes them susceptible to user service interruptions when facing node damage caused by space debris or malicious attacks. Although traditional heuristic algorithms have fast computation speeds, they are difficult to ensure the global optimality of reconnection schemes in large-scale networks. However, precise algorithms such as backtracking often face the problem of high computational complexity. This paper designs a Deep Reinforcement Learning (DRL) algorithm for user reconnection in software-defined large-scale LEO satellite networks. The satellites use the software-defined network (SDN) controller to achieve global programmability of the network, and cooperates with the ground SDN data layer to ensure basic transmission, forming an integrated air ground network model. We formulate the total system cost minimization as an optimization problem which is solved by the DRL algorithm under multi-scenarios with node damage. Meanwhile, a backtracking method based on pruning is introduced as a baseline for a theoretical optimal solution to the problem, and compared with the solution of traditional heuristic algorithms. Experimental results show that the DRL algorithm proposed in this paper can approximate the optimal solution of backtracking method with low time cost, and is superior to traditional heuristic algorithms, effectively balancing solution accuracy and real-time performance.

Index Terms—DRL; Non-Terrestrial Network; SDN; Satellite Communication; User Reconnection

I. INTRODUCTION

With the development of digital economy, the coverage capacity of traditional terrestrial network has reached the limit. Non-Terrestrial Network (NTN) is a communication technology based on 3GPP R17 and subsequent standards [1]. By integrating satellite communication, air network and ground network, NTN is committed to building a seamless connection architecture covering the sky, ground and near-earth space, expanding the service range from limited ground to wide area three-dimensional space, and

providing all-round and multi-dimensional connection support for various application scenarios. Its architecture includes two modes: transparent forwarding (the satellite is only used as signal relay) and base station on satellite (the satellite carries the base station function). It not only supports the smart phone directly connected to the satellite to realize emergency communication and global voice data services, but also provides wide area connection for Internet of Things (IoT) terminals, which is widely used in marine monitoring, emergency rescue, remote area coverage and other scenarios. As an important supplement to cellular communication, NTN is continuously evolving with 5G-Advanced, and will become the core technology for building three-dimensional networks in the 6G era, promoting the realization of the goal of global ubiquitous access.

Facing the complex challenges of high dynamic changes in NTN topology and heterogeneous resource management, the introduction of SDN architecture has become the key to breaking the situation. By deeply integrating its concept of "separation of control and forwarding" into the satellite communication system, it gives full play to the global intelligent scheduling advantage of SDN [2]. The centralized controller is used to uniformly arrange a large number of satellite nodes, which can not only dynamically optimize the on-board resource allocation, but also calmly cope with the frequent switching caused by the high-speed movement of low earth orbit satellites, so as to realize the intelligent routing and seamless connection of cross domain links, so as to break the bottleneck of the rigid function of the traditional satellite network and provide the core support for the construction of the intelligent communication network of air ground integration.

However, the satellite operating environment is

complex, and node damage (such as space debris collision, malicious attack interference, and electronic component failure [3]) can cause a large number of users under its coverage to instantly lose contact, seriously threatening the survivability of the network. Currently, research is dedicated to building a comprehensive recovery system that integrates algorithms and cutting-edge technologies, utilizing LEO satellites, high-altitude platform stations (HAPS), and other technologies to construct NTN as critical backup or alternative means to ensure rapid recovery of connections in the event of ground network unavailability. Reference [4] proposes a robust routing algorithm based on extended path regions to resist link failures. When a link failure occurs on the main path, the algorithm can prioritize multi hop mode for route recovery, effectively improving delivery rate. Reference [5] proposes a recovery model based on Network Function Virtualization to address the possibility of node failure in the integrated network of heaven and earth. Tasks can be restored by using matching games to quickly select alternative resources and redeploy the Service Function Chain. Reference [6] utilizes Constrained Multi Agent Reinforcement Learning (CMADR) to adjust routing strategies, minimizing communication delay and achieving dynamic self-healing of the network when congestion or local interruptions occur. Reference [7] proposes a diffusion model-based artificial intelligence (AI) algorithm that utilizes LEO satellites in NTN for round-trip transmission in case of terrestrial network interruption or overload, ensuring maximum transmission capacity in emergency situations and restoring data connectivity in disaster areas. Reference [8] explores a solution that can restore connectivity through satellite links, ensuring lower latency communication even under challenging conditions, and supporting public safety and emergency management through forward-looking disaster recovery solutions.

With the development of AI, deep reinforcement learning (DRL) has been introduced into satellite network resource management. Reference [9] designs an autonomous decision-making algorithm based on deep reinforcement learning, which enables LEO satellite network to perceive real-time network status, notably improving end-to-end performance and resource utilization efficiency in dynamic environments. Reference [10] proposed a centralized LSN architecture that utilizes DRL to achieve long-term optimization of LSN energy consumption performance while satisfying end-to-end delay constraints for each data stream. However, existing research lacks a strict "optimal solution" benchmark to evaluate the approximation degree of DRL algorithm solutions. Meantime, how to quickly and cost effectively re-

connect damaged users to neighboring satellites by utilizing the global view advantage of SDN in a short period of time is the current research difficulty. The existing methods for user reconnection can be mainly divided into two types: one is heuristic algorithms based on distance or load (such as greedy algorithms), which are computationally simple but prone to getting stuck in local optima, leading to uneven network load; The other type is based on exhaustive exact algorithms (such as backtracking), which can obtain the global optimum, but the computational complexity increases exponentially when faced with large-scale user reconnection, making it difficult to meet real-time requirements. Thus, this paper aims to explore a solution that falls between "optimality" and "real-time", and proposes a user reconnection optimization method based on DRL. This method uses "separation of control and forwarding", the feature of SDN to endow LEO satellites with global perception and centralized control capabilities. The system monitors real-time user status through satellites. When it is determined that the damage has reached the threshold that affects the basic functions of the network, a backup channel is quickly established to achieve seamless switching of communication services and efficient use of communication resources.

The main contributions of this paper are as follows: **Firstly**, a satellite-user network model is constructed, fully leveraging the advantages of wide area coverage of satellites and global scheduling of SDN, and modeling the user reconnection problem after satellite damage as a system total cost minimization problem under multiple constraints. **Secondly**, an optimization method for user reconnection based on DRL is designed, which models the problem of "damaged user reconnection" as a sequential decision-making process. By training a Deep Q-network (DQN), the optimal node is selected for each user considering the current load and future impact, thereby minimizing the total system cost of the entire reconnection process and achieving fast decision-making. **At the same time**, in order to quantify the accuracy of the DRL algorithm, experiments are conducted in a simulation environment using the DRL algorithm, a traditional heuristic algorithm, and the backtracking method. The experimental results demonstrate that the DRL algorithm performs better than heuristic algorithms in multiple scenarios, and can approach the optimal solution obtained by backtracking with low time cost, filling the gap that lacks comparison of "Ground Truth".

The reminder of this paper is organized as follows. Section II presents system model. In Section III, we conduct problem formulation and solve this problem. In Section IV, we provide experimental results. We conclude this paper in Section V.

II. SYSTEM MODEL

A. Network Model

As shown in Fig. 1, we establish an LEO satellite network model based on SDN, which consists of S satellites and U users. The architecture gets over the limitations of traditional layered isolation by deeply integrating the satellite layer and ground layer to construct a communication system with full coverage and dynamic collaboration. The entire network is logically divided into a control plane and a data plane, utilizing the global view ability of SDN to achieve unified scheduling and intelligent orchestration of heterogeneous resources, laying a topological foundation for responding to complex and changing emergency scenarios.

The top satellite layer consists of a collaborative constellation of multiple LEO satellites, serving as the "space-based backbone" of the network with the advantage of wide area coverage. Some of the satellites are responsible for basic communication, while others are responsible for the global control signaling and real-time perception of network status, playing a critical role as redundant backup links in extreme scenarios. When the communication link is interrupted, the satellite layer can quickly activate the satellite ground direct connection channel, providing a reliable data transmission path for the lower layer and ensuring the continuity of core instructions in the risk of network disconnection.

The bottom ground layer gathers diverse heterogeneous facilities, including ground user terminals, communication base stations, and ground control centers. Among them, the ground control center serves as the brain of the entire system, continuously monitoring the operational status and link quality of nodes. Once the local link interruption caused by node failure is detected, the control center immediately starts the recovery program using the global view, issues flow table instructions to trigger the topology reconstruction strategy, and commands satellite nodes to replenish resources.

Facing emergency scenarios such as sudden disasters or large-scale node failures, this model demonstrates excellent self-healing capabilities. The system has achieved a basic transformation from "single point dependence" to "multi-dimensional supplement" through dynamic reorganization of links between satellite and ground and optimized scheduling of transmission resources. This mechanism significantly enhances survival capability in dynamic and uncertain environments, providing vital communication support for disaster relief and infrastructure protection.

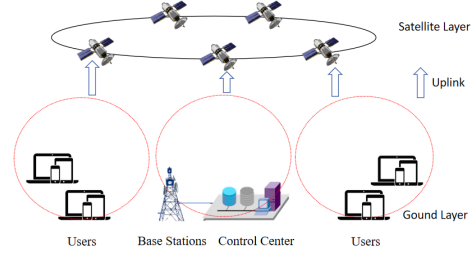


Fig. 1. LEO satellite network model based on SDN

B. Communication Model

When constructing a communication model for SDN based LEO satellite network, due to the existence of stable direct view path signals in the communication scenario between satellites and ground users, the Rayleigh distribution is used to model their channels. Although there is a stable direct line of sight path in the Rayleigh channel, satellite systems using frequency reuse technology will face serious inter beam interference and inter satellite interference, which can limit system capacity and even cause signal blockage. To achieve efficient multi-user access and overcome wireless channel interference, the system adopts multiple antenna technology and Orthogonal Frequency Division Multiple Access (OFDMA) technology at the receiving end. By decomposing the high-speed data stream into multiple parallel low-speed subcarriers for transmission, it effectively suppresses Inter Symbol Interference (ISI) and improves spectrum utilization; Meanwhile, linked to strict coordination mechanisms between satellite ground and inter satellite spectra, the frequency band resources of each beam are allocated according to real-time traffic requirements to ensure the reliability of the dominant communication path, thereby providing reliable support for multi-user concurrent communication. Additionally, priority allocation mechanism is activated to prioritize available channels for high priority users (such as emergency scenarios and critical instructions) and ensure the continuity of important information.

III. PROBLEM FORMULATION AND SOLUTION

A. Problem Formulation

Fig. 2 shows the optimization way for user reconnection based on SDN, which adopts an adaptive scheduling strategy to achieve hierarchical self-healing from "local recovery" to "collaborative takeover". When the system detects a failure in a satellite node through real-time monitoring, it will activate a hierarchical recovery mechanism based on the damage assessment algorithm: if local or mild damage is detected, the system will prioritize the use of an ad hoc network-based recovery strategy. Through dynamic routing reconstruction and task

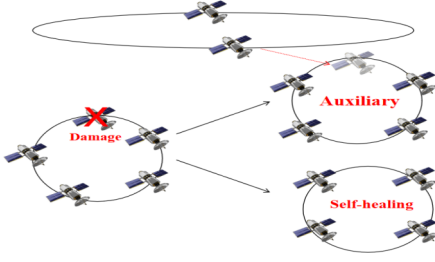


Fig. 2. An optimization method for user reconnection based on SDN

reallocation, communication tasks of the damaged node will be taken over by neighboring satellite nodes to maintain regional service continuity; If severe damage occurs and remaining satellite resources cannot meet users' needs of communication, the auxiliary communication mode will be triggered. At this time, the system will establish a communication link with the ground user using a new satellite to ensure uninterrupted business. This adaptive self-healing strategy ensures that the core business can obtain highly reliable transmission guarantees. We define C as the total cost incurred by all users establishing connections and communicating with satellites, calculated by summing the cost of each satellite.

$$C = \sum_{s=1}^S C_s \quad (1)$$

where C_s represents communication cost incurred by satellite s and its connected users, calculated as the sum of receiving cost and processing cost.

$$C_s = \sum_{u=1}^U \left(L_u \cdot E_{per_Bit} + \frac{L_u}{R_{u,s}} \cdot E_{per_Second} \right) \quad (2)$$

where E_{per_Bit} is energy consumption per bit of data received, E_{per_Second} is energy consumption for processing data per second. L_u is the packet length of user u , while $R_{u,s}$ is the transmission rate between user u and satellite s , calculated according to Shannon's theorem.

$$R_{u,s} = B \log_2 (1 + \text{SINR}_{u,s}) \quad (3)$$

where B is the channel bandwidth, while $\text{SINR}_{u,s}$ is the Signal to Interference plus Noise Ratio of user u communicating with satellite s , calculated according to its definition.

$$\text{SINR}_{u,s} = \frac{P_u g_{u,s}}{\sum_{s' \in S \setminus \{s\}} \zeta_{s'} + \sigma^2} \quad (4)$$

where $g_{u,s}$ is the transmission power of user u , and $g_{u,s}$ is the channel gain. σ^2 is the noise power during transmission, and ζ_s is the interference power of satellite s , calculated according to the free space loss formula.

$$\zeta_s = 92.45 + 20 \log_{10}(f) + 20 \log_{10}(d) \quad (5)$$

where f is the frequency of the electromagnetic wave, and d is the propagation distance of the electromagnetic wave.

When constructing the system performance evaluation model, the reception cost of satellite during transmission, as well as the processing cost during data demodulation and forwarding, are considered. By calculating the consumption of these two key resources, a comprehensive optimization equation is constructed. Seeking the optimal solution for resource allocation and energy efficiency optimization while meeting system performance indicators. The optimization equation is as follows.

$$\min C \quad (6)$$

$$\text{s.t.} \quad \sum_{s=1}^S x_{u,s} = 1, \forall u \in U \quad (7)$$

$$x_{u,s} \leq 1 - \delta_s, \forall s \in S, \forall u \in U \quad (8)$$

$$\sum_{s=1}^S x_{m,s} \sum_{s=1}^S x_{n,s}, \forall m, n \text{ s.t. } p_m > p_n \quad (9)$$

$$\sum_{j=1}^J z_{i,j} = 1, \forall i \quad (10)$$

$$\sum_{i=1}^I z_{i,j} \leq q, \forall j \quad (11)$$

Equation (7) defines the main association constraint, while each user can only choose one satellite as the main service node. $x_{u,s}$ represents user satellite association variables. $x_{u,s}=1$ indicates that user u establishes a connection with satellite s , otherwise it is 0. Equation (8) defines the satellite constraint, while damaged satellites cannot serve users. δ_s represents the fault state of satellite s . $\delta_s=0$ indicates that satellite s is normal, otherwise it is 1. Equation (9) defines the priority constraints, while high priority users are given priority in obtaining services. p_u represents the priority of user u . Equations (10) and (11) define the channel constraint, while each user can only be assigned to one sub channel, and each sub channel allows q users to share at most. $z_{i,j}$ represents the channel associated variable. $z_{i,j}=1$ indicates that user i is assigned to sub channel j , otherwise it is 0.

B. Problem Solution

To solve the above optimization equation, this paper proposes a DRL algorithm based on DQN, which utilizes the idea of Time Difference (TD) learning to drive model convergence by minimizing the mean square error between Q predicted values and Q target values. The reward function of the algorithm is designed as the core element, with the goal of minimizing the total system cost. When the agent chooses a legal action, the reward value is negative for the total cost to encourage minimizing the cost; If an illegal action is chosen, a fixed penalty value

is applied to constrain the legality of the action. The solving process is divided into four stages. Firstly, in the environment construction stage, the user reconnection problem is modeled as a Markov decision process, and the state space, action space, and reward function are defined; Next, during the network training phase, the DQN architecture and experience replay mechanism are utilized to enable the agent to steadily update the Q-value through multiple rounds of "exploration utilization" cycles, continuously trial and error, and gradually approach the optimal strategy with the help of a dual network structure; Subsequently, in the inference and decision-making stage, the trained intelligent agent processes disconnected users in the testing environment in priority order, and generates a preliminary reconnection plan based on legitimacy verification; Finally, in the result integration stage, the minimum cost of reconnection is added up with the fixed cost of undamaged users, and a complete system recovery view, including connection relationships, load statistics, and total energy consumption is output. The $Q(s_t, a_t)$ represents the long-term cumulative reward prediction value for selecting action a_t under state s_t to gradually approach the optimal action value function through the collaborative updating of the Q-network and the target network and guide the agent to balance exploration and utilization. The specific steps of the algorithm are shown in the table below.

Algorithm 1: DQN-based User Reconnection Algorithm for Damaged Satellite Network

Input: Damaged Satellite, Episode E , Time T
Output: Minimum Cost C
for episode $e = 1$ **to** E **do**
 Initialize state $s_{t=1}$ by resetting the environment
 for time $t = 1$ **to** T **do**
 With probability ϵ : $a_{t+1} \leftarrow \text{Random}(0, d - 1)$
 Otherwise: $a_{t+1} \leftarrow \arg \max Q(s_t, a_t; \theta)$
 Perform a_t , observe reward r_t and next state s_{t+1}
 Store transition (s_t, a_t, r_t, s_{t+1}) in Replay Buffer D
 Calculate Loss: $L \leftarrow E[(r + \gamma \max Q_{\text{target}}(s_{t+1}) - Q(s_t, a_t))^2]$
 if Terminal state reached **then break**
 end for
end for
return E

IV. EXPERIMENTAL RESULTS

To verify the accuracy of the algorithm proposed in this paper, we use the optimal solution obtained by backtracking as the baseline for performance evaluation and compare it with the solution obtained by traditional heuristic algorithms. The backtracking method uses a depth first search strategy to systematically traverse all possible user satellite allocation combinations, ensuring that the global optimal solution can be found. In the search process, a pruning strategy based on load constraints is introduced: when it is found during the recursive process that the

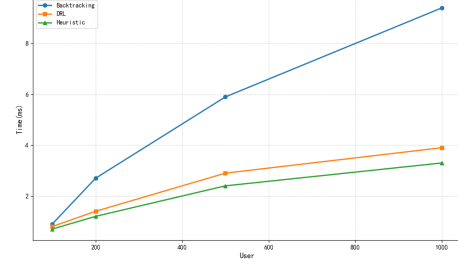


Fig. 3. Average runtime comparison

load of the current partial solution has exceeded the satellite carrying capacity, or the current cumulative cost has exceeded the known feasible solution (such as the solution obtained by greedy algorithm), the search for that branch is immediately terminated and backtracked, effectively reducing unnecessary calculations. Although the backtracking method can ensure the optimality of the solution, its time complexity increases exponentially, the computational cost is high, and it is difficult to meet real-time need. In contrast, traditional heuristic algorithms adopt a greedy strategy, assigning users to the satellites with the lowest current load and the ability to establish connections based on the principle of the closest distance between users.

Although this algorithm has high computational efficiency, it is prone to getting stuck in local optima and cannot guarantee obtaining global optima, especially in scenarios with dense user distribution or uneven satellite load, where performance loss is more pronounced. This experiment conducted 100 experiments in multi-user scenarios using backtracking, the DRL algorithm, and the heuristic algorithm, and obtained the average runtime and average system energy consumption for each experiment. The experimental results are shown in the following figures.

From the comparison of average runtime in Fig. 3, it can be seen that when the number of users is small, the runtime of DRL algorithm is about 11% lower than that of the backtracking method and about 13% higher than that of heuristic algorithm. As the number of users gradually increases, the runtime of DRL algorithm is about 55% lower than the backtracking method and about 16% higher than the heuristic algorithm.

This indicates that as the number of users increases, the DRL algorithm is significantly lower than the backtracking method and has runtime similar to heuristic algorithms. This is because the backtracking method requires traversing all possible user satellite allocation combinations, and the time complexity is usually $O(k^n)$ (where k is the number of satellites and n is the number of users), so an increase in the number of users will lead to a significant increase in

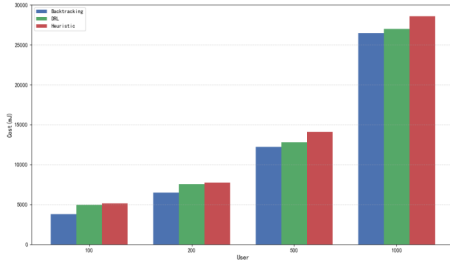


Fig. 4. Average cost comparison

computational complexity; Heuristic algorithms typically adopt a greedy strategy, with a time complexity of $O(n)$ or $O(n \log n)$, resulting in slow time growth. The DRL algorithm proposed in this paper performs fast inference through a single forward propagation of neural networks, with controllable time complexity. Therefore, it outperforms backtracking in terms of running efficiency and is close to heuristic algorithms.

From the comparison of average system energy consumption in Fig. 4, when the number of users is small, the DRL algorithm has a system energy consumption that is about 8% higher than the backtracking method and about 12% lower than the heuristic algorithm; As the number of users gradually increases, the system energy consumption of the DRL algorithm is about 3% higher than the backtracking method and about 16% lower than the heuristic algorithm. This indicates that the energy consumption of the DRL algorithm is slightly higher than that of the backtracking method, and the gap decreases with the increase of the number of users. The energy consumption is lower than that of heuristic algorithms, and the gap widens as the number of users increases. This is because the backtracking method can search for all possible solutions and select the optimal solution, resulting in the lowest energy consumption. The heuristic algorithm, on the other hand, is prone to getting stuck in local optima due to its simple rules such as nearest distance or minimum load, leading to low resource allocation efficiency and high energy consumption. The DRL algorithm proposed in this paper can find a strategy that is close to the optimal one by learning historical data and interacting with the environment, so the energy consumption is close to the backtracking method.

In practical engineering applications, system design often needs to seek the optimal balance between performance and efficiency. In contrast with the backtracking method and the heuristic algorithm, the DRL algorithm combines better running efficiency than the backtracking method and lower system energy consumption than the heuristic algorithm, becoming the best compromise solution that balances performance and efficiency.

V. CONCLUSION

This paper focuses on the optimization problem of user reconnection in the scenario of node damage in LEO satellite network. This paper proposes a solution based on DRL to address the contradiction between the poor quality of traditional heuristic algorithms and the difficulty of precise algorithms in meeting real-time requirements. By introducing the pruning based backtracking method as the "theoretical optimal solution" benchmark, the approximation ability of DRL algorithm is verified, and it is proved that it is superior to traditional heuristic algorithms in multiple scenarios. Multi scale scenario experiments have shown that the DRL algorithm effectively reduces the total system cost while maintaining a faster speed than the backtracking method. Future work will notice the cascading effect of multi satellite concurrent failures and explore distributed DRL training strategies under federated learning frameworks to further enhance model generalization ability.

REFERENCES

- [1] L. Chhangte and A. S. Patel, "Integrated terrestrial and non-terrestrial-network architecture for edge caching and coded delivery in 3gpp system," in *2025 17th International Conference on COMMunication Systems and Networks (COM-SNETS)*, (Bengaluru, India), pp. 1226–1229, 2025.
- [2] J. Dafda and M. Subhedar, "Dynamic load balancing in sdn using energy aware routing and optimization algorithm," in *2022 IEEE Bombay Section Signature Conference (IBSSC)*, (Mumbai, India), pp. 1–6, 2022.
- [3] M. Sheng, D. Zhou, S.-J. Ji, W.-G. Bai, Y. Zhu, J.-Y. Liu, and J.-D. Li, "Effects of space environment on satellite mega-constellations: From nodes and links to network performance," *Engineering*, vol. 54, pp. 93–102, 2025.
- [4] H. Nie *et al.*, "A robust routing algorithm against link failures for leo satellite networks," in *2024 IEEE 100th Vehicular Technology Conference (VTC2024-Fall)*, (Washington, DC, USA), pp. 1–5, 2024.
- [5] Z. Jia *et al.*, "Nfv-enabled service recovery in space-air-ground integrated networks: A matching game-based approach," *IEEE Transactions on Network Science and Engineering*, vol. 12, no. 3, pp. 1732–1744, 2025.
- [6] C. R. Prasad, A. Maniatus, and A. Katti, "Rerouting and load balancing on leo satellite networks," in *2024 IEEE Cloud Summit*, (Washington, DC, USA), pp. 87–94, 2024.
- [7] F. Khoramnejad and E. Hossain, "Carrier aggregation, load balancing, and backhauling in non-terrestrial networks: Generative diffusion model-based optimization," *IEEE Transactions on Wireless Communications*, vol. 24, no. 5, pp. 4483–4499, 2025.
- [8] H. Gupta, N. R. Kuraku, R. Tripathi, and T. S., "Design and implementation of a resilient communication network for disaster recovery using satellite links," in *2025 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*, (Chennai, India), pp. 1–8, 2025.
- [9] J. Qiao, "Dynamic resource management and routing algorithm for low-earth orbit satellite networks based on deep reinforcement learning," in *2025 6th International Symposium on Computer Engineering and Intelligent Communications (ISCEIC)*, (Chongqing, China), pp. 44–48, 2025.
- [10] Y.-H. Hsu, J.-I. Lee, and F.-M. Xu, "A deep reinforcement learning based routing scheme for leo satellite networks in 6g," in *2023 IEEE Wireless Communications and Networking Conference (WCNC)*, (Glasgow, United Kingdom), pp. 1–6, 2023.