Scalable Beam Hopping via Hierarchical Multi-Agent Reinforcement Learning

Qiaolin Ouyang [1], Neng Ye [2], Aihua Wang [1] and Wonjae Shin[3]

[1]School of Information and Electronics, Beijing Institute of University, Beijing, China
[2]School of Cyberspace Science and Technology, Beijing Institute of Technology, Beijing, China
[3]Department of Communication Engineering, Korea University, Seoul, Republic of Korea
E-mail: qlouyang@bit.edu.cn, ianye@bit.edu.cn, wah@bit.edu.cn ,wjshin@korea.ac

Abstract

Beam hopping efficiently addresses non-uniform traffic distribution in satellite networks, and finer beam granularity significantly improves system performance. However, applying adaptive beam hopping with multi-agent deep reinforcement learning (MADRL) under fine granularity inevitably increases the complexity of observation and action spaces, making it difficult for agents to learn effectively. In this paper, we propose a hierarchical MADRL-based beam hopping strategy, which progressively narrows down decision granularity through a layered structure. Additionally, we introduce agent reuse across hierarchical layers to improve learning efficiency and scalability. Simulation results demonstrate that our approach effectively overcomes learning inefficiencies under fine-grained beam pointing scenarios. Compared with conventional MADRL methods, the hierarchical strategy achieves better adaptability, scalability, and higher performance, providing a promising solution for scalable fine-grained beam hopping in emerging satellite communication systems.

I. Introduction

The rapid evolution of satellite communication has become crucial in delivering global connectivity and supporting high-capacity data transmission, especially in geographically isolated and underserved areas [1], [2]. Multi-beam satellites (MBS) with flexible beam hopping (BH) technology efficiently allocate narrow, high-gain beams to terrestrial user terminals. Such dynamic and targeted beam allocation effectively enhances communication quality and addressing the uneven and unpredictable traffic demand distribution.

Recently, reinforcement learning (RL) approaches have gained traction as powerful tools for optimizing BH scheduling, enabling satellites to dynamically adapt to changing conditions through continuous online learning [3],[5]. Given that traffic-adaptive BH scheduling can be effectively modeled as a Markov Decision Process (MDP), deep reinforcement learning (DRL) methods have increasingly been adopted to navigate these complex decision-making scenarios. Although DRL techniques are proficient at generating near-optimal schedules under dynamic conditions, their practical implementation is significantly hindered by the rapid growth in complexity arising from expanding observation and action spaces when finer service granularity is pursued. This complexity negatively impacts learning efficiency and scalability.

To address these challenges, this paper introduces a hierarchical DRL framework aimed at reducing the complexity associated with fine-grained beam pointing. The proposed hierarchical architecture decomposes the decision-making process into multiple layers, systematically refining beam decisions at each stage. Moreover, by leveraging agent reuse across these hierarchical layers, the approach significantly improves computational efficiency and accelerates learning processes. Experimental simulations demonstrate that the hierarchical method effectively resolves the scalability and learning difficulties faced by conventional DRL approaches, ultimately providing enhanced adaptability and performance in scenarios requiring detailed beam hopping control.

II. System Model

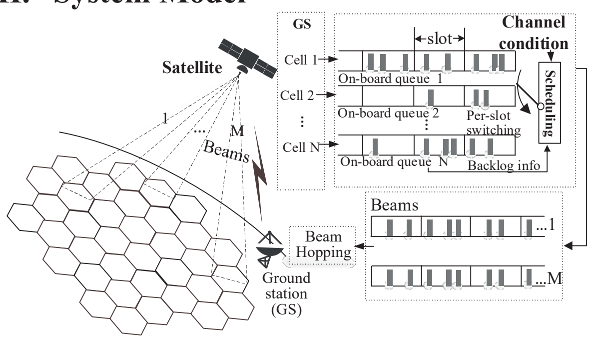

Figure 1. Satellite Beam hopping system.

We consider a beam-hopping communication system in which a satellite employs M beams with full-frequency reuse to serve a specified geographical region, as illustrated in Fig. 1. Data packets are transmitted from a ground station (GS) to the satellite and subsequently stored in corresponding transmission queues onboard. The satellite then delivers the queued data to terrestrial user terminals located within distinct coverage cells through beam hopping.

The signal-to-noise plus interference (SINR) level between the satellite and the terminals located in the -th cell is assumed to be uniform for simplicity, and is represented mathematically as :

$$\Gamma_{n,t} = \frac{|v_{n,n,t}\sqrt{p_{n,t}}|^2}{\sigma_n^2 B + \sum_{n' \in \mathcal{N}_i^- \setminus \{n\}} |v_{n,n',t}\sqrt{p_{n',t}}|^2},$$

where $v_{n,n',t}$ refers to the channel gain for the satellite beam towards the n'-th cell at the direction of the n-th cell, σ_n^2 is the noise level, B is the bandwidth, and $p_{n,t}$ is the transmit power. We define $\mathcal{N}_t^- = \{\mathrm{n} \mid \mathrm{x}_{\mathrm{m,t}}^{\mathrm{b}} = \mathrm{n}\}$ as the set of cells that are served by the satellite at time slot t,t.e., the beam hopping pattern. Then, the corresponding rate can be given as

$$R_{n,t} = I(n \in \mathcal{N}_t^-) B \log_2\left(1 + \Gamma_{n,t}\right)$$

The objective is to minimize the sum of gap between the required rate and the served rate,

$$\min_{x^{\mathrm{b}}} \frac{1}{T} \sum_{t=1}^{T} \sum_{n \in \mathcal{N}} \left(R_n^{\mathrm{req}} - R_{n,t} \right)$$
$$s.t.\ x_{m,t}^{\mathrm{b}} \in \mathcal{N}, \forall m \in \mathcal{M}, \forall t \in \mathcal{T}$$

III. Hierarchical Multi-Agent Reinforcement Learning for Scalable Beam Hopping

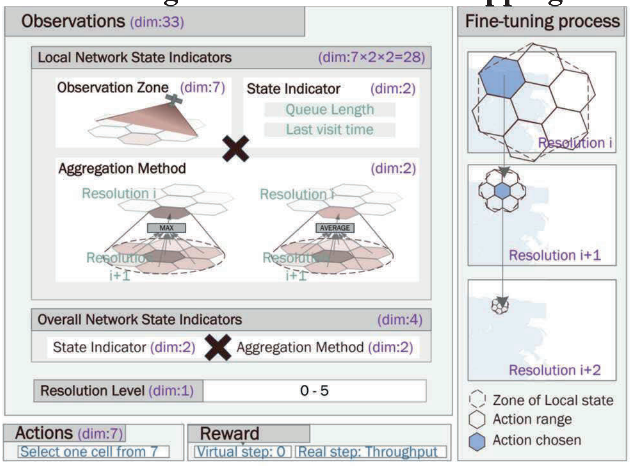

Figure 2. Hierarchical MADRL for Beam Hopping.

To achieve fine-grained beam control with scalable beam hopping, we propose a hierarchical MADRL framework that decomposes the decision space and reuses agents across layers, as shown in Fig. 2.

Using H3 spatial indexing, the system aggregates geographic information across resolutions to enable top-down beam scheduling with reduced complexity. The H3-based hierarchy enables agents to first make coarse-resolution decisions and then progressively refine them. Each agent operates at multiple resolutions using shared parameters, allowing it to generalize knowledge and reduce training cost. This layered refinement ensures scalable scheduling across large areas.

Each agent's observation includes global and local network states, other agents' recent actions, and the current resolution level. This structure ensures consistency and scalability across resolutions. Rewards are only issued at the lowest resolution based on actual service rates.

Finally, to improve generalization, agents are shared across resolutions. This cross-layer sharing leverages the scale-invariant nature of traffic and interference patterns, enhancing training efficiency and performance in heterogeneous DS2D networks.

IV. Evaluation Result

We evaluate the proposed hierarchical MADRL framework in a DS2D beam hopping system under realistic traffic conditions. The simulation considers a satellite at 600 km altitude covering around 5000 H3 cells. Each episode randomly selects a coverage area over China, with Poisson-distributed traffic based on real population data and an average demand of 2 Gbps.

We compare our hierarchical MAPPO method (H-MAPPO) against three baselines: Greedy scheduling based on queue length, vanilla MAPPO with flat decision space, and greedy algorithm. This setup allows us to assess both the scalability and cooperation efficiency of our approach.

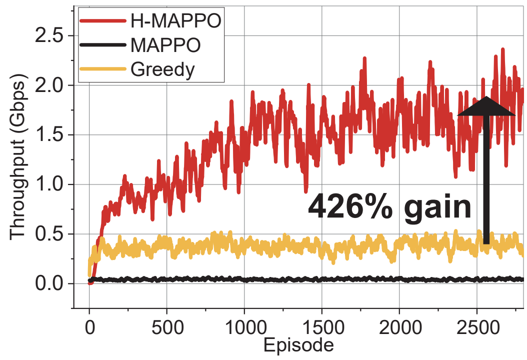

Figure 3. Evaluation result.

As shown in Fig.3, H-MAPPO achieves faster convergence and significantly better runtime efficiency than the baselines. While vanilla MAPPO suffers from large state/action spaces and slow learning, H-MAPPO benefits from hierarchical abstraction and agent reuse, maintaining high throughput.

V. Conclusion

This paper presents a hierarchical MADRL framework for scalable and fine-grained beam hopping in large-scale DS2D satellite networks. By leveraging H3-based spatial indexing, layered decision decomposition, and cross-resolution agent reuse, the proposed approach effectively reduces the complexity of DRL-based scheduling while maintaining high performance. Experimental results demonstrate that our method achieves superior training efficiency and system throughput compared to conventional baselines. The framework offers a practical and adaptable solution for next-generation satellite communication systems facing heterogeneous traffic demands and large-scale deployment scenarios.

Acknowledgement

This work was supported in part by the China Scholarship Council (CSC) grants (202406030123), and the Institute of Information & Communications Technology Planning \& Evaluation (IITP) grants (No.2024-00436887 and No.2021-0-00467)

References

[1] Z. M. Bakhsh, Y. Omid, G. Chen, F. Kayhan, Y. Ma, and R. Tafazolli, "Multi-satellite MIMO systems for direct satellite-to-device communications: A survey," *IEEE Commun. Surv. Tutorials*, pp. 1–1, 2024.
[2] G. M. Capez, S. Henn, J. A. Fraire, and R. Garello, "Sparse satellite con- stellation design for global and regional direct-to-satellite IoT services," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 5, pp. 3786–3801, 2022.
[3] G. Wang, F. Yang, J. Song, and Z. Han, "Resource allocation and load balancing for beam hopping scheduling in satellite-terrestrial communications: A cooperative satellite approach," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2024.
[4] Z. Lin, Z. Ni, L. Kuang, C. Jiang, and Z. Huang, "Satellite-terrestrial coordinated multi-satellite beam hopping scheduling based on multi- agent deep reinforcement learning," *IEEE Trans. Wireless Commun.*, vol. 23, no. 8, pp. 10091–10103, Aug. 2024.
[5] G. Xu, F. Tan, Y. Ran, Y. Zhao, and J. Luo, "Joint beam-hopping scheduling and coverage control in multibeam satellite systems," *IEEE Wireless Commun. Lett.*, vol. 12, no. 2, pp. 267–271, Feb. 2023.