SCIENCE CHINA
Information Sciences

• LETTER •

September 2026, Vol. 69, Iss. 9, 190309:1–190309:2
https://doi.org/10.1007/s11432-025-5052-0

Special Topic: User-Centric Intelligent Networking

Role-customized cross-domain task orchestration in satellite-terrestrial integrated networks

Yu ZHAO[1,2,3], Cheng WANG[1,2,3], Chungang YAN[1,2,3] & Changjun JIANG[1,2,3]*

[1]*School of Computer Science and Technology, Tongji University, Shanghai 201804, China*
[2]*Key Laboratory of Embedded System and Service Computing, Ministry of Education, Shanghai 201804, China*
[3]*Shanghai Artificial Intelligence Laboratory, Shanghai 200030, China*

Received 23 October 2025/Revised 13 March 2026/Accepted 20 July 2026/Published online 18 August 2026

Citation Zhao Y, Wang C, Yan C G, et al. Role-customized cross-domain task orchestration in satellite-terrestrial integrated networks. Sci China Inf Sci, 2026, 69(9): 190309, https://doi.org/10.1007/s11432-025-5052-0

The global network is evolving from isolated local connectivity toward wide-area intelligent collaboration. Satellite-terrestrial integrated networks (STINs), which integrate terrestrial infrastructures, low-Earth-orbit (LEO) satellites, and cloud centers, provide a promising architecture for ubiquitous service provisioning in areas with limited terrestrial coverage, constrained infrastructure, or heavy traffic loads [1]. LEO satellites complement terrestrial resources by offering wide-area connectivity and cross-domain communication-computation support for delay-sensitive and computation-intensive services in remote, maritime, and congestion-prone scenarios [2]. However, due to the time-varying visibility, dynamic link states, and limited onboard resources of LEO satellites, LEO-assisted task orchestration needs to be carefully balanced against their resource scarcity and service uncertainty. These features introduce stronger coupling among task partitioning, cross-domain offloading, and resource allocation in STINs. Moreover, as user service requests become increasingly concurrent and spatio-temporally dynamic, adaptive coordination among terrestrial edge, LEO satellite, and cloud resources becomes essential for improving service efficiency and user experience.

Existing multi-agent deep reinforcement learning (MADRL)-based task orchestration methods have demonstrated effectiveness for distributed decision-making in dynamic networks, as summarized in Appendix A. However, most studies either focus on terrestrial-only or satellite-only scenarios, or consider cross-domain task orchestration under relatively unified agent structures. Although some studies model different types of network entities as agents, they often rely on global information sharing or centralized value estimation, which may introduce scalability and adaptability challenges. The functional differences among satellite-terrestrial computing nodes, such as service responsibilities, executable actions, and operational constraints, are not explicitly characterized for cross-domain task orchestration.

To address these challenges, we propose a role-customized multi-agent cooperative learning framework for cross-domain task orchestration in STINs. The core idea is to map the functional roles of different agents to their observation design, action space, feasibility constraints, and reward feedback, constructing a cooperative learning model that supports local decision-making and cross-domain collaborative optimization. Specifically, role-customized agents are deployed at LEO satellites and base stations (BSs) for autonomous local decision-making, and these distributed agents are coordinated through limited information exchange and feedback integration, thereby balancing local autonomy with network-level cross-domain optimization. The overall architecture of the proposed framework is shown in Figure 1(a). The system model, including the task, communication, and computation models, is described in Appendix B, followed by the formulation of the cross-domain task orchestration problem in Appendix C.

Methods. To capture the dynamic and partially observable characteristics of STINs, we model $\mathcal{P}_1$ as a decentralized partially observable Markov decision process (Dec-POMDP) $\langle \mathcal{N}, \mathcal{M}, \mathcal{O}, \mathcal{A}, \mathcal{R}, P, \gamma \rangle$, where $\mathcal{N}$ and $\mathcal{M}$ denote the sets of LEO and BS agents, and $\mathcal{O} = \{\mathcal{O}_n\}_{n\in\mathcal{N}} \cup \{\mathcal{O}_m\}_{m\in\mathcal{M}}$ and $\mathcal{A} = \{\mathcal{A}_n\}_{n\in\mathcal{N}} \cup \{\mathcal{A}_m\}_{m\in\mathcal{M}}$ denote the role-specific observation and action spaces. We consider two heterogeneous roles, namely, the space domain (LEO) and the terrestrial domain (BS), and adopt a distributed training and execution paradigm. Each agent maintains its own actor-critic networks without parameter sharing. Accordingly, role customization is realized in three aspects: (i) role-dependent hybrid action parameterization for partitioning, routing, and resource allocation; (ii) role-specific operational constraints embedded in feasible decisions; and (iii) a local-global reward trade-off for cross-role cooperation.

Each agent observes common local information, including the task demands and spatial distribution of users in its current coverage area, local queue length, and remaining computing resources. Moreover, LEO agent $n \in \mathcal{N}$ observes its orbital position and neighboring inter-satellite link (ISL) states, and receives lightweight summaries of the available resources of neighboring LEO satellites and the cloud center through periodic message exchange. In contrast, BS agent $m \in \mathcal{M}$ observes the positions of its connected LEO satellites and obtains summaries of the available resources of neighboring BSs, connected LEO satellites, and the cloud center. Given its observation, each agent selects a hybrid action from its role-specific action space. LEO agent n selects $a_n(t) = \{\boldsymbol{\alpha}_n(t), \boldsymbol{\beta}_n(t), \boldsymbol{w}_n(t)\} \in \mathcal{A}_n$, while BS agent m selects

* Corresponding author (email: cjjiang@tongji.edu.cn)

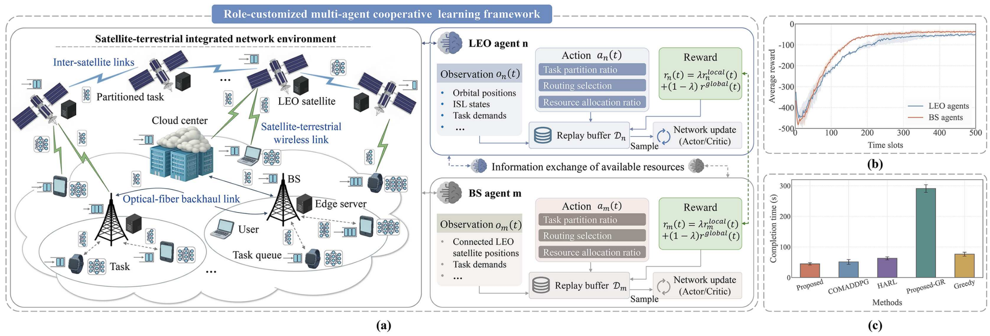

Figure 1 (Color online) (a) The overall architecture of the role-customized multi-agent cooperative learning framework for cross-domain task orchestration in STINs; (b) convergence performance; (c) task completion time comparison.

$a_m(t) = \{\boldsymbol{\alpha}_m(t), \boldsymbol{\beta}_m(t), \boldsymbol{w}_m(t)\} \in \mathcal{A}_m$. Here, $\boldsymbol{\alpha}_{n/m}(t)$, $\boldsymbol{\beta}_{n/m}(t)$, and $\boldsymbol{w}_{n/m}(t)$ denote task partitioning, next-hop routing, and computing resource allocation decisions for tasks served by the agent, respectively. Due to constrained onboard resources and dynamic satellite connectivity, the feasible offloading choices of LEO agents are further restricted by the remaining onboard resources, as well as the availability of ISLs and satellite-terrestrial links, making their actions more constrained than those of BS agents.

After executing the action, each agent receives an immediate reward that balances local autonomy and cross-domain efficiency, defined as

$$\begin{aligned} r_n(t) &= (1-\lambda) r_n^{\mathrm{local}}(t) + \lambda r^{\mathrm{global}}(t), \quad n \in \mathcal{N}, \\ r_m(t) &= (1-\lambda) r_m^{\mathrm{local}}(t) + \lambda r^{\mathrm{global}}(t), \quad m \in \mathcal{M}, \end{aligned} \tag{1}$$

where $r_n^{\mathrm{local}}(t)$ and $r_m^{\mathrm{local}}(t)$ are defined as the negative average completion times of tasks served by LEO agent n and BS agent m at slot t, respectively; $r^{\mathrm{global}}(t)$ is the negative system-wide average completion time over all tasks at slot t; and $\lambda \in [0,1]$ controls the trade-off between autonomy and global cooperation. The transition function P models the stochastic evolution of task queues, available computing resources, and time-varying communication channels, and $\gamma \in [0,1)$ is the discount factor.

Each agent learns a role-specific policy parameterized by neural networks and interacts with the environment and peers to optimize long-term performance under partial observability. In the fully distributed paradigm, the cooperative objective of the policy set $\boldsymbol{\pi} = \{\pi_n\}_{n\in\mathcal{N}} \cup \{\pi_m\}_{m\in\mathcal{M}}$ is to maximize the expected cumulative discounted reward averaged across all agents, given by

$$\max_{\boldsymbol{\pi}} J(\boldsymbol{\pi}) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \left(\frac{\sum_{n\in\mathcal{N}} r_n(t) + \sum_{m\in\mathcal{M}} r_m(t)}{|\mathcal{N}| + |\mathcal{M}|}\right)\right]. \tag{2}$$

Based on the Dec-POMDP model, we develop a role-customized MADRL mechanism for cross-domain task orchestration. LEO agent n and BS agent m update their role-specific actor-critic networks using the corresponding replay buffers $\mathcal{D}_n$ and $\mathcal{D}_m$, respectively. The critic evaluates the long-term utility of role-aware decisions, and the actor is optimized to produce higher-value feasible actions. Through iterative policy evaluation and improvement, local operational preferences are preserved while system-level efficiency is reinforced by the global reward. Combined with low-overhead cross-domain coordination via periodic information exchange, the proposed mechanism enables satellite-terrestrial agents to learn complementary behaviors under differentiated capabilities and constraints. The pseudo-code of the training procedure is provided in Appendix D.

Experiments. We simulate a network scenario with multiple users, BSs, LEO satellites, and a cloud center, where LEO satellites are modeled via Poliastro to capture realistic orbital dynamics. The proposed framework adopts the soft actor-critic (SAC) algorithm with entropy regularization, clipped double-Q learning, and adaptive temperature adjustment for cross-domain policy learning. All results are averaged over five independent runs. Figure 1(b) shows that both BS and LEO agents achieve stable convergence, while LEO agents exhibit slight fluctuations due to orbital motion, limited onboard computing resources, and dynamic visibility. Figure 1(c) compares the proposed framework with convex optimization task decomposition-based multi-agent deep deterministic policy gradient (COMADDPG), heterogeneous-agent reinforcement learning (HARL), global-reward variant of the proposed framework (Proposed-GR), and Greedy in terms of task completion time. COMADDPG, HARL, Proposed-GR, and Greedy denote a representative MADRL variant, a heterogeneous-agent reinforcement learning baseline, the global-reward variant of our framework, and a non-deep reinforcement learning (non-DRL) baseline, respectively. The proposed framework achieves the lowest completion time, reducing it by 12.26%, 28.70%, and 41.41% compared with COMADDPG, HARL, and Greedy, respectively, while also outperforming Proposed-GR. These results verify the effectiveness of the proposed framework in task orchestration efficiency and cross-domain adaptability, with further comparisons and ablation studies provided in Appendix E.

Conclusion. This study proposes a role-customized multi-agent cooperative learning framework for cross-domain task partitioning, offloading, and resource allocation in STINs. Each BS and LEO satellite is associated with a domain-specific agent that autonomously learns role-tailored orchestration policies under dynamic and partially observable environments. To coordinate heterogeneous satellite-terrestrial agents, a role-customized MADRL mechanism is developed to balance local autonomy and cross-domain coordination through limited information exchange and cooperative feedback. Experimental results demonstrate that the proposed framework improves cross-domain task orchestration efficiency and achieves lower task completion time than baseline methods.

Acknowledgements This work was supported by National Key Research and Development Program of China (Grant No. 2023YFB3002201) and Discipline Breakthrough Pioneer Program of the Ministry of Education of China (Grant No. JYB2025XDXM122).

Supporting information Appendixes A–E. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Xie B, Cui H, Ho I W H, et al. Computation offloading and resource allocation in LEO satellite-terrestrial integrated networks with system state delay. IEEE Trans Mobile Comput, 2024, 24: 1372–1385

2 Shi Y M, Zhu J Y, Jiang C X, et al. Satellite edge artificial intelligence with large models: architectures and technologies. Sci China Inf Sci, 2025, 68: 170302