TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 13/24 pp1463–1480
DOI: 10.26599/TST.2023.9010131
Volume 29, Number 5, October 2024

Data-Driven Collaborative Scheduling Method for Multi-Satellite Data-Transmission

Xiaoyu Chen, Weichao Gu, Guangming Dai*, Lining Xing, Tian Tian,
Weilai Luo, Shi Cheng, and Mengyun Zhou

Abstract: With continuous expansion of satellite applications, the requirements for satellite communication services, such as communication delay, transmission bandwidth, transmission power consumption, and communication coverage, are becoming higher. This paper first presents an overview of the current development status of Low Earth Orbit (LEO) satellite constellations, and then conducts a demand analysis for multi-satellite data transmission based on LEO satellite constellations. The problem is described, and the challenges and difficulties of the problem are analyzed accordingly. On this basis, a multi-satellite data-transmission mathematical model is then constructed. Combining classical heuristic allocating strategies on the features of the proposed model, with the reinforcement learning algorithm Deep Q-Network (DQN), a two-stage optimization framework based on heuristic and DQN is proposed. Finally, by taking into account the spatial and temporal distribution characteristics of satellite and facility resources, a multi-satellite scheduling instance dataset is generated. Experimental results validate the rationality and correctness of the DQN algorithm in solving the collaborative scheduling problem of multi-satellite data transmission.

Key words: relay satellite; scheduling; data transmission; Deep Q-Network (DQN); Genetic Algorithm (GA)

1 Introduction

The development of Low Earth Orbit (LEO) satellite constellations has a significant historical background. Notable projects in this field include the "Iridium" project[1] by Motorola in the United States, SpaceX's "Starlink" project, Amazon's "Project Kuiper"[2], Telesat's "Lightspeed" project[3] in Canada, the LeoSat constellation project by Luxembourg-based LeoSat company, as well as LEO satellite constellations proposed by space internet company OneWeb and rocket manufacturer Astra. These initiatives collectively aim to offer users worldwide access to high-speed and high-quality internet connectivity and data communication services. China has also devised plans for LEO satellite networking, namely the "Hongyan Constellation" consisting of 300 LEO satellites and a global data processing center, and the "Hongyun Project"[4] comprising 156 LEO satellites.

With the increasing number of satellites being launched into orbit, there is an inevitable rise in the need for control, operation, and maintenance tasks for these satellites. Effectively managing tasks, such as satellite networking, command uploading, data transmission, and task distribution among the

• Xiaoyu Chen, Weichao Gu, Guangming Dai, Tian Tian, Weilai Luo, and Mengyun Zhou are with School of Computer Science, China University of Geosciences, Wuhan 430074, China. E-mail: xiaoyu.chen@cug.edu.cn; Gu_WeiChao@outlook.com; cugdgm@126.com; tiantiandouble@hotmail.com; weilai_luo@cug.edu.cn; myzhou@cug.edu.cn.
• Lining Xing is with School of Electronic Engineering, Xidian University, Xi'an 710071, China. E-mail: lnxing@xidian.edu.cn.
• Shi Cheng is with School of Computer Science, Shaanxi Normal University, Taiyuan 710119, China. E-mail: cheng@snnu.edu.cn.
* To whom correspondence should be addressed.
Manuscript received: 2023-07-13; revised: 2023-10-11; accepted: 2023-10-25

numerous satellite, present a significant challenge in the field of satellite applications. This challenge includes addressing large-scale satellite scheduling and combinatorial optimization problems.

Precise algorithms, including branch and bound and dynamic programming, are commonly employed to address scheduling problems. Furthermore, heuristic-based approximate algorithms, such as Simulated Annealing (SA), Genetic Algorithms (GA), Differential Evolution (DE) algorithms, Particle Swarm Optimization (PSO) algorithms, and recently, Reinforcement Learning (RL) algorithms based on artificial intelligence techniques, have emerged as effective approaches. These methods aim to obtain optimal or near-optimal solutions for scheduling problems, leading to enhanced task completion efficiency, reduced task costs, and improved resource utilization optimization.

In various fields of scheduling problems, the scheduling optimization algorithms have been widely applied. For example, Zhang et al.[5] proposed a hybrid algorithm that integrates PSO, GA, and heuristic task interleaving algorithm to solve the task processing problem in digital array radar systems. Fu et al.[6] provided a literature review on distributed scheduling problems in intelligent manufacturing systems. Hu and Li[7] proposed an improved heuristic job scheduling method. Jiang et al.[8] addressed a mixed-integer linear programming model that has been proposed to solve the workshop scheduling problem. Yan et al.[9] proposed a short-term task scheduling strategy based on multi-provider Deep Q-Network (DQN) for frequent task scheduling problems in cloud network integrated environments. Chen et al.[10] addressed resource allocation and scheduling problems in mixed-service systems in 5G/B5G communication systems, creatively handling non-convex optimization problems with the innovative use of dueling DQN.

In the field of satellite scheduling and planning, many experts and scholars have made significant contributions. For the agile imaging satellite scheduling problem, Wang et al.[11] designed two sets of heuristics to solve the multi-satellite dynamic scheduling problem by targeting the dynamic characteristics such as task attribute changes, and also considered the task priority, as well as time window blocking and task overlap. Chen et al.[12] constructed a mixed-integer linear programming model and proposed a two-stage heuristic algorithm[13]. For the relay satellite data transmission scheduling problem, Huang et al.[14] proposed a satellite data transmission scheduling algorithm based on an improved ant colony system, to maximize the weighted success rate of scheduled tasks. Yan[15] proposed a hybrid particle swarm scheduling algorithm based on the rolling window to achieve dynamic scheduling of the relay satellite system based on pre-planning. Xiang et al.[16] proposed a parallel genetic algorithm with dual-threshold control to achieve global optimization of satellite data transmission. Fan[17] proposed an improved particle swarm algorithm combined with heuristic rules to address the integrated scheduling problem of satellite data transmission tasks and measurement and control tasks. Dong[18] conducted in-depth research on satellite network data transmission and satellite task scheduling, and proposed solutions to these problems. Du[19] researched satellite task scheduling engines, focusing on the generalization of satellite task scheduling modeling and solution methods. He proposed a generalized modeling method for satellite task scheduling, an adaptive parallel modal evolution algorithm for routine satellite task scheduling, and a distributed dynamic rolling optimization algorithm for emergency satellite task scheduling.

Reinforcement learning, as a popular machine learning method, has been introduced into various scheduling problem domains in recent years. Wang[20] addressed the online scheduling problem of satellites in a centralized structure and established a satellite online scheduling model based on Markov Decision Processes (MDP). He proposed a reinforcement-learning-based satellite online scheduling algorithm. Inspired by Recurrent Neural Networks (RNN) and attention mechanisms, Chen et al.[21] proposed an end-to-end framework based on Deep Reinforcement Learning (DRL). This model treats neural networks as complex heuristic methods and constructs them by observing reward signals and following feasible rules. Bao et al.[22] studied the online scheduling problem of satellite missions. They first established a model of MDP, and then applied the reinforcement-learning-based Asynchronous Advantage Actor-Critic (A3C) algorithm to assign arriving tasks to different satellites. Liu et al.[23] studied satellite scheduling problems using competitive learning strategies. They proposed a Q-network-based solution to solve the single satellite scheduling problem, and introduced a profit-based competition strategy to address the inherent scheduling conflicts of multi-satellite tasks. Ren et al.[24] proposed a Q-learning-based reinforcement learning algorithm to achieve fast response in emergency task scheduling, enabling real-time scheduling of tasks while maximizing scheduling stability. He et al.[25] verified the advantage of Q-network in fitting long-term returns in the scheduling problem of imaging satellites, and combined heuristic algorithms and an improved DQN for solving it.

The remainder of this article is organized as follows. Section 2 examines the task requirements for data transmission in future satellite application scenarios, specifically focusing on LEO satellite constellations. The key difficulties and challenges in scheduling planning for such tasks are also analyzed. A mathematical model for the multi-satellite data transmission problem is presented in Section 3. On this basis, Section 4 introduces a two-stage framework He-DQN algorithm, which combines GA, heuristic strategies, and DQN algorithms to address and resolve the task scheduling problem at hand. Additionally, Section 5 outlines the construction of specific simulation scenarios and provides corresponding experiment examples. The experimental results are analyzed to evaluate the effectiveness of the proposed algorithm and examine the factors that influence problem-solving for different examples. Finally, Section 6 concludes this article.

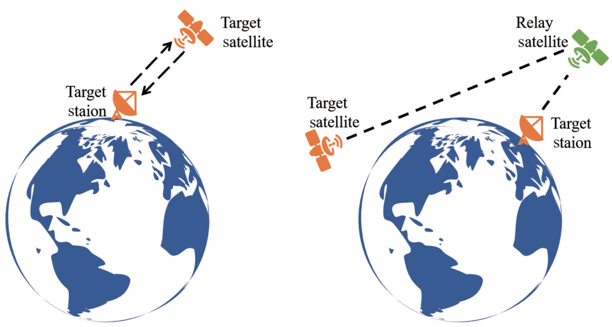
2 Problem Description and Analysis

Satellite services have become integral to various aspects of everyday life, encompassing earth observation, communication, navigation and positioning, internet access, scientific research, and military security, among many other domains[26]. In numerous application scenarios, such as communication[27], navigation[28], and remote sensing[29], the exchange of data between satellites and ground stations plays a vital role. This necessitates research on collaborative multi-satellite data transmission mission planning. By appropriately sequencing task execution, considering the temporal and spatial dependencies among tasks, and effectively allocating resources like energy, storage, and communication bandwidth, potential resource conflicts and wastage can be circumvented. Moreover, such planning can augment the overall completion rate of the entire data transmission process.

In our study on data-driven collaborative multi-satellite data transmission, we initially present an abstract description of the physical process involved in transmitting data between satellites and ground stations within this specific application scenario. Subsequently, we analyze the task requirements to identify the primary difficulties and challenges associated with this scheduling problem.

The data transmission scenario can be explained with simplicity as follows. During a specified time window when the antenna payloads of a ground station and a satellite are within communication range, data can be either downlinked from the satellite to the ground or uploaded from the ground to the satellite. For instance, satellite imaging data can be transmitted to a ground station, or control instructions can be uploaded from a ground station to a specific satellite, as illustrated in Fig. 1a. However, practical scenarios often involve more complex transmission situations. On one hand, in many application scenarios, there may be no direct communication time window between the ground station and the satellite. For example, achieving 100% coverage of an LEO spacecraft at a height of 300 km would require deploying over 100 stations on earth's surface, which is practically infeasible[30]. In such cases, the use of relay satellites becomes necessary for achieving the transmission goal, as shown in Fig. 1b. On the other hand, in a given application scenario, multiple ground stations and satellites may participate in the data transmission process, requiring considerations of task conflicts and time window resource conflicts among different ground stations and satellites.

In traditional data transmission scenarios, relay services utilizing high-orbit relay satellites are commonly employed. These high-orbit relay satellites are typically positioned in geostationary orbits approximately 36 000 km above the earth's surface.



(a) Direct transmission (b) Relay transmission

Fig. 1 Data transmission.

They have the capability to oversee and track spacecraft operating in medium-to-low orbits, facilitating real-time transmission of the acquired data back to ground stations[30].

Currently, high-orbit geostationary satellites are the predominant technology for data relay in space. However, in future applications, the drawbacks of high-orbit data relay satellites, such as high latency, signal attenuation, complexity, and power consumption, may become more significant. On the contrary, with the rapid deployment and progress of LEO satellite constellations, LEO satellites offer several advantages, including low latency, minimal signal attenuation, high flexibility, high reliability, and lower per-satellite costs. These advantages make LEO satellites highly promising for specific data transmission scenarios.

Using two future complex application scenarios as examples: LEO satellite navigation information augmentation[31] and high-intensity surveillance remote sensing data transmission in hot spot areas, we can anticipate the value demonstrated by LEO satellite constellations in the field of satellite data transmission services.

LEO satellite constellations have the advantages of high ground receiver signal strength and rapid geometric variation. They can complement medium to high-earth Global Navigation Satellite System (GNSS) constellations, providing significant advantages in enhancing the accuracy, integrity, continuity, and availability of GNSS systems[32]. Taking China's "Hongyan" LEO satellite constellation as an example (as shown in Fig. 2), it enables precise orbit determination and clock bias determination through the joint collaboration of medium to high-earth navigation satellites and low-earth communication satellites. This in turn enables dynamic sub-decimeter-level and static centimeter-level global precise single-point positioning[33].

In the future, the ability to achieve large-scale, continuous, and low-latency global monitoring is an important aspect of competition in the space domain (as shown in Fig. 3). The United States, through the deployment of a large number of low-cost, scalable, and small satellites in LEO, is establishing the next-generation space architecture regulatory layer to enable time-sensitive target monitoring missions, enhance space reconnaissance tactical support capabilities, and increase space resilience[34]. By utilizing LEO satellites

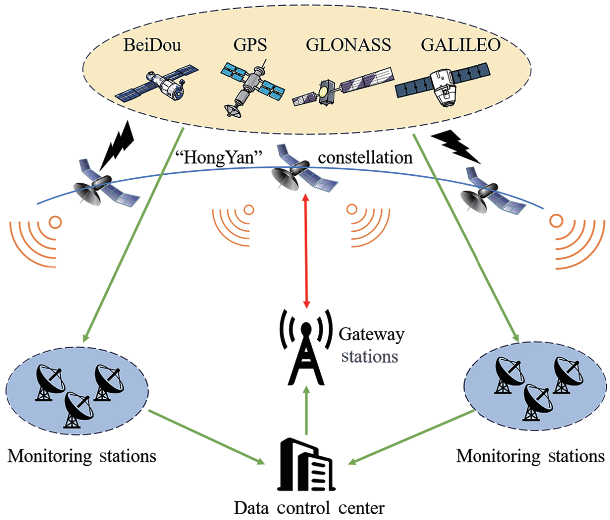


Fig. 2 "HongYan" constellation service[33].

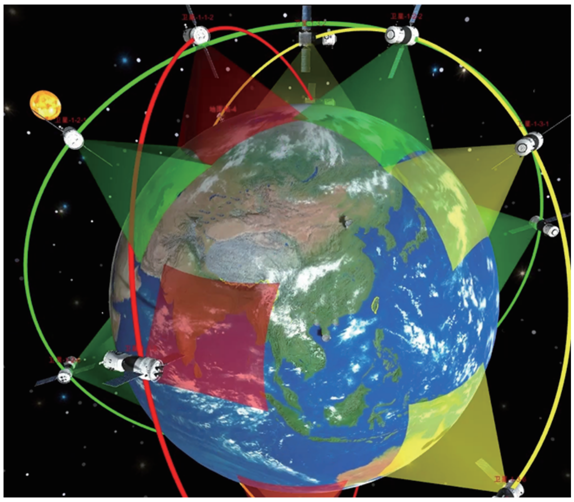


Fig. 3 Global surveillance (simulated by China Satellite Tool Kit (CSTK)).

for high-intensity remote sensing surveillance of hot spot areas, the demand for continuous monitoring with low latency can be effectively met. In this process, a large-scale LEO satellite constellation can ensure high-frequency satellite imaging observations of hot spot areas. Subsequently, by allocating ground station and relay satellite resources properly, the satellite-to-ground and inter-satellite connection relationships and time intervals can be determined, and satellite-ground and inter-satellite links can be used to receive surveillance data.

Through the above two complex application scenarios, it is evident that data relay transmission will become an important component of future LEO satellite applications. Therefore, research on data transmission tasks oriented towards LEO satellite constellations is highly necessary. For future large-scale LEO satellite constellations, efficient utilization of these relay resources will inevitably involve the combinatorial optimization problem of large-scale satellite scheduling in such a context.

As mentioned earlier, when multiple ground stations and satellites are involved in the data transmission process within an application scenario, it becomes crucial to consider the conflicts between tasks of different ground stations and different satellites, as well as conflicts regarding the utilization of time window resources. The following summarizes some key challenges and difficulties to consider in this problem context.

(1) Direct or relay transmission. Priority is given to direct transmission between target ground stations and target satellites. If there is no direct time window, and it is necessary to relay through a relay satellite, generally only one relay is considered.

(2) Conflict avoidance in resource selection. When selecting time window resources for a task, it is important to consider whether it will have an impact on other tasks in the same time period and potentially cause resource conflicts.

(3) Uniqueness in antenna data transmission. A single antenna can only perform one transmission task at a time.

(4) Constraints on data transmission volume within the time window. Consider excluding time window resources that cannot achieve the required data transmission volume, such as those with short duration or slow transmission rates.

(5) For each individual data transmission activity, there is a variety of options when it comes to selecting the objective function for scheduling and planning.

(6) Similarly, for an overall data transmission scenario, there are also multiple choices for the objective function.

3 Multi-Satellite Data-Transmission Scheduling Model

Based on potential mission requirements in future scenarios, this paper models and formally describes the problem of task scheduling for data transmission in LEO satellite constellations.

3.1 Definition of terms

- **Data transmission scenario.** In this context, it is assumed that there exists a data transmission scenario, where multiple satellites and ground stations need to perform data transmission activities during the mission period. This includes data uploading activities from several ground stations to the satellites and data downloading activities from several satellites to the ground stations.
- **Data transmission meta-task.** The data transmission activities conducted between each pair of satellites and ground stations are defined as meta-tasks in the data transmission mission. A data transmission mission scenario consists of multiple meta-tasks, where each meta-task specifies the target satellite and target ground station for a data transmission activity, along with the corresponding data transmission volume requirement and activity priority level.
- **Time window resource.** Data transmission activities between satellites and ground stations, as well as between satellites, must be conducted within the time windows when their respective antenna payloads can communicate. In the proposed model of this paper, it is assumed that each satellite and ground station is equipped with only one antenna for data transmission. The time window resources in the data transmission mission are divided into three parts: the inter-satellite communication time window resources between all relay satellites and all target satellites, the satellite-to-ground communication time window resources between all relay satellites and all target ground stations, and the direct communication time window resources between all target satellites and all target ground stations.

3.2 Notations

(1) Data transmission mission period $\left[S_{\text{beg}}, S_{\text{end}}\right]$. All data transmission activities are completed within the data transmission mission period.

(2) Data transmission meta-task M_i. It can be represented by a 6-tuple as follows:

$$M_i = \{\text{mID}_i, w_i, \text{ts}_i, g_i, \text{dir}_i, \text{sz}_i\},$$

where the meanings of these symbols are as follows:

- mID_i: meta-task ID;
- w_i: meta-task weight;
- ts_i: target satellite identification;
- g_i: target ground station identification;
- dir_i: transmission direction;
- sz_i: transmission data size.

(3) Meta-task set $M = \{M_1, \ldots, M_i, \ldots, M_{N_\text{m}}\}$. In a

data transmission scenario, there is multiple data transmission activities involving different ground stations and satellites. Each transmission activity is described as a meta-task M_i, and the total number of meta-tasks is N_{m}. The set of meta-tasks constitutes the entire data transmission mission requirement.

(4) Target satellite set $\mathrm{TS}=\{\mathrm{ts}_1, \ldots, \mathrm{ts}_i, \ldots, \mathrm{ts}_{N_{\mathrm{t}}}\}$. It contains all target satellites that are required for data transmission. ts_i identifies a particular target satellite and the total number of target satellites is N_{t}.

(5) Target ground station set $G=\left\{g_1, \ldots, g_i, \ldots, g_{N_{\mathrm{g}}}\right\}$. It contains all target ground stations that are required for data transmission. g_i identifies a particular target ground station and the total number of target ground stations is N_{g}.

(6) Relay satellite set $\mathrm{RS}=\{\mathrm{rs}_1, \ldots, \mathrm{rs}_i, \ldots, \mathrm{rs}_{N_{\mathrm{r}}}\}$. It contains all satellites that are used as relay resources for data transmission. rs_i identifies a particular relay satellite and the total number of relay satellites is N_{r}.

(7) Inter-satellite link resources. $\mathrm{TR}=\mathrm{TR}_{1,1}\cup\cdots\cup\mathrm{TR}_{i,j}\cup\cdots\cup\mathrm{TR}_{N_{\mathrm{t}},N_{\mathrm{r}}}$, where $i\in\{1, 2, \ldots, N_{\mathrm{t}}\}$, $j\in\{1, 2, \ldots, N_{\mathrm{r}}\}$, and the $\mathrm{TR}_{i,j}=\left\{\mathrm{tr}_{i,j}^1, \ldots, \mathrm{tr}_{i,j}^k \ldots, \mathrm{tr}_{i,j}^{\mathrm{trN}_{i,j}}\right\}$ indicates the set of communication time windows between target satellite i and relay satellite j, where $\mathrm{trN}_{i,j}$ represents the total number of communication time windows between them. $\mathrm{tr}_{i,j}^k$ can be represented by a 6-tuple as follows:

$$\mathrm{tr}_{i,j}^k=\left\{\mathrm{rID}_{i,j}^k, \mathrm{ts}_i, \mathrm{rs}_j, \mathrm{twBeg}_{i,j}^k, \mathrm{twEnd}_{i,j}^k, \mathrm{trRate}_{i,j}^k\right\},$$

where the meanings of these symbols are as follows:

- $\mathrm{rID}_{i,j}^k$: resource ID;
- ts_i: target satellite identification;
- rs_j: relay satellite identification;
- $\mathrm{twBeg}_{i,j}^k$: transmittable start time;
- $\mathrm{twEnd}_{i,j}^k$: transmittable end time;
- $\mathrm{trRate}_{i,j}^k$: inter-satellite link transmission rate.

(8) Satellite-ground station link resources.

$$\mathrm{SG}=\mathrm{RSG}\cup\mathrm{TSG},$$
$$\mathrm{RSG}=\mathrm{RSG}_{1,1}\cup\cdots\cup\mathrm{RSG}_{i,j}\cup\cdots\cup\mathrm{RSG}_{N_{\mathrm{r}},N_{\mathrm{g}}},$$
$$\mathrm{TSG}=\mathrm{TSG}_{1,1}\cup\cdots\cup\mathrm{TSG}_{i,j}\cup\cdots\cup\mathrm{TSG}_{N_{\mathrm{t}},N_{\mathrm{g}}}.$$

where $i\in\{1,2,\ldots,N_{\mathrm{t}}\}, j\in\left\{1,2,\ldots,N_{\mathrm{g}}\right\}$.

$$\mathrm{RSG}_{i,j}=\left\{\mathrm{rsg}_{i,j}^1, \ldots, \mathrm{rsg}_{i,j}^k, \ldots, \mathrm{rsg}_{i,j}^{\mathrm{rsgN}_{i,j}}\right\}$$

indicates the set of communication time windows between the relay satellite i and the target ground station j.

$$\mathrm{TSG}_{i,j}=\left\{\mathrm{tsg}_{i,j}^1, \ldots, \mathrm{tsg}_{i,j}^k, \ldots, \mathrm{tsg}_{i,j}^{\mathrm{tsgN}_{i,j}}\right\}$$

indicates the set of communication time windows between the target satellite i and the target ground station j. $\mathrm{rsgN}_{i,j}$ and $\mathrm{tsgN}_{i,j}$ indicate the total number of communication time windows between each other, respectively.

$\mathrm{rsg}_{i,j}^k$ and $\mathrm{tsg}_{i,j}^k$ can be represented by the following esimilar 6-tuples, respectively:

$$\mathrm{rsg}_{i,j}^k=\left\{\mathrm{rID}_{i,j}^k, \mathrm{rs}_i, g_j, \mathrm{twBeg}_{i,j}^k, \mathrm{twEnd}_{i,j}^k, \mathrm{rsgRate}_{i,j}^k\right\},$$

$$\mathrm{tsg}_{i,j}^k=\left\{\mathrm{rID}_{i,j}^k, \mathrm{ts}_i, g_j, \mathrm{twBeg}_{i,j}^k, \mathrm{twEnd}_{i,j}^k, \mathrm{tsgRate}_{i,j}^k\right\},$$

where $\mathrm{rsgRate}_{i,j}^k$ denotes relay-satellite to ground link rate and $\mathrm{tsgRate}_{i,j}^k$ denotes target-satellite to ground link rate.

(9) The whole scheduling scheme for this problem can be described as $\mathrm{Scheme}=\{\mathrm{sc}_1, \ldots, \mathrm{sc}_i, \ldots, \mathrm{sc}_{\mathrm{m}}\}$, sc_i indicates the scheduling scheme of meta-task M_i, and it can be described as

$$\mathrm{sc}_i=\begin{cases}\varnothing, & \text{Case (1)};\\ \mathrm{tsg}_{\mathrm{ts}_i,g_i}^k, & \text{Case (2)};\\ \mathrm{tr}_{\mathrm{ts}_i,\mathrm{rs}_i}^p+\mathrm{rsg}_{\mathrm{rs}_i,g_i}^q, & \text{Case (3)}.\end{cases}$$

where p, q, and k refer to a particular time window number between the corresponding two antennas, respectively. Case (1) means that the meta-task is not scheduled for execution, Case (2) means that the meta-task is scheduled to be executed via direct transmission, and Case (3) means that the meta-task is scheduled to be executed via a relay satellite.

3.3 Variables

In a data transmission scenario, there are multiple data transmission activities between multiple ground stations and satellites. There are time window selection conflicts between different meta-tasks. Therefore, the optimization variables in this problem model are the selection states of time window resources for each meta-task.

(1) There are two ways to complete a meta-task: direct transmission and relay transmission. The meta-task M_i specifies a target satellite ts_i and a target ground station g_i. When there exists available direct communication time window $\mathrm{TSG}_{\mathrm{ts}_i,g_i}$ between them, it is advisable to arrange as many meta-tasks as possible to be completed through this way, in order to save relay satellite resources and improve task

completion rate. In this case, the execution status of meta-task M_i on $\mathrm{tsg}^k_{\mathrm{ts}_i,\,g_i}$ is denoted as a boolean variable $\overleftrightarrow{x}^k_{\mathrm{ts}_i,\,g_i}$, and $\overleftrightarrow{x}^k_{\mathrm{ts}_i,\,g_i}=1$ represents that $\mathrm{tsg}^k_{\mathrm{ts}_i,\,g_i}$ will be the scheduling scheme for this meta-task.

(2) The meta-task M_i specifies a target satellite ts_i and target ground station g_i, if there are no available direct communication time windows or all available direct communication time windows are deprecated due to conflicts, it is necessary to arrange relay satellite resources to assist in completing the transmission activities. In this case, the problem can be described as a multi-phase and multi-decision problem. As shown in Fig. 4a, in the first phase, it is necessary to first select a relay satellite to execute relay transmission; then, based on the intersection and union characteristics of the time windows, a suitable time window resource needs to be chosen from all time windows between the target ground station and the selected relay satellite; finally, in the selection of time windows between the selected relay satellite and the target satellite, it is necessary to track the state of the preceding link's selection. After the completion of the previous stage, the time window selection for the third stage is conducted to complete the relay transmission (as shown in Fig. 4b).

In this case, the execution status of meta-task M_i on $\mathrm{tr}^p_{\mathrm{ts}_i,\,\mathrm{rs}_i}$ is denoted as a boolean variable $\overrightarrow{x}^p_{\mathrm{ts}_i,\,\mathrm{rs}_i}$, and the execution status on $\mathrm{rsg}^q_{\mathrm{rs}_i,\,g_i}$ is denoted as a boolean variable $\underrightarrow{x}^q_{\mathrm{rs}_i,\,g_i}$.

Based on the definition above, if there exists p and q, such that $\overrightarrow{x}^p_{\mathrm{ts}_i,\,\mathrm{rs}_i}=1 \wedge \underrightarrow{x}^q_{\mathrm{rs}_i,\,g_i}=1$, it represents that $\left(\mathrm{tr}^p_{\mathrm{ts}_i,\,\mathrm{rs}_i},\, \mathrm{rsg}^q_{\mathrm{rs}_i,\,g_i}\right)$ will be the scheduling scheme for this meta-task.

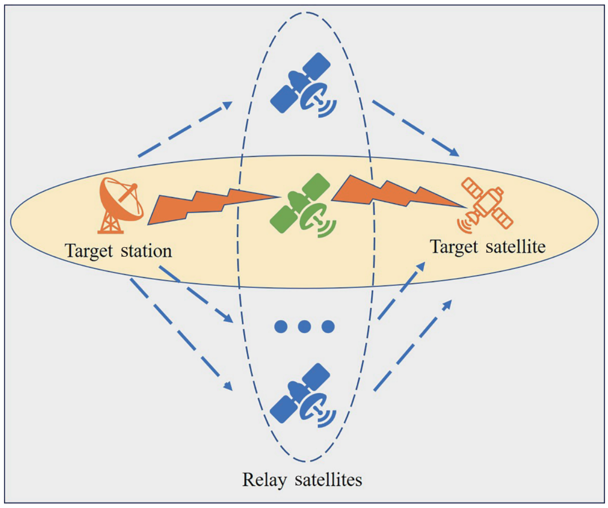
3.4 Objectives

The direct transmission scheduling status of the meta-task M_i is denoted as a boolean variable $\overleftrightarrow{E_i}$, and the relay transmission scheduling status is denoted as a boolean variable $\underrightarrow{\overrightarrow{E_i}}$,

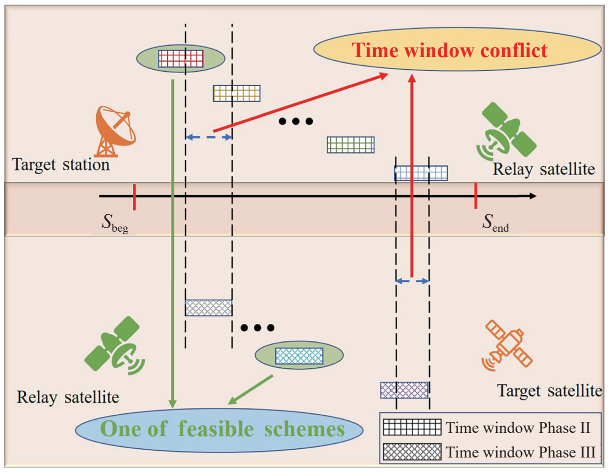
$$\overleftrightarrow{E_i}=\begin{cases}1, & \exists\, \overleftrightarrow{x}^k_{\mathrm{ts}_i,\,g_i}=1;\\ 0, & \forall\, \overleftrightarrow{x}^k_{\mathrm{ts}_i,\,g_i}=0\end{cases} \tag{1}$$

$$\underrightarrow{\overrightarrow{E_i}}=\begin{cases}1, & \exists\left(\overrightarrow{x}^p_{ts_i,\,rs_i}=1 \wedge \underrightarrow{x}^q_{rs_i,\,g_i}=1\right);\\ 0, & \neg\left(\exists\left(\overrightarrow{x}^p_{ts_i,\,rs_i}=1 \wedge \underrightarrow{x}^q_{rs_i,\,g_i}=1\right)\right)\end{cases} \tag{2}$$

From these, we can obtain the schedulable status E_i of the meta-task,



(a) Relay transmission Phase I



(b) Relay transmission Phases II&III

Fig. 4 Relay transmission workflows.

$$E_i=\begin{cases}1, & \overleftrightarrow{E_i}=1 \vee \underrightarrow{\overrightarrow{E_i}}=1;\\ 0, & \overleftrightarrow{E_i}=0 \wedge \underrightarrow{\overrightarrow{E_i}}=0\end{cases} \tag{3}$$

Based on the status of E_i, we have optimization objectives as follows:

(1) Maximize the total number of completed meta-tasks,

$$\max \sum_{M_i\in M} E_i.$$

(2) Maximize the total weights of executing meta-tasks,

$$\max \sum_{M_i\in M} w_i\cdot E_i.$$

3.5 Constraints

(1) Satellite-to-ground communication constraint.

Within the constraints of ground station properties, set the minimum antenna communication elevation angle to 30 degrees, to impose restrictions on the time windows for satellite-to-ground communication.

(2) Data transmission task integrity constraint. The transmission data volume of each meta-task needs to be transmitted all at once within a time window, without considering multiple transmissions in batches.

(3) Time window resource selection constraint. To avoid data transmission interruptions that may occur when two or more tasks are executed consecutively within a communication time window, the optimization problem is modeled as a 0/1 integer programming model. This means that the time window resources selected for meta-tasks are indivisible, and within the duration of a time window, they cannot be used to execute other meta-tasks.

(4) Exclusivity and availability constraints of bilateral devices during data transmission. In the planning process, a relay satellite may serve multiple meta-tasks, but when data transmission occurs between satellite and ground or between satellites, one antenna can only execute one transmission activity at a time. As shown in Fig. 5, in this case, conflicting tasks should be deprecated and re-scheduled for them.

(5) Antenna switching time constraint. For the same ground station or the same satellite, when switching the data transmission equipment, the antenna angle needs to be adjusted. Therefore, there exists an antenna switching duration constraint when switching data transmission equipment between different meta-tasks (as shown in Fig. 5).

(6) Data volume constraint for each time window transmission.

$$\mathrm{Rate}_{i,j}^{k} \times \left(\mathrm{twEnd}_{i,j}^{k} - \mathrm{twBeg}_{i,j}^{k}\right) \geqslant \mathrm{sz}_i \tag{4}$$

(7) Relay satellite storage capacity constraint. After receiving the transmission data of a meta-task, the relay transmission of the second stage must be performed before executing the next meta-task.

4 He-DQN: A Two-Stage Optimization Framework Based on Heuristic and DQN

Based on the multi-stage and multi-decision characteristics of the data transmission task, we will solve the problem through a two-stage scheduling process. In the first stage, we need to schedule and plan the meta-tasks that can be directly transmitted. Due to the problem's 0/1 integer programming nature and the relatively small decision space of the direct transmission process, we directly utilize a genetic algorithm to plan the meta-tasks that can be completed through direct transmission. The remaining meta-tasks are considered as tasks that require relay transmission.

In the second stage, to schedule the meta-tasks that require relay transmission, we divide the scheduling process into two phases. In the first phase, we propose a heuristic strategy that considers the number of relay satellite time windows, total communication time, and load balancing. This strategy is used to assign relay satellites for conducting relay transmission for each meta-task and determines the communication links for relay transmission. In the second phase, based on the determined communication links from the first phase and all feasible time windows, we use the DQN algorithm to arrange the specific time window for the communication links.

4.1 Direct transmission meta-tasks

Based on the analysis of the problem model, it is known that there exists a time window set between the target objects specified by each meta-task. However, due to the presence of transmission conflicts among different meta-tasks, for example, between meta-task $M_i = \{\mathrm{mID}_i, w_i, \mathrm{ts}_i, g_i, \mathrm{dir}_i, \mathrm{sz}_i\}$ and $M_j = \{\mathrm{mID}_j, w_j, \mathrm{ts}_j, g_j, \mathrm{dir}_j, \mathrm{sz}_j\}$, if $\mathrm{ts}_i = \mathrm{ts}_j$ or $g_i = g_j$, then the time

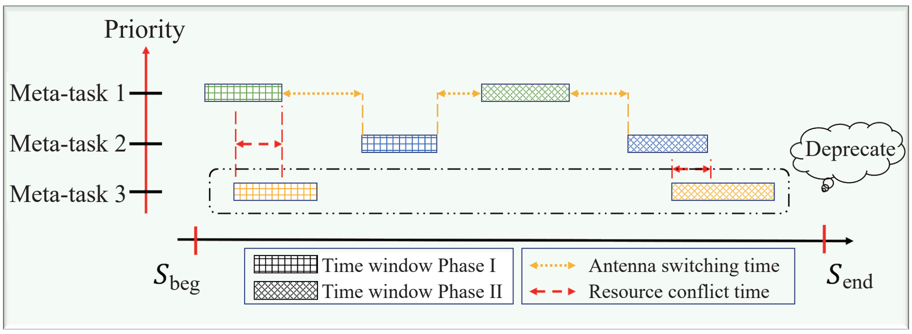


Fig. 5 Conflict and switching time constraints.

windows $\left[\mathrm{twBeg}^{p}_{\mathrm{ts}_i, g_i}, \mathrm{twEnd}^{p}_{\mathrm{ts}_i, g_i}\right]$ and $\left[\mathrm{twBeg}^{q}_{\mathrm{ts}_j, g_j}, \mathrm{twEnd}^{q}_{\mathrm{ts}_j, g_j}\right]$ in respective communication resources scheduled for them cannot have intersections.

To achieve the objective of "maximizing the total numbers of completed tasks" or "maximizing the total weights of task execution", while satisfying this constraint and other relevant constraints, the scheduling process for this stage can be achieved using a genetic algorithm[35].

The direct communication resource set $\mathrm{TSG}_{i,j} = \left\{\mathrm{tsg}^{1}_{i,j}, \ldots, \mathrm{tsg}^{k}_{i,j}, \ldots, \mathrm{tsg}^{\mathrm{tsgN}_{i,j}}_{i,j}\right\}$ between the target satellite i and the target ground station j specified in each data transmission meta-task can be obtained from the resource set. Additionally, the number of available resources for each meta-task, $\mathrm{tsgN}_{i,j}$ is also known. Therefore, each gene position in an individual can be encoded as a selection index of the communication resource between the target satellite and target ground station for a specific meta-task. Each individual in the population corresponds to a feasible scheduling solution.

The main steps to solve this problem using genetic algorithm are described as follows:

(1) Initialization. First, by iterating through all the resources in the resource set, we obtain the direct communication resources for each meta-task. Based on this, randomly select communication resources for gene encoding and generating individuals, and form the initial population.

(2) Conflict elimination. Due to the existence of constraints, when encoding individuals in the population, there may be conflicts (time window overlaps) in the communication resources planned for each gene position of the individual. Therefore, in computing the fitness of individuals in the population, conflict elimination operations are required. Depending on different objective functions, different elimination strategies are used to eliminate conflicting meta-tasks one by one until the individual satisfies all the constraints and becomes a feasible solution.

(3) Mutation operator. The mutation operation increases the diversity of the population. With a certain probability defined as the mutation probability, the mutation operator alters the resource selection index of a gene position in an individual, which helps to improve the local search ability and obtain better feasible solutions.

(4) Crossover operator. The crossover operation increases the diversity of the population by exchanging a certain segment of genes between two individuals in the population. Similarly, with a certain probability defined as the crossover probability, the crossover operator selects individuals and gene positions to perform the exchange operation, thereby enhancing the global search ability of the problem.

(5) Fitness function. The fitness function is used to evaluate the fitness value of individuals in the population, and higher fitness values are retained during subsequent selection. Based on conflict elimination, feasible scheduling solutions for the problem can be obtained, and accordingly, the fitness value of each individual can be calculated based on different objective functions.

(6) Selection operator. After generating new individuals through crossover and mutation operations, in order to maintain a stable population size in each iteration, we need to select and eliminate individuals in the population. A roulette wheel selection strategy is used, where individuals are selected to form the new population based on their fitness values and the cumulative probability of the total fitness.

According to above-mentioned steps, we repeat a series of evolutionary iterations. After each population iteration, we record the individual with the highest fitness value. Finally, when the specified number of evolutionary generations is reached, the optimal feasible solution is obtained. With the optimal feasible solution, the direct transmission scheduling status of each meta-task can be obtained. For meta-tasks that cannot be completed through direct transmission, a heuristic strategy combined with the DQN algorithm is proposed for relay transmission scheduling process.

4.2 Relay transmission meta-tasks

Through the first stage of the genetic algorithm, we can obtain the scheduling results of meta-tasks for direct transmission and the set of meta-tasks waiting for relay transmission. The relay transmission scheduling problem is a complex optimization problem with multiple stages and decisions.

In a data transmission scenario with N_{r} relay satellites, for each meta-task (assuming the target satellite is i and the target ground station is j), there are $\mathrm{tr}_{N_{i,l}}$ communication time windows between target satellite i and a particular relay satellite l, and $\mathrm{rsg}_{N_{l,j}}$ communication time windows between the particular relay satellite l and target ground station j, thus the search space for a meta-task comes to $\sum_{k=1}^{N_r} \mathrm{tr}_{N_{i,l}} \times \mathrm{rsg}_{N_{l,j}}$, that requires relay transmission is enormous and complex.

In this regard, we propose using a heuristic strategy in the first phase to select a relay satellite for each meta-task. Based on the selection of relay satellites, we can construct "target satellite − target relay satellite − target ground station" relay transmission link and obtain all Feasible Time Window Arrangements (FTWAs) for this link. In the second phase, based on the first phase results, we need to determine the specific time window arrangements for each communication link. In this phase, the DQN algorithm is used to train and learn the decision planning, guiding the agent to select the optimal decision action given a certain state, thus achieving scheduling for the continuous state space.

4.2.1 Heuristic strategy

For a certain meta-task, when the target satellite i and target ground station j are clear, the impact of selecting different relay satellites l to complete the task can be considered from the following two aspects:

- The number of time windows $\mathrm{tr}_{N_{i,l}}$ and total communication duration in target satellite i and relay satellite l communication resources $\mathrm{TR}_{i,l}$.
- The number of time windows $\mathrm{rsg}_{N_{l,j}}$ and total communication duration in relay satellite l to target ground station j communication resources $\mathrm{RSG}_{l,j}$.

Clearly, for a meta-task, if a certain relay satellite has more communication time windows for both the target satellite i and the target ground station j, then this relay satellite l should be given priority for selection. If multiple relay satellites have the same number of communication time windows for the same meta-task, then the one with a shortest total communication duration should be selected for relay transmission. This helps reduce the occupation of relay satellite resources, release satellite resources, and minimize satellite energy consumption. It also allows the selected relay satellite to have more available time to execute other meta-tasks.

On the other hand, if we aim to complete as many meta-tasks as possible, we need to utilize the available relay satellite resources to the maximum extent. However, the aforementioned heuristic strategy may result in one relay satellite servicing multiple meta-tasks (as shown in Fig. 6). This can lead to excessive conflicts between transmission task links, exceeding the service capacity limit of the relay satellite, and

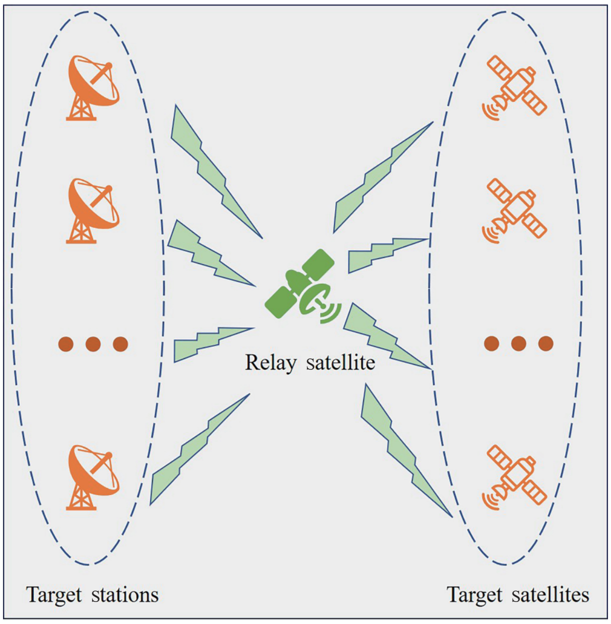


Fig. 6 Relay satellite service.

affecting the completion of meta-tasks in the entire data transmission scenario. Therefore, achieving a balanced distribution of relay satellite resources and ensuring a certain load balance when executing meta-tasks for each relay satellite is also an important consideration.

Based on the comprehensive considerations mentioned above, in the first phase of relay transmission scheduling, we propose a heuristic strategy that takes into account the number of time windows, total communication duration, and load balance for selecting relay satellites for each relay transmission meta-task. The strategy is described as follows.

Firstly, we calculate the number of FTWAs for each relay satellites in each meta-task. For the relay transmission link "target satellite − relay satellite − target ground station", just assume there are p time windows in the first phase and q time windows in the second phase. Theoretically, there are $p \times q$ possible combinations of time windows. However, considering constraints, such as time window availability and time window sequencing, we need to obtain all FTWAs that satisfy these constraints from the $p \times q$ combinations of time windows for each link. By performing this process for all relay transmission meta-tasks, we can determine the number of FTWAs in the communication links through each relay satellite for each meta-task.

Based on the number of FTWAs, we prioritize assigning the relay satellite with the maximum number to each meta-task. In cases where multiple meta-tasks are assigned to the same relay satellite, if there are other relay satellites that have not been assigned any meta-task, we should consider load balancing for the satellite. This can be done by resolving conflicts based on the total communication duration, with the relay satellite given priority to execute the meta-task with a smaller total communication duration.

For the remaining meta-tasks that generate conflicts, the relay satellite with the fewer FTWAs should be selected for execution. The process of relay satellite selection and conflict elimination is repeated until all meta-tasks are scheduled to be executed by a particular relay satellite.

Through this heuristic strategy, we can maximize the utilization of relay satellite resources and allocate a relay satellite with the highest redundancy in feasible link selections to execute each meta-task. This effectively reduces contention conflicts among meta-tasks for relay satellite resources, thereby improving the success rate of subsequent phase of meta-task scheduling.

Additionally, with the determination of relay satellites, we can obtain the clear FTWAs for each communication link. These FTWAs serve as the foundation for subsequent scheduling processes.

4.2.2 DQN

DQN is a deep reinforcement learning algorithm based on Q-learning. It leverages the powerful approximation and generalization capabilities of deep neural networks to effectively handle continuous state spaces. During the training process, the algorithm continuously adjusts the weights of the neural network to improve the accuracy of value function estimation. This guides the agent to learn to choose the optimal actions given specific states, enabling it to make decisions and plans for tasks in continuous state spaces.

Through the scheduling arrangement in the first stage, we not only determine the selection of relay satellites for each meta-task, but also obtain all FTWAs for the communication links. This provides the foundational conditions for the subsequent time window planning stage.

To plan the communication time window selection for each meta-task, we utilize a multi-layer feedforward neural network based DQN architecture. The design of this network architecture aims to extract useful features from the input states through successive layers of non-linear transformations and parameter learning. It generates corresponding outputs, which represent the time window scheduling strategies for each meta-task's respective transmission link.

In the second stage, we abstract the task planning problem into an MDP. The MDP decision process can be described using the following formula:

$$V_\pi(S_i, a) = R(S_i, a_i) + \gamma \sum_{s_i \in S} \Big(P(S_i, a, S_{i+1}) \times \max V_\pi(S_{i+1}, a_{i+1}) \Big) \tag{5}$$

where $V_\pi(S_i, a_i)$ describes the overall expected reward in a reinforcement learning environment when an agent chooses an action a_i according to the policy π in state S_i. The expected reward consists of two parts: (1) function $R(S_i, a_i)$ denotes the immediate reward directly obtained when performing action a_i in state S_i. (2) the rest part accounts for the future rewards, which is achieved by calculating the weighted average expected reward for transferring from the current state S_i to all possible next states S_{i+1}. In which, the function $P(S_i, a_i, S_{i+1})$ represents the probability transitioning from state S_i to state S_{i+1}. For each possible next state, we calculate the value of taking the optimal action a_{i+1} and have the $\max V_\pi(S_{i+1}, a_{i+1})$. The parameter γ is a discount factor, introduced to prevent the occurrence of state loops and to account for the uncertainty in future predictions. It discounts the consideration of future rewards, with a value closer to 1 indicating a longer-term perspective.

At the beginning, the system is in the initial state S_0. The agent takes action A_0 based on S_0, and interacts with the environment, thus receiving a reward R_1 for executing action A_0 and transitioning to the next state S_1. The agent continuously outputs actions A_i based on state S_i until reaching a terminal state. Each interaction between the agent and the environment is referred to as a time step.

- **State.** The state space S is a collection of states, denoted as $S = \{S_0, S_1, \ldots, S_i, \ldots\}$, where each state consists of attributes related to tasks and resources, describing the situation at a particular time step during time window selection. The details of the states are as follows:

$$S_i = \Big\{ x_i^t = \Big(d_i^{t,j}, n_i^t, \mathrm{sz}_i^t, w_i^t, \mathrm{twBeg}_i^{t,j}, \mathrm{twEnd}_i^{t,j}, g_i^t \Big), \; t = 0, 1, \ldots, N \Big\} \tag{6}$$

For a meta-task t at the specified resource j at time step i, the variables are as follows: d_i^t is the set of remaining communication time windows; n_i^t is the number of remaining time windows in the set; sz_i^t is the transmission data volume; w_i^t is the execution reward; $\mathrm{twBeg}_i^{t,j}$ and $\mathrm{twEnd}_i^{t,j}$ are the starting and ending times of the feasible time window, respectively; and g_i^t is the label indicating whether meta-task t has been selected before time step i.

In addition, if there are no remaining time windows for all meta-tasks, then $d_i^{t,j} = \varnothing, \forall t \in \{1, 2, \ldots, n\}$, it is defined as the terminal state.

- **Action.** Actions refer to the decisions made for meta-task t at time step i. Since in the aforementioned first phase, we have already assigned a resource for each task, the model's actions can be defined as combinations of all feasible time windows $A_i = \{a_0, a_1, \ldots, a_s\}$. In each step, the agent chooses an action through "exploration" or "exploitation". If it is "exploration", the agent randomly selects an action. If it is "exploitation", the agent uses the current time step i's state S_i as input to the neural network and selects the action with the highest Q-value.

- **Reward.** In the scheduling process, it is essential to arrange data task transmissions that satisfy the constraint conditions in order to maximize the benefits. Therefore, based on the optimization variables we designed earlier, to achieve the maximum profit, the reward function is designed to be the incremental total profit after taking the action, $\mathrm{reward} = \{R_1, R_2, \ldots, R_i, \ldots | R_i = f(x_i) - f(x_{i-1})\}$, in which $f(x_i)$ represents the total reward at time step i.

- **Value functions.** Based on the definitions of state, action, and reward mentioned above, we can introduce the value function. The value function measures the desirability of each state, helping the agent choose actions with the highest value to optimize the task planning process. According to Bellman equation, the value function is defined as $Q^*(s, a) = \mathrm{reward} + \gamma \times Q(s, a)$.

During the training phase, two separate neural networks are used: the current network, which calculates the Q-values for the current state, and the target network, which calculates the target Q-values. In the training process, the state S_i is used as the input to the neural network, and the network computes the Q-values for each action. These Q-values represent the expected returns for choosing each action in the current state. Based on the computed Q-values, the agent selects an action A_i according to a certain strategy and interacts with the environment, receiving a reward and obtaining information about the next state S_{i+1}.

Next, $\mathrm{loss} = \mathrm{TDerror}^2$ is used to update the parameters of the current network. The goal is to minimize the loss, allowing the current network to approximate the estimated values of the target network. This process enables the current network to make more accurate predictions of action-value functions and enhance the decision-making ability of the agent.

In conclusion, by constructing and training the neural network, we can optimize the network's parameters and minimize the error between the network's output and the expected output. This enables the intelligent agent to gradually learn and optimize task scheduling strategies, allowing it to generate accurate and efficient scheduling decisions in the problem of selecting feasible time window arrangements for communication links.

5 Instance and Experiment

5.1 Generate instance

In order to analyze the solving capability and effectiveness of the proposed algorithm in this problem, we generate corresponding test instances based on the richness of resources and the level of conflicts among meta-tasks. Each test instance includes a set of data transmission meta-tasks and various sets of satellite-to-ground and inter-satellite time window resources.

Firstly, we set up a data transmission scenario with a period of one day. Taking into account the characteristics of ground station selection, we designate ten target ground (Grd) stations located uniformly within China for data transmission, as shown in Table 1.

As the Target Satellite (TS) constellation for data transmission, we construct a Walker constellation consisting of 4 orbital planes, with each plane having 5 satellites. The phase factor is set to 1, resulting in a total of 20 satellites. The attributes of these target satellites are shown in Table 2.

As the Relay Satellite (RS) constellation for data transmission, we construct a Walker constellation consisting of 10 orbital planes, with each plane having 10 satellites. The phase factor is set to 1, resulting in a total of 100 satellites. The attributes of these relay satellites are shown in Table 2.

Table 1 Ground stations information.

Ground station tag	Longitude & Latitude
GS-BJ	(116.412 53°, 39.909 60°)
GS-KM	(102.846 42°, 24.890 17°)
GS-KS	(75.994 55°, 39.497 53°)
GS-SY	(109.518 10°, 18.254 54°)
GS-CD	(104.081 06°, 30.578 24°)
GS-HEB	(126.549 33°, 45.812 15°)
GS-LS	(91.123 16°, 29.656 04°)
GS-LZ	(103.846 02°, 36.065 30°)
GS-SH	(121.483 21°, 31.239 24°)
GS-WH	(114.311 31°, 30.598 06°)

Through all combinations of the 10 target ground stations and the 20 target satellites, a maximum of 200 data transmission meta-tasks can be set in this scenario. We assume that all data transmission tasks involve data downlink from the target satellites to the target ground stations. The execution weight of the meta-tasks are set as random integers ranging from [10, 100], and the data transmission demanding data sizes are set as random integers ranging from [2000, 3000], with units in Mb (megabits). All the inter-satellite and satellite-to-ground communication time windows can be obtained through calculations. In each communication resource, the communication transmission rate is set as a random integer ranging from [10, 20], with units in Mbps (megabits per second).

To demonstrate the influence of meta-task quantity and resource abundance on problem-solving, we create 6 instances with different meta-task scales and resource richness using the aforementioned data transmission scenario. The statistical information for all scenarios is presented in Table 3. In which |TS|, |Grd|, and |RS| are respectively denoted as the number of the target satellite, the number of the target ground station, and the number of the relay satellite. To better compare and analyze the impact of resource abundance in the relay satellite constellation, for two scenarios with the same meta-task quantity, their meta-task sets $M=\{M_1, \ldots, M_i, \ldots, M_m\}$ and the Target Satellite-to-Ground station communication resources $\mathrm{TSG}_{i,j}=\left\{\mathrm{tsg}_{i,j}^{1}, \ldots, \mathrm{tsg}_{i,j}^{k}, \ldots, \mathrm{tsg}_{i,j}^{\mathrm{tsgN}_{i,j}}\right\}$ are entirely identical, with only variations in the number of relay satellites.

5.2 Results and discussion

5.2.1 Stage-1: Direct transmission scheduling

During the first stage of direct transmission meta-tasks scheduling using genetic algorithm, our objective function is set to maximize the total weights of direct transmission meta-tasks. We use a population size of 50 to perform 100 generations of iterations, and set the crossover rate to 0.6 and the mutation rate to 0.4. After the first-stage scheduling process in each instance, the obtained optimal results are listed in Table 4.

From the results, we can observe that through the scheduling process, a considerable portion of meta-tasks in each test instance can be completed through the direct communication time windows between target satellites and ground stations (labeled as set M_{exe}, and $|M_{\mathrm{exe}}|$ refers to the number of set M_{exe}), the other meta-tasks (labeled as set M_{rmn}, and $|M_{\mathrm{rmn}}|$ refers to the number of set M_{rmn}) will remain there to be dealt with. Additionally, it is evident that as the number of meta-tasks increases, i.e., when the same number of target ground stations need to receive data from more target satellites, the proportion of meta-tasks that can be directly transmitted decreases. This is because for a given target ground station, the more target satellites are involved in data transmission, the higher the likelihood of time window conflicts in the meta-task scheduling results. In which $|M|$ is denoted as the number of meta-tasks.

5.2.2 Stage-2: Relay transmission scheduling

Through Stage-1 processing, we have differentiated the transmission states of all meta-tasks in the scenarios, dividing them into sets of meta-tasks (M_{exe}) that can be directly transmitted and those (M_{rmn}) that require relay transmission. For the meta-tasks that can be directly transmitted, we obtained the scheduling results directly ($\sum w_i^{\mathrm{exe}}$). For the remaining meta-tasks (M_{rmn}) that need relay transmission, we first used the proposed heuristic strategy to select relay satellites. The resulting relay satellite scheduling is shown in Fig. 7, where the x-axis represents the number of meta-tasks that a single relay satellite needs to serve, and the y-axis represents the number of relay satellites in corresponding states.

Table 2 Satellite properties.

Type	Orbital category	Number of recursive laps	Orbital period (s)	Inclination (deg)	Orbital height (km)
Target satellite	Sun-synchronous	14	6164.46	98.9873	888.322
Relay satellite	Recursive	15	5684.23	60.0000	505.852

Table 3 Instances overview.

Number	Instance tag	|TS|	|Grd|	|RS|	$\sum w_i$
1	M50-R10	5	10	10	3105
2	M50-R20	5	10	20	3105
3	M100-R30	10	10	30	5967
4	M100-R50	10	10	50	5967
5	M200-R60	20	10	60	12 120
6	M200-R100	20	10	100	12 120

Table 4 Stage-1 optimal results.

Number	$\|M\|$	$\sum w_i$	$\|M_{\text{exe}}\|$	$\sum w_i^{\text{exe}}$	$\|M_{\text{rmn}}\|$
1&2	50	3105	19	1520	31
3&4	100	5967	29	2304	71
5&6	200	12 120	43	3446	157

From Fig. 7, we can observe that, under the condition of the same number of meta-tasks, the more abundant the relay satellite resources, the fewer meta-tasks a single satellite needs to serve, with more relay satellites serving only 1 and 2 meta-tasks. However, when the relay satellite resources decrease, more and more relay satellites need to serve 3 and 4 meta-tasks. On one hand, this increases the communication burden on the satellites, and on the other hand, it adds difficulty to the subsequent feasible time window scheduling.

With the determination of relay satellites for each meta-task, we can establish the satellite-ground relay transmission links as "target satellite – relay satellite – target ground station" and obtain all FTWAs for these links by satisfying the constraints. The number of FTWAs obtained for each meta-task in 6 test instances is shown in Fig. 8.

From the results, it can be observed that after determining the transmission links, each link has multiple FTWAs. As the number of relay satellite resources increases, each meta-task can obtain relay satellites with higher scheduling redundancy, resulting in an increase in the number of FTWAs for the corresponding transmission links.

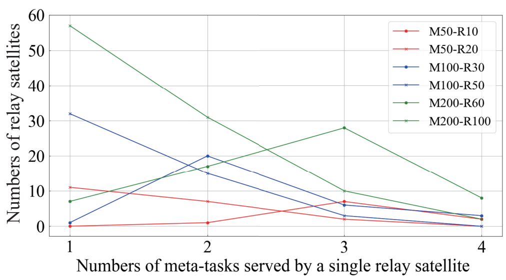
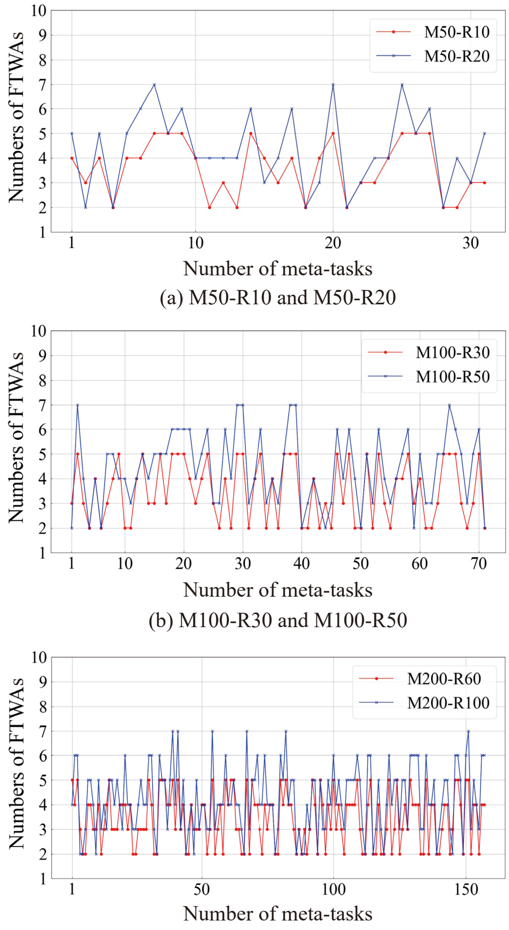


Fig. 7 Relay satellite loading status.



(a) M50-R10 and M50-R20

(b) M100-R30 and M100-R50

(c) M200-R60 and M200-R100

Fig. 8 Quantities of FTWAs in each meta-task.

The next step is to apply DQN for the subsequent FWTA scheduling. In order to demonstrate the impact of different resource abundances on the overall scheduling revenue, we conducted separate training using Instance 5: M200-R60 and Instance 6: M200-R100. From the 157 meta-tasks that require relay transmission in each instance, we randomly select the same 100 meta-tasks to form 1000 test samples for training. Each test sample has a total weights ranging from [5800, 6200]. The results are shown in Fig. 9.

From Fig. 8, we can observe that the total weight of scheduling meta-tasks sharply increases at the beginning of the training, then fluctuates around a certain value, continuously exploring and trying different actions until it stabilizes and converges towards the end of the training. The variation of the samples during training does not significantly affect the convergence of the total weight. This indicates that the trained Q-network has good generalization capability and can be effectively applied to unknown scenarios.

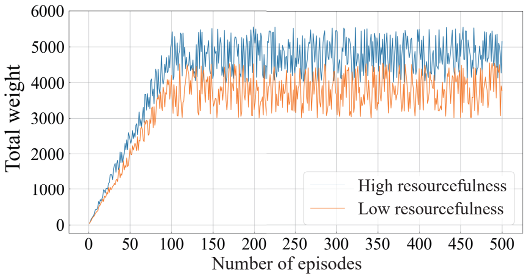


Fig. 9 Total weight of per episode in samples.

The 6 instances, M50-R10, M50-R20, M100-R30, M100-R50, M200-R60, and M200-R100, have undergone preliminary scheduling and obtained abbreviated instances. We apply the trained model to each of these prepped instances, and the scheduling results are shown in Table 5.

In Table 5, $\sum w_i^{\text{rmn}}$ refers to the total weight of the remaining meta-tasks M_{rmn}, the meta-task set that can be scheduled is labeled as $M_{\text{exe}'}$, the total weight result is labeled as $\sum w_i^{\text{exe}'}$, and "Rate" indicates the rate of achievement of the total weights.

5.2.3 Discussions

Combining the scheduling results in Stage-1 and Stage-2, we can summarize the overview for all instances in Table 6, which showcases the completion status of meta-tasks at different stages and the achievement rates of profit after two stages.

At the same time, the achievement rates of scheduling results in terms of weight for each instance is shown in Fig. 10.

Combining Tables 5 and 6 along with Fig. 10, we can easily observe that as the number of meta-tasks increases, it implies that the same number of target ground stations needs to handle more meta-tasks through relayed data transmission from multiple target satellites. For each ground station, the conflicts arising from completing meta-tasks will inevitably increase, leading to a decrease in the revenue attainment rate for completing tasks. On the other hand, by increasing the number of relay satellite resources, the revenue attainment rates of meta-tasks can be significantly improved.

Table 5 Relay transmission FTWA scheduling results.

Intance	$\lvert M_{\text{rmn}}\rvert$	$\sum w_i^{\text{rmn}}$	$\lvert M_{\text{exe}'}\rvert$	$\sum w_i^{\text{exe}'}$	Rate (%)
M50-R10	31	1585	23	1356	85.5
M50-R20	31	1585	29	1505	95.0
M100-R30	71	3663	55	2972	81.1
M100-R50	71	3663	64	3245	88.6
M200-R60	157	8674	118	6573	75.8
M200-R100	157	8674	132	6956	80.2

Table 6 Overall scheduling results.

Number	$\sum w_i$	$\lvert M_{\text{exe}}\rvert$	$\sum w_i^{\text{exe}}$	$\lvert M_{\text{exe}'}\rvert$	$\sum w_i^{\text{exe}'}$	Rate (%)
1	3105	19	1520	23	1356	92.6
2	3105	19	1520	29	1505	97.4
3	5967	29	2304	55	2972	88.4
4	5967	29	2304	64	3245	93.0
5	12 120	43	3446	118	6573	82.7
6	12 120	43	3446	132	6956	85.8

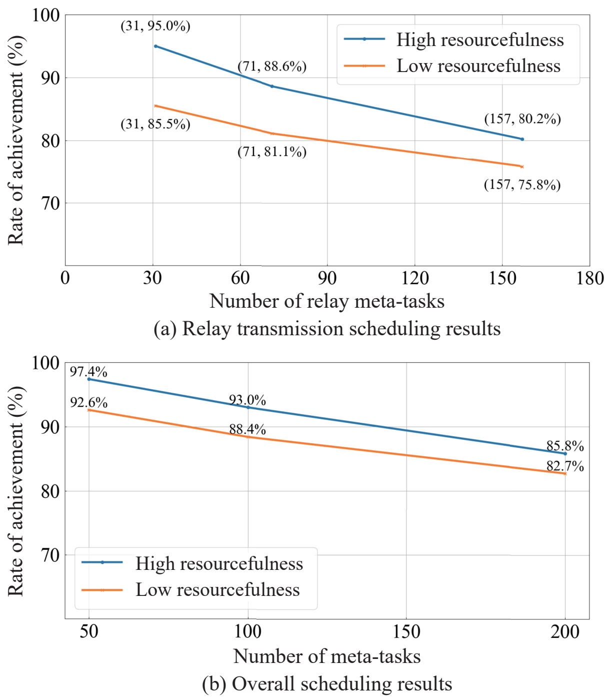


Fig. 10 Achievement rates of scheduling results.

In terms of the execution efficiency of the algorithm, the genetic algorithm Stage-1 needs to spend a minute level of time for obtaining the preprocessed results due to the large number of iterative searches; in Stage-2 of the DQN solution, due to a large number of data training to obtain the solution model, the scheduling results can be generated in the second level of time, which can achieve a higher efficiency of the solution.

6 Conclusion

In this paper, a data-driven cooperative scheduling model is proposed for multi-satellite data transmission, focusing on the potential large-scale challenges faced by LEO satellite data transmission in future application scenarios. To facilitate the study, a set of problem test instances is generated by methodical design, taking into account the problem attributes and scale of the solutions. A two-stage optimization framework, named He-DQN, is introduced to address test instances of various task sizes and resource quantities, combining the classical genetic algorithm, heuristic strategies, and the advanced DQN algorithm. Initially, the genetic algorithm is applied to determine the direct-transmission data meta-tasks set. Subsequently, a greedy strategy is utilized to assign relay satellites to meta-tasks requiring relay transmission, allowing for the identification of feasible time window arrangements for relay transmission links. Finally, the DQN algorithm is employed to schedule precise time window resource allocation for each meta-task. Experimental results demonstrate the effectiveness and accuracy of the He-DQN algorithm in solving the multi-satellite data transmission scheduling problem. The algorithm achieves a high solution quality, with over 80% mission gain attainment rate in all six instances, and it excels in scenarios involving intense resource contention conflicts.

In future research, we can carry out further research in terms of problem model and solution algorithm: (1) In the construction of the model, we can consider the segmented execution of the transmission task and the slicing of the time window, thus utilizing the time window more efficiently. (2) In terms of algorithmic solution, we can consider the reduction of the algorithmic execution steps in future research, which is conducive to the search for the globally optimal solution.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Nos. 42271391, 62006214, and 42101439), the Joint Funds of Equipment Pre-Research and Ministry of Education of China (No. 8091B022148), the 14th Five-Year Pre-Research Project of Civil Aerospace of China, the Hubei Excellent Young and Middle-Aged Science and Technology Innovation Team Plan Project (No. T2021031), and the Open Research Project of Hubei Key Laboratory of Intelligent Geo-Information Processing (No. KLIGIP-2022-B09).

References

[1] Be prepared, stay connected, https://www.iridium.com/, 2020.

[2] X. Lin, S. Cioni, G. Charbit, N. Chuberre, S. Hellsten, and J. F. Boutillon, On the path to 6G: Embracing the next wave of low earth orbit satellite access, *IEEE Commun. Mag.*, vol. 59, no. 12, pp. 36–42, 2021.

[3] F. S. Prol, R. M. Ferre, Z. Saleem, P. Välisuo, C. Pinell, E. S. Lohan, M. Elsanhoury, M. Elmusrati, S. Islam, K. Çelikbilek, et al, Position, navigation, and timing (PNT) through low earth orbit (LEO) satellites: A survey on current status, challenges, and opportunities, *IEEE Access*, vol. 10, pp. 83971–84002, 2022.

[4] S. Chen, Analysis of LEO satellite communication and suggestions for its development strategy in China, (in Chinese), *Telecommun. Sci.*, vol. 36, no. 6, pp. 1–13, 2020.

[5] H. Zhang, J. Xie, J. Ge, J. Shi, and Z. Zhang, Hybrid particle swarm optimization algorithm based on entropy theory for solving DAR scheduling problem, *Tsinghua Science and Technology*, vol. 24, no. 3, pp. 282–290, 2019.

[6] Y. Fu, Y. Hou, Z. Wang, X. Wu, K. Gao, and L. Wang, Distributed scheduling problems in intelligent manufacturing systems, *Tsinghua Science and Technology*, vol. 26, no. 5, pp. 625–645, 2021.

[7] Z. Hu and D. Li, Improved heuristic job scheduling method to enhance throughput for big data analytics, *Tsinghua Science and Technology*, vol. 27, no. 2, pp. 344–357, 2022.

[8] E. Jiang, L. Wang, and J. Wang, Decomposition-based multi-objective optimization for energy-aware distributed hybrid flow shop scheduling with multiprocessor tasks, *Tsinghua Science and Technology*, vol. 26, no. 5, pp. 646–663, 2021.

[9] Y. Yan, K. Du, L. Wang, H. Niu, and X. Wen, MP-DQN based task scheduling for RAN QoS fluctuation minimizing in public clouds, in *Proc. 2022 IEEE Int. Conf. Communications Workshops* (*ICC Workshops*), Seoul, Republic of Korea, 2022, pp. 878–884.

[10] G. Chen, R. Shao, F. Shen, and Q. Zeng, Slicing resource allocation based on dueling DQN for eMBB and URLLC hybrid services in heterogeneous integrated networks, *Sensors*, vol. 23, no. 5, p. 2518, 2023.

[11] M. Wang, G. Dai, and M. Vasile, Heuristic scheduling algorithm oriented dynamic tasks for imaging satellites, *Math. Probl. Eng.*, vol. 2014, p. 234928, 2014.

[12] X. Chen, G. Reinelt, G. Dai, and M. Wang, Priority-based and conflict-avoidance heuristics for multi-satellite scheduling, *Appl. Soft Comput.*, vol. 69, pp. 177–191, 2018.

[13] X. Chen, G. Reinelt, G. Dai, and A. Spitz, A mixed integer linear programming model for multi-satellite scheduling, *Eur. J. Oper. Res.*, vol. 275, no. 2, pp. 694–707, 2019.

[14] S. Huang, D. Ma, D. Fang, and T. Cui, Satellite data transmission scheduling based on improved ant colony system, (in Chinese), *Radio Eng.*, vol. 45, no. 7, pp. 27–30 &58, 2015.

[15] J. Yan, Research on dynamic scheduling approach of relay satellites, (in Chinese), Master dissertation, Xidian University, Xi’an, China, 2017.

[16] Y. Xiang, W. Zhang, and M. Tian, Satellite data transmission integrated scheduling and optimization, (in

Chinese), *Syst. Eng. Electron.*, vol. 40, no. 6, pp. 1288–1293, 2018.
[17] H. Fan, Research on resource scheduling of ground station based on satellite data transmission and telemetry tracing and control tasks, (in Chinese), Master dissertation, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China, 2021.
[18] G. Dong, Research on satellite network data transmission and task scheduling, (in Chinese), Master dissertation, Guilin University of Electronic Technology, Guilin, China, 2022.
[19] Y. Du, Research on the general-purpose scheduling engine for satellite task scheduling problems, (in Chinese), PhD dissertation, National University of Defense Technology, Changsha, China, 2021.
[20] H. Wang, Massive scheduling method under online situation for satellites based on reinforcement learning, (in Chinese), PhD dissertation, University of Chinese Academy of Sciences, Beijing, China, 2018.
[21] M. Chen, Y. Chen, Y. Chen, and W. Qi, Deep reinforcement learning for agile satellite scheduling problem, in *Proc. 2019 IEEE Symp. Series on Computational Intelligence* (*SSCI*), Xiamen, China, 2019, pp. 126–132.
[22] X. Bao, S. Zhang, and X. Zhang, An effective method for satellite mission scheduling based on reinforcement learning, in *Proc. 2020 Chinese Automation Congress* (*CAC*), Shanghai, China, 2020, pp. 4037–4042.
[23] Y. Liu, Q. Chen, C. Li, and F. Wang, Mission planning for earth observation satellite with competitive learning strategy, *Aerosp. Sci. Technol.*, vol. 118, p. 107047, 2021.
[24] L. Ren, X. Ning, and J. Li, Hierarchical reinforcement-learning for real-time scheduling of agile satellites, *IEEE Access*, vol. 8, pp. 220523–220532, 2020.
[25] Y. He, L. Xing, Y. Chen, W. Pedrycz, L. Wang, and G. Wu, A generic Markov decision process model and reinforcement learning method for scheduling agile earth observation satellites, *IEEE Trans. Syst. Man Cybern. Syst.*, vol. 52, no. 3, pp. 1463–1474, 2022.
[26] M. Zuo, G. Dai, L. Peng, and M. Wang, An envelope curve-based theory for the satellite coverage problems, *Aerosp. Sci. Technol.*, vol. 100, p. 105750, 2020.
[27] X. Chen, Z. Xu, and L. Shang, Satellite internet of things: Challenges, solutions, and development trends, *Front. Inf. Technol. Electron. Eng.*, vol. 24, no. 7, pp. 935–945, 2023.
[28] Z. Zhang, Y. Bai, and C. Li, Overview of developments in global navigation satellite systems, (in Chinese), *Sci. Technol. Innovation*, no. 9, pp. 150–152, 2023.
[29] W. Sun, G. Yang, C. Chen, M. Chang, K. Huang, X. Meng, and L. Liu, Development status and literature analysis of China's earth observation remote sensing satellites, (in Chinese), *Natl. Remote Sens. Bull.*, vol. 24, no. 5, pp. 479–510, 2020.
[30] C. Fan, X. Chen, A. Liu, and J. Luo, Read the "sky chain": A text to understand China's "data relay satellite corps", (in Chinese), *Space Int.*, no. 8, pp. 7–10, 2022.
[31] Z. Zhang, D. Wang, X. Man, and Y. Zhang, Application prospect and challenge of low orbit internet system in navigation enhancement service, (in Chinese), *World Sci-Tech R&D*, vol. 45, no. 3, pp. 266–275, 2023.
[32] X. Zhang and F. Ma, Review of the development of LEO navigation-augmented GNSS, (in Chinese), *Acta Geod. Cartogr. Sin.*, vol. 48, no. 9, pp. 1073–1087, 2019.
[33] Y. Meng, L. Bian, Y. Wang, T. Yan, W. Lei, M. He, and X. Li, Global navigation augmentation system based on Hongyan satellite constellation, (in Chinese), *Space Int.*, no. 10, pp. 20–27, 2018.
[34] X. Wu, D. Qin, Y. Zhang, and Y. Li, Supporting capability for custody layer of the next-generation space architecture, (in Chinese), *J. Inf. Eng. Univ.*, vol. 24, no. 3, pp. 379–384, 2023.
[35] P. Zhao and Z. Chen, An adapted genetic algorithm applied to satellite autonomous task scheduling, (in Chinese), *Chin. Space Sci. Technol.*, vol. 36, no. 6, pp. 47–54, 2016.

Xiaoyu Chen received the BEng degree in information security, the MEng degree in computer technology, and the PhD degree in geoinformation engineering from China University of Geosciences, Wuhan, China in 2012, 2014, and 2018, respectively. He is currently an associate professor at School of Computer Science, China University of Geosciences, Wuhan, China. His research interests include combinatorial optimization and satellite scheduling.

Weichao Gu received the BEng degree from China University of Geosciences, Wuhan, China in 2021. He is currently a master student in computer technology at School of Computer Science, China University of Geosciences, Wuhan, China. His research interests are satellite scheduling and machine learning.

Guangming Dai received the BEng and MEng degrees in applied geophysics from China University of Geosciences, Wuhan, China in 1985 and 1996, respectively, and the PhD degree in computer science and technology from Huazhong University of Science and Technology, Wuhan, China in 2004. He is currently a professor at School of Computer Science, China University of Geosciences, Wuhan, China. His research interests include space information network and military artificial intelligence applications.

Lining Xing received the double BEng degrees from Xi'an Jiaotong University, China in 2002, and the PhD degree in management science from National University of Defense Technology, China in 2009. He is currently a professor at School of Electronic Engineering, Xidian University, China. His research interests include evolutionary algorithm and combinatorial optimization.

Tian Tian received the BEng degree from Southwest University of Science and Technology, China in 2022. She is currently a master student in electronic information at School of Computer Science, China University of Geosciences, Wuhan, China. Her research interests are scheduling problem and deep reinforcement learning.

Weilai Luo received the BEng degree from China University of Geosciences, Wuhan, China in 2022. He is currently a master student in electronic information at School of Computer Science, China University of Geosciences, Wuhan, China. His research interests are intelligent optimization problems and algorithms.

Shi Cheng received the BEng degree in mechanical and electrical engineering from Xiamen University, Xiamen, China in 2005, the MEng degree in software engineering from Beihang University, Beijing, China in 2008, and the PhD degree in electrical engineering and electronics from University of Liverpool, Liverpool, UK in 2013. He is currently an associate professor at Shaanxi Normal University, Xi'an, China. His research interests include swarm intelligence, multiobjective optimization, and data mining techniques.

Mengyun Zhou received the BEng degree in geosciences from Yangtze University, Jingzhou, China in 2010, and the PhD degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China in 2016. She is currently an associate professor at School of Computer Science, China University of Geosciences, Wuhan, China. Her research interests include space information network and spatio-temporal big data analysis.