Hierarchical Cross-Domain Satellite Resource Management: An Intelligent Collaboration Perspective

Hongmei He, Di Zhou, *Member, IEEE*, Min Sheng, *Senior Member, IEEE*, and Jiandong Li, *Fellow, IEEE*

***Abstract*— The expansion of satellite applications induces the formation of the multi-domain satellite system (MDSS) containing multiple domains with specific applications such as earth resource remote sensing and the Internet of remote things. Resource management is pivotal in enhancing the scheduling capability of the MDSS. However, this is challenging since the dynamic buffer space and communication opportunity, as well as the uncertain data traffic, exacerbate the difficulty of matching satellite resources with data traffic. Moreover, the coexistence of resource competition and collaboration across domains aggravates the dilemma of cross-domain collaboration. In this paper, we propose a hierarchical cross-domain collaborative resource management framework that can flexibly allocate the mission data through local intra-domain and global cross-domain scheduling. Then, to match the uncertain demands of missions with dynamic and limited resources, we propose a multi-agent reinforcement learning-based resource management method to guide collaboration for multi-satellite data carry-forward in a domain. Further, considering resource competition and collaboration in MDSS, we propose a domain-satellite nested matching game data scheduling algorithm to achieve pair-wise stable collaboration of cross-domain satellites. The simulation results indicate that the proposed algorithm improves the amount of offloaded data by 64.4% and 12.7% compared to the non-collaborative and the non-cross-domain schemes, respectively.**

***Index Terms*— Multi-domain satellite system, hierarchical resource management, multi-agent collaboration, matching game.**

I. Introduction

A. Motivation

RECENTLY the rapid development of space exploration significantly catalyzes the application of satellites, which are widely applied to Earth observation, Internet of remote things (IoRT), and communications with their seamless and ubiquitous coverage [1], [2], [3], [4]. In the future, satellites will even be employed in the envisioned construction of the interplanetary Internet and the Internet of space things [5], [6], [7], [8], [9]. Currently, various countries and commercial companies are scrambling to deploy satellites, presenting a novel trend of the coexistence of multiple satellite domains [10], [11]. The domain is a collection of satellites that realize a specific application, e.g., earth resource satellites constitute the earth resource remote sensing domain, IoRT satellites constitute the IoRT domain, and meteorological satellites make up the meteorological observation domain. With the increase in the number of domains and the scope of applications, the independent multi-domain characteristics of the satellite networks are becoming increasingly prominent, which inevitably leads to low utilization of scarce network resources and poor network utility. In this paper, we concentrate on resource management for the multi-domain satellite system (MDSS).

In reality, the amount of mission data varies among domains, resulting in both resource shortage and resource waste in MDSS. Fortunately, the cross-domain collaborative resource management for data scheduling (CDC-RMDS) can transmit mission data from one domain to another [12]. However, an effective mechanism for cross-domain collaboration is lacking due to the difficulty of matching resources and missions with different properties among multiple domains [10], [13], [14]. In addition, the exponential increase of data, and the unprecedented growth of applications such as the satellite Internet of Things [9], [15], [16], aggravate the mismatch between mission data and resources. Consequently, it is essential to investigate the collaborative resource management of the MDSS to provide theoretical support for the construction, operation, and management of future satellite networks.

The current satellite network resource management mechanisms can be mainly divided into two categories, namely, non-collaborative satellite resource management (NCSRM) and collaborative satellite resource management (CSRM). The NCSRM mechanism lacking in collaboration by inter-satellite links (ISLs) can only offload data by the satellite-ground links (SGLs). As a result, it is inefficient when the deployment of ground stations (GSs) is limited [17]. In terms of the CSRM mechanism, ISLs provide opportunities for data exchange among satellites [18]. Consequently, the satellite with limited

Manuscript received 15 May 2022; revised 21 October 2022 and 19 January 2023; accepted 21 January 2023. Date of publication 31 January 2023; date of current version 18 April 2023. This work was supported in part by the Natural Science Foundation of China under Grant U19B2025, 62121001, and 62001347, in part by Key Research and Development Program of Shaanxi (Program No. 2022ZDLGY05-02), and in part by Young Talent Support Program of Xi'an Association for Science and Technology (No. 095920221337). The associate editor coordinating the review of this article and approving it for publication was W. Saad. *(Corresponding author: Di Zhou.)*

The authors are with the State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China (e-mail: hehongmei@stu.xidian.edu.cn; zhoudi@xidian.edu.cn; msheng@mail.xidian.edu.cn; jdli@mail.xidian.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCOMM.2023.3241185.

Digital Object Identifier 10.1109/TCOMM.2023.3241185

GS offloading resources can load data to satellites with abundant offloading resources by ISLs [19]. It is a promising solution to improve resource utilization through inter-satellite (e.g., intra-domain and cross-domain) collaborative resource management using ISLs. However, developing an effective CDC-RMDS strategy for the MDSS has several challenges:

- *Multi-Domain Independent Resource Collaboration*: The CDC-RMDS requires elaborate planning of resources and mission data distributed across multiple domains. The competition for resources among busy domains, as well as the collaboration between busy and idle domains, results in the coexistence of collaboration and competition among multiple domains. Therefore, the CDC-RMDS among satellites in multiple domains is tricky.
- *Intra-Domain Stochastical Mission Data Arrival*: The unpredictability of arrived mission data leads to incomplete information about data arrival in the satellite networks, which means that the demand for resources in each domain is not completely known [20], [21]. Therefore, a full information-based resource management approach is not applicable. However, designing dynamic data scheduling methods to cope with random data traffic in the satellite networks exacerbates the difficulty of efficient intra-domain multi-satellite collaborative resource management.
- *Intra-Domain Dynamic and Limited Multi-Dimensional Resource Management*: On the one hand, the dynamic connectivity of the network inevitably leads to intermittent communication opportunities for satellites [22]. The dynamic and limited resources increase the urgent demand for efficient resource management. On the other hand, the data offloading is closely related to the allocation of multi-dimensional resources, including onboard processing (OBP) resources, storage resources, and transmission resources [23]. Considering the causality of the scheduling process, it is necessary to rationalize the sequence of data arrival, processing storage and transmission [24].

B. Literature Review

Resource management plays an essential role in enhancing the data scheduling capability of satellite system. Some prominent works alleviate the transmission inefficiency resulting from the intermittent connections between satellites and GSs [10], [19], [20], [21], [25], [26], [27], [28]. In this section, we detail these recent investigations on resource management with inter-satellite collaboration by ISLs, which fall into the following three main categories.

The first class of investigations consider the predictable dynamic resource and models the inter-satellite collaborative data offloading as a multidimensional resource joint management problem [19], [25], [26]. Nevertheless, all of these works ignore the unpredictability of arrived mission data. To deal with this problem, considering the time-varying characteristic of the satellite network topology, the authors in [25] proposed a multi-hop transmission data offloading method to enhance the amount of offloaded data in finite communication opportunities. Further, by jointly considering data offloading from satellites to the GS and data forwarding among satellites to allocate the finite offloading time, [26] developed an iterative optimization algorithm to significantly improve the volume of offloaded data. The authors in [19] achieved a significant improvement in energy efficiency with guaranteed amount of offloaded data by multi-power level transmission resource management.

Further, in the second category, some stochastic optimization techniques are proposed to settle with the inter-satellite collaborative scheduling under the stochastic arrived mission data [20], [21], [27]. Specifically, considering the impact of current scheduling decisions on the amount of offloaded data in the future, [20] proposed an MDP-based inter-satellite store-carry-forward scheduling framework to maximize long-term network efficiency. In addition, given the unpredictability of long-term data arrival distribution and the dynamics of network resources, in [27], a robust optimization-based approach was proposed to improve the data transmission performance of multi-satellite networks. The recent work [21] focused on the long-term optimization problem of joint data offloading of satellite clusters. Then decomposed the problem by Lyapunov optimization theory and proposed an online resource optimization scheme for data offloading. However, these investigations only concentrate on the unified scheduling of data in an application-specific domain, while ignoring the collaborative scheduling among multiple domains.

Moreover, the third type of investigations concentrate on collaborative inter-satellite scheduling in a MDSS [10], [28]. Recent work [28] modeled the dynamic satellite resources and proposed an on-demand scheduling algorithm with intra-domain and cross-domain missions. However, [28] ignored the performance degradation resulting from inter-domain resource imbalance. To achieve the collaboration of resources among independent satellite systems and improve resource utilization, [10] proposed a scheduling architecture to realize interoperability among different satellite systems. The feasibility of cross-domain mission scheduling by deep reinforcement learning under this framework is demonstrated. Nonetheless, in practice, [10] still faces challenges such as dynamic resource and stochastic data traffic.

To sum up, most existing works neglect either stochastic data traffic or multi-domain collaboration. Different from these works, we investigate collaborative resource scheduling of intra-domain and cross-domain satellites, while considering the long-term resource scheduling under stochastic data traffic.

C. Contributions

In this work, we concentrate on the efficient CDC-RMDS of multi-domain resources to obtain the maximum amount of offloaded data. To this end, we consider the features of the satellite networks in the modeling of the resource management problem for the MDSS data scheduling and investigate the CDC-RMDS problem from a hierarchical collaboration perspective (i.e., intra-domain and cross-domain).

For clarity, the major contributions of this work are listed as follows:

- *Practical MDSS and Resource Management Framework*: We propose a collaborative MDSS optimization model considering practical issues such as unpredictable arrived mission data, limited buffer and OBP, and dynamic communication opportunities. Further, we propose a hierarchical collaborative resource management framework that can hierarchically and flexibly allocate the available resources and mission data within and between domains by ISLs to facilitate the capabilities of data scheduling.
- *Dynamic Intra-Domain Multi-Satellite Management Strategy*: We model the problem of multi-satellite resources management with ISLs in a domain as a problem of maximizing the amount of successfully offloaded data with multi-agent collaboration. On this basis, we develop a multi-satellite scheduling decision-making system based on the actor-critic framework and propose a multi-satellite collaborative scheduling (MSCS) algorithm. Through centralized training and decentralized execution, we obtain the multi-satellite scheduling policy in a domain that adapts to the dynamic environment and guides efficient collaborative resource scheduling among satellites.
- *Cross-Domain Collaborative Management Algorithm*: We propose a cross-domain scheduling algorithm based on the domain-satellite nested matching game (DSNMG) to realize multi-domain matching and cross-domain satellite collaboration. In the inner layer, we consider the storage and transmission resource to achieve stable two-way selection collaboration among satellites of different domains. Based on this, in the outer layer, we design the cross-domain matching preference lists according to the matching results of the inner layer to obtain a stable cross-domain matching collaboration scheme. We alleviate the inappropriate allocation of resources and data in the MDSS by designing stable and efficient cross-domain collaboration scheduling algorithms for cross-domain data transmission.
- *Verification*: Extensive simulations are implemented to evaluate our developed algorithms. The proposed hierarchical resource management algorithm improves the amount of offloaded data compared to benchmark algorithms. We also investigate the impacts of several parameters on the data scheduling performance and the resource utilization of the MDSS, which can provide a guide for the future design of MDSS resource management.

D. Organization

The remainder of the paper is organized as follows. The elaborated reference system model is in section II. In Section III we formulate the resource management problem in MDSS and further formulate it by exploiting the multi-agent reinforcement learning (MARL) framework. Then, we propose a hierarchical cross-domain resource management algorithm in Section IV. Simulation results are presented and analyzed in Section V, followed by a conclusion of this work in the final Section VI.

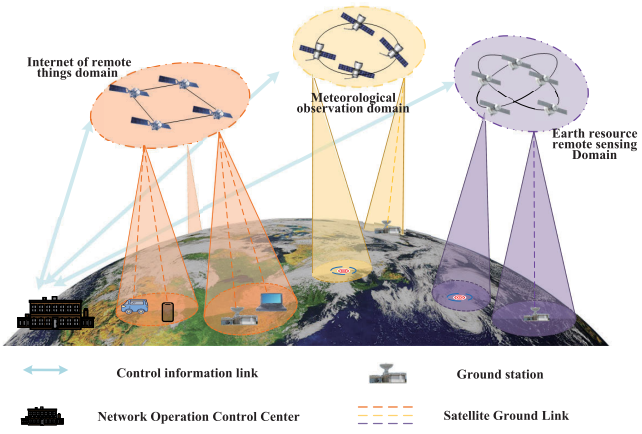


Fig. 1. A multi-domain satellite system scenario.

II. System Model

The goal of network resource scheduling is to maximize the amount of offloaded mission data in MDSS through the collaboration of satellites. Therefore, in this section, we describe the system model including the network model, data traffic model, and resource model to facilitate the formulation of resource scheduling. Specifically, the network model specifies basic composition of satellite network and the data traffic model describes the distribution of randomly arrived mission data on satellite. The resource model represents the transmission, storage, and OBP resources on the satellite. Table I summarizes the key notations and abbreviations in this paper.

A. Network Model

We consider a MDSS $\mathcal{D}$ consisting of K domains $\mathcal{D} = \{DOM_1, DOM_2, \cdots, DOM_K\}$, each of which contains a set of intra-domain satellites $U_k = \{u_1, u_2, \cdots, u_N\}$, where N represents the number of satellites in a domain. $\mathcal{U} = \{U_1, U_2, \cdots, U_K\}$ represents the set of all satellites of the MDSS. In addition, there is a group of GSs represented by $\mathcal{GS} = \{GS_1, GS_2, \cdots, GS_M\}$, which is responsible for receiving data from each domain. The network operation control center (NOCC) on the ground uploads control information to satellites. Fig. 1 shows an example of the MDSS containing three domains. In the considered scenario, the satellites in each domain continuously store, process, and transmit data traffic. Then satellites offload the processed data to the GSs via inter-satellite storage and forwarding. The scheduling period $\mathcal{T}$ is divided into T time slots as $\Gamma = \{1, 2, 3, \cdots, T\}$, and we assume that in each time slot the network topology is fixed [20]. The visible satellite of i represents the satellites in its field-of-view, that is, there is a communication opportunity between them. A set of available communication links is denoted as $\mathcal{E}$, which includes ISLs and SGLs.

B. Data Traffic Model

The distribution of satellite mission traffic is closely related to the type of domain and geography. We classify the traffic

TABLE I
THE KEY NOTATIONS AND ABBREVIATIONS

Notations/Abbreviations	Description/Full name
DOM_k, $\mathcal{U}$, $\mathcal{GS}$, $\mathcal{E}$, e_{ij}^t	The k-th domain in MDSS; the set of satellites and GSs in the MDSS; the set of available communication links; the link from i to j in t-th time slot.
T, τ, $\mathcal{T}$	The total number of time slots in the scheduling period; the length of time slot; scheduling period.
λ, γ, ψ	The average volume of arrived mission data in a time slot; the discount factor; the neighbor discount factor.
$d_b^{i,t}$, $d_r^{i,t}$, $\delta_{e_{ij}^t}$	The amount of random arrived data in satellite i in t-th slot; the amount of data received by satellite i from neighbor satellites in t-th slot; boolean variable indicates whether the link e_{ij}^t is active.
CB_i^t, CB_{max}, B_i^t	The remaining storage capacity of satellite i; the buffer capacity; the volume of data in the satellite buffer.
ρ, a_i^t, a_{vir}, $\mathcal{N}_i^t$	Maximum amount of data processed by the OBP in a time slot; the action of satellite i in t-th time slot; the virtual action; the set of neighbor nodes of satellite i in t-th time slot.
$d_p^{i,t}$, $d_{un\_p}^{i,t}$	Processed data on satellite i; unprocessed data on satellite i.
$\tilde{s}_i^t$, $\tilde{r}^t$	The overall state of satellite i; the overall reward in a domain.
$R_{e_{ij}}^t$, $c(e_{ij}{}^t)$, $x_{e_{ij}}^t$	The data rate of link e_{ij}; the capacity of link e_{ij}; the amount of data transmit from i to j in t-th time slot.
$\mathscr{M}_i^t : \tilde{s}_i^t \rightarrow a_i^t$	A mapping from a state to action in satellite i.
$\mathscr{R}$, $\mathcal{R}$, $\mathbb{R}$	The cumulative expected reward of a satellite, a domain, and a MDSS.
ϕ, φ	The matching of a domain; the matching of satellites in a domain.
π, $\mathcal{P}$, $\mathbb{P}$, Π	The strategy of a satellite; the strategy of each satellite in a domain; the strategy of each satellite in the MDSS; the set of all feasible satellite connect strategies.
$\mathbb{E}[\cdot]$, $\mathrm{P}(\cdot)$	The expectation function; the preference list.
MDSS, SGLs, ISLs	Multi-domain satellite system; satellite ground links; inter-satellite links.
GSs, NOCC, IoRT	Ground stations; network operation control center; Internet of remote things.
CDC-RMDS, MSCS	Cross-domain collaborative resource management for data scheduling; multi-satellite collaborative scheduling.
NCGT, BCT, DSNMG	Non-collaborative greedy transmission scheme; blind collaboration transmission scheme; domain-satellite nested matching game.
MARL, MAPPO, MAA2C	Multi-agent reinforcement learning; multi-agent proximal policy optimization; multi-agent advantage actor-critic.

intensity h_i into different levels, i.e., $H = \{h_1, h_2, \ldots h_I\}$. In a domain, the mission traffic intensity of the satellite can be expressed as $h_i = \mathcal{F}_{dom}(x_{long}^t, x_{lat}^t)$, where x_{long}^t and x_{lat}^t denote the longitude and latitude of the subsatellite point of the satellite in t-th time slot, respectively. $\mathcal{F}_{dom}(\cdot)$ denotes the traffic intensity distribution function of the domain.

Further, the mission traffic is also stochastic [20]. The volume of arrived mission data can be represented as

$$d_b^{i,t} = n_i^t * \varpi_m, \tag{1}$$

where n_i^t and ϖ_m represent the number of mission bursts on satellite i at t-th time slot and the data amount of the mission, respectively. We assume that n_i^t is independently and identically distributed (i.i.d.) in each time slot. We define $\Pr[n_i^t]$ as the probability density of n_i^t, and $n_i^t \in \{0, 1, 2, \ldots\}$. $\Pr[n_i^t]$ is different in various domains. For example, in the IoRT domain and the earth resource remote sensing domain, we assume that [29] and [30]

$$\Pr[n_i^t] = \frac{\left[u^{h_i}\right]^{n_i^t}}{n_i^t!} e^{-\left[u^{h_i}\right]}, \tag{2}$$

where u^{h_i} denotes the average number of arrived missions per time slot in a satellite at level h_i. In meteorological observation domain, we assume that [31]

$$\Pr[n_i^t] = \begin{cases} \dfrac{1}{2\sigma}, & u^{h_i} - \sigma \leqslant n_i^t \leqslant u^{h_i} + \sigma \\ 0, & otherwise \end{cases}, \tag{3}$$

where $u^{h_i} - \sigma$ and $u^{h_i} + \sigma$ represent the lower and upper bounds of the number of arrived mission data, respectively. The average arrived mission data of satellite i in a time slot is $\mathbb{E}[d_b^{i,t}] = \lambda_i$.

C. Resource Model

The specified applications of different domains are determined by satellite resources in that domain. These applications include meteorological observation, IoRT, earth resource remote sensing, and other applications that require carrying and offloading data via satellites. We divide the resources of each domain in a MDSS into three categories: transmission, storage, and OBP resources. The amount of transmission resources refer to the transmission capacity between satellites and between satellites and the ground. The amount of storage resources are the buffer capacity on the satellite, which are indicated as CB_{max}. In addition, the amount of OBP resources represents the ability to process application-specific data in a time slot, which is indicated as the OBP rate ρ.

Link rate is used to represent the transmission capability of the link and can be represented as [20]

$$R_{e_{ij}^t} = \frac{P_{itr} G_{itr} G_{jre} L^t_{f_{ij}}}{kT_s \cdot (E_b/N_0)_{req} \cdot M}, \tag{4}$$

wherein, the free space loss $L^t_{f_{ij}}$ is

$$L^t_{f_{ij}} = (\frac{c}{4\pi \cdot S(e^t_{ij}) \cdot f})^2. \tag{5}$$

Here, P_{itr} and G_{itr} are the satellite transmission power (in W) and satellite transmitting antenna gain of the link e^t_{ij}. The receiving antenna gain of satellites or GSs is indicated as G_{jre}. Furthermore, k and T_s are the Boltzmann constant (in JK^{-1}) and the total system noise temperature (in K). $(E_b/N_0)_{req}$ is the required ratio of received energy-per-bit to noise-density. M is the link margin which indicates the minimum additional gain required to exceed the link budget [1]. Additionally, c is the speed of light (in km/s), and f denotes the communication center frequency (in Hz) of link e^t_{ij}. $S(e^t_{ij})$ represents the slant range (in km). Therefore, the maximum amount of data for link e^t_{ij} in a time slot can reach

$$c(e^t_{ij}) = R_{e^t_{ij}} \cdot \tau, \quad \forall e^t_{ij} \in \mathcal{E}. \tag{6}$$

At extremely high frequency (EHF), there is serious atmospheric precipitation that leads to communication interruption [32]. For a SGL with low attenuation the value for a link margin is 2-3 dB. By contrast, for EHF-band systems (e.g., Ka-band), a high level of link margin is required (10 dB or more) for intensive atmospheric effects [1].

III. Problem Formulation

A. Maximize the Amount of Offloaded Data

1) Constraints: There are differences in the type of OBP resources ρ required for different missions. For example, missions in the meteorological observation domain require OBP resource ρ_m for meteorological parameter correction, missions in the IoRT domain require OBP resources ρ_r for normalizing IoRT data from different devices, missions in the earth resources remote sensing domain require OBP resources ρ_e for image correction, etc. Therefore, these constraints are mission-specific. The mission data $d^{i,t}$ received by satellite i in t-th time slot can be divided into the arrived mission data of itself $d^{i,t}_b$ and the data received from other satellites $d^{i,t}_r$, that is,

$$d^{i,t} = d^{i,t}_b + d^{i,t}_r. \tag{7}$$

The data in the satellite buffer can be divided into processed data and unprocessed data, which can be represented by

$$B^t_i = d^{i,t}_p + d^{i,t}_{un\_p}, \tag{8}$$

where $d^{i,t}_{un\_p}$ represents the amount of unprocessed data of satellite i at the beginning of the t-th time slot. Considering the causality of the system, the arrived data traffic at the current time slot can only be processed in the later time slot. Therefore, $d^{i,t}_{un\_p}$ contains a part of historical unprocessed mission data $d^{t-1,i}_{un\_p}$ that exceeds the processing capacity ρ_k in a time slot and the amount of arrived mission data on satellite i in $t-1$-th time slot, that is,

$$d^{i,t}_{un\_p} = \max\{d^{i,t-1}_{un\_p} - \rho_k, 0\} + d^{i,t-1}_b, \quad \forall i \in DOM_k, \forall DOM_k \in \mathcal{D}. \tag{9}$$

Similarly, the amount of processed data of satellite i at the beginning of the t-th time slot is represented as

$$d^{i,t}_p = \min\{d^{i,t-1}_{un\_p}, \rho_k\} + d^{i,t-1}_r, \forall i \in DOM_k, \forall DOM_k \in \mathcal{D}. \tag{10}$$

Moreover, The constraints for different domain missions also include some general constraints. We consider a practical network with a limited amount of transmission resources and storage resources. We assume that only processed data can be transmitted between satellites. Therefore, the amount of data $x_{e^t_{ij}}$ on link e^t_{ij} cannot exceed either the link capacity $c(e^t_{ij})$ or the amount of processed data $d^{i,t}_p$ on satellite i, i.e.,

$$\delta_{e^t_{ij}} \cdot x_{e^t_{ij}} \leqslant \min\{c(e^t_{ij}), d^{i,t}_p\}, \quad \forall e^t_{ij} \in \mathcal{E}, t \in \Gamma. \tag{11}$$

It is evident that, the total amount of data stored on satellite i cannot exceed its buffer capacity $CB_{\max}$, i.e.,

$$B^t_i + d^{i,t}_b + d^{i,t}_r - \sum_{j \in \mathcal{U} \cup \mathcal{GS}} x_{e^{t+1}_{ij}} \cdot \delta_{e^{t+1}_{ij}} \leqslant CB_{\max}, \quad \forall d^{i,t}_b \in \mathbf{d}^i_b, i \in \mathcal{U}, t \in \{1, 2, \ldots T-1\}, \tag{12}$$

where, $\{\mathbf{d}^i_b = d^{i,t}_b | t \in \Gamma\}$. Due to the limitation of the number of onboard transponders, we assume that the satellite i can only select one node at a time slot to establish a link for data transmission, even if i has more visible satellites or GSs [33]. In addition, satellite i can only receive data from one satellite at a time slot. That is,

$$\sum_{j \in \mathcal{U} \cup \mathcal{GS}, i \neq j} \delta_{e^t_{ij}} \leqslant 1, \forall i \in \mathcal{U}, t \in \Gamma, \tag{13}$$

and

$$\sum_{i \in \mathcal{U}, i \neq j} \delta_{e^t_{ij}} \leqslant 1, j \in \mathcal{U}, t \in \Gamma. \tag{14}$$

The decision variable $\delta_{e^t_{ij}} \in \{0, 1\}$ indicates whether a link $e^t_{ij} \in \mathcal{E}$ is established between i and j. If there is a link established between i and j, $\delta_{e^t_{ij}} = 1$. Otherwise, $\delta_{e^t_{ij}} = 0$. Moreover, all incoming data flow and all outgoing data flow on a satellite node must satisfy flow conservation in each satellite. Therefore, we have the following constraints,

$$\begin{cases} \sum_{j \in \mathcal{U}} \delta_{e^t_{ji}} \cdot x_{e^t_{ji}} + d^{i,t}_b = \sum_{j \in \mathcal{U} \cup \mathcal{GS}} \delta_{e^t_{ij}} \cdot x_{e^t_{ij}} + x_{e^t_{ii}}, \\ \qquad \forall d^{i,t}_b \in \mathbf{d}^i_b, i \in \mathcal{U}, i \neq j, t = 1, \\ \sum_{j \in \mathcal{U}} \delta_{e^t_{ji}} \cdot x_{e^t_{ji}} + d^{i,t}_b + x_{e^{t-1}_{ii}} = \sum_{j \in \mathcal{U} \cup \mathcal{GS}} \delta_{e^t_{ij}} \cdot x_{e^t_{ij}} + x_{e^t_{ii}}, \\ \qquad \forall d^{i,t}_b \in \mathbf{d}^i_b, i \in \mathcal{U}, i \neq j, t \in \{2, 3, \ldots, T\}, \end{cases} \tag{15}$$

where $x_{e^t_{ii}}$ represents the data stored in the buffer of satellite i at the beginning of the $t+1$-th time slot.

2) Formulation: According to the relevant constraints, we construct the CDC-RMDS problem to maximize the amount of scheduled data in MDSS.

$$\mathbf{CDC-RMDS}: \max_{\delta} \sum_{t\in\Gamma}\sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{GS}} x_{e_{ij}^t} \tag{16}$$

$$\text{s.t. } (9)-(15).$$

$$\delta_{e_{ij}^t} \in \{0,1\}, \forall e_{ij}^t \in \mathcal{E}. \tag{17}$$

B. Problem Formulation Based on Multi-Agent Reinforcement Learning

In the CDC-RMDS, the inter-satellite transmission is mutually coupled in each time slot. Naturally, considering the satellite resource state and the volume of offloaded data, we characterize the decision-making of satellite transmission connection as an action-state-reward driven MDP process. According to the data traffic model, the stochastic arrived mission data $d_b^{i,t}$ in constraints can not be accurately predicted. This causes failure to directly solve the problem by large-scale linear programming. Therefore, considering the random fluctuation of data traffic during scheduling, we adopt a model-free reinforcement learning framework to conduct inter-satellite transmission scheduling according to the resource state of satellites. In this way, the decision-making of inter-satellite transmission connection adapts to the random data traffic, such that the amount of offloaded data is maximized over the entire scheduling period.

1) State: The state of satellites in domain k in t-th time slot can be represented as

$$S_k^t = \{s_1^t, s_2^t, \cdots, s_N^t\}_k, \tag{18}$$

in which s_i^t is the state of satellite i in t-th time slot and can be expressed as

$$s_i^t \triangleq \left\langle CB_i^t, t^* \right\rangle. \tag{19}$$

Here, the buffer states of the satellite i are represented by CB_i^t. Continuous variable $CB_i^t \in [0, CB_{\max}]$ represents the remaining storage capacity of satellite i. The maximum common multiple of the periodic operation of the satellite and the Earth's rotation is set as the scheduling period. The orbital relationship and link situation will be reproduced after one scheduling period. The state characteristic t^* represents the current link state of the satellite. In the considered multi-satellite scenario, the overall state of satellite i includes not only its storage and link state s_i^t, but also the state of neighbor node [34]. The overall state of satellite i can be expressed as

$$\tilde{s}_i^t = [s_i^t] \cup [s_j^t]_{j\in\mathcal{N}_i^t}, \tag{20}$$

where the $\mathcal{N}_i^t$ is the set of neighbor nodes of i in t-th, including the satellites and GSs within i's line-of-sight.

2) Action: The action of a satellite is selecting a neighbor node for data transmission. The action of satellite i in t-th time slot is represented as a_i^t, where $a_i^t \in \mathcal{N}_i^t$. The action space of each satellite is composed of the neighbor nodes $\mathcal{N}_i^t$. However, the communication opportunities between satellites are constantly changing, resulting in a variation in the action space. Therefore, the action space size of satellite i is determined by the maximum number of the neighbor node of satellite i during the scheduling period. Due to the limited number of onboard transponders [33], we assume that a satellite can only construct one transmitting link and one receiving link in a time slot. Therefore, some neighbor nodes may not be available for data receiving or transmission. The invalid actions are defined as the actions that should be excluded to avoid conflicted transmission. The connection constraints of satellites can be written as

$$\sum_{j\in\mathcal{N}_i^t} a_{ij}^t \le 1,\ \forall i \in \mathcal{U}, t \in \Gamma, \tag{21}$$

$$a_{ij}^t + a_{ji}^t \le 1,\ i \ne j, \forall i, j \in \mathcal{U}\cup\mathcal{GS}, t \in \Gamma, e_{ij}^t \in \mathcal{E}, \tag{22}$$

$$\sum_{j\in\mathcal{U}} a_{ji}^t \le 1,\ \forall i \in \mathcal{U}, t \in \Gamma, e_{ij}^t \in \mathcal{E}, \tag{23}$$

$$\sum_{i\in\mathcal{U}} a_{ij}^t \le \mathcal{F},\ \forall j \in \mathcal{GS}, t \in \Gamma, e_{ij}^t \in \mathcal{E}, \tag{24}$$

where $a_{ij}^t \in \{0,1\}$ represents whether the link from the satellite node i to j is established. The maximum number of satellites connected with the GS in a time slot is denoted as $\mathcal{F}$.

3) Reward: Each satellite makes decisions for connection action selection according to the policy $\pi(\tilde{s}_i^t)$ and its overall states $\tilde{s}_t^i$, but some actions are unrealistic due to the real constraints including buffer, link rate, transponder, etc. Therefore, we carefully formulate the reward as a solution to a linear programming problem that excludes unrealistic actions.

In the process of data transmission during a time slot, the volume of data sent on the selected link should be less than its capacity, i.e.,

$$x_{e_{ji}^t} \cdot a_{ji}^t \le c(e_{ji}^t),\ \forall j \in \mathcal{U}, t \in \Gamma, e_{ji}^t \in \mathcal{E}. \tag{25}$$

The data received by satellite i minus the data sent by satellite i does not exceed its remaining capacity in a time slot, which can be represented as

$$d_b^{i,t} + d_r^{i,t} - \sum_{j\in\mathcal{U}\cup\mathcal{GS}} (x_{e_{ij}^t} \cdot a_{ij}^t) \le CB_i^t,\ \forall i \in \mathcal{U}, t \in \Gamma, e_{ij}^t \in \mathcal{E}. \tag{26}$$

The $d_r^{i,t}$ can be represented as

$$d_r^{i,t} = \sum_{j\in\mathcal{U}, j\ne i} x_{e_{ji}^t}. \tag{27}$$

Moreover, there is a constraint that needs to be satisfied to ensure that the volume of data transmitted on the link is less than the amount of processed data, i.e.,

$$x_{e_{ji}^t} \le d_p^{j,t},\ \forall j \in \mathcal{U}, t \in \Gamma, e_{ij}^t \in \mathcal{E}. \tag{28}$$

In a domain, multiple satellites participate in collaborative scheduling to optimize the amount of total offloaded data. The overall reward can be expressed as $r^t = \sum_{i\in U_k} r_i^t$, in which the reward of each agent r_i^t can be easily measured as $r_i^t = \sum_{j\in\mathcal{GS}} x_{e_{ij}}^t$. Therefore, the reward of the system under resource

constraints can be expressed as

$$r^t = \max_{a_{ij}^t} \sum_{j \in GS} \sum_{i \in U_k} x_{e_{ij}}^t \quad \text{s.t.} \quad (21) - (26),\ (28). \tag{29}$$

Furthermore, if each satellite merely considers its own reward, it may be counterproductive to global optimization. Considering the influence of neighbor satellites, we introduce a neighbor discount factor $\psi \in [0, 1]$. The overall reward of each satellite can be further represented as

$$\tilde{r}_i^t = r_i^t + \psi \sum_{i \in \mathcal{N}_i^t} r_j^t, \tag{30}$$

and the overall reward of a domain can be expressed as

$$\tilde{r}^t = \sum_{i \in U_k} \tilde{r}_i^t. \tag{31}$$

4) Formulation: We consider the sequential decision resource scheduling in the CDC-RMDS problem as an MDP, where the changes in satellite storage resources can be regarded as the state transition in the MDP. Furthermore, in view of the interaction of data transmission between satellites, we formulate the resource scheduling problem in the MDSS as a fully cooperative MARL task for multiple satellites, where each satellite serves as an agent. The satellite decides the data transmission action according to its state and the state of neighbour satellites, to maximize the amount of offloaded data to $\mathcal{GS}$ in the entire MDSS. The mapping from the satellite state to action in t-th time slot is $\mathscr{M}_i^t : \tilde{s}_i^t \rightarrow a_i^t$. The continuous mapping over a period is called a policy, and the policy of satellite i is represented as

$$\pi_i = \{\mathscr{M}_i^1(\tilde{s}_i^1), \cdots, \mathscr{M}_i^T(\tilde{s}_i^T)\}. \tag{32}$$

The scheduling policy of satellite i indicates which of the visible satellites should be selected by the satellite to establish a connection and transmit data. The policy of each satellite in each domain constitutes the scheduling policy of the entire MDSS. For a domain DOM_k, each satellite has its own policy, and the set of strategies of domain DOM_k can be represented as

$$\mathcal{P}_k = \{\pi_1, \cdots, \pi_N\}^k. \tag{33}$$

The total policy set of the MDSS can be expressed as

$$\mathbb{P} = \{\mathcal{P}_1, \mathcal{P}_2, \cdots, \mathcal{P}_K\}. \tag{34}$$

The cumulative expected reward of satellite i can be expressed as

$$\mathscr{R}(\tilde{s}_i^t, \pi_i) = \mathbb{E}\left[\sum_{m=0}^{T-t} \gamma^m \cdot \tilde{r}(\tilde{s}_i^{t+m}, \mathscr{M}_{t+m}^{\pi_i}(\tilde{s}_i^{t+m}))\right]. \tag{35}$$

Here, $\mathbb{E}[\cdot]$ denotes the expectation function and γ represents the discount factor. The total reward of the system is the sum reward of each domain. For each domain the expected reward is

$$\mathcal{R}(DOM_k, \mathcal{P}_k) = \sum_{i=1}^{N} \mathscr{R}(\tilde{s}_i, \pi_i),\ \pi_i \in \mathcal{P}_k. \tag{36}$$

Furthermore, for the MDSS we have

$$\mathbb{R}(\mathcal{D}, \mathbb{P}) = \sum_{k=1}^{K} \mathcal{R}(DOM_k, \mathcal{P}_k),\ \mathcal{P}_k \in \mathbb{P}. \tag{37}$$

The goal of resource scheduling is to find the optimal policy $\mathbb{P}^*$ to maximize the amount of offloaded data in MDSS,

$$\mathbb{P}^* = \arg\max_{\mathbb{P} \in \Pi, \mathbf{s}, \mathbf{a}, \tilde{\mathbf{r}}} \mathbb{R}(\mathcal{D}, \mathbb{P}) \quad \text{s.t.} \quad (21) - (26),\ (28). \tag{38}$$

In a scheduling period, the set of all feasible strategies is Π, which remains unchanged in any period.

IV. Hierarchical Cross-Domain Satellite Resource Scheduling Strategy

In this section, to effectively solve the satellite resource scheduling problem formulated in Section III, we design a hierarchical MDSS resource scheduling algorithm. Firstly, we are concerned with utilizing ISLs for multi-satellite collaboration to promote the volume of offloaded data in a domain. A multi-satellite intra-domain collaboration method based on MARL is proposed in Subsection IV-A. Then, to balance the data between busy and idle domains with limited cross-domain transmission resources and promote MDSS resource utilization, we design a DSNMG-based cross-domain scheduling algorithm in Subsection IV-B. Furthermore, in Subsection IV-C, the intra-domain and cross-domain scheduling processes are embedded in the hierarchical framework.

A. Intra-Domain Multi-Satellite Collaboration Scheduling Method

We develop the MSCS algorithm by adopting the framework with centralized training and decentralized execution. The proximal policy optimization (PPO) is an on-policy reinforcement learning algorithm that optimizes the policy function π_θ with a trust-region method. Due to the random data traffic and the time-varying inter-satellite communication opportunities in MDSS, the training process is unstable. To improve the training stability without reducing the data efficiency, PPO tactfully eliminates the part of the data that makes the network parameters vary drastically by the clip function and limits the update of policy to a certain range [35]. On account of this, PPO is widely applied in resource control and scheduling [36], [37], [38].

We consider a centralized training decentralized execution framework as shown in the top right corner of Fig. 2. The centralized training estimates the joint value function based on global information, while decentralized execution only makes the action decisions based on the local environmental observations of satellites. After the training phase, the agents no longer need global information, and they can choose their actions based on the local information in a decentralized manner. In the top right corner of Fig. 2, the blue part is executed only during the training phase.

1) Actors: Each satellite is configured with an actor network served as a policy function, which is composed of multiple

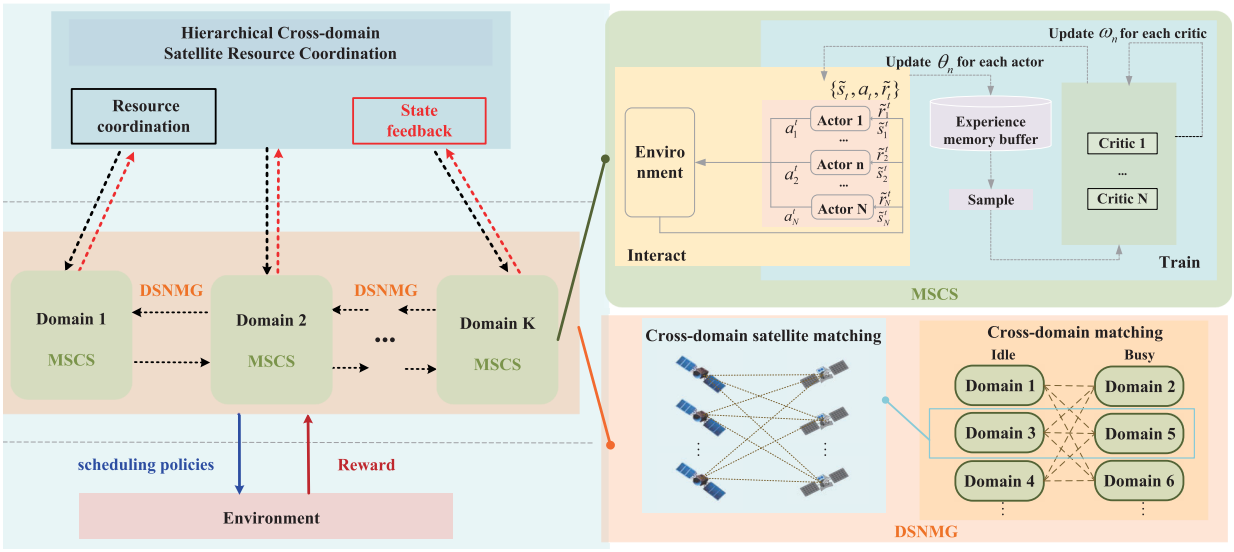


Fig. 2. Overview of our MDSS hierarchical scheduling framework.

layers of perceptron (MLP). For each satellite, we build a stochastic policy function π_{θ^i} with parameter θ^i, such that $\pi_{\theta^i}(a_i, \tilde{s}_i) \in [0, 1]$ offers the probabilistic distribution of choosing the action decision a_i given a state $\tilde{s}_i$. This policy function π_{θ^i} is represented to be a stochastic policy when starting to explore the optimal operational decisions throughout the state and action spaces. The stochasticity of the policy function π_{θ^i} decreases with the training process. The policy function π_{θ^i} is optimized through the learning process to approximate the optimal policy of (36).

The goal of the training phase of proposed MSCS is to maximize the expected discount return, i.e., $\max_{\mathbb{P}} \mathbb{R}(\mathbb{P})$. π_i^{old} denotes the current policy of satellite i, for the updated policy π_i, we have $\mathbb{R}(\pi_i) = \mathbb{R}(\pi_i^{old}) + \frac{1}{1-\gamma}\mathbb{E}_{\pi_i}\{A^{\pi_i^{old}}(\tilde{s}_i^t, a_i^t)\}$, which guarantees a nondecreasing policy update by $\mathbb{E}_{\pi_i}\{A^{\pi_i^{old}}(\tilde{s}_i^t, a_i^t)\} \geq 0$, i.e., $\mathbb{R}(\pi_i) \geq \mathbb{R}(\pi_i^{old})$.

The parameter θ^i can be updated by

$$\Delta\theta^i = \nabla_{\theta^i} E_t\{L(\theta^i)\}, \tag{39}$$

where $L(\theta^i)$ is the loss function and can be defined as:

$$L(\theta^i) = min(\beta(\theta^i)A^{\pi_i^{old}}(\tilde{s}_i^t, a_i^t), clip(\beta(\theta^i), 1-\epsilon, 1+\epsilon)A^{\pi_i^{old}}(\tilde{s}_i^t, a_i^t)), \tag{40}$$

The preset parameter $\epsilon \in (0, 1)$ with the interval $[1-\epsilon, 1+\epsilon]$, which defines the upper bound and the lower bound for $\beta(\theta^i)$ to limit the scope of parameter update. The $\beta(\theta^i)$ can be represented as

$$\beta(\theta^i) = \frac{\pi_{\theta^i}(a_i|\tilde{s}_i)}{\pi_{\theta^i_{old}}(a_i|\tilde{s}_i)}. \tag{41}$$

2) Critics: The critics are also required to update to approximate the value function $Q_{\omega^i}(\tilde{s}_i^t)$, which is employed to estimate the expected total reward. The parameter ω^i is optimized iteratively by

$$\Delta\omega^i = \frac{1}{B}\sum_{t=1}^{B}\left\{\nabla_{\omega^i}(Q_{\omega^i}(\tilde{s}_i^t) - \hat{V}_i^t)^2\right\}, \tag{42}$$

Algorithm 1 MSCS

Input: MDSS network topology obtained by satellite simulator

Output: Trained MSCS policy of actors π_{θ^i}

Initialization: actors π_{θ^i} and critics Q_{ω^i} with θ^i, ω^i; current actors $\pi_{\theta^i}^{old}$ and target critics $Q_{\omega^i}^{old}$ with $\theta_{old}^i \leftarrow \theta^i$, $\omega_{old}^i \leftarrow \omega^i$
for $Iteration = 1, 2, \cdots, L$ **do**
  Let $\left\langle CB_i^t, t^*\right\rangle_{t=0}$ be the initial state of satellite i
  for $j = 0, 1, \cdots, \frac{T}{B} - 1$ **do**
    for $t = 1 + Bj, \cdots, B(j+1)$ **do**
      Each satellite i conducts action according to $\pi_{\theta^i}^{old}(a_i^t|\tilde{s}_i^t)$
      Get the reward $\tilde{r}_i^t$ and $\tilde{r}^t$ through (29)
    end for
    Collect the trajectory of each satellite i, $\varsigma_i = \left\{\tilde{s}_i^t, a_i^t, \tilde{r}_i^t\right\}_{t=1+Bj}^{(j+1)B}$
    Compute $\left\{A^{\pi_i^{old}}(\tilde{s}_i^t, a_i^t)\right\}_{t=1+Bj}^{(j+1)B}$ and $\left\{\hat{V}^t\right\}_{t=1+Bj}^{(j+1)B}$
    for Each satellite i, $i \in 1, 2, \cdots, N$ **do**
      Compute $\Delta\theta^i$ and $\Delta\omega^i$ by (39) and (42)
    end for
  end for
  Update networks parameters for each satellite θ_{old}^i, ω_{old}^i
end for

where

$$\hat{V}_i^t = \tilde{r}_i^{t+1} + \gamma\tilde{r}_i^{t+2} + \cdots + \gamma^{T-t-1}\tilde{r}_i^T. \tag{43}$$

Both the minimization of the actor network (39) and the critic network (42) are iteratively solved by exploiting stochastic gradient search with the Adam optimizer.

The pseudo-code of the MSCS algorithm is presented in Algorithm 1. We construct an actor π_{θ^i} with parameter θ^i and a critic Q_{ω^i} with parameter ω^i for each satellite. The

parameters of the current actors $\pi_{\theta^i}^{old}$ and target critics $Q_{\omega^i}^{old}$ are updated by $\Delta\theta^i$ and $\Delta\omega^i$, respectively. B is the batch size and T_l is the step of the satellite planning period. The algorithm consists of L iterations round. In each iteration, each satellite selects actions based on the policy $\pi(\tilde{s}_i^t)$ and its overall states $\tilde{s}_t^i$, which includes the buffer and link states of its own and neighbor nodes. By solving a linear optimization problem, we can eliminate the actions that do not conform to the actual situations and obtain the reward $\tilde{r}^t$. After that, the state $\tilde{s}_{t+1}^i$ is updated for the next time slot and the new action a_i^{t+1} is obtained. We collect the trajectory $\varsigma_i = \{\tilde{s}_i, a_i, \tilde{r}\}$ of each satellite, which is obtained by satellite i following the policy function $\pi_{\theta^i}^{old}$. Then in each mini-batch, the parameters are updated by the collected trajectory ς_i.

B. Cross-Domain Scheduling Based on Domain-Satellite Nested Matching Game

The matching game provides a solution to combinatorial problems, where two sets of players are assigned to each other based on the two-side preference information. Recently, the matching game has been widely applied to solve resource scheduling owing to its ability to produce mutually satisfying and stable bilateral cooperation [39], [40].

In cross-domain scheduling, we focus on how to maximize the amount of offloaded data in the MDSS by sharing storage and transmission resources across domains. There is resource competition between busy domains and collaboration between busy and idle domains in the MDSS. In addition, the domain consists of multiple satellites, so cross-domain collaboration is not only to determine the collaboration relationship between domains but also to determine the cross-domain satellite collaboration connection. Inspired by the matching game, we regard cross-domain collaboration as a two-stage matching game process. Such collaboration and competition coexist in cross-domain collaboration scheduling and can be viewed as a matching game process. Furthermore, we propose a domain-satellite nested matching game algorithm to solve the matching of busy and idle domains and the collaborative cross-domain connection of satellites in these domains. We construct a matching preference list of cross-domain satellites and then leverage the matching results of cross-domain satellites as the basis for the preference list for cross-domain matching.

We first divide the domain set $\mathcal{D}$ into the idle domain set $\mathcal{D}^I$ and the busy domain set $\mathcal{D}^B$ according to the buffer utilization of satellites, i.e., $\mathcal{D}^I \cup \mathcal{D}^B = \mathcal{D}$ and $\mathcal{D}^I \cap \mathcal{D}^B = \emptyset$. Then we define the cross-domain matching ϕ and the cross-domain satellite matching φ. We point out that these matchings should be one-to-one due to the limitation of the number of onboard transponders. Considering that cross-domain collaboration is discussed in a time slot, for clarity, the time slot t is omitted in the formulation.

Definition 1: The matching ϕ maps from the idle domain set $\mathcal{D}^I = \{DOM_1^I, DOM_2^I, \cdots\}$ to the busy domain set $\mathcal{D}^B = \{DOM_1^B, DOM_2^B, \cdots\}$. The matching ϕ is one-to-one matching, and we call $\phi(DOM_k^I)$ is the match of DOM_k^I, $\phi(DOM_k^I) \in \mathcal{D}^B$. $\phi(DOM_{k'}^B) = DOM_k^I$, if and only if $\phi(DOM_k^I) = DOM_{k'}^B$.

Definition 2: The matching φ maps from the set of idle domain satellite $DOM_k^I = \{u_1^{I,k}, u_2^{I,k}, \cdots, u_N^{I,k}\}$ to the set of busy domain satellite $DOM_k^I = \{u_1^{B,k}, u_2^{B,k}, \cdots, u_N^{B,k}\}$. The matching φ of satellites between domains is one-to-one matching, call $\varphi(u_i^{I,k})$ is a match of $u_i^{I,k}$, then $\varphi(u_i^{I,k})$ is in $DOM_{k'}^B$. $\varphi(u_i^{B,k'}) = u_i^{I,k}$, if and only if $\varphi(u_i^{I,k}) = u_i^{B,k'}$.

The matching of cross-domain satellites in the idle domain and the busy domain is a two-way selection process, according to the preference of each satellite. Satellites in DOM_k^I prefer the satellites in $DOM_{k'}^B$ with close distances, while satellites in $DOM_{k'}^B$ prefer the satellites in DOM_k^I with a large available buffer. $\mathrm{P}(u_i)$ denotes the preference list of satellite u_i. For example, $\mathrm{P}(u_4^{I,k}) = u_2^{B,k'}, u_1^{B,k'}, u_4^{I,k}, \cdots, u_7^{B,k'}$ indicates that the first and second choice of satellite node $u_4^{I,k}$ in the idle domain DOM_k^I are $u_2^{B,k'}$, $u_1^{B,k'}$. The third option is the $u_4^{I,k}$ itself, which indicates the satellite does not establish a match, while the following are the invisible satellite in $DOM_{k'}^B$. $u_2^{B,k'} \underset{u_4^{I,k}}{\succ} u_1^{B,k'}$ means $u_4^{I,k}$ prefers $u_2^{B,k'}$ than $u_1^{B,k'}$. Each busy domain $DOM_{k'}^B$ in $\mathcal{D}^B$ also has a preference $\mathrm{P}(DOM_{k'}^B)$ for the satellite in idle domains. The same goes for idle domains DOM_k^I, and $\mathrm{P}(DOM_k^I)$ is its preference list.

The satellites in DOM_k can be expressed as $\mathcal{U}_k = \{u_1, u_2, \cdots, u_N\}$. A specific multi-domain matching triplet can be expressed as $(\mathcal{D}^I, \mathcal{D}^B, \mathbf{P}(\mathcal{D}))$, and a specific cross-domain satellite matching triplet is denoted as $(DOM_k^I, DOM_{k'}^B, \mathbf{P}(DOM))$. The preference relation $\mathbf{P}$ in triplet $(\mathcal{I}, \mathcal{J}, \mathbf{P})$ is defined as a reflexive, complete and transitive relation between $\mathcal{I}$ and $\mathcal{J}$, which is a set of preference list of each node in $\mathcal{I}$ and $\mathcal{J}$. The preference list of satellites in DOM_k^I is constructed by sorting the available buffer of satellites in $DOM_{k'}^B$ in descending order. The preference list of satellites in $DOM_{k'}^B$ is constructed by sorting the inter-satellite link rate in (4) in descending order. In this way, the outcome of the matching game $(DOM_k^I, DOM_{k'}^B, \mathbf{P}(DOM))$ is produced by the individual desire of satellites. In the cross-domain satellite matching $(DOM_k^I, DOM_{k'}^B, \mathbf{P}(DOM))$, satellites obtain matching proposals by accept-reject procedure in [41] based on preference lists. Further, In cross-domain matching $(\mathcal{D}^I, \mathcal{D}^B, \mathbf{P}(\mathcal{D}))$. The preference lists for $\mathcal{D}^I$ and $\mathcal{D}^B$ are constructed by sorting the number of cross-domain satellite connections in descending order, that is to say, cross-domain matching prefers to connect idle and busy domains that have more inter-satellite connections. For simplicity, we only represent the preference elements in the preference list that are better than not match, e.g., the aforementioned $\mathrm{P}(u_4^{I,k})$ can be written as $\mathrm{P}(u_4^{I,k}) = u_2^{B,k'}, u_1^{B,k'}$.

Definition 3: If there does not exist the blocking pair $(u_i^{I,k}, u_{i'}^{B,k'})$, where $u_i^{I,k} \in DOM_k^I$, $u_{i'}^{B,k'} \in DOM_{k'}^B$, such that $u_{i'}^{B,k'} \underset{u_i^{I,k}}{\succ} \varphi(u_i^{I,k})$ and $u_i^{I,k} \underset{u_{i'}^{B,k'}}{\succ} \varphi(u_{i'}^{B,k'})$ where $\varphi(u_{i'}^{B,k'})$ and $\varphi(u_i^{I,k})$ represent the current matched partners of $u_{i'}^{B,k'}$ and $\varphi(u_i^{I,k})$, respectively, the matching φ is stable. The matching ϕ is stable in the same way.

The specific DSNMG algorithm is listed in Algorithm 2. The cross-domain scheduling is divided into the matching of domains and the matching of cross-domain satellites.

Algorithm 2 DSNMG

Input: The state and network topology of the MDSS $\mathcal{D}$
Output: ϕ, φ, the next state of all satellites in MDSS $\tilde{S}^{t+1}$

1: Initialization: Construct the preference lists of satellites in idle domains and busy domains $\mathbf{P}(u_i^{I,k})$, $\mathbf{P}(u_{i'}^{B,k'})$
2: **for** Each domain $DOM_{k'}^B$ in $\mathcal{D}^B$ **do**
3: **for** Each domain DOM_k^I in $\mathcal{D}^I$ **do**
4: **if** There exists unmatched satellite $u_{i'}^{B,k'}$ in $DOM_{k'}^B$ **then**
5: **for** Each unmatched satellites in $DOM_{k'}^B$ **do**
6: Selects the favorite satellite in DOM_k^I and presents for matching proposal according to $\mathbf{P}(u_{i'}^{B,k'})$
7: **end for**
8: **for** Each satellite in DOM_k^I **do**
9: Establish matching with one of the current favorite satellite based on $\mathbf{P}(u_i^{I,k})$ and reject the other satellite matching
10: **end for**
11: **end if**
12: Save the matching result φ of the current busy-idle-domains pair and the preference of domain.
13: **end for**
14: **end for**
15: Get the preference list $\mathbf{P}(DOM_k)$ of each domain through counting the number of established matching φ
16: **if** There exists unmatched domain $DOM_{k'}^B$ in $\mathcal{D}^B$ **then**
17: **for** Each unmatched satellites in $\mathcal{D}^B$ **do**
18: Selects the favorite domain in $\mathcal{D}^I$ and requests for matching according to $\mathbf{P}(DOM_{k'}^B)$
19: **for** Each domain DOM_k^I in $\mathcal{D}^I$ **do**
20: Establish matching with one of the current favorite domain based on $\mathbf{P}(DOM_k^I)$ and reject the other satellite matching
21: **end for**
22: **end for**
23: **end if**
24: Get the cross-domain matching ϕ and the corresponding inter-satellite matching φ
25: Update the state of the MDSS $\tilde{S}^t \rightarrow \tilde{S}^{t+1}$

Specifically, during the matching of cross-domain satellites (Algorithm 2 steps 2-14), we match the satellites within the pair of busy-idle domains according to their preferences. Then, the satellite matching results for this pair of busy-idle domains are saved, and we traverse over all the busy-idle domain matching pairs. We adopt the number of inter-satellite matches within each matching pair as the preference for cross-domain matching to determine the matching among domains.

Theorem 1: The proposed DSNMG can obtain stable and weak Pareto optimal matching results among domains and among cross-domain satellites within a limited number of iterations.

Proof. In Algorithm 2, referring to the proof process of [41] and [42], the loop in steps 4-11 stops when there is no unmatched satellite. In each iteration of steps 5-10, satellite $u_i^{I,k}$ is matched to a satellite in domain $DOM_{k'}^B$ by an accept-reject procedure based on its preference list. When a matched satellite $u_{i'}^{B,k'}$ in $DOM_{k'}^B$ is deleted and a prioritized satellite in DOM_k^I is matched to a better satellite in $DOM_{k'}^B$ in step 9. Since the number of satellites in a domain and the elements in the preference list is finite, the iteration terminates in limited steps. Then the ultimate stable matching of Algorithm 2 in steps 4-11 is proved by contradiction. We assume that for a satellite $u_{i'}^{B,k'}$ in $DOM_{k'}^B$, there is $\tilde{\varphi}$ that satisfies $\tilde{\varphi}(u_{i'}^{B,k'}) \underset{u_{i'}^{B,k'}}{\succ} \varphi(u_{i'}^{B,k'})$. It means that $u_{i'}^{B,k'}$ can be matched to a better potential satellite in DOM_k^I under $\tilde{\varphi}$ compared to φ. Algorithm 2 will not terminate in such cases according to Definition 3, which contradicts the assumption. Accordingly, we can conclude that the resulting φ is stable and weak Pareto optimal for each satellite in $DOM_{k'}^B$. In the same way, the matching of domains ϕ can also be proven to be stable (steps 16-24). Therefore, the matching result of Algorithm 2 is stable and weak Pareto optimal. □

C. Intra-Domain and Cross-Domain Hierarchical Resource Scheduling

After investigating the intra-domain multi-agent collaborative RMDS and the multi-domain collaborative RMDS, we combine these two proposed algorithms MSCS and DSNMG through the hierarchical framework to obtain the DSNMG-MSCS. As shown in Fig. 2, the left side is the structure of the MDSS hierarchical cross-domain resource management, and the right side is the schematic diagram of intra-domain and cross-domain scheduling, respectively.

The pseudo-code for the details is presented in Algorithm 3, and the hierarchical DSNMG-MSCS structure is divided into two layers. Each domain in the bottom layer is independent of other domains, and the satellites within the domain are scheduled through the MSCS collaborative resource scheduling to ensure the maximization of the cumulative expected gain within the domain. At the top layer, we design the DSNMG algorithm for collaborating inter-domain data. Step 9 of Algorithm 3 is performed for each domain to obtain the cooperative pairs between domains and determine the specific inter-satellite connection choices in the cooperative pairs. We adjust the amount of mission data in each domain through cross-domain link data scheduling to alleviate the resource mismatch between domains and improve the amount of offloaded data for the MDSS.

Intuitively, the data transmission of cross-domain affects the policy training of the multi-agent reinforcement learning based MSCS method within the domain. To ensure the effectiveness of the method, we extend the action space to include a virtual action a_{vir}. When the satellite executes the cross-domain transmission, the intra-domain transmission should not be conducted. Therefore, the intra-domain action of the satellite is defined as a_{vir}. We append the triplet $\{\tilde{s}_i^t, a_{vir}, \tilde{r}_i^t\}$ to the trajectory ς_i.

The implementation of the proposed algorithm can be described as follows. First, the NOCC on the ground senses the satellite states (including buffer states, and data arrival

Algorithm 3 DSNMG-MSCS

Input: Network topology of the MDSS obtained by satellite simulator

Output: Trained scheduling policy π_{θ^i} for each actor in the MDSS

Initialization: Actors π_{θ^i} and critics Q_{ω^i} with θ^i, ω^i; current actors $\pi_{\theta^i}^{old}$ and target critics $Q_{\omega^i}^{old}$ with θ_{old}^i and ω_{old}^i for each domain
for $Iteration = 1, 2, \cdots, L$ **do**
  Let $\left\langle CB_i^t, t^* \right\rangle_{t=0}$ be the initial state of satellite i
  for $j = 0, 1, \cdots, \frac{T}{B} - 1$ **do**
    for $t = 1 + Bj, \cdots, B(j+1)$ **do**
      if All domains are idle or busy **then**
        Each satellite i executes action according to $\pi_{\theta^i}^{old}(a_i^t|\tilde{s}_i^t)$
      else
        Execute **Algorithm. 2** for each domain
        $a_i^t = a_{vir}$
      end if
      Get the reward $\tilde{r}_i^t$, and the next state $\tilde{s}_{t+1}$
    end for
    Execute **Algorithm 1 steps** 9∼ 15
  end for
  Each domain update target networks $\theta_{old}^i \leftarrow \theta^i$, $\omega_{old}^i \leftarrow \omega^i$
end for

TABLE II
MAIN SIMULATION PARAMETERS

Parameter	Value	Parameter	Value
Satellites per domain	10	P_{itr}	80W
Orbit period	5744.27s	$\mathcal{T}$	2h
Inclination	86.4deg	T_s	300K
Satellite altitude [20]	554.8km	τ	300s
The number of domain	1,2,3,4,5	M [1]	1.85
G_{jre}	16 dB	f	6.45GHz
ϵ [37]	0.2	B	32
T_l	288	ψ	0.2
learning rate	10^{-4}	γ	0.99

information) by collecting satellite state information. After that, the NOCC conducts centralized training according to the collected historical state information to obtain the policy of each satellite [43]. Then, the trained satellite scheduling policies are uploaded to each satellite. The satellites collaboratively perform data scheduling according to the current network status and policies.

V. Simulation Results and Discussion

In this section, we perform a series of simulations in the MDSS with the satellite parameters from the satellite simulator. First, the detailed settings of the simulation scenario and the algorithm parameters are presented. Then the effectiveness of the proposed intra-domain collaboration and cross-domain collaboration methods are verified.

A. Simulation Configuration

Satellite scenario parameters: We conduct simulations in the MDSS with five different domains, each of which contains 10 satellites. There are 50 satellites distributed at a height of 554.8km and an inclination of 86.4° with the right ascension of the ascending node at 45°, 75°, 105°, 135°, and 165°, respectively. Three GSs are located at Beijing (40°N,116°E), Kashi (39.5°N,76°E), and Sanya (18°N,109.5°E), respectively. We build up such a MDSS network with the above parameters based on Matlab and a satellite simulator. The simulation scheduling horizon ranges from 22 Jun. 2021 04:00:00 to 23 Jun. 2021 04:00:00 with a duration of 300 seconds as a time slot. The radio frequency is 6.45GHz (C-band). The learning rate of the Adam optimizer is 10^{-4}, and the discount factor γ is 0.99. The clip range ϵ is 0.2 [37]. The actor networks and critic networks are parameterized by constructing the Rectified Linear Unit MLP with 16 and 256 units respectively. The maximum step T_l of each iteration is 288, and the minibatch size B is 32. The main parameters operated in the simulations are listed in Table II. The establishment of the multi-domain satellite environment and the simulation of the algorithm program are operated in python 3.8. We utilize Pytorch with version 1.7.1 to build and train the neural networks.

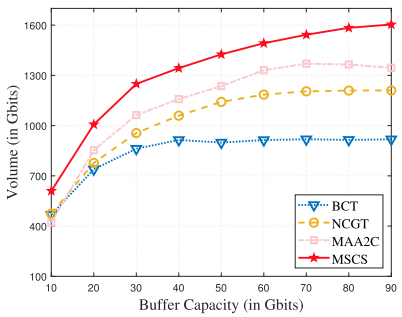
B. Simulation Benchmark Design

To evaluate the performance of the proposed MSCS algorithm, we mainly consider the following three benchmark scheduling schemes (i.e., the Non-collaborative greedy transmission scheme, the blind collaboration transmission scheme, and the MAA2C-based collaboration transmission scheme). Then we evaluate their performance under the same conditions.

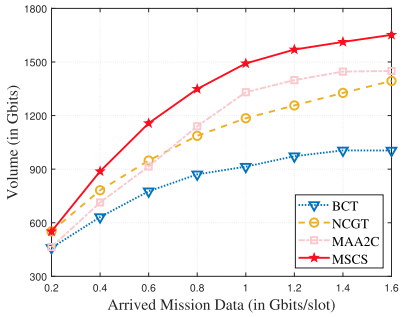
- **Non-collaborative greedy transmission scheme (NCGT).** The NCGT myopically determines the transmission connection for each satellite to offload the maximum amount of data at each slot. Therefore, the NCGT is going to greedily offload as much data as possible to the GS in each communication opportunity. This means that future rewards are ignored in the NCGT scheme.
- **Blind collaboration transmission scheme (BCT).** The BCT performs satellite resource scheduling actions randomly, and satellites select feasible connections with the same probability, which means that collaboration among satellites is unguided and blind in the BCT.
- **MAA2C.** In the training process of MAA2C, each satellite is treated as an agent. To ensure the fairness of the comparison, the MAA2C applies the same definitions of state, action, and reward in Section III-B.

Furthermore, to verify the performance of the proposed DSNMG-MSCS algorithm in the MDSS, we demonstrate simulation results by comparing our work with the following benchmark algorithms.

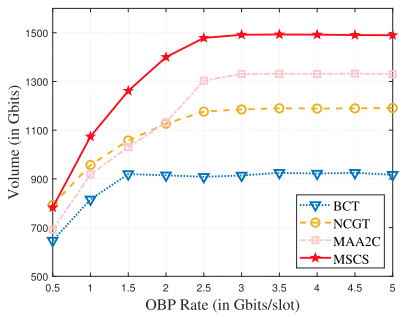
- **Domain isolation (DI).** In the MDSS, domains are isolated from each other and do not collaborate.
- **Buffer oriented matching (BOM).** Cross-domain collaboration by matching satellites with the largest receiving satellite buffer.



(a) Network performance of different schemes with varying buffer capacity $CB_{\max}$ (ρ=3, λ=1).



(b) Network performance of different schemes with varying arrived mission data λ ($CB_{\max}$=60Gbits, ρ=3).



(c) Network performance of different schemes with varying OBP rate ρ ($CB_{\max}$=60Gbits, λ=1).

Fig. 3. The intra-domain RMDS performance.

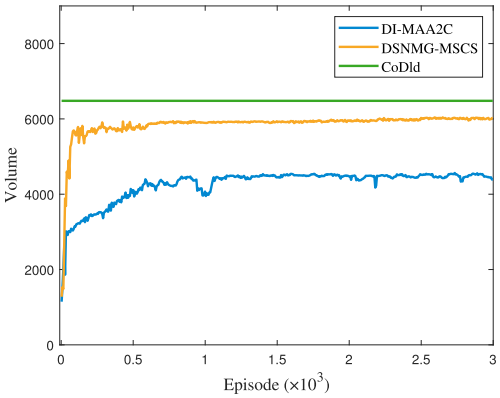


Fig. 4. The volume of offloaded data in the training process ($CB_{\max}$=60Gbits, ρ=3, λ = 2.1).

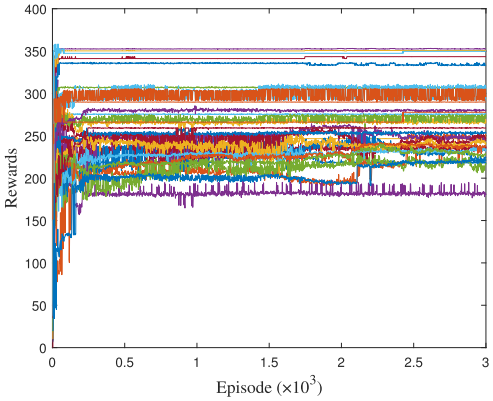


Fig. 5. The reward of each satellite in DSNMG-MSCS ($CB_{\max}$=60Gbits, ρ=3, λ = 2.1).

- **Link rate oriented matching (LROM).** Cross-domain collaboration by matching satellites with the highest ISL rate.

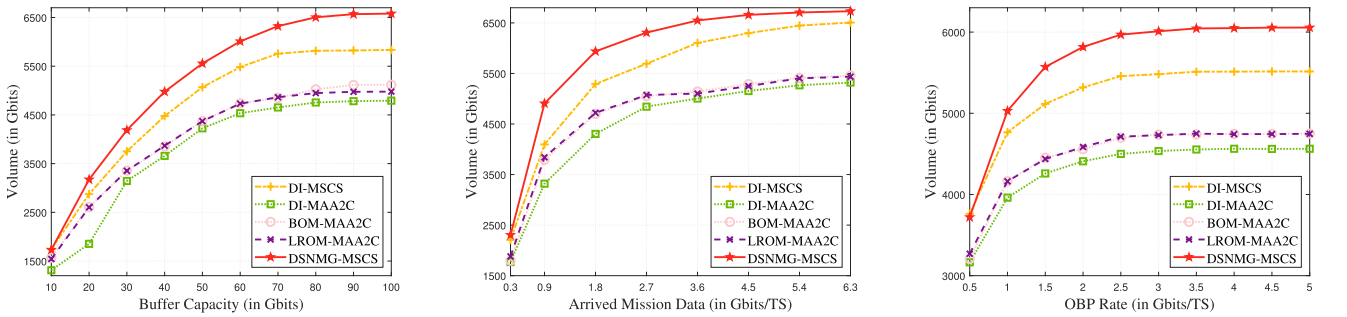
C. Performance of Intra-Domain Satellite Scheduling

In this simulation, we evaluate the network performance of a domain with 10 satellites under different network parameters, i.e., satellite buffer, arrived mission data, and OBP rate.

We investigate the impacts of the buffer capacity on the volume of offloaded data, as shown in Fig. 3a. We can see that the offloaded data of the four schemes show an upward trend as the buffer capacity increases at the beginning. This is because a larger buffer can store more data traffic, to increase the probability of successfully offloading data in the future. It can also be noted that the amount of offloaded data does not increase indefinitely as buffer capacity increases, this is due to the fact that other resources such as transmission resources and OBP resources become bottlenecks. Notably, since the proposed MSCS scheme can coordinate the resources of multiple satellites, it is superior to the BCT, NCGT, the MAA2C in terms of the volume of offloaded data.

In Fig. 3b, we investigate the impact of arrived mission data on the amount of offloaded data under the four schemes. We can notice that the amount of offloaded data in the domain increases significantly as the λ increases before the offloaded data increases to a certain level. However, the growth reaches a bottleneck due to the limitation of other resources, such as storage resources and transmission resources. In terms of the maximum amount of offloaded data, the proposed method improves 64.4%, 18.5%, and 14% compared to the BCT, the NCGT, and the MAA2C, respectively. We attribute this to the efficient inter-satellite collaboration of the MSCS to accomplish more offloaded data when the arrived mission data is large.

As shown in Fig. 3c, with the OBP rate ρ increasing, the total amount of offloaded data of the four schemes shows an increasing trend. In the case of low a OBP rate, the increase in the OBP rate significantly promotes the amount of offloaded data since all the processed data can be offloaded. When the OBP rate is large enough, the amount of offloaded data increases slowly as the OBP rate increases and eventually becomes stable. In this case, other resources become bottlenecks. Additionally, it can be found that the proposed algorithm is superior to the NCGT and the BCT in terms of the volume of offloaded data, which is because the NCGT and the BCT do not utilize the ISLs efficiently. The proposed MSCS scheme can balance the resources of satellites by ISLs, which in turn improves resource utilization and increases the volume of offloaded data in the future.



(a) Network performance of different schemes with varying buffer capacity $CB_{\max}$ (ρ=3, λ=2.1).

(b) Network performance of different schemes with varying arrived mission data λ ($CB_{\max}$=60Gbits, ρ=3).

(c) Network performance of different schemes with varying OBP rate ρ ($CB_{\max}$=60Gbits, λ=2.1).

Fig. 6. The performance of cross-domain RMDS (three domains, the percentage of arrived data in busy and idle domains is 80% and 20%, respectively).

D. Performance of Cross-Domain Satellite Scheduling

In the following experiments, we investigate the performance of different algorithms in the MDSS with different network parameters, including the buffer, OBP rate, arrived mission data, and the number of domains.

To verify that the algorithm can converge to an effective solution, we compare the proposed algorithm with the CoDld in which data information is supposed to be fully known. In some conditions, the CoDld can find the optimal solution [26]. It can be seen from Fig. 4 that the DSNMG-MSCS is closer to the upper bound (CoDld) and the training processing is more stable. Fig. 5 shows the reward for each satellite in DSNMG-MSCS. As the training process proceeds, the rewards of each satellite fluctuate and then stabilize. It can be noticed that the rewards of some satellites drop to maximize the reward of the entire system.

In Fig. 6a, we investigate the effect of satellite buffer on the amount of offloaded data under different algorithms. It is noteworthy that the difference in the amount of offloaded data between the proposed DSNMG-MSCS algorithm and the DI-MSCS algorithm becomes larger with the increase in the buffer. This is because a larger buffer provides more options for data forwarding across domains, which in turn enables better cross-domain decisions. In particular, the DSNMG-MSCS algorithm has a performance improvement of 12.7%, 28.6%, 32.2%, and 37.3% compared to the DI-MSCS, BOM-MAA2C, LORM-MAA2C, DI-MAA2C, respectively, when the buffer capacity $CB_{\max}$ = 100Gbits. This is attributed to the efficient intra-domain collaborative scheduling and the joint consideration of resource coordination across multiple domains.

We can see from Fig. 6b that the amount of offloaded data in the algorithms gradually increases as the arrived mission data increases. We can find that as λ increases, the difference between the volume of offloaded data of the cross-domain approach compared to the non-cross-domain approach increases at first and then decreases. This is predictable because when the λ is low, the resources of each domain are relatively abundant, and the performance gap between the volume of offloaded data in cross-domain and non-cross-domain is not noticeable. However, since cross-domain scheduling can better guide and coordinate the resources of multiple domains, the gap will gradually become significant as the λ increases. When the λ increases to a certain level, this gap decreases as the load on each domain become heavier.

Next, in Fig. 6c we investigate the effect of the OBP rate on the performance of the offloaded data volume. Predictably, since OBP resources are not shared across domains, with the increase in the OBP rate, the convergence trend of the five algorithms is the same. The offloaded data volume tends to increase for all algorithms as the OBP rate increases, due to the fact that the increase in the OBP rate makes it possible to process an increased amount of data that can be transmitted and offloaded by inter-satellite collaboration. It is worth noting that the gain gradually plateaus when the OBP rate is sufficiently large, due to bottlenecks in other resource constraints (e.g., transmission, buffer). In particular, when the trend of performance curves is steady, the proposed method is superior to the benchmark algorithms. This is because of the efficient intra-domain and cross-domain collaborative scheduling. The DSNMG-MSCS shows a performance improvement of 9.8%, 27.5%, 27.6%, and 32.8% compared to the DI-MSCS, BOM-MAA2C, LROM-MAA2C, DI-MAA2C, respectively.

We further investigate the impact of the number of domains on the MDSS. It can be seen from Fig. 7 that the amount of offloaded data gradually increases with the increase of the number of domains. We can notice that the DI-MSCS outperforms the DI-MAA2C because the proposed MSCS algorithm can obtain a better intra-domain scheduling policy to accomplish more data offloading compared to the MAA2C. The DSNMG-MSCS can further improve the performance of offloaded data volume as the number of domains increases compared to DI-MSCS. We attribute it to the proposed domain-satellite nested matching game cross-domain collaboration approach, which can perform effective cross-domain satellite scheduling and balance the data of different domain missions by referring to the buffer and link information of different domains.

To further investigate the relationship between performance gain and resource utilization, Fig. 8 depicts the utilization of buffer, OBP, and transmission resources for different methods. The average storage resource utilization of the three methods is

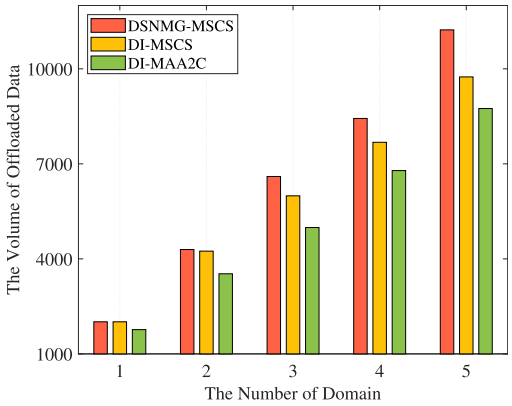


Fig. 7. Performance versus the number of domain ($CB_{\max}$=60Gbits, ρ=3, λ=4.5, the percentage of arrived data in busy and idle domains is 80% and 20%, respectively).

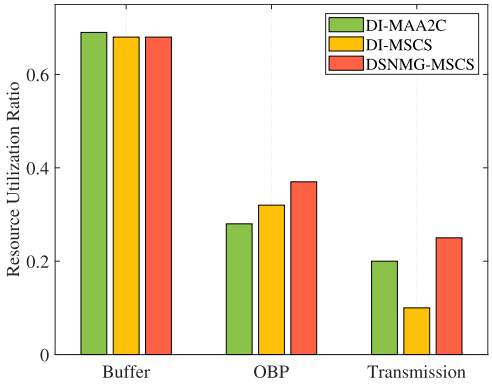


Fig. 8. Resource utilization ratio of different resources (three domains, $CB_{\max}$=60Gbits, ρ=3, λ=2.1, the percentage of arrived data in busy and idle domains is 80% and 20%, respectively).

almost the same. However, the DI-MAA2C and the DI-MSCS are worse than the DSNMG-MSCS in terms of the amount of offloaded data in Fig. 6. This is because more data in the DI-MAA2C and the DI-MSCS algorithm are trapped in the buffer. OBP resources are consumed when there is unprocessed data in the satellite buffer. The proposed DSNMG-MSCS scheme has higher resource OBP utilization than other algorithms, under its efficient coordination of multi-satellite multi-domain buffer and transmission resources, which in turn storages and processes more data. However, since OBP resources are not shared between domains, resource utilization improvement is limited. It can be intuitively seen that the cross-domain scheduling approach can improve transmission resource utilization, which means the data are more likely to be forwarded and offloaded. It is worth noting that the transmission utilization of the DI-MAA2C is higher than that of the DI-MSCS. Combined with Fig. 6, we can conclude that the inefficient inter-satellite collaboration of the DI-MAA2C results in the consumption of transmission resources but less volume of offloaded data.

VI. Conclusion

In this paper, we investigate the CDC-RMDS problem in the MDSS from a hierarchical collaboration perspective (i.e., intra-domain and cross-domain) to improve the capability of data offloading. Targeting the intra-domain satellite collaboration, we model the multi-satellite collaborative RMDS problem as an action-state-reward driven MDP process and propose the MSCS algorithm to achieve efficient collaboration of intra-domain satellites. For cross-domain, the DSNMG is proposed to collaborate the resources of multiple domains to address the resource imbalance between domains and further enhance the MDSS scheduling capability. The simulation results show that our proposed algorithm improves 64.4% compared to the non-collaboration approach and about 12.7% in comparison to the non-cross-domain algorithm in terms of the amount of offloaded data. We also investigate and analyze the impact of some network parameters such as buffer capacity and OBP rate on network performance, which can guide further research in the field of MDSS resource management.

References

[1] J. N. Pelton, S. Madry, and S. Camacho-Lara, *Handbook of Satellite Applications*. New York, NY, USA: Springer, 2017.

[2] O. Kodheli et al., "Satellite communications in the new space era: A survey and future challenges," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 1, pp. 70–109, 1st Quart., 2021.

[3] J. Liu, Y. Shi, Z. M. Fadlullah, and N. Kato, "Space-air-ground integrated network: A survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2714–2741, 4th Quart., 2018.

[4] Y. Wang et al., "Multi-resource coordinate scheduling for earth observation in space information networks," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 2, pp. 268–279, Feb. 2018.

[5] N. Saeed, A. Elzanaty, H. Almorad, H. Dahrouj, T. Y. Al-Naffouri, and M.-S. Alouini, "CubeSat communications: Recent advances and future challenges," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1839–1862, 3rd Quart., 2020.

[6] A. Kak and I. F. Akyildiz, "Designing large-scale constellations for the Internet of Space Things with CubeSats," *IEEE Internet Things J.*, vol. 8, no. 3, pp. 1749–1768, Feb. 2021.

[7] I. F. Akyildiz and A. Kak, "The Internet of Space Things/CubeSats," *IEEE Netw.*, vol. 33, no. 5, pp. 212–218, Sep. 2019.

[8] F. Song, Y.-T. Zhou, L. Chang, and H.-K. Zhang, "Modeling space-terrestrial integrated networks with smart collaborative theory," *IEEE Network*, vol. 33, no. 1, pp. 51–57, Jan./Feb. 2019.

[9] S. Mao, S. He, and J. Wu, "Joint UAV position optimization and resource scheduling in space-air-ground integrated networks with mixed cloud-edge computing," *IEEE Syst. J.*, vol. 15, no. 3, pp. 3992–4002, Sep. 2021.

[10] B. Deng, C. Jiang, H. Yao, S. Guo, and S. Zhao, "The next generation heterogeneous satellite communication networks: Integration of resource management and deep reinforcement learning," *IEEE Wireless Commun.*, vol. 27, no. 2, pp. 105–111, Apr. 2020.

[11] I. Sanad and D. G. Michelson, "A framework for heterogeneous satellite constellation design for rapid response earth observations," in *Proc. IEEE Aerosp. Conf.*, Jul. 2019, pp. 1–10.

[12] Y. Yang, X. Chen, R. Tan, and Y. Xiao, "Cross-domain resource management frameworks," in *Intelligent IoT for the Digital World: Incorporating 5G Communications and Fog/Edge Computing Technologies*. Hoboken, NJ, USA: Wiley, 2021, pp. 97–148.

[13] B. Li, Z. Fei, C. Zhou, and Y. Zhang, "Physical-layer security in space information networks: A survey," *IEEE Internet Things J.*, vol. 7, no. 1, pp. 33–52, Jan. 2020.

[14] H. Wu and J. Yan, "QoS provisioning in space information networks: Applications, challenges, architectures, and solutions," *IEEE Netw.*, vol. 35, no. 4, pp. 58–65, Jul. 2021.

[15] M. Centenaro, C. E. Costa, F. Granelli, C. Sacchi, and L. Vangelista, "A survey on technologies, standards and open challenges in satellite IoT," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 3, pp. 1693–1720, 3rd Quart., 2021.

[16] B. Mao, F. Tang, Y. Kawamoto, and N. Kato, "Optimizing computation offloading in satellite-UAV-served 6G IoT: A deep learning approach," *IEEE Netw.*, vol. 35, no. 4, pp. 102–108, Jul. 2021.

[17] D. Zhou, M. Sheng, J. Wu, J. Li, and Z. Han, "Gateway placement in integrated satellite–terrestrial networks: Supporting communications and Internet of Remote Things," *IEEE Internet Things J.*, vol. 9, no. 6, pp. 4421–4434, Mar. 2022.

[18] M. Sheng, Y. Wang, J. Li, R. Liu, D. Zhou, and L. He, "Toward a flexible and reconfigurable broadband satellite network: Resource management architecture and strategies," *IEEE Wireless Commun.*, vol. 24, no. 4, pp. 127–133, Aug. 2017.

[19] M. Zhang and W. Zhou, "Energy-efficient collaborative data downloading by using inter-satellite offloading," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–6.

[20] D. Zhou, M. Sheng, J. Luo, R. Liu, J. Li, and Z. Han, "Collaborative data scheduling with joint forward and backward induction in small satellite networks," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3443–3456, May 2019.

[21] S. Zhang, G. Cui, and W. Wang, "Joint data downloading and resource management for small satellite cluster networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 1, pp. 887–901, Jan. 2022.

[22] I. Leyva-Mayorga, B. Soret, and P. Popovski, "Inter-plane inter-satellite connectivity in dense LEO constellations," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3430–3443, Jun. 2021.

[23] S. Mao and Y. Zhang, "Aerial edge computing for 6G," *J. China Univ. Posts Telecommun.*, vol. 29, no. 1, pp. 50–63, 2022.

[24] S. Fu, J. Gao, and L. Zhao, "Collaborative multi-resource allocation in terrestrial-satellite network towards 6G," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7057–7071, Nov. 2021.

[25] S. Tani, M. Hayama, H. Nishiyama, N. Kato, K. Motoyoshi, and A. Okamura, "Multi-carrier relaying for successive data transfer in earth observation satellite constellations," in *Proc. IEEE Global Commun. Conf.*, Dec. 2017, pp. 1–5.

[26] X. Jia, T. Lv, F. He, and H. Huang, "Collaborative data downloading by using inter-satellite links in LEO satellite networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1523–1532, Mar. 2017.

[27] D. Zhou, M. Sheng, B. Li, J. Li, and Z. Han, "Distributionally robust planning for data delivery in distributed satellite cluster network," *IEEE Trans. Wireless Commun.*, vol. 18, no. 7, pp. 3642–3657, Jul. 2019.

[28] L. Chen, F. Tang, Z. Li, L. T. Yang, J. Yu, and B. Yao, "Time-varying resource graph based resource model for space-terrestrial integrated networks," in *Proc. IEEE INFOCOM*, May 2021, pp. 1–10.

[29] W. Lin, Z. Deng, Q. Fang, N. Li, and K. Han, "A new satellite communication bandwidth allocation combined services model and network performance optimization," *Int. J. Satell. Commun. Netw.*, vol. 35, no. 3, pp. 263–277, May 2017.

[30] Y. Zhu, M. Sheng, J. Li, and R. Liu, "Performance analysis of intermittent satellite links with time-limited queuing model," *IEEE Commun. Lett.*, vol. 22, no. 11, pp. 2282–2285, Nov. 2018.

[31] Z. Gao, A. Liu, C. Han, and X. Liang, "Max completion time optimization for Internet of Things in LEO satellite-terrestrial integrated networks," *IEEE Internet Things J.*, vol. 8, no. 12, pp. 9981–9994, Jun. 2021.

[32] D. Zhou, M. Sheng, R. Liu, Y. Wang, and J. Li, "Channel-aware mission scheduling in broadband data relay satellite networks," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 5, pp. 1052–1064, May 2018.

[33] D. Zhou, M. Sheng, X. Wang, C. Xu, R. Liu, and J. Li, "Mission aware contact plan design in resource-limited small satellite networks," *IEEE Trans. Commun.*, vol. 65, no. 6, pp. 2451–2466, Jun. 2017.

[34] T. Chu, J. Wang, L. Codecà, and Z. Li, "Multi-agent deep reinforcement learning for large-scale traffic signal control," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 3, pp. 1086–1095, Mar. 2020.

[35] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *CoRR*, vol. abs/1707.06347, pp. 1–12, Jul. 2017.

[36] T. Zhang, K. Zhu, J. Wang, and Z. Han, "Cost-efficient beam management and resource allocation in millimeter wave backhaul HetNets with hybrid energy supply," *IEEE Trans. Wireless Commun.*, vol. 21, no. 5, pp. 3291–3306, May 2022.

[37] D. Guo, L. Tang, X. Zhang, and Y.-C. Liang, "Joint optimization of handover control and power allocation based on multi-agent deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13124–13138, Nov. 2020.

[38] C.-S. Ying, A. H. F. Chow, Y.-H. Wang, and K.-S. Chin, "Adaptive metro service schedule and train composition with a proximal policy optimization approach based on deep reinforcement learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 6895–6906, Jul. 2022.

[39] Y. Gu, Y. Zhang, M. Pan, and Z. Han, "Matching and cheating in device to device communications underlying cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 10, pp. 2156–2166, Oct. 2015.

[40] Y. Yuan, T. Yang, H. Feng, and B. Hu, "Learning for matching game in cooperative D2D communication with incomplete information," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 7174–7178, Jul. 2019.

[41] A. E. Roth and M. Sotomayor, "Two-sided matching," in *Handbook of Game Theory With Economic Applications*, vol. 1. 1992, pp. 485–541.

[42] D. Wu, L. Zhou, Y. Cai, H.-C. Chao, and Y. Qian, "Physical–social-aware D2D content sharing networks: A provider–demander matching game," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7538–7549, Aug. 2018.

[43] Y. Shi, Y. Cao, J. Liu, and N. Kato, "A cross-domain SDN architecture for multi-layered space-terrestrial integrated networks," *IEEE Netw.*, vol. 33, no. 1, pp. 29–35, Jan./Feb. 2019.

Hongmei He received the B.E. degree in communication engineering from Xidian University, Xi'an, China, in 2019, where she is currently pursuing the Ph.D. degree in communication and information systems. Her research interests include dynamic resource allocation and mission planning in satellite networks.

Di Zhou (Member, IEEE) received the B.E. and Ph.D. degrees in communication and information systems from Xidian University, Xi'an, China, in 2013 and 2019, respectively. She was a Visiting Ph.D. Student with the Department of Electrical and Computer Engineering, University of Houston, from 2017 to 2018. She is currently an Associate Professor with the State Key Laboratory of Integrated Service Networks, Xidian University. Her research interests include dynamic resource allocation, mission planning, performance evaluation in space-terrestrial integration networks, and space information networks.

Min Sheng (Senior Member, IEEE) joined Xidian University in 2000, where she is currently a Full Professor and the Director of the State Key Laboratory of Integrated Services Networks. She has published over 200 refereed papers in international leading journals and key conferences in the area of wireless communications and networking. Her current research interests include space-terrestrial integration networks, intelligent wireless networks, and mobile ad hoc networks. She received the China National Funds for Distinguished Young Scientists in 2018. She is the Vice Chair of IEEE Xi'an Section. She is an Editor of IEEE COMMUNICATIONS LETTERS and IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.

Jiandong Li (Fellow, IEEE) received the B.E., M.S., and Ph.D. degrees in communications engineering from Xidian University, Xi'an, China, in 1982, 1985, and 1991, respectively. Since 1985, he has been a Faculty Member with the School of Telecommunications Engineering, Xidian University, where he is currently a Professor and the Vice Director of the Academic Committee of State Key Laboratory of Integrated Service Networks. He was a Visiting Professor at the Department of Electrical and Computer Engineering, Cornell University, from 2002 to 2003. His major research interests include wireless communication theory, cognitive radio, and signal processing. He was awarded as a Distinguished Young Researcher from NSFC and a Changjiang Scholar from the Ministry of Education, China. He served as the General Vice Chair for ChinaCom 2009 and the TPC Chair of IEEE ICCC 2013.