# Enhancing Earth Observation Throughput Using Inter-Satellite Communication

Peng Wang, Hongyan Li, *Member, IEEE*, Binbin Chen, *Member, IEEE*, and Shun Zhang, *Senior Member, IEEE*

***Abstract*— Earth observation systems play important roles in many critical applications. The rapid increase of the number of satellites and their sensing capability, however, makes it challenging to send the massive amount of observed data back to the Earth. One promising direction to enhance the earth observation throughput is to use inter-satellite communication. Towards this, we identify two key design factors: 1) the capability to support on-demand scheduling of inter-satellite communication; and 2) the capability to co-optimize the scheduling of observation and transmission missions. For both, rigorous study is needed to determine whether they provide sufficient throughput gain to justify their additional complexity. Our work formulates a generic earth observation and transmission problem to study the maximum network throughput under different settings. By succinctly modeling the different constraints using a generalized time-varying graph representation, we can efficiently find the optimal scheduling solutions. We conduct an extensive study, which shows that using 40 relay satellites from the "starlink" constellation can increase the throughput of 10 sensing satellites from the "Gaofen" constellation by more than 400%. In particular, on-demand scheduling under heavy load and co-optimization of observation/transmission under light but time-critical load can improve the throughput by more than 180% and 100%, respectively.**

***Index Terms*— Satellite network, earth observation, inter-satellite communication scheduling, joint earth observation and transmission.**

## I. Introduction

EARTH observation systems gather various information about our planet Earth using remote-sensing satellites from orbit. They have been used in many important applications, including agriculture monitoring, mapping, weather forecasting, and emergency observation of earthquake and hurricane areas [1]. According to the open UCS satellite database [2], there are around 900 operational Earth observation satellites in orbit at the end of 2020, most of which are low-Earth orbit (LEO) satellites in either polar or near-polar orbits. A variety of observation technologies are in use today, including both passive and active observation solutions, ranging from optical, radar (radio detection and ranging), LiDAR (Laser Imaging Detection and Ranging), to multi-spectral/hyper-spectral imaging [3], whose spatial resolution has also increased from tens of meters to decimeters [4].

With the rapid increase of the number of observation satellites and their sensing capability, it has become increasingly a challenge to send back the massive amount of Earth observation data from the satellites to the ground stations. It was reported that year 2015 already saw 27.4 Terabyte (TB)/day of data [5]. In the year of 2019, the volume of data produced by only Landsat-7, Landsat-8, MODIS (Terra and Aqua units), and Sentinel satellites (Sentinel-1, -2 and -3) has reached 5 Petabyte (PB) [6], and Landsat-8 itself can generate 0.7 TB image data per day with full observation capacity [7]. Most observation satellites in operation today can only transmit their data directly to ground stations. However, the total contact period every day between a LEO satellite and any given ground station can be as short as around 1 hour. Deploying more ground stations at different locations can increase the total contact periods. However, as it is extremely costly (if not totally impossible) to setup ground stations in open ocean (70% of the earth's surface), deserts, or foreign territories, it is difficult to increase the total contact period between a given satellite and all accessible ground stations beyond a certain threshold (e.g., 30% of time). As a result, current Earth observation system faces huge pressure on meeting the increasing data transmission demands with their limited capability.

One promising direction to enhance the earth observation throughput is to use inter-satellite communication. Indeed, there are more than 3000 communications satellites in orbits by the end of May 2021, out of which more than 1700 are "starlink" satellites [2], [8]. In fact, for "starlink" system alone, it may consist of dozens of thousands of satellites in near future. Furthermore, the inter-satellite communication technology has become increasingly mature. High-throughput

Manuscript received 2 August 2021; revised 29 December 2021; accepted 22 March 2022. Date of publication 6 April 2022; date of current version 11 October 2022. The work of Peng Wang, Hongyan Li, and Shun Zhang was supported by the National Natural Science Foundation of China under Grant 61871456. The work of Binbin Chen was supported by the Singapore University of Technology and Design (SUTD) Start-Up Research Grant (SRG) under Award SRG ISTD 2020 157. The associate editor coordinating the review of this article and approving it for publication was K. Navaie. *(Corresponding author: Hongyan Li.)*

Peng Wang is with the State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China, and also with the Pillar of Information Systems Technology and Design, Singapore University of Technology and Design, Singapore 487372 (e-mail: pengwangclz@163.com).

Hongyan Li and Shun Zhang are with the State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China (e-mail: hyli@xidian.edu.cn; zhangshunsdu@xidian.edu.cn).

Binbin Chen is with the Pillar of Information Systems Technology and Design, Singapore University of Technology and Design, Singapore 487372 (e-mail: binbin_chen@sutd.edu.sg).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2022.3163389.

Digital Object Identifier 10.1109/TWC.2022.3163389

inter-satellite communication has been successfully demonstrated in China's "BeiDou" [9] constellation and the US "Iridium" [10] constellation. They will also be supported in the "starlink" [11] and "Hongyan" [12] constellations, owned by American SpaceX company and China Aerospace Science and Technology Corporation, respectively. The data rate of inter-satellite communications increases as inter-satellite communications adopts higher frequency (e.g., from L(1-2GHz), S(2-4GHz), Ku(12-18GHz), K(18-27GHz), to Ka(27-40GHz) band). Using Ka band, the data rate for small satellites has reached 320 Mbps [13]. Furthermore, the laser-based inter-satellite communication technology has demonstrated data rate above 1Gbps [13]. By leveraging inter-satellite communication, an observation satellite that is not in direct contact with any ground station can relay its data to the ground via one (or multiple) relay satellite(s).

In this work, we aim to understand the potential benefit of using inter-satellite communication for enhancing earth observation throughput. Intuitively, higher throughput can be achieved by increasing the number of relay satellites. The amount of throughput gain, however, depends on a couple of key design factors, specifically:

*On demand scheduling capability:* Given transmission missions, whether the inter-satellite communication can be scheduled on-demand, or they must follow some predetermined patterns. The flexibility of on-demand scheduling can potentially bring higher throughput, especially for dynamic observation missions (e.g., the emergency monitoring of natural disasters). However, it also incurs extra implementation overhead in practice. In particular, one satellite may be in contact with multiple satellites at any given time. Due to a satellite's limited number of transceivers and the fact that both the transmission and the reception satellites need to track the location of each other, it is easier to pre-schedule all the communicating satellite pairs in advance.

*Co-optimization of observation and transmission capability:* If multiple observation satellites can observe one observation mission, is the system able to optimize the scheduling of the observation and transmission missions in a joint manner? Again, co-scheduling can potentially improve the system's throughput, but it implies that the scheduling of two different sub-systems needs to be tightly coordinated.

Fig. 1 illustrates the impact of these two factors using a simple example, where the observation and transmission opportunities are drawn as red solid lines with arrows. Suppose communication satellite $C_2$ can be scheduled to communicate to at most 2 other satellites. The highest throughput can be obtained in the following way: in the first time window, mission $M_1$ and $M_2$ are observed by observation satellite $O_1$ and $O_3$ respectively. In the second time window, mission $M_1$ is transmitted to the ground station $G_1$ directly while mission $M_2$ is transmitted back to ground station $G_2$ through communication satellite $C_2$ and $C_1$. However, if communication satellite $C_2$ is pre-scheduled to communicate with observe satellite $O_2$ in the second time window (i.e., no on-demand scheduling), it is easy to see that only mission $M_1$ can be finished in these two time windows. Now let us consider the benefit of joint optimization of observation and transmission:

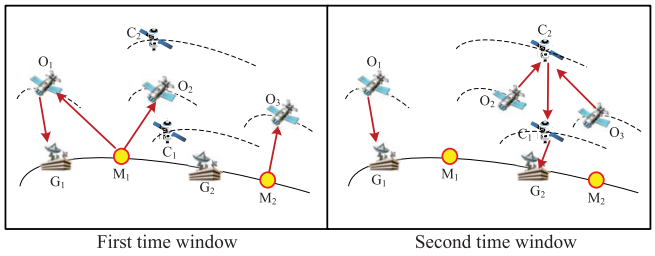


Fig. 1. The topology of time-varying satellite network.

Without considering the transmission opportunity, there is no difference in scheduling observation satellite $O_1$ or $O_2$ to carry out mission $M_1$. However, if mission $M_1$ is observed by satellite $O_2$ in the first time window, then in the second time window, communication satellite $C_2$ needs to choose between the competing transmission requests from observation satellite $O_2$ or $O_3$, resulting in finishing at most 1 mission.

Earth-observation scheduling has been extensively studied (e.g., in [14]–[23]). However, these works do not consider the transmission bandwidth constraints. On the other hand, there are also a large body of work (e.g., [10], [24]–[29]) that study the scheduling of inter-satellite communication problem without considering observation tasks. More specifically, [26]–[28] schedules the inter-satellite communication to provide each satellite with equal chances to communicate with each other, without considering dynamic traffic loads of different satellites. In comparison, references [25], [29] consider a given traffic demand and model the problem as a Mixed Integer Linear Programming, which is hard to solve due to its NP-completeness. The authors in [10] models the on-demand scheduling problem using a space-time topology graph. Although the problem is solved based on the bipartite graph matching algorithm, it still suffers from high computation complexity as the constraints are modeled for every time window that corresponds to the transmission of a single packet. In addition, its model assumes that a maximum of one relay satellite is used, which may result in sub-optimal network throughput. Finally, [30] studies the joint observation and transmission scheduling problem, without considering inter-satellite communication. As a result, the transmission is constrained to only the direct communication links between observation satellites and ground stations.

To rigorously study the potential gain of using inter-satellite communications for improving earth observation throughput, especially to evaluate the impact of the two design factors we have identified, our work formulates a generic earth observation and transmission problem, which is a linear programming (LP) problem. Solutions for generic LP problem may still incur high computational overhead and can be time-consuming when the network size is large. To enable easy deployment to real-world system, we transform the LP problem into an equivalent standard maximum flow problem, which can be solved in a more efficient manner. Our transformation adopts a generalized time-varying graph and introduces auxiliary vertices and edges in the graph to succinctly represent different resource constraints. The graph allows us

to efficiently and optimally derive the scheduling solutions under different settings (i.e., predetermined scheduling vs. on-demand scheduling, and independent scheduling vs. co-optimization of observation and transmission missions ). Based on our problem formulation, we conduct extensive numerical experiments based on real-world "Gaofen" and "starlink" satellite constellations. We reduce the running time of LP method by hundreds of times via the maximum flow algorithm. Our result shows that using 40 relay satellites from the "starlink" constellation can increase the throughput of 10 sensing satellites from the "Gaofen" constellation by more than 400%. They can also reduce the mission completion delay of 90% of 200 missions from 11 hours to 2 hours. The contribution of on-demand scheduling and observation-transmission co-optimization depends on the traffic load, and they can increase the throughput by more than 180% and 100% under a heavy and light but time-critical traffic load respectively. One interesting observation from our study is that, the observation-transmission co-optimization does not increase the throughput under heavy load or when there are abundant storage resources to support delay-tolerant missions. We also find that the satellite's on-board storage plays a more important role when the network communication capacity falls in a certain range (i.e., neither too high nor too low).

The remainder of the paper is organized as follows. We discuss related works on earth observation and transmission scheduling in Section II. In Section III, we introduce the background of earth observation and present our problem statement. In Section IV, we model the problem to maximize the early observation throughput as a linear programming (LP) problem. In Section V, we construct the generalized time-varying graph and introduce auxiliary vertices and edges to represent the constraints of the formulated LP problem, which allows us to use the maximum flow algorithm to optimally solve the problem. We present the evaluation results and conclusions in Section VI and VII, respectively.

## II. Related Work

### A. Earth Observation Scheduling

Many existing efforts that study how to schedule earth observation missions do not consider how to send the observed data to ground stations. Instead, they focus on the efficient use of the observation resources and time windows, either for a single satellite setting [14]–[17] or over multiple satellites [18]–[23]. For a single satellite, the decision is about how to best choose among competing missions and to decide the scheduling among them. For the multiple satellites setting, it also decides which satellite to observe which mission. Various schemes have been proposed to achieve different goals. Specially, to eliminate conflicts, Wu *et al.* [18] designed a metric to represent the opportunity cost for each mission such that the mission with higher cost will be discarded when it competes for resources with other missions. References [19] and [20] used group division strategy to aggregate multiple observation missions into single one for reducing the on/off switch times of satellite cameras to enhance their lifespan. To cater to missions with different priority levels, Xu *et al.* [21] designed priority-based indicators to evaluate benefits for missions and opportunity costs for missions' observation time windows, and then used a heuristic algorithm to solve the problem. Reference [22] modeled the earth observation scheduling problem as a mixed integer linear programming problem for maximizing the weighted number of completed missions. They introduced auxiliary linear variables and constraints in their model to reduce the number of integer variables and hence the computational cost. Du *et al.* [23] used the historical information of scheduling result to predict the probability that a mission can be fulfilled by different satellites. They then proposed to assign a mission to the satellite that provides the highest probability.

### B. Satellite Data Transmission Scheduling

Many existing efforts studying data transmission over satellite networks focus on inter-satellite communication scheduling [10], [24], [26]–[29]. Given all possible contacts among satellites, the inter-satellite communication scheduling problem selects a subset of contacts to establish inter-satellite communication, under each satellite's transceiver constraint. For instance, consider a satellite $C_1$ with a single transceiver that is within the communication range of multiple satellites (e.g., $C_2$ and $C_3$) at the same time. $C_1$ then has to choose between $C_2$ and $C_3$. [26]–[28] designed fair inter-satellite communication scheduling methods in centralized and distributed manners. The goal is to provide similar communication opportunities among different pairs of satellites, based on an underlying satellite contact graph. Such fair inter-satellite communication scheduling, however, does not take into consideration the traffic demand. Hence they cannot ensure maximum throughput for transferring earth observation data. Reference [29] considered different traffic demands, and modeled the inter-satellite communication scheduling problem as a mixed integer linear programming problem. However, the problem cannot be solved efficiently when the network scale is large. In [10], the authors designed an efficient algorithm for solving on-demand inter-satellite communication scheduling problem based on bipartite graph matching. However, its model assumes that a maximum of single relay satellite is used, which can cause sub-optimal network throughput.

There are also works that focus on the design of routing schemes over a given contact graph without considering satellite transceiver constraints. Here the contact graph models the contacts between satellites. Specifically, it depicts a contact between two satellites as a vertex and links two vertices by a directional edge if they satisfy certain conditions. Contact graph routing (CGR) [31] aims at finding the shortest path (i.e., a sequence of contacts) from a source to a destination over a given contact graph. The Consultative Committee for Space Data Systems (CCSDS) used CGR in their Schedule Aware Bundle Routing (SABR) algorithm [32]. SABR calculates the route over a satellite network where the inter-satellite communications are pre-determined. Zhang *et al.* proposed a delay bounded routing algorithm [33], where they used a storage time aggregated graph (STAG) [34] that represents satellites as vertices and used edges to represent the com-

munication links between two satellites. Specially, by splitting the time horizon into successive time windows and integrating the data rate of the communication links during each time window as a sequence, STAG represents the pre-determined communications schedule of satellite network. Since these routing algorithms aim at finding the shortest path for a given mission, they are not designed for optimizing the throughput of a satellite network given a set of observation missions.

### C. Joint Scheduling of Satellite Observation and Transmission Missions

The authors in [25] modeled the joint satellite observation and transmission scheduling problem as a mixed integer linear programming problem. Their problem aims to maximize the weighted data volume under both transceiver and energy resources constraints. However, they did not consider the observation capacity limits of satellites. In [30], the authors used an event-driven time expanded graph to model the joint scheduling problem. Specifically, they split the time horizon into equal-length time periods and the satellites and contacts among them for each time period are depicted as vertices and edges, respectively. Two vertices representing the same satellite in two adjacent time periods are linked by a directional edge, which characterizes the storage of that satellite. Xiao *et al.* [35] modeled the joint scheduling problem as a mixed integer linear programming one, and solved it as a two-stage flow shop scheduling problem. Neither [30] nor [35] considered the use of inter-satellite communication in their problem formulation.

## III. Problem Statement

In an earth observation system, observation satellites use their remote sensing equipment to observe their targets on earth. The obtained data needs to be transmitted to ground stations either directly or via relay satellites. A data processing center then gathers all the data for processing. Given a time horizon, the (average) throughput of an earth observation system measures the amount of observation data that can be collected for a given set of observation missions. In existing earth observation systems, an observation satellite can only transmit data directly to ground stations without using the help of other relay satellites. In this work, we aim to study how the extra communication opportunities provided by relay satellites can enhance the throughput of an earth observation system. In the following, we will first introduce some concepts for defining the problem.

- Given a time horizon $T$, the *inter-satellite communication scheduling problem* is about deciding the time windows inside $T$ for different satellites to communicate with one another. Specially, two conditions need to be fulfilled for two satellites to communicate: 1) They are in each other's communication range. Since satellites orbit the Earth in a fixed manner, the exact time interval when two satellites are in each other's communication range can be calculated in advance. 2) They have available transceiver for each other.
- *Predetermined inter-satellite communication scheduling* means the communication time windows among satellites are arranged in advance (i.e., without considering the traffic demands of observation missions).
- *On-demand inter-satellite communication scheduling* is conducted according to the transmission demand of observation satellites. Specifically, if an observation satellite $\mathbb{O}$ has data to send, the on-demand communication schedule will arrange communication time windows among satellites to help relay $\mathbb{O}$'s data to ground stations.
- A *path* means a series of successive communications from source to destination, such as path $(\mathbb{O}, \mathbb{A}, \mathbb{G})$, where $\mathbb{O}$ sends data to $\mathbb{A}$, $\mathbb{A}$ then sends data to $\mathbb{G}$.
- *Co-optimization of observation and transmission* means the joint scheduling of observation and transmission missions to optimize the overall throughput. Specifically, given observation missions in time horizon $T$, the following decisions are jointly made: (1) which observation missions to be carried out by which observation satellite in which time windows; (2) the inter-satellite and satellite-to-ground-station communication schedule; and (3) how much data to send during a communication window.

To understand the throughput gain of an earth observation system by using inter-satellite communication, two factors need to be considered: 1) The first is the flexibility level of the inter-satellite communications scheduling. As on-demand communication scheduling considers the transmission requests of observation satellites and the predetermined scheduling does not, the on-demand communication scheduling can potentially provide higher throughput. However, the on-demand scheduling needs prior information of missions and incurs extra coordination overhead for implementation in practice. 2) The second factor is the co-optimization of observation and transmission. Again, co-scheduling can potentially improve the system's throughput, but it implies that the scheduling of two different sub-systems needs to be tightly coordinated.

In the next section, we will consider both the on-demand inter-satellite communication scheduling and co-optimization of observation and transmission in formulating our problem, and show how it can be modelled and solved efficiently.

## IV. Problem Formulation

### A. Basic Notations for Satellite Network

For clarity, all the notations and their descriptions in this paper are collected in TABLE I. Specifically, we consider a satellite network (SN) that includes observation and relay satellites, and ground stations, which are denoted by $\mathcal{O} = \{\mathbb{O}_1, \ldots, \mathbb{O}_P\}$, $\mathcal{C} = \{\mathbb{C}_1, \ldots, \mathbb{C}_Q\}$ and $\mathcal{G} = \{\mathbb{G}_1, \ldots \mathbb{G}_K\}$, respectively. Here $P$, $Q$ and $K$ are the numbers of elements in $\mathcal{O}$, $\mathcal{C}$ and $\mathcal{G}$, respectively. SN also includes a data processing center (DPC) $\mathbb{D}$. Given a time horizon $T$, we use $\mathcal{M} = \{\mathbb{M}_1, \ldots, \mathbb{M}_N\}$ to denote the $N$ *observation missions* under consideration. Each mission is defined as $\mathbb{M}_i = \{\upsilon(\mathbb{M}_i), T_{start}(\mathbb{M}_i)\}$, where $\upsilon(\mathbb{M}_i)$ denotes the maximum amount of data that the mission could generate and $T_{start}(\mathbb{M}_i)$ denotes the mission's arrival time. In this work, we focus on the throughput of earth observation system, by assuming

TABLE I
NOTATIONS USED IN THIS PAPER

| Notations | Descriptions |
|---|---|
| $T$ | the time horizon for the scheduling problem |
| $\mathcal{O}$, $\mathcal{C}$, $\mathcal{G}$, $\mathcal{M}$ | the set of observation and communication satellites, ground stations and missions |
| $\mathbb{O}_i$, $\mathbb{C}_j$, $\mathbb{G}_k$, $\mathbb{M}_p$ | observation satellite i, communication satellite j, ground station k and observation mission p. |
| $v(\mathbb{M}_i)$,$T_{start}(\mathbb{M}_i)$ | the data volume and release time of mission $i$, respectively |
| $\delta = [t_s, t_e]$ | a time window, with $t_s$ and $t_e$ representing its start and end time, respectively |
| $\lvert\delta\rvert$ | the length of time window $\delta$, which equals to $t_e - t_s$. |
| $r^{\delta}_{\mathbb{M}_i,\mathbb{O}_j}$ | indicate the available observation rate of $\mathbb{O}_j$ when observe mission $\mathbb{M}_i$ in time window $\delta$ |
| r | the observation rate of a satellite's camera |
| $s^{\delta}_{\mathbb{M}_i,\mathbb{O}_j}$ | indicate the decided rate at which $\mathbb{O}_j$ observes $\mathbb{M}_i$ in $\delta$. |
| $r^{\delta}_{\mathbb{A},\mathbb{B}}$, $s^{\delta}_{\mathbb{A},\mathbb{B}}$ | the available and determined transmission rate from $\mathbb{A}$ to $\mathbb{B}$ in $\delta$ |
| $Q_{\mathbb{A}}$ | the available number of transmitters and receivers of $\mathbb{A}$ |
| $S_{\mathbb{A}}(\delta)$, $R_{\mathbb{A}}(\delta)$ | the total amount of data $\mathbb{A}$ can send and receive in $\delta$, respectively |
| $N(\mathbb{O}_i)$ | the number of cameras carried by satellite $\mathbb{O}_i$ |
| $\beta_{\mathbb{A}}$ | the available storage size of $\mathbb{A}$ in $T$ |
| $\beta_{\mathbb{A}}(\delta_i, \delta_{i+1})$ | the amount of data $\mathbb{A}$ stores and carries from $\delta_i$ to $\delta_{i+1}$ |
| $\mathcal{V}, \mathcal{E}$ | the set of vertices and edges, respectively |
| $(\mathbb{M}_i, \mathbb{O}^{\delta}_j)$ | a directional edge, on which the data goes from $\mathbb{M}_i$ to $\mathbb{O}_j$ in $\delta$ |
| $C_{\mathbb{M}_i,\mathbb{O}^{\delta}_j}$ | the maximum amount of data $\mathbb{O}_j$ can observe for $\mathbb{M}_i$ in $\delta$ |
| $[\mathbb{A}^{\delta}, \mathbb{B}^{\delta}]$ | a bidirectional edge consists of two directional edges $(\mathbb{A}^{\delta}, \mathbb{B}^{\delta})$ and $(\mathbb{B}^{\delta}, \mathbb{A}^{\delta})$ |
| $C_{\mathbb{A}^{\delta},\mathbb{B}^{\delta}}$ | the maximum amount of data $\mathbb{A}$ can send to $\mathbb{B}$ in $\delta$ |
| $(\mathbb{A}^{\delta_i}, \mathbb{A}^{\delta_{i+1}})$ | a directional edge indicating that the data is stored from $\delta_i$ to $\delta_{i+1}$ at vertex $\mathbb{A}$ |
| $C_{\mathbb{A}^{\delta_i},\mathbb{A}^{\delta_{i+1}}}$ | the maximum amount of data $\mathbb{A}$ can store from $\delta_i$ to $\delta_{i+1}$ |
| $f_{\mathbb{A}^{\delta_i},\mathbb{B}^{\delta_j}}$ | the determined amount of data flowing from vertex $\mathbb{A}^{\delta_i}$ to $\mathbb{B}^{\delta_j}$ |

all the missions are of equal importance and there are no specific deadlines for finishing them. We will leave the study of missions with different priority level and deadlines to future work.

Due to the mobility of satellites [36], inter-satellite communications and those between satellites and ground stations are *intermittent*. Similarly, a mission can only be carried out by an observation satellite when the observation target is within the satellite's observation range. We assume all the potential *communication and observation opportunities* can be calculated in advance based on the satellites' orbit parameters and the coordinates of the mission list.

In our problem formulation, we will further divide the time horizon $T$ into variable length time windows. We denote a time window $\delta \subseteq T$ by $\delta = [t_s, t_e]$, where $t_s$ and $t_e$ are the start and end time of $\delta$ and $|\delta| = t_e - t_s$ is its length. We split $T$ in such a way that in each time window $\delta$, the set of observation and communication opportunities remain unchanged. We will elaborate the time division mechanism in following subsection.

If a mission is within the observation range of a satellite's camera, there is observation opportunity. The maximum observation data rate at a satellite $\mathbb{O}_j \in \mathcal{O}$ for a mission $\mathbb{M}_i \in \mathcal{M}$ in time window $\delta$ is denoted by $r^{\delta}_{\mathbb{M}_i,\mathbb{O}_j}$, and the allocated observation rate is defined as $s^{\delta}_{\mathbb{M}_i,\mathbb{O}_j}$. In addition, considering $\mathbb{A}$ and $\mathbb{B}$ in $\mathcal{O} \cup \mathcal{C} \cup \mathcal{G}$, if the distance between $\mathbb{A}$ and $\mathbb{B}$ is less than their transmission ranges during the whole $\delta$, there is communication opportunity between $\mathbb{A}$ and $\mathbb{B}$. We denote the available and allocated transmission rate form $\mathbb{A}$ to $\mathbb{B}$ in $\delta$ by $r^{\delta}_{\mathbb{A},\mathbb{B}}$ and $s^{\delta}_{\mathbb{A},\mathbb{B}}$, respectively. Note that, two ground stations can communicate through terrestrial networks and possess the permanent communication opportunity.

### B. Models of Wireless Channels

In this subsection, we elaborate the channel model we adopt for both inter-satellite links and satellite-to-ground stations links. We start with the notations. $P_{tr}$ is the constant transmission power (unit:$W$), $G_{tr}$ is the transmission antenna gain (unit:$dB$), and $G_{re}$ represents the reception antenna gain (unit:$dB$). Besides, $L^{\delta}_f$ means the free space loss related to communication center frequency (unit:$Hz$) $f$ of the channel and time window $\delta$, while $L^{\delta}_p$ represents the atmospheric attenuation in $\delta$. In particular, the free space loss [37], [38] $L^{\delta}_f = (\frac{c}{4\pi s(\delta) f})^2$, where $c$ is the speed of light (unit:$km/s$), $s(\delta)$ represents the slant range (unit:$km$) in delta. In addition, the atmospheric attenuation [39] $L^{\delta}_p = L^{\delta} \times \gamma^{\delta}_R$, where $L^{\delta}$ represents the equivalent effective slant-path length (unit: $km$) in time window $\delta$. It can be obtained by Recommendation ITU-R P.618-12 [39] and ITU-R P.839 [40]. Besides, $\gamma^{\delta}_R = \rho \times R^{\eta}$ (unit:$dB/km$) is the attenuation per kilometer in slot $\delta$, where $R$ is the rainfall intensity (unit:$mm/h$), while $\rho$ and $\eta$ can be obtained in Recommendation ITU-R P.839-3 [41].

We use the inter-satellite channel model as in [37] to elaborate its achievable data rate $r^{\delta}$ in time window $\delta$,

$$r^{\delta} = \frac{P_{tr} G_{tr} G_{re} L^{\delta}_f}{k T_s [\frac{E_b}{N_o}]_{req} M}, \tag{1}$$

where $k$ and $T_s$ respectively represents the Boltzmann's constant (unit:$JK^{-1}$) and total system noise temperature(unit:$K$), and $[\frac{E_b}{N_o}]_{req}$ means the required ratio of the reception energy per bit to the noise density. $M$ is the inter-satellite link margin. As a result, the achievable inter-satellite link rate changes with the slant range between the two satellites.

Regarding the achievable data rate for a down link between a satellite and a ground station, we first calculate the $SNR(\delta)$

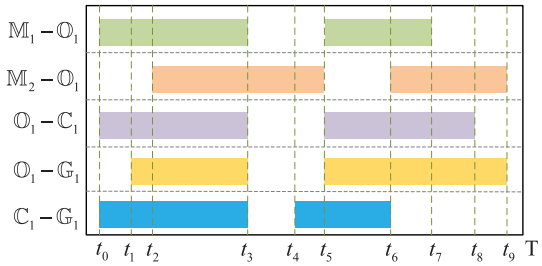


Fig. 2. An example to illustrate the time division mechanism for determining time windows.

in a time window $\delta$ as:

$$SNR(\delta) = \frac{P_{tr} G_{tr} G_{re} L_f^\delta L_p^\delta}{N}, \tag{2}$$

where $N$ is the additional Gaussian white noise power. Thus, we adopt Shannon formula to calculate the achievable data rate $r_{ij}^\delta$ of the satellite $j$ to ground station $j$ link channel,

$$r_{ij}^\delta = B \log_2(1 + SNR(\delta)), \tag{3}$$

where $B$ is the available bandwidth of the satellite to ground station link.

As a result, the achievable data rate between satellites and ground stations will be affected by the time-varying weather and the relative positions between satellites and ground stations. Following work [42], we assume that the rain attenuation does not change for a time period and the satellite-to-ground channel states of that period can be known at the beginning of the period. According to [43], the time period can be 30 minutes considering all the technical limitations today (e.g., regarding the weather forecasting accuracy).

### C. Time Division Mechanism

Due to the intermittent communication and observation opportunities, the satellite network topology is time-varying. Reference [10] adopted the transmission time of one packet as the length of a time unit to describe the topology. Suppose a packet lasts for 10 millisecond, that approach will then need to divide a one-hour time horizon into 360,000 time units, and one day into more than 8.6 million units, resulting in very high scheduling complexity.

To overcome the problem, we adopt the time division mechanism as in [44]. Fig. 2 illustrates how our approach works using a simple example that consists of two missions $\{\mathbb{M}_1, \mathbb{M}_2\}$, one observation satellite $\mathbb{O}_1$, one communication satellite $\mathbb{C}_1$ and one ground station $\mathbb{G}_1$. All the available observation and communication opportunities are depicted along time horizon $T$ as color-coded blocks in a Gantt chart. Based on that, we can identify all the time moments when there is any change (i.e., an opportunity starts or ends). In the example, these moments are denoted by $t_0$ to $t_9$, in an ascending order. Two consecutive moments will then form the start and end time of one time window. As such, we can divide the time into 9 time windows with the $i$-th time window being $[t_{i-1}, t_i]$. While the set of observation and communication opportunities change over time, they remain unchanged in each time window.

By applying this mechanism, we split a horizon $T$ into $n$ time windows $\{\delta_1, \ldots, \delta_n\}$ with $\delta_i = [t_{i-1}, t_i]$, where the length of time window varies. As one time window can span multiple seconds or even minutes, such a time division mechanism can significantly reduce the scale of problems we need to solve, as compared to the fixed-interval-length approach used by [10].

### D. Modeling the Constraints

In the following, we will define the observation constraints, the constraints for inter-satellite communication scheduling (i.e., the transceiver resource constraints), and the constraints ensuring joint scheduling (i.e., flow conservation and capacity constraints).

*1) Inherent Constraints of Earth Observation:* The observation opportunity only occurs when the special geographical constraints are satisfied (i.e., the observation target is under the satellite's observation range wherein the satellite can rotate its sensor to observe the target) [20]. We assume this knowledge of observation opportunities is given in advance.

We use $v(\mathbb{M}_i)$ to denote the maximum data volume of $\mathbb{M}_i$' when it is observed under the highest image resolution. Hence, for any mission $\mathbb{M}_i \in \mathcal{M}$,

$$\sum_{\delta \subseteqq T} \sum_{\mathbb{O}_j \in \mathcal{O}} |\delta| \times s_{\mathbb{M}_i, \mathbb{O}_j}^\delta \leq v(\mathbb{M}_i), \quad \forall \mathbb{M}_i \in \mathcal{M}. \tag{4}$$

The rate $s_{\mathbb{M}_i, \mathbb{O}_j}^\delta$ at which $\mathbb{O}_j$ observes mission $\mathbb{M}_i$ in $\delta$ should satisfy two conditions: 1) It is less than the available observation rate $r_{\mathbb{M}_i, \mathbb{O}_j}^\delta$; 2) There is observation opportunity between $\mathbb{O}_j$ and $\mathbb{M}_i$ in $\delta$. We can simply set $r_{\mathbb{M}_i, \mathbb{O}_j}^\delta = 0$ when there is no observation opportunity. Thus,

$$s_{\mathbb{M}_i, \mathbb{O}_j}^\delta \leq r_{\mathbb{M}_i, \mathbb{O}_j}^\delta, \quad \forall \mathbb{M}_i \in \mathcal{M}, \mathbb{O}_j \in \mathcal{O}, t_s \geq T_{start}(\mathbb{M}_i). \tag{5}$$

Furthermore, one observation satellite may have multiple *observation opportunities* to observe different missions at the same time. To avoid competition for cameras among missions, many works [30], [45]–[47] constrained that one camera could only observe one mission at the same time. We assume a satellite $\mathbb{O}_j$ carries $N(\mathbb{O}_j)$ cameras and each camera can provide the same observation rate $r$. We further assume that within time window $\delta \subseteqq T$, more than one missions can be observed by sharing the satellite's cameras. Hence, the sum of the missions' observation rate cannot exceed the satellites' available observation rate.[1] Namely,

$$\sum_{\mathbb{M}_i \in \mathcal{M}} s_{\mathbb{M}_i, \mathbb{O}_j}^\delta \leq r \times N(\mathbb{O}_j), \quad \forall \mathbb{O}_j \in \mathcal{O}, \delta \subseteqq T. \tag{6}$$

*2) Constraints for Inter-Satellite Communication Scheduling:* Consider a satellite (or a ground station) $\mathbb{A} \in \mathcal{O} \cup \mathcal{C} \cup \mathcal{G}$, the number of other satellites it can communicate with at the same time is limited by the number of transceivers it carries. We consider each satellite (or ground station) carries limited number of transceivers as in [24], [25], [48]. Also, we assume the transmission and reception antennas for satellites are

[1]We do not model the setup time for a camera to switch between two different missions, as they are usually much shorter than the length of a time window.

separated as in [49]. Hence, the scheduling of transmission and reception are independent.

Since phased-array antennas have been widely used in satellites such as WINDS, Spaceway3, WGS, Globestar and Iridium-NEXT [50], we assume the communication switching delay (i.e., the delay for a satellite $\mathbb{A}$ to switch communication to satellite $\mathbb{B}$ from $\mathbb{C}$) can be ignored. Furthermore, through combining the phased-array antennas with time-division multiple access technology as in [49], one satellite can communicate to multiple satellites in a time window $\delta$.

As a result, for $\mathbb{A}$ in any time window $\delta$, its transmission and reception will be constrained by data volume $S_{\mathbb{A}}(\delta)$ and $R_{\mathbb{A}}(\delta)$, respectively. Specially, $S_{\mathbb{A}}(\delta)$ means the total amount of data all the $Q_{\mathbb{A}}$ transmitters can transmit in whole $\delta$. Similarly, all $\mathbb{A}$'s receivers can receive $R_{\mathbb{A}}(\delta)$ data in $\delta$. Without loss of generality, we assume $\mathbb{A}$'s transmission links have the same transmission rate in $\delta$, and the reception links also have the same reception rate. As such, for any $\mathbb{A} \in \mathcal{O} \cup \mathcal{C} \cup \mathcal{G}$ in $\delta$, $S_{\mathbb{A}}(\delta) = Q_{\mathbb{A}} \times r^{\delta}_{\mathbb{A},\mathbb{B}} \times |\delta|, \forall \mathbb{B}^{\delta} \in \{\mathbb{B}^{\delta} \in \mathcal{V} | (\mathbb{A}^{\delta}, \mathbb{B}^{\delta}) \in \mathcal{E}\}$, $R_{\mathbb{A}}(\delta) = Q_{\mathbb{A}} \times r^{\delta}_{\mathbb{B},\mathbb{A}} \times |\delta|, \forall \mathbb{B}^{\delta} \in \{\mathbb{B}^{\delta} \in \mathcal{V} | (\mathbb{B}^{\delta}, \mathbb{A}^{\delta}) \in \mathcal{E}\}$. Thus, the linear transmission constraint is defined as follows,

$$\sum_{\mathbb{B}^{\delta}:(\mathbb{A}^{\delta},\mathbb{B}^{\delta})\in\mathcal{E}} s^{\delta}_{\mathbb{A},\mathbb{B}} \times |\delta| \leq S_{\mathbb{A}}(\delta), \quad \forall \mathbb{A} \in \mathcal{O} \cup \mathcal{C} \cup \mathcal{G}, \delta \subseteqq T. \tag{7}$$

Similarly, the reception constraint can also be formulated as follow,

$$\sum_{\mathbb{B}^{\delta}:(\mathbb{B}^{\delta},\mathbb{A}^{\delta})\in\mathcal{E}} s^{\delta}_{\mathbb{B},\mathbb{A}} \times |\delta| \leq R_{\mathbb{A}}(\delta), \quad \forall \mathbb{A} \in \mathcal{O} \cup \mathcal{C} \cup \mathcal{G}, \delta \subseteqq T. \tag{8}$$

Furthermore, when two entities $\mathbb{A}$ and $\mathbb{B}$ are scheduled to communicate within time window $\delta \subseteqq T$, the decided transmission rate $s^{\delta}_{\mathbb{A},\mathbb{B}}$ cannot exceed $r^{\delta}_{\mathbb{A},\mathbb{B}}$. Namely,

$$s^{\delta}_{\mathbb{A},\mathbb{B}} \leq r^{\delta}_{\mathbb{A},\mathbb{B}}, \quad \forall \mathbb{A}, \mathbb{B} \in \mathcal{O} \cup \mathcal{C} \cup \mathcal{G}, \delta \subseteqq T. \tag{9}$$

Note that when there is no communication opportunities between $\mathbb{A}$ and $\mathbb{B}$ during time window $\delta$, we can simply set $r^{\delta}_{\mathbb{A},\mathbb{B}} = 0$.

*3) Constraints Ensuring Joint Scheduling:* Firstly, for all missions $\mathbb{M}_i \in \mathcal{M}$, their observed data amount equals to that transmitted to data processing center, namely,

$$\sum_{\mathbb{M}_i \in \mathcal{M}} \sum_{\delta \subseteqq T} \sum_{\mathbb{O}^{\delta}_j \in \mathcal{O}} (s^{\delta}_{\mathbb{M}_i,\mathbb{O}_j} \times |\delta|) = \sum_{\delta \subseteqq T} \sum_{\mathbb{G}_i \in \mathcal{G}} (s^{\delta}_{\mathbb{G}_i,\mathbb{D}} \times |\delta|), \tag{10}$$

which ensures that the observed data must be transmitted to ground.

Next, for the entity other than the missions $\mathcal{M}$ and the data processing center $\mathbb{D}$, the total amount of data flowing into it should be equal to that out of it in time horizon $T$, namely,

$$\sum_{\mathbb{A} \in \mathcal{N}} \sum_{\delta \subseteqq T} (s^{\delta}_{\mathbb{A},\mathbb{B}} \times |\delta|) = \sum_{\mathbb{A} \in \mathcal{N} \cup \{\mathbb{D}\}} \sum_{\delta \subseteqq T} (s^{\delta}_{\mathbb{B},\mathbb{A}} \times |\delta|), \forall \mathbb{B} \in \mathcal{O} \cup \mathcal{C} \cup \mathcal{G}, \tag{11}$$

where $\mathcal{N} = \mathcal{M} \cup \mathcal{O} \cup \mathcal{C} \cup \mathcal{G}$. This constraint ensures that the data scheduling is lossless.

In fact, the satellites of SN can store, carry and forward the data. For $\mathbb{A} \in \mathcal{O} \cup \mathcal{C} \cup \mathcal{G}$, we define its available storage size in $T$ as $\beta_{\mathbb{A}}$. For clarity, we use $\delta_i$ to denote the $i$-th time window of $T$ and $\beta_{\mathbb{A}}(\delta_i, \delta_{i+1})$ to denote the amount of data that $\mathbb{A}$ stores from $i$-th time window to its adjacent time window. Note that, when $\delta_i$ is the first and last time window of $T$, there is $\beta_{\mathbb{A}}(\delta_{i-1}, \delta_i) = 0$ and $\beta_{\mathbb{A}}(\delta_i, \delta_{i+1}) = 0$, respectively. In particular, $|\delta_i|(\sum_{\mathbb{B} \in \mathcal{N}} s^{\delta_i}_{\mathbb{B},\mathbb{A}} - \sum_{\mathbb{B} \in \mathcal{N} \cup \{\mathbb{D}\}} s^{\delta_i}_{\mathbb{A},\mathbb{B}}) + \beta_{\mathbb{A}}(\delta_{i-1}, \delta_i) = \beta_{\mathbb{A}}(\delta_i, \delta_{i+1})$. That is, the total amount of data $\mathbb{A}$ receives in $\delta_i$ and stores from time window $\delta_{i-1}$ to $\delta_i$, equals to the amount $\mathbb{A}$ transmits in $\delta_i$ and stores from $\delta_i$ to $\delta_{i+1}$. Naturally, the amount of data $\mathbb{A}$ stores from one time window to the adjacent window cannot exceed $\mathbb{A}$'s storage size, namely,

$$\beta_{\mathbb{A}}(\delta_i, \delta_{i+1}) \leq \beta_{\mathbb{A}}, \delta_i \subseteqq T, \quad \forall \mathbb{A} \in \mathcal{O} \cup \mathcal{C} \cup \mathcal{G}. \tag{12}$$

With the above constraints, we denote the successfully observed and transmitted volume of data for $\mathbb{M}_i$ as $F(\mathbb{M}_i)$, namely,

$$F(\mathbb{M}_i) = \sum_{\delta \subseteqq T} \sum_{\mathbb{O}_j \in \mathcal{O}} (|\delta| \times s^{\delta}_{\mathbb{M}_i,\mathbb{O}_j}), \quad \forall \mathbb{M}_i \in \mathcal{M}. \tag{13}$$

As a result, (10) constrains that the observed data must be transmitted to data processing center. (11) allows the intermediate satellites to help transmit the observed data of observation satellite to ground stations. Besides, (12) allows data to be stored and carried by a satellite before being forwarded, which supports the joint use of transmission and storage resources.

### *E. Problem Formulation*

Let $s$ denote the union of all observation rate $s^{\delta}_{\mathbb{M}_i,\mathbb{O}_j}$ and transmission rate $s^{\delta}_{\mathbb{A},\mathbb{B}}$ over all $\delta \in T$. With the above constraints, we formulate the problem to maximize the observed and transmitted mission data amount (i.e., the throughput of SN) as follow,

$$\textbf{P1:} \max_{s} \sum_{\mathbb{M}_i \in \mathcal{M}} F(\mathbb{M}_i)$$
$$s.t.\ (4) - (13).$$

Although the problem **P1**, being a linear programming problem, can be optimally solved in polynomial time, the required complexity is still high for large time horizon $T$ and when the network consists of a large number of missions and satellites. In the next section, we will transform **P1** into a well-investigated network flow problem, which can reduce the running time for solving **P1** by hundreds of times as in Section VI. Figure 3 summarizes the overall diagram we take to efficiently solve **P1**.

### *F. Different Settings of Problem P1*

Before we proceed to solving **P1**, it is worth noting that both observation and inter-satellite transmission scheduling are embedded in the problem formulation of **P1**, which considers on-demand inter-satellite communications and conducts transmission-observation co-optimization with the goal of maximizing the throughput of SN.

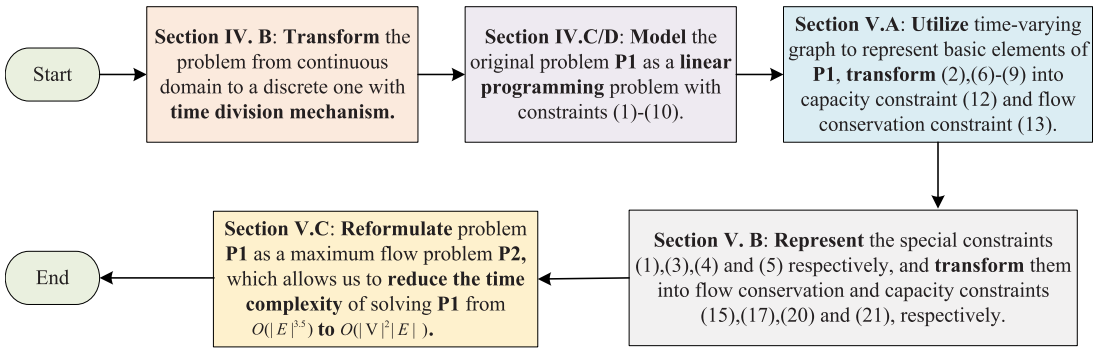


Fig. 3. The diagram for efficiently solving P1.

If the inter-satellite communications are pre-determined, the communication scheduling between two satellites $\mathbb{A}$ and $\mathbb{B}$ is decided without knowing the traffic demand. Problem **P1** removing constraints (4)-(6) can be viewed as the pre-determined inter-satellite communication problem. After solving it, the observation-transmission co-optimization is achieved through solving simplified **P1** by excluding its constraints (7) and (8).

With respect to the independent scheduling of observation and transmission, the observation and transmission scheduling can be viewed as two sub-problems of **P1**. Specifically, the problem for maximizing the amount of observed data can be viewed as the **P1** that drops constraints (7)-(13). Correspondingly, **P1** without constraints (4)-(6) is to solve the transmission problem with the goal to maximize the throughput of the given observation data.

## V. Graph Based Representation for Throughput Maximization

To efficiently solve problem **P1**, we first design a time-varying graph to model its basic elements. Next, we introduce virtual vertices and edges to transform the unique constraints of problem **P1** as flow conservation and capacity ones, which allows us to solve problem **P1** efficiently as a maximum flow problem.

### *A. Time-Varying Graph-Based Representation for Basic Elements*

The scheduling time horizon $T$ is split into $n$ time windows $\{\delta_1, \ldots, \delta_n\}$ by adopting the time division mechanism in Section IV-C. Next, the time-varying graph $G = (\mathcal{V}, \mathcal{E})$ is designed, representing all the observation and communication opportunities in each time window $\delta \in T$, and the storage process for each satellite. Specifically, all the entities, i.e., missions $\mathcal{M}$, observation satellites $\mathcal{O}$, communication satellites $\mathcal{C}$, ground stations $\mathcal{G}$ and DPC $\mathbb{D}$, are represented as vertices, while all the *observation and transmission opportunities* for any two vertices are modelled as edges. Meanwhile, we also represent the storage resources of satellites and ground stations as edges. We illustrate the detailed process in the following.

*1) Vertex Representation:* For each element $\mathbb{A} \in \mathcal{O} \cup \mathcal{C} \cup \mathcal{G}$, we create $n$ vertices, one to represent $\mathbb{A}$ in each of the $n$ time windows. Specially, vertex $\mathbb{A}$ in $\delta_i$ is denoted as $\mathbb{A}^{\delta_i}$. For each mission in $\mathcal{M}$, one vertex is created. We create a single vertex for the DPC $\mathbb{D}$ too. Finally, all the vertices are collected into the vertex set $\mathcal{V}$.

*2) Observation and Transmission Opportunity Representation:* In each time window $\delta_q \in T$, we represent all the observation opportunities of the satellite network as edges. Specifically, for $\forall \mathbb{M}_i \in \mathcal{M}, \mathbb{O}_j \in \mathcal{O}$ in $\delta_q \in T$, if the available observation rate $r^{\delta_q}_{\mathbb{M}_i,\mathbb{O}_j} > 0$, it is represented as a directional edge $(\mathbb{M}_i, \mathbb{O}_j^{\delta_q})$, where the direction is from $\mathbb{M}_i$ to $\mathbb{O}_j^{\delta_q}$. The edge's weight, describing the maximum volume of data that $\mathbb{O}_j$ can observe for $\mathbb{M}_i$ in $\delta_q$, is defined as $C_{\mathbb{M}_i,\mathbb{O}_j^{\delta_q}}$ and $C_{\mathbb{M}_i,\mathbb{O}_j^{\delta_q}} = r^{\delta_q}_{\mathbb{M}_i,\mathbb{O}_j} \times |\delta_q|$. Similarly, in each time window $\delta_i \in T$, if $r^{\delta_i}_{\mathbb{A},\mathbb{B}} > 0$ for $\forall \mathbb{A}, \mathbb{B} \in \mathcal{O} \cup \mathcal{C} \cup \mathcal{G}$, we model it as one bidirectional edge $[\mathbb{A}^{\delta_i}, \mathbb{B}^{\delta_i}]$, which is the combination of two directional edges $(\mathbb{A}^{\delta_i}, \mathbb{B}^{\delta_i})$ and $(\mathbb{B}^{\delta_i}, \mathbb{A}^{\delta_i})$. Moreover, each directional edge $(\mathbb{A}^{\delta_i}, \mathbb{B}^{\delta_i})$ has weight $C_{\mathbb{A}^{\delta_i},\mathbb{B}^{\delta_i}}$ to characterize the maximum amount of data that $\mathbb{A}$ can transmit to $\mathbb{B}$ in $\delta_i$ and $C_{\mathbb{A}^{\delta_i},\mathbb{B}^{\delta_i}} = r^{\delta_i}_{\mathbb{A},\mathbb{B}} \times |\delta_i|$. Finally, all the observation and communication edges are put into $\mathcal{E}$.

*3) Storage Constraint Representation:* To represent the process that one satellite stores data from one time window to the adjacent time window, storage edges are introduced. Specifically, for $\forall \mathbb{A} \in \mathcal{V} - \mathcal{M} - \{\mathbb{D}\}$ and $\forall i \in [1, n-1]$, one directional edge $(\mathbb{A}^{\delta_i}, \mathbb{A}^{\delta_{i+1}})$ is added between $\mathbb{A}^{\delta_i}$ and $\mathbb{A}^{\delta_{i+1}}$. The weight of the edge represents the data amount $\mathbb{A}$ can store from $\delta_i$ to $\delta_{i+1}$, which is denoted as $C_{\mathbb{A}^{\delta_i},\mathbb{A}^{\delta_{i+1}}}$.

To illustrate our approach, let us look at a concrete example satellite network with 3 missions, 2 observation satellites, 2 communication satellites, 1 ground station and a DPC, as shown in Fig. 4. The time horizon $T$ is split into 2 time windows, and the missions, satellites, ground stations and DPC are represented as vertices. Besides, in Fig. 4, the observation and transmission opportunities are separately marked as green directional edges and orange bidirectional edges, and the storage edges are modeled as dotted directional edges. For instance, edges $(\mathbb{M}_1, \mathbb{O}_1^{\delta_1})$ and $[\mathbb{O}_2^{\delta_1}, \mathbb{C}_1^{\delta_1}]$ are the observation and transmission links respectively. Besides, the weights labeled on the edge (e.g., $C_{\mathbb{M}_1,\mathbb{O}_1^{\delta_1}} = 3$, $C_{\mathbb{O}_2^{\delta_1},\mathbb{C}_1^{\delta_1}} = 6$), characterize the maximum amount of data that can be observed or transmitted within a time window. In addition, the storage edge (e.g., $(\mathbb{O}_1^{\delta_1}, \mathbb{O}_1^{\delta_2})$) represents the satellite's available storage size from $\delta_1$ to $\delta_2$.

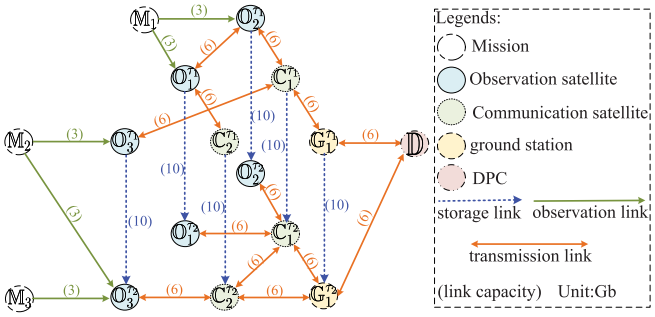


Fig. 4. Graph representation for the network.

With the constructed graph $G = (\mathcal{V}, \mathcal{E})$, the decision variables $s$ can be mapped into a new decision variable $f$, which represents the scheduled amount of data flows from one vertex to another. Explicitly, for any observation edge $(\mathbb{M}_i, \mathbb{O}_j^{\delta_q})$ and communication edge $(\mathbb{A}^{\delta_i}, \mathbb{B}^{\delta_i})$ in $\mathcal{E}$, the new decision variable will be

$$\begin{cases} f_{\mathbb{M}_i,\mathbb{O}_j^{\delta_q}} = s_{\mathbb{M}_i,\mathbb{O}_j^{\delta_q}} \times |\delta_q|, & \forall (\mathbb{M}_i, \mathbb{O}_j^{\delta_q}) \in \mathcal{E}, \\ f_{\mathbb{A}^{\delta_i},\mathbb{B}^{\delta_i}} = s_{\mathbb{A}^{\delta_i},\mathbb{B}^{\delta_i}} \times |\delta_i|, & \forall (\mathbb{A}^{\delta_i}, \mathbb{B}^{\delta_i}) \in \mathcal{E}, \end{cases} \tag{14}$$

where the data volume of $\mathbb{M}_i$ that $\mathbb{O}_j$ wants to observe in $\delta_q$ is defined as $f_{\mathbb{M}_i,\mathbb{O}_j^{\delta_q}}$, while $f_{\mathbb{A}^{\delta_i},\mathbb{B}^{\delta_i}}$ is the amount of data that $\mathbb{A}$ transmits to $\mathbb{B}$ in $\delta_i$.

Naturally, the flow across one edge cannot exceed the edge's capacity, namely,

$$0 \le f_{\mathbb{A}^{\delta_i},\mathbb{B}^{\delta_j}} \le c_{\mathbb{A}^{\delta_i},\mathbb{B}^{\delta_j}}, \quad \forall (\mathbb{A}^{\delta_i}, \mathbb{B}^{\delta_i}) \in \mathcal{E}, \delta_i, \delta_j \in T, \tag{15}$$

it can be easily checked that the satisfaction of (15) indicates the satisfaction of (5) and (9) for problem **P1**. To ensure no loss of information,

$$\begin{aligned} &\sum_{\mathbb{A}^{\delta_i} \in \mathcal{V}} f_{\mathbb{A}^{\delta_i},\mathbb{B}^{\delta_i}} - \sum_{\mathbb{A}^{\delta_i} \in \mathcal{V}-\mathcal{M}} f_{\mathbb{B}^{\delta_i},\mathbb{A}^{\delta_i}} \\ &= \begin{cases} f_{\mathbb{B}^{\delta_i},\mathbb{B}^{\delta_{i+1}}}, i = 1, & \forall \mathbb{B} \in \mathcal{O} \cup \mathcal{C} \cup \mathcal{G}, \\ f_{\mathbb{B}^{\delta_i},\mathbb{B}^{\delta_{i+1}}} - f_{\mathbb{B}^{\delta_{i-1}},\mathbb{B}^{\delta_i}}, & \forall i \in [2, n-1], \mathbb{B} \in \mathcal{O} \cup \mathcal{C} \cup \mathcal{G}, \\ -f_{\mathbb{B}^{\delta_{i-1}},\mathbb{B}^{\delta_i}}, i = n, & \forall \mathbb{B} \in \mathcal{O} \cup \mathcal{C} \cup \mathcal{G}. \end{cases} \end{aligned} \tag{16}$$

The above constraint allows satellite to store data from current time window to adjacent window and transmit the stored data. Next, for problem **P1**, the satisfaction of (16) indicates the satisfaction of (10) and (11), and the satisfaction of both (15) and (16) implies the satisfaction of (12).

A typical network flow problem (e.g., maximum flow) only involves capacity flow constraints and flow conservation ones. With the help of the representations we presented earlier, the satisfaction of capacity constraints (15) and flow conservation constraints (16) will indicate the satisfaction of (5), (9) - (12) for problem **P1**. However, the remaining constraints (4), (6), (7) and (8) for problem **P1** still do not hold with the satisfaction of capacity and flow conservation ones. To overcome this problem, we introduce graph operations to transform these unique constraints in the next subsection.

### *B. Graph-Based Representation for the Remaining Constraints of **P1***

We now introduce some virtual vertices and edges to represent the unique constraints of problem **P1**.

*1) Redundant Observation Avoidance Representation:* As redundant observation will cause resource wastes, it is necessary to ensure the observed data volume of mission $\mathbb{M}_i$ does not exceed that mission's maximum data amount $v(\mathbb{M}_i)$. To represent this, we introduce a virtual vertex $\mathbb{M}$ and we connect every mission $\mathbb{M}_i$ with $\mathbb{M}$ using an *observation constraint edge* $(\mathbb{M}, \mathbb{M}_i)$. Specially, the mission $\mathbb{M}_i$'s data volume is labeled on the virtual edge $(\mathbb{M}, \mathbb{M}_i)$ as its capacity,

$$C_{\mathbb{M},\mathbb{M}_i} = v(\mathbb{M}_i), \quad \forall \mathbb{M}_i \in \mathcal{M}. \tag{17}$$

Furthermore, with the help of the flow conservation on each mission $\mathbb{M}_i \in \mathcal{M}$ and edge capacity constraints of edges, we can obtain

$$\sum_{\mathbb{A}^{\delta_q} \in \mathcal{V}} f_{\mathbb{M}_i,\mathbb{A}^{\delta_q}} = f_{\mathbb{M},\mathbb{M}_i} \le C_{\mathbb{M},\mathbb{M}_i} = v(\mathbb{M}_i), \quad \forall \mathbb{M}_i \in \mathcal{M}, \tag{18}$$

which means the data volumes flowing into $\mathbb{M}_i$ equals to those flowing out of $\mathbb{M}_i$ (i.e., the observed data volume of $\mathbb{M}_i$) and cannot exceed $v(\mathbb{M}_i)$. With the network topology in Fig. 4, we present one example to illustrate the representation in Fig. 5(a), where the inserted virtual edges are $(\mathbb{M}, \mathbb{M}_1)$, $(\mathbb{M}, \mathbb{M}_2)$ and $(\mathbb{M}, \mathbb{M}_3)$, respectively. After the representation operation in Fig. 5(a), the redundant observation avoidance can be achieved through satisfying common flow conservation and capacity constraints (18), which indicates the satisfaction of (4).

*2) Observation Competition Elimination Characterization:* One observation satellite could have multiple observation opportunities in a time window $\delta$, (15) and (16) do not limit the amount of mission data one satellite can observe. As a result, (6) could be violated.

To overcome this problem, in each time window $\delta_q \in T$, as in [44], we introduce a virtual vertex $\mathbb{O}_{i'}^{\delta_q}$ in $G$ for each observation satellite $\mathbb{O}_i^{\delta_q}$ that satisfies the condition, where multiple mission vertices are connected to it by edges (i.e., it has multiple observation opportunities in $\delta_q$). Next, we delete the edges between mission vertices and $\mathbb{O}_i^{\delta_q}$ from $\mathcal{E}$, link the mission vertices to $\mathbb{O}_{i'}^{\delta_q}$. The corresponding capacity of the new edge $(\mathbb{M}_i, \mathbb{O}_{i'}^{\delta_q})$ is the same as that of the removed edge $(\mathbb{M}_i, \mathbb{O}_i^{\delta_q})$. Furthermore, we connect the virtual vertex $\mathbb{O}_{i'}^{\delta_q}$ with original vertex $\mathbb{O}_i^{\delta_q}$ to construct the *observation constraint edge* $(\mathbb{O}_{i'}^{\delta_q}, \mathbb{O}_i^{\delta_q})$ with the capacity as

$$C_{\mathbb{O}_{i'}^{\delta_q},\mathbb{O}_i^{\delta_q}} = r \times N(\mathbb{O}_i) \times |\delta_q|, \tag{19}$$

which is the total volume of data that satellite can observe in $\delta_q$. With Fig. 4, an example of the conflict resolution graph is shown in Fig. 5(b). Hence, with the flow conservation of the new inserted vertex $\mathbb{O}_{i'}^{\delta_q}$ and capacity constraints of its involving edges, there will be

$$\sum_{\mathbb{M}_i \in \mathcal{M}} f_{\mathbb{M}_i,\mathbb{O}_{i'}^{\delta_q}} = f_{\mathbb{O}_{i'}^{\delta_q},\mathbb{O}_i^{\delta_q}} \le r \times N(\mathbb{O}_i) \times |\delta_q|, \tag{20}$$

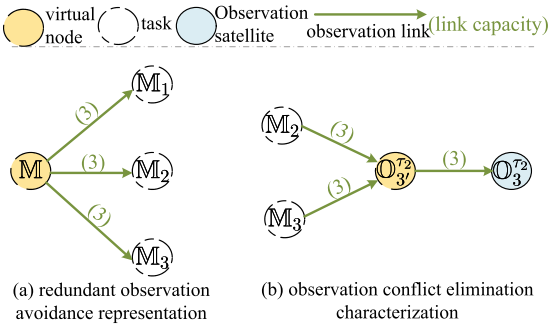


Fig. 5. Graph-based constraints representation in the observation phase.

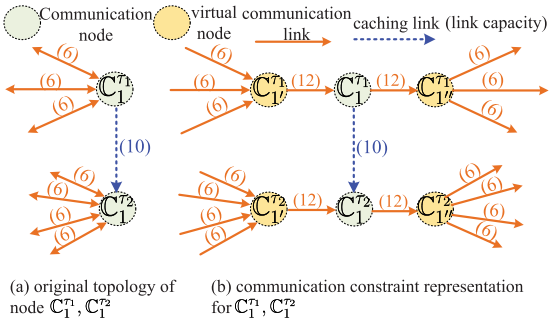


Fig. 6. Communication constraints representation example for $\mathbb{C}_1$.

where the planned observation data $\sum_{\mathbb{M}_i \in \mathcal{M}} f_{\mathbb{M}_i, \mathbb{O}_{i'}^{\delta_q}}$ for $\mathbb{O}_i^{\delta_q}$ will not exceed its observation capacity. Hence, with the help of the characterization, the observation ability constraint (6) will be fulfilled by satisfying the flow conservation and capacity constraints (20).

*3) Communication Constraints Depiction:* Next, we introduce virtual vertices and edges to represent the constraints similar to [48]. Particularly, in each time window $\delta_i \in T$, for any entity $\mathbb{A}^{\delta_i} \in \mathcal{V}$ with more than 1 communication edge, we separate each bidirectional edge $[\mathbb{A}^{\delta_i}, \mathbb{B}^{\delta_i}]$ of $\mathbb{A}^{\delta_i}$ into two directional edges $(\mathbb{A}^{\delta_i}, \mathbb{B}^{\delta_i})$ and $(\mathbb{B}^{\delta_i}, \mathbb{A}^{\delta_i})$. Then we introduce the virtual reception and transmission vertices $\mathbb{A}_{'}^{\delta_i}$ and $\mathbb{A}_{''}^{\delta_i}$ for $\mathbb{A}^{\delta_i}$. Besides, all the incoming directional edges (e.g., $(\mathbb{B}^{\delta_i}, \mathbb{A}^{\delta_i})$) for $\mathbb{A}^{\delta_i}$ are removed from $\mathcal{E}$, the ending vertices (e.g., $\mathbb{B}^{\delta_i}$) of all the incoming edges (e.g., $(\mathbb{B}^{\delta_i}, \mathbb{A}^{\delta_i})$) are linked to $\mathbb{A}^{\delta_i}$'s reception vertex $\mathbb{A}_{'}^{\delta_i}$. The capacity of the new edge is the same as the removed edge's capacity (e.g., $C_{\mathbb{B}^{\delta_i}, \mathbb{A}_{'}^{\delta_i}} = C_{\mathbb{B}^{\delta_i}, \mathbb{A}^{\delta_i}}$). Similarly, the outgoing directional edges (e.g., $(\mathbb{A}^{\delta_i}, \mathbb{B}^{\delta_i})$) from $\mathbb{A}^{\delta_i}$ are also removed, the transmission virtual vertex $\mathbb{A}_{''}^{\delta_i}$ is linked to the head vertices (e.g., $\mathbb{B}^{\delta_i}$) of the removed outgoing edges, and the capacity of the new edges are equal to those of the removed edges (e.g., $C_{\mathbb{A}_{''}^{\delta_i}, \mathbb{B}^{\delta_i}} = C_{\mathbb{A}^{\delta_i}, \mathbb{B}^{\delta_i}}$). Importantly, we link the reception vertex $\mathbb{A}_{'}^{\delta_i}$ with $\mathbb{A}^{\delta_i}$ to construct the *reception constraint edge*, and the capacity of the edge is,

$$C_{\mathbb{A}_{'}^{\delta_i}, \mathbb{A}^{\delta_i}} = R_{\mathbb{A}}(\delta_i). \tag{21}$$

Correspondingly, the *transmission constraint edge* can be obtained through linking $\mathbb{A}^{\delta_p}$ to its transmission vertex $\mathbb{A}_{''}^{\delta_i}$, and its capacity is,

$$C_{\mathbb{A}^{\delta_i}, \mathbb{A}_{''}^{\delta_i}} = S_{\mathbb{A}}(\delta_i). \tag{22}$$

For instance, in Fig. 4, $\mathbb{C}_1$ has multiple communication edges in both $\delta_1$ and $\delta_2$. Take $\mathbb{C}_1$ as an example, we represent the communication constraint edges for $\mathbb{C}_1$ in Fig. 6. Fig. 6(a) shows the original topology of $\mathbb{C}_1^{\delta_1}$ and $\mathbb{C}_1^{\delta_2}$, while Fig. 6(b) shows the details of the constraint edges. Note that the reception and transmission virtual vertices do not have storage edges and we assume the transceiver number $Q_{\mathbb{C}_1} = 2$, and $S_{\mathbb{C}_1}(\delta_2) = R_{\mathbb{C}_1}(\delta_2) = 2 \times 6 = 12$.

With the help of flow conservation on the reception virtual vertex $\mathbb{A}_{'}^{\delta_i}$ and capacity constraints on edge $(\mathbb{A}_{'}^{\delta_i}, \mathbb{A}^{\delta_i})$, we can obtain that

$$\sum_{\mathbb{B}^{\delta_i} \in \mathcal{V}} f_{\mathbb{B}^{\delta_i}, \mathbb{A}_{'}^{\delta_i}} = f_{\mathbb{A}_{'}^{\delta_i}, \mathbb{A}^{\delta_i}} \leq R_{\mathbb{A}}(\delta_i). \tag{23}$$

Thus, the reception constraint (8) for $\mathbb{A}^{\delta_i}$ can be transformed into the normal constraints. Similarly, from the flow conservation of the transmission vertex $\mathbb{A}_{''}^{\delta_i}$ and $(\mathbb{A}^{\delta_i}, \mathbb{A}''^{\delta_i})$'s capacity constraints, the transmission constraint (7) will be met through satisfying the following constraint

$$S_{\mathbb{A}}(\delta_i) \geq f_{\mathbb{A}^{\delta_i}, \mathbb{A}_{''}^{\delta_i}} = \sum_{\mathbb{B}^{\delta_i} \in \mathcal{V}} f_{\mathbb{A}_{''}^{\delta_i}, \mathbb{B}^{\delta_i}}. \tag{24}$$

Hence, the transmission and reception constraints (8) and (7) are also transformed into the normal flow conservation and capacity constraints (23) and (24), respectively.

### C. Maximum Flow Formulation

As a result of the representation, all the constraints for problem **P1** are transformed into common flow conservation constraints and edge capacity constraints. The successfully observed and transmitted data (13) for each mission can be rewritten as follows,

$$F'(\mathbb{M}_i) = f_{\mathbb{M}, \mathbb{M}_i}, \quad \forall \mathbb{M}_i \in \mathcal{M}. \tag{25}$$

As such, problem **P1** can be viewed as finding the maximum flow from the newly inserted vertex $\mathbb{M}$ to the data processing center vertex $\mathbb{D}$. Then, problem **P1** can be reconstructed as follow,

$$\textbf{P2:} \max_{f} \sum_{\mathbb{M}_i \in \mathcal{M}} F'(\mathbb{M}_i) \tag{26}$$
$$s.t., (15), (16), (18), (20), (23) - (25).$$

As a result, we transform the linear programming problem **P1** into a maximum flow one with the help of the graph representation, since all the constraints of **P2** are exactly the same as those of a standard maximum flow problem as in Section 6 of reference [51]. Note that, **P1** and **P2** are equivalent since the satisfaction of all constraints in **P2** indicate the satisfaction of the ones of **P1**, and vice versa. **P2** can be optimally solved by the maximum flow algorithm (e.g., push-relabel algorithm [51]), which has been proved by the max-flow min-cut theorem [51]. Suppose there are $|\mathcal{V}|$ vertices and $|\mathcal{E}|$ edges, the computation complexity of linear programming to solve **P1** is $O(|\mathcal{E}|^{3.5})$ [25], while the

computation complexity of using the maximum flow algorithm as in [51] to solve **P2** is $O(|\mathcal{V}|^2|\mathcal{E}|)$, rendering the algorithm more deployable in real-world systems compared to algorithms for solving original **P1**.

### D. Deployment

We assume all the scheduling decisions are made in a centralized manner, where the state of satellite resources are known. Specifically, all Earth observation requests are submitted to a satellite control center, where our proposed algorithm is run to obtain the scheduling decisions. Next, the satellite control center sends the scheduling decisions to geostationary orbit data relay satellites (e.g., "Tianlian" satellites of China) as in [52], [53], which can provide global coverage. Via the geostationary orbit satellites, the satellite control center can transmit the decisions to all involved observation and communication satellites in near real time using dedicated control channels. Storing the orbit parameters of itself and relevant satellites onboard, a satellite can calculate the rotation angles and other parameters for its cameras or antennas according to the scheduling decisions. To reduce the impact of weather forecasting inaccuracy (e.g., regarding the level of rain attenuation that may effect the satellite to ground station link), the satellite control center can calculate the scheduling decisions periodically (e.g., every 30 minutes) based on its most up-to-date knowledge and upload any update regarding the scheduling decisions to targeted satellites through the geostationary orbit satellites.

## VI. Simulation

### A. Simulation Setup

The simulations are conducted on a satellite network, consisting of 10 "Gaofen" satellites, 40 "starlink" satellites, and 4 ground stations.[2] The 4 ground stations are located at Kashi (39.5°N, 76°E), Sanya (18°N, 109.5°E), Xi'an (34°N, 108°E) and Miyun (40°N,116°E). We created 200 observation missions, corresponding to the capital cities of 200 different countries. For each mission, the area of the observation target is set to be $50 \times 50$ square kilometers. We randomly set the spatial resolution for the Gaofen satellites using value from $\{1.5, 3, 6, 9\}$ meter. For example, one pixel (under the 3 meters spatial resolution setting) represents an area of $3 \times 3$ square meters on the ground. The spectral and radiometric resolution is set as 3 and 8 bits, respectively. As a result, the data volume of an observation mission can be up to $\{25, 6.21, 1.55, 0.69\}$ Gb, depending on the chosen spatial resolution.

The observation and transmission opportunities of the studied satellite network are obtained from the Systems Tool Kit (STK) simulator [54]. For the observation satellites, their observation rate is set as 300Mbps as in [55]. Although our problem formulation allows using time-varying link rate for inter-satellite links and satellite-to-ground-station links, for simplicity, we assume the weather can be perfectly forecast and we use a constant average data rate of 300Mbps as in [55] for both inter-satellite links and satellite-to-ground-station links. Besides, we use the maximum amount of data that can be transmitted in a time window as the edge weight, which is proportional to the length of the time window. Each starlink satellite carries a single transceiver, similar to the setup in [10], [25]. The storage size for each satellite is set to be 100 GB. All simulations are conducted on a Hewlett-Packard Z640 Tower workstation (Intel Xeon E5-2620 v3 2.40GHz, 32GB RAM, O.S. Windows 7 Professional 64bits).

[2]The 10 "Gaofen" satellites contains Gaofen 1, 2, 3, 5, 6, 7, 8, 9-01, 12 and 14. Since the "Hongyan" constellation is still under construction and there are already hundreds of "starlink" satellites in-orbit, we adopt "starlink" satellites for evaluation. The 40 "starlink" satellites are randomly chosen from the system tool kit (STK) database.

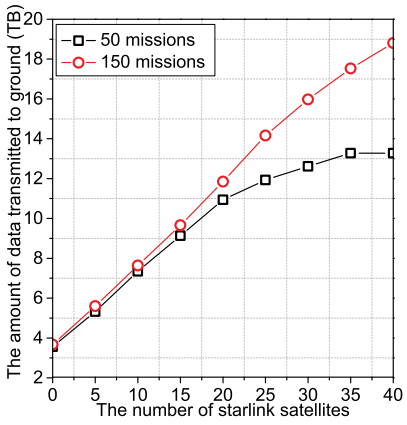


Fig. 7. Throughput versus the starlink satellites number.

### B. Simulation Results and Analysis

*1) Running Time Evaluation:* To show the necessity of our scheme, we evaluate the running time of solving **P1** via standard linear programming solver CVXPY [56] and that of solving **P2** using maximum flow algorithms. We conduct the experiment when the network expands, where the number of involved "starlink" changes from 10 to 80 and the length of $T$ increases from 2 hours to 4 hours. In first column of TABLE II, we use "P1" or "P2" followed by "-" and a number to represent that the corresponding row is the running time of solving **P1** or **P2**, where the number represents the length of $T$. The increase of either the length of $T$ or the number of involved "starlink" satellites leads to a fast increase of running time for solving **P1**, while the time of our scheme remains around 1 second under these settings. When using CVXPY to solve **P1**, "-" means that we encounter memory overflow issues and cannot obtain the final results as the network size increases to some extent. Besides, when length of $T$ is 2 hours, the running time of solving **P1** is irregularly large when 10 and 40 "starlink" satellites are involved, which shows that the linear programming algorithm is unstable. Particularly, we solve **P2** when length of $T$ reaches 24 hours as highlighted in TABLE II, the running time only reaches 20 seconds. This experiment shows the memory efficiency and computational efficiency of our scheme. In contrast, the standard solution for solving **P1** is not practical when the network size becomes large.

*2) The Performance Evaluation by Using Inter-Satellite Communications:* In Fig. 7, we answer the question about how much throughput gain SN can achieve through introducing

TABLE II
THE RUNNING TIME VERSUS DIFFERENT NUMBER OF INVOLVED "STARLINK" SATELLITES

| "starlink" number | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
|---|---|---|---|---|---|---|---|---|
| **P1**-2h | 286s | 36s | 51s | 1531.6s | 131s | 162s | 221s | 370s |
| **P2**-2h | 0.3s | 0.5s | 0.7s | 0.7s | 0.7s | 0.6s | 0.5s | 0.6s |
| **P1**-3h | 62s | 84s | 114s | 184s | 328s | 483s | 671s | - |
| **P2**-3h | 0.6s | 0.9s | 1.2s | 1.2s | 0.7s | 0.7s | 0.7s | 0.9s |
| **P1**-4h | 116s | 160s | 254s | 281s | 653s | - | - | - |
| **P2**-4h | 0.8s | 1.3s | 1.6s | 1.9s | 1s | 1s | 1.2s | 1.3s |
| **P2**-24h | 8.6s | 14.9s | 21s | 22s | 16s | 10.8s | 11.4s | 12.0s |

starlink satellites. The scheduling time horizon $T$ is set to be $T = 24$ hours. Two scenarios are constructed, where 50 missions and 150 mission are respectively inserted from the beginning of $T$ and carried out repeatedly. Recall that **P2** models both the on-demand scheduling and co-optimization of observation and transmission, we evaluate the throughput of SN by solving **P2** with the involved number of starlink satellites varying from 0 to 40 with a step of 5 under the 2 mission scenarios.

As shown in Fig. 7, the throughput of SN increases as the number of starlink satellite increases. This is because the communication capacity increases as the number of starlink satellites increases. For the 50-mission scenario, the throughput eventually stops increasing when the number of starlink satellites reaches 35, this is because all the missions can be transmitted. In comparison, the throughput of 150-mission scenario keeps growing as the number of starlink satellites increases, this shows there are missions that are not fully transmitted even with 40 satellites introduced. It is worth noting that, the Gaofen satellites themselves can only transmit back around 30000 Gb data. However, with the help of 10 starlink satellites, the throughput increases to 2 times. It further increases to 3 times when 20 starlink satellites are used. Finally, the throughput when 40 starlink satellites are introduced in 150-mission scenario is 5 times that of the throughput when no inter-satellite communication is used.

Note that the performance gain reported here is based on the assumption that both the on-demand inter-satellite communication scheduling and the co-optimization of observation and transmission tasks are supported. Next, we will study the performance gain introduced by the on-demand inter-satellite communication scheduling and the co-optimization of observation and transmission missions, respectively.

*3) The Effects of the Flexibility of Inter-Satellite Communication Schedule on Throughput:* Recall that our **P2** formulation models the on-demand scheduling. To compare it with pre-determined scheduling schemes, we consider the following 4 different schemes as baselines. The first two are fair-scheduling schemes proposed in [28]. Under the first *probabilistic-fair scheme*, when multiple transmission opportunities compete for the transceiver resource, one transmission opportunity is randomly chosen. A certain level of fairness is achieved by choosing among all candidate transmission opportunities with equal probability. In comparison, under the second *deterministic-fair scheme*, among all competing transmission opportunities, the one that has been selected the least number of times will be chosen.

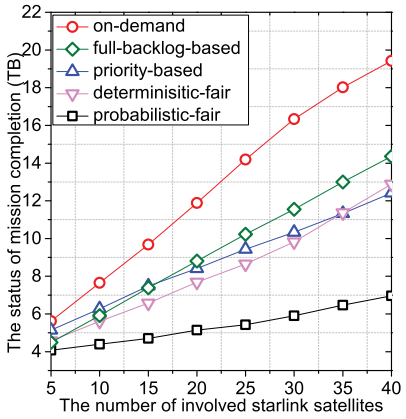


Fig. 8. Throughput of different scheduling methods.

Besides these two schemes proposed in [28], we further propose the following two heuristic-based pre-determined scheduling schemes as our baselines. Our third scheme is a *priority-based scheme*, where the transmission opportunities with the ground station and that between a communication satellite and an observation satellite are given higher priority compared to those between two communication satellites. More specifically, an opportunity with higher priority will always be scheduled before those with lower priority. If there are multiple opportunities with the same level of priority, equal bandwidth will be allocated among them in a pre-determined manner. Our fourth baseline is a *full-backlog-based scheduling scheme*. Under this scheme, we assume the actual observation missions are not known by the scheduling schemes, instead, the scheduler takes a pessimistic approach by assuming there are always backlogs in all the observation satellites. Specifically, we conduct on-demand scheduling based on the assumption that there is always traffic backlog at all the observation satellites and pre-determine the transmission opportunities so as to optimize the throughput under such a setting.

In the following, we will compare the on-demand scheduling scheme as formulated in **P2** with the four baseline scheduling methods. In our experiment, we set the scheduling time horizon to be 24 hours.

In Fig. 8, all the 200 observation missions are introduced from the beginning of the time horizon, and each of them needs to be repeatedly carried out over the 24 hours. This aims to continuously create high demand to the studied earth observation system. We evaluate the total amount of data volume that is transmitted back to the ground over the

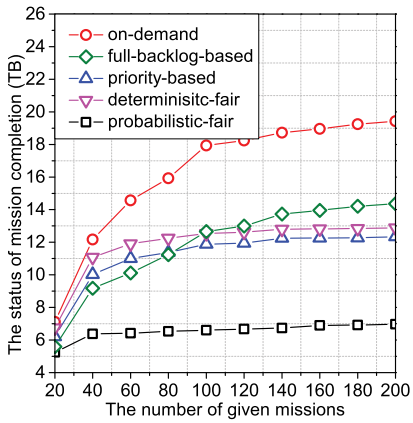


Fig. 9. Throughput versus different missions number.

time horizon, by varying the number of involved starlink satellites from 5 to 40. As shown in the figure, under all the 5 schemes, the total volume transmitted increases with the number of involved starlink satellites increases. This is expected as the transmission capacity of the network increases with more communication satellites. Among the four baseline pre-determined schemes, the probabilistic-fair communication schedule has the lowest throughput, while the full-backlog-based scheme achieves the highest throughput. This is due to the fact that the full-backlog-based scheme can select paths from observation satellites to ground stations even though it just uses an assumed traffic pattern, while all the other three schemes are less effective in doing that. However, as the traffic pattern assumed by the full-backlog-based scheme can be different from the actual one, it may also select transmission opportunities that are sub-optimal, i.e., cannot be used. As expected, the on-demand topology schedule outperforms all the other 4 schemes. As shown in the figure, it can further increase the throughput by around 50%, comparing with the best pre-determined scheduling scheme (i.e., the full-backlog-based scheme).

Next, we study the maximum mission data volume that the 5 different schemes can accommodate using 40 starlink satellites. We start from 20 missions, and increase the load by adding 20 missions each time. As shown in Fig. 9, the data volume transmitted by all the schemes increases with the increasing number of missions. However, the increase rate slows down as the total data volume increases, which eventually approaches some limit. This happens as the increasing mission loads occupies most of the available communication resources. Our on-demand scheme consistently outperforms all the 4 pre-determined schemes. The percentage of gain is smaller when the number of missions is smaller (e.g., at 20 or 40), this is because under such a light load, most schemes (except for the probabilistic-fair scheme) provides abundant transmission opportunities to support the missions. Once the load increases beyond 60 missions, the throughput gain provided by on-demand scheduling over the best performing pre-determined scheme becomes more significant and increases to around 50% for larger number of missions.

In Fig. 10, all the 200 missions are introduced from the beginning of $T$ and repeatedly carried out. Each method's name is followed by a "-" and a number (e.g., on-demand-40) in the legend of Fig. 10(a) to indicate the number of "starlink" satellites that are involved while conducting that method. To understand the source of the throughput gain, we examine the number of paths with different number of hops under the 5 schemes. As shown in Fig. 10(a), the number of paths when 40 starlink satellites are involved is much more than that with only 20 starlink satellites, this is because fewer satellites provide fewer communication opportunities. When the path-hops are 0 (i.e., when Gaofen satellites directly transmit data to ground stations), the path number of different schemes are equal, because the observation satellites preferentially communicate with ground stations in all 5 schemes. In addition, under all schemes, the number of paths always decreases as the number of hops increases. They become rather small when the number of hops is greater than 3. Note that the number of 2-hops and 3-hops paths are both considerable, different from [10]. In particular, *on-demand scheme* arrange more paths in each kind of paths(i.e., paths with different hops) than other methods, which explains why *on-demand scheme* outperforms other schemes, as shown in Fig. 8.

Fig. 10(b) plots the total amount of data volume delivered via paths of different number of hops, under all the five schemes. The trend in Fig. 10(b) is similar to that of 10(a), which is expected as each path accommodates similar amount of data volume as in our setting.

*4) The Throughput Gain Due to the Co-Optimization of Observation and Transmission:* Recall that **P2** models the co-optimization of observation and transmission, we call it a *simultaneous scheduling* scheme. To compare it with independent observation and transmission scheduling, we design a *sequential scheduling* scheme as follows. For the observation phase, we adopt the maximum flow algorithm [51] to maximize the data volume that can be observed by all the observation satellites. Next, we use the observed data of each observation satellite as the traffic input of SN. Again, we use the maximum flow algorithm [51] to maximize the amount of data transmitted to ground stations.

In TABLE III, we evaluate a few different settings for a time horizon of 24 hours and with 15 starlink satellites. We evaluate the throughput of SN with the number of missions varying from 10 to 200. These mission are inserted at the beginning of $T$ and repeatedly carried out. As shown in the Table, when the satellites can store data, the advantage of *simultaneous scheduling* is rather limited, as shown in the 3rd and 4th rows of TABLE III. However, when we set the storage size of satellites to 0 (to simulate the observation of urgent missions) *simultaneous scheduling* improves the throughput of *sequential scheduling* by 42.5% when 10 missions are inserted. As the number of missions increase, the gap between the throughput of *simultaneous scheduling* and the *sequential scheduling* sharply narrows. This trend can be explained as the observation scheduling is no longer critical when the missions increase to some extent (i.e., 20). In addition, the throughput does not increase much and fast as the number of missions increases from 40 to 200. This is because the transmission

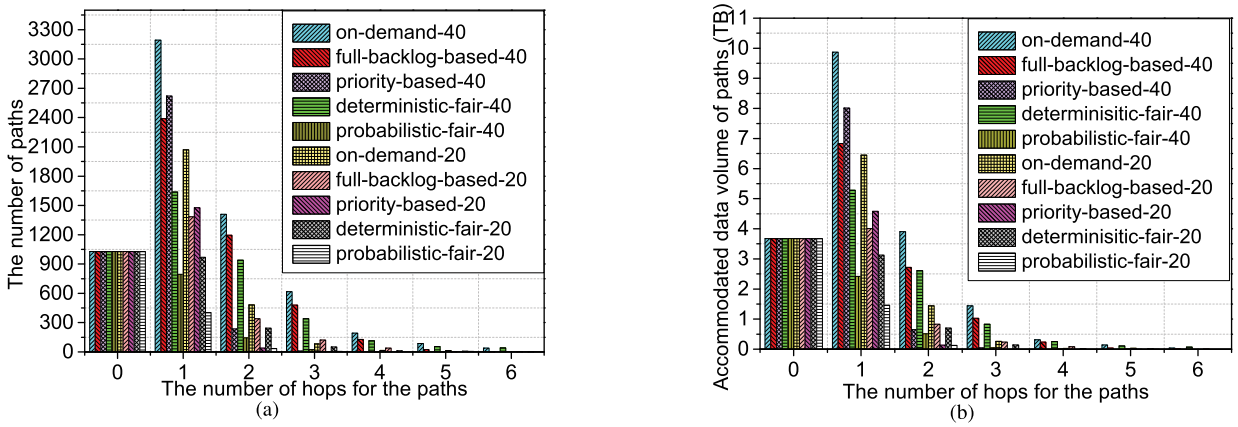


Fig. 10. (a) The number of paths with different number of hops, used by different schemes; (b) The amount of data transmitted via paths with different number of hops, under different schemes.

TABLE III

THE THROUGHPUT OF THE TWO SCHEMES WITH HELP OF 15 STARLINK SATELLITES(UNIT: GB)

| mission number | 10 | 20 | 40 | 60 | 80 | 100 | 200 |
|---|---|---|---|---|---|---|---|
| sequential scheduling without storage | 12700 | 26144 | 41840 | 42952 | 45264 | 48016 | 53000 |
| simultaneous scheduling without storage | 18096 | 27328 | 42160 | 43136 | 45416 | 48104 | 53024 |
| sequential scheduling with storage | 20536 | 31616 | 47600 | 48960 | 52960 | 57240 | 63688 |
| simultaneous scheduling with storage | 20896 | 31920 | 47944 | 49160 | 53160 | 57416 | 63760 |

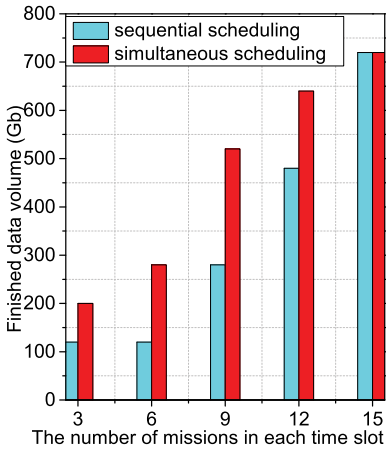


Fig. 11. The amount of transmitted data versus the number of given missions.

capacity of SN becomes the bottleneck when the mission load becomes heavy (e.g., beyond 40 missions). Importantly, the throughput of *simultaneous scheduling* and *sequential scheduling* are consistently the same when the mission number is beyond 20, this shows that the *simultaneous scheduling* can no longer improve the throughput of *sequential scheduling* when the mission load increases to some extent.

To further study the impact of these factors, in the following we focus on a light-load scenario with a smaller number of time-critical missions. We simulate time-critical missions by not allowing any storage in the satellite. Further, we reduce the time horizon to 1 hour. We evaluate the throughput of the two schemes with the number of missions varying from 3 to 15 with a step of 3. As shown in Fig. 11, when the number of missions increases, the throughput gap between the two schemes first becomes larger, then the gap becomes smaller until disappearing at 15 missions. To understand this, let us first look at what causes the gap between the two schemes. As the sequential scheduling does not consider the transmission constraints when scheduling the observation, it could result in the case that some Gaofen satellites that do not have sufficient transmission opportunities are assigned the observation opportunities, while the other Gaofen satellites are underused. When the mission load is very light, the transmission opportunities are relatively abundant, so the gap is small. When the network is under a medium load (e.g., when there are 9 missions), *sequential scheduling* can choose the Gaofen satellites that cause much congestion in the transmission phase, while the *simultaneous scheduling* can better coordinate that. Under such settings, the *simultaneous scheduling* can increase the throughput of *sequential scheduling* by more than $100\%$. With more missions entering, the throughput of *sequential scheduling* gradually catches up with that of the *simultaneous scheduling* as the transmission opportunity becomes the sole bottleneck, and the observation scheduling becomes less critical as long as the transmission opportunities can be fully utilized.

*5) The Throughput Gain Under Varying Storage Size Constraints and Mission Completion Delay Constraints:* Inserting 200 missions at the beginning of $T$, we first evaluate the effects of different satellite storage size on the mission completion ratio. As shown in Figure 12, the satellites storage varies from 0 to 350 Gb. Note that the two numbers connected by '-' in legend represent the number of involved "starlink" satellites and the length of $T$, respectively. Overall, as the storage size increases, the mission completion ratio increases little and

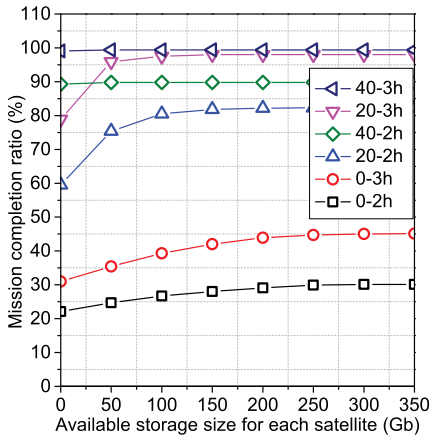


Fig. 12. The curves of mission completion ratio versus different storage size.

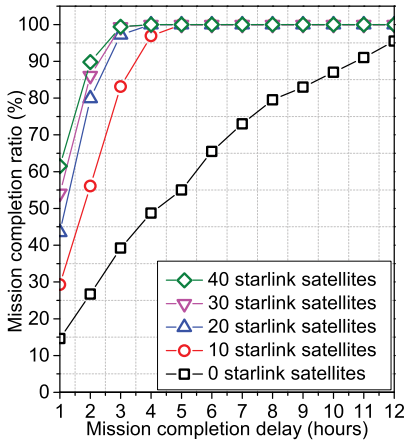


Fig. 13. The curves of mission completion ratio versus different delay.

slowly when no "starlink" involves. With the involvement of 20 "starlink", mission completion ratio increases a lot and stops increasing when the size reaches 150Gb. Particularly, the increase is negligible with the help of 40 "starlink" satellites. In addition, when no "starlink" satellites help, the mission completion ratio increases more when the length of $T$ is bigger. It is worth noting that when no "starlink" satellites help, it requires 250Gb satellite storage to increase mission completion ratio by less than 15%. Interestingly, with 20 "starlink" satellites, the mission completion ratio can be increased by more than 20% with only 100Gb storage size. The earth observation system no longer need storage resources when the communication capacity is high.

Given 200 missions at the beginning of $T$, we evaluate their delay. We adopt our on-demand scheduling scheme for the evaluation, where 10 "Gaofen" observation satellites and different number of "starlink" satellites (ranging from 10 to 40) are used. Fig. 13 plots how the mission completion ratio increases with the allowed time to complete them. In particular, 90% missions can be completed with less than 2 hours of delay in the network with 40 "starlink" satellites involved. In comparison, it takes 11 hours to complete 90% missions without the help of "starlink" satellites. If we compare the number of completed missions under a given delay requirement, e.g., 2 hours, the network with 40 "starlink" satellites can complete 90% missions while the network without "starlink" satellites can only complete less than 30% of missions. The reason behind above phenomenon is that without or with few "starlink" satellites, an observation satellite that is not in direct contact with any ground station can only wait to fly over the ground stations to offload its observed data. On the contrary, the observation satellite does not need to wait and can directly relay its observed data to ground stations via multiple "starlink" satellites.

## VII. Conclusion

In this work, we study the problem of using inter-satellite communication to enhance the throughput of earth observation systems. We identified two important design factors, i.e., the capability to provide on-demand inter-satellite communication scheduling and the capability of conducting co-optimization of transmission and observation scheduling. We formulate the throughput optimization problem under both constraints as a linear programming problem. We then construct a generalized time-varying graph and introduce auxiliary constraint edges to help solve the formulated problem efficiently. Our simulation results based on the "starlink" and "Gaofen" constellation show that both factors, if enabled, can significantly enhance the throughput of an earth observation system. In our future work, we will study earth observation scheduling problems that involve more real-world constraints, such as different QoS requirements of missions.

## References

[1] J. Du, C. Jiang, Q. Guo, M. Guizani, and Y. Ren, "Cooperative earth observation through complex space information networks," *IEEE Trans. Wireless Commun.*, vol. 23, no. 2, pp. 136–144, Apr. 2016.

[2] Union of Concern Scientists. *UCS Satellite Database*. Accessed: Dec. 10, 2020. [Online]. Available: https://www.ucsusa.org/resources/satellite-database.

[3] R. Tomás and Z. Li, "Earth observations for geohazards: Present and future challenges," *Remote Sens.* vol. 9, no. 3, p. 194, Mar. 2017.

[4] A. S. Belward and J. O. Skøien, "Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites," *ISPRS J. Photogramm. Remote Sens.*, vol. 103, pp. 115–128, May 2015.

[5] H. K. Ramapriyan, "The role and evolution of NASA's earth science data systems," NASA Goddard Space Flight Center, Greenbelt, MD, USA, Tech. Rep. GSFC-E-DAA-TN24713, 2015.

[6] V. Gomes, G. Queiroz, and K. Ferreira, "An overview of platforms for big earth observation data management and analysis," *Remote Sens.*, vol. 12, no. 8, p. 1253, Apr. 2020.

[7] P. Soille, A. Burger, D. De Marchi, P. Kempeneers, D. Rodriguez, V. Syrris, and V. Vasilev, "A versatile data-intensive computing platform for information retrieval from big geospatial data," *Future Gener. Comput. Syst.*, vol. 81, pp. 30–40, Apr. 2018.

[8] Jonathan. *Jonathan's Space Pages*. Accessed: Jul. 5, 2021. [Online]. Available: https://planet4589.org/space/stats/star/starstats.html

[9] D. Yang, J. Yang, G. Li, Y. Zhou, and C. Tang, "Globalization highlight: Orbit determination using BeiDou inter-satellite ranging measurements," *GPS Solutions*, vol. 21, no. 3, pp. 1395–1404, Jul. 2017.

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

[11] *High Speed Internet Access Across the Globe*. Accessed: Oct. 5, 2020. [Online]. Available: https://www.starlink.com/

[12] *China's New Space Race: First Satellite of Casc's Hongyan Leo Satcom Constellation to Launch by End of 2018*. Accessed: Oct. 5, 2020. [Online]. Available: https://spacewatch.global/

[13] NASA. *State of the Art of Small Spacecraft Technology*. Accessed: Jul. 5, 2021. [Online]. Available: https://www.nasa.gov/

[14] M. Vasquez and J.-K. Hao, "A 'logic-constrained' knapsack formulation and a tabu algorithm for the daily photograph scheduling of an earth observation satellite," *Comput. Optim. Appl.*, vol. 20, no. 2, pp. 137–157, 2001.

[15] V. Gabrel and D. Vanderpooten, "Enumeration and interactive selection of efficient paths in a multiple criteria graph for scheduling an earth observing satellite," *Eur. J. Oper. Res.*, vol. 139, no. 3, pp. 533–542, 2002.

[16] W.-C. Lin, D.-Y. Liao, C.-Y. Liu, and Y.-Y. Lee, "Daily imaging scheduling of an earth observation satellite," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 35, no. 2, pp. 213–223, Mar. 2005.

[17] D. Y. Liao and Y. T. Yang, "Imaging order scheduling of an earth observation satellite," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 37, no. 5, pp. 794–802, Sep. 2007.

[18] G. Wu, M. Ma, J. Zhu, and D. Qiu, "Multi-satellite observation integrated scheduling method oriented to emergency tasks and common tasks," *J. Syst. Eng. Electron.*, vol. 23, no. 5, pp. 723–733, Oct. 2012.

[19] G. Wu, J. Liu, M. Ma, and D. Qiu, "A two-phase scheduling method with the consideration of task clustering for earth observing satellites," *Comput. Oper. Res.*, vol. 40, no. 7, pp. 1884–1894, Jul. 2013.

[20] P. Li, J. Li, H. Li, S. Zhang, and G. Yang, "Graph based task scheduling algorithm for earth observation satellites," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–7.

[21] R. Xu, H. Chen, X. Liang, and H. Wang, "Priority-based constructive algorithms for scheduling agile earth observation satellites with total priority maximization," *Expert Syst. Appl.*, vol. 51, pp. 195–206, Jun. 2016.

[22] 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.

[23] Y. Du, T. Wang, B. Xin, L. Wang, Y. Chen, and L. Xing, "A data-driven parallel scheduling approach for multiple agile earth observation satellites," *IEEE Trans. Evol. Comput.*, vol. 24, no. 4, pp. 679–693, Aug. 2020.

[24] J. A. Fraire and J. M. Finochietto, "Design challenges in contact plans for disruption-tolerant satellite networks," *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 163–169, May 2015.

[25] 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. Commu.*, vol. 65, no. 6, pp. 2451–2466, Mar. 2017.

[26] J. A. Fraire, P. G. Madoery, and J. M. Finochietto, "On the design and analysis of fair contact plans in predictable delay-tolerant networks," *IEEE Sensors J.*, vol. 14, no. 11, pp. 3874–3882, Nov. 2014.

[27] J. Fraire and J. M. Finochietto, "Routing-aware fair contact plan design for predictable delay tolerant networks," *Ad-Hoc Netw.*, vol. 25, pp. 303–313, Feb. 2015.

[28] Z. Yan *et al.*, "Distributed contact plan design for GNSSs," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 1, pp. 660–672, Feb. 2020.

[29] J. A. Fraire, P. G. Madoery, and J. M. Finochietto, "Traffic-aware contact plan design for disruption-tolerant space sensor networks," *Ad Hoc Netw.*, vol. 47, pp. 41–52, Sep. 2016.

[30] Y. Wang, M. Sheng, W. Zhuang, S. Zhang, N. Zhang, R. Liu, and J. Li, "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.

[31] G. Araniti *et al.*, "Contact graph routing in DTN space networks: Overview, enhancements and performance," *IEEE Commun. Mag.*, vol. 53, no. 3, pp. 38–46, Mar. 2015.

[32] CCSDS. *Schedule-Aware Bundle Routing*. Accessed: Jul. 5, 2021. [Online]. Available: https://public.ccsds.org/Publications/BlueBooks.aspx

[33] T. Zhang, H. Li, S. Zhang, J. Li, and H. Shen, "STAG-based QoS support routing strategy for multiple missions over the satellite networks," *IEEE Trans. Commun.*, vol. 67, no. 10, pp. 6912–6924, Oct. 2019.

[34] H. Li, T. Zhang, Y. Zhang, K. Wang, and J. Li, "A maximum flow algorithm based on storage time aggregated graph for delay-tolerant networks," *Ad Hoc Netw.*, vol. 59, pp. 63–70, May 2017.

[35] Y. Xiao, S. Zhang, P. Yang, M. You, and J. Huang, "A two-stage flow-shop scheme for the multi-satellite observation and data-downlink scheduling problem considering weather uncertainties," *Rel. Eng. Syst. Saf.*, vol. 188, pp. 263–275, Aug. 2019.

[36] M. Sheng, D. Zhou, R. Liu, Y. Wang, and J. Li, "Resource mobility in space information networks: Opportunities, challenges, and approaches," *IEEE Netw.*, vol. 33, no. 1, pp. 128–135, Jan. 2019.

[37] A. Golkar and I. L. I. Cruz, "The federated satellite systems paradigm: Concept and business case evaluation," *Acta Astron.*, vol. 111, pp. 230–248, Jun. 2015.

[38] 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.

[39] *Propagation Data and Prediction Methods Required for the Design of Earth-Space Telecommunication Systems*, document Rec. ITU-R 12-618, 2015.

[40] *Rain Height Model For Prediction Methods*, document ITU-R 839-3-2001, 2001.

[41] *Specific Attenuation Model For Use in Prediction Methods*, document ITU-R 839-3-2005, 2005.

[42] A. Destounis and A. D. Panagopoulos, "Dynamic power allocation for broadband multi-beam satellite communication networks," *IEEE Commun. Lett.*, vol. 15, no. 4, pp. 380–382, Apr. 2011.

[43] A. Paraboni *et al.*, "Meteorology-driven optimum control of a multibeam antenna in satellite telecommunications," *IEEE Trans. Antennas Propag.*, vol. 57, no. 2, pp. 508–519, Mar. 2009.

[44] J. Li, P. Wang, H. Li, and K. Shi, "Enhanced time-expanded graph for space information network modeling," *Sci. China Inf. Sci.*, to be published.

[45] L. A. Wolsey and G. L. Nemhauser, *Integer Combination Optimization*, vol. 55. Hoboken, NJ, USA: Wiley, 1999.

[46] G. Wu *et al.*, "An adaptive simulated annealing-based satellite observation scheduling method combined with a dynamic task clustering strategy," *Comput. Sci.*, 2014.

[47] R. Liu, W. Wu, Q. Yang, D. Zhou, and W. Zhang, "Exploring the information capacity of remote sensing satellite networks," *IEEE Access*, vol. 8, pp. 34056–34070, 2020.

[48] P. Wang, X. Zhang, S. Zhang, H. Li, and T. Zhang, "Time-expanded graph-based resource allocation over the satellite networks," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 360–363, Apr. 2019.

[49] J. Ren, H. Zhou, N. Zhou, and Q. Liu, "Application of phased array antenna and fixed multibeam antenna in communications satellite systems," *Space Int.*, vol. 443, no. 11, pp. 55–60, 2015.

[50] X. Chen and J. Wan, "Development status and proposals for multi-beam antennas of communications satellites," *Space Electron. Technol.*, vol. 13, no. 2, pp. 54–60, 2016.

[51] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, "Network flows: Theory, algorithms, and applications," *J. Oper. Res. Soc.*, vol. 45, no. 11, pp. 791–796, 1993.

[52] D. Fan, M. Qiu, X. Xu, and M. Chen, "Research on the integration demonstration system of space information network based on TDRSS," *J. Phys. Conf. Ser.*, vol. 1187, no. 5, 2019, Art. no. 052025.

[53] J. Wang, "Proposal for developing China's data relay satellite system," *Spacecraft Eng.*, vol. 20, no. 2, pp. 1–8, 2011.

[54] An Ansys Company. *The Synthesis Toolkit (STK)*. Accessed: Jul. 5, 2021. [Online]. Available: https://www.agi.com/products/stk

[55] W. Fu, J. Ma, P. Chen, and F. Chen, "Remote sensing satellites for digital earth," in *Manual of Digital Earth*. Singapore: Springer, 2020, pp. 55–123.

[56] S. Diamond and S. Boyd, "CVXPY: A Python-embedded modeling language for convex optimization," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2909–2913, Jan. 2016.

**Peng Wang** received the B.Eng. degree in telecommunication engineering from Xidian University, Xi'an, China, in 2017, where he is currently pursuing the Ph.D. degree with the State Key Laboratory of Integrated Service Networks.

Since 2021, he has been a Visiting Ph.D. Student with the Pillar of Information System of Technology and Design, Singapore University of Technology and Design, Singapore. His research interests include satellite networks, time-varying graph theory, routing, and scheduling algorithms.

**Hongyan Li** (Member, IEEE) received the M.S. degree in control engineering from Xi'an Jiaotong University, Xi'an, China, in 1991, and the Ph.D. degree in signal and information processing from Xidian University, Xi'an, in 2000.

She is currently a Professor with the State Key Laboratory of Integrated Service Networks, Xidian University. Her research interests include deterministic networks, computation-aware networks, spatial information networks, cognitive networks, and integration of heterogeneous networks.

**Binbin Chen** (Member, IEEE) received the B.Sc. degree in computer science from Peking University and the Ph.D. degree in computer science from the National University of Singapore. He is currently an Associate Professor with the Pillar of Information Systems Technology and Design (ISTD), Singapore University of Technology and Design (SUTD). His current research interests include wireless networks, cyber-physical systems, and cyber security for critical infrastructures.

**Shun Zhang** (Senior Member, IEEE) received the B.S. degree in communication engineering from Shandong University, Jinan, China, in 2007, and the Ph.D. degree in communications and signal processing from Xidian University, Xi'an, China, in 2013.

He is with the State Key Laboratory of Integrated Services Networks, Xidian University, where he is currently a Professor. His research interests include massive MIMO, millimeter wave systems, RIS assisted communications, deep learning for communication systems, orthogonal time frequency space (OTFS) systems, and multiple access techniques. He has authored or coauthored more than 80 journal and conference papers. He is the inventor of 16 granted patents (including a PCT patent authorized by U.S. patent and trademark office). He has received two best paper awards in conferences and the two Prize Award from Natural Sciences for Research Excellence by both the China Institute of Communications and the Chinese Institute of Electronics. He is an Editor for *Physical Communication*.