Collaborative Data Scheduling With Joint Forward and Backward Induction in Small Satellite Networks

Di Zhou, Min Sheng, *Senior Member, IEEE*, Jie Luo, *Senior Member, IEEE*, Runzi Liu, *Member, IEEE*, Jiandong Li, *Senior Member, IEEE*, and Zhu Han, *Fellow, IEEE*

***Abstract*— Small satellite networks (SSNs) have attracted intensive research interest recently and have been regarded as an emerging architecture to accommodate the ever-increasing space data transmission demand. However, the limited number of on-board transceivers restricts the number of feasible contacts (i.e., an opportunity to transmit data over a communication link), which can be established concurrently by a satellite for data scheduling. Furthermore, limited battery space, storage space, and stochastic data arrivals can further exacerbate the difficulty of the efficient data scheduling design to well match the limited network resources and random data demands, so as to the long-term payoff. Based on the above motivation and specific characteristics of SSNs, in this paper, we extend the traditional dynamic programming algorithms and propose a finite-embedded-infinite two-level dynamic programming framework for optimal data scheduling under a stochastic data arrival SSN environment with joint consideration of contact selection, battery management, and buffer management while taking into account the impact of current decisions on the infinite future. We further formulate this stochastic data scheduling optimization problem as an infinite-horizon discrete Markov decision process (MDP) and propose a joint forward and backward induction algorithm framework to achieve the optimal solution of the infinite MDP. Simulations have been conducted to demonstrate the significant gains of the proposed algorithms in the amount of downloaded data and to evaluate the impact of various network parameters on the algorithm performance.**

***Index Terms*— Small satellite networks, stochastic data arrival, dynamic data scheduling, Markov decision process, resource management.**

Manuscript received October 26, 2018; revised January 26, 2019; accepted February 11, 2019. Date of publication February 19, 2019; date of current version May 15, 2019. This work was supported in part by the Natural Science Foundation of China under Grant 91638202, Grant 61725103, and Grant 61701365; in part by the Key R&D Program – The Key Industry Innovation Chain of Shaanxi under Grant 2017ZDCXL-GY-04-04; and in part by US MURI AFOSR MURI 18RT0073, NSF CNS-1717454, CNS- 1731424, CNS-1702850, CNS-1646607. The associate editor coordinating the review of this paper and approving it for publication was J. Luo. *(Corresponding author: Min Sheng.)*

D. Zhou, M. Sheng, R. Liu, and J. Li are with the State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China (e-mail: dizhou@stu.xidian.edu.cn; msheng@mail.xidian.edu.cn; jdli@mail.xidian.edu.cn; rzliu@xidian.edu.cn).

J. Luo is with the Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO 80523 USA (e-mail: rockey@colostate.edu).

Z. Han is with the Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77004 USA, and also with the Department of Computer Science and Engineering, Kyung Hee University, Seoul 446-701, South Korea (e-mail: zhan2@uh.edu).

Color versions of one or more of the figures in this paper are available online at http:ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2019.2900316

I. Introduction

A. Motivation

THE increasing worldwide demand for space exploration and research is creating new opportunities for the deployment of small satellites in low earth orbit (LEO). Small satellites have become fashionable and are catalyzing new applications and business models [1]. Small satellites, deployed as a sensor network in LEO, can further form a small satellite network (SSN) [2]. The formed SSN has an advantage of handling burst missions in near-real-time because of its potential to perform coordinated observations and data offloading using inter-satellite (IS) links before they come to the contact (i.e., an opportunity to send data over a communication link) with a ground station [3], [4]. Consequently, SSN enabled IS links is considered as a promising solution to download the massive on-board data and very popular in environment surveillance, intelligence reconnaissance, and target surveillance [5], [6].

One important problem in SSNs is to deliver the data from areas of interest collected by small satellites to ground stations with the objective of matching the network resources and data arrival demand. It remains technically challenging to develop efficient data scheduling strategies for SSNs, due to the following two aspects: 1) Random and bursty data arrival. The data arrival in SSNs is often unpredictable in real-world applications, which means that the accurate data arrival information is often unavailable. Therefore, dynamic data scheduling planning is needed to accommodate the random and bursty data arrival, which further exacerbate the difficulty in designing efficient data scheduling planning. 2) Limited and dynamic network resources. On one hand, the inherent highly dynamic network topology can inevitably result in intermittent transmission windows for IS contacts and satellite downlink (SD) contacts for data transmission. On the other hand, to reduce the size and cost of the satellite payload, the on-board resources for each satellite is limited. Specifically, each satellite is equipped with a limited number of transceivers for communicating with its neighboring satellites concurrently [7]. Besides, the on-board storage device embedded in each satellite has limited capacity for storing data while they await their transmission [8]. Furthermore, batteries with limited capacity are installed on satellites for establishing communication links for data transmission and reception [9].

Fortunately, a satellite has the ability to harvest solar energy and store energy in the rechargeable battery to supply the forthcoming communications [10]. However, since no energy can be generated during eclipse periods, there may not be persistent energy supply from solar panels [11].

The data scheduling strategy is dedicated to efficiently matching the on-board network resources in the network with the uneven and stochastic data arrivals to maximize the long-term network performance. However, the aforementioned practical issues in SSNs show that the design of the data scheduling strategies for stochastically arrived data is influenced by a couple of factors such as the finite transmitters, battery capacity, storage capacity, dynamic energy supply from solar panels at satellites, and the time-varying IS channel conditions. It is therefore highly non-trivial to design efficient data scheduling strategies to well match the dynamic multi-dimensional resources to stochastically arrived space data.

B. Literature Review

There is a growing research interest in small satellites to support integrated satellite-terrestrial systems [12], [13] and some emerging applications such as internet of things (IoTs) [14]. Due to the intermittent IS and SD contacts, the store-carry-forward paradigm in delay tolerant networks (DTN) is utilized for data transmission [15] in SSNs, wherein each individual node is assumed to be ready to forward data for others. However, due to the existence of selfish or even malicious nodes, the expected link may not be established normally for data transmission [16]. Therefore, some countermeasures are proposed to deal with such potential security attacks [16]–[18]. Specifically, a secure multilayer credit-based incentive scheme was proposed in [16] to stimulate bundle forwarding cooperation among DTN nodes. Besides, Zhu *et al.* [17] devised a novel jammer inference-based jamming defense framework to enhance the effectiveness of a series of the proposed anti-jamming strategies. By exploiting the buffering characteristic in DTN nodes, an opportunistic batch bundle authentication scheme was proposed in [18] to achieve efficient bundle authentication.

On the basis of the store-carry-forward paradigm for data transmission, some efforts have been focusing on the design of data scheduling strategies to efficiently exploit the limited network resources, which are mainly classified into two main categories. The first kind of algorithm focuses on the data scheduling planning for predictable data arrival [19]–[21]. Specifically, a primal decomposition method was proposed in [19] to efficiently utilize the energy resource to achieve high network profit. Besides, recent work [20] proposed a collaborative data scheduling scheme to achieve optimal throughput by jointly scheduling data offload among the satellites and data downloading from satellites to the ground station. Furthermore, based on time-evolving resource graph, an optimal resource allocation strategy was proposed in [21] to facilitate efficient cooperation among various resources. However, these algorithms are static optimization and cannot be applied directly to the data scheduling design with the stochastic data arrival. Therefore, the second kind of algorithm devises data scheduling planning with random data arrival [22], [23]. Considering the future data arrivals, a predictive back-pressure algorithm was proposed in [22] for efficient resource allocation to minimize the time average cost of the system. Moreover, recent work [23] proposed an optimal data scheduling scheme based on Lyapunov optimization to maximize the downloaded data for the randomly arrived data. However, all these existing algorithms make a finite-horizon data scheduling planning, which leads to the fact that they cannot be directly applied to solve the infinite-horizon stochastic data scheduling problem in SSNs.

Such existing finite-horizon data scheduling planning algorithms neglect its corresponding impact on the next planning horizon, which may lead to a poor network resource status (e.g., energy and storage status) at the end of the current planning horizon to support the stochastically arrived data in the next planning horizon. Therefore, dynamic programming algorithms, such as Markov decision process (MDP) and semi-Markov decision process (SMDP) which have been extensively investigated for scheduling and resource allocation in land terrestrial communication environment, are considered promising solutions for infinite-horizon stochastic data scheduling problem. Specifically, recent work [24] reviewed numerous applications of the MDP framework for developing adaptive algorithms and protocols for wireless sensor networks. By exploiting MDP to model the transmission process in wireless network, a policy-restricted MDP and an induced MDP were proposed in [25] to reduce the state space to further reduce complexity. Besides, a channel resource allocation scheme based on SMDP was proposed to maximize the overall system rewards in vehicular ad hoc networks in [26]. Li *et al.* [27] proposed two SMDP based coordinated virtual machine allocation schemes to balance the tradeoff between the high cost of providing services by the remote cloud and the limited computing capacity of the local fog. The attentive design of the infinite-horizon stochastic data scheduling optimization in SSNs using a dynamic programming framework is still missing to the best of knowledge of the authors.

Dynamic programming has to be specialized for applications in satellite scenarios with the consideration of the following two satellite characteristics: 1) the satellite situation, e.g., network topology, is time variant but not random; 2) the solution for each iteration is not a number, i.e., cannot be obtained directly by calculating an explicit expression, but an optimization problem. Based on this motivation, to bridge these research gaps, in this paper, we consider the practical issues of stochastic data arrival and dynamic and limited network resources in an SSN environment. We extend the traditional dynamic programming framework and propose a finite-embedded-infinite two-level dynamic programming framework with joint consideration of contact selection, battery management, and buffer management specifically for the optimization of stochastic data scheduling in SSNs.

C. Contributions

To the best of our knowledge, our finite-embedded-infinite two-level dynamic programming framework is the first approach to formulate the data scheduling process over stochastic data arrival regarding contact selection. In this paper, we formulate the stochastic data scheduling optimization problem as a discrete infinite MDP. By exploiting the specific problem structure, we propose a joint forward and backward induction (JFBI) algorithm framework to resolve the infinite MDP iteratively using a value iteration approach. The calculation of the value in each iteration in forward induction part can be regarded as a finite horizon MDP for each cycle. We further propose a backward induction based cycle reward (BICR) algorithm to calculate this value in each iteration in the forward induction part. Based on the JFBI algorithm framework, a value based JFBI (VB-JFBI) algorithm and a policy based JFBI (PB-JFBI) algorithm are proposed to achieve the optimal feasible contact selection policy. Finally, we evaluate our proposed two JFBI algorithms by simulations on an SSN with real dynamic network topology traces obtained using Satellite Tool Kit (STK). The results validate the effectiveness of our approaches and demonstrate the performance gain over existing schemes.

In a nutshell, we summarize the contributions of this work as follows.

- We propose a finite-embedded-infinite two-level dynamic programming framework with considerations of the practical issues of stochastic data arrival and time-varying IS contacts specifically in SSN environment.
- We formulate the stochastic data scheduling optimization problem as a discrete infinite MDP problem to maximize the discounted infinite-horizon network reward considering the special feature of the SSN situation which is time variant but not stochastic.
- Two JFBI algorithms, i.e., VB-JFBI and PB-JFBI algorithms, are proposed to efficiently resolve the proposed infinite MDP problem. In particular, compared with conventional MDP, the value for each iteration cannot be obtained directly by calculating an explicit expression but needs to solve an optimization problem.
- Our proposed two JFBI algorithms can give the optimal contact selection policy in accordance with IS channel condition, battery condition, and storage condition for any initial resource status (e.g., storage and battery) compared with other state-of-the-art data scheduling schemes which may only provide contact selection strategies for a specific initial network resource status.
- We obtain the following interesting results through extensive simulations: 1) By considering the infinite horizon planning in the proposed VB-JFBI and the PB-JFBI algorithms, a more judicious IS and SD contacts selection can be achieved to match network resources and data demand; 2) Our proposed two JFBI algorithms can efficiently collaborate on multi-dimensional network resources so as to boost the long-term network performance.

D. Organization

The remainder of the paper is organized as follows. In Section II, the reference SSN system model is described. A finite-embedded-infinite two-level MDP framework is given to formulate the stochastic data scheduling problem in Section III. We further design two JFBI algorithms, i.e., the VB-JFBI and the PB-JFBI algorithms, to solve the proposed MDP problem in Section IV. We summarize the performance evaluation settings, the adopted VB-JFBI and PB-JFBI algorithms, and the simulation results in Section V. Finally, concluding remarks are given in Section VI.

II. System Model

A. Network Model

We consider an SSN with two types of components: 1) a set $\mathcal{U} = \{u_1, \cdots, u_U\}$ of U small satellites moving in LEOs. These small satellites collect space data by their onboard imagers or sensors and then transmit the space data to ground stations; 2) a set $\mathcal{G} = \{g_1, \cdots g_G\}$ of G ground stations which serve as the sinks for the data collected in small satellites. Due to the periodic movement of satellites, the network topology periodically repeats. The periodic change of the network topology is called a cycle and the time duration of each cycle is denoted as $\mathcal{T}$[1]. $\mathcal{T}$ is slotted as T slots with constant duration τ and the network topology is fixed during each slot (without loss of generality) [28]. We further utilize $z_{l,t}$ to represent the time slot (TS) which belongs to the t-th slot of the l-th cycle in system. An example SSN with four small satellites and two ground stations is depicted in Fig. 1. Besides, we define $\mathcal{G}\left(\mathcal{V}, \mathcal{E}\left(z_{l,t}\right)\right)$ as the predictable network topology graph in TS $z_{l,t}$, where $\mathcal{V}$ represents the set of network nodes and $\mathcal{E}\left(z_{l,t}\right)$ denotes the set of available communication links which include the set of IS links $\mathcal{E}_{ss}\left(z_{l,t}\right)$ and SDs $\mathcal{E}_{sg}\left(z_{l,t}\right)$ in TS $z_{l,t}$, i.e., $\mathcal{E}\left(z_{l,t}\right) = \mathcal{E}_{ss}\left(z_{l,t}\right) \cup \mathcal{E}_{sg}\left(z_{l,t}\right)$. Besides, $e_{ij}\left(z_{l,t}\right)$ is defined as the communication links from nodes v_i to v_j in TS $z_{l,t}$. Since the network topology is repeated every cycle, $\mathcal{E}\left(z_{l,t}\right)$, $\mathcal{E}_{ss}\left(z_{l,t}\right)$, $\mathcal{E}_{sg}\left(z_{l,t}\right)$, and $e_{ij}\left(z_{l,t}\right)$ remain unchanged for any cycle l. For the sake of brevity, we use $\mathcal{E}_t$, $\mathcal{E}_{ss}^t$, $\mathcal{E}_{sg}^t$, and e_{ij}^t instead of $\mathcal{E}\left(z_{l,t}\right)$, $\mathcal{E}_{ss}\left(z_{l,t}\right)$, $\mathcal{E}_{sg}\left(z_{l,t}\right)$, and $e_{ij}\left(z_{l,t}\right)$ for any cycle l, respectively. The key notations are summarized in Table I.

B. Stochastic Data Arrival Model

Due to the random distribution of the targets of interest and their differentiation, there is a bursty data source for each satellite. Let $\mathbf{r}\left(z_{l,t}\right) = \left\{r_i\left(z_{l,t}\right)\tau | v_i \in \mathcal{U}\right\}$ be the random arrivals (number of bits) collected by satellites at the end of TS $z_{l,t}$.[2] Assume that $r_i\left(z_{l,t}\right)$ is independent and identically distributed (i.i.d.) across TSs [23] according to a general distribution $\Pr\left[r_i\right]$ for satellite u_i. Besides, $r_i\left(z_{l,t}\right)$ is independent with regard to i. The moment generating functions of r_i exist with $\mathbb{E}\left[r_i\right] = \lambda_i$ which is partitioned (quantized) into $H+1$

[1] $\mathcal{T}$ is the least common multiple of the period of the Earth's rotation and the period in which the satellite moves along its orbit.

[2] We assume that the transmitters are causal so that the data arrived at the TS is not observed when the actions of this TS are performed.

TABLE I
KEY NOTATIONS

Symbol	Definition
$\mathcal{U}$, $\mathcal{G}$	The set of small satellites and ground stations in the network
$\mathcal{T}$, T, τ, $z_{l,t}$	The time duration of each cycle, the number of time slots in a cycle $\mathcal{T}$, the duration of each time slot, and the time slot which belongs to the t-th slot of the l-th cycle in system
$\mathcal{V}$, $\mathcal{E}\left(z_{l,t}\right)$, $\mathcal{G}\left(\mathcal{V},\mathcal{E}\left(z_{l,t}\right)\right)$	The set of network nodes, the set of available communication links which include the set of IS links $\mathcal{E}_{ss}\left(z_{l,t}\right)$ and SDs $\mathcal{E}_{sg}\left(z_{l,t}\right)$ in TS $z_{l,t}$, and the predictable network topology graph in TS $z_{l,t}$
$\mathbf{r}\left(z_{l,t}\right)$, $C\left(e_{ij}^t\right)$	The random arrivals (number of bits) collected by satellites at the end of TS $z_{l,t}$ and the capacity of link e_{ij}^t $\left(e_{ij}^t \in \mathcal{E}\left(z_{l,t}\right)\right)$
P_{ss}, P_{sg}, P_r, P_h	The corresponding transmission powers for IS and SD contacts, the nominal operation power, and the energy collection rate
$E_{it}\left(z_{l,t}\right)$, $E_{ir}\left(z_{l,t}\right)$, $E_{io}\left(z_{l,t}\right)$, $E_{ih}\left(z_{l,t}\right)$	The energy consumption of transmitting, receiving, and the nominal energy consumption and the harvested energy at u_i during the time slot $z_{l,t}$
$\boldsymbol{s}\left(z_{l,t}\right)$, $a\left(z_{l,t}\right)$	The system state and action (the feasible network contact selection) in TS $z_{l,t}$
$\mathbf{B}\left(z_{l,t}\right)$, $\mathbf{EB}\left(z_{l,t}\right)$	The system storage state and battery energy state
B_{max}, EB_{max}	The storage capacity and battery capacity
$R\left(\boldsymbol{s}\left(z_{l,t}\right), a\left(z_{l,t}\right)\right)$	The immediate reward of taking action $a\left(z_{l,t}\right)$ in state $\boldsymbol{s}\left(z_{l,t}\right)$ at TS $z_{l,t}$
$y_l\left(e_{ij}^t\right)$	$= 1$ if link e_{ij}^t is active in cycle l, and 0 otherwise
$p\left(\boldsymbol{s}\left(z_{l,t+1}\right) \mid \boldsymbol{s}\left(z_{l,t}\right), a\left(z_{l,t}\right)\right)$	The state transition probability (the probability that the system will be in state $\boldsymbol{s}\left(z_{l,t+1}\right)$ in the $(t+1)$-th slot in cycle l, given that it was in state $\boldsymbol{s}\left(z_{l,t}\right)$ and action $a\left(z_{l,t}\right)$ in the t-th slot in cycle l)
$\mathcal{D}\left(z_{l,t}\right)$, $\mathcal{C}\left(z_{l,t}\right)$	The total amount of the data downloaded to ground stations in TS $z_{l,t}$ and the total amount of the data which remains in the satellites' storage at the end of TS $z_{l,t}$
$\kappa_t : \boldsymbol{s}\left(z_{l,t}\right) \to a\left(z_{l,t}\right)$	A mapping from a state to an action
$\boldsymbol{\vartheta}_l$, Π	A sequence of decisions for each TS in cycle l and the set of all feasible network contact policies during a cycle
$\varPhi_{\Sigma}^{\boldsymbol{\vartheta}_l}(\boldsymbol{s}'\left(z_{l+1,1}\right), \boldsymbol{\vartheta}_l, \boldsymbol{s}\left(z_{l,1}\right))$	The expected total reward in a cycle l
$\phi_l^*\left(\boldsymbol{s}(z_{l,1})\right)$	The maximum achievable network reward at state $\boldsymbol{s}(z_{l,1})$
$Q_l\left(\boldsymbol{s}\left(z_{l,1}\right), \boldsymbol{\vartheta}_l\right)$	The value of taking policy $\boldsymbol{\vartheta}_l$ at state $\boldsymbol{s}(z_{l,1})$
$\psi_t^{\boldsymbol{\vartheta}_l}\left(\boldsymbol{s}(z_{l,t})\right)$	The reward-to-go function in the t-th slot during cycle l
$\varTheta^*$, $\phi_l^*\left(\boldsymbol{s}(z_{l,1})\right)$	The optimal policy of the proposed infinite MDP problem and the optimal value

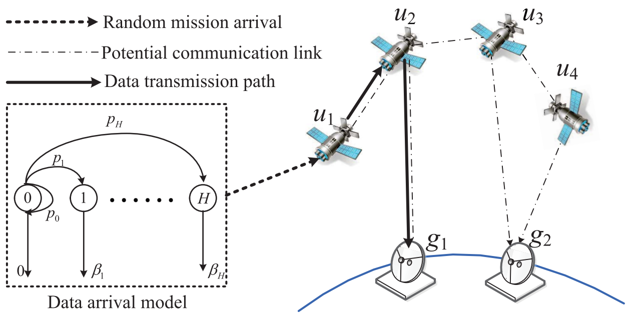

Fig. 1. Small satellite network model and data arrival model.

levels, denoted as $\mathcal{R} = \{0, \beta_1, \beta_2, \cdots, \beta_H\}$ shown in Fig. 1, where 0 represents that there is no data arrival. During TS $z_{l,t}$, each satellite u_i is associated with an average data arrival rate $r_i\left(z_{l,t}\right) = \beta_n$ with some probability p_n. The specific data arrival distribution $\Pr\left[r_i\right]$ can be captured by mission management center (MMC) [29]. In practical applications, MMC can monitor the data arrival information for a long period of time, and then the data arrival distribution $\Pr\left[r_i\right]$ will be obtained according to the statistical characteristics of the obtained data arrival information.

Data is served by the SSN through store-carry-forward paradigm [15] which consists of two types as shown in Fig. 1: 1) store-transmission: small satellites can first store the stochastically arrived data onboard and then carry them forward until a transmission opportunity is available; and 2) store-relay-transmission: small satellites can transmit the stored data onboard to other small satellites for assistance to download the data.

C. Inter-Satellite Contact Model

Due to the periodic orbiting movement of satellites, the IS contacts are dynamic which is reflected in the following two aspects: 1) the existence of IS links is dynamic; and 2) the achievable data rate (in bps) of IS links is dynamic. Since the network topology is repeated every cycle, we utilize t to represent $\left(z_{l,t}\right)$ for brevity.

According to [30] and [31], the achievable data rate (in bps) of an IS link from a small satellite u_i to u_j in TS $\left(z_{l,t}\right)$, denoted by $Ds_{ij}\left(z_{l,t}\right)$, can be expressed as follows:

$$Ds_{ij}\left(z_{l,t}\right) = \frac{P_{ts} G_{tri} G_{rej} L_{fij}\left(z_{l,t}\right)}{kT_s \cdot \left({}^{E_b}/{}_{N_0}\right)_{req} \cdot \varOmega}, \tag{1}$$

where the free space loss $L_f\left(z_{l,t}\right)$ is

$$L_{fij}\left(z_{l,t}\right) = \left(\frac{c}{4\pi \cdot S_{ij}\left(z_{l,t}\right) \cdot f}\right)^2. \tag{2}$$

P_{ts} is the constant transmission power (in W) of satellites for IS links. G_{tri} and G_{rej} are the transmitting antenna gain and receiving antenna gain for the two ends of the link e_{ij}^t, respectively. Besides, k and T_s are the Boltzmann's constant (in JK^{-1}) and total system noise temperature (in K). $({}^{E_b}/{}_{N_0})_{req}$ and Ω are the required ratio of received energy-per-bit to noise-density and link margin. $S_{ij}(z_{l,t})$ is the slant range (in km) in TS $(z_{l,t})$. c and f are the speed of light (in km/s) and communications center frequency (in Hz) of IS links. Since the network topology is repeated every cycle, $S_{ij}(z_{l,t})$ remains the same for the same slot t in different cycle l, which makes that $Ds_{ij}(z_{l,t})$ remains constant for the same slot t in different cycle l. Therefore, we utilize $Ds_{ij}(t)$ to replace with $Ds_{ij}(z_{l,t})$ for brevity.

D. Energy Dynamics Model

In this subsection, we present the energy dynamics model for small satellites [32]. To model the energy consumption of satellites, we first define $x_l(e_{ij}^t, u_k)$ as the amount of data originating from satellite u_k passing through link e_{ij}^t during TS $z_{l,t}$. Furthermore, the corresponding IS link capacity for link e_{ij}^t is denoted as $C(e_{ij}^t)$, which is defined as the maximum amount of data that can be transmitted by the corresponding link and represented as $C(e_{ij}^t) = Ds_{ij}(t)\cdot\tau, \forall e_{ij}^t \in \mathcal{E}_{ss}^t$. Similarly, the capacity link for an SD $e_{ij}^t \in \mathcal{E}_{sg}^t$ is $C(e_{ij}^t) = Ds_{ij}\cdot\tau, \forall e_{ij}^t \in \mathcal{E}_{sg}^t$, where Ds_{ij} is the constant data rate of SD $e_{ij}^t \in \mathcal{E}_{sg}^t$. Hereafter, we elaborate on the specific energy consumption and harvesting model for small satellites.

The energy consumption of satellite u_i for transmitting data during TS $z_{l,t}$, denoted by $E_{it}(z_{l,t})$, can be expressed as

$$E_{it}(z_{l,t}) = \sum_{u_k\in\mathcal{U}}\left(\sum_{j:(e_{ij}^t)\in\mathcal{E}_{ss}^t} P_{ss}\cdot\frac{x_l(e_{ij}^t,u_k)}{C(e_{ij}^t)} + \sum_{j:(e_{ij}^t)\in\mathcal{E}_{sg}^t} P_{sg}\cdot\frac{x_l(e_{ij}^t,u_k)}{C(e_{ij}^t)}\right)\cdot\tau, \quad (3)$$

which includes two parts, i.e., energy consumption for IS links and SDs. P_{ss} and P_{sg} are the corresponding transmission powers.

We denote $E_{ir}(z_{l,t})$ as energy consumption for receiving data at satellite u_i during TS $z_{l,t}$ with reception power P_r. $E_{ir}(z_{l,t})$ is shown as follows

$$E_{ir}(z_{l,t}) = \sum_{u_k\in\mathcal{U}}\sum_{j:(e_{ji}^t)\in\mathcal{E}_{ss}^t} P_r\cdot\frac{x_l(e_{ji}^t,u_k)}{C(e_{ji}^t)}\cdot\tau. \quad (4)$$

Moreover, we use $E_{io}(z_{l,t})$ to denote the energy consumption for nominal operation at satellite u_i during TS $z_{l,t}$. $E_{io}(z_{l,t})$ can be expressed as

$$E_{io}(z_{l,t}) = P_o\cdot\tau, \quad (5)$$

where P_o is the nominal operation power.

The total energy consumption at satellite u_i during TS $z_{l,t}$, denoted by $E_{ic}(z_{l,t})$, consists of the above three energy consumption items and is shown as

$$E_{ic}(z_{l,t}) = E_{it}(z_{l,t}) + E_{ir}(z_{l,t}) + E_{io}(z_{l,t}). \quad (6)$$

We denote $E_{ih}(z_{l,t})$ as the harvested energy at satellite u_i over TS $z_{l,t}$ which can be determined in advance based on orbital dynamics. Since $E_{ih}(z_{l,t})$ are identical for the same slot t in different cycle l, i.e., $E_{ih}(z_{l',t'}) = E_{ih}(z_{l,t}), \forall l'$, $t' = t$, we use E_{ih}^t instead of $E_{ih}(z_{l,t})$ for brevity. E_{ih}^t[3] can be expressed as follows

$$E_{ih}^t = P_h \times \max\left\{0, \tau - d_i^t\right\}, \quad (7)$$

where P_h is the energy collection rate. $d_i^t = 0$ when satellite u_i is exposed to the sun while when satellite u_i is is eclipsed by the Earth, d_i^t is the remaining time that satellite u_i still stay in shadow at the beginning of the t-th slot in a cycle. Observe from (7) that E_{ih}^t is positive when d_i^t is smaller than τ; otherwise, $E_{ih}^t = 0$.

III. Problem Formulation Based on MDP Framework

In this section, we propose a finite-embedded-infinite two-level MDP framework based on the dynamic programming algorithms [33] to formulate the stochastic data scheduling problem, where we assume a priori knowledge of the statistics of data arrival. Specifically, we first elaborate on the technical definition of the MDP for SSN. Then, we further formulate the proposed stochastic data scheduling problem as an infinite horizon MDP.

A. A Discrete Finite-Embedded-Infinite Two-Level MDP Framework for SSNs

Data scheduling in SSN is a sequential decision problem, which can be cast as an infinite MDP described as follows.

1) State: The system state in TS $z_{l,t}$, denoted as $\boldsymbol{s}(z_{l,t})$, is represented by a 2-tuple, i.e.,

$$\boldsymbol{s}(z_{l,t}) \triangleq \langle \mathbf{B}(z_{l,t}), \mathbf{EB}(z_{l,t})\rangle, \quad (8)$$

where $\mathbf{B}(z_{l,t}) = \{B_i(z_{l,t})|i\in\mathcal{U}\}$ and $\mathbf{EB}(z_{l,t}) = \{EB_i(z_{l,t})|i\in\mathcal{U}\}$ are the system storage state and battery energy state, respectively. $B_i(z_{l,t})$ and $EB_i(z_{l,t})$ are the data stored and residual energy in satellite u_i at the beginning of TS $z_{l,t}$, respectively. We partition the satellite storage and battery capacity into M_1 and M_2 intervals with equal lengths, i.e.,

$$[0, B_{max}] = \left[0, \frac{B_{max}}{M_1}\right) \cup\cdots\cup \left[(M_1-1)\frac{B_{max}}{M_1}, B_{max}\right]$$

and

$$[EB_l, EB_{max}] = \left[EB_l, EB_l + \frac{EB_a}{M_2}\right) \cup\cdots\cup \left[EB_l + (M_2-1)\cdot\frac{EB_a}{M_2}, EB_{max}\right],$$

where B_{max} and EB_{max} are the storage capacity and battery capacity, respectively. $EB_l = (1-\eta)\cdot EB_{max}$ and $EB_a = EB_{max} - EB_l$. η is the maximum discharge depth of the battery. The satellite buffer state and energy state in each interval are represented by their mid-values, i.e.,

$$B_i(z_{l,t}) \in \{B_{max}/2M_1, \cdots, (2M_1-1)B_{max}/2M_1\}$$

and

$$EB_i(z_{l,t}) \in \{EB_l + EB_a/2M_2, \cdots, EB_{max} - EB_a/2M_2\}.$$

[3] We assume that the onboard batteries are causal so that the energy harvested in TS $z_{l,t}$ is not observed when the actions of this TS are performed.

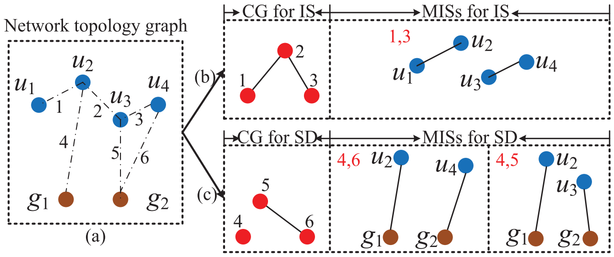

Fig. 2. Maximum feasible IS and SD contacts selection schematic.

2) Action: The action in the MDP problem is $a(z_{l,t})$, which is the feasible network contact selection in TS $z_{l,t}$. Due to the limited number of transceivers, a potential contact may not be feasible for data transmission. We introduce a set of boolean variables $y_l\left(e_{ij}^t\right)=\{0,1\}, \forall e_{ij}^t \in \mathcal{E}_{ss}^t \cup \mathcal{E}_{sg}^t$ to model the contention caused by limited transponders in TS $z_{l,t}$, where $y_l\left(e_{ij}^t\right)=1$ if link e_{ij}^t is active in cycle l, and 0 otherwise. We assume a small satellite can only simultaneously establish one IS link and one SD in each TS, i.e.,

$$\sum_{j:e_{ij}^t \in \mathcal{E}_{ss}^t} y_l\left(e_{ij}^t\right) \leq 1, \quad \forall u_i \in \mathcal{U}, l, t, \tag{9}$$

$$y_l\left(e_{ij}^t\right) = y_l\left(e_{ji}^t\right), \ \forall e_{ij}^t \in \mathcal{E}_{ss}^t, l, t, \tag{10}$$

and

$$\sum_{j:e_{ij}^t \in \mathcal{E}_{sg}^t} y_l\left(e_{ij}^t\right) \leq 1, \ \forall u_i \in \mathcal{U}, l, t. \tag{11}$$

(10) restricts bi-directionality on the IS contact selection. Besides, a ground station can only establish one SD in each TS, i.e.,

$$\sum_{i:e_{ij}^t \in \mathcal{E}_{sg}^t} y_l\left(e_{ij}^t\right) \leq 1, \ \forall g_j \in \mathcal{G}, l, t. \tag{12}$$

The action space is all the potential feasible network contacts. Since the action space remains unchanged for any cycle, we let A_t denote the feasible action set in the t-th slot in any cycle, which can be obtained by finding the feasible IS and SD contacts. To efficiently exploit the communication resource, we find the maximum feasible IS and SD contacts. We first construct conflict graphs (CGs) based on the network topology for the IS and SD contacts, respectively. Then, an efficient algorithm, such as the Bron-Kerbosch (BK) algorithm [34], can be employed to discover the maximum independent sets (MISs) for the CGs to attain the corresponding feasible IS and SD contacts. Note that in our proposed MIS problems, to get all possible maximum feasible contacts, the weight of the edge in the aforementioned CGs is set at 1, which means that the resulting feasible IS and SD contacts may not be unique. An example is shown in Fig. 2, where Fig. 2a is the network graph for the example SSN in Fig. 1. Fig. 2b and Fig. 2c are the corresponding feasible IS and SD contacts, respectively.

3) Reward: $R(\boldsymbol{s}(z_{l,t}), a(z_{l,t}))$ denotes the immediate reward of taking action $a(z_{l,t})$ in state $\boldsymbol{s}(z_{l,t})$ at TS $z_{l,t}$. Different from the conventional MDP wherein the reward can be defined as the function of state and action, and can be calculated directly, $R(\boldsymbol{s}(z_{l,t}), a(z_{l,t}))$ in our MDP model needs to be calculated by solving a linear programming (LP) optimization problem.

In the process of data transmission for each time slot, flow conservation should be satisfied. Flow conservation balances the change in the storage occupancy of a node against the incoming data for every TS. To model the flow conservation, we define $p_i(u_k, z_{l,t})$ (in Mbit) as the volume of data which origins from satellite u_k and is stored in the buffer of satellite u_i at the end of TS $z_{l,t}$. Thus, we have the following equations:

$$\sum_{j:e_{ij}^t \in \mathcal{E}_t} x_l\left(e_{ij}^t, u_k\right) + p_i\left(u_k, z_{l,t}\right) = B_k\left(z_{l,t}\right), \quad \forall u_k \in \mathcal{U}, u_i = u_k, \tag{13}$$

$$\sum_{j:e_{ji}^t \in \mathcal{E}_t} x_l(e_{ji}^t, u_k) = \sum_{j:e_{ij}^t \in \mathcal{E}_t} x_l(e_{ij}^t, u_k) + p_i\left(u_k, z_{l,t}\right), \quad \forall u_k \in \mathcal{U}, u_i \neq u_k. \tag{14}$$

$$\sum_{j:e_{ji}^t \in \mathcal{E}_{sg}^t} x_l(e_{ji}^t, u_k) = p_i\left(u_k, z_{l,t}\right), \quad \forall u_k \in \mathcal{U}, g_i \in \mathcal{G}. \tag{15}$$

Constraints are also needed to ensure that the amount of data sent over an active link is limited by its capacity over that TS, and thus we have

$$\sum_{u_k \in \mathcal{U}} x_l\left(e_{ij}^t, u_k\right) \leq C\left(e_{ij}^t\right) \cdot y_l\left(e_{ij}^t\right), \quad \forall e_{ij}^t \in \mathcal{E}_t. \tag{16}$$

Similarly, the storage at satellite node does not exceed the specified limit, i.e.,

$$B_i'\left(z_{l,t}\right) = \sum_{u_k \in \mathcal{U}} p_i\left(u_k, z_{l,t}\right) \leq B_{max}. \tag{17}$$

Herein, $B_i'(z_{l,t})$ is the amount of buffered data to be downloaded in satellite u_i at the end of TS $z_{l,t}$ which does not include newly arrived data during TS $z_{l,t}$. Besides, energy consumption during TS $z_{l,t}$ cannot exceed the specific limit. Since only energy that has been buffered up to $z_{l,t}$ can be used in the current TS, there exists a causality constraint on $EB_i'(z_{l,t})$ satisfying

$$EB_i'\left(z_{l,t}\right) = EB_i\left(z_{l,t}\right) - E_{ic}\left(z_{l,t}\right) \geq EB_{max} \cdot (1-\eta), \tag{18}$$

where $EB_i'(z_{l,t})$ denotes the residual energy of satellite u_i at the end of TS $z_{l,t}$ which does not include the harvested energy during TS $z_{l,t}$. Besides, we define $\mathbf{B}'(z_{l,t}) = \{B_i'(z_{l,t}) | v_i \in \mathcal{U}\}$ and $\mathbf{EB}'(z_{l,t}) = \{EB_i'(z_{l,t}) | v_i \in \mathcal{U}\}$, respectively.

Given state $\boldsymbol{s}(z_{l,t})$ (i.e., $\mathbf{B}(z_{l,t})$ and $\mathbf{EB}(z_{l,t})$) and action $a(z_{l,t})$ (i.e., $y_l\left(e_{ij}^t\right), \forall e_{ij}^t \in \mathcal{E}_t$) in TS $z_{l,t}$, the corresponding reward $R(\boldsymbol{s}(z_{l,t}), a(z_{l,t}))$ can be obtained by solving the following linear programming (LP) problem:

$$\text{LP}: \max_{\mathbf{x}(z_{l,t}), \mathbf{B}'(z_{l,t}), \mathbf{EB}'(z_{l,t})} \mathcal{D}\left(z_{l,t}\right) - \omega \mathcal{C}\left(z_{l,t}\right) \quad \text{s.t. } (13)-(18), \tag{19}$$

where

$$\mathcal{D}\left(z_{l,t}\right) = \sum_{e_{ij}^t \in \mathcal{E}_{sg}^t} \sum_{k=1}^{k=U} x_l\left(e_{ij}^t, u_k\right)$$

represents the total amount of the data downloaded to ground stations in TS $z_{l,t}$, and

$$\mathcal{C}\left(z_{l,t}\right) = \sum_{k=1}^{k=U} b_i'\left(z_{l,t}\right)$$

denotes the total amount of the data which remains in the satellites' storage. ω here is a weight parameter which can be regarded as the penalty factor, mimicking the soft delay constraint. By finding the data transmission variables $\mathbf{x}\left(z_{l,t}\right) = \left\{x_l(e_{ij}^t, u_k) | e_{ij}^t \in \mathcal{E}_t, u_k \in \mathcal{U}\right\}$, $\mathbf{B}'\left(z_{l,t}\right)$, and $\mathbf{EB}'\left(z_{l,t}\right)$ to solve the LP problem, we can then obtain the potential transferred system states in the next TS.

4) State Transition Probabilities: The state transition probability in the same cycle l, denoted as $p\left(\boldsymbol{s}\left(z_{l,t+1}\right)|\boldsymbol{s}\left(z_{l,t}\right), a\left(z_{l,t}\right)\right)$ is the probability that the system will be in state $\boldsymbol{s}\left(z_{l,t+1}\right)$ in the $(t+1)$-th slot in cycle l, given that it was in state $\boldsymbol{s}\left(z_{l,t}\right)$ and action $a\left(z_{l,t}\right)$ in the t-th slot in cycle l. The buffer and energy at satellite u_i evolves according to $B_i\left(z_{l,t+1}\right) = Q_1\left(B_i'\left(z_{l,t}\right) + r_i\left(z_{l,t}\right)\right)$ and $EB_i\left(z_{l,t+1}\right) = Q_2\left(EB_i'\left(z_{l,t}\right) + E_{ih}^t\right)$, where

$$\begin{aligned} &Q_1\left(\varrho_1\right) \\ &= \left(2\min\left\{\left\lfloor\frac{M_1 \min\left\{\varrho_1, B_{max}\right\}}{B_{max}}\right\rfloor + 1, M_1\right\} - 1\right) \cdot \frac{B_{max}}{2M_1}, \end{aligned} \tag{20}$$

and

$$\begin{aligned} &Q_2\left(\varrho_2\right) \\ &= EB_l + \left(2\min\left\{\left\lfloor\frac{M_2 \min\left\{\varrho_2 - EB_l, EB_a\right\}}{EB_a}\right\rfloor + 1, M_2\right\} - 1\right) \\ &\quad \cdot \frac{EB_a}{2M_2}. \end{aligned} \tag{21}$$

Herein, $\lfloor x \rfloor$ is the floor function. $Q_1\left(\varrho_1\right)$ and $Q_2\left(\varrho_2\right)$ are two quantization functions mapping $B_i\left(z_{l,t}\right)$ and $EB_i\left(z_{l,t}\right)$ to the buffer state and battery energy state, which are nondecreasing over ϱ_1 and ϱ_2. Similarly, the state transition probability between two different cycles can be denoted as $p\left(\boldsymbol{s}\left(z_{l+1,1}\right)|\boldsymbol{s}\left(z_{l,T}\right), a\left(z_{l,T}\right)\right)$, which represents the state transition from the T-th slot in cycle l to the first slot in cycle $(l+1)$.

Proposition 1: For the state transition probability in the same cycle l, with given $\boldsymbol{s}\left(z_{l,t}\right)$ and $a\left(z_{l,t}\right)$, the state in TS $z_{l,t+1}$ depends only on the present $\mathbf{B}'\left(z_{l,t}\right)$, $\mathbf{r}\left(z_{l,t}\right)$, $\mathbf{EB}'\left(z_{l,t}\right)$, and $\mathbf{E}_h\left(t\right)$. In other words, the system state transition process has the Markov property. Besides, the state transition probability $p\left(\boldsymbol{s}\left(z_{l,t+1}\right)|\boldsymbol{s}\left(z_{l,t}\right), a\left(z_{l,t}\right)\right) = p\left(\mathbf{r}\left(z_{l,t}\right)\right)$, where $\mathbf{r}\left(z_{l,t}\right) = \left\{r_i\left(z_{l,t}\right)\tau | v_i \in \mathcal{U}\right\}$ denotes the data arrival vector in TS $z_{l,t}$. The state transition process between two different cycles also has the Markov property and the state transition probability $p\left(\boldsymbol{s}\left(z_{l+1,1}\right)|\boldsymbol{s}\left(z_{l,T}\right), a\left(z_{l,T}\right)\right) = p\left(\mathbf{r}\left(z_{l,T}\right)\right)$, where $\mathbf{r}\left(z_{l,T}\right) = \left\{r_i\left(z_{l,T}\right)\tau | v_i \in \mathcal{U}\right\}$.

Proof: Define $\mathbf{E}_h\left(t\right) = \left\{E_{ih}^t | v_i \in \mathcal{U}\right\}$. The specific derivation for $p\left(\boldsymbol{s}\left(z_{l,t+1}\right)|\boldsymbol{s}\left(z_{l,t}\right), a\left(z_{l,t}\right)\right)$ is shown as follows:

$$\begin{aligned} &p\left(\boldsymbol{s}\left(z_{l,t+1}\right)|\boldsymbol{s}\left(z_{l,t}\right), a\left(z_{l,t}\right)\right) \\ &= p\left(\left(\mathbf{B}\left(z_{l,t+1}\right), \mathbf{EB}\left(z_{l,t+1}\right)\right)|\boldsymbol{s}\left(z_{l,t}\right), a\left(z_{l,t}\right)\right) \\ &= p\left(\left(\mathbf{B}'(z_{l,t}) + \mathbf{r}(z_{l,t}), \mathbf{EB}'(z_{l,t}) + \mathbf{E}_h\left(t\right)\right)|\boldsymbol{s}\left(z_{l,t}\right), a\left(z_{l,t}\right)\right) \\ &= p\left(\mathbf{r}\left(z_{l,t}\right)\right) \\ &= \prod_{i=1}^{i=U} p\left(r_i\left(z_{l,t}\right)\right). \end{aligned} \tag{22}$$

Note that the energy harvesting process is a deterministic process. Besides, $\mathbf{B}'\left(z_{l,t}\right)$ and $\mathbf{EB}'\left(z_{l,t}\right)$ can be obtained by solving a deterministic LP. Consequently, the system state in the next TS depends only on the data arrival vector $\mathbf{r}\left(z_{l,t}\right)$. Since the derivation of $p(\boldsymbol{s}(z_{l+1,1})|\boldsymbol{s}(z_{l,T}), a(z_{l,T}))$ is similar to that of $p(\boldsymbol{s}(z_{l,t+1})|\boldsymbol{s}(z_{l,t}), a(z_{l,t}))$, the corresponding part of the proof is omitted for brevity. According to the definition given in [35], a stochastic process has the Markov property if the conditional probability distribution of future states of the process (conditional on both past and present states) depends only upon the present state, not on the sequence of events that preceded it. Therefore, the proposed state transition process is a Markov process. ■

B. Problem Formulation

In the proposed finite-embedded-infinite two-level MDP framework, a decision is a mapping from a state to an action, i.e., $\kappa_t : \boldsymbol{s}\left(z_{l,t}\right) \rightarrow a\left(z_{l,t}\right)$. A "policy" $\boldsymbol{\vartheta}_l$ is a sequence of decisions for each TS in cycle l, i.e., $\boldsymbol{\vartheta}_l = \left\{\kappa_1^{\boldsymbol{\vartheta}_l}\left(\boldsymbol{s}\left(z_{l,1}\right)\right), \cdots, \kappa_T^{\boldsymbol{\vartheta}_l}\left(\boldsymbol{s}\left(z_{l,T}\right)\right)\right\}$. The set of all feasible network contact policies during a cycle is denoted as Π. Specifically, Π is the combination of the feasible network contact decision for each TS in a cycle and remains unchanged for any cycle due to the periodic repetition of the network topology. Given the initial state $\boldsymbol{s}\left(z_{l,1}\right)$ in the first TS in a cycle l, the expected total reward in a cycle l denotes the sum of the slot reward in a cycle l gained by taking policy $\boldsymbol{\vartheta}_l$ in state $\boldsymbol{s}\left(z_{l,1}\right)$ and proceeding to state $\boldsymbol{s}'\left(z_{l+1,1}\right)$ in the next cycle. Specifically, we use $\Phi_{\Sigma}^{\boldsymbol{\vartheta}_l}(\boldsymbol{s}'\left(z_{l+1,1}\right), \boldsymbol{\vartheta}_l, \boldsymbol{s}\left(z_{l,1}\right))$ to represent the expected total reward in a cycle l and it is expressed as

$$\begin{aligned} &\Phi_{\Sigma}^{\boldsymbol{\vartheta}_l}(\boldsymbol{s}'\left(z_{l+1,1}\right), \boldsymbol{\vartheta}_l, \boldsymbol{s}\left(z_{l,1}\right)) \\ &\qquad = \mathbb{E}_{\boldsymbol{\vartheta}_l, \boldsymbol{s}_l}\left[\sum_{t=1}^{T} R\left(\boldsymbol{s}\left(z_{l,t}\right), \kappa_t^{\boldsymbol{\vartheta}_l}\left(\boldsymbol{s}(z_{l,t})\right)\right)\right], \end{aligned} \tag{23}$$

where $\mathbb{E}\left[\cdot\right]$ is the expectation function and the expectation is over all possible state sequences $\boldsymbol{s}_l = \left\{\boldsymbol{s}\left(z_{l,1}\right), \cdots, \boldsymbol{s}\left(z_{l,T}\right)\right\}$ in cycle l induced by $\boldsymbol{\vartheta}_l$. Consequently, the discounted infinite horizon total reward associated with a given policy Θ and initial state $\boldsymbol{s}\left(z_{1,1}\right)$ (i.e., the state of the first slot in cycle 1) is given by

$$\begin{aligned} &J^{\Theta}\left(\boldsymbol{s}\left(z_{1,1}\right)\right) \\ &= \mathbb{E}\left[\sum_{l=1}^{\infty} \alpha^{l-1} \Phi_{\Sigma}^{\boldsymbol{\vartheta}_l}\left(\boldsymbol{s}'(z_{l+1,1}), \boldsymbol{\vartheta}_l^{\Theta}, \boldsymbol{s}(z_{l,1})\right)|\boldsymbol{s}\left(z_{1,1}\right)\right], \end{aligned} \tag{24}$$

where $\boldsymbol{s}(z_{l,1})$ and $\boldsymbol{\vartheta}_l^{\Theta}$ denote the state visited at TS $z_{l,1}$ and policy taken on cycle l based on state $\boldsymbol{s}(z_{l,1})$, according to policy Θ. $\alpha\,(0 < \alpha < 1)$ is the discounting factor which can weigh the myopic or foresighted decisions. Note that the impact of the initial state on the discounted infinite horizon total reward, i.e., $J^{\Theta}\left(\boldsymbol{s}\left(z_{1,1}\right)\right)$, is dissolved over time.

The goal of the infinite horizon MDP is to find the optimal feasible network contacts for each slot in one cycle, i.e., $\Theta\left(\boldsymbol{s}\right)$, in state $\boldsymbol{s}$ that maximizes the discounted infinite horizon total reward $J^{\Theta}\left(\boldsymbol{s}\left(z_{1,1}\right)\right)$ as follows

$$V\left(\boldsymbol{s}\left(z_{1,1}\right)\right) = \max_{\Theta \in \Pi} J^{\Theta}\left(\boldsymbol{s}\left(z_{1,1}\right)\right). \tag{25}$$

The infinite horizon optimal data scheduling is to determine an optimal policy Θ^* that maximizes the value function.

IV. Joint Forward and Backward Induction Based Solution

In this section, we first propose a JFBI algorithm framework, which consists of two layers to solve the proposed infinite MDP problem with consideration of the special characteristics of SSNs, i.e., periodic characteristics. Specifically, a forward induction algorithm is proposed at the outer layer to solve the proposed infinite MDP problem. The value calculation in each iteration in forward induction, i.e., the network reward in each cycle, consists of T slots, which can be regarded as a finite-horizon MDP. We further propose a backward induction algorithm at the inner layer to obtain this value. Then, we analyze the complexity of the proposed JFBI algorithm framework.

A. Algorithm Design

Based on the proposed JFBI algorithm framework, we first propose a value based JFBI (VB-JFBI) algorithm to obtain the optimal contact selection policy, wherein the iteration stops until the improvement to the value in forward induction is less than a predetermined small threshold $\epsilon > 0$. Furthermore, a policy based JFBI (PB-JFBI) algorithm is proposed to achieve the optimal and stable contact selection policy for a cycle, wherein the iteration stops until the obtained policies in the two consecutive cycles are identical.

Definition 1: (Bellman's optimal equations for infinite cycles): Define $\phi_l^*(\boldsymbol{s}(z_{l,1}))$ as the maximum achievable network reward at state $\boldsymbol{s}(z_{l,1})$ during cycle l. We can find $\phi_l^*(\boldsymbol{s}(z_{l,1}))$ at every state recursively by solving the following Bellman's optimal equations [36],

$$\begin{aligned}&\phi_l^*\left(\boldsymbol{s}\left(z_{l,1}\right)\right)\\&= \varPhi_{\Sigma}^{\boldsymbol{\vartheta}_l}\left(\boldsymbol{s}'\left(z_{l+1,1}\right), \boldsymbol{\vartheta}_l, \boldsymbol{s}\left(z_{l,1}\right)\right)\\&\quad + \alpha \cdot \sum_{\boldsymbol{s}'(z_{l+1,1})} p\left(\boldsymbol{s}'\left(z_{l+1,1}\right) | \boldsymbol{s}\left(z_{l,1}\right), \boldsymbol{\vartheta}_l\right) \phi_{l+1}^*\left(\boldsymbol{s}'\left(z_{l+1,1}\right)\right).\end{aligned} \tag{26}$$

Based on the optimal Bellman equations in (26), we propose a forward induction method, which is a value iteration approach, to solve the proposed infinite horizon discounted MDP, i.e., to obtain $V\left(\boldsymbol{s}\left(z_{1,1}\right)\right)$. The idea is to divide the optimization problem based on the number of steps to go. In particular, given an optimal policy for $(l-1)$ time steps to go, we calculate the Q-values for l-steps to go. After that, we can obtain the optimal policy Θ^* based on the following equations:

$$\begin{aligned}&Q_l\left(\boldsymbol{s}\left(z_{l,1}\right), \boldsymbol{\vartheta}_l\right)\\&= \varPhi_{\Sigma}^{\boldsymbol{\vartheta}_l}(\boldsymbol{s}'(z_{l-1,1}), \boldsymbol{\vartheta}_l, \boldsymbol{s}\left(z_{l,1}\right))\\&\quad + \alpha \cdot \sum_{\boldsymbol{s}'(z_{l-1,1})} p(\boldsymbol{s}'(z_{l-1,1}) | \boldsymbol{s}(z_{l,1}), \boldsymbol{\vartheta}_l)\, \phi_{l-1}^*\left(\boldsymbol{s}'(z_{l-1,1})\right),\end{aligned} \tag{27}$$

$$\begin{cases}\phi_l^*(\boldsymbol{s}(z_{l,1})) = Q_l^*(\boldsymbol{s}(z_{l,1}),\boldsymbol{\vartheta}_l) = \max\limits_{\boldsymbol{\vartheta}_l \in \Pi} Q_l(\boldsymbol{s}(z_{l,1}),\boldsymbol{\vartheta}_l),\\ \Theta_l^*\left(\boldsymbol{s}(z_{l,1})\right) = \arg\max\limits_{\boldsymbol{\vartheta}_l \in \Pi} Q_l^*\left(\boldsymbol{s}\left(z_{l,1}\right), \boldsymbol{\vartheta}_l\right),\end{cases} \tag{28}$$

where $\Theta_l\left(\boldsymbol{s}(z_{l,1})\right)$ is the value of state $\boldsymbol{s}(z_{l,1}$ and $Q_l\left(\boldsymbol{s}\left(z_{l,1}\right), \boldsymbol{\vartheta}_l\right)$ is the value of taking policy $\boldsymbol{\vartheta}_l$ at state $\boldsymbol{s}(z_{l,1})$. The optimal policy Θ^* of this MDP problem can be obtained by solving (27) recursively with the forward induction algorithm, in which, the Q-values are evaluated over all possible states in the first slot in each cycle. Since the cycle horizon is infinite, this iteration procedure in forward induction continues till a convergence condition is met (e.g., $\left\|\phi_l^*(\boldsymbol{s}(z_{l,1})) - \phi_{l-1}^*(\boldsymbol{s}(z_{l-1,1}))\right\| < \epsilon$) [36].

From (27), we can see that the decisions from the first slot to the T-th slot in cycle l need to be selected. In other words, the calculation of $Q_l^*(\boldsymbol{s}(z_{l,1}),\boldsymbol{\vartheta}_l)$ for cycle l can be regarded as a finite-horizon MDP problem. Consequently, we proposed a BICR algorithm to calculate this Q-value $Q_l^*(\boldsymbol{s}(z_{l,1}),\boldsymbol{\vartheta}_l)$ by recursively obtaining the optimal actions for earlier slots back to the first slot. In each slot during cycle l, the optimal action balances the immediate reward and the future reward by selecting a suitable network contacts topology.

The reward-to-go function in the t-th slot during cycle l is defined as the sum of the expected total reward from the t-th slot to the T-th slot during cycle l, which can be expressed as shown in (29).

Definition 2 (Bellman's Optimal Equations for Finite Slots in One Cycle): A feasible network contacts selection policy for cycle l, i.e., $\boldsymbol{\vartheta}_l$ is optimal if and only if it achieves the maximum value of the total expected reward during cycle l. In other words, the optimal reward-to-go functions for cycle l should satisfy the Bellman's optimal equations in (30) [36].

By summarizing the above descriptions, we first present the detailed procedure of the proposed VB-JFBI algorithm as shown in Algorithm 1, wherein the iteration stops until the improvement to the value in forward induction is less than a very small value $\epsilon > 0$. Specifically, the immediate reward matrix and the state transition probability matrix can be obtained by solving a set of LP problems in (19).[4] Then, the optimal policy Θ^* of the proposed infinite MDP problem can be obtained by solving (27) recursively with the forward induction approach. It is noteworthy that the optimal value $\phi_l^*(\boldsymbol{s}(z_{l,1}))$ during each iteration in forward induction should be calculated by utilizing the proposed BICR scheme in Algorithm 2, wherein, the reward-to-go functions are evaluated over all possible states in each TS during a cycle. The

[4] Due to the periodic orbiting movement of satellites, the immediate reward matrix and the state transition probability matrix remain the same for different cycles.

Algorithm 1 Value Based Joint Forward and Backward Induction Algorithm

```
1: Input: $\mathcal{T}$, $\mathcal{G}(\mathcal{V},\mathcal{E}_t)\,\forall t\in\mathcal{T}$, data arrival process ($\mathcal{R}$ and the corresponding probability for each of them).
2: Output: optimal policy matrix $\Theta^*$ and optimal value matrix $V_{opt}$.
3: Initialization: Set $l=1$, initial immediate reward matrix as $R(|S|\times|A|)=\mathbf{0}$, state transition probability matrix as $P(|S|\times|S|\times|A|)=\mathbf{0}$, $\phi^*_{l-1}(\boldsymbol{s}(z_{l-1,1}))(|S|\times 1)=\mathbf{0}$, $V_{opt}(|S|\times|T|)=\mathbf{0}$, and $\Theta^*(|S|\times|T|)=\mathbf{0}$.
4: Define the state set $S$ according to (20) and (21).
5: Utilize BK algorithm to find the set of actions $A=\{A_t,\forall t\in\mathcal{T}\}$ for each decision epoch according to $\mathcal{G}(\mathcal{V},\mathcal{E}_t),\forall t\in\mathcal{T}$.
6: for $t=1$ to $T$ do
7:     Solve the LP problem in (19) for each $\boldsymbol{s}(z_{l,t})$ and $a(z_{l,t})$ to obtain $R(\boldsymbol{s}(z_{l,t}),a(z_{l,t}))$, $\mathbf{B}'(z_{l,t})$, and $\mathbf{EB}'(z_{l,t})$. Put $R(\boldsymbol{s}(z_{l,t}),a(z_{l,t}))$ to the corresponding locations in $R$.
8:     According to $\mathbf{B}'(z_{l,t})$, $\mathbf{EB}'(z_{l,t})$, $\mathbf{r}(z_{l,t})$, and $\mathbf{E}_h(t)$, obtain the potential states in the TS $(z_{l,t+1})\,(t<T)$ and TS $(z_{l+1,1})\,(t=T)$ and their corresponding transition probability $p(\boldsymbol{s}(z_{l,t+1})\,|\boldsymbol{s}(z_{l,t}),a(z_{l,t}))$ during the same cycle and $p(\boldsymbol{s}(z_{l+1,1})\,|\boldsymbol{s}(z_{l,T}),a(z_{l,T}))$ between different cycles based on (22). Put $p(\boldsymbol{s}(z_{l,t+1})\,|\boldsymbol{s}(z_{l,t}),a(z_{l,t}))$ and $p(\boldsymbol{s}(z_{l+1,1})\,|\boldsymbol{s}(z_{l,T}),a(z_{l,T}))$ to the corresponding locations in $P$.
9: end for
10: Repeat
11: Exploit the proposed BICR scheme in Algorithm 2 to calculate the optimal value matrix $\psi^*_l$ for cycle $l$.
12: $\phi^*_l(\boldsymbol{s}(z_{l,1}))=\psi^*_l(|S|,1)$ (i.e., assign the cumulative network reward of the first slot in cycle $l$ for all states obtained by Algorithm 2 to $\phi^*_l(\boldsymbol{s})$ in (28)).
13:  $l=l+1$
14: Until $\left\|\phi^*_l(\boldsymbol{s}(z_{l,1}))-\phi^*_{l-1}(\boldsymbol{s}(z_{l-1,1}))\right\|<\epsilon$.
15: $\Theta^*=\boldsymbol{\vartheta}^*_l$, $V_{opt}=\psi^*_l$.
```

iterations in forward induction continues until the difference between $\phi^*_l(\boldsymbol{s}(z_{l,1}))$ and $\phi^*_{l-1}(\boldsymbol{s}(z_{l-1,1}))$ is below a given threshold.

Algorithm 2 Backward Induction Based Cycle Reward Algorithm

```
1: Input: Optimal value vector $\phi^*_{l-1}(\boldsymbol{s})$, the set of actions $A=\{A_t,\forall t\in\mathcal{T}\}$, the state set $S$, immediate reward matrix as $R(|S|\times|A|)$, state transition probability matrix as $P(|S|\times|S|\times|A|)$, data arrival process ($\mathcal{R}$ and the corresponding probability for each of them), cycle $l$.
2: Output: optimal policy matrix $\boldsymbol{\vartheta}^*$ and optimal reward-to-go value matrix $V_{opt}$.
3: Initialization: Set $t=T$, $\psi^*_l(|S|\times|T|)=\mathbf{0}$, and $\boldsymbol{\vartheta}^*_l(|S|\times|T|)=\mathbf{0}$.
4: while $t>0$ do
5:     for each system state $\boldsymbol{s}(z_{l,t})\in S$ do
6:         Find the optimal action $a_{opt}(z_{l,t})$ for $\boldsymbol{s}(z_{l,t})$ and compute $\psi^*_t(\boldsymbol{s}(z_{l,t}))$ according to Bellman's optimal equations in (30) and put them to the corresponding locations in $\boldsymbol{\vartheta}^*_l$ and $\psi^*_l$.
7:     end for
8:     $t=t-1$
9: end while
```

Based on the VB-JFBI algorithm framework, we further proposed a PB-JFBI algorithm to obtain a stable contact selection policy. The only difference between the PB-JFBI algorithm and VB-JFBI algorithm is the termination condition in step (14) of the algorithm 1. To achieve stable contact selection policy, we replace the termination condition in step (14) of the algorithm 1 with $\boldsymbol{\vartheta}^*_l=\boldsymbol{\vartheta}^*_{l-1}$. We choose to skip the detail of the PB-JFBI algorithm for brevity.

It is noteworthy that once Θ^* is obtained, it acts as a look-up table. During data scheduling, based on the system state, the optimal network contact operation can be determined immediately by referring to the associated entry in such a table. The corresponding optimal contact selection actions in each TS obtained by the proposed JFBI framework can provide an upper bound of network reward to any data scheduling designs of the SSNs.

Remark 1: In practical applications, data arrivals often remain stationary over a significantly long time duration (e.g., dozens of days) [29]. The proposed VB-JFBI and PB-JFBI algorithms can be utilized to obtain the optimal contact selection strategies according to different resource status (storage

$$\psi^{\boldsymbol{\vartheta}_l}_t(\boldsymbol{s}(z_{l,t}))=\begin{cases}R\Big(\boldsymbol{s}(z_{l,t}),\kappa^{\boldsymbol{\vartheta}_l}_t(\boldsymbol{s}(z_{l,t}))\Big)+\sum\limits_{\boldsymbol{s}'(z_{l,t+1})}p\Big(\boldsymbol{s}'(z_{l,t+1})\,|\boldsymbol{s}(z_{l,t}),\kappa^{\boldsymbol{\vartheta}_l}_t(\boldsymbol{s}(z_{l,t}))\Big)\psi^{\boldsymbol{\vartheta}_l}_{t+1}(\boldsymbol{s}'(z_{l,t+1})), & t<T,\\ R\left(\boldsymbol{s}(z_{l,T}),\kappa^{\boldsymbol{\vartheta}}_T(\boldsymbol{s}(z_{l,T}))\right)+\alpha\cdot\sum\limits_{\boldsymbol{s}'(z_{l-1,1})}p\left(\boldsymbol{s}'(z_{l-1,1})\,|\boldsymbol{s}(z_{l,T}),\kappa^{\boldsymbol{\vartheta}}_T(\boldsymbol{s}(z_{l,T}))\right)\phi^*_{l-1}(\boldsymbol{s}'(z_{l-1,1})), & t=T,l>1,\\ R\left(\boldsymbol{s}(z_{l,T}),\kappa^{\boldsymbol{\vartheta}}_T(\boldsymbol{s}(z_{l,T}))\right), & t=T,l=1.\end{cases}\tag{29}$$

$$\psi^*_t(\boldsymbol{s}(z_{l,t}))=\begin{cases}\max\limits_{a_t\in A_t}\left\{R(\boldsymbol{s}(z_{l,t}),a(z_{l,t}))+\sum\limits_{\boldsymbol{s}'(z_{l,t+1})}p(\boldsymbol{s}'(z_{l,t+1})\,|\boldsymbol{s}(z_{l,t}),a(z_{l,t}))\psi^*_{t+1}(\boldsymbol{s}'(z_{l,t+1}))\right\}, & t<T,\\ \max\limits_{a_T\in A_T}\left\{R(\boldsymbol{s}(z_{l,T}),a(z_{l,T}))+\alpha\cdot\sum\limits_{\boldsymbol{s}'(z_{l-1,1})}p(\boldsymbol{s}'(z_{l-1,1})\,|\boldsymbol{s}(z_{l,T}),a(z_{l,T}))\,\phi^*_{l-1}(\boldsymbol{s}'(z_{l-1,1}))\right\}, & t=T,\ l>1,\\ \max\limits_{a_T\in A_T}\{R(\boldsymbol{s}(z_{l,T}),a(z_{l,T}))\}, & t=T,\ l=1.\end{cases}\tag{30}$$

and battery). Small disturbance of data arrival may not affect the contact selection for achieving the optimal or a near optimal network reward. If MMC detects a change in the data arrival distribution, the proposed algorithms can be re-executed to obtain the optimal contact selection actions.

B. Complexity Analysis

In this subsection, we analyze the complexity of the proposed JFBI algorithm framework. Note that the difference between the complexity of the VB-JFBI and PB-JFBI algorithms is only the number of outer iterations, i.e., the iterations in forward induction part (step 10-14 in Algorithm 1). Therefore, we only analyze the complexity of the proposed VB-JFBI algorithm. Hereinafter, we give the specific analysis of the complexity of all operations for VB-JFBI respectively as follows:

- Firstly, we analyze the complexity of the computation of the immediate reward for each TS in one cycle, i.e., the complexity of constructing the matrix $R\left(|S| \times |A|\right)$. We assume that the number of the feasible network contact combinations, i.e., the number of actions, in each TS is γ. Besides, we consider the worst case, where each satellite can establish communication links with each other satellite and ground station during a TS. Hence, to obtain the immediate reward for each TS, we need to solve $|S| \cdot \gamma \cdot T$ LP problems as shown in (19) and $\zeta = U \cdot (U + G) \cdot U$ variables need to be solved in each LP problem. Consequently, the complexity to solve these LP problems [37] is
$$\mathcal{C}_{lp} = \mathcal{O}(|S| \cdot \gamma \cdot T \cdot \zeta^{3.5}) = \mathcal{O}(|S| \cdot \gamma \cdot T \cdot U^7 \cdot (U + G)^{3.5}).$$
- Then, we give the complexity analysis of each outer iteration for cycles, i.e., the forward induction part. We utilize the BICR algorithm in Algorithm 2 to calculate the optimal value for each cycle. Owing to fast memory-retrieving step, the complexity of the backward induction in BICR algorithm is $\mathcal{O}(T)$ [38]. Therefore, the complexity of the operations in the forward induction part is $\mathcal{C}_{fi} = \varpi \cdot \mathcal{O}(T)$, where ϖ is the number of outer iteration for cycles.

Thus, we can derive that the complexity of the VB-JFBI algorithm in the worst case is

$$\begin{aligned} \mathcal{C}_{lp} + \mathcal{C}_{fi} &= \mathcal{O}(|S| \cdot \gamma \cdot T \cdot U^7 \cdot (U + G)^{3.5}) + \varpi \cdot \mathcal{O}(T) \\ &= \mathcal{O}(|S| \cdot \gamma \cdot T \cdot U^7 \cdot (U + G)^{3.5}). \quad (31) \end{aligned}$$

We can note that the complexity of the VB-JFBI algorithm is an algebraic expression about T.

Remark 2: We can observe from (31) that the complexity of the VB-JFBI algorithm is mainly caused by the computation of the reward matrix $R\left(|S| \times |A|\right)$, wherein, $|S|$ grows rapidly as the number of the storage state and battery state of each satellite, i.e., M_1 and M_2, increase. However, it is noteworthy that this paper is dedicated to giving an optimal contact selection strategies for any initial resource status (e.g., storage and battery) according to different resource status (storage and battery). In other words, once we obtain the optimal policy Θ^*, it acts as a look-up table, which means that the complexity of online use is quite low.

V. Simulation Results and Discussions

In this section, we present extensive experiments which are conducted using the real-world satellite parameters supplied from the STK and Matlab simulator to evaluate the performance of the proposed VB-JFBI and PB-JFBI algorithms for SSNs. For performance comparison, we adopt the following two baseline schemes.

- **Fair Contact Selection (FCS)**: In this scheme, contacts tend to be established fairly, which guarantees that each satellite has fair opportunity to establish IS and SD contacts. In other words, feasible contacts selection in this scheme neglects the storage, buffer status and also neglects the time-varying IS channel condition.
- **Myopia**: The myopia algorithm determines the feasible contacts establishment policy aiming at maximizing the instantaneous reward received at each TS for each transmission opportunity. Hence, contact selection policy according to the myopia algorithm is to download as much data as possible to the ground stations in each TS. In other words, future demands are completely ignored in the myopia scheme.

A. Simulation Configuration

We conduct experiments on an SSN scenario with six small satellites and three ground stations. Specifically, three small satellites are distributed over a sun-synchronous orbits at a height of 554.8km and with inclination $86.4°$ and other three small satellites are distributed over a sun-synchronous orbits at a height of 554.8km and with inclination $93.6°$, wherein the orbital period of the each satellite is one hour. Three ground stations are located at Beijing (40°N, 116°E), Kashi (39.5°N, 76°E), and Qingdao (36°N, 120°E). The time duration of each cycle in the proposed SSN scenario is about $\mathcal{T} = 24$hours. Furthermore, we utilize STK to obtain all the potential contacts from 14 Aug. 2018 04:00:00 to 15 Aug. 2018 04:00:00 which can be regarded as the action set during a cycle. In the simulations, we set $\tau = 300$s $P_{ss} = 20$W, $P_{sg} = 20$W, $P_r = 10$W, $P_o = 5$W, $P_h = 20$W, $\eta = 80\%$ [30]. Besides, we set $M_1 = M_2 = 8$, the penalty factor ω in (19) at $\omega = 0.02$, and the parameter ϵ in termination condition in step (14) of the algorithm 1 at $\epsilon = 0.05$. We set the baseline data arrival amount in each TS at ρ (in Gbits). Data arrival for each satellite is ρ with probability 0.7 or 3ρ with probability 0.3 in each TS. Moreover, we set the SD capacity at 30Mbps. According to (1) [30], the IS link capacity is distributed within $[20, 50]$Mbps. The discounting factor α is set at $\alpha = 0.8$, except otherwise stated.

B. Performance Evaluation

We present the simulation results from two aspects. First, we investigate the discounted infinite horizon total reward $V\left(\boldsymbol{s}\left(z_{1,1}\right)\right)$ and the convergence property of the proposed VB-JFBI and PB-JFBI algorithms. Then, we compare the performance of the proposed algorithms with baseline algorithms

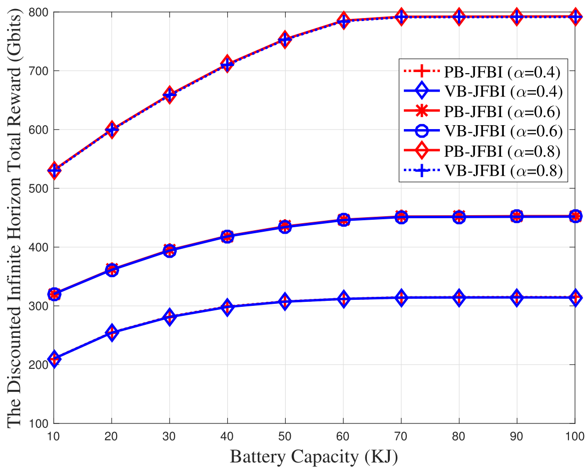

Fig. 3. Performance evaluation for different discounting factors α. The Discounted Infinite Horizon Total Reward $V\left(\boldsymbol{s}\left(z_{1,1}\right)\right)$ versus Battery Capacity EB_{max} with $B_{max} = 30$Gbits.

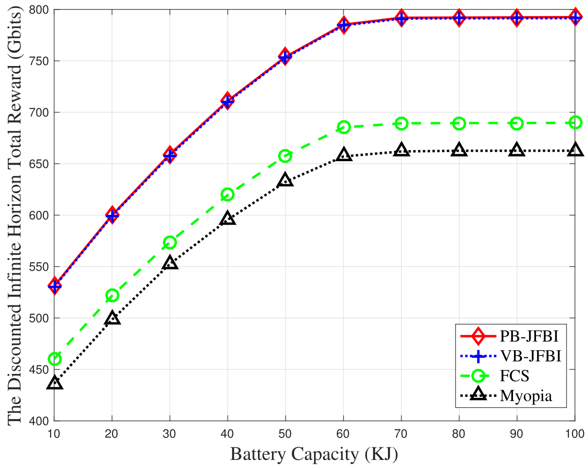

Fig. 4. The Discounted Infinite Horizon Total Reward $V\left(\boldsymbol{s}\left(z_{1,1}\right)\right)$ versus Battery Capacity EB_{max} with $B_{max} = 30$Gbits.

TABLE II
CONVERGENCE PROPERTY OF THE PROPOSED SCHEMES WITH $B_{max} = 30$GBITS, $EB_{max} = 60$KJ

Algorithms	Number of iterations	$V\left(\boldsymbol{s}\left(z_{1,1}\right)\right)$ (Gbits)
VB-JFBI ($\alpha = 0.4$)	8	311.8
PB-JFBI ($\alpha = 0.4$)	41	311.9
VB-JFBI ($\alpha = 0.6$)	12	445.9
PB-JFBI ($\alpha = 0.6$)	63	446.3
VB-JFBI ($\alpha = 0.8$)	23	784.4
PB-JFBI ($\alpha = 0.8$)	80	785.1
VB-JFBI ($\alpha = 0.99$)	86	2839.5
PB-JFBI ($\alpha = 0.99$)	334	2841.4

and investigate impacts of different network parameters on the performance of the proposed algorithms.

Fig. 3 shows that the discounted infinite horizon total reward $V\left(\boldsymbol{s}\left(z_{1,1}\right)\right)$ performances of the two types of optimal policies VB-JFBI and PB-JFBI, corresponding to these two policy values with an initial state $\boldsymbol{s}\left(z_{1,1}\right)$ (initial storage and battery status of every satellite are 10Gbits and 10KJ, respectively). Specifically, we investigate the performance impacts of the battery storage on the proposed schemes under different discounting factors. As expected, with the increase in discounting factor, the discounted infinite horizon total reward performance gain increases. Moreover, it can be observed that $V\left(\boldsymbol{s}\left(z_{1,1}\right)\right)$ performance of the VB-JFBI scheme is quite close to that of the PB-JFBI scheme, which is clearly shown in Table II.

We also investigate the convergence property of the proposed schemes for different discounting factors and present their corresponding performance in Table II. We can observe that the larger the discounting factor, the larger the number of iterations for both two schemes. Table II also clearly depicts that, as expected under the same discounting factor α, the VB-JFBI algorithm performs almost as well as the PB-JFBI algorithm. Specifically, $V\left(\boldsymbol{s}\left(z_{1,1}\right)\right)$ of the PB-JFBI scheme is slightly larger than that of the VB-JFBI scheme for the same discounting factor α, which can further demonstrate and clearly show the small performance difference between the VB-JFBI scheme and the PB-JFBI scheme in Fig. 3. This is accounted by the fact that the number of iterations in the VB-JFBI scheme is larger than that in the PB-JFBI scheme.

Hereafter, we elaborate on the performance comparison for the proposed algorithms and two baseline algorithms. Since the FCS makes contact selection policy decision just according to the topology information in each TS during a cycle, it neglects the battery and storage condition in each TS, which means that the feasible contact selection policy in different cycles remains unchanged. Similarly, the myopia scheme makes the feasible contact selection decision which can download data as much as possible in the current TS and ignore its impact on the future. In other words, the feasible contact selection policy for a cycle also remains unchanged for different cycles. To compare fairly, we adopt the number of iterations in VB-JFBI algorithm to calculate the discounted infinite horizon total reward $V\left(\boldsymbol{s}\left(z_{1,1}\right)\right)$ for the FCS and the myopia algorithms.

Fig. 4 illustrates the discounted infinite horizon total reward $V\left(\boldsymbol{s}\left(z_{1,1}\right)\right)$ performances of different algorithms for different battery capacity EB_{max}. The value function of the discounted infinite horizon total reward $V\left(\boldsymbol{s}\left(z_{1,1}\right)\right)$ is computed with an initial state $\boldsymbol{s}\left(z_{1,1}\right)$ (initial storage and battery status of every satellite are 10Gbits and 10KJ, respectively), for the different policies. Potentially larger battery storage at a satellite results in a larger discounted infinite horizon total reward $V\left(\boldsymbol{s}\left(z_{1,1}\right)\right)$ until the network storage and communication resources become the bottleneck of the system for all data scheduling algorithms. This is explained by the fact that as $EBmax$ increases, the onboard battery can fully harvest and store the solar energy for the forthcoming communications when the solar energy is available. Besides, it can be observed that the proposed VB-JFBI and PB-JFBI algorithms

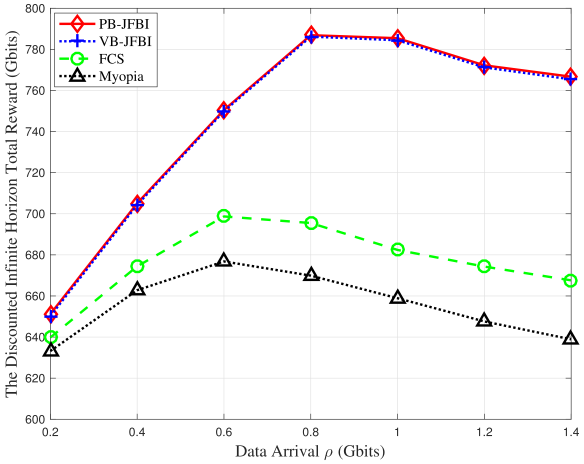

Fig. 5. The Discounted Infinite Horizon Total Reward $V\left(\boldsymbol{s}\left(z_{1,1}\right)\right)$ versus Data Arrival ρ with $B_{max}=30$Gbits and $EB_{max}=60$KJ.

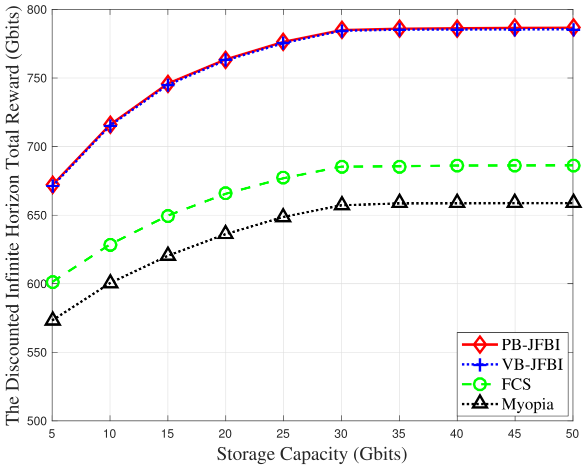

Fig. 6. The Discounted Infinite Horizon Total Reward $V\left(\boldsymbol{s}\left(z_{1,1}\right)\right)$ versus Storage Capacity B_{max} with $EB_{max}=60$KJ.

outperform the myopia and FCS algorithms in terms of $V\left(\boldsymbol{s}\left(z_{1,1}\right)\right)$, which is accounted by the fact that myopia and FCS schemes neglect the future to make contacts selection decision in the current TS. We attribute the result to the balanced matching between dynamic contacts and network state in the VB-JFBI and PB-JFBI algorithms, which boosts resource utilization and prevents energy shortage for future gains.

Fig. 5 shows the discounted infinite horizon total reward $V\left(\boldsymbol{s}\left(z_{1,1}\right)\right)$ for different contact selection policies. The value function of the discounted infinite horizon total reward $V\left(\boldsymbol{s}\left(z_{1,1}\right)\right)$ is computed with an initial state $\boldsymbol{s}\left(z_{1,1}\right)$ (initial storage and battery status of every satellite are 10Gbits and 10KJ, respectively), for the different policies. It can be seen that with the increase of data arrival ρ, the discounted infinite horizon total reward $V\left(\boldsymbol{s}\left(z_{1,1}\right)\right)$ of all four algorithms increases first and then decreases. This trend is expected, because the network storage, battery, and communication resources may become the bottleneck with the increase of data arrival ρ, which can lead to the fact that more data may be stored in satellites. According to the calculation of the immediate reward in each slot shown in (19), the more data satellites store, the smaller $V\left(\boldsymbol{s}\left(z_{1,1}\right)\right)$ the algorithms may achieve. Moreover, the gaps between the proposed algorithms and other two schemes also increase. This is explained by the fact that the network resources may be adequate for the small ρ, wherein even non-optimal contact decisions can still download the stored data. By contrast, for large ρ, effective feasible contact selection strategies play a key role in matching the dynamic network contacts and data demand to maximize $V\left(\boldsymbol{s}\left(z_{1,1}\right)\right)$, which can be demonstrated by the observation that the proposed VB-JFBI and PB-JFBI algorithms yield higher $V\left(\boldsymbol{s}\left(z_{1,1}\right)\right)$ than the myopia and FCS algorithms. In summary, the proposed VB-JFBI and PB-JFBI algorithms can efficiently exploit contact opportunities according to the dynamic network states including storage and battery status, and therefore can achieve better network performance.

We also investigate the impact of $Bmax$ on the system performance as shown in Fig. 6. The value function of the discounted infinite horizon total reward $V\left(\boldsymbol{s}\left(z_{1,1}\right)\right)$ is calculated with an initial state $\boldsymbol{s}\left(z_{1,1}\right)$ (initial storage and battery status of every satellite are 5Gbits and 10KJ, respectively), for the different policies. The use of larger data storage implicitly delays the transmission of some data and stores more data in storage for future transmission that can contribute to $V\left(\boldsymbol{s}\left(z_{1,1}\right)\right)$ if is sufficiently large. Thus, with the increase of the storage capacity $Bmax$, the discounted infinite horizon total reward $V\left(\boldsymbol{s}\left(z_{1,1}\right)\right)$ increases for all schemes. Moreover, it can be observed that there exists the saturated structure, i.e., $V\left(\boldsymbol{s}\left(z_{1,1}\right)\right)$ is gradually saturated for all schemes due to the fact that network battery and communication resources become the bottleneck under this circumstance. Furthermore, due to joint optimization of multi-dimensional resources (i.e., the communication, storage, energy resources), the proposed VB-JFBI and PB-JFBI algorithms can achieve better network performance compared to the FCS and myopia algorithms.

VI. Conclusion

In this paper, we investigate a dynamic data scheduling optimization problem with joint consideration of contact selection, battery management, and buffer management under a stochastic data arrival SSN environment. We propose a finite-embedded-infinite two-level dynamic programming framework and further formulate the stochastic data scheduling problem as an infinite horizon MDP to maximize the discounted infinite-horizon network reward. A JFBI algorithm framework is proposed to efficiently resolve the infinite MDP, wherein a BICR scheme is devised to calculate the value of each iteration in forward induction part. Based on the JFBI algorithm framework, we further propose two specific algorithms, i.e., VB-JFBI and PB-JFBI to obtain the optimal contact selection policy for a cycle. Extensive simulations have been conducted to investigate the impact of different factors, such as battery capacity and storage capacity on network performance and further validate the efficiency of the newly proposed VB-JFBI and PB-JFBI algorithms in terms of the discounted infinite-horizon network reward. This gives us confidence that the proposed MDP framework can be widely adopted in dynamic data scheduling in future satellite systems.

References

[1] M. N. Sweeting, "Modern small satellites-changing the economics of space," *Proc. IEEE*, vol. 106, no. 3, pp. 343–361, Mar. 2018.

[2] J. Alvarez and B. Walls, "Constellations, clusters, and communication technology: Expanding small satellite access to space," in *Proc. IEEE Aerosp. Conf.* Big Sky, MT, USA, Mar. 2016, pp. 1–11.

[3] 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.

[4] D. N. Amanor, W. W. Edmonson, and F. Afghah, "Intersatellite communication system based on visible light," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 6, pp. 2888–2899, May 2018.

[5] F. Aulí-Llinás, M. W. Marcellin, V. Sanchez, J. Bartrina-Rapesta, and M. Hernández-Cabronero, "Dual link image coding for earth observation satellites," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5083–5096, Sep. 2018.

[6] G. Palermo, A. Golkar, and P. Gaudenzi, "Earth orbiting support systems for commercial low earth orbit data relay: Assessing architectures through tradespace exploration," *Acta Astronautica*, vol. 111, pp. 48–60, Jun./Jul. 2015.

[7] 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.

[8] K. Kaneko, Y. Kawamoto, H. Nishiyama, N. Kato, and M. Toyoshima, "An efficient utilization of intermittent surface–satellite optical links by using mass storage device embedded in satellites," *Perform. Eval.*, vol. 87, pp. 37–46, May 2015.

[9] K. B. Chin *et al.*, "Energy storage technologies for small satellite applications," *Proc. IEEE*, vol. 106, no. 3, pp. 419–428, Mar. 2018.

[10] R. Liu, M. Sheng, K.-S. Lui, X. Wang, Y. Wang, and D. Zhou, "An analytical framework for resource-limited small satellite networks," *IEEE Commu. Lett.*, vol. 20, no. 2, pp. 388–391, Feb. 2016.

[11] C. Benson, "Design options for small satellite communications," in *Proc. IEEE Aerosp. Conf.* Big Sky, MT, USA, Mar. 2017, pp. 1–7.

[12] G. Araniti, I. Bisio, M. De Sanctis, A. Orsino, and J. Cosmas, "Multimedia content delivery for emerging 5G-satellite networks," *IEEE Trans. Broadcast.*, vol. 62, no. 1, pp. 10–23, Mar. 2016.

[13] Y. Shi, J. Liu, Z. M. Fadlullah, and N. Kato, "Cross-layer data delivery in satellite-aerial-terrestrial communication," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 138–143, Jun. 2018.

[14] M. de Sanctis, E. Cianca, G. Araniti, I. Bisio, and R. Prasad, "Satellite communications supporting Internet of remote things," *IEEE Internet Things J.*, vol. 3, no. 1, pp. 113–123, Feb. 2016.

[15] 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.

[16] H. Zhu, X. Lin, R. Lu, Y. Fan, and X. Shen, "SMART: A secure multilayer credit-based incentive scheme for delay-tolerant networks," *IEEE Trans. Veh. Technol.*, vol. 58, no. 8, pp. 4628–4639, Oct. 2009.

[17] H. Zhu, C. Fang, Y. Liu, C. Chen, M. Li, and X. S. Shen, "You can jam but you cannot hide: Defending against jamming attacks for Geolocation database driven spectrum sharing," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 10, pp. 2723–2737, Oct. 2016.

[18] H. Zhu, X. Lin, R. Lu, X. Shen, D. Xing, and Z. Cao, "An opportunistic batch bundle authentication scheme for energy constrained DTNs," in *Proc. IEEE INFOCOM*. San Diego, CA, USA, Mar. 2010, pp. 1–9.

[19] 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.

[20] 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.

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

[22] J. Du, C. Jiang, Y. Qian, Z. Han, and Y. Ren, "Resource allocation with video traffic prediction in cloud-based space systems," *IEEE Trans. Multimedia*, vol. 18, no. 5, pp. 820–830, May 2016.

[23] Y. Wang, M. Sheng, J. Li, X. Wang, R. Liu, and D. Zhou, "Dynamic contact plan design in broadband satellite networks with varying contact capacity," *IEEE Commu. Lett.*, vol. 20, no. 16, pp. 2410–2413, Dec. 2016.

[24] M. Abu Alsheikh, D. T. Hoang, D. Niyato, H.-P. Tan, and S. Lin, "Markov decision processes with applications in wireless sensor networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 3, pp. 1239–1267, Apr. 2015.

[25] M. Shifrin, A. Cohen, O. Weisman, and O. Gurewitz, "Coded retransmission in wireless networks via abstract MDPs: Theory and algorithms," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4292–4306, Jun. 2016.

[26] M. Li, L. Zhao, and H. Liang, "An SMDP-based prioritized channel allocation scheme in cognitive enabled vehicular ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 7925–7933, Sep. 2017.

[27] Q. Li, L. Zhao, J. Gao, H. Liang, L. Zhao, and X. Tang, "SMDP-based coordinated virtual machine allocations in cloud-fog computing systems," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1977–1988, Jun. 2018.

[28] J. Du, C. Jiang, J. Wang, Y. Ren, S. Yu, and Z. Han, "Resource allocation in space multiaccess systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 2, pp. 598–618, Apr. 2017.

[29] M. Jamilkowski, K. D. Grant, and S. W. Miller, "Support to multiple missions in the joint polar satellite system (JPSS) common ground system (CGS)," in *Proc. AIAA SPACE Conf. Expo.* Pasadena, CA, 2015, pp. 1–5.

[30] A. Golkar and I. Lluch i Cruz, "The Federated Satellite Systems paradigm: Concept and business case evaluation," *Acta Astron.*, vol. 111, pp. 230–248, Jun. 2015.

[31] 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.

[32] Y. Yang *et al.*, "Towards energy-efficient routing in satellite networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3869–3886, Dec. 2016.

[33] D. P. Bertsekas, *Dynamic Programming Optima Control*, 3rd ed. Belmont, MA, USA: Athena Scientific, 2011.

[34] C. Bron and J. Kerbosch, "Algorithm 457: Finding all cliques of an undirected graph," *Commun. ACM*, vol. 16, no. 9, pp. 575–577, Sep. 1973.

[35] C. Gardiner, *Stochastic Methods*, vol. 4. Berlin, Germany: Springer, 2009.

[36] R. Bellman, *Dynamic Programming*. Mineola, NY, USA: Dover, 2013.

[37] Y. F. Liu, Y. H. Dai, and Z. Q. Luo, "Joint power and admission control via linear programming deflation," *IEEE Trans. Signal Process.*, vol. 61, no. 6, pp. 1327–1338, Mar. 2013.

[38] B. Varan and A. Yener, "Delay constrained energy harvesting networks with limited energy and data storage," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1550–1564, May 2016.

Di Zhou received the B.E. degree in information engineering from Xidian University, Xi'an, China, in 2013, where she is currently pursuing the Ph.D. degree in communication and information systems. She was also a Visiting Ph.D. Student with the Department of Electrical and Computer Engineering, University of Houston, from 2017 to 2018. Her research interests include routing, resource allocation, and mission planning in space information networks.

Min Sheng (M'03–SM'16) received the M.Eng. and Ph.D. degrees in communication and information systems from Xidian University, Shaanxi, China, in 2000 and 2004, respectively. She is currently a Full Professor with the Broadband Wireless Communications Laboratory, School of Telecommunication Engineering, Xidian University. Her general research interests include mobile ad hoc networks, wireless sensor networks, wireless mesh networks, third generation/fourth generation mobile communication systems, dynamic radio resource management for integrated services, cross-layer algorithm design and performance evaluation, cognitive radio and networks, cooperative communications, and medium access control protocols. She has published two books and more than 50 papers in refereed journals and conference proceedings. She was selected as the New Century Excellent Talents in University by the Ministry of Education of China, and obtained the Young Teachers Award by the Fok Ying-Tong Education Foundation, China, in 2008.

Jie Luo (M'03–SM'12) received the B.S. and M.S. degrees in electrical engineering from Fudan University, Shanghai, China, in 1995 and 1998, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Connecticut in 2002. From 2002 to 2006, he was a Research Associate with the Institute for Systems Research, University of Maryland at College Park, College Park. In 2006, he joined Colorado State University, Fort Collins, where he is currently an Associate Professor with the Electrical and Computer Engineering Department. His research focuses on cross-layer design of wireless communication networks, with an emphasis on the bottom several layers. His general areas of research interests include wireless communications, wireless networks, information theory, and signal processing.

Runzi Liu received the B.Eng. degree in telecommunications engineering and the Ph.D. degree in communication and information systems from Xidian University, Xi'an, China, in 2011 and 2016, respectively. Since 2016, she has been with the Broadband Wireless Communications Laboratory, School of Telecommunications Engineering, Xidian University, where she currently holds a faculty postdoctoral position. Her research interests include routing and capacity analysis in ad hoc networks and space networks.

Jiandong Li (M'96–SM'05) received the B.E., M.S., and Ph.D. degrees in communications engineering from Xidian University, Xi'an, China, in 1982, 1985, and 1991, respectively. He has been a Faculty Member with the School of Telecommunications Engineering, Xidian University, since 1985, where he is currently a Professor and the Vice Director of the Academic Committee of the State Key Laboratory of Integrated Service Networks. He was a Visiting Professor with the Department of Electrical and Computer Engineering, Cornell University, from 2002 to 2003. His major research interests include wireless communication theory, cognitive radio, and signal processing. He received the Distinguished Young Researcher Award from the NSFC and the Changjiang Scholar Award from the Ministry of Education, China, respectively. He served as the General Vice Chair for CHINACOM 2009 and a TPC Chair of IEEE ICCC 2013.

Zhu Han (S'01–M'04–SM'09–F'14) received the B.S. degree in electronic engineering from Tsinghua University, in 1997, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Maryland at College Park, College Park, in 1999 and 2003, respectively.

From 2000 to 2002, he was an R&D Engineer at JDSU, Germantown, MD, USA. From 2003 to 2006, he was a Research Associate with the University of Maryland. From 2006 to 2008, he was an Assistant Professor with Boise State University, ID, USA. He is currently a Professor with the Electrical and Computer Engineering Department and the Computer Science Department, University of Houston, TX, USA. His research interests include wireless resource allocation and management, wireless communications and networking, game theory, big data analysis, security, and smart grid. He received an NSF Career Award in 2010, the Fred W. Ellersick Prize of the IEEE Communication Society in 2011, the EURASIP Best Paper Award for the *Journal on Advances in Signal Processing* in 2015, the IEEE Leonard G. Abraham Prize in communications systems (Best Paper Award in IEEE JSAC) in 2016, and several best paper awards in IEEE conferences. He is currently an IEEE Communications Society Distinguished Lecturer.