On the Topology Scaling of Interplanetary Networks

Xiaojian Tian[1] and Zuqing Zhu[1,†]

[1]School of Information Science and Technology, University of Science and Technology of China, Hefei, China

[†]Email: {zqzhu}@ieee.org

Abstract—The increase of deep space (DS) exploration missions suggests that it would be difficult for the existing interplanetary networks (IPNs) to cope with the growing interplanetary data transfer (IP-DT) demands without topology scaling. Therefore, this paper studies how to expand an IPN by deploying new relay satellites and optimizing their orbit parameters in account of the routing and scheduling of IP-DTs, such that the improvement on the IPN's overall performance can be maximized. We formulate an optimization model to tackle the IPN topology scaling problem and propose an effective heuristic for time-efficient problem-solving. Simulations verify the performance of our proposal.

I. INTRODUCTION

The rapid advances on deep space (DS) exploration and interplanetary science missions have led to a growing interest in interplanetary networks (IPNs) [1], which are built for the communications among various DS objects, such as ground stations, satellites, landers, rovers, *etc*. However, other than their counterparts in the networks on/around Earth [2–9], interplanetary data transfer (IP-DT) have to face several unique challenges. First, an IPN normally has a dynamic and unstable topology, which is caused by the movement and shields of spacecrafts and celestial bodies. Second, the communications in IPNs usually have limited bandwidth and long delays, due to the vast distances involved in transmitting signals through DS. Fortunately, delay tolerant networking (DTN) [10] can potentially address these challenges, which enables efficient transmission of data over unreliable and high-latency links by adopting the "store-carry-forward (SCF)" scheme, and thus can greatly improve the reliability and efficiency of IPNs.

Currently, IPNs employ relatively simple architectures with low topology density. However, with the continuous development of DS exploration missions, the number and complexity of spacecraft such as space station, space probes, and satellites will be increasing, imposing higher demands on the bandwidth and capacity of IPNs. Therefore, future IPNs will require a unified network infrastructure to enlarge their topologies and accommodate rising traffic loads. This motivates us to consider the topology scaling of IPNs, *i.e.*, how to expand the topology and link capacity of an IPN effectively. Nevertheless, as the links in an IPN are usually extremely long and always exposed to severe electro-magnetic interferences in the universe, it will be difficult to improve their capacities directly [11, 12]. Hence, to alleviate the bottlenecks caused by rapid traffic growth, it would be more reasonable to improve link capacities by deploying relay satellites [13], as shown in Fig. 1.

Note that, how to expand an IPN with the deployment of relay satellites is a rather complex problem. This is because the orbit parameters of the satellites need to be optimized jointly
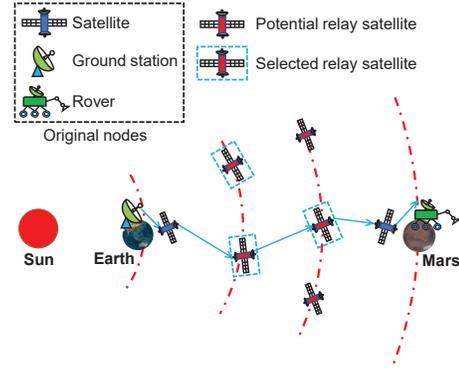

Fig. 1. An example on IPN topology scaling.

with the routing and data scheduling in the IPN. Currently, the studies on the routing and data scheduling in IPN are still in their early stage, with the most representative algorithm as the contact graph routing (CGR) algorithm [14] by the National Aeronautics and Space Administration (NASA). In our previous work [15], we proposed a distributed routing and data scheduling algorithm that can outperform CGR.

All the existing studies on the routing and data scheduling of IP-DTs are under the assumption that the IPN topology is fixed. Nevertheless, the rapid development of DS missions inevitably leads to IPN topology scaling with relay satellite deployments [13]. In [13], the authors analyzed a few schemes for deploying relay satellites, including the commutating ring, minimal Earth ring, elliptical transfer between planetary orbits, *etc.*, and confirmed the merits of multi-hop DS communications with relay satellites. The study in [16] proposed to deploy relay satellites to form a linear-circular commutating chain topology for realizing a broadband multi-hop IPN for Earth-Mars communications. The two-hop relay schemes based on Sun-Earth L4/L5 Lagrange points have been studied in [17, 18]. Wan *et al.* [19] addressed how to design a Solar System satellite relay constellation network to boost up the bandwidth between Earth and Mars. However, these studies all treated the deployment of relay satellites as a static geometric problem just to reduce the lengths of links in the expanded IPN, but as they did not optimize the orbit parameters of relay satellites jointly with the routing and data scheduling of IP-DTs, their performance gains might not be maximized.

To the best of our knowledge, the joint optimization of IPN topology scaling and routing and data scheduling of IP-DTs has not been studied in the literature yet. Hence, in this paper, we tackle the problem and our main contributions are:

- We solve the joint optimization of IPN topology scaling and routing and data scheduling of IP-DTs.

- We formulate and solve an optimization to plan the orbits of new relay satellites, such that the performance gain on IP-DTs can be maximized under a fixed expansion cost.
- We verify the performance of our proposals with extensive simulations. The results show that our approach can achieve cost-effective IPN topology scaling by deploying relay satellites and planning their orbits properly.

The rest of the paper is organized as follows. We explain the network model, and formulate and solve the optimization of IPN topology scaling in Section II. In Section III, we propose a heuristic algorithm to solve our optimization model more time-efficiently. We discuss numerical simulations in Section IV, to evaluate our proposals for IPN topology scaling. Finally, Section V summarizes the paper.

II. IPN TOPOLOGY SCALING IN CONSIDERATION OF IP-DT PERFORMANCE

In this section, we first explain the network model of IPN topology scaling, and then formulate and solve its optimization, which determines the number of relay satellites and plans their orbits, such that the maximum performance gain can be achieved on IP-DTs under a fixed expansion cost.

A. Network Model

To expand the topology of an IPN, we need to consider two types of nodes, which are 1) the original ones that have already been deployed and 2) the relay satellites that can potentially be added into the IPN[1]. Then, the time-varying topology of the expanded IPN can be denoted as $G^t(V, V^R, E^t)$, where V is the set of original nodes, V^R is the set of relay satellites selected from the potential ones, and E^t denotes the set of temporal links at time t. Each temporal link in E^t is defined as $e^t(u, v, t^s, t^e, r, \tau)$, where u and v are its end nodes ($u, v \in V \cup V^R$), t^s and t^e denote the start and end time of its contact, respectively, r is its data-rate, and τ is its transmission latency.

We model the IPN as a discrete-time system that operates on time-slots (TS'), each of which has a fixed duration of Δt. Therefore, each node sets its IP-DT scheme at the beginning of each TS (i.e., $t = 0, \Delta t, 2\Delta t, \cdots$), which means that the system time can be simplified as $t \in \mathcal{T}, \mathcal{T} = \{0, 1, 2, \cdots\}$ after normalization [15]. Note that, each topology scaling is a macro-level operation on the IPN, which may affect it for years before the next one. Hence, although we consider the performance gain on IP-DTs brought by a topology scaling, we will not analyze bundle-level operations on each node here.

Specifically, when analyzing the performance gain on IP-DTs, we consider the average performance over a relatively long time period. Meanwhile, we divide the IPN into several subsystems $Z = \{\tilde{v}, \tilde{u}, \cdots\}$, according to the celestial bodies in it, i.e., each subsystem consists of the landers, rovers, and satellites located on or around a same celestial body. Then, the spatial scale in a subsystem and that between two subsystems can have several magnitudes of difference (i.e., $\sim10^4$ km

---

[1]For simplicity, we assume that all the potential relay satellites use circular orbits centered on the Sun, and the orbits of all the objects related to the IPN (i.e., including the celestial bodies such as Earth and Mars) are coplanar.

versus $\sim10^7$ km or longer). Hence, for a topology scaling, we abstract each subsystem as a virtual node ($\tilde{v} \in Z$) to ignore the bundle-level operations on the nodes in it, and focus on the IP-DTs among the subsystems to analyze the performance gain brought by deploying new relay satellites.

B. Problem Formulation

Our problem formulation uses a polar coordinate system, where the Sun is the pole and the ray from Sun to Earth at $t = 0$ is the polar axis. The length unit in the polar coordination system is an astronomical unit (AU) (i.e., the average distance between Sun and Earth, and $AU \approx 1.496 \times 10^8$ km). Then, the location of a node v at time t can be represented as $P(v, t) = (\rho(v, t), \theta(v, t))$. As the topology scaling needs to select potential relay satellites and determine the routing paths among subsystems, we denote the set of potential relay satellites as R and define two sets of decision variables.

Variables:

- $\mathcal{I} = \{I_v, v \in R\}$: the set of boolean variables for relay satellite selection, where I_v equals 1 if a potential relay satellite $v \in R$ is selected, and 0 otherwise. Note that, the orbit parameters of each potential relay satellite are preset, and the radius and initial phase of the potential relay satellite v are O_v^r and O_v^ϕ, respectively.
- $x_{e^t}^{\tilde{u}, \tilde{v}}$: the boolean variable for inter-subsystem routing, which equals 1 if e^t is included in the routing path between subsystems $\tilde{u}$ and $\tilde{v}$ (i.e., $\tilde{u}$ and $\tilde{v}$ are the virtual nodes after topology abstraction) at TS t, and 0 otherwise.

Optimization Objective:

$$\text{Minimize} \quad \frac{1}{|\mathcal{T}|} \sum_{\{\tilde{u}, \tilde{v} \in Z, \tilde{u} \neq \tilde{v}\}} \sum_{t \in \mathcal{T}} \mathcal{L}_t^{\tilde{u}, \tilde{v}}, \quad (1)$$

where $\mathcal{L}_t^{\tilde{u}, \tilde{v}}$ denotes the IP-DT latency between subsystems $\tilde{u}$ and $\tilde{v}$ at time t. Hence, the optimization objective is to minimize the time average value of the total IP-DT latency among all the subsystems in the expanded IPN.

Then, in order to obtain $\mathcal{L}_t^{\tilde{u}, \tilde{v}}$, we model the state of the outgoing queue on subsystem $\tilde{u}$ for the IP-DT to subsystem $\tilde{v}$ as a birth-death Markov chain with the $M/G/1$ queuing model. Specifically, we assume that the number of bundles arriving at the outgoing queue within a TS follows the Poisson distribution, while the service time of each bundle follows an arbitrary distribution whose mean and variance are known (i.e., τ_s and σ_s^2, respectively). This is because it would be hard to estimate the distribution of bundle lengths in an arbitrary IPN. According to the Pollaczek–Khintchine formula [20], the average delay of a bundle in the $M/G/1$ queue is

$$T_{M/G/1} = \frac{1 + c_s^2}{2} \cdot T_{M/M/1}, \quad (2)$$

where $c_s^2 = \frac{\sigma_s^2}{\tau_s^2}$ is the squared coefficient of variation of service time, and $T_{M/M/1}$ is the average delay of the related $M/M/1$ queue. Specifically, the number of bundles arriving at the related $M/M/1$ queue follows the same Poisson distribution, and the average value of its service time is also the same as

that of the $M/G/1$ queue. Hence, we can obtain $T_{M/M/1}$ as

$$T_{M/M/1} = \frac{1}{\mu - \lambda} \tag{3}$$

where λ and μ are the arrival and service rates of bundles, respectively. Then, by combining Eqs. (2) and (3), we have

$$
\begin{aligned}
T_{M/G/1} &= \frac{1 + c_s^2}{2} \cdot \frac{1}{\mu - \lambda} = \left(\frac{1 + \frac{\sigma_s^2}{\tau_s^2}}{2} \right) \cdot \left(\frac{1}{\frac{1}{\tau_s} - \lambda} \right) \\
&= \left(\frac{1 + \frac{\sigma_l^2}{r^2} \cdot \frac{r^2}{\tau_l^2}}{2} \right) \cdot \left(\frac{1}{\frac{r}{\tau_l} - \lambda} \right) \\
&= \left(\frac{\tau_l^2 + \sigma_l^2}{2 \cdot \tau_l} \right) \cdot \left(\frac{1}{r - \lambda \cdot \tau_l} \right),
\end{aligned}
\tag{4}
$$

where τ_l and σ_l^2 are the mean and variance of bundle size, respectively, and r is the outgoing queue's data-rate.

On the right side of Eq. (4), the only variable is r, which is determined by how we deploy relay satellites between subsystems $\tilde{u}$ and $\tilde{v}$ and select the routing path over them for the related IP-DTs. Hence, we have $T_{M/G/1} \propto \frac{1}{r}$ approximately, *i.e.*, the average delay that bundles experience in the outgoing queue is roughly proportional to $\frac{1}{r}$. Meanwhile, we have

$$\mathcal{L}_t^{\tilde{u},\tilde{v}} = T_{M/G/1} + \frac{L_t^{\tilde{u},\tilde{v}}}{c}, \quad \{\tilde{u}, \tilde{v} \in Z, \tilde{u} \neq \tilde{v}\}, \forall t \in \mathcal{T}, \tag{5}$$

where $L_t^{\tilde{u},\tilde{v}}$ is the length of the routing path between subsystems $\tilde{u}$ and $\tilde{v}$ at time t, and c denotes the speed of light. In Eq. (5), the first term is for the queuing delay and the second one is for the propagation delay. Hence, in order to minimize $\mathcal{L}_t^{\tilde{u},\tilde{v}}$, we need to jointly minimize $\frac{1}{r}$ and $L_t^{\tilde{u},\tilde{v}}$. Note that, r just equals the data-rate of the bottleneck link on the routing path between subsystems $\tilde{u}$ and $\tilde{v}$ at time t.

Constraints:

1) Constraints on Budget of Relay Satellites:

$$\sum_{v \in R} I_v \leq N. \tag{6}$$

Eq. (6) ensures that the total number of newly-deployed relay satellites cannot exceed a preset upper-limit, according to the budget of the IPN topology scaling.

2) Constraints on Inter-Subsystem Routing: We formulate the inter-subsystem routing as a hitting set problem [21], by noticing that any path between subsystems $\tilde{s}$ and $\tilde{d}$ has to intersect the cut $\delta(Z')$ of every set Z', where Z' only contains the virtual nodes for subsystems and potential relay satellites and satisfies $\tilde{s} \in Z'$ and $\tilde{d} \notin Z'$. Meanwhile, the cut $\delta(Z')$ is the link set defined as $\delta(Z') = \{e^t(u, v, t^s, t^e, r, \tau) \in E^t : u \in Z', v \notin Z'\}$. Then, we define a super set $\mathcal{Z}$, each element of which is a possible set Z' that satisfies the aforementioned definition, and obtain the following constraints for routing.

$$\sum_{e^t \in \delta(Z')} x_{e^t}^{\tilde{u},\tilde{v}} \geq 1, \{\tilde{u}, \tilde{v} \in Z, \tilde{u} \neq \tilde{v}\}, \forall t \in \mathcal{T}, \forall Z' \in \mathcal{Z}. \tag{7}$$

Meanwhile, only the links that are provided by the selected relay satellites can be used for inter-subsystem routing, as

$$
\begin{aligned}
& x_{e^t}^{\tilde{u},\tilde{v}} \leq (I_u + I_v)/2, \ \{\tilde{u}, \tilde{v} \in Z, \tilde{u} \neq \tilde{v}\}, \forall t \in \mathcal{T}, \\
& \{e^t(u, v, t^s, t^e, r, \tau) \in E^t : u, v \in Z \cup R, u \neq v\}.
\end{aligned}
\tag{8}
$$

3) Constraints on Link Length: The movement of an object v (*i.e.*, a subsystem or a relay satellite) satisfies the following

dynamics equations, as it moves in a circular orbit.

$$
\begin{cases}
\rho(v, t) = O_v^r, \\
\theta(v, t) = O_v^\phi + \omega_v \cdot t,
\end{cases}
\tag{9}
$$

where ω_v is the angular velocity of v. Then, ω_v can be determined with the Kepler's third law (*i.e.*, $\omega \propto R^{-\frac{3}{2}}$) as

$$\omega_v = \omega_0 \cdot \left(\frac{O_v^r}{R_0} \right)^{-\frac{3}{2}}, \tag{10}$$

where R_0 and ω_0 are the radius and angular velocity of a nearby celestial body (*e.g.*, Earth) that can be used as the reference node of v, respectively. Hence, the link length between two objects v and u can be obtained as

$$L_t^{v,u} = \sqrt{\rho_v^2 + \rho_u^2 - 2\rho_v \cdot \rho_u \cdot \cos(\theta_v - \theta_u)}, \tag{11}$$

where we denote $\rho(v, t)$ as ρ_v, and so on, for simplicity.

4) Constraints on Channel Data-Rate: As a sophisticated model of physical channels in the universe is not the focus and beyond the scope of this work, we calculate the data-rate of links in an IPN according to the Shannon Theory [22]:

$$C = W \cdot \log_2 \left(1 + \frac{E_b}{N_0} \right), \tag{12}$$

where C denotes the channel capacity, W is the channel bandwidth, and $\frac{E_b}{N_0}$ is the signal-to-noise-ratio (SNR). For DS communications, we generally have [23] $\frac{E_b}{N_0} \propto \frac{1}{L^2}$, where L is the channel's physical length. Therefore, we can approximate the capacity of the link between two IPN nodes v and v as

$$
\begin{aligned}
C_t^{v,u} &= W \cdot \log_2 \left[1 + \epsilon \cdot \frac{1}{(L_t^{v,u})^2} \right] \\
&\approx W \cdot \log_2 \left[\epsilon \cdot \frac{1}{(L_t^{v,u})^2} \right],
\end{aligned}
\tag{13}
$$

where ϵ denotes the ratio between the SNR and $\frac{1}{(L_t^{v,u})^2}$, which is related to the physical parameters of the link such as antenna gain, antenna power, *etc.*

Then, as practical DS communications usually use channel encoding, the actual data-rate of a link whose capacity is $C_t^{v,u}$ can be calculated as $r_t^{v,u} = \eta \cdot C_t^{v,u}$, where η represents the channel capacity utilization. In order to further remove physical parameters in our optimization model, we use a link whose data-rate is known as the reference for calculating the data-rates of the remaining links.

$$r_t^{v,u} - r_0 = \eta \cdot W \cdot \left\{ \log_2 \left[\epsilon \cdot \frac{1}{(L_t^{v,u})^2} \right] - \log_2 \left(\epsilon \cdot \frac{1}{L_0^2} \right) \right\}, \tag{14}$$

where r_0 and L_0 denote the data-rate and length of the reference link, respectively. Therefore, we can obtain the data-rate of a link e^t at time t as

$$r_{e^t} = r_0 + \eta \cdot W \cdot \log_2 \left(\frac{L_0^2}{L_{e^t}^2} \right), \ \forall t \in \mathcal{T}, \forall e^t \in E^t. \tag{15}$$

Next, the length of the routing path between two subsystems $\tilde{u}$ and $\tilde{v}$ at time t can be obtained as

$$L_t^{\tilde{u},\tilde{v}} = \sum_{\forall e^t \in E^t} x_{e^t}^{\tilde{u},\tilde{v}} \cdot L_{e^t}, \ \{\tilde{u}, \tilde{v} \in Z, \tilde{u} \neq \tilde{v}\}, \forall t \in \mathcal{T}. \tag{16}$$

Finally, the data-rate of the routing path should satisfy

$$r_t^{\tilde{u},\tilde{v}} \leq x_{e^t}^{\tilde{u},\tilde{v}} \cdot r_{e^t}, \ \{\tilde{u}, \tilde{v} \in Z, \tilde{u} \neq \tilde{v}\}, \forall t \in \mathcal{T}, \forall e^t \in E^t. \tag{17}$$

III. HEURISTIC ALGORITHM DESIGN

The optimization model formulated in the previous section is a mixed integer nonlinear programming (MINLP) problem, which can become intractable when there are many subsystems or/and the number of potential relay satellites is relatively large. Therefore, we design a time-efficient heuristic, namely, the loop-orbit topology scaling (LOTS) algorithm, in the following to solve it. Specifically, LOTS tries to address the selection of potential relay satellites and the calculation of time-varying inter-subsystem routing paths properly to approximate the exact solution of the MINLP.
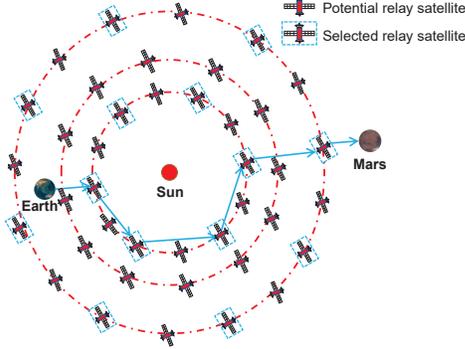

Fig. 2. Example on orbits of potential relay satellites designed for LOTS.

LOTS works on a number of potential relay satellites whose orbits are specially designed, *i.e.*, they are moving on several loop-orbits in groups and the potential relay satellites on the same orbit are placed at equal intervals, as shown in Fig. 2. This setting can greatly simplified the calculation of inter-subsystem routing paths in the expanded IPN, as each inter-subsystem routing path can be tackled with on-loop and inter-loop routing, as indicated by the solid blue lines in Fig. 2. Deploying relay satellites in a same orbit or different orbits helps to provide continuous coverage and improve the IPN's overall connectivity, while the actual deployment should be chosen based on mission demands and the original topology.

Algorithm 1 shows the procedure of the LOTS algorithm. *Line* 1 is for the initialization, where $\mathbf{O}^r$ stores all the radii of the potential relay satellites in R, and V_0^R is a temporary node set to store intermediate results. Then, the for-loop that cover *Lines* 2-28 selects two best orbits and distributes the quota of N new relay satellites on them (*i.e.*, the numbers are N_1 and N_2, respectively). Here, the reason why we assume that the IPN topology scaling only involves relay satellites on two new orbits is mainly because of the budget of a topology scaling (*i.e.*, launching relay satellites centered on the Sun would be fairly expensive). Then, for each possible combination of N_1 and N_2, the topology scaling is divided into two phases: 1) orbit selection, and 2) satellite selection and routing.

5) Orbit Selection: In *Lines* 4-15, the for-loop checks each pair of possible orbits defined by the radii in $\mathbf{O}^r$ to find the two best orbits that help to minimize the objective in Eq. (1). First, for each pair of subsystems $\tilde{u}$ and $\tilde{v}$ in the IPN and each TS, we calculate K shortest paths between $\tilde{u}$ and $\tilde{v}$ in an augmented time-varying topology $G(V \cup R, E^t)$, which includes all the potential relay satellites and the links provisioned by them

---

Algorithm 1: Loop-Orbit Topology Scaling Algorithm

Input: $\{O_u^r, O_u^\phi : u \in Z \cup R\}$, N
Output: $\mathcal{I}$, $\{x_{e^t}^{\tilde{u},\tilde{v}}\}$

1   $\mathbf{O}^r = \{O_u^r : \forall u \in R\}$, $V_0^R = V^R = \emptyset$;
2   **for** *each* $N_1 \in [1, N-1]$ **do**
3     $N_2 = N - N_1$;
4   **for** *each pair of* $\{O_u^r, O_v^r \in \mathbf{O}^r, O_u^r > O_v^r\}$ **do**
5       $y_{u,v} = 0$;
6   **for** *each pair of* $\{\tilde{u}, \tilde{v} \in Z, \tilde{u} \neq \tilde{v}\}$ **do**
7   **for** *each TS* $t \in \mathcal{T}$ **do**
8   find K shortest paths between subsystems $\tilde{u}$ and $\tilde{v}$ in $G(V \cup R, E^t)$ to store in $P_{\tilde{u},\tilde{v}}$;
9   calculate $\mathcal{L}_{t,p}^{\tilde{u},\tilde{v}}$ of each path $p \in P_{\tilde{u},\tilde{v}}$;
10   choose the path p^* with the smallest $\mathcal{L}_{t,p}^{\tilde{u},\tilde{v}}$;
11          $\mathcal{L}_t^{\tilde{u},\tilde{v}} = \mathcal{L}_{t,p^*}^{\tilde{u},\tilde{v}}$, $y_{u,v} = y_{u,v} + \mathcal{L}_t^{\tilde{u},\tilde{v}}$;
12   **end**
13   **end**
14   **end**
15   choose $O_{u^*}^r$ and $O_{v^*}^r$ with the smallest $y_{u,v}$, and include all the relay satellites on the orbits in V_0^R;
16   **for** *each pair of* $O_i^\phi \in [0, \frac{2\pi}{N_1}]$ *and* $O_j^\phi \in [0, \frac{2\pi}{N_2}]$ **do**
17       $z_{i,j} = 0$;
18   **for** *each pair of* $\{\tilde{u}, \tilde{v} \in Z, \tilde{u} \neq \tilde{v}\}$ **do**
19   **for** *each TS* $t \in \mathcal{T}$ **do**
20   find K shortest paths between $\tilde{u}$ and $\tilde{v}$ in $G(V \cup V_0^R, E^t)$ to store in $P'_{\tilde{u},\tilde{v}}$;
21   calculate $\mathcal{L}_{t,p}^{\tilde{u},\tilde{v}}$ of each path $p \in P'_{\tilde{u},\tilde{v}}$;
22   choose the path p^* with the smallest $\mathcal{L}_{t,p}^{\tilde{u},\tilde{v}}$;
23           $\mathcal{L}_t^{\tilde{u},\tilde{v}} = \mathcal{L}_{t,p^*}^{\tilde{u},\tilde{v}}$, $z_{i,j} = z_{i,j} + \mathcal{L}_t^{\tilde{u},\tilde{v}}$;
24   **end**
25   **end**
26   **end**
27   choose $O_{i^*}^\phi, O_{j^*}^\phi$ with the smallest $z_{i,j}$, and insert the corresponding relay satellites in V^R;
28 **end**
29 determine $\mathcal{I}$ and $\{x_{e^t}^{\tilde{u},\tilde{v}}\}$ according to V^R;

(*Line* 8). Fig. 3 and Table I explain the routing path calculation. Fig. 3 shows the positions of the subsystems and two loop-orbits of potential relay satellites. For such a setting, we can calculate $K = 5$ shortest paths between $\tilde{u}$ and $\tilde{v}$ as listed in Table I, where "→" and "⇢" denote an inter-loop link and an on-loop link, respectively. Specifically, the first and last routing paths in Table I are also plotted in Fig. 3, where we omit the actual relay satellites on the orbits for clear illustration.

Next, *Line* 9 calculates the IP-DT latency $\mathcal{L}_{t,p}^{\tilde{u},\tilde{v}}$ of each path p with Eq. (5). Here, the $T_{M/G/1}$ in Eq. (5) can be approximated as $T_{M/G/1} = a \cdot \frac{1}{r_{t,p}^{\tilde{u},\tilde{v}}}$, where a is a positive constant and $r_{t,p}^{\tilde{u},\tilde{v}}$ is the data-rate of path p at TS t, which can be obtained with Eq. (17). We choose the path p^* that can provide the smallest IP-DT latency (*Line* 10), and use its IP-DT latency as that of the routing path between subsystems $\tilde{u}$ and $\tilde{v}$ at TS t (*Line* 11). Here, $y_{u,v}$ is the temporary variable to store the summation of the IP-DT latencies of routing paths

between all the subsystem pairs when orbits O_u^r and O_v^r are chosen. Finally, *Line* 15 chooses the two best orbits O_{u*}^r and O_{v*}^r that provide the smallest $y_{u,v}$, and put all the potential relay satellites on the orbits in set V_0^R.

TABLE I
K SHORTEST ROUTING PATHS FROM $\tilde{u}$ TO $\tilde{v}$

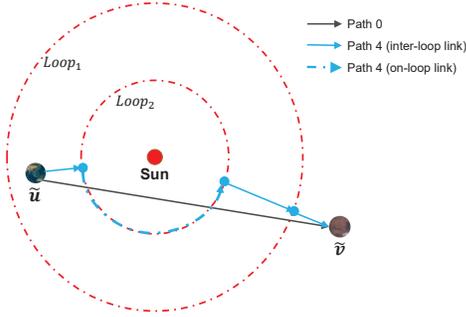| Index	Routing Path
0	$\tilde{u} \to \tilde{v}$
1	$\tilde{u} \to Loop_1 \dashrightarrow Loop_1 \to \tilde{v}$
2	$\tilde{u} \to Loop_2 \dashrightarrow Loop_2 \to \tilde{v}$
3	$\tilde{u} \to Loop_1 \to Loop_2 \dashrightarrow Loop_2 \to \tilde{v}$
4	$\tilde{u} \to Loop_2 \dashrightarrow Loop_2 \to Loop_1 \to \tilde{v}$

Fig. 3. Example on routing paths in augmented topology $G(V \cup R, E^t)$.

6) Satellite Selection and Routing: After selecting the orbits, *Algorithm* 1 proceeds to choose the relay satellites on them and finalize the node-level routing path for each subsystem pair at each TS t. Specifically, as the selected relay satellites on an orbit should be separated at equal intervals, we only need to select the initial phase of one of them and then the others can be determined. Hence, the for-loop of *Lines* 16-26 checks each pair of possible initial phases on the two orbits to finalize the satellite selection and inter-subsystem routing. The procedure here is similar to that in *Lines* 4-15, which chooses the initial phases to minimize the total IP-DT latency of routing paths between all the subsystem pairs. Finally, *Line* 27 selects the two best initial phases and inserts the corresponding relay satellites in V^R to accomplish the IPN topology scaling.

7) Complexity Analysis: The time complexity of *Algorithm* 1 can be analyzed as follows. First, the for-loop of *Lines* 2-28 will run $N-1$ times. For the orbit selection phase, $|\mathbf{O}^r|$ is the number of all the radii of the potential relay satellites, which is a constant. The complexity of the for-loop of *Lines* 6-13 is $O(|Z|^2)$. The for-loop of *Lines* 7-12 will execute $|\mathcal{T}|$ times. The satellite selection and routind phase is overall similar to the orbit selection phase, except that the for-loop of *Lines* 16-26 can run $\frac{|\Phi|^2}{N}$ times at most, where $|\Phi|$ is a constant that depends on the initial phase interval of the potential relay satellites, and the time complexity of computing the K shortest paths is $O(K \cdot N^2)$. Finally, the overall time complexity of *Algorithm* 1 is $O(N \cdot (|Z|^2 \cdot |\mathcal{T}| + \frac{1}{N} \cdot |Z|^2 \cdot |\mathcal{T}| \cdot K \cdot N^2))$, *i.e.*, $O(|Z|^2 \cdot |\mathcal{T}| \cdot K \cdot N^2)$.

IV. PERFORMANCE EVALUATIONS ON TOPOLOGY SCALING

In this section, we discuss the simulation results on IPN topology scaling to evaluate the performance of our proposal.
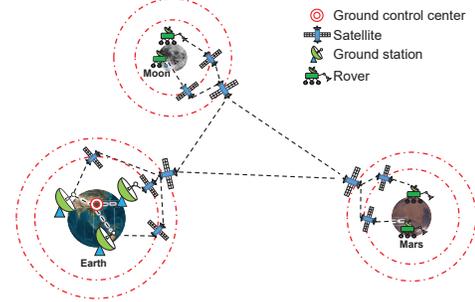

Fig. 4. IPN before topology scaling.

A. Simulation Setup

The topology of the original IPN is shown in Fig. 4, which consists of 18 nodes, including one ground control center, 3 ground stations, 4 rovers (located on Moon or Mars), and 10 satellites (four around Earth, three around Moon and three around Mars). We assume the simulation time cover the period between the last two Mars oppositions (from Oct. 13, 2020 to Dec. 8, 2022, around 780 days), and leverage the Satellite Tool Kit (STK) [24] to calculate and simulate node motions in the IPN. Meanwhile, we design simulations with the realistic settings in STK. For instance, the three ground stations are selected as the deep space station (DSS) 25 in USA, DSS 34 in Australia and DSS 65 in Spain, respectively, and the bandwidth and capacity utilization of each channel is assumed to be $W = 1$ MHz and $\eta = 0.1$, respectively.

As for the potential relay satellites, we set the interval of their orbit radii as $0.01AU$, and the interval of their initial phases is selected as $\frac{\pi}{100}$. We assume that the bundles are in three priorities: 1) $B(q) = 2$ (high priority) is for a bundle that contains the control information sent by the ground control center to a satellite/rover, and its bundle size is within $[1, 8]$ KByte, 2) $B(q) = 1$ (middle priority) is for a bundle that contains the status information returned by a satellite/rover to the ground control center, and its bundle size is within $[16, 64]$ KByte, and 3) $B(q) = 0$ (low priority) is for a bundle that contains the scientific data for being transmitted in the IPN (*i.e.*, the ratios of the bundles for Earth-Moon, Earth-Mars and Moon-Mars are set as 20%, 40%, and 40%, respectively), and its bundle size is fixed as 1024 KByte. The ratio of the bundles in the three priorities is set as $1:1:10$. The lifetime of each bundle follows the negative exponential distribution with an average of $7,200$ seconds. Since we do not consider fine-grained routing and data scheduling in the simulations discussed in this section, we just use CGR [14] for routing and data scheduling in the original and expanded IPNs. To ensure sufficient statistical accuracy, the simulations average the results of 5 independent runs to get each data point.

B. Effects of IPN Topology Scaling

We first compare the performance of IP-DTs in the original and expanded IPNs. Here, we obtain the expanded IPN with both the MINLP in Section II-B and our LOTS algorithm (*i.e.*, *Algorithm* 1), with $N \in \{1, 2\}$, which means that one or two relay satellites will be added in the expanded IPNs. Fig. 5 shows the simulation results, where "BE" denotes scenario with the original IPN before topology scaling, and the algorithms with "-1" and "-2" refer to the scenarios with one and two relay satellites, respectively.

(a) Average E2E latency

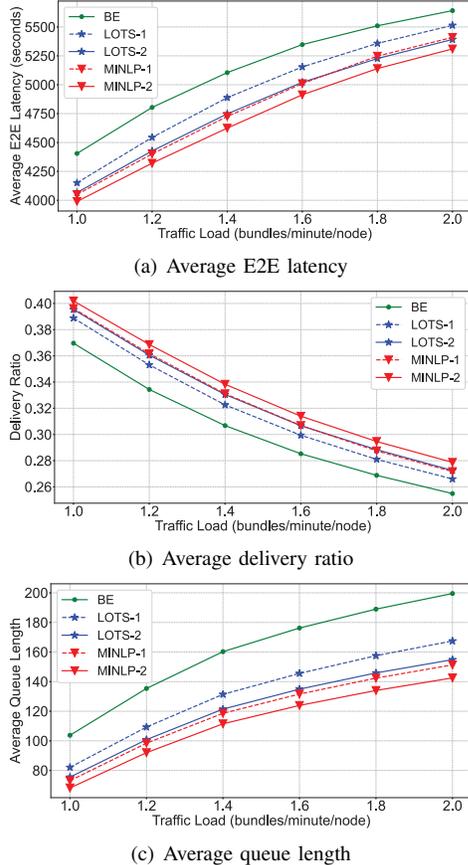(b) Average delivery ratio

(c) Average queue length

Fig. 5. Results of IPN topology scaling (with one and two relay satellites).

Figs. 5(a) and 5(b) plot the results on average E2E latency and average delivery ratio of IP-DTs, respectively. It can be seen that topology scaling indeed can improve the performance of IP-DTs effectively, *i.e.*, the average E2E latency is reduced and the average delivery ratio is increased in all the expanded IPNs. Meanwhile, we observe that there are noticeable performance difference between the IPNs expanded with MINLP and LOTS. Specifically, the average E2E latency and average delivery ratio of IP-DTs in the IPN expanded with LOTS-2 are similar to those in the IPN expanded with MINLP-1. The main reason for the performance gap is that the topology scaling only considers a relatively small number of relay satellites, and thus the approximation made by LOTS might not be good enough. This analysis can be verified by the fact that in Figs. 5(a) and 5(b), the performance gap between MINLP-2 and LOTS-2 is smaller than that between MINLP-1 and LOTS-1. Fig. 5(c) compares the average queue length of

IPN nodes before and after topology scaling. We observe that the average queue length can be significantly reduced after the IPN topology scaling, and the reduction actually becomes more obvious when the traffic load is higher. This further confirms the necessity of IPN topology scaling, especially for addressing the ever-increasing traffic demands in future IPNs.

V. CONCLUSION

This work studied how to expand an IPN with the deployment of relay satellites, such that the orbit parameters of relay satellites can be optimized jointly with the routing and scheduling of IP-DTs for maximizing the improvement on the IPN's overall performance. We tackled the problem by formulating and solving an MINLP model to plan the orbits of new relay satellites, and also proposed an effective heuristic for time-efficient problem-solving. Extensive simulations verified the performance of our proposals. Specifically, our proposal achieved cost-effective IPN topology scaling by deploying relay satellites and planning their orbits properly.

REFERENCES

[1] A. Alhilal, T. Braud, and P. Hui, "The sky is NOT the limit anymore: Future architecture of the interplanetary Internet," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 34, pp. 22–32, Aug. 2019.

[2] Z. Zhu, W. Lu, L. Zhang, and N. Ansari, "Dynamic service provisioning in elastic optical networks with hybrid single-/multi-path routing," *J. Lightw. Technol.*, vol. 31, pp. 15–22, Jan. 2013.

[3] L. Gong *et al.*, "Efficient resource allocation for all-optical multicasting over spectrum-sliced elastic optical networks," *J. Opt. Commun. Netw.*, vol. 5, pp. 836–847, Aug. 2013.

[4] P. Lu *et al.*, "Highly-efficient data migration and backup for Big Data applications in elastic optical inter-datacenter networks," *IEEE Netw.*, vol. 29, pp. 36–42, Sept./Oct. 2015.

[5] L. Gong and Z. Zhu, "Virtual optical network embedding (VONE) over elastic optical networks," *J. Lightw. Technol.*, vol. 32, pp. 450–460, Feb. 2014.

[6] S. Li *et al.*, "Protocol oblivious forwarding (POF): Software-defined networking with enhanced programmability," *IEEE Netw.*, vol. 31, pp. 58–66, Mar./Apr. 2017.

[7] J. Liu *et al.*, "On dynamic service function chain deployment and readjustment," *IEEE Trans. Netw. Serv. Manag.*, vol. 14, pp. 543–553, Sept. 2017.

[8] R. Proietti *et al.*, "Experimental demonstration of machine-learning-aided QoT estimation in multi-domain elastic optical networks with alien wavelengths," *J. Opt. Commun. Netw.*, vol. 11, pp. A1–A10, Jan. 2019.

[9] S. Tang *et al.*, "Sel-INT: A runtime-programmable selective in-band network telemetry system," *IEEE Trans. Netw. Serv. Manag.*, vol. 17, pp. 708–721, Jun. 2020.

[10] S. Burleigh *et al.*, "Delay-tolerant networking: an approach to interplanetary Internet," *IEEE Commun. Mag.*, vol. 41, pp. 128–136, Jun. 2003.

[11] D. Rogstad, A. Mileant, and T. Pham, *Antenna arraying techniques in the deep space network*. John Wiley & Sons, 2005.

[12] H. Kaushal and G. Kaddoum, "Optical communication in space: Challenges and mitigation techniques," *IEEE Commun. Surveys Tuts.*, vol. 19, pp. 57–96, First Quarter 2017.

[13] J. Breidenthal, "The merits of multi-hop communication in deep space," in *Proc. of AESS 2000*, pp. 211–222, Mar. 2000.

[14] G. Araniti *et al.*, "Contact graph routing in DTN space networks: overview, enhancements and performance," *IEEE Commun. Mag.*, vol. 53, pp. 38–46, Mar. 2015.

[15] X. Tian and Z. Zhu, "On the distributed routing and data scheduling in interplanetary networks," in *Proc. of ICC 2022*, pp. 1109–1114, May 2022.

[16] S. Haque, "A broadband multi-hop network for Earth-Mars communication using multi-purpose interplanetary relay satellites and linear-circular commutating chain topology," in *Proc. of AIAA 2011*, pp. 1–28, Jan. 2011.

[17] E. Butte, L. Chu, and J. Miller, "An enhanced architecture for the next generation NASA SCaN study," in *Proc. of AIAA 2016*, pp. 1–28, Oct. 2016.

[18] B. Du, F. Gao, and J. Xu, "The analysis of topology based on Lagrange points L4/L5 of Sun-Earth system for relaying in Earth and Mars communication," in *Proc. of ICCSN 2017*, pp. 533–537, May 2017.

[19] P. Wan and Y. Zhan, "A structured Solar System satellite relay constellation network topology design for Earth-Mars deep space communications," *Int. J. Satell. Commun. New.*, vol. 37, pp. 292–313, Oct. 2019.

[20] G. Jain and K. Sigman, "A Pollaczek–Khintchine formula for M/G/1 queues with disasters," *J. Appl. Probab.*, vol. 33, pp. 1191–1200, Dec. 1996.

[21] M. Goemans and D. Williamson, *Approximation Algorithms for NP-hard Problems*. PWS Publishing Co., 1996.

[22] C. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, Jul. 1948.

[23] J. Mukherjee and B. Ramamurthy, "Communication technologies and architectures for space network and interplanetary Internet," *IEEE Commun. Surveys Tuts.*, vol. 15, pp. 881–897, Second Quarter 2013.

[24] Satellite tool kit. [Online]. Available: http://www.agi.com/products/stk/.