Dynamic Planning of Inter-Plane Inter-Satellite Links in LEO Satellite Networks

Jiahao Pi‡, Yongyi Ran*†, Hao Wang‡, Yanyun Zhao‡, Ruili Zhao‡, Jiangtao Luo*†
School of Communication and Information Engineering
Chongqing University of Posts and Telecommunications, Chongqing, China
‡{S200101200, S190101069, S200101135, D200101029}@stu.cqupt.edu.cn, †{Ranyy, Luojt}@cqupt.edu.cn

ICC 2022 - IEEE International Conference on Communications | 978-1-5386-8347-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICC45855.2022.9838251

***Abstract*—Low Earth Orbit (LEO) satellite constellations are promising to provide global coverage and low latency communication by deploying a large number of small satellites and widely establishing Inter-Satellite Links (ISLs). However, due to the high motion of the LEO satellites, fixed inter-plane ISLs cannot provide long-time continuous connectivities and guarantee high-throughput communication performance. The existing dynamic planning approaches almost only consider part of the constellation information and cannot derive the optimal inter-plane ISLs. This paper proposes a dynamic Inter-plane Inter-satellite Links Planning method based on Multi-Agent deep reinforcement learning (MA-IILP) to optimize the total throughput and inter-plane ISL switching rate. We formulate a Partially Observable Markov Decision Process (POMDP) model with taking into account the Euclidean distance, communication rate and link switching cost. We derive the optimal strategy by utilizing Multi-Agent Deep Deterministic Policy Gradient (MADDPG) algorithm with a centralized training and distributed execution paradigm. Finally, extensive experiments are carried out and the results illustrate that our proposed approach can increase the total throughput of the target constellation by 2.8%∼7.2%, and decrease the inter-plane ISL switching rate by 30.7%∼68.4% compared to the state-of-the-art baseline algorithms.**

***Index Terms*—Low Earth Orbit (LEO) satellites, satellite constellations, inter-plane ISLs, deep reinforcement learning**

I. INTRODUCTION

In recent years, Low Earth Orbit (LEO) satellite constellations have become an emerging and promising technology to provide low latency, broadband communications and global coverage for ground users, and are expected to play a vital role in 6G communication. Many leading companies such as SpaceX, OneWeb, and Amazon are attempting to deploy a mega LEO satellite constellation for stable broadband Internet services. The LEO satellites can be connected through Inter-Satellite Links (ISLs) by exploiting optical or visible light communication systems, which include: intra-plane ISLs, connecting neighboring satellites from the same orbital plane, and inter-plane ISLs, connecting satellites from different orbital planes. The intra-plane ISLs are rather stable because inter-satellite distances can be sustained a long time within an orbital plane. However, the inter-satellite distances between different orbital planes are time-variant: longest when satellites are over the Equator, and shortest over the polar region boundaries. Besides, the orbital periods will be different if the planes are deployed at different altitudes, leading to aperiodic topologies. Therefore, any connectivity schemes with fixed inter-plane ISLs cannot well satisfy the change of constellation topology, it is critical to plan the inter-plane ISLs dynamically.

Due to the environmental characteristics and hardware limitations of LEO constellations, it is challenging to determine the inter-plane ISLs dynamically. First, the dynamic movement and high dimensions of LEO constellations make the planning of inter-plane ISLs dramatically complex. Hundreds of satellites move at around 7.5 km/s relative to ground users, and each of them has a set of inter-plane neighbors for establishing ISLs. This incurs that the planning of inter-plane ISLs is NP-hard and faces the problem of "curse of dimensionality". Second, only partial information can be observed by each satellite due to the limited Line-of-Sight (LoS) distance. It is costly to collect global constellation information in real-time, while partial information can easily cause sub-optimization. An efficient approach should be designed to derive a globally optimal planning scheme. Third, one satellite may be eligible to establish an ISL with its several neighboring satellites, and there exists competition and cooperation among the satellites in the same orbit. Competing with other satellites for one satellite is to improve the quality of its inter-plane ISL, while cooperating is to maximize the total throughput of the constellation. It is essential to achieve a good tradeoff between competition and cooperation for the whole constellation.

Most existing work focuses on analyzing the features and models of the inter-plane ISLs without planning the inter-satellite connectives. For example, a power budget model is proposed in [1] to analyze the effect of slant range on power requirements and a thorough analysis of ISL connectivity is provided in [2] by investigating the visibility between satellites and their antenna steering capabilities. These efforts only derive some references for inter-plane ISLs and not provide any specific ISL planning schemes. The basic ISL planning algorithms are heuristic [3]–[5], which derive schemes through greedy, simulated annealing and other methods according to partial information of a LEO constellation, so it is easy to result in local optimality. Another representative approach is proposed by [6], which models the ISLs network with Finite State Automation (FSA) and solves it with Integer Linear Programming (ILP). However, this algorithm requires huge

This work is supported by National Natural Science Foundation of China (No.62171072, No.62172064, No.62003067), Natural Science Foundation of Chongqing (cstc2021jcyj-msxmX0586).

computation efforts, and is not suitable for high-dimensional and highly dynamic LEO constellations.

In order to address the aforementioned issues, we propose a dynamic Inter-plane Inter-satellite Links Planning method based on Multi-Agent deep reinforcement learning (MA-IILP) to optimize the total throughput and inter-plane ISL switching rate. We firstly formulate a Partially Observable Markov Decision Process (POMDP) model with designing state space, action space and reward function according to Euclidean distance, LoS distance, communication rate and switching cost. Then, we utilize Multi-Agent Deep Deterministic Policy Gradient (MADDPG) to find the optimal strategy. MADDPG works in a centralized training and decentralized execution paradigm, where each agent makes actions based on its local observations and is trained with observations and actions of all the agents. Our main contributions are summarized as follows:

- We formulate the optimization objective as a utility function, which consists of inter-plane ISL communication rate and antenna switching cost. The antenna switching cost is for establishing the new ISLs.
- We model the optimization problem as a POMDP and solve it using MA-IILP algorithm with centralized training and distributed execution. We design an additional penalty mechanism to achieve a good tradeoff between competition and cooperation for the whole constellation.
- In order to handle the problem of "curse of dimensionality" and accelerate the convergence, we train the dynamic planning algorithm orbit by orbit.
- Extensive experiments are carried out and the results show that MA-IILP algorithm can significantly reduce the switching rate, improve the total throughput as well as the average number of established inter-plane ISLs.

II. Related Work

In this section, we review the recent research on inter-plane ISLs, which can be classified into two categories: the feature and model analysis of inter-plane ISLs [1], [2], [7] and the dynamic planning of inter-plane ISLs [3]–[6].

A. The feature and model analysis of inter-plane ISLs

Many existing studies have focused on analyzing the diverse features and models of inter-plane ISLs. In [1], the authors analyzed the power budget for communication between CubeSats by investigating how the slant range determines the power requirements. Visibility between neighboring satellites and their antenna steering capabilities were investigated by Lee *et al.* [2] to analyze the time-varying ISL connectivity. In addition, the authors in [2] pointed out that the antenna steering angle is a feasible solution for estimating the antenna operation cost, which provides a direction for estimating the inter-plane ISL switching cost in this paper. Chen *et al.* [7] presented a computational model for geometric parameters of inter-plane ISLs, including the maximum and minimum distances of inter-plane ISLs.

B. The dynamic planning of inter-plane ISLs

With the development of research on inter-plane ISL features and models, a few inter-plane ISL planning methods have been proposed. Leyva-Mayorga *et al.* [3] proposed a Greedy Independent Experiments Matching (GIEM) algorithm based on greedy algorithm to research inter-plane ISL planning with the objective of maximizing the total throughput. After verifying the Markovianity, Greedy Markovian satellite Matching (GMM) algorithm was proposed to plan inter-plane ISLs with the primary objective of reducing switching rate. To research the planning of laser ISLs in navigation constellations, the authors in [4] set up a multi-objective model to balance the inter-satellite communication and orbit determination performance. Liu *et al.* [5] investigated the assignment of hybrid laser/radio ISLs network, and adopted the Multi-Objective Simulated Annealing algorithm (MOSA) to plan laser ISLs. The methods mentioned above are all heuristic and are only capable of finding sub-optimal solutions. In order to find the optimal delay and throughput of data transmission in Global Navigation Satellite Systems (GNSSs), the authors in [6] modeled the satellite network as a FSA and solved it based on ILP. However, this method is limited by the huge computational demand.

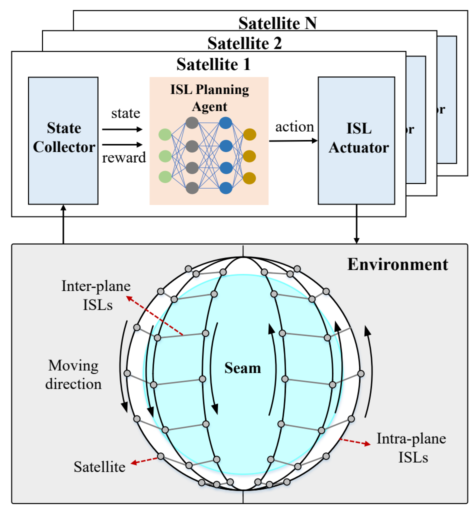

Fig. 1. Satellite ISLs topology and ISL decision networks.

III. System Model and Problem formulation

A. System Architecture

As shown in Fig. 1, we consider a polar orbit constellation, where N satellites are evenly distributed in M planes. Each orbital plane $m \in \{1, 2, \ldots, M\}$ is deployed at a given altitude h_m, inclination ϵ_m near 90 deg. Each orbital plane consists of N_m evenly distributed satellites. Besides, we denote the position of satellite u as (x_u, y_u, z_u) in a rectangular coordinate system, and define $p(u)$ as the orbital plane of satellite u. Generally, each satellite is equipped with a total of four ISLs. Two intra-plane ISLs connect neighboring satellites from the same plane, whereas two inter-plane ISLs connect satellites from different planes.

Each satellite has an ISL decision network, in which the ISL planning agent is trained through the received rewards and states until convergence. The state collector obtains the states and rewards by interacting with other satellites in the environment. The ISL planning agent makes decisions based on the state information collected by the state collector, and the ISL actuator establishes inter-plane ISLs with the corresponding satellites according to the instructions from the ISL planning agent.

B. Communication Model

This paper assumes a decision period T_d , and the number of decisions N_d = T/T_d, where T is the period of constellation. At any decision moment, the constellation can be expressed as a undirected graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N}$ is the vertex set, representing the satellites, and $\mathcal{E}$ is the edge set, representing the ISLs. We define the relative direction of vertex v w.r.t. vertex u as

$$d\left(uv\right) = \begin{cases} -, & p(u) > p(v) \\ 0, & p(u) = p(v) \\ +, & p(u) < p(v) \end{cases}, \tag{1}$$

and respectively denote the number of neighbor vertices in $+$ and $-$ directions relative to u as $\deg_{\mathcal{G}}^{+}(u)$, $\deg_{\mathcal{G}}^{-}(u)$. We refer to a pair of source-destination satellites u and v as satellite pair uv. Furthermore, we define the source satellite as standard satellite and the destination satellite as target satellite.

Limited by Doppler effect and LoS, inter-plane ISLs cannot be established between some satellite pairs in LEO constellations. If a satellite pair can establish an inter-plane ISL, we define the satellite pair as eligible satellite pair. In the following, we will filter the set of eligible satellite pairs.

We denote the Euclidean distance between a satellite pair uv as

$$||uv||= \sqrt{\left(x_u - x_v\right)^2 + \left(y_u - y_v\right)^2 + \left(z_u - z_v\right)^2}. \tag{2}$$

If the Euclidean distance between a satellite pair is longer than their LoS distance, the LoS will be sheltered by the Earth. We define the LoS distance between a satellites pair uv as $l(uv)$, and if $||uv||> l(uv)$, the satellite pair is not eligible satellite pair. The latter can be written as

$$l(uv) = \sqrt{h_{p(u)}\left(h_{p(u)} + 2\mathrm{R_E}\right)} + \sqrt{h_{p(v)}\left(h_{p(v)} + 2\mathrm{R_E}\right)}, \tag{3}$$

where $\mathrm{R_E}$ is the radius of the earth.

The satellites in the first plane and the *M*-th plane move in opposite directions with large relative velocities. It is challenging to maintain ISLs in “seam” area, and cross-seam ISLs will not be considered throughout our analyses. Since we focus on inter-plane ISLs, satellite pairs located in the same plane are not eligible satellite pairs. According to the above analysis, the set of eligible satellite pairs can be denoted as

$$E = \left\{uv : |p(u) - p(v)| \notin \{0, M-1\}, ||uv||< l(uv)\right\}. \tag{4}$$

Satellites communicate in a free-space environment. Therefore, inter-satellite communication is mainly affected by free-space path loss (FSPL) and the thermal noise assumed to be additive white Gaussian (AWGN) [8]. For eligible satellite pairs, we consider their features as follows.

The FSPL between an eligible satellite pair uv is given as

$$L(u,v) = \left(\frac{4\pi||uv||f}{c}\right)^2, \tag{5}$$

where c is the speed of light and f is the carrier frequency. At any moment, the SNR between an eligible satellite pair uv can be expressed as

$$\mathrm{SNR}(u,v) = \frac{P_t G_t G_r}{k_B \tau B L(u,v)}, \tag{6}$$

where P_t is the transmission power, G_t and G_r are the transmitter antenna gain and receiver antenna gain respectively, k_B is the Boltzmann constant, τ is the thermal noise in Kelvin, and B is the channel bandwidth in Hertz.

This paper assumes that all satellites have sufficiently narrow antenna beams and have precise beam alignment capabilities. Therefore, satellites can communicate in an interference-free environment. The maximum data rate that satellite u can choose to communicate with satellite v in an interference-free environment is given as

$$R_{\mathrm{SNR}}(u,v) = B\log_2\Big(1 + \mathrm{SNR}(u,v)\Big). \tag{7}$$

C. Switching Cost Model

The antenna steering angle of satellite u from aligning satellite v_1 to aligning satellite v_2 is calculated as

$$\theta_u^{v_1 v_2} = \arccos\left(\frac{\|uv_1\|^2 + \|uv_2\|^2 - \|v_1 v_2\|^2}{2\|uv_1\| \cdot \|uv_2\|}\right). \tag{8}$$

To measure the impact of inter-plane ISL switching cost, We define the average antenna steering angle $\bar{\theta}_u$ for each satellite u. The latter can be given as

$$\bar{\theta}_u = \frac{\sum\limits_{v_1 \neq v_2 \in E_u^+} \theta_u^{v_1 v_2} + \sum\limits_{v_1 \neq v_2 \in E_u^-} \theta_u^{v_1 v_2}}{\binom{N_u^+}{2} + \binom{N_u^-}{2}}, \tag{9}$$

where E_u^+, E_u^- are the sets of satellites v in $+$ and $-$ directions relative to u which satisfy the condition $uv \in E$ respectively, and N_u^+, N_u^- are the number of sets E_u^+, E_u^- respectively.

For the n-th decision, we denote all edges connected between eligible pairs in the graph as $\mathcal{G}_{\mathcal{M}_n}$. For the edges in $\mathcal{G}_{\mathcal{M}_n}$, we define $\theta_{uv}\left(n\right)$ as the antenna steering angle of the edge between u and v. The latter can be given as

$$\theta_{uv}\left(n\right) = \begin{cases} 0, & uv \in \mathcal{G}_{\mathcal{M}_{n-1}} \\ \bar{\theta}_u + \bar{\theta}_v, & uv \notin \mathcal{G}_{\mathcal{M}_{n-1}} \end{cases}. \tag{10}$$

D. Problem Formulation

In order to guarantee high total throughput of the constellation and to minimize inter-plane ISL switching rate, we consider the joint optimization problem of the total throughput and inter-plane ISL switching cost over period T.

At every decision, the establishment of inter-plane ISLs can be treated as a matching problem. For the matching graph $\mathcal{G}_{\mathcal{M}_n}$ at the n-th decision, we define the utility function $w(\mathcal{G}_{\mathcal{M}_n})$ as the achievable profit subtracted by the switching cost [9], which can be written as

$$w\left(\mathcal{G}_{\mathcal{M}_n}\right) = \sum_{uv \in \mathcal{G}_{\mathcal{M}_n}} \rho R_{\mathrm{SNR}}(u,v) - \lambda \theta_{uv}\left(n\right), \tag{11}$$

where ρ is the profit of each communicate rate, λ is the unit operating cost of antenna steering angle.

Hence, the optimization problem is to maximize the satellite network utility by obtaining the optimal matching graph set $\mathcal{G}_{\mathcal{M}} = \{\mathcal{G}_{\mathcal{M}_1}, \ldots, \mathcal{G}_{\mathcal{M}_{N_d}}\}$, which can be formulated as

$$\begin{aligned} &\max \sum_{n=1}^{N_d} w(\mathcal{G}_{\mathcal{M}_n}) = \sum_{n=1}^{N_d} \sum_{uv \in \mathcal{G}_{\mathcal{M}_n}} \rho R_{\mathrm{SNR}}(u,v) - \lambda \theta_{uv}\left(n\right) \\ &s.t. \quad \begin{cases} \deg^{d}_{\mathcal{G}_{\mathcal{M}_n}}(u) \in \{0,1\}, \quad \forall u \in \mathcal{N}, d \in \{-,+\} \\ uv \in E. \end{cases} \end{aligned} \tag{12}$$

IV. Algorithm Design

The above optimization problem is modeled as a POMDP, where the information observed by each satellite is only the partial glimpse of the constellation state. In this section, MA-IILP algorithm is proposed to solve the optimization problem of the total throughput and inter-plane ISL switching cost.

A. POMDP

In order to handle the problem of "curse of dimensionality" and train the planning algorithm orbit by orbit, we select agents from satellite networks and design appropriate actions for them. During the motion of satellite u, we define its side near the plane $((p\left(u\right)+1) \bmod M)$ as the positive side, and the other side as the negative side. Each satellite actively decides the positive side ISL, while the negative side ISL passively accepts the decisions from satellites on the negative side. Due to the existing of "seam", there is no need for satellites in the M-th plane to actively decide the positive side ISLs. Therefore, except for the satellites in the M-th plane, all satellites are independent agents. Then, we will define each element of reinforcement learning for each agent.

State space. We define the state space of agent i as $S_i = \{D_i, L_i, R_i\}$, where D_i is the set of distance between agent i and satellites in the next plane within its LoS. L_i is the target satellite connecting with its positive side ISL, R_i is the rate of the positive side ISL. At different decisions, the state space of each agent is time-variant due to the motion of satellites.

Action space. We define the action space of agent i as $A_i = \{\mathcal{V}_i, K\}$, where $\mathcal{V}_i$ is the set of satellites in the next plane within its LoS and K is keeping silent. Once agent i selects the action $a_i \in \mathcal{V}_i$, the agent will establish an inter-plane ISL with the corresponding target satellite on its positive side. If agent i chooses the action $a_i = K$, the agent will not establish its positive side ISL.

Reward. We define the contribution of agent i as r_i, which can be calculated as

$$r_i = \begin{cases} 0, & a_i = K \\ \alpha_i \rho R_{\mathrm{SNR}}(i, a_i) - \lambda \theta_{i a_i}, & a_i \in \mathcal{V}_i \end{cases}, \tag{13}$$

where α_i is decision conflict discount of agent i. Since each agent makes decisions independently based on its partial observations, agents in the same plane may choose the same target satellite. Therefore, we select a trainer to reconsider the contribution of each agent based on the partial observations and actions received from all the agents. That is, agents whose decisions have no conflict with other agents have $\alpha_i = 1$, while agents whose decisions conflict with other agents have the following design: agents are added to different lists according to their target satellites. For each agent i, if the communication rate of its inter-plane ISL established with the target satellite is the largest in the list, then $\alpha_i = 0.8$, otherwise, $\alpha_i = 0.1$.

Since each agent cooperatively maximizes the same optimization goal, each agent has the same reward $\sum\limits_{i=1}^{N_a} r_i$, where $N_a = N - N_m$ is the number of agents.

B. MA-IILP Algorithm

We obtain inter-plane ISL planning method by utilizing MADDPG, which combines strategy-based learning and value-based learning [10]. MADDPG works in a centralized training and decentralized execution paradigm. Therefore, after the training converges, each agent can make decisions independently according to its partial observation. The pseudo-code of MA-IILP is shown in Algorithm 1, where initializations and training processes are the same for each agent.

Algorithm 1 The training of MA-IILP

```
for agent i = 1, N_a do
    Initialize the policy network π_i with random weights θ_i, the value
    network Q_i^{π_i} and corresponding target network
end for
Initialize an experience buffer D
for episode = 1 to M' do
    Initialize a Gumbel-Softmax distribution for exploration
    Receive initial state x_t = (s_1, ..., s_{N_a})
    for t = 1 to max-episode-length do
        for agent i = 1, N_a do
            Select and execute action a_i ~ π_i(·|s_i)
            Establish an ISL with the corresponding target satellite
            Oberserbe reward r_t and next state x_{t+1}
        end for
        Store transition (x_t, a_t, r_t, x_{t+1}) into D
        Set x_t ← x_{t+1}
        for agent i = 1, N_a do
            Sample a random batch (x_j, a_j, r_j, x_{j+1}) from D
            Set y_i^j = r_i^j + γ Q_i^{π'}(x_{j+1}, a_1' ..., a_{N_a}')|_{a_k' ~ π'(s_{k,j})}
            Update value network by minimizing the loss L(θ_i) in (16)
            Update policy network by the policy gradient method in (17)
            Update target network parameters in (18)
        end for
    end for
end for
```

Each agent i has its own policy network π_i with the weight θ_i that producing differentiable samples through a Gumbel-Softmax distribution [11]. Each agent i has a value network $Q_i^{\pi_i}(\mathbf{x}, a_1, \ldots, a_{N_a})$, where $\mathbf{x} = (s_1, \ldots, s_{N_a})$. Besides, each agent i has a corresponding target policy network π_i' with the weight θ_i' and a target value network $Q_i^{\pi_i'}(\mathbf{x}, a_1, \ldots, a_{N_a})$. An experience memory $\mathcal{D}$ is also initialized to restore environment transitions, and a random min-batch will be sampled for the update of policy and value networks from the experience memory.

For each decision epoch t, agent i will observe local distance information D_i, the positive side ISL information L_i and R_i. Based on current strategy π_i, current state $s_{i,t} = \{D_i, L_i, R_i\}$, and gumbel noise, agent i will select and execute an action $a_{i,t}$. The assignment of $a_{i,t}$ is given by:

$$a_{i,t} \sim \pi_i\left(\cdot | s_{i,t}\right). \tag{14}$$

Then the inter-plane ISL is established between agent i and the corresponding target satellite. Each agent will transit current state $s_{i,t}$ to $s_{i,t+1}$ and receive reward $r_{i,t}$. After obtaining all these information, the experience memory will record a transition $(\mathbf{x}_t, \mathbf{x}_{t+1}, \mathbf{a}_t, \mathbf{r}_t)$, where $\mathbf{a}_t = (a_{1,t}, \ldots, a_{N_a,t})$ and $\mathbf{r}_t = (r_{1,t}, \ldots, r_{N_a,t})$. Next, a minibatch $(\mathbf{x}_j, \mathbf{x}_{j+1}, \mathbf{a}_j, \mathbf{r}_j)$ is sampled randomly from $\mathcal{D}$. At the end of the decision epoch t, the policy networks are updated with the policy gradient method. The target value y_i^j can be written as

$$y_i^j = r_i^j + \gamma Q_i^{\boldsymbol{\pi}'}(\mathbf{x}_{j+1}, a_1', \ldots, a_{N_a}')|_{a_k' \sim \boldsymbol{\pi}'(s_{k,j})}, \tag{15}$$

where $\boldsymbol{\pi}' = \{\pi_1', \ldots, \pi_{N_a}'\}$ is the set of target policy networks with delayed parameters θ_i'. The value networks are updated by minimizing the loss

$$\mathcal{L}(\theta_i) = \mathbb{E}_{\mathbf{x}_j, a, r, \mathbf{x}_{j+1}}\left[\left(Q_i^{\boldsymbol{\pi}}\left(\mathbf{x}_j, a_1, \ldots, a_{N_a}\right) - y_i^j\right)^2\right], \tag{16}$$

where $\boldsymbol{\pi} = \{\pi_1, \ldots, \pi_{N_a}\}$. Moreover, the weights of the policy networks can be updated with the policy gradient method

$$\nabla_{\theta_i} J\left(\theta_i\right) = \mathbb{E}_{\mathbf{x}_j \sim \mathcal{D}, a_i \sim \pi_i}\left[\nabla_{a_i} Q_i^{\pi}\left(\mathbf{x}_j, a_1, \ldots, a_{N_a}\right) \nabla_{\theta_i} a_i\right]. \tag{17}$$

Then, with (16) and (17), the weights of the target networks can be updated by

$$\theta_i' \leftarrow \beta \theta_i + (1-\beta)\,\theta_i'. \tag{18}$$

The network parameters of each agent are trained until convergence.

V. Evaluation

A. Simulation Setup

Simulator for the MA-IILP algorithm has been developed in Python 3. Simulation parameters are shown in Table. I. In MA-IILP algorithm, one-hidden-layer fully-connected neural networks are parameterized for policy and value networks and the corresponding target networks. We update the network parameters after every 1000 samples added to the experience memory. With the length of experience memory $\mathcal{D}$ 10000, the mini-batch size 1024, the learning rate 0.01, and the discount rate $\gamma = 0.95$, the proposed MA-IILP algorithm is trained until convergence. Fig. 2 illustrates the average rewards after 75000 episodes attained by MA-IILP when orbital plane numbers $M = 6, 8, 10$.

TABLE I
PARAMETER SETTINGS FOR EVALUATION

Parameter	Symbol	Value
Number of satellites per plane	N_m	11
Number of orbital planes	M	$\{6, 8, 10\}$
Altitude of orbital planes	h_m	780 km
Longitude of orbital planes	ϵ_m	86.4 deg
Equivalent isotropically radiated power	EIRP	8912.5 W
Carrier frequency in the Ka-band	f	23.28 GHz
Carrier bandwidth	B	15 MHz
Quality factors	G/τ	8 dB/K
Number of inter-plane transceivers	Q	2
Decision period	T_d	300 s

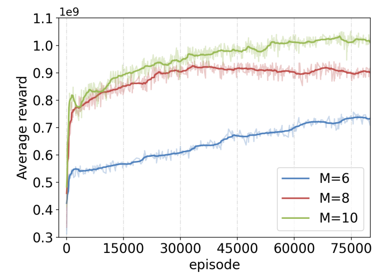

Fig. 2. Agents average reward after 75000 episodes with $M = 6, 8, 10$

B. Baseline Algorithms

We compare our algorithm (MA-IILP) in the same environment with the following algorithms [3].

1) GIEM: A dynamic inter-plane ISL planning algorithm based on greedy algorithm.
2) GMM: An extension of the GIEM algorithm where inter-plane ISLs are maintained for as long as possible.
3) Geographical matching algorithm (GEO): The latitude of the constellation is divided into N_m logical regions, and inter-plane ISLs are established for satellites belonging to the same logical region.

C. Experiment Results

Fig. 3 shows that MA-IILP has the largest mean number of inter-plane ISLs in each decision among four algorithms. The main reason is that MA-IILP always establishes an inter-plane ISL for each satellite on each side. However, GIEM, GMM and GEO are unable to make global considerations when making inter-plane ISL decisions. For some satellites, there is no satellite being eligible to establish ISLs with them. In this case, satellites cannot establish inter-plane ISLs.

Fig. 4 showcases the improvement of the total throughput provided by MA-IILP. Although maximizing the total throughput is the primary goal of GIEM, it only reaches approximate optimality because it is based on greedy algorithm. Therefore, when considering the appropriate switching cost, the performance of GIEM is not as good as MA-IILP. Since the primary

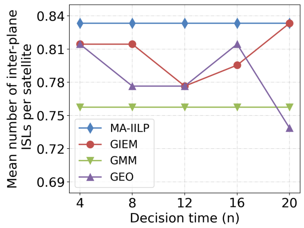

Fig. 3. Mean number of inter-plane ISLs per satellite per four decisions with $M = 6$

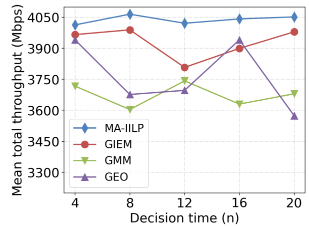

Fig. 4. Mean total throughput as a function of the SNR per four decisions with $M = 6$

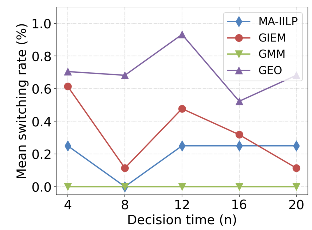

Fig. 5. Mean switching rate of inter-plane ISLs per four decisions with $M = 6$

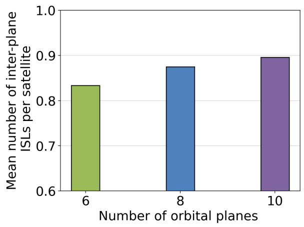

Fig. 6. Mean number of inter-plane ISLs per satellite per decision with MA-IILP

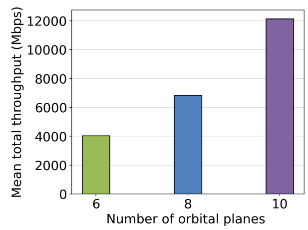

Fig. 7. Mean total throughput as a function of the SNR per decision with MA-IILP

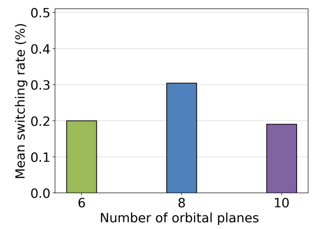

Fig. 8. Mean switching rate of inter-plane ISLs per decision with MA-IILP

goal of GMM is to reduce switching rate, as expected, the total throughput obtained by this algorithm is much smaller than MA-IILP in most decisions.

Fig. 5 illustrates the switching rate of inter-plane ISLs achieved with four algorithms respectively. We can see that in most decisions MA-IILP is much less than GIEM. The reason is that MA-IILP considers the tradeoff between the total throughput and switching rate. The GMM is always minimum among four algorithms, because its primary objective is minimizing switching cost. The GEO algorithm always has the largest switching rate among the four algorithms, because satellites in the same logical region are time-varying due to the motion of the satellites.

D. Sensitivity Analysis

Fig. 6 demonstrates that the mean number of inter-plane ISLs per satellite achieved by MA-IILP shows an increase as the number of planes increases. Except for "seam" region, MA-IILP always creates nearly fully connected inter-plane ISLs for the constellation globally. As the size of the constellation increases, the percentage of "seam" region becomes smaller. Therefore, the mean number of inter-plane ISLs per satellite increases. Fig. 7 demonstrates the massive gains in the total throughput of the constellation as the number of orbital planes increases. There are two main reasons for this: 1) MA-IILP is able to establish more inter-plane ISLs, and 2) the communication rate per ISL is higher. The inter-plane ISL switching rate performance is given in Fig. 8. As the orbital plane number increases, the inter-plane ISL switching rate of the MA-IILP is consistently below 30.4%.

VI. Conclusion

In this paper, we have proposed a MA-IILP algorithm to achieve the optimal inter-plane ISL planning strategy, which jointly optimizes the total throughput and inter-plane ISL switching cost in LEO constellations. The optimization problem is investigated to achieve the maximum expected discounted reward. Based on the target networks and experience memory, the MA-IILP algorithm can efficiently learn the optimal strategy and each satellite can distributedly decide inter-plane ISLs. Experiment results are presented to demonstrate the better performance of MA-IILP algorithm compared to baseline algorithms.

References

[1] O. Popescu, "Power budgets for cubesat radios to support ground communications and inter-satellite links," *Ieee Access*, vol. 5, pp. 12 618–12 625, 2017.

[2] Y. Lee and J. P. Choi, "Connectivity analysis of mega constellation satellite networks with optical inter-satellite links," *IEEE Transactions on Aerospace and Electronic Systems*, pp. 1–1, 2021.

[3] I. Leyva-Mayorga, B. Soret, and P. Popovski, "Inter-plane inter-satellite connectivity in dense leo constellations," *IEEE Transactions on Wireless Communications*, vol. 20, no. 6, pp. 3430–3443, 2021.

[4] W. Chengzhuo, L. Suyang, G. Xiye, and Y. Jun, "Dynamic optimization of laser inter-satellite link network topology based on genetic algorithm," in *2019 14th IEEE International Conference on Electronic Measurement & Instruments (ICEMI)*. IEEE, 2019, pp. 1331–1342.

[5] S. Liu, J. Yang, X. Guo, and L. Sun, "Inter-satellite link assignment for the laser/radio hybrid network in navigation satellite systems," *GPS Solutions*, vol. 24, no. 2, pp. 1–14, 2020.

[6] Z. Yan, G. Gu, K. Zhao, Q. Wang, G. Li, X. Nie, H. Yang, and S. Du, "Integer linear programming based topology design for gnsss with inter-satellite links," *IEEE Wireless Communications Letters*, vol. 10, no. 2, pp. 286–290, 2020.

[7] Q. Chen, L. Yang, X. Liu, B. Cheng, J. Guo, and X. Li, "Modeling and analysis of inter-satellite link in leo satellite networks," in *2021 13th International Conference on Communication Software and Networks (ICCSN)*. IEEE, 2021, pp. 134–138.

[8] 3GPP, "Solutions for NR to support Non-Terrestrial Networks (NTN)," 3rd Generation Partnership Project (3GPP), Technical report (TR) 38.821, 01 2020, version 16.0.0.

[9] N. Zhao, Y. Cheng, Y. Pei, Y.-C. Liang, and D. Niyato, "Deep reinforcement learning for trajectory design and power allocation in uav networks," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.

[10] R. Lowe, Y. Wu, A. Tamar, J. Harb, P. Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," *Neural Information Processing Systems (NIPS)*, pp. 6382–6393, 2017.

[11] E. Jang, S. Gu, and B. Poole, "Categorical reparameterization with gumbel-softmax," in *International Conference on Learning Representations (ICLR)*, 2017.