DeepISL: Joint Optimization of LEO Inter-Satellite Link Planning and Power Allocation via Parameterized Deep Reinforcement Learning

Yue Li‡, Jiangtao Luo*†, Yongyi Ran†, Jiahao Pi‡
School of Communications and Information Engineering,
Chongqing University of Posts and Telecommunications, Chongqing, China
‡{S210131117, S200101200}@stu.cqupt.edu.cn, †{Luojt, Ranyy}@cqupt.edu.cn

GLOBECOM 2023 - 2023 IEEE Global Communications Conference | 979-8-3503-1090-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/GLOBECOM54140.2023.10437157

***Abstract*—The Low Earth Orbit (LEO) satellite constellation has been recognized as an important component of the future 6G network. Due to the high speed movement and limited on-board energy of LEO satellites, as well as the uneven distribution of service requests on the ground, it is difficult to achieve optimal satellite communication performance using a static inter-satellite links (ISLs) scheme and a fixed transmission power per link. To solve this problem, this paper proposes a joint optimization algorithm based on parameterized deep reinforcement learning (named DeepISL) for dynamic planning of ISLs between different planes and transmit power allocation per link. First, a partially observable Markov decision process (POMDP) is established by modeling the communication, energy, and overall energy efficiency as well as the antenna steering costs. Second, to solve the hybrid action space problem with discrete action variables for ISL planning and continuous action variables for power allocation, deep multi-agent reinforcement learning with parameterized action space is used to obtain the optimal joint strategy. Finally, extensive experiments illustrate that our proposed algorithm can improve the energy efficiency of the constellation by $4.2\% \sim 10.5\%$ compared to the comparison algorithms, and can also achieve better performance in terms of throughput and ISLs switching ratio.**

***Index Terms*—LEO satellite, inter-satellite-link planning, power allocation, parameterized deep reinforcement learning.**

I. Introduction

The Low Earth Orbit (LEO) satellite constellation is an emerging and promising technology to provide broadband communications, low latency services and global coverage for ground users [1]. In order to provide efficient communication among users in different satellite coverage areas without the relay of any ground station, inter-satellite links (ISLs) are usually established between satellites. However, because of the high-speed motion of satellites, inter-plane ISLs connecting satellites at different orbits cannot be maintained for a long time and need to be regulated and switched dynamically according to the real-time satellite constellation states. At the same time, a high transmission power for an ISL usually can achieve a higher data transmission rate, but the available energy of a satellite is limited. Therefore, it is critical to jointly plan the inter-plane ISLs and regulate their the power allocation with the aim of improving the performance of the LEO satellite networks.

However, it is challenging to achieve the above objective due to the dynamic environmental states of the LEO satellite constellation. First, there are a large number of satellites in the LEO constellation, and each satellite has a group of neighboring satellites used to build ISLs, training a large amount of data can have a negative impact, which leads to the "curse of dimensionality". Second, there may exist "switching conflicts" when one satellite switches its inter-plane ISL from one connected satellite to other candidate satellites. If more than one satellites from the same orbit require to establish an ISL with the same satellite at the adjacent orbit, the "switching conflicts" will occur. Third, the joint optimization of planning the inter-plane ISLs and regulating their the power allocation will encounter the issue of hybrid action space, because planning ISLs is a discrete action while power allocation is a continuous action. The joint optimization algorithm should well deal with the issues of hybrid action space.

Currently, most existing works focus only on the dynamic planning of inter-plane ISLs without considering the power allocation of the ISLs. The greedy matching algorithm is proposed in [2], which selects the inter-plane ISL with the highest throughput and tends to result in poor energy efficiency. Finite state automation (FSA) is used to model ISLs in [3] and solve inter-plane ISLs planning based on integer linear programming (ILP), but this requires a large amount of computation. A multi-agent deep reinforcement learning scheme is proposed in [4], which enables optimal planning decisions for ISLs. The ISLs assignment problem is treated as a general graph matching by [5], and the ISLs assignment strategy is established based on Signed Variance and Blossom algorithm. Although the above-mentioned algorithms can achieve a matching strategy with excellent performance of ISLs, they fail to consider the dynamic power allocation, and thus cannot flexibly switch ISLs and allocate power according to the actual amount of data.

To solve the above problem, we propose a joint optimization algorithm for ISLs planning and power allocation based on parameterized deep reinforcement learning, named DeepISL. In this algorithm, each agent makes decisions based on its own

*The corresponding author is Jiangtao Luo.

*This work is jointly supported by National Natural Science Foundation of China (No.62171072, 62172064, 62003067), Natural Science Foundation of Chongqing (cstc2021jcyj-msxmX0586) and Chongqing Postgraduate Research and Innovation Project (CYB21204).

observations and is trained with its own information. After the algorithm converges, each agent can make optimal decisions. Our major contributions are summarized below:

- The joint optimization problem is formulated as a partially observable Markov decision process (POMDP) since each satellite can only observe partial information of the satellite constellation. To avoid "curse of dimensionality", we train the algorithm orbit-by-orbit and a "switching conflicts" penalty mechanism is designed to weigh the decisions between satellites.
- We propose a parametrized Deep Q-Network (P-DQN) based algorithm to solve joint optimization problem with a hybrid discrete-continuous action space, where ISLs planning is a discrete action and power allocation is a continuous action.
- Extensive experiments are carried out and the results show that the DeepISL can improve the energy efficiency of the LEO constellation while increasing the throughput and reducing the ISL switching ratio.

II. RELATED WORK

In this section, we review the recent research on dynamic planning and power allocation of inter-plane ISLs.

A. Dynamic Planning for Inter-Plane ISLs

Most research algorithms on dynamic planning of ISLs focus on heuristic algorithms, linear integer programming algorithms and deep reinforcement learning. In [2], a greedy matching algorithm is proposed to build inter-plane ISLs with the objective of maximizing the throughput. The heuristic-based algorithm is easy to implement, but it tends to result in poor energy efficiency. Yan *et al.* [3] modeled the network of ISLs with finite state automation (FSA) and solved the planning problem of inter-plane ISLs based on integer linear programming (ILP), however, this algorithm is extremely complex and unsuitable for high-dimensional satellite constellations. To solve the problem of high complexity, Pi *et al.* [4] proposed an approach based on multi-agent deep reinforcement learning to train the algorithm orbit by orbit.

The existing approaches mentioned above can partially handle with the dynamic characteristics of the LEO satellite constellations and improve the performance of ISLs, but almost all of them fail to take into account the dynamic power allocation, thus they cannot effectively and flexibly achieve the ISL switching and power allocation according to the actual data transmission demands.

B. Dynamic Power Allocation for Inter-Plane ISLs

The dynamic power allocation of inter-plane ISLs can flexibly regulate the power according to actual service demand to achieve the purpose of saving energy. To optimize the communication quality, Jia *et al.* [6] designed an MRR-ANC (modified retro reflectors-analog network coding) system in two way relay channel(TWRC) and proposed a power allocation scheme to maximize throughput. Chen *et al.* [7] studied the optimal power control problem for spectrum sharing in cognitive satellite-terrestrial networks(CSTNs) and solved the proposed power control problem by a game theoretic approach to maximize the throughput.

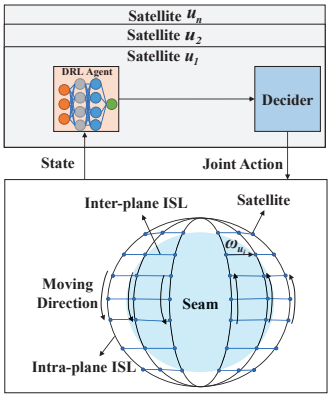

Fig. 1: LEO satellite constellation topology and decision network

The purpose of the above methods is to maximize throughput without taking into account energy efficiency, which may lead to over-allocation of power, resulting in a waste of power resources.

III. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Architecture

As shown in Fig. 1, we consider a polar orbit constellation. The satellite constellation has N satellites and M orbital planes, with N_m satellites uniformly distributed on each orbital plane $m \in \{1, ...M\}$ at an inclination ϵ_m, and the altitude of m is H. For a satellite u, we define its Cartesian coordinates as (x_u, y_u, z_u) and the orbital plane it is located in as m_u. At any time, the constellation can be represented as an undirected graph $g = (V, E)$, where V denotes the set of vertices (satellites), E denotes the set of edges (ISLs), and e_{uv} denotes the link between satellite u and satellite v.

B. Communication Model

We define the source-target satellites u and v as the satellite pair uv. For satellite u, we define the plane $((m_u + 1) \mod M)$ as its positive side plane and the plane on the other side as its negative side plane. Herein we assume that each satellite can keep two intra-plane ISLs and two inter-plane ISLs. The intra-plane ISLs of a satellite connect neighboring satellites within the same orbital plane, while the inter-plane ISLs of a satellite connect neighboring satellites from the adjacent orbital planes.

For the satellite pair uv, the line-of-sight (LoS) distance between satellite u and satellite v is [2]

$$l_{|uv|} = 2\sqrt{H\left(H + 2R_E\right)}, \tag{1}$$

where R_E denotes the radius of the Earth, H is the altitude of orbital planes. And the Euclidean distance can be written as

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

If inter-plane ISL can be established between satellite pairs uv, we call such a satellite pair as eligible satellite pair. However, some satellite pairs are unable to establish inter-plane ISL because their LoS is sheltered by the Earth, and such a satellite pair is defined as $d_{|uv|} > l_{|uv|}$.

The region between 1-st and M-th orbital planes is called "Seam". The satellites in the 1-st and M-th orbital planes move in opposite directions with very high velocities. Therefore, it is extremely challenging to maintain inter-plane ISLs in the "Seam" region, and the establishment of inter-plane ISLs across "Seam" will not be considered in this paper. Then, the set of eligible satellite pairs can be expressed as [4]

$$Y = \left\{ uv : m_u - m_v \notin \{0, M-1\} \ \ and \ d_{|uv|} < l_{|uv|} \right\}. \quad (3)$$

Since satellites communicate in a free-space environment, satellites communication is mainly affected by free-space path loss (FSPL) and additive white Gaussian noise (AWGN). For a eligible satellite pair uv, FSPL is defined as

$$L_{uv} = \left[\left(4\pi d_{|uv|} f \right) / c \right]^2, \quad (4)$$

where f denotes the carrier frequency and c denotes the speed of light. The SNR between satellite pair of uv is

$$SNR_{uv} = \frac{P_{tran} G_{tran} G_{rec}}{k_B T_e B L_{uv}}, \quad (5)$$

where P_{tran} is the transmitted power, k_B is the Boltzmann constant, T_e is the thermal noise in Kelvin, and B is the channel bandwidth, G_{tran} and G_{rec} are the gain of the transmitter antenna and receiver antenna, respectively.

In this paper, we assumed that the satellites have enough narrow antenna beams with accurate beam alignment. As a result, the satellites can communicate in an interference-free environment. The maximum communication rate between an eligible satellite pair of uv is

$$R_{e_{uv}} = B \log_2 \left(1 + SNR_{uv} \right). \quad (6)$$

C. Energy Model

We assume that the system is time-slotted, the duration of the time slot is δ, the t-th time slot is denoted as $\delta(t)$ and a satellite period has N_d time slots. The solar panels equipped on the satellite collect energy from the solar radiation when the satellite is in sunlight and the energy consumed by the satellite is preferentially extracted from the collected energy. However, when the satellite is on the backside, the satellite cannot collect energy because the Earth shields it from the sun's rays, and the consumed energy needs to be extracted from the battery. We express the power of the energy collected by the satellite u for the t-th time slot as [8]

$$P_{u,t}^{harvest} = \tau \cdot \varphi \cdot \varpi \cdot A_e \cdot \sin \sigma, \quad (7)$$

where τ denotes the energy collection constant, $\tau = 1$ when the satellite is on the sunny side and $\tau = 0$ when the satellite is on the backside, φ denotes the energy conversion efficiency of the solar panel to convert solar energy into electrical energy, ϖ is the solar irradiation per unit area, A_e is the area of the solar panel, and σ denotes the angle between the solar panel and the sunlight. The energy collected by the satellite u for the t-th time slot can be expressed as

$$E_{u,t}^{harvest} = P_{u,t}^{harvest} \cdot \delta(t). \quad (8)$$

For the entire satellite period, we assume that the maximum capacity of the battery of satellite u is C_u^{max} and the upper limit of allocated power is P_u^{max}. For the t-th time slot, the energy of satellite u at the beginning of the time slot is $C_{u,t}$, the maximum real-time distributable power $P_{u,t}^{max}$ is

$$P_{u,t}^{max} = \begin{cases} \min\left(C_{u,t}/\delta(t), P_u^{max} \right), & \tau = 0 \\ \min\left(P_{u,t}^{harvest} + C_{u,t}/\delta(t), P_u^{max} \right), & \tau = 1 \end{cases}. \quad (9)$$

Assuming that the total power allocated to the intra-plane ISLs and other operation is P_0, the power allocated by satellite u to the inter-plane ISL in its positive side plane direction is $P_{e_{uv},t}$, the total power allocated by satellite u can be expressed as

$$P_{u,t}^{consume} = P_{e_{uv},t} + P_0 \le P_{u,t}^{max}. \quad (10)$$

Therefore, the energy consumed by the satellite u is

$$E_{u,t}^{consume} = P_{u,t} \cdot \delta(t). \quad (11)$$

Then, the difference between the energy collected and the energy consumed during $\delta(t)$ is $\delta\left(E_{u,t}\right) = E_{u,t}^{harvest} - E_{u,t}^{consume}$. Thus, the capacity of the battery of the $(t+1)$-th time slot is [9]

$$C_{u,t+1} = \begin{cases} \min\left(C_{u,t} + \delta\left(E_{u,t}\right), C_u^{max} \right), & \delta\left(E_{u,t}\right) > 0 \\ \max\left(C_{u,t} + \delta\left(E_{u,t}\right), 0 \right), & \delta\left(E_{u,t}\right) < 0 \end{cases} \quad (12)$$

D. Energy Efficiency Model

In order to reasonably allocate the power for satellites, we introduce energy efficiency in this paper. For the t-th time slot, we assume that the number of packet arrivals at satellite u obeys a Poisson distribution with mean ρ, the size of each packet is F_f. The amount of data arriving at satellite u is $\omega_{u,t}$, and the amount of data actually sent by satellite u to satellite v is $\omega_{e_{uv},t}$, the energy efficiency of e_{uv} is defined as

$$E_{eff,t}^{uv} = \omega_{e_{uv},t} / E_{e_{uv},t} = R_{e_{uv},t} / P_{e_{uv},t}, \quad (13)$$

where $E_{e_{uv},t}$ denote the energy consumed during $\delta(t)$, $R_{e_{uv},t}$ is communication rate and $P_{e_{uv},t}$ is the transmitted power.

In addition, the communication rate of each inter-plane ISL must be doubly constrained in order to transmit most of the data and to avoid wasting energy by communicating at a rate that exceeds the service requirements. This constraint is

$$\lambda \omega_{u,t}/\delta(t) \le R_{e_{uv},t} \le \omega_{u,t}/\delta(t), \quad (14)$$

where λ is the satisfaction factor.

E. Antenna Steering Cost Model

In this paper, we estimate the antenna steering cost by antenna steering angle, the antenna steering angle for switching the inter-plane ISL of satellite u from v_1 to v_2 is

$$\theta_u = \arccos\left(\frac{\left(d_{|uv_1|}\right)^2 + \left(d_{|uv_2|}\right)^2 - \left(d_{|v_1 v_2|}\right)^2}{2 \cdot d_{|uv_1|} \cdot d_{|uv_2|}} \right). \quad (15)$$

To calculate the antenna steering cost of inter-plane ISLs, we define the average antenna steering angle $\hat{\theta}_u$ for each satellite u. For the t-th time slot, $\hat{\theta}_{u,t}$ is defined as [4]

$$\hat{\theta}_{u,t} = \frac{\sum_{v_1 \ne v_2 \in Y_{u,t}^+} \theta_u + \sum_{v_1 \ne v_2 \in Y_{u,t}^-} \theta_u}{\binom{N_{u,t}^+}{2} + \binom{N_{u,t}^-}{2}}, \quad (16)$$

where $Y_{u,t}^+$ and $Y_{u,t}^-$ represent the set of satellites v that satisfy the condition $uv \in Y_t$ in the positive and negative side planes of the satellite u, respectively, Y_t is the set of eligible satellite

pairs for the t-th time slot. And $N_{u,t}^{+}$, $N_{u,t}^{-}$ represent the number of eligible satellite pairs in sets $Y_{u,t}^{+}$ and $Y_{u,t}^{-}$.

For the t-th time slot, in order to establish the inter-plane ISL e_{uv} , we define the antenna steering angle as

$$\theta_{uv,t} = \begin{cases} 0 & , e_{uv} \in E_{t-1} \\ \hat{\theta}_{u,t} + \hat{\theta}_{v,t} & , e_{uv} \notin E_{t-1} \end{cases}. \tag{17}$$

where E_t denotes the set of ISLs at the t-th time slot.

F. Problem Formulation

To improve energy efficiency and throughput and reduce ISLs switching costs, we need to make decisions about inter-plane ISLs planning and power allocation. For the t-th time slot, we define the utility function $\varphi(t)$ as

$$\varphi(t) = \sum_{e_{uv}} \left(\alpha_1 E_{eff,t}^{uv} + \alpha_2 R_{e_{uv},t} \right) - \sum_{e_{uv}} \alpha_3 \theta_{uv,t}. \tag{18}$$

where α_1, α_2 and α_3 are weight factors. Thus, the optimization problem can be formulated as maximizing the utility of the satellite network as follow

$$\begin{aligned} &\max \sum_{t=1}^{N_d} \varphi(t) \\ &s.t. \quad \begin{cases} uv \in Y_t \\ e_{uv} \in E_t \\ \lambda \omega_{u,t}/\delta(t) \le R_{e_{uv},t} \le \omega_{u,t}/\delta(t) \\ P_{e_{uv},t} \le P_{u,t}^{max} - P_0 \\ \alpha_1, \alpha_2, \alpha_3 \end{cases} . \end{aligned} \tag{19}$$

IV. THE PROPOSED DEEPISL ALGORITHM

This section introduces the deep reinforcement learning model and the DeepISL algorithm.

A. DRL Model

In this paper, we propose the DeepISL algorithm to address the problem of joint optimization of inter-plane ISLs planning and power allocation. To prevent the "curse of dimensionality", we train algorithm orbit-by-orbit. Each satellite actively decides to establish ISLs with satellites in the positive side plane, and passively accepts requests to establish ISLs with satellites from the negative side plane. Due to the "Seam" issue, there is no need for satellites in the M-th plane to establish ISLs. So all satellites are independent agents except the satellites in the M-th plane. Then, we define the state space, action space and reward function for each agent.

State Space. For the t-th time slot, we define $D_{i,t}$ is the set of distances between satellite u_i and the satellites in the positive side plane within its LoS, $C_{i,t}$ denotes the battery energy state of satellite u_i, and $\omega_{i,t}$ is the amount of data to be sent by satellite u_i to its positive side plane. Then the state space of agent i is $S_{i,t} = \{D_{i,t}, C_{i,t}, \omega_{i,t}\}$.

Action Space. For agent i, we treat the ISLs planning as discrete actions $v_{i,t}$ and $p_{i,t}$ represents the assigned power, then the decision action can be defined as $a_{i,t} = (v_{i,t}, p_{i,t})$. All possible $a_{i,t}$ form the action space of agent i.

Reward. Since all agents cooperate to maximize the same optimization objective, we define the reward as $\sum_{i=1}^{N_n} = r_{i,t}$, $N_n = N - N_m$ represents the number of satellites except the M-th orbital plane and $r_{i,t}$ represents the contribution of agent i, expressed as

$$r_{i,t} = \kappa_i \left(\alpha_1 E_{eff,t}^{iv_{i,t}} + \alpha_2 R_{e_{iv_{i,t},t}} \right) - \alpha_3 \theta_{iv_{i,t},t}. \tag{20}$$

In addition, we designed a "switching conflicts" resolution mechanism during the training process, aiming to align the reward according to the conflict. The conflict factor $\kappa_i = 1$ for the agents that have no "switching conflicts". For the agents that have "switching conflicts", if the weighted sum of energy efficiency and throughput of the inter-plane ISL established with the target satellite is greater than that all other agents, then $\kappa_i = 0.6$, otherwise $\kappa_i = 0.05$. The reward will finally guide all agents to make rational decisions.

B. The Proposed DeepISL Algorithm

Since the action space is discrete-continuous hybrid action space, DQN and DDPG are suitable for the problems of discrete and continuous actions, respectively, we introduce the parameterized action space, which is rewritten as [10]

$$A_{i,t} = \left\{ \left(v_{i,t}, p_{v_{i,t}} \right) | v_{i,t} \in V_{i,t}, p_{v_{i,t}} \in P_{V_{i,t}} \right\}, \tag{21}$$

where $V_{i,t}$ is the set of satellites in the positive side plane within its LoS, $P_{V_{i,t}}$ denotes the power assigned to the ISLs. Once agent i selects action $a_{i,t} = \left(v_{i,t}, p_{v_{i,t}} \right) \in \left(V_{i,t}, P_{V_{i,t}} \right)$, agent i will establish an inter-plane ISL with the target satellite $v_{i,t}$ and assign transmission power $p_{v_{i,t}}$ to the established ISL.

For each agent i, we denote the action value function as $Q_i(s_i, a_i) = Q_i(s_i, v_i, p_{v_i})$. Assuming that agent i selects action $a_{i,t} = \left(v_{i,t}, p_{v_{i,t}} \right)$ at state $s_{i,t}$, the Bellman equation is

$$\begin{aligned} &Q(s_{i,t}, v_{i,t}, p_{v_{i,t}}) = \\ &\mathbb{E}\left[r_{i,t} + \gamma \max_{v_{i,t} \in V_{i,t}} \sup_{p_{v_{i,t}} \in P_{V_{i,t}}} Q\left(s_{i,t+1}, v_i, p_{v_i} \right) | s_{i,t} = s_i \right], \end{aligned} \tag{22}$$

when the Q function is fixed, for any given v_i,

$$x_{v_i}^{Q}(S) = arg \sup_{p_{v_{i,t}} \in P_{V_{i,t}}} Q\left(s_i, v_i, p_{v_i} \right) \tag{23}$$

is a function of the state s_i. Therefore, we use a deep neural network $Q_i\left(s_{i,t}, v_{i,t}, p_{v_{i,t}}; w_i \right)$ with network weight w_i to approximate $Q\left(s_i, v_i, p_{v_i} \right)$ and a deterministic policy network $\mu_{v_i}\left(s_{i,t}; \theta_i \right)$ with parameter θ_i to approximate $x_{v_i}^{Q}$. In other words, when w_i is fixed, we want to find θ_i which satisfies

$$Q_i\left(s_{i,t}, v_{i,t}, \mu_{v_i}\left(s_{i,t}; \theta_i \right); w_i \right) \approx \sup_{p_{v_{i,t}} \in P_{V_{i,t}}} Q\left(s_i, v_i, p_{v_i}; w_i \right). \tag{24}$$

Similar to DQN, network parameters w_i are updated with the gradients of the least squares loss function. In addition, to find θ_i that maximizes $Q_i\left(s_i, v_i, \mu_{v_i}\left(s_i; \theta_i \right); w_i \right)$ when w_i is fixed, set the loss functions of w_i and θ_i as

$$\ell_t(w_i) = \frac{1}{2}\left[y_{i,t} - Q\left(s_{i,t}, v_i, p_{v_{i,t}}; w_{i,t} \right) \right]^2, \tag{25}$$

$$\ell_t(\theta_i) = -\sum_{v_i \in V_i} Q\left(s_{i,t}, v_i, p_{v_i}\left(s_{i,t}; \theta_i \right); w_{i,t} \right), \tag{26}$$

where $y_{i,t}$ is the target value, expressed as

$$y_{i,t} = r_{i,t} + \gamma \max_{v_i \in V_i} Q\left(s_{i,t+1}, v_i, \mu_{v_i}\left(s_{i,t+1}; \theta_i \right); w_{i,t} \right). \tag{27}$$

Then, w_i and θ_i can be updated by the following equation

$$w_{i,t+1} \leftarrow w_{i,t} - \alpha \nabla_{w_i} \ell_t\left(w_{i,t} \right) \tag{28}$$

$$\theta_{i,t+1} \leftarrow \theta_{i,t} - \beta \nabla_{\theta_i} \ell_t\left(\theta_{i,t} \right). \tag{29}$$

where α and β are the learning rates.

The pseudo-code of training DeepISL algorithm is shown in Algorithm 1, where the initialization and training processes are the same for each agent.

Algorithm 1: Training process of DeepISL

for *agent* $i = 1, N_n$ **do**
  Initialize deterministic strategy network $\mu_{v_i}(\theta_i)$ and value network $Q_i(w_i)$, learning rate α,β and probability ξ . Initialize the experience pool $\mathbf{\Gamma}$
end
for *episode* $= 1$ *to* M' **do**
  for *agent* $i = 1, N_n$ **do**
    Observe the state $s_{i,t}$
    Obtain continuous parameter $p_{v_{i,t}} \leftarrow \mu_{v_i}(\theta_i)$.
    Obtain discrete action by $v_{i,t} = argmax_{v_i \in V_i} Q(s_{i,t}, (V_i, p_{V_i}); w_i)$
    Select action $a_{i,t}$ according to ξ-greedy strategy
    Execute $a_{i,t}$ and observe $r_{i,t}$ and $s_{i,t+1}$
    Store transition $[s_{i,t}, s_{i,t+1}, a_{i,t}, r_{i,t}]$ into $\mathbf{\Gamma}$
  end
end
for *agent* $i = 1, N_n$ **do**
  Randomly draw a batch of $[s_b, s_{b+1}, a_b, r_b]_{b \in \bar{B}}$ from $\mathbf{\Gamma}$
  $y_b = r_b + \gamma \max_{v \in V} Q\left(s_{b+1}, v, \mu_{v_i}(s_{b+1}; \theta_t); w_t\right)$
  Calculate $\ell_t(w_i)$ and $\ell_t(\theta_i)$ according to Equations (25) and (26)
  Update the network parameters w_i and θ_i according to Equations (28) and (29)
end

V. Experiment and Analysis

A. Simulation Setup

In this paper, we use Python 3.9.15 and Pytorch 1.10.0 to build a simulation platform and conduct simulation experiments to verify the feasibility and effectiveness of the proposed algorithm. In our experiments, we set up the neural network of $\mu_{v_i}(\theta_i)$ and $Q_i(w_i)$ to contain two fully connected hidden layers with 64 neurons and 'ReLU' was used as the activation function. For each time slot, the number of packet arrivals for each satellite obeys a Poisson distribution with mean ρ. To reflect the performance of different transmission demands, the value of ρ will be in the range of $350k - 600k$, but to simulate communication scenarios with different transmission demands, the value of ρ will be controlled. The main simulation parameters of the system are given in Table I.

B. Performance Metrics and Comparison Algorithms

Performance Metrics: 1) Mean energy effciency of ISL: The ratio of the sum of the energy efficiency of each inter-plane ISL to the total number of inter-plane ISLs. 2) Mean throughtput of ISL: The ratio of the sum of the throughput of each inter-plane ISL to the total number of inter-plane ISLs. 3) Inter-plane ISLs Switching ratio: The ratio of switched inter-plane ISLs to the total inter-plane ISLs.

Comparison Algorithms: 1) GIEM: A dynamic inter-plane ISL planning algorithm based on greedy algorithm with a fixed allocation of transmission power[11]. 2) DY-DQN: Relaxation of continuous action into discrete actions, dynamic planning of ISLs and dynamic allocation of transmission power. 3) FP-DQN: An algorithm for dynamically planning inter-plane ISL based on DQN with a fixed allocation of transmission power.

TABLE I: PARAMETER SETTINGS FOR EVALUATION

Parameter	Symbol	Value
Number of satellites	N	66
Number of orbital planes	M	6
Altitude of orbital planes	H	780 Km
Inclination of orbital planes	ϵ_m	86.4 deg
Carrier frequency in the Ka-band	f	23.28 GHz
Carrier bandwidth	B	15 MHz
Quality factors	G_{rec}/T_e	8 dB/K
The size of each packet	F_f	$1500B$
The duration of the time slot	$\delta(t)$	300 s
Number of inter-plane transceivers	Q	2
Satisfaction factor	λ	$\{0.85, 0.9, 0.95\}$
Probability of greedy strategy	ξ	0.8
Size of the Mini-batch	$\bar{B}$	1024
Capacity of the experience memory	$Memory$	10000
Lerning rate	α,β	0.0095
Discount factor	γ	0.95
Weight factors	α_1,α_2,α_3	$1, 0.1, 1$

C. Experiment Results and Analysis

1) Convergence graph analysis: Fig. 2 represents the reward convergence diagram for $\rho = 400k, 500k, 600k$ and $\lambda = 0.9$. It can be seen that at the early stage of training, the decision of the satellite isn't optimal because the parameters of the neural network are randomly generated, but after the neural network continuously learns and updates its own parameters, the algorithm basically reaches convergence after 15000 training period.

2) Algorithm Comparison Analysis: Fig. 3 shows that the energy efficiency of ISL decreases and the magnitude of decrease grows as the ρ rises. Compared with the comparison algorithm, DeepISL has the best energy efficiency performance. Since the DY-DQN algorithm slackens the continuous action, while the FP-DQN algorithm adopts a fixed power allocation scheme, both fail to obtain the optimal solution. Meanwhile, GIEM disregards energy efficiency and selects the ISL with the highest throughput, it has the worst energy efficiency performance. Due to the limited energy resources of the satellite, the energy efficiency remains stable after the ρ reaches $550k$. Fig. 4 shows that as the ρ rises, the throughput of the satellite increases, but the magnitude of the increase is diminished. The DeepISL algorithm outperforms DY-DQN and FP-DQN due to the fact that DY-DQN and FP-DQN cannot obtain the optimal solution. However, DeepISL is slightly inferior to GIEM due to the fact that the GIEM algorithm selects the ISL with the highest throughput, while DeepISL still has to consider the energy efficiency and the switching cost of the ISL. In addition, the throughput remains almost constant after the ρ reaches $550k$ due to the limited energy of the satellite. Fig. 5 illustrates the ISLs switching ratio achieved with four algorithms respectively. We can find that all three algorithms, DeepISL, DY-DQN and FP-DQN, perform closely to each other. This is because in all three algorithms, the ISLs switching cost is considered. Since the GIEM algorithm selects the inter-plane ISL with the highest throughput and without considering the switching cost of ISL, the GIEM algorithm has the poorest performance.

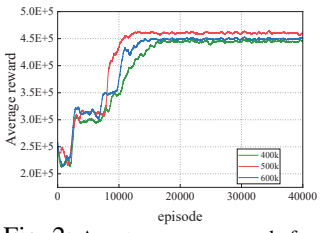

Fig. 2: Agents average reward after 40000 episodes while $\lambda = 0.9$

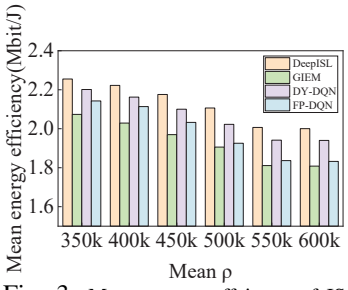

Fig. 3: Mean energy effciency of IS while $\lambda = 0.9$

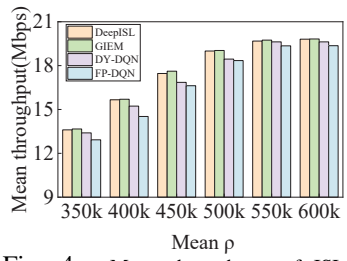

Fig. 4: Mean throughtput of ISL while $\lambda = 0.9$

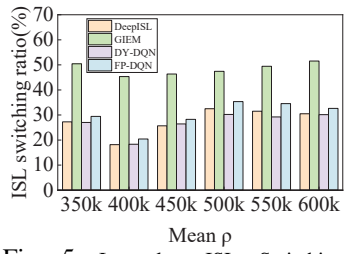

Fig. 5: Inter-plane ISLs Switching ratio while $\lambda = 0.9$

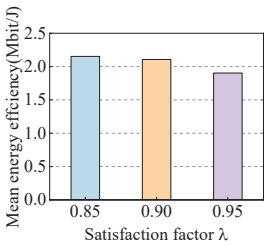

Fig. 6: Mean energy effciency of ISL with ρ = 500k

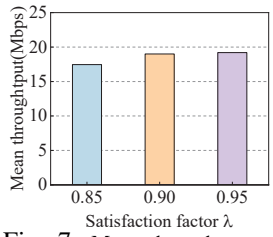

Fig. 7: Mean throughtput of ISL with ρ = 500k

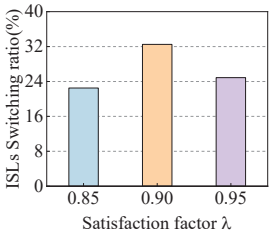

Fig. 8: ISLs Switching ratio with ρ = 500k

3) Sensitivity Analysis: Fig. 6 demonstrates that the mean energy effciency of ISL achieved by DeepISL shows an decrease as the λ grows. As the λ grows, satellites need to allocate higher power, but the percentage increase in throughput is less than the percentage increase in power, so energy efficiency decreases as the λ grows with increasing magnitude. Fig. 7 shows that throughput increases as the λ grows. As λ increases, it will lead to a sharp decrease in energy efficiency, therefore, the magnitude of the throughput increasing decreases in order to obtain a higher reward. Fig. 8 demonstrates that the ISL switching ratio is highest at $\lambda = 0.9$. This is because the performance of energy efficiency is sacrificed for low ISL switching ratio at $\lambda = 0.95$, while the throughput property is sacrificed at $\lambda = 0.85$. The respective performance is well-balanced at $\lambda = 0.9$.

VI. Conclusion

In this paper, we investigated the joint optimization problem of inter-plane ISLs planning and power allocation in the LEO constellation to improve weighted benefits of overall energy efficiency and total throughput of the constellation while reducing the cost of switching ISLs. To achieve this goal, we model the optimization objective as a POMDP, and a DeepISL algorithm is employed to obtain the optimal decision since joint optimization is a discrete-continuous hybrid action space problem. Experimental results show that compared with the comparison algorithms, our proposed DeepISL algorithm can achieve better performance.

References

[1] P. Chini, G. Giambene, and S. Kota, "A survey on mobile satellite systems," *International Journal of Satellite Communications and Networking*, vol. 28, no. 1, pp. 29–57, 2010.

[2] I. Leyva-Mayorga, M. Röper, B. Matthiesen *et al.*, "Inter-plane inter-satellite connectivity in leo constellations: Beam switching vs. beam steering," in *2021 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6. IEEE, 2021.

[3] Z. Yan, G. Gu, K. Zhao *et al.*, "Integer linear programming based topology design for gnsss with inter-satellite links," *IEEE Wireless Communications Letters*, vol. 10, no. 2, pp. 286–290, 2020.

[4] J. Pi, Y. Ran, H. Wang, Y. Zhao, R. Zhao, and J. Luo, "Dynamic planning of inter-plane inter-satellite links in leo satellite networks," in *ICC 2022-IEEE International Conference on Communications*, pp. 3070–3075. IEEE, 2022.

[5] B. Xu, K. Han, Q. Ren *et al.*, "An optimized strategy for inter-satellite links assignments in gnss," *Advances in Space Research*, vol. 71, no. 1, pp. 720–730, 2023.

[6] J. Yanmei, L. Congmin, S. Pengfei, and L. Lu, "Inter-satellite optical analog network coding using modulated retro reflectors," in *Science and Technologies for Smart Cities: 7th EAI International Conference, SmartCity360°, Virtual Event, December 2-4, 2021, Proceedings*, pp. 15–26. Springer, 2022.

[7] Z. Chen, D. Guo, G. Ding, X. Tong, H. Wang, and X. Zhang, "Optimized power control scheme for global throughput of cognitive satellite-terrestrial networks based on non-cooperative game," *IEEE Access*, vol. 7, pp. 81 652–81 663, 2019.

[8] M. Marchese and F. Patrone, "E-cgr: Energy-aware contact graph routing over nanosatellite networks," *IEEE Transactions on Green Communications and Networking*, vol. 4, no. 3, pp. 890–902, 2020.

[9] Y. Yang, M. Xu, D. Wang, and Y. Wang, "Towards energy-efficient routing in satellite networks," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3869–3886, 2016.

[10] J. Xiong, Q. Wang, Z. Yang, P. Sun, L. Han, Y. Zheng, H. Fu, T. Zhang, J. Liu, and H. Liu, "Parametrized deep q-networks learning: Reinforcement learning with discrete-continuous hybrid action space," *arXiv preprint arXiv:1810.06394*, 2018.

[11] 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.