Article

A Graph Reinforcement Learning-Based Handover Strategy for Low Earth Orbit Satellites under Power Grid Scenarios

Haizhi Yu [1], Weidong Gao [1,*] and Kaisa Zhang [2]

1   School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China; yhz_study@bupt.edu.cn
2   School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China; kaisa@bupt.edu.cn
*   Correspondence: gaoweidong@bupt.edu.cn

Abstract: Amidst the escalating need for stable power supplies and high-quality communication services in remote regions globally, due to challenges associated with deploying a conventional power communication infrastructure and its susceptibility to natural disasters, LEO satellite networks present a promising solution for broad geographical coverage and the provision of stable and high-speed communication services in remote regions. Given the necessity for frequent handovers to maintain service continuity, due to the high mobility of LEO satellites, a primary technical challenge confronting LEO satellite networks lies in efficiently managing the handover process between satellites, to guarantee the continuity and quality of communication services, particularly for power services. Thus, there is a critical need to explore satellite handover optimization algorithms. This paper presents a handover optimization scheme that integrates deep reinforcement learning (DRL) and graph neural networks (GNN) to dynamically optimize the satellite handover process and adapt to the time-varying satellite network environment. DRL models can effectively detect changes in the topology of satellite handover graphs across different time periods by leveraging the powerful representational capabilities of GNNs to make optimal handover decisions. Simulation experiments confirm that the handover strategy based on the fusion of message-passing neural network and deep Q-network algorithm (MPNN-DQN) outperforms traditional handover mechanisms and DRL-based strategies in reducing handover frequency, lowering communication latency, and achieving network load balancing. Integrating DRL and GNN into the satellite handover mechanism enhances the communication continuity and reliability of power systems in remote areas, while also offering a new direction for the design and optimization of future power system communication networks. This research contributes to the advancement of sophisticated satellite communication architectures that facilitate high-speed and reliable internet access in remote regions worldwide.

Keywords: LEO satellite; handover strategy; power grid scenario; graph reinforcement learning

1. Introduction

The escalating demand for electricity in our society has led to a significant surge in pressure on the power supply, causing the expansion of both the scale of the power grids and the number of substations of the power grid, thus imposing greater demands on communication quality. The prevailing wired data communication method in contemporary power systems is optical fiber communication [1], with wireless alternatives encompassing power wireless private network and power wireless public network communication. Optical fiber communication is susceptible to natural disasters and human activities, and installing optical cables in remote mountainous regions is challenging [2]. Although wireless public network communication and other wireless data communication methods offer high bandwidth, low latency, and extensive coverage communication network services to the power system, most current power communication technologies depend on infrastructure

such as base stations for information exchange. However, geographical constraints and construction costs make it difficult to expand the ground communication infrastructure in remote mountain and desert regions [3]. Due to the incomplete coverage and unstable communication quality of power wireless private networks and public networks, many transmission lines in certain areas may be located in communication blind spots, leading to insufficient communication services. Furthermore, significant natural disasters can severely damage ground communication facilities, disrupting the transmission of essential information. Consequently, the inability to monitor the real-time conditions of power lines presents significant safety risks [4]. The large-scale power generation, the distribution process of the grid, and the long-distance transmission work process must ensure the safe operation of the power business [5]. This poses a significant challenge to the safe operation of the smart grid, because mainstream power communication methods cannot guarantee the reliability and security of the power business in specific power scenarios. Therefore, traditional mainstream power communication methods cannot ensure stable signal coverage, making it difficult to achieve high-quality communication in remote areas with complex terrain and vulnerability to natural disasters. These methods are inadequate for addressing current challenges faced by power communication enterprises, and fail to meet the industry's service requirements.

With the increasing interconnection of power grids, satellite communications have been extensively adopted in the electric power systems sector, assuming a progressively pivotal role. As an essential supplement to electric power data communication methods, satellite communications effectively overcome the deficiencies of traditional mainstream electric power communications. Satellite communication technology can extend the coverage of traditional mainstream communication of electric power to remote areas such as oceans and deserts due to its features of high reliability, global communication coverage, and independence from harsh environmental conditions [6]. This technology enables the transmission of electricity in remote areas [7], ensuring continuous communication services. In power emergency scenarios, ensuring emergency protection, seamless communication, timely warnings, and post-disaster recovery are critical issues for power communication enterprises. The satellite communication network provides a range of services, such as data, voice, and video, to the power emergency command center, enabling real-time information exchange between disaster-stricken areas and the command center. This offers essential solutions for emergency communication [8]. Additionally, satellite communications can effectively offload terrestrial traffic to alleviate congestion in terrestrial networks, particularly when the traffic exceeds the capacity of the terrestrial link [9]. The integration of satellite communication and ground networks is projected to become a cornerstone of the sixth-generation (6G) wireless communication system, enabling the achievement of ubiquitous Internet of Things (IoT) [10]. In future communication systems, non-ground networks and satellite communication technologies are gaining increasing attention and emerging as a prevailing trend in power communication development.

Based on different orbital altitudes, satellites are classified into Geosynchronous Earth Orbit (GEO), Medium Earth Orbit (MEO), and Low Earth Orbit (LEO) satellites. Among them, LEO satellites have drawn significant attention from the academic community because of their global coverage, low transmission latency, low power links, and robust resistance to destruction. Compared to medium- and high-orbit satellites, LEO satellites operate at lower orbital altitudes, resulting in smaller path losses and shorter propagation delays. This enables them to provide better data transmission rates and higher throughput [11], reduce transmission power, better mitigate signaling attenuation, and contribute to reducing the dimensions of terminal equipment [12], lowering the energy consumption of communication equipment, and enhancing the overall effectiveness of the satellite network. LEO satellites have extensive development prospects. As shown in Figure 1, constructing a wireless transmission network centered on LEO satellite communications, featuring strong anti-destructive properties and low latency, can provide flexible, efficient, and reliable

communication solutions for scenarios such as grid data collection and back transmission, emergency response, and so on.
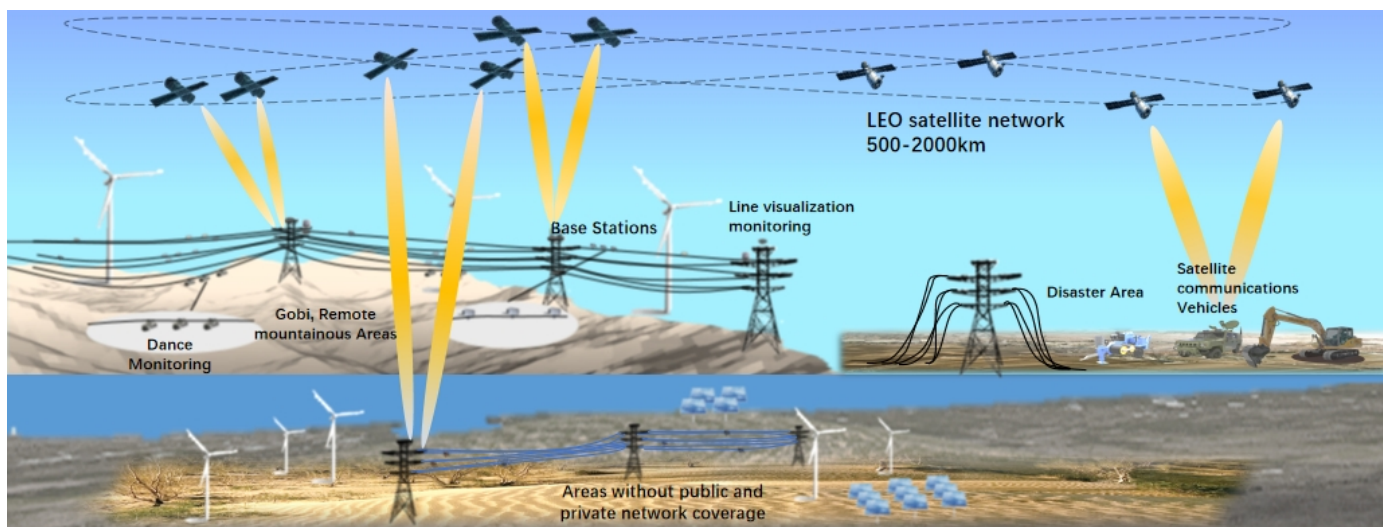

Figure 1. LEO satellite communication-assisted grid scenario figure.

Unlike terrestrial cellular networks, LEO satellites move at high speeds around the Earth in set orbits, each offering a limited service duration to fixed ground terminals [13]. The challenges of exploring inter-satellite optimized handover algorithms in LEO satellite networks primarily arise from the high-speed movement of satellites and the rapid changes in their relative positions, resulting in continuous fluctuations in the signaling environment and network state. Due to the high-speed movement of satellites, users must selectively hand over among candidate satellites to maintain communication continuity, necessitating handover algorithms to minimize communication interruptions and delays. Furthermore, during real-world communication scenarios, the duration of user communication typically exceeds the visibility period of a single satellite, and due to the time-varying characteristics of satellites, it is necessary to analyze the dynamic topology of the satellite network, which varies at different moments [14], leading to differences in the handover diagrams formed by the set of candidate satellites. As the satellite's movement causes the user to move out of the coverage area of the current servicing satellite, ground users must continuously hand over to other visible satellites during the communication process to ensure communication continuity. Because of the high-speed mobility and short coverage period of LEO satellites, handover between grid users and satellites occurs frequently, which may increase communication delays and reduce connection stability. At present, one crucial and urgent research issue is the development of a feasible inter-satellite handover strategy in LEO satellite communication systems. This strategy is vital for minimizing the handover frequency between satellites and power terminals, ensuring stable communication links, enabling seamless services, and maintaining Quality of Service (QoS) for users.

In terrestrial cellular networks, users typically base their handover decisions on received signal strength from terrestrial base stations. However, for users accessing satellite networks, only considering the received signal strength is insufficient. In satellite networks, factors such as the remaining service time of the access satellite and the system's load balance significantly affect the user's QoS. Therefore, the decision-making process for satellite handover is intricate and demanding [15]. When formulating the handover strategy, one can categorize the criteria for handover satellite–ground links based on factors such as the elevation angle and service time of the candidate satellites mentioned earlier, and subsequently propose various handover schemes; distinct choices corresponding to diverse optimization objectives. The grid user terminals continuously select service satellites using the proposed handover strategy to ensure continuous communication. Current satellite handover algorithms are categorized into single-attribute handover and multi-attribute

handover according to decision-making. Single-attribute handover algorithms make satellite handover decisions based on a single attribute, whereas multi-attribute handover algorithms consider multiple attributes simultaneously.

Currently, the research on single-attribute handover algorithms is relatively mature due to their ease of implementation and flexibility in selecting different handover algorithms for varying scenarios. In [16,17], a handover strategy based on the maximum elevation angle is proposed, in which the user terminal prioritizes the satellite with the greatest elevation angle among the candidate satellites. Due to atmospheric and geographic factors affecting the channel, the maximum elevation angle of the satellites may not accurately reflect the actual link quality. In [18–20], a strategy based on the maximum received signal strength was proposed, where user terminals prioritize satellites with the highest received signal strength, but which overlooks other handover factors, such as the candidate satellite load conditions, potentially leading to an increased handover frequency. In [21,22], a handover strategy based on the longest remaining service time was proposed. The user terminal prioritizes satellites with the longest remaining service time among the candidates, effectively reducing the frequency of the handover. However, longer service times may indicate smaller satellite elevation angles, particularly in poorer channel conditions, which hinder effective communication. Zhou et al. [23] proposed a handover strategy based on the maximum load of satellites to mitigate service quality degradation resulting from satellite overload. However, this strategy proves ineffective in scenarios with higher user density. While these single-attribute-based interstellar handover algorithms are simpler, they possess shortcomings and struggle to meet user QoS requirements in actual satellite communication scenarios.

To comprehensively assess the impact of various handover factors on interstellar handover, the current research emphasizes multi-attribute-based interstellar handover strategies. In [24], a handover strategy was proposed based on integrated weighting of the service quality, incorporating three variables: the service time, elevation angle, and number of idle channels. Miao et al. [25] proposed a multi-attribute decision-making handover strategy that comprehensively considers three influencing factors: the received signal strength, remaining service time, and satellite idle channels. Given that the relationship between satellites and users can be modeled as a graph, numerous studies have utilized graph theory. Wu et al. [26] proposed a satellite handover framework based on graph theory. Here, the satellite handover problem can be transformed into finding the shortest path in a weighted directed graph, and the shortest path algorithm is then employed to derive the user's optimal inter-satellite handover strategy. In consideration of the time-varying characteristics of satellite topology caused by satellite motion, Hu et al. [27] proposed a real-time inter-satellite handover prediction framework based on the time-evolving graph (TEG) and a shortest path dynamic updating algorithm. This framework effectively reduces the handover failure rates and avoids unnecessary handovers. Hozayen et al. [28] proposed a graph-based customizable handover framework that considers both the handover time and the target in selecting a sequence, ensuring QoS. By finding shortest paths in a time-based graph, the framework determines the optimal handover sequence and time to meet desired QoS. Li et al. [29] proposed an intelligent handover strategy employing a multi-attribute graph (MAG) and a genetic algorithm to optimize the handover process, consequently reducing communication delay and handover time. Currently, the integration of traditional handover algorithms with machine learning algorithms has become a prominent research area. Shadab M. proposed a Deep Q-Network (DQN) model to optimize the handover process of LEO satellites in non-terrestrial networks (NTNs). This model considers various handover criteria and demonstrates convergence. Wang et al. [30] proposed a satellite handover strategy based on deep reinforcement learning (DRL), reducing satellite handover time by simultaneously considering multiple factors. Zhang et al. [31] proposed a convolutional neural network-based handover strategy. Here, users can make near-optimal handover decisions based on historical signal strength, extracted using a convolutional neural network to reveal potential optimal handover strategies.

Existing research generally falls into two categories: the handover algorithm considers relatively individual factors, which overlooks the impact of other handover factors, resulting in a higher handover failure rate; or focuses solely on selecting the next moment handover satellite, leading to local optima, which may not be the optimal choice for the entire communication duration, resulting in unnecessary handovers [32]. Regarding this, this paper proposes a satellite handover optimization algorithm, termed MPNN-DQN, based on the combination of graph neural networks and reinforcement learning, building upon the satellite handover framework grounded in graph theory. This algorithm comprehensively considers multiple attributes, including the remaining service time, propagation delay, and user data rate, with the objective of maximizing the quality of service. This approach enables the handover algorithm to make optimal decisions throughout the entire communication duration, minimizing the number of switches while satisfying user demand.

The main contributions of the paper can be summarized as follows.

In this paper, we first construct the satellite handover directed graph using graph theory, considering the remaining service time of the satellite and the potential handover between grid users and the satellite. Secondly, to maximize the service quality for power users, we propose a handover strategy that comprehensively weighs the service quality. This strategy assigns edge weights to the satellite handover directed graph based on three factors: the remaining service time, propagation delay, and user data rate. Next, within the graph-based LEO satellite handover framework, the processing capability of graph neural networks can enhance adaptability to changes in the satellite handover directed graph's topology. Learning the satellite handover graph structure and edge weight information enables us to obtain the satellite network state representation. The satellite handover algorithm, employing the DQN algorithm, models and analyzes the handover process throughout the communication duration, with grid users' service quality serving as a reward function to determine the optimal path in the satellite handover directed graph. Finally, experimental simulations are conducted to validate the effectiveness of the proposed algorithm.

2. Materials and Methods

2.1. System Model

Figure 2 shows the system model proposed in this paper, depicting a typical scenario of a network of LEO satellites serving tracking areas (TAs) located in remote regions on earth. The TAs here can be viewed as an entire communication user node with respect to the satellite. With a satellite ephemeris and satellite constellation design for the payload, the TAs are ensured continuous coverage by at least one satellite during switchover. This solution, proposed by 3GPP in Release 17, addresses the key issue of "Mobility management with large satellite coverage areas" in non-geostationary satellite orbit (NGSO) mobility management [33]. The system model comprises three components: the satellite network, the communication relay station, and the power equipment. The satellite network comprises N LEO satellites capable of covering the grid relay station. The communication relay station receives information from the power equipment within its vicinity, and establishes communication links with the LEO satellites. Satellite coverage depends on the minimum elevation angle between the terminal and the satellite. Additionally, the grid relay station's communication duration may exceed the LEO satellites' service time. When the elevation angle of Sat1 is about to decrease below the minimum elevation angle, the relay station must select between Sat2 and Sat3 to maintain ongoing communication. Therefore, this article focuses on designing a rational satellite handover strategy to achieve seamless handover and ensure user QoS within the communication duration of power users.
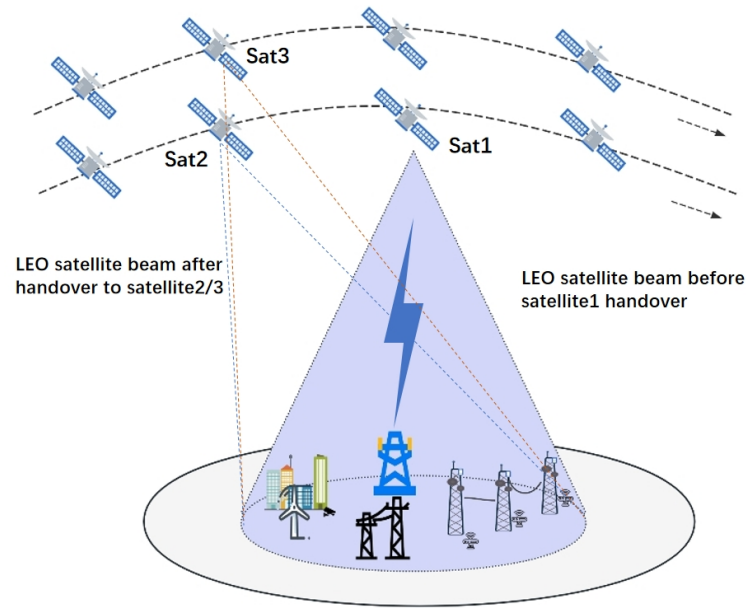
Figure 2. System model.

A single TA can be served by multiple satellites, resulting in multiple candidate satellite options during the handover process. The user's position can be determined using the Global Positioning System (GPS), while the LEO satellite orbit data can be obtained through Two-Line Element (TLE) sets using the simplified general perturbations model (SGP4), as described in [34]. The service period of each satellite to the terminal can be determined based on the satellite ephemeris information and the terminal location information as follows:

$$T = \left\{ \left(t_s^1, t_e^1 \right), \left(t_s^2, t_e^2 \right), \cdots, \left(t_s^i, t_e^i \right), \cdots, (t_s^n, t_e^n) \right\} \tag{1}$$

where (t_s^i, t_e^i) denotes the start service time and the end service time of satellite i to the power terminal. The satellite network comprises numerous satellites, and each LEO satellite can only provide coverage to the ground for a certain period due to its high-speed mobility. Different satellites have varying coverage periods. Therefore, if there are areas of overlap in the coverage periods of different satellites, it indicates that the user terminal can switch between the satellites.

As shown in Figure 3, a region bounded by a dotted line represents overlapping coverage time between satellites. For example, when $t_s^1 < t_s^2 < t_e^1$, it signifies that the service time of LEO Satellite 1 and LEO Satellite 2 to the terminal overlaps, allowing the terminal to switch from LEO Satellite 1 to LEO Satellite 2. It is evident that there is no overlapping coverage time between Satellite 1 and Satellite 3, thus handover cannot occur at the terminal.

Based on the prediction of overlapping service cycles of satellites to terminals in the coming period, a collection of candidate satellites that can provide services during the user's communication hours can be obtained. As shown in Figure 4, the virtual start node (Sat0) is introduced, with each node representing a satellite, and the directed edges indicating handover relationships. The presence of a directed edge between two satellite nodes means that the terminal can switch between these two satellites. The satellite handover process can be represented as finding a path in the satellite handover directed graph, where the handover criterion is translated into weights applied to the directed edges [26]. Thus, in this paper, to address the multi-objective optimization problem, we seek optimal handover paths in the directed graph, aiming to maximize the quality of user service.
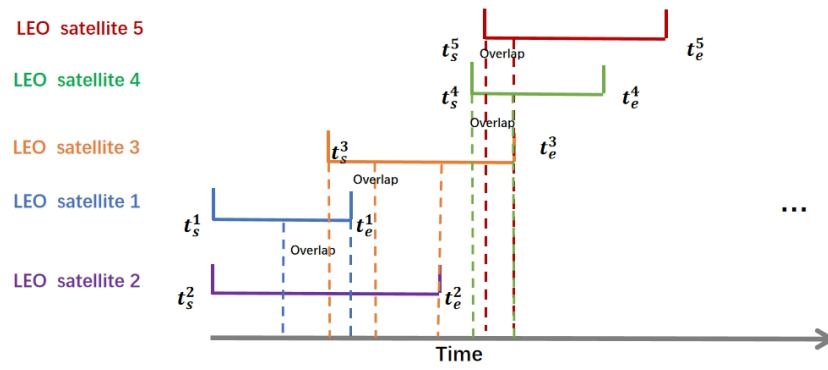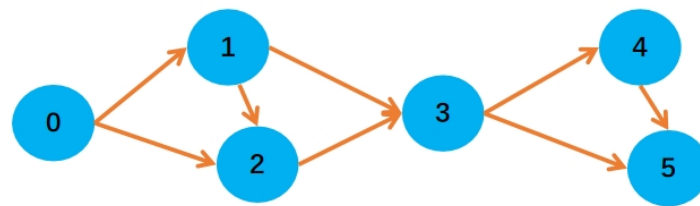
Figure 3. Satellite service cycle overlap map.

Figure 4. Satellite handover directed graph.

2.2. Analysis of Handover Decision Factors

LEO satellite communications play a vital role in grid scenarios in remote areas, ensuring the provision of reliable power business services. When optimizing handover algorithms, careful consideration must be given to the impact of three key factors on power services: the transmission delay, handover frequency, and data rate. Transmission latency directly affects the real-time availability of monitoring and control commands, particularly in remote areas where the power system may encounter higher latency challenges. Therefore, minimizing the transmission latency is crucial to ensuring real-time responses. Optimizing the number of satellite handovers is essential for reducing communication outages and service disruptions. Frequent handovers can result in signal loss and communication instability, thus impacting the continuity of grid operations. The data rate determines the quantity and quality of data that can be transmitted. For grid monitoring and control, a higher data rate is necessary to support complex data transmission requirements. Thus, optimization of the handover algorithm must consider these three factors to ensure that the LEO satellite communication effectively meets the real-time, continuity, and high-efficiency requirements of power grid operations.

2.2.1. Remaining Service Time

When designing the satellite handover algorithm for the power system, it is crucial to fully consider the service time of the satellite. The service time provided by the satellite to grid users is related to the number of service switches. A longer service time of the candidate satellite results in fewer switches for power grid user terminals engaged in communication. The service time of the satellite is closely related to the reliability and stability of the power system. It is crucial for real-time monitoring, control, and other key functions. Frequent handovers may result in long service interruptions, causing loss or delay of information, thereby affecting the system's real-time monitoring and control capability. Therefore, the service time of the satellite is selected as a handover factor. When the grid user is ready to handover to another satellite, the user sends position information to the ground station, which then returns the matrix T_{serv}:

$$T_{serv} = \begin{bmatrix} t_1^s t_2^s t_i^s \dots t_N^s \\ t_1^e t_2^e t_i^e \dots t_N^e \end{bmatrix} \tag{2}$$

T_{serv} is a 2 × N dimensional matrix, where t_i^s in the first row denotes the start service time of satellite j to TAs, and t_i^e in the second row denotes the end service time of satellite j to TAs (1 ≤ i ≤ N; N is the number of LEO satellites in the satellite network). Therefore, the remaining service time of satellite i to user at moment t can be expressed as

$$T_i(t) = t_i^e - t \tag{3}$$

2.2.2. Transmission Delay

When designing the satellite handover algorithm for the power system, it is essential to fully consider the transmission delay of the star–ground link. This is because the transmission delay of the star–ground link directly impacts the real-time and accuracy of data transmission. In smart grid applications, a significant amount of power data requires timely collection and analysis to achieve real-time monitoring and management of the power grid operation status. An increase in transmission delay may result in delayed information transmission, thereby affecting the real-time monitoring and control capability of the system. Especially in the event of system state changes, faults, or emergencies, a short transmission delay is crucial to ensure the timely implementation of necessary measures. Therefore, the design of satellite handover algorithms in power scenarios should consider the transmission delay of the star–ground link to ensure that the power system can promptly and reliably meet the requirements of real-time communication and control. d is the propagation distance between the satellite and the power terminal, calculated as:

$$d = \sqrt{h^2 + (x - o_x)^2 + (y - o_y)^2} \tag{4}$$

where (o_x, o_y) is the position directly below the satellite, (x, y) is the coordinate position of the user node, and h represents the vertical height of the satellite from the ground. The transmission delay expression can be defined as:

$$PD = \frac{d}{c_{light}} \tag{5}$$

2.2.3. Data Rate

In the power system, considering the grid user data rate is essential because the power system requires responsiveness to real-time monitoring and control needs, and the user data rate directly impacts the efficiency of information transmission. Therefore, the consideration of power data rate in the design of satellite handover algorithms is an important factor in ensuring the efficiency of power system communication and data transmission.

Due to their distance from the ground, satellites experience significant effects from the surrounding environment during signal transmission, leading to various losses and fading phenomena, resulting in a weakened signal strength. The channel model mainly comprises free-space propagation loss, atmospheric loss, shadow fading, and other factors, based on the causes of loss and fading. Among these factors, free-space propagation loss is considered the primary source of loss during wireless signal transmission. Combined with the Star–Earth link distance given in Section 2.2.2, the free-space propagation loss L_F can be defined as [35]:

$$L_F = 20lg(\frac{4\pi df}{\lambda}) \tag{6}$$

where c is the speed of light; f is the carrier frequency in GHz; L_a denotes the signal loss generated by the atmosphere, rainfall, etc.; and L_o is the other losses and fading. Then, the total loss L_p of the signal during transmission can be expressed as:

$$L_P = L_F + L_a + L_o \tag{7}$$

The received power is defined as:

$$P_r = P_t - L_p + G_r + G_t \tag{8}$$

where P_t is the transmit power and G_t and G_r are the antenna gains of the transmitter and receiver, respectively. The user data rate is given by Shannon's capacity theorem:

$$R = B log(1 + \frac{P_r}{P_N}) \tag{9}$$

where B is the channel bandwidth and P_N is the noise power.

2.3. Problem Description

In an LEO satellite communication system, the problem of selecting the handover satellite is essentially a multi-criteria optimization problem. We address the optimization problem within a specific communication duration T, which is divided into M different time periods, denoted as $[(t_0, t_1), (t_1, t_2), \ldots, (t_n, t_{n+1}), (t_{M-1}, t_M)]$. The satellite network topology remains fixed in each time slot. The UE measures information from the LEO satellite network to obtain the directed graph of visible satellites capable of covering the user in that time slot. Subsequently, it selects the handover path that satisfies the user's QoS in each time slot. The aim of this paper is to minimize the number of long-term switches over the entire communication duration T, select satellites with higher data rates and lower transmission delays, and maintain the load balance of the satellites. This constitutes a multi-objective optimization problem involving three handover factors. In this paper, we employ a normalized function of the remaining service time, data rate, and transmission delay of the satellite as the objective function for these three attributes.

Assuming a switchover at the service cutoff moment of the current serving satellite, the candidate satellite can provide communication to the user for the length of $t_j^e - t_i^e$, so the normalization function of the remaining service time is defined as:

$$N(t_i) = \frac{t_j^e - t_i^e}{t_{max}}, t_e^j - t_e^i \leq t_{max} \tag{10}$$

where t_{max} is the maximum service time of the candidate satellite. Both the transmission delay and data rate are obtained using the Min–Max normalization method to obtain the objective function, defined as, respectively:

$$N(d_i) = \frac{d_i - \min(d)}{\max(d) - \min(d)} \tag{11}$$

$$N(r_i) = \frac{r_i - \min(r)}{\max(r) - \min(r)} \tag{12}$$

wherein d_i denotes the current transmission delay of the candidate satellite, and $\min(d)$ denotes the minimum transmission delay of the candidate satellite. r_i denotes the current data rate of the candidate satellite, and $\max(r)$ denotes the maximum data rate of the candidate satellite.

In path screening, each objective function is assigned corresponding weights. These objective functions are then multiplied by their respective weights and summed to obtain a new objective function, which is subsequently used to address the multi-objective optimization problem as a single-objective task. To achieve a balance between the three objectives, we formulate the multi-objective optimization problem as follows:

$$Z = max\{\sum_{i=1}^{n}(w_1 N(d_i) + w_2 N(r_i) + w_3 N(t_i))\} \tag{13}$$

where w_1, w_2, and w_3 denote the weight values corresponding to different attributes, respectively, which are subsequently calculated based on the Analytic Hierarchy Process (AHP), and n denotes the number of candidate satellites for handover paths in the satellite handover directed graph.

3. The Proposed DRL + GNN-Based Handover Scheme

3.1. GNN Architecture

A graph neural network (GNN), introduced in [36], is a deep learning model designed for processing graph data. The GNN model comprises multiple graph convolutional layers, with each layer updating the representation of a node by aggregating both the node and its neighborhood data. Through this process of information aggregation, the model can capture features of both local nodes and global graph structures. This capability allows GNN to excel in processing graph data. GNNs leverage their "black-box" nature to learn the relationships between nodes and edges, iterating over the states of nodes and edges in the process. A message passing neural network (MPNN) forms the foundational framework of graph neural networks [37], employing an iterative message passing algorithm to propagate information between nodes and edges of a graph.

In the algorithm proposed in this paper, DQN agents do not directly interact with the LEO satellite network environment; instead, they interact with it through a graph neural network. Specifically, the graph neural network learns the representation of the LEO satellite handover graph state, which serves as the interaction environment for the deep reinforcement learning agents. Given that the edge features of the satellite handover graph define the optimization problem for satellite handover selection, we employ an enhanced MPNN to conduct the message passing process between all edges in the graph. The structure of MPNN message passing is illustrated in Figure 5.
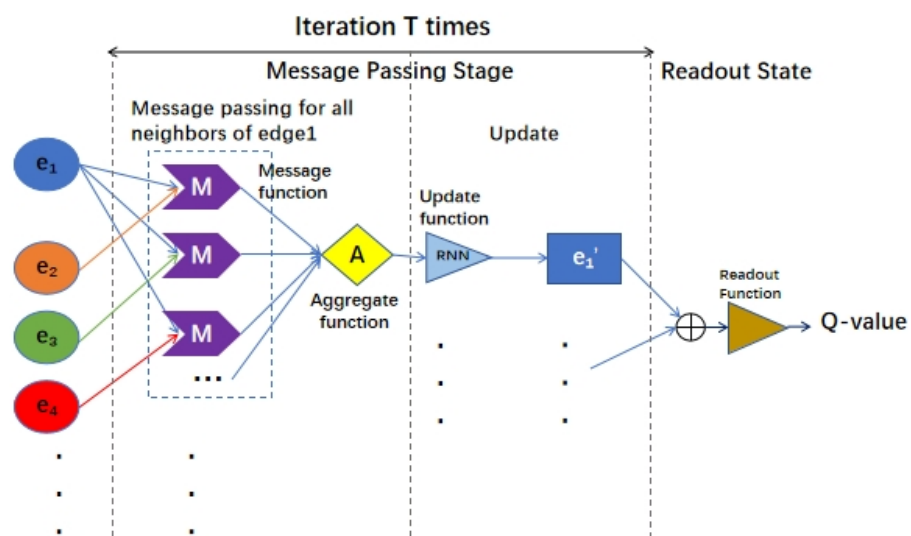

Figure 5. MPNN passing architecture.

MPNN is divided into message passing phase and readout phase. The message passing phase can be defined by message function, aggregation function and update function. The message function $M()$ is used to generate the message of the neighboring edge to the central edge, and its input is the state of the central edge and the neighboring edges; the aggregation function $A()$ is used to aggregate all the messages generated by the message function, and obtain the post-aggregation message as M. Its inputs are the messages generated by all the neighboring edges to the central edge; the update function $U()$ is used to update the state representation of the edges, and its inputs are the state of the edge's last iteration and the the aggregated message.

In the message passing phase, the states of all the edges in the graph are first generated by the message function $M()$, then all the messages are aggregated by the aggregation function A, and then the new state of each node is calculated using the update function $U()$. The functions $m()$ and $U()$ can be learned by neural network. The specific message passing process is as follows: (1) combine the features of each edge with the features of the neighboring edges using full connectivity to generate the messages from the neighboring edges to this edge; (2) aggregate the messages generated by each edge with all its neighboring edges using element-by-element summation to obtain the aggregated message m; (3) use a recurrent neural network (RNN) to update the state h of each edge; the input of the RNN model is the state of the edge's previous iteration and the aggregated message m.

The message-passing process can be expressed as

$$m^{K+1}(c) = A_{e \in N(c)}[M(h_e^k, h_c^k)] \tag{14}$$

$$h_c^{k+1} = U[h_e^k, M^{K+1}(c)] \tag{15}$$

where h_e^k and h_c^k denote the state of edge e and c after the kth iteration, respectively, and $N(c)$ denotes that the set of all neighboring edges of c; $m^{K+1}(c)$ is the aggregation of messages generated by all neighboring edges to the central edge c in the $k + 1$st iteration.

Readout phase: the q-estimate corresponding to the current state and action are output through the readout function $R()$. Finally, the DRL agnet selects the action with the highest q-value by comparing the q-estimate of a set of actions in the current state. The input to the readout function is the state representation of all edges, and the readout process can be expressed as follows:

$$O = R(\sum_{c \in E} h_c) \tag{16}$$

where $R()$ denotes the readout function and o is the state representation of the output, which uses a fully connected deep neural network (DNN), and E denotes the set of all edges. The flow of the algorithm is shown in Algorithm 1. The inputs to the algorithm are the characteristic feature information of the LEO satellite handover graph topology and edges, and the feature information includes the remaining service time, data rate, and transmission delay. The algorithm performs T message passing steps, and then outputs the q-estimate.

Algorithm 1 LEO satellite handover graph state representation learning

---

Input: Topology of LEO satellite handover map, Directed edge state information
Initialization: Randomly initialize graph neural network parameters w
for i in range(K):
    For each edge e:
      $M^{K+1}(s) = A_{u \in N(s)}[m(h_u^k, h_s^k)]$
      $h_s^{k+1} = U[h_s^k, M^{K+1}(s)]$
Output: State representation of all edges after aggregation update
A q-estimate is obtained from $R()$ function

3.2. DQN Framework

Deep reinforcement learning (DRL) is an approach that integrates reinforcement learning with deep learning techniques. Through learning via interactions with the environment, AI can adaptively make decisions to maximize an objective function in an optimization problem. The DRL framework defines state spaces and action spaces, represented by sets of states (S) and actions (A), respectively, along with reward functions. DRL agents then seek the optimal policy by iteratively exploring the state and action spaces.

The objective of Q-learning [38] is to facilitate the learning of a policy π :S→A by an intelligent agent. The algorithm generates a Q-table comprising all feasible combinations of states and actions, and throughout the training process, the intelligent agent updates the Q-values in the table according to the rewards obtained from action selection. The formula

for updating Q values during the training process in Q-learning is obtained from the Bellman equation:

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha(R(s_t, a_t) + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)) \tag{17}$$

where Q(s_t, a_t) is the Q-value function at time t, α For the learning rate, $R(s_t, a_t)$ is the reward value for taking action a_t in state s_t.

Deep Q-Network (DQN) [39] is an enhanced algorithm derived from Q-learning, integrating a deep neural network (DNN) to approximate the q-value function. The Q-table in q-learning stores the associations between q-values, states, and actions. As the Q-table grows in size, q-learning encounters challenges in learning optimal policies within high-dimensional state and action spaces. Consequently, DQN employs Q-networks in lieu of Q-tables and depends on the generalization capability of DNNs to estimate the Q-values for states and actions that are not pre-defined. The implementation process of the DQN algorithm is shown in Figure 6.
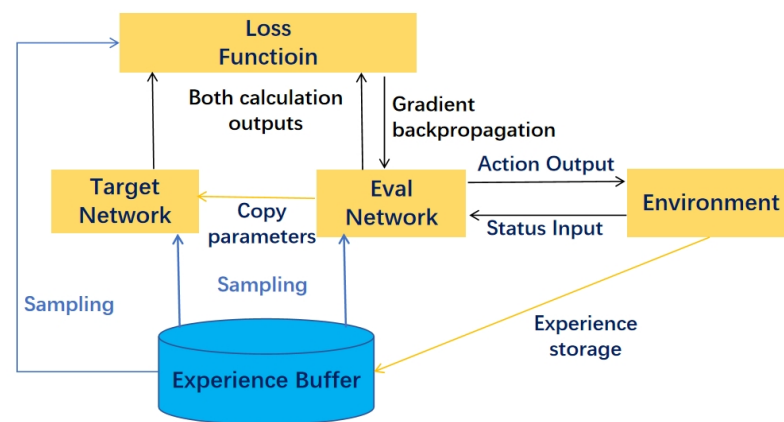

Figure 6. Process of DQN algorithm.

DQN comprises two Q networks with identical structures: one is the Q estimation network employed for action selection, denoted as $Q(s, a; \theta)$; the other is the Q target network utilized for training, denoted as $Q(s, a; \theta^-)$. θ and θ^- represent the parameters of the Q estimation network and the Q-target network, respectively. Furthermore, throughout DQN training, the intelligent agents utilize an experience replay buffer D to retain past experiences (i.e., tuples containing s_t, a_t, r_t, s_{t+1}). For each training process, a certain number of experiences are randomly selected from D as a small batch, and the loss function for training DQN is defined as:

$$L(\theta) = E_{<s_t a_t r_t s_{t+1}> \in D}[(r_t + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}; \theta^-) - Q(s_t, a_t; \theta))^2] \tag{18}$$

where r_t denotes the immediate reward value and γ denotes the discount factor.

Combined with the research scenario of this paper, there are three key elements in our proposed DQN algorithm, which are state space, action space, and reward function, and the specific descriptions of the three elements are shown below.

- State space

The system state space comprises information regarding the environment perceived by the intelligent agent, encompassing the changes in the environment following the execution of the agent's actions. From a reinforcement learning perspective, the system state space serves as the foundation for intelligent agents to make decisions and assess their long-term rewards. Consequently, the design of the system state space directly influences the final performance and convergence speed of the DQN algorithm.

The entire communication duration is partitioned into k equal time slots, and the satellite network state at different time slots is denoted as $S(t)$, aiming to define the state

space as finite. The state is defined as the combination of the inherent state characteristics of the satellite handover graph with zero-padding vectors, which autonomously determines the size of the state space N. These characteristics include propagation delay, remaining service time, and user data rate. The specific description is as follows:

$$S(t) = \{\{d_1, r_1, t_1, padding\}, \{d_2, r_2, t_2, padding\}, \ldots, \{d_k, r_k, t_k, padding\}\} \quad (19)$$

The size of the state space is configured to exceed the number of features, with vectors padded with zeros, enabling each edge to retain its own information along with aggregated information from all neighboring edges. Nevertheless, it should not be excessively large to avoid the state space becoming bigger, which could lead to the GNN model becoming overly large and thus prone to overfitting [40]. The feature description is shown in Table 1:

Table 1. Description of characteristics.

Symbol	Description
d_k	Transmission delay
r_k	Data Rate
t_k	remaining service time
$x_4 - x_n$	Zero padding

- Action space

During the path selection for satellite handover graphs, we calculate the number of paths and the set of nodes for each path (i.e., all nodes traversed by each path) from the virtual start node, on a hop-by-hop basis. We take advantage of the generalization capability of GNNs to introduce actions to intelligent agents in the form of graphs. Successfully trained GNNs can comprehend actions on various graphical structures (i.e., different network states and topologies).

$$a(t) \in \{1, 2, \ldots, k\} \quad (20)$$

where k denotes the number of paths from the virtual start node.

- Reward function

The reward function of the system should be related to the optimization objective, based on the handover decision problem studied in this paper, the reward function is defined with respect to the quality of the user's service, thus the reward function is defined as:

$$R(s, a) = w_1 N(r_i) - w_2 N(d_i) + w_3 N(t_i) \quad (21)$$

where $N(d_i)$ is a normalized function of the transmission delay on the candidate path; $N(r_i)$ is a normalized function of the user data rate on the candidate path; and $N(t_i)$ is a normalized function of the remaining service time on the candidate path.

We used an Analytic Hierarchy Process (AHP) analysis to calculate the weights of different handover factors in the reward function with the following steps.

Step 1: Constructing the judgment matrix. We employ the consistent matrix method to construct a judgment matrix to compare the handover factors with each other. Among all handover factors, transmission delay holds the highest significance, as real-time monitoring of services by the power system is crucial for timely reaction to faults. Service time ranks as the second-most critical handover factor, as it dictates the handover frequency to mitigate power data loss caused by frequent handover, subsequently impacting the system's real-time monitoring and other functionalities. User data rate stands as the third-most significant handover factor, as it influences the efficient transmission of power data information, albeit with a lesser impact on real-time services. Based on the importance to different handover factors, a judgment matrix A is constructed, where each row from top to bottom and each

column from left to right represents a different handover factor, which are transmission delay, remaining service time, and data rate in that order.

$$A = \begin{bmatrix} 1 & 1/3 & 1/5 \\ 3 & 1 & 1/2 \\ 5 & 2 & 1 \end{bmatrix} \tag{22}$$

Step 2: Calculating the weights of the handover factors. We obtain the eigenvector corresponding to the largest eigenroot λ_{max} in the judgment matrix by $AW = \lambda_{max}W$ and normalize this eigenvector to obtain the weight vector W:

$$W = [0.1095, 0.3090, 0.5815]^{\mathrm{T}} \tag{23}$$

After analyzing and calculating the weights of each handover factor, the reward function is further expressed as:

$$R(s,a) = \{\sum_{i=1}^{n}(0.5815N(r_i) - 0.1095N(d_i) + 0.3090N(t_i))\} \tag{24}$$

3.3. MPNN-DQN Based Handover Scheme

This paper proposes a DQN + MPNN architecture for inter-satellite handover, aiming to address the satellite handover decision problem encountered by grid users. This architecture represents handover users as intelligent agents and reformulates the satellite handover decision problem into a multi-criteria optimization problem. The corresponding set of decision strategies is derived by maximizing the long-term cumulative expected reward function.

Throughout the learning iterations of the MPNN-DQN algorithm, the agents receive graph-structured state observations from the environment at various time steps. The GNN constructs a graph representation in which the directed edges of the graph represent graph entities. The topology of the GNN graph is invariant, since the state space size is predefined, ensuring it maintains a consistent length across different topology sizes. Using the graph structure, an iterative message passing algorithm operates between the states of the edges. The output of this algorithm is aggregated into a global state that encodes the graph topology information. Subsequently, the output is processed by the DNN to estimate the q-value. Through comparing the q-value estimates of a set of actions in the current state, the agent can select the most efficient action. The overall framework of the satellite handover decision algorithm designed in this article is shown in Figure 7.
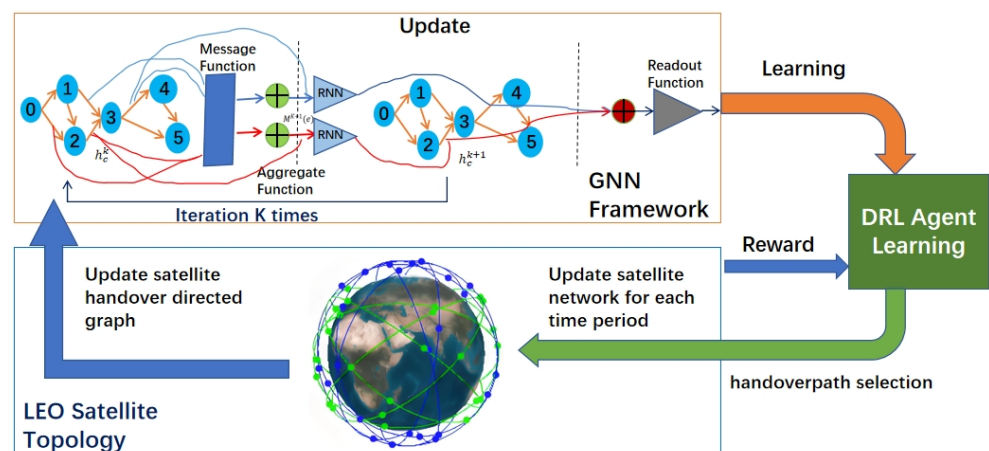

Figure 7. The overall framework of satellite handover decision algorithm.

The specific implementation flow of the handover algorithm framework is illustrated in Algorithm 2. Initially, we initialize the environment and the DRL intelligence, set the reward to zero, create an experience replay buffer, etc., and set the end-of-training flag, DONE, to false. Subsequently, we enter a while loop, where the DRL intelligence learns by interacting with the environment until the most suitable handover paths are discovered for each time slot within the communication duration. For each time slot in the satellite handover graph topology, we calculate the q-value of each state–action pair after obtaining all possible handover paths from the virtual start node. Subsequently, we transition to the new satellite network topology state s′ for the next time slot, obtain the reward r and the completion marker DONE value according to the reward function, and store the state, action, etc., in the experience replay buffer M. These stored memories are subsequently utilized to train the GNN model every 15 iterations.

Algorithm 2 Satellite handover decision algorithm based on MPNN-DQN

---

1. Initialization

(1) Low Earth orbit satellite network environment: Satellite handover graph topology, transmission delay, data rate, remaining service time, etc.

(2) GNN Parameters: Model Parameters w

(3) DQN intelligent agent: Set the reward at 0, the estimated network parameter for Q is θ, and the target network parameter θ^- for Q is equal to θ; Empty the experience replay pool; Set the end of training to indicate that DONE is false;

1. Training

while not DONE:

    Obtain the initial state of edges in the LEO satellite handover graph

    Calculate all candidate paths p and detailed path information

    For each edge p:

        Calculate the network state corresponding to the selected path p

    Choose action a according to the greedy strategy

    if np.random.rand() > self.epsilon:

        Based on Algorithm 1 (refer to Equations (14)–(16)), obtain the estimated value of q

        for the current state and action, and select the action with the maximum estimated value

of q as the path to be selected

    else:

        Randomly select a candidate path as action a

    Generate new status, instantly reward r (refer to Equation (24)) and mark Done

    Store the information of state transition to the experience replay pool

    reward = reward + r

    steps += 1

if steps % M == 0 and steps%15 == 0:

    Sampling batch data from experience replay pool

    Calculate loss and gradient (refer to Equation (18))

    Update network parameters w and θ

if episode%50 == 0:

    Update network parameters θ^-

In the MPNN-DQN-based satellite handover algorithm, DQN agents interact with the LEO satellite environment during the training phase, using the robust characterization and generalization capabilities of GNNs to learn the optimal strategy for satellite handover paths; and during the testing phase, they utilize the trained GNNs to determine the optimal handover paths within the specified communication duration.

4. Results

4.1. Experimental Setup

The hardware configuration used for the experiment is as follows: an Intel Core i7 processor with four cores at a clock speed of 2.60 GHz, an NVIDIA GeForce MX350 GPU (NVIDIA, Santa Clara, CA, USA) with 2 GB of memory, 16 GB of DDR4 RAM, and 512 GB of SSD storage (SSSTC, Milpitas, CA, USA). The operating system used was Windows 11, and the experiments utilized Python 3.10 and TensorFlow 2.10.0. This setup adequately met the computational demands of training deep reinforcement learning models.

We detail the experimental setup, including the set of environment and DQN model parameters, along with their chosen values for the training. Our experimental environment simulates a single user covered by various satellites at different timestamps. A satellite is considered to cover the User Equipment (UE) when its elevation angle relative to the UE is at least 10°. Table 2 summarizes the LEO satellite mobile environment parameters. Table 3 summarizes the system model parameters.

Table 2. Summary of LEO satellite mobility environment parameters.

Parameter	Value
UE position (Latitude, Longitude, Altitude)	(−62°, 50°, 0 m)
Simulation time (minutes)	30
Number of total time slots	60
Number of total satellites providing coverage	15
Satellite altitude (km)	400–600
Minimum coverage elevation angle	10°
Simulation starting time	1 May 2023 09:30 a.m. (UTC)

Table 3. Summary of DQN framework parameters.

Parameter	Value
Discount factor	0.95
Learning rate	0.001
Experience replay pool size	4000
Initial exploration rate	1
Termination of exploration rate	0.005
Training batch size	32
Q-target network parameter update step size (episodes)	50
DQN iterations	1600
Loss Function	Mean-Squared Error (MSE)
Optimizer	Stochastic Gradient Descent (SGD)

4.2. Learning Convergence Analysis

Within the reinforcement learning framework, convergence of the learning process occurs when the agent's reward function stabilizes and no longer undergoes significant changes over time, signifying that the agent has acquired the optimal policy.

As shown in Figure 8, this graph illustrates the learning convergence of the MPNN-DQN-based interstellar handover algorithm proposed in this paper. It is evident that the average reward attained by the agent gradually increases as training progresses, and it tends to stabilize after approximately 1700 episodes. This stabilization indicates that the learning process has converged to the optimal handover scheme.
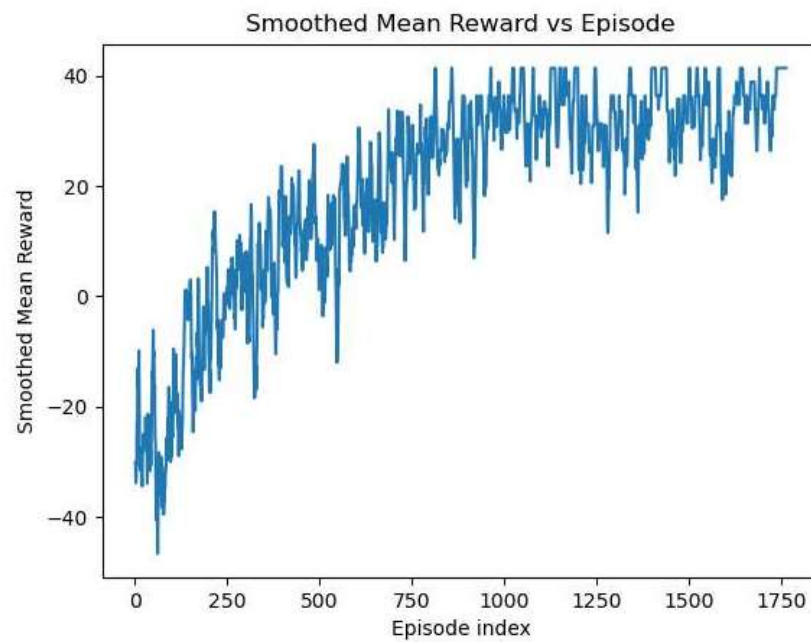
Figure 8. Convergence of mean rewards with episodes.

4.3. Comparison of Algorithm Performance

4.3.1. Handover Frequency Comparison

In power scenarios in remote areas or disaster-stricken areas, the grid user terminals must communicate with satellites, and frequent satellite handovers can cause a series of serious problems. Increased communication interruptions and delays: communication links must be re-established whenever a satellite handover occurs, leading to interruptions and delays that affect real-time performance and stability. Increased energy consumption: satellite handover consumes power from grid terminal equipment, particularly in remote areas with limited energy supplies. Increased system complexity: frequent satellite handover necessitates more complex algorithms and protocols to manage the connection-handover process. Increased data loss and errors: each handover carries a risk of data loss or transmission errors. In power scenarios in disaster areas, transmitted data, crucial for monitoring and control, when lost or erroneous, can lead to misjudgments or operational failures.

We compared the handover algorithm proposed in this study with the handover frequency of the DRL strategy, the maximum elevation angle strategy, and the maximum remaining service time strategy. As shown in Figure 9, the diamonds represent different experimental values, and the curves indicate normal distribution, which serves to help analyze and interpret the distribution characteristics of the data; the higher the peak and the steeper the curve, the higher the degree of concentration of handover frequency near the mean; the flatter the curve, the greater the degree of dispersion of handover frequency. In addition, the curve shows a long tail or skew, which may indicate the existence of some abnormal data points. Thus, we are able to obtain the average handover frequency of multiple experiments with different handover strategies as well as clearer experimental results. It can be seen that the handover number resulting from the maximum elevation angle based handover strategy is the largest. This occurs because the strategy only considers the impact of the elevation angle on handover, ignoring other factors such as service time, which leads to frequent inter-satellite handover. Our proposed handover strategy exhibits a similar handover frequency as the strategy based on maximum remaining service time, while the frequency in the DRL-based strategy is slightly higher than our proposal. This primarily results from the previously mentioned algorithm selecting only the next-hop satellites for interplanetary handover, which only represents the current optimum, leading to unnecessary handover. Meanwhile, our algorithm focuses on the global number of

handover times in the whole communication duration to realize the optimal selection while satisfying the user's QoS.
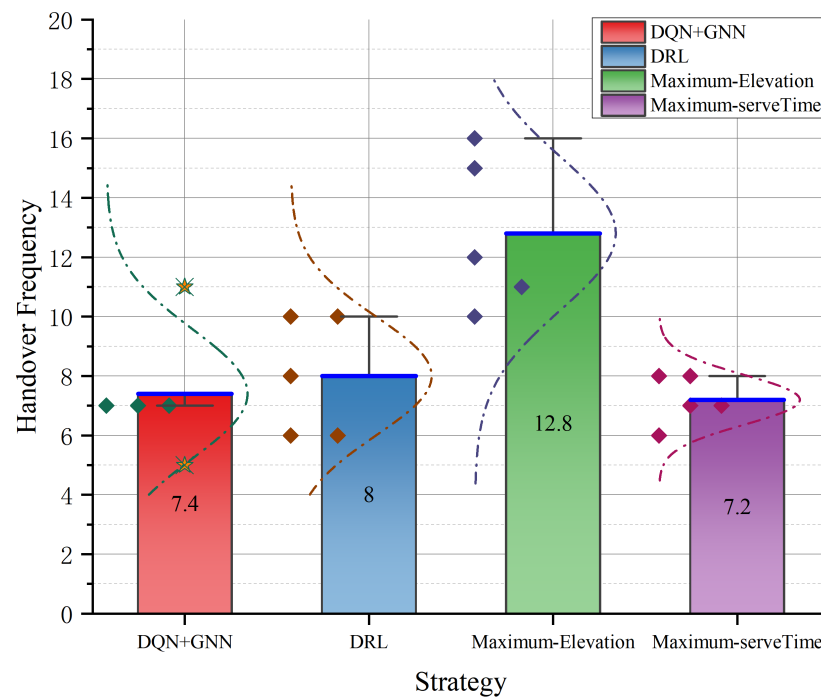

Figure 9. Comparison chart of satellite handover frequency.

4.3.2. Data Rate Comparison

In remote area grid scenarios utilizing satellite communications for data transmission, inter-satellite handovers must prioritize the user data rate. This emphasis is crucial, as the data rate directly impacts the user's quality of service (QoS), which is essential for ensuring communication reliability, power grid stability, and the accurate, timely transmission of monitoring and control commands. Furthermore, higher data rates minimize the impact of handover, prevent data rate fluctuations from causing disruptions to grid operations or poor decision making, and avoid wasted resources.

We compared the data rates of the GNN+DQN-based handover strategy proposed in this study with other strategies. As shown in Figure 10, the pink squares represent the average value of multiple experiments, the red stars represent abnormal experimental values. The maximum service time-based strategy exhibits the lowest user data rate. This is due to the strategy prioritizing satellites with the longest service times to minimize handover, resulting in the service time of satellites to be too long, when the channel conditions will be poor and unfavorable for communication. Our proposed strategy's data rate is close to that of the maximum elevation angle-based strategy and exceeds the data rate of the DQN-based strategy by approximately 0.7 Mbps. This is because we start from the perspective of global optimization and focus on the impact of user data rate, so that we can better ensure the stable operation of the power grid in remote areas and efficient data communication. Additionally, the maximum elevation angle-based strategy achieves the highest data rate because it consistently selects the satellite closest to the ground power terminals, and the main loss in the signaling process of our design is the free-space propagation loss. More realistic channel conditions will be designed later.
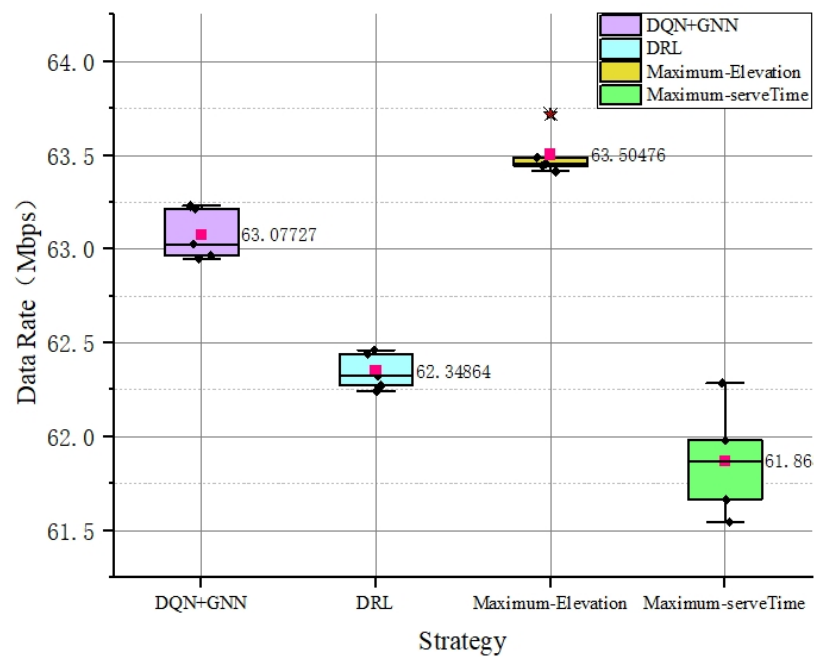
Figure 10. Comparison box plot of satellite user data rate.

4.3.3. Handover Delay Comparison

Following the handover management in terrestrial networks, the whole handover management scheme will be divided into three phases: handover measurement, handover judgment, and handover execution [41].

(1) Handover measurement phase: the star-carrying base station instructs the grid terminal to upload measurement information periodically.

(2) Handover judgment phase: the star-carrying base station compares the handover measurement information uploaded by the grid terminal with the preset handover threshold value, judges whether the handover conditions are met, and selects candidate satellites according to the deployed handover algorithm.

(3) Handover execution stage: the star-carrying base station initiates the handover request, and under the control of the ground core network, the grid terminal, the source satellite and the target satellite interact with each other in terms of control signaling and data.

We compared the total handover process delay and transmission delay of three algorithms. As shown in Figure 11, the average results from various experiments indicate that our GNN+DQN-based handover strategy has a lower transmission delay than the DRL-based strategy and, relative to other handover strategies, it minimizes the total delay of the handover process over the entire communication duration. This is attributed to the three phases of the handover process, including the delay caused by the handover execution (affected by the handover frequency) as well as the data/signaling transmission delay of the star–ground link. Our scheme comprehensively considers two key factors—transmission delay and handover frequency—and optimizes both across the entire communication duration to reduce the number of frequent handover times as much as possible and, at the same time, reduces the transmission delay. The maximum elevation angle-based strategy consistently selects the satellite with the highest elevation angle, thereby minimizing transmission delay due to the shortest star–ground transmission distance. However, it overlooks other factors, leading to frequent handover and increased total delay.
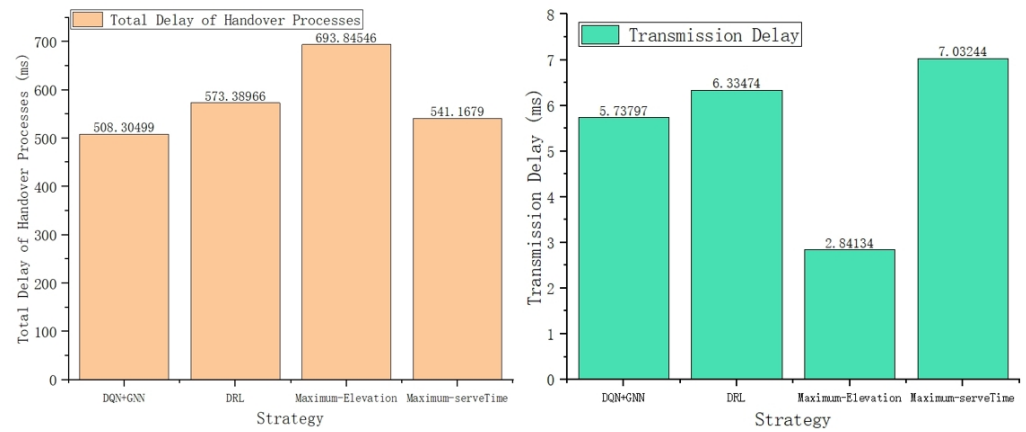
Figure 11. Comparison chart of delay.

4.3.4. Complexity Analysis

The handover strategy proposed in this paper combines the MPNN and DQN algorithms, and the time complexity of each message passing of the message passing neural network is O(E), where E is the number of edges. Assuming K message-passing iterations, the time complexity is O(E × K), and the time complexity of the deep Q network is mainly in the Q-value computation and updating, and the computational complexity of each time step is O(F), where F is the number of parameters of the neural network. Assuming there are T moments, the time complexity is O(T × F). Since our proposed strategy does not learn every moment, but learns a time slot, the time complexity of the strategy is O(E × K + T × F/5). The MPNN part stores the edges and the information on the edges, and the space complexity is O(E).The DQN part of the neural network has a storage complexity of O(F). Then, the space complexity of the strategy is O(E + F). The time complexity of the handover strategy based on the DQN algorithm is O(T × F) and the space complexity is O(F). The handover strategy based on maximum elevation angle, and based on maximum remaining service time performs one conditional judgment and selection operation per time step and, assuming that there are N satellite candidates, the time complexity of each time step is O(N). With T time moments, the total time complexity is O(N × T), and the conventional algorithms usually do not require additional complex data structures with a space complexity of O(1).

The proposed MPNN-DQN method has higher computational and time complexity compared to traditional heuristic methods and some simpler DRL-based methods due to the additional processing of GNN layers. However, this complexity is justified by the improved performance and adaptability in dynamic satellite network environments, as shown in the experimental results.

The training size and computational time for each strategy is given in Table 4. The machine learning-based algorithms (DQN + GNN and DRL) involve a large amount of training time and data, and require higher computational resources while delivering better performance. In contrast, handover strategies based on traditional methods such as Max Elevation and Max Servetime are computationally inexpensive, but may not achieve the same level of optimization.

Table 4. Computing Time and Training Size.

Method	Training Size	Computing Time	Training Required
DQN + GNN	5000 episodes	24 h	Yes
DRL	5000 episodes	18 h	Yes
Max-Elevation	N/A	N/A	No
Max-ServeTime	N/A	N/A	No

5. Discussion

This investigation introduces an innovative methodology designed to optimize the satellite handover process within Low Earth Orbit (LEO) constellations through the integration of graph neural network (GNN) and deep reinforcement learning (DRL) technologies, thereby ensuring uninterrupted handover and enhancing the quality of service. The outcomes section underscores the model's proficiency in mitigating handover failures and curtailing transmission latency, factors that are paramount to ensuring the reliability and efficiency of satellite communications. Prior investigations in this domain have predominantly concentrated on the deployment of conventional handover mechanisms. The method presented herein is advanced by the amalgamation of graph neural network (GNN) and deep reinforcement learning (DRL) algorithms, leveraging GNN's robust representational capabilities to adapt to evolving satellite network conditions, and employing DRL to effectuate real-time adaptive handover decisions.

6. Conclusions

This paper presents an innovative approach to satellite handover strategies within LEO satellite constellations, leveraging the synergistic potential of GNN and DRL. Our proposed MPNN-DQN algorithm addresses the dynamic and complex nature of satellite networks, offering a robust solution to ensure high-quality, continuous communication services. By comparing simulation with traditional handover strategies, our method reduces handover delay and frequency while ensuring user data rate as much as possible. This plays a important role in emergency communication scenarios for power systems.

Author Contributions: Conceptualization, H.Y. and K.Z.; methodology, H.Y. and K.Z.; investigation, H.Y. and W.G.; writing—original draft preparation, H.Y.; writing—review and editing, W.G., H.Y. and K.Z.; project administration, K.Z. and W.G. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Data will be made available on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Yu, H.; Li, P.; Zhang, L.; Zhu, Y.; Al-Zahrani, F.A.; Ahmed, K. Application of optical fiber nanotechnology in power communication transmission. *Alex. Eng. J.* **2020**, *59*, 5019–5030. [CrossRef] [CrossRef]
2. Cao, J.; Liu, J.; Li, X.; Zeng, L.; Wang, B. Performance analysis of a new power wireless private network in intelligent distribution networks. In Proceedings of the 2012 Power Engineering and Automation Conference, Wuhan, China, 18–20 September 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 1–4. [CrossRef]
3. Li, Z.; Wang, Y.; Liu, M.; Sun, R.; Chen, Y.; Yuan, J.; Li, J. Energy efficient resource allocation for UAV-assisted space-air-ground Internet of remote things networks. *IEEE Access* **2019**, *7*, 145348–145362. [CrossRef] [CrossRef]
4. Zhu, Q.; Sun, F.; Hua, Z. Research on hybrid network communication scheme of high and low orbit satellites for power application. In Proceedings of the 2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, China, 11–13 December 2020; IEEE: Piscataway, NJ, USA, 2020; Volume 9, pp. 460–466. [CrossRef]
5. Berger, L.T.; Iniewski, K. *Smart Grid Applications, Communications, and Security*; John Wiley & Sons: Hoboken, NJ, USA, 2012. [CrossRef]
6. Liu, J.; Shi, Y.; Fadlullah, Z.M.; Kato, N. Space-air-ground integrated network: A survey. *IEEE Commun. Surv. Tutor.* **2018**, *20*, 2714–2741. [CrossRef] [CrossRef]
7. De Sanctis, M.; Cianca, E.; Araniti, G.; Bisio, I.; Prasad, R. Satellite communications supporting internet of remote things. *IEEE Internet Things J.* **2015**, *3*, 113–123. [CrossRef] [CrossRef]
8. Li, B.; Li, Z.; Zhou, H.; Chen, X.; Peng, Y.; Yu, P.; Wang, Y.; Feng, X. A system of power emergency communication system based BDS and LEO satellite. In Proceedings of the 2021 Computing, Communications and IoT Applications (ComComAp), Shenzhen, China, 26–28 November 2021; IEEE: Hoboken, NJ, USA, 2021; pp. 286–291. [CrossRef]
9. Liu, J.; Shi, Y.; Zhao, L.; Cao, Y.; Sun, W.; Kato, N. Joint placement of controllers and gateways in SDN-enabled 5G-satellite integrated network. *IEEE J. Sel. Areas Commun.* **2018**, *36*, 221–232. [CrossRef] [CrossRef]

10. Giordani, M.; Zorzi, M. Satellite communication at millimeter waves: A key enabler of the 6G era. In Proceedings of the 2020 International Conference on Computing, Networking and Communications (ICNC), Big Island, HI, USA, 17–20 February 2020; IEEE: Hoboken, NJ, USA, 2020; pp. 383–388. [CrossRef]

11. Liu, Z.; Zha, X.; Ren, X.; Yao, Q. Research on Handover Strategy of LEO Satellite Network. In Proceedings of the 2021 2nd International Conference on Big Data and Informatization Education (ICBDIE), Hangzhou, China, 2–4 April 2021; IEEE: Hoboken, NJ, USA, 2021; pp. 188–194. [CrossRef]

12. Gong, Y. A Review of Low Earth Orbit Satellite Communication Mobility Management. *Commun. Technol.* **2023**, *56*, 923–928.

13. Ilčev, S.D. *Global Mobile Satellite Communications Applications*; Springer: Berlin/Heidelberg, Germany, 2018. [CrossRef]

14. Han, Z.; Xu, C.; Zhao, G.; Wang, S.; Cheng, K.; Yu, S. Time-varying topology model for dynamic routing in LEO satellite constellation networks. *IEEE Trans. Veh. Technol.* **2022**, *72*, 3440–3454. [CrossRef] [CrossRef]

15. Jia, M.; Zhang, X.; Sun, J.; Gu, X.; Guo, Q. Intelligent resource management for satellite and terrestrial spectrum shared networking toward B5G. *IEEE Wirel. Commun.* **2020**, *27*, 54–61. [CrossRef] [CrossRef]

16. Bottcher, A.; Werner, R. Strategies for handover control in low earth orbit satellite systems. In Proceedings of the IEEE Vehicular Technology Conference (VTC), Stockholm, Sweden, 8–10 June 1994; IEEE: Hoboken, NJ, USA, 1994; pp. 1616–1620. [CrossRef]

17. Xu, J.; Wang, Z.; Zhang, G. Design and transmission of broadband LEO constellation satellite communication system based on high-elevation angle. *Commun. Technol.* **2018**, *51*, 1844–1849.

18. Papapetrou, E.; Karapantazis, S.; Dimitriadis, G.; Pavlidou, F.N. Satellite handover techniques for LEO networks. *Int. J. Satell. Commun. Netw.* **2004**, *22*, 231–245. [CrossRef] [CrossRef]

19. Gkizeli, M.; Tafazolli, R.; Evans, B.G. Hybrid channel adaptive handover scheme for non-GEO satellite diversity based systems. *IEEE Commun. Lett.* **2001**, *5*, 284–286. [CrossRef] [CrossRef]

20. Wang, Z.; Li, L.; Xu, Y.; Tian, H.; Cui, S. Handover control in wireless systems via asynchronous multiuser deep reinforcement learning. *IEEE Internet Things J.* **2018**, *5*, 4296–4307. [CrossRef] [CrossRef]

21. Duan, C.; Feng, J.; Chang, H.; Song, B.; Xu, Z. A novel handover control strategy combined with multi-hop routing in LEO satellite networks. In Proceedings of the 2018 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), Vancouver, BC, Canada, 21–25 May 2018; IEEE: Hoboken, NJ, USA, 2018; pp. 845–851. [CrossRef]

22. Rehman, T.; Khan, F.; Khan, S.; Ali, A. Optimizing satellite handover rate using particle swarm optimization (pso) algorithm. *J. Appl. Emerg. Sci.* **2017**, *7*, 53–63. [CrossRef]

23. Zhou, J.; Ye, X.; Pan, Y.; Xiao, F.; Sun, L. Dynamic channel reservation scheme based on priorities in LEO satellite systems. *J. Syst. Eng. Electron.* **2015**, *26*, 1–9. [CrossRef] [CrossRef]

24. Huang, F.; Xu, H.; Zhou, H.; Wu, S.Q. QoS based average weighted scheme for LEO satellite communications. *J. Electron. Inf. Technol.* **2008**, *30*, 2411–2414. [CrossRef] [CrossRef]

25. Miao, J.; Wang, P.; Yin, H.; Chen, N.; Wang, X. A multi-attribute decision handover scheme for LEO mobile satellite networks. In Proceedings of the 2019 IEEE 5th International Conference on Computer and Communications (ICCC), Chengdu, China, 6–9 December 2019; IEEE: Hoboken, NJ, USA, 2019; pp. 938–942. [CrossRef]

26. Wu, Z.; Jin, F.; Luo, J.; Fu, Y.; Shan, J.; Hu, G. A graph-based satellite handover framework for LEO satellite communication networks. *IEEE Commun. Lett.* **2016**, *20*, 1547–1550. [CrossRef] [CrossRef]

27. Hu, X.; Song, H.; Liu, S.; LI, X.; Wang, W.; Wang, C. Real-time prediction and updating method for LEO satellite handover based on time evolving graph. *J. Commun.* **2018**, *39*, 43–51. [CrossRef]

28. Hozayen, M.; Darwish, T.; Kurt, G.K.; Yanikomeroglu, H. A graph-based customizable handover framework for LEO satellite networks. In Proceedings of the 2022 IEEE Globecom Workshops (GC Wkshps), Rio de Janeiro, Brazil, 4–8 December 2022; IEEE: Hoboken, NJ, USA, 2022; pp. 868–873. [CrossRef]

29. Li, H.; Liu, R.; Hu, B.; Ni, L.; Wang, C. A multi-attribute graph based handover scheme for LEO satellite communication networks. In Proceedings of the 2022 IEEE 10th International Conference on Computer Science and Network Technology (ICCSNT), Dalian, China, 22–23 October 2022; IEEE: Hoboken, NJ, USA, 2022; pp. 127–131. [CrossRef]

30. Wang, J.; Mu, W.; Liu, Y.; Guo, L.; Zhang, S.; Gui, G. Deep reinforcement learning-based satellite handover scheme for satellite communications. In Proceedings of the 2021 13th International Conference on Wireless Communications and Signal Processing (WCSP), Changsha, China, 20–22 October 2021; IEEE: Hoboken, NJ, USA, 2021; pp. 1–6. [CrossRef]

31. Zhang, C.; Zhang, N.; Cao, W.; Tian, K.; Yang, Z. An AI-based optimization of handover strategy in non-terrestrial networks. In Proceedings of the 2020 ITU Kaleidoscope: Industry-Driven Digital Transformation (ITU K), Ha Noi, Vietnam, 7–11 December 2020; IEEE: Hoboken, NJ, USA, 2020; pp. 1–6. [CrossRef]

32. Liang, J.; Zhang, D.; Qiu, F. Multi-attribute Handoff Control Methodfor LEO Satellite Internet. *J. Army Eng. Univ.* **2022**, *1*, 14–20. [CrossRef]

33. 3GPP. Study on Architecture Aspects for Using Satellite Access in 5G. Technical Report, 3rd Generation Partnership Project (3GPP), Technical Specifcation (TS) 23.737, March 2021. Available online: https://itecspec.com/archive/3gpp-specification-tr-23-737/ (accessed on 23 June 2024).

34. Vallado, D.; Crawford, P.; Hujsak, R.; Kelso, T. Revisiting spacetrack report# 3. In Proceedings of the AIAA/AAS Astrodynamics Specialist Conference and Exhibit, Big Sky, MT, USA, 21–24 August 2006; p. 6753. [CrossRef]

35. Liu, M. Research on Handover Strategy in Low Earth Orbit Mobile Satellite Network. Master's Thesis, Chongqing University of Posts and Telecommunications, Chongqing, China, 2021. [CrossRef]

36.   Scarselli, F.; Gori, M.; Tsoi, A.C.; Hagenbuchner, M.; Monfardini, G. The graph neural network model. *IEEE Trans. Neural Networks* **2008**, *20*, 61–80. [CrossRef] [CrossRef]

37.   Gilmer, J.; Schoenholz, S.S.; Riley, P.F.; Vinyals, O.; Dahl, G.E. Neural message passing for quantum chemistry. In Proceedings of the INTERNATIONAL Conference on Machine Learning, PMLR, Sydney, Australia, 6–11 August 2017; pp. 1263–1272. [CrossRef]

38.   Watkins, C.J.; Dayan, P. Q-learning. *Mach. Learn.* **1992**, *8*, 279–292. [CrossRef] [CrossRef]

39.   Mnih, V.; Kavukcuoglu, K.; Silver, D.; Graves, A.; Antonoglou, I.; Wierstra, D.; Riedmiller, M. Playing atari with deep reinforcement learning. *arXiv* **2013**, arXiv:1312.5602. [CrossRef]

40.   Almasan, P.; Suárez-Varela, J.; Rusek, K.; Barlet-Ros, P.; Cabellos-Aparicio, A. Deep reinforcement learning meets graph neural networks: Exploring a routing optimization use case. *Comput. Commun.* **2022**, *196*, 184–194. [CrossRef] [CrossRef]

41.   Zhang, Z. Research on LEO Satellite Handover Technology Based on 5G Architecture. Master's Thesis, Beijing University of Posts and Telecommunications, Beijing, China, 2023. [CrossRef]