Ultra-Dense LEO Satellite-Aircraft Access and Service Management in Civil Aviation

Wang Yilei[1], Ma Ting[2], Liu Xiaoyu[1], Gao Zhuxuan[1], Zhou Haibo[1,*], Shen Xuemin[3]

[1] School of Electronic Science and Engineering, Nanjing University, Nanjing 210046, China
[2] School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
[3] The Department of Electrical and Computer Engineering, University of Waterloo, Waterloo N2L 3G1, Canada
[*] The corresponding author, email: haibozhou@nju.edu.cn

Cite as: Y. Wang, T. Ma, *et al*., "Ultra-dense leo satellite-aircraft access and service management in civil aviation," *China Communications*, 2025, vol. 22, no. 1, pp. 277-292. **DOI**: 10.23919/JCC.fa.2023-0072.202501

Abstract: With the deployment of ultra-dense low earth orbit (LEO) satellite constellations, LEO satellite access network (LEO-SAN) is envisioned to achieve global Internet coverage. Meanwhile, the civil aviation communications have increased dramatically, especially for providing airborne Internet services. However, due to dynamic service demands and onboard LEO resources over time and space, it poses huge challenges in satellite-aircraft access and service management in ultra-dense LEO satellite networks (UDLSN). In this paper, we propose a deep reinforcement learning-based approach for ultra-dense LEO satellite-aircraft access and service management. Firstly, we develop an airborne Internet architecture based on UDLSN and design a management mechanism including medium earth orbit satellites to guarantee lightweight management. Secondly, considering latency-sensitive and latency-tolerant services, we formulate the problem of satellite-aircraft access and service management for civil aviation to ensure service continuity. Finally, we propose a proximal policy optimization-based access and service management algorithm to solve the formulated problem. Simulation results demonstrate the convergence and effectiveness of the proposed algorithm with satisfying the service continuity when applying to the UDLSN.

Keywords: civil aviation; deep reinforcement learning; satellite-aircraft access; service management; ultra-dense LEO satellite network

I. INTRODUCTION

The International Air Transport Association (IATA) has predicted that by 2036, there will be 7.8 billion civil aviation passengers world-wide, roughly doubling from the current level [1]. At the same time, it also brings an increase in the communication and service needs of aircraft passengers, such as real-time communication needs for safety in the cockpit and broadband communication needs for entertainment in the cabin [2, 3]. From the perspective of driving safety and passenger experience, it has become an inevitable trend for civil aviation to provide in-flight networking services [4, 5]. At present, there are two main technical schemes for the airborne communication system: air-to-ground (ATG) communication and satellite communication [6, 7]. Base stations on the ground offer a better Internet experience but have the natural disadvantage of being unable to support cross-ocean flights. Furthermore, as an indispensable technical means of 6G [8–10], the layout of LEO satellite access network (LEO-SAN) is progressing rapidly, and many high-tech companies such as SpaceX [11] have invested in the field of low earth orbit (LEO) satellite

Received: Jan. 31, 2023
Revised: Jun. 12, 2023
Editor: Xiao Zhenyu

communications, proposing ultra-dense LEO satellite networks (UDLSNs) such as Starlink to achieve global Internet coverage. The EMEA Satellites Operators Association (ESOA) states that the integration of ultra-dense satellite communications as a key technology for future communications development can provide extensive and reliable broadband connectivity [12]. Therefore, in order to fully meet the needs of global aviation users, airborne Internet will gradually transition from ground base station access to satellite forwarding access [13, 14].

Meanwhile, along with the booming air transport industry, the communication requirements of civil aviation have increased exponentially, and people are desperately seeking fast connections and high-capacity communication all the time [15]. Therefore, there are various sorts of services that civil aviation Internet needs to carry, and in order to guarantee the different service requirements of civil aviation, aircraft must select the appropriate satellite to access on time and manage the service effectively and flexibly to ensure a long-term connectivity experience [16]. However, unlike traditional terrestrial networks, the large scale and rapidly changing topology of UDLSN pose a significant challenge to service management, while limited on-board resources for LEO satellites and more demanding user service requirements make it urgent to propose a timely and effective method for aircraft-satellite access and service management. Access and service management methods have been studied in recent years. Lin et al. in [17] adopt multicast transmission to increase the received rate for all customers under the constraints of the quality of service (QoS) requirements. Jia et al. in [18] propose an adaptive access control and resource allocation scheme that enables the duration of the update gap to vary according to the flow. Zhang et al. in [19] conduct resource allocation optimization for terrestrial-satellite links within the constraints of the backhaul rate and the QoS of the users. However, the complexity of the algorithm grows exponentially as the size of the satellite increases, so these algorithms cannot be applied to UDLSN. Furthermore, these studies also do not take into account the diversity of user requirements and do not enable high-capacity communications. In addition, most of the existing algorithms are one-off optimization algorithms, which cannot be adapted to complex and dynamic satellite topologies and have no robustness to ensure stable connectivity through fast updating access and service delivery decisions based on dynamic network information. Thus, it is necessary but challenging to efficiently conduct satellite-aircraft access and service management in the UDLSN to meet different service demands.

In this paper, based on the diversity of civil aviation carrying services, we consider two types of airborne Internet service requirements: latency-sensitive voice services, such as telephone calls and video calls, and latency-tolerant streaming services, such as video on demand (VoD) and interactive personality TV (IPTV). We propose an optimization problem of satellite access and service management for civil aviation in UDLSN. To cope with the complexity of network nodes as well as their flexibility and variability, we adopt a lightweight management architecture in UDLSN. The deep reinforcement learning (DRL) method is utilized to address the problem, which enables efficient and intelligent management of airborne satellite communication systems.

The main contributions and innovations of this paper are listed below:

- We propose an airborne network architecture that considers the ultra-dense LEO satellite constellation as the LEO-SAN. In this architecture, we use a two-level control management mechanism combining medium earth orbit (MEO) satellites and partial LEO satellites to guarantee lightweight control, and the aircraft access decisions are made by the two-level controllers.
- To achieve efficient ultra-dense LEO satellite-aircraft access and service management in civil aviation, we propose a proximal policy optimization-based access and service management (PPO-ASM) algorithm, which meets the demands of two types of civil aviation services with different latency sensitivities and balances the satellite load.
- To evaluate the performance of the proposed PPO-ASM algorithm, we compare it to other commonly used methods. The numerical results demonstrate that the proposed method can successfully deal with the dynamic challenges of humongous nodes, achieve efficient service management, and enhance user satisfaction.

The remainder of this paper are structured as follows. Section II contains the related work about satellite access and service management. Section III describes the system depiction. Section IV depicts the problem formulation, Section V gives the PPO-ASM algorithm. Section VI presents the performance evaluation. Section VII concludes our work.

II. RELATED WORKS

Given the burgeoning of civil aviation, the requirement for airborne Internet is growing, and many researchers have explored the communication need of civil aviation in recent years.

In [20], the authors compared the performance of delay-aware resource control using dual-connected and single-connected modes for aviation network scenarios. The conclusion is that the dual-connected mode performs only slightly better than the single-connected mode but takes up more resources and is more complex, so it only makes more sense to use the single-connected mode. Deng et al. in [21] proposed a channel model for aircraft-satellite terahertz band communications and applied the model to estimate the characteristics of the aviation terahertz link under different conditions. The final results show that the capacity of the airborne satellite terahertz link can reach 50–150 Gbps, which is sufficient to enable all passengers and crew on board to obtain data transmission rates equivalent to those of ground base station communications throughout the flight. Nowadays, ultra-dense LEO satellite constellations have become a potential solution to address the diversity of communication services for dense users, and many scholars have explored ultra-dense constellation configurations. The authors in [22] and [23] constructed constellation configurations that minimize the network's total amount of satellites while considering the user terminal backhaul requirements as well as the overall service efficiency, respectively. Dense constellations are becoming a fundamental component of future technologies for 6G and beyond. On this basis, we set the civil aviation Internet to communicate in a single connection mode and use a UDLSN as a communication network for real-time communication as well as service management.

Unlike traditional terrestrial networks, satellite-based airborne network in which aircraft and LEO satellites are in high-speed motion, the topology changes rapidly and carries a wide variety of services, so the traditional static access and resource allocation methods will no longer be applicable, and there is an urgent requirement to find management methods to adapt the complex network of highly dynamic multi-nodes. For satellite networks, a user-centric switching method is presented in [24] to address the issue of frequent switching and to provide a higher QoS, which provides an improvement in throughput and end-to-end delay over the traditional methods. YAN et al. in [25] designed a space-ground integration architecture based on a software-defined network (SDN). To manage users' access with different priorities and simultaneously meet the requirements on the probability of successful access, a dynamic channel resource allocation method is provided. A new global coverage LEO/MEO double-layer satellite network is designed using the street-of-coverage technique in [26], whose simulations indicate that the proposed constellation configuration provides complete coverage of China on average, and covers 99.9% of the global area. Therefore, we adopt double-layer satellite network for network management in the civil aviation satellite Internet scenario, where MEOs and some LEOs act as global controllers and local controllers responsible for mobility management and content distribution, respectively, thus improving network robustness and ensuring lightweight management. For the access of civil aircraft and the allocation of time and frequency resources of satellites, we consider the communication requirements of two types of civil aviation classical services, namely voice services and streaming services, in order to guarantee the satisfaction of all civil aviation users.

Most of the existing studies use one-time optimization algorithms, such as convex optimization or genetic algorithms, which cannot be adapted to changes in the environment to achieve access decisions and maintain flexible network control continuously. The performance of these techniques will be constrained as the issue's complexity rises. Control management for UDLSNs is more flexible attributable of the adoption of DRL methods. By interacting with the environment, it can adjust to situations that change over time and make prompt decisions. Moreover, DRL methods can further improve the performance of satellite networks to maximize long-term benefits. To decrease the frequency of satellite switching and increase long-term system throughput, [27] proposed a scheme based on terminal-driven DRL, in which the deep Q-network parameters are trained by a centralized intelligent agent installed on the backhaul side of the satellite base station, and at the end of the training each terminal can independently make appropriate access decisions based on the trained parameters. In [28], the authors adopted SDN architecture to operate the satellite network and implement a deep Q-learning technique to resource management to enhance network performance. For the limited available frequency bandwidth and fast-changing user requirements, the authors in [29] proposed a satellite resource allocation algorithm based on the DRL algorithm for satellite time-frequency resource allocation to meet the time-sensitive requirements of users. Compared with conventional optimization algorithms, the application of DRL brings more flexibility to the resource allocation of satellite networks. It can adapt to time-varying situations and dynamically allocate resources by interacting with the environment. Based on the trained model with the management architecture, the aircraft can make timely access actions to guarantee the continuity of services and effectively solve the high complexity of the ultra-dense constellation scenario [30]. Therefore, we use the DRL algorithm to solve the optimization problem and effectively cope with the high dynamics of large-scale nodes in the satellite network, so that realize the intelligent management of the airborne network.

Existing studies rarely consider the UDLSN and civil aviation Internet services. Based on the above discussion, this paper considers the UDLSN as LEO-SAN for civil aviation and adopts layered lightweight management architecture to control large-scale nodes timely and effectively. According to the demand for civil aviation, we consider voice service and streaming service as civil aviation primary services, and the DRL algorithm is applied to learn the satellite access selection and perform service management so as to maximize the satisfaction of civil aviation passengers.

III. SYSTEM MODEL

3.1 Network Model

The UDLSN-based airborne internet architecture is depicted in Figure 1. Inspired by [31], we use a double-layer hybrid network architecture to ensure the reliability of network management, where MEO satellites are selected as the network manager to centrally control and maintain LEO satellites in the coverage area. Ultra-dense LEO satellite constellation builds LEO-SAN to provide connectivity services for aircraft during their flight missions. In order to prevent the MEO satellite controllers from overloading, we further divide the LEO satellites into multiple clusters and select one LEO satellite in each cluster as a local manager to control all LEO satellites in the cluster, and only the local LEO managers can communicate with MEOs, thus effectively alleviating the burden of MEOs and realizing light weight management. In the two-level control architecture proposed above, the access decision of the aircraft is made by the MEO controller or the local LEO controller. While the interference impact between aircraft is usually presented in a limited area, aircraft are usually managed by the same local LEO controller. In this scenario, the state information of the aircraft is uploaded from each aircraft to the current access satellite and then relayed between LEO satellites until it reaches the local controller. The local controller LEO satellite then performs model training based on the received state information, followed by sending decision instructions back to the aircraft to complete the policy update. If a small probability event occurs and the aircraft flies across domains, the local controller of both domains can further upload information to the global controller MEO for model training, and the global controller MEO then send it down to the local controller and further to the aircraft.

Within the selected time period $\mathcal{T}$, the selected area $\mathcal{R}$ is covered by a MEO satellite c, and all LEOs in the airspace $\mathcal{R}$ are in the same cluster, where LEO l is the manager inside the cluster. Based on our proposed architecture, the LEO satellite l serves as a local controller, which is responsible for collecting environmental status information of aircraft and LEO satellites in the region $\mathcal{R}$ so as to make access decisions, and broadcasts them to the other satellites in the region $\mathcal{R}$, which are then distributed to the aircraft.

Due to the movement of civil aircraft and satellites, the network topology changes dynamically. Thus, time periods are equally divided into multiple small time slots in this paper, which are expressed as $\mathcal{T} = \{1, 2, ...t..., T\}$. The set of LEO satellites and civil

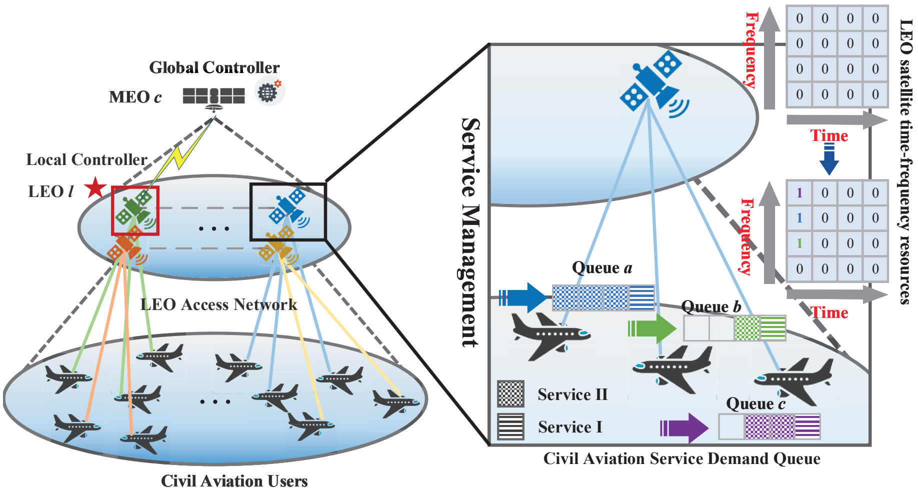


Figure 1. *Two-level control-based airborne network.*

aircraft in the selected region $\mathcal{R}$ are represented as $\mathcal{N} = \{1, 2, ...n..., N\}$ and $\mathcal{M} = \{1, 2, ...m..., M\}$ respectively, where N and M are the total numbers of LEO satellites and civil aircraft, respectively. Denote $x_{m,n}[t] \in \{0, 1\}$ represents whether the civil aircraft m is connected to LEO satellite n in time slot t, $x_{m,n}[t] = 1$ represents that the civil aircraft m is connected to LEO satellite n in time slot t, $x_{m,n}[t] = 0$ represents that the civil aircraft m is not connected to the LEO satellite n in time slot t. Assume that each time slot is of short duration, the network topology does not change in any time slot t.

3.2 Service Model

This section selects two classic aviation service types for research: one is a secure, delay-sensitive voice service, and the other is a high-capacity, delay-tolerant stream service. The data arrival models of these two services will be described respectively.

3.2.1 Voice Service

Voice service is a real-time service and has strict requirements for latency. Suppose that the data arrival of voice service of user u obeys the Poisson process from the active state ($l^u[t] = 1$) to the silent state ($l^u[t] = 0$), that is, ON/OFF model [32]. The state of time slot t is $l^u[t]$, and the transition of ON/OFF state occurs in different time slots. The state transition probability is shown in Figure 2, where τ is the duration of a time slot and τ_1 and τ_2 are respectively the durations of the active and silent states.

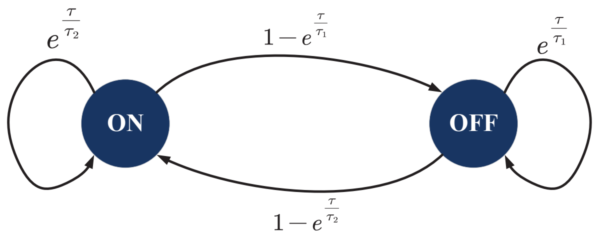


Figure 2. *Voice service data arrival ON/OFF state transition probability diagram.*

Therefore, when the state of the voice service i in time slot t is $l_i^u[t]$ and the unit data volume in the active state is q^u, the data volume of the voice service i in the time slot t is

$$Q_i^u[t] = l_i^u[t] * q^u. \tag{1}$$

3.2.2 Streaming Service

Civil aircraft passengers expect to enjoy VoD at any time during the flight, so streaming service is another main service demand of civil aircraft passengers, which is not as sensitive to latency as session class service. Suppose that the data arrival of streaming service is subject to the Markov process of discrete state in continuous time, and the data quantity arriving at each time slot is quantified as $G + 1$ level, and the unit data quantity is q^e.

Then the data quantity arriving at each level [33] is $2^0 * q^e, 2^1 * q^e, ..., 2^G * q^e$. The state level of time slot t is $l^e[t] = g$, $g \in \{0, 1, ..., G\}$ and the transition of the state level takes place in different time slots.

Then the transition probability of the state level in the next time slot $t+1$ is

$$P(l^e[t+1]=g+\rho|l^e[t]=g)$$
$$=\begin{cases}\frac{\omega}{\omega+\mu}(1-e^{-(\omega+\mu)\tau}), \rho=1\\ \frac{\mu}{\omega+\mu}(1-e^{-(\omega+\mu)\tau}), \rho=-1\\ e^{-(\omega+\mu)\tau}, \rho=0,\end{cases}$$

where $\mu=g*\alpha$, $\omega=(G-g)*\beta$, α, and β represent transition rates; $\rho=1$ indicates that the state level of the next time slot jumps to a higher level; $\rho=-1$ indicates that the state level of the next time slot jumps to a lower level; $\rho=0$ indicates that the state level of the next time slot does not change.

Therefore, when the state level of the streaming service j in time slot t is $l_j^e[t]$, the data volume of the streaming service j in the time slot t is

$$Q_j^e[t]=2^{l_j^e[t]}*q^e. \quad (2)$$

It is assumed that the total number of voice and streaming services generated in each time slot on aircraft m is U_m, the proportions are r_m^u and r_m^e respectively, and $r_m^u, r_m^e \in [0,1]$, $r_m^u+r_m^e=1$. Then the number of voice and streaming services generated by aircraft m is $U_m*r_m^u$ and $U_m*r_m^e$, respectively.

Denote $Q_m^u[t]$ and $Q_m^e[t]$ as the newly arrived data volumes of voice services and streaming services of civil aircraft m at time t, respectively. Then $Q_m^u[t]$ and $Q_m^e[t]$ can be expressed as

$$Q_m^u[t]=\sum\nolimits_{i=1}^{U_m*r_m^u} Q_i^u[t], \quad (3)$$

$$Q_m^e[t]=\sum\nolimits_{j=1}^{U_m*r_m^e} Q_j^e[t]. \quad (4)$$

3.3 Communication Model

This paper mainly considers the uplink data transmission process from civil aircraft to LEO satellite, the downlink transmission is similar to the uplink, and the communication model will be formulated below.

3.3.1 Channel Model

In air-space links, path fading is serious and the link loss is high, so it is crucial to guarantee the quality of link communication. The signal-to-noise ratio (SNR) of data transmission from civil aircraft m to LEO satellite n in the time slot t is expressed as

$$\varUpsilon_{m,n}[t]=|h_{m,n}[t]|^2*d_{m,n}^{-\delta}[t]*\frac{P_m}{N_0}, \quad (5)$$

where $d_{m,n}[t]$ is the distance between aircraft m and LEO satellite n in time slot t, δ is the path fading index, P_m is the transmitting power of civil aircraft m, and N_0 is the additive Gaussian noise variance [34]. The satellite-aircraft links are modeled as Shadowed-Rician fading channels [35]. Thus, according to the Shannon formula, the data transmission rate of the uplink communication link between civil aircraft m and LEO satellite n in time slot t is

$$R_{m,n}[t]=\frac{B_n}{C_n}*\log_2(1+\varUpsilon_{m,n}[t]), \quad (6)$$

where B_n represents the total bandwidth of the LEO satellite n in the uplink and C_n is the number of carrier channels of the LEO satellite n in the uplink.

3.3.2 Delay Definition

Latency is an important metric that affects the user service experience. We consider latency as one of the key factors to measure system performance. The total delay of uplink communication between civil aircraft and LEO satellite includes propagation delay, transmission delay, and handover delay. The following three types of delays are described, respectively.

Propagation Delay: The uplink propagation delay from aircraft m to LEO satellite n in time slot t is

$$D_{m,n}^p[t]=x_{m,n}[t]*(\frac{d_{m,n}[t]}{c}), \quad (7)$$

where $d_{m,n}[t]$ denotes the distance between satellite n and aircraft m in time slot t, and c is the speed of light.

Transmission Delay: The uplink transmission delay from aircraft m to LEO satellite n in time slot t is

$$D_{m,n}^t[t]=x_{m,n}[t]*\frac{Q_m[t]}{R_{m,n}[t]}, \quad (8)$$

where $Q_m[t]$ is the total amount of data to be transmitted by civil aircraft m in time slot t. It is assumed that all data arriving can be transmitted within one time slot, so

$$Q_m[t] = Q_m^u[t] + Q_m^e[t]. \tag{9}$$

Handover Delay: Because of the simultaneous mobility of aircraft and satellites, civil aircraft will constantly switch between the ultra-dense LEO satellite constellations, so the handover delay is also taken into account in this paper. The uplink handover delay from civil aircraft m to LEO satellite n in time slot t is

$$D_{m,n}^h[t] = D_H \mathbb{I}_{\{t>1 \& x_{m,n}[t]=1 \& x_{m,n}[t-1]=0\}}, \tag{10}$$

where D_H is the single handover delay, and when the logical variable condition in $\mathbb{I}_{\{condition\}}$ is True, $\mathbb{I}_{\{condition\}} = 1$. On the contrary, $\mathbb{I}_{\{condition\}} = 0$, that is, only when $t > 1$, $x_{m,n}[t-1] = 0$ and $x_{m,n}[t] = 1$, there will be handover delay.

Therefore, the total time delay of uplink communication between civil aircraft m and LEO satellite n in time slot t is:

$$D_{m,n}[t] = D_{m,n}^p[t] + D_{m,n}^t[t] + D_{m,n}^h[t]. \tag{11}$$

Then the uplink communication delays for voice and streaming services are as follows:

$$D_{m,n}^u[t] = D_{m,n}^p[t] + x_{m,n}[t] * \frac{Q_m^u[t]}{R_{m,n}[t]} + D_{m,n}^h[t], \tag{12}$$

$$D_{m,n}^e[t] = D_{m,n}^p[t] + x_{m,n}[t] * \frac{Q_m^e[t]}{R_{m,n}[t]} + D_{m,n}^h[t]. \tag{13}$$

IV. PROBLEM FORMULATION

In this section, we will define the decision variable, objective function, and constraints for the problem of satellite-aircraft access and service management for the different communication requirements of the two types of civil aviation services.

We need to choose which LEO satellite access n for each civil aircraft m in each time slot t according to LEO satellite time and frequency resource state information and civil aircraft service demand state information, that is, to determine the decision variables $x_{m,n}[t]$. After determining $x_{m,n}[t]$, one carrier channel of LEO satellite n is allocated to each civil aircraft m for the transmission of data in the service queue, so as to meet the service requirements of civil aircraft m as far as possible.

The optimization objective of this paper is to meet the service requirements of all civil users as far as possible, and because of the diversity of civil aviation service types, we need to consider the different service requirements of both voice and streaming service users.

For users of delay-sensitive voice services, we need to meet the maximum tolerable delay of the service as much as possible to ensure user satisfaction. Therefore, for this type of service we define the optimization objective as minimizing the number of times that the voice service communication delay is higher than the maximum tolerable delay, and define the number of times that the voice service communication delay of civil aircraft m over the maximum tolerable delay D_{TH}^u at time slot t as

$$T_m^u[t] = \sum\nolimits_{n=1}^{N} \left(\frac{Q_m^u[t]}{q^u}\right) \mathbb{I}_{D_{m,n}^u[t] > D_{TH}^u}, \tag{14}$$

so for all civil aircraft at time t, the optimization objective for voice services is

$$T^u[t] = \sum_{m=1}^{M} T_m^u[t]. \tag{15}$$

For delay-tolerant streaming service users, the requirement is to maintain the fairness of user communication and the high efficiency of resource utilization as much as possible, so the optimization objective should be to minimize the average communication delay of the whole user. Define the average time delay of all civil aircraft streaming services in time slot t as

$$T^e[t] = \frac{1}{M} \sum\nolimits_{m=1}^{M} \sum\nolimits_{n=1}^{N} D_{m,n}^e[t]. \tag{16}$$

We set the optimization goal to meet the service needs of all civil aircraft in the entire time slot T as much as possible. Therefore, in order to meet the service needs of civil aviation users and bring them an improved service experience, this paper establishes a multi-objective optimization problem of UDLSN access and service management for two types of civil aviation services, namely, delay-sensitive and delay-tolerant services, using as $x_{m,n}[t]$ decision variables which is shown as

$$\begin{aligned} &\min_{x_{m,n}[t]} \left(\frac{1}{T} \sum_{t=1}^{T} \left(w_1 * \frac{T^u[t]}{N_{th}} + w_2 * \frac{T^e[t]}{D_{th}^e} \right) \right), \\ &s.t.\ C1: \sum\nolimits_{n=1}^{N} x_{m,n}[t] \le 1, \forall t \in \mathcal{T}, m \in \mathcal{M}, \\ &\quad C2: \sum\nolimits_{m=1}^{M} x_{m,n}[t] \le C_n, \forall t \in \mathcal{T}, n \in \mathcal{N}, \end{aligned} \tag{17}$$

where w_1 and w_2 are weight coefficients used to measure the importance of voice and streaming services, $w_1 + w_2 = 1$, N_{th} is the threshold for the number of voice services that exceed the tolerated delay, and D_{th}^e is the set delay threshold for streaming services.

It is stipulated that each time slot can only be connected to one LEO satellite at most, and the total number of aircraft connected to the same LEO satellite at the same time cannot exceed the carrier channel number of the uplink of this LEO satellite.

V. ALGORITHM DESIGN

To address the optimization problem raised above, this paper proposes a proximal policy optimization-based access and service management (PPO-ASM) algorithm using a DRL approach called PPO [36] to obtain the optimal access decision for civil aircraft. The architecture, process, and details of the PPO-ASM algorithm are given in this section.

5.1 Algorithm Architecture

Traditional reinforcement learning methods, such as Q-learning [37] and Sarsa [38], store and update action value estimation in the form of tables. However, when the problem solution space dimension is too high, it will encounter great bottlenecks in storage space and computational efficiency, so it cannot solve many practical problems. The method of the DRL combines the perception ability of the deep learning method with the decision-making ability of reinforcement learning method, which has successfully solved many challenging problems recently.

The PPO algorithm is a DRL algorithm proposed by the OpenAI team in 2017 to overcome the shortcomings of the Trust Region Policy Optimization (TRPO) algorithm [39], which modifies the constraints in the objective function of the TRPO algorithm to a penalty term that penalizes larger policy updates, thus eliminating the conjugate gradient calculation. Compared with other DRL algorithms, the PPO algorithm has higher data validity and robustness, which makes it ideal for application in UDLSNs to cope with the large number of nodes and the pursuit of efficient and reliable civil aviation service needs. Therefore, this paper solves the proposed problem based on the PPO algorithm, copes with the dynamization of civil aviation satellite network nodes, and proposes the PPO-ASM algorithm to satisfy the different service demands of two types of delay sensitivity while maximizing the long-term benefits to users and achieving dynamic and efficient control management in UDLSN scenario.

In the proposed architecture, the UDLSN is considered an environment interacting with the Deep Neural Networks (DNN), and MEO satellite c or local LEO controller l acts as intelligent agent using a chain decision method to update the aircraft access decision one by one. At time slot t, satellite controller needs to make the access decision for each civil aircraft m based on the current civil aircraft's service requirements and the satellite time-frequency resource status information $s_m[t]$, and the environment returns a reward $r_m[t]$ based on the civil aircraft's action $a_m[t]$ and moves to the next state $s_{m+1}[t]$. The states, actions, rewards, and next states at time t constitute experienced data $\langle s_m[t], a_m[t], r_m[t], s_{m+1}[t] \rangle$ and are stored in the experience buffer B_u. When the number of experienced data reaches an empirical threshold C_B, the network will randomly select a batch of experienced data for learning and adjust the network parameters according to the loss function so as to effectively enhance the satisfaction of civil aviation users.

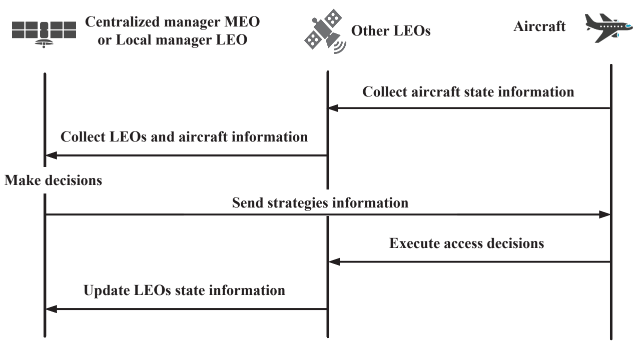


Figure 3. *Centralized decision updating process.*

Based on a large number of UDLSN satellite nodes and a large amount of information, this paper designs a centralized decision-updating process. A trained learning model is deployed in the civil aviation satellite network environment, and the satellite controller determines aircraft update strategy. The decision updating process is shown in Figure 3. The state information of the aircraft is uploaded from the aircraft to the current access satellite and then passed to the local LEO controller through the cluster. The local controller makes access decisions based on the trained model and the collected state information, then sends the decisions to the current access satellite of the aircraft through the inter-satellite link, and then sends them back to the aircraft to complete the policy update. In case of a small probability event such as an aircraft flying across domains, the local LEO controller of both domains can further upload information to the global controller MEO for model training, and then send it down to the local LEO controller and back to the aircraft.

5.2 Algorithm Procedure

Based on the above proposed architecture, the specific steps of the PPO-ASM algorithm will be given, starting with the specific definitions of state S, action A and reward R respectively.

5.2.1 State

For each time slot, the satellite controller needs to observe the state of the environment and make access decisions for each aircraft based on information such as the service volumes of the aircraft and the state of visible satellites. Consequently, the observation state of civil aircraft m in time slot t includes three parts:

- Information about civil aircraft m ($s_m^i[t]$) : Identifier of civil aircraft m, amount of data to be transmitted for voice services $Q_m^u[t]$ and streaming services $Q_m^e[t]$;
- Information about visible LEO satellites set $N_m[t]$ ($s_m^n[t]$) : Identifier of LEO satellite n, the distance between LEO satellite n and aircraft m $d_{m,n}[t]$, LEO satellite uplink channel information $|h_{m,n}[t]|^2$, number of idle channels $C_n[t]$;
- The satellite accessed by aircraft m in the last time slot, that is, the action of aircraft m in the last time slot $a_m[t-1]$, and when $t=0$, the last time slot action does not exist, $a_m[t-1]=-1$.

Therefore, the observation state space of civil aircraft m in time slot t is expressed as

$$s_m[t] = \left\{ s_m^i[t], \{s_m^n[t]\}_{n \in N_m[t]}, a_m[t-1] \right\}. \tag{18}$$

5.2.2 Action

The civil aircraft needs to select a visible satellite to access at each time slot and the satellite to be accessed is not at full payload, so the action space for time slot t civil aircraft m is expressed as

$$a_m[t] = \{ n_m[t] \,|\, n_m[t] \in N_m[t] \}, \tag{19}$$

where the action of civil aircraft m must satisfy $C_m[t] \neq 0$ to guarantee effective satellite access.

5.2.3 Reward

At each time slot t, after the civil aircraft m executes an action, a reward $r_m[t]$ needs to be given back to evaluate the goodness of the action. Based on the service types of the civil aircraft, the reward needs to take into account both streaming and voice service requirements. So we define the reward expression for the civil aircraft m at time slot t, after executing the action $a_m[t]$ in state $s_m[t]$, as follows:

$$r_m[t] = -\omega_1 * \frac{T_m^u[t]}{N_{th}} - \omega_2 * \frac{D_{m,n}^e[t]}{D_{th}^e}. \tag{20}$$

For each of the two service types, the reward gives a penalty based on the service characteristics and contains the average streaming service delay per passenger over the entire time slot $\mathcal{T}$ and the proportion of voice services that exceed the tolerated delay. Then the cumulative discount reward R is defined as:

$$R = \frac{1}{M * T} \sum_{t=1}^{T} \sum_{m=1}^{M} r_m[t]. \tag{21}$$

Thus, when minimizing R, the service needs of all aircraft passengers are maximized in the long run, and this is what we are working on.

The PPO-ASM algorithm adopts an Actor-Critic framework with two neural networks, Actor and Critic, where the Actor network estimates the policy function $\pi_{\theta_a}(a[t]|s[t])$ to perform the action based on the current environment state, and the Critic network estimates the state value function $V_{\theta_c}(s[t])$ based on the current environment state and the next state after performing the action, so as to evaluate the state. The Actor network and the Critic network optimize their respective networks by updating network parameters θ_a and θ_c. θ_a and θ_c are updated by the loss functions L_a and L_c respectively. The expressions for the loss functions L_a, L_c are as follows

$$L_c = (A_m[t])^2, \tag{22}$$

$$\begin{aligned} L_a = -min(&\frac{\pi_{\theta_a^{new}}[a_m[t]|s_m[t]]}{\pi_{\theta_a^{old}}[a_m[t]|s_m[t]]}A_m[t], \\ clip(&\frac{\pi_{\theta_a^{new}}[a_m[t]|s_m[t]]}{\pi_{\theta_a^{old}}[a_m[t]|s_m[t]]}, 1-\epsilon, 1+\epsilon)A_m[t])). \end{aligned} \tag{23}$$

$A_m[t]$ is the advantage function, denoted as

$$A_m[t] = r_m[t] + \gamma * V_{\theta_c}(s_m[t+1]) - V_{\theta_c}(s_m[t]), \tag{24}$$

in which $\gamma \in [0,1]$ is the discount factor, and $r_m[t]$ is the reward obtained by the civil aircraft m after taking action $a_m[t]$ in the environmental state $s[t]$, θ_a^{old} and θ_a^{new} are the vectors of policy parameters before and after the update, respectively, $\pi_{\theta_a^{old}}[a_m[t]|s_m[t]]$ and $\pi_{\theta_a^{new}}[a_m[t]|s_m[t]]$ denote the probability of the strategy before and after the update, respectively; ϵ is the hyperparameter, usually taken as 0.2. The second term of Eq. (23) represents a clipping of $\frac{\pi_{\theta_{new}}(\boldsymbol{a}[t]|s[t])}{\pi_{\theta_{old}}(\boldsymbol{a}[t]|s[t])}$ that restricts it to the interval $[1-\epsilon, 1+\epsilon]$.

Based on the above algorithm characteristics, the detailed procedure of the PPO-ASM algorithm for UDLSN are given in Algorithm 1.

Algorithm 1. *PPO-based access and service management algorithm.*

Input: Information about civil aircraft and LEO satellites $\mathcal{S}$, the learning rate of the actor and critic networks to R_a , R_c.
Output: Civil aircraft access action set $\mathcal{A}$.

Initialize the network parameters θ_a, θ_c;
for episode i in $\{1, 2, \cdots, N_e\}$ **do**
  for time t in $\{1, 2, \cdots, T\}$ **do**
    for civil aircraft m in $\{1, 2, \cdots, M\}$ **do**
      Obtain observation state $s_m[t]$ from the civil aircraft m and LEOs and input it into the actor network;
      The actor network returns an action $a_m[t]$ based on status $s_m[t]$;
      Obtain reward $r_m[t]$ and the state $s_{m+1}[t]$ of aircraft $m+1$ in the time slot t;
      Store the experienced data $\langle s_m[t], a_m[t], r_m[t], s_{m+1}[t] \rangle$ in experience buffer B_u;
    end for
  end for
  if experienced data in buffer $B_u > C_u$ **then**
    Randomly sample experience data from the buffer B_u;
    for update episode i_u in $\{1, 2, \cdots, N_e\}$ **do**
      Calculate the cumulative award R;
      Calculate the advantage function $A_m[t]$;
      Update the parameter θ_a and θ_c by maximizing L_a and minimizing L_c;
    end for
  end if
end for

VI. PERFORMANCE EVALUATION

In this section, we perform a trajectory-driven simulation to evaluate the performance of the PPO-ASM algorithm. The trajectory data of civil aircraft is provided by [40]. We select the track of civil aircraft over the Indian Ocean for the Europe-North America East Coast flight for simulation. We adopt Starlink's full period of 11,927 LEO satellites for the LEO-SAN. Total simulation time T is 8 s, where each time slot τ is 100 ms. Then we set the total number of voice and streaming services for each aircraft passenger in each time slot as 110, and the ratio is $1:10$. The path fading is 2.1 and the noise density is -174dBm/hz [34]. Other major simulation parameters are shown in Table 1. For practical application reasons, we set Starlink's first period 4,408 satellites as the LEO-SAN and the number of aircraft as 50. The average number of visible satellites per aircraft is 60.

Table 1. *Simulation parameters.*

Parameter	Value	Parameter	Value
τ_1	60 ms	τ_2	100 ms
q^u	30 KB	q^e	30 KB
G	4	β	$\frac{3.9}{1+5.04458/G}$
α	$3.9-\beta$	C_n	3
P_m	100 W	B_n	1800 MHz
ω_1	0.5	ω_2	0.5
D_{HO}	30 ms	γ	0.9

6.1 Convergence

Figure 4 shows the convergence trend of the PPO-ASM algorithm at different learning rates. The convergence performance of the PPO-ASM algorithm changes from superior to inferior: $r_a = 0.0001, r_c = 0.00005 > r_a = 0.0002, r_c = 0.0001 > r_a = 0.0004, r_c = 0.0002$. When $r_a = 0.0001, r_c = 0.00005$, the average reward at convergence can reach the highest value, But its convergence speed is slowest, and when $r_a = 0.0004, r_c = 0.0002$, the convergence speed is fastest in all three cases, it can reach convergence at epoch 27. It can be seen that when the learning rate is relatively large, the speed of convergence can be accelerated, and conversely when the learning rate is small, the convergence result can be improved. Thus, in order to balance convergence speed and performance, we set $r_a = 0.0002, r_c = 0.0001$ in future simulations.

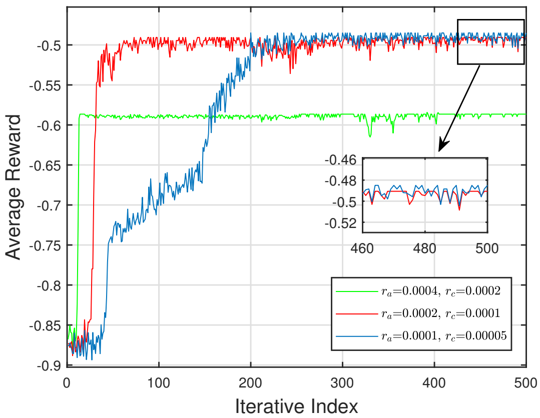


Figure 4. *Convergence at different learning rates.*

6.2 Performance Analysis

To verify the effectiveness of the algorithm, we compared the performance of the PPO-ASM algorithm with the following three algorithms:

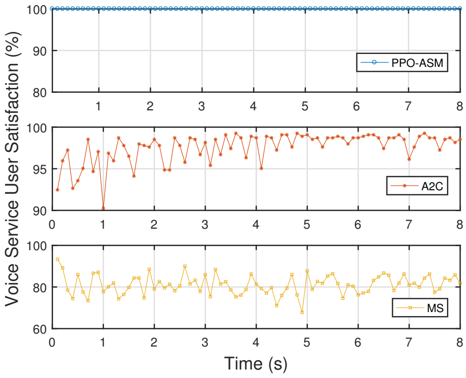
- A2C (Advantage Actor-Critic) algorithm [41]: On the basis of the AC (Actor-Critic) algorithm which uses advantage functions instead of the primitive rewards in the Critic network to measure the quality of the selected action.
- MS (Max SNR) algorithm: Each aircraft selects the satellite with the highest SNR value for the link to access.
- MD (Minimize Distance) algorithm: Each aircraft chooses the satellite with the shortest distance.
- MT (Maximum Time) algorithm: Each aircraft selects the satellite with the longest service time.
- Random (RAND) algorithm: Each aircraft randomly selects a satellite with an available channel.

In the following, different LEO satellite access and data management algorithms are compared in terms of user performance to verify the effectiveness of the algorithm proposed in this paper. Figure 5 and Figure 6 show the average streaming service latency and the average voice service user demand satisfaction for civil aviation users within the simulated time slot, respectively.

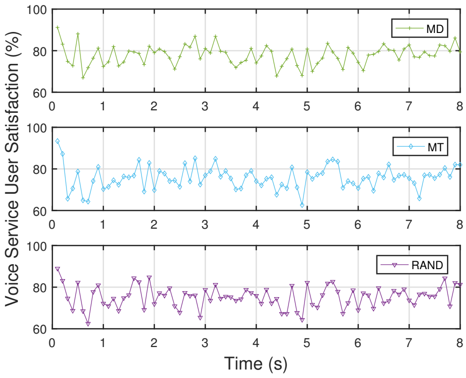
As can be indicated in Figure 5, the PPO-ASM algorithm achieves the highest user satisfaction for voice services of all the compared algorithms and can meet the voice service needs of all users, while the MS algorithm only focuses on a portion of the latency, which deteriorates if requiring satellite switching. The MD and MT algorithms, similarly, only consider propagation delay and switching delay, respectively, and cannot satisfy voice service requirements.

It can be found from Figure 6 that the PPO-ASM algorithm can obtain the shortest delay for streaming services. The A2C algorithm often leads to disruptive and drastic policy updates , so the final performance is worse than the PPO-ASM algorithm. The MS algorithm only focuses on transmission latency and does not take into account switching latency and propagation latency, resulting in higher streaming class latency. The MD and MT algorithms only consider propagation delay and switching delay respectively, which cannot meet the user's service requirements.

It can be postulated that the PPO-ASM algorithm can effectively solve the problem of dynamic changes in the nodes in UDLSN and meet the different ser-

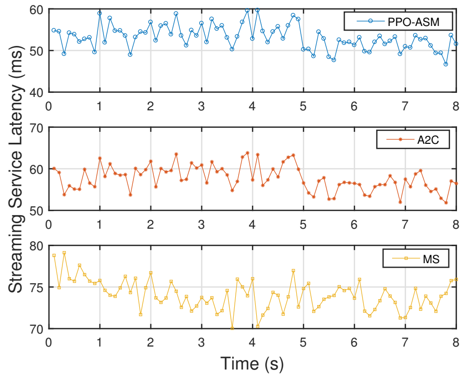


(a) *The average voice service user demand satisfaction obtained by PPO-ASM, A2C, and MS algorithms.*

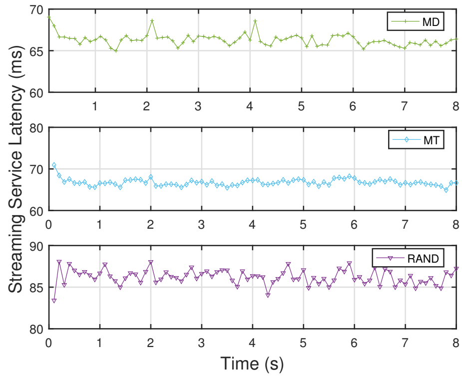


(b) *The average voice service user demand satisfaction obtained by MD, MT, and RAND algorithms.*

Figure 5. *The average voice service user demand satisfaction per time slot.*



(a) *The average streaming service latency obtained by PPO-ASM, A2C, and MS algorithms.*



(b) *The average streaming class service latency obtained by MD, MT, and RAND algorithms.*

Figure 6. *The average streaming service latency per time slot.*

vice requirements while achieving dynamic and efficient allocation of LEO satellite resources.

6.3 Impact of Satellite Scale

Technology implementation related to achieving an airborne Internet comparable to terrestrial cellular networks is still in the exploratory stage. With the rapid deployment of dense constellations, it is necessary to explore the scale of the network required for the Internet in civil aviation. Therefore, we compared the performance of the first group of the first phase, all the constellations of the first phase, and all the satellites of the full phase of Starlink for three different constellation scales in total. Thus, Figure 7 and Figure 8 show the average streaming delay and user satisfaction with voice services for the entire simulation time slot obtained by different algorithms for different satellite scales.

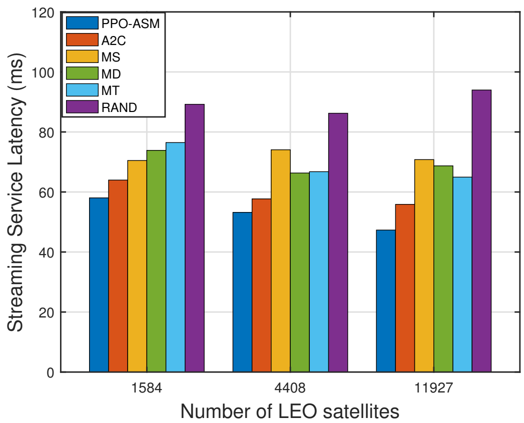


Figure 7. *The average streaming service latency at different satellite sizes.*

Figure 7 indicates that the PPO-ASM algorithm achieves the shortest streaming service delay for different size constellations and that the delay obtained

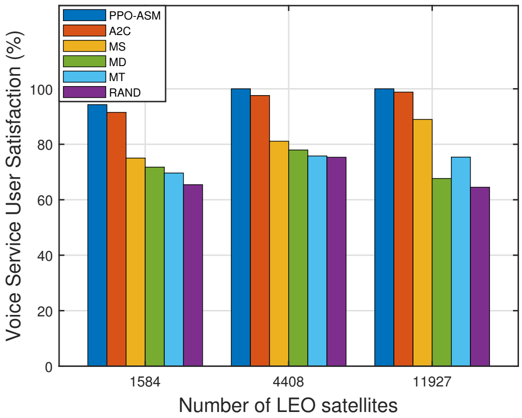


Figure 8. *The average voice service user demand satisfaction at different satellite sizes.*

by the PPO-ASM algorithm decreases as the number of satellites increases. The increased number of satellites means more link options for aircraft. Aircraft can choose closer satellites and satellites with better channel conditions for communication, which in turn reduces communication latency. Obviously, the neural network model of the PPO-ASM algorithm performs better, and when the number of satellites is increased from 4,408 to 11,927, the PPO-ASM algorithm can still significantly reduce the latency of the streaming service, while the performance improvement of the A2C algorithm is minimal. But the latency obtained by the other compared one-off optimization algorithms (e.g. the MD algorithm) does not decrease with increasing constellation size, we can see that because the increase in the number of satellites adds satellites that are closer to the aircraft, but their channel conditions are not necessarily better, and thus the delay obtained is not always shorter.

Figure 8 demonstrates an overall upward trend in user satisfaction with voice services as the LEO satellite constellation size increases and that the PPO-ASM algorithm achieves the highest satisfaction among the algorithms compared. This is mainly due to the increase in the LEO satellite constellation providing satellites with better channel conditions for the PPO-ASM algorithm, and the PPO-ASM algorithm allows to explore better access options, hence voice services user satisfaction also increases. Furthermore, the PPO-ASM algorithm can achieve close to 100% satisfaction at a satellite size of 1584, and when the satellite size is further expanded, the PPO-ASM algorithm can fully satisfy the voice service requirements of all users, which cannot be achieved by other algorithms. It is worth noting that increasing the size of the network does not guarantee performance improvements in the MS, MD, MT, and RAND algorithms. Because these algorithms are single-minded and consider only one-off optimization, which is highly influenced by the environment and poorly robust. In summary, it is clear that the compared algorithms are not suitable for UDLSNs, while the PPO-ASM algorithm is remarkably robust and can be adapted to various scenarios.

6.4 Impact of Aircraft Numbers

According to the characteristics of the airborne internet, the density of aircraft changes over time, Sometimes there is a lot of aircraft in airspace, sometimes very few. So Figure 9 and Figure 10 show the effect of the number of aircraft on the performance obtained by the different algorithms.

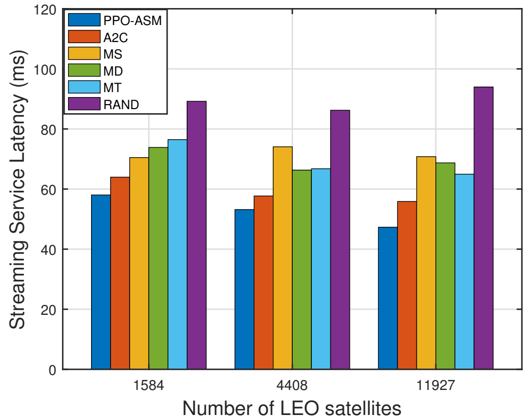


Figure 9. *The average streaming service latency at different aircraft sizes.*

Figure 9 indicates the streaming service latency obtained by different algorithms when the number of aircraft changes from 10 to 100. The PPO-ASM algorithm achieves the shortest streaming service delay among the compared algorithms, and the streaming service delay fluctuates less as the number of aircraft changes, which demonstrates the good adaptability of the PPO-ASM algorithm. As the number of aircraft increases, limited resources, and multi-user selection scheduling complicate access and service management algorithms, as a result, streaming service latency increases. The performance gap between the

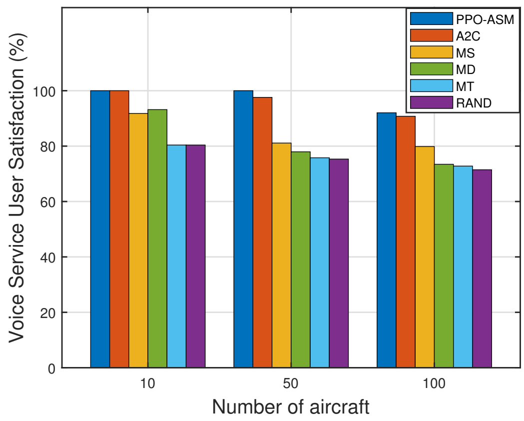


Figure 10. *The average voice service user demand satisfaction at different aircraft sizes.*

A2C algorithm and the PPO-ASM algorithm becomes larger as the number of aircraft increases. It is because the A2C algorithm is prone to massively disruptive update strategies, so when the number of aircraft increases, many aircraft update their relatively poor-performing strategies, resulting in higher latency for streaming services. The MS, MD, MT, and RAND algorithms cause higher latency because they each focus on only one aspect that affects the latency. Thus, the performance obtained by other algorithms is strongly influenced by environmental factors (e.g. aircraft traffic, satellite channel conditions), while the PPO-ASM algorithm has good adaptability.

From Figure 10 we can see that the PPO-ASM algorithm consistently achieves the highest user satisfaction for voice services and can reach 100% when the number of aircraft is 10 and 50. When the number of aircraft increases, the service demand also increases, and the limited network resources may not be able to meet the requirements of all aircraft, as a result, voice service satisfaction decreases. Other algorithms can not satisfy all users in any case and the AC algorithm achieves close to 100% satisfaction only when the number of aircraft is 10. Other algorithms are unable to meet the needs of multi-user voice services because they only consider a single factor.

The PPO-ASM algorithm and A2C algorithm have monotonic characteristics as the scale of satellites increases and the number of aircraft decreases. The reason is that the PPO-ASM algorithm and A2C algorithm always take the metrics related to streaming and voice services as the long-term optimization target to guarantee the service experience of aircraft. Thus, when the scale of LEO satellites increases, more satellites provide more available resources and network capacity, or when the number of aircraft decreases, there is less competition for satellite resources, and aircraft have more possibilities to get a better service experience. The RAND algorithm has strong randomness, so it does not have monotonic properties. The MS, MD, and MT algorithms are all single metric algorithms, which all consider only one performance metric. However, the factors affecting the performance of streaming and voice services are multidimensional, so they are susceptible to performance fluctuations due to environmental factors and do not have monotonic characteristics.

VII. CONCLUSION

In this paper, we have investigated a lightweight airborne Internet architecture leveraging the ultra-dense LEO satellite networks for the global Internet access of civil aviation, and a satellite-aircraft access and data transmission problem has been formulated considering two types of different civil aviation services. We have proposed a PPO-based access and service management algorithm to meet the dynamic and diverse requirements of the airborne Internet. We have designed a PPO-ASM algorithm to solve the formulated problem to guarantee effective real-time network management. We have verified the good performance of the PPO-ASM algorithm in terms of communication latency and user satisfaction, which would be effectively applied in the UDLSN scenario. We believe that this study will shed light on the satellite Internet access for aircraft passengers and would provide a method reference for the future deployment of UDLSN. In future work, we will design the satellite-aircraft access mechanism guaranteeing the quality of experience of access users.

ACKNOWLEDGEMENT

This work was supported in part by the National Key R&D Program of China under Grant 2020YFB1806104, in part by Innovation and Entrepreneurship of Jiangsu Province High-level Talent Program, in part by Natural Sciences and Engineering Research Council of Canada (NSERC). The authors also would like to thank the support from Huawei.

References

[1] M. Price, "Current and emerging trends in the aerospace sector," *Atkins, London, UK*, 2018.
[2] Z. Na, Y. Liu, *et al.*, "Uav-supported clustered noma for 6g-enabled internet of things: Trajectory planning and resource allocation," *IEEE Internet of Things Journal*, 2021, vol. 8, no. 20, pp. 15 041–15 048.
[3] Q. Chen, W. Meng, *et al.*, "Service-oriented fair resource allocation and auction for civil aircrafts augmented space-air-ground integrated networks," *IEEE Transactions on Vehicular Technology*, 2020, vol. 69, no. 11, pp. 13 658–13 672.
[4] J. Liu, Y. Shi, *et al.*, "Space-air-ground integrated network: A survey," *IEEE Communications Surveys & Tutorials*, 2018, vol. 20, no. 4, pp. 2714–2741.
[5] L. Abouzaid, E. Sabir, *et al.*, "The meshing of the sky: Delivering ubiquitous connectivity to ground internet of things," *IEEE Internet of Things Journal*, 2021, vol. 8, no. 5, pp. 3743–3757.
[6] N. Cheng, W. Quan, *et al.*, "A comprehensive simulation platform for space-air-ground integrated network," *IEEE Wireless Communications*, 2020, vol. 27, no. 1, pp. 178–185.
[7] M. Meng, B. Hu, *et al.*, "Cooperative user-scheduling and resource allocation optimization for intelligent reflecting surface enhanced leo satellite communication," *China Communications*, 2024, vol. 21, no. 2, pp. 227–244.
[8] C. Lei, W. Feng, *et al.*, "Secrecy rate maximization for 6g cognitive satellite-uav networks," *China Communications*, 2023, vol. 20, no. 1, pp. 246–260.
[9] H. Wu, J. Chen, *et al.*, "Resource management in space-air-ground integrated vehicular networks: Sdn control and ai algorithm design," *IEEE Wireless Communications*, 2020, vol. 27, no. 6, pp. 52–60.
[10] J. Yu, X. Liu, *et al.*, "3d channel tracking for uav-satellite communications in space-air-ground integrated networks," *IEEE Journal on Selected Areas in Communications*, 2020, vol. 38, no. 12, pp. 2810–2823.
[11] M. Albulet, "Spacex non-geostationary satellite system," *FCC Application SATLOA2016111500118*, 2016.
[12] E. Association, *et al.*, "Esoa satellite action plan for 5g standards," *White Paper*, 2021.
[13] T. Ma, H. Zhou, *et al.*, "Uav-leo integrated backbone: A ubiquitous data collection approach for b5g internet of remote things networks," *IEEE Journal on Selected Areas in Communications*, 2021, vol. 39, no. 11, pp. 3491–3505.
[14] X. Cao, P. Yang, *et al.*, "Airborne communication networks: A survey," *IEEE Journal on Selected Areas in Communications*, 2018, vol. 36, no. 9, pp. 1907–1926.
[15] T. Darwish, G. Kurt, *et al.*, "Location management in internet protocol-based future leo satellite networks: A review," *IEEE Open Journal of the Communications Society*, 2022, vol. 3, pp. 1035–1062.
[16] J. Chen, H. Zhou, *et al.*, "Service-oriented dynamic connection management for software-defined internet of vehicles," *IEEE Transactions on Intelligent Transportation Systems*, 2017, vol. 18, no. 10, pp. 2826–2837.
[17] Z. Lin, M. Lin, *et al.*, "Supporting iot with rate-splitting multiple access in satellite and aerial-integrated networks," *IEEE Internet of Things Journal*, 2021, vol. 8, no. 14, pp. 11 123–11 134.
[18] H. Jia, C. Jiang, *et al.*, "Adaptive access control and resource allocation for random access in ngso satellite networks," *IEEE Transactions on Network Science and Engineering*, 2022, vol. 9, no. 4, pp. 2721–2733.
[19] Y. Zhang, H. Zhang, *et al.*, "Resource allocation in terrestrial-satellite-based next generation multiple access networks with interference cooperation," *IEEE Journal on Selected Areas in Communications*, 2022, vol. 40, no. 4, pp. 1210–1221.
[20] D. Wang, Y. Wang, *et al.*, "On delay-aware resource control with statistical qos provisioning by dual connectivity in heterogeneous aeronautical network," *IEEE Transactions on Vehicular Technology*, 2020, vol. 69, no. 3, pp. 2915–2927.
[21] J. Kokkoniemi, J. Jornet, *et al.*, "Channel modeling and performance analysis of airplane-satellite terahertz band communications," *IEEE Transactions on Vehicular Technology*, 2021, vol. 70, no. 3, pp. 2047–2061.
[22] R. Deng, B. Di, *et al.*, "Ultra-dense leo satellite constellations: How many leo satellites do we need?" *IEEE Transactions on Wireless Communications*, 2021, vol. 20, no. 8, pp. 4843–4857.
[23] H. Jiang, H. Wang, *et al.*, "Dynamic user association in scalable ultra-dense leo satellite networks," *IEEE Transactions on Vehicular Technology*, 2022, vol. 71, no. 8, pp. 8891–8905.
[24] J. Li, K. Xue, *et al.*, "A user-centric handover scheme for ultra-dense leo satellite networks," *IEEE Wireless Communications Letters*, 2020, vol. 9, no. 11, pp. 1904–1908.
[25] L. Yan, X. Ding, *et al.*, "Dynamic channel allocation aided random access for sdn-enabled leo satellite iot," *Journal of Communications and Information Networks*, 2021, vol. 6, no. 2, pp. 134–141.
[26] Y. Li, J. Wu, *et al.*, "A novel two-layered optical satellite network of leo/meo with zero phase factor," *Science China Information Sciences*, 2010, vol. 53, no. 6, pp. 1261–1276.
[27] Y. Cao, S. Lien, *et al.*, "Deep reinforcement learning for multi-user access control in non-terrestrial networks," *IEEE Transactions on Communications*, 2021, vol. 69, no. 3, pp. 1605–1619.
[28] C. Qiu, H. Yao, *et al.*, "Deep q-learning aided networking, caching, and computing resources allocation in software-defined satellite-terrestrial networks," *IEEE Transactions on Vehicular Technology*, 2019, vol. 68, no. 6, pp. 5871–5883.
[29] Y. He, B. Sheng, *et al.*, "Multi-objective deep reinforcement learning based time-frequency resource allocation for multi-beam satellite communications," *China Communications*, 2022, vol. 19, no. 1, pp. 77–91.
[30] D. Zhou, M. Sheng, *et al.*, "Aerospace integrated networks innovation for empowering 6g: A survey and future challenges," *IEEE Communications Surveys and Tutorials*, 2023, p. 1.
[31] T. Ma, B. Qian, *et al.*, "Satellite-terrestrial integrated 6g: An ultra-dense leo networking management architecture," *IEEE Wireless Communications*, 2022.
[32] M. Marvi, A. Aijaz, *et al.*, "Integrating stochastic geometry

and on/off traffic models: Toward spatio-temporal analysis of wireless networks with heterogeneous services," *IEEE Transactions on Network Science and Engineering*, 2022, vol. 9, no. 3, pp. 1668–1679.

[33] J. So, "Adaptive traffic prediction based access control in wireless cdma systems supporting integrated voice/-data/video services," *IEEE Communications Letters*, 2004, vol. 8, no. 12, pp. 703–705.

[34] B. Di, H. Zhang, *et al.*, "Ultra-dense leo: Integrating terrestrial-satellite networks into 5g and beyond for data offloading," *IEEE Transactions on Wireless Communications*, 2019, vol. 18, no. 1, pp. 47–62.

[35] R. Swaminathan, S. Sharma, *et al.*, "Haps-based relaying for integrated space–air–ground networks with hybrid fso/rf communication: A performance analysis," *IEEE Transactions on Aerospace and Electronic Systems*, 2021, vol. 57, no. 3, pp. 1581–1599.

[36] J. Schulman, F. Wolski, *et al.*, "Proximal policy optimization algorithms," 2017, p. arXiv:1707.06347.

[37] M. Qin, Q. Yang, *et al.*, "Machine learning aided context-aware self-healing management for ultra dense networks with qos provisions," *IEEE Transactions on Vehicular Technology*, 2018, vol. 67, no. 12, pp. 12 339–12 351.

[38] A. Asghari, M. Sohrabi, *et al.*, "Task scheduling, resource provisioning, and load balancing on scientific workflows using parallel sarsa reinforcement learning agents and genetic algorithm," *The Journal of Supercomputing*, 2021, vol. 77, no. 3, pp. 2800–2828.

[39] J. Schulman, S. Levine, *et al.*, "Trust region policy optimization," in *International Conference on Machine Learning*, 2015, pp. 1889–1897.

[40] Y. Chen, J. Sun, *et al.*, "Hybrid n-inception-lstm-based aircraft coordinate prediction method for secure air traffic," *IEEE Transactions on Intelligent Transportation Systems*, 2022, vol. 23, no. 3, pp. 2773–2783.

[41] M. Babaeizadeh, I. Frosio, *et al.*, "Reinforcement learning through asynchronous advantage actor-critic on a gpu," *arXiv e-prints*, 2016, p. arXiv:1611.06256.

Biographies

Wang Yilei received the B.S. degree in electronic information science and technology from Central South University (CSU) in 2021. She is currently working towards the doctor degree in School of Electronic Science and Engineering, Nanjing University. Her current research is space-air-ground integrated network, access and handover management.

Ma Ting received the B.S., M.S., and Ph.D degrees in statistics from Sichuan University, Chengdu, China, in 2013, 2016, and 2020, respectively. She is currently a Post-Doctoral Fellow with the School of Electronic Science and Engineering, Nanjing University, Nanjing, China. Her main current research interests include robust hypothesis testing, space-air-ground integrated networks, convex optimization theory, and game theory.

Liu Xiaoyu received the B.S. degree in electronic and information engineering from Xidian University, Xi'an, China, in 2021. He is currently pursuing the M.S. degree with the School of Electronic Science and Engineering, Nanjing University, China. His research interests include space-air-ground integrated networks, network simulation and deep reinforcement learning.

Gao Zhuxuan received the M.S. degree in electronics and communication engineering from Nanjing University in 2022. Her research interest is satellite network access and resource allocation.

Zhou Haibo received the Ph.D degree in information and communication engineering from Shanghai Jiao Tong University, Shanghai, China, in 2014. He is currently an Professor with the School of Electronic Science and Engineering, Nanjing University, Nanjing, China. His research interests include resource management and protocol design in vehicular ad hoc networks, cognitive networks, and space-air-ground integrated networks.

Shen Xuemin received the Ph.D degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 1990. He is currently a University Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His current research interests include network resource management, wireless network security, the Internet of Things, 5G and beyond, and vehicular ad-hoc and sensor networks.