Secure Federated Learning for Multi-UAV Networks: A Framework Based on Cooperative Beamforming and Participant Selection

Xiujuan Zhang[1], Zhenyu Zheng[1], Yu Du[2,3], Xin Fan[2,3,*], Jin Qian[4], Chuanwen Luo[2,3,*]

{ xiujuanzhang@qfnu.edu.cn, zhenyu.zheng@qfnu.edu.cn,
nagasakiianno@bjfu.edu.cn, fanxin@bjfu.edu.cn,
qianjin@tzu.edu.cn, chuanwenluo@bjfu.edu.cn }

[1]School of Computer Science, Qufu Normal University, Rizhao 276826, China

[2]School of Information Science and Technology, Beijing Forestry University, Beijing 100083, China

[3]Hebei Key Laboratory of Smart National Park, Beijing 100083, China

[4]College of Information Engineering, Taizhou University, Taizhou 225300, China

[*]Corresponding authors

Abstract. This paper addresses the severe eavesdropping threat in UAV-assisted Federated Learning (FL) networks. We propose a cooperative secure framework that utilizes a novel dynamic participant selection mechanism to partition the UAV swarm into learning and jamming groups. The jamming group employs Cooperative Beamforming (CB) to actively suppress eavesdroppers, while the learning group is optimized to balance data quality and system cost. We formulate a joint optimization problem aiming to maximize the secrecy rate while minimizing system latency and energy consumption. To solve this non-convex problem with a high-dimensional mixed-action space, we propose a Deep Reinforcement Learning (DRL) algorithm named Graph Attention Multi-Head Actor-Soft Actor-Critic (GAMA-SAC). Extensive simulations demonstrate that the proposed scheme significantly enhances communication security and accelerates model convergence compared to baseline approaches.

Keywords: Federated Learning, Unmanned Aerial Vehicle (UAV), CB, Physical Layer Security, Artificial Noise

1 Introduction

1.1 Background and Related Work

With the explosive growth of the Internet of Things (IoT), Unmanned Aerial Vehicles (UAVs) are increasingly recognized as ideal aerial mobile edge computing platforms or data aggregation nodes. This is due to their exceptional maneuverability, on-demand deployment flexibility, and

IIKI 2025, December 18-21, Chengdu, People's Republic of China
Copyright © 2026 EAI
DOI 10.4108/eai.18-12-2025.2365263

capability to establish reliable Line-of-Sight (LoS) communication links in complex environments [1–3]. To address the bandwidth and privacy concerns of traditional centralized machine learning, integrating Federated Learning (FL) into UAV networks has become a promising paradigm [4–6]. By keeping raw data local and only exchanging model parameters, UAV-assisted FL significantly mitigates the risks of data privacy breaches while reducing communication overhead. However, the practical deployment of such systems faces a series of key challenges that need to be addressed, specifically in balancing communication security, system cost, and model accuracy.

In terms of communication security, although FL provides inherent privacy protection at the data content level, the model parameters exchanged over the air still face the risk of being eavesdropped or reconstructed [7, 8]. To counter this, cryptographic methods and differential privacy are often employed [9, 10], but they inevitably introduce high computational complexity. Alternatively, physical layer security (PLS) techniques[11–13], such as artificial noise (AN) injection, have been explored. While [14, 15] utilized the UAV as a jammer, the limited power and jamming capability of a single UAV make it difficult to cope with complex or colluding eavesdropping environments.

In terms of system cost, which encompasses energy consumption and latency, the constraints are particularly stringent for battery-powered UAVs and IoT devices. Extensive research has been conducted on the joint scheduling of communication and computation resources to minimize costs [16, 17]. For mobile FL architectures, [18] proposed an energy-efficient scheme balancing local computation and wireless transmission. However, most existing works focus on transmission and computing energy, often overlooking the additional energy overhead required to maintain security (e.g., jamming power) under eavesdropping threats.

Regarding the assurance of model accuracy, wireless channel fading and background noise significantly impact FL convergence. Scholars have extensively investigated the convergence bounds of FL over noisy channels [19–21]. Furthermore, the frequency of global aggregation is a critical parameter; [22] demonstrated that adaptively adjusting the aggregation frequency can enhance final model accuracy by effectively balancing the utilization of local updates against the noise accumulation from frequent transmissions.

In summary, most existing studies typically focus on only one or two of these challenges in isolation. There is a lack of research that systematically balances the intrinsic trade-offs among system cost, model accuracy, and communication security within a unified framework, which is essential for the comprehensive optimization of future UAV-assisted FL networks.

1.2 Contributions

To address the above challenges, this paper proposes a multi-UAV cooperative secure FL framework. The core of this framework lies in a novel UAV selection mechanism that dynamically partitions the UAV swarm into a learning UAV group and a jamming UAV group. The learning UAVs are responsible for performing local training and parameter uploading, while the jamming UAV group provides physical layer security for the learning process by using cooperative beamforming (CB) [23, 24] technology to precisely focus AN on the eavesdropper. To achieve the optimal trade-off among communication security, system cost, and FL performance within this framework, we formulate this problem as a complex joint optimization problem and design a deep reinforcement learning

algorithm named Graph Attention Multi-Head Actor-Soft Actor-Critic (GAMA-SAC) to seek the optimal solution. The main contributions of this study are as follows:

- We propose a multi-UAV secure FL framework that integrates dynamic role partitioning, UAV selection, and CB. The UAV selection mechanism in this framework not only enhances communication security through CB but also aims to improve FL efficiency and reduce overall system overhead by intelligently balancing learning gains and costs.
- To address the non-convexity of the problem and its high-dimensional mixed-action space, we design the GAMA-SAC deep reinforcement learning algorithm. This algorithm utilizes a Graph Attention Network (GAT) to effectively capture the dynamic spatial cooperative relationships among UAVs and employs a multi-head actor network to directly handle the complex mixed-action space.
- Extensive simulation experiments verify the convergence and superiority of the algorithm.

2 System Model

2.1 Network Model

We consider a UAV-assisted secure FL system consisting of a ground Base Station (BS), a mobile Eavesdropper (Eve), and a swarm of N UAVs. All entities operate in a 3D Cartesian coordinate system. In each FL round i, the UAV swarm is dynamically partitioned into two disjoint sets: a learning group $\mathscr{S}_L$ with Q UAVs, and a jamming group $\mathscr{S}_J$ with $N-Q$ UAVs. The learning UAVs perform local training and upload model parameters to the BS, while the jamming UAVs transmit Artificial Noise (AN) via Cooperative Beamforming (CB) to disrupt Eve.

Let p_{BS}, p_E^i, $p_k^{L,i}$, and $p_j^{J,i}$ denote the 3D coordinates of the BS, Eve, the k-th learner ($k \in \mathscr{S}_L$), and the j-th jammer ($j \in \mathscr{S}_J$) in the i-th round, respectively. The BS is fixed at $z = 0$. We assume Eve's position p_E^i is static within each time slot but varies across slots, and is known to the BS for optimization purposes.

2.2 UAV Selection Mechanism

To balance FL efficiency, system cost, and cooperative jamming performance, the BS selects Q optimal UAVs as learners ($\mathscr{S}_L$) based on a comprehensive score S_k. This score integrates three metrics: 1) *Learning Gain (G_k)*, defined as the average loss on the local dataset to quantify data value; 2) *Learning Cost (C_k)*, representing the weighted sum of computation/communication latency and energy consumption; and 3) *Dispersion Index (I_k)*, the average distance from u_k to other UAVs. A smaller I_k indicates a central position ideal for forming a compact jamming array (reducing mobility energy for jammers), thus UAVs with larger I_k are preferred for learning. The total score for UAV k is defined as:

$$S_k = \alpha G_k - \beta C_k + \gamma I_k, \tag{1}$$

where α, β, γ are weighting coefficients. The Q UAVs with the highest S_k form $\mathscr{S}_L$, while the remaining become jammers $\mathscr{S}_J$.

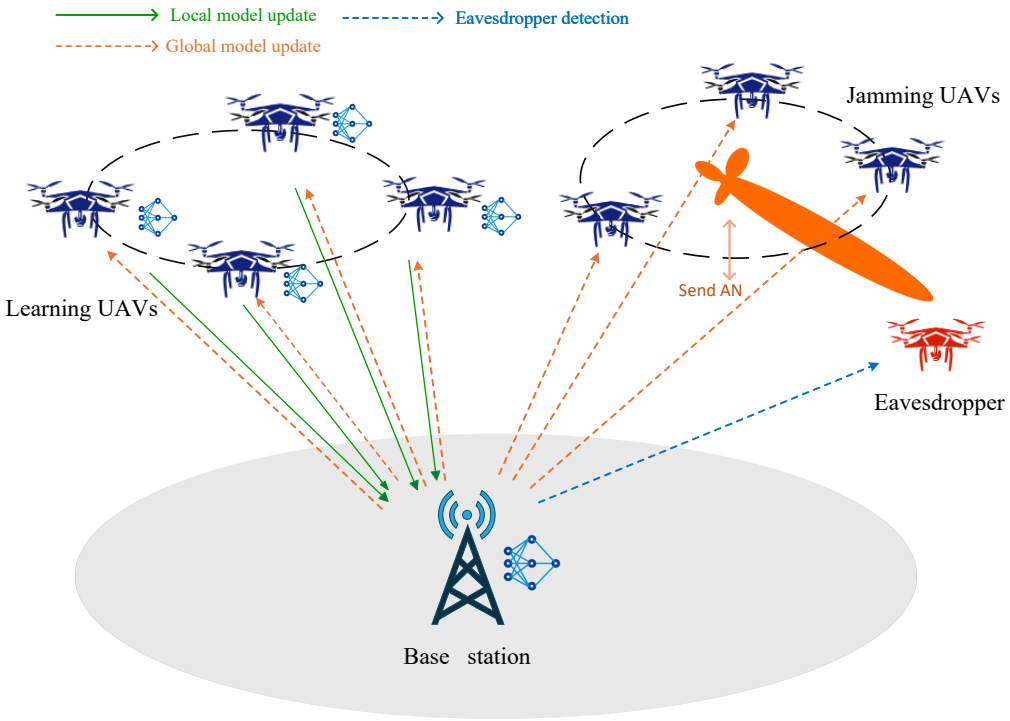

Fig. 1. The system model diagram.

2.3 Communication and Security Model

We adopt a probabilistic Air-to-Ground (A2G) channel model for the links between learning UAVs and the BS, considering both Line-of-Sight (LoS) and Non-Line-of-Sight (NLoS) propagation probabilities dependent on the elevation angle. The average path loss from learner u_k to the BS is denoted as $\overline{PL}_{k,BS}$. The Air-to-Air (A2A) channels (UAV-to-UAV and UAV-to-Eve) are modeled as LoS links.

2.3.1 SINR Analysis

The BS receives the model update from u_k with transmission power P_k. The received Signal-to-Interference-plus-Noise Ratio (SINR) at the BS is:

$$\text{SINR}^i_{k,BS} = \frac{P_k / \overline{PL}^i_{k,BS}}{\sigma^2}, \tag{2}$$

where σ^2 is the noise power. Simultaneously, the jamming group $\mathscr{S}_J$ transmits AN using CB to disrupt Eve. The interference power at Eve is $P^i_{AN,E} = \frac{P^i_{jam} G^i_{J,E}}{PL^j_{J,E}}$, where P^i_{jam} is the total jamming power determined by excitation weights $\{I_j\}$, and $G^i_{J,E}$ is the array beamforming gain in Eve's direction. Consequently, the SINR at Eve intercepting u_k is:

$$\mathrm{SINR}^i_{k,E} = \frac{P_k / PL^i_{k,E}}{P^i_{AN,E} + \sigma^2_E}. \tag{3}$$

2.3.2 Secrecy Rate

The achievable secrecy rate for UAV u_k is the difference between the capacity of the legitimate channel and the wiretap channel:

$$R^i_{sec,k} = \left[B\log_2(1+\mathrm{SINR}^i_{k,BS}) - B\log_2(1+\mathrm{SINR}^i_{k,E}) \right]^+, \tag{4}$$

where B is the bandwidth and $[x]^+ = \max(0,x)$.

2.4 Latency Model

2.4.1 Local Training Latency

This latency is the time required for UAV u_k to complete τ epochs of local training. It depends on the size of the UAV's local dataset $|D_k|$, the number of CPU cycles required to train each data sample c_k, the number of local training epochs τ, and the CPU frequency f^i_k allocated by the UAV for the computation task. This latency is calculated as

$$T^i_{k,\mathrm{train}} = \frac{c_k |D_k| \tau}{f^i_k}. \tag{5}$$

2.4.2 Communication Latency

This latency is the time required for the learning UAV u_k to upload its trained model parameters to the BS. It depends on the data size of the model parameters to be transmitted, Ω, and the uplink data transmission rate of the UAV. Therefore, the communication latency is given by

$$T^i_{k,\mathrm{comm}} = \frac{\Omega}{B\log_2(1+\mathrm{SINR}^i_{k,\mathrm{BS}})}. \tag{6}$$

2.4.3 Total Latency

For the FL system, the time required to complete one round of global iteration is a key performance indicator of its efficiency. In our framework, since all learning UAVs perform their local training and model uploading tasks in parallel, the total latency for each round, T^i_{round}, depends on

the last learning UAV to complete its task. Therefore, the total latency for the i-th round is defined as the maximum of the task completion times among all learning UAVs $u_k \in \mathscr{S}_L$, expressed as

$$T^i_{\text{round}} = \max_{k \in \mathscr{S}_L} \{T^i_{k,\text{train}} + T^i_{k,\text{comm}}\}. \tag{7}$$

2.5 Energy Consumption Model

The energy consumption consists of two parts: the learning group and the jamming group. For any learner $u_k \in \mathscr{S}_L$, the energy consumption $E^i_{k,L}$ includes local computation, wireless transmission, and hovering energy to maintain position:

$$E^i_{k,L} = P_{k,comp} T^i_{k,train} + P_k T^i_{k,comm} + P_{hover} T^i_{round}, \tag{8}$$

where $P_{k,comp}$ and P_{hover} are the computation and hovering power, respectively. For any jammer $u_j \in \mathscr{S}_J$, the energy consumption $E^i_{j,J}$ comprises the jamming signal transmission and the propulsion energy required to move to the optimal jamming position $p^{opt,i}_j$:

$$E^i_{j,J} = |I^i_j|^2 P_{max} T^i_{round} + E^i_{j,prop}. \tag{9}$$

Here, $E^i_{j,prop}$ is calculated based on the standard rotary-wing UAV power consumption model $P(v)$. Specifically, $E^i_{j,prop} = \int_0^{T_{move}} P(v(t))dt + m_j g \Delta h$, accounting for both aerodynamic power and potential energy changes. The total system energy consumption for round i is the sum of all UAVs' energy:

$$E^i_{total} = \sum_{k \in \mathscr{S}_L} E^i_{k,L} + \sum_{j \in \mathscr{S}_J} E^i_{j,J}. \tag{10}$$

2.6 Problem Formulation

Our goal is to find a balance among the conflicting objectives of energy consumption, latency, security, and FL accuracy. This requires us to jointly optimize the transmit power $\{P_k\}$ and local training epochs τ of the learning UAVs, as well as the target positions $\{\mathbf{p}^{\text{opt}}_j\}$ and jamming transmit weights $\{I_j\}$ of the jamming UAVs. We first define a comprehensive system cost C^i_{sys}, which consists of the weighted sum of the total latency and total energy consumption per round, as

$$C^i_{\text{sys}} = \omega_T \cdot T^i_{\text{round}} + \omega_E \cdot E^i_{\text{total}}, \tag{11}$$

where ω_T and ω_E are the weight coefficients for time and energy costs, respectively. We formulate this multi-objective joint optimization problem optimization problem (P1) as follows:

$$\text{(P1)} \quad \min_{\mathscr{Z}} \quad \omega_1 C^i_{\text{sys}} - \omega_2 G_{\text{acc}} - \omega_3 \mathbb{E}[R^i_{\text{sec},k}] \tag{12}$$

$$
\begin{aligned}
\text{s.t.} \quad & R^i_{\text{sec},k} \geq R_{\min}, \quad \forall u_k \in \mathscr{S}_L, && \text{(12a)}\\
& 0 \leq P_k \leq P_t^{\max}, \quad \forall u_k \in \mathscr{S}_L, && \text{(12b)}\\
& \tau_{\min} \leq \tau \leq \tau_{\max}, \quad \tau \in \mathbb{Z}^+, && \text{(12c)}\\
& T^i_{j,\text{move}} \leq T^i_{\text{round}}, \quad \forall u_j \in \mathscr{S}_J, && \text{(12d)}\\
& \mathbf{p}_j^{\text{opt},i} \in \mathscr{S}_{\text{space}}, \quad \forall u_j \in \mathscr{S}_J, && \text{(12e)}\\
& E^i_{L,\text{total}} \leq E^i_{k,\text{avail}}, \quad \forall u_k \in \mathscr{S}_L, && \text{(12f)}\\
& E^i_{J,\text{total}} \leq E^i_{j,\text{avail}}, \quad \forall u_j \in \mathscr{S}_J, && \text{(12g)}\\
& \|\mathbf{p}^i_a - \mathbf{p}^i_b\| \geq d_{\text{safe}}, \quad \forall a \neq b, && \text{(12h)}\\
& G_{acc}(\{P_k\}, \tau) \geq \varepsilon_{acc}, && \text{(12i)}
\end{aligned}
$$

where $\mathscr{X} = \{\{P_k\}, \tau, \{\mathbf{p}_j^{\text{opt}}\}, \{I_j\}\}$ is the set of optimization variables. The objective function (12) aims to minimize a comprehensive cost composed of three weighted parts: C^i_{sys} is the system cost (including energy and latency); G_{acc} represents the FL accuracy; and $\mathbb{E}[R^i_{\text{sec},k}]$ is the average secrecy rate of the system. $\omega_1, \omega_2, \omega_3$ are the weight coefficients for each term.

Constraint (12a) ensures the communication security of all learning UAVs. Constraints (12b) and (12c) limit the ranges for transmit power and local training epochs, respectively. Constraint (12d) ensures synchronization between the jamming and learning tasks, where the jamming UAVs' movement time $T^i_{j,\text{move}}$ is determined by their maximum horizontal and vertical speeds. Constraint (12e) stipulates that the UAVs must operate within a predefined safe geographical space $\mathscr{S}_{\text{space}}$. Constraints (12f) and (12g) ensure that the task energy consumption for learning and jamming UAVs does not exceed their currently available battery energy, respectively. Finally, (12h) is the safety distance constraint to avoid collisions between UAVs. Constraint (12i), which requires the accuracy determined by the optimization variables to be above a preset threshold ε_{acc}, thereby guaranteeing the convergence performance of FL. Since the objective function is non-convex, the variables are highly coupled, and the action space is a mix of high-dimensional continuous and discrete variables, traditional optimization methods struggle to solve it efficiently. To this end, we will introduce an advanced deep reinforcement learning algorithm in the next section to address these challenges.

3 Algorithm Design

To efficiently solve problem (P1), we model it as a Markov Decision Process (MDP) and design an algorithm named GAMA-SAC to solve it.

3.1 Markov Decision Process Modeling

We set the BS as the central agent. The core elements of the MDP, $\langle \mathscr{S}, \mathscr{A}, \mathscr{R} \rangle$, are defined as follows.

3.1.1 State Space ($\mathscr{S}$)

At the beginning of the i-th round, the state $s_i \in \mathscr{S}$ observed by the agent is a high-dimensional vector designed to comprehensively describe the system's situation. This state vector s_i consists of the following parts:

$$\begin{aligned} s_i = \{ & \{\mathbf{p}_k^i\}_{k=1}^N, \{E_k^i\}_{k=1}^N, \mathbf{p}_E^i, \mathbf{M}^i, \\ & \{\mathbf{p}_{k,\text{rel}}^i\}_{k=1}^N, \{d_{k,\text{rel}}^i\}_{k=1}^N, \mathbf{p}_{C_J}^i, B_{\text{rem}}^i \}, \end{aligned} \tag{13}$$

where $\{\mathbf{p}_k^i\}$ are the 3D positions of all UAVs; $\{E_k^i\}$ is their remaining energy; $\mathbf{p}_E^i$ is the eavesdropper's position; $\mathbf{M}^i$ is a mask vector identifying the UAVs' roles; $\{\mathbf{p}_{k,\text{rel}}^i\}$ and $\{d_{k,\text{rel}}^i\}$ are the relative position vectors and distances of the UAVs to the eavesdropper, respectively; $\mathbf{p}_{C_J}^i$ is the geometric center of the jamming swarm; and B_{rem}^i is the total energy budget for the current episode.

3.1.2 Action Space ($\mathscr{A}$)

The agent makes decisions in a mixed-action space. The action $a_i \in \mathscr{A}$ selected in state s_i consists of a discrete part and multiple continuous parts:

$$a_i = \{\tau^i, \{P_k^i\}_{k\in\mathscr{S}_L}, \{I_j^i\}_{j\in\mathscr{S}_J}, \{\mathbf{p}_j^{\text{opt},i}\}_{j\in\mathscr{S}_J}\}, \tag{14}$$

where $\tau^i \in \{1, ..., \tau_{\max}\}$ is the discrete number of local training epochs; $\{P_k^i\}$ is the continuous vector of learner transmission powers; $\{I_j^i\}$ is the continuous vector of jammer excitation weights; and $\{\mathbf{p}_j^{\text{opt},i}\}$ is the continuous vector of jammer target positions.

3.1.3 Reward Function ($\mathscr{R}$)

We define the immediate reward $R(s_i, a_i)$ as the negative of the total cost C_{total}^i incurred in the i-th time slot, given by

$$R(s_i, a_i) = -C_{\text{total}}^i. \tag{15}$$

The total cost C_{total}^i is a weighted sum that incorporates all optimization objectives, with its expanded form as follows:

$$\begin{aligned} C_{\text{total}}^i = & \omega_T T_{\text{round}}^i + \omega_E E_{\text{total}}^i + \omega_{\text{acc}} G_{\text{acc}}^i \\ & - \omega_{\text{sec}} \left(\frac{1}{Q} \sum_{k\in\mathscr{S}_L} R_{\text{sec},k}^i \right) + P_{\text{penalty}}^i, \end{aligned} \tag{16}$$

where P_{penalty}^i is a hard constraint penalty term. We introduce a "safety shield" mechanism: if the agent attempts an action that would cause the inter-UAV distance to be less than d_{safe}, the action is blocked, and the agent receives a severe penalty. By maximizing this reward function, the agent is guided to learn an optimal policy that balances communication security, system efficiency, learning performance, and flight safety.

3.2 GAMA-SAC Algorithm Framework

To effectively solve the aforementioned MDP, we designed an agent named GAMA-SAC. This agent employs an advanced neural network architecture aimed at efficiently handling the dynamic topological relationships of the UAV swarm and the high-dimensional mixed-action space. Its core architecture consists of the following three key components.

3.2.1 GNN-based State Encoder

In our problem, the spatial topology of the UAV swarm is critical for decision-making. Standard Multi-Layer Perceptron (MLP) networks are inefficient at processing such dynamically changing graph-structured data. To this end, we model the UAV swarm at each moment as a dynamic graph $G^i = (\mathscr{V}, \mathscr{E}^i)$, where the node set $\mathscr{V}$ represents all UAVs, and the edge set $\mathscr{E}^i$ is dynamically constructed based on the distances between UAVs. We employ a GAT as the state encoder. Through its self-attention mechanism, GAT can learn the importance weights of different neighboring nodes relative to a central node, thereby generating a highly informative feature embedding vector $\mathbf{h}_k$ for each UAV k, which contains both its local cooperative relationships and global position information. This encoding method effectively captures the spatial cooperative relationships between UAVs, providing a richer representation for subsequent decision-making.

3.2.2 Multi-Head Actor Network

Our MDP features a high-dimensional mixed-action space, which includes both discrete actions and multiple sets of high-dimensional continuous actions. To handle this complex action structure directly and avoid information loss caused by "flattening" operations, our Actor network adopts a multi-head output architecture. After being processed by the GNN encoder and shared MLP layers, the UAV feature embeddings are fed into different "Decision Heads." The Global Decision Head uses aggregated information from all UAV features to decide on global parameters, i.e., the discrete local training epochs τ. The Role-Specific Decision Heads separate the embeddings of the learning UAVs $\{\mathbf{h}_k | k \in \mathscr{S}_L\}$ and the jamming UAVs $\{\mathbf{h}_j | j \in \mathscr{S}_J\}$, feeding them into their respective decision heads to output their exclusive actions in parallel.

3.2.3 Soft Actor-Critic Training Framework

To improve the stability and robustness of the policy, we use the Soft Actor-Critic (SAC) framework to train the GAMA network. SAC is a DRL algorithm based on the maximum entropy (Maximum Entropy) principle. Its objective function maximizes the cumulative reward while also maximizing the policy's entropy $\mathscr{H}$, is given by

$$J(\pi) = \sum_{i=0}^{T} \mathbb{E}_{(s_i,a_i)\sim\pi} \left[R(s_i,a_i) + \alpha \mathscr{H}(\pi(\cdot|s_i)) \right], \tag{17}$$

where α is the temperature coefficient, used to adjust the strength of the entropy regularization. The entropy regularization term $\alpha\mathscr{H}$ encourages the policy to maintain a certain level of randomness, preventing premature convergence to a narrow local optimum.

Table 1: System Parameters

Parameter	Value	Parameter	Value	Parameter	Value
$P_{k,\text{comp}}$	2.0 W	$v_{\text{horz}}^{\max}$	20.0 m/s	v_{tip}	200.0 m/s
$P_t^{\max}$	1.0 W	$v_{\text{vert}}^{\max}$	5.0 m/s	v_{h0}	15.0 m/s
P_{hover}	150.0 W	m_j	2.0 kg	ρ	1.2 kg/m³
f_c	2.4 GHz	E_{avail}	5e7 J	d_0	0.012
B	1.0 MHz	P_B	120.0 W	s	0.05
σ^2	-114 dBm	P_I	160.0 W	A	0.503 m²

4 Experimental Simulation

4.1 Experiment Setup

Scenario Configuration: Our simulation environment is set in a 3D space, containing one BS, one aerial mobile Eve, and a swarm of 7 UAVs. At the beginning of each round, the UAV swarm is dynamically partitioned into 4 learning UAVs and 3 jamming UAVs.
Parameter Settings: The parameters used in the simulation are listed in Table 1 for brevity.

4.2 Baseline Algorithms

To fully verify the superiority of our proposed GAMA-SAC algorithm, we conduct a comprehensive comparison against the following four representative algorithms: Standard SAC (Std-SAC), Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and Random Decision Making (RDM).

4.3 Results and Discussion

Fig. 2 shows the cumulative reward curves during the training process for our proposed GAMA-SAC algorithm and the standard Std-SAC algorithm. The faint, thin lines in the figure represent the raw evaluation rewards, which exhibit high variance, while the thicker lines represent the smoothed average curves, which more clearly illustrate the learning trend. It is clear from the figure that both curves show a good convergence trend, proving the effectiveness of the DRL framework in solving such complex problems. However, GAMA-SAC's convergence speed is significantly faster than that of Std-SAC, and it ultimately converges to a higher reward level. This preliminarily demonstrates that our designed GAT state encoder and multi-head actor network architecture can more effectively understand the cooperative relationships of the UAV swarm and the complex action space compared to a simple MLP, thereby learning a superior policy.

Fig. 3 provides a comprehensive comparison of the performance of all five algorithms across the three key dimensions of average secrecy rate, total energy consumption, and average latency.

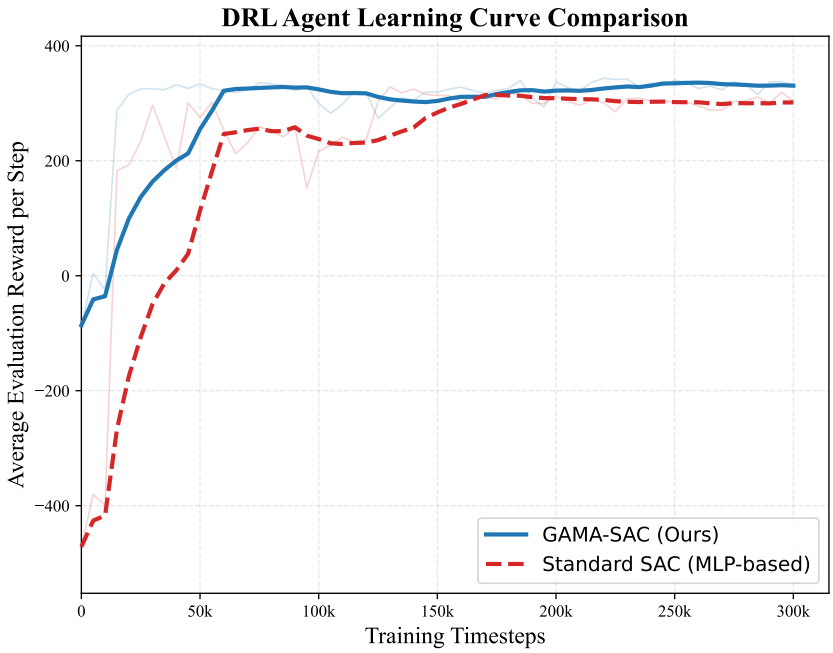

Fig. 2. DRL agent learning curve comparison.

In terms of security performance, the two DRL algorithms, GAMA-SAC and Std-SAC, show overwhelming advantages, achieving average secrecy rates far higher than the heuristic algorithms and the random policy. This indicates that the DRL agent successfully learned how to use precise CB to actively suppress the eavesdropper. To achieve this superior security performance, the DRL algorithms incurred higher energy costs. This reflects our algorithm's ability to make intelligent performance trade-offs, i.e., willingness to spend reasonable and controllable extra energy in exchange for a significant improvement in security performance. In terms of time latency, all algorithms kept the per-round delay at a low level. Notably, the Std-SAC algorithm produced significantly higher latency, which again confirms the inefficiency of its policy.

Fig. 4 compares the FL model accuracy curves over the same number of global communication rounds under the five different algorithm policies. As shown in the figure, the FL model convergence curve under the GAMA-SAC policy consistently leads throughout the entire training process. This is thanks to our algorithm introducing the theory-driven "Accuracy Cost" term into the MDP modeling, enabling the agent to intelligently accelerate FL convergence while optimizing for security and energy consumption.

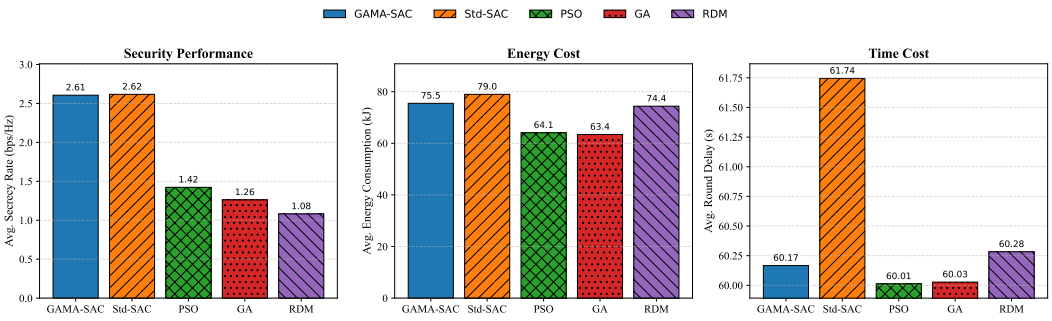

Fig. 3. Comparison of key performance indicators (security, energy cost, latency) for different algorithms.

5 Conclusion

This paper proposes and evaluates a cooperative secure Federated Learning (FL) framework designed for multi-UAV networks, which are inherently vulnerable to eavesdropping attacks. To address these security challenges, we introduce a dynamic participant selection mechanism that intelligently partitions the available UAVs into a learning group dedicated to collaborative model training and a jamming group responsible for emitting artificial noise to confound potential eavesdroppers. Furthermore, we formulate a comprehensive joint optimization problem to balance the conflicting objectives of secrecy rate, system cost, and learning accuracy. Given the inherent complexity and the high-dimensional mixed-action space of this optimization problem, we develop a novel Graph Attention Multi-Head Actor-Soft Actor-Critic (GAMA-SAC) algorithm. By leveraging graph attention networks, this algorithm efficiently captures the dynamic topological relationships among the participating UAVs. Extensive simulation results demonstrate the superiority of the proposed framework over existing baseline approaches. Specifically, the cooperative beamforming strategy effectively suppresses eavesdropping capabilities, while the GAMA-SAC algorithm reduces the system cost and notably accelerates the convergence rate of the federated model. Ultimately, this research provides a robust and scalable solution for secure data sharing in multi-UAV environments.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 62502036, 62572058, U24A20244, and 62472251), the China Postdoctoral Science Foundation (No. 2024M750199), the Outstanding Youth Team of Central Universities (No. QNTD202504), the Young Elite Scientists Sponsorship Program by the China Association for Science and Technology (No. 2023QNRC001), the Taizhou Science and Technology Support Program (Social Development) (No. TSL202519), the General Project for Philosophy and Social Sciences Research in Jiangsu Universities (No. 2025SJYB1697), the Jiangsu Province's "Qinglan Project" (2025): The Program for Middle-aged and Young Academic Leaders, and the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (No. 23KJB510033).

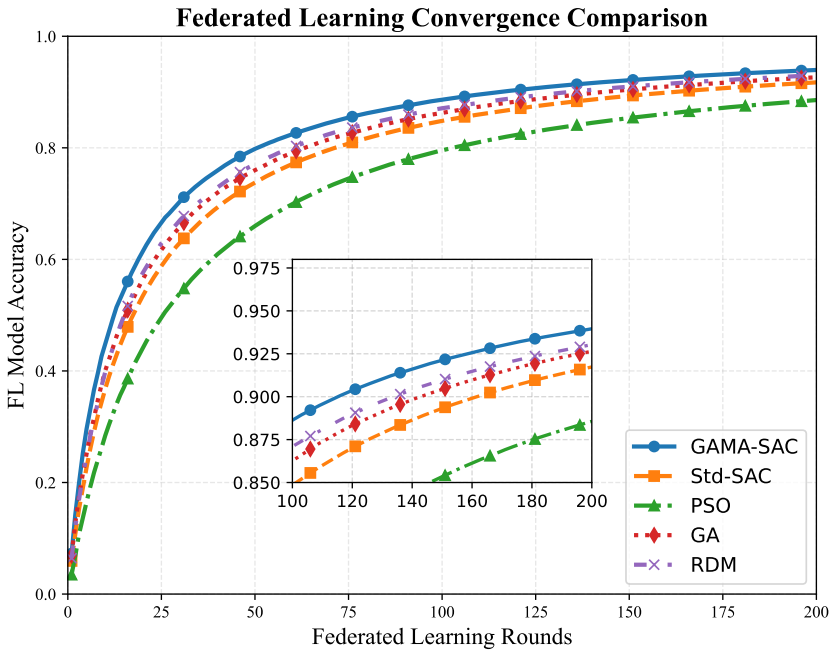

Fig. 4. FL convergence comparison.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] Ning Z, Hu H, Wang X, Guo L, Guo S, Wang G, et al. Mobile edge computing and machine learning in the internet of unmanned aerial vehicles: a survey. ACM Computing Surveys. 2023;56(1):1-31.

[2] Fan X, Li G, Li J, Wang Y, Luo C, Hong Y, et al. A hybrid uplink-downlink secure transmission scheme for UAV-aided coordinated multi-point networks. Digital Communications and Networks. 2025;11(3):925-36.

[3] Fan X, Li G, Luo C, Hong Y, Chen Z, Chen T, et al. UAV-Assisted Heterogeneous Federated Learning over the Air Against Byzantine Attacks. Tsinghua Science and Technology. 2026;31(2):904-19.

[4] Fan X, Wang Y, Huo Y, Tian Z. Joint optimization of communications and federated learning over the air. IEEE Transactions on Wireless Communications. 2021;21(6):4434-49.

[5] Fan X, Wang Y, Huo Y, Tian Z. BEV-SGD: Best effort voting SGD against Byzantine attacks for analog-aggregation-based federated learning over the air. IEEE Internet of Things Journal. 2022;9(19):18946-59.

[6] Fan X, Wang Y, Zhang W, Li Y, Cai Z, Tian Z. GANFed: GAN-Based Federated Learning with Non-IID Datasets in Edge IoTs. In: ICC 2024-IEEE International Conference on Communications. IEEE; 2024. p. 5443-8.

[7] Hu K, Gong S, Zhang Q, Seng C, Xia M, Jiang S. An overview of implementing security and privacy in federated learning. Artificial intelligence review. 2024;57(8):204.

[8] Mothukuri V, Khare P, Parizi RM, Pouriyeh S, Dehghantanha A, Srivastava G. Federated-learning-based anomaly detection for IoT security attacks. IEEE Internet of Things Journal. 2021;9(4):2545-54.

[9] Wei K, Li J, Ding M, Ma C, Yang HH, Farokhi F, et al. Federated learning with differential privacy: Algorithms and performance analysis. IEEE transactions on information forensics and security. 2020;15:3454-69.

[10] Xie Q, Jiang S, Jiang L, Huang Y, Zhao Z, Khan S, et al. Efficiency optimization techniques in privacy-preserving federated learning with homomorphic encryption: A brief survey. IEEE Internet of Things Journal. 2024;11(14):24569-80.

[11] Yan Y, Jing T, Gao Q, Huo Y, Fan X, Wang Y, et al. An Initial Phase-added and RIS-assisted Physical Layer Security Scheme Based on Deep Reinforcement Learning. IEEE Transactions on Vehicular Technology. 2025.

[12] Li H, Li M, Lin Y, Li T, Li R, Fan X. Physical layer secure transmission of AI models in UAV-enabled edge AIoT. Electronics. 2025;14(17):3450.

[13] Yu W, Li J, Fan X, Wang G, Li G, Luo C, et al. 3D physical layer secure transmission for UAV-assisted mobile communications without locations of eavesdroppers. In: International Conference on Wireless Artificial Intelligent Computing Systems and Applications. Springer; 2024. p. 355-66.

[14] Khan WU, Lagunas E, Ali Z, Javed MA, Ahmed M, Chatzinotas S, et al. Opportunities for physical layer security in UAV communication enhanced with intelligent reflective surfaces. IEEE Wireless Communications. 2022;29(6):22-8.

[15] Rao H, Xiao S, Yan S, Wang J, Tang W. Optimal geometric solutions to UAV-enabled covert communications in line-of-sight scenarios. IEEE Transactions on Wireless Communications. 2022;21(12):10633-47.

[16] Hua H, Li Y, Wang T, Dong N, Li W, Cao J. Edge computing with artificial intelligence: A machine learning perspective. ACM Computing Surveys. 2023;55(9):1-35.

[17] Yang Z, Chen M, Saad W, Hong CS, Shikh-Bahaei M. Energy efficient federated learning over wireless communication networks. IEEE Transactions on Wireless Communications. 2020;20(3):1935-49.

[18] Zeng Q, Du Y, Huang K, Leung KK. Energy-efficient resource management for federated edge learning with CPU-GPU heterogeneous computing. IEEE Transactions on Wireless Communications. 2021;20(12):7947-62.

[19] Xu C, Liu S, Yang Z, Huang Y, Wong KK. Learning rate optimization for federated learning exploiting over-the-air computation. IEEE Journal on Selected Areas in Communications. 2021;39(12):3742-56.

[20] Zhao H, Ji F, Li Q, Guan Q, Wang S, Wen M. Federated meta-learning enhanced acoustic radio cooperative framework for ocean of things. IEEE Journal of Selected Topics in Signal Processing. 2022;16(3):474-86.

[21] Fan X, Wang Y, Huo Y, Tian Z. 1-bit compressive sensing for efficient federated learning over the air. IEEE transactions on wireless communications. 2022;22(3):2139-55.

[22] Wang S, Tuor T, Salonidis T, Leung KK, Makaya C, He T, et al. Adaptive federated learning in resource constrained edge computing systems. IEEE journal on selected areas in communications. 2019;37(6):1205-21.

[23] Liu S, Sun G, Li J, Liang S, Wu Q, Wang P, et al. UAV-enabled collaborative beamforming via multi-agent deep reinforcement learning. IEEE Transactions on Mobile Computing. 2024;23(12):13015-32.

[24] Zheng X, Sun G, Li J, Liang S, Wu Q, Yin M, et al. Reliable and energy-efficient communications via collaborative beamforming for UAV networks. IEEE Transactions on Wireless Communications. 2024;23(10):13235-51.