drones

MDPI

Article

Collaborative Beamforming for Secure UAV Swarm Communications: An End-to-End MAPPO-Based Framework Against Mobile Eavesdroppers

Runze Dong [1], Jieyong Zhang [1,*], Buhong Wang [1], Cunqian Feng [2], Jiacai Jiang [3] and Jiwei Tian [4,5]

[1] School of Information and Navigation, Air Force Engineering University, Xi'an 710077, China; drzaxx@buaa.edu.cn (R.D.); wbhgroup@aliyun.com (B.W.)
[2] School of Air Defense and Antimissile, Air Force Engineering University, Xi'an 710100, China; fengcunqian@sina.com
[3] Southwest China Institute of Electronic Technology, Chengdu 610036, China; jiacaijiang@126.com
[4] School of Cyber Science and Engineering, Xi'an Jiaotong University, Xi'an 710049, China; tianjiwei2016@163.com
[5] School of Air Traffic Control and Navigation, Air Force Engineering University, Xi'an 710100, China
* Correspondence: dumu3110728@126.com

Highlights

What are the main findings?

- This paper explores the potential of joint beamforming and trajectory optimization from the perspective of enhancing the security of downlink transmissions from an unmanned aerial vehicle (UAV) swarm to terrestrial friendly nodes.
- This paper proposed a novel optimization framework based on multi-agent proximal optimization (MAPPO) to address the cooperative transmission requirements among UAVs, in addition to resolving the security challenges resulting from the mobility of an eavesdropper.

What are the implications of the main findings?

- The security of transmissions between the UAV swarm and legitimate users is significantly enhanced through optimized collaborative beamforming and cooperative trajectories.
- The proposed MAPPO-based approach outperforms both multi-agent and single-agent reinforcement learning benchmarks in terms of convergence speed, training stability, and total reward, thereby validating its effectiveness.

Abstract

Unmanned aerial vehicles (UAVs) are expected to serve as core nodes for next-generation communication networks, while the broadcast nature of line-of-sight (LoS) links makes the security of transmissions a server problem, which is more prominent for a mobile eavesdropper scenario. In this paper, the security enhancement of UAV swarm communication is considered. Specifically, a UAV swarm with aerial base stations attempts to transmit confidential information to terrestrial nodes, and a mobile eavesdropper lurking nearby tries to approach better receiving points to intercept communications. For the purpose of enhancing the security of transmissions utilizing spatial freedom, a virtual antenna array is formed by the UAV swarm, and a multi-agent proximal policy optimization (MAPPO)-based approach is developed to jointly optimize the collaborative beamforming and cooperative trajectories of the UAV swarm under a maximum power constraint. The simulation results demonstrate the capability of the proposed method to direct the UAV swarm to transmit mission information directionally and validate the superiority of security performance compared to benchmarks.

Check for updates

Academic Editor: Petros S. Bithas

Received: 3 April 2026
Revised: 16 May 2026
Accepted: 19 May 2026
Published: 25 May 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Keywords: unmanned aerial vehicle (UAV); physical-layer security; collaborative beamforming; trajectory optimization; multi-agent proximal policy optimization (MAPPO)

1. Introduction

1.1. Backgrounds

Due to their advantages in terms of high mobility, flexible deployment, and cost-effectiveness, unmanned aerial vehicles (UAVs) are envisioned as a pivotal component of the forthcoming sixth-generation (6G) wireless networks, specifically in providing sensitive coverage and ubiquitous connectivity as aerial base stations [1–4]. While a single UAV has shown potential in enhancing wireless connectivity, the fundamental constraints on size, weight, and energy cannot be neglected.

A solitary UAV usually carries limited energy and a small-scale antenna array, which bottlenecks the transmission range, achievable data rate, and, more critically, the ability to perform sharp beamforming for security purposes. Compared to single UAVs, a UAV swarm offers superior resilience and flexibility by coordinating multiple low-cost nodes, dynamically reconfigures its topology according to the current environment, and maintains service continuity even if individual nodes malfunction [5,6]. As such, the paradigm migration from a single UAV to a UAV swarm is natural for breaking through the aforementioned physical limitations [7].

However, the dominant line-of-sight (LoS) characteristics of air-to-ground (A2G) and air-to-air (A2A) channels, which are typical link properties of UAV-involved networks, while beneficial for transmission quality, inherently result in signal exposure to prospective eavesdroppers due to their broadcast nature. Consequently, enhancing the security of UAV transmissions, particularly from the perspective of the physical layer, has become a critical issue in current research [8,9]. Physical-layer security improves the security of connections via signal processing techniques, such as beamforming, artificial noise (AN), power allocation, etc. This kind of approach eliminates the requirement for key storage and distribution, and it is well-suited to energy-constrained UAV platforms [10,11].

Moreover, given that UAV communication optimization is intrinsically dynamic and sequential, reinforcement learning (RL), especially multi-agent reinforcement learning (MARL), has emerged as a promising paradigm for autonomous decision-making in UAV-oriented networks [12–16].

1.2. Related Works

The integration of UAVs with emerging communication paradigms, such as reconfigurable intelligent surfaces (RISs) and integrated sensing and communication (ISAC), has become an active research focus in recent years [17–20]. Existing studies have demonstrated that UAV mobility can be leveraged to improve communication quality through joint optimization of power allocation, sensing, trajectory planning, and phase control. For example, ref. [17] investigated the joint power and phase design of UAV-mounted RISs for sum spectral efficiency maximization, while ref. [18] considered a joint sensing, beamforming, and trajectory design for UAV base stations. Although these studies provide useful optimization frameworks for UAV-assisted communications, their primary objective lies in throughput or efficiency enhancement rather than adversarial secure transmission. More importantly, they do not explicitly address the challenge that the communication environment may evolve in response to the transmitter's current actions.

Recent research has further extended the focus from single-UAV systems to UAV swarms, motivated by their advantages in spatial diversity, resilience, and cooperative

operation [21–23]. In [21], a MARL-based sensing-communication resource allocation strategy was proposed for improving the total energy efficiency of a UAV swarm. Ref. [22] analyzed the rate performance of UAV swarms from the perspective of the sum rate and max–min rate, while ref. [23] studied UAV swarms as aerial relays in hierarchical federated learning (FL) systems. These works confirm that multi-UAV cooperation can significantly enlarge the available spatial degrees of freedom (DoFs). However, the cooperative gain is mostly exploited for communication efficiency, coverage, or learning performance rather than for secure distributed beamforming against intelligent adversaries. As such, the security value of swarm-level spatial cooperation remains insufficiently explored.

To address the severe vulnerability of UAV communications to eavesdropping, physical-layer security techniques have attracted extensive attention [24–26]. Existing works have considered a variety of security enhancement mechanisms, including beamforming, artificial noise, cooperative jamming, and RIS-assisted transmission. For instance, ref. [24] investigated RIS-assisted secure UAV transmission, ref. [25] considered artificial-noise-aided countermeasures against a UAV eavesdropper, and ref. [26] studied cooperative secure transmission and computation in UAV-assisted mobile edge computing (MEC). Despite these advances, most prior works still rely on simplified threat models in which the eavesdropper is static, follows a pre-defined trajectory, or is otherwise non-reactive to the legitimate transmitter. Under these assumptions, the secure design remains essentially a static or slowly time-varying optimization problem. In contrast, if the eavesdropper actively adapts its position in response to the instantaneous signal environment, the transmitter encounters a fundamentally distinct challenge: its current beamforming and trajectory decisions directly shape the future threat state.

From a methodological perspective, learning-based approaches have also been explored for UAV communication optimization [12–15,27,28]. Their main advantage lies in handling high-dimensional, continuous, and sequential decision-making problems that are difficult to solve using conventional optimization tools. However, even in existing learning-based secure communication studies, the eavesdropper is commonly modeled as static or only weakly dynamic, and the learning objective is often limited to generic resource allocation rather than tightly coupled secure beamforming and trajectory control. Moreover, the partially observable nature of practical secure transmission is rarely emphasized: in realistic scenarios, the eavesdropper's real-time position and channel state information (CSI) are not perfectly available during execution, which makes the environment both reactive and only partially observable.

In summary, the existing literature reveals three major limitations. First, most prior studies do not model the eavesdropper as a truly reactive adversary whose movement is dictated by the received signal conditions. Second, although multi-UAV and swarm-based systems have been extensively investigated, their distributed spatial DoFs have not been fully exploited for collaborative secure beamforming under active threats. Third, existing learning-based approaches fail to explicitly formulate the problem as a partially observable sequential secure control task, in which multiple UAVs must coordinate their mobility and beamforming decisions against a dynamically evolving adversary. These limitations collectively motivate the present work.

1.3. Motivations and Contributions

The above discussion reveals that the principal gap in the existing literature lies not merely in the absence of additional optimization variables or alternative learning algorithms but in the lack of a unified framework for secure UAV-swarm transmission against a reactive eavesdropper. In the considered scenario, the UAV swarm must simultaneously exploit two tightly coupled sources of adaptability: spatial mobility for repositioning the

distributed array and collaborative beamforming for shaping the confidential signal in space. Once the eavesdropper becomes mobile and reacts to the received power, these two design dimensions can no longer be treated independently, since the current trajectory and beamforming decisions jointly determine both the immediate secrecy rate and the future adversarial state. This makes the problem fundamentally different from conventional static secure communication designs or pre-planned UAV trajectory optimization.

Meanwhile, this problem also exhibits a natural multi-agent and partially observable structure. Each UAV only has access to local and delayed information during execution, whereas effective defense against the eavesdropper requires globally coordinated actions from the whole swarm. Therefore, the distinctive contribution of this work is not simply the integration of beamforming, trajectory optimization, and MARL but the formulation and solution of a coupled secure control problem in which a UAV swarm must jointly exploit distributed spatial DoFs against a reactive adversary under partial observability. In light of the aforementioned analyses, this paper develops an end-to-end multi-agent proximal policy optimization (MAPPO)-based approach for secure UAV swarm communication with collaborative beamforming and cooperative trajectory planning. The primary contributions are summarized as follows:

- Reactive Adversarial Security Modeling: Different from prior secure UAV communication studies that mainly assume static, pre-defined, or non-reactive eavesdroppers, we establish a secure UAV-swarm communication model with a reactive mobile eavesdropper. Specifically, the eavesdropper updates its position based on the spatial gradient of the received signal power, thereby transforming the secure transmission task from a conventional static optimization problem into a sequential adversarial interaction. Based on this model, the received signal expressions and secrecy rate metrics for both legitimate users and the eavesdropper are derived.
- Strongly Coupled Joint Secure Control Formulation: We formulate a time-averaged minimum secrecy rate maximization problem by jointly optimizing the collaborative beamforming vectors and cooperative trajectories of the UAV swarm under transmit power and trajectory constraints. Unlike existing works that primarily focus on static security designs or loosely coupled optimization variables, the proposed formulation explicitly captures the tight coupling between swarm mobility and distributed spatial beam control under an active threat. This formulation reveals that defending against a reactive eavesdropper fundamentally requires the joint exploitation of mobility and distributed spatial degrees of freedom.
- MAPPO-Based Solution Under Partial Observability: To solve the resulting non-convex and dynamic problem, we formulate it as a partially observable Markov decision process (POMDP) and develop an end-to-end MAPPO-based framework. The role of MAPPO here is not merely that of a generic learning optimizer; rather, it is specifically adopted to match the intrinsic structure of the problem, including multi-agent cooperation, delayed local observations, and the reactive evolution of the adversarial environment. In this way, the UAV swarm learns a coordinated control policy that jointly adapts trajectory and beamforming against dynamically changing threats.

The rest of this paper is organized as follows. Section 2 introduces the system model, and Section 3 formulates and analyzes the optimization problem. Section 4 details the proposed MAPPO-based approach. The simulation results are presented in Section 5. Finally, Section 6 concludes this paper.

2. System Model

2.1. Network Model

As illustrated by Figure 1, the considered UAV swarm communication system consists of U UAVs, L terrestrial legitimate users, and an active eavesdropper, which attempts to dynamically adjust its location to obtain better channel conditions. Each UAV is equipped with an N_t-element uniform linear array (ULA) to act as a distributed aerial base station. The time-varying coordinate of UAV n in time slot n is denoted by $\boldsymbol{q}_u[n] = [x_u[n], y_u[n], H]^T$, where H is the fixed flying altitude. The UAV swarm's collective trajectory is subject to the following mobility constraint:

$$\|\boldsymbol{q}_u[n+1] - \boldsymbol{q}_u[n]\| \leqslant d_{max}, \forall u, n. \tag{1}$$

In order to avoid collision among UAVs, the minimum distance constraint should be guaranteed:

$$\|\boldsymbol{q}_i[n] - \boldsymbol{q}_j[n]\| \geqslant d_{min}, \forall n, i \neq j. \tag{2}$$

All legitimate users are assumed to be equipped with a single antenna and be static during the transmission, whose locations are specified by $\boldsymbol{q}_l = [x_l, y_l, 0]^T$. In contrast, the eavesdropper is equipped with N_e antennas; it attempts to actively adjust its location for improving interception performance; its coordinate in time slot n is given by $\boldsymbol{q}_e[n] = [x_e[n], y_e[n], 0]^T$. Critically, from the perspective of the UAV swarm, Eve's location is initially known only at $n = 0$, i.e., $\boldsymbol{q}_e[0]$ is available. The subsequent trajectory of Eve is unknown and must be inferred by the UAV swarm through interactions with the whole environment, which forms a key part of the learning problem.

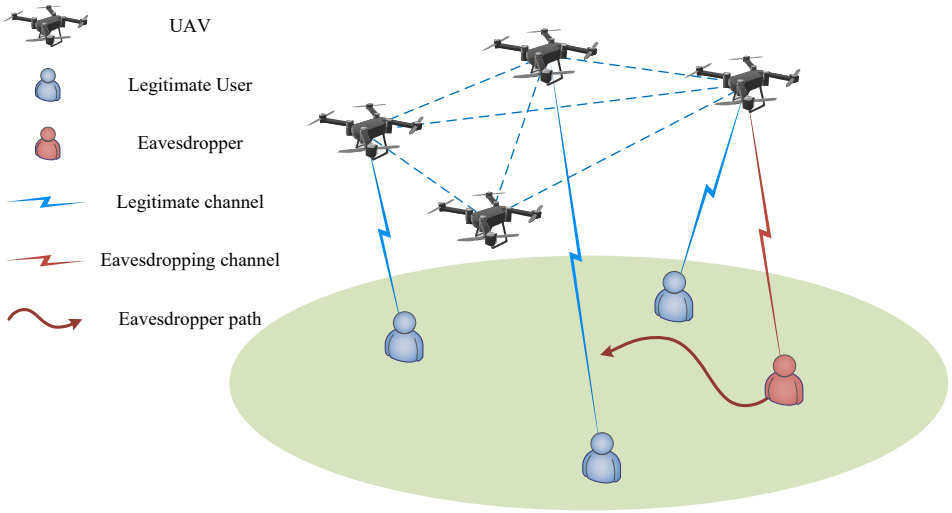


Figure 1. UAV swarm communication model.

2.2. Communication Model

The UAV swarm is expected to form a distributed antenna array to transmit the common confidential signal $s[n]$ with $\mathbb{E}[|s[n]|^2] = 1$. For the purpose of collaborative beamforming analysis, this paper assumes perfect phase synchronization among the distributed UAVs. In practice, synchronization errors (e.g., phase offsets) can degrade coherent information, combining main-lobe gain and increasing side-lobe leakage to the eavesdropper and thereby penalizing the achievable secrecy rate. Thus, assuming perfect synchronization allows us to isolate and highlight the spatial security enhancements of the proposed joint optimization framework. Each UAV applies its own distinctive beamforming vec-

tor $\boldsymbol{w}_u[n] \in \mathbb{C}^{N_t \times 1}$. As such, the received signal models for legitimate users and the eavesdropper are given by the following:

$$y_l[n] = \sum_{u=1}^{U} \boldsymbol{h}_{u,l}^H[n]\boldsymbol{w}_u[n]s[n] + n_l[n], \tag{3}$$

$$\boldsymbol{y}_e[n] = \sum_{u=1}^{U} \boldsymbol{H}_{u,e}^H[n]\boldsymbol{w}_u[n]s[n] + \boldsymbol{n}_e[n], \tag{4}$$

where $\boldsymbol{h}_{u,l}^H[n] \in \mathbb{C}^{N_t \times 1}$ is the channel gain from the u-th UAV to the l-th legitimate user, and $\boldsymbol{H}_{u,e}^H[n] \in \mathbb{C}^{N_t \times N_e}$ is the channel gain from the u-th UAV to the eavesdropper. $n_l[n] \sim \mathcal{CN}(0, \sigma_l^2)$ and $\boldsymbol{n}_e[n] \sim \mathcal{CN}(0, \sigma_e^2 \boldsymbol{I}_{N_e})$ are the additive white Gaussian noise (AWGN) corresponding to the l-th legitimate user and the eavesdropper.

The air-to-ground channel is modeled by a dominant LoS path, which is a common simplification in UAV-related studies given the typical altitude and operating environment of UAVs [18,25]. Under Rician fading channels with a high K-factor (typical for high-altitude UAVs), the proposed algorithm is expected to sustain its performance well since the deterministic LoS path dominates. Conversely, severe multi-path (low K-factor) would disrupt phase coherence among distributed antennas and degrade the array gain. This allows us to isolate and highlight the impact of the proposed collaborative beamforming and trajectory design, while the inclusion of more realistic statistical channel characteristics, such as Rician fading, is left for future investigation. As such, the channel vector from UAV u to legitimate user l in time slot n is modeled as follows:

$$\boldsymbol{h}_{u,l}[n] = \rho_{u,l}[n]\mathbf{a}(\phi_{u,l}[n]), \tag{5}$$

where $\rho_{u,l}[n] = \rho_0 d_{u,l}^{-\alpha}[n]$ is the path loss, with ρ_0, $d_{u,l}[n]$, and α representing the reference path loss at a distance of 1 m, the Euclidean distance between UAV u and legitimate user l, and the path loss exponent, respectively. ρ_0 could be further determined via $\rho_0 = \frac{c}{4\pi f_c}$. Given the azimuth angle of departure (AoD) ϕ, the steering vector for the horizontal ULA is given by the following:

$$\mathbf{a}_{tx}(\phi) = \frac{1}{\sqrt{N_t}}\left[1, e^{j\frac{2\pi d}{\lambda}sin\phi}, \ldots, e^{j\frac{2\pi d}{\lambda}(N_t-1)sin\phi}\right]^T, \tag{6}$$

where d is the antennas spacing, and λ is the carrier wavelength. Moreover, the aggregate channel from the entire UAV swarm to legitimate user l is $\boldsymbol{h}_l[n] \triangleq [\boldsymbol{h}_{1,l}^T[n], \boldsymbol{h}_{2,l}^T[n], \ldots, \boldsymbol{h}_{U,l}^T[n]]^T \in \mathbb{C}^{UN_t \times 1}$.

Similarly to the legitimate channel, the channel matrix between UAV U and the eavesdropper $\boldsymbol{H}_{u,e}[n] \in \mathbb{C}^{N_t \times N_e}$ is modeled as follows:

$$\boldsymbol{H}_{u,e}[n] = \rho_{u,e}[n]\mathbf{a}_{tx}(\phi_{u,e}[n])\mathbf{a}_{rx}^H(\theta_{u,e}[n]), \tag{7}$$

where the definition of $\rho_{u,e}[n]$ is similar to $\rho_{u,l}[n]$, $\phi_{u,e}$ is the AoD from the UAV, and $\theta_{u,e}$ is the angle of arrival (AoA) at the eavesdropper. $\mathbf{a}_{tx}$ and $\mathbf{a}_{rx}$ are the steering vectors for the UAV and the eavesdropper, and the structure of $\mathbf{a}_{rx}$ could refer to (6). After that, the consolidated eavesdropping channel gain from the entire UAV swarm to the eavesdropper is structured as $\boldsymbol{H}_e[n] \triangleq [\boldsymbol{H}_{1,e}^T[n], \boldsymbol{H}_{2,e}^T[n], \ldots, \boldsymbol{H}_{U,e}^T[n]]^T \in \mathbb{C}^{UN_t \times N_e}$.

To enhance the comprehension of the collaborative relationship among UAV swarm and to integrate the dispersed ULA into a unified virtual antenna array, the individual beamforming vectors could be integrated into an overall beamforming vector as follows:

$$\boldsymbol{w}[n] = \left[\boldsymbol{w}_1^T[n], \boldsymbol{w}_2^T[n], \ldots, \boldsymbol{w}_U^T[n]\right]^T, \tag{8}$$

and accordingly, the received signal models (3) and (4) could be rewritten as follows:

$$y_l[n] = \boldsymbol{h}_l^H[n]\boldsymbol{w}[n]s[n] + n_l[n], \tag{9}$$

$$\boldsymbol{y}_e[n] = \boldsymbol{H}_e^H[n]\boldsymbol{w}[n]s[n] + \boldsymbol{n}_e[n], \tag{10}$$

and the achievable for legitimate user b in time slot n is

$$R_l[n] = \log_2\left(1 + \frac{\left|\boldsymbol{h}_l^H[n]\boldsymbol{w}[n]\right|^2}{\sigma_l^2}\right). \tag{11}$$

Assuming that the eavesdropper employs a maximum ratio combining (MRC) the receiver to maximize the received signal to noise (SNR), the corresponding achievable rate in time slot n is given by

$$R_e[n] = \log_2 \det\left(\boldsymbol{I}_{N_e} + \frac{1}{\sigma_e^2}\boldsymbol{H}_e^H[n]\boldsymbol{w}[n]\boldsymbol{w}^H[n]\boldsymbol{H}_e[n]\right). \tag{12}$$

As such, the secrecy rate is $R_s[n] = [R_L[n] - R_e[n]]^+$, where $R_L[n]$ is the achievable minimum rate for legitimate users:

$$R_L[n] = \min_{l \in L} R_l[n]. \tag{13}$$

2.3. Moving Strategy and Intercept Model of the Eavesdropper

As an intelligent adversary, the eavesdropper could actively move to a location favorable for interception, aiming at minimizing the secrecy rate $R_s[n]$. Following common practices in previous work [26,29], we model the mobility of the eavesdropper as a practical gradient-ascent strategy. That is, in each time slot, the eavesdropper estimates the received signal power from the UAV swarm, calculates the gradient of that received power with respect to its own position, and then moves in the direction of the steepest ascent. Specifically, the position update follows

$$\boldsymbol{q}_e[n+1] = \boldsymbol{q}_e[n] + v_e \frac{\nabla_{\boldsymbol{q}_e} P_r[n]}{||\nabla_{\boldsymbol{q}_e} P_r[n]||}, \tag{14}$$

where $P_r[n] \triangleq ||\boldsymbol{H}_e^H[n]\boldsymbol{w}[n]||^2$ is the received power at the eavesdropper, and v_e is its speed. The detailed derivation of the gradient $\nabla_{\boldsymbol{q}_e} P_r[n]$ is provided in Appendix A. Following this model, a strong and adaptive threat is illustrated.

3. Optimization Problem Formulation and Analysis

In light of the aforementioned system model and formula of the secrecy rate, the following time-averaged minimum secrecy rate maximization problem is formulated in order to enhance the security performance of UAV swarm transmission. This is achieved by

jointly optimizing the collaborative beamforming vectors, $\boldsymbol{w}[n]$, and the trajectories, $\boldsymbol{q}_U[n]$, of the UAV swarm over N time slots:

$$\max_{\boldsymbol{w}[n],\boldsymbol{q}_U[n]} \frac{1}{N}\sum_{n=1}^{N} R_s[n] \tag{15}$$

$$s.t.\ \|\boldsymbol{w}_u[n]\|_2^2 \leqslant P_t, \forall u, n, \tag{16}$$

$$\boldsymbol{q}_u[n] \in \mathbb{Q}, \forall u, n, \tag{17}$$

$$(1),(2) \tag{18}$$

in which (16) represents the transmit power constraint, and P_t denotes the maximum transmit power budget for each UAV. Constraint (17) ensures that the UAV swarm remains within the designated flight region $\mathbb{Q}$, defined as

$$\mathbb{Q} \triangleq \{(x, y, z) \in \mathbb{R}^3 | x_{min} \leqslant x \leqslant x_{max}, y_{min} \leqslant y \leqslant y_{max}, z = 80\}. \tag{19}$$

The objective of optimization problem (15) is to maximize the instantaneous secrecy rate $R_s[n]$ in each time slot, which is defined as the difference between the legitimate multicast rate $R_L[n]$ and the achievable rate of the eavesdropper. Note that $R_L[n]$ is determined by the worst-case achievable rate among all legitimate users, i.e., $\min_{l \in L} R_l[n]$, ensuring fairness and robustness in the secure multicast scenario.

The time-averaged minimum secrecy rate maximization problem (15) presents a formidable challenge to solve utilizing conventional optimization techniques due to the following reasons:

- Non-Convexity: The optimization problem is inherently non-convex. The objective function involves the difference of two logarithmic functions, which is neither convex nor concave. Furthermore, the collision avoidance constraint (2) is non-convex, and the complex coupling between the UAV trajectory variables $\boldsymbol{q}_U[n]$ and beamforming vectors $\boldsymbol{w}[n]$ within the channel expressions adds severe non-linearity. While techniques like successive convex approximation (SCA) or block coordinate descent (BCD) can handle static non-convex problems, they are computationally prohibitive for large-scale swarms and often trapped in sub-optimal solutions.
- Dynamic and Reactive Environment: Most critically, the considered scenario involves an active eavesdropper whose trajectory $\boldsymbol{q}_e[n]$ is not pre-determined but evolves dynamically based on the UAV swarm's actions. Specifically, $\boldsymbol{q}_e[n+1]$ is a function of the gradient $\nabla_{\boldsymbol{q}_e} P_r[n]$, which in turn depends on the UAVs' beamforming and positions in time slot n. This creates a time-sequential dependency where the current decision influences the future state of the environment. Traditional optimization methods, which typically assume a static or predictably fading environment, fail to capture this interactive game-theoretic dynamic.
- High Dimensionality: As the number of UAVs and antennas increases, the joint action space, i.e., continuous flight control and high-dimensional beamforming, expands exponentially. Centralized optimization algorithms scale poorly in such high-dimensional spaces, incurring unacceptable computational complexity for real-time applications.

Given these challenges, problem (15) is more naturally modeled as a POMDP due to the actuality that the instant location of the eavesdropper is unknown for the UAV swarm. This perspective allows us to leverage MARL, which excels in handling sequential decision-making tasks in dynamic, continuous environments. Consequently, in the subsequent section, an approach based on the MAPPO algorithm is illustrated, which adopts the

framework of centralized training with decentralized execution (CTDE), enabling the UAV swarm to learn a cooperative policy that proactively suppresses the active eavesdropper while ensuring robust communication.

4. Proposed MAPPO-Based Approach

In this section, we first formulate the POMDP corresponding to the time-averaged minimum secrecy rate maximization problem (15), considering the dynamic nature of the eavesdropper. After that, the MAPPO-based approach is detailed to solve it.

4.1. POMDP Formulation

Recall the composition of the UAV swarm communication system, the entire electromagnetic propagation environment together with all legitimate users. The dynamic eavesdropper will be the environment, and each UAV is an agent. Consequently, the same indicator, u, is employed for both agents and UAVs. Utilizing the CTDE framework, the POMDP for the UAV swarm is defined by the tuple $\langle \mathcal{S}, \mathcal{O}, \mathcal{A}, \mathcal{R}, \gamma \rangle$, where $\mathcal{S}$ is the global state space, $\mathcal{O}$ is the observation space for each individual agent, $\mathcal{A}$ is the action space for each agent, $\mathcal{R}$ is the reward function, and γ is the discount factor.

4.1.1. State Space $\mathcal{S}$ and Observation Space $\mathcal{O}$

For normal Markov decision processes (MDPs), only state space $\mathcal{S}$ is involved. Different from that, the MARL approach in the CTDE manner requires an observation space $\mathcal{O}$ for each agent, i.e., each UAV. As such, the global state $s[n] \in \mathcal{S}$ is available to the critic network during training for evaluating the value function, while each UAV operates based on its local observation $o_u[n] \in \mathcal{O}$.

The state at step n includes the precise locations of all entities and all channel gains:

$$s[n] = \{\boldsymbol{Q}_U[n], \boldsymbol{Q}_l, \boldsymbol{q}_e[n], \boldsymbol{H}_L[n], \boldsymbol{H}_e[n]\}, \tag{20}$$

where $\boldsymbol{Q}_U[n] \triangleq \{\boldsymbol{q}_1[n], \ldots, \boldsymbol{q}_U[n]\}$ is the overall trajectory of the UAV swarm, $\boldsymbol{Q}_L \triangleq \{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_L\}$ is the combined location of legitimate users, and $\boldsymbol{H}_L[n] \triangleq \{\boldsymbol{h}_1[n], \ldots, \boldsymbol{h}_L[n]\}$ is the compound legitimate channel gain. Crucially, $\boldsymbol{s}[n]$ concludes the current position of the eavesdropper and the global CSI, which is available to the centralized critic during training.

Reflecting the partial observability, each agent makes decisions based on its local observation $\boldsymbol{o}_u[n] \in \mathcal{O}$, which is given by

$$o_u[n] = \{\boldsymbol{q}_u[n], \boldsymbol{Q}_l, \boldsymbol{q}_e[n-1], \boldsymbol{H}_{u,L}[n], \boldsymbol{H}_{u,e}[n-1]\}, \tag{21}$$

where $\boldsymbol{H}_{u,L}[n] \triangleq \{\boldsymbol{h}_{u,1}[n], \ldots, \boldsymbol{h}_{u,L}[n]\}$ is the aggregated legitimate channel from UAV u to the legitimate users. The distinction between $\boldsymbol{H}_L[n]$ and $\boldsymbol{H}_{u,L}[n]$ is that the former contains all the legitimate channel gains corresponding to all UAVs, whereas in the latter, only those of individual UAV u are present. More intuitively, the combination of $\{\boldsymbol{H}_{1,L}[n], \boldsymbol{H}_{2,L}[n], \ldots, \boldsymbol{H}_{U,L}[n]\}$ will ultimately manifest as $\boldsymbol{H}_L[n]$. A similar relationship is evident in $H_{u,e}[n]$ and $\boldsymbol{H}_e[n]$.

A fundamental distinction between $\mathcal{S}$ and $\mathcal{O}$ lies in the scope and timeliness of the available information. While the state space encompasses the real-time global configuration, the observation space is strictly confined to local visible information. Notably, the eavesdropper's location and CSI are only observable with a one-step delay, indicating the realistic constraints of information acquisition.

4.1.2. Action Space $\mathcal{A}$

As the optimization variables of problem (15) are transmit beamformers and trajectories of the UAV swarm, the action at step n of agent u is outlined as follows:

$$a_u[n] = \{\boldsymbol{w}_u[n], \boldsymbol{v}_u[n]\}, \tag{22}$$

where $\boldsymbol{v}_u[n] \triangleq \{\delta_{x,u}[n], \delta_{y,u}[n]\}$ is the velocity specified by the displacement along the x-axis and y-axis. Considering the output of neural network, the combination of $\boldsymbol{w}_u[n]$ should be completed via imaginary parts, which render the dimension of $a_u[n]$ as $2N_t + 2$. Moreover, additional normalization procedures will be conducted to ensure the transmit power constraint.

4.1.3. Reward Function $\mathcal{R}$

The reward at time step n comprises these three parts: secrecy rate $R_{sec}[n]$, collision avoidance penalty $R_{col}[n]$, and boundary crossing penalty $R_{bou}[n]$, i.e., $R[n] = R_{sec}[n] + R_{col}[n] + R_{bou}[n]$:

- Secrecy Rate: As the objective of optimization problem (15) is to maximize the time-averaged minimum secrecy rate, the minimum secrecy rate, which is also the core component of $R[n]$, could be defined as $R_{sec}[n] = R_s[n]$. It should be noted that, although the reward corresponding to each single step n does not directly equate to the time-averaged minimum secrecy rate, the characteristics of reinforcement learning focusing on cumulative rewards will make the sum secrecy rate the actual optimization goal. The calculation of the total secrecy rate is divided by the number of time slots of transmission or steps of POMDP to derive the time-averaged minimum secrecy rate. Consequently, such a value of $R_{sec}[n]$ is equivalent to the time-averaged minimum secrecy rate. Furthermore, this design is intended to encourage all UAV agents to improve the security performance of transmission.
- Collision Avoidance Penalty: In order to ensure that each UAV maintains a safety distance with each other, i.e., satisfy the constraint (2), the collision avoidance penalty is defined as follows:

$$R_{col}[n] = \begin{cases} -\zeta, & \left\|\boldsymbol{q}_i[n] - \boldsymbol{q}_j[n]\right\| < d_{min}, \forall n, i \neq j, \\ 0, & \text{otherwise}, \end{cases} \tag{23}$$

  where ζ is the collision penalty hyperparameter.
- Boundary Crossing Penalty: The penalty term is established to ensure the designated flight region $\mathbb{Q}$ for UAV swarm, i.e., constraint (17):

$$R_{bou}[n] = \begin{cases} -\mu, & \boldsymbol{q}_u[n] \notin \mathbb{Q}, \forall u, \\ 0, & \text{otherwise}, \end{cases} \tag{24}$$

  where μ is the boundary penalty hyperparameter.

It is important to note that the fairness among multiple legitimate users in terms of received rates is inherently addressed by the multicast rate $R_L[n]$. Specifically, assuming that the physical constraints are satisfied (i.e., $R_{col}[n] = 0$ and $R_{bou}[n] = 0$), the primary driving force for the policy update is exclusively governed by $R_{sec}[n]$. Since $R_{sec}[n]$ is strictly formulated as the minimum current secrecy rate among all users, it essentially implements a max–min fairness criterion. Consequently, considerations of transmission fairness are not required during the design of $R[n]$, and dividing $R[n]$ into U parts yields the reward corresponding to each agent.

The formulation of POMDP is hereby finished with predefined discount factor γ. After that, the framework of the MAPPO-based approach will be presented for solving it.

4.2. Framework of the MAPPO-Based Approach

Considering the swarm intelligence of the UAV swarm and the complexity of solving (15) through traditional algorithms such as convex optimization, we leverage MARL as the basic frame, and we develop a MAPPO-based approach accordingly.

Next, we first propose the network structure of actor and critic, and the overall structure of the MAPPO-based algorithm. After that, the training strategy and rules followed are elaborated and summarized in Algorithm 1.

Algorithm 1: MAPPO-based algorithm

Input: Actor networks π_{ϑ_u}, critic network V_φ, maximum iteration number K, environment step N

Output: Optimized actor networks $\pi_{\hat{\vartheta}_u}$, $V_{\hat{\varphi}}$

```
for k ← 1 to K do
    for n ← 1 to N do
        for u ← 1 to U do
            Observe local observation o_u[n];
            Sample action a_u[n] ~ π_{ϑ_u}(a_u[n]|o_u[n]);
            Execute actions {a_1[n], a_2[n], ..., a_u[n]};
            Received reward R[n], next global state s[n+1], and next local
             observations o_u[n+1];
            Store {s[n], o_u[n], a_u[n], R[n], s[n+1], o_u[n+1]}in the buffer D;
    Sample a batch of M trajectories from D;
    Update the critic network V_φ following (26);
    for u ← 1to U do
        Update the actor network π_{ϑ_u} following (27).
```

4.2.1. Network Structure

As the adopted CTDE framework, there will be multiple actor networks for MAPPO, each corresponding to one UAV agent. Different with this, only one centralized critic network is possessed to evaluate all the policies provided by multiple actor networks:

- Actor Network $\pi_{\vartheta_u}(a_u[n]|o_u[n])$ (Decentralized): As illustrated in Figure 2a, each agent should be allocated an individual actor network, which takes the corresponding observation at each step $o_u[n]$ as input, and it outputs a probability distribution over action $a_u[n]$. Each network consists of fully connected layers with nonlinear activations, typically using LeakyReLU in the hidden layers. As shown in Figure 2a, the input layer of the actor network contains $3(2+L)+2N_t(L+N_e)$ neurons. It is followed by two fully connected hidden layers, each with 128 neurons, and an output layer of $2N_t+2$ neurons. The actor networks operate fully decentralized during execution, mapping only local observations to actions.
- Critic Network $V_\varphi(s[n])$ (Centralized): Different from actor networks, a single centralized critic network receives the global state $s[n]$ as input and outputs a scalar value estimating the expected return. The specific structure of the critic network is analogous to that of the actor network, except for the scale of hidden layers, as illustrated by Figure 2b. Specifically, the input layer the of critic network comprises $3(1+L+U)+2UN_t(L+N_e)$ neurons, followed by two fully connected hidden layers with 1024 and 256 neurons, respectively, and it outputs a scalar value through a

single-neuron output layer. The incorporation of more hidden layer neurons has been demonstrated to facilitate the critic network to more accurately capture the implicit information contained in the global state. This, in turn, has been shown to result in a more precise state value evaluation, thereby ensuring the provision of a more reliable gradient benchmark for the actor.

It should be noted that batch normalization is applied to both the actor and critic networks to enhance training stability.

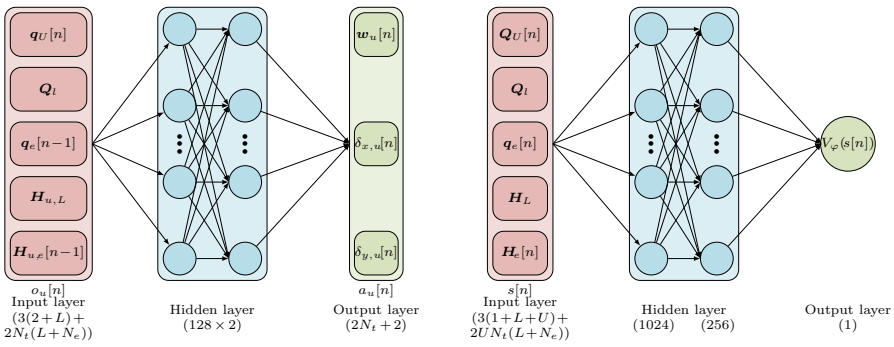


Figure 2. (**a**) Network structure of the decentralized actor networks. (**b**) Network structure of the centralized critic network.

4.2.2. Algorithm Architecture

At each step n, all agents independently observe the environment to obtain its local partial observation $o_u[n]$; then, the action $a_u[n]$ sampled from its policy network $\pi_{\vartheta_u}(a_u[n]|o_u[n])$ is decided. After this, agents interact with the environment by action $a_u[n]$, and feedback consists of the reward corresponding to each agent, i.e., $R_u[n]$, along with the transitioned observation $o_u[n+1]$. During the interaction, the tuple $(\mathcal{S}_n, \mathcal{O}_n, \mathcal{A}_n, \mathcal{R}_n, \mathcal{S}_{n+1}, \mathcal{O}_{n+1})$ is collected and stored in a replay buffer for subsequent training.

After the replay buffer is full, a set of M trajectories will be sampled from the replay buffer for the network update. The states $\{s[m], s[m+1]\}$ of length M are delivered to the centralized critic network for computing values $V_\varphi(s[m])$ and $V_\varphi(s[m+1])$. Generalized advantage estimation (GAE) is then utilized to compute the target for the critic network and estimate advantages for policy updates, as it optimally balances the bias–variance trade-off in advantage estimation. The advantage estimate at step m, which leverages $V_\varphi(s[m])$, $V_\varphi(s[m+1])$, and $R[m]$ and is computed via GAE to reduce variance and stabilize learning, is defined as

$$A[m] = \sum_{l=0}^{\infty} (r\lambda)^l \delta[m+l], \tag{25}$$

where $\delta[m] = R[m] + \gamma V_\varphi(s[m+1]) - V_\varphi(s[m])$ is the temporal difference (TD) error, and λ is the GAE parameter balancing bias and variance. In the considered fully cooperative setting, the shared reward and centralized value function—estimated by the centralized critic $V_\varphi(s[m])$ using the global state—yield identical advantage estimations for all agents, i.e., $A_u[m] = A[m]$ for each u. Within the CTDE framework, the centralized critic leverages global information, including the true location of the eavesdropper, to compute low-variance value estimates and the advantage function. These globally aware gradient signals are then used to optimize the policy networks of the actors, thereby guiding the agents to gradually learn a robust policy capable of implicitly reasoning about the unobserved dynamics of the eavesdropper.

The critic network parameters, φ, are updated by minimizing the value function regression loss over the sampled batch:

$$L^{critic}(\varphi) = \frac{1}{M}\sum_{m=1}^{M}\left[V_{\varphi}(s[m]) - V_{target}[m]\right]^2, \tag{26}$$

where $V_{target}[m]$ could be calculated via $V_{target}[m] = A[m] + V_{\varphi_{old}}(s[m])$.

Meanwhile, the decentralized actor network is updated only from its local observation, and using the proximal policy optimization (PPO)-Clip objective to ensure monotonic improvement,

$$L^{actor}(\vartheta_u) = \hat{\mathbb{E}}_m\left[\min(r_m(\vartheta_u)\hat{A}_u[m], \text{clip}(r_m(\vartheta_u), 1-\epsilon, 1+\epsilon)\hat{A}_u[m])\right], \tag{27}$$

where $r_m(\vartheta_u) = \frac{\pi_{\vartheta_u}(a_u[m]|o_u[m])}{\pi^{old}_{\vartheta_u}(a_u[m]|o_u[m])}$ is the importance sampling ratio between the current and previous actor policies for agent u, and ϵ is the PPO clip threshold.

This training mechanism ensures that all actor networks are updated based on its unique trajectory and computed advantage while still benefiting from the stability of centralized value estimation and advantage computation via the centralized critic network. Eventually, the UAV swarm is expected to learn coordinated maneuvers that trap or confuse the reactive eavesdropper based on inferred trajectories.

The overall algorithm structure is outlined in Figure 3.

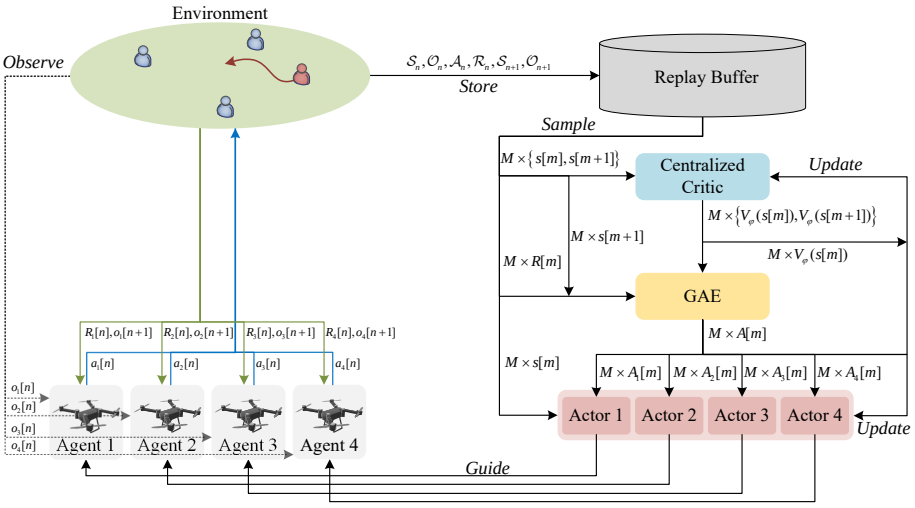


Figure 3. Algorithm structure of the MAPPO-based approach, taking $U = 4$ as an example.

4.2.3. Computational Complexity Analysis

To assess the practical feasibility of the proposed MAPPO-based algorithm, this section provides a comprehensive analysis of its computational complexity. The analysis evaluates the computational requirements across both the offline training and online execution phases by examining three key aspects: training complexity, inference latency, and communication overhead under the CTDE framework:

- Training Complexity: The per-iteration training complexity is empirically dominated by the forward and backward passes of the U actor networks and the single centralized critic network. Let the input and output dimensions be D_{in} and D_{out}, respectively. For a neural network with two hidden layers (H_1, H_2), the complexity of one pass scales as $\mathcal{O}(D_{\text{in}}H_1 + H_1H_2 + H_2D_{\text{out}})$. During each training iteration, let N_{ep} be the number of

PPO update epochs and M be the mini-batch size. The total complexity per iteration scales as follows:

$$\mathcal{O}\left(N_{ep}M\left(\sum_{u=1}^{U} C_{\text{actor},u} + C_{\text{critic}}\right)\right) \tag{28}$$

where $C_{\text{actor},u}$ and C_{critic} represent the per-sample forward pass complexities for the actor and critic, respectively. In our implementation, the critic employs significantly larger hidden layers ($H_1 = 1024, H_2 = 256$) than the actors ($H_1 = 128, H_2 = 128$) to efficiently process the global state. Regardless of this expansion, the resultant FLOPs are comfortably feasible for offline processing on standard GPU infrastructures.

- Inference Complexity: During the execution phase, the centralized critic is organically discarded, and each UAV exclusively runs its local decentralized actor network. The inference complexity for a single decision is minimal and bounded by $\mathcal{O}(D_{\text{in}}H_1 + H_1H_2 + H_1D_{\text{out}})$. On an embedded processor (e.g., NVIDIA Jetson), this forward-pass inference typically executes in a few milliseconds. This is substantially shorter than the system time slot interval $\Delta_t = 0.5$ s, rigorously satisfying the real-time requirements for trajectory and beamforming adaptation.
- Communication Overhead: Under the CTDE framework, inter-agent communication is solely required during the centralized training phase to aggregate trajectories and global states. During the decentralized execution phase, each UAV independently maps its local observation to an action without requiring real-time information exchange with other agents or a central controller. This results in zero inter-agent communication overhead during practical operation, ensuring exceptional scalability for large UAV swarms.

Details of the MAPPO-based approach are outlined in Algorithm 1. It is worth noting that unlike traditional supervised learning models that may overfit to static training data, the proposed MAPPO-based framework naturally mitigates overfitting through continuous interactions with dynamically updating memories in the replay buffer.

5. Experimental Results

In this section, multiple experiments are illustrated to verify the effectiveness and superiority of the proposed MAPPO-based approach for the time-averaged minimum secrecy rate maximization problem. The UAV swarm consists of four UAVs, and the initial locations are $\{[20, 20, 80]^T, [20, 30, 80]^T, [30, 30, 80]^T, [30, 20, 80]^T\}$. The flight altitude is fixed at $H = 80$ m to ensure dominant LoS links and maintain computational tractability by focusing on 2D horizontal trajectory optimization. There are five legitimate users located at $\{[100, 50, 0]^T, [174, 76, 0]^T, [144, 170, 0]^T, [56, 170, 0]^T, [26, 96, 0]^T\}$. For the eavesdropper, its initial position is $[120, 160, 0]^T$. $\mathbb{Q}$ is set to be the square in the range $[0, 300]$. For all the ULAs, the antenna space is set as half of the wavelength. To better visualize the spatial distribution of the system, the 3D coordinates of the UAV swarm, the legitimate users, and the eavesdropper are illustrated in Figure 4. It should be noted that for nodes with mobility, such as UAVs and the eavesdropper, the locations depicted in Figure 4 represent their initial states.

The hyperparameters are selected as follows. The standard DRL parameters, including the discount factor γ, the PPO clip parameter ϵ, and the GAE parameter λ, are set to their commonly used values from the foundational PPO work [30] to ensure training stability. The learning rates for the actor and critic networks are chosen based on a coarse search to balance convergence speed and stability. Currently, the penalty coefficients ζ and μ are tuned via grid search to enforce safety constraints without destabilizing the learning of the primary secrecy rate objective. Considering adaptive tuning to enhance robustness across different scenarios is a valuable direction for future work. The core simulation

parameters are summarized in Table 1. All simulations were executed on a laptop equipped with an Intel Core Ultra 9 CPU and an NVIDIA RTX 5070Ti GPU using the PyTorch 2.5.0 framework. To ensure the statistical reliability of the results, all simulations involving training convergence and secrecy performance are repeated over 20 independent runs with different random seeds.

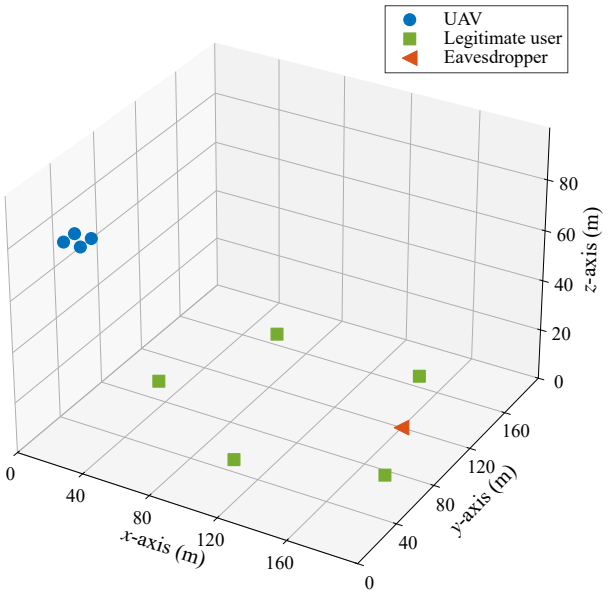


Figure 4. 3D spatial distribution of network nodes.

Table 1. Primary simulation parameters.

Parameter	Value
Transmit power of UAV P_t	20 dBm
Antenna number of UAV N_t	16
Antenna number of eavesdropper N_e	4
Maximum moving distance per time slot d_{max}	2 m
Minimum distance among UAVs d_{min}	3 m
Velocity of the eavesdropper v_e	1 m/s
Time slot interval Δ_t	0.5 s
Carrier frequency f_c	28 GHz
Path loss exponent α	1.2
Noise power $\sigma_i^2, i \in \{l, e\}$	-110 dBm
Maximum iteration number K	10,000
Environment step N	300
Discount factor γ	0.9
Clip parameter ϵ	0.2
GAE parameter λ	0.95
Learning rate of actor network	1×10^{-4}
Learning rate of critic network	3×10^{-4}
Collision penalty hyperparameter ζ	5
Boundary penalty hyperparameter μ	5
Size of the replay buffer $\mathcal{D}$	1200
Batch size M	200
PPO epochs per update N_{ep}	3

The convergence performance of the proposed MAPPO-based algorithm is presented in Figure 5. To verify the superiority of the proposed MAPPO-based algorithm, we compare it with several state-of-the-art and representative RL algorithms, including PPO [30], deep deterministic policy gradient (DDPG) [31], multi-agent DDPG (MADDPG) [32], and soft actor–critic (SAC) [33]. For all algorithms, the maximum training episodes is set to 10,000. The MADDPG algorithm also utilizes the CTDE framework.

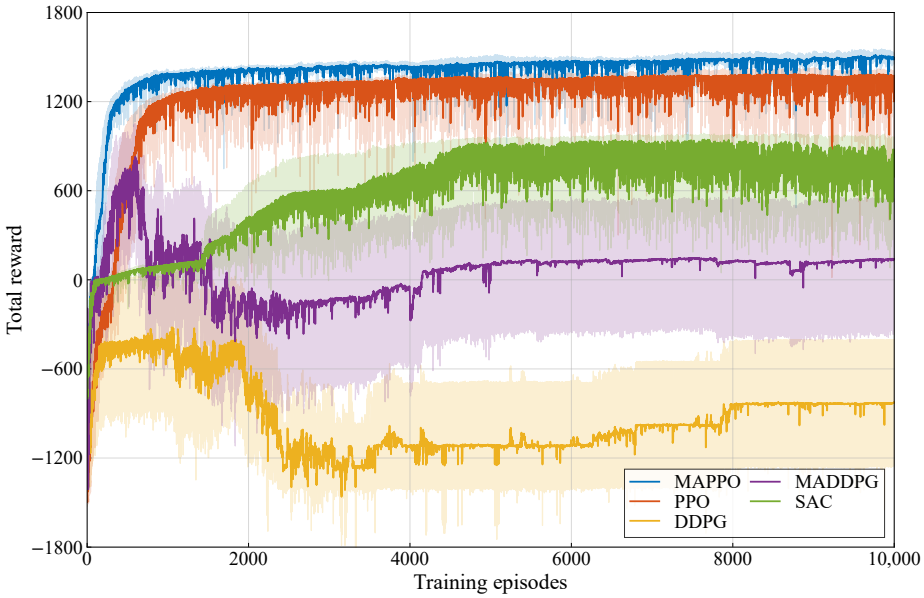


Figure 5. Convergence performance comparison of various algorithms.

It could be summarized from Figure 5 that the proposed MAPPO-based algorithm outperforms benchmark algorithms in terms of convergence speed, stationarity of training, and the final total reward, which suggests that it could obtain the highest time-averaged minimum secrecy rate. In addition, during the training process, each algorithm exhibited oscillations to varying degrees, particularly PPO and SAC. This is because the mobility of the eavesdropper and the outdated location information make it difficult for the agent to accurately grasp the state transition dynamics of the propagation environment and the action–value relationships.

In order to evaluate the isolated impact of the dynamic eavesdropper on security performance, we take the two best-performing algorithms in Figure 5, i.e., MAPPO and PPO, and the training curve indicating the difference between dynamic and static eavesdropper is shown in Figure 6. Furthermore, we quantitatively evaluate the performance sensitivity resulting from imperfect CSI by incorporating the AoD estimation error into Figure 6. The estimation error is modeled as $\hat{\phi} = \phi + \Delta\phi$, where ϕ is the actual angle and $\Delta\phi \sim \mathcal{N}(0, \sigma_\phi^2)$ represents the estimation error. Specifically, the standard deviation of the estimation error, σ_ϕ, is set to 1° and 3°.

Figure 6 suggests that as expected, the total reward corresponding to static eavesdropper is higher than that of the dynamic eavesdropper situation. Both MAPPO and PPO are capable of effectively handling scenarios with a dynamic eavesdropper involved, as their total rewards remain comparable to those achieved in static eavesdropper scenarios. For the proposed MAPPO-based algorithm, the dynamic eavesdropper scenario achieves a performance level nearly identical to that of the static eavesdropper scenario, while the latter exhibits faster convergence. This demonstrates the robustness of the proposed approach against mobile eavesdropping threats. Moreover, under imperfect CSI with angular estimation errors of 1° and 3°, the total rewards of the MAPPO algorithm only degrade by

5.04% and 18.44%, respectively, compared to the perfect CSI scenario. This indicates that, despite the loss in spatial nulling precision, the proposed algorithm remains robust and highly practical under such environmental uncertainties. From Figures 5 and 6, it can be observed that the proposed MAPPO-based algorithm achieves stable convergence, as the total reward consistently converges to steady-state values with minimal oscillation after the initial learning phase.

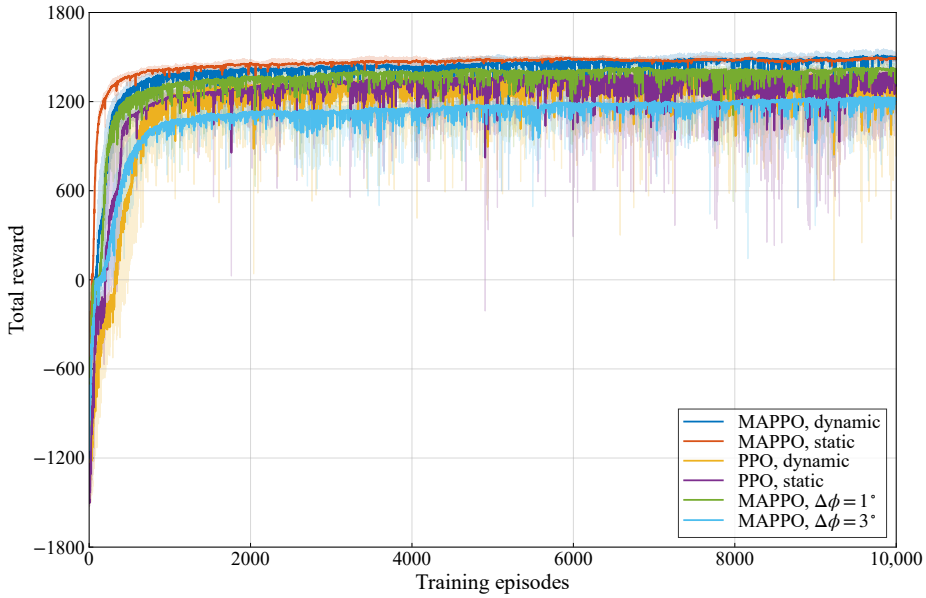


Figure 6. Verification of performance changes brought by dynamic eavesdropper.

To further validate the necessity of the joint optimization architecture, ablation studies are conducted by decoupling the trajectory and beamforming design. Two baselines are introduced: (1) Fixed Trajectory (FT-MAPPO): All UAVs move with a constant velocity of 1 m/s in the direction 45° from the positive x-axis, while solely optimizing the collaborative beamforming. (2) Fixed Beamforming (FB-MAPPO): All UAVs perform static beamforming directed toward the geometric center of all legitimate users, and only the trajectories are optimized. The convergence curves of the proposed MAPPO-based algorithm and the two baselines are presented in Figure 7.

As illustrated in Figure 7, the experimental results clearly demonstrate that decoupling these two aspects leads to significant degradation in the total reward. FT-MAPPO fails to effectively evade the dynamic eavesdropper via adaptive maneuvering, while FB-MAPPO lacks the spatial DoF required to achieve sharp nulling toward the eavesdropper. In contrast, joint optimization fully exploits the available DoF to counter dynamic threats. Specifically, when optimizing the trajectory alone (FB-MAPPO), the total reward converges to approximately 0, indicating that trajectory design by itself fails to yield a positive secrecy rate. This validates the critical importance of the joint optimization architecture in effectively maximizing the time-averaged minimum secrecy rate in the presence of a dynamic eavesdropper.

In the following, the optimized variables of optimization problem (15) through Algorithm 1 are presented and analyzed. The cooperative trajectories of the UAV swarm are illustrated in Figure 8, in which the path of the eavesdropper is also provided. Figure 8 illustrates that each UAV proceeds toward its optimized destination and hovers upon arrival to perform secure transmission. Meanwhile, the eavesdropper loiters in proximity to its initial location without approaching the legitimate users, as the UAV swarm adaptively counters its mobility strategy through the joint design of beamforming and trajectory.

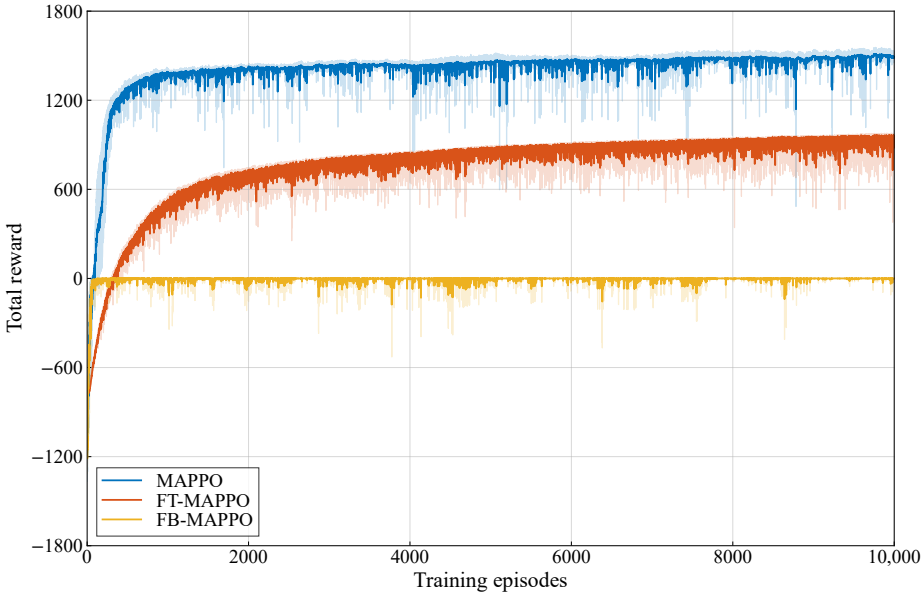
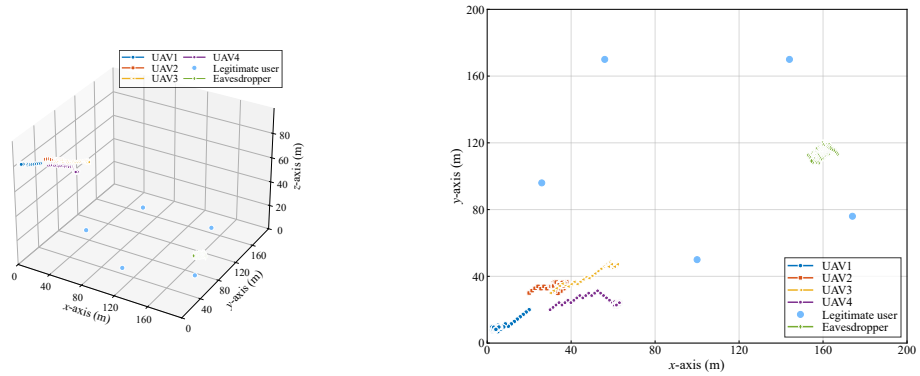


Figure 7. Ablation study of the action space comparing joint optimization and decoupled designs.

The transmit beampatterns of four UAVs at step $n = 300$ are plotted in Figure 9 on a decibel (dB) scale. The directions of the legitimate users are marked by green arrows, while the direction of the eavesdropper is indicated with a red arrow. The transmit gain of UAV u in direction ϕ is calculated as $G(\phi) = 20\lg(|\boldsymbol{w}_u^H[n]\mathbf{a}_{tx}(\phi)|^2)$. Figure 9 demonstrated that based on the proposed MAPPO-based algorithm, the UAV swarm achieves effective collaborative beamforming. Specifically, the main lobe of each beam is directed toward one or several legitimate users. For instance, the specific transmit gains directed toward the eavesdropper for UAVs 1 through 4 are explicitly suppressed to −33.8 dB, −23.0 dB, −29.7 dB, and −37.4 dB, respectively, resulting in an average gain of approximately −27 dB. This noticeable attenuation compared to the main lobes effectively degrades the quality of the eavesdropping channel. It is this effective collaborative beamforming that ensures secure transmission for legitimate users while simultaneously confusing the eavesdropper.



(**a**) (**b**)

Figure 8. Optimized trajectories of the UAV swarm and the moving path of the eavesdropper: (**a**) 3D version; (**b**) 2D version.

To more intuitively reveal the secrecy performance differences among the algorithms in Figure 5, the time-averaged minimum secrecy rate $R_s[n]$ calculated via $R_L[n]$ in terms of time slot n is depicted in Figure 10. Furthermore, the time-averaged average secrecy rate coming from $R'_s[n] = \frac{1}{L}\sum_{l=1}^{L}[R_l[n] - R_e[n]]^+$ is also presented as an additional reference. It is evident from Figure 10 that the proposed MAPPO-based algorithm outperforms the

other benchmark algorithms in terms of security performance indicated by secrecy rate. In addition, the fairness of service quality among all legitimate users could also be proved, as the time-averaged minimum secrecy rate is only 5.8% lower than the time-averaged average secrecy rate.

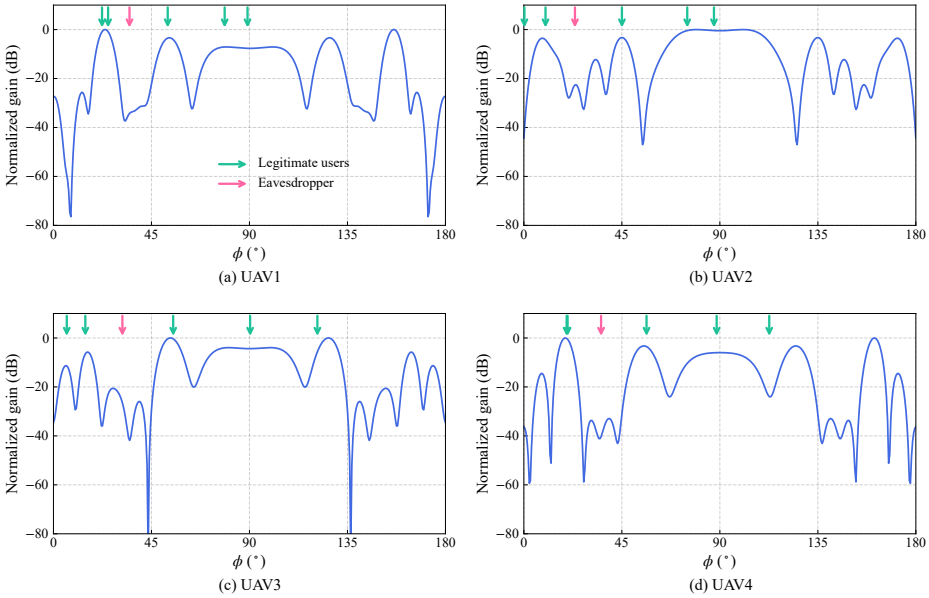


Figure 9. Transmit beampatterns of all UAVs.

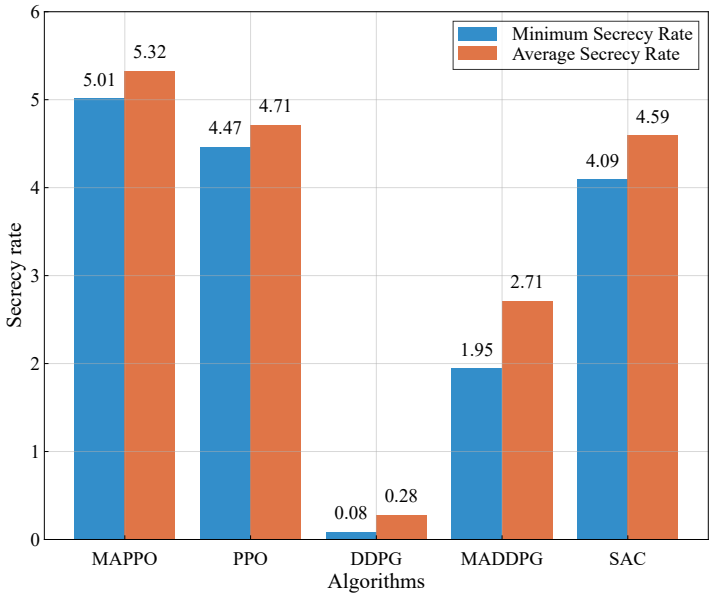


Figure 10. Time-averaged minimum and average secrecy rate comparison over different algorithms.

To thoroughly address the impacts of varied network scales and parameters, we verify the generalization ability and robustness of the proposed MAPPO-based framework. First, to enrich the diversity of the downlink communication mission, we evaluate the training convergence across five different scenarios, as illustrated in Figure 11: (1) four UAVs and five legitimate users (the fundamental setup); (2) five UAVs and five legitimate users; (3) six UAVs and five legitimate users; (4) six UAVs and six legitimate users; and (5) the fundamental setup with $P_t = 30$ dBm. The two additional UAVs are initially located at $[20, 40, 80]^T$ and $[30, 40, 80]^T$, respectively, while the extra legitimate user is positioned at $[110, 140, 0]^T$.

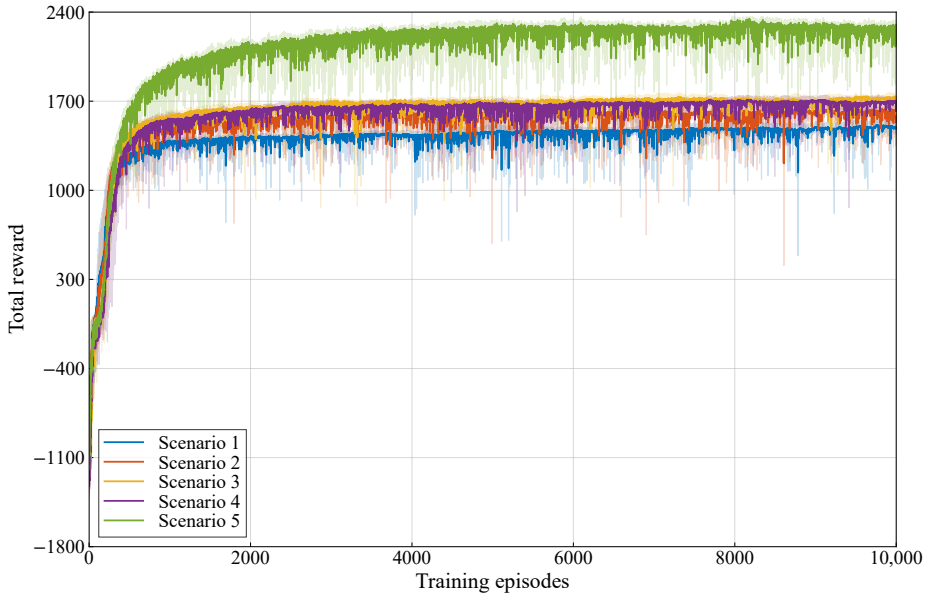


Figure 11. Convergence performance comparison among different transmission scenarios.

It can be observed from Figure 11 that the proposed algorithm maintains stable monotonic convergence trends across all configurations, demonstrating its scalability despite the expansion in agent numbers and state spaces, as well as varying network parameters. Furthermore, as the number of UAVs increases from 4 to 6, the total reward exhibits a consistent upward trend. This improvement stems from the enlarged DoF of the distributed antenna array formed by the UAV swarm, which enables more precise and higher-gain collaborative beamforming towards the legitimate user, thereby enhancing the received signal quality and secrecy performance. However, when the number of legitimate users increases from 5 to 6, i.e., transitioning from scenario 3 to scenario 4, a slight decrease in the total reward is observed. This is because, with a fixed number of UAVs, serving multiple geographically dispersed users forces the swarm to split its spatial resources, resulting in a reduction in effective DoF per user and potential beamforming trade-offs that compromise the optimal secrecy rate achievable for any single user. For scenario 5, a significant increase in the total reward is observed, highlighting the positive impact of the elevated transmit power on the system's secrecy performance. Accordingly, evaluations spanning UAV and user densities together with rising transmit power validate the algorithm's scalability and its robustness in dynamically managing distributed antenna resources amid growing network complexity.

Building upon these generalized scenarios, Figure 12 presents the impact of specific system parameters on the achieved minimum and average secrecy rates. We select the five representative scenarios illustrated in Figure 11 for this security performance evaluation. The results in Figure 12 confirm that increasing the number of UAVs by one yields a slight, yet positive, improvement in the achieved secrecy rate. This plateaued gain occurs because while the additional UAV provides extra spatial DoFs to enhance the legitimate channel via collaborative beamforming, the eavesdropper may also capture residual signals from this new transmission source. Conversely, when the transmit power budget is increased to 30 dBm, the minimum secrecy rate rises significantly by 50.7%. Collectively, these evaluations verify that the proposed MAPPO-based algorithm is highly generalized and scalable to dynamic variations in the size of the UAV swarm and hardware configurations. Interestingly, under a fixed number of UAVs, increasing the number of legitimate users leads to a marginal decline in secrecy performance, which corresponds to the observations in Figure 11. This degradation arises because the need to serve all users simultaneously

increases the difficulty of steering the collaborative beam away from the eavesdropper. Specifically, systematically varying the swarm size explicitly evaluates the impact of the total number of distributed transmit antennas, demonstrating how enriched spatial DoFs enhance the capabilities of secure collaborative beamforming.

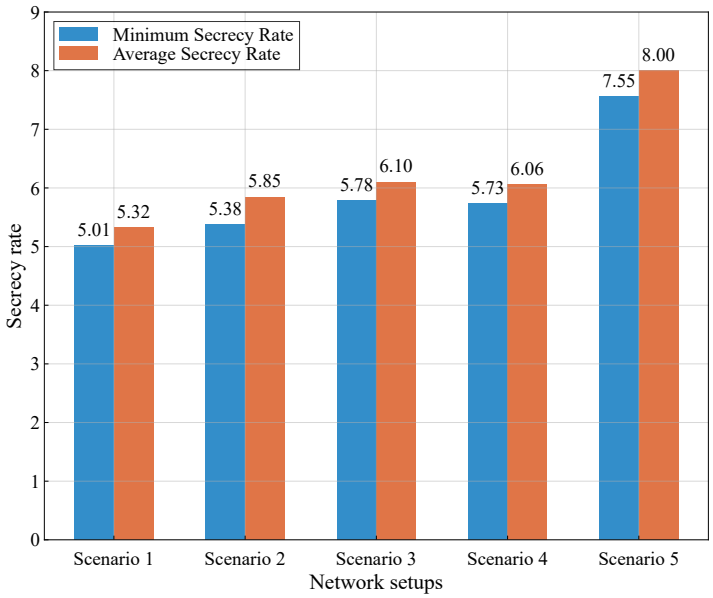


Figure 12. Time-averaged minimum and average secrecy rate comparison for different UAV numbers and transmit powers.

6. Conclusions

In this paper, the MARL-driven collaborative beamforming and cooperative trajectories planning framework against dynamic security threats was developed for a UAV swarm downlink transmission scenario. Multiple legitimate users receive the confidential signal from the UAV swarm in the presence of an active eavesdropper, which attempts to improve its eavesdropping quality through moving to a location exhibiting stronger signal receiving power. With the aim of maximizing the time-averaged minimum secrecy rate, a beamforming and trajectory joint optimization problem was formulated. In light of the challenges encountered in addressing this optimization problem through the utilization of traditional convex optimization-based methods, a framework founded upon MARL was developed. In the case that only the outdated eavesdropper location is known for the actor network, the optimization problem was transformed into a POMDP. Based on elaborately designed actor and critic networks, we develop an end-to-end MAPPO-based algorithm to solve the formulated problem. Its effectiveness in enhancing security performance and optimizing collaborative beamforming and trajectories is evaluated through comprehensive simulations. The results demonstrate that the proposed MAPPO-based framework significantly outperforms existing benchmarks, improving the time-averaged minimum secrecy rate by 12.1% compared with the PPO baseline. In order to address the challenge posed by persistently intelligent adversaries, future studies will concentrate on game-theoretic or multi-agent adversarial pursuit–evasion models for eavesdroppers. Furthermore, investigating robust strategies under more realistic imperfect CSI models is also an important direction for future work.

Author Contributions: Conceptualization, R.D. and J.Z.; methodology, B.W. and C.F.; software, J.T.; validation, R.D., J.J., and J.T.; formal analysis, R.D.; investigation, R.D.; resources, R.D., B.W., and C.F.; data curation, R.D.; writing—original draft preparation, R.D.; writing—review and editing, J.Z.;

visualization, R.D.; supervision, B.W.; project administration, R.D. and J.J.; funding acquisition, R.D., B.W., and J.J. All authors have read and agreed to the published version of this manuscript.

Funding: This research was funded by the National Natural Science Foundation of China under Grants 62301600 and 62472437 and by the China Postdoctoral Science Foundation under Grant 2025M784512.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author. The data are not publicly available due to privacy reasons.

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

Appendix A. Derivation of the Gradient $\nabla_{q_e} P_r$

To facilitate the gradient derivation of the received power P_r with respect to the eavesdropper's position $\boldsymbol{q}_e$, we first define the aggregated received signal vector (omitting the time slot index n for brevity) as

$$\boldsymbol{u} = \boldsymbol{H}_e^H \boldsymbol{w} = \sum_{u=1}^{U} \boldsymbol{H}_{u,e}^H \boldsymbol{w}_u. \tag{A1}$$

Consequently, the received power can be expressed as $P_r = ||\boldsymbol{u}||^2 = \boldsymbol{u}^H \boldsymbol{u}$. By applying the chain rule for complex-valued vector differentiation, the gradient with respect to the position $\boldsymbol{q}_e$ is derived as

$$\nabla_{\boldsymbol{q}_e} P_r = \frac{\partial (\boldsymbol{u}^H \boldsymbol{u})}{\partial \boldsymbol{q}_e} = 2\Re\left\{ \left(\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{q}_e} \right)^H \boldsymbol{u} \right\}, \tag{A2}$$

where the partial derivative of the aggregated signal vector is given by the summation of individual channel derivatives, i.e., $\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{q}_e} = \sum_{u=1}^{U} \frac{\partial \boldsymbol{H}_{u,e}^H}{\partial \boldsymbol{q}_e} \boldsymbol{w}_u$.

Recall the channel matrix $\boldsymbol{H}_{u,e} = \rho_{u,e} \mathbf{a}_{tx}(\phi_{u,e}) \mathbf{a}_{rx}^H(\theta_{u,e})$, for which its Hermitian transpose is $\boldsymbol{H}_{u,e}^H = \rho_{u,e} \mathbf{a}_{rx}(\theta_{u,e}) \mathbf{a}_{tx}^H(\phi_{u,e})$. Let $\boldsymbol{A}_u \triangleq \mathbf{a}_{rx}(\theta_{u,e}) \mathbf{a}_{tx}^H(\phi_{u,e})$, representing the spatial array response matrix incorporating both the AoD and AoA, and the derivative of the Hermitian channel matrix $\boldsymbol{H}_{u,e}^H = \rho_{u,e} \boldsymbol{A}_u$ with respect to $\boldsymbol{q}_e$ is formulated as follows:

$$\frac{\partial \boldsymbol{H}_{u,e}^H}{\partial \boldsymbol{q}_e} = \underbrace{\frac{\partial \rho_{u,e}}{\partial \boldsymbol{q}_e} \boldsymbol{A}_u}_{\text{Amplitude Variation}} + \underbrace{\rho_{u,e} \frac{\partial \boldsymbol{A}_u}{\partial \boldsymbol{q}_e}}_{\text{Phase Variation}}. \tag{A3}$$

Focusing first on the amplitude variation term, we calculate the derivative of the path loss. Given the Euclidean distance $d_{u,e} = ||\boldsymbol{q}_e - \boldsymbol{q}_u||$, the gradient flows through the distance function:

$$\frac{\partial \rho_{u,e}}{\partial \boldsymbol{q}_e} = \frac{\partial (\rho_0 d_{u,e}^{-\alpha})}{\partial d_{u,e}} \cdot \frac{\partial d_{u,e}}{\partial \boldsymbol{q}_e} = -\alpha \rho_0 d_{u,e}^{-\alpha-1} \frac{\boldsymbol{q}_e - \boldsymbol{q}_u}{d_{u,e}} = -\frac{\alpha \rho_{u,e}}{d_{u,e}^2} (\boldsymbol{q}_e - \boldsymbol{q}_u). \tag{A4}$$

It is worth noting that while $\boldsymbol{q}_e$ and $\boldsymbol{q}_u$ are defined in 3D space, the effective gradient update for the ground-based eavesdropper only utilizes the horizontal components (first two dimensions) of the resulting vector.

Subsequently, for the phase variation term in (A3), the matrix $\boldsymbol{A}_u$ depends on $\boldsymbol{q}_e$ through the AoD $\phi_{u,e}$ and AoA $\theta_{u,e}$; thus, we have

$$\frac{\partial \boldsymbol{A}_u}{\partial \boldsymbol{q}_e} = \frac{\partial \mathbf{a}_{rx}(\theta_{u,e})}{\partial \boldsymbol{q}_e} \mathbf{a}_{tx}^H(\phi_{u,e}) + \mathbf{a}_{rx}(\theta_{u,e}) \frac{\partial \mathbf{a}_{tx}^H(\phi_{u,e})}{\partial \boldsymbol{q}_e}. \tag{A5}$$

Generally, we compute the derivatives via the angle ψ, which represents either $\phi_{u,e}$ or $\theta_{u,e}$. By the chain rule, we have

$$\frac{\partial \mathbf{a}(\psi)}{\partial \boldsymbol{q}_e} = \frac{\partial \mathbf{a}}{\partial \psi} \cdot (\nabla_{\boldsymbol{q}_e} \psi)^T. \tag{A6}$$

For a ULA, the derivative with respect to the angle is $\frac{\partial \mathbf{a}(\psi)}{\partial \psi} = j\frac{2\pi d}{\lambda} \cos(\psi) \boldsymbol{D}\mathbf{a}(\psi)$, where $\boldsymbol{D} = \mathrm{diag}(0, 1, \ldots, N-1)$. For the conjugate transpose vector $\mathbf{a}_{tx}^H$, the derivative is the Hermitian transpose of the standard derivative, i.e., $\frac{\partial \mathbf{a}_{tx}^H}{\partial \psi} = (\frac{\partial \mathbf{a}_{tx}}{\partial \psi})^H$.

Assuming that the ULA is aligned with the x-axis, the angles are defined by geometry as $\cos\phi_{u,e} = \frac{x_e - x_u}{d_{u,e}}$ and $\cos\theta_{u,e} = \frac{x_u - x_e}{d_{u,e}}$. The geometric derivatives of these angles with respect to the horizontal coordinates of $\boldsymbol{q}_e$ are

$$\nabla_{\boldsymbol{q}_e} \phi_{u,e} = \frac{1}{d_{u,e}^2} \begin{bmatrix} -(y_e - y_u) \\ (x_e - x_u) \\ 0 \end{bmatrix}, \tag{A7}$$

$$\nabla_{\boldsymbol{q}_e} \theta_{u,e} = \frac{1}{d_{u,e}^2} \begin{bmatrix} -(y_u - y_e) \\ (x_u - x_e) \\ 0 \end{bmatrix}, \tag{A8}$$

which complete the extraction of (A6).

Finally, by substituting (A4) and the detailed angular derivatives back into (A3) and then into (A2), we obtain the complete analytical expression for the gradient $\nabla_{\boldsymbol{q}_e} P_r$, which is utilized by the eavesdropper to update its location.

References

1. Yao, J.; Yang, Z.; Yang, Z.; Xu, J.; Quek, T.Q.S. UAV-Enabled Secure ISAC Against Dual Eavesdropping Threats: Joint Beamforming and Trajectory Design. *IEEE Wirel. Commun. Lett.* **2025**, *14*, 3199–3203. [CrossRef]
2. Wang, C.; Tang, X.; Xiong, Z.; Zhai, D.; Zhang, R.; Niyato, D.; Han, Z. UAV-Assisted Federated Learning with Robust Resource and Trajectory Optimization under Location Uncertainties. *IEEE Trans. Cogn. Commun. Netw.* **2026**, *12*, 1282–1296. [CrossRef]
3. Shi, J.; Li, Z.; Sun, W.; Bai, Z.; Wang, F.; S. Quek, T.Q. Multi-agent SAC approach aided joint trajectory and power optimization for multi-UAV assisted wireless networks with safety constraints. *Chin. J. Aeronaut.* **2025**, *38*, 103368. [CrossRef]
4. Tian, J.; Wang, B.; Guo, R.; Wang, Z.; Cao, K.; Wang, X. Adversarial Attacks and Defenses for Deep-Learning-Based Unmanned Aerial Vehicles. *IEEE Internet Things J.* **2022**, *9*, 22399–22409. [CrossRef]
5. Wang, Z.; Yao, H.; Mai, T.; Xiong, Z.; Wu, X.; Wu, D.; Guo, S. Learning to Routing in UAV Swarm Network: A Multi-Agent Reinforcement Learning Approach. *IEEE Trans. Veh. Technol.* **2023**, *72*, 6611–6624. [CrossRef]
6. Shen, Y.; Qu, Y.; Dong, C.; Zhou, F.; Wu, Q. Joint Training and Resource Allocation Optimization for Federated Learning in UAV Swarm. *IEEE Internet Things J.* **2023**, *10*, 2272–2284. [CrossRef]
7. Pei, J.; Chen, H.; Tse, C.K. UAV Swarm for Connectivity Enhancement of Multiple Isolated Sensor Networks for Internet of Things Application. *IEEE Trans. Veh. Technol.* **2023**, *72*, 3914–3929. [CrossRef]
8. Li, S.; Dong, H.; Shan, C.; Fang, X.; Wu, W.; Li, Z. Secure Hybrid Beamforming Design for mmWave Integrated Sensing and Communication Systems. *IEEE Trans. Veh. Technol.* **2025**, *74*, 10622–10638. [CrossRef]
9. Dang-Ngoc, H.; Nguyen, D.N.; Ho-Van, K.; Hoang, D.T.; Dutkiewicz, E.; Pham, Q.V.; Hwang, W.J. Secure Swarm UAV-Assisted Communications with Cooperative Friendly Jamming. *IEEE Internet Things J.* **2022**, *9*, 25596–25611. [CrossRef]
10. Lv, L.; Fu, J.; Feng, Y.; Yang, L.; Nallanathan, A. ISAC-Enhanced Secrecy Communication: A Two-Stage Artificial Noise-Assisted Secure Beamforming Design. *IEEE Commun. Lett.* **2026**, *30*, 243–247. [CrossRef]
11. Zhang, Y.; Gang, Y.; Wu, P.; Fan, G.; Xu, W.; Ai, B.; Wu, Q. Integrated Sensing, Communication, and Computation in SAGIN: Joint Beamforming and Resource Allocation. *IEEE Trans. Cogn. Commun. Netw.* **2025**, *11*, 3128–3143. [CrossRef]
12. Lei, J.; Zhang, T.; Mu, X.; Liu, Y. NOMA for STAR-RIS Assisted UAV Networks. *IEEE Trans. Commun.* **2024**, *72*, 1732–1745. [CrossRef]
13. Wang, P.; Yang, H.; Han, G.; Yu, R.; Yang, L.; Sun, G.; Qi, H.; Wei, X.; Zhang, Q. Decentralized Navigation with Heterogeneous Federated Reinforcement Learning for UAV-Enabled Mobile Edge Computing. *IEEE Trans. Mob. Comput.* **2024**, *23*, 13621–13638. [CrossRef]

14. Dong, R.; Wang, B.; Cao, K.; Tian, J.; Cheng, T. Secure Transmission Design of RIS Enabled UAV Communication Networks Exploiting Deep Reinforcement Learning. *IEEE Trans. Veh. Technol.* **2024**, *73*, 8404–8419. [CrossRef]
15. Zhang, Z.; Wang, J.; Chen, J.; Fu, H.; Tong, Z.; Jiang, C. Diffusion-Based Reinforcement Learning for Cooperative Offloading and Resource Allocation in Multi-UAV Assisted Edge-Enabled Metaverse. *IEEE Trans. Veh. Technol.* **2025**, *74*, 11281–11293. [CrossRef]
16. Xie, Z.; Wang, Z.; Zhang, Z.; Wang, J.; Jiang, Z.; Han, Z. Distributed UAV Swarm for Device-Free Integrated Sensing and Communication Relying on Multi-Agent Reinforcement Learning. *IEEE Trans. Veh. Technol.* **2024**, *73*, 19925–19930. [CrossRef]
17. Ayub, M.S.; Rosa, R.L.; Koo, I. Alternating Optimization-Based Joint Power and Phase Design for RIS-Empowered FANETs. *Drones* **2026**, *10*, 66. [CrossRef]
18. Zhang, X.; Peng, M.; Liu, C. Sensing-Assisted Beamforming and Trajectory Design for UAV-Enabled Networks. *IEEE Trans. Veh. Technol.* **2024**, *73*, 3804–3819. [CrossRef]
19. Gupta, A.; Fernando, X. Performance Analysis of Unmanned Aerial Vehicle-Assisted and Federated Learning-Based 6G Cellular Vehicle-to-Everything Communication Networks. *Drones* **2025**, *9*, 771. [CrossRef]
20. Hu, Y.; Zhuo, X.; Meng, Z.; Wu, W.; Lu, W.; Tang, L.; Qu, F.; Bu, Z. Collaborative Positioning Optimization for Multiple Moving Users in UAV-Enabled ISAC. *IEEE Trans. Cogn. Commun. Netw.* **2025**, *11*, 3016–3030. [CrossRef]
21. Zhu, Q.; Liu, R.; Liu, Q.; Chen, C. Resource Allocation for UAV Swarm-Assisted Green ISAC Networks via Multi-Agent RL. *IEEE Trans. Green Commun. Netw.* **2025**, *9*, 1354–1367. [CrossRef]
22. Yang, B.; Dang, Y.; Taleb, T.; Shen, S.; Jiang, X. Sum Rate and Max-Min Rate for Cellular-Enabled UAV Swarm Networks. *IEEE Trans. Veh. Technol.* **2023**, *72*, 1073–1083. [CrossRef]
23. Wang, T.; Huang, X.; Wu, Y.; Qian, L.; Lin, B.; Su, Z. UAV Swarm-Assisted Two-Tier Hierarchical Federated Learning. *IEEE Trans. Netw. Sci. Eng.* **2024**, *11*, 943–956. [CrossRef]
24. Lv, J.; Cheng, J.; Li, P. Secure Energy Efficiency Maximization for IRS-Assisted UAV Communication: Joint Beamforming Design and Trajectory Optimization. *Drones* **2025**, *9*, 648. [CrossRef]
25. Wei, Z.; Liu, F.; Liu, C.; Yang, Z.; Ng, D.W.K.; Schober, R. Integrated Sensing, Navigation, and Communication for Secure UAV Networks with a Mobile Eavesdropper. *IEEE Trans. Wirel. Commun.* **2024**, *23*, 7060–7078. [CrossRef]
26. Zhao, M.; Wang, Z.; Guo, K.; Zhang, R.; Quek, T.Q.S. Against Mobile Collusive Eavesdroppers: Cooperative Secure Transmission and Computation in UAV-Assisted MEC Networks. *IEEE Trans. Mob. Comput.* **2025**, *24*, 5280–5297. [CrossRef]
27. Zheng, L.; He, J.; Zhu, X.; Zhang, Y.; Shen, Y.; Taleb, T. RSMA-Enabled Multi-UAV Secure Communication via MARL with Multi-Task Attention DRNN. *IEEE Trans. Inf. Forensics Secur.* **2025**, *20*, 12624–12639. [CrossRef]
28. Guo, L.; Jia, J.; Chen, J.; Yang, S.; Xue, Y.; Wang, X. RIS-Aided Secure A2G Communications With Coordinated Multi-UAVs: A Hybrid DRL Approach. *IEEE Trans. Netw. Sci. Eng.* **2024**, *11*, 4536–4550. [CrossRef]
29. Savkin, A.V.; Huang, H.; Ni, W. Securing UAV Communication in the Presence of Stationary or Mobile Eavesdroppers via Online 3D Trajectory Planning. *IEEE Wirel. Commun. Lett.* **2020**, *9*, 1211–1215. [CrossRef]
30. Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal Policy Optimization Algorithms. *arXiv* **2017**, arXiv:1707.06347. [CrossRef]
31. Lillicrap, T.P.; Hunt, J.J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D.; Wierstra, D. Continuous Control with Deep Reinforcement Learning. *arXiv* **2016**, arXiv:1509.02971.
32. Ryan, L.; Yi, W.; Aviv, T.; Jean, H.; Pieter, A.; Igor, M. Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments. In *Proceedings of the Neural Information Processing Systems (NIPS)*; Curran Associates, Inc.: Long Beach, CA, USA, 2017.
33. Haarnoja, T.; Zhou, A.; Abbeel, P.; Levine, S. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. *arXiv* **2018**, arXiv:1801.01290.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.