Resilient Path Planning for UAVs in Data Collection under Adversarial Attacks

Xueyuan Wang and M. Cenk Gursoy

***Abstract*—In this paper, we investigate jamming-resilient UAV path planning strategies for data collection in Internet of Things (IoT) networks, in which the typical UAV can learn the optimal trajectory to elude such jamming attacks. Specifically, the typical UAV is required to collect data from multiple distributed IoT nodes under collision avoidance, mission completion deadline, and kinematic constraints in the presence of jamming attacks. We first design a fixed ground jammer with continuous jamming attack and periodical jamming attack strategies to jam the link between the typical UAV and IoT nodes. Defensive strategies involving a reinforcement learning (RL) based virtual jammer and the adoption of higher SINR thresholds are proposed to counteract against such attacks. Secondly, we design an intelligent UAV jammer, which utilizes the RL algorithm to choose actions based on its observation. Then, an intelligent UAV anti-jamming strategy is constructed to deal with such attacks, and the optimal trajectory of the typical UAV is obtained via dueling double deep Q-network (D3QN). Simulation results show that both non-intelligent and intelligent jamming attacks have significant influence on the UAV's performance, and the proposed defense strategies can recover the performance close to that in no-jammer scenarios.**

***Index Terms*—UAV path planning, IoT networks, jamming attack, reinforcement learning.**

I. Introduction

Unmanned aerial vehicles (UAVs) or drones are aircrafts piloted by remote control or embedded computer programs without human onboard. Owing to their mobility, autonomy, and flexibility, UAVs are expected to be utilized extensively in different use cases in the next decade [1]. For example, they are considered as critical components in Internet of Things (IoT) scenarios [2], in which devices often have small transmit power and may not be able to communicate over a long range [3]. In such cases, UAVs can be used to assist IoT applications on e.g., data gathering [4], disaster mitigation and recovery [5]. Efficient trajectory control enables the UAV to achieve higher network performance with limited terrestrial infrastructure [6], [7]. However, the broadcast nature of wireless transmissions makes the UAV-enabled wireless communication systems vulnerable to jamming attacks [1], [8]–[10], leading to one of the major and serious threats to UAV-aided communications, especially when the jammer is mobile [11]. In addition, the trajectory control problem in hostile environments faces the following challenges, which make the UAV path planning hard to perform: i) multiple practical constraints should be jointly considered; ii) there exists uncertainty in the information on the jammer; and iii) malicious jamming makes the environment time-varying and non-stationary, especially when intelligent and mobile jammers are considered. Motivated by these observations, this paper investigates the effective jamming-resilient policies to safeguard UAV-enabled data collection networks by designing the UAV trajectory under multiple practical constraints in adversarial settings.

Xueyuan Wang is with School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou 213164, China (e-mail: xywang@cczu.edu.cn).

M. Cenk Gursoy is with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY, 13244 USA (e-mail: mcgursoy@syr.edu).

The material in this paper has been presented in part at the 2022 IEEE International Conference on Communications.

A. Related Works

Several studies have recently addressed ground jamming attacks in UAV-enabled networks. Particularly, in [9], a received signal strength based jammer localization algorithm is proposed to help the UAV plan its path. In [10], by exploiting the block coordinate descent (BCD) and successive convex approximation (SCA) techniques, an iterative algorithm was proposed to solve the anti-jamming three-dimensional UAV trajectory design problem. In [12], the authors considered multiple ground jammers in a multi-UAV path planning problem, and proposed two BCD based algorithms to obtain sub-optimal solutions with the aid of slack variables, SCA technique and S-procedure. In addition, the same method was constructed to solve the UAV path planning in a uplink communication system with turning and climbing angle constraints in [13]. The authors in [14] considered a UAV-enable relay network under malicious ground jamming attacks, BCD and SCA techniques were utilized to optimize the UAV trajectory and the transmit powers of both the UAV and the source node. In [15], by introducing the slack variables and leveraging the SCA technique, the authors designed a trajectory planning method to optimize the UAV's 3D position, and cases with a single jammer and also with multiple jammers were discussed. The authors in [16] investigated UAV swarm communication in the presence of jammers, an iterative algorithm was constructed based on BCD and SCA technique to optimize the UAVs' trajectories.

Above mentioned works all considered ground jammers and mainly utilized traditional optimization techniques to solve the UAV path planning problem. These optimization-based methods typically require prior knowledge of the jammer, and lack the ability to adapt to different jamming environments, e.g., in which the jammer's location is changed, or jamming is performed in a time-varying fashion. To tackle this challenge, the authors in [17] developed a reinforcement learning (RL)

based automatic flight control algorithm to perform UAV trajectory design in a coordinated satellite-UAV communication system in the presence of ground jamming attacks. The jammer launched attacks according to a jamming probability. However, this jammer was not smart either. The authors in [18] designed a deep Q-network (DQN) based UAV trajectory and power control scheme against attacks from a ground jammer, which could change its path and transmit power levels. But the jammer did not have an intelligent policy to adjust its condition to perform adaptive attacks.

Due to the expanded application of UAVs, malicious jamming attacks may also come from the sky. Consequently, UAV jamming attacks have also been considered recently in the literature. For instance, the authors in [19] formulated a zero-sum pursuit-evasion game to compute optimal trajectory strategies by a team of UAVs to evade the attack of an aerial jammer on the communication channel between UAVs. The authors in [20] considered a Bayesian Stackelberg game to formulate the competitive relations between UAVs and an aerial jammer, where the jammer and the UAVs aim to complete their missions by selecting their optimal power control strategies.

RL has also been utilized to obtain solutions against aerial jamming attacks. For example, in [21], the authors utilized the non-cooperative game theory to propose a Q-learning based power control algorithm to obtain an adaptive policy against a smart UAV jammer, which executes multiple attack types, such as eavesdropping, jamming, and spoofing. In [22], a static and smart attacker, which made subjective decisions to choose the attack types was taken into account, and DQN based UAV power allocation strategy was proposed against the attack. In [23], the ground users aimed to learn the optimal anti-jamming policy to protect its communication with a ground base station. The optimal jamming trajectory and user communication trajectory were obtained via deep recurrent Q-network and DQN, respectively. In [24], the authors considered a task-based anti-jamming scenario, in which a UAV swarm cooperated to detect a fixed ground target, and a cluster of UAV jammers cooperated to interfere with the area around the target. A knowledge-based RL algorithm was proposed for the UAV swarm to learn jamming-resilient trajectories. In [25], the authors considered a maritime communication scheme, which applied a UAV as relay to forward the message between ships against smart aerial jamming attacks. Q-learning and dueling neutral networks were utilized to select power control policies for the jammer and the UAV, respectively. The authors in [26] considered both a fixed jammer and a mobile UAV jammer with a fixed trajectory. A modified Q-learning algorithm based on multi-parameter programming was proposed for the UAVs to tune antenna beam to improve the overall communication quality.

In above mentioned related works, smart aerial jamming attacks were considered. However, none of the studies considered jamming attacks in UAV data collection networks with multiple practical constraints, e.g., collision avoidance constraint, kinematic constraint, communication constraint, flight duration constraint. An intelligent reflecting surface (IRS)-assisted UAV data collection network under malicious jamming was taken into account in [27]. An alternating optimization based algorithm was proposed by leveraging the Dinkelbach's algorithm, SCA, and BDC method, which requires prior knowledge of the jammer information. The jammer in this work was on the ground and non-intelligent. Also, collision avoidance and kinematic constraints were not taken into account. Note that different networks require very different algorithm designs.

B. Contributions

In this paper, different from prior studies, we consider a general noncooperative multi-UAV setting and address decentralized UAV trajectory designs for data collection in the presence of adversarial jamming attackers while also avoiding collision with other non-adversarial UAVs and considering multiple practical constraints. The main contributions are summarized as follows:

- A practical environment involving multiple constraints is considered. In particular, a practical setting that includes a fixed/mobile jammer and multiple non-cooperative and non-adversarial UAVs is addressed. Collision avoidance, mission completion deadline, kinematic, and transmission constraints are taken into account.
- A fixed ground jammer is designed with both continuous jamming attack and periodical jamming attack strategies to jam the link between the typical UAV and IoT nodes. Information on the jammer (e.g., its location) and the channel is unavailable to the typical UAV. Based on the UAV path planning algorithm proposed in [28], defensive strategies involving virtual jammers and higher SINR thresholds are proposed against both attack strategies.
- An RL-based intelligent UAV jammer is designed, by which the jammer follows the typical UAV and injects interference. Subsequently, an intelligent jamming-resilient strategy is constructed, with which the optimal trajectory of the typical UAV is devised via dueling double deep Q-network (D3QN) with designed state parameterization process. Sophisticated reward functions is designed to find the balance between the motion, mission and communication performance.

We further note that the proposed anti-jamming algorithms are completely based on observable data from the environment, which is more realistic than that in previous studies that assume the position of the jammer is fixed and known or part of the channel information is known. In addition, practical constraints including collision avoidance, mission completion deadline and kinematic constraints are taken into account even for the intelligent UAV jammer (in addition to the typical UAV), leading to more realistic and practical models.

The remainder of the paper is organized as follows: Section II provides the details of the considered system model. Section III introduces the ground jamming attack strategies and the RL-based anti-jamming algorithm. Section IV describes the intelligent mobile jamming attack algorithm. Section V presents the details of defense algorithm against the intelligent jamming attack. Section VI focuses on numerical and simulation results to evaluate the performance of the proposed algorithms. Finally, concluding remarks are provided in Section VII.

II. System Model

A. Network

We assume that the area of interest is a cubic volume, which can be specified by $\mathbb{C} : \mathbb{X} \times \mathbb{Y} \times \mathbb{Z}$ and $\mathbb{X} \triangleq [x_{\min}, x_{\max}]$, $\mathbb{Y} \triangleq [y_{\min}, y_{\max}]$, and $\mathbb{Z} \triangleq [z_{\min}, z_{\max}]$. There are multiple no-fly zones (obstacles) in the area through which UAVs cannot fly. And the no-fly zones are denoted as $\mathbb{N} : \mathbb{X}^N \times \mathbb{Y}^N \times \mathbb{Z}$. An illustration of the system model is provided in Fig. 1.

1) UAV: In the considered multi-UAV scenario, one UAV is chosen as the typical one, whose mission is to collect data from multiple ground IoT nodes. The UAV is modeled as disc-shaped with radius r. Let $\mathbf{p}^V = [p_x, p_y, H_V]$ denote the 3D position of the UAV, where H_V is the altitude of the UAV.

The typical UAV's information forms a vector that consists of the UAV's position, current velocity $\mathbf{v} = [v_x, v_y]$, radius r, destination $\mathbf{p}^D$, maximum speed $v_{\max}$, and orientation ϕ, i.e., $\mathbf{s}^V = [\mathbf{p}^V, \mathbf{v}, r, \mathbf{p}^D, v_{\max}, \phi] \in \mathbb{R}^{11}$. In this multi-UAV scenario, there are also J^o other non-cooperative and non-adversarial UAVs traveling within region $\mathbb{C}$. None of the UAVs communicate with each other. Therefore, the missions, destinations, movements, and decision-making policies of other UAVs are unknown. It is assumed that the typical UAV is equipped with low-cost sensors, with which it is able to sense the existence of other UAVs when they are closer than a certain distance. The circular sensing region is denoted by $\mathbb{O}$.

2) Jammer: One jammer also exists in the environment, which transmits jamming signals to interfere the links between the typical UAV and the IoT nodes. The jammer can be a ground jammer with height $H_J = 0$ or a moving UAV jammer with height H_J.

3) IoT Nodes: In this UAV-assisted network, there are N IoT nodes that need to upload finite amount data D_{n0}^L to the typical UAV via uplink transmission. The n^{th} node has transmit power P^n, and is located at ground position $\mathbf{p}^n = [p_{x_n}, p_{y_n}]$. The IoT nodes have two modes: active mode, if the node still has data to be transmitted; and silent mode, if data upload is completed.

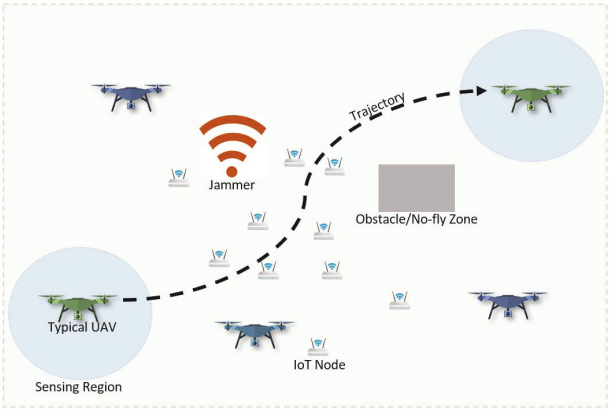


Fig. 1: An illustration of a UAV-assisted data collection network with a jammer, which can be either a ground jammer or a moving UAV jammer.

B. Channel Model

Due to the high UAV altitude, we assume that all links between the typical UAV and IoT nodes, and link between the typical UAV and the jammer are line-of-sight (LOS). Then, the path loss can be expressed as

$$L(d) = \left(d^2 + H^2\right)^{\alpha/2} \tag{1}$$

where α is the path loss exponent, d is the horizontal distance between the typical UAV and a node or a jammer, and H is the height difference between the typical UAV and the IoT node (for which case we have $H = H_V$) or a jammer (for which case we have $H = |H_V - H_J|$).

The IoT nodes and the jammer are assumed to have the omni-directional antenna gains of $G_n = 0$dB and $G_J = 0$dB, respectively. The UAVs are assumed to be equipped with a receiver with a horizontally oriented antenna, and a simple analytical approximation for antenna gain provided by UAV can be expressed as [29]

$$G_V(d) = \sin(\theta) = \frac{H}{\sqrt{d^2 + (H)^2}} \tag{2}$$

where θ is elevation angle between the UAV and a node. We note that even though specific antenna gains are considered for the sake of being concrete, the subsequent analysis is applicable to any type of antenna pattern.

C. Signal-to-Interference-plus-Noise Ratio (SINR)

The received signal from the n^{th} node to the typical UAV can be expressed as $P_n^r = P^n G_V(d_n) L^{-1}(d_n)$. With this, the SINR at the UAV if it is communicating with the n^{th} IoT node can be formulated as

$$S_n^V \triangleq \frac{P^n G_V(d_n) L^{-1}(d_n)}{\mathcal{N}_s + I^J} \tag{3}$$

where $\mathcal{N}_s$ is the noise power, and I^J is the interference from the jammer, which can be expressed as

$$I^J = P^J G_V(d_{JV}) L^{-1}(d_{JV}) \tag{4}$$

where P^J is the transmit power of the jammer, and d_{JV} is the horizontal distance between the UAV and the jammer.

D. Rate

The maximum achievable information rate if the typical UAV is connected with the n^{th} node is

$$R_n^{\max} = \log_2(1 + S_n^V). \tag{5}$$

To support data flows, UAV has to maintain a reliable communication link to the IoT nodes. To achieve this, it is assumed that the SINR at the UAV when connected with a node should be larger than a certain threshold $\mathcal{T}_s^V$. Then, the UAV can communicate with the node successfully. Otherwise, the UAV is not able to collect data from the node. Therefore, the effective information rate according to the SINR threshold $\mathcal{T}_s^V$ can be given as

$$R_n^V = \begin{cases} R_n^{\max}, & \text{if } S_n^V \geqslant \mathcal{T}_s^V, \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

E. Scheduling

Standard time-division multiple access (TDMA) model is adopted. Hence, the UAV can communicate with at most one node at each time. Using $q_n^V \in \{0,1\}$ to indicate the connection with the n^{th} node, we have

$$\sum_{n}^{N} q_n^V \leqslant 1. \tag{7}$$

The scheduling is according to the largest SINR strategy, meaning that the UAV is connected with the active node providing the largest S_n^V. We can mathematically express the scheduling strategy as

$$q_n^V = \begin{cases} 1, & \text{if } n = \underset{n' \in \{\text{active nodes}\}}{\operatorname{argmax}} S_{n'}^V, \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

A summary of the notations are provided in Table I.

TABLE I: Table of Notations

Notations	Description
$\mathbb{C}$, $\mathbb{N}$	Area of interest; no-fly zones/obstacles
$\mathbb{O}$	The typical UAV's sensing region
N, J^o	The number of IoT nodes; the number of other non-cooperative and non-adversarial UAVs
$\mathbf{p}^*$	Position, where $* \in \{V, J, n\}$ and V, J, n stand for the typical UAV, the jammer, and the n^{th} IoT node, respectively
$\mathbf{v}^*, \phi^*, r^*$	Velocity, orientation, radius, where $* \in \{V, J\}$
L, G_*	Path loss, and antenna gain, where $* \in \{V, J, n\}$
H_*	Flying altitude, where $* \in \{V, J\}$
D_n^L	The amount of data left at the n^{th} IoT node
P^*	Transmit power, where $* \in \{J, n\}$
P_n^r	Received signal power from the n^{th} IoT node
I^J	Interference from the jammer
S_n^V, R_n^V	SINR and effective information rate of the typical UAV when connected with the n^{th} IoT node
q_n^V	Connection indicator of the typical UAV with the n^{th} IoT node
$\mathcal{S}^*, \mathcal{A}^*, \mathcal{R}^*$	State space, action space, reward, where $* \in \{V, J\}$
T^*	Number of total time steps, where $* \in \{V, J\}$
$\mathcal{T}_s^V$	SINR threshold of the typical UAV
$\mathcal{T}_t^*$	Mission completion time threshold, where $* \in \{V, J\}$
$\mathcal{T}_r^*$	Maximum rotation angle in unit time duration, where $* \in \{V, J\}$
$v_{\max}^*$	Maximum speed, where $* \in \{V, J\}$
$\mathcal{N}_s$	The noise power
Δt	One time step duration
π^*	Policy, where $* \in \{V, J\}$
τ	The jamming period in periodic jamming attack strategy

III. Ground Jamming Attacks and Defenses

In this section, we consider a network with a fixed ground jammer, which is located on the ground at $\mathbf{p}^J$ and is assumed to have transmit power P^J and omni-directional antenna pattern with $G_J = 1 = 0$dB. Therefore, the interference from the jammer to the typical UAV can be expressed as $I^J = P^J \left(d_{JV}^2 + H_V^2\right)^{-\alpha/2} \frac{H_V}{\sqrt{d_{JV}^2 + H_V^2}}$. Different non-learning based attack strategies are designed for the jammer to jam the links between the typical UAV and the IoT nodes, and different defense strategies are also designed for the typical UAV against these jamming attacks.

A. Ground Jamming Attack Strategies

1) Continuous Jamming Attack Strategy: It is designed that the jammer transmits at a fixed transmit power P_l^J at a fixed location all the time.

2) Periodic Jamming Attack Strategy: It is designed that the jammer works periodically at a fixed location with relatively higher transmit power P_h^J, and the jamming time duration is τ_h^J seconds per minute. For fairness in the comparison with the continuous attack strategy, it is assumed that $P_l^J \times \tau_l^J = P_h^J \times \tau_h^J$, and $\tau_l^J = 60$s.

Note that the jammer's information and strategy are unknown to the typical UAV.

B. Problem Formulation for Defense

The goal of the typical UAV is to design efficient trajectories to maximize the collected data from the IoT nodes under several constraints in the presence of jamming attacks. Specifically, the optimization problem can be formulated as

$$\begin{aligned} (\text{PV}): \underset{\{\mathbf{p}_t^V, \forall t\}}{\operatorname{argmax}} \quad & \sum_{t=0}^{T^V} \sum_{n=1}^{N} q_{nt}^V \Delta t R_{nt}^V \\ s.t. \quad & ||\mathbf{p}_t^V - \mathbf{p}_{jt}||_2 > r^V + r_j, \forall j, \forall t && \text{(PV.a)} \\ & T^V \cdot \Delta t \leqslant \mathcal{T}_t^V && \text{(PV.b)} \\ & v_{s_t}^V \leqslant v_{\max}^V, \forall t && \text{(PV.c)} \\ & |\phi_t^V - \phi_{t-1}^V| \leqslant \Delta t \cdot \mathcal{T}_r^V, \forall t && \text{(PV.d)} \\ & \sum_{n}^{N} q_n^V \leqslant 1, \forall t && \text{(PV.e)} \\ & \mathbf{p}_0^V = \mathbf{p}_V^S, \mathbf{p}_T^V = \mathbf{p}_V^D, && \text{(PV.f)} \\ & \mathbf{p}_t^V \notin \mathbb{N}, \forall t, && \text{(PV.g)} \end{aligned}$$

where $\mathbf{p}_t^V$ is the typical UAV's position at t. In the above formulation, we have collision avoidance constraints in (PV.a) and (PV.g), which restrict that the distance between two UAVs should be large than the sum of their radii all the time and the typical UAV should not collide with the obstacles/no-fly zones. Mission completion deadline in (PV.b) requires the typical UAV to finish its mission in allowed time duration. Kinematic constraints in (PV.c) and (PV.d) show the maximum speed and maximum rotation angel in unit time duration limitations. (PV.e) is TDMA constraint, and (PV.f) indicates the start and destination locations constraint.

C. Reinforcement Learning Formulation

Typically, a sequential decision making problem can be formulated as a Markov decision process (MDP) [30], which can be described by tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma \rangle$, representing the state space, action space, state-transition model, the reward function, and a discount factor that trades off the importance

of the immediate and future rewards. Therefore, the trajectory optimization problem, as a sequential decision making problem, can be formulated as an MDP constructed as follows:

1) State Space $\mathcal{S}^V$: The state is $\mathbf{s}_t^{Vjn} = [\mathbf{s}_t^V, \mathbf{s}_t^o, \mathbf{s}_t^n, s_{tt}^V]$ with the following components:

- $\mathbf{s}_t^V = [\mathbf{p}^V, \mathbf{v}, r, \mathbf{p}^D, v_{\max}, \phi]$ is the typical UAV's full information vector at time step t.
- $\mathbf{s}_t^o = [[p_{x_{jt}}, p_{y_{jt}}, H_V, v_{x_{jt}}, v_{y_{jt}}, r_j] : j \in \{1, 2, ..., J_t^o\}]$ is the joint information vector of observed other non-cooperative and non-adversarial UAVs at the same height. $J_t^o \geqslant 0$ is the number observed other UAVs.
- $\mathbf{s}_t^n = [\mathbf{s}_{nt}^n : n \in \{1, ..., N\}]$ is the joint information vector of all IoT nodes. $\mathbf{s}_{nt}^n = [\mathbf{p}^n, D_{nt}^L, P_{nt}^r]$ consists of the location information $\mathbf{p}^n$, the amount of remaining data D_{nt}^L (which can be obtained from $D_{n,t-1}^L$, P_{nt}^r, and the scheduling parameter q_n^V), and the received signal power P_{nt}^r from each node.
- s_{tt}^V is the available time left for the given mission.

It's worth noting that information of the jammer is unknown.

2) Action Space $\mathcal{A}^V$: The action $\mathbf{a}^V$ is the index of each velocity in a velocity-set, which consists of permissible velocities sampled according to the kinematic constraints.

3) State-Transition Model $\mathcal{P}^V$: In an MDP, the state transition of an agent follows a Markov chain. Each agent takes action according to the current state, and then turns into next state after interacting with the environment. The transition probability distribution is related to the applied algorithm.

4) Reward $\mathcal{R}^V$: The reward function of the typical UAV in the considered scenario can be expressed as

$$\mathcal{R}_t^V = \mathcal{R}_{dt}^V + \mathcal{R}_{ct}^V + \mathcal{R}_{ot}^V + \mathcal{R}_{tt}^V + \mathcal{R}_{gt}^V + \mathcal{R}_{st}^V. \tag{9}$$

The first term $\mathcal{R}_{dt}^V$ is related to the data collected from the nodes during next time duration Δt, and can be expressed as

$$\mathcal{R}_{dt}^V = \alpha_1 \times \left(\sum_{n=1}^{N} D_{nt}^L - \sum_{n=1}^{N} D_{n,t+1}^L \right). \tag{10}$$

$\mathcal{R}_{ct}^V$ indicates the "repulsive force" from other agents, and is introduced to encourage the typical UAV to stay further away from others to avoid collision, It is given by

$$\mathcal{R}_{ct}^V = \begin{cases} -\alpha_2, & \text{if } d_{t_{\min}}^V \leqslant r^V + r_j, \\ -\alpha_2 \times (1 - \frac{d_{t_{\min}}^V - r^V - r_j}{d_b^V}), & \\ & \text{if } r^V + r_j < d_{t_{\min}}^V \leqslant d_b^V + r^V + r_j, \\ 0, & \text{otherwise}, \end{cases} \tag{11}$$

where $d_{t_{\min}}^V$ is the minimum distance from the typical UAV to other UAVs at the same height during next time duration Δt, and d_b^V is a constant that denotes the distance buffer, inside which the typical UAV will receive a penalty that depends on $d_{t_{\min}}^V$. $\mathcal{R}_{ot}^V$ is to penalize the collision with fixed obstacles or entering non-fly zones, and can be expressed as

$$\mathcal{R}_{ot}^V = \begin{cases} -\alpha_3, & \text{if } \mathbf{p}_{t+1}^V \in \mathbb{N}, \\ 0, & \text{otherwise}. \end{cases} \tag{12}$$

$\mathcal{R}_{tt}^V$ represents the "attractive force" from the destination to encourage the typical UAV to arrive at its destination within the allowed duration of time, and can be formulated as

$$\mathcal{R}_{tt}^V = \begin{cases} \alpha_4 \times (s_{t,t+1}^V - T_{g,t+1}^{V\min}), & \text{if } s_{t,t+1}^V < T_{g,t+1}^{V\min}, \\ 0, & \text{otherwise}, \end{cases} \tag{13}$$

where $s_{t,t+1}^V$ is the available time left for the given mission, $T_{g,t+1}^{V\min} = d_{g,t+1}^V / v_{\max}^V$ is the minimum time duration needed to reach destination, and $d_{g,t+1}^V$ is the distance to destination at time step $t+1$. R_{gt}^V is the reward given for arriving at the destination, and

$$\mathcal{R}_{gt}^V = \begin{cases} \alpha_5, & \text{if } \mathbf{p}_{t+1}^V = \mathbf{p}_V^D, \\ 0, & \text{otherwise}. \end{cases} \tag{14}$$

The last term $\mathcal{R}_{st}^V = -\alpha_6$ is a step penalty for each movement, and it is used to encourage fast arrival. Note that $\alpha_{1\sim6}$ are positive constants, and can be varied to adjust the weight or emphasis of each reward term to adapt to different mission priorities.

RL is a class of machine learning methods that can be utilized for solving sequential decision making problems with unknown state-transition dynamics [31] [32]. RL can also be utilized to develop a jamming-resilient method that does not need to model the environment [26]. Dueling double deep Q-network (D3QN) is a combination of dueling deep Q-network (DQN) and double deep Q-network (DDQN), and is a more effective and stable learning strategy. Thus, D3QN is used to learn the typical UAV path planning policy.

D. Anti-Ground-Jamming Strategies

Due to the uncertainty in jammer's location, the typical UAV's policy should be trained with consideration on the influence from jamming attacks. Therefore, two D3QN-based defense strategies are proposed as follows:

1) Defense with a Virtual Jammer in Training (VJ): With this strategy, we assume that a virtual jammer exists in the environment in training, which transmits all the time at a fixed transmit power $P^{J'}$ at location $\mathbf{p}^{J'}$. The location of the virtual jammer can be chosen arbitrarily or randomly (and this location does not need to match the location of the real jammer, whose knowledge is assumed to be not available in training). For instance, the virtual jammer's location can be chosen according to the distribution of the IoT nodes (e.g., the geometric center of the IoT node groups). Then, a policy can be learned in this environment with virtual jammer present.

2) Defense with Higher SINR Threshold (HST): As noted before, the transmission is reliable when the experienced SINR at the typical UAV is larger than a certain threshold, i.e., $S^V \geqslant \mathcal{T}_s^V$. With this defense strategy, we impose, in training, an SINR threshold $\mathcal{T}_s^{V'}$ that is larger than what is needed, i.e., $\mathcal{T}_s^{V'} > \mathcal{T}_s^V$, and we have the typical UAV learn a policy using this higher SINR threshold $\mathcal{T}_s^{V'}$. This leads to resiliency to increased interference inflicted by the jammer.

The main algorithm is provided in Algorithm 1.

Algorithm 1: UAV Path Planning Algorithm Against Ground Jamming Attacks

Input: $\mathcal{T}_s^V$, $\mathcal{T}_t^V$, $v_{\max}^V$, $\mathcal{T}_r^V$

Initialize replay memory $\mathcal{D}$
Initialize evaluation network $\boldsymbol{\xi}$ (including $\boldsymbol{\xi}^V$ and $\boldsymbol{\xi}^A$)
Initialize target network $\boldsymbol{\xi}^-$ (including $\boldsymbol{\xi}^{V-}$ and $\boldsymbol{\xi}^{A-}$) by copping from $\boldsymbol{\xi}$
$\mathcal{A}^V \leftarrow$ sampleActionSpact$(v_{\max}^V, \mathcal{T}_r^V)$
Choose defense strategy DS
if *DS is VJ* **then**
  $\mathbf{p}^{J'} \leftarrow$ randomGenerate in $\mathbb{C}$
else if *DS is HST* **then**
  $\mathcal{T}_s^{V'} \leftarrow \mathcal{T}_s^V + c_s$
else
  No defense
for *episode = 0: total episode N_e* **do**
  $\mathcal{E} \leftarrow$ resetEnvironment()
  while *not done* **do**
    $\mathbf{s}_t^{Vjn} \leftarrow$ observeEnvironment$(\mathcal{E})$
    $\widetilde{\mathbf{s}}_t^{Vjn} \leftarrow$ parameterizeState$(\mathbf{s}_t^{Vjn})$
    $c \leftarrow$ randomSample(Uniform (0,1))
    if $c \leqslant \epsilon$ **then**
      $\mathbf{a}_t^V \leftarrow$ randomSample$(\mathcal{A}^V)$
    else
      $\mathbf{a}_t^V \leftarrow \underset{\mathbf{a}^{V'} \in \mathcal{A}}{\operatorname{argmax}} Q(\widetilde{\mathbf{s}}_t^{jn}, \mathbf{a}^{V'}; \boldsymbol{\xi})$
    $\mathcal{R}_t^V, \mathbf{s}_{t+1}^{Vjn} \leftarrow$ executeAction$(\mathbf{a}_t^V, \mathbf{p}^{J'}$ or $\mathcal{T}_s^{V'})$
    $\widetilde{\mathbf{s}}_{t+1}^{Vjn} \leftarrow$ parameterizeState$(\mathbf{s}_{t+1}^{Vjn})$
    Uptate D with tuple $(\widetilde{\mathbf{s}}_t^{Vjn}, \mathbf{a}_t^V, \mathcal{R}_t^V, \widetilde{\mathbf{s}}_{t+1}^{Vjn})$
    Sample a minibatch of N_b tuples $(\mathbf{s}, \mathbf{a}, \mathcal{R}, \mathbf{s}') \sim$ Uniform(D)
    for *each tuple j* **do**
      Calculate target
      $$y_j = \begin{cases} \mathcal{R}, & \text{if } \mathbf{s}' \text{ is terminal,} \\ \mathcal{R} + \gamma Q(\mathbf{s}', \underset{\mathbf{a}'}{\operatorname{argmax}} Q(\mathbf{s}', \mathbf{a}'; \boldsymbol{\xi}); \boldsymbol{\xi}^-), & \text{o.w.} \end{cases}$$
    Do a gradient descent step with loss $E[(y_j - Q(\mathbf{s}, \mathbf{a}; \boldsymbol{\xi}))^2]$
    Update $\boldsymbol{\xi}^- \leftarrow \boldsymbol{\xi}$ every N_r steps
return ξ

IV. Intelligent Mobile Jamming Attack

In this section, we design an intelligent UAV jammer to jam the transmission between the typical UAV and the IoT nodes based on the observations from the environment.

A. UAV Jammer Model

In this setting, an intelligent UAV jammer is assumed to have transmit power P^J, height H_J, and certain departure and landing points. The jammer is equipped with sensors (e.g., radar) in order to sense nearby UAVs and track the typical UAV[1]. The jammer is further assumed to be able to eavesdrop/learn[2]: 1) location information of the active ground nodes assigned to the typical UAV; and 2) the typical UAV's continuous reference signal received power (RSRP) and reference signal received quality (RSRQ) reports.

If the jammer travels at the same height, it needs to avoid collision with the typical UAV while trying to get close to the UAV to increase the interference. In addition, if the sine antenna pattern of the typical UAV is adopted[3], it will not receive interference from the jammer (or the interference is really small) if the jammer travels at exactly the same height as the typical UAV. With this consideration, a strong jammer is designed to fly at a different height compared to the typical UAV. Then, the interference from the jammer can be formulated as

$$\begin{aligned} I^J &= P^J G_V(d_{JV})(d_{JV}^2 + (H_V - H_J)^2)^{-\frac{\alpha}{2}} \\ &= P^J |H_V - H_J| (d_{JV}^2 + (H_V - H_J)^2)^{-\frac{\alpha+1}{2}} \end{aligned} \tag{15}$$

where H_V and H_J are the heights of the typical UAV and the jammer, respectively.

B. Problem Formulation for Intelligent Attack

The objective of the jammer is to reduce the SINR of the typical UAV subject to collision avoidance constraints, maximum travel time constraint, kinematic constraints and the start and destination constraints, similar to what has been described in Section III-B for the typical UAV. We can formulate the optimization problem as

$$(\text{PJ}): \underset{\{\mathbf{p}_t^J, \forall t\}}{\operatorname{argmax}} \quad \mathbb{E}\left[\sum_{t=0}^{T^J} \sum_{n=1}^{N} q_{nt}^V \frac{1}{S_{nt}^V} \middle| \pi^V \right]$$

$$s.t. \quad ||\mathbf{p}_t^J - \mathbf{p}_{jt}||_2 > r^J + r_j, \forall j, \forall t \tag{PJ.a}$$

$$T^J \cdot \Delta t \leqslant \mathcal{T}_t^J \tag{PJ.b}$$

$$v_{s_t}^J \leqslant v_{\max}^J, \forall t \tag{PJ.c}$$

$$|\phi_t^J - \phi_{t-1}^J| \leqslant \Delta t \cdot \mathcal{T}_r^J, \forall t \tag{PJ.d}$$

$$\mathbf{p}_0^J = \mathbf{p}_J^S, \mathbf{p}_T^J = \mathbf{p}_J^D, \tag{PJ.e}$$

$$\mathbf{p}_t^J \notin \mathbb{N}, \forall t, \tag{PJ.f}$$

where the expectation $\mathbb{E}[]$ in the objective function is with respect to the typical UAV's decision making policy π^V, $\mathbf{p}_t^J$ is the position of the jammer at t, and T^J is the total flight time of the jammer. S_{nt}^V is the typical UAV's SINR if it is connected with the n^{th} IoT node at t, and q_{nt}^V is the association indicator of the typical UAV at time step t. Hence, in the above optimization problem, we have collision avoidance constraints in (PJ.a) and (PJ.f), mission completion deadline constraint in (PJ.b), kinematic constraints in (PJ.c) and (PJ.d), and start

[1]Note that this assumption can be realized in practice by equipping with low-cost sensors and radars.

[2]Note that the jammer, which is able to obtain these information, is a strong adversary, and consequently makes the defense more difficult. If the typical UAV can defend against this strong jammer, it can defend other jammers better. Therefore, this work considers the worst-case scenario and provides the corresponding defense strategies.

[3]Note that other antenna patterns can also be utilized, only leading to different formulation of the interference I^J.

and destination locations constraint in (PJ.e) for the intelligent UAV jammer.

C. Reinforcement Learning Formulation

The problem of trajectory design for the intelligent UAV jammer is also a sequential decision making problem, and thus can be formulated as an MDP and solved via RL. The tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{R} \rangle$ is formulated below.

1) State Space $\mathcal{S}^J$: In this network, the jammer can obtain the following information vectors:

- The full information vector of itself, $\mathbf{s}_t^J = [p_x^J, p_y^J, H_J, v_x^J, v_y^J, r^J, p_{gx}^J, p_{gy}^J, H_J, v_{\max}^J, \phi^J]$, at time step t.
- The observable information vector of other UAVs at H_J in its sensing region, i.e., $\mathbf{s}_t^{Jo} = [[p_{x_{jt}}, p_{y_{jt}}, H_J, v_{x_{jt}}, v_{y_{jt}}, r_j] : j \in \{1, 2, ..., J_t^{Jo}\}]$.
- The typical UAV's observable state $\mathbf{s}_t^V = [p_{x_t}^V, p_{y_t}^V, H_V, v_{x_t}^V, v_{y_t}^V, r^V]$.
- The location information of active IoT nodes, i.e., $\mathbf{s}_t^{Jn} = [[p_{x_n}^n, p_{y_n}^n] : n \in \{1, ..., N^c\}]$.
- The available time left for the jammer, s_{tt}^J.

The observed information vectors can be parameterized by following process:

- The first two information vectors are transformed into jammer-centric coordinates, in which the jammer's current location is the origin and the direction to the jammer's destination is the x-axis, i.e.,

$$\widetilde{\mathbf{s}}_t^J = [\tilde{v}_{x_t}^J, \tilde{v}_{y_t}^J, \tilde{p}_{gx_t}^J, \tilde{p}_{gy_t}^J, d_{g_t}^J, a_{g_t}^J, r^J, v_{\max}^J, \theta_t^J]$$
$$\widetilde{\mathbf{s}}_t^{Jo} = [[\tilde{p}_{x_{jt}}, \tilde{p}_{y_{jt}}, \tilde{v}_{x_{jt}}, \tilde{v}_{y_{jt}}, d_{jt}^o, a_{jt}^o, r_j], \text{ for } j \in \{1, 2, ..., J_t^{Jo}\}].$$

- The information vector of the typical UAV and the IoT nodes from the past τ time steps can be parameterized and utilized to learn the typical UAV's policy, i.e.,

$$\widetilde{\mathbf{s}}_t^V = [\tilde{p}_{x_t}^V, \tilde{p}_{y_t}^V, H_{VJ}, \tilde{v}_{x_t}^V, \tilde{v}_{y_t}^V]$$
$$\widetilde{\mathbf{s}}_{nt}^{Jn} = [[\tilde{p}_{x_n}, \tilde{p}_{y_n}, d_{nt}^V, a_{nt}^V] : n \in \{1, ..., N^c\}]$$

where N^c is the number of active nodes, d_{nt}^V, a_{nt}^V are the distance and azimuth angle of the n^{th} IoT node with respect to the typical UAV's location, and the node's information vector in $\widetilde{\mathbf{s}}_{nt}^n$ is listed in the smallest d_{nt}^V to the largest order. Then, we have

$$\widetilde{\mathbf{s}}_{nt}^V = [[\widetilde{\mathbf{s}}_{t'}^V, \widetilde{\mathbf{s}}_{nt'}^{Jn}], t' \in [t - \tau, t]].$$

The parameterized state vector of the jammer can be jointly expressed as

$$\widetilde{\mathbf{s}}_t^{Jjn} = [\widetilde{\mathbf{s}}_t^J, \widetilde{\mathbf{s}}_t^{Jo}, \widetilde{\mathbf{s}}_{nt}^V, s_{tt}^J]. \tag{16}$$

2) Action Space $\mathcal{A}^J$: Based on the jammer's kinematic constraints, permissible velocities can be sampled to build a velocity-set. The jammer's action a^J is the index of each velocity in the velocity-set.

3) Reward $\mathcal{R}^J$: The reward function of the jammer is designed based on the objective function and the constraints, i.e.,

$$\mathcal{R}_t^J = \mathcal{R}_{st}^J + \mathcal{R}_{ct}^J + \mathcal{R}_{ot}^J + \mathcal{R}_{tt}^J + \mathcal{R}_{gt}^J + \mathcal{R}_{dt}^J. \tag{17}$$

The first term is the related to the SINR experienced at the typical UAV, and it can be expressed as

$$\mathcal{R}_{st}^J = \begin{cases} \alpha_1^J \times \frac{1}{S_{t+1}^V} & \text{if } S_{t+1}^V > S_b^V \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

where S_{t+1}^V can be obtained from the typical UAV's RSRP and RSRQ reports, and S_b^V is a positive constant which is smaller than the SINR threshold. $\mathcal{R}_{ct}^J, \mathcal{R}_{ot}^J, \mathcal{R}_{tt}^J, \mathcal{R}_{gt}^J$ are the reward terms for collision avoidance, fixed obstacle avoidance, maximum travel time constraint, and arrival-to-the-destination goal, respectively, and are similar to the reward terms in (11), (12), (13) and (14), respectively. The last term, $\mathcal{R}_{dt}^J$, is a reward term based on the distance between the jammer and the typical UAV, and is formulated as

$$\mathcal{R}_{dt}^J = d_{JV_t} - d_{JV_{t+1}}. \tag{19}$$

D. Intelligent Jamming Attack Algorithm

The jammer's action space is sampled to be discrete, and thus Q value based RL algorithms, e.g., DQN, DDQN, D3QN, can be used to learn its policy. Since D3QN is more effective and stable, we choose D3QN to learn a strong jammer policy. The training procedure can be performed using Algorithm 1 by eliminating lines 5-11 and utilizing the designed $\mathcal{S}^J, \mathcal{A}^J, \mathcal{R}^J$ in Section IV-C.

V. Defense Against Intelligent Jamming Attack

In this section, we aim to design a defense algorithm against the intelligent jamming attacks.

A. Reinforcement Learning Formulation

The goal of the typical UAV is to maximize the collected data from all IoT nodes in the presence of intelligent jamming attacks, and the objective function is the same as in (PV). Due to the jammer's existence, state space $\mathcal{S}^V$ and reward function $\mathcal{R}^V$ described in Section III-C should be updated correspondingly.

1) State Space $\mathcal{S}^V$: Since the jammer injects interference and is generally close to the typical UAV, and the typical UAV is able to observe nearby UAVs in its sensing region, we assume that the typical UAV is able to detect the jammer all the time[4]. Therefore, the location information of the jammer, $\mathbf{p}_t^J$, can be obtained by the typical UAV. The jammer's locations in the past τ time steps can be used to estimate the jammer's next movement, and therefore we have

$$\widetilde{\mathbf{s}}_t^J = [\mathbf{p}_{t'}^J, t' \in [t - \tau, t]].$$

[4]The assumption can be removed. Discussions are provided in Section VI-C-3).

The observed information vectors $\mathbf{s}_t^V, \mathbf{s}_t^o, \mathbf{s}_t^n$ (described in Section III-C) can be transformed into typical UAV-centric coordinates and parameterized into

$$\begin{aligned}
\widetilde{\mathbf{s}}_t^V &= [\tilde{v}_{x_t}^V, \tilde{v}_{y_t}^V, \tilde{p}_{gx_t}^V, \tilde{p}_{gy_t}^V, d_{g_t}^V, a_{g_t}^V, r^V, v_{\max}^V, \theta_t^V] \\
\widetilde{\mathbf{s}}_{jt}^o &= [\tilde{p}_{x_{jt}}, \tilde{p}_{y_{jt}}, \tilde{v}_{x_{jt}}, \tilde{v}_{y_{jt}}, d_{jt}^o, a_{jt}^o, r_j], \\
&\qquad\qquad \text{for } j \in \{1, 2, ..., J^c\} \\
\widetilde{\mathbf{s}}_{nt}^n &= [\tilde{p}_{x_{nt}}, \tilde{p}_{y_{nt}}, d_{nt}^n, a_{nt}^n, D_{nt}^L, P_{nt}^r], \quad \text{for } n \in \{1, ..., N^c\}.
\end{aligned}$$

Therefore, the state of the typical UAV is updated to

$$\widetilde{\mathbf{s}}_t^{Vjn} = [\widetilde{\mathbf{s}}_t^V, [\widetilde{\mathbf{s}}_{jt}^o, j \in \{1, ..., J^c\}], [\widetilde{\mathbf{s}}_{nt}^n, n \in \{1, ..., N^c\}], \widetilde{\mathbf{s}}_t^J, s_{tt}^V]. \tag{20}$$

2) Reward $\mathcal{R}^V$: To encourage the typical UAV to fly away from the jammer, an additional reward term is added to the original reward function in (9), which is

$$\mathcal{R}_{Jt}^V = \begin{cases} -\alpha_7 \times (1 - \frac{d_{JV_{t+1}}}{d_{b2}^V}) & \text{if } d_{JV_{t+1}} \leqslant d_{b2}^V \\ 0, & \text{otherwise} \end{cases} \tag{21}$$

where $d_{JV_{t+1}}$ is the distance between the typical UAV and the jammer at the next time step $t+1$, and d_{b2}^V is the distance buffer which essentially defines the safe distance between the typical UAV and the jammer.

B. Defense Against Intelligent Attack Algorithm

With the modified state and reward functions, the typical UAV's policy can be retrained using modified Algorithm 1.

VI. Simulation Results

In this section, we provide simulation results to show the performance of ground/mobile jamming attack strategies and the defense strategies. We choose the following performance metrics: 1) success rate (SR), which is the portion of successful trajectories among all trajectories (and a successful trajectory means that the typical UAV arrives at its destination within mission completion deadline without collisions); 2) data collection rate (DR), which is the percentage of collected data within successful trajectories; 3) arriving on time rate (TR); and 4) collision rate (CR), and a collision event occurs when the typical UAV collides with any of the other UAVs in the environment. In the figures, we use yellow areas to show the reliable transmission region, inside which the UAVs can achieve $S^V \geqslant \mathcal{T}_s^V$. The blue triangles are the IoT nodes, and the red triangle is the jammer. The blue and green areas are the departure and landing areas, respectively. The destination of the typical UAV is denoted as a black cross in the landing area. The gray areas are the fixed obstacles or no-fly zones. In the figures of trajectories, black-doted lines and red doted lines display the trajectories of the typical UAV and other non-cooperative and non-adversarial UAVs, respectively. The orange-doted lines depict the trajectories of the mobile jammer. The typical UAV flies at 50m, and the transmit power of the IoT node is 10 dBm (10^{-2}W). Other UAVs are assumed to use optimal reciprocal collision avoidance (ORCA) [33] in choosing actions and determining their trajectories.

The typical UAV's policy is designed as a three-layer DNN of size (256, 256, 128), and the jammer's policy is designed as a two-layer DNN of size (256, 128). In the DNNs, ReLU function is used as the activation function, and batch-normalization is used for each layer. Adam optimizer is used to update the parameters with learning rate 0.0003. Batch size is set to be 256, and the regularization parameter is 0.0001. The exploration parameter ϵ decays linearly from 0.5 to 0.1. The replay memory capacity is 1000000.

A. Continuous Jamming Attack Scenario

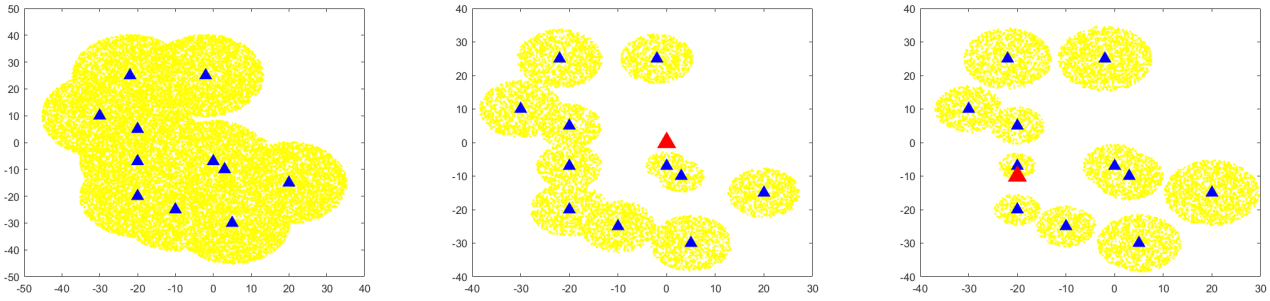
In this subsection, the jammer is located at a fixed location and transmits at a fixed power level all the time. The transmit power of the jammer is $P_l^J = 10^{-3}/3$W, and the SINR threshold for the typical UAV is $\mathcal{T}_s^V = 3.5$.

1) Attack Performance: Fig. 2 depicts the reliable transmission regions when the jammer is absent (in Fig. 2(a)) and the jammer is located at different locations (Figs. 2(b) and 2(c)). We immediately notice that the existence of the jammer significantly reduces the reliable transmission region, and different jammer locations have varying impact.

Table II provides the attack performance in testing when the jammer is located at different locations $\mathbf{p}^J$ (as well as when it is absent). Note that the numerical results are averaged over 5000 testing episodes, and in each episode, the number of IoT nodes is randomly chosen from $N \in [5, 10]$, the number of other UAVs is $J = 2$, the locations of nodes, the start and destination points of the typical UAV, and the start and destination points of other UAVs are randomly generated. From Table II, we observe that the existence of the jammer substantially reduces the typical UAV's reward, SR and TR, and slightly reduces the DR. The decline in SR and TR indicates that the typical UAV needs more time to arrive at its destination. This performance degradation is due to two reasons: 1) because of the reduction in the reliable transmission region (i.e., yellow areas in the figures), the UAVs needs to get closer to each IoT node to collect data successfully, leading to a longer trajectory and longer mission completion time; and 2) the interference inflicted by the jammer changes the SINR, and this change makes the UAV get confused and not choose the optimal actions, leading to longer trajectories as well. The slight decrease in DR means that in trajectories with successful arrivals, the typical UAV can still collect the vast majority (over 96%) of the data in the presence of a fixed jammer. Overall, we can state that the jammer prevents the UAV from completing its mission to a certain extent.

TABLE II: Performance of continuous jamming attack.

	SR(%)	DR(%)	TR(%)	CR(%)	Reward
No-Jammer	99.4	100	100	0.6	81.9
$\mathbf{p}^J$=(0,0)	84.8	98.1	85.3	0.5	18.66
(20,20)	95.1	98.7	96	0.9	41.89
(-20,30)	97.2	99.2	97.7	0.5	53.44
(-20,-10)	90.8	99.8	91.2	0.4	7.61
(10,-10)	87.2	96.1	87.8	0.6	24.83



(a) No jammer (b) Jammer at (0,0) (c) Jammer at (-20,-10)

Fig. 2: Illustrations of reliable transmission region when the jammer is absent or is located at different locations with $P_l^J = 10^{-3}/3$W.

2) Performance with Defense Utilizing a Virtual Jammer: Now, we deploy defensive measures and assume that, in training, a virtual jammer is located at $\mathbf{p}^{J'} = (0, 0)$ on the ground with transmit power $P^{J'} = 10^{-3}/3$W. The performance results in testing (achieved by the learned policy in the presence of the virtual jammer) are presented in Table III. From this table, we observe that the SR, TR and DR can be recovered close to those in the no-jammer scenario. On the other hand, since the typical UAV needs to fly closer to the IoT nodes to get reliable connection in the presence of a jammer, the trajectories become longer. Thus, the reward with the defense strategy is still smaller than that of the no-jammer case, since we introduce a negative reward term R_{st}^V for each step (as noted at the end of Section III-C).

TABLE III: Performance with defense strategy using a virtual jammer.

	SR(%)	DR(%)	TR(%)	CR(%)	Reward
No-Jammer	99.4	100	100	0.6	81.9
$\mathbf{p}^J$=(0,0)	98.8	99.8	99	0.3	68.69
(20,20)	98.6	99.9	98.8	0.1	58.58
(-20,30)	99.1	100	99.6	0.5	73.84
(-20,-10)	98.7	99.8	99.1	0.4	66.8
(10,-10)	98.8	99.9	99.3	0.5	66.18

3) Performance with Defense Using a Higher SINR Threshold: With this defense strategy, we assume that the SINR threshold is $\mathcal{T}_s' = 3.9$ in training, while $\mathcal{T}_s = 3.5$ in testing. The testing performance is presented in Table IV. With this strategy, the SR, TR and DR performances can be recovered close to those of the no-jammer case, and we can observe performances similar to those of the defense with the virtual jammer.

4) Influence of the Transmit Power Levels : We also present the attack and defense results in Table V considering different transmit powers for the IoT nodes P^n and the jammer P^J. From Table V, we observe that if P^n is larger, the jammer expectedly needs to transmit at a high power level to have better attack performance. With the proposed strategy, i.e., defense with higher SINR threshold, the typical UAV can successfully defend the attacks and recover the performance.

TABLE IV: Performance with defense strategy using a higher SINR threshold.

	SR(%)	DR(%)	TR(%)	CR(%)	Reward
No-Jammer	99.4	100	100	0.6	81.9
$\mathbf{p}^J$=(0,0)	97.7	100	99.5	0.8	67.46
(20,20)	98.2	99.8	98.5	0.3	49.49
(-20,30)	98.8	99.9	99.3	0.4	65
(-20,-10)	98.8	99.9	99.1	0.3	64.83
(10,-10)	98.8	99.9	99.3	0.4	61.34

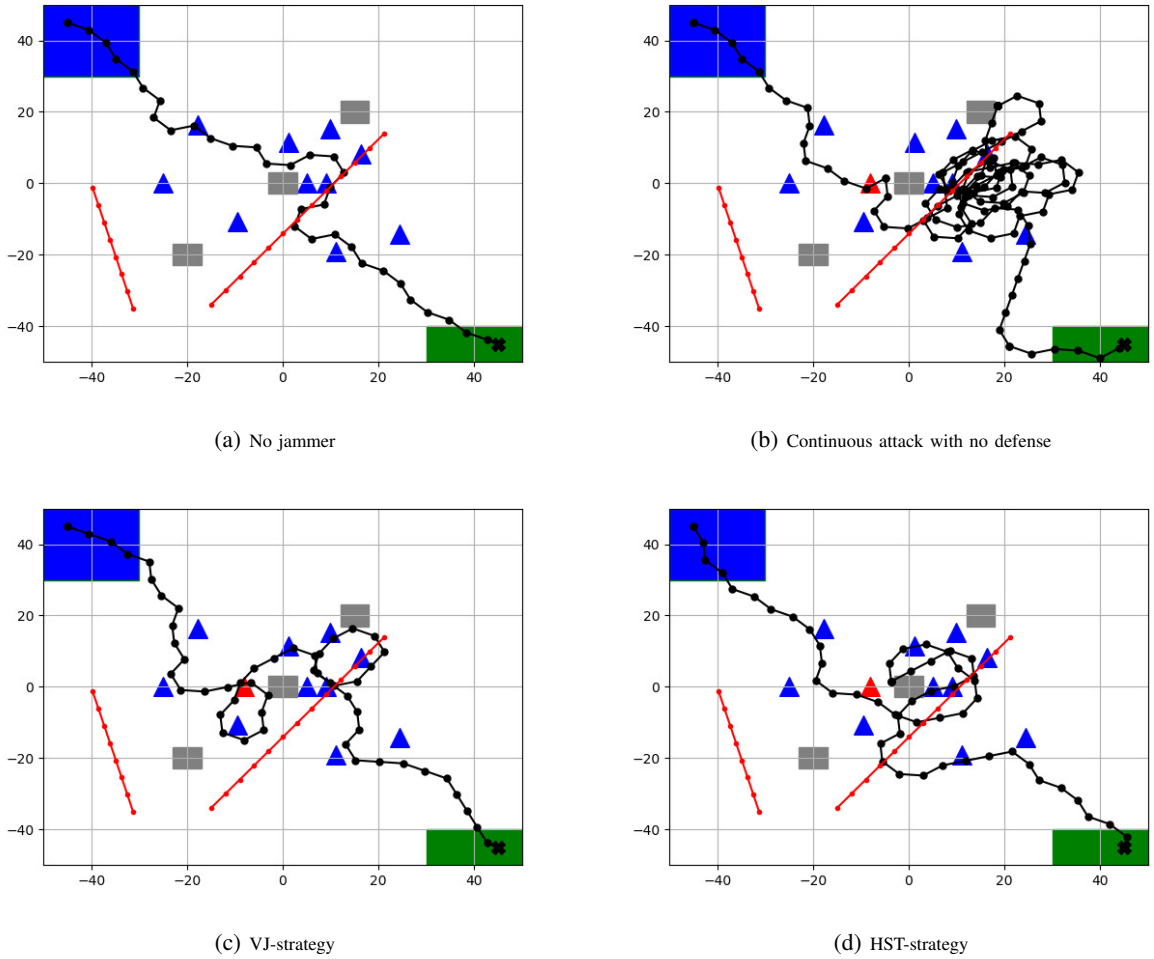
TABLE V: Jamming attack performance and defense performance when $\mathbf{p}^J$=(0,0).

		SR(%)	DR(%)	Reward
$P^n = 10^{-1.8}$	No-Jammer	99.8	100	126.09
	$P^J = 10^{-3}$	91	96.8	31.8
	$P^J = 2 * 10^{-3}$	85.4	93.3	-1.21
	Defense	99	99.7	110.6
$P^n = 10^{-1.6}$	No-Jammer	99.6	100	124.27
	$P^J = 2 * 10^{-3}$	90.7	98	40.9
	$P^J = 4 * 10^{-3}$	86.2	96.6	3.68
	Defense	98.6	99.8	106.3

5) Trajectory Designs: Fig. 3 presents the UAV trajectories in no-jammer, continuous jamming attack when $P_l^J = 10^{-3}/3$W (with no defense), virtual jammer defense strategy (VJ-strategy), and higher SINR threshold defense strategy (HST-strategy) scenarios. Fig. 3(a) shows that the typical UAV can find an efficient trajectory to complete its mission when there is no jammer. Fig. 3(b) shows that the typical UAV trajectory becomes curvy (with several loops) due to the existence of the jammer at (-8,0). Figs. 3(c) and (d) demonstrate that with the two defense strategies, the typical UAV is able to complete its mission in shorter trajectories under continuous jamming attacks, while the trajectories are still relatively longer than that in the no-jammer scenario.

B. Periodic Jamming Attack Scenario

In this subsection, it is assumed that the jammer interferes periodically with transmit power P_h^J and jamming period τ_h^J.



(a) No jammer

(b) Continuous attack with no defense

(c) VJ-strategy

(d) HST-strategy

Fig. 3: Examples of typical UAV trajectory in different scenarios.

We adopt two transmit power levels $P_{h1}^J = 10^{-3}/2.5$W and $P_{h2}^J = 10^{-3}/2$W. Since $P_l^J \times 60 = P_h^J \times \tau_h^J$, we have $\tau_{h1}^J = \frac{P_l^J \times 60}{P_{h1}^J} = 50$s and $\tau_{h2}^J = 40$s. In other words, the jammer transmits with P_{h1}^J for 50s and becomes silent for 10s per minute, or transmits with P_{h2}^J for 40s and becomes silent for 20s per minute. Fig. 4 illustrates examples of the reliable transmission region when the jammer has different transmit powers. As we see from the figures, the larger the transmit power is, the greater influence the jammer exerts, e.g., when $P_{Jh}^2 = 10^{-3}/2$W, majority of the connections are blocked during jamming. Table VI provides the performances of the periodic jamming attack and two aforementioned defense strategies. The results in the table indicate that the influence of the periodic jamming attack is not significant, due to the reason that the typical UAV is able to wait for the jammer to become silent and then collect data from the IoT nodes. However, overall the SR and TR are still reduced due to the longer mission completion time caused by waiting. Overall, the collision rate is under 0.6%, thus is not listed in the table. Using the proposed defense strategies, the performance can again be recovered to levels close to those in the no-jammer scenario.

TABLE VI: Periodic jamming attack performance and defense performance.

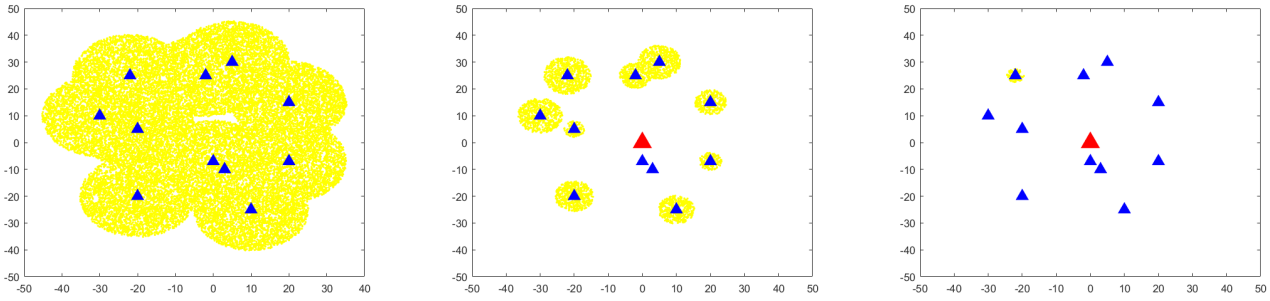
		SR(%)	DR(%)	TR(%)	Reward
No-Jammer		99.4	100	100	81.904
Attack	$P^J = \frac{10^{-3}}{2}$	99.1	99.3	99.5	72.355
	$P^J = \frac{10^{-3}}{2.5}$	96.9	99.3	97.3	50.257
Defense	$P^J = \frac{10^{-3}}{2}$	99.6	100	99.9	74.269
	$P^J = \frac{10^{-3}}{2.5}$	99.1	100	99.8	63.151

C. Intelligent Jamming Attack Scenarios

1) Original Policies of the Typical UAV: The typical UAV has two original policies learned in the no-jammer scenarios for different mission completion deadlines $\mathcal{T}_t^V$, and these policies are denoted as

- π_1^V when $\mathcal{T}_t^V = 100$s;
- π_2^V when $\mathcal{T}_t^V = 200$s.

The performances of these two original policies in a no-jammer scenario are provided in Table VII. From the table, we observe that the SR is at least 98.9% and DR is close to 100%, indicating that the typical UAV can complete its



(a) No jammer (b) $P_{h1}^J = 10^{-3}/2.5$W (c) $P_{h2}^J = 10^{-3}/2$W

Fig. 4: Illustrations of the reliable transmission region when the jammer is absent or is located at (0,0) with different transmit powers.

mission with 98% success rate if no jammer exists. In addition, with the looser mission completion deadline of $\mathcal{T}_t^V = 200$s, the overall performance can be increased. Note that since the reward function is modified within the defense algorithm, we do not compare the reward performances in this section.

TABLE VII: Performance of typical UAV's policies in the absence of jamming attacks.

	SR(%)	DR(%)	TR(%)	CR(%)
π_1^V	98.9	99.9	99.4	0.4
π_2^V	99.4	100	100	0.6

2) Intelligent Jamming Attack Performance: In this subsection, the jammer flies at height $H_J = 30$m with transmit power $P^J = 10^{-3}/3$W or 10^{-4}W. Four jammers are trained to attack the typical UAV with different transmit power levels P^J, and these jammers are described in more detail below:

- Jammer 1 (J1) is trained to attack π_1^V with transmit power $P^J = 10^{-3}/3$W, and its policy is denoted by π_1^J;
- Jammer 2 (J2) is trained to attack π_2^V with transmit power $P^J = 10^{-3}/3$W, and its policy is denoted by π_2^J;
- Jammer 3 (J3) is trained to attack π_1^V with transmit power $P^J = 10^{-4}$W, and its policy is denoted by π_3^J;
- Jammer 4 (J4) is trained to attack π_2^V with transmit power $P^J = 10^{-4}$W, and its policy is denoted by π_4^J.

The attack performances of the jammers are provided in Table VIII. By comparing the typical UAV's SR, DR, and TR in the no-jammer scenario (provided in Table VII) and in the presence of different jammers (provided in Table VIII), we observe that each jammer can significantly reduce the SR, DR and TR. The substantial decrease in DR is due to the reason that the jammer is encouraged to get close to the typical UAV and thus interference from the jammer can be very large, leading to the result that most connections are blocked and the typical UAV fails to collect data from some nodes. This is also the reason that the DR in the presence of jammers J1 and J2 (where $P^J = 10^{-3}/3$W) is much smaller than DR in the presence of jammers J3 and J4 (where $P^J = 10^{-4}$W). In addition, the SR and TR decrease due to the following two reasons: 1) the reliable transmission region is substantially reduced with the existence of the jammer, and thus the typical UAV needs more time to collect data from the nodes; and 2) the reliable transmission region is dynamically changing due to the movement of the jammer, and that leads the typical UAV not to choose the optimal action and generally need more time to arrive at its destination, and therefore violating its mission completion deadline. These are also the reasons for why SR and TR when $\mathcal{T}_t^V = 100$s (with jammers J1 and J3 in Table VIII) are smaller than SR and TR when $\mathcal{T}_t^V = 200$s (with jammers J2 and J4 in Table VIII).

TABLE VIII: Performance of typical UAV in the presence of different jammers.

Jammer	SR(%)	DR(%)	TR(%)	CR(%)
J1	0.7	13.8	1.3	0.7
J2	7.6	5.3	8.1	0.5
J3	2.1	49.8	2.6	0.5
J4	33.7	77.8	34.4	0.7

3) Defense Performance: Using the proposed defense algorithm with updated state space and reward function, policies can be re-trained against the intelligent jammers. To defend against the jammers designed in the previous subsection, we re-train the typical UAV's policy. The re-trained policies are listed and described below:

- π_1^{Vd} is trained with the existence of J1, i.e., π_1^{Ud} is trained to defend against J1;
- π_2^{Vd} is trained to defend against J2;
- π_3^{Vd} is trained to defend against J3;
- π_4^{Vd} is trained to defend against J4.

Since the typical UAV needs more time to finish its mission due to the significant reduction in the reliable transmission region, we loosen the mission completion deadline $\mathcal{T}_t^V$ in defensive strategies. The performances of defense polices are provided in Table IX. From the rows in boldface (in which we have the performance results of the retrained policies against the corresponding jamming attackers), we observe that the performance of the typical UAV in terms of SR, DR and TR is considerably restored. More specifically, the DR is recovered to above 80% when defending against J1, J3, and J4, and above 70% when defending against J2. Also, the SR and TR are recovered to above 94%. The reasons for this significant improvement are the following: 1) with loosened mission completion deadline, the typical UAV is allowed to

use more time to collect data from the IoT nodes; 2) with the presence of the jammer in training, the UAV learns the dynamically varying reliable transmission regions; and 3) the typical UAV's policy is updated and re-trained, and thus the jammer cannot predict the typical UAV's movement well.

In addition, we also use the policies π_2^{Vd} and π_4^{Vd} to defend against other jammers (with respect to which the defensive policies have not been retrained). It is observed that both policies can recover the performance to some extent, especially when using π_2^{Vd}. Even though the performances are generally not as good as the case of using the matching defense policy, the SR and TR are above 80% and DR is above 70%, which are much higher than the performance without any defense.

TABLE IX: Performances of defense policies in the intelligent jamming attack scenarios.

	Jammer	SR(%)	DR(%)	TR(%)	CR(%)
π_1^{Vd}	**J1**	**94.7**	**87.7**	**95.6**	**0.9**
π_2^{Vd}	J1	94.8	82.1	94.6	1.1
	J2	**94.8**	**71.5**	**95.4**	**0.6**
	J3	98.8	89.2	99.4	0.6
	J4	96.9	78.4	98.1	1.1
π_3^{Vd}	**J3**	**98.1**	**84.6**	**98.5**	**0.3**
π_4^{Vd}	J1	83	88.3	83.9	0.9
	J2	81.9	79.8	82.7	0.8
	J3	87.2	84.1	88.2	0.9
	J4	**98.4**	**82.3**	**99.1**	**0.7**

The proposed algorithm can be extended and utilized in more realistic scenarios, e.g., in a scenario in which the typical UAV is not able to detect the jammer all the time. Particularly, if the jammer is in sensing region $\mathbb{O}$, its position, velocity and orientation can be sensed. Otherwise, the typical UAV fails to sense this information. In this more practical setting, in order to predict the jammer's information, a velocity filter is designed to obtain the estimated next velocity $\hat{\mathbf{v}}_{t+1}^J$ using the jammer's velocities in the past τ time steps, and the estimated velocity can now be expressed as

$$\hat{\mathbf{v}}_{t+1}^J = \frac{1}{\tau} \sum_{t'=t-\tau}^{t} \mathbf{v}_{t'}^J, \qquad \text{if } \mathbf{p}_{t+1}^J \notin \mathbb{O} \tag{22}$$

where $\mathbf{p}_{t+1}^J$ is the jammer's position at time step $t+1$. Then, the next estimated position is $\hat{\mathbf{p}}_{t+1}^J = \mathbf{p}_t^J + \hat{\mathbf{v}}_{t+1}^J \times \Delta t$ and the next estimated orientation is $\hat{\phi}_{t+1} = \arctan \hat{v}_{t+1,y}^J / \hat{v}_{t+1,x}^J$. Fig. 5 plots the reward values in training when considering the scenario with assumption-1 (in which the jammer's information is detected all the time) and the scenario with assumption-2 (in which the jammer's information is estimated using the velocity filter if it is outside $\mathbb{O}$). From Fig. 5, we observe that we can achieve comparable performance in the scenario with assumption-2 compared to that with assumption-1. This follows from two reasons. First, due to the kinematic constraints, sudden drastic changes in the jammer's velocity are not allowed, and thus the past movements provide relatively accurate indications on its near-term future mobility. Secondly, if the jammer is outside of the typical UAV's sensing region $\mathbb{O}$, it is far away from the typical UAV, and correspondingly its interference is small, leading to small influence on the typical UAV's transmission. Thus, the estimation error does not impact the typical UAV's performance substantially.

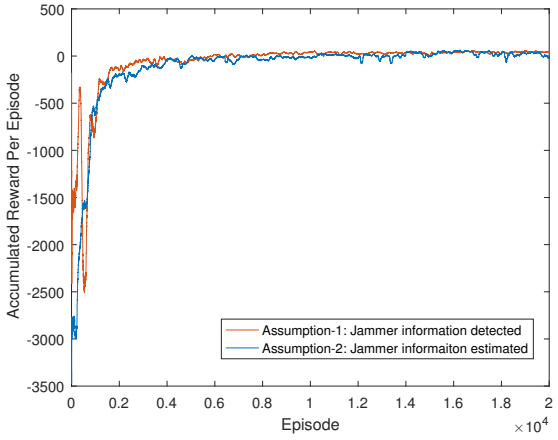


Fig. 5: Comparison of accumulated reward per episode between two scenarios. In the scenario with assumption-1, the jammer's information can be detected all the time, while in the scenario with assumption-2, the jammer's information is estimated when it is outside of $\mathbb{O}$.

In the literature, Q-learning (e.g., [21], [26]) and DQN (e.g., [22], [23]) have been used to defend against jamming attacks. Due to the large size of the state space, Q-learning is typically infeasible to be used in such studies. Figure 6 depicts the reward when utilizing DQN, DDQN and D3QN to train the typical UAV's defense policy. It can be observed that D3QN is more rewarding and converges much faster. Since the dueling architecture is able to learn which states are valuable without learning the effect of each action for each state, it has the ability to identify the correct action more quickly during policy evaluation [34].

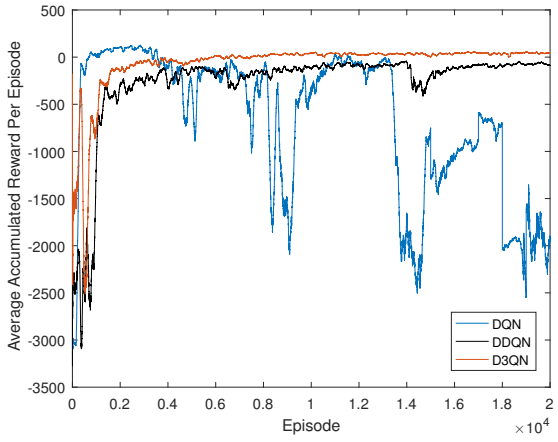
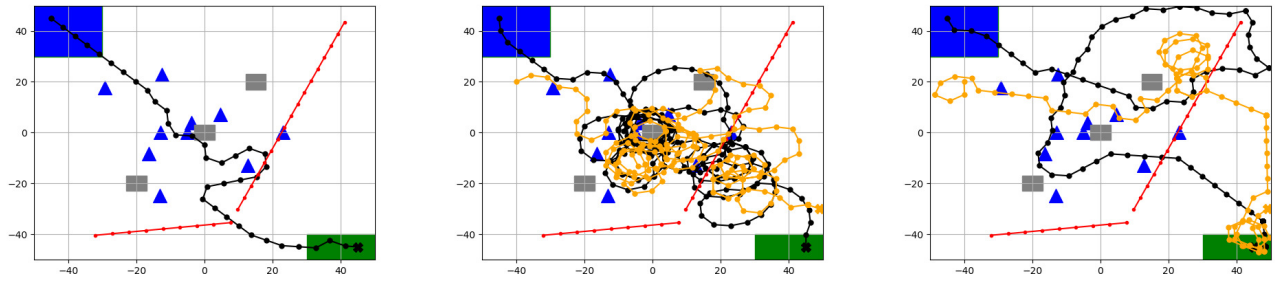


Fig. 6: Comparison of accumulated reward per episode among DQN, DDQN and D3QN.

4) Trajectory Designs: We provide examples of UAV trajectories in Fig. 7 in the no-jammer scenario, in the scenario with intelligent jamming attack and no defense, and in the scenario in which defensive policy is employed. Note that the



(a) No jammer (b) Intelligent jamming attack (with J2) (c) Defense (with π_2^{Vd})

Fig. 7: Examples of typical UAV trajectory in different scenarios.

orange-dotted lines are the intelligent jammer's trajectories. In Fig. 7(b), we observe that the intelligent jammer can follow the typical UAV closely if the UAV does not implement the defense policy, and the jammer makes the typical UAV trajectory really curvy and long (compared with the trajectory in Fig. 7(a) where no jammer exists). In addition, Fig. 7(c) shows that if the defense strategy is utilized and the typical UAV's policy is updated, the intelligent jammer is not able to follow the UAV well. Therefore, the typical UAV can find an efficient trajectory to complete its mission (e.g., a short trajectory not exceeding the mission completion deadline, and being close to the IoT nodes but away from the jammer in order to collect data). This observation further verifies the effectiveness of the proposed and implemented defensive measures.

VII. Conclusion

In this paper, we have investigated jamming-resilient UAV path planning strategies for data collection in IoT networks, in which the typical UAV can learn the optimal trajectory to elude such jamming attacks. Specifically, the typical UAV is required to collect data from multiple distributed IoT nodes under collision avoidance, mission completion deadline, and kinematic constraints in the presence of jamming attacks. We have first designed a fixed ground jammer with continuous jamming attack and periodic jamming attack strategies to inject interference into the link between the typical UAV and IoT nodes. RL-based defensive strategies that utilize a virtual jammer and adopt a higher SINR threshold are proposed against these attacks. Secondly, we have designed an intelligent UAV jammer, which uses an RL algorithm to choose actions based on its observation. Finally, an intelligent UAV anti-jamming strategy is developed to defend against such intelligent jamming attacks. The optimal trajectory of the typical UAV is obtained via D3QN. Simulation results have shown that both fixed jamming and intelligent UAV jamming attacks have significant influence on the typical UAV's performance, and the proposed defense strategies can recover the performance close to that in the no-jammer scenario.

References

[1] Y. Zeng, Q. Wu, and R. Zhang, "Accessing from the sky: A tutorial on UAV communications for 5G and beyond," *Proceedings of the IEEE*, vol. 107, no. 12, pp. 2327–2375, 2019.

[2] F. Syed, S. K. Gupta, S. Hamood Alsamhi, M. Rashid, and X. Liu, "A survey on recent optimal techniques for securing unmanned aerial vehicles applications," *Transactions on Emerging Telecommunications Technologies*, vol. 32, no. 7, p. e4133, 2021.

[3] M. Mozaffari, W. Saad, M. Bennis, Y. Nam, and M. Debbah, "A tutorial on UAVs for wireless networks: Applications, challenges, and open problems," *IEEE Communications Surveys Tutorials*, pp. 1–1, 2019.

[4] S. Goudarzi, N. Kama, M. H. Anisi, S. Zeadally, and S. Mumtaz, "Data collection using unmanned aerial vehicles for Internet of Things platforms," *Computers & Electrical Engineering*, vol. 75, pp. 1–15, 2019.

[5] S. H. Alsamhi, O. Ma, M. S. Ansari, and F. A. Almalki, "Survey on collaborative smart drones and Internet of Things for improving smartness of smart cities," *IEEE Access*, vol. 7, pp. 128 125–128 152, 2019.

[6] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Mobile unmanned aerial vehicles (UAVs) for energy-efficient Internet of Things communications," *IEEE Transactions on Wireless Communications*, vol. 16, no. 11, pp. 7574–7589, 2017.

[7] N. C. Coops, T. R. Goodbody, and L. Cao, "Four steps to extend drone use in research," 2019.

[8] D. Wang, B. Bai, W. Zhao, and Z. Han, "A survey of optimization approaches for wireless physical layer security," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1878–1911, 2018.

[9] B. Duan, D. Yin, Y. Cong, H. Zhou, X. Xiang, and L. Shen, "Anti-jamming path planning for unmanned aerial vehicles with imperfect jammer information," in *2018 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 2018, pp. 729–735.

[10] B. Duo, Q. Wu, X. Yuan, and R. Zhang, "Anti-jamming 3D trajectory design for UAV-enabled wireless sensor networks under probabilistic LoS channel," *IEEE Transactions on Vehicular Technology*, 2020.

[11] D. Darsena, G. Gelli, I. Iudice, and F. Verde, "Detection and blind channel estimation for UAV-aided wireless sensor networks in smart cities under mobile jamming attack," *IEEE Internet of Things Journal*, vol. 9, no. 14, pp. 11 932–11 950, 2022.

[12] Y. Wu, W. Fan, W. Yang, X. Sun, and X. Guan, "Robust trajectory and communication design for multi-UAV enabled wireless networks in the presence of jammers," *IEEE Access*, vol. 8, pp. 2893–2905, 2019.

[13] Y. Gao, Y. Wu, Z. Cui, H. Chen, and W. Yang, "Robust design for turning and climbing angle-constrained UAV communication under malicious jamming," *IEEE Communications Letters*, vol. 25, no. 2, pp. 584–588, 2020.

[14] Y. Wu, W. Yang, X. Guan, and Q. Wu, "UAV-enabled relay communication under malicious jamming: Joint trajectory and transmit power optimization," *IEEE Transactions on Vehicular Technology*, 2021.

[15] H. Wang, J. Chen, G. Ding, and J. Sun, "Trajectory planning in UAV communication with jamming," in *2018 10th International Conference on Wireless Communications and Signal Processing (WCSP)*. IEEE, 2018, pp. 1–6.

[16] Y. Wu, X. Guan, W. Yang, and Q. Wu, "UAV swarm communication under malicious jamming: Joint trajectory and clustering design," *IEEE Wireless Communications Letters*, vol. 10, no. 10, pp. 2264–2268, 2021.

[17] C. Han, A. Liu, K. An, H. Wang, G. Zheng, S. Chatzinotas, L. Huo, and X. Tong, "Satellite-assisted UAV trajectory control in hostile jamming environments," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 4, pp. 3760–3775, 2021.

[18] Z. Lin, X. Lu, C. Dai, G. Sheng, and L. Xiao, "Reinforcement learning based UAV trajectory and power control against jamming," in *International Conference on Machine Learning for Cyber Security*. Springer, 2019, pp. 336–347.

[19] S. Bhattacharya and T. Başar, "Game-theoretic analysis of an aerial jamming attack on a UAV communication network," in *Proceedings of the 2010 American Control Conference*. IEEE, 2010, pp. 818–823.

[20] Y. Xu, G. Ren, J. Chen, Y. Luo, L. Jia, X. Liu, Y. Yang, and Y. Xu, "A one-leader multi-follower Bayesian-Stackelberg game for anti-jamming transmission in UAV communication networks," *IEEE Access*, vol. 6, pp. 21 697–21 709, 2018.

[21] C. Li, Y. Xu, J. Xia, and J. Zhao, "Protecting secure communication under UAV smart attack with imperfect channel estimation," *IEEE Access*, vol. 6, pp. 76 395–76 401, 2018.

[22] L. Xiao, C. Xie, M. Min, and W. Zhuang, "User-centric view of unmanned aerial vehicle transmission against smart attacks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 4, pp. 3420–3430, 2017.

[23] N. Gao, Z. Qin, X. Jing, Q. Ni, and S. Jin, "Anti-intelligent UAV jamming strategy via deep Q-networks," *IEEE Transactions on Communications*, vol. 68, no. 1, pp. 569–581, 2019.

[24] Z. Li, Y. Lu, X. Li, Z. Wang, W. Qiao, and Y. Liu, "UAV networks against multiple maneuvering smart jamming with knowledge-based reinforcement learning," *IEEE Internet of Things Journal*, 2021.

[25] K. Liu, P. Li, C. Liu, L. Xiao, and L. Jia, "UAV-aided anti-jamming maritime communications: a deep reinforcement learning approach," in *2021 13th International Conference on Wireless Communications and Signal Processing (WCSP)*. IEEE, 2021, pp. 1–6.

[26] J. Peng, Z. Zhang, Q. Wu, and B. Zhang, "Anti-jamming communications in UAV swarms: A reinforcement learning approach," *IEEE Access*, vol. 7, pp. 180 532–180 543, 2019.

[27] Z. Ji, J. Tu, X. Guan, W. Yang, W. Yang, and Q. Wu, "Energy efficient design in IRS-assisted UAV data collection system under malicious jamming," *arXiv preprint arXiv:2208.14751*, 2022.

[28] X. Wang, M. C. Gursoy, T. Erpek, and Y. E. Sagduyu, "Learning-based UAV path planning for data collection with integrated collision avoidance," *IEEE Internet of Things Journal*, vol. 9, no. 17, pp. 16 663–16 676, 2022.

[29] J. Chen, D. Raye, W. Khawaja, P. Sinha, and I. Guvenc, "Impact of 3D UWB antenna radiation pattern on air-to-ground drone connectivity," in *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*, Aug 2018, pp. 1–5.

[30] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[31] Y. F. Chen, M. Liu, M. Everett, and J. P. How, "Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning," in *2017 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2017, pp. 285–292.

[32] M. Everett, Y. F. Chen, and J. P. How, "Collision avoidance in pedestrian-rich environments with deep reinforcement learning," *IEEE Access*, vol. 9, pp. 10 357–10 377, 2021.

[33] J. Van Den Berg, S. J. Guy, M. Lin, and D. Manocha, "Reciprocal n-body collision avoidance," in *Robotics research*. Springer, 2011, pp. 3–19.

[34] Z. Wang, T. Schaul, M. Hessel, H. Hasselt, M. Lanctot, and N. Freitas, "Dueling network architectures for deep reinforcement learning," in *International conference on machine learning*. PMLR, 2016, pp. 1995–2003.

Xueyuan Wang received the B.S. degree in electrical and electronics engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 2013, the M.S. degree in electrical engineering from Syracuse University, Syracuse, NY, in 2016, and the Ph.D. degree in electrical and computer engineering from Syracuse University in 2021. She is currently an instructor in the School of Computer Science and Artificial Intelligence at Changzhou University. Her primary research interests include unmanned aerial vehicles-enabled networks, 5G and beyond communications, Internet of Things networks, multi-agent joint control, and reinforcement learning.

M. Cenk Gursoy (Senior Member, IEEE) received the B.S. degree with high distinction in electrical and electronics engineering from Bogazici University, Istanbul, Turkey, in 1999 and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, in 2004. He was a recipient of the Gordon Wu Graduate Fellowship from Princeton University between 1999 and 2003. He is currently a Professor in the Department of Electrical Engineering and Computer Science at Syracuse University. His research interests are in the general areas of wireless communications, information theory, communication networks, signal processing, and machine learning. He is a member of the editorial boards of IEEE Transactions on Wireless Communications and IEEE Transactions on Communications, and he is an Area Editor for IEEE Transactions on Vehicular Technology. He also served as an editor for IEEE Transactions on Green Communications and Networking between 2016 and 2021, IEEE Transactions on Wireless Communications between 2010 and 2015, IEEE Communications Letters between 2012 and 2014, IEEE Journal on Selected Areas in Communications - Series on Green Communications and Networking (JSAC-SGCN) between 2015 and 2016, Physical Communication (Elsevier) between 2010 and 2017, and IEEE Transactions on Communications between 2013 and 2018. He has been the co-chair of the 2017 International Conference on Computing, Networking and Communications (ICNC) - Communication QoS and System Modeling Symposium, the co-chair of 2019 IEEE Global Communications Conference (Globecom) - Wireless Communications Symposium, the co-chair of 2019 IEEE Vehicular Technology Conference Fall - Green Communications and Networks Track, and the co-chair of 2021 IEEE Global Communications Conference (Globecom), Signal Processing for Communications Symposium. He received an NSF CAREER Award in 2006. More recently, he received the EURASIP Journal of Wireless Communications and Networking Best Paper Award, 2020 IEEE Region 1 Technological Innovation (Academic) Award, 2019 The 38th AIAA/IEEE Digital Avionics Systems Conference Best of Session (UTM-4) Award, 2017 IEEE PIMRC Best Paper Award, 2017 IEEE Green Communications & Computing Technical Committee Best Journal Paper Award, UNL College Distinguished Teaching Award, and the Maude Hammond Fling Faculty Research Fellowship. He is a Senior Member of IEEE, and is the Aerospace/Communications/Signal Processing Chapter Co-Chair of IEEE Syracuse Section.