参赛队员姓名：周城锴

中学：成都市第七中学（高新校区）

省份：四川省

国家/地区：中 国

指导教师姓名：Kadhim Mustafa Raad Kadhim

指导教师单位：电子科技大学

论文题目：对抗环境下多无人机的数据采集路径规划算法

论文题目：对抗环境下多无人机的数据采集路径规划算法 (Multi-UAVs Path Planning for Data Harvesting in Adversarial Scenarios)

作者：周城锴(Zhou, Chengkai)

论文摘要：两年前，我和父母一起在海南旅行时，第一次亲自尝试了操作无人机飞行。我发现虽然无人机能够全景拍照、自动降落，带给我更多平时无法触及的视野，但是，让无人机高空徘徊、绕开障碍却非常困难，这些困难极大影响了我的操作体验。此后，我便对如何让无人机智能地避开障碍、完成各种任务产生了浓厚的兴趣。在经常关注无人机应用的相关新闻后，我意识到我操作无人机时遇到的问题，在军事、救灾等场景下同样具有重要意义。

去年冬天，我有幸入选了中国科协和教育部共同组织实施的中学生“英才计划”，进入高校参加科研训练，并和指导老师讨论了我关于无人机的思考。在学习过程中，我了解到无人机在学术界被称为“Unmanned Aerial Vehicle”，大量的工作聚焦于如何让无人机群从城市、森林等区域采集数据信息，通过无人机群的路径规划，使其躲避障碍、防止碰撞，在采集信息后返回着陆区域，这些方法也解答了我长久的疑问。同时，我产生了新的疑问，在战场等对抗环境下，这些方法是否仍然有效呢。我尝试验证自己的想法，通过模拟对抗情况，让敌人可以设置陷阱来捕获我方无人机，结果表明现有方法无法让无人机绕开陷阱进行信息收集。

在和老师进行深入讨论后，我设计了对抗环境下的多无人机路径规划方法。整个方法基于强化学习框架，我定义了其中的状态空间和奖励函数，并且设计了安全通信方法，应对敌人设置的陷阱诱捕和信息窃听两种威胁方式。整个框架采用 DDQN 进行训练，并且考虑到了人在回路的无人机控制方式。根据测试结果分析，我设计的方法能够在对抗环境中安全、全面的采集数据信息，并返回着陆区，相较于对比方法，数据采集覆盖率和安全返航率提升超过 60%。

关键词：无人机协同，无人机路径规划，深度强化学习，对抗任务环境

Multi-UAVs Path Planning for Data Harvesting in Adversarial Scenarios

Chengkai Zhou
Chengdu No.7 High School, Gaoxin
Chengdu, China

***Abstract*—In recent years, unmanned aerial vehicles (UAVs) have been widely utilized for data harvesting in scenarios where reliable communication infrastructure is absent. Among existing solutions, deep Q-network (DQN) models have been extensively applied for multi-UAV path planning owing to its superiority in representing the surrounding environment and action space for UAVs. However, the specific designs for UAV path planning in non-cooperative and adversarial environment is still seldom mentioned. This fact has reduced the performance for multi-UAV data collection in harsh environment and tasks, like the rescue in disaster area and the investigation in conflicting regions. Therefore, this paper introduces a novel DQN framework named MPDA addressing the critical issues in adversarial environments. First, I assume and formulate the existence of trapping and eavesdropping attacks in monitoring regions. Second, a novel Deep Q-learning model is designed which encodes the surrounding threats into the state space and Q-networks, by combining them with the battery capacity, the safe flying distance among UAVs and the volume of collected data. Besides, a novel reward function is also designed to further refine the action series of UAVs for better data coverage among devices and safety landing. Third, I design a Shamir threshold method based mechanism for secure information sharing between IoT devices and UAVs. Finally, extensive simulation results have demonstrated the advanced performance for multi-UAV data collection in adversarial environment, which outperform baseline method by 60% on multiple metrics.**

***Index Terms*—Multi UAVs, Path planning, Deep Q-learning, Adversarial environment**

I. BACKGROUND

Due to their efficiency, flexibility, and low resource consumption [1], grouped Unmanned Aerial Vehicles (UAVs) form a promising approach for information harvesting within regions where infrastructure and stable network connection are unavailable [2] [3] [4]. UAVs are deployed to carry out diverse tasks including real-time identification and localization, sustainable urban-scale sensing of targets in surrounding area, and application-aware content and network optimization. Moreover, the extensive operation of grouped UAVs also demonstrates that appropriate collaboration strategies could improve overall mission execution; flying a swarm/group of UAVs supports the cooperation ability between UAVs by sharing important information, which assists in finishing the given tasks easier and faster, even in a much harsher and more complicated environment.

Recently, the main stream of research in grouped UAV task planning concentrates on the adoption of the reinforcement learning technique in various circumstances. Among these attempts, the Deep Q-learning (DQN) based methods have achieved advanced performance. The deep neural networks in DQN models provides superior capabilities in representing high-dimensional state spaces for capturing elements in monitoring area, while the Q-learning mechanism could derive sophisticated strategies and make decisions from previous actions. Both features have enhanced the overall task execution and collaboration among UAVs. For example, Zeng *et al.* studied a DQN-based solution realizing simultaneous UAV navigation and radio mapping in 3D space. The proposed model implemented a dueling deep DQN algorithm based on Markov decision process (MDP) [5]. Walker *et al.* proposed a deep reinforcement learning (DRL) framework for autonomous UAV indoor navigation based on partially observable Markov decision process (POMDP) [6]. Bayerlein *et al.* introduced a benchmark DQN approach based on the decentralized partially observable Markov decision process (Dec-POMDP) to work out multi-UAVs path planning strategy in random dense urban environments [7].

Despite their remarkable achievements, most of current DQN models for UAV planning are still constrained in handling the mission process in non-cooperative and competing scenarios. This kind of situation is somehow common and pivotal as UAVs are more frequently applied for challenging tasks like investigation in battlefields or disaster area. In these scenarios, some adversaries and rivals may exist in the monitoring area, like the enemies from non-cooperative parties and the uncertain dangers in the disaster area. The adversaries try to interrupt the normal mission process of UAVs. Specifically, they may raise two types of threats. First, they can conduct eavesdropping attacks by deploying malicious UAVs cruising in the area. These UAVs can steal sensing data by overhearing the communication channels between UAVs and the sensing devices. Second, normal UAVs are also in danger of being captured or destroyed when adversaries set some traps within the task area, which is denoted as trapping attacks. Both attacking models cloud bring severe threats to the execution of data collection tasks and the retaining of UAVs.

Therefore, I study the problem of multi-UAV cooperation in an antagonistic mission environment with both eavesdropping and trapping attacks. In this case, I need to consider the power capacity of UAVs, the complicated environment, and the potential threats from enemies. Initially, the data collector would deploy some IoT devices in the non-cooperative area to monitor and collect information about the surroundings. Then a group of UAVs will be released to collect data from IoT devices in the antagonistic region and try to land back to the starting area. The path planning for UAVs should be properly derived such that the total data volumes and the coverage of IoT devices and both guaranteed, while the UAVs are protected against both attacks.

According to the problem formulation, I propose a novel DQN model for multi-UAVs path planning in an adversarial scenario to address these issues. I formulate the problem by transforming a complicated environment into a grid world and various kinds of regions into different sets of cell grids marked by different labels. Then, I propose a DQN model that can derive the action plans for UAVs in distributed manners, aiming to simplify the problem and lower the computation load. I design a novel state space consisting of the physical space, the adversarial space representing the adversarial attack, and the return space representing the corresponding values of collecting data from different devices. I follow the typical mapping strategies to generate the Q-network. Further, the reward function is improved via inserting penalties when UAVs are captured by traps or continuously collect data from few IoT devices. In this way, the model could facilitate UAVs to bypass traps and collect data from distinct devices. Finally, I also design a multi-round Shamir secret-sharing mechanism to secure the message against eavesdropping UAVs. Simulation results suggest that our model can excellently handle the difficulties within the mission of collecting data under the threat of enemy attack and significantly boost the return success rate (RSR). The main contribution of this work includes:

- A novel system model for multi-UAV path planning in non-cooperative environment is constructed. The formulated model includes the eavesdropping and trapping attacks, which are essential for the design of secure and appropriate solutions.
- A DDQN-based model for multi-UAVs path planning in an adversarial scenario is designed to derive the optimal movement plans, where novel state spaces and reward functions are studied to enhance the resistance against trapping attacks as well as improve the data coverage among IoT devices.
- A distributed multi-round data sharing strategy is proposed for secure data collection against eavesdropping attacks. The strategy is based on Shamir threshold method and guarantees the original data would hardly be obtained.
- Extensive numerical results are introduced and my method performs better in adversarial environments than baseline models.

The organization of this paper is as follow. Section II briefly reviews existing solutions on Multi-UAV path planning. Section III introduces the validation and observation on applying the baseline model to adversarial environment. Section IV introduces the proposed reinforcement learning model. Section V gives the implementation of the model, and section VI introduces the evaluation results. Finally, section VII concludes the whole paper.

II. Related Work

A. Reinforcement Learning

Reinforcement learning (RL) is widely used in tasks where interaction with the real environment exists, especially the UAV path planning task. The main idea that is the UAV's actions interact with the environment to obtain Reward and state changes, so that the local model learns the environment properties and selects the optimal action strategy for the UAV. Pham *et al.* proposed a Q-learning algorithm [8] for autonomous UAV navigation. The work let Q-learning control the proportional Cintegral Cderivative (PID) controller parameters willing to improve the navigation performance of UAV in a 2D indoor space. Considering convergence, in their subsequent work, the Q-learning with function approximation based on fixed sparse representation (FSR) and a model termed the honey badger algorithm (HBA) were integrated.

Since then, the optimization of various aspects of RL methods for UAV missions has become a hot research topic. In [9], Hu *et al.* investigated optimization, premature convergence, and the traditional HBA feasible and efficient paths issues. This work proposed SaCHBA-PDN to sort out the above issues by the Bernoulli shift map, piece-wise optimal decreasing neighborhood, and horizontal crossing with strategy adaptation. Xu *et al.* proposed RL-based model to realize a more effective path planning method [1]. They integrated the comprehensive learning [10] (CLPSO) algorithm with the dynamic multi-swarm PSO (DMSPSO) algorithm denoted as CL-DMSPSO. Poudel *et al.* proposed an actor-critic multi-agent RL model for multi-UAVs operations [11]. They also proposed a priority-aware task assignment and path planning (AMTP) algorithm, which surpassed recent methods in many aspects. Liu *et al.* investigated the computation offloading problem (COP) [12]. The suggested solution was based on multi-objective MDP and their Q-network structure combined Double Deep Q Network (DDQN) with Dueling Deep Q Network (Dueling DQN) to improve the optimization efficiency.

While the aforementioned works are excellent contributions to the optimization of the methodology, they are all based on idealized discourse assumptions about the real environment. These methods lack consideration of the adversarial environment.

B. Attacker Model

It is important for UAVs to have the ability to combat risky environments in the course of their missions. In the military, for example, UAVs attempting to collect data from IoT devices may be able to capture or eavesdrop on the private data collected.

Existing work, researchers have proposed some real risks and solutions. For the information security (IS), [13] assumed UAVs could be protected from eavesdropping by sending interference signals to ground eavesdroppers. Thus, a multi-UAVs RL method based on multi-agent depth deterministic strategy gradient (MADDPG) have been proposed where the safety capacity maximized by jointly optimizing the UAV jammer, trajectory, and launcher's respective power. A proof-of-stake (PoS) blockchain-secured UAV-IoT data collection method is suggested by [14]. The model is based on generating blocks from the collected data and performing block audits with the PoS consensus mechanism. Meanwhile, improving robustness is also an essential way for models to adapt to diverse environments. Bai *et al.* proposed a resilience guarantee framework for multi-UAVs collaborative quality of service (QoS) management, which can provide resilient and guaranteed communication QoS [15]. Wu *et al.* investigated a UAV enabled wireless communication, where a number of ground nodes (GNs) are scheduled to communicate with UAVs in the existence of jammers with imperfect location information [16]. In [17], Wang *et al.* suggested a practical algorithm to prevent the secondary eavesdropper (SE) wiretapping information as much as possible and proves the security of the UAV itself is a great significance. The Dec-POMDP-based DRL approach to make path planning for UAVs have been proposed in [7]. It succeed in making movement decisions trading off data volume goals, flight-time efficiency, and navigation constraints. Finally, Wang *et al.* suggested a two-stage RL train policy to plan a collision-free trajectory by leveraging local noisy observations [18].

However, the existing methods are just considered part of the factors and still cannot obtain the best performance in the real task scenario. Our approach considers a combination of multiple aspects in the adversarial environment and gives advanced solutions.

III. Problem Inquiry

Building upon what I have learned about the current developments and progress in drone path planning, I have identified two concerns regarding the safety of drones in battlefield environments: being captured by traps and incomplete data collection. To validate whether the mentioned concerns have a real impact, my mentor told me to conduct simulation experiments. First, I selected a well-structured and high-performance model as a baseline for the experiments [19]. Then, I made modifications to the model's environment to meet the conditions I wanted to verify. Next, I recorded and analyzed the simulation experiment results to draw relevant conclusions. Finally, I listed all the impacting factors that affect the performance and safety of drone missions in a battlefield environment.

The baseline scenario involves multiple drones collectively collecting data from IoT devices in the environment, with the objective of reaching a landing zone before running out their battery power. Regarding obstacles in the environment, the baseline scenario specifies tall buildings that drones cannot fly over and low buildings that they can pass. However, it does not include provisions for traps capable of capturing drones. Therefore, I introduced traps into the paths where drones are allowed to fly (excluding tall building areas and landing zones). Drones entering these traps would be captured. The results of the simulation experiments have been visualized in Fig. 1, Fig. 1(a)-1(c) and 1(d)-1(f) represent the simulated paths on the two maps, respectively. The blue areas represent the start/landing areas. The yellow areas represent tall buildings that drones find difficult to fly over, while the pink square areas depict low buildings that drones can easily fly over. The purple squares represent trap areas. The circles in the figures represent IoT devices, from which drones need to collect data. The drones' movement paths are represented by continuous arrows.

The practical results validate my concerns, as the drones experienced a high loss rate in trap-infested environments. Taking Fig. 1(b) as an example, both drones intended to pass through the sensor-dense right channel to collect data, but they both fell into traps along this path. Additionally, as depicted in Fig. 1(c) and 1(f), even with the presence of three drones, there are still instances where data from at least one IoT device has not been collected. Clearly, the results of the simulation experiments validate my hypotheses regarding the safety issues of drones in battlefield environments. The initial intention of drone path planning is to collect a larger and more comprehensive volume of sensing data, which is why paths tend to concentrate in device-dense areas. However, this has inadvertently led to a greater risk from traps. On the other hand, excessive focus on avoiding traps can lead to a decrease in the drone's data collection capabilities.

In reality, there are far more factors that influence a drone's ability to execute tasks in real-world environments than just the two factors I mentioned above. At this point, I believe that analyzing and clarifying the issues that need to be considered in drone path planning, as well as the importance of each factor, is essential.

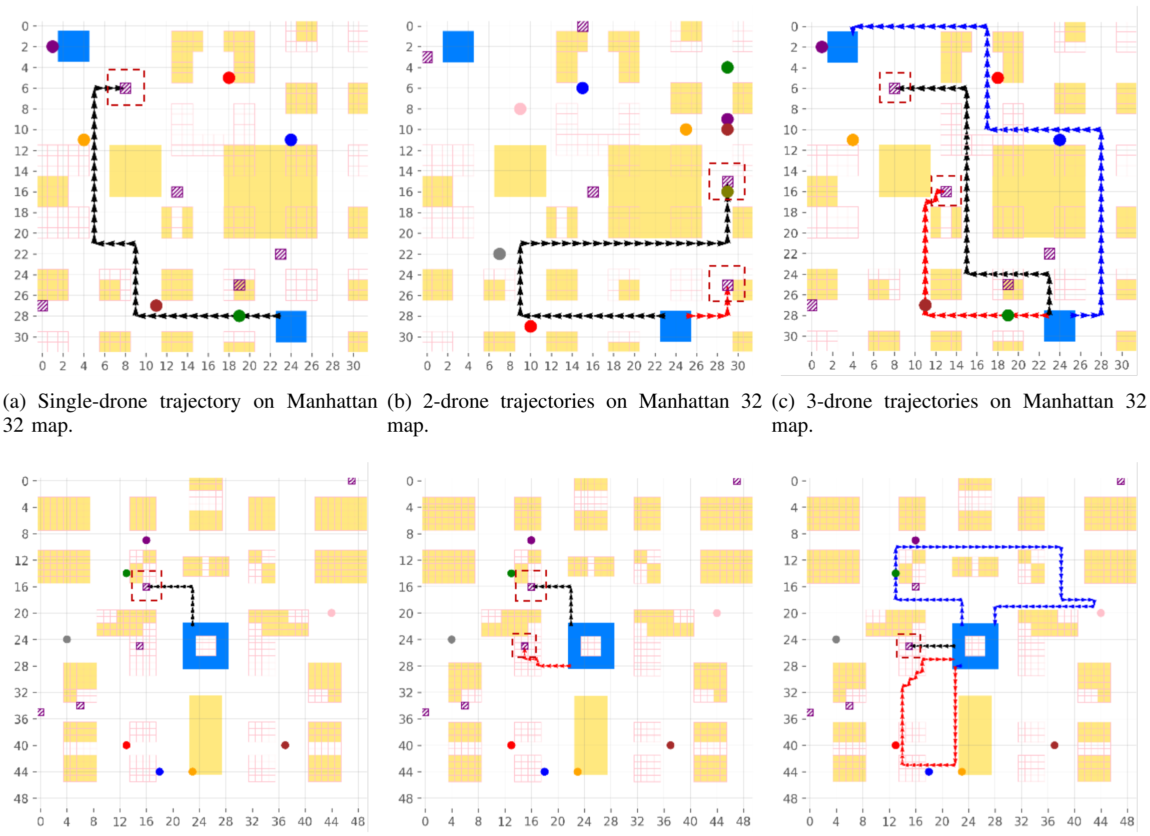
To comprehensively consider the factors, I'm analyzing the interactions between the objects present in drone missions from the following three perspectives [20] [21] [22]:

- The objectives that drones themselves aim to achieve.
- The relationships among multiple drones.
- The interaction between drones and their environment.

Regarding the drones themselves, their inherent characteristics include their: 1) Current position. 2) Battery capacity. 3) Self-condition (e.g., landing, damaged, or in-flight). For multiple drones to work in coordination, it's essential to consider the amount of data they collectively need to collect while also avoiding collisions between them.

The interaction between drones and their environment necessitates considering factors such as: 1) The presence of buildings in the environment (to avoid collisions with tall, impassable structures). 2) The presence of traps in the environment. 3) The distribution of IoT devices in the environment. 4) The data that needs to be collected within the environment. 5) Security considerations for the transmission of data between drones and IoT devices are also a crucial aspect to account for in drone operations.

Observing the factors listed above, it becomes apparent that there are conflicting factors, such as drone battery life and the volumes of data to be collected from the environment, all within the constraints of ensuring a safe return. Balancing



(a) Single-drone trajectory on Manhattan 32 map. (b) 2-drone trajectories on Manhattan 32 map. (c) 3-drone trajectories on Manhattan 32 map.

(d) Single-drone trajectory on Urban 50 map. (e) 2-drone trajectories on Urban 50 map. (f) 3-drone trajectories on Urban 50 map.

Fig. 1. Illustration of UAV being captured in battlefield environment under baseline model. The two rows of images shows the flight trajectories of drones on the Manhattan 32 and Urban 50 maps, where purple grids refer to traps. In all cases, UAVs are likely to fuse at dense regions and easily being captured (marked by red dotted frames).

these conflicting factors is a challenging aspect of effective drone path planning.

IV. Problem Resolving

After discussing my analysis with my mentor, he suggested that I consider using reinforcement learning (RL) as a method to address these issues [23] [24]. RL is a powerful approach because it can interact with the environment, define rewards based on those interactions, and optimize decisions based on these rewards [25].

Previously, the work I referenced when conducting the simulation happened to use the RL method [19]. The work used multiple maps representing different information from various perspectives (including physical environment map, flying times map, operational status map, device map) and then stacked these maps together as input into a convolutional neural network (CNN) [26]. Additionally, this work defined a reward function that encompassed multiple constraints and aspects: the total data collected by all drones during the current time period, penalties for drone collisions, penalties for drones failing to return successfully, and a constant motion penalty to inspire drones to minimize their movement time and prioritize efficient paths. Based on this reward function, the CNN is trained in multi rounds to output appropriate motion paths for the UAV based on the input superimposed map.

I agreed that while the work had a strong overall approach, it lacked consideration for environmental diversity. **Building upon this, I made improvements by considering physical traps in the environment, the risk of eavesdropping, the coverage of data collection equipment by drones.** Generally, I have developed an improved model called **M**ulti-UAVs **P**ath planning model for **D**ata harvesting in **A**dversarial scenarios (**MPDA** for short). The development in my model includes:

- An extended state space for Q-networks, which captures more information in environment.
- An improved reward function, which balances multiple conflicting factors like the limited battery capacity and data collection from remote IoT devices.
- A secure communication method to defend the eavesdropping attacks.

Next, I will provide a detailed overview of the overall process of the method I proposed. Fig. 2 illustrates the overview of MPDA model, where each UAV acts as an agent on a certain map. The process can be described as an agent taking a move according to the state through the action function, which would be evaluated and rewarded. Meanwhile, the taken movement

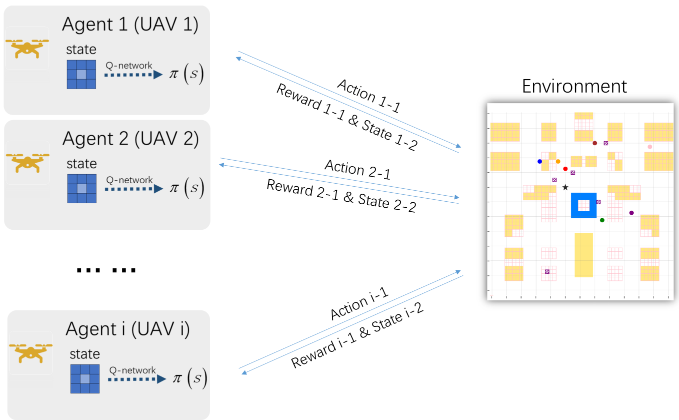
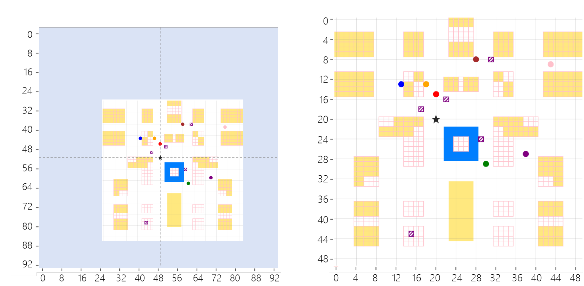


Fig. 2. Illustration of the MPDA model



(a) (b)

Fig. 3. Illustration of centered and non-centered input maps. The position of UAV is represented by the star and also the intersection of the dashed lines.

will also make the state change correspondingly, which can influence the next move. To understand the mechanisms of the proposed model, I need to introduce the following parts: system setup, map handling, return rate, state space, Q-network structure, reward function design, training process, and secure communication mechanism.

- **System Setup.** To build our assumption, I describe formally the entities that appear in the environment and their characteristics.

  a) Physical Environment: The physical environment demands a gird world $\mathcal{M}$ with a size of $M \times M \in \mathbb{N}^2$ where this virtual grid world contains some main sets to validate the UAV model when it face difficulties or issues. First, $\mathcal{L}$ is defined as the allowed starting/landing area for UAVs, which consists of L positions given by the set:

$$\mathcal{L} = \left\{ \left[x_i^l, y_i^l\right]^{\mathrm{T}}, i = 1, ..., L, : \left[x_i^l, y_i^l\right]^{\mathrm{T}} \in \mathcal{M} \right\}, \quad (1)$$

  where x and y represent the abscissa and the ordinate in the grid world $\mathcal{M}$, and T means transforming the matrix. Meanwhile, the $\mathcal{Z}$ set represents the tall buildings and the no-fly zone (NFZ) which UAVs cannot pass across and it defined as follows:

$$\mathcal{Z} = \left\{ [x_i^z, y_i^z]^{\mathrm{T}}, i = 1, ..., Z, : [x_i^z, y_i^z]^{\mathrm{T}} \in \mathcal{M} \right\}. \quad (2)$$

  The obstacles areas that block wireless links/signals, these areas stored into $\mathcal{B}$ set. Equation 3 declare the idea of obstacles blocking connections and I have considered the short buildings that UAVs are allowed to fly over even if the connection are blocked.

$$\mathcal{B} = \left\{ \left[x_i^b, y_i^b\right]^{\mathrm{T}}, i = 1, ..., B, : \left[x_i^b, y_i^b\right]^{\mathrm{T}} \in \mathcal{M} \right\}. \quad (3)$$

  It is assumed that $\mathcal{T}$ is a set of traps positions inside the grid world given by Equation 4, it refers to the UAVs positions where they might be captured by the enemy with a certain probability.

$$\mathcal{T} = \left\{ \left[x_i^{tr}, y_i^{tr}\right]^{\mathrm{T}}, i = 1, ...Tr, : \left[x_i^{tr}, y_i^{tr}\right]^{\mathrm{T}} \in \mathcal{M} \right\}. \quad (4)$$

  b) UAVs: The total number of deployed UAVs moving within the limit of the grid world $\mathcal{M}$ is I, and all the deployed UAVs form a set $\mathcal{I}$. The state of the i-th UAV is described via following parameters:

  – Position $p_i(t) = [x_i(t), y_i(t), z_i(t)]^{\mathrm{T}} \in \mathbb{R}^3$ where the parameter $z_i(t)$ refers to the altitude which start from normally start from 0 to h altitude depends on UAV ability.
  – The operational status $\phi_i(t) \in \{0, 1\}$, which are inactive and active, respectively.
  – Battery energy level $b_i(t) \in \mathbb{N}$

  The action space (movement) $\mathcal{A}$ of each UAV describe precisely as follows:

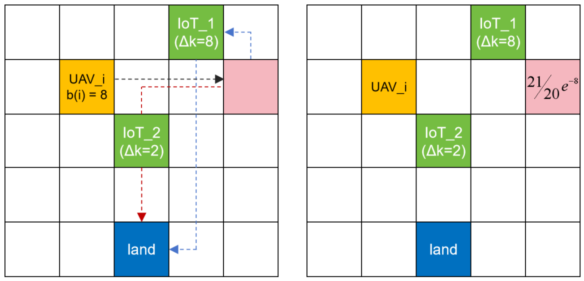
$$\mathcal{A} = \left\{ \begin{matrix} \underbrace{\begin{bmatrix}0\\0\\0\end{bmatrix}}_{\text{hover}}, & \underbrace{\begin{bmatrix}c\\0\\0\end{bmatrix}}_{\substack{\text{east}\\\text{move}}}, & \underbrace{\begin{bmatrix}0\\c\\0\end{bmatrix}}_{\substack{\text{north}\\\text{move}}}, & \underbrace{\begin{bmatrix}-c\\0\\0\end{bmatrix}}_{\substack{\text{west}\\\text{move}}}, & \underbrace{\begin{bmatrix}0\\-c\\0\end{bmatrix}}_{\substack{\text{south}\\\text{move}}}, \\ \underbrace{\begin{bmatrix}2c\\0\\0\end{bmatrix}}_{\substack{\text{east}\\\text{sprint}}}, & \underbrace{\begin{bmatrix}0\\2c\\0\end{bmatrix}}_{\substack{\text{north}\\\text{sprint}}}, & \underbrace{\begin{bmatrix}-2c\\0\\0\end{bmatrix}}_{\substack{\text{west}\\\text{sprint}}}, & \underbrace{\begin{bmatrix}0\\-2c\\0\end{bmatrix}}_{\substack{\text{south}\\\text{sprint}}}, & \underbrace{\begin{bmatrix}0\\0\\-h\end{bmatrix}}_{\text{land}} \end{matrix} \right\}, \quad (5)$$

  where the matrix included hovering, east-moving, north-moving, west-moving, south-moving, east-sprinting, north-sprinting, west-sprinting, south-sprinting and landing sets, respectively. In $\mathcal{A}$, the unit distance for each step is represented as c. Thus, the purpose to define the sprinting movement is to allow UAVs to rush across traps.

- **Map Handling.** Following the setup inspired by the [27], the maps are processed with each drone at the center. The maps are divided into a local map, which focuses on the immediate surroundings, and a global map, which encompasses the broader perspective [28].

  −**Turning 3D space into 2D space**. The process is shown in Equation 6 and 7,

$$\tilde{u}_k = \left[\begin{pmatrix} \frac{1}{c} & 0 & 0 \\ 0 & \frac{1}{c} & 0 \end{pmatrix} u_k \right] \quad (6)$$



(a) (b)

Fig. 4. Illustration of the reward rate estimation. The figure illustrates the process of computing the reward rate when a UAV moves to a specific grid (like the pink grid). The estimation considers the steps moving to the grid, the distance among the grid, each IoT device (green grids), and the landing zone (blue grid), as well as the remaining data on each IoT device.

$$\tilde{p}_i = \left[\begin{pmatrix} \frac{1}{c} & 0 & 0 \\ 0 & \frac{1}{c} & 0 \end{pmatrix} p_i\right], \tag{7}$$

where u_k and p_i refer to the 3D coordinates of the k-th UAV and the i-th IoT device respectively, while $\tilde{u}_k$ and $\tilde{p}_i$ represent the corresponding 2D coordinates.

$-$**Defining the general mapping function**. The centering and global-local mapping algorithms are based on map layer representations of the state space. To represent any state with a spatial aspect given by a position and a corresponding value as a map layer, I define a general mapping function in Equation 8.

$$f_{mapping} : \mathbb{N}^{Q\times 2} \times \mathbb{R}^{Q} \mapsto \mathbb{R}^{M\times M} \tag{8}$$

The map layer $A \in \mathbb{R}^{M\times M}$ is defined as:

$$A = f_{mapping}\left(\{\tilde{p}_q\}, \{v_q\}\right), \tag{9}$$

where $a_{\tilde{p}_q,0,\tilde{p}_q,1} = v_q,\ \forall q \in [0, ..., Q-1]$. The elements of A whose index is not in the grid coordinates are assigned as 0.

$-$**Central mapping**. The centering function is defined as:

$$f_{center} : \mathbb{R}^{\mathrm{M\times M\times n}} \times \mathbb{N}^2 \times \mathbb{R}^{n} \mapsto \mathbb{R}^{M_c\times M_c\times n} \tag{10}$$

where M_c can be expressed by M, given as $M_c = 2M-1$. A centered tensor $B \in \mathbb{R}^{M_c\times M_c\times n}$ can be valued with a tensor $A \in \mathbb{R}^{\mathrm{M\times M\times n}}$, as shown in Equation 11.

$$B = f_{center}\left(A, \tilde{p}, x_{pad}\right) \tag{11}$$

The elements of B with respect to the elements of A are defined as:

$$b_{i,j} = \begin{cases} \tilde{a}_{i,j}, & M \leq i + \tilde{p}_0 + 1 < 2\mathrm{M} \\ & \wedge M \leq j + \tilde{p}_1 + 1 < 2M \\ x_{pad}, & \text{otherwise}, \end{cases} \tag{12}$$

where $\tilde{a}_{i,j} = a_{i+\tilde{p}_0-M+1, j+\tilde{p}_1-M+1}$. $a_{i,j}$, $b_{i,j}$ and x_{pad} are vector valued of dimension $\mathbb{R}^n$. The map layers of A can be padded with the padding value x_{pad} via this Equation. After the centering process, the construction of the observation state for each UAV is illustrated in Figure 3a, compared to the construction without centering shown in Figure 3b.

As shown, the centered map which takes the position of the UAV as the center and fills the empty grids with x_{pad}. The centered map will completely cover the previous map even if the UAV is deployed at a corner of the non-centered map.

$-$**Global-local mapping**. The tensor $B \in \mathbb{R}^{M_c\times M_c\times n}$ processed with the centering function will be processed again with local and global mapping. The local map function is defined by

$$f_{local} : \mathbb{R}^{M_c\times M_c\times n} \times \mathbb{N} \mapsto \mathbb{R}^{l\times l\times \mathrm{n}} \tag{13}$$

while the global map function is defined by Equation 14.

$$f_{global} : \mathbb{R}^{M_c\times M_c\times n} \times \mathbb{N} \mapsto \mathbb{R}^{\left\lceil \frac{M_c}{g} \right\rceil \times \left\lceil \frac{M_c}{g} \right\rceil \times \mathrm{n}} \tag{14}$$

The local mapping and the global mapping are, respectively, based on

$$X = f_{local}\left(B, l\right) \tag{15}$$

$$Y = f_{global}\left(B, g\right) \tag{16}$$

The respective elements of X and Y concerning the elements of B are defined as:

$$x_{i,j} = b_{i+M-\left\lceil \frac{l}{2} \right\rceil, j+M-\left\lceil \frac{l}{2} \right\rceil} \tag{17}$$

$$y_{i,j} = \frac{1}{g^2} \sum_{u=0}^{g-1} \sum_{v=0}^{g-1} b_{gi+u, gj+v} \tag{18}$$

The process steps in Equation 17 and 18 can be regarded as an average pooling operation with pooling cell size g. Parameters l and g are introduced to determine the size of the local and global maps. During the whole process, the size of the map will increase by increasing l or decreasing g.

- **Return Rate.** Return rate takes into account factors that are of a conflicting nature during UAV operations: battery capacity, motion consumption and data to be collected:

$$w(p(\hat{i})) = \sum_{i\in I} e^{-b(i)} \sum_{k\in K} \frac{\Delta k}{dis(p(i), p(\hat{i})) + dis_{\text{back}}\left(p(\hat{i}), p(\mathrm{IoT}(k))\right)}, \tag{19}$$

where $p(\hat{i})$ is the position of any grid on the map. Δk is the difference value between the respective data sizes of devices that have been collected at least once and the not collected devices. $dis(p(i), p(\hat{i}))$ refers to the distance to the target device, while $dis_{back}(p(\hat{i}), IoT(k))$ is the distance that the current drone passes through $p(\hat{i})$ and finally returns to $\mathcal{L}$ area. $e^{-b(i)}$ is used to dynamically adjust the UAV action tendency based on the remaining power. If the remaining electric quantity is relatively high, the power-consuming value of $e^{-b(i)}$ changes slowly, which will not impact $w(p(\hat{i}))$. Otherwise, it will have a

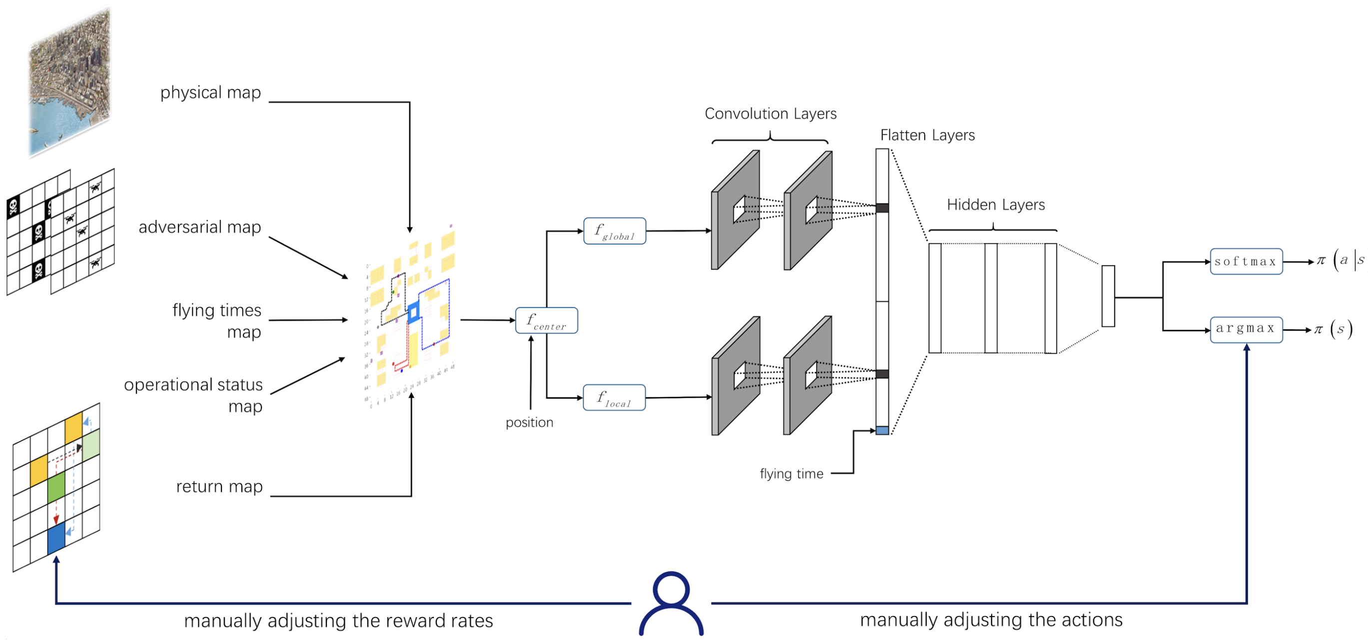


Fig. 5. Structure of the DQN in MPDA. Different information in the region will be encoded into the convolutional Q-network. Meanwhile, the operator can also add his expert experience and controlling orders into the reward rates or the action part, which will be a man-in-the-loop system.

much stronger effect on $w(p(\hat{i}))$. An illustration is shown in Fig. 4.

- **State Space** The state space in a typical DQN model indicates all combinations of situations within the environment. The grid world is the basis for the construction of state space, and each grid will be marked with one number representing one aspect of the state in this grid. I consider the scenario focus of MPDA to constitute the state space in terms of three dimensions:

$$\Omega = \Omega_{phy} \times \Omega_{\text{adversarial}} \times \Omega_{\text{return}}, \tag{20}$$

where Ω_{phy} represents the physical map that can describe the $\mathcal{L}$, $\mathcal{B}$, and $\mathcal{Z}$ areas. $\Omega_{adversarial}$ is the adversarial map, includes the entrapping, eavesdropping, i.e. $\mathcal{T}$ area and area of hostile drones. Ω_{return} represents the return rate map, which is based on the return rate function I designed.
In the proposed model MPDA, both local and global aspects are considered for each state map:

$$\begin{aligned} s_i(t) = \quad & (M_{l,i}(t), \mathcal{T}_{l,i}(t), \mathcal{R}_{l,i}(t), \Phi_{l,i}(t), \\ & M_{g,i}(t), \mathcal{T}_{g,i}(t), \mathcal{R}_{g,i}(t), \Phi_{g,i}(t), b(i)) \end{aligned} \tag{21}$$

The states can be derived in Equation 22 to 29:

$$M_{l,i}(t) = f_{local}(f_{center}(M, p_i(t), [0,1,1]^T), l), \tag{22}$$
$$M_{g,i}(t) = f_{global}(f_{center}(M, p_i(t), [0,1,1]^T), g), \tag{23}$$
$$\mathcal{T}_{l,i}(t) = f_{local}(f_{center}(\mathcal{T}, p_i(t), 0^T), l), \tag{24}$$
$$\mathcal{T}_{g,i}(t) = f_{global}(f_{center}(\mathcal{T}, p_i(t), 0^T), g), \tag{25}$$
$$\mathcal{R}_{l,i}(t) = f_{local}(f_{center}(\mathcal{R}, p_i(t), 0^T), l), \tag{26}$$
$$\mathcal{R}_{g,i}(t) = f_{global}(f_{center}(\mathcal{R}, p_i(t), 0^T), g), \tag{27}$$
$$\Phi_{l,i}(t) = f_{local}(f_{center}(\Phi, p_i(t), 0^T), l), \tag{28}$$
$$\Phi_{g,i}(t) = f_{global}(f_{center}(\Phi, p_i(t), 0^T), g), \tag{29}$$

where $M_{l,i}(t)$ refers to the position of the UAV in the local map. $M_{g,i}(t)$ is the position in the global map, and $p_i(t)$ is the operation status. For different types of states, I construct the trap map $\mathcal{T}$, the return map $\mathcal{R}$, data collection map Φ, and the remaining electric quantity b.

- **Q-network structure.** Based on the processing of the map, the Q-network selects optimal actions for the UAVs according to the previous states. I follow the idea proposed in [19] to construct the deep Q-network, as shown in Figure 5. The structure of the network can be described as:
  - Integrated all state maps into one hyper map. Then, the environment map is stacked with physical map, adversarial map, flying time map, operational status map and return map that handles together the current position by the centering function.
  - Pass the result from the first step to the global and local mapping functions.
  - The obtained results of the second step fed into the neutral network to gain the best action of the current

and next states by obtaining the maximum value and normalized exponential function, respectively.

Furthermore, the input reward map to the network can have its grid reward values altered manually based on geographic importance, allowing for real-time control of drone flight tendencies. Additionally, the final output actions can also be adjusted by tuning weights manually. Both facts allow for the man-in-the-loop control, where operators can add their expert experience into the operation of UAVs.

- **Reward Function.** The reward function is an abstraction that transfers the optimization goal to the achievable goal, which assists the model to judge the performance of an action. The process of this proposed formulation is given in Equation 30.

$$r_i(t) = \alpha \sum_{k \in K} D_k(t) + w(i) + \beta_i(t) + \gamma_i(t) + \varphi_i(t) + \varepsilon, \tag{30}$$

The meaning represented by each item of reward function is given separately:

  - $\alpha \sum\limits_{k \in K} D_k(t)$: α is a hyper-parameter, the values of α in each UAV should be the same at a particular time. $\Sigma_{k \in K} D_k(t)$ refers to the total amount of data collected from the k-th device within t time slots.
  - $w(i)$: The summation of all grid return rate, which is defined in Equation 31.

$$w(i) = \sum_{p(\hat{i}) \in M^2} w(p(\hat{i})) \tag{31}$$

  - $\beta_i(t)$: When the drone does not collide with an accident, a reward β is given.
  - $\varphi_i(t)$: A reward η is given when the UAV is not caught in a trap.
  - $\gamma_i(t)$: This item denotes the punishment γ when the UAV does not land in time.
  - ε: This item refers to the constant punishment for UAVs' movement, which aims to force UAVs to reduce the flying time and seek the most efficient path to finish missions.

The reward function takes into account multiple realistic environmental factors to help Q-networks make optimal decisions.

- **Training Process.** After learning several DQN models, I choose DDQN and use a multi-round training paradigm for the proposed path planning model, which maximizes the expectation of accumulated rewards and improves our training model. DDQN uses an experience replay buffer for stable learning, applying epsilon-greedy exploration, and implementing Double Q-Learning to reduce overestimation bias [29]. Loss is calculated by reward function $r_i(t)$ and used to update the online network's weights via gradient descent. The target network is periodically updated to enhance training stability. This process continues until convergence or a predefined number of episodes, resulting in a more robust reinforcement learning agent.
- **Secure Communication Mechanism.** To defend against the eavesdropping attacks, I apply the idea of Shamir secret sharing mechanism demonstrated by [30]. In this method, a certain butch of data D would be encrypted by the Shamir threshold equation, and an attacker needs to get at least k values of $f(x_i)$ to restore the original data. The critical process can shown in the following equation:

$$f(x) = D + a_1 x + a_2 x^2 + ... + a_{k-1} x^{k-1} \tag{32}$$

The essential idea of Adi Shamir's threshold scheme is that 2 points are sufficient to define a line, 3 points are sufficient to define a parabola, 4 points to define a cubic curve and so on. Thus, it takes k points to define a polynomial of degree $k-1$. This idea could alleviate the threat of information eavesdropping.

Accordingly, an IoT device will encrypt each batch of sensing data into n folds, and sends one copy per time slot when a UAV flies into its grid. In this case, it allow the existence of malicious UAVs eavesdropping the message, while only the normal UAVs are guaranteed to successfully decrypt the original data. Moreover, as the number k can be determined by data requestors before the deployment of IoT devices, adversaries can hardly estimate the necessary scales of malicious UAVs for overhearing.

V. Experiments

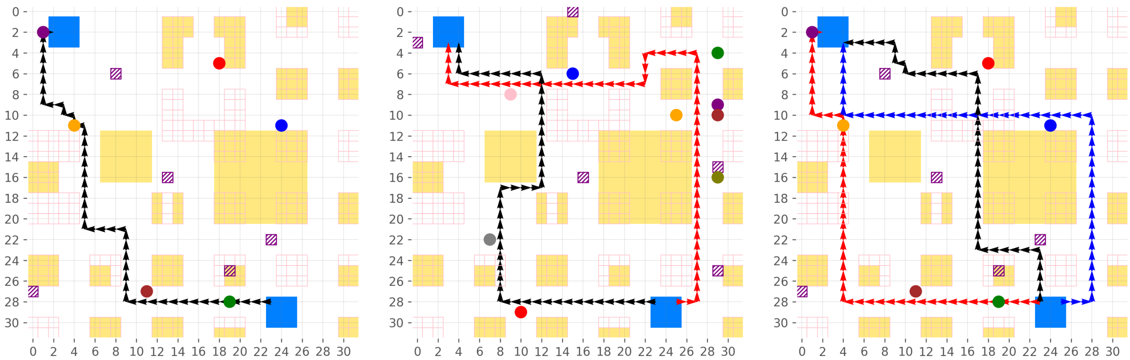
Experimental Platform

Following the README documentation of the open-source code for [19] on Github, I conducted simulation experiments in the following environments. I used an A4000 server GPU and a ROG Magician 7 Plus laptop. I established a connection between the laptop and the server using RSA key authentication through Visual Studio Code. The experiments were conducted using the Python programming language and the TensorFlow machine learning framework. The version of Python is 3.6.2, and the version of TensorFlow is 2.5.0. Additionally, I also utilized some dependent code packages. Here are the names and versions of the dependencies listed: numpy 1.19.5; keras 2.4.3; matplotlib 3.3.0; scikit-image 0.17.2; tqdm 4.45.0. The authors in [19] have provided modules to handle maps and simulate multiple drone flights in the code, and I have added simulations for traps and eavesdropping. At the same time, I changed the corresponding modules in the code related to traps and eavesdropping, such as the reward function.

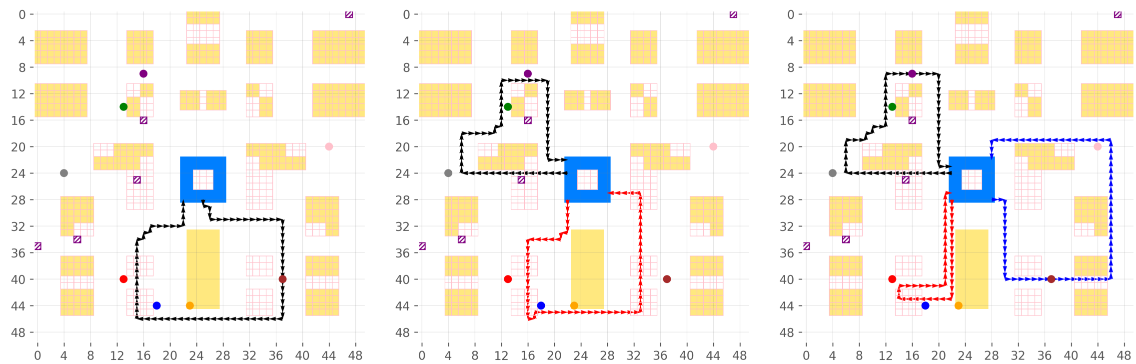
Evaluation Metrics

I conduct experiments to validate the proposed method and analyze it in the following ways:

- Drone Trajectory.
- Percentage of IoT device coverage.
- Impact of number of drones on results.
- Influence of the number of IoT devices on the results.
- Impact of map trap proportion.



(a) Single-drone trajectory in Manhattan 32 map. (b) 2-drone trajectories in Manhattan 32 map. (c) 3-drone trajectories in Manhattan 32 map.



(d) Single-drone trajectory in Urban 50 map. (e) 2-drone trajectories in Urban 50 map. (f) 3-drone trajectories in Urban 50 map.

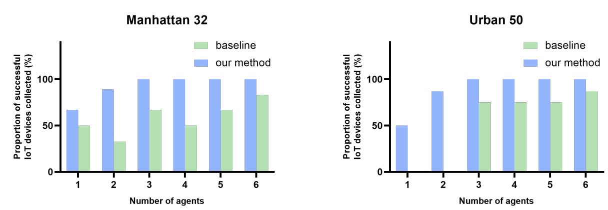
Fig. 6. Illustration of the drone path planning under my model MPDA in battlefield environments. In all images, UAVs can more or less bypass the traps and avoid being captured.

I use the ratio of the amount of data eventually collected by all drones collecting to the total amount of data to be collected as an evaluation metric, called "collection ratio".

In addition, as the method in [19] is referenced and improved upon, I use the paper's method as baseline to compare it with my proposed method.

VI. Analysis

Fig. 6 shows the UAV movement trajectories when 1, 2, and 3 UAVs are dispatched on the two maps, respectively. According to the above figures, method in this paper is effective in preventing UAVs from entering traps in all types of situations. For example, as shown in Fig. 1(b) and Fig. 6(b), when the UAVs plan to pass through the road with dense IoT devices on the right side to collect data, both of UAVs in Fig. 1(b) enter into the trap, while UAVs in Fig. 6(b) effectively circumvents the trap. The method in this paper is able to comprehensively consider the devices scattered all over the map. As shown in Fig. 7, the action trajectories of the three UAVs of our method cover all the areas in the map where the IoT devices are located. To further illustrate the coverage of the devices, I show the percentage of IoT devices with different number of UAVs whose data are effectively collected in Figure 7. The method described in this paper demonstrates a

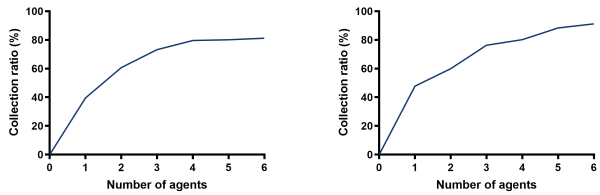


(a) Manhattan 32 (b) Urban 50

Fig. 7. IoT device collection situation.

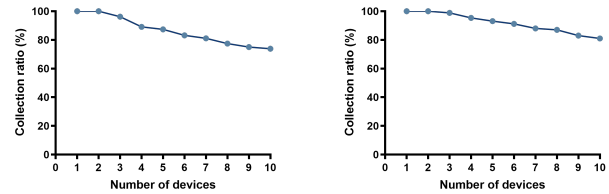
high capability for data collection in a battlefield environment. When number of drones reaches three or more, the effective data collection approaches nearly 100%.

I also compared the performance with baseline model. For the Manhattan 32 scenario, as Table I, my model has about 60% greater RSR, collection ratio and collection ratio and landed than baseline. And for the Urban50, as Table II, our model has about 65% greater RSR, 75% greater collection ratio and 70% greater collection ratio and landed than baseline. Since the baseline did not consider the attacker model and its algorithm did not take the number of the devices that data are collected from as a considerable variable. Whereas, my model



(a) Manhattan 32 (b) Urban 50

Fig. 8. Impact of the number of agents.



(a) Manhattan 32 (b) Urban 50

Fig. 9. Impact of the number of IoT devices.

takes the return and the avoidance of traps into account, having a much better performance in the training in the adversarial military mission circumstance.

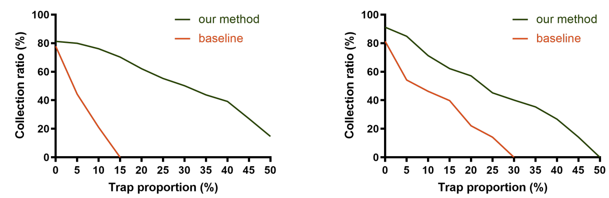
Furthermore, I conducted observations on the model's performance under varying conditions, including different numbers of drones, IoT devices, and traps. First, Figure 8 shows the effect of the number of UAVs dispatched on the final collection ratio for the same parameter settings. The method in this paper grows rapidly in the early stage and achieves a higher collection ratio with fewer UAVs. In the manhattan 32 map, as shown in Figure 8(a), the number of drones dispatched stabilizes after 4, and in the urban 50 map, as shown in Figure 8(b), the number of drones dispatched stabilizes after 5. Then, Figure 9 shows the effect of the number of IoT devices on the data collection ratio. When the number of devices is small, method in this paper is able to achieve 100% collection ratio. However, the collection rate decreases with the growth of the number of devices. As shown in Figure 9(a) and Figure 9(b), the metric decreases more slowly and smoothly under my model MPDA, and the model has good stability. Finally, Figure 10 shows the effect of trap percentage in the map, to show the performance of method in this paper, the figure compares method in this paper with baseline. A high trap proportion tends to lead to the destruction of drones or to avoid traps. Both sprinting and detouring require higher power cost. Therefore, as the trap rate increases, the collection ratio decreases. However, the collection ratio of our method consistently decreases more slowly than baseline, and the green curve representing the performance of our method is always located at the upper right of the red curve representing the baseline method in Figure 10(a) and Figure 10(b). The higher resistance of method in this paper to traps in the map ensures the safety of UAV operations.

TABLE I
FOR MANHATTTAN32

Method	baseline	MPDA
RSR	12.3%	74.0%
Collection Ratio	21.5%	81.5%
Collection Ratio and Landed	2.6%	64.3%

TABLE II
FOR URBAN50

Method	baseline	MPDA
RSR	15.7%	78.6%
Collection Ratio	18.6%	91.7%
Collection Ratio and Landed	2.9%	72.1%



(a) Manhattan 32 (b) Urban 50

Fig. 10. Impact of map trap proportion.

VII. CONCLUSION

In this paper, I study the problem of multi-UAV path planning in non-cooperative environment. The malicious attackers are assumed to conduct two types of attacks, i.e., eavesdropping and trapping attacks. The design goal of path planning is to achieve a balance on the volume of collected data, the maximum size of IoT devices with sufficient data collected, and the security and safety of UAVs. The improved model MPDA has success in executing the military battlefield mission, avoiding all the obstacles and traps and defending from all eavesdropping

Meanwhile, throughout the entire process of working out this project, I have experienced an extraordinary excitement of taking advantage of my hobby to have an even further study on it. Also, I have learned how to efficiently reading scientific papers and extracting useful information, as well as how to seeking out and succinctly expressing actual issues. What I believe is that my project must be one of the most important concerns for my country.

With this project, I learned a lot of knowledge on artificial intelligence algorithms, which would benefit me in my future study and career. I believe that I can have a much deeper study on this topic with more professional knowledge when I graduate and join in universities. I will also keep pursuing my dream in applying my research results to serve my country.

References

[1] L. Xu, X. Cao, W. Du, and Y. Li, "Cooperative path planning optimization for multiple uavs with communication constraints," *Knowledge-Based Systems*, vol. 260, p. 110164, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705122012606

[2] S. Aggarwal and N. Kumar, "Path planning techniques for unmanned aerial vehicles: A review, solutions, and challenges," *Computer Communications*, vol. 149, pp. 270–299, 2020.

[3] Y. Zhao, Z. Zheng, and Y. Liu, "Survey on computational-intelligence-based uav path planning," *Knowledge-Based Systems*, vol. 158, pp. 54–64, 2018.

[4] M. Jones, S. Djahel, and K. Welsh, "Path-planning for unmanned aerial vehicles with environment complexity considerations: A survey," *ACM Computing Surveys*, vol. 55, no. 11, pp. 1–39, 2023.

[5] Y. Zeng, X. Xu, S. Jin, and R. Zhang, "Simultaneous navigation and radio mapping for cellular-connected uav with deep reinforcement learning," *IEEE Transactions on Wireless Communications*, vol. 20, no. 7, pp. 4205–4220, 2021.

[6] O. Walker, F. Vanegas, F. Gonzalez, and S. Koenig, "A deep reinforcement learning framework for uav navigation in indoor environments," *2019 IEEE Aerospace Conference*, pp. 1–14, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:182107388

[7] H. Bayerlein, M. Theile, M. Caccamo, and D. Gesbert, "Multi-uav path planning for wireless data harvesting with deep reinforcement learning," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 1171–1187, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:225062058

[8] H. X. Pham, H. M. La, D. Feil-Seifer, and L. Van Nguyen, "Reinforcement learning for autonomous uav navigation using function approximation," in *2018 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, 2018, pp. 1–6.

[9] G. Hu, J. Zhong, and G. Wei, "Sachba_pdn: Modified honey badger algorithm with multi-strategy for uav path planning," *Expert Systems with Applications*, vol. 223, p. 119941, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417423004426

[10] J. Liang, A. Qin, P. Suganthan, and S. Baskar, "Comprehensive learning particle swarm optimizer for global optimization of multimodal functions," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 3, pp. 281–295, 2006.

[11] S. Poudel and S. Moh, "Priority-aware task assignment and path planning for efficient and load-balanced multi-uav operation," *Vehicular Communications*, vol. 42, p. 100633, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214209623000633

[12] X. Liu, Z.-Y. Chai, Y.-L. Li, Y.-Y. Cheng, and Y. Zeng, "Multi-objective deep reinforcement learning for computation offloading in uav-assisted multi-access edge computing," *Information Sciences*, vol. 642, p. 119154, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025523007399

[13] Y. Zhang, Z. Mou, F. Gao, J. Jiang, R. Ding, and Z. Han, "Uav-enabled secure communications by multi-agent deep reinforcement learning," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 10, pp. 11 599–11 611, 2020.

[14] X. Tang, X. Lan, L. Li, Y. Zhang, and Z. Han, "Incentivizing proof-of-stake blockchain for secured data collection in uav-assisted iot: A multi-agent reinforcement learning approach," *IEEE Journal on Selected Areas in Communications*, vol. 40, pp. 3470–3484, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:250311187

[15] C. Bai, P. Yan, X. Yu, and J. Guo, "Learning-based resilience guarantee for multi-uav collaborative qos management," *Pattern Recognit.*, vol. 122, p. 108166, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:239647322

[16] 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, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:210695465

[17] Z. Wang, J. Guo, Z. Chen, L. Yu, Y. Wang, and H. Rao, "Robust secure uav relay-assisted cognitive communications with resource allocation and cooperative jamming," *J. Commun. Networks*, vol. 24, pp. 139–153, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:247101230

[18] D. Wang, T. Fan, T. Han, and J. Pan, "A two-stage reinforcement learning approach for multi-uav collision avoidance under imperfect sensing," *IEEE Robotics and Automation Letters*, vol. 5, pp. 3098–3105, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:212645615

[19] H. Bayerlein, M. Theile, M. Caccamo, and D. Gesbert, "Multi-uav path planning for wireless data harvesting with deep reinforcement learning," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 1171–1187, 2021.

[20] J. Wu, H. Wang, N. Li, P. Yao, Y. Huang, and H. Yang, "Path planning for solar-powered uav in urban environment," *Neurocomputing*, vol. 275, pp. 2055–2065, 2018.

[21] M. Petrlík, T. Báča, D. Heřt, M. Vrba, T. Krajník, and M. Saska, "A robust uav system for operations in a constrained environment," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2169–2176, 2020.

[22] F. Mohammed, A. Idries, N. Mohamed, J. Al-Jaroodi, and I. Jawhar, "Uavs for smart cities: Opportunities and challenges," in *2014 international conference on unmanned aircraft systems (ICUAS)*. IEEE, 2014, pp. 267–273.

[23] Y. Matsuo, Y. LeCun, M. Sahani, D. Precup, D. Silver, M. Sugiyama, E. Uchibe, and J. Morimoto, "Deep learning, reinforcement learning, and world models," *Neural Networks*, vol. 152, pp. 267–275, 2022.

[24] Q. Liu, L. Shi, L. Sun, J. Li, M. Ding, and F. Shu, "Path planning for uav-mounted mobile edge computing with deep reinforcement learning," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 5, pp. 5723–5728, 2020.

[25] S. Meyn, *Control systems and reinforcement learning*. Cambridge University Press, 2022.

[26] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: analysis, applications, and prospects," *IEEE transactions on neural networks and learning systems*, 2021.

[27] M. Theile, H. Bayerlein, R. Nai, D. Gesbert, and M. Caccamo, "Uav path planning using global and local map information with deep reinforcement learning," in *2021 20th International Conference on Advanced Robotics (ICAR)*. IEEE, 2021, pp. 539–546.

[28] H. Bayerlein, M. Theile, M. Caccamo, and D. Gesbert, "Uav path planning for wireless data harvesting: A deep reinforcement learning approach," in *GLOBECOM 2020-2020 IEEE Global Communications Conference*. IEEE, 2020, pp. 1–6.

[29] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 26–38, 2017.

[30] S. A. A. Hakeem and H.-C. Kim, "Centralized threshold key generation protocol based on shamir secret sharing and hmac authentication," *Sensors (Basel, Switzerland)*, vol. 22, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:245869632

致谢

感谢Mustafa老师在研究过程中对我进行悉心的学术指导。Mustafa老师从最初的选题到最终的实验结果分析、报告撰写，一直给我提供无私的帮助。他牺牲了自己无数的周末休息时间，来和我讨论、给我指导。并且，他经常在深夜与我探讨方法模型、实验设置等方面的问题，给予我极大的支持和力量。

感谢我的母校成都七中，以及我参与科研学习的电子科技大学、四川大学。感谢这些学校为我提供了珍贵的学习机会和绝佳的学习平台。在母校老师的帮助下，我能够平衡好学习和科研的时间分配，全心全意地完成研究。在高校，我有幸和很多优秀的老师、师兄师姐学习、讨论，这使我能够快速打开自己的视野，令我受益匪浅。同时，电子科技大学和四川大学提供的计算资源，也是我完成科研任务的重要保障。

感谢父亲母亲以及各位长辈的支持与鼓励。在我不自信、迷茫的时候，是我的父亲母亲告诉我要相信自己，激励我克服困难，不断进步。

感谢丘成桐中学科学奖，实现了一个高中生体验科研工作的梦想。在整个学习过程中，我学到了强化学习、深度学习、加密通信等很多理论知识和技术方法，也训练了自己的代码能力和科学技术论文书写能力。更重要的是，我体会到了计算机科学如何用数学模型和方法一步步的解决一个实际问题，拓宽了我的眼界。

其它情况说明：

选题来源于同指导老师的讨论以及自己的个人兴趣。我在随家人旅游的过程中，对无人机操作产生了浓厚的兴趣。随后，我有幸入选了中国科协和教育部共同组织实施的中学生“英才计划”，在

高校进行科研学习。指导老师在得知我对无人机的兴趣后，指导我了解了无人机领域的研究情况，并且在讨论后选择《对抗环境下多无人机的数据采集路径规划算法》(Multi-UAVs Path Planning for Data Harvesting in Adversarial Scenarios)作为题目。

多无人机路径规划是计算机技术同无人机控制结合的一个重要任务，通过确定每架无人机的飞行轨迹，让无人机高效和安全的收集数据信息，在灾难救险、战场侦察等方面有很显著的意义。经过前期学习，我发现其他学者的研究没有充分考虑环境中的恶劣情况，比如敌对方设置的陷阱、通信干扰等，但这对于无人机的实际应用又是有重要价值的。在和指导老师确认想法后，我确定了自己的研究题目。

在课题研究过程中，我承担了问题建模、模型设计、代码编程和实验测试分析等任务，并且完成了论文撰写。

Kadhim Mustafa Raad Kadhim老师是我的科研指导老师，指导我完成论文的选题、算法模型研究、代码编程和实验测试分析，并且指导我完成论文的撰写。指导为无偿指导。

Curriculum Vitae

Personal Details

Title	Surname / Family name	First / given names	Sex	Photograph
Dr.	Kadhim	Mustafa Raad Kadhim	Male	

Current Address	
四川省成都市	
Date of birth	**Nationality**
1991-05-21	IRAQ
Place of Birth	**Marital Status**
Baghdad -Iraq	Single
Contact telephone number / Skype Name	
+8613084428131	
Email	
Kadhim_mustafa@uestc.edu.cn	

Current and Previous Employment

Employer name and address	Main responsibilities
University of Electric Science and Technology of China	The main job responsibilities are research in multiple machine learning domains, specifically unsupervised learning models for UAVs and data clustering and classification. Also, mentor students during their research process as well as reviewing their scientific papers. Currently, working on important projects that would help society in many aspects.
Job title	
Scientific researcher	
Dates of Employment	
Jan. 2023- Present	
Reason for Leaving	

Employer name and address	Main responsibilities
New vision Co., Ltd.	Worked as a manager of the software department. Within the period of our working there, our team had finished a lot of projects for the government, companies, and banks. The tasks of data collection, analysis, and structuring, as well as dividing the work among the team, were one of Dr. Mustafa's tasks. As well, software development training lectures for the new employees were given to engage them in teamwork.
Job title	
Manager of software department	
Dates of Employment	
Jan. 2014- Oct. 2015	
Reason for Leaving	
To continue studying	

Education, Skills, Professional Development Introduction:

Currently, Dr. Mustafa is a full-time scientific researcher at the University of Electric Science and Technology of China; earlier, obtained the Ph.D. degree from the same university. Thus, the Master's degree was obtained from China and the Bachelor's degree from Iraq. However, during the bachelor's study, Dr. Mustafa took part in several university completions and gained multiple awards; as a result, the third-best student prize was gained at graduation.

Dr. Mustafa has very good knowledge about the machine learning domain. The recent publication was based on investigating clustering and cluster ensemble real-world issues and finding decent solutions, aiming to come out with a strong framework. The research papers that were published by Dr. Mustafa investigated multiple issues that most recent researchers have not considered, also introduced critical solutions that improve the performance of machine learning in multiple aspects. Finally, several evaluation standardizations were proposed as well.

As recent work experience, Dr. Mustafa worked as a freelancer and full-time manager, where multiple real-world projects for governments and companies were developed and managed. The two most important developed projects are the project to management systems of dead people's information and the DNA checking progress management system. These projects were developed for the Ministry of Health's Forensic Medicine Department, and they are still working.

Briefly, Dr. Mustafa's personality is very passionate about learning and ready to work under pressure to obtain more knowledge and give better achievements; also, working with a team always leads to better service and creative ideas.

Education and Qualifications

Year of Start - Award	*Institution and Department*
2009 - 2013	AL-Rafidain University College
Qualification and Grade	*Title of Award*
Bachelor degree	Software engineering

Year of Start - Award	*Institution and Department*
2015 - 2018	Southwest Jiaotong University
Qualification and Grade	*Title of Award*
Master’s Degree	Computer science and technology

Year of Start - Award	*Institution and Department*
2018 - 2022	University of Electronic Science and Technology of China
Qualification and Grade	*Title of Award*
PhD. Degree	Software engineering

Language skills

First Language	*Second Language*	*Third Language*
Arabic	English	Chinese
Level	*Level*	*Level*
Mother language	Good	Good, have HSK4 Certificate

Research Publications

#	Paper Name	Publisher	Type/Date	Author Order
1	A novel self-directed learning framework for cluster ensemble	Journal of King Saud University - Computer and Information Sciences	Journal / 2022 Impact factor: 8.839 Rank: Q1/JCR1	1
2	A Novel Side-Information for Unsupervised Cluster Ensemble	2021 18th International Computer Conference on Wavelet Active Media Technology and Information Processing	EI Conference / 2021	1
3	Rapid Clustering with Semi-Supervised Ensemble Density Centers	2019 16th International Computer Conference on Wavelet Active Media Technology and Information Processing	EI Conference / 2019	1
4	Semi-supervised cluster ensemble based on density peaks	Data Science and Knowledge Engineering for Sensing Decision Support	EI Conference / 2018	1
5	A Novel Cluster Ensemble based on a Single Clustering Algorithm	16th Conference on Computer Science and Intelligence Systems	EI Conference / 2021	3
7	Comparison of Time Interval Statistic and Pulse Shape Discrimination in Fast Neutron Detection Method with Liquid Scintillation Detector Loaded Gd	Applied Physics Frontier	Journal / 2017 Impact factor: 3.56 Rank: Q2	7

Technical skills

1	Analysing information and Database structuring	5	Python
2	Programming using C#	6	HTML
3	Programming using C++	7	CSS 3
4	Programming using Visual Basic.Net (VB.net)	8	SQL server Database Query (DBQ)
5	ASP.net With C#	9	SQL server Database Administrator (DBA)
6	MATLAB	10	Web Services