

COUNTER UAV SWARMS: CHALLENGES, CONSIDERATIONS, AND FUTURE DIRECTIONS IN UAV WARFARE

Michelle Sherman, Sihua Shao, Xiang Sun, and Jun Zheng

ABSTRACT

Modern advances in unmanned aerial vehicle (UAV) technology have widened the scope of commercial and military applications. However, the increased dependency on wireless communications exposes UAVs to potential attacks and introduces new threats, especially from UAVs designed with the malicious intent of targeting vital infrastructures. Significant efforts have been made from researchers and other United States (U.S.) Department of Defense (DoD) agencies for developing countermeasures for detection, interception, or destruction of the malicious UAVs. One promising countermeasure is the use of a counter UAV (CUAV) swarm to detect, track, and neutralize the malicious UAV. This article aims to recognize the state-of-the-art swarm intelligence algorithms for achieving cooperative capture of a mobile target UAV. The major design and implementation challenges for swarm control, algorithm architecture, and safety protocols are considered. A prime challenge for UAV swarms is a robust communication infrastructure to enable accurate data transfer between UAVs for efficient path planning. A multi-agent deep reinforcement learning approach is applied to train a group of CUAVs to intercept a faster malicious UAV, while avoiding collisions among other CUAVs and non-cooperating obstacles (i.e., other aerial objects maneuvering in the area). The impact of the latency incurred through UAV-to-UAV communications is showcased and discussed with preliminary numerical results.

INTRODUCTION

Although unmanned aerial vehicles (UAVs) offer many positive benefits (e.g., surveillance [1] and mobile infrastructure [2]), there is a significant rise in UAVs used for malicious intent (such as flying over high-profile locations or military compounds) that poses threats to national security. According to the testimony of Samantha Vinograd, Department of Homeland Security (DHS), in 2022 before the U.S. Senate Committee on Homeland Security & Government Affairs, “The Transportation Security Administration (TSA), since 2021, has reported nearly 2,000 drone sightings near U.S. airports, including incursions at major airports nearly every day,” [3]. In addition, UAV technol-

ogy has been exploited to intercept communications, obtain/collect sensitive information or intelligence, smuggling of substances, conduct attacks by arming the UAV with small explosive payloads, or used as one-way attackers.

Current counter UAV (CUAV) technologies, or countermeasures, are divided into aerial or ground based platforms that fall into three main categories: *detection and tracking*, *neutralization*, and *cyber-attack* [4]. Both platforms rely on precise sensing for detection and estimation of the malicious UAV’s trajectory using radar, radio frequency (RF), acoustic or optical sensors/detectors. However, most of these techniques are short ranged, have a high false positive rate (especially distinguishing micro UAVs from birds), and are inefficient in noisy and dynamic scenarios [5]. Neutralization solutions can be divided into physical or non-physical systems that may disable a UAV or disrupt its communication systems, respectively. Physical systems may include launching a projectile or warfighter UAVs or interceptors, firing a net to capture the UAV, or using directional high-energy lasers or high-power microwave devices to produce a high voltage and current that destroys the main electronic circuitry in the UAV. Non-physical systems utilize RF jammers or global positioning system (GPS) spoofers to disrupt the control link and navigation of the UAV. Cyber-attack methods involve obtaining control or hijacking of the UAV.

Though many of these technologies have proven to be very effective at the neutralization of large UAVs, there are still several challenges when facing small/micro UAVs, UAVs with weak emissions, or weak radio signal reflection [6]. A multi-sensor network may be required to improve the detection capability. Moreover, most development efforts place an emphasis on the neutralization of a single target, not on the adaption to multiple targets or swarms of targets which may overwhelm ground-based solutions.

The U.S. military has invested in the development of defensive swarm tactics to detect, track, and attack an enemy swarm, even one that has evasive maneuvering capabilities. Swarming can be defined as the coordination of multiple UAVs that cooperate to achieve a common task. These UAVs can be equipped with various sensors and collect/

Michelle Sherman and Jun Zheng are with New Mexico Tech, USA;

Sihua Shao (corresponding author) is with Colorado School of Mines, USA; Xiang Sun is with University of New Mexico, USA.

Digital Object Identifier: 10.1109/MWC.003.2400047

share information to enable the swarm to autonomously make decisions. Since 2017, Defense Advanced Research Projects Agency (DARPA) has conducted several field experiments to evaluate the effectiveness and integration ability of the developed swarm tactics in both physical and virtual environments. The emerging development of swarm intelligence has led to the increased adoption of machine learning to detect and deter/capture enemy entities more rapidly.

The existing literature in CUAV swarms primarily focus on three types of cooperative formations: target-oriented (e.g., CUAVs track and capture an enemy UAV), defense-oriented (e.g., CUAVs form defensive boundaries to prevent an enemy from entering), and air-combat (e.g., CUAVs engage with the enemy swarm). The majority utilize optimization methods or deep learning strategies to improve the tracking efficiency but lack in scalability or do not consider practical scenarios (e.g., assuming no collisions will take place, CUAVs have higher speeds than the target UAV for faster capture, CUAVs have a global view of the environment, assume CUAVs have no communication infrastructure, etc.). To that end, there is a need for understanding the limitations and challenges of implementing a feasible cooperative solution. Therefore, in this article, we incorporate the practical constraints into a deep reinforcement learning (DRL) framework to demonstrate the effectiveness of CUAV communication-enabled cooperation to track and capture a target UAV. Thus, the contributions of this article are as follows:

- We categorize the cooperative CUAV swarm algorithms into three groups and identified the common cooperative design challenges, especially related to communication, collision avoidance, and scalability.
- We identified the training/learning paradigms and execution and efficient reward function design in a multi-agent deep reinforcement learning (MADRL) algorithm as the key technologies to incorporate the practical constraints in a cooperative CUAV scenario.
- We use the visualization of the CUAV trajectories to evaluate the detection, tracking, and neutralization behaviors and demonstrate the effectiveness of MADRL in a practical setting.
- We highlight the superiority of a communication-enabled scenario and numerically analyze the impact of latency on the performance.

ALGORITHMS FOR COOPERATIVE CUAV SWARMS

The general structure of a swarm intelligence algorithm [7] tasked with detection, localization, and interception of mobile targets involves three phases, which are graphically displayed in Fig. 1: *search*, *detection*, and *coordination*. This is sometimes referred to in the literature as the “chaser-target” or “pursuer-evader” problem, where the CUAVs are the pursuers and the targets/enemy UAVs are the evaders. “Pursuer-evader” problems generally only consider that the evader is trying to avoid capture. Here we also consider when the evader is tasked with some mission to reach a restricted zone while simultaneously avoiding capture from pursuers.

In the *search* phase, the CUAVs utilize on-board sensor readings to monitor the airspace within their search area. In the *detection* phase, the CUAVs

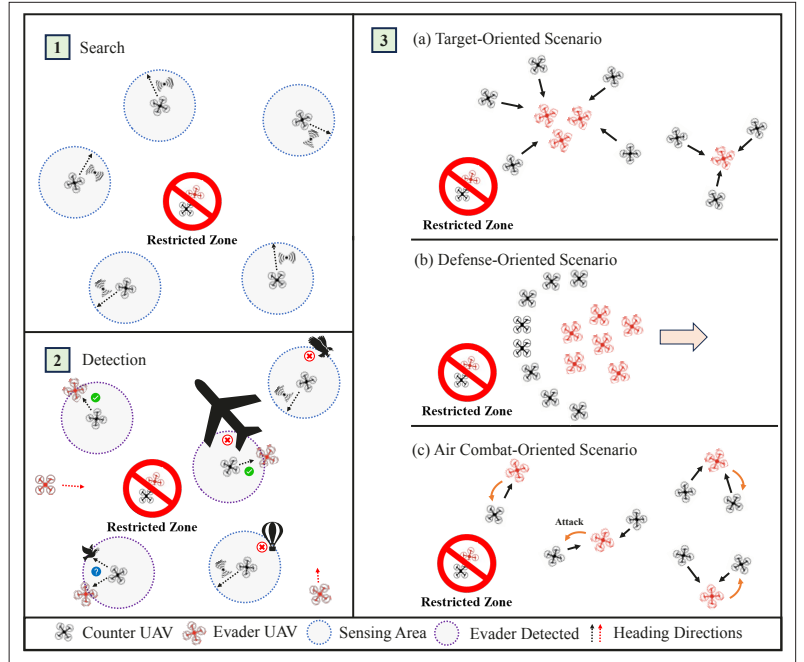


FIGURE 1. Example scenarios for Cooperative CUAV Swarm Algorithms.

detect any aerial objects within their search area, and must utilize a classification algorithm to distinguish between a target UAV and other aerial obstacles, such as birds, planes, helicopters, weather balloons, and so on. Finally, in the *coordination* or *converge* phase, the CUAVs track the identified target UAV to form a prediction of its flight path and configure themselves based on the predefined objective. The objective can fall into the categories of: target-oriented, defense oriented, or air combat.

TARGET-ORIENTED ALGORITHMS

In swarm intelligence algorithms tasked with target-oriented coordination, a counter swarm traps/captures an enemy swarm (e.g., by firing nets), or disables the enemy swarm (e.g., by crashing into it, firing projectiles, spoofed signals, or jammers) and causes them to crash as shown in Fig. 1: 3a. This scenario is most closely related to various hunting strategies and predator-prey interactions.

The formation a CUAV swarm will take on depends on the initial positions of the CUAVs, the flight path of the target, and the speed of the CUAVs and the target. The authors in [8] experimented with three different formations to track a target UAV: follow, surround (encirclement), and cone. Initially, a swarm may exhibit follow behavior as they try to catch up to a target. If the CUAVs are able to communicate with each other, neighboring CUAVs can attempt to surround the target and bring it to a standstill or prohibit any further movement, assuming the target also has some collision avoidance protocol. However, if the priority is to reduce the likelihood of CUAV collisions, the swarm may take on a cone/triangle shape around the target as they chase it.

DEFENSE-ORIENTED ALGORITHMS

A non-destructive approach of swarm intelligence is to form a defensive barrier to prevent the target from reaching the restricted zone, as shown in Fig. 1: 3b and restrict its flight path. The authors in [9]

In a dynamic environment with continuous mobility of CUAVs, collision avoidance is challenging. CUAVs must maintain relative observation information of other CUAVs and other aerial objects that reside in its sensing area.

proposed a CUAV system following a clustering procedure to intercept and escort a target away from a restricted zone. This is achieved through a modular design consisting of the phases: patrolling/deployment, clustering/formation for interception, and chase/escort. The formation of the swarm is governed by cluster heads (CHs) or "parent" UAVs which have branches or "child" UAVs. The type of action the swarm will take on (chase/escort) is governed by the distance between the CHs and the target. The authors in [10] took a different defensive approach by inserting a small number of CUAVs to infiltrate a target swarm with the objective of misleading the CUAVs from their original task. However, this work assumes that the CUAVs are able to utilize the targets' communication channels.

AIR COMBAT-ORIENTED ALGORITHMS

In an environment where the CUAVs and the targets are equipped with tools for counter attacks, an essential issue becomes modeling the air-combat behavior, as depicted in Fig. 1: 3c. The CUAVs need to exhibit offensive and defensive strategies while simultaneously coordinating as a unit. The authors in [11] developed a UAV swarm vs. swarm combat system where each CUAV's decision-making strategy to attack a target is guided by a profit metric in the target-allocation decision problem and the swarm motion decision problem.

Summary: CUAV control in each category of cooperative swarm algorithms has its own set of formation constraints and coordination strategy. However, there are some key considerations (discussed below) that are common to the three groups.

LIMITATIONS AND CHALLENGES

In the literature, various challenges related to enabling cooperative CUAV swarms with efficient search, track, and capture of a mobile target have been considered.

COMMUNICATION AMONG CUAVS

One common assumption in the pursuer-evader problem is to assume that the pursuers do not communicate or exchange information with other pursuers. That is, the CUAVs can act independently based on their own acquired observation data, but this may lead to a low target capture rate. However, in the later sections, it will be shown that by enabling communication, the behavior of a CUAV has the potential to be influenced given the context of other CUAVs' observed sensor data via some communication protocol. That is, a CUAV can choose to make decisions based on its own observations and the observations made by other CUAVs. The CUAVs can share observation data by broadcasting positional information of the detected target. Enabling communication among CUAVs is a crucial design consideration to enable timely capture of a target, especially under noisy conditions and a dynamic environment. In addition, the freshness and availability of the observation data is impacted by the latency incurred and the communication range of the CUAVs.

COLLISION AVOIDANCE

One key challenge to achieve efficient swarm coordination is designing a collision avoidance safety protocol. In a dynamic environment with

continuous mobility of CUAVs, collision avoidance is challenging. CUAVs must maintain relative observation information of other CUAVs and other aerial objects that reside in its sensing area. The safe distance protocol can be designed with a virtual attractive-repulsive potential function to achieve the various coordinated path-planning strategies as discussed in the previous section. Optimizing the design of the safety protocol becomes critical as it can effect the capture rate of the target, especially in the target-oriented swarm algorithms. For example, referring to Fig. 1: 3a, the designed algorithm may choose to not allow the CUAVs to proceed any closer to the targets or otherwise they risk collision. Meanwhile, since the CUAVs come to a standstill, the targets may be able to escape between the CUAVs and reach the restricted zone. In addition, the design protocol must be scalable and require limited processing on the CUAV end to achieve appropriate decision-making.

CUAV-TO-TARGET-UAV RATIO

When considering tracking and capturing a single target, the main issue becomes combining the time-varying observation measurements from multiple CUAVs to improve the target tracking accuracy. The work in [12] utilizes a DeepSet Neural Network to process the collected observation data of CUAVs and extract relevant features when they are available. When considering multiple targets, a few challenges arise depending on the ratio of the number of CUAVs to targets. If multiple targets are detected within the sensing area of a CUAV, observation data association becomes a challenge and the CUAV must follow some decision-making strategy on which target to follow. When there are enough CUAVs to pursue each target individually, then the problem becomes an assignment and hand-off problem. That is, a CUAV may pursue a target, but it may also hand-off this task to another CUAV that is closer and has a better chance of capturing the target. If there are not enough CUAVs to track and capture all of the targets, then a possible strategy is to enable some CUAVs to pursue the center of densely populated clusters of targets in order to spatially divert them whereby they can be more easily captured by other CUAVs.

COMPLEXITY OF AERIAL SPACE ENVIRONMENT

In this work, we are concerned with tracking and capturing moving targets. Most work has focused on mobile targets on a two-dimensional (2D) plane. The reasoning for this assumption is that the target will eventually descend to the same altitude of the restricted zone it is trying to reach. It is also assumed that the CUAVs are monitoring the area at this same altitude for all time. In real applications, the target may have maneuvering capabilities or is combat equipped in which 2D agility of the CUAVs would not be sufficient to protect the restricted zone.

CUAV COORDINATION AND CONTROL

The approaches for swarm control can be categorized into centralized and decentralized methods. Centralized coordination involves utilizing a back-end system to perform all of the processing of observation and state data from all of the

CUAVs, after which point a control decision is made and is distributed to the CUAVs. Although centralized approaches can achieve a better performance since the quality of the decision is based on a global view of the CUAVs, they are also limited in terms of scalability. Decentralized approaches generate decisions using only the information available to the CUAV regarding its state. In this case, distributed methods scale, but suffer from decision optimality, and are resource constrained. Hybrid methods are currently being investigated to combine the benefits offered by both approaches to enable neighboring CUAVs to exchange observation, state, and resource information, but the computation regarding the decision making is done on the back-end.

PRIOR KNOWLEDGE OF TARGET UAV'S TRAJECTORY AND OPERATION

In tracking methods, the motion of a target UAV can be categorized as deterministic or random. We may have some priori information about the policy of the target such that we can estimate its trajectory and reduce the search time. However, in real-life this may not always be the case. The trajectory of the target can be probabilistic or completely unknown.

KEY TECHNOLOGIES TOWARD COOPERATIVE SWARMS

Currently in the existing literature, the control framework for a cooperative UAV swarm is carried out by successive convex optimization or control theory methods. These methods are generally ideal for simple scenarios since their methodology requires precisely known model dynamics of the environment, which is not easily attainable or scalable for complex stochastic environments, and thus many of the practical constraints mentioned in the previous section are ignored in the algorithm design.

The thriving advances in DRL methods have empowered them to handle the nonlinear dynamics, uncertainties, and practical challenges associated with learning coordinated strategies among multiple agents. In multi-agent DRL (MADRL), the agents seek to maximize their own expected return through interactions with the environment. Since each agent is making sequential decisions in a shared environment, their interests may or may not be aligned. The algorithm must balance between exploring and exploiting the different interests of the CUAVs to see how they affect the overall performance of the system.

In general, the training/learning paradigm and execution can take on a centralized or distributed/decentralized form. The authors in [13] discuss the different training schemes, namely centralized training centralized execution (CTCE), distributed training distributed execution (DTDE), and centralized training distributed execution (CTDE). CTDE methods have gained considerable attention in the recent years since it leans more toward a practical scenario where agents have access to information gained by other agents that have the potential to influence their decisions; however, the agents will act in a distributive manner based on their derived policies.

For many real-world scenarios, MADRL approaches face a challenge with sparse, delayed, and dense rewards [14]. Sparse rewards are rewards that are received rarely, which can

impede the learning for an agent, as it may not have enough feedback from the environment to converge to an optimal policy. Delayed rewards imply the agent must wait for some period before it learns the impact from its actions. Dense rewards occur frequently which can speed up the learning, but may lead to a sub-optimal policy. Thus, a key problem in MADRL is developing a well-designed reward function for promoting cooperative behavior among the agents.

DRL ARCHITECTURE FOR TARGET-ORIENTED CUAV SWARM

To evaluate the design challenges, particularly in communications, collision avoidance, CUAV-to-target ratio, coordination and control, we consider a target-oriented CUAV swarm scenario as a paradigm. Notably, the developed DRL architecture can also be adapted to the other two CUAV swarm strategies, namely defense-oriented and air combat-oriented, with a few modifications to the observation space, reward function design, and associated formation constraints. The simulation environment is set up as a multi-agent system where each CUAV is considered an agent i , for $i \in [1, N]$, tasked with determining the best actions to take based on its current state and available observation data in order to capture the target UAV before it reaches a restricted zone. In addition, the agents must avoid collisions with a number of non-cooperating obstacles. The considered 2D topology is shown in Fig. 2.

The target follows its own virtual attractive/repulsive force policy to derive the optimal linear and angular velocity to reach the restricted zone quickly while avoiding capture or collisions with agents or obstacles.

In real-world scenarios, it may be possible that the target UAV is traveling at a higher velocity than the agents. Thus, we consider the problem of optimizing the linear and angular velocity for each agent to minimize the expectation of the euclidean distance between the agent and the target every time-step in order to achieve timely capture. The agents are subject to collision avoidance constraints and range of linear/angular velocity values.

In addition to the non-communication configurations studied in our previous works in [12] and [15], we investigate the performance when the agents are *enabled to communicate with each other* to share their observation data via a wireless link. However, this CUAV-to-CUAV link may suffer from network latency and thus we must analyze the performance of the models when considering the sharing of outdated sensing information.

REINFORCEMENT LEARNING (RL) FRAMEWORK

The CUAV swarm control is formulated as a MADRL problem following a CTDE paradigm and the partially observable Markov decision process (MDP), which is characterized by states, observations, actions, rewards, and associated transition and observation probabilities. At each step, the agent receives an interpretation of the environment (the state) and takes an action by following a certain policy. As a consequence, one step later, the agent receives a numerical reward and transitions to a new state where it receives an observation. The goal for each agent is to follow

To evaluate the design challenges, particularly in communications, collision avoidance, CUAV-to-target ratio, coordination and control, we consider a target-oriented CUAV swarm scenario as a paradigm. Notably, the developed DRL architecture can also be adapted to the other two CUAV swarm strategies, namely defense-oriented and air combat-oriented, with a few modifications to the observation space, reward function design, and associated formation constraints.

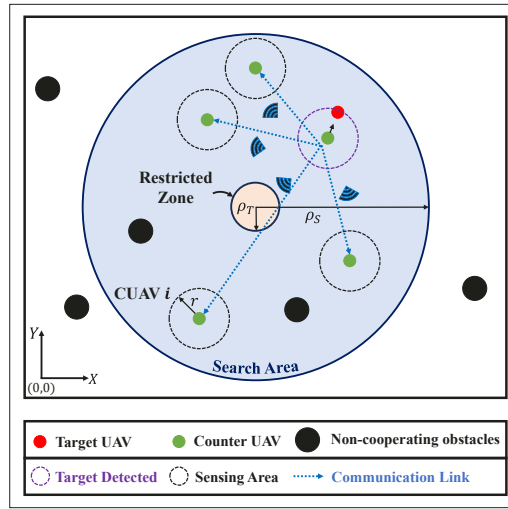


FIGURE 2. Topology of the considered target-oriented CUAV Swarm.

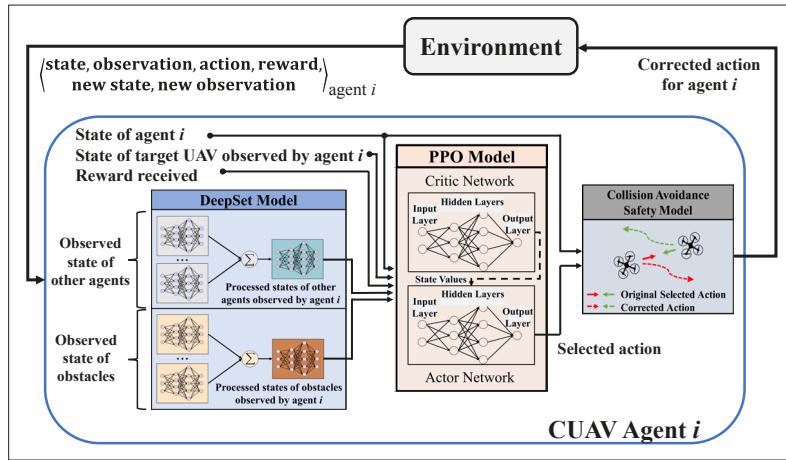


FIGURE 3. Block diagram of the designed DRL algorithm architecture.

polices that maximize its discounted expected return (essentially a cumulative sum of rewards the agent expects to receive following a policy).

State Space: The state of an agent is a tuple composed of its current: 2D location, heading angle, linear and angular velocity, and the distance and angle between the agent and the center of the restricted zone.

Observation Space: The observation of an agent is a nested tuple composed of multiple elements that capture the observation data (i.e., distance and angular information) relative to the target, other agents, and the obstacles. When assuming the CUAVs do not communicate with each other, the observation is only populated when a target, other UAV(s), or obstacle enters its sensing area. If the CUAVs are able to share information, then the time-varying observation is populated depending on the communication latency. Note: if in the current time-step t an agent i senses the target, its observation is populated with the current information at time t . In this work, we consider three scenarios:

- **Instantaneous communication:** once the target enters the sensing area of any UAV, the observation of all UAVs is populated with the derived information at time t .
- **Fixed Latency:** assumes all UAVs have the same

communication delay. At time t , if UAV i has not observed a target, then the observation is populated with the corresponding data that was derived some τ delay ago as long as the target was sensed by at least one UAV j (s.t. $j \neq i$).

- **Variable Latency:** assumes each UAV has a different backoff delay τ_i for $i \in [1, N]$. At time t , the observation for UAV i is populated with the corresponding data if the target was sensed τ_i ago by any UAV j (s.t. $j \neq i$).

Action Space: At each step, the action of an agent corresponds to adjusting its linear and angular velocity. The actions derived from the proximal policy optimization (PPO) actor network are then mapped to feasible actions subject to collision avoidance constraints.

Reward Function: The immediate reward function for each agent is characterized by six conditionals. A positive reward is assigned if the agent captures the malicious UAV. Penalties are assigned if the agent enters the restricted zone or collides with other agents or obstacles. Otherwise, if the malicious UAV enters the sensing area of the agent, a weighted negative penalty is assigned depending on the euclidean distance between the two entities.

DEEP REINFORCEMENT LEARNING ARCHITECTURE

The DRL architecture, shown in Fig. 3, is composed of three elements to solve the RL problem:

- A DeepSet network which comprises four networks to pre-process the dynamic observation tuple relative to other agents and to obstacles.
- A PPO algorithm is implemented to maximize the expected discounted future return. PPO is an on-policy based method which is comprised of two neural networks: a critic network, which evaluates the state-value function for a given policy (probability distribution of selecting each action), which is derived by an actor network.
- A safety algorithm designed to map the derived actions from the PPO actor network to a feasible action that will meet collision and search boundary constraints. Through numerous interactions with the environment, the CUAVs learn an optimal policy with the aim of maximizing the target capture rate when subject to collision safety constraints.

NUMERICAL RESULTS AND PERFORMANCE EVALUATION

The performance of the MADRL PPO-based CUAV Swarm control algorithm while considering cooperative agents is characterized by the convergence of the target UAV capture rate. Table 1 lists the main simulation setup and algorithm hyperparameters. During training, CUAVs are tasked with capturing the target UAV while avoiding collisions with each other or other aerial obstacles, as shown in Fig. 2. At the start of each episode, the positions of the aerial obstacles are randomly initialized within the $X \times Y$ grid area. The CUAVs are randomly initialized somewhere between the circular restricted zone and the search area. The location of the target UAV is randomly initialized somewhere outside of the $X \times Y$ grid area. An episode terminates if the maximum allotted time is reached, the target UAV is captured, or a CUAV enters the restricted zone. An iteration of training is defined as the collection and storage of up

to 65,536 transitions in a replay buffer, whereby mini-batches of size 4,096 are utilized to update the DeepSet, Actor, and Critic networks. Here, we investigate the impact of enabling CUAV-to-CUAV communications on the capture rate performance of the malicious UAV and analyzing the impact of communication latency.

NO COMMUNICATION VS. INSTANTANEOUS CUAV-TO-CUAV COMMUNICATION

We compare the performance of the instantaneous communication-enabled MADRL PPO-based algorithm versus a baseline non-communication system where the CUAVs know only their local environment state information. Figure 4 compares these algorithms for the mean target UAV capture rate in a training iteration, which is characterized as the number of episodes where the malicious UAV is captured over the number of episodes in a training iteration. The simulation considers the cases when there are $N = 4$ and $N = 8$ CUAVs.

From the plots, we can see that the communication enabled scenarios achieved a significantly better performance compared to the no communication scenarios. In fact, it was observed that the no communication scenarios seem to experience the lazy agent problem. Agents do not have direct access to global information which limits their ability to explore and converge to an optimal joint policy. Alternatively, when enabling CUAV-to-CUAV communication, we can achieve a significantly higher capture rate for both N cases due to the positional information exchange between the CUAVs once the target has reached the sensing area of a CUAV. However, these results assume that the information exchange is instantaneous. In the next two sections, we investigate the performance of the learning algorithm when there is communication delay caused by wireless contention.

CUAV-TO-CUAV COMMUNICATION WITH FIXED LATENCY

Here, we analyze the effect of fixed delayed information exchange between CUAVs. Figure 5a shows the mean target UAV capture rate for latency $\tau = 0.03$ sec. (equivalent to a delay of three time-steps) and $\tau = 0.1$ sec. (equivalent to a delay of ten time-steps) when $N = 8$ CUAVs. We can see that the impact of the additional latency has a slightly reduced performance in both cases compared to when the CUAVs have instantaneous communication. The reason is because, once a CUAV detected a target, then after some τ time later, the other CUAVs will start chasing the tail of the target's flight path since their decisions are derived from outdated information. As the delay increases, it is more likely that CUAVs are too late to capture the target before it reaches the restricted zone.

CUAV-TO-CUAV COMMUNICATION WITH VARIABLE LATENCY

Figure 5b compares the target capture rate performance for fixed latency and variable latency, where the communication delay for each CUAV i is given by τ_i such that the total average delay is 0.1 sec. among the $N = 8$ CUAVs per time-step. We can see that when considering the variable delay among the CUAVs, the capture rate is slightly reduced compared to the fixed latency results, since not all CUAVs will receive the notification that a target was detected at the same

Parameter	Value
Geographical Area $X \times Y$	160 m $\times$ 120 m
Number of CUAVs (N)	8
Number of Obstacles (O)	5
Search Area Radius (ρ_S)	60 m
Restricted Zone Radius (ρ_T)	5 m
CUAV Sensing Radius (r)	20 m
Collision Safety Distance (ϵ)	1.5 m
Maximum Length of Episode	40 sec.
Time-step size (t_s)	0.01 sec.
Replay Buffer Size	65536
Mini-batch Size	4096
Learning Rate for Actor & Critic (θ)	8.4856×10^{-5}

TABLE 1. Simulation parameter values.

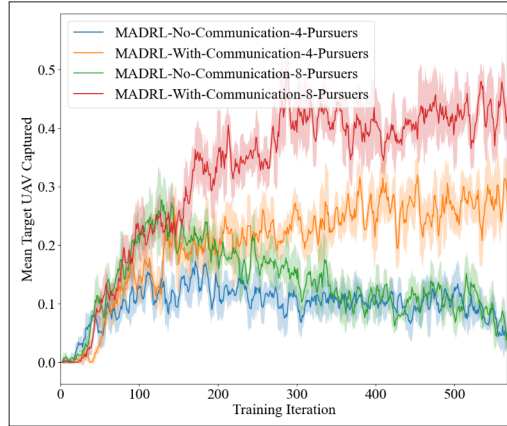


FIGURE 4. Target UAV capture rate training performance considering cooperative and non-cooperative CUAVs and assuming no communication latency.

time. We note that even with the impact of the variable latency, the performance results are still significantly better than when the CUAVs act independently and are not equipped for CUAV-to-CUAV communication.

CONCLUSION AND FUTURE DIRECTION

The control framework design challenges that are common to difference cooperative CUAV swarm strategies are related to communication, collision avoidance, complexity of the environment, and scalability. DRL can effectively incorporate the practical constraints into non-convex optimization problems through CTDE MADRL architecture and a well-structured reward function. The target UAV capture rate can be significantly enhanced by enabling sensing information exchange between CUAVs, at the expense of reduced performance as the average latency incurred between CUAVs increases. The communication latency incurred will exacerbate with a larger number of CUAVs and ultimately impair the swarm's ability to effectively coordinate and lead to sub-optimal policies. One of the future directions is to investigate delay-aware DRL architectures for managing the different delays associated with CUAV swarm control. For example, the communication and

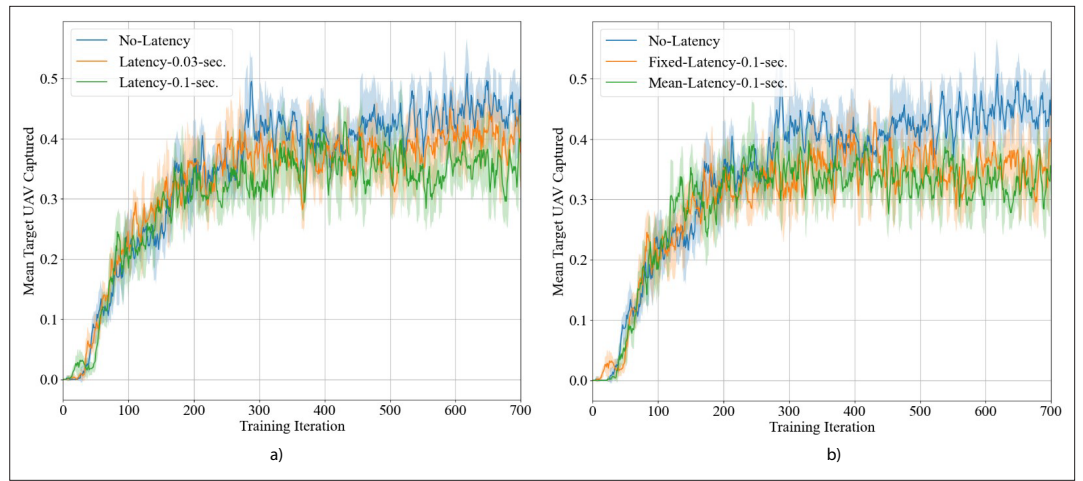


FIGURE 5. Target UAV capture rate training performance considering cooperative CUAVs with: a) fixed communication latency; b) variable communication latency $\tau_i, i \in [1, N]$ (see the attached supplementary media for a demonstration of the trained algorithm when the delay is fixed for $\tau_i = 0.03$ sec. $\forall i$).

processing delays associated with data transmission, which impact the freshness of an agent's absolute and relative observation(s) of the environment, and the delays associated with the control (action feedback and actuation delay), which will impact the reaction time of the agents and the learning process.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation under Award under grant no. CNS-2323050 and CNS-2148178, where CNS-2148178 is supported in part by funds from federal agency and industry partners as specified in the Resilient & Intelligent NextG Systems (RINGS) program.

REFERENCES

- [1] M. Sherman *et al.*, "UAV Assisted Cellular Networks With Renewable Energy Charging Infrastructure: A Reinforcement Learning Approach," *Proc. IEEE Military Commun. Conf.*, 2021, pp. 495–502.
- [2] M. Sherman *et al.*, "Optimizing Aol in UAVRIS Assisted IoT Networks: Off Policy vs. On Policy," *IEEE Internet of Things J.*, vol. 10, no. 14, 2023, pp. 12,401–15.
- [3] S. Vinograd, "Protecting the Homeland from Unmanned Aircraft Systems, HSGAC Congress," available: <https://www.hsgac.senate.gov/hearings/protecting-the-homeland-from-unmanned-aircraft-systems/>.
- [4] H. Kang *et al.*, "Protect Your Sky: A Survey of Counter Unmanned Aerial Vehicle Systems," *IEEE Access*, vol. 8, 2020, pp. 168,671–710.
- [5] V. Chamola *et al.*, "A Comprehensive Review of Unmanned Aerial Vehicle Attacks and Neutralization Techniques," *Ad Hoc Networks*, vol. 111, 2021, p. 102324.
- [6] A. Dudush *et al.*, "State of the Art and Problems of Defeat of Low, Slow and Small Unmanned Aerial Vehicles," *Advances in Military Technology*, vol. 13, no. 2, 2018, pp. 157–71.
- [7] J. Tang, H. Duan, and S. Lao, "Swarm Intelligence Algorithms for Multiple Unmanned Aerial Vehicles Collaboration: A Comprehensive Review," *Artificial Intelligence Review*, vol. 56, no. 5, 2023, pp. 4295–4327.
- [8] C. Arnold and J. Brown, "Performance Evaluation for Tracking a Malicious UAV using an Autonomous UAV Swarm," *Proc. IEEE Annual Ubiquitous Computing, Electronics Mobile Commun. Conf.*, 2020, pp. 0707–12.
- [9] M. R. Brust *et al.*, "Swarm-Based Counter UAV Defense System," *Discover Internet of Things*, vol. 1, 2021, pp. 1–19.
- [10] J. Simonjan, S. R. Probst, and M. Schranz, "Inducing Defenders to Mislead an Attacking UAV Swarm," *Proc. IEEE Int'l. Conf. Distributed Computing Systems Workshops*, 2022, pp. 278–83.
- [11] S. Shahid *et al.*, "Offense-Defense Distributed Decision Making for Swarm vs. Swarm Confrontation While Attacking

the Aircraft Carriers," *Drones*, vol. 6, no. 10, 2022.

- [12] J.-E. Pierre, X. Sun, and R. Fierro, "Multi-Agent Partial Observable Safe Reinforcement Learning for Counter Uncrewed Aerial Systems," *IEEE Access*, vol. 11, 2023, pp. 78,192–206.
- [13] S. Gronauer and K. Diepold, "Multi-Agent Deep Reinforcement Learning: A Survey," *Artificial Intelligence Review*, vol. 55, 2022, pp. 895–943.
- [14] B. Liu *et al.*, "Lazy Agents: A New Perspective on Solving Sparse Reward Problem in Multi-Agent Reinforcement Learning," *Proc. Int'l. Conf. Machine Learning*, 2023, pp. 21,937–50.
- [15] J.-E. Pierre *et al.*, "Multi-agent Deep Reinforcement Learning for Countering Uncrewed Aerial Systems," *Proc. Int'l. Symposium on Distributed Autonomous Robotic Systems*, 2022, pp. 394–407.

BIOGRAPHIES

MICHELLE SHERMAN [S'18] (michelle.sherman@student.nmt.edu) received the B.S. degrees in Mathematics and Electrical Engineering with a Minor in Optical Science & Engineering in 2020, and the M.S. degree in Mathematics with Specialization in Analysis in 2023 from New Mexico Tech, where she is currently finishing her Ph.D. degree in Electrical Engineering. Her research interests include machine learning, deep learning, UAV-assisted wireless communication and sensing networks, and UAV Swarms. Ms. Sherman was awarded as one of the Top Three Outstanding Engineering Students of the Year in 2020 by the Society of Professional Engineers at New Mexico Tech.

SIHUA SHAO [M'18, SM'23] (sihua.shao@mines.edu) received his Ph.D. degree and the Hashimoto Prize for best doctoral dissertation from the New Jersey Institute of Technology in 2018. He received NSF CRII Award and New Mexico EPSCoR Mentor Award in 2023. Currently, he is an assistant professor with the Department of Electrical Engineering at Colorado School of Mines. His research interests include wireless communications, wireless networks, and machine learning.

XIANG SUN [S'13, M'18] (sunxiang@unm.edu) is an assistant professor with the Department of Electrical and Computer Engineering at the University of New Mexico. He received his Ph.D. degree in Electrical Engineering from New Jersey Institute of Technology in 2018. His research interests include deep reinforcement learning, distributed machine learning, generative AI, Internet of Things, free space optics, wireless networks, and edge computing. Some of his recognitions include 2016 IEEE Int'l. Conf. Communications (ICC) Best Paper Award, 2018 NJIT Hashimoto Prize, 2018 InterDigital Innovation Award on IoT Semantic Mashup, and 2019 IECE Communications Society Best Tutorial Paper Award.

JUN ZHENG [M'05] (junezheng@nmt.edu) is a Professor in the Department of Computer Science and Engineering at the New Mexico Institute of Mining and Technology. He received his Ph.D. degree in Computer Engineering from the University of Nevada, Las Vegas, in 2005. His current research interests include cybersecurity and applied machine learning.