Deception in Drone Surveillance Missions: Strategic vs. Learning Approaches

Zelin Wan
zelin@vt.edu
Virginia Tech
Falls Church, VA, USA

Jin-Hee Cho
jicho@vt.edu
Virginia Tech
Falls Church, VA, USA

Mu Zhu
mzhu5@ncsu.edu
North Carolina State University
Raleigh, NC, USA

Ahmed H. Anwar
ahmed.h.hemida.ctr@army.mil
DEVCOM Army Research Laboratory
Adelphi, MD, USA

Charles Kamhoua
charles.a.kamhoua.civ@mail.mil
DEVCOM Army Research Laboratory
Adelphi, MD, USA

Munindar P. Singh
mpsingh@ncsu.edu
North Carolina State University
Raleigh, NC, USA

ABSTRACT

Unmanned Aerial Vehicles (UAVs) have been used for surveillance operations, search and rescue missions, and delivery services. Given their importance and versatility, they naturally become targets for cyberattacks. Denial-of-Service (DoS) attacks are commonly considered to exhaust their resources or crash UAVs (or drones). This work proposes a unique proactive defense using honey drones (HD) for UAVs during surveillance operations. These HDs use lightweight virtual machines to lure and redirect potential DoS attacks. Both the choice of target by the attacker and the HD's deceptive tactics are influenced by the strength of the radio signal. However, a critical trade-off exists in that stronger signals can deplete battery life, while weaker signals can negatively affect the connectivity of a drone fleet network. To address this, we formulate an optimization problem to select the best strategies for an attacker or defender in selecting their signal strength level. We propose a novel HD-based defense to identify the optimal setting using deep reinforcement learning (DRL) or game theory and compare their performance with that of non-HD-based methods, such as Intrusion Detection Systems and ContainerDrone. Our experiments demonstrate the unique benefits and superior efficacy of each HD-based defense across various attack scenarios.

CCS CONCEPTS

• **Computing methodologies** → **Reinforcement learning**; • **Security and privacy** → **Denial-of-service attacks**; • **Theory of computation** → **Algorithmic game theory**.

KEYWORDS

Honey drone, defensive deception, unmanned aerial vehicle, mission effectiveness, game theory, deep reinforcement learning

ACM Reference Format:
Zelin Wan, Jin-Hee Cho, Mu Zhu, Ahmed H. Anwar, Charles Kamhoua, and Munindar P. Singh. 2023. Deception in Drone Surveillance Missions: Strategic vs. Learning Approaches. In *The Twenty-fourth International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing (MobiHoc '23), October 23–26, 2023, Washington, DC, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3565287.3616525

MobiHoc '23, October 23–26, 2023, Washington, DC, USA

ACM ISBN 978-1-4503-9926-5/23/10.
https://doi.org/10.1145/3565287.3616525

1 INTRODUCTION

Our work presents a novel approach to mitigate DoS attacks by employing defensive deception (DD) tactics within Unmanned Aerial Vehicles (UAVs) systems [23]. We propose a surveillance mission system that utilizes honey drones (HDs), a specialized form of a drone-based honeypot, to combat Denial-of-Service (DoS) attacks while performing relay service. Unlike techniques using Raspberry Pi to emulate static drones [6], our HDs are drone-based, equipped with specifically vulnerable software and dynamic signal strength. These HDs function as proactive decoys, attracting and disorienting cyber attackers, collecting crucial attack intelligence, and dynamically reconfiguring system settings as a response.

This work identifies optimal settings under which game theory or deep reinforcement learning (DRL)-based HD defense is used while investigating the advantages and constraints of each strategy. This work has the following **key contributions:** (1)

- **Defensive Deception using Honey Drones**: We design a surveillance mission system where HDs, serving as drone-based mobile honeypots with intentionally vulnerable software, aim to attract DoS attacks. These drones act as proactive decoys to collect crucial attack intelligence and allow responses to the detected threat.
- **Intelligent Attack-Defense Game Modeling Under Uncertainty**: We create an attack-defense interaction model, which allows both the attacker and defender to adopt intelligent strategies. This intelligent strategy selection will enrich defensive deception research by introducing promising proactive defense strategies under diverse cyber games.
- **Extensive Comparative Performance Validation & Analyses**: We validate the performance of the developed HD-based defenses via extensive experiments and demonstrate their superiority under various attack scenarios in mission performance and energy conservation.

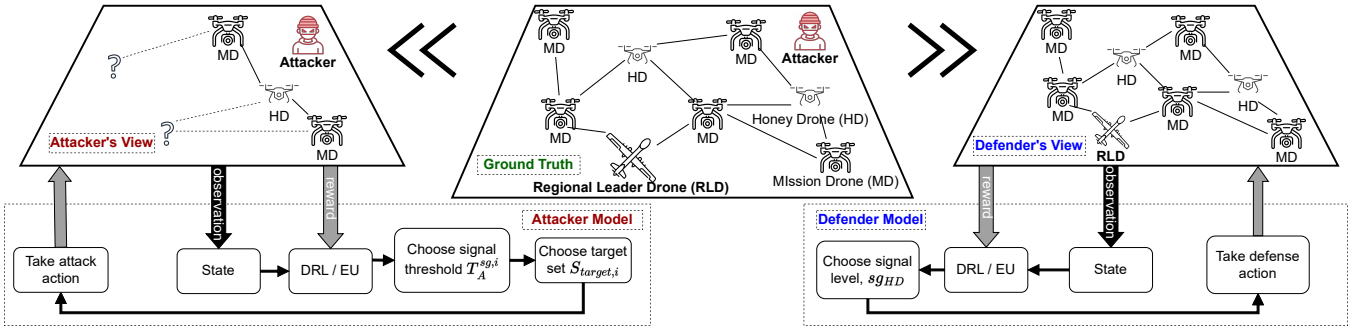


Figure 1: The proposed model of honey drone mission systems under the scenario where both the DoS attacker and defender leverage either DRL or GT to choose their most effective strategies, specifically the optimal signal strength.

2 RELATED WORK

The existing body of research has presented various strategies for defending UAVs against DoS attacks. Those include a resource allocation technique [4], hierarchical detection system [16], intrusion detection systems (IDS) [12] , and injection detector [17].

Game-Theoretic Defensive Deception: Various defensive deception (DD) applications have adopted game theory, such as Cumulative Prospect Theory [22] or a multi-stage Stackelberg game [2], to analyze attacker behavior. Hypergame Theory, another game-theoretic approach, has been employed to develop more advanced defensive deception techniques [1, 19, 20]. These studies underscore the potential of Hypergame Theory in devising robust strategies against cyber threats. In addition, no prior works above have considered UAV contexts with high dynamics, resource-constraints, and time-sensitive tasks.

DRL-based Defensive Deception: DRL has also been popularly leveraged for optimizing the effectiveness of DD approaches and examining various vulnerabilities [10] . In addition, deceptive signals as a defense against the vulnerabilities of RL have been examined [8]. In addition, the optimal selection of proactive defenses, such as moving target defense and DD, was studied using DRL [3]. DRL was also used to strengthen UAV communications [11].

Limitations of the Related State-of-the-Art Defensive Deception Research: Although defensive deception techniques often employ either game theory or Deep Reinforcement Learning (DRL), they are seldom used in conjunction or compared directly to investigate the advantage of each approach. Since each approach using game theory or DRL to solve a given problem has not been compared, there is a lack of understanding of how to best leverage the merit of each technique under certain conditions. We will fill this gap and contribute to identifying the best approach to be leveraged depending on environmental and system conditions.

3 SYSTEM MODEL

3.1 Network Model

We envisage a drone fleet for surveillance operations within a targeted region, aiming to maximize mission effectiveness while defending against DoS attacks. The network design includes a regional leader drone (RLD) communicating with the ground control station (GCS) through a satellite network [5], and mission drones (MDs) and honey drones (HDs) connected to the RLD via WiFi, forming a *flying ad hoc network* (FANET) [9]. Each drone maintains the Neighbor Table (NT) and Fleet Table (FT) for location and mission crew status. On receipt of a *hello* message from another drone, a connection setup procedure starts, encompassing a TCP handshake and data transmission over UDP[9].

3.2 Node Model

Our network incorporates a Ground Control Station (GCS) for task distribution, a charging station (CS) for drone power replenishment, and UAVs composed of an RLD, multiple MDs, and multiple HDs. The GCS monitors the mission's progression, the CS charges the drones, and the RLD adjusts the drones' routes and signal strength levels dynamically. MDs adhere to a specified path and transmit data to the RLD via multi-hop communication, while HDs can serve as mobile drone-based honeypots to lure DoS attacks.

3.3 Energy Model

For our simulation, we deploy Crazyflie 2.X quadrotor drones and consider the energy utilization of both MDs and HDs following a model based on their different operational rates [13].

3.4 Threat Model

We focus on DoS attacks, a prevalent and severe threat for UAVs. These attacks execute by sending numerous simultaneous JSON connection requests to a drone under attack, disrupting its network connectivity and leading to a drone crash [7]. To evaluate a drone's software vulnerability, we employ the Common Vulnerability Scoring System (CVSS), symbolized as a real value, $\text{vul}_\kappa \in [0, 1]$, signifying the probability of a successful attack compromising drone κ[18]. Figure1 portrays the high-level conceptual layout of our proposed honey drone mission system and the way agents choose attack/defense tactics.

4 STRATEGY SELECTION

This section tackles the challenge in mission systems arising from the absence of pre-existing information about attack patterns, necessitating an autonomous decision-making mechanism. We employ game theory and DRL to address this. We characterize the time taken for mission completion as T_M, with an upper limit denoted

as $T_M^{\max}$. The mission is divided into several rounds of interaction between the attacker and the defender.

Both attacker and defender have ten interaction strategies. This number was determined through preliminary experiments, where we assessed various strategy counts. Ten strategies emerged as the optimal balance, ensuring computational efficiency without oversimplifying the environmental conditions. We tested other numbers of strategies and found there was no significant difference. The design of the three subgames also considered the number of rounds a game plays before the mission terminates, ensuring players accumulate sufficient interaction experiences to form beliefs. Introducing too many subgames may dilute these experiences and make belief formation challenging.

4.1 Attacker Model

4.1.1 **Attacker's Action Space.** The attacker observes the drones' signal strengths and selects its attack strategy, $AS_i \in \{AS_1, \ldots, AS_{10}\}$, where each action corresponds to a range of received signal strengths $[sg_i^l, sg_i^u]$. The signal strength decreases as the distance between the transmitter and receiver increases and is estimated by the signal attenuation formula $P_{dBm}(d) = P_{dBm}(d_0) - \eta \cdot 10 \cdot \log_{10}(\frac{d}{d_0})$ where $\eta = 4$, and $P_{dBm}(d)/P_{dBm}(d_0)$ is the observed signal strength at a distance d/d_0.

4.1.2 **Attack Strategy Selection using Game Theory (GT).** The game theory (GT) agent for the attacker identifies an optimal attack strategy AS_i based on the expected utility yielded by choosing strategy i. This is calculated as:

$$
\begin{aligned}
EU^A(AS_i, C_\Sigma^A) &= \sum_{j=1}^{m} S_j^A \cdot u_{ij}^A, \text{ where } u_{ij}^A = G_{ij}^A - L_{ij}^A, \\
G_{ij}^A &= \mathrm{ai}_{ij}^A + \mathrm{dc}_{ij}^A, \quad L_{ij}^A = \mathrm{di}_{ij}^A + \mathrm{ac}_{ij}^A, \\
\mathrm{ai}_{ij}^A &= \frac{\sum_{\kappa_j \in S_{\text{target},i}^A} \mathcal{ASR}'_{\kappa_j} \mathrm{C}_{\kappa_j}}{\zeta}, \\
\mathrm{ac}_{ij}^A = e^{|S_{\text{target},i}| - \zeta}, \quad \mathrm{di}_{ij}^A &= 1 - \mathrm{ai}_{ij}^A, \quad \mathrm{dc}_{ij}^A = \frac{j}{\mathrm{sig}_{max}} + \mathrm{ai}_{ij}^A,
\end{aligned}
\tag{1}
$$

where S_j^A is the attacker's belief regarding the defender's strategy choice j. The terms G_{ij}^A and L_{ij}^A refer to the attacker's gain and loss respectively. $S_{\text{target},i}^A$ indicates the set of target drones when the attacker picks strategy i, while ζ refers to the attack budget. $\mathcal{ASR}'_{\kappa_j}$ is the anticipated attack success ratio for drone κ_j. The term C_{κ_j} denotes the criticality of drone κ_j, and sig_{max} is indicative of the maximum signal strength (normalized to 10).

4.1.3 **Attack Strategy Selection using DRL.** The objective of the attacker DRL agent is to select the optimal attack strategy to maximize the accumulated reward, G^A. The decision-making process of this agent, including its state, action set, and reward function, is described as follows:

- The **State** ($\mathcal{S}t^A$) can be expressed as $\mathcal{S}t^A = (NTR^t)$, where NTR^t is the count of drones within each signal strength range at round t.
- The **Action Set** ($\mathcal{A}^A$) is defined by $\mathcal{A}^A = \{a_1, \ldots, a_i, \ldots, a_n\}$, where each a_i equates to AS_i and determines the subset of target drones, represented as $S_{\text{target},i}$. The action i executed by the attacker DRL agent in round t is denoted as a_i^t.
- The **Reward Function** ($\mathcal{R}_t^A(a_i^t)$) is the immediate reward an attacker obtains by executing a_i^t, and is given by $\mathcal{N}_{MNC}^t$, which is the number of unfulfilled mission tasks in round t. The accumulated attack reward, symbolized as G^A, is computed by $G^A = \sum_{t=0}^{\infty} (\gamma^A)^t \cdot \mathcal{R}_t^A$, where γ^A represents the decay factor of the attacker.

4.2 Defender Model

4.2.1 **Defender's Action Space.** Our defense strategy $DS_j \in DS_1, \ldots, DS_{10}$ controls the HDs' signal strength sg_{HD}. The signal strength of MDs is set as $sg_{MD} = sg_{HD} - \rho$, where ρ is a predefined integer to ensure a stronger signal strength for HDs. The signal transmission range is uniformly divided from 100m to 1000m. Game theory and DRL are utilized to find the optimal defensive strategy, sg_{HD}, to modulate the signal strength levels of both MDs and HDs.

4.2.2 **Defense Strategy Selection using Game Theory (GT).** The defender's expected utility when taking defense strategy j is computed based on the multiplication of the defender's belief, S_i^D, in an attacker choosing attack strategy i and the utility of the defender for every defense strategy against each attack strategy, u_{ji}^D. This utility signifies the difference between the defender's gain and loss. The gain takes into account the decreased security vulnerability by defense strategy j and the attack cost by selecting attack strategy i. Conversely, the loss comprises the negative effect introduced by attack strategy i and defense cost by opting for defense strategy j. The defender's utility when choosing strategy j is:

$$
\begin{aligned}
EU^D(DS_j, C_\Sigma^D) &= \sum_{i=1}^{n} S_i^D \cdot u_{ji}^D, \quad u_{ji}^D = G_{ji}^D - L_{ji}^D, \\
G_{ji}^D &= \mathrm{di}_{ji}^D + \mathrm{ac}_{ji}^D, \quad L_{ji}^A = \mathrm{ai}_{ji}^D + \mathrm{dc}_{ji}^D, \\
\mathrm{di}_{ji}^D &= 1 - \frac{\sum_{\kappa \in S_{\text{target},i}^D} \mathrm{vul}_\kappa}{\zeta} + \frac{N'_{connect,j}}{N_{drone}}, \\
\mathrm{dc}_{ji}^D = e^{j - \mathrm{sig}_{max}}, \quad \mathrm{ai}_{ji}^D &= 1 - \mathrm{di}_{ji}^D, \quad \mathrm{ac}_{ji}^D = \frac{|S_{\text{target},i}^D|}{\zeta},
\end{aligned}
\tag{2}
$$

where C_Σ^D is the defender's beliefs toward attack strategies. The G_{ji}^D and L_{ji}^D denote the defender's gain and loss. The vul_κ refers to the vulnerability level of drone κ in range $[0, 1]$, as mentioned in Section 3.4. The number of target drones perceived by the defender in $S_{\text{target},i}^D$ is based on their experience. The defender maintains a record of which drones are targeted when the attacker opts for strategy i. The ζ is the attack budget. $N'_{connect,j}$ is the anticipated number of connected drones after choosing DS_j, and N_{drone} is the total number of drones initially allocated to the mission team. The sig_{max} denotes the maximum signal level (i.e., 10).

4.2.3 **Defense Strategy Selection using DRL.** The defender DRL agent aims to optimize the drones' signal strength, including both MDs and HDs, by maximizing the total accumulated reward. The state, action set, and reward used by the defender DRL agent are as follows:

- **State** ($\mathcal{S}_t^D$) is composed of the mission completion ratio and the scan progress map, defined as $\mathcal{S}_t^D = (\mathcal{R}_{MC}^t, \mathcal{M}_{SP}^t)$ where $\mathcal{R}_{MC}^t$ is the ratio of completed mission tasks at round t in range $[0, 1]$. $\mathcal{M}_{SP}^t$ is a map showing the scan progress for each cell at round t. Each cell value in the target area reflects the level of scanning progress, providing a detailed overview of the surveillance status of the target area.
- **Action Set** ($\mathcal{A}^D$) is denoted by $\mathcal{A}^D = \{a_1, \cdots, a_j, \cdots, a_m\}$ where each action a_j signifies a defense strategy DS_j indicating the signal strength of the HDs. For MDs, the descriptions in Section 4.2.1 apply. The action j selected by the defender in round t is represented as a_j^t.
- **Reward Function** ($\mathcal{R}_t^D(a_j^t)$), a defender's immediate reward by executing action a_j^t, is given by $\mathcal{N}_{MC}^t$, the number of mission tasks completed in round t. The accumulated defense reward, G^D, is calculated by $G^D = \sum_{t=0}^{\infty} (\gamma^D)^t \cdot \mathcal{R}_t^D$, where γ^D is the defender's decay factor.

5 EXPERIMENT SETUP

5.1 Simulation Environment Setup

Experiments were conducted in a Python 3.10 simulated environment using PyTorch and NetworkX. The surveillance area is a 750m x 750m grid divided into 25 cells. The drone fleet consists of 15 MDs and 5 HDs. Drones with low battery return to the charging station, and if no additional MDs are available, the mission proceeds with fewer MDs. We employed the A2C algorithm for DRL and used a memory buffer storing up to 10,000 transitions. Prioritized experience replay [15] was also integrated to emphasize high TD error transitions.

5.2 Metrics

For experimental verification, we consider the following criteria: (1) **Ratio of Completed Mission Tasks ($\mathcal{R}_{MC}$)**, which quantifies the proportion of completed cells among all assigned cells during the mission duration; (2) **Energy Consumption ($\mathcal{EC}$)** accounts for the cumulative energy utilization by all drones, encompassing both HDs and MDs; and (3) **Number of Active, Connected Drones ($\mathcal{N}_{AC}$)** estimates the number of non-compromised MDs participating in the mission execution.

5.3 Comparing Schemes

We compare the performance of the following schemes: (1) **HD-F**: HD-based approach using a fixed signal strength level (i.e., 5); (2) **HD-DRL**: HD-based approach with the optimal signal strength level identified by DRL; (3) **HD-GT**: HD-based approach with the optimal signal strength level identified by GT; (4) **IDS** [14]: Intrusion detection system-based approach to detect and isolate DoS attacks; (5) **CD**: ContainerDrone [4] which stops working when detecting DoS attacks; and (6) **No-Defense**: No defense is used.

6 NUMERICAL RESULTS AND ANALYSES

6.1 Ratio of Mission Completion

Figure 2 presents the performance of various defense schemes (see Section 5.3) against DoS attacks based on the ratio of completed mission tasks ($\mathcal{R}_{MC}$). Key observations include: (1) HD-DRL excels against fixed or DRL-based attack strategies. Clearer action patterns in these attacks, especially with DRL, allow the defender to counteract more effectively. The defender's reward, linked to completed mission tasks (Section 4.2.3), results in a higher $\mathcal{R}_{MC}$ when the attacker's patterns are more discernible. (2) HD-GT performs well in early game rounds due to GT's strategic forecasting based on a rapidly formed payoff matrix. However, its advantage diminishes over time due to GT's limited adaptability. Conversely, DRL strategies, while initially slower due to their learning curve, improve over time, eventually outpacing GT-based approaches. (3) HD-DRL shows performance fluctuations, particularly against intelligent adversaries using GT or DRL. These adversaries amplify the defender's optimization complexity, causing more explorative behaviors in the DRL agent. Additionally, the presence of an intelligent opponent introduces non-stationarity challenges. In this multi-agent environment, strategies evolve and change the goals of the optimal status with fluctuating learning curve.

6.2 Energy Consumption

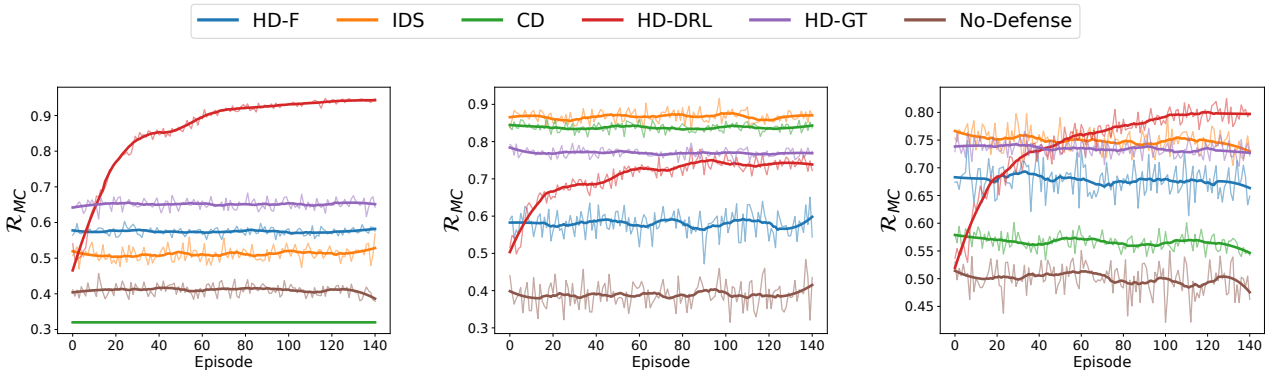
Figure 3 illustrates the energy consumption of various defense strategies (i.e., HD-F, HD-DRL, HD-GT, IDS, CD, and No-Defense) against DoS attacks, measured using the $\mathcal{EC}$ metric. Based on Figure 3, we observe: (1) HD-GT demonstrates the most energy-efficient strategy among the various defense mechanisms considered. This is attributed to the fact that the GT-based agent takes signal strength-based defense costs into account while deciding on a strategy (see the details in Section 4.2.2), leading to the lowest $\mathcal{EC}$ and hence conserving energy. (2) Notably, when the attacker employs GT or DRL to select its attack strategy, HDs employing intelligent defense mechanisms (i.e., HD-DRL and HD-GT) show lower $\mathcal{EC}$ than other defense techniques (i.e., IDS and CD). As seen in Figure 2, intelligent HD-based strategies can effectively balance mission performance with energy conservation, showing a high $\mathcal{R}_{MC}$ and maintaining a lower energy cost in $\mathcal{EC}$.

6.3 Number of Active, Connected Drones

Figure 4 presents the number of active, connected drones for various defense strategies (i.e., HD-F, HD-DRL, HD-GT, IDS, CD, and No-Defense) in response to DoS attacks, measured using the $\mathcal{N}_{AC}$ metric. Based on Figure 4, we make the following observations: (1) HD-based defenses (i.e., HD-F, HD-DRL, and HD-GT) effectively maintain the connectivity of the drone fleet. This is because HDs can serve as relays, facilitating mission drones' connection to the regional leader drone. (2) High connectivity, ensured by HDs, does not necessarily lead to increased energy consumption, particularly when intelligent strategies are employed. Similar to Figure 3, HD-based defenses demonstrate high $\mathcal{N}_{AC}$ while HD-DRL and HD-GT exhibit lower energy consumption in Figures 3b and 3c. These intelligent strategies do not rely solely on high signal strength. Instead, they occasionally utilize lower signal strengths to avoid being targeted by attackers, resulting in efficient energy conservation.

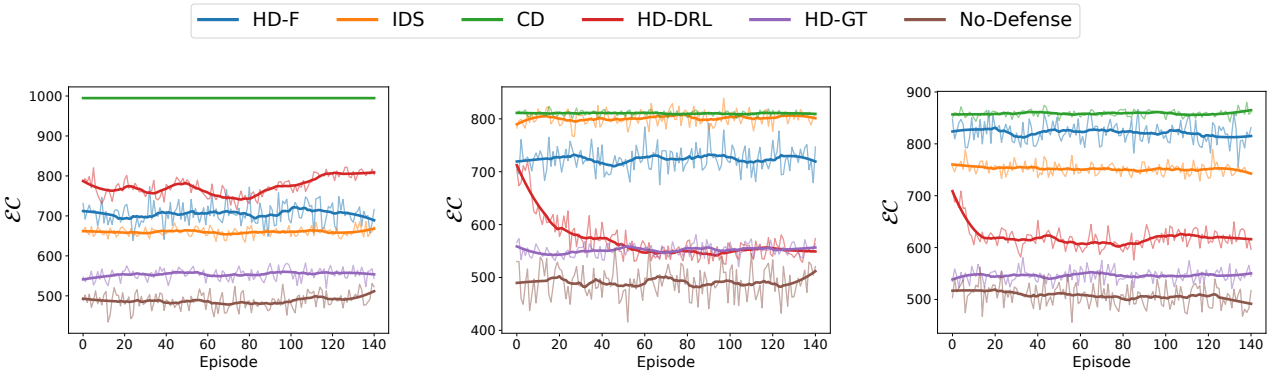
7 CONCLUSION & FUTURE WORK

This study compared various HD-based defenses with non-HD-based counterparts (i.e., IDS and CD) when intelligent strategy



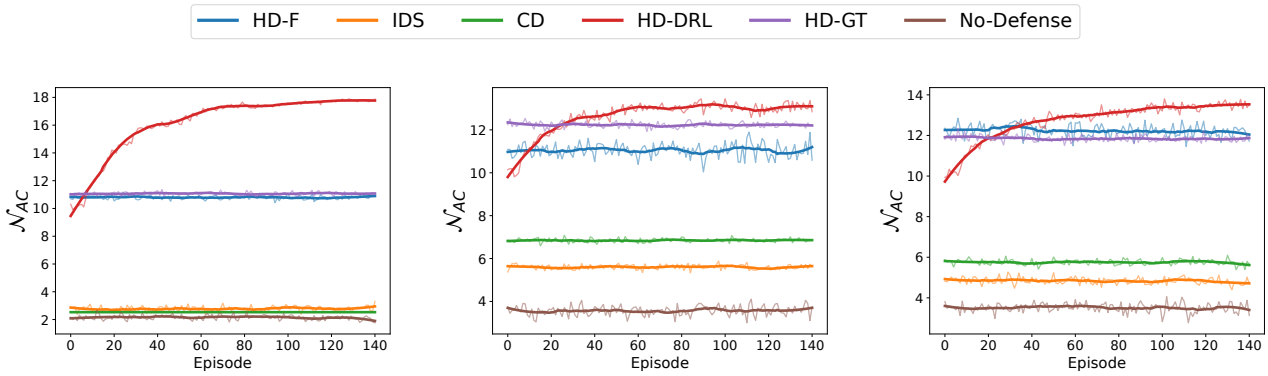
(a) $\mathcal{R}_{MC}$ under fixed attack. (b) $\mathcal{R}_{MC}$ under GT attack. (c) $\mathcal{R}_{MC}$ under DRL attack.

Figure 2: Performance analysis of HD-DRL, HD-GT, IDS, CD, and fixed defense, given an attack strategy with respect to the ratio of completed mission tasks ($\mathcal{R}_{MC}$).



(a) $\mathcal{EC}$ under fixed attack. (b) $\mathcal{EC}$ under GT attack. (c) $\mathcal{EC}$ under DRL attack.

Figure 3: Performance analysis of HD-DRL, HD-GT, IDS, CD, and fixed defense, given an attack strategy with respect to energy consumption ($\mathcal{EC}$).



(a) $\mathcal{N}_{AC}$ under fixed attack. (b) $\mathcal{N}_{AC}$ under GT attack. (c) $\mathcal{N}_{AC}$ under DRL attack.

Figure 4: Performance analysis of HD-DRL, HD-GT, IDS, CD, and fixed defense, given an attack strategy, with respect to the number of active, connected drones ($\mathcal{N}_{AC}$).

selection methods are used based on deep reinforcement learning or game theory in terms of the mission completion ratio, energy consumption, and the number of active, connected drones.

Via the extensive experiments, we obtained the following **key findings**: (1) HD-based defenses outperformed IDS and CD, where both HD-DRL and HD-GT offer distinctive advantages. These HD-based defenses efficiently maintained the connectivity of the drone fleet, ensured high mission completion ratios, and regulated energy consumption effectively by properly defending against DoS attacks. This was primarily achieved by intelligently varying the signal strength to minimize security vulnerabilities while maintaining strong mission performance and energy efficiency. (2) When comparing HD-DRL and HD-GT, each displayed unique strengths depending on the specific context. HD-DRL showed superior performance under fixed or DRL-based attack strategy, which tends to exhibit clearer action patterns which made it easier for HD-DRL to identify and counter the attacks. DRL's autonomous and continuous learning capabilities based on complex neural networks enabled the HD-DRL strategy to gradually improve its performance, eventually surpassing other strategies. (3) On the other hand, HD-GT demonstrated initial advantages, particularly in the early stages of the game, due to the strategic forecasting abilities of game theory. However, as the game advanced, this advantage diminished due to GT's limited explorability for optimal solutions under high dynamics. Nevertheless, HD-GT stood out in energy efficiency, consuming less energy than other defenses while maintaining high mission completion rates and drone connectivity. Hence, HD-GT can provide its high merit under resource-constrained environments, which requires significantly low energy consumption. (4) Overall, the choice between HD-DRL and HD-GT would depend on the specific circumstances. For scenarios with predictable or fixed attacker strategies or where long-term learning and adaptation are required or allowed, HD-DRL would be a preferred choice. Conversely, for situations requiring immediate effective defenses or where energy efficiency is a paramount concern, HD-GT can provide a feasible, attractive solution.

As for **future work directions**, we aim to extend this work by (1) exploring different types of cyberattacks beyond DoS, to ascertain the effectiveness of HD-based deceptive defense techniques; (2) incorporating *transfer learning* [21] to counteract the initial performance drop in RL, and to evaluate its advantages over GT in aspects such as mission efficacy and efficiency (including computational burden like training duration); (3) designing more realistic and methodical mechanisms to evaluate agents' perceived uncertainty that aligns with real-world scenarios; (4) conducting sensitivity analysis by varying scenario settings, to thoroughly assess the robustness and scalability of our approach across different environments; and (5) exploring other application scenarios to evaluate and validate our technique further. This will provide a more comprehensive understanding of the limitations and potential of our method, aligning it closer with practical implementations.

ACKNOWLEDGMENT

This work is partly supported by the Army Research Office and Army Research Laboratory under Grant Contract Numbers W911NF-20-2-0140, W911NF-19-2-0150, W911NF-17-1-0370, and W911NF-23-2-0012. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes, notwithstanding any copyright notation herein.

REFERENCES

[1] A. H Anwar et al. Honeypot-based cyber deception against malicious reconnaissance via hypergame theory. In *GLOBECOM 2022-2022 Global Communications Conference*, pages 3393–3398. IEEE, 2022.
[2] A. Basak et al. Identifying stealthy attackers in a game theoretic framework using deception. In *International Conference on Decision and Game Theory for Security*, pages 21–32. Springer, 2019.
[3] A. Charpentier, N. Boulahia Cuppens, F. Cuppens, and R. Yaich. Deep reinforcement learning-based defense strategy selection. In *Proceedings of the 17th International Conference on Availability, Reliability and Security*, pages 1–11, 2022.
[4] J. Chen et al. A container-based DoS attack-resilient control framework for real-time UAV systems. In *2019 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, pages 1222–1227. IEEE, 2019.
[5] K.R. Chevli et al. Blue force tracking network modeling and simulation. In *MILCOM 2006-2006 Military Communications Conference*, pages 1–7, Washington, DC, USA, 2006. IEEE.
[6] J. Daubert, D. Boopalan, M. Mühlhäuser, and E. Vasilomanolakis. HoneyDrone: A medium-interaction unmanned aerial vehicle honeypot. In *NOMS: Network Operations and Management Symposium*, pages 1–6, Taiwan, China, 2018. IEEE.
[7] M. Hooper et al. Securing commercial wifi-based UAVs from common security attacks. In *MILCOM 2016-2016 IEEE Military Communications Conference*, pages 1213–1218, Baltimore, MD, USA, 2016. IEEE.
[8] Y. Huang and Q. Zhu. Deceptive reinforcement learning under adversarial manipulations on cost signals. In *Decision and Game Theory for Security: 10th International Conference, GameSec 2019, Stockholm, Sweden, October 30–November 1, 2019, Proceedings 10*, pages 217–237. Springer, 2019.
[9] G.-H. Kim et al. Multi-drone control and network self-recovery for flying ad hoc networks. In *2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN)*, pages 148–150, Vienna, Austria, 2016. IEEE.
[10] H. Li, Y. Guo, S. Huo, H. Hu, and P. Sun. Defensive deception framework against reconnaissance attacks in the cloud with deep reinforcement learning. *Science China Information Sciences*, 65(7):170305, 2022.
[11] F. O Olowononi et al. Deep reinforcement learning for deception in IRS-assisted UAV communications. In *MILCOM 2022-2022 IEEE Military Communications Conference (MILCOM)*, pages 763–768. IEEE, 2022.
[12] S. Ouiazzane, M. Addou, and F. Barramou. A multiagent and machine learning based denial of service intrusion detection system for drone networks. *Geospatial Intelligence: Applications and Future Trends*, pages 51–65, 2022.
[13] J. Panerati et al. Learning to fly—a gym environment with pybullet physics for reinforcement learning of multi-agent quadcopter control. *2021 International Conference on Intelligent Robots and Systems (IROS)*, 0:7512–7519, 2021.
[14] K. Rahman et al. Detection of security attacks using intrusion detection system for UAV networks: A survey. In *Big Data Analytics and Computational Intelligence for Cybersecurity*, pages 109–123. Springer, 2022.
[15] T. Schaul, J. Quan, I. Antonoglou, and D. Silver. Prioritized experience replay. *arXiv preprint arXiv:1511.05952*, 2015.
[16] H. Sedjelmaci, S. M. Senouci, and N. Ansari. A hierarchical detection and response system to enhance security against lethal cyberattacks in UAV networks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 48(9):1594–1606, 2017.
[17] H. Sedjelmaci, S. M Senouci, and M.-A. Messous. How to detect cyber-attacks in unmanned aerial vehicles network? In *2016 Global Communications Conference (GLOBECOM)*, pages 1–6. IEEE, 2016.
[18] Special Interest Group (SIG). Common Vulnerability Scoring System (CVSS), 2021. Accessed: 11-19-2021.
[19] Z. Wan et al. Foureye: Defensive deception against advanced persistent threats via hypergame theory. *IEEE Transactions on Network and Service Management*, 19(1):112–129, 2021.
[20] Z. Wan et al. Resisting multiple advanced persistent threats via hypergame-theoretic defensive deception. *IEEE Transactions on Network and Service Management*, 2023.
[21] K. Weiss, T. M Khoshgoftaar, and D. Wang. A survey of transfer learning. *Journal of Big data*, 3(1):1–40, 2016.
[22] L. Xiao et al. Attacker-centric view of a detection game against advanced persistent threats. *IEEE Transactions on Mobile Computing*, 17(11):2512–2523, 2018.
[23] M Zhu et al. A survey of defensive deception: Approaches using game theory and machine learning. *IEEE Communications Surveys & Tutorials*, 23(4):2460–2493, 2021.