eMTD: Energy-Aware Moving Target Defense for Sustainable Solar-powered Sensor-based Smart Farms

Dian Chen* , Ing-Ray Chen* , Dong Sam Ha† , and Jin-Hee Cho*
*Department of Computer Science, Virginia Tech, Falls Church, VA, USA
†Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, USA

***Abstract*—Smart farms, as a way for better productivity and efficiency, have yet to be thoroughly studied for their service quality amid cyber threats. This work introduces a proactive security approach, Moving Target Defense (MTD), to the smart farm system to proactively address diverse cyber threats. Specifically, we develop an energy-aware MTD, termed `eMTD`, using port hopping for a sustainable smart farm network. Leveraging deep reinforcement learning (DRL), we identify the optimal MTD strategy capable of ensuring high monitoring quality of animal conditions and sufficient energy levels for solar-powered sensors on the farm. Our experiments demonstrate a significant improvement by approximately 15% in monitoring quality and remaining energy compared to other schemes.**

***Index Terms*—Smart farm, moving target defense, port hopping, deep reinforcement learning, energy-aware.**

I. Introduction

As smart farming systems have been adopted extensively in agriculture industries, the efficiency, security, and cost of such smart systems have become more critical. Solar-powered sensor-based animal monitoring systems offer an energy-efficient solution for cost reduction on farms. Leveraging sensors, IoT, edge computing, and cloud technologies [1], the monitoring systems play a pivotal role. However, current smart farm research overlooks attack-resilient, energy-adaptive systems vital for sustainability, especially in energy-constrained or fluctuating environments prone to cyberattacks.

Given the limited energy from solar cells, maintaining sensor functionality during cyber attacks is crucial. We aim to develop an energy-aware moving target defense (MTD) strategy based on port hopping. This approach maximizes monitoring quality while ensuring adequate energy levels for sensor tasks. The proposed MTD aims to determine the optimal time to adjust service ports in smart farm networks.

This work makes the following **key contributions**:

- We are the first to design an energy-aware MTD for solar-powered sensor-based smart farm systems to ensure system performance despite cyberattacks and energy fluctuations.
- We utilize deep reinforcement learning (DRL) algorithms [20, 25] to determine the optimal defense strategy for solar-powered sensors. We aim to maximize monitoring quality while preserving energy levels. We compare our DRL-based MTD with other rule-based MTD and non-MTD solutions to highlight its effectiveness.
- We rigorously validate the robustness and efficacy of our proposed DRL agents through comprehensive experimentation with real datasets [3]. Our results demonstrate that the proposed MTD outperforms other counterparts by about 15% in monitoring quality and energy sustainability.

II. Related Work

Smart farms boost agricultural productivity by monitoring animal conditions and the environment, leveraging Internet-of-Things (IoT) and edge cloud computing. Yet, increased connectivity poses risks like Denial-of-Service (DoS) and data transit attacks [23, 32]. Ferrag et al. [9] outlined cyber threats in IoT-based agriculture and proposed solutions leveraging blockchain technology. Gayathri et al. [10] introduced access control rules and MTD techniques within the Amazon Web Services framework to counter IoT attacks in smart environments. El-Ghamry et al. [6] introduced a convolutional neural network (CNN)-based detection system to mitigate the associated threat. Furthermore, Eldosouky et al. [7] mathematically analyzed and mitigated the impact of GPS spoofing attacks on unmanned aerial vehicles (UAVs). Seo et al. [26] presented the drone-based defensive deception game framework to reduce the potential attack surface and security vulnerabilities of drone systems. Woo et al. [30] introduced Controller Area Network (CAN) ID shuffling to mitigate security vulnerabilities in vehicular systems.

Cyber defense in network operations and management systems (NOMS) is indispensable to safeguard against sophisticated cyber threats and vulnerabilities while managing network infrastructure. Celdrán et al. [2] developed an IoT-focused framework, employing behavioral fingerprinting to identify and categorize preliminary malicious stages of cryptojackers targeting single-board computers. Similarly, Hajizadeh et al. [12] introduced the Flow-based Self-Active Intrusion Detection System (FSA-IDS) integrating active learning (AL) into self-learning to minimize labeling costs and enhance IDS effectiveness using real-world network traffic datasets. Lübben and Pahl [19] proposed a DNN-based anomaly detection model improving the detection performance while reducing the computational cost. Kapetanidou et al. [15] explored the suitability of two security methods, specifically the cryptography-based approach, and a more lightweight reputation-based alternative, within ad hoc information-centric networks.

Port hopping, a common technique in MTD, involves dynamically associating a service's port with an unallocated pseudo-random port. This strategy is intended to confuse potential attackers. Shi et al. [27] demonstrated that port hopping effectively maintains system operation even under high rates of DoS attack traffic. Hari and Dohi [13] showed that port hopping can improve communication success rates amidst various DoS attack patterns. Lee and Thing [16] observed a significant benefit of port hopping in traffic reception during DoS attacks. Additionally, Fan et al. [8] proposed an end-hopping scheme for IoT, employing fixed hopping timeslots and a robust time synchronization strategy based on MTD principles. Giraldo et al. [11] introduced a decentralized MTD with a dual-layered uncertainty approach to enhance microgrid security by replicating essential sensory and control signals. Zhang et al. [31] employed a hidden MTD to detect false-data injection and parameter derivation attacks.

While previous works have contributed to our understanding of defenses in Smart IoT, there is still ample room for additional research and exploration. First, energy-aware defenses have not been considered in [2, 6, 10, 19] for smart farm environments using solar-powered sensors, which are used in our work. Further, no prior port hopping-based MTD has been applied in smart farm environments whose prior concern is to save or maintain energy to ensure system sustainability. Additionally, existing approaches above [7, 8, 11, 12, 15, 26, 30, 31] have not been sufficiently applied and validated in energy-constrained smart environments. They neglected the challenges associated with balancing defense effectiveness and associated costs. In contrast, our approach addresses the dynamic behaviors of animals and energy fluctuations within smart farm environments.

III. Problem Statement

The proposed smart farm system utilizes DRL to identify the optimal defense strategy of port hopping, aiming to achieve high monitoring quality and extend the system's lifetime in the presence of attacks. We formulate this problem as a scalarization-based multi-objective optimization (MOO) function [5] by:

$$\text{maximize} \quad \mathcal{MQ}(s^*) + \mathcal{RE}(s^*) \tag{1}$$

Here $\mathcal{MQ}(s^*)$ is the monitoring quality of animal conditions taking a set of port hopping defense strategies s^*, $\mathcal{RE}(s^*)$ refers to the remaining energy of the entire system (e.g., solar-powered sensors) by performing s^*, with $\mathcal{MQ}(s^*)$ and $\mathcal{RE}(s^*)$ scaled in $[0, 1]$. Here $\mathcal{MQ}(s^*)$ can be calculated by:

$$\mathcal{MQ}(s^*) = \frac{\sum_{t=1}^{T_{\text{cur}}(s^*)} \sum_{i=1}^{X} \sum_{j=1}^{d} \text{mq}(i,j)}{X \times d}, \tag{2}$$

where $T_{\text{cur}}(s^*)$ is the current system's operation time step while taking s^*, X is the number of sensed data, d is the number of attributes for each animal as detailed in Table I, $GT_{i,j}$ is the ith ground truth data for jth attribute, and $x_{i,j}$ is our observed data. The $\text{mq}(i,j)$ term indicates the degree of monitoring quality in a jth attribute compared to the ith ground truth data and returns 1 when $x_{i,j} == GT_{i,j}$; 0 otherwise. On the other hand, $\mathcal{RE}(s^*)$ is formulated by:

$$\begin{aligned} \mathcal{RE}(s^*) &= 1 - \left(\mathcal{E}_S + \mathcal{E}_D + \mathcal{E}_{\text{active}} + \mathcal{E}_{\text{sleep}}\right) \\ &= 1 - \left(\frac{e_S}{E_S} + \frac{e_D(s^*)}{E_S} + \frac{T_u}{E_S}(d_{\text{active}} + d_{\text{sleep}})\right), \end{aligned} \tag{3}$$

where e_S and $e_D(s^*)$ are the energy consumed per data transmission and energy consumed by hopping given s^*, respectively. The d_{active} and d_{sleep} are energy levels consumed per second in active and sleep modes. The E_S term indicates the energy level when a sensor is fully charged.

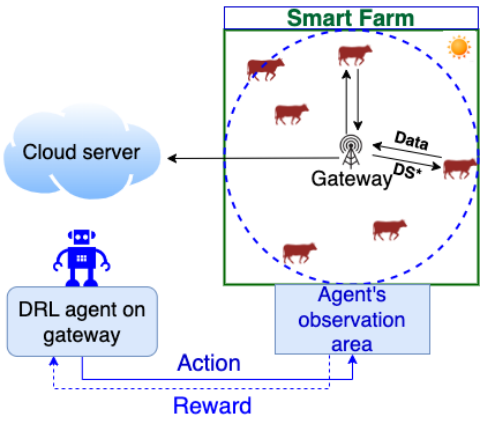


Fig. 1. An example of the considered wireless solar-powered sensor-based smart farm network (DS* refers to a set of port hopping defense strategies.)

TABLE I
EVD Dataset Description

Metric	Description
Serial	A unique animal identifier
HR	Heart Rate of the animal
Average temperature	Average body temperature in Celsius
Min-temperature	Minimum temperature in Celsius
Max-temperature	Maximum temperature in Celsius
Average-activity	Average activity recorded by the number of steps taken
Battery-level	Residual battery life
Timestamp	Date and time of transmission

IV. System Model

Network Model. The network includes solar-powered sensors, a long-range (LoRa) gateway, and a cloud server, as shown in Fig 1. Sensors attached to animals like cows collect and transmit data to the gateway, which aggregates and sends it to the cloud server. The gateway acts as middleware, facilitating sensor-cloud connectivity for cost-effective IoT devices with extended ranges. A deep reinforcement learning (DRL) model on the gateway optimizes defense strategies requested from the sensors to maintain system operations. Due to IoT resource constraints, the system may exacerbate certain cyber threats and pose unique challenges for implementing effective security measures, leaving wireless sensor-to-gateway

communication vulnerable to cyberattacks. Consequently, multiple adversarial attacks may affect data quality and protocol deployment. Refer to the Threaat Model below. Our study examines the robustness of our approach to ensure monitoring quality amidst these threats. For scalability, we can use more gateways covering a larger operation area. Since each gateway is equipped by a DRL agent, a multi-agent DRL approach can be applied to achieve the system goal [17].

Node Model. In the smart farm network, sensors periodically transmit data to the LoRa gateway to monitor animal conditions. Solar-powered sensors experience energy fluctuations due to environmental factors such as animal positions, weather conditions, and sunlight levels. Hence, the system must withstand the dynamic smart farm environment, including energy fluctuations and potential adversarial attacks. We define sensor node i at time t, denoted by sn_t^i, with four attributes:

$$\mathrm{sn}_t^i = [\mathrm{temp}_t^i, \mathrm{hb}_t^i, \mathrm{ma}_t^i, \mathrm{bl}_t^i], \tag{4}$$

where temp_t^i is sensor node i's temprature at time t, hb_t^i is node i's heartbeat at time t, ma_t^i refers to node i's moving activity at time t, and bl_t^i means sensor node i's battery life. The ma_t^i and bl_t^i parameters are scaled in $[0, 100]$ as %.

In the proposed system, the LoRa protocol enables long-distance transmission, spanning 5 to 15 km, with a data rate of 27 $kbps$. Energy consumption varies: the LoRa radio of SAM R34/35 dissipates around 170 mW during transmission. Sensor nodes start with an initial energy level of 5 kWh, with charging efficiency depending on lighting conditions: around 10 mW/cm^2 outdoors and 0.1 mW/cm^2 indoors [29].

Threat Model. We consider cyberattacks on sensor nodes, assuming full trust in the gateway and cloud server.

- **False data injection** [22]: A compromised sensor executes this attack by transmitting falsified or modified data to the gateway.
- **Non-compliance to protocol** [21]: A compromised sensor can execute this attack by employing an undesired action, such as a defense strategy.
- **Send data obstruction** [28]: This is one consequence of de-authentication attacks [24], a significant availability threat in the smart farm environment. This attack disrupts sensor nodes' connection to the network, resulting in a loss of real-time communication with LoRa gateways.

Typically, an outside attacker begins by conducting host probing and port scanning as the initial steps to uncover vulnerabilities within a network [4]. A sensor node in the wireless sensor network will have an open port when it attempts to establish communication with the gateway for data transmission and network coordination. We assume that by default, 30% of sensor nodes are under threat, with attackers attempting to identify vulnerable ports. Once an attacker identifies a vulnerable port on a sensor node, they will compromise the sensor and execute corresponding attacks.

V. Proposed approach: eMTD

Our proposed approach comprises two main components: one focuses on establishing the relationship between port hopping and attack success rate. At the same time, the other involves designing features for our DRL agent, which aims to identify the optimal defense strategy.

A. Modeling Port Hopping

We employ the model abstraction from [18] to establish the correlation between port hopping and the attack success rate. In our scenario, there are two entities: a server responsible for maintaining a set of open ports to provide network services and an attacker targeting the server. The attacker seeks to conduct reconnaissance on the host, while our objective is to conceal the attributes of currently active ports. The attacker succeeds if it identifies an active port, making this reconnaissance process akin to an urn statistic model [14].

An urn problem refers to a class of probability problems involving drawing objects from an urn (a container) without replacement. We can liken our network server host to such a model, containing v black balls and $n-v$ white balls, totaling n balls. Here, the number of balls represents the available service ports, with black balls denoting vulnerable service ports and white balls representing secure ports. Combining this urn model with our environment in the presence of attackers, we simulate the attack process as the attacker draws k balls at each time step. If there is at least one black ball, we consider the attack attempt successful. Upon a successful attack, the corresponding sensor node is compromised and executes the attacks outlined in our Attack model in Section IV.

Based on the above urn model, the attack success rate (ASR) is given by:

$$ASR = P(X = x) = \binom{v}{x} p^x (1-p)^{k-x}. \tag{5}$$

In this context, X denotes a random variable representing the number of black balls within k draws, and $p = \frac{v}{n}$ represents the probability of drawing a black ball, signifying the likelihood of encountering a vulnerable port. The Attack Success Rate (ASR) in our environment is defined as the probability of drawing at least one black ball, formulated as follows:

$$ASR = P(X \geq x) = 1 - P(X = 0) = 1 - (1-p)^k \tag{6}$$

Assuming the attacker can probe $k = n$ times, the following parameters are defined:

- N: Maximum port pool (i.e., 64,512)
- m: Number of probes allowed before one port hopping
- Hopping frequency: Normalized between $[0, 1]$, ranging from no port hopping (e.g., static port: $m = N$) to perfect port hopping (e.g., perfect hopping: $m = 1$)
- $\frac{N}{m}$: Total hopping events over the lifetime of reconnaissance, probing the entire port pool

Hence, ASR can be written as:

$$\begin{aligned} P(X > 0) &= 1 - P(X = 0) \\ &= 1 - P(X_1 = 0)P(X_2 = 0)...P(X_{\frac{N}{m}} = 0) \\ &= 1 - [\frac{\binom{N-v}{m}}{\binom{N}{m}}]^{\frac{N}{m}} \end{aligned} \tag{7}$$

We utilize this equation to quantify ASR, the probability of a sensor node being compromised successfully and subsequently executing attacks in Section IV (Threat Model).

B. DRL-based MTD Strategy Selection

We employ DRL to discern the optimal defense strategy, namely the optimal hopping frequency, to maximize monitoring quality while preserving the remaining energy level within our proposed system. The DRL agent operates on the gateway, adjusting the optimal defense strategy at each time step. The design features of our DRL agent are outlined as follows:

- **State space** $\mathcal{S}_{t-1,t}$: For the DRL agent to determine the optimal action (e.g., defense strategy) based on the current environment, we define the state space as $\mathcal{S}_{t-1,t} = (s_t, a_{t-1})$. Here, s_t represents the state at the current time t, and a_{t-1} denotes the previous action taken by the agent at time $t-1$. Incorporating the previous action into the state enables the model to capture temporal dependencies and learn patterns influenced by the agent's recent history. Moreover, the previous action serves as a relevant contextual cue for the current action, facilitating the agent's adaptation to environmental changes. By encompassing the previous action, the state space effectively replaces the information requirements of both monitoring quality and remaining energy, reducing the state space from two dimensions to one dimension. This reduction significantly decreases the computational complexity during training.
- **Action space** $\mathcal{A}_t$: Once the initial hopping frequency is given, the DRL agent will optimally determine whether to increase or decrease the frequency with a certain value τ or stay the same based on the current system state at each step during operation. Hence we define the action space $\mathcal{A}_t = \{\text{increase, decrease, stay}\}$. A high hopping frequency leads to low ASR resulting in fewer compromised sensor nodes and consequently better monitoring quality of animal status at the cost of depleting the remaining energy level, and vice-versa. Moreover, hopping more frequently will influence the service availability of data transmission, which may increase latency and impact the monitoring quality (i.e. freshness of data). A rule-based approach is lacking in handling the high complexity and dynamics of such systems, and may not handle inherent uncertainty with deterministic rules. Therefore, we deploy DRL to identify the optimal action to achieve both high monitoring quality and the remaining energy level of sensors in the presence of energy fluctuations and cyber attacks.
- **Immediate reward** (r_t): The DRL agent receives immediate reward upon taking action at time t, formulated as $r_t = \mathcal{MQ}(a_t) + \mathcal{RE}(a_t)$ defined by Eqs. 2 and 3, respectively.
- **Accumulated reward** $(\mathcal{R}_t)$: The DRL agent aims to select an action that maximizes the accumulated expected return, expressed as $\mathcal{R}t = \sum t = 0^T \gamma^t r_t$, where r_t denotes the reward at time t, γ represents a discount factor, and T signifies the duration of an episode.

TABLE II
KEY DESIGN PARAMETERS AND DEFAULT VALUES

Notation	Description	Default Value
n	Total number of sensors (cows)	20
P^1_{mv}	Probability of cow i to move	[0.3,0.7]
τ	Adjust step size when an agent takes action	0.1
T_u	Time interval for a sensor to send sensed data	30 s
T_a	Time interval for an agent to select an action	60 s
E_{init}	Initial energy level of sensors	[0.1,0.2]
df_{init}	Initial hopping frequency	[0.3,0.6]
P_A	Percentage of sensor nodes being attacked	0.3
α	Sun expose rate	0.8

VI. EXPERIMENT RESULT

A. Parameterization

We utilize real datasets from the smart farm operated by Virginia Tech's College of Agriculture and Life Sciences, which hosts the Smart Farm Innovation Network (TM). This network serves as a centralized platform for aggregating and analyzing data from various farms across Virginia [3]. These datasets were gathered from multiple devices, including EmbediVet Implantable Temperature Devices (EVD), Halter Sensors, Heart Rate Sensors, and Implantable Temperature Sensors, whose attributes are summarized in Table I. Our proposed system utilizes these datasets and sensor information to simulate and evaluate the monitoring quality of the proposed system. Compromised datasets were generated based on the original data and threat models, resulting in semi-synthetic datasets that inject threats into the real datasets.

The farm under consideration spans 40 acres (approximately 160 km^2), with each side 400 meters long. There are 20 cows on the farm, monitored by a single gateway to ensure comprehensive coverage and efficiency. The entire monitoring simulation extends over 24 hours to demonstrate the efficacy of the DRL agent in discovering optimal policies for the solar-powered system under varying conditions, including both daylight and nighttime scenarios. Additionally, we denote P^i_{mv} as the moving probability for cow i, assuming that cows move randomly with speeds distributed normally, with an average of 1.5 m/s and a standard deviation of 0.1 m/s. The typical ranges of average activity lie within 1 to 2 meters per second. Each sensor node has random initial battery levels within the range of $[0.1, 0.2]$, denoted by E_{init}. The DRL agent is deployed on the gateway and tasked with selecting the optimal action at each time step T_a based on the current system state. The objective is to maximize the monitoring quality as well as the remaining energy of the entire system. Table II provides an overview of the key design parameters, their interpretations, and the default values utilized in our simulations.

B. Metrics

We use the following metrics to validate our approach:

- **Accumulated reward** $(\mathcal{R})$: This metric calculates the sum of immediate rewards over the entire simulation period.
- **Monitoring quality** $(\mathcal{MQ})$: As defined by Expression (2), this metric reflects the accuracy of the monitoring during

system operation, based on the quantity of true sensed data received regarding attributes of each animal.

- **Remaining energy** ($\mathcal{RE}$): This metric quantifies the level of remaining energy in the sensor network during simulations, as defined by Expression (3).
- **Convergence time** ($\mathcal{C}_\mathcal{T}$): This denotes the duration from the initiation of the training process to the attainment of a converged state, estimated by:

$$C_T = T_c - T_b, \tag{8}$$

where T_c represents the convergence time of the model measured in the unit of the number of episodes, and T_b denotes the initiation time of the training process. Specifically, we define T_c by the point when the immediate reward stabilizes within a range of $[-1, +1]$ around its final value for a minimum of 10 consecutive episodes.

C. Schemes for Performance Analysis

We compare the schemes for performance analysis below:

- **Deep Q-Network based MTD (DQN-MTD) [20]**: DRL agents select the best action from the learned Q-table.
- **Proximal Policy Optimization based MTD (PPO-MTD) [25]**: PPO enables the DRL agents to learn an optimal policy by employing an actor-critic algorithm with multiple echoes of stochastic gradient.
- **Random-MTD**: Agents randomly select an action from the action space at each step.
- **Greedy-MTD**: Agents choose an action based on immediate reward. This scheme can also be considered a rule-based/heuristic approach since the agent selects a solution to achieve a local optimal value of Eq. (1).
- **Fixed-MTD**: Agents deploy a fixed hopping frequency (i.e., 0.6) throughout the entire simulation. The frequency is determined based on empirical evidence that results in the best performance relative to alternative options.
- **Static-HF**: The system does not change the port number while operating.

DQN-MTD and PPO-MTD denote our proposed schemes, whereas Random-MTD, Greedy-MTD, Fixed-MTD, and Static-MTD function as baseline schemes in the comparative performance analysis.

D. Comparative Performance Analysis

Fig. 2 illustrates the learning process of six schemes outlined in Section VI-B, with $P_A = 0.3$ as default. Notably, Static-MTD, Fixed-MTD, Random-MTD, and Greedy-MTD do not engage in the learning process. Consequently, their training curve remains a horizontal line, indicating no measure of the convergence time.

The results demonstrate that our proposed PPO-based schemes outperform others in terms of accumulated rewards ($\mathcal{R}$) (Fig. 2(a)) and monitoring quality ($\mathcal{MQ}$) (Fig. 2(b)), while showing reversed trends for the remaining energy ($\mathcal{RE}$) (Fig. 2(c)). This discrepancy arises due to the conflicting objectives of the system, where higher monitoring quality may lead to lower remaining energy levels and vice versa, resulting in the adoption of different policies. Moreover, we observe that the improvement in monitoring quality achieved by more frequent hopping outweighs the energy sacrificed by hopping. For instance, PPO-MTD, despite maintaining the lowest remaining energy level among all schemes, achieves the highest accumulated rewards during learning. Additionally, the results underscore the characteristics of the PPO algorithm, which consistently updates the policy with a small step size, minimizing the probability of overlooking the optimal state and being trapped in a non-optimal state. Figs. 2(a) and 2(d) show that while DQN-MTD exhibits a lower convergence time than PPO-MTD, it converges with a lower reward.

E. Sensitivity Analyses

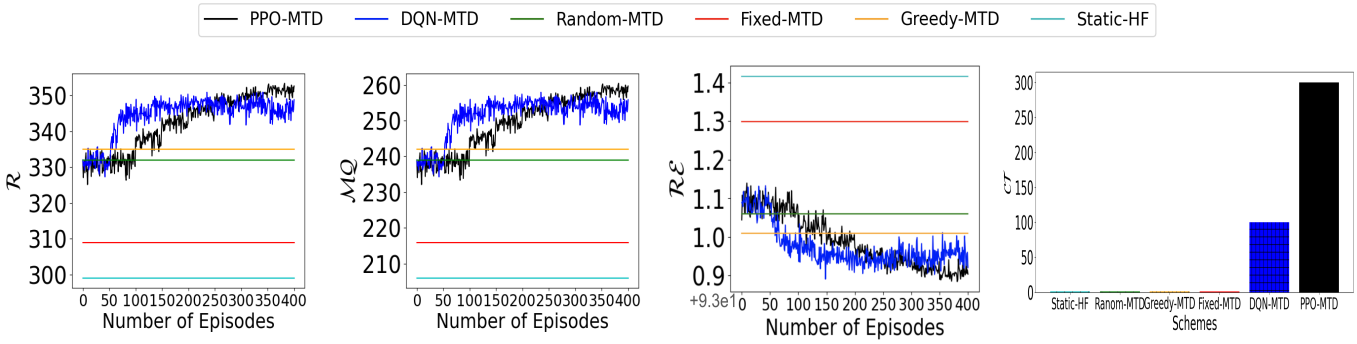
1) **Effect of Attack Severity (P_A):** Fig. 3 shows the effect of attack severity (P_A) on performance. As P_A increases, the monitoring quality ($\mathcal{MQ}$) diminishes due to the increased presence of compromised data. Furthermore, elevated P_A yields greater remaining energy ($\mathcal{RE}$) across all schemes, as more frequent hopping becomes undesirable, given the diminishing efficacy of hopping. Consequently, less energy is expended while maintaining the effectiveness of the defense strategy (i.e., MTD). The PPO-based DRL scheme consistently outperforms DQN and baseline schemes, a result that aligns well with the findings of the comparative analysis in Section VI-D. Further, the PPO-based DRL scheme demonstrates accelerated learning as P_A increases (which increases the occurrence of attack events in the system), because there are more samples for DRL to learn from at each time step.

2) **Effect of Initial Sensor Energy Level (E_{init}):** Fig. 4 shows the effect of the initial sensor energy level (E_{init}) on performance. As E_{init} increases, the accumulated reward ($\mathcal{R}$) also increases, indicating that more energy is available for data transmission and defense. However, this trend is not observed for monitoring quality ($\mathcal{MQ}$) when the initial energy level continues to rise, such as in the ranges of $[0.15, 0.25)$ and $[0.2, 0.3)$. In such cases, if the energy level is already sufficient to achieve optimal monitoring quality, further increases in the initial energy level do not enhance monitoring quality. A higher initial sensor energy level leads to a higher remaining energy level after convergence. On the other hand, the convergence times for both DRL-based schemes under different E_{init} remain identical, indicating that the efficiency of learning is unaffected by varying the initial sensor energy level.

VII. Conclusion & Future work

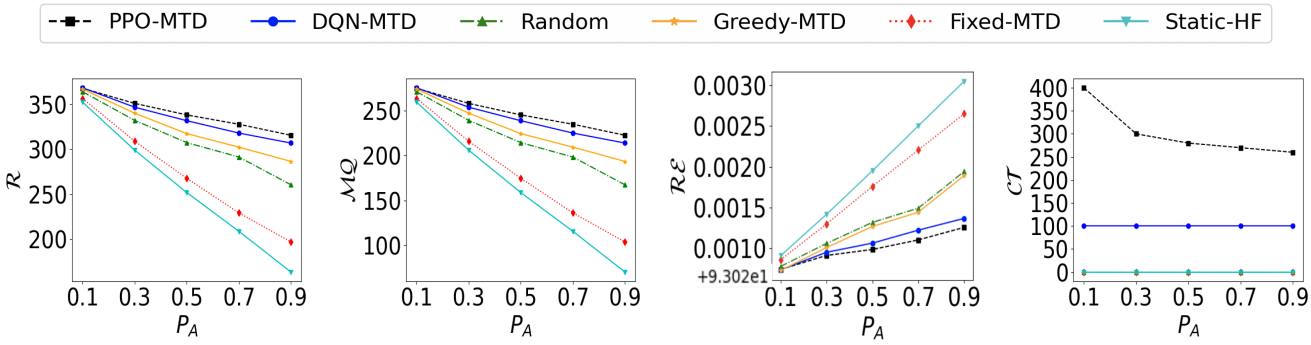
In this work, we introduced an energy-efficient DRL-based defense strategy for moving target defense, namely `eMTD`, in the smart farm environment. Our experimental results showcase the superiority of our proposed scheme over baseline methods concerning system performance. The **key findings** obtained from the experiment outcomes are as follows:

- PPO demonstrates superior performance over DQN in accumulated reward ($\mathcal{R}$) and monitoring quality ($\mathcal{MQ}$) without compromising too much on energy sustainability as the



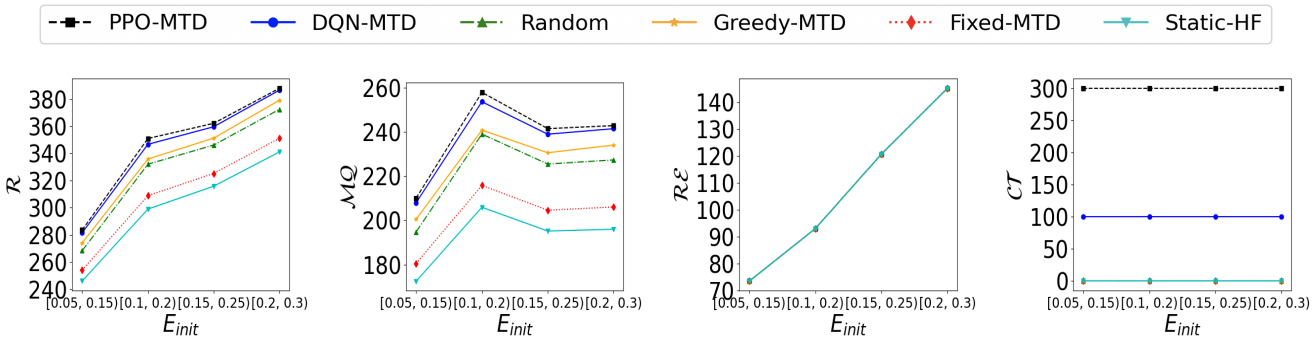
(a) Accumulated Reward ($\mathcal{R}$) (b) Monitoring Quality ($\mathcal{MQ}$) (c) Remaining Energy ($\mathcal{RE}$) (d) Convergence Time ($\mathcal{CT}$)

Fig. 2. Comparative performance analysis during training time



(a) Accumulated Reward ($\mathcal{R}$) (b) Monitoring Quality ($\mathcal{MQ}$) (c) Remaining Energy ($\mathcal{RE}$) (d) Convergence Time ($\mathcal{CT}$)

Fig. 3. Effect of varying the attack probability (P_A)



(a) Accumulated Reward ($\mathcal{R}$) (b) Monitoring Quality ($\mathcal{MQ}$) (c) Remaining Energy ($\mathcal{RE}$) (d) Convergence Time ($\mathcal{CT}$)

Fig. 4. Effect of varying the initial sensor energy level (E_{init})

remaining sensor energy level after convergence for both schemes is approximately the same.

- Despite DQN's tendency to converge faster and require less training time compared to PPO, it converges to a local optimal state, resulting in lower rewards relative to PPO.
- The increase in monitoring quality resulting from uncompromised data reception at sensor nodes significantly outweighs the additional energy consumption cost associated with data transmission due to port hopping.
- The port hopping frequency influences both ASR and data freshness (induced by delay), contributing to fluctuations in monitoring quality.

Moving forward, we will conduct comprehensive experiments to perform sensitivity analyses with additional parameters, including sun exposure and other environmental conditions. In addition, we plan to scale up our proposed system, considering the integration of additional gateways with a multi-agent DRL approach and extending the duration of the simulations. In this approach, DRL agents can collaborate to control sensor nodes within their respective transmission ranges, thus enhancing the system's scalability and adaptability.

References

[1] R. Bogue, "Solar-powered sensors: A review of products and applications," *Sensor Review*, 2012.

[2] A. H. Celdrán, J. v. d. Assen, K. Moser, P. M. S. Sánchez, G. Bovet, G. M. Pérez, and B. Stiller, "Early detection of cryptojacker malicious behaviors on iot crowdsensing devices," in *NOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium*, 2023, pp. 1–8.

[3] Center for Advanced Innovation in Agriculture (CAIA). (2023) Virginia tech smartfarm innovation network (TM). [Online]. Available: https://caia.cals.vt.edu/caia-s-research-platforms/vtsmartfarm.html

[4] C.-M. Chen, S.-C. Hsu, and G.-H. Lai, "Defense denial-of service attacks on ipv6 wireless sensor networks," in *Genetic and Evolutionary Computing: Proceedings of the Ninth International Conference on Genetic and Evolutionary Computing, August 26-28, 2015, Yangon, Myanmar-Volume 1*. Springer, 2016, pp. 319–326.

[5] J.-H. Cho, Y. Wang, R. Chen, K. S. Chan, and A. Swami, "A survey on modeling and optimizing multi-objective systems," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1867–1901, 2017.

[6] A. El-Ghamry, A. Darwish, and A. E. Hassanien, "An optimized cnn-based intrusion detection system for reducing risks in smart farming," *Internet of Things*, vol. 22, p. 100709, 2023.

[7] A. Eldosouky, A. Ferdowsi, and W. Saad, "Drones in distress: A game-theoretic countermeasure for protecting uavs against gps spoofing," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 2840–2854, 2020.

[8] Y. Fan, G. Wu, K.-C. Li, and A. Castiglione, "Robust end hopping for secure satellite communication in moving target defense," *IEEE Internet of Things Journal*, vol. 9, no. 18, pp. 16 908–16 916, 2022.

[9] M. A. Ferrag, L. Shu, X. Yang, A. Derhab, and L. Maglaras, "Security and privacy for green iot-based agriculture: Review, blockchain solutions, and challenges," *IEEE Access*, vol. 8, pp. 32 031–32 053, 2020.

[10] R. Gayathri, S. Usharani, M. Mahdal, R. Vezhavendhan, R. Vincent, M. Rajesh, and M. Elangovan, "Detection and mitigation of IoT-based attacks using snmp and moving target defense techniques," *Sensors*, vol. 23, no. 3, 2023.

[11] J. Giraldo, M. E. Hariri, and M. Parvania, "Decentralized moving target defense for microgrid protection against false-data injection attacks," *IEEE Transactions on Smart Grid*, vol. 13, no. 5, pp. 3700–3710, 2022.

[12] M. Hajizadeh, S. Barua, and P. Golchin, "Fsa-ids: A flow-based self-active intrusion detection system," in *NOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium*, 2023, pp. 1–9.

[13] K. Hari and T. Dohi, "Sensitivity analysis of random port hopping," in *2010 7th International Conference on Ubiquitous Intelligence & Computing and 7th International Conference on Autonomic & Trusted Computing*, 2010, pp. 316–321.

[14] N. Johnson and S. Kotz, *Urn Models and Their Application: An Approach to Modern Discrete Probability Theory*, ser. Approach to Modern Discrete Probability Theory. Wiley, 1977. [Online]. Available: https://books.google.com/books?id=ZBfvAAAAMAAJ

[15] I. A. Kapetanidou, P. Mendes, and V. Tsaoussidis, "Enhancing security in information-centric ad hoc networks," in *NOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium*, 2023, pp. 1–9.

[16] H. Lee and V. Thing, "Port hopping for resilient networks," in *IEEE 60th Vehicular Technology Conference, 2004. VTC2004-Fall. 2004*, vol. 5, 2004, pp. 3291–3295 Vol. 5.

[17] R. Lowe, Y. I. Wu, A. Tamar, J. Harb, O. Pieter Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," *Advances in neural information processing systems*, vol. 30, 2017.

[18] Y.-B. Luo, B.-S. Wang, and G.-L. Cai, "Effectiveness of port hopping as a moving target defense," in *2014 7th International Conference on Security Technology*, 2014, pp. 7–10.

[19] C. Lübben and M.-O. Pahl, "Distributed device-specific anomaly detection using deep feed-forward neural networks," in *NOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium*, 2023, pp. 1–9.

[20] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[21] J. Ophoff, K. Renaud, and e. Bui, Tung X., "Revealing the cyber security non-compliance "attribution gulf"," in *Proceedings of the 54th Annual Hawaii International Conference on System Sciences, HICSS 2021*, ser. Proceedings of the Annual Hawaii International Conference on System Sciences. USA: University of Hawaii at Manoa, 2021, pp. 4557–4566.

[22] M. A. Rahman and H. Mohsenian-Rad, "False data injection attacks with incomplete information against smart power grids," in *2012 IEEE Global Communications Conference (GLOBECOM)*, 2012, pp. 3153–3158.

[23] A. Rettore de Araujo Zanella, E. da Silva, and L. C. Pessoa Albini, "Security challenges to smart agriculture: Current state, key issues, and future directions," *Array*, vol. 8, p. 100048, 2020.

[24] L. P. Rondon, L. Babun, A. Aris, K. Akkaya, and A. S. Uluagac, "Survey on enterprise internet-of-things systems (e-iot): A security perspective," *Ad Hoc Networks*, vol. 125, p. 102728, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1570870521002171

[25] J. Schulman, et al., "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[26] S. Seo, H. Moon, S. Lee, D. Kim, J. Lee, B. Kim, W. Lee, and D. Kim, "D3gf: A study on optimal defense perfor-

mance evaluation of drone-type moving target defense through game theory," *IEEE Access*, vol. 11, pp. 59 575–59 598, 2023.

[27] L. Shi, C. Jia, S. Lü, and Z. Liu, "Port and address hopping for active cyber-defense," in *Intelligence and Security Informatics*, C. C. Yang, D. Zeng, M. Chau, K. Chang, Q. Yang, X. Cheng, J. Wang, F.-Y. Wang, and H. Chen, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 295–300.

[28] S. Sontowski, M. Gupta, S. S. Laya Chukkapalli, M. Abdelsalam, S. Mittal, A. Joshi, and R. Sandhu, "Cyber attacks on smart farming infrastructure," in *2020 IEEE 6th International Conference on Collaboration and Internet Computing (CIC)*, 2020, pp. 135–143.

[29] *CC2640R2F SimpleLink™ Bluetooth® 5.1 Low Energy Wireless MCU*, Texas Instruments, 2016, rev. C. [Online]. Available: https://www.ti.com/product/CC2640R2F

[30] S. Woo, D. Moon, T.-Y. Youn, Y. Lee, and Y. Kim, "Can id shuffling technique (cist): Moving target defense strategy for protecting in-vehicle can," *IEEE Access*, vol. 7, pp. 15 521–15 536, 2019.

[31] Z. Zhang, R. Deng, D. K. Y. Yau, P. Cheng, and J. Chen, "On hiddenness of moving target defense against false data injection attacks on power grid," *ACM Trans. Cyber-Phys. Syst.*, vol. 4, no. 3, mar 2020. [Online]. Available: https://doi.org/10.1145/3372751

[32] K. Zhao and L. Ge, "A survey on the internet of things security," pp. 663–667, 2013.