Energy-aware moving target defense strategy based on port hopping for smart farms: Deep reinforcement learning approach

Dian Chen
Computer Science
Virginia Tech
Falls Church, US
dianc@vt.edu

Ming-Ju Kuo
Computer Science
Virginia Tech
Falls Church, US
mingkuo@vt.edu

Zhuochen Ying
Computer Engineering
Virginia Tech
Blacksburg, US
yingzhuochen@vt.edu

***Abstract*—The smart farm system contributed to the productivity and efficiency of farms. However, these systems suffered cybersecurity vulnerabilities and energy constraints. As there was no research studying both the defense capability and energy constraints of the smart farm systems, we developed a moving target defense (MTD) based energy-aware defense strategy, which focuses on the trade-off between monitoring quality and energy consumption speed of the system. We defined our task as an optimization problem and used deep reinforcement learning algorithms PPO and DQN to find the optimal solution. We conducted comprehensive experiments on the application of defense strategy in the smart farm environment and measured the performance of our strategies. We compared our optimal strategies with random/greedy strategies and showed advancements in our strategies. In the end, we concluded our work and summarized our contributions.**

***Index Terms*—smart farm, moving target defense, port hopping, deep reinforcement learning, energy awareness**

I. Introduction

Smart farming, also known as digital agriculture, is a newly developed agricultural form in recent years. It adopts new information and communication technologies on the farm, including the Internet of Things (IoT), Machine Learning (ML), and Computer Vision (CV). With real-time environment awareness, wireless data transmission, and integrated computing, the smart farm network is able to monitor the farm environment and make good choices. The application of a digital agricultural system in the farm contributes to its productivity and security, which eases the food shortage caused by the exponential increase in population worldwide.

As smart farming systems are adopted into more farms, the efficiency, security, and cost of these systems should be considered seriously. The real-world smart farming network is composed of sensors and central gateways, with wireless data communication between them. The framework of the smart farm shows the potential cybersecurity vulnerabilities of the system, as attackers are affected to attack both the sensors and gateways, as well as the data transmission process, by destroying, polluting, or blocking the data in these nodes. Therefore, the smart farm should be able to perform defense against hostile attacks. Besides, as most of the sensors in the farm work on limited energy and monitoring the farm needs to consumes most of the energy, the smart farm should trade off the defense effectiveness and energy consumption of defense actions, to ensure both the monitoring quality and security of the smart farm system.

We conducted a comprehensive literature review on the topics of energy-aware defense strategy applications on smart farms, IoT, and networks. We found that in the area of smart farm, it lacks research that considers the defense capability in an energy-constrained environment. Besides, we didn't find prior work that proposes the trade-off between the defense capability and energy cost. Additionally, when we did research on the defense strategies in the smart farm environment, we found that the port hopping strategy for moving target defense was not studied.

We developed an energy-aware defense strategy for moving target defense (MTD), which maximized the monitoring quality of the system while also maintaining a sufficient energy level for the tasks of sensor nodes. Our approach focused on the "when-to-move" principle of MTD, to suggest the moment that the port hopping MTD strategy should be adjusted in the smart farm environment. The following is our **key contributions** of this project:

- We developed an intelligent, energy-aware defense strategy for moving target defense to maintain the system's performance in the presence of cyber-attacks and energy fluctuations.
- We leveraged two deep reinforcement learning algorithms (PPO & DQN) to solve our multi-objective problem in the designed environment.
- We conducted comprehensive experiments to show both the effectiveness and efficiency of our proposed method.
- We demonstrated the insights from our proposed approach and experiment results.

For the remaining sections of our paper, Section II summarizes the work of our comprehensive literature review. Section III demonstrates the optimal problem that we need to solve during the experiment. Section IV illustrates the network and attack model considered in the experiment. Section V introduces our approach applied in the experiment. Section VI enumerates the experimental results as well as our findings and discussions. Section VII concisely concludes our work in this paper and lists potential future work.

II. Literature review

A. Defense strategy in Smart environment

Smart farms have the potential to enhance agricultural productivity through the monitoring of both animal conditions and the environment, employing technologies such as the IoT and edge cloud computing. The transmission of data within the smart farm is facilitated by the utilization of the Long Range (LoRa) protocol and Bluetooth Low Energy (BLE). However, the increased connectivity also introduces the risk of cyber attacks, including Denial-of-Service (DoS) and data transit attacks[14][21]. In their academic work, Ferrag *et al.* delineate cyber threats within the context of IoT-based agriculture and present corresponding solutions utilizing blockchain technology[6]. Gayathri *et al.* introduced access control rules (ACL) and moving target defense (MTD) techniques within the Amazon Web Services (AWS) framework to address IoT attacks within smart environments. In the context of IoT devices, prevalent threats include Denial of Service (DoS) attacks and false data injection. To mitigate these threats, the authors proposed the utilization of a network management protocol (SNMP) and kernel learning detection (KLD) methods for effective attack detection[7]. El-Ghamry *et al.* have presented a convolutional neural network (CNN)-based detection system as a measure to mitigate the associated threat[3].

B. Defense strategy in Internet-of-Things

IoT technology finds application in various domains, including but not limited to drones, automotive vehicles, satellites, and smart farms. Given that the smart farm operates within the IoT domain, numerous defensive strategies are applicable to the system. Fan *et al.* proposed an end-hopping scheme that relies on a fixed hopping timeslot and a rigorous time synchronization strategy, utilizing MTD principles[5]. Eldosouky *et al.* proposed a mathematical framework designed for the analysis and mitigation of the impact of GPS spoofing attacks on unmanned aerial vehicles (UAVs)[4]. Seo *et al.* presented the Drone-Based Defensive Deception Game Framework (D3GF) in an effort to diminish the potential attack surface and security vulnerabilities inherent in drone systems[16]. Woo et al. introduced the concept of Controller Area Network (CAN) ID shuffling as a measure to mitigate security vulnerabilities in vehicular systems[19].

DoS could be the one of major attacks on IoT. Dahiya *et al.* suggested the utilization of Bayesian pricing and auction mechanisms to attain Bayesian Nash Equilibrium points in various scenarios, where the integration of probabilistic information proves advantageous for both legitimate users and service providers[2].

For false-data injection attacks, Giraldo *et al.* have introduced the Decentralized Moving Target Defense via Data Replication (DMTDR) framework, which employs a dual-layered approach of uncertainty to augment the security of microgrids. This strategy aims to reduce the success rate, contributing to the overall robustness of the smart farm system within the IoT ecosystem[8]. They proposed another method to harness the flexibility of IoT networks to strengthen the security of cyber-physical systems (CPSs) through the replication of pertinent sensory and control signals[9]. The utilization of another MTD, a hidden Moving Target Defense, can enhance the defender's ability to identify false-data injection attacks and prevent the attacker from deducing new parameters[20].

C. Energy-efficient port hopping

Port hopping serves as a common MTD mechanism, dynamically associating a service's port with an unallocated pseudo-random port. This approach is designed to introduce confusion among potential attackers. Research conducted by Shi *et al.* has demonstrated that the implementation of port hopping effectively sustains the normal operation of a system even when subjected to a high rate of DoS attack traffic[17]. Hari *et al.* presented in their research that the practice of port hopping has the potential to enhance communication success rates in the presence of diverse DoS attack patterns[10]. The experiment conducted by Lee *et al.* showed a substantial decrease in the reception of desirable traffic when port hopping was not employed. Moreover, the findings indicated that the implementation of port hopping can enhance the reception of desirable traffic in the context of a Denial of Service (DoS) attack[11]. Various factors may impact the Attack Success Rate (ASR), including the size of the port pool, the number of probes, the count of vulnerable services, and the hopping frequency. In the context of a smart farm operating within an energy-constrained environment, it becomes imperative to opt for strategies that consume minimal energy among these influencing factors. The research conducted by Lou et al. serves as an evaluation of the effectiveness of these factors in mitigating ASR[12].

III. Problem statement

To achieve high monitoring quality and extend the system's lifetime in the presence of attacks, the smart farm system leverages deep reinforcement learning (DRL) to identify the optimal defense strategy of port hopping based on the current system state. We formulate this problem as a multi-objective optimization problem, formulated as a scalarization-based multi-objective optimization (MOO) function [1], 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 defense strategies s^*, $\mathcal{RE}(s^*)$ refers to the remaining energy of the entire system (e.g. solar sensors) by performing s^*, with $\mathcal{MQ}(s^*)$ and $\mathcal{RE}(s^*)$ scaled in $[0, 1]$. The proposed system aims to maximize $\mathcal{MQ}(s^*)$ while maintaining $\mathcal{RE}(s^*)$ in an acceptable level. Therefore, the proposed

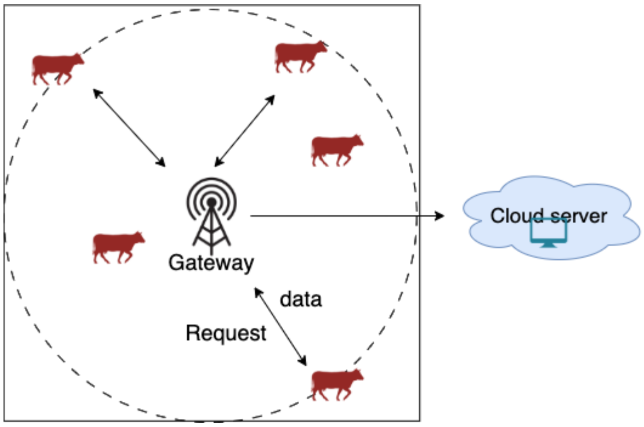


Fig. 1. Wireless Solar Sensor-based Smart Farm Network.

system will aim to build a sustainable smart farm using the DRL approach. This involves effective and efficient designs of defense strategy and design features of DRL under adversarial attacks, as detailed in Section IV-C

IV. System model

A. Network model

The proposed network consists of a group of solar-powered sensors, the long-range (LoRa) gateway, and a cloud server, as described in Fig 1. In the network, each animal (e.g. cow) is attached by a solar-powered sensor which periodically collects the animal conditions and transmits them to the gateway aggregates the received data for each animal, and sends them to the cloud server. In this case, a gateway can be regarded as a middleware connecting sensors and the cloud server, enabling connectivity for IoT devices to be less expensive and have a longer range for transmission. A DRL model will be deployed on the gateway to maintain the monitoring system's normal operation by requesting the sensors to perform optimal defense strategy. We assume that the communication between wireless sensors and LoRa gateways may be vulnerable to cyberattacks as the system does not use data encryption because encryption is not a viable solution under severe resource constraints in IoT environments. Therefore, multiple adversarial attacks (see Section IV-C) may exist in data transmission to influence the quality of sensed data transferred and properly deploy the protocol. Our work investigates the robustness of our proposed approach to ensure monitoring quality under such threats.

B. Node model

In the smart farm network, sensors can periodically transmit sensed data to the LoRa gateway for monitoring animal conditions. Due to the inherent nature of solar-powered sensors, the sensors' energy levels fluctuate throughout the day. They are influenced by environmental factors, such as animals' positions, weather conditions, or the amount of sunshine under different seasons. Therefore, the system must be robust against the dynamic smart farm environment under energy fluctuation and adversarial attacks.

In this network, each sensor transmits its sensed data to nearby clients through Bluetooth Low Energy (BLE) while a client transmits its local update to the LoRa gateways. The communication network deploys the LoRa protocol for long-distance transmission, covering distances typically ranging from 5 to 15 km and achieving a data transfer speed of 27 $kbps$. On the other hand, the BLE protocol is utilized for short-distance communication, spanning up to 100 meters with a transfer speed of 2 $Mbps$. Regarding energy consumption, the LoRa radio of SAM R34/35 dissipates around 170 mW when transmitting data, whereas the BLE radio consumes approximately 11 mW. A Raspberry Pi consumes 0.117 W per sec while idling and 0.172 W per sec with workloads. A fully charged sensor node has an initial energy level of 5 kW. The efficiency of charging the solar-powered sensors depends on the lighting conditions, which is approximately 10 mW/cm^2 under outdoor light and 0.1 mW/cm^2 under indoor light [18].

C. Threat model

In the proposed system, we consider the following cyberattacks being performed on sensor nodes while assuming both the gateway and cloud server are fully trusted.

- **False data injection**: A compromised sensor will perform this attack by sending falsified or modified data to the gateway.
- **Non-compliance to protocol**: This attack can be performed by a compromised sensor that utilizes an undesired action (e.g. defense strategy)
- **Send data obstruction**: The attacker can prevent the sensor from connecting to the internet and the sensor cannot send its collected data to the gateway.

We consider there are 30% of sensor nodes are compromised at the beginning of system operation. The attack success rate of performing each attack will be discussed in Section V-A.

V. Proposed approach

Our proposed approach consists of two main parts: one is how to make the relationship between hopping frequency and attack success rate while the other is the design features of our DRL agent that aims to identify the optimal defense strategy.

A. Model abstraction of port hopping

We leverage the model abstraction from [12] to correlate the hopping frequency to the attack success rate. Assume our scenario has two entities: a server and an attacker. The server is responsible for maintaining a set of ports open to provide network services while the attacker targets the server. The attacker aims to perform a reconnaissance attack on the host while our task is to hide the attributes of currently active ports. The success of the attacker is achieved if it finds an active port, and thus this process of reconnaissance can be regarded as an urn model in statistic theory.

An urn problem refers to a class of probability problems that involve drawing objects from an urn (a container) without replacement. We then can regard our network server host as such a model, which contains v black and $n - v$ white balls for a total of n balls. In this case, the number of balls represents the number of available service ports, with black balls representing vulnerable services' ports while white balls represent secure ports. Combining this urn model and 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 say the process of this attack attempt succeeds. Once the attack succeeds, we say the corresponding sensor node is compromised and will perform attacks we discussed in our threat model.

Based on the above urn model, the attack success rate (ASR) can be formulated as:

$$ASR = P(X = x) = \binom{v}{x} p^x (1-p)^{k-x} \tag{2}$$

where X is a random variable representing the number of black balls within k draws, and $p = \frac{v}{n}$ is the probability of drawing a black ball (e.g. finding a vulnerable port). In our environment, the ASR can be defined as the probability of drawing at least one black ball, formulated by:

$$ASR = P(X \geq x) = 1 - P(X = 0) = 1 - (1-p)^k \tag{3}$$

Luo et al. [12] also demonstrates the relationship between hopping frequency and ASR. Assume the attacker is allowed to probe $k = n$ probes, they define the following parameters:

- N: Maximum port pool (i.e. 64512)
- m: Number of probes allowed before one port hopping
- hopping frequency: Normoized between [0,1] which is from no hopping (e.g. static port: m = N) to perfect port hopping (e.g. perfect hopping: m = 1)
- $\frac{N}{m}$: hopping events in the whole reconnaissance lifetime of probing the entire port pool

Our new ASR then becomes:

$$P(X > 0) = 1 - P(X = 0) \tag{4}$$

$$= 1 - P(X_1 = 0)P(X_2 = 0)...P(X_{\frac{N}{m}} = 0) \tag{5}$$

$$= 1 - [\frac{\binom{N-v}{m}}{\binom{N}{m}}]^{\frac{N}{m}} \tag{6}$$

We will deploy this equation to measure the ASR, which will be used as the probability of a sensor node being compromised.

B. Deep reinforcement learning-based defense strategy

We leverage the deep reinforcement learning (DRL) technique to identify the optimal defense strategy, which is optimal hopping frequency, to maximize monitoring quality while maintaining the remaining energy level of our proposed system. The DRL agent will be deployed on the gateway adjusting the optimal defense strategy at each time step. We propose the design features of our DRL agent as follows:

TABLE I
KEY DESIGN PARAMETERS, THEIR MEANINGS, AND DEFAULT VALUES

Notation	Meaning	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^{LES}_{init}	Initial energy level of low energy sensors	[0.1,0.2]
df_{init}	Initial hopping frequency	[0.3,0.6]
P_A	Percentage of sensor nodes being compromised	0.3
α	Sun expose rate	0.8

- State space: $s^t_i = \{\{\text{re}^t_1\}, \ldots, \{\text{re}^t_i\}\}$. Here, re^t_i represents the remaining energy of sensor i at time t.
- Action space: $\mathcal{A}_t = \{increase, decrease, stay\}$. After the initial hopping frequency is given, the DRL agent will optimally determine whether to increase, or decrease the frequency with a certain value or stay the same based on the current system state at each step during operation.
- Immediate reward: $r_t = \mathcal{MQ}(a_t) + \mathcal{RE}(a_t)$.

VI. EXPERIMENTAL RESULT

A. Parameterization

This study used the sample datasets from the smart farm operated by Virginia Tech's College of Agriculture and Life Sciences, as our home institution, has the SmartFarm Innovation Network (TM), which functions as a centralized platform to aggregate and analyze data from numerous farms throughout Virginia. The normal ranges of average activity are in $[1, 2]$ meters per second. Moreover, cow i will move with the probability P^i_{mv} at a random speed following the normal distribution of their speeds with an average of 1.5 m/s and a standard deviation of 0.1 m/s. Our proposed system utilized these datasets and sensors' information to simulate and evaluate the monitoring quality of the proposed system. The compromised datasets were generated based on the original data and threat models described in Section IV-C. Therefore, our work uses semi-synthetic datasets, injecting threats on top of the real datasets.

The considered farm covers an area of 40 acres (i.e., $\sim 160K$ square meters), with each side measuring 400 meters in length. We consider 20 cows on the farm and one gateway to monitor the entire expanse efficiently and effectively. The entire monitoring simulation spans 24 hours. The DRL agent is deployed on the gateway which selects the optimal action at each time step corresponding to the current system state to maximize monitoring quality and remaining energy of the entire system.

B. Metrics

We use the following metrics to evaluate both the effectiveness and efficiency of our proposed system:

- **Accumulated reward** ($\mathcal{R}$): This calculates the sum of immediate reward over the entire simulation.

- **Monitoring quality** ($\mathcal{MQ}$): This measures the accuracy of received data by the gateway compared with the initial sensed data on sensors.
- **Remaining energy** ($\mathcal{RE}$): This measures the degree of remaining energy in the sensor network per time interval, T_u. $\mathcal{RE}$ is formulated by:

$$\begin{aligned} \mathcal{RE} &= 1 - \Big(\mathcal{E}_S + \mathcal{E}_D + \mathcal{E}_{active} + \mathcal{E}_{sleep}\Big) \\ &= 1 - (\frac{e_S}{E_S} + \frac{e_D}{E_S} + \frac{T_u}{E_S}(d_{active} + d_{sleep})), \end{aligned} \tag{7}$$

where e_S and e_D are the energy consumed per data transmission and energy consumed per hopping, respectively. The d_{active} and d_{sleep} are energy levels consumed per second in active and sleep modes. The E_S indicates the energy level when a sensor is fully charged.
- **Convergence time** ($\mathcal{C_T}$): Time spent from the beginning of the training process to converged state, estimated by:

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

C. Comparing Schemes

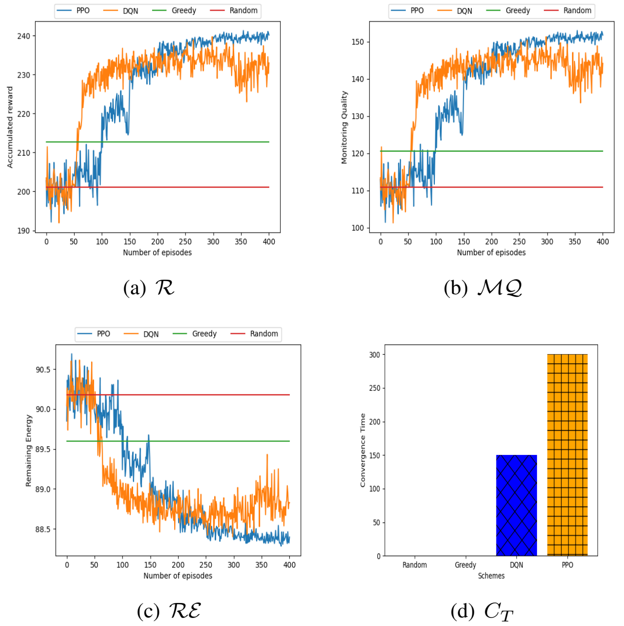
We consider the four schemes below for performance comparison. The DQN and PPO are for our proposed scheme while the other two, Random and Greedy, are baseline schemes for performance analysis.

- **Deep Q-Network (DQN)** [13]: DRL agents select the best action from the learned Q-table.
- **Proximal Policy Optimization (PPO)** [15]: DRL agents select the optimal actions based on learned policy. The PPO uses an actor-critic style algorithm deploying multiple echos of stochastic gradient ascent to update the policy.
- **Random**: Agents will randomly select an action from the action space at each step.
- **Greedy**: Agents will choose an action based on the immediate reward.

D. Comparative performance analysis

The following are **key findings** based on the experimental results:

- PPO outperforms DQN for metrics accumulated reward ($\mathcal{R}$), monitoring quality ($\mathcal{MQ}$), while it is just the opposite in the remaining energy ($\mathcal{RE}$) and convergence time (C_T).
- Though DQN often converges faster and has less training time than PPO, it is possible to converge to a local optimal state, which leads to lower rewards compared to PPO.
- The increase in monitoring quality by receiving data from a sensor node is much larger than the cost of that sensor to send data.
- The hopping frequency influences both ASR and freshness of data (caused by delay) resulting in the variation of monitoring quality.



(a) $\mathcal{R}$

(b) $\mathcal{MQ}$

(c) $\mathcal{RE}$

(d) C_T

Fig. 2. Comparative performance Analysis During Training Time

VII. Conclusion & Future work

In this work, we proposed an energy-efficient DRL-based defense strategy for moving target defense in the smart farm environment. The experimental results demonstrate our proposed scheme outperforms baseline methods in terms of both effectiveness and efficiency. In addition, in two DRL techniques, PPO outperforms DQN with a higher accumulated reward while taking longer to converge. In the future, we will provide more comprehensive experiments to show the sensitivity analysis with various parameters, such as attack severity. Moreover, we consider the larger scale of the proposed system. For example, we will consider more gateways with a multi-agent DRL approach, which DRL agents can learn together controlling sensor nodes in their own transmission range respectively.

References

[1] Jin-Hee Cho, Yating Wang, Ray Chen, Kevin S Chan, and Ananthram Swami. A survey on modeling and optimizing multi-objective systems. *IEEE Communications Surveys & Tutorials*, 19(3):1867–1901, 2017.

[2] Amrita Dahiya and Brij B. Gupta. A reputation score policy and bayesian game theory based incentivized mechanism for ddos attacks mitigation and cyber defense. *Future Generation Computer Systems*, 117:193–204, 2021. ISSN 0167-739X. doi: https://doi.org/10.1016/j.future.2020.11.027.

[3] Amir El-Ghamry, Ashraf Darwish, and Aboul Ella Hassanien. An optimized cnn-based intrusion detection system for reducing risks in smart farming. *Internet of Things*, 22:100709, 2023. ISSN 2542-6605. doi: https://doi.org/10.1016/j.iot.2023.100709.

[4] AbdelRahman Eldosouky, Aidin Ferdowsi, and Walid Saad. Drones in distress: A game-theoretic countermea-

sure for protecting uavs against gps spoofing. *IEEE Internet of Things Journal*, 7(4):2840–2854, 2020. doi: 10.1109/JIOT.2019.2963337.

[5] Yongkai Fan, Guodong Wu, Kuan-Ching Li, and Arcangelo Castiglione. Robust end hopping for secure satellite communication in moving target defense. *IEEE Internet of Things Journal*, 9(18):16908–16916, 2022. doi: 10.1109/JIOT.2022.3144971.

[6] Mohamed Amine Ferrag, Lei Shu, Xing Yang, Abdelouahid Derhab, and Leandros Maglaras. Security and privacy for green iot-based agriculture: Review, blockchain solutions, and challenges. *IEEE Access*, 8:32031–32053, 2020. doi: 10.1109/ACCESS.2020.2973178.

[7] Rajakumaran Gayathri, Shola Usharani, Miroslav Mahdal, Rajasekharan Vezhavendhan, Rajiv Vincent, Murugesan Rajesh, and Muniyandy Elangovan. Detection and mitigation of iot-based attacks using snmp and moving target defense techniques. *Sensors*, 23(3), 2023. ISSN 1424-8220. doi: 10.3390/s23031708.

[8] Jairo Giraldo, Mohamad El Hariri, and Masood Parvania. Decentralized moving target defense for microgrid protection against false-data injection attacks. *IEEE Transactions on Smart Grid*, 13(5):3700–3710, 2022. doi: 10.1109/TSG.2022.3176246.

[9] Jairo A. Giraldo, Mohamad El Hariri, and Masood Parvania. Moving target defense for cyber–physical systems using iot-enabled data replication. *IEEE Internet of Things Journal*, 9(15):13223–13232, 2022. doi: 10.1109/JIOT.2022.3144937.

[10] Kousaburou Hari and Tadashi Dohi. Sensitivity analysis of random port hopping. In *2010 7th International Conference on Ubiquitous Intelligence & Computing and 7th International Conference on Autonomic & Trusted Computing*, pages 316–321, 2010. doi: 10.1109/UIC-ATC.2010.69.

[11] H.C.J. Lee and V.L.L. Thing. Port hopping for resilient networks. In *IEEE 60th Vehicular Technology Conference, 2004. VTC2004-Fall. 2004*, volume 5, pages 3291–3295 Vol. 5, 2004. doi: 10.1109/VETECF.2004.1404672.

[12] Yue-Bin Luo, Bao-Sheng Wang, and Gui-Lin Cai. Effectiveness of port hopping as a moving target defense. In *2014 7th International Conference on Security Technology*, pages 7–10, 2014. doi: 10.1109/SecTech.2014.9.

[13] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.

[14] Angelita Rettore de Araujo Zanella, Eduardo da Silva, and Luiz Carlos Pessoa Albini. Security challenges to smart agriculture: Current state, key issues, and future directions. *Array*, 8:100048, 2020. ISSN 2590-0056. doi: https://doi.org/10.1016/j.array.2020.100048.

[15] John Schulman, et al. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[16] Sang Seo, Heaeun Moon, Sunho Lee, Donghyeon Kim, Jaeyeon Lee, Byeongjin Kim, Woojin Lee, and Dohoon Kim. D3gf: A study on optimal defense performance evaluation of drone-type moving target defense through game theory. *IEEE Access*, 11:59575–59598, 2023. doi: 10.1109/ACCESS.2023.3278744.

[17] Leyi Shi, Chunfu Jia, Shuwang Lü, and Zhenhua Liu. Port and address hopping for active cyber-defense. In Christopher C. Yang, Daniel Zeng, Michael Chau, Kuiyu Chang, Qing Yang, Xueqi Cheng, Jue Wang, Fei-Yue Wang, and Hsinchun Chen, editors, *Intelligence and Security Informatics*, pages 295–300, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg. ISBN 978-3-540-71549-8.

[18] *CC2640R2F SimpleLink™ Bluetooth® 5.1 Low Energy Wireless MCU*. Texas Instruments, 2016. URL `https://www.ti.com/product/CC2640R2F`. Rev. C.

[19] Samuel Woo, Daesung Moon, Taek-Young Youn, Yousik Lee, and Yongeun Kim. Can id shuffling technique (cist): Moving target defense strategy for protecting in-vehicle can. *IEEE Access*, 7:15521–15536, 2019. doi: 10.1109/ACCESS.2019.2892961.

[20] Zhenyong Zhang, Ruilong Deng, David K. Y. Yau, Peng Cheng, and Jiming Chen. On hiddenness of moving target defense against false data injection attacks on power grid. *ACM Trans. Cyber-Phys. Syst.*, 4(3), mar 2020. ISSN 2378-962X. doi: 10.1145/3372751. URL `https://doi.org/10.1145/3372751`.

[21] Kai Zhao and Lina Ge. A survey on the internet of things security. pages 663–667, 2013. doi: 10.1109/CIS.2013.145.