Computers, Materials & Continua

Tech Science Press

Doi:10.32604/cmc.2025.069332

ARTICLE

A Dynamic Deceptive Defense Framework for Zero-Day Attacks in IIoT: Integrating Stackelberg Game and Multi-Agent Distributed Deep Deterministic Policy Gradient

Shigen Shen[1,2], **Xiaojun Ji**[1,*] **and Yimeng Liu**[1]

[1]School of Information Engineering, Huzhou University, Huzhou, 313000, China
[2]Zhejiang Key Laboratory of Industrial Solid Waste Thermal Hydrolysis Technology and Intelligent Equipment, Huzhou University, Huzhou, 313000, China

*Corresponding Author: Xiaojun Ji. Email: 2023388416@stu.zjhu.edu.cn

Received: 20 June 2025; Accepted: 14 August 2025; Published: 23 September 2025

ABSTRACT: The Industrial Internet of Things (IIoT) is increasingly vulnerable to sophisticated cyber threats, particularly zero-day attacks that exploit unknown vulnerabilities and evade traditional security measures. To address this critical challenge, this paper proposes a dynamic defense framework named Zero-day-aware Stackelberg Game-based Multi-Agent Distributed Deep Deterministic Policy Gradient (ZSG-MAD3PG). The framework integrates Stackelberg game modeling with the Multi-Agent Distributed Deep Deterministic Policy Gradient (MAD3PG) algorithm and incorporates defensive deception (DD) strategies to achieve adaptive and efficient protection. While conventional methods typically incur considerable resource overhead and exhibit higher latency due to static or rigid defensive mechanisms, the proposed ZSG-MAD3PG framework mitigates these limitations through multi-stage game modeling and adaptive learning, enabling more efficient resource utilization and faster response times. The Stackelberg-based architecture allows defenders to dynamically optimize packet sampling strategies, while attackers adjust their tactics to reach rapid equilibrium. Furthermore, dynamic deception techniques reduce the time required for the concealment of attacks and the overall system burden. A lightweight behavioral fingerprinting detection mechanism further enhances real-time zero-day attack identification within industrial device clusters. ZSG-MAD3PG demonstrates higher true positive rates (TPR) and lower false alarm rates (FAR) compared to existing methods, while also achieving improved latency, resource efficiency, and stealth adaptability in IIoT zero-day defense scenarios.

KEYWORDS: Industrial internet of things; zero-day attacks; Stackelberg game; distributed deep deterministic policy gradient; defensive spoofing; dynamic defense

1 Introduction

Securing data in Industrial Internet of Things (IIoT) control systems is becoming increasingly challenging due to the growing threat of zero-day attacks [1–3]. These attacks capitalize on undisclosed vulnerabilities, fundamentally compromising conventional security infrastructures—including static intrusion detection mechanisms and signature-dependent perimeter defenses—which prove particularly ineffective within the dynamic and resource-constrained operational matrices of IIoT deployments [4,5]. Such legacy approaches exhibit inherent rigidity that prevents adaptation to the polymorphic and evasive characteristics of zero-day incursions, culminating in disproportionate resource expenditures and inadequate protection postures. Although hybrid deep learning-based Intrusion Detection Systems (IDSs) have achieved high accuracy on

known attacks, they still lack adaptability against intelligent and evolving threats like zero-day and Advanced Persistent Threat (APT) attacks [6]. As a result, advanced adaptive defense strategies, especially those leveraging game-theoretic models, Defensive Deception (DD), and Multi-Agent Reinforcement Learning (MARL), have emerged as promising solutions for enhancing resilience and securing IIoT control systems [7]. The Stackelberg game provides a hierarchical decision-making framework that effectively models the dynamic interactions between defenders and attackers in IIoT control systems [8,9].

The Stackelberg game provides a hierarchical decision-making framework that effectively models the dynamic interactions between defenders and attackers in IIoT control systems [8,9]. Unlike traditional static defenses, Stackelberg games explicitly consider leader-follower roles, where defenders (leaders) proactively adapt their defense strategies while anticipating the attackers' (followers) responses [10,11]. This leader-follower interaction enables defenders to strategically allocate limited resources and optimize packet sampling policies to effectively contain zero-day attacks. In IIoT environments, such hierarchical modeling allows defenders to dynamically adjust detection and mitigation strategies, achieving a fast-converging attack-defense equilibrium even when attackers modify their tactics. Consequently, the Stackelberg game framework addresses the shortcomings of conventional defenses, offering a promising solution to tackle the unpredictability and stealthiness of zero-day attacks in IIoT systems [12].

However, solely relying on Stackelberg game models has limitations in capturing the nonlinear and high-dimensional interactions inherent in IIoT control systems, particularly under complex and dynamic zero-day attack scenarios. Compared with existing Internet of Things (IoT) security solutions that predominantly rely on static rules, signature-based detection, or purely reactive defense mechanisms, the proposed framework specifically addresses the stealthiness, adaptability, and uncertainty of zero-day attacks. Conventional methods often fail to promptly recognize and mitigate previously unknown vulnerabilities, leading to delayed responses and increased system risk. In contrast, our framework introduces a hierarchical game-theoretic approach combined with continuous learning via reinforcement learning, enabling proactive adaptation to evolving threats in real time. Furthermore, the integration of dynamic deception and behavioral fingerprinting empowers the system to both confuse attackers and accurately identify abnormal behaviors at an early stage, effectively closing the gap left by traditional defenses. To address these challenges, we propose the Zero-day-aware Stackelberg Game-based Multi-Agent Distributed Deep Deterministic Policy Gradient (ZSG-MAD3PG) framework. This model integrates the Stackelberg game with the Multi-Agent Distributed Deep Deterministic Policy Gradient (MAD3PG) algorithm, providing a robust foundation for continuous learning and dynamic strategy adaptation [13,14]. Its ability to handle continuous action spaces and to coordinate across distributed agents makes it particularly effective in managing the evolving attack patterns and defense requirements of IIoT networks [15]. A key advancement of ZSG-MAD3PG lies in its capacity to learn optimal defense policies that adapt in real time to changes in attack strategies and environmental conditions, ensuring a rapid convergence to secure and efficient communication [16,17].

Zero-day attacks characterized by their stealthiness, rapid propagation, and lack of known signatures, pose substantial threats to the confidentiality, integrity, and availability of IIoT control systems, particularly targeting data communication links with sophisticated evasion tactics [18–20]. The dynamic and adaptive nature of these attacks renders static defense policies and fixed detection rules largely ineffective, emphasizing the need for advanced, dynamic, and context-aware defense strategies. Traditional game-theoretic and ML-based approaches often fail to fully account for the nuanced perceptual asymmetries and nonlinear interactions present in IIoT zero-day attack scenarios. Addressing these challenges requires a comprehensive framework that combines dynamic game-based deception, continuous adaptation through MARL, and context-aware behavioral fingerprinting, thereby ensuring more robust and scalable protection for IIoT data communication networks.

To this end, this paper proposes the ZSG-MAD3PG framework—a Stackelberg-game-based dynamic defense architecture tailored to zero-day attacks in IIoT control systems. In this framework, the defender acts as the leader in the Stackelberg game, dynamically adjusting the packet sampling strategy using a Long Short-Term Memory (LSTM)-enhanced MAD3PG algorithm to address evolving zero-day attack patterns [21–23]. Meanwhile, the attacker acts as the follower, modifying its attack strategy to reach a rapid and stable attack-defense equilibrium [24–26]. Additionally, the integration of a dynamic defensive deception system—comprising adaptive baiting systems and adversarially generated obfuscated traffic—enhances the detection and mitigation of zero-day attacks by shortening concealment times and reducing resource consumption. Specifically, compared with traditional static defense schemes that typically incur significant CPU and memory overhead and introduce noticeable latency in IIoT scenarios, the proposed ZSG-MAD3PG framework effectively reduces resource consumption and latency while maintaining high stealthiness and detection accuracy. Finally, a lightweight behavioral fingerprinting detection mechanism ensures real-time zero-day attack identification within industrial device clusters [27–29].

The main contributions of this paper are summarized below:

1. **Zero-Day-Aware Dynamic Defense Framework.** We propose the ZSG-MAD3PG framework, which synergistically combines Stackelberg game theory with an LSTM-enhanced MAD3PG algorithm to effectively address the evolving complexity of zero-day attacks in IIoT data communication. The incorporation of the LSTM module enables the framework to capture temporal dependencies and sequential patterns in network traffic data, thereby facilitating more informed and adaptive adjustments to the packet sampling strategy. This capability significantly accelerates the convergence rate of the MAD3PG algorithm, leading to a rapid and stable attack-defense equilibrium within the leader-follower Stackelberg game setting. Consequently, this framework provides a realistic and robust approach for securing IIoT environments against sophisticated zero-day threats.
2. **Adaptive Deceptive Techniques to Counter Evolving Threats.** We integrate dynamic defensive deception techniques—comprising adaptive baiting and adversarially generated obfuscated traffic—to strategically mislead attackers, reducing the concealment time of zero-day attacks and lowering the system's resource consumption.
3. **Lightweight Detection Mechanism for Real-Time Zero-Day Attack Defense.** We design and evaluate a lightweight detection mechanism based on behavioral fingerprinting, demonstrating improved detection latency, enhanced resource efficiency, and strong stealthiness against zero-day attacks. Performance analyses validate the framework's significant advancements in IIoT zero-day attack protection, providing an effective balance between real-time security and system resource constraints.

The remainder of this paper is organized as follows: Section 2 surveys relevant literature. Section 3 elaborates on the system architecture and modeling assumptions. Section 4 introduces the Stackelberg-game-based dynamic defense framework, describing the attacker and defender models and the ZSG-MAD3PG algorithm with dynamic deceptive defenses. Section 5 provides comprehensive numerical evaluations and performance comparisons. Section 6 concludes the paper.

2 Related Work

In recent years, zero-day attacks have emerged as a critical threat to data communication security in IIoT systems, driving extensive research into diverse defense strategies [30]. Existing studies systematically review and compare Machine Learning (ML) and Deep Learning (DL)-based zero-day attack detection methods, focusing on their performance in terms of accuracy, recall, and generalization, while highlighting key challenges and future research directions [31–33]. Some works introduce zero-shot learning frameworks that

map known and unknown behaviors through semantic attributes and propose new metrics, such as the Zero-day Detection Rate (Z-DR), to reveal performance gaps of Network Intrusion Detection Systems (NIDS) in detecting zero-day attacks [34]. To address the complexity of zero-day attacks in IIoT environments, several studies propose advanced frameworks—for example, a 5G-enabled dual autoencoder-based federated learning model that integrates anomaly detection with collaborative self-learning, outperforming traditional ML approaches [35]. The hybrid detection framework in [36] leverages neural algorithmic reasoning and meta-learning to effectively detect and classify zero-day attacks even under limited training data conditions. Furthermore, DL-based intrusion detection systems combining open set recognition, clustering, and model updating techniques have been developed to enhance the detection of unknown attacks and support practical labeling and adaptation, demonstrating superior results on benchmark datasets [37]. Additionally, some research shifts the defense paradigm from attacker-driven to asset-centered strategies by employing virtual machine introspection and system call interception to monitor critical assets, thereby enabling timely detection of zero-day ransomware attacks [38]. In the Internet of Vehicles (IoV) domain, the Zero-X framework combines blockchain-enabled federated learning, deep neural networks, and open set recognition to ensure privacy-preserving collaboration while achieving high detection accuracy and reducing false positives for both zero-day and known attacks [39]. Although these studies advance zero-day attack detection through innovative frameworks and evaluation metrics, they often suffer from data scarcity, high algorithmic complexity, and limited real-time processing capability, making practical deployment in large-scale IIoT environments challenging. In particular, many approaches rely heavily on extensive labeled datasets, involve high computational overhead, and lack adaptive mechanisms to continuously respond to stealthy and evolving zero-day attack strategies. These limitations highlight the need for a more scalable and adaptive defense framework.

To systematically optimize defensive strategies against zero-day threats, game-theoretic models have garnered considerable attention. Among these, Stackelberg games are particularly well-suited for modeling the hierarchical interactions between defenders and adaptive attackers in IIoT environments. The Stackelberg game framework effectively captures the dynamic interplay of decisions made by defenders and attackers, offering a structured approach to enhance security and resource allocation. For instance, a Stackelberg game-based hierarchical decision-making framework has been employed to model cooperative incentive interactions between cloud servers and edge servers in cloud computing networks, leading to efficient resource trading and improved quality of service [40]. Recent work has extended the leader–Stackelberg game decision-making mechanism by introducing cooperative incentive-based models, defining a critical cooperativeness threshold to ensure unique Stackelberg equilibrium solutions. A consensus model is established, and numerical validation demonstrates its applicability and parameter sensitivity [41]. In cognitive radio networks, Stackelberg games have been leveraged to optimize transmission power and spectrum-sharing strategies, addressing resource allocation challenges and outlining future research directions [42]. Similarly, a two-stage Stackelberg game approach has been applied to mobile crowdsensing systems, designing an incentive mechanism that balances high-quality data contributions from mobile users with revenue maximization for service providers [43]. In the context of distributed cold-chain logistics, a blockchain-based Stackelberg game resource allocation model enables secure data management and efficient pricing, ultimately enhancing revenue streams and data value [44]. Within cyber-physical systems, a Stackelberg game model has been used to analyze and optimize Denial of Service (DoS) attack strategies, incorporating self-adaptive particle swarm optimization and online algorithms to compute Stackelberg equilibrium solutions and manage energy resources effectively [45]. In Unmanned Aerial Vehicle (UAV)-aided mobile-edge computing networks, a Stackelberg game framework has been formulated to optimize computation offloading strategies managed by an edge service provider, achieving a unique Nash equilibrium

via backward induction and gradient-based iterative algorithms, leading to improved utility and performance compared to benchmark methods [46]. In addition to these applications, Stackelberg game models have also been integrated with reinforcement learning techniques to address robust optimal tracking control in uncertain nonlinear systems with disturbances and actuator saturation. This approach incorporates prescribed performance constraints and reinforcement learning-based online approximation of optimal controllers, validated through numerical simulations and hardware experiments on quadcopter systems [47]. Moreover, a no-regret online learning algorithm for discrete-time dynamic Stackelberg games has been proposed, addressing scenarios with unknown but linearly parameterized follower utilities. This algorithm adaptively adjusts the leader's randomized strategies based on past follower responses, achieving sublinear regret independent of the state space size and outperforming existing model-free reinforcement learning approaches [48]. While these studies demonstrate the versatility and effectiveness of Stackelberg games in different domains, they generally focus on static or semi-static scenarios and often lack integration with adaptive learning mechanisms needed to address stealthy and dynamic zero-day attacks in real-time. Furthermore, high model complexity and substantial computational overhead limit their scalability and practical application in large-scale IIoT environments.

With the increasing complexity and frequency of cybersecurity threats in IoT systems, there is a growing interest in leveraging MAD3PG-based DD strategies that incorporate MARL algorithms to enhance the resilience of IoT environments. Recent studies have proposed various MARL-based methods, such as Empirical Clustering Layer-based Multi-Agent Distributed Deep Deterministic Policy Gradient (ECL-MAD3PG), which uses an empirical clustering layer-based multi-agent dual dueling policy gradient algorithm to reduce value overestimation and improve stability in high-dimensional environments, achieving slightly higher task completion rates than Multi-Agent Deep Deterministic Policy Gradient (MADDPG) in complex UAV cooperative combat scenarios [49]. Other works focus on integrating MAD3PG with consensus mechanisms to enhance blockchain security and efficiency in UAV-assisted 6G networks, jointly optimizing node selection and blockchain configuration as a Markov decision process and outperforming benchmarks in throughput, incentive mechanisms, and latency [50]. In vehicular edge computing, a flexible Remote Aerial Vehicle (RAV)-enabled framework that uses multi-agent deep reinforcement learning to optimize task partitioning and resource allocation has been proposed, showing superior performance compared to traditional optimization techniques [51]. Multi-Agent Federated Deep Reinforcement Learning (MAF-DRL) integrates federated learning with MARL to enhance attack detection in wireless sensor networks, offering decentralized and privacy-preserving learning and demonstrating improved energy efficiency and system robustness [52]. Further studies have explored adversarial MARL-based secure offloading in UAV-enabled edge computing, treating intelligent eavesdroppers as smart agents in a zero-sum game setting to enhance security and energy efficiency [53]. In hierarchical federated learning, a DRL-based reputation mechanism using Stackelberg Zero-sum Game (SZG)-MADDPG has been shown to improve global model accuracy and stability by better-evaluating trust and adapting to client behaviors [54]. A blockchain-secured UAV-assisted IoT data collection framework has also been introduced, employing game-theoretic and MARL approaches to jointly optimize transmission, incentives, and UAV deployment for improved system performance [55]. A broader review of MARL applications in next-generation networks—including 5G, UAV, vehicular networks, and IoT—underscores the benefits of multi-agent modeling for dynamic and decentralized decision-making, while highlighting challenges in robustness, scalability, and deployment [56]. Finally, privacy concerns in collaborative deep learning for heterogeneous IoT surveillance have been addressed by reformulating data distribution control as an MARL problem, effectively balancing privacy protection and latency performance [57]. Despite these advancements, many existing MARL- and MAD3PG-based approaches still face

limitations such as data scarcity, high algorithmic complexity, insufficient real-time policy adaptation, and high resource consumption, which hinder their scalability and practical deployment in IIoT scenarios.

Based on these insights, our work proposes new defense ideas and improvements in three areas. First, we propose a dynamic defense architecture based on the Stackelberg game, which utilizes the dynamic characteristics of the leader-follower game and the DD technique to mislead the attacker's cognition and quickly converge to a better defense strategy. Second, we propose the ZSG-MAD3PG algorithm, which effectively captures time-dependent and long-term interaction properties to enhance the temporal modeling capability of zero-day attack defense. Third, we comprehensively validate the performance of the defense framework through numerical simulation, and the results show that the method possesses higher robustness and adaptive capability in IIoT control systems, while reducing the False Alarm Probability (FAR), and improving the True Positive Rate (TPR) and System Reliability (SR), which fully proves its effectiveness and practical application potential in adversarial scenarios.

3 System Model

In this section, we introduce an edge-based IIoT network as well as its network architecture, Stackelberg game model, computation model, and equilibrium analysis.

3.1 Network Model

As shown in Fig. 1, we present a Stackelberg game-based defensive deception architecture designed to counter zero-day attacks in IIoT environments. In this framework, a Stackelberg Defender operates as the leader in a hierarchical decision-making game, orchestrating optimal defense strategies in response to potential attack actions initiated by zero-day attackers targeting the IIoT system. These attacks may originate from compromised IoT/IIoT devices attempting to breach critical assets by targeting control points (CP_1, CP_2) and the Central Controller (CC), which are part of the Deceptive Control Surface integrated within the Industrial Master Control System.

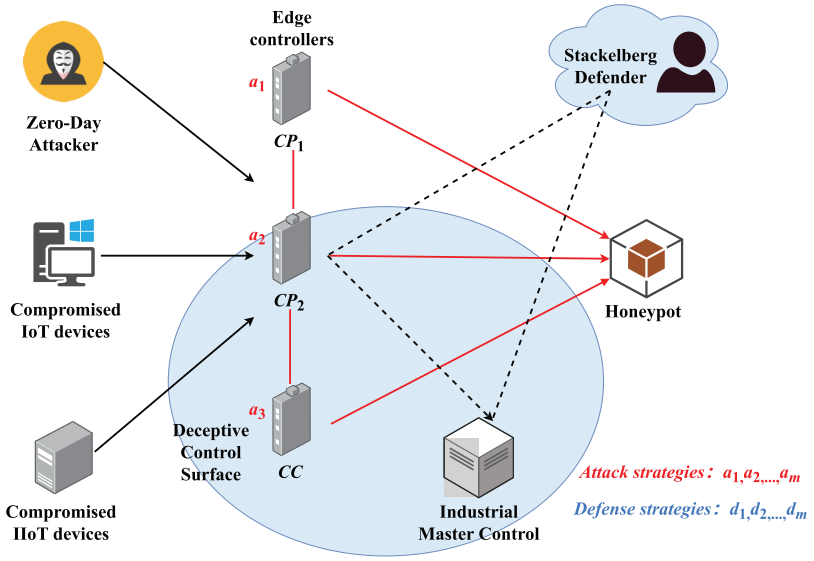

Figure 1: Stackelberg game-based deceptive defense against zero-day attacks in IIoT. The red lines show attack traffic. $a_1, a_2, \ldots, a_M$ are attack strategies on control points (CPs); $d_1, d_2, \ldots, d_M$ are corresponding deception-based defenses. The blue area marks the deceptive control surface with edge and central controllers

The architecture leverages adaptive deception by dynamically adjusting the connection paths and behaviors of CPs and CCs. By doing so, it enhances the resilience of the system against unknown threats, reduces attacker success rates, and facilitates real-time decision-making at the edge. Furthermore, the Stackelberg framework ensures that defensive actions are optimized in anticipation of attacker behavior, providing a game-theoretic foundation for proactive cybersecurity in IIoT systems.

Moreover, the architecture incorporates deceptive vulnerability disclosure techniques to manipulate the attacker's perception of system weaknesses. By selectively revealing fabricated or misleading system information, the defender increases the perceived cost of exploitation, thereby deterring attackers while preserving the functional integrity of real assets. Through continuous feedback between edge-level deception mechanisms and the centralized decision engine, the proposed framework provides a robust and adaptive solution to securing IIoT systems against zero-day vulnerabilities.

To enhance NIDS performance in IIoT environments, we design a Stackelberg game-based defensive deception architecture, integrated with a Bayesian learning framework. The defender (leader) proactively selects optimal deception strategies, while the attacker (follower) responds to maximize attack success, forming a bi-level optimization problem.

The Bayesian module adaptively updates detection metrics—FAR and TPR—in response to new deception intelligence. Let D denote the observed evidence. The posterior distributions are computed as

$$P(\mathrm{FAR} \mid D) = \frac{P(D \mid \mathrm{FAR}) \cdot P(\mathrm{FAR})}{P(D)}, \tag{1}$$

$$P(\mathrm{DR} \mid D) = \frac{P(D \mid \mathrm{DR}) \cdot P(\mathrm{DR})}{P(D)}, \tag{2}$$

where priors $P(\mathrm{FAR})$ and $P(\mathrm{DR})$ are updated via evidence from sources such as honeypots or redirected traffic.

We define two evidence integration windows: T_n for negative updates affecting FAR and T_P for positive updates affecting DR. Expected values over these windows are

$$E[\mathrm{FAR}(T_n)] = \frac{\sum_{i=1}^{T_n} \mathrm{FN}_i}{\sum_{i=1}^{T_n} (\mathrm{FN}_i + \mathrm{TP}_i)}, \tag{3}$$

$$E[\mathrm{DR}(T_p)] = \frac{\sum_{i=1}^{T_p} \mathrm{TP}_i}{\sum_{i=1}^{T_p} (\mathrm{TP}_i + \mathrm{FN}_i)}. \tag{4}$$

This adaptive framework enables real-time deception optimization and improves NIDS capability against adaptive zero-day attacks by anticipating and shaping attacker behavior through game-theoretic feedback.

To formalize the adaptive interaction between the defender and zero-day attacker within the proposed architecture, a Stackelberg game framework is adopted. Given the limitations of traditional models in handling dynamic and partially observable threats, we integrate deep learning to enhance the defender's ability to anticipate and optimize deception strategies under uncertainty.

3.2 Stackelberg Game Model

In IIoT environments, zero-day attacks pose significant threats to data communication security and system integrity. Traditional Stackelberg game models lack adaptive learning capabilities, making them insufficient to address dynamic and unknown zero-day threats. To overcome this limitation, we propose a deep neural network (DNN)-based Stackelberg game framework that models the adversarial interaction between a defender and a zero-day attacker, as illustrated in Fig. 2.

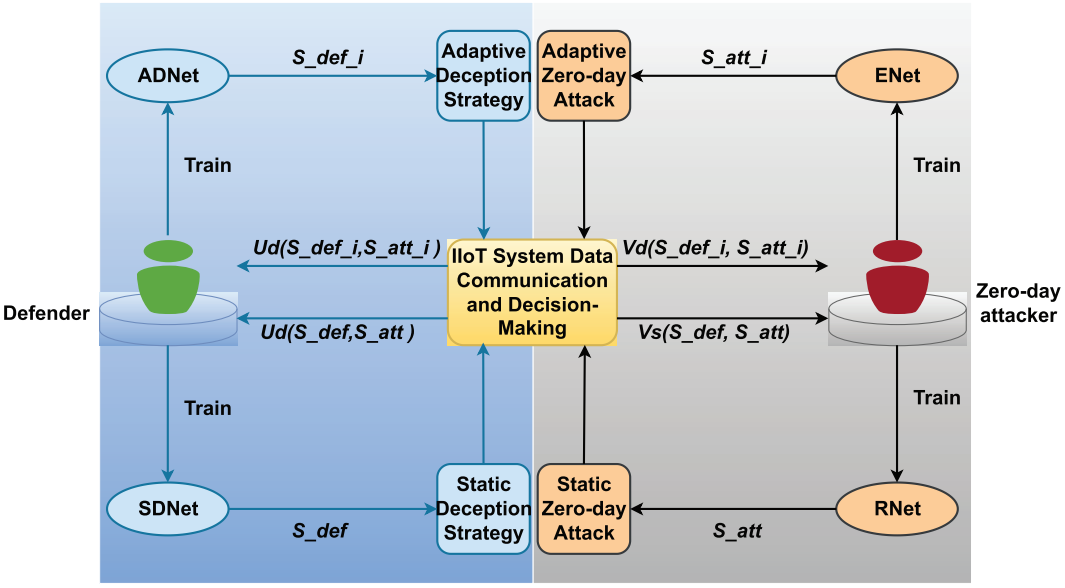

Figure 2: Framework of the proposed DNN-based Stackelberg game against zero-day attacks in IIoT environments

In this framework, the defender acts as the leader, proactively selecting deception strategies, while the attacker acts as the follower, adapting attack strategies in response. The game is formulated as

$$G = (P, S_{def}, S_{att}, N_{def}, N_{att}, U), \tag{5}$$

where:

- $P = \{\text{Defender, Attacker}\}$ represents the set of players.
- S_{def} denotes the defender's deception strategy set, including deploying honeypots, injecting fake vulnerabilities, and rerouting traffic.
- S_{att} represents the attacker's adaptive zero-day strategy set, designed to evade deception and exploit unknown system weaknesses.
- N_{def} is the DNN trained by the defender to dynamically generate optimal deception strategies based on environmental observations.
- N_{att} is the DNN trained by the attacker to infer and optimize zero-day attack strategies against the defender's deception mechanisms.
- $U = \{U_{def}, U_{att}\}$ is the set of utility functions: the defender maximizes its utility by misleading or detecting attacks with minimal disruption, while the attacker maximizes its utility by successfully compromising the system without detection.

We use two deep neural networks—Deception Strategy Network (DSnet) for the defender and Adaptive Zero-day Network (AZnet) for the attacker—to model strategy generation. The defender selects single-channel or multi-channel deception strategies, while the attacker generates corresponding adaptive attack strategies. The central computation module simulates the IIoT communication system under adversarial interaction. This DNN-Stackelberg framework enables both players to dynamically update their strategies in response to each other, effectively capturing the nature of evolving zero-day threats and improving the robustness of IIoT systems.

3.2.1 Defender's Strategy (Leader)

The defender adopts a twofold strategy that includes deceptive defense and proactive protection mechanisms. The decision involves:

- **Deceptive Strategy:** Deploy misleading information, honeypots, or decoy services to confuse or mislead attackers, thereby increasing their uncertainty and attack cost.
- **Defense Strategy:** Select appropriate reactive mechanisms like traffic filtering, network segmentation, or intrusion prevention systems.

The defender's optimization goal is

$$S_{defender} = \arg\min \left(\sum_i D_{dec,i} + C_{attack} \right), \tag{6}$$

where $D_{dec,i}$ represents the cost of deception, primarily including the deployment of honeypots (decoy systems mimicking real industrial devices) and the generation of fake network traffic to mislead attackers, and C_{attack} is the expected loss due to successful attacks.

3.2.2 Attacker's Response (Follower)

The attacker observes the defender's deception and defense deployment and selects an attack method to maximize the success probability under uncertainty. The reaction function is

$$A_{attacker} = \arg\max \left(P_{attack}(S_{defender}) \right), \tag{7}$$

where $P_{attack}(S_{defender})$ represents the success probability of an attack given the defender's deceptive and reactive strategies.

Deception strategies including static and dynamic spoofing incur varying degrees of operational costs but are effective in reducing the efficiency of attackers. Static deception provides a low-cost solution with limited adaptability, while dynamic deception incurs higher costs due to real-time interactions, but proves more effective in countering adaptive threats. By utilizing the Stackelberg game model, IIoT systems can strategically balance these deception techniques with proactive protection mechanisms to proactively disrupt attacker plans, reduce attack efficiency, and increase system resilience against zero-day exploits.

3.2.3 Equilibrium Analysis

The Stackelberg Equilibrium (SE) characterizes the optimal interaction outcome between the defender and the attacker. It ensures that, given the strategy of the opposing party, neither player can unilaterally deviate to improve their own utility. Formally, the equilibrium analysis is defined as follows.

Let $(S_{defender}, A_{attacker})$ be a strategy pair, where the defender's strategy is defined as $S_{defender} = \{Or_m, D_m\}$, with Or_m representing the adopted deception tactics and D_m denoting the selected defense mechanisms.

The defender's utility function $U_{\text{defender}}(S_{\text{defender}})$, which reflects the trade-off between protection effectiveness and resource cost, is greater than or equal to any alternative strategy $\tilde{S}_{\text{defender}} \in S_{\text{defender}}$; and simultaneously, the attacker's utility function $U_{\text{attacker}}(A_{\text{attacker}})$, which measures the success probability and expected gains of the attack, is no less than that under any other strategy $\tilde{A}_{\text{attacker}} \in A_{\text{attacker}}$. Formally,

$$U_{\text{defender}}(S_{\text{defender}}) \geq U_{\text{defender}}(\tilde{S}_{\text{defender}}), \quad \forall\, \tilde{S}_{\text{defender}} \in S_{\text{defender}}, \tag{8}$$

$$U_{\text{attacker}}(A_{\text{attacker}}, S_{\text{defender}}) \geq U_{\text{attacker}}(\tilde{A}_{\text{attacker}}, S_{\text{defender}}), \quad \forall\, \tilde{A}_{\text{attacker}} \in A_{\text{attacker}}. \tag{9}$$

Here, S_{defender} and A_{attacker} denote the strategy spaces of the defender and attacker, respectively. The pair of strategies constitutes a Stackelberg equilibrium of the game G if and only if the above conditions are satisfied.

In this Stackelberg game, the defender acts as the leader by proactively deploying deceptive and defensive strategies to mislead the attacker and reduce the likelihood of a successful intrusion. The attacker, as the follower, observes the defender's strategy and adapts its attack accordingly to maximize its utility. The resulting Stackelberg equilibrium thus represents a dynamic and stable outcome in which neither party can improve their payoff through unilateral strategy adjustment.

3.3 Adaptive Zero-Day Attacker Modeling in IIoT Stackelberg Games

3.3.1 Attacking Decision

In the proposed Stackelberg game framework for IIoT zero-day attack–defense confrontation, the attacker is modeled as the follower who observes the defender's deployed deceptive strategies and selects an optimal attack target accordingly. Let the attacker's action be defined by a decision vector $\mathbf{a} = \{a_0, a_1, ..., a_N\}$, where $a_N = 1$ indicates that the attacker chooses to launch a zero-day attack on node n (e.g., an edge server), and $a_N = 0$ otherwise. To reflect realistic constraints, we consider a single-target attack setting, enforcing $\sum_{n=0}^{N} a_N \leq 1$. Here, $a_N = 1$ represents a no-attack action, where the attacker strategically refrains from launching any attack.

3.3.2 Impact of Attacker Behavior

In the process of modeling zero-day attack behavior, the Impact Factor (I_a) serves as a critical metric for evaluating the potential damage or disruption an attacker may cause upon successfully compromising a target. To assess the relative impact of different attack targets, this study defines the impact factor $I_a(n)$ as a function of several key attributes of the target node n:

$$I_a(n) = \alpha R(n) + \beta E(n) + \gamma X(n). \tag{10}$$

Here α, β, and γ are weight coefficients representing the relative importance of each attribute. $R(n)$ represents the importance or critical role of the target node within the system. A compromise of critical nodes—such as core controllers, gateways, or databases—could severely threaten system stability and operational continuity. $E(n)$ denotes the ease with which a vulnerability on the node can be exploited, higher scores indicate greater ease of exploitation. $X(n)$ reflects the exposure level of the target node in the network architecture, including the number of external interfaces, connectivity with other nodes, and lack of isolation mechanisms.

3.3.3 Attacker Cost

Launching a zero-day attack in a hostile IIoT environment inevitably incurs costs associated with resource consumption and detection risk. To maintain computational simplicity while preserving the underlying cost factors, we model the cost of an attack $C_a(n)$ for the target node n as the linear sum of resource expenditures $\Omega(n)$ and exposure risk $\Phi(n)$:

$$C_a(n) = f\big(C_{ex}(n), \Phi_A(n)\big) = C_{ex}(n) + \sqrt{\Phi_A(n)}, \tag{11}$$

where $C_{ex}(n)$ quantifies the resource cost for successfully compromising node n, including time and computational effort, and $\Phi_A(n)$ represents the detection or tracing risk during the attack. The square root function applied to $\Phi_A(n)$ models diminishing marginal impact of increasing exposure risk on the total cost, reflecting that beyond a certain threshold, additional detection risk incrementally adds less to the overall cost.

This formulation maintains computational tractability while capturing nonlinear cost interactions, thus enabling more realistic and flexible integration within the attacker's utility optimization framework in dynamic IIoT security scenarios.

3.3.4 Uncertainty of Attacker Behavior

To characterize the attacker's strategy selection behavior under conditions of incomplete or imperfect information, this study introduces an attacker uncertainty formulation based on the Stackelberg game model. This metric quantifies the variability in the attacker's strategic decisions when confronted with complex defense mechanisms. The uncertainty of the attacker is defined as

$$U_A = 1 - \exp\left(-\frac{(1-\Phi_A)\cdot \Pi_D + \Theta \cdot \Lambda}{T_A \cdot \Omega_A}\right). \tag{12}$$

Here, T_A denotes the dwell time in the system, while $\Phi_A \in [0,1]$ represents the probability of detection, characterizing the likelihood that the attacker is identified by the defender. The parameter $\Pi_D \in [0,1]$ reflects the complexity or intensity of the deployed defense strategy, where higher values indicate more sophisticated or multi-layered defenses. Θ captures the attacker's cognitive bias, indicating misjudgment of the defense. Λ measures the intensity of the attack–defense confrontation, and Ω_A represents the attacker's risk aversion, reflecting adaptability to changing defenses. Together, these variables shape the attacker's decision-making under uncertainty.

This formulation captures how the attacker's behavior is constrained by their understanding of the environment, the intensity of confrontation, and their tolerance to risk. Notably, when the attacker entirely ignores the presence of the defense (i.e., $\Theta = 0$), the uncertainty reaches its maximum value: $U_S = 1$. This implies that the attacker is unable to adjust its behavior based on the defensive dynamics, resulting in complete strategic ambiguity.

3.4 Adaptive Defense in IIoT Stackelberg Games

3.4.1 Defense Strategy

In the Stackelberg game framework for IIoT security, the defender acts as the leader who commits to a defensive strategy in anticipation of the attacker's adaptive responses. The defender's strategy involves two key components: defensive deception and bundled defense mechanisms. Let the defender's decision be denoted by a strategy vector $\mathbf{d} = [d_1, d_2, \ldots, d_N]$, where $d_n \in [0,1]$ indicates whether a deceptive or protective defense is activated on node n.

Deceptive defense aims to mislead the attacker by introducing uncertainty about real system states, whereas bundled defense integrates traditional security controls—such as Role-Based Access Control (RBAC) for managing user permissions and machine learning-based anomaly detection for identifying abnormal communication patterns—with decoys or obfuscation techniques to increase the overall defense complexity. The joint deployment strategy enforces a resource constraint $\sum_{n=1}^{N} C_n \cdot d_n \leq B_D$, where C_n is the deployment cost at node n, and B_D is the defender's total budget.

3.4.2 Impact of Defender Behavior

To reduce model complexity, the Defensive Impact Factor $I_D(n)$ is expressed by aggregating positive and negative components:

$$I_D(n) = x \cdot (J(n) + Q(n) + Z(n)) - y \cdot (\Psi(n) + \sigma_D). \tag{13}$$

Here, $J(n)$ represents the monitoring capability of node n, reflecting the frequency and depth of security inspections. $Q(n)$ denotes the intensity of deployed deception mechanisms, including honeypots and decoy data. $Z(n)$ captures the adaptive response ability of the defense system, measuring how dynamically the defense strategy adjusts upon detecting attacks. $\Psi(n)$ quantifies network communication instability, including bandwidth fluctuations, latency, and packet loss, which negatively impact defense performance. σ_D reflects the variance in defense strategy selection, indicating instability or inconsistency in the defense approach. The positive weighting coefficients x and y tune the relative importance of each factor. Together, these elements comprehensively characterize the overall defense impact at node n.

3.4.3 Defender Cost

To simplify cost modeling, the defender's cost at node n is defined as the sum of baseline resource consumption and management complexity:

$$C_D(n) = C_{PS}(n) + \Gamma \cdot AL_n^{\delta}, \tag{14}$$

where $C_{PS}(n)$ is the resource cost at node n, $AL_n \in [0,1]$ denotes the defense activation level, $\Gamma > 0$ is the complexity coefficient, and $\delta > 1$ controls the nonlinear growth of complexity. The total defense cost is subject to the following budget constraint:

$$\sum_{n=1}^{N} C_D(n) = \sum_{n=1}^{N} \left(C_{PS}(n) + \Gamma \cdot AL_n^{\delta} \right) \leq B_D, \tag{15}$$

which ensures cost-efficiency while preserving key defense overhead factors.

3.4.4 Uncertainty in the Defender's Behavior

Modeling uncertainty in defender's behavior to quantify the uncertainty in a defender's response under IIoT environments, we design a behavioral uncertainty model based on the principles of stochastic game theory and risk sensitivity. The proposed uncertainty function for the defender is defined as

$$U_D = 1 - \exp\left(-\frac{\mu_D \cdot \Phi_D \cdot \Psi}{T_D} \right) \cdot (1 + \sigma_D). \tag{16}$$

Specifically, $\mu_D \in [0,1]$ denotes the defender's sensitivity to uncertainty, also referred to as the risk aversion coefficient. The parameter Φ_D denotes the probability of detecting the attacker. $\Psi \in \mathbb{R}^+$ accounts

for communication instability, which can be quantified by the average bandwidth variation, delay, and packet loss rate. T_D indicates the defender's monitoring duration, which reflects the defender's capability to anticipate and respond to potential attacks. Finally, σ_D represents the uncertainty associated with the choice of different strategies by the defender, which can be further formalized by modeling σ_D as the variance of the defense's strategy choice. Specifically, we define

$$\sigma_D = \frac{1}{M}\sum_{i=1}^{M}(d_i - \bar{d})^2, \tag{17}$$

where d_i denotes the encoding (or performance metric) of the i-th selected defense strategy, $\bar{d}$ is the mean of all selected strategies, and M represents the total number of strategy samples observed over time. This variance-based formulation captures the fluctuation in the defender's strategic choices, thereby quantifying the degree of unpredictability from the attacker's perspective.

3.5 Strategic Utility Formulation in IIoT Stackelberg Games

In the proposed Stackelberg game framework for IIoT security, we model the interaction between a zero-day attacker and a defender employing bundled deceptive strategies. The defender, as the leader, first deploys defense strategies across critical nodes, while the attacker, as the follower, observes the defense configuration and selects the optimal target to attack. To analyze strategic behaviors under uncertainty and resource constraints, we formulate utility functions for both players that reflect their goals, limitations, and the impact of their decisions.

3.5.1 Attacker Utility Function

The goal of the attacker is to maximize the net benefit from a successful zero-day attack while minimizing the associated costs and behavioral uncertainty. The utility function consists of three core components: the impact of a successful attack, the costs incurred during the attack, and the impact of decision uncertainty due to incomplete information and deceptive defenses, and is formulated as

$$EU_A(n) = \left(I_a(n)\cdot(1 - P_D(n)) - C_a(n)\right)\cdot(1 - U_S). \tag{18}$$

Here $I_a(n)$ represents the impact factor of attacking node n, incorporating its criticality, exploitability, and exposure level within the network. The term $P_D(n) \in [0,1]$ denotes the effectiveness of the defense strategy deployed at node n, which reduces the probability of a successful attack. $C_a(n)$ accounts for the cost of launching the attack, including both the resource expenditure and the associated exposure risk. Lastly, $U_S \in [0,1]$ reflects the attacker's behavioral uncertainty, capturing the degradation of rational decision-making caused by complex and deceptive defense mechanisms.

This formulation ensures that the attacker's utility decreases with stronger defense, higher cost, and greater uncertainty. Notably, if the uncertainty reaches its maximum ($U_S = 1$), the attacker's expected utility becomes zero, indicating complete strategic ambiguity.

3.5.2 Defender Utility Function

In the proposed Stackelberg game-based confrontation model for IIoT environments, the defender aims to minimize system losses resulting from successful zero-day attacks while ensuring efficient allocation of limited resources under bundled defense strategies. To reflect this dual objective, the defender's utility

function is formulated by considering both the expected impact of attacks and the total defense cost. The utility is expressed as

$$EU_D(n) = -\sum_{n=1}^{N} a_n \cdot L(n) - \pi \cdot \sum_{n=1}^{N} C_D(n). \tag{19}$$

Here, $a_n \in \{0,1\}$ is a binary decision variable indicating whether node n is targeted by the attacker. The expected loss at node n is defined as

$$L(n) = (1 - P_D(n)) \cdot I_a(n), \tag{20}$$

where $P_D(n) \in [0,1]$ represents the effectiveness of the deployed defense at node n, and $I_a(n)$ is the impact factor reflecting the node's importance, exploitability, and exposure. $C_D(n)$ denotes the defense cost at node n, incorporating resource usage and management complexity. The parameter $\pi > 0$ is a weighting coefficient that adjusts the defender's sensitivity to defense expenditure.

This utility formulation balances the need to reduce the consequences of successful attacks and the imperative to optimize defensive resource deployment. It serves as the defender's optimization objective in the Stackelberg game, supporting rational strategic planning against adversarial threats under constrained conditions.

3.5.3 Stackelberg Game-Based Utility Optimization

We model the strategic interaction between a zero-day attacker and a defender using bundled deception in an IIoT system as a Stackelberg game. The defender (leader) allocates defense resources across critical nodes, anticipating the attacker's (follower) best response. The bilevel optimization is

$$\max_{\{P_D(n)\}} \quad EU_D = -\sum_{n=1}^{N} a_n^* \cdot L(n) - \pi \cdot \sum_{n=1}^{N} C_D(n), \tag{21}$$

$$\text{subject to:} \quad a_n^* = \arg\max_{a_n \in \{0,1\}} \{EU_A(n)\}, \tag{22}$$

$$EU_A(n) = (I_a(n)(1 - P_D(n)) - C_a(n))(1 - U_S). \tag{23}$$

All variables follow the notation defined previously. This formulation captures the hierarchical decision-making structure, enabling the defender to proactively deploy deception-based defense strategies under uncertainty and limited resources.

To address the limitations of static game-theoretic models in dynamic IIoT environments, we integrate deep reinforcement learning into the Stackelberg framework. The following section introduces ZSG-MAD3PG, a practical algorithm for learning adaptive deception strategies against zero-day attacks under uncertainty.

4 Practical MAD3PG-Based Deceptive Defense against IIoT Zero-Day Attacks

To effectively derive optimal defensive responses in the face of asymmetric and evolving threats, we integrate deep reinforcement learning into the Stackelberg game framework. Specifically, we propose a novel algorithm named ZSG-MAD3PG, which enables the defender, as the game leader, to learn and refine deception and protection strategies within a continuous action space. This integration supports real-time adaptive decision-making by considering the uncertainty and dynamic evolution of zero-day

attacks. By combining game-theoretic modeling with deep reinforcement learning, the proposed ZSG-MAD3PG approach significantly improves the defender's capability to anticipate and mitigate hidden or previously unknown threats, thereby enhancing the resilience and integrity of IIoT data communications under zero-day conditions.

Building upon this foundation, and in order to further enhance data communication security in IIoT control systems under an incomplete-information Stackelberg game environment, we design ZSG-MAD3PG as a novel algorithm tailored to optimizing dynamic defense strategies against sophisticated cyber threats. This framework is particularly suited to addressing zero-day attacks, which are often stealthy, abrupt, and unpredictable.

In such an adversarial Stackelberg game environment, attackers continuously evaluate potential attack paths by estimating their expected utilities (EU_A) while considering the defender's utility (EU_D), thereby formulating adaptive strategies for launching zero-day attacks. Meanwhile, defenders must develop optimal response strategies despite lacking full knowledge of the attacker's goals, tactics, and timing. The proposed ZSG-MAD3PG algorithm facilitates this process by enabling proactive policy updates, ultimately improving system robustness and maintaining operational stability in the presence of strategic and high-impact adversaries.

The ZSG-MAD3PG framework integrates dual Q-learning, a dyadic network architecture, and a policy steering mechanism to efficiently learn optimal defense policies in IIoT environments. Its architecture comprises a state-strategy network φ, dominance network Δ, value network ψ, a target network φ_{target}, and strategy replicas. At each time step t, the input state s_t is composed of features that capture zero-day attack characteristics, quantified uncertainty levels, historical interaction records, environmental cost, and defense perception indicators. The defender selects action a_t via φ, receives reward r_t and next state s_{t+1}, and stores the transition (s_t, a_t, r_t) in the replay buffer D. Once the buffer exceeds a predefined threshold B, mini-batches are sampled to compute current and target Q-values, construct the target $r + \gamma Q(s', a'; \varphi_{\text{target}})$, and update parameters via gradient descent. To ensure training stability, φ_{target} is periodically synchronized with φ. Furthermore, the focused training procedure employs a decentralized implementation to enhance scalability and reduce communication overhead, which is crucial for IIoT scenarios with distributed nodes. The overall computational complexity per training iteration is approximately $\mathcal{O}(B \cdot |A| \cdot |S|)$, where B is the mini-batch size, $|A|$ the action space dimension, and $|S|$ the state space dimension. By leveraging parallel updates and experience replay, the framework maintains a feasible computational burden and achieves fast convergence, enabling the defender to improve long-term performance and adaptiveness under diverse and unknown attack strategies, especially zero-day threats, while ensuring practical deployability in real-world IIoT environments.

As shown in Algorithm 1, the proposed ZSG-MAD3PG algorithm adopts a centralized training with decentralized execution framework to enhance data communication security in IIoT environments. Each agent is equipped with three neural components: a policy network φ, a value stream ψ, and an advantage stream Δ, which together construct a dueling Deep Q-Network (DQN) architecture for robust decision-making.

Algorithm 1: ZSG-MAD3PG: Stackelberg game-based MAD3PG for defending IIoT zero-day attacks

1: Randomly initialize the policy network φ, value stream ψ, and advantage stream Δ;
2: Initialize target network $\varphi_{\text{target}} \leftarrow \varphi$ with same weights;
3: Set up the replay buffer D with capacity M;
4: Configure optimizer and parameters: learning rate λ, batch size B, discount factor γ, and target update frequency C;

(Continued)

Algorithm 1 (continued)

for each episode $k = 1, 2, \cdots, K_{\max}$ **do**
  Reset the IIoT attacker-defender environment, and obtain initial state s_0;
  Set $s_t = s_0$;
  for each time step $t = 1, 2, \cdots, T_{\max}$ **do**
    Select action $a_t = \arg\max Q(s_t, a; \varphi)$ using current policy network;
    Execute a_t, and observe reward r_t and next state s_{t+1};
    Store transition (s_t, a_t, r_t) into replay buffer D;
    Set $s_t = s_{t+1}$;
  end for
  if replay buffer D has at least B samples **then**
    Sample a mini-batch of transitions *(s, a, r)* from D;
    Compute current Q-values $Q(s, a; \varphi)$ using φ;
    Compute next actions $a' = \arg\max Q(s, a; \varphi)$;
    Compute target Q-values $Q(s, a'; \varphi_{\text{target}})$;
    Compute expected targets $r + \gamma Q(s, a'; \varphi_{\text{target}})$;
    Perform gradient descent on loss between predicted and target Q-values;
    if $t \bmod C = 0$ **then**
      Update target network: $\varphi_{\text{target}} \leftarrow \varphi$;
    end if
  end if
end for

To address adversarial interactions in IIoT systems, we model the attacker as the leader and the defender as the follower in a Stackelberg game. The attacker selects its action first, while the defender observes and responds strategically. Each episode begins with resetting the IIoT environment and observing the initial state s_0. At each time step t, the agent selects an action a_t by maximizing the Q-value estimated by the current policy network, i.e., $a_t = \arg\max_a Q(s_t, a; \varphi)$. After executing the action, the agent observes the reward r_t and the next state s_{t+1}, and stores the transition (s_t, a_t, r_t) into the replay buffer D.

Once the replay buffer contains sufficient experience (at least B samples), a mini-batch of transitions is sampled to update the policy. The current Q-values are computed using φ, while the target Q-values are computed using the target network φ_{target}. The expected Q-target that follows the double DQN strategy to mitigate overestimation bias is computed as

$$y_q = r + \gamma Q(s, a'; \varphi_{\text{target}}), \tag{24}$$

where $a' = \arg\max_a Q(s, a; \varphi)$. The loss function is defined as the mean squared error (MSE) between the predicted Q-values and target Q-values:

$$\mathcal{L}(\varphi) = \mathbb{E}_{(s,a,r)\sim D}\left[\left(Q(s, a; \varphi) - y_q\right)^2\right]. \tag{25}$$

The model parameters are updated via gradient descent, and the target network φ_{target} is periodically synchronized with the policy network every C steps.

In the Stackelberg game context, the reward function r_t captures the response effectiveness of the defender (follower) under the strategy of the attacker (leader). We define the reward as

$$r_t = U_{\text{follower}}(s_t, a_t^{\text{follower}} \mid a_t^{\text{leader}}) - U_{\text{leader}}(s_t, a_t^{\text{leader}}). \tag{26}$$

To clarify, the advantage stream Δ mentioned in Algorithm 1 and the advantage network Δ in the architecture refer to the same component within the Dueling DQN design. Specifically, in Dueling DQN, the Q-value function $Q(s,a)$ is decomposed into a state-value function $V(s)$ and an advantage function $A(s,a)$. The advantage network Δ estimates $A(s,a)$, highlighting the relative importance of actions given a state, while the value stream ψ estimates $V(s)$, representing the intrinsic value of the state itself. These are combined as

$$Q(s,a) = V(s) + \left(A(s,a) - \frac{1}{|\mathcal{A}|}\sum_{a'} A(s,a')\right). \tag{27}$$

This decomposition helps stabilize learning and improves policy differentiation. The policy network φ integrates these outputs and determines optimal actions. The target network φ_{target} serves as a delayed copy of φ for stable target value computation. In addition, strategy replicas refer to copies of networks maintained during training to enhance exploration and convergence.

Fig. 3 illustrates the dynamic attacker–defender interactions within the Stackelberg framework, highlighting how ZSG-MAD3PG adapts to diverse real-time attack strategies. As the Stackelberg leader, the defender proactively anticipates the attacker's responses and iteratively refines its policies using historical data and real-time feedback. By integrating deep reinforcement learning with game-theoretic modeling, ZSG-MAD3PG enables adaptive optimization of deception and protection strategies, ensuring rapid responses to zero-day attacks and maintaining robust operational resilience in complex IIoT environments. This design encourages agents to learn optimal responses in adversarial settings, guided by the Stackelberg hierarchy. This learning framework enables each agent to adaptively evolve its policy under adversarial threats, optimizing long-term rewards and enhancing defense robustness in IIoT communication networks.

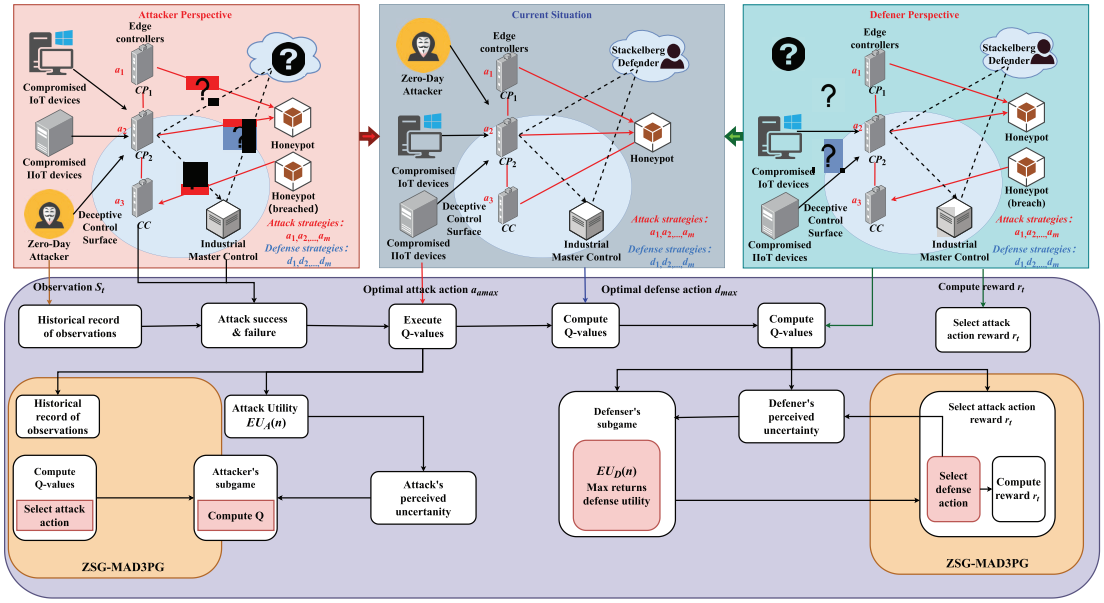

Figure 3: Interaction between the attacker (leader) and defender (follower) in a multi-agent IIoT environment under the Stackelberg game, integrated with the ZSG-MAD3PG algorithm for strategic policy learning

5 Numerical Results and Analysis

This section outlines our IIoT data communication network simulation setup, incorporating zero-day attack scenarios. We then evaluate the performance of the ZSG-MAD3PG algorithm across different

learning rates and attack success levels. Finally, we compare ZSG-MAD3PG with representative baselines, demonstrating its effectiveness against zero-day threats in dynamic IIoT settings.

5.1 Simulation Settings

We established a simulated IIoT data communication network environment, consisting of legitimate IIoT devices alongside 30 strategically deployed honeypots to implement defensive deception strategies. The choice of 30 honeypots was based on balancing computational feasibility and sufficient coverage to realistically represent large-scale IIoT networks, thus allowing assessment of the deception strategy's scalability. Within this environment, the attacker and defender iteratively interact under a Stackelberg game framework: the attacker advances upon detecting honeypots and exits after fulfilling their objectives, while the defender, leveraging insights into attacker behavior, dynamically deploys deception strategies targeting both legitimate devices and honeypots, accounting for inherent response delays. The simulation framework models these interactions from both parties' perspectives using Stackelberg game theory, further enhanced by the ZSG-MAD3PG algorithm to optimize strategic adaptation. We conducted 2000 simulation rounds utilizing Python 3.8 and PyTorch 1.10, demonstrating the robustness of our approach and evaluating the effectiveness of various strategic responses.

Specifically, we employed a range of metrics to assess performance, including SR, FAR, and TPR to evaluate the accuracy and reliability of the IIoT NIDS, Uncertainty-Attack (U-A), and Uncertainty-Defense (U-D) to quantify decision-making uncertainties in the attack-defense interactions, and Expected Uncertainty-Attack (EU-A) and Expected Uncertainty-Defense (EU-D) to measure the expected utilities of the IIoT zero-day attackers and defenders under deceptive defense strategies. Furthermore, we compared the performance of ZSG-MAD3PG against multiple baselines, including Stackelberg game-based strategies (SG), traditional game-based strategies (TG), and randomized strategies (RG), as well as traditional ML-based approaches, both with DD and without DD (ND) mechanisms. These combinations yielded eight different experimental scenarios: SG/ZSG-MAD3PG/DD, SG/ML/DD, RG/ZSG-MAD3PG/DD, RG/ML/DD, SG/DD, RG/DD, SG/ND, and RG/ND. The evaluation metrics (SR, FAR, TPR, U-A, U-D, EU-A, EU-D) provide a comprehensive assessment of the framework's performance, and a visual table representation further illustrates the superiority of ZSG-MAD3PG over SG, TG, and RG schemes. Table 1 summarizes the eight experimental scenarios designed to evaluate the performance of various combinations of game-theoretic strategies, learning algorithms, and the use of defensive deception mechanisms.

Table 1: Experimental scenarios: strategy combinations for IIoT defense evaluation

Scenario	Game strategy	Algorithm	Defense deception (DD)
SG/ZSG-MAD3PG/DD	Stackelberg (SG)	ZSG-MAD3PG	Enabled
SG/ML/DD	Stackelberg (SG)	ML-based	Enabled
RG/ZSG-MAD3PG/DD	Randomized (RG)	ZSG-MAD3PG	Enabled
RG/ML/DD	Randomized (RG)	ML-based	Enabled
SG/DD	Stackelberg (SG)	None	Enabled
RG/DD	Randomized (RG)	None	Enabled
SG/ND	Stackelberg (SG)	None	Disabled
RG/ND	Randomized (RG)	None	Disabled

5.2 *Influence of Step Size on the Performance of ZSG-MAD3PG*

As shown in Fig. 4, we explore how the learning rate influences the SR, TPR, and FAR of the ZSG-MAD3PG algorithm. The findings demonstrate that the ZSG-MAD3PG approach consistently surpasses other comparative methods. Fig. 4a highlights the changes in Mean Time Between Failures (MTBF) under varying learning rates. Specifically, the MTBF steadily increases as the step size varies from 0.00001 to 0.001. However, when the step size rises to 0.01, the SR begins to decline, reaching its lowest point at 0.1. This observation indicates that excessively high learning rates can significantly compromise the stability and effectiveness of the IIoT control system, resulting in a diminished SR. Such behavior is likely caused by the fact that larger learning rates lead to more abrupt and substantial parameter updates, thereby introducing instability and increasing the risk of overfitting during training.

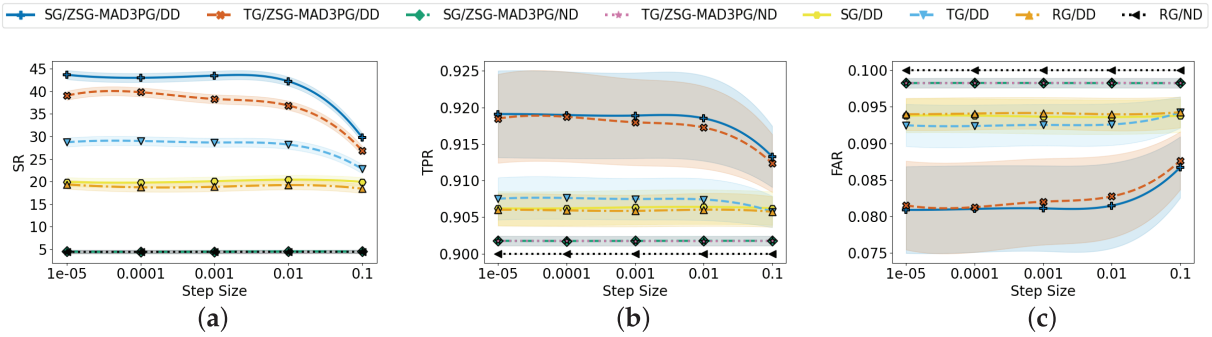

Figure 4: Comparison of step size effects across different algorithms for (**a**) SR, (**b**) TPR, and (**c**) FAR

Furthermore, this behavior is reflected in the TPR and FAR metrics, as depicted in Fig. 4b and c. At lower learning rates, the TPR remains consistently high, while the FAR is relatively low. As the learning rate increases, the TPR initially rises before eventually decreasing, whereas the FAR shows a steady upward trend. These findings underscore that although larger learning rates may accelerate the convergence process, they can also hinder detection accuracy and contribute to higher false alarm rates.

In practice, selecting an appropriate learning rate is essential for balancing training efficiency and model stability, ultimately enhancing the ZSG-MAD3PG algorithm's performance in IIoT scenarios. Our experimental results reveal that the optimal trade-off for SR, TPR, and FAR occurs at a step size of 0.001. Moreover, the ZSG-MAD3PG algorithm demonstrates consistent convergence performance across various hyperparameter settings, indicating its robustness and suitability for practical deployments. Based on these insights, we adopt a step size of 0.001 in our subsequent experiments to ensure high detection accuracy, reduced false alarms, and prolonged SR, thus providing reliable and stable performance in real-world IIoT environments.

5.3 *Comparative Performance*

To validate the effectiveness of the ZSG-MAD3PG method, we set the disruption success rate of the zero-day attack from 0.05 to 0.5 in 10-fold increments. Other parameters are set as described above.

Fig. 5 provides a comparative analysis of the SR, TPR, and FAR of the ZSG-MAD3PG algorithm and other representative methods under different IIoT zero-day compromise success rates. The results highlight that protection methods vary in effectiveness against different attackers. Fig. 5a reveals that the ZSG-MAD3PG algorithm, based on Stackelberg game theory, consistently outperforms traditional game-theoretic and ML methods in dynamic and complex IIoT environments due to its superior adaptability

and scalability. In contrast, ML algorithms can outperform ZSG-MAD3PG in simpler and static scenarios by leveraging predefined rules and datasets. The relative decline in ZSG-MAD3PG performance in such contexts is linked to its complex structure and intensive training requirements, which can result in overfitting or under-convergence.

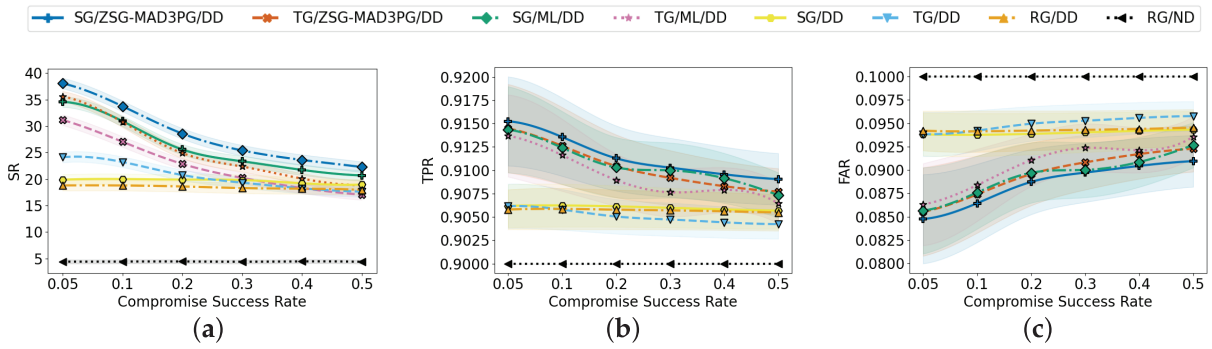

Figure 5: Algorithm efficacy assessment under varying zero-day compromise success rates across (**a**) SR, (**b**) TPR, and (**c**) FAR metrics

The DD strategy also demonstrates notable effectiveness in reducing false positives and negatives within NIDS, consistently surpassing the ND strategy across different attack intensities. Notably, as the success rate of zero-day attacks rises, the SR of all methods except ND shows a general decrease, suggesting that zero-day attacks remain feasible even in low-security conditions. Despite these trends, the ZSG-MAD3PG algorithm based on Stackelberg game theory consistently exhibits robust performance, confirming its effectiveness against advanced threats in dynamic IIoT environments.

Fig. 5b compares the performance of the eight schemes in terms of TPR under different IIoT zero-day compromise success rates. Overall, the ZSG-MAD3PG approach significantly outperforms the other compared schemes in terms of detection accuracy, showing excellent robustness and adaptability. In contrast, the scheme based on traditional ML methods performs stably in the face of static and rule-fixed attack environments, and can apply pre-trained models and fixed feature sets better. However, in highly dynamic, complex, and uncertain IIoT attack environments, traditional ML models lack flexible adaptation mechanisms, leading to a significant dip in their detection rates.

In addition, the RG scheme adopts a random selection of DD strategies at the defender's end, but it lacks the targeting and dynamic optimization capabilities to identify the diversified means of attackers in a timely and effective manner, thus limiting the detection performance. This result further highlights the advantages of dynamic game-theoretic approach, especially in the face of complex and varied IIoT attack threats.

Fig. 5c compares the FAR performance of eight detectors under varying IIoT zero-day compromise success rates. ZSG-MAD3PG with Stackelberg game theory exhibits significantly lower FAR across all scenarios, confirming its adaptive advantage in dynamic IoT environments. While ML-based schemes achieve better FAR control in static conditions, their pursuit of higher TPR typically increases FAR due to static dataset dependence and limited adaptability to evolving zero-day patterns. FAR trends across all methods correlate with TPR, indicating consistent detection-false alarm trade-offs. ZSG-MAD3PG demonstrates superior robustness and outperforms its counterparts in game-theoretic attack detection scenarios.

Fig. 6a and b presents the adversarial cost dynamics across IIoT zero-day compromise scenarios. Stackelberg's game-based approaches incur moderately higher costs than classical game-theoretic methods due to computational complexity in strategic decision-making. Conversely, RG schemes demonstrate the highest

expenditure, attributable to randomized defense policies causing resource inefficiency and suboptimal attack mitigation. Notably, the ZSG-MAD3PG framework with Stackelberg modeling maintains elevated attack costs while significantly reducing defensive expenditure. This indicates its capacity for adaptive resource allocation and protection optimization. Traditional game and ML methods exhibit intermediate cost profiles with limited defensive efficacy and subpar resource utilization compared to ZSG-MAD3PG. Critically, the ZSG-MAD3PG solution achieves optimal cost equilibrium, demonstrating superior adaptability and practical viability in dynamic threat environments.

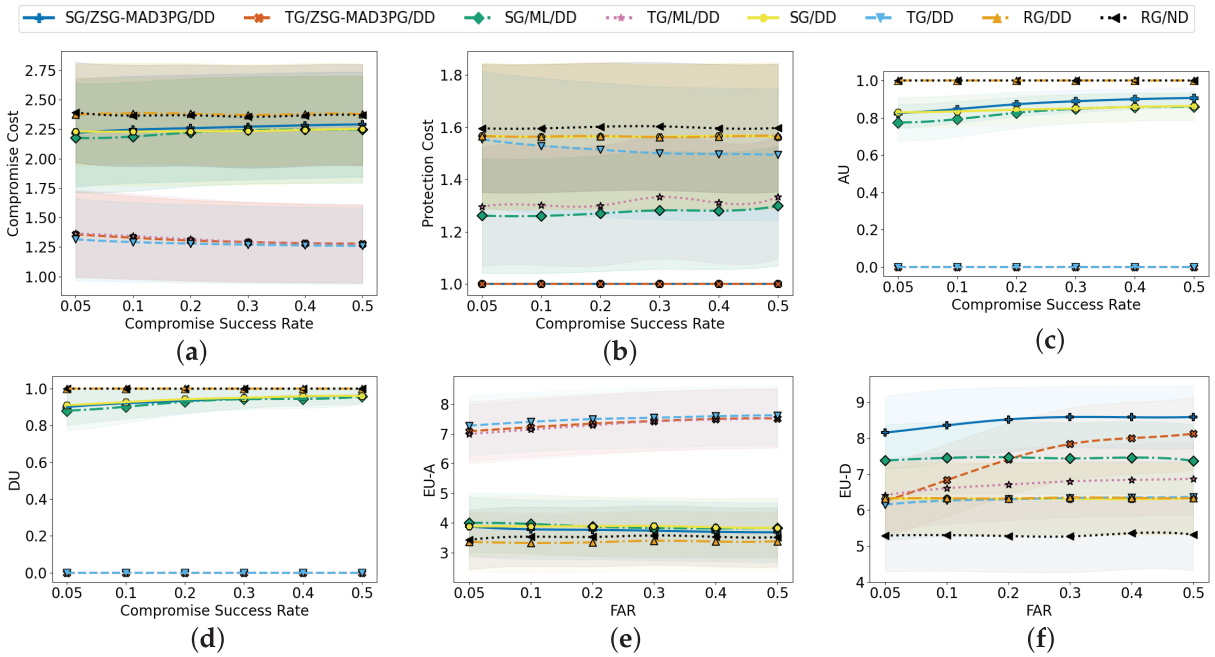

Figure 6: Algorithm performance comparison under different IIoT APT attack success rates in terms of (**a**) attack cost, (**b**) defense cost, (**c**) attack uncertainty, (**d**) defense uncertainty, (**e**) EU-A, and (**f**) EU-D

Fig. 6c and d delineates the uncertain evolution of IIoT zero-day attackers and defenders across eight schemes under varying compromise rates. Three characteristic patterns emerge: RG-based schemes demonstrate complete uncertainty for both adversarial and defensive entities, while TG approaches maintain near-zero uncertainty levels. SG methods consistently occupy the intermediate uncertainty spectrum between these extremes. Notably, SG uncertainty exhibits a positive correlation with increasing compromise rates, contrasting with the static uncertainty profiles observed in other schemes. This dynamic primarily originates from the strategic interdependence inherent in SG frameworks, where both participants continuously recalibrate actions based on perceived opponent behavior. As compromise probabilities escalate, these reciprocal adjustments create self-reinforcing feedback loops that systematically amplify uncertainty, particularly through evolving information asymmetries that impair defenders' capability to accurately model attacker potential.

Fig. 6e and f illustrates the dynamic evolution of EU-A and EU-D utilities with increasing IIoT zero-day compromise rates. Three distinct EU-A patterns emerge: RG-based schemes show the lowest utility due to high uncertainty, TG approaches reach maximum EU-A via deterministic strategies, and SG methods achieve intermediate values. Specifically, the SG/ ZSG-MAD3PG/DD configuration demonstrates reduced EU-A, thanks to its dynamic deception mechanisms that disrupt attacker decision-making. In contrast, ZSG-MAD3PG consistently achieves superior EU-D values, outperforming ML and classical

game-theoretic solutions. This EU-D advantage aligns with lower adversarial costs (Fig. 6a,b) and superior detection metrics (Fig. 5b,c). Furthermore, ZSG-MAD3PG's adaptability enables it to expand EU-D gains as compromise rates rise, underscoring its robust resource allocation and real-time defense adjustments in evolving threat environments.

Overall, the results demonstrate that ZSG-MAD3PG consistently outperforms other approaches across dynamic IIoT scenarios, offering superior detection accuracy, reduced false alarms, and more balanced attack-defense costs. Its adaptability and dynamic deception strategies ensure robust protection against evolving zero-day threats, confirming its effectiveness as a promising defense solution for complex IIoT environments.

6 Conclusion

In the face of critical vulnerabilities in IIoT data integrity exposed by zero-day intrusions, this research pioneers a stackelberg game-theoretic framework for countermeasure synthesis. The model formally derives optimal defense resource allocation strategies against industrial zero-day threats, culminating in the ZSG-MAD3PG algorithm. This novel algorithm combines multi-agent reinforcement learning and dynamic game equilibrium principles. This hybrid architecture enables autonomous policy optimization and provable convergence in a volatile IIoT ecosystem, generating real-time defense policies through emerging agent coordination. Empirical simulations validate that the proposed defense scheme combining the Stackelberg game and the ZSG-MAD3PG algorithm is more efficient in detection and more effective in defense than traditional ML-based approaches, and demonstrates unprecedented capabilities in protecting industrial control data streams from evolving IIoT zero-day attack vectors.

Acknowledgement: Not applicable.

Funding Statement: This research was funded in part by the Humanities and Social Sciences Planning Foundation of Ministry of Education of China under Grant No. 24YJAZH123, National Undergraduate Innovation and Entrepreneurship Training Program of China under Grant No. 202510347069, and the Huzhou Science and Technology Planning Foundation under Grant No. 2023GZ04.

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Shigen Shen and Xiaojun Ji; methodology, Xiaojun Ji; software, Xiaojun Ji; validation, Yimeng Liu; formal analysis, Shigen Shen and Xiaojun Ji; investigation, Yimeng Liu; data curation, Yimeng Liu; writing—original draft preparation, Xiaojun Ji; writing—review and editing, Shigen Shen and Yimeng Liu; visualization, Yimeng Liu; supervision, Shigen Shen; funding acquisition, Shigen Shen and Yimeng Liu. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Data available on request from the authors.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Ibrahim Hairab B, Aslan HK, Elsayed MS, Jurcut AD, Azer MA. Anomaly detection of zero-day attacks based on CNN and regularization techniques. Electronics. 2023;12(3):573. doi:10.3390/electronics12030573.
2. Shen S, Cai C, Li Z, Shen Y, Wu G, Yu S. Deep Q-network-based heuristic intrusion detection against edge-based SIoT zero-day attacks. Appl Soft Comput. 2024;150:111080. doi:10.1016/j.asoc.2023.111080.
3. Touré A, Imine Y, Semnont A, Delot T, Gallais A. A framework for detecting zero-day exploits in network flows. Comput Netw. 2024;248(1):110476. doi:10.1016/j.comnet.2024.110476.

4. Shen S, Cai C, Shen Y, Wu X, Ke W, Yu S. Joint mean-field game and multiagent asynchronous advantage actor-critic for edge intelligence-based IoT malware propagation defense. IEEE Trans Depend Secure Comput. 2025;22(4):3824–38. doi:10.1109/TDSC.2025.3542104.
5. Zhu W, Liu X, Liu Y, Shen Y, Gao X-Z, Shen S. RT-A3C: real-time asynchronous advantage actor–critic for optimally defending malicious attacks in edge-enabled Industrial Internet of Things. J Inf Secur Appl. 2025;91(17):104073. doi:10.1016/j.jisa.2025.104073.
6. Mir M, Trik M. A novel intrusion detection framework for industrial IoT: GCN-GRU architecture optimized with ant colony optimization. Comput Electr Eng. 2025;126(7):110541. doi:10.1016/j.compeleceng.2025.110541.
7. Liu Y, Li W, Dong X, Ren Z. Resilient formation tracking for networked swarm systems under malicious data deception attacks. Int J Robust Nonlinear Control. 2025;35(6):2043–52. doi:10.1002/rnc.7777.
8. Cheng Y, Deng X, Li Y, Yan X. Tight incentive analysis of Sybil attacks against the market equilibrium of resource exchange over general networks. Games Econ Behav. 2024;148(2):566–610. doi:10.1016/j.geb.2024.10.009.
9. Saheed YK, Misra S. CPS-IoT-PPDNN: a new explainable privacy preserving DNN for resilient anomaly detection in Cyber-Physical Systems-enabled IoT networks. Chaos Solitons Fractals. 2025;191:115939. doi:10.1016/j.chaos.2024.115939.
10. Zheng Z, Li Z, Huang C, Long S, Shen X. Defending data poisoning attacks in DP-based crowdsensing: a game-theoretic approach. IEEE Trans Mobile Comput. 2025;24(3):1859–76. doi:10.1109/TMC.2024.3486689.
11. Zhang Y, Malacaria P. Dealing with uncertainty in cybersecurity decision support. Comput Secur. 2025;148(1):104153. doi:10.1016/j.cose.2024.104153.
12. Alshahrani R, Shabaz M, Khan MA, Kadry S. Enabling intrinsic intelligence, ubiquitous learning and blockchain empowerment for trust and reliability in 6G network evolution. J King Saud Univ Comput Inf Sci. 2024;36(4):102041. doi:10.1016/j.jksuci.2024.102041.
13. Du X, Chen H, Xing Y, Yu PS, He L. A Contrastive-enhanced ensemble framework for efficient multi-agent reinforcement learning. Expert Syst Appl. 2024;245:123158. doi:10.1016/j.eswa.2024.123158.
14. Shen S, Xie L, Zhang Y, Wu G, Zhang H, Yu S. Joint differential game and double deep Q-networks for suppressing malware spread in industrial Internet of Things. IEEE Trans Inf Forensics Secur. 2023;18:5302–15. doi:10.1109/TIFS.2023.3307956.
15. Hu B, Zhang W, Gao Y, Du J, Chu X. Multiagent deep deterministic policy gradient-based computation offloading and resource allocation for ISAC-aided 6G V2X networks. IEEE Internet Things J. 2024;11(20):33890–902. doi:10.1109/JIOT.2024.3432728.
16. Yu S, Zhai R, Shen Y, Wu G, Zhang H, Yu S, et al. Deep Q-network-based open-set intrusion detection solution for Industrial Internet of Things. IEEE Internet Things J. 2024;11(7):12536–50. doi:10.1109/JIOT.2023.3333903.
17. Shen Y, Shepherd C, Ahmed CM, Shen S, Yu S. SGD3QN: joint stochastic games and dueling double deep Q-networks for defending malware propagation in edge intelligence-enabled Internet of Things. IEEE Trans Inf Forensics Secur. 2024;19:6978–90. doi:10.1109/TIFS.2024.3420233.
18. Deldar F, Abadi M. Deep learning for zero-day malware detection and classification: a survey. ACM Comput Surv. 2023;56(2):1–37. doi:10.1145/3580945.
19. Cen M, Deng X, Jiang F, Doss R. Zero-Ran Sniff: a zero-day ransomware early detection method based on zero-shot learning. Comput Secur. 2024;142(3):103849. doi:10.1016/j.cose.2024.103849.
20. Saheed YK, Chukwuere JE. CPS-IIoT-P2Attention: explainable privacy-preserving with scaled dot-product attention in Cyber-Physical System-Industrial IoT network. IEEE Access. 2025;13(3):81118–42. doi:10.1109/ACCESS.2025.3566980.
21. Wu G, Zhang Y, Zhang H, Yu S, Yu S, Shen S. SIHQR model with time delay for worm spread analysis in IIoT-enabled PLC network. Ad Hoc Netw. 2024;160(7):103504. doi:10.1016/j.adhoc.2024.103504.
22. Ye J, Cheng W, Liu X, Zhu W, Wu X, Shen S. SCIRD: revealing infection of malicious software in edge computing-enabled IoT networks. Comput Mater Contin. 2024;79(2):2743–69. doi:10.32604/cmc.2024.049985.
23. Yu S, Wang X, Shen Y, Wu G, Yu S, Shen S. Novel intrusion detection strategies with optimal hyper parameters for industrial Internet of Things based on stochastic games and double deep Q-networks. IEEE Internet Things J. 2024;11(17):29132–45. doi:10.1109/JIOT.2024.3406386.

24. Standen M, Kim J, Szabo C. Adversarial machine learning attacks and defences in multi-agent reinforcement learning. ACM Comput Surv. 2025;57(5):1–35. doi:10.1145/3571234.
25. Shen S, Cai C, Shen Y, Wu X, Ke W, Yu S. MFGD3QN: enhancing edge intelligence defense against DDoS with mean-field games and dueling double deep Q-network. IEEE Internet Things J. 2024;11(13):23931–45. doi:10.1109/JIOT.2024.3387090.
26. Adawadkar AMK, Kulkarni N. Cyber-security and reinforcement learning—a brief survey. Eng Appl Artif Intell. 2022;114:105116. doi:10.1016/j.engappai.2022.105116.
27. Zeng L, Qiu D, Sun M. Resilience enhancement of multi-agent reinforcement learning-based demand response against adversarial attacks. Appl Energy. 2022;324(2):119688. doi:10.1016/j.apenergy.2022.119688.
28. Rajaee M, Mazlumi K. Multi-agent distributed deep learning algorithm to detect cyber-attacks in distance relays. IEEE Access. 2023;11:10842–49. doi:10.1109/ACCESS.2023.3247570.
29. Sepehrzad R, Faraji MJ, Al-Durra A, Sadabadi MS. Enhancing cyber-resilience in electric vehicle charging stations: a multi-agent deep reinforcement learning approach. IEEE Trans Intell Transp Syst. 2024;25(11):18049–62. doi:10.1109/TITS.2024.3408238.
30. Ahmad R, Alsmadi I, Alhamdani W, Tawalbeh L. Zero-day attack detection: a systematic literature review. Artif Intell Rev. 2023;56(10):10733–811. doi:10.1007/s10462-023-10437-z.
31. Guo Y. A review of machine learning-based zero-day attack detection: challenges and future directions. Comput Commun. 2023;198(10):175–85. doi:10.1016/j.comcom.2022.11.001.
32. Ali S, Rehman SU, Imran A, Adeem G, Iqbal Z, Kim K-I. Comparative evaluation of AI-based techniques for zero-day attacks detection. Electronics. 2022;11(23):3934. doi:10.3390/electronics11233934.
33. Yee Por L, Dai Z, Juan Leem S, Chen Y, Yang J, Binbeshr F, et al. A systematic literature review on AI-based methods and challenges in detecting zero-day attacks. IEEE Access. 2024;12:144150–63. doi:10.1109/ACCESS.2024.3455410.
34. Sarhan M, Layeghy S, Gallagher M, Portmann M. From zero-shot machine learning to zero-day attack detection. Int J Inf Secur. 2023;22(4):947–59. doi:10.1007/s10207-023-00652-7.
35. Verma P, Bharot N, Breslin JG, O'Shea D, Vidyarthi A, Gupta D. Zero-day guardian: a dual model enabled federated learning framework for handling zero-day attacks in 5G enabled IIoT. IEEE Trans Consum Electron. 2024;70(1):3856–66. doi:10.1109/TCE.2023.3335385.
36. Cevallos M, Jús F, Rizzardi A, Sicari S, Porisini AC. NERO: neural algorithmic reasoning for zeRO-day attack detection in the IoT: a hybrid approach. Comput Secur. 2024;142(12):103898. doi:10.1016/j.cose.2024.103898.
37. Soltani M, Ousat B, Jafari Siavoshani M, Jahangir AH. An adaptable deep learning-based intrusion detection system to zero-day attacks. J Inf Secur Appl. 2023;76(1):103516. doi:10.1016/j.jisa.2023.103516.
38. Azzedin F, Suwad H, Rahman MM. An asset-based approach to mitigate zero-day ransomware attacks. Comput Mater Contin. 2022;73(2):3003–20. doi:10.32604/cmc.2022.028646.
39. Korba AA, Boualouache A, Ghamri-Doudane Y. Zero-X: a blockchain-enabled open-set federated learning framework for zero-day attack detection in IoV. IEEE Transact Vehic Technol. 2024;73(9):12399–414. doi:10.1109/TVT.2024.3385916.
40. Zhou H, Wang Z, Cheng N, Zeng D, Fan P. Stackelberg-game-based computation offloading method in cloud–edge computing networks. IEEE Internet Things J. 2022;9(17):16510–20. doi:10.1109/JIOT.2022.3153089.
41. Tang M, Liao H, Wu X. A Stackelberg game model for large-scale group decision making based on cooperative incentives. Inf Fusion. 2023;96(5):103–16. doi:10.1016/j.inffus.2023.03.013.
42. Chowdhury S. Resource allocation in cognitive radio networks using Stackelberg game: a survey. Wirel Pers Commun. 2022;122(1):807–24. doi:10.1007/s11277-021-08926-x.
43. Hu C-L, Lin K-Y, Chang CK. Incentive mechanism for mobile crowdsensing with two-stage Stackelberg game. IEEE Trans Serv Comput. 2023;16(3):1904–18. doi:10.1109/TSC.2022.3198436.
44. Zhang Y, Li C, Xin X. Stackelberg game-based resource allocation with blockchain for cold-chain logistics system. Comput Mater Contin. 2023;75(2):2429–42. doi:10.32604/cmc.2023.037139.
45. Wang Z, Shen H, Zhang H, Gao S, Yan H. Optimal DoS attack strategy for cyber-physical systems: a Stackelberg game-theoretical approach. Inf Sci. 2023;642(1):119134. doi:10.1016/j.ins.2023.119134.

46. Zhou H, Wang Z, Min G, Zhang H. UAV-aided computation offloading in mobile-edge computing networks: a Stackelberg game approach. IEEE Internet Things J. 2023;10(8):6622–33. doi:10.1109/JIOT.2022.3197155.
47. Tan J, Xue S, Li H, Guo Z, Cao H, Li D. Prescribed performance robust approximate optimal tracking control via Stackelberg game. IEEE Trans Autom Sci Eng. 2025;22:12871–83. doi:10.1109/TASE.2025.3549114.
48. Lauffer N, Ghasemi M, Hashemi A, Savas Y, Topcu U. No-regret learning in dynamic Stackelberg games. IEEE Trans Autom Control. 2024;69(3):1418–31. doi:10.1109/TAC.2023.3330797.
49. Yang Y, Li J, Hou J, Wang Y, Zhao H. A policy gradient algorithm to alleviate the multi-agent value overestimation problem in complex environments. Sensors. 2023;23(23):9520. doi:10.3390/s23239520.
50. Nahom Abishu H, Sun G, Habtamu Yacob Y, Owusu Boateng G, Ayepah-Mensah D, Liu G. Multiagent DRL-based consensus mechanism for blockchain-based collaborative computing in UAV-assisted 6G Networks. IEEE Internet Things J. 2025;12(4):4331–48. doi:10.1109/JIOT.2024.3484005.
51. Liang H, Zhang H, Ale L, Hong X, Wang L, Jia Q, et al. Joint Task Partitioning and resource allocation in RAV-enabled vehicular edge computing based on deep reinforcement learning. IEEE Internet Things J. 2025;12(11):15453–66. doi:10.1109/JIOT.2025.3527929.
52. Moudoud H, Abou El Houda Z, Brik B. Advancing security and trust in WSNs: a federated multi-agent deep reinforcement learning approach. IEEE Trans Consum Electron. 2024;70(4):6909–18. doi:10.1109/TCE.2024.3440178.
53. Li X, Huangfu W, Xu X, Huo J, Long K. Secure offloading with adversarial multi-agent reinforcement learning against intelligent eavesdroppers in UAV-enabled mobile edge computing. IEEE Trans Mob Comput. 2024;23(12):13914–28. doi:10.1109/TMC.2024.3439016.
54. Al-Maslamani NM, Abdallah M, Ciftler BS. Reputation-aware multi-agent DRL for secure hierarchical federated learning in IoT. IEEE Open J Commun Soc. 2023;4:1274–84. doi:10.1109/OJCOMS.2023.3280359.
55. Tang X, Lan X, Li L, Zhang Y, Han Z. Incentivizing proof-of-stake blockchain for secured data collection in UAV-assisted IoT: a multi-agent reinforcement learning approach. IEEE J Sel Areas Commun. 2022;40(12):3470–84. doi:10.1109/JSAC.2022.3213360.
56. Li T, Zhu K, Luong NC, Niyato D, Wu Q, Zhang Y, et al. Applications of multi-agent reinforcement learning in future internet: a comprehensive survey. IEEE Commun Surv Tutor. 2022;24(2):1240–79. doi:10.1109/COMST.2022.3160697.
57. Baccour E, Erbad A, Mohamed A, Hamdi M, Guizani M. Multi-agent reinforcement learning for privacy-aware distributed CNN in heterogeneous IoT surveillance systems. J Netw Comput Appl. 2024;230(3):103933. doi:10.1016/j.jnca.2024.103933.