Engineering Applications of Artificial Intelligence

Explainable anti-jamming defense for edge intelligence-enabled social Internet of Things networks via FlipIt game-based deep reinforcement learning

--Manuscript Draft--

Manuscript Number:	EAAI-25-17157
Article Type:	Research paper
Keywords:	Social Internet of Things; Anti-jamming; Multi-agent reinforcement learning; FlipIt game; Explainable artificial intelligence
Abstract:	Jamming attacks pose a significant threat to the operational stability of resource-constrained edge intelligence (EI)-enabled Social Internet of Things (SIoT) systems. To address this challenge, we propose a proactive and explainable anti-jamming framework that integrates game theory with multi-agent deep reinforcement learning. Specifically, we model the dynamic attack–defense confrontation between jammers and EI devices as a FlipIt game to capture continuous channel-control competition. We then develop a FlipIt Game-based Multi-Agent Attention Distributed Deep Deterministic Policy Gradient (FG-MAD3PG) algorithm. This approach enables multiple EI devices to collaboratively learn optimal anti-jamming strategies in continuous action spaces, while its integrated attention mechanism provides inherent explainability by offering transparent insights into the agents' decision-making focus. Extensive experiments demonstrate that FG-MAD3PG achieves substantial gains in cumulative reward and anti-jamming success rate over strong baselines. These results validate the effectiveness of our approach, showing it not only enhances communication resilience but also yields interpretable and trustworthy anti-jamming defense policies for EI-enabled SIoT systems.

Explainable anti-jamming defense for edge intelligence-enabled social Internet of Things networks via FlipIt game-based deep reinforcement learning

Abstract

Jamming attacks pose a significant threat to the operational stability of resource-constrained edge intelligence (EI)-enabled Social Internet of Things (SIoT) systems. To address this challenge, we propose a proactive and explainable anti-jamming framework that integrates game theory with multi-agent deep reinforcement learning. Specifically, we model the dynamic attack–defense confrontation between jammers and EI devices as a FlipIt game to capture continuous channel-control competition. We then develop a FlipIt Game-based Multi-Agent Attention Distributed Deep Deterministic Policy Gradient (FG-MAD3PG) algorithm. This approach enables multiple EI devices to collaboratively learn optimal anti-jamming strategies in continuous action spaces, while its integrated attention mechanism provides inherent explainability by offering transparent insights into the agents' decision-making focus. Extensive experiments demonstrate that FG-MAD3PG achieves substantial gains in cumulative reward and anti-jamming success rate over strong baselines. These results validate the effectiveness of our approach, showing it not only enhances communication resilience but also yields interpretable and trustworthy anti-jamming defense policies for EI-enabled SIoT systems.

Keywords: Social Internet of Things, Anti-jamming, Multi-agent reinforcement learning, FlipIt game, Explainable artificial intelligence

1. Introduction

Edge intelligence (EI)-enabled social Internet of Things (SIoT) is emerging as a foundational paradigm for connecting massive, resource-constrained devices over wireless links. Such SIoT, driven by low-power wireless interfaces, has become the core infrastructure for massive connectivity (Jiang and Zeng, 2025; van 't Schip, 2025; Pinto et al., 2025), and has established a dominant position in modern communication networks. However, the broadcast nature of wireless channels exposes SIoT networks to eavesdropping and jamming attacks. Since traditional cryptography-based security solutions demand substantial computational resources while SIoT devices are resource-constrained, these schemes are often unsuitable for SIoT environments. Integrating artificial intelligence (AI) into edge computing has therefore become a prominent trend (Amanatidis et al., 2024; Amin et al., 2025), enabling AI-assisted edge intelligence devices to construct effective learning mechanisms for EI-enabled SIoT systems.

The design philosophy of EI-enabled SIoT devices prioritizes low cost, wide coverage, and high energy efficiency, which inherently subjects them to strict resource and power constraints. Within this context, jamming attacks have emerged as a primary threat vector for disrupting normal SIoT operations. These attacks force devices to perform repeated transmission retries, causing communication interruptions and rapid battery depletion (Di Bonito et al., 2025; Zhang et al., 2024). Although various anti-jamming techniques have been proposed, most are designed for traditional networks with abundant resources. Their high computational complexity and power consumption requirements make them unsuitable for resource-constrained SIoT environments (Qiao et al., 2022; Li et al., 2022).

In the evolving landscape of wireless security threats, traditional passive defense strategies can no longer effectively defend against sophisticated jamming attacks, making proactive anti-jamming defense mechanisms a critical research priority (Jin et al., 2021; Shen et al., 2023b). In dynamic adversarial environments, there exists continuous strategic gaming between jammers and edge intelligence devices, where both parties attempt to maximize their benefits by optimizing their respective strategies. This essential characteristic of attack-defense confrontation makes game theory an ideal tool for analyzing and modeling such problems (Zhou et al., 2025; Shen et al., 2025b).

Many game-theoretic methods have achieved notable accomplishments in the anti-jamming field (Shen et al., 2023a, 2024a), effectively analyzing attack-defense equilibrium strategies in static environments and providing a theoretical foundation for understanding the basic interaction relationships between jammers and edge intelligence devices. Classical game models based on complete information assumptions have successfully addressed resource allocation and power control problems in simple scenarios (Shen et al., 2025a; Wu et al., 2023b), providing preliminary strategic guidance for anti-jamming system design. These studies have laid important theoretical groundwork for subsequent complex game modeling.

However, these game-theoretic analysis methods face significant limitations when applied to modern SIoT environments. First, they struggle with computational complexity when dealing with large-scale, high-dimensional continuous action spaces. In realistic SIoT environments, edge intelligence devices need to make real-time decisions within continuous power ranges (Kil et al., 2024; Wu et al., 2023a), multi-dimensional spectrum resources (Schultz et al., 2025), and dynamic network topologies (Vieth et al., 2025), which far exceeds the analytical capabilities of traditional game theory. Second, the non-stationarity and partial observability inherent in multi-agent environments further exacerbate the difficulty of strategy solving. Last, existing game-theoretic approaches often fail to capture the stealthy and continuous nature of modern jamming attacks, where control over communication channels can switch repeatedly between attackers and defenders.

To address these challenges, we propose the Multi-Agent Attention Distributed Deep Deterministic Policy Gradient (FG-MAD3PG) algorithm integrating with the FlipIt game model. The FlipIt game, originally designed for modeling stealthy cyber-attacks, is particularly suitable for our scenario as it captures the continuous struggle for channel control where neither party has perfect information about the opponent's state. FG-MAD3PG extends the traditional MADDPG framework by incorporating attention mechanisms to dynamically weight the importance of different environmental features and neighboring agents' information (Fan et al., 2021). The attention mechanism enables each agent to focus on the most relevant information in complex multi-agent EI-enabled SIoT scenarios, while the distributed training architecture ensures scalability across multiple edge intelligence devices while maintaining convergence stability. This algorithm combines Actor-Critic architecture with deterministic policy gradient methods, enabling edge intelligence devices to learn optimal anti-jamming strategies in continuous action spaces while effectively handling the partial observability and non-stationarity inherent in multi-agent EI-enabled SIoT environments.

While deep reinforcement learning algorithms have achieved significant breakthroughs in performance, their black-box characteristics result in a lack of transparency and interpretability in decision-making processes. This opacity may pose potential risks in safety-critical anti-jamming applications where understanding the rationale behind decisions is crucial (Mir and Rizvi, 2025). To address this issue, we introduce Local Interpretable Model-agnostic Explanations (LIME) technology (Jang and Yoon, 2025) to enhance the interpretability of the FG-MAD3PG algorithm. LIME, as a post-hoc explanation method, can explain the prediction results of arbitrary machine learning models through local linear approximation, providing an effective approach for understanding complex neural network decision processes (Zhan et al., 2025).

In our SIoT anti-jamming framework, LIME serves multiple critical functions. By generating perturbed samples near decision points and observing changes in model outputs, LIME identifies features that have the most significant impact on anti-jamming decisions, including jammer signal strength, jamming patterns, channel quality indicators, and spatial-temporal correlations (Lin et al., 2024; Suwattananuruk and Chien, 2025; Wu et al., 2023c). The interpretability analysis provided by LIME offers substantial practical value: it helps system designers optimize sensor configurations and data collection strategies, identifies system vulnerabilities and potential attack vectors, and validates algorithm rationality to improve system trustworthiness (Kandpal et al., 2025; Rahmat et al., 2025). In multi-agent EI-enabled SIoT environments, LIME reveals collaboration patterns and strategic differences among edge intelligence devices, identifying role divisions in collaborative defense and providing insights for coordination optimization. This interpretability is crucial for ensuring that edge intelligence devices in EI-enabled SIoT can continuously adapt to dynamically changing interference environments while providing trustworthy and verifiable defense strategies.

Our contributions are summarized as follows:

- We propose a novel FlipIt game-theoretic framework that captures the stealthy and continuous nature of channel control competition between jammers and edge intelligence devices in EI-enabled SIoT systems, providing a more realistic model than traditional game approaches for dynamic jamming-defense scenarios.

- We develop the FG-MAD3PG algorithm that incorporates attention mechanisms for feature importance weight-

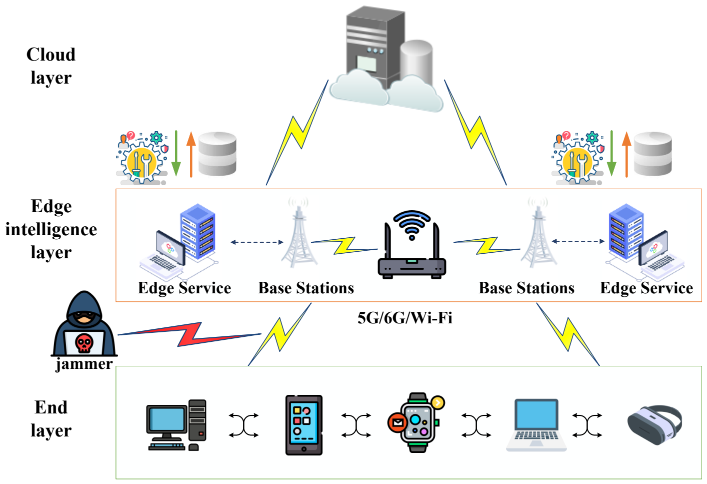


Fig. 1: An edge intelligence-enabled SIoT system with jammers.

ing and distributed training for scalability, enabling multiple edge intelligence devices to learn cooperative anti-jamming strategies in continuous action spaces.

- We conduct comprehensive experimental evaluations demonstrating that our approach achieves substantial improvements in cumulative reward and anti-jamming success rate compared to state-of-the-art baselines including Dueling Double Deep Q-Network (D3QN) and Multi-Agent Proximal Policy Optimization (MAPPO), with detailed ablation studies validating the contribution of attention mechanisms and distributed training to overall system performance.

The remainder of this article is organized as follows. Section 2 reviews related work in game-theoretic anti-jamming, multi-agent reinforcement learning, and explainable AI. Section 3 presents our system model and FlipIt game formulation. Section 4 describes the FG-MAD3PG algorithm design and LIME integration methodology. Section 5 provides experimental evaluations and interpretability analysis. Section 6 concludes the paper with discussions on limitations and future research directions.

2. Related work

To counter jamming attacks in IoT, existing research proposes a spectrum of defenses spanning real-time detection. Savva et al. (2025) constructed a fuzzy logic intrusion detection system for efficient, real-time detection of jamming attacks in IoT environments, achieving high accuracy in identifying constant, deceptive, random, reactive, and complex jammers. Yao et al. (2025) proposed a lightweight and scalable wireless signal identity detection system for IoT named SE-MobileNet. Their model is a lightweight convolutional neural network that replaces the repeated bottleneck modules in MobileNetV2 with squeeze-and-excitation modules, thereby strengthening channel recalibration and enhancing representation capacity. Shen et al. (2022) introduced a friendly jamming handover scheme that, with the help of multiple cooperative jammers, protects downlink transmissions from a controller to an actuator against a mobility-aware eavesdropper. They also studied self-organizing IoT networks that can autonomously construct and maintain connectivity to resist jamming. Nourildean and Hassib (2024) conducted an extensive validation of an SIR epidemic model to study the propagation of jamming attacks in IoT wireless networks.

From the perspectives of game theory, multi-agent reinforcement learning, and adversarial machine learning, recent studies have enriched anti-jamming theory and methods for IoT systems. Alsulami et al. (2021) investigated coalition game-theoretic defense against combined jamming and eavesdropping threats, where attackers simultaneously degrade signal quality and compromise secrecy rates. Feng et al. (2023) modeled ground user-UAV jammer interactions as a Markov game and proposed ETMAPPO, an event-triggered variant of MAPPO with Beta policies that reduces communication overhead and improves convergence efficiency compared to baseline multi-agent deep reinforcement learning (MADRL) approaches. Wang et al. (2025) explored adversarial machine learning techniques for covert communication jamming, employing gray-box attacks to craft imperceptible perturbations against DQN-based anti-jamming systems through iterative policy querying and random-walk heuristics. Nugraha et al. (2020) studied multi-agent jamming attacks through Dynamic Resilient Graph Games, analyzing strategic interactions in networked environments. While these approaches advance anti-jamming research from different angles, they lack the integrated framework combining stealthy game modeling, attention-based learning, and interpretability analysis that our FG-MAD3PG approach provides.

The fusion of game theory and deep reinforcement learning has emerged as a promising paradigm for anti-jamming and resilient networking, particularly in next-generation wireless systems. Ibrahim et al. (2024) integrated deep reinforcement learning with Bayesian game theory for dynamic packet scheduling and resource allocation in 6G environments, addressing complex multi-objective optimization challenges. Peng et al. (2025) formulated radio countermeasures as a non-zero-sum game combined with deep reinforcement learning, developing intelligent jamming schemes that optimize the trade-off between success rate and power efficiency. Xue et al. (2022) designed a potential game framework coupled with reinforcement learning to model sensor-attacker conflicts in wireless estimation systems, enabling strategic defense planning. Huang et al. (2025) applied game-theoretic modeling with deep reinforcement learning to enhance resilience in smart manufacturing systems through dynamic reconfiguration. However, these hybrid approaches primarily focus on single-agent scenarios or assume centralized coordination, lacking the distributed architecture and attention mechanisms necessary for scalable multi-agent anti-jamming defense in resource-constrained SIoT environments.

Recent advances in explainable AI have demonstrated the importance of interpretability in mission-critical systems. Rjoub et al. (2025) proposed a trust-aware scheduling framework that integrates DQN with LIME, achieving efficient task allocation while providing transparent trustworthiness assessment. Gupta and Singh (2024) developed a multilayer machine learning pipeline with SHAP-based explainability for botnet detection, successfully balancing accuracy with interpretability. While these works showcase the value of explainable AI in networked systems, they have not been applied to the specific challenges of multi-agent anti-jamming scenarios where understanding cooperative behaviors and strategic decisions is crucial for system reliability.

Despite significant contributions in anti-jamming defenses, game-theoretic modeling, and explainable AI, existing literature lacks a unified framework integrating these critical aspects. Current approaches cannot model stealthy channel control competition through formal game theory while simultaneously employing distributed multi-agent reinforcement learning in continuous action spaces and providing interpretable explanations for emergent behaviors. To address this gap, we propose a comprehensive framework that integrates the FlipIt game model with our FG-MAD3PG algorithm for effective adaptive anti-jamming defense, while leveraging LIME to reveal the underlying decision logic of learned policies. This integration delivers an anti-jamming defense mechanism that achieves superior performance while maintaining the transparency and trustworthiness essential for practical deployment in EI-enabled SIoT systems.

3. Models

3.1. System Model

Jamming attacks in EI-enabled SIoT can pose a number of threats including communication failures, service interruptions, and the erosion of social trust. As cyberattacks continue to grow in complexity, deploying AI on edge intelligence devices close to data sources for real-time jamming analysis and communication protection has become urgent. As shown in Fig. 1, the jammer targets the wireless access links between SIoT devices and the edge infrastructure over 5G, 6G or Wi-Fi, disrupting both uplink and downlink and degrading end-to-end services to and from the cloud. The combination of edge computing and AI technology allows multi-agent reinforcement learning models to be deployed on edge intelligence devices, optimizing their training and inference capabilities for intelligent anti-jamming.

We consider a heterogeneous network consisting of N edge intelligence devices and M malicious jamming nodes engaged in a continuous struggle for channel control. The jamming nodes attempt to disrupt the normal operation of the network by occupying communication channels and interfering with legitimate transmissions through continuous signal blocking. To mitigate these jamming attacks, edge intelligence devices act as defenders, employing self-adaptive learning algorithms to detect and counter malicious behaviors. The key challenge lies in the asymmetric information structure: while jammers can actively probe channel states, edge intelligence devices must make decisions based on partial observations affected by interference.

Different from traditional reactive security models, our approach enables edge intelligence devices to proactively adapt their communication strategies based on learned jamming patterns. Edge intelligence devices continuously monitor the network environment, identify jamming behaviors, and adjust their transmission parameters including power levels, channel selection, and timing. Through repeated interactions, edge intelligence devices learn to predict jammer actions and preemptively switch to safer channels or adjust transmission power to maintain communication quality. This competitive interaction continues until reaching a strategic equilibrium where neither jamming nodes nor edge intelligence devices have incentive to unilaterally deviate from their current strategies.

To this end, we model this interactive procedure on the basis of the FlipIt game, which can guide us in seeking the optimal coordinated anti-jamming strategy. The FlipIt game is particularly suitable for our scenario as it captures three essential characteristics: first, it models the stealthy nature of channel control where possession can switch between SIoT jammers and edge intelligence defenders without immediate detection; second, it incorporates the cost-benefit tradeoffs inherent in both jamming and anti-jamming operations; third, it represents the continuous and asynchronous nature of the competition, matching the persistent struggle for communication channels within the SIoT system. This game-theoretic foundation provides the basis for developing our FG-MAD3PG algorithm, where multiple edge intelligence devices learn cooperative anti-jamming strategies through distributed reinforcement learning.

3.2. Game Model

We now formally define the strategic interactions between edge intelligence defenders and SIoT jammers within the FlipIt game framework, establishing the mathematical foundation for our anti-jamming strategy optimization.

Definition 3.1 (FlipIt Anti-Jamming Game (FG-AJ))**.** *The FG-AJ for the EI-enabled SIoT is described as a ten-tuple $\mathcal{G} = (\mathcal{V}, \mathcal{W}, \mathcal{X}, \mathcal{A}, \mathcal{J}, \mathcal{O}, \psi, u, \mathcal{T}, \gamma)$, whose components are defined as follows:*

- $\mathcal{V} = \{1, 2, \ldots, N\}$ *is the set of defending edge intelligence devices.*
- $\mathcal{W} = \{1, 2, \ldots, M\}$ *is the set of SIoT jammers.*
- $\mathcal{X}$ *represents the state space characterizing channel conditions. Each state $x \in \mathcal{X}$ encodes control status and quality metrics, specifically $x = (c, q, p)$, where $c \in \{0, 1\}$ indicates defender control of the channel (1=controlled, 0=compromised), $q \in [0, 1]$ is a normalized channel-quality indicator (e.g., SINR-based), and $p \in \mathbb{R}_+$ is the aggregate interference power.*
- $\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2 \times \ldots \times \mathcal{A}_N$ *is the joint continuous action space for edge intelligence defenders. Each $\mathcal{A}_i$ is defender i's compact action space, including transmission power and channel-selection distribution.*
- $\mathcal{J} = \mathcal{J}_1 \times \mathcal{J}_2 \times \ldots \times \mathcal{J}_M$ *is the joint action space for SIoT jammers. Each $\mathcal{J}_m$ is jammer m's compact action repertoire, including jamming power levels and frequency selections.*
- $\mathcal{O} = \mathcal{O}_1 \times \mathcal{O}_2 \times \ldots \times \mathcal{O}_N$ *is the joint observation space for edge intelligence defenders. Each observation $o_i \in \mathcal{O}_i$ represents defender i's partially observable channel measurements, including RSSI, SINR, channel busy flags, ACK/NACK feedback, and neighbor summaries.*
- $\psi : \mathcal{X} \times \mathcal{A} \times \mathcal{J} \to \Delta(\mathcal{O})$ *is the observation kernel, yielding a probability distribution over edge intelligence defender observations. Here, $\Delta(\cdot)$ denotes the probability simplex over its argument.*
- $u : \mathcal{X} \times \mathcal{A} \times \mathcal{J} \to \mathbb{R}^N$ *is the per-stage utility function for edge intelligence defenders. The jammer coalition utility is $u_J = -\sum_{i=1}^{N} u_i$, establishing a zero-sum interaction between the defender and jammer coalitions.*
- $\mathcal{T} : \mathcal{X} \times \mathcal{A} \times \mathcal{J} \to \Delta(\mathcal{X})$ *is the state transition kernel, specifying probabilities of channel control transitions.*
- $\gamma \in [0, 1)$ *is the discount factor, balancing immediate rewards against long-term objectives.*

To embed FlipIt semantics into the MADRL framework, we shape the per-stage reward for each defender. The reward function is designed to balance the benefit of maintaining channel control against the costs of actions and service degradation. For defender i at time step t, its reward $r_{i,t}$ is formulated as

$$r_{i,t}(x_t, a_{i,t}) = \alpha \cdot \mathrm{SINR}_{i,t} - \beta \cdot P(a_{i,t}) - \kappa \cdot C_i^{\mathrm{flip}}(a_{i,t}) - \delta \cdot \mathbb{I}_{\mathrm{violated},t}, \tag{1}$$

where each component maps to a core concept of the FlipIt game:

- $\mathrm{SINR}_{i,t}$ is the measured signal-to-interference-plus-noise ratio, serving as a tangible proxy for channel quality and control. A higher SINR directly correlates with the *time-under-control* objective in the FlipIt game.
- $P(a_{i,t})$ is the transmit power cost induced by action $a_{i,t}$, representing the energy expenditure for maintaining or recapturing the channel.
- $C_i^{\mathrm{flip}}(a_{i,t}) \geq 0$ is an intervention cost charged when an action triggers control reacquisition (e.g., frequency hopping), modeling the cost of a *flip*.
- $\mathbb{I}_{\mathrm{violated},t}$ is an indicator function that equals 1 if a quality-of-service (QoS) violation occurs (e.g., high latency or packet loss), representing a penalty for losing channel control.
- $\alpha, \beta, \kappa, \delta > 0$ are design weights that balance these competing objectives.

The time-under-control (ToC) metric over a discrete time horizon H is defined as $\mathrm{ToC}(H) = \frac{1}{H} \sum_{t=0}^{H-1} \mathbb{I}[c_t = 1]$, where $\mathbb{I}[c_t = 1]$ is an indicator function that is 1 if the defender controls the channel at time step t.

For each edge intelligence defender $i \in \mathcal{V}$, let $\mathcal{N}_i \subseteq \mathcal{V} \setminus \{i\}$ denote its communication neighborhood, which represents the set of edge intelligence devices within its communication range. To facilitate strategic analysis, we define the strategy spaces for both coalitions. Let $\Sigma = \{\sigma : \mathcal{O} \to \Delta(\mathcal{A})\}$ denote the space of stationary mixed strategies for the edge intelligence defender coalition under partial observability, and let $\Theta = \{\tau : \mathcal{X} \to \Delta(\mathcal{J})\}$ denote the corresponding space for the SIoT jammers coalition, reflecting their superior state information through active probing capabilities.

Theorem 3.1 (Existence of stationary mixed-strategy equilibrium)**.** *The induced discounted FlipIt game admits a value and stationary mixed strategies $(\sigma^\star, \tau^\star)$ that form a Nash equilibrium, when the following conditions are satisfied: (i) $\mathcal{X}$ is a Borel space and $\mathcal{A}, \mathcal{J}$ are nonempty compact metric spaces; (ii) the utility function u is bounded and continuous in (x, a, j); (iii) the transition kernel $\mathcal{T}(\cdot|x, a, j)$ and observation kernel $\psi(\cdot|x, a, j)$ are weakly continuous in (x, a, j); and (iv) the interaction is zero-sum with jammer reward $u_J = -\sum_{i=1}^N u_i$.*

Proof sketch. We establish equilibrium existence by leveraging the FlipIt framework's structure. For a simplified finite-horizon, undiscounted version of the game, each player i chooses a strategy s_i that specifies decision parameters. The payoff function for defender i over a time horizon $[0, T]$, denoted by $U_i(s_i, s_{-i})$, is formulated as

$$U_i(s_i, s_{-i}) = \int_0^T \mathbb{I}[\text{defender } i \text{ controls channel at } t] \cdot r_i(t) dt - \sum_k c_i^{\mathrm{flip}}, \tag{2}$$

where $\mathbb{I}[\cdot]$ is the indicator function, $r_i(t)$ is the instantaneous reward rate at time t, s_{-i} denotes the strategies of all players except player i, c_i^{flip} is the cost incurred by defender i for each channel flip operation, and the summation is over the index k of all flip operations.

Under the assumption of compact strategy spaces and continuous payoff functions, Glicksberg's theorem guarantees the existence of a mixed-strategy Nash equilibrium $s^* = (s_1^*, s_2^*, \ldots)$, such that for every player i, the following condition holds for any alternative strategy $s_i' \in S_i$:

$$U_i(s_i^*, s_{-i}^*) \geq U_i(s_i', s_{-i}^*), \tag{3}$$

where S_i represents the strategy space for player i. This result for a simpler setting provides the theoretical foundation. For the full discounted stochastic extension with a zero-sum structure and weak continuity assumptions, standard results for FlipIt games over Borel spaces guarantee the existence of stationary optimal strategies and a game value. □

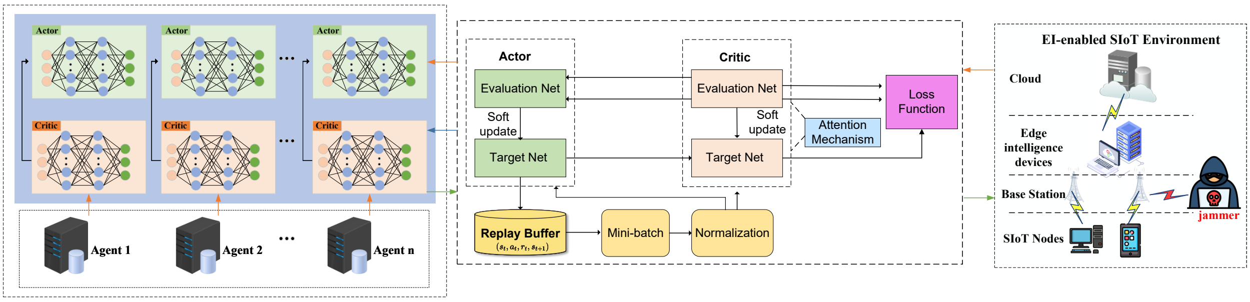


Fig. 2: Architecture of the FG-MAD3PG framework for interference mitigation in edge-enhanced SIoT environments.

The establishment of Theorem 3.1 provides essential theoretical foundations for our FG-MAD3PG-based defense architecture. It confirms that a stable strategic configuration emerges from the adversarial dynamics, ensuring that our distributed learning approach can converge to meaningful defense policies rather than oscillating indefinitely. This theoretical guarantee enables the design of our multi-agent reinforcement learning algorithm with confidence in its convergence properties under the FlipIt-guided framework, where per-agent observations $o_{i,t}$ feed the actors, centralized critics condition on compact summaries of (o_t, a_t) for centralized training with decentralized execution, and rewards r_i encode both ToC benefits and flip costs.

4. Solution

The existence of a stable equilibrium in the FG-AJ game provides theoretical guarantees, yet it does not directly yield closed-form strategies that are tractable in deployment. To bridge this gap, we develop the FG-MAD3PG algorithm that learns equilibrium strategies through distributed reinforcement learning and augments them with post-hoc explainability via LIME.

4.1. Integration of FlipIt Game with FG-MAD3PG

The FlipIt game maps naturally to our reinforcement learning formulation through three aspects that we describe in turn. First, to capture stealth, each agent observation o_i is a noisy channel measurement that may lag the true control state. Second, to encode both overt defenses such as power adaptation and proactive strategies such as frequency hopping, the action a_i is continuous. Third, to reflect asynchronous feedback, the reward combines immediate terms such as power cost and delayed terms such as channel control, which is realized only after acknowledgments.

4.2. FG-MAD3PG Architecture

We consider N defending edge devices indexed by $i \in \{1, \ldots, N\}$. At each time step t, agent i receives a local observation o_i^t from its observation space O_i and selects a continuous action a_i^t from its action space $\mathcal{A}_i$. The architecture in Fig. 2 has three components that work together in a distributed manner, and we introduce them in order so the notation is consistent.

Each agent i maintains a deterministic policy μ_i, which maps its local observation o_i^t to an action a_i^t. This relationship is given by

$$a_i^t = \mu_i(o_i^t; \theta_i^\mu). \tag{4}$$

Here, o_i^t is the current observation of agent i, $\mu_i(\cdot)$ is the actor (policy) network of agent i, and θ_i^μ denotes the parameters of μ_i.

Each agent also maintains a local critic Q_i, which estimates the action-value based on the local pair (o_i, a_i) and an attention-derived context vector h_i aggregated from its neighbors $\mathcal{N}_i \subseteq \{1, \ldots, N\}$. The critic's output is represented as

$$Q_i(o_i, a_i, h_i; \theta_i^Q). \tag{5}$$

Here, Q_i is the critic network of agent i, θ_i^Q are its parameters, and $\mathcal{N}_i$ is the set of neighbor indices for agent i.

To compute the attention context h_i, we first produce embeddings e_i for the agent's own information tuple (o_i, a_i) and e_j for each neighbor's tuple (o_j, a_j). These are generated using embedding networks with shared parameters, according to the functions

$$e_i = f_{\text{self}}(o_i, a_i; \theta_{\text{emb}}) \in \mathbb{R}^{d_e}, \tag{6}$$

$$e_j = f_{\text{nbr}}(o_j, a_j; \theta_{\text{emb}}) \in \mathbb{R}^{d_e}, \quad j \in \mathcal{N}_i. \tag{7}$$

Here, e_i and e_j are the embedding vectors for agent i and neighbor j respectively, f_{self} and f_{nbr} are the embedding functions with shared parameters θ_{emb}, and d_e is the embedding dimension.

Next, we compute raw attention scores α_{ij} for each neighbor $j \in \mathcal{N}_i$ using scaled dot-product attention, formulated as

$$\alpha_{ij} = \frac{(\mathbf{W}_q e_i)^\top (\mathbf{W}_k e_j)}{\sqrt{d_k}}. \tag{8}$$

Here, $\mathbf{W}_q \in \mathbb{R}^{d_k \times d_e}$ and $\mathbf{W}_k \in \mathbb{R}^{d_k \times d_e}$ are the learnable projection matrices for the query and key, respectively, and d_k is the key dimension.

These scores are then normalized via a softmax function to obtain the final attention weights β_{ij}, computed by

$$\beta_{ij} = \frac{\exp(\alpha_{ij})}{\sum_{l \in \mathcal{N}_i} \exp(\alpha_{il})}. \tag{9}$$

Here, the denominator represents the sum of exponentiated scores over all neighbors l in the set $\mathcal{N}_i$.

Finally, the context vector h_i is obtained as a weighted sum of value projections from the neighbors, given by

$$h_i = \sum_{j \in \mathcal{N}_i} \beta_{ij} \mathbf{W}_v e_j. \tag{10}$$

Here, $\mathbf{W}_v \in \mathbb{R}^{d_k \times d_e}$ is the learnable projection matrix for the value.

The training protocol is distributed to limit overhead and improve scalability. Each critic exchanges information only within its neighborhood $\mathcal{N}_i$, which reduces communication load and removes the need for a central coordinator.

4.3. Learning Objectives and Update Rules

We maintain a distributed replay buffer $\mathcal{D}$ that stores transitions $(o^t, a^t, r^t, o^{t+1}, d^t)$. In each transition, o^t collects the joint observations $(o_1^t, \ldots, o_N^t)$, a^t collects the joint actions $(a_1^t, \ldots, a_N^t)$, $r^t = \{r_i^t\}_{i=1}^N$ are the per-agent rewards computed using Eq. (1), o^{t+1} is the next joint observation, and $d^t \in \{0, 1\}$ is a flag indicating episode termination at time t.

For agent i, the critic parameters θ_i^Q are updated by minimizing the mean squared temporal difference (TD) error, defined by the loss function $\mathcal{L}_i(\theta_i^Q)$ as

$$\mathcal{L}_i(\theta_i^Q) = \mathbb{E}_{(o,a,r,o',d)\sim\mathcal{D}} \left[\left(y_i - Q_i(o_i, a_i, h_i; \theta_i^Q) \right)^2 \right]. \tag{11}$$

Here, the expectation is taken over a mini-batch of transitions sampled from the replay buffer $\mathcal{D}$, and y_i is the TD target. The context h_i is computed from the neighbor tuples within the same mini-batch.

The TD target y_i uses target networks for stability and is defined as

$$y_i = r_i + \gamma(1 - d)\, Q_i'(o_i', \mu_i'(o_i'), h_i'; \theta_i^{Q'}). \tag{12}$$

Here, r_i is the reward for agent i, γ is the discount factor, d is the termination flag, o_i' is the next observation, h_i' is the next-step attention context, Q_i' is the target critic with parameters $\theta_i^{Q'}$, and μ_i' is the target actor.

We update the target networks using Polyak averaging to improve stability, according to the rules

$$\theta_i^{Q'} \leftarrow \tau\, \theta_i^Q + (1 - \tau)\, \theta_i^{Q'}, \qquad \theta_i^{\mu'} \leftarrow \tau\, \theta_i^{\mu} + (1 - \tau)\, \theta_i^{\mu'}. \tag{13}$$

Algorithm 1 FG-MAD3PG with LIME-based Explainability for EI-enabled SIoT Anti-Jamming Defense

```
Initialize: Actor networks $\{\mu_i\}_{i=1}^N$, Critic networks $\{Q_i\}_{i=1}^N$, Target networks $\{\mu_i', Q_i'\}_{i=1}^N$;
Initialize: Attention modules, Replay buffer $\mathcal{D}$, LIME explainer $\mathcal{E}$, Update interval $U$, Minimum buffer size $B_{min}$;
for episode = 1 to $E$ do
    Reset environment and obtain initial observations $\{o_i^0\}_{i=1}^N$;
    for $t = 0$ to $T - 1$ do
        for each agent $i \in \{1, \dots, N\}$ do
            Select action $a_i^t = \mu_i(o_i^t) + \epsilon_t$;
        end for
        Execute joint action $a^t$ and observe rewards $r^t$, next observations $o^{t+1}$, and done flag $d^t$;
        Store transition $(o^t, a^t, r^t, o^{t+1}, d^t)$ in $\mathcal{D}$;
        if $t$ mod $U = 0$ and $|\mathcal{D}| \geq B_{min}$ then
            Sample minibatch $\mathcal{B}$ from $\mathcal{D}$;
            for each agent $i$ do
                Compute attention context $h_i$ from neighborhood tuples in $\mathcal{B}$;
                Update critic by minimizing $\mathcal{L}_i(\theta_i^Q)$;
                Update actor using the deterministic policy gradient;
                Softly update target networks with rate $\tau$;
            end for
        end if
    end for
end for
Post-training Analysis:
Generate LIME explanations $\{\xi_i\} = \mathcal{E}(\{\mu_i\}, \mathcal{D}_{test})$ on test scenarios;
Aggregate and analyze feature-importance scores;
Output: Trained policies $\{\mu_i\}_{i=1}^N$ and explanation sets $\{\xi_i\}$.
```

Here, $\tau \in (0, 1)$ is the soft update rate, while θ_i^μ and θ_i^Q are the current actor and critic parameters, and $\theta_i^{\mu'}$ and $\theta_i^{Q'}$ are their corresponding target network parameters.

The actor parameters θ_i^μ are optimized using the deterministic policy gradient. The gradient of the expected return J with respect to θ_i^μ is

$$\nabla_{\theta_i^\mu} J = \mathbb{E}_{o\sim\mathcal{D}}\left[\nabla_{a_i} Q_i(o_i, a_i, h_i)\Big|_{a_i=\mu_i(o_i)} \nabla_{\theta_i^\mu}\mu_i(o_i)\right]. \tag{14}$$

Here, the expectation is approximated over a mini-batch of observations, and the inner gradient $\nabla_{a_i} Q_i$ is the gradient of the critic's output with respect to the action a_i, evaluated at the action selected by the current policy, $a_i = \mu_i(o_i)$.

4.4. LIME Integration for Explainability

We integrate LIME for post-hoc inspection so that local decisions can be explained without altering the learned policies. The explainer $\mathcal{E}$ samples perturbed observations near the current input, queries the policy, and fits a local linear surrogate to reveal feature importance. Algorithm 1 presents the complete procedure.

Here, E is the total number of training episodes, T is the maximum number of steps per episode, ϵ_t is the exploration noise added at time t, and ξ_i denotes the vector of local feature-importance weights returned by the LIME explainer $\mathcal{E}$ for agent i.

4.5. Convergence and Stability Properties

The FG-MAD3PG algorithm inherits convergence properties of actor–critic methods and is designed to approach a local Nash equilibrium of the FG-AJ game. Under standard conditions including Lipschitz-continuous function approximators, bounded rewards, and sufficiently rich exploration, the contraction of the Bellman operator for fixed

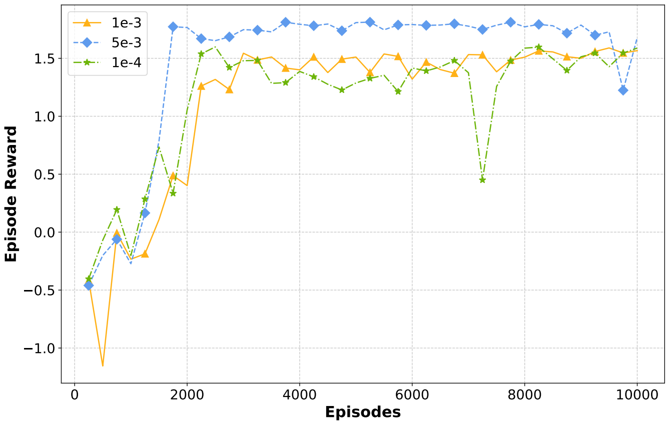


Fig. 3: Impact of learning rate on episode reward based on FG-MAD3PG.

policies and the deterministic policy gradient theorem support convergence to stable policies. The attention mechanism further reduces gradient variance by assigning larger weights to informative neighbors, which improves sample efficiency and helps track nonstationary jamming. The distributed design scales to large networks, preserves privacy by restricting communication to neighborhoods, and remains robust when a subset of agents fails.

5. Experimental Analyses

We develop the FG-MAD3PG algorithm utilizing Python to conduct experimental episodes for resisting jamming attacks via game-theoretic anti-jamming strategies in EI-enabled SIoT systems based on the PyTorch framework. We execute $10,000$ episodes and mark a point every 25 episodes, visualizing the preferred parameter settings and demonstrating the superiority of the proposed FG-MAD3PG approach. Our experimental evaluation was conducted using Python 3.8.10 on an Ubuntu 20.04 system, utilizing PyTorch 1.13.0 for deep reinforcement learning implementation. We modified the multi-agent environment framework[1] to suit our research needs, extending it with custom modifications to accommodate our SIoT anti-jamming environment and explainable reinforcement learning requirements. To handle the computational requirements of MADRL, we employed a high-performance workstation. The empirical evaluation comprised $10,000$ distinct simulation episodes, ensuring statistical significance and robust performance assessment across various anti-jamming configurations and attack scenarios.

5.1. Hyperparameter Sensitivity Analysis

This section investigates how critical hyperparameters affect the performance of the FG-MAD3PG algorithm in EI-enabled SIoT anti-jamming scenarios.

5.1.1. Impact of Learning Rate

Fig. 3 shows the evolution of the average episode reward during training. The aggressive learning rate of $5e-3$ achieves the highest asymptotic performance, reaching approximately 1.7 by around 2000 episodes. The episode rewards for $1e-3$ and $1e-4$ are very similar, both eventually stabilizing at about 1.5. However, the $1e-4$ run exhibits pronounced instability around 7000 episodes, with a sharp and notably poor performance drop. Therefore, for episode reward, $5e-3$ is the best choice, and $1e-3$ is preferred over $1e-4$ due to better stability.

Fig. 4 illustrates the anti-jamming success rate evolution under learning rates of $1e-4$, $1e-3$, and $5e-3$. The aggressive rate $5e-3$ demonstrates rapid convergence, reaching approximately 0.67 by 2000 episodes and maintaining stable performance throughout training. The moderate rate $1e-3$ shows similar final performance, stabilizing near 0.66 after initial fluctuations. The conservative rate $1e-4$ exhibits the most volatile behavior, with a significant performance drop around 7000 episodes before recovering. Therefore, in terms of anti-jamming success rate, $5e-3$ is the preferred configuration due to its superior stability and convergence speed.

[1] `https://github.com/xlws1/MTD-MA3C`

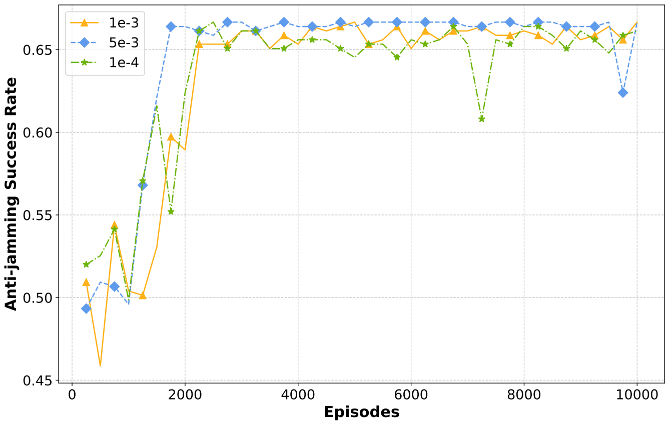


Fig. 4: Impact of learning rate on anti-jamming success rate on FG-MAD3PG.

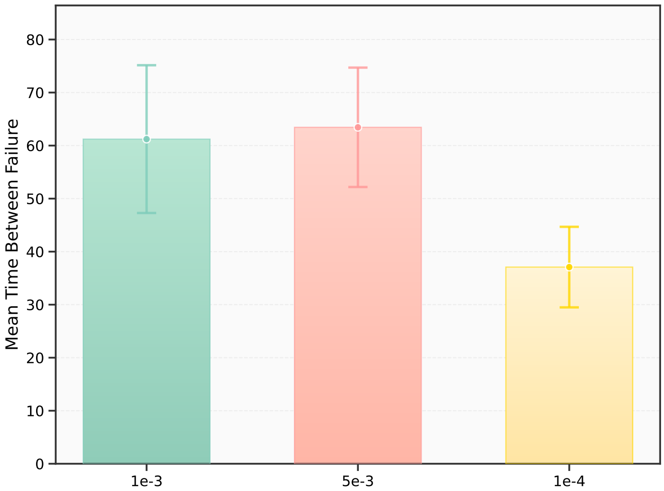


Fig. 5: Impact of learning rate on mean time between failure during anti-jamming operations on FG-MAD3PG.

Fig. 5 presents the mean time between failure (MTBF) for different learning rates, a key reliability metric in EI-enabled SIoT anti-jamming systems. The conservative learning rate of $1e-4$ yields the lowest MTBF at roughly 38. Increasing the rate to $1e-3$ improves reliability to about 50. The aggressive rate of $5e-3$ provides the highest MTBF, around 52. Therefore, for MTBF, $5e-3$ is the best configuration.

Overall, considering convergence speed, stability, and reliability, $5e-3$ emerges as the optimal learning rate, achieving the highest episode reward, best anti-jamming success rate, and longest MTBF.

5.1.2. Impact of Exploration Rate

Fig. 6 shows the average episode reward evolution for exploration rates of $\epsilon = 0.1$, 0.2, and 0.5. All three rates exhibit a steep learning curve, reaching a high reward plateau of approximately 1.8 within the first few thousand episodes. The conservative rate $\epsilon = 0.1$ reaches the plateau the fastest. The highest exploration rate $\epsilon = 0.5$ converges slightly slower initially but maintains the most stable rewards thereafter. In contrast, $\epsilon = 0.2$ attains a similar peak but suffers a noticeable drop near 9600 episodes before recovering.

Fig. 7 displays the anti-jamming success rate for exploration rates ϵ of 0.1, 0.2, and 0.5. All three settings approach a final success rate around 0.67. The highest exploration rate $\epsilon = 0.5$ attains and maintains the highest and most stable success rate. The conservative rate $\epsilon = 0.1$ is slightly lower near the end of training. The balanced rate $\epsilon = 0.2$ exhibits a marked performance collapse around 9600 episodes before returning to the plateau.

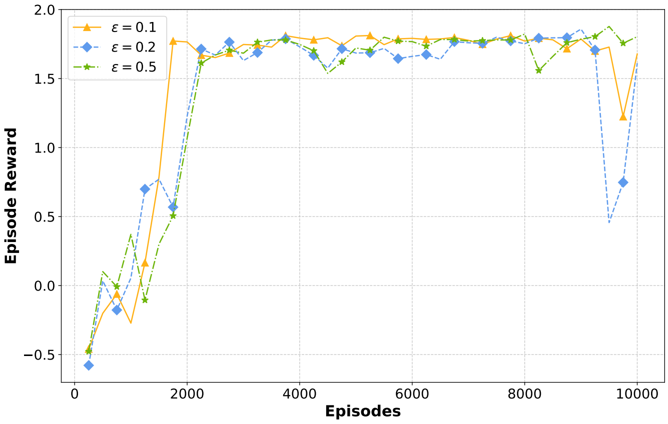


Fig. 6: Impact of exploration rate ϵ on episode reward based on FG-MAD3PG.

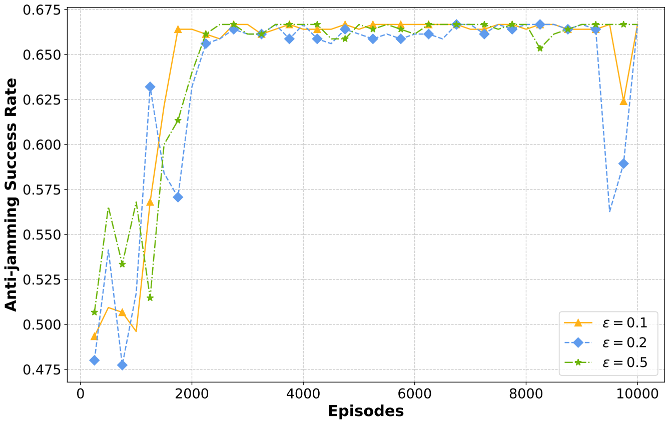


Fig. 7: Impact of exploration rate ϵ on anti-jamming success rate based on FG-MAD3PG.

Fig. 8 presents the MTBF for different exploration rates. Conservative exploration with $\epsilon = 0.1$ achieves the highest reliability with MTBF around 63. Both $\epsilon = 0.2$ and $\epsilon = 0.5$ show moderate reliability near 40. Therefore, for reliability, $\epsilon = 0.1$ is the best choice.

Overall, taking the three metrics jointly, $\epsilon = 0.1$ offers the best overall trade-off: it reaches the reward plateau quickly and achieves a final success rate comparable to the other settings, while delivering the highest MTBF.

5.2. *Comparison with Baseline Algorithms*

After evaluating hyperparameter influences on FG-MAD3PG, this section presents comparative performance analysis against established MADRL algorithms, namely D3QN (Shen et al., 2024b) and MAPPO (Jiang et al., 2024).

Fig. 9 compares average episode rewards of FG-MAD3PG, D3QN, and MAPPO over 10,000 training episodes. MAPPO exhibits the poorest performance, with rewards consistently remaining around −1.0 throughout training. D3QN achieves moderate performance, quickly reaching and stabilizing between 0.2 and 0.5. FG-MAD3PG improves rapidly within the first 2000 episodes and then maintains the highest average reward near 1.5, significantly outperforming both baseline algorithms. Therefore, FG-MAD3PG demonstrates superior learning efficiency and performance.

Fig. 10 shows anti-jamming success rates during training. MAPPO maintains poor performance with success rates around 0.32–0.37. D3QN exhibits a transient early spike but quickly drops and remains around 0.35–0.38. In contrast, FG-MAD3PG rises to a plateau near 0.65–0.67 by roughly 2000 episodes, with only a brief dip later in

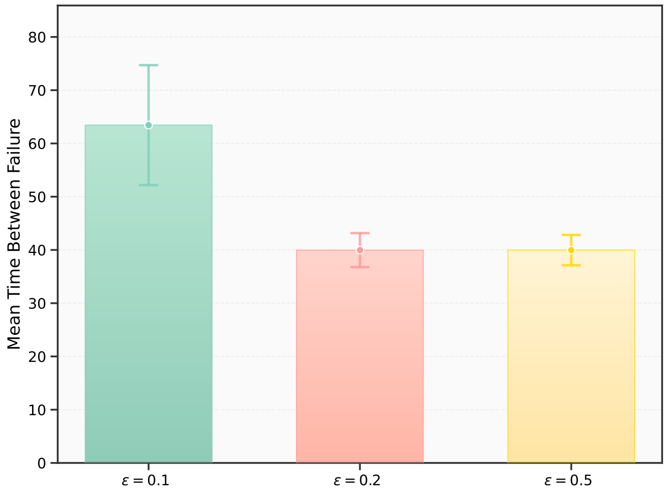


Fig. 8: Impact of exploration rate ϵ on mean time between failure on FG-MAD3PG.

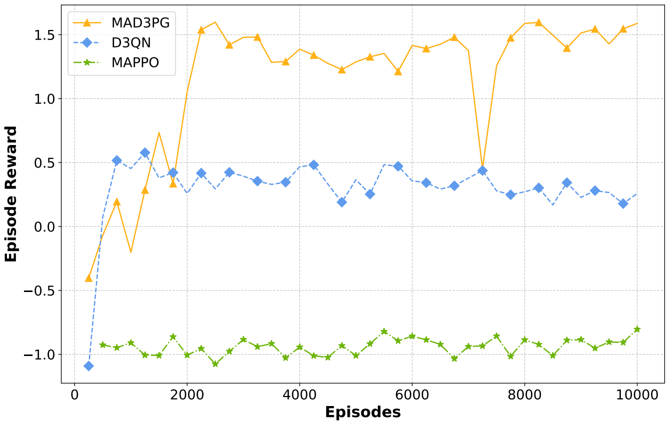


Fig. 9: Comparison of episode reward among FG-MAD3PG, D3QN, and MAPPO.

training, demonstrating clear superiority in anti-jamming effectiveness. Thus, FG-MAD3PG achieves the best anti-jamming success rate among all algorithms.

Fig. 11 compares the MTBF across algorithms. FG-MAD3PG achieves the highest reliability with MTBF near 37. D3QN demonstrates moderate reliability around 25, while MAPPO shows the lowest reliability near 15. These results indicate that FG-MAD3PG not only delivers superior performance but also maintains better system reliability under jamming attacks. Therefore, FG-MAD3PG provides the best reliability metric.

Overall, FG-MAD3PG significantly outperforms both D3QN and MAPPO across all evaluation metrics. It achieves approximately three times higher episode rewards than D3QN and nearly double the anti-jamming success rate. The superior MTBF further confirms FG-MAD3PG's robustness and reliability in defending against jamming attacks. These results validate the effectiveness of integrating factor graphs with attention mechanisms for distributed anti-jamming defense in EI-enabled SIoT systems.

5.3. *LIME-based Interpretability Analysis*

To understand learned anti-jamming strategies, we apply LIME analysis to interpret agent decision-making processes.

Fig. 12 presents LIME feature importance scores for edge intelligence device 1. The analysis reveals that self channel frequency shows the highest positive importance at 0.067, followed by jammer jamming power level at 0.047.

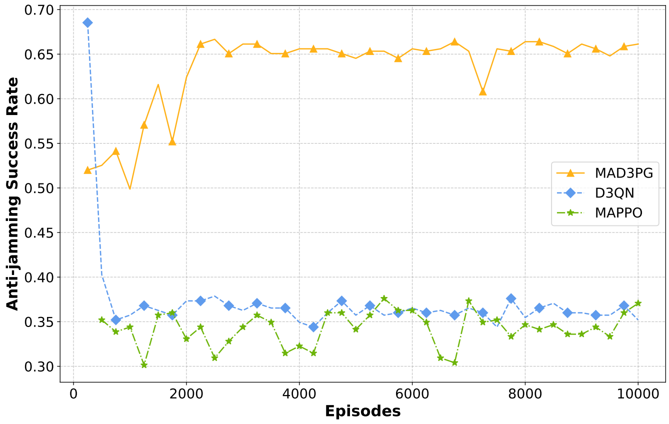


Fig. 10: Comparison of anti-jamming success rates among FG-MAD3PG, D3QN, and MAPPO.

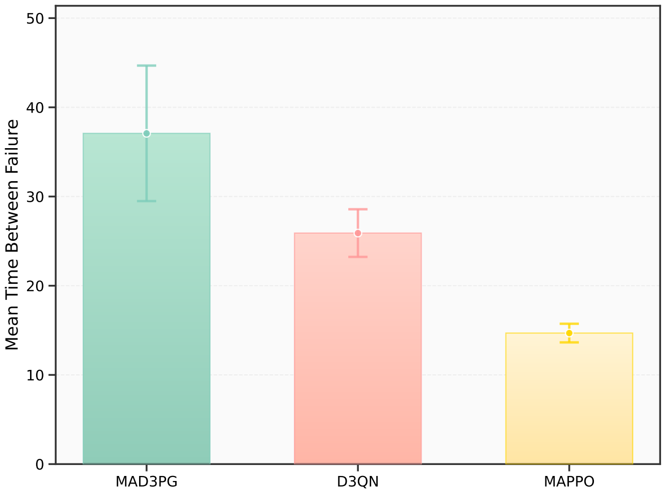


Fig. 11: Comparison of mean time between failure among FG-MAD3PG, D3QN, and MAPPO.

Jammer interference intensity exhibits positive influence at 0.020, while jammer distance shows minimal negative impact at -0.002. Negative importance factors include self channel quality score at -0.155 and self node mobility at -0.089. This pattern indicates a proactive channel hopping strategy where the device prioritizes frequency selection and responds to jamming power levels while considering channel quality degradation as a trigger for defensive actions.

Fig. 13 shows LIME analysis for edge intelligence device 2. Signal strength demonstrates the highest positive importance at 0.016, with detected jamming frequency at 0.014. Jammer distance shows positive influence at 0.009, while relative distance contributes 0.004. Negative factors include interference intensity at -0.035 and channel status at -0.019. This pattern reveals an adaptive power control strategy where the device emphasizes signal strength maintenance and spatial awareness, adjusting transmission parameters based on jamming conditions and distance metrics.

Through comprehensive LIME analysis across all agents, we identify three distinct defensive strategies learned by FG-MAD3PG. First, proactive channel hopping emerges when agents prioritize frequency selection and jamming power monitoring, using channel quality degradation as triggers for preemptive switching. Second, collaborative spectrum sharing manifests through attention mechanisms that weight neighbor channel status and communication patterns. Third, adaptive power control appears when agents emphasize signal strength maintenance and spatial relationships, dynamically adjusting transmission power based on interference levels and distance metrics. These

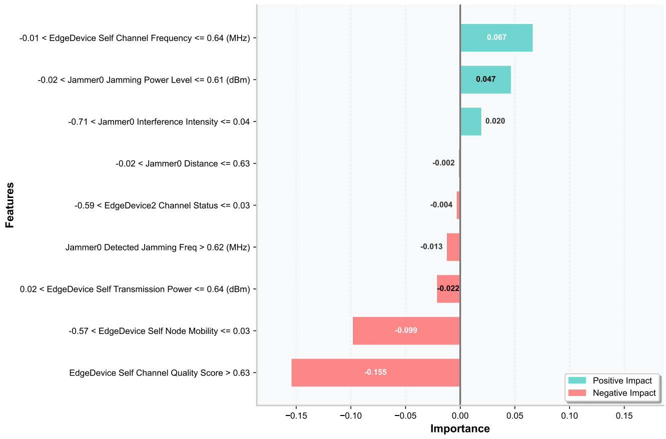


Fig. 12: LIME feature importance for edge intelligence device 1 showing proactive channel hopping strategy.

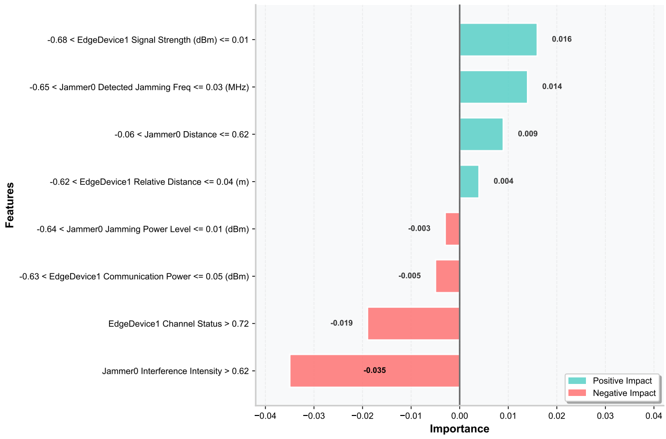


Fig. 13: LIME feature importance for edge intelligence device 2 showing adaptive power control strategy.

interpretable strategies validate that FG-MAD3PG learns meaningful anti-jamming behaviors rather than exploiting spurious correlations.

6. Conclusion

In this paper, we have proposed and analyzed the game-theoretic strategy assisted by FG-MAD3PG approach for defensively mitigating jamming attacks in EI-enabled SIoT systems. In our framework, edge intelligence devices and jammers are two players in a FlipIt-oriented game model, in which edge intelligence devices proactively defend against jammers by learning optimal channel control strategies, making it possible to understand the attack patterns of jammers and protect the communication channels in EI-enabled SIoT systems. Note that we have taken attention mechanisms and distributed training into account to achieve more accurate threat perception and prediction based on multi-agent reinforcement learning. Integrating explainable AI techniques into traditional MADRL, we have next built the corresponding FG-MAD3PG based on the proposed FG-AJ game to self-adaptively pursue the optimal anti-jamming defense strategy in practice. Our experimental evaluation reveals that the learning rate and exploration rate are critical hyperparameters affecting decision-making performance. It further demonstrates that FG-MAD3PG-aided edge intelligence devices significantly outperform D3QN- and MAPPO-aided counterparts in terms of anti-jamming success rate within EI-enabled SIoT systems.

Data availability

Data will be made available on request.

Acknowledgements

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

References

Alsulami, B.S., Bajracharya, C., Rawat, D.B., 2021. Game theory-based attack and defense analysis in virtual wireless networks with jammers and eavesdroppers. Digital Communications and Networks 7, 327–334.

Amanatidis, P., Karampatzakis, D., Michailidis, G., Lagkas, T., Iosifidis, G., 2024. Adaptive reverse task offloading in edge computing for AI processes. Computer Networks 255, 110844.

Amin, A., Chimba, D., Hasan, K., 2025. Integrating AI and edge computing for advanced safety at railroad grade crossings. Journal of Rail Transport Planning & Management 33, 100501.

Di Bonito, L.P., Campanile, L., Iacono, M., Di Natale, F., 2025. An explainable artificial intelligence framework to predict marine scrubbers performances. Engineering Applications of Artificial Intelligence 160, 111860.

Fan, D., Shen, H., Dong, L., 2021. Multi-agent distributed deep deterministic policy gradient for partially observable tracking. Actuators 10, 268.

Feng, Z., Huang, M., Wu, Y., Wu, D., Cao, J., Korovin, I., Gorbachev, S., Gorbacheva, N., 2023. Approximating Nash equilibrium for anti-UAV jamming Markov game using a novel event-triggered multi-agent reinforcement learning. Neural Networks 161, 330–342.

Gupta, S., Singh, B., 2024. An intelligent multi-layer framework with SHAP integration for botnet detection and classification. Computers & Security 140, 103783.

Huang, S., Mo, G., Jing, S., Leng, J., Li, X., Gu, X., Yan, Y., Wang, G., 2025. Digital twin-driven self-adaptive reconfiguration planning method of smart manufacturing systems using game theory and deep Q-network for Industry 5.0. Journal of Industrial Information Integration 47, 100901.

Ibrahim, A., Chen, Z., Eljailany, H.A., Yu, G., Ipaye, A.A., Abouda, K.A., Idress, W.M., 2024. Advancing 6G-IoT networks: Willow catkin packet transmission scheduling with AI and Bayesian game-theoretic approach-based resource allocation. Internet of Things 25, 101119.

Jang, H., Yoon, B., 2025. An explainable artificial intelligence – human collaborative model for investigating patent novelty. Engineering Applications of Artificial Intelligence 154, 110984.

Jiang, B., Du, J., Jiang, C., Han, Z., Debbah, M., 2024. Underwater searching and multiround data collection via AUV swarms: An energy-efficient AoI-aware MAPPO approach. IEEE Internet of Things Journal 11, 12768–12782.

Jiang, Q., Zeng, H., 2025. Novel meta learning-artificial intelligence model for systematically detecting potential pan cancer-related targets. Engineering Applications of Artificial Intelligence 160, 111976.

Jin, H., Li, Z., Zou, D., Yuan, B., 2021. DSEOM: A framework for dynamic security evaluation and optimization of MTD in container-based cloud. IEEE Transactions on Dependable and Secure Computing 18, 1125–1136.

Kandpal, S., Anas, M., Yadav, S., Geed, S.R., Kamsonlian, S., Sawarkar, A.N., 2025. Machine learning models for co-gasification of petcoke with biomass/coal: A comparative analysis of bagging, boosting, and neural networks with model interpretation using Shapley additive and local interpretable model-agnostic explanations. Journal of Environmental Chemical Engineering 13, 118206.

Kil, H.J., Kim, J.H., Lee, K., Kang, T.U., Yoo, J.H., ho Lee, Y., Park, J.W., 2024. A self-powered and supercapacitive microneedle continuous glucose monitoring system with a wide range of glucose detection capabilities. Biosensors and Bioelectronics 257, 116297.

Li, K., Jiu, B., Pu, W., Liu, H., Peng, X., 2022. Neural fictitious self-play for radar antijamming dynamic game with imperfect information. IEEE Transactions on Aerospace and Electronic Systems 58, 5533–5547.

Lin, W., Zhu, M., Zhou, X., Zhang, R., Zhao, X., Shen, S., Sun, L., 2024. A deep neural collaborative filtering based service recommendation method with multi-source data for smart cloud-edge collaboration applications. Tsinghua Science and Technology 29, 897–910.

Mir, A.N., Rizvi, D.R., 2025. Advancements in deep learning and explainable artificial intelligence for enhanced medical image analysis: A comprehensive survey and future directions. Engineering Applications of Artificial Intelligence 158, 111413.

Nourildean, S.W., Hassib, M.D., 2024. IoT-based MANET performance improvement against jamming attackers in different mobile applications. e-Prime - Advances in Electrical Engineering, Electronics and Energy 8, 100615.

Nugraha, Y., Cetinkaya, A., Hayakawa, T., Ishii, H., Zhu, Q., 2020. Dynamic resilient graph games for state-dependent jamming attacks analysis on multi-agent systems. IFAC-PapersOnLine 53, 3421–3426. 21st IFAC World Congress.

Peng, X., Xu, H., Qi, Z., Wang, D., Zhang, Y., Pang, Y., 2025. Energy-efficient strategy generation for smart jammer in non-zero-sum games: A deep reinforcement learning approach. Computer Networks 270, 111520.

Pinto, R.P., Silva, B.M., Inácio, P.R., 2025. Federated learning for anomaly detection on Internet of Medical Things: A survey. Internet of Things 33, 101677.

Qiao, T., Cao, Y., Tang, J., Zhao, N., Wong, K.K., 2022. IRS-aided uplink security enhancement via energy-harvesting jammer. IEEE Transactions on Communications 70, 8286–8297.

Rahmat, F., Zulkafli, Z., Ishak, A.J., Abdul Rahman, R.Z., Tahir, W., Ab Rahman, J., Jayaramu, V., De Stercke, S., Ibrahim, S., Ismail, M., 2025. Interpretable spatio-temporal prediction using deep neural network - local interpretable model-agnostic explanations: A case study on leptospirosis outbreaks in Malaysia. Engineering Applications of Artificial Intelligence 151, 110665.

Rjoub, G., Elmekki, H., Bentahar, J., Pedrycz, W., Kassaymeh, S., Almobydeen, S.B., Dssouli, R., 2025. Enhanced dynamic deep Q-network for federated learning scheduling policies on IoT devices using explanation-driven trust. Knowledge-Based Systems 318, 113574.

Savva, M., Ioannou, I., Vassiliou, V., 2025. Fuzzy logic-based IDS (FLIDS) for the detection of different types of jamming attacks in IoT networks. Computer Communications 241, 108251.

van 't Schip, M., 2025. The Internet of Forgotten Things: European cybersecurity regulation and the cessation of Internet of Things manufacturers. Computer Law & Security Review 57, 106152.

Schultz, J.D., Parker, K.A., Sbaiti, B., Beratan, D.N., 2025. Using machine learning to map simulated noisy and laser-limited multidimensional spectra to molecular electronic couplings. Digital Discovery 4, 1912–1924.

Shen, D., Huo, Y., Gao, Q., 2022. A friendly jamming handover scheme for a conscious mobile eavesdropper in IoT systems. Procedia Computer Science 202, 15–20. International Conference on Identification, Information and Knowledge in the internet of Things, 2021.

Shen, S., Cai, C., Li, Z., Shen, Y., Wu, G., Yu, S., 2024a. Deep Q-network-based heuristic intrusion detection against edge-based SIoT zero-day attacks. Applied Soft Computing 150, 111080.

Shen, S., Cai, C., Shen, Y., Wu, X., Ke, W., Yu, S., 2024b. MFGD3QN: Enhancing edge intelligence defense against DDoS with mean-field games and dueling double deep Q-network. IEEE Internet of Things Journal 11, 23931–23945.

Shen, S., Cai, C., Shen, Y., Wu, X., Ke, W., Yu, S., 2025a. Joint mean-field game and multiagent asynchronous advantage actor-critic for edge intelligence-based IoT malware propagation defense. IEEE Transactions on Dependable and Secure Computing 22, 3824–3838.

Shen, S., Hong, T., Shen, Y., Wu, X., Dong, J., Wu, J., 2025b. RMAAC: Joint Markov games and robust multiagent actor-critic for explainable malware defense in social IoT. IEEE Transactions on Dependable and Secure Computing Early Access, https://doi.org/10.1109/TDSC.2025.3595097.

Shen, S., Wu, X., Sun, P., Zhou, H., Wu, Z., Yu, S., 2023a. Optimal privacy preservation strategies with signaling Q-learning for edge-computing-based IoT resource grant systems. Expert Systems with Applications 225, 120192.

Shen, S., Xie, L., Zhang, Y., Wu, G., Zhang, H., Yu, S., 2023b. Joint differential game and double deep Q-networks for suppressing malware spread in industrial Internet of Things. IEEE Transactions on Information Forensics and Security 18, 5302–5315.

Suwattananuruk, B., Chien, C.F., 2025. Denoising variational autoencoders for smart inspection of wafer probe card PCB channels for advancing quality control for semiconductor manufacturing. Computers & Industrial Engineering 209, 111453.

Vieth, J., Westphal, J., Speerforck, A., 2025. District heating network topology optimization and optimal co-planning using dynamic simulations. Advances in Applied Energy 19, 100233.

Wang, T., Niu, Y., Zhou, Z., 2025. Adversarial attacks against intelligent anti-jamming communication: An adaptive gray-box attack method. Physical Communication 72, 102716.

Wu, G., Wang, H., Zhang, H., Zhao, Y., Yu, S., Shen, S., 2023a. Computation offloading method using stochastic games for software-defined-network-based multiagent mobile edge computing. IEEE Internet of Things Journal 10, 17620–17634.

Wu, G., Xie, L., Zhang, H., Wang, J., Shen, S., Yu, S., 2023b. STSIR: An individual-group game-based model for disclosing virus spread in social Internet of Things. Journal of Network and Computer Applications 214, 103608.

Wu, G., Xu, Z., Zhang, H., Shen, S., Yu, S., 2023c. Multi-agent DRL for joint completion delay and energy consumption with queuing theory in MEC-based IIoT. Journal of Parallel and Distributed Computing 176, 80–94.

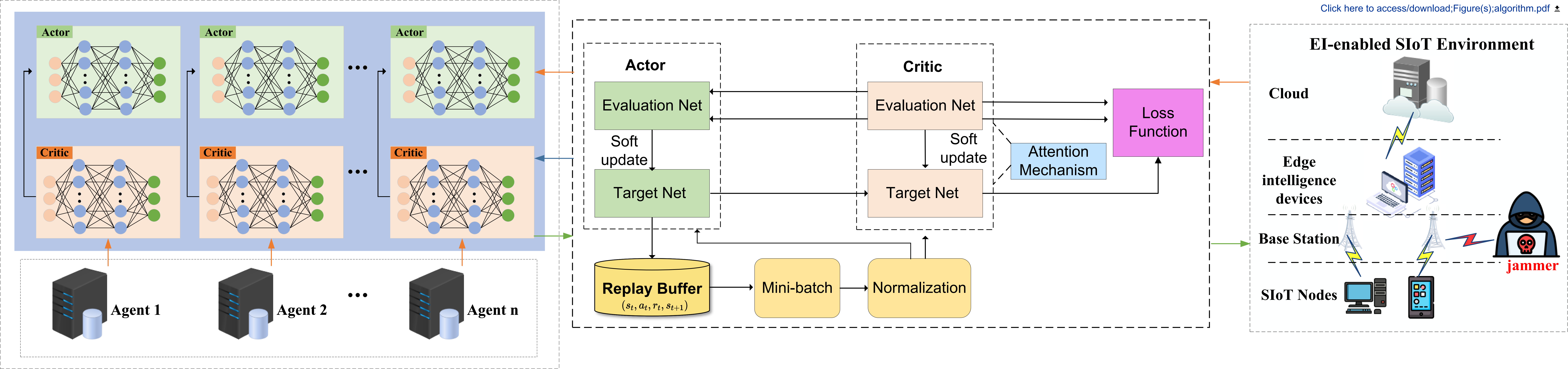
Xue, L., Ma, B., Liu, J., Yu, Y., 2022. Jamming attack against remote state estimation over multiple wireless channels: A reinforcement learning based game theoretical approach. ISA Transactions 130, 1–9.

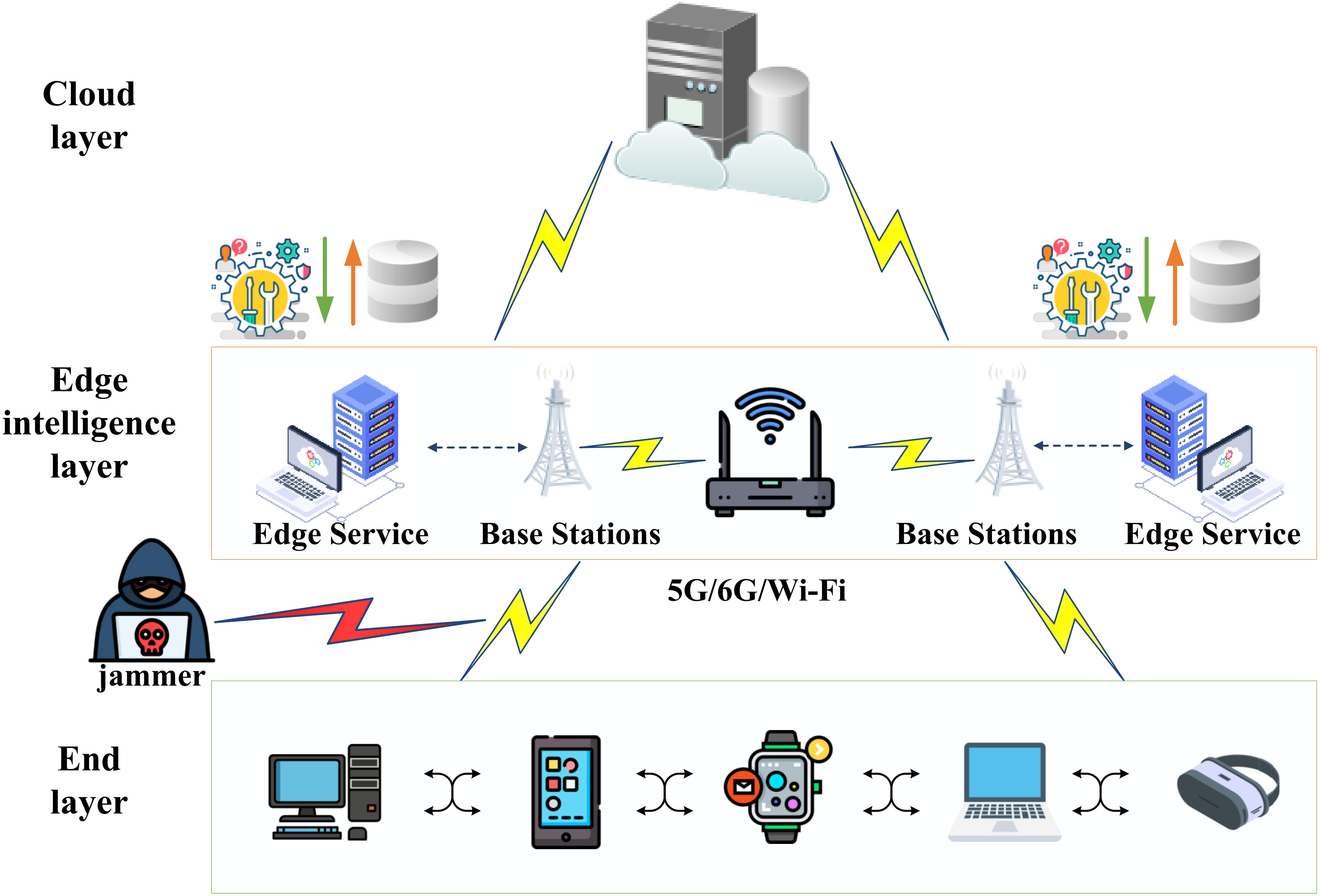
Yao, X., Yang, H., Zeng, W., 2025. Lightweight physical-layer authentication for IoT devices access against jamming attacks. Physical Communication 72, 102787.

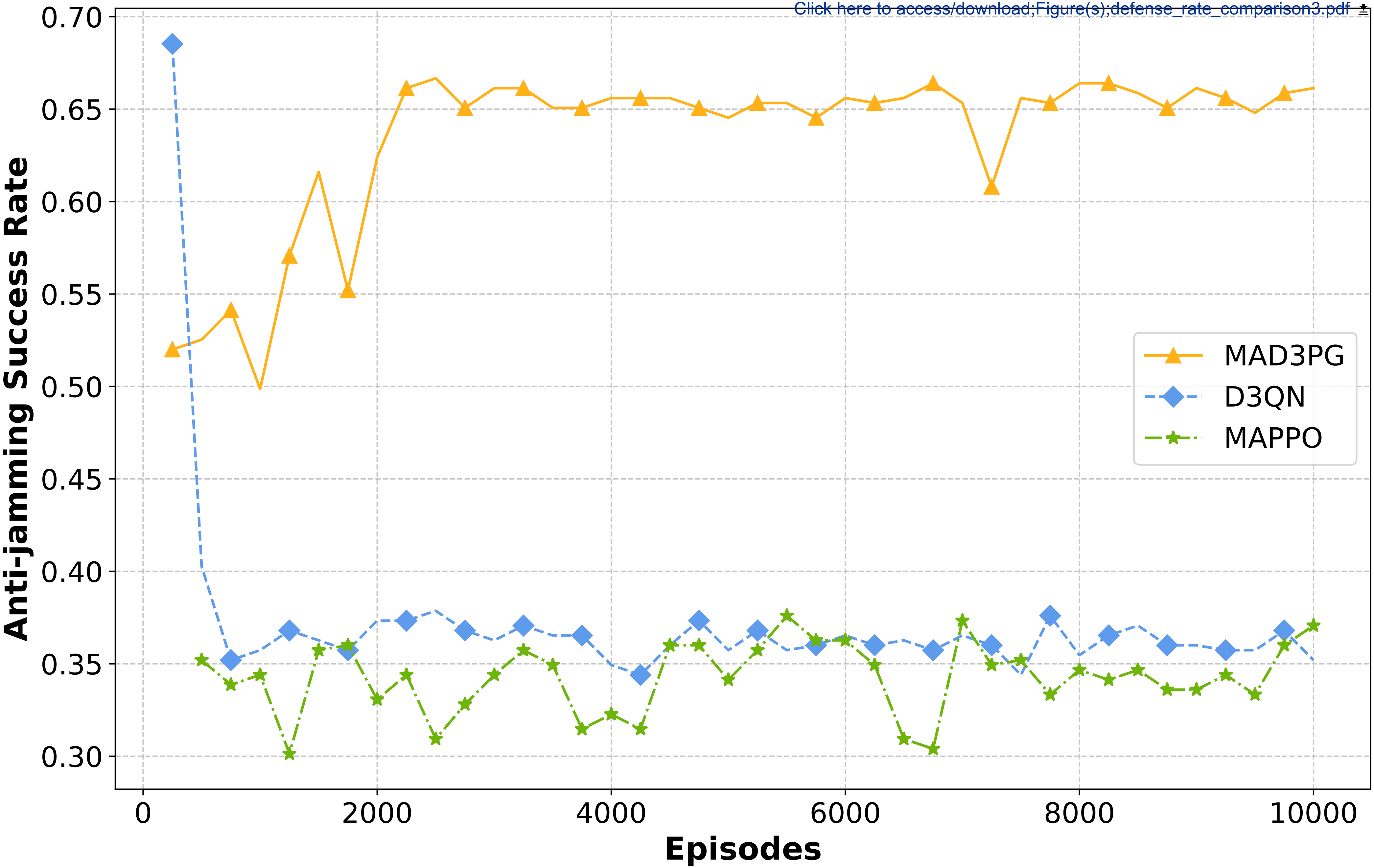
Zhan, Y., Zheng, B., Liu, D., Deng, B., Yang, X., 2025. Exploring black-box adversarial attacks on interpretable deep learning systems. Computer Vision and Image Understanding 259, 104423.

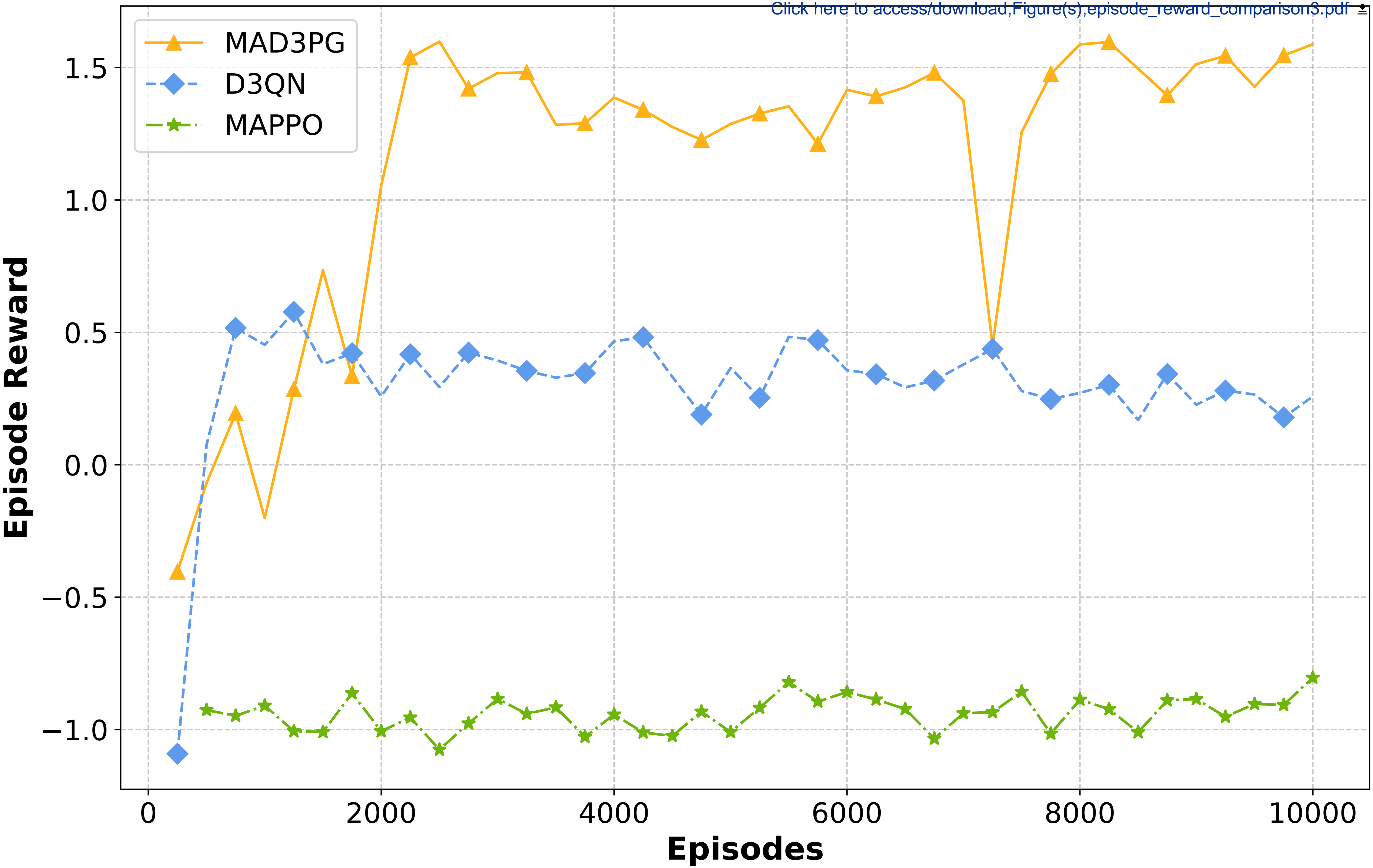
Zhang, P., Chen, N., Shen, S., Yu, S., Wu, S., Kumar, N., 2024. Future quantum communications and networking: A review and vision. IEEE Wireless Communications 31, 141–148.

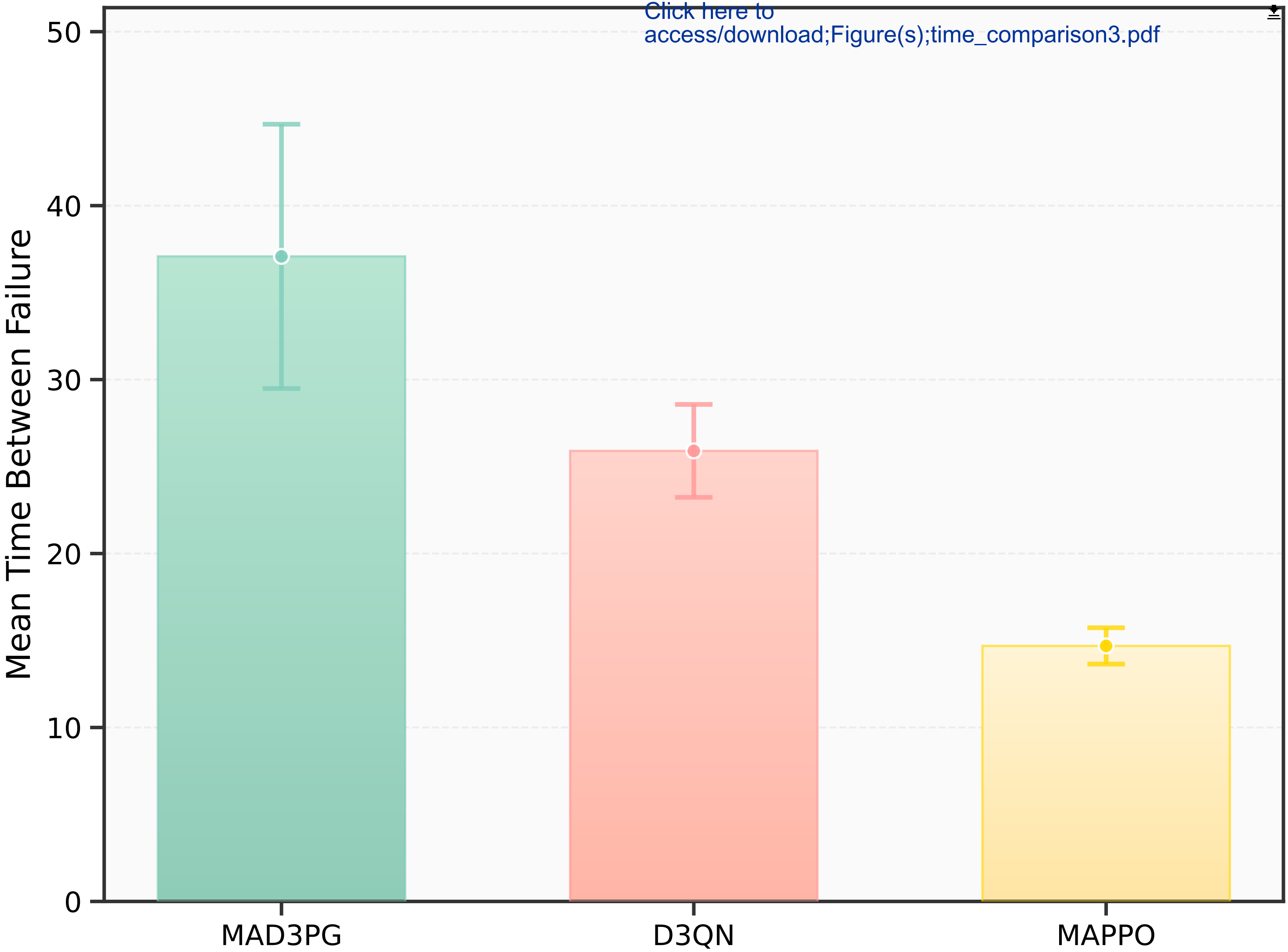
Zhou, M., Han, L., Che, X., 2025. Strengthening edge defense: A differential game-based edge intelligence strategy against APT attacks. Computers & Security 157, 104580.

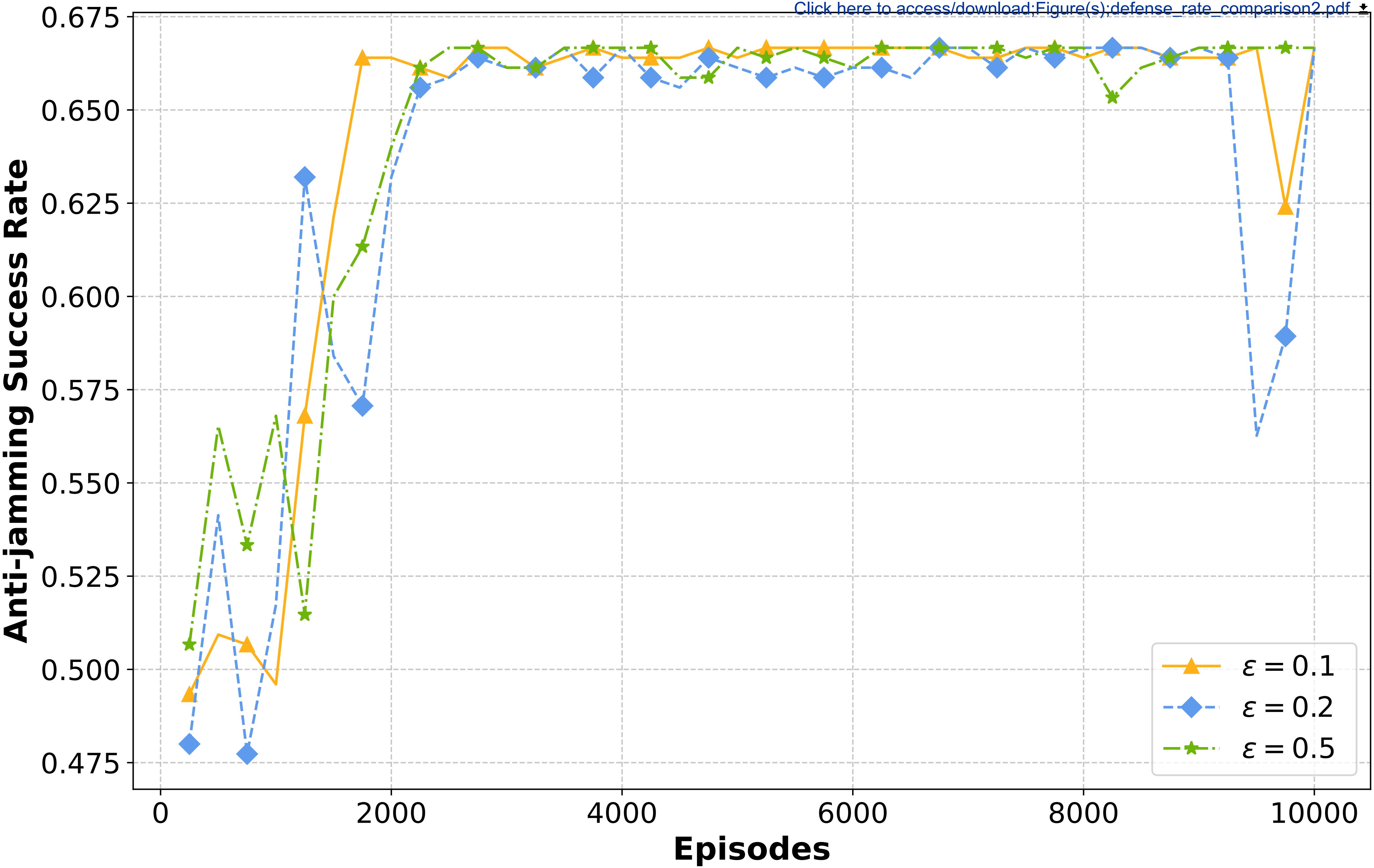

Actor
Actor
Actor
Critic
Critic
Critic
Agent 1
Agent 2
Agent n
Actor
Evaluation Net
Soft update
Target Net
Critic
Evaluation Net
Soft update
Target Net
Attention Mechanism
Loss Function
Replay Buffer
(s_t, a_t, r_t, s_{t+1})
Mini-batch
Normalization
EI-enabled SIoT Environment
Cloud
Edge intelligence devices
Base Station
jammer
SIoT Nodes

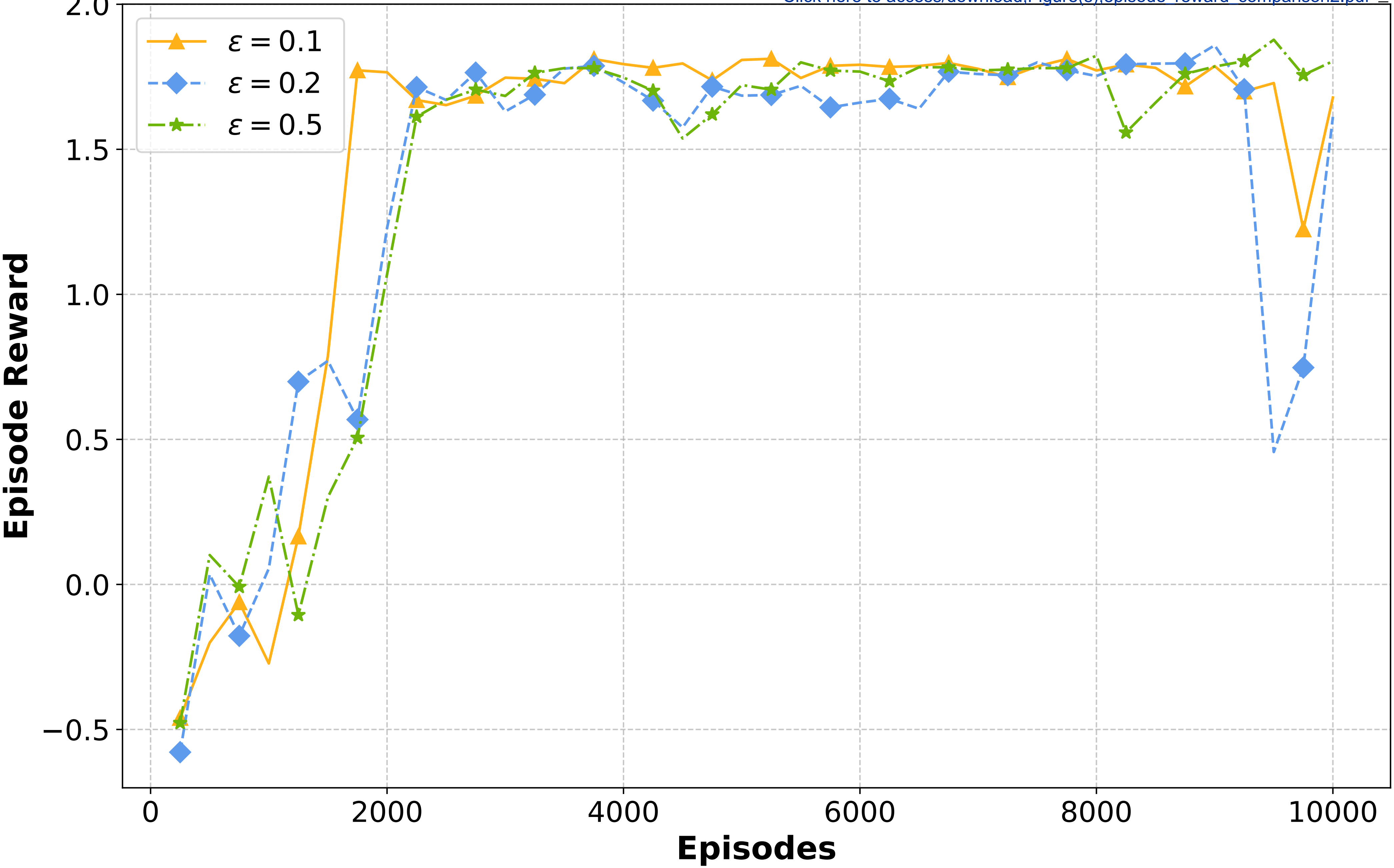

Cloud layer
Edge intelligence layer
Edge Service
Base Stations
Base Stations
Edge Service
5G/6G/Wi-Fi
jammer
End layer

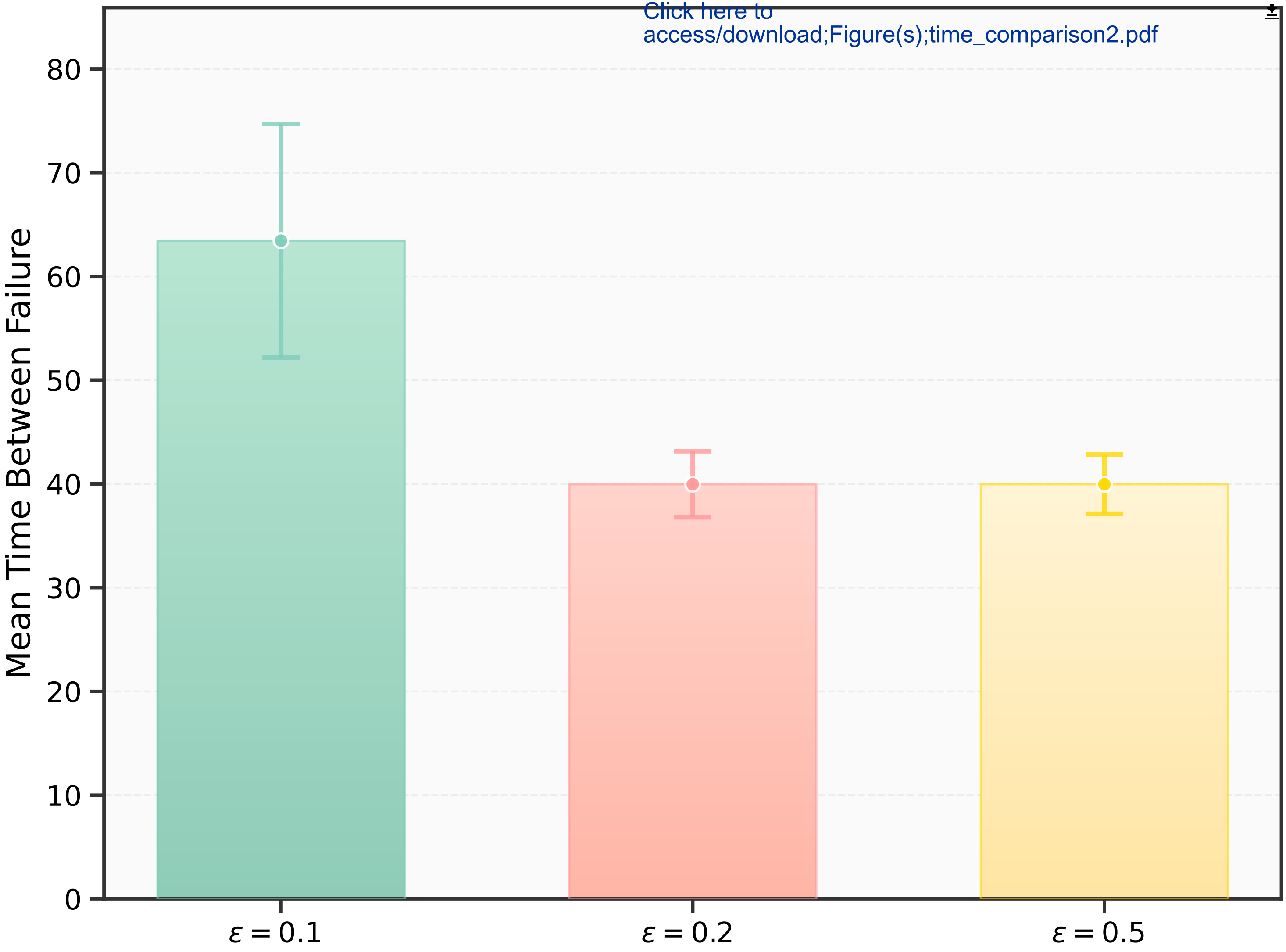

Anti-jamming Success Rate
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0
2000
4000
6000
8000
10000
Episodes
MAD3PG
D3QN
MAPPO

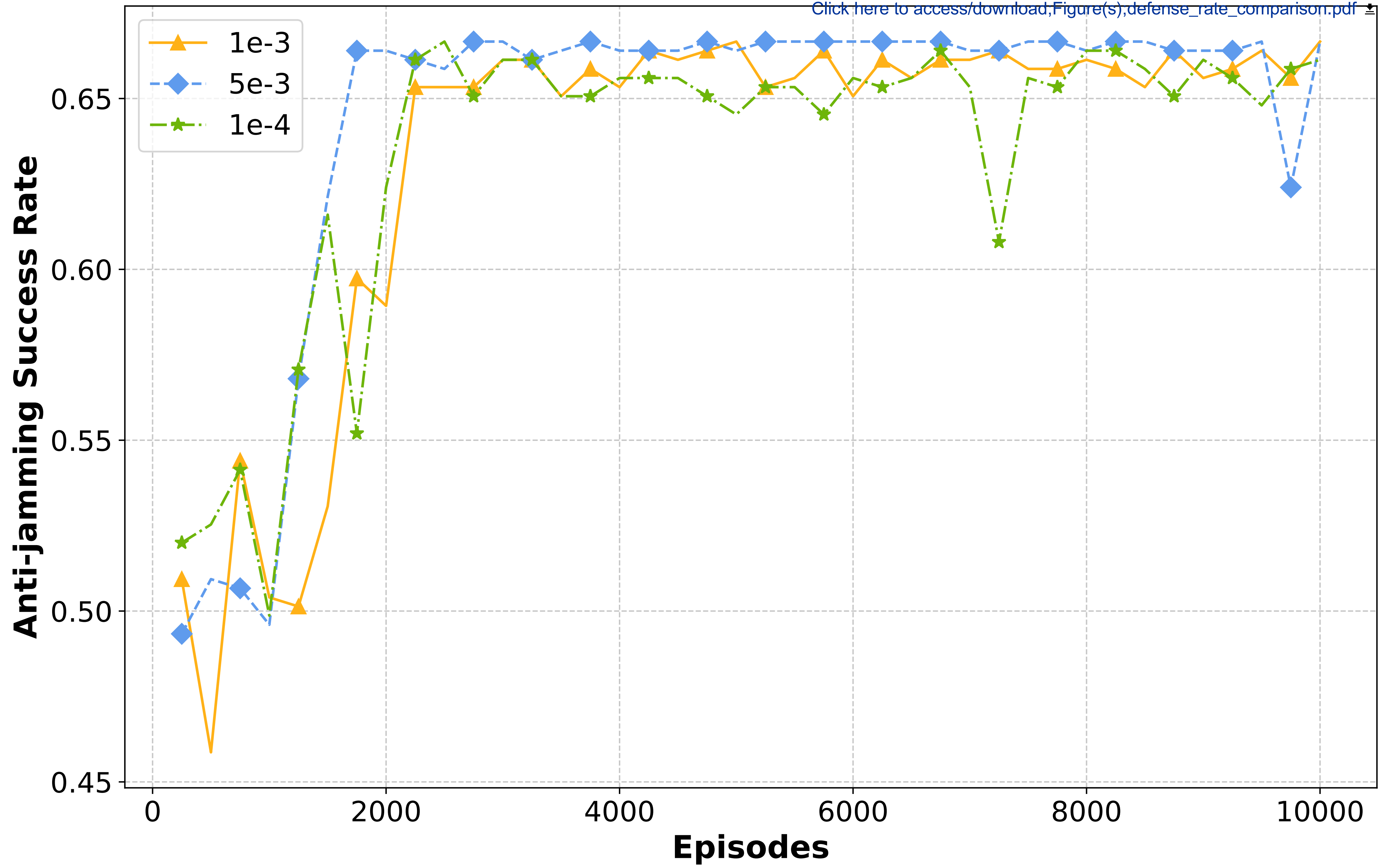

MAD3PG
D3QN
MAPPO
1.5
1.0
0.5
0.0
−0.5
−1.0
Episode Reward
0
2000
4000
6000
8000
10000
Episodes

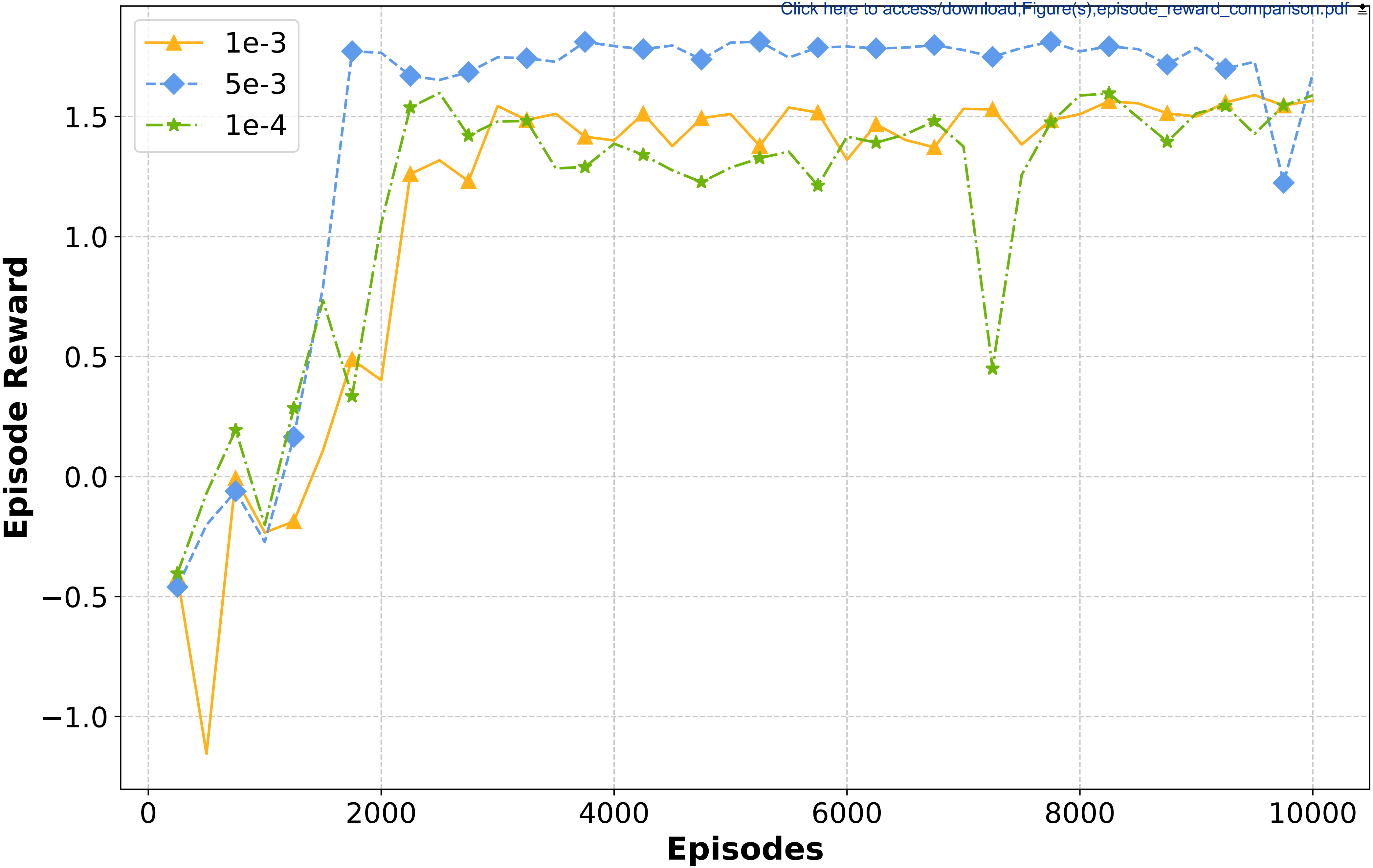

50
40
30
20
10
0
Mean Time Between Failure
MAD3PG
D3QN
MAPPO

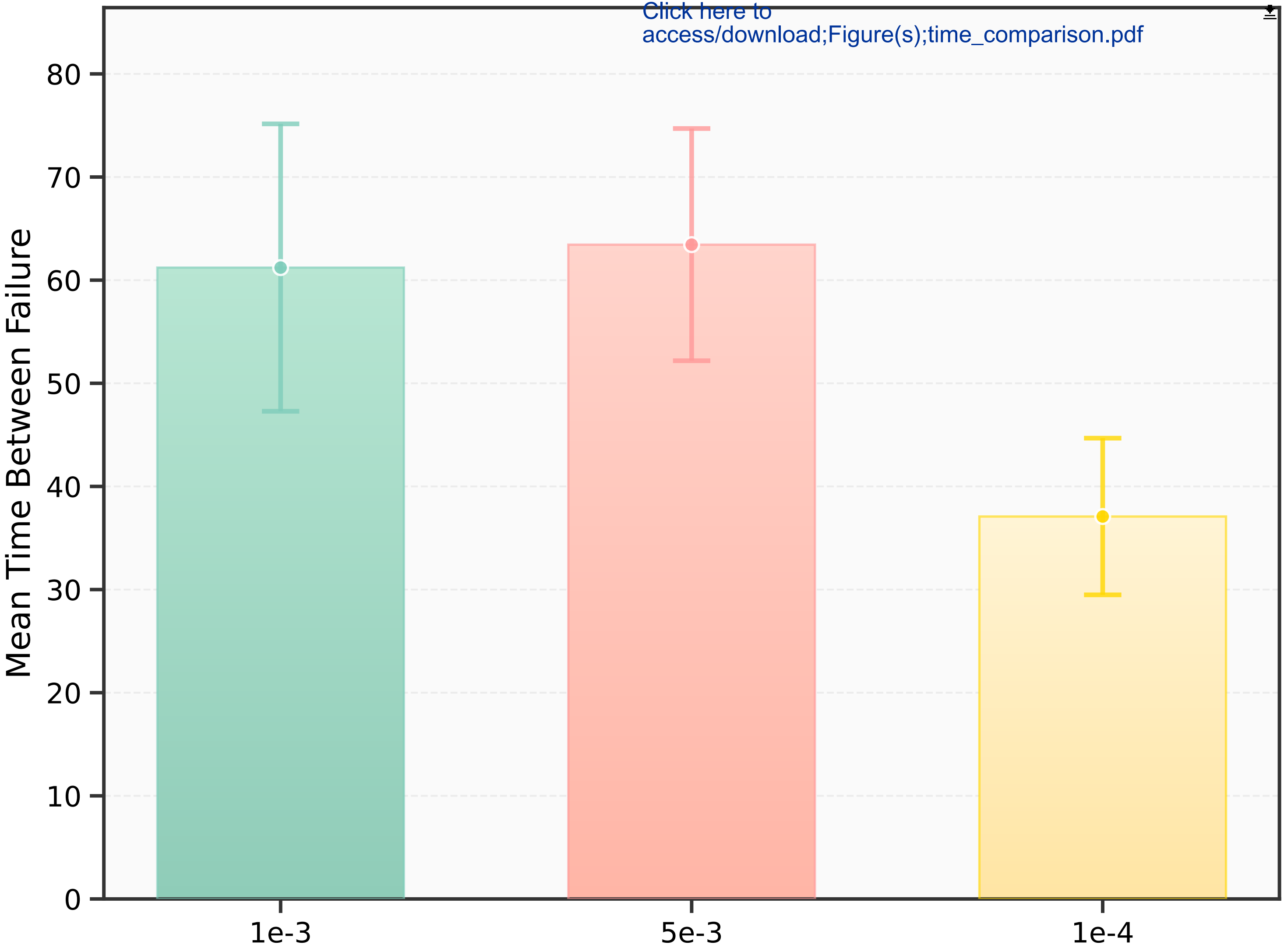

Anti-jamming Success Rate
0.675
0.650
0.625
0.600
0.575
0.550
0.525
0.500
0.475
0
2000
4000
6000
8000
10000
Episodes
$\varepsilon = 0.1$
$\varepsilon = 0.2$
$\varepsilon = 0.5$

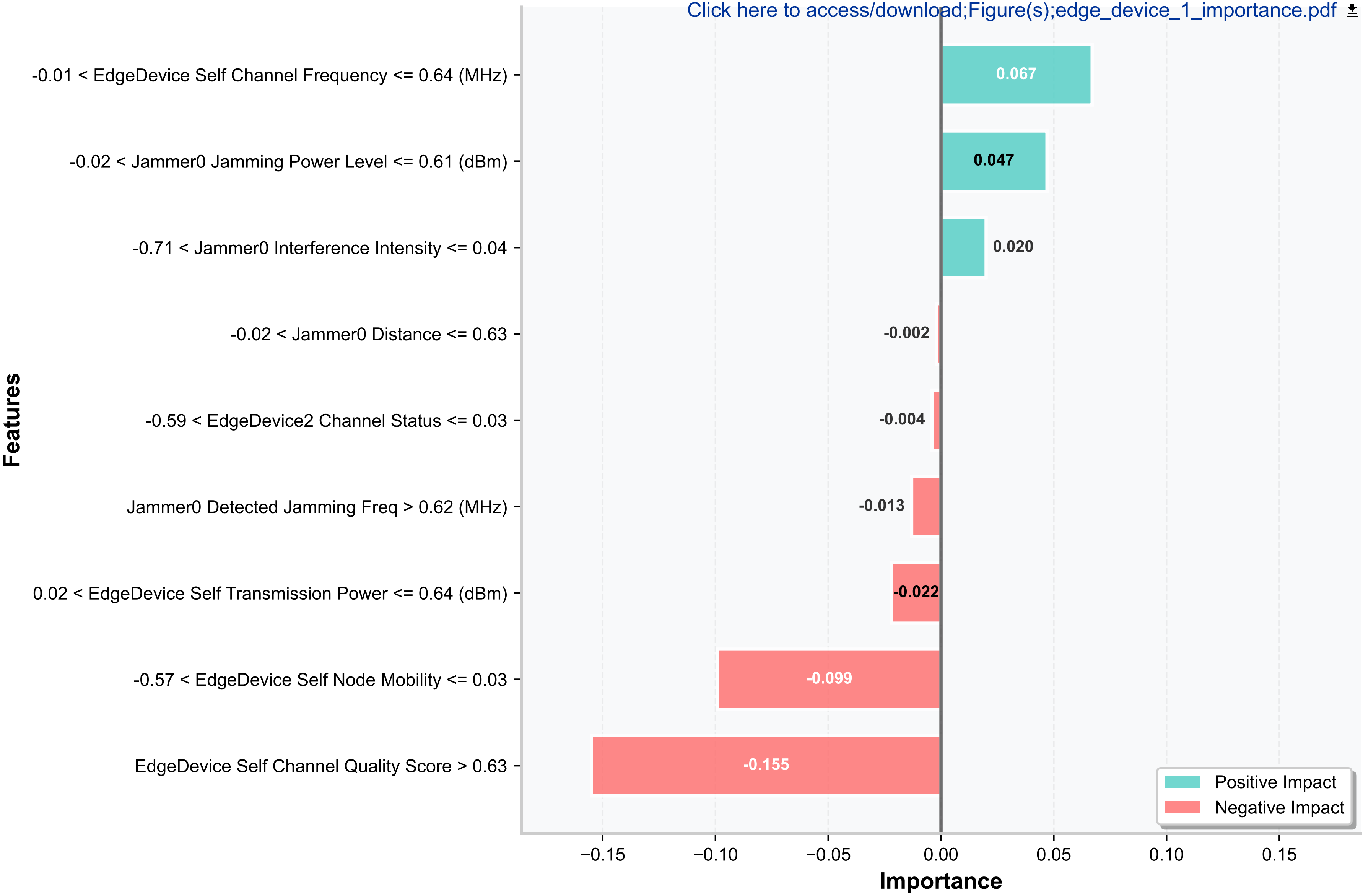

$\varepsilon = 0.1$
$\varepsilon = 0.2$
$\varepsilon = 0.5$
Episode Reward
2.0
1.5
1.0
0.5
0.0
−0.5
Episodes
0
2000
4000
6000
8000
10000

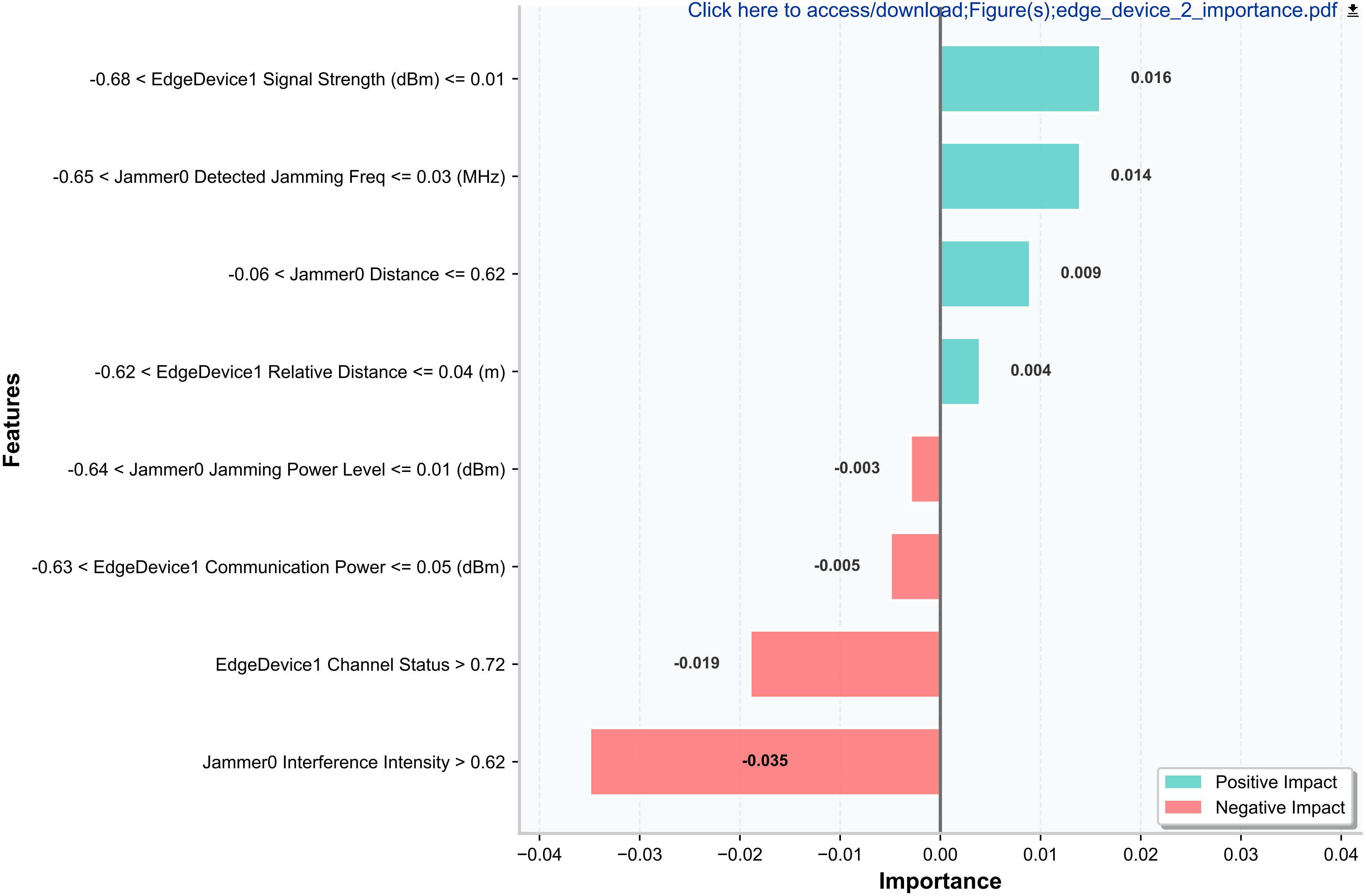

Mean Time Between Failure
0
10
20
30
40
50
60
70
80
$\varepsilon = 0.1$
$\varepsilon = 0.2$
$\varepsilon = 0.5$


1e-3
5e-3
1e-4
Anti-jamming Success Rate
0.45
0.50
0.55
0.60
0.65
Episodes
0
2000
4000
6000
8000
10000


1e-3
5e-3
1e-4
Episode Reward
1.5
1.0
0.5
0.0
−0.5
−1.0
0
2000
4000
6000
8000
10000
Episodes


Mean Time Between Failure
0
10
20
30
40
50
60
70
80
1e-3
5e-3
1e-4


-0.01 < EdgeDevice Self Channel Frequency <= 0.64 (MHz)
0.067
-0.02 < Jammer0 Jamming Power Level <= 0.61 (dBm)
0.047
-0.71 < Jammer0 Interference Intensity <= 0.04
0.020
-0.02 < Jammer0 Distance <= 0.63
-0.002
-0.59 < EdgeDevice2 Channel Status <= 0.03
-0.004
Jammer0 Detected Jamming Freq > 0.62 (MHz)
-0.013
0.02 < EdgeDevice Self Transmission Power <= 0.64 (dBm)
-0.022
-0.57 < EdgeDevice Self Node Mobility <= 0.03
-0.099
EdgeDevice Self Channel Quality Score > 0.63
-0.155
Features
−0.15
−0.10
−0.05
0.00
0.05
0.10
0.15
Importance
Positive Impact
Negative Impact


-0.68 < EdgeDevice1 Signal Strength (dBm) <= 0.01
-0.65 < Jammer0 Detected Jamming Freq <= 0.03 (MHz)
-0.06 < Jammer0 Distance <= 0.62
-0.62 < EdgeDevice1 Relative Distance <= 0.04 (m)
-0.64 < Jammer0 Jamming Power Level <= 0.01 (dBm)
-0.63 < EdgeDevice1 Communication Power <= 0.05 (dBm)
EdgeDevice1 Channel Status > 0.72
Jammer0 Interference Intensity > 0.62
0.016
0.014
0.009
0.004
-0.003
-0.005
-0.019
-0.035
Features
−0.04
−0.03
−0.02
−0.01
0.00
0.01
0.02
0.03
0.04
Importance
🟩 Positive Impact
🟥 Negative Impact