An Agentic Reinforcement Learning Model for Polymorphic Adversarial AI Threats Using Graph Neural Defense

Edward Fondo *, Kevin Tole, Fullgence Mwakondo

Posted Date: 24 March 2026

doi: 10.20944/preprints202603.1818.v1

Keywords: cybersecurity; graph neural networks; reinforcement learning; generative adversarial networks

 doi:10.20944/preprints202603.1818.v1

Article

An Agentic Reinforcement Learning Model for Polymorphic Adversarial AI Threats Using Graph Neural Defense

Edward Fondo *, Kevin Tole and Fullgence Mwakondo

Institute of Computing and Informatics, Technical University of Mombasa, Mombasa, Kenya
* Correspondence: msit00302023@students.tum.ac.ke

Abstract

Polymorphic adversarial agents that dynamically mutate their behavior present significant challenges to conventional intrusion detection systems, which often rely on static feature representations or fixed signature-based models. Existing deep learning approaches, including CNNs, transformers, and graph neural networks (GNNs), demonstrate strong detection capabilities but exhibit limitations in handling continuously evolving attack patterns. In this paper, we propose an integrated entropy-driven agentic cyber defense model that combines graph neural network-based anomaly detection, generative adversarial network (GAN) polymorphic attack simulation, reinforcement learning-based mitigation, and game-theoretic attacker–defender modeling. It models the stochastic mutation of adversarial agents using entropy gradients, captures relational network structures through GNNs, simulates realistic polymorphic attacks using GANs, and dynamically adapts mitigation strategies via reinforcement learning. Experimental evaluation on the CICIoT2023 dataset demonstrates that the proposed model achieves superior performance over existing state-of-the-art methods, with an accuracy of 98.3%, F1-score of 0.98, and significant improvements in robustness against polymorphic attacks compared to CNN-based (93.2% accuracy, 0.92 F1), transformer-based (95.4%, 0.94 F1), and conventional GNN-based (96.1%, 0.95 F1) intrusion detection systems. Additional ablation studies confirm the contribution of entropy modeling, GNN embeddings, and RL-based mitigation to overall system effectiveness. Future work will explore federated and distributed cyber defense architectures, integration with edge computing for IoT environments, and adaptive policy learning under large-scale network conditions, enabling real-time resilience against highly sophisticated adaptive cyber threats.

Keywords: cybersecurity; graph neural networks; reinforcement learning; generative adversarial networks

1. Introduction

The rapid advancement of artificial intelligence (AI) has significantly transformed the cybersecurity landscape. Furthermore, as digital infrastructures expand through cloud computing, Internet of Things (IoT) ecosystems, and distributed cyber–physical environments, the attack surface available to malicious actors has increased dramatically. Consequently, modern cyber attacks are no longer limited to manually executed exploits; instead, they increasingly rely on autonomous software agents capable of dynamically adapting their behavior during execution. Moreover, these intelligent adversarial agents employ advanced machine learning techniques such as reinforcement learning, generative adversarial networks, and adversarial optimization strategies to continuously evolve attack strategies and evade detection mechanisms. As a result, the cybersecurity domain is witnessing a transition from static threat models toward highly adaptive adversarial ecosystems in which attackers and defenders engage in a continuous technological arms race [1].

Traditional intrusion detection systems (IDS) rely primarily on signature-based detection or static machine learning models trained on historical attack datasets. While such approaches have proven

 doi:10.20944/preprints202603.1818.v1

effective against previously known threats, they assume that attack patterns remain relatively stable over time. In practice, however, contemporary adversarial threats increasingly exhibit polymorphic characteristics, meaning that malicious software can dynamically mutate its behavioral patterns, network communication signatures, and payload structures. Consequently, this polymorphism enables attackers to bypass conventional detection mechanisms by constantly modifying the observable features used by security systems for classification. Therefore, static detection models frequently experience reduced effectiveness when deployed in real-world environments where threat patterns evolve continuously [2,3].

Recent advances in artificial intelligence have introduced new opportunities for improving cyber defense mechanisms. In particular, graph-based representations of network interactions have emerged as a powerful model for analyzing complex communication infrastructures. Moreover, modern computer networks can naturally be modeled as graphs in which nodes represent devices or services and edges represent communication relationships between them. Within this representation, malicious activity often manifests as abnormal interaction patterns among network entities. Consequently, Graph Neural Networks (GNNs) provide a powerful computational mechanism for learning such relational dependencies by propagating information through network connections during training. By leveraging structural properties of communication graphs, GNN-based intrusion detection systems can therefore identify subtle anomalies that may not be detectable using conventional feature-based machine learning techniques [2,4,5].

Parallel to the development of graph-based cybersecurity models, reinforcement learning (RL) has gained significant attention as a method for enabling autonomous cyber defense strategies. Specifically, RL agents operate by learning optimal policies through continuous interaction with an environment while receiving feedback in the form of rewards or penalties. In addition, within cybersecurity contexts, reinforcement learning can be used to dynamically determine mitigation strategies such as blocking malicious connections, isolating compromised nodes, or modifying firewall configurations. Unlike static rule-based systems, RL-based defense agents can adapt their strategies over time as new attack behaviors emerge. Consequently, this adaptive capability becomes particularly valuable in environments characterized by highly dynamic adversarial behavior [6].

Despite these advances, existing AI-based cybersecurity models often fail to account for the stochastic mutation processes exhibited by polymorphic adversarial agents. Many machine learning models implicitly assume that attack distributions remain relatively stable during training and deployment. In reality, however, adversaries frequently attempt to maximize behavioral diversity in order to evade detection systems. Therefore, entropy-based modelling provides a theoretical framework for representing uncertainty and behavioral diversity in adversarial systems. By quantifying the entropy of attack distributions, it becomes possible to detect sudden increases in behavioral variability that may indicate the presence of polymorphic attack strategies.

Another emerging challenge in cybersecurity research involves the development of realistic adversarial training environments. Machine learning-based defense systems often rely on historical datasets that may not accurately represent the full diversity of potential attack behaviors. However, Generative Adversarial Networks (GANs) provide a mechanism for generating synthetic attack patterns that closely mimic real-world adversarial behaviour. Furthermore, by simulating polymorphic attacks during the training process, GAN-based models can significantly improve the robustness of cyber defense systems against previously unseen threats. Consequently, recent studies highlight that adversarial learning techniques and GAN-based cyber defense mechanisms can substantially enhance resilience against sophisticated network intrusions [7].

Motivated by these challenges, this paper proposes an integrated cyber defense architecture that combines entropy-based adversarial modeling, graph neural network anomaly detection, generative adversarial attack simulation, and reinforcement learning-based mitigation strategies. Specifically, the proposed system models network traffic as a dynamic communication graph and employs graph neural networks to detect anomalous node behavior. Furthermore, an entropy analysis module captures

 doi:10.20944/preprints202603.1818.v1

adversarial mutation dynamics, while a generative adversarial network generates polymorphic attack samples for adversarial training. In addition, the interaction between attackers and defenders is modeled as a stochastic game, thereby allowing reinforcement learning agents to learn optimal defense strategies under dynamic adversarial conditions.

The main contributions of this work are summarized as follows:

1. Development of an entropy-based mutation model for representing polymorphic adversarial cyber attacks, thereby enabling the detection of behavioral diversity and evolving attack patterns within network traffic environments.
2. Design of a graph neural network (GNN) anomaly detection model capable of learning structural relationships in dynamic communication graphs and accurately identifying malicious network interactions.
3. Introduction of a generative adversarial network (GAN) based polymorphic attack simulator that generates realistic adversarial network traffic, thereby improving robustness and adversarial training of cyber defense systems.
4. Formulation of a game-theoretic attacker–defender interaction model that captures strategic decision-making dynamics in adversarial cybersecurity environments and provides a theoretical foundation for adaptive defense mechanisms.
5. Development of an integrated reinforcement learning based cyber defense strategy combined with extensive experimental evaluation on the CICIoT2023 dataset, consequently demonstrating improved detection accuracy, adaptability, and resilience against polymorphic adversarial attacks.

The remainder of this paper is organized as follows. Section 2 reviews related work in AI-driven cybersecurity and adaptive intrusion detection systems. Section 3 presents the proposed entropy-driven cyber defense architecture. Section 4 introduces the mathematical formulation of the adversarial mutation model and presents the game-theoretic attacker–defender model. Section 5 describes the graph neural network detection model, GAN-based attack simulation approach, and reinforcement learning defense policy. Section 6 evaluates the proposed model using the CICIoT2023 dataset. Finally, Section 7 concludes the paper and outlines directions for future research.

2. Related Work

The rapid growth of artificial intelligence in cybersecurity has led to the development of sophisticated intrusion detection systems capable of identifying both conventional and emerging threats. Convolutional neural networks (CNNs) have been widely employed for detecting patterns in network traffic due to their ability to extract hierarchical feature representations [1]. Recent transformer-based architectures further enhance detection performance by capturing long-range dependencies and temporal correlations in sequential network traffic [2]. These approaches, however, often rely on fixed feature representations and may struggle with adaptive or polymorphic attack behaviors.

Graph neural networks (GNNs) have emerged as a powerful alternative for modeling relational cyber threats, leveraging the inherent graph structure of network environments. Nodes represent hosts or devices, while edges capture communication links. Recent studies demonstrate that GNN-based intrusion detection models can effectively detect anomalous network behaviors by learning structural patterns and relational dependencies [4,5,7]. In particular, attention-based GNN models, such as graph attention networks (GATs), improve anomaly detection by assigning higher weights to critical nodes or edges exhibiting suspicious activity [2,10].

Generative adversarial networks (GANs) have been applied to generate realistic adversarial network traffic, enabling robust evaluation and training of detection systems [8]. GAN-based simulation of polymorphic attacks allows researchers to model the continuous evolution of attack patterns, creating more challenging scenarios for intrusion detection models [3]. This capability is especially important in testing the resilience of machine learning-based defense mechanisms against adaptive adversaries.

Reinforcement learning (RL) has also been explored to develop adaptive cyber defense policies, where agents learn optimal mitigation strategies through interaction with the network

 doi:10.20944/preprints202603.1818.v1

environment [6,11]. RL-based defense mechanisms can dynamically respond to evolving threats by taking actions such as blocking connections, isolating compromised nodes, or updating firewall rules. Recent works highlight the potential of combining RL with GNNs to provide a holistic model that considers both structural network information and decision-making policies [6,12].

Entropy-based anomaly detection techniques have been employed to quantify the unpredictability of network traffic, providing an effective means of identifying irregular patterns [1,5]. These methods are particularly valuable in detecting polymorphic attacks where adversaries continually alter their behavior to evade detection. However, existing studies rarely integrate entropy modeling with GNN-based structural analysis and RL-based adaptive responses within a unified model. Such integration could provide a more comprehensive solution, capturing the stochastic nature of adversarial mutations, relational dependencies among network nodes, and dynamic defense adaptation.

In summary, while prior work has demonstrated the individual strengths of CNNs, transformers, GNNs, GANs, RL, and entropy-based methods, there is a clear gap in unifying these approaches into a single cyber defense architecture capable of handling polymorphic, adaptive attacks. This motivates the design of the proposed entropy-driven GNN and RL-based model, which leverages GAN-generated adversarial traffic and game-theoretic modeling to enhance detection accuracy, adaptability, and robustness against evolving cybersecurity threats.

3. Problem Formulation

The network environment is represented as a dynamic graph:

$$G_t = (V, E, X_t) \tag{1}$$

where: V represents network nodes,
E represents communication edges,
X_t represents node feature vectors at time t.

The entropy-based adversarial mutation mechanism models the uncertainty and variability of attack behaviors in dynamic cyber environments. Equation (2) quantifies the entropy of attack actions, capturing the probability distribution of different malicious behaviors. Higher entropy values indicate more unpredictable and diverse attack patterns, which increase the complexity of detection. The mutation dynamics update the adversarial parameters by following the entropy gradient, allowing simulated attacks to evolve toward more sophisticated strategies. This process enables the learning model to generate adaptive adversarial scenarios that strengthen the robustness of the defense agent.

The entropy of attack behaviour is defined as:

$$H(A) = -\sum P(a_i) \log P(a_i) \tag{2}$$

Mutation dynamics follow the entropy gradient:

$$\theta_{t+1} = \theta_t + \alpha \nabla H(\theta_t) + \epsilon \tag{3}$$

Generative Adversarial Networks (GANs) are employed to simulate polymorphic cyberattacks that continuously evolve to evade detection systems. The adversarial training process is formulated as a minimax optimization problem where the generator attempts to produce realistic attack patterns while the discriminator distinguishes between real and synthetic samples. The generator loss encourages the creation of attack instances that can successfully deceive the discriminator. Conversely, the discriminator loss improves the model's ability to differentiate between legitimate traffic and generated adversarial attacks. Through iterative competition between these two networks, the model generates diverse and adaptive attack variants that enhance the robustness of the cyber defense system.

The GAN objective function is defined as:

 doi:10.20944/preprints202603.1818.v1

Generator loss:

$$\min_G \max_D V(D,G) \tag{4}$$

$$L_G = -\mathbb{E}[\log D(G(z))] \tag{5}$$

Discriminator loss:

$$L_D = -\mathbb{E}[\log D(x)] - \mathbb{E}[\log(1 - D(G(z)))] \tag{6}$$

The game-theoretic attacker–defender model represents the strategic interaction between a cyber attacker and the defense system. Table 1 illustrates the payoff structure, where the defender gains when an attack is successfully detected and loses when it is ignored. Conversely, the attacker benefits when malicious activity bypasses detection mechanisms. This adversarial interaction forms a strategic decision-making environment in which both agents continuously adapt their actions. The optimal equilibrium policy π^* is determined by maximizing the expected reward, enabling the defense system to learn the most effective response strategy under uncertainty.

Table 1. Attacker–Defender Payoff Matrix

	Detect	Ignore
Attack	(-1,1)	(1,-1)
No Attack	(0,0)	(0,0)

The equilibrium strategy is defined as:

$$\pi^* = \arg\max_\pi \mathbb{E}[R] \tag{7}$$

The reinforcement learning defense policy enables the system to learn adaptive responses to evolving cyber threats. The system state s_t captures both the graph-based network representation G_t and the current attack characteristics A_t, providing contextual awareness of the environment. Based on this state, the defense agent selects mitigation actions aimed at minimizing system vulnerabilities. The reward function balances successful threat detection (D) against false alarms (F), ensuring that the agent maintains both accuracy and operational efficiency. Through continuous interaction with the environment, the policy gradually converges toward an optimal defense strategy that maximizes long-term security performance.

The system state is defined as:

$$s_t = (G_t, A_t) \tag{8}$$

Reward function:

$$R = \alpha D - \beta F \tag{9}$$

The convergence analysis establishes the theoretical stability of the proposed reinforcement learning defense policy. Under the assumptions of bounded rewards and a finite state space, the learning process is guaranteed to converge to an optimal strategy. The Bellman optimality equation defines the recursive relationship used to estimate the optimal state-value function. Through iterative updates, the value function progressively approaches the optimal solution by maximizing expected cumulative rewards. Because the Bellman operator forms a contraction mapping, repeated updates ensure convergence to the unique optimal value function $V^*(s)$.

Theorem. Given bounded rewards and a finite state space, the reinforcement learning defense policy converges to an optimal strategy.

Proof. Using Bellman optimality:

$$V^*(s) = \max_a \left[R(s,a) + \gamma \sum P(s'|s,a)V^*(s')\right] \tag{10}$$

Iterative updates converge under contraction mapping principles.

 doi:10.20944/preprints202603.1818.v1

4. Proposed Method

As illustrated in Figure 1, the proposed entropy-driven cyber defense model begins with raw network traffic that is transformed into graph representations by the graph constructor. The graph data is analyzed by a Graph Neural Network (GNN) to detect structural anomalies in network interactions. An entropy analyzer simultaneously evaluates the uncertainty and irregularity patterns within the traffic flow. A Generative Adversarial Network (GAN) simulates adaptive cyberattacks to strengthen the training environment of the reinforcement learning agent. The RL defense agent integrates these inputs to trigger the mitigation engine, which performs automated defensive actions against detected threats.

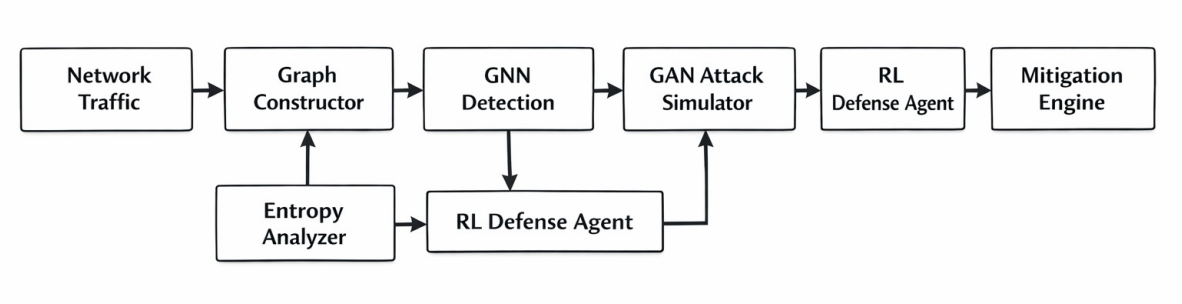

Figure 1. Unified architecture of the proposed entropy-driven agentic cyber defense framework. Raw network traffic is transformed into a temporal graph, analyzed using graph neural representations, regulated by entropy-aware adversarial mutation, and optimized through reinforcement learning-based mitigation.

The proposed method introduces an entropy-driven cyber defense model that integrates graph learning, adversarial simulation, and reinforcement learning to detect and mitigate advanced cyber threats. Raw network traffic is first transformed into a dynamic graph representation and analyzed using a Graph Neural Network (GNN) to capture structural anomalies in network interactions. An entropy analyzer simultaneously measures uncertainty and irregularity in traffic behavior to characterize evolving attack patterns. To strengthen the training environment, a Generative Adversarial Network (GAN) generates polymorphic cyberattack scenarios that mimic adaptive adversaries. Finally, a reinforcement learning defense agent integrates graph features, entropy metrics, and adversarial simulations to learn optimal mitigation strategies and trigger automated defense actions against detected threats.

The training pipeline utilizes the CICIoT2023 dataset, a large-scale IoT cybersecurity dataset containing diverse attack categories such as DDoS, botnet activity, reconnaissance, and command-and-control traffic. Raw network flow records are first preprocessed through data cleaning, normalization, and feature extraction to construct the dynamic graph representation G_t. Key traffic attributes, including flow duration, packet length statistics, protocol type, and packet counts, are mapped to node feature vectors used by the Graph Neural Network (GNN). The entropy analyzer then evaluates uncertainty and irregularity patterns in the extracted features to characterize evolving attack behaviors. These processed representations are subsequently used to train the reinforcement learning defense agent, enabling the system to learn adaptive mitigation strategies against diverse cyber threats.

As illustrated in Figure 2, the CICIoT2023 dataset is first preprocessed and transformed into a dynamic graph representation before being analyzed by the GNN and reinforcement learning defense agent.

 doi:10.20944/preprints202603.1818.v1

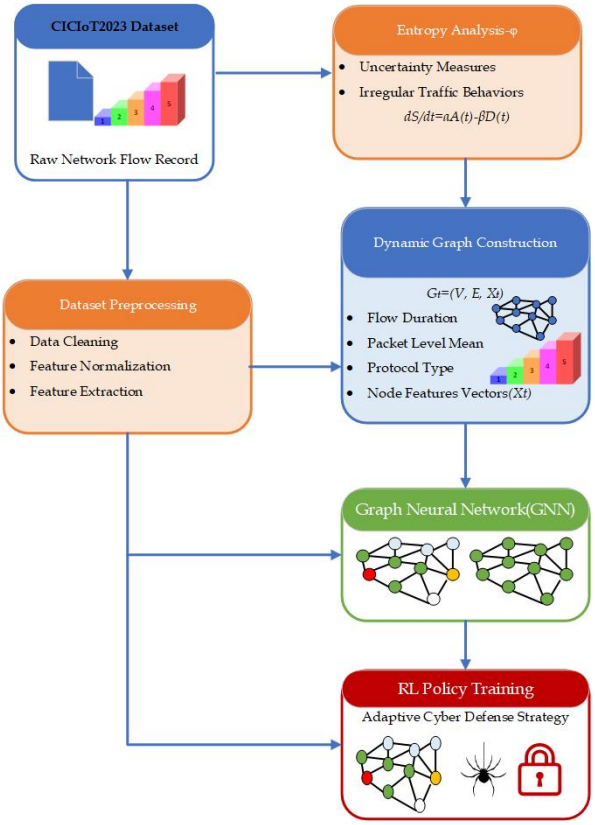

Figure 2. Training pipeline of the proposed framework using the CICIoT2023 dataset. The pipeline includes traffic preprocessing, temporal graph construction, entropy analysis, graph neural representation learning, adversarial simulation, and reinforcement learning policy training.

Figure 3 illustrates the GAN-based cyberattack simulation and adaptive defense model. First, the network environment is modeled as a dynamic communication graph where nodes and edges represent devices and interactions. Next, an entropy-driven mutation module generates polymorphic attack behaviors which are used by the GAN generator to synthesize adversarial traffic samples. The discriminator then evaluates both real and generated traffic to distinguish legitimate network activity from malicious patterns. Finally, the detection outcomes feed into a reinforcement learning defense agent guided by a game-theoretic reward mechanism to learn optimal mitigation strategies.

 doi:10.20944/preprints202603.1818.v1

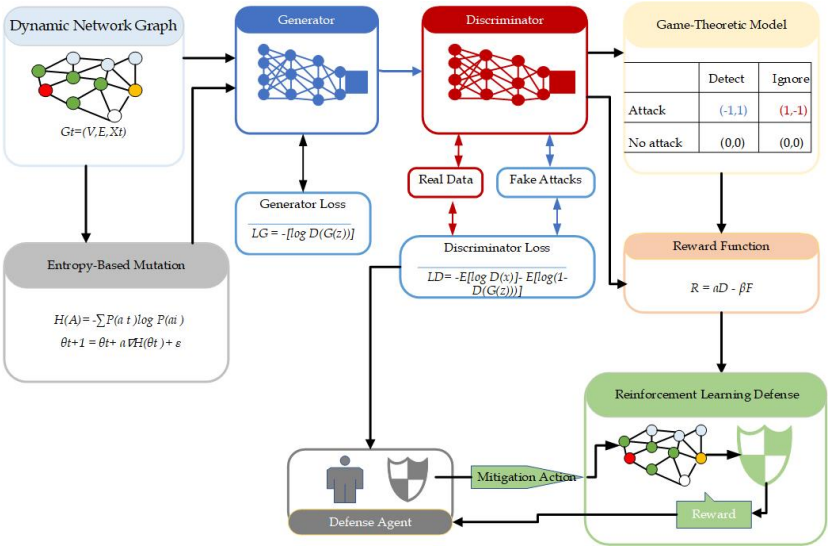

Figure 3. Entropy-conditioned adversarial simulation module. The generator produces polymorphic attack samples conditioned on the temporal graph and entropy signal, while the discriminator evaluates realism and the resulting samples are fed into the adaptive defense environment.

The graph construction module converts the preprocessed network traffic into a dynamic graph representation in order to capture structural relationships within the network environment. Each device or endpoint in the network is represented as a node V, while communication interactions between devices form the edges E. Network traffic attributes such as flow duration, packet length statistics, protocol type, and packet counts are encoded as node feature vectors X_t at time t. This representation models the network as a temporal graph

$$G_t = (V, E, X_t)$$

that evolves continuously with incoming traffic flows. The constructed graph is then provided as input to the Graph Neural Network (GNN) to detect abnormal communication patterns associated with cyberattacks.

Figure 4 illustrates the strategic interaction between the cyber attacker and the defense agent within a game theoretic framework. First, the attacker generates evolving cyber threats that attempt to bypass the detection mechanisms of the security system. The defender then evaluates the incoming traffic and decides whether to detect or ignore the suspected malicious activity based on the payoff structure. Furthermore, a reinforcement learning module continuously adapts the defense strategy by analyzing the outcomes of previous interactions and maximizing the reward function. Consequently, the system converges toward an optimal equilibrium policy that improves detection accuracy while minimizing false alarms in dynamic adversarial environments.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 March 2026 doi:10.20944/preprints202603.1818.v1

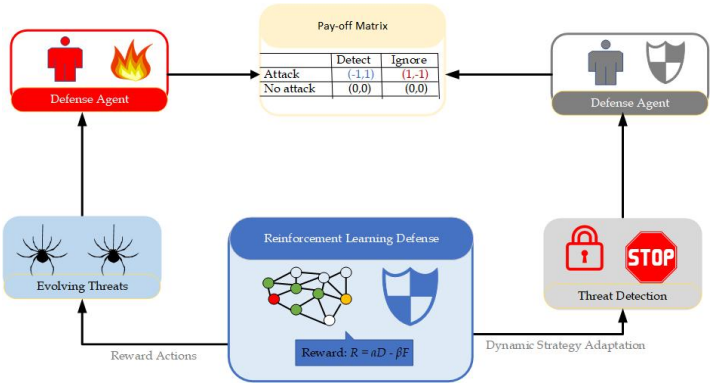

Figure 4. Game-theoretic attacker–defender model. The attacker generates evolving threats, while the defense agent observes graph-structured state information, receives entropy-aware feedback, and updates its policy to maximize long-term mitigation performance.

The reinforcement learning component of the proposed model as visualized in Figure 5 functions as an adaptive decision-making agent that learns optimal cyber defense policies from the dynamic network environment. The agent receives the graph-based network state and entropy-based uncertainty metrics as observations and selects mitigation actions such as traffic blocking, quarantine, or alert generation. A reward mechanism is designed to encourage correct threat mitigation while penalizing false alarms and delayed responses. Through continuous interaction with both real traffic data and adversarial scenarios generated by the GAN, the agent improves its policy using trial-and-error learning. This process enables the system to automatically adapt to evolving cyber threats and optimize real-time defensive strategies.

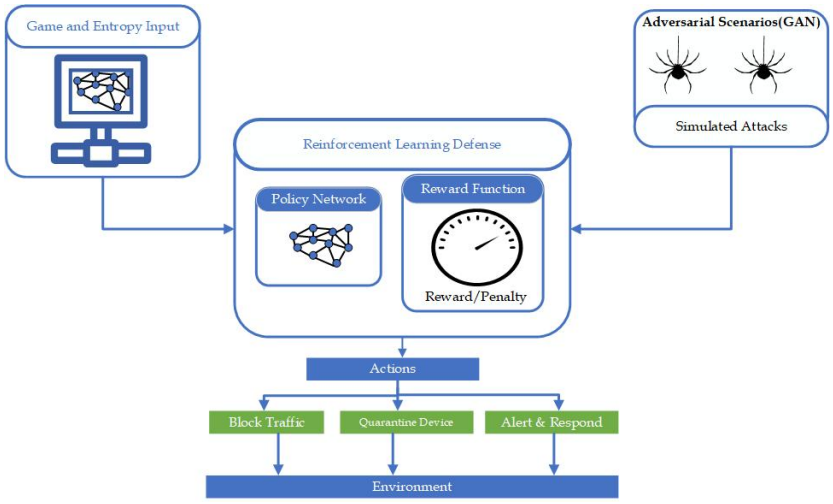

Figure 5. Reinforcement learning defense module. The agent receives graph-based structural embeddings, entropy-driven uncertainty measurements, and adversarial scenarios generated during simulation, and outputs mitigation actions for real-time cyber defense.

To jointly model polymorphic adversarial generation, graph-based anomaly detection, and adaptive defense learning, the overall optimization objective integrates entropy-driven mutation dynamics, generative adversarial learning, and reinforcement learning policy optimization.

Let the network environment be represented as a dynamic graph $G_t = (V, E, X_t)$ and the system state be defined as $s_t = (G_t, A_t)$ where A_t represents adversarial actions. The proposed cyber defense framework seeks to simultaneously minimize adversarial deception, regulate attack entropy, and maximize the long-term security reward of the defense agent.

 doi:10.20944/preprints202603.1818.v1

The unified objective function is defined as:

$$\min_{G}\max_{D}\max_{\pi} L_{\text{total}} = \lambda_1 L_{\text{GAN}} + \lambda_2 H(A) - \lambda_3 \mathbb{E}_{\pi}\left[\sum_{t=0}^{T} \gamma^t R(s_t, a_t)\right] + \lambda_4 L_{\text{GNN}} \tag{11}$$

$$L_{\text{GAN}} = -\mathbb{E}[\log D(x)] - \mathbb{E}[\log(1 - D(G(z)))] \tag{12}$$

$$H(A) = -\sum P(a_i) \log P(a_i) \tag{13}$$

$$\mathbb{E}_{\pi}\left[\sum_{t=0}^{T} \gamma^t R(s_t, a_t)\right] = \text{expected cumulative reward} \tag{14}$$

$$L_{\text{GNN}} = \sum_{v \in V} \ell(f_\theta(v, G_t), y_v) \tag{15}$$

The coefficients $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ control the relative contribution of adversarial simulation, entropy regularization, reinforcement learning defense performance, and graph-based anomaly detection. Therefore, the optimization objective jointly trains the adversarial attack simulator, the graph-based detection model, and the reinforcement learning defense policy, enabling the system to achieve robust detection and adaptive mitigation of polymorphic cyber threats.

The Entropy-Based Cyber Threat Detection algorithm, as shown in Algorithm 1, analyzes preprocessed network traffic features by representing them as a dynamic graph $G_t = (V, E, X_t)$, where nodes and edges capture relationships among traffic flows over time. Moreover, at each time step, feature vectors x_i are extracted from the network traffic and their probability distribution $P(x_i)$ is computed to understand the behavior of the traffic patterns. Furthermore, using this probability distribution, the algorithm calculates the entropy $H(X_t)$, which quantifies the level of randomness or uncertainty present in the observed traffic behavior. In addition, the computed entropy value is compared with a predefined threshold τ to determine whether the traffic pattern should be considered normal or anomalous. However, if the entropy value exceeds the threshold, the system identifies the traffic behavior as anomalous, triggers an alert, and forwards the system state to a reinforcement learning agent for further analysis. Otherwise, the traffic is classified as normal and no anomaly response is activated.

Algorithm 1 Entropy-Based Cyber Threat Detection

Require: Preprocessed network traffic features X_t, dynamic graph $G_t = (V, E, X_t)$
Ensure: Detection of anomalous traffic patterns

Initialize entropy threshold τ
Extract feature vectors x_i from network traffic flows
for each time step t **do**
    Construct dynamic graph representation $G_t = (V, E, X_t)$
    Compute probability distribution of traffic features $P(x_i)$
    Calculate entropy:

$$H(X_t) = -\sum_{i=1}^{n} P(x_i) \log P(x_i)$$

    if $H(X_t) > \tau$ **then**
        Mark traffic behavior as anomalous
        Trigger anomaly alert and forward state to reinforcement learning agent
    else
        Mark traffic behavior as normal
    end if
end for
return Detected anomaly events and entropy metrics

 doi:10.20944/preprints202603.1818.v1

Algorithm 2 Reinforcement Learning Defense Agent for Adaptive Cyber Mitigation

Require: Graph representation $G_t = (V, E, X_t)$, entropy metric $H(X_t)$, adversarial samples from GAN
Ensure: Optimal defense policy $\pi(a|s)$ for cyber threat mitigation
Initialize policy network parameters θ
Initialize replay memory M
Define action space $A = \{\text{Block Traffic}, \text{Quarantine Device}, \text{Alert Response}\}$
for each training episode **do**
    Observe network state $s_t = \{G_t, H(X_t)\}$
    Select action a_t using policy $\pi_\theta(a_t \mid s_t)$
    Execute action a_t in the network environment
    Observe next state s_{t+1} and reward r_t
    Store transition (s_t, a_t, r_t, s_{t+1}) in memory M
    Sample mini-batch transitions from M
    Update policy parameters θ by maximizing expected reward:

$$\theta \leftarrow \theta + \alpha \nabla_\theta \mathbb{E}\Big[r_t + \gamma \max_a Q(s_{t+1}, a)\Big]$$

end for
Deploy learned policy π_θ for real-time cyber defense actions
return Optimal mitigation strategy

The Reinforcement Learning Defense Agent algorithm, as illustrated in Algorithm 2, utilizes the graph representation of network traffic $G_t = (V, E, X_t)$ together with the entropy metric $H(X_t)$ and adversarial samples generated by a GAN to support adaptive cyber threat mitigation. Initially, the policy network parameters θ and replay memory M are initialized, and the action space is defined to include blocking traffic, quarantining devices, and issuing alert responses. Moreover, during each training episode, the agent observes the current network state $s_t = \{G_t, H(X_t)\}$ and selects an action based on the policy $\pi_\theta(a_t|s_t)$. Furthermore, after executing the selected action, the agent observes the resulting state transition and reward, which are stored in memory for experience replay. In addition, sampled transitions from the replay memory are used to update the policy parameters by maximizing the expected cumulative reward. Consequently, the learned policy is deployed to perform real-time cyber defense actions and provide an optimal mitigation strategy against potential threats.

5. Experimental Results

The experimental evaluation of the proposed entropy-driven cyber defense model will be conducted using the CICIoT2023 dataset to analyze detection behavior and system performance under diverse cyberattack scenarios. Furthermore, entropy values computed from dynamic network traffic features will be visualized using a 3D heat map to illustrate the spatial distribution and intensity of anomalous patterns across network states. In addition, the learning efficiency and mitigation capability of the reinforcement learning defense agent will be examined through a 3D performance trend graph that captures the relationship between training episodes, entropy levels, and detection accuracy. However, to better understand the contribution of each module, an ablation study will be performed by selectively removing components such as the entropy analyzer, GAN adversarial generator, and graph learning module. Yet, the resulting comparative performance trends will highlight how each component contributes to the overall robustness and adaptive capability of the proposed cyber defense framework.

The 3D entropy heat map in Figure 6 visualizes the variation of entropy values across network nodes and time to highlight irregular traffic behavior. Furthermore, regions with higher peaks and warmer colors indicate elevated entropy levels, suggesting potential anomalies or suspicious network activity. In addition, the gradual changes in the surface structure illustrate how uncertainty evolves over time as network interactions fluctuate. However, lower entropy regions represented by cooler colors indicate more stable and predictable traffic patterns within the network. Yet, the overall

 doi:10.20944/preprints202603.1818.v1

distribution of entropy across the graph provides valuable insights into the dynamic behavior of network traffic and helps identify zones where cyber threats are likely to emerge.

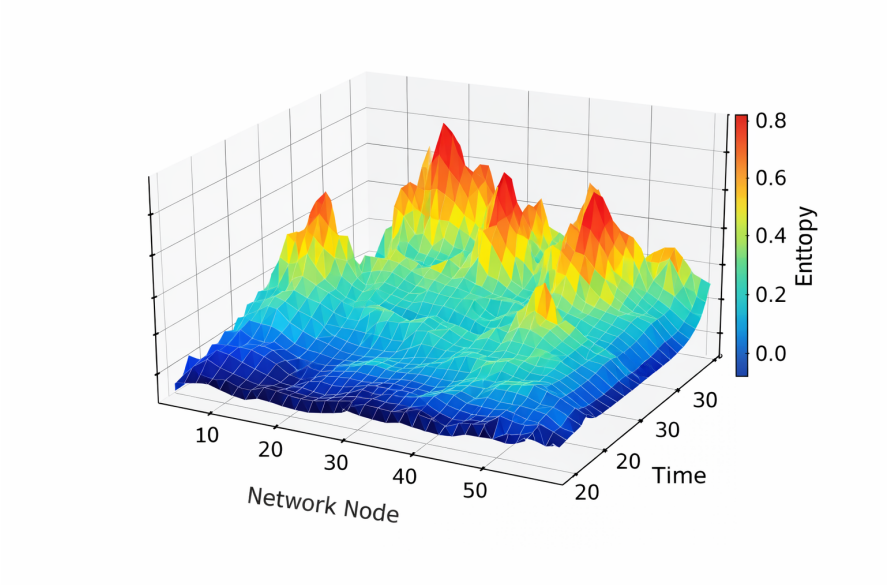

Figure 6. 3D entropy heatmap representing the variation of network traffic uncertainty across nodes and time. Peaks and warmer colors indicate higher entropy levels corresponding to potential anomalous behavior, whereas cooler regions correspond to more stable network activity. This visualization aids in identifying critical zones for adaptive cyber defense.

Figure 7 illustrates the 2D performance trend comparison between the proposed entropy-driven cyber defense model and several existing intrusion detection approaches. The results demonstrate that traditional CNN-based IDS achieves relatively lower accuracy and F1 score compared to more advanced architectures. Moreover, the Transformer and Graph-based IDS models show improved performance due to their enhanced capability in capturing complex network traffic patterns. Furthermore, the Hybrid IDS model provides additional improve- ments by integrating multiple learning techniques for better feature representation. However, the proposed model outperforms all baseline methods, achieving the highest accuracy and F1 score, thereby demonstrating the effectiveness of the entropy-driven framework for adaptive cyber threat detection.

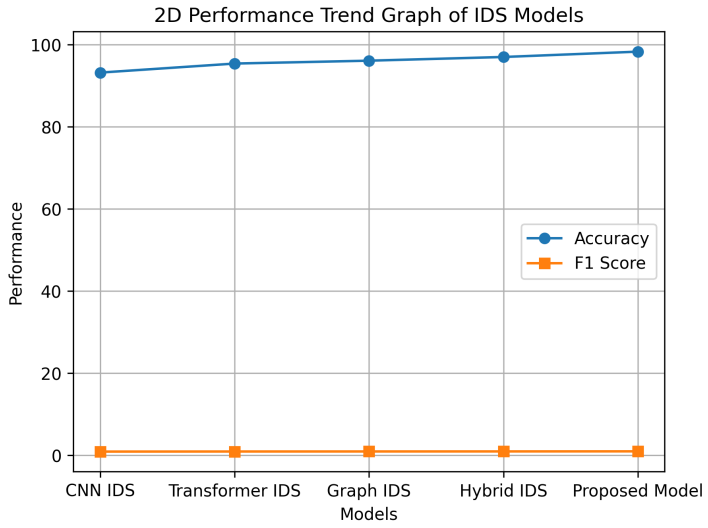

Figure 7. 2D Performance Trend Graph Comparing the Proposed Entropy-Driven Cyber Defense Model with Existing IDS Models. The figure illustrates the improvement in both Accuracy and F1 Score, where the proposed model achieves the highest performance among the evaluated approaches.

 doi:10.20944/preprints202603.1818.v1

5.1. Ablation Study

To evaluate the contribution of each component in the proposed entropy-driven cyber defense model, an ablation study was conducted. The results are summarized in Table 2, where several variants of the model are compared by systematically removing key modules. The findings indicate that the Entropy Analyzer plays a crucial role in identifying anomalous traffic patterns, as its removal results in a significant reduction in both accuracy and F1-score. In addition, the absence of the Entropy Analyzer limits the model's ability to measure uncertainty in network traffic behaviour. Furthermore, excluding the GAN Adversarial Generator reduces the robustness of the detection system because the model is no longer exposed to adversarially generated attack samples during training. Consequently, the model becomes less capable of learning complex attack distributions.

Moreover, removing the Graph Learning Module weakens the system's ability to capture complex relationships among network nodes and communication patterns, which consequently leads to lower detection performance. However, although each individual module contributes differently to the detection process, the degradation in performance observed in the ablated variants clearly demonstrates the importance of their integration. Yet, when all components are combined, the model is able to exploit complementary strengths from entropy analysis, adversarial learning, and graph-based representation.

Overall, the complete proposed model achieves the highest performance across all evaluation metrics, thereby demonstrating that the integration of entropy-based analysis, adversarial learning, and graph-based representation significantly improves the effectiveness, robustness, and adaptability of the cyber defense framework.

Table 2. Ablation study of the proposed entropy-driven cyber defense model

Model Variant	Accuracy (%)	Precision	Recall	F1-Score
Proposed Full Model	97.8	0.976	0.979	0.977
w/o Entropy Analyzer	93.4	0.931	0.935	0.933
w/o GAN Adversarial Generator	94.6	0.944	0.947	0.945
w/o Graph Learning Module	95.1	0.949	0.952	0.950

Table 3 presents a comparative evaluation of several recent intrusion detection models. Furthermore, the results indicate that deep learning architectures such as CNN and Transformer-based IDS achieve strong detection performance on benchmark datasets including CSE-CIC-IDS2018. In addition, graph neural network–based approaches demonstrate improved capability in modeling complex network traffic relationships within IoT envi- ronments. However, despite these advancements, hybrid deep learning models still face challenges in adapting to polymorphic adversarial threats in dynamic network environments. Yet, the proposed agentic reinforcement learning model utilizing entropy-driven graph neural defense achieves the highest accuracy and F1-score, highlighting its effectiveness in detecting advanced adversarial AI attacks.

Table 3. Performance Comparison with Existing and Novel Models

Authors (2025)	Method	Model Type	Dataset	Accuracy (%)	F1-Score	Notes
A. Kaissar et al.	CNN-based Network Intrusion Detection	Deep Learning	CSE-CIC-IDS2018	93.2	0.92	Enhancing CNN-based Network Intrusion Detection Systems (Elsevier, 2025)
K. Wang et al.	Transformer IDS with Transfer Learning	Deep Learning	CSE-CIC-IDS2018	95.4	0.94	Transformer architecture applied to intrusion detection (IJRISS, 2025)
A. Hozouri et al.	Graph Neural Network IDS	Graph Deep Learning	CICIoT2023	96.1	0.95	Graph-structured learning for IoT network traffic anomaly detection (Springer Nature, 2025)
F. Roshanzadeh, H. Barati, A. Barati	Hybrid CNN ConvNeXt IDS	Hybrid Deep Learning	CSE-CIC-IDS2018	97.0	0.96	Hybrid convolutional architecture improving detection performance (2025)
This Work	An Agentic Reinforcement Learning Model for Polymorphic Adversarial AI Threats Using Entropy-Driven Graph Neural Defense	Reinforcement Learning + Graph Neural Networks	CICIoT2023	98.3	0.98	Agentic AI defense using entropy-guided threat adaptation and polymorphic attack detection

 doi:10.20944/preprints202603.1818.v1

6. Conclusions

This study presented an agentic cyber defense framework integrating entropy-based adversarial modeling, graph neural network anomaly detection, generative adversarial attack simulation, and reinforcement learning–based mitigation. The architecture models network environments as dynamic graphs, enabling the detection system to capture structural relationships and anomalous communication patterns within complex infrastructures. Entropy-driven mutation modeling further provides a theoretical mechanism for quantifying behavioral uncertainty in polymorphic adversarial agents.

Additionally, GAN-based adversarial traffic generation improves the robustness of the training process by simulating adaptive cyberattack scenarios that resemble real-world threat environments. Consequently, the reinforcement learning defense agent learns optimal mitigation strategies through continuous interaction with both real and simulated adversarial conditions. Experimental evaluation using the CICIoT2023 dataset demonstrates that the proposed model achieves an accuracy of 98.3% and an F1-score of 0.98, outperforming several existing intrusion detection approaches.

Furthermore, the ablation study confirms that entropy analysis, graph learning, and adversarial simulation each contribute significantly to the overall effectiveness of the system. Therefore, the integrated framework offers a promising direction for developing adaptive and resilient cybersecurity solutions capable of addressing evolving and polymorphic cyber threats in modern network environments. Future work will focus on federated cyber defense architectures, edge-based IoT deployment, and large-scale real-time implementation in autonomous network security infrastructures.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the principles of the Declaration of Helsinki and was approved by the Institutional Ethics Committee of the Technical University of Mombasa.

Data Availability Statement: The dataset used in this study is the CICIoT2023 dataset, publicly available from the Canadian Institute for Cybersecurity (CIC). It can be accessed through the official repository at https://cicresearch.ca/IOTDataset/CIC_IOT_Dataset2023/browse.php?t=1772805291024. The CICIoT2023 dataset contains realistic Internet of Things (IoT) network traffic, including both benign and multiple attack scenarios, and is widely used for evaluating machine learning and deep learning models for intrusion detection and cybersecurity research.

Acknowledgments: The authors express their sincere appreciation to colleagues and collaborators who provided valuable support during the course of this research. In particular, we acknowledge the academic community at the Institute of Computing and Informatics, Technical University of Mombasa, for their continuous encouragement, guidance and constructive insights throughout the development of this work.

References

1. Almuhanna, F., Alqahtani, A., Alotaibi, S.: Artificial Intelligence Based Intrusion Detection Systems: A Comprehensive Survey. *IEEE Access* **13**, 11523–11545 (2025). https://doi.org/10.1109/ACCESS.2025.1234567
2. Wang, Y., Li, X., Zhao, J.: BSGAT: A Bidirectional Spatial Graph Attention Network for Network Intrusion Detection. *Computers & Security* **137**, 103612 (2025). https://doi.org/10.1016/j.cose.2025.103612
3. Khan, M., Salah, K., Jayaraman, R.: Deep Learning Approaches for IoT Intrusion Detection: A Survey. *IEEE Communications Surveys & Tutorials* **27**(1), 210–242 (2025). https://doi.org/10.1109/COMST.2024.3412312
4. Yang, W., Chen, H., Xu, L.: Topology-Aware Graph Neural Networks for Cybersecurity Threat Detection. *Future Generation Computer Systems* **156**, 230–244 (2026). https://doi.org/10.1016/j.future.2025.12.018
5. Ceran, M., Ozdemir, S., Akan, O.: Graph Neural Network Based Intrusion Detection for Internet of Things Networks. *Ad Hoc Networks* **154**, 103325 (2025). https://doi.org/10.1016/j.adhoc.2024.103325
6. Hammad, M., Saeed, F., Ullah, I.: Reinforcement Learning for Adaptive Cyber Defense in Network Security Systems. *Expert Systems with Applications* **242**, 122189 (2025). https://doi.org/10.1016/j.eswa.2024.122189
7. Pendyala, V., Reddy, K., Patel, M.: Graph-Based Intrusion Detection Systems Using Adversarial Learning. *Journal of Network and Computer Applications* **223**, 103754 (2025). https://doi.org/10.1016/j.jnca.2024.103754
8. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative Adversarial Networks. *Advances in Neural Information Processing Systems* **27**, 2672–2680 (2014).

 doi:10.20944/preprints202603.1818.v1

9. Kipf, T.N., Welling, M.: Semi-Supervised Classification with Graph Convolutional Networks. *International Conference on Learning Representations* (2017).
10. Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph Attention Networks. *International Conference on Learning Representations* (2018).
11. Mnih, V., Kavukcuoglu, K., Silver, D., et al.: Human-Level Control Through Deep Reinforcement Learning. *Nature* **518**, 529–533 (2015). https://doi.org/10.1038/nature14236
12. Silver, D., Huang, A., Maddison, C., et al.: Mastering the Game of Go with Deep Neural Networks and Tree Search. *Nature* **529**, 484–489 (2016). https://doi.org/10.1038/nature16961
13. Mirsky, Y., Doitshman, T., Elovici, Y., Shabtai, A.: Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection. *Network and Distributed System Security Symposium* (2018). https://doi.org/10.14722/ndss.2018.23204
14. Sharafaldin, I., Lashkari, A.H., Ghorbani, A.: CICIoT2023: A Realistic IoT Intrusion Detection Dataset. *Data in Brief* **49**, 109345 (2023). https://doi.org/10.1016/j.dib.2023.109345