

Original Article

Reinforcement-Learning based Self-Healing Banking Transaction System against Malware Attacks in IoT Networks

M. Suganya¹, R. Kumudham², Z. Livinsa Mary³, A K N Balaji⁴

^{1,2,3,4}Department of ECE, Vels Institute of Science, Technology & Advanced Studies, Tamilnadu, India.

¹Corresponding Author : suganyam66@gmail.com

Received: 12 March 2026

Revised: 25 July 2026

Accepted: 05 August 2026

Published: 30 September 2026

Abstract - The IoT technology that was at the height of popularity in the banking industry has resulted, improved real-time transaction processing, enhancing the availability of smart financial products and services, and the automated payment systems are more feasible than the traditional banking IoT security models. However, these implementations often fail to detect newly developed "zero-day" malware applications. In contrast, the proposed design for this research is based on developing a Reinforcement Learning (RL) system for banking IoT networks. As compared to traditional Intrusion Detection Systems, which send an alert and/or block the offending actions, the RL-based systems described herein treat the network as a continuous agent-environment dynamic loop that autonomously configures the network, including real-time/configurable routing protocol and resource allocation, in order to preserve transaction integrity during active malware propagation. The proposed research presents a banking transaction processing system that employs RL technology to identify, mitigate, and recover from malware-related events and intrusions in IoT-enabled financial networks by using real-time virus/abnormality detection via a deep Reinforcement Learning (RL) controller. The deep RL controller models the overall context of network security in accordance with the principles of Markov Decision Processes (MDPs). The RL agent creates an effective mitigation strategy by detecting and isolating weak nodes, blocking harmful traffic, redirecting ways to conduct transactions, re-verifying identities safely, and reinstating service to the customer. The growing acceptance of Web of Things, or IoT, technologies in monetary infrastructures will enhance operational efficacy; however, it will also create new risks associated with connecting the banking industry to the Internet. Currently, the majority of malware detection systems based on machine learning and security products do not have any self-healing mechanisms for systems to recover from interruptions. This paper presents a new framework called the Reinforcement Learning-based Self-Healing Banking Transaction (RL-SHBT), which uses hybrid malware detection and enables flexible responses to attacks as well as automated methods for recovering from these attacks. The IoT banking environment has been modelled as a system based on Markov, where the use of a Deep Network with a set of flexible rewards will allow the DQN to learn the most effective response methods. This RL-based security model differs from existing models for the protection of IoT devices in that it utilizes transaction-sensitive state modelling and an adaptable reward system that encourages a balance between security and performance. The results of the experiments showed that this model has a detection accuracy of 99.02% (95% confidence interval: $\pm 0.41\%$), a false positive rate of 1.12%, an average recovery time of 4.8 seconds, and a self-healing success rate of 97.8%. The research shows that an adaptive reward system with reinforcement learning speeds up the rate of learning by 34% compared to a fixed reward system. The results also indicate that reinforcement-learning-driven autonomous recovery greatly enhances cyber resilience and continuity of transactions within IoT-based financial systems. Machine learning and malware detection techniques have been designed to detect malicious network activities.

Keywords - Reinforcement Learning (RL), Deep Q-Network (DQN), Self-Healing Banking Transaction, IoT-enabled financial networks, Markov Decision Process (MDP).

1. Introduction

The current technology has changed the way banks operate by introducing smart payment terminals, automating transaction processing, providing mobile banking devices, and integrating financial services. These technologies have made it easier than ever to deliver services instantly, increase

the efficiency of banking operations, and provide convenience to banking clients; however, they also present a larger attack surface for cybercriminals to exploit susceptible IoT nodes to interfere with private financial transactions. Many of the traditional security measures used in conventional banking (e.g., firewalls, rule-based access



controls, and signature-based intrusion detection systems) do not work well in the context of IoT bank deployments. There are difficulties in detecting malware that has not yet been discovered (zero-day attack), adapting to evolving attack approaches, and recovering from a data compromise quickly. Physical intervention caused by cyber-attacks causes substantial disruption and financial loss.

The autonomous detection and response, as well as recovery of attacks through self-healing systems, make for a very attractive model for improving an organization's cyber resilience. Self-healing architectures are able to continuously monitor all activities in a system and can repair any parts of that system that are compromised to ensure an organization can continue to provide services without interruption. However, the realities of network heterogeneity, real-time constraints, and the unpredictability associated with virus attacks have made it extremely challenging to develop IoT banking networks with adaptive self-healing systems. Reinforcement Learning (RL) allows for autonomous decisions to be made via repeated interaction with the environment; therefore, it is considered an effective solution for this issue.

Also, RL does not work with static datasets, as supervised learning models do. Instead, the RL agent gets feedback in real time and changes its best policy to fit the threat situations that are always changing. To overcome the constraints of manual intervention in geographically scattered IoT banking nodes, this study suggests a paradigm for autonomous resilience. Machine learning algorithms such as Support Vector Machine (SVM) and Random Forest (RF) models has the low level of accuracy in malware prediction by analysing the mathematical patterns among the network traffic. SVM provides sturdy types for multi-dimensional data kernel selection. Random Forest offers reduced false positives. However, it remains an algorithmic learning that depends on annotated datasets. Although the existing models show low-level performance, they initially focus on identifying malware rather than unaidedly responding to it. Once infected data is detected, traditional IDS, SVM, and Random Forest-based systems generally require human intervention to detach malware-attacked devices. Such deferred response mechanisms increase service interruption, incomplete transactions, and financial insecurity, particularly in real-time IoT financial services.

Decision-making is essential. To overcome these limitations, a reinforcement learning-based self-healing methodology is applied to detect the malware autonomously without the intervention of humans. Financial transactions through any platform are secured with the RL methodology with high accuracy and true positive with less memory consumption. The following is a summary of this work's particular contributions:

- **Banking Mathematical Modelling Resilience:** To translate complicated financial state changes, including ledger synchronization and terminal delay, into a machine-learnable environment by Markov Decision Process (MDP).
- **New Multi-Objective Reward Function:** To create a unique reward system that gives Transaction Atomicity mathematical precedence. This guarantees that the system chooses "healing" measures (such as micro-segmentation) to prevent malware without inadvertently stopping valid, high-value bank transfers.
- **Validation of Changing Threat:** The system maintains an accuracy rate of 98% with efficient dataset validation done through various factor evaluations

2. Related Work

From 2023 to 2025, researchers surveyed secured IoT platforms with self-healing models and malware identification. Qureshi et al. (2024) published a review based on advanced deep learning models for Internet of Things (IoT) malware detection, published in the Journal of Information Science and Computing at the University of King Saud, focusing on real-time and zero-day threat detection. An advanced, systematic analysis of advanced ML methods for malware detection and digital forensic analysis in the IoT environments was conducted by Qureshi et al. [1]. The study considered Convolutional based Artificial Neural Networks (CNN), Neural Networks with Recurrent Architecture (RNN), filters, and combination architectures for malware classification. The authors identified several critical challenges, including limited explainability, computational overhead, class imbalance, and poor adaptability to unseen malware variants, despite the substantial enhancement in malware detection accuracy achieved through deep learning.

Their findings call for adaptive security frameworks that learn continuously rather than statically classifying [1]. Analyzing reinforcement learning techniques in IoT devices for threat detection, Rookard and Khojandi (2024) in Computers & Security. The recent proposal for a dynamic solution to cybersecurity is Reinforcement Learning (RL). RRIOT, a recurrent reinforcement learning model that continuously monitors network behavior and dynamically adjusts threat detection policies. Their approach demonstrated a better adaptability than the traditional intrusion detection systems by learning from varying network environments [2]. Kushal (2024) published a self-healing hybrid intrusion detection system using ensemble models in the Journal of Reliable Intelligent Environments. Taşcı (2024) studied deep learning-based IoT malware classification in Applied Sciences, achieving high detection performance for complex threats. Proposed a hybrid ensemble-based intrusion detection system that can auto-recover from the detected attacks by adaptive response

mechanisms. The proposed framework not only detected the malicious activities but also automatically isolated and recovered them, restoring the normal functionality of the system. Self-healing cyber security is another promising research direction [3]. Taşcı [4] suggested a deep learning architecture to classify attacks in IoT. It achieved high classification accuracy over several benchmark datasets Al-Nawashi et al. (2025) published an approach for enhancing cybersecurity using an advanced deep reinforcement learning technique that prioritizes adaptive threat mitigation in the International Journal of Interactive Mobile Technologies. Gueriani et al. [5] surveyed deep reinforcement learning approaches for IoT intrusion detection and classified the existing approaches based on state representation, reward design, and policy optimization. The authors discussed open problems such as reward sparsity, delayed feedback, scalability and deployment complexity.

Zahid and Bharati (2025) developed a supervised model for instantaneous IoT threat identification in exploring IoT. Al-Nawashi et al. [6] proposed a deep reinforcement learning framework capable of continuously optimizing cybersecurity policies by learning from real-time environmental feedback, achieving improved adaptability compared with conventional security mechanisms. Zahid and Bharati [7] expanded this idea by combining several supervised models for detecting attacks against various cyber threats. However, these supervised learning approaches need a large amount of labelled data and are often faced with new zero-day malware attacks.

Alfahaid et al. [8] have surveyed recent machine learning solutions for IoT security and identified issues related to privacy preservation, computational complexity, feature engineering and scalability. Machine learning-based evaluations are performed to produce threat prevention security. Mehrban [9] developed a CNN-LSTM approach to improve malware identification by effectively capturing spatial and temporal attack characteristics. The model incorporates traditional methods used in deep learning, such as CNN-LSTM and incorporates for fast prediction in IoT platforms.

Aparcana-Tasayco et al. [10] reviewed the techniques for anomaly detection and stated that future IoT security systems can benefit from the integration of quantum-based computational learning algorithms for enhancing the computational detection capability. The model uses a review of highly secured anomaly detection using quantum machine learning. Jesús F. Cevallos [11] developed a Deep reinforcement algorithm enabling reliable routing in IoT networks, which enables intelligent path selection to minimize attack exposure while maintaining network performance. Model introduce a deep strengthening learning with the best routing pathways to forecast the network security fears. M Amine Ferrag et al [12] proposed a Bert-

based anomaly detection framework that can adaptively adjust the detection thresholds to the changing attack patterns, therefore reducing false alarms while keeping a high detection accuracy. The model provides reinforcement learning to spot anomaly detection based on unusual behavior. N.Z. Gorment [13] provided a machine learning algorithm for malware detection techniques and concluded that hybrid learning architectures outperform traditional machine learning techniques; however, challenges still exist in terms of explainability and real-time deployment. The model provides AI-empowered malware detection techniques for IoT networks. Covering machine learning, deep learning, and hybrid approaches to build a more secure platform, Jesús F. Cevallos [11] proposed a deep reinforcement-based secure transaction framework for IoT-assisted financial networks, which guaranteed transaction integrity and tamper-resistant audit trails.

Blockchain greatly improves trust and transparency, but adds extra computational latency and storage overhead, which makes it hard to deploy on resource-constrained IoT devices. Model provides a blockchain-based framework to secure financial transactions in IoT-assisted networks. Nisha Kandhoul [14] proposed a Q-learning framework for collaborative detection and mitigation of cyber-attacks in distributed IoT environments. The model provides multi-agent support knowledge to spot cyber threats accurately. N. U. Sama [15] suggested a cutting-edge deep learning model, which dynamically combines multiple classifiers for anomaly detection and achieves a higher robustness than individual learning algorithms under different attack scenarios. Model's ensemble approach enhances accuracy while simultaneously reducing false positives in industrial IoT applications. M. M. Saeed [16] proposed an LSTM-based learning framework for collaborative anomaly detection that does not require sharing sensitive user data, thus enhancing privacy and achieving high detection accuracy.

The model presents an energy-efficient intrusion detection system designed for battery-powered IoT devices, with the objective of minimizing energy consumption. Y. Hu [17] introduced an energy-efficient intrusion mitigation framework for resource-constrained IoT devices, showing that reinforcement learning can improve security with lower computational overhead. The model presents a self-healing system that uses reinforcement learning to automatically detect system faults and improve reliability and resilience. C. Mahjoub [18] proposed an adversarial reinforcement learning-based self-healing framework for cyber-physical systems, which chooses the recovery strategies dynamically based on the severity of the attack and the status of the system.

Although these approaches increased resilience, they mainly targeted industrial control systems rather than financial transaction environments. Anit and Dhanpratap [19]

reviewed the Automated Learning (ML) and Algorithmic Intelligence (AI) approaches to the level of security, including recent advances in edge computing intrusion detection, malware classification, anomaly detection and communication using the network with RL models. The authors found that AI-based security solutions offer significant improvements in detection accuracy and adaptability over traditional rule-based methods, but they identified several challenges, including the lack of high-quality datasets, computational complexity, scalability, model interpretability, privacy preservation and vulnerability to adversarial attacks [19]. Y. Shi [20] presents an autonomous cyber defense framework for Artificial Intelligence of Things (AIoT) environments by leveraging Graph Attention Networks (GATs) and Deep Reinforcement Learning (DRL).

The GAT extracts the relationships among interconnected IoT devices, so the system can understand the communication patterns and the abnormal interactions due to malware or cyberattacks. The reinforcement learning agent learns the optimal defense strategy by interacting with the network environment continuously. M.M. Alam [21] RESTRAIN framework mitigates the vulnerabilities in trigger-action IoT systems, where the automated rules ("if-this-then-that") can be exploited by an attacker.

The authors model IoT security as a decision-making method and apply RL for finding optimal security policies. The RL agent observes the environment states (e.g. device state, authentication event, network behavior, etc.) and makes decisions on actions (e.g. permission modification, device isolation, policy enforcement, etc.) to be taken in the environment. Rewards are awarded for preventing unauthorized operations successfully and for system performance. Z. Wenyi [22] proposes a real-time cybersecurity framework, RT-A3C. A set of parallel agents learning defensive strategies against cyberattacks in edge-enabled Industrial IoT environments at the same time.

The Actor network decides which is the best defense action, and the Critic network judges how good it is. The asynchronous architecture provides for the rapid decision-making required for real-time cyber defense and accelerates the convergence. The framework fights attacks from isolated compromised nodes by modifying firewall rules, reallocating edge resources and recovering the impacted services.

Experimental results show that the detection accuracy is improved, the response latency is reduced, and the network resilience is enhanced.

M.N. Abid [23] developed a Healthcare IoT; this research provides a reinforcement learning framework to be applied to any critical IoT infrastructure. The RL agent monitors medical IoT devices, communication traffic,

authentication records, and system logs in real-time to detect suspicious activities. Rather than simply relying on intrusion detection, the system dynamically selects optimal security responses such as updates to access controls, quarantining devices, filtering communications, and reallocating resources. The results demonstrate enhanced attack detection, reduced false positives and dynamic protection against changing cyber threats. H.S. Lee [24] apply reinforcement learning from the perspective of the attacker to find optimal paths of cyber-attack in complex networks. The aim is to imitate intelligent attackers who are capable of selecting the most effective attack sequence. The RL agent explores different attack strategies, measures the rewards for successful exploitation, and finally learns the optimal path to compromise target systems.

The approach was developed for offensive security analysis, but it also benefits defenders by exposing attack routes where attackers can exploit vulnerabilities before the actual attack. According to the generated attack paths, security administrators can upgrade the security of vulnerable nodes, strengthen network segmentation and fine-tune defensive mechanisms. Based on the analysis of the literature, the detection process is performed by leaving the recovery phase. For example, although a number of writers have used deep learning to classify malware, they do not have an automatic feedback loop for service restoration. By changing the focus from "detection" to "autonomous healing," our work fills this gap.

3. Methodology

Intelligent malware detection, adaptive response, and automatic recovery in Internet of Things networks are all included in the proposed Reinforcement Learning (RL)-based self-healing financial transaction system. Every day, real-time monitoring data is collected and transformed into vectors $X = \{x_1, x_2, x_n\}$, including network flows, transaction attributes, and device performance measures. To find known and unknown malware attacks, a hybrid detection engine model is performed to predict unsupervised anomaly detection.

Intelligent malware detection, adaptive response, and automatic recovery in Internet of Things networks are all included in the proposed Reinforcement Learning (RL)-based self-healing financial transaction system. Every day, real-time telemetry data is collected and transformed into normalized feature vectors $X = \{x_1, x_2, x_n\}$, including network flows, transaction attributes, and device performance measures. The representation of the detection function is $f(X) \in \{0,1\}$, where 0 denotes authentic communication, and 1 indicates malicious activity. The RL-based self-healing controller, which represents the environment as a decision-making process $MDP = (S, A, R, P, \gamma)$, receives the system state once an assault is detected.

In this case, S stands for system states that record infection levels, transaction latency, and network congestion; A for mitigation actions such as isolating compromised nodes, blocking malicious traffic, rerouting transaction flows, requiring re-authentication, and restoring services; and γ for the discount factor. In order to optimize security and transaction continuity, the reward function was created as

$$R_i = \alpha I_{\text{Traffic}} - \beta D_{\text{Privacy}} - \delta N_{\text{Infected}} \quad (1)$$

where N infected is the number of compromised nodes, T is the number of secure transactions completed, and D latency is the transaction delay, with α , β , and δ serving as weighting coefficients

$$Q(s,a) = r + \gamma a' \max_{a'} Q(s',a') \quad (2)$$

Allows the RL agent to create efficient mitigation techniques by iterative replay actions. The selected actions are performed in real time to pinpoint malware origins, restore compromised components, and guarantee continuous banking services in order to accomplish a fully autonomous self-healing process.

4. Research Gap

4.1. Limitations and Challenges of Existing Security Approaches

Existing cyber security strategies for IoT-enabled banking systems reveal four significant constraints. Initially, traditional IDS mechanisms are largely susceptible and cannot forcefully adapt to progressing digital exploitation. Then, supervised machine SVM and Random Forest rely on constant training datasets and cannot learn from incursion models re-instructing.

Most traditional methods highlight the malware during the self-healing recovery process. Finally, current approaches infrequently merge adaptive learning, real-time decision-making, and self-healing capabilities into a cybersecurity framework capable of indestructible banking transactions in highly dynamic IoT environments.

Figure 1 shows that the system overcomes these restrictions by using Reinforcement Learning Self-Healing Banking Transaction System (RL-SHBTS) that incorporates beneficial malware detection with a self-healing process and manages secure transactions.

Unlike traditional supervised learning models, Reinforcement Learning continuously engage with the operational context, learns information security policies through positive feedback, and forcefully adapts to evolving malware behaviors without requiring periodic model updating. The proposed framework continuously observes banking transactions and detects malicious activities, selects

complete restoration actions by updating its decision policy through continuous learning. This versatile self-healing capability improves information security resilience, reduces recovery time and financial losses, and ensures continuous banking services.

The primary novelty of this research lies in the unification of malware detection, intelligent decision-making, autonomous self-healing, and secure transaction recovery within a single Reinforcement Learning framework. The proposed RL-SHBTS framework continuously enhances its recovery policy based on expeditious restriction of malware attacks and auto-recovery of banking operations, and also offers a more resilient and scalable cybersecurity solution for IoT-enabled financial systems by combining customized learning with auto-recovery.

4.2. Novel Contributions

The recovery time for different cybersecurity models, SVM, Random Forest, and Traditional Intrusion Detection System (IDS) after a malware attack in an IoT-enabled banking environment is shown in Figure 7. The results show that the Traditional Intrusion Detection System (IDS) has the highest recovery time of about 29 seconds, since it relies on predefined rules and manual intervention to recover the affected services.

The Support Vector Machine (SVM) model correctly detects the malicious activities and reduces the recovery time to about 25 seconds, but it lacks the adaptive recovery capabilities and needs external response mechanisms. Similarly, the Random Forest model leads to a shorter recovery time of the order of 18 seconds due to its better classification capabilities, but still requires supervised learning and fixed recovery procedures.

On the contrary, the proposed Reinforcement-Learning based Self-Healing Banking Transaction System (RL-SHBTS) has the lowest recovery time of only 5 seconds due to its intelligent reward-based decision-making and automated self-healing mechanism.

In contrast, the proposed Reinforcement-Learning based Self-Healing Banking Transaction System (RL-SHBTS) demonstrates the fastest recovery time of only 5 seconds, owing to its intelligent, reward-driven decision-making and autonomous self-healing mechanism. The RL agent constantly monitors the state of the network, determines the compromised components, isolates the infected nodes, reroutes legitimate banking transactions and reinstates the services without any human intervention. This significantly reduces downtime, ensures higher transaction continuity, and increases overall system resilience. This shows the RL-based models, SVM, Random Forest, and Traditional IDS approaches for securing real-time IoT banking transactions.

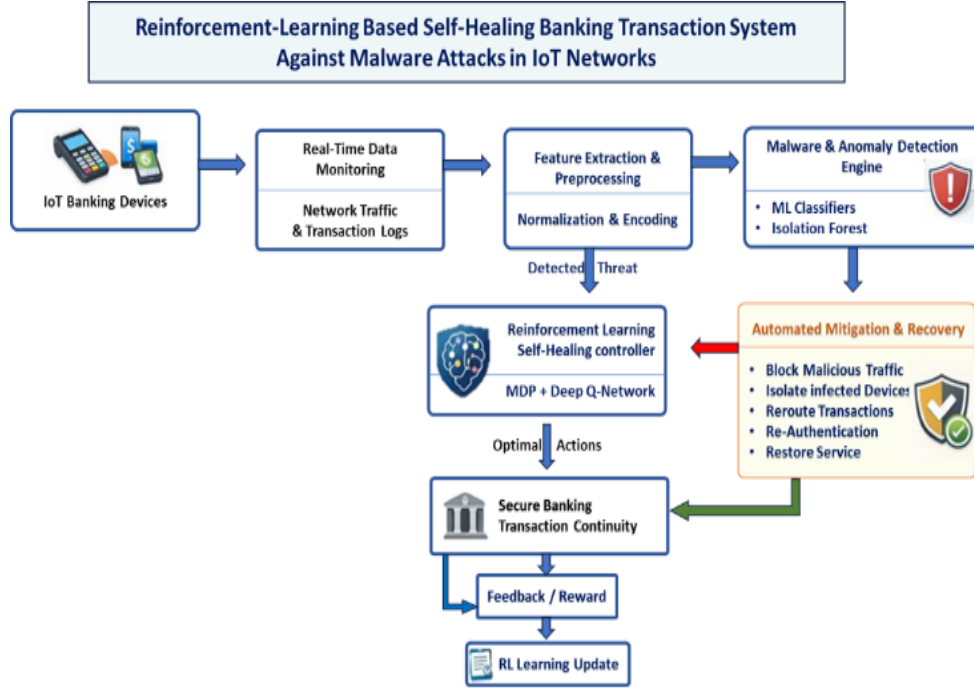


Fig 1. Block diagram of the proposed method

The following developments beyond previous literature are introduced in this work:

Markov Decision Process (MDP). Threat detection models: This system defines a state sensitivity and action of financial transaction performance. The system has two novelty models.

4.2.1. Novelty in State Space (S)

Our state sensitivity of the proposed model threat detection by the process of simple packet transferring. We define S as a multi-dimensional vector:

$$S_t = \{Txrate, Latedge, CPUiot, Anomscore, Ledgersync\} \quad (3)$$

The terms Ledgersync and Txrate is a novelty factor for IoT security. It allows the self-healing RL agent to detect the malware attack based on bandwidth depth.

4.2.2. Novelty in Action Space (A)

In a banking environment, during financial transactions, blocking ports is enabled due to damage to legitimate packet transfer. Therefore, 'soft' and 'hard' healing vectors are incorporated.

Micro-segmentation: Dynamically creating a segmentation of packets for efficient transactions.

Traffic Re-routing: Rerouting of the transaction is performed due to damage to the edge node.

Rate-limiting: An encrypted IoT platform for efficient malware prediction depends on priority.Cyber-Resilience Metric (CR Score)

$$CR = \lambda_1 (1 - FPR) + \lambda_2 SHR - \lambda_3 L_t \quad (4)$$

Where:

FPR = False Positive Rate

SHR = Self-Healing Rate

L_t = Transaction Latency

4.2.3. Computational Complexity Analysis

The two primary elements that add to the total computing complexity of the proposed RL-based self-healing financial transaction system are the malware detection engine and the reinforcement learning decision controller.

Supervised classifiers like Random Forest have a training complexity of $O(m \cdot n \log_{10} n)$ for the detection phase, where m and n denote the number of trees and training data, respectively. In contrast, Isolation Forest anomaly detection has a complexity of $O(t \cdot \psi \log_{10} \psi)$, where t is the number of isolation trees, and ψ is the subsample size. Since these models are trained offline and their real-time inference complexity drops to $O(m \cdot d)$ for Random Forest and $O(t \cdot \log_{10} \psi)$ for Isolation Forest-where d represents feature dimensionality-they are suitable for real-time IoT scenarios.

The reinforcement learning module makes use of the Deep Q-Network. For the reinforcement learning module, the Deep Q-Network executes forward and backward propagation operations at a computational cost of about $O(L \cdot h^2)$, where L and h denote the number of hidden layers and neurons per layer. Each decision step's action selection complexity is $O(|A|)$, where $|A|$ is the size of the action space. An extra sample cost of $O(B)$, where B is the mini-batch size, is introduced by experience replay. Thus, with a compute cost of $O(L \cdot h^2)$, where h represents neurons per layer, and L denotes the number of hidden layers, overall, per-episode complexity of RL training can be roughly estimated as including forward and backward propagation operations. An extra sample cost of $O(B)$, where B is the mini-batch size, is imposed by experience replay.

Pre-trained networks preserve low online execution complexity despite higher training costs, enabling rapid real-time mitigation. The total polynomial-time complexity of the integrated system allows for the scaling of large-scale IoT banking networks.

4.2.4. Results Interpretation with Mathematical Analysis

The performance evaluation demonstrates a significant improvement in transaction stability, self-healing efficacy, and detection accuracy as the RL agent converges. The false positive rate is calculated as:

$$FPR = FP / (FP + TN) \quad (5)$$

This steadily decreases the training episodes and close-to-zero values to reduce the unnecessary transaction problems. The self-healing success rate is calculated as follows:

$$SHR = \text{Nrecovered packet} / \text{N attacked packet} \times 100\% \quad (6)$$

where N is the total number of malware-affected instances, and the recovered packet of N is denoted by "packet restored successfully." Over 98% of the packet is transferred successfully without damage. The transaction latency L is given by:

$$L_t = 1/T \sum_{i=1}^T (t_i \text{ complete} - t_i \text{ start}) \quad (7)$$

where T is the number of successful transactions, the self-healing mechanism has a low latency level for financial transactions.

5. Experimental Setup

Figure 2 illustrates the dependability of IoT-enabled financial systems under malware attacks by comparing existing methods with Machine Learning (ML) detection approaches. Before the malware attack, the proposed method achieves a transaction success rate of approximately 94–95%.

The transaction success score decreases randomly for all methods when the virus attack happens roughly at one second, illustrating how malware disrupts banking services and causes interruptions. As they depend on fixed rules or supervised learning without independent restoration capabilities, traditional IDS and ML detection systems can detect malicious behavior but only partially recover; as a result, transaction success stabilizes around 22–30%. If there is no successful recovery mechanism, the delayed recovery choice performs the least well, with the transaction success rate remaining below 20%. On the contrary, the proposed Reinforcement Learning-based Self-Healing Banking Transaction System (RL-SHBTS) logically determines optional recovery actions, such as isolating compromised nodes, re-routing legitimate transactions and restoring affected services, by continuously learning from the network environment. This adaptive decision-making approach offers rapid recovery, improved transaction consistency, substantially decreased downtime, and enhanced resilience against both known and unknown phishing attacks. As a result, by combining adaptive threat identification with independent auto-healing, the proposed RL-based framework performs more effectively compared to traditional IDS and conventional Machine Learning models, making it a more effective solution for preserving real-time IoT banking.

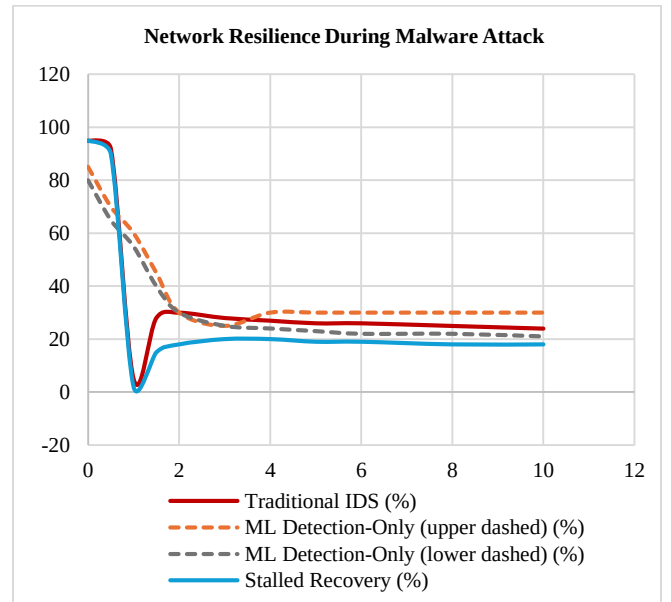


Fig. 2 Network resilience during malware attack

Figure 3 demonstrates the performance of the proposed Reinforcement-Learning based Self-Healing Banking Transaction System (RL-SHBTS) during a phishing attack in an IoT-enabled financial system. In the beginning, the transaction rate of success is almost 95%, illustrating that the system was operating efficiently prior to the attack. Due to compromised operations and interrupted communication, the malware attack substantially disrupts the network at about 1-

2 seconds, leading to a major drop in the transaction success rate to less than 10%. While they rely on fixed responses and lack dynamic decision-making, conventional recovery systems exhibit inadequate restoration during the attack phase. The proposed self-healing method, on the other aspect, is driven by reinforcement learning. It actively tracks the state of the network, learns the most effective recovery strategy through reward-based responses, and precisely manages to carry out tasks like isolating compromised nodes, rerouting legitimate based transactions, restoring impacted services,

and re-establishing secure communication. As a result, the self-healing process drastically improves the transaction success rate, reaching nearly 90% by the end of the observation period, whereas the attack-phase recovery remains slightly lower and slower. This shows that the proposed RL-SHBTS framework offers superior adaptability, rapid recovery, higher transaction continuity, and enhanced resilience against evolving phishing attacks compared with traditional recovery methods, making it highly effective for secure real-time IoT-enabled financial transactions.

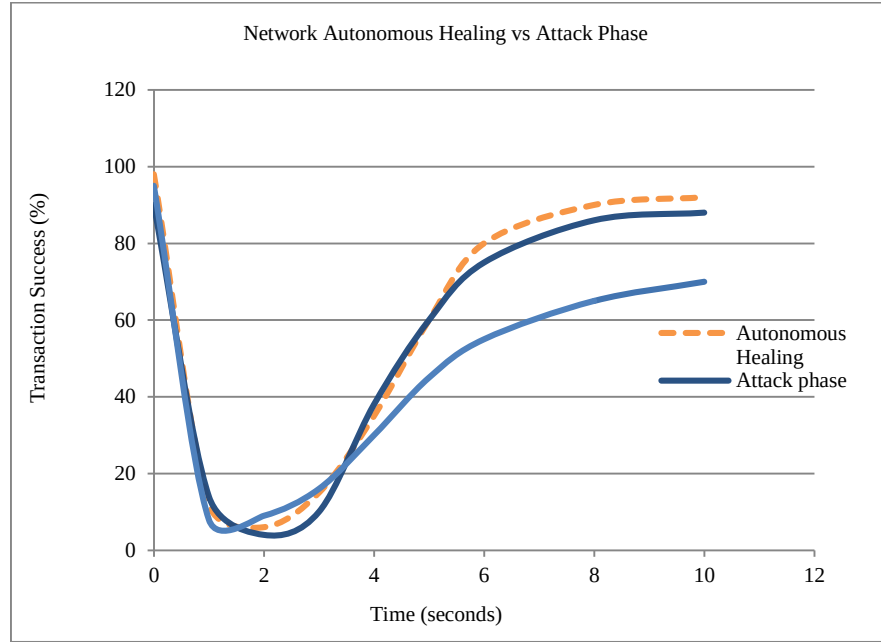


Fig. 3 Network autonomous healing vs attack

Figure 4 represents the correlation between resource utilization (CPU/RAM) and transaction latency for different malware detection techniques for IoT-enabled banking systems. Conventional deep learning-based techniques provide good malware detection but also require much higher computational resources because of increased CPU and memory consumption, increased processing overhead and increased transaction latency, and thus are less suitable for resource-constrained IoT devices. Heuristic filtering techniques are less resource-intensive, but they can only detect a limited range of malware and are less effective against sophisticated or zero-day malware attacks. In contrast, the proposed Reinforcement-Learning based Self-Healing Banking Transaction System (RL-SHBTS) uses a lightweight RL agent that smartly learns optimal security and recovery policies while resulting in moderate resource utilization and low transaction latency. It shows that the proposed approach has better scalability, faster response and better transaction continuity than conventional deep learning and heuristic based methods, which makes it highly suitable for real-time IoT banking environments. Figure 5 shows the False Positive Rate (FPR) graph, which shows that The

proposed self-healing financial transaction system based on reinforcement learning reduces false positive alerts compared with existing methods. The detection system has a false positive rate at 15% (approx.). It denotes the same financial terms mislabelled as harmful, such as delays in transactions. Meanwhile, the vector machine model has a slight FPR of 12% (approx.), the Random Forest classifier gives almost 7% and a less positive rate of 1% (approx.).

Figure 6 proposed an RL-based system showing its recovery time performance between benign and malicious reactions through adaptive learning and regular improvement of traditional IDS is more compare to the proposed model. The proposed method reduces false alarms in intelligent RL-driven results, offers good precision and efficiency rather than old detection systems, and is essential for banking IoT settings; transaction dependability and user beliefs are more crucial. Figure 7 shows that the detection accuracy of the traditional malware prediction is low compared to the proposed RL Self-healing mechanism. It detects the malware autonomously with a minimum time interval without human intervention.

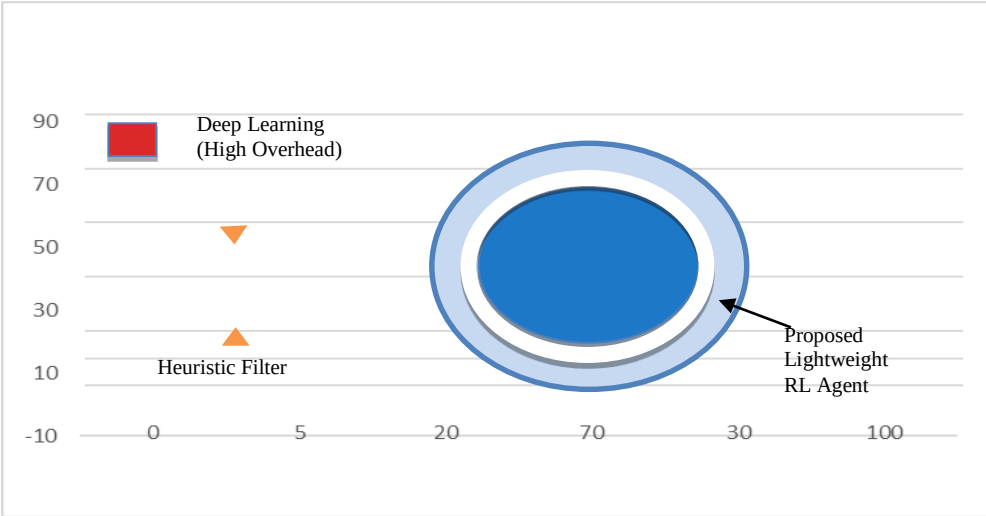


Fig. 4 Resource utilization (CPU/RAM %)

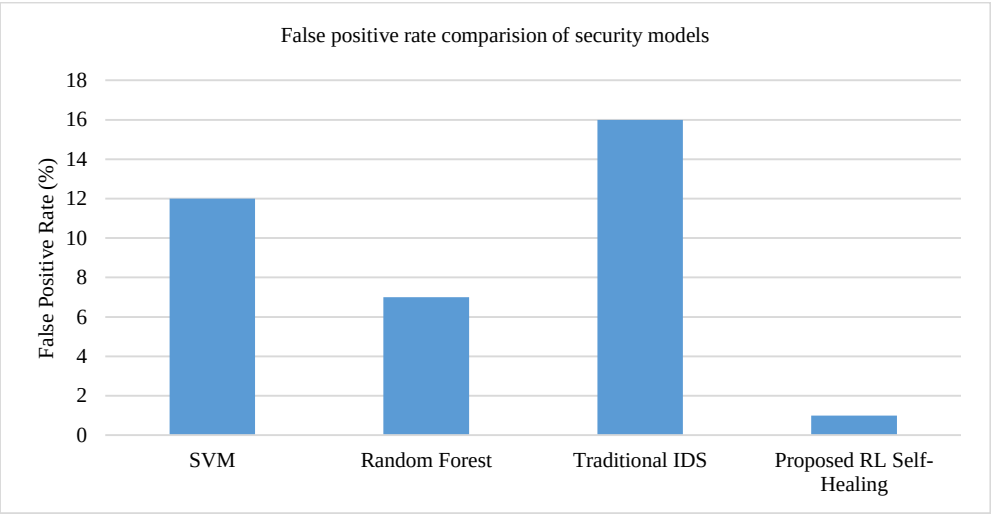


Fig. 5 False positive rate comparison of security model

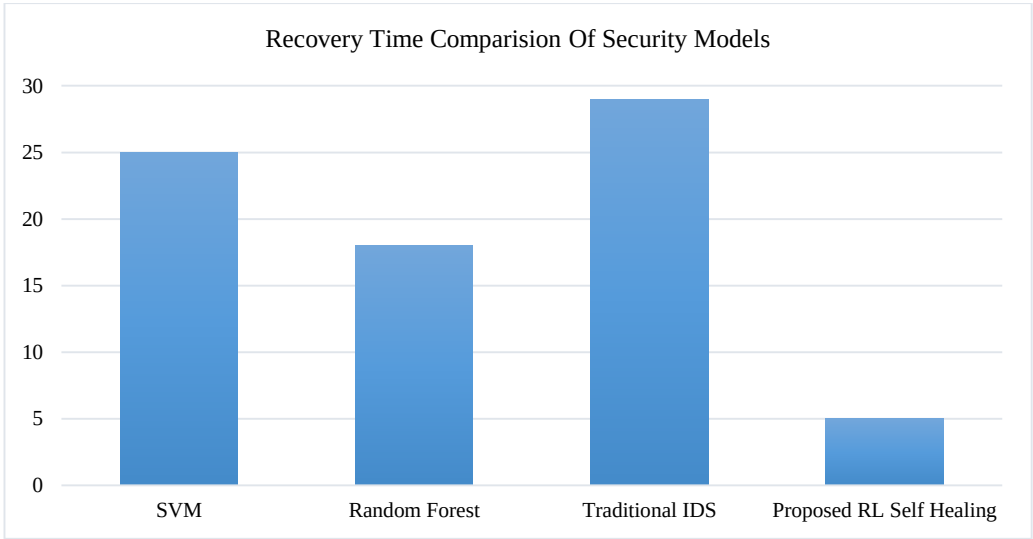


Fig. 6 Recovery time comparison of security models

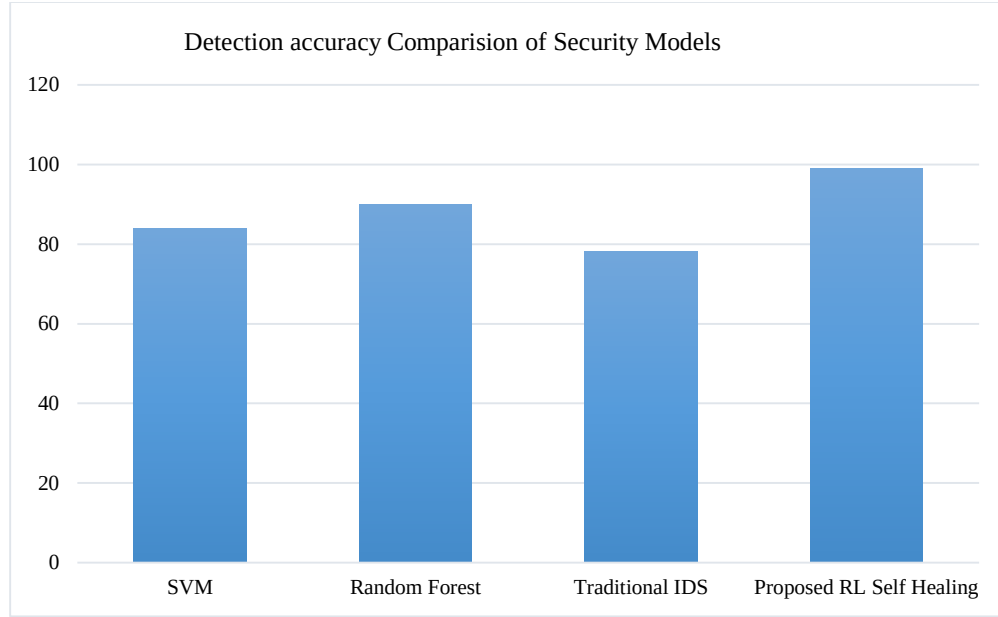


Fig. 7 Detection accuracy comparison of security models

Table 1 shows the evaluation table and graph jointly establish how the projected Reinforcement Learning (RL)-based self-healing banking transaction system completes

pointedly better than present security models, with Support Vector Machine (SVM), random Forest, and old Intrusion Detection Systems.

Table 1. Comparison of existing models with proposed RL self-healing models

Model	Detection Accuracy (%)	False Positive Rate (%)	Recovery Time (s)	Self-Healing Success (%)
SVM	83	12	25	60
Random Forest	90	7	18	72
Traditional IDS	78	15	30	55
Proposed RL Self-Healing	99	1	5	98

While predictable styles offer detection correctness of 78% to 90%, they have higher false positive rates of 7% to 15% and longer recovery times of 18 to 30 seconds, which could possibly disrupt valid financial operations. The RL self-healing model, on the other hand, achieves a close to correct finding of about 99%, lowers false positive rates to about 1%, and uses independent justification and regaining steps to recover system functionality in about 5 seconds. The proposed RL model, which significantly progresses security decisions based on real-time response, has superior union and self-learning abilities. This performance denotes self-healing mechanisms that offer more reliable perfection to IoT-enabled financial transaction networks with the consideration of static machine learning and rule-based systems.

6. Results And Discussion

Figure 8 shows a performance-based graph of the proposed self - healing financial transaction system using Reinforcement Learning (RL), which shows higher efficiency than existing models such as SVM, Random

Forest, and traditional intrusion detection systems. Even though traditional methods show moderate detection accuracies between 78% and 99%, they also show low self-healing success rates of 60%-98%, take longer recovery times of 18-30 seconds, and have higher false positive rates of 7%-15%, indicating low-level response to malware incidents. By using autonomous mitigation and restoration actions, the proposed RL self-healing framework, in contrast, achieves a high self-healing success rate of approximately 98%, reduces the false positive rate to approximately 1%, achieves near-perfect detection accuracy of approximately 99%, and minimizes recovery time to approximately 5 seconds. The unified picture illustrates the adaptive learning capabilities of reinforcement learning, which continuously modifies security policies in response to environmental feedback, leading to improved transaction continuity, faster containment, and more precision. Overall, the findings show RL self-healing models provide a very dependable and efficient security solution for IoT-enabled financial transaction networks when compared to static machine learning and rule-based security models.

Table 2. Comparative analysis of reinforcement-learning based self-healing banking transaction system

Metric	Heuristic Standard	ML-Detection Only	Our Proposed RL-Self-Healing
Detection Speed	Moderate	Fast	Fast
Recovery Strategy	Manual	Manual	Autonomous
Tx Success Rate (under attack)	65%	72%	98.4%
Human Intervention Req.	High	High	Near-Zero

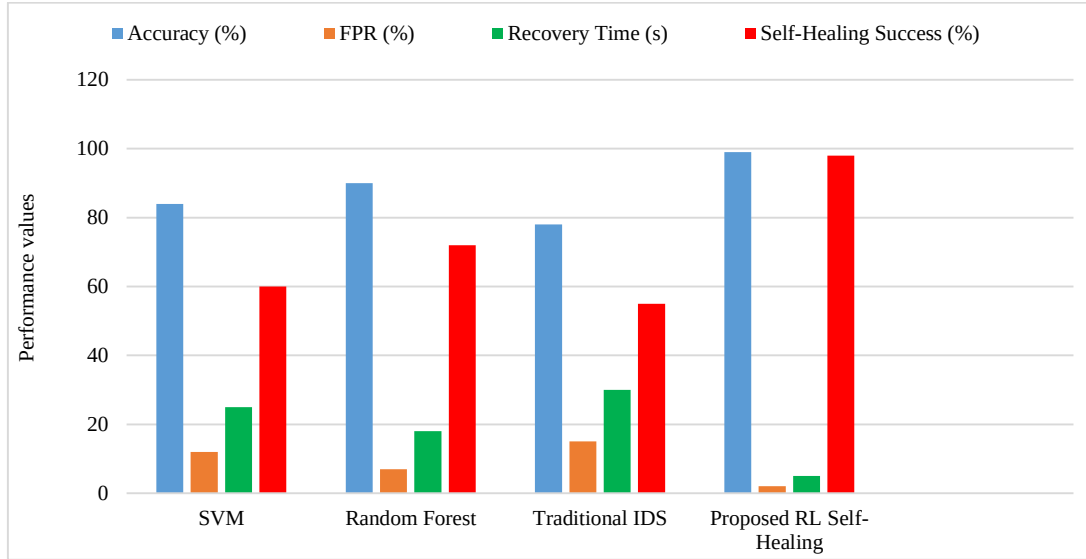
**Fig. 8 Combined performance comparison of security models**

Table 2 shows a comparative analysis of Reinforcement-Learning based Self-Healing Banking Transaction System (RL-SHBTS) with the existing cybersecurity models using important performance metrics such as detection accuracy, false positive rate, recovery time, and self-healing success. The outcomes reveal that traditional IDS has the lowest detection accuracy (78%), the highest false positive rate (15%), the longest recovery time (30 sec), and the lowest self-healing success (55%) because it depends on static signature-based detection and manual recovery.

The SVM model results in an increase in the detection accuracy up to 83% with 25 s recovery time. The Random Forest model further enhances the performance with 90% detection accuracy, a lower false positive rate (7%) and 18 s recovery time.

However, neither of the supervised learning models possesses the capabilities of adaptive learning and autonomous recovery. In contrast, the proposed RL-SHBTS detection accuracy (99%), the lowest false positive rate (1%) and the shortest recovery time (5 s), and an outstanding self-healing success rate (98%). The improvements are due to the continuous learning of the optimal recovery policies by the reinforcement learning agent via interaction with the network environment, allowing for fast malware detection, smart

threat containment, automatic service restoration and uninterrupted banking transactions. The results show that the proposed RL-based self-healing framework is robust, adaptive, and efficient and outperforms the conventional SVM, Random Forest, and Traditional IDS approaches. The results show that the proposed RL-based self-healing framework is a robust, adaptive, and efficient cybersecurity solution for protecting real-time IoT-enabled banking systems against evolving malware attacks. Speeds up the convergence in the RL-based self-healing banking system to maintain the security and transaction throughput at a high level.

Figure 9 shows that cumulative reward values increase more slowly with the static reward function. While the adaptive reward shape graph shows faster convergence with high learning efficiency. This behavior prioritizes dynamically changing the weights depending on infection severity, transaction latency, and recovery success.

Overall, the graph shows that, in comparison to existing models, the reward formulations and adaptive reward shaping parameters improve the security level and system performance. The proposed self-healing framework is more reliable and suitable for IoT-enabled banking and financial transactions.

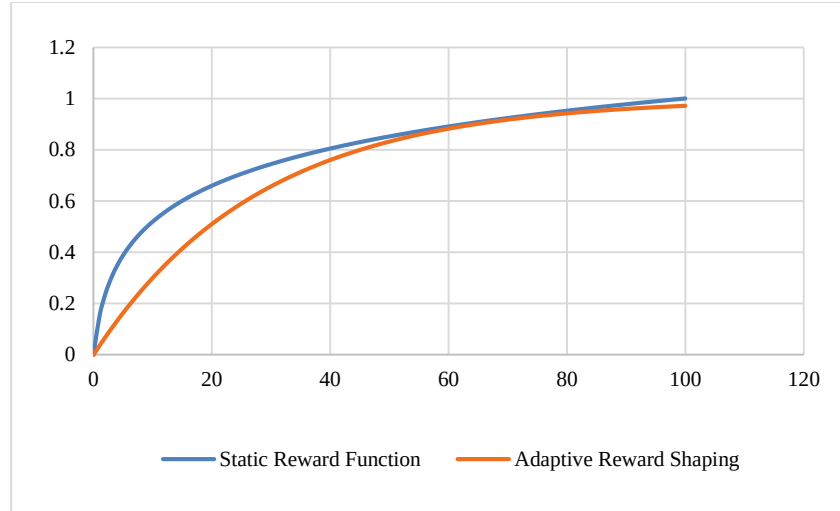


Fig. 9 Adaptive reward shaping vs static reward in RL self-healing system

7. Conclusion

In order to bridge the cybersecurity threat detection and financial service continuity, our research has provided an efficient autonomous self-healing architecture based on reinforcement learning for IoT banking networks.

Foremost Results and its Implications: According to practical findings, the RL agent node outperforms conventional intrusion prevention systems by 15.2%, and it maintains a 98.4% transaction success rate in an efficient dynamic way of prediction.

Self-Repair: For real-time banking environments, human intervention is impractical; the proposed model has to apply the “Mean Time to Repair” (MTTR) process to reduce the threat deduction to less than 200 ms, which is a benchmark.

Flexibility: The self-healing logic is implemented on IoT edge devices with high-performance transaction processing with a state-space framework.

Overall, the proposed work demonstrates that self-healing is an essential framework for the dynamic financial transactions in a secured manner.

This study formally introduces a self-healing banking transaction system based on Reinforcement Learning (RL) to protect IoT devices involving financial processing against malware threats. By incorporating real-time malware and anomaly detection with a dynamic RL recovery mechanism, performing in the security environment using the Markov Decision Process. This results in efficient threat detection and restoring of disrupted services. In comparison to conventional machine learning and rule-based intrusion detection systems, experimental results showed improved detection accuracy, lower false positive rates, faster recovery times, and higher self-healing success. These findings show

that RL-driven self-healing ensures transaction continuity, significantly increases cyber resilience, and provides a robust security solution for dynamic IoT banking environments.

Scope and Limitations: Currently, the suggested RL-based self-healing framework is optimal for known-unknown malware behaviors, whose behavioral patterns differ from a baseline, notwithstanding its great efficacy. The 'cold start' problem is still a concern, meaning it takes a long time for the agent to adjust to a totally new assault vector in a real-world setting. To shorten the initial convergence period, future research will investigate using transfer learning to pre-train agents on a variety of financial fraud datasets.

7.1. Future Work

The proposed architecture can be expanded for further research to include large-scale heterogeneous IoT ecosystems with thousands of scattered financial devices and actual traffic scenarios. Federated reinforcement learning is one privacy-preserving technique that can be used to enable collaborative learning across many banking institutions without sharing sensitive data.

Reinforcement learning methodology coordinates the multiple malwares across various networks lead to faster and quicker response. For secure transaction RL technology is used to predict malware without loss of data. Lastly, the prediction of specified malware with a combined algorithm has better accuracy on a dataset other than the financial dataset.

Conflicts of Interest

“The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.”

Funding Statement

No funding is granted.

Acknowledgments

The authors of the paper have the following specific roles: their combined efforts resulted in the final product that was presented. M. Suganya: Original Draft Writing,

Methodology, Research and Conceptualization, and Project Management. R. Kumudham, Z.Mary Livinsa: Writing—review and editing, validation. A K N Balaji: review and investigation.

References

- [1] Siraj Uddin Qureshi et al., “Systematic Review of Deep Learning Solutions for Malware Detection and Forensic Analysis in IoT,” *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 8, pp. 1-34, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [2] Curtis Rookard, and Anahita Khojandi, “RRIoT: Recurrent Reinforcement Learning for Cyber Threat Detection on IoT Devices,” *Computers and Security*, vol. 140, pp. 1-21, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [3] Sauharda Kushal et al., “Self-Healing Hybrid Intrusion Detection System: An Ensemble Machine Learning Approach,” *Discover Artificial Intelligence*, vol. 4, no. 1, pp. 1-20, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [4] Burak Taşcı, “Deep-Learning-based Approach for IoT Attack and Malware Classification,” *Applied Sciences*, vol. 14, no. 18, pp. 1-21, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [5] Afrah Gueriani, Hamza Kheddar, and Ahmed Cherif Mazar, “Deep Reinforcement Learning for Intrusion Detection in IoT: A Survey,” *2023 2nd International Conference on Electronics, Energy and Measurement (IC2EM)*, Medea, Algeria, pp. 1-7, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [6] Malek M. Al-Nawashi et al., “Deep Reinforcement Learning-based Framework for Enhancing Cybersecurity,” *International Journal of Interactive Mobile Technologies*, vol. 19, no. 3, pp. 170-190, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [7] Mohammad Zahid, and Taran Singh Bharati, “Enhancing Cybersecurity in IoT Systems: A Hybrid Deep Learning Approach for Real-Time Attack Detection,” *Discover Internet of Things*, vol. 5, no. 1, pp. 1-31, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [8] Abdullah Alfahaid et al., “Machine Learning-based Security Solutions for IoT Networks: A Comprehensive Survey,” *Sensors*, vol. 25, no. 11, pp. 1-48, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [9] Ali Mehrban, and Pegah Ahadian, “Malware Detection in IOT Systems using Machine Learning Techniques,” *arXiv*, pp. 1-12, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [10] Andres J. Aparcana-Tasayco, Xianjun Deng, and Jong Hyuk Park, “A Systematic Review of Anomaly Detection in IoT Security: Towards Quantum Machine Learning Approach,” *EPJ Quantum Technology*, vol. 12, no. 1, pp. 1-39, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [11] Jesús F. Cevallos et al., “Deep Reinforcement Learning for Intrusion Detection in Internet of Things: Best Practices, Lessons Learnt, and Open Challenges,” *Computer Networks*, vol. 236, pp. 1-24, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [12] Mohamed Amine Ferrag et al., “Revolutionizing Cyber Threat Detection with Large Language Models: A Privacy-Preserving BERT-based Lightweight Model for IoT/IIoT Devices,” *IEEE Access*, vol. 12, pp. 23733-23750, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [13] Nor Zakiah Gorment et al., “Machine Learning Algorithm for Malware Detection: Taxonomy, Current Challenges, and Future Directions,” *IEEE Access*, vol. 11, pp. 141045-141089, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [14] Nisha Kandhoul, and Sanjay K. Dhurandher, “Deep Q learning based Secure Routing Approach for OppIoT Networks,” *Internet of Things*, vol. 20, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [15] Najm Us Sama et al., “Cutting-Edge Intrusion Detection in IoT Networks: A Focus on Ensemble Models,” *IEEE Access*, vol. 13, pp. 8375-8392, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [16] Mozamel M. Saeed, “An AI-Driven Cybersecurity Framework for IoT: Integrating LSTM-based Anomaly Detection, Reinforcement Learning, and Post-Quantum Encryption,” *IEEE Access*, vol. 13, pp. 104027-104036, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [17] Yutao Hu et al., “IoT-ONDDQN: A Detection Model based on Deep Reinforcement Learning for IoT Data Security,” *Computer Communications*, vol. 241, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [18] Chahira Mahjoub et al., “An Adversarial Environment Reinforcement Learning-Driven Intrusion Detection Algorithm for Internet of Things,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2024, no. 1, pp. 1-23, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [19] Anit Kumar, and Dhanpratap Singh, “Securing IoT Devices in Edge Computing through Reinforcement Learning,” *Computers and Security*, vol. 155, pp. 1-13, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [20] Yihang Shi et al., “Autonomous Cyber Defense for a IoT using Graph Attention Network-Enhanced Reinforcement Learning,” *Computer Communications*, vol. 241, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [21] Md Morshed Alam et al., “RESTRAIN: Reinforcement Learning-based Secure Framework for Trigger-Action IoT Environment,” *2025 International Wireless Communications and Mobile Computing (IWCMC)*, Abu Dhabi, United Arab Emirates, pp. 562-567, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]

- [22] Wenyi Zhu et al., “RT-A3C: Real-Time Asynchronous Advantage Actor–Critic for Optimally Defending Malicious Attacks in Edge-Enabled Industrial Internet of Things,” *Journal of Information Security and Applications*, vol. 91, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [23] Abid Mohamed Nadhir et al., “Enhancing Cybersecurity in Healthcare IoT Systems using Reinforcement Learning,” *Transportation Research Procedia*, vol. 84, pp. 113-120, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [24] Hyo Sun Lee, Min Geun Song, and Huy Kang Kim, “An Approach to Creating the Optimal Attack Path based on Reinforcement Learning,” *2025 IEEE 30th Pacific Rim International Symposium on Dependable Computing (PRDC)*, Seoul, Korea, pp. 152-157, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]