

Adaptive CyberDefense Against Advanced Persistent Threats Using Game Theory and Reinforcement Learning

D. Navya¹, Dr. I. Ravi Prakash Reddy^{2*}

¹M.Tech Student, India, 1G. Narayanamma Institute of Technology and Science
Navyavarma3010@gmail.com

²Professor, India, G. Narayanamma Institute of Technology and Science
IRPReddy@gnits.ac.in

*Corresponding Author: Dr. I. Ravi Prakash Reddy (irpreddy@gnits.ac.in)

ABSTRACT

Advanced Persistent Threats (APTs) are a great challenge to the world of Cybersecurity. The paper makes a contribution to the world of Cybersecurity by developing an adaptive defense system using Game Theory and Reinforcement Learning. In the endeavor to identify anomalies and perform IP blocking and firewall activation, the system proposed monitors real-time data points such as CPU usage, network traffic, and failed login attempts. The system uses a Q-learning engine to optimize its decision-making process, resulting in faster and more accurate responses. The proposed adaptive and learning-driven approach outperforms the existing static security system against emerging APTs, as shown in the experiments.

INTRODUCTION

Because of the increase in digital infrastructure enhancement, cloud computing, and network connectivity, the ability of systems to be attacked by cybercriminals has increased. The most dangerous type of attack according to cybersecurity professionals is the use of Advanced Persistent Threats (APTs). They use a high degree of stealth, take a long-term approach to attaining their objectives, and are adaptive to changes made by their targets. Traditional cybersecurity methods, including firewalls, antivirus software, and signature-based intrusion detection systems, are based on a fixed structure. Thus, they provide an effective way to prevent known attacks, but they do not provide an effective way to prevent or stop adaptive attacks where the attacker continually modifies their tactics as they seek to evade detection while obtaining sensitive data or otherwise achieving strategic control over the system(s).

A critical limitation of traditional static cyber security systems is the need for intelligence-driven defense systems that can learn from their experiences and dynamically alter their behaviour in response to changes in the attackers' tactics and techniques. Recent research demonstrates that a key method of modelling cybersecurity can be viewed as a strategic interaction between defenders and attackers. Therefore, attack/defend (or game) theory is an effective way to mathematically model the strategic interaction between these two groups. Also, reinforcement learning provides mechanisms for intelligent systems to learn from prior experience and refine their decision-making processes through repeated interactions. Using these ideas as a basis, this paper proposes an adaptive cyber defence system that utilises game theory and reinforcement learning to provide an improved means of defending against APTs.

1.1 Problem Statement

Advanced Persistent Threats can still evade our ways of defending ourselves. This happens because Advanced Persistent Threats can change and endure for some time. Most security systems employ security measures like rules and signatures that remain static. This makes it very easy for the attackers to identify what the system is doing. This allows the attackers to always come up with ways of attacking. Advanced Persistent Threats are problematic because they can change and endure. We need to act on this problem concerning Advanced Persistent Threats because most of the times when the Advanced Persistent Threats strike, the security systems are unable to defeat them. This happens because the security systems remain static because the systems which are responding to the attack are not automated enough. So, the attack can keep going for a while. This can result in a lot of damage to the system. Advanced Persistent Threats are very good, at what they do.

The Research Problem:

How can an adaptive cyber defense framework be designed to continuously learn from attacker behavior and automatically respond to Advanced Persistent Threats in real time?

1.2 Objectives

- ☐ To design an adaptive cyber defense system to counter Advanced Persistent Threats
- ☐ To apply game theory to simulate the interaction between attackers and defenders.
- ☐ To apply reinforcement learning to adaptively choose effective defensive strategies.
- ☐ To allow automatic detection and response to system conditions in real time.
- ☐ To apply the proposed system to a simulated cyber environment with monitoring assistance.

LITERATURE REVIEW

cyber attacks are getting really bad. They are very sneaky. Can trick our computer systems. Advanced Persistent Threats or APTs for short are a problem for us. They make it hard for us to keep our cyber systems safe. So people who study this stuff are trying to figure out what these attackers do and how they think. They want to find ways to catch the guys when they are doing something wrong and make our systems stronger. Two things that are helping us are game theory and machine learning. Cyber attacks and these two tools are important because they can help us understand how attackers and defenders work together. Cyber attacks are a deal and game theory and machine learning can help make our systems better, at fighting against cyber attacks. Zhang and his team came up with a way to think about how attackers and defenders interact with each other on computers and the internet. They showed that if we can understand what attackers do we can help defenders figure out the ways to keep things safe[1]. The way they did it had

some problems. They only used plans that they had already thought of. They could not learn new things, which made it not very good, at dealing with Advanced Persistent Threats that are always changing.

The people who made Foureye, Li and the others they came up with this thing to trick people when they are trying to figure out what is going on and when they are moving around inside[2]. This way of doing things helped us see what the bad people were doing and understand what they were trying to do. But the main thing it did was try to trick them and watch what they did it did not really try to stop them or learn from what happened when the bad people attacked Foureye was mostly, about deception and watching.

Chen and his team looked into using reinforcement learning to help defend against cyber attacks in grid systems. They found out that systems that learn from experience can get better at defending themselves over time by paying attention to the types of attacks they see.. Their work was mostly about protecting the basic infrastructure from attacks and did not really think about how attackers and defenders interact with each other in a bigger picture of security [3]. Chen and his team didn't fully explore how reinforcement learning can be used to understand what attackers are trying to do and how defenders can respond to those actions, in grid environments. Zhao and his team looked into using intelligence that people can understand to find complex cyber threats. They found a way to make security decisions more clear and easy to understand. It needed a lot of computer power and people had to get involved to choose the right actions[4]. This makes it not very useful, for defense systems that need to work fast.

Zhang and colleagues looked at ways to defend against attacks in the Industrial Internet of Things environments. They found that their model was better, at handling attacks that happened in time. However their model was based on the idea that attacks would not change much. It did not look at how it would work over a long period of time when the attacks could change and get stronger[5]. Zhang and colleagues did this research on the Industrial Internet of Things. Reinforcement learning is based on some ideas that were first talked about by Sutton and Barto. They showed how agents can figure out the things to do by trying different actions and seeing what happens in their environment. Sutton and Barto did not focus on cybersecurity. Their work gives us the theory we need to use Q-learning in systems that defend against cyber attacks. This is

important for cyber defense systems as seen in reference [6].

Reinforcement learning and Q-learning are tools, for this type of work. Based on the literature surveyed, it is clear that the traditional intrusion detection approach is not adaptable, and the current game-theoretic or reinforcement learning method is only addressing a particular issue in the context of APT. Most of the current approaches are either not real-time or do not combine strategic modeling with continuous learning. In this context, the proposed research work will focus on an adaptive cyber defense strategy that combines game-theoretic modeling with reinforcement learning.

MATERIALS AND METHODS

3.1 System Overview

The cybersecurity defense is like a game between the bad guys who try to attack and the good guys who try to defend. The good guys watch what is happening in the system all the time. They check to see if everything is okay, with the security. Then they pick the way to defend against the bad guys. When the bad guys try ways to attack the system gets smarter. It figures out which ways of defending work in different situations. The cybersecurity defense Improves at protecting against the bad guys.

The framework illustrates how attackers and defenders act by using concepts from games. It also employs a type of learning that helps make good decisions when things change. We can observe what is happening in time so we have the right information. If something bad occurs, the safety rules activate automatically in the background. There is a part of the framework that we can observe on the internet. It assists us in understanding what is happening in the system and alerts us when we need to know something. The framework integrates game concepts and unique learning to make it effective. The observation in real time and the safety rules are components of the framework. The internet component is also very helpful, for observing the system activity and receiving alerts.

3.2 System Architecture

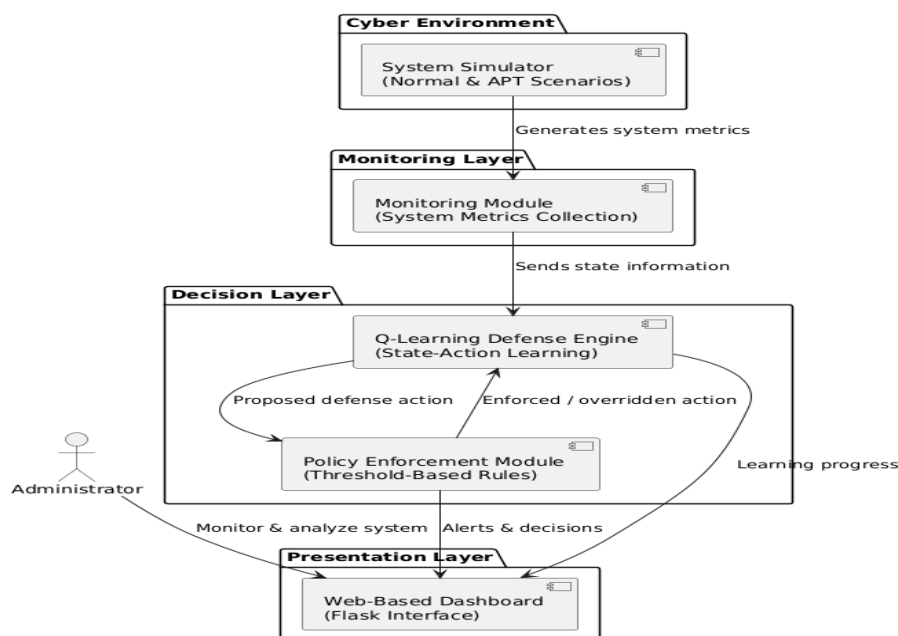


Figure 1. System architecture

The proposed system consists of multiple interconnected modules that collectively enable adaptive cyber defense, as shown in Figure 1. The system is made of parts that work together to keep an eye on what the system is doing check if it is safe and take action against Advanced Persistent Threats. It helps us see what is

happening now make good decisions and learn all the time. The system does this with Advanced Persistent Threats, in mind so it can deal with Advanced Persistent Threats properly

The system has also been structured in a way that allows it to have four major components. The components include the cyber environment, the monitoring layer, the decision layer, and the presentation layer. All these components ensure that the cyber environment and other layers work together to keep everything. The cyber environment and other layers are also able to adapt and make decisions whenever there is a challenge.

The cyber environment is where things happen it is, like, the place where normal things go on and where bad people try to do things. In this Cyber environment, we actually use a system simulator that shows us what normal looks like and what happens when there are things like Advanced Persistent Threats or APTs. The system simulator makes up events that look like real events, so we do not need to put it on a real network to test it.

The role of the monitoring layer entails gathering and processing system metrics provided by the system simulator. Here, various parameters are constantly monitored by the monitoring module related to system state and security status, converting collected events into useful state information. The generated state information is sent to the decision layer.

The decision layer is at the core of the proposed framework. It contains a Q-learning defense engine to learn optimal actions for defense, depending on state observations. The learning engine tests various state-action pairs and offers suitable defense strategies. For effectiveness and safety, there is a need to integrate a policy enforcement module within the decision layer. This module enforces threshold-based rules to test proposed actions for defending against potential security constraints.

The presentation layer offers a user interface with a dashboard using a web-based interface developed using Flask. The presentation layer is used to display the status of a system, alerts, as well as defense decisions. The feedback about learning status, including execution of a defense system, is visualized using a dashboard.

3.3 Simulator and Monitoring Module

The function of the simulator and monitoring module is of critical importance to the effective creation of adaptive cyber defense by providing real-world system observations through a series of benign and malicious conditions, as well as APT-like cyber attacks.

The metrics produced by the simulator correspond to the current system and security state of the environment. The metrics were influenced by the characteristics of the benchmark intrusion datasets, such as CIC-IDS 2017 and CIC-APT 2021. This enabled the framework to effectively track the behavior of realistic attacks.

This monitoring module continuously collects the generated metrics and analyzes them to create meaningful indicators such as threat level and anomaly intensity. This information is mapped into various system state representations that accurately indicate the current state of the system.

Then the derived state information is sent to the Q-learning defense engine located in the decision layer. Also, by using this framework of decoupling the simulations and monitors from the learning rules, modularity is ensured along with maximum flexibility of adaptation.

3.4 Adaptive Methodology and State Definition

This section explains the methodology that was used to implement the adaptive decision-making process for the proposed cyber defense framework. Here, the methodology is centered on the system's state, the feasible defense strategies, an efficient reward system, as well as the implementation of reinforcement learning to generate the optimal defense strategies while adhering to the policy.

State Representation

The system state reflects the current state of the security in the cyber environment as observed by the

monitoring module. The relevant system metrics, such as the level of attacks, anomalies, or system status, are processed to create a succinct view of the environment. The relevant metrics are discretized into a certain number of states, making the learning task easier. Each state reflects the overall threat level in the system at any particular time. It serves as the input to the learning engine.

Action Space

The action space essentially refers to the set of defense response measures that are available to the system and which the system is capable of undertaking. Every action is associated with a particular response measure, which may include measures such as raising security alerts, strengthening the security measures, and keeping the security measures at their current state. The actions are formulated to ensure that they are mutually exclusive as well as collectively exhaustive. This allows for a stable learning process by keeping the action space feasible.

Reward Function Design

The reward function helps to facilitate the learning process by expressing the quality of the defense actions chosen. Rewards are given for successful mitigation or safety, while negative rewards are used to penalize the system if the defense action chosen is not effective. Such a formulation will enable the learning agent to balance the system while accounting for the effectiveness of the actions chosen.

Q-Learning Strategy

In this context, it can be observed that Q-learning has been used as a method of reinforcement learning for the derivation of optimized defensive strategies by interacting with the environment. In the concept of reinforcement learning, it can thus be observed that the learning agent makes use of a Q-table to identify the expected value of the reward. This procedure of learning is based on the improvement of the agent by securing increasing levels of reward while interacting with the environment. The interactive element of this algorithm is based on the philosophy of adaptability while learning in response to variable attacks that may occur. This is possible without the need for any knowledge of the signature of the attacks; thus, the framework can learn to become more effective in the delivery of defensive strategies.

Policy-Based Safeguards

To ensure system reliability, policy-based safeguard mechanisms are incorporated into the decision-making process. The policy enforcement module does this by using a set of threshold-based policy rules to validate actions proposed by the learning engine. Thus, if an action violates the proposed actions as a result of a learning constraint, the action is overridden based on a more secure action that prevents violation of the proposed action.

3.5 Implementation Details

We implement the proposed adaptive cyber defense framework in a simulator-driven environment in order to evaluate its practical feasibility under controlled conditions. The implementation unifies system simulation with monitoring, adaptive decision-making, and policy enforcement into a single workflow. Real-time interaction among these components will enable dynamic defense responses against evolving attack scenarios. It utilizes a web-based interface for visualizing the system behaviors and defense decisions. The implementation focuses on functional validation, not performance evaluation.

3.5.1 Runtime Environment

The proposed adaptive cyber defense framework is implemented in Python on a simulator-based environment, which can emulate realistic cyber-attack scenarios. The simulator enables controlled generation of normal and malicious activities without even using a live network to deploy the framework for testing purposes.

The monitoring component processes the simulator outputs to extract system state information, and this information is delivered to the Q-learn-based defense engine. The learning engine will interact with the

simulator to choose appropriate defense actions according to system state information and learned experience. A visualization interface is designed using a web-based interface with the Flask framework, where the system status, alerts, as well as defense decision-making are visualized. The proposed implementation approaches the system by creating a modular structure, which facilitates smooth interaction between simulation, monitoring, learning, and visualization modules.

3.5.2 Actual Implementation Outputs

The features of the implemented adaptive cyber defense framework have provided real-time visibility and understanding of the system during various stages of attack defense. The system displays information regarding the security situation together with the action selected under threat levels detected within the system. The generated alerts and logs reflect the overall capabilities of the system in terms of attack activity detection and the execution of an adaptive response. The system keeps a summary of different attack types that have been detected during the execution of the system, and this reflects a broad range of attack types within the overall framework.

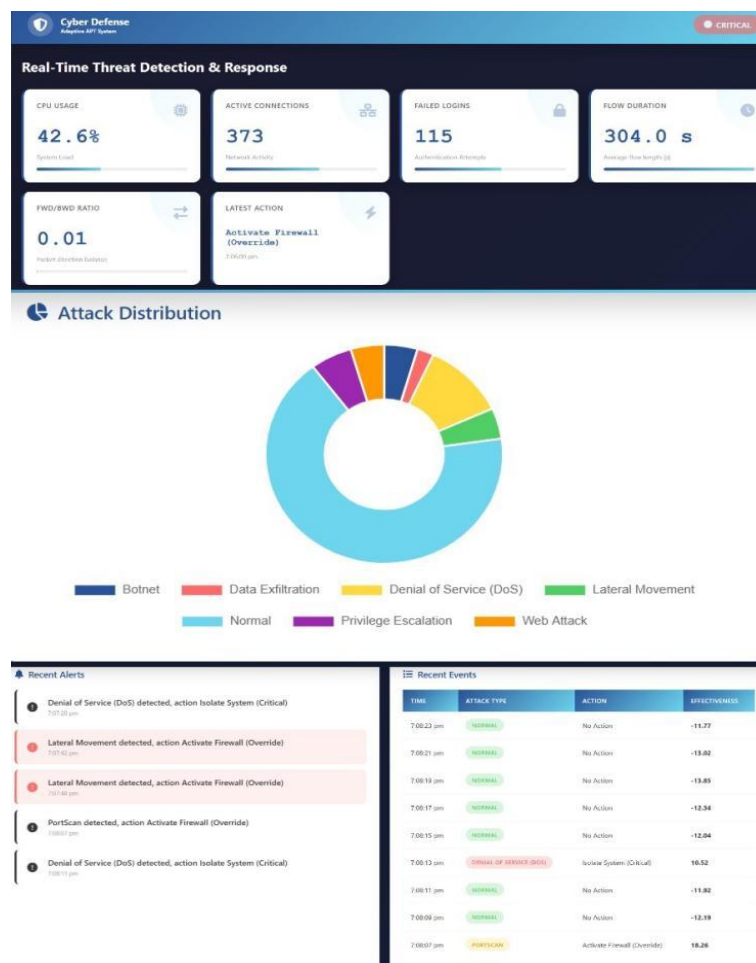


Figure 2. Implementation outputs of the proposed adaptive cyber defense system.

The modularity design allows for a seamless integration of the simulation, monitoring, learning, policy enforcement, and visualization modules.

The results of trying out the cyber defense framework show that it really works well when things are happening in time. The system is good, at learning and defending itself against kinds of attacks. When you combine reinforcement learning and policy-based defenses the cyber defense framework can adapt to situations and stay stable. Trying out the cyber defense framework confirms that the proposed adaptive cyber defense framework is an idea and can actually work. The cyber defense framework is a way to defend against cyber attacks.

Figure 2 demonstrates some representative outputs extracted during the execution of the proposed adaptive cyber defense framework. The depicted views prove the interaction between system monitoring, alert generation, and adaptive defense-making decisions. Such outputs verify the practical realization of the proposed framework.

RESULTS AND DISCUSSION

4.1 Experimental Setup and Performance Metrics

The new cyber defense system was tested using a cyber world that was created based on the CIC-IDS-2017 and CIC-APT-2021 datasets. This fake world was made to act like computer traffic and also to show different kinds of attacks like when someone tries to shut down the system or get special access or move from one computer to another or even steal data.

To make sure the test was fair all the methods were tried in the same way. The things that were measured to see how well they worked included accuracy, precision, recall, overall performance, and response time, as summarized in Table 1. The cyber defense system must be able to look for problems and also respond quickly. This is because the cyber defense systems are used in a world that is constantly changing, and therefore they must be able to look for problems and respond quickly.

The Q-learning agent was trained using the cyber world. As the agent improved, the results became more consistent, which means that the agent was learning well. The cyber defense system was tested using the Q-learning agent and the fake cyber world. The results showed that the system was improving at looking for problems and responding quickly. The Q-learning agent and the cyber defense system are significant because they help ensure that the cyber world is safe from individuals who attempt to shut down systems or steal information.

Algorithm	Accuracy	Precision	Recall	F1	Time (s)
Q-Learning (Proposed)	0.9825	1.0000	0.9760	0.9879	0.001
Rule-Based	0.9300	1.0000	0.7760	0.8739	0.000
Random Forest	1.0000	1.0000	1.0000	1.0000	0.042
XGBoost	1.0000	1.0000	1.0000	1.0000	0.003
Isolation Forest	0.9800	0.9835	0.9520	0.9675	0.011

Table1. Performance metrics

4.2 Comparative Analysis

The new defense system based on Q-learning was compared with approaches such as rule-based detection, Random Forest, XGBoost, and Isolation Forest algorithms as shown in figure 3.

Some approaches such as Random Forest and XGBoost are really effective in finding the answers because they were trained on labeled data. These approaches are known as learning techniques and they perform really well. However, they have some issues. They are trained on data that does not change. They need to be trained again when new attack patterns emerge.

Isolation Forest is an approach that can identify behavior without being trained on labeled data. This is known as anomaly detection. It performs really well in identifying behavior but does not know how to react to it.

Rule-based systems are really fast because they are simple. They can make decisions very quickly. However, they are not very effective in handling attacks because they are not flexible.

The Q-learning approach is different. It learns as it goes. It updates its strategy based on what happens. It is not as accurate as other approaches but it is still very effective in detecting problems. It can also react very quickly. It makes good decisions.

The Q-learning defense system is an option because it can adapt to new threats and respond quickly. The Q-learning method is very useful, for defending against attacks that are always evolving.

The Q-learning agent gets better and better as it works with the environment. We can see this because the performance metrics become more stable over time. This means the Q-learning agent is learning and finding the policy. The Q-learning agent is different from models that make decisions based on fixed rules. The Q-learning agent changes its defense strategy based on what it learns so it can deal with different attacks. This is very important for cybersecurity because attacks are always changing and we do not know what they will be. The Q-learning agent is good at dealing with the Q-learning agents performance in these situations. The Q-learning agent is a choice, for real-world cybersecurity because the Q-learning agent can adapt to new attacks.

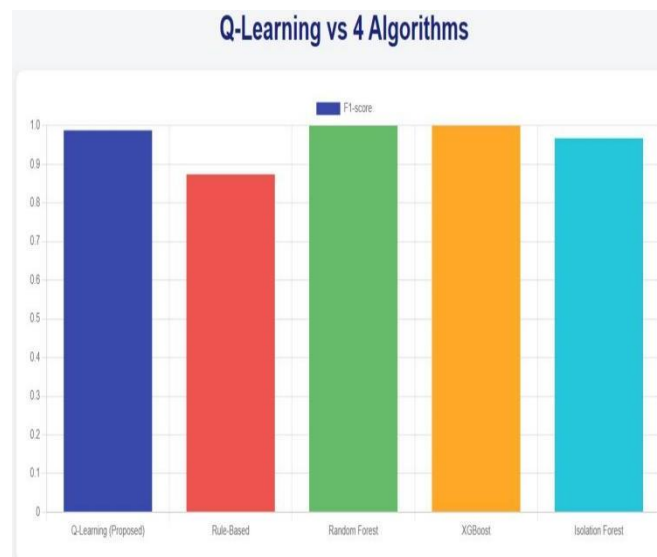


Figure 3.Performance comparison of Q-learning with existing algorithms.

5. Conclusion and Future work

This paper proposed an adaptive defense mechanism using Q-learning, which contributes to the selection of mitigation actions according to specific states perceived within the system. The paper effectively presents the effectiveness of various scenarios of cyber attacks using the proposed solution, which incorporates real-time monitoring, alerting, and decision-making processes. According to the experimental results, this defense strategy strikes a good balance with respect to addressing various types of attacks as opposed to static and supervised approaches.

However, though the high performances of supervised learning approaches have been confirmed, their inability to adapt and cope with dynamic and changing environments due to their offline nature is a significant issue. In contrast, the proposed method, which relies on reinforcement learning, has the advantage of learning from interacting with the environment.

Future work will include extending the framework to incorporate multi-agent reinforcement learning for cooperative defense, deep reinforcement learning for dealing with larger state spaces, and validation of the framework using real-world network traffic data sets. Other extensions could be the optimization of policies

as well as the integration with distributed intrusion detection systems.

REFERENCES

- [1] T. Zhang and S. Rass, "A game-theoretic framework for analyzing interactions between attackers and defenders in cyber systems," *Computers & Security*, vol. 97, 2020.
- [2] H. Li, P. Ning, and M. Reiter, "Foureyeye: Defensive deception against advanced persistent threats," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 1248–1263, 2021.
- [3] J. Chen, Q. Huang, and Z. Li, "Reinforcement learning-based cyber defense in smart grid environments," *IEEE Transactions on Smart Grid*, vol. 13, no. 2, pp. 1456–1468, 2022.
- [4] Y. Zhao, H. Liu, and K. Chen, "Explainable artificial intelligence for advanced cyber threat detection," *IEEE Access*, vol. 11, pp. 45678–45690, 2023.
- [5] X. Zhang, L. Wang, and M. Li, "Adaptive defense strategies for Industrial Internet of Things security," *IEEE Internet of Things Journal*, vol. 10, no. 4, pp. 3124–3136, 2023.
- [6] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. MIT Press, 2018.
- [7] A. S. Alshamrani, S. Myneni, A. Chowdhary, and D. Huang, "A survey on advanced persistent threats: Techniques, solutions, challenges, and research opportunities," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1851–1877, 2019.
- [8] M. Ring, D. Schlör, D. Landes, and A. Hotho, "A survey of network-based intrusion detection datasets," *Computers & Security*, vol. 86, pp. 147–167, 2019.
- [9] M. Tavallaei et al., "A detailed analysis of the CIC-IDS-2017 dataset," in *Proc. IEEE Int. Conf. Information Systems Security*, 2018.
- [10] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [11] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016.
- [12] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *Proc. IEEE Int. Conf. Data Mining*, 2008.
- [13] K. Lye and J. Wing, "Game strategies in network security," *International Journal of Information Security*, vol. 4, no. 1–2, pp. 71–86, 2005.
- [14] R. Sommer and V. Paxson, "Outside the closed world: On using machine learning for network intrusion detection," in *Proc. IEEE Symposium on Security and Privacy*, 2010.
- [15] W. Lee and S. J. Stolfo, "A framework for constructing features and models for intrusion detection systems," *ACM Transactions on Information and System Security*, vol. 3, no. 4, pp. 227–261, 2000.
- [16] N. Hubballi and V. Suryanarayanan, "False alarm minimization techniques in intrusion detection systems," *Computer Communications*, vol. 49, pp. 1–17, 2014.
- [17] A. Patcha and J. Park, "An overview of anomaly detection techniques: Existing solutions and latest technological trends," *Computer Networks*, vol. 51, no. 12, pp. 3448–3470, 2007.

Declarations

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper. The research was conducted independently as part of academic requirements, and no financial, commercial, or personal relationships have influenced the outcomes of this study. The authors confirm that the work presented is original and has not been submitted elsewhere for publication.

Funding Information

This research was conducted as part of the M.Tech academic curriculum in Computer Networks and Information Security at G. Narayanamma Institute of Technology and Science, Hyderabad, India. The study did not receive any external financial support, grants, or sponsorship from funding agencies in the public, commercial, or not-for-profit sectors. All resources used for experimentation and implementation were part of the institutional academic infrastructure.

Data Availability

The data used in this study were generated through a controlled simulated cyber attack environment developed to evaluate the proposed adaptive cyber defense framework under varying threat intensities. The experimental setup, configuration parameters, performance logs, and evaluation metrics were created specifically for research and validation purposes. Supporting data and implementation details are available from corresponding author Dr. I. Ravi Prakash Reddy upon reasonable request for academic and research use.

Author Contribution Statement

The authors were responsible for the complete conceptualization and design of the adaptive cyber defense framework. This includes system architecture development, implementation of dynamic attack probability modeling, integration of adaptive strategy mechanisms, development of performance evaluation modules, and experimental validation. The authors conducted data generation, performance analysis, comparative evaluation of defense strategies, and prepared, reviewed, and finalized the manuscript for submission

Authors Biography

Dongari Navya is currently pursuing M.Tech in Computer Networks and Information Security at G. Narayanamma Institute of Technology and Science, Hyderabad, India. Her research interests include adaptive cyber defense systems, intelligent intrusion detection mechanisms, network security analytics, and reinforcement learning-based security frameworks. She has worked on projects involving dynamic attack probability modeling, strategy optimization, and real-time security visualization dashboards. Her academic focus is on designing intelligent and resilient cybersecurity systems capable of adapting to evolving threat landscapes.

Dr. I. Ravi Prakash Reddy is Dean of Placements & Corporate Relations at G. Narayanamma Institute of Technology & Science (Autonomous), Hyderabad. He holds a Ph.D. and has been with the institute since 2001, having served as Head of the Department for Information Technology. His academic interests include Distributed Systems, Software Agents, Operating Systems, and related areas. Dr. Reddy has authored 30+ research publications, contributed two book chapters, and written two books. He is also credited with three registered patents and actively fosters industry-academia linkages.