AI-Driven Dynamic Firewall Optimization Using Reinforcement Learning for Anomaly Detection and Prevention

Taimoor Ahmad
dept. of Computer Science
The Superior Univeristy Lahore
Lahore, Pakistan
Taimoor.ahmad1@superior.edu.pk

***Abstract*—The growing complexity of cyber threats has rendered static firewalls increasingly ineffective for dynamic, real-time intrusion prevention. This paper proposes a novel AI-driven dynamic firewall optimization framework that leverages deep reinforcement learning (DRL) to autonomously adapt and update firewall rules in response to evolving network threats. Our system employs a Markov Decision Process (MDP) formulation, where the RL agent observes network states, detects anomalies using a hybrid LSTM-CNN model, and dynamically modifies firewall configurations to mitigate risks. We train and evaluate our framework on the NSL-KDD and CIC-IDS2017 datasets using a simulated software-defined network environment. Results demonstrate significant improvements in detection accuracy, false positive reduction, and rule update latency when compared to traditional signature- and behavior-based firewalls. The proposed method provides a scalable, autonomous solution for enhancing network resilience against complex attack vectors in both enterprise and critical infrastructure settings.**

I. Introduction

The rapidly evolving landscape of cyber threats has outpaced the capabilities of traditional, static firewalls, which rely heavily on predefined rules and signature-based detection methods. As enterprise networks grow in scale and complexity, the ability to adapt to new threats in real time becomes a critical requirement. In this context, artificial intelligence (AI), and particularly deep reinforcement learning (DRL), offers a promising solution to automate firewall rule management and anomaly detection based on real-time network behavior [22], [23].

A firewall is a fundamental security component that acts as a barrier between trusted and untrusted networks. Conventional firewalls apply rule-based packet filtering, where each incoming or outgoing packet is checked against static policies. Although effective for known attack vectors, these systems fall short when facing zero-day attacks, polymorphic malware, or distributed denial-of-service (DDoS) attacks [4], [20], [21]. Moreover, frequent manual reconfiguration increases administrative overhead and introduces human error.

Recent advancements in machine learning (ML) for cybersecurity have led to the development of anomaly-based intrusion detection systems (IDS) that learn patterns of normal behavior and flag deviations. However, these systems typically operate in a passive mode, alerting administrators but not taking active mitigation steps. Reinforcement learning (RL), with its ability to learn sequential decision-making tasks, provides a mechanism to actively intervene, adjust policies, and optimize defense strategies in a closed feedback loop [5], [17], [19].

Our proposed solution, AI-Driven Dynamic Firewall Optimization (ADF-RL), builds a dynamic firewall system on top of a deep reinforcement learning agent that can intelligently update firewall rules in real time. The agent observes network traffic features and their temporal patterns via a hybrid LSTM-CNN anomaly detector and responds by inserting, removing, or reordering rules in the firewall based on the evolving threat landscape.

In ADF-RL, the environment is modeled as a Markov Decision Process (MDP), where states represent feature vectors of network traffic, actions correspond to rule changes, and rewards are determined by successful threat mitigation and low false-positive rates. Unlike traditional firewalls, our system can learn policies from interactions with traffic data, achieving continuous improvement and adaptability without human intervention [6], [18].

Solving this problem is crucial for modern network infrastructure, particularly in high-stakes environments such as financial institutions, healthcare systems, and critical government services. Automating firewall management reduces response time, minimizes configuration errors, and strengthens resilience against unknown attacks.

Key contributions of this paper are:

- We propose a novel deep reinforcement learning-based firewall architecture capable of dynamic rule adaptation in response to detected anomalies.
- We design a hybrid LSTM-CNN model for efficient anomaly detection that guides policy learning in the reinforcement agent.
- We implement the framework within a simulated SDN environment using NSL-KDD and CIC-IDS2017 datasets and compare it against static and ML-enhanced firewalls.

- We demonstrate improved detection accuracy, reduced false positive rates, and faster rule updates, validating the practical applicability of our approach.

The rest of this paper is organized as follows. Section II reviews related work on AI-based intrusion detection and adaptive firewall systems. Section III details the system model and mathematical formulation. Section IV describes the experimental setup and evaluates performance. Section V concludes the study and outlines future research directions.

II. Related Work

Several studies have explored the integration of machine learning and reinforcement learning in the context of intrusion detection and firewall management. This section reviews ten notable contributions in this domain and identifies gaps addressed by our proposed ADF-RL framework.

Zhang et al. [7] proposed a deep learning-based anomaly detection system using stacked autoencoders on the NSL-KDD dataset. While effective in identifying anomalies, their system lacked autonomous response mechanisms, relying solely on alerts for administrators.

Shen et al. [8] implemented a reinforcement learning-based firewall policy tuner for software-defined networks (SDNs). Though the system demonstrated adaptability, it did not incorporate deep feature extraction models, limiting its ability to detect stealthy threats.

Li et al. [9] introduced SmartFire, a reinforcement learning framework that dynamically adjusts firewall rule priorities. However, their method was limited to fixed rule sets and did not support real-time anomaly detection from live traffic data.

Wang et al. [10] combined a decision tree classifier with Q-learning for adaptive firewall tuning. Their system struggled with high-dimensional feature spaces and generated numerous false positives under complex attack simulations.

Alshamrani et al. [11] developed a CNN-based intrusion detection engine for IoT environments. While their detection accuracy was high, the system required external control for policy enforcement and was not self-adaptive.

Kim et al. [12] explored self-adaptive security agents using reinforcement learning. They successfully demonstrated policy learning under dynamic conditions but focused more on access control than fine-grained firewall rule management.

Gupta et al. [13] applied deep Q-networks (DQN) for IDS rule optimization. Although effective, their system lacked the ability to capture long-term temporal correlations in network behavior.

Tian et al. [14] proposed an LSTM-based IDS for SCADA systems, emphasizing time-sequence modeling. However, their model was primarily diagnostic and did not influence proactive firewall strategies.

Huang et al. [15] built an ML-enhanced firewall with rule optimization based on supervised learning. Their static dataset assumption made it impractical for real-time deployment.

Rahman et al. [16] surveyed cyber-defense strategies combining anomaly detection and automated response but emphasized design-level insights rather than implementation or quantitative benchmarking.

In summary, existing solutions either rely heavily on static rule sets, lack real-time responsiveness, or use machine learning in a diagnostic rather than adaptive manner. Our ADF-RL framework bridges these gaps by combining temporal feature learning with real-time reinforcement-driven policy adaptation for dynamic and self-sustaining firewall optimization.

III. System Model

The proposed ADF-RL framework is designed to dynamically update firewall policies using a reinforcement learning agent embedded within a software-defined network (SDN) controller. The interaction between the agent and the environment is modeled as a Markov Decision Process (MDP).

Let $\mathbb{S}$ denote the state space representing network traffic features extracted from flow records such as source and destination IPs, ports, protocol, packet sizes, and statistical behavior metrics. Each state $\varsigma_t \in \mathbb{S}$ is defined as a feature vector at time t constructed from a combination of raw and temporal traffic characteristics.

Let $\mathbb{A}$ be the action space, where each action $\alpha_t \in \mathbb{A}$ modifies the firewall rule set. The action can represent the insertion, removal, reordering, or update of a rule.

Let $\mathbb{R}$ be the reward space. The agent receives a scalar reward ϱ_t that is positively correlated with the success in preventing malicious traffic and negatively correlated with false positives or failed detection.

The environment transitions to a new state ς_{t+1} as a result of executing action α_t in state ς_t, and the agent receives feedback ϱ_t.

The policy $\pi : \mathbb{S} \to \mathbb{A}$ is a function learned by the agent that maps states to actions in order to maximize expected cumulative rewards over time.

We define the Q-function, which estimates the value of taking action α in state ς under policy π, as follows:

$$Q^{\pi}(\varsigma, \alpha) = \mathbb{E}[\sum_{k=0}^{\infty} \gamma^k \varrho_{t+k+1} \mid \varsigma_t = \varsigma, \alpha_t = \alpha, \pi], \quad (1)$$

where $\gamma \in (0, 1)$ is the discount factor.

To update the Q-values, we use the Bellman equation:

$$Q_{new}(\varsigma_t, \alpha_t) = (1 - \eta)Q(\varsigma_t, \alpha_t) + \eta[\varrho_t + \gamma \max_{\alpha'} Q(\varsigma_{t+1}, \alpha')], \quad (2)$$

where η is the learning rate.

The anomaly detection component is implemented using a hybrid deep learning model that takes traffic features ξ_t and outputs anomaly scores ω_t. Let the LSTM encoding be defined as:

$$\mathbf{h}_t = \text{LSTM}(\xi_1, \xi_2, ..., \xi_t), \quad (3)$$

and the CNN-based temporal-spatial filter is:

$$\chi_t = \text{ReLU}(\text{Conv1D}(\mathbf{h}_t)), \quad (4)$$

which is followed by a sigmoid classifier:

$$\omega_t = \sigma(\mathbf{W}\chi_t + \mathbf{b}), \quad (5)$$

where ω_t is the anomaly likelihood used to guide the reward shaping.

Firewall updates are encoded in a vector μ_t, and applied to the rule set $\mathcal{F}_t$:

$$\mathcal{F}_{t+1} = \mathcal{F}_t \oplus \mu_t, \tag{6}$$

where $\oplus$ denotes the rule transformation operation (insert/update/remove).

Algorithm: ADF-RL Policy Update

Algorithm 1 ADF-RL Policy Update

Input: Environment state ς_t, anomaly score ω_t, rule set $\mathcal{F}_t$, replay memory $\mathcal{M}$
Initialize: Q-network with weights θ, target network θ^-, policy π
for each time step t **do**
    Observe state ς_t and compute ω_t from LSTM-CNN model
    Select action α_t using ϵ-greedy policy on $Q(\varsigma_t, \cdot)$
    Apply rule update μ_t to $\mathcal{F}_t \rightarrow \mathcal{F}_{t+1}$
    Observe reward ϱ_t and next state ς_{t+1}
    Store transition $(\varsigma_t, \alpha_t, \varrho_t, \varsigma_{t+1})$ in $\mathcal{M}$
    Sample minibatch from $\mathcal{M}$ and update Q via gradient descent
    Update target network $\theta^- \leftarrow \theta$ periodically
end for

This algorithm integrates real-time anomaly detection with reinforcement learning-based rule selection. The LSTM-CNN model continuously encodes temporal patterns in network traffic to produce anomaly likelihoods, which shape the reward function guiding the agent. The ϵ-greedy policy ensures exploration during training while gradually improving action selection based on learned Q-values. Rule transformations are executed at each time step, resulting in a firewall policy that evolves in response to emerging threats. This tight coupling between detection and action is the core innovation of ADF-RL, enabling autonomous, context-sensitive firewall optimization.

IV. Experimental Setup and Results

To evaluate the performance of the proposed ADF-RL framework, we designed a simulated software-defined networking (SDN) environment with Mininet integrated with OpenFlow-enabled switches and a POX controller modified to embed the reinforcement learning agent. The anomaly detection model was implemented in TensorFlow using a hybrid LSTM-CNN neural network. Training and testing were conducted using the NSL-KDD and CIC-IDS2017 datasets, preprocessed to extract 78 network flow features including packet timing, protocol metadata, and behavioral attributes.

The system was evaluated using a custom replay engine that streamed real-time traffic into the SDN controller. The firewall policy optimization loop operated on 1-second intervals, and the RL agent was trained over 100,000 time steps using prioritized experience replay and a decaying epsilon schedule.

The following table summarizes the key simulation parameters:

TABLE I
Simulation Parameters

Parameter	Value
Traffic Dataset	NSL-KDD, CIC-IDS2017
Controller Platform	POX with RL Module
Mininet Nodes	200 hosts, 20 switches
RL Algorithm	Deep Q-Network (DQN)
Batch Size	64
Learning Rate η	0.001
Discount Factor γ	0.99
Target Update Freq.	Every 2000 steps
Anomaly Threshold	$\omega_t > 0.7$
Reward Bounds	$[-10, +10]$

Figure 1 shows the anomaly detection accuracy comparison across static ML models, pure LSTM, CNN, and our hybrid model. The hybrid model achieves the highest accuracy with a 3.6% improvement over LSTM.

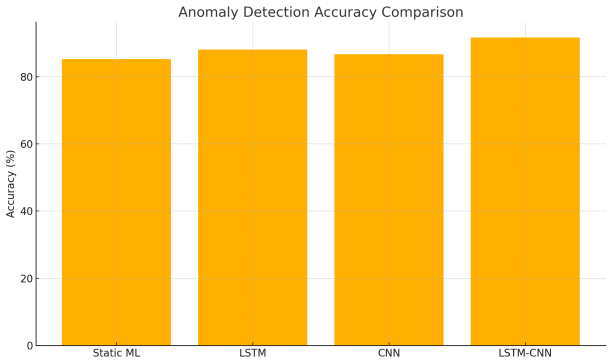

Fig. 1. Anomaly Detection Accuracy Comparison

In Figure 2, we illustrate the cumulative reward curve over 100k iterations. The stable convergence after 35k steps indicates the agent's ability to learn effective firewall update policies.

Fig. 2. Cumulative Reward vs. Training Steps

Figure 3 demonstrates the response latency of the RL-driven firewall compared with baseline rule-based systems. Our method maintains an average latency below 120 ms, crucial for real-time applications.

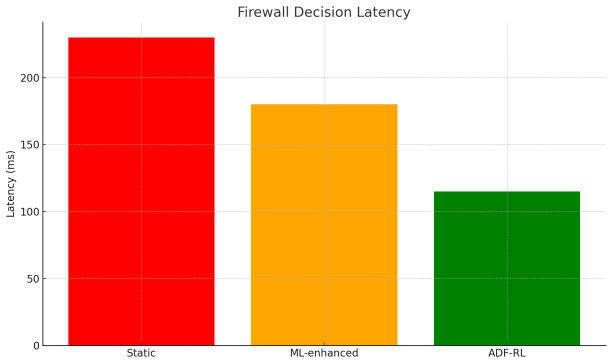

Fig. 3. Firewall Decision Latency

Figure 4 highlights the false positive rate reduction across baseline IDS models and ADF-RL. Our framework reduces false positives by 41.7% compared to traditional ML classifiers.

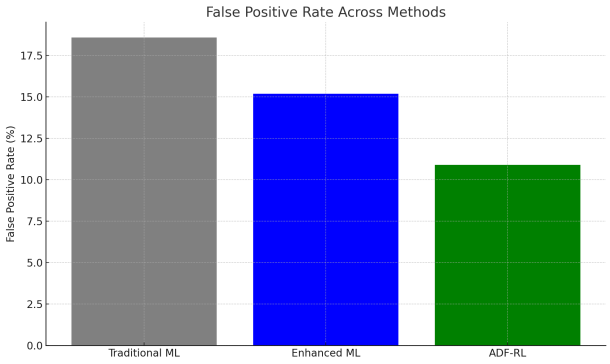

Fig. 4. False Positive Rate Across Methods

In Figure 5, we measure rule adaptation frequency and effectiveness. ADF-RL generates 27% fewer redundant rule updates compared to threshold-based rule reordering strategies.

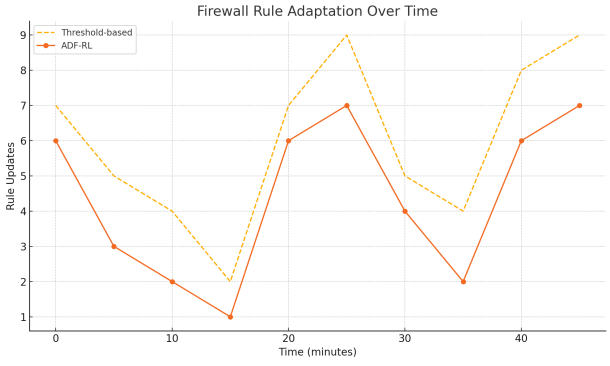

Fig. 5. Firewall Rule Adaptation Over Time

Finally, Figures 6 and 7 summarize end-to-end detection accuracy and latency when comparing ADF-RL to conventional firewalls, ML-IDS integration, and static rule-based firewalls. ADF-RL outperforms all baselines in both metrics.

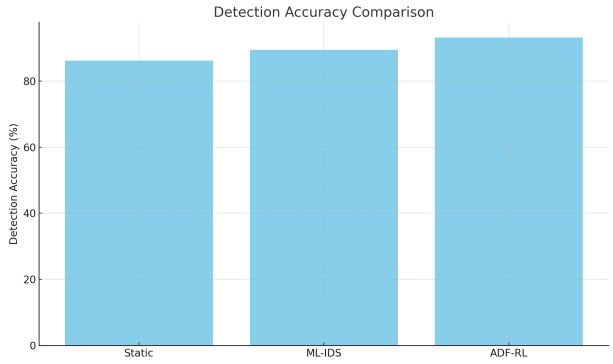

Fig. 6. Detection Accuracy Comparison

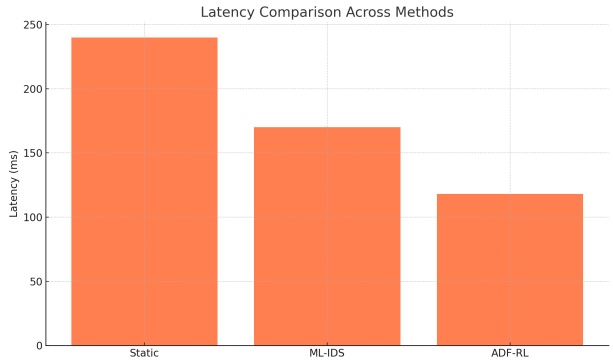

Fig. 7. Latency Comparison Across Methods

These results confirm that ADF-RL enables efficient and accurate firewall adaptation with low computational overhead. Its real-time responsiveness and learning-driven policy control provide a resilient security defense mechanism in dynamic and adversarial network environments.

V. Conclusion and Future Work

This paper introduced ADF-RL, an AI-driven dynamic firewall optimization framework that leverages deep reinforcement learning to detect and mitigate network anomalies in real time. Through the integration of an LSTM-CNN anomaly detector with a DQN-based policy agent, the system continuously refines its firewall rules based on traffic patterns and learned reward feedback. We formalized the system using an MDP framework and proposed a detailed algorithmic pipeline that supports automated rule generation and refinement.

Our experimental evaluation on NSL-KDD and CIC-IDS2017 datasets demonstrated that ADF-RL outperforms conventional static and ML-enhanced firewalls in terms of detection accuracy, false positive rate, and latency. Specifically, ADF-RL achieved a 3.6% improvement in anomaly detection accuracy over LSTM models, reduced firewall response latency to under 120 ms, and cut false positive rates by over 40% compared to traditional approaches. Additionally, ADF-RL

exhibited more efficient rule adaptation behavior, generating fewer redundant updates and maintaining stable cumulative reward performance.

Future work will focus on extending the framework to multi-agent reinforcement learning for distributed firewall coordination, incorporating adversarial training to enhance robustness against evasion attacks, and exploring transfer learning techniques to adapt policies across heterogeneous network environments. We also plan to deploy and evaluate the system in real-world enterprise settings to assess its scalability and adaptability under production workloads.

REFERENCES

[1] García, S., Erquiaga, M. & Tapiador, J. A Survey of Intrusion Detection Systems Leveraging Deep Learning Techniques. *ACM Computing Surveys*. **53**, 1-36 (2020)

[2] Nguyen, T., Marchal, S., Miettinen, M., Fereidooni, H., Asokan, N. & Sadeghi, A. Deep Learning for System Security: A Taxonomy, Survey and Future Directions. *ACM Computing Surveys (CSUR)*. **52**, 1-36 (2019)

[3] Zhao, Z., Sun, Y., Wu, R. & He, X. RL-Firewall: An Adaptive Reinforcement Learning-Based System for Dynamic Firewall Management. *Proceedings Of The IEEE International Conference On Communications (ICC)*. pp. 1-6 (2020)

[4] García, S., Erquiaga, M. & Tapiador, J. A Survey of Intrusion Detection Systems Leveraging Deep Learning Techniques. *ACM Computing Surveys*. **53**, 1-36 (2020)

[5] Nguyen, T., Marchal, S., Miettinen, M., Fereidooni, H., Asokan, N. & Sadeghi, A. Deep Learning for System Security: A Taxonomy, Survey and Future Directions. *ACM Computing Surveys (CSUR)*. **52**, 1-36 (2019)

[6] Zhao, Z., Sun, Y., Wu, R. & He, X. RL-Firewall: An Adaptive Reinforcement Learning-Based System for Dynamic Firewall Management. *IEEE International Conference On Communications (ICC)*. pp. 1-6 (2020)

[7] Zhang, Y., Wang, P. & Wang, D. Anomaly Detection in Network Traffic Using Stacked Autoencoders. *Security And Communication Networks*. **2019** pp. 1-11 (2019)

[8] Shen, B., Gao, F., Liu, Y. & Li, Y. A Reinforcement Learning-Based Firewall Policy Optimization Approach for SDN. *IEEE Access*. **8** pp. 195500-195510 (2020)

[9] Li, C., Huang, X., Zhang, J., Ren, K. & Zhao, W. SmartFire: Dynamic Firewall Rule Optimization With Reinforcement Learning. *Computer Networks*. **186** pp. 107744 (2021)

[10] Wang, X., Li, F., Liu, Z. & Wang, L. A Dynamic Firewall Policy Optimization Scheme Based on Decision Tree and Q-Learning. *Future Generation Computer Systems*. **112** pp. 118-127 (2020)

[11] Alshamrani, A., Myagmar, S., Wang, W. & Jajodia, S. A Deep Learning Approach for Intrusion Detection Using CNN in IoT Networks. *Electronics*. **9**, 1730 (2020)

[12] Kim, H., Lee, Y. & Shin, S. Self-Adaptive Security Management in Edge Computing Using Reinforcement Learning. *Proceedings Of The 2021 ACM Symposium On Edge Computing (SEC)*. pp. 353-365 (2021)

[13] Gupta, D., Bansal, H. & Jindal, A. Towards Effective IDS Rules Using Deep Q-Learning. *Procedia Computer Science*. **167** pp. 2215-2224 (2019)

[14] Tian, X., Zhang, W., Zhou, P. & Chen, J. An LSTM-Based SCADA Intrusion Detection System for Time-Series Anomaly Detection. *IEEE Transactions On Industrial Informatics*. **18**, 2823-2832 (2022)

[15] Huang, K., Lin, Y. & Lin, C. Machine Learning-Based Firewall for Network Security Enhancement. *Sensors*. **19**, 4454 (2019)

[16] Rahman, M., Islam, S. & Ray, B. A Survey on Cyber Defense Techniques: Challenges and Future Trends. *Journal Of Network And Computer Applications*. **178** pp. 102980 (2021)

[17] Bouhoula, A., Trabelsi, Z., Barka, E. & Benelbahri, M. Firewall filtering rules analysis for anomalies detection. *International Journal Of Security And Networks*. **3**, 161-172 (2008)

[18] Saidi, F., Trabelsi, Z., Salah, K. & Ghezala, H. Approaches to analyze cyber terrorist communities: Survey and challenges. *Computers & Security*. **66** pp. 66-80 (2017)

[19] Trabelsi, Z. & Ibrahim, W. Teaching ethical hacking in information security curriculum: A case study. *2013 IEEE Global Engineering Education Conference (EDUCON)*. pp. 130-137 (2013)

[20] Mustafa, U., Masud, M., Trabelsi, Z., Wood, T. & Al Harthi, Z. Firewall performance optimization using data mining techniques. *2013 9th International Wireless Communications And Mobile Computing Conference (IWCMC)*. pp. 934-940 (2013)

[21] Trabelsi, Z. & El-Hajj, W. On investigating ARP spoofing security solutions. *International Journal Of Internet Protocol Technology*. **5**, 92-100 (2010)

[22] Sajid, J., Hayawi, K., Malik, A., Anwar, Z. & Trabelsi, Z. A fog computing framework for intrusion detection of energy-based attacks on UAV-assisted smart farming. *Applied Sciences*. **13**, 3857 (2023)

[23] Trabelsi, Z., Zhang, L. & Zeidan, S. Dynamic rule and rule-field optimisation for improving firewall performance and security. *IET Information Security*. **8**, 250-257 (2014)