Article

Explainable Reinforcement Learning for Adaptive Cyber Defense in Encrypted Networks

Muhammad Abubakar *

Posted Date: 10 November 2025

doi: 10.20944/preprints202511.0668.v1

Keywords: explainable reinforcement learning; adaptive cyber defense; encrypted traffic; network security automation; interpretable machine intelligence; intrusion response; visibility- limited environments

 doi:10.20944/preprints202511.0668.v1

Article

Explainable Reinforcement Learning for Adaptive Cyber Defense in Encrypted Networks

Muhammad Abubakar

Independent Researcher, Bangladesh; sanisco92@gmail.com

Abstract

The growing adoption of encrypted traffic in enterprise and critical-infrastructure networks has created a defensive paradox: while encryption protects data in transit, it also limits the visibility defenders traditionally rely on to detect and respond to threats. This paper explores an explainable reinforcement learning (XRL) framework designed to support adaptive cyber defense in such visibility-constrained environments. The proposed approach treats the network as a dynamic decision space in which an RL agent learns to anticipate suspicious behavior, reconfigure defenses, and allocate monitoring resources without decrypting traffic. To maintain operational trust, we integrate explainability mechanisms that translate the agent's actions into interpretable cues—highlighting the observed patterns, state transitions, and reward dynamics that influenced each decision. Through this combination of adaptability and transparency, the framework aims to support more resilient intrusion response, reduce false positives, and offer network analysts a clearer understanding of machine-driven defense strategies. The insights presented here outline how XRL can help modernize cyber defense practices for encrypted networks, where traditional signature-driven and static rule-based methods increasingly fall short.

Keywords: explainable reinforcement learning; adaptive cyber defense; encrypted traffic; network security automation; interpretable machine intelligence; intrusion response; visibility-limited environments

1. Introduction

1.1. Background

The rapid expansion of encrypted network traffic has reshaped how organizations protect their digital infrastructures. Encryption now secures the majority of global data-in-transit, shielding users from eavesdropping and tampering. Yet the same protections also restrict the visibility that security teams once relied upon to detect malicious activity. Instead of inspecting payloads directly, defenders must infer intent from metadata, flow behavior, and statistical patterns that are often subtle and easily obscured. As adversaries refine their tactics, traditional signature-based systems and rule-driven policies struggle to keep pace.

1.2. Problem Context: Encrypted Traffic and Defensive Blindness

Encrypted networks create an operational dilemma. Defenders must respond quickly to anomalous or coordinated attacks, but the encrypted nature of the traffic limits the amount of information accessible for decision-making. Tools built for clear-text environments tend to generate high false-positive rates when applied to encrypted flows, and adaptive attackers exploit this uncertainty to blend into ordinary traffic. This defensive gap is particularly visible in large, dynamic environments such as cloud platforms, enterprise data centers, and critical infrastructure networks where traffic patterns evolve continuously.

 doi:10.20944/preprints202511.0668.v1

1.3. Gaps in Current Cyber Defense Approaches

Several approaches attempt to compensate for the loss of visibility, ranging from statistical traffic analysis to machine-learning classifiers trained on flow features. While these methods provide partial support, they still depend heavily on static rules and pre-defined detection models. They rarely adapt to novel adversarial behavior, and even when they perform well, they offer little explanation for their outputs. This lack of transparency fuels operational hesitation: analysts are reluctant to trust automated systems that cannot justify their actions, especially when decisions may disrupt legitimate encrypted sessions or reconfigure network defenses.

1.4. Contribution of This Work

This paper presents an explainable reinforcement learning (XRL) framework designed specifically for adaptive cyber defense in encrypted network environments. The framework treats the network as a dynamic ecosystem in which decisions must be made with incomplete information. The reinforcement learning agent continuously explores defensive strategies, adjusting its policy as it encounters new traffic behaviors or adversarial patterns. To address the trust and accountability issues that often limit the deployment of autonomous systems, we integrate a layer of explainability that clarifies why certain actions are taken. These explanations help security teams understand how the agent interprets network states, what signals influence its decisions, and how its actions relate to long-term defensive goals.

1.5. Structure of the Paper

The remainder of this paper is organized as follows. Section 2 reviews prior work on encrypted network defense, machine-learning-based intrusion detection, reinforcement learning in security contexts, and explainable AI. Section 3 describes the proposed methodology, including the system model, RL formulation, and explainability mechanisms. Section 4 presents experimental results demonstrating the performance of the proposed framework. Section 5 discusses the implications, limitations, and future research directions. Section 6 concludes the paper, followed by required statements and references.

2. Related Work

2.1. Encrypted Network Defense Strategies

Research on defending encrypted networks has expanded rapidly as encryption has become the default mode of communication. Early approaches largely relied on traffic fingerprinting and statistical profiling to approximate what traditional deep-packet inspection once offered. These methods examine flow characteristics such as timing, byte distribution, and packet sequences to infer potential threats. While such analyses provide partial visibility, they often struggle when dealing with sophisticated adversaries who deliberately manipulate flow properties to mimic benign behavior. More recent studies have explored encrypted traffic analytics (ETA) tools that combine flow metadata with lightweight cryptographic indicators, offering improved detection without breaking encryption. However, ETA by itself does not fully resolve the challenge of adapting to evolving adversarial strategies.

2.2. Machine Learning in Cyber Defense

Machine learning has played an increasingly important role in intrusion detection and network monitoring. Classical models such as decision trees, support vector machines, and ensemble classifiers have shown promise in identifying unusual flow patterns derived from encrypted traffic features. Despite their effectiveness in controlled settings, these models typically operate in static environments and require periodic retraining, limiting their ability to respond to real-time changes in attacker behavior. Additionally, many models act as black boxes, offering limited insight into why

 doi:10.20944/preprints202511.0668.v1

certain flows are classified as malicious. This opacity becomes a serious constraint when the output of an algorithm may trigger automated defensive actions that affect live network operations.

2.3. Reinforcement Learning for Security Automation

Reinforcement learning (RL) has attracted significant attention as a tool for automating cyber defense tasks due to its ability to learn adaptive policies. RL agents have been applied to tasks such as intrusion response, firewall rule optimization, and network reconfiguration. These systems allow defenders to move beyond static rule sets by enabling the policy to evolve through interaction with the environment. However, most RL-based security studies focus on environments with full visibility, where the agent has access to detailed packet-level information or explicit attacker models. When RL is applied to encrypted networks, the lack of rich observable features complicates the learning process, potentially slowing convergence or leading to unstable behavior. Furthermore, the inherent opacity of RL decision-making poses a significant barrier to adoption in operational settings where trust and accountability are critical.

2.4. Explainable AI in Security Systems

Explainability has become an important focus within cybersecurity machine learning research. Techniques such as feature attribution, local surrogate models, attention mechanisms, and rule extraction have been explored to help analysts interpret model outputs. In the security domain, explanations serve dual purposes: they enhance operator trust and help identify model biases or blind spots that adversaries could exploit. However, most explainable AI (XAI) research concentrates on classifiers, leaving a gap when it comes to RL agents whose decisions unfold sequentially over time. The challenge is not only to justify individual actions but also to convey how each action contributes to long-term defensive goals. Integrating explainability directly into the RL decision loop remains an emerging research area.

2.5. Summary of Literature Gaps

Taken together, the literature highlights three persistent gaps. First, existing encrypted traffic defense tools provide only partial visibility and lack mechanisms for continuous adaptation. Second, machine learning models designed for security tasks often fail to generalize effectively in dynamic environments where attacker behavior evolves rapidly. Third, although explainability has gained traction, its application to reinforcement learning, particularly in the context of encrypted networks, remains limited. These gaps motivate the need for a framework that unifies adaptive learning with interpretable decision-making, enabling defenders to operate effectively even when network visibility is constrained.

3. Methods

3.1. System Model and Threat Environment

The proposed framework is designed for network environments where encryption is pervasive and packet payloads are inaccessible to monitoring systems. The defender observes only flow-level and metadata-derived features such as packet sizes, timing intervals, session duration, handshake patterns, and cryptographic negotiation details. Adversaries are assumed to operate within these constraints, attempting to embed malicious activities inside traffic that resembles legitimate encrypted sessions. We model the environment as a partially observable system in which the defender must infer underlying threat conditions from limited observations. This setting mirrors realistic enterprise and cloud infrastructures where defenders must respond to anomalies without decrypting traffic or violating privacy requirements.

 doi:10.20944/preprints202511.0668.v1

3.2. Reinforcement Learning Formulation

To support adaptive defensive behavior, we employ a reinforcement learning formulation in which the agent interacts with the network environment over discrete time steps. At each step, the agent receives observable signals, selects a defensive action, and obtains a reward that reflects the quality of the response. Over time, the agent seeks to refine its strategy to maximize long-term defensive effectiveness rather than short-term gains.

3.2.1. State Representation in Encrypted Networks

The state representation is constructed from encrypted traffic characteristics and contextual indicators. These include statistical summaries of recent flows, anomaly scores generated by lightweight detectors, resource utilization metrics from network sensors, and inferred behavioral patterns such as burstiness or protocol irregularities. Because detailed payload data is unavailable, the state emphasizes temporal correlations and structural patterns rather than content-level information. To manage the partial observability inherent in encrypted traffic, the framework allows the agent to maintain an internal representation (for example, through recurrent encodings) that captures longer-term dependencies across flow sequences.

3.2.2. Action Space and Defense Policies

The action space consists of a set of defensive adjustments the agent may apply. These include prioritizing certain flows for deeper statistical inspection, reallocating monitoring resources to different segments of the network, temporarily tightening access-control rules, or initiating an alert for human review. The actions are deliberately defined at a coarse operational level so that the agent influences the defense posture without performing intrusive or sensitive interventions, such as decrypting traffic or terminating sessions prematurely. In this way, the framework supports real-world deployment constraints where automated systems must avoid disrupting legitimate encrypted communications.

3.2.3. Reward Design and Adaptation

The reward function balances several objectives: detecting malicious activity, minimizing false alarms, preserving normal network performance, and avoiding excessive resource consumption. Rewards are shaped to reflect both immediate outcomes and longer-term effects. For example, correctly identifying a suspicious flow yields positive reinforcement, while unnecessary interventions incur penalties. A small discount factor encourages consistent defensive behavior rather than aggressive short-term actions. Through continuous interaction with the environment, the agent learns a policy that adapts to evolving threat characteristics and varying traffic conditions.

3.3. Explainability Layer

Explainability is incorporated as an integral component of the framework rather than an afterthought. Each decision produced by the RL agent is accompanied by a set of interpretable cues indicating the factors that influenced the chosen action. This dual output action plus explanation enables analysts to follow the logic of the system and assess its reliability.

3.3.1. Attribution-Based Explanations

Attribution-based techniques are used to highlight which features or state components contributed most strongly to the agent's decision. Instead of exposing model internals directly, the system produces human-oriented summaries that point to notable changes in flow patterns, shifts in anomaly indicators, or deviations from historical baselines. These summaries allow analysts to understand the underlying triggers without requiring deep familiarity with RL theory.

 doi:10.20944/preprints202511.0668.v1

3.3.2. Policy-Level Explanations

Beyond individual actions, the framework provides higher-level insights into the agent's overall policy. These may include descriptions of recurring behaviors, preferred responses under certain conditions, or broad defensive themes that emerge over time. Such policy-level explanations help practitioners evaluate whether the learned strategy aligns with organizational policies and operational expectations.

3.4. Integration With Encrypted Traffic Analytics (ETA)

The RL agent operates alongside encrypted traffic analytics rather than replacing them. ETA tools supply additional metadata-derived signals that enrich the state representation. These include cryptographic handshake features, endpoint fingerprinting indicators, and statistical models built from historical encrypted communications. The integration ensures the agent benefits from domain-specific insights while maintaining the adaptability inherent to RL.

3.5. Overall Architecture of the Proposed Framework

The complete framework consists of four main components: the observation module, which collects and preprocesses encrypted traffic features; the RL engine, which selects actions based on the evolving policy; the explainability layer, which generates interpretable outputs; and the feedback module, which evaluates the consequences of actions and updates the reward structure. Together, these components form a continuous decision-making loop capable of adapting to dynamic environments while maintaining transparency and operational trust. The architecture is flexible enough to integrate with existing monitoring systems and can be scaled to cloud-based or distributed infrastructures.

4. Results

4.1. Simulation Setup and Datasets

To examine the behavior of the proposed framework, we developed a simulation environment that approximates an enterprise network with a mix of benign and adversarial encrypted traffic. Rather than relying on packet payloads, the dataset used in the experiments consists of flow-level features commonly accessible in encrypted environments packet size sequences, timing intervals, session durations, TLS handshake metadata, and anomaly indicators derived from statistical models. Both synthetic flows and publicly available encrypted traffic traces were incorporated to ensure diversity in traffic patterns. Malicious activity included command-and-control communications, data exfiltration attempts masked within encrypted channels, and reconnaissance behaviors designed to mimic ordinary user traffic. These adversarial flows were interspersed with large volumes of routine encrypted communication, such as web browsing, API calls, and background service traffic. The resulting environment provided a realistic blend of noise, complexity, and uncertainty for the reinforcement learning agent.

4.2. Performance Metrics

Evaluation focused on two categories of outcomes: defensive effectiveness and interpretability quality. Defensive effectiveness was measured using metrics such as detection rate, false-positive rate, response latency, and resource utilization. Because the agent operates under partial observability, we also tracked its stability across different traffic conditions and its ability to converge to a consistent strategy. Interpretability quality was assessed through metrics that reflect the clarity, usefulness, and consistency of the generated explanations, including whether the explanations matched observable state changes and whether they produced actionable insights for human analysts.

 doi:10.20944/preprints202511.0668.v1

4.3. Baseline Models

To contextualize performance, the RL-based defender was compared against several baselines. The first was a static rule-based configuration representative of traditional intrusion detection setups adapted for encrypted traffic. The second baseline consisted of a machine-learning classifier trained on flow metadata; although not adaptive, this model provided a useful benchmark for pattern-recognition capability. A third baseline simulated a non-explainable RL agent to isolate the impact of the explainability layer. These comparisons allowed us to assess not only absolute performance but also how adaptability and interpretability contributed to overall defensive value.

4.4. Evaluation Results

4.4.1. Adaptiveness Under Encrypted Traffic

The RL agent demonstrated an ability to adjust its behavior as traffic patterns shifted. When benign flows changed due to workload fluctuations or service updates, the agent gradually reweighted its expectations and avoided the spike in false positives seen in the static baseline. Under adversarial changes, such as attackers altering timing patterns to evade detection, the agent responded by exploring alternative defensive strategies, preserving detection performance without relying on predefined signatures.

4.4.2. Detection Improvements

Across multiple test scenarios, the RL-based framework achieved higher detection rates than both the rule-based baseline and the static classifier. While the improvement was modest in stable environments, the gains became more pronounced in dynamic conditions where adversarial behavior evolved gradually. The agent's long-term policy allowed it to identify subtle deviations in encrypted flows that short-term models tended to overlook. Importantly, the false-positive rate remained manageable, suggesting that the system did not compensate for increased sensitivity by generating excessive alerts.

4.4.3. Interpretability Outcomes

The addition of the explainability layer produced explanations that aligned with observable changes in the network environment. Analysts reviewing the generated summaries reported that the explanations made it easier to understand why certain flows were prioritized for inspection or why resources were shifted across monitoring points. Policy-level insights revealed recurring defensive themes, such as heightened vigilance during periods of bursty traffic or cautious resource allocation during ambiguous states. These explanations not only aided analyst trust but also helped identify instances where the agent initially converged toward suboptimal strategies, allowing corrective adjustments during training.

4.5. Comparison With State-of-the-Art

When compared with recent research on encrypted traffic defense, the proposed framework offers two practical advantages. First, its adaptive nature reduces dependence on frequent model retraining, a limitation often observed in supervised machine learning approaches. Second, the integrated explainability mechanisms address a persistent concern within RL-based security systems, namely that their decision processes are opaque and difficult to validate. Although the overall detection performance aligns with the best-reported results in metadata-based intrusion detection, the dual emphasis on adaptability and transparency distinguishes the framework from prior work.

 doi:10.20944/preprints202511.0668.v1

5. Discussion

5.1. Implications for Cyber Defense Operations

The results highlight several implications for real-world cyber defense. First, the agent's ability to adapt under shifting network conditions shows that reinforcement learning can serve as a practical complement to existing monitoring tools, especially in environments dominated by encrypted traffic. Traditional systems often require manual rule updates or periodic retraining, which can lag behind the pace of threat evolution. By contrast, the RL agent updates its understanding continuously, reducing the operational burden on analysts and narrowing the window in which attackers can exploit outdated detection logic. The framework also emphasizes the importance of flow-level behavioral patterns rather than payload content. This focus aligns well with privacy-preserving operational requirements, particularly in sectors where decrypting user traffic is not legally or ethically acceptable. The agent's reliance on metadata and statistical indicators suggests that effective defense does not necessarily require deep inspection of packet contents; instead, meaningful insights can be drawn from the structure and dynamics of encrypted sessions.

5.2. Practical Considerations for Deployment

Despite the promise shown by the framework, several practical factors must be considered before deployment in production environments. One challenge concerns resource allocation: RL agents that explore a wide range of defensive actions may temporarily allocate monitoring resources inefficiently during early learning phases. While this issue diminishes as the policy matures, organizations may need to enforce safeguards that constrain exploration in mission-critical networks. Another consideration is the integration with existing tools. Many organizations already rely on layered security architectures composed of firewalls, anomaly detectors, endpoint sensors, and SIEM platforms. The proposed framework is designed to complement these systems, but integration requires careful tuning to avoid redundant alerts or conflicting actions. The explainability layer helps mitigate this by clarifying why particular actions are taken, making it easier for analysts to verify compatibility with existing policies. Finally, the adoption of RL-based defense systems introduces governance questions. Although the agent improves over time, its behavior must align with organizational risk tolerance and regulatory requirements. The ability to generate understandable explanations serves as an important checkpoint, ensuring that autonomous decisions remain transparent and accountable.

5.3. Limitations

This work has several limitations that should be acknowledged. The simulation environment provides a controlled way to evaluate the system, but real-world encrypted traffic exhibits far greater diversity, especially in large-scale cloud infrastructures. Unexpected traffic bursts, sudden protocol shifts, or previously unseen attack strategies may challenge the agent's ability to generalize. Additionally, although the explainability layer greatly improves transparency, the explanations remain abstractions of the underlying model behavior and may not capture all nuances of the decision process. Another limitation is the dependency on high-quality metadata. While encrypted networks expose useful flow features, some environments deliberately minimize metadata exposure to maximize privacy. In such settings, the available signals may be too sparse for the agent to form reliable policies without additional contextual sources.

5.4. Future Research Directions

Several avenues for future research emerge from this study. One direction is to extend the agent's capabilities to multi-agent settings where defenders share information or coordinate actions across different segments of the network. This could improve resilience against distributed attacks and reduce the burden on any single monitoring point. Another promising direction is the

 doi:10.20944/preprints202511.0668.v1

incorporation of adversarial RL techniques that model attacker behavior more realistically, allowing the defender to train against a stronger and more adaptive opponent. Enhancing the explainability layer is also a priority. Future work could explore interactive explanations that adjust to analyst preferences or provide deeper narratives linking sequences of actions to broader defensive strategies. Finally, rigorous testing in real operational networks, particularly those with heterogeneous traffic and strict performance requirements, would be essential to validate the framework's readiness for deployment.

6. Conclusions

Encrypted networks have become the backbone of modern digital communication, yet their protective qualities also restrict the visibility traditionally used for threat detection. This work presented an explainable reinforcement learning framework designed to operate effectively in such environments. By focusing on flow-level behavioral cues and adapting its policy through continuous interaction, the RL agent demonstrated an ability to detect subtle malicious patterns embedded within encrypted traffic. The integrated explainability layer further strengthened the framework by providing interpretable insights into both short-term decisions and longer-term defensive strategies, helping to bridge the trust gap that often limits the adoption of autonomous security systems. The findings suggest that adaptability and transparency can coexist within a single defensive architecture, offering a realistic path forward for organizations seeking to modernize their security posture without compromising privacy or operational stability. While the results are encouraging, broader evaluation across heterogeneous, high-volume networks is needed to fully assess generalizability. Nonetheless, the framework outlined in this study offers a foundation for more responsive and accountable cyber defense systems tailored to the realities of encrypted communication.

Data Availability Statement: The data used or generated during the study can be provided upon reasonable request, unless restricted by confidentiality or security requirements. When possible, datasets derived from public encrypted-traffic repositories were utilized to support reproducibility. Authors are encouraged to deposit supplementary data in recognized repositories following FAIR data principles.

Conflicts of Interest: The authors declare that they have no known financial or personal conflicts of interest that could have influenced the work presented in this manuscript.

References

1. Hussain, M. K., Rahman, M. M., Soumik, M. S., & Alam, Z. N. (2025). Business Intelligence-Driven Cybersecurity for Operational Excellence: Enhancing Threat Detection, Risk Mitigation, and Decision-Making in Industrial Enterprises. *Journal of Business and Management Studies*, *7*(6), 39-52.
2. Hussain, M. K., Rahman, M. M., Soumik, M. S., Alam, Z. N., & RAHAMAN, M. A. (2025). Applying Deep Learning and Generative AI in US Industrial Manufacturing: Fast-Tracking Prototyping, Managing Export Controls, and Enhancing IP Strategy. *Journal of Business and Management Studies*, *7*(6), 24-38.
3. Rahman, M. M., Soumik, M. S., Farids, M. S., Abdullah, C. A., Sutrudhar, B., Ali, M., & HOSSAIN, M. S. (2024). Explainable Anomaly Detection in Encrypted Network Traffic Using Data Analytics. *Journal of Computer Science and Technology Studies*, *6*(1), 272-281.
4. Soumik, M. S., Omim, S., Khan, H. A., & Sarkar, M. (2024). Dynamic Risk Scoring of Third-Party Data Feeds and Apis for Cyber Threat Intelligence. *Journal of Computer Science and Technology Studies*, *6*(1), 282-292.
5. Rjoub, G., Bentahar, J., Abdel Wahab, O., Mizouni, R., Song, A., Cohen, R., & Otrok, H. (2023). *A Survey on Explainable Artificial Intelligence for Cybersecurity*. arXiv preprint.
6. Zhang, Z., Al Hamadi, H., Damiani, E., Yeun, C. Y., & Taher, F. (2022). *Explainable Artificial Intelligence Applications in Cyber Security: State-of-the-Art in Research*. IEEE Access.

 doi:10.20944/preprints202511.0668.v1

7. Premakumari, S. B. N., Sundaram, G., Rivera, M., Wheeler, P., & Pérez Guzmán, R. E. (2025). *Reinforcement Q-Learning-Based Adaptive Encryption Model for Cyberthreat Mitigation in Wireless Sensor Networks*. Sensors, 25(7), 2056.
8. Alnfiai, M. M. (2025). *AI-powered Cyber Resilience: A Reinforcement Learning Approach for Automated Threat Hunting in 5G Networks*. EURASIP Journal on Wireless Communications and Networking, 2025:68.
9. Abouhawwash, M. (2024). *Innovations in Cyber Defense with Deep Reinforcement Learning: A Concise and Contemporary Review*. Artificial Intelligence in Cybersecurity, 1, 44-51.