

Reinforcement Learning-Based Framework for Cybersecurity in Online Examination Systems

Anurag Shah¹, Manoj Kumar², Sarika Jain¹

¹Amity Institute of Information Technology, Amity University Uttar Pradesh, Noida, India

²Faculty of Engineering & Information Sciences, University of Wollongong, Dubai, UAE,

anurag@nielit.gov.in, manojkumar@uowdubai.ac.ae, sjain@amity.edu

Abstract. As cyber security threats are becoming more and more complex, sophisticated and persistent, the methods of traditional defence mechanism are becoming obsolete and are struggling hard in keeping pace cost-effectively to preventing the black swan^[1] threats of the today's cyberworld, that are poised with the advent of Artificial Intelligence, hence need is being felt for the adaptive capabilities of the cybersecurity tools along with the dynamic decision-making abilities to mitigate the ever evolving cybersecurity risks associated not only in the cyberworld but also with the online examination systems being used today in every step of the learning process as well as the shortlisting process of any institute/organisation. Machine Learning (ML) must be used to counter the attacks on the system today, as the same are increasing exponentially day by day. For this, Reinforcement Learning (RL), which is a powerful subset of ML, must be effectively used to create strategies in dynamic environments with varied hardware and software in different scenarios to effectively detect and control the ever-evolving cybersecurity risks and threats while effectively addressing the fatigue in manual processes and providing real-time decision-making for the stakeholders. In this paper, we are investigating extensively the application of RL techniques to provide adaptive decision-making processes by enabling systems with proactive decision-making in real-time changing threat scenarios. This paper evaluates the present state-of-the-art RL-based methods, while highlighting their strengths and weaknesses, and it also identifies the areas that require further improvements to propose a novel RL-based cybersecurity framework tailored to the specific demands of online examination platforms.

Keywords: Cybersecurity, Reinforcement Learning, Online Examination, Intrusion Detection, Adaptive Defense, Zero-day attack.

(Article history: Selected papers from NIELIT's International Conference on Communication, Electronics and Digital Technologies (NICE-DT'26), 29-30Jan, 2026, Gorakhpur, India)

1 INTRODUCTION

Since the current security environment remains largely stagnant, it is hard to adequately address the ever-changing cyber security threat due to complexity and sophistication, posing a serious challenge towards the means of cybersecurity in practice as it is today. The currently used solutions to conventional cybersecurity are typically defined by the utilisation of predetermined rules and heuristics and cannot fully adapt and react to the quickly evolving attack patterns. The traditional approaches usually fail when they face advanced zero-day attacks as well as new attacks and hence cause vulnerabilities that make systems vulnerable to compromise. Although Reinforcement Learning (RL), as a strong subdivision of ML, offers a dynamic alternative to cybersecurity as the deployed systems can learn in real-time to apply the best defence strategies by dealing with the interactions, reviewing feedback, and constant analysis of the system in the real-time running environment [2][6]. The traditional methods are fixed and rely on the existing rules and heuristics, but the RL models are able to dynamically modify the RL defence mechanisms of the deployed cybersecurity system in real-time so as to efficiently identify and react to the cyber threats on the zero-day [1][4].

Following the COVID-19 pandemic, there is a fast-tracked movement to adapt to a digitally assisted education framework that has significantly contributed to the dependence on web-based examination frameworks. These remote proctored online examination systems were not even thought of as a system that could be implemented in the next decade, but rather an overnight adoption by the institutions and academia. These systems are already handling sensitive academic and personal data and have become the subjects of the elaborate and sophisticated cyberattacks that involve impersonation, session hijacking, injection attacks, unauthorised access and distributed denial-of-service (DDoS) attacks. Current security systems (firewalls, intrusion detection systems (IDS) and traditional authentication mechanisms) are not serving well in cost-effective manner because of their lack of flexibility and adaptability in these dynamically changing complex threat situations. The current traditional measures lack independent detection and response capabilities and, therefore, could lead to a lack of responsiveness in the mitigation of threats, which may be too late or ineffective.

Reinforcement Learning (RL) has been in the limelight as being highly capable of independently learning to find the best decisions after iteratively engaging in cyber environments and emerging as a potential solution to new cybersecurity [6]. As a

result of continuous learning through real-time feedback, RL agents prove to be very useful in detecting cyber threats, dynamically adapting to defence strategies, and significantly enhancing cybersecurity capabilities, especially in sensitive and dynamic systems like online test systems [2].

The objective of this study is to address critical gaps in the body of knowledge in the field of cybersecurity using RL approaches, with special emphasis on the following main areas:

- **Adaptive Threat Mitigation:** The RL-based agents will be capable of dynamically changing defensive strategies based on the current threat analysis and adaptability of attack strategies, thereby enhancing the efficiency with which the system detects and responds to advanced cybersecurity threats.
- **Real-Time decision making:** RL models reduce response times considerably in situations when emergencies need to be addressed because of their capacity to make quick, automatic decisions in real-time. This capability to act fast is particularly helpful when it comes to such tasks as online tests, where authenticity and immediate action play a crucial role.
- **Compatibility with Existing Infrastructure:** The proposed RL cybersecurity model should be interoperable with the existing security hardware, e.g., Intrusion Detection Systems (IDS), encryption, and biometric authentication systems as well, to make the implementation of this model in existing educational technology infrastructure as easy and feasible as possible.

The paper provides a detailed overview of the existing state-of-the-art reinforcement learning (RL) methods in the field of cybersecurity and contains both advantages and disadvantages of the field, as well as provides unequivocal research directions for the future. The efficacy and strength of the developed RL-based adaptive framework are confirmed with the help of empirical tests conducted based on the general benchmarking data collections, in particular, the CIC-IDS-2018[7] dataset. The results are dramatic as the number of false positives can be reduced by up to 30% and the detection of a threat can be improved by approximately 22 per cent relative to traditional rule-based methods.

2 LITERATURE REVIEW

2.1 Introduction to Reinforcement Learning.

Reinforcement Learning (RL) is a subdiscipline of machine learning that is interested in goal-directed decision making, i.e. in the case of intelligent agents having a dialogue with the environment to know what action to take that will provide the highest rewards. The RL operates on a cycle of operation that includes the situation, the behaviour, as well as the return nodes. Through this process, the agents can decide their policy of decision-making in a step-by-step manner by utilising the feedback they get with the help of the environment. This adaptive-based learning process can help RL to deal with dynamic and complex cybersecurity threats since it can respond autonomously to changing threat conditions.

2.2 Cyber security Applications

Reinforcement Learning is one of the important technological methods to enhance various aspects of cybersecurity [13]. It has particularly been in use with threat monitoring and pre-emptive defence. For example:

- **Intrusion Detection:** RL algorithms have led to tremendous advancement of the intrusion-detecting power through their ability to alter their mechanisms in response to the current attack patterns [3]. Indicatively, Q-Learning has proven to be very useful in identifying and removing malicious network traffic automatically, therefore capable of deploying defences more timely and precise than the traditional static methods.
- **Malware Detection:** The malware detection has improved significantly with RL-based cybersecurity solutions because this type of system has the ability to tune its malware detection strategies on the basis of newly observed malware behaviours and signatures. This is what provides RL systems with the flexibility to identify malware threats in a timely and appropriate manner and remove them when signature-based defences might be ineffective.
- **Network Defence:** Dynamic network defence policies have been constructed successfully using network reinforcement learning (DRL). The defence systems enabled with DRL are able to modify strategies based on the behaviour of the network and any changes in the attack methods observed, and provide strong protection against various and advanced cybersecurity threats [12][15].

2.3 Reinforcement Learning Advances in Cybersecurity

The outcomes of RL have seen very important breakthroughs in a diversity of cybersecurity systems [8]:

- **Deep Q-Networks (DQN):** DQN is one of the success stories in the domain of autonomous penetration testing, in which it has succeeded in achieving an accuracy of up to about 95 per cent in the detection of potential vulnerabilities and consequently, the level of cybersecurity has been significantly increased [9][16].

- **Double Deep Q-Networks (DDQN):** The DDQNs have managed to achieve false-positive warnings in the domain of intrusion detection to a minimal level of 18, as evidenced by carrying out research with NSL-KDD data. This upgrade raises the reliability and productivity of the cybersecurity systems to an extremely high level.
- **Policy Gradient Methods:** These RL techniques can enhance the flexibility and timeliness of cybersecurity systems to a great extent, which makes them especially useful in settings where the network conditions vary, such as during periods of increased demand during online examinations.

2.4 Security Challenges in Online Examination Systems

There are unique cybersecurity threats to online examination platforms, and the defences need to be sophisticated and adaptive:

- **Zero-Day Threats:** The traditional security mechanisms based on known threat signatures are not sufficient to address the new and unknown cyber threats, which present significant risks to exam integrity, which cannot be compromised.
- **High Security Stakes:** This means the security concerns, done to keep the dignity of the exam situation, where the cybersecurity strategies, which are essential to be applied. The ones that have the best accuracy in detecting threats and the lowest number of false detections are mainly those that are the most effective at detecting the threat and minimising the risk of fraud through the tests.
- **Scalability of the system:** The requirement to provide a large number of users and traffic in the network simultaneously, particularly at the peak period of examination, necessitates cyber security solutions that are extremely scalable and hard-wearing that guaranteeing the appropriate and safe system functioning.

3 METHODOLOGY

The reinforcement learning (RL) model that the author uses to make cybersecurity decisions represents a systematic approach that includes several different components:

3.1 States:

States are elaborate statements of the current cybersecurity landscape, which contain live data regarding the state of the system, the count of security breaches that have ever taken place, the count of unusual traffic patterns and the indicators of possible attacks. To be guided properly throughout the process of decision-making, state descriptions incorporate the data provided in numerous monitoring platforms and the information distributed in the logs of the system.

In particular, the RL approach defines states in terms of:

- **Network Traffic Metrics:** Records the accurate information like packet Flow rates, Diversity of source IP addresses, pattern of use of a protocol and indicators that verify normal traffic as opposed to anomalies.
- **User Behavioural Analysis:** Here, user behaviour analysis consists of user-centric metrics, which are the frequency of login attempts, success in authentication, biometric check results, and user timing patterns.
- **System Performance Indicators:** System performance indicators sphere involves server load, CPU load, latency, response times, memory usage, and the stability in general. It is an appraisal reflecting the performance of these resources available as well as the resources upon which the service relies.

3.2 Actions:

The activities in this RL model consist of the cybersecurity operational responses to the existing threats. The possible responses include blocking of suspicious IP addresses, isolating infected network segments, regulating or restricting suspicious traffic, launching a targeted security scan, as well as dynamically updating the firewall settings. These are situational responses, and they are strategically planned to adequately respond to the perceived threats.

3.3 Rewards:

The reward structure is used to define the way in which Rewards can be used as a motivational tool to make the most effective security measures in the cyber area. They provide stimuli to actions, which would be effective to counter cyber threats and secure the system, as well as minimise false alerts. On the other hand, negative incentives are used to deter such behaviours that lead to unnecessary interruption, slow reaction, or exposure to vulnerability. The reward criteria will consist of the following:



Positive Rewards:

Successful prevention of a threat will be rewarded with +10, which serves as an incentive to take preventive and adequate steps.

False-negative alert degradation will be rewarded with +5, which will consequently become an incentive towards proper identification of threats.

Penalties:

A larger penalty of -15 is the cost of not detecting real threats (false negatives) and, as such, will not allow the improper detections.

Interruptions within the examinations underway are punished by -20, and it highlights the necessity of continuous system functioning.

3.4 Algorithms:

Q-Learning, Deep Q-Networks (DQN), and policy gradient algorithms have been identified as the best RL algorithms to use in this application because they have demonstrated effectiveness in dynamic and complex cybersecurity scenarios. These algorithms are superior in that they both learn optimal decision-making policies via active interactions with cybersecurity situations and simultaneously optimise the exploration (assessing new actions) and exploitation (exploiting already known successful strategies) processes.

4 PROPOSED FRAMEWORK

The customized cyber security system that is implemented by RL to protect online learning websites and implementing is made up of three separate modules:

4.1 Threat Detection Module

Applies the Double Deep Q-Networks (DDQN) in combination with the method of the prioritised experience replay to allow not only identifying but also classifying the abnormal behaviour in examination network traffic in real time.

Priority on experience replay provides a better learning experience by prioritising significant or uncommon security events. This allows adaptation to new threats to be done faster.

This module constantly examines the incoming traffic data, assimilative determines any possible security risk, detects with higher accuracy, accelerates the speed of response, and at the same time minimises the false-positive notifications.

4.2 Response Selection Module

Uses the Proximal Policy Optimisation (PPO), which is a policy gradient approach that does not require the identification of the threats in the initial stage.

PPO is suitable in decision-making situations that are complicated and with high-dimensional representations of states, and gives stable and robust policy improvements.

After identifying threats, this module dynamically picks appropriate security responses such as blocking of IP addresses, selective network isolation, throttling of suspicious activities, and real-time firewall rule changes. These kinds of strategies quickly suppress threats and reduce the operational effect.

4.3 Recovery Module

- Uses Monte Carlo Tree Search (MCTS) to carefully analyse and select optimal recovery strategies after cybersecurity breaches^[10].
- MCTS performs a detailed analysis of different recovery strategies as well as the involved long-term aspects, identifying the ones that least disrupt the system and require fewer resources, and thus, facilitate quick recovery of normal operations.
- The said method will not only lead to a fast recovery but will also incorporate the lessons learnt into future cybersecurity practices, thus enhancing the long-term resilience of the system.

The result is a connected and agile cybersecurity framework that - besides actively detecting and effectively eliminating cyber threats - quickly returns to normal, all in one process, and is an online examination environment-specific for integrity and security.

5 IMPLEMENTATION AND RESULTS

5.1 Dataset and Training

The Intrusion Detection System (IDS) developed in this work is based on Reinforcement Learning (RL) and was trained and tested on the CIC-IDS-2018 [7] dataset. This data set is commonly accepted as a standard benchmark that represents actual network traffic scenarios correctly; hence, it is primarily appropriate in simulating the network environment that arises when conducting online exams. Besides the authentic forms of attacks that have been factored into the dataset, including DDoS, brute force entry, and SQL injection, the RL-based IDS is testable and applicable in the resolution of cybersecurity issues that are commonly encountered in high-stakes academic exams.

The hyper parameters were also configured in a specific manner to assist the RL agent to learn in a more efficient manner, namely:

- Size of Batch: 64 (to ensure that the agent is efficient and still stable in its training simultaneously)
- Learning Rate: 0.001 (chosen to ensure stable convergence)
- Discount Factor (g): 0.95 (offering an effective compromise between short-term and long-term rewards)

5.2 Performance Metrics

To test the efficiency of the RL-based IDS strictly, the performance of the method was contrasted with the important parameters of the traditional approach to IDS, which is rule-based:

Metric	RL-Based System	Traditional IDS
Detection Accuracy (%)	83.2	76.5
False Positive Rate (%)	8.8	12.4
Response Time (ms)	240	450

The RL-based IDS was significantly better than the conventional rule-based method, with a detection accuracy of 83.2 and a significant improvement of 6.7. This development highlights the enhanced ability of the RL system to properly detect actual security threats in a variety of different and complex attack scenarios. Besides, the RL agent significantly reduced false positives to a minimum of 8.8, which significantly reduced the unwarranted user interference, which is essential, especially in the case of online examinations.

It is not to say that the RL-based IDS has succeeded in ensuring the average response time is lowered to 240 ms, which is a significant novelty in comparison to the response time of the traditional IDS, which was 450 ms. Heavy reduction of the response time is highly significant where a high-speed threat resolver is required, e.g. in online examinations in real-time to maintain the process intact and the system operational with continuous traffic. This can be the case in an event where an urgent reaction will be needed.

6 MATHEMATICAL FORMULATION

The recommended Intrusion Detection System (IDS) is an RL-based system that can be modelled using a Markov Decision Process [11][14] (MDP), which can be represented by the following:

$$M = (S, A, P, R, \gamma)$$

- S denotes the set of observable states, representing the extracted features from network traffic flows.
- A is the set of possible actions (e.g., label as benign or various attack types).
- $P(s'|s, a)$ defines the transition probability from state s to s' upon taking action a.
- $R(s, a)$ is the reward received after acting a in state s.
- $\gamma \in [0,1]$ represents the discount factor (0.95 by default) that makes a trade-off between the value of immediate and future rewards.

The task of the RL agent is to figure out an optimal policy $\pi^*(a|s)$ that will get the expected cumulative discounted reward maximised:

$$\pi^* = \arg \max_{\pi} E_{\pi} [\sum (\gamma^t * R(s_t, a_t))]$$

The reward function to be used to penalise false positives is as follows:

$$R(s, a) = \{ \begin{aligned} &+1, \text{ if a correctly detects an attack or benign traffic} \\ &-1, \text{ if a misclassifies (FN or FP)} \\ &-\lambda, \text{ if a causes a false positive, } \lambda > 1 \end{aligned} \}$$

The agent's policy is optimised using the Q-learning update rule:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha [R(s_t, a_t) + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)]$$

Where $\alpha = 0.001$ is the learning rate used to balance convergence speed and stability.

Detection effectiveness is evaluated using standard classification metrics:

$$\text{Detection Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

$$\text{False Positive Rate} = FP / (FP + TN)$$

With such mathematical grounds, the RL agent can outperform the traditional IDS models in accuracy and response time, as well as learn adaptive dynamic methods of intrusion detection without much human participation.

7 CHALLENGES AND FUTURE WORK

More than impressive outcomes in terms of securing exam-day networks can be achieved through the RL application in the sphere of cybersecurity; numerous issues associated with the nature and flaws should be addressed.

7.1 Challenges

Reward Function Complexity: The development of a suitable reward function for intrusion detection based on RL is always challenging. Rewarding system must be stable enough to monitor the short-term detection and long-term stability. The results of cybersecurity may, however, be context sensitive, uncertain or delayed, and therefore the specification of rewards becomes difficult.

Partial Visibility: Cybersecurity network environments are often not fully visible because of encryption, the network design is segmented, or the traffic operations are camouflaged. This incomplete observability prevents the precision of the learning of an RL agent, which can lead to suboptimal policy making, especially in the case of advanced persistent threats.

Exploration-Exploitation Dilemma: The fact that RL must be successful depending on how far it can go to balance exploration and exploitation (i.e. attempt new strategies and continue to utilise already proven successful actions) is its nature. Security can, however, act as a catalyst in the exploration process, should it be a cause of erroneous or timid reaction within the course of the training process, in the case of an online exam, for example. Safe exploration has been a problem to be tackled and not yet achieved.

Computational Resource Requirements: The training of complex RL models, such as deep Q-networks and policy gradient algorithms, is highly-computationally expensive and generally requires the use of a GPU accelerator. This may limit the practicality of such computations in live online exam settings where timely flexibility is essential unless optimally optimised.

Absence of Specialised Datasets: There is an evident lack of good-quality datasets that represent real, exam-specific cyberattacks such as targeted denial-of-service, exam fraud, or impersonation attempts. This makes the models more difficult to extend and less stable, which typically requires applying synthetic data generation or domain adaptation methods.

Threats to Adversarial Manipulation: The policies obtained through reinforcement learning are not resistant to attacks. Defences can be reverse-engineered by trained attackers to design inputs or reverse-engineer them to pose spurious false positives or false negatives of the RL models. Consequently, the formulation of adversarial robust rules is a research question of significance to remains unsolved.

7.2 Future Directions

Federated Learning to Decentralised Detection: Federated learning may help to improve cybersecurity by allowing different organisations or learning systems to share their resources in shared efforts to train RL-oriented models without jeopardising the privacy of sensitive information. Although the training process is decentralised, the players are also able to make the model resistant to various conditions of the network by improving the quality of the input data they provide.

Hybrid Reinforcement Learning Architectures: A combination of RL and supervised or unsupervised machine learning algorithms may form hybrid systems that possess greater robustness, as well as greater interpretability. The hybrid models may be dynamically able to switch between learning paradigms depending on network conditions or threat signatures to improve the detection abilities as well as response accuracy.

Cross-Platform Compatibility: The next stage of the development will be to reach the Higher Performance Framework-based Intrusion Detection Systems that will be compatible with other infrastructures and test settings. It will be critical in this case to have policy frameworks that are generalised to an extent that they are capable of working even when the terms of network topologies, security features and traffic configurations and even do not require retraining with high frequencies.

Real-Time, Lightweight Inference: Future efforts ought to be directed towards the creation of lightweight RL agents that can execute within a lightweight or edge computing platform to have the capability of managing real-time detection and response without the requirement to utilise substantial backend computing units [5]. Such approaches as knowledge distillation, model compression and hardware-conscious design are all possible options to reduce the inference latency.

8 CONCLUSION

Reinforcement learning, or RL, has become the new leader of cybersecurity with its remarkable ability to devise flexible and smart solutions under the circumstances of constantly evolving and complex cyberthreats. The fact that RL is the reverse of conventional rule-based or fixed machine learning systems is that RL actively learning and adapts to the interactions with the environment. This aspect allows the development of proactive defence measures, which are self-evolved with time. The new RL-based approach presented in this paper aims to enhance digital cheating prevention, which belongs to the educational industry and, regrettably, is the most frequent victim of cyberattacks.

In the meantime, the policy gradient methodology with sophisticated RL models, and specifically, the Double Deep Q-Networks (DDQN) in the given cybersecurity framework, provide outstanding detection rates (83.2) and, simultaneously, a significant decrease in false-positive notifications and response times. The features are critical towards the integrity and effectiveness of exam platforms, particularly where high real-time conditions are enforced. This fact is further confirmed by the adoption of the CIC-IDS-2018 [7] dataset, which confirms the resilience of the system to all the common cybersecurity-related attacks, such as DDoS attacks, brute force attacks, and SQL injection attacks.

One of the key strengths of the chosen framework is its flexibility. Due to this, it is in a good position to handle both known and unknown cyber threats. Consequently, it can achieve so much in terms of developing cybersecurity solutions that are flexible and scalable. As a matter of fact, reinforcement learning (RL)-based intrusion detection systems have already demonstrated that they can be superior to the traditional ones, particularly regarding more rapid response and more precise threat detection.

Nonetheless, one can still face some significant obstacles. It is not an easy task to design a reward function that directs the system to work effectively, and fighting adversarial attacks is not a small task as well. In addition, such systems may need a large portion of computing power. The solutions to these issues are needed in case RL can be used extensively in real-world cybersecurity.

This paper brings out the potential of RL as an essential component of the digital future of cybersecurity, particularly in digital education. It lays the foundation for the further work on the creation of intelligent and self-adaptable defence systems that should be able to secure any delicate digital infrastructure in the constantly evolving cyberspace.

9 REFERENCES

1. Ale, B. J. M., Hartford, D. N. D., & Slater, D. H. (2020). Dragons, black swans and decisions. *Environmental Research*, 183. <https://doi.org/10.1016/j.envres.2020.109127>
2. Al-Nawashi, M. M., Al-hazaimah, O. M., Nedal, T. M., Gharaibeh, N., Abu-Ain, W. A., & Abu-Ain, T. (2025). Deep Reinforcement Learning-Based Framework for Enhancing Cybersecurity. *International Journal of Interactive Mobile Technologies (IJIM)*, 19(03), 170–190. <https://doi.org/10.3991/ijim.v19i03.50727>
3. Chen, H., Lai, Y., Liu, J., & Wanyan, H. (2024). Interpretable Cross-Layer Intrusion Response System Based on Deep Reinforcement Learning for Industrial Control Systems. *IEEE Transactions on Industrial Informatics*. <https://doi.org/10.1109/TII.2024.3388672>
4. Ibrahim, M., & Elhafiz, R. (2023). Security Analysis of Cyber-Physical Systems Using Reinforcement Learning. *Sensors (Basel, Switzerland)*, 23(3), 1634. <https://doi.org/10.3390/s23031634>
5. Mao, B., Qiu, J., & Kato, N. (2023). On An Intelligent Task Offloading Model to Jointly Optimise Latency and Energy for Electric Connected Vehicles. *IEEE Transactions on Vehicular Technology*. <https://doi.org/10.1109/TVT.2023.3edge333241>
6. Mavroudis, V., Palmer, G., Farmer, S., Whitehead, K. S., Foster, D., Price, A., Miles, I., Caron, A., & Pasteris, S. (2025). Guidelines for Applying RL and MARL in Cybersecurity Applications. *arXiv*. <https://doi.org/10.48550/arXiv.2503.04262>
7. Mohamed, Abdissalam (2024), “CSE-CIC-IDS2018”, Mendeley Data, V1, doi: 10.17632/29hdbdxx2r.1
8. Ogundairo, O., & Brooklyn, P. (2024). EasyChair Preprint Enhancing Cybersecurity Protocols with Reinforcement Learning Enhancing Cybersecurity Protocols with Reinforcement Learning. <https://easychair.org/publications/preprint/MfKF>
9. Oh, S. H., Jeong, M. K., Kim, H. C., & Park, J. (2023). Applying Reinforcement Learning for Enhanced Cybersecurity against Adversarial Simulation. *Sensors*, 23(6). <https://doi.org/10.3390/s23063000>
10. Ponnolly, J., Puthenveetil, J., & D’Urso, P. (2024). Prescriptive Analytics-based Robust Decision-Making Model for Cyber Disaster Risk Reduction. 2024 IEEE 3rd International Conference on AI in Cybersecurity, ICAIC 2024. <https://doi.org/10.1109/ICAIC60265.2024.10433837>
11. Ravari, A., Ghoreishi, S. F., & Imani, M. (2024). Optimal Inference of Hidden Markov Models Through Expert-Acquired Data. *IEEE Transactions on Artificial Intelligence*. <https://doi.org/10.1109/TAI.2024.3358261>
12. Ren, J., & Qin, T. (2024). *Decentralised Blockchain-Based and Trust-Aware Task Offloading Strategy for Healthcare IoT. IEEE Internet of Things Journal*, 11(1), 829–847. <https://doi.org/10.1109/JIOT.2023.3286900>
13. Security, H. N. (2024). Reinforcement learning is the path forward for AI integration into cybersecurity. In *Help Net Security*. <https://www.helpnetsecurity.com/2024/03/26/ai-reinforcement-learning/>
14. Tan, Y., Liu, J., & Wang, J. (2024). 5G End-to-End Slice Embedding Based On Heterogeneous Graph Neural Network and Reinforcement Learning. *IEEE Transactions on Cognitive Communications and Networking*. <https://doi.org/10.1109/TCCN.2024.3349452>
15. Ye, D., Zhu, T., Gao, K., & Zhou, W. (2024). Defending Against Label-Only Attacks via Meta-Reinforcement Learning. *IEEE Transactions on Information Forensics and Security*, 19, 3295–3308. <https://doi.org/10.1109/TIFS.2024.3357292>
16. Yu, Y., Liu, J., Guo, H., Mao, B., & Kato, N. (2024). A Spatiotemporal Backdoor Attack Against Behaviour-Oriented Decision Makers in Metaverse: From the Perspective of Autonomous Driving. *IEEE Journal on Selected Areas in Communications*, 42(4), 948–962. <https://doi.org/10.1109/JSAC.2023.3345379>