


Paper Type: Review Article

Innovations in Cyber Defense with Deep Reinforcement Learning: A Concise and Contemporary Review

Mohamed Abouhawwash^{1,2,*} 

¹ Department of Computational Mathematics, Science, and Engineering (CMSE), College of Engineering, Michigan State University, East Lansing, MI 48824, USA; abouhaww@msu.edu.

² Department of Mathematics, Faculty of Science, Mansoura University, Mansoura 35516, Egypt.

Received: 20 Feb 2024

Revised: 16 May 2024

Accepted: 11 Jun 2024

Published: 14 Jun 2024

Abstract


This study presents a concise review for exploring the burgeoning intersection of Deep Reinforcement Learning (DRL) and cybersecurity, delving into its basics, applications, and open challenges. In particular, DRL is introduced as a dynamic approach to cybersecurity, permitting adaptive threat detection, intrusion prevention, and occurrence response through continuous learning and decision-making. Nevertheless, there are many technical, operational, and ethical challenges that obstruct its widespread adoption, including data scarcity, computational complexity, vulnerability to adversarial attacks, and privacy concerns. To deal with these obstacles, researchers and practitioners must work together to come up with strong and ethical DRL-based security solutions. However, despite these difficulties, integrating DRL into cybersecurity frameworks may be a promising way to improve resilience against evolving cyber threats. Through tackling its limitations and utilizing its promise, we can create a more robust, quick-reacting cyberspace.


Keywords: Deep Reinforcement Learning, Cybersecurity, Cyber Defense, Artificial Intelligence, Machine Learning, Autonomous Security Systems, Threat Detection, Intrusion Detection Systems, Network Security, Adaptive Security.

1 | Introduction

The digital interconnection of today's world has made cyber security a major issue for people, companies, and countries [1]. With the rapid growth in technology as well as the increasing sophistication of cyber threats, novel approaches should be developed to safeguard sensitive information and maintain the integrity of digital infrastructure. Conventional measures in cybersecurity are inadequate due to ever-changing cyber-attack methods [2]. This has attracted great interest in advanced artificial intelligence (AI) techniques that can boost cyber defenses [3]. One of the interesting research directions is Deep Reinforcement Learning (DRL), which is a subfield of machine learning that combines reinforcement learning's decision-making powers with deep learning's strong feature extraction capabilities. Through interaction with their environment, DRL has been highly successful in areas like gaming, robotics, and autonomous systems whose agents learn optimal policies. The use of DRL could revolutionize cyber security by adaptively reacting toward threats, acquiring knowledge from new attack patterns, and independently making choices to reduce risks [4-5].

 Corresponding Author: abouhaww@msu.edu

 <https://doi.org/10.61356/j.aics.2024.1298>

 Licensee **Artificial Intelligence in Cybersecurity**. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0>).

This mini-review seeks to travel the joining of DRL and cybersecurity, providing a comprehensive overview of how DRL techniques are being applied to enhance cyber defense strategies [6-7]. This study starts by outlining the foundational concepts of DRL and then examines its current applications in cybersecurity, including threat detection, intrusion prevention, and automated incident response. This also discusses notable real-world implementations to illustrate the practical benefits and challenges of integrating DRL into cybersecurity frameworks [8].

Furthermore, we will address the limitations and ethical considerations associated with deploying DRL-based systems in security-critical environments. Finally, we will highlight emerging trends and future research directions, offering insights into how DRL can continue to evolve and contribute to a more secure digital future as shown in Figure 1.

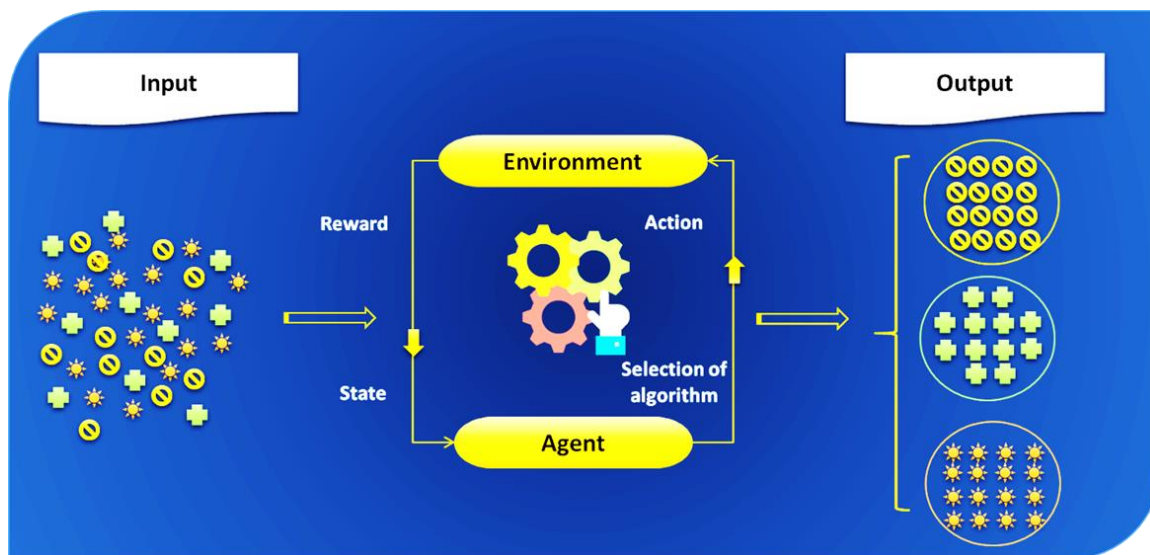


Figure 1. Visualization of the basic idea of RL.

2 | Foundations of Reinforcement Learning

Reinforcement Learning (RL) is a branch of machine learning concerned with how agents ought to take actions in an environment to maximize some notion of cumulative reward. Unlike supervised learning, where the model learns from a fixed dataset of labeled examples, RL is based on the idea of learning through interaction with an environment. This learning process is typically framed as a Markov Decision Process (MDP) and involves several key components:

- **Agent:** The learner or decision-maker that interacts with the environment.
- **Environment:** Everything the agent interacts with and learns from. The environment provides feedback to the agent in the form of rewards and new states.
- **State (s):** A representation of the current situation of the agent in the environment.
- **Action (a):** Choices available to the agent that affect the state of the environment.
- **Reward (r):** A scalar feedback signal received after taking an action, guiding the agent's learning process.
- **Policy (π):** The strategy that the agent employs to determine its actions based on the current state.
- **Value Function (V or Q):** Estimates the expected cumulative reward for states (V) or state-action pairs (Q), guiding the agent in choosing actions that maximize long-term rewards.

The agent's objective is to learn a policy that maximizes the expected cumulative reward, often referred to as the return. The return is typically discounted over time to prioritize immediate rewards over distant ones [9-11].

2.1 | Deep Learning Integration with RL

Deep Learning (DL) has revolutionized many fields by providing powerful techniques to automatically learn representations from raw data. By integrating DL with RL, we can tackle complex problems with high-dimensional state and action spaces that are intractable for traditional RL algorithms. This integration leverages the ability of deep neural networks to approximate complex functions, enabling RL to scale to real-world problems [12]. In traditional RL, value functions or policies are typically represented using tabular methods or simple function approximators. However, these approaches become impractical in large or continuous state spaces. Deep Learning addresses this limitation by using neural networks as function approximators. These networks can represent value functions (Q-values), policies, or both, allowing RL algorithms to generalize across vast and complex state spaces [13].

2.2 | Key Algorithms in Deep Reinforcement Learning

Q-Learning is a value-based RL algorithm that aims to find the optimal action-selection policy by learning the Q-value function, which estimates the expected cumulative reward of taking an action a in a state s and following the optimal policy thereafter. The Q-Learning update rule is given as follows:

$$Q(s, a) \leftarrow Q(s, a) + \alpha \left[r + \gamma \max_{a'} Q(s', a') - Q(s, a) \right] \quad (1)$$

Where $Q(s, a)$ is the current estimate of the Q-value for state s and action a . α is the learning rate. r is the reward received after taking action a in state s . γ is the discount factor. s' is the next state after taking action a . $\max_{a'} Q(s', a')$ is the maximum Q-value for the next state? s' [14].

2.2.1 | Deep Q-Networks (DQN)

Deep Q-Networks extend Q-Learning by using deep neural networks to approximate the Q-value function, allowing the algorithm to handle high-dimensional state spaces such as images. The DQN update rule is given as follows:

$$\theta \leftarrow \theta + \alpha \nabla_{\theta} \left[r + \gamma \max_{a'} Q(s', a'; \theta^-) - Q(s, a; \theta) \right]^2 \quad (2)$$

Where θ are the parameters of the Q-network. θ^- are the parameters of the target network (a copy of θ updated periodically). The Q-values $Q(s, a; \theta)$ and $Q(s', a'; \theta^-)$ re-approximated by neural networks. Key components in DQN. First Experience Replay, which stores the agent's experiences (s, a, r, s') in a replay buffer and samples mini-batches during training to break correlation. Second, the Target Network, which is a copy of the Q-network that is updated less frequently to stabilize training [15].

2.2.2 | Policy Gradients

Policy Gradient methods directly optimize the policy $\pi_{\theta}(a_t | s_t)$ by regulating the parameters θ in the direction that maximizes expected rewards. The REINFORCE mechanism can be expressed as:

$$\begin{aligned} \theta &\leftarrow \theta + \alpha \nabla_{\theta} J(\theta) \\ J(\theta) &= \mathbb{E}_{\pi_{\theta}} \left[\sum_{t=0}^T r_t \right] \end{aligned} \quad (3)$$

Where $J(\theta)$ is the expected return (cumulative reward). The gradient of $J(\theta)$ with respect to θ is given by:

$$\nabla_{\theta} J(\theta) = \mathbb{E}_{\pi_{\theta}} \left[\sum_{t=0}^T \nabla_{\theta} \log \pi_{\theta}(a_t | s_t) G_t \right] \quad (4)$$

Where $G_t = \sum_{k=t}^T \gamma^{k-t} r_k$ is the return from time step t .

2.2.3 | Actor-Critic Methods

Actor-critic methods combine value-based and policy-based approaches. An actor network updates the policy directly, while a critic network evaluates the action by estimating the value function. In mathematical terms, the Actor Update Rule is expressed as follows:

$$\theta \leftarrow \theta + \alpha \nabla_{\theta} \log \pi_{\theta}(a_t | s_t) A(s_t, a_t) \quad (5)$$

While Critic Update Rule is expressed as follows:

$$w \leftarrow w + \beta \nabla_w (r_t + \gamma V(s_{t+1}; w) - V(s_t; w))^2 \quad (6)$$

Where $\pi_{\theta}(a | s)$ is the policy parameterized by θ . $V(s; w)$ is the value function parameterized by w . $A(s_t, a_t) = r_t + \gamma V(s_{t+1}; w) - V(s_t; w)$ is the advantage function, representing how much better action a_t is compared to the average action at state s_t .

3 | Applications

3.1 | Threat Detection and Mitigation

One of the primary applications of DRL in cybersecurity is in threat detection and mitigation. Traditional rule-based systems often fail to detect new and sophisticated attacks due to their static nature. DRL, however, offers a dynamic approach to threat detection by continuously learning and adapting to emerging threats.

Threat Detection: DRL algorithms can be trained to identify anomalies in network traffic and user behavior. Through modeling the normal behavior of systems, DRL agents can detect deviations that may indicate potential security breaches. For instance, DRL can be used to monitor network packets and flag unusual patterns that signify the presence of malware or other malicious activities. The deep learning component enables the agent to understand complex patterns in large datasets, enhancing the accuracy of threat detection [16]. Once a threat is detected, DRL can be employed to automatically implement mitigation strategies. This includes isolating affected systems, blocking malicious IP addresses, and deploying patches. The ability of DRL to learn optimal actions through trial and error allows it to develop effective responses to various types of cyber attacks. For example, DRL can optimize the sequence of actions needed to contain a ransomware attack, minimizing the damage and recovery time.

3.2 | Intrusion Detection Systems

Intrusion Detection Systems (IDS) are crucial for monitoring and analyzing network traffic to detect unauthorized access and potential security threats. DRL has been increasingly applied to enhance IDS capabilities by improving detection rates and reducing false positives.

Network-Based IDS: DRL can be applied to network-based IDS to analyze traffic patterns and identify intrusions. Through training on large datasets of network traffic, DRL models can differentiate between normal and malicious behavior. The continuous learning capability of DRL ensures that the IDS can adapt to new types of attacks that were not previously encountered during the training phase [17].

Host-Based IDS: For host-based IDS, DRL can monitor system calls, file accesses, and other activities on individual devices. Through learning the typical behavior of applications and users, DRL can detect anomalies that may indicate a security breach. For example, if a normally benign application suddenly starts accessing sensitive files or making unusual network connections, the DRL-based IDS can flag this behavior for further investigation.

3.3 | Malware Analysis and Classification

Malware analysis is another critical area where DRL has shown significant promise. Traditional malware detection methods often rely on signature-based detection, which is ineffective against new, unknown malware variants. DRL offers a more flexible and adaptive approach.

Dynamic Malware Analysis: DRL can be used to automate the dynamic analysis of malware by interacting with potentially malicious software in a controlled environment. Through observing the behavior of the malware, the DRL agent can classify it based on its actions. This approach allows for the detection of previously unseen malware variants that do not match any known signatures.

Behavioral Analysis: DRL can also analyze the behavior of software to detect malicious activities. Through monitoring the actions taken by software, such as file modifications, network connections, and registry changes, DRL can identify patterns indicative of malware. This method is particularly effective against sophisticated malware that employs obfuscation techniques to evade traditional detection methods.

3.4 | Network Traffic Analysis and Anomaly Detection

Network traffic analysis is essential for maintaining the security and performance of networks. DRL has been applied to this area to identify anomalies that could indicate security threats or performance issues.

Real-Time Traffic Monitoring: DRL can be used to monitor network traffic in real-time, identifying unusual patterns that may signify an ongoing attack. Through continuous learning from network traffic data, DRL agents can detect anomalies such as Distributed Denial of Service (DDoS) attacks, data exfiltration, and unauthorized access attempts [18].

Anomaly Detection: DRL-based systems can be trained to recognize normal network behavior and flag deviations. This approach is particularly useful for detecting zero-day attacks, where no prior signature or pattern is available.

3.5 | Automated Incident Response

Automated incident response is a critical application of DRL in cybersecurity, aimed at reducing the time and effort required to respond to security incidents.

Incident Triage: DRL can be used to automate the triage process by analyzing alerts and determining their severity and priority. This helps security teams focus on the most critical threats, improving response times and reducing the workload on human analysts.

Remediation Actions: Once an incident is identified, DRL can automate the remediation process. This includes actions such as isolating infected systems, rolling back malicious changes, and applying security patches [19]. Through learning the most effective response strategies, DRL can minimize the impact of security incidents and speed up recovery.

4 | Challenges

In this section, we will explore the technical, operational, and ethical challenges that currently hinder the widespread adoption of DRL in cybersecurity.

4.1 | Technical Challenges

Data Scarcity and Quality: One of the primary technical challenges in applying DRL to cybersecurity is the scarcity and quality of labeled training data. High-quality, labeled datasets are essential for training effective DRL models, but such data can be difficult to obtain due to privacy concerns and the sensitive nature of cybersecurity incidents. Furthermore, the data that is available may be imbalanced or incomplete, which can negatively impact the performance and reliability of DRL models [15].

Computational Complexity: DRL algorithms, particularly those involving deep learning, require substantial computational resources for training and deployment. The high dimensionality of state and action spaces in cybersecurity tasks often leads to extensive training times and the need for powerful hardware, such as GPUs or TPUs. This computational demand can be a significant barrier for organizations with limited resources, hindering the practical implementation of DRL solutions.

Scalability Issues: Scaling DRL models to handle real-world cybersecurity environments presents significant challenges. These environments are typically dynamic and complex, with a vast number of possible states and actions. Ensuring that DRL models can scale effectively to manage large networks and diverse security scenarios without degrading performance is a critical concern. Additionally, real-time processing requirements necessitate efficient algorithms that can make quick decisions without compromising accuracy [13].

4.2 | Adversarial Attacks on DRL Systems

Vulnerability to Adversarial Attacks: DRL models, like other machine learning systems, are susceptible to adversarial attacks where malicious actors manipulate inputs to deceive the model. In cybersecurity, adversaries can exploit these vulnerabilities to evade detection or mislead the DRL agent into making suboptimal decisions. Adversarial attacks pose a significant threat to the reliability and trustworthiness of DRL-based security systems.

Robustness and Resilience: Ensuring the robustness and resilience of DRL models against adversarial manipulation is an ongoing research challenge. Developing methods to detect and mitigate the impact of adversarial attacks is crucial for maintaining the integrity of DRL-based cybersecurity solutions. Techniques such as adversarial training, which involves training models on adversarial examples, and incorporating robust optimization strategies, are potential approaches to enhance the resilience of DRL systems [12].

4.3 | Ethical and Privacy Concerns

Privacy Issues: The use of DRL in cybersecurity often involves processing sensitive and personal data. Ensuring the privacy and security of this data during training and deployment is paramount. There are concerns regarding how data is collected, stored, and used, particularly in light of stringent data protection regulations such as the General Data Protection Regulation (GDPR). Balancing the need for comprehensive data to train effective DRL models with the imperative to protect individual privacy is a complex issue.

Bias and Fairness: DRL models can inadvertently learn and propagate biases present in the training data. In cybersecurity, biased models may lead to unfair treatment of certain users or the overlooking of specific types of threats. Addressing issues of bias and ensuring fairness in DRL systems is essential to developing equitable and effective cybersecurity solutions. This requires careful consideration of training data and the implementation of fairness-aware learning algorithms [10].

4.4 | Scalability and Real-Time Processing

Real-Time Decision Making: Cybersecurity threats often require immediate responses to mitigate potential damage. DRL models must be capable of making real-time decisions, which can be challenging given the complexity and scale of cybersecurity environments. Ensuring that DRL systems can process data and execute actions quickly enough to be effective in real-time scenarios is a significant hurdle.

Deployment and Integration: Integrating DRL models into existing cybersecurity infrastructures poses practical challenges. Ensuring compatibility with legacy systems, maintaining operational efficiency, and minimizing disruptions during deployment are critical considerations. Additionally, continuous monitoring and updating of DRL models are necessary to adapt to evolving threats, requiring robust maintenance and support mechanisms [15].

5 | Conclusions

In this paper, we provide a concise, but, contemporary review of DRL in the realm of cybersecurity aiming to reveal a landscape rich with potential for transformative innovation. Although challenged by things like data scarcity, computational complexity, and adversarial attacks, DRL can be seen to have the potential for a variety of uses that include threat detection, intrusion prevention, and incident response. To address these challenges therefore calls for joint efforts from scientists, operators, and policymakers to create resilient and ethical DRL-based cybersecurity solutions. As we push for more advanced types of DRL methodologies, the integration of DRL into cybersecurity frameworks might offer a way of making security systems better able to address changing cyber threats.

Acknowledgments

The author is grateful to the editorial and reviewers, as well as the correspondent author, who offered assistance in the form of advice, assessment, and checking during the study period.

Funding

This research has no funding source.

Data Availability

The datasets generated during and/or analyzed during the current study are not publicly available due to the privacy-preserving nature of the data but are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that there is no conflict of interest in the research.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

- [1] Nguyen, T. T., & Reddi, V. J. (2021). Deep reinforcement learning for cyber security. *IEEE Transactions on Neural Networks and Learning Systems*, 34(8), 3779–3795.
- [2] Sewak, M., Sahay, S. K., & Rathore, H. (2023). Deep reinforcement learning in the advanced cybersecurity threat detection and protection. *Information Systems Frontiers*, 25(2), 589–611.
- [3] Oh, S. H., Kim, J., Nah, J. H., & Park, J. (2024). Employing Deep Reinforcement Learning to Cyber-Attack Simulation for Enhancing Cybersecurity. *Electronics*, 13(3), 555.
- [4] Adawadkar, A. M. K., & Kulkarni, N. (2022). Cyber-security and reinforcement learning—A brief survey. *Engineering Applications of Artificial Intelligence*, 114, 105116.
- [5] Sewak, M., Sahay, S. K., & Rathore, H. (2021). Deep reinforcement learning for cybersecurity threat detection and protection: A review. *International Conference On Secure Knowledge Management In Artificial Intelligence Era*, 51–72.
- [6] Oh, S. H., Jeong, M. K., Kim, H. C., & Park, J. (2023). Applying Reinforcement Learning for Enhanced Cybersecurity against Adversarial Simulation. *Sensors*, 23(6), 3000.
- [7] Fard, N. E., Selmic, R. R., & Khorasani, K. (2023). A Review of Techniques and Policies on Cybersecurity Using Artificial Intelligence and Reinforcement Learning Algorithms. *IEEE Technology and Society Magazine*, 42(3), 57–68.
- [8] Mesadieu, F., Torre, D., & Chennameneni, A. (2024). Leveraging Deep Reinforcement Learning Technique for Intrusion Detection in SCADA Infrastructure. *IEEE Access*.
- [9] Cengiz, E., & Gök, M. (2023). Reinforcement learning applications in cyber security: A review. *Sakarya University Journal of Science*, 27(2), 481–503.

- [10] Maddireddy, B. R., & Maddireddy, B. R. (2024). The Role of Reinforcement Learning in Dynamic Cyber Defense Strategies. *International Journal of Advanced Engineering Technologies and Innovations*, 1(2), 267–292.
- [11] Bailey, T., Johnson, J., & Levin, D. (2021). Deep reinforcement learning for online distribution power system cybersecurity protection. *2021 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, 227–232.
- [12] Shashkov, A., Hemberg, E., Tulla, M., & O'Reilly, U.-M. (2023). Adversarial agent-learning for cybersecurity: a comparison of algorithms. *The Knowledge Engineering Review*, 38, e3.
- [13] Zhao, J., Hu, F., & Hei, X. (2023). Defensive Schemes for Cyber Security of Deep Reinforcement Learning. In *AI, Machine Learning and Deep Learning* (pp. 139–149). CRC Press.
- [14] Alturkistani, H., & El-Affendi, M. A. (2022). Optimizing cybersecurity incident response decisions using deep reinforcement learning. *International Journal of Electrical and Computer Engineering*, 12(6), 6768.
- [15] Liu, X., Ospina, J., & Konstantinou, C. (2020). Deep reinforcement learning for cybersecurity assessment of wind integrated power systems. *IEEE Access*, 8, 208378–208394.
- [16] Selim, A., Zhao, J., Ding, F., Miao, F., & Park, S.-Y. (2023). Adaptive Deep Reinforcement Learning Algorithm for Distribution System Cyber Attack Defense With High Penetration of DERs. *IEEE Transactions on Smart Grid*.
- [17] Tareq, I., Elbagoury, B. M., El-Regaily, S. A., & El-Horbaty, E.-S. M. (2024). Deep Reinforcement Learning Approach for Cyberattack Detection. *International Journal of Online & Biomedical Engineering*, 20(5).
- [18] Abid, M. N., Beggas, M., & Laouid, A. (2024). Reinforcement Learning Approach for IoT Security using CyberBattleSim: A Simulation-based Study. *2024 6th International Conference on Pattern Analysis and Intelligent Systems (PAIS)*, 1–7.
- [19] Kabanda, G., CHIPFUMBU, C. T., & Chingoriwo, T. (2023). A Reinforcement Learning Paradigm for Cybersecurity Education and Training. *Oriental Journal of Computer Science and Technology*, 12–45.

Disclaimer/Publisher's Note: The perspectives, opinions, and data shared in all publications are the sole responsibility of the individual authors and contributors, and do not necessarily reflect the views of Sciences Force or the editorial team. Sciences Force and the editorial team disclaim any liability for potential harm to individuals or property resulting from the ideas, methods, instructions, or products referenced in the content.