Vol.4, No.4, Page 2025- 2033, 2026 https://doi.org/10.59628/jast.v4i4.2574 ISSN 2958-9568 JAST

SUN PUBLISHER

مجلة جامعة صنعاء للعلوم التطبيقية والتكنولوجيا

Sana'a University Journal of Applied Sciences and Technology

https://journals.su.edu.ye/index.php/jast/

Smart cybersecurity strategies based on deep reinforcement learning : A Literature Review

Aleah Abdulkaher Alshamiri * **and Ghaleb H. Al Gaphari**

Department of Computer Science, Faculty of Computer and Information Technology, Sana'a University, Sana'a, Yemen

*Corresponding author: aleah.alshamiri@gmail.com

Abstract

This paper presents a literature review on the application of deep reinforcement learning (DRL) in cybersecurity. It focuses on key domains, including intrusion detection, adaptive cyber defense, multi-agent coordination, and automated penetration testing. A total of 18 peer-reviewed studies published between 2022 and 2025 were selected through a structured review process and analyzed using performance metrics such as accuracy, precision, recall, and F1-score.The results indicate that multi-agent DRL approaches generally outperform single-agent models in dynamic attack environments. Hybrid DRL models that integrate deep learning techniques, such as convolutional and recurrent neural networks and attention mechanisms, show improved detection accuracy and adaptability. DRL-based penetration testing methods also demonstrate the ability to autonomously explore vulnerabilities and optimize attack strategies. However, challenges remain, including limited generalization to real-world scenarios, high computational costs, low interpretability, and the lack of standardized datasets. Addressing these issues can enable the development of more adaptive, efficient, and reliable cybersecurity systems.

Article Info

Keywords:
Deep Reinforcement Learning, Cybersecurity, Threat Detection, Automated Penetration Testing, Explainable Artificial Intelligence, Multi-Agent Systems.

Article History:
Received: 27-November-2025,
Revised: 31-January-2026,
Accepted: 14-February-2026,
Published: 28 April 2026.

1. INTRODUCTION

The rapid growth of interconnected systems, such as industrial control networks, intelligent infrastructure, and cloud services, has increased cybersecurity complexity. These systems face sophisticated cyber threats, including multistage intrusions, advanced persistent threats (APTs), and coordinated attacks. Traditional rule-based security mechanisms are insufficient because modern attacks are adaptive, stealthy, and capable of bypassing static detection rules [1–3].

To address these limitations, the cybersecurity community has shifted toward intelligent, automated, and adaptive defense mechanisms capable of operating in dynamic, rapidly changing environments [4, 5]. Building on this, learning-based approaches that train systems on data and experience enable adaptation to observed interactions rather than relying solely on predefined rules. Among these approaches, deep reinforcement learning (DRL), which combines deep learning (using neural networks to learn representations) and reinforcement learning (learning through trial and error with feedback), has emerged as a promising paradigm for enhancing cybersecurity.

DRL combines the representational power of deep neural networks with the sequential decision-making framework of reinforcement learning, enabling agents to learn optimal defensive or offensive strategies directly through interactions with their environment [1, 2, 6]. This makes DRL particularly well-suited to cybersecurity tasks that require continuous adaptation to evolving attack behaviors, which are commonly observed in real-world cyberattack scenarios [7, 8]. Consequently, DRL has been increasingly applied to a wide range of cybersecurity applications, including intrusion and anomaly detection, automated penetration testing, cyberattack simulation, and adaptive defense strategies [6, 7, 9, 10].

Recent studies have demonstrated that DRL-based systems can proactively detect stealthy cyberattacks, dynamically adjust defense policies, and automate complex penetration testing processes [3]. In addition, multi-

©2026 The Author(s). This article is licensed under CC By 4.0

agent DRL frameworks (systems where several agents use DRL to interact and coordinate) enable coordinated decision-making among multiple learning agents in complex cyber environments [3]. Among various DRL algorithms, deep Q-networks (DQNs, an algorithm that uses deep networks to select actions in discrete settings) remain the most widely adopted owing to their effectiveness in discrete action spaces and suitability for sequential cyber decision-making tasks [2, 3, 7, 8]. Other algorithms, such as the advantage actor–critic (A2C, a model combining value-based and policy-based learning), proximal policy optimization (PPO, an RL algorithm that balances stability and efficiency), and deep deterministic policy gradient (DDPG, for continuous-action environments), are more useful for continuous-action environments. The performance of these methods is further enhanced when combined with auxiliary techniques, including attention mechanisms (systems that focus on relevant input parts) and knowledge-driven reward design (using domain expertise to shape reward signals).

Despite these advances, several challenges remain. DRL models often struggle to generalize to unseen or real-world environments [2, 11], require extensive computing resources during training, and offer limited interpretability of learned policies [2]. These limitations highlight the need for generalizable, transparent, and domain-aware DRL-based cybersecurity solutions.

In response to these challenges, this study reviews and analyzes recent literature on DRL in cybersecurity and addresses the following research question:

RQ1: To what extent can deep reinforcement learning methods enhance cybersecurity performance across key application domains, such as intrusion detection, automated penetration testing, and multi-agent defense coordination?

To answer this question, this study examines recent DRL-based cybersecurity research, compares algorithmic approaches and reported performance outcomes, and identifies open research challenges that may inform the development of more scalable, efficient, and intelligent cyber-defense strategies. The remainder of this paper is organized as follows: Section 2 describes the study selection methodology; Section 3 reviews related work; Section 4 presents a comparative analysis; Section 5 discusses challenges and future research directions; and Section 6 concludes the paper.

2. METHODOLOGY

This study uses a structured literature review to emphasize transparency and reproducibility in analyzing recent developments in deep reinforcement learning DRL based cybersecurity solutions. The review aims to structure, compare, and critique emerging research trends, techniques, and challenges in DRL for cybersecurity, with an analysis focus rather than an exhaustive review or meta-analysis.

The review focuses on identifying dominant methodological paradigms (standard approaches and methods), performance trends, and open research challenges across four core application areas: intrusion detection systems (which identify unauthorized access), adaptive cyber defense (which uses AI to adjust protections in real time), automated penetration testing (which simulates cyberattacks to assess system vulnerabilities), and multi-agent reinforcement learning–based cybersecurity frameworks (where multiple AI agents collaborate or compete to enhance security).

2.1. Review Design and Scope

This review includes peer-reviewed studies from 2018 to 2025 and focuses on works since 2022 to capture recent developments in DRL-based cyber defense.

This review prioritizes methodological diversity, technical rigor, and practical relevance in adaptive and intelligent cybersecurity systems.

Accordingly, rather than aiming for exhaustive, protocol-driven coverage of all studies, this review provides a focused comparative analysis of representative, methodologically relevant works.

2.2. Data Sources and Search Strategy

Relevant studies identified through targeted searches conducted across the following major scientific databases:

- IEEE Xplore
- SpringerLink
- ScienceDirect
- Wiley Online Library
- Google Scholar

Search queries were constructed using combinations of the following keywords:

- Deep Reinforcement Learning AND *Cybersecurity*
- *Intrusion Detection Systems*
- *Adaptive Cyber Defense*
- *Automated Penetration Testing*
- *Multi-Agent Reinforcement Learning*

The search strategy targeted titles, abstracts, and keywords to ensure that selected studies explicitly addressed DRL-based cybersecurity applications.

2.3. Inclusion and Exclusion Criteria

To ensure consistency and relevance, explicit inclusion and exclusion criteria were applied during the study selection process.

©2026 The Author(s). This article is licensed under CC By 4.0

Inclusion Criteria

- Peer-reviewed journal articles and conference papers
- Studies applying DRL or MARL techniques to cybersecurity problems
- Research on intrusion detection, cyber defense, penetration testing, and attack–defense modeling

Exclusion Criteria

- Non-peer-reviewed articles, tutorials, surveys, or opinion papers
- Studies focusing exclusively on classical machine learning without reinforcement learning components
- Works lacking experimental evaluation or empirical validation
- Papers unrelated to cybersecurity or defensive applications

After applying these criteria, 18 primary studies were selected for detailed analysis and comparison. Table 1 *outlines* the number of records retained at each stage of the study selection process, from initial identification to final inclusion.

Table 1. Records at Each Stage of the Study Selection Process

Stage	Description	Number
N1	Initial records identified	582
..	From ScienceDirect	167
..	From Google Scholar	50
..	From IEEE Xplore	147
..	From ACM Digital Library	100
..	From SpringerLink	118
Ns	Duplicates removed	218
	Records after duplicates removed	364
N2	Screened (title and abstract)	364
N6	Excluded at screening	264
N3	Full-text articles assessed	100
	Excluded after full-text review	60
N4	Final included primary studies	18

2.4. Quality Assessment of Selected Studies

Quality assessment ensured the rigor and relevance of the selected studies. This step verified that all works met basic scientific quality and relevance to DRL-based cybersecurity, not ranked studies.

Each study was evaluated against five predefined quality criteria: topic relevance, methodological soundness, evaluation rigor, environmental transparency, and practical relevance. A clear scoring system minimized assessment heterogeneity. Only studies that met the quality standards were analyzed in detail, resulting in the selection of 18 primary studies.

Table 2 outlines the quality assessment criteria used to assess the methodological strength and relevance of the included studies.

Table 2. Quality Assessment Criteria for Included Studies

Criterion Code	Quality Criterion	Description
Q1	Relevance to DRL-based Cybersecurity	The study explicitly applies deep reinforcement learning (or MARL) to cybersecurity-related problems such as intrusion detection, cyber defense, or penetration testing.
Q2	Methodological Soundness	The study clearly describes the DRL model, learning algorithm, reward design, and experimental setup.
Q3	Evaluation and Performance Metrics	The study reports quantitative evaluation metrics (e.g., accuracy, precision, recall, F1-score, response time).
Q4	Experimental Environment and Dataset Clarity	The dataset or simulation environment is clearly specified and appropriate for the cybersecurity task.
Q5	Practical Relevance and Insight	The study provides insights into applicability, limitations, or real-world relevance of the proposed approach.

2.5. Study Classification and Analysis Framework

The selected studies were categorized into four main thematic groups:

1 DRL-based Intrusion Detection Systems
2 Adaptive Cyber Defense and Attack–Defense Modeling
3 Automated Penetration Testing using DRL
4 Multi-Agent and Hybrid DRL-Based Cybersecurity Models

Within each category, studies were analyzed based on:

- Reinforcement learning algorithms employed
- Application domain and threat model
- Dataset or simulation environment
- Performance metrics used for evaluation
- Identified strengths and limitations

This classification is grounded in three methodological criteria:

(i) the primary cybersecurity objective addressed by each study,
(ii) the functional role of DRL within the system,
(iii) the underlying system architecture (single-agent, multi-agent, or hybrid).

This framework is designed to facilitate a structured com-

©2026 The Author(s). This article is licensed under CC By 4.0

parison among the studies and to highlight key methodological differences, making it easier to understand the purpose and significance of each thematic group.

Figure 1 illustrates the conceptual classification framework adopted in this review, highlighting the main dimensions used to analyze DRL-based cybersecurity systems, including learning paradigms, application domains, DRL algorithms, evaluation environments, and explainability aspects

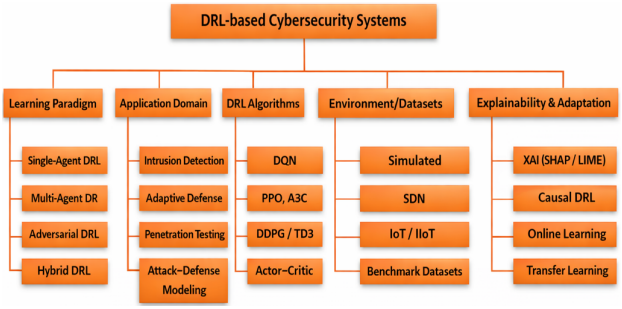


Figure 1. Hierarchical taxonomy of DRL-based cybersecurity systems across learning paradigms, application domains, algorithms, environments, and explainability mechanisms.

2.6. Comparative Evaluation Strategy

A comparative evaluation was conducted using commonly reported performance metrics, including accuracy, precision, recall, and the F1-score, as presented in the original studies. These metrics were employed to support a qualitative comparison of performance trends across different DRL-based approaches. Given that the reviewed studies used heterogeneous datasets, threat models, and experimental environments, performance comparisons are interpreted qualitatively rather than as direct benchmarks. Contextual factors, such as attack diversity, dataset imbalance, and online versus offline evaluation settings, received special attention. Accordingly, the reported performance metrics should be interpreted as indicative trends rather than definitive evidence of algorithmic superiority.

3. RELATED WORK

The cybersecurity landscape has evolved rapidly in recent years because of the increasing sophistication, diversity, and persistence of modern cyber threats. Traditional security mechanisms, which often rely on static rules or predefined signatures, have become insufficient to address complex attack behaviors that continuously adapt to defensive measures. In this context, deep reinforcement learning (DRL) has emerged as a prominent paradigm for designing intelligent cybersecurity systems that learn directly from their environments and adapt their actions in real time with minimal human intervention. Consequently, DRL has been widely applied to various cybersecurity tasks, including threat detection, adaptive defense, automated penetration testing, and multi-agent coordination. This section reviews representative studies that have shaped the evolution of DRL-based cybersecurity research and critically examines their methodological contributions and limitations.

3.1. Generalization Challenges and Environmental Constraints

Generalization to unseen environments remains one of the most critical challenges for DRL-based cybersecurity systems. Recent studies emphasize that DRL agents trained in limited or homogeneous environments often fail to perform effectively when exposed to novel attack scenarios [11]. One notable study demonstrated that the Advantage Actor–Critic (A2C) algorithm outperformed DQN and PPO across diverse vulnerability assessment and penetration testing (VAPT) environments, highlighting the influence of environmental bias on learning outcomes. The authors further recommended using diverse training environments, transfer learning, and causal modeling to improve generalization. These findings are reinforced by a complementary study [12], which incorporated causal modeling into PPO through a causal reward mechanism, resulting in a 40% increase in win rate and an 18% improvement in network isolation. Together, these studies suggest that generalization and interpretability are closely related objectives rather than independent design goals.

3.2. Security Posture Enhancement and Internet of Things (IoT) Applications

The application of DRL in IoT and cyber–physical systems has gained increasing attention owing to the large attack surface and resource constraints inherent in such environments. One study [13] proposed a DRL-based intrusion detection framework for IoT systems that integrates data collection, feature extraction using the central moment momentum (CMM) algorithm, and network packet classification, achieving 98.82% accuracy on the CIC-IDS2018 dataset. In contrast, another study [1] introduced an adaptive defense framework based on DQN, PPO, and TD3 to dynamically execute security actions, such as blocking suspicious network entities, and reported an overall accuracy of 92%. Similarly, the Deep IDPS model [8] combined a CNN, long short-term memory (LSTM), and attention mechanisms to enhance intrusion detection and prevention in SDN environments, significantly reducing false-alarm rates. These studies demonstrate the effectiveness of hybrid DRL architectures in improving detection accuracy and operational robustness in IoT and SDN settings.

©2026 The Author(s). This article is licensed under CC By 4.0

3.3. Multi-Agent Cooperation for Defensive and Offensive Decision-Making

Multi-agent reinforcement learning (MARL) has been increasingly explored to address the distributed and cooperative nature of cybersecurity operations. A hierarchical MARL (HMARL) framework [13] demonstrated improved defensive coordination by organizing interactions between strategic and operational agents, thereby reducing attack success rates and leading to the emergence of stable defense equilibria. Similarly, the DMARL-based approach [14] assigned defensive responsibilities to autonomous agents within a CIRA-CPS framework, achieving higher detection accuracy, fewer false positives, and improved adaptability in real-time environments. These findings highlight the potential of MARL to enhance scalability and coordination in complex cyber-defense scenarios. It should be noted that the reported performance advantages of MARL are often obtained under simplified or fully observable simulation settings and may not directly translate to large-scale, partially observable, or resource-constrained real-world networks.

3.4. Performance Enhancement through Domain Knowledge and Supporting Techniques

Several studies have attempted to improve DRL performance by integrating domain-specific security knowledge and complementary analytical techniques. For instance, the KI-APT framework [7] leverages CVE and CPE vulnerability information and employs PPO within the cyborg simulation environment, resulting in higher penetration success rates and reduced redundant attack steps. Similarly, another study [15] investigated an IMO–DDPG-based intrusion detection approach that incorporated Z-score normalization and principal component analysis (PCA), achieving an accuracy of 94.5%. Furthermore, combining DQN with isolation forest was shown to improve anomaly detection performance in smart grid environments [14]. Collectively, these approaches demonstrate that incorporating prior security knowledge and feature engineering techniques can significantly enhance DRL effectiveness while reducing computational overhead.

3.5. Reinforcement Learning–Driven Defensive and Offensive Strategies

Deep reinforcement learning (DRL) has increasingly been employed to support both defensive and offensive cybersecurity strategies, particularly in automated attack simulation and penetration testing. In this context, an offensive agent based on the deep deterministic policy gradient (DDPG) was evaluated in dynamic environments in [4], achieving an average performance of approximately 85% and outperforming classical Q-learning approaches under the same conditions.

Similarly, the dynamic penetration (DynPen) framework proposed in [9] leverages proximal policy optimization (PPO) for automated penetration testing within the Cyber Operations Research Gym (Cyborg) environment. The reported results indicate improved effectiveness compared to the advantage actor–critic (A2C) and deep Q-network (DQN) baseline methods in the same simulation setting.

Additional studies [5, 8] have investigated the use of DQN to identify vulnerable attack paths and enable rapid responses to evolving threats. Although these approaches demonstrate the feasibility of DRL-based penetration testing and attack simulation, their performance is typically evaluated in controlled or simulated environments, which may limit their direct generalization to real-world operational networks. Overall, these studies reflect the growing technical maturity of DRL-driven offensive security research and highlight the need to carefully consider deployment constraints and evaluation realism.

3.6. Methodological Insights from Existing Reviews

Several recent review studies [2, 3, 16] provide structured and in-depth examinations of DRL applications in cybersecurity. They spotlight three main categories: threat detection, active defense, and penetration testing. These reviews repeatedly expose critical challenges, including the scarcity of labeled datasets, the interpretability gap in learned policies, and weak generalization across environments. They urge the adoption of realistic simulation platforms and hybrid architectures to strike the right balance between detection accuracy, explainability, and operational efficiency. Together, these insights provide a clear picture of current DRL-based cybersecurity limitations.

3.7. Emerging Research Directions

By synthesizing the reviewed literature, three major research directions emerge.

First, improving generalization and interpretability through causal modeling, multi-scenario training environments, and transfer learning has emerged as a priority [9, 11, 12].

Second, enhancing real-time adaptation and multi-agent cooperation is critical for effective defense and offense in dynamic cyber environments [1, 6, 9, 13].

Third, integrating complementary techniques, such as formal security knowledge, anomaly detection methods, and advanced feature engineering, can help achieve an effective trade-off among accuracy, response time, and

©2026 The Author(s). This article is licensed under CC By 4.0

computational cost [4, 10, 14, 15].

Overall, existing studies indicate that DRL-based cybersecurity systems are evolving toward more autonomous, generalizable, and interpretable solutions. Addressing adversarial robustness and reducing computational complexity remain key challenges and promising avenues for future research.

4. RESULTS AND ANALYSIS

To support a comparative analysis of the selected studies, we adopted a structured analytical framework that considers the application domain, learning paradigm, algorithmic design, experimental environment, and reported performance metrics. Quantitative metrics, such as accuracy, precision, recall, and F1-score, provide useful insights; however, they must be interpreted in light of the diverse experimental environments across the reviewed works.

A prominent observation is the frequent adoption of DQN-based approaches, especially in intrusion detection and adaptive defense. This prevalence is mainly due to DQN's architectural simplicity, ease of implementation, and suitability for discrete action spaces. The apparent dominance of DQN-based approaches in Table 3 reflects publication and accessibility bias in the existing literature rather than clear evidence of intrinsic algorithmic superiority. However, several studies implicitly indicate limitations, including sensitivity to reward design, sample inefficiency, and degraded performance when transferring policies across environments. These limitations partly explain the increasing use of policy gradient methods, such as PPO and DDPG, in penetration testing and attack–defense modeling, where continuous action spaces and smoother policy updates are required.

The comparative results also indicate that multi-agent reinforcement learning (MARL) often outperforms single-agent DRL models in terms of accuracy and F1-score. This gain is largely attributed to MARL's ability to decompose complex defense tasks, enable cooperative decision-making, and improve the coverage of large and dynamic state spaces. Nevertheless, this reported superiority should be interpreted cautiously. Most MARL-based studies are evaluated in controlled or simulated environments and often do not account for practical challenges, such as communication overhead, coordination complexity, and scalability constraints. These factors may limit real-world deployments.

Hybrid DRL models that integrate attention mechanisms, causal reasoning, or domain knowledge (e.g., CVE/CPE repositories) exhibit improved robustness and interpretability compared to purely model-free approaches. These models benefit from incorporating structured prior knowledge, which can mitigate overfitting and enhance decision consistency in dynamic attack scenarios. However, hybridization typically increases computational complexity and training costs, raising concerns about deployment feasibility in resource-constrained environments.

Overall, while the quantitative comparison suggests that MARL and hybrid DRL architectures achieve superior performance metrics, these gains remain highly dependent on the experimental context, dataset characteristics, and evaluation protocols. The lack of standardized benchmarks and heavy reliance on simulated environments further constrain direct comparability. These findings underscore the need for future research to focus not only on performance optimization but also on the generalization, explainability, computational efficiency, and real-world applicability of DRL-based cybersecurity solutions.

Accordingly, the following comparative summary should be interpreted with caution. Reported performance values are reproduced for contextual reference only and should not be interpreted as directly comparable benchmarks. Table 3 presents a comparative analysis of the selected studies. It highlights how algorithmic design, application context, and evaluation environments influence reported performance rather than implying absolute superiority among approaches.

5. DISCUSSION

Deep learning reinforcement (DRL) has made significant strides in cybersecurity; however, several methodological and practical obstacles prevent the widespread adoption of these methods in the real world. One of the most critical limitations is generalization, as many DRL models demonstrate strong performance in controlled training environments but struggle to maintain effectiveness when exposed to heterogeneous, dynamic, and previously unseen real-world attack scenarios. This gap between simulated training conditions and operational environments limits the practical reliability of current DRL-based solutions.

A second major challenge relates to data availability and quality. DRL models typically require large volumes of representative training data; however, in cybersecurity applications, access to diverse real-world datasets is often restricted due to privacy concerns, security policies, and the sensitive nature of attack data. Consequently, many existing studies rely heavily on synthetic or simulated datasets, which further exacerbates generalization issues. Interpretability represents another fundamental challenge. Most DRL models operate as black-box systems, making it difficult for security analysts to understand, verify, and trust the decisions they produce. In security-critical environments, where accountability and transparency are essential, the lack of explainability significantly hinders the adoption of DRL-based solutions.

In addition, the computational cost and scalability of DRL training pose substantial barriers. Modeling large-

©2026 The Author(s). This article is licensed under CC By 4.0

Table 3. Comparative summary of selected DRL-based cybersecurity studies, presenting security objectives, application domains, employed algorithms, reported performance outcomes, and key contributions.

Ref	Year	Security Objective	Application Domain	Algorithms Used	Main Results	Key Contributions
[12]	2024	Hybrid (Causal DRL)	Explainable cyber defense	PPO + causal modeling	∼40% performance improvement with causal explanations	Integration of causal reasoning with DRL decisions
[1]	2024	Adaptive Defense	Real-time adaptive cyber defense	DQN, PPO, TD3	High F1-score and robust real-time performance	Multi-layered adaptive defense model
[13]	2024	Adaptive Defense	Defense against cyber attacks	DDPG	Improved defense efficiency and reduced network losses	Self-learning offensive/defensive agent design
[15]	2025	Intrusion Detection	Intrusion detection systems	CNN, LSTM, Attention + DRL	Improved detection accuracy and reduced false alarms	Hybrid attention-based IDS architecture
[14]	2025	Intrusion Detection	Smart network security	DQN + Isolation Forest	Improved flexibility and robustness of detection	Hybrid DRL–ML security model
[4]	2025	Penetration Testing	Knowledge-based penetration testing	PPO + CVE knowledge base	Improved efficiency and reduced redundant attack steps	Integration of security knowledge with DRL for intelligent attacks
[9]	2024	Penetration Testing	Automated penetration testing	PPO	High success rate and incremental learning of attack paths	Two-level DRL framework for adaptive penetration testing
[5]	2024	Penetration Testing	Cyberattack strategy optimization	DDPG	85% success rate in target penetration	Modeling cyberattacks as DRL-based optimization problems
[8]	2025	Penetration Testing	Penetration testing automation	DQN	Accurate identification of critical attack paths in reduced time	Automated vulnerability exploitation using DRL
[16]	2023	Intrusion Detection	Intrusion detection and response	DQN, PPO, Actor–Critic	High detection accuracy with flexible policy learning	Comprehensive DRL-based IDS framework
[6]	2025	Multi-Agent / Hybrid	Multi-agent attack detection	DMARL	Increased detection accuracy and reduced false positives	Collaborative multi-agent detection framework
[10]	2025	Intrusion Detection	Intrusion detection in SDN	CNN–LSTM + DRL	Superior accuracy and reduced unnecessary control messages	Deep hybrid IDS for SDN environments
[17]	2025	Adaptive Defense	Detection and response to cyber attacks	DQN	Reduced detection time and improved response efficiency	Intelligent adaptive cyber defense framework based on DRL
[18]	2024	Adaptive Defense	Dynamic cyber defense strategies	DQN	96.5% accuracy and superior adaptability compared to classical ML	Demonstrated effectiveness of DRL under dynamic attacks

scale cyber environments, particularly those involving adversarial or multi-agent interactions, requires considerable computational resources and prolonged training times. These requirements raise concerns about the feasibility of deploying such models in real-time or resource-constrained operational settings.

Importantly, these challenges also offer several promising research opportunities. Improving generalization and adaptability through techniques such as transfer learning, domain adaptation, and continual learning represents a key direction for future work. Similarly, enhancing data efficiency via synthetic data generation, simulation-to-real (sim-to-real) transfer, and federated or privacy-preserving learning can help mitigate data scarcity while respecting privacy constraints.

The integration of explainable artificial intelligence (XAI) and causal modeling into DRL frameworks offers a viable pathway for improving transparency and analyst trust. Concrete explainability techniques, such as policy distillation, reward attribution analysis, saliency-based visualization of state–action importance, and causal graph representations, have recently emerged as promising tools for interpreting DRL policies in cybersecurity contexts. By incorporating domain knowledge from human security experts or causal reasoning mechanisms, future models can enable more interpretable and context-aware

©2026 The Author(s). This article is licensed under CC By 4.0

decision-making.

Furthermore, optimizing computational efficiency through lightweight architectures, hierarchical learning, and distributed or multi-agent coordination strategies may reduce resource consumption and support real-time deployment. The use of multimodal data sources, including network traffic, logs, alerts, and contextual metadata, also offers opportunities to improve detection accuracy and capture complex, multidimensional attack behaviors.

Finally, addressing these challenges effectively requires multidisciplinary collaboration across cybersecurity, artificial intelligence, data science, systems engineering, and the social sciences. Such collaboration can enable the development of robust, ethical, and operationally viable DRL-based cyber defense systems. Collectively, these challenges and opportunities define a rich research agenda essential for advancing the next generation of intelligent, adaptive, and trustworthy cybersecurity solutions.

6. CONCLUSION

This review provides a structured, critical synthesis of recent research on deep reinforcement learning (DRL)–based cybersecurity solutions. It highlights both methodological trends and persistent research challenges. A comparative analysis demonstrates that multi-agent DRL (MARL) approaches generally outperform single-agent and hybrid models in complex, dynamic cyber-defense scenarios. This is primarily because MARL enables cooperative decision-making and scalable defense strategies. At the same time, hybrid DRL architectures—those integrating auxiliary models such as CNNs, LSTMs, attention mechanisms, or domain knowledge from vulnerability intelligence sources—offer improved interpretability, robustness, and adaptability. This is particularly true in dynamic-threat environments. Despite these advances, the review reveals several unresolved limitations that hinder the widespread adoption of DRL-based cybersecurity systems. These include limited generalization across heterogeneous environments, high computational and training costs, insufficient interpretability, and the lack of standardized real-world benchmark datasets. Such constraints underscore the need for more rigorous evaluation frameworks and realistic experimental settings. From a research perspective, the findings of this review point toward several promising directions. These include causality-aware reinforcement learning, multimodal DRL architectures, real-time adaptive defense mechanisms, and the development of large-scale, standardized cybersecurity benchmarks. Addressing these challenges is essential for bridging the gap between experimental success and operational deployment. From a practical standpoint, this review provides a structured and practical roadmap for researchers and practitioners. It clarifies design trade-offs, identifies dominant methodological patterns, and highlights critical gaps in existing DRL-based cybersecurity research. As cyber threats continue to increase in scale and sophistication, DRL-enabled defense mechanisms are expected to play an increasingly central role in the development of intelligent, adaptive, and autonomous cybersecurity systems.

REFERENCES

[1] A. A. Hammad, S. R. Ahmed, M. K. Abdul-Hussein, M. R. Ahmed, D. A. Majeed, and S. Algburi, "Deep reinforcement learning for adaptive cyber defense in network security," in *Proceedings of ACM International Conference*, May 2024, pp. 292–297. DOI: 10.1145/3660853.3660930.

[2] W. Yang, A. Acuto, Y. Zhou, and D. Wojtczak, "A survey of deep reinforcement learning-based network intrusion detection," *arXiv preprint arXiv:2410.07612*, Sep. 2024.

[3] A. M. K. Adawadkar and N. Kulkarni, "Cyber-security and reinforcement learning: A brief survey," *Eng. Appl. Artif. Intell.*, vol. 114, 2022. DOI: 10.1016/j.engappai.2022.105116.

[4] Y. Li, H. Dai, and J. Yan, "Knowledge-informed auto-penetration testing based on reinforcement learning with reward machines," in *Proceedings of IEEE International Conference*, May 2024. DOI: 10.1109/10650368.

[5] K. Bum-Sok, H.-W. Suk, C. Yong-Hoon, M. Dae-Sung, and K. Min-Suk, "Optimal cyber attack strategy using reinforcement learning based on common vulnerability scoring system," *Comput. Model. Eng. & Sci.*, vol. 141, no. 2, p. 1551, 2024.

[6] A. Manikandan and S. D. Rajan, "Cyber attack detection using deep multi-agent reinforcement learning with beth dataset," *SN Comput. Sci.*, vol. 6, no. 5, Jun. 2025. DOI: 10.1007/s42979-025-03981-8.

[7] S. H. Oh, J. Kim, J. H. Nah, and J. Park, "Employ deep reinforcement learning for cyber attack simulation to enhance cybersecurity," *Electronics*, vol. 13, no. 3, Feb. 2024. DOI: 10.3390/electronics13030555.

[8] I. Jabr, Y. Salman, M. Shqair, and A. Hawash, "Penetration testing and attack automation simulation: A deep reinforcement learning approach," *An-Najah Univ. J. for Res. A (Natural Sci.*, vol. 39, no. 1, pp. 7–14, Feb. 2025. DOI: 10.35552/anujr.a.39.1.2231.

[9] Q. e. a. Li, "Dynpen: Automated penetration testing in dynamic network scenarios using deep reinforcement learning," *IEEE Trans. on Inf. Forensics Secur.*, vol. 19, pp. 8966–8981, 2024. DOI: 10.1109/TIFS.2024.3461950.

[10] N. Niknami and J. Wu, "Deepidps: An adaptive drl-based intrusion detection and prevention system for sdn," *IEEE Access*, 2024.

[11] A. Venturi, M. Andreolini, M. Marchetti, and M. Colajanni, "Assessing generalizability of deep reinforcement learning algorithms for automated vulnerability assessment and penetration testing," *Array*, vol. 24, Dec. 2024. DOI: 10.1016/j.array.2024.100365.

[12] T. Purves, K. G. Kyriakopoulos, S. Jenkins, I. Phillips, and T. Dudman, "Causally aware reinforcement learning agents for autonomous cyber defence," *Knowledge-Based Syst.*, vol. 304, Nov. 2024. DOI: 10.1016/j.knosys.2024.112521.

[13] Y. Tang, J. Sun, H. Wang, J. Deng, L. Tong, and W. Xu, "A method of network attack-defense game and collaborative defense decision-making based on hierarchical multi-agent reinforcement learning," *Comput. & Secur.*, vol. 142, Jul. 2024. DOI: 10.1016/j.cose.2024.103871.

[14] B. R. Maddireddy and B. R. Maddireddy, "The role of reinforcement learning in dynamic cyber defense strategies," *Int. J. Adv. Eng. Trends Innov.*, 2024.

©2026 The Author(s). This article is licensed under CC By 4.0

[15] Y. Ma, C. Li, Y. Wang, and Y. Wang, "Application of deep reinforcement learning algorithms for automatic threat detection and response in dynamic network environments," *J. Comput. Methods Sci. Eng.*, vol. 25, no. 3, pp. 2112–2125, May 2025. DOI: 10.1177/14727978241309550.

[16] T. T. Nguyen and V. J. Reddi, "Deep reinforcement learning for cyber security," *IEEE Trans. on Neural Networks Learn. Syst.*, vol. 34, no. 8, pp. 3779–3795, Aug. 2023. DOI: 10.1109/TNNLS.2021.3121870.

[17] M. M. Al-Nawashi, O. M. Al-Hazaimeh, N. M. Tahat, N. Gharaibeh, W. A. Abu-Ain, and T. Abu Ain, "Deep reinforcement learning-based framework for enhancing cybersecurity," *Int. J. Interact. Mob. Technol.*, vol. 19, no. 3, pp. 170–190, Feb. 2025. DOI: 10.3991/ijim.v19i03.50727.

[18] H. S. AlSagri, S. S. Sohail, and S. Sebastian, "The role of deep reinforcement learning in developing adaptive cybersecurity defenses for smart grid systems," *J. Inf. Optim. Sci.*, vol. 45, no. 8, pp. 2299–2307, 2024. DOI: 10.47974/JIOS-1807.

©2026 The Author(s). This article is licensed under CC By 4.0