Adaptive Reinforcement Learning Cyber Defense Strategies using CyberBattleSim

Bum-Sok Kim[1], Yonghoon Choi[2], Minsuk Kim[*]
Sangmyung University, Dept. of Electronic Information System Engineering[1],
Sangmyung University, Dept. of Human Intelligence & Robot Engineering[2*],

qjatod132@naver.com[1], choiyonghoon1103@gmail.com[2], *minsuk.kim@smu.ac.kr[*]

Abstract

This paper develops and evaluates two defender agents, CredentialScanAndReimage (CSR) and AdaptiveFirewall (AF), using CyberBattleSim, a reinforcement learning-based tool for cyber-attack-defense simulation. We designed a scenario named 'CIA Triad-based Capture the Flag (CT2F)' to enhance confidentiality, integrity, and availability, incorporating a 'Honeypot' for cyber deception. We also evaluated cyber defense strategies within the 'CT2F' scenario using Q-Learning, DQN, and A2C. Experimental results show that A2C with CSR defenders outperforms DQN and Q-Learning in cumulative rewards, though DQN generally achieves higher success rates without defenders. In particular, in the absence of defenders, DQN's maximum success rate was 60%, significantly higher than Q-Learning's and A2C's. However, DQN's maximum success rate decreased by 27% with the implementation of CSR, and by 33% with AF. A2C with CSR achieved a 21% reduction in attacker success rates. These results emphasize the potential of reinforcement-learning-based autonomous defense mechanisms in dynamic environments.

Ⅰ. Introduction

In recent years, cybersecurity has faced significant challenges due to evolving cyber-attacks, particularly Distributed Denial of Service (DDoS) attacks in the gaming industry [1]. These attacks, which overload servers using multiple sources, lead to economic losses and reputational damage. For instance, a recent e-sports event experienced a significant delay due to a DDoS attack, resulting in the league's suspension. DDoS attacks are difficult to prevent as they flood networks with traffic, and increasing server capacity is inefficient and costly. To address these issues, a Reinforcement Learning (RL)-based cyber-attack defense simulation environment has gained attention. RL enables agents to adapt strategies through interactions with dynamic threats, simulating attackers' tactics to identify vulnerabilities and test defenses. This paper develops defender agents and evaluates strategies using CyberBattleSim, an RL-based simulation tool developed by the Microsoft Defender team [2]. We introduce the 'CIA Triad-based Capture the Flag (CT2F)' scenario, an adaptation of the original 'ToyCTF' scenario that integrates the CIA Triad and a cyber deception element, the 'Honeypot'. The effectiveness of these strategies is compared using RL algorithms like Q-Learning, Deep Q-Network (DQN), and Advantage Actor-Critic (A2C) aiming to enhance intelligent security measures.

Ⅱ. Proposed Method

2.1. CIA Triad-based Capture the Flag Scenario with Cyber Deception

The original 'ToyCTF' scenario in CyberBattleSim simulates security vulnerabilities and attacks but lacks the complexity of real-world networks [2]. To address these limitations, we developed the 'CT2F' scenario, which incorporates the CIA Triad principles of Confidentiality, Integrity, and Availability. Confidentiality is enhanced by restricting access to critical nodes exclusively to HTTPS and by enhancing firewall rules. Integrity is maintained through the application of the least privilege principle and the identification of additional vulnerabilities. Availability is ensured by reassessing critical infrastructure nodes. Furthermore, a 'Honeypot' was integrated to lure attackers, gather threat intelligence, and increase the costs of attacks by imposing penalties on compromises within the 'CT2F' scenario. These improvements enable the 'CT2F' scenario to more effectively simulate dynamic security environments, enhancing system security and reliability.

2.2. Cyber Defense Strategy in CT2F

CyberBattleSim, an RL-based tool, simulates interactions between attackers and defenders. It provides agents with node information to develop effective defense strategies. Our goal is to analyze and optimize defense strategies using RL. Fig. 1 shows the architecture and interaction between attacker and defender agents, where attackers perform local and remote attacks, and defenders monitor credentials, revoke vulnerabilities, and adjust firewall rules. We developed the CredentialScanAndReimage (CSR) and AdaptiveFirewall (AF) agents. The CSR agent monitors credentials and reimages hosts as needed, enhancing credential management. In contrast, the AF agent dynamically adjusts firewall rules in response to suspicious traffic patterns, optimizing resources and improving network security. These dynamic

adjustment processes enable real-time responses to emerging security threats and significantly enhance the security of the network.

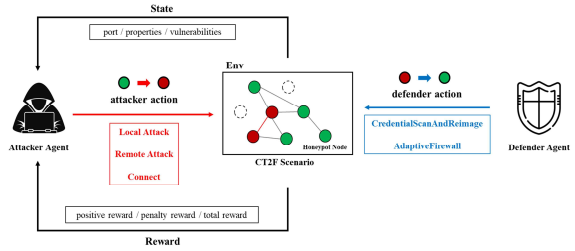

Fig. 1 CyberBattleSim Architecture Overview

Ⅲ. Experiment and Results

This paper evaluates the effectiveness of the CSR and AF defender agents in the 'CT2F' scenarios, using RL algorithms such as Q-Learning, DQN, and A2C. Fig. 2 shows the cumulative rewards of DQN, Q-Learning, and A2C in the 'CT2F' scenario with CSR and AF defenders. In particular, the A2C with CSR (A2C_CSR) defender achieved the highest cumulative rewards, clearly outperforming the others. The Q-Learning with CSR (QL_CSR) and DQN with CSR (DQN_CSR) defender also showed competitive performance, though slightly less effective than A2C_CSR. In contrast, the performance of AF defenders was consistently lower across all algorithms. Specifically, the DQN with AF (DQN_AF) defender, Q-Learning with AF (QL_AF) defender, and A2C with AF (A2C_AF) defender registered significantly lesser rewards in the 'CT2F' scenario.

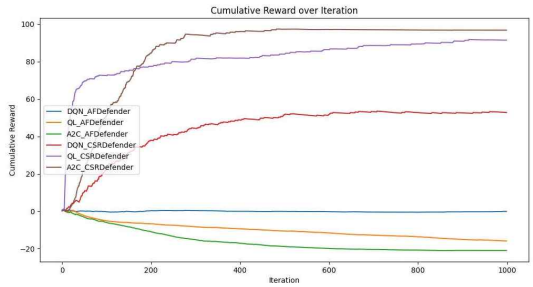

Fig. 2 Cumulative Reward in CT2F

Overall, these results emphasize the significant impact of defender agents on RL algorithms. Although DQN demonstrated superior performance relative to Q-Learning in the scenario without defenders, achieving a maximum rate of 0.60, both A2C and Q-Learning lagged with maximum rates of 0.18 and 0.12, respectively. In the CSR scenario, DQN still outperformed the others with a maximum rate of 0.27, followed by A2C at 0.21 and Q-Learning at 0.14. In the AF scenario, DQN maintained the highest success rate of 0.33, while A2C and Q-Learning achieved maximum rates of 0.19 and 0.08, respectively. Table 1 shows the success rates for DQN, Q-Learning, and A2C in various scenarios. These results highlight the adaptability of DQN in dynamic environments, while also demonstrating the competitiveness of A2C. The results underscore the importance of selecting appropriate algorithms and defender agents to optimize RL strategies in cyber-dynamic environments.

Table 1 Success Rate in CT2F

Scenario	Algorithm	Max	Mean
C2TF without Defender	Q-Learning	0.12	0.01
	DQN	0.60	0.04
	A2C	0.18	0.03
C2TF with CSR	Q-Learning	0.14	0.01
	DQN	0.27	0.03
	A2C	0.21	0.04
C2TF with AF	Q-Learning	0.08	0.01
	DQN	0.33	0.05
	A2C	0.19	0.03

Ⅳ. Conclusion

In this paper, we developed the CSR and AF defense agents using CyberBattleSim to validate their effectiveness through RL. The CSR agent managed credential-related threats, while the AF agent dynamically adjusted to network traffic to enhance security. We developed the 'CT2F' scenario based on the CIA Triad, integrating a 'Honeypot' to increase attack costs and gather threat intelligence. Our analysis focused on the impact of defender agents on attacker performance using Q-learning, DQN, and A2C in the 'CT2F' scenario. The CSR agent reduced attacker success rates by up to 27% with DQN, while the AF agent reduced attacker success rates by up to 33% with DQN in the 'CT2F' scenario. Additionally, A2C with the CSR agent achieved a success rate reduction of up to 21%, demonstrating competitive performance compared to DQN. These results demonstrate the effectiveness of the proposed defense techniques in deterring attackers and emphasize the potential of RL for autonomous cybersecurity defense. Future work will explore additional RL algorithms, more complex scenarios, and the integration of multi-agent systems to better model attacker-defender interactions.

ACKNOWLEDGMENT

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation grant funded by the Korean government (No. RS-2022-II220961)

참 고 문 헌

[1] Bennani, H. "Cybersecurity, Cybercrime and the Video Gaming Industry," Utica University, May. 2022.

[2] Microsoft, "CyberBattleSim," 2021, (https://github.com/microsoft/CyberBattleSim).