AI-Based Automated Threat Scenario Generation Leveraging the ATT&CK Framework*

Sarang Na[1], Seongmin Park[1], Daeun Kim[1], Jingang Kim[1], and Joonhyung Lim[1]

Korea Internet & Security Agency, Republic of Korea
no.1.nasa@kisa.or.kr, smpark@kisa.or.kr, whale53@kisa.or.kr,
kimjg@kisa.or.kr, lim@kisa.or.kr

Abstract

This study proposes an AI-based method for automated threat scenario generation using the MITRE ATT&CK framework and Deep Q-Learning (DQN). By mapping ATT&CK techniques to reinforcement learning actions, the system simulates realistic attack paths and adaptive behaviors. It achieved over 98% success in multi-step attack simulations, demonstrating its potential to enhance realism and efficiency in cyber defense training.

1 Introduction

Modern warfare has evolved beyond traditional physical domains to encompass the digital battlespace, as recent conflicts such as the Russia–Ukraine and Hamas–Israel wars demonstrate the strategic importance of information and network-based capabilities. Offensive operations—ranging from critical infrastructure disruption to data exfiltration and influence campaigns—produce far-reaching consequences that demand systematic preparation and realistic training. Exercises in the information domain strengthen national defense by improving operational readiness and enabling early detection of security and operational weaknesses; however, conventional approaches—often passive in nature—remain inadequate for today's rapidly shifting threat environment. To address these limitations, this study presents an AI-driven framework for generating realistic and adaptive threat scenarios to support next-generation defense training.

2 Related Work

MITRE ATT&CK provides a widely adopted knowledge base that organizes 14 tactics and 196 adversary techniques, supporting realistic modeling of attacker behavior to inform defense hardening and incident response [**?**]. Atomic Red Team supplies an open-source collection of ATT&CK–mapped attack scripts for lightweight scenario testing [**?**], and MITRE's Caldera platform exposes ATT&CK via APIs while integrating third-party tools (including Atomic) to automate adversary emulation in training environments [**?**]. NASimEmu is a penetration-testing framework that generates diverse network attack scenarios and supports AI-driven training by simulating vulnerabilities and privilege escalation in realistic virtual environments [**?**]. In this paper, we extend NASimEmu to simulate ATT&CK techniques for the automated generation and execution of adaptive threat scenarios.

*Proceedings of the 9th International Conference on Mobile Internet Security (MobiSec'25), Article No. P-28, December 16-18, 2025, Sapporo, Japan.

3 AI-based ATT&CK simulation method

Automated scenario generation is achieved through the application of Deep Q-Learning (DQN). The Q-network receives the current environment state as input and produces Q-values for all possible actions, effectively approximating the state–action value function. This approach enables efficient policy learning even within high-dimensional state spaces. The overall simulation architecture consists of several key components:

- `Environment` – the target system or simulated network under attack
- `Agent` – the learning entity that interacts with the environment to maximize rewards
- `State` – variables describing the system's condition
- `Action` – possible operations the agent can take within the environment
- `Reward` – feedback on the outcome of an executed action (goal: +100, failure: 0, cost: –1)
- `Policy` – decision strategy for selecting actions in each state

The trained model generates threat scenarios by updating host states—such as Discovered, Compromised, and Accessed—according to the success or failure of simulated attacks. To enable diverse and realistic attack paths, individual ATT&CK techniques are mapped to the agent's action space, allowing the agent to select and execute specific techniques (e.g., scanning (4), exploitation (1), privilege escalation (2), and impact operations (5)) as actionable steps within the environment. Sequential dependencies between techniques are incorporated, and each scenario iterates until an objective—such as data exfiltration or system shutdown—is achieved.

We implemented a simulation module using NASimEmu, comprising up to three subnets with one or two nodes each. Each action was executed with an 80% probability under the defined objective of disrupting the target system. The trained model was then used to generate threat scenarios. In 5,000 runs, the attacks succeeded in roughly three steps with a probability exceeding 98%. The most common sequence was: (1) compromise the currently accessible node (R) (90%) $\rightarrow$ (2) perform subnet scanning (90%) $\rightarrow$ (3) execute T1499 against the target node (S) (99.8%). Outcomes varied depending on firewall settings, open ports, and process configurations, underscoring the flexibility of the simulation environment.

4 Conclusion

We propose an AI-based method for simulating ATT&CK techniques, enabling automated cyberattack training across diverse environments and threat scenarios. Future work will extend the simulation and emulation capabilities to support more complex lateral movements and advanced attack objectives in IT–OT converged environments.

5 Acknowledgments

This work was supported by the Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korean government (MSIT) (RS-2023-00220303).

References

[1] Caldera. https://caldera.mitre.org.

[2] Jaromír Janisch, Tomáš Pevný, and Viliam Lisý. NASimEmu: Network attack simulator & emulator for training agents generalizing to novel scenarios, 2023. arXiv:2305.17246.

[3] Atomic Red Team. https://atomicredteam.io.

[4] MITRE ATT&CK. https://attack.mitre.org.