전자통신동향분석, 41권 1호 2026, 82-95

강화학습 기반 사이버 공격 시뮬레이션 및 에뮬레이션 환경 기술 동향

Trends and Comparative Analysis on Reinforcement Learning-Based Cyber Attack Simulation and Emulation Environments

김범석 (B.S. Kim, bskim97@etri.re.kr) 지능형네트워크보안연구실 석사후연수연구원
구기종 (K.J. Koo, kjkoo@etri.re.kr) 지능형네트워크보안연구실 책임연구원
최양서 (Y.S. Choi, yschoi92@etri.re.kr) 지능형네트워크보안연구실 책임연구원
유재학 (J.H. Yoo, dbzzang@etri.re.kr) 지능형네트워크보안연구실 책임연구원
문대성 (D.S. Moon, daesung@etri.re.kr) 지능형네트워크보안연구실 책임연구원

ABSTRACT

With the increasing sophistication and automation of cyberattacks, simulation and emulation environments that replicate attack processes in virtual settings have become essential for evaluating defensive strategies. Moreover, research has been conducted on the application of (RL) in cyber ranges to enable autonomous penetration testing. This study provides a comparative analysis of the characteristics and limitations of RL-based cyber-attack simulation environments, including Network Attack Simulation, Cyber Battle Simulation,and the Autonomous Pentesting framework based on Reinforcement Learning, as well as emulation environments such as Cyber Game for Intelligent Learning and Pentesting Gym. In addition, this paper summarizes the characteristics of hybrid environments, including Network Attack Simulation and Emulation and the Generalizable Autonomous Pentesting Framework. In particular, the analysis includes structural differences, Markov decision process design, scalability to large-scale networks, and generalization capability across diverse scenarios. In the future, providing high-fidelity learning environments that closely resemble real networks will require integrated training architectures that combine simulation and emulation, along with expansion toward multi-agent environments.

KEYWORDS Cyber Attack, Emulation, Generalization, Penetration Testing, Reinforcement Learning, Scalability, Simulation

* DOI: https://doi.org/10.22648/ETRI.2025.J.410108
* 본 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임[No. 2022-0-00961, 자가진화형 AI 기반 사이버 공방 핵심원천기술 개발].

I. 서론

오늘날 인공지능 기술의 발전으로 산업 전반에 자동화와 지능화를 이루었지만, 동시에 보안 취약점을 악용한 사이버 공격은 더욱 정교하고 다양하게 진화하고 있다[1]. 더 나아가 네트워크 인프라, 클라우드, IoT 환경으로 사이버 공간이 확장되면서 공격 표면이 넓어지고 있다[2]. 이로 인해 기존의 수동적 침투 테스트 방법만으로는 고도화된 위협에 대해 신속하고 효과적으로 대응하기 점점 어려워지고 있다.

전통적으로 침투 테스트는 전문가의 경험과 수작업 절차에 의존해왔다[3]. 그러나 수작업 중심의 방식은 상당한 시간과 비용이 요구되므로 대규모 네트워크나 동적으로 변화하는 환경에 대해 실시간으로 대응하기는 매우 어렵다.

수작업에 의존하는 기존 침투 테스트의 한계를 보완하기 위해 초기에는 네트워크 구성을 바탕으로 공격 그래프를 생성하고 가능한 모든 공격 경로를 열거하는 방법이 활용되었다[4]. 그러나 공격 그래프 생성 과정은 네트워크 규모가 커질수록 경로 수가 기하급수적으로 증가하기 때문에 실질적인 적용을 하는 데 어려움이 있다. 이후 복잡한 공격 경로를 사전에 열거하는 방식에서 벗어나 머신러닝을 활용하여 네트워크의 특성을 입력으로 받아 효과적인 공격 경로를 직접 예측하는 방향으로 전환되었다[5]. 특히 강화학습은 공격자가 수행하는 일련의 단계를 순차적 의사결정 모델로 모델링할 수 있어 자율 침투 테스트 분야에서 효과적인 모델링 방법으로 사용되고 있다. 강화학습 기반 자율 침투 테스트는 환경과의 지속적인 상호작용을 하며 보상을 통해 최적의 공격 전략을 학습하므로 복잡한 공격 경로 탐색을 자동화하는 데 활용될 수 있다.

최근에는 사이버 레인지에 강화학습을 적용하여 자율적으로 침투 전략을 학습하고 훈련하는 연구가 진행되고 있다[6]. 초기 연구들은 강화학습 기반 사이버 공격 시뮬레이션 환경을 중심으로 발전해 왔다[7,8]. 시뮬레이션 환경은 가상 네트워크 모델에서 동작하기 때문에 빠르고 효율적인 학습이 가능하지만, 추상화 수준이 높아 실제 네트워크의 복잡한 동작을 충분히 반영하기 어렵다는 한계를 가진다. 시뮬레이션의 한계를 보완하기 위해 강화학습 기반 사이버 공격 에뮬레이션 환경이 등장하였다[9]. 에뮬레이션 환경은 실제 운영 체제와 공격 도구를 사용하여 높은 현실성을 제공하기 때문에 정책의 실제 환경 전이 성능을 평가하기에 적합하지만, 높은 구축 비용과 연산 자원 소모로 인해 학습 효율성이 낮아진다. 시뮬레이션과 에뮬레이션의 상반된 특성으로 인해 최근 연구들은 시뮬레이션과 에뮬레이션을 결합한 하이브리드 구조를 제안하고 있다[10]. 하이브리드 구조는 현실성과 효율성이 모두 높다는 장점이 있지만, 특정 환경 구성에 과도하게 적응한 정책이 새로운 시나리오에서 성능이 저하될 수 있다. 또한, 대규모 네트워크나 다양한 취약점 조합으로 인해 상태 및 행동 공간이 확장될 경우 학습 안정성과 전이 성능을 유지하기 어렵다. 이에 따라 강화학습 기반 자율 침투 테스트 분야에서는 대규모 환경 변화에 견딜 수 있는 확장성[11]과 다양한 시나리오에서도 정책의 일관성을 유지할 수 있는 일반화 능력[12]이 중요한 도전 과제로 떠오르고 있다.

이에 본고에서는 강화학습 기반 사이버 공격 시뮬레이션 환경과 에뮬레이션 환경의 기술 동향을 분석하고 각 환경의 구조와 특성을 비교하였다. 아울러 확장성과 정책의 일반화 성능을 달성하기 위한 최근 연구 방향을 고찰함으로써 향후 실세계에서 활용 가능한 지능형 침투 테스트 환경의 구축 가능성을 제시한다.

II. 사이버 공격 시뮬레이션 환경

본 장에서는 사이버 공격 시뮬레이션의 개념을 정리하고 대표적인 강화학습 기반 사이버 공격 시뮬레이션 환경인 Network Attack Simulation(NASim)[13], Cyber Battle Simulation(CyberBattleSim)[14] 그리고 Autonomous Pentesting framework based on Reinforcement Learning(APRIL)[15]을 중심으로 각 환경의 특징과 한계를 제시한다.

1. 사이버 공격 시뮬레이션 개요

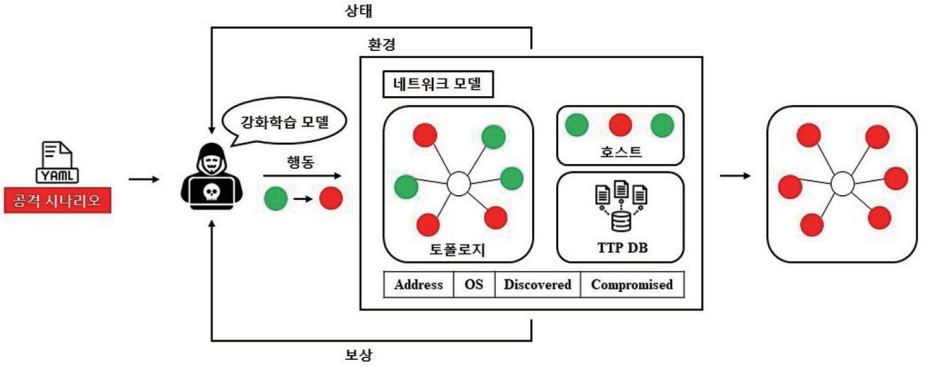
사이버 공격 시뮬레이션 환경은 실제 공격 과정을 가상화된 환경에서 재현하여 보안 시스템의 취약점을 진단하고 대응 전략의 효과를 검증하기 위한 환경이다. 그림 1은 시뮬레이션 환경에서 공격 에이전트가 상태를 관찰하고 행동을 수행하며 보상을 받는 학습 과정의 전체적인 구조를 나타낸다. 그림 1에서 볼 수 있듯이 사이버 공격 시뮬레이션 환경에서는 침투 테스트 시나리오를 기반으로 네트워크 내 공격 절차를 모의실험하고 네트워크 토폴로지, 호스트 구성 그리고 공격 행위의 모델링을 통해 실제 시스템에 영향을 주지 않고 다양한 공격 시나리오를 반복 학습하여 정책을 검증할 수 있다.

2. 사이버 공격 시뮬레이션 개발 현황

2.1 Network Attack Simulation

강화학습 기반 사이버 공격 시뮬레이션 연구 초기에는 실제 네트워크의 복잡성을 효율적으로 재현하기 어렵다는 제약이 존재하여 단순화된 가상 공격 환경[16,17]들이 제안되었다. 이로 인해 초기 연구들은 강화학습 기법의 적용 가능성을 검증하는 데 초점을 두면서 Schwartz 등[9]은 NASim을 제안하였다. NASim은 취약성을 가진 컴퓨터 네트워크 환경을 모사하여 침투 테스트 과정을 시뮬레이션하도록 설계된 대표적인 강화학습 기반 사이버 공격 시뮬레이션 환경이다. NASim에서는 MITRE ATT&CK 프레임워크 기반의 다양한 시나리오를 제공하며 공격자의 Tactics, Techniques and Procedures (TTPs)를 단순화된 형태로 모델링한다.

NASim은 호스트, 운영체제, 서비스, 방화벽 등 네트워크 구성요소를 추상화된 상태로 표현하며 스캔, 익스플로잇, 권한 상승과 같은 주요 공격 단계를 행동으로 정의한다. 보상 함수는 자산 발견, 공격 성공, 목표 달성 여부 등을 기준으로 구성되며 상태 전



아이콘 출처 Flaticon.com

그림 1 사이버 공격 시뮬레이션 환경의 강화학습 에이전트 학습 구조

이는 시나리오에 따라 확률적으로 결정된다. 또한, NASim은 불완전한 관측 상황을 반영하기 위해 전체 환경을 POMDP로 모델링하여 에이전트가 제한된 정보하에서 공격 경로를 탐색하도록 설계되었다.

NASim에서는 Q-Learning과 Deep Q-Network (DQN) 같은 Off-Policy 계열 강화학습 알고리즘을 적용하여 공격 정책 학습 성능을 분석하였다. 실험 결과 두 알고리즘 모두 목표 호스트까지의 공격 경로를 안정적으로 학습하는 것으로 확인되었다.

NASim은 추상화된 네트워크 자산과 공격 절차를 규칙 기반으로 구성하여 강화학습 알고리즘을 평가할 수 있는 시뮬레이션 환경을 제공한다. NASim의 구조는 다양한 공격 시나리오에서 정책 학습 성능을 검증하는 데 활용되며 자율 침투 테스트 에이전트 개발과 보안 전략 평가 연구를 위한 환경으로 사용된다.

2.2 Cyber Battle Simulation

NASim을 포함한 초기 강화학습 기반 사이버 공격 시뮬레이션 연구들[8,18]은 단순화된 자산 모델과 제한된 공격 단계 정의로 인해 실제 공격 절차의 복잡성을 충분히 반영하지 못하였다. 이로 인해 실제 네트워크 침투 과정의 복잡도를 반영할 수 있는 환경이 요구되면서 Microsoft 365 Defender 연구팀[14]은 CyberBattleSim을 제안하였다. CyberBattleSim은 엔터프라이즈 네트워크 내 침투 과정에서 자주 수행되는 탐색, 권한 상승, 측면 이동 등의 단계를 더욱 세부적으로 모델링한 강화학습 기반 시뮬레이션 환경이다. CyberBattleSim은 여러 개의 호스트 노드로 구성된 네트워크를 정의하며 각 노드는 취약점, 자격 증명, 권한 수준, 서비스와 같은 다양한 속성을 갖도록 설계되어 있다.

CyberBattleSim은 공격자가 확보한 접근 권한 수준, 접근 가능한 호스트 목록, 발견된 자격 증명 등 공격 진행 상황을 반영하는 정보로 상태를 구성하며, 포트 스캔, 서비스 탐색, 자격 증명 획득, 권한 상승, 측면 이동과 같은 공격 절차를 행동으로 정의한다. 보상 함수는 새로운 호스트 침투나 권한 획득과 같은 공격 목표 달성도를 기준으로 구성되며 상태 전이는 고정된 규칙에 의해 결정된다. 또한, CyberBattleSim은 NASim과 동일하게 부분 관측 환경을 가정하며 전체 환경을 POMDP 기반 의사결정 구조로 모델링한다.

CyberBattleSim[19]에서는 NASim과 달리 Off-Policy 계열 강화학습 알고리즘인 Q-Learning과 DQN뿐만 아니라 On-Policy 계열 강화학습 알고리즘인 Proximal Policy Optimization(PPO)와 Advantage Actor-Critic(A2C)도 적용하여 공격 정책 학습 성능을 비교하였다. 실험 결과 단순한 네트워크 구성에서는 Off-Policy 알고리즘이 On-Policy 알고리즘보다 빠르게 수렴하는 경향을 보였으나 복잡한 네트워크 구성에는 On-Policy 알고리즘이 Off-Policy 알고리즘에 비해 상대적으로 안정적인 성능을 보였다.

CyberBattleSim은 엔터프라이즈 네트워크 요소를 학습 가능한 형태로 구성하여 복잡하고 다양한 침투 경로 탐색 전략을 분석할 수 있는 환경으로 활용된다.

2.3 Autonomous Pentesting framework based on Reinforcemnt Learning

NASim과 CyberBattleSim을 포함한 기존 강화학습 기반 사이버 공격 시뮬레이션 환경[11,20]은 비교적 단순한 네트워크 구성이나 제한된 공격 단계에서 효과적이지만, 네트워크 규모가 커질수록 상태 공간과 행동 공간이 기하급수적으로 확장된다. 이로 인해 이산 공간에서 대규모의 네트워크 구성 요소를 직접 처리할 경우 조합 폭발 문제가 발생하여 수천 개의 비정형 침투 행동을 효율적으로 학습하기 어려워지

면서 Zhou 등[15]은 APRIL을 제안하였다. APRIL은 대규모 이산 행동 공간을 연속 임베딩 공간으로 변환하여 행동 간의 의미적 유사성을 정책 학습 과정에 반영할 수 있도록 설계한 환경이다.

APRIL의 상태 공간은 전역 네트워크 구조를 사용하지 않고 포트, 실행 중인 서비스, 권한 수준, 알려진 취약점 등 호스트 단위의 지역 정보로 구성된다. 각 호스트의 원시정보는 Sentence Embedding과 Word Embedding을 통해 고정 길이의 벡터로 변환되어 네트워크 규모 확장에 따른 상태 표현 문제를 해결한다. 행동 공간은 포트 스캔, 서비스 탐색, 익스플로잇, 권한 상승 등 수천 개 이상의 침투 관련 행동으로 이루어진 대규모 이산 집합으로 정의된다. APRIL은 National Vulnerability Database(NVD)와 Wikipedia 텍스트를 기반으로 사전 학습된 Transformer-based Denoising AutoEncoder(TSDAE) 임베딩 모델을 통해 의미적으로 유사한 행동이 연속 임베딩 공간에서 가깝게 배치되도록 구성함으로써 정책이 행동 간 의미적 유사성을 활용하여 탐색 효율을 높일 수 있도록 한다. 보상 함수는 정보 획득 또는 취약점 익스플로잇 성공 시 부여되는 양의 보상과 행동 비용을 반영한 음의 보상으로 구성된다.

더 나아가 APRIL은 정책 학습 단계에서 에이전트가 연속 임베딩 공간에서 Proto-Action을 생성한 후 Approximate Nearest Neighbors(ANN)를 통해 실제 이산 행동 집합에서 k개의 후보 행동을 검색한다. 이후 Twin Critic 네트워크가 후보 행동의 장기 가치를 평가하고 Upper Confidence Bound(UCB) 기반 점수를 통해 최종 행동을 선택한다. 선택된 행동은 환경에 적용되어 리플레이 버퍼에 저장되고 Off-Policy 방식으로 정책이 업데이트된다. 추가적으로 APRIL은 거리 인식 손실과 대조 손실을 함께 사용하여 임베딩 공간의 의미적 구조를 유지되도록 학습함으로써 수렴 안정성과 일반화 성능을 향상시켰다.

APRIL에서는 DQN, PPO, Heuristically Assisted Deep Reinforcement Learning(HA-DRL) 등의 다양한 강화학습 알고리즘을 활용하여 단일 호스트 및 체인 시나리오에서 확장성과 일반화 성능을 평가하였다. 단일 호스트 50개를 활용하여 행동 공간의 크기를 변화시키며 확장성을 평가한 결과 APRIL은 기존 알고리즘인 DQN, PPO, HA-DRL 대비 빠르게 수렴하며 안정적인 학습 성능을 보였다. 체인 시나리오 기반 일반화 실험에서는 네트워크 구조가 변경된 환경에서도 APRIL이 다른 알고리즘 대비 우수한 전이 성능을 보였다. 또한, 후보 행동 탐색 과정에서 최근접 이웃 수를 변화시키며 분석한 결과 k값이 100일 때 탐색 다양성과 계산 비용 간의 균형이 가장 적절한 것으로 확인되었다. 마지막으로 UCB 탐색과 Distance-Aware Loss 적용 여부를 실험한 결과 두 요소 모두 정책 안정성과 성능 향상에 기여하는 것으로 나타났다.

APRIL은 대규모 이산 행동 공간을 효율적으로 처리하기 위한 다양한 전략을 결합함으로써 강화학습 기반 자율 침투 테스트 분야에서 높은 확장성과 일반화 성능을 동시에 달성할 수 있는 환경으로 활용된다.

3. 사이버 공격 시뮬레이션 한계

사이버 보안 침투 테스트 분야에서 강화학습 기반 사이버 공격 시뮬레이션 환경은 빠른 학습 속도와 높은 실험 재현성을 제공하는 유용한 도구로 활용되고 있다. 그러나 기존 환경들은 단순화된 전이 모델과 추상화된 자산 정보를 사용하기 때문에 실제 네트워크의 복잡한 동작을 충분히 반영하지 못한다. 대부분의 시뮬레이션은 포트, 서비스, 취약점과 같은 정적인 속성을 중심으로 상태를 구성하며 행동의 성공 여부 또한 사전에 정의된 규칙 또는 고

정 확률에 의해 결정된다. 이러한 구조는 네트워크 트래픽 변동이나 프로토콜 상호작용과 같은 비결정적 요소를 반영하기 어렵게 만들어 학습된 정책이 실제 환경에서 성능이 저하되는 Reality Gap 문제로 이어진다. 또한, 시뮬레이션은 높은 샘플 효율을 얻기 위해 Off-Policy 계열 알고리즘을 주로 활용하지만, 추상화된 전이 모델에서 수집된 경험을 반복적으로 재사용하는 과정에서 Q-Value 과추정, 정책 불안정성, 특정 행동에 대한 편향 탐색이 나타난다. 반면 On-Policy 방식은 상대적으로 안정적인 정책 업데이트를 제공하지만 단순화된 관측 구조로 인해 복잡한 공격 절차를 정밀하게 모사하기 어렵고 탐색 비용이 증가한다.

시뮬레이션의 구조적 한계는 현실적인 공격 절차를 충분히 재현하기 어렵다는 것을 보여주며 실제 시스템 기반의 높은 현실성을 제공하는 강화학습 기반 에뮬레이션의 필요성을 부각시켰다. 이에 따라 실제 네트워크상에서 공격을 직접 실행할 수 있는 환경들이 후속 연구[21,22]로 제안되었다.

III. 사이버 공격 에뮬레이션 환경

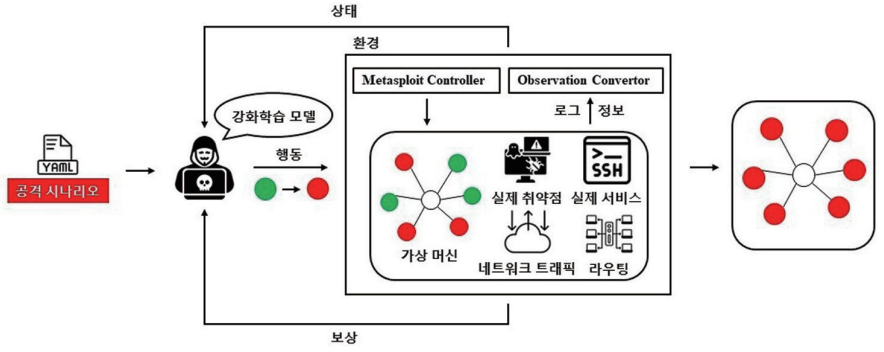
본 장에서는 사이버 공격 에뮬레이션의 개념을 정리하고 대표적인 강화학습 기반 사이버 공격 에뮬레이션 환경인 Cyber Game for Intelligent Learning (CyGIL)[23]과 Pentesting Gym(PenGym)[24]을 중심으로 각 환경의 특징과 한계를 제시한다.

1. 사이버 공격 에뮬레이션 개요

사이버 공격 에뮬레이션 환경은 실제 네트워크 환경과 유사한 가상 인프라에서 공격 행위를 직접 실행하여 보안 시스템의 대응 성능을 검증하기 위한 환경이다. 그림 2는 에뮬레이션 환경에서 공격 에이전트가 실제 취약점, 서비스, 네트워크 트래픽과 상호작용을 하며 학습하는 전체적인 구조를 나타낸다. 그림 2에서 볼 수 있듯이 사이버 공격 에뮬레이션 환경에서는 스캐닝, 취약점 익스플로잇, 권한 상승과 같은 공격 절차가 실제 도구와 서비스 위에서 동작하여 현실적인 공격 시나리오를 바탕으로 정책의 효과와 보안 대응 체계를 평가할 수 있다.

2. 사이버 공격 에뮬레이션 개발 현황

최근 사이버 보안 분야에서 실제 환경 수준의 공



아이콘 출처 Flaticon.com

그림 2 사이버 공격 에뮬레이션 환경의 강화학습 에이전트 학습 구조

격 대응 평가가 요구되면서 강화학습 기반의 사이버 공격 에뮬레이션 환경에 관한 연구가 진행되고 있다. 이에 따라 본 절에서는 호스트 및 네트워크 수준의 행동을 모두 지원하는 CyGIL과 실제 네트워크상에서 공격자의 행동을 직접 수행하는 PenGym을 제시한다.

2.1 Cyber Game for Intelligent Learning

단순화된 전이 모델을 기반으로 하는 시뮬레이션 환경[25,26]만으로는 최신 사이버 위협 환경의 복잡성과 산업용 공격 도구의 실제 동작을 충분히 반영하기 어렵다. 이로 인해 실제 시스템 수준의 상호작용을 포함하는 에뮬레이션 기반 훈련 환경의 필요성이 증가하면서 Li 등[23]은 CyGIL을 제안하였다. CyGIL은 엔터프라이즈급 네트워크에서 발생하는 실제 공격 절차를 강화학습 과정에 직접 반영할 수 있도록 설계된 사이버 작전 에뮬레이션 환경이다.

CyGIL은 최신 레드팀 도구와 오픈소스 네트워크 인프라를 모듈형 구조로 통합하여 다양한 에뮬레이션 환경과의 호환성을 제공한다. CyGIL의 학습 환경은 Networked Cyber Physical System(NCPS), Action Actuator(AA), Sensor View(SV)로 구성되며, 각각 네트워크 자산 관리, 공격 절차 실행, 관측 정보 제공 역할을 수행한다.

CyGIL의 관측 공간은 부분 관측 구조를 따르며 권한 수준, 사용자 계정, 서비스 구성, 네트워크 속성 등 호스트 단위 정보를 정규화된 형태로 제공한다. 행동 공간은 산업 현장에서 사용되는 실제 공격 도구와 MITRE ATT&CK 기반 TTP를 반영하여 구성되며 Metasploit 연동 및 CALDERA 확장 기능을 통해 사용자 정의 침투 절차를 추가함으로써 확장 가능하다. 보상 함수는 행동 비용 페널티를 반영한 음의 보상과 목표 달성 시 부여되는 양의 보상으로 구성되어 효율적인 침투 전략 학습을 유도하도록 설계되었다.

CyGIL에서는 DQN과 Cross-Entropy(CE) 알고리즘을 활용하여 두 가지 시나리오에서 에이전트의 침투 전략 학습 성능을 평가하였다. 단순한 시나리오에서는 수집 및 유출 절차로 구성된 제한된 행동 공간을 기반으로 목표 파일 유출 여부를 평가하였으며, 두 알고리즘 모두 사전 지식 없이도 최적의 행동 시퀀스를 학습하였다. 반면 복잡한 시나리오에서는 탐색, 자격 증명 접근, 권한 상승, 측면 이동으로 이어지는 다단계 공격 절차를 대상으로 평가한 결과 서브넷 구조와 자격 증명 의존 관계로 인해 성공률은 낮았으나 에이전트는 최종적으로 도메인 관리자 계정에 도달하는 전략을 학습하였다.

CyGIL은 최신 산업용 공격 도구와 실제 네트워크 상호작용을 통합하여 복잡한 네트워크 사이버 작전을 다루기 위한 확장 가능한 에뮬레이션 기반 강화학습 환경을 제공한다.

2.2 Pentesting Gym

기존 에뮬레이션 환경은 실제 시스템 동작을 반영할 수 있지만, 환경 구성과 초기화 및 재생성 과정의 자동화 수준이 낮고 운영 비용이 높아 강화학습에 필요한 대규모 반복 실험 환경을 구축하기 어렵다는 한계가 있었다. 이로 인해 현실성과 자동화를 동시에 갖춘 학습 환경의 필요성이 커지면서 Nguyen 등[24]은 PenGym을 제안하였다. PenGym은 실제 네트워크 환경을 자동으로 생성하고 관리할 수 있는 실제 환경 기반 강화학습 사이버 공격 에뮬레이션 환경이다.

PenGym은 학습 환경과 실제 공격 절차가 수행되는 사이버 레인지가 분리된 구조로 설계되어 있으며 두 환경은 Action-State 모듈을 통해 연결된다. Action-State 모듈은 에이전트가 선택한 논리적 행동

을 실제 공격 명령으로 변환하여 KVM 기반 가상머신 환경에서 실행하고, 그 결과를 관측 정보와 보상으로 재구성하여 학습 환경으로 전달한다.

PenGym의 상태 공간은 도달 가능한 호스트, 발견된 취약점, 획득한 권한 수준 등 부분 관측 정보로 구성되며 행동 공간에는 스캔, 서비스 탐색, 취약점 실행, 권한 상승 등 실제 공격 절차가 포함된다. 보상 함수는 Common Vulnerability Scoring System(CVSS) 기반 복잡도 점수와 행동 비용을 결합하여 침투 성과를 극대화하도록 설계되며 전이는 기존 시뮬레이션 환경과 달리 확률적 모델이 아닌 실제 공격 도구 실행 결과에 의해 결정된다.

더 나아가 PenGym은 자동 사이버 레인지 생성기를 통해 NASim 시나리오 구성 정보를 분석하여 CyRIS 기반 배포 파일로 변환하고 KVM 환경에 운영체제와 서비스를 자동 배포함으로써 현실적인 네트워크 환경을 구성한다. 또한, 서브넷 간 트래픽 처리 과정에서 브리지 제어 기법을 적용하여 실제 라우팅 흐름과 유사한 네트워크 동작을 구현한다.

PenGym에서는 Q-Learning, Q-Learning에 Experience Replay가 적용된 QL-Replay 그리고 DQN을 사용하여 NASim과의 학습 성능을 비교하였다. 실험 결과 단순한 시나리오에서는 두 환경 모두 유사한 수렴 성능을 보였으나 시나리오 복잡도가 증가하면 PenGym은 실제 시스템 동작 기반 전이로 인해 탐색 비용이 증가하여 NASim보다 수렴 속도는 느렸지만 더 현실성 있는 정책을 학습하는 것으로 나타났다.

PenGym은 실제 동작 기반 행동 실행 구조와 자동화된 사이버 레인지 생성 기술을 결합하여 현실성 높은 강화학습 기반 사이버 공격 에뮬레이션 환경을 제공한다. 이러한 구조는 실제 네트워크 기반 자율 침투 테스트의 학습 및 검증을 위한 신뢰성 높은 벤치마크 환경으로 활용된다.

3. 사이버 공격 에뮬레이션 한계

강화학습 기반 사이버 공격 에뮬레이션 환경은 실제 운영 체제, 서비스, 공격 도구 등을 직접 실행하여 높은 현실성을 제공한다. 그러나 기존 에뮬레이션 환경[27,28]은 공통적으로 구조적인 제약을 가지고 있어 대규모 자율 침투 테스트 환경으로 확장하기 어렵다.

대부분의 에뮬레이션은 VM 기반 자원을 사용하기 때문에 시나리오 규모가 커질수록 CPU 및 메모리 요구량이 급증하고 초기화 및 복구 단계에서 큰 오버헤드가 발생하여 학습 효율이 저하된다. 또한, 에뮬레이션에서 실제 도구를 사용함에도 불구하고 프로세스 스캔, 서비스 버전 식별, 방화벽 규칙, 자격 증명 처리와 같은 세부 절차가 단순화되어 관측 공간과 행동 공간이 실제 공격 절차를 충분히 반영하지 못한다. 마지막으로 특정 운영체제 또는 서비스 구성에 의존하는 구조적 편향이 존재하여 정책이 특정 플랫폼에 과적합될 가능성이 존재한다.

에뮬레이션의 구조적 한계는 대규모 네트워크 구성에서의 확장성, 다양한 시나리오에 대한 일반화 능력, 환경 변화에 대한 정책의 강건성 한계로 이어진다. 이에 따라 최근 연구에서는 시뮬레이션과 에뮬레이션을 결합한 하이브리드 구조[10,29]를 통해 확장성과 일반화 능력을 동시에 확보하는 방향으로 발전하고 있다.

IV. 사이버 공격 하이브리드 환경

본 장에서는 사이버 공격 하이브리드의 개념을 정리하고 대표적인 강화학습 기반 사이버 공격 하이브리드 환경인 Network Attack Simulation and Emulation(NASimEmu)[30]과 Generalizable Autonomous Pentesting framework(GAP)[31]을 중심

으로 각 환경의 특징과 한계를 제시한다.

1. 사이버 공격 하이브리드 개요

사이버 공격 하이브리드 환경은 시뮬레이션 환경과 에뮬레이션 환경을 결합하여 실제 네트워크와 유사한 조건에서 정책을 학습할 수 있도록 설계된 통합형 학습 환경이다. 사이버 공격 하이브리드 환경에서는 시뮬레이션 단계에서 정책을 효율적으로 학습하고 에뮬레이션 단계에서 실제 시스템 동작 기반의 전이를 통해 정책의 신뢰성을 검증하여 현실성 격차를 최소화하고 대규모 반복 학습과 고충실도 평가를 동시에 수행할 수 있다.

2. 사이버 공격 하이브리드 개발 현황

최근 강화학습 기반 침투 테스트 연구에서는 시뮬레이션 환경에서 대규모 반복 학습을 수행하는 방법과 에뮬레이션 환경에서 실제 시스템 동작을 기반으로 정책을 검증하는 방법을 결합하여 현실성 격차를 줄이고 일반화 성능을 향상시키려는 연구가 진행되고 있다. 이에 따라 본 절에서는 NASim을 확장하여 시뮬레이션 단계에서 학습된 정책을 에뮬레이션 단계에서 직접 검증할 수 있도록 설계된 NASimEmu과 시뮬레이션 환경과 실환경 간의 격차를 줄이기 위해 제안된 GAP을 제시한다.

2.1 Network Attack Simulation and Emulation

강화학습 기반 침투 테스트 연구에서 시뮬레이션 환경은 빠른 반복 학습과 다양한 시나리오 생성에 유리하지만, 추상화된 공격 절차로 인해 실제 환경과의 괴리가 존재한다. 반면 에뮬레이션 환경은 높은 현실성을 제공하지만, 구축 비용이 크고 반복 실험이 어렵다. 시뮬레이션 환경과 에뮬레이션 환경의 상반된 특성으로 인해 시뮬레이션과 에뮬레이션을 결합한 통합형 학습 환경이 요구되면서 Janisch 등[30]은 NASimEmu을 제안하였다. NASimEmu는 NASim 기반 시뮬레이션 환경과 산업 표준 도구 기반 에뮬레이션 환경을 공통 학습 인터페이스로 통합한 하이브리드 학습 환경이다.

NASimEmu의 목적은 시뮬레이션 환경에서 학습된 정책이 에뮬레이션 환경에서도 유효하게 동작하는지를 평가하여 추상화 수준에 따른 정책의 일반화 가능성을 검증하는 데 있다. 이를 위해 NASimEmu는 두 환경에서 동일한 학습 절차가 수행될 수 있도록 인터페이스를 통합하였다. 먼저 시뮬레이션 단계에서 NASim을 기반으로 정적, 랜덤 그리고 동적 시나리오를 자동 생성하여 구조적으로 상이한 환경에서 병렬 학습이 가능하도록 확장하였다. 이후 에뮬레이션 단계에서는 Vagrant를 통해 시나리오를 배포하고 Virtual Box상에서 Windows 및 Linux 기반 호스트를 구성하며 RouterOS를 통해 서브넷을 분리하여 실제 네트워크 운영 구조와 유사한 환경을 구현한다.

NASimEmu의 관측 공간은 호스트 도달 여부, 취약점 발견 상태, 침해 여부, 운영체제 유형, 서비스 구성, 프로세스 정보, 노드 가치 등으로 구성된 벡터 형태의 부분 관측 정보로 표현된다. 행동 공간은 스캔, 익스플로잇, 권한 상승 등으로 이루어진 이산 행동 집합으로 정의된다. 보상 함수는 각 행동의 비용을 반영한 음의 보상과 목표 달성 시 부여되는 양의 보상으로 설계되어 불필요한 행동을 최소화하도록 유도하였다. 또한, NASimEmu는 에피소드 종료 조건을 자동으로 처리하지 않고 별도의 종료 행동을 통해 종료하도록 설계하여 실행 전략의 자율성을 높였다.

NASimEmu에서는 PPO를 사용하여 두 가지

정책 모델인 Multi Layer Perceptron(MLP) 모델과 Size-Invariant 모델을 비교하였다. MLP 모델은 입력 차원을 고정하기 위해 모든 시나리오의 호스트 수를 30개로 제한하고 부족한 부분을 패딩으로 처리하는 방법을 사용한다. 그러나 이 방법은 입력 순서 변화에 민감하며 패딩 영역에서 과적합이 발생하여 새로운 시나리오로 전이될 때 성능이 크게 저하되는 문제가 발생하였다. 이와 달리 Size-Invariant 모델은 모든 호스트 정보를 공유 임베딩 함수로 변환한 뒤 평균 및 최대 연산을 통해 전역 상태 벡터로 요약하는 구조를 사용하여 입력 크기 및 순서 변화의 영향을 받지 않도록 설계되었다. 더 나아가 Sine-Cosine 기반의 Positional Embedding을 도입하여 호스트 탐색 순서와 공격 경로 정보를 추가로 인코딩함으로써 시나리오 간 전이성과 일반화 성능이 유의미하게 향상시켰다.

NASimEmu는 시뮬레이션 환경에서 학습된 정책을 실제 시스템 기반의 에뮬레이션 환경에서 직접 검증할 수 있도록 설계된 하이브리드 학습 환경으로 서로 다른 추상화 수준 간 정책의 전이 가능성과 일반화 성능을 평가하는 데 활용된다.

2.2 Generalizable Autonomous Pentesting Framework

강화학습 기반 자율 침투 테스트 연구에서는 시뮬레이션 환경이 높은 샘플 효율을 제공하지만 실제 환경으로의 전이가 어렵다. 반면, 에뮬레이션 환경은 높은 현실성을 제공하지만 데이터 수집 비용과 위험으로 인해 반복 학습이 제한되는 문제가 발생한다. 또한, 학습된 정책이 환경 구조나 취약점 구성이 조금만 변해도 성능이 급격히 저하되는 문제가 나타나면서 Zhou 등[31]은 참고문헌 [15]를 확장하여 GAP을 제안하였다. GAP은 현실 데이터를 활용한 시뮬레이션 도메인 확장과 메타 강화학습 기반의 전이 학습 구조를 결합하여 다양한 환경 변화에도 적응 가능한 일반화된 자율 침투 테스트 에이전트를 학습하기 위한 환경이다.

GAP은 현실 데이터를 시뮬레이션에 반영하여 정책을 학습하고 학습된 정책을 실제 환경에서 검증하는 순환적 학습 구조인 Real-to-Sim-to-Real 파이프라인을 중심으로 설계된 환경이다. GAP은 현실 데이터 기반 시뮬레이션 구축, 도메인 랜덤화 그리고 메타 강화학습을 결합하여 다양한 환경 변화에도 견고하게 작동하는 정책을 학습하도록 한다. GAP의 구조는 실제 또는 에뮬레이션 환경에서 수집된 호스트 구성, 서비스 정보, 취약점 지문 등을 시뮬레이션 파라미터로 변환하여 현실성을 갖춘 학습 환경을 생성하는 Real-to-Sim 단계와 시뮬레이션에서 학습된 정책을 실제 또는 고충실도 테스트 환경에서 검증하는 Sim-to-Real 단계로 구성된다. 또한, GAP은 LLM 기반 합성 환경 생성과 도메인 랜덤화를 통해 다양한 네트워크 구성과 취약점 배치를 변형하여 폭넓은 환경 분포를 제공하며, Meta-RL을 통해 이전에 보지 못한 환경에서도 빠르게 적응할 수 있는 일반화된 정책을 학습하도록 설계되었다.

GAP의 관찰 공간은 에이전트가 스캐닝을 통해 수집하는 포트 정보, 서비스 유형, 운영체제 지문, 웹 핑거 프린트 등의 정보로 정의된다. 이러한 관찰은 대부분 텍스트 로그 형태로 주어지기 때문에 GAP에서는 사전학습된 Sentence-BERT(SBERT) 인코더를 사용하여 중복되고 비정형적인 원시 로그를 벡터 형태의 상태 표현으로 변환한다. 행동 공간은 정보 수집 명령, 서비스 및 취약점 스캔, 익스플로잇 실행, 권한 상승 등 실제 공격 절차를 구성하는 명령으로 이루어진다. 이때 각 행동은 API를 통해 실제 명령으로 실행되며, 실행 결과는 다음 상태의 관찰 정보로 수집된다. 보상 함수는 정보 획득, 취약점 발

견, 성공적인 침해 등 목표 달성과 관련된 행동에 대해 양의 보상을 부여하고 실행 비용이 많이 드는 행동이나 불필요한 반복 행동에 대해 음의 보상을 적용하도록 설계되어 에이전트가 의미 있는 공격 경로를 탐색하도록 유도한다.

GAP은 Vulhub 기반 고충실도 취약 환경에서 PPO, ARPIL, GAP을 비교하여 정책 학습 및 전이 성능을 평가하였다. GAP에서는 추가 학습 없이 새로운 환경에서도 즉시 정책을 시행하는 Zero-Shot 일반화 능력과 새로운 환경에서 소량의 상호작용만으로 빠르게 적응하는 Few-Shot 적응 능력을 중심으로 분석하였다. 실험 결과 행동 공간의 크기가 증가할수록 학습 곡선의 수렴 속도가 감소하고 학습 시간은 증가하였다. 또한, 동일한 취약점을 공유하나 호스트 구성만 변형된 유사 환경에서 Zero-Shot 전이 성능을 평가한 결과 PPO와 APRIL 대비 일반화 간극을 유의미하게 줄이고 더 높은 성공률을 보였다. 마지막으로 서로 다른 취약점을 가진 상이 환경에서 Few-Shot 정책 적응 성능을 분석한 결과 GAP-Transfer는 초기 Zero-Shot 성능과 소량의 상호작용만으로 빠르게 수렴하는 적응 성능 모두에서 가장 우수한 성능을 보였다. 특히 평균 학습 시간을 약 40% 단축하였으며 APRIL 대비 약 22% 빠른 적응 속도를 달성하였다.

GAP은 다양한 환경 변화에도 견고한 정책 학습과 전이를 가능하게 하여 시뮬레이션과 실환경 간의 간극을 줄이고 일반화 가능한 자율 침투 테스트 방법을 평가하기 위한 벤치마크 환경으로 활용된다.

V. 사이버 공격 시뮬레이션 및 에뮬레이션 환경 설계를 위한 고려사항과 향후 연구 방향

표 1은 강화학습 기반 사이버 공격 시뮬레이션과 에뮬레이션 환경의 특징을 비교한 것이다. 표 1에서 볼 수 있듯이 본고에서는 NASim, CyberBattleSim 그리고 APRIL 등의 3가지 시뮬레이션 환경과 2가지 에뮬레이션 환경인 CyGIL과 PenGym뿐만 아니라 2가지 하이브리드 환경인 NASimEmu와 GAP을 분

표 1 강화학습 기반 사이버 공격 환경 특성 비교

환경 / 특성	NASim	CyberBattleSim	APRIL	CyGIL	PenGym	NASimEmu	GAP
시뮬레이션 기반	○	○	○	×	×	○	○
에뮬레이션 기반	×	×	×	○	○	○	○
호스트 기반 익스플로잇	○	○	○	○	○	○	○
네트워크 기반 익스플로잇	△	△	△	×	○	○	○
확률적 전이	○	○	○	○	○	○	○
부분 관측성	○	○	○	○	○	○	○
확장성	○	△	○	△	△	△	○
현실성	×	×	△	○	○	○	○
일반화 능력	△	△	○	△	○	○	○
소스코드 제공 여부	○	○	○	×	○	○	○

석하여 각 환경의 구조적 특성과 차이점을 정리하였다.

이러한 비교 및 분석을 바탕으로 시뮬레이션 및 에뮬레이션 환경을 효과적으로 설계하기 위해서는 다양한 기술적 고려 사항이 요구된다. 먼저 시뮬레이션 환경의 경우 실제 네트워크에서 발생하는 비동기 이벤트, 동적 트래픽, 프로토콜 상호작용 등을 반영할 수 있는 고도화된 상태 전이 모델의 개발이 필요하다. 또한, 특정 시나리오에 과도하게 적응하는 문제를 완화하기 위해 도메인 랜덤화나 지속 학습 방법을 적용하여 정책의 일반화 성능을 향상시킬 필요가 있다. 더불어 보상 설계의 주관성을 줄이기 위해 역강화학습 기반 자동 보상 학습 방법을 도입하는 방안도 고려될 수 있다.

이와 달리 에뮬레이션 환경은 VM 기반 구조로 인해 자원 오버헤드가 크기 때문에 경량 가상화 기술, 스냅샷 기반 빠른 상태 복구, 분산 실행 구조와 같은 시스템 최적화 전략이 필요하다. 또한, 실제 공격 절차의 세부 단계를 정밀하게 모델링하여 관측 공간과 행동 공간이 실제 네트워크 동작을 충분히 반영할 수 있도록 개선하고 특정 구성에 편향되지 않은 범용적 환경 설계가 요구된다.

향후 연구 방향으로는 시뮬레이션 환경과 에뮬레이션 환경을 결합한 하이브리드 학습 환경이 주목받고 있다. 특히 디지털 트윈 기반 네트워크 모델링은 실제 시스템의 동작을 정밀하게 모사할 수 있어 현실성과 확장성을 동시에 확보할 방법으로 평가된다. 더 나아가 실제 사이버 공방 환경에서 발생하는 상호작용은 단일 에이전트 기반 학습만으로는 충분히 재현하기 어렵기 때문에 협력과 경쟁이 존재하는 멀티 에이전트 강화학습 구조로의 확장이 필요하다. 이러한 연구는 대규모 네트워크에서도 일반화 가능한 정책을 학습하는 지능형 자율 침투 테스트 환경의 실현 가능성을 높일 것으로 기대된다.

VI. 결론

최근 사이버 공간이 확장됨에 따라 위협은 이전보다 지능적이고 자동화된 형태로 진화하고 있다. 이에 대응하기 위해 사이버 레인지에 강화학습을 결합하여 지능형 자율 침투 테스트를 수행하는 시뮬레이션 및 에뮬레이션 환경을 개발하는 연구가 진행되고 있다. 본고에서는 대표적인 강화학습 기반 사이버 공격 시뮬레이션 환경인 NASim, CyberBattlesim 그리고 APRIL과 에뮬레이션 환경인 CyGIL과 PenGym뿐만 아니라 하이브리드 환경인 NASimEmu과 GAP의 특징과 한계를 분석하였다.

강화학습 기반 시뮬레이션 환경은 높은 학습 효율성을 제공하지만 현실성이 떨어진다. 반면, 에뮬레이션 환경은 높은 현실성을 제공하지만 시스템 자원 소모가 크고 학습 효율과 확장성이 떨어진다.

따라서 향후에는 시뮬레이션 환경과 에뮬레이션 환경을 결합한 하이브리드 환경이 요구되는 동시에 실제 환경과의 괴리를 최소화할 수 있는 고충실도의 취약 환경을 구성해야 한다. 이를 통해 강화학습 에이전트가 실제와 유사한 조건에서 안정적으로 학습하고 다양한 공격 시나리오에서도 일관된 성능을 유지할 수 있는 확장성과 일반화 능력을 얻을 수 있을 것으로 기대된다. 더 나아가 협력과 경쟁이 공존하는 사이버 공방을 재현하기 위해서는 멀티 에이전트 강화학습 기반의 연구가 지속적으로 이루어져야 한다.

용어해설

강화학습 에이전트가 환경과 상호작용을 하며 상태, 행동, 보상을 반복적으로 경험함으로써 누적 보상을 극대화하는 최적 정책을 학습하는 방법

MDP(Markov Decision Process) 에이전트가 완전한 환경 상태를 기반으로 상태 전이와 보상 함수를 통해 누적 보상을 최대화하는 최적 정책을 학습하는 강화학습의 기본 수학적 모델

PODMP(Partially Observable Markov Decision Process) 에이전트가 환경의 전체 상태를 직접 관측할 수 없는 상황에서 불완전한 관찰 정보를 통해 상태를 추정하고 의사결정을 수행하는 MDP의 확장 모델

일반화 훈련에 사용되지 않은 새로운 환경이나 시나리오에서도 학습된 정책이 안정적으로 성능을 유지하며 적용되는 능력

확장성 네트워크 규모, 호스트 수, 행동 공간 등 환경의 복잡도 증가에도 시스템이 성능 저하 없이 정상적으로 학습될 수 있는 능력

강건성 환경 변화나 불확실한 조건에서도 일관된 성능을 유지하는 시스템 또는 모델의 안정적 특성

참고문헌

[1] Ö. Aslan et al., "A comprehensive review of cyber security vulnerabilities, threats, attacks, and solutions," Electronics, vol. 12, no. 6, 2023, p. 1333.

[2] M.A.I. Mallick and R. Nath, "Navigating the cyber security landscape: A comprehensive review of cyber-attacks, emerging trends, and recent developments," World Scientific News, vol. 190, no. 1, 2024, pp. 1-69.

[3] T. Wilhelm, "Professional penetration testing: Creating and learning in a hacking lab," Elsevier, 2025.

[4] O. Sheyner et al., "Automated generation and analysis of attack graphs," in Proc. IEEE Symp. Security Privacy, (Berkeley, CA, USA), May. 2002, pp. 273-284.

[5] M. Aljabri et al., "Intelligent techniques for detecting network attacks: Review and research directions," Sensors, vol. 21, no. 21, 2021, p. 7070.

[6] T.T. Nguyen and V.J. Reddi, "Deep reinforcement learning for cyber security," IEEE Trans. Neural Netw. Learn. Syst., vol. 34, no. 8, 2021, pp. 3779-3795.

[7] B.S. Kim et al., "A Study of Reinforcement Learning-based Cyber Attack Prediction using Network Attack Simulator (NASim)," J. Semicond. Display Technol., vol. 22, no. 3, 2023, pp. 112-118.

[8] S.H. Oh et al., "Employing deep reinforcement learning to cyber-attack simulation for enhancing cybersecurity," Electronics, vol. 13, no. 3, 2024, p. 555.

[9] A. Velazquez et al., "Cyber Operations Gyms to Train Autonomous Cyber Defense Agents for NATO," in Proc. Int. Conf. Military Commun. Inf. Syst., (Oerias, Portugal), May. 2025, pp. 1-10.

[10] B.L. Semage, "Robust and Efficient Reinforcement Learning for Physics Tasks," Ph.D. dissertation, Deakin University, 2023.

[11] M.C. Ghanem et al., "Hierarchical reinforcement learning for efficient and effective automated penetration testing of large networks," J. Intell. Inf. Syst., vol. 60, no. 2, 2023, pp. 281-303.

[12] M.C. Ghanem, "Towards an efficient automation of network penetration testing using model-based reinforcement learning," Ph.D, dissertation, City, University of London, 2023.

[13] J. Schwartz and H. Kurniawati, "Autonomous penetration testing using reinforcement learning," arXiv preprint, 2019. doi: 10.48550/arXiv.1905.05965

[14] Microsoft Threat Intelligence, "Gamifying machine learning for stronger security and AI models," Microsoft Security Blog, 2021. 4. 8. https://www.microsoft.com/en-us/security/blog/2021/04/08/gamifying-machine-learning-for-stronger-security-and-ai-models/

[15] S. Zhou et al., "APRIL: towards scalable and transferable autonomous penetration testing in large action space via action embedding," IEEE Trans. Dependable Secure Comput., vol. 22, no. 3, 2024, pp. 2443-2459.

[16] Z. Hu et al., "Automated penetration testing using deep reinforcement learning," in Proc. IEEE Eur. Symp. Security Privacy Workshops, (Genoa, Italy), Sep. 2020, pp. 2-10.

[17] H.S. Anderson et al., "Learning to evade static PE machine learning malware models via reinforcement learning," arXiv preprint, 2018. doi: 10.48550/arXiv.1801.08917

[18] S.H. Oh et al., "Applying reinforcement learning for enhanced cybersecurity against adversarial simulation," Sensors, vol. 23, no. 6, 2023, p. 3000.

[19] B.S. Kim et al., "Optimal Cyber Attack Strategy Using Reinforcement Learning Based on Common Vulnerability Scoring

System," Comput. Model. Eng. Sci., vol. 141, no. 2, 2024.

[20] K. Tran et al., "Cascaded reinforcement learning agents for large action spaces in autonomous penetration testing," Appl. Sci., vol. 12, no. 21, 2022, p. 11265.

[21] M. Jeong et al., "Cyber Environment Test Framework for Simulating Command and Control Attack Methods with Reinforcment Learning," Appl. Sci., vol. 15, no. 4, 2025, p. 2120.

[22] J. Claypoole et al., "Interpreting Agent Behaviors in Reinforcment-Learning-Based Cyber-Battle Simulation Platforms," arXiv preprint, 2025. doi: 10.48550/arXiv.2506.08192

[23] L. Li et al., "Cygil: A cyber gym for training autonomous agents over emulated network systems," arXiv preprint, 2021. doi: 10.48550/arXiv.2109.03331

[24] H.P.T. Nguyen et al., "PenGym: Realistic training environment for reinforcement learning pentesting agents," Comput. Secur., vol. 148, 2025, p. 104140.

[25] U.U. Izuazu et al., "Explainable and perturbation-resilient model for cyber-threat detection in industrial control systems Networks," Discover Internet Things, vol. 5, no. 1, 2025, p. 9.

[26] A. Tantawy et al., "Model-based risk assessment for cyber physical systems security," Comput. Secur., vol. 96, 2020, p. 101864.

[27] J.D. Yoo et al., "Cyber attack and defense emulation agents," Appl. Sci., vol. 10, no. 6, 2020, p. 2140.

[28] A.F. Browne et al., "Development of an architecture for a cyber-physical emulation test range for network security testing," IEEE Access, vol. 6, 2018, pp. 73273-73279.

[29] Y. Chen et al., "Multiscale emulation technology based on the integration of virtualization, physical and simulation networks," in Proc. IEEE Int. Conf. Data Sci. Cyberspace, June 2019, (Hangzhou, China), pp. 396-402.

[30] J. Janisch et al., "NASimEmu: Network attack simulator & emulator for training agents generalizing to novel scenarios," in Proc. Eur. Symp. Res. Comput. Secur., (The Hague, The Netherlands), Sep. 2023, pp. 589-608.

[31] S. Zhou et al., "Mind the Gap: Towards Generalizable Autonomous Penetration Testing via Domain Randomization and Meta-Reinforcement Learning," arXiv prerpint, 2024. doi: 10.48550/arXiv.2412.04078