강화학습을 활용한 MITRE ATT&CK 기반 네트워크 공격 시뮬레이션 개발

김범석 [1], 김정현 [1], 구기종 [2], 김민석 [3]
[1] 상명대학교 전자정보시스템공학과
[2] 한국전자통신연구원 지능형네트워크보안연구실
[3] 상명대학교 휴먼지능로봇공학과

qjatod132@naver.com, minsuk.kim@smu.ac.kr

Development of Reinforcement Learning-Based Network Attack Simulation Using the MITRE ATT&CK Framework

Bum-Sok Kim[1], Jung-Hyun Kim[1], Min-Suk Kim[2], Ki-Jong Koo[3]
[1]Dept. of Electronic Information System Engineering, Sangmyung University
[2]Intelligent Convergence Research Laboratory, Electronics and Telecommunications Research Institute
[3]Dept. of Human Intelligence & Robot Engineering, Sangmyung University

요 약

본 논문에서는 강화학습을 활용하여 MITRE ATT&CK 프레임워크를 기반으로 한 사이버 공방 시뮬레이션 환경을 개발하였다. 본 논문의 목적은 실제와 유사한 네트워크 환경에서 방어자의 존재 여부가 공격 시퀀스의 복잡성과 학습 효율성에 미치는 영향을 분석하는 것이다. 본 논문에서 개발된 방어 전략은 공격 테크닉의 사전 및 사후 조건을 기반으로 하여 공격의 조건 및 결과와 반대되는 효과를 생성하였다. 이를 통해 공격자 및 방어자 간의 상호작용을 실시간으로 동기화하고, 네트워크 보안의 동적인 적응을 가능하게 한다. 이를 위해 Deep Q-Network (DQN), Proximal Policy Optimization (PPO), 그리고 Soft Actor-Critic (SAC) 등 다양한 강화학습 알고리즘을 사용하였다. 실험 결과, 방어자가 없는 시나리오에서는 알고리즘들이 간단한 시퀀스로 빠르게 수렴하는 경향을 보였다. 이와 달리, 방어자가 포함된 시나리오에서는 DQN 과 PPO 가 반복적 행동으로, SAC 는 보다 다양한 행동 패턴으로 최대 스텝 수에 이르렀다. 이러한 결과는 방어자의 존재가 사이버 공격 전략의 복잡성을 크게 증가시키며, 강화학습 기반의 사이버 보안 전략 개발에 중요한 역할을 한다.

1. 서론

최근 오늘날 디지털 기술의 급속한 발전으로 인해 사이버 보안의 중요성은 계속해서 증가하고 있다. 이러한 발전은 우리에게 많은 편리함을 제공하는 동시에 보안 위협의 범위와 복잡성을 확대하고 있다. 특히, 데이터 유출과 개인정보 침해 등의 사이버보안 문제가 심화되면서 전통적인 보안 조치만으로는 진화하는 사이버 위협에 대응하기 점차 어려워지고 있다. 이러한 상황에서 사이버 공격의 형태가 다양화되고 광범위해지며 개인, 기업, 국가 수준에서의 피해가 증가하고 있다. 그러나 실제 환경에서 보안 테스트는 제약과 위험을 수반하기 때문에 시뮬레이션을 통한 가상 환경에서의 사전 테스트는 필수적이다. 이에 따라 사이버 공방 시뮬레이션의 중요성이 강조되고 있다.

본 논문에서는 강화학습을 기반으로 한 사이버 공격 시뮬레이션 환경인 Network Attack Simulation (NASim)을 활용한다.[1] NASim 은 네트워크 공격 시나리오를 모델링하는데 사용되며, 주로 레드 팀(공격자)의 활동에 중점을 두지만, 본 논문에서는 방어 에이전트를 새롭게 통합하여 고정된 규칙을 따르는 공격자 에이전트와 동적으로 반응하는 방어 에이전트 간의 상호작용을 통해 보다 정교한 방어 전략의 효과를 평가하였다. 이러한 상호작용을 통해 레드 팀과 블루 팀(방어자)은 실시간으로 전략을 개발하고 적응하며 상대방의 움직임에 효과적으로 대응한다. 또한 본 논문에서는 실제와 유사한 사이버 공격 시나리오에서 MITRE ATT&CK 프레임워크를 기반으로 한 공

격 테크닉을 활용하고 실환경 적용을 위해 공격 테크닉의 사전 및 사후 상태 특징을 정의하였다. 이 과정에서 강화학습 알고리즘인 Deep Q-Network (DQN)[2], Proximal Policy Optimization (PPO)[3] 그리고 Soft Actor-Critic (SAC)[4]를 사용하여 시나리오에서의 성능을 비교 분석한다.

2. 관련연구

2.1 전통적인 사이버 보안 전략

초기의 사이버 보안 전략은 네트워크와 데이터를 보호하는 필수적인 수단으로, 주로 침입 탐지 시스템, 침입 방지 시스템, 그리고 안티바이러스 소프트웨어에 의존하였다. 이러한 시스템들은 알려진 위협의 서명을 감지하는 서명 기반 방법을 활용하여 네트워크의 보안을 유지하고자 하였다. 이 방법은 기존에 알려진 공격 패턴을 효과적으로 차단하는 데는 유용하였으나, 새롭게 변형된 공격 유형인 제로데이 공격에는 취약함을 드러냈다. 특히, 고도로 맞춤화된 멀웨어, 소셜 엔지니어링 공격, 그리고 분산 서비스 거부 공격과 같은 현대의 복잡한 공격 방법에 대해서는 더욱 취약한 모습을 보였다.

2.2 머신러닝을 활용한 사이버 보안 전략

사이버 보안 환경이 고도화됨에 따라 전통적인 사이버 보안 전략만으로는 새롭게 등장하는 위협들에 대해서 효과적으로 대응하기 어려워지고 있다. 이에 대응하여 지도학습, 비지도학습 그리고 강화학습을 포함한 머신러닝 기술을 융합하는 새로운 보안 전략이 주목받고 있다. 지도 학습은 레이블이 지정된 데이터를 사용하여 악성 행동과 정상 행동을 구분한다. 이 방법은 알려진 공격 패턴을 효과적으로 탐지하는 데 매우 유용하며, 이미 알려진 위협에 대해 높은 탐지 정확도를 제공한다. 그러나 지도 학습은 대량의 레이블이 지정된 데이터에 크게 의존하고 많은 시간과 비용을 요구한다. 또한 제로데이 공격과 같은 알려지지 않은 새로운 공격 유형에 대해서는 탐지 능력이 제한적일 수 있다. 비지도 학습은 레이블이 지정되지 않은 데이터를 분석하여 비정상적인 행동을 탐지하고 이상 징후를 식별한다. 이 방법은 지도 학습보다 더 넓은 범위의 데이터를 학습할 수 있고, 알려지지 않은 공격 유형을 탐지하는 데 유리하다. 그러나 이는 종종 높은 오경보율을 보이며, 정상적인 활동을 잘못된 이상 행동으로 분류할 수 있어 신뢰성 문제가 발생할 수 있다. 반면에, 강화학습은 동적인 사이버 보안 환경에서 환경과의 지속적인 상호작용을 통해 최적의 대응 전략을 학습할 수 있는 기술이다. 강화학습을 활용한 실시간 네트워크 시나리오 모델링은 복잡한 네트워크 환경에서 중요한 역할을 하며, 기존의 서명 기반 탐지 방식과 달리 새로운 유형의 위협에 효과적으로 대응할 수 있는 가능성을 제공한다.

3. 제안 방법

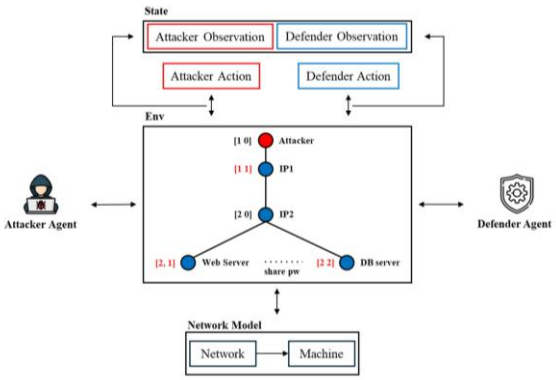
3.1 MITRE ATT&CK 프레임워크 기반의 사이버공방 시뮬레이션 설계

본 논문에서는 강화학습 기반의 사이버 공방 시뮬레이션을 통해 실제 네트워크와 유사한 사이버 전장을 설계하였다. 이 환경은 서브넷, 토폴로지 그리고 방화벽 등의 다양한 네트워크 구성요소와 운영체제, 서비스, 취약점, 파일 등의 다양한 호스트 자산으로 구성되어 있다. 이 시나리오의 목표는 네트워크 내에서 호스트와 서비스를 발견하고 발견된 취약점을 이용하여 주요 호스트를 장악하는 것이다. 이를 위해 MITRE ATT&CK 프레임워크를 기반으로 한 다양한 공격 테크닉을 적용하였다. 공격 테크닉은 호스트의 상태가 공격 테크닉의 사전 조건을 만족할 때 적용되며, 성공적으로 수행될 경우 사후 조건에 정의된 값으로 호스트 상태가 업데이트된다. 이러한 사전 조건과 사후 조건은 각 공격 테크닉의 성공 여부를 평가하는 데 중요한 기준으로 사용된다. 공격 테크닉의 사전 조건은 타겟 호스트의 공격 가능성을 평가하는 데 필요한 상태를 나타내며 사후 조건은 공격이 성공적으로 수행되었을 때 변화하는 타겟 호스트의 상태를 반영한다. 이러한 사전 조건 및 사후 조건의 정의는 제 환경에서 공격 테크닉의 효과를 정확하게 판단하고, 공격의 결과를 정확하게 반영하기 위해 필수적이다. 이를 통해 실제와 유사한 사이버 공방 환경에서 다양한 상황을 모의할 수 있다. 이는 사이버 보안 전략을 보다 효과적으로 수립하고 실행하는 데 기여한다.

3.2 강화학습 기반의 사이버 방어 전략

본 논문에서는 강화학습 기반의 사이버 방어 시스템 내에서 공격 기술과 1:1 로 대응하는 방어 기술을 개발하였다. 그림 1 은 시뮬레이션 환경 내에서 공격자와 방어자의 상호작용 과정이다. 이 그림은 각 에이전트의 관측과 행동 그리고 그에 따른 환경의 변화를 보여준다. 이러한 방어 기술들은 각 공격 기술의 사전 조건과 사후 조건 값을 기준으로 설계하였으며 공격의 조건 및 결과와 정반대될 수 있도록 구성하였다. 본 논문에서 개발된 방어 전략은 크게 두 가지 형태로 구현하였다. 첫 번째 전략인 사전 차단 전략은 공격이 호스트에 도달하기 전에 미리 감지하고 차단함으로써 중지시키는 기법이다. 이 전략은 네트워크의 전반적인 보안 상태, 중요 자산의 보안 수준, 그

리고 전체적인 보안 정책의 적용 상태를 고려하여 실행된다. 두 번째 전략인 추후 조치는 공격 후 발생할 수 있는 피해를 최소화하고 시스템을 빠르게 복구하는데 초점을 맞춘다. 이는 특정 서버의 보안 패치 적용 여부, 애플리케이션의 취약점 상태와 같은 구체적인 요소들을 포함하는 서브 상태를 기반으로 작동한다. 각각의 방어 기술들은 공격 테크닉과 정확히 매칭되어 공격이 성공적으로 수행되지 못하도록 설계되는 반대 효과를 생성한다. 예를 들어, 공격 기술이 특정 네트워크 취약점을 이용할 때 방어 기술은 이를 감지하고 해당 취약점을 보호하거나 공격을 중단시킬 수 있다. 본 논문에서는 공격 및 방어 기술이 적용될 때 메인 상태와 서브 상태 간의 동기화를 통해 네트워크 및 호스트의 상태 변화를 정확하게 관리하고 추적한다. 이 동기화 과정은 변경사항이 하위 시스템에 미치는 영향과 그 결과가 상위 시스템에 어떻게 반영되는지를 명확히 파악하는 데 중요한 역할을 한다. 즉 공격자와 방어자 간의 상호작용을 통해 발생하는 상태 변화를 실시간으로 동기화함으로써 각 에이전트 사이의 간극을 최소화하고 사이버 공방 시나리오를 보다 효과적으로 모의할 수 있다.

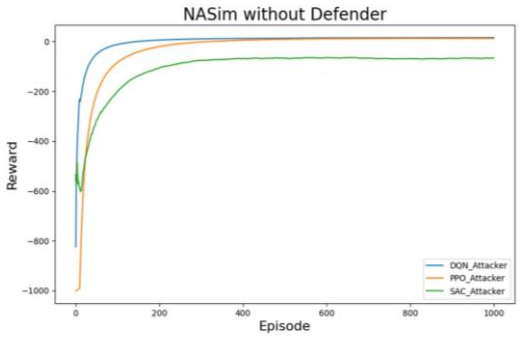


(그림 1) 사이버 공방 시뮬레이션에서의 공격자와 방어자의 상호작용

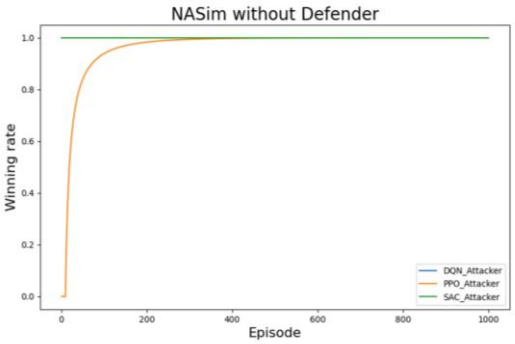
4. 실험 결과

4.1 Scenario without Defender

본 절에서는 방어자가 없는 시나리오에서 DQN, PPO, 그리고 SAC 강화학습 알고리즘을 사용하여 최적의 공격 시퀀스 생성하는 실험을 진행하였다. 이 실험을 통해 수동으로 정의된 행동 시퀀스와의 비교 분석도 진행하였다. 이때 기존의 수동 시퀀스는 13 개의 행동으로 구성되어 있으며, 이는 각 알고리즘의 비교 기준으로 사용되었다. 그림 2 와 3 은 각각 방어자가 없는 시나리오에서의 각 알고리즘별로 수렴하는 보상의 경향과 각 알고리즘의 승률을 나타낸다. 표 1 은 방어자의 유무에 따라 각 알고리즘의 공격 테크닉 시퀀스 개수를 보여준다. 표 1 에서 알 수 있듯이, 방어자가 없는 시나리오에서는 DQN 과 PPO 는 각각 15 와 14 의 행동으로 수동 시퀀스에 가까운 결과를 보여주는 반면, SAC 는 42 의 행동으로 상대적으로 긴 시퀀스를 생성하였다. 실험 결과, 각 알고리즘은 안정적으로 수렴하는 경향을 보였다. DQN 과 SAC 은 일부 행동을 다른 행동으로 대체하여 성공적으로 공격을 수행한 반면 PPO 는 수동 시퀀스와 유사한 형태로 학습이 완료되었다. 특히 방어자가 없는 시나리오에서는 수동 시퀀스와 비교했을 때 공격 시퀀스의 복잡성이 큰 변동 없이 상대적으로 간단한 모습을 보였다. 이러한 결과는 방어자의 유무가 공격 알고리즘의 행동 패턴과 복잡성에 미치는 영향을 보여준다.



(그림 2) 방어자가 없는 시나리오에서의 보상 변화

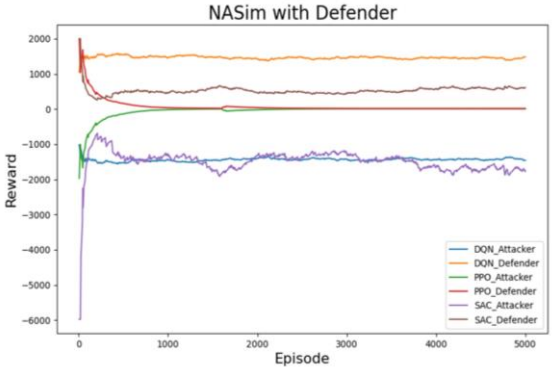


(그림 3) 방어자가 없는 시나리오에서의 승률 변화

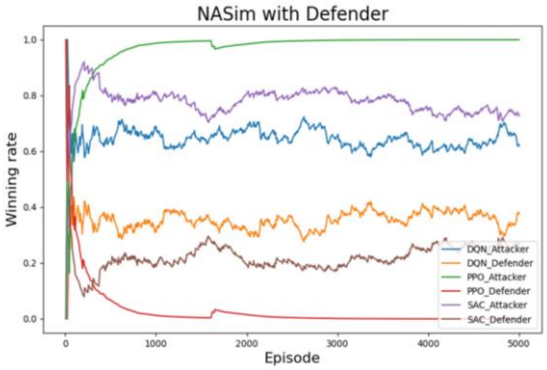
4.2 Scenario with Defender

본 절에서는 사이버 공방 시뮬레이션 환경에 방어자 에이전트를 통합하여 공격자와 방어자가 동시에 작동할 수 있도록 환경을 조정하고 실험을 수행하였다. 이 과정에서 먼저 공격자를 학습시킨 후, 학습된 공격자를 기반으로 방어자를 학습을 진행하였다. 그림 4 와 그림 5 는 각각 방어자가 있는 시나리오에서 각 알고리즘의 보상 값 변화와 각 알고리즘의 승률을 나타낸다. 이를 통해 공격자와 방어자 간의 상호작용과 방어자의 개입에 따른 공격자 자신의 전략 조정 과정을 관찰하였다. 특히, 방어 기법의 존재 유무에

따라 각각의 알고리즘에서 생성된 공격 테크닉 시퀀스의 복잡성이 어떻게 변화하는지 분석하였다. 표 1 에서 볼 수 있듯이, 방어자가 포함된 시나리오에서는 각 알고리즘은 반복적이고 다양한 공격 패턴을 통해 공격 시퀀스를 생성하여 공격자의 행동에 대응하려는 경향을 보였다. DQN 은 일부 행동을 반복하여 최대 스텝 수에 도달하였으며, PPO 는 한가지 액션만을 반복하는 경향을 보였다. 이와 달리, SAC 는 다양한 액션을 반복하며 최대 스텝인 1000 스텝에 도달하였다. 이를 통해 방어 기법이 공격 행위에 미치는 영향 입증하고 각 알고리즘의 적응력과 효율성을 평가하였다. 이 과정에서 공격자와 방어자는 각각 다른 상태 정보를 기반으로 학습을 진행하며 공격자와 방어자 사이의 상태 정보가 일치하도록 매 단계마다 동기화를 진행하였다. 이를 통해 두 에이전트는 실시간으로 변화하는 상태 정보를 공유하며 상호 작용하는 동안의 학습 진행 상황을 파악할 수 있다. 실험 결과, Value-Based 알고리즘인 DQN 이 가장 안정적으로 수렴하는 모습을 보였다. 반면, Policy-Based 알고리즘인 PPO 와 SAC 는 DQN 에 비해 상대적으로 낮은 보상으로 수렴하는 경향을 보였다. 이러한 실험을 통해 방어자의 존재가 알고리즘의 성능에 미치는 영향을 명확히 보여주며 공격자와 방어자의 상호작용이 각 알고리즘 성능에 어떻게 영향을 미치는지를 입증한다.



(그림 4) 방어자가 존재하는 시나리오에서의 보상 변화



(그림 5) 방어자가 존재하는 시나리오에서의 승률 변화

<표 1> 방어자 유무에 따른 생성된 공격 시퀀스 개수

Scenario \ Algorithm	DQN	PPO	SAC
Manual without Defender	15	14	42
Manual with Defender	1000	1000	1000

5. 결론

본 논문에서는 강화학습 기반의 사이버 공방 시뮬레이션 환경을 활용하여 실제와 유사한 사이버 위협 시나리오를 구현하였다. MITRE ATT&CK 프레임워크 기반의 공격 테크닉들을 적용하여 방어자의 유무에 따른 공격 및 방어 전략의 효과와 각 알고리즘의 성능 변화를 평가하였다. 본 논문에서 개발된 방어 전략은 공격 테크닉의 사전 조건과 사후 조건을 기준으로 설계되었으며, 이를 통해 공격의 조건 및 결과와 정반대되는 효과를 생성하였다. 이러한 방어 기술은 네트워크의 보안 상태와 중요 자산의 보안 수준을 고려하여 사전 차단 및 추후 조치 전략을 실행함으로써 효과적인 방어 메커니즘을 제공한다. 실험 결과, 방어자가 없는 시나리오에서는 알고리즘들이 상대적으로 간단한 시퀀스를 통해 빠르게 수렴하였다, 반면, 방어자가 통합된 시나리오에서는 DQN 과 PPO 가 반복적인 행동을 통해 최대 스텝에 도달하였으며, SAC 는 다양한 행동을 반복함으로써 최대 스텝까지 도달하여 복잡한 공격 시퀀스를 생성하였다. 이러한 실험들을 통해 강화학습 기반의 동적 방어 메커니즘이 다양한 보안 위협에 신속하고 효과적으로 대응할 수 있음을 보여주었다. 추후 연구에서는 더욱 다양한 공방 시나리오를 통합하고 진화된 학습 알고리즘을 적용하여 사이버 보안의 효율성을 극대화하고자 한다.

사사문구

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

참고문헌

[1] Jjschwartz, “NetworkAttackSimulator,” 2019, [Online]. Available: https://github.com/Jjschwartz/NetworkAttackSimulator.

[2] Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D. et al., "Playing Atari with deep reinforcement learning," Proc. NIPS Deep Learning Workshop, pp. 5-10, 2013.

[3] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O., "Proximal Policy Optimization Algorithms," arXiv preprint arXiv:1707.06347, 2017.

[4] Haarnoja, T., Zhou, A., Abbeel, P., Levine, S., "Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor," Proceedings of the 35th International Conference on Machine Learning (ICML), pp. 1861-1870, 2018.