신경망 및 학습 시스템에 대한 IEEE 트랜잭션, DOI: 10.1109/TNNLS.2021.3121870, 조기 액세스

사이버 보안을 위한 심층 강화 학습

탄 티 응우옌(Thanh Thi Nguyen)과 비자이 자나파 레디(Vijay Janapa Reddi)

개요 - 인터넷에 연결된 시스템의 규모가 상당히 증가했으며 이러한 시스템은 그 어느 때보다 사이버 공격에 노출되고 있습니다. 사이버 공격의 복잡성과 역학으로 인해 대응력, 적응력, 확장성이 뛰어난 보호 메커니즘이 필요합니다. 이러한 문제를 해결하기 위해 기계 학습, 더 구체적으로 DRL(심층 강화 학습) 방법이 널리 제안되었습니다. 기존 RL에 딥 러닝을 통합함으로써 DRL은 복잡하고 역동적이며 특히 고차원적인 사이버 방어 문제를 해결할 수 있는 능력이 뛰어납니다. 이 문서에서는 사이버 보안을 위해 개발된 DRL 접근 방식에 대한 조사를 제시합니다. 사이버 물리 시스템을 위한 DRL 기반 보안 방법, 자율 침입 탐지 기술, 사이버 공격에 대한 방어 전략을 위한 다중 에이전트 DRL 기반 게임 이론 시뮬레이션 등 다양한 중요한 측면을 다룹니다. DRL 기반 사이버 보안에 대한 폭넓은 논의와 향후 연구 방향도 제시된다. 우리는 이 포괄적인 검토가 점점 더 복잡해지는 사이버 보안 문제에 대처하기 위해 새로운 DRL의 잠재력을 탐구하는 향후 연구의 기반을 제공하고 촉진할 것으로 기대합니다.

색인 용어 - 설문 조사, 검토, 심층 강화 학습, 딥 러닝, 사이버 보안, 사이버 방어, 사이버 공격, 사물 인터넷, IoT.

I. 소개

사물인터넷(IoT)은 지능형 기술에 활용되고 있습니다.
통신, 물류, 의료, 교육, 금융, 정부, 엔터테인먼트 등 다양한 분야에서 광범위하게 활용되고 있습니다. IoT에 다양한 정보통신기술(ICT) 도구가 융합되면서 IoT의 기능과 서비스가 사용자에게 새로운 수준으로 향상되었습니다. ICT는 지난 10년 동안 시스템 설계, 네트워크 아키텍처, 지능형 장치 측면에서 눈부신 발전을 이루었습니다. 예를 들어 인지 무선 네트워크와 5G 셀룰러 네트워크[1],[2], 소프트웨어 정의 네트워크(SDN)[3], 클라우드 컴퓨팅[4], (모바일) 엣지 캐싱 등의 혁신으로 ICT가 발전했습니다. [5], [6] 및 포그 컴퓨팅 [7].

이러한 발전과 함께 사이버 공격에 대한 취약성이 증가하고 있습니다. 사이버 공격은 컴퓨터 정보 시스템, 네트워크 인프라 또는 개인용 컴퓨터 장치를 표적으로 삼기 위해 하나 이상의 컴퓨터가 실행하는 모든 유형의 공격 기동으로 정의됩니다. 사이버 공격은 경제적 경쟁업체나 국가 후원 공격자에 의해 시작될 수 있습니다. 따라서 이러한 공격의 영향을 완화하고 제거하기 위한 사이버 보안 기술 개발이 절실히 필요합니다[8].

TT Nguyen은 Deakin University, Melbourne Burwood Campus, Burwood, VIC 3125, Australia의 정보 기술 학교에 재직하고 있습니다. 이메일: thanh.nguyen@deakin.edu.au.

VJ Reddi는 Harvard University, Cambridge, MA 02138, USA의 John A. Paulson School of Engineering and Applied Sciences에 재직하고 있습니다. 이메일: vj@eecs.harvard.edu.

인공지능(AI), 특히 머신러닝(ML)은 사이버 공간의 공격과 방어 모두에 적용됐다. 공격자 측에서는 ML을 활용하여 방어 전략을 손상시킵니다. 사이버 보안 측면에서는 ML을 사용하여 보안 위협에 대한 강력한 저항력을 발휘하여 발생하는 영향이나 피해를 적응적으로 예방하고 최소화합니다. 이러한 ML 애플리케이션 중에서 비지도 학습 방법과 지도 학습 방법은 침입 탐지[9]-[11], 악성 코드 탐지[12]-[14], 사이버 물리 공격[15]-[17], 데이터 개인 정보 보호 [18].

원칙적으로 비지도 방법은 레이블을 사용하지 않고 데이터의 구조와 패턴을 탐색하는 반면, 지도 방법은 데이터 레이블을 기반으로 한 예제를 통해 학습합니다. 그러나 이러한 방법으로는 사이버 공격, 특히 새로운 위협이나 지속적으로 진화하는 위협에 대해 동적이고 순차적인 대응을 제공할 수 없습니다. 또한, 공격 이후에 공격의 흔적을 수집, 분석하여 탐지 및 방어 대응이 이뤄지는 경우가 많아 선제적 방어 솔루션이 저해되는 경우가 많습니다. 통계적 연구에 따르면 공격의 62%가 사이버 시스템에 심각한 피해를 입힌 후에 인지된 것으로 나타났다[19].

ML의 한 갈래인 강화 학습(RL)은 미지의 환경을 탐색하고 활용하여 자신의 경험을 통해 학습할 수 있기 때문에 인간 학습과 가장 가까운 형태입니다. RL은 환경에 대한 제한된 사전 지식 없이 또는 제한된 사전 지식 없이 최적으로 순차적 조치를 취하도록 자율 에이전트를 모델링할 수 있으므로 실시간 및 적대적 환경에서 특히 적응 가능하고 유용합니다.

함수 근사화 및 표현 학습의 힘으로 딥러닝은 RL 방법에 통합되어 많은 복잡한 문제를 해결할 수 있게 되었습니다[20]-[24]. 따라서 딥 러닝과 RL의 결합은 사이버 공격이 점점 더 정교하고 빠르며 유비쿼터스화되는 사이버 보안 애플리케이션에 탁월한 적합성을 나타냅니다[25]-[28].

DRL의 출현은 실제로 비디오 게임 영역(예: Atari [29], [30]), Go 게임 [31], [32], 실시간 전략 게임 StarCraft II [33] 등 다양한 분야에서 큰 성공을 거두었습니다. [36], 3D 멀티 플레이어 게임 Quake III Arena Capture the Flag [37] 및 팀워크 게임 Dota 2 [38]를 로봇 공학 [39], 자율 차량 [40], 자율 수술 [41]과 같은 실제 응용 프로그램에 적용 , [42], 자연어 처리 [43], 생물학적 데이터 마이닝 [44] 및 약물 설계 [45]. 최근에는 IoT 분야의 다양한 문제를 해결하기 위해 DRL 기법도 적용되고 있다. 예를 들어, 스마트 시티 애플리케이션을 위한 네트워킹, 캐싱 및 컴퓨팅 기능을 통합하는 DRL 기반 자원 할당 프레임워크가 [46]에서 제안되었습니다. DRL 알고리즘, 즉 Double Dueling Deep Q-network[47,48]는 기지국의 동적 변화 상태, 모바일 에지 캐싱(MEC)으로 구성된 대규모 상태 공간을 포함하기 때문에 이 문제를 해결하는 데 사용됩니다. 서버

신경망 및 학습 시스템에 대한 IEEE 트랜잭션, DOI: 10.1109/TNNLS.2021.3121870, 조기 액세스

그리고 캐시. 프레임워크는 SDN의 프로그래밍 가능한 제어 원리와 정보 중심 네트워킹의 캐싱 기능을 기반으로 개발되었습니다. 또는 Zhu et al. [49]에서는 사용자의 상황 정보와 트래픽 패턴 통계를 나타내는 상황 인식 개념을 이용하여 MEC 정책을 탐색하였다. 모바일 네트워크 에지에서 AI 기술을 사용하면 운영 환경을 지능적으로 활용하고 적절한 콘텐츠를 무엇을, 어디서, 어떻게 캐시할지에 대한 올바른 결정을 내릴 수 있습니다. 캐싱 성능을 높이기 위해 DRL 접근 방식, 즉 비동기식 장점 행위자-비평가 알고리즘[50]을 사용하여 오프로딩 트래픽 최대화를 목표로 하는 최적의 정책을 찾습니다.

현재 조사 결과에 따르면 사이버 환경에서 DRL 적용은 일반적으로 IoT 애플리케이션의 통신 및 네트워킹 기능 최적화 및 강화(예: [51]-[59])와 사이버 공격 방어라는 두 가지 관점으로 분류됩니다. 이 문서에서는 사이버 공격이나 위협이 있을 때 사이버 보안 문제를 해결하기 위해 DRL 방법을 사용하는 나중에 초점을 맞춥니다. 다음 섹션에서는 DRL 방법의 배경을 제공하고 섹션 III에서는 사이버 보안에 DRL 적용에 대한 자세한 조사를 진행합니다. 우리는 이러한 애플리케이션을 사이버 물리 시스템을 위한 DRL 기반 보안 솔루션, 자율 침입 탐지 기술, 사이버 보안을 위한 DRL 기반 게임 이론을 포함하여 세 가지 주요 범주로 분류합니다.

섹션 IV에서는 사이버 보안을 위한 DRL에 대한 광범위한 논의와 향후 연구 방향으로 논문을 마무리합니다.

II. 심층 강화 학습 예비

ML의 다른 인기 있는 분야, 즉 예시를 통해 학습하는 지도 방법과 달리 RL은 환경과 직접 상호 작용하여 자체 학습 경험을 생성하여 에이전트의 특성을 지정합니다. RL은 상태, 행동, 보상의 개념으로 설명됩니다(그림 1). 이는 에이전트가 두 가지 변화를 일으키는 각 시간 단계에서 조치를 취하는 시행착오 접근 방식입니다. 즉, 환경의 현재 상태가 새로운 상태로 변경되고 에이전트가 환경으로부터 보상 또는 페널티를 받습니다. 상태가 주어지면 보상은 에이전트에게 행동이 얼마나 좋은지 나쁜지 알려줄 수 있는 함수입니다.

에이전트는 받은 보상을 바탕으로 좋은 행동을 더 많이 취하고 나쁜 행동을 점차 걸러내는 방법을 학습합니다.

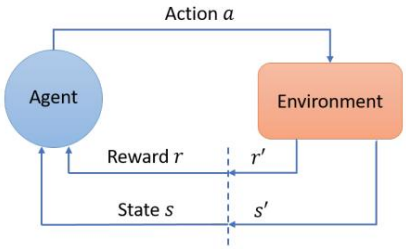

그림 1. 상태, 행동, 보상으로 특징지어지는 RL의 에이전트와 환경 간의 상호 작용. 현재 상태 s와 보상 r을 기반으로 에이전트는 최적의 조치를 취하여 상태와 보상이 변경됩니다. 그런 다음 에이전트는 다음 작업을 결정하기 위해 환경으로부터 다음 상태 s와 보상 r을 수신하여 에이전트-환경 상호 작용의 반복 프로세스를 만듭니다.

널리 사용되는 RL 방법은 Q-learning으로, 그 목표는 할인된 누적 보상을 기반으로 최대화하는 것입니다.
벨만 방정식 [60]:

$$Q(st, at) = E[rt+1 + \gamma rt+2 + \gamma^{2} {}_{rt+3+\ldots|st, at}] \quad (1)$$

할인 요인 γ ∈ [0, 1]은 미래 보상의 중요도 수준을 관리합니다. 학습 융합을 분석하기 위한 수학적 트릭으로 적용됩니다. 실제로 확률론적 환경의 부분적인 관찰 가능성이나 불확실성으로 인해 할인이 필요합니다.

Q-학습은 일련의 상태에 따라 행동의 예상 보상(Q-값)을 저장하기 위해 조회 테이블 또는 Q-테이블을 사용해야 합니다.
이는 상태 공간과 행동 공간이 증가할 때 큰 메모리를 필요로 합니다. 실제 문제에는 연속적인 상태나 행동 공간이 포함되는 경우가 많으므로 Q-learning은 이러한 문제를 해결하는 데 비효율적입니다. 다행스럽게도 딥러닝은 기존 RL 기술을 훌륭하게 보완하는 강력한 도구로 등장했습니다. 딥러닝 방법에는 두 가지 일반적인 기능, 즉 함수 근사화와 표현 학습이 있는데, 이는 원시 고차원 데이터의 컴팩트한 저차원 표현을 효과적으로 학습하는 데 도움이 됩니다[61].

딥러닝과 RL의 결합은 구글 딥마인드가 시작하고 개척한 연구 방향이었습니다.

그들은 Q-학습이 고차원 감각 입력을 처리할 수 있도록 심층 신경망(DNN)을 사용하여 심층 Q-네트워크(DQN)를 제안했습니다[29], [62].

그러나 DNN을 사용하여 Q 함수를 근사화하는 것은 관측 시퀀스 간의 상관관계와 Q 값 Q(s, a)와 대상 값 Q(s , a) 간의 상관 관계로 인해 불안정합니다. Mnihet al. [29]는 이 문제를 해결하기 위해 두 가지 새로운 기술, 즉 경험 재생 메모리와 대상 네트워크의 사용을 제안했습니다(그림 2). 한편으로, 경험 메모리는 에이전트와 환경의 상호 작용에서 얻은 학습 경험 튜플(s, a, r, s)의 광범위한 목록을 저장합니다. 에이전트의 학습 프로세스는 연속적인 경험의 상관 관계를 피하기 위해 이러한 경험을 무작위로 검색합니다. 반면, 타겟 네트워크는 기술적으로 추정 네트워크의 복사본이며,

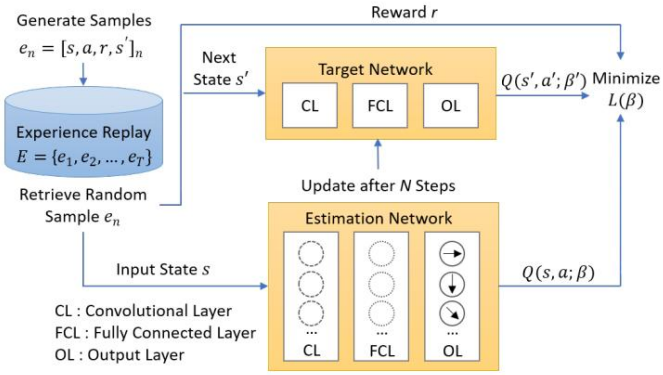

그림 2. L(β) = E[(r+ a |β) Q(s, a|β))2]로 설명되는 손실 함수를 사용하는 DQN 및 대상 심층 신 아키텍처 여기서 β와 β는 γ maxa Q(s)의 매개변수 입니다 . , 추정 경망 각각 에이전트가 수행한 각 작업은 현재 상태, 작업 a, 보상 r 및 다음 상태로 구성된 경험을 생성합니다. 이러한 학습 경험(샘플)은 경험 재생 메모리에 저장됩니다. , 안정적인 학습 과정을 위해 무작위로 검색됩니다.

신경망 및 학습 시스템에 대한 IEEE 트랜잭션, DOI: 10.1109/TNNLS.2021.3121870, 조기 액세스

하지만 해당 매개변수는 고정되어 있으며 일정 기간이 지난 후에만 업데이트됩니다. 예를 들어, 목표 네트워크는 [29]에 설명된 것처럼 추정 네트워크가 10,000번 업데이트된 후에 업데이트됩니다. DQN은 RL 에이전트가 게임 보드의 원시 이미지 픽셀만 사용하여 49개의 Atari 게임을 플레이하면서 인간 수준의 성능을 제공할 수 있는 최초의 획기적인 발전을 이루었습니다.

DQN은 가치 기반 방법으로 훈련 시간이 오래 걸리고 연속적인 행동 공간으로 문제를 해결하는 데 한계가 있습니다. 일반적으로 가치 기반 방법은 Q-값 함수를 사용하여 상태에 따른 행동의 장점을 평가합니다. 상태나 동작의 수가 크거나 무한할 경우 비효율적이거나 심지어 비실용성을 나타냅니다. 또 다른 유형의 RL, 즉 정책 그라데이션 방법이 이 문제를 효과적으로 해결했습니다.

이러한 방법은 가능한 모든 행동에 대한 확률 분포인 정책 π(s, a)를 학습하여 직접 행동을 도출하는 것을 목표로 합니다. REINFORCE[64], 바닐라 정책 그라데이션[65], 신뢰 영역 정책 최적화(TRPO)[66] 및 근접 정책 최적화(PPO)[67]는 주목할만한 정책 그라데이션 방법입니다.

그러나 기울기 추정은 종종 큰 변동으로 인해 어려움을 겪습니다 [68]. 가치 기반 방법과 정책 구배 방법의 조합은 두 가지 방법의 장점을 통합하고 단점을 근절하기 위해 개발되었습니다.

이러한 종류의 조합은 또 다른 유형의 RL, 즉 배우-비평 방법을 구성했습니다. 이 구조는 두 가지 구성 요소, 즉 DNN으로 특징지어질 수 있는 행위자와 비평가로 구성됩니다. 배우는 비평가로부터 피드백을 받아 정책을 배우려고 시도합니다. 이러한 반복 프로세스는 행위자가 전략을 개선하고 최적의 정책으로 수렴하는 데 도움이 됩니다.

DDPG(Deep deterministic Policy Gradient)[69], D4PG(Distributed Distribution DDPG)[70], A3C(Asynchronous Advantage Actor-Critic)[50], UNREAL(Unsupervised Reinforcement and Auxiliary Learning)[71]은 행위자를 활용하는 방법입니다. -비평 프레임워크. 널리 사용되는 알고리즘 A3C의 예시적인 아키텍처가 그림 3에 나와 있습니다. A3C의 구조는 마스터 학습 에이전트(글로벌)와 개별 학습자(작업자)의 계층 구조로 구성됩니다. 마스터 에이전트와 개별 학습자 모두 DNN으로 모델링되며 각각 두 개의 출력(비평가용 출력과 행위자용 출력)이 있습니다. 첫 번째 출력은 주어진 상태 V(s)의 예상 보상을 나타내는 스칼라 값이고, 두 번째 출력은 가능한 모든 행동 π(s, a)에 대한 확률 분포를 나타내는 값의 벡터입니다.

비평가의 가치 손실 함수는 다음과 같이 지정됩니다.

$$L1 = (R \quad V(초))2 \tag{2}$$

여기서 R = r + γV(s)는 할인된 미래 보상입니다. 또한 행위자는 다음과 같은 정책 손실 함수의 최소화를 추구하고 있습니다.

$$L2 = \quad \log(\pi(a|s)) * A(s) \qquad H(\pi) \tag{삼}$$

여기서 A(s) = R V(s)는 추정된 이점 함수이고, H(π)는 엔트로피 정규화의 강도를 제어하는 하이퍼파라미터 를 사용하여 에이전트의 탐색 능력을 처리하는 엔트로피 항입니다. 이점 함수 A(s)는 에이전트가 특정 상태에 있을 때 얼마나 유리한지를 보여줍니다. A3C의 학습 프로세스는 각 학습자가 별도의 환경과 상호 작용하고 마스터 네트워크를 독립적으로 업데이트하기 때문에 비동기식입니다.

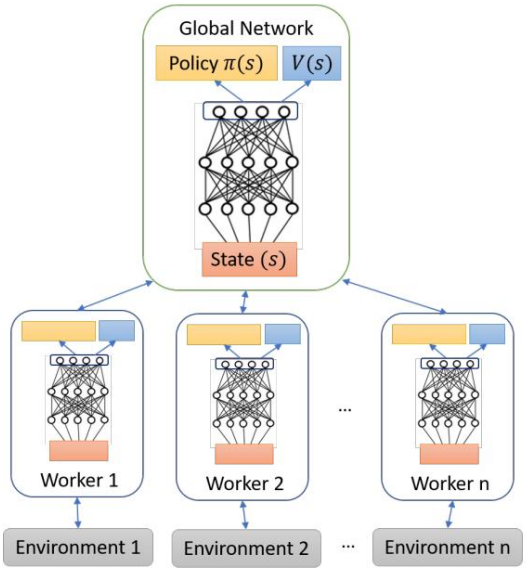

그림 3. 글로벌 네트워크와 다수의 작업자 에이전트로 구성된 A3C의 학습 아키텍처. 각 작업자는 처음에 매개변수를 글로벌 네트워크의 매개변수로 재설정하고 학습 환경 사본과 상호작용합니다. 이러한 개별 학습 프로세스에서 얻은 그라데이션은 글로벌 네트워크를 비동기적으로 업데이트하는 데 사용됩니다. 이는 개별 작업자 에이전트가 얻은 경험이 독립적이기 때문에 학습 속도를 높이고 글로벌 네트워크에서 학습하는 경험을 다양화합니다.

이 프로세스는 반복되며 학습이 완료되면 마스터 네트워크가 사용됩니다.

표 1에는 가치 기반, 정책 그라데이션 및 행위자 비판 방법의 비교 가능한 기능과 일반적인 예제 알고리즘이 요약되어 있습니다. 가치 기반 방법은 전문가와 같은 다른 소스의 데이터를 활용할 수 있기 때문에 정책 그라데이션 방법보다 표본 효율성이 더 높습니다[72]. DRL에서 가치 함수 또는 정책 함수는 일반적으로 이산 또는 연속 상태를 입력으로 사용할 수 있는 (심층) 신경망과 같은 범용 함수 근사기로 근사화됩니다. 따라서 상태 공간 모델링은 DRL의 작업 공간을 처리하는 것보다 더 간단합니다. 값 기반 방법은 모든 작업을 명시적으로 평가하고 이러한 평가를 기반으로 각 시간 단계에서 작업을 선택하므로 이산 작업 공간 문제에 적합합니다. 반면, 정책 그라데이션 및 행위자 비평 방법은 정책(상태와 동작 간의 매핑)을 동작에 대한 확률 분포로 설명하기 때문에 연속 동작 공간에 더 적합합니다. 연속성 특성은 이산적 행동 공간과 연속적 행동 공간의 주요 차이점입니다. 개별 행동 공간에서 행동은 상호 배타적인 옵션 집합으로 특징지어지는 반면, 연속 행동 공간에서는 행동이 특정 범위나 경계의 값을 갖습니다[73].

III. 사이버 보안 의 DRL : 설문 조사

데이터 프라이버시에서 중요한 인프라 보호에 이르기까지 사이버 보안의 다양한 측면에 대한 RL의 수많은 적용이 문헌에서 제안되었습니다. 그러나 기존 RL의 단점으로 인해 복잡하고 대규모의 사이버 보안 문제를 해결하는 능력이 제한되었습니다.

최근 몇 년 동안 연결된 IoT 장치의 수가 증가함에 따라 사이버 범죄의 수가 크게 증가했습니다.

신경망 및 학습 시스템에 대한 IEEE 트랜잭션, DOI: 10.1109/TNNLS.2021.3121870, 조기 액세스

표 I

DRL 유형 의 특징 과 주목할만한 방법 요약

DRL 유형	가치 기반 - 상태	정책 그라데이션	행위자-비평가
특징	Q(s, a)가 주어지면 행동의 가치를 계산합니다. - 학습된 명시적 정책이 없습니다. - 효율적인 샘플 [72].	- 가치 기능이 필요하지 않습니다. - 명시적 정책이 구축됩니다. - 표본이 비효율적입니다 [72].	- 행위자는 정책 π(s, a)를 생성합니다. - 평론가는 V(s)로 액션을 평가한다. - 종종 가치 기반 또는 정책 그라데이션 방법보다 더 나은 성능을 발휘합니다.
일반적인 방법	- DQN [29] - 이중 DQN [47] - 결투 Q-네트워크 [48] - 우선 경험 재생 DQN [63]	- 강화 [64] - 바닐라 폴리시 그라데이션 [65] - TRPO [66] - PPO [67]	- DDPG [69] - D4PG [70] - A3C [50] - 언리얼 [71]
애플리케이션 널리 사용되는 OpenAI Gym 툴킷[74]에 설명 및 구현된 Acrobot, CartPole 및 MountainCar와 같은 고전적인 제어 작업과 같은 개별 작업 공간 문제에 적합합니다.		OpenAI Gym 툴킷(MountainCarContinuous and Pendulum[74] 또는 BipedalWalker 및 CarRacing 문제[75])에 설명 및 구현된 고전적인 제어 작업과 같은 연속 작업 공간 문제에 더 적합합니다.	

공격 인스턴스와 그 복잡성. 딥 러닝의 출현과 RL과의 통합으로 사이버 물리 시스템에 대한 위조된 데이터 주입[86], 자율 시스템에 대한 속임수 공격과 같은 정교한 유형의 사이버 공격을 탐지하고 이에 맞서 싸울 수 있는 DRL 방법 클래스가 탄생했습니다. [93], 분산 서비스 거부 공격 [114], 호스트 컴퓨터 또는 네트워크에 대한 침입 [125], 재밍 [142], 스푸핑 [157], 악성 코드 [161], 적대적인 네트워킹 환경에서의 공격 [168] 등이 있습니다. 이 섹션에서는 사이버 물리 시스템에 대한 방어 방법부터 자율 침입 탐지 접근 방식, 게임 이론 기반 솔루션에 이르기까지 사이버 보안을 위한 최첨단 DRL 기반 솔루션에 대한 포괄적인 조사를 제공합니다.

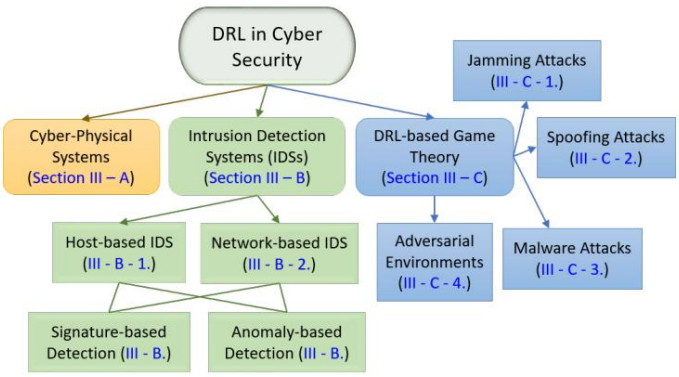

그림 4. 사이버 보안 분야 DRL에 관한 설문조사의 다양한 (하위)섹션.

설문 조사의 구조는 그림 4에 나와 있습니다. 우리는 설문 조사를 사이버 보안에 대한 DRL의 기존 적용으로 제한합니다. 사이버 보안에는 DRL이 적용되지 않은 다른 주제가 있으므로 섹션 IV(논의 및 향후 연구 방향)에서 논의됩니다.

이러한 잠재적 주제에는 사이버 보안에 대한 다중 에이전트 DRL 접근 방식, 호스트 기반 및 네트워크 기반 침입 탐지 시스템 결합, 모델 기반 DRL, 사이버 보안 애플리케이션을 위한 모델 프리 및 모델 기반 DRL 방법 결합, 다음을 수행할 수 있는 방법 조사가 포함됩니다. 사이버 환경의 지속적인 행동 공간, 공격적인 AI, 딥페이크, 기계 학습 중독, 적대적 기계 학습, Human-On-The-Loop 아키텍처 내의 인간-기계 팀 구성, 비트 앤 피스 분산 서비스 거부 처리 암호화 알고리즘을 해독하기 위한 양자 물리학 기반의 강력한 컴퓨터에 의한 공격뿐만 아니라 잠재적인 공격도 가능합니다.

A. 사이버 물리 시스템을 위한 DRL 기반 보안 방법 사이버 공격에 대한 사이버

물리 시스템(CPS)의 방어 방법에 대한 조사는 사이버 보안 연구계로부터 상당한 관심과 관심을 받아왔습니다. CPS는 인터넷 통합을 통해 촉진되는 컴퓨터 기반 알고리즘에 의해 제어되는 메커니즘입니다. 이 메커니즘은 공유 네트워크를 통해 분산된 물리적 시스템을 효율적으로 관리합니다. 인터넷과 제어 기술의 급속한 발전으로 CPS는 제조[76], 건강 모니터링[77],[78], 스마트 그리드[79]-[81], 교통[79]-[81] 등 다양한 분야에서 광범위하게 사용되었습니다. 82], [83]. 인터넷에 널리 노출되면서 이러한 시스템은 사이버 공격에 점점 더 취약해지고 있습니다[84]. 2015년 해커들은 피싱 이메일을 통해 로그인 자격 증명을 획득하여 독일의 한 제철소 제어 시스템을 공격했습니다.

이 공격으로 인해 공장이 부분적으로 폐쇄되고 수백만 달러의 피해가 발생했습니다. 마찬가지로, 2015년 12월 말 우크라이나의 전력망에 대한 비용이 많이 드는 사이버 공격이 발생하여 수십만 명의 최종 소비자에 대한 전력 공급이 중단되었습니다[85].

CPS에 대한 사이버 공격을 연구하려는 노력의 일환으로 Feng et al. [85] 수학적 모델을 통해 사이버 상태 역학을 특성화했습니다.

$$x \cdot (t) = f(t, x, u, w; \theta(t, a, d)); x(t0) = x0 \quad (4)$$

여기서 x, u 및 w는 물리적 상태, 제어 입력 및 이에 따른 외란을 나타냅니다(그림 5 참조). 또한 θ(t, a, d)는 시점 t에서의 사이버 상태를 나타내며 a와 d는 각각 사이버 공격과 방어를 나타냅니다.

그런 다음 CPS 방어 문제는 각 시간 단계에서 플레이어의 효용을 0으로 합산하는 2인 제로섬 게임으로 모델링됩니다. 방어자는 배우 평론가 DRL 알고리즘으로 표현됩니다. 시뮬레이션 결과는 [85]에서 제안한 방법이 알려지지 않은 사이버 공격으로부터 CPS를 적시에 정확하게 방어하기 위한 최적의 전략을 학습할 수 있음을 보여줍니다.

자율 자동차, 화학 공정, 자동 조종 항공 전자 공학 및 스마트 그리드와 같은 중요한 안전 영역에 CPS를 적용하려면 특정 정확성 수준이 필요합니다.

Akazakiet al. [86]은 CPS 모델에 대한 위조된 입력(반례)을 찾기 위해 DRL, 즉 이중 DQN 및 A3C 알고리즘의 사용을 제안했습니다. 이를 통해 CPS 결함을 효과적이고 자동으로 감지할 수 있습니다. 무한한 상태로 인해

신경망 및 학습 시스템에 대한 IEEE 트랜잭션, DOI: 10.1109/TNNLS.2021.3121870, 조기 액세스

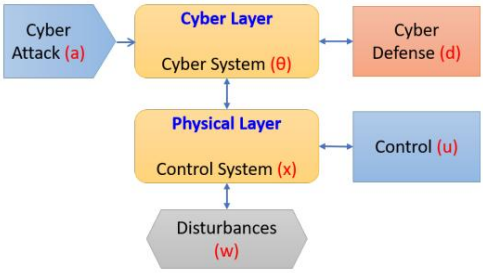

그림 5. 사이버 물리 시스템의 공격 및 방어 역학.
물리적 계층은 교란으로 인해 종종 불확실한 반면, 사이버 공격은 방어 전략 d를 구현해야 하는 사이버 계층에 직접적인 영향을 미칩니다. θ(t, a, d)로 특징지어지는 공격-방어 역학은 기존 물리 시스템에 주입되어 Eq.에 제시된 사이버-물리 공동 모델링을 개발합니다. (4).

CPS 모델의 공간에서는 시뮬레이션된 어닐링[87] 및 교차 엔트로피[88]와 같은 기존 방법이 비효율적인 것으로 나타났습니다.
실험 결과는 더 적은 수의 시뮬레이션 실행 측면에서 이러한 방법에 비해 DRL 알고리즘 사용의 우월성을 보여줍니다. 이는 CPS의 소프트웨어 및 물리적 시스템이 매우 복잡함에도 불구하고 CPS 모델의 결함에 대한 보다 실용적인 감지 프로세스로 이어집니다.

미래의 스마트 시티에서 작동하는 자율주행차(AV)에는 카메라, 레이더, 도로변 스마트 센서, 차량 간 비커닝 등 차량 내 센서의 강력한 처리 장치가 필요합니다. 이러한 의존은 감각 데이터를 조작하고 시스템의 신뢰성에 영향을 주어 사고 위험을 높이거나 차량 흐름을 줄이는 등 AV를 제어하려는 사이버 물리적 공격에 취약합니다.

Ferdowsiet al. [89]는 공격자가 AV의 센서 판독값에 잘못된 데이터를 삽입하는 반면 AV(방어자)는 AV를 강력하게 제어하기 위해 해당 문제를 처리해야 하는 시나리오를 조사했습니다. 구체적으로, 다른 자동차를 바짝 따라가는 자동차의 자율 제어에 초점을 맞춘 자동차 추종 모델[90]이 고려된다. 방어자는 센서 판독값을 기반으로 선두 차량의 속도를 학습하는 것을 목표로 합니다.

공격자의 목표는 뒤따르는 차량이 최적의 안전 간격에서 벗어나도록 유도하는 것입니다. 공격자와 방어자 사이의 상호작용은 게임이론적 문제를 특징으로 합니다. 대화형 게임 구조와 DRL 솔루션은 그림 6에 도식화되어 있습니다. 혼합 전략 내쉬 균형 분석을 기반으로 솔루션을 직접 도출하는 대신 저자는 이 동적 게임을 해결하기 위해 DRL을 사용할 것을 제안했습니다. 장단기 기억(LSTM)[91]은 환경의 시간적 역학을 포착할 수 있으므로 방어 에이전트와 공격 에이전트 모두에 대한 Q 함수를 근사화하는 데 사용됩니다.

마찬가지로, Rasheed et al. [92]는 5G 통신 링크가 장착된 자율주행차에서 데이터 주입 공격에 대처하기 위해 LSTM과 GAN(Generative Adversarial Network) 모델을 통합한 적대적 DRL 기법을 소개했다. 공격자는 자율주행차 사이의 안전 거리 간격에 영향을 미치기 위해 잘못된 데이터를 주입하려고 시도하는 반면 자율주행차는 이러한 편차를 최소화합니다.

LSTM은 생성기로 사용되는 반면 CNN(Convolutional Neural Network)은 판별기로 사용되어 이전 시간적 동작을 캡처할 수 있는 GAN 구조와 유사합니다.

자율주행차와 공격자의 이전 거리 편차는 물론, 이러한 관찰을 바탕으로 자율주행차가 충돌과 사고를 피하기 위한 최적의 동작(적절한 속도)을 선택하는 DRL 알고리즘을 제안합니다.

자율 시스템은 통신 채널의 소음, 센서 오류, 센서 측정 판독 오류, 패킷 오류, 특히 사이버 공격과 같은 다양한 소스로 인해 비효율성에 취약할 수 있습니다. 자율 시스템에 대한기만 공격은 센서와 명령 센터 사이의 통신 채널에 소음을 주입하려는 공격자에 의해 시작되므로 널리 퍼져 있습니다. 이러한 종류의 공격은 손상된 정보를 명령 센터로 전송하여 결국 시스템 성능을 저하시킵니다. Gupta와 Yang [93]은 시스템이 적대적인 예를 사용하여 학습할 수 있도록 하여 자율 시스템의 견고성을 높이는 방법을 연구했습니다. 문제는 플레이어가 지휘 센터(관찰자)가 되고 적이 되는 제로섬 게임으로 공식화됩니다. Roboschool[67]의 역진자 문제가 시뮬레이션 환경으로 사용됩니다.

TRPO 알고리즘은 측정 손상 측면에서 적대적 공격을 안정적으로 감지하고 그 영향을 자동으로 완화할 수 있는 관찰자를 설계하는 데 사용됩니다.

B. DRL 기반 침입 탐지 시스템 침입을 탐지하기 위해 보

안 전문가는 일반적으로 애플리케이션 추적, 네트워크 트래픽 흐름, 사용자 명령 데이터 등의 감사 데이터를 관찰하고 검사하여 정상 동작과 비정상 동작을 구별해야 합니다. 그러나 네트워크 규모가 커지면 감사 데이터의 양도 급격히 늘어난다.

이로 인해 수동 감지가 어렵거나 불가능해집니다. 침입 탐지 시스템(IDS)은 호스트 컴퓨터나 네트워크 장비에 설치되어 감사 데이터를 분석하여 비정상적이거나 악의적인 활동을 탐지하고 관리자에게 보고하는 소프트웨어 또는 하드웨어 플랫폼입니다. 침입 탐지 및 예방 시스템은 악의적인 활동의 영향을 줄이기 위해 즉시 적절한 조치를 취할 수 있습니다.

다양한 유형의 감사 데이터에 따라 IDS는 호스트 기반 IDS와 네트워크 기반 IDS라는 두 가지 범주로 분류됩니다. 호스트 기반 IDS는 일반적으로 호스트 컴퓨터의 로그 파일이나 설정을 관찰하고 분석하여 비정상적인 동작을 발견합니다. 네트워크 기반 IDS는 스니퍼를 사용하여 네트워크에서 전송 패킷을 수집하고 트래픽을 검사합니다.

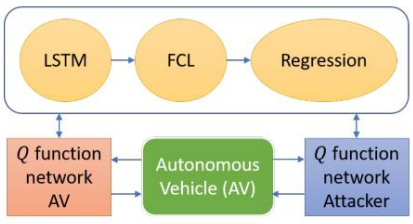

그림 6. 강력한 AV 제어를 위한 적대적 DRL 알고리즘의 아키텍처. LSTM(장단기 메모리), FCL(완전 연결 계층) 및 회귀로 구성된 심층 신경망(DNN)은 플레이어의 결과가 포함된 대규모 데이터 세트 내에서 장기 종속성을 학습하는 데 사용됩니다. 과거 상호작용. DNN은 플레이어, 즉 AV(방어자), 특히 AV 센서 판독값에 잘못된 데이터를 주입하려는 공격자에 대한 최적의 조치를 찾기 위해 Q 함수를 근사화할 수 있습니다.

신경망 및 학습 시스템에 대한 IEEE 트랜잭션, DOI: 10.1109/TNNLS.2021.3121870, 조기 액세스

침입탐지 데이터입니다. 호스트 기반 시스템은 일반적으로 크로스 플랫폼 지원이 부족하며 이를 구현하려면 호스트 운영 체제 및 구성에 대한 지식이 필요합니다.

네트워크 기반 시스템은 특정 네트워크 세그먼트의 트래픽을 모니터링하는 것을 목표로 하며 운영 체제에 독립적이며 호스트 기반 시스템보다 이식성이 뛰어납니다. 따라서 네트워크 기반 시스템을 구현하는 것이 더 쉽고 호스트 기반 시스템보다 더 많은 모니터링 기능을 제공합니다. 그러나 네트워크 기반 IDS는 네트워크 세그먼트를 통과하는 모든 패킷을 검사해야 하기 때문에 트래픽이 많고 고속 네트워크를 처리하는 데 어려움을 겪을 수 있습니다.

IDS 유형에 관계없이 시그니처 기반 탐지와 이상 징후 기반 탐지라는 두 가지 일반적인 탐지 방법이 사용됩니다.
시그니처 탐지에는 알려진 공격의 패턴을 저장하고 가능한 공격의 특성을 데이터베이스의 특성과 비교하는 작업이 포함됩니다. 이상 탐지는 시스템의 정상적인 동작을 관찰하고, 예상치 못한 트래픽 속도 증가(예: 초당 IP 패킷 수)와 같이 정상에서 벗어난 활동이 발견되면 관리자에게 경고합니다. 비지도 클러스터링 및 지도 분류 방법을 포함한 기계 학습 기술은 적응형 IDS를 구축하는 데 널리 사용되었습니다[94]-[98]. 이러한 방법은 예를 들어 신경망 [99], k-최근접 이웃 [99], [100], 지원 벡터 머신(SVM) [99], [101], 랜덤 포레스트 [102] 및 최근 딥 러닝 [103]입니다. 그러나 일반적으로 기존 사이버 공격의 고정된 기능에 의존하므로 새로운 공격이나 변형된 공격을 탐지하는 데 부족합니다.

동적 침입에 대한 즉각적인 대응이 부족하면 감독되지 않거나 감독되는 기술로 인해 비효율적인 솔루션이 생성될 수도 있습니다. 이와 관련하여 RL 방법은 다양한 IDS 응용에서 효과적으로 입증되었습니다 [106].

다음 하위 섹션에서는 호스트 기반 및 네트워크 기반 IDS 모두에서 DRL 방법의 사용을 검토합니다.

1) 호스트 기반 IDS: 감사 데이터의 양과 침입 행위의 복잡성이 증가함에 따라 적응형 침입 탐지 모델은 그 효율성이 제한적입니다.

일시적으로 격리된 레이블이 있거나 레이블이 없는 데이터만 처리할 수 있습니다. 실제로 많은 복잡한 침입은 동적 행동의 시간적 순서로 구성됩니다. Xu와 Xie[107]는 이 문제를 처리할 수 있는 RL 기반 IDS를 제안했습니다. 시스템 호출 추적 데이터는 상태 값을 사용하여 호스트 프로세스의 비정상적인 시간적 동작을 감지할 수 있는 Markov 보상 프로세스에 공급됩니다. 따라서 침입 탐지 문제는 Markov 체인의 상태 값 예측 작업으로 변환됩니다.

상태값 예측 모델로는 선형 시간차(TD) RL 알고리즘[108]이 사용되며, 그 결과는 미리 정해진 임계값과 비교되어 정상적인 추적과 공격 추적을 구별합니다. TD 학습 알고리즘은 실제 값과 추정 값 사이의 오류를 사용하는 대신 연속 근사치의 차이를 사용하여 상태 값 함수를 업데이트합니다. 시스템 호출 추적 데이터를 사용하여 얻은 실험 결과는 제안된 RL 기반 IDS가 SVM, 은닉 마르코프 모델 및 기타 기계 학습 또는 데이터 마이닝 방법에 비해 더 높은 정확도와 더 낮은 계산 비용 측면에서 우월함을 보여줍니다. 그러나 선형 기반 함수를 기반으로 제안된 방법은 순차적 침입 동작이 매우 비선형적인 경우 단점이 있습니다. 따라서 커널 기반 RL 접근 방식은 다음을 사용합니다.

최소 제곱 TD [109]는 [110], [111]에서 침입 탐지를 위해 제안되었습니다. 커널 방법을 사용하여 TD RL의 일반화 기능이 특히 고차원 및 비선형 특징 공간에서 향상되었습니다. 따라서 커널 최소제곱TD 알고리즘은 이상 확률을 정확하게 예측할 수 있어 특히 다단계 사이버 공격을 처리할 때 IDS의 탐지 성능을 향상시키는 데 도움이 됩니다.

2) 네트워크 기반 IDS: Deokar와 Hazarnis [112]는 이상 기반 탐지 방법과 시그니처 기반 탐지 방법 모두의 단점을 지적했습니다. 한편, 이상 탐지는 사용자가 거의 수행하지 않는 활동을 이상으로 분류할 수 있기 때문에 잘못된 경보 비율이 높습니다. 반면, 시그니처 탐지는 잘 알려진 공격 패턴의 데이터베이스를 사용하기 때문에 새로운 유형의 공격을 발견할 수 없습니다. 이에 저자는 로그 파일을 활용해 이상 징후 탐지와 시그니처 탐지 기능을 결합해 알려진 공격과 알려지지 않은 공격을 효과적으로 식별할 수 있는 IDS를 제안했다.

제안된 IDS는 RL 방법, 연관 규칙 학습 및 로그 상관 기술의 협력을 기반으로 합니다. RL은 이상 징후나 공격 징후가 포함된(또는 포함되지 않은) 로그 파일을 선택할 때 시스템에 보상(또는 페널티)을 제공합니다. 이 절차를 통해 시스템은 공격 추적을 검색할 때 보다 적절한 로그 파일을 선택할 수 있습니다.

현재 인터넷이 직면한 가장 어려운 과제 중 하나는 DDoS(Distributed Denial of Service) 위협에 대한 대처입니다. 이는 DoS 공격이지만 분산적 성격을 갖고 있어 대용량 트래픽에서 발생하며 다수의 데이터를 침해하는 공격입니다. 호스트의. Mialalis와 Kudenko [113], [114]는 처음에 SARSA 알고리즘 [115]을 기반으로 하는 다중 에이전트 라우터 조절 방법을 도입하여 여러 에이전트를 학습하여 피해자 서버에 대한 트래픽의 속도를 제한하거나 조절함으로써 DDoS 공격을 해결했습니다. 그러나 이 방법은 확장성 측면에서 제한된 기능을 가지고 있습니다. 따라서 그들은 언급된 단점을 제거하기 위해 분할 정복 패러다임을 기반으로 하는 원래의 다중 에이전트 라우터 제한에 대한 조정된 팀 학습 설계를 추가로 제안했습니다. 제안된 접근 방식은 작업 분해, 계층적 팀 기반 통신, 팀 보상이라는 세 가지 메커니즘을 통합하여 DDoS 공격의 홍수를 막거나 줄이기 위해 서로 다른 위치에 있는 여러 방어 노드를 포함합니다. Yau 등의 연구를 기반으로 네트워크 에뮬레이터가 개발되었습니다. 제한 접근 방식을 평가합니다. 시뮬레이션 결과는 서로 다른 공격 역학을 갖는 다양한 시나리오에서 제안된 방법의 복원력과 적응성이 경쟁 방법인 기본 라우터 조절 및 가산 증가/증가 감소 조절 알고리즘[116]보다 우수하다는 것을 보여줍니다. 제안된 방법의 확장성은 최대 100개의 RL 에이전트를 사용하여 성공적으로 실험되었으며, 이는 대규모 인터넷 서비스 제공자 네트워크에 배포할 수 있는 큰 잠재력을 가지고 있습니다.

또는 Bhosale et al. [117]은 복잡한 공격에 대해 빠른 대응이 가능하도록 RL과 Influence Diagram[119]을 활용한 다중 에이전트 지능형 시스템[118]을 제안했다. 각 에이전트는 로컬 데이터베이스와 다른 에이전트로부터 받은 정보(예: 결정 및 이벤트)를 기반으로 정책을 학습합니다.

Shamshirbandet al. [120]에서는 무선센서를 이용한 침입탐지 및 방지 시스템을 출시하였다.

신경망 및 학습 시스템에 대한 IEEE 트랜잭션, DOI: 10.1109/TNNLS.2021.3121870, 조기 액세스

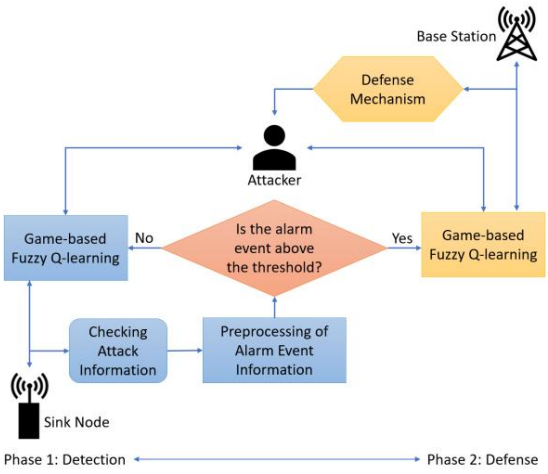

그림 7. 게임 이론 접근법과 퍼지 Q-러닝을 기반으로 한 2단계 침입 탐지 및 예방 시스템. 1단계에서 싱크 노드는 퍼지 Q-학습을 사용하여 공격자가 피해자 노드에 발생하는 이상 현상을 탐지합니다. 악성 정보는 싱크 노드에 의해 전처리되고 임계값에 대해 검사된 후 기지국이 퍼지 Q-학습을 사용하여 최적의 방어 조치를 선택하는 2단계로 전달됩니다.

게임 이론 접근 방식을 기반으로 하는 네트워크(WSN)를 개발하고 퍼지 Q-학습 알고리즘[121], [122]을 사용하여 플레이어를 위한 최적의 정책을 얻었습니다. 싱크 노드, 기지국 및 공격자는 특히 애플리케이션 계층에서 DDoS 공격에 대한 방어 전략을 도출하기 위해 싱크 노드와 기지국이 조정되는 3인 게임을 구성합니다.

IDS는 탐지와 방어의 두 단계에 참여하는 퍼지 Q-학습 알고리즘을 기반으로 향후 공격을 탐지합니다(그림 7). 게임은 공격자가 WSN의 피해자 노드에 DDoS 공격으로 특정 임계값을 초과하는 압도적인 양의 플러딩 패킷을 보낼 때 시작됩니다.

대표적인 WSN 프로토콜인 LEACH(Low Energy Adaptive Clustering Hierarchy)를 사용하여 제안하는 방법의 성능을 평가하고 기존 소프트 컴퓨팅 방법과 비교한다. 결과는 감지 정확도, 에너지 소비 및 네트워크 수명 측면에서 제안된 방법의 효율성과 실행 가능성을 보여줍니다.

또 다른 접근법에서는 Caminero et al. [124]는 네트워크 침입 탐지를 위한 분류기를 구현하는 데 RL 이론을 통합하기 위해 RL을 사용하는 적대적 환경이라는 모델을 제안했습니다. 레이블이 지정된 네트워크 침입 데이터 세트에서 추출한 무작위 샘플이 RL 상태로 처리되는 시뮬레이션 환경이 생성됩니다. 적대적 전략은 오버샘플링 메커니즘을 통해 훈련 편향을 피하고 과소 대표 클래스의 분류 오류를 줄이는 데 도움이 되므로 불균형 데이터 세트를 처리하는 데 사용됩니다.

마찬가지로 [125]의 연구에서는 네트워크 침입 탐지를 위해 DQN, 이중 DQN(DDQN), 정책 기울기 및 행위자 비판 모델과 같은 DRL 방법을 적용했습니다. 여러 가지 조정과 적응을 통해 DRL 알고리즘은 레이블이 지정된 침입 데이터를 분류하는 감독 방식으로 사용할 수 있습니다. DRL 정책 네트워크는 간단하고 빠르며, 진화하는 환경에 따른 현대 데이터 네트워크의 온라인 학습 및 신속한 대응에 적합합니다. 두 개의 침입 탐지 데이터 세트에서 얻은 결과는 DDQN이 사용된 네 가지 DRL 알고리즘 중에서 가장 좋은 알고리즘임을 보여줍니다. DDQN의 성과 어떤 경우에는 기존의 많은 기계 학습 방법과 동일하고 훨씬 더 좋습니다. 최근 Saeed et al. [126]은 RL 알고리즘을 활용하는 여러 접근 방식을 포함하여 기존 다중 에이전트 IDS 아키텍처를 조사했습니다. RL 방법의 적응 기능은 IDS가 환경 변화에 효과적으로 대응하는 데 도움이 될 수 있습니다. 그러나 다중 에이전트 시스템의 융합이 어렵기 때문에 최적의 솔루션이 보장되지는 않습니다.

다. 사이버보안을 위한 DRL 기반 게임이론

방화벽, 바이러스 백신 소프트웨어 또는 침입 탐지와 같은 전통적인 사이버 보안 방법은 일반적으로 수동적이고 일방적이며 동적 공격에 뒤처집니다. 사이버 공간은 다양한 사이버 구성요소를 포함하므로 안정적인 사이버 보안을 위해서는 이러한 구성요소 간의 상호 작용에 대한 고려가 필요합니다. 특히, 구성 요소에 적용되는 보안 정책은 다른 구성 요소가 내리는 결정에 일정한 영향을 미칩니다. 따라서 시스템 규모가 클 경우 가정 시나리오가 많아 결정 공간이 상당히 늘어납니다. 게임이론은 많은 시나리오를 검토하여 각 플레이어에게 가장 적합한 정책을 도출할 수 있기 때문에 이러한 대규모 문제를 해결하는 데 효과적으로 입증되었습니다[127]-[131]. 게임 플레이어의 효용이나 보상은 자신의 행동뿐만 아니라 다른 플레이어의 활동에 따라 달라집니다. 즉, 사이버 방어 전략의 효율성은 공격자의 전략과 다른 네트워크 사용자의 행동을 고려해야 합니다. 게임 이론은 공격자와 방어자가 관련된 사이버 보안 문제의 활동과 유사한 지능적인 의사 결정자 간의 갈등과 협력을 모델링할 수 있습니다. 이러한 유사성은 게임 이론이 여러 경쟁 메커니즘의 복잡한 행동을 수학적으로 설명하고 분석할 수 있게 해주었습니다. 다음에서는 방해 전파, 스푸핑, 악성 코드 및 적대적 환경에서의 공격을 포함한 다양한 공격 시나리오에서 사이버 보안 문제를 특성화하는 여러 DRL 에이전트가 포함된 게임 이론 모델을 제시합니다.

1) 재밍(Jamming) 공격: 재밍 공격은 DoS 공격의 특별한 경우로 간주될 수 있으며, 이는 예상되는 기능을 수행하는 네트워크의 용량을 감소시키거나 근절하는 모든 이벤트로 정의됩니다[132]-[134]. 방해 전파는 네트워킹에 있어서 심각한 공격이며 이 문제를 해결하기 위해 기계 학습이나 특히 RL을 사용한 연구자들의 큰 관심을 끌었습니다(예: [135]-[141]). 최근 딥 러닝의 발전으로 재밍 처리 또는 완화를 위해 DRL을 사용하는 것이 쉬워졌습니다. Xiaoet al. [142]은 MEC 시스템의 보안 문제를 연구하고 재밍 공격에 대비하여 에지 노드에 안전한 오프로드를 제공하는 RL 기반 솔루션을 제안했습니다. MEC는 클라우드 컴퓨팅 기능이 셀룰러 네트워크 또는 일반적으로 모든 네트워크의 에지 노드에서 발생하도록 하는 기술입니다. 이 기술은 사용자가 셀룰러 고객에게 더 가까운 엣지에 캐시된 콘텐츠에 액세스하도록 요청할 때 네트워크 트래픽을 줄이고 오버헤드와 대기 시간을 줄이는 데 도움이 됩니다. 그러나 MEC 시스템은 클라우드 서버나 데이터베이스 센터에 비해 프로토콜 보안이 덜하여 사용자와 공격자에게 물리적으로 더 가깝기 때문에 사이버 공격에 취약합니다. [142]에서는 RL 방법론을 사용하여 방어 수준을 선택하고 중요

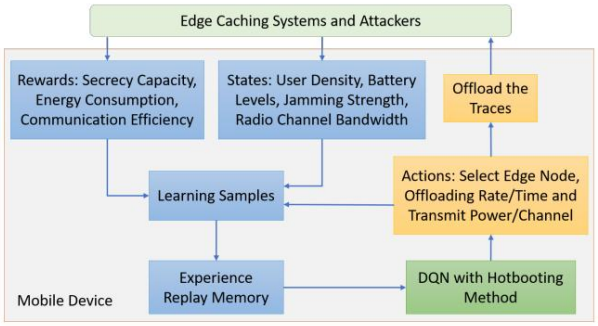

그림 8. 핫부팅 기술을 사용한 DQN 기반 MEC의 안전한 오프로딩 방법. DQN 에이전트의 작업은 모바일 장치가 그에 따라 트레이스를 에지 노드로 오프로드하기 위한 오프로딩 속도, 전력 및 채널과 같은 최적의 매개변수를 찾는 것입니다. 공격자는 이 프로세스를 방해하기 위해 전파 방해, 스푸핑, DoS 또는 스마트 공격을 배포할 수 있습니다. 엣지 캐싱 시스템과 상호 작용함으로써 에이전트는 이전 작업의 보상을 평가하고 새로운 상태를 획득하여 다음 최적의 작업을 선택할 수 있습니다.

오프로딩 속도 및 시간, 전송 채널 및 전력과 같은 매개변수. 네트워크 상태 공간이 크기 때문에 저자는 그림 8과 같이 고차원 데이터를 처리하기 위해 DQN 사용을 제안했습니다. DQN은 높은 계산 복잡성과 메모리가 필요한 Q 함수를 근사화하기 위해 CNN을 사용합니다. 이러한 단점을 완화하기 위해 핫부팅 기술이라는 전이 학습 방법이 사용됩니다. 핫부팅 방법은 유사한 상황에서 학습된 경험을 활용하여 CNN의 가중치를 보다 효율적으로 초기화하는 데 도움이 됩니다. 이렇게 하면 각 에피소드가 시작될 때 무작위 탐색을 방지하여 학습 시간이 줄어듭니다. 시뮬레이션 결과는 제안된 방법이 MEC 시스템의 보안 및 사용자 개인정보 보호 측면에서 효과적이며 낮은 오버헤드로 다양한 유형의 스마트 공격에 맞서 시스템을 보호할 수 있음을 보여줍니다.

반면 Aref et al. [143]은 WACR(광대역 자율 인지 무선 장치)에서 방해 전파 방지 통신을 처리하기 위해 다중 에이전트 RL 방법을 도입했습니다. WACR은 무선 상태를 감지할 수 있는 고급 무선 장치입니다.

주파수 스펙트럼 및 네트워크를 파악하고 인지된 상태에 따라 작동 모드를 자동으로 최적화합니다.
그러나 인지 통신 프로토콜은 의도하지 않은 간섭이나 고의적인 전파 방해로 안정적인 통신을 방해하려는 악의적인 사용자가 있는 경우 어려움을 겪을 수 있습니다. 각 무선 장치의 노력은 사용 가능한 공통 광대역 스펙트럼을 최대한 점유하고 전체 스펙트럼 대역에 영향을 미치는 방해 전파의 스위핑 신호를 피하는 것입니다. [143]에서 제안된 다중 에이전트 RL 접근 방식은 신호 방해 및 다른 무선 장치의 방해를 방지하는 것을 목표로 각 무선 장치에 대한 최적의 정책을 학습하여 적절한 하위 대역을 선택합니다. 비교 연구는 제안된 방법이 무작위 정책에 비해 상당한 우위를 보인다는 것을 보여줍니다. 제안된 방법의 단점은 재머가 인지 무선 기술을 사용하여 적응형 재밍을 수행할 수 있지만 재머가 WACR 전략에 응답할 때 고정된 전략을 사용한다는 가정입니다. [144]에서는 현재 스펙트럼 하위 대역이 재머에 의해 간섭을 받는 경우 Q-learning을 사용하여 가능한 한 오랫동안 중단 없이 전송이 가능한 새로운 하위 대역을 최적으로 선택합니다. Q-learning 에이전트의 보상 구조는 다음과 같이 정의됩니다.

방해 전파 또는 간섭 요인이 WACR 전송을 방해하는 데 걸리는 시간입니다. Hardware-in-the-loop 프로토타입 시뮬레이션을 사용한 실험 결과는 에이전트가 재밍 패턴을 감지하고 재밍 방지를 위한 최적의 하위 대역 선택 정책을 성공적으로 학습할 수 있음을 보여줍니다. 이 방법의 명백한 단점은 제한된 수의 환경 상태로 Q-table을 사용한다는 것입니다.

스펙트럼(또는 보다 일반적으로 리소스)에 대한 액세스 권한은 CRN(인지 무선 네트워크)과 기존 무선 기술 간의 주요 차이점입니다. 일반적으로 RL 또는 Q-learning은 인지 무선 노드가 무선 주파수 환경과 상호 작용하는 최적의 정책을 생성하기 위해 조사되었습니다[145]. Attaret al. [146]은 두 CRN 아키텍처, 즉 인프라 기반(예: IEEE 802.22 표준) 및 인프라 없는(예: 임시 CRN)에 대한 공격에 대해 RL 솔루션을 조사했습니다. 공격자는 스펙트럼 감지 프로세스를 조작하여 인프라가 없는 CRN에서 보안 위협의 주요 소스를 발생시키려고 시도할 수 있습니다. 외부 적대 노드는 CRN의 일부가 아니지만 이러한 공격자는 재밍 공격을 통해 임시 CRN의 작동에 영향을 미칠 수 있습니다. 인프라 기반 CRN에서는 외인성 공격자가 기존 에뮬레이션을 탑재하거나 센서 전파 방해 공격을 수행할 수 있습니다. 공격자는 특정 대역의 가용성에 대한 IEEE 802.22 기지국의 결정에 영향을 미치기 위해 로컬 허위 경보 확률을 높일 수 있습니다. 재밍 공격은 단기 및 장기 효과를 모두 가질 수 있습니다. 왕 외. [147]은 각 무선 장치가 사용 가능한 채널의 상태와 품질을 관찰하고 그에 따라 결정을 내리는 방해 전파 전략을 관찰하는 CRN의 전파 방해에 맞서 싸우기 위한 게임 이론적 프레임워크를 개발했습니다. CRN은 minimax-Q 학습 정책[148]을 사용하여 최적의 채널 활용 전략을 학습할 수 있으며, 채널 전환 전략과 함께 데이터에 사용할 채널 수와 패킷을 제어하는 문제를 해결할 수 있습니다. 스펙트럼 효율적인 처리량을 통해 표현되는 minimax-Q 학습의 성능은 즉각적인 보상에 우선 순위를 두고 환경 역학과 공격자의 인지 능력을 무시하는 근시 학습 방법보다 우수합니다.

CRN에서 2차 사용자(SU)는 1차 사용자(PU)의 통신 중단을 피해야 할 의무가 있으며 허가된 스펙트럼은 PU가 점유하지 않는 경우에만 액세스할 수 있습니다. SU의 기회주의적 접근과 SU의 전송 전략을 탐지할 수 있는 스마트 방해 전파의 출현으로 인해 CRN에서 재밍 공격이 발생합니다.

Xiaoet al. [149] 스마트 방해 전파가 PU가 아닌 SU를 방해하는 것을 목표로 하는 시나리오를 연구했습니다. 따라서 SU와 방해 전파는 결정을 내리기 전에 PU의 존재를 확인하기 위해 채널을 감지해야 합니다. 구성된 시나리오는 데이터 패킷을 보조 수신 노드로 전송하기 위해 릴레이 노드가 지원하는 보조 소스 노드로 구성됩니다. 스마트 재머는 SU의 주파수와 전송 전력을 신속하게 학습할 수 있지만 SU는 기본 동적 환경에 대한 완전한 지식을 갖고 있지 않습니다. SU와 방해 전파 사이의 상호 작용은 협력적인 전송 전력 제어 게임으로 모델링되었으며 SU의 최적 전략은 Stackelberg 평형을 기반으로 도출되었습니다 [150]. SU 플레이어의 목적은 전파 방해가 있는 경우 데이터 메시지를 효율적으로 보내기 위해 적절한 전송 전력을 선택하는 것입니다.

신경망 및 학습 시스템에 대한 IEEE 트랜잭션, DOI: 10.1109/TNNLS.2021.3121870, 조기 액세스

공격. 방해 전파의 유틸리티 이득은 SU의 손실이고 그 반대도 마찬가지입니다. RL 방법, 즉 Q-learning[60] 및 WoLF-PHC[151]은 스마트 방해기에 대처하기 위한 지능형 에이전트로 SU를 모델링하는 데 사용됩니다. WoLF-PHC는 Win 또는 Learn Fast 알고리즘과 정책 언덕 오르기 방법의 조합을 나타냅니다. 학습 속도를 조정하여 게임 균형에 대한 수렴을 촉진하기 위해 다양한 학습 속도를 사용합니다 [151].

시뮬레이션 결과는 SINR(Signal to Interference Plus Noise Ratio) 측면에서 제안한 방법의 재밍 방지 성능이 향상되었음을 보여줍니다. Stackelberg 게임에서 달성되는 최적의 전략은 최악의 시나리오에서 재머로 인해 발생하는 피해를 최소화할 수 있습니다.

최근 Han et al. [152]는 주파수 공간 전파 방해 통신 게임 기반의 DQN 알고리즘을 사용하는 CRN용 전파 방해 방지 시스템을 소개했습니다. 이 게임은 SU의 진행 중인 전송을 방해하기 위해 방해 신호를 주입하는 수많은 방해 전파 환경을 시뮬레이션합니다. SU는 PU의 통신을 방해해서는 안 되며 스마트 방해 전파를 물리쳐야 합니다. 이 통신 시스템은 주파수 호핑과 사용자 이동성을 모두 활용하는 2차원 통신 시스템입니다. RL 상태는 PU, SU, 방해 전파 및 서비스 기지국/액세스 포인트로 구성된 무선 환경입니다.

DQN은 SU가 전파 방해가 심한 지역을 떠나야 하는지 또는 신호를 보낼 채널을 선택해야 하는지를 결정하는 최적의 주파수 호핑 정책을 도출하는 데 사용됩니다.

실험 결과는 빠른 수렴 속도, SINR 증가, 방어 비용 절감, SU 활용성 향상 측면에서 Q-learning 기반 전략에 비해 DQN 기반 방법의 우월성을 보여줍니다. 핵심 구성요소 CNN을 포함하는 DQN은 벤치마크 Q-학습 방법에 비해 주파수 채널 수가 많은 시스템의 학습 속도를 높이는 데 도움이 됩니다.

Han et al.의 작업을 개선합니다. [152], Liu et al. [153]은 또한 DRL 방법을 사용하지만 다양하고 광범위한 기여를 하는 방해 전파 방지 통신 시스템을 제안했습니다. 특히 Liu et al. [153]은 [152]에서와 같이 SINR 및 PU 점유를 사용하는 대신 환경 상태를 특성화하기 위해 스펙트럼 폭포[154]로 알려진 시간적 특징이 있는 원시 스펙트럼 정보를 사용했습니다.

이 때문에 Liu et al.의 모델은 재밍 패턴과 재머 매개변수에 대한 사전 지식을 필요로 하지 않고 오히려 국지적 관측 데이터를 사용합니다. 이는 모델의 정보 손실을 방지하고 동적 환경에 대한 적응성을 촉진합니다. 더욱이 Liu 등의 연구에서는 재머가 [152]에서와 같이 사용자와 동일한 채널-슬롯 전송 구조를 취해야 한다고 가정하지 않습니다. 재귀적 CNN은 재귀적 특성을 갖는 스펙트럼 폭포로 표현되는 복잡한 무한 환경 상태를 처리하는 데 활용됩니다. 이 모델은 스위핑 재밍, 콤 재밍, 동적 재밍, 지능형 콤 재밍을 포함한 여러 가지 재밍 시나리오를 사용하여 테스트되었습니다. Han et al.의 단점. Liu 등의 방법은 한 명의 사용자에 대해서만 최적의 정책을 도출할 수 있다는 것이며, 이는 여러 사용자의 시나리오에 초점을 맞춘 향후 연구 방향에 영감을 줍니다.

2) 스푸핑 공격: 스푸핑 공격은 공격자가 미디어 액세스 제어와 같은 위조된 ID를 사용하여 다른 노드라고 주장하는 무선 네트워크에서 널리 사용됩니다. 불법적으로 네트워크에 접속하는 행위. 이러한 불법적인 침투는 중간자 공격, 즉 DoS 공격으로 이어질 수 있습니다[155]. Xiaoet al. [156], [157]은 합법적인 수신자와 스푸퍼 간의 상호 작용을 제로섬 인증 게임으로 모델링하고 스푸핑 탐지 문제를 해결하기 위해 Q-learning 및 Dyna-Q [158] 알고리즘을 활용했습니다. 수신기 또는 스푸퍼의 유용성은 스푸핑 탐지에서 예상되는 보상인 베이지안 위험을 기반으로 계산됩니다. 수신기는 PHY 계층 스푸핑 탐지에서 최적의 테스트 임계값을 선택하는 것을 목표로 하고 스푸퍼는 최적의 공격 주파수를 선택해야 합니다. 충돌을 방지하기 위해 스푸퍼는 협력하여 수신기를 공격합니다. 시뮬레이션과 실험 결과는 고정된 테스트 임계값을 사용한 벤치마크 방법에 비해 제안된 방법의 향상된 성능을 보여줍니다. 제안된 접근 방식의 단점은 동작 공간과 상태 공간이 모두 지정된 간격 내에서 경계를 이루는 이산 수준으로 양자화되어 지역적으로 최적의 솔루션을 얻을 수 있다는 것입니다.

3) 악성 코드 공격: 모바일 장치의 가장 까다로운 악성 코드 중 하나는 공개적으로 알려지지 않은 보안 취약점을 악용하는 제로 데이 공격이며, 이러한 공격이 억제되거나 완화될 때까지 해커는 이미 컴퓨터 프로그램, 데이터 또는 네트워크에 악영향을 끼쳤을 수 있습니다. [159], [160]. 이러한 공격을 방지하려면 애플리케이션에서 생성된 추적이나 로그 데이터를 실시간으로 처리해야 합니다. 제한된 컴퓨팅 성능, 배터리 수명 및 무선 대역폭으로 인해 모바일 장치는 처리를 위해 특정 맬웨어 탐지 작업을 클라우드의 보안 서버로 오프로드하는 경우가 많습니다. 강력한 컴퓨팅 리소스와 더욱 업데이트된 악성 코드 데이터베이스를 갖춘 보안 서버는 작업을 보다 빠르고 정확하게 처리한 다음 지연 시간을 최소화하여 탐지 보고서를 모바일 장치로 다시 보낼 수 있습니다. 따라서 오프로드 프로세스는 클라우드 기반 악성 코드 탐지 성능에 영향을 미치는 핵심 요소입니다. 예를 들어 너무 많은 작업이 클라우드 서버에 오프로드되면 무선 네트워크 정체가 발생하여 감지 지연이 길어질 수 있습니다. Wan et al. [161]은 이전에 제안된 [162] 게임 모델을 개선하여 모바일 오프로딩 성능을 향상시켰습니다. 최적의 오프로딩 속도를 선택하기 위해 [162]에서 사용된 Q-learning 접근 방식은 네트워크 크기가 증가하거나 선택할 수 있는 가능한 오프로딩 속도가 많을 때 고차원성의 저주를 겪습니다.

Wan et al. [161]은 핫부팅 Q-learning과 DQN의 사용을 주창하였으며, 표준 Q-learning에 비해 악성코드 탐지 정확도와 속도 측면에서 향상된 성능을 보여주었다. 최적의 오프로딩 속도를 선택하기 위해 DQN을 사용한 클라우드 기반 악성 코드 탐지 접근 방식은 그림 9에 나와 있습니다.

4) 적대적 환경에서의 공격: 기존 네트워크는 클라이언트 애플리케이션과 서버 간의 직접 통신을 용이하게 하며, 각 네트워크에는 스위치 제어 기능이 있어 네트워크 재구성 작업에 시간이 많이 걸리고 비효율적입니다. 이 방법은 여러 서버가 포함된 둘 이상의 데이터베이스에서 요청된 데이터를 검색해야 할 수도 있기 때문에 불리합니다. 소프트웨어 정의 네트워크는 네트워크를 적응적으로 재구성할 수 있는 차세대 네트워킹 기술이다. 네트워크 아키텍처의 글로벌 뷰를 통해 제어 기능을 프로그래밍할 수 있으므로 SDN은 네트워크 리소스를 효과적으로 관리하고 최적화할 수 있습니다.

RL은 문헌에서 강력한 것으로 광범위하게 입증되었습니다.

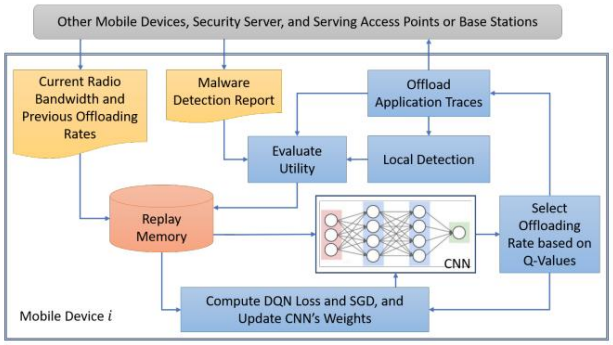

그림 9. SGD(확률적 경사하강법) 방법을 사용하여 CNN의 가중치를 업데이트하는 DQN을 사용한 클라우드 기반 악성 코드 탐지.
악성 탐지는 모바일 기기보다 강력한 컴퓨팅 자원을 갖춘 클라우드 서버에서 수행됩니다. DQN 에이전트는 네트워크 정체 및 감지 지연을 방지하기 위해 모바일 장치에 대한 최적의 작업 오프로드 속도를 선택하는 데 도움이 됩니다. 네트워크 상태를 관찰하고 서버의 맬웨어 탐지 보고서를 기반으로 유틸리티를 평가함으로써 에이전트는 일련의 최적 작업, 즉 동적 오프로드 속도를 생성하는 데 사용되는 상태와 보상을 공식화할 수 있습니다.

SDN 제어 방법(예: [163]~[167]).

SDN 제어에서 RL의 성공 사례는 풍부하지만, 공격자가 적대적 환경에서 네트워크 제어 알고리즘을 알고 있다면 방어자의 훈련 프로세스를 위조할 수 있습니다. 이 문제를 해결하기 위해 Han et al. [168]은 SDN을 위한 자율 방어 시스템을 구축하기 위해 적대적 RL의 사용을 제안했습니다. 공격자는 네트워크에서 중요한 노드를 선택하여 손상시킵니다(예: 백본 네트워크의 노드 또는 대상 서브넷의 노드). 공격자는 네트워크를 통해 전파함으로써 결국 중요한 서버를 손상시키려고 시도하는 반면, 방어자는 서버가 손상되는 것을 방지하고 영향을 받지 않는 노드를 최대한 많이 보존합니다. 이러한 목표를 달성하기 위해 RL 방어자는 "격리", "패치", "다시 연결" 및 "마이그레이션"으로 구성된 네 가지 가능한 조치를 취합니다. 두 가지 유형의 DRL 에이전트는 다양한 네트워크 상태에 따라 적절한 조치를 선택하기 위해 방어자(예: 이중 DQN 및 A3C)를 모델링하도록 훈련되었습니다.

보상은 크리티컬 상태에 따라 특성화됩니다.

서버, 보존된 노드 수, 마이그레이션 비용 및 취해진 조치의 유효성. 해당 연구에서는 공격자가 보상 기호를 뒤집거나 상태를 조작하여 RL 방어자의 학습 프로세스에 침투할 수 있는 시나리오를 고려했습니다. 이러한 원인 공격은 방어자의 훈련 과정을 방해하고 최적이 아닌 행동을 수행하게 만듭니다. 적대적 훈련 접근 방식은 인기 있는 네트워크 에뮬레이터인 Mininet[169]을 사용한 여러 실험을 통해 입증된 탁월한 성능으로 중독 공격의 영향을 줄이기 위해 적용됩니다.

적대적인 환경에서 방어자는 공격 유형, 공격 대상, 빈도, 위치 등 공격자의 개인 정보를 알지 못할 수 있습니다. 따라서 예를 들어 방어자는 공격자의 대상이 아닌 자산을 보호하기 위해 상당한 자원을 할당할 수 있습니다. 방어자는 침입자의 복잡성과 비용을 증가시키기 위해 방어 전략을 동적으로 재구성해야 합니다. Zhu et al. [170]은 방어자와 공격자가 반복적으로 방어와 공격 전략을 변경할 수 있는 모델을 도입했습니다. 방어자는 공격자에 대해 사전 지식이 없습니다.

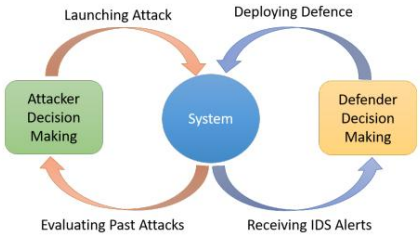

그림 10. 방어자와 공격자는 방어 및 공격 주기를 포함하는 적대적 환경에서 침입 탐지 시스템(IDS)을 통해 상호 작용합니다.
이 두 사이클을 이용하여 방어자와 공격자는 방어와 공격 전략을 반복적으로 변경할 수 있습니다. 이 모델은 버퍼 오버 읽기 공격[171] 및 코드 재사용 공격[172]과 같은 다양한 종류의 공격에 대한 방어 전략을 연구하는 데 사용할 수 있습니다.

시작된 공격 및 공격 정책. 그러나 공격자 클래스를 인식하고 방어 및 공격 활동이 공동으로 제공하는 시스템 유틸리티에 액세스할 수 있습니다. [170]에서는 사이버 방어를 위해 두 가지 대화형 RL 방법, 즉 적응형 RL과 강력한 RL이 제안되었습니다. 적응형 RL은 탐색 속도가 감소하는 공격(비지속적 공격자)을 처리하는 반면, 강력한 RL은 탐색 속도가 일정한 침입자(지속적 공격자)를 처리합니다. 방어자와 공격자 사이의 상호작용은 그림 10과 같이 공격과 방어 사이클을 통해 설명됩니다. 공격자와 방어자는 동시에 조치를 취하지 않고 비동기적으로 조치를 취합니다. 공격 주기에서 공격자는 필요한 경우 새 공격을 시작하기 전에 이전 공격을 평가합니다. 방어 사이클에서 방어자는 경보를 받은 후 최신 공격에 대한 메타 분석을 수행하고 해당 유틸리티를 계산한 후 필요한 경우 새로운 방어를 배치합니다. 이 시스템 모델의 장점은 공격자에 대한 기본 모델을 가정하지 않고 대신 공격 전략을 블랙박스로 취급한다는 것입니다.

또는 Elderman et al. 정보가 불완전하고 부분적으로 관찰할 수 있는 두 명의 에이전트(공격자 1명, 방어자 1명)가 포함된 확률론적 마르코프 게임으로 네트워킹에서 시뮬레이션된 사이버 보안 문제. 공격자는 네트워크 토폴로지를 알지 못하지만 귀중한 자산이 포함된 위치에 접근하여 액세스를 시도합니다. 방어자는 내부 네트워크를 알고 있지만 침입자의 공격 유형이나 위치는 볼 수 없습니다. 플레이어는 관찰할 수 없는 상대를 물리치기 위해 전략을 조정해야 하기 때문에 이것은 도전적인 사이버 보안 게임입니다[174]. 몬테카를로 학습, Q-학습, 신경망 등 다양한 알고리즘을 사용하여 방어자와 공격자를 모두 학습합니다. 시뮬레이션 결과는 소프트맥스 탐색을 통한 몬테카를로 학습이 공격 전략과 방어 전략을 모두 학습하는 가장 좋은 방법임을 보여줍니다. 신경망 알고리즘은 적대적 학습 능력이 제한되어 있으므로 Q-learning 및 Monte Carlo 학습 기술보다 성능이 뛰어납니다. 이 시뮬레이션은 실제 사이버 보안 문제를 하나의 자산으로 두 명의 플레이어만 하는 게임으로 단순화하는 단점이 있습니다.

실제로는 귀중한 데이터가 저장된 서버에 여러 명의 해커가 동시에 침투할 수 있습니다. 또한 네트워크는 시뮬레이션된 단일 위치가 아닌 다른 위치에 유용한 데이터를 포함할 수 있습니다.

신경망 및 학습 시스템에 대한 IEEE 트랜잭션, DOI: 10.1109/TNNLS.2021.3121870, 조기 액세스

IV. 토론 및 향후 연구 방향

DRL은 인간 또는 초인적 AI 에이전트를 설계하고 생성하는 가장 성공적인 방법 중 하나로 최근 몇 년 동안 등장했습니다. 이러한 성공의 대부분은 복잡하고 고차원적인 순차적 의사 결정 문제를 해결하기 위해 DNN을 기존 RL의 프레임워크에 통합하는 데 의존했습니다. 따라서 IoT, 사이버 보안 등 다양한 분야에서 DRL 알고리즘의 적용이 발견되고 있습니다. 오늘날 컴퓨터와 인터넷은 엔터테인먼트, 통신, 교통, 의료, 쇼핑 등 우리 삶의 여러 영역에서 중요한 역할을 하고 있습니다. 우리의 많은 개인 정보와 중요한 데이터는 온라인에 저장됩니다.

은행, 모기지 회사, 중개 회사 등 금융 기관도 온라인으로 비즈니스를 운영합니다. 따라서 해커가 컴퓨터 시스템에 접근하는 것을 방지하기 위한 보안 계획을 마련하는 것이 필수적입니다. 이 문서는 DRL 방법과 사이버 보안 문제에 대한 적용에 대한 포괄적인 조사를 제시했으며 주목할만한 예가 표 II에 요약되어 있습니다. 사이버 시스템의 적대적인 환경은 여러 DRL 에이전트를 포함하는 게임 이론 모델의 다양한 제안을 촉발했습니다. 우리는 이러한 종류의 응용 프로그램이 사이버 보안 문제에 대한 DRL과 관련된 문헌에서 논문의 주요 비율을 차지한다는 것을 발견했습니다.

A. CPS 보안 솔루션에 DRL 적용에 대한 과제와 향후 작업

새로운 영역은 사이버-물리 시스템용 보안 솔루션에 DRL을 사용하는 것입니다[175], [176]. 환경 모니터링 네트워크, 전기 스마트 그리드 시스템, 운송 관리 네트워크, 사이버 제조 관리 시스템 등 CPS의 대규모 및 복잡한 특성으로 인해 대응력과 정확성이 뛰어난 보안 솔루션이 필요합니다. 이는 TRPO 알고리즘 [93], LSTM-Q 학습 [89], 이중 DQN 및 A3C [86]와 같은 다양한 DRL 접근 방식으로 해결되었습니다. CPS 보안 솔루션을 위한 DRL 알고리즘을 구현할 때 가장 큰 과제 중 하나는 현실적인 CPS 시뮬레이션이 부족하다는 것입니다. 예를 들어 [86]의 작업에서는 Matlab/Simulink CPS 모델링을 사용하고 이를 OpenAI Gym 환경에 포함해야 했습니다. 이 구현은 OpenAI Gym 라이브러리에 Matlab/Simulink CPS 시뮬레이션을 통합함으로써 발생하는 오버헤드로 인해 계산 시간 측면에서 비용이 많이 듭니다. 따라서 향후 작업에서는 DRL 지원 환경에 직접 포함된 CPS 모델의 보다 적절한 시뮬레이션이 권장됩니다. DRL 알고리즘을 적용할 때 흔히 발생하는 또 다른 과제는 훈련된 정책을 시뮬레이션에서 실제 환경으로 전환하는 것입니다. 시뮬레이션은 DRL 에이전트 교육에 저렴하고 안전하지만 모델링의 부정확성과 오류로 인한 현실 격차로 인해 전송이 어려워집니다. 이는 CPS의 복잡성, 역학 및 대규모 규모로 인해 CPS 모델링에 더욱 중요합니다. 이러한 방향의 연구, 즉 CPS를 위한 DRL 기반 보안 솔루션의 시뮬레이션-실제 전송은 DRL 에이전트의 교육 과정에서 시간과 비용을 줄이고 안전성을 높이고 결국 비용이 많이 드는 실수를 줄이는 데 도움이 될 수 있으므로 조사할 가치가 있습니다. 실제 환경에서 실행할 때.

B. IDS에 DRL 적용에 대한 과제와 향후 작업

IDS에 대한 전통적인 RL 방법의 적용이 많이 있었지만 이러한 종류의 적용을 위한 DRL 알고리즘에 대한 작업은 적었습니다. 이는 아마도 최근에 딥러닝과 RL 방법의 통합이 지속되었기 때문일 것입니다. 침입 탐지 문제의 복잡성과 역동성은 딥 러닝의 강력한 표현 학습 및 함수 근사 기능과 기존 RL의 최적 순차 의사 결정 기능을 결합한 DRL 방법을 통해 효과적으로 해결될 것으로 예상됩니다.

IDS에 DRL을 적용하려면 에이전트를 대화식으로 교육하기 위한 시뮬레이션 또는 실제 침입 환경이 필요합니다. 훈련을 위해 실제 환경을 사용하는 것은 비용이 많이 드는 반면, 시뮬레이션된 환경은 현실과 거리가 멀 수 있기 때문에 이것은 큰 도전입니다.
침입 탐지를 위한 DRL에 대한 기존 연구의 대부분은 게임 기반 설정(예: 그림 7) 또는 레이블이 지정된 침입 데이터 세트에 의존했습니다. 예를 들어 [125]의 작업에서는 레이블이 지정된 침입 샘플의 두 데이터 세트를 사용하고 감독 학습 방식으로 이러한 데이터 세트에서 작동하도록 DRL 기계를 조정했습니다.

이러한 종류의 애플리케이션에는 라이브 환경이 부족하고 DRL 에이전트와 환경 간의 적절한 상호 작용이 부족합니다. 따라서 DRL 에이전트의 작업에 실시간으로 대응하고 DRL의 기능을 최대한 활용하여 복잡하고 정교한 사이버 침입 탐지 문제를 해결할 수 있는 보다 현실적인 환경을 만드는 방법에 대한 향후 연구에는 공백이 있습니다. 또한 호스트 기반 및 네트워크 기반 IDS에는 장점과 단점이 모두 있으므로 이러한 시스템을 결합하는 것이 논리적인 접근 방식이 될 수 있습니다. 이러한 종류의 통합 시스템을 위한 DRL 기반 솔루션은 또 다른 흥미로운 향후 연구일 것입니다.

C. 모델 기반 DRL 방법의 기능 탐색

지금까지 사이버 방어에 사용된 대부분의 DRL 알고리즘은 모델 프리(model-free) 방식으로 많은 양의 훈련 데이터가 필요하기 때문에 비효율적입니다. 실제 사이버 보안 실무에서는 이러한 데이터를 얻기가 어렵습니다. 연구자들은 일반적으로 제안된 접근 방식을 검증하기 위해 시뮬레이터를 활용하지만 이러한 시뮬레이터는 IoT 시스템의 실제 사이버 공간의 복잡성과 역동성을 완전히 특성화하지 못하는 경우가 많습니다. 모델 기반 DRL을 사용하면 확장 가능한 방식으로 데이터를 쉽게 수집할 수 있으므로 훈련 데이터를 제한적으로 사용할 수 있는 경우 모델 기반 DRL 방법이 모델 없는 방법보다 더 적합합니다. 따라서 모델 기반 DRL 방법의 탐색 또는 사이버 방어를 위한 모델 기반 및 모델 없는 방법의 통합은 흥미로운 향후 연구입니다. 예를 들어, 함수 근사기는 실제 고차원 및 부분적으로 관찰 가능한 환경의 프록시 모델을 학습하는 데 사용될 수 있으며[177]-[179], 몬테카를로 트리 검색 기술과 같은 계획 알고리즘을 배포하는 데 사용될 수 있습니다. [180], 최적의 동작을 도출합니다. 대안으로, 계획 기능을 갖춘 모델 없는 정책[181], [182] 또는 모델 기반 예측 검색[31]과 같은 모델 기반 및 모델 없는 조합 접근 방식을 두 방법의 장점을 통합하여 사용할 수 있습니다. 반면, 사이버 보안에 DRL을 적용하는 것에 관한 현재 문헌은 종종 행동 공간을 구분하는 데 제한이 있습니다.

표 2
사이버 보안의 일반적인 DRL 애플리케이션 요약

응용	목표/목표	알고리즘	상태	조치 다	보상
견고성 기반 CPS 위조 [86]	CPS에 대한 위조 입력(반례) 찾기	더블 DQN 그리고 A3C	는 다음의 출력으로 정의됩니다. 체계.	음 입력 값을 선택하십시오 조각별로 일정한 입력 신호 세트에서.	의 기능이 특징 과거 의존적 평생 속성, 출력 신호 및 시 간.
보안 및 안전 자율주행차 시스템에서 [89]	견고함을 극대화 사이버 물리 공격에 대한 AV 역학 제어 센서 판독값에 잘못된 데이터가 있습니다.	Q-러닝 LSTM으로	AV 자신의 위치와 속도 거리와 속도와 함께 근처에 있는 일부 물체의 예: AV의 선두주자.	적절한 속도를 취하십시오. 사이의 안전한 간격을 유지 AV.	다음과 같은 유틸리티 함수를 사용하여 최적의 안전과의 편차를 고려합니다. 간격.
증가 견고성 ~의 그런큼 자발적인 반대하는 시스템 적대적인 공격 [93]	필터링 방식을 고안하세요 손상된 것을 감지하기 위해 측정 (기만 공격)을 완화하고 적대적 오류의 영향.	TRPO	센서로 특징지어짐 측정 및 작동 소음.	아떤 추정을 결정 손상된 데이터베이스로부터 추정된 상태를 생성하는 데 사용하는 규칙 상태.	다음과 같은 함수를 통해 정의됩니다. 상태 특성을 입력으로 사용합니다.
안전한 오프로드 모바일 엣지에서 캐싱 [142]	모바일 정책 알아보기 데이터를 안전하게 오프로드하는 장치 재밍 및 스마트 공격에 대비하 여 노드를 에지합니다.	DQN 핫부팅 중 옮기다 학습 기술.	사용자 밀도, 배터리의 조합으로 표현 레벨, 재밍 강도, 무선 채널 대역폭.	상담원의 작업에는 다음이 포함됩니다. 엣지 노드를 선택하고, 오프로드 속도 선택 및 시간, 전력 전송 및 채널.	비밀을 기준으로 계산됨 용량, 에너지 소비, 그리고 의사소통 능률.
방해 전파 방지 통신 방식 CRN의 경우 [152]	최적의 주파수 도출 CRN SU에 대한 호핑 정책 스마트 재머를 물리치기 위해 주파수 공간 방해 전파 방지 게임 에 관한 것입니다.	CNN을 활용한 DQN	현재 상태는 다음과 같습니다. PU 및 SINR 정보 시간 t 1에 다음으로부터 수신됨 기지국 또는 액세스 제공 가리키다.	SU는 다음을 남기기 위해 조치를 취합니다. 무거운 지리적 영역 스마트로 인해 전파 방해가 방해됨 방해 전파를 제거하거나 주파수 채널을 선택하여 신호를 보냅니 다.	SINR 및 전송 비용을 기반으로 유틸리티 기능을 통해 표현됩니 다.
방해 전파 방지 의사소통 방법 [153], 개선 이전 작업 [152]	스마트한 방해 전파 방지 제안 [152]와 유사한 체계 두 가지 주요 차이점: Spec-trum Waterfall이 상태 및 방해 전파는 사용자와 채널-슬롯 전송 구조가 다릅니 다.	DQN 재귀적 CNN 기반 특 재귀 성에 스펙트럼의 폭포.	시간적, 스펙트럼 정보(예: 스펙트럼) 사 용 네트워크 환경의 주파수 및 시간 도메인 정보를 모두 포함하는 워터폴입니다.	에이전트의 작업은 다음을 선택하는 것입니다. 미리 정의된 세트에서 이산화된 전송 주파 수.	SINR 기반 전송과 관련된 기능으로 정의 주파수에 대한 요율 및 비용 스위칭.
스푸핑 감지 무선으로 네트워크 [156], [157]	최적의 인증 임계값을 선택합니다.	Q-러닝과 다이나-Q	잘못된 경보 비율을 포함하고 놓친 탐지율 스푸핑 감지 시스템 시간 t - 1	액션 세트에는 다음이 포함됩니다. 다양한 이산 선택 인증 수준 임계값은 지정된 간격.	베이지안을 기반으로 계산된 효용함수 사 용 예상되는 위험 스푸핑 탐지의 대가.
모바일 오프로딩 클라우드 기반의 경우 악성코드 탐지 [161] 개선 중 이전 작업 [162]에서	맬웨어 탐지 개선 정확성과 속도.	핫부팅 Q-러닝과 DQN.	현재 무선 대역폭과 이전 오프로딩으로 구 성됩니다. 다른 장치의 요금.	최적의 오프로드 속도 선택 모바일 기기별 레벨입니다.	다음을 기반으로 계산된 효용 함수로 표 현됩니다. 감지 정확도, 응답 속도, 전송 비용.
자발적인 SDN의 방어 [168]	중독 공격에 대처하세요 상태를 조작하거나 뒤집는 것 동안 보상 신호 RL 기반의 훈련 과정 국방 요원.	더블 DQN 그리고 A3C	배열로 표현됨 0과 1을 보여주는 네트워크 상태(노드가 손상되었거나 링크가 켜짐/꺼짐으로 전환됩니다). 배열 길이는 여러 개와 같습니다. 노드와 여러 링크.	공격자는 다음을 선택하는 방법을 배웁니다. 노드가 타협하는 동안 방어자는 격리, 패치, 재연결의 네 가 지 조치를 취할 수 있습니다. 서버 보호를 위해 마이그레이션 최대한 많은 노드를 보존합니다. 가능한.	상태를 기반으로 모델링됨 중요한 서버의 번호 보존 노드, 마이그레이션 비용과 행동의 타당성 찍은.
안전한 모바일 군중 감지 (MCS) 시스템 [193]	결제 정책을 최적화하여 공식화하여 가짜 감지 공격에 대한 감지 성능을 향상시킵니다. 스태켈버그 게임.	DQN	이전 센싱 품질과 결제로 구성 정책.	서버의 최적 선택 스마트폰 결제 벡터 사용자.	다음과 같은 유틸리티 함수를 사용하여 에 대한 총 지불액이 포함됩니다. 사용자와 서버의 이점 다양한 정확도 수준의 감지 보고서를 통해
자동화된 URL 기반 피싱 감지 [198]	악성 웹사이트 탐지 (URL)	DQN	벡터의 특징 HTTPS와 같은 웹사이트 기능의 공간 표 현 IP 주소를 갖는 프로토콜, URL의 접두사 또는 접미사.	무해하거나 피싱 URL에 해당하는 0 또는 1을 선택합니다.	분류 작업에 따라 보상은 1과 같습니다. 또는 URL이 분류된 경우 -1 옳든 그르든.

실제 문제에 대한 DRL 솔루션의 전체 기능.
최적의 선택을 위해 DRL을 적용하는 것이 그 예입니다.
행동 공간이 있는 [161], [162]의 모바일 오프로딩 비율
비율이 조금만 변경되더라도 분할되었습니다.
주로 클라우드 기반 악성 코드의 성능에 영향을 미칩니다.
탐지 시스템. 대처할 수 있는 방법에 대한 조사
사이버 환경의 지속적인 행동 공간(예: 정책)
그라디언트 및 배우 평론가 알고리즘은 또 다른 고무적입니다.
연구방향.

D. 적대적인 사이버 환경에서 DRL 훈련

AI는 사이버 공격을 방어하는 데 도움이 될 수 있지만
위험한 공격, 즉 공격적인 AI를 촉진합니다. 해커는 다음과 같은 작업을 수행할 수 있습니다.

AI를 활용해 공격을 더 스마트하고 더 많이 수행
컴퓨터 시스템이나 네트워크에 침투하기 위해 탐지 방법을 우회하도록 정교
합니다. 예를 들어, 해커는 다음을 사용할 수 있습니다.
사용자의 정상적인 행동을 관찰하고 사용하는 알고리즘
사용자의 패턴을 분석하여 추적 불가능한 공격 전략을 개발합니다.
기계 학습 기반 시스템은 인간을 모방하여 공예품을 만들 수 있습니다.
대규모 피싱 공격에 활용되는 유력한 가짜 메시지입니다. 마찬가지로, 매우 현실감 있게
만들어서
AI 발전을 기반으로 한 가짜 비디오 또는 오디오 메시지(예:
딥페이크[183]), 해커는 선거에서 거짓 뉴스를 퍼뜨릴 수 있습니다.
또는 금융 시장을 조작합니다[184]. 대안적으로 공격자는
딥 러닝 훈련에 사용되는 데이터 풀을 해칠 수 있음
방법(예: 기계 학습 중독) 또는 공격자가

상태나 정책을 조작하고, RL의 보상 신호 중 일부를 위조하여 에이전트가 최적이 아닌 조치를 취하도록 속여 에이전트가 손상되도록 합니다[185]. 이러한 종류의 공격은 AI 시스템 간의 전투의 일부이기 때문에 예방, 탐지 및 대응이 어렵습니다. 적대적 기계 학습, 특히 지도 방법은 사이버 보안에서 광범위하게 사용되어 왔지만[186], 적대적 RL을 사용하는 연구는 거의 발견되지 않았습니다[187]. 다양한 적대적 사이버 환경에서 훈련된 적대적 DRL 또는 DRL 알고리즘은 점점 복잡해지는 공격 AI 시스템에 맞서 싸울 수 있는 솔루션이 될 수 있으므로 포괄적인 조사 가치가 있습니다[188]-[190].

E. 인간 온 더 루프(Human-On-The-Loop) 모델을 사용한 인간-기계 팀 구

성 AI 시스템의 지원으로 사이버 보안 전문가는 더 이상 사이버 공격을 탐지하고 방어하기 위해 막대한 양의 공격 데이터를 수동으로 검사하지 않습니다. 보안 팀만으로는 규모를 유지할 수 없기 때문에 이는 많은 장점이 있습니다. AI 기반 방어 전략은 자동화되어 신속하고 효율적으로 구축될 수 있지만 이러한 시스템만으로는 새로운 위협이 도입될 때 창의적인 대응을 할 수 없습니다. 더욱이, 사이버 범죄나 사이버 전쟁의 배후에는 항상 인간의 적들이 있습니다. 따라서 사이버 방어를 위해서는 인간의 지성과 기계의 결합이 매우 필요합니다. 인간-기계 통합을 위한 전통적인 인간-인-루프(Human-In-The-Loop) 모델은 자율 에이전트가 작업의 일부를 수행하고 작업을 완료하기 전에 인간의 응답을 기다려야 하기 때문에 사이버 방어 시스템에 빠르게 적응하는 데 어려움을 겪습니다. 현대의 인간 온 더 루프(Human-On-The-Loop) 모델은 미래의 인간-기계 팀 구성 사이버 보안 시스템을 위한 솔루션이 될 것입니다. 이 모델을 사용하면 에이전트는 자동으로 작업을 수행할 수 있으며, 인간은 필요한 경우에만 에이전트의 작업을 모니터링하고 개입할 수 있습니다. 사이버 방어를 위한 Human-On-The-Loop 모델에서 인간 지식을 DRL 알고리즘[191]에 통합하는 방법은 흥미로운 연구 문제입니다.

F. 다중 에이전트 DRL 방법의 기능 탐색

해커들이 컴퓨터 시스템과 네트워크를 공격하기 위해 점점 더 정교하고 대규모의 접근 방식을 활용함에 따라 방어 전략도 더욱 지능적이고 대규모화되어야 합니다. 멀티에이전트 DRL은 이 문제를 해결하기 위해 탐구할 수 있는 연구 방향입니다. 본 문서에서 검토된 사이버 보안을 위한 게임 이론 모델에는 여러 에이전트가 포함되어 있지만 에이전트 간의 의사소통, 협력 및 조정이 제한된 몇 명의 공격자와 방어자로 제한됩니다. 효과적인 대규모 방어 계획을 위해서는 사이버 보안 문제에서 다중 에이전트 DRL의 이러한 측면을 철저하게 조사해야 합니다. 다중 에이전트 DRL 자체의 문제는 비정상성, 부분 관찰성 및 효율적인 다중 에이전트 훈련 방식과 같은 해결되어야 합니다[192]. 한편, 재밍(Jamming), 스푸핑(Spoofing), 허위 데이터 주입, 악성 코드, DoS, DDoS, 무차별 대입, Heartbleed, 봇넷, 웹 공격, 침투 공격 등 다양한 사이버 공격에 대처하기 위해 RL 방법론이 적용되어 왔다[193]. [198]. 그러나 최근 등장했거나 새로운 유형의 공격은 대부분 해결되지 않았습니다. 이 새로운 것 중 하나는 유형은 비트 앤 피스 DDoS 공격입니다. 이 공격은 주소당 정크가 너무 적기 때문에 많은 탐지 방법을 우회할 수 있도록 다수의 IP 주소가 있는 합법적인 트래픽에 작은 정크를 주입합니다. 예를 들어, 또 다른 새로운 공격은 컴퓨팅 클라우드에서 다른 회사의 IT 시스템을 관리하거나 서버에서 다른 회사의 데이터를 호스팅하는 회사의 시스템을 침해하는 공격입니다. 또는 해커는 양자 물리학 기반의 강력한 컴퓨터를 사용하여 현재 다양한 유형의 귀중한 데이터를 보호하는 데 사용되는 암호화 알고리즘을 해독할 수 있습니다[184]. 따라서 이러한 새로운 유형의 공격을 해결하기 위한 향후 연구가 권장됩니다.

참고자료

[1] I. Kakalou, KE Psannis, P. Krawiec 및 R. Badea, "5G를 향한 인지 무선 네트워크 및 네트워크 서비스 체인: 과제 및 요구 사항", IEEE Communications Magazine, vol. 55, 아니. 11, pp. 145-151, 2017.

[2] Y. Huang, S. Li, C. Li, YT Hou 및 W. Lou, "5G NR의 동적 eMBB/URLLC 다중화에 대한 심층 강화 학습 기반 접근 방식", IEEE Internet of Things Journal, vol. 7, 아니. 7, 페이지 6439-6456, 2020.

[3] P. Wang, LT Yang, X. Nie, Z. Ren, J. Li 및 L. Kuang, "데이터 기반 소프트웨어 정의 네트워크 공격 탐지: 최첨단 및 관점", 정보 과학 , 권. 513, pp. 65-83, 2020.

[4] A. Botta, W. De Donato, V. Persico 및 A. Pescape, "클라우드 컴퓨팅과 사물 인터넷의 통합: 설문 조사", Future Generation Computer Systems, vol. 56, pp. 684-700, 2016.

[5] O. Krestinskaya, AP James 및 LO Chua, "에지 컴퓨팅을 위한 신경기억 회로: 검토", 신경망 및 학습 시스템에 대한 IEEE 트랜잭션, vol. 31, 아니. 1, pp. 4-23, 2020.

[6] N. Abbas, Y. Zhang, A. Taherkordi 및 T. Skeie, "모바일 에지 컴퓨팅: 설문 조사", IEEE Internet of Things Journal, vol. 5, 아니. 1, 450-465페이지, 2018.

[7] AV Dastjerdi, R. Buyya, "포그 컴퓨팅: 사물 인터넷의 잠재력 실현 지원", Computer, vol. 49, 아니. 8, pp. 112-116, 2016.

[8] B. Geluvaraj, PM Satwik 및 TA Kumar, "사이버 보안의 미래: 사이버 공간에서 인공 지능, 기계 학습 및 딥 러닝의 주요 역할", 컴퓨터 네트워크 및 통신 기술에 관한 국제 회의, 2019, 739-747페이지.

[9] AL Buczak 및 E. Guven, "사이버 보안 침입 탐지를 위한 데이터 마이닝 및 기계 학습 방법에 대한 조사", IEEE Communications Surveys and Tutorials, vol. 18, 아니. 2, pp. 1153-1176, 2016.

[10] G. Apruzzese, M. Colajanni, L. Ferretti, A. Guido 및 M. Marchetti, "사이버 보안을 위한 기계 및 딥 러닝의 효율성", 사이버 분쟁에 관한 국제 컨퍼런스(CyCon), 2018, 371~390쪽.

[11] Y. Xin, L. Kong, Z. Liu, Y. Chen, Y. Li, H. Zhu, M. Gao, H. Hou, C. Wang, "사이버 보안을 위한 머신 러닝 및 딥 러닝 방법", IEEE Access, vol. 6, 페이지 35365-35381, 2018.

[12] N. Milosevic, A. Dehghantanha 및 KKR Choo, "기계 학습을 통한 Android 악성 코드 분류", 컴퓨터 및 전기 공학, vol. 61, pp. 266-274, 2017.

[13] R. Mohammed Harun Babu, R. Vinayakumar 및 KP Soman, "사이버 보안을 위한 딥 러닝 적용에 대한 간단한 검토", arXiv 사전 인쇄 arXiv:1812.06292, 2018.

[14] DS Berman, AL Buczak, JS Chavis, CL Corbett, "사이버 보안을 위한 딥러닝 방법에 대한 조사", Information, vol. 10, 아니. 4, 122페이지, 2019.

[15] S. Paul, Z. Ni 및 C. Mu, "사이버 물리 전력 시스템에서 적대적인 반복 게임을 위한 학습 기반 솔루션", 신경망 및 학습 시스템에 대한 IEEE 트랜잭션, DOI: 10.1109/TNNLS.2019.2955857, 2020.

[16] D. Ding, QL Han, Y. Xiang, X. Ge, XM Zhang, "산업용 사이버 물리 시스템의 보안 제어 및 공격 탐지에 관한 조사", Neurocomputing, vol. 275, pp. 1674-1683, 2018.

[17] M. Wu, Z. Song, YB Moon, "기계 학습 방법을 사용하여 CyberManufacturing 시스템에서 사이버 물리 공격 탐지," Journal of Intelligent Manufacturing, vol. 30, 아니. 3, pp. 1111-1123, 2019.

[18] L. Xiao, X. Wan, X. Lu, Y. Zhang 및 D. Wu, "기계 학습을 기반으로 한 IoT 보안 기술", arXiv 사전 인쇄 arXiv:1801.06275, 2018.

[19] A. Sharma, Z. Kalbarczyk, J. Barlow 및 R. Iyer, "대규모 컴퓨팅 조직의 보안 데이터 분석", DSN(Dependable Systems and Networks), IEEE/IFIP 41차 국제 컨퍼런스, 2011 , 506-517페이지.

[20] ND Nguyen, T. Nguyen 및 S. Nahavandi, "심층 강화 학습을 사용하는 인간 수준 에이전트를 위한 시스템 설계 관점: 설문 조사", IEEE Access, vol. 5, 페이지 27091-27102, 2017.

[21] Z. Sui, Z. Pu, J. Yi 및 S. Wu, "모델 기반 데모를 사용한 심층 강화 학습을 통한 충돌 방지를 통한 형성 제어", 신경망 및 학습 시스템에 대한 IEEE 트랜잭션, DOI: 10.1109/TNNLS.2020.3004893, 2020.

[22] A. Tsantekidis, N. Passalis, AS Toufa, K. Saitas-Zarkias, S. Chairis-tanidis, A. Tefas, "심층 강화 학습을 사용한 금융 거래의 가격 추적", 신경망 및 학습에 대한 IEEE 트랜잭션 시스템, DOI: 10.1109/TNNLS.2020.2997523, 2020.

[23] TT Nguyen, "다목적 심층 강화 학습 프레임워크", arXiv 사전 인쇄 arXiv:1803.02965, 2018.

[24] X. Wang, Y. Gu, Y. Cheng, A. Liu 및 CP Chen, "대략적인 정책 기반 가속 심층 강화 학습", 신경망 및 학습 시스템에 대한 IEEE 트랜잭션, vol. 31, 아니. 6, pp. 1820-1830, 2020.

[25] MH Ling, KLA Yau, J. Qadir, GS Poh, Q. Ni, "인지 무선 네트워크의 보안 강화를 위한 강화 학습 적용", Applied Soft Computing, vol. 37, pp. 809-829, 2015.

[26] Y. Wang, Z. Ye, P. Wan, J. Zhao, "인지 무선 네트워크의 강화 학습 알고리즘을 기반으로 한 동적 스펙트럼 할당에 대한 조사", Artificial Intelligence Review, vol. 51, 아니. 3, pp. 493-506, 2019.

[27] X. Lu, L. Xiao, T. Xu, Y. Zhao, Y. Tang 및 W. Zhuang, "VANET에 대한 강화 학습 기반 PHY 인증", IEEE Transactions on Vehicle Technology, vol. 69, 아니. 3, pp. 3068-3079, 2020.

[28] M. Alauthman, N. Aslam, M. Al-Kassassbeh, S. Khan, A. Al-Qerem 및 KKR Choo, "효율적인 강화 학습 기반 봇넷 탐지 접근 방식", Journal of Network and Computer Application, 권. 150, 페이지 102479, 2020.

[29] V. Mnih, K. Kavukcuoglu, D. Silver, AA Rusu, J. Veness, MG Bellemare, ... 및 S. Petersen, "심층 강화 학습을 통한 인간 수준 제어", Nature, vol. 518, 아니. 7540, pp. 529-533, 2015.

[30] ND Nguyen, S. Nahavandi 및 T. Nguyen, "심층 강화 학습에 대한 인간 혼합 전략 접근 방식", 2018년 IEEE 국제 컨퍼런스 SMC(시스템, 인간 및 사이버네틱스), 2018, 페이지 4023-4028 .

[31] D. Silver, A. Huang, CJ Maddison, A. Guez, L. Sifre, G. Van Den Driessche, ... 및 S. Dieleman, "심층 신경망과 트리 검색으로 바둑 게임을 마스터하기, " 자연, vol. 529, 아니. 7587, pp. 484-489, 2016.

[32] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, ... 및 Y. Chen, "인간 지식 없이 바둑 게임 마스터하기", Nature, vol. 550, 아니. 7676, pp. 354-359, 2017.

[33] O. Vinyals, T. Ewalds, S. Bartunov, P. Georgiev, AS Vezhnevets, M. Yeo, ... 및 J. Quan, "스타크래프트 II: 강화 학습을 위한 새로운 도전", arXiv 사전 인쇄 arXiv:1708.04782, 2017.

[34] P. Sun, X. Sun, L. Han, J. Xiong, Q. Wang, B. Li, ... 및 T. Zhang, "TStarBots: 스타크래프트 II에 내장된 부정 행위 수준 AI를 완전히 물리치기 게임," arXiv 사전 인쇄 arXiv:1809.07193, 2018.

[35] ZJ Pang, RZ Liu, ZY Meng, Y. Zhang, Y. Yu 및 T. Lu, "스타크래프트 전체 게임에 대한 강화 학습", arXiv 사전 인쇄 arXiv:1809.09095, 2018.

[36] V. Zambaldi, D. Raposo, A. Santoro, V. Bapst, Y. Li, I. Babuschkin 및 M. ⋯ Shanahan, "관계형 심층 강화 학습", arXiv 사전 인쇄 arXiv:1806.01830, 2018.

[37] M. Jaderberg, WM Czarnecki, I. Dunning, L. Marris, G. Lever, A. G. Castaneda, ... 및 N. Sonnerat, "인구 기반 심층 강화 학습을 사용한 1인칭 멀티플레이어 게임에서 인간 수준의 성능", arXiv 사전 인쇄 arXiv:1807.01281, 2018.

[38] OpenAI, "OpenAI Five", [온라인]. 사용 가능: https://openai.com/five/, 2019년 3월 1일.

[39] S. Gu, E. Holly, T. Lillicrap 및 S. Levine, 2017 IEEE International Conference on Robotics and Automation(ICRA)에서 "비동기 정책 외 업데이트를 통한 로봇 조작에 대한 심층 강화 학습", 2017, pp. 3389-3396.

[40] D. Isele, R. Rahimi, A. Cosgun, K. Subramanian 및 K. Fujimura, "심층 강화 학습을 사용하여 자율 차량으로 폐쇄된 교차로 탐색", 2018 IEEE 로봇 공학 및 자동화 국제 컨퍼런스(ICRA), 2018, 2034~2039페이지.

[41] TT Nguyen, ND Nguyen, F. Bello 및 S. Nahavandi, "수술 패턴 절단을 위한 심층 강화 학습을 사용한 새로운 인장 방법", 2019 IEEE 국제 산업 기술 회의(ICIT), DOI: 10.1109/ICIT .2019.8755235, 2019.

[42] ND Nguyen, T. Nguyen, S. Nahavandi, A. Bhatti 및 G. Guest, "자율 로봇 수술을 위한 심층 강화 학습을 통한 연조직 조작", 2019 IEEE 국제 시스템 회의(SysCon), DOI : 10.1109/SYSCON.2019.8836924, 2019.

[43] Y. Keneshloo, T. Shi, N. Ramakrishnan 및 CK Reddy, "시퀀스 간 모델을 위한 심층 강화 학습", 신경망 및 학습 시스템에 대한 IEEE 트랜잭션, vol. 31, 아니. 7, pp. 2469-2489, 2020.

[44] M. Mahmud, MS Kaiser, A. Hussain 및 S. Vassanelli, "생물학적 데이터에 대한 심층 학습 및 강화 학습의 응용", 신경망 및 학습 시스템에 대한 IEEE 거래, vol. 29, 아니. 6, pp. 2063-2079, 2018.

[45] M. Popova, O. Isayev 및 A. Tropsha, "새로운 약물 설계를 위한 심층 강화 학습", Science Advances, vol. 4, 아니. 7, 페이지 eaap7885, 2018.

[46] Y. He, FR Yu, N. Zhao, VC Leung 및 H. Yin, "스마트 시티를 위한 모바일 에지 컴퓨팅 및 캐싱을 갖춘 소프트웨어 정의 네트워크: 빅 데이터 심층 강화 학습 접근 방식", IEEE Communications Magazine, 권. 55, 아니. 12, pp. 31-37, 2017.

[47] HV Hasselt, A. Guez 및 D. Silver, "이중 Q 학습을 통한 심층 강화 학습", 제30차 AAAI 인공 지능 회의, 2016년, pp. 2094-2100.

[48] Z. Wang, T. Schhaul, M. Hessel, H. Hasselt, M. Lanctot 및 N. Freitas, "심층 강화 학습을 위한 결투 네트워크 아키텍처", 기계 학습에 관한 국제 컨퍼런스, 2016년, 1995페이지 -2003.

[49] H. Zhu, Y. Cao, W. Wang, T. Jiang, S. Jin, "모바일 에지 캐싱을 위한 심층 강화 학습: 검토, 새로운 기능 및 공개 문제", IEEE Network, vol. 32, 아니. 6, pp. 50-57, 2018.

[50] V. Mnih, AP Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, ... 및 K. Kavukcuoglu, "심층 강화 학습을 위한 비동기 방법", 기계 학습에 관한 국제 컨퍼런스, 2016, pp. 1928-1937.

[51] Y. Zhang, J. Yao, H. Guan, "심층 강화 학습을 통한 지능형 클라우드 리소스 관리", IEEE Cloud Computing, vol. 4, 아니. 6, pp. 60-69, 2017.

[52] J. Zhu, Y. Song, D. Jiang, H. Song, "인지 사물 인터넷을 위한 새로운 딥 Q 학습 기반 전송 스케줄링 메커니즘", IEEE Internet of Things Journal, vol. 5, 아니. 4, pp. 2375-2385, 2017.

[53] R. Shafin, H. Chen, YH Nam, S. Hur, J. Park, J. Zhang, ... 및 L. Liu, "자체 조정 부문화: 심층 강화 학습이 방송 빔 최적화를 충족함", IEEE 무선 통신 거래, vol. 19, 아니. 6, pp. 4038-4053, 2020.

[54] D. Zhang, X. Han, C. Deng, "스마트 그리드의 딥 러닝 및 강화 학습 연구 및 실습에 대한 검토", CSEE Journal of Power and Energy Systems, vol. 4, 아니. 3, pp. 362-370, 2018.

[55] X. He, K. Wang, H. Huang, T. Miyazaki, Y. Wang, S. Guo, "콘텐츠 중심 IoT의 심층 강화 학습을 기반으로 한 녹색 자원 할당", 신흥 주제에 대한 IEEE Transactions in 컴퓨팅, DOI: 10.1109/TETC.2018.2805718, 2018.

[56] Y. He, C. Liang, R. Yu 및 Z. Han, "컴퓨팅, 캐싱 및 통신을 갖춘 신뢰 기반 소셜 네트워크: 심층 강화 학습 접근 방식", IEEE Transactions on Network Science and Engineering, DOI: 10.1109/TNSE.2018.2865183, 2018.

[57] NC Luong, DT Hoang, S. Gong, D. Niyato, P. Wang, YC Liang, DI Kim, "통신 및 네트워킹에 심층 강화 학습 적용: 설문 조사", IEEE 통신 설문 조사 및 튜토리얼. DOI: 10.1109/COMST.2019.2916583, 2019.

[58] Y. Dai, D. Xu, S. Maharjan, Z. Chen, Q. He 및 Y. Zhang, "블록체인과 심층 강화 학습으로 지능형 5G를 넘어설 수 있었습니다."
IEEE 네트워크, vol. 33, 아니. 3, 2019년 10~17페이지.

[59] AS Leong, A. Ramaswamy, DE Quevedo, H. Karl 및 L. Shi, "사이버 물리 시스템의 무선 센서 스케줄링을 위한 심층 강화 학습", Automata, vol. 113, 페이지 108759, 2020.

[60] CJ Watkins, P. Dayan, "Q-learning", 기계 학습, vol. 8, 아니. 3-4, pp. 279-292, 1992.

[61] K. Arulkumaran, MP Deisenroth, M. Brundage 및 AA Bharath, "심층 강화 학습: 간단한 조사", IEEE Signal Process Magazine, vol. 34, 아니. 6, 26~38페이지, 2017.

[62] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra 및 M. Riedmiller, "심층 강화 학습으로 Atari 재생", arXiv 사전 인쇄 arXiv:1312.5602, 2013.

[63] T. Schhaul, J. Quan, I. Antonoglou 및 D. Silver, "우선순위 경험 재생", arXiv 사전 인쇄 arXiv:1511.05952, 2015.
[64] RJ Williams, “연결주의 강화 학습을 위한 간단한 통계적 기울기 추종 알고리즘,” Machine Learning, vol. 8, 아니. 3-4, pp. 229-256, 1992.

[65] RS Sutton, DA McAllester, SP Singh 및 Y. Mansour, "함수 근사를 통한 강화 학습을 위한 정책 구배 방법", 신경 정보 처리 시스템의 발전, 2000, 페이지 1057-1063.

[66] J. Schulman, S. Levine, P. Abbeel, M. Jordan 및 P. Moritz, "신뢰 지역 정책 최적화", International Conference on Machine Learning, 2015, pp. 1889-1897.

[67] J. Schulman, F. Wolski, P. Dhariwal, A. Radford 및 O. Klimov, "Prox-imal 정책 최적화 알고리즘", arXiv 사전 인쇄 arXiv:1707.06347, 2017.

[68] C. Wu, A. Rajeswaran, Y. Duan, V. Kumar, AM Bayen, S. Kakade 및 P. Abbeel, ... "행동 종속 인수분해 기준을 사용한 정책 그라데이션에 대한 분산 감소", arXiv 사전 인쇄 arXiv: 1803.07246 , 2018.

[69] TP Lillicrap, JJ Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, ... 및 D. Wierstra, "심층 강화 학습을 통한 지속적인 제어", arXiv 사전 인쇄 arXiv:1509.02971 , 2015.

[70] G. Barth-Maron, MW Hoffman, D. Budden, W. Dabney, D. Horgan, A. Muldal, ... 및 T. Lillicrap, "분산 분포 결정론적 정책 그라데이션", arXiv 사전 인쇄 arXiv:1804.08617 , 2018.
[71] M. Jaderberg, V. Mnih, WM Czarnecki, T. Schaul, JZ Leibo, D. Sil-ver 및 K. Kavukcuoglu, “비지도 보조 작업을 통한 강화 학습”, arXiv 사전 인쇄 arXiv:1611.05397 , 2016.
[72] O. Nachum, M. Norouzi, K. Xu 및 D. Schuurmans, "가치와 정책 기반 강화 학습 사이의 격차 해소", 신경 정보 처리 시스템의 발전, 2017, pp. 2775-2785.

[73] M. Lapan, 심층 강화 학습 실습: 심층 Q 네트워크, 값 반복, 정책 그라데이션, TRPO, AlphaGo Zero 등을 사용하여 최신 RL 방법 적용, Packt Publishing Ltd., 2018.

[74] OpenAI Gym Toolkit 문서. 고전적 제어: 고전적인 RL 문헌의 제어 이론 문제. 2020년 12월 14일에 검색됨: https://gym.openai.com/envs/#classic control [75] OpenAI Gym Toolkit 문서. Box2D: Box2D 시뮬레이터의 연속 제어 작
업입니다. 2020년 12월 14일 검색: https://gym.openai.com/envs/#box2d

[76] L. Wang, M. Torngren 및 M. Onori, "제조 분야의 사이버 물리 시스템의 현재 상태 및 발전", Journal of Manufacturing Systems, vol. 37, pp. 517-527, 2015.

[77] Y. Zhang, M. Qiu, CW Tsai, MM Hassan 및 A. Alamri, "Health-CPS: 클라우드 및 빅 데이터가 지원하는 의료 사이버 물리 시스템"
IEEE 시스템 저널, vol. 11, 아니. 1, pp. 88-95, 2017.
[78] PM Shakeel, S. Baskar, VS Dhulipala, S. Mishra 및 MM Jaber, “학습 기반 딥 Q 네트워크를 사용하여 의료 시스템의 보안 및 개인 정보 보호 유지,” Journal of Medical Systems, vol. 42, 아니. 10, 186페이지, 2018.

[79] MH Cintuglu, OA Mohammed, K. Akkaya 및 AS Uluagac, "스마트 그리드 사이버 물리 시스템 테스트베드에 대한 조사", IEEE 통신 조사 및 튜토리얼, vol. 19, 아니. 1, 446-464페이지, 2017.
[80] Y. Chen, S. Huang, F. Liu, Z. Wang, X. Sun, "자동 전압 제어에 대한 강화 학습 기반 허위 데이터 주입 공격 평가", IEEE Transactions on Smart Grid, vol. 10, 아니. 2, pp. 2158-2169, 2018.

[81] Z. Ni 및 S. Paul, "스마트 그리드 보안의 다단계 게임: 강화 학습 솔루션", 신경망 및 학습 시스템에 대한 IEEE 트랜잭션, DOI: 10.1109/TNNLS.2018.2885530, 2019.

[82] A. Ferdowsi, A. Eldosouky 및 W. Saad, "안전한 지능형 교통 시스템을 위한 상호의존성 인식 게임 이론 프레임워크", IEEE Internet of Things Journal. DOI: 10.1109/JIOT.2020.3020899, 2020.

[83] Y. Li, L. Zhang, H. Zheng, X. He, S. Peeta, T. Zheng 및 Y. Li, “교통-사이버-전기 자동차를 위한 비차선 규율 기반 자동차 추종 모델 물리적 시스템,” 지능형 교통 시스템에 관한 IEEE 거래, vol. 19, 아니. 1, pp. 38-47, 2018.

[84] C. Li 및 M. Qiu, 사이버 물리 시스템을 위한 강화 학습: 사이버 보안 사례 연구 포함, CRC Press, 2019.
[85] M. Feng, H. Xu, "알 수 없는 사이버 공격이 있을 때 사이버 물리 시스템을 위한 심층 강화 학습 기반 최적 방어", Computational Intelligence(SSCI), 2017 IEEE 심포지엄 시리즈, 2017, pp. 1 -8.

[86] T. Akazaki, S. Liu, Y. Yamagata, Y. Duan 및 J. Hao, "심층 강화 학습을 사용한 사이버 물리 시스템의 위조", 형식 방법에 관한 국제 심포지엄, 2018, pp. 456-465.

[87] H. Abbas 및 G. Fainekos, "안전 특성의 시뮬레이션된 어닐링 위조에 대한 수렴 증명", 2012년 통신, 제어 및 컴퓨팅에 관한 제50회 연례 Allerton 회의(Allerton), 2012년, pp. 1594-1601 .

[88] S. Sankaranarayanan 및 G. Fainekos, "교차 엔트로피 방법을 사용한 하이브리드 시스템의 시간적 속성 위조", 하이브리드 시스템에 관한 제15차 ACM 국제 회의: 계산 및 제어, 2012, pp. 125-134.

[89] A. Ferdowsi, U. Challita, W. Saad 및 NB Mandayam, "자율 차량 시스템의 보안 및 안전을 위한 강력한 심층 강화 학습", 2018년 제21차 지능형 교통 시스템(ITSC)에 관한 국제 회의, 307-312쪽.

[90] X. Wang, R. Jiang, L. Li, YL Lin, FY Wang, “장기 기억이 중요합니다: 딥러닝 기반 자동차 추적 모델에 대한 테스트 연구,”
Physica A: 통계 역학 및 그 응용, vol. 514, pp. 786-795, 2019.

[91] S. Hochreiter 및 J. Schmidhuber, "장기 단기 기억", 신경 계산, vol. 9, 아니. 8, pp. 1735-1780, 1997.
[92] I. Rasheed, F. Hu, L. Zhang, "LSTM-GAN을 사용하여 보안과 안전을 유지하기 위한 자율주행차 시스템에 대한 심층 강화 학습 접근 방식", Vehcular Communications, vol. 26, 페이지 100266, 2020.

[93] A. Gupta 및 Z. Yang, "사이버 공격을 받는 자율 시스템의 관찰자 설계를 위한 적대적 강화 학습", arXiv 사전 인쇄 arXiv:1809.06784, 2018.

[94] A. Abubakar 및 B. Pranggono, "소프트웨어 정의 네트워크를 위한 기계 학습 기반 침입 탐지 시스템", 2017년 제7차 신흥 보안 기술(EST)에 관한 국제 회의, 2017, pp. 138-143.

[95] S. Jose, D. Malathi, B. Reddy 및 D. Jayaseeli, "이상 기반 호스트 침입 탐지 시스템에 대한 조사", Journal of Physics: Conference Series, vol. 1000, 아니. 1, p. 012049, 2018.
[96] S. Roshan, Y. Miche, A. Akusok 및 A. Lendasse, "클러스터링 및 익스트림 학습 기계를 사용한 적응형 및 온라인 네트워크 침입 탐지 시스템," Journal of the Franklin Institute, vol. 355, 아니. 4, pp. 1752-1779, 2018.

[97] S. Dey, Q. Ye, S. Sampalli, "이기종 클라이언트 네트워크를 포함하는 모바일 클라우드에서 데이터 융합을 위한 기계 학습 기반 침입 탐지 체계", Information Fusion, vol. 49, pp. 205-215, 2019.

[98] D. Papamartzivanos, FG Marmol 및 G. Kambourakis, "딥 러닝 자가 적응형 오용 네트워크 침입 탐지 시스템 소개", IEEE Access, vol. 7, 페이지 13546-13560, 2019.

[99] W. Haider, G. Creech, Y. Xie 및 J. Hu, "제로 데이 및 스텔스 공격에 대한 호스트 기반 침입 탐지 시스템(IDS)의 견고성을 평가하기 위한 Windows 기반 데이터 세트", Future Internet, 권. 8, 아니. 2016년 3월 29일.

[100] P. Deshpande, SC Sharma, SK Peddoju 및 S. Junaid, "HIDS: 클라우드 컴퓨팅 환경을 위한 호스트 기반 침입 탐지 시스템", 시스템 보증 엔지니어링 및 관리 국제 저널, vol. 9, 아니. 3, pp. 567-576, 2018.

[101] M. Nobakht, V. Sivaraman 및 R. Boreli, "Open-Flow를 사용하는 스마트 홈 IoT를 위한 호스트 기반 침입 탐지 및 완화 프레임워크", 제11차 ARES(가용성, 신뢰성 및 보안에 관한 국제 회의), 2016, pp. 147-156.

[102] PAA Resende 및 AC Drummond, "침입 탐지 시스템을 위한 Random Forest 기반 방법에 대한 조사", ACM Computing Surveys(CSUR), vol. 51, 아니. 3, 48페이지, 2018.

[103] G. Kim, H. Yi, J. Lee, Y. Paek, S. 윤, "호스트 기반 침입 탐지 시스템 설계를 위한 LSTM 기반 시스템 호출 언어 모델링 및 강력한 앙상블 방법", arXiv 사전 인쇄 arXiv :1611.01726, 2016.

[104] A. Chawla, B. Lee, S. Fallon 및 P. Jacob, "CNN/RNN 모델이 결합된 호스트 기반 침입 탐지 시스템", 데이터베이스의 기계 학습 및 지식 발견에 관한 유럽 연합 컨퍼런스, 2018 , pp. 149-158.

[105] MM Hassan, A. Gumaei, A. Alsanad, M. Alrubaian 및 G. Fortino, "빅 데이터 환경에서 효율적인 침입 탐지를 위한 하이브리드 딥 러닝 모델", Information Sciences, vol. 513, pp. 386-396, 2020.
[106] A. Janagam, S. Hossen, “기계 학습 알고리즘(심층 강화 학습 알고리즘)을 사용한 네트워크 침입 탐지 시스템 분석,” 논문, Blekinge Institute of Technology, 스웨덴, 2018.

[107] X. Xu 및 T. Xie, "시스템 호출 시퀀스를 사용한 호스트 기반 침입 탐지를 위한 강화 학습 접근 방식", International Conference on Intelligent Computing, 2005, pp. 995-1003.

[108] RS Sutton, "시간적 차이 방법으로 예측하는 방법 학습", Machine Learning, vol. 3, 아니. 1, 9-44페이지, 1988.
[109] X. Xu, "강화 학습을 위한 희소 커널 기반 최소 제곱 시간차 알고리즘", International Conference on Natural Computation, 2006, pp. 47-56.

[110] X. Xu 및 Y. Luo, "침입 탐지의 동적 동작 모델링에 대한 커널 기반 강화 학습 접근 방식", 신경망 국제 심포지엄, 2007, pp. 455-464.

[111] X. Xu, "시간차 학습을 기반으로 한 순차적 이상 탐지: 원리, 모델 및 사례 연구", Applied Soft Computing, vol. 10, 아니. 3, pp. 859-867, 2010.

[112] B. Deokar, A. Hazarnis, "로그 파일과 강화 학습을 사용한 침입 탐지 시스템", International Journal of Computer Application, vol. 45, 아니. 19, pp. 28-35, 2012.
[113] K. Malialis, "네트워크 침입 대응을 위한 분산 강화 학습", 영국 요크 대학교 박사 학위 논문, 2014년.
[114] K. Malialis 및 D. Kudenko, "다중 에이전트 강화 학습을 사용한 네트워크 침입에 대한 분산 대응," 인공 지능의 엔지니어링 응용, vol. 41, pp. 270-284, 2015.

[115] RS Sutton, AG Barto, "강화 학습 소개,"
MIT Press Cambridge, MA, USA, 1998.
[116] DK Yau, JC Lui, F. Liang 및 Y. Yam, "최대 최소 공정 서버 중심 라우터 제한을 사용하여 분산 서비스 거부 공격 방어", 네트워킹에 대한 IEEE/ACM 트랜잭션, vol. 13, 아니. 1, pp. 29-42, 2005.

[117] R. Bhosale, S. Mahajan 및 P. Kulkarni, "침입 탐지 시스템을 위한 협력적 기계 학습", 국제 과학 및 공학 연구 저널, vol. 5, 아니. 1, pp. 1780-1785, 2014.

[118] A. Herrero 및 E. Corchado, "네트워크 침입 탐지를 위한 다중 에이전트 시스템: 검토", 정보 시스템 보안의 컴퓨팅 인텔리전스, 2009, pp. 143-154.

[119] A. Detwarasiti, RD Shachter, "팀 의사결정 분석을 위한 영향 다이어그램", 의사결정 분석, vol. 2, 아니. 4, pp. 207-228, 2005.
[120] S. Shamshirband, A. Patel, NB Anuar, MLM Kiah 및 A. Abra-ham, "무선 센서 네트워크의 침입을 탐지하고 방지하기 위해 퍼지 Q-학습을 사용한 협력 게임 이론 접근 방식"

인공지능의 공학적 응용, vol. 32, pp. 228-241, 2014.

[121] P. Munoz̃, R. Barco 및 I. de la Bandera, "차세대 무선 네트워크를 위한 퍼지 Q-학습을 사용한 로드 밸런싱 최적화,"
애플리케이션을 갖춘 전문가 시스템, vol. 40, 아니. 4, pp. 984-994, 2013.
[122] S. Shamshirband, NB Anuar, MLM Kiah 및 A. Patel, "다중 에이전트 시스템 기반 협력 무선 침입 탐지 전산 지능 기술의 평가 및 설계," 인공 지능의 엔지니어링 응용, vol. 26, 아니. 9, pp. 2105-2127, 2013.

[123] S. Varshney 및 R. Kuma, "WSN의 LEACH 라우팅 프로토콜 변형: 비교 분석," 제8차 클라우드 컴퓨팅, 데이터 과학 및 엔지니어링에 관한 국제 컨퍼런스(Confluence), 2018, pp. 199-204 .

[124] G. Caminero, M. Lopez-Martin 및 B. Carro, "침입 탐지를 위한 적대적 환경 강화 학습 알고리즘,"
컴퓨터 네트워크, vol. 159, 페이지 96-109, 2019.
[125] M. Lopez-Martin, B. Carro 및 A. Sanchez-Esguevillas, "지도 문제에 대한 침입 탐지에 심층 강화 학습 적용", Expert Systems with Application, vol. 141, 112963, 2020.
[126] IA Saeed, A. Selamat, MF Rohani, O. Krejcar 및 JA Chaudhry, "다중 에이전트 침입 탐지에 대한 체계적인 최첨단 분석", IEEE Access, vol. 8, pp. 180184-180209, 2020.

[127] S. Roy, C. Ellis, S. Shiva, D. Dasgupta, V. Shandilya 및 Q. Wu, "네트워크 보안에 적용되는 게임 이론 조사", 제43차 하와이 시스템 과학 국제 회의, 2010, pp. 1-10.

[128] S. Shiva, S. Roy 및 D. Dasgupta, "사이버 보안을 위한 게임 이론", 사이버 보안 및 정보 지능 연구에 관한 제6차 연례 워크숍, 2010, p. 34.

[129] K. Ramachandran 및 Z. Stefanova, "사이버 보안의 동적 게임 이론", 동적 시스템 및 응용 국제 회의, 2016년, vol. 7, pp. 303-310.

[130] Y. Wang, Y. Wang, J. Liu, Z. Huang 및 P. Xie, "사이버 보안을 위한 게임 이론 방법 조사", IEEE First International Conference on Data Science in Cyberspace(DSC), 2016, pp. 631-636.
[131] Q. Zhu 및 S. Rass, "게임 이론과 네트워크 보안의 만남: 튜토리얼", 컴퓨터 및 통신 보안에 관한 2018 ACM SIGSAC 컨퍼런스, 2018, pp. 2163-2165.

[132] A. Mpitziopoulos, D. Gavalas, C. Konstantopoulos 및 G. Pantziou, "WSN의 전파 방해 공격 및 대책에 대한 조사", IEEE 통신 조사 및 자습서, vol. 11, 아니. 4, 42-56페이지, 2009.

[133] S. Hu, D. Yue, X. Xie, X. Chen 및 X. Yin, "주기적인 DoS 방해 공격 하에서 네트워크 제어 시스템의 탄력적인 이벤트 트리거 컨트롤러 합성", IEEE Transactions on Cybernetics, DOI: 10.1109 /TCYB.2018.2861834, 2018.

[134] H. Boche 및 C. Deppe, "수동적 도청자 및 능동적 전파 방해 공격 하에서 안전한 식별", 정보 법의학 및 보안에 관한 IEEE Transactions, vol. 14, 아니. 2, 472-485페이지, 2019.
[135] Y. Wu, B. Wang, KR Liu 및 TC Clancy, "다채널 인지 무선 네트워크의 전파 방해 게임", 통신의 선택 영역에 관한 IEEE Journal, vol. 30, 아니. 1, 4-15페이지, 2012.

[136] S. Singh 및 A. Trivedi, "강화 학습 알고리즘을 사용한 인지 무선 네트워크의 전파 방해 방지", 무선 및 광 통신 네트워크(WOCN), 2012년 제9차 국제 컨퍼런스, pp. 1-5.

[137] Y. Gwon, S. Dastangoo, C. Fossa 및 HT Kung, "경쟁하는 모바일 네트워크 게임: 강화 학습을 통한 방해 전파 방지 및 전파 방해 전략 수용", IEEE Conference on Communications and Network Security(CNS), 2013, pp 28-36.

[138] WG Conley 및 AJ Miller, MILCOM 2013-2013 IEEE 군사 통신 컨퍼런스, 2013, pp. 1176-1182에서 "임시 인지 무선 네트워크에 동적으로 대응하기 위한 인지 방해 게임".

[139] K. Dabcevic, A. Betancourt, L. Marcenaro 및 CS Regazzoni, 2014년 음향, 음성 및 신호 처리에 관한 IEEE 국제 컨퍼런스에서 "인지 무선 통신에 대한 전파 방해 공격을 완화하기 위한 가상의 플레이 기반 게임 이론적 접근 방식" (ICASSP), 2014, 페이지 8158-8162.

[140] F. Slimeni, B. Scheers, Z. Chtourou 및 V. Le Nir, 2015년 군사 통신 및 정보에 관한 국제 회의에서 "수정된 Q-학습 알고리즘을 사용한 인지 무선 네트워크의 전파 방해 완화" 시스템(ICMCIS), 2015, 페이지 1-7.

[141] L. Xiao, X. Lu, D. Xu, Y. Tang, L. Wang 및 W. Zhuang, "강화 학습을 통한 스마트 전파 방해에 대비한 VANET의 UAV 릴레이", IEEE Transactions on Vehicle Technology, vol. 67, 아니. 5, 4087-4097페이지, 2018.

[142] L. Xiao, X. Wan, C. Dai, X. Du, X. Chen, M. Guizani, "강화 학습을 통한 모바일 에지 캐싱의 보안", IEEE Wireless Communications, vol. 25, 아니. 3, pp. 116-122, 2018.

[143] MA Aref, SK Jayaweera 및 S. Machuzak, "다중 에이전트 강화 학습 기반 인지 방해 방지", WCNC(Wireless Communica-tions and Networking Conference), 2017, 페이지 1-6.
[144] S. Machuzak 및 SK Jayaweera, "광대역 자율 인지 라디오를 이용한 강화 학습 기반 전파 방해 방지", 2016 IEEE/CIC 중국 통신 국제 회의(ICCC), 2016, pp. 1-5.

[145] MD Felice, L. Bedogni, L. Bononi, "인지 무선 네트워크를 위한 강화 학습 기반 스펙트럼 관리: 문헌 검토 및 사례 연구", Handbook of Cognitive Radio, 2019, pp. 1849-1886.
[146] A. Attar, H. Tang, AV Vasilakos, FR Yu 및 VC Leung, "인지 무선 네트워크의 보안 문제 조사: 솔루션 및 향후 연구 방향," IEEE 회보, vol. 100, 아니. 12, pp. 3172-3186, 2012.

[147] B. Wang, Y. Wu, KR Liu 및 TC Clancy, "인지 무선 네트워크를 위한 방해 전파 확률론적 게임", 통신의 선택 영역에 관한 IEEE Journal, vol. 29, 아니. 4, pp. 877-889, 2011.

[148] ML Littman, "다중 에이전트 강화 학습을 위한 프레임워크로서의 마르코프 게임", The 11th International Conference on Machine Learning, 1994, pp. 157-163.

[149] L. Xiao, Y. Li, J. Liu 및 Y. Zhao, "재밍에 대한 협력적 인지 무선 네트워크에서 강화 학습을 통한 전력 제어", The Journal of Supercomputing, vol. 71, 아니. 9, pp. 3237-3257, 2015.
[150] D. Yang, G. Xue, J. Zhang, A. Richa 및 X. Fang, "무선 네트워크의 스마트 재머 대처: Stackelberg 게임 접근 방식,"
무선 통신에 관한 IEEE 거래, vol. 12, 아니. 8, 4038-4047페이지, 2013.

[151] M. Bowling, M. Veloso, "가변 학습률을 사용한 다중 에이전트 학습", 인공 지능, vol. 136, 아니. 2, pp. 215-250, 2002.
[152] G. Han, L. Xiao, HV Poor, "심층 강화 학습을 기반으로 한 2차원 방해 전파 방지 통신", 음향, 음성 및 신호 처리에 관한 제42차 IEEE 국제 회의, 2017, pp. 2087-2091.

[153] X. Liu, Y. Xu, L. Jia, Q. Wu 및 A. Anpalagan, "스펙트럼 워터풀을 사용한 방해 전파 통신: 심층 강화 학습 접근 방식", IEEE Communications Letters, vol. 22, 아니. 5, 페이지 998-1001, 2018.

[154] W. Chen, X. Wen, "패턴 모양 인식 알고리즘의 지각 스펙트럼 폭포", 제18차 국제 첨단 통신 기술 회의(ICACT), 2016, pp. 382-389.

[155] K. Zeng, K. Govindan 및 P. Mohapatra, "무선 네트워크의 비암호화 인증 및 식별", IEEE Wireless Communications, vol. 17, 아니. 5, 56-62페이지, 2010.

[156] L. Xiao, Y. Li, G. Liu, Q. Li 및 W. Zhuang, "무선 네트워크에서 강화 학습을 통한 스푸핑 탐지", Global Communica-tions Conference(GLOBECOM), 2015년, 1페이지 -5.

[157] L. Xiao, Y. Li, G. Han, G. Liu 및 W. Zhuang, "무선 네트워크에서 강화 학습을 통한 PHY 계층 스푸핑 탐지", IEEE Transactions on Vehicle Technology, vol. 65, 아니. 12, pp. 10037-10047, 2016.

[158] RS Sutton, "대략적인 동적 프로그래밍을 기반으로 한 학습, 계획 및 반응을 위한 통합 아키텍처", The 7th International Conference on Machine Learning, 1990, pp. 216-224.

[159] X. Sun, J. Dai, P. Liu, A. Singhal 및 J. Yen, "제로 데이 공격 경로의 확률적 식별을 위한 베이지안 네트워크 사용", 정보 법의학 및 보안에 관한 IEEE Transactions, vol. 13, 아니. 10, pp. 2506-2521, 2018.

[160] Y. Afek, A. Bremler-Barr 및 SL Feibish, "대량 공격을 위한 제로데이 서명 추출", 네트워크 작업에 대한 IEEE/ACM 트랜잭션, DOI: 10.1109/TNET.2019.2899124, 2019.

[161] X. Wan, G. Sheng, Y. Li, L. Xiao 및 X. Du, "클라우드 기반 악성 코드 탐지를 위한 강화 학습 기반 모바일 오프로딩", GLOBECOM 2017-IEEE 글로벌 커뮤니케이션 컨퍼런스, 2017 , 1-6페이지.

[162] Y. Li, J. Liu, Q. Li 및 L. Xiao, "학습을 통한 맬웨어 탐지를 위한 모바일 클라우드 오프로딩", IEEE Conference on Computer Communications Workshops, 2015, pp. 197-201.

[163] MA Salahuddin, A. Al-Fuqaha 및 M. Guizani, "차량 인터넷을 지원하는 RSU 클라우드용 소프트웨어 정의 네트워킹,"
IEEE 사물 인터넷 저널, vol. 2, 아니. 2, pp. 133-144, 2015.

[164] R. Huang, X. Chu, J. Zhang 및 YH Hu, "강화 학습을 사용한 소프트웨어 정의 무선 센서 네트워크의 에너지 효율적인 모니터링: 프로토타입", International Journal of Distributed Sensor Networks, vol. 11, 아니. 2015년 10월 360428일.

[165] S. Kim, J. Son, A. Talukder, CS Hong, "SDN의 효율적인 라우팅을 위한 Q-leaning 기반의 혼잡 방지 메커니즘," International Conference on Information Networking(ICOIN), 2016, pp. 124 -128.

[166] SC Lin, IF Akyildiz, P. Wang 및 M. Luo, "다층 계층적 소프트웨어 정의 네트워크의 QoS 인식 적응형 라우팅: 강화 학습 접근 방식", 2016 IEEE International Conference on Services Computing(SCC), 2016, pp. 25-33.

[167] A. Mestres, A. Rodriguez-Natal, J. Carner, P. Barlet-Ros, E. Alarcon, M. Sole, ... 및 G. Estrada, "지식 정의 네트워킹" ACM
SIGCOMM 컴퓨터 통신 검토, vol. 47, 아니. 3, pp. 2-10, 2017.

[168] Y. Han, BI Rubinstein, T. Abraham, T. Alpcan, O. De Vel, S.
Erfani, ... 및 P. Montague, "소프트웨어 정의 네트워킹의 자율적 방어를 위한 강화 학습", 국제 보안 결정 및 게임 이론 회의, 2018, pp. 145-165.

[169] B. Lantz 및 B. O'Connor, "분산 SDN 개발을 위한 미니넷 기반 가상 테스트베드", ACM SIGCOMM Computer Communication Review, vol. 45, 아니. 4, pp. 365-366, 2015.

[170] M. Zhu, Z. Hu 및 P. Liu, "Heartbleed에 대한 적응형 사이버 방어를 위한 강화 학습 알고리즘", 이동 표적 방어에 관한 첫 번째 ACM 워크샵, 2014, pp. 51-58.

[171] J. Wang, M. Zhao, Q. Zeng, D. Wu 및 P. Liu, "버퍼 "Heartbleed" 과다 읽기 취약성의 위험 평가", 제45차 연례 IEEE/IFIP 의존 가능한 시스템에 대한 국제 컨퍼런스 및 네트워크, 2015, pp. 555-562.

[172] B. Luo, Y. Yang, C. Zhang, Y. Wang 및 B. Zhang, "코드 재사용 공격 및 방어에 대한 조사", 지능형 및 대화형 시스템 및 응용 프로그램에 관한 국제 컨퍼런스, 2018, pp. 782-788.

[173] R. Elderman, LJ Pater, AS Thie, MM Drugan, M. Wiering, "사이버 보안 시뮬레이션의 적대적 강화 학습", ICAART(International Conference on Agents and Artificial Intelligence), 2017, vol. 2, pp. 559-566.

[174] K. Chung, CA Kamhoua, KA Kwiat, ZT Kalbarczyk 및 R.
K. Iyer, "사이버 보안 모니터링을 위한 학습을 통한 게임 이론", IEEE 17차 HASE(High Assurance Systems Engineering) 국제 심포지엄, 2016, pp. 1-8.

[175] F. Wei, Z. Wan, H. He, "심층 강화 학습 기반 스마트 그리드를 위한 사이버 공격 복구 전략", IEEE Transactions on Smart Grid, vol. 11, 아니. 3, pp. 2476-2486, 2020.

[176] XR Liu, J. Ospina 및 C. Konstantinou, "풍력 통합 전력 시스템의 사이버 보안 평가를 위한 심층 강화 학습", arXiv 사전 인쇄 arXiv:2007.03025, 2020.

[177] J. Oh, X. Guo, H. Lee, RL Lewis, S. Singh, "아타리 게임에서 딥 네트워크를 사용한 동작 조건부 비디오 예측", 신경 정보 처리 시스템의 발전, 2015년, 2863페이지 -2871.

[178] M. Mathieu, C. Couprie 및 Y. LeCun, "평균 제곱 오류를 뛰어넘는 심층 다중 규모 비디오 예측", arXiv 사전 인쇄 arXiv:1511.05440, 2015.

[179] A. Nagabandi, G. Kahn, RS Fearing 및 S. Levine, "모델 없는 미세 조정을 통한 모델 기반 심층 강화 학습을 위한 신경망 역학", 2018 IEEE International Conference on Robotics and Automation(ICRA)), 2018, pp. 7559-7566.

[180] CB Browne, E. Powley, D. Whitehouse, SM Lucas, PI Cowling, P. Rohlfshagen, ... 및 S. Colton, "몬테카를로 트리 검색 방법에 대한 조사", 전산 지능 및 AI에 관한 IEEE 트랜잭션 게임에서, vol. 4, 아니. 1, pp. 1-43, 2012.

[181] A. Tamar, Y. Wu, G. Thomas, S. Levine 및 P. Abbeel, "가치 반복 네트워크", 신경 정보 처리 시스템의 발전, 2016, pp. 2154-2162.

[182] R. Pascanu, Y. Li, O. Vinyals, N. Heess, L. Buesing, S. Racaniere, ... 및 P. Battaglia, "처음부터 모델 기반 계획 학습", arXiv 사전 인쇄 arXiv: 1707.06170 , 2017.

[183] TT Nguyen, CM Nguyen, DT Nguyen, DT Nguyen 및 S.
Nahavandi, "딥페이크 생성 및 탐지를 위한 딥 러닝: 설문 조사", arXiv 사전 인쇄 arXiv:1909.11573, 2019.

[184] M. Giles, "2019년에 걱정해야 할 5가지 새로운 사이버 위협", MIT Technology Review. [온라인]. 이용 가능: https://www.technologyreview.com/2019/01/04/66232/five-emerging-cyber-threats-2019/, 2019년 1월 4일.

[185] V. Behzadan 및 A. Munir, "정책 유도 공격에 대한 심층 강화 학습의 취약성", 패턴 인식의 기계 학습 및 데이터 마이닝에 관한 국제 컨퍼런스, 2017, pp. 262-275.

[186] V. Duddu, "사이버 전쟁에서의 적대적 기계 학습에 대한 조사,"
국방과학저널, vol. 68, 아니. 4, pp.356-366, 2018.

[187] T. Chen, J. Liu, Y. Xiang, W. Niu, E. Tong 및 Z. Han, "강화 학습의 적대적 공격 및 방어 - AI 보안 관점에서,"
사이버 보안, vol. 2, 아니. 1, p. 2019년 11월 11일

[188] I. Ilahi, M. Usama, J. Qadir, MU Janjua, A. Al-Fuqaha, DT Hoang 및 D. Niyato, "심층 강화 학습에 대한 적대적 공격에 대한 과제 및 대응책", arXiv 사전 인쇄 arXiv: 2001.09684 , 2020.

[189] L. Tong, A. Laszka, C. Yan, N. Zhang 및 Y. Vorobeychik, "움직이는 건초 더미에서 바늘 찾기: 적대적 강화 학습으로 경고 우선 순위 지정", Proceedings of the AAAI Conference on Artificial Intelligence, 2020년, vol. 34, 아니. 1, 946-953페이지.

[190] J. Sun, T. Zhang, X. Xie, L. Ma, Y. Zheng, K. Chen 및 Y. Liu, "심층 강화 학습에 대한 은밀하고 효율적인 적대적 공격", arXiv 사전 인쇄 arXiv:2005.07099 , 2020.

[191] T. Nguyen, ND Nguyen 및 S. Nahavandi, "인간 전략을 사용한 다중 에이전트 심층 강화 학습", 2019 IEEE 국제 산업 기술 회의(ICIT), 2019, DOI: 10.1109/ICIT.2019.8755032 .

[192] TT Nguyen, ND Nguyen 및 S. Nahavandi, "다중 에이전트 시스템을 위한 심층 강화 학습: 과제, 솔루션 및 애플리케이션 검토", IEEE Transactions on Cybernetics, vol. 50, 아니. 9, pp. 3826-3839, 2020.

[193] L. Xiao, Y. Li, G. Han, H. Dai 및 HV Poor, "심층 강화 학습을 갖춘 안전한 모바일 크라우드센싱 게임", 정보 법의학 및 보안에 관한 IEEE 거래, vol. 13, 아니. 1, pp. 35-47, 2017.

[194] Y. Liu, M. Dong, K. Ota, J. Li 및 J. Wu, IEEE 23차 컴퓨터 지원 모델링 국제 워크숍에서 "소프트웨어 정의 네트워크에서 DDoS 범람에 대한 심층 강화 학습 기반 스마트 완화" 및 통신 링크 및 네트워크 설계(CAMAD), 2018, 페이지 1-6.

[195] Y. Xu, G. Ren, J. Chen, X. Zhang, L. Jia 및 L. Kong, "UAV 통신 네트워크에서 간섭 인식 협력 방해 전파 방지 분산 채널 선택", Applied Sciences, vol. 8, 아니. 2018년 1911년 10월 10일.

[196] F. Yao 및 L. Jia, "무선 네트워크의 공동 다중 에이전트 강화 학습 방해 전파 방지 알고리즘", IEEE Wireless Communications Letters, DOI: 10.1109/LWC.2019.2904486, 2019.

[197] Y. Li, X. Wang, D. Liu, Q. Guo, X. Liu, J. Zhang, and Y. Xu, "지능형 재머에 맞서는 심층 강화 학습 기반 재밍 방지 방법의 성능에 대해, " 응용과학, vol. 9, 아니. 7, 1361페이지, 2019.

[198] M. Chatterjee 및 AS Namin, "심층 강화 학습을 통해 피싱 웹 사이트 탐지", IEEE 43차 연례 컴퓨터 소프트웨어 및 애플리케이션 컨퍼런스(COMPSAC), 2019, vol. 2, pp. 227-232.

Thanh Thi Nguyen은 2015년 미국 캘리포니아주 스탠포드 대학교 컴퓨터 공학과, 2019년 미국 매사추세츠주 하버드 대학교 존 A. 폴슨 공학 및 응용과학 대학 엣지 컴퓨팅 연구소의 객원학자였습니다. 그는 2016년에 Alfred Deakin 박사후 연구 펠로우십을 받았고, 2018년에는 유럽 위원회로부터 ICT 전문가 교환 프로그램을 위한 유럽-태평양 파트너십 상을 받았으며, 호주-인도 전략 연구 기금 조기 및 중간 경력 펠로우십을 받았습니다. 2020년 과학 아카데미. Nguyen 박사는 2013년 호주 모나쉬 대학교에서 수학 및 통계학 박사 학위를 취득했으며 인공 지능, 딥 러닝, 심층 강화 학습, 사이버 보안, IoT, 데이터 과학 등 다양한 분야에 대한 전문 지식을 보유하고 있습니다. 그는 현재 호주 빅토리아주 디킨대학교 정보기술대학원의 수석 강사로 재직하고 있습니다.

Vijay Janapa Reddi는 2010년 Harvard University에서 컴퓨터 과학 박사 학위를 취득했습니다. 그는 NAE(National Academy of Engineering) Gilbreth Lecturer Honor(2016), IEEE TCCA Young Computer Architect Award(2016) 등 여러 상을 받았습니다. 2016), Intel Early Career Award(2013), Google Faculty Research Awards(2012, 2013, 2015, 2017), 2005 마이크로아키텍처 국제 심포지엄 최우수 논문, 2009 고성능 컴퓨터 아키텍처 국제 심포지엄 최우수 논문, IEEE Computer Architecture Awards의 최고 추천 제품(2006, 2010, 2011, 2016, 2017).

Reddi 박사는 현재 Harvard University의 John A. Paulson 공학 및 응용 과학 대학 부교수로 재직하며 Edge Computing Lab을 이끌고 있습니다. 그의 연구 관심 분야는 컴퓨터 아키텍처와 시스템 소프트웨어 설계, 특히 기계 학습을 기반으로 한 모바일 및 엣지 컴퓨팅 플랫폼의 맥락입니다.