(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2024년07월09일
(11) 등록번호 10-2683123
(24) 등록일자 2024년07월04일

(51) 국제특허분류(Int. Cl.)
G06N 3/092 (2023.01) *G06F 21/55* (2013.01)
G06N 3/096 (2023.01) *G06N 3/0985* (2023.01)
(52) CPC특허분류
G06N 3/092 (2023.01)
G06F 21/55 (2013.01)
(21) 출원번호 10-2024-0023307
(22) 출원일자 2024년02월19일
심사청구일자 2024년02월19일
(56) 선행기술조사문헌
KR1020220154785 A*
(뒷면에 계속)

(73) 특허권자
세종대학교산학협력단
서울특별시 광진구 능동로 209 (군자동, 세종대학교)
(72) 발명자
김문석
서울특별시 광진구 능동로 209 (군자동)
(74) 대리인
수안특허법인

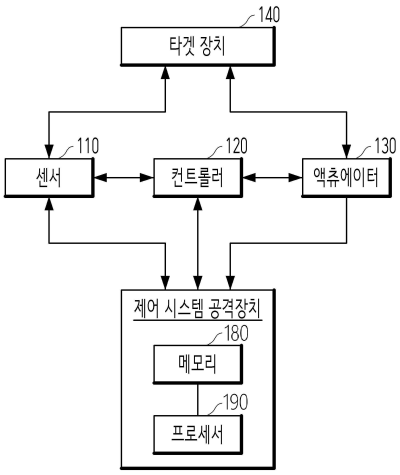
전체 청구항 수 : 총 8 항 심사관 : 신지선

(54) 발명의 명칭 **강화 학습 기반의 제어 시스템 공격 방법 및 장치**

(57) 요 약

강화 학습을 이용하여, 제어 시스템을 공격하는 방법 및 장치가 개시된다. 개시된 제어 시스템 공격 방법은 타겟 장치에 대한 제어 시스템에 포함된 센서의 측정값 및 상기 제어 시스템에 포함된 액츄에이터에 대한 설정값 중 적어도 하나를 수집하는 단계; 및 미리 학습된 강화 학습 모델을 이용하여, 상기 제어 시스템을 공격하기 위한 공격용 측정값 및 공격용 설정값 중 적어도 하나를 결정하는 단계를 포함하며, 상기 강화 학습 모델의 상태 정보는 제1시점에서의 상기 측정값 및 설정값 중 적어도 하나를 포함하며, 상기 강화 학습 모델의 액션 정보는 상기 제1시점의 다음 시점인 제2시점에서의 상기 공격용 측정값 및 상기 공격용 설정값 중 적어도 하나를 포함한다.

대 표 도 - 도1



(52) CPC특허분류
G06N 3/096 (2023.01)
G06N 3/0985 (2023.01)

(56) 선행기술조사문헌
KR102372934 B1*
US20220147815 A1
US20230393543 A1
KR1020220049674 A
*는 심사관에 의하여 인용된 문헌

이 발명을 지원한 국가연구개발사업

과제고유번호	1711198792
과제번호	00165231
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	가동원전안전성향상핵심기술개발사업
연구과제명	APR1400 대상 AI 탐지 알고리즘 및 시뮬레이션 기반 공격 패킷생성 기술 개발
과제수행기관명	세종대학교산학협력단
연구기간	2024.01.01 ~ 2024.12.31

명 세 서

청구범위

청구항 1

타겟 장치에 대한 제어 시스템에 포함된 센서의 측정값 및 상기 제어 시스템에 포함된 액츄에이터에 대한 설정값 중 적어도 하나를 수집하는 단계; 및

미리 학습된 강화 학습 모델을 이용하여, 상기 제어 시스템을 공격하기 위한 공격용 측정값 및 공격용 설정값 중 적어도 하나를 결정하는 단계를 포함하며,

상기 강화 학습 모델의 상태 정보는

제1시점에서의 상기 측정값 및 설정값 중 적어도 하나를 포함하며,

상기 강화 학습 모델의 액션 정보는

상기 제1시점의 다음 시점인 제2시점에서의 상기 공격용 측정값 및 상기 공격용 설정값 중 적어도 하나를 포함하며,

상기 강화 학습 모델의 학습을 위한 보상값은

상기 타겟 장치의 피해 정도 및 상기 제어 시스템을 위한 공격 탐지 모델의 공격 탐지 여부에 따라서 결정되는

제어 시스템 공격 방법.

청구항 2

삭제

청구항 3

제 1항에 있어서,

상기 보상값은

상기 피해 정도에 비례하도록 결정되며,

상기 피해 정도는

하기 [수학식]에 따라 계산되는

제어 시스템 공격 방법.

[수학식]

$$\text{피해 정도} = r_l / d_l$$

$$r_l = H_l - L_l$$

$$d_l = \begin{cases} min(|s_l - L_l|, |s_l - H_l|) & L_l < s_l < H_l \\ \delta_l & otherwise \end{cases}$$

여기서, H_l은 상기 측정값의 정상 범위 중 최대값, L_l은 상기 측정값의 정상 범위 중 최소값, s_l은 상기 측정값, δ_l은 미리 설정된 상수값, l은 상기 센서의 인덱스를 나타냄.

청구항 4

제 3항에 있어서,

상수값은

상기 공격 탐지 모델에 의해 공격이 탐지된 경우의 상수값인

제어 시스템 공격 방법.

청구항 5

제 4항에 있어서,

상기 상수값은

상기 최대값과 상기 최소값의 차이값의 절반보다 큰 값인

제어 시스템 공격 방법.

청구항 6

제 1항에 있어서,

상기 측정값에 따라서 상기 설정값을 결정하는 상기 제어 시스템의 컨트롤러로, 상기 공격용 측정값을 전송하는 단계

를 더 포함하는 제어 시스템 공격 방법.

청구항 7

제 1항에 있어서,

상기 설정값에 따라서 동작하는 상기 액츄에이터로, 상기 공격용 설정값을 전송하는 단계

를 더 포함하는 제어 시스템 공격 방법.

청구항 8

미리 학습된 강화 학습 모델을 저장하는 메모리; 및

상기 메모리와 전기적으로 연결된 적어도 하나의 프로세서를 포함하며,

상기 프로세서는,

타겟 장치에 대한 제어 시스템에 포함된 센서의 측정값 및 상기 제어 시스템에 포함된 액츄에이터에 대한 설정값 중 적어도 하나를 상기 강화 학습 모델에 입력하여, 상기 제어 시스템을 공격하기 위한 공격용 측정값 및 공격용 설정값 중 적어도 하나를 결정하며,

상기 강화 학습 모델의 상태 정보는

제1시점에서의 상기 측정값 및 설정값 중 적어도 하나를 포함하며,

상기 강화 학습 모델의 액션 정보는

상기 제1시점의 다음 시점인 제2시점에서의 상기 공격용 측정값 및 상기 공격용 설정값 중 적어도 하나를 포함하며,

상기 강화 학습 모델의 학습을 위한 보상값은

상기 타겟 장치의 피해 정도 및 상기 제어 시스템을 위한 공격 탐지 모델의 공격 탐지 여부에 따라서 결정되는

제어 시스템 공격 장치.

청구항 9

삭제

청구항 10

제 8항에 있어서,

상기 보상값은

상기 피해 정도에 비례하도록 결정되며,

상기 피해 정도는

하기 [수학식]에 따라 계산되는

제어 시스템 공격 장치.

[수학식]

$$\text{피해 정도} = r_l / d_l$$

$$r_l = H_l - L_l$$

$$d_l = \begin{cases} min(|s_l - L_l|, |s_l - H_l|) & L_l < s_l < H_l \\ \delta_l & otherwise \end{cases}$$

여기서, H_l은 상기 측정값의 정상 범위 중 최대값, L_l은 상기 측정값의 정상 범위 중 최소값, s_l은 상기 측정값, δ_l 은 미리 설정된 상수값, l은 상기 센서의 인덱스를 나타냄.

발명의 설명

기 술 분 야

[0001] 본 발명은 제어 시스템을 공격하는 방법 및 장치에 관한 것으로서, 더욱 상세하게는 강화 학습 기반으로, 제어 시스템을 공격하는 방법 및 장치에 관한 것이다.

배 경 기 술

[0003] 산업 제어 시스템(Industrial Control System) 등 다양한 종류의 제어 시스템은 외부의 공격으로부터 시스템을 보호하기 위한 방어 방법을 이용하고 있다. 이러한 방어 방법은 일예로서, 공격자의 공격을 탐지하고 공격을 탐지한 경우 공격 트래픽을 차단하는 방식일 수 있다.

[0004] 제어 시스템에 대한 공격 기법은 날로 발전하고 있으며, 공격 기법의 발전은 역설적으로 제어 시스템의 방어 성능을 강화하는 역할을 수행하고 있다. 개발자들은 제어 시스템에 대한 다양한 공격 기법을 연구함으로써, 공격 차단 성능이 강화된 방어 방법을 개발할 수 있다. 최근 인공 지능 기술의 발전과 함께, 인공 지능 기술을 이용한 공격 기법이 개발되고 있으며, 인공 지능 기술을 이용한 공격 기법의 개발은 제어 시스템의 방어 성능을 더욱 강화시킬 수 있따.

[0005] 관련 선행문헌으로 특허문헌인 대한민국 등록특허 제10-2100328호, 제10-2508071호, 제10-2596738호가 있으며,

비특허문헌인 "실시간 분산 제어시스템 보안, 신인철, 정보보호학회지 제24권 제2호, 2014. 10"이 있다.

발명의 내용

해결하려는 과제

[0007] 본 발명은, 강화 학습을 이용하여, 제어 시스템을 공격하는 방법 및 장치를 제공하기 위한 것이다.

[0008] 또한 본 발명은, 제어 시스템의 공격 방어에 활용될 수 있는, 제어 시스템 공격 방법 및 장치를 제공하기 위한 것이다.

과제의 해결 수단

[0010] 상기한 목적을 달성하기 위한 본 발명의 일 실시예에 따르면, 타겟 장치에 대한 제어 시스템에 포함된 센서의 측정값 및 상기 제어 시스템에 포함된 액츄에이터에 대한 설정값 중 적어도 하나를 수집하는 단계; 및 미리 학습된 강화 학습 모델을 이용하여, 상기 제어 시스템을 공격하기 위한 공격용 측정값 및 공격용 설정값 중 적어도 하나를 결정하는 단계를 포함하며, 상기 강화 학습 모델의 상태 정보는 제1시점에서의 상기 측정값 및 설정값 중 적어도 하나를 포함하며, 상기 강화 학습 모델의 액션 정보는 상기 제1시점의 다음 시점인 제2시점에서의 상기 공격용 측정값 및 상기 공격용 설정값 중 적어도 하나를 포함하는 제어 시스템 공격 방법이 제공된다.

[0011] 또한 상기한 목적을 달성하기 위한 본 발명의 다른 실시예에 따르면, 미리 학습된 강화 학습 모델을 저장하는 메모리; 및 상기 메모리와 전기적으로 연결된 적어도 하나의 프로세서를 포함하며, 상기 프로세서는, 타겟 장치에 대한 제어 시스템에 포함된 센서의 측정값 및 상기 제어 시스템에 포함된 액츄에이터에 대한 설정값 중 적어도 하나를 상기 강화 학습 모델에 입력하여, 상기 제어 시스템을 공격하기 위한 공격용 측정값 및 공격용 설정값 중 적어도 하나를 결정하며, 상기 강화 학습 모델의 상태 정보는 제1시점에서의 상기 측정값 및 설정값 중 적어도 하나를 포함하며, 상기 강화 학습 모델의 액션 정보는 상기 제1시점의 다음 시점인 제2시점에서의 상기 공격용 측정값 및 상기 공격용 설정값 중 적어도 하나를 포함하는 제어 시스템 공격 장치가 제공된다.

발명의 효과

[0013] 본 발명의 일실시예에 따르면, 제어 시스템의 공격 탐지를 회피할 수 있는 제어 시스템 공격 방법 및 장치가 제공될 수 있다.

[0014] 또한 본 발명의 일실시예에 따르면, 제어 시스템에 대한 공격을 탐지하고 방어하는 방법을 개발하는데 이용될 수 있는, 제어 시스템 공격 방법 및 장치가 제공될 수 있다.

도면의 간단한 설명

[0016] 도 1은 본 발명의 일실시예에 따른 제어 시스템 공격 장치를 설명하기 위한 도면이다.

도 2는 본 발명의 일실시예에 따른 제어 시스템 공격 방법을 설명하기 위한 도면이다.

도 3은 본 발명의 일실시예에 따른 강화 학습 모델을 설명하기 위한 도면이다.

도 4는 본 발명의 일실시예에 따른 강화 학습 모델의 업데이트 방법을 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0017] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세한 설명에 상세하게 설명하고자 한다. 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다. 각 도면을 설명하면서 유사한 참조부호를 유사한 구성요소에 대해 사용하였다.

[0018] 이하에서, 본 발명에 따른 실시예들을 첨부된 도면을 참조하여 상세하게 설명한다.

[0020] 도 1은 본 발명의 일실시예에 따른 제어 시스템 공격 장치를 설명하기 위한 도면으로서, 도 1에서는 산업 제어 시스템을 공격하는 공격 장치가 일실시예로서 설명된다.

[0021] 제어 시스템 예컨대, 산업 제어 시스템은 도 1에 도시된 바와 같이, 센서(110), 컨트롤러(120) 및 액츄에이터(130, actuator)를 포함하며, 타겟 장치(140)에 대한 제어를 수행한다. 센서(110)의 측정값은

PLC(Programmable Logic Controller)와 같은 컨트롤러(120)로 입력되고, 컨트롤러(120)는 측정값에 따라서 액츄에이터(130)를 위한 설정값을 결정하여, 액츄에이터(130)로 전송한다. 그리고 액츄에이터(130)는 설정값에 따라서 동작을 수행한다. 센서(110)는 주기적으로 측정값을 생성하고, 주기적으로 생성되는 측정값에 따라서, 설정값 역시 주기적으로 생성된다.

[0022] 예컨대, 타겟 장치(140)는 물탱크, 센서(110)는 물탱크의 수위를 측정하는 센서, 액츄에이터(130)는 물탱크로 물을 주입하거나 물탱크로부터 물을 배수시키는 펌프일 수 있다. 이러한 예시에서, 컨트롤러(120)는 센서(110)가 전송하는 물탱크(140)의 수위 측정값에 따라서 펌프(130)에 대한 설정값을 생성하여, 펌프(130)로 전송한다. 만일 물탱크(140)의 수위가 정상 범위를 벗어나 매우 낮다면 컨트롤러(120)는 펌프(130)가 물탱크(140)로 물을 주입할 수 있도록 설정값을 생성하여 펌프(130)로 전송하며, 물탱크(140)의 수위가 정상 범위를 벗어나 매우 높다면, 펌프(130)가 물탱크(140)로부터 물을 배수할 수 있도록 설정값을 생성하여 펌프(130)로 전송할 수 있다. 물탱크의 수위가 정상 범위내라면, 컨트롤러(120)는 펌프(130)가 동작하지 않도록 설정값을 생성하여 펌프(130)로 전송할 수 있다.

[0023] 본 발명의 일실시예에 따른 제어 시스템 공격 장치는 메모리(180) 및 메모리(180)와 전기적으로 연결된 적어도 하나의 프로세서(190)를 포함한다. 메모리(180)는 미리 학습된 강화 학습 모델을 저장하며, 프로세서(190)는 강화 학습 모델을 이용하여, 제어 시스템에 대한 공격을 수행한다.

[0024] 프로세서(190)는 타겟 장치(140)에 대한 제어 시스템에 포함된 센서(110)의 측정값 및 제어 시스템에 포함된 액츄에이터(130)에 대한 설정값 중 적어도 하나를 강화 학습 모델에 입력하여, 제어 시스템을 공격하기 위한 공격용 측정값 및 공격용 설정값 중 적어도 하나를 결정할 수 있다. 즉, 프로세서(190)는 강화 학습 모델을 이용하여, 조작된 측정값인 공격용 측정값과, 조작된 설정값인 공격용 설정값을 생성할 수 있다. 구체적인 공격용 측정값 및 공격용 설정값 결정 방법은 도 2에서 자세히 설명된다.

[0025] 센서(110)의 측정값 및 액츄에이터에 대한 설정값은, 제어 시스템의 네트워크에 대한 해킹 등을 통해 수집될 수 있으며, 공격용 측정값 및 공격용 설정값 역시, 제어 시스템의 네트워크에 대한 해킹 등을 통해 제어 시스템으로 전달될 수 있다.

[0026] 따라서, 제어 시스템의 컨트롤러(120)는 센서(110)의 측정값이 아닌 공격용 측정값에 따라서, 액츄에이터(130)를 위한 설정값을 결정하게 됨으로써, 제어 시스템에 대한 공격이 이루어질 수 있다. 또한 제어 시스템의 액츄에이터(130)는 컨트롤러(120)에 의해 결정된 설정값이 아닌 공격용 설정값에 따라서 동작을 수행하게 됨으로써, 제어 시스템에 대한 공격이 이루어질 수 있다.

[0027] 본 발명의 일실시예에 따른 제어 시스템 공격 방법은, 제어 시스템에 대한 공격을 탐지하고 방어하는 방법을 개발하는데 이용될 수 있다.

[0029] 도 2는 본 발명의 일실시예에 따른 제어 시스템 공격 방법을 설명하기 위한 도면이며, 도 3은 본 발명의 일실시예에 따른 강화 학습 모델을 설명하기 위한 도면이다.

[0030] 본 발명의 일실시예에 따른 제어 시스템 공격 방법은 메모리 및 프로세서를 포함하는 컴퓨팅 장치에서 수행될 수 있으며, 전술된 제어 시스템 공격 장치는 이러한 컴퓨팅 장치의 일예이다.

[0031] 도 2를 참조하면 본 발명의 일실시예에 따른 제어 시스템 공격 장치는 타겟 장치에 대한 제어 시스템에 포함된 센서의 측정값 및 제어 시스템에 포함된 액츄에이터에 대한 설정값 중 적어도 하나를 수집(S210)한다. 그리고 미리 학습된 강화 학습 모델을 이용하여, 제어 시스템을 공격하기 위한 공격용 측정값 및 공격용 설정값 중 적어도 하나를 결정(S220)한다.

[0032] 제어 시스템 공격 장치는, 측정값에 따라서 설정값을 결정하는 제어 시스템의 컨트롤러로, 공격용 측정값을 전송함으로써, 제어 시스템에 대한 공격을 수행하거나, 또는 설정값에 따라서 동작하는 액츄에이터로, 공격용 설정값을 전송함으로써, 제어 시스템에 대한 공격을 수행할 수 있다. 공격용 측정값 및 공격용 설정값은 전술된 바와 같이, 제어 시스템 네트워크에 대한 해킹 등을 통해 제어 시스템으로 전송될 수 있으며, 센서의 측정값 및 액츄에이터에 대한 설정값 역시, 제어 시스템 네트워크에 대한 해킹 등을 통해 수집될 수 있다.

[0033] 강화 학습 모델은 일실시예로서 도 3의 액터-크리틱(actor-critic) 모델일 수 있다. 강화 학습 모델에서 에이전트(RL Agent)는, 제어 시스템 공격 장치에 대응되며, 환경(Environment)은 제어 시스템에 대응된다. 그리고 에이전트는 액터와 크리틱으로 구성될 수 있다. 에이전트, 즉 제어 시스템 공격 장치는 현재 제어 시스템의 상태(state)를 관찰(observation)하며, 관찰된 상태에서 보상(reward)이 최대가 되는 액션(action)을 수행한다. 여

기서, 액션은 공격용 측정값 및 공격용 설정값 중 적어도 하나를 결정하는 것에 대응되며, 액터와 크리틱은, TD(Temporal Difference) 에러를 통해 업데이트될 수 있다.

[0034] 여기서, 강화 학습 모델의 상태 정보는 제1시점에서의 측정값 및 설정값 중 적어도 하나를 포함하며, 강화 학습 모델의 액션 정보는 제1시점의 다음 시점인 제2시점에서의 공격용 측정값 및 공격용 설정값 중 적어도 하나를 포함한다. 즉 강화 학습 모델은 현재 시점에서의 제어 시스템의 측정값 및 설정값 중 적어도 하나를 입력받아, 제어 시스템에 대한 다음 시점에서의 공격용 측정값 및 공격용 설정값 중 적어도 하나를 출력할 수 있다. 다음 시점에서 생성된 센서의 측정값과 액추에이터에 대한 설정값 대신, 공격용 측정값 및 공격용 설정값이 컨트롤러 및 액츄에이터 각각으로 제공됨으로써, 공격이 이루어질 수 있다.

[0035] 제어 시스템의 센서의 개수를 N_s라고 하고, 액츄에이터의 개수를 N_a라고 할 때, 현재 시점 T_i에서의 상태 정보(S)는 [수학식 1]과 같이 표현될 수 있다.

수학식 1

[0036] $$S = \{s_1, s_2, \cdots, s_l, \cdots, s_{N_s}, a_1, a_2, \cdots, a_m, \cdots, a_{N_a}\}$$

[0037] 여기서, l은 센서의 인덱스를 나타내며, m은 액츄에이터의 인덱스를 나타낸다.

[0038] 그리고 다음 시점 T_{i+1}에서의 액션 정보는 [수학식 2]와 같이 표현될 수 있다.

수학식 2

[0039] $$\mathring{A} = \{s'_1, s'_2, \cdots, s'_l, \cdots, s'_{N_s}, a'_1, a'_2, \cdots, a'_m, \cdots, a'_{N_a}\}$$

[0040] 여기서, s'_l은 제어 시스템 공격 장치에 의해 조작된 공격용 측정값을 나타내며, a'_m은 제어 시스템 공격 장치에 의해 조작된 공격용 설정값을 나타낸다.

[0041] 강화 학습 모델의 학습을 위한 보상값은 타겟 장치의 피해 정도 및 제어 시스템을 위한 공격 탐지 모델의 공격 탐지 여부에 따라서 결정될 수 있다. 보상값은 피해 정도에 비례하도록 결정될 수 있으며, 또한 공격 탐지가 회피될 수 있도록 결정될 수 있다. 그리고 피해 정도는 일실시예로서 [수학식 3]과 같이 계산될 수 있다.

수학식 3

[0042] $$\text{피해 정도} = r_l / d_l$$

$$r_l = H_l - L_l$$

$$d_l = \begin{cases} min(|s_l - L_l|, |s_l - H_l|) & L_l < s_l < H_l \\ \delta_l & otherwise \end{cases}$$

[0043] 여기서 , H_l은 센서의 측정값의 정상 범위 중 최대값, L_l은 센서의 측정값의 정상 범위 중 최소값, s_l은 센서의 측정값, δ_l은 미리 설정된 상수값, l은 센서의 인덱스를 나타낸다. 그리고 보상값 계산에 이용되는 센서의 측정값은 T_{i+2} 시점에서의 센서의 측정값일 수 있다.

[0044] 예컨대, 물탱크의 수위를 측정하는 센서의 측정값이 1.1m이고, 물탱크 수위의 정상 범위가 1m(최소값)에서 2m

(최대값)라면, d_l은 0.1로 계산될 수 있고, 피해 정도는 10으로 계산될 수 있다.

[0045] 만일 복수의 센서가 제어 시스템에 포함된 경우, 보상값은 센서 별로 계산된 피해 정도의 합으로 결정될 수 있다.

[0046] 그리고 상수값은 공격 탐지 모델에 의해, 공격용 측정값 및 공격용 설정값을 이용한 공격이 탐지된 경우의 상수값일 수 있으며, 제어 시스템 공격 장치가 공격 탐지를 회피할 수 있는 공격용 측정값과 공격용 설정값을 결정할 수 있도록, 센서의 측정값이 정상범위 내에 위치할 때의 최대 d_l보다 큰 값으로 설정될 수 있다.

[0047] 상수값은 일실시예로서 센서의 측정값의 정상 범위 중 최대값과 정상 범위중 최소값의 차이값의 절반보다 큰 값으로 설정될 수 있다. 전술된 예시에서, 최대값과 최소값의 차이값은 1이므로, 상수값은 0.5보다 큰 값으로 결정될 수 있다.

[0049] 도 4는 본 발명의 일실시예에 따른 강화 학습 모델의 업데이트 방법을 설명하기 위한 도면이다.

[0050] 본 발명의 일실시예에 따른 제어 시스템 공격 장치는, 강화 학습 모델의 가치 함수와 정책 함수 딥러닝 네트워크에 대해 업데이트를 수행한다. 즉, 제어 시스템 공격 장치는 강화 학습 모델을 이용해 공격용 측정값 및 공격용 설정값을 결정하면서, 수집된 데이터, 즉 센서의 측정값 및 액츄에이터에 대한 설정값을 이용하여 강화 학습 모델에 대한 학습을 수행하여, 딥러닝 네트워크를 업데이트한다.

[0051] 본 발명의 일실시예에 따른 강화 학습 모델에서 에이전트의 액터는 정책 함수를 이용하여 액션, 즉 공격용 측정값 및 공격용 설정값의 결정을 수행한다. 그리고 에이전트의 크리틱(비평가)은 가치 함수를 이용하여 액터의 액션, 즉 공격용 측정값 및 공격용 설정값의 결정이 적절한지 여부를 평가한다. 정책 함수와 가치 함수는 딥러닝 네트워크로 구현될 수 있으며, 가치 함수의 평가 결과가, 정책 함수 딥러닝 네트워크의 파라미터(가중치) 업데이트에 이용된다. 그리고 가치 함수 딥러닝 네트워크는, 액션의 수행 결과에 대한 보상에 따라 파라미터가 업데이트된다.

[0052] 도 4를 참조하면, 제어 시스템 공격 장치는 정책 함수 및 가치 함수에 대한 딥러닝 네트워크를 이용하여 정책 함수 및 가치 함수의 출력값을 획득(S410)한다.

[0053] 정책 함수($\pi^{\theta}(S,A)$)는 특정 상태(S)에서 특정 액션(A)이 결정될 확률을 계산하기 위한 함수로서, 여기서 θ는 정책 함수 딥러닝 네트워크의 가중치를 나타낸다. 정책 함수 딥러닝 네트워크는 상태 정보를 입력받아, 입력된 상태 정보에 대해서 특정 액션을 결정할 확률에 대한, 평균($\mu(S)$)과 표준 편차($\sigma(S)$)를 출력하도록 학습된 네트워크이다. 제어 시스템 공격 장치는 정책 함수 딥러닝 네트워크의 출력값으로부터 정규 분포($\mathbb{N}(A|\mu(S),\sigma(S))$)를 생성하여, 특정 액션을 결정할 확률을 계산할 수 있다. 예컨대, 전술된 업데이트 정보 중 가장 높은 확률에 대응되는 정보에 따라서, 제어 시스템 공격 장치는, 공격용 측정값 및 공격용 설정값을 결정할 수 있다.

[0054] 가치 함수($V^{\nu}(S)$)는 입력된 상태에 대한 보상의 기대값을 계산하기 위한 함수이며, 여기서 ν는 가치 함수 딥러닝 네트워크의 가중치를 나타낸다. 가치 함수 딥러닝 네트워크는 상태 정보를 입력받아, 입력된 상태 정보에 대한 보상 기대값을 출력하도록 학습된 네트워크이다.

[0055] 제어 시스템 공격 장치는 정책 함수 및 가치 함수 딥러닝 네트워크의 출력값을 손실 함수에 적용하여, 손실 함수의 손실값이 최소가 되도록 정책 함수 및 가치 함수 딥러닝 네트워크의 가중치를 업데이트(S420)한다.

[0056] 제어 시스템 공격 장치는 일실시예로서, 시간차 학습(Temporal Difference Learning)을 이용해, 정책 함수 및 가치 함수 딥러닝 네트워크의 가중치를 업데이트할 수 있다. 시간 t 시점에서 상태를 S_t 라 하고 해당 시점에서의 보상을 Γ_t 라고 할 때, $V^{\nu}(S_t)$ 의 실제 값에 대한 추정치 TD 타겟(Temporal Difference target, y_t)은 [수학식 4]를 통해 계산될 수 있다.

수학식 4

[0057] $$y_t = \Gamma_t + \gamma \cdot V^{\nu}(S_{t+1})$$

[0058] 여기서, γ 은 감가율(Discount Factor)을 나타낸다.

[0059] 가치 함수 딥러닝 네트워크는 [수학식 5]와 같이 계산되는 TD 에러값(ε_t)이 최소가 되도록 업데이트될 수 있다.

수학식 5

[0060] $$\varepsilon_t = \left(y_t - V^{\nu}(S_t)\right)^2$$

[0061] 그리고 정책 함수 딥러닝 네트워크는 [수학식 6]과 같이 표현되는 손실 함수의 손실값이 최소가 되도록 업데이트될 수 있다.

수학식 6

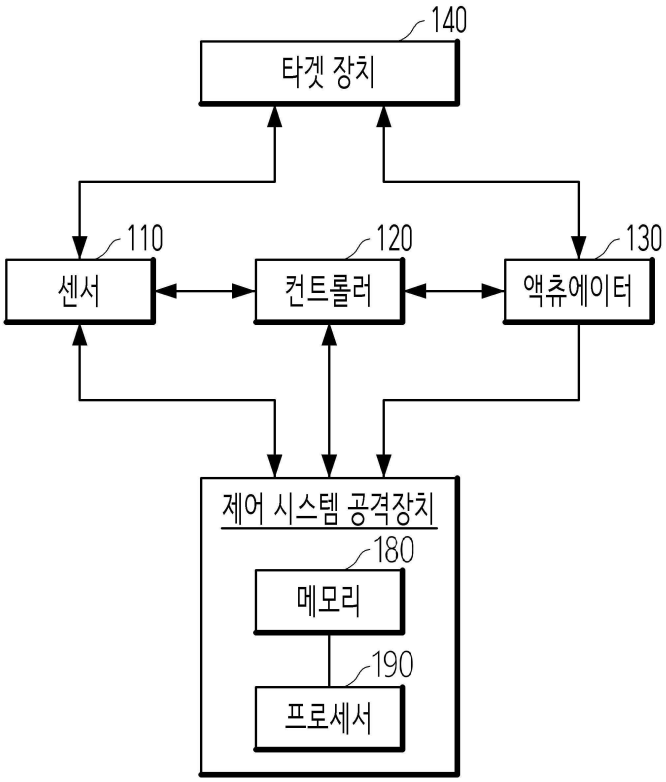
[0062] $$Loss = -\log(\mathbb{N}(A|\mu(S_t), \sigma(S_t))) \cdot \varepsilon_t$$

[0064] 앞서 설명한 기술적 내용들은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예들을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다. 하드웨어 장치는 실시예들의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

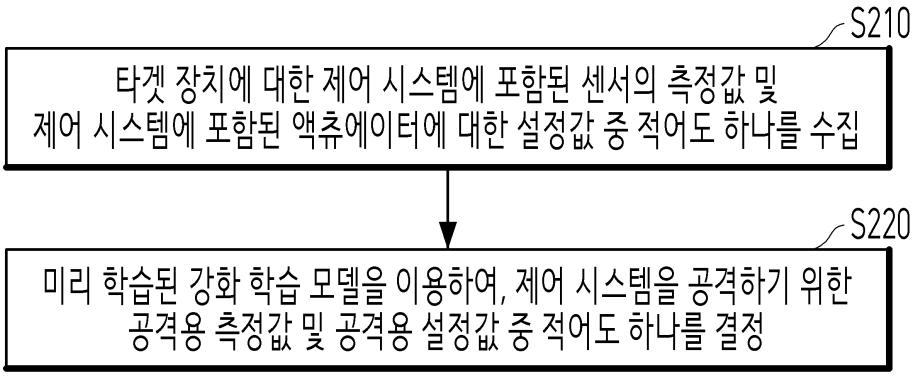
[0066] 이상과 같이 본 발명에서는 구체적인 구성 요소 등과 같은 특정 사항들과 한정된 실시예 및 도면에 의해 설명되었으나 이는 본 발명의 보다 전반적인 이해를 돕기 위해서 제공된 것일 뿐, 본 발명은 상기의 실시예에 한정되는 것은 아니며, 본 발명이 속하는 분야에서 통상적인 지식을 가진 자라면 이러한 기재로부터 다양한 수정 및 변형이 가능하다. 따라서, 본 발명의 사상은 설명된 실시예에 국한되어 정해져서는 아니되며, 후술하는 특허청구범위뿐 아니라 이 특허청구범위와 균등하거나 등가적 변형이 있는 모든 것들은 본 발명 사상의 범주에 속한다고 할 것이다.

도면

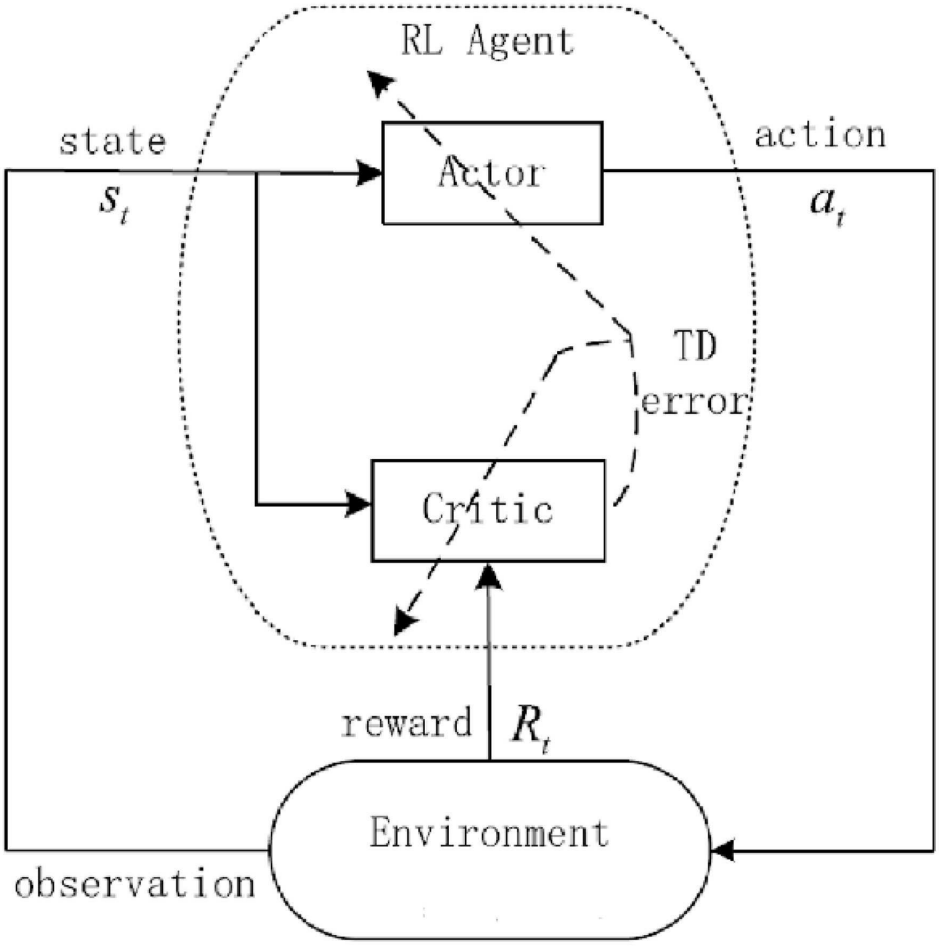
도면1


140
타겟 장치
110
센서
120
컨트롤러
130
액츄에이터
제어 시스템 공격장치
180
메모리
190
프로세서

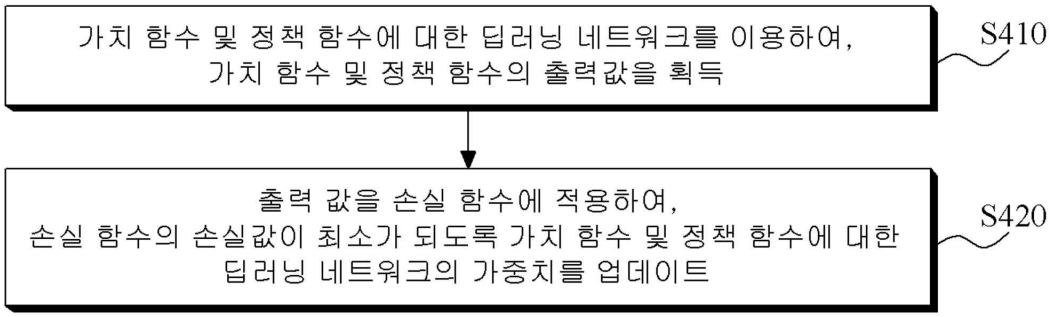
도면2


S210
타겟 장치에 대한 제어 시스템에 포함된 센서의 측정값 및
제어 시스템에 포함된 액츄에이터에 대한 설정값 중 적어도 하나를 수집
S220
미리 학습된 강화 학습 모델을 이용하여, 제어 시스템을 공격하기 위한
공격용 측정값 및 공격용 설정값 중 적어도 하나를 결정

도면3



도면4



【심사관 직권보정사항】

【직권보정 1】

【보정항목】 청구범위

【보정세부항목】 청구항 3

【변경전】

제 1항에 있어서,

상기 보상값은

상기 피해 정도에 비례하도록 결정되며,

상기 피해 정도는

하기 [수학식]에 따라 계산되는

제어 시스템 공격 방법.

[수학식]

$$\text{피해 정도} = r_l / d_l$$

$$r_l = H_l - L_l$$

$$d_l = \begin{cases} min(|s_l - L_l|, |s_l - H_l|) & L_l < s_l < H_l \\ \delta_l & otherwise \end{cases}$$

여기서, H_l은 상기 측정값의 정상 범위 중 최대값, L_l은 상기 측정값의 정상 범위 중 최소값, 상기 s_l은 상기 측정값, δ_l은 미리 설정된 상수값, l은 상기 센서의 인덱스를 나타냄.

【변경후】

제 1항에 있어서,

상기 보상값은

상기 피해 정도에 비례하도록 결정되며,

상기 피해 정도는

하기 [수학식]에 따라 계산되는

제어 시스템 공격 방법.

[수학식]

$$\text{피해 정도} = r_l / d_l$$

$$r_l = H_l - L_l$$

$$d_l = \begin{cases} min(|s_l - L_l|, |s_l - H_l|) & L_l < s_l < H_l \\ \delta_l & otherwise \end{cases}$$

여기서, H_l은 상기 측정값의 정상 범위 중 최대값, L_l은 상기 측정값의 정상 범위 중 최소값, s_l은 상기 측정값, δ_l은 미리 설정된 상수값, l은 상기 센서의 인덱스를 나타냄.

【직권보정 2】

【보정항목】 청구범위

【보정세부항목】 청구항 10

【변경전】

제 8항에 있어서,

상기 보상값은

상기 피해 정도에 비례하도록 결정되며,

상기 피해 정도는

하기 [수학식]에 따라 계산되는

제어 시스템 공격 장치.

[수학식]

$$\text{피해 정도} = r_l / d_l$$

$$r_l = H_l - L_l$$

$$d_l = \begin{cases} min(|s_l - L_l|, |s_l - H_l|) & L_l < s_l < H_l \\ \delta_l & otherwise \end{cases}$$

여기서, H_l은 상기 측정값의 정상 범위 중 최대값, L_l은 상기 측정값의 정상 범위 중 최소값, 상기 s_l은 상기 측정값, δ_l은 미리 설정된 상수값, l은 상기 센서의 인덱스를 나타냄.

【변경후】

제 8항에 있어서,

상기 보상값은

상기 피해 정도에 비례하도록 결정되며,

상기 피해 정도는

하기 [수학식]에 따라 계산되는

제어 시스템 공격 장치.

[수학식]

$$\text{피해 정도} = r_l / d_l$$

$$r_l = H_l - L_l$$

$$d_l = \begin{cases} min(|s_l - L_l|, |s_l - H_l|) & L_l < s_l < H_l \\ \delta_l & otherwise \end{cases}$$

여기서, H_l은 상기 측정값의 정상 범위 중 최대값, L_l은 상기 측정값의 정상 범위 중 최소값, s_l은 상기 측정값, δ_l은 미리 설정된 상수값, l은 상기 센서의 인덱스를 나타냄.