



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2024년11월29일

(11) 등록번호 10-2736334

(24) 등록일자 2024년11월26일

(51) 국제특허분류(Int. Cl.)
H04L 9/40 (2022.01) G06N 3/045 (2023.01)(52) CPC특허분류
H04L 63/1416 (2013.01)
G06N 3/045 (2023.01)

(21) 출원번호 10-2024-0040363

(22) 출원일자 2024년03월25일

심사청구일자 2024년03월25일

(56) 선행기술조사문헌

KR1020210153784 A*

KR1020220079683 A*

KR102648565 B1*

*는 심사관에 의하여 인용된 문헌

(73) 특허권자

세종대학교산학협력단

서울특별시 광진구 능동로 209 (군자동, 세종대학교)

(72) 발명자

김문석

서울특별시 광진구 능동로 209 (군자동)

조부성

서울특별시 광진구 능동로 209 (군자동)

(74) 대리인

수안특허법인

전체 청구항 수 : 총 11 항

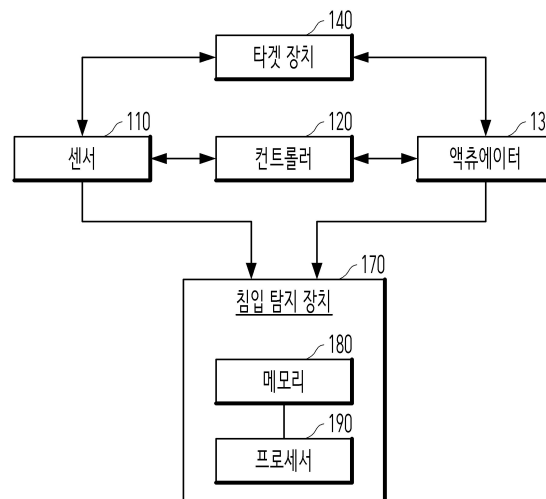
심사관 : 문형섭

(54) 발명의 명칭 강화 학습을 이용하는 제어 시스템에 대한 침입 탐지 방법

(57) 요약

강화 학습 기반으로, 제어 시스템에 대한 침입을 탐지하는 방법이 개시된다. 개시된 제어 시스템에 대한 침입 탐지 방법은 타겟 장치에 대한 제어 시스템에 포함된 센서의 측정값 및 상기 제어 시스템에 포함된 액추에이터에 대한 설정값을 수집하는 단계; 및 미리 학습된 제1강화 학습 모델을 이용하여, 상기 측정값에 대한 변조 공격을 탐지하는 단계를 포함하며, 상기 제1강화 학습 모델의 상태 정보는 현재 시점에서의 상기 측정값, 상기 설정값 및 이전 시점에서 생성된 상기 현재 시점에 대한 예측 측정값을 포함하며, 상기 제1강화 학습 모델의 액션 정보는 상기 예측 측정값 및 상기 변조 공격의 탐지 여부를 나타내는 판별값을 포함한다.

대표도 - 도1



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

명세서

청구범위

청구항 1

타겟 장치에 대한 제어 시스템에 포함된 센서의 측정값 및 상기 제어 시스템에 포함된 액추에이터에 대한 설정값을 수집하는 단계; 및

미리 학습된 제1강화 학습 모델을 이용하여, 상기 측정값에 대한 변조 공격을 탐지하는 단계를 포함하며,

상기 제1강화 학습 모델의 상태 정보는

현재 시점에서의 상기 측정값, 상기 설정값 및 이전 시점에서 생성된 상기 현재 시점에 대한 예측 측정값을 포함하며,

상기 제1강화 학습 모델의 액션 정보는

상기 예측 측정값 및 상기 변조 공격의 탐지 여부를 나타내는 판별값을 포함하며,

상기 제1강화 학습 모델의 학습을 위한 보상값은

상기 측정값의 변조 정도를 나타내는 피해 반경값에 따라 결정되며,

상기 피해 반경값은

상기 측정값 및 상기 예측 측정값의 차이에 따라 결정되는

제어 시스템에 대한 침입 탐지 방법.

청구항 2

삭제

청구항 3

제 1항에 있어서,

상기 피해반경값은

하기 [수학식]에 따라 계산되는

제어 시스템에 대한 침입 탐지 방법.

[수학식]

$$\xi = \sum_{l=1}^{N_s} w_l \times \left| \frac{s_l - s_l^*}{s_l} \right|$$

여기서, ξ 는 상기 피해 반경값, 상기 N_s 는 상기 센서의 개수, s_l 은 1번째 센서의 측정값, s_l^* 은 상기 1번째 센서에 대한 예측 측정값, w_l 은 1번째 센서에 대한 가중치를 나타냄.

청구항 4

제 1항에 있어서,

상기 보상값은

상기 변조 공격이 미발생한 상황에서 상기 변조 공격이 미탐지된 경우, 상기 피해 반경값에 반비례하도록 결정되며,

상기 변조 공격이 오탐지된 경우, 미리 설정된 음수값으로 결정되는

제어 시스템에 대한 침입 탐지 방법.

청구항 5

제 1항에 있어서,

상기 제1강화 학습 모델에 의해 상기 변조 공격이 미탐지되고, 상기 타겟 장치에 대한 모니터링 결과 상기 변조 공격이 확인된 경우, 상기 현재 시점에서의 측정값 및 상기 설정값을 상기 제1강화 학습 모델에 대한 훈련 데이터로 저장하는 단계

를 더 포함하는 제어 시스템에 대한 침입 탐지 방법.

청구항 6

제 1항에 있어서,

미리 학습된 제2강화 학습 모델을 통해 생성된 상기 제어 시스템을 공격하기 위한 공격용 측정값 및 공격용 설정값을 상기 상태 정보로 이용하여, 상기 측정값에 대한 변조 공격을 탐지하는 단계; 및

상기 공격용 측정값 및 공격용 설정값 중 상기 변조 공격을 위한 측정값 및 설정값으로 미탐지된 공격용 측정값 및 공격용 설정값을 이용하여, 상기 제1강화 학습 모델에 대한 학습을 수행하는 단계

제어 시스템에 대한 침입 탐지 방법.

청구항 7

타겟 장치에 대한 제어 시스템에 포함된 센서의 측정값 및 상기 제어 시스템에 포함된 액추에이터에 대한 설정값을 수집하는 단계; 및

미리 학습된 제1강화 학습 모델을 이용하여, 상기 설정값에 대한 변조 공격을 탐지하는 단계를 포함하며,

상기 제1강화 학습 모델의 상태 정보는

현재 시점에서의 상기 측정값, 상기 설정값, 상기 타겟 장치에 대한 피해 정도를 나타내는 피해비용값 및 이전 시점에서 생성된 상기 현재 시점에 대한 피해비용 예측값을 포함하며,

상기 제1강화 학습 모델의 액션 정보는

상기 피해비용 예측값 및 상기 변조 공격의 탐지 여부를 나타내는 판별값을 포함하는

제어 시스템에 대한 침입 탐지 방법.

청구항 8

제 7항에 있어서,

상기 제1강화 학습 모델의 학습을 위한 보상값은

상기 피해비용값 및 상기 피해비용 예측값의 차이에 따라서 결정되는 제어 시스템에 대한 침입 탐지 방법.

청구항 9

제 8항에 있어서,

상기 피해비용값은

하기 [수학식]에 따라 계산되는,

제어 시스템에 대한 침입을 탐지하는 방법.

[수학식]

$$\text{피해 비용값} = r_l / d_l$$

$$r_l = H_l - L_l$$

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

여기서, H_l 은 상기 측정값의 정상 범위 중 최대값, L_l 은 상기 측정값의 정상 범위 중 최소값, 상기 s_l 은 상기 측정값, δ_l 은 미리 설정된 상수값, l 은 상기 센서의 인덱스를 나타냄.

청구항 10

제 8항에 있어서,

상기 보상값은

변조 공격 여부와 상기 판별값이 대응될 경우, 상기 피해비용값 및 상기 피해비용 예측값의 차이에 반비례하도록 결정되며,

상기 변조 공격이 오탐지된 경우, 미리 설정된 음수값으로 결정되는

제어 시스템에 대한 침입 탐지 방법.

청구항 11

제 7항에 있어서,

상기 제1강화 학습 모델에 의해 상기 변조 공격이 미탐지되고, 상기 타겟 장치에 대한 모니터링 결과 상기 변조 공격이 확인된 경우, 상기 현재 시점에서의 측정값 및 상기 설정값을 훈련 데이터로 저장하는 단계

를 더 포함하는 제어 시스템에 대한 침입 탐지 방법.

청구항 12

제 7항에 있어서,

미리 학습된 제2강화 학습 모델을 통해 생성된 상기 제어 시스템을 공격하기 위한 공격용 측정값 및 공격용 설정값을 상기 상태 정보로 이용하여, 상기 설정값에 대한 변조 공격을 탐지하는 단계; 및

상기 공격용 측정값 및 공격용 설정값 중 상기 변조 공격을 위한 측정값 및 설정값으로 미탐지된 공격용 측정값 및 공격용 설정값을 이용하여, 상기 제1강화 학습 모델에 대한 학습을 수행하는 단계를 포함하는 제어 시스템에 대한 침입 탐지 방법.

발명의 설명

기술 분야

[0001] 본 발명은 제어 시스템에 대한 침입을 탐지하는 방법에 관한 것으로서, 더욱 상세하게는 강화 학습 기반으로, 제어 시스템에 대한 침입을 탐지하는 방법에 관한 것이다.

배경 기술

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

[0004] 최근 인공지능 기술의 발달과 함께, 제어 시스템에 대한 침입을 탐지하기 위해 인공지능 기술이 활용되고 있다.

[0005] 관련 선행문헌으로 특허문헌인 대한민국 등록특허 제10-2100328호, 제10-1538709호, 대한민국 공개특허 제2022-0079683호가 있으며, 비특허문헌으로 "방어 자산의 가용성 상태를 활용한 강화학습 기반 APT 공격 대응 기법, 김형록, 최창희, 정보보호학회논문지 제33권 제6호, 2013.12"가 있다.

발명의 내용

해결하려는 과제

[0007] 본 발명은, 강화 학습을 이용하여, 제어 시스템에 대한 침입을 탐지하는 방법을 제공하기 위한 것이다.

[0008] 특히 본 발명은, 강화 학습을 이용하여, 제어 시스템에 포함된 센서의 측정값에 대한 변조 공격 또는 제어 시스템에 포함된 액추에이터의 설정값에 대한 변조 공격을 탐지하는 방법을 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 액추에이터에 대한 설정값을 수집하는 단계; 및 미리 학습된 제1강화 학습 모델을 이용하여, 상기 측정값에 대한 변조 공격을 탐지하는 단계를 포함하며, 상기 제1강화 학습 모델의 상태 정보는 현재 시점에서의 상기 측정값, 상기 설정값 및 이전 시점에서 생성된 상기 현재 시점에 대한 예측 측정값을 포함하며, 상기 제1강화 학습 모델의 액션 정보는 상기 예측 측정값 및 상기 변조 공격의 탐지 여부를 나타내는 판별값을 포함하는 제어 시스템에 대한 침입 탐지 방법이 제공된다.

[0011] 또한 상기한 목적을 달성하기 위한 본 발명의 다른 실시예에 따르면, 시스템에 포함된 액추에이터에 대한 설정값을 수집하는 단계; 및 미리 학습된 제1강화 학습 모델을 이용하여, 상기 설정값에 대한 변조 공격을 탐지하는 단계를 포함하며, 상기 제1강화 학습 모델의 상태 정보는 현재 시점에서의 상기 측정값, 상기 설정값, 상기 타겟 장치에 대한 피해 정도를 나타내는 피해비용값 및 이전 시점에서 생성된 상기 현재 시점에 대한 피해비용 예측값을 포함하며, 상기 제1강화 학습 모델의 액션 정보는 상기 피해비용 예측값 및 상기 변조 공격의 탐지 여부를 나타내는 판별값을 포함하는 제어 시스템에 대한 침입 탐지 방법이 제공된다.

발명의 효과

[0013] 본 발명의 일실시예에 따르면, 강화 학습을 통해, 제어 시스템에 포함된 센서의 측정값에 대한 변조 공격 또는 제어 시스템에 포함된 액추에이터의 설정값에 대한 변조 공격이 효율적으로 탐지될 수 있다.

[0014] 또한 본 발명의 일실시예에 따르면, 변조 공격 탐지에 실패한 데이터를 이용해 강화 학습 모델을 추가로 학습함으로써, 제어 시스템에 대한 침입 탐지 성능이 더욱 향상될 수 있다.

[0015] 또한 본 발명의 일실시예에 따르면, 제어 시스템에 대한 공격용 측정값 및 공격용 설정값을 이용해 제어 시스템

을 공격하고, 강화 학습 모델에 대한 학습을 수행함으로써, 제어 시스템에 대한 침입 탐지 성능이 더욱 향상될 수 있다.

도면의 간단한 설명

- [0017] 도 1은 본 발명의 일실시예에 따른 제어 시스템에 대한 침입 탐지 장치를 설명하기 위한 도면이다.
- 도 2는 본 발명의 일실시예에 따른 제어 시스템에 대한 침입 탐지 방법을 설명하기 위한 도면이다.
- 도 3은 본 발명의 일실시예에 따른 강화 학습 모델을 설명하기 위한 도면이다.
- 도 4는 본 발명의 다른 실시예에 따른 제어 시스템에 대한 침입 탐지 방법을 설명하기 위한 도면이다.
- 도 5는 본 발명의 일실시예에 따른 침입 탐지 시스템을 설명하기 위한 도면이다.
- 도 6은 본 발명의 다른 실시예에 따른 침입 탐지 시스템을 설명하기 위한 도면이다.
- 도 7은 본 발명의 일실시예에 따른 강화 학습 모델의 업데이트 방법을 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0018] 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.
- [0019] 이하에서, 본 발명에 따른 실시예들을 첨부된 도면을 참조하여 상세하게 설명한다.
- [0021] 도 1은 본 발명의 일실시예에 따른 제어 시스템에 대한 침입 탐지 장치를 설명하기 위한 도면으로서, 도 1에서는 산업 제어 시스템에 대한 침입을 탐지하는 장치가 일실시예로서 설명된다.
- [0022] 제어 시스템 예컨대, 산업 제어 시스템은 도 1에 도시된 바와 같이, 센서(110), 컨트롤러(120) 및 액추에이터(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 액추에이터(130)를 위한 설정값을 결정하여, 액추에이터(130)로 전송한다. 그리고 액추에이터(130)는 설정값에 따라서 동작을 수행한다. 센서(110)는 주기적으로 측정값을 생성하고, 주기적으로 생성되는 측정값에 따라서, 설정값 역시 주기적으로 생성된다.
- [0023] 예컨대, 타겟 장치(140)는 물탱크, 센서(110)는 물탱크의 수위를 측정하는 수위 센서, 액추에이터(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.
- [0024] 이러한 제어 시스템에 대한 공격자는, 제어 시스템에 침입하여, 센서(110)의 측정값이나 액추에이터(130)의 설정값을 변경 즉 변조하는 공격을 수행할 수 있다. 일례로서 센서(110)의 측정값이 변조됨으로써, 물탱크(140)의 수위가 정상 범위를 벗어난 환경에서도, 물탱크(140)의 물이 펌프(130)에 의해 배수되지 않을 수 있다.
- [0025] 본 발명의 일실시예에 따른 침입 탐지 장치(170)는, 이러한 제어 시스템에 대한 침입을 탐지하며, 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 침입을 탐지한다.
- [0026] 프로세서(190)는 타겟 장치(140)에 대한 제어 시스템에 포함된 센서(110)의 측정값 및 제어 시스템에 포함된 액추에이터(130)에 대한 설정값을 수집하고, 수집된 데이터를 미리 학습된 강화 학습 모델에 입력하여, 제어 시스템에 대한 침입을 탐지할 수 있다. 프로세서(190)는 센서(110)의 측정값에 대한 변조 공격이나, 액추에이터(130)의 설정값에 대한 변조 공격을 탐지할 수 있으며, 구체적인 침입 탐지 방법은 도 2 내지 도 4에서 자세히 설명된다.

- [0027] 한편, 강화 학습 모델은 학습 단계에서 훈련 데이터를 통해 학습되며, 훈련 데이터를 통해 학습된 강화 학습 모델은, 침입 탐지 단계에서 별도의 훈련없이 제어 시스템에 대한 침입 탐지에 이용될 수 있다.
- [0028] 실시예에 따라서, 강화 학습 모델의 학습을 위한 훈련 데이터는, 변조 공격을 모니터링하는 별도의 모니터링 장치의 모니터링 결과 또는 제어 시스템 공격 장치를 이용한 제어 시스템 공격을 통해, 획득될 수 있으며, 모니터링 장치와 제어 시스템 공격 장치는 도 5 및 도 6에서 자세히 설명된다.
- [0030] 도 2는 본 발명의 일실시예에 따른 제어 시스템에 대한 침입 탐지 방법을 설명하기 위한 도면이며, 도 3은 본 발명의 일실시예에 따른 강화 학습 모델을 설명하기 위한 도면이다.
- [0031] 본 발명의 일실시예에 따른 제어 시스템에 대한 침입 탐지 방법은 메모리 및 프로세서를 포함하는 컴퓨팅 장치에서 수행될 수 있으며, 전송된 침입 탐지 장치는 이러한 컴퓨팅 장치의 일예이다.
- [0032] 도 2를 참조하면 본 발명의 일실시예에 따른 침입 탐지 장치는 타겟 장치에 대한 제어 시스템에 포함된 센서의 측정값 및 제어 시스템에 포함된 액추에이터에 대한 설정값을 수집(S210)한다. 그리고 미리 학습된 강화 학습 모델을 이용하여, 측정값에 대한 변조 공격을 탐지(S220)한다.
- [0033] 강화 학습 모델은 일실시예로서 도 3의 액터-크리틱(actor-critic) 모델일 수 있다. 강화 학습 모델에서 에이전트(RL Agent)는, 침입 탐지 장치에 대응되며, 환경(Environment)은 제어 시스템에 대응된다. 그리고 에이전트는 액터와 크리틱으로 구성될 수 있다. 에이전트, 즉 침입 탐지 장치는 현재 제어 시스템의 상태(state) 정보를 관찰(observation)하며, 관찰된 상태에서 보상(reward)이 최대가 되는 액션(action) 정보를 생성한다. 액터와 크리틱은, TD(Temporal Difference) 에러를 통해 업데이트될 수 있으며, 정책 함수와 가치 함수는 딥러닝 네트워크로 구현될 수 있다.
- [0034] 강화 학습 모델의 상태 정보는 현재 시점에서의 측정값, 현재 시점에서의 설정값을 포함하고, 이전 시점에서 생성된 현재 시점에 대한 예측 측정값을 포함한다. 예측 측정값은, 이전 시점에서 생성된, 현재 시점에서의 센서의 측정값에 대한 예측값을 나타낸다.
- [0035] 이전 시점에서의 측정값과 설정값으로부터 현재 시점에서의 측정값은 용이하게 예측될 수 있으므로, 본 발명의 일실시예는 측정값에 대한 변조 공격을 탐지하게 위해 예측 측정값을 이용한다. 일례로서, 이전 시점에서의 물탱크에 대한 수위 측정값이 5m이고, 펌프가 물을 배수하지 않고 있다면, 현재 시점에서의 측정값은 5m 이상일 것으로 용이하게 예측될 수 있다.
- [0036] 현재 시점(T_i)에서의 상태 정보(S)는 일실시예로서, [수학식 1]과 같이 표현될 수 있다.

수학식 1

$$S = \{s_1, s_2, \dots, s_l, \dots, s_{N_s}, a_1, a_2, \dots, a_m, \dots, a_{N_a}, s^*_1, s^*_2, \dots, s^*_l, \dots, s^*_{N_s}\}$$

- [0037]
- [0038] 여기서, s_l 은 1번째 센서의 측정값, a_m 은 m번째 액추에이터의 설정값, s^*_l 은 1번째 센서에 대한 예측 측정값, l은 센서에 대한 인덱스, m은 액추에이터에 대한 인덱스, N_s 는 센서의 개수, N_a 는 액추에이터의 개수를 나타낸다.
- [0039] 예측 측정값은 이전 시점(T_{i-1})에서 생성된 액션 정보로부터 제공되며, T_i 가 0인 경우, 즉 예측 측정값이 존재하지 않는 경우, 측정값이 예측 측정값으로 이용될 수 있다.
- [0040] 강화 학습 모델의 액션 정보($\hat{A}$)는 측정값에 대한 변조 공격의 탐지 여부를 나타내는 판별값 및 예측 측정값을 포함하며, 현재 시점(T_i)에서의 액션 정보는 일실시예로서 [수학식 2]와 같이 표현될 수 있다.

수학식 2

$$\mathbf{\hat{A}} = \{s^*_{1}, s^*_{2}, \dots, s^*_{l}, \dots, s^*_{N_s}, \chi\}$$

여기서, χ 는 판별값을 나타내며, 이진값으로 표현될 수 있다. 예컨대, 변조 공격이 탐지된 경우, 판별값은 1, 변조 공격이 미탐지된 경우 판별값은 0일 수 있다. 그리고 현재 시점에서 생성된 예측 측정값은, 다음 시점(T_{i+1})에서의 상태 정보로 이용된다.

강화 학습 모델의 학습을 위한 보상값은 측정값의 변조 정도를 나타내는 피해 반경값에 따라 결정되며, 피해 반경값은 측정값 및 예측 측정값의 차이에 따라 결정될 수 있다. 그리고 일실시예로서 피해 반경값(ξ)은 [수학식 3]과 같이 계산될 수 있다.

수학식 3

$$\xi = \sum_{l=1}^{N_s} w_l \times \left| \frac{s_l - s^*_l}{s_l} \right|$$

여기서, w_l 은 1번째 센서에 대한 가중치를 나타내며, 센서별 가중치 즉, N_s 개의 가중치의 총합이 1이 되도록, 센서에 가중치가 할당될 수 있다.

그리고 보상값은 측정값과 예측 측정값의 차이가 최소가 되면서 변조 공격이 탐지될 수 있도록, 변조 공격이 미발생한 상황에서 침입 탐지 장치에 의해 변조 공격이 미탐지된 경우에는, 피해 반경값에 반비례하도록 결정된다. 그리고 침입 탐지 장치에 의해 변조 공격이 오탐지된 경우에는, 미리 설정된 음수값으로 보상값이 결정될 수 있다. 또한 변조 공격이 발생한 상황에서 침입 탐지 장치에 의해 변조 공격이 탐지된 경우에는, 측정값이 비정상 상태임을 감안하여, 보상값이 0으로 결정될 수 있다.

일예로서, 변조 공격이 미발생한 상황에서 침입 탐지 장치에 의해 변조 공격이 미탐지된 경우의 보상값은, $\exp(-\xi)$ 로 결정될 수 있다.

도 4는 본 발명의 다른 실시예에 따른 제어 시스템에 대한 침입 탐지 방법을 설명하기 위한 도면이다.

도 4를 참조하면 본 발명의 일실시예에 따른 침입 탐지 장치는 타겟 장치에 대한 제어 시스템에 포함된 센서의 측정값 및 제어 시스템에 포함된 액추에이터에 대한 설정값을 수집(S410)한다. 그리고 미리 학습된 강화 학습 모델을 이용하여, 설정값에 대한 변조 공격을 탐지(S420)한다. 여기서, 강화 학습 모델은 도 3의 강화 학습 모델일 수 있다.

강화 학습 모델의 상태 정보는 현재 시점에서의 측정값, 현재 시점에서의 설정값, 현재 시점에서의 타겟 장치에 대한 피해 정도를 나타내는 피해비용값을 포함하고, 이전 시점에서 생성된 현재 시점에 대한 피해비용 예측값을 포함한다. 피해비용 예측값은, 이전 시점에서 생성된, 현재 시점에서의 피해비용값에 대한 예측값을 나타낸다.

측정값에 대한 변조 공격을 탐지하는 경우와 달리, 설정값에 대한 변조 공격 탐지를 위해 예측 설정값을 이용하는 것은 적절치 않다. 설정값은 측정값 이외에도 다양한 변수에 의해 결정되며, 설정값이 사용자의 정책에 의한 값인지 공격에 의한 값인지 여부를 판단하기도 쉽지 않기 때문이다. 이에 본 발명의 일실시예는 설정값에 대한 변조 공격 탐지를 위해, 타겟 장치의 피해 정도를 나타내는 피해비용값을 이용한다.

[0053] 현재 시점(T_i)에서의 상태 정보(S)는 일실시예로서, [수학식 4]와 같이 표현될 수 있다.

수학식 4

$$S = \{s_1, s_2, \dots, s_l, \dots, s_{N_s}, a_1, a_2, \dots, a_m, \dots, a_{N_a}, \sum_{l \in I_D} \frac{r_l}{d_l}, (\sum_{l \in I_D} \frac{r_l}{d_l})^* \}$$

[0054]

[0055] 여기서, $\sum_{l \in I_D} \frac{r_l}{d_l}$ 는 피해비용값, $(\sum_{l \in I_D} \frac{r_l}{d_l})^*$ 는 피해비용 예측값, I_D 는 센서의 인덱스 집합을 나타낸다.

[0056] 그리고 센서 각각에 대한 피해비용값은 [수학식 5]와 같이 계산될 수 있으며, 센서별 피해비용값의 합산값이, 상태 정보로 이용될 수 있다.

수학식 5

$$\text{피해 비용값} = r_l/d_l$$

$$r_l = H_l - L_l$$

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

[0057]

[0058] 여기서, H_l 은 센서의 측정값의 정상 범위 중 최대값, L_l 은 센서의 측정값의 정상 범위 중 최소값, δ_l 은 미리 설정된 상수값을 나타낸다.

[0059] 예컨대, 물탱크의 수위를 측정하는 센서의 측정값이 1.1m이고, 물탱크 수위의 정상 범위가 1m(최소값)에서 2m(최대값)라면, d_l 은 0.1로 계산될 수 있고, 피해 비용값은 10으로 계산될 수 있다.

[0060] 피해비용 예측값은 이전 시점(T_{i-1})에서 생성된 액션 정보로부터 제공되며, T_i 가 0인 경우, 즉 피해비용 예측값이 존재하지 않는 경우, 피해비용값이 피해비용 예측값으로 이용될 수 있다.

[0061] 강화 학습 모델의 액션 정보($\hat{A}$)는 피해비용 예측값 및 설정값에 대한 변조 공격의 탐지 여부를 나타내는 판별값을 포함하며, 현재 시점(T_i)에서의 액션 정보는 일실시예로서 [수학식 6]과 같이 표현될 수 있다.

수학식 6

$$\hat{A} = \{(\sum_{l \in I_D} \frac{r_l}{d_l})^*, \psi\}$$

[0062]

[0063] 여기서, ψ 는 판별값을 나타내며, 이진값으로 표현될 수 있다. 예컨대, 변조 공격이 탐지된 경우, 판별값은 1, 변조 공격이 미탐지된 경우 판별값은 0일 수 있다. 그리고 현재 시점에서 생성된 피해비용 예측값은, 다음 시점(T_{i+1})에서의 상태 정보로 이용된다.

[0064] 강화 학습 모델의 학습을 위한 보상값은 피해비용값 및 피해비용 예측값의 차이에 따라서 결정될 수 있다. 보상값은, 피해비용값과 피해비용 예측값의 차이가 최소가 되면서 변조 공격이 탐지될 수 있도록, 변조 공격 상황과

탐지값이 대응될 경우, 피해비용값 및 피해비용 예측값의 차이에 반비례하도록 결정될 수 있다. 즉, 변조 공격이 발생한 상황에서 변조 공격이 탐지되었거나 변조 공격이 미발생한 상황에서 변조 공격이 미탐지된 경우에 보

$$\exp\left(-\left|\sum_{l \in I_D} \frac{r_l}{d_l} - \left(\sum_{l \in I_D} \frac{r_l}{d_l}\right)^*\right|\right)$$

상값은 일실시예로서, 로 결정될 수 있다. 그리고 변조 공격이 오탐지된 경우, 보상값은 미리 설정된 음수값으로 결정될 수 있다.

- [0066] 도 5는 본 발명의 일실시예에 따른 침입 탐지 시스템을 설명하기 위한 도면이다.
- [0067] 도 5를 참조하면, 본 발명의 일실시예에 따른 침입 탐지 시스템은 제어 시스템(510), 침입 탐지 장치(520) 및 모니터링 장치(530)를 포함한다.
- [0068] 침입 탐지 장치(520)는 제어 시스템(510)에 대한 침입을 탐지하며, 강화 학습 모델을 이용하여, 센서의 측정값이나 액추에이터의 설정값에 대한 변조 공격을 탐지한다. 이 때, 침입 탐지 장치(520)는 센서의 측정값에 대한 변조 공격을 탐지하기 위한 제1강화 학습 모델과, 액추에이터의 설정값에 대한 변조 공격을 탐지하기 위한 제2강화 학습 모델을 이용할 수 있다.
- [0069] 모니터링 장치(530)는 제어 시스템(510)의 타겟 장치를 모니터링하여, 침입 탐지 장치(520)와 별도로 변조 공격의 여부를 판단한다. 모니터링 장치(530)는 미리 설정된 규칙 기반으로 변조 공격의 여부를 판단할 수 있으며, 예컨대, 물탱크의 물이 넘치는 상황이 발생한 경우에, 변조 공격이 발생한 것으로 판단할 수 있다.
- [0070] 침입 탐지 장치(520)는 강화 학습 모델에 의해 변조 공격이 미탐지되고, 타겟 장치에 대한 모니터링 결과 변조 공격이 확인된 경우, 현재 시점에서의 측정값 및 설정값을 강화 학습 모델에 대한 훈련 데이터로 저장한다. 측정값 및 설정값은 공격 시점과 함께 훈련 데이터로 저장되며, 저장된 훈련 데이터는 변조 공격이 라벨링된 후, 강화 학습 모델의 학습에 이용될 수 있다. 즉, 측정값 변조 공격이 미탐지된 경우에는 저장된 훈련 데이터가 제1강화 학습 모델의 학습에 이용되며, 설정값 변조 공격이 미탐지된 경우에는 저장된 훈련 데이터가 제2강화 학습 모델의 학습에 이용된다.
- [0071] 침입 탐지 과정에서, 침입 탐지 장치(520)는 제1강화 학습 모델을 이용하여, 제어 시스템에 대한 데이터로부터 1차적으로 센서의 측정값에 대한 변조 공격을 탐지하고, 제2강화 학습 모델을 이용하여, 변조 공격이 탐지되지 않은 나머지 데이터로부터 2차적으로 액추에이터의 설정값에 대한 변조 공격을 탐지할 수 있다.
- [0072] 본 발명의 일실시예에 따르면, 변조 공격 탐지에 실패한 데이터를 이용해 강화 학습 모델을 추가로 학습함으로써, 제어 시스템에 대한 침입 탐지 성능이 더욱 향상될 수 있다.
- [0074] 도 6은 본 발명의 다른 실시예에 따른 침입 탐지 시스템을 설명하기 위한 도면이다.
- [0075] 도 6을 참조하면, 본 발명의 일실시예에 따른 침입 탐지 시스템은 제어 시스템(610), 침입 탐지 장치(620) 및 제어 시스템 공격 장치(630)를 포함한다.
- [0076] 침입 탐지 장치(620)는 제어 시스템(610)에 대한 침입을 탐지하며, 강화 학습 모델을 이용하여, 센서의 측정값이나 액추에이터의 설정값에 대한 변조 공격을 탐지한다.
- [0077] 제어 시스템 공격 장치(630)는 미리 학습된 강화 학습 모델을 이용하여, 제어 시스템(610)을 공격하기 위한 공격용 측정값 및 공격용 설정값을 생성한다. 제어 시스템(610)의 컨트롤러가, 센서의 측정값이 아닌 공격용 측정값에 따라서, 액추에이터를 위한 설정값을 결정하게 됨으로써, 제어 시스템(610)에 대한 공격이 이루어질 수 있다. 또한 제어 시스템(610)의 액추에이터가 컨트롤러에 의해 결정된 설정값이 아닌 공격용 설정값에 따라서 동작을 수행하게 됨으로써, 제어 시스템(610)에 대한 공격이 이루어질 수 있다.
- [0078] 제어 시스템 공격 장치(630)는 제어 시스템(610)에 포함된 센서의 측정값 및 제어 시스템에 포함된 액추에이터에 대한 설정값을 수집한다. 그리고 미리 학습된 공격용 강화 학습 모델을 이용하여, 제어 시스템(610)을 공격하기 위한 공격용 측정값 및 공격용 설정값을 생성한다. 공격용 강화 학습 모델은 도 3의 액터-크리틱(actor-critic) 모델일 수 있다.
- [0079] 여기서 공격용 강화 학습 모델의 상태 정보는 현재 시점(T_i)에서의 측정값 및 설정값을 포함하며, 공격용 강화 학습 모델의 액션 정보는 다음 시점(T_{i+1})에서의 공격용 측정값 및 공격용 설정값을 포함한다. 즉 공격용 강화

학습 모델은 현재 시점(T_i)에서의 제어 시스템(610)의 측정값 및 설정값 중 적어도 하나를 입력받아, 제어 시스템(610)에 대한 다음 시점(T_{i+1})에서의 공격용 측정값 및 공격용 설정값을 출력할 수 있다. 다음 시점(T_{i+1})에서 생성된 센서의 측정값과 액추에이터에 대한 설정값 대신, 공격용 측정값 및 공격용 설정값이 컨트롤러 및 액추에이터 각각으로 제공됨으로써, 공격이 이루어질 수 있다.

[0080] 현재 시점(T_i)에서의 공격용 강화 학습 모델의 상태 정보(S)는 [수학식 7]과 같이 표현될 수 있다.

수학식 7

$$S = \{s_1, s_2, \dots, s_l, \dots, s_{N_s}, a_1, a_2, \dots, a_m, \dots, a_{N_a}\}$$

[0081]

[0082] 그리고 다음 시점(T_{i+1})에서의 공격용 강화 학습 모델의 액션 정보($\hat{A}$)는 [수학식 8]과 같이 표현될 수 있다.

수학식 8

$$\hat{A} = \{s'_1, s'_2, \dots, s'_l, \dots, s'_{N_s}, a'_1, a'_2, \dots, a'_m, \dots, a'_{N_a}\}$$

[0083]

[0084] 여기서, s'_l 은 제어 시스템 공격 장치(630)에 의해 조작된 공격용 측정값을 나타내며, a'_m 은 제어 시스템 공격 장치(630)에 의해 조작된 공격용 설정값을 나타낸다.

[0085] 강화 학습 모델의 학습을 위한 보상값은 타겟 장치의 피해비용값 및 침입 탐지 장치의 침입 탐지 여부에 따라서 결정될 수 있다. 제어 시스템 공격 장치(630)의 공격이, 침입 탐지 장치(620)에 의해 탐지되지 않은 경우, 보상값은 피해비용값에 비례하도록 결정될 수 있으며, 복수의 센서가 제어 시스템에 포함된 경우, 보상값은 센서 별로 계산된 피해비용값의 합산값으로 결정될 수 있다. 제어 시스템 공격 장치(630)의 공격이 침입 탐지 장치(620)에 의해 탐지된 경우, 보상값은 미리 설정된 음수값으로 설정될 수 있다.

[0086] 침입 탐지 장치(620)는, 미리 학습된 공격용 강화 학습 모델을 통해 생성된, 제어 시스템(610)을 공격하기 위한 공격용 측정값 및 공격용 설정값을 상태 정보로 이용하여, 측정값 또는 설정값에 대한 변조 공격을 탐지한다. 그리고 공격용 측정값 및 공격용 설정값 중, 변조 공격을 위한 측정값 및 설정값으로 미탐지된 공격용 측정값 및 공격용 설정값을 이용하여, 제1 및 제2강화 학습 모델에 대한 학습을 수행할 수 있다.

[0087] 본 발명의 일실시예에 따르면, 제어 시스템을 위한 공격용 측정값 및 공격용 설정값을 이용해 제어 시스템에 대한 공격을 수행함으로써, 제어 시스템에 대한 침입 탐지 성능이 더욱 향상될 수 있다.

[0089] 도 7은 본 발명의 일실시예에 따른 강화 학습 모델의 업데이트 방법을 설명하기 위한 도면이다.

[0090] 본 발명의 일실시예에 따른 침입 탐지 장치는, 강화 학습 모델의 가치 함수와 정책 함수의 딥러닝 네트워크에 대해 업데이트를 수행한다. 즉, 침입 탐지 장치는 제어 시스템으로부터 수집된 데이터, 즉 센서의 측정값 및 액추에이터에 대한 설정값과, 예측값 등을 이용하여 강화 학습 모델에 대한 학습을 수행하여, 딥러닝 네트워크를 업데이트한다.

[0091] 본 발명의 일실시예에 따른 강화 학습 모델에서 에이전트의 액터는 정책 함수를 이용하여 액션 정보를 생성한다. 그리고 에이전트의 크리틱(비평가)은 가치 함수를 이용하여 액터의 액션, 즉 액션 정보가 적절한지 여부를 평가한다. 정책 함수와 가치 함수는 전술된 바와 같이, 딥러닝 네트워크로 구현될 수 있으며, 가치 함수의 평가 결과가, 정책 함수 딥러닝 네트워크의 파라미터(가중치) 업데이트에 이용된다. 그리고 가치 함수 딥러닝 네트워크의 파라미터는, 액션의 수행 결과에 대한 보상에 따라서, 업데이트된다.

[0092] 도 7을 참조하면, 침입 탐지 장치는 정책 함수 및 가치 함수에 대한 딥러닝 네트워크를 이용하여 정책 함수 및 가치 함수의 출력값을 획득(S710)한다.

[0093] 정책 함수($\pi^\theta(S, A)$)는 특정 상태(S)에서 특정 액션(A)이 결정될 확률을 계산하기 위한 함수로서, 여기서 θ 는 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.

[0094] 침입 탐지 장치는 정책 함수 딥러닝 네트워크의 출력값으로부터 정규 분포($N(A|\mu(S), \sigma(S))$)를 생성하여, 특정 액션을 결정할 확률을 계산할 수 있다. 예컨대, 전송된 업데이트 정보 중 가장 높은 확률에 대응되는 정보에 따라서, 침입 탐지 장치는, 측정 예측값, 피해비용 예측값, 판별값 등을 결정할 수 있다.

[0095] 가치 함수($V^v(S)$)는 입력된 상태에 대한 보상의 기대값을 계산하기 위한 함수이며, 여기서 v 는 가치 함수 딥러닝 네트워크의 가중치를 나타낸다. 가치 함수 딥러닝 네트워크는 상태 정보를 입력받아, 입력된 상태 정보에 대한 보상 기대값을 출력하도록 학습된 네트워크이다.

[0096] 침입 탐지 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(S720)한다.

[0097] 침입 탐지 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 보상

을 Γ_t 라고 할 때, $V^v(S_t)$ 의 실제 값에 대한 추정치 TD 타겟(Temporal Difference target, y_t)은 [수학식 9]를 통해 계산될 수 있다.

수학식 9

[0098]
$$y_t = \Gamma_t + \gamma \cdot V^v(S_{t+1})$$

[0099] 여기서, γ 은 감가율(Discount Factor)을 나타낸다.

[0100] 가치 함수 딥러닝 네트워크는 [수학식 10]과 같이 계산되는 TD 에러값(ϵ_t)이 최소가 되도록 업데이트될 수 있다.

수학식 10

[0101]
$$\epsilon_t = (y_t - V^v(S_t))^2$$

[0102] 그리고 정책 함수 딥러닝 네트워크는 [수학식 11]과 같이 표현되는 손실 함수의 손실값이 최소가 되도록 업데이트될 수 있다.

수학식 11

$$Loss = -\log(N(A|\mu(S_t), \sigma(S_t))) \cdot \varepsilon_t$$

[0103]

[0105]

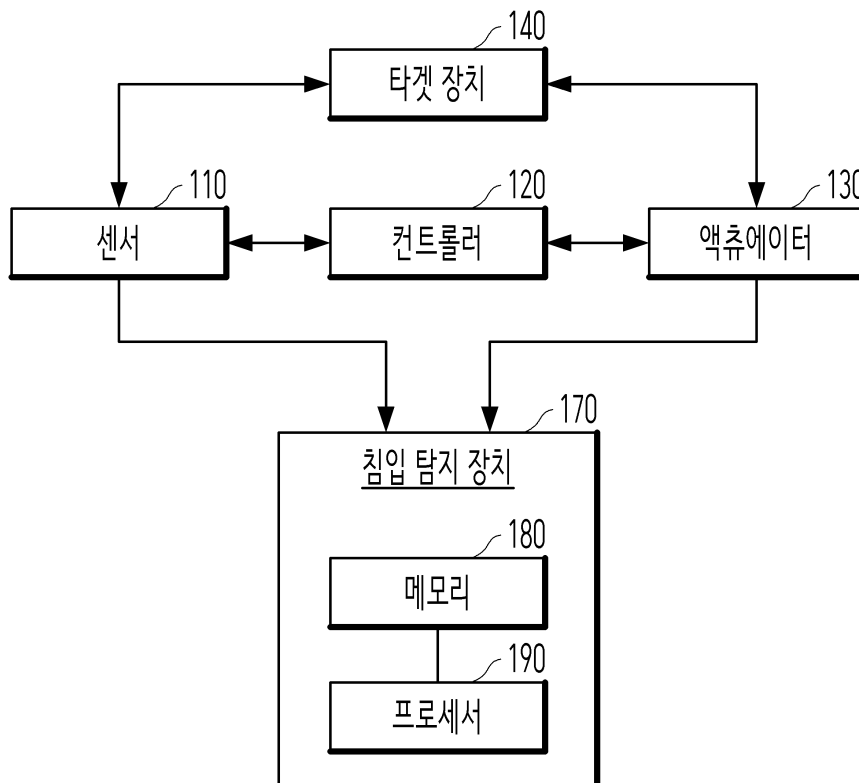
앞서 설명한 기술적 내용들은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예들을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic 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.

[0107]

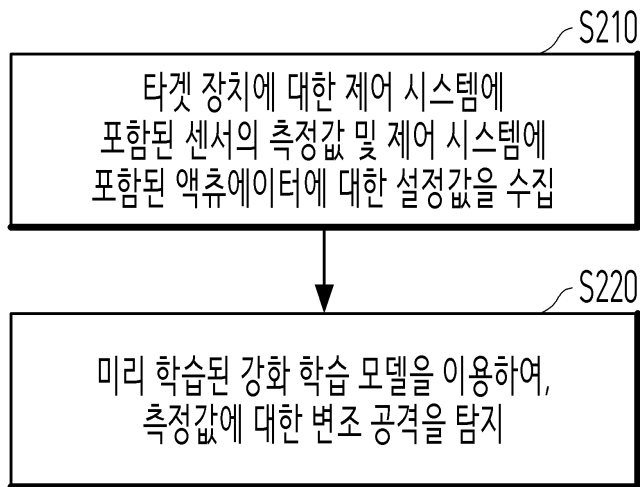
이상과 같이 본 발명에서는 구체적인 구성 요소 등과 같은 특정 사항들과 한정된 실시예 및 도면에 의해 설명되었으나 이는 본 발명의 보다 전반적인 이해를 돕기 위해서 제공된 것일 뿐, 본 발명은 상기의 실시예에 한정되는 것은 아니며, 본 발명이 속하는 분야에서 통상적인 지식을 가진 자라면 이러한 기재로부터 다양한 수정 및 변형이 가능하다. 따라서, 본 발명의 사상은 설명된 실시예에 국한되어 정해져서는 아니되며, 후술하는 특허청구범위뿐 아니라 이 특허청구범위와 균등하거나 등가적 변형이 있는 모든 것들은 본 발명 사상의 범주에 속한다고 할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

도면

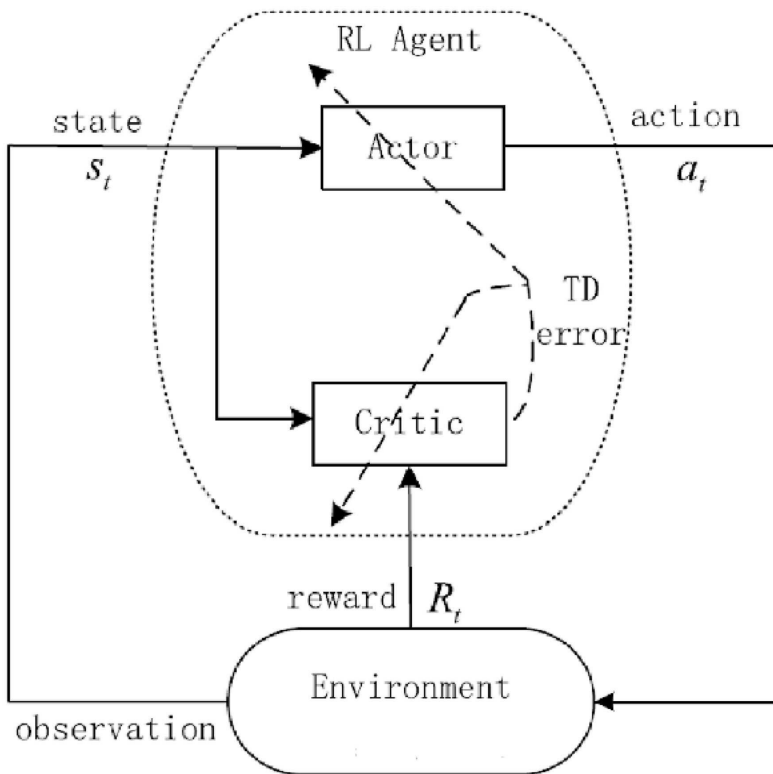
도면1



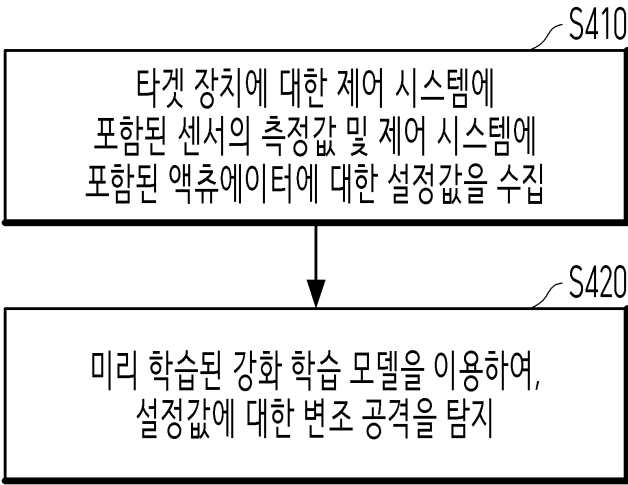
도면2



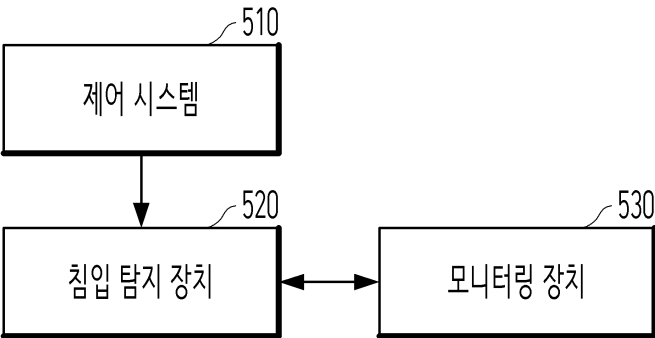
도면3



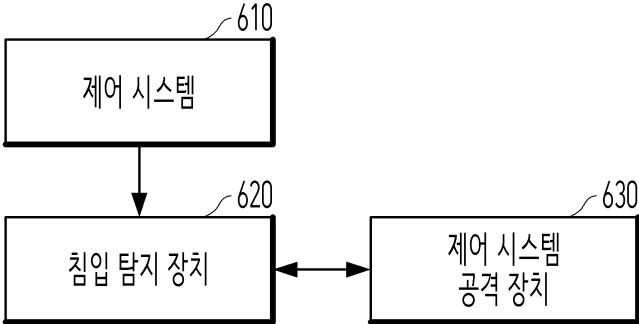
도면4



도면5



도면6



도면7

