사이버-물리 융합보안 연구 동향

김상준, 박경준
대구경북과학기술원

요 약

사이버물리시스템(Cyber-physical system, CPS)은 소프트웨어가 유무선 네트워크를 통해 다양한 물리시스템을 제어하는 형태의, 밀접하게 결합된 분산 제어시스템을 의미하며, 최근 스마트 빌딩, 스마트 팩토리, 자율 주행차, 무인열차 시스템 등 다양한 사회 기반시설이 이에 해당된다. 특히 사이버물리시스템의 경우 물리시스템의 동역학 정보를 이용한 악의적인 공격의 가능성이 존재한다. 이는 기존의 사이버 공격과는 근본적인 차이점을 지니기 때문에 기존의 탐지기법으로는 탐지에 한계가 있다. 이러한 공격에 대비하는 방법론으로 사이버-물리 융합보안에 관한 연구가 최근 관심을 받고 있다. 본고에서는 사이버-물리 공격의 수학적 표현법과 이상탐지기 기반의 사이버-물리 공격 탐지 방법, 이상탐지기 상에서 검출 불가능한 특수한 형태의 사이버-물리 공격에 대해 알아본다.

Ⅰ. 서 론

사이버물리시스템은 물리공간에 존재하는 물리시스템과 컴퓨터 내에 존재하는 사이버세계의 소프트웨어가 네트워크를 통해 밀접하게 결합된 분산 제어 시스템을 의미한다[1]. 사이버물리시스템의 네트워크는 물리시스템의 상태 정보를 사이버세계의 소프트웨어로 전달하며, 소프트웨어는 이 정보를 바탕으로 제어신호를 계산하여 물리시스템을 실시간으로 제어하는 피드백-제어시스템(feedback-control system)의 정보전달의 역할을 수행한다. 사이버물리시스템은 지능형 교통시스템, 발전소, 스마트 공장, 스마트 빌딩과 같은 사회기반시설에 적용되는 핵심기술일 뿐만 아니라 의료장비, 로봇과 같은 여러 분야에 걸쳐 응용되고 있다[1][2].

〈그림 1〉과 같이 사이버물리시스템은 네트워크의 도입으로 하나의 컴퓨터가 다수의 물리시스템을 제어하는 등 시스템 구성적인 이점을 주었지만, 네트워크를 통해 시스템 외부에서 사이버 물리시스템 동작에 직접적인 영향을 미치는 사이버-물리 공격에 노출되었다[3].

사이버-물리 공격이란 단순히 사이버시스템을 공격하는 기존의 사이버 공격과 달리, 공격자가 물리시스템의 동역학 정보를 활용하여 사이버를 통해 물리시스템을 공격하는 완전히 새로운 형태의 공격기법을 의미한다. 이에 따라 사이버물리시스템은 사이버-물리 공격에 대한 높은 수준의 보안을 요구한다. 기존의 네트워크 공격은 개인정보 및 기밀문서 유출과 도청이 목적이며, 네트워크 보안 역시 데이터 암호화 및 네트워크 접근 인증을 대상으로 연구가 진행되었다. 하지만, 사이버-물리 공격은 사이버 물리시스템이 제어하는 물리시스템의 오작동을 유발시키는 것이 목적이며, 센서 정보 및 제어입력정보를 실시간으로 변조하여 물리시스템이 불안정한 상태가 되도록 유도한다[4]. 사이버-물리 공격에 대응하기 위해서는 네트워크 보안적인 요소 뿐만 아니라 물리시스템의 동적 변화(dynamics)를 모두 고려하는 사이버-물리 융합보안 기법이 도입되어야 한다.

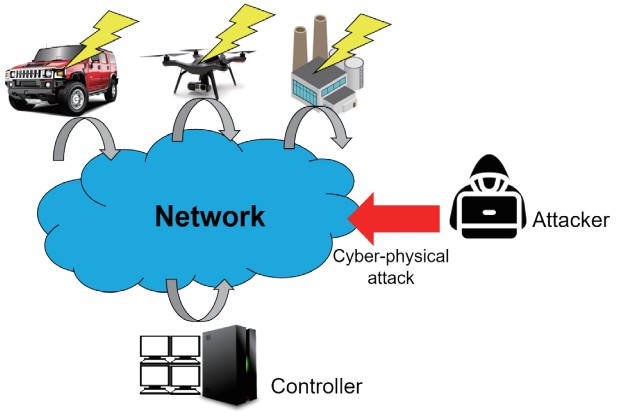

그림 1. 네트워크를 통한 사이버-물리 공격

기존에는 제어 이론의 이상탐지기(anomaly detector)를 이용하여 센서 출력 정보의 조작 및 제어 입력 정보의 조작 여부를 판단하였으나, 최근 제어 이론 분야에서 동적 시스템 지식을 기반으로 하는 특수한 형태의 공격들이 연구되고 있다[5]. 이러한 특수 형태의 공격은 기존의 이상탐지기에 의해 검출되지 않아 사

이버물리시스템에 큰 위협이 되고 있다.

본 고에서는 사이버-물리 공격의 수학적인 표현에 대하여 알아보고, 사이버-물리 공격이 물리시스템에 적용되면 어떤 영향이 생기는지에 대해 알아본다.

Ⅱ. 사이버-물리 공격의 표현 및 검출 방법

본 장에서는 사이버-물리 시스템을 대상으로 하는 공격의 수학적 표현과 이상탐지기에서의 공격 검출 방법에 대해 알아본다.

사이버-물리 공격은 제어 이론 분야에서 주로 다루어지며, 물리시스템과 제어기 간의 교환되는 센서 출력 또는 제어 입력 값에 공격 신호를 더하는 방식으로 표현된다. 사이버-물리 공격을 표현하기 위해서 다음과 같은 단입력-단출력의 선형시스템을 고려한다.

$$\dot{x}(t) = Ax(t) + Bu(t)$$
$$y(t) = Cx(t) \quad (1)$$

식(1)에서 $A \in \mathbb{R}^{n\times n}$는 시스템 행렬, $B \in \mathbb{R}^{n}$는 입력 행렬, $C \in \mathbb{R}^{1\times n}$, 출력 행렬을 의미하며, 물리시스템의 상태 변수, $x(t) \in \mathbb{R}^{n}$입력 $u(t) \in \mathbb{R}$와 출력 $y(t) \in \mathbb{R}$는 시간에 따라 변화하는 값이다. 물리시스템의 모든 상태 변수 측정은 보통 불가능하기 때문에, 관측기(observer) 기반의 제어를 수행하게 된다. 관측기는 시스템 모델을 기반으로 물리시스템의 상태를 추정하는 구성 요소이며, 제어기는 추정된 상태 변수를 기반으로 물리시스템의 제어 입력을 계산하는 역할을 수행한다. 관측기와 제어기는 식(2)와 같이 표현된다.

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + L(y(t) - C\hat{x}(t))$$
$$u(t) = -K\hat{x}(t) \quad (2)$$

수식에서 $\hat{x}(t) \in \mathbb{R}^{n}$는 관측기에 의해 추정되는 상태 변수를 의미하며, 관측기 $L \in \mathbb{R}^{n}$과 제어기 $K \in \mathbb{R}^{1\times n}$는 적당한 방법에 의해 선정될 수 있다.

사이버-물리 공격은 식(3)과 같이 제어 입력 또는 센서 출력에 공격 신호를 더하는 방식으로 표현된다. 이때, 제어 입력에 추가적인 신호 $a^{u}(t)$를 더하는 공격을 제어기 공격, 센서 출력에 공격자의 신호 $a^{y}(t)$를 더하는 공격을 센서 공격이라고 한다. $\tilde{x}(t)$와 $\tilde{y}(t)$는 각각 공격자에 의해 영향을 받은 상태 변수와 센서 출력을 의미한다.

$$\dot{\tilde{x}}(t) = A\tilde{x}(t) + B(u(t) + a^{u}(t)) \quad (3)$$
$$\tilde{y}(t) = Cx(t) + a^{y}(t) \quad (4)$$

센서 공격 및 제어기 공격으로 인해 물리시스템의 제어 성능이 저하되거나 상태 변수가 발산하여 물리시스템이 불안정해질 수 있다. 제어 입력 및 센서 출력 정보를 조작하는 사이버-물리 공격을 판단하기 위해서 제어 이론에서는 이상 탐지기를 사용한다. 이상탐지기란 관측기에서 추정되는 물리시스템의 센서 출력과 실제 센서 출력의 오차를 기반으로 시스템의 이상 여부를 판단하는 요소를 의미한다. 이상탐지기는 식(5)과 같이 표현된다[4].

$$r(t) = \|y(t) - C\hat{x}(t)\|$$
$$\Gamma(t) = \begin{cases} 1, & if\ r(t) > \delta \\ 0, & otherwise \end{cases} \quad (5)$$

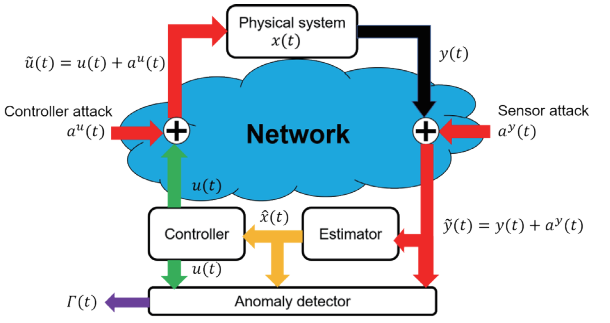

그림 2. 사이버-물리 공격과 이상 관측

만약, 식(2)의 관측기에서 예측된 센서 출력 $\hat{y}(t)$와 실제 센서 출력$y(t)$가 공격 검출 판단 기준 오차 δ이상인 경우 식(5)의 이상 탐지기 상태 $\Gamma(t)$는 1이 되며, 제어기는 물리시스템에 공격 또는 외부 요인에 의한 문제가 있다고 판단한다.

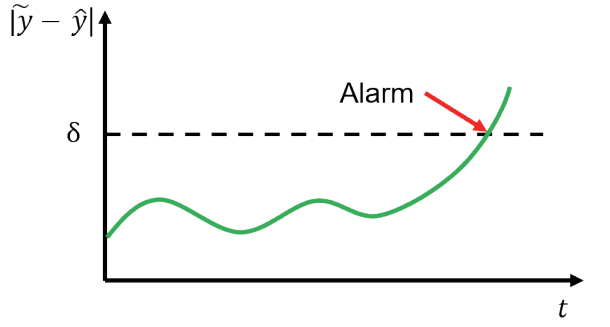

그림 3. 이상 탐기지에서의 공격 탐지

네트워크 전송 지연과 패킷 손실, 센서의 잡음들은 센서 출력 값과 관측기의 물리시스템 모델 간의 오차 $r(t)$를 증가시키는 요인이 된다. 만약, 공격 검출 판단기준 δ가 너무 작은 경우, 공격이 실제로는 존재하지 않음에도 불구하고 이상탐지기가 공격이라고 판단하는 오탐(false positive)이 발생할 수 있으며, δ를

너무 크게 설정하는 경우에는 공격이 발생했음에도 불구하고 공격이 검출되지 않는 미탐(false negative)이 발생하거나 공격 검출 시간의 지연을 야기한다. 그러므로, 공격 검출 판단기준 δ는 네트워크 환경과 센서의 잡음을 고려하여 설정되어야 한다.

Ⅲ. 사이버-물리 공격의 분류와 stealthy한 사이버-물리 공격의 표현

본 장에서는 사이버-물리 공격을 분류하고 이상 탐지기에서 검출 불가능한 stealthy 공격을 분석한다.

1. 사이버-물리 공격의 분류

사이버물리시스템에 대한 공격은 〈그림 4〉와 같이 시스템 지식, 시스템 신호 유출, 시스템 파괴의 3요소로 분류할 수 있다[5].

시스템 지식이란 물리시스템 모델, 즉 물리시스템의 동역학 방정식에 대한 정보를 의미한다. 시스템에 대한 지식 없이 무작위로 생성되는 제어기·센서 공격 또는 통신을 방해하는 공격은 이상 관측기에 의해서 검출 가능하다. 반면, 정교한 물리시스템 모델 기반의 제어기·센서 공격은 이상탐지기에 검출되지 않는다.

시스템 신호 유출은 네트워크 상에서 물리시스템과 제어기 사이의 제어 입력과 센서 출력 값의 획득 여부를 의미한다. 기존 네트워크 보안 연구에서 다루어지는 공격 기법인 감청(Eavesdropping)은 제어 입력과 센서 출력 값을 직접적으로 알아낼 수는 있으나, 물리시스템에 직접적인 손상을 입힐 수 없다. 하지만, 물리시스템의 정교한 모델을 보유하지 못한 공격자의 경우에도 제어 입력 신호와 센서 출력 신호를 이용하여 이상 탐지기에 검출되는 속도를 늦추는 정교한 공격을 시도할 수 있다.

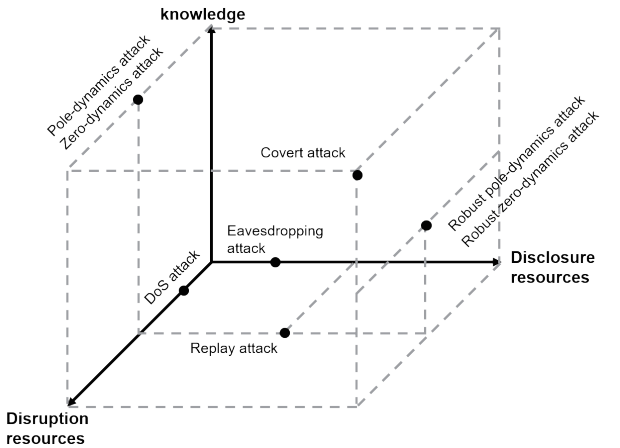

그림 4. 사이버-물리 공격의 분류

시스템 파괴는 실제적으로 물리시스템의 상태를 불안정하게 만드는 요소를 의미한다. 감청 공격의 경우에는 시스템에 아무 영향이 없으나, 서비스 거부(Denial of service, DoS) 공격과 같은 통신 방해 공격이나 제어 입력 신호 또는 센서 출력 값을 변조하는 공격은 실제 물리시스템을 불안정하게 만들 수 있다.

〈표 1〉은 사이버-물리 공격 별 시스템 모델의 필요 여부, 공격 위치, 이상 탐지기 검출 여부를 분류를 나타낸다.

표 1. 시스템 모델의 필요 여부 및 공격 위치, 이상탐지기 검출 여부에 따른 사이버-물리 공격 분류

공격 분류	시스템 모델 필요	센서공격	제어기 공격	이상 탐지기로 검출가능 여부
DoS	×	×	×	○
Replay	×	○	○	×
Zero-dynamics	○	×	○	×
Pole-dynamics	○	○	×	×
Covert	○	○	○	×

〈표 1〉에서는 〈그림 4〉의 사이버-물리 공격 중 물리시스템의 동작에 영향을 미치는 공격들을 대상으로 시스템 모델 필요 여부, 네트워크 상의 공격 대상 신호 및 이상탐지기에 의한 공격 검출 여부를 분류하였다. 〈표 1〉의 서비스 거부 공격은 제어 명령 및 센서 신호 정보의 전달을 수행하는 통신을 방해하여 물리시스템의 제어 성능 저하 및 불안정을 유발하는 공격이다[6]. 서비스 거부 공격은 단순히 통신을 방해하기 때문에 물리시스템의 수학적 모델을 보유하지 않더라도 수행 가능한, 난이도가 낮은 공격에 속한다[5]. 하지만, 서비스 거부 공격은 이상탐지기에 의해 쉽게 탐지 가능하며 공격에 대한 여러 대처와 관련된 연구가 수행되었다.

서비스 거부 공격에 비해, 다른 사이버-물리 공격들은 물리시스템의 특성을 활용하여 개발된 공격이기 때문에 이론적으로 이상탐지기에 의해 검출되지 않는다. 이상탐지기에 검출되지 않는 공격을 stealthy 공격이라고 한다.

2. Replay 공격

Replay 공격[7]은 물리시스템과 제어기가 네트워크 상에서 교환하는 센서 출력 신호를 기록하였다가 특정시점에서 기록된 과거의 센서 신호를 제어기로 입력하는 공격이다. Replay 공격은 두 단계의 공격 수행 단계를 거친다. 첫번째 단계에서는 〈그림 5〉와 같이 공격자가 네트워크 상에서 센서 출력 값을 도청하여 기록한다. 센서 출력 값을 기록한 후, 두 번째 단계에서 〈그림 6〉과 같이 공격자는 특정 시점에 물리시스템과 제어기 사이의 네트워크 링크를 단절한 후, 물리 시스템-공격자-제어기 네트워크

연결을 수립한다. 공격자는 공격 시점으로부터 T시간 이전의 센서 출력 값을 제어기에 입력하며, 이와 동시에 물리시스템에 손상을 입힐 수 있는 제어 입력 신호를 전송한다.

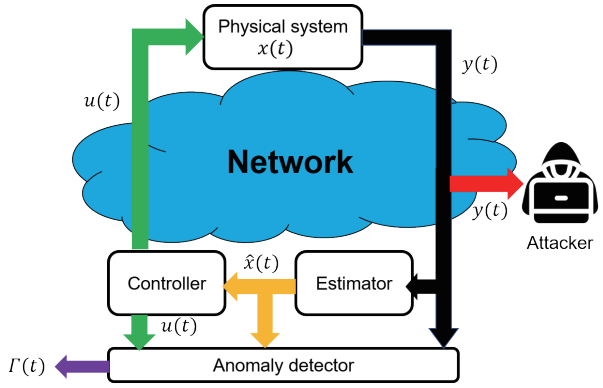

그림 5. Replay 공격을 위한 센서 값 기록

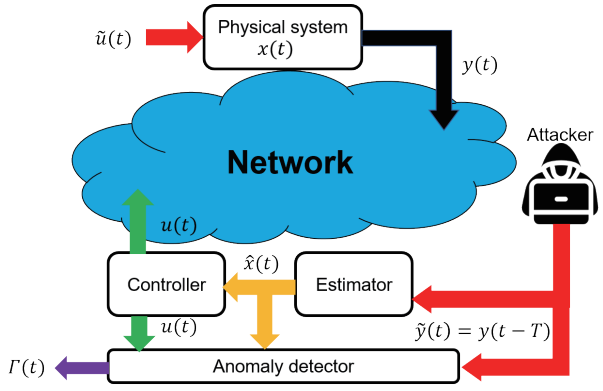

그림 6. 네트워크 상의 제어 입력과 센서 값 전달경로 단절 후, 기록된 센서 값과 조작된 제어입력신호를 전달하는replay 공격

제어기는 공격자의 센서 신호가 이전에 받은 센서 출력과 동일하므로, 공격자의 센서 공격 신호를 정상으로 판단하게 되고, 물리 시스템은 잘못된 제어 입력신호를 받기 때문에, 물리시스템의 상태 변수는 불안정해진다.

Replay 공격은 과거의 신호를 기록하였다가 재생하기 때문에 물리시스템에 대한 정확한 모델을 요구하지 않는다.

Replay 공격을 검출하기 위해서 '워터마킹(watermarking)' 신호를 이용한 공격 검출 기법이 제안되었다[8]. '워터마킹'이란 제어 입력 신호에 무작위 잡음 신호를 섞어서 함께 보내는 방법이다. 물리시스템에 무작위의 워터 마킹 신호가 입력되면, 입력된 '워터마킹' 신호에 대한 응답이 센서 출력 신호에 반영되어 제어기에 입력되게 된다. 하지만, replay 공격이 수행되는 경우에는 과거의 센서 출력 신호가 들어오기 때문에 워터 마킹 신호에 대한 응답이 올바르지 않아 공격 여부를 판단할 수 있다.

3. Zero-dynamics 공격

Zero-dynamics 공격[5]은 불안정한 영점과 관련된 시스템 동역학을 대상으로 제어 입력 신호를 〈그림 7〉과 같이 변조하는 제어기 공격이다. 이 공격을 적용하기 위해서는 물리 시스템에 불안정한 영점이 반드시 존재하여야 한다.

물리시스템에 입력되는 신호 중, 불안정한 영점을 상쇄하는 공격 신호에 대한 응답은 센서 출력에 나타나지 않는다. 하지만, 실제 물리시스템의 상태 변수는 발산하게 되어 물리시스템의 상태가 불안정하게 된다. 상태공간에서 영동역학 표현을 위해 시스템 식(1)에 zero-dynamics 공격을 반영하여, 식(6)과 같이 Byrnes-Isidori normal form으로 표현한다[9][10]. 식(6)의 행렬 A_v, B_v C_v는 식(7)과 같다[10].

$$\begin{aligned}\begin{bmatrix}\dot{z}_s(t)\\ \dot{z}_u(t)\end{bmatrix} &= \begin{bmatrix}S_s & 0\\ 0 & S_u\end{bmatrix}\begin{bmatrix}z_s(t)\\ z_u(t)\end{bmatrix} + \begin{bmatrix}P_s\\ P_u\end{bmatrix} y(t)\\ \dot{\eta}(t) &= A_v\eta(t) + B_v(\psi^T\eta(t) + \varphi_s^T z_s(t)\\ &\quad + \varphi_u^T z_u(t) + gu(t) + a^u(t))\\ y(t) &= C_v\eta(t)\end{aligned} \tag{6}$$

$$\begin{aligned}A_v &= \begin{bmatrix}0_{v-1} & I_{v-1}\\ 0 & 0_{v-1}\end{bmatrix}\\ B_v &= \begin{bmatrix}0_{v-1}\\ 1\end{bmatrix}\\ C_v &= [1\ 0_{v-1}^T]\end{aligned} \tag{7}$$

$$\begin{aligned}\dot{z}^a(t) &= S_u z_u^a(t)\\ a^u(t) &= -\frac{1}{g}\varphi_u^T z_u^a(t)\end{aligned} \tag{8}$$

식(6)에서 불안정한 영동역학 상태 변수$z_u(t)$를 상쇄하기 위한 제어기 공격 신호 $a^u(t)$는 식(8)과 같이 설계될 수 있다. 불안정한 영동역학 상태 변수 요소가 상쇄된 센서 신호는 안정한 영동역학적 요소만 반영되므로 제어기는 물리시스템이 안정적으로 동작한다고 판단한다. 하지만, $\lim_{t\to\infty}\|a^u(t)\| = \infty$ 이므로, 시스템 식(1)에 제어 입력 $u(t) + a^u(t)$를 반영하면 실제 물리시스템의 상태 변수는 발산한다.

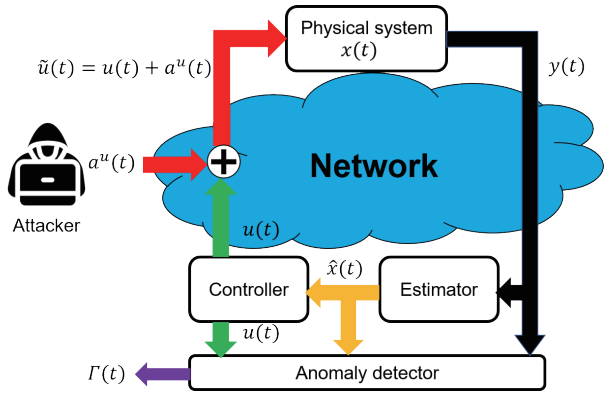

그림 7. Zero-dynamics 공격의 제어 입력 변조

현실적으로 물리시스템의 완벽한 수학적 모델을 얻는 것은 불가능하며, 실제 물리시스템과 모델의 오차가 존재한다면 이상탐지기에 검출되게 된다. 또한, 물리시스템을 구동하는 액츄에이터의 출력 한계로 인해 제어 입력의 포화가 발생하여 이상탐지기에 공격이 검출된다. 이러한 한계를 극복하기 위해서 robust zero-dynamics 공격이 개발되었다. Robust zero-dynamics 공격[9]은 물리시스템의 완벽한 모델을 알지 못하는 공격자가 제어입력신호와 센서 정보를 모두 알 수 있다면 불안정한 영점을 상쇄하는 신호를 발생시켜 제어입력신호에 추가하는 공격을 의미한다. 또한, 이상탐지기에 검출되는 속도를 늦추어 물리시스템을 불안정하게 만들 수 있다.

Zero-dynamics 공격을 검출하기 위해 시스템 행렬 A, B, C를 변경하여 공격을 검출하는 방법이 연구되었다[11]. 또한, 0차 홀드(Zero-order hold)를 이용하는 디지털 제어시스템에서 발생하는 불안정한 영점을 대상으로 하는 공격을 generalized hold 기법으로 탐지하는 연구가 진행되었다[12].

4. Pole-dynamics 공격

Pole-dynamics 공격[13]은 물리시스템의 불안정한 상태 변수를 대상으로 하는 센서 공격으로, 〈그림 8〉과 같이 표현된다. 이 공격이 수행되기 위해서는 반드시 물리시스템에 불안정한 상태 변수가 존재하여야 한다.

상태공간에서 물리시스템의 안정한 상태 변수와 불안정한 상태 변수를 분리하고, pole-dynamics 공격을 표현하기 위해서 식(1)을 식(9)와 같이 재구성할 수 있다. 식(9)에서 $x_s(t)$와 $x_u(t)$는 각각 안정한 상태 변수와 불안정한 상태 변수를 의미하며, $a^y(t)$는 pole-dynamics에 해당하는 센서 공격을 의미한다. 센서 출력 $y(t)$는 안정한 상태 변수 와 불안정한 상태 변수 $x_u(t)$가 반영되어 제어기에 입력된다.

$$\begin{bmatrix}\dot{x_s}(t)\\ \dot{x_u}(t)\end{bmatrix} = \begin{bmatrix}A_s & 0\\ 0 & A_u\end{bmatrix}\begin{bmatrix}x_s(t)\\ x_u(t)\end{bmatrix} + \begin{bmatrix}B_s\\ B_u\end{bmatrix}u(t) \qquad (9)$$
$$y(t) = C_s x_s(t) + C_u x_u(t) + a^y(t)$$

$$x^a(t) = A_u x_u^a(t)$$
$$a^y(t) = -C_u x_u^a(t) \qquad (10)$$

Pole-dynamics 공격은 센서 정보를 구성하는 상태 변수 요소 중에서 불안정한 상태 변수가 반영된 센서 출력 요소 $C_u x_u(t)$를 상쇄하는 신호를 더한다. Pole-dynamics 공격은 식(10)과 같이 설계될 수 있다. 공격자에 의해 변조된 센서 출력은 $\tilde{y} = C_s x_s(t) + C_u(x_u(t) - x_u^a(t))$가 되며, $\lim_{t\to\infty}\|x_u(t) - x_u^a(t)\| = 0$가 되어 제어기 입장에서는 센서 출력이 안정되게 보인다. 하지만, 공격자에 의해 생성되는 신호에 대한 상태 변수가 $\lim_{t\to\infty}\|x_u^a(t)\| = \infty$이므로 시간이 지남에 따라 공격자의 센서 변조 신호는 무한대로 발산하게 되며, 불안정한 상태 변수 역시 $\lim_{t\to\infty}\|x_u(t)\| = \infty$가 되므로, 불안정한 상태 변수 역시 발산하게 된다.

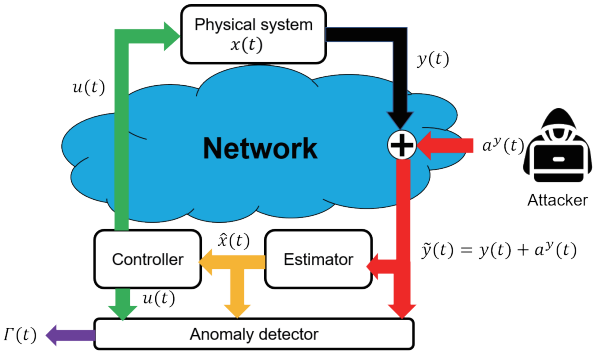

그림 8. Pole-dynamics 공격의 센서 출력 변조

Zero-dynamics 공격과 마찬가지로, pole-dynamics 공격도 완벽한 물리시스템의 모델을 요구하기 때문에 실제적으로 공격을 구현하기 어렵다. 하지만, 불완전한 물리시스템의 모델과 제어입력신호, 센서 출력을 이용하여 센서 출력의 불안정한 상태 변수 요소를 상쇄하는 robust pole-dynamics 공격이 개발되었다. 불안정한 물리시스템의 모델로 인해서 시간이 지남에 따라 이상탐지기에 의해 공격이 검출되지만, 검출되는 속도를 늦추어 물리시스템을 불안정하게 만들 수 있다.

현재까지 pole-dynamics 공격에 대한 탐지 기법 및 공격 대응 기법은 개발되지 않았다.

5. Covert 공격

Covert 공격[14][15]은 〈그림 9〉와 같이 정교한 물리시스템 모

델을 기반으로 제어기와 물리시스템간의 제어입력신호 전달경로와 센서 출력 전달 경로를 탈취 및 조작하는 공격을 의미한다. 정교한 물리시스템의 모델을 가지고 있는 공격자는 제어기에서 물리시스템으로 전달되는 제어 입력신호를 탈취하여 물리시스템 모델을 기반으로 거짓된 센서 값을 만들어내고, 공격자에 의해 만들어진 센서 정보를 다시 제어기에 입력하는 공격을 의미한다. 동시에, covert 공격은 물리시스템의 제어 권한을 탈취하여 공격자의 의도대로 물리시스템을 제어하거나 물리시스템의 오작동을 야기한다.

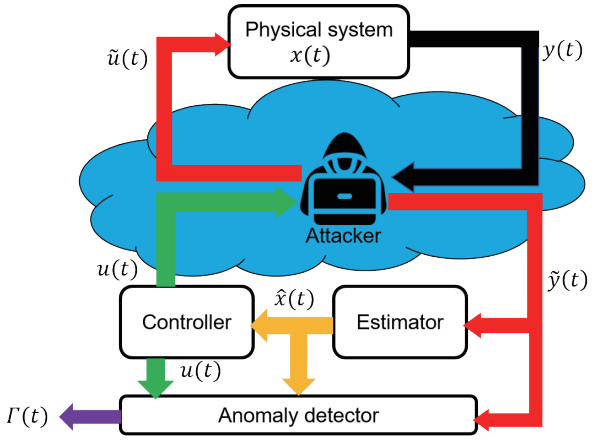

그림 9. Covert 공격의 동작

Ⅳ. 네트워크 중간자 공격 기반 사이버-물리 공격의 구현

본 장에서는 네트워크 상에서 사이버-물리 공격이 실제로 구현될 수 있는 중간자 공격에 대해 분석한다.

중간자 공격은 네트워크를 통해 정보를 교환하는 특정 호스트들 간의 통신제어권을 탈취하는 공격을 의미한다[16]. 통신제어권을 탈취한 공격자는 호스트들 간의 통신 내용을 도청할 수 있으며, 주고받는 정보를 변조할 수 있다[17][18]. 중간자 공격과 물리 시스템 모델을 기반으로 센서 정보 및 제어입력신호 변조가 수행될 수 있으며, 이는 zero-dynamics 공격 또는 pole-dynamics 공격과 같은 stealthy한 공격으로 이어질 수 있다.

중간자 공격은 주로 네트워크 프로토콜의 취약점을 기반으로 공격이 수행되며[19], 〈표 2〉와 같이OSI 계층 모델의 모든 계층에서 공격이 발생할 수 있다.

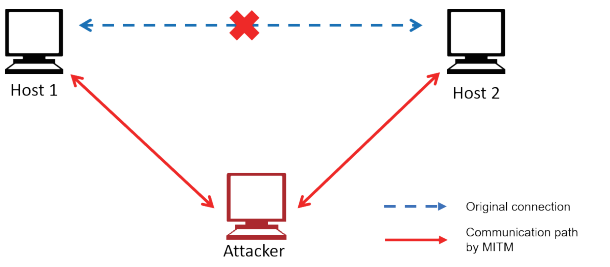

그림 10. 중간자 공격의 동작

표 2. 네트워크 계층에 따른 중간자 공격의 분류[16]

계층	공격 구현 방법
Application	BGP MITM, DHCP spoofing, DNS spoofing
Presentation	SSL/TLS MITM
Transport	IP spoofing
Network	
Data Link	ARP spoofing

또한, GSM이나 LTE 와 같은 셀룰러 네트워크에 대해서도 거짓 기지국(false base station) 기반의 중간자 공격을 통해 거짓된 신호를 입력하는 공격이 수행될 수 있다. 사이버물리시스템을 운영하는 네트워크 관리자는 중간자 공격에 대한 예방책을 미리 수립해야 할 것이다.

Ⅴ. Stealthy 사이버-물리 공격 시뮬레이션

본 장에서는 replay 공격과 zero-dynamics 공격, pole-dynamics 공격의 예제를 중심으로 사이버-물리 시스템에 stealthy 공격이 적용되는 상황에서 물리시스템의 동작 및 이상탐지기에 의한 공격 검출 여부에 대하여 알아본다.

사이버-물리 공격의 영향을 알아보기 위해서 다음과 같이 zero-dynamics 공격과 pole-dynamics 공격이 모두 수행될 수 있는 2차 선형 시불변 시스템을 고려한다. 시스템 (12)는 두 개의 불안정한 극점 $p_0, p_1 = 0.5 \pm j1.94$과 한 개의 불안정한 영점 $z_0 = 1$을 가지므로 zero-dynamics 공격 및 pole-dynamics 공격을 모두 수행될 수 있다.

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 \\ -4 & 1 \end{bmatrix} x(t) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(t)$$
$$y(t) = [1\ 0]x(t) \qquad (12)$$

1. Replay 공격.

〈그림 11〉은 replay 공격 수행 시의 물리시스템의 상태, 제어기 및 센서 공격 신호, 이상탐지기의 반응을 나타낸다. 이상탐지기의 공격 검출 판단 기준 δ는 0.08로 설정한다. 시뮬레이션 시간 3.5초 시점에서 0.6초 시점의 센서 값을 제어기에 다시 입력하는 replay공격이 물리시스템에 수행되며, 물리시스템에는 크기 1의 계단 함수 형태의 제어 입력 신호가 입력된다. 공격 수행 이후의 물리시스템의 상태변수들은 발산한다.

공격 시작시점에서 replay 공격에 의해 조작된 센서 정보와 물리 시스템 모델에서 예측한 센서 정보의 오차로 인해서 0.08에 가까운 이상탐지기 응답이 발생한다. 하지만, 실제 시스템에서 네트워크 지연 및 패킷 손실, 센서 잡음으로 인해 이상탐지기의 응답은 항상 0이 아니기 때문에, 공격 시작시점에서 발생한 이상탐지기 응답은 잡음으로 간주될 수 있다. Replay 공격에 의한 영향 측면에서, 공격 시작 시점 이후 빠른 속도로 이상탐지기의 값이 0으로 수렴하기 때문에 이상탐지기로는 공격 검출이 불가능하다. 또한, replay 공격의 센서 값 재생 시점 설정을 통해 공격 시작 시의 이상탐지기의 응답을 더욱 작게 만들 수 있다.

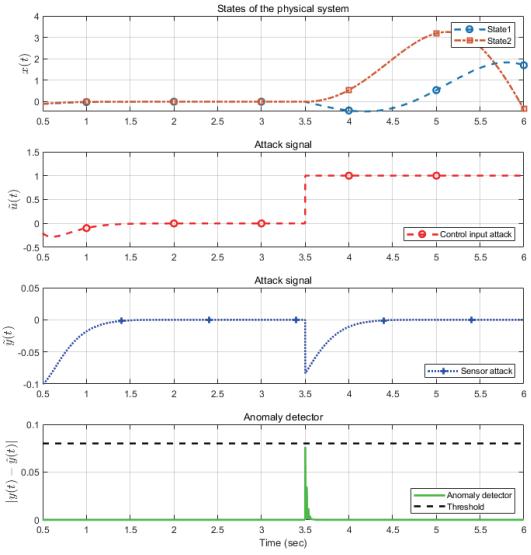

그림 11. 물리시스템에 적용된 replay 공격 영향과 이상탐지기의 응답

2. Zero-dynamics 공격

〈그림 12〉는 zero-dynamics 공격 수행 시의 물리시스템의 상태, 제어기 공격 신호, 센서 정보, 이상탐지기의 반응을 나타낸다. 공격 영향의 확인을 위해서 물리시스템이 정상상태인 3초 이후에 공격을 시도한다. 이상탐지기의 공격 검출 판단 기준 δ는 0.008로 설정한다. 시뮬레이션 시간 4초 시점에서 zero-dynamics 공격이 수행되며, 공격 수행 시점 이후의 물리시스템의 상태변수 중 하나가 지수적으로 발산한다.

불안정한 영점으로 인해서 지수적으로 발산하는 공격 신호가 물리시스템에 입력됨에도 불구하고 발산하는 상태변수가 센서 정보에 반영되지 않는다. 공격 수행시점에서, zero-dynamics공격의 초기값에 의해 센서 값이 약간 증가하는 양상을 보이지만, 공격이 수행됨에 따라서 센서 값에 보이는 공격의 영향은 0으로 수렴하게 되며, 이상탐지기의 응답 역시 공격 검출 판단 기준 를 넘지 않기 때문에 센서의 잡음 또는 네트워크 전송 지연으로 간주되어, zero-dynamics 공격은 이상탐지기에 검출되지 않는다.

Zero-dynamics 공격 신호 초기값에 따라서 물리시스템의 발산 속도와 이상 탐지기의 응답 크기 조절이 가능하다. 공격 신호의 초기값이 클수록 이상탐지기의 응답이 커지는 한편, 더욱 빠른 속도로 물리시스템의 상태변수를 발산시킬 수 있다.

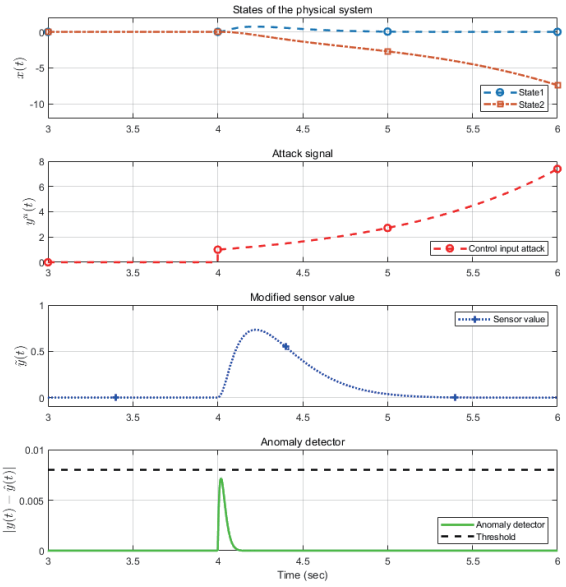

그림 12. 물리시스템에 적용된 Zero-dynamics 공격 영향과 이상탐지기의 응답

3. Pole-dynamics 공격

〈그림 13〉은 pole-dynamics 공격 수행 시의 물리시스템의 상태, 센서 공격 신호, 센서 정보, 이상탐지기의 반응을 나타낸다. 공격 영향의 확인을 위해서 물리시스템이 정상상태인 3초 이후에 공격을 시도한다. 이상탐지기의 공격 검출 판단기준 δ는 0.08로 설정한다. 시뮬레이션 시간 4초 시점에서 pole-dynamics 공격이 수행되며, 공격 수행 시점 이후의 물리시스템의 불안정한 상태변수들이 발산한다.

시뮬레이션에 이용된 물리시스템의 상태변수 두 개는 모두 불

안정하며, 공격 시점 이후에 제어기에 입력되는 센서 값은 〈그림 13〉의 공격 신호와 같이 원래의 센서 값에서 불안정한 상태변수에 해당하는 센서 값이 상쇄된 신호가 된다. 공격 시점에서 공격 신호의 초기값에 의해 센서 값은 다소 감소하는 추세를 보이나, 시간이 지남에 따라 제어기에 입력되는 센서 값은 0으로 수렴하게 된다. 또한, 이상탐지기의 응답 역시 공격 검출 판단기준 δ를 넘지 않고 시간에 지남에 따라 0으로 수렴하기 때문에 이상탐지기 기법으로는 pole-dynamics 공격을 검출할 수 없다.

Pole-dynamics 역시 zero-dynamics 공격과 마찬가지로, 공격 신호 초기값에 따라서 물리시스템의 발산 속도와 이상 탐지기의 응답 크기 조절이 가능하며, 공격 신호의 초기값이 클수록 이상탐지기의 응답이 커지는 한편, 더욱 빠른 속도로 물리시스템의 상태변수를 발산시킬 수 있다.

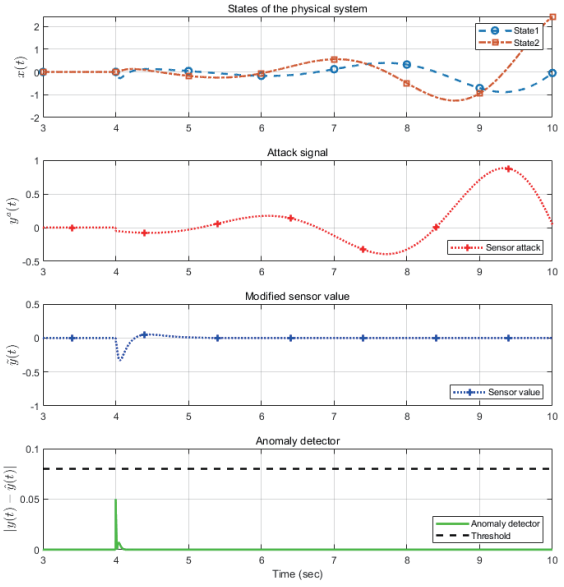

그림 13. 물리시스템에 적용된 Pole-dynamics 공격 영향과 이상탐지기의 응답

통신을 방해하는 서비스 거부 공격이나 제어 입력 또는 센서 출력에 임의의 값을 더하는 공격과는 달리, stealthy한 사이버-물리 공격은 물리시스템에 대한 지식을 기반으로 공격이 수행되며 이상탐지기에 의해 검출이 되지 않는다. 공격 시작 시점에서 stealthy 공격에 대한 이상탐지기 응답이 발생하나, 공격 신호 초기값에 따라서 이상탐지기 초기 응답의 크기를 조절할 수 있으며, 공격 시간이 지남에 따라서 센서 출력과 이상탐지기 응답은 안정된 값으로 수렴하는 특성이 있음을 시뮬레이션을 통해 확인할 수 있다.

Ⅵ. 결 론

본고에서는 사이버-물리 공격의 표현법에 대해 알아보고 공격의 세 가지 성질과 이상탐지기에 의한 검출 가능성 여부를 기준으로 사이버-물리 공격을 구체적으로 분류하였다. 또한, 사이버-물리 공격의 구현을 위한 중간자 공격에 대해 분석하고 물리시스템에 stealthy한 사이버-물리 공격이 적용되는 경우에는 물리시스템에 어떠한 변화가 발생하는지에 대해 시뮬레이션 결과를 바탕으로 확인하였다.

사이버물리시스템은 사회 기반 시설 전 분야에 적용되는 핵심 기술이며, 이에 따라 사이버물리시스템은 높은 수준의 보안이 요구된다. 사이버-물리 공격은 기존의 네트워크를 대상으로 하는 사이버 공격과는 달리 물리시스템에 대한 지식과 네트워크 지식이 융합되어 있는 정밀한 공격이며 기존의 사이버보안 기법으로는 검출이 어렵기 때문에 사이버-물리 공격의 검출과 방어를 위해서는 물리시스템의 동역학적 요소와 네트워크의 성질을 모두 고려해야 한다. 사회기반시설에 적용되는 사이버물리시스템의 안정적인 운영을 위해서는 사이버-물리 융합보안 연구를 위한 보다 집중적인 노력이 요구된다.

Acknowledgement

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (2014-0-00065, Resilient Cyber-Physical Systems Research)

참 고 문 헌

[1] K.-J. Park, R. Zheng, and X. Liu, "Cyber-physical systems: Milestones and research challenges," Elsevier Computer Communications, vol. 36, issue 1, pp.1-7, December 2012.

[2] K.-J. Park, J. Kim, H. Lim and Y. Eun, "Robust path diversity for network quality of service in cyber-physical systems," IEEE Transactions on Industrial Informatics, vol. 10, no. 4, pp. 2204-2215, November 2014.

[3] F. Pasqualetti, F. Dörfler and F. Bullo, "Attack detection and identification in cyber-physical systems," IEEE Transactions Automatic Control, vol. 58, no. 11, pp. 2715 - 2729, November 2013.

[4] J. Giraldo et el., "A Survey of physics-based attack detection in cyber-physical systems," ACM Computing Surveys, vol. 5, no. 4, pp. 76:1-76:36, July 2018.
[5] A. Teixeira, D. Perez, H. Sandberg and K. H. Johansson, "Attack models and scenarios for networked control systems," Conference on High Confidence Networked Systems (HiCoNS), Beijing, China, 2012.
[6] H. Zhang, P. Cheng, L. Shi and J. Chen, "Optimal DoS attack scheduling in wireless networked control system," IEEE Transactions on Control Systems Technology, vol. 24, no. 3, pp. 843-852, May 2016.
[7] Y. Mo and B. Sinopoli, "Secure control against replay attacks," 47th Annual Allerton Conference on Communication, Control, and Computing, IL, USA, 2009.
[8] Y. Mo, S. Weerakkody and B. Sinopoli, "Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs," vol. 35, issue 1, pp. 93-109, February 2015
[9] G. Park, H. Shim, C. Lee, Y. Eun and K. H. Johansson, "When adversary encounters uncertain cyber-physical systems: Robust zero-dynamics attack with disclosure resources," IEEE 55th Conference on Decision and Control (CDC), NV, USA, 2016.
[10] H. K. Khalil, Nonlinear Systems (3rd ed), Prentice Hall, 1996.
[11] A. Teixeira, I. Shames, H. Sandberg and K. H. Johansson, "Revealing stealthy attacks in control systems," Allerton Conference Communications, Control Computing, IL, USA, 2012.
[12] J. Back, J. Kim, C. Lee, G. Park and H. Shim, "Enhancement of security against zero dynamics attack via generalized hold," IEEE 56th Conference on Decision and Control (CDC), Melbourne, Australia, 2017.
[13] H. Jeon and Y. Eun, "A stealthy sensor attack for uncertain cyber-physical systems," IEEE Internet of Things Journal, vol. 6, no. 4, pp. 6345-6352, August 2019.
[14] R. S. Smith, "Covert misappropriation of networked control systems: Presenting a feedback structure," IEEE Control Systems Magazine, vol. 35, no. 1, pp. 82-92, February 2015.
[15] A. O. de Sá, L. F. R. d. C. Carmo and R. C. S. Machado, "Covert attacks in cyber-physical control systems," IEEE Transactions on Industrial Informatics, vol. 13, no. 4, pp. 1641-1651, August 2017.
[16] M. Conti, N. Dragoni and V. Lesyk, "A survey of man in the middle attacks," IEEE Communications Surveys & Tutorials, vol. 18, no. 3, pp. 2027-2051, August 2016.
[17] S. Kim, Y. Won, I. Park, Y. Eun and K.-J. Park, "Cyber-physical vulnerability analysis of communication-based train control," IEEE Internet of Things Journal, vol. 6, no. 4, pp. 6353-6362, Aug. 2019.
[18] Y. Won et al., "An attack-resilient CPS architecture for hierarchical control: A Case study on train control systems," IEEE Computer, vol. 51, no. 11, pp. 46-55, Nov. 2018.
[19] N. Hubballi, S. Biswas, S. Roopa, R. Ratti and S. Nandi, "LAN attack detection using discrete event systems," ISA Transactions, vol. 50, no. 1, pp. 119-130, January 2011.

약 력

김 상 준

2017년 부경대학교 공학사
2017년~현재 대구경북과학기술원 정보통신융합전공 석·박사통합과정
관심분야: Cyber-physical systems, Software defined networking

박 경 준

1998년 서울대학교 공학사
2000년 서울대학교 공학석사
2005년 서울대학교 공학박사
2005년~2006년 삼성전자 정보통신총괄 책임연구원
2006년~2010년 일리노이대학교(어버나-샴페인) 박사후연구원
2011년~현재 대구경북과학기술원 정보통신융합전공 교수
관심분야: Resilient cyber-physical systems, Smart factory optimization