논문 2019-14-27

칼만필터를 이용한 사이버 물리 시스템의 자율 복원성 확보 기법 및 자율주행차량 적용 연구

(Kalman Filter Based Resilient Cyber-Physical System and its Application to an Autonomous Vehicle)

김 재 훈, 김 동 길, 이 동 익*
(Jae- Hoon Kim, Dong- Gil Kim, Dong- Ik Lee)

Abstract : Recently, successful attacks on cyber-physical systems have been reported. As existing network security solutions are limited in preventing the system from malicious attacks, appropriate countermeasures are required from the perspective of the control. In this paper, the cyber and physical attacks are interpreted in terms of actuator and sensor attacks. Based on the interpretation, we suggest a strategy for designing Kalman filters to secure the resilience and safety of the system. Such a strategy is implemented in details to be applied for the lateral control of autonomous driving vehicle. A set of simulation results verify the performance of the proposed Kalman filters.

Keywords : Resilient control, Kalman filter, Actuator attack, Sensor attack, Autonomous vehicle

Ⅰ. 서 론

임베디드시스템은 가전제품, 자동차, 로봇, 의료기기 등 특정한 임무 수행을 목적으로 하는 독립적인 개체로 개발되어져 왔으며, 지속적으로 활용범위가 넓혀지고 있다. 향후의 임베디드시스템은 독립적으로 동작하는 전통적인 임베디드시스템에서 탈피하여, 네트워크 및 클라우드 등으로 지칭되는 사이버 세계와의 상호작용이 강조되는 사이버물리시스템 (Cyber Physical System, CPS)으로 발전될 것으로 기대되고 있다 [1].

CPS 형태의 임베디드시스템의 발전은 생활의 질을 한 단계 더 향상시킬 것으로 예상된다. 구체적으로 살펴보면 다음과 같다 [2]:

- 정보의 다양성: 물리세계에 대한 많은 정보를 네트워크와 클라우드를 통하여 언제든지 제공받을 수 있으며, 이를 활용하여 물리세계에 대한 이해도를 높일 수 있다.
- 시스템 자율성: 제한된 정보만을 이용하여 판단하고 행위를 취하던 임베디드시스템이 클라우드의 도움으로 타 객체가 수집한 정보 등 다양한 정보의 활용이 가능하고, 이는 높은 수준의 자율성 (autonomy) 구현이 가능하도록 한다.
- 시스템 안전성: 물리세계에 대한 충분한 정보는 임베디드시스템이 보다 빠르고 정확한 반응이 가능하도록 한다. 예를 들면, 교차로에 진입단계에서 주변의 다양한 센서에서 수집한 정보를 이용하여 위험요소를 빠르게 판별하고 적절한 반응으로 자동차의 안전성을 높일 수 있다.
- 유연한 복구성: 네트워크로 연결된 물리세계는 풍부한 정보를 클라우드에 전송할 수 있다. 한편으로는 클라우드로부터 언제든지 펌웨어 및 미들웨어 다운로드 받을 수 있으며, 이를 활용하면 긴급한 문제로부터 유연하면서 신속하게 복구를 가능하게 한다.

*Corresponding Author (dilee@ee.knu.ac.kr)
Received: Sep. 6, 2019, Revised: Sep. 22, 2019, Accepted Sep. 30, 2019.
J.H. Kim, D.I. Lee: Kyungpook National University.
D.G. Kim: Kyungil University.
※ 본 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(NRF 2017R1A2B4003008)과 2019년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF 2017R1C1B5076020).

ⓒ IEMEK J. Embed. Sys. Appl. 2019 Oct. 14(5) 239-247
ISSN : 1975-5066
http://dx.doi.org/10.14372/IEMEK.2019.14.5.239

그러나 네트워크 및 클라우드와 연결되어 있는 CPS는 외부로부터의 악의적 공격 가능성에 노출되어 있는 문제점이 있으며, 외부 (사이버) 공격에 의해서 물리적 시스템이 조작되어 경제적 손실을 야기한 사례가 보고된 바 있다. 잘 알려진 사례로는 사이버 공격에 의해 하수처리시스템의 밸브가 임의 조작되어 백만리터 이상의 하수가 건물, 공원 등으로 넘쳐 매우 큰 경제적 사회적 손실을 야기하였다 [3]. 그리고 산업용 제어시스템으로 널리 사용되는 Siemens의 제어 및 데이터획득시스템을 표적으로 하는 StuxNet 바이러스 공격도 잘 알려져 있다 [4]. 또한, Tencent 사의 킨 보안 연구소 해커팀은 원격에서 테슬라 차량의 통제권을 탈취하는 사례를 시연해 보였다 [5]. 그 밖의 사이버물리시스템의 보안 취약점을 악용한 사례들을 [6]에서 찾아 볼 수 있다.

근래에 들어 사이버 공격에 대한 연구가 활발히 이루어지고 있으며, 사이버 공격에 대한 모델 및 시나리오에 대한 연구가 이루어지고 있다 [7]. 사이버 공격 모델은 (1) 반복 공격 (relay attack), (2) 영 동역학 공격 (zero dynamics attack), (3) 바이어스 공격 (bias injection attack), (4) 서비스거부 공격 (denial of service, DoS) 그리고 (5) stealthy attack 등과 같은 공격 시나리오가 알려져 있다.

한편 사이버 공격을 탐지하여 공격을 방어하거나, 사이버 공격에 대처하여 가능한 제어 성능을 유지하기 위한 자율복원 제어 (Resilient Control)에 대한 연구도 활발히 이루어지고 있다 [8]. 자율복원 제어는 사이버공격에 대한 위험을 제거하는 전략에 따라서 크게 3가지로 분류할 수 있다: (1) 방지 (prevention), (2) 제한 (isolation), 그리고 (3)완화 (mitigation). 본 논문에서는 완화전략에 대해서 중점적으로 다룬다. [9]는 재밍에 의한 주기적인 DoS 공격에 대해서 안정성을 확보하는 방법에 대해서 제안하였다. [10]은 센서 및 무선네트워크의 공격에 대해서 receding-horizon stackelberg 제어를 이용한 안전성 확보에 대해서 연구하였다. [11]은 UAV (Unmanned aerial vehicle)의 센서 및 액추에이터를 목표로 하는 공격에 대해서 추정오차비용을 산출하고, 제어성능에 미치는 영향을 분석하였다. [12]는 센서 및 액추에이터에 대한 공격을 영향을 최소화하기 위해서 관측기 기반의 LQ제어기를 설계하고 성능을 분석하였다. 상기 논문은 사이버공격의 유무 또는 사이버공격에 의해서 나타는 영향을 상태변수 추정으로 판단하고 있다. 그러나 상태변수 추정은 외란으로 인하여 (물리공격은 외란으로 간주될 수 있다. 기술상 편의를 위해, 이하, 외란=물리공격), 정확한 추정이 어려우며, 특히 외란에 의한 상태변수의 변화를 사이버공격에 의한 변화라고 단정 지어서 부적절한 경보와 복구가 수행될 수 있다.

본 논문에서는 칼만필터 기반의 관측기를 이용하여 시스템의 상태변수를 추정할 뿐만 아니라 사이버공격과 물리 공격을 함께 고려하여 대처 가능한 자율복원 제어 기법을 제안한다. 제안한 기법은 자율주행차량의 조향제어 모델에 적용하여 사이버공격 탐지 및 제어보상 가능성에 대하여 확인하였다.

II. CPS 공격 모델 및 대처 방안

1. CPS 모델

본 논문에서는 다음과 같은 이산 시간 선형 시불변 (Linear Time Invariant) 사이버 물리 시스템을 고려한다.

$$x(k+1) = Ax(k) + Bu(k) + B_\alpha \alpha(k) \quad (1)$$

$$z(k) = Cx(k) + D_\beta \beta(k) \quad (2)$$

여기서, $x(k) \in R^n$, $u(k) \in R^m$, $z(k) \in R^p$는 각각 시스템의 상태 변수 벡터, 시스템 입력 벡터, 측정 벡터이다. 행렬 A, B, 그리고 C는 모두 알려져 있다. $\alpha(k) \in R^q$는 액추에이터 공격 (Actuator Attack) 벡터, $\beta(k) \in R^r$는 센서 공격 (Sensor attack) 벡터이다. B_α와 D_β는 임의의 상수 행렬이다.

일반성을 잃지 않고 우리는 식 (1)의 시스템 입력 $u(k)$를 액추에이터와 관련된 제어 입력 (Control input) $u_1(k) \in R^{m_1}$과 기준 입력 (Reference input) $u_2(k) \in R^{m_2}$로 구분할 수 있다. 여기서, $m_1 + m_2 = m$ 이다. 즉, $u(k) = [u_1^T(k)\ u_2^T(k)]^T$이고 따라서 $B = [B_1\ B_2]$이다. 상기 시스템이 (A, C)에 대해 관측가능 (Observable)하고, (A, B_1)에 대해 제어가능 (Controllable)하다면, 관측기 기반 상태 궤환 제어 입력 $u_1(k)$은 아래와 같이 나타낼 수 있다.

$$u_1(k) = -K\hat{x}(k) + v(k) \quad (3)$$

여기서, $\hat{x}(t)$는 관측기의 상태 추정 벡터, K는 극 배치 방법 (Pole placement method)을 이용하

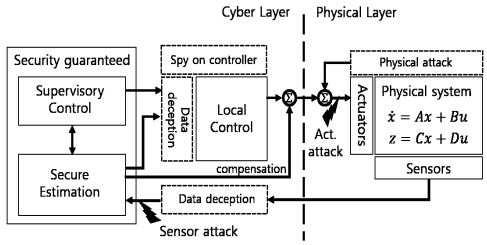

그림 1. 자율복원 제어시스템 아키텍처
Fig. 1 Resilient control system architecture

여 구할 수 있는 제어 이득 행렬, $v(t)$는 새로운 입력 벡터이다.

사이버 물리 시스템은 사용자 몰래 설치된 악성 코드나 스파이칩의 위협에 언제든 노출될 수 있기 때문에 그로 인해 야기될 성능 저하를 최대한 복구하고 피해를 최소화할 수 있어야 한다. False data injection을 가능하게 하는 또는 그렇게 불리기도 하는 Data deception 공격은 전형적인 악의적 공격 패턴 중 하나로 센서 및 관측기 등의 데이터를 조작하여 제어시스템에 심각한 영향을 미친다. 그림 1은 악의적 공격에 대항한 자율복원 제어시스템 구조를 나타낸다. CPS는 크게 사이버계층과 물리계층으로 나뉘며 사이버공격과 물리공격은 최종적으로 액추에이터 입력과 센서 출력을 변조하는 형태로 해석될 수 있다. 여기서 Secure Estimation은 보안 문제와 관련된 다양한 추정 기법 또는 관측기를 의미하며 상태변수를 추정함과 동시에 액추에이터 및 센서 공격에 따른 비정상적인 데이터 신호를 추정 및 식별하는 기능을 가진다. 감시 제어 (Supervisory Control)와 보안관련 추정 (Secure Estimation) 사이는 신뢰할 수 있는 양방향 데이터 전송이 가능하다고 가정하였으며, 이에 따라 로컬 제어 루프의 제어 신호의 흐름은 단방향으로 구성될 수 있다.

관측가능한 시스템에 대해 액추에이터 공격 및 센서 공격 검출을 위한 조건은 다음과 같다 [13].

$$B_\alpha \neq 0,\ D_\beta = 0,\ \text{and}\ q \leq p \tag{4}$$
$$\text{or},\ B_\alpha = 0,\ D_\beta \neq 0. \tag{5}$$

식 (4)와 (5)는 액추에이터와 센서가 동시에 공격당하지 않는다면 적절한 추정 기법을 이용하여 각각의 공격에 대해 검출 및 복구가 가능함을 의미한다. 특히, 식 (4)의 $q \leq p$조건은 이산 시스템에 대해 다음의 조건과 동치이다.

$$rank(B_\alpha) = q \tag{6}$$
$$rank(CB_\alpha) = rank(B_\alpha) \tag{7}$$

식 (6)은 B_α가 full column rank임을 의미하며 식 (6)과 (7)의 조건이 모두 만족되면 미지의 입력 관측기 (Unknown input obsever) 설계가 가능하고 [14], 이는 액추에이터 공격 식별을 위해 활용될 수 있다.

2. 액추에이터 공격 모델 및 대처 방안

본 절에서는 식 (4)의 조건을 만족하는 관측기 기반 상태 궤환 제어시스템을 고려한다. 액추에이터 공격은 제어 입력에 영향을 미치는 모든 요소를 대상으로 한다. 이러한 관점에서 볼 때, 광의적으로 기준 입력 $u_2(k)$에 대한 공격도 액추에이터 공격으로 간주될 수 있다. 다만, 물리 공격의 경우 실제 액추에이터와 관련된 제어 입력 $u_1(k)$에만 영향을 미칠 수 있다. 즉, $u_1(k)$을 구성하는 변수 $\{K, \hat{x}(k), v(k)\}$와 기준 입력 $u_2(k)$는 모두 사이버 공격에 의해 조작될 수 있다고 가정한다. 물리 공격을 포함한 액추에이터 공격 모델은 다음과 같다.

$$\begin{aligned} u_1(k) &= -(K+\Delta K)(\hat{x}(k)+\Delta\hat{x}(k)) \\ &\quad +(v(k)+\Delta v(k))+d(k) \\ &= -K\hat{x}(k)+v(k)+u_{1,\alpha}(k) \\ &= u_{1,des}(k)+u_{1,\alpha}(k) \end{aligned} \tag{8}$$
$$u_2(k) = u_{2,des}(k)+u_{2,\alpha}(k) \tag{9}$$

여기서, $d(k)$는 물리 공격 벡터, $u_{1,des}(k)$와 $u_{2,des}(k)$는 요구되는 (Desired) 입력 벡터, $u_{1,\alpha}(k)$와 $u_{2,\alpha}(k)$는 액추에이터 공격에 의해 발생되며 요구되는 시스템 입력을 교란하는 미지의 입력 벡터이다.

액추에이터 공격 하에, 칼만필터 기반의 관측기가 상태변수 추정을 위해 사용하는 입력 벡터 $u(k)$에는 상기 미지의 입력이 포함되어 있거나 포함되어 있지 않을 수 있다. 전자의 경우 사이버 공격과 관련이 있으며, 칼만필터도 시스템과 동일한 공격을 받고 있음으로 추정 값이 발산하지는 않으나 그만큼 사이버 공격을 식별하기가 더 어려워진다. 후자의 경우 물리 공격과 관련이 있으며, 미지의 입력을 적절히 고려하지 않고 설계된 칼만필터는 시스템에 인가되는 입력과 칼만필터에 인가되는 입력이 상이하기 때문에 액추에이터가 공격받는 즉시 발산한다.

물론, 사이버 공격과 물리 공격이 공존할 수 있으므로 상기 두 가지 문제점을 함께 다루는 통합된

해결책이 요구된다. 우선, 상태 변수 $x(k)$와 시스템 입력 $u(k)$를 동시에 추정하는 칼만필터를 설계함으로써, 일차적으로 필터가 발산하는 것을 방지할 수 있다. 나아가 감시 제어 레벨에서 생성되는 $v(k)$와 $u_2(k)$는 데이터 변조 없이 칼만필터에 제공되며, K는 기 설계된 값으로 칼만필터에 연역적 정보로 알려져 있다고 가정함으로써 액추에이터 공격 ($u_\alpha(k)$)을 식별하고 복구하는 것이 가능하다. 이와 같이 설계된 칼만필터를 이용하면 실시간 자율복원이 가능하다. 덧붙여, 식 (8) 그리고/또는 식 (9)을 식 (1)에 대입하면 필터 설계에 필요한 상태 공간 방정식이 명시 될 수 있다. 만약, $q=m$ 이면, $\alpha(k)=u_\alpha(k)$ 이고 $B_\alpha=B$ 이다.

3. 센서 공격 모델 및 대처 방안

칼만필터 기반의 상태 궤환 제어시스템에서 제어 입력은 칼만필터가 추정한 상태 변수에 의존한다. 그리고 칼만필터는 시스템 모델과 측정값을 기반으로 시스템의 상태변수를 추정하기 때문에, 센서 공격에 의해 측정값이 변조되면 추정값을 신뢰할 수 없게 되고 결국 비정상적인 제어 입력이 초래된다.

편의 상 $r=p$ 로 가정하면, $D_\beta=I_p$ (여기서, I_p는 $p\times p$ 크기의 단위행렬)이고 센서 공격을 포함한 측정 방정식은 아래와 같이 간단히 나타낼 수 있다.

$$z(k)=Cx(k)+\beta(k) \tag{10}$$

식 (8)을 $\beta(k)$에 대해 전개하면 다음과 같다.

$$\begin{aligned}\beta(k)&=z(k)-Cx(k)\\&=z(k)-C(Ax(k-1)+Bu(k-1))=\ \cdots\\&=z(k)-CA^k x(0)+\sum_{i=1}^{k}CA^{i-1}Bu(k-i)\end{aligned} \tag{11}$$

식 (5)의 조건에 따라 $B_\alpha=0$ 이면 액추에이터 공격이 없기 때문에, $k\geq 1$에 대해 시스템 입력 $u(k-i)$는 알려져 있다. 따라서 식 (11)는 시스템의 초기 상태 변수 $x(0)$를 칼만필터가 정확히 알고 있을 경우에 (즉, $\hat{x}(0)=x(0)$), 센서 공격을 식별하고 상쇄시킬 수 있음을 보여준다 [15].

Ⅲ. 응용 시스템

1. 자율주행 차량의 조향 제어시스템 모델

조향 제어를 위한 차량의 측방향 (lateral) 동역학 모델은 다음과 같이 bicycle (half-vehicle) 모델로 나타낼 수 있다 [16].

$$m(\ddot{y}+\dot{\psi}V_x)=F_{yf}+F_{yr} \tag{12}$$

$$I_z\ddot{\psi}=l_fF_{yf}-l_rF_{yr} \tag{13}$$

$$F_{yf}=2C_{\alpha f}(\delta-\theta_{Vf}) \tag{14}$$

$$F_{yr}=2C_{\alpha r}(-\theta_{Vr}) \tag{15}$$

$$\theta_{Vf}=\frac{\dot{y}+l_f\dot{\psi}}{V_x} \tag{16}$$

$$\theta_{Vf}=\frac{\dot{y}-l_r\dot{\psi}}{V_x} \tag{17}$$

여기서, $\dot{y}$은 차량의 측방향 속도, $\dot{\psi}$은 차량의 요 각속도 (Yaw rate), F_{yf}와 F_{yr}은 각각 전후면 측방향 타이어력 (Tire force), δ는 조향각 (Steer angle)을 제어하는 액추에이터 입력, $C_{\alpha f}$와 $C_{\alpha r}$은 각각 전후면 타이어의 코너링 강성도 (Cornering stiffness), θ_{Vf}와 θ_{Vr}은 각각 전후면 타이어의 슬립각 (Slip angle)을 유발하는 차량 종축 (Longitudinal axis 또는 x-axis)에 대한 속도각, I_z는 요 관성 모멘트, l_f와 l_r은 각각 차량의 무게중심으로부터 전후면 타이어까지의 종방향 거리, m은 차량의 질량, V_x는 차량의 종방향 주행 속도, 그리고 식 (14)와 (15)의 계수 2는 전후면에 위치한 두 개의 휠 (Wheel)을 의미한다.

일정한 종방향 속도 V_x를 가지는 차량이 회전반경 R인 원형 도로를 따라 주행한다고 가정하면 차량의 요 각속도를 위한 기준 입력 (Reference input)은 다음과 같이 설정될 수 있다.

$$\dot{\psi}_{ref}=\frac{V_x}{R} \tag{18}$$

식 (18)를 이용하여 아래의 두 오차 방정식을 정의할 수 있다.

$$\dot{e}_1=\dot{y}+V_x(\psi-\psi_{ref}) \tag{19}$$

$$e_2=\psi-\psi_{ref} \tag{20}$$

여기서 e_1는 측방향 위치 오차, e_2는 요각 (Yaw angle) 오차를 나타낸다. 시스템 방정식 (12) - (17)와 오차 방정식 (19) - (20)을 적절히 조합하여 이산화 과정을 거치면 다음과 같은 선형 시불변 상태 공간 방정식을 유도할 수 있다.

$$x(k+1)=Ax(k)+B_1\delta(k)+B_2\dot{\psi}_{ref}(k) \tag{21}$$

$$z(k)=Cx(k) \tag{22}$$

여기서, 상태 벡터는 $x=[e_1, \dot{e}_1, e_2, \dot{e}_2]^T$이고 측정 벡터는 $z=e_1$이다. 이 때, (A, C)은 관측가능(observable)하다.

식 (21)과 (22)로 정의된 측방향 주행 차량 시스템은 기준입력 $\dot{\psi}_{ref}(k)$과 상태 궤환 제어 입력 $\delta(k)$에 의해 제어된다. (A, B_1)은 제어가능 (Controllable) 함으로 상태 궤환 제어 입력 $\delta(k)$는 아래와 같이 정의 할 수 있다.

$$\delta(k) = -K\hat{x}(k) + \delta_{ff} \tag{23}$$

여기서, δ_{ff}는 회전 반경 R의 원형도로에 진입한 차량이 정상상태에 도달했을 때, e_1이 0이 되도록 정상상태 오차값을 보정하는 피드포워드 (Feed forward) 입력으로 상수이다. 참고로, 선회하는 차량이 정상상태에 도달하면 e_2는 차량의 슬립각에 수렴한다. 액추에이터 공격을 고려한 제어 입력은 다음과 같다.

$$\delta(k) = -K\hat{x}(k) + \delta_{ff} + \delta_{\alpha}(k) \tag{24}$$

여기서, $\delta_{\alpha}(k)$는 액추에이터 공격에 따른 제어 입력을 교란하는 미지의 입력이다.

기준 입력이 바뀌면 제어 입력도 바뀌므로, 기준 입력 $\dot{\psi}_{ref}(k)$의 변조 역시 액추에이터 공격으로 간주될 수 있다. 시스템 (21)은 선형 시불변이기 위해 차량의 종방향 속도 V_x는 항상 일정하다고 가정하므로, $\dot{\psi}_{ref}(k)$에 대한 공격은 도로의 회전 반경에 관한 정보가 조작된 것으로 가정한다.

$$\dot{\psi}_{ref}(k) = \frac{V_x}{R+\Delta R} \Leftrightarrow \Delta R = \frac{V_x - \dot{\psi}_{ref}(k)R}{\dot{\psi}_{ref}(k)} \tag{25}$$

따라서 기준 입력 $\dot{\psi}_{ref}(k)$에 대한 공격 모델은 다음과 같다.

$$\dot{\psi}_{ref}(k) = \frac{V_x}{R} + \frac{-V_x\Delta R}{R(R+\Delta R)} = \frac{V_x}{R} + \dot{\psi}_{ref,\alpha}(k) \tag{26}$$

마찬가지로, $\dot{\psi}_{ref,\alpha}$는 액추에이터 공격에 따른 기준 입력을 변화시키는 미지의 입력이다.

또한, 센서 공격을 포함한 측정 방정식은 다음과 같다.

$$z(k) = Cx(k) + \beta(k) = e_1(k) + e_{1,\beta}(k) \tag{27}$$

여기서, $e_{1,\beta}(k)$은 악의적으로 인가된 센서값이다.

2. 액추에이터 공격 관측기 설계

제어 입력 $\delta(k)$에 대한 공격을 검출하기 위한 칼만필터 모델 (F_1)은 다음과 같다.

$$F_1 : \begin{cases} x_a(k+1) = A_a x_a(k) + B_{2a}\dot{\psi}_{ref}(k) + w_p(k) & (28) \\ z(k) = C_a x_a(k) + w_m(k) & (29) \end{cases}$$

$$x_a(k) = [x^T(k) \;\; \delta(k)]^T \tag{30}$$

$$A_a = \begin{bmatrix} A & B_1 \\ 0_{1,4} & 1 \end{bmatrix}, \;\; B_{2a} = \begin{bmatrix} B_2 \\ 0 \end{bmatrix}, \;\; C_a = [C \;\; 0], \tag{31}$$

여기서 $x_a(k)$는 증강된 (Augmented) 상태 벡터로 정의된다. $w_p(k)$와 $w_m(k)$는 영평균 가우시안 잡음 (Zero mean Gaussian noise)이며, 각각 공정잡음 (Process noise)과 측정잡음 (Measurement noise)을 나타낸다. 또한, 추가된 제어 입력의 상태 천이 모델은 파라메터 추정 문제와 유사하게 $\delta(k+1)=\delta(k)$로 가정되었다. 칼만필터 (F_1)는 상태 변수를 추정함과 동시에 액추에이터 공격으로 인해 변조된 제어 입력의 추정 값을 제공한다. 칼만필터가 추정한 $\hat{x}(k)$와 $\hat{\delta}(k)$, 그리고 알려진 정보 K와 δ_{ff}를 이용하여 액추에이터 공격을 상쇄하기 위한 보상 값은 다음과 같이 구할 수 있다.

$$\delta_c(k) = -\delta_{\alpha}(k) = -(\hat{\delta}(k) + K\hat{x}(k) - \delta_{ff}) \tag{32}$$

제어 입력을 보상하게 되면 시스템 모델 (21)과 필터 모델 (28)은 각각 아래와 같이 변경됨에 유의해야 한다.

$$x(k+1) = Ax(k) + B_1(\delta(k) + \delta_c(k)) + B_2\dot{\psi}_{ref}(k) \tag{33}$$

$$\begin{aligned} x_a(k+1) = {} & A_a x_a(k) + B_{1a}\delta_c(k) \\ & + B_{2a}\dot{\psi}_{ref}(k) + w_p(k) \end{aligned} \tag{34}$$

여기서 $B_{1a} = [B_1^T \;\; 0]^T$이다.

기준 입력 $\dot{\psi}_{ref}(k)$에 대한 공격을 검출하기 위한 필터 모델 (F_2)은 다음과 같다.

$$F_2 : \begin{cases} x_a(k+1) = A_a x_a(k) + B_{1a}\delta(k) + w_p(k) & (35) \\ z(k) = C_a x_a(k) + w_m(k) & (36) \end{cases}$$

$$x_a(k) = [x^T(k) \;\; \dot{\psi}_{ref}(k)]^T \tag{37}$$

$$A_a = \begin{bmatrix} A & B_2 \\ 0_{1,4} & 1 \end{bmatrix}, \;\; B_{1a} = \begin{bmatrix} B_1 \\ 0 \end{bmatrix}, \;\; C_a = [C \;\; 0], \tag{38}$$

여기서, 기준 입력의 상태 공간에서의 천이 모델

은 $\dot{\psi}_{ref}(k+1)=\dot{\psi}_{ref}(k)$로 가정되었다. 칼만필터가 추정한 $\hat{\dot{\psi}}_{ref}(k)$와 알려진 정보 V_x와 R을 이용하여 $\dot{\psi}_{ref}$에 대한 보상 값은 아래와 같이 구해진다.

$$\dot{\psi}_{ref,c}(k)=-\dot{\psi}_{ref,\alpha}(k)=-(\hat{\dot{\psi}}_{ref}(k)-\frac{V_x}{R}) \quad (39)$$

마찬가지로, 기준 입력을 보상함에 따라 시스템 모델 (21)와 필터 모델 (35)은 다음과 같이 변경된다.

$$\begin{aligned} x(k+1) &= Ax(k)+B_1\delta(k) \\ &\quad +B_2(\dot{\psi}_{ref}(k)+\dot{\psi}_{ref,c}(k)) \end{aligned} \quad (40)$$

$$\begin{aligned} x_a(k+1) &= A_a x_a(k)+B_{1a}\delta(k) \\ &\quad +B_{2a}\dot{\psi}_{ref,c}(k)+w_p(k) \end{aligned} \quad (41)$$

여기서 $B_{2b}=[B_2^T\ 0]^T$이다.

한편, 측방향 주행 제어시스템에서 측정값은 e_1 하나이기 때문에 $\delta(k)$와 $\dot{\psi}_{ref}(k)$가 동시에 공격받는 상황은 식 (4)의 미지의 입력 관측기 설계 조건 ($q\le p$)을 만족하지 않는다. 만약 상기 두 입력이 동시에 공격당할 경우, 이를 검출하고 복구하기 위해서는 추가적인 센서 정보를 활용할 수 있어야 하며 식 (4)의 조건이 만족되면 제안된 칼만필터 기반의 액추에이터 공격 추정 기법은 적절히 확장될 수 있다.

3. 센서 공격 관측기 설계

센서 공격은 식 (5)의 조건을 만족함과 더불어 Ⅱ.3절에서 논의 된 바와 같이 시스템의 초기 상태 값을 필터가 알고 있을 경우에 식별될 수 있다. 센서 공격 벡터 $e_{1,\beta}(k)$ 검출을 위한 칼만필터 모델 (F_3)은 다음과 같다.

$$F_3: \begin{cases} x_a(k+1)=A_a x_a(k)+B_a u(k)+w_p(k) & (42) \\ z(k)=C_a x_a(k)+w_m(k) & (43) \end{cases}$$

$$x_a(k)=[x^T(k)\ \ e_{1,\beta}(k)]^T \quad (44)$$

$$u_a(k)=[\delta(k)\ \ \dot{\psi}_{ref}(k)]^T \quad (45)$$

$$A_a=\begin{bmatrix} A & 0_{4,1} \\ -CA+C & 1 \end{bmatrix},\ B_a=\begin{bmatrix} B_1 & B_2 \\ -CB_1 & -CB_2 \end{bmatrix},\ C_a=[C\ \ 1], \quad (46)$$

여기서, $e_{1,\beta}(k)$의 상태 천이 모델은 정상상태 조건에서 $z(k+1)\cong z(k)$라고 가정하면 얻을 수 있다. 액추에이터 공격 식별 및 보상 메커니즘과는 달리, 센서 공격은 $e_{1,\beta}(k)$를 직접적으로 추정함으로써 칼만필터 내부에서 식별되고 상쇄된다.

Ⅳ. 시뮬레이션

본 절에서는 상기 자율주행 차량의 조향제어 시스템에 대해 액추에이터 및 센서 공격 예시에 따라 표준 칼만필터 (F_0)와의 비교를 통해 제안된 칼만필터 F_1, F_2, F_3의 성능을 평가한다. 시뮬레이션은 Matlab/Simulink를 이용하여 수행되었다.

1. 주행 및 공격 시나리오

일정한 종방향 속도를 가지는 자율주행 차량이 직선 주행하다가 $t_k=5s$에서 회전반경이 R인 원형 도로에 진입한다. 고려된 주행 시나리오에 따른 기준 입력 $\dot{\psi}_{ref}(k)$는 다음과 같다.

$$\dot{\psi}_{ref}(t_k)=\begin{cases} 0, & 0\le t_k<5s, \\ \dfrac{V_x}{R}, & 5s\le t_k. \end{cases} \quad (47)$$

또한, 정상상태의 측방향 위치 오차 $e_{1,ss}$를 보정하는 피드포워드 입력 δ_{ff}는 다음과 같다.

$$\delta_{ff}=\begin{cases} 0, & \text{if } \dot{\psi}_{ref}(t_k)=0 \\ \dfrac{l_f+l_r}{R}+K_V a_y+k_3 e_{2,ss}, & \text{if } \dot{\psi}_{ref}(t_k)\neq 0. \end{cases} \quad (48)$$

$$a_y=\frac{V_x^2}{R} \quad (49)$$

$$K_V=\frac{l_r m}{2C_{\alpha f}(l_f+l_r)}-\frac{l_f m}{2C_{\alpha r}(l_f+l_r)} \quad (50)$$

$$e_{2,ss}=-\frac{l_r}{R}+\frac{l_f}{2C_{\alpha f}(l_f+l_r)}\frac{mV_x^2}{R}, \quad (51)$$

여기서, a_y는 정상상태에서의 차량의 측방향 관성가속도 (Inertial acceleration), K_V는 차량의 무게중심과 타이어의 성능 특성에 따라 결정되는 Understeer gradient 계수, $e_{2,ss}$는 요 각의 정상상태 오차 값이다. k_3은 상태 궤환 제어 이득 행렬 K의 3번째 성분값이다. 이 때, K를 설계하기 위한 폐루프 시스템의 요구되는 극점은 [-5-3i, -5+3i, -7, -10]으로 가정하였다. 그 외, 시뮬레이션을 위한 시스템 파라메터는 $m=1573$, $I_z=2873$, $l_f=1.1$, $l_r=1.58$, $C_{\alpha f}=80000$, $C_{\alpha r}=80000$ 이다[16]. 이 값들을 제안한 칼만필터 모델 F_1, F_2, F_3에 대입하여 관측가능성을 판별하면 모두 관측가능함을 알 수 있다.

시뮬레이션에서 고려된 제어 입력과 관련된 액추에이터 공격은 다음과 같이 인가된다: 사이버 공

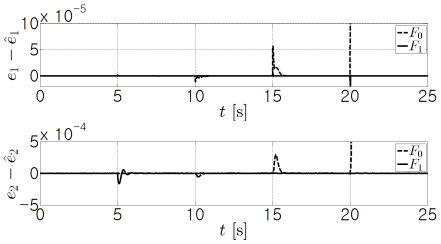

그림 2. 액추에이터 공격에 따른 추정 오차
Fig. 2 Estimation errors under actuator attack

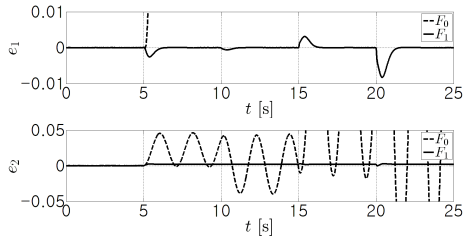

그림 3. 액추에이터 공격에 따른 성능 저하 및 복구
Fig. 3 Control performance degradation and reconstruction under actuator attack

격에 의해 $t_k = 1s$에서 폐루프 시스템의 극점이 [-3i, 3i, 0, -10]이 되도록 K가 변조되고 δ_{ff}와 $\hat{x}(k)$가 각각 $t_k = 10s, 15s$에서 추가적으로 변조된다. 이 후 $t_k = 20s$에서 물리 공격이 추가되었다. 기준 입력과 관련된 액추에이터 공격 또는 센서 공격은 $t_k = 10s$에서 인가된다. 모든 공격은 스텝 입력 신호로 가정하였다.

2. 시뮬레이션 결과

그림 2와 3은 제어 입력과 관련된 액추에이터 공격에 따른 추정 오차 및 제어 성능을 보여준다. 그림 2에서 표준 칼만필터의 추정값은 세 차례의 사이버 공격이 인가되었음에도 불구하고 발산하지 않다가 물리 공격이 인가되는 시점인 $t_k = 20s$에서 발산한다. 그러나 그림 3에서 표준 칼만필터 기반의 제어시스템은 $t_k = 5s$에서 발산하였음을 확인할 수 있다. 이러한 현상은 표준 칼만필터의 경우 시스템과 동일한 사이버 공격을 받기 때문에 나타나며 그 결과 추정값의 정확도와 관계없이 제어시스템은 발산할 수 있음을 보여준다. 표준 칼만필터 기반의 제어시스템이 $t_k = 5s$에서 발산한 이유는 K의 변조는 직선 주행과 무관하며 따라서 그 영향은 차량이

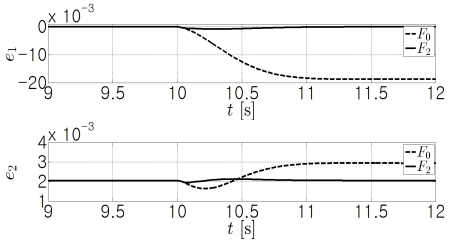

그림 4. 기준입력 변조에 따른 성능 저하 및 복구
Fig. 4 Performance degradation and reconstuction under reference input deception attack

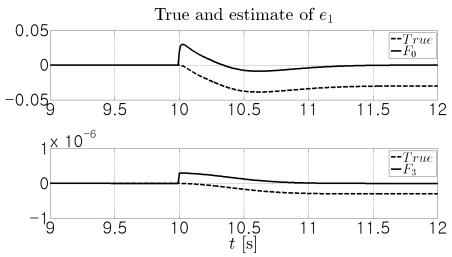

그림 5. 센서 공격에 따른 추정 결과
Fig. 5 Estimation results under sensor attack

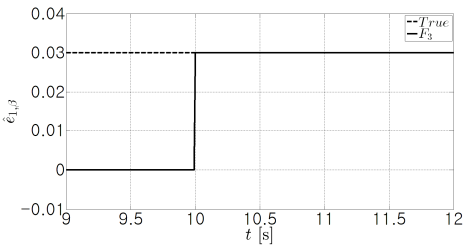

그림 6. 바이어스 인가 센서 공격 추정
Fig. 6 Estimation of bias injection attack

$t_k = 5s$에서 원형도로에 진입하는 순간 발현되었기 때문이다. 반면, 제안한 칼만필터 F_1 기반의 제어시스템은 모든 공격에 대해 적절한 보상 메커니즘을 통해 성공적으로 복구된다.

기준 입력에 대한 액추에이터 사이버 공격에 따른 표준 칼만필터 (F_0)와 제안한 칼만필터 (F_2) 기반의 제어 성능은 그림 4에 나타내었다. 악의적으로 조작된 기준입력을 신뢰하는 표준 칼만필터는 제어시스템에 정상상태 오차를 야기하였고 제안한 칼만필터 F_2는 조작된 기준입력을 식별하고 이를 보상

함으로써 저하된 제어 성능을 빠르게 복구한다.

그림 5는 센서 공격에 대한 표준 칼만필터 (F_0)와 제안한 칼만필터 (F_3)의 e_1에 대한 추정 결과를 보여준다. 표준 칼만필터의 경우, $t_k = 10s$에서 공격자에 의해 인가된 Bias 성분만큼 제어시스템의 정상상태 오차를 유발한다. 반면 모델 F_3을 기반으로 하는 칼만필터의 추정 오차 및 제어시스템의 정상상태 오차는 무시할만한 수준으로 줄어든다. 상쇄되지 못한 Bias 성분은 모델 F_3을 도출함에 있어 정상상태 가정으로 인해 나타나는 잔여 성분이다. 그림 6는 제안한 칼만필터 (F_3)가 추정한 Bias 성분 값을 보여준다.

V. 결 론

본 논문에서는 사이버 물리 시스템의 안전성 및 보안성 문제와 관련하여 사이버 공격 및 물리 공격을 액추에이터와 센서 관점에서 분석하고 이를 식별하고 대처하기 위한 칼만필터 기반의 추정 기법을 제시하였다. 제안된 추정 기법은 상태 궤환 제어를 위해 상태변수를 추정할 뿐만 아니라 자율복원 제어를 위해 악의적인 공격에 의해 야기되는 비정상적인 신호를 추정한다. 상기 칼만필터 기반의 추정 기법은 자율주행 자동차의 조향제어 시스템에 대해 고려된 공격 유형에 따라 구체화되었으며 그 성능을 시뮬레이션을 통해 검증하였다.

References

[1] M. Bhrugubanda, "A Review on Applications of Cyber Physical Systems," Journal of Innovative Research in Science, Engineering and Technology, Vol. 2, No. 6, pp. 728-730, 2015.

[2] Y.S. Eun, K.J. Park, M.G. Won, T.J. Park, S.H. Son, "Recent Trends in Cyber-Physical Systems Research," Journal of Communications of the Korea Institute of Information Scientists and Engineers, Vol. 30, No. 12, pp. 8-15, 2013 (in Korean).

[3] J. Slay, M. Miller, "Lessons Learned From the Maroochy Water Breach," Journal of Critical Infrastructure Protection, Vol. 253, pp. 73-82, 2007.

[4] Avaliable on : https://www.computerworld.com/article/2515570

[5] Avaliable on : https://keenlab.tencent.com/en/2016/09/19/Keen-Security-Lab-of-Tencent-Car-Hacking-Research-Remote-Attack-to-Tesla-Cars/

[6] A. Cárdenas, S. Amin, S. Sastry, "Research Challenges for the Security of Control Systems," Proceedings of 3rd Conference Hot Topics Security, pp. 1-6, 2008.

[7] A. Teixeira, D. Pérez, H. Sandberg, K.H. Johansson, "Attack Models and Scenarios for Networked Control System," Proceedings of 1st Conference High Confidence Networked Systems, pp. 55-64, 2012.

[8] Y.Z. Lun, A. D'Innocenzo, F. Smarra, I. Malavolta, M.D. Di Bendetto, "State of the Art of Cyber-physical Systems Security: An Automatic Control Perspective," Journal of System Software, Vol. 149, pp. 174-216, 2019.

[9] H. Shisheh Foroush, S. Martínez, "On Multi-input Controllable Linear Systems Under Unknown Periodic dos Jamming Attacks," Proceedings of SIAM Conference Control and its Applications, 2013.

[10] M. Zhu, S. Martínez, "Stackelberg-game Analysis of Correlated Attacks in Cyber-physical systems," Proceedings of American Control Conference, pp. 4063-4068, 2011.

[11] C. Won, I. Hwang, "Analytical Analysis of Cyber Attacks on Unmanned Aerial Systems," Proceedings of AIAA Conference on Guidance Navigation and Control, 2013.

[12] S.M. Djouadi, A.M. Melin, E.M. Ferragut, J.A. Laska, J. Dong, "Finite Energy and Bounded Attacks on Control System Sensor Signals," Proceedings of American Control Conference (ACC), pp. 1716-1722, 2014.

[13] W. Ao, Y.D. Song, C.Y. Wen, "Adaptive Cyber-physical System Attack Detection and Reconstruction With Application to Power System," Journal of IET Control Theory and Applications, Vol. 10, No. 12, pp. 1458-1468, 2016.

[14] M. Darouach, M. Zasadzinski, A.B. Onana, S. Nowakowski, "Kalman Filtering With Unknown Inputs via Optimal State Estimation of Singular Systems," Journal of

Systems Science, Vol. 26, No. 10, pp. 2015-2028, 1995.
[15] H. Fawzi, P. Tabuada, S. Diggavi, "Secure Estimation and Control for Cyber-Physical Systems Under Adversarial Attacks," Journal of IEEE Transactions on Automatic Control, Vol. 59, No. 6, pp. 1454-1467, 2014.
[16] R. Rajamani, Vehicle Dynamics and Control, New York, Sringer-Verlag, 2005.

Jae-Hoon Kim (김 재 훈)

He received B.S. and M.S. degrees in Electronics Engineering from Kyungpook National University, Korea, in 2011 and 2013 respectively. He is currently a Ph.D. student in Electronics engineering from Kyungpook National University, Korea. His current research interests include control and estimation with regard to diagnostics, reliability, and safety of nonlinear systems.

Email: jaehoon87kim@gmail.com

Dong-Gil Kim (김 동 길)

He received B.S., M.S., and Ph. D. degrees in Electronics Engineering from Kyungpook National University, Korea, in 2006, 2008, and 2015 respectively. He worked as senior researcher at Defence Agency for Technology and Quality from 2014 to 2017. He is currently an associate professor with Robot Engineering at Kyungil University, Korea from 2017.

Email: eastroad@gmail.com; dgkim@kiu.kr

Dong-Ik Lee (이 동 익)

He received B.S. and M.S. degrees in Electronics Engineering from Kyungpook National University, Korea, in 1987 and 1990 respectively. He worked as researcher at Agency for Defense Development from 1990 to 1997. He received Ph.D. degree in Department of Automatic Control and Systems Engineering at the University of Sheffield, England, 2002. He worked as common founder and CTO at DRTS Ltd England from Jan. 2002 to March 2005. He is currently an professor with Electronics Engineering at Kyungpook National University, Korea from 2005.

Email: dilee@ee.knu.ac.kr