Synchronous Detection Technique for Actuator/Sensor Health Monitoring in Control Systems under Deception Attacks

Maruthi T. Ravichandran and Liang Zhang

***Abstract*—Resilient control systems are feedback systems, which maintain an acceptable level of performance in the presence of attacks on the sensors and actuators. This paper addresses the design of control systems, which are resilient under a deception attack on the DC gains of the sensor and/or the actuator. The approach is based on the method of *synchronous detection*, which is widely used in communication systems. It consists of two steps: First, the "health" of the sensor and the actuator, which are assumed to be under attack, is assessed. Second, based on this assessment, the controller is modified (if possible) so that the effects of the attacks on the closed-loop system response are eliminated. The above approach is applied to a model of uranium enrichment centrifuge control system, and, using simulations, is shown to provide effective protection against the attacks.**

***Index Terms*—Resilient control systems; Deception attack; Attack detection; Synchronous detection**

I. Introduction

Resilient control systems are feedback systems, which are capable of detecting and mitigating malicious attacks on their sensors and actuators, wherein the attacks are intended to force the plant output to deviate substantially from the reference signal. In the absence of appropriate detection and mitigation strategies, attacks may lead to unwanted consequences, such as damage to the plant. For example, consider the drive system of a Uranium gas enrichment centrifuge, which typically consists of a three-phase AC induction motor, a controller, and a speed sensor. Since this system operates in a closed-loop configuration, an attack on the sensor that forces it to project a "low" speed may lead to the actual centrifuge rotational speed taking dangerously high values.

With growing concern over the security of critical infrastructure systems such as nuclear power plants, water distribution network, etc., the research on attack detection and mitigation in control systems has gained significant attention recently [1]–[5]. The security of cyber-physical systems, involved in these critical infrastructures, typically consists of *confidentiality* (preventing information disclosure to attackers), *integrity* (preventing data/resource modification or deletion by attackers), and *availability* (preventing communication and control software/hardware failures) [6], [7]. In this paper, we focus on the attacks that aim at compromising the integrity of the system at hand, often referred to as *deception attacks*.

To defend against deception attacks in control systems, different types of attack detectors may be developed. Among existing results in the literature, the most commonly used approach is to design a state estimator, and detect the attack based on the estimation residue, i.e., the difference between the measurement data and the estimator output (see, for instance, [8]–[13]). This approach, however, is not effective when the attacks are hidden in the transient dynamics or aligned with the noise statistics. Another approach is to inject a randomly generated "watermark" signal – stationary and Gaussian – into the control system, and search for the watermark in the measured outputs (by evaluating the correlation between the output signals and the injected watermark signal) [14], [15]. It is shown that this approach is effective against *replay attacks*, i.e., attacks wherein valid data transmission is maliciously repeated. On the other hand, due to its reliance on statistical correlation, this approach is not capable of detecting other deception attacks (e.g., multiplying the outputs with a constant).

In the current paper, we assume that the attacker *modifies the DC gain/s of the sensor and/or the actuator* of the control system, and propose a novel approach to monitor these DC gains in real-time. This approach is based on the method of *synchronous detection* [16], which is used in analog communication systems for separating the carrier and the information signals. Using this approach, one can effectively detect, identify, and mitigate the attack. Other types of deception attacks and other classes of attacks will be considered elsewhere.

The remainder of this paper is organized as follows: Section II presents the models of the system and the attacker. The synchronous detection-based approach to detect and mitigate the attack is introduced in Section III, and is discussed in details in Sections IV and V. An application to a model of uranium enrichment centrifuge control system is shown in Section VI. Finally, the conclusions and directions of future work are provided in Section VII. All proofs are included in the Appendix.

II. Modeling

A. System model

Consider the nominal (non-attacked) feedback linear time-invariant system shown in Fig. 1, where $K(s)$, $A(s)$, $P(s)$, and $S(s)$ represent the transfer functions of the controller, actuator, plant, and sensor, respectively, and the gain of the

M. T. Ravichandran is currently a member of the research staff at Ford Motor Company, Dearborn, MI 48120, USA (email: mravich5@ford.com).

L. Zhang (corresponding author) is with the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT 06269, USA (e-mail: liang.zhang@uconn.edu).

pre-compensator S_0 is equal to the DC gain of the sensor, i.e., $S_0 = \lim_{s\to 0} S(s)$. In addition, let A_0 denote the nominal DC gain of the actuator. Note that the pre-compensator is used in the system so that the plant output y tracks the reference input r. Clearly, the closed-loop transfer function from r to y is given by

$$G(s) = \frac{S_0 K(s)A(s)P(s)}{1 + K(s)A(s)P(s)S(s)}. \tag{1}$$

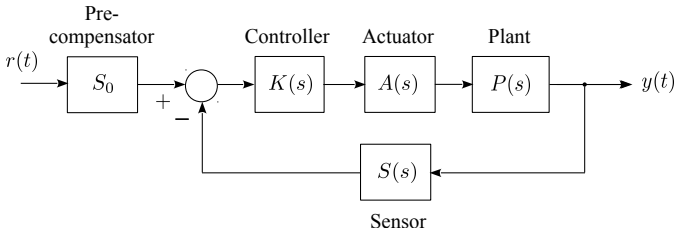


Fig. 1. Nominal feedback control system

B. Attacker model

Assume that a deception attack is imposed on the sensor and/or the actuator of the feedback control system shown in Fig. 1. Depending on the attacker's action, we assume that a deception attack can be categorized as follows:

1) Type 1 deception attack: The transfer functions of the sensor and/or the actuator are modified to $\widetilde{S}(s)$ and $\widetilde{A}(s)$, respectively (see Fig. 2(a)).

2) Type 2 deception attack: External deceptive signals are projected as the outputs of the sensor and/or the actuator (see Fig. 2(b)).

3) Type 3 deception attack: External deceptive signals are added to the outputs of the plant and/or the actuator (see Fig. 2(c)).

In this paper, we focus on a special case of Type 1 deception attack, while other types of attacks will be studied elsewhere. Specifically, for the Type 1 deception attack considered in this paper, we assume that the DC gains of the sensor and/or the actuator are modified from S_0 to S_a and from A_0 to A_a, respectively. Thus, the transfer functions of the sensor and the actuator are modified to

$$\widetilde{S}(s) = \frac{S(s)S_a}{S_0}, \qquad \widetilde{A}(s) = \frac{A(s)A_a}{A_0}. \tag{2}$$

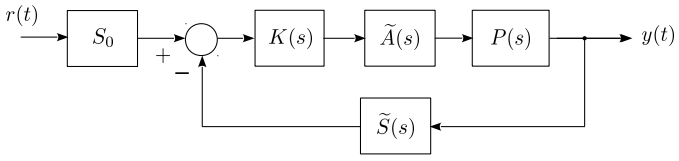
In addition, the following assumptions are made to facilitate the subsequent analysis:

Assumption 1: *a) The controller, plant, nominal actuator, and nominal sensor are open-loop asymptotically stable. b) The attacked actuator and the attacked sensor are open-loop asymptotically stable. c) The nominal and the attacked closed-loop systems are asymptotically stable.*

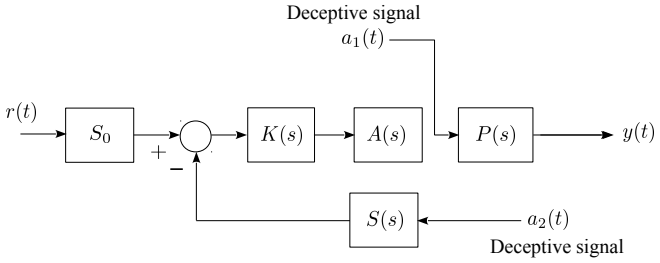
To identify and mitigate such a deception attack, a novel approach is proposed in this paper, and is discussed next.

III. Approach

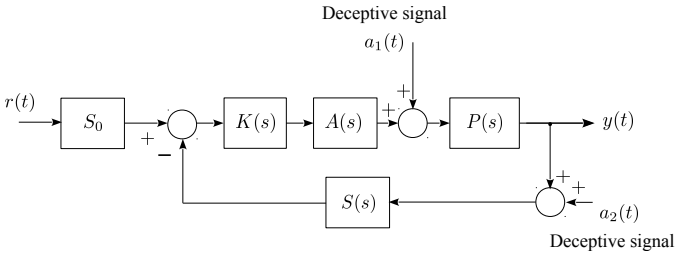
The approach proposed is based on the method of synchronous detection, which, as mentioned before, is used in analog communication systems to recover the information signal from the modulated one. As illustrated in Fig. 3, the recovery is achieved by first multiplying the modulated signal $m(t)\sin(\omega t)$ by $\sin(\omega t)$ and then passing the resulting signal through a lowpass filter.



(a) Type 1 deception attack



(b) Type 2 deception attack



(c) Type 3 deception attack

Fig. 2. Deception attack on a feedback control system

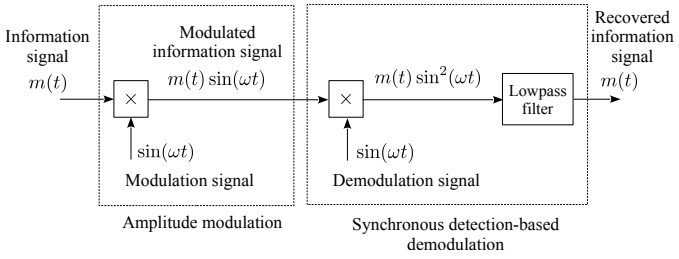


Fig. 3. Synchronous detection method in analog communication

In this paper, we apply this method to identifying a deception attack in a control system. Specifically, as illustrated in Fig. 4, this approach consists of the following:

- **Step 1:** Adding two sinusoidal signals $\alpha\sin(\omega t)$ and $\beta\sin(2\omega t)$ to the reference level r.
- **Step 2:** Multiplying the outputs of the actuator (v_1) and the sensor (v_2) by the same sinusoidal signals.
- **Step 3:** Computing the moving average of the signals resulting from Step 2 over time interval $T = 2\pi/\omega$ to obtain signals z_{11}, z_{12}, z_{21}, and z_{22}.

Note that the above procedure involves two carrier signals (one with frequency ω and the other with frequency 2ω). As

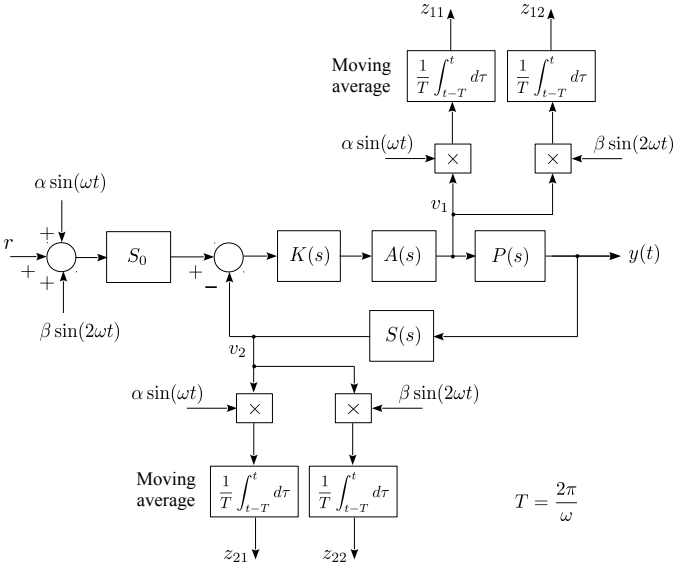


Fig. 4. Identification of deception attacks using synchronous detection

explained in Section IV, this is necessary to uniquely identify the attacker's behavior. In addition, since the signals resulting from Step 2 are all periodic with T being the smallest common period, the lowpass filter in Fig. 3 is replaced by the moving average blocks, where the moving averages are computed over time interval T for the sake of faster computation of z_{ij}'s.

Following the above procedure, the outputs of the moving average blocks, z_{11}, z_{12}, z_{21}, and z_{22}, are analyzed from the point of view of their consistency with the nominal values. As discussed in Section IV, an attack on the sensor and the actuator leads to these z_{ij}'s taking steady state values that differ from their nominal ones, which gives rise to the detection of the attacker's action. Moreover, the introduction of parameters α, β, and ω allows the user to change the nominal values of the z_{ij}'s, and, thus, have the potential to significantly increase the system's resilience to attacks.

Regarding the mitigation of a deception attack, it is based on the results of the detection procedure. Specifically, the DC gains of the attacked sensor and actuator are calculated based on the values of z_{11}, z_{12}, z_{21}, and z_{22}. Then, the controller and the pre-compensator are modified accordingly to ensure that the plant output is close to the reference signal.

IV. Attack detection and identification

A. Nominal system operation

Consider the feedback control system with synchronous detection as shown in Fig. 4. Let $v_1(t)$ and $v_2(t)$ denote the output signals of the actuator and the sensor, respectively, and let $G_i(s)$, $i = 1, 2$, denote the transfer functions from the reference r to outputs v_i, $i = 1, 2$, i.e.,

$$V_1(s) = G_1(s)R(s), \quad V_2(s) = G_2(s)R(s), \tag{3}$$

where $R(s)$ and $V_i(s)$ are the Laplace transforms of r and v_i, respectively. Then, based on the system block diagram, we obtain

$$G_1(s) = \frac{S_0K(s)A(s)}{1 + K(s)A(s)P(s)S(s)}, \tag{4}$$

$$G_2(s) = \frac{S_0K(s)A(s)P(s)S(s)}{1 + K(s)A(s)P(s)S(s)}. \tag{5}$$

Under Assumption 1, the steady state of signals v_1 and v_2 exists and is given by:

$$\begin{aligned} v_{1,ss}(t) &= rG_1(0) + \alpha|G_1(j\omega)|\sin(\omega t + \angle G_1(j\omega)) \\ &\quad + \beta|G_1(j2\omega)|\sin(2\omega t + \angle G_1(j2\omega)), \\ v_{2,ss}(t) &= rG_2(0) + \alpha|G_2(j\omega)|\sin(\omega t + \angle G_2(j\omega)) \\ &\quad + \beta|G_2(j2\omega)|\sin(2\omega t + \angle G_2(j2\omega)). \end{aligned}$$

Thus, with $T = 2\pi/\omega$ denoting the period of the oscillations $\sin(\omega t)$, the steady state value of the moving average signal z_{11} can be calculated as:

$$\begin{aligned} z_{11,ss} &= \frac{1}{T}\int_{t-T}^{t} \alpha v_{1,ss}(\tau)\sin(\omega\tau)d\tau \\ &= \frac{rG_1(0)\alpha}{T}\int_{t-T}^{t} \sin(\omega\tau)d\tau + \\ &\quad \frac{\alpha^2|G_1(j\omega)|}{T}\int_{t-T}^{t} \sin(\omega\tau + \angle G_1(j\omega))\sin(\omega\tau)d\tau + \\ &\quad \frac{\alpha\beta|G_1(j2\omega)|}{T}\int_{t-T}^{t} \sin(2\omega\tau + \angle G_1(j2\omega))\sin(\omega\tau)d\tau \\ &= \frac{\alpha^2|G_1(j\omega)|}{2T}\cdot \\ &\quad \int_{t-T}^{t} [\cos(\angle G_1(j\omega)) - \cos(2\omega\tau + \angle G_1(j\omega))]\, d\tau \\ &= \frac{\alpha^2}{2}|G_1(j\omega)|\cos(\angle G_1(j\omega)) \\ &= \frac{\alpha^2}{2}\mathrm{Re}\{G_1(j\omega)\}. \end{aligned} \tag{6}$$

Similarly,

$$\begin{aligned} z_{12,ss} &= \frac{\beta^2}{2}\mathrm{Re}\{G_1(j2\omega)\}, \\ z_{21,ss} &= \frac{\alpha^2}{2}\mathrm{Re}\{G_2(j\omega)\}, \\ z_{22,ss} &= \frac{\beta^2}{2}\mathrm{Re}\{G_2(j2\omega)\}. \end{aligned} \tag{7}$$

The above expressions (6) and (7) can be rewritten as

$$\begin{aligned} z_{11,ss} = \frac{\alpha^2 S_0}{2D_1(\omega)}\Big(&\mathrm{Re}(\overline{KA}(j\omega))\left[1 + \mathrm{Re}(\overline{KAPS}(j\omega))\right] + \\ &\mathrm{Im}(\overline{KA}(j\omega))\mathrm{Im}(\overline{KAPS}(j\omega))\Big), \end{aligned} \tag{8}$$

$$\begin{aligned} z_{12,ss} = \frac{\beta^2 S_0}{2D_1(2\omega)}\Big(&\mathrm{Re}(\overline{KA}(j2\omega))\left[1 + \mathrm{Re}(\overline{KAPS}(j2\omega))\right] + \\ &\mathrm{Im}(\overline{KA}(j2\omega))\mathrm{Im}(\overline{KAPS}(j2\omega))\Big), \end{aligned} \tag{9}$$

$$z_{21,ss} = \frac{\alpha^2 S_0}{2D_1(\omega)}\cdot\Big(\mathrm{Re}(\overline{KAPS}(j\omega)) + |\overline{KAPS}(j\omega)|^2\Big), \tag{10}$$

$$z_{22,ss} = \frac{\beta^2 S_0}{2D_1(2\omega)} \cdot \Big(\text{Re}(\overline{KAPS}(j2\omega)) + |\overline{KAPS}(j2\omega)|^2\Big), \tag{11}$$

where

$$\begin{aligned}
&\overline{KA}(jx) = K(jx)A(jx),\\
&\overline{KAPS}(jx) = K(jx)A(jx)P(jx)S(jx),\\
&D_1(x) = [1 + \text{Re}(\overline{KAPS}(jx))]^2 + [\text{Im}(\overline{KAPS}(jx))]^2,\\
&\qquad x = \omega, 2\omega.
\end{aligned}$$

As one may observe from the above (8)-(11), the steady state z_{ij}'s are functions of the amplitudes (α, β) and frequencies (ω, 2ω) of the sinusoidal signals, and are *independent* of the reference level r. Clearly, when an attacker modifies the transfer function of any of the system components, the steady state values of the z_{ij}'s will change from the nominal values, (8)-(11), thus leading to the detection of the attack. Moreover, this approach provides system operators the flexibility to change the nominal values of $z_{11,ss}$, $z_{12,ss}$, $z_{21,ss}$ and $z_{22,ss}$ when needed by using a new set of values for α, β, and ω.

B. Attacked system operation

As mentioned in Subsection II-A, we assume that the attacker modifies the DC gains of the sensor and the actuator, i.e., $\lim_{s\to 0}\widetilde{S}(s) = S_a$ and $\lim_{s\to 0}\widetilde{A}(s) = A_a$. Then, the transfer functions from the reference r to outputs $v_1(t)$ and $v_2(t)$ become

$$G_{1,a}(s) = \frac{S_0^2 A_a K(s)A(s)}{S_0A_0 + S_aA_aK(s)A(s)P(s)S(s)}, \tag{12}$$

$$G_{2,a}(s) = \frac{S_0 S_a A_a K(s)A(s)P(s)S(s)}{S_0A_0 + S_aA_aK(s)A(s)P(s)S(s)}. \tag{13}$$

Thus, under the attack, the steady state values of signals z_{11}, z_{12}, z_{21} and z_{22} can be calculated as:

$$\begin{aligned}
z_{11,ss,a} &= \frac{\alpha^2}{2}\text{Re}\{G_{1,a}(j\omega)\}\\
&= \frac{\alpha^2 S_0^2 A_a}{2D_2(\omega)}\Big(\text{Re}(\overline{KA}(j\omega))\left[S_0A_0 + S_aA_a\text{Re}(\overline{KAPS}(j\omega))\right] +\\
&\qquad S_aA_a\text{Im}(\overline{KA}(j\omega))\text{Im}(\overline{KAPS}(j\omega))\Big),
\end{aligned} \tag{14}$$

$$\begin{aligned}
z_{12,ss,a} &= \frac{\beta^2}{2}\text{Re}\{G_{1,a}(j2\omega)\}\\
&= \frac{\beta^2 S_0^2 A_a}{2D_2(2\omega)}\Big(\text{Re}(\overline{KA}(j2\omega))\left[S_0A_0 + S_aA_a\text{Re}(\overline{KAPS}(j2\omega))\right]\\
&\qquad + S_aA_a\text{Im}(\overline{KA}(j2\omega))\text{Im}(\overline{KAPS}(j2\omega))\Big),
\end{aligned} \tag{15}$$

$$\begin{aligned}
z_{21,ss,a} &= \frac{\alpha^2}{2}\text{Re}\{G_{2,a}(j\omega)\}\\
&= \frac{\alpha^2 S_0 S_a A_a}{2D_2(\omega)}\Big(S_0A_0\text{Re}(\overline{KAPS}(j\omega)) +\\
&\qquad S_aA_a|\overline{KAPS}(j\omega)|^2\Big),
\end{aligned} \tag{16}$$

$$\begin{aligned}
z_{22,ss,a} &= \frac{\beta^2}{2}\text{Re}\{G_{2,a}(j2\omega)\}\\
&= \frac{\beta^2 S_0 S_a A_a}{2D_2(2\omega)}\Big(S_0A_0\text{Re}(\overline{KAPS}(j2\omega)) +\\
&\qquad S_aA_a|\overline{KAPS}(j2\omega)|^2\Big),
\end{aligned} \tag{17}$$

where

$$\begin{aligned}
D_2(x) &= [S_0A_0 + S_aA_a\text{Re}(\overline{KAPS}(jx))]^2 +\\
&\quad [S_aA_a\text{Im}(\overline{KAPS}(jx))]^2, \qquad x = \omega, 2\omega.
\end{aligned} \tag{18}$$

Clearly, if the above steady state values of z_{ij}'s under the attack, (14)-(17), are different from the nominal ones, (8)-(11), then it is possible to detect the attack. However, if these steady state values are the same, i.e., $z_{ij,ss,a} = z_{ij,ss}$, for all $i, j = 1, 2$, then the attack is *undetectable*. The scenario under which an undetectable attack takes place is characterized by:

Proposition 1: Consider a Type 1 deception attack defined by (2). The attack is undetectable if and only if

- both S_0A_0 and S_aA_a are roots of the following quadratic equation:

$$x^2 + b_1x + b_2 = 0, \tag{19}$$

where

$$\begin{aligned}
b_1 &= \frac{S_0A_0\text{Re}(\overline{KAPS}(j\omega))\text{Re}(\overline{KAPS}(j2\omega))}{\text{Re}(\overline{KAPS}(j\omega)) - \text{Re}(\overline{KAPS}(j2\omega))}\cdot\\
&\qquad \left[\frac{1}{|\overline{KAPS}(j2\omega)|^2} - \frac{1}{|\overline{KAPS}(j\omega)|^2}\right],
\end{aligned} \tag{20}$$

$$\begin{aligned}
b_2 &= \frac{(S_0A_0)^2}{\text{Re}(\overline{KAPS}(j\omega)) - \text{Re}(\overline{KAPS}(j2\omega))}\cdot\\
&\qquad \left[\frac{\text{Re}(\overline{KAPS}(j2\omega))}{|\overline{KAPS}(j2\omega)|^2} - \frac{\text{Re}(\overline{KAPS}(j\omega))}{|\overline{KAPS}(j\omega)|^2}\right],
\end{aligned} \tag{21}$$

- and $F_1(\omega) = F_1(2\omega)$, where

$$\begin{aligned}
F_1(x) &= \frac{A_0\text{Re}(G_1(jx))}{S_0D_3(x)}\cdot\Big([1 + \rho\text{Re}(\overline{KAPS}(jx))] +\\
&\qquad [\rho\text{Im}(\overline{KAPS}(jx))]^2\Big),
\end{aligned} \tag{22}$$

with ρ being the ratio between the two roots of equation (19) and

$$\begin{aligned}
D_3(x) &= \text{Re}(\overline{KA}(jx))\left[1 + \rho\text{Re}(\overline{KAPS}(jx))\right] +\\
&\qquad \rho\text{Im}(\overline{KA}(jx))\text{Im}(\overline{KAPS}(jx)).
\end{aligned}$$

Proof: See the Appendix.

Clearly, to ensure that an attack is detectable, the values of parameters α, β, and ω must be selected to avoid the undetectable scenario. Based on Proposition 1, this can be achieved by choosing the above parameters so that S_0A_0 is not a root of (19).

C. Attack Identification

After an attack is detected, the goal is to identify the attacked DC gains of the sensor and actuator, S_a and A_a. To accomplish this, rewrite equation (16) as

$$c_{11}(S_aA_a)^2 + c_{12}(S_aA_a) + c_{13} = 0, \tag{23}$$

where

$$c_{11} = \left(\frac{2z_{21,ss,a}}{\alpha^2} - S_0\right)|\overline{KAPS}(j\omega)|^2,$$

$$c_{12} = S_0 A_0 \text{Re}(\overline{KAPS}(j\omega)) \left(\frac{4z_{21,ss,a}}{\alpha^2} - S_0 \right),$$
$$c_{13} = S_0^2 A_0^2 \frac{2z_{21,ss,a}}{\alpha^2}.$$

Next, solve (23) for $S_a A_a$ to obtan

$$S_a A_a = \frac{-c_{12} \pm \sqrt{c_{12}^2 - 4c_{11}c_{13}}}{2c_{11}}. \quad (24)$$

Similarly, rewrite (17) as

$$c_{21}(S_a A_a)^2 + c_{22}(S_a A_a) + c_{23} = 0, \quad (25)$$

where

$$c_{21} = \left(\frac{2z_{22,ss,a}}{\beta^2} - S_0 \right) |\overline{KAPS}(j2\omega)|^2,$$
$$c_{22} = S_0 A_0 \text{Re}(\overline{KAPS}(j2\omega)) \left(\frac{4z_{22,ss,a}}{\beta^2} - S_0 \right),$$
$$c_{23} = S_0^2 A_0^2 \frac{2z_{22,ss,a}}{\beta^2},$$

and solve (25) for $S_a A_a$ to get

$$S_a A_a = \frac{-c_{22} \pm \sqrt{c_{22}^2 - 4c_{21}c_{23}}}{2c_{21}}. \quad (26)$$

Clearly, equations (23) and (25) may have either one or both roots in common. Each case is discussed below:

[**Case 1**] If (23) and (25) have exactly one root in common, then this common root is assigned as the value of $S_a A_a$. Next, using (14), we solve for A_a as

$$A_a = \frac{2z_{11,ss,a} D_2(\omega)}{\alpha^2 S_0^2 D_4(\omega)}. \quad (27)$$

where $D_2(\cdot)$ is given in (18) and

$$D_4(x) = \text{Re}(\overline{KA}(jx))(S_0 A_0 + S_a A_a \text{Re}(\overline{KAPS}(jx))) + S_a A_a \text{Im}(\overline{KA}(jx))\text{Im}(\overline{KAPS}(jx)). \quad (28)$$

Note that A_a can be obtained from (15) in the same manner as the above, i.e.,

$$A_a = \frac{2z_{12,ss,a} D_2(2\omega)}{\beta^2 S_0^2 D_4(2\omega)}. \quad (29)$$

Finally, since both $S_a A_a$ and A_a are known, S_a is computed by dividing the former by the latter.

[**Case 2**] When (23) and (25) have two roots in common, and these roots are unique, it implies that two different values of $S_a A_a$ will result in the same $(z_{21,ss,a}, z_{22,ss,a})$. Clearly, $S_a A_a$ cannot be uniquely determined using signals $z_{21,ss,a}$ and $z_{22,ss,a}$ alone, implying that the other signals $z_{11,ss,a}$ and $z_{12,ss,a}$ have to be utilized as well.

Let x_1 and x_2 denote the above mentioned candidate values of $S_a A_a$. Substitute these x_i's, $i = 1, 2$, in the expressions for $z_{11,ss,a}$ and $z_{12,ss,a}$, (14) and (15), and solve for A_a to get

$$A_{a,11}^{(i)} = \frac{2z_{11,ss,a} D_5^{(i)}(j\omega)}{\alpha^2 S_0^2 D_6^{(i)}(j\omega)}, \quad i \in \{1, 2\}, \quad (30)$$

$$A_{a,12}^{(i)} = \frac{2z_{12,ss,a} D_5^{(i)}(j2\omega)}{\beta^2 S_0^2 D_6^{(i)}(j2\omega)}, \quad i \in \{1, 2\}, \quad (31)$$

where

$$D_5^{(i)}(\xi) = [S_0 A_0 + x_i \text{Re}(\overline{KAPS}(j\xi))]^2 + [x_i \text{Im}(\overline{KAPS}(j\xi))]^2,$$
$$D_6^{(i)}(\xi) = \text{Re}(\overline{KA}(j\xi))(S_0 A_0 + x_i \text{Re}(\overline{KAPS}(j\xi))) + x_i \text{Im}(\overline{KA}(j\xi))\text{Im}(\overline{KAPS}(j\xi)), \quad i = 1, 2. \quad (32)$$

Now, consider the following scenarios:

- If only one x_i leads to $A_{a,11}^{(i)} = A_{a,12}^{(i)}$, then assign this x_i as the value of $S_a A_a$ and $A_{a,11}^{(i)}$ as the value of A_a. Finally, the value of S_a can be calculated as before.
- If both x_i's lead to $A_{a,11}^{(i)} = A_{a,12}^{(i)}$, $i = 1, 2$, then the synchronous detection-based method cannot uniquely identify the attacked sensor and actuator DC gains.

To generalize the above scenario, consider two deception attacks, Attack 1 and Attack 2, which take place according to (2), and which lead to the sensor and the actuator DC gains (S_{a1}, A_{a1}) and (S_{a2}, A_{a2}), respectively. As in (14)-(17), let $z_{ij,ss,a}^{(l)}$ be the steady state value of z_{ij} under Attack l, $i, j, l = 1, 2$. Then, Attacks 1 and 2 are said to be *indistinguishable* if $z_{ij,ss,a}^{(1)} = z_{ij,ss,a}^{(2)}$, for all $i, j = 1, 2$. To characterize such attacks, consider:

Proposition 2: Assume that Attacks 1 and 2 take place as described above. These attacks are indistinguishable if and only if

- both $S_{a1}A_{a1}$ and $S_{a2}A_{a2}$ are roots of quadratic equation (19),
- and $F_2(\omega) = F_2(2\omega)$, where

$$F_2(\xi) = \frac{S_0 A_0 \text{Re}(\overline{KA}(j\xi)) + x_2 \text{Re}(\overline{KA}(j\xi)\overline{KAPS}(-j\xi))}{S_0 A_0 \text{Re}(\overline{KA}(j\xi)) + x_1 \text{Re}(\overline{KA}(j\xi)\overline{KAPS}(-j\xi))} \cdot \frac{S_0 A_0 \text{Re}(\overline{KAPS}(j\xi))x_1 + x_1^2 |\overline{KAPS}(j\xi))|^2}{S_0 A_0 \text{Re}(\overline{KAPS}(j\xi))x_2 + x_2^2 |\overline{KAPS}(j\xi))|^2}, \quad (33)$$

and where, as before, x_1 and x_2 are the unique candidate values of $S_a A_a$, which are obtained as the common roots of the quadratic equations (23) and (25).

Proof: See the Appendix.

Thus, this proposition indicates that, to avoid indistinguishable attacks, the parameters α, β, and ω should be selected so that one or both roots of (19) are infeasible (e.g., beyond practical range, being complex), or if this cannot be achieved, then select these parameters to force function $F_2(\xi)$ to take different values at $\xi = \omega$ and $\xi = 2\omega$. Finally, it should be noted that if we only use one sinusoidal signal for attack identification (i.e., when β is 0), then it is all but impossible to avoid indistinguishable attacks caused by the dual roots of (24).

As described above, by using the synchronous detection method with appropriately selected parameters, a deception attack can be detected by monitoring the moving average signals z_{11}, z_{12}, z_{21}, and z_{22} and comparing them with their nominal values. In addition, the DC gains of the sensor and the actuator under attack can be uniquely identified online based on the steady state values of z_{11}, z_{12}, z_{21}, and z_{22}

and equations (24)-(27). This also provides the possibility to mitigate the attack, which is discussed next.

V. Attack mitigation and Timing Issues

As described above, under a Type 1 deception attack on the sensor and the actuator, the steady state values of z_{11}, z_{12}, z_{21} and z_{22} can be used to calculate the values of the attack-modified gains S_a and A_a. These, in turn, can be utilized to compensate for the effects of the attack by modifying the controller from $K(s)$ to $\frac{S_0 A_0}{S_a A_a} K(s)$, and modifying the pre-compensator from S_0 to S_a. This is illustrated in Fig. 5. As

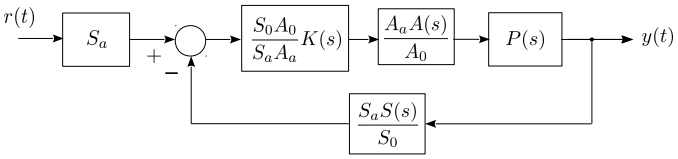


Fig. 5. Mitigation of deception attack

a result, the closed-loop transfer function from the reference signal r to the plant output y is restored to its nominal form (1). In addition, the transfer functions from r to outputs v_i's, $i = 1, 2$, are restored to their corresponding nominal forms (4) and (5) as well. This also renders the moving average signals z_{11}, z_{12}, z_{21} and z_{22} back to their nominal values (8)-(11).

As far as the timing issue is concerned, in some applications, there may exist a "critical" time duration, T_c, beyond which it is undesirable for the plant output to be substantially different from the reference. Obviously, it is necessary that the time required to complete the identification and mitigation procedures be less than T_c. Below, we examine the duration of the former under a deception attack.

Let the transient response of the resilient control system be partitioned into the following three time intervals:

Time interval 1: Time required for the plant output to go close to the new steady state value, after the attack takes place.

Time interval 2: Time required to calculate the new steady state values of z_{11}, z_{12}, z_{21} and z_{22}, after the time interval 1.

Time interval 3: Time required for the plant output to go close to the reference signal, after the identification and mitigation procedures are applied.

To quantify the above time intervals, assume that the nominal closed-loop system is a first-order system with time constant τ_n. In addition, assume that the attack modifies the system's time constant to τ_a. Then, the duration of the first time interval is $3\tau_a$. As for the second time interval, its minimum duration is T, where T is the smallest common period of the sinusoidal oscillations in Fig. 4. Finally, as before, the duration of the third time interval is $3\tau_n$. Thus, the *minimum* time duration, T_{idm}, required to identify and mitigate a deception attack is

$$T_{\text{idm}} = 3\tau_a + T + 3\tau_n. \tag{34}$$

In addition, it should be noted that since the attack is unknown beforehand in practice, we have no knowledge of τ_a. Thus, it may take additional time to be assured that the plant has entered the modified steady state in order to correctly identify the values of the system parameters under attack and apply the mitigation.

VI. Application to Uranium Enrichment Centrifuge Control System

In this section, we apply the deception attack identification and mitigation method developed above to a uranium gas enrichment centrifuge control system. Specifically, a uranium gas enrichment centrifuge typically consists of a three-phase AC induction motor, a controller, and a speed sensor. Consider the three-phase induction motor, whose transfer function from the input voltage to the rotational speed is given by (see [17]):

$$P(s) = \frac{157}{4s + 1}.$$

Assume that this motor is operated in the closed-loop configuration of Fig. 4 with a proportional controller $K(s)$ = 20, a static sensor $S(s) = 1$ and a static actuator $A(s) = 2$. Thus, the nominal closed-loop transfer function is given by

$$G(s) = \frac{1570}{s + 1570.25},$$

which has time constant $\tau_n = 6.368 \times 10^{-4}$sec. The reference value of the system is $r = 528$ rad/s. The parameters of the carrier signals are selected as $\alpha = \beta = 10$ and $\omega = 800$. Note that the amplitudes of the carrier signals are less than 2% of the system reference. In addition, the undetectable attacks and indistinguishable attacks do not exist under these parameters. Using (8)-(11), the nominal steady state values of z_{11}, z_{12}, z_{21} and z_{22} can be computed as

$$z_{11,ss} = 412.4,\ z_{12,ss} = 1018.9,$$
$$z_{21,ss} = 39.7,\ z_{22,ss} = 24.5.$$

Further, assume that an attacker conducts a deception attack on the sensor at $t = 15$ seconds, with the DC gain of the sensor modified to $S_a = 0.5$. Given these data, the closed-loop transfer function under the attack becomes

$$G_a(s) = \frac{1570}{s + 785.25},$$

with time constant $\tau_a = 1.274 \times 10^{-3}$sec. The steady state values of moving average signals z_{11}, z_{12}, z_{21} and z_{22} under the attack can be calculated as

$$z_{11,ss,a} = 1018.9,\ z_{12,ss,a} = 1611.9,$$
$$z_{21,ss,a} = 24.53,\ z_{22,ss,a} = 9.70.$$

The changes in these signals will lead to the detection of the attack. Then, based on the values of $z_{ij,ss,a}$, the attack-modified DC gain of the sensor can be calculated.

The trajectory of the plant output, y, is illustrated in Fig. 6. As seen in this figure, y deviates from the reference signal, r, after the attack takes place at time $t = 15$sec. As described in Section V, the time required by y to reach the new steady state is $3\tau_a = 3.822 \times 10^{-3}$sec. Further, a duration of $T = 7.854 \times 10^{-3}$sec is required to calculate the new steady state values of z_{ij}'s and identify the attack. Thus, the attack is identified during the interval of $3\tau_a + T = 0.0117$sec. In this example, another $T = 7.854 \times 10^{-3}$sec is added to

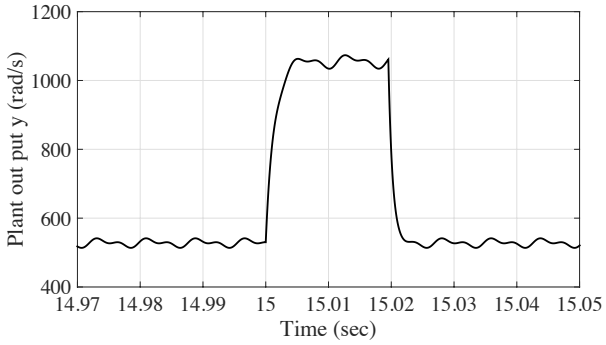


Fig. 6. Trajectory of the plant output, y

confirm that the moving average signals have entered the new steady state. As a result, the application of the mitigation procedure starts at $t = 15.0195$sec, which causes y to begin approaching the reference signal. Finally, after a further $3\tau_n = 0.0019$sec, normal operation of the plant is achieved (at $t = 15.0214$sec). As one can see, the synchronous detection-based method developed in this paper can successfully detect, identify, and mitigate the above attack within a very short period of time. This is of significant importance in protecting critical infrastructure systems agains malicious attacks.

VII. Conclusions and Future Work

In this paper, we developed a technique for actuator and sensor health monitoring in a resilient feedback control system. Specifically, using the method of synchronous detection, we designed a procedure that can detect, identify, and mitigate deception attacks on the DC gains of the sensor and/or the actuator. The procedure involves adding two sinusoidal carrier signals to the reference input and measuring the modulation of the sinusoidal oscillations at the outputs of the actuator and the sensor. Closed-form formulas are derived to calculate the values of the system parameters modified by the attack. The efficacy of the technique is demonstrated through an application to a uranium enrichment centrifuge control system using simulations.

Future work of this research include:

- Investigation of synchronous detection-based techniques using other modulation signals (deterministic and random).
- Design of synchronous detection-based approaches for identification of other types of deception attacks and other classes of attacks (e.g., replay attack, covert attack, denial-of-service attack).
- Development of rigorous methods for resilient control systems design.
- Extension of the methods to systems with MIMO subplants.

Appendix

Proof of Proposition 1: Comparing (10) and (11) with (16) and (17), we can see that if $S_aA_a = S_0A_0$, then $z_{21,ss} = z_{21,ss,a}$ and $z_{22,ss} = z_{22,ss,a}$. In other words, the steady state values of signals z_{21} and z_{22} remain the same before and after the attack occurs in this case. It should be noted, however, that when $S_aA_a = S_0A_0$, we have $z_{11,ss} \neq z_{11,ss,a}$ and $z_{12,ss} \neq z_{12,ss,a}$ unless $A_a = A_0$ and $S_a = S_0$. This ensures that such an attack can be detected. On the other hand, it is also possible that $z_{21,ss} = z_{21,ss,a}$ and $z_{22,ss} = z_{22,ss,a}$ when $S_aA_a \neq S_0A_0$. This is characterized as follows:

Let x_1 and x_2 denote two arbitrary complex numbers and $x_1 \neq x_2$. Let $z^{(i)}_{2k,ss,a}$, $i,k \in \{1,2\}$ denote the steady state value of moving average signal z_{2k} with $S_aA_a = x_i$. Assume that $z^{(1)}_{21,ss,a} = z^{(2)}_{21,ss,a}$ and $z^{(1)}_{22,ss,a} = z^{(2)}_{22,ss,a}$. From (16) and (17), we obtain that

$$\frac{S_0A_0\text{Re}(\overline{KAPS}(j\xi))x_1 + |\overline{KAPS}(j\xi)|^2x_1^2}{(S_0A_0)^2 + 2S_0A_0\text{Re}(\overline{KAPS}(j\xi))x_1 + |\overline{KAPS}(j\xi)|^2x_1^2} = \frac{S_0A_0\text{Re}(\overline{KAPS}(j\xi))x_2 + |\overline{KAPS}(j\xi)|^2x_2^2}{(S_0A_0)^2 + 2S_0A_0\text{Re}(\overline{KAPS}(j\xi))x_2 + |\overline{KAPS}(j\xi)|^2x_2^2}$$

holds for $\xi = \omega$ and $\xi = 2\omega$. Re-organizing the terms in the above equation leads to

$$\begin{aligned} S_0A_0|\overline{KAPS}(j\xi)|^2(x_1+x_2)+ \\ |\overline{KAPS}(j\xi)|^2\text{Re}(\overline{KAPS}(j\xi))x_1x_2+ \\ (S_0A_0)^2\text{Re}(\overline{KAPS}(j\xi)) = 0, \end{aligned}$$

and, thus,

$$x_1 + x_2 = -\frac{\text{Re}(\overline{KAPS}(j\xi))}{S_0A_0}x_1x_2 - \frac{S_0A_0\text{Re}(\overline{KAPS}(j\xi))}{|\overline{KAPS}(j\xi)|^2}. \tag{35}$$

Since (35) holds for both $\xi = \omega$ and $\xi = 2\omega$, i.e.,

$$\frac{\text{Re}(\overline{KAPS}(j\omega))}{S_0A_0}x_1x_2 + \frac{S_0A_0\text{Re}(\overline{KAPS}(j\omega))}{|\overline{KAPS}(j\omega)|^2} = \frac{\text{Re}(\overline{KAPS}(j2\omega))}{S_0A_0}x_1x_2 + \frac{S_0A_0\text{Re}(\overline{KAPS}(j2\omega))}{|\overline{KAPS}(j2\omega)|^2},$$

we have $x_1x_2 = b_2$, where b_2 is given in (21). Then, replacing x_1x_2 in (35) with b_2 for either $\xi = \omega$ or $\xi = 2\omega$ leads to $x_1 + x_2 = -b_1$, where b_1 is given in (20). This implies that x_1 and x_2 are roots of quadratic equation (19).

Therefore, both S_aA_a and S_0A_0 must be roots of (19) to have $z_{21,ss} = z_{21,ss,a}$ and $z_{22,ss} = z_{22,ss,a}$. In this case, to have $z_{11,ss} = z_{11,ss,a}$ and $z_{12,ss} = z_{12,ss,a}$, it follows from (8), (14) and (9), (15) that

$$\frac{\text{Re}(G_1(j\xi))}{S_0} = \frac{A_a}{A_0}\cdot \left(\frac{\text{Re}(\overline{KA}(j\xi))\left[1 + \frac{S_aA_a}{S_0A_0}\text{Re}(\overline{KAPS}(j\xi))\right]}{[1 + \frac{S_aA_a}{S_0A_0}\text{Re}(\overline{KAPS}(j\xi))]^2 + [\frac{S_aA_a}{S_0A_0}\text{Im}(\overline{KAPS}(j\xi))]^2} + \frac{\text{Im}(\overline{KA}(j\xi))\text{Im}(\overline{KAPS}(j\xi))}{[1 + \frac{S_aA_a}{S_0A_0}\text{Re}(\overline{KAPS}(j\xi))]^2 + [\frac{S_aA_a}{S_0A_0}\text{Im}(\overline{KAPS}(j\xi))]^2} \right), \quad \xi = \omega, 2\omega.$$

Since S_aA_a and S_0A_0 are roots of (19), let ρ denote the ratio of the roots. Then,

$$A_a = F_1(\omega) = F_1(2\omega),$$

where function $F_1(\cdot)$ is defined in (22). Therefore, undetectable attack exists if and only if S_aA_a and S_0A_0 are both roots of (19) and $F_1(\omega) = F_1(2\omega)$.

□

Proof of Proposition 2: It follows immediately from the proof of Proposition 1 that both $S_{a1}A_{a1}$ and $S_{a2}A_{a2}$ must be roots of (19) to have $z^{(1)}_{21,ss,a} = z^{(2)}_{21,ss,a}$ and $z^{(1)}_{22,ss,a} = z^{(2)}_{22,ss,a}$. Without loss of generality, let x_1 and x_2 denote the roots of (19) and assume $S_{a1}A_{a1} = x_1$ and $S_{a2}A_{a2} = x_2$. In addition, to have $z^{(1)}_{11,ss,a} = z^{(2)}_{11,ss,a}$ and $z^{(1)}_{12,ss,a} = z^{(2)}_{12,ss,a}$, it follows from (14) and (15) that

$$\begin{aligned}
&\frac{A_{a1}\mathrm{Re}(\overline{KA}(j\xi))\left[1+\mathrm{Re}(\overline{KAPS}(j\xi))x_1\right]}{(S_0A_0)^2+2S_0A_0\mathrm{Re}(\overline{KAPS}(j\xi))x_1+|\overline{KAPS}(j\xi)|^2x_1^2}+\\
&\frac{A_{a1}\mathrm{Im}(\overline{KA}(j\xi))\mathrm{Im}(\overline{KAPS}(j\xi))x_1}{(S_0A_0)^2+2S_0A_0\mathrm{Re}(\overline{KAPS}(j\xi))x_1+|\overline{KAPS}(j\xi)|^2x_1^2}=\\
&\frac{A_{a2}\mathrm{Re}(\overline{KA}(j\xi))\left[1+\mathrm{Re}(\overline{KAPS}(j\xi))x_2\right]}{(S_0A_0)^2+2S_0A_0\mathrm{Re}(\overline{KAPS}(j\xi))x_2+|\overline{KAPS}(j\xi)|^2x_2^2}+\\
&\frac{A_{a2}\mathrm{Im}(\overline{KA}(j\xi))\mathrm{Im}(\overline{KAPS}(j\xi))x_2}{(S_0A_0)^2+2S_0A_0\mathrm{Re}(\overline{KAPS}(j\xi))x_2+|\overline{KAPS}(j\xi)|^2x_2^2},\\
&\xi=\omega,2\omega.
\end{aligned}$$

Thus,

$$\frac{A_{a1}}{A_{a2}} = F_2(\omega) = F_2(2\omega),$$

where function $F_2(\cdot)$ is defined in (33). Therefore, indistinguishable attacks exist if and only if $S_{a1}A_{a1}$ and $S_{a2}A_{a2}$ are both roots of (19) and $F_2(\omega) = F_2(2\omega)$.

□

REFERENCES

[1] A. A. Cardenas, S. Amin, and S. Sastry, "Secure control: Towards survivable cyber-physical systems," in *The 28th International Conference on Distributed Computing Systems Workshops*. IEEE, 2008, pp. 495–500.

[2] Y.-L. Huang, A. A. Cárdenas, S. Amin, Z.-S. Lin, H.-Y. Tsai, and S. Sastry, "Understanding the physical and economic consequences of attacks on control systems," *International Journal of Critical Infrastructure Protection*, vol. 2, no. 3, pp. 73–83, 2009.

[3] M. Krotofil and A. A. Cárdenas, "Resilience of process control systems to cyber-physical attacks," in *Secure IT Systems*. Springer, 2013, pp. 166–182.

[4] S. Sridhar and M. Govindarasu, "Model-based attack detection and mitigation for automatic generation control," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 580–591, 2014.

[5] B. Genge, I. Kiss, and P. Haller, "A system dynamics approach for assessing the impact of cyber attacks on critical infrastructures," *International Journal of Critical Infrastructure Protection*, vol. 10, pp. 3–17, 2015.

[6] A. Cardenas, S. Amin, B. Sinopoli, A. Giani, A. Perrig, and S. Sastry, "Challenges for securing cyber physical systems," in *Workshop on future directions in cyber-physical systems security*, 2009, p. 5.

[7] E. K. Wang, Y. Ye, X. Xu, S.-M. Yiu, L. C. K. Hui, and K.-P. Chow, "Security issues and challenges for cyber physical system," in *Proceedings of the 2010 IEEE/ACM International Conference on Green Computing and Communications & International Conference on Cyber, Physical and Social Computing*. IEEE Computer Society, 2010, pp. 733–738.

[8] A. Teixeira, S. Amin, H. Sandberg, K. H. Johansson, and S. S. Sastry, "Cyber security analysis of state estimators in electric power systems," in *49th IEEE Conference on Decision and Control (CDC)*. IEEE, 2010, pp. 5991–5998.

[9] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security (TISSEC)*, vol. 14, no. 1, p. 13, 2011.

[10] F. Pasqualetti, F. Dorfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2715–2729, 2013.

[11] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Transactions on Automatic Control*, vol. 59, no. 6, pp. 1454–1467, 2014.

[12] A. Teixeira, K. C. Sou, H. Sandberg, and K. H. Johansson, "Secure control systems: A quantitative risk management approach," *IEEE Control Systems*, vol. 35, no. 1, pp. 24–45, 2015.

[13] J.-Y. Keller, K. Chabir, and D. Sauter, "Input reconstruction for networked control systems subject to deception attacks and data losses on control signals," *International Journal of Systems Science*, vol. 47, no. 4, pp. 814–820, 2016.

[14] Y. Mo, R. Chabukswar, and B. Sinopoli, "Detecting integrity attacks on SCADA systems," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 4, pp. 1396–1407, 2014.

[15] Y. Mo, S. Weerakkody, and B. Sinopoli, "Physical authentication of control systems: designing watermarked control inputs to detect counterfeit sensor outputs," *IEEE Control Systems*, vol. 35, no. 1, pp. 93–109, 2015.

[16] A. B. Carlson and P. Crilly, *Communication Systems*, 5th ed. McGraw-Hill, 2009.

[17] S.-J. Lee, "Multiple simultaneous specifications (MSS) control design method of a high-speed AC induction motor," Master's thesis, University of Toronto, 2000.