On the Vulnerabilities Due to Manipulative Zero-Stealthy Attacks in Cyber-Physical Systems

Takumi SHINOHARA * and Toru NAMERIKAWA **

Abstract: In this paper, we analyze the vulnerabilities due to integrity cyber attacks named zero-stealthy attacks in cyber-physical systems, which are modeled as a stochastic linear time invariant (LTI) system equipped with a Kalman filter, an LQG controller, and a $\chi^2$ failure detector. The attacks are designed by a sophisticated attacker so that the measurement residual of the compromised system coincides with the healthy one, and thus it is impossible to detect the attacks. First, we characterize and analyze an existence condition of the attacks from an attacker’s standpoint. Then, we extend the attacks into an attacker’s goal: The scenario when the adversary wishes to manipulate the systems to an objective designed by him/her. Our results provide that the attacker can manipulate the compromised system to the objective without accessing the networks of real-time sensor or actuator data. Finally, we verify the dangerousness of the attacks through a simple numerical example.

Key Words: cyber-physical systems, system security, vulnerability analysis.

1. Introduction

From the viewpoints of increasing the efficiency of social systems, creating new industries, and improving productivity, cyber-physical systems (CPS) which integrate data sensing, actuation, control, communication, and computation via networks have attracted much attention in recent years. Many applications are considered as CPS such as transportation, manufacturing, medical devices, and energy systems [1]–[3]. However, CPS are severely dependent on networking and computing devices, and hence they have great potential security threats and vulnerabilities due to various cyber attacks [4]–[6]. In fact, there were several reports on cyber attacks targeting control systems, which carried significant damage to the physical plants, for instance the Stuxnet incident [7] and the Maroochy water breach [4],[8]. Therefore, comprehending and analyzing vulnerabilities due to malicious cyber attacks in CPS are significant issues to improve the security of them.

The analysis of vulnerabilities of CPS to the malicious cyber attacks has been discussed by many research communities. The general studies focus on specific cyber attacks against particular system, such as replay, covert, zero-dynamics, and false data injection attacks. These attacks are categorized as integrity cyber attacks, which aim at compromising the data integrity, and have recently received considerable attention [9]. Replay attacks, which are carried out by stealing information on a wireless network and imitating a legitimate sender to use the stolen information to falsify correct information, are studied in [10],[11]. In [12], the effect of covert attacks against network control systems is considered. The covert attacker is modeled as knowledgeable of the system and accessible to all sensor and actuator channels. Zero-dynamics attacks, which are injected into both the actuators and sensors networks by an adversary without altering the sensor measurement, are studied in [13]–[15]. In [16], Liu et al. had introduced false data injection attacks against a static state estimator in electric power grid, and Mo et al. [17]–[19] extended these attacks into generally stochastic linear time invariant (LTI) systems.

In this paper, we analyze the special case of false data injection attacks named zero-stealthy attacks, which indicate that the measurement residual of the compromised system coincides with the healthy one. In this situation, any failure detector cannot detect the attacks. Though the prior work [20] has already discussed the zero-stealthy attacks in noiseless systems, the relationship between the number of sensors and actuators were limited. Additionally, more realistic scenarios accounting for a noisy environment should be expected, and hence our paper deals with the attacks in a common stochastic LTI system. We first show an existence condition of the attacks using invariant subspace properties, and the condition shows that the vulnerability due to the attacks exists in all CPS which are modeled as a stochastic LTI system. Previous researches about replay or false data injection attacks discussed a malicious attacker whose goal is to destabilize the system. In a more realistic scenario, however, attackers may have specific goals to gain some benefit, for instance driving the system to a particular state. Therefore, in this paper, we deal with zero-stealthy attacks with a realistic attacker’s goal where the compromised system state is moved to an objective which is designed by the adversary. We name the attacks manipulative zero-stealthy attacks, and a feasible condition of attack objectives and a design procedure of the attacks are addressed.

The rest of this paper is organized as follows. In Section 2, we describe the physical system model and the cyber system model which is equipped with a linear filter, a feedback controller, and a failure detector. In Section 3, we develop the zero-stealthy attack model mathematically, and an existence
condition and a design procedure of the attacks are provided. In Section 4, we consider the case when the adversary wants to manipulate the systems, namely manipulative zero-stealthy attacks. Regarding the attacks, we derive a condition of feasible attack objectives and an attack design. In Section 5, we check the dangerousness of the attacks through illustrating examples, and finally in Section 6, we conclude this work.

Notation and Terminology
We write \( \mathbb{N}_0 \) for \( \{0, 1, 2, \ldots\} \). \( \mathbb{R} \) and \( \mathbb{R}^n \) stand for the set of real numbers and the \( n \)-dimensional Euclidean space, respectively. For a linear map \( A \), \( A^\top \) is the transpose of \( A \). \( \ker A \) and \( \im A \) are the kernel and image of \( A \), respectively. If \( A \) is a linear map from \( X \) to \( Y \) and \( V \) is a subspace of \( Y \), then the inverse image of \( V \) under \( A \) is the subspace of \( X \) denoted by \( A^{-1}V \), that is, \( A^{-1}V = \{x \in X \mid Ax \in V \} \). \( A^{-1} \) is the inverse of \( A \) and \( A^\dagger \) is the Moore-Penrose pseudoinverse of \( A \). \( I \) is the identity matrix of appropriate dimensions. For two subspaces \( V_1 \) and \( V_2 \) of \( \mathbb{R}^n \), we denote \( V_1 + V_2 \) and \( V_1 \times V_2 \) as follows:

\[
\begin{align*}
V_1 + V_2 & = \{v_1 + v_2 : v_1 \in V_1, v_2 \in V_2\}, \\
V_1 \times V_2 & = \{(v_1, v_2) : v_1 \in V_1, v_2 \in V_2\}.
\end{align*}
\]

A symbol \( O \) is used to denote the null subspace. Furthermore, the linear span of a subspace \( V \) generated by vectors \( v_1, \ldots, v_r \in V \) is defined as

\[
\langle v_1, \ldots, v_r \rangle = \{a_1v_1 + \cdots + a_r v_r : a_1, \ldots, a_r \in \mathbb{R}\}.
\]

2. Problem Formulation

2.1 CPS Model
Consider a cyber-physical system which is modeled as a stochastic LTI system equipped with a Kalman filter, an LQG controller, and a \( \chi^2 \) failure detector as depicted in Fig. 1 [10], [11],[18],[19].

As indicated in Fig. 1, we assume that the physical plant is represented as follows:

\[
\begin{align}
x_{k+1} &= Ax_k + Bu_k + w_k, \\ y_k &= Cx_k + v_k,
\end{align}
\]

where \( x_k \in \mathbb{R}^n \), \( u_k \in \mathbb{R}^q \), and \( y_k \in \mathbb{R}^m \) are the physical state vector, the control input, and the measurement from sensors at time \( k \), respectively. \( w_k \sim \mathcal{N}(0, W) \) and \( v_k \sim \mathcal{N}(0, V) \) are i.i.d. Gaussian random variables as process noise and measurement noise, respectively. We assume that the pair \((C, A)\) is detectable and the pairs \((A, B)\) and \((A, W^\top)\) are reachable.

The measurement of the physical plant is transmitted to cyber space via a network, then an optimal input is calculated based on the value. In this paper, we assume that a Kalman filter is used to estimate the true state and an LQG feedback controller is used to stabilize the physical plant. It is well known that there exists a positive semidefinite matrix \( P \), which satisfies \( P_{k+1} = P_{k+1} = P \) with regard to error covariances of Kalman filter if the system is reachable and detectable. Thus, there exists the converged error covariance \( P \) and Kalman gain \( K \):

\[
\begin{align}
K &= P_{k+1}(CP_{k+1}^\top + V)^{-1}, \\
P &= APA^\top + W - APCA^\top(PCA^\top + V)^{-1}CPA^\top.
\end{align}
\]

In this paper, therefore, the state estimation is derived from a steady Kalman filter:

\[
\begin{align}
\hat{x}_{0:k-1} &= \hat{x}_0, \\
\hat{x}_{k:k-1} &= A\hat{x}_{k-1} + Bu_{k-1}, \\
\hat{x}_k &= \hat{x}_{k:k-1} + K(y_k - C\hat{x}_{k:k-1}),
\end{align}
\]

where \( \hat{x}_k \in \mathbb{R}^n \) is the a posteriori minimum mean square error (MMSE) estimate of \( x_k \). For future analysis, we define the measurement residual \( z_k \) and the estimation error \( e_k \) as follows:

\[
\begin{align}
z_k &= y_k - C\hat{x}_{k:k-1}, \\
e_k &= x_k - \hat{x}_k.
\end{align}
\]

The LQG controller is a feedback system which uses a Kalman filter to estimate the state of the system and uses the estimated values to perform a state feedback control. The control input given to the system at time \( k \) is represented by the following:

\[
u_k = L\hat{x}_k,
\]

where \( L \) is the fixed feedback gain. The evaluation function \( J \) of the regulator which performs the feedback control takes the following form:

\[
J = \lim_{T \to \infty} \frac{1}{T} \sum_{k=0}^{T-1} \left( e_k^\top Q e_k + u_k^\top R u_k \right),
\]

where \( \mathbb{E}[-] \) is defined as the expectation of a random variable and \( Q \) and \( R \) are the positive semidefinite and the positive definite matrix, respectively. It is well known that the positive semidefinite solution \( S \) which satisfies the following discrete-time Riccati equation is uniquely determined, and \( L \) which minimizes the evaluation function \( J \) is given as follows:

\[
\begin{align}
S &= A^\top S A + Q - A^\top S B (B^\top S B + R)^{-1} B^\top S A, \\
L &= -(B^\top S B + R)^{-1} B^\top S A.
\end{align}
\]

2.1.1 \( \chi^2 \) failure detector [21]
Failure detectors are widely used in CPS to detect anomalous operations and monitor the system behavior. We assume that “Failure Detector” in Fig. 1 has a \( \chi^2 \) failure detector, which is computed as follows:

\[
\begin{align}
x_k &= x_{k:k-1}^\top (CP_{k:k-1}^\top + V)^{-1} x_k \geq \eta
\end{align}
\]


\[ g_k \triangleq z_k^T \mathcal{P}^{-1} z_k, \quad \mathcal{P} \triangleq CPC^T + V. \]  

(10)

Since \( z_k \) is a Gaussian distribution, \( g_k \) is a \( \chi^2 \) distribution with \( m \) degrees of freedom. The \( \chi^2 \) failure detector compares \( g_k \) with a threshold value:

\[ g_k \overset{\text{H}_1}{\sim} \eta, \]  

(11)

where \( \eta \) is a threshold. If \( g_k \) is greater than the threshold, then the alternative hypothesis \( \mathcal{H}_1 \) is accepted and a failure alarm will be triggered.

### 2.2 Attack Model

Next, we discuss a cyber threat model in the system. In this paper, as indicated below, we assume the worst-case and most powerful attacker who has the perfect knowledge of the system model:

**Assumption 1**

1. The malicious attacker has the knowledge of the system parameters, namely, \( A, B, C, K, L \). Moreover, he/she can leverage these matrices to attack design.

2. The attacker can compromise a subset of both sensors and actuators and can add arbitrary values to change the measurement and input values.

Let us define \( \{i_1, \ldots, i_l\} \subseteq [1, \ldots, m] \) as a subset of compromised sensors. Figure 2 shows the system subject to the integrity attacks.

The modified measurement which the estimator receives is defined as follows:

\[ y_k' = Cx_k' + \Gamma y_k' + v_k, \]  

(12)

where \( \Gamma \) is the sensor selection matrix that the attacker can define as

\[ \Gamma \triangleq [a_{i_1}, \ldots, a_{i_l}] \in \mathbb{R}^{m \times l}, \]  

(13)

where \( a_i \) is the \( i \)th vector of the canonical basis of \( \mathbb{R}^m \). The arbitrary value which is injected by the attacker \( y_k' \) is defined as

\[ y_k' \triangleq [y_{i_1}, \ldots, y_{i_l}]^T \in \mathbb{R}^l, \]  

(14)

where \( y_k' \) indicates the injected value on sensor \( i \) at time \( k \). Similarly, we express the attacked system equation as

\[ x_{k+1}' = Ax_k' + B(u_k' + u_k^0) + w_k, \]  

(15)

where \( u_k^0 \in \mathbb{R}^q \) is the arbitrary value which is designed by the attacker. Without loss of generality, we assume that the injection of arbitrary control inputs starts at \( k = 0 \) and the manipulation of sensor measurements starts at \( k = 1 \), i.e., \( u_k^0 = 0, \forall k \leq -1 \) and \( y_k' = 0, \forall k \leq 0 \). In this paper, we use the superscript ‘\( \Delta \)’ in order to indicate the compromised system value, e.g., \( x_k', y_k', u_k' \).

Furthermore, we define the errors between the compromised and healthy system as follows:

\[ \Delta x_k \triangleq x_k' - x_k, \quad \Delta \xi_k \triangleq \xi_k' - \xi_k, \quad \Delta u_k \triangleq u_k' - u_k, \quad \Delta y_k \triangleq y_k' - y_k, \quad \Delta \eta_k \triangleq \eta_k' - \eta_k, \]  

(16)

One can indicate the difference system model as follows:

\[ \Delta x_{k+1} = (A + BL) \Delta x_k - BL \Delta \xi_k + B u_k^0, \]  

(17)

\[ \Delta z_{k+1} = C \Delta \xi_k + C B u_k^0 + \Gamma \Delta y_{k+1}, \]  

(18)

\[ \Delta \xi_{k+1} = (A - KCA) \Delta \xi_k + (B - KCB) u_k^0 - K T \gamma_{k+1}', \]  

(19)

For the sake of simplicity, we define as

\[ \bar{A} \triangleq A - KCA \in \mathbb{R}^{m \times m}, \]  

\[ \bar{B} \triangleq \begin{bmatrix} B - KCB & -K \Gamma \end{bmatrix} \in \mathbb{R}^{m \times (q + l)}, \]  

\[ \bar{C} \triangleq CA \in \mathbb{R}^{m \times m}, \]  

\[ \bar{D} \triangleq \begin{bmatrix} C B & K \Gamma \end{bmatrix} \in \mathbb{R}^{m \times (q + l)}, \]  

(20)

and the attacker’s arbitrary injection as

\[ \Delta \gamma_{k} \triangleq \left[ u_k^0 \right]_{y_{k+1}'} \in \mathbb{R}^{q + l}. \]  

(21)

Therefore, we can rewrite (19) and (18) as

\[ \Delta \xi_{k+1} = \bar{A} \Delta \xi_k + \bar{B} \Delta \gamma_k, \]  

(22)

\[ \Delta \xi_{k+1} = \bar{C} \Delta \xi_k + \bar{D} \gamma_k. \]  

(23)

We denote the system (22) and (23) as a system \( \Sigma \). In addition, let us define \( \mathcal{A} \triangleq \left[ a_{i_0}^0, a_{i_1}^0, \ldots \right] \) as an infinite attack sequence which is designed by the malicious attacker.

### 3. Zero-Stealthy Attacks

Needless to say, the attacker wishes to inject the arbitrary bias to the system without triggering a failure alarm. The simplest expression of this attack is defined as the following definition.

**Definition 1 (Zero-stealthy attacks)** An attack sequence \( \mathcal{A} \) is zero-stealthy attacks if the following holds:

\[ \Delta \xi_k = 0, \quad \forall k \in \mathbb{N}_0. \]  

(24)

Obviously, if the attacker injects zero-stealthy attacks to the systems, then \( \Delta \xi_k = 0, \forall k \in \mathbb{N}_0 \), that is, it is impossible to detect the attacks. Thus, there is no chance for the failure detector to detect the attacks. First, we devote to formulate an existence condition of the attacks.

---

Fig. 2 System diagram subject to zero-stealthy attacks.
3.1 Existence Condition of Zero-Stealthy Attacks

In order to introduce the existence condition of the zero-stealthy attacks, we utilize the following linear subspace [22].

**Definition 2 (The weakly unobservable subspace)** For the system \( \Sigma \), a point \( \Delta \varepsilon_0 \) is called a weakly unobservable point if there exists a malicious input sequence \( \alpha \) such that the corresponding output function satisfies \( \Delta \varepsilon_k = 0, \forall k \in \mathbb{N}_0 \). The set of all weakly unobservable points of \( \Sigma \) is denoted by \( \mathcal{V}^* \) and is called the weakly unobservable subspace of \( \Sigma \).

A subspace \( \mathcal{V}^* \) is the largest subspace \( \mathcal{V} \) of \( \mathbb{R}^n \) for which the following equation holds [22]:

\[
\begin{bmatrix}
\tilde{A} & \mathcal{V} \subseteq (\mathcal{V} \times \mathcal{O}) + \text{im} \begin{bmatrix} \tilde{B} & \tilde{D} \end{bmatrix}
\end{bmatrix},
\]  

(25)

or equivalently, for some \( \tilde{F} \in \mathbb{R}^{(q+1)\times n} \),

\[
(\tilde{A} + \tilde{B}\tilde{F})\mathcal{V} \subseteq \mathcal{V} \subseteq \ker (\tilde{C} + \tilde{D}\tilde{F}).
\]

(26)

Regarding the invariance of \( \mathcal{V}^* \), the following lemma holds.

**Lemma 1** Let \( \Delta \varepsilon_0 \in \mathcal{V}^* \) and let \( \alpha \) be an input sequence such that corresponding output function satisfies \( \Delta \varepsilon_k = 0, \forall k \in \mathbb{N}_0 \). Then, the associated state satisfies \( \Delta \varepsilon_k \in \mathcal{V}^*, \forall k \in \mathbb{N}_0 \).

**Proof.** This lemma can be proved using [22, Lemma 7.9], and therefore we omit the proof. \( \square \)

From Lemma 1, if \( \mathcal{V}^* = \mathbb{R}^n \), then there exists an attack sequence \( \alpha \) which leads to \( \Delta \varepsilon_k = 0 \), \( \forall k \in \mathbb{N}_0 \) for all \( \Delta \varepsilon_j \in \mathbb{R}^n \). In other words, there exist zero-stealthy attacks if \( \mathcal{V}^* = \mathbb{R}^n \). Thus, in this paper, we explore the condition that \( \mathcal{V}^* = \mathbb{R}^n \) instead of the existence condition of zero-stealthy attacks. In order to introduce the condition, we obtain the following lemma regarding the relationship between the weakly unobservable subspace \( \mathcal{V}^* \) and the unobservable subspace of the system \( \Sigma \).

**Lemma 2** For the system \( \Sigma \),

\[ \mathcal{V}^* = \ker \mathcal{U}, \]

(27)

where

\[ \ker \mathcal{U} = \bigcap_{j=0}^{n-1} \ker \left( \tilde{C} + \tilde{D}\tilde{F} \right) \in \mathbb{R}^n \]

is the unobservable subspace of \( \Sigma \) for some \( \tilde{F} \in \mathbb{R}^{(q+1)\times n} \).

**Proof.** This lemma can be proved using [23, Lemma 1], and therefore we omit the proof. \( \square \)

Now, the existence condition of zero-stealthy attacks can be derived by the following theorem.

**Theorem 1** Suppose that Assumption 1 holds. For the system \( \Sigma \), if \( \tilde{D} \) has full row rank, then

\[ \mathcal{V}^* = \mathbb{R}^n, \]

(28)

which indicates there exists a zero-stealthy attack sequence.

**Proof.** Let \( \tilde{F} = -\tilde{D}^*\tilde{C} \). Then,

\[ \ker \mathcal{U} = \bigcap_{j=0}^{n-1} \ker \left( \tilde{I} - \tilde{D}\tilde{D}^* \right) \tilde{C} \left( \tilde{A} + \tilde{B}\tilde{F} \right) \].

(29)

If \( \tilde{D} \) has full row rank, then \( \tilde{D}\tilde{D}^* = I \). Hence,

\[ \ker \mathcal{U} = \mathbb{R}^n, \]

(30)

namely, according to Lemma 2,

\[ \mathcal{V}^* = \mathbb{R}^n. \]

(31)

The following remark states the vulnerability due to the zero-stealthy attacks in CPS.

**Remark 1** It is worth noticing that \( \tilde{D} \) depends on \( \Gamma \), which is designed by the attacker. Obviously, \( \tilde{D} \) has full row rank if the adversary inserts the arbitrary bias to all sensors, i.e., \( \Gamma = I \). Hence, there exist zero-stealthy attacks in any systems when the adversary attacks to all sensors.

Depending on the system, the adversary can enforce zero-stealthy attacks by attacking only one sensor. In contrast, if \( CB = 0 \), then the adversary has to manipulate all measurements (i.e., \( \Gamma = I \)), and this situation is more difficult as the number of the sensors becomes larger. Thus, \( CB = 0 \) is an ideal situation for the system supervisor, which implies that they ought to design the matrices \( B \) and \( C \) as a minimum countermeasure so that the attacker has to compromise more sensors.

For future analysis, we obtain the following proposition regarding the system \( \Sigma \) subject to zero-stealthy attacks.

**Proposition 1** Suppose that Assumption 1 holds. For the system \( \Sigma \) subject to zero-stealthy attacks, \( \Delta \xi_k = \Delta \xi_k, \forall k \in \mathbb{N}_0 \).

**Proof.** For the system \( \Sigma \), if \( \Delta \varepsilon_k = 0, \forall k \in \mathbb{N}_0 \), then

\[ \Delta \tilde{\xi}_k+1 = (A + BL) \Delta \xi_k, \forall k \in \mathbb{N}_0. \]

Taking the fact that \( \Delta \xi_0 = 0 \) into account, we have \( \Delta \xi_k = 0, \forall k \in \mathbb{N}_0 \) and \( \Delta \xi_k = \Delta \xi_k - \Delta \xi_k = \Delta \xi_k, \forall k \in \mathbb{N}_0 \).

This proposition indicates that \( \Delta \xi_k \) has the dynamics of the system \( \Sigma \) when the attacker asserts the zero-stealthy attacks. Moreover, since

\[ \Delta \xi_{k+1} = \Delta \xi_{k+1} - C\Delta \xi_{k+1}, \forall k \in \mathbb{N}_0, \]

and \( \Delta \xi_k = 0, \forall k \in \mathbb{N}_0 \), one can easily derive that \( \Delta \xi_2 = 0, \forall k \in \mathbb{N}_0 \), which implies that the compromised measurement also coincides with the benign one under the zero-stealthy attacks.

3.2 Design Procedure of Zero-Stealthy Attacks

In the previous section, the existence condition of zero-stealthy attacks is derived. Then, we devote to introduce a design procedure of the attacks. The following lemma provides a necessary and sufficient condition pertaining to a design of an input sequence which leads to zero output.

**Lemma 3** For the system \( \Sigma \), let \( \tilde{F} \in \mathbb{R}^{(q+1)\times n} \) satisfy the following equation:

\[ (\tilde{A} + \tilde{B}\tilde{F})\mathcal{V}^* \subseteq \mathcal{V}^* \subseteq \ker (\tilde{C} + \tilde{D}\tilde{F}). \]

In addition, let \( \tilde{L} \) be a linear map such that

\[ \text{im} \tilde{L} = \ker \tilde{D} \cap \tilde{B}^{-1}\mathcal{V}^*. \]
Then, the output $\Delta z_k$ resulting from a malicious input $\xi_k^\ast$ is zero if and only if $\xi_k^\ast$ takes the following form:

$$\xi_k^\ast = F \Delta x_k + Lv_k^\ast$$  \hspace{1cm} (34)

for some function $v_k^\ast$.

**Proof.** The proof is given in Appendix to improve legibility. \hfill $\square$

Using Lemma 3, one can easily derive the following theorem regarding the design procedure of zero-stealthy attacks.

**Theorem 2** Suppose that Assumption 1 holds and $\tilde{D}$ has full row rank. Then, the following attack sequence is zero-stealthy attacks if and only if

$$\xi_k^\ast = -D^\ast C \Delta x_k + \alpha_k,$$  \hspace{1cm} (35)

where $\alpha_k \in \mathbb{R}^{d_f}$ is some sequence satisfying $\alpha_k \in \ker \tilde{D}$, $\forall k \in \mathbb{N}_0$. Moreover, $\Delta x_k$ subject to the attacks takes the following form:

$$\Delta x_k = \sum_{j=0}^{k-1} (\tilde{A} - BD^\ast C)^{k-1-j} B \alpha_j,$$  \hspace{1cm} (36)

**Proof.** As is the case with Theorem 1, let $\tilde{F} = -D^\ast C$. Then, according to Lemma 3 and Proposition 1, we have

$$\xi_k^\ast = -D^\ast C \Delta x_k + \alpha_k,$$  \hspace{1cm} (37)

where $\alpha_k \in \ker \tilde{D} \cap \tilde{B}^{-1} \mathbb{R}^{d_f}$. Using the result of Theorem 1, we have $\mathbb{V}^\ast = \mathbb{R}^n$ when $\tilde{D}$ has full row rank. Thus, we can rewrite the above equation as

$$\xi_k^\ast = -D^\ast C \Delta x_k + \alpha_k,$$  \hspace{1cm} (38)

where $\alpha_k \in \mathbb{R}^{d_f}$ is some sequence satisfying $\alpha_k \in \ker \tilde{D}$, $\forall k \in \mathbb{N}_0$.

Since this attack sequence leads to $\Delta x_k = 0$, $\forall k \in \mathbb{N}_0$, $\Delta x_k = \Delta x_k$. Remember that $\Delta x_0 = 0$, then we have

$$\Delta x_k = \tilde{A} \Delta x_{k-1} + \tilde{B} \alpha_{k-1} = \tilde{A} \Delta x_{k-1} - BD^\ast C \Delta x_{k-1} + B \alpha_{k-1} = \sum_{j=0}^{k-1} (\tilde{A} - BD^\ast C)^{k-1-j} B \alpha_j,$$  \hspace{1cm} (39)

which concludes the proof. \hfill $\square$

### 4. Manipulative Zero-Stealthy Attacks

This section is devoted to zero-stealthy attacks with an attacker’s objective, namely manipulative zero-stealthy attacks. As mentioned before, attackers would have specific goals to obtain some benefit. From the viewpoint of attackers, they want to retain the compromised state at the objective without triggering a failure alarm. Thus, manipulative zero-stealthy attacks are defined as follows.

**Definition 3 (Manipulative zero-stealthy attacks)** An attack sequence $\mathcal{A}$ is manipulative zero-stealthy attacks if the following equations hold for some $T \in (n, \infty)$:

$$\mathbb{E} \left[ x^\ast_T \right] = x^\ast, \; \forall k \geq T, \; \Delta z_k = 0, \; \forall k \in \mathbb{N}_0, \; (40)$$

where $x^\ast$ is a finite attack objective designed by the malicious attacker.

In the system where we now discuss, the original system state $x_k$ satisfies $\mathbb{E} [x_k] = 0$ in steady state by the LQG controller. Hence, (40) can be rewritten as

$$\Delta x_k = x^\ast, \; \forall k \geq T, \; \Delta z_k = 0, \; \forall k \in \mathbb{N}_0, \; (41)$$

### 4.1 Condition of Feasible Attack Objectives

We proved that the weakly unobservable subspace $\mathbb{V}$ of $\Sigma$ is equivalent to $\mathbb{R}^n$ if $\tilde{D}$ has full row rank. However, the attacker cannot manipulate the system state to all $x^\ast \in \mathbb{R}^n$ since the difference system state $\Delta x_k$ has to remain $\Delta x_k = x^\ast$ in $\forall k \geq T$. Hence, we explore the condition of feasible attack objectives. At first, the feasible attack objectives and their subspace are characterized by the following definition.

**Definition 4 (Feasible objective subspace)** For the system $\Sigma$, if there exist a zero-stealthy attack sequence $\mathcal{A}$ and an objective $x^\ast \in \mathbb{R}^n$ which satisfy (41), then the attack objective is feasible. The set of all feasible attack objectives is denoted by $\mathcal{F}^\ast$ and is called feasible objective subspace.

Regarding the shape of the feasible objective subspace, we obtain the following theorem.

**Theorem 3** Suppose that Assumption 1 holds and $\tilde{D}$ has full row rank. Then, we have

$$\mathcal{F}^\ast = (I - \tilde{A} + \tilde{B} \tilde{D}^\ast C)^{-1} \tilde{B} \ker \tilde{D}.$$  \hspace{1cm} (42)

**Proof.** Consider a feasible attack objective $x^\ast \in \mathcal{F}^\ast$. Then, there exists $\Delta x_T$ such that $\Delta x_T = \Delta x_{T+1} = \cdots = x^\ast$. From (36),

$$x^\ast = (\tilde{A} - \tilde{B} \tilde{D}^\ast C) x^\ast + \tilde{B} \alpha_k,$$

$$\Rightarrow (I - \tilde{A} + \tilde{B} \tilde{D}^\ast C) x^\ast = \tilde{B} \alpha_k.$$  \hspace{1cm} (43)

Thus,

$$x^\ast \in (I - \tilde{A} + \tilde{B} \tilde{D}^\ast C)^{-1} \tilde{B} \ker \tilde{D},$$  \hspace{1cm} (44)

which implies

$$\mathcal{F}^\ast \subseteq (I - \tilde{A} + \tilde{B} \tilde{D}^\ast C)^{-1} \tilde{B} \ker \tilde{D}. \hspace{1cm} (45)$$

On the other hand, recall Theorem 1; we know $\Delta x_k \in \mathbb{R}^n$, $\forall k \in \mathbb{N}_0$ if $\tilde{D}$ has full row rank. Thus, for a time instance $T > n$, there exist a reachable state

$$\Delta x_T \in (I - \tilde{A} + \tilde{B} \tilde{D}^\ast C)^{-1} \tilde{B} \ker \tilde{D} \subseteq \mathbb{R}^n,$$

and a zero-stealthy attack sequence $\mathcal{A}$. Thus, using some input $\alpha \in \ker \tilde{D}$, we can formulate $\Delta x_T$ as

$$\Delta x_T = (\tilde{A} - \tilde{B} \tilde{D}^\ast C) \Delta x_T + \tilde{B} \alpha.$$  \hspace{1cm} (46)

Here, for $t \in \mathbb{N}_0$, we obtain

$$\Delta x_{T+t} = (\tilde{A} - \tilde{B} \tilde{D}^\ast C) \left( \Delta x_T + \sum_{j=0}^{t} (\tilde{A} - \tilde{B} \tilde{D}^\ast C)^j \tilde{B} \alpha \right)$$

$$= \Delta x_T,$$  \hspace{1cm} (47)

which indicates that $\Delta x_T$ is a feasible objective. Therefore, $\Delta x_T \in \mathcal{F}^\ast$ and
\[(I - \breve{A} + \breve{B}D^t\breve{C})^{-1} \breve{B} \ker \breve{D} \subset \mathcal{F}^*, \]  

which concludes this proof. □

From this theorem, the adversary knows whether an attack objective is feasible or not in advance. On the other hand, from the viewpoint of the system supervisors, they can get the knowledge about which kind of state is subject to manipulative zero-stealthy attacks.

### 4.2 Design Procedure of Manipulative Zero-Stealthy Attacks

In this subsection, we show a design procedure of manipulative zero-stealthy attacks. Hereinafter, suppose that \( \breve{D} \) has full row rank and \( x^* \in \mathcal{F}^* \).

#### 4.2.1 Attack Design in \([0, T - 1]\)

At first, we consider an attack design to converge to the objective \( x^* \) in the interval \([0, T - 1]\). Let us define a bounded attack vector \( x^* \), the extended controllability matrix \( \Phi_T \), and the input-output matrix \( \Psi_T \) of the system \( \Sigma \) as follows:

\[
\Phi_T \doteq \begin{bmatrix} (\xi_0^T) & (\xi_1^T) & \cdots & (\xi_{T-1}^T) \end{bmatrix}^T \in \mathbb{R}^{T(q + h)},
\]

\[
\Psi_T \doteq \begin{bmatrix} \breve{A} & 0 & 0 & \cdots & 0 \\ \breve{C} & \breve{B} & \breve{D} & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \breve{C}\breve{A}^{-1} & \breve{C}\breve{A}^{-2} & \cdots & \breve{C} & \breve{D} \end{bmatrix} \in \mathbb{R}^{T \times (T(q + h))}.
\]

The attacker wishes to achieve \( \Delta x_T = x^* \), i.e., \( \Phi_T x_T = x^* \), under the zero-stealthy constraints \( \Delta x_k = 0, \ k \in [0, T - 1] \), namely \( \Psi_T x_T = 0 \). Hence, the adversary calculates the attack vector \( x_T \) satisfying the following simultaneous equation:

\[
\begin{bmatrix} \Phi_T \\ \Psi_T \end{bmatrix} \begin{bmatrix} x_T \\ y_T \end{bmatrix} = \begin{bmatrix} x^* \\ 0 \end{bmatrix} .
\]  

(49)

Since \( \breve{D} \) has full row rank, \( \Psi_T \) also has full row rank. Moreover, we assume that the original system (1) is reachable, and thus the system \( \Sigma \) is also reachable, which implies \( \Phi_T \) has full rank. Consequently, \( \Xi_T \) has full row rank and the above simultaneous equation is an underdetermined system which has many choices of \( x_T \). One particular solution is

\[
\Xi_T = \Xi_T^k y = \Xi_T^k (\Xi_T\Xi_T^k)^{-1} y ,
\]  

which is the solution of (49) that minimizes \( \|x_T\| \) [24].

#### 4.2.2 Attack Design in \([T, \infty)\)

Then, we devote to introduce the attack procedure after \( x_T \) converges to \( x^* \), that is the interval \([T, \infty)\). Since \( \Delta x_T \) has to remain \( x^* \) during this interval, malicious input \( \xi_T \) has to satisfy

\[
x^* = \breve{A} x^* + \breve{B} \xi_T \Rightarrow \breve{B} \xi_T = (I - \breve{A}) x^*, \ k \geq T.
\]

Moreover, for zero-stealthiness, \( \xi_T \) also needs to satisfy the equation \( \breve{C} x^* + \breve{D} \xi_T = 0 \). Thus, the adversary solves the following simultaneous equation:

\[
\begin{bmatrix} B \\ D \end{bmatrix} \xi_T = \begin{bmatrix} (I - \breve{A}) x^* \\ -\breve{C} x^* \end{bmatrix} , \ k \geq T.
\]

(51)

In order to solve this equation, the following lemma is provided.

**Lemma 4** Consider the equation

\[
Ax_k = b_k, \ \forall k \in \mathbb{N}_0,
\]

where \( A \in \mathbb{R}^{m \times n} \) and suppose that there exists at least one solution. Then, any solution of (52) can be expressed as

\[
x_k = A^* b_k + x^*_k,
\]

where \( x^*_k \) is a solution of the equation \( Ax^*_k = 0 \).

**Proof.** The proof can be found in [25, Lemma 3.3]. □

Using this lemma, one can easily derive the malicious sequence in the interval \([T, \infty)\) as

\[
\xi_T = \Theta^* \omega + \theta, \ \theta \in \ker \Theta, \ k \geq T.
\]

(54)

We summarize that the adversary can calculate the attack sequence using (50) and (54) in advance. In other words, he/she does not need the information about the real-time sensor or actuator data for this attack.

### 5. Numerical Example

We illustrate the vulnerabilities due to manipulative zero-stealthy attacks with a simple example of a remote vehicle control borrowed from [18],[19]. The vehicle is moving along one-dimensional line defined as x-axis, and the state space includes its position \( p^* \) and velocity \( v^* \) of the vehicle. As a consequence, the discrete-time system dynamics are as follows:

\[
\begin{bmatrix} v_{k+1}^* \\ p_{k+1}^* \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v_k^* \\ p_k^* \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} w_{k,1},
\]

(55)

\[
p_{k+1}^* = p_k^* + v_k^* + w_{k,2},
\]

(56)

where \( w_{k,1} \) and \( w_{k,2} \) are process noises, which can be written as

\[
x_{k+1} = Ax_k + Bu_k + w_k,
\]

(57)

where

\[
x_k = \begin{bmatrix} v_k^* \\ p_k^* \end{bmatrix} , \ A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} , \ B = \begin{bmatrix} 1 \\ 0 \end{bmatrix} , \ w_k = \begin{bmatrix} w_{k,1} \\ w_{k,2} \end{bmatrix}.
\]

Suppose that two sensors are measurable about its velocity and position, respectively. Then, the output equation is shown as

\[
y_k = x_k + v_k,
\]

(58)

where \( v_k \) is a measurement noise. Moreover, we assume that the covariance matrices of the noise are \( Q = R = I \) and the weight matrices of the LQG costs are \( W = I \) and \( U = 1 \). Therefore, the steady state Kalman gain and the LQG gain are given by

\[
K = \begin{bmatrix} 0.5939 & 0.0793 \\ 0.0793 & 0.6944 \end{bmatrix} , \ L = \begin{bmatrix} -1.2439 \\ -0.4221 \end{bmatrix}.
\]

In this case, \( \breve{D} \) has full row rank only if all sensors are attacked. Hence, we consider a malicious adversary who compromises two sensors, i.e., \( \Gamma = I \). One can easily derive that

\[
F^* = \begin{bmatrix} 0 \\ 1 \end{bmatrix},
\]

(59)

which implies that the attacker can only compromise the vehicle position to the desired value without triggering a failure alarm. In this example, we define \( x^* = [0, 10]^T \) and \( T = 10 \).
Similarly, the state goes to the objective without triggering a failure alarm. The measurement residual between the compromised and healthy system remains this value in the steady state to the objective which achieves to manipulate the compromised and healthy system.

In manipulative zero-stealthy attacks, we can confirm the vulnerabilities due to manipulative zero-stealthy attacks, and an existence subspace of feasible attack objectives in CPS. Then, we tackled to formulate manipulative zero-stealthy attacks as a stochastic LTI system. First, we introduced an existence condition and a design procedure of zero-stealthy attacks against state estimation in wireless sensor networks, Proc. IEEE, Vol. 100, No. 1, pp. 210–224, 2012.

Figure 3 shows the attack sequence calculated by (50) and (54), and Fig. 4 shows the difference system between the compromised and healthy system $\Delta x_k$ and $\Delta z_k$. Obviously, the system state goes to the objective $x^* = [0, 10]'$ at $k = 10$ and remains this value in $k \geq T$. However, the difference of measurement residual between the compromised and healthy systems does not vary by the attack, i.e., $\Delta x_k = 0$, $\forall k \in \mathbb{N}_0$. Similarly, $\chi^2$ detector $g_k$ in the attacked system which is depicted in Fig. 5 coincides with the healthy one. From these results, we can confirm the vulnerabilities due to manipulative zero-stealthy attacks which achieve to manipulate the compromised state to the objective without triggering a failure alarm.

6. Conclusion

In this paper, we have discussed the vulnerabilities due to manipulative zero-stealthy attacks in CPS, which are modeled as a stochastic LTI system. First, we introduced an existence condition and a design procedure of zero-stealthy attacks using invariant subspace properties. The existence condition implies that the vulnerabilities of zero-stealthy attacks exist in all CPS. Then, we tackled to formulate manipulative zero-stealthy attacks, and an existence subspace of feasible attack objectives and a design procedure are derived. For a sophisticated attacker who has the knowledge of the system, he/she knows whether an attack objective is feasible or not and calculates the attack sequence in advance. Note that the adversary does not need to access the real-time networks to get sensor or actuator data. Finally, we checked the dangerousness of the attack through an illustrative example. Future work will concentrate on the secure/resilient estimation of CPS under malicious attacks.

Acknowledgments

This work was supported by JST CREST Grant Number JP-MJCR15K2, Japan.

References

[1] K.D. Kim and P.R. Kumar: Cyber-physical systems: A perspective at the centennial, Proc. IEEE, Vol. 100, Special Centennial Issue, pp. 1287–1308, 2012.
[2] J. Sztipanovits, X. Koutsoukos, G. Karsai, N. Kottenstette, P. Antsaklis, V. Gupta, B. Goodwine, J. Baras, and S. Wang: Toward a science of cyber-physical system integration, Proc. IEEE, Vol. 100, No. 1, pp. 29–44, 2012.
[3] Y. Mo, T.H.-J. Kim, K. Brancik, D. Dickinson, H. Lee, A. Perrig, and B. Sinopoli: Cyber-physical security of a smart grid infrastructure, Proc. IEEE, Vol. 100, No. 1, pp. 195–209, 2012.
[4] F. Pasqualetti, F. Dörfler, and F. Bullo: Control-theoretic methods for cyberphysical security: Geometric principles for optimal cross-layer resilient control systems, IEEE Control Systems Magazine, Vol. 35, No. 1, pp. 110–127, 2015.
[5] S. Sridhar, A. Hahn, and M. Govindarasu: Cyber-physical system security for the electric power grid, Proc. IEEE, Vol. 100, No. 1, pp. 210–224, 2012.
[6] A. Banerjee, K.K. Venkatasubramanian, T. Mukherjee, and S.K.S. Gupta: Ensuring safety, security, and sustainability of mission-critical cyber-physical systems, Proc. IEEE, Vol. 100, No. 1, pp. 283–299, 2012.
[7] D.P. Eidler: Was Stuxnet an act of war? Decoding a cyber-attack, IEEE Security Privacy, Vol. 9, No. 4, pp. 56–59, 2011.
[8] J. Slay and M. Miller: Lesson learned from the Maroochy water branch, Critical Infrastructure Protection, Vol. 253, pp. 73–82, 2007.
[9] A. Teixeira, K.C. Sou, H. Sandberg, and K.H. Johansson: Secure control systems: A quantitative risk management approach, IEEE Control Systems Magazine, Vol. 35, No. 1, pp. 34–45, 2015.
[10] Y. Mo, W. Sean, and B. Sinopoli: Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs, IEEE Control Systems Magazine, Vol. 35, No. 1, pp. 93–109, 2015.
[11] Y. Mo, R. Chabukswar, and B. Sinopoli: Detecting integrity attacks on SCADA systems, IEEE Trans. Control Systems Technology, Vol. 22, No. 4, pp. 1396–1407, 2014.
[12] R.H. Smith: Covert misappropriation of networked control systems: Presenting a feedback structure, IEEE Control Systems Magazine, Vol. 35, No. 1, pp. 82–92, 2015.
[13] F. Pasqualetti, F. Dörfler, and F. Bullo: Attack detection and identification in cyber-physical systems, IEEE Trans. Automatic Control, Vol. 58, No. 11, pp. 2715–2729, 2013.
[14] A. Teixeira, I. Shames, H. Sandberg, and K.H. Johansson: Revealing stealthy attacks in control systems, Proc. of 50th Annual Allerton Conference, pp. 1806–1813, 2012.
[15] A. Teixeira, I. Shames, H. Sandberg, and K.H. Johansson: A secure control framework for resource-limited adversaries, Automatica, Vol. 51, pp. 135–148, 2015.
[16] Y. Liu, M. Reiter, and P. Ning: False data injection attacks against state estimation in electric power grids, Proc. 49th IEEE Conference on Decision and Control, pp. 5967–5972, 2010.
[17] Y. Mo and B. Sinopoli: Integrity attacks on cyber-physical systems, Proc. 1st International Conference on High Confidence Networked System, pp. 47–54, 2012.
[18] Y. Mo and B. Sinopoli: On the performance degradation of cyber-physical systems under stealthy integrity attacks, IEEE
Due to Lemma 1, we can derive
\[
\Delta x_k+1 = (\bar{A} + \bar{B}F)\Delta x_k + \bar{B}a_k,
\]
\[
\Delta z_k+1 = (\bar{C} + \bar{D}F)\Delta x_k + \bar{D}a_k = 0.
\]
(A.1)

Since $(\bar{A} + \bar{B}F)\Delta x_k$ and $\Delta x_k+1$ lie in $\mathcal{V}^*$ for all $k$, we can find that $B_0a_k \in \mathcal{V}^*$. Similarly, since $\Delta x_k \in \ker(\bar{C} + \bar{D}F)$, we can find that $a_k \in \ker \bar{D}$. Thus, $a_k \in \ker \bar{D} \cap \bar{B}^{-1}\mathcal{V}^*$, which indicates that there exist a function $w_k^\ast$ such that $a_k \doteq \bar{L}w_k^\ast$, $\forall k \in \mathbb{N}_0$, and hence $z_k^\ast = F\Delta x_k + \bar{L}w_k^\ast$.

On the other hand, for the necessity, assume that $z_k^\ast = F\Delta x_k + \bar{L}w_k^\ast$. Then, we have
\[
\Delta x_k = \bar{A}\Delta x_{k-1} + \bar{B}z_{k-1}^\ast
= (\bar{A} + \bar{B}F)\Delta x_{k-1} + \bar{B}\bar{L}w_{k-1}^\ast
= \sum_{j=0}^{k-1} (\bar{A} + \bar{B}F)^{k-1-j} \bar{B}\bar{L}w_j^\ast,
\]
(A.2)

where we use the fact that $\Delta x_0 = 0$. Since $(\bar{A} + \bar{B}F)\mathcal{V}^* \subseteq \mathcal{V}^*$ and $\operatorname{im} \bar{L}$ is a subspace of $\mathcal{V}^*$, we must have $\Delta x_k \in \mathcal{V}^*$, $\forall k \in \mathbb{N}_0$. Then, the resulting output is equal to
\[
\Delta z_k+1 = \bar{C}\Delta x_k + \bar{D}z_k^\ast
= (\bar{C} + \bar{D}F)\Delta x_k + \bar{D}\bar{L}w_k^\ast.
\]
(A.3)

Since $\Delta x_k \subseteq \ker(\bar{C} + \bar{D}F)$, the first term on the right-hand side of (A.3) is 0. Additionally, from the definition of $\bar{L}$, we see that $\operatorname{im} \bar{L}$ is a subspace of $\ker \bar{D}$. Thus, the second term on the right-hand side of (A.3) is also 0, and as a consequence, we obtain $\Delta z_k = 0$, $\forall k \in \mathbb{N}_0$, which concludes the proof. \hfill \square