Preprints of the 20th World Congress
The International Federation of Automatic Control
Toulouse, France, July 9-14, 2017

Consequence Analysis of Innovation-based Integrity Attacks with Side Information on Remote State Estimation ⋆

Ziyang Guo * **Dawei Shi** ** **Karl Henrik Johansson** ***
Ling Shi *

* *Electronic and Computer Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong (e-mail: zguoae@ust.hk, eesling@ust.hk).*
** *Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA (e-mail: dawei.shi@outlook.com).*
*** *ACCESS Linnaeus Centre, School of Electrical Engineering, Royal Institute of Technology, Stockholm, Sweden (e-mail: kallej@kth.se).*

Abstract: In this work, we study the worst-case consequence of innovation-based integrity attacks with side information in a remote state estimation scenario. A new type of linear attack strategy based on both intercepted and sensing data is proposed and a corresponding stealthiness constraint is characterized. The evolution of the remote estimation error covariance is derived in the presence of the proposed malicious attack, based on which the worst-case attack policy is obtained in closed form. Furthermore, the system estimation performance under the proposed attack is compared with that under the existing attack strategy to determine which attack is more critical in deteriorating system functionality. Simulation examples are provided to illustrate the developed results.

Keywords: Cyber-Physical System Security, Remote State Estimation, Integrity Attack

1. INTRODUCTION

The widespread implementation of cyber-physical systems (CPS) in critical infrastructures ranging from national power grids to manufacturing processes has reinforced the safety and security requirements in the control system design. Due to the interconnection between different components and technologies, CPS are vulnerable to cyber threats which may cause severe consequences on national economy, social security or even loss of human lives (Kim and Kumar, 2012; Mo et al., 2015). Hence, security is of fundamental importance to ensure the safe operation of CPS and has attracted considerable interest from both academic and industrial communities.

The cyber-physical attack space can be divided according to adversary's system knowledge, disclosure resources and disruption resources (Sandberg et al., 2015). False-data injection attacks, a particular type of integrity attack, were initially proposed for electric power grids on measurement data in Liu et al. (2011). The consequence of false-data injection attacks on remote state estimation was investigated in Mo et al. (2010) and a quantitative measure of system resilience to such an attack was proposed. The explicit trade-off between attack stealthiness and system performance degradation was analyzed for control signal injection attack in Kung et al. (2016). Furthermore, the false-data injection attacks on system state dynamics and secure estimation problems were investigated in Shi et al. (2016a,b). Replay attack, which degrades system performance by recording and replaying the sensor data without the knowledge of system parameters, were studied in Mo and Sinopoli (2009), Mo et al. (2014), Miao et al. (2013). Specifically, the feasibility conditions and countermeasures were considered for LQG control systems in Mo and Sinopoli (2009) and Mo et al. (2014), while the attack detection problem was investigated under a stochastic game framework in Miao et al. (2013). Denial-of-Service (DoS) attack attempts to block the communication channel and prevent legitimate access between system components. Since jamming is a power-intensive activity and the available power of a jammer might be limited, DoS models were studied for resource-constrained attackers (Gupta et al., 2010; Zhang et al., 2015). Besides the aforementioned works which only focus on one side, i.e., either the attacker or the defender, game-theoretic approaches were proposed to investigate the optimal transmission scheduling and power scheduling problems taking both sides into consideration (Li et al., 2015, 2016).

An innovation-based linear integrity attack, which is designed based on the intercepted innovation sequence from being noticed by the χ^2 false-data detector, was proposed in Guo et al. (2016). The evolution of the estimation error covariance and the worst-case attack strategy were

⋆ The work by Z. Guo and L. Shi is supported by an HKUST KTH Partnership FP804. The work by D. Shi is supported by Natural Science Foundation of China (61503027). The work by K. H. Johansson is supported by the Knut and Alice Wallenberg Foundation and the Swedish Research Council.

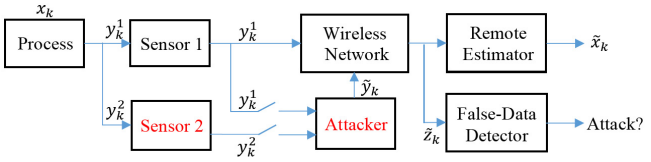

Fig. 1. System Architecture: The sensor transmits the measurement to the remote estimator through a wireless communication network. The attacker is able to modify the transmitted data based on both the intercepted and the sensing information without being detected.

obtained. In this work, we consider the scenario where the malicious agent is able to take an extra measurement of the system state besides the previous intercepted measurement. In this case, the linear attack strategy can be designed based on both the intercepted and sensing data, which is different from Guo et al. (2016). For the proposed linear attack policy, the remote estimation error covariance is derived and the worst-case strategy is obtained in closed form. It is further proved that the attack consequence using both the intercepted and sensing data is more severe than that only using the intercepted data (Guo et al., 2016) when the system is stable.

The remainder of the paper is organized as follows. Section II introduces the system architecture. Section III presents the innovation-based linear attack strategy and the stealthiness constraint. Section IV derives iteration of the remote estimation error covariance. Section V obtains the worst-case attack in closed form and compares with the existing result. Numerical examples are provided in Section VI. Some concluding remarks are given in the end.

2. SYSTEM ARCHITECTURE

2.1 Process Model

As shown in Fig. 1, we consider a first-order discrete-time linear time-invariant (LTI) process described by

$$x_{k+1} = Ax_k + w_k, \tag{1}$$
$$y_k^1 = C_1 x_k + v_k^1, \tag{2}$$

where $A, C \in \mathbb{R}$, $C_1 \neq 0$, $k \in \mathbb{N}$ is the time index, $x_k \in \mathbb{R}$ is the system state, $y_k^1 \in \mathbb{R}$ is the sensor measurement, $w_k \in \mathbb{R}$ and $v_k^1 \in \mathbb{R}$ are zero-mean i.i.d. Gaussian noises with covariances $Q \geq 0$ and $R_1 > 0$, respectively. The initial state x_0 is zero-mean Gaussian with covariance $\Pi_0 \geq 0$ and independent of w_k and v_k^1 for all $k \geq 0$.

2.2 Remote Estimator

At each time instant, the sensor sends its measurement to a remote estimator through a wireless communication network. To estimate the system state, the following standard Kalman filter is adopted by the remote estimator:

$$\hat{x}_k^{1-} = A\hat{x}_{k-1}^1, \tag{3}$$
$$P_k^{1-} = A^2 P_{k-1}^1 + Q, \tag{4}$$
$$K_k^1 = \frac{C_1 P_k^{1-}}{C_1^2 P_k^{1-} + R_1}, \tag{5}$$
$$\hat{x}_k^1 = \hat{x}_k^{1-} + K_k^1 (y_k^1 - C_1 \hat{x}_k^{1-}), \tag{6}$$
$$P_k^1 = P_k^{1-} - K_k^1 C_1 P_k^{1-}, \tag{7}$$

where $\hat{x}_k^{1-}$ and $\hat{x}_k^1$ are the *a priori* and the *a posteriori* minimum mean squared error (MMSE) estimates of the state x_k, respectively, and P_k^{1-} and P_k^1 the corresponding error covariances.

It is well known that the Kalman filter converges from any initial condition exponentially fast (Anderson and Moore, 2012). Thus, we define the steady-state value as

$$P_1 \triangleq \lim_{k \to +\infty} P_k^{1-}, \tag{8}$$
$$K_1 \triangleq \frac{C_1 P_1}{C_1^2 P_1 + R_1}, \tag{9}$$

where P_1 is the unique positive semi-definite solution of the Riccati equation $P_1 = A^2 P_1 + Q - A^2 C_1^2 P_1^2 / (C_1^2 P_1 + R_1)$. For the ease of presentation, we assume that the system starts from the steady state, i.e., $\Pi_0 = P_1$, which results in a fixed-gain estimator with $K_k = K_1, \ \forall k$.

2.3 False-Data Detector

A false-data detector is equipped at the remote side to monitor system behavior and detect the existence of potential cyber attacks. According to Anderson and Moore (2012), the innovation sequence $z_k^1 = y_k^1 - C_1 \hat{x}_k^{1-}$ has a steady-state Gaussian distribution $\mathcal{N}(0, M)$ with $M = C_1^2 P_1 + R_1$ and $\mathbb{E}[z_i^1 z_j^1] = 0$ for all $i \neq j$. Hence, its statistical characteristics (mean and covariance) are used to diagnose the system anomalies.

The χ^2 false-data detector is widely used for fault detection in practice (Mason and Young, 2002; Pouliezos and Stavrakakis, 2013; Mo and Sinopoli, 2009; Mo et al., 2014; Miao et al., 2013). It makes a decision based on the sum of the normalized innovation sequence, i.e., at time k, the detection criterion follows the hypothesis testing:

$$g_k = \sum_{i=k-J+1}^{k} z_i^1 M^{-1} z_i^1 \underset{H_1}{\overset{H_0}{\lessgtr}} \delta, \tag{10}$$

where J is the window size of detection, δ is the threshold, the null hypotheses H_0 means that the system is operating normally, while the alternative hypotheses H_1 means that the system is under attack. The normalized sum in (10) satisfies the χ^2 distribution with mJ degrees of freedom. Thus, it is easy to calculate the false alarm rate from the χ^2 distribution. If g_k is greater than the threshold, an alarm will be triggered.

3. ATTACK STRATEGY AND STEALTHINESS CONSTRAINT

In this section, we consider a malicious agent who intentionally launches attacks to degrade the system estimation performance. The attacker is not only able to intercept the transmitted data packet, but also has an extra private sensor to measure the system state itself. In this case, we introduce the innovation-based attack policies and analyze the stealthiness constraints needed from being detected.

3.1 Linear Attack Strategy

Similar to the attack models in existing works (Mo et al., 2015; Smith, 2015; Callegati et al., 2009), we assume that the attacker has full knowledge of the process model and

is capable of intercepting and modifying the transmitted measurements. It is worth noticing that the attacker can work equivalently with the measurement and the innovation under above assumptions. Specifically, based on the system knowledge, the attacker is able to implement a filter to first calculate the innovation z_k according to $z_k = y_k - C\hat{x}_k^-$ with $\hat{x}_k^- = \mathbb{E}[x_k|y_{1:k-1}]$, then generate the compromised innovation $\tilde{z}_k$, and finally go back to the manipulated measurement $\tilde{y}_k$ according to $\tilde{y}_k = \tilde{z}_k + C\tilde{x}_k^-$ with $\tilde{x}_k^- = \mathbb{E}[x_k|\tilde{y}_{1:k-1}]$. This procedure $y_k \to z_k \to \tilde{z}_k \to \tilde{y}_k$ means that generating $\tilde{y}_k$ is equivalent to generating $\tilde{z}_k$. To simplify the subsequent discussion, we design the attack strategy in terms of z_k.

When the malicious attacker is only capable of intercepting the transmitted data, the attack strategy is based on the system innovation z_k^1. Recall the linear attack strategy studied in Guo et al. (2016), which is defined as

$$\tilde{z}_k^1 = T_k^1 z_k^1 + b_k^1, \tag{11}$$

where $z_k^1 \in \mathbb{R}$ is the currently intercepted innovation, $\tilde{z}_k^1 \in \mathbb{R}$ is the innovation modified by the attacker, $T_k^1 \in \mathbb{R}$ is an arbitrary number, and $b_k^1 \in \mathbb{R}$ is a zero-mean i.i.d. Gaussian random variable with covariance L_k^1 and independent of z_k^1. In this case, $\tilde{z}_k^1$ is zero-mean Gaussian distributed with covariance $T_k^1(C_1^2 P_1 + R_1)T_k^1 + L_k^1$.

However, if the malicious attacker is able to measure the system state itself according to

$$y_k^2 = C_2 x_k + v_k^2,$$

with $C_2 \neq 0$ and $v_k^2 \in \mathbb{R}$ being a zero-mean i.i.d. Gaussian noise with covariance $R_2 > 0$, the attack strategy can be designed based on the intercepted data and the sensing data together, i.e.,

$$\tilde{z}_k^3 = T_k^3 z_k^3 + b_k^3, \tag{12}$$

where $z_k^3 = y_k - C\hat{x}_k^{3-} = C(x_k - \hat{x}_k^{3-}) + v_k \in \mathbb{R}^2$ is the innovation calculated by the malicious attacker using a Kalman filter with $y_k = [y_k^1,\ y_k^2]' \in \mathbb{R}^2$, $C = [C_1,\ C_2]' \in \mathbb{R}^2$, and $v_k = [v_k^1,\ v_k^2]' \in \mathbb{R}^2$ being i.i.d. Gaussian noise with covariance $R = \mathrm{Diag}\{R_1,\ R_2\}$, $\tilde{z}_k^3 \in \mathbb{R}^2$ is the compromised innovation, $T_k^3 = [T_k^{31},\ T_k^{32}] \in \mathbb{R}^{1\times 2}$ is an arbitrary attack vector, and $b_k^3 \in \mathbb{R}^2$ is zero-mean Gaussian distributed with covariance L_k^3 and independent of z_k^3. It can be observed that $\tilde{z}_k^3$ is Gaussian distributed with zero mean and covariance $T_k^3(CP_3C' + R)T_k^{3\prime} + L_k^3$.

3.2 Stealthiness Constraint

For the aforementioned two types of linear attack strategies, the goal of the malicious attacker is to degrade the system estimation performance as much as possible and simultaneously remain stealthy to the false-data detector. According to the detection criterion (10), if the modified innovation sequence $\tilde{z}_k^i = T_k^i z_k^i + b_k^i$, $i = 1, 3$ preserves the same statistical characteristic as the original innovation z_k^1, the detection rate of the proposed linear attack is the same as that without attack. As mentioned above, $z_k^1 \sim \mathcal{N}(0, M)$ with $M = C_1^2 P_1 + R_1$, $\tilde{z}_k^1 \sim \mathcal{N}(0, T_k^1 M T_k^1 + L_k^1)$ and $\tilde{z}_k^3 \sim \mathcal{N}(0, T_k^3(CP_3C' + R)T_k^3 + L_k^3)$. Thus, to avoid being detected, the attack strategies (11) and (12) need to satisfy the stealthiness constraints

$$T_k^1 M T_k^1 + L_k^1 \leq M, \tag{13}$$

$$T_k^3(CP_3C' + R)T_k^{3\prime} + L_k^3 \leq M, \tag{14}$$

respectively. The notation P_3 stands for the steady-state value of the covariance matrix $\mathbb{E}[(x_k - \hat{x}_k^{3-})(x_k - \hat{x}_k^{3-})']$, which corresponds to the unique semi-definite solution of Riccati equation $X = A^2X + Q - A^2X^2C'(CXC' + R)^{-1}C$.

3.3 Problem of Interest

Based on the system model and proposed attack strategies, the problems we are interested in contain the following:

(1) How does the estimation error covariance evolve in the presence of the linear attack?
(2) What is the worst-case attack policy that yields the largest error covariance?
(3) What is the degradation of estimation performance under different attack strategies?

The detailed mathematical formulations and solutions to these problems will be introduced in the following sections.

4. PERFORMANCE ANALYSIS

Let $\tilde{x}_k^{i-}$ and $\tilde{x}_k^i$ be the *a priori* and the *a posteriori* MMSE estimates at the remote estimator in the presence of the proposed linear attack $\tilde{z}_k^i = T_k^i z_k^i + b_k^i$, $i = 1, 3$, which can be obtained from the recursion

$$\tilde{x}_k^{i-} = A\tilde{x}_{k-1}^i, \tag{15}$$

$$\tilde{x}_k^i = \tilde{x}_k^{i-} + K_1 \tilde{z}_k^i, \tag{16}$$

where the fixed gain K_1 is given in (9).

Since the false-data detector is unaware of the malicious attack if the stealthiness constraint is satisfied, the state estimate $\tilde{x}_k^i$ produced by the remote estimator will deviate from the true system state. To quantify the system performance, we define $\tilde{P}_k^{i-}$ and $\tilde{P}_k^i$ as the *a priori* and the *a posteriori* MMSE error covariance at the remote estimator under linear attack $\tilde{z}_k^i = T_k^i z_k^i + b_k^i$, $i = 1, 3$. The evolution of the error covariance under different types of attacks is summarized in the following.

4.1 Error Covariance under Attack using Intercepted Data

Lemma 1. (Guo et al., 2016) When the process (1)–(2) is under linear attack $\tilde{z}_k^1 = T_k^1 z_k^1 + b_k^1$, the estimation error covariance at the remote estimator follows the recursion

$$\tilde{P}_k^1 = A^2 \tilde{P}_{k-1}^1 + Q + (1 - 2T_k^1)\Delta_1,$$

where $\Delta_1 = C_1^2 P_1^2/(C_1^2 P_1 + R_1)$.

4.2 Error covariance under Attack using Intercepted and Sensing Data

Theorem 2. When the process (1)–(2) is under linear attack $\tilde{z}_k^3 = T_k^3 z_k^3 + b_k^3$, the estimation error covariance at the remote estimator follows the recursion

$$\tilde{P}_k^3 = A^2 \tilde{P}_{k-1}^3 + Q + \Delta_1 - 2K_1 T_k^3 C P_3, \tag{17}$$

where P_3 is the unique positive semi-definite solution of $X = A^2X + Q + A^2X^2C'(CXC' + R)^{-1}C$ with $C = [C_1, C_2]'$, $R = \mathrm{Diag}\{R_1, R_2\}$.

Proof. According to the process model (1)–(2) and state estimate iteration (15)–(16), the estimation error when the system is under linear attack $\tilde{z}_k^3 = T_k^3 z_k^3 + b_k^3$ follows

$$x_k - \tilde{x}_k^{3-} = A(x_{k-1} - \tilde{x}_{k-1}^3) + w_{k-1},$$

$$x_k - \tilde{x}_k^3 = x_k - \tilde{x}_k^{3-} - K_1 \tilde{z}_k^3,$$

based on which the error covariance at the remote estimator can be represented as

$$\begin{aligned}\tilde{P}_k^{3-} &= A^2\tilde{P}_{k-1}^3 + Q,\\ \tilde{P}_k^3 &= \tilde{P}_k^{3-} + \Delta_1 - 2\mathbb{E}[(x_k - \tilde{x}_k^{3-})\tilde{z}_k^3 K_1]\\ &= \tilde{P}_k^{3-} + \Delta_1 - 2\mathbb{E}[(x_k - \tilde{x}_k^{3-})(x_k - \hat{x}_k^{3-})T_k^3 C K_1],\end{aligned} \tag{18}$$

where the last equality is due to the fact that

$$\tilde{z}_k^3 = T_k^3 C(x_k - \hat{x}_k^{3-}) + T_k^3 v_k + b_k^3. \tag{19}$$

To obtain the explicit error iteration, we need to figure out the last term of (18). It is worth noticing that the corrupted innovation $\tilde{z}_k^3$ is used to update the state estimate in the presence of the attack while the true innovation z_k^1 is adopted in the absence of the attack. These two situations are considered separately as follows.

When the system is under linear attack $\tilde{z}_k^3 = T_k^3 z_k^3 + b_k^3$, according to (19), one has

$$\begin{aligned}x_k - \tilde{x}_k^{3-} =& A(x_{k-1} - \tilde{x}_{k-1}^{3-}) + w_{k-1} - AK_1T_{k-1}^3 C(x_{k-1}\\ &- \hat{x}_{k-1}^{3-}) - AK_1T_{k-1}^3 v_{k-1} - AK_1 b_{k-1}^3,\\ x_k - \hat{x}_k^{3-} =& A(1-KC)(x_{k-1} - \hat{x}_{k-1}^{3-}) + w_{k-1} - AKv_{k-1},\end{aligned}$$

from which the correlation of the estimation error between the estimator and the attacker is given by

$$\begin{aligned}P_k^{EA-} &= \mathbb{E}[(x_k - \tilde{x}_k^{3-})(x_k - \hat{x}_k^{3-})]\\ &= A^2(1-KC)P_{k-1}^{EA-} + Q\\ &\quad - A^2K_1T_k^3C(1-KC)P_3 + A^2K_1T_k^3RK'\\ &= A^2(1-KC)P_{k-1}^{EA-} + Q,\end{aligned} \tag{20}$$

where the second equality follows from $\mathbb{E}[(x_{k-1}-\hat{x}_{k-1}^{3-})^2] = P_3$ and $\mathbb{E}[v_{k-1}v_{k-1}'] = R$. The last equality follows from the fact that $K = P_3C'(CP_3C'+R)^{-1}$.

In the absence of the attack, the innovation z_k^1 is used to estimate the system state, i.e.,

$$\begin{aligned}&x_k - \tilde{x}_k^{3-}\\ &= A(1-K_1C_1)(x_{k-1} - \hat{x}_{k-1}^{1-}) + w_{k-1} - AK_1v_{k-1}^1.\end{aligned}$$

The correlation between the estimator and the attacker in this case follows

$$\begin{aligned}P_k^{ea-} &= \mathbb{E}[(x_k - \tilde{x}_k^{3-})(x_k - \hat{x}_k^{3-})]\\ &= A^2(1-K_1C_1)(1-KC)P_{k-1}^{ea-} + Q + A^2K_1\bar{K}_1R_1\\ &= A^2(1-KC)P_{k-1}^{ea-} + Q - A^2K_1C_1(1-KC)P_{k-1}^{ea-}\\ &\quad + A^2K_1C_1(1-KC)P_3,\end{aligned} \tag{21}$$

where the last equality follows from $K = [\bar{K}_1, \bar{K}_2] = (1-KC)P_3C'R^{-1} = [(1-KC)P_3C_1/R_1, (1-KC)P_3C_2/R_2]$.

According to the steady-state assumption, when the malicious attack occurs, the evolution of the correlation term between the estimator and the attacker follows (20) with $P_0^{EA-} = \lim_{k\to\infty} P_k^{ea-}$, i.e., the initial value of the correlation in the presence of the attack is the steady-state value of that in the absence of the attack. It can be observed from (21) that $\lim_{k\to\infty} P_k^{ea-} = P_3$. Note that P_3 is the unique positive semi-definite solution of $X = A^2X + Q + A^2X^2C'(CXC'+R)^{-1}C$, which coincides with the solution of (20). Hence, there is no dynamic in the evolution of the correlation term, i.e., $\mathbb{E}[(x_k-\tilde{x}_k^-)(x_k-\hat{x}_k^{3-})] = P_3, \forall k \in \mathbb{N}$. Therefore, the remote estimation error covariance (18) is obtained as

$$\tilde{P}_k^3 = A^2\tilde{P}_{k-1}^3 + Q + \Delta_1 - 2K_1T_k^3CP_3,$$

which completes the proof. ■

5. WORST-CASE LINEAR ATTACK STRATEGY

Based on the error covariance iteration obtained in last section, we derive a closed-form expression of the worst-case linear attack strategy and compare the attack consequence between different strategies in this section.

5.1 Worst-case Attack using Intercepted Data

Lemma 3. (Guo et al., 2016) For process (1)–(2) with the linear attack $\tilde{z}_k^1 = T_k^1 z_k^1 + b_k^1$, $T_k^1 = -1$ and $b_k^1 = 0$ is the worst-case linear attack strategy in the sense that the remote estimation error covariance is maximized.

5.2 Worst-case Attack using Intercepted and Sensing Data

Theorem 4. For process (1)–(2) with the linear attack $\tilde{z}_k^3 = T_k^3 z_k^3 + b_k^3$, the worst-case linear attack strategy which maximizes the remote estimation error covariance is $T_k^3 = -\sqrt{\frac{\Delta_1}{\Delta_3}\frac{K}{K_1}}$, where $\Delta_1 = C_1^2P^2/(C_1^2P_1 + R_1)$ and $\Delta_3 = P_3^2C'(CP_3C'+R)^{-1}C$.

Proof. In the proof, we first show that whether there exists malicious attacks during the past time instants or not, the optimal estimation gain at time k is the steady-state gain K. Then, the feasibility of the optimal gain is verified taking the stealthiness constraint into consideration, based on which we derive the closed-form expression of the worst-case linear attack strategy.

According to state estimate iteration (15)–(16), one has

$$x_k - \tilde{x}_k^3 = A(x_{k-1} - \tilde{x}_{k-1}^3) + w_{k-1} - \tilde{K}_k z_k^3,$$

where $\tilde{K}_k = K_1[T_k^3 + b_k^3(z_k^{3\prime}z_k^3)^{-1}z_k^3]$. Then, the estimation error covariance at the remote estimator is obtained as

$$\begin{aligned}\tilde{P}_k^3 &= A^2\tilde{P}_{k-1}^3 + Q + \tilde{K}_k(CP_3C'+R)\tilde{K}_k'\\ &\quad - 2\mathbb{E}\{\tilde{K}_k z_k^3[A(x_{k-1} - \tilde{x}_{k-1}^3) + w_{k-1}]\}.\end{aligned} \tag{22}$$

Note that to find the optimal state estimate at time k which minimizes the estimation error covariance $\tilde{P}_k^3$ is equivalent to find the optimal gain $\tilde{K}_k$. Now we focus on calculating the last term of (22). We first evaluate that

$$\begin{aligned}&A(x_{k-1} - \tilde{x}_{k-1}^3) + w_{k-1}\\ &= A^k(x_0 - \hat{x}_0^1) + \sum_{i=0}^{k-2} A^{k-1-i}w_i + w_{k-1} - \sum_{i=1}^{k-1} A^{k-i}\tilde{K}_i z_i^3,\end{aligned}$$

where the last equality follows from the assumption $\tilde{x}_0^3 = \hat{x}_0^1$. It can be also obtained that

$$\begin{aligned}z_k^3 &= CA[(1-KC)A]^{k-1}(x_0 - \hat{x}_0^3)\\ &\quad + CA\sum_{i=0}^{k-2}[(1-KC)A]^{k-2-i}(1-KC)w_i + Cw_{k-1}\\ &\quad + CA\sum_{i=0}^{k-1}[(1-KC)A]^{k-1-i}Kv_i + v_k.\end{aligned}$$

Due to the fact that $\mathbb{E}[z_i^3 z_j^{3\prime}] = 0, \forall i \neq j$, it now follows that the last term of (22) can be further simplified as

$$\begin{aligned}
&\mathbb{E}\left[\tilde{K}_k z_k^3[A(x_{k-1}-\tilde{x}_{k-1}^3)+w_{k-1}]\right]\\
&=\tilde{K}_k CA^2\Big\{(1-KC)^{k-1}A^{2(k-1)}P^{EA}\\
&\quad+\sum_{i=0}^{k-2}(1-KC)^{k-2-i}A^{2(k-2-i)}(1-KC)Q\Big\}+\tilde{K}_k CQ\\
&=\tilde{K}_k CA^2P^{EA}+\tilde{K}_k CQ\\
&=\tilde{K}_k CP_3, \qquad (23)
\end{aligned}$$

where the first equality follows from $\mathbb{E}[w_i^2]=Q$, $\forall i\in\mathbb{N}$ and $\mathbb{E}[(x_0-\hat{x}_0^3)(x_0-\hat{x}_0^1)]=P^{EA}$. The second equality is due to the fact that P^{EA} is the unique positive semi-definite fixed point of $X=(I-KC)A^2X+(I-KC)Q$. Hence, the remote estimation error covariance is given by

$$\begin{aligned}
\tilde{P}_k^3&=A^2\tilde{P}_{k-1}^3+Q+\tilde{K}_k(CP_3C'+R)\tilde{K}_k'-2\tilde{K}_kCP_3\\
&=A\tilde{P}_{k-1}^3A'+Q+(\tilde{K}_k-\Lambda)(CP_3C'+R)(\tilde{K}_k-\Lambda)'\\
&\quad-\Lambda(CP_3C'+R)\Lambda' \qquad (24)
\end{aligned}$$

with $\Lambda=P_3C'(CP_3C'+R)^{-1}=K$.

Based on above derivation, we now move to the stage of finding the worst-case linear attack strategy. It can be observed from (24) that no matter what $\tilde{P}_{k-1}^3$ is, the optimal state estimate at time k which minimizes the error covariance $\tilde{P}_k^3$ is obtained when $\tilde{K}_k=K_1T_k^3=K$, i.e., $T_k^3=\frac{K}{K_1}$. However, due to the existence of the false-data detector, the feasibility of T_k^3 needs to be verified. Multiplying K_1 on both sides of (14), one has

$$\begin{aligned}
&K_1T_k^3(CP_3C'+R)T_k^{3\prime}K_1=K(CP_3C'+R)K'=\Delta_3\\
&\le K_1(C_1^2P_1+R_1)K_1=\Delta_1,
\end{aligned}$$

which provides a feasibility condition for the optimal state estimate.

For the case that $\Delta_3>\Delta_1$, the optimal estimation gain cannot be achieved. Without lose of generality, we assume that $\tilde{K}_k=K_1T_k^3=K_1[T_k^{31},\ T_k^{32}]=[\lambda_1\bar{K}_1,\ \lambda_2\bar{K}_2]$. The estimation error covariance (17) and the stealthiness constraint (14) in this case can be represented as

$$\begin{aligned}
&\tilde{P}_k^3=A^2\tilde{P}_{k-1}^3+Q+\Delta_1-2\lambda_1\bar{K}_1C_1P_3-2\lambda_2\bar{K}_2C_2P_3,\\
&\lambda_1^2\bar{K}_1^2M_1+\lambda_2^2\bar{K}_2^2M_2+2\lambda_1\lambda_2\bar{K}_1\bar{K}_2M_{12}+L_k^3\le M,
\end{aligned}$$

where $M_1=C_1^2P_3+R_1$, $M_2=C_2^2P_3+R_2$, $M_{12}=C_1C_2P_3$ and $M=C_1^2P_1+R_1$. Note that the worst attack consequence is achieved when $L_k^3=0$ and $T_k^3(CP_3C'+R)T_k^{3\prime}=M$. In this case, finding optimal attack strategy is equivalent to solving problem

$$\begin{aligned}
\min_{\lambda_1,\lambda_2}\quad&\lambda_1\bar{K}_1C_1P_3+\lambda_2\bar{K}_2C_2P_3\\
s.t.\quad&\lambda_1^2\bar{K}_1^2M_1+\lambda_2^2\bar{K}_2^2M_2+2\lambda_1\lambda_2\bar{K}_1\bar{K}_2M_{12}=M.
\end{aligned}$$

Let the derivative of the Lagrangian with respect to λ_1 and λ_2 equal to zero, it can be obtained that $\lambda_1=\lambda_2$.

Hence, the optimal estimation gain when $\Delta_3>\Delta_1$ is in the form of $\tilde{K}_k=K_2T_k^3=\lambda K$. According to the stealthiness constraint $\lambda^2\Delta_3=\Delta_1$, the optimal estimator at time k is obtained when $T^*=T_k^3=\lambda\frac{K}{K_1}=\sqrt{\frac{\Delta_1}{\Delta_3}}\frac{K}{K_1}$, i.e., for any $T^\dagger=T^*+\Gamma$ with $\Gamma\in\mathbb{R}^{1\times2}$ being an arbitrary vector satisfying the constraint

$$T^\dagger(CP_3C'+R)T^{\dagger\prime}<C_1^2P_1+R_1,$$

one has

$$\tilde{P}_k^3(T^\dagger)-\tilde{P}_k^3(T^*)=-2K_1\Gamma CP_3\ge0.$$

Similarly, for any $T^\ddagger=-T^*-\Gamma$ different from $-T^*$, it can be observed that the stealthiness constraint

$$T^\ddagger(CP_3C'+R)T^{\ddagger\prime}=T^\dagger(CP_3C'+R)T^{\dagger\prime}<C_1^2P_1+R_1$$

is satisfied and

$$\tilde{P}_k^3(-T^*)-\tilde{P}_k^3(T^\ddagger)=-2K_1\Gamma CP_3\ge0.$$

Therefore, the worst-case linear attack strategy in the case that $\Delta_3>\Delta_1$ is $T_k^3=-\sqrt{\frac{\Delta_1}{\Delta_3}}\frac{K}{K_1}$ and $b_k^3=0$.

For the case where the constraint of the false-data detector is satisfied, i.e., $\Delta_3\le\Delta_1$, the optimal estimation gain is achievable, i.e., $T_k^3=\frac{K}{K_1}$. However, the worst-case linear attack strategy is not the simple negative of T_k^3. As we mentioned before, the remote estimation error covariance is maximized when $L_k^3=0$ and $T_k^3(CP_3C'+R)T_k^{3\prime}=M$. To obtain the worst-case attack policy, we assume that $T_k^3=[-\eta_1T_k^{31},\ -\eta_2T_k^{32}]$ without loss of generality. It can be derived that $\eta_1=\eta_2=\eta$ in a similar way of the case $\Delta_3>\Delta_1$. Hence, the stealthiness constraint becomes $\eta^2\Delta_3=\Delta_1$, which leads to the worst-case linear attack strategy $T_k^3=-\eta\frac{K}{K_1}=-\sqrt{\frac{\Delta_1}{\Delta_3}}\frac{K}{K_1}$. ■

5.3 Comparison between Different Attack Consequence

In this subsection, we compare the worst attack consequence for different linear attack strategies. The result is summarized in the following theorem.

Theorem 5. For process (1)–(2), the worst-case error covariance at the remote estimator under linear attack strategy $\tilde{z}_k^3=T_k^3z_k^3+b_k^3$ is

(1) larger than that under attack $\tilde{z}_k^1=T_k^1z_k^1+b_k^1$ if $|A|<1$;
(2) smaller than that under attack $\tilde{z}_k^1=T_k^1z_k^1+b_k^1$ if $|A|>1$;
(3) equal to that under attack $\tilde{z}_k^1=T_k^1z_k^1+b_k^1$ if $|A|=1$.

Proof. According to the worst-case error covariance iteration at the remote estimator

$$\tilde{P}_k^3=A^2\tilde{P}_{k-1}^3+Q+\Delta_1+2|\lambda|\Delta_3$$

when the system is under linear attack $\tilde{z}_k^3=T_k^3z_k^3+b_k^3$ and

$$\tilde{P}_k^1=A^2\tilde{P}_{k-1}^1+Q+3\Delta_1$$

when the system is under linear attack $\tilde{z}_k^1=T_k^1z_k^1+b_k^1$, we then focus on comparing the terms $|\lambda|\Delta_3$ and Δ_1. Note that P_1 and P_3 are the unique solutions of algebraic Riccati equations

$$P_1=A^2P_1+Q-A^2\Delta_1, \qquad (25)$$

$$P_3=A^2P_3+Q-A^2\Delta_3, \qquad (26)$$

and $P_1>P_3>0$. It can observed from (25)–(26) that $\Delta_3>\Delta_1$ if $|A|<1$, $\Delta_3<\Delta_1$ if $|A|>1$ and $\Delta_3=\Delta_1$ if $|A|=1$. Due to the worst-case stealthiness constraint $\lambda^2\Delta_3=\Delta_2$, one has $|\lambda|=\sqrt{\frac{\Delta_1}{\Delta_3}}<1$ if $|A|<1$, $|\lambda|=\sqrt{\frac{\Delta_1}{\Delta_3}}>1$ if $|A|>1$ and $|\lambda|=\sqrt{\frac{\Delta_1}{\Delta_3}}=1$ if $|A|=1$, which leads to the results $|\lambda|\Delta_3>\Delta$ if $|A|<1$, $|\lambda|\Delta_3<\Delta$ if $|A|>1$ and $|\lambda|\Delta_3=\Delta$ if $|A|=1$. ■

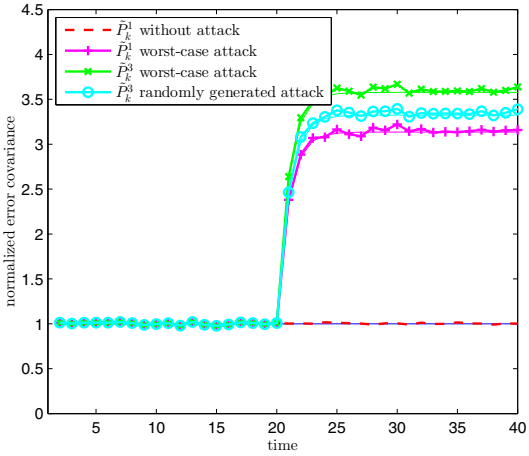

Fig. 2. Remote estimation error covariances under different worst-case linear attack strategies when $|A| < 1$.

6. SIMULATION EXAMPLE

To demonstrate the analytical results, we provide numerical examples in this section. We consider a stable process with parameters $A = 0.6$, $Q = 0.5$, $C_1 = 1$, $C_2 = 1$, $R_1 = 2$, $R_2 = 0.5$. As shown in Fig. 2, the green x-mark line, cyan circle line, magenta plus line and red dashed line represent the normalized error covariances under the worst-case attack using intercepted and sensing data, randomly generated attack using intercepted and sensing data, and the worst-case attack using intercepted data and without attack respectively. The malicious attacks start from the steady state. Observed from Fig. 2, the worst-case attack strategy yields a larger degradation of system estimation performance than a randomly generated attack. It can be also observed that for stable systems, the linear attack depends on the knowledge of two sensors is much powerful than that based on one sensor, which is consistent with the analytical results obtained in Theorem 5.

7. CONCLUSION

In this paper, we proposed an innovation-based integrity attack based on both intercepted and sensing data. The evolution of remote estimation error covariance was investigated in the presence of attack, based on which the worst-case strategy was obtained in closed form. Furthermore, we compared the attack consequence with the existing work to determine which strategy leads to a worse estimation performance. Simulation were provided to demonstrate the analytical results. Future works contain the consequence analysis for high dimensional systems and the influence of model uncertainty on the process $y_k \to z_k \to \tilde{z}_k \to \tilde{y}_k$.

REFERENCES

Anderson, B.D. and Moore, J.B. (2012). *Optimal filtering*. Courier Corporation.

Callegati, F., Cerroni, W., and Ramilli, M. (2009). Man-in-the-middle attack to the HTTPS protocol. *IEEE Security and Privacy Magazine*, (1), 78–81.

Guo, Z., Shi, D., Johansson, K.H., and Shi, L. (2016). Optimal linear cyber-attack on remote state estimation. *IEEE Transactions on Control of Network Systems*.

Gupta, A., Langbort, C., and Basar, T. (2010). Optimal control in the presence of an intelligent jammer with limited actions. In *IEEE Conference on Decision and Control*, 1096–1101.

Kim, K. and Kumar, P.R. (2012). Cyber–physical systems: A perspective at the centennial. *Proceedings of the IEEE*, 100(Special Centennial Issue), 1287–1308.

Kung, E., Dey, S., and Shi, L. (2016). The performance and limitations of-stealthy attacks on higher order systems. *IEEE Transactions on Automatic Control*.

Li, Y., Quevedo, D.E., Dey, S., and Shi, L. (2016). Sinr-based dos attack on remote state estimation: A game-theoretic approach. *IEEE Transactions on Control of Network Systems*.

Li, Y., Shi, L., Cheng, P., Chen, J., and Quevedo, D.E. (2015). Jamming attacks on remote state estimation in cyber-physical systems: A game-theoretic approach. *IEEE Transactions on Automatic Control*, 60(10), 2831–2836.

Liu, Y., Ning, P., and Reiter, M.K. (2011). False data injection attacks against state estimation in electric power grids. *ACM Transactions on Information and System Security*, 14(1), 13.

Mason, R.L. and Young, J.C. (2002). *Multivariate statistical process control with industrial applications*, volume 9. Siam.

Miao, F., Pajic, M., and Pappas, G.J. (2013). Stochastic game approach for replay attack detection. In *52nd IEEE Conference on Decision and Control*, 1854–1859.

Mo, Y., Chabukswar, R., and Sinopoli, B. (2014). Detecting integrity attacks on scada systems. *IEEE Transactions on Control Systems Technology*, 22(4), 1396–1407.

Mo, Y., Garone, E., Casavola, A., and Sinopoli, B. (2010). False data injection attacks against state estimation in wireless sensor networks. In *49th IEEE Conference on Decision and Control*, 5967–5972.

Mo, Y. and Sinopoli, B. (2009). Secure control against replay attacks. In *47th Annual Allerton Conference on Communication, Control, and Computing*, 911–918.

Mo, Y., Weerakkody, S., and Sinopoli, B. (2015). Physical authentication of control systems: designing watermarked control inputs to detect counterfeit sensor outputs. *IEEE Control Systems*, 35(1), 93–109.

Pouliezos, A. and Stavrakakis, G.S. (2013). *Real time fault monitoring of industrial processes*, volume 12. Springer Science & Business Media.

Sandberg, H., Amin, S., and Johansson, K.H. (2015). Cyberphysical security in networked control systems: An introduction to the issue. *IEEE Control Systems Magazine*, 35(1), 20–23.

Shi, D., Chen, T., and Darouach, M. (2016a). Event-based state estimation of linear dynamic systems with unknown exogenous inputs. *Automatica*, 69, 275–288.

Shi, D., Elliott, R.J., and Chen, T. (2016b). On finite-state stochastic modeling and secure estimation of cyber-physical systems.

Smith, R.S. (2015). Covert misappropriation of networked control systems: Presenting a feedback structure. *IEEE Control Systems Magazine*, 35(1), 82–92.

Zhang, H., Cheng, P., Shi, L., and Chen, J. (2015). Optimal denial-of-service attack scheduling with energy constraint. *IEEE Transactions on Automatic Control*, 60(11), 3023–3028.