Worst-Case Innovation-Based Integrity Attacks With Side Information on Remote State Estimation

Ziyang Guo, Dawei Shi, Karl Henrik Johansson, *Fellow, IEEE*, and Ling Shi, *Senior Member, IEEE*

***Abstract*—In this paper, we study the worst-case consequence of innovation-based integrity attacks with side information in a remote state estimation scenario where a sensor transmits its measurement to a remote estimator equipped with a false-data detector. If a malicious attacker is not only able to compromise the transmitted data packet but also able to measure the system state itself, the attack strategy can be designed based on the intercepted data, the sensing data, or alternatively the combined information. Surprisingly, we show that launching attacks using the combined information are not always optimal. First, we characterize the stealthiness constraints for different types of attack strategies to avoid being noticed by the false-data detector. Then, we derive the evolution of the remote estimation error covariance in the presence of attacks, based on which the worst-case attack policies are obtained by solving convex optimization problems. Furthermore, the closed-form expressions of the worst-case attacks are obtained for scalar systems and the attack consequences are compared with the existing work to determine which strategy is more critical in deteriorating system performance. Simulation examples are provided to illustrate the analytical results.**

***Index Terms*—Cyber-physical system (CPS) security, integrity attack, remote state estimation.**

I. Introduction

INCREASING applications of cyber-physical systems (CPS) in critical infrastructures ranging from national power grids to manufacturing processes have reinforced the safety and security requirements in the control system design. Due to the interconnection between different components and technologies, CPSs are vulnerable to cyber threats that may cause severe consequences on national economy, social security, or even loss of human lives [1], [2]. Recently reported accidents (e.g., StuxNet malware [3], Maroochy water bleach [4]) evidently indicate the fundamental importance of security in CPS. In this regard, worst-case consequence analysis and defense mechanism design for CPS have attracted considerable interest from both academic and industrial communities [5], [6].

The cyber-physical attack space can be divided according to an adversary's system knowledge, disclosure resources, and disruption resources [7]. False-data injection attacks, a particular type of integrity attack, were initially proposed for electric power grids in [8]. The consequence of such an attack was investigated in a remote estimation scenario and a quantitative measure of system resilience too was proposed [9]. Furthermore, the tradeoff between attack stealthiness and estimation quality was analyzed for control signal injection in [10]. Replay attacks degrade system performance by recording and replaying the sensor data without knowledge of system parameters. The feasibility conditions and countermeasures of replay attacks were investigated for LQG control systems in [11]. The tradeoff between system performance and detection rate for replay attacks was studied under a stochastic game framework in [12]. Denial-of-service (DoS) attacks attempt 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 [13]–[15]. Besides the above works which only focused on either the attacker or the defender, game-theoretic approaches were proposed to investigate the optimal scheduling problems by taking both sides into consideration [16], [17].

An innovation-based linear integrity attack, which was designed based on the intercepted innovation sequence from being noticed by the χ^2 false-data detector, was proposed in [18]. The evolution of the estimation error covariance and the worst-case attack strategy were derived. Different from [18], this paper focuses on the scenario where the malicious attacker has side information about the system, which is formulated as an extra measurement. This side information provides more freedom for the attack policy design. Specifically, we investigate attack strategies characterized by the information sets available

Manuscript received September 25, 2017; revised December 5, 2017; accepted January 1, 2018. Date of publication January 15, 2018; date of current version March 14, 2019. The work of Z. Guo and L. Shi was supported by the Hong Kong RGC General Research Fund 16222716. The work of D. Shi was supported by the Natural Science Foundation of China under Grant 61503027. The work of K. H. Johansson was supported by the Knut and Alice Wallenberg Foundation and the Swedish Research Council. Recommended by Associate Editor S. Dey. *(Corresponding author: Dawei Shi.)*

Z. Guo and L. Shi are with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong (e-mail: zguoae@ust.hk; eesling@ust.hk).

D. Shi is with the Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA (e-mail: dawei.shi@outlook.com).

K. H. Johansson is with the ACCESS Linnaeus Centre, School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm 11428, Sweden (e-mail: kallej@kth.se).

Digital Object Identifier 10.1109/TCNS.2018.2793664

to the attacker and derive the stealthiness constraints needed to avoid being detected by the false-data detector. The estimation performance for different attack scenarios is analyzed, based on which the worst-case attack policy is further investigated. Moreover, the attack consequences between different strategies are compared to determine which strategy is more critical to hamper system performance. To our best knowledge, this is the first work that considers malicious attacks with side information. A preliminary version of these results is available in [19]. Different from [19], we investigate more comprehensive attack scenarios in this paper and analyze the estimation performance and the worst-case attack policy for multivariable systems rather than scalar systems, which involves more elaborated mathematical treatment.

The contributions of this paper are summarized as follows.

1) We propose innovation-based linear attack strategies based on the intercepted data (Scenario I), the sensing data (Scenario II), as well as the combined information (Scenario III). The stealthiness constraints for the proposed attack policies are investigated to avoid being noticed by the false-data detector.
2) We derive the evolution of the estimation error covariance when the system is under attack (*Lemmas 2 and 3*), based on which the worst-case linear attack strategy is obtained for multivariable systems by solving a convex optimization problem (*Theorems 1 and 2*).
3) For scalar systems, the worst-case linear attack strategy is obtained in a closed form (*Propositions 1 and 2*). The estimation error under different attacks is compared to determine which strategy is more critical in deteriorating the system performance (*Corollaries 1 and 2*).

The remainder of this paper is organized as follows. Section II introduces the system architecture. Section III presents three types of innovation-based linear attack strategies and corresponding stealthiness constraints. Section IV derives the evolution of the remote estimation error covariance in the presence of attacks. Section V investigates the worst-case attack policy for multivariable systems. Section VI obtains the closed-form worst-case attack for scalar systems and compares different attack consequences. Section VII discusses the mitigation strategies of linear attacks. Numerical examples are provided in Section VIII. Some concluding remarks are made in Section IX.

Notations: $\mathbb{N}$ and $\mathbb{R}$ denote the sets of non-negative integers and real numbers, respectively. $\mathbb{R}^n$ is the n-dimensional Euclidean space. $\mathbb{S}_+^n$ ($\mathbb{S}_{++}^n$) is the set of $n \times n$ positive semidefinite (definite) matrices. When $X \in \mathbb{S}_+^n$ ($\mathbb{S}_{++}^n$), we simply write $X \geq 0$ ($X > 0$). The superscript $'$ and $\mathrm{Tr}(\cdot)$ stand for the transpose and the trace of a matrix, respectively. $\mathrm{Diag}\{\cdot\}$ represents a block-diagonal matrix.

II. System Architecture

The system block diagram is shown in Fig. 1. A sensor measures a physical plant and transmits the measurement data to a remote estimator through a wireless network. The attacker attempts to intercept and modify the transmitted data, which may degrade the estimation performance without being noticed by the false-data detector. The detailed system model is presented in this section, while the considered attack policy is introduced in Section III.

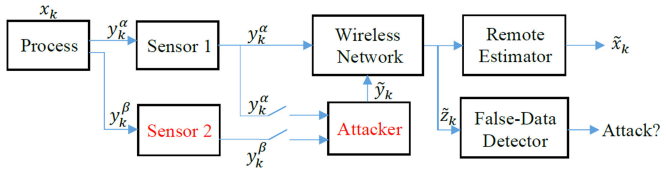


Fig. 1. System setup with an attacker having access to two sensors when conducting an integrity attack over a wireless network.

A. Process Model

We consider a linear time-invariant process described by

$$x_{k+1} = Ax_k + w_k \tag{1}$$

$$y_k^\alpha = C_\alpha x_k + v_k^\alpha \tag{2}$$

where $k \in \mathbb{N}$ is the time index, $x_k \in \mathbb{R}^n$ is the system state, $y_k^\alpha \in \mathbb{R}^m$ is the sensor measurement, and $w_k \in \mathbb{R}^n$ and $v_k^\alpha \in \mathbb{R}^m$ are zero-mean independent identically distributed (i.i.d.) Gaussian noises with covariances $Q \geq 0$ and $R_\alpha > 0$, respectively. The initial state x_0 is zero-mean Gaussian with covariance matrix $\Pi_0 \geq 0$ and independent of w_k and v_k^α for all $k \geq 0$. The pair (A, C_α) is detectable and $(A, \sqrt{Q})$ is stabilizable. The superscript $^\alpha$ is used to represent quantities related to the measurement from sensor 1.

B. 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, a Kalman filter is adopted by the remote estimator to process the received data

$$\begin{aligned}
\hat{x}_k^{\alpha-} &= A\hat{x}_{k-1}^\alpha \\
P_k^{\alpha-} &= AP_{k-1}^\alpha A' + Q \\
\hat{x}_k^\alpha &= \hat{x}_k^{\alpha-} + K_k^\alpha (y_k^\alpha - C_\alpha \hat{x}_k^{\alpha-}) \\
K_k^\alpha &= P_k^{\alpha-} C_\alpha' (C_\alpha P_k^{\alpha-} C_\alpha' + R_\alpha)^{-1} \\
P_k^\alpha &= (I - K_k^\alpha C_\alpha) P_k^{\alpha-}
\end{aligned}$$

where $\hat{x}_k^{\alpha-}$ and $\hat{x}_k^\alpha$ are, respectively, the *a priori* and the *a posteriori* minimum mean squared error (MMSE) estimates of the state x_k, and $P_k^{\alpha-}$ and P_k^α the corresponding estimation error covariances. The recursion starts from $\hat{x}_k^{\alpha-} = 0$ and $P_k^{\alpha-} = \Pi_0$. It is well known that the Kalman filter converges exponentially fast from any initial condition [20]. We denote the steady-state values as

$$P_\alpha \triangleq \lim_{k \to +\infty} P_k^{\alpha-}$$

$$K_\alpha \triangleq P_\alpha C_\alpha' (C_\alpha P_\alpha C_\alpha' + R_\alpha)^{-1}$$

where P_α is the unique positive semidefinite solution of $X = AXA' + Q - AXC_\alpha'(C_\alpha XC_\alpha' + R_\alpha)^{-1} C_\alpha XA'$. Without loss of generality, we assume that the remote estimator starts from the steady state, that is, $P_k^{\alpha-} = P_\alpha \ \forall k \in \mathbb{N}$.

C. False-Data Detector

A false-data detector is equipped at the remote side to monitor system behavior and detect the existence of potential malicious attacks. The innovation sequence $z_k^\alpha = y_k^\alpha - C_\alpha \hat{x}_k^{\alpha-}$ has a steady-state Gaussian distribution $\mathcal{N}(0, C_\alpha P_\alpha C_\alpha' + R_\alpha)$ and $\mathbb{E}[z_i^\alpha z_j^{\alpha\prime}] = 0$ for all $i \neq j$ [20]. Hence, its statistical characteristics (mean and covariance) can be used to diagnose the system anomalies.

The χ^2 false-data detector is a residue-based detector widely used for fault detection in the process industry and studied in the research community [11], [12], [21], [22]. It makes a decision based on the sum of the normalized innovation sequence, that is, at time k, the detection criterion follows the hypothesis test:

$$g_k = \sum_{i=k-J+1}^{k} z_i^{\alpha\prime}(C_\alpha P_\alpha C_\alpha' + R_\alpha)^{-1} z_i^\alpha \underset{H_1}{\overset{H_0}{\lessgtr}} \delta \tag{3}$$

where J is the detection window size, 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 (3) satisfies the χ^2 distribution with mJ degrees of freedom. Thus, the false alarm rate in the absence of the attack can be easily calculated.

III. Attack Strategy and Stealthiness Constraint

In this section, we consider the scenario where there exists a malicious agent who intentionally launches cyber attacks to degrade the estimation performance of the system described in Section II. The attacker is not only able to intercept the transmitted data packet from the sensor to the remote estimator, but also has an extra sensor to measure the system state itself, see Fig. 1. We introduce three different attack strategies and analyze the stealthiness constraints for these attacks from being detected by the false-data detector. Without loss of generality, we assume that the attack starts from time instant $k = 0$.

A. Linear Attack Strategy

Motivated by attack models in the literature [2], [23], [24], we assume that the attacker has knowledge of the process model and is capable of intercepting and modifying the transmitted measurement. It is worth noticing that modifying the measurement is equivalent to modifying the innovation under this assumption. Specifically, based on the system parameters and the available measurement information y_k, the attacker is able to implement a filter to first calculate the true innovation $z_k = y_k - C\hat{x}_k^-$ 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 $\tilde{y}_k = \tilde{z}_k + C\tilde{x}_k^-$ with $\tilde{x}_k^- = \mathbb{E}[x_k|\tilde{y}_{1:k-1}]$. The procedure $y_k \to z_k \to \tilde{z}_k \to \tilde{y}_k$ means that generating the manipulated measurement $\tilde{y}_k$ is equivalent to generating the compromised innovation $\tilde{z}_k$. Moreover, similar to [18], [19], and [25], we consider the innovation-based attack in this paper, that is, the information set for the attacker at time k is the innovation z_k. Thus, a general attack strategy can be defined as

$$\tilde{z}_k = f_k(z_k) \tag{4}$$

where f_k is an arbitrary function with appropriate domain. A malicious attacker considered in this paper aims at maximizing the remote estimation error covariance while simultaneously remaining stealthy to the false-data detector. Hence, in order to achieve this objective, the attacker needs to characterize the stealthiness constraint and quantify the attack effect on the system estimation quality. However, if f_k is a nonlinear function or does not have an explicit expression, it is difficult for the malicious attacker to guarantee attack stealthiness, not to mention launching an optimal attack. Instead, if a linear attack policy is considered, it is possible for the malicious attacker to analytically quantify the stealthiness constraint and the attack consequence such that the optimal stealthy attack can be launched. Moreover, a linear attack might be much easier to implement in practice. Motivated by the aforementioned observations, we focus on the subset of all linear attack strategies in this initial study where f_k is an affine function of the innovation z_k as

$$\tilde{z}_k = T_k z_k + b_k \tag{5}$$

where T_k is an arbitrary matrix with appropriate dimension and b_k is the Gaussian random variable. Due to the fact that the attack policy varies with the information set available at the malicious attacker, we further introduce three different attack scenarios specified by the superscripts as follows.

1) Scenario I: For the case where the malicious attacker is capable of intercepting the transmitted data, the attack strategy is designed based on the system innovation z_k^α as

$$\tilde{z}_k^\alpha = T_k^\alpha z_k^\alpha + b_k^\alpha \tag{6}$$

where $z_k^\alpha = y_k^\alpha - C_\alpha \hat{x}_k^{\alpha-} \in \mathbb{R}^m$ is the currently intercepted innovation, $\tilde{z}_k^\alpha \in \mathbb{R}^m$ is the innovation modified by the attacker, $T_k^\alpha \in \mathbb{R}^{m\times m}$ is an arbitrary matrix, and $b_k^\alpha \in \mathbb{R}^m$ is a zero-mean i.i.d. Gaussian random variable with covariance L_k^α and independent of z_k^α. This is corresponds to the attack strategy studied in [18]. It can be observed that $\tilde{z}_k^\alpha$ is zero-mean Gaussian distributed with covariance $T_k^\alpha(C_\alpha P_\alpha C_\alpha' + R_\alpha)T_k^\alpha + L_k^\alpha$.

2) Scenario II: For the case where the malicious attacker cannot successfully eavesdrop the transmitted data but is instead able to measure the system state using an extra sensor, that is, $y_k^\beta = C_\beta x_k + v_k^\beta$, where (A, C_β) is detectable and $v_k^\beta \in \mathbb{R}^m$ is a white Gaussian noise with zero mean and covariance $R_\beta > 0$, the linear attack strategy is designed with respect to the attacker's own information as

$$\tilde{z}_k^\beta = T_k^\beta z_k^\beta + b_k^\beta \tag{7}$$

where $z_k^\beta = y_k^\beta - C_\beta \hat{x}_k^{\beta-} \in \mathbb{R}^m$ is the innovation calculated by the attacker (i.e., using a Kalman filter and measurement y_k^β), $\tilde{z}_k^\beta \in \mathbb{R}^m$ is the corrupted innovation, $T_k^\beta \in \mathbb{R}^{m\times m}$ is an arbitrary attack matrix, and $b_k^\beta \in \mathbb{R}^m$ is a zero-mean i.i.d. Gaussian random variable with covariance L_k^β and independent of z_k^β. It is worth noticing that $\tilde{z}_k^\beta \sim \mathcal{N}(0, T_k^\beta(C_\beta P_\beta C_\beta' + R_\beta)T_k^{\beta\prime} + L_k^\beta)$ since $z_k^\beta \sim \mathcal{N}(0, C_\beta P_\beta C_\beta' + R_\beta)$, where P_β is the steady-state value of the covariance matrix $\mathbb{E}[(x_k - \hat{x}_k^{\beta-})(x_k - \hat{x}_k^{\beta-})']$, that is, the unique positive semidefinite solution of $X = AXA' + Q - AXC_\beta'(C_\beta XC_\beta' + R_\beta)^{-1}C_\beta XA'$.

3) Scenario III: For the case where the malicious agent is simultaneously able to intercept the transmitted data and measure the system state with an extra sensor, the information owned by the attacker is given by $y_k = Cx_k + v_k$, where $y_k = [y_k^\alpha \; y_k^\beta]' \in \mathbb{R}^{2m}$, $C = [C_\alpha', C_\beta']' \in \mathbb{R}^{2m\times n}$, $v_k = [v_k^\alpha, v_k^\beta] \in \mathbb{R}^m$ is white Gaussian noise with covariance matrix $R = \mathrm{Diag}\{R_\alpha, R_\beta\}$, and (A, C) is detectable. In this case, the attack strategy can be designed based on the intercepted and the sensing data together, which is defined as

$$\tilde{z}_k^\gamma = T_k^\gamma z_k^\gamma + b_k^\gamma \tag{8}$$

where $z_k^\gamma = y_k - C\hat{x}_k^{\gamma-} \in \mathbb{R}^{2m}$ is the innovation calculated by the malicious attacker based on measurement y_k, $\tilde{z}_k^\gamma \in \mathbb{R}^{2m}$ is the corrupted innovation, $T_k^\gamma = [T_k^{\gamma 1}, T_k^{\gamma 2}] \in \mathbb{R}^{m\times 2m}$ is an arbitrary attack matrix, and $b_k^\gamma \in \mathbb{R}^{2m}$ is a zero-mean i.i.d. Gaussian random variable with covariance L_k^γ and independent of z_k^γ. Hence, $\tilde{z}_k^\gamma$ is Gaussian distributed with zero mean and covariance $T_k^\gamma(CP_\gamma C' + R)T_k^{\gamma\prime} + L_k^\gamma$ with P_γ being the steady-state value of the covariance matrix $\mathbb{E}[(x_k - \hat{x}_k^{\gamma-})(x_k - \hat{x}_k^{\gamma-})']$, which corresponds to the unique positive semidefinite solution of $X = AXA' + Q - AXC'(CXC' + R)^{-1}CXA'$.

B. Stealthiness Constraints

For the aforementioned three types of attack strategies, the objective of the malicious agent is to degrade the estimation performance as much as possible and simultaneously remain stealthy to the false-data detector. Recalling the binary hypothesis test in (3), we define p_k^F as the false alarm rate in the absence of the attack (decide H_1 when H_0 is true) and p_k^D as the detection rate in the presence of the attack (decide H_1 when H_1 is true). In detection theory, the performance of the detector can be characterized by the tradeoff between p_k^F and p_k^D [26]. Note that for the considered detection criterion (3), the proposed linear attack strategy $\tilde{z}_k^i = T_k^i z_k^i + b_k^i$, $i = \alpha, \beta, \gamma$ is strictly stealthy if $p_k^D = p_k^F \; \forall k > 0$. Namely, the malicious attacker remains undetectable to the false-data detector if the covariance of the corrupted innovation $\tilde{z}_k^i$ is equal to the covariance of the steady-state innovation z_k^α. Due to the fact that $z_k^\alpha \sim \mathcal{N}(0, C_\alpha P_\alpha C_\alpha' + R_\alpha)$, the stealthiness constraints for linear attack strategies (6), (7), and (8) are obtained as follows:

$$T_k^\alpha(C_\alpha P_\alpha C_\alpha' + R_\alpha)T_k^{\alpha\prime} + L_k^\alpha = C_\alpha P_\alpha C_\alpha' + R_\alpha \tag{9}$$

$$T_k^\beta(C_\beta P_\beta C_\beta' + R_\beta)T_k^{\beta\prime} + L_k^\beta = C_\alpha P_\alpha C_\alpha' + R_\alpha \tag{10}$$

$$T_k^\gamma(CP_\gamma C' + R)T_k^{\gamma\prime} + L_k^\gamma = C_\alpha P_\alpha C_\alpha' + R_\alpha \tag{11}$$

where P_α, P_β, and P_γ are the steady-state values of the error covariances $\mathbb{E}[(x_k - \hat{x}_k^{\alpha-})(x_k - \hat{x}_k^{\alpha-})']$, $\mathbb{E}[(x_k - \hat{x}_k^{\beta-})(x_k - \hat{x}_k^{\beta-})']$, and $\mathbb{E}[(x_k - \hat{x}_k^{\gamma-})(x_k - \hat{x}_k^{\gamma-})']$, respectively.

C. Problems of Interest

Based on the system model and the proposed attack strategy, the problems we are interested in consist as follows.

1) How does the estimation error covariance evolve in the presence of the attack?
2) What is the worst-case attack policy that yields the largest error covariance?
3) What is the worst-case performance gap between the attack strategies?

The detailed mathematical formulations and solutions to these problems will be introduced in Sections IV and V.

IV. Performance Analysis

For the considered process (1), (2) under the linear attack $\tilde{z}_k^i = T_k^i z_k^i + b_k^i$, $i = \alpha, \beta, \gamma$, the remote state estimate follows:

$$\tilde{x}_k^{i-} = A\tilde{x}_{k-1}^i \tag{12}$$

$$\tilde{x}_k^i = \tilde{x}_k^{i-} + K_\alpha \tilde{z}_k^i \tag{13}$$

where K_α is the steady-state gain, $\tilde{x}_k^{i-}$ and $\tilde{x}_k^i$ are, respectively, the *a priori* and the *a posteriori* MMSE state estimates in the presence of the linear attack.

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 $\tilde{P}_k^i$, respectively, as the *a priori* and the *a posteriori* MMSE error covariance at the remote estimator when the system is under linear attack $\tilde{z}_k^i = T_k^i z_k^i + b_k^i$, $i = \alpha, \beta, \gamma$. In addition, we denote $P_{k,i}^{EA-} \triangleq \mathbb{E}[(x_k - \tilde{x}_k^{i-})(x_k - \hat{x}_k^{i-})']$ as the correlation of the estimation error between the estimator and the attacker in the presence of the attack and $P_{k,i}^{ea-} \triangleq \mathbb{E}[(x_k - \tilde{x}_k^{i-})(x_k - \hat{x}_k^{i-})']$ as the same amount in the absence of the attack. The evolution of the estimation error covariance under different attack strategies is investigated in the following section.

A. Error Covariance Under Attack Using Intercepted Data

For Scenario I where the attack strategy is designed based on the system innovation z_k^α, the error covariance at the remote side is summarized in the following lemma.

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

$$\begin{aligned}\tilde{P}_k^\alpha &= A\tilde{P}_{k-1}^\alpha A' + Q + P_\alpha C_\alpha'(C_\alpha P_\alpha C_\alpha' + R_\alpha)^{-1}C_\alpha P_\alpha \\ &\quad - P_\alpha C_\alpha' T_k^{\alpha\prime} K_\alpha' - K_\alpha T_k^\alpha C_\alpha P_\alpha\end{aligned} \tag{14}$$

where P_α is the unique positive semidefinite solution of $X = AXA + Q + AXC_\alpha'(C_\alpha XC_\alpha' + R_\alpha)^{-1}C_\alpha XA'$.

Proof: See [18, Th. 1]. ■

B. Error Covariance Under Attack Using Intercepted and Sensing Data

In this section, we focus on the attack strategy based on both the intercepted and the sensing data, that is, Scenario III. The estimation error covariance iteration under such an attack is obtained in the following lemma.

Lemma 2: When the process (1), (2) is under attack $\tilde{z}_k^\gamma = T_k^\gamma z_k^\gamma + b_k^\gamma$, the estimation error covariance at the remote

estimator follows the recursion:

$$\tilde{P}_k^\gamma = A\tilde{P}_{k-1}^\gamma A' + Q + P_\alpha C_\alpha'(C_\alpha P_\alpha C_\alpha' + R_\alpha)^{-1}C_\alpha P_\alpha - P_\gamma C' T_k^{\gamma\prime} K_\alpha' - K_\alpha T_k^\gamma C P_\gamma \quad (15)$$

where P_γ is the unique positive semidefinite solution of $X = AXA' + Q + AXC'(CXC' + R)^{-1}CXA'$, $C = [C_\alpha, C_\beta]'$, $R = \mathrm{Diag}\{R_\alpha, R_\beta\}$.

Proof: According to the process model (1), (2) and the state estimate update (12), (13), the estimation error when the system is under attack $\tilde{z}_k^\gamma = T_k^\gamma z_k^\gamma + b_k^\gamma$ follows:

$$x_k - \tilde{x}_k^\gamma = A(x_{k-1} - \tilde{x}_{k-1}^\gamma) + w_{k-1} - K_\alpha \tilde{z}_k^\gamma$$

based on which the error covariance is represented as

$$\begin{aligned}\tilde{P}_k^\gamma &= \mathbb{E}[(x_k - \tilde{x}_k^\gamma)(x_k - \tilde{x}_k^\gamma)'] \\ &= A\tilde{P}_{k-1}^\gamma A' + Q + K_\alpha(C_\alpha P_\alpha C_\alpha' + R_\alpha)K_\alpha' \\ &\quad - \mathbb{E}[(x_k - \tilde{x}_k^{\gamma-})(x_k - \hat{x}_k^{\gamma-})' C' T_k^{\gamma\prime} K_\alpha'] \\ &\quad - \mathbb{E}[K_\alpha T_k^\gamma C(x_k - \hat{x}_k^{\gamma-})(x_k - \tilde{x}_k^{\gamma-})'] \end{aligned} \quad (16)$$

where the second equality is due to the fact that

$$\tilde{z}_k^\gamma = T_k^\gamma C(x_k - \hat{x}_k^{\gamma-}) + T_k^\gamma v_k + b_k^\gamma. \quad (17)$$

To obtain the explicit error iteration, we need to expand the last two terms of (16). It is worth noticing that the corrupted innovation $\tilde{z}_k^\gamma$ is used to update the state estimate in the presence of the attack, while the true innovation z_k^α is adopted in the absence of the attack. These two situations are considered separately as follows.

When the system is under attack $\tilde{z}_k^\gamma = T_k^\gamma z_k^\gamma + b_k^\gamma$, according to (17), one has

$$\begin{aligned} x_k - \tilde{x}_k^{\gamma-} &= A(x_{k-1} - \tilde{x}_{k-1}^{\gamma-}) + w_{k-1} - AK_\alpha T_{k-1}^\gamma v_{k-1} \\ &\quad - AK_\alpha T_{k-1}^\gamma C(x_{k-1} - \hat{x}_{k-1}^{\gamma-}) - AK_\alpha b_{k-1}^\gamma \\ x_k - \hat{x}_k^{\gamma-} &= A(I - KC)(x_{k-1} - \hat{x}_{k-1}^{\gamma-}) + w_{k-1} - AKv_{k-1} \end{aligned}$$

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

$$\begin{aligned} P_{k,\gamma}^{EA-} &= \mathbb{E}[(x_k - \tilde{x}_k^{\gamma-})(x_k - \hat{x}_k^{\gamma-})'] \\ &= AP_{k-1,\gamma}^{EA-}(I - KC)'A' + Q \\ &\quad - AK_\alpha T_{k-1}^\gamma C P_\gamma (I - KC)'A' + AK_\alpha T_{k-1}^\gamma R K' A' \\ &= AP_{k-1,\gamma}^{EA-}(I - KC)'A' + Q \end{aligned} \quad (18)$$

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

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

$$\begin{aligned} x_k - \tilde{x}_k^{\gamma-} &= A(I - K_\alpha C_\alpha)(x_{k-1} - \tilde{x}_{k-1}^{\gamma-}) \\ &\quad + w_{k-1} - AK_\alpha v_{k-1}^\alpha. \end{aligned}$$

In this case, the correlation of the estimation error between the estimator and attacker follows:

$$\begin{aligned} &P_{k,\gamma}^{ea-} \\ &= \mathbb{E}[(x_k - \tilde{x}_k^{\gamma-})(x_k - \hat{x}_k^{\gamma-})'] \\ &= A(I - K_\alpha C_\alpha)P_{k-1,\gamma}^{ea-}(I - KC)'A' + Q + AK_\alpha R_\alpha \bar{K}_1' A' \\ &= AP_{k-1,\gamma}^{ea-}(I - KC)'A' + Q - AK_\alpha C_\alpha P_{k-1,\gamma}^{ea-}(I - KC)'A' \\ &\quad + AK_\alpha C_\alpha P_\gamma (I - KC)'A' \end{aligned} \quad (19)$$

where the last equality follows from $K \triangleq [\bar{K}_1, \bar{K}_2] = (I - KC)P_\gamma C' R^{-1} = [(I - KC)P_\gamma C_\alpha' R_\alpha^{-1}, (I - KC)P_\gamma C_\beta' R_\beta^{-1}]$.

Due to the fact that P_γ is the unique positive semidefinite solution of $X = AX(I - KC)'A' + Q$ with $K = XC'(CXC' + R)^{-1}$, it can be observed that if the initial value $P_{0,\gamma}^{EA-}$ of (18) satisfies $P_{0,\gamma}^{EA-} = P_\gamma$, the correlation term $P_{k,\gamma}^{EA-}$ will be time-invariant, that is, $P_{k,\gamma}^{EA-} = P_\gamma \ \forall k \in \mathbb{N}$, which leads to error covariance recursion (15) and finishes the proof. Hence, to show Lemma 2, it suffices to show $P_{0,\gamma}^{EA-} = P_\gamma$. Note that the initial value of the correlation in the presence of attacks is the steady-state value of that in the absence of attacks, that is, $P_{0,\gamma}^{EA-} = \lim_{k\to\infty} P_{k,\gamma}^{ea-}$. Consequently, showing $P_{0,\gamma}^{EA-} = P_\gamma$ is equivalent to showing $\lim_{k\to\infty} P_{k,\gamma}^{ea-} = P_\gamma$ for any initial value P_0^{ea-}.

By denoting $\zeta_k \triangleq P_k^{ea-} - P_\gamma \in \mathbb{R}^{n\times n}$, the above problem is equivalent to showing $\lim_{k\to\infty}\zeta_k = 0$ for any initial value ζ_0. From (19), the recursion of ζ_k can be obtained as

$$\begin{aligned}\zeta_k &= A(I - K_\alpha C_\alpha)\zeta_{k-1}(I - KC)'A' \\ &= [A(I - K_\alpha C_\alpha)]^k \zeta_0 [(I - KC)'A']^k.\end{aligned}$$

According to [27, Th. 5.6.12], for a matrix $X \in \mathbb{R}^{n\times n}$, $\lim_{k\to\infty} X^k = 0$ if and only if $\rho(X) < 1$, where $\rho(X)$ denotes the spectral radius of X. Due to the fact that $\rho(A(I - K_\alpha C_\alpha)) < 1$ and $\rho(A(I - KC)) < 1$ [20], it follows that $\lim_{k\to\infty}\zeta_k = 0 \, \forall \zeta_0 \in \mathbb{R}^{n\times n}$. Hence, $P_k^{ea-} = P_\gamma + \zeta_k$ converges to P_γ from any initial condition. ■

C. Error Covariance Under Attack Using Sensing Data

Building on the obtained result, we consider Scenario II in this section. In this case, the malicious agent launches attacks on the system based on its private information, that is, the local innovation z_k^β. The error covariance recursion in the presence of attacks is summarized in the following lemma.

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

$$\begin{aligned}\tilde{P}_k^\beta &= A\tilde{P}_{k-1}^\beta A' + Q + P_\alpha C_\alpha'(C_\alpha P_\alpha C_\alpha' + R_\alpha)^{-1}C_\alpha P_\alpha \\ &\quad - P_{k,\beta}^{EA-} C_\beta' T_k^{\beta\prime} K_\alpha' - K_\alpha T_k^\beta C_\beta P_{k,\beta}^{EA-}\end{aligned} \quad (20)$$

where $P_{k,\beta}^{EA-}$ follows the recursion:

$$P_{k,\beta}^{EA-} = AP_{k-1,\beta}^{EA-}(I - K_\beta C_\beta)'A' + Q \quad (21)$$

with the initial value $P_{0,\beta}^{EA-}$ being the unique solution of

$$X = A(I - K_\alpha C_\alpha)X(I - K_\beta C_\beta)'A' + Q. \tag{22}$$

Proof: Similar to the proof of Lemma 2, the remote estimation error covariance when the system is under attack $\tilde{z}_k^\beta = T_k^\beta z_k^\beta + b_k^\beta$ can be represented as

$$\begin{aligned}\tilde{P}_k^\beta = {} & A\tilde{P}_{k-1}^\beta A' + Q + K_\alpha(C_\alpha P_\alpha C_\alpha' + R_\alpha)K_\alpha' \\ & - \mathbb{E}[(x_k - \tilde{x}_k^{\beta-})(x_k - \hat{x}_k^{\beta-})'C'T_k^{\beta\prime}K_\alpha'] \\ & - \mathbb{E}[K_\alpha T_k^\beta C(x_k - \hat{x}_k^{\beta-})(x_k - \tilde{x}_k^{\beta-})'].\end{aligned}$$

In this case, the correlation term in the presence of attack follows the iteration:

$$P_{k,\beta}^{EA-} = AP_{k-1,\beta}^{EA-}(I - K_\beta C_\beta)'A' + Q. \tag{23}$$

Note that the initial value $P_{0,\beta}^{EA-}$ of (23) is the steady-state value of the correlation in the absence of attack. According to

$$P_{k,\beta}^{ea-} = A(I - K_\alpha C_\alpha)P_{k-1,\beta}^{ea-}(I - K_\beta C_\beta)'A' + Q \tag{24}$$

it can be observed that $P_{0,\beta}^{EA-} = \lim_{k\to\infty} P_{k,\beta}^{ea-}$ is the unique solution of $X = A(I - K_\alpha C_\alpha)X(I - K_\beta C_\beta)'A' + Q$. Due to the fact that $\lim_{k\to\infty} P_{k,\beta}^{ea-} \neq P_\beta$ and $\lim_{k\to\infty} P_{k,\beta}^{EA-} = P_\beta$, where P_β is the unique solution of $X = AX(I - K_\beta C_\beta)'A' + Q$, the correlation term $P_{k,\beta}^{EA-}$ is time-varying, converging to P_β, as given in (21) and (22). ■

So far, we have obtained the error covariance iterations for three attack policies. They provide the basis for the worst-case analysis and attack consequence comparison investigated in Sections V and VI.

V. Worst-Case Attack Strategy

In this section, we investigate the worst-case attack consequence when the process (1), (2) is under attack $\tilde{z}_k^i = T_k^i z_k^i + b_k^i$, $i = \alpha, \beta, \gamma$ utilizing the error covariance iterations obtained in Section IV. The worst-case attack parameters T_k^i, L_k^i for multivariable systems ($n \geq 1$, $m \geq 1$) are obtained by solving convex optimization problems.

A. Worst-Case Linear Attack Using Intercepted Data

For Scenario I where the attack strategy is based on system innovation z_k^α, the worst-case linear attack is obtained in the closed form and the result is summarized as follows.

Lemma 4: For the process (1), (2) with the attack $\tilde{z}_k^\alpha = T_k^\alpha z_k^\alpha + b_k^\alpha$, the choice $T_k^\alpha = -I$ and $b_k^\alpha = 0$ is the worst-case strategy in the sense that the remote estimation error covariance is maximized.

Proof: See [18, Th. 2]. ■

B. Worst-Case Linear Attack Using Intercepted and Sensing Data

When the linear attack strategy is based on the combined information of the intercepted and the sensing data, namely, Scenario III, the worst-case attack policy T_k^γ, L_k^γ can be numerically calculated by maximizing the estimation error covariance (15). The obtained result is summarized in the following theorem.

Theorem 1: For the process (1), (2) with the attack $\tilde{z}_k^\gamma = T_k^\gamma z_k^\gamma + b_k^\gamma$, the worst case T_k^γ, which yields the largest remote estimation error covariance, is given by the solution to the convex optimization problem

$$\begin{aligned}\mathbf{P_1}: \min_{T_k^\gamma \in \mathbb{R}^{m\times 2m}} \quad & \operatorname{Tr}(CP_\gamma K_\alpha T_k^\gamma) \\ \text{s.t.} \quad & \begin{bmatrix} C_\alpha P_\alpha C_\alpha' + R_\alpha & T_k^\gamma \\ T_k^{\gamma\prime} & (CP_\gamma C' + R)^{-1} \end{bmatrix} \geq 0.\end{aligned}$$

The corresponding b_k^γ is a zero-mean Gaussian process with covariance $L_k^\gamma = C_\alpha P_\alpha C_\alpha' + R_\alpha - T_k^\gamma(CP_\gamma C' + R)T_k^{\gamma\prime}$.

Proof: The iteration of the estimation error covariance when the system is under attack $\tilde{z}_k^\gamma = T_k^\gamma z_k^\gamma + b_k^\gamma$ is obtained in (15). It can be observed that maximizing $\operatorname{Tr}(\tilde{P}_k^\gamma)$ is equivalent to maximizing $\operatorname{Tr}(-P_\gamma C'T_k^{\gamma\prime}K_\alpha' - K_\alpha T_k^\gamma CP_\gamma)$. According to the stealthiness constraint (11) and the fact that $\operatorname{Tr}(X) = \operatorname{Tr}(X')$ and $\operatorname{Tr}(XYZ) = \operatorname{Tr}(YZX) = \operatorname{Tr}(ZXY)$, the worst-case linear attack can be obtained by solving the optimization problem

$$\begin{aligned}\max_{T_k^\gamma \in \mathbb{R}^{m\times 2m}} \quad & -2\operatorname{Tr}(CP_\gamma K_\alpha T_k^\gamma) \\ \text{s.t.} \quad & T_k^\gamma(CP_\gamma C' + R)T_k^{\gamma\prime} - (C_\alpha P_\alpha C_\alpha' + R_\alpha) \leq 0.\end{aligned}$$

Using the Schur complement, the above constraint can be rewritten into a linear matrix inequality, which leads to problem $\mathbf{P_1}$. ■

C. Worst-Case Linear Attack Using Sensing Data

Different from Scenario III, the error covariance recursion (20) of Scenario II involves a time-varying correlation term. Hence, the worst-case linear attack strategy is obtained by dynamically solving a convex optimization problem at each time instant, which is summarized in the following theorem.

Theorem 2: For the process (1), (2) with the attack $\tilde{z}_k^\beta = T_k^\beta z_k^\beta + b_k^\beta$, the worst-case attack strategy T_k^β, which maximizes the remote estimation error covariance, is obtained by solving the following problem for each k:

$$\begin{aligned}\mathbf{P_2}: \min_{T_k^\beta \in \mathbb{R}^{m\times m}} \quad & \operatorname{Tr}(C_\beta P_{k,\beta}^{EA-} K_\alpha T_k^\beta) \\ \text{s.t.} \quad & \begin{bmatrix} C_\alpha P_\alpha C_\alpha' + R_\alpha & T_k^\beta \\ T_k^{\beta\prime} & (C_\beta P_\beta C_\beta' + R_\beta)^{-1} \end{bmatrix} \geq 0\end{aligned}$$

where $P_{k,\beta}^{EA-}$ follows the recursion:

$$P_{k,\beta}^{EA-} = AP_{k-1,\beta}^{EA-}(I - K_\beta C_\beta)'A' + Q$$

with the initial value $P_{0,\beta}^{EA-}$ being the unique solution of

$$X = A(I - K_\alpha C_\alpha)X(I - K_\beta C_\beta)'A' + Q.$$

The corresponding b_k^β is a zero-mean Gaussian with covariance $L_k^\beta = C_\alpha P_\alpha C_\alpha' + R_\alpha - T_k^\beta(C_\beta P_\beta C_\beta' + R_\beta)T_k^{\beta\prime}$.

Algorithm 1: Calculating the Optimal Attack Signal.

1: Process begins;
2: **for** $k = 1 : 1 : \infty$ **do**
3: Calculate T_k^i $(i = \beta, \gamma)$ by solving problem $\mathbf{P_1}$, $\mathbf{P_2}$
4: Calculate z_k^i, $\tilde{z}_k^i$ based on the knowledge of T_k^i, C_i, y_k^i, $\hat{x}_k^{i-}$, $\tilde{x}_k^{i-}$ according to $z_k^i = y_k^i - C_i\hat{x}_k^{i-}$, $\tilde{z}_k^i = T_k^i z_k^i + b_k^i$;
4: Calculate the worst-case linear attack strategy $\tilde{y}_k^i$ by $\tilde{y}_k^i = \tilde{z}_k^i + C_i\tilde{x}_k^{i-}$;
5: Update the prior state estimates according to $\hat{x}_{k+1}^{i-} = A(\hat{x}_k^{i-} + K_i z_k^i)$, $\tilde{x}_{k+1}^{i-} = A(\tilde{x}_k^{i-} + K_\alpha \tilde{z}_k^i)$;
6: **end for**

Proof: For each time instant k, the proof is similar to that of Theorem 1 and is omitted here. ■

Remark 1: Problems $\mathbf{P_1}$ and $\mathbf{P_2}$ are semidefinite programming problems and can be efficiently solved using the CVX toolbox in MATLAB [28]. Note that the obtained worst-case attack parameters $T_k^{\gamma *}$ and $b_k^{\gamma *}$ are time-invariant since P_γ is a constant while $T_k^{\beta *}$ and $b_k^{\beta *}$ are time-varying since $P_{k,\beta}^{EA-}$ changes with time k.

D. Calculation of Worst-Case Attack Signal

The worst possible action of the attacker on the system measurement y_k^i is given by Algorithm 1. At each time instant k, the malicious attacker first solves the optimization problem $\mathbf{P_1}$ or $\mathbf{P_2}$ based on its knowledge of the system parameters, from which the true innovation z_k^i and the corrupted innovation $\tilde{z}_k^i$ can be obtained. According to the relationship between the measurement and the innovation, the worst-case attack signal $\tilde{y}_k^i$ is obtained. Finally, the attacker updates the priori state estimates for the original and compromised processes, which it then uses in the next iteration. This characterization helps us in Section VI to explicitly compare attack consequences and reason about the need for various protection mechanisms.

VI. Worst-Case Attack for Scalar Systems

In this section, we focus on scalar systems ($n = m = 1$) with $C_\alpha \neq 0$ and $C_\beta \neq 0$. In this case, the closed-form expression of the worst-case linear attack strategy can be derived.

A. Closed-Form Expression of Worst-Case Linear Attack

According to Lemma 4, for process (1), (2) with $n = m = 1$ under linear attack $\tilde{z}_k^\alpha = T_k^\alpha z_k^\alpha + b_k^\alpha$ (Scenario I), the remote estimation error covariance is simply maximized when $T_k^\alpha = -1, b_k^\alpha = 0$. Let us focus on the worst-case analysis for the attack strategy $\tilde{z}_k^i = T_k^i z_k^i + b_k^i$, $i = \beta, \gamma$ and summarize the results in the following propositions.

Proposition 1: For a scalar process (1)–(2) with the attack $\tilde{z}_k^\gamma = T_k^\gamma z_k^\gamma + b_k^\gamma$, the worst-case strategy that maximizes the estimation error covariance is $T_k^\gamma = -\sqrt{\frac{\Delta_\alpha}{\Delta_\gamma}}\frac{K}{K_\alpha}$, $b_k^\gamma = 0$, where $\Delta_\alpha = P_\alpha C_\alpha'(C_\alpha P_\alpha C_\alpha' + R_\alpha)^{-1} C_\alpha P_\alpha$, $\Delta_\gamma = P_\gamma C'(CP_\gamma C' + R)^{-1} CP_\gamma$.

Proof: When the scalar process (1), (2) is under attack $\tilde{z}_k^\gamma = T_k^\gamma z_k^\gamma + b_k^\gamma$, the stealthiness constraint (11) becomes

$$
\begin{aligned}
&T_k^\gamma (CP_\gamma C' + R) T_k^{\gamma\prime} + L_k^\gamma \\
&\overset{(a)}{=} \begin{bmatrix} T_k^{\gamma 1} & T_k^{\gamma 2} \end{bmatrix} \begin{bmatrix} C_\alpha^2 P_\gamma + R_\alpha & C_\alpha C_\beta P_\gamma \\ C_\alpha C_\beta P_\gamma & C_\beta^2 P_\gamma + R_\beta \end{bmatrix} \begin{bmatrix} T_k^{\gamma 1} \\ T_k^{\gamma 2} \end{bmatrix} + L_k^\gamma \\
&= (T_k^{\gamma 1})^2 (C_\alpha^2 P_\gamma + R_\alpha) + \left(\frac{T_k^{\gamma 2} C_\beta}{C_\alpha}\right)^2 \left(C_\alpha^2 P_\gamma + \frac{R_\beta C_\alpha^2}{C_\beta^2}\right) \\
&\quad + 2T_k^{\gamma 1} \frac{T_k^{\gamma 2} C_\beta}{C_\alpha} C_\alpha^2 P_\alpha + L_k^\gamma \\
&\overset{(b)}{=} C_\alpha^2 P_\alpha (U_{k,1} + U_{k,2})^2 + R_\alpha U_{k,1}^2 + \frac{R_\beta C_\alpha^2}{C_\beta^2} U_{k,2}^2 + L_k^\gamma \\
&= C_\alpha^2 P_\alpha + R_\alpha \qquad (25)
\end{aligned}
$$

where equalities (a) and (b) follow from $T_k^\gamma = [T_k^{\gamma 1}, T_k^{\gamma 2}]$ and change of variables $U_{k,1} = T_k^{\gamma 1}, U_{k,2} = \frac{T_k^{\gamma 2} C_\beta}{C_\alpha}$, respectively. The remote estimation error covariance (15) becomes

$$
\begin{aligned}
\tilde{P}_k^\gamma &= A^2 \tilde{P}_{k-1}^\gamma + Q + \frac{C_\alpha^2 P_\alpha^2}{C_\alpha^2 P_\alpha + R_\alpha} - 2K_\alpha T_k^\gamma CP_\gamma \\
&= A^2 \tilde{P}_{k-1}^\gamma + Q + \Delta_\alpha - 2K_\alpha T_k^{\gamma 1} C_\alpha P_\gamma - 2K_\alpha T_k^{\gamma 2} C_\beta P_\gamma \\
&= A^2 \tilde{P}_{k-1}^\gamma + Q + \Delta_\alpha - 2K_\alpha C_\alpha P_\gamma (U_{k,1} + U_{k,2}) \qquad (26)
\end{aligned}
$$

where the second equality follows from the fact that $\Delta_\alpha = P_\alpha C_\alpha'(C_\alpha P_\alpha C_\alpha' + R_\alpha)^{-1} C_\alpha P_\alpha$, $C = [C_\alpha, C_\beta]'$ and $T_k^\gamma = [T_k^{\gamma 1}, T_k^{\gamma 2}]$. It can be observed from (25) and (26) that $U_{k,1} + U_{k,2}$ must be negative, otherwise one can always flip the sign of $U_{k,1}$ and $U_{k,2}$ such that a larger estimation error covariance is obtained under the same stealthiness constraint. Thus, the smaller the term $U_{k,1} + U_{k,2}$, the larger the error covariance $\tilde{P}_k^\gamma$. To make $U_{k,1} + U_{k,2}$ negative, at least one of the elements should be negative. Without loss of generality, we assume that $U_{k,1}$ is negative. In this case, for any $L_k^\gamma > 0, U_{k,1} < 0$ and $U_{k,2}$, which satisfy the last equality of (25), one can always reduce L_k^γ to zero and find an $\varepsilon < 0$ satisfying $C_\alpha^2 P_\alpha (U_{k,1} + U_{k,2} + \varepsilon)^2 + R_\alpha (U_{k,1} + \varepsilon)^2 + \frac{R_\beta C_\alpha^2}{C_\beta^2} U_{k,2}^2 = C_\alpha^2 P_\alpha + R_\alpha$, which leads to a larger $\tilde{P}_k^\gamma$ since $U_{k,1} + U_{k,2} + \varepsilon < U_{k,1} + U_{k,2} < 0$. Hence, the worst attack strategy that maximizes the remote estimation error covariance is achieved when $L_k^\gamma = 0$ and $T_k^\gamma (CP_\gamma C' + R) T_k^{\gamma\prime} = C_\alpha^2 P_\alpha + R_\alpha$.

According to the iteration of the remote state estimate (12) and (13), one has

$$
\begin{aligned}
\tilde{x}_k^\gamma &= A\tilde{x}_{k-1}^\gamma + K_\alpha \tilde{z}_k^\gamma \\
&= A\tilde{x}_{k-1}^\gamma + K_\alpha [T_k^\gamma + b_k^\gamma (z_k^{\gamma\prime} z_k^\gamma)^{-1} z_k^{\gamma\prime}] z_k^\gamma \\
&= A\tilde{x}_{k-1}^\gamma + \tilde{K}_k z_k^\gamma
\end{aligned}
$$

where $\tilde{K}_k = K_\alpha [T_k^\gamma + b_k^\gamma (z_k^{\gamma\prime} z_k^\gamma)^{-1} z_k^{\gamma\prime}]$. Since the largest estimation error is achieved when $L_k^\gamma = 0$, i.e., $b_k^\gamma = 0$, the worst-case linear attack strategy satisfies $\tilde{K}_k = K_\alpha T_k^\gamma$. Without loss of generality, we assume that $\tilde{K}_k = [K_\alpha T_k^{\gamma 1}, K_\alpha T_k^{\gamma 2}] =$

$[\lambda_1\bar{K}_1, \lambda_2\bar{K}_2]$, where $K \triangleq [\bar{K}_1, \bar{K}_2]$. In this case, the stealthiness constraint (25) becomes

$$\begin{aligned} &K_\alpha T_k^\gamma (CP_\gamma C' + R) T_k^{\gamma\prime} K_\alpha' \\ &= \begin{bmatrix} \lambda_1\bar{K}_1 & \lambda_2\bar{K}_2 \end{bmatrix} \begin{bmatrix} M_1 & M_{12} \\ M_{12} & M_2 \end{bmatrix} \begin{bmatrix} \lambda_1\bar{K}_1 \\ \lambda_2\bar{K}_2 \end{bmatrix} \\ &= \lambda_1^2\bar{K}_1^2 M_1 + \lambda_2^2\bar{K}_2^2 M_2 + 2\lambda_1\lambda_2\bar{K}_1\bar{K}_2 M_{12} \\ &= \Delta_\alpha \end{aligned} \tag{27}$$

where $M_1 = C_\alpha^2 P_\gamma + R_\alpha$, $M_2 = C_\beta^2 P_\gamma + R_\beta$, $M_{12} = C_\alpha C_\beta P_\gamma$. The estimation error covariance (26) becomes

$$\tilde{P}_k^\gamma = A^2 \tilde{P}_{k-1}^\gamma + Q + \Delta_\alpha - 2\lambda_1\bar{K}_1 C_\alpha P_\gamma - 2\lambda_2\bar{K}_2 C_\beta P_\gamma.$$

Thus, optimization problem $\mathbf{P_1}$ can be represented as

$$\begin{aligned} \min_{\lambda_1,\lambda_2} \quad & \lambda_1\bar{K}_1 C_\alpha P_\gamma + \lambda_2\bar{K}_2 C_\beta P_\gamma \\ \text{s.t.} \quad & \lambda_1^2\bar{K}_1^2 M_1 + \lambda_2^2\bar{K}_2^2 M_2 + 2\lambda_1\lambda_2\bar{K}_1\bar{K}_2 M_{12} = \Delta_\alpha \end{aligned} \tag{28}$$

based on which we define the Lagrangian as

$$\begin{aligned} \mathcal{L}_p = &\lambda_1\bar{K}_1 C_\alpha P_\gamma + \lambda_2\bar{K}_2 C_\beta P_\gamma \\ &+ \mu(\lambda_1^2\bar{K}_1^2 M_1 + \lambda_2^2\bar{K}_2^2 M_2 + 2\lambda_1\lambda_2\bar{K}_1\bar{K}_2 M_{12} - \Delta_\alpha) \end{aligned}$$

where μ is the Lagrangian multiplier. Set the derivative of $\mathcal{L}_p$ with respect to λ_1 and λ_2 equal to zero

$$\begin{aligned} \frac{\partial\mathcal{L}_p}{\partial\lambda_1} &= \bar{K}_1 C_\alpha P_\gamma + 2\mu(\lambda_1\bar{K}_1^2 M_1 + \lambda_2\bar{K}_1\bar{K}_2 M_{12}) = 0 \\ \frac{\partial\mathcal{L}_p}{\partial\lambda_2} &= \bar{K}_2 C_\beta P_\gamma + 2\mu(\lambda_2\bar{K}_2^2 M_2 + \lambda_1\bar{K}_1\bar{K}_2 M_{12}) = 0. \end{aligned}$$

It can be observed that $\mu \neq 0$, otherwise one has $\bar{K}_1 C_\alpha P_\gamma = P_\gamma^2 C_\alpha^2 / R_\alpha = 0$, which contradicts with the fact that $P_\gamma > 0$, $C_\alpha \neq 0$, and $R_\alpha > 0$. Multiplying $\bar{K}_2 C_\beta$ and $\bar{K}_1 C_\alpha$ to above two equations, respectively, and subtracting gives

$$\begin{aligned} &C_\beta(\lambda_1\bar{K}_1 M_1 + \lambda_2\bar{K}_2 M_{12}) = C_\alpha(\lambda_2\bar{K}_2 M_2 + \lambda_1\bar{K}_1 M_{12}) \\ &\Leftrightarrow \lambda_1\bar{K}_1(C_\beta M_1 - C_\alpha M_{12}) = \lambda_2\bar{K}_2(C_\alpha M_2 - C_\beta M_{12}) \\ &\Leftrightarrow \lambda_1\bar{K}_1 C_\beta R_\alpha = \lambda_2\bar{K}_2 C_\alpha R_\beta \\ &\Leftrightarrow \lambda_1 = \lambda_2 \triangleq \lambda \end{aligned} \tag{29}$$

where the last equivalence is due to that $\bar{K}_1 = (I - KC)P_\gamma C_\alpha / R_\alpha$ and $\bar{K}_2 = (I - KC)P_\gamma C_\beta / R_\beta$. Hence, the largest estimation error covariance is achieved when $\tilde{K}_k = K_\alpha T_k^\gamma = \lambda K$. The stealthiness constraint (27) then becomes $\lambda^2\Delta_\gamma = \Delta_\alpha$. Therefore, the worst-case linear attack strategy at time k is $T_k^\gamma = -\sqrt{\frac{\Delta_\alpha}{\Delta_\gamma}}\frac{K}{K_\alpha}$ and $b_k^\gamma = 0$. ■

Remark 2: According to Proposition 1, the worst-case linear attack strategy for Scenario III is achieved by simply flipping the sign of the innovation z_k^γ calculated by the malicious attacker and multiplying by the constant $\sqrt{\frac{\Delta_\alpha}{\Delta_\gamma}}\frac{K}{K_\alpha}$.

We now investigate the worst-case linear attack policy when using sensing information only.

Proposition 2: For the scalar process (1), (2) with the attack $\tilde{z}_k^\beta = T_k^\beta z_k^\beta + b_k^\beta$, the worst-case strategy that maximizes the estimation error covariance is $T_k^\beta = -\sqrt{\frac{\Delta_\alpha}{\Delta_k^\beta}}\frac{K_k^\beta}{K_\alpha}$, $b_k^\beta = 0$, where $K_k^\beta = P_{k,\beta}^{EA-} C_\beta'(C_\beta P_\beta C_\beta' + R_\beta)^{-1}$, $\Delta_\alpha = P_\alpha C_\alpha'(C_\alpha P_\alpha C_\alpha' + R_\alpha)^{-1} C_\alpha P_\alpha$, $\Delta_k^\beta = P_{k,\beta}^{EA-} C_\beta'(C_\beta P_\beta C_\beta' + R_\beta)^{-1} C_\beta P_{k,\beta}^{EA-}$, and $P_{k,\beta}^{EA-}$ follows the recursion

$$P_{k,\beta}^{EA-} = AP_{k-1,\beta}^{EA-}(I - K_\beta C_\beta)'A' + Q \tag{30}$$

with the initial value $P_{0,\beta}^{EA-}$ being the unique solution of

$$X = A(I - K_\alpha C_\alpha)X(I - K_\beta C_\beta)'A' + Q. \tag{31}$$

Proof: The proof is similar to Proposition 1 and is omitted here. ■

Remark 3: The worst-case linear attack strategy for Scenario II is to flip the sign of the innovation z_k^β at each time instant and multiply by a time-varying coefficient $\sqrt{\frac{\Delta_\alpha}{\Delta_k^\beta}}\frac{K_k^\beta}{K_\alpha}$. This time variation is the main difference to Scenario III.

B. Strategy Comparison

In this section, we compare the system estimation performance under the worst-case attack policies for Scenarios I–III. We first introduce a preliminary lemma needed for the subsequent derivation.

Lemma 5: For scalar processes, the steady-state error covariances P_α, P_β, P_γ, and $P_{0,\beta}^{EA-}$ have the following relationship:

1) $P_\alpha \geq P_\beta \geq P_\gamma, P_\alpha \geq P_\beta \geq P_{0,\beta}^{EA-}$, if $C_\alpha^2/R_\alpha \leq C_\beta^2/R_\beta$;
2) $P_\beta \geq P_\alpha \geq P_\gamma, P_\beta \geq P_\alpha \geq P_{0,\beta}^{EA-}$, if $C_\alpha^2/R_\alpha \geq C_\beta^2/R_\beta$.

Proof: To show $P_\alpha \geq P_\gamma$ and $P_\beta \geq P_\gamma$, we first prove that $C_\alpha^2(C_\alpha^2 X + R_\alpha)^{-1} \leq C'(CXC' + R)^{-1}C$ for any $X \geq 0$, where $C = [C_\alpha, C_\beta]'$, $R = \mathrm{Diag}\{R_\alpha, R_\beta\}$. Note that

$$C'(CXC' + R)^{-1}C = C' \begin{bmatrix} W_{11} & W_{12} \\ W_{12} & W_{22} \end{bmatrix}^{-1} C$$

where $W_{11} = C_\alpha^2 P_\gamma + R_\alpha > 0$, $W_{12} = C_\alpha C_\beta P_\gamma$, and $W_{22} = C_\beta^2 P_\gamma + R_\beta > 0$. The Schur complement of W_{11} is denoted $S \triangleq W_{22} - W_{12}^2 W_{11}^{-1} > 0$, based on which we obtain

$$\begin{aligned} &C'(CXC' + R)^{-1}C \\ &= C' \begin{bmatrix} W_{11}^{-1} + W_{11}^{-2}W_{12}^2 S^{-1} & -W_{11}^{-1}W_{12}S^{-1} \\ -W_{11}^{-1}W_{12}S^{-1} & S^{-1} \end{bmatrix} C \\ &= C_\alpha^2 W_{11}^{-1} + (C_\alpha W_{11}^{-1} W_{12} S^{-\frac{1}{2}} - C_\beta S^{-\frac{1}{2}})^2 \\ &\geq C_\alpha^2 W_{11}^{-1} = C_\alpha^2(C_\alpha^2 X + R_\alpha)^{-1}. \end{aligned}$$

Then, it is easy to obtain $P_k^{\alpha-} \geq P_k^{\gamma-}$ for any initial condition by induction, i.e., $P_\alpha \geq P_\gamma$. Similarly, one has $P_\beta \geq P_\gamma$.

For the case where $C_\alpha^2/R_\alpha \leq C_\beta^2/R_\beta$, we now prove that $P_\alpha \geq P_\beta \geq P_{0,\beta}^{EA-}$. It is worth noticing that $P_k^{\alpha-}$, $P_k^{\beta-}$, and $P_{k,\beta}^{ea-}$ converges to its steady-state value P_α, P_β, and $P_{0,\beta}^{EA-}$ from any the initial condition [20]. Without loss of generality,

we assume that the error covariance P_k^{i-}, $i=\alpha,\beta$ and $P_{k,\beta}^{ea-}$ evolves from the same initial point. According to

$$\begin{aligned} P_k^{i-} &= A^2(1-K_k^iC_i)P_{k-1}^{i-}+Q \\ &= A^2\frac{1}{\frac{C_i^2}{R_i}P_{k-1}^{i-}+1}P_{k-1}^{i-}+Q \end{aligned}$$

it can be observed that the larger C_i^2/R_i, the smaller P_k^{i-}. Hence, one has $P_\alpha \geq P_\beta$, if $C_\alpha^2/R_\alpha \leq C_\beta^2/R_\beta$. Then, due to the fact that $0<1-K_k^iC_i=\frac{R_i}{C_i^2P_k^{i-}+R_i}\leq 1\ \forall i=\alpha,\beta$, and the recursion

$$\begin{aligned} P_k^{i-} &= A^2(1-K_k^iC_i)P_{k-1}^{i-}+Q,\ \ i=\alpha,\beta \\ P_{k,\beta}^{ea-} &= A^2(I-K_k^\alpha C_\alpha)(I-K_k^\beta C_\beta)P_{k-1,\beta}^{ea-}+Q \end{aligned}$$

we obtain that $P_\alpha \geq P_{0,\beta}^{EA-}$ and $P_\beta \geq P_{0,\beta}^{EA-}$. For the case $C_\alpha^2/R_\alpha \geq C_\beta^2/R_\beta$, the proof is similar and omitted here. ■

Remark 4: As we know, a smaller sensor noise level leads to a more accurate state estimate and more information contributes to a smaller estimation error. The results obtained in Lemma 5 are consistent with these intuitions.

Recall the worst-case attack strategies $\tilde{z}_k^\gamma=T_k^\gamma z_k^\gamma+b_k^\gamma$, $\tilde{z}_k^\beta=T_k^\beta z_k^\beta+b_k^\beta$ and $\tilde{z}_k^\alpha=T_k^\alpha z_k^\alpha+b_k^\alpha$ with T_k^γ, b_k^γ given in Proposition 1, T_k^β, b_k^β given in Proposition 2, and T_k^α, b_k^α given in Lemma 4. We now compare their consequences between above scenarios in the following two corollaries.

Corollary 1: For a scalar process (1), (2), the worst-case error covariance at the remote estimator under attack $\tilde{z}_k^\gamma=T_k^\gamma z_k^\gamma+b_k^\gamma$ is

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

Proof: The worst-case error covariance iteration at the remote estimator follows

$$\tilde{P}_k^\gamma = A^2\tilde{P}_{k-1}^\gamma+Q+\Delta_\alpha+2|\lambda|\Delta_\gamma \tag{32}$$

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

$$\tilde{P}_k^\alpha = A^2\tilde{P}_{k-1}^\alpha+Q+3\Delta_\alpha \tag{33}$$

under attack $\tilde{z}_k^\alpha=T_k^\alpha z_k^\alpha+b_k^\alpha$. Since the initial conditions of (32) and (33) are the same, the relationship between $\tilde{P}_k^\gamma$ and $\tilde{P}_k^\alpha$ depends on the magnitudes of $|\lambda|\Delta_\gamma$ and Δ_α. Hence, we now focus on comparing these two terms, where λ is defined in (29). Note that P_α and P_γ are the unique solutions of the algebraic Riccati equations

$$\begin{aligned} P_\alpha &= A^2P_\alpha+Q-A^2\Delta_\alpha \\ P_\gamma &= A^2P_\gamma+Q-A^2\Delta_\gamma \end{aligned}$$

based on which one has

$$(1-A^2)(P_\alpha-P_\gamma)=A^2(\Delta_\gamma-\Delta_\alpha) \tag{34}$$

with $P_\alpha \geq P_\gamma$. It can be observed from (34) that $\Delta_\gamma>\Delta_\alpha$ if $|A|<1$, $\Delta_\gamma<\Delta_\alpha$ if $|A|>1$, and $\Delta_\gamma=\Delta_\alpha$ if $|A|=1$. According to the stealthiness constraint $\lambda^2\Delta_\gamma=\Delta_\alpha$, it can further be obtained that $|\lambda|=\sqrt{\frac{\Delta_\alpha}{\Delta_\gamma}}<1$ if $|A|<1$, $|\lambda|=\sqrt{\frac{\Delta_\alpha}{\Delta_\gamma}}>1$ if $|A|>1$, and $|\lambda|=\sqrt{\frac{\Delta_\alpha}{\Delta_\gamma}}=1$ if $|A|=1$. Then, dividing both sides of $\lambda^2\Delta_\gamma=\Delta_\alpha$ by λ leads to the results $|\lambda|\Delta_\gamma>\Delta_\alpha$ if $|A|<1$, $|\lambda|\Delta_\gamma<\Delta_\alpha$ if $|A|>1$, and $|\lambda|\Delta_\gamma=\Delta_\alpha$ if $|A|=1$. ■

Corollary 2: For a scalar process (1), (2) with $|A|<1$, the steady-state worst-case error covariance at the remote estimator under attack strategy $\tilde{z}_k^\beta=T_k^\beta z_k^\beta+b_k^\beta$ is

1) larger than that under attack $\tilde{z}_k^\alpha=T_k^\alpha z_k^\alpha+b_k^\alpha$ but smaller than that under attack $\tilde{z}_k^\gamma=T_k^\gamma z_k^\gamma+b_k^\gamma$ if $C_\alpha^2/R_\alpha<C_\beta^2/R_\beta$;
2) smaller than that under attacks $\tilde{z}_k^\alpha=T_k^\alpha z_k^\alpha+b_k^\alpha$ and $\tilde{z}_k^\gamma=T_k^\gamma z_k^\gamma+b_k^\gamma$ if $C_\alpha^2/R_\alpha>C_\beta^2/R_\beta$;
3) equal to that under attack $\tilde{z}_k^\alpha=T_k^\alpha z_k^\alpha+b_k^\alpha$ and smaller than that under attack $\tilde{z}_k^\gamma=T_k^\gamma z_k^\gamma+b_k^\gamma$ if $C_\alpha^2/R_\alpha=C_\beta^2/R_\beta$.

Proof: According to Lemma 3 and Lemma 5, one has $\lim_{k\to\infty}P_k^{EA-}=P_\beta>P_k^{EA-}$, based on which the steady-state error covariance iteration of $\tilde{P}_k^\beta$ for stable systems with $|A|<1$ follows

$$\begin{aligned} \tilde{P}_k^\beta &= A\tilde{P}_{k-1}^\beta A'+Q+\Delta_\alpha-P_\beta C_\beta' T_k^{\beta\prime}K_\alpha'-K_\alpha T_k^\beta C_\beta P_\beta \\ &= A^2\tilde{P}_{k-1}^\beta+Q+\Delta_\alpha+2|\eta|\Delta_\beta. \end{aligned} \tag{35}$$

When $C_\alpha^2/R_\alpha<C_\beta^2/R_\beta$, one has $P_\alpha>P_\beta>P_\gamma$ and $\Delta_\gamma>\Delta_\beta>\Delta_\alpha$. Due to the worst-case stealthiness constraints $\eta^2\Delta_\beta=\Delta_\alpha$ and $\lambda^2\Delta_\gamma=\Delta_\alpha$, it can be obtained that $|\lambda|\Delta_\gamma>|\eta|\Delta_\beta>\Delta_\alpha$, which results in $\tilde{P}_k^\gamma>\tilde{P}_k^\beta>\tilde{P}_k^\alpha$ in the steady-state value. Similarly, for the case that $C_\alpha^2/R_\alpha>C_\beta^2/R_\beta$, one has $P_\beta>P_\alpha>P_\gamma$, which leads to $\Delta_\gamma>\Delta_\alpha>\Delta_\beta$. According to the stealthiness constraint, finally we obtain the steady-state error covariance $\tilde{P}_k^\gamma>\tilde{P}_k^\alpha>\tilde{P}_k^\beta$. The proof of the last case is similar. ■

Remark 5: Practically, the attacker might consider all three attack scenarios simultaneously and choose the policy which yields the largest estimation error covariance. For example, according to Corollary 1, the attacker will launch an attack based on z_k^α rather than z_k^γ when the system is unstable.

VII. Discussion on Mitigation Strategies

The previous sections focus on the worst-case attack analysis. In this section, we will discuss possible countermeasures. In particular, we consider mitigation strategies for linear attacks from three aspects, adopted from the literature.

One efficient methodology to authenticate the correct operation of a control system under replay attacks was first proposed in [2] and [11]. In this case, to detect the existence of a malicious attack, a Gaussian "watermark" signal was added to the control input. The system sacrifices control performance to increase the detection probability of the attack. A detection scheme based on such an authentication signal can be adopted in our case. Suppose that the sensor adds a random authentication signal to the transmitted innovation. Meanwhile, the remote estimator generates this signal using the same seed and subtracts it from the received innovation. Under this scheme, the system will not be affected by the additive authentication signal in the absence of attacks. However, in the presence of attacks, the alarm rate at the false-data detector will increase and the estimation error

covariance will be larger compared with that in the absence of attacks. Hence, the authentication signal can be carefully designed to tradeoff the system performance with the security level.

Second, if a multisensor system is considered, it is possible to design countermeasures without degrading system performance since more information is available for the mitigation design. Detections against linear integrity attacks were investigated for multisensor systems, with only a portion of sensors compromised, in [29]. Although the corrupted data preserve the same statistical features as the original ones, they cannot successfully bypass the false-data detector designed based on the information extracted from the trusted sensors and the correlations among the sensors. Moreover, the concept of transfer entropy in information theory was utilized for anomaly detection in multisensor systems in [30]. The causal relationship between system variables is reflected by the transfer entropy and the change of the entropy implies the existence of a malicious attack.

Third, instead of detecting the malicious attack, it is also meaningful to design new detection algorithms or data fusion schemes to make the system more robust to cyber attacks. To achieve this goal, a stochastic χ^2 false-data detector with a random threshold was proposed in [31], under which the remote estimator determines whether to fuse the received data or not based on the data importance. Note that such a detection method do not check the exact existence of the attacks, but efficiently increases the robustness of the system.

VIII. Simulation Example

To demonstrate the aforementioned results, we provide some numerical simulations in this section. We first consider a stable process with parameters

$$A=\begin{bmatrix}0.8 & 0.6 & 0\\ 0 & 0.5 & 0.3\\ 0 & 0 & 0.7\end{bmatrix}, C_\alpha=\begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}, C_\beta=\begin{bmatrix}1 & 0 & 1\\ 0 & 1 & 0\end{bmatrix}.$$

The process and measurement noise covariances are $Q=\mathrm{Diag}\{0.8,1.2,0.5\}$, $R_\alpha=\mathrm{Diag}\{2,1.2\}$, and $R_\beta=\mathrm{Diag}\{0.8,0.5\}$. The normalized error covariance at the remote estimator when the system is under worst case and randomly generated linear attack strategies is shown in Fig. 2. During time period $[0,34]$, the system has entered steady state. The malicious attack starts at $k=35$. It can be observed that the estimation error in the presence of the worst-case linear attack is larger than when the system is under randomly generated attack for all the attack scenarios (Scenarios I–III). Moreover, the worst attack consequences can be compared numerically by solving optimization problems $\mathbf{P_1}$ and $\mathbf{P_2}$. For the considered system, the malicious attack based on both the intercepted and sensing data is more critical than that based on the sensing data only, and the latter is more severe than that based on the intercepted data only. This agrees with the theory developed in Section V.

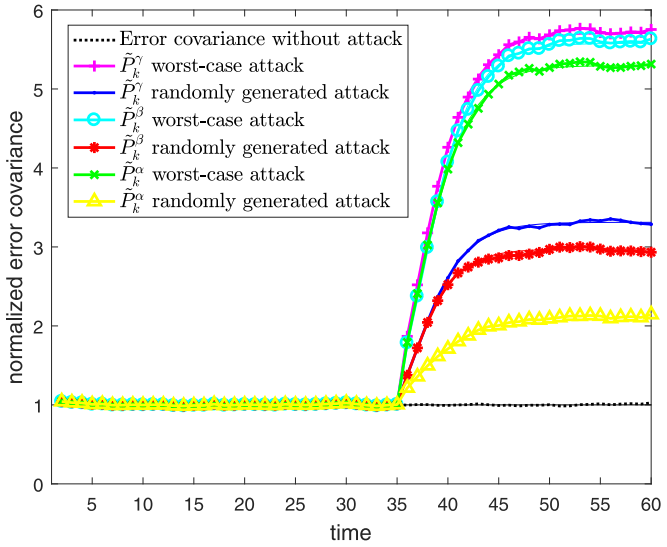


Fig. 2. Normalized estimation error covariances when a stable system is under worst case and randomly generated linear attacks for Scenarios I–III.

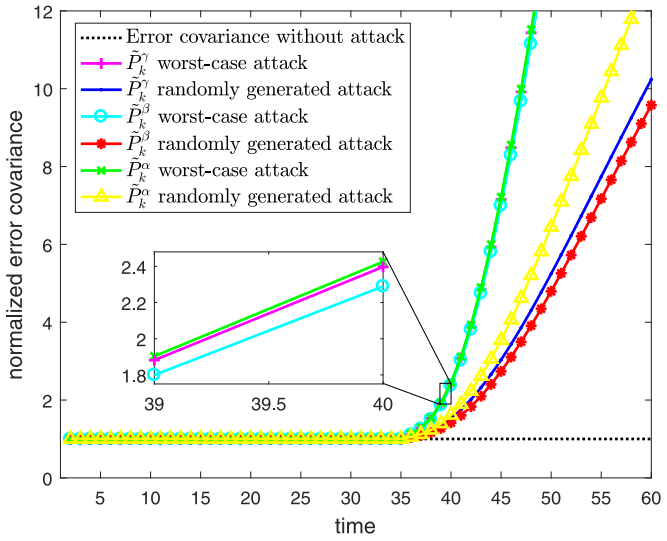


Fig. 3. Normalized estimation error covariances when an unstable system is under worst case and randomly generated linear attacks for Scenarios I–III.

We also consider an unstable process with

$$A=\begin{bmatrix}1+\epsilon & 0.6 & 0\\ 0 & 0.9 & 0.3\\ 0 & 0 & 0.7\end{bmatrix}$$

where ϵ is the floating point relative accuracy in MATLAB. Other system parameters are the same as above. The simulation result is shown in Fig. 3. In this case, although the attack based on the intercepted data yields the worst estimation performance, the error covariances under all worst-case attacks diverge exponentially fast, which correspond to the green, magenta, and cyan lines shown in the zoomed plot of Fig. 3.

To demonstrate the closed-form comparison results in Section VI, we now consider scalar processes with parameters

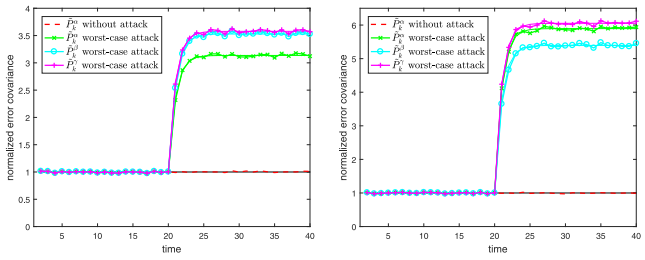


Fig. 4. Normalized remote estimation error covariances under worst case linear attack strategies. (a) $|A| < 1$ and $C_\alpha^2/R_\alpha < C_\beta^2/R_\beta$; (b) $|A| < 1$ and $C_\alpha^2/R_\alpha > C_\beta^2/R_\beta$.

$A = 0.6$, $Q = 0.5$, $C_\alpha = 1$, $C_\beta = 1$, $R_\alpha = 2$, $R_\beta = 0.5$ and $A = 0.6$, $Q = 0.5$, $C_\alpha = 1$, $C_\beta = 1$, $R_\alpha = 0.8$, $R_\beta = 2$. The simulation results are shown in Fig. 4. The malicious attacks are injected to the system from the steady state at time instant $k = 20$. The magenta plus, cyan circle, green x-mark, and red dashed lines represent the normalized estimation error covariances under the worst-case attack using intercepted and sensing data, the worst-case attack using sensing data only, the worst-case attack using intercepted data only, and without attack, respectively. Note that $C_\alpha^2/R_\alpha = 0.5 < C_\beta^2/R_\beta = 2$ for the first considered process. According to Corollary 2, the estimation error covariance of Scenario II should be larger than that of Scenario I while smaller than that of Scenario III, which is consistent with the results observed in Fig 4(a). For the second process with $C_\alpha^2/R_\alpha = 1.25 > C_\beta^2/R_\beta = 0.5$, it can be observed from Fig. 4(b) that the error covariance of Scenario III is larger than that of Scenario I and the latter is larger than that of Scenario II, which confirms the results obtained in both Corollary 1 and Corollary 2.

IX. Conclusion

In this paper, the worst-case consequences for three innovation-based integrity attacks were analyzed. We considered scenarios where the attack strategy is designed based on the intercepted data, the sensing data, or both of them. We investigated the remote estimation error covariance evolutions in the presence of the proposed attacks, based on which the worst-case attack policies were obtained by solving convex optimization problems. Furthermore, we derived closed-form expressions of the worst-case linear attacks for scalar systems. The attack consequences were compared to determine which attack leads to worse estimation performance. Simulation examples were provided to demonstrate the analytical results. Future work includes development of detection mechanisms and efficient mitigation schemes for the proposed attacks.

References

[1] K. Kim and P. R. Kumar, "Cyber–physical systems: A perspective at the centennial," *Proc. IEEE*, vol. 100, no. Special Centennial Issue, pp. 1287–1308, May 2012.

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

[3] J. P. Farwell and R. Rohozinski, "Stuxnet and the future of cyber war," *Survival*, vol. 53, no. 1, pp. 23–40, 2011.

[4] J. Slay and M. Miller, *Lessons Learned From the Maroochy Water Breach*. New York, NY, USA: Springer-Verlag, 2007.

[5] S. H. Ahmed, G. Kim, and D. Kim, "Cyber physical system: Architecture, applications and research challenges," in *Proc. IFIP Wireless Days*, 2013, pp. 1–5.

[6] D. Shi, R. J. Elliott, and T. Chen, "On finite-state stochastic modeling and secure estimation of cyber-physical systems," *IEEE Trans. Autom. Control*, vol. 62, no. 1, pp. 65–80, Jan. 2017.

[7] 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.

[8] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Trans. Inf. Syst. Security*, vol. 14, no. 1, 2011, Art. no. 13.

[9] Y. Mo, E. Garone, A. Casavola, and B. Sinopoli, "False data injection attacks against state estimation in wireless sensor networks," in *Proc. 49th IEEE Conf. Decis. Control*, 2010, pp. 5967–5972.

[10] E. Kung, S. Dey, and L. Shi, "The performance and limitations epsilon-stealthy attacks on higher order systems," *IEEE Trans. Autom. Control*, vol. 62, no. 2, pp. 941–947, Feb. 2017.

[11] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *Proc. 47th Annu. Allerton Conf. Commun., Control, Comput.*, 2009, pp. 911–918.

[12] F. Miao, M. Pajic, and G. J. Pappas, "Stochastic game approach for replay attack detection," in *Proc. 52nd IEEE Conf. Decis. Control*, 2013, pp. 1854–1859.

[13] S. Amin, A. A. Cárdenas, and S. S. Sastry, "Safe and secure networked control systems under denial-of-service attacks," in *Proc. Hybrid Syst., Comput. Control*, 2009, pp. 31–45.

[14] H. Zhang, P. Cheng, L. Shi, and J. Chen, "Optimal denial-of-service attack scheduling with energy constraint," *IEEE Trans. Autom. Control*, vol. 60, no. 11, pp. 3023–3028, Nov. 2015.

[15] H. Zhang, P. Cheng, L. Shi, and J. Chen, "Optimal DoS attack scheduling in wireless networked control system," *IEEE Trans. Control Syst. Technol.*, vol. 24, no. 3, pp. 843–852, May 2016.

[16] Y. Li, L. Shi, P. Cheng, J. Chen, and D. E. Quevedo, "Jamming attacks on remote state estimation in cyber-physical systems: A game-theoretic approach," *IEEE Trans. Autom. Control*, vol. 60, no. 10, pp. 2831–2836, Oct. 2015.

[17] Y. Li, D. E. Quevedo, S. Dey, and L. Shi, "SINR-based DoS attack on remote state estimation: A game-theoretic approach," *IEEE Trans. Control Netw. Syst.*, vol. 4, no. 3, pp. 632–642, Sep. 2017.

[18] Z. Guo, D. Shi, K. H. Johansson, and L. Shi, "Optimal linear cyber-attack on remote state estimation," *IEEE Trans. Control Netw. Syst.*, vol. 4, no. 1, pp. 4–13, Mar. 2017.

[19] Z. Guo, D. Shi, K. H. Johansson, and L. Shi, "Consequence analysis of innovation-based integrity attacks with side information on remote state estimation," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 8399–8404, 2017.

[20] B. D. Anderson and J. B. Moore, *Optimal Filtering*. North Chelmsford, MA, USA: Courier Corp., 2012.

[21] R. L. Mason and J. C. Young, *Multivariate Statistical Process Control With Industrial Applications*. Philadelphia, PA, USA: SIAM, 2002.

[22] A. Pouliezos and G. S. Stavrakakis, *Real Time Fault Monitoring of Industrial Processes*, vol. 12. New York, NY, USA: Springer-Verlag, 2013.

[23] U. Meyer and S. Wetzel, "A man-in-the-middle attack on UMTS," in *Proc. 3rd ACM Workshop Wireless Security*, 2004, pp. 90–97.

[24] F. Callegati, W. Cerroni, and M. Ramilli, "Man-in-the-middle attack to the HTTPS protocol," *IEEE Security Privacy*, vol. 7, no. 1, pp. 78–81, Jan./Feb. 2009.

[25] Z. Guo, D. Shi, K. H. Johansson, and L. Shi, "Worst-case stealthy innovation-based linear attack on remote state estimation," *Automatica*, vol. 89, pp. 117–124, 2018.

[26] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York, NY, USA: Springer-Verlag, 2013.

[27] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2013.

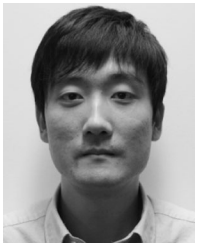
[28] M. Grant, S. Boyd, V. Blondel, and H. Kimura, "CVX: Matlab software for disciplined convex programming, version 2.0," CVX Res., Inc., 2011.

[29] Y. Li, L. Shi, and T. Chen, "Detection against linear deception attacks on multi-sensor remote state estimation," *IEEE Trans. Control Netw. Syst.*, to be published, 2017, doi: 10.1109/TCNS.2017.2648508.

[30] D. Shi, Z. Guo, K. H. Johansson, and L. Shi, "Causality countermeasures for anomaly detection in cyber-physical systems," *IEEE Trans. Autom. Control*, vol. 63, no. 2, pp. 386–401, 2018.
[31] Y. Li and T. Chen, "Stochastic detector against linear deception attacks on remote state estimation," in *Proc. 55th IEEE Conf. Decis. Control*, 2016, pp. 6291–6296.

Ziyang Guo received the B.Eng. degree (Hons.) from the College of Control Science and Engineering, Zhejiang University, Hangzhou, China, in 2014. She is currently working toward the Ph.D. degree in electronic and computer engineering at the Hong Kong University of Science and Technology, Hong Kong.

Her research interests include cyber-physical system security, state estimation, and networked control systems.



Dawei Shi received the B.Eng. degree in electrical engineering and automation from the Beijing Institute of Technology, Beijing, China, in 2008, and the Ph.D. degree in control systems from the University of Alberta, Edmonton, AB, Canada, in 2014.

In 2014, he was appointed as an Associate Professor with the School of Automation, Beijing Institute of Technology, Beijing, China. Since 2017, he has been with the John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA, as a Postdoctoral Fellow. His research interests include event-based control and estimation, robust model predictive control and tuning, and wireless-sensor networks.

Dr. Shi received the Best Student Paper Award in the IEEE International Conference on Automation and Logistics in 2009. He is a Reviewer for a number of international journals, including the IEEE TRANSACTIONS ON AUTOMATIC CONTROL and *Automatica*, and *Systems and Control Letters*.

Karl Henrik Johansson (F'13) received the M.Sc. and Ph.D. degrees in electrical engineering from Lund University, Lund, Sweden, in 1992 and 1997, respectively.

He is the Director of the Stockholm Strategic Research Area ICT The Next Generation and a Professor with the School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm, Sweden. He has held visiting positions at the UC Berkeley, Caltech, NTU, HKUST Institute of Advanced Studies, and NTNU. His research interests include networked control systems, cyber-physical systems, and applications in transportation, energy, and automation.

Dr. Johansson is a member of the IEEE Control Systems Society Board of Governors, the IFAC Executive Board, and the European Control Association Council. He has received several best paper awards and other distinctions. He is a Distinguished Professor with the Swedish Research Council and a Wallenberg Scholar and he has received the Future Research Leader Award from the Swedish Foundation for Strategic Research and the triennial Young Author Prize from IFAC. He is a Fellow of the Royal Swedish Academy of Engineering Sciences and is also an IEEE Distinguished Lecturer.

Ling Shi (SM'17) received the B.S. degree in electrical and electronic engineering from Hong Kong University of Science and Technology, Kowloon, Hong Kong, in 2002, and the Ph.D. degree in control and dynamical systems from the California Institute of Technology, Pasadena, CA, USA, in 2008.

He is currently an Associate Professor with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology. His research interests include cyber-physical systems security, networked control systems, sensor scheduling, and event-based state estimation.

Dr. Shi was an editorial board member for The European Control Conference 2013–2016. He has been serving as a Subject Editor for the *International Journal of Robust and Nonlinear Control* since 2015, an Associate Editor for the IEEE TRANSACTIONS ON CONTROL OF NETWORK SYSTEMS since 2016, and an Associate Editor for the IEEE CONTROL SYSTEMS LETTERS since 2017. He also served as an Associate Editor for a special issue on Secure Control of Cyber Physical Systems in the IEEE TRANSACTIONS ON CONTROL OF NETWORK SYSTEMS in 2015–2017. He serves as the General Chair of the 23rd International Symposium on Mathematical Theory of Networks and Systems.