2024 IEEE 63rd Conference on Decision and Control (CDC)
December 16-19, 2024. MiCo, Milan, Italy

Optimal Linear Deception Attacks on Remote State Estimation with Constrained Alarm Rates: A Low-Dimensional Case

Jun Shang, Hanwen Zhang, Jing Zhou, and Tongwen Chen

Abstract— **This study addresses linear attacks on remote state estimation within the context of a constrained alarm rate. Smart sensors, which are equipped with local Kalman filters, transmit innovations instead of raw measurements through a wireless communication network. This transmission is vulnerable to malicious data interception and manipulation by attackers. The aim of this research is to identify the optimal attack strategy that degrades the system performance while adhering to stealthiness constraints. A notable innovation of this paper is the direct association of the attack's stealthiness with the alarm rate, diverging from traditional approaches that rely on the covariance of the innovation or the Kullback–Leibler divergence, which are conventional metrics that have been extensively explored in previous studies. Our findings reveal that the optimal attack strategy exhibits some structural characteristics in systems of low dimensions. The performance of the proposed attack strategy is demonstrated through numerical examples.**

I. Introduction

The advancement of modern industrial applications has significantly encouraged the use of wireless network technologies. However, these advancements also introduce new challenges related to cybersecurity threats [1]–[3]. Among these threats, denial-of-service (DoS) attacks [4] and false data injection attacks [5] are particularly noteworthy. In certain scenarios, malicious entities have the capability to intercept and alter data packets during transmission, aiming to impair the performance of systems. Specifically, the manipulation of data to maximize performance degradation while evading detection by anomaly detectors is termed an optimal deception attack. This concept has attracted significant research interest over the past decade.

Deception attacks on remote state estimation have garnered significant attention in recent years. In scenarios where sensors measure system states and transmit packets to a remote end, attackers can cleverly alter the transmitted data to mislead Kalman filters into making suboptimal estimations. A landmark development was the introduction of the optimal innovation-based linear attack, characterized by simply inverting the sign of nominal innovations [6]. This approach has been expanded to various contexts, including scenarios where attackers utilize additional sensors to acquire side information about system states [7], [8], employ historical information [9], [10], and launch attacks with relaxed stealthiness constraints [11], [12], as well as on event-based estimation [13] and using nonlinear mappings [14]. Despite recent discoveries that the optimal information-based attack is an affine function of the minimum mean square error (MMSE) estimate of prediction errors [15], linear attacks continue to be a focal point of interest due to their simplicity and effectiveness in evading whiteness detectors [16], [17].

Stealthy attacks are characterized by varying definitions based on the approach to stealthiness constraint formulation, typically categorized into deterministic and stochastic frameworks. Deterministic stealthiness involves measuring the discrepancy between the post-attack and nominal system outputs, aiming to keep this difference within a predefined boundary. The stochastic approach, on the other hand, focuses on analyzing the residual sequence from the compromised system. In the context of Kalman filters, maintaining the unchanged probability density function (pdf) of the innovation (residual) sequence is considered the most stringent form of stealthiness constraint, a principle employed in [6]. Maintaining the unchanged pdf of the innovation sequence, given that the autocorrelation of the nominal innovation sequence is zero, equates to preserving both the unchanged pdf of each innovation and the sequence's whiteness property. Although whiteness was not emphasized in studies like [6], [7], [11], [12], [14], the inherent innovation-based nature of their attacks ensures the generation of white residuals. Conversely, strategies leveraging historical information inherently produce auto-correlated residuals, as seen in [10], [15], [18], [19].

Despite extensive research on innovation-based attacks with various stealthiness constraints, the question of the optimal linear attack that preserves the alarm rate remains unanswered. The approach in [6] maintains the alarm rate using a highly strict stealthiness constraint, which is a sufficient but not a necessary condition for keeping the alarm rate unchanged. Current literature lacks comprehensive solutions to this problem. This paper aims to bridge this gap by exploring the optimal linear attack strategy within the context of a constrained alarm rate, especially in low-

J. Shang's work was supported by the National Natural Science Foundation of China under Grant 62303353, and Shanghai Municipal Science and Technology Major Project under Grant 2021SHZDZX0100. H. Zhang's work was supported by the National Natural Science Foundation of China under Grant 62273030. J. Zhou and T. Chen's work was supported by the Natural Sciences and Engineering Research Council of Canada.

J. Shang is with the Department of Control Science and Engineering, Shanghai Research Institute for Intelligent Autonomous Systems, National Key Laboratory of Autonomous Intelligent Unmanned Systems, and Frontiers Science Center for Intelligent Autonomous Systems, Tongji University, Shanghai 200092, China. shangjun@tongji.edu.cn

H. Zhang is with the Key Laboratory of Knowledge Automation for Industrial Processes of Ministry of Education, School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China. zhanghanwen@ustb.edu.cn

J. Zhou and T. Chen are with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, T6G 1H9, Canada. jzhou15@ualberta.ca; tchen@ualberta.ca

dimensional systems.

The remainder of this paper is organized as follows. Section II formulates the problem. Section III depicts the design of stealthy linear attacks. Section IV gives numerical examples to verify the theoretical results. Finally, Section V concludes this paper.

Notations: $\mathbb{S}_{++}^n$ ($\mathbb{S}_{+}^n$) denotes the set of $n \times n$ positive (semi)-definite matrices. If $X \in \mathbb{S}_{+}^n$, denote $X \succeq 0$ (or $X \succ 0$ if $X \in \mathbb{S}_{++}^n$). $X \succeq Y$ if $X - Y \in \mathbb{S}_{+}^n$. For $X \in \mathbb{R}^{m\times n}$, $X^{\mathrm{T}} \in \mathbb{R}^{n\times m}$ is the transpose of X, and $\sigma_i(X)$ denotes the ith largest singular value of X. For $X \in \mathbb{R}^{m\times m}$, $\mathrm{Tr}(X)$ represents the trace of X. I denotes the identity matrix. For $x \in \mathbb{R}^m$, $\|x\|$ represents the Euclidean norm of x. $\Pr(\cdot)$ denotes the probability of an event. $\mathbb{E}[\cdot]$ denotes the expectation of a random variable. $\mathrm{cov}(\cdot)$ is the covariance of a random vector. $[\![a, b]\!] = \{x \in \mathbb{Z} | a \le x \le b\}$.

II. Problem Formulation

The system architecture explored in this study is depicted in Fig. 1. The smart sensor can process measurements using a local Kalman filter, with the resulting innovations transmitted sequentially via a wireless network. This setup is vulnerable to malicious attackers capable of intercepting and tampering with the data in transit. The attacker's goal is to impair the system's performance without triggering the anomaly detection mechanisms.

A. Process Model

In this study, we focus on a linear time-invariant (LTI) process, described as follows:

$$x_{k+1} = Ax_k + w_k \tag{1a}$$
$$y_k = Cx_k + v_k \tag{1b}$$

where $x_k \in \mathbb{R}^n$ is the system state, $y_k \in \mathbb{R}^m$ is the sensor measurement, and $w_k \in \mathbb{R}^n$ and $v_k \in \mathbb{R}^m$ represent zero-mean independent and identically distributed (i.i.d.) Gaussian noises with covariances $Q \in \mathbb{S}_{+}^n$ and $R \in \mathbb{S}_{++}^m$, respectively. The initial state $x_0 \in \mathbb{R}^n$ is zero-mean Gaussian. For all $k \ge 0$, w_k, v_k, and x_0 are mutually independent. Assume that $\mathrm{rank}(C) = m$, the pair (A, C) is observable, and $(A, \sqrt{Q})$ is controllable.

For system (1), the following Kalman filter is adopted to estimate the system state:

$$\hat{x}_k^- = A\hat{x}_{k-1} \tag{2a}$$
$$P_k^- = AP_{k-1}A^{\mathrm{T}} + Q \tag{2b}$$
$$K_k = P_k^- C^{\mathrm{T}}(CP_k^- C^{\mathrm{T}} + R)^{-1} \tag{2c}$$
$$z_k = y_k - C\hat{x}_k^- \tag{2d}$$
$$\hat{x}_k = \hat{x}_k^- + K_k z_k \tag{2e}$$
$$P_k = P_k^- - K_k C P_k^-. \tag{2f}$$

In (2), $\hat{x}_k^-$ denotes the *a priori* MMSE estimate of x_k, and P_k^- is the *a priori* estimation error covariance; similarly, $\hat{x}_k$ is the *a posteriori* MMSE estimate of x_k, and P_k denotes the corresponding estimation error covariance.

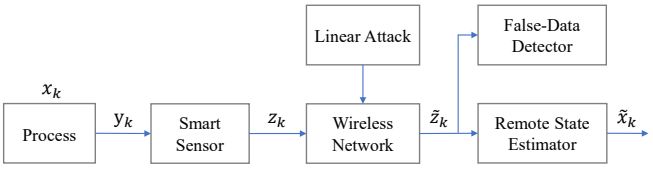

Fig. 1. Deception attacks on remote state estimation.

For notational simplicity, we define the following two matrix-valued functions $h, \tilde{g}\colon \mathbb{S}_{+}^n \to \mathbb{S}_{+}^n$ as

$$h(X) \triangleq AXA^{\mathrm{T}} + Q \tag{3}$$
$$\tilde{g}(X) \triangleq X - XC^{\mathrm{T}}(CXC^{\mathrm{T}} + R)^{-1}CX. \tag{4}$$

Thus, the recursion of the estimation error covariance in (2) simplifies to $P_k = \tilde{g}(P_k^-)$ and $P_{k+1}^- = h(P_k)$. Let $\bar{P}$ denote the steady-state *a priori* estimation error covariance, which is the unique positive semidefinite solution of $h[\tilde{g}(X)] = X$. We assume for simplicity that the filter has reached the steady state when $k = 0$. Consequently, the Kalman gain in (2c) simplifies to a constant matrix:

$$K = \bar{P}C^{\mathrm{T}}(C\bar{P}C^{\mathrm{T}} + R)^{-1}. \tag{5}$$

As indicated in Fig. 1, innovations are transmitted rather than raw measurements, facilitated by a local Kalman filter incorporated within the smart sensor.

B. Anomaly Detector

From (2), the innovation z_k follows a zero-mean Gaussian distribution. The steady-state covariance of z_k satisfies $\Sigma = C\bar{P}C^{\mathrm{T}} + R$. Additionally, the sequence $\{z_k\}$ exhibits zero autocorrelation. To monitor the transmitted data, a commonly utilized anomaly detector is the χ^2 detector. The detection index for this detector is expressed by

$$d_k = z_k^{\mathrm{T}}\Sigma^{-1}z_k. \tag{6}$$

Here, d_k adheres to a χ^2 distribution with m degrees of freedom. To assess the system's condition, we use the following hypothesis test:

$$\begin{cases} H_0 : d_k \le \delta_\alpha \\ H_1 : d_k > \delta_\alpha \end{cases} \tag{7}$$

where H_0 denotes normal system operation, H_1 indicates an anomaly, and δ_α is the threshold, which can be determined by the following equation:

$$F_m(\delta_\alpha) = \frac{1}{\Gamma(m/2)}\gamma\left(\frac{m}{2}, \frac{\delta_\alpha}{2}\right) = 1 - \alpha. \tag{8}$$

Here, $F_m(\cdot)$ signifies the cumulative distribution function of the χ^2-distribution with m degrees of freedom, Γ and γ denote the gamma function and the lower incomplete gamma function, respectively, and α is the preset significance level. Therefore, the false alarm rate when the system is operating normally is $\Pr(d_k > \delta_\alpha) = \alpha$.

C. Deception Attacks

The primary aim of the attacker is to degrade the estimation accuracy of the system while ensuring that the attack remains undetected. The linear attack model introduced in [6] is considered as follows:

$$\tilde{z}_k = T_k z_k + b_k \tag{9}$$

where $\tilde{z}_k$ represents the manipulated innovation, $T_k \in \mathbb{R}^{m\times m}$ is the attack coefficient matrix, and $b_k \in \mathbb{R}^m$ is a zero-mean i.i.d. Gaussian noise with covariance Π_k.

To effectively remain undetected by the anomaly detector at each moment, attackers must judiciously select T_k and Π_k for crafting $\tilde{z}_k$. The most straightforward approach is adopting the so-called strict stealthiness as described in [6], where $\tilde{z}_k$ and z_k have an identical pdf. This approach ensures that the alarm rates of the χ^2 detector remain unchanged, preserving the stealthiness of the attack. Moreover, the sequence $\tilde{z}_k$, produced under the attack model, retains zero autocorrelation, thus maintaining the innovation sequence's whiteness. This simplicity offers an advantage over some more complex strategies such as in [9], [10].

In our study, while adhering to the attack model specified, we explore a new version of the stealthiness constraint. Unlike the constraints based on the Kullback–Leibler (KL) divergence seen in work like [11], [12], [20], our focus is directly on the alarm rate triggered by the χ^2 detector. Considering the system's normal operational alarm rate is α, an attack that does not alter this rate is considered stealthy under our framework. Therefore, the criterion for our proposed deception attack is defined as:

$$\Pr(\tilde{z}_k^{\mathrm{T}}\Sigma^{-1}\tilde{z}_k > \delta_\alpha) = \alpha. \tag{10}$$

It is important to note that the stringent stealthiness condition set forth in [6] is merely a sufficient, not a necessary, condition for meeting the above criterion. For χ^2 detectors, directly monitoring the alarm rate offers a clear indicator of the system's state, distinguishing between normal and compromised conditions.

With $\tilde{z}_k$ in (9), the state estimate at the remote end will be calculated as follows:

$$\tilde{x}_k^- = A\tilde{x}_{k-1} \tag{11a}$$

$$\tilde{x}_k = \tilde{x}_k^- + K\tilde{z}_k. \tag{11b}$$

To measure the impact of the deception attack, we consider the following estimation error covariances:

$$\tilde{P}_k^- = \mathbb{E}[(x_k - \tilde{x}_k^-)(x_k - \tilde{x}_k^-)^{\mathrm{T}}] \tag{12a}$$

$$\tilde{P}_k = \mathbb{E}[(x_k - \tilde{x}_k)(x_k - \tilde{x}_k)^{\mathrm{T}}]. \tag{12b}$$

From (12b), we see that

$$\mathbb{E}[\|x_k - \tilde{x}_k\|_2^2] = \mathrm{Tr}(\tilde{P}_k). \tag{13}$$

Assuming the attacker has the capability to initiate consecutive attacks over the time interval $[\check{k}, \hat{k}]$, their objective is to maximize the accumulative trace of the estimation error covariance, denoted as $\sum_{k=\check{k}}^{\hat{k}} \mathrm{Tr}(\tilde{P}_k)$. This term, also called accumulative estimation error for simplicity, serves as a key metric for assessing the effectiveness of the attacks in degrading the system's estimation accuracy.

D. Problem of Interest

In this study, our goal is to identify the optimal linear deception attack strategy that preserves the alarm rate, focusing on maximizing the accumulative estimation error. Technically, we need to optimize the attack coefficients (T_k, Π_k) to maximize $\sum_{k=\check{k}}^{\hat{k}} \mathrm{Tr}(\tilde{P}_k)$ while satisfying constraint (10).

III. Main Results

To delve into the optimal attack strategy, we need to first obtain the recursion of the estimation error covariance induced by attacks. From (12b), the estimation error covariance can be rewritten as

$$\begin{aligned}\tilde{P}_k = {} & \mathbb{E}[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^{\mathrm{T}}] + \mathbb{E}[(x_k - \hat{x}_k)(\hat{x}_k - \tilde{x}_k)^{\mathrm{T}}] \\ & + \mathbb{E}[(\hat{x}_k - \tilde{x}_k)(x_k - \hat{x}_k)^{\mathrm{T}}] + \mathbb{E}[(\hat{x}_k - \tilde{x}_k)(\hat{x}_k - \tilde{x}_k)^{\mathrm{T}}].\end{aligned}$$

With the law of iterated expectations, we have

$$\begin{aligned}&\mathbb{E}[(x_k - \hat{x}_k)(\hat{x}_k - \tilde{x}_k)^{\mathrm{T}}] \\ &= \mathbb{E}\left[\mathbb{E}[(x_k - \hat{x}_k)(\hat{x}_k - \tilde{x}_k)^{\mathrm{T}}|z_{0:k}]\right] = 0.\end{aligned}$$

It follows that

$$\tilde{P}_k = P + \mathbb{E}[(\hat{x}_k - \tilde{x}_k)(\hat{x}_k - \tilde{x}_k)^{\mathrm{T}}] \tag{14}$$

where $P = \tilde{g}(\bar{P})$. From (2a), (2e), and (11), one can obtain

$$\hat{x}_k - \tilde{x}_k = A(\hat{x}_{k-1} - \tilde{x}_{k-1}) + K(z_k - \tilde{z}_k).$$

With $\hat{x}_{\check{k}}^- = \tilde{x}_{\check{k}}^-$, we have

$$\hat{x}_k - \tilde{x}_k = \sum_{i=0}^{k-\check{k}} A^i K(z_{k-i} - \tilde{z}_{k-i}). \tag{15}$$

Note that $\forall i \neq 0$, $\mathbb{E}[z_k z_{k-i}^{\mathrm{T}}] = 0$. Substituting (9) into (15) yields that

$$\begin{aligned}\mathbb{E}[(\hat{x}_k - \tilde{x}_k)(\hat{x}_k - \tilde{x}_k)^{\mathrm{T}}] = {} & \sum_{i=0}^{k-\check{k}} A^i K \mathbb{E}[(z_{k-i} - T_{k-i}z_{k-i} \\ & - b_{k-i})(z_{k-i} - T_{k-i}z_{k-i} - b_{k-i})^{\mathrm{T}}]K^{\mathrm{T}}(A^i)^{\mathrm{T}}.\end{aligned}$$

Combining the above equation with (14), we obtain

$$\begin{aligned}\tilde{P}_k = P + \sum_{i=0}^{k-\check{k}} A^i K \big(&\Pi_{k-i} + \Sigma - T_{k-i}\Sigma - \Sigma T_{k-i}^{\mathrm{T}} \\ &+ T_{k-i}\Sigma T_{k-i}^{\mathrm{T}}\big) K^{\mathrm{T}}(A^i)^{\mathrm{T}}.\end{aligned} \tag{16}$$

For notational simplicity, we define the following variable:

$$\Psi_k = \sum_{i=0}^{\hat{k}-k} K^{\mathrm{T}}(A^i)^{\mathrm{T}} A^i K \tag{17}$$

which has a recursion

$$\Psi_k = \Psi_{k+1} + K^{\mathrm{T}}(A^{\mathrm{T}})^{\hat{k}-k} A^{\hat{k}-k} K. \tag{18}$$

We then obtain

$$\sum_{k=\check{k}}^{\hat{k}} \mathrm{Tr}(\tilde{P}_k) = (\hat{k}-\check{k}+1)\mathrm{Tr}(P) + \sum_{k=\check{k}}^{\hat{k}} \mathrm{Tr}[\Psi_k(\Pi_k + \Sigma - T_k\Sigma - \Sigma T_k^{\mathrm{T}} + T_k\Sigma T_k^{\mathrm{T}})]. \quad (19)$$

Owing to the structure of (19), the design of the optimal attack sequence $\{T_k\}$ can be decoupled. That is, the optimal attack coefficient at instant k, T_k^*, has no impact on the optimal solution for T_{k+1}. So we can solve the coefficients separately, which is different from the coupled problem in [12], [15].

To sum up, the optimal deception attack that maximizes $\sum_{k=\check{k}}^{\hat{k}} \mathrm{Tr}(\tilde{P}_k)$ can be obtained by solving the following optimization problem:

$$\begin{aligned} \max_{T_k,\Pi_k} \quad & \mathrm{Tr}[\Psi_k(\Pi_k + T_k\Sigma T_k^{\mathrm{T}} - T_k\Sigma - \Sigma T_k^{\mathrm{T}})] \\ \text{s.t.} \quad & \Pr[(T_k z_k + b_k)^{\mathrm{T}}\Sigma^{-1}(T_k z_k + b_k) > \delta_\alpha] = \alpha \\ & \mathbb{E}[b_k b_k^{\mathrm{T}}] = \Pi_k. \end{aligned} \quad (20)$$

Denote the objective function of problem (20) as J_k. Note that the constant part in $\mathrm{Tr}(\tilde{P}_k)$, namely $\mathrm{Tr}(P + \Psi_k\Sigma)$, is not included in J_k. Denote the optimal solution to (20) as $\{T_k^*, \Pi_k^*\}$; the corresponding objective value is J_k^*. Then the covariance of $\tilde{z}_k^*$ satisfies

$$\mathrm{cov}(\tilde{z}_k^*) = T_k^*\Sigma(T_k^*)^{\mathrm{T}} + \Pi_k^*. \quad (21)$$

If we consider another attack with coefficient $T_k^\star \in \mathbb{R}^{m\times m}$ and $\Pi_k^\star = 0$, the attack model is $\tilde{z}_k^\star = T_k^\star z_k$. Let $\mathrm{cov}(\tilde{z}_k^\star) = \mathrm{cov}(\tilde{z}_k^*)$, i.e.,

$$T_k^*\Sigma(T_k^*)^{\mathrm{T}} + \Pi_k^* = T_k^\star\Sigma(T_k^\star)^{\mathrm{T}}. \quad (22)$$

It is easy to see that $T_k^\star$ satisfies the constraint of (20). Note that both T_k^* and $-T_k^*$ satisfies the constraint, and thus, we see from the objective function of (20) that

$$\mathrm{Tr}[\Psi_k(T_k^*\Sigma + \Sigma(T_k^*)^{\mathrm{T}})] \le 0, \mathrm{Tr}[\Psi_k(T_k^\star\Sigma + \Sigma(T_k^\star)^{\mathrm{T}})] \le 0.$$

For T_k^* and $T_k^\star$, the difference between the objective values satisfies

$$J_k^\star - J_k^* = \mathrm{Tr}\{\Psi_k[(T_k^* - T_k^\star)\Sigma + \Sigma(T_k^* - T_k^\star)^{\mathrm{T}}]\}.$$

Note that (22) leads to $(-T_k^\star)\Sigma(-T_k^\star)^{\mathrm{T}} \succeq (-T_k^*)\Sigma(-T_k^*)^{\mathrm{T}}$. Using the similar proof in [12, Th.1], one can verify that there exists $T_k^\star$ such that $\mathrm{Tr}[\Psi_k(T_k^* - T_k^\star)\Sigma] \ge 0$, and thus $J_k^\star - J_k^* \ge 0$. We then conclude that the optimal Π_k can be represented as $\Pi_k^* = 0$. It follows that $b_k = 0$ is an optimal solution, and the attack model reduces to $\tilde{z}_k = T_k z_k$. Accordingly, problem (20) can be rewritten as

$$\begin{aligned} \max_{T_k} \quad & \mathrm{Tr}[\Psi_k(T_k\Sigma T_k^{\mathrm{T}} - T_k\Sigma - \Sigma T_k^{\mathrm{T}})] \\ \text{s.t.} \quad & \Pr[z_k^{\mathrm{T}} T_k^{\mathrm{T}}\Sigma^{-1} T_k z_k > \delta_\alpha] = \alpha. \end{aligned} \quad (23)$$

Let $H_k = \Sigma^{-\frac{1}{2}} T_k \Sigma^{\frac{1}{2}}$. Then we have

$$\begin{aligned} J_k &= \mathrm{Tr}[\Psi_k(\Sigma^{\frac{1}{2}} H_k H_k^{\mathrm{T}}\Sigma^{\frac{1}{2}} - \Sigma^{\frac{1}{2}} H_k\Sigma^{\frac{1}{2}} - \Sigma^{\frac{1}{2}} H_k^{\mathrm{T}}\Sigma^{\frac{1}{2}})] \\ &= \mathrm{Tr}[\Sigma^{\frac{1}{2}}\Psi_k\Sigma^{\frac{1}{2}}(H_k H_k^{\mathrm{T}} - H_k - H_k^{\mathrm{T}})]. \end{aligned} \quad (24)$$

The constraint of (23) becomes

$$\Pr(z_k^{\mathrm{T}}\Sigma^{-\frac{1}{2}} H_k^{\mathrm{T}} H_k \Sigma^{-\frac{1}{2}} z_k > \delta_\alpha) = \alpha. \quad (25)$$

Let $\xi_k = \Sigma^{-\frac{1}{2}} z_k$. With this whitening, we have $\mathbb{E}[\xi_k\xi_k^{\mathrm{T}}] = I$. In fact, ξ_k is a vector composed of m i.i.d. random variables that follow standard normal distributions. We rewrite (25) as

$$\Pr(\xi_k^{\mathrm{T}} H_k^{\mathrm{T}} H_k \xi_k \le \delta_\alpha) = 1 - \alpha. \quad (26)$$

Consider the following singular value decomposition:

$$H_k = \bar{U}_k\bar{\Sigma}_k\bar{V}_k^{\mathrm{T}} \quad (27)$$

Then $H_k^{\mathrm{T}} H_k = \bar{V}_k\bar{\Sigma}_k^2\bar{V}_k^{\mathrm{T}}$ and $H_k H_k^{\mathrm{T}} = \bar{U}_k\bar{\Sigma}_k^2\bar{U}_k^{\mathrm{T}}$. It is clear that $\bar{U}_k$ has no impact on the constraint of (26), which can be rewritten as

$$\Pr(\xi_k^{\mathrm{T}}\bar{V}_k\bar{\Sigma}_k^2\bar{V}_k^{\mathrm{T}}\xi_k \le \delta_\alpha) = 1 - \alpha. \quad (28)$$

With $\hat{\xi}_k = \bar{V}_k^{\mathrm{T}}\xi_k$, (28) becomes $\Pr(\hat{\xi}_k^{\mathrm{T}}\bar{\Sigma}_k^2\hat{\xi}_k \le \delta_\alpha) = 1 - \alpha$. Additionally, $\mathbb{E}[\hat{\xi}_k\hat{\xi}_k^{\mathrm{T}}] = \bar{V}_k^{\mathrm{T}}\bar{V}_k = I$. We see that $\hat{\xi}_k$ and ξ_k share the same pdf. Then (28) reduces to

$$\Pr(\xi_k^{\mathrm{T}}\bar{\Sigma}_k^2\xi_k \le \delta_\alpha) = 1 - \alpha \quad (29)$$

which is independent of $\bar{V}_k$ as well. Therefore, if $\bar{\Sigma}_k$ satisfies (29), then we can arbitrarily select $\bar{U}_k$ and $\bar{V}_k$. We now use the following spectral decomposition:

$$\Sigma^{\frac{1}{2}}\Psi_k\Sigma^{\frac{1}{2}} = \bar{\Phi}_k\bar{\Lambda}_k\bar{\Phi}_k^{\mathrm{T}} \quad (30)$$

where $\bar{\Lambda}_k = \mathrm{diag}\{\bar{\lambda}_{k1}, \bar{\lambda}_{k2}, \ldots, \bar{\lambda}_{km}\}$ denote the eigenvalues, arranged in nonincreasing order. We then rewrite (24) as

$$J_k = \mathrm{Tr}\left[\bar{\Phi}_k\bar{\Lambda}_k\bar{\Phi}_k^{\mathrm{T}}(\bar{U}_k\bar{\Sigma}_k^2\bar{U}_k^{\mathrm{T}} - \bar{U}_k\bar{\Sigma}_k\bar{V}_k^{\mathrm{T}} - \bar{V}_k\bar{\Sigma}_k\bar{U}_k^{\mathrm{T}})\right].$$

Since we have shown that $\bar{U}_k$ and $\bar{V}_k$ have no impact on the constraint, we just need to consider the objective function. Based on the properties of singular values, one can prove

$$\begin{aligned} J_k &\le \sum_{i=1}^{k} \sigma_i\left[\bar{\Phi}_k\bar{\Lambda}_k\bar{\Phi}_k^{\mathrm{T}}(\bar{U}_k\bar{\Sigma}_k^2\bar{U}_k^{\mathrm{T}} - \bar{U}_k\bar{\Sigma}_k\bar{V}_k^{\mathrm{T}} - \bar{V}_k\bar{\Sigma}_k\bar{U}_k^{\mathrm{T}})\right] \\ &\le \sum_{i=1}^{k} \sigma_i(\bar{\Lambda}_k)\sigma_i(\bar{\Sigma}_k^2 + 2\bar{\Sigma}_k) \\ &= \mathrm{Tr}[\bar{\Lambda}_k\bar{\Sigma}_k^2 + 2\bar{\Lambda}_k\bar{\Sigma}_k]. \end{aligned}$$

It is easy to verify that $\bar{U}_k = -\bar{\Phi}_k$ and $\bar{V}_k = \bar{\Phi}_k$ yields $J_k = \mathrm{Tr}[\bar{\Lambda}_k\bar{\Sigma}_k^2 + 2\bar{\Lambda}_k\bar{\Sigma}_k]$, and thus $\bar{V}_k = -\bar{U}_k = \bar{\Phi}_k$ is an optimal solution.

Now the optimization variable reduces to $\bar{\Sigma}_k$ and the optimization problem can be simplified as

$$\begin{aligned} \max_{\bar{\Sigma}_k} \quad & \mathrm{Tr}(\bar{\Lambda}_k\bar{\Sigma}_k^2 + 2\bar{\Lambda}_k\bar{\Sigma}_k) \\ \text{s.t.} \quad & \Pr(\xi_k^{\mathrm{T}}\bar{\Sigma}_k^2\xi_k \le \delta_\alpha) = 1 - \alpha. \end{aligned} \quad (31)$$

Notice that $\bar{\Sigma}_k \in \mathbb{R}^{m\times m}$ is a diagonal matrix and it has only m degrees of freedom. Compared with the original optimization problem with respect to $T_k \in \mathbb{R}^{m\times m}$, which has m^2 degrees of freedom, this is a large reduction of

complexity. We then rewrite $\bar{\Sigma}_k = \mathrm{diag}\{\bar{\sigma}_{k1}, \bar{\sigma}_{k2}, \ldots, \bar{\sigma}_{km}\}$, $\xi_k = [\xi_{k1}, \xi_{k2}, \ldots, \xi_{km}]^{\mathrm{T}}$, and (31) becomes

$$\begin{aligned} \max_{\{\bar{\sigma}_{ki}\}_{i=1}^m} \quad & \sum_{i=1}^{m} (\bar{\lambda}_{ki}\bar{\sigma}_{ki}^2 + 2\bar{\lambda}_{ki}\bar{\sigma}_{ki}) \\ \text{s.t.} \quad & \Pr\left(\sum_{i=1}^{m} \xi_{ki}^2 \bar{\sigma}_{ki}^2 \le \delta_\alpha\right) = 1-\alpha. \end{aligned} \tag{32}$$

Because of the constraint in (32), it is not easy to obtain an analytical solution. However, for low-dimensional systems, say when $m = 2$, numerical solutions are readily attainable. In such a case, the decision variables reduce to $\bar{\sigma}_{k1}$ and $\bar{\sigma}_{k2}$. Given that $\{\bar{\lambda}_{ki}\}$ is arranged in nonincreasing order, the objective function of (32) implies $\bar{\sigma}_{k1} > \bar{\sigma}_{k2}$. Notice that $\bar{\sigma}_{ki} = 1$, $\forall i \in [\![1, m]\!]$ satisfies the constraint of (32). With some elementary deduction, we establish upper and lower bounds for $\bar{\sigma}_{ki}$:

$$0 \le \bar{\sigma}_{k2} \le 1 \le \bar{\sigma}_{k1} \le \sqrt{\frac{\delta_\alpha}{F_1^{-1}(1-\alpha)}}$$

where $F_1^{-1}(\cdot)$ is the inverse cumulative distribution function of the χ^2 distribution with a single degree of freedom, as detailed earlier in (8). We then need to numerically obtain the feasible region about $\bar{\sigma}_{k1}$ and $\bar{\sigma}_{k2}$, applicable uniformly across all k. Since the objective function of (32) is rather simple, we can obtain the optimal values of $\bar{\sigma}_{k1}$ and $\bar{\sigma}_{k2}$ efficiently through recursive algorithms. Moreover, since these computations can be executed offline, they do not impinge upon the real-time execution demands of deploying the attack sequence.

Now denote the optimal solution to (31) as $\bar{\Sigma}_k^*$. From (27), we have $H_k^* = -\bar{\Phi}_k \bar{\Sigma}_k^* \bar{\Phi}_k^{\mathrm{T}}$, and accordingly, the optimal attack coefficient satisfies

$$T_k^* = -\Sigma^{\frac{1}{2}} \bar{\Phi}_k \bar{\Sigma}_k^* \bar{\Phi}_k^{\mathrm{T}} \Sigma^{-\frac{1}{2}} \tag{33}$$

$$b_k = 0. \tag{34}$$

The corresponding optimal accumulative estimation error is

$$\sum_{k=\check{k}}^{\hat{k}} \mathrm{Tr}(\tilde{P}_k) = \sum_{k=\check{k}}^{\hat{k}} \mathrm{Tr}[\bar{\Lambda}_k(\bar{\Sigma}_k^* + I)^2] + (\hat{k} - \check{k} + 1)\mathrm{Tr}(P).$$

The time-varying nature of the attack coefficient T_k is evident. As discussed previously, setting $\bar{\Sigma}_k = I$ meets the constraint detailed in (31). Substituting $\bar{\Sigma}_k^*$ with the identity matrix simplifies T_k to $-I$, aligning with the strategy highlighted in [6]. This approach yields the accumulative estimation error as

$$\sum_{k=\check{k}}^{\hat{k}} \mathrm{Tr}(\tilde{P}_k) = 4\sum_{k=\check{k}}^{\hat{k}} \mathrm{Tr}(\bar{\Lambda}_k) + (\hat{k} - \check{k} + 1)\mathrm{Tr}(P).$$

If $\bar{\Sigma}_k^* \neq I$, then the approach delineated in [6] emerges as merely a suboptimal solution with respect to the alarm rate constraint. The notable distinction lies in the requirement from [6] for the innovations' covariance under attack, $\mathrm{cov}(\tilde{z}_k)$, to match that of the original, $\mathrm{cov}(z_k)$, a stipulation absent in our discussion where the focus is solely on maintaining the alarm rate. Further intuitive comparisons are forthcoming in the subsequent section.

IV. Numerical Examples

In this section, we consider a two-dimensional LTI system. The system parameters are given as follows:

$$\begin{aligned} A &= \begin{bmatrix} 0.55 & 0.25 \\ -0.05 & 0.64 \end{bmatrix}, \; C = \begin{bmatrix} 0.9 & -0.8 \\ 0.1 & 0.7 \end{bmatrix} \\ Q &= \mathrm{diag}\{0.5, 0.7\}, \; R = \mathrm{diag}\{0.6, 0.8\}. \end{aligned} \tag{35}$$

To estimate the system state in (35), a Kalman filter is used. In nominal conditions, the steady-state estimation error covariance satisfies

$$\bar{P} = \begin{bmatrix} 0.727 & 0.132 \\ 0.132 & 0.872 \end{bmatrix}, \; P = \begin{bmatrix} 0.463 & 0.215 \\ 0.215 & 0.449 \end{bmatrix}.$$

To monitor the system, we employ a χ^2 detector with a permissible false alarm rate set at 1%, corresponding to a significance level, α, of 0.01. The detection threshold, δ_α, is determined to be 9.21 following (8).

The attacker initiates data manipulation from the 21st sampling instant, continuing until the 50th instant, designated as $\check{k} = 21$ and $\hat{k} = 50$. Using the steady-state Kalman gain K and matrix A, we can calculate Ψ_k for each relevant k. For the duration between $k = 21$ and $k = 50$, we derive the optimal attack coefficient T_k following (33). To assess the efficacy of this attack, 10^6 Monte Carlo simulations are conducted to average the estimation error covariances. The outcomes of this attack, depicted in Fig. 2, are then contrasted with the conventional attack strategy from [6]. For a comprehensive comparison, we also examine the attack strategy from [11] that employs the KL divergence as a metric for stealthiness. It is crucial to note, however, that the attack from [11] would modify the alarm rate with any nonzero KL divergence. To minimize the change in the alarm rate, a minimal KL divergence of 10^{-4} is chosen for this comparison.

From Fig. 2, it is evident that the attack strategy we propose results in the most significant deterioration of estimation performance, leading to an accumulative estimation error of 167.3. In comparison, the attacks described in [6] and [11] lead to accumulative estimation errors of 163.9 and 165.0, respectively.

To confirm that our proposed attack strategy does not affect the alarm rate, we calculate the average alarm rates resulting from the different attacks, as shown in Fig. 3. Notably, while the attack from [11] causes a change in the alarm rate, the other two strategies do not. This outcome affirms that our attack strategy consistently satisfies the constraint regarding the alarm rate.

For a clearer understanding, we also present a single-instance simulation of the proposed attack strategy in Fig. 4, illustrating the detection index as per (6) alongside the designated threshold δ_α. It is apparent that the proposed attack maintains the alarm rate unchanged.

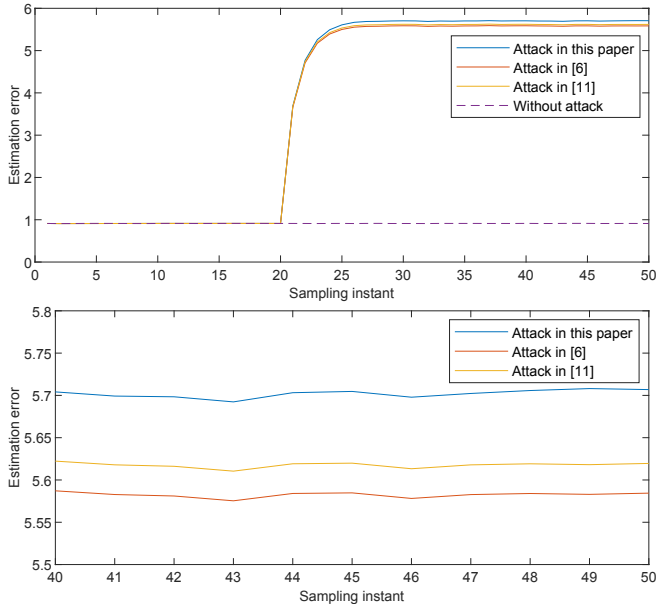

Fig. 2. Attack performance of different strategies.

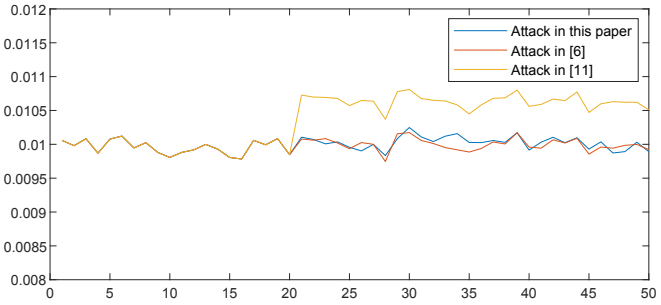

Fig. 3. Average alarm rates of different attacks.

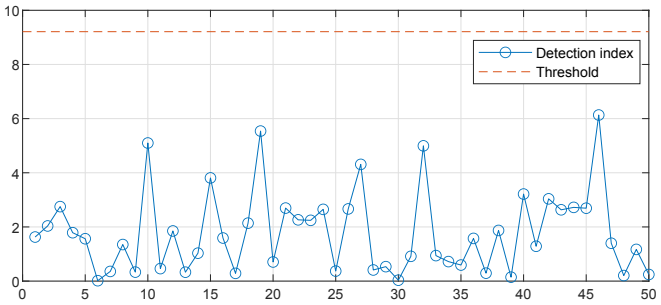

Fig. 4. One-shot simulation of detection indices.

V. Conclusion and Future Work

In this paper, we explore the optimal linear attack strategy for remote state estimation. Unlike previous studies that focus on stealthiness defined through innovation covariance or the KL divergence, we anchor our stealthiness constraint in the alarm rate. We demonstrate that, for low-dimensional systems, the optimal linear attack reveals intriguing structural properties, negating the need for a random compensation term in its design. Our work also establishes the relationship between the proposed strategy and prior approaches, highlighting the superiority of our method through numerical simulations. Future directions include extending the analysis to attacks on high-dimensional systems under constrained alarm rates and investigating defense mechanisms against these stealthy attacks, potentially employing Stackelberg games to determine optimal defense strategies.

References

[1] D. Ding, Q. L. Han, X. Ge, and J. Wang, "Secure state estimation and control of cyber-physical systems: A survey," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 1, pp. 176–190, 2021.

[2] A. Naha, A. M. Teixeira, A. Ahlén, and S. Dey, "Quickest detection of deception attacks on cyber–physical systems with a parsimonious watermarking policy," *Automatica*, vol. 155, 2023, Art. no. 111147.

[3] J. Zhou, J. Shang, and T. Chen, "Cybersecurity landscape on remote state estimation: A comprehensive review," *IEEE/CAA J. Autom. Sinica*, vol. 11, no. 4, pp. 851–865, 2024.

[4] J. Zhou, J. Shang, Y. Li, and T. Chen, "Optimal DoS attack against LQR control channels," *IEEE Trans. Circuits Syst. II*, vol. 68, no. 4, pp. 1348–1352, 2021.

[5] A. Y. Lu and G. H. Yang, "False data injection attacks against state estimation without knowledge of estimators," *IEEE Trans. Autom. Control*, vol. 67, no. 9, pp. 4529–4540, 2022.

[6] 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, 2017.

[7] ——, "Worst-case innovation-based integrity attacks with side information on remote state estimation," *IEEE Trans. Control Netw. Syst.*, vol. 6, no. 1, pp. 48–59, 2019.

[8] J. Zhou, J. Shang, and T. Chen, "On information fusion in optimal linear FDI attacks against remote state estimation," *IEEE Trans. Control Netw. Syst.*, vol. 10, no. 4, pp. 2085–2096, 2023.

[9] Y. Li and G. Yang, "Optimal stealthy innovation-based attacks with historical data in cyber-physical systems," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 6, pp. 3401–3411, 2021.

[10] J. Shang and T. Chen, "Optimal stealthy integrity attacks on remote state estimation: The maximum utilization of historical data," *Automatica*, vol. 128, 2021, Art. no. 109555.

[11] 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.

[12] J. Shang, H. Yu, and T. Chen, "Worst-case stealthy innovation-based linear attacks on remote state estimation under Kullback–Leibler divergence," *IEEE Trans. Autom. Control*, vol. 67, no. 11, pp. 6082–6089, 2022.

[13] ——, "Worst-case stealthy attacks on stochastic event-based state estimation," *IEEE Trans. Autom. Control*, vol. 67, no. 4, pp. 2052–2059, 2021.

[14] J. Shang, J. Zhou, and T. Chen, "Nonlinear stealthy attacks on remote state estimation," *Automatica*, vol. 167, 2024, Art. no. 111747.

[15] J. Zhou, J. Shang, and T. Chen, "Optimal deception attacks against remote state estimation: An information-based approach," *IEEE Trans. Autom. Control*, vol. 68, no. 7, pp. 3947–3962, 2023.

[16] H. Liu, Y. Ni, L. Xie, and K. H. Johansson, "How vulnerable is innovation-based remote state estimation: Fundamental limits under linear attacks," *Automatica*, vol. 136, 2022, Art. no. 110079.

[17] J. Shang, M. Chen, and T. Chen, "Optimal linear encryption against stealthy attacks on remote state estimation," *IEEE Trans. Autom. Control*, vol. 66, no. 8, pp. 3592–3607, 2021.

[18] Y. Li and G. Yang, "Optimal stealthy false data injection attacks in cyber-physical systems," *Inf. Sci.*, vol. 481, pp. 474–490, 2019.

[19] J. Shang, J. Zhou, and T. Chen, "Single-dimensional encryption against innovation-based stealthy attacks on remote state estimation," *Automatica*, vol. 136, 2022, Art. no. 110015.

[20] C. Bai, F. Pasqualetti, and V. Gupta, "Data-injection attacks in stochastic control systems: Detectability and performance tradeoffs," *Automatica*, vol. 82, pp. 251–260, 2017.