2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Stealthy Linear Deception Attacks against Kalman Filtering with Partially Secured Measurements

Jing Zhou and Tongwen Chen

Abstract— **In this paper, we investigate an optimal strategy for malicious agents to compromise remote state estimators where only a portion of the transmitted packets is secured. First, the analysis of the performance evolution and stealthiness properties of innovation-based linear attacks that can compromise unsafe transmission channels is provided. An optimal attack policy is then derived by numerically solving an optimization problem step by step. Different from the scenario where all links are vulnerable to cyber-attacks, the existence of secured channels poses a tighter stealthiness constraint and thus can significantly reduce the worst-case attack impact. Additionally, it is shown that the well-studied *flipping-sign-attack* in existing work cannot remain stealthy. Finally, a numerical example and comparative studies are included to verify the effectiveness of the proposed method.**

I. INTRODUCTION

The state-of-the-art in modern industrial applications has greatly promoted the deployment of wireless network techniques. Yet theses advancements also bring forward new challenges regarding threats from cyber-attacks. In various situations, malicious agents are capable of intercepting and modifying data packets in unreliable transmission channels, with the goal to degrade system performance [1], [2], [3], [4], [5]. Specially, the data manipulation that achieves the maximum performance degradation and deceives anomaly detectors is known as an "optimal stealthy attack" and has stimulated extensive research interests in the past decade.

The deception attack against remote state estimation has gained particular attention in recent years. In the scenario where different sensors measure system states and send packets to the remote end, attackers may modify the transmitted data sophisticatedly to mislead a Kalman filter into non-optimal estimation. If all measurements could be manipulated and the compromised innovation was required to be identically and independent distributed (i.i.d.) Gaussian, the optimal innovation-based linear attack was simply flipping the sign of nominal innovations [6]. The result was extended to the cases where attackers deployed extra sensors to gain side information of system states and attacks with relaxed stealthiness [7], [8], [9]. Though the latest research found the optimal information-based attack was an affine function of the minimum mean-square error (MMSE) estimation of prediction errors [10], linear attacks still reside at the center of our concern owing to their simple form and the capacity to deceive whiteness detectors [11], [12], [13].

This work was supported by the Natural Sciences and Engineering Research Council of Canada.

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

Almost all existing work on innovation-based linear attacks focuses on the scenario that entire data channels can be compromised. However in practical cases measurements may have various levels of confidentiality or be transmitted to the remote end through different channels. For the links with high reliability (e.g., cable transmission), it is rather difficult or even impossible for adversaries to manipulate data. For the insecure channels with low protection (e.g., wireless transmission), malicious agents may modify the data freely according to their interests. When only a portion of transmitted packets is vulnerable, there is little work on discussing how to design optimal attacks from an adversary's perspective to worsen system performance; the guideline to make full utilization of secured channels for system protection is also lacking. Li and Chen [14] proposed a sequential data fusion algorithm to defend against deception attacks in unsecured channels, while the underlying assumption was that the adversary adopted the so-called *flipping-sign-attack* in [6]. A countermeasure against this type of attacks using learning algorithms was given in [15]. As will be shown in this paper, flipping the sign of nominal innovations from unsafe channels will lead to changes of statistical properties of innovations from the stacked measurements, thus the attack can be easily detected by a χ^2 detector.

To fill the gap in this research topic, we investigate the optimal linear strategy for malicious agents to compromise remote state estimators when partial transmitted packets are secured. The detailed analysis on the attack performance evolution and stealthiness properties is given. The optimal attack coefficients are then obtained by solving a constrained optimization problem repeatedly. A numerical example shows that protecting only a small portion of transmission channels can significantly reduce the worst-case attack impact, thus providing new insights on designing defensive measures for networked systems. The finding is also consistent with [16], where encrypting a single data channel is sufficient to prevent most of stealthy deception attacks. The advantage over [16] is that the method in this work does not require an encryption/decryption module but only an additional χ^2 anomaly detector.

The remainder of this paper is organized as follows. Section II describes the deception attack problem. Section III shows the design procedure for stealthy linear attacks. Section IV gives a numerical example to verify the theoretical results. Finally, Section V concludes this paper.

Notations: $\mathcal{N}(0,\Pi)$ stands for a zero-mean Gaussian distribution. $h(X) \triangleq AXA^{\mathrm{T}} + Q$ is the Lyapunov operator. $g_{[C,R]}(X) \triangleq X - XC^{\mathrm{T}}(CXC^{\mathrm{T}} + R)^{-1}CX$ is the Riccati operator.

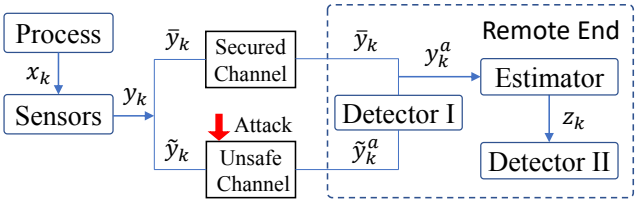

Fig. 1. Deception attacks against remote state estimation.

II. Problem Formulation

The system architecture of remote state estimation with secured and unsafe channels is illustrated in Fig. 1.

A. Process Model

In this work, we consider an LTI process

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

where $x_k \in \mathbb{R}^n$ is the system state; $y_k \in \mathbb{R}^m$ denotes the measurement; $w_k \in \mathbb{R}^n$ and $v_k \in \mathbb{R}^m$ are zero-mean i.i.d. Gaussian noises with covariance $Q \in \mathbb{S}_+^n$ and $R \in \mathbb{S}_{++}^m$, respectively. The initial state $x_0 \in \mathbb{R}^n$ is zero-mean Gaussian with covariance $\Pi_0 \in \mathbb{S}_+^n$, independent of w_k and $v_k, \forall k \in \mathbb{N}$. As depicted in Fig. 1, the raw measurement is partitioned into safe and unsafe parts. Define

$$y_k = \begin{bmatrix} \bar{y}_k \\ \tilde{y}_k \end{bmatrix},\ C = \begin{bmatrix} \bar{C} \\ \tilde{C} \end{bmatrix},\ v_k = \begin{bmatrix} \bar{v}_k \\ \tilde{v}_k \end{bmatrix},\ R = \begin{bmatrix} \bar{R} & S \\ S^{\mathrm{T}} & \tilde{R} \end{bmatrix},$$

where $\bar{y}_k = \bar{C}x_k + \bar{v}_k \in \mathbb{R}^{\bar{m}}$ is transmitted through a secured channel and thus cannot be altered, while $\tilde{y}_k = \tilde{C}x_k + \tilde{v}_k \in \mathbb{R}^{\tilde{m}}$ may be modified by malicious agents. Assume the pair $(A, \tilde{C})$ is detectable and $(A, \sqrt{Q})$ is controllable.

B. Anomaly Detector

Once the data has been received by the remote end, two χ^2 detectors can be utilized to detect anomalies. The first one is deployed to reveal abnormal events (faults/attacks) in unsafe channels. The second one is a byproduct of the remote state estimator. Each χ^2 detector is associated with a Kalman filter, which generates MMSE state estimation based on the received data and produces an i.i.d. Gaussian sequence [17]. Without loss of generality, we assume both filters are in steady state. Whenever $\tilde{g}_k$ or g_k exceeds a given threshold, an alarm will be raised.

i) $\tilde{g}_k = \tilde{z}_k^{\mathrm{T}} \tilde{\Sigma}^{-1} \tilde{z}_k$, where $\tilde{z}_k$ is the innovation corresponding to the unsecured measurement $\tilde{y}_k$, i.e.,

$$\tilde{x}_{k|k-1} = A\tilde{x}_{k-1|k-1}, \tag{3a}$$
$$\tilde{z}_k = \tilde{y}_k - \tilde{C}\tilde{x}_{k|k-1}, \tag{3b}$$
$$\tilde{x}_{k|k} = \tilde{x}_{k|k-1} + \tilde{K}\tilde{z}_k, \tag{3c}$$

ii) $g_k = z_k^{\mathrm{T}} \Sigma^{-1} z_k$, where z_k is the innovation corresponding to the stacked measurement y_k, i.e.,

$$x_{k|k-1} = Ax_{k-1|k-1}, \tag{4a}$$
$$z_k = y_k - Cx_{k|k-1}, \tag{4b}$$
$$x_{k|k} = x_{k|k-1} + Kz_k, \tag{4c}$$

where $\tilde{x}_{k|k-1}$ and $x_{k|k-1}$ are the *a priori* state estimates; $\tilde{x}_{k|k}$ and $x_{k|k}$ are the corresponding *a posteriori* state estimates. $\tilde{K} = \tilde{P}\tilde{C}^{\mathrm{T}}\tilde{\Sigma}^{-1}$ and $K = PC^{\mathrm{T}}\Sigma^{-1}$ represent the estimator gains; $\tilde{\Sigma} = \tilde{C}\tilde{P}\tilde{C}^{\mathrm{T}} + \tilde{R}$, $\Sigma = CPC^{\mathrm{T}} + R$. $\tilde{P}$ and P denote solutions of the Riccati equations $X = h[g_{[\tilde{C},\tilde{R}]}(X)]$ and $X = h[g_{[C,R]}(X)]$, respectively. In steady state, the two innovations in (3b) and (4b) satisfy $\tilde{z}_k \sim \mathcal{N}(0_{\tilde{m}\times 1}, \tilde{\Sigma})$ and $z_k \sim \mathcal{N}(0_{m\times 1}, \Sigma)$.

C. Deception Attacks

From instant $\bar{k}$, we assume the attacker adopts the following innovation-based linear attack to modify the measurements in unsafe channels:

$$\tilde{z}_k^a = T_k\tilde{z}_k + b_k,\ b_k \sim \mathcal{N}(0_{\tilde{m}\times 1}, \Phi_k),$$

where $\tilde{z}_k^a$ is the compromised innovation of filter (3); $b_k \in \mathbb{R}^{\tilde{m}}$ is an i.i.d. Gaussian noise to compensate the stealthiness property. We also make the standard assumption that the attacker knows all system parameters and the initial state of filter (3); thus he/she can run a duplicated Kalman filter as (3) to obtain the nominal innovation $\tilde{z}_k$. At each instant, the attacker should choose $T_k \in \mathbb{R}^{\tilde{m}\times\tilde{m}}$ and $\Phi_k \in \mathbb{S}_+^{\tilde{m}}$ to design $\tilde{z}_k^a$, which is equivalent to modifying $\tilde{y}_k$ as

$$\tilde{y}_k^a = \tilde{C}\tilde{x}_{k|k-1}^a + T_k\tilde{z}_k + b_k,\ b_k \sim \mathcal{N}(0_{\tilde{m}\times 1}, \Phi_k), \tag{5}$$

where $\tilde{x}_{k|k-1}^a$ is the compromised *a priori* state estimate and can be determined by the following equation:

$$\tilde{x}_{k|k-1}^a = A^{k-\bar{k}}\tilde{x}_{\bar{k}|\bar{k}-1} + \sum_{i=\bar{k}}^{k-1} A^{k-i}K\tilde{z}_i^a.$$

The attack performance is measured by the trace of the *a posteriori* estimation error covariance, namely

$$P_{k|k}^a = \mathbb{E}[(x_k - x_{k|k}^a)(x_k - x_{k|k}^a)^{\mathrm{T}}],$$

where $x_{k|k}^a$ is the compromised estimation of filter (4).

D. Problem of Interest

In this work, we aim to design the optimal attack coefficients (T_k, Φ_k) to maximize $\mathrm{Trace}(P_{k|k}^a)$, meanwhile the following constraints are fulfilled:

$$\tilde{z}_k^a \sim \mathcal{N}(0_{\tilde{m}\times 1}, \tilde{\Sigma}), \tag{6a}$$
$$z_k^a \sim \mathcal{N}(0_{m\times 1}, \Sigma), \tag{6b}$$

where z_k^a is the compromised innovation of filter (4). Constraint (6) guarantees that the alarm rates of two χ^2 detectors will not change; hence the attack can remain stealthy.

III. Main Results

With presence of deception attacks, the corrupted data received by the remote state estimator is

$$y_k^a = \begin{bmatrix} \bar{y}_k \\ \tilde{y}_k^a \end{bmatrix} = \begin{bmatrix} \bar{C}x_k + \bar{v}_k \\ \tilde{C}\tilde{x}_{k|k-1}^a + T_k\tilde{z}_k + b_k \end{bmatrix},$$

thus the innovation of filter (4) becomes

$$z_k^a = y_k^a - Cx_{k|k-1}^a = \begin{bmatrix} \bar{y}_k - \bar{C}x_{k|k-1}^a \\ \tilde{y}_k^a - \tilde{C}x_{k|k-1}^a \end{bmatrix} = \begin{bmatrix} \bar{C}(x_k - x_{k|k-1}^a) + \bar{v}_k \\ \tilde{C}(\tilde{x}_{k|k-1}^a - x_{k|k-1}^a) + T_k\tilde{z}_k + b_k \end{bmatrix}.$$

Define $\tilde{e}_{k|k-1}=x_k-\tilde{x}_{k|k-1}$ and $e_{k|k-1}=x_k-x_{k|k-1}$ as the prediction errors of the Kalman filters in (3) and (4), respectively. $\tilde{e}^a_{k|k-1}=x_k-\tilde{x}^a_{k|k-1}$ and $e^a_{k|k-1}=x_k-x^a_{k|k-1}$ are their counterparts when there exist deception attacks. Note that $\tilde{z}_k=\tilde{C}\tilde{e}_{k|k-1}+\tilde{v}_k$; define the following matrices

$$\hat{C}_k=\begin{bmatrix}0_{\bar{m}\times n} & 0_{\bar{m}\times n} & \bar{C}\\ T_k\tilde{C} & -\tilde{C} & \tilde{C}\end{bmatrix},\ \hat{T}_k=\begin{bmatrix}I_{\bar{m}} & 0_{\bar{m}\times\tilde{m}}\\ 0_{\tilde{m}\times\bar{m}} & T_k\end{bmatrix},$$
$$E=\begin{bmatrix}0_{\bar{m}\times\tilde{m}}\\ I_{\tilde{m}}\end{bmatrix},\ \eta_k=\begin{bmatrix}\tilde{e}^{\mathrm{T}}_{k|k-1} & (\tilde{e}^a_{k|k-1})^{\mathrm{T}} & (e^a_{k|k-1})^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}},$$

then z^a_k can be rewritten in a compact form as

$$z^a_k=\hat{C}_k\eta_k+\hat{T}_kv_k+Eb_k. \tag{7}$$

Next we analyze the influence of the modified data y^a_k on the covariance of z^a_k and the trace of $P^a_{k|k}$.

A. The Dynamics of Prediction Errors

From (1) and (3), we have

$$\tilde{e}_{k|k-1}=A\tilde{e}_{k-1|k-2}-A\tilde{K}\tilde{z}_{k-1}+w_{k-1}.$$

Substituting $\tilde{z}_{k-1}=\tilde{C}\tilde{e}_{k-1|k-2}+\tilde{v}_{k-1}$ yields

$$\tilde{e}_{k|k-1}=(A-A\tilde{K}\tilde{C})\tilde{e}_{k-1|k-2}-A\tilde{K}\tilde{v}_{k-1}+w_{k-1}. \tag{8}$$

For the compromised filter in (3), we have

$$\tilde{e}^a_{k|k-1}=A\tilde{e}^a_{k-1|k-2}-A\tilde{K}\tilde{z}^a_{k-1}+w_{k-1}.$$

Substituting $\tilde{z}^a_k=T_k\tilde{z}_k+b_k$ and $\tilde{z}_k=\tilde{C}\tilde{e}_{k|k-1}+\tilde{v}_k$ yields

$$\begin{aligned}\tilde{e}^a_{k|k-1}=&A\tilde{e}^a_{k-1|k-2}-A\tilde{K}T_{k-1}\tilde{C}\tilde{e}_{k-1|k-2}\\ &-A\tilde{K}T_{k-1}\tilde{v}_{k-1}-A\tilde{K}b_{k-1}+w_{k-1}.\end{aligned} \tag{9}$$

For the compromised filter in (4), we have

$$e^a_{k|k-1}=Ae^a_{k-1|k-2}-AKz^a_{k-1}+w_{k-1}.$$

The filter gain can be partitioned as $K=\begin{bmatrix}K_{\mathrm{I}} & K_{\mathrm{II}}\end{bmatrix}$, where $K_{\mathrm{I}}\in\mathbb{R}^{n\times\bar{m}}$, $K_{\mathrm{II}}\in\mathbb{R}^{n\times\tilde{m}}$. Then substituting (7) into $e^a_{k|k-1}$ and with some straightforward derivations, we have

$$\begin{aligned}e^a_{k|k-1}=&\begin{bmatrix}-AK_{\mathrm{II}}T_{k-1}\tilde{C} & AK_{\mathrm{II}}\tilde{C} & A-AKC\end{bmatrix}\eta_{k-1}\\ &-\begin{bmatrix}AK_{\mathrm{I}} & AK_{\mathrm{II}}T_{k-1}\end{bmatrix}v_{k-1}-AK_{\mathrm{II}}b_{k-1}+w_{k-1}\\ =&-AK_{\mathrm{II}}T_{k-1}\tilde{C}\tilde{e}_{k-1|k-2}+AK_{\mathrm{II}}\tilde{C}\tilde{e}^a_{k-1|k-2}\\ &+(A-AKC)e^a_{k-1|k-2}-AK_{\mathrm{I}}\bar{v}_{k-1}-AK_{\mathrm{II}}T_{k-1}\tilde{v}_{k-1}\\ &-AK_{\mathrm{II}}b_{k-1}+w_{k-1}.\end{aligned} \tag{10}$$

B. The Covariance of z^a_k

Notice that v_k,b_k are zero-mean and independent of η_k; from (7), the covariance of z^a_k is given by

$$\mathscr{C}_{z^a_k}\triangleq\mathbb{E}[z^a_k(z^a_k)^{\mathrm{T}}]=\hat{C}_k\mathbb{E}[\eta_k\eta_k^{\mathrm{T}}]\hat{C}_k^{\mathrm{T}}+\hat{T}_kR\hat{T}_k^{\mathrm{T}}+E\Phi_kE^{\mathrm{T}}. \tag{11}$$

Now we evaluate the recursion of $\mathscr{C}_{\eta_k}=\mathbb{E}[\eta_k\eta_k^{\mathrm{T}}]$, which is partitioned as a 3-by-3 block matrix. The $(1,1)$th block denotes the covariance of $\tilde{e}_{k|k-1}$. Since the filter has already been in steady state before instant $\bar{k}$, we have $\mathscr{C}^{11}_{\eta_k}=\tilde{P},\forall k\geq\bar{k}$. From (8)–(9), we have that the $(1,2)$th block satisfies

$$\begin{aligned}\mathscr{C}^{12}_{\eta_k}=&\mathbb{E}[\tilde{e}_{k|k-1}(\tilde{e}^a_{k|k-1})^{\mathrm{T}}]\\ =&-(A-A\tilde{K}\tilde{C})\mathbb{E}[\tilde{e}_{k-1|k-2}\tilde{e}^{\mathrm{T}}_{k-1|k-2}]\tilde{C}^{\mathrm{T}}T^{\mathrm{T}}_{k-1}\tilde{K}^{\mathrm{T}}A^{\mathrm{T}}\\ &+(A-A\tilde{K}\tilde{C})\mathbb{E}[\tilde{e}_{k-1|k-2}(\tilde{e}^a_{k-1|k-2})^{\mathrm{T}}]A^{\mathrm{T}}\\ &+A\tilde{K}\mathbb{E}[\tilde{v}_{k-1}\tilde{v}^{\mathrm{T}}_{k-1}]T^{\mathrm{T}}_{k-1}\tilde{K}^{\mathrm{T}}A^{\mathrm{T}}+\mathbb{E}[w_{k-1}w^{\mathrm{T}}_{k-1}]\\ =&(A-A\tilde{K}\tilde{C})(\mathscr{C}^{12}_{\eta_{k-1}}-\tilde{P}\tilde{C}^{\mathrm{T}}T^{\mathrm{T}}_{k-1}\tilde{K}^{\mathrm{T}})A^{\mathrm{T}}\\ &+A\tilde{K}\tilde{R}T^{\mathrm{T}}_{k-1}\tilde{K}^{\mathrm{T}}A^{\mathrm{T}}+Q\\ =&(A-A\tilde{K}\tilde{C})\mathscr{C}^{12}_{\eta_{k-1}}A^{\mathrm{T}}+Q,\end{aligned} \tag{12}$$

where the last equality is from the equation

$$\begin{aligned}&-(A-A\tilde{K}\tilde{C})\tilde{P}\tilde{C}^{\mathrm{T}}T^{\mathrm{T}}_{k-1}\tilde{K}^{\mathrm{T}}A^{\mathrm{T}}+A\tilde{K}\tilde{R}T^{\mathrm{T}}_{k-1}\tilde{K}^{\mathrm{T}}A^{\mathrm{T}}\\ =&A[\tilde{K}(\tilde{C}\tilde{P}\tilde{C}^{\mathrm{T}}+\tilde{R})-\tilde{P}\tilde{C}^{\mathrm{T}}]T^{\mathrm{T}}_{k-1}\tilde{K}^{\mathrm{T}}A^{\mathrm{T}}\\ =&A(\tilde{K}\tilde{\Sigma}-\tilde{P}\tilde{C}^{\mathrm{T}})T^{\mathrm{T}}_{k-1}\tilde{K}^{\mathrm{T}}A^{\mathrm{T}}=0_n.\end{aligned} \tag{13}$$

Since the *a priori* state estimation at step $\bar{k}$ is not affected by deception attacks, $\tilde{e}_{\bar{k}|\bar{k}-1}$ and $\tilde{e}^a_{\bar{k}|\bar{k}-1}$ are identical. Thus $\mathscr{C}^{12}_{\eta_{\bar{k}}}=\mathscr{C}^{11}_{\eta_{\bar{k}}}=\tilde{P}$. It follows directly from (12) that $\mathscr{C}^{12}_{\eta_k}=\tilde{P},\forall k\geq\bar{k}$. For the $(2,2)$th block, we have

$$\begin{aligned}\mathscr{C}^{22}_{\eta_k}=&\mathbb{E}[\tilde{e}^a_{k|k-1}(\tilde{e}^a_{k|k-1})^{\mathrm{T}}]\\ =&A\mathbb{E}[\tilde{e}^a_{k-1|k-2}(\tilde{e}^a_{k-1|k-2})^{\mathrm{T}}]A^{\mathrm{T}}+A\tilde{K}\mathbb{E}[b_{k-1}b^{\mathrm{T}}_{k-1}]\tilde{K}^{\mathrm{T}}A^{\mathrm{T}}\\ &+A\tilde{K}T_{k-1}\tilde{C}\mathbb{E}[\tilde{e}_{k-1|k-2}\tilde{e}^{\mathrm{T}}_{k-1|k-2}]\tilde{C}^{\mathrm{T}}T^{\mathrm{T}}_{k-1}\tilde{K}^{\mathrm{T}}A^{\mathrm{T}}\\ &+A\tilde{K}T_{k-1}\mathbb{E}[\tilde{v}_{k-1}\tilde{v}^{\mathrm{T}}_{k-1}]T^{\mathrm{T}}_{k-1}\tilde{K}^{\mathrm{T}}A^{\mathrm{T}}+A\tilde{K}\mathbb{E}[b_{k-1}b^{\mathrm{T}}_{k-1}]\tilde{K}^{\mathrm{T}}A^{\mathrm{T}}\\ &-A\mathbb{E}[\tilde{e}^a_{k-1|k-2}\tilde{e}^{\mathrm{T}}_{k-1|k-2}]\tilde{C}^{\mathrm{T}}T^{\mathrm{T}}_{k-1}\tilde{K}^{\mathrm{T}}A^{\mathrm{T}}\\ &-A\tilde{K}T_{k-1}\tilde{C}\mathbb{E}[\tilde{e}_{k-1|k-2}(\tilde{e}^a_{k-1|k-2})^{\mathrm{T}}]A^{\mathrm{T}}+\mathbb{E}[w_{k-1}w^{\mathrm{T}}_{k-1}]\\ =&A\mathscr{C}^{22}_{\eta_{k-1}}A^{\mathrm{T}}+A\tilde{K}T_{k-1}\tilde{\Sigma}T^{\mathrm{T}}_{k-1}\tilde{K}^{\mathrm{T}}A^{\mathrm{T}}+A\tilde{K}\Phi_{k-1}\tilde{K}^{\mathrm{T}}A^{\mathrm{T}}\\ &-A\mathscr{C}^{21}_{\eta_{k-1}}\tilde{C}^{\mathrm{T}}T^{\mathrm{T}}_{k-1}\tilde{K}^{\mathrm{T}}A^{\mathrm{T}}-A\tilde{K}T_{k-1}\tilde{C}\mathscr{C}^{12}_{\eta_{k-1}}A^{\mathrm{T}}+Q\\ =&A\mathscr{C}^{22}_{\eta_{k-1}}A^{\mathrm{T}}+A\tilde{K}\tilde{\Sigma}\tilde{K}^{\mathrm{T}}A^{\mathrm{T}}-A\tilde{P}\tilde{C}^{\mathrm{T}}T^{\mathrm{T}}_{k-1}\tilde{K}^{\mathrm{T}}A^{\mathrm{T}}\\ &-A\tilde{K}T_{k-1}\tilde{C}\tilde{P}A^{\mathrm{T}}+Q,\end{aligned} \tag{14}$$

where the last equality is from the stealthiness constraint $T_{k-1}\tilde{\Sigma}T^{\mathrm{T}}_{k-1}+\Phi_{k-1}=\tilde{\Sigma}$ and $\mathscr{C}^{12}_{\eta_{k-1}}=\mathscr{C}^{21}_{\eta_{k-1}}=\tilde{P}$. Since $\tilde{e}_{\bar{k}|\bar{k}-1}$ and $\tilde{e}^a_{\bar{k}|\bar{k}-1}$ are identical, the initial state of the above recursion is $\mathscr{C}^{22}_{\eta_{\bar{k}}}=\tilde{P}$. From (8) and (10), the $(1,3)$th block is

$$\begin{aligned}\mathscr{C}^{13}_{\eta_k}=&\mathbb{E}[\tilde{e}_{k|k-1}(e^a_{k|k-1})^{\mathrm{T}}]\\ =&-(A-A\tilde{K}\tilde{C})\mathbb{E}[\tilde{e}_{k-1|k-2}\tilde{e}^{\mathrm{T}}_{k-1|k-2}]\tilde{C}^{\mathrm{T}}T^{\mathrm{T}}_{k-1}K_{\mathrm{II}}A^{\mathrm{T}}\\ &+(A-A\tilde{K}\tilde{C})\mathbb{E}[\tilde{e}_{k-1|k-2}(\tilde{e}^a_{k-1|k-2})^{\mathrm{T}}]\tilde{C}^{\mathrm{T}}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}}\\ &+(A-A\tilde{K}\tilde{C})\mathbb{E}[\tilde{e}_{k-1|k-2}(e^a_{k-1|k-2})^{\mathrm{T}}](A-AKC)^{\mathrm{T}}\\ &+A\tilde{K}\mathbb{E}[\tilde{v}_{k-1}\bar{v}^{\mathrm{T}}_{k-1}]K^{\mathrm{T}}_{\mathrm{I}}A^{\mathrm{T}}+A\tilde{K}\mathbb{E}[\tilde{v}_{k-1}\tilde{v}^{\mathrm{T}}_{k-1}]T^{\mathrm{T}}_{k-1}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}}\\ &+\mathbb{E}[w_{k-1}w^{\mathrm{T}}_{k-1}]\\ =&-(A-A\tilde{K}\tilde{C})\tilde{P}\tilde{C}^{\mathrm{T}}T^{\mathrm{T}}_{k-1}K_{\mathrm{II}}A^{\mathrm{T}}+A\tilde{K}S^{\mathrm{T}}K^{\mathrm{T}}_{\mathrm{I}}A^{\mathrm{T}}\\ &+A\tilde{K}\tilde{R}T^{\mathrm{T}}_{k-1}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}}+(A-A\tilde{K}\tilde{C})\mathscr{C}^{12}_{\eta_{k-1}}\tilde{C}^{\mathrm{T}}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}}\\ &+(A-A\tilde{K}\tilde{C})\mathscr{C}^{13}_{\eta_{k-1}}(A-AKC)^{\mathrm{T}}+Q\\ =&(A-A\tilde{K}\tilde{C})\mathscr{C}^{13}_{\eta_{k-1}}(A-AKC)^{\mathrm{T}}\\ &+(A-A\tilde{K}\tilde{C})\tilde{P}\tilde{C}^{\mathrm{T}}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}}+A\tilde{K}S^{\mathrm{T}}K^{\mathrm{T}}_{\mathrm{I}}A^{\mathrm{T}}+Q,\end{aligned} \tag{15}$$

where in the last equality we use $\mathscr{C}^{12}_{\eta_{k-1}} = \tilde{P}$ and a similar equation as (13) to eliminate two terms. To obtain the initial state of the above recursion, we consider the case without attacks, i.e., $T_k = I_{\tilde{m}},\ b_k = 0_{\tilde{m}\times 1}, \tilde{e}_{k-1|k-2} = \tilde{e}^a_{k-1|k-2}$. Then the dynamics in (10) become:

$$\begin{aligned} e_{k|k-1} =&(A-AKC)e_{k-1|k-2} - AK_{\mathrm{I}}\bar{v}_{k-1} - AK_{\mathrm{II}}\tilde{v}_{k-1} + w_{k-1} \\ =&(A-AKC)e_{k-1|k-2} - AKv_{k-1} + w_{k-1}. \end{aligned} \tag{16}$$

It follows from (8) and (16) that

$$\begin{aligned} \mathbb{E}[\tilde{e}_{k|k-1}e^{\mathrm{T}}_{k|k-1}] =&(A-A\tilde{K}\tilde{C})\mathbb{E}[\tilde{e}_{k-1|k-2}e^{\mathrm{T}}_{k-1|k-2}](A-AKC)^{\mathrm{T}} \\ &+A\tilde{K}\mathbb{E}[\tilde{v}_{k-1}v^{\mathrm{T}}_{k-1}]K^{\mathrm{T}}A^{\mathrm{T}} + \mathbb{E}[w_{k-1}w^{\mathrm{T}}_{k-1}]. \end{aligned}$$

Since both $A-A\tilde{K}\tilde{C}$ and $A-AKC$ are stable, the steady-state value of $\mathbb{E}[\tilde{e}_{k|k-1}e^{\mathrm{T}}_{k|k-1}]$ will converge to the solution of the matrix equation

$$\begin{aligned} X =&(A-A\tilde{K}\tilde{C})X(A-AKC)^{\mathrm{T}} \\ &+A\tilde{K}\begin{bmatrix} S^{\mathrm{T}} & \tilde{R}\end{bmatrix}K^{\mathrm{T}}A^{\mathrm{T}} + Q. \end{aligned} \tag{17}$$

The solution of (17) is denoted as $\bar{X}$. For the two terms in (15), we have

$$\begin{aligned} &(A-A\tilde{K}\tilde{C})\tilde{P}\tilde{C}^{\mathrm{T}}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} + A\tilde{K}S^{\mathrm{T}}K^{\mathrm{T}}_{\mathrm{I}}A^{\mathrm{T}} \\ =&A[(\tilde{P}\tilde{C}^{\mathrm{T}} - \tilde{K}\tilde{C}\tilde{P}\tilde{C}^{\mathrm{T}})K^{\mathrm{T}}_{\mathrm{II}} + \tilde{K}S^{\mathrm{T}}K^{\mathrm{T}}_{\mathrm{I}}]A^{\mathrm{T}} \\ \overset{(a)}{=}&A[(\tilde{K}\tilde{\Sigma} - \tilde{K}\tilde{C}\tilde{P}\tilde{C}^{\mathrm{T}})K^{\mathrm{T}}_{\mathrm{II}} + \tilde{K}S^{\mathrm{T}}K^{\mathrm{T}}_{\mathrm{I}}]A^{\mathrm{T}} \\ =&A\tilde{K}[(\tilde{\Sigma} - \tilde{C}\tilde{P}\tilde{C}^{\mathrm{T}})K^{\mathrm{T}}_{\mathrm{II}} + S^{\mathrm{T}}K^{\mathrm{T}}_{\mathrm{I}}]A^{\mathrm{T}} \\ =&A\tilde{K}\begin{bmatrix} S^{\mathrm{T}} & \tilde{R}\end{bmatrix}K^{\mathrm{T}}A^{\mathrm{T}}, \end{aligned}$$

where equality (a) is from $\tilde{P}\tilde{C}^{\mathrm{T}} = \tilde{K}\tilde{\Sigma}$. With this equality, we can verify from (15) and (17) that $\mathscr{C}^{13}_{\eta_k}$ is already in its steady state at instant $\bar{k}$. Thus $\mathscr{C}^{13}_{\eta_k} = \bar{X}, \forall k \geq \bar{k}$. From (9) and (10), the $(2,3)$th block satisfies

$$\begin{aligned} \mathscr{C}^{23}_{\eta_k} =&\mathbb{E}[\tilde{e}^a_{k|k-1}(e^a_{k|k-1})^{\mathrm{T}}] \\ =&-A\mathbb{E}[\tilde{e}^a_{k-1|k-2}\tilde{e}^{\mathrm{T}}_{k-1|k-2}]\tilde{C}^{\mathrm{T}}T^{\mathrm{T}}_{k-1}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} \\ &+A\mathbb{E}[\tilde{e}^a_{k-1|k-2}(\tilde{e}^a_{k-1|k-2})^{\mathrm{T}}]\tilde{C}^{\mathrm{T}}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} \\ &+A\mathbb{E}[\tilde{e}^a_{k-1|k-2}(e^a_{k-1|k-2})^{\mathrm{T}}](A-AKC)^{\mathrm{T}} \\ &+A\tilde{K}T_{k-1}\tilde{C}\mathbb{E}[\tilde{e}_{k-1|k-2}\tilde{e}^{\mathrm{T}}_{k-1|k-2}]\tilde{C}^{\mathrm{T}}T^{\mathrm{T}}_{k-1}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} \\ &-A\tilde{K}T_{k-1}\tilde{C}\mathbb{E}[\tilde{e}_{k-1|k-2}(\tilde{e}^a_{k-1|k-2})^{\mathrm{T}}]\tilde{C}^{\mathrm{T}}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} \\ &-A\tilde{K}T_{k-1}\tilde{C}\mathbb{E}[\tilde{e}_{k-1|k-2}(e^a_{k-1|k-2})^{\mathrm{T}}](A-AKC)^{\mathrm{T}} \\ &+A\tilde{K}T_{k-1}\mathbb{E}[\tilde{v}_{k-1}\bar{v}^{\mathrm{T}}_{k-1}]K^{\mathrm{T}}_{\mathrm{I}}A^{\mathrm{T}} + \mathbb{E}[w_{k-1}w^{\mathrm{T}}_{k-1}] \\ &+A\tilde{K}T_{k-1}\mathbb{E}[\tilde{v}_{k-1}\tilde{v}^{\mathrm{T}}_{k-1}]T^{\mathrm{T}}_{k-1}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} \\ &+A\tilde{K}\mathbb{E}[b_{k-1}b^{\mathrm{T}}_{k-1}]K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} \\ =&-A\mathscr{C}^{21}_{\eta_{k-1}}\tilde{C}^{\mathrm{T}}T^{\mathrm{T}}_{k-1}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} + A\mathscr{C}^{22}_{\eta_{k-1}}\tilde{C}^{\mathrm{T}}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} \\ &+A\mathscr{C}^{23}_{\eta_{k-1}}(A-AKC)^{\mathrm{T}} + A\tilde{K}T_{k-1}\tilde{C}\mathscr{C}^{11}_{\eta_{k-1}}\tilde{C}^{\mathrm{T}}T^{\mathrm{T}}_{k-1}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} \\ &-A\tilde{K}T_{k-1}\tilde{C}\mathscr{C}^{12}_{\eta_{k-1}}\tilde{C}^{\mathrm{T}}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} - A\tilde{K}T_{k-1}\tilde{C}\mathscr{C}^{13}_{\eta_{k-1}}(A-AKC)^{\mathrm{T}} \\ &+A\tilde{K}T_{k-1}S^{\mathrm{T}}K^{\mathrm{T}}_{\mathrm{I}}A^{\mathrm{T}} + A\tilde{K}T_{k-1}\tilde{R}T^{\mathrm{T}}_{k-1}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} \\ &+A\tilde{K}\Phi_{k-1}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} + Q. \end{aligned}$$

Then substituting $\mathscr{C}^{11}_{\eta_{k-1}} = \mathscr{C}^{12}_{\eta_{k-1}} = \mathscr{C}^{21}_{\eta_{k-1}} = \tilde{P},\ \mathscr{C}^{13}_{\eta_{k-1}} = \bar{X}$, and the stealthiness constraint $T_{k-1}\tilde{\Sigma}T^{\mathrm{T}}_{k-1} + \Phi_{k-1} = \tilde{\Sigma}$, we can simplify the above recursion as

$$\begin{aligned} \mathscr{C}^{23}_{\eta_k} =&A\mathscr{C}^{23}_{\eta_{k-1}}(A-AKC)^{\mathrm{T}} - A\tilde{K}T_{k-1}\tilde{C}\bar{X}(A-AKC)^{\mathrm{T}} \\ &+A\mathscr{C}^{22}_{\eta_{k-1}}\tilde{C}^{\mathrm{T}}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} + A\tilde{K}T_{k-1}S^{\mathrm{T}}K^{\mathrm{T}}_{\mathrm{I}}A^{\mathrm{T}} + A\tilde{K}\tilde{\Sigma}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} \\ &-A\tilde{P}\tilde{C}^{\mathrm{T}}T^{\mathrm{T}}_{k-1}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} - A\tilde{K}T_{k-1}\tilde{C}\tilde{P}\tilde{C}^{\mathrm{T}}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} + Q. \end{aligned} \tag{18}$$

Again, since $\tilde{e}_{\bar{k}|\bar{k}-1}$ and $\tilde{e}^a_{\bar{k}|\bar{k}-1}$ are identical, the initial state is $\mathscr{C}^{23}_{\eta_{\bar{k}}} = \bar{X}$. Finally, for the $(3,3)$th block, we have

$$\begin{aligned} \mathscr{C}^{33}_{\eta_k} =&\mathbb{E}[e^a_{k|k-1}(e^a_{k|k-1})^{\mathrm{T}}] \\ =&AK_{\mathrm{II}}T_{k-1}\tilde{C}\mathscr{C}^{11}_{\eta_{k-1}}\tilde{C}^{\mathrm{T}}T^{\mathrm{T}}_{k-1}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} + AK_{\mathrm{II}}\tilde{C}\mathscr{C}^{22}_{\eta_{k-1}}\tilde{C}^{\mathrm{T}}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} \\ &+(A-AKC)\mathscr{C}^{33}_{\eta_{k-1}}(A-AKC)^{\mathrm{T}} \\ &-AK_{\mathrm{II}}T_{k-1}\tilde{C}\mathscr{C}^{12}_{\eta_{k-1}}\tilde{C}^{\mathrm{T}}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} + AK_{\mathrm{II}}\tilde{C}\mathscr{C}^{23}_{\eta_{k-1}}(A-AKC)^{\mathrm{T}} \\ &-AK_{\mathrm{II}}T_{k-1}\tilde{C}\mathscr{C}^{13}_{\eta_{k-1}}(A-AKC)^{\mathrm{T}} \\ &-AK_{\mathrm{II}}\tilde{C}\mathscr{C}^{21}_{\eta_{k-1}}\tilde{C}^{\mathrm{T}}T^{\mathrm{T}}_{k-1}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} + (A-AKC)\mathscr{C}^{32}_{\eta_{k-1}}\tilde{C}^{\mathrm{T}}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} \\ &-(A-AKC)\mathscr{C}^{31}_{\eta_{k-1}}\tilde{C}^{\mathrm{T}}T^{\mathrm{T}}_{k-1}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} + AK_{\mathrm{II}}T_{k-1}S^{\mathrm{T}}K^{\mathrm{T}}_{\mathrm{I}}A^{\mathrm{T}} \\ &+AK_{\mathrm{I}}\bar{R}K^{\mathrm{T}}_{\mathrm{I}}A^{\mathrm{T}} + AK_{\mathrm{II}}T_{k-1}\tilde{R}T^{\mathrm{T}}_{k-1}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} \\ &+AK_{\mathrm{II}}\Phi_{k-1}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} + AK_{\mathrm{I}}ST^{\mathrm{T}}_{k-1}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} + Q. \end{aligned}$$

Then using the stealthiness constraint $T_{k-1}\tilde{\Sigma}T^{\mathrm{T}}_{k-1} + \Phi_{k-1} = \tilde{\Sigma}$, $\mathscr{C}^{11}_{\eta_{k-1}} = \mathscr{C}^{12}_{\eta_{k-1}} = \mathscr{C}^{21}_{\eta_{k-1}} = \tilde{P}$, $\mathscr{C}^{13}_{\eta_{k-1}} = \bar{X}$, and $\mathscr{C}^{31}_{\eta_{k-1}} = \bar{X}^{\mathrm{T}}$, we can simplify $\mathscr{C}^{33}_{\eta_k}$ as

$$\begin{aligned} \mathscr{C}^{33}_{\eta_k} =&(A-AKC)\mathscr{C}^{33}_{\eta_{k-1}}(A-AKC)^{\mathrm{T}} + AK_{\mathrm{II}}\tilde{C}\mathscr{C}^{22}_{\eta_{k-1}}\tilde{C}^{\mathrm{T}}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} \\ &+(A-AKC)\mathscr{C}^{32}_{\eta_{k-1}}\tilde{C}^{\mathrm{T}}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} + AK_{\mathrm{II}}\tilde{C}\mathscr{C}^{23}_{\eta_{k-1}}(A-AKC)^{\mathrm{T}} \\ &+AK_{\mathrm{II}}\tilde{\Sigma}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} + AK_{\mathrm{I}}\bar{R}K^{\mathrm{T}}_{\mathrm{I}}A^{\mathrm{T}} + AK_{\mathrm{I}}ST^{\mathrm{T}}_{k-1}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} \\ &-AK_{\mathrm{II}}T_{k-1}\tilde{C}\tilde{P}\tilde{C}^{\mathrm{T}}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} - AK_{\mathrm{II}}\tilde{C}\tilde{P}\tilde{C}^{\mathrm{T}}T^{\mathrm{T}}_{k-1}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} \\ &-AK_{\mathrm{II}}T_{k-1}\tilde{C}\bar{X}(A-AKC)^{\mathrm{T}} \\ &-(A-AKC)\bar{X}^{\mathrm{T}}\tilde{C}^{\mathrm{T}}T^{\mathrm{T}}_{k-1}K^{\mathrm{T}}_{\mathrm{II}}A^{\mathrm{T}} \\ &+AK_{\mathrm{II}}T_{k-1}S^{\mathrm{T}}K^{\mathrm{T}}_{\mathrm{I}}A^{\mathrm{T}} + Q, \end{aligned} \tag{19}$$

the initial state is $\mathscr{C}^{33}_{\eta_{\bar{k}}} = P$ since the filter in (4) is not altered and is in steady state at instant $\bar{k}$.

Now all nine blocks of $\mathscr{C}_{\eta_k} = \mathbb{E}[\eta_k\eta^{\mathrm{T}}_k]$ are obtained with the above recursions. It is interesting to notice that $\mathscr{C}^{12}_{\eta_k}$ and $\mathscr{C}^{13}_{\eta_k}$ are constant and not affected by linear attacks. Substituting these block matrices into (11), we can obtain the covariance of z^a_k, which plays a central role in the analysis of stealthiness constraints.

C. Stealthiness Constraints and Attack Performance

First, we partition $\mathscr{C}_{z^a_k}$ as a 2-by-2 block matrix according to the dimensions of $\bar{y}_k$ and $\tilde{y}_k$. Substituting $\mathbb{E}[\eta_k\eta^{\mathrm{T}}_k]$ into (11), we obtain the $(1,1)$th block,

$$\mathscr{C}^{11}_{z^a_k} = \bar{C}\mathscr{C}^{33}_{\eta_k}\bar{C}^{\mathrm{T}} + \bar{R}, \tag{20}$$

the $(1,2)$th block,

$$\mathscr{C}^{12}_{z^a_k} = (\bar{C}\mathscr{C}^{31}_{\eta_k}\tilde{C}^{\mathrm{T}} + S)T^{\mathrm{T}}_k + \bar{C}(\mathscr{C}^{33}_{\eta_k} - \mathscr{C}^{32}_{\eta_k})\tilde{C}^{\mathrm{T}}, \tag{21}$$

and the $(2,2)$th block,

$$\begin{aligned} \mathscr{C}^{22}_{z^a_k} =&T_k\tilde{C}(\mathscr{C}^{13}_{\eta_k} - \mathscr{C}^{12}_{\eta_k})\tilde{C}^{\mathrm{T}} + \tilde{C}(\mathscr{C}^{31}_{\eta_k} - \mathscr{C}^{21}_{\eta_k})\tilde{C}^{\mathrm{T}}T^{\mathrm{T}}_k \\ &+\tilde{C}(\mathscr{C}^{33}_{\eta_k} + \mathscr{C}^{22}_{\eta_k} - \mathscr{C}^{23}_{\eta_k} - \mathscr{C}^{32}_{\eta_k})\tilde{C}^{\mathrm{T}} + \tilde{\Sigma}. \end{aligned} \tag{22}$$

The constraint on $\mathscr{C}^{21}_{z^a_k}$ is omitted owing to the symmetry of $\mathscr{C}_{z^a_k}$. It is observed from (20) that $\mathscr{C}^{11}_{z^a_k}$ is determined by $\mathscr{C}^{33}_{\eta_k}$, which is not affected directly by T_k but by the attack coefficient in the previous step, namely, T_{k-1}, by recursion (19). Thus when designing T_k, we should take its impact on the stealthiness constraint in the subsequent step into account. Therefore, (6b) becomes

$$\mathscr{C}^{11}_{z^a_{k+1}} = \Sigma^{11} \triangleq \bar{C}P\bar{C}^{\mathrm{T}} + \bar{R}, \tag{23a}$$

$$\mathscr{C}^{12}_{z^a_k} = \Sigma^{12} \triangleq \bar{C}P\tilde{C}^{\mathrm{T}} + S, \tag{23b}$$

$$\mathscr{C}^{22}_{z^a_k} = \Sigma^{22} \triangleq \tilde{C}P\tilde{C}^{\mathrm{T}} + \tilde{R}. \tag{23c}$$

Note that (23a) is for ensuring the stealthiness property at instant $k+1$ by imposing an equality constraint on T_k and $\mathscr{C}^{11}_{z^a_{\bar{k}}} = \bar{C}P\bar{C}^{\mathrm{T}} + \bar{R}$ is fulfilled automatically since $\mathscr{C}^{33}_{\eta_{\bar{k}}} = P$. All above constraints are linear with respective to T_k.

The constraint in (6a) leads to $T_k\tilde{\Sigma}T_k^{\mathrm{T}} + \Phi_k = \tilde{\Sigma}$, which can be reformulated as the following LMI by eliminating Φ_k and applying the Schur complement:

$$\begin{bmatrix} \tilde{\Sigma} & T_k \\ T_k^{\mathrm{T}} & \tilde{\Sigma}^{-1} \end{bmatrix} \succeq 0. \tag{24}$$

For the remote state estimator, the *a posteriori* estimation error covariance evolves according to

$$\begin{aligned} P^a_{k|k} &= \mathbb{E}[(x_k - x^a_{k|k-1} - Kz^a_k)(x_k - x^a_{k|k-1} - Kz^a_k)^{\mathrm{T}}] \\ &= P^a_{k|k-1} + K\mathbb{E}[z^a_k(z^a_k)^{\mathrm{T}}]K^{\mathrm{T}} - \mathbb{E}[e^a_{k|k-1}(z^a_k)^{\mathrm{T}}]K^{\mathrm{T}} \\ &\quad - K\mathbb{E}[z^a_k(e^a_{k|k-1})^{\mathrm{T}}], \end{aligned}$$

where $P^a_{k|k-1} = \mathbb{E}[e^a_{k|k-1}(e^a_{k|k-1})^{\mathrm{T}}] = \mathscr{C}^{33}_{\eta_k}$. Since $\mathbb{E}[z^a_k(z^a_k)^{\mathrm{T}}] = \Sigma$ is constant and $P^a_{k|k-1}$ is not affected by the attack policy at instant k, in order to maximize $\mathrm{Trace}(P^a_{k|k})$, we need only to minimize the index function

$$\begin{aligned} J_k &= \mathrm{Trace}\{K\mathbb{E}[z^a_k(e^a_{k|k-1})^{\mathrm{T}}]\} \\ &= \mathrm{Trace}\{K\mathbb{E}[(\hat{C}_k\eta_k + \hat{T}_k v_k + Eb_k)(e^a_{k|k-1})^{\mathrm{T}}]\} \\ &= \mathrm{Trace}\{K\hat{C}_k\mathbb{E}[\eta_k(e^a_{k|k-1})^{\mathrm{T}}]\}. \end{aligned}$$

Substituting $\hat{C}_k$ and η_k yields

$$J_k = \mathrm{Trace}(K_{\mathrm{II}}T_k\tilde{C}\mathscr{C}^{13}_{\eta_k}) - \mathrm{Trace}(K_{\mathrm{II}}\tilde{C}\mathscr{C}^{23}_{\eta_k} - KC\mathscr{C}^{33}_{\eta_k}).$$

The last term is not affected by T_k, thus the objective is to design T_k that minimizes the first term.

D. Optimal Attack Policy

Summarizing the analysis in the previous section, the attacker needs to solve the following optimization problem step by step to obtain the optimal attack strategy:

$$\mathbf{P}_1: \quad \min_{T_k \in \mathbb{R}^{\tilde{m}\times\tilde{m}}} \ \mathrm{Trace}(\tilde{C}\mathscr{C}^{13}_{\eta_k}K_{\mathrm{II}}T_k) \quad \text{s.t. (23) and (24).}$$

The problem has a linear objective and all constraints are convex, thus can be solved efficiently with the Matlab CVX toolbox. At instant $\bar{k}$, constraint (23a) is dropped. After T^*_k is solved, we have $\Phi^*_k = \tilde{\Sigma} - T^*_k\tilde{\Sigma}(T^*_k)^{\mathrm{T}}$. The coefficients are substituted into (14), (18), and (19) to obtain $\mathscr{C}^{22}_{\eta_k}$, $\mathscr{C}^{23}_{\eta_k}$, and $\mathscr{C}^{33}_{\eta_k}$, respectively; then all parameters in (20)–(22) and J_k are updated and the attacker readily proceeds to solve $\mathbf{P}_1$ in the subsequent step. The attack performance is evaluated by

$$\begin{aligned} P^a_{k|k} &= AP^a_{k-1|k-1}A^{\mathrm{T}} + K\Sigma K^{\mathrm{T}} - K_{\mathrm{II}}T_k\tilde{C}\mathscr{C}^{13}_{\eta_k} - (\mathscr{C}^{31}_{\eta_k})\tilde{C}^{\mathrm{T}}T_k^{\mathrm{T}}K_{\mathrm{II}}^{\mathrm{T}} \\ &\quad + (K_{\mathrm{II}}\tilde{C}\mathscr{C}^{23}_{\eta_k} - KC\mathscr{C}^{33}_{\eta_k}) + (K_{\mathrm{II}}\tilde{C}\mathscr{C}^{23}_{\eta_k} - KC\mathscr{C}^{33}_{\eta_k})^{\mathrm{T}} + Q. \end{aligned} \tag{25}$$

Remark 1. $\mathbf{P}_1$ may have multiple optimal solutions; then there exist different optimal policies that can achieve the same attack performance. However there can be the case that the optimal solution is an identity matrix. In fact, if the unique solution at instant $\bar{k}$ is $T^*_{\bar{k}} = I_{\tilde{m}}$, then it is easy to verify $T^*_k = I_{\tilde{m}}, \forall k \geq \bar{k}$. It indicates that there does not exist such a deception attack that can deceive two anomaly detectors simultaneously and also degrade the estimation performance. These situations arise because the equality constraints in (23) are overly restrictive. For example, if $m \geq 2$ and $\tilde{m} = 1$, there exists only one unsafe channel that can be compromised by attackers; $T_k \in \mathbb{R}$ is a scalar. One can verify that $\bar{X} = P$ and $T^*_k = 1$ is the only coefficient that satisfies all equality constraints in (23). Then $\Phi^*_k = 0$, $\tilde{z}^a_k = \tilde{z}_k$, $\tilde{y}^a_k = \tilde{y}_k$, and $y^a_k = y_k$. No stealthy deception attacks can be launched in this case.

Remark 2. It can be verified that $T_{\bar{k}} = -I_{\tilde{m}}$ does not satisfy (23), thus flipping the sign of nominal innovations in the unsafe channel is not a stealthy policy and can be easily detected by anomaly detector II.

Remark 3. Compared with the scenario where all transmission links are vulnerable, the deployment of secured channels imposes stricter stealthiness constraints, thus limiting the set of feasible policies for adversaries. It is clear that securing all channels can completely prevent stealthy attacks but also induces extra costs. In practical cases, system defenders should consider the optimal tradeoff between the resources required for channel protection and performance degradation caused by malicious attacks.

IV. Numerical Examples

A stable LTI process is utilized to verify the effectiveness of the proposed strategy:

$$A = \begin{bmatrix} 0.4 & 0 & 0 & 0 & 0 \\ 0.08 & 0.4 & 0 & 0 & 0 \\ 0.73 & 0.39 & 0.8 & 0 & 0 \\ 0.87 & 0.98 & 0.48 & 0.67 & 0 \\ 0.46 & 0.94 & 0.95 & 0.81 & 0.4 \end{bmatrix},$$

$$C = \mathrm{diag}\{\begin{bmatrix} 1 & 3 & 2 & 2 & 4 \end{bmatrix}\},$$

$$Q = \mathrm{diag}\{\begin{bmatrix} 0.8 & 1 & 1.2 & 2 & 1.6 \end{bmatrix}\},$$

$$R = \mathrm{diag}\{\begin{bmatrix} 1 & 2 & 8 & 12 & 6 \end{bmatrix}\}.$$

In nominal conditions, the summation of steady state estimation errors is $\mathrm{Trace}(P_{k|k}) = 3.37$. We assume $\bar{m} = 1, \tilde{m} = 4$, $\bar{k} = 40$; there is only one secured channel and the remaining four can be compromised. The blue curve in Fig. 2 shows the evolution of $\mathrm{Trace}(P^a_{k|k})$ with attack (5). As a comparison,

the red curve is the performance of the optimal linear attack ($z_k^a = -z_k$) in [6], where all five channels are vulnerable to malicious agents. The steady state values of Trace($P_{k|k}^a$) drop from 631 to 123 in the two cases. It is seen that even if only a small portion (20%) of original data is secured, the worst-case attack effects can be significantly reduced. This observation highlights the necessity and effectiveness of investing limited resources to securing partial transmission channels in real applications.

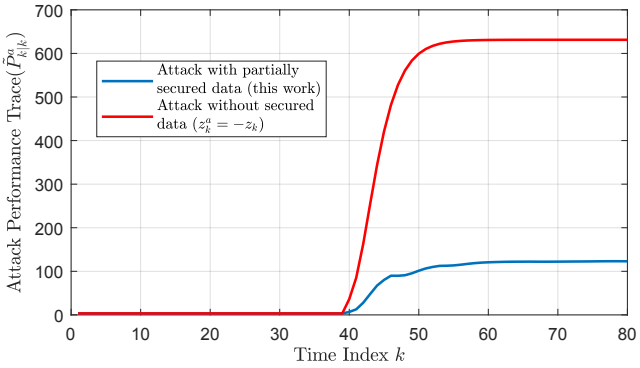

Fig. 2. Optimal attack performances with/without secured channel.

Let the threshold of anomaly detector II be 6, which leads to a theoretical false alarm rate $\alpha = 0.3062$. The LTI process runs for 20000 times with randomly generated noises. Fig. 3 indicates the alarm rate (AR) before and after the attack occurs. The empirical AR is always consistent with the theoretical one, illustrating the stealthiness property. As a comparison, if the attacker adopts the optimal strategy in [6] and simply flips the sign of nominal innovations in the unsafe channel ($\tilde{z}_k^a = -\tilde{z}_k$), Fig. 4 shows that this attack cannot remain stealthy as a greater AR is induced. Note that both attacks can completely deceive anomaly detector I.

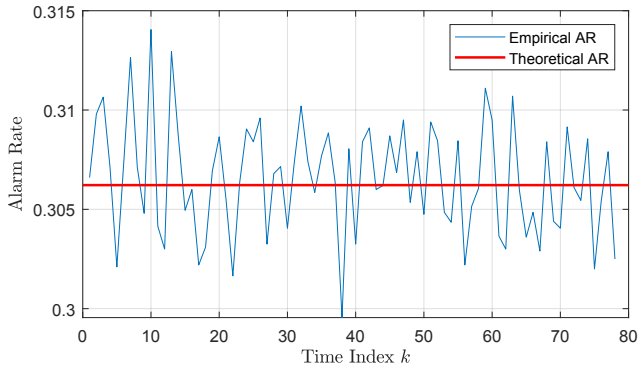

Fig. 3. Alarm rate of the optimal attack in (5).

V. Conclusions

In this work, we study the optimal linear attack strategy and mainly focus on analyzing how data manipulation in unsafe channels impacts the innovation of the remote state estimator. The worst-case attack policy is obtained and it shows that protecting a small portion of channels can significantly reduce the worst-case attack impact. The future work would include investigating the feasibility and methods to make full utilization of secured channels to completely prevent these stealthy deception attacks.

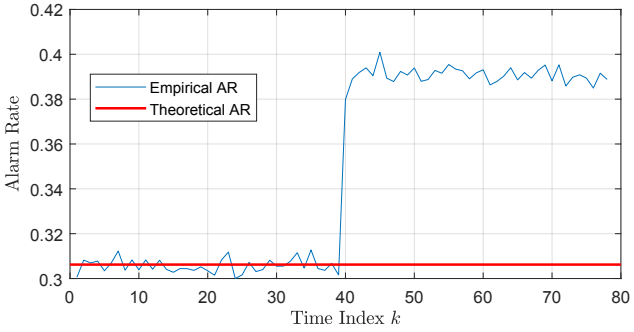

Fig. 4. Alarm rate of the attack policy $T_k = -I_{\tilde{m}}, \Phi_k = 0_{\tilde{m}}$.

References

[1] H. Menouar, I. Guvenc, K. Akkaya, A. S. Uluagac, A. Kadri, and A. Tuncer, "Uav-enabled intelligent transportation systems for the smart city: Applications and challenges," *IEEE Commun. Mag.*, vol. 55, no. 3, pp. 22–28, 2017.

[2] Y. Nakahira and Y. Mo, "Attack-resilient H_2, H_∞, and L_1 state estimator," *IEEE Trans. Autom. Control*, vol. 63, no. 12, pp. 4353–4360, 2018.

[3] R. Romagnoli, S. Weerakkody, and B. Sinopoli, "A model inversion based watermark for replay attack detection with output tracking," in *Proc. Am. Control Conf.*, Philadelphia, USA, 2019, pp. 384–390.

[4] G. Hug and J. A. Giampapa, "Vulnerability assessment of AC state estimation with respect to false data injection cyber-attacks," *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1362–1370, 2012.

[5] D. G. Eliades and M. M. Polycarpou, "A fault diagnosis and security framework for water systems," *IEEE Trans. Control Syst. Technol.*, vol. 18, no. 6, pp. 1254–1265, 2009.

[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 stealthy innovation-based linear attack on remote state estimation," *Automatica*, vol. 89, pp. 117–124, 2018.

[8] ——, "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.

[9] J. Zhou, J. Shang, and T. Chen, "Optimal linear FDI attacks with side information: A comparative study," in *Proc. 4th Int. Conf. Ind. Cyber-Phys. Syst.*, Vancover, CA, 2021, pp. 138–143.

[10] ——, "Optimal deception attacks against remote state estimation: An information-based approach," *IEEE Trans. Autom. Control*, 2022.

[11] 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.

[12] H. Liu, Y. Ni, L. Xie, and K. H. Johansson, "An optimal linear attack strategy on remote state estimation," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 3527–3532, 2020.

[13] 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.

[14] Y. Li, L. Shi, and T. Chen, "Detection against linear deception attacks on multi-sensor remote state estimation," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 3, pp. 846–856, 2018.

[15] A. Chattopadhyay and U. Mitra, "Security against false data injection attack in cyber-physical systems," *IEEE Trans. Control Netw. Syst.*, vol. 7, no. 2, pp. 1015–1027, 2020.

[16] 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.

[17] B. D. Anderson and J. B. Moore, *Optimal Filtering*. Courier Corporation, 2012.