This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TAC.2026.3672354, IEEE Transactions on Automatic Control

Remote State Estimation under Stochastic Stealthy Attacks: Short-Term Optimization and Long-Term Convergence Analysis

Lubin Zhang, Jun Shang, Zidong Wang, and Qinyuan Liu

***Abstract*—This paper investigates the problem of remote state estimation in cyber-physical systems subject to stochastic stealthy attacks. Unlike existing studies that assume persistent intrusion, the attack success is modeled as a stochastic process, thereby providing a more realistic characterization of adversarial capabilities. A comprehensive analysis is conducted from both short-term and long-term perspectives. In the short-term analysis, the evolution of the estimation error covariance is examined, and optimal attack strategies are derived under explicit stealthiness constraints, which limit the detection probability of the attacker. In the long-term analysis, the conditions under which the expected estimation error covariance diverges or converges are explored as a function of the attack success rate and strategy. Rigorous necessary, sufficient, and equivalent conditions for error covariance divergence are established. Moreover, the convergence behavior of the estimation process is characterized under various attack designs, revealing critical thresholds and trade-offs between attack frequency and intensity. Simulation results are provided to validate the theoretical findings and to illustrate the quantitative impact of attack parameters on estimation performance degradation.**

***Index Terms*—Remote state estimation, stochastic stealthy attacks, Kalman filter, short-term optimization, long-term convergence.**

I. Introduction

Cyber-Physical Systems (CPSs) are composed of integrated computational and physical components that are interconnected through advanced sensing, communication, and control mechanisms, thereby enabling intelligent automation across critical infrastructures [5], [54]. However, the widespread deployment of CPSs, particularly in power grids, water distribution systems, and industrial control networks, has raised increasing concerns regarding cybersecurity risks [41], [59]. Among these risks, remote state estimation (RSE), which serves as a foundational technology in CPSs, has been identified as particularly vulnerable due to the inherent insecurity of wireless communication channels [3], [4]. As a result, the security of RSE has attracted significant research attention over the past decade. Existing attack models targeting RSE are primarily classified into two categories: Denial-of-Service (DoS) attacks and integrity attacks [44]. DoS attacks impair system performance by jamming communication channels and thereby blocking the transmission of critical data [18], [33], [51], [56]. In contrast to the typically detectable nature of DoS attacks, deception-based integrity attacks manipulate transmitted data while remaining covert, thereby posing a more insidious threat to the reliability and security of CPSs [14], [17], [30], [48].

With the widespread deployment of residual-based false data detectors in modern control systems, increasing research attention has been directed toward stealthy attacks that are specifically designed to evade detection while maximizing the mean square error (MSE) of state estimates, either instantaneously or over a finite time horizon. Existing studies on stealthy attacks can be classified according to their stealthiness constraints into two main categories: strict stealthiness [10], [11], [36] and relaxed stealthiness [1], [12], [19], [21], [37], [52]. Under strict stealthiness, it is required that the residual sequences under attack follow the same probability distribution as those observed during normal operation. In contrast, relaxed stealthiness permits bounded deviations in residual behavior, typically quantified by the Kullback–Leibler (KL) divergence being constrained below a predefined threshold δ. These studies have demonstrated the considerable adaptability and disruptive potential of stealthy attacks in degrading the performance of remote state estimation. However, it has also been noted that the actual capabilities of adversaries (particularly in terms of eavesdropping and data tampering) may have been overestimated in the literature, especially under scenarios involving secured communication channels [8], [9], [20], [49] or when subject to energy limitations [6], [25]–[27], [35], [57].

While existing studies have relaxed attacker assumptions by incorporating resource constraints (e.g., energy budgets) and structural constraints (e.g., partially secured channels), the resulting models still implicitly assume deterministic attack outcomes. A critical yet underexplored dimension of stealthy attacks lies in their inherently stochastic success, which arises from uncertainties in attack execution. Due to limitations in wireless transmission and the computational capabilities of adversarial devices, consistent signal interception cannot be

This work was supported in part by the National Natural Science Foundation of China under Grants 62222312, 62473285, and 62303353, in part by the Fundamental Research Funds for the Central Universities of China, in part by the Royal Society of the UK, and the Alexander von Humboldt Foundation of Germany. *(Corresponding author: Qinyuan Liu.)*

Lubin Zhang and Qinyuan Liu are with the School of Computer Science and Technology, Tongji University, Shanghai 201804, China, and also with the Key Laboratory of Embedded System and Service Computing, Ministry of Education, Shanghai 200092, China (e-mail: liuqy@tongji.edu.cn).

Jun Shang is with the Department of Control Science and Engineering, Shanghai Institute of Intelligent Science and Technology, State Key Laboratory of Autonomous Intelligent Unmanned Systems, and Frontiers Science Center for Intelligent Autonomous Systems, Tongji University, Shanghai 200092, China (e-mail: shangjun@tongji.edu.cn).

Zidong Wang is with the Department of Computer Science, Brunel University of London, Uxbridge, Middlesex, UB8 3PH, United Kingdom (e-mail: zidong.Wang@brunel.ac.uk).

For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

guaranteed, nor can it be ensured that tampered signals are always accepted by the remote state estimator. To account for this, the success of each attack (i.e., whether signals transmitted over wireless channels are successfully compromised) is modeled as a random variable, thereby capturing the uncontrollable randomness characteristic of real-world attack scenarios. Another limitation of the existing literature is its predominant emphasis on short-term perspectives, often focused on maximizing the estimation error covariance over one-step or finite-step horizons. However, the convergence behavior of the estimation covariance, particularly in unstable systems, constitutes a fundamental criterion for evaluating system security and attack severity. An investigation into the factors influencing convergence or divergence is thus essential, as it provides critical insights for the development of effective defensive strategies or for enhancing the effectiveness of adversarial attack policies.

Stochastic DoS attacks on remote state estimation have been extensively investigated over the past decades. Under the Bernoulli packet dropout model, the fundamental stability threshold for linear systems has been derived [38], and the exact critical attack probability for detectable systems has been established [34]. Furthermore, the Markovian dropout model has also been explored to capture temporal dependencies, leading to results on stability criteria and critical phase transitions [16], [46], [47], [53]. While existing literature provides a theoretical foundation, a critical limitation exists: DoS attacks are non-stealthy. The resulting measurement outage is readily detectable via residual-based false data detectors.

Compared to DoS attacks, stochastic deception attacks pose unique challenges by covertly compromising data integrity rather than merely disrupting availability. Early research models these threats as false data injection processes governed by Bernoulli sequences [7], [28], [40]. To address more complex scenarios, subsequent studies have investigated hybrid frameworks integrating deception strategies with channel anomalies [32], [50], [55]. However, the existing literature predominantly treats stochastic deception attacks as non-strategic bounded disturbances. Consequently, the analysis of worst-case performance degradation under stealthiness constraints, the investigation of the coupling effect between attack strategies and success rates on system convergence, and the identification of the minimal system knowledge required to design divergence-inducing attacks, remain unexplored. Addressing these issues provides theoretical guidelines for the security of remote state estimation.

The introduction of stochastic attacks complicates the system analysis, as the current estimation error becomes coupled with the entire history of stochastic attack outcomes. Consequently, the derivation of the resulting stochastic estimation error covariance becomes non-trivial and requires rigorous treatment. Furthermore, the system's long-term stability characterization is significantly complicated by the varied and adaptable attack strategies employed by persistent stealthy attacks, demanding a comprehensive evaluation of the system's stability properties.

Motivated by the above considerations, remote state estimation under stochastic stealthy attacks is systematically examined, with attention given to both short-term dynamics and long-term convergence behavior. The principal contributions of this work are summarized as follows.

1) The closed-form recursive solution for the estimation error covariance is derived under stochastic stealthy attacks, through which the ill-posed stochastic optimization problem for determining the one-step optimal attack is analytically decoupled. Furthermore, the one-step optimal attack strategy under relaxed stealthiness constraints is characterized.
2) It is proven that a finite sequence of stealthy attacks cannot lead to divergence in the evolution of the error covariance, even in the absence of defensive countermeasures.
3) The impact of persistent stochastic stealthy attacks on the convergence of remote state estimation is analyzed. Specifically: (i) a necessary condition, (ii) a sufficient condition, and (iii) a complete necessary and sufficient condition for convergence are established. Moreover, closed-form expressions are developed to quantify how attack success probabilities and specific attack designs influence convergence behavior.

The remainder of the paper is organized as follows. Section II presents the problem formulation. In Section III, the evolution of the estimation error covariance under stochastic stealthy attacks is analyzed, and the one-step optimal attack strategy under stealthiness constraints is derived. Section IV investigates the convergence properties of remote state estimation, considering both finite attack scenarios and persistent stochastic attacks. Numerical examples are provided in Section V to illustrate the theoretical results. Finally, the paper is concluded in Section VI.

Notations: $\mathbb{N}$ and $\mathbb{R}$ are the sets of natural and real numbers, respectively. $\mathbb{R}^p$ is the set of all p-dimensional real column vectors and $\mathbb{R}^{n\times n}$ is the set of all $n\times n$ real matrices. Let $\mathbb{S}^n_{++}(\mathbb{S}^n_{+})$ denote the set of $n\times n$ positive definite (semi-definite) matrices. $X\in\mathbb{S}^n_{++}(\mathbb{S}^n_{+})$ indicates $X\succ 0(X\succeq 0)$. $\mathrm{Tr}(X), |X|, \sigma_i(x)$, $\rho(X)$, and $\|X\|_2$ denote the trace, determinant, the ith largest eigenvalue (in modulus), spectral radius, and the matrix 2-norm of X, respectively. $X^\top$ is the transpose of X. $\mathcal{N}(\mu,\Sigma)$ represents the Gaussian distribution with mean μ and covariance Σ.

II. Problem Formulation

A. Remote Estimator

Consider the linear discrete time-invariant system:

$$\begin{aligned} x_{k+1} &= Ax_k+\omega_k \\ y_k &= Cx_k+v_k \end{aligned} \tag{1}$$

where $x_k\in\mathbb{R}^n$ is the system state, $y_k\in\mathbb{R}^m$ is the sensor measurement, and $\omega_k\in\mathbb{R}^n$ and $v_k\in\mathbb{R}^m$ are zero-mean independent and identically distributed (i.i.d.) Gaussian noises with covariance $Q\succ 0, R\succ 0$, respectively. The initial state $x_0\in\mathbb{R}^n$ follows a zero-mean Gaussian distribution with covariance $\Pi_0\succeq 0$, independent of ω_k and $v_k, \forall k\in\mathbb{N}$. It is assumed that the pair (A,C) is detectable and that $\mathrm{rank}(C)=m$.

At each time instant k, the measurement y_k is transmitted through a wireless channel to a remote estimation center, where a standard Kalman filter is employed for optimal state estimation as follows:

$$\begin{aligned} \hat{x}_k^- &= A\hat{x}_{k-1} \\ P_k^- &= AP_{k-1}A^\top + Q \\ K_k &= P_k^- C^\top (CP_k^- C^\top + R)^{-1} \\ z_k &= y_k - C\hat{x}_k^- \\ \hat{x}_k &= \hat{x}_k^- + K_k z_k \\ P_k &= (I - K_k C)P_k^- \end{aligned} \tag{2}$$

Here, $\hat{x}_k^-$ and $\hat{x}_k$ denote the *a priori* and a *posteriori* state estimates, respectively, with P_k^- and P_k denoting their corresponding estimation error covariances. The estimation error covariance of the Kalman filter is known to converge to a unique steady-state value, irrespective of the initial condition. Let the steady-state *a priori* estimation error covariance be defined as

$$\bar{P} = \lim_{k\to\infty} P_k^- \tag{3}$$

where $\bar{P}$ is the unique positive semi-definite solution of

$$X = h \circ \tilde{g}(X) \tag{4}$$

with

$$\begin{aligned} h(X) &\triangleq AXA^\top + Q, \\ \tilde{g}(X) &\triangleq X - XC^\top (CXC^\top + R)^{-1} CX. \end{aligned}$$

It is assumed that the Kalman filter operates at steady state prior to the onset of any attack, i.e., $P_0^- = \bar{P}$. Let $\bar{\Sigma} = C\bar{P}C^\top + R$ denote the covariance matrix of z_k and let $\bar{K} = \bar{P}C^\top \bar{\Sigma}^{-1}$ represent the steady-state Kalman gain. Given the Gaussian assumption for the initial state and system noises, the innovation z_k is also Gaussian distribution, i.e., $z_k \sim \mathcal{N}(0, \bar{\Sigma})$.

B. Stochastic Stealthy Attacks

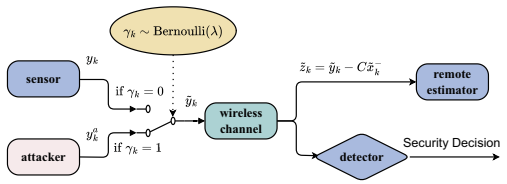


Fig. 1. System architecture for remote state estimation under stochastic stealthy attacks

As illustrated in Fig. 1, the wireless communication channel is characterized by inherent unreliability and vulnerability. First, the data transmitted from the sensor to the remote estimator are subject to eavesdropping by an attacker. The attacker then attempts to alter y_k into a manipulated signal y_k^a. Due to the probabilistic nature of attack execution, the remote estimator receives the manipulated signal y_k^a if $\gamma_k = 1$, or the original signal y_k if $\gamma_k = 0$. To denote potentially compromised estimation variables, a tilde notation is adopted for vectors and covariances, such as $\tilde{y}_k,\ \tilde{z}_k,\ \tilde{x}_k^-\ ,\tilde{x}_k,\ \tilde{e}_k,\ \tilde{e}_k^-,\ \tilde{P}_k,\ \tilde{P}_k^-$.

The primary objective of the attacker is to degrade estimation performance, which is quantified by $\mathrm{Tr}(\tilde{P}_k)$. It is assumed that the attack begins at the sampling instant $k = 0$. Prior to proceeding, the following assumptions are introduced.

Assumption 1. *At the onset of the attack, the attacker has complete knowledge of all system parameters as well as the initial state estimate at the remote estimator, i.e., $\hat{x}_0$.*

Assumption 2. *Through eavesdropping on the communication channels, the attacker can acquire sensor information and potentially manipulate the transmitted data y_k.*

Although Assumption 1 is strong, it facilitates the derivation of the one-step optimal attack strategy and reveals the worst-case impact on remote state estimation. This worst-case scenario serves as a theoretical benchmark. Furthermore, the necessity of this assumption for inducing estimation divergence will be examined later.

Most existing literature assumes deterministic attack success, under which malicious attempts are always effective when initiated. However, this idealization neglects practical constraints in real-world scenarios. To more accurately characterize the influence of stealthy attacks, the success of an attack at each time step k is modeled probabilistically. Specifically, the event that a communication channel is compromised is represented by a Bernoulli process γ_k with parameter λ:

$$\gamma_k = \begin{cases} 1, & \text{if the communication is compromised,} \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

The sequence $\{\gamma_k\}$ is independent over time and statistically independent of the process noise $\{\omega_k\}$ and measurement noise $\{v_k\}$. In this setting, the signal $\tilde{y}_k$ received by the remote estimator is given by:

$$\tilde{y}_k = \gamma_k y_k^a + (1 - \gamma_k) y_k. \tag{6}$$

That is, the original measurement y_k is received when no attack occurs, while the compromised signal y_k^a is received during a successful attack. According to the Kalman filter (2), the corresponding innovation $\tilde{z}_k$ derived from $\tilde{y}_k$ is expressed as:

$$\tilde{z}_k = \gamma_k (y_k^a - C\tilde{x}_k^-) + (1 - \gamma_k)(y_k - C\tilde{x}_k^-). \tag{7}$$

For notational convenience, let $z_k^a = y_k^a - C\tilde{x}_k^-$, which represents the manipulated innovation that the attacker injects into the filter. Under Assumptions 1 and 2, the attacker has the capability to precisely generate and inject any desired z_k^a. This is achieved by manipulating the measurement as $y_k^a = z_k^a + C\tilde{x}_k^-$ using its knowledge of C and $\tilde{x}_k^-$. The specific model describing how the attacker designs z_k^a will be formally introduced later in (10). Accordingly, the Kalman filter in the presence of deception attacks is formulated as:

$$\begin{aligned} \tilde{x}_k^- &= A\tilde{x}_{k-1} \\ \tilde{z}_k &= \gamma_k z_k^a + (1 - \gamma_k)(y_k - C\tilde{x}_k^-) \\ \tilde{x}_k &= \tilde{x}_k^- + \bar{K}\tilde{z}_k \\ &= \tilde{x}_k^- + \gamma_k \bar{K} z_k^a + (1 - \gamma_k)\bar{K}(y_k - C\tilde{x}_k^-) \end{aligned} \tag{8}$$

where the Kalman gain is fixed as $K = \bar{K}$, since the estimation error covariance has reached steady state.

The definitions of the prior and posterior estimation errors and their corresponding covariances under deception attacks are given by:

$$\begin{aligned} \tilde{e}_k^- &= x_k - \tilde{x}_k^-, \quad \tilde{P}_k^- = \mathbb{E}[\tilde{e}_k^-(\tilde{e}_k^-)^\top] \\ \tilde{e}_k &= x_k - \tilde{x}_k, \quad \tilde{P}_k = \mathbb{E}[\tilde{e}_k(\tilde{e}_k)^\top]. \end{aligned} \tag{9}$$

Inspired by [10], a linear attack model is adopted:

$$z_k^a = T_k z_k + b_k \tag{10}$$

where the matrix $T_k \in \mathbb{R}^{m\times m}$ is to be designed and b_k is Gaussian white noise with zero mean and covariance Θ_k, which is also a design parameter chosen by the attacker. The covariance matrix of z_k^a is:

$$\Sigma_k^a = T_k \bar{\Sigma} T_k^\top + \Theta_k. \tag{11}$$

Based on the Bernoulli attack model in (5) and the linear attack structure in (10), the stochastic linear attack model considered in this paper is expressed as:

$$\tilde{z}_k = \gamma_k(T_k z_k + b_k) + (1-\gamma_k)(y_k - C\tilde{x}_k^-). \tag{12}$$

C. Stealthiness Constraints

The relaxed stealthiness constraint is measured by the following Kullback–Leibler divergence:

$$D_{\mathrm{KL}}(z_k^a \| z_k) = \int_{\mathbb{R}^m} f_{z_k^a}(t) \ln \frac{f_{z_k^a}(t)}{f_{z_k}(t)} dt \tag{13}$$

where $f_{z_k^a}(t)$ and $f_{z_k}(t)$ denote the probability density functions of z_k^a and z_k, respectively. A linear attack (10) is said to be stealthy if

$$D_{\mathrm{KL}}(z_k^a \| z_k) \le \delta \tag{14}$$

where δ is the fixed threshold.

D. Problem of Interest

In this work, attack-induced performance degradation is evaluated using the estimation error covariance matrix:

$$\tilde{P}_k = \mathbb{E}[(x_k - \tilde{x}_k)(x_k - \tilde{x}_k)^\top].$$

Both short-term and long-term performance aspects are considered, as discussed below.

Short-term perspective issues: A central focus in existing literature has been placed on optimizing the posterior estimation error covariance [8]–[10], [12], [37]. One-step optimization is particularly important, as it facilitates the assessment of immediate performance degradation and serves as a foundation for analyzing long-term convergence properties. Therefore, the first problem addressed in this study is formulated in (15), which defines the one-step optimal attack as the solution that maximizes the trace of the posterior estimation error covariance while satisfying the stealthiness constraint imposed by the KL divergence bound δ.

$$\begin{aligned} \textbf{P1:} \quad & \max \quad \mathrm{Tr}(\tilde{P}_k) \\ & \text{s.t.} \quad D_{\mathrm{KL}}(z_k^a \| z_k) \le \delta. \end{aligned} \tag{15}$$

Here, $\tilde{P}_k$ denotes a random matrix jointly determined by $\gamma_{1:k} = \{\gamma_1, \gamma_2, \cdots, \gamma_k\}$. Before optimizing $\mathrm{Tr}(\tilde{P}_k)$, it is necessary to derive the relationship between $\tilde{P}_k^-$ and $\tilde{P}_k$ in terms of $\gamma_{1:k}$ and the prior attack design, i.e., $T_{1:k-1}$ and $\Theta_{1:k-1}$), since the coupling effect of stochastic stealthy attacks remains analytically unclear. Optimizing the stochastic objective function $\mathrm{Tr}(\tilde{P}_k)$ in P1 seems to be ill-posed. However, as will be demonstrated in Theorem 1, we can prove that P1 is equivalent to a deterministic optimization problem P2. A detailed analysis will be presented in Section III.

Long-term perspective issues: An equally critical issue involves analyzing the convergence behavior of $\tilde{P}_k$, which serves as a fundamental metric for evaluating both estimator robustness and the long-term efficacy of stealthy attacks. This aspect has been largely neglected in existing research due to the widespread assumption of deterministic attack success, under which $\tilde{P}_k$ diverges inevitably. However, under persistent stochastic stealthy attacks, convergence behavior presents a challenging and open research question.

In the extreme scenario where $\lambda = 0$, the system reduces to a standard Kalman filter, and convergence is guaranteed. Conversely, when $\lambda = 1$, divergence is ensured. This raises the fundamental question: how does the convergence of the estimation error covariance depend on the attack success probability $\lambda \in [0, 1]$?

Before examining the relationship between λ and the convergence of $\tilde{P}_k^-$, another essential issue must be addressed: whether it is possible for an attacker to design a finite-step attack that prevents convergence, even in the absence of subsequent attacks. If such an attack exists, the necessity of persistent attacks becomes a secondary concern.

In summary, this study addresses three core research challenges:

1) Characterizing the evolution of the estimation error covariance matrix under stochastic stealthy attacks and deriving one-step optimal stealthy attack strategies.
2) Establishing whether finite-step stealthy attacks can induce divergence of the estimation error covariance.
3) Identifying critical thresholds for the attack success probability λ and attack design parameters that govern convergence or divergence, and quantifying the relative disruption caused by one-step optimal attacks in comparison to generic malicious strategies.

III. One-Step Optimization of Attacks

In this section, the evolution of the estimation error covariance under stochastic stealthy attacks is analyzed, and the one-step optimal attack strategy under stealthiness constraints is derived. Specifically, a one-step optimal attack at time step k is defined as an attack that maximizes the trace of the error covariance matrix $\tilde{P}_k$. The analysis is conducted from the perspective of the remote estimator to evaluate the maximum potential impact of the attack, thereby establishing a theoretical foundation for the subsequent convergence analysis.

It is noted that the estimation error covariance at time instant k depends not only on the current attack parameters, but also on the cumulative influence of all preceding attack designs from time steps 1 to $k-1$. This dependency introduces significant complexity due to the stochastic nature of attack

success at each time step. To facilitate a rigorous analysis of how historical attack designs propagate through the estimation error covariance, the following lemma is established as an analytical foundation.

Lemma 1. *For the system depicted in Fig. 1, under the stochastic linear attack model* (12)*, the following holds:*

$$\mathbb{E}[z_i z_k^\top] = 0 \text{ and } \mathbb{E}[\tilde{z}_i (z_k^a)^\top] = 0, \text{ for } 0 \le i < k. \tag{16}$$

Proof: According to the properties of Kalman filtering, the innovation sequence $\{z_k\}$ is a zero-mean white sequence. Therefore, for $i < k$, we have $\mathbb{E}[z_i z_k^\top] = 0$. Since γ_i is independent of y_i, z_k^a and $\tilde{x}_i$, one obtains:

$$\begin{aligned}
&\mathbb{E}[\tilde{z}_i (z_k^a)^\top] \\
&= \lambda \mathbb{E}[z_i^a (z_k^a)^\top] + (1-\lambda)\mathbb{E}[(z_i + C\hat{x}_i^- - C\tilde{x}_i^-)(z_k^a)^\top] \\
&= -(1-\lambda) C \mathbb{E}[(\tilde{x}_i - \hat{x}_i^-)(z_k^a)^\top]
\end{aligned}$$

where the last equality follows from $\mathbb{E}[z_i z_k^\top] = 0$ and the independence of the white noise b_k from z_i for $\forall i < k$. Furthermore, the following relation holds:

$$\begin{aligned}
&z_i - \tilde{z}_i \\
&= (\gamma_i + 1 - \gamma_i) z_i - \gamma_i (T_i z_i + b_i) - (1-\gamma_i)(y_i - C\tilde{x}_i^-) \\
&= \gamma_i (I - T_i) z_i - \gamma_i b_i + (1-\gamma_i) C (\tilde{x}_i^- - \hat{x}_i^-).
\end{aligned}$$

Substituting the above expression into $\tilde{x}_i^- - \hat{x}_i^-$ yields:

$$\begin{aligned}
\tilde{x}_i^- - \hat{x}_i^- &= A(\tilde{x}_{i-1}^- - \hat{x}_{i-1}^-) + A\bar{K}(\tilde{z}_{i-1} - z_{i-1}) \\
&= A[I - (1-\gamma_{i-1})\bar{K}C](\tilde{x}_{i-1}^- - \hat{x}_{i-1}^-) \\
&\quad -\gamma_{i-1} A\bar{K}(I - T_{i-1}) z_{i-1} + \gamma_{i-1} A\bar{K} b_{i-1}.
\end{aligned}$$

Assuming $\mathbb{E}[(\tilde{x}_{i-1}^- - \hat{x}_{i-1}^-)(z_k^a)^\top] = 0$, it follows that:

$$\begin{aligned}
&\mathbb{E}[(\tilde{x}_i^- - \hat{x}_i^-)(z_k^a)^\top] \\
&= \mathbb{E}[A(I - (1-\gamma_{i-1})\bar{K}C)(\tilde{x}_{i-1}^- - \hat{x}_{i-1}^-)(z_k^a)^\top] \\
&\quad -\mathbb{E}[(\gamma_{i-1} A\bar{K}(I - T_{i-1}) z_{i-1} - \gamma_{i-1} A\bar{K} b_{i-1})(z_k^a)^\top] = 0.
\end{aligned}$$

Moreover, when $i = 0$, it holds that $\mathbb{E}[(\tilde{x}_i^- - \hat{x}_i^-)(z_k^a)^\top] = 0$ as the attack has not yet affected the estimator. By induction, it is concluded that $\mathbb{E}[(\tilde{x}_i^- - \hat{x}_i^-)(z_k^a)^\top] = 0$ for all $i < k$, which implies $\mathbb{E}[\tilde{z}_i (z_k^a)^\top] = 0$. ∎

Lemma 2. *([37]) The optimization problem*

$$\begin{aligned}
\max_{T_k, \Sigma_k^a} \quad & \mathrm{Tr}(\bar{K}\Sigma_k^a \bar{K}^\top) - 2\mathrm{Tr}(\bar{K} T_k C \bar{P}) \\
\text{s.t.} \quad & T_k \bar{\Sigma} T_k^\top - \Sigma_k^a \preceq 0 \\
& D_{\mathrm{KL}}(z_k^a \| z_k) \le \delta
\end{aligned} \tag{17}$$

has an optimal solution

$$\begin{aligned}
T_k^* &= -\left(I - \frac{1}{\eta} C \bar{P}^2 C^\top \bar{\Sigma}^{-1} \right)^{-1} \\
b_k^* &= 0
\end{aligned} \tag{18}$$

where η is determined by

$$\sum_{i=1}^{m} \left[(1 - \hat{\lambda}_i^{\frac{1}{2}}/\eta)^{-2} + 2\ln(1 - \hat{\lambda}_i^{\frac{1}{2}}/\eta) \right] - m = 2\delta \tag{19}$$

with $\hat{\lambda}_i$ denoting the ith largest eigenvalue of $\bar{\mathcal{M}}^{\frac{1}{2}} \bar{\Sigma} \bar{\mathcal{M}}^{\frac{1}{2}}$ and $\bar{\mathcal{M}} = \bar{K}^\top \bar{P} C^\top \bar{\Sigma}^{-1} C \bar{P} \bar{K}$. The covariance of z_k^a is given by

$$(\Sigma_k^a)^* = \bar{\mathcal{M}}^{-\frac{1}{2}} \left((\bar{\mathcal{M}}^{\frac{1}{2}} \bar{\Sigma} \bar{\mathcal{M}}^{\frac{1}{2}})^{-\frac{1}{2}} - I/\eta \right)^{-2} \bar{\mathcal{M}}^{-\frac{1}{2}} \tag{20}$$

Theorem 1. *For the system illustrated in Fig. 1, under the stochastic linear attack defined in* (12) *and subject to the stealthiness constraint* (14)*, the posterior estimation error covariance at the remote estimator, conditioned on γ_k, is given by*

$$\begin{aligned}
\tilde{P}_k &= \tilde{P}_k^- + \gamma_k (\bar{K}\Sigma_k^a \bar{K}^\top - \bar{P} C^\top T_k^\top \bar{K}^\top - \bar{K} T_k C \bar{P}) \\
&\quad + (1-\gamma_k)(\bar{K}(C\tilde{P}_k^- C^\top + R)\bar{K}^\top - \bar{K} C \tilde{P}_k^- - \tilde{P}_k^- C^\top \bar{K}^\top)
\end{aligned} \tag{21}$$

Proof: We begin by expanding the posterior estimation error covariance as

$$\begin{aligned}
\tilde{P}_k &= \mathbb{E}[(x_k - \tilde{x}_k)(x_k - \tilde{x}_k)^\top] \\
&= \tilde{P}_k^- + \bar{K}\mathbb{E}[\tilde{z}_k \tilde{z}_k^\top]\bar{K}^\top - \bar{K}\mathbb{E}[\tilde{z}_k (\tilde{e}_k^-)^\top] - \mathbb{E}[\tilde{e}_k^- \tilde{z}_k^\top]\bar{K}^\top.
\end{aligned} \tag{22}$$

Substituting (6) into (22) yields

$$\begin{aligned}
&\bar{K}\mathbb{E}[\tilde{z}_k \tilde{z}_k^\top]\bar{K}^\top \\
&= \gamma_k \bar{K}\Sigma_k^a \bar{K}^\top + (1-\gamma_k)\bar{K}(C\tilde{P}_k^- C^\top + R)\bar{K}^\top
\end{aligned} \tag{23}$$

where the binary property of γ_k has been used: $\gamma_k^2 = \gamma_k$, $(1-\gamma_k)^2 = (1-\gamma_k)$, and $\gamma_k(1-\gamma_k) = 0$.

Next, we compute the cross-covariance term:

$$\begin{aligned}
\mathbb{E}[\tilde{e}_k^- \tilde{z}_k^\top]\bar{K}^\top &= \mathbb{E}[\tilde{e}_k^- (\gamma_k z_k^a + (1-\gamma_k)(y_k - C\tilde{x}_k^-))^\top]\bar{K}^\top \\
&= \gamma_k \mathbb{E}[\tilde{e}_k^- z_k^a]\bar{K}^\top + (1-\gamma_k)\tilde{P}_k^- C^\top \bar{K}^\top.
\end{aligned} \tag{24}$$

To deal with the term $\mathbb{E}[\tilde{e}_k^- z_k^a]$ in (22), we first compute the term $\tilde{e}_k^-$ given by

$$\begin{aligned}
\tilde{e}_k^- &= Ax_{k-1} + \omega_{k-1} - A(\tilde{x}_{k-1}^- + \bar{K}\tilde{z}_{k-1}) \\
&= A^k (x_0 - \hat{x}_0^-) + \sum_{i=0}^{k-1} A^i \omega_{k-1-i} - \sum_{i=0}^{k-1} A^{i+1} \bar{K} \tilde{z}_{k-1-i}
\end{aligned}$$

and the term z_k^a given by

$$\begin{aligned}
z_k^a &= T_k C(x_k - \hat{x}_k^-) + T_k v_k + b_k \\
&= T_k C[A(I - \bar{K}C)]^k (x_0 - \hat{x}_0^-) + \sum_{i=0}^{k-1} T_k C[A(I - \bar{K}C)]^i \\
&\quad \times \omega_{k-1-i} + \xi_k
\end{aligned}$$

where

$$\xi_k = T_k v_k + b_k - \sum_{i=0}^{k-1} T_k C[A(I - \bar{K}C)]^i A K v_{k-1-i}.$$

By applying Lemma 1, it follows that

$$\begin{aligned}
&\mathbb{E}[\tilde{e}_k^- \tilde{z}_k^\top]\bar{K}^\top \\
&= \mathbb{E}[\tilde{e}_k^- (\gamma_k z_k^a + (1-\gamma_k)(y_k - C\tilde{x}_k^-))^\top]\bar{K}^\top \\
&= \gamma_k \mathbb{E}\big[(A^k (x_0 - \hat{x}_0^-) + \sum_{i=0}^{k-1} A^i \omega_{k-1-i} - \sum_{i=0}^{k-1} A^{i+1} \bar{K} \tilde{z}_{k-1-i}) \\
&\quad \times (z_k^a)^\top \big] \bar{K}^\top + (1-\gamma_k)\tilde{P}_k^- C^\top \bar{K}^\top
\end{aligned}$$

$$
\begin{aligned}
&= \gamma_k \mathbb{E}\big[(A^k(x_0 - \hat{x}_0^-) + \sum_{i=0}^{k-1} A^i \omega_{k-1-i})(z_k^a)^\top\big] \bar{K}^\top \\
&\quad + (1-\gamma_k)\tilde{P}_k^- C^\top \bar{K}^\top \\
&= \gamma_k \big[A^k \bar{P} (I - \bar{K}C)^{k^\top} A^{k^\top} + \sum_{i=0}^{k-1} A^i Q (I - \bar{K}C)^{i^\top} A^{i^\top}\big] \\
&\quad \times C^\top T_k^\top \bar{K}^\top + (1-\gamma_k)\tilde{P}_k^- C^\top \bar{K}^\top \\
&= \gamma_k \bar{P} C^\top T_k^\top \bar{K}^\top + (1-\gamma_k)\tilde{P}_k^- C^\top \bar{K}^\top \qquad (25)
\end{aligned}
$$

where the last equality has been obtained by using the fact that $\bar{P}$ is the unique positive semi-definite solution of $X = h \circ \tilde{g}(X)$, i.e.,

$$
\begin{aligned}
\bar{P} &= (h \circ \tilde{g})^k(\bar{P}) \\
&= A^k \bar{P} (A^k (I - KC)^k)^\top + \sum_{i=0}^{k-1} A^i Q (A^i (I - \bar{K}C)^i)^\top.
\end{aligned}
$$

Substituting (23)–(25) into (22), we obtain the desired expression (21), which completes the proof. ∎

Unlike in the standard Kalman filtering framework, the posterior estimation error covariance $\tilde{P}_k$ in (21) is a random matrix due to the stochastic nature of γ_k, which represents the success or failure of the attack at time k. Consequently, optimizing the objective function $\mathrm{Tr}(\tilde{P}_k)$ in problem P1 (15) is generally not a well-defined problem. However, using the expression of $\tilde{P}_k$ (21), derived in Theorem 1, we can reformulate this stochastic problem into an equivalent, deterministic optimization problem P2.

Based on the expression for $\tilde{P}_k$ in (21), the objective function $\mathrm{Tr}(\tilde{P}_k)$ of problem P1 is a linear sum of terms that are constant at time step k and variable terms that depend on the optimization T_k and Σ_k^a. Maximizing $\tilde{P}_k$ is equivalent to maximizing only the variable part, and therefore the original problem P1 (15) is equivalent to the following problem P2:

$$
\begin{aligned}
\textbf{P2:} \quad \max_{T_k, \Sigma_k^a} \quad & \mathrm{Tr}(\bar{K}\Sigma_k^a \bar{K}^\top - \bar{P}C^\top T_k^\top \bar{K}^\top - \bar{K}T_k C\bar{P}) \\
\text{s.t.} \quad & D_{\mathrm{KL}}(z_k^a \| z_k) \le \delta
\end{aligned} \qquad (26)
$$

From the linear attack formulation in (11), since $\Theta_k \succeq 0$, the following inequality is inherently satisfied:

$$
T_k \bar{\Sigma} T_k^\top - \Sigma_k^a \preceq 0
$$

By invoking Lemma 2, the optimal value of problem (26) can be expressed as

$$
M = \mathrm{Tr}(\bar{K}(\Sigma_k^a)^* \bar{K}^\top - \bar{P}C^\top (T_k^*)^\top \bar{K}^\top - \bar{K}T_k^* C\bar{P}). \qquad (27)
$$

Here, the optimal value M remains time-invariant for all $k > 0$ since the structure of the optimization problem does not vary with time.

With the one-step optimal attack strategy established, a fundamental question arises for the remote estimation system: Is this particular optimal strategy essential to destabilize the convergence of the estimation error covariance $\tilde{P}_k$? If so, detection and resilience mechanisms can be specifically tailored against such structured attacks. Conversely, if general classes of persistent stealthy attacks can produce a comparable effect, then the problem becomes significantly more challenging. To address this issue, we proceed to analyze the long-term convergence behavior of $\tilde{P}_k$ under various attack scenarios.

IV. Convergence Analysis

In this section, the convergence behavior of the estimation error covariance is examined under two classes of attack scenarios: (i) a finite number of stochastic stealthy attacks, and (ii) persistent stochastic stealthy attacks. The latter refers to attack sequences that begin at $k = 0$ and persist for all $k \ge 0$. To facilitate the subsequent analysis, the following lemma is introduced.

Lemma 3. *([38]) If a linear operator $\mathcal{L}(\cdot) : \mathbb{S}_+^n \to \mathbb{S}_+^n$ satisfies*

1) *$\mathcal{L}(tX) = t\mathcal{L}(X),\ t > 0$;*
2) *$\forall X,\ Y \in \mathbb{S}_+^n, X \succeq Y$ indicates $\mathcal{L}(X) \succeq \mathcal{L}(Y)$;*
3) *there exists $Y \in \mathbb{S}_+^n,\ s.t.\ Y \succ \mathcal{L}(Y)$.*

then the following conclusions hold:

1) *$\forall W \in \mathbb{S}_+^n, \lim_{k\to\infty} \mathcal{L}^k(W) = 0$;*
2) *$\forall U,\ W_0 \in \mathbb{S}_+^n, W_{k+1} = \mathcal{L}(W_k) + U$, sequence $\{W_k\}$ is bounded.*
3) *If the abovementioned sequence $\{W_k\}$ satisfies $W_0 \preceq W_1$, then $\{W_k\}$ is nondecreasing and convergent.*

For convenience, Lemma 3 consolidates several results originally presented in distributed form throughout [38]. It provides a foundational tool for analyzing the asymptotic behavior of matrix recursions involving monotonicity and boundedness properties, which are central to our subsequent convergence proofs.

A. Convergence after a Finite Number of Attacks

In this subsection, we examine the convergence properties of remote state estimation when the system is subjected to a finite number of stochastic stealthy attacks. We impose no constraints on the attack duration, design, or frequency, except that they must satisfy the stealthiness constraint (14) and terminate at a finite time instant.

Theorem 2. *Consider the system in Fig. 1 under linear deception attacks satisfying the stealthiness constraint* (14)*. If the attacks occur only during a finite interval $[0, k^*]$, then the prior estimation error covariance matrix converges to the steady-state prior covariance matrix $\bar{P}$ of the standard Kalman filter, i.e.,*

$$
\lim_{k\to\infty} \tilde{P}_k^- = \bar{P}. \qquad (28)
$$

Proof: Assume that the attacks are active for $0 \le k \le k^*$, and stop thereafter. For $k > k^*$, we set $\gamma_k \equiv 0$ ($k > k^*$) in (21), and the posterior estimation error covariance then simplifies to

$$
\begin{aligned}
\tilde{P}_k &= \tilde{P}_k^- + \bar{K}(C\tilde{P}_k^- C^\top + R)\bar{K}^\top - \bar{K}C\tilde{P}_k^- - \tilde{P}_k^- C^\top \bar{K}^\top \\
&= (I - \bar{K}C)\tilde{P}_k^- (I - \bar{K}C)^\top + \bar{K}R\bar{K}^\top.
\end{aligned}
$$

With $\tilde{P}_{k+1}^- = A\tilde{P}_k A^\top + Q$, the prior covariance iteration for $k \ge k^*$ becomes:

$$
\tilde{P}_{k+1}^- = A(I - \bar{K}C)\tilde{P}_k^- (I - \bar{K}C)^\top A^\top + A\bar{K}R\bar{K}^\top A^\top + Q.
$$

Define the operators $g_0(\cdot) : \mathbb{S}_+^n \to \mathbb{S}_+^n$ and $\mathcal{L}_0(\cdot) : \mathbb{S}_+^n \to \mathbb{S}_+^n$:

$$
g_0(X) = A(I - \bar{K}C)X(I - \bar{K}C)^\top A^\top + Q^+
$$

$$\mathcal{L}_0(X) = A(I-\bar{K}C)X(I-\bar{K}C)^\top A^\top$$

where $Q^+ = A\bar{K}R\bar{K}^\top A^\top + Q \succeq Q \succ 0$.

Clearly, $\tilde{P}_{k+1}^- = g_0(\tilde{P}_k^-) = \mathcal{L}_0(\tilde{P}_k^-) + Q^+$. Since $\bar{K}$ is the steady-state Kalman gain, $\bar{P}$ is a fixed point of $g_0(\cdot)$ that satisfies

$$\bar{P} = g_0(\bar{P}) = A(I-\bar{K}C)\bar{P}(I-\bar{K}C)^\top A^\top + Q^+ \succ 0$$
$$\Rightarrow \ \bar{P} \succ \mathcal{L}_0(\bar{P}) = A(I-\bar{K}C)\bar{P}(I-\bar{K}C)^\top A^\top,$$

so there exists $\bar{P} \succ 0$ such that $\bar{P} \succ \mathcal{L}_0(\bar{P})$. It is straightforward to verify

1) $\mathcal{L}_0(tX) = t\mathcal{L}_0(X),\ t > 0$;
2) $\forall X,\ Y \in \mathbb{S}_+^n, X \succeq Y$ indicates $\mathcal{L}_0(X) \succeq \mathcal{L}_0(Y)$.

Therefore, $\mathcal{L}_0$ satisfies the conditions of Lemma 3.

Let $V_0 = 0_{n\times n}$ and $V_k \triangleq g_0(V_{k-1}) = \mathcal{L}_0(V_{k-1}) + Q^+$. Obviously,$V_0 \preceq V_1$. By Lemma 3, $\{V_k\}$ is non-decreasing and

$$\lim_{k\to\infty} V_k = \bar{V}. \tag{29}$$

For any positive semi-definite matrix $U_0 \succeq \bar{V}$, defining the sequence $\{U_k\}_{k\geq 0}$ as

$$U_k \triangleq g_\lambda(U_{k-1}), \tag{30}$$

we have

$$\begin{aligned} U_k &= U_k - \bar{V} + \bar{V} = g_0(U_{k-1}) - g_0(\bar{V}) + \bar{V} \\ &= \mathcal{L}_0(U_{k-1} - \bar{V}) + \bar{V} = \mathcal{L}_0(g_0(U_{k-2}) - g_0(\bar{V})) + \bar{V} \\ &= \mathcal{L}_0(\mathcal{L}_0(U_{k-2} - \bar{V})) + \bar{V} = \cdots = \mathcal{L}_0^k(U_0 - \bar{V}) + \bar{V} \end{aligned}$$

As k tends to infinity, we have

$$\lim_{k\to\infty} U_k = \bar{V} \tag{31}$$

Let $V_0 = 0$ and $U_0 = \bar{V} + \tilde{P}_{k^*}^-$. Then, $\tilde{P}_k^-\ (k \geq k^*)$ can be bounded by V_{k-k^*} and U_{k-k^*} as follows:

$$V_{k-k^*} \preceq \tilde{P}_k^- \preceq U_{k-k^*},\ \forall k \geq k^*,$$

which implies

$$\lim_{k\to\infty} \tilde{P}_k^- = \bar{V} \tag{32}$$

Similarly, by setting $V_0 = 0$ and $U_0 = \bar{V} + \bar{X}$, where $\bar{X}$ is an arbitrary positive semi-definite fixed point of $g_0(\cdot)$ such as $\bar{P}$, it can be demonstrated that

$$\bar{X} = \bar{V}, \tag{33}$$

which shows the uniqueness of the fixed point of $g_0(\cdot)$. Since $\bar{P} = g_0(\bar{P})$, we have

$$\lim_{k\to\infty} \tilde{P}_k^- = \bar{V} = \bar{P}. \tag{34}$$

■

B. Convergence Under Persistent stochastic Stealthy Attacks

This section investigates the convergence behavior of the expected prior error covariance, $\mathbb{E}[\tilde{P}_k^-]$, under persistent stochastic stealthy attacks. Building on the formulation of the one-step optimal attack strategy presented in Section III, we now turn our attention to the long-term implications of such attacks on remote state estimation. Specifically, we aim to address the following critical questions: Which factor plays a more decisive role in disrupting convergence? Is it the optimal design of the attacks or their success rate λ? As will be rigorously demonstrated in Section IV, it is the latter (i.e., the attack success rate) that fundamentally determines the convergence or divergence behavior of the system.

To facilitate this analysis, we introduce the matrix Δ_k to characterize the instantaneous degradation in estimation performance due to a successful attack at time step k, i.e., when $\gamma_k = 1$. Specifically, Δ_k is defined as the incremental increase in the estimation error covariance resulting from the attack:

$$\Delta_k \triangleq \bar{K}\Sigma_k^a\bar{K}^\top - \bar{P}C^\top T_k^\top \bar{K}^\top - \bar{K}T_kC\bar{P}. \tag{35}$$

This expression captures the net effect of the attack on the posterior error covariance, which serves as a key quantity in the subsequent convergence analysis.

We consider the attackers to be *malicious*, meaning that their injected signals are intentionally designed to degrade the estimation performance of the remote estimator. As a result, each successful attack inevitably leads to a non-decreasing posterior error covariance, i.e., $\Delta_k \succeq 0$. Furthermore, the one-step optimal attacks derived in Theorem 1 and Lemma 2 provide a theoretical upper bound on the impact that any such stealthy attack can induce, and this motivates the following definition. We name it "Malicious Attack" to emphasize the adversarial intent (via $\Delta_k \succeq 0$), since its "bounded" characteristic is a natural consequence of the preceding modeling.

Definition 1 (Malicious attack)**.** *A linear stealthy attack design is said to be a malicious attack if the resulting one-step degradation in the posterior error covariance satisfies*

$$0_{n\times n} \preceq \Delta_k \preceq MI_{n\times n} \tag{36}$$

where M is a scalar constant corresponding to the maximal impact, as determined by the one-step optimization problem (27)*.*

Given that any malicious attack sequence Δ_k is bounded in the sense of Definition 1, it follows that there exist symmetric matrices $\underline{\Delta}$, $\overline{\Delta}$ such that

$$\underline{\Delta} \preceq \Delta_k \preceq \overline{\Delta}.$$

This boundedness allows us to frame the convergence analysis of $\mathbb{E}[\tilde{P}_k^-]$ under general attack sequences by comparing it to the behavior of systems subjected to constant attack profiles. Therefore, to facilitate rigorous analysis, we consider a constant malicious attack model where the impact of each successful attack remains invariant over time:

$$\Delta_k = \Delta,\ k > 0. \tag{37}$$

To analyze the convergence behavior of the estimation error covariance under persistent malicious attacks, we first introduce two matrix operators that capture the stochastic dynamics of the system under attack. Specifically, let $\mathcal{L}_\lambda(X)$ and $g_\lambda(X)$ be operators mapping $\mathbb{S}_+^n \to \mathbb{S}_+^n$:

$$\begin{aligned}\mathcal{L}_\lambda(X) &= \lambda AXA^\top + (1-\lambda)A(I-\bar{K}C)X(I-\bar{K}C)^\top A^\top \\ g_\lambda(X) &= \mathcal{L}_\lambda(X) + Q + (1-\lambda)A\bar{K}R\bar{K}^\top A^\top + \lambda A\Delta A^\top.\end{aligned} \tag{38}$$

For notational simplicity, we introduce the matrix

$$Q_\lambda \triangleq Q + (1-\lambda)A\bar{K}R\bar{K}^\top A^\top + \lambda A\Delta A^\top.$$

With definitions of $\mathcal{L}_\lambda(X)$ and $g_\lambda(X)$, we now state Theorem 3.

Theorem 3. *For the system in Fig. 1 under persistent stochastic stealthy attacks satisfying* (36) *and* (37)*, the following conditions are equivalent:*

1) $\mathbb{E}[\tilde{P}_k^-]$ converges to $P_\lambda \in \mathbb{S}_+^n$;
2) there exists $\bar{X} \in \mathbb{S}_+^n,\ s.t.\ \bar{X} = g_\lambda(\bar{X})$;
3) there exists $Y \in \mathbb{S}_+^n,\ s.t.\ Y \succeq \mathcal{L}_\lambda(Y) + Q$.

Proof: To begin with, we show that condition 2) can be derived from condition 1). Under the linear deception attack that satisfies (36), the estimation error covariance can be written as

$$\begin{aligned}\tilde{P}_k = \tilde{P}_k^- + \gamma_k\Delta + (1-\gamma_k)[&\bar{K}(C\tilde{P}_k^-C^\top + R)\bar{K}^\top \\ &-\bar{K}C\tilde{P}_k^- - \tilde{P}_k^-C^\top\bar{K}^\top].\end{aligned}$$

The iterative formula from $\tilde{P}_k^-$ to $\tilde{P}_{k+1}^-$ is given by

$$\begin{aligned}\tilde{P}_{k+1}^- = \gamma_k A\tilde{P}_k^-A^\top + (1-\gamma_k)A(I-\bar{K}C)\tilde{P}_k^-(I-\bar{K}C)^\top A^\top \\ +Q + \gamma_k A\Delta A^\top + (1-\gamma_k)A\bar{K}R\bar{K}^\top A^\top.\end{aligned} \tag{39}$$

Clearly, $\mathbb{E}[\tilde{P}_{k+1}^-]$ can be expressed in terms of $\mathbb{E}[\tilde{P}_0^-]$ and $g_\lambda(\cdot)$ as follows

$$\begin{aligned}&\mathbb{E}[\tilde{P}_{k+1}^-] = \mathbb{E}_{\gamma_k}\left[\mathbb{E}[\tilde{P}_{k+1}^-|\gamma_k]\right] \\ &= \mathbb{E}_{\gamma_k}\big[\gamma_k A\mathbb{E}[\tilde{P}_k^-]A^\top + (1-\gamma_k)A(I-\bar{K}C)\mathbb{E}[\tilde{P}_k^-](I-\bar{K}C)^\top \\ &\quad \times A^\top + Q + \gamma_k A\Delta A^\top + (1-\gamma_k)A\bar{K}R\bar{K}^\top A^\top\big] \\ &= \lambda A\mathbb{E}[\tilde{P}_k^-]A^\top + (1-\lambda)A(I-\bar{K}C)\mathbb{E}[\tilde{P}_k^-](I-\bar{K}C)^\top A^\top \\ &\quad +Q + \lambda A\Delta A^\top + (1-\lambda)A\bar{K}R\bar{K}^\top A^\top \\ &= g_\lambda(\mathbb{E}[\tilde{P}_k^-]) = g_\lambda(g_\lambda(\mathbb{E}[\tilde{P}_{k-1}^-])) = \cdots = g_\lambda^{k+1}(\tilde{P}_0^-) \\ &= g_\lambda^{k+1}(\bar{P}).\end{aligned}$$

The above equation indicates that analyzing the convergence of $\mathbb{E}[\tilde{P}_k^-]$ is equivalent to examining the convergence of the sequence $\{g_\lambda^k(\bar{P})\}_{k\geq 0}$.

According to condition 1), i.e., $\mathbb{E}[\tilde{P}_k^-]$ converges to P_λ, we have

$$\lim_{k\to\infty} g_\lambda^k(\bar{P}) = \lim_{k\to\infty}\mathbb{E}[\tilde{P}_k^-] = P_\lambda.$$

Obviously, P_λ is a fixed point of $g_\lambda(\cdot)$ as follows

$$\begin{aligned}g_\lambda(P_\lambda) &= g_\lambda(\lim_{k\to\infty} g_\lambda^k(\bar{P})) = \lim_{k\to\infty} g_\lambda(g_\lambda^k(\bar{P})) \\ &= \lim_{k\to\infty} g_\lambda^k(\bar{P}) = P_\lambda\end{aligned}$$

Therefore, the fixed point $\bar{X} \in \mathbb{S}_+^n$ in condition 2) exists.

In what follows, we show that condition 3) can be derived from condition 2). We know that $\bar{X}$, the solution of $X = g_\lambda(X)$, also satisfies the inequality in condition 3):

$$\begin{aligned}\bar{X} = g_\lambda(\bar{X}) &= \mathcal{L}_\lambda(\bar{X}) + Q_\lambda \\ &= \mathcal{L}_\lambda(\bar{X}) + Q + (1-\lambda)A\bar{K}R\bar{K}^\top A^\top + \lambda A\Delta A^\top \\ &\succeq \mathcal{L}_\lambda(\bar{X}) + Q.\end{aligned}$$

Therefore, condition 3) holds.

At last, we show that condition 1) can be derived from condition 3). Since $Q \succ 0$, from condition 3), we have

$$Y \succeq Q \succ 0 \text{ and } Y \succ \mathcal{L}_\lambda(Y) \succ 0.$$

It is straightforward to verify that $\mathcal{L}_\lambda(\cdot)$ satisfies the first two conditions of Lemma 3. Therefore, applying Lemma 3 yields

1) $\forall W \in \mathbb{S}_+^n,\ \lim_{k\to\infty}\mathcal{L}_\lambda^k(W) = 0$;
2) $g_\lambda^k(W_0) \preceq M_{W_0}$ since Q_λ is a constant positive semi-definite matrix when the attack design (T_k, Θ_k) and λ are fixed.

Letting $V_0 = 0$ and defining $V_k \triangleq g_\lambda(V_{k-1})$, the sequence $\{V_k\}_{k\geq 0}$ is non-decreasing. By Lemma 3, it follows that

$$\lim_{k\to\infty} V_k = \bar{V}.$$

For any positive semi-definite matrix $U_0 \succeq \bar{V}$, define the sequence $\{U_k\}_{k\geq 0}$ as

$$U_k \triangleq g_\lambda(U_{k-1}). \tag{40}$$

It can be shown that

$$\begin{aligned}U_k &= U_k - \bar{V} + \bar{V} = g_\lambda(U_{k-1}) - g_\lambda(\bar{V}) + \bar{V} \\ &= \mathcal{L}_\lambda(U_{k-1} - \bar{V}) + \bar{V} = \mathcal{L}_\lambda(g_\lambda(U_{k-2}) - g_\lambda(\bar{V})) + \bar{V} \\ &= \mathcal{L}_\lambda(\mathcal{L}_\lambda(U_{k-2} - \bar{V})) + \bar{V} = \cdots = \mathcal{L}_\lambda^k(U_0 - \bar{V}) + \bar{V}\end{aligned}$$

As $k \to \infty$, it follows that

$$\lim_{k\to\infty} U_k = \bar{V}. \tag{41}$$

By choosing $U_0 = \bar{V} + \mathbb{E}[\tilde{P}_0^-]$, the expected error covariance $\mathbb{E}[\tilde{P}_k^-]$ is bounded as

$$V_k \preceq \mathbb{E}[\tilde{P}_k^-] \preceq U_k,\ k \geq 0.$$

Therefore, one has

$$\lim_{k\to\infty}\mathbb{E}[\tilde{P}_k^-] = \bar{V} \tag{42}$$

which indicates that the convergence point P_λ in condition 1) coincides with $\bar{V}$ in (42), and is thus unique. ■

Theorem 3 establishes two equivalent formulations for the convergence of $\mathbb{E}[\tilde{P}_k^-]$: one based on a fixed-point condition and the other on a linear matrix inequality. The fixed-point formulation provides a sufficient basis for analyzing both convergence and divergence. In contrast, the inequality formulation reveals that convergence is determined solely by the success rate λ, independent of the specific design of linear attack strategies.

Corollary 1 (Necessary condition for convergence)**.** *For the system in Fig. 1 under persistent stochastic stealthy attacks*

satisfying (36) *and* (37)*, if the expectation of the prior error covariance matrix* $\mathbb{E}[\tilde{P}_k^-]$ *converges, then*

$$\lambda \leq \frac{1}{[\rho(A)]^2}. \tag{43}$$

where $\rho(A)$ *denotes the spectral radius of matrix* A*.*

Proof: Define the Lyapunov-type operator

$$h_\lambda(X) = \lambda AXA^\top + Q \preceq g_\lambda(X),\ \forall X \in \mathbb{S}_+^n.$$

Assuming $h_\lambda^k(X) \preceq g_\lambda^k(X)$, one has

$$\begin{aligned} h_\lambda^{k+1}(X) &= h_\lambda(h_\lambda^k(X)) \preceq h_\lambda(g_\lambda^k(X)) \\ &\preceq g_\lambda(g_\lambda^k(X)) = g_\lambda^{k+1}(X). \end{aligned}$$

Therefore, by induction, we have $h_\lambda^k(X) \preceq g_\lambda^k(X), \forall X \in \mathbb{S}_+^n,\ k = 1, 2, \cdots$.

According to Theorem 3, the convergence of $\mathbb{E}[\tilde{P}_k^-]$ is equivalent to the existence of a fixed point $\bar{X} \in \mathbb{S}_+^n$ of $g_\lambda(\cdot)$, which enables us to establish a uniform upper bound for $h_\lambda^k(\bar{X})$

$$h_\lambda^k(\bar{X}) \preceq g_\lambda^k(\bar{X}) = \bar{X},$$

while the closed form of $h_\lambda^k(\bar{X})$ is

$$h_\lambda^k(\bar{X}) = A_\lambda^k \bar{X} (A_\lambda^\top)^k + \sum_{i=0}^{k-1} A_\lambda^i Q (A_\lambda^\top)^i,$$

where $A_\lambda = \sqrt{\lambda} A$. Perform the Jordan decomposition $A_\lambda = S^{-1} J S$. Since $Q \succ 0$, there always exists a sufficiently small $\varepsilon > 0$ such that $Q \succeq \varepsilon I$. Therefore, one has

$$\begin{aligned} \bar{X} \succeq h_\lambda^k(\bar{X}) &\succeq \sum_{i=0}^{k-1} A_\lambda^i Q (A_\lambda^\top)^i \succeq \varepsilon \sum_{i=0}^{k-1} A_\lambda^i (A_\lambda^\top)^i \\ &= \varepsilon \sum_{i=0}^{k-1} (S^{-1} J S)^i [(S^{-1} J S)^\top]^i \\ &= \varepsilon S^{-1} [\sum_{i=0}^{k-1} J^i (J^\top)^i] S. \end{aligned} \tag{44}$$

Note that $\sum_{i=0}^{k-1} J^i (J^\top)^i \preceq \epsilon^{-1} S \bar{X} S^{-1}$ implies that the spectral radius $\rho(J) \leq 1$, otherwise $J^i (J^\top)^i$ would diverge as $i \to \infty$. This leads to a result equivalent to (43):

$$\rho(A) = \rho(\frac{1}{\sqrt{\lambda}}\ A_\lambda) = \frac{1}{\sqrt{\lambda}} \rho(A_\lambda) = \frac{1}{\sqrt{\lambda}} \rho(J) \leq \frac{1}{\sqrt{\lambda}}, \tag{45}$$

and the proof is complete. ■

Corollary 2 (Sufficient condition for convergence)**.** *For the system in Fig. 1 under persistent stochastic stealthy attacks satisfying* (36) *and* (37)*, if the following inequality is satisfied:*

$$\lambda < \frac{1 - \|F\|\|F^\top\|}{\|A\|\|A^\top\| - \|F\|\|F^\top\|}, \tag{46}$$

then the expectation of the prior error covariance matrix $\mathbb{E}[\tilde{P}_k^-]$ *converges, where* $F = A(I - \bar{K}C)$*.*

Proof: It can be observed that

$$\begin{aligned} \|g_\lambda(X)\| &= \|\lambda AXA^\top + (1-\lambda) FXF^\top + Q_\lambda\| \\ &\leq \lambda \|AXA^\top\| + (1-\lambda)\|FXF^\top\| + \|Q_\lambda\| \\ &\leq (\lambda \|A\|\|A^\top\| + (1-\lambda)\|F\|\|F^\top\|)\ \|X\| + \|Q_\lambda\|. \end{aligned}$$

Let $V_0 = 0_{n\times n}$ and $V_k \triangleq g_\lambda(V_{k-1})$. Denote $t \triangleq \lambda\|A\|\|A^\top\| + (1-\lambda)\|F\|\|F^\top\|$. Then, the following recursive inequality holds:

$$\begin{aligned} \|V_k\| = \|g_\lambda(V_{k-1})\| &\leq t\|V_{k-1}\| + \|Q_\lambda\| \\ &\leq \cdots \leq t^k \|V_0\| + \sum_{i=0}^{k-1} t^i \|Q_\lambda\|. \end{aligned}$$

By applying (46), it is obtained that

$$\begin{aligned} t &= (\lambda\|A\|\|A^\top\| + (1-\lambda)\|F\|\|F^\top\|) \\ &< \frac{1 - \|F\|\|F^\top\|}{\|A\|\|A^\top\| - \|F\|\|F^\top\|} \|A\|\|A^\top\| \\ &\quad + (1 - \frac{1 - \|F\|\|F^\top\|}{\|A\|\|A^\top\| - \|F\|\|F^\top\|}) \|F\|\|F^\top\| \\ &= 1. \end{aligned}$$

Therefore, the sequence $\{V_k\}$ is bounded. Since $\{V_k\}$ is also non-decreasing, it must converge, and $\lim_{k\to\infty} V_k$ is a fixed point of $g_\lambda(\cdot)$, which ensures that $\mathbb{E}[\tilde{P}_k^-]$ converges. ■

In some cases, the sufficient condition of Corollary 2 cannot be satisfied, as $\frac{1-\|F\|\|F^\top\|}{\|A\|\|A^\top\|-\|F\|\|F^\top\|}$ is negative under various norms (e.g., 2-norm, L_p-norm). The subsequent Corollary 3 presents an equivalent condition for convergence, which overcomes this limitation via more complex computations.

Corollary 3 (Equivalent condition for convergence)**.** *For the system in Fig. 1 under persistent stochastic stealthy attacks satisfying* (36) *and* (37)*, the expectation of the prior error covariance matrix* $\mathbb{E}[\tilde{P}_k^-]$ *converges if and only if there exists* $0 \prec Z \preceq I_n$ *such that*

$$\begin{pmatrix} Z & \sqrt{1-\lambda} ZA(I - \bar{K}C) & \sqrt{\lambda} ZA \\ \sqrt{1-\lambda}(I - \bar{K}C)^\top A^\top Z & Z & 0 \\ \sqrt{\lambda} A^\top Z & 0 & Z \end{pmatrix} \succ 0.$$

Proof: By adopting an approach analogous to that in [38, Th. 5], the Schur complement is utilized to establish an equivalent condition for convergence. According to Theorem 3, the convergence of $\mathbb{E}[\tilde{P}_k^-]$ is ensured if and only if there exists a solution $Y \in \mathbb{S}_+^n$ satisfying the inequality $Y \succeq \mathcal{L}_\lambda(Y) + Q$. Here, $\mathcal{L}_\lambda(Y) = \lambda AYA^\top + (1-\lambda)A(I-\bar{K}C)Y(I-\bar{K}C)^\top A^\top$. Owing to the homogeneity of Y and $\mathcal{L}_\lambda(Y)$ and the condition $Q \succ 0$, the feasibility of the inequality $Y \succeq \mathcal{L}_\lambda(Y) + Q$ is equivalent to that of

$$Y \succ \lambda AYA^\top + (1-\lambda)A(I - \bar{K}C)Y(I - \bar{K}C)^\top A^\top$$

First, by applying the Schur complement to the above inequality, we obtain the following equivalent matrix inequality:

$$\begin{pmatrix} Y - \lambda AYA^\top & \sqrt{1-\lambda} A(I - \bar{K}C) \\ \sqrt{1-\lambda}(I - \bar{K}C)^\top A^\top & Y^{-1} \end{pmatrix} \succ 0.$$

Next, applying the Schur complement decomposition again to the top-left block of the preceding matrix inequality above, we arrive at the final LMI condition (47):

$$\begin{pmatrix} Y & \sqrt{1-\lambda} A(I - \bar{K}C) & \sqrt{\lambda} A \\ \sqrt{1-\lambda}(I - \bar{K}C)^\top A^\top & Y^{-1} & 0 \\ \sqrt{\lambda} A^\top & 0 & Y^{-1} \end{pmatrix} \succ 0. \tag{47}$$

Subsequently, by left- and right-multiplying both sides of (47) with $\mathrm{diag}(Y^{-1},\ I_n,\ I_n)$ and setting $Z = Y^{-1}$, the following equivalent inequality is obtained:

$$\begin{pmatrix} Z & \sqrt{1-\lambda}ZA(I-\bar{K}C) & \sqrt{\lambda}ZA \\ \sqrt{1-\lambda}(I-\bar{K}C)^\top A^\top Z & Z & 0 \\ \sqrt{\lambda}A^\top Z & 0 & Z \end{pmatrix} \succ 0. \tag{48}$$

Since $Y \succeq 0$, it follows that $Z \succeq 0$. Given that inequality (48) is linear in Z, the constraint $0 \preceq Z \preceq I_n$ can be imposed without loss of generality. ■

Corollaries 1 and 2 provide computationally efficient necessary and sufficient conditions, respectively. They offer efficient assessments tied directly to system parameters, such as the norms and spectral radius of A and F, allowing for the rapid verification of straightforward cases. However, these conditions are inconclusive if λ falls into the "undetermined region" between their two thresholds. Corollary 3 resolves this ambiguity by providing a definitive equivalent condition. Therefore, in practice, Corollaries 1 and 2 serve as efficient initial criteria. Corollary 3 is then reserved for the more complex cases that fall within the gap, thus providing a complete and synergistic framework for the convergence analysis.

Corollary 4. *When a remote estimation system is subjected to malicious attacks, the convergence or divergence of the expectation of its estimation covariance $\mathbb{E}[\tilde{P}_k^-]$ is independent of the specific attack design, namely, T_k, Θ_k, and Σ_k^a.*

Proof: Theorem 3 provides two conditions equivalent to the convergence in expectation of the system's error covariance. One of these conditions is: there exists $Y \in \mathbb{S}_+^n$, s.t.

$$Y \succeq \lambda AYA^\top + (1-\lambda)A(I-\bar{K}C)Y(I-\bar{K}C)^\top A^\top + Q$$

However, any variables associated with the attacker's specific design $T_k, \Theta_k, \Sigma_k^a$ and its corresponding impact Δ are absent from this condition. This implies that as long as the attack satisfies the premises of Theorem 3, namely, persistent stochastic stealthy attacks satisfying (36) and (37), the convergence or divergence outcome is independent of the specific attack design $T_k, \Theta_k, \Sigma_k^a$. Consequently, Corollary 4 holds. ■

Corollary 4 reveals a critical security principle for remote estimation systems: given a fixed attack success rate λ, stealthy malicious attacks can be constructed to induce divergence in the estimation process with effectiveness comparable to that of optimally designed attacks. Crucially, this construction only requires the vector $C\tilde{x}_k^-$ and the covariance $\bar{\Sigma}$, rather than the full system knowledge and sensor data y_k as outlined in Assumptions 1 and 2. Specifically, a stealthy white-noise sequence attack can be straightforwardly constructed using these terms as follows

$$y_k^a = C\tilde{x}_k^- + b_k \tag{49}$$

with the parameters set to be $T_k = 0$, $\Theta_k = \bar{\Sigma}$, and $\Delta = \bar{K}\bar{\Sigma}\bar{K}^\top$.

This result highlights the necessity for both enhanced protection of system parameters and reinforced integrity of communication channels against tampering, offering specific guidance for countermeasures:

1) From a long-term perspective, investing resources to protect the integrity of communication channels, and thereby reduce the attack success rate can ensure the convergence of the expectation of the error covariance, even in the presence of optimal stealthy attacks. Conversely, if a sufficiently low attack success rate cannot be guaranteed, divergence cannot be prevented, even if the attacker possesses limited information and strict detection mechanisms are in place.

2) An attacker does not necessarily require complete system information to launch a stealthy attack. Consequently, the protection of specific system matrices (e.g., $C\tilde{x}_k^-$ and $\bar{\Sigma}$ in the Kalman filter) must be prioritized.

Corollary 5. *Under different attack designs (resulting in Δ and Δ^*), if $\mathbb{E}[\tilde{P}_k^-]$ converges, the difference between the corresponding convergence points P_λ and P_λ^* is given by*

$$P_\lambda - P_\lambda^* = \lambda \sum_{i=0}^{\infty} \mathcal{L}_\lambda^i(A(\Delta - \Delta^*)A^\top) \tag{50}$$

where the operator $\mathcal{L}_\lambda(\cdot)$ is defined in Theorem 3.

Proof: The discrepancy between Δ and Δ^* introduces differences in Q_λ and Q_λ^*, expressed as

$$\begin{aligned} Q_\lambda &= Q + \lambda A\Delta A^\top + (1-\lambda)A\bar{K}R\bar{K}^\top A^\top \\ Q_\lambda^* &= Q + \lambda A\Delta^* A^\top + (1-\lambda)A\bar{K}R\bar{K}^\top A^\top \end{aligned}$$

which further result in different operators g_λ and g_λ^*:

$$\begin{aligned} g_\lambda(X) &= \mathcal{L}_\lambda(X) + Q_\lambda \\ g_\lambda^*(X) &= \mathcal{L}_\lambda(X) + Q_\lambda^*. \end{aligned}$$

According to the proof of Theorem 3, we have

$$\begin{aligned} P_\lambda = \bar{V} &= \lim_{i\to\infty} g_\lambda^i(0_{n\times n}) = \sum_{i=0}^{\infty} \mathcal{L}_\lambda^i(Q_\lambda) \\ P_\lambda^* = \bar{V}^* &= \lim_{i\to\infty} (g_\lambda^*)^i(0_{n\times n}) = \sum_{i=0}^{\infty} \mathcal{L}_\lambda^i(Q_\lambda^*). \end{aligned}$$

Since $\mathcal{L}(\cdot)$ is a linear operator, its iterated forms $\mathcal{L}^i(\cdot)$ are also linear. Hence, the difference can be written as

$$\begin{aligned} P_\lambda - P_\lambda^* &= \sum_{i=0}^{\infty} \mathcal{L}_\lambda^i(Q_\lambda) - \mathcal{L}_\lambda^i(Q_\lambda^*) \\ &= \sum_{i=0}^{\infty} \mathcal{L}_\lambda^i(Q_\lambda - Q_\lambda^*) \\ &= \sum_{i=0}^{\infty} \mathcal{L}_\lambda^i(\lambda A\Delta A^\top - \lambda A\Delta^* A^\top) \\ &= \lambda \sum_{i=0}^{\infty} \mathcal{L}_\lambda^i(A(\Delta - \Delta^*)A^\top) \end{aligned} \tag{51}$$

The proof is now complete. ■

Similarly, if $\mathbb{E}[\tilde{P}_k^-]$ diverges, the difference between $\mathbb{E}[\tilde{P}_k^-]$ and $\mathbb{E}[(\tilde{P}_k^-)^*]$ can be expressed as:

$$\mathbb{E}[\tilde{P}_k^-] - \mathbb{E}[(\tilde{P}_k^-)^*] = \lambda \sum_{i=0}^{k-1} \mathcal{L}_\lambda^i(A(\Delta - \Delta^*)A^\top). \tag{52}$$

Although the general condition for convergence/divergence depends only on λ according to Corollary 4, the one-step optimal attack leads to a significant degradation in long-term estimation performance. Equations (50) and (52) quantify this impact, showing that one-step optimal attacks: 1) increase the convergence point of $\mathbb{E}[\tilde{P}_k^-]$ to a larger value in the convergent case, and 2) accelerate the divergence rate of $\mathbb{E}[\tilde{P}_k^-]$ in the divergent case.

Remark 1. As compared to the existing literature on secure remote estimation under stealthy attacks, this paper introduces several distinctive novelties.

1) A stochastic attack model is proposed, where each attack succeeds with a fixed probability λ, enabling a more realistic characterization of random attack outcomes not addressed in prior deterministic frameworks.
2) A deterministic transformation framework is developed to decouple the stochastic attack process from the estimation error evolution. By establishing the orthogonality between the attack signal and the potentially compromised historical innovations, the original ill-posed stochastic optimization problem is converted into a tractable equivalent deterministic formulation.
3) A variational formulation of one-step optimal stealthy attacks is developed, which maximizes the trace of the posterior error covariance under Kullback–Leibler divergence constraints, offering a tractable upper bound for attack impact.
4) Novel Lyapunov-type operators, $\mathcal{L}_\lambda(\cdot)$ and $g_\lambda(\cdot)$, are introduced to explicitly characterize the covariance evolution dynamics under stochastic stealthy attacks. By incorporating a covariance degradation increment, these operators facilitate a rigorous stability analysis that is invariant to specific attack strategies.
5) The convergence behavior of the expected estimation error covariance is rigorously analyzed under persistent attacks, leading to both necessary and sufficient conditions for convergence. These conditions are expressed through fixed-point characterizations, Lyapunov-type inequalities, and an equivalent matrix inequality condition, all of which are independent of specific attack parameters.
6) This paper establishes that the convergence or divergence of the estimation error is determined solely by the attack success rate λ, revealing a fundamental system vulnerability and guiding more robust design of secure estimation schemes.

V. Numerical Experiments

In this section, we first verify Corollaries 1-2 and Theorem 3, and then examine the filtering performance under different attack designs and values of λ.

A. Verification of Corollaries 1 and 2

Corollaries 1 and 2 establish necessary and sufficient conditions for the convergence of $\mathbb{E}[\tilde{P}_k^-]$, respectively. To validate these theoretical results, we conduct a numerical analysis. Firstly, we randomly generate $10,000$ distinct linear time-invariant systems. We set the dimensions as $m = 3$ and $n = 3$. The eigenvalues of the random matrix A are uniformly distributed over the interval $[1, 10]$. The other system matrices (e.g., C, Q, R, etc.) are random matrices that satisfy the assumptions herein. For each system, we establish necessary and sufficient conditions by using the matrix 2-norm in (46) and compute the critical threshold λ_c for the transition in the expected error covariance between convergent and divergent behavior. When $\lambda < \lambda_c$, the expected error covariance converges; when $\lambda > \lambda_c$, it diverges. The existence of this critical value λ_c is guaranteed for single-input single-output systems. For multiple-input multiple-output systems, extensive random experiments consistently indicate the existence of such a λ_c.

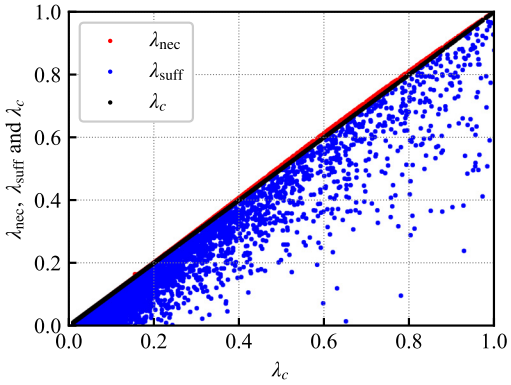


Fig. 2. Scatter plot of the sufficient condition $\lambda_{\rm suff}$, necessary condition $\lambda_{\rm nec}$, and the convergence-divergence transition λ_c.

Fig. 2 presents a scatter plot of $10,000$ data points for $\lambda_{\rm nec}$ (red), $\lambda_{\rm suff}$ (blue), and λ_c (black). The black points form a boundary that clearly separates the red and blue point clouds, which illustrates the relationship $\lambda_{\rm nec} \geq \lambda_c \geq \lambda_{\rm suff}$. For any given randomly generated system, let the attack success rate be λ. If $\lambda > \lambda_{\rm nec}$, it follows that $\lambda > \lambda_c$, causing the system to diverge. This confirms that Corollary 1 provides a necessary condition for convergence. Similarly, the inequality $\lambda \geq \lambda_{\rm suff}$ shows that if $\lambda < \lambda_{\rm suff}$, convergence is guaranteed. This validates that Corollary 2 provides a sufficient condition for convergence. Notably, a few blue points lie at a substantial distance from the boundary, constituting a minimal number of outliers across the $10,000$-point sample.

B. Verification of Theorem 3

Theorem 3 presents three equivalent conditions: 1) convergence of expectation of covariance, 2) existence of a fixed point, and 3) solvability of a linear matrix inequality. It is evident that condition 2) follows from condition 1), and condition 3) follows from condition 2), while the equivalence between 1) and 3) requires numerical verification. We validate Theorem 3 by comparing the differences in λ_c obtained from conditions 1) and 3).

In this numerical experiment, we set the system dimensions to $n = 3$ and $m = 3$. We stochastically generate $1,000$ matrices A with spectral radius distributed across intervals $[1, 2]$, $[2, 3]$, ..., $[9, 10]$. The matrices Q and R are parametrized as $0.1I_3$, $0.1I_3$.

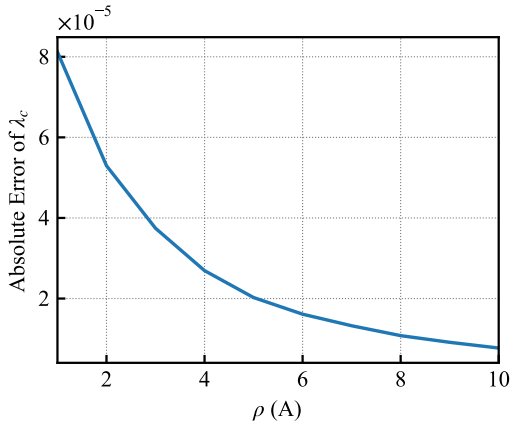


Fig. 3. Difference in λ_c according to conditions 1) and 3)

In Fig. 3, regardless of the spectral radius of the matrix A, the order of magnitude of the gap between the critical lambda points derived from conditions 1) and 3) is 10^{-5}. The error, on the order of 10^{-5}, is negligible relative to λ_c itself and can be attributed to machine precision errors during matrix convergence and linear inequality solving processes. Therefore, we consider conditions 1) and 3) to be equivalent, and Theorem 3 is accordingly verified.

C. Verification of Corollary 4

Corollary 4 states that the convergence of the expected covariance is related to the attack success rate λ, independent of the specific attack designs. We validate Corollary 4 through an example with the following parameter settings:

$$A = \begin{bmatrix} 1.25 & 0 \\ 1 & 1.1 \end{bmatrix}, \; C = \begin{bmatrix} 1 & 1 \end{bmatrix}, \; Q = 20I_2, \; R = 2.5$$

and three kinds of attack impact Δ_1, Δ_2, and Δ_3 are equal to 0, $1.5\bar{K}\bar{\Sigma}\bar{K}^\top$, and $3\bar{K}\bar{\Sigma}\bar{K}^\top$, respectively.

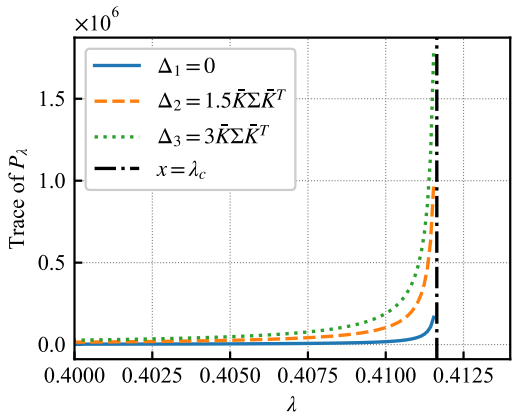


Fig. 4. The convergence point of the covariance expectation P_λ varies with λ under three attack designs.

As illustrated in Fig. 4, despite variations in the convergence point P_λ across three different attack designs, a consistent pattern emerges: P_λ demonstrates a rapid increase when the attack success rate λ exceeds 0.41, ultimately approaching infinity at the critical value λ_c. On the other hand, when $\lambda > \lambda_c$, $\mathbb{E}[\tilde{P}_k^-]$ diverges under all three attack designs. These simulation results corroborate Corollary 4, indicating that the convergence or divergence of $\mathbb{E}[\tilde{P}_k^-]$ is independent of the specific attack design employed.

D. Performance of Filtering

To evaluate the filtering performance, we consider a linearized discrete-time SEIR epidemic model. The state vector is defined as $x_k = [E_k, I_k, R_k]^\top$, representing the Exposed, Infected, and Removed populations, respectively. With a sampling interval $\Delta t = 1$ day, the state transition matrix A and measurement matrix C are given by:

$$A = \begin{bmatrix} 1-\sigma & \beta S_0/N & 0 \\ \sigma & 1-\gamma & 0 \\ 0 & \gamma & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

where the assumption $S_k \approx S_0$ is used for linearization. The parameters are set to $\sigma = 0.2$ (incubation rate) and $\gamma = 0.1$ (recovery rate). The effective transmission rate is set to $\beta S_0/N = 0.3$, corresponding to a basic reproduction number $R_0 = 3$. The process and measurement noise covariances are set to $Q = 5I_3$ and $R = 10I_2$, respectively.

The critical point, λ_c, is approximately 0.74267 (i.e., $\mathbb{E}[\tilde{P}_k^-]$ diverges when λ exceeds λ_c, otherwise it converges). The tested values of λ include 0.2, 0.7, and 0.8, which are significantly below, slightly below, and slightly above λ_c, respectively. Three different attack designs were employed:

1) White noise sequence attack. The parameters are set as $T_k = 0$, $\Theta_k = \bar{\Sigma}$, and $\Delta = \bar{K}\bar{\Sigma}\bar{K}^\top$. This strategy maintains strict stealthiness without real-time measurements y_k or full knowledge of system parameters, relying solely on the predicted innovation $C\tilde{x}_k^-$ and the error covariance $\bar{\Sigma}$.
2) One-step optimal attack with stealthiness constraint $\delta = 0$. The parameters are set as $T_k = -I$, $\Theta_k = 0$, and $\Delta = 3\bar{K}\bar{\Sigma}\bar{K}^\top$. This strategy requires complete knowledge of the system parameters and real-time measurements, as outlined in Assumptions 1 and 2, respectively. The design is optimized subject to a strict stealthiness constraint $\delta = 0$.
3) One-step optimal attack with stealthiness constraint $\delta = 1$. The parameters are set as

$$T_k^* = -\left(I - \frac{1}{\eta} C\bar{P}^2 C^\top \bar{\Sigma}^{-1}\right)^{-1}, \quad \Theta_k = 0,$$
$$(\Sigma_k^a)^* = \bar{\mathcal{M}}^{-\frac{1}{2}} \left((\bar{\mathcal{M}}^{\frac{1}{2}} \bar{\Sigma} \bar{\mathcal{M}}^{\frac{1}{2}})^{-\frac{1}{2}} - I/\eta\right)^{-2} \bar{\mathcal{M}}^{-\frac{1}{2}},$$

where the computation of η and $\bar{\mathcal{M}}$ is detailed in Lemma 2. The information structure is identical to the second case, differing only in the relaxed stealthiness constraint, i.e., $\delta = 1$.

Based on the attack design and the corresponding success rate λ, we conduct comparative analyses across three distinct scenarios. Each scenario is subjected to 5×10^7 Monte Carlo simulations to ensure statistical reliability:

1) Performance of different attack designs under low success rate λ;
2) Performance of different attack designs under high success rate λ;
3) Performance of the same attack design across varying λ values.

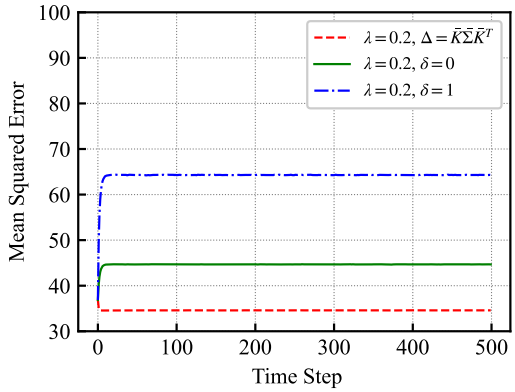


Fig. 5. Performance of different attack designs under low success rate λ

In Fig. 5, when $\lambda = 0.2$, significantly below λ_c, the MSE of remote state estimation shows substantial variations across three attack scenarios. Specifically, the MSE increases markedly when the stealthiness constraint $\delta = 1$ compared to $\delta = 0$, whereas the MSE under white noise sequences remains significantly lower than both cases. These simulation results clearly indicate that at low λ values, the attack design can substantially affect the performance of remote state estimation.

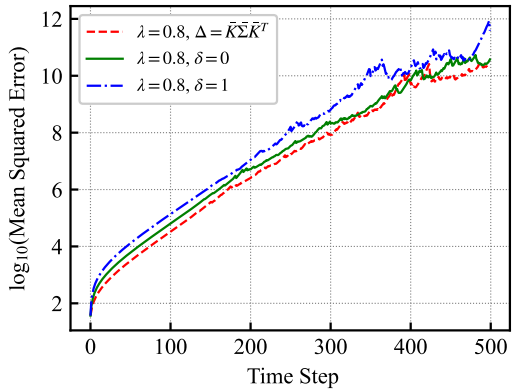


Fig. 6. Performance of different attack designs under high success rate λ

In Fig. 6, when λ slightly exceeds the critical value λ_c, the MSE on a logarithmic scale, $\log_{10}(\text{MSE})$, exhibits a remarkably similar divergent behavior across the three distinct attack designs (white noise sequence attack, one-step optimal attacks with stealthiness constraint $\delta = 0$ and $\delta = 1$). The performance curves obtained from Monte Carlo simulations exhibit oscillations and crossovers at large time steps (e.g. near $k = 400$). This phenomenon is attributed to the cumulative effect of stochasticity introduced by the probabilistic nature of the attack outcome at each step. The experimental results demonstrate that while the one-step optimal attack design accelerates the divergence compared to general attacks, it does not determine the system's convergence outcome. This finding substantiates the conclusion presented in Corollary 4.

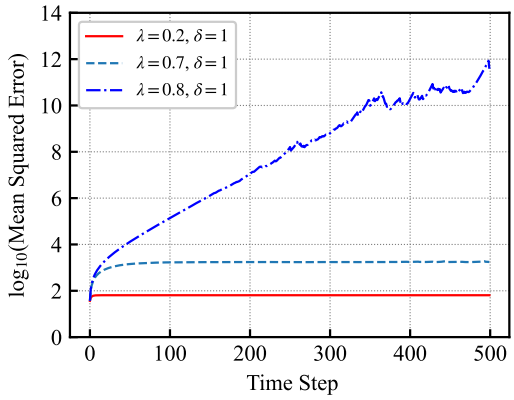


Fig. 7. Performance of the same attack design across varying λ

In Fig. 7, as λ increases from 0.2 to 0.7 (an increment of 0.5), $\log_{10}(\text{MSE})$ still remains stable, with variations within two orders of magnitude. In contrast, when $\lambda = 0.8$ (an increment of 0.1), $\log_{10}(\text{MSE})$ exhibits approximately linear growth with the time step after the initial 50 steps. Within the first 500 time steps, the MSE for $\lambda = 0.8$ exceeds that for $\lambda = 0.7$ by eight orders of magnitude. These results confirm the decisive role of λ in convergence/divergence and highlight the rapid change in system estimation error characteristics near λ_c.

VI. Conclusion

This paper has investigated stealthy attacks with stochastic success rates in the context of remote state estimation systems. The evolution of the remote estimation error covariance has been analyzed, and it has been shown that a finite number of attacks has not compromised asymptotic convergence. This finding has motivated the central focus of the study: the convergence properties of the expected estimation error covariance. A sufficient condition, a necessary condition, and a necessary and sufficient condition for divergence have been established. Furthermore, it has been demonstrated that the asymptotic convergence behavior has depended solely on the attack success rate and has remained invariant to the specific design of the attacks.

Future research may explore extensions to multi-agent and distributed estimation frameworks, where the interplay between attack propagation and network topology becomes critical [2], [39], [42]. Another promising direction lies in developing real-time detection and mitigation strategies based on data-driven inference of attack success rates [29], [43], [45]. Moreover, robustness against adaptive and learning-based stealthy attackers remains an open and practically significant challenge [13].

REFERENCES

[1] 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.

[2] R. Caballero-Águila, M. P. Frías-Bustamante and A. Oya-Lechuga, Least-squares linear estimation for multirate uncertain systems subject to DoS attacks, *International Journal of Network Dynamics and Intelligence*, vol. 4, no. 2, art. no. 100014, Jun. 2025.

[3] R. Caballero-Águila, J. Hu and J. Linares-Pérez, Distributed estimation for uncertain systems subject to measurement quantization and adversarial attacks, *Information Fusion*, vol. 120, art. no. 103044, Aug. 2025.

[4] R. Caballero-Águila and J. Linares-Pérez, Centralized fusion estimation in networked systems: addressing deception attacks and packet dropouts with a zero-order hold approach, *International Journal of Network Dynamics and Intelligence*, vol. 3, no. 4, art. no. 100021, Dec. 2024.

[5] W. Chen, J. Hu, Z. Wu and S. Ma, A survey on fault detection for networked systems under communication constraints, *Systems Science & Control Engineering*, vol. 13, no. 1, art. no. 2460434, 2025.

[6] X. Chen, L. Peng, Y. Yi, Z. Ji, Optimal stealthy deception attack strategy under energy constraints, *Communications in Nonlinear Science and Numerical Simulation*, vol. 140, Part 2, p. 108423, Jan. 2025.

[7] D. Ding, Z. Wang, D. W. C. Ho, and G. Wei, Distributed recursive filtering for stochastic systems under uniform quantizations and deception attacks through sensor networks, *Automatica*, vol. 78, pp. 231-240, Apr. 2017.

[8] H. Guo, J. Sun, and Z.-H. Pang, Stealthy false data injection attacks with resource constraints against multi-sensor estimation systems, *ISA Transactions*, vol. 127, pp. 32–40, Aug. 2022.

[9] H. Guo, J. Sun, and Z.-H. Pang, Residual-based false data injection attacks against multi-sensor estimation systems, *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 5, pp. 1181–1191, May. 2023.

[10] Z. Guo, D. Shi, K. H. Johansson, and L. Shi, Optimal linear cyber-attack on remote state estimation, *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 4–13, Mar. 2017.

[11] Z. Guo, D. Shi, K. H. Johansson, and L. Shi, Worst-case innovation-based integrity attacks with side information on remote state estimation, *IEEE Transactions on Control of Network Systems*, vol. 6, no. 1, pp. 48–59, Mar. 2019.

[12] 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.

[13] A. Hasan, I. Kuncara, A. Widyotriatmo, O. Osen and R. T. Bye, Secure state estimation and control for autonomous ships under cyberattacks, *Systems Science & Control Engineering*, vol. 13, no. 1, art. no. 2518964, 2025.

[14] J. Hu, B. Xu, R. Caballero-Águila, C. Jia and H. Dong, Distributed state estimation for nonlinear dynamical networks with stochastic topological structures subject to random deception attacks and bit-rate constraints, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 55, no. 6, pp. 3976–3988, Jun. 2025.

[15] J. Huang, Y. Tang, W. Yang and F. Li, Resilient consensus-based distributed filtering: convergence analysis under stealthy attacks, *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4878–4888, Jul. 2020.

[16] M. Huang and S. Dey, Kalman filtering with Markovian packet losses and stability criteria, *Proceedings of the 45th IEEE Conference on Decision and Control*, pp. 5621-5626, 2006.

[17] X. Li, P. Zhang and H. Dong, A robust covert attack strategy for a class of uncertain cyber-physical systems, *IEEE Transactions on Automatic Control*, vol. 69, no. 3, pp. 1983–1990, Mar. 2024.

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

[19] Y. Li and G. Yang, Optimal stealthy false data injection attacks in cyberphysical systems, *Information Sciences*, vol. 481, pp. 474–490, Nov. 2019.

[20] Y. Li, Y. Yang, Z. Zhao, J. Zhou, and D. E. Quevedo, Deception attacks on remote estimation with disclosure and disruption resources, *IEEE Transactions on Automatic Control*, vol. 68, no. 7, pp. 4096–4112, Jul. 2023.

[21] Y. Li and G. Yang, Optimal deception attacks against remote state estimation in cyber-physical systems, *Journal of the Franklin Institute*, vol. 357, no. 3, pp. 1832–1852, 2020.

[22] Q. Liu, Y. Nie, Z. Wang, H. Dong and C. Jiang, Binary-encoding-based quantized Kalman filter: an approximate MMSE approach, *IEEE Transactions on Automatic Control*, vol. 70, no. 5, pp. 3181–3196, May. 2025.

[23] Q. Liu, Z. Wang, H. Dong and C. Jiang, Recursive Bayesian estimation for discrete-time systems with state-dependent packet dropouts: A cross-coupled method, *IEEE Transactions on Automatic Control*, vol. 69, no. 6, pp. 3705–3716, Jun. 2024.

[24] Q. Liu, Z. Wang, X. He and D. H. Zhou, On Kalman-consensus filtering with random link failures over sensor networks, *IEEE Transactions on Automatic Control*, vol. 63, no. 8, pp. 2701–2708, Aug. 2018.

[25] X. Liu and G.-H. Yang, Optimal stealthy attacks with energy constraint against remote state estimation, *IEEE Transactions on Cybernetics*, vol. 54, no. 6, pp. 3577–3587, June 2024.

[26] X. Liu and G. Yang, Optimal intermittent deception attacks with energy constraints for cyber-physical systems , *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 54, no. 10, pp. 5889–5900, Oct. 2024.

[27] X. Liu, G. Yang, Optimal design and allocation of stealthy attacks against remote state estimation for cyber-physical systems, *Information Sciences*, vol. 676, p. 120859, 2024.

[28] Y.-W. Lv and G.-H. Yang, An adaptive cubature Kalman filter for nonlinear systems against randomly occurring injection attacks, *Applied mathematics and computation*, vol. 418, p. 126834, 2022.

[29] L. Ma, H. Zhang, G. Wang, C. Yang and L. Zhou, Security coordination control for the belt conveyor systems with false data injection attacks, *International Journal of Network Dynamics and Intelligence*, vol. 4, no. 1, art. no. 100001, Mar. 2025.

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

[31] Y. Nie, Z. Wang and Q. Liu, Expectation–maximization-based remote state estimation under state-dependent packet dropouts: from maximum a posteriori perspective, *IEEE Transactions on Automatic Control*, vol. 70, no. 3, pp. 1608–1622, Mar. 2025.

[32] C. Peng, J. Wu and E. Tian, Stochastic event-triggered H_∞ control for networked systems under denial of service attacks, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 7, pp. 4200-4210, July 2022,.

[33] L. Peng, L. Shi, X. Cao, and C. Sun, Optimal attack energy allocation against remote state estimation, *IEEE Transactions on Automatic Control*, vol. 63, no. 7, pp. 2199–2205, Jul. 2018.

[34] K. Plarre and F. Bullo, On Kalman Filtering for Detectable Systems With Intermittent Observations, *IEEE Transactions on Automatic Control*, vol. 54, no. 2, pp. 386-390, Feb. 2009.

[35] F. Qu, E. Tian, and X. Zhao, Chance-constrained H_∞ state estimation for recursive neural networks under deception attacks and energy constraints: The finite-horizon case, *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 9, pp. 6492–6503, Sep. 2023.

[36] J. Shang and T. Chen, Optimal stealthy integrity attacks on remote state estimation: The maximum utilization of historical data, *Automatica*, vol. 128, p. 109555, 2021.

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

[38] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, and S. S. Sastry, Kalman filtering with intermittent observations, *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1453–1464, Sept. 2004.

[39] B. Song, B. Sun, Z. Shang, J. Guo and X. Bai, Ultimately bounded control for nonlinear systems under denial-of-service attacks: an accumulation-based event-triggered mechanism, *International Journal of Systems Science*, vol. 56, no. 8, pp. 1713–1725, 2025.

[40] W. Song, Z. Wang, J. Wang, F. E. Alsaadi, and J. Shan, Secure particle filtering for cyber-physical systems with binary sensors under multiple attacks, *IEEE Systems Journal*, vol. 16, no. 1, pp. 603-613, 2021.

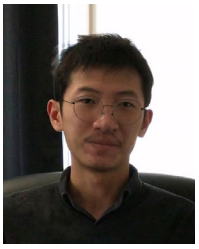
[41] Y. Song, B. Zhang, C. Wen, D. Wang and G. Wei, Model predictive control for complicated dynamic systems: a survey, *International Journal of Systems Science*, vol. 56, no. 9, pp. 2168–2193, 2025.

[42] L. Sun, T. Wu, Y. Yi, Q. Wang and Y. Zhang, Stealthy false data injection attacks against distributed multi-agent systems, *International Journal of Systems Science*, vol. 56, no. 9, pp. 2082–2096, 2025.

[43] Y.-C. Sun, K. Gao, L. Chen, F. Yang and L. Yao, Optimal power schedule for distributed Kalman filtering under DoS attacks: a Stackelberg game strategy, *International Journal of Systems Science*, vol. 56, no. 9, pp. 2067–2081, 2025.

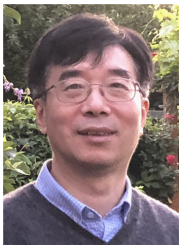
[44] A. Teixeira, D. Pérez, H. Sandberg, and K. H. Johansson, Attack models and scenarios for networked control systems, In: *Proc. 1st international conference on High Confidence Networked Systems (HiCoNS'12)*, pp. 55–64, 2012.
[45] Y. Wang, C. Zhang, J. Wu and E. Tian, Critical-information-based attack-defense strategy for nonlinear systems: an encoding-decoding scheme, *International Journal of Systems Science*, vol. 56, no. 7, pp. 1606–1615, 2025.
[46] J. Wu, G. Shi, B. D. O. Anderson and K. H. Johansson, Kalman filtering over Gilbert–Elliott channels: stability conditions and critical curve, *IEEE Transactions on Automatic Control*, vol. 63, no. 4, pp. 1003-1017, Apr. 2018.
[47] J. Wu, L. Shi, L. Xie, and K. H. Johansson, An improved stability condition for Kalman filtering with bounded Markovian packet losses, *Automatica*, vol. 62, pp. 32-38, Dec. 2015.
[48] M. Xie, D. Ding, X. Ge, Q.-L. Han, H. Dong and Y. Song, Distributed platooning control of automated vehicles subject to replay attacks based on proportional integral observers, *IEEE/CAA Journal of Automatica Sinica*, vol. 11, no. 9, pp. 1954–1966, Sept. 2024.
[49] H. Xu, Y. Yang, J. Shang, J. Fu, and Y. Li, Integrity attacks on remote estimation with spatial-temporal information sources, *Automatica*, vol. 155, p. 111172, Sep. 2023.
[50] Y. Xu, S. Chai, P. Shi, B. Zhang, and Y. Wang, Resilient and event-triggered control of stochastic jump systems under deception and denial of service attacks, *International Journal of Robust and Nonlinear Control*, vol. 33, no. 3, pp. 1821-1837, 2023.
[51] M. Yao, G. Wei, D. Ding and Y. Ju, Reset PI controller design of cyber-physical systems under constrained bit rate and DoS attacks: a hybrid system framework, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 55, no. 2, pp. 1298–1308, Feb. 2025.
[52] D. Ye and J. Wang, False data injection attack design in multi-sensor systems based on KL divergence theory, In: *Proc. 2019 IEEE 8th Data Driven Control and Learning Systems Conference (DDCLS)*, pp. 333–337, 2019.
[53] K. You, M. Fu, and L. Xie, Necessary and sufficient conditions for stability of Kalman filtering with Markovian packet losses, *IFAC Proceedings Volumes*, vol. 44, no. 1, pp. 12465-12470, Jan. 2011.
[54] H. Yu, Z. Wang, L. Zou and Y. Wang, A survey on security control and estimation for cyber-physical systems under cyber-attacks: advances, challenges and future directions, *Artificial Intelligence Science and Engineering*, vol. 1, no. 1, pp. 1–16, 2025.
[55] H. Yue, J. Zhang, J. Xia, J. H. Park and X. Xie, Adaptive intelligent control for nonlinear stochastic cyber-physical systems with unknown deception attacks: switching event-triggered scheme, *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 71, no. 12, pp. 5571-5581, Dec. 2024.
[56] H. Zhang, P. Cheng, L. Shi, and J. Chen, Optimal DoS attack scheduling in wireless networked control system, *IEEE Transactions on Control Systems Technology*, vol. 24, no. 3, pp. 843–852, May. 2016.
[57] Y. Zhang, Z. Peng, G. Wen, J. Wang and T. Huang, Optimal stealthy linear Man-in-the-Middle attacks with resource constraints on remote state estimation, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 54, no. 1, pp. 445–456, Jan. 2024.
[58] J. Zhou, J. Shang, and T. Chen, On information fusion in optimal linear FDI attacks against remote state estimation, *IEEE Transactions on Control of Network Systems*, vol. 10, no. 4, pp. 2085–2096, Dec. 2023.
[59] L. Zou, B. Song, J. Suo, N. Li and C. Wang, A survey on outlier-resistant state estimation and its applications, *Systems Science & Control Engineering*, vol. 13, no. 1, art. no. 2474471, 2025.

Lubin Zhang received the B.S. degree in mathematics and applied mathematics from the Department of Mathematics, Tongji University, Shanghai, China, in 2024. He is currently working toward the Ph.D. degree in the Department of Computer Science and Technology at the same university. His research interests include filtering theory and cyber-physical security.



Jun Shang (Senior Member, IEEE) received the B.Eng. degree in control science and engineering from Harbin Institute of Technology, Harbin, China, in 2013, and the Ph.D. degree in control science and engineering from Tsinghua University, Beijing, China, in 2018.

From September 2018 to January 2023, he was a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada. He is currently a Professor with Tongji University, Shanghai, China. His research interests include fault diagnosis, cyber-physical security, and networked control.



Zidong Wang (Fellow, IEEE) received the B.Sc. degree in mathematics from Suzhou University, Suzhou, China, in 1986, and the M.Sc. degree in applied mathematics and the Ph.D. degree in electrical engineering from Nanjing University of Science and Technology, Nanjing, China, in 1990 and 1994, respectively.

From 1990 to 2002, he held teaching and research appointments in universities in China, Germany, and U.K. He is currently a Professor of dynamical systems and computing with the Department of Computer Science, Brunel University London, Uxbridge, U.K. He has authored a number of articles in international journals. His research interests include dynamical systems, signal processing, bioinformatics, and control theory and applications.

Prof. Wang is a member of the Academia Europaea, European Academy of Sciences and Arts, and program committee for many international conferences, an Academician of the International Academy for Systems and Cybernetic Sciences, and a fellow of the Royal Statistical Society. He holds the Alexander von Humboldt Research Fellowship of Germany, the JSPS Research Fellowship of Japan, and the William Mong Visiting Research Fellowship of Hong Kong. He serves (or has served) as the Editor-in-Chief for *International Journal of Systems Science, Neurocomputing*, and *Systems Science and Control Engineering*, and an Associate Editor for 12 international journals including IEEE TRANSACTIONS ON AUTOMATIC CONTROL, IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY, IEEE TRANSACTIONS ON NEURAL NETWORKS, IEEE TRANSACTIONS ON SIGNAL PROCESSING, and IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS - PART C: APPLICATIONS AND REVIEWS.

Qinyuan Liu received the B.Eng. degree in measurement and control technology and instrumentation from Huazhong University of Science and Technology, Wuhan, China, in 2012, and the Ph.D. degree in control science and engineering from Tsinghua University, Beijing, China, in 2017.

He is currently a professor in the School of Computer Science and Technology, Tongji University, Shanghai, China. From Jul. 2015 to Sep. 2016, he was a Research Assistant in the Department of Electronic & Computer Engineering, Hong Kong University of Science and Technology, Hong Kong, China. From Jan. 2016 to Jan. 2017, he was an international researcher in the Department of Computer Science, Brunel University of London, UK. His research interests include networked control systems, multi-agent systems, and distributed filtering. He is an active reviewer for many international journals.