Computers, Materials & Continua

Tech Science Press

Doi:10.32604/cmc.2025.065946

ARTICLE

Optimized Attack and Detection on Multi-Sensor Cyber-Physical System

Fangju Zhou[1], Hanbo Zhang[2], Na Ye[1], Jing Huang[1] and Zhu Ren[1,*]

[1]School of Information Science and Engineering (School of Cyber Science and Technology), Zhejiang Sci-Tech University, Hangzhou, 310018, China

[2]School of Mechanical Engineering, Zhejiang Sci-Tech University, Hangzhou, 310018, China

*Corresponding Author: Zhu Ren. Email: zhuren@zstu.edu.cn

Received: 25 March 2025; Accepted: 20 May 2025; Published: 30 July 2025

ABSTRACT: This paper explores security risks in state estimation based on multi-sensor systems that implement a Kalman filter and a χ^2 detector. When measurements are transmitted via wireless networks to a remote estimator, the innovation sequence becomes susceptible to interception and manipulation by adversaries. We consider a class of linear deception attacks, wherein the attacker alters the innovation to degrade estimation accuracy while maintaining stealth against the detector. Given the inherent volatility of the detection function based on the χ^2 detector, we propose broadening the traditional feasibility constraint to accommodate a certain degree of deviation from the distribution of the innovation. This broadening enables the design of stealthy attacks that exploit the tolerance inherent in the detection mechanism. The state estimation error is quantified and analyzed by deriving the iteration of the error covariance matrix of the remote estimator under these conditions. The selected degree of deviation is combined with the error covariance to establish the objective function and the attack scheme is acquired by solving an optimization problem. Furthermore, we propose a novel detection algorithm that employs a majority-voting mechanism to determine whether the system is under attack, with decision parameters dynamically adjusted in response to system behavior. This approach enhances sensitivity to stealthy and persistent attacks without increasing the false alarm rate. Simulation results show that the designed leads to about a 41% rise in the trace of error covariance for stable systems and 29% for unstable systems, significantly impairing estimation performance. Concurrently, the proposed detection algorithm enhances the attack detection rate by 33% compared to conventional methods.

KEYWORDS: Cyber-physical system; kalman filter; remote state estimation; Chi-square detection; linear deception attack

1 Introduction

The deep integration of information technology with industrialization has significantly enhanced the intelligence and networking capabilities of next-generation production systems, establishing higher standards for traditional single-point technologies. In this context, Cyber-Physical Systems (CPS) have emerged as advanced systems that seamlessly integrate the physical environment, communication infrastructure, and computational resources [1,2]. CPS integrate advanced technologies in computing, communications, and control to enable dynamic regulation and real-time perception. These capabilities support the delivery of information-centric services across complex engineering domains, thereby enhancing system reliability, operational efficiency, and responsiveness [3]. These systems have found extensive applications in environmental monitoring, industrial automation, navigation, and target tracking. Furthermore, network

communication technology facilitates network-based control of physical processes, streamlining system design and deployment while enabling more flexible and efficient management [4].

Unlike traditional physical systems that operate in relatively isolated environments, modern CPS systems inherently rely on openness and interconnectivity to maximize efficiency and scalability. However, this openness also introduces vulnerabilities, exposing CPS susceptible to cyber threats within the information layer. Such threats can trigger cascading effects, potentially leading to hardware failures and significant system damage. Such as the SQL Slammer worm and the Stuxnet attacks, illustrate the serious risks that CPS security vulnerabilities pose to national security and public safety [5,6].

The CPS attack model is structured into three aspects: the adversary's *apriori* system model knowledge, disclosure abilities, and disruption resources [7]. Based on this modeling, attacks are generally grouped into three major types: 1) Denial of Service attack; 2) Replay attack; 3) False Data Injection attack [8]. Wei et al. investigated sequential DoS attacks against finite impulse response (FIR) systems, developing a parameter identification algorithm to formulate optimal attack strategies based on the covariance matrix of estimation error [9]. To mitigate impact of DoS attacks, Zhao et al. proposed an adaptive event-triggered communication mechanism. This mechanism reduces communication resource consumption and alleviates network bandwidth pressure by only transmitting data when necessary, based on specific event triggers rather than continuous transmission. In addition, they developed a combined design method to jointly tune the controller gain, and event-triggered weighting matrix [10]. Mo et al. examined how replay attacks influence system behavior and evaluated whether such attacks can succeed under certain conditions. They further proposed injecting minor variations into the control commands to help the system recognize and detect these attacks [11]. To address periodic replay threats in CPS, Li et al. designed an encryption-based method that ensures complete detection during an attack [12]. Naha et al. propose a detection method for replay attacks that integrates signal watermarking with cumulative sum testing to enhance system resilience. By optimizing the watermark signal's variance to maximize the KLD, the method significantly shortens the latency in identifying replay attacks [13]. Ni et al. investigated how reset attacks affect CPS, presented basic and advanced reset attacks, and demonstrated validity of these attacks [14]. Remote state estimation with an active eavesdropper, Ding et al. introduced a unified framework such attacks and proposed a stealthiness metric derived from the estimator's packet reception rate [15]. Pang et al. proposed a partial FDI attack strategy aimed at networked stochastic systems. This strategy degrade the performance of a Kalman filter-based output tracking control system by manipulating certain sensor measurements [16]. Xu et al. addressed event-based remote state estimation attacks by proposing a false data injection strategy aimed at evading the Chi-squared data detector while reducing the impact of the scheduler. They developed a two-channel, scheduler-oriented false data injection method by altering the numerical characteristics of the innovation signal [17]. Taking into account multiple forms of detection feedback, Li et al. proposed a novel estimation framework designed to defend against false data injection attacks [18].

Existing research on stealthy attacks has made substantial progress in the field of CPS security. Anomaly detection techniques have evolved significantly, with advanced methods leveraging deep learning models to enhance adaptability and detection sensitivity. Alzubi proposed a GRU-based detection framework that demonstrates improved performance in dynamic environments by effectively capturing temporal dependencies [19]. Furthermore, Alzubi et al. introduced a deep learning-driven detection scheme that integrates Frechet and Dirichlet distributions to enhance intrusion detection accuracy in industrial wireless sensor networks [20]. Despite recent advances, residual-based detection remains highly relevant for resource-constrained systems, owing to its minimal computational demands and ease of implementation. However, linear deception attack strategies remain limited by strict feasibility constraints, often enforcing tight residual conditions. Guo et al. formulated a linear attack with the condition that residual covariance remains

unchanged [21]. An attack strategy aimed at maximizing system degradation was developed by Liu et al, while strictly satisfying constraint $T_k \mathscr{P} T_k^T + \mathscr{B} = \mathscr{P}$ [22]. Li et al. extended their research on detecting linear deception attacks in multi-sensor remote state estimation [23]. However, their work did not investigate the impact of relaxing feasibility constraints on attack performance. To address these limitations, this paper proposes a linear deception attack framework with relaxed feasibility constraints, enabling the attacker to introduce controlled statistical deviations into the innovation signal. Furthermore, considering the characteristics of multi-sensor systems, we design an adaptive detection algorithm that compares distributed state estimates, thereby improving detection sensitivity under parameter uncertainty. The primary contributions as follows:

1. We adopt a linear deception attack form broadly applicable to multi-sensor remote estimation systems and establish the corresponding feasibility constraint. Recognizing the inherent statistical variability of the detection function derived from the χ^2 detector, we strategically relax this constraint to allow controlled deviations from the nominal innovation distribution. The recursive formulation of the error covariance under the proposed broadened constraint is rigorously derived using Kalman filtering theory, thereby enabling a precise quantification of the attack's detrimental effects on the system's estimation accuracy.
2. Under the broadened constraint scenario, we incorporate the predefined permissible deviation into the estimation error covariance to construct an optimization-based objective function. Consequently, the determination of attack parameters is transformed into a structured optimization problem. Comparative analysis demonstrates that the proposed attack strategy markedly outperforms existing approaches, yielding significantly larger estimation error covariance and thereby severely degrading system performance.
3. Since single-sensor detection methods cannot be directly applied to multi-sensor scenarios, we propose a novel adaptive detection algorithm specifically designed for such settings. This algorithm dynamically adjusts detection parameters and leverages discrepancies in inter-sensor state estimation to identify linear deception attacks. Simulation results demonstrate that our adaptive detection approach significantly reduces the missed detection rate compared to traditional fixed-parameter detection algorithms, thereby enhancing the reliability of multi-sensor CPS.

The structure of the remainder of this paper is as follows. Section 2 outlines the setup of CPS and briefly reviews essential concepts. In Section 3, we examine the features of linear attacks and derive a targeted attack strategy. A detailed account of the proposed attack detection method can be found in Section 4. Section 5 illustrates numerical results and simulation experiments. Lastly, Section 6 summarizes the contributions of study then discusses possible directions for future research.

2 Cyber-Physical System Setup

The system configuration designed to support remote state estimation under cyber attack conditions, illustrated in Fig. 1, comprises six core components: physical process, sensors, adversary, wireless communication network, remote estimator, and a false data detection mechanism. Sensors collect data from the physical process and transmit measurements to the remote estimator through the wireless network. In this configuration, the remote estimator transmits a centralized prior estimate to the sensor at each time step through a dedicated feedback channel. Although this design slightly increases communication overhead, it significantly reduces the computational burden requirements for a single sensor. To better support the design of the attack strategies and detection algorithms, it is necessary to review representative classical and emerging CPS attack and detection methods.

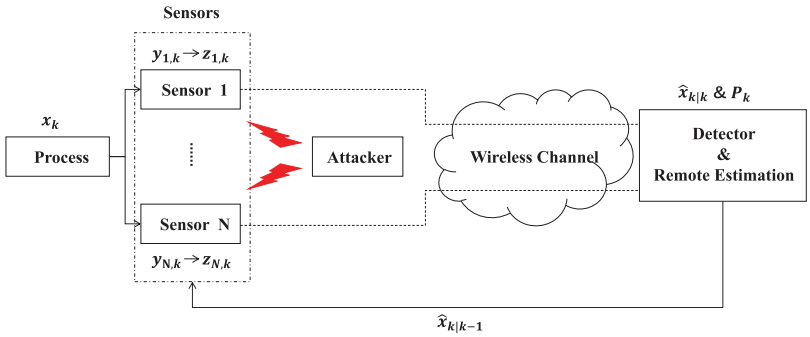

Figure 1: System architecture

As shown in Table 1, a range of attack strategies have been proposed to compromise the state estimation of CPS. Replay attack and DoS attack are simple to implement but often limited in their impact. Machine learning-based adversarial attacks demonstrate strong performance, but they generally rely on prior knowledge of training data or model structure, making them difficult to implement in real CPS environments. To address these limitations, this paper proposes a linear attack strategy under relaxed feasibility constraints. This design enables the attacker to degrade estimation performance while remaining undetectable within acceptable statistical bounds.

Table 1: Overview of CPS attack strategies

Technique	Description	Evaluation metrics	Datasets/Environments
Replay attack	Reuses previously recorded sensor data to bypass detectors	Detection rate, false alarm rate	Simulated LTI systems
DoS attack	Prevent timely state updates	Packet loss rate, estimation error	Networked CPS simulations
False data injection attack	Injects false data into sensor	Estimation deviation, detection rate	LTI systems
ML-based attack	Generates adversarial inputs	Attack success rate, detection robustness	ICS simulation datasets

Note: LTI = Linear Time-Invariant; ICS = Industrial Control System.

As summarized in Table 2, recent years have seen the development of a variety of detection techniques tailored to CPS.

Table 2: Overview of CPS detection methods

Technique	Description	Evaluation metrics	Datasets/Environments
Graph signal processing-based	Uses graph residual energy to detect	Node-level accuracy, graph residual energy	Smart grid simulation networks
ML-based detection	Learns temporal patterns of anomalies using deep neural networks	Precision, recall, F1-score	ICS benchmark datasets

(Continued)

Table 2 (continued)

Technique	Description	Evaluation metrics	Datasets/Environments
Federated learning-based	Distributed anomaly detection without centralized data sharing	Global accuracy, communication overhead	Multi-agent CPS networks
Kalman residual detection	Monitors innovation residuals	Estimation error, detection rate	Simulated linear Gaussian systems

Despite notable progress, existing detection methods face several limitations. Graph-based approaches can become computationally intensive for large-scale networks. Deep learning models require substantial training data and may struggle with limited generalization. Federated learning introduces communication and synchronization complexity, and Kalman residual methods often fail to detect stealthy or low-magnitude attacks under noisy conditions. To address these challenges, we design a detection mechanism based on adaptive that enhances sensitivity to persistent threats while maintaining a low false alarm rate. Furthermore, the approach avoids large-scale model training or distributed coordination, making it suitable for real-time deployment in noisy and resource-constrained CPS environments.

2.1 Process Model

We consider a networked system consisting of N wireless sensors and a single remote estimator, that communicate in real time. Each sensor $i \in \mathcal{N} \triangleq \{1, 2, 3, \ldots, N\}$ observes the output of a linear time-invariant process denoted by $\{x(k)\}$.

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

$$y_{i,k} = C_i x_k + v_{i,k} \tag{2}$$

$k \in \mathbb{N}$ denotes the discrete-time index, $x_k \in \mathbb{R}^n$ represents system state vector, $y_{i,k} \in \mathbb{R}^{m_i}$ indicates the measurement vector collected by sensor i. The system matrix is given by $A \in \mathbb{R}^{n\times n}$ and the observation matrix corresponding to sensor i is denoted as $C_i \in \mathbb{R}^{m_i \times n}$. The variables $w_k \in \mathbb{R}^n$ denote the process noise and $v_{i,k} \in \mathbb{R}^{m_i}$ denote the measurement noise. Both are zero-mean, independent, and identically distributed (i.i.d.) Gaussian random variables with associated covariance matrices.

$$E[w_k {w_l}^T] = \delta_{kl} Q (Q \geq 0)$$
$$E[v_{i,k} v_{j,l}^T] = \delta_{ij}\delta_{kl} R_i (R_i > 0)$$
$$E[w_k {v_{i,l}}^T] = 0, \forall k, l \in \mathbb{N}, i, j = 1, 2, 3, \ldots, N$$

The initial state x_0 is a zero-mean Gaussian random vector with a positive definite covariance matrix $\Pi_0 > 0$. It is assumed to be statistically independent of both the noise w_k and the noise $v_{i,k}$ for all $k \geq 0$.

When sensors transmit observations to a centralized fusion unit, the system behaves equivalent to that of a single sensor directly communicating with a remote estimator under real-time conditions [24]. By defining

$$C \triangleq \begin{bmatrix} C_1^T & C_2^T & C_3^T & \cdots & C_N^T \end{bmatrix}^T$$
$$y_k \triangleq \begin{bmatrix} y_{1,k}^T & y_{2,k}^T & y_{3,k}^T & \cdots & y_{N,k}^T \end{bmatrix}^T$$
$$v_k \triangleq \begin{bmatrix} v_{1,k}^T & v_{2,k}^T & v_{3,k}^T & \cdots & v_{N,k}^T \end{bmatrix}^T$$

$$R \triangleq diag\{R_1, R_2, R_3, \cdots, R_N\}$$

The total measurement equation is

$$y_k = Cx_k + v_k \tag{3}$$

v_k denotes a zero-mean Gaussian noise sequence with covariance matrix R [25]. The system is assumed to satisfy the detectability condition for the pair (A, C), the controllability condition holds for the pair $(A, \sqrt{Q})$.

2.2 Remote Estimation

Given the demands of real-time performance and high accuracy, it is computationally inefficient for each sensor to independently calculate its prior estimate using only local information. Consequently, the centralized prior estimate feedback mechanism employed in this study provides significant advantages that outweigh the minor increase in communication overhead.

At each discrete time step, sensors transmit their local measurements to the remote estimator over a wireless communication network. The estimator employs a Kalman filter to perform real-time state estimation by minimizing mean squared error. This technique operates by recursively updating state estimates through the fusion of prior predictions and incoming measurements The Kalman filtering process involves two key steps: predicting the system state and correcting it using the latest observations.

$$\hat{x}_{k|k-1} = A\hat{x}_{k-1}$$
$$P_{k|k-1} = AP_{k-1}A^T + Q$$
$$K_k = P_{k|k-1}C^T(CP_{k|k-1}C^T + R)^{-1}$$
$$\hat{x}_k = \hat{x}_{k|k-1} + K_k(y_k - C\hat{x}_{k|k-1})$$
$$P_k = (I - K_kC)P_{k|k-1}$$

where $\hat{x}_{k|k-1}$ and $\hat{x}_k$ are the *a priori* and the *a posteriori* minimum mean squared error (MMSE) estimates of the state x_k in the Kalman filter, and $P_{k|k-1}$ and P_k are the corresponding estimation error covariances. The recursion is initialized with $\hat{x}_0 = 0$ and $P_0 = \Pi_0 > 0$. The gain matrix K_k determines the weighting of the current measurement in updating the state estimate For notational clarity in the following analysis, we introduce:

$$h(X) \triangleq AXA^T + Q$$
$$\tilde{g}_i(X) \triangleq X - XC_i^T(C_iXC_i^T + R_i)^{-1}C_iX$$
$$\tilde{g}(X) \triangleq X - XC^T(CXC^T + R)^{-1}CX$$

Although Kalman filter employs a time-varying gain K_k, both the estimation error covariance and the gain matrix converge exponentially to a unique steady-state solution, irrespective of the initial conditions, provided that the pair (A, C) is detectable and $(A, \sqrt{Q})$ is controllable. The steady-state values corresponding to the local and centralized Kalman filters are defined as follows:

$$\overline{P}_i \triangleq \lim_{k\to\infty} P_{i,k|k-1}, P_i \triangleq \lim_{k\to\infty} P_{i,k}$$
$$\overline{P} \triangleq \lim_{k\to\infty} P_{k|k-1}, P \triangleq \lim_{k\to\infty} P_k$$

The matrices $\overline{P}_i$, P_i, $\overline{P}$, P represent the unique solutions to the corresponding equations, each being positive semi-definite:

$$h \circ \tilde{g}_i(X) = X, \tilde{g}_i \circ h(X) = X$$
$$h \circ \tilde{g}(X) = X, \tilde{g} \circ h(X) = X$$

The fixed-gain representations for both local and centralized Kalman filter are derived below, without loss of generality:

$$K_i \triangleq \overline{P}_i C_i^T (C_i \overline{P}_i C_i^T + R_i)^{-1}$$
$$K \triangleq \overline{P} C^T (C \overline{P} C^T + R)^{-1}$$

Under these conditions, Kalman filter operates with a fixed gain, recursive update of $\hat{x}_k$ is given by the following expression.

$$\hat{x}_k = \hat{x}_{k|k-1} + K(y_k - C\hat{x}_{k|k-1}) \tag{4}$$

In absence of attack, the communication link between sensors and remote estimator is assumed to be ideal, meaning that no packet loss, delay, or quantization distortion occurs under normal conditions. All transmitted innovation sequences are reliably received by the estimator. This assumption guarantees that any anomalies detected in the innovation statistics can be attributed solely to potential malicious attacks, rather than network-related factors.

For local Kalman filters, the innovation corresponding to sensor i is defined as $z_{i,k} \triangleq y_{i,k} - C_i \hat{x}_{i,k|k-1}$. In practical scenarios, each intelligent sensor processes its own raw measurement locally and then sends the resulting innovation to remote estimator. In distributed multi-sensor systems, individual sensors are unable to independently compute their local *a priori* estimates $\hat{x}_{i,k|k-1}$ due to the absence of information from other nodes. To address this, a more efficient strategy involves the remote estimator broadcasting a centralized *a priori* estimate $\hat{x}_{k|k-1}$ at each time instant, thereby significantly reducing communication overhead [26]. Under this strategy, the innovation for each sensor is redefined as $z_{i,k} = y_{i,k} - C_i \hat{x}_{k|k-1}$. During nominal conditions, the innovation sequence follows an independent and identically distributed (i.i.d.) pattern, characterized by a zero-mean Gaussian distribution with a specific covariance structure. In the case of centralized Kalman filtering, the corresponding innovation can be concisely expressed as follows:

$$z_k = y_k - C\hat{x}_{k|k-1}$$

Transmitting innovations instead of raw measurements offers significant advantages, as innovations typically demonstrate lower average signal amplitudes. This leads to reduced communication bandwidth requirements and decreased sensor energy consumption, thereby improving overall communication efficiency [27]. Additionally, because the innovation sequence inherently follows a zero-mean white Gaussian distribution, it offers a statistical foundation for false data detectors to reliably ascertain whether the system is subject to cyber-attacks or data anomalies.

2.3 False Data Detector

Although machine learning-based detection methods have gained popularity in recent years due to their flexibility and adaptability, the χ^2 detector remains better suited to the problem addressed in this study. χ^2 detector is constructed based on statistical distribution properties of the innovation sequence,

with clear mathematical derivations and explicit assumptions and χ^2 detection typically computed as $g_k = z_k^\top P^{-1} z_k$, involves only simple matrix operations, ensuring low computational complexity that is well suited for real-time and embedded system applications. In contrast, machine learning methods generally impose considerable computational burdens, making them less appropriate for resource-constrained environments. Moreover, the Chi-square detector does not require large volumes of training data, thereby avoiding common issues associated with machine learning, such as data availability challenges, overfitting, and generalization errors. Since the innovation sequence itself is a key statistical quantity derived from the Kalman filter, the Chi-square detector naturally integrates with the estimation framework. But machine learning methods require additional feature extraction and data processing steps, which increase system complexity. Therefore, the Chi-square detector offers a more efficient, reliable, and theoretically grounded choice for anomaly detection in multi-sensor remote estimation systems.

Theorem 1. *Consider the LTI system governed by Eqs. (1) and (2) under Kalman filtering. In this setting, the innovation $z_{i,k}$ corresponding to the i-th local Kalman filer follows a steady-state Gaussian distribution. $\mathcal{N}(0, C_i \overline{P}_i C_i^T + R_i)$ and $E[z_{i,k} z_{i,l}^T] = 0$ for all $k \neq l$.*

Proof of Theorem 1. Noting $\hat{e}_k = x_k - \hat{x}_{k|k-1}$, according to Eq. (2), rewrite $z_{i,k}$:

$$\begin{aligned} z_{i,k} &= y_{i,k} - C_i \hat{x}_{k|k-1} \\ &= C_i x_k + v_{i,k} - C_i \hat{x}_{k|k-1} \\ &= C_i \hat{e}_k + v_{i,k} \end{aligned}$$

The error covariance becomes

$$\begin{aligned} E[z_{i,k} z_{i,k}^T] &= C_i E[\hat{e}_k \hat{e}_k^T] C_i^T + E[v_{i,k} v_{i,k}^T] \\ &= C_i \overline{P}_i C_i^T + R_i \end{aligned}$$

□

In the same way, in the centralized Kalman filter framework, the innovation term $z_k = y_k - C\hat{x}_{k|k-1}$ also follows $\mathcal{N}(0, C\overline{P}C^T + R)$ with cross-time expectations $E[z_k z_l^T] = 0$ for all $k \neq l$.

The χ^2 detector identifies anomalies by evaluating the cumulative sum of the normalized innovation sequence. The detection procedure adheres to a hypothesis testing criterion at each step k

$$g_k = \sum_{j=k-J+1}^{k} z_j^T \mathscr{P}^{-1} z_j \underset{H_1}{\overset{H_0}{\lessgtr}} \eta \tag{5}$$

where $\mathscr{P} = C\overline{P}C^T + R$, J represents the detection window size, and η denotes an appropriately chosen detection threshold. With the aim of regulate the false alarm rate of detection strategy, the threshold η is determined according to a predefined significance level α.η is selected as the $(1-\alpha)$ quantile of the Chi-squared distribution $\chi^2(mJ)$, such that it satisfies the specified confidence requirement

$$\mathbb{P}(g_k > \eta \mid H_0) = \alpha$$

Null hypothesis H_0 indicates that system operates under normal conditions, whereas alternative hypothesis H_1 corresponds to an ongoing attack. The normalized detection statistic defined in Eq. (5) follows the χ^2 distribution with mJ degrees of freedom where $m = \sum_{i=1}^{N} m_i$ [28]. If the statistic g_k exceeds the predefined η, the detector issues an alarm. Otherwise, the measurement is considered normal and passes detector.

3 Linear Attack Strategy

This section formulates a linear deception attack and revisits conventional feasibility constraint. To account for the variability of the χ^2 detection statistic, we propose a generalized extension of the original constraint and its impact on system performance is analyzed. Finally, establish an objective function in conjunction with the selected deviation from the distribution of the innovation to determine specific attack parameters.

3.1 Linear Deception Attack

Consider an attacker with full knowledge of the system model and the capability to intercept and modify measurement data in real time. Given this assumption, the attacker can manipulate the innovation sequence to any desired value [29]. The corresponding strategy is expressed as

$$\tilde{z}_k = f_k(z_k) + b_k$$

$\tilde{z}_k$ denotes the innovation term that has been altered by the attacker, f_k represents a general function defined over a suitable domain, and $b_k \in \mathbb{R}^m$ is a Gaussian random vector that is independent of z_k.

However, if the function f_k is nonlinear, it becomes difficult to rigorously analyze the impact of the attack, as statistical properties of modified innovation sequence cannot be precisely characterized. In contrast, adopting a linear attack strategy enables explicit quantification of both the stealthiness constraint and the attack's effect, thereby facilitating the design of effective stealth attacks. Consequently, this study focuses on linear deception attacks, where f_k is defined as a linear operator acting on the innovation signal z_k. The corresponding attack mechanism is as follows:

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

$T_k \in \mathbb{R}^{m\times m}$ denotes a configurable attack matrix. The attacker can compute the steady-state Kalman filter configuration along with the corresponding innovation statistics. Under this assumption, the forged innovation $\tilde{z}_k$ follows an i.i.d. zero-mean Gaussian distribution with covariance $T_k \mathscr{P} T_k^T + \mathscr{B}$. Attacker is capable of intercepting and altering innovation sequences in real time, without incurring observable delays. Given limited disruption capacity, the attacker may only target a subset of sensor channels by imposing structural constraints on T_k. If forged innovation $\tilde{z}_k$ matches statistical profile of nominal innovation z_k, then the linear attack defined in Eq. (6) can evade detection, as it satisfies the test condition of Eq. (5). That is to say, $\tilde{z}_k$ must conform to the $\mathscr{N}(0, \mathscr{P})$ which means that attack condition can be defined as

$$T_k \mathscr{P} T_k^T + \mathscr{B} = \mathscr{P} \tag{7}$$

In previous work, the feasibility condition for stealthy attacks was defined as a zero-deviation constraint, indicating that the residual distribution during an attack must precisely align with the normal case. But the detection statistic g_k defined in Eq. (5) is a random variable that inherently fluctuates due to its underlying χ^2 distribution with mJ degrees of freedom. Even under normal conditions, the statistic exhibits a mean of $\mathbb{E}[g_k] = mJ$ and a variance of $\mathrm{Var}(g_k) = 2mJ$, implying that g_k does not remain constant but varies within a probabilistic confidence interval. As long as the statistics after the attack satisfy $\mathbb{P}(\tilde{g}_k \le \eta) \approx \mathbb{P}(g_k \le \eta)$, the purpose of bypassing the detector can be achieved. We express the new constraint as follows:

$$\| T_k P T_k^{\top} + \mathscr{B} - \mathscr{P} \|_F \le \varepsilon$$

After attack, $\tilde{\mathscr{P}} = T_k \mathscr{P} T_k^T + \mathscr{B}$ is the covariance of $\tilde{z}_k$. The expected detection statistics are as follows:

$$\mathbb{E}[\tilde{g}_k] = \sum_{j=k-J+1}^{k} \mathbb{E}[\tilde{z}_j^\top P^{-1} \tilde{z}_j] = \sum_{j=k-J+1}^{k} \mathrm{tr}(P^{-1}\tilde{P}) = J \cdot \mathrm{tr}(P^{-1}\tilde{P})$$

The deviation can be estimated by the trace inequality

$$|\mathbb{E}[\tilde{g}_k] - mJ| = J \cdot |\mathrm{tr}(\mathscr{P}^{-1}(\tilde{\mathscr{P}} - \mathscr{P}))| \le J \cdot \|\mathscr{P}^{-1}\|_F \cdot \|\tilde{\mathscr{P}} - \mathscr{P}\|_F \tag{8}$$

From Eq. (8), deviation of the expected detection statistic is proportional to the Frobenius norm of the covariance perturbation. Thus, by choosing ε

$$\varepsilon \le \frac{\sqrt{2mJ}}{J \cdot \|\mathscr{P}^{-1}\|_F} \tag{9}$$

Eq. (9) ensures that the detection statistics after the attack are still within the fluctuation range of normal system operation, thereby maintaining the concealment of the attack in a statistical sense.

3.2 Performance Analysis

Malicious attackers often formulate strategies aimed at undermining system reliability by introducing substantial estimation errors into the remote estimator. Given the LTI system described in Eqs.(1) and (3), and considering a linear deception attack as specified in Eq. (6), the resulting state estimate evolves as follows:

$$\tilde{x}_{k|k-1} = A\hat{x}_{k-1} \tag{10}$$

$$\tilde{x}_k = \tilde{x}_{k|k-1} + K\tilde{z}_k \tag{11}$$

When the χ^2 detector fails to identify an anomaly, and the system is mistakenly considered to be operating normally, allowing the remote estimator to continue functioning. In such cases, due to the use of compromised data, the estimated state gradually deviates from the true state, ultimately degrading overall system performance.

To quantify this deviation, we define *a priori* error $\tilde{e}_{k|k-1}$ as difference between x_k and *a priori* state estimate $\tilde{x}_{k|k-1}$ after an attack, one has $\tilde{e}_{k|k-1} = x_k - \tilde{x}_{k|k-1}$. Similarly, the *a posteriori* error $\tilde{e}_k$ is defined as the deviation between the x_k and the updated estimate $\tilde{x}_k$, given by $\tilde{e}_k = x_k - \tilde{x}_k$.

Then, the *a priori* error covariance matrix $\tilde{P}_{k|k-1}$ can be expressed as $\tilde{P}_{k|k-1} = E[\tilde{e}_{k|k-1}\tilde{e}_{k|k-1}^T]$. And the *a posteriori* error covariance matrix $\tilde{P}_k$ can be expressed as

$$\begin{aligned}
\tilde{P}_k &= E[\tilde{e}_k \tilde{e}_k^T] \\
&= E[(x_k - \tilde{x}_k)(x_k - \tilde{x}_k)^T] \\
&= E[((x_k - \tilde{x}_{k|k-1}) - K\tilde{z}_k)((x_k - \tilde{x}_{k|k-1}) - K\tilde{z}_k)^T] \\
&= \tilde{P}_{k-1} + K(T_k \mathscr{P} T_k^T + \mathscr{B})K^T - E[K\tilde{z}_k(x_k - \tilde{x}_{k|k-1})^T] - E[(x_k - \tilde{x}_{k|k-1})\tilde{z}_k^T K^T]
\end{aligned} \tag{12}$$

To obtain the last two terms of Eq. (12), substituting Eq. (11) into $\tilde{e}_{k|k-1} = x_k - \tilde{x}_{k|k-1}$, we can obtain

$$\begin{aligned}\tilde{e}_{k|k-1} &= x_k - \tilde{x}_{k|k-1} \\ &= Ax_{k-1} + w_{k-1} - A\tilde{x}_{k-1} \\ &= A(x_k - \tilde{x}_{k-1|k-2}) - AK\tilde{z}_{k-1} + w_{k-1} \\ &= A^k(x_0 - \hat{x}_{0|-1}) + \sum_{i=1}^{k} A^{i-1}w_{k-i} - \sum_{i=1}^{k} A^i K\tilde{z}_{k-i}\end{aligned} \tag{13}$$

Introducing the *a priori* error $\hat{e}_{k|k-1}$ between the x_k and the *a priori* state estimate $\hat{x}_{k|k-1}$, one has

$$\begin{aligned}\hat{e}_{k|k-1} &= x_k - \hat{x}_{k|k-1} \\ &= Ax_{k-1} + w_{k-1} - A\hat{x}_{k-1} \\ &= Ax_{k-1} + w_{k-1} - A(\hat{x}_{k-1|k-2} + Kz_{k-1}) \\ &= A(I - KC)(x_{k-1} - \hat{x}_{k-1|k-2}) + w_{k-1} - AKv_{k-1}\end{aligned} \tag{14}$$

Substituting Eqs. (3) and (14) into Eq. (6) for expansion and iteration, we can get

$$\begin{aligned}\tilde{z}_k &= T_k z_k + b_k \\ &= T_k(Cx_k + v_k - C\hat{x}_{k|k-1}) + b_k \\ &= T_k C(A(I - KC)(x_{k-1} - \hat{x}_{k-1|k-2}) + w_{k-1} - AKv_{k-1}) + T_k v_k + b_k \\ &= T_k C(A(I - KC))^k(x_0 - \hat{x}_{0|-1}) + \sum_{i=1}^{k} T_k C(A(I - KC))^{i-1}w_{k-i} \\ &- \sum_{i=1}^{k} T_k C(A(I - KC))^{i-1}AKv_{k-i} + T_k v_k + b_k\end{aligned} \tag{15}$$

It is known that $E[\tilde{z}_k\tilde{z}_l^T] = 0$ for all $k \neq l$, so the last term of Eq. (13) and $\tilde{z}_k$ are independent of each other. Since x_0, w_k, v_k and b_k are mutually independent, the last three terms of Eq. (15) and the first two terms of Eq. (13) are also independent of each other.

Based on the above analysis, the third term of Eq. (12) is obtained

$$\begin{aligned}E[K\tilde{z}_k(x_k - \tilde{x}_{k|k-1})^T] &= E[K(T_k C(A(I - KC))^k(x_0 - \hat{x}_{0|-1}) \\ &+ \sum_{i=1}^{k} T_k C(A(I - KC))^{i-1}w_{k-i}) \times (A^k(x_0 - \hat{x}_{0|-1}) + \sum_{i=1}^{k} A^{i-1}w_{k-i})^T] \\ &= KT_k C((A(I - KC))^k \times E[(x_0 - \hat{x}_{0|-1})(x_0 - \hat{x}_{0|-1})^T](A^k)^T \\ &+ \sum_{i=1}^{k} (A(I - KC))^{i-1}E[w_{k-i}w_{k-i}^T](A^{i-1})^T) \\ &= KT_k C((A(I - KC))^k\overline{P}(A^k)^T + \sum_{i=1}^{k} (A(I - KC))^{i-1}Q(A^{i-1})^T \\ &= KT_k C\overline{P}\end{aligned} \tag{16}$$

Similarly, the fourth term of Eq. (12) follows:

$$E[(x_k - \tilde{x}_{k|k-1})\tilde{z}_k^T K^T] = \overline{P}C^T T_k^T K^T \tag{17}$$

Therefore, the error covariance can be expressed as follows:

$$\tilde{P}_k = A\tilde{P}_{k-1}A^T + Q + K(T_k\mathscr{P}T_k^T + \mathscr{B})K^T - KT_kC\overline{P} - \overline{P}C^TT_k^TK^T \tag{18}$$

3.3 Computation of the Optimal Attack Strategy

When $T_k\mathscr{P}T_k^T + \mathscr{B} \neq \mathscr{P}$, to achieve optimal impact, attacker aims to maximize $\tilde{P}_k$, as defined in Eq. (18), under linear deception attack. Specifically, the objective is to maximize $tr(\tilde{P}_k)$.

Attacker predefines the random variable b_k to follow a Gaussian distribution characterized by zero mean and covariance $\mathscr{B}$. By selecting a minor deviation from the innovation distribution, the attack matrix T_k can be determined under the widening constraint by formulating the objective function presented in Eq. (19).

$$\begin{aligned} &\max_{T_k\in\mathbb{R}^{m\times m}} tr(\tilde{P}_k) \\ &s.t.\|T_k\mathscr{P}T_k^T + \mathscr{B} - \mathscr{P}\|_F = \varepsilon \end{aligned} \tag{19}$$

where ε represents the deviation from the distribution of the innovation chosen by the attacker and given by Eq. (9).

Further analysis of Eqs. (18) and (19), it can be seen that maximizing the trace of the error covariance matrix in Eq. (18) is mathematically equivalent to optimizing the objective function $tr(K(T_k\mathscr{P}T_k^T + \mathscr{B})K^T - KT_kC\overline{P} - \overline{P}C^TT_k^TK^T)$. As a result, the problem of solving the attack strategy under the widening constraint shown in Eq. (19) can be transformed into an optimization problem as follows:

$$\begin{aligned} &\max_{T_k\in\mathbb{R}^{m\times m}} tr(K(T_k\mathscr{P}T_k^T + \mathscr{B})K^T - KT_kC\overline{P} - \overline{P}C^TT_k^TK^T) \\ &s.t.\|T_k\mathscr{P}T_k^T + \mathscr{B} - \mathscr{P}\|_F = \varepsilon \end{aligned} \tag{20}$$

The attacker through several means, such as insider threats or the leakage of system parameters by staff can obtain system matrices A and C. The noise covariances Q and R can be estimated through statistically analyzing the measurement sequences collected during periods of normal system operation. With knowledge of these parameters, the attacker is able to compute the steady-state covariance matrix $\overline{P}$ by solving the associated Riccati equation. Additionally, the attacker can derive the innovation sequence, which is modeled as a Gaussian distribution with zero mean and covariance $\mathscr{P}$. In practical engineering applications, when dimensions of attack matrix $T_k \in \mathbb{R}^{m\times n}$ are moderate, obtaining a closed-form solution for the optimization problem can be computationally challenging. Here a gradient-based numerical optimization approach is employed to approximate the optimal solution of the attack matrix.

To facilitate optimization, we present a Lagrangian formulation

$$\mathscr{L}(T_k,\lambda) = \mathrm{tr}(KMK^\top - KT_kC\tilde{P} - \tilde{P}C^\top T_k^\top K^\top) + \lambda\left(\|M-\mathscr{P}\|_F^2 - \varepsilon^2\right) \tag{21}$$

where $M = T_k\mathscr{P}T_k^\top + \mathscr{B}$ and $E = M - \mathscr{P}$. The Eq. (21) becomes

$$\mathscr{L}(T_k,\lambda) = \mathrm{tr}(KT_kPT_k^\top K^\top) - 2\mathrm{tr}(KT_kC\tilde{P}) + \lambda\cdot\mathrm{tr}(E^\top E) + \mathrm{tr}(K\mathscr{B}K^\top)$$

The gradient is obtained by differentiating the objective function $\mathscr{L}$ with respect to T_k

$$\begin{aligned} \frac{\partial\mathscr{L}}{\partial T_k} &= \frac{\partial}{\partial T_k}\left(\mathrm{tr}(KT_k\mathscr{P}T_k^\top K^\top) - \mathrm{tr}(KT_kC\tilde{P}) - \mathrm{tr}(\tilde{P}C^\top T_k^\top K^\top) + \lambda\cdot\mathrm{tr}(E^\top E)\right) \\ &= 2K^\top KT_k\mathscr{P} - K^\top C\tilde{P}^\top - K^\top C\tilde{P}^\top + 2\lambda(E)\mathscr{P}T_k \end{aligned}$$

$$= K^\top K T_k \mathscr{P} - K^\top C \tilde{P}^\top + 2\lambda (T_k \mathscr{P} T_k^\top + \mathscr{B} - \mathscr{P}) \mathscr{P} T_k$$

where $E = T_k \mathscr{P} T_k^\top + \mathscr{B} - \mathscr{P}$. A gradient descent scheme is then applied, updating T_k iteratively as follows:

$$T_k^{(i+1)} = T_k^{(i)} - \eta \cdot \nabla_{T_k} \mathscr{L}(T_k^{(i)}, \lambda)$$

The step size η determines the magnitude of each update during the iterative optimization. The procedure is terminated when any of the following conditions is met: Frobenius norm of the gradient satisfies $\|\nabla_{T_k} \mathscr{L}\|_F < \delta$; or the constraint $\|T_k P T_k^\top + \mathscr{B} - \mathscr{P}\|_F \approx \varepsilon$ is approximately fulfilled. This design of optimized attack strategies is of broad relevance in CPS security, particularly in domains such as electric vehicle infrastructure [30].

The proposed attack framework does not require simultaneous interference with all sensor channels. Instead, the attacker can selectively target a subset of sensor innovations by strategically combining attack resources. While the attack matrix $T_k \in \mathbb{R}^{m \times m}$ is obtained though solving the optimization problem in Eq. (20), the problem formulation itself is under the attacker's control. Structural constraints can be imposed on T_k to enable structured attacks.

The solution obtained is denoted as T_k^*. The attack algorithm at **Algorithm 1**.

Algorithm 1: Attack Algorithm

Input: $\hat{x}_0$, P_0, $\mathscr{B}$
Output: $\tilde{x}_k$, $\tilde{P}_k$
/*During normal operation of the system*/
for $k = 1 : step$ **do**
  $\hat{x}_{k|k-1} = A\hat{x}_{k-1}$;
  $P_{k|k-1} = AP_{k-1}A^T + Q$;
  $K_k = P_{k|k-1}C^T(CP_{k|k-1}C^T + R)^{-1}$;
  $\hat{x}_k = \hat{x}_{k|k-1} + K_k(y_k - C\hat{x}_{k|k-1})$;
  $P_k = (I - K_k C)P_{k|k-1}$;
end
/*System reaches steady state*/
$\overline{P} = P_{k|k-1}$;
$\mathscr{P} = C\overline{P}C^T + R$;
$K = \overline{P}C^T(C\overline{P}C^T + R)^{-1}$;
/*System reaches steady state*/
for $k = 1 : step$ **do**
  if *Launch Attack* **then**
    Obtain the T_k^* by solving Eq. (21);
    Replace the original innovation z_k by $\tilde{z}_k = T_k^* z_k + b_k$;
  else
    Keep the original innovation at step k;
  end
end

After solving Eq. (20) to obtain T_k^* and designing a specific attack strategy based on $\tilde{z}_k = T_k^* z_k + b_k$, the attacker can first compute the value of the χ^2 detection function under attack using Eq. (5) to determine whether it falls within the normal fluctuation range before deciding to execute the attack. If an attack occurs

and the χ^2 detection function value remains below the threshold established by the system, the detector will interpret the system as operating normally.

The proposed attack strategy relaxes the traditional strict feasibility constraint by introducing a small deviation bounded by a tolerance parameter ε, as shown in Eq. (20). Traditional detection schemes based on residual monitoring are insufficient, because they assume strict adherence to the nominal distribution and lack mechanisms to detect small but systematic deviations. The attack discussed in this paper precisely exploits this statistical uncertainty. Effective defense would either require significantly tightening the detection thresholds, which would inevitably lead to a higher false alarm rate, or introducing complex multi-dimensional detection frameworks, which increase the risk of missed detections due to difficulties in parameter tuning.

Moreover, the attack formulation explicitly integrates a bounded relaxation of the detection feasibility constraint. As defined in Eq. (20), the objective is to maximize the degradation of remote estimator's error covariance, subject to the relaxed constraint $|T_k \mathscr{P} T_k^{\top} + \mathscr{B} - \mathscr{P}|_F \leq \varepsilon$. The parameter ε explicitly controls the allowable statistical deviation, thus providing a trade-off mechanism: a larger ε permits more powerful attacks but increases the risk of detection, whereas a smaller ε ensures better stealthiness but limits the attack impact. This design enables the attacker to flexibly balance between effectiveness and stealth.

Real-world CPS face several practical constraints, such as communication noise, limited computational resources, and strict real-time requirements. However, the proposed methods remain practical. Matrix operations involved in Eqs. (8) and (21), such as trace evaluations and Frobenius norm calculations, scale quadratically with the number of sensors, which keeps the computational burden manageable for embedded processors typically used in smart grid substations.

4 Detection of Linear Attack

As previously discussed, a linear attack can evade detection by conventional detectors. To determine whether any sensors have been compromised, a Kalman filter can be employed to estimate the measurements of each individual sensor.

At each time k, note $\Delta\hat{x}_{ij,k} = \hat{x}_{i,k} - \hat{x}_{j,k} = (\hat{x}_{i,k} - x_k) - (\hat{x}_{j,k} - x_k)$ where $\hat{x}_{i,k}$ represents the *a posteriori* state estimate of the i-th sensor and $\hat{x}_{j,k}$ represents the *a posteriori* state estimate of the j-th sensor.

In the absence of attacks and under steady-state conditions, we are able to obtain that $\hat{x}_{i,k} - x_k$ and $\hat{x}_{j,k} - x_k$ are zero-mean Gaussian. Based on statistical knowledge, $\Delta\hat{x}_{ij,k}$ follows Gaussian distribution $\mathcal{N}(0, P_{ij,k})$ where the covariance term $P_{ij,k}$ can be obtained in advance through process simulation. However, in presence of an attack, the *a posteriori* compromised sensor will deviate from its nominal distribution, causing a statistically significant shift in the value of $\Delta\hat{x}_{ij,k}$. This shift disrupts the expected Gaussian consistency between sensors, allowing anomalies to be detected through inter-sensor discrepancies.

Therefore, we consider the security issues in this case and propose a method to detect whether the system is under attack by comparing the change of a new detection indicator. Specifically, any pair of distinct sensors, denoted as the i-th and j-th sensors ($i, j = 1, 2, \ldots, N,\ i \neq j$), can be arbitrarily selected. The corresponding detection indicator is defined as follows:

$$G_{ij,k}^{\mathscr{J}} = \sum_{h=k-\mathscr{J}+1}^{k} (\Delta\hat{x}_{ij,h})^T P_{ij,h}^{-1} (\Delta\hat{x}_{ij,h}) \underset{H_1}{\overset{H_0}{\lessgtr}} \delta_{ij,k} \tag{22}$$

Let $\mathscr{J}$ denote the detection window size, and $\delta_{ij,k}$ denotes the threshold. For the two sensors, the normalized sum in Eq. (22) conforms to a χ^2 distribution with $n\mathscr{J}$ degrees of freedom under normal

conditions. However, if an attack occurs, the distribution characteristics are expected to deviate from the nominal pattern. Eq. (22) does not directly rely on the innovation sequence, it utilizes the statistical consistency among multiple sensors by comparing the posterior estimates $\hat{x}_{i,k}$ and $\hat{x}_{j,k}$ obtained from different sensors.

Combined with Eq. (22), we propose a dual-stage detection method to balance these trade-offs by adjusting the detection window length L, the effective rejection threshold M, and the single-sample detection threshold η maintain a low false positive rate while minimizing the probability of missed detections.

Firstly, according to **Algorithm 2**, we need to set the parameters maximum detection window length L and effective rejection threshold M. The choice of L should reflect the dynamic characteristics of the system. For systems with rapidly varying states, a smaller L enables prompt detection of anomalies to ensure timely detection of anomalies. But systems with higher noise levels require a larger L to effectively smooth out random fluctuations. The optimal value of L can be determined empirically through simulation under the assumed attack model. The parameter M controls the rejection threshold, a lower value (e.g., $M = [L/2]$) is suitable for systems requiring higher sensitivity. Conversely, for applications that prioritize reliability and low false alarm rates, a more conservative threshold (e.g., $M = \lfloor 2L/3 \rfloor$ to $L-1$) is recommended. When L and M are determined, the detection procedure incrementally increases the window length $\mathscr{J}$ from 1 to L, allowing the detector to adaptively accumulate evidence over multiple time scales.

In contrast to fixed-window χ^2 detection, **Algorithm 2** dynamically adjusts the window length. Smaller windows provide rapid response to strong anomalies, while larger windows accumulate evidence to capture weak or stealthy deviations. To reduce the risk of misjudgment from a single detection window, a multi-window voting scheme is used. The system is considered under attack only if at least M out of L window-based tests report anomalies. The additional overhead compared to traditional χ^2 detectors is minor, making the proposed method suitable for real-time implementation in resource-constrained environments. After obtaining preliminary results, conduct 100 cycles to confirm the final system status.

Algorithm 2: Detection Algorithm

Step 1: Given $2 \le L \le 20, L, M \in \mathbb{Z}^+$.
Step 2: Let $\mathscr{J}$ =1, $i, j \in N$, when $G^1_{ij,k} > \delta^1_{ij,k}$, where $\delta^1_{ij,k}$ is the fractile with confidence level α, we reject H_0.
Step 3: If $G^1_{ij,k} \le \delta^1_{ij,k}$, try to test $G^{\mathscr{J}}_{ij,k}$ and $\delta^{\mathscr{J}}_{ij,k}$ where $\mathscr{J} = 2, 3, \ldots, L$.
Step 4: When L-M of the tests are false, we state H_0 cannot be rejected.

While the **Algorithm 2** improves sensitivity to stealthy linear attacks by aggregating residual decisions over a window, certain real-world scenarios may still limit its effectiveness. If the injected attack signals are correlated, the residual inconsistencies could be masked, violating the statistical assumptions in Eq. (22). Additionally, if an attacker can adapt its strategy based on detection outcomes in real time, the fixed-length window aggregation might not react quickly enough to capture rapid changes. Furthermore, our approach assumes that measurement noise and packet losses across different sensors are independent. In practical systems where disturbances are correlated or bursty failures occur, the detection sensitivity could degrade. Although **Algorithm 2** improves the detection rate, it inherently introduces a longer decision window to ensure robustness against random fluctuations. As trade-off between rapid detection and reliable decision-making is observed, and optimizing this trade-off remains a topic for future investigation.

5 Simulation Examples

This section presents simulation results that evaluate the effectiveness of the proposed linear deception attack and its associated detection approach.

5.1 Stable Process under Linear Attack

We consider a dynamic model characterized by the following parameters

$$A = \begin{bmatrix} 0.6 & 0.4 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.4 & 0.3 \\ 0 & 0 & 0 & 0.2 \end{bmatrix},$$

$C_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}$,
$C_2 = \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix}$,
$C_3 = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}$,
$C_4 = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}$,
$R_1 = 0.1$, $R_2 = 0.2$, $R_3 = 0.3$, $R_4 = 0.4$ and $\hat{x}_0 = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}^T$.

When the system operates in a safe and steady state, remote estimator employs Kalman filter to perform state estimation and derives the traces of the system state and its corresponding estimation error covariance, as illustrated in Figs. 2 and 3.

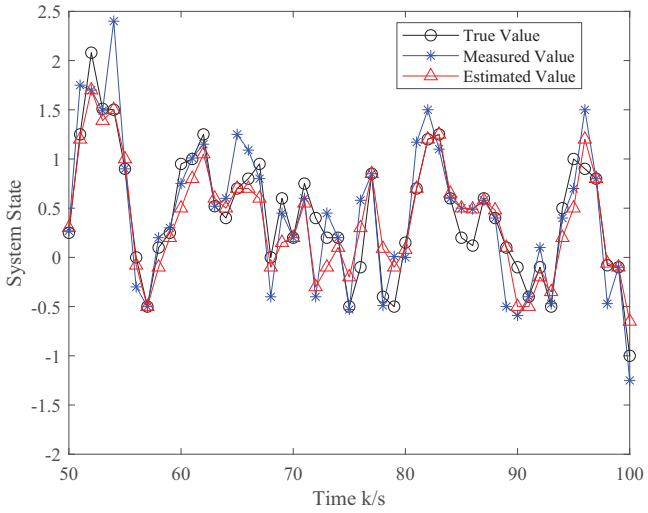

Figure 2: System status

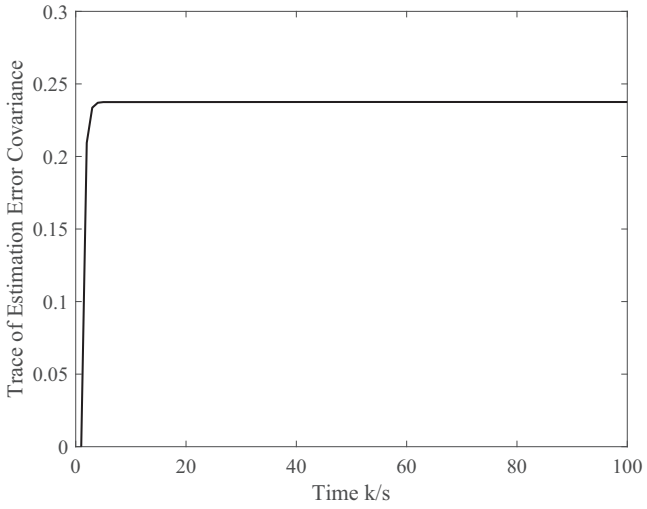

Figure 3: Trace of estimation error covariance

During the interval $[0, 50]$, the remote estimator operates under the Kalman filtering framework and attains a steady-state condition. To initiate a cyber attack, the adversary injects falsified innovation signals, specifically $\tilde{z}_k = -Iz_k$ and $\tilde{z}_k = T_k^* z_k$, over the interval $[83, 93]$. The corresponding simulation results, including the system state estimation and the trace of the error covariance matrix, are illustrated in Figs. 4 and 5.

As shown by the purple and red curves in Fig. 4, it is evident that a linear attack using $\tilde{z}_k = -Iz_k$ or $\tilde{z}_k = T_k^* z_k$ results in the state estimate gradually deviating from both the true system state x_k and the Kalman filter estimate $\hat{x}_k$. The red and yellow curves in Fig. 5 indicate that under a linear attack, the $\mathrm{tr}(\tilde{P}_k)$ exceeds the value observed during normal system operation, and the error covariance will converge. The Figs. 4 and 5 demonstrate that both attack strategies effectively disrupt system performance. Additionally, Fig. 5 shows that $\mathrm{tr}(\tilde{P}_k)$ under the attack $\tilde{z}_k = T_k^* z_k$ is larger than that corresponding to attack with $\tilde{z}_k = -Iz_k$ during the same time period. Although the error covariance increases under attack, it remains bounded, indicating that the system maintains practical stability without exhibiting divergent behavior.

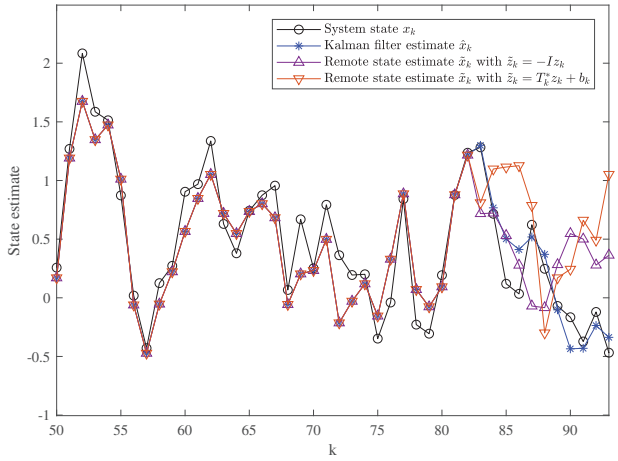

Figure 4: State estimate

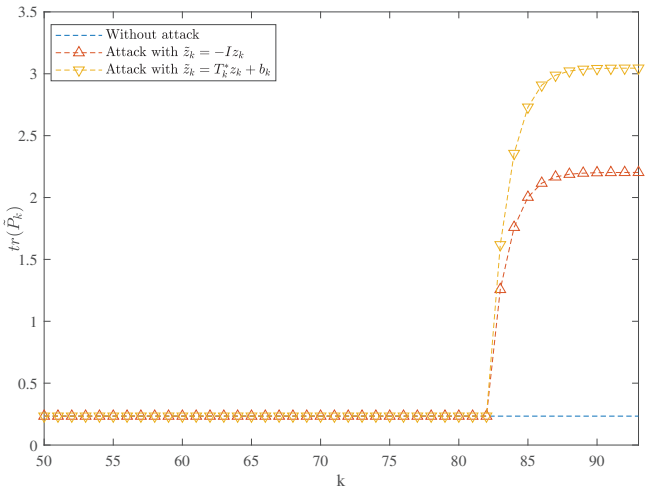

Figure 5: The trace of remote estimation error covariance

Detection statistic values based on the Chi-square detector, calculated according to Eq. (5) under different system operating conditions, are shown in Fig. 6.

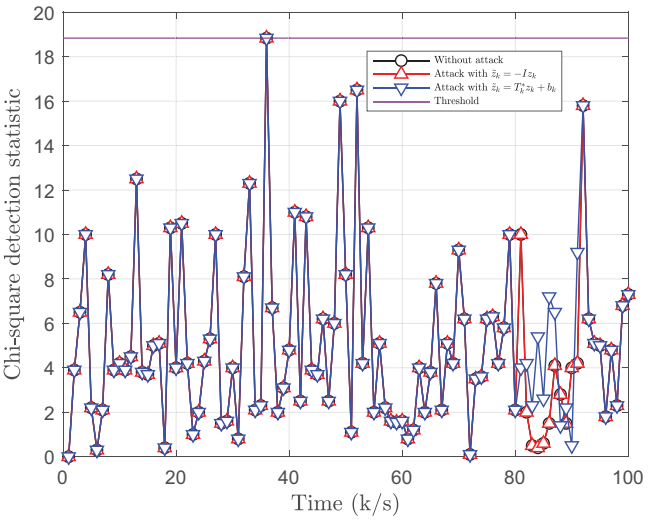

Figure 6: Detection function based on Chi-square detector

As shown in Fig. 6, during normal operation, the maximum detection statistic value reaches 18.8346. This value is selected as the detection threshold, i.e., $\eta = 18.8346$. When an attacker implements the strategy $\tilde{z}_k = T_k^* z_k$ during the interval $k = 83$ to $k = 93$, the detection statistic remains below the threshold η. The detector erroneously classifies the system as operating normally and fails to trigger an alarm, allowing the remote estimator to continue updating the state estimates using the Kalman filter. This result demonstrates that, in a stable system, the proposed attack $\tilde{z}_k = T_k^* z_k$ can successfully evade the Chi-square detector at certain time steps, thereby verifying the stealthiness of the proposed attack strategy.

5.2 *Unstable Process under Linear Attack*

We consider a dynamic model characterized by the following parameters

$$A = \begin{bmatrix} 1 & 0.1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0.1 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$C_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}$,
$C_2 = \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix}$,
$C_3 = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}$,
$C_4 = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}$,
$R_1 = 0.1$, $R_2 = 0.2$, $R_3 = 0.3$, $R_4 = 0.4$ and $\hat{x}_0 = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}^T$.

During the interval $[0, 50]$, the remote estimator runs the Kalman filter and reaches a steady state. The attacker employs false data, specifically $\tilde{z}_k = -Iz_k$ and $\tilde{z}_k = T_k^* z_k$, during the period $[83, 93]$ to execute a cyber attack. The simulation results for the state estimate and the $\mathrm{tr}(\tilde{P}_k)$ are shown in Figs. 7 and 8.

The purple and red curves in Fig. 7 indicate that a linear attack using either $\tilde{z}_k = -Iz_k$ or $\tilde{z}_k = T_k^* z_k$ results in a gradual deviation of the state estimate from the true system state x_k and the Kalman filter estimate $\hat{x}_k$. The red and yellow curves in Fig. 8 demonstrate that, under unstable conditions, a linear attack results in a trace of the error covariance $\tilde{P}_k$ that exceeds its value under normal system operation, resulting in exponential divergence of the error covariance. Both Figs. 7 and 8 illustrate that both attack strategies effectively disrupt system performance. Furthermore, Fig. 8 reveals that the trace of error covariance $\tilde{P}_k$ for the attack using $\tilde{z}_k = T_k^* z_k$ exceeds that observed under the attack $\tilde{z}_k = -Iz_k$ during the same time period. This observation indicates that, in certain instances, the proposed attack method in this study may lead to a more severe degradation in the performance of the unstable system.

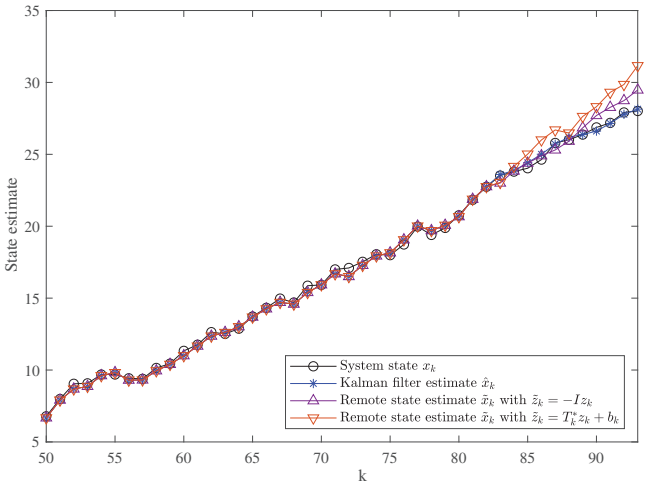

Figure 7: State estimate

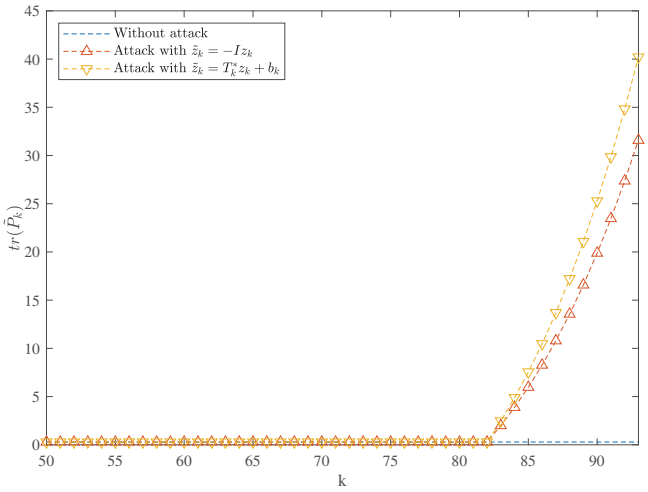

Figure 8: The trace of remote estimation error covariance

5.3 Detection of Linear Attack

We consider a dynamic model characterized by the following parameters

$$A = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}, C_1 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, C_2 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix},$$
$$Q = \begin{bmatrix} 0.01 & 0 \\ 0 & 0.01 \end{bmatrix}, R_1 = R_2 = \begin{bmatrix} 0.01 & 0 \\ 0 & 0.01 \end{bmatrix},$$
$\hat{x}_0 = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$, $L = 6$ and $M = 4$.

During the time interval $[0, 200]$, the remote estimator operates under the Kalman filtering algorithm and gradually reaches a steady-state condition. It is assumed that the attacker employs the false data $\tilde{z}_k$ to execute a linear attack on the first sensor at time steps $k = 38$, $k = 39$, and $k = 40$. Subsequently, we utilize two different algorithms for detection, each executed 100 times. The first approach involves fixing $\mathscr{J}$ in Eq. (22) to 1 and conducting direct detection. The second method employs **Algorithm 2**. A detection output of 0 indicates acceptance of H_0, signifying that no attack has occurred. Conversely, a result of 1 indicates acceptance of H_1, confirming that the system has been compromised. The simulation results for $k = 30$ and $k = 40$ are presented in Figs. 9 and 10.

It can be known from the setting of the simulation parameters that when $k = 30$, the system is not actually attacked. From Fig. 9, it can be found that in the 100 tests, the algorithm with $\mathscr{J} = 1$ considers the number of times that the system is attacked at the current moment is 1. Algorithm 2 considers the number of times to be 2. Both are within acceptable limits. This figure also illustrates the feasibility of Table 2 when the system is not attacked. Fig. 10 indicates that when the system is attacked at time 40, **Algorithm 2** detects that the system is attacked significantly more times than the algorithm that only uses $\mathscr{J} = 1$ for detection. This proves that **Algorithm 2** is more efficient in detection and reduce the missed detection rate when the system is under attack.

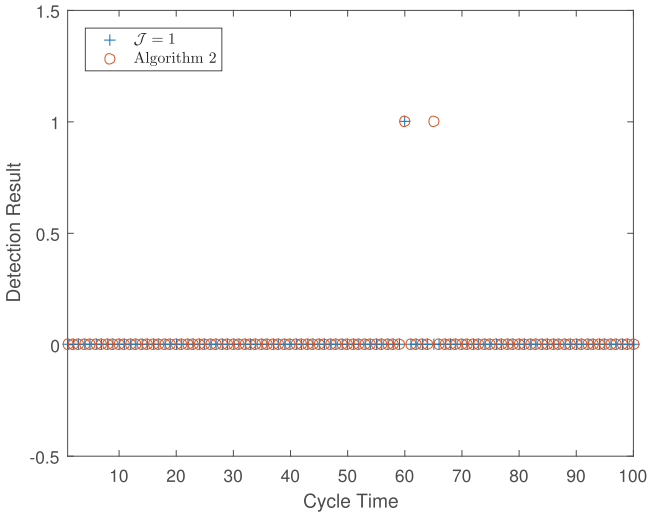

Figure 9: Comparison of the two algorithms (k = 30)

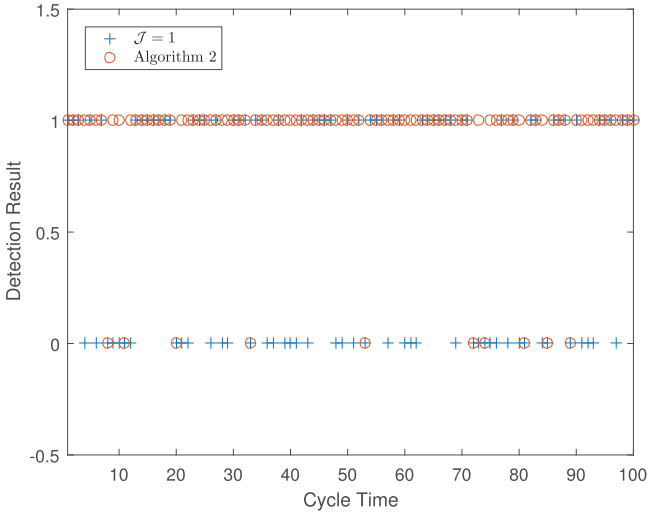

Figure 10: Comparison of the two algorithms (k = 40)

6 Conclusion

To address linear attacks, this study proposes a novel attack parameter design method with a broadened traditional feasibility constraint. Simulation comparisons demonstrate that at certain time steps, the attack strategy can successfully evade the χ^2 detector, leading to greater deviation in the state estimate of the remote estimator, which consequently results in more significant damage to system performance. The generalized feasibility constraint presented in this paper offers a more realistic foundation for modeling stealthy attacks within real-world detection systems. Furthermore, we propose a new detection algorithm. Analysis through simulation and comparison indicates that the index of the detection method increases only when the system is under attack, thereby validating the effectiveness of this algorithm in detecting the presence of an attack. Future work will focus on enhancing detection efficiency. Additionally, we will explore the system's

performance under various attack strategies and investigate new detection schemes to effectively mitigate these threats.

Acknowledgement: We are grateful to our families and friends for their unwavering understanding and encouragement.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors contributed to the study as follows: study conception and design: Fangju Zhou, Zhu Ren; simulations: Fangju Zhou, Na Ye, Jing Huang; analysis and interpretation of results: Fangju Zhou, Hanbo Zhang, Na Ye, Jing Huang, Zhu Ren; draft manuscript preparation: Fangju Zhou, Hanbo Zhang, Zhu Ren. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Data available on request from the authors.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Humayed A, Lin JQ, Li FJ, Luo B. Cyber-physical systems security—a survey. IEEE Internet Things J. 2017 Dec;4(6):1802–31. doi:10.1109/JIOT.2017.2703172.
2. Fawzi H, Tabuada P, Diggavi S. Secure estimation and control for cyber-physical systems under adversarial attacks. IEEE Trans Automat Contr. 2014 Jun;59(6):1454–67. doi:10.1109/tac.2014.2303233.
3. Zhang XM, Han QL, Ge XH, Ding L. Resilient control design based on a sampled-data model for a class of networked control systems under denial-of-service attacks. IEEE Trans Cybern. 2020 Aug;50(8):3616–26. doi:10.1109/tcyb.2019.2956137.
4. Gu CY, Zhu JW, Zhang WA, Yu L. Sensor attack detection for cyber-physical systems based on frequency domain partition. IET Control Theory Appl. 2020 Jul;14(11):1452–66. doi:10.1049/iet-cta.2019.1140.
5. Ayas MŞ. A brief review on attack design and detection strategies for networked cyber-physical systems. Turkish J Eng. 2021;5(1):1–7.
6. Hasan MK, Habib AKMA, Shukur Z, Ibrahim F, Islam S, Razzaque MA. Review on cyber-physical and cyber-security system in smart grid: standards, protocols, constraints, and recommendations. J Netw Comput Appl. 2023;209(23):103540. doi:10.1016/j.jnca.2022.103540.
7. Teixeira A, Pérez D, Sandberg H, Johansson KH, Acm. Attack models and scenarios for networked control systems. In: 1st ACM International Conference on High Confidence Networked Systems; 2012 Apr 17–19; Beijing, China. p. 55–64.
8. Ye D, Zhang TY. Summation detector for false data-injection attack in cyber-physical systems. IEEE Trans Cybern. 2020 Jun;50(6):2338–45. doi:10.1109/tcyb.2019.2915124.
9. Wei JL, Jia RZ, Song Y, Jing FW, Guo J. Binary observation-based FIR system identification under sequence denial of service attacks. Int J Robust Nonlinear Control. 2024 Mar;34(5):3442–63. doi:10.1002/rnc.7146.
10. Zhao N, Shi P, Xing W, Lim CP. Event-triggered control for networked systems under denial of service attacks and applications. IEEE Trans Circuits Syst I: Regular Papers. 2022 Feb;69(2):811–20. doi:10.1109/tcsi.2021.3116278.
11. Mo YL, Sinopoli B. Secure control against replay attacks. In: Proceeding of the 47th Annual Allerton Conference on Communication, Control, and Computing; 2009 Sep; Monticello, IL, USA.
12. Li TX, Wang ZD, Zou L, Chen B, Yu L. A dynamic encryption-decryption scheme for replay attack detection in cyber-physical systems. Automatica. 2023;151(1):110926. doi:10.1016/j.automatica.2023.110926.
13. Naha A, Teixeira A, Ahlén A, Dey S. Sequential detection of replay attacks. IEEE Trans Automat Contr. 2023 Mar;68(3):1941–8. doi:10.1109/tac.2022.3174004.

14. Ni YQ, Guo ZY, Mo YL, Shi L. On the performance analysis of reset attack in cyber-physical systems. IEEE Trans Automat Contr. 2020 Jan;65(1):419–25. doi:10.1109/tac.2019.2914655.
15. Ding KM, Ren XQ, Leong AS, Quevedo DE, Shi L. Remote state estimation in the presence of an active eavesdropper. IEEE Trans Automat Contr. 2021 Jan ;66(1):229–44. doi:10.1109/tac.2020.2980730.
16. Lu AY, Yang GH. False data injection attacks against state estimation without knowledge of estimators. IEEE Trans Automat Contr. 2022 Sep;67(9):4529–40. doi:10.1109/tac.2022.3161259.
17. Xu QL, Xiong JL. Scheduler-pointed false data injection attack for event-based remote state estimation. Automatica. 2024;162(10):111523. doi:10.1016/j.automatica.2024.111523.
18. Li L, Yang H, Xia YQ, Yang HJ. State estimation for linear systems with unknown input and random false data injection attack. IET Control Theory Appl. 2019 Apr;13(6):823–31. doi:10.1049/iet-cta.2018.5954.
19. Alzubi OA. A deep learning-based frechet and dirichlet model for intrusion detection in IWSN. J Intell Fuzzy Syst. 2022;42(2):873–83. doi:10.3233/JIFS-189756.
20. Alzubi OA, Qiqieh I, Alzubi JA. Fusion of deep learning based cyberattack detection and classification model for intelligent systems. Cluster Comput. 2023;26(2):1363–74. doi:10.1007/s10586-022-03686-0.
21. Guo Z, Shi D, Johansson KH, Shi L. Worst-case stealthy innovation-based linear attack on remote state estimation. Automatica. 2018 Mar;89(1):117–24. doi:10.1016/j.automatica.2017.11.018.
22. Liu H, Ni Y, Xie L, Johansson KH. How vulnerable is innovation-based remote state estimation: fundamental limits under linear attacks. Automatica. 2022 Feb;136(12):110079. doi:10.1016/j.automatica.2021.110079.
23. Li Y, Shi L, Chen T. Detection against linear deception attacks on multi-sensor remote state estimation. IEEE Trans Control Netw Syst. Jan 2017;5(3):846–56. doi:10.1109/TCNS.2017.2648508.
24. Li Y, Yang Y, Zhao Z, Zhou J, Quevedo DE. Deception attacks on remote estimation with disclosure and disruption resources. IEEE Trans Automat Contr. 2023 Jul;68(7):4096–112. doi:10.1109/tac.2022.3202981.
25. Gupta A, Sikdar A, Chattopadhyay A, IEEE. Quickest detection of false data injection attack in remote state estimation. In: IEEE International Symposium on Information Theory (ISIT); 2021 Jul 12–20; Melbourne, VIC, Australia. p. 3068–73.
26. Li YZ, Yang YK, Chai TY, Chen TW. Stochastic detection against deception attacks in CPS: performance evaluation and game-theoretic analysis. Automatica. 2022 Oct;144(1):110461. doi:10.1016/j.automatica.2022.110461.
27. Guo ZY, Shi DW, Johansson KH, Shi L. Optimal linear cyber-attack on remote state estimation. IEEE Trans Control Netw Syst. 2017 Mar;4(1):4–13. doi:10.1109/tcns.2016.2570003.
28. Guo ZY, Shi DW, Quevedo DE, Shi L. Secure state estimation against integrity attacks: a gaussian mixture model approach. IEEE Trans Signal Process. 2019 Jan;67(1):194–207. doi:10.1109/tsp.2018.2879037.
29. Liu HX, Ni YQ, Xie LH, Johansson KH. An optimal linear attack strategy on remote state estimation. IFAC-PapersOnLine. 2020;53(2):3527–32. doi:10.1016/j.ifacol.2020.12.1719.
30. Pandey R, Koranga M, Thakur SN, Khan H, Singh SK, Ravikumar RN. Securing vehicle-to-grid communications: a cyber-physical approach. In: Optimized energy management strategies for electric vehicles. Hershey, PA, USA: IGI Global Scientific Publishing; 2025. p. 301–18.