Vulnerability Analysis of Distributed State Estimator Under False Data Injection Attacks

Pengyu Li and Dan Ye, *Senior Member, IEEE*

***Abstract*— This paper focuses on the vulnerability and strict vulnerability of distributed state estimators under false data injection (FDI) attacks, where adversaries aim to exert unbounded effects on the estimation error dynamics by injecting malicious data into sensor nodes, communication links, or both. In particular, a distributed system is characterized as vulnerable (or strictly vulnerable) if there exists an unbounded FDI attack that leads to bounded changes (or no changes) in residuals. By utilizing invertibility theory and carefully designed attack sequences, we establish the conditions for systems to be (strictly) vulnerable, based on different attack scenarios. Additionally, we provide a comparative analysis to evaluate the varying impact of different attacks on system security. Finally, a three-area grid system model is presented to illustrate the validity of the theoretical results.**

***Index Terms*— Cyber-physical systems, distributed state estimation, vulnerability, false data injection attacks.**

I. Introduction

SINCE the high integration of computing, communication, and storage functions, cyber-physical systems (CPSs) enable real-time sensing, information fusion, and dynamic control of distributed state estimators [1], [2], [3], [4], [5]. Although the openness of distributed networks facilitates real-time data interaction between different sensor nodes, it also provides opportunities for attackers to intrude on control centers and corrupt data. In this scenario, identifying and profiling the vulnerabilities of distributed networks is crucial for defenders, which is conducive to clarifying which communication channels or sensors are prioritized for protection.

In the existing literatures, precisely designed attacks can be broadly categorized into two main types: false data injection (FDI) attacks and denial-of-service (DoS) attacks [6]. In general, DoS attackers disrupt data availability by deliberately allocating attack resources [7], [8]. They exploit vulnerabilities in transport protocols to deny targeted users access to databases. In contrast to DoS attacks, FDI attackers aim to compromise the integrity of data to undermine the normal operation of CPSs [9], [10]. FDI attacks, by eliminating the impact on detection residuals, are more covert and destructive in nature [11], [12]. As a result, the investigations of FDI attack behaviors that lead to system insecurity have become a major concern.

For centralized CPSs, FDI attacks destabilize systems by corrupting sensors, actuators, or both [13], [14]. Specifically, an FDI attack is said to be stealthy (or strictly stealthy) when its impact on residuals is bounded (or zero). It's important to note that strictly stealth attacks are theoretically invisible to almost all anomaly detectors [15]. In addition to stealthiness, another critical aspect of CPS security is the effectiveness of attacks. In [16], a class of unbounded FDI attacks with complete stealthiness was studied, where attackers could exert unbounded effects on CPSs. To characterize the coupling between the stealthiness and effectiveness of an attack, the concepts of vulnerability and strict vulnerability were introduced in [17]. A system is regarded as vulnerable (or strictly vulnerable) if a stealthy (or strictly stealthy) FDI attack can destabilize it. Along this line, operators can more easily discover flaws in defense mechanisms.

In addition to centralized systems, many researchers have also explored the vulnerability of distributed CPSs under FDI attacks. To assess the overall vulnerability of state omniscience, [18] introduced a decentralized FDI attack model aimed at compromising distributed observers. In the case of linearized static and dynamic CPSs, corresponding vulnerabilities were analyzed across four attack scenarios [19]. While these attacks can be relatively straightforward to implement (simply corrupting measurements or estimators in sensor nodes), the resulting vulnerability conditions also impose stringent constraints on system structures. Subsequent work in [20] and [21] revealed the vulnerability of individual nodes, where attackers can freely tamper with the transmitted edge information. It's important to note that the investigations mentioned above primarily focus on single-FDI attack scenarios, where adversaries attack either sensor nodes or edges.

However, in a real-world communication network, every component and communication link in CPSs may potentially be compromised [14]. Cunning attackers can not only hijack

Manuscript received 25 February 2023; revised 14 November 2023; accepted 26 April 2024. Date of publication 6 May 2024; date of current version 10 May 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62173071 and Grant U22A2067, in part by Liaoning Distinguished Young Funds under Grant 2023020359-JH6/1005, and in part by the Fundamental Research Funds for the Central Universities under Grant N2204008. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Alexey Vinel. *(Corresponding author: Dan Ye.)*

Pengyu Li is with the College of Information Science and Engineering, Northeastern University, Shenyang 110819, China (e-mail: 1337513823@qq.com).

Dan Ye is with the College of Information Science and Engineering and the State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110819, China (e-mail: yedan@ise.neu.edu.cn).

Digital Object Identifier 10.1109/TIFS.2024.3396634

communication channels but also directly breach control centers. In such scenarios, a distributed system may exhibit robustness to single-FDI attacks but vulnerability to joint-FDI attacks. Consequently, an urgent task is to analyze and compare the impact of different attacks on system security.

Motivated by the aforementioned works, this paper focuses on the vulnerability and strict vulnerability of distributed state estimators under single-FDI attacks and joint-FDI attacks. The primary contributions can be summarized as follows:

(1) By employing invertibility theory and the careful construction of attack sequences, we provide necessary and sufficient conditions for identifying (strictly) vulnerable systems across different scenarios. Based on the derived results, defenders have an explicit insight into which edges or sensor nodes are more critical to security.

(2) We systematically compare vulnerability and strict vulnerability conditions under different attack scenarios. Specifically, state estimates are more critical than measurement data. In addition, when only state estimates are under attack, it is more cost-effective to protect state estimates within nodes instead of those within edges.

(3) In contrast to the existing single-FDI attacks [19], [20], [21], the proposed joint-FDI attacks have weaker requirements for system structures due to the relaxed implement conditions. In this sense, single-FDI attack sacrifices stealthiness to enhance the attack effectiveness, while joint-FDI attacks disrupts more components of CPSs to achieve the attack objective.

The rest of this paper is organized as follows. Section II presents problem formulation and preliminaries, including the system model, the attack model, the definition of vulnerabilities and strict vulnerabilities. In Section III, the sufficiency and necessity for (strictly) vulnerable systems are investigated. Section IV provides an example to demonstrate the validity of the theoretical results. Finally, conclusions are given in Section V.

Notations: $\mathbb{R}^n$ and $\mathbb{R}^{n\times m}$ denote the n-dimensional Euclidean space and the set of $n \times m$ real matrices, respectively. The symbol $\otimes$ denotes the Kronecker product. For a matrix X, span(X) represents its column spanning space and Ker(X) denotes its kernel space, respectively. For a vector x, $\|x\|$ stands for its Euclidean norm.

II. Problem Formulation and Preliminaries

In this section, we first present the system model and the attack model, and then introduce the definitions of vulnerability and strict vulnerability.

A. System Model

Consider a discrete-time linear time-invariant (LTI) system including s sensors:

$$\begin{aligned} x_{k+1} &= Ax_k + w_k, \\ y_{i,k} &= C_i x_k + v_{i,k}, \end{aligned} \tag{1}$$

where $x_k \in \mathbb{R}^n$ denotes the physical plant state, $y_{i,k} \in \mathbb{R}^m$ denotes the measurement of the sensor i, $i \in \{1, \cdots, s\}$. In addition, $w_k \sim \mathcal{N}(0, Q)$ and $v_{i,k} \sim \mathcal{N}(0, R_i)$ are the uncorrelated process noise and measurement noise, respectively.

In particular, let a directed graph $\mathcal{G} = (\mathcal{V}, \xi)$ characterize the communication topology of the sensor network. $\mathcal{V} = \{1, \cdots, s\}$ and $\xi \subseteq \mathcal{V} \times \mathcal{V}$ denote the sets of sensor nodes and the corresponding communication links, respectively. Define $N_i = \{j : (i, j) \in \xi\}$ be the neighbors of the ith sensor with the cardinality $|N_i| = l_i$. In addition, let $\xi_{i,j} = 1$ if node i is directly connected to node j, otherwise $\xi_{ij} = 0$. Accordingly, two assumptions are given below.

Assumption 1: The matrix pairs $(A, \sqrt{Q})$ are controllable.

Assumption 2: The graph $\mathcal{G}$ is strongly connected.

Based on the above assumptions, a distributed state estimator in [21] is adopted to estimate the state of node i, it follows

$$\hat{x}_{i,k+1} = A\hat{x}_{i,k} + K_i(y_{i,k} - C_i\hat{x}_{i,k}) - \rho A \sum_{j\in N_i}(\hat{x}_{i,k} - \hat{x}_{j,k}), \tag{2}$$

where $\hat{x}_{i,k}$ and $\hat{x}_{j,k}$ are the estimates of the ith node and the jth node, respectively. K_i is the steady-state estimator gain, and $\rho \in (0, \min\{l_i^{-1}\})$ is the consensus coefficient. In addition, the residue $z_{i,k}$ can be computed by $z_{i,k} = y_{i,k} - C_i\hat{x}_{i,k}$.

B. Attack Model

Based on where the attack is launched, we introduce single-FDI attacks and joint-FDI attacks in this section. To characterize the worst case, assume that adversaries can adopt all the system knowledge to design their attack strategies. In particular, some symbols will be abused to improve the legibility of the article.

In the following, we first present three single-FDI attacks, where attackers can compromise either the measurement data of the target node (Type-II attack), the state estimates of the target node (Type-III attack), or the state estimates of the neighboring nodes transmitted over the communication edges (Type-I attack).

1) **Type-I attack**: The malicious data is injected into edge (i, j) to tamper with $\hat{x}_{j,k}$, and the compromised estimator and residue are updated by

$$\begin{aligned} \hat{x}'_{i,k+1} &= A\hat{x}'_{i,k} + K_i(y_{i,k} - C_i\hat{x}'_{i,k}) - \rho A \sum_{j\in N_i}(\hat{x}'_{i,k} - \hat{x}'^{a}_{j,k}) \\ z'_{i,k} &= y_{i,k} - C_i\hat{x}'_{i,k}, \end{aligned} \tag{3}$$

where $\hat{x}'^{a}_{j,k} = \hat{x}'_{j,k} + \Xi_{ij}\alpha_{ij,k}$ and $\Xi_{ij}\alpha_{ij,k}$ is the injected attack signal. In particular, Ξ_{ij} is designed to be full column rank.

2) **Type-II attack**: The malicious data is injected into node i to tamper with $y_{i,k}$, and this implies that

$$\begin{aligned} \hat{x}'_{i,k+1} &= A\hat{x}'_{i,k} + K_i(y_{i,k} - C_i\hat{x}'_{i,k}) - \rho A \sum_{j\in N_i}(\hat{x}'_{i,k} - \hat{x}'_{j,k}) \\ &\quad + K_i\Xi_i\alpha^a_k, \\ z'_{i,k} &= y_{i,k} - C_i\hat{x}'_{i,k} + \Xi_i\alpha^a_k, \end{aligned} \tag{4}$$

where $\Xi_i\alpha^a_k$ is the injected attack signal and Ξ_i is designed to be full column rank.

3) **Type-III attack**: The malicious data is injected into node i to tamper with $\hat{x}_{i,k}$, then

$$\hat{x}'_{i,k+1} = A\hat{x}'_{i,k} + K_i(y_{i,k} - C_i\hat{x}'_{i,k}) - \rho A \sum_{j\in N_i}(\hat{x}'_{i,k} - \hat{x}'_{j,k})$$

$$
\begin{aligned}
&\quad + (A - K_iC_i - \rho A l_i)\Xi_i\alpha_{i,k} + \rho A \sum_{j\in N_i} \Xi_j\alpha_{j,k},\\
z'_{i,k} &= y_{i,k} - C_i\hat{x}'_{i,k} - C_i\Xi_i\alpha_{i,k}, \qquad (5)
\end{aligned}
$$

where $\Xi_i\alpha_{i,k}$ and $\Xi_j\alpha_{j,k}$ are the injected attack signals, Ξ_i and Ξ_j are designed to be full column rank.

Compared to single-FDI attacks, joint-FDI attacks enable attackers to simultaneously compromise both the target node and its communication links. Specifically, adversaries can compromise the state estimates (Type-IV attack) or measurement data (Type-V attack) of the target node while simultaneously modifying the state estimates of the neighboring nodes transmitted over the communication edges.

4) **Type-IV attack**: The malicious data is injected into node i and edge (i, j) to tamper with $\hat{x}_{i,k}$ and $\hat{x}_{j,k}$, leading to

$$
\begin{aligned}
\hat{x}'_{i,k+1} &= A\hat{x}'_{i,k} + K_i(y_{i,k} - C_i\hat{x}'_{i,k}) - \rho A \sum_{j\in N_i}(\hat{x}'_{i,k} - \hat{x}'_{j,k})\\
&\quad + (A - K_iC_i - \rho A l_i)\Xi_i\alpha_{i,k} + \rho A \sum_{j\in N_i} \Xi_j\alpha_{j,k}\\
&\quad + \rho A \sum_{j\in N_i} \theta_{ij}\beta_{ij,k},\\
z'_{i,k} &= y_{i,k} - C_i\hat{x}'_{i,k} - C_i\Xi_i\alpha_{i,k}, \qquad (6)
\end{aligned}
$$

where $\Xi_i\alpha_{i,k}$, $\Xi_j\alpha_{j,k}$, and $\theta_{ij}\beta_{ij,k}$ are the injected attack signals, Ξ_i, Ξ_j, and θ_{ij} are designed to be full column rank.

5) **Type-V attack**: The malicious data is injected into node i and edge (i, j) to tamper with $y_{i,k}$ and $\hat{x}_{j,k}$, and it can be verified that

$$
\begin{aligned}
\hat{x}'_{i,k+1} &= A\hat{x}'_{i,k} + K_i(y_{i,k} - C_i\hat{x}'_{i,k}) - \rho A \sum_{j\in N_i}(\hat{x}'_{i,k} - \hat{x}'_{j,k})\\
&\quad + \rho A \sum_{j\in N_i} \theta_{ij}\beta_{ij,k} + K_i\Xi_i\alpha_{i,k},\\
z'_{i,k} &= y_{i,k} - C_i\hat{x}'_{i,k} + \Xi_i\alpha_{i,k}, \qquad (7)
\end{aligned}
$$

where $\Xi_i\alpha_{i,k}$ and $\theta_{ij}\beta_{ij,k}$ are the injected attack signals, Ξ_i and θ_{ij} are designed to be full column rank.

In this paper, the influences caused by attacks are characterized by

$$
\Delta\hat{x}_{i,k+1} \triangleq \hat{x}'_{i,k+1} - \hat{x}_{i,k+1}, \qquad \Delta z_{i,k} \triangleq z'_{i,k} - z_{i,k}, \qquad (8)
$$

in which $\Delta\hat{x}_{i,k+1}$ is generally adopted to quantify the effectiveness of attacks, while $\Delta z_{i,k}$ is employed to reflect the stealthiness of attacks.

To clarify which sensor nodes and communication links are more critical for the security of distributed CPSs, we investigate three single-FDI attacks and two joint-FDI attacks. In contrast to centralized systems, the mathematical complexity arising from the strong coupling in distributed systems presents difficulties and challenges in problem analysis.

C. Vulnerability and Strict Vulnerability

In this section, we utilize vulnerability and strict vulnerability to characterize the security of distributed state estimation under attacks. Inspired by [16], the definitions for vulnerability and strict vulnerability are presented below.

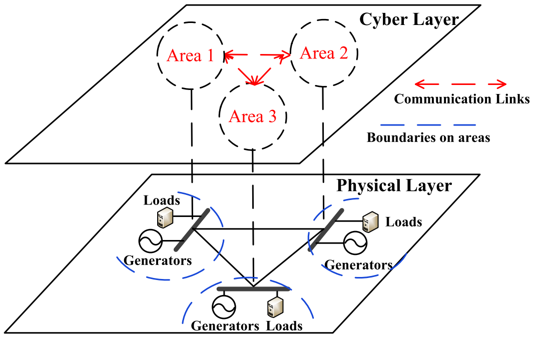

Fig. 1. Cyber-physical structure of a three-area power system.

Definition 1: System (1) equipped with the distributed state estimator (2) is vulnerable if there exists an attack sequence that satisfies the following two conditions:

(1) The state estimation difference $\Delta\hat{x}_k$ satisfies

$$
\|\Delta\hat{x}_k\| > \sigma, \qquad \forall\sigma \in \mathbb{R}. \qquad (9)
$$

(2) The residual difference Δz_k satisfies

$$
\|\Delta z_k\| < \delta, \qquad \exists\delta > 0,\ \delta \in \mathbb{R}, \qquad (10)
$$

where $\Delta\hat{x}_k = [(\Delta\hat{x}_{1,k})^{\mathrm{T}}, \cdots, (\Delta\hat{x}_{s,k})^{\mathrm{T}}]^{\mathrm{T}}$ and $\Delta z_k = [(\Delta z_{1,k})^{\mathrm{T}}, \cdots, (\Delta z_{s,k})^{\mathrm{T}}]^{\mathrm{T}}$.

Definition 2: System (1) equipped with the distributed state estimator (2) is strictly vulnerable if there exists an attack sequence that satisfies the following two conditions:

(1) The state estimation difference $\Delta\hat{x}_k$ satisfies

$$
\|\Delta\hat{x}_k\| > \sigma, \qquad \forall\sigma \in \mathbb{R}. \qquad (11)
$$

(2) The residual difference Δz_k satisfies

$$
\|\Delta z_k\| = 0. \qquad (12)
$$

From Definitions 1 and 2, a distributed system is characterized as vulnerable (or strictly vulnerable) if there exists an unbounded FDI attack that leads to bounded changes (or no changes) in residuals. By examining vulnerability and strict vulnerability, defenders can more readily pinpoint weaknesses in distributed CPSs.

Remark 1: At present, the application of distributed CPSs is widespread, covering areas such as smart grids, transportation systems, smart manufacturing and industrial automation, etc. Motivated by [22], a typical three-area power system model is introduced in Fig. 1.

The three-area power system is composed of three balancing control areas interconnected by tie lines [23]. As one of the critical closed-loop control systems in smart grids, load frequency control can automatically adjust the generation reference value according to the load changes, so as to maintain the frequency and the power of the connecting lines in each control area at a predetermined value. As depicted in Fig. 1, three areas in the physical layer are designated for load frequency control, and measurement signals are transmitted from remote telemetry units to the area controller via wireless networks. The detailed structure of the physical layer can be seen in Fig. 2, which depicts the IEEE 118-buses grid system [24].

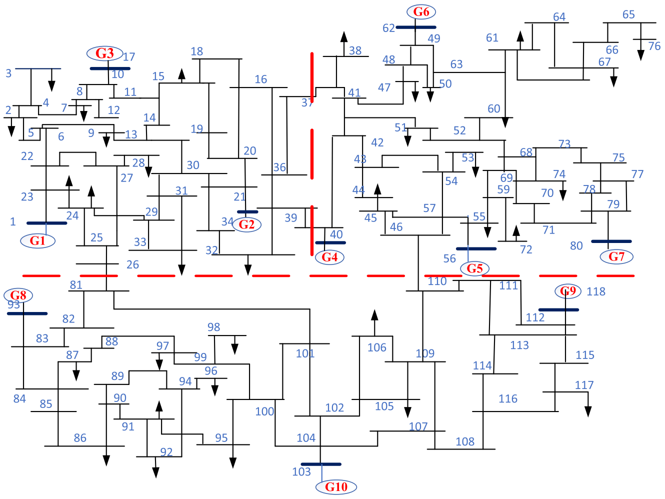

Fig. 2. IEEE 118-buses grid system.

From Fig. 2, the model including s buses can be represented by the graph $\mathcal{G} = (\mathcal{V}, \xi)$, where $\mathcal{V} = \{1, \cdots, s\}$ and $s = 10$. Let each node in the graph represent a synchronous generator and the edges represent connecting lines. For simplicity, assume that these connecting lines correspond to generator buses at both ends, i.e., s buses in the grid are connected to synchronous generators. Therefore, the physical dynamics of generators can be represented by

$$\begin{cases} \dot{\delta}_i(t) = w_i(t), \\ M_i \dot{w}_i(t) = P_{mi}(t) - P_{i,j}(t) - D_i w_i(t), \end{cases}$$

where $i = 1, 2, \cdots, s$. In addition, $w_i(t)$ is the angular frequency of the i-th generator, D_i is the damping coefficient of the i-th generator's angular frequency, M_i is the inertia coefficient of the i-th generator, $\dot{\delta}_i(t)$ is the rotor angle of the i-th generator, $P_{mi}(t)$ is the mechanical power input of the i-th generator, and $P_{i,j}(t)$ is the active power generated by the i-th generator.

For a lossless power system, $P_{i,j}(t)$ satisfies

$$P_{i,j}(t) = \sum_{j \in N_i} k_{ij} \sin[\delta_i(t) - \delta_j(t)].$$

Assuming that the relative differences in rotor angles are sufficiently small, the above model can be linearized as follows:

$$M_i \ddot{\delta}_i(t) + D_i \delta_i(t) = P_{mi}(t) + w_i(t) - \sum_{j \in N_i} k_{ij} [\delta_i(t) - \delta_i(t)].$$

In this paper, we are only concerned with the internal physical state changes of generators, thus the mechanical input power is assumed to be constant, i.e., $P_{mi}(t) = 0$. For an electrical grid system consisting of N grid sub-areas, the dynamic model for the i-th grid sub-area can be expressed by

$$\begin{cases} \dot{x}^{(I)}(t) = A^{(I)} x^{(I)}(t) + w^{(I)}(t), \\ y^{(I)}(t) = C^{(I)} x^{(I)}(t) + v^{(I)}(t), \quad I = 1, 2, \cdots, N, \end{cases}$$

where $\dot{x}^{(I)}(t) = [\delta_1(t), \cdots, \delta_s(t), \dot{\delta}_1(t), \cdots, \dot{\delta}_s(t)]$ denotes the state and the corresponding first-order derivative of generators. $A^{(I)} = \begin{bmatrix} 0_N & I_N^{(I)} \\ -M^{(I)} \Gamma_{\mathcal{G}}^{(I)} & -M^{(I)} D^{(I)} \end{bmatrix}$, $D^{(I)} = \mathrm{diag}(d_1, \cdots, d_s)$, $M^{(I)} = \mathrm{diag}(1/m_1, \cdots, 1/m_s)$, $\Gamma_{\mathcal{G}}$ denote the Laplacian matrix of graph $\mathcal{G}$, $C^{(I)}$ is the observation matrix with proper dimension, $w^{(I)}(t)$ and $v^{(I)}(t)$ are the uncorrelated process noise and measurement noise.

In practice, the continuous variable $\dot{x}^{(I)}(t)$ can be discretized with a small time period Δt [25]:

$$\dot{x}_{t+1}^{(I)} = A_d^{(I)} x_t^{(I)} + w_t^{(I)},$$

where $A_d^{(I)} = e^{A^{(I)} \Delta t}$.

In summary, a nonlinear continuous system model can be linearised and discretised into an LTI system, thereby simplifying the analysis process.

Remark 2: In real-world distributed systems, attackers not only hijack communication channels but also directly invade control centers. In this scenario, a distributed system may exhibit robustness to single-FDI attacks but vulnerability to joint-FDI attacks. Compared with the single-FDI attack scenario [14], the joint-FDI attack is a more common attack behavior that may pose a serious threat to the operation of systems. However, despite its clear engineering insight, the security problem under joint-FDI attacks has thus far been largely overlooked. This may be due to the mathematical complexities induced by the strong coupling of joint attacks, and it serves as one of the motivations for our current investigation. In the subsequent sections, we analyze and compare the effects of single-FDI attacks and joint-FDI attacks on system security by providing necessary and sufficient conditions.

III. Vulnerability and Strict Vulnerability for Different Attack Cases

In this section, we investigate the necessary and sufficient conditions of vulnerability and strict vulnerability under different attack scenarios. For convenience, the following lemma is acquired.

Lemma 1: [26] Considering an LTI system with initial value $x_0 = 0$:

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k), \\ y(k) &= Cx(k) + Du(k), \end{aligned} \tag{13}$$

where $x(k) \in \mathbb{R}^t$ is the state, $u(k) \in \mathbb{R}^r$ is the input, and $y(k) \in \mathbb{R}^m$ is the output.

System (13) is defined as *invertible* if $y(k) = 0$ means $u(k) = 0$, $\forall k \in \mathbb{N}$. In particular, system (13) is *invertible* if and only if

$$\mathrm{rank}(M_t) - \mathrm{rank}(M_{t-1}) = r, \tag{14}$$

in which M_t is calculated by

$$M_t = \begin{pmatrix} D & 0 & 0 & \cdots & 0 \\ CB & D & 0 & \cdots & 0 \\ CAB & CB & D & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ CA^{t-1}B & CA^{t-2}B & CA^{t-3}B & \cdots & D \end{pmatrix}.$$

We now illustrate the feasibility conditions for vulnerability and strict vulnerability with respect to single-FDI attacks and joint-FDI attacks. The results are summarized in the following theorems and corollaries.

1) **Type-I attack**: From (3), the dynamics of $\Delta\hat{x}_k$ and Δz_k can be calculated by

$$\Delta\hat{x}_{k+1} = [H + \rho(\xi\otimes A)]\Delta\hat{x}_k + \rho(\xi\otimes A)\Xi^a\alpha_k^a, \tag{15a}$$

$$\Delta z_k = -C\Delta\hat{x}_k,$$

$$H = \begin{bmatrix} A - K_1C_1 - \rho A l_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A - K_sC_s - \rho A l_s \end{bmatrix},$$

$$\xi = \begin{bmatrix} \xi_{11} & \cdots & \xi_{1s} \\ \vdots & \ddots & \vdots \\ \xi_{s1} & \cdots & \xi_{ss} \end{bmatrix}, \quad \Xi^a = \begin{bmatrix} \Xi_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Xi_s \end{bmatrix},$$

$$C = \begin{bmatrix} C_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & C_s \end{bmatrix}, \quad \alpha_k^a = (\alpha_{1,k}^{\mathrm{T}}, \cdots, \alpha_{s,k}^{\mathrm{T}})^{\mathrm{T}}, \tag{15b}$$

where $\xi_{ij} = 1$ if node i is directly connected to node j, otherwise $\xi_{ij} = 0$. In addition, $\Xi_i\alpha_{i,k} = \sum_{j\in N_i}\Xi_{ij}\alpha_{ij,k}$, and $\Xi_i = I$ if $\sum_{j\in N_i}\Xi_{ij}\alpha_{ij,k} \neq 0$. In particular, the initial state $\Delta\hat{x}_0 = \hat{x}_0' - \hat{x}_0 = 0$.

Theorem 1: System (1) is vulnerable under Type-I attack if and only if there exist $Q \in \mathbb{R}^{ns\times ns}$ and $v \in \mathbb{R}^{ns}$ such that:

(1) $v \in \mathrm{Ker}(C)$;

(2) v is an unstable eigenvector of $H+\rho(\xi\otimes A)+\rho(\xi\otimes A)\Xi^aQ$, i.e., $[H + \rho(\xi\otimes A) + \rho(\xi\otimes A)\Xi^aQ]v = \lambda v$ with $|\lambda| \geq 1$;

(3) v is reachable for $(H + \rho(\xi\otimes A), \rho(\xi\otimes A)\Xi^a)$.

Proof: See Appendix A. ■

Theorem 2: System (1) under Type-I attack is strictly vulnerable if and only if

(1) the dynamic differences of estimators, i.e., system (15), is *non-invertible*;

(2) there exists a vector $v \in \mathbb{R}^{ns}$ satisfying $\rho(\xi\otimes A)\Xi^av \neq 0$.

Proof: See Appendix B. ■

Remark 3: In Theorem 2, we investigate necessary and sufficient conditions for strict vulnerability. In contrast to centralized systems, the designed strict vulnerability conditions become more stringent due to the condition $\rho(\xi\otimes A)\Xi^av \neq 0$. This is a consequence of the strong coupling inherent in distributed systems.

Theorems 1 and 2 provide evaluation criteria for the vulnerability and strict vulnerability of distributed CPSs under Type-I attack. In other words, if any of the necessary and sufficient conditions (as stated in Theorems 1 and 2) are not met, the distributed CPSs are resilient to such attacks. By employing a similar derivation result, we can obtain the following corollaries. For the sake of readability, the proof process has been omitted.

2) **Type-II attack**: From (4), the dynamics of $\Delta\hat{x}_k$ and Δz_k can be calculated by

$$\Delta\hat{x}_{k+1} = [H + \rho(\xi\otimes A)]\Delta\hat{x}_k + K\Xi^a\alpha_k^a,$$

$$\Delta z_k = -C\Delta\hat{x}_k + \Xi^a\alpha_k^a, \tag{16}$$

where $K = \mathrm{diag}\{K_1, \cdots, K_s\}$, $\Xi^a = (\Xi_1^{\mathrm{T}}, \cdots, \Xi_s^{\mathrm{T}})^{\mathrm{T}}$, and Ξ_i is full column rank. In particular, $\Xi_i = 0$ if node i is not attacked.

Corollary 1: System (1) is vulnerable under Type-II attack if and only if there exists $v \in \mathbb{R}^{ns}$ such that:

(1) $Cv \in \mathrm{span}(\Xi^a)$;

(2) v is an unstable eigenvector of $H + \rho(\xi\otimes A) + KC$, i.e., $[H + \rho(\xi\otimes A) + KC]v = \lambda v$ with $|\lambda| \geq 1$;

(3) v is reachable for $(H + \rho(\xi\otimes A), K\Xi^a)$.

Theorem 3: System (1) under Type-II attack is not strictly vulnerable.

Proof: Note that (16) is *invertible*, demonstrating that zero output is uniquely determined by zero input. In other words, $\Delta z_k = 0$ if and only if $\alpha_k^a = 0$ for all $k \in \mathbb{N}$. In this scenario, any non-zero attack sequence α_k^a will result in $\Delta z_k \neq 0$, thereby violating the definition of strict vulnerability. ■

Corollary 1 and Theorem 3 indicate that exclusively manipulating the measurement data within the target node will consistently produce a bounded effect on residuals. In simpler terms, this form of attack is more easily detectable by defenders.

Remark 4: Based on the coupling between multiple different nodes, Corollary 1 illustrates the vulnerability of distributed CPSs under Type-II attack. In this sense, several existing works [21], [26] that primarily focus on the vulnerability of individual nodes can be seen as a special case of this paper.

3) **Type-III attack**: From (5), the dynamics of $\Delta\hat{x}_k$ and Δz_k can be calculated by

$$\Delta\hat{x}_{k+1} = [H + \rho(\xi\otimes A)](\Delta\hat{x}_k + \Xi^a\alpha_k^a),$$

$$\Delta z_k = -C(\Delta\hat{x}_k + \Xi^a\alpha_k^a), \tag{17}$$

where

$$\Xi^a = \begin{bmatrix} \Xi_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Xi_s \end{bmatrix}, \quad \alpha_k^a = \left(\alpha_{1,k}^{\mathrm{T}}, \ \cdots, \ \alpha_{s,k}^{\mathrm{T}}\right)^{\mathrm{T}},$$

and Ξ_i is full column rank. In particular, $\Xi_i = 0$ if node i is not compromised.

Theorem 4: System (1) under Type-III attack is vulnerable if and only if there exists $v \in \mathbb{R}^{ns}$ such that $Cv = 0$ and $[H + \rho(\xi\otimes A)]v \neq 0$.

Proof: See Appendix C. ■

Corollary 2: System (1) under Type-III attack is strictly vulnerable if and only if

(1) the dynamic differences of estimators, i.e., system (17), is *non-invertible*;

(2) there exists a vector $v \in \mathbb{R}^{ns}$ satisfying $[H + \rho(\xi\otimes A)]\Xi^av \neq 0$.

Remark 5: When only measurement data is tampered with (Type-II attack), the corresponding system is not strictly vulnerable. In this sense, state estimates are more deserving of protection than measurement data. However, the implement conditions of Type-I attack and Type-III attack depend on a crucial requirement: the measurement matrix C is full column rank.

In addition, compared with Type-III attack, the implementation conditions of Type-I attack are more strict. In some specific cases, directly attacking sensor nodes instead of communication links is a more cost-effective approach, as demonstrated in the Simulation Examples.

4) **Type-IV attack**: From (6), the dynamics of $\Delta\hat{x}_k$ and Δz_k can be calculated by

$$\begin{aligned}\Delta\hat{x}_{k+1} &= [H+\rho(\xi\otimes A)]\Delta\hat{x}_k\\ &\quad + [(H+\rho(\xi\otimes A))\Xi^a \quad \rho(\xi\otimes A)\theta^a]\begin{bmatrix}\alpha_k^a\\ \beta_k^a\end{bmatrix},\\ \Delta z_k &= -C\Delta\hat{x}_k - [C\Xi^a \quad 0]\begin{bmatrix}\alpha_k^a\\ \beta_k^a\end{bmatrix},\end{aligned} \tag{18}$$

where

$$\Xi^a = \begin{bmatrix}\Xi_1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \Xi_s\end{bmatrix}, \quad \theta^a = \begin{bmatrix}\theta_1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \theta_s\end{bmatrix},$$

$$\alpha_k^a = \left(\alpha_{1,k}^{\mathrm{T}}, \ \cdots, \ \alpha_{s,k}^{\mathrm{T}}\right)^{\mathrm{T}}, \ \beta_k^a = \left(\beta_{1,k}^{\mathrm{T}}, \ \cdots, \ \beta_{s,k}^{\mathrm{T}}\right)^{\mathrm{T}},$$

and $\theta_{i,k}\beta_{i,k} = \sum_{j\in N_i}\theta_{ij}\beta_{ij,k}$. In particular, $\theta_{i,k} = I$ if $\sum_{j\in N_i}\theta_{ij}\beta_{ij,k} \neq 0$. In addition, Ξ_i is full column rank, and $\Xi_i = 0$ if node i is not under attack.

Corollary 3: System (1) is vulnerable under Type-IV attack if and only if there exists $v \in \mathbb{R}^{ns}$ and Q, $W \in \mathbb{R}^{ns\times ns}$ such that:

(1) $v \in \mathrm{Ker}(I_{ns} + \Xi^a Q)$;

(2) v is an unstable eigenvector of $\rho(\xi\otimes A)\theta^a W$, i.e., $\rho(\xi\otimes A)\theta^a W v = \lambda v$ with $|\lambda| \geq 1$;

(3) v is reachable for $(H+\rho(\xi\otimes A), [(H+\rho(\xi\otimes A))\Xi^a \quad \rho(\xi\otimes A)\theta^a])$.

Corollary 4: System (1) under Type-IV attack is strictly vulnerable if and only if

(1) the dynamic differences of estimators, i.e., system (18), is *non-invertible*;

(2) there exists a vector $v \in \mathbb{R}^{2ns}$ satisfying

$$[(H+\rho(\xi\otimes A))\Xi^a \quad \rho(\xi\otimes A)\theta^a]v \neq 0.$$

5) **Type-V attack**: From (7), the dynamics of $\Delta\hat{x}_k$ and Δz_k can be calculated by

$$\begin{aligned}\Delta\hat{x}_{k+1} &= [H+\rho(\xi\otimes A)]\Delta\hat{x}_k + [K\Xi^a \quad \rho(\xi\otimes A)\theta^a]\begin{bmatrix}\alpha_k^a\\ \beta_k^a\end{bmatrix},\\ \Delta z_k &= -C\Delta\hat{x}_k + [\Xi^a \quad 0]\begin{bmatrix}\alpha_k^a\\ \beta_k^a\end{bmatrix},\end{aligned} \tag{19}$$

where

$$\Xi^a = \begin{bmatrix}\Xi_1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \Xi_s\end{bmatrix}, \quad \theta^a = \begin{bmatrix}\theta_1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \theta_s\end{bmatrix},$$

$$\alpha_k^a = \left(\alpha_{1,k}^{\mathrm{T}}, \ \cdots, \ \alpha_{s,k}^{\mathrm{T}}\right)^{\mathrm{T}}, \quad \beta_k^a = \left(\beta_{1,k}^{\mathrm{T}}, \ \cdots, \ \beta_{s,k}^{\mathrm{T}}\right)^{\mathrm{T}},$$

and $\theta_{i,k}\beta_{i,k} = \sum_{j\in N_i}\theta_{ij}\beta_{ij,k}$. In particular, $\theta_{i,k} = I$ if $\sum_{j\in N_i}\theta_{ij}\beta_{ij,k} \neq 0$. In addition, Ξ_i is full column rank, and $\Xi_i = 0$ if node i is not under attack.

Corollary 5: System (1) is vulnerable under Type-V attack if and only if there exists $v \in \mathbb{R}^{ns}$ and Q, $W \in \mathbb{R}^{ns\times ns}$ such that:

TABLE I

Summary of Attack Scenarios

Scenario	Attack mode	Edge	Node	Vulnerable	Strict vulnerable	$rank(C) = ns$
Type-I	single-FDI	✓	–	✓	✓	✕
Type-II	single-FDI	–	✓	✓	✕	✓
Type-III	single-FDI	–	✓	✓	✓	✕
Type-IV	joint-FDI	✓	✓	✓	✓	✓
Type-V	joint-FDI	✓	✓	✓	✓	✓

✓feasible ✕ not feasible

(1) $v \in \mathrm{Ker}(C - \Xi^a Q)$;

(2) v is an unstable eigenvector of $H+\rho(\xi\otimes A)+KC+\rho(\xi\otimes A)\theta^a W$, i.e., $(H+\rho(\xi\otimes A)+KC+\rho(\xi\otimes A)\theta^a W)v = \lambda v$ with $|\lambda| \geq 1$;

(3) v is reachable for $(H+\rho(\xi\otimes A), [K\Xi^a \quad \rho(\xi\otimes A)\theta^a])$.

Corollary 6: System (1) under Type-V attack is strictly vulnerable if and only if

(1) the dynamic differences of estimators, i.e., system (19), is *non-invertible*;

(2) there exists a vector $v \in \mathbb{R}^{2ns}$ satisfying $[K\Xi^a \quad \rho(\xi\otimes A)\theta^a]v \neq 0$.

By simultaneously attacking (though not necessarily all) nodes and edges, it is always possible to render the system vulnerable and strictly vulnerable under joint-FDI attacks regardless of system structures. Compared with single-FDI attacks, joint-FDI attacks pose a more serious threat to the normal operation of systems.

Then, we systematically compare the vulnerability and strict vulnerability conditions under different attack scenarios, which are stated in Table I.

From Table I, joint-FDI attacks have weaker requirements for system structures due to the relaxed implementation conditions. Compared to the former, single-FDI attacks often fall short in terms of the scale of damage or the capacity to remain stealthy. In other words, single-FDI attacks sacrifice stealthiness to enhance attack effectiveness, while joint-FDI attacks disrupt more components of CPSs to achieve attack objectives. Based on the comparison results, defenders have an explicit insight into which edges or sensor nodes are more critical to the security of systems.

IV. Simulation Examples

In this section, we choose the first sub-area of the three-region power system to demonstrate theoretical results. According to Fig. 2, the first sub-area contains 3 generators and 3 buses. A simplified version of the information interaction graph is shown in Fig. 3. Let each node in the graph represent a synchronous generator, and the edges represent connecting lines. For simplicity, assume that these connecting lines correspond to generator buses at both ends.

By using linearization and discretization techniques, the system parameters can be calculated as

$$A = \begin{pmatrix} 0.9942 & 0.0029 & 0.0029 & 0.0988 & 0.0001 & 0.0001\\ 0.0018 & 0.9965 & 0.0018 & 0.0001 & 0.0953 & 0.0001\\ 0.0013 & 0.0013 & 0.9974 & 0.0000 & 0.0000 & 0.0994\\ -0.1149 & 0.0574 & 0.0575 & 0.9735 & 0.0028 & 0.0029\\ 0.0348 & -0.0697 & 0.0349 & 0.0018 & 0.9054 & 0.0018\\ 0.0259 & 0.0259 & -0.0518 & 0.0013 & 0.0013 & 0.9871 \end{pmatrix},$$

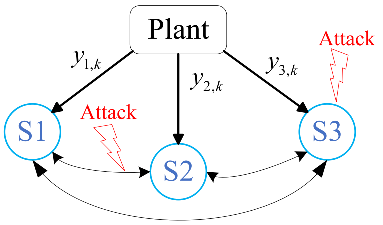

Fig. 3. Information interaction graph of sensor networks under attacks.

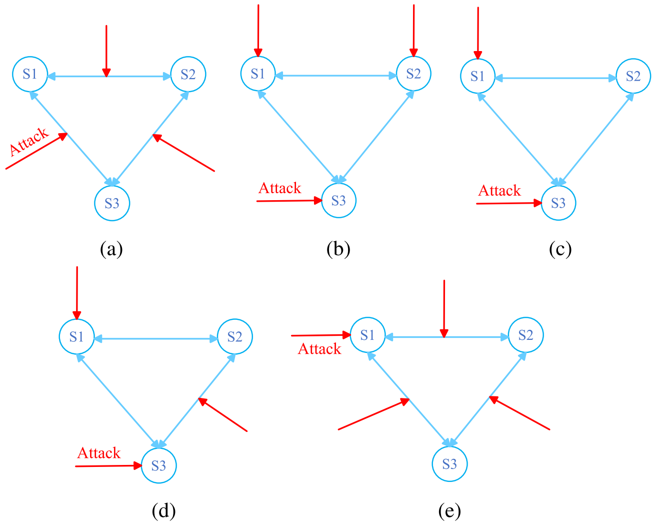

Fig. 4. Topologies when C is not full column rank. (a) Type-I attack. (b) Type-II attack. (c) Type-III attack. (d) Type-IV attack. (e) Type-V attack.

assume that the consensus coefficient ($\rho = 0.1$) and the noise signals ($Q = R_i = I$, where $i = \{1, 2, 3\}$) satisfy the specified conditions.

In what follows, we test the security of the system under different attacks. In particular, the simulation results are conducted by selecting different measurement matrices.

A. C Is Not Full Column Rank

Let $C_1 = (0\ 1\ 1\ 1\ 1\ 1)$, $C_2 = (1\ 1\ 1\ 1\ 1\ 1)$ and $C_3 = (1\ 1\ 1\ 1\ 1\ 1)$ to ensure that $C = \operatorname{diag}\{C_1,\ C_2,\ C_3\}$ is not full column rank. The topologies under different attacks are given in Fig. 4.

Now, we check the (strict) vulnerability of the above topologies. Fig. 5 illustrates the variation of state estimation difference $\Delta\hat{x}_k$ and the residual difference Δz_k, respectively. From Fig. 5(a), the effects of attacks are unbounded, i.e., $\lim_{k\to\infty} \|\Delta\hat{x}_k\| \to \infty$. In Fig. 5(b), the residual bias Δz_k is always zero except under Type-II attack. Therefore, the topologies in Figs. 4(a), (c), (d), (e) are strictly vulnerable, while the one in Fig. 4(b) is vulnerable but not strictly vulnerable. In other words, state estimates are more deserving of protection than measurement data. Furthermore, when only state estimates are under attack, it is more cost-effective to protect state estimates within nodes (as shown in Fig. 4(c)) rather than those within edges (as shown in Fig. 4(a)).

B. C Is Full Column Rank

To make the comparison relatively fair, we still adopt the topologies in Fig. 4, except that $C_1 = (1\ 1\ 1\ 1\ 1\ 1)$. The evolutions of $\Delta\hat{x}_k$ and Δz_k are depicted in Fig. 6. It can be seen that the residual bias Δz_k under single-FDI attacks is non-zero. In other words, the distributed system exhibits resilience to single-FDI attacks but strict vulnerability to joint-FDI attacks. Take Fig. 4(a) as an example, even if all edges are compromised, the attackers are still exposed to anomaly detectors due to divergent residuals. However, if an additional attack signal is injected into Ⓢ1, such attack behaviors would escape monitors, as shown in Fig. 4(e).

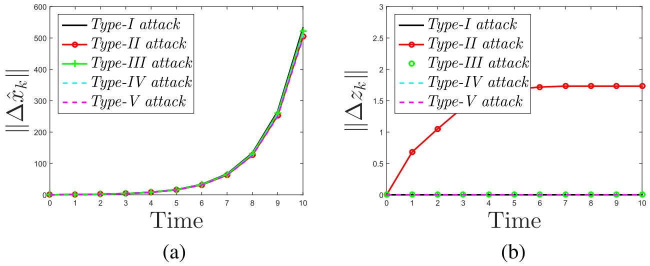

Fig. 5. Evolutions of $\Delta\hat{x}_k$ and Δz_k when $\operatorname{rank}(C) < ns$. (a) State estimation difference $\Delta\hat{x}_k$. (b) Residual difference Δz_k.

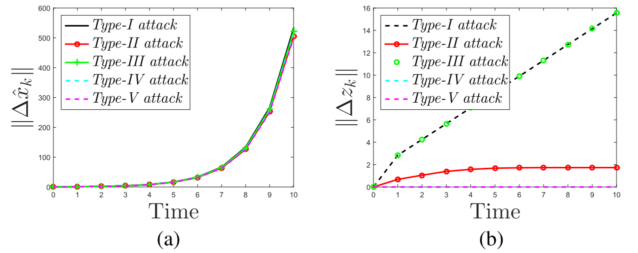

Fig. 6. The evolutions of $\Delta\hat{x}_k$ and Δz_k when $\operatorname{rank}(C) = ns$. (a) State estimation difference $\Delta\hat{x}_k$. (b) Residual difference Δz_k.

In summary, single-FDI attacks sacrifice stealthiness to enhance the attack effectiveness, and joint-FDI attacks disrupt more components of CPSs to achieve the attack objective.

V. Conclusion

In this paper, the vulnerability and strict vulnerability of distributed state estimators under single-FDI attacks and joint-FDI attacks are investigated, respectively. For each attack scenario, we provide necessary and sufficient conditions for vulnerability and strict vulnerability. Compared with the common single-FDI attacks, the derived joint-FDI attacks pose a more serious threat to the operation of CPSs due to the relatively relaxed insecurity conditions. To illustrate the superiority of joint-FDI attacks, a systematic comparison is given to clarify which edges or nodes are more critical to the vulnerability. In the future, we can construct defense strategies and compensation mechanisms to reduce the effect of attacks.

Appendix A
Proof of Theorem 1

(*Sufficiency:*) Note that $v \in \operatorname{Ker}(C)$, it means $\Delta z_k = Cv = 0$, thus the stealthy conditions are satisfied.

Considering that v is reachable for $(H+\rho(\xi\otimes A), \rho(\xi\otimes A)\Xi^a)$, based on (15), we can construct $\Delta\hat{x}_k = v$.

Then, taking $\alpha_k^a = Q\Delta\hat{x}_k$, it follows

$$\begin{aligned}\Delta\hat{x}_{k+1} &= [H + \rho(\xi \otimes A)]\Delta\hat{x}_k + \rho(\xi \otimes A)\Xi^a\alpha_k^a,\\ &= [H + \rho(\xi \otimes A) + \rho(\xi \otimes A)\Xi^a]Qv = \lambda v\end{aligned}$$

with $|\lambda| \geq 1$. In this scenario, it can be seen that $\lim_{k\to\infty} \|\Delta\hat{x}_k\| \to \infty$, i.e., the system (1) under *Type-I attack* is vulnerable.

(*Necessity:*) The proof of necessity can be explained in two steps.

1) If the system (15) is *non-invertible*:

In this scenario, we can construct a nonzero attack sequence $\{\alpha_0^a = \alpha_0^*, \cdots, \alpha_{ns-1}^a = \alpha_{ns-1}^*\}$ satisfying $\{\Delta\hat{x}_1 = \Delta\hat{x}_1^*, \cdots, \Delta\hat{x}_{ns-1} = \Delta\hat{x}_{ns-1}^*, \Delta\hat{x}_{ns} = a_1\Delta\hat{x}_1^* + \cdots + a_1\Delta\hat{x}_{ns-1}^*\}$ and $\{\Delta z_0 = \cdots = \Delta z_{ns-1} = 0\}$, where $\alpha_0^* \neq 0$.

Due to the property of linear systems, if

$$\begin{aligned}&\alpha_0^a = \alpha_0^*, \alpha_1^a = \alpha_1^* - a_{ns-1}\alpha_0^*, \cdots, \alpha_{ns-1}^a\\ &= \alpha_{ns-1}^* - \sum_{i=0}^{ns-2} a_{i+1}\alpha_i^*,\end{aligned}$$

it will lead to $\Delta\hat{x}_{ns} = 0$.

Define a matrix Q such that $\alpha_k^a = Q\Delta\hat{x}_k$, then we construct an attack signal

$$\alpha_k' = \sum_{i=0}^{ns-1} \lambda^{-i}\alpha_i^a. \tag{20}$$

Substituting (20) into (15), let $v = \sum_{i=0}^{ns-1} \lambda^{-i}\Delta\hat{x}_i$, it results in

$$\begin{aligned}&[H + \rho(\xi \otimes A) + \rho(\xi \otimes A)\Xi^a Q]v\\ &= \sum_{i=0}^{ns-1} \lambda^{-i}[(H + \rho(\xi \otimes A))\Delta\hat{x}_i\\ &\quad + \rho(\xi \otimes A)\Xi^a\alpha_i^a]\\ &= \sum_{i=0}^{ns-1} \lambda^{-i}\Delta\hat{x}_{i+1} = \lambda \sum_{i=0}^{ns-1} \lambda^{-i}\Delta\hat{x}_i = \lambda v,\\ -Cv &= -\sum_{i=0}^{ns-1} \lambda^{-i}C\Delta\hat{x}_i = 0.\end{aligned}$$

Since the system (1) under *Type-I attack* is vulnerable, it means that v is an unstable eigenvector of $H + \rho(\xi \otimes A) + \rho(\xi \otimes A)\Xi^a Q$, i.e., $[H + \rho(\xi \otimes A) + \rho(\xi \otimes A)\Xi^a Q]v = \lambda v$ with $|\lambda| \geq 1$.

In addition, the condition $Cv = 0$ is equivalent to $v \in \mathrm{Ker}(C)$. Combining with that $v = \sum_{i=0}^{ns-1} \lambda^{-i}\Delta\hat{x}_i$ is reachable for $(H + \rho(\xi \otimes A), \rho(\xi \otimes A)\Xi^a)$, hence conditions (1)-(3) are all satisfied.

2) If the system (15) is *invertible*:

Similar to [16], construct the following attack signal

$$\alpha_k^a = \sum_{i=0}^{ns-1} K_i[\Delta z_{k+i} - C(H + \rho(\xi \otimes A))^i\Delta\hat{x}_k],$$

in which K_i $(i = 0, 1, \cdots, ns-1)$ are the gain coefficients.

Based on the compatibility of norms, it is easy to see

$$\begin{aligned}&\frac{\|\alpha_k^a\|}{\|\Delta\hat{x}_k\| + 1}\\ &\leq \sum_{i=0}^{ns-1} \|K_i\|[\frac{\|\Delta z_{k+i}\|}{\|\Delta\hat{x}_k\| + 1} + \|C(H + \rho(\xi \otimes A))^i\|\frac{\Delta\hat{x}_k}{\|\Delta\hat{x}_k\| + 1}]\\ &\leq \sum_{i=0}^{ns-1} \|K_i\|(\delta + \|C(H + \rho(\xi \otimes A))^i\|).\end{aligned} \tag{21}$$

Since (21) is bounded, there exists a convergent subsequence $\{i_k : k \in \mathbb{N}\}$ such as

$$\lim_{k\to\infty} \frac{\Delta\hat{x}_{i_k+q}}{\|\Delta\hat{x}_{i_k+q}\| + 1} = \breve{x}_q, \quad \lim_{k\to\infty} \frac{\Delta\alpha_{i_k+q}}{\|\Delta\hat{x}_{i_k+q}\| + 1} = \breve{\alpha}_q, \tag{22}$$

where $\lim\sup_{k\to\infty} \|\Delta\hat{x}_{i_k+q}\| \to \infty$ with $\|\Delta\hat{x}_{i_k}\| \leq \|\Delta\hat{x}_{i_k+q}\|$ for arbitrary $q = -p, \cdots, 0, \cdots, ns-1$.

From $\lim\sup_{k\to\infty} \|\Delta\hat{x}_{i_k+q}\| \to \infty$ and $\|\Delta z_{i_k+q}\| \leq \delta$, then

$$\begin{aligned}&[H + \rho(\xi \otimes A)]\breve{x}_q + \rho(\xi \otimes A)\Xi^a\breve{\alpha}_q =\\ &\lim_{k\to\infty} \frac{\|\Delta\hat{x}_{i_k+q+1}\| + 1}{\|\Delta\hat{x}_{i_k+q}\| + 1}\breve{x}_{q+1} = \lim_{k\to\infty} c_q\breve{x}_{q+1}\end{aligned} \tag{23}$$

and

$$C\breve{x}_q = \lim_{k\to\infty} \frac{\Delta z_{i_k+q}}{\|\Delta\hat{x}_{i_k+q}\| + 1} = 0. \tag{24}$$

Note that $\Delta\hat{x}_k$ is ns-dimensional, thus $\breve{x}_{-p}, \cdots, \breve{x}_{ns-1}$ are linearly dependent. Hence, one can construct

$$V = span[\breve{x}_{-p}, \cdots, \breve{x}_t],$$

where $0 \leq t \leq ns - 1$ and $\mathrm{span}[\breve{x}_{-p}, \cdots, \breve{x}_t] = \mathrm{span}[\breve{x}_{-p}, \cdots, \breve{x}_{t+1}]$, which implies that $\breve{x}_{t+1} \in V$.

Considering that $\breve{x} = b_{-p}\breve{x}_{-p} + \cdots + b_{-p}\breve{x}_{-p} \in V$, it follows that

$$\begin{aligned}&[H + \rho(\xi \otimes A)]\breve{x} + \rho(\xi \otimes A)\Xi^a[b_{-p}\breve{\alpha}_{-p} + \cdots + b_{-p}\breve{\alpha}_{-p}]\\ &= b_{-p}c_{-p}\breve{x}_{-p+1} + \cdots + b_d c_d\breve{x}_{t+1} \in V\end{aligned}$$

and $C\breve{x} = b_{-p}C\breve{x}_{-p} + \cdots + b_d C\breve{x}_t = 0$. Then, $\breve{x}_{-p}, \cdots, \breve{x}_t$ is reachable for $(H + \rho(\xi \otimes A), \rho(\xi \otimes A)\Xi^a)$.

Due to the property of linear systems, we can choose a unique matrix Q such that $\breve{\alpha}_q = Q\breve{\alpha}_q$ holds for all $q = -p, \cdots, 0$.

In what follows, we prove that $H + \rho(\xi \otimes A) + \rho(\xi \otimes A)\Xi^a Q$ is unstable by contradiction. Assume that $H + \rho(\xi \otimes A) + \rho(\xi \otimes A)\Xi^a Q$ is stable, which means that

$$\|[H + \rho(\xi \otimes A) + \rho(\xi \otimes A)\Xi^a Q]^p v\| \leq \|v\|. \tag{25}$$

Accordingly, (25) can be rewritten as

$$\begin{aligned}&[H + \rho(\xi \otimes A)]\breve{x}_q + \rho(\xi \otimes A)\Xi^a\breve{\alpha}_q\\ &= \lim_{k\to\infty} \frac{\|\Delta\hat{x}_{i_k+q+1}\| + 1}{\|\Delta\hat{x}_{i_k+q}\| + 1}\breve{x}_{q+1}\\ &= [H + \rho(\xi \otimes A) + \rho(\xi \otimes A)\Xi^a Q]\breve{x}_{q+1}.\end{aligned}$$

By iterations, it can be derived that

$$\begin{aligned}&\lim_{k\to\infty} \frac{\|\Delta\hat{x}_{i_k}\| + 1}{\|\Delta\hat{x}_{i_k-p}\| + 1}\breve{x}_0\\ &= [H + \rho(\xi \otimes A) + \rho(\xi \otimes A)\Xi^a Q]^p\breve{x}_{-p}.\end{aligned}$$

Note that $\|\Delta\hat{x}_{i_k}\| \geq \|\Delta\hat{x}_{i_k-p}\|$, thus $\lim_{k\to\infty} \frac{\|\Delta\hat{x}_{i_k}\|+1}{\|\Delta\hat{x}_{i_k-p}\|+1} \geq 1$. Combine with $\|\breve{x}_{-p}\| = \|\breve{x}_0\| = 1$, it is easy to see

$$\begin{aligned}\|\breve{x}_{-p}\| = \|\breve{x}_0\| &\leq \|\lim_{k\to\infty} \frac{\|\Delta\hat{x}_{i_k}\| + 1}{\|\Delta\hat{x}_{i_k-p}\| + 1}\breve{x}_0\|\\ &= \|H + \rho(\xi \otimes A) + \rho(\xi \otimes A)\Xi^a Q]^p\breve{x}_{-p}\|,\end{aligned}$$

which contradicts with the assumption in (25). Hence, $H + \rho(\xi \otimes A) + \rho(\xi \otimes A)\Xi^a Q$ is unstable on V, i.e., there exists a vector $v \in V$ satisfying $[H + \rho(\xi \otimes A) + \rho(\xi \otimes A)\Xi^a Q]v = \lambda v$ with $|\lambda| \geq 1$.

In summary, conditions (1)-(3) are all satisfied, thus the proof is complete.

Appendix B
Proof of Theorem 2

(*Sufficiency:*) Since system (15) is *non-invertible*, we can construct an attack sequence $\{\alpha_k^a\}$ such that $\Delta z_k = 0$, $\forall k \in \mathbb{N}$. Without loss of generality, let $\{\alpha_0^a\} \neq 0$, otherwise one can always remove the zero signal in the attack $\{\alpha_k^a\}$. In this scenario, it can be seen that

$$\begin{aligned} \begin{bmatrix} \Delta\hat{x}_1 \\ \Delta z_0 \end{bmatrix} &= \begin{bmatrix} H + \rho(\xi \otimes A) \\ -C \end{bmatrix} \Delta\hat{x}_0 + \begin{bmatrix} \rho(\xi \otimes A)\Xi^a \\ 0 \end{bmatrix} \alpha_0^a \\ &= \begin{bmatrix} \rho(\xi \otimes A)\Xi^a \\ 0 \end{bmatrix} \alpha_0^a. \end{aligned}$$

Let $\alpha_0^a = v$, it will lead to $\Delta\hat{x}_1 = \rho(\xi \otimes A)\Xi^a v \neq 0$. For convenience, we can take $\|\Delta\hat{x}_1\| = 1$ by elaborately designing the vector v.

Note that if $\limsup_k \|\Delta\hat{x}_k\| \to \infty$, the proof will be completed immediately. If $\limsup_k \|\Delta\hat{x}_k\| \leq \delta$, we can also complete the proof by contradiction.

Consider the following attack signal

$$\alpha_k' = \sum_{s=0}^{k} (2\delta + 1)^{k-s} \alpha_s^a.$$

By using the inductive method, it can be derived that

$$\begin{aligned} \Delta\hat{x}_{k+1}' &= \sum_{s=0}^{k} (2\delta + 1)^{k-s} \Delta\hat{x}_s, \\ \Delta z_k' &= \sum_{s=0}^{k} (2\delta + 1)^{k-s} \Delta z_s. \end{aligned}$$

Then, one can easily obtain that

$$\begin{aligned} \|\Delta\hat{x}_{k+1}'\| &\geq (2\delta + 1)^{k-1} \|\Delta\hat{x}_1\| - \sum_{s=2}^{k} (2\delta + 1)^{k-s} \|\Delta\hat{x}_s\| \\ &\geq (2\delta + 1)^{k-1} - \sum_{s=2}^{k} (2\delta + 1)^{k-s} \delta \\ &\geq \frac{(2\delta + 1)^{k-1} + 1}{2}, \\ \Delta z_k' &= \sum_{s=0}^{k} (2\delta + 1)^{k-s} \Delta z_s = 0, \end{aligned}$$

which is equivalent to

$$\limsup_{k\to\infty} \|\Delta\hat{x}_k'\| \to \infty, \quad \text{and} \quad \|\Delta z_k{}'\| = 0, \ \forall k \in \mathbb{N}.$$

From Definition 2, it follows that the system (1) under *Type-I attack* is strictly vulnerable.

(*Necessity:*) The necessity proof is illustrated by two steps.

1) The proof of condition (1) can be obtained directly from the definition of strict vulnerability.

Considering the system (1) is strictly vulnerable, we can construct an attack sequence $\{\alpha_k^a\}$ such that $\|\Delta z_k\| = 0$, $\forall k \in \mathbb{N}$. Note that Δz_k is a column vector, thus $\Delta z_k = 0$ holds. Clearly, the system (1) is *non-invertible*.

2) We prove condition (2) by contradiction. Suppose that the system (1) is strictly vulnerable, while for every $v \in \mathbb{R}^{ns}$ satisfies $\rho(\xi \otimes A)\Xi^a v = 0$. Note that

$$\begin{aligned} \begin{bmatrix} \Delta\hat{x}_1 \\ \Delta z_0 \end{bmatrix} &= \begin{bmatrix} H + \rho(\xi \otimes A) \\ -C \end{bmatrix} \Delta\hat{x}_0 + \begin{bmatrix} \rho(\xi \otimes A)\Xi^a \\ 0 \end{bmatrix} \alpha_0^a \\ &= \begin{bmatrix} \rho(\xi \otimes A)\Xi^a \\ 0 \end{bmatrix} \alpha_0^a = 0, \\ \begin{bmatrix} \Delta\hat{x}_2 \\ \Delta z_1 \end{bmatrix} &= \cdots = \begin{bmatrix} \Delta\hat{x}_{k+1} \\ \Delta z_k \end{bmatrix} = 0, \end{aligned}$$

which implies that $\Delta\hat{x}_k$ is not norm-unbounded, thus violating the definition of strict vulnerability. The proof is complete.

Appendix C
Proof of Theorem 4

(*Sufficiency:*) We need to prove that the above condition, i.e., there exists a vector $v \in \mathbb{R}^{ns}$ such that $Cv = 0$ and $[H + \rho(\xi \otimes A)]v \neq 0$, implies vulnerability.

Consider the following attack signal

$$\begin{cases} \Xi_1^a \alpha_1^a = \eta_1 v, \\ \Xi^a \alpha_k^a = (\eta_k I_{ns} - \eta_{k-1}[H + \rho(\xi \otimes A)])v, \quad k \geq 2, \end{cases} \tag{26}$$

where $\{\eta_k\}$ is a non-zero real-number sequence.

Note that

$$\begin{aligned} \Delta\hat{x}_1 &= [H + \rho(\xi \otimes A)][0 + (\eta_1 v)] \\ &= \eta_1 [H + \rho(\xi \otimes A)]v, \\ &\vdots \\ \Delta\hat{x}_{k+1} &= [H + \rho(\xi \otimes A)][\Delta\hat{x}_k + \Xi_{k+1}^a \alpha_{k+1}^a v] \\ &= \eta_{k+1} [H + \rho(\xi \otimes A)]v, \end{aligned}$$

which shows that $\limsup_{k\to\infty} \|\Delta\hat{x}_{k+1}'\| \to \infty$ if $\limsup_{k\to\infty} |\eta_{k+1}| \to \infty$. In addition, it can be seen that

$$\|\Delta z_k\| = \| - \eta_k C v\| = 0 \leq \delta,$$

which implies that the system (1) is vulnerable.

(*Necessity:*) We prove necessity by contradiction. Suppose that the system (1) is vulnerable if

1) for every $v \in \mathbb{R}^{ns}$ satisfies $Cv = 0$ and $[H + \rho(\xi \otimes A)]v = 0$. In this scenario, it follows $\|\Delta\hat{x}_k\| = \|\Delta z_k\| = 0$, which violates the definition of vulnerability.

2) for every $v \in \mathbb{R}^{ns}$ satisfies $Cv \neq 0$ and $[H + \rho(\xi \otimes A)]v = 0$. In this scenario, it follows $\|\Delta\hat{x}_k\| = 0$ and $\|\Delta z_k\| \neq 0$, which violates the definition of vulnerability.

3) for every $v \in \mathbb{R}^{ns}$ satisfies $Cv \neq 0$ and $[H + \rho(\xi \otimes A)]v \neq 0$. In this scenario, the matrix C is full column rank. Denote the left inverse matrix of C as D, such as

$$DCv = v.$$

Then, $||\Delta\hat{x}_{k+1}||$ can be calculated by

$$\begin{aligned}||\Delta\hat{x}_{k+1}|| &= ||[H+\rho(\xi\otimes A)]D\Delta z_k||\\ &\le ||H+\rho(\xi\otimes A)||\,||D||\,\delta,\end{aligned}$$

which means that $\Delta\hat{x}_{k+1}$ is not norm-unbounded, thus violating the definition of vulnerability. The proof is complete.

References

[1] S. Ghosh, M. R. Bhatnagar, W. Saad, and B. K. Panigrahi, "Defending false data injection on state estimation over fading wireless channels," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 1424–1439, 2021.

[2] X. Ge, Q.-L. Han, M. Zhong, and X.-M. Zhang, "Distributed Krein space-based attack detection over sensor networks under deception attacks," *Automatica*, vol. 109, Nov. 2019, Art. no. 108557.

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

[4] D. Wang, J. Huang, Y. Tang, and F. Li, "A watermarking strategy against linear deception attacks on remote state estimation under K-L divergence," *IEEE Trans. Ind. Informat.*, vol. 17, no. 5, pp. 3273–3281, May 2021.

[5] R. Zhang and P. Venkitasubramaniam, "Stealthy control signal attacks in linear quadratic Gaussian control systems: Detectability reward tradeoff," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 7, pp. 1555–1570, 2017.

[6] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "A secure control framework for resource-limited adversaries," *Automatica*, vol. 51, pp. 135–148, Jan. 2015.

[7] X. Shao and D. Ye, "Fuzzy adaptive event-triggered secure control for stochastic nonlinear high-order MASs subject to DoS attacks and actuator faults," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 12, pp. 3812–3821, Dec. 2021.

[8] H. Zhang and W. X. Zheng, "Denial-of-service power dispatch against linear quadratic control via a fading channel," *IEEE Trans. Autom. Control*, vol. 63, no. 9, pp. 3032–3039, Sep. 2018.

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

[10] D. Ye and T.-Y. Zhang, "Summation detector for false data-injection attack in cyber-physical systems," *IEEE Trans. Cybern.*, vol. 50, no. 6, pp. 2338–2345, Jun. 2020.

[11] P. Li and D. Ye, "Detection and performance compensation for linear ϵ-stealthy attacks in cyber-physical systems," *IEEE Trans. Control Netw. Syst.*, vol. 10, no. 3, pp. 1338–1349, Sep. 2023.

[12] H. Song, P. Shi, C.-C. Lim, W.-A. Zhang, and L. Yu, "Attack and estimator design for multi-sensor systems with undetectable adversary," *Automatica*, vol. 109, Nov. 2019, Art. no. 108545.

[13] Q. Zhang, K. Liu, and Y. Xia, "Optimal stealthy deception attack against cyber-physical systems," *IEEE Trans. Cybern.*, vol. 50, no. 9, pp. 3963–3972, Sep. 2020.

[14] P. Li and D. Ye, "Measurement-based optimal stealthy attacks on remote state estimation," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 3365–3374, 2022.

[15] W. Xu, Z. Wang, L. Hu, and J. Kurths, "State estimation under joint false data injection attacks: Dealing with constraints and insecurity," *IEEE Trans. Autom. Control*, vol. 67, no. 12, pp. 6745–6753, Dec. 2022, doi: 10.1109/TAC.2021.3131145.

[16] T.-Y. Zhang and D. Ye, "False data injection attacks with complete stealthiness in cyber–physical systems: A self-generated approach," *Automatica*, vol. 120, Oct. 2020, Art. no. 109117.

[17] T. Sui, Y. Mo, D. Marelli, X. Sun, and M. Fu, "The vulnerability of cyber-physical system under stealthy attacks," *IEEE Trans. Autom. Control*, vol. 66, no. 2, pp. 637–650, Feb. 2021.

[18] T.-Y. Zhang, D. Ye, and Y. Shi, "Decentralized false-data injection attacks against state omniscience: Existence and security analysis," *IEEE Trans. Autom. Control*, vol. 68, no. 8, pp. 1–15, Sep. 2022.

[19] T. Sui and X.-M. Sun, "The vulnerability of distributed state estimator under stealthy attacks," *Automatica*, vol. 133, Nov. 2021, Art. no. 109869.

[20] J. Zhou, W. Yang, H. Zhang, W. X. Zheng, Y. Xu, and Y. Tang, "Security analysis and defense strategy of distributed filtering under false data injection attacks," *Automatica*, vol. 138, Apr. 2022, Art. no. 110151.

[21] J. Zhou, W. Yang, W. Ding, W. X. Zheng, and Y. Xu, "Watermarking-based protection strategy against stealthy integrity attack on distributed state estimation," *IEEE Trans. Autom. Control*, vol. 68, no. 1, pp. 628–635, Jan. 2023.

[22] S. Liu and P. X. Liu, "Distributed model-based control and scheduling for load frequency regulation of smart grids over limited bandwidth networks," *IEEE Trans. Ind. Informat.*, vol. 14, no. 5, pp. 1814–1823, May 2018.

[23] X. Chen, S. Hu, Y. Li, D. Yue, C. Dou, and L. Ding, "Co-estimation of state and FDI attacks and attack compensation control for multi-area load frequency control systems under FDI and DoS attacks," *IEEE Trans. Smart Grid*, vol. 13, no. 3, pp. 2357–2368, May 2022.

[24] X. Wang, X. Luo, M. Zhang, Z. Jiang, and X. Guan, "Detection and isolation of false data injection attacks in smart grid via unknown input interval observer," *IEEE Internet Things J.*, vol. 7, no. 4, pp. 3214–3229, Apr. 2020.

[25] B. Qu, Z. Wang, B. Shen, and H. Dong, "Distributed state estimation for renewable energy microgrids with sensor saturations," *Automatica*, vol. 131, Sep. 2021, Art. no. 109730.

[26] M. Sain and J. Massey, "Invertibility of linear time-invariant dynamical systems," *IEEE Trans. Autom. Control*, vols. AC-14, no. 2, pp. 141–149, Apr. 1969.

Pengyu Li received the B.S. degree from the North University of China, Taiyuan, China, in 2017, and the M.S. degree from the School of Information Science and Engineering, Lanzhou University, Lanzhou, China, in 2020. He is currently pursuing the Ph.D. degree in control science and engineering with the College of Information Science and Engineering, Northeastern University, Shenyang, China. His current research interests include the security of cyber-physical systems and distributed systems.

Dan Ye (Senior Member, IEEE) received the B.S. and M.S. degrees in mathematics and applied mathematics from Northeast Normal University, China, in 2001 and 2004, respectively, and the Ph.D. degree in control theory and engineering from Northeastern University, China, in 2008. From 2008 to 2010, she was a Lecturer with Northeastern University. She is currently a Professor with the College of Information Science and Engineering, Northeastern University. Her research interests include fault-tolerant control, robust control, adaptive control, and security of cyber-physical systems.