SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

April 2026, Vol. 69, Iss. 4, 142108:1–142108:15
https://doi.org/10.1007/s11432-025-4567-3

Stealthy integrity attacks and optimal selective protection for distributed state estimation

Jiayu ZHOU[1,2], Cheng QIAO[1], Yanbin SUN[1], Wen YANG[2] & Zhihong TIAN[1*]

[1]*Cyberspace Institute of Advanced Technology, Guangzhou University, Guangzhou 510006, China*
[2]*Key Laboratory of Smart Manufacturing in Energy Chemical Process, East China University of Science and Technology, Shanghai 200237, China*

Received 6 May 2025/Revised 3 July 2025/Accepted 25 August 2025/Published online 5 March 2026

Abstract This paper studies the security of distributed state estimation under integrity attacks. Considering cost and energy constraints, only a subset of the sensors within the network is protected. To reveal the vulnerability of such partial protection schemes, we investigate integrity attacks in distributed networks that can bypass anomaly detectors and lead to unbounded estimation errors. First, the feasibility analyses of such attacks under both strict and non-strict stealthiness are derived using matrix decomposition theory, along with the corresponding necessary conditions and sufficient conditions. Second, an explicit form of integrity attack is presented, and an offline algorithm for generating attack sequences is provided. These reflect how an adversary can undermine the estimation performance by targeting only unprotected sensors. It is demonstrated that the deviations of the detected quantities induced by the attack sequences decay exponentially, suggesting that the stealthiness of the proposed attack is enhanced compared to previous methods, particularly with an increased detection interval. Third, a sufficient condition for securing sensor networks is established. Utilizing the constraints determined by this condition, the optimization of the selective protection strategy is formulated, which identifies both the minimum number of sensors to protect and which specific sensors they are. Finally, a numerical simulation is conducted to validate the proposed results.

Keywords distributed network, state estimation, integrity attack, vulnerability, selective protection

Citation Zhou J Y, Qiao C, Sun Y B, et al. Stealthy integrity attacks and optimal selective protection for distributed state estimation. Sci China Inf Sci, 2026, 69(4): 142108, https://doi.org/10.1007/s11432-025-4567-3

1 Introduction

Thanks to advances in ubiquitous sensing and cloud-edge computing, the scale of cyber-physical systems (CPS) has expanded dramatically, leading to a wide range of applications in many fields such as power, transportation, and manufacturing [1]. In CPS, a large number of sensing devices, interconnected via wireless networks, are typically deployed to leverage their information interaction for executing intricate engineering tasks [2]. Therefore, how to accurately monitor the operational status of CPS is critical to the completion of the tasks. In this regard, distributed state estimation has attracted a great deal of attention for providing reliable technical support [3–5]. The openness of sensor networks, while enhancing communication efficiency, also presents one of the primary avenues for adversaries to compromise the accuracy of state estimation. Depending on the purpose of the attack, they can be primarily categorized into eavesdropping attacks, denial-of-service (DoS) attacks, and integrity attacks [6–8]. In contrast to the first two types of attacks, integrity attacks aim to tamper with data. Such attacks compromise the precision of state estimation, resulting in erroneous decisions and control actions within CPS.

So far, many efforts have been made on data integrity in state estimation by studying the possible strategies adopted by adversaries. Niu et al. [9] proposed an innovation-based stealthy attack on distributed networks and determined the optimal parameters through semi-definite programming. Based on the input and output signals of the system, Park et al. [10] constructed a zero-dynamic attack to destabilize the internal state of the system. In order to make the compromised estimates of the nodes as close as possible to the predetermined values, Choraria et al. [11] developed an algorithm to find the optimal values for a class of linear attacks and demonstrated its convergence. For the scenario where only a subset of network sensors are reliable, Xu et al. [12] proposed a stealthy attack strategy based on historical data. Moreover, the authors showed that the strategy is capable of circumventing detection mechanisms that rely on mutual verification of information between secure and suspicious sensors. The

* Corresponding author (email: tianzhihong@gzhu.edu.cn)

concept of perfect attack, which aims at diverging the estimation error, has also attracted great interest from researchers in the field. Hu et al. [13] derived a necessary and sufficient condition under which a remote estimator cannot withstand perfect attacks and provided an algorithm for generating attack sequences. In the case where the estimates can be manipulated, Ni et al. [14] established a necessary and sufficient condition for the existence of a perfect attack on remote estimation systems. Similar work was studied in [15], where Jin et al. introduced a completely stealthy attack to impair the estimation performance.

In addition to investigating the potential attack strategies that adversaries might employ, some researchers have delved deeper into the design and implementation of defense mechanisms. Regarding deception attacks on measurement data, Xiao et al. [16] developed an adaptive event-triggered scheme to address the state estimation problem in large-scale systems. In order to resist the influence of unbounded attack signals on the system, Kircher et al. [17] proposed an estimator based on the resolution of a convex optimization problem and proved the boundedness of the resulting estimation error. In view of the complexity and diversity of integrity attacks, the design of detection mechanisms for wireless communication data has garnered significant attention. Li et al. [18] presented three data validation methods that leverage the correlation between reliable and suspicious sensors to identify false data. To improve the detection of stealthy attacks, Zhou et al. [19] proposed a watermarking-based data transmission strategy and analyzed its effect on stealthy attacks in different scenarios. For the purpose of excluding suspicious sensors under false data injection attacks, Suo et al. [20] put forward an attack detection scheduling algorithm and discussed the boundedness of the estimation error.

Despite the advancements made in the previously discussed studies, essential challenges remain to be addressed. To begin with, the attack strategies suggested by previous studies sacrifice a certain level of stealthiness for achieving attack utility, potentially increasing the risk of detection for adversaries who lack comprehensive system information. In addition, the current research does not fully consider the impact of resource constraints, such as cost and energy, on sensor networks, resulting in inadequate scalability of the protection schemes. These gaps highlight the need to utilize limited resources to provide protection against integrity attacks. Given the practical consideration of protecting only a subset of sensors in large-scale distributed networks, it is imperative to investigate the reliability and security of such selective protection schemes. Consequently, three important issues are worth discussing. (i) How to assess the vulnerability of such selective protection schemes? (ii) What attack sequences would an adversary design to impair system performance? (iii) Which sensors need to be protected to defend against integrity attacks? In light of the aforementioned issues, this paper undertakes a comprehensive investigation, and the main contributions are as follows.

(1) We demonstrate the infeasibility of strictly stealthy integrity attacks in sensor networks with partial protection. In addition, we establish the necessary conditions and sufficient conditions for the feasibility of non-strictly stealthy attacks, enhancing the insights into security vulnerabilities in sensor networks.

(2) From the perspective of an adversary, we derive an explicit form of the attack that circumvents the anomaly detector and leads to unbounded estimation errors. Additionally, we demonstrated that the deviations of the detected quantities induced by the attack exhibit exponential decay, indicating that the proposed attack possesses a higher degree of stealth compared to the method in [13].

(3) From the perspective of a defender, we determine a sufficient condition for securing sensor networks. By formulating this condition as an optimization problem, we propose a defense scheme that strategically selects the smallest set of sensors for protection, thereby enabling a more efficient and targeted allocation of resources towards countermeasure implementation.

Notations: $\mathbb{R}^n$ and $\mathbb{R}^{m\times n}$ represent the n-dimensional Euclidean space and $m \times n$ matrices, respectively. For a matrix X, $\mathrm{Col}(X)$, $\mathrm{Nul}(X)$, $\mathrm{Rank}(X)$, $\rho(X)$ and $\Psi(X)$ denote its column space, null space, rank, spectral radius and the set consisting of all its columns, respectively. $X \succ 0$ (or $X \succeq 0$) means that X is a positive definite (or positive semi-definite) matrix. $|\cdot|$ indicates the determinant of a matrix or the absolute value of a scalar, and $\emptyset$ stands for the empty set. $\mathrm{Span}(\cdot)$ represents the set of all possible linear combinations of a given collection of vectors, and $\mathrm{Diag}(\cdot)$ refers to a diagonal matrix or a block diagonal matrix. For a set of vectors or matrices $x_1, \ldots, x_N$, we sometimes abbreviate $\mathrm{Diag}(x_1, \ldots, x_N)$ to $\mathrm{Diag}(x_i)$ for ease of expression. $\mathbf{1}$ denotes a column vector where all entries are equal to 1 and $\|\cdot\|$ stand for the Euclidean norm or spectral norm.

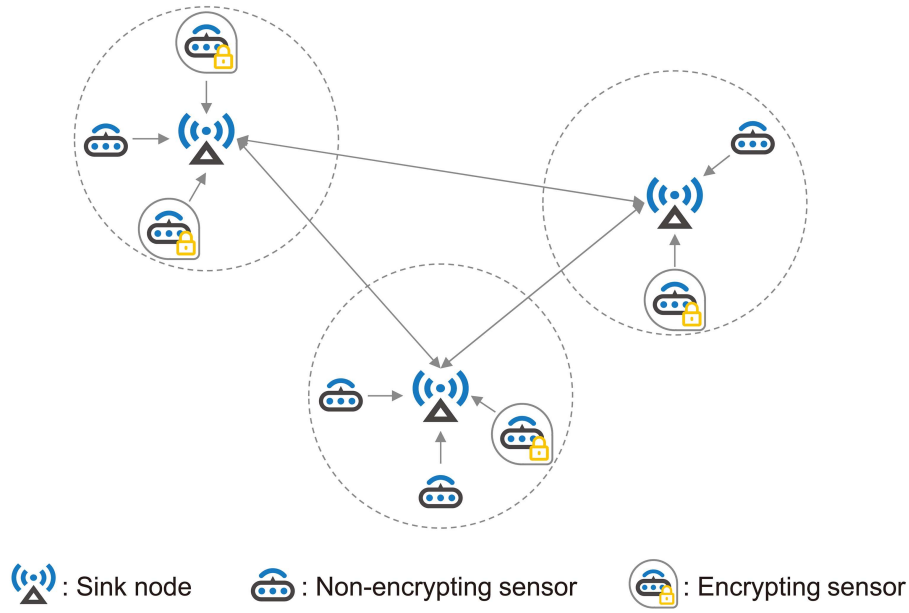

Figure 1 (Color online) Communication diagram for clustered sensor networks.

2 Problem formulation

2.1 Distributed sensor network and attack model

Consider a discrete-time system characterized by the following state-space model:

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

where $x_k \in \mathbb{R}^n$ is the system state, and $w_k \in \mathbb{R}^n$ is the Gaussian noise with zero-mean and covariance $Q \succeq 0$. A clustered sensor network containing N sink nodes is deployed for estimating the state of the system (1). To obtain the measurement output of the system, each sink node communicates with a set of wireless sensors. Letting $M_i = \{1, \ldots, m_i\}$ denote the set of sensors corresponding to the sink node i, the measurement equation for the s-th sensor is given as follows:

$$y_{i,k}^{(s)} = H_i^{(s)} x_k + \nu_{i,k}^{(s)}, \ s \in M_i, \tag{2}$$

where $y_{i,k}^{(s)} \in \mathbb{R}$ is the measurement output and $\nu_{i,k}^{(s)} \in \mathbb{R}$ is the Gaussian noise with zero-mean and covariance $R_i > 0$. It is assumed that all the measurement noises are mutually uncorrelated and are uncorrelated with process noise. The data transmission process of the clustered sensor network is shown in Figure 1. At each time step, each sink node conducts state estimation by integrating its own measurements with those sent from neighboring nodes.

For the sake of clarity, we utilize a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ to represent the network topology of the sink nodes. The set of sink nodes is defined as $\mathcal{V} = \{1, \ldots, N\}$, and the relationship between nodes is represented by the edge set $\mathcal{E} = \mathcal{V} \times \mathcal{V}$. If sink node i can directly receive information from sink node j, then $(i, j) \in \mathcal{E}$. The in-neighboring set of the sink node i is defined as $\mathcal{N}_i = \{j \in \mathcal{V} | (i, j) \in \mathcal{E}\}$. The adjacency of the graph $\mathcal{G}$ is defined as $\mathcal{A} = [a_{ij}]$, where

$$a_{ij} = \begin{cases} 1, & \text{if } (i, j) \in \mathcal{E}, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

Data encryption is widely recognized as an important defense against cyber attacks. However, the large scale of sensor networks, where sensors may rely on battery power, renders limitations such as cost and energy to be factors that cannot be ignored. In view of this, it is a feasible approach to use sensors with encryption capability only for some critical measurement data. Compared to protecting all sensors, this approach is cost-effective because it reduces overall security expenditures by focusing resources where they are most needed. Such a method is similar to those presented in [12,18], which assumed that only some of the sensors in the network are reliable. To facilitate the presentation, we define a binary variable that signifies the encryption status of the measurement as follows:

$$\eta_i^{(s)} = \begin{cases} 1, & \text{the measurement } y_{i,k}^{(s)} \text{ is encrypted}, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

The measurement of the encrypted sensor is defined as

$$\psi_{i,k}^{(s)} = \mathscr{C}\big(y_{i,k}^{(s)}, \delta_{i,k}\big), \tag{5}$$

where $\psi_{i,k}^{(s)}$ is the ciphertext, $\mathscr{C}(\cdot,\cdot)$ is the encryption function, and $\delta_{i,k}$ is the secret key. To reduce the computational complexity, the encryption mode is considered to be symmetric, i.e., the encryption and decryption operations both utilize the same secret key. After the sink node receives the encrypted data, it decrypts the ciphertext using the secret key $\delta_{i,k}$ to recover the original data:

$$y_{i,k}^{(s)} = \mathscr{C}^{-1}\big(\psi_{i,k}^{(s)}, \delta_{i,k}\big). \tag{6}$$

Overall, the deployed sensors are classified into two categories based on their capabilities: one category is endowed with encryption functionality, while the other lacks this feature. In addition to encrypting data, utilizing communication channels with higher security to transmit data is also one approach to protect the integrity of data [21]. This approach typically involves the use of dedicated network links, such as through private networks or trusted provider services, to reduce the risks associated with data transmission over public networks. In terms of the research issue of this paper, the transmission of data from the sensors to the sink node over secure channels can also be modeled as shown in (4).

Since the sink node is capable of communication, it can transmit its measurement to neighboring nodes to increase their accuracy. Let $y_{i,k} = [y_{i,k}^{(1)}, \ldots, y_{i,k}^{(m_i)}]^\top$ and $H_i = [H_i^{(1)\top}, \ldots, H_i^{(m_i)\top}]^\top$. For a given topology, each sink node estimates the state of the system by utilizing the following distributed recursive estimator [9]:

$$\begin{aligned}\hat{x}_{i,k+1} = {} & A\hat{x}_{i,k} + AK_{i,k}(y_{i,k} - H_i\hat{x}_{i,k}) \\ & + AG_{i,k}\sum_{j\in\mathcal{N}_i}(y_{j,k} - H_j\hat{x}_{i,k}),\end{aligned} \tag{7}$$

where $\hat{x}_{i,k}$ is the estimate generated by sink node i, and $K_{i,k}$ and $G_{i,k}$ are the gain matrices.

Assumption 1. The pair $(A, \breve{H}_i)$ is observable, where $\breve{H}_i = [H_i^\top, \sum_{j\in\mathcal{N}_i} H_j^\top]^\top$.

The observability mentioned in the above assumption suggests that the sensor network is regionally observable, which is weaker than that in the existing studies [22,23]. Under Assumption 1, the error covariance of the estimator (7) converges to a unique steady-state value with arbitrary initial state [24]. Moreover, the gain matrices that lead to the minimum mean square error are given by

$$\begin{aligned}\big[K_{i,k}, G_{i,k}\big] &= \breve{P}_{i,k}\breve{H}_i^\top(\breve{H}_i\breve{P}_{i,k}\breve{H}_i^\top + \breve{R}_i)^{-1}, \\ \breve{R}_i &= \mathrm{Diag}\Big[R_i, \sum_{j\in\mathcal{N}_i} R_j\Big], \\ \breve{P}_{i,k+1} &= A\breve{P}_{i,k}A^\top + Q - A\big[K_{i,k}, G_{i,k}\big]\breve{H}_i\breve{P}_{i,k}A^\top.\end{aligned} \tag{8}$$

Without loss of generality, we assume that the error dynamics have achieved a steady state at $k = 0$. Therefore, the subscripts for $K_{i,k}$ and $G_{i,k}$ are omitted in the rest of the paper.

2.2 Integrity attack

Since only a portion of the sensor data is encrypted, a malicious attacker could tamper with the unencrypted data to threaten the security of the sensor network. Dolev et al. [25] introduced an attack model that characterizes an adversary's capability to intercept and tamper with data as it is transmitted over the network. Inspired by this, we establish the following assumptions regarding the adversary's capabilities.

(1) The adversary has knowledge of the system's parameters A, H_i, K_i and G_i.

(2) The adversary is able to tamper with any unencrypted data.

In light of (4), the compromised measurement is modeled as the following form:

$$y_{i,k}^a = y_{i,k} + (I - \eta_i)\phi_{i,k}, \tag{9}$$

where $\eta_i = \mathrm{Diag}(\eta_i^{(1)}, \ldots, \eta_i^{(m_i)})$, and $\phi_{i,k}$ is the attack vector designed by the adversary.

Remark 1. As the adversary is not aware of the encryption key, tampering with the data indiscriminately increases the risk of being detected by security measures. Hence, a sophisticated adversary should avoid tampering with the encrypted data without knowing the secret key of the encryption mechanism to prevent unpredictable decryption results from undermining the stealthiness of the attack. This setting follows similar assumptions to those in [12,18], in which only a subset of the sensors' data within the system is considered trustworthy.

2.3 Anomaly detector

It is well known that the χ^2 detector is a widely used detector for disclosing anomalies in the system [26, 27]. Its essence is to discriminate the authenticity of the data based on the fact that normal data exhibits a stationary distribution. Driven by this motivation, a residual-based detector is employed in each sink node to verify the integrity of the measurements. Define $\varepsilon_{i\ell,k} \triangleq y_{\ell,k} - H_\ell \hat{x}_{i,k}$, where $\ell \in \mathcal{N}_i \cup \{i\}$. The detected quantity in the sink node i is given by

$$\zeta_{i\ell,k} \triangleq \sum_{h=k-c+1}^{k} \varepsilon_{i\ell,h}^\top \Omega \varepsilon_{i\ell,h} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \gamma, \tag{10}$$

where $c > 0$ is the detection interval, $\Omega \succ 0$ is the weight matrix, and γ is the detection threshold. At each time step, the anomaly detector identifies the authenticity of the measurements sent by the neighboring nodes. The hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ are used to represent whether the data has been tampered with. Specifically, $\zeta_{i\ell,k} \leqslant \gamma$ indicates that the measurement $y_{\ell,k}$ is normal, while $\zeta_{i\ell,k} > \gamma$ signifies that it has been tampered with.

2.4 Problem of interests

In this paper, we consider a more realistic scenario where only a portion of the sensor data is protected. Such a scenario complicates the analysis of attack modeling and the assessment of network security. In view of this, we aim to investigate the following problems.

Problem 1. In the case where only part of the sensor data is susceptible to tampering, what are the conditions under which an integrity attack remains stealthy and leads to estimation error divergence?

Problem 2. If the adversary is unaware of the detection threshold, what tactics will it adopt to stay as stealthy as possible in the attack?

Problem 3. With limited costs and budgets, how to determine which sensors need to be protected to balance data integrity and resource efficiency?

3 Main results

To enhance the relevance of defense measures, it is imperative to study the formulation and implementation of attack strategies from the perspective of an adversary. In this section, we first survey the conditions for the existence of integrity attacks and derive an explicit form of the feasible attack sequences. Subsequently, we investigate deployment schemes for encrypted sensors to ensure the security of distributed systems.

3.1 Feasibility of integrity attacks

In what follows, we focus on the vulnerability of sensor networks with partial protection. The evolution of the distributed recursive estimator under integrity attack is given by

$$\begin{aligned} \hat{x}_{i,k+1}^* = {} & A\hat{x}_{i,k}^* + AK_i\big(y_{i,k}^a - H_i\hat{x}_{i,k}^*\big) \\ & + AG_i \sum_{j\in\mathcal{N}_i} \big(y_{j,k}^a - H_j\hat{x}_{i,k}^*\big), \end{aligned} \tag{11}$$

where $\hat{x}_{i,k}^*$ is the estimate induced by compromised measurements $y_{i,k}^a$ and $y_{j,k}^a$. Accordingly, the detected quantity induced by the attack in sink node i is

$$\zeta_{i\ell,k}^* \triangleq \sum_{h=k-c+1}^{k} \varepsilon_{i\ell,h}^{*\top} \Omega \varepsilon_{i\ell,h}^*, \tag{12}$$

where $\varepsilon_{i\ell,h}^* = y_{\ell,h}^a - H_\ell \hat{x}_{i,h}^*$. Letting $e_{i,k}^* \triangleq x_k - \hat{x}_{i,k}^*$, the estimation error under integrity attack is expressed as

$$\begin{aligned} e_{i,k+1}^* = {} & \Big(A - AK_iH_i - AG_i \sum_{j\in\mathcal{N}_i} H_j\Big)e_{i,k}^* + w_k \\ & - AK_i(I-\eta_i)\phi_{i,k} - AG_i \sum_{j\in\mathcal{N}_i} (I-\eta_j)\phi_{j,k} \\ & - AG_i \sum_{j\in\mathcal{N}_i} \nu_{j,k} - AK_i\nu_{i,k}. \end{aligned} \tag{13}$$

To quantify the performance of integrity attacks, the definitions regarding perfection and stealthiness of the attack are introduced as follows [28, 29].

Definition 1 (Perfection)**.** The integrity attack in (9) is perfect if the following condition holds:

$$\lim_{k\to\infty} \|e^*_{i,k}\| \to \infty,\ i \in \{1,\dots,N\}. \tag{14}$$

Definition 2 (Stealthiness)**.** The integrity attack in (9) is stealthy if the following condition is satisfied:

$$\left\|\tilde{\varepsilon}_{i\ell,k}\right\| \leqslant \gamma',\ \ell \in \mathcal{N}_i \cup \{i\}, \tag{15}$$

where $\tilde{\varepsilon}_{i\ell,k} \triangleq \varepsilon^*_{i\ell,k} - \varepsilon_{i\ell,k}$, and $\gamma' > 0$ is a sufficiently small constant. In particular, the integrity attack is strictly stealthy if $\gamma' = 0$.

The above two definitions reflect two important attributes required for integrity attacks. The concept of perfection encapsulates the effectiveness of an attack in compromising data integrity. A perfect integrity attack can lead to a complete loss of accuracy and credibility of the data, which in the case of state estimation is manifested as the divergence of the estimation error. Stealthiness characterizes the capability of the attack in evading anomaly detectors. A smaller deviation in the residuals indicates that the false data does not elicit a significant change in the detected quantity. Stealthy attacks can be particularly challenging to defend against because they may not trigger alarms or raise suspicion until significant damage has been done.

Remark 2. The aforementioned definitions align with existing studies [13, 14], in which the authors consider a system insecure if the attack is both perfect and stealthy. In addition to perfect attacks, there are also non-perfect attacks whose bias on the estimates is bounded. Comparatively, perfect attacks are deemed more destructive as they can result in unpredictable consequences for the system.

Theorem 1. Under the anomaly detector (10), an integrity attack cannot simultaneously satisfy both perfection and strict stealthiness in sensor networks with partial protection.

Proof. By integrating all $e^*_{i,k}$ into a compact form, the estimation error for the whole network under integrity attack is given by

$$\begin{aligned} e^*_{k+1} = &\big(I\otimes A - \Lambda\bar{H} - \Upsilon\mathrm{Diag}(\mathcal{A}_i\otimes I)(I\otimes H)\big)e^*_k \\ &- \Lambda(I-\bar{\eta})\phi_k - \Upsilon(\mathcal{A}\otimes I)(I-\bar{\eta})\phi_k \\ &- \Lambda\nu_k - \Upsilon(\mathcal{A}\otimes I)\nu_k + \mathbf{1}w_k, \end{aligned} \tag{16}$$

where

$$\begin{aligned} &e^*_k = \left[e^{*\top}_{1,k},\dots,e^{*\top}_{N,k}\right]^\top, && \nu_k = \left[\nu^\top_{1,k},\dots,\nu^\top_{N,k}\right]^\top, && \Lambda = \mathrm{Diag}(AK_i), \\ &\Upsilon = \mathrm{Diag}(AG_i), && \bar{H} = \mathrm{Diag}(H_i), && \bar{\eta} = \mathrm{Diag}(\eta_i), \\ &H = [H_1^\top,\dots,H_N^\top]^\top, && \mathcal{A}_i = [a_{i1},\dots,a_{iN}], && \phi_k = [\phi^\top_{1,k},\dots,\phi^\top_{N,k}]^\top. \end{aligned}$$

Let the deviation between the estimation errors e^*_k and e_k be $\tilde{e}_k \triangleq e^*_k - e_k$. In view of (16), one has

$$\begin{aligned} \tilde{e}_{k+1} = &\big(I\otimes A - \Lambda\bar{H} - \Upsilon\mathrm{Diag}(\mathcal{A}_i\otimes I)(I\otimes H)\big)\tilde{e}_k \\ &- \Lambda(I-\bar{\eta})\phi_k - \Upsilon(\mathcal{A}\otimes I)(I-\bar{\eta})\phi_k. \end{aligned} \tag{17}$$

To streamline later analysis, we further define

$$\begin{aligned} &\epsilon_k \triangleq \left[\tilde{\varepsilon}^\top_{11,k},\dots,\tilde{\varepsilon}^\top_{NN,k}\right]^\top,\ \epsilon'_k \triangleq \left[\epsilon'^\top_{1,k},\dots,\epsilon'^\top_{N,k}\right]^\top, \\ &\epsilon'_{i,k} \triangleq \left[\epsilon'^\top_{i1,k},\dots,\epsilon'^\top_{iN,k}\right]^\top,\ \epsilon'_{ij,k} \triangleq \begin{cases} \tilde{\varepsilon}_{ij,k}, & \text{if } j\in\mathcal{N}_i, \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{18}$$

Based on the above definitions, it is derived that

$$\begin{aligned} &\epsilon_k = \bar{H}\tilde{e}_k + (I-\bar{\eta})\phi_k, \\ &\epsilon'_k = \mathrm{Diag}(\bar{\mathcal{A}}_i\otimes I)(I\otimes H)\tilde{e}_k + \mathrm{Diag}(\bar{\mathcal{A}}_i\otimes I)\big(\mathbf{1}\otimes(I-\bar{\eta})\big)\phi_k, \end{aligned} \tag{19}$$

where $\bar{\mathcal{A}}_i = \mathrm{Diag}(a_{i1}, \ldots, a_{iN})$. Substituting (19) into (17) yields

$$\begin{aligned}\tilde{e}_{k+1} &= (I \otimes A)\tilde{e}_k - \Lambda\epsilon_k - \Upsilon\mathrm{Diag}(\mathcal{A}_i \otimes I)\epsilon'_k \\ &= \sum_{r=0}^{k}(I \otimes A)^{k-r}\big(-\Lambda\epsilon_r - \Upsilon\mathrm{Diag}(\mathcal{A}_i \otimes I)\epsilon'_r\big),\end{aligned} \tag{20}$$

where $\tilde{e}_0 = 0$. When $\tilde{\varepsilon}_{i\ell,k} = 0$ for all $i \in \mathcal{V}$ and $\ell \in \mathcal{N}_i \cup \{i\}$, it follows that $\epsilon_k = \epsilon'_k = 0$. In this case, we can deduce that $\tilde{e}_k = 0$, $\forall k > 0$. Overall, an integrity attack cannot be guaranteed to be both perfect and strictly stealthy in sensor networks with partial protection. The proof is completed.

Theorem 1 demonstrates the infeasibility of executing a perfect attack with strict stealthiness in networks where only a subset of sensors is protected. It is noteworthy that the threshold of anomaly detectors is typically not set too low to prevent undue sensitivity to noise in detection residuals. Consequently, a sophisticated adversary may still be able to undermine the system by designing false data sequences that achieve a general level of stealth. In light of this, we further study whether non-strictly stealthy attacks exist in partially protected networks when $\gamma' > 0$.

Define $\mathcal{D}$ as the sum of the adjacency matrix $\mathcal{A}$ and the identity matrix I, where d_{ij} represents the (i, j)-th entry of $\mathcal{D}$, for all $i, j \in \{1, \ldots, N\}$. Assume that the system matrix A possesses v distinct eigenvalues, ordered such that $|\lambda_1| \leqslant |\lambda_2| \leqslant \cdots \leqslant |\lambda_v|$. Before proceeding further, let us introduce the Jordan canonical form of the system matrix $I \otimes A$:

$$J = P(I \otimes A)Q, \tag{21}$$

where $J = \mathrm{Diag}\{J_1, \ldots, J_v\}$, $P = [P_1, \ldots, P_v]$ and $Q = [Q_1, \ldots, Q_v]$. In (21), J_ι, $\iota \in \varkappa \triangleq \{1, \ldots, v\}$, is the Jordan block associated with the eigenvalue λ_ι, P is the matrix consisting of the eigenvectors and generalized eigenvectors of the matrix $I \otimes A$, and $Q = P^{-1}$. The following theorem presents the necessary conditions for the existence of non-strictly stealthy attacks in sensor networks with partial protection.

Theorem 2. If there exists an integrity attack that is both perfect and stealthy in networks with partial sensor protection, then the following conditions are satisfied.

(1) The spectral radius of A is not less than 1, i.e., $\rho(A) \geqslant 1$.

(2) $\exists\bar{\vartheta} \in \Psi([Q_q, \ldots, Q_v])$ such that $\mathrm{Diag}(\mathcal{D}_i \otimes I)(I \otimes H)\bar{\vartheta} \in \mathrm{Col}(\mathbf{1} \otimes (I - \bar{\eta}))$, where $\mathcal{D}_i = \mathrm{Diag}(d_{i1}, \ldots, d_{iN})$ and $q = \min\{\iota \in \varkappa | 1 \leqslant |\lambda_\iota|\}$.

Proof. We prove the above theorem by contradiction. First, it is assumed that the integrity attack is perfect and stealthy and that $\rho(A) < 1$. Based on the definition of stealthiness, it can be derived that ϵ_k and ϵ'_k are norm-bounded, i.e., $\|\epsilon_k\| \leqslant \gamma_1$ and $\|\epsilon'_k\| \leqslant \gamma_2$, where γ_1 and γ_2 are positive constants. In view of (20), we can obtain

$$\begin{aligned}\lim_{k\to\infty}\|\tilde{e}_k\| \leqslant \lim_{k\to\infty}\sum_{r=0}^{k-1}\|I \otimes A\|^{k-r-1}\big(&\gamma_1\|\Lambda\| \\ &+ \gamma_2\|\Upsilon\mathrm{Diag}(\mathcal{A}_i \otimes I)\|\big).\end{aligned} \tag{22}$$

Based on Gelfand's formula, if all eigenvalues in the matrix A satisfy $|\lambda_\iota| < 1$, then we deduce that

$$\lim_{k\to\infty}\|\tilde{e}_k\| \leqslant \frac{\upsilon\big(\gamma_1\|\Lambda\| + \gamma_2\|\Upsilon\mathrm{Diag}(\mathcal{A}_i \otimes I)\|\big)}{1 - \rho(A)} < \infty, \tag{23}$$

where $\upsilon > 0$ is a constant. Since the error dynamics without attack is asymptotically stable, the boundedness of $\|\tilde{e}_k\|$ in (23) suggests that the integrity attack is not perfect, which contradicts the aforementioned assumption. Therefore, it can be concluded that there exists at least one eigenvalue of the system matrix A such that $|\lambda_\iota| \geqslant 1$.

Next, we demonstrate that there exists no attack that is both perfect and stealthy if condition (2) does not hold. Since the system matrix A has at least one eigenvalue with absolute value greater than 1, there exists a positive integer q such that

$$|\lambda_1| \leqslant \cdots \leqslant |\lambda_{q-1}| < 1 \leqslant |\lambda_q| \leqslant \cdots \leqslant |\lambda_v|. \tag{24}$$

According to (24), the system matrix A can be expressed in its Jordan canonical form as follows:

$$I \otimes A = QJP = \begin{bmatrix}\breve{Q} & \bar{Q}\end{bmatrix}\begin{bmatrix}\breve{J} & \\ & \bar{J}\end{bmatrix}\begin{bmatrix}\breve{P} \\ \bar{P}\end{bmatrix}, \tag{25}$$

where

$$\breve{Q}=\big[Q_1,\ldots,Q_{q-1}\big],\qquad \bar{Q}=\big[Q_q,\ldots,Q_v\big],\qquad \breve{P}=\big[P_1,\ldots,P_{q-1}\big],$$
$$\bar{P}=\big[P_q,\ldots,P_v\big],\qquad \breve{J}=\mathrm{Diag}\big(J_1,\ldots,J_{q-1}\big),\qquad \bar{J}=\mathrm{Diag}\big(J_q,\ldots,J_v\big).$$

Based on (25), one has

$$\begin{aligned}\lim_{k\to\infty}\sum_{r=0}^{k}J^rP\xi_r&=\lim_{k\to\infty}\sum_{r=0}^{k}\begin{bmatrix}\breve{J}^r\breve{P}\\ \bar{J}^r\bar{P}\end{bmatrix}\xi_r\\&=\infty\cdot P\varsigma+P\varsigma',\end{aligned}\tag{26}$$

where $\xi_r=-\Lambda\epsilon_r-\Upsilon\mathrm{Diag}(\mathcal{A}_i\otimes I)\epsilon'_r$, and ς and ς' are constant vectors. It follows from $\rho(\breve{J})<1$ that $\sum_{r=0}^{\infty}\|\breve{J}^r\breve{P}\xi_r\|$ is bounded. Therefore, $\sum_{r=0}^{\infty}\|\bar{J}^r\bar{P}\xi_r\|\to\infty$ holds if $\sum_{r=0}^{\infty}\|J^rP\xi_r\|\to\infty$. The above deduction implies that $\bar{P}\varsigma\neq 0$ and $\breve{P}\varsigma=0$. Since $\bar{P}\varsigma\neq 0$, there exists a nonzero vector b such that $\bar{P}\varsigma=b$. By premultiplying the matrix $\bar{Q}$ to both sides of the aforementioned equation, we obtain

$$\bar{Q}b=\varsigma.\tag{27}$$

Eq. (27) shows that ς belongs to the column space of $\bar{Q}$, i.e., $\varsigma\in\mathrm{Col}(\bar{Q})$. Given the preceding analysis, the vector ς can be represented as a linear combination of the column vectors of the matrix $\bar{Q}$, expressed formally as

$$\varsigma=\sum_{r=1}^{z}c_r\vartheta_r,\tag{28}$$

where z is the number of columns in matrix $\bar{Q}$, ϑ_r is the r-th column of the matrix $\bar{Q}$, and c_r, $\forall r\in\{1,\ldots,z\}$, are constants that are not all zero. Combining (19), (26) and (28) yields

$$\begin{aligned}\lim_{k\to\infty}\epsilon_k&=\infty\cdot\bar{H}\sum_{r=1}^{z}c_r\vartheta_r+(I-\bar{\eta})\phi_k+\varsigma',\\\lim_{k\to\infty}\epsilon'_k&=\infty\cdot\mathrm{Diag}(\bar{\mathcal{A}}_i\otimes I)(I\otimes H)\sum_{r=1}^{z}c_r\vartheta_r\\&\quad+\mathrm{Diag}(\bar{\mathcal{A}}_i\otimes I)\big(\mathbf{1}\otimes(I-\bar{\eta})\big)\phi_k+\varsigma'.\end{aligned}\tag{29}$$

If $\mathrm{Diag}(\mathcal{D}_i\otimes I)(I\otimes H)\vartheta_r\notin\mathrm{Col}(\mathbf{1}\otimes(I-\bar{\eta}))$ for any $\vartheta_r\in\Psi(\bar{Q})$, then one has $\lim_{k\to\infty}\|\epsilon_k\|\to\infty$ and $\lim_{k\to\infty}\|\epsilon'_k\|\to\infty$. In this case, we can further deduce that $\lim_{k\to\infty}\|\tilde{\varepsilon}_{i\ell,k}\|\to\infty$, which contradicts the stealthiness in Definition 2. The proof is complete.

Theorem 2 outlines the necessary conditions for an adversary to perform an integrity attack in sensor networks with partial protection. In general, the first condition is determined by calculating the spectral radius of the system matrix A. The second condition is assessed by judging whether $\mathrm{Rank}(\mathbf{1}\otimes(I-\bar{\eta}))$ is equal to $\mathrm{Rank}([\mathbf{1}\otimes(I-\bar{\eta}),\mathrm{Diag}(\mathcal{D}_i\otimes I)(I\otimes H)\bar{\vartheta}])$. If any of the conditions in Theorem 2 is not satisfied, it implies that the sensor network with partial protection is capable of resisting integrity attacks. Conversely, satisfaction of these conditions may provide an adversary with the opportunity to launch integrity attacks that evade detection by the anomaly detector and lead to divergence of estimation errors. Following this revelation, we extend our investigation to explore the explicit forms these attacks may devise and their potential impacts.

To ensure that the attack is perfect and stealthy, the false data sequence is designed in the following form:

$$\phi_{i,k}=-\varphi_i\alpha^k\varrho_i,i\in\mathcal{V},\tag{30}$$

where $\varphi_i\in\mathbb{R}$, $\alpha\in\mathbb{R}$ and $\varrho_i\in\mathbb{R}^{m_i}$ are the constants and vectors to be determined, respectively. Intuitively, these parameters serve different purposes for achieving integrity attacks. To be specific, the scalar φ_i is designed to adjust the magnitude of the attack sequence. A small value of φ_i indicates that the perturbation in the data due to an integrity attack is relatively minor at each time step. The role of α^k is to achieve the perfectness of the attack, i.e., as k tends to infinity, the false data are able to drive the estimation error to diverge. The function of the ϱ_i is to realize the stealthiness of the attack, namely, it has to be carefully designed to ensure that the perfect attack is executed without alerting the detector. Based on the attack sequence in (30), we provide sufficient conditions for implementing a perfect and stealthy attack in partially protected networks through the following theorem.

Theorem 3. For the attack sequence presented in (30), an integrity attack in the partially protected sensor network is both perfect and stealthy if the following conditions hold:

(1) $\alpha \in \Xi(A) \triangleq \{\lambda_\iota \in \mathbb{R} || A - \lambda_\iota I| = 0,\ |\lambda_\iota| > 1\}$;

(2) $\exists \omega' \in \mathrm{Span}(\Theta(A))$ such that $\mathrm{Rank}(\mathbf{1} \otimes (I - \bar{\eta})) = \mathrm{Rank}([\mathbf{1} \otimes (I - \bar{\eta}), \mathrm{Diag}(\mathcal{D}_i \otimes I)(I \otimes H)\omega'])$, where $\Theta(A) \triangleq \{\omega | (I \otimes A)\omega = \alpha\omega\}$;

(3) $\max(|\varphi_i|) \leqslant \gamma' \|H\|^{-1} \|\omega'\|^{-1}$.

Proof. To facilitate the analysis, we set $\hbar_{k+1} \triangleq \tilde{e}_{k+1} - \tilde{e}_k$. Substituting (30) into (20), one has

$$\begin{aligned} \hbar_{k+1} = {} & \Phi \hbar_k + \Lambda(I - \bar{\eta})\mathrm{Diag}\big(\varphi_i(\alpha^k - \alpha^{k-1})I\big)\varrho \\ & + \Upsilon \mathrm{Diag}(\bar{\mathcal{A}}_i \otimes I)\big(\mathbf{1} \otimes (I - \bar{\eta})\big) \\ & \times \mathrm{Diag}\big(\varphi_i(\alpha^k - \alpha^{k-1})I\big)\varrho, \end{aligned} \tag{31}$$

where $\Phi = I \otimes A - \Lambda \bar{H} - \Upsilon \mathrm{Diag}(\mathcal{A}_i \otimes I)(I \otimes H)$ and $\varrho = \big[\varrho_1^\top, \ldots, \varrho_N^\top\big]^\top$. The second condition in Theorem 3 suggests that there exists a nonzero vector τ such that the following linear equation is consistent:

$$\big(\mathbf{1} \otimes (I - \bar{\eta})\big)\tau = \mathrm{Diag}(\mathcal{D}_i \otimes I)(I \otimes H)\omega'. \tag{32}$$

By solving (32) and setting $\varrho = \tau$, the incremental error $\hbar_{k+1}$ is given by

$$\begin{aligned} \hbar_{k+1} = {} & \Phi\big[\hbar_k - \mathrm{Diag}\big(\varphi_i(\alpha^k - \alpha^{k-1})I\big)\omega'\big] \\ & + \mathrm{Diag}\big(\varphi_i(\alpha^{k+1} - \alpha^k)I\big)\omega'. \end{aligned} \tag{33}$$

Noting that $\hbar_0 = 0$, Eq. (33) can be rewritten in the following power form:

$$\begin{aligned} \hbar_k = {} & \Phi^{k-1}\mathrm{Diag}(\varphi_i I)(I - \Phi)\omega' \\ & + \mathrm{Diag}\big(\varphi_i(\alpha^k - \alpha^{k-1})I\big)\omega'. \end{aligned} \tag{34}$$

Accordingly, the differential error $\tilde{e}_k$ induced by the integrity attack is expressed as

$$\tilde{e}_k = \sum_{r=1}^{k} \Big[\Phi^{r-1}\mathrm{Diag}(\varphi_i I)(I - \Phi) + \mathrm{Diag}\big(\varphi_i(\alpha^r - \alpha^{r-1})I\big)\Big]\omega'. \tag{35}$$

Since the error covariance of the estimator (7) converges to a unique steady-state value under Assumption 1, it follows that Φ is stable. In other words, the spectral radius of the matrix Φ is less than 1, i.e., $\rho(\Phi) < 1$. Hence, the limit of $\tilde{e}_k$ as k approaches ∞ is

$$\lim_{k\to\infty} \tilde{e}_k = \lim_{k\to\infty} \sum_{r=1}^{k} \mathrm{Diag}\big(\varphi_i(\alpha^r - \alpha^{r-1})I\big)\omega' + \mathrm{Diag}(\varphi_i I)\omega'. \tag{36}$$

If $\alpha \in \Xi(A)$, then we can deduce that

$$\begin{aligned} & \lim_{k\to\infty} \Big\| \sum_{r=1}^{k} \mathrm{Diag}\big(\varphi_i(\alpha^r - \alpha^{r-1})I\big)\omega' \Big\| \\ & = \big\| \mathrm{Diag}\big(\varphi_i(\alpha^\infty - 1)I\big)\omega' \big\| \\ & \to \infty. \end{aligned} \tag{37}$$

Applying the triangle inequality, one has

$$\begin{aligned} \lim_{k\to\infty} \|\tilde{e}_k\| \geqslant {} & \lim_{k\to\infty} \Big\| \sum_{r=1}^{k} \mathrm{Diag}\big(\varphi_i(\alpha^r - \alpha^{r-1})I\big)\omega' \Big\| \\ & - \|\mathrm{Diag}(\varphi_i I)\omega'\| \\ \to {} & \infty. \end{aligned} \tag{38}$$

Algorithm 1 Integrity attack in partially protected sensor networks.

Require: System parameters A, H and $\mathcal{A}$, and encrypted index $\bar{\eta}$;
Ensure: Attack vectors $\phi_{i,k}$, $\forall i \in \mathcal{V}$;
Based on (24), determine all the eigenvalues of the matrix A with absolute values greater than 1, i.e., λ_ι, $\iota \in \{p, \ldots, v\}$;
for $\iota \in \{p, \ldots, v\}$ **do**
  Calculate the generalized solution for $(I \otimes A - \lambda_\iota I)\omega = 0$, where ω is the solution vector;
  Construct ω' based on $\mathrm{Span}(\Theta(A))$;
  if $\mathrm{Rank}(\mathbf{1} \otimes (I - \bar{\eta})) = \mathrm{Rank}([\mathbf{1} \otimes (I - \bar{\eta}), \mathrm{Diag}(\mathcal{D}_i \otimes I)(I \otimes H)\omega'])$ **then**
    Solve the linear equations $\big(\mathbf{1} \otimes (I - \bar{\eta})\big)\tau = \big[\mathbf{1} \otimes (I - \bar{\eta}), \mathrm{Diag}(\mathcal{D}_i \otimes I)(I \otimes H)\omega'\big]$;
    Set $\alpha = \lambda_\iota$, $\varrho = \tau$;
    Determine φ_i based on condition (3) in Theorem 3;
    Set $\phi_{i,k} = -\varphi_i \alpha^k \varrho_i$;
    break
  end if
end for

Additionally, we define $\sigma_k \triangleq \epsilon_k - \epsilon_{k-1}$ and $\sigma'_k \triangleq \epsilon'_k - \epsilon'_{k-1}$, respectively. It follows from (19) and (32) that

$$\begin{aligned} \sigma_k &= \bar{H}\Phi^{k-1}\mathrm{Diag}(\varphi_i I)(I - \Phi)\omega', \\ \sigma'_k &= \mathrm{Diag}(\bar{\mathcal{A}}_i \otimes I)(I \otimes H)\Phi^{k-1}\mathrm{Diag}(\varphi_i I)(I - \Phi)\omega'. \end{aligned} \tag{39}$$

Based on (30), we can infer that

$$\begin{aligned} \epsilon_0 &= -\bar{H}\mathrm{Diag}(\varphi_i I)\omega', \\ \epsilon'_0 &= \mathrm{Diag}(\bar{\mathcal{A}}_i \otimes I)(I \otimes H)\mathrm{Diag}(\varphi_i I)\omega'. \end{aligned} \tag{40}$$

Combining (39) and (40) leads to

$$\begin{aligned} \epsilon_k &= \bar{H}\sum_{r=1}^{k}\Phi^{r-1}\mathrm{Diag}(\varphi_i I)(I - \Phi)\omega' - \bar{H}\mathrm{Diag}(\varphi_i I)\omega', \\ \epsilon'_k &= \mathrm{Diag}(\bar{\mathcal{A}}_i \otimes I)(I \otimes H)\Big(-\sum_{r=1}^{k}\Phi^{r-1}\mathrm{Diag}(\varphi_i I) \\ &\quad \times (I - \Phi)\omega' + \mathrm{Diag}(\varphi_i I)\omega'\Big). \end{aligned} \tag{41}$$

By means of the properties of the norm inequality, we have

$$\begin{aligned} \|\epsilon_k\| &\leqslant \Big\|\bar{H}\sum_{r=1}^{k}\Phi^{r-1}\mathrm{Diag}(\varphi_i I)(I - \Phi)\omega'\Big\| + \|\bar{H}\mathrm{Diag}(\varphi_i I)\omega'\| \\ &\leqslant 2\max(|\varphi_i|)\|\bar{H}\|\|\omega'\|. \end{aligned} \tag{42}$$

Similarly, the upper bound of $\|\epsilon'_k\|$ is given by

$$\|\epsilon'_k\| \leqslant \max(|\varphi_i|)\|H\|\|\omega'\|. \tag{43}$$

Since $\|\bar{H}\| = \max(\|H_1\|, \ldots, \|H_N\|)$ and $H^\top H = H_1^\top H_1 + \cdots + H_N^\top H_N$, it is easy to derive that $\|\bar{H}\| \leqslant \|H\|$. Consequently, the attack will be stealthy if the following inequality holds:

$$\max|\varphi_i| \leqslant \gamma'(\|H\|\|\omega'\|)^{-1}. \tag{44}$$

The proof is complete.

Algorithm 1 provides the steps to implement a perfect and stealthy attack in a partially protected sensor network. The attack sequence is evidently generated in an offline manner, implying that its construction does not necessitate reliance on real-time measurements. Such an attack is easier to execute with lower computational and communication overheads compared to the attack algorithm described in [10], which requires online generation.

In summary, Theorems 1–3 as well as Algorithm 1 provide a comprehensive and definitive response to Problem 1. These findings reveal that a distributed network, despite partial encryption of sensor data, remains vulnerable to integrity attacks when certain specific conditions are satisfied. According to Definition 2, the key to carrying out stealthy integrity attacks is to accurately determine the magnitude of the γ'. Thus, an interesting question is whether a partially protected sensor network is secure if the detection threshold is kept secret from the adversary. This question is actually formulated as Problem 2 in Section 2.4. By the following theorem, we show that an attack sequence of the form (30) is still able to maintain a certain degree of stealthiness.

Theorem 4. If condition (2) of Theorem 3 is satisfied, then the residual deviation $\tilde{\varepsilon}_{i\ell,k}$ decays exponentially to zero for all $i \in \mathcal{V}$ and $\ell \in M_i$ as k increases.

Proof. The sum of the first k terms of Φ^{k-1} is given by

$$\sum_{r=1}^{k} \Phi^{r-1} = (I - \Phi^k)(I - \Phi)^{-1}. \tag{45}$$

The second equal sign holds because the matrix $I - \Phi$ is positive definite and invertible. Substituting (45) into (41) yields

$$\begin{aligned} \epsilon_k &= -\bar{H}\mathrm{Diag}(\varphi_i I)\Phi^k \omega', \\ \epsilon'_k &= \mathrm{Diag}(\bar{\mathcal{A}}_i \otimes I)(I \otimes H)\Phi^k \omega'. \end{aligned} \tag{46}$$

Based on the analysis of Theorem 3, it follows that $\rho(\Phi) < 1$. Hence, the limits of ϵ_k and ϵ'_k as k approaches infinity are

$$\begin{aligned} \lim_{k\to\infty} \epsilon_k &= \lim_{k\to\infty} -\bar{H}\mathrm{Diag}(\varphi_i I)\Phi^k \omega' \\ &= 0, \\ \lim_{k\to\infty} \epsilon'_k &= \mathrm{Diag}(\bar{\mathcal{A}}_i \otimes I)(I \otimes H)\Phi^k \omega' \\ &= 0. \end{aligned} \tag{47}$$

From (18), we observe that $\tilde{\varepsilon}_{i\ell,k}$ decays exponentially to zero as k increases. The proof is complete.

According to Theorem 4, the deviations of the residuals induced by the integrity attack decay exponentially over time. In other words, the proposed attack sequence in (30) enables the alarm rate to gradually converge to the false positive rate, which is superior to the one in [13] in terms of stealthiness. Furthermore, it can be deduced that the larger the detection interval, the smaller the deviation of the detected quantity caused by the attack sequence. Thus, even if the adversary is not aware of the detection threshold, it can damage the estimation system while maintaining stealth by virtue of the attack sequence presented in (30).

3.2 Countermeasures to integrity attacks

The previous results analyze the security vulnerabilities that exist in partially protected sensor networks. To address Problem 3, we focus on determining the selection of sensor data for encryption to safeguard the distributed network against integrity attacks. To this end, we first determine the sufficient conditions under which an adversary cannot perform a perfect and stealthy attack.

Theorem 5. Perfect integrity attacks with stealthiness cannot be achieved in partially protected networks if $\mathrm{Diag}(\mathcal{D}_i \otimes I)(I \otimes H)\varpi \notin \mathrm{Nul}(\mathbf{1} \otimes \bar{\eta})$ for all $\varpi \in \Psi([Q_q, \ldots, Q_v])$.

Proof. The theorem can be proven by contradiction. First, we assume that a perfect and stealthy attack can be implemented on a partially protected sensor network even if $\mathrm{Diag}(\mathcal{D}_i \otimes I)(I \otimes H)\varpi \notin \mathrm{Nul}(\mathbf{1}^\top \otimes \bar{\eta})$ for all $\varpi \in \Psi([Q_q, \ldots, Q_v])$. By the definition of $\eta_i^{(s)}$ in (4), it can be derived that

$$\bar{\eta}(I - \bar{\eta}) = 0. \tag{48}$$

Recalling Theorem 2, one of the prerequisites for a perfect and stealthy attack to be feasible in a partially protected sensor network is that $\mathrm{Diag}(\mathcal{D}_i \otimes I)(I \otimes H)\varpi \in \mathrm{Col}(\mathbf{1} \otimes (I - \bar{\eta}))$, where $\varpi \in \Psi([Q_q, \ldots, Q_v])$. This condition implies that there exists a nonzero vector κ such that

$$\big(\mathbf{1} \otimes (I - \bar{\eta})\big)\kappa = \mathrm{Diag}(\mathcal{D}_i \otimes I)(I \otimes H)\varpi. \tag{49}$$

Combining (48) and (49) yields

$$\big(\mathbf{1} \otimes \bar{\eta}\big)^\top \mathrm{Diag}(\mathcal{D}_i \otimes I)(I \otimes H)\varpi = 0. \tag{50}$$

Eq. (50) shows that $\mathrm{Diag}(\mathcal{D}_i \otimes I)(I \otimes H)\varpi \in \mathrm{Nul}(\mathbf{1}^\top \otimes \bar{\eta})$, which contradicts the previous assumption. The proof is complete.

Theorem 5 reflects that the execution of perfect and stealthy attacks depends on the binary variable $\bar{\eta}$, which indicates whether the sensor data are encrypted. In order to protect the system from integrity attacks, the parameter $\bar{\eta}$ needs to be carefully designed to meet the criteria specified in Theorem 5. For this purpose, we model the scheme for determining the encrypted sensors as the following optimization problem:

$$\begin{aligned} (\mathcal{P}_1): \ & \arg\min_{\bar{\eta}} \ \mathrm{Tr}(\bar{\eta}) \\ \text{s.t. } & (\mathbf{1}^\top \otimes \bar{\eta})\beta\varpi \neq 0, \ \forall \varpi \in \Psi(\bar{Q}), \\ & \eta_i^{(s)} \in \{0,1\}, \ \forall i \in \mathcal{V} \text{ and } s \in M_i, \end{aligned} \tag{51}$$

where $\beta = \mathrm{Diag}(\mathcal{D}_i \otimes I)(I \otimes H)$. Notice that $\bar{\eta}$ is a diagonal matrix with diagonal entries that are either 0 or 1. Minimizing the trace of $\bar{\eta}$ is equivalent to minimizing the number of sensors that are to be encrypted. Therefore, the solution to the above optimization problem identifies the minimal set of sensors whose data require protection.

Given that the problem involves constraints on binary variables, the solution space is discrete since the decision variables are required to take values of either 0 or 1. Consequently, this problem is classified as a mixed-integer programming (MIP) problem. The discreteness and combinatorial nature of MIP problems result in high complexity, often requiring the exploration of many possible combinations to find an optimal solution. Solving this type of problem typically requires specialized algorithms such as heuristic methods, branch-and-bound, or branch-and-cut. Utilizing the Gurobi optimizer in MATLAB, the above optimization problem can be solved efficiently. In summary, the results in Section 3.2 provide a comprehensive and definitive response to Problem 3. These findings provide a practical strategy for improving system efficiency while ensuring the security of sensor networks under integrity attacks.

Remark 3. There are various types of data transmitted in distributed state estimation, such as measurement, estimate, and innovation. It is worth noting that the results obtained in Section 3 can be extended to the case where the transmission data are estimates or innovations, and this will be our future work.

4 Simulation

In this section, a numerical example is presented to verify the theoretical results. Consider a four-degree-of-freedom aircraft model with the system dynamics represented as follows [30]:

$$\begin{aligned} \dot{V} &= -g\mu + \frac{P_{max}\delta_T^c}{m \cdot 100} - \frac{1}{m}\Big(qS_aC_{D_0} + km^2g^2\frac{n_y^2}{qS_a}\Big), \\ \dot{\mu} &= \frac{g}{V} \cdot n_y^c - \frac{g}{V}, \\ \dot{\theta} &= \frac{g}{V} \cdot n_y \cdot \gamma^c, \end{aligned} \tag{52}$$

where V, μ, θ, δ_T^c, n_y^c and γ^c are the velocity, track angle, track azimuth, normal overload, roll angle, and thrust of the aircraft, respectively. The model is linearized with

$$A = \begin{bmatrix} 1.25 & 1 & 0 \\ 0 & 0.9 & 7 \\ 0 & 0 & 0.6 \end{bmatrix}. \tag{53}$$

It can be calculated that the unstable eigenvalue of the system matrix A is 1.125. Hence, the value of α is set to 1.125. To monitor the state of the system, there are 6 sink nodes that collect their sensor measurements and communicate with other sink nodes. The communication topology of the sink nodes is illustrated in Figure 2, and the other system parameters are selected as follows:

$$\begin{aligned} & Q = 0.1I, \ R = 0.15I, \ \breve{P}_{i,0} = I, \ \Omega = I, \\ & c = 1, \ \varphi_i = 0.5, \ \bar{M} = 400. \end{aligned} \tag{54}$$

Without loss of generality, all simulation results are derived from 400 Monte Carlo trials. It is assumed that the adversary starts attacking at $k = 0$. To quantify the impact of integrity attacks on the estimation performance, the

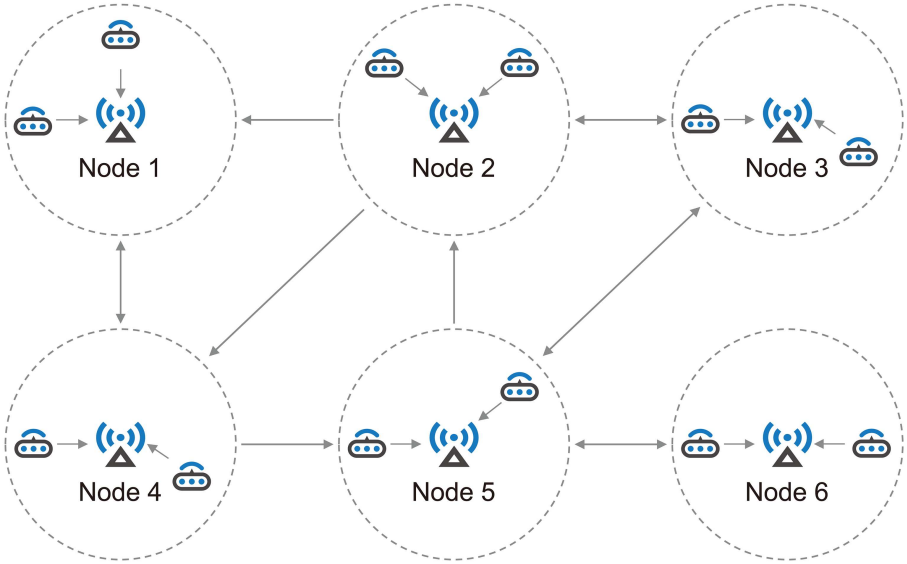

Figure 2 (Color online) Topology of sink nodes.

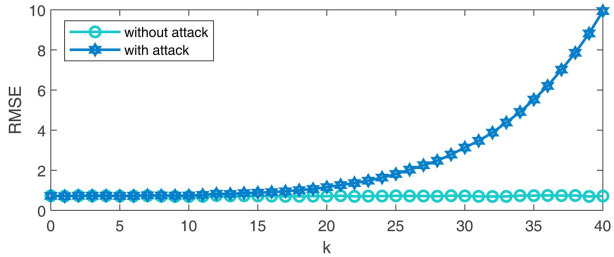

Figure 3 (Color online) Comparison of RMSEs for estimators with and without integrity attacks.

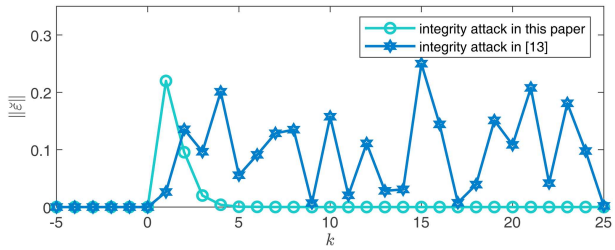

Figure 4 (Color online) Evolutions of $\breve{\varepsilon}_k$ under integrity attacks in this paper and in [13].

root mean square error (RMSE) is introduced as follows:

$$\mathrm{RMSE}(k) = \sqrt{\frac{1}{\bar{M}} \sum_{r=1}^{\bar{M}} \|x_k^{(r)} - \bar{x}_k^{(r)}\|^2}, \tag{55}$$

where $\bar{x}_k^{(r)} = \frac{1}{N} \sum_{i=1}^{N} \hat{x}_{i,k}^{(r)}$, and the superscript (r) denotes the r-th Monte Carlo trial. Figure 3 presents a comparison of the RMSE of the estimators with and without integrity attacks. In the absence of attacks, the RMSE is observed to be relatively low and stable, which indicates that the estimator is accurate and reliable under normal conditions. Conversely, the RMSE of the estimator under integrity attack exhibits a significant and rapid increase, suggesting that the proposed attack algorithm is a perfect attack strategy. To assess the stealthiness of the attack sequence, the following average residual deviation is introduced:

$$\breve{\varepsilon}_k = \frac{1}{N + \bar{m}} \sum_{i=1}^{N} \sum_{\ell \in M_i \cup \{i\}} \tilde{\varepsilon}_{i\ell,k}, \quad \forall i \in \mathcal{V}, \tag{56}$$

where $\bar{m} = \sum_{i=1}^{N} m_i$.

Figure 4 exhibits the evolution of $\breve{\varepsilon}_k$ under the attack scheme proposed in this paper as well as the one in [13]. It can be seen that the deviation in the residual induced by the proposed attack sequence decays exponentially, demonstrating superior stealthiness compared to the existing work in [13]. In addition, Figure 5 illustrates the trend of the average residual deviation for different values of φ_i. A smaller value of φ_i is more conducive to reducing the deviation of the residual induced by the integrity attack. Consequently, a malicious adversary can evade the anomaly detector by strategically selecting a smaller value for φ_i. Figure 6 presents the relationship between the

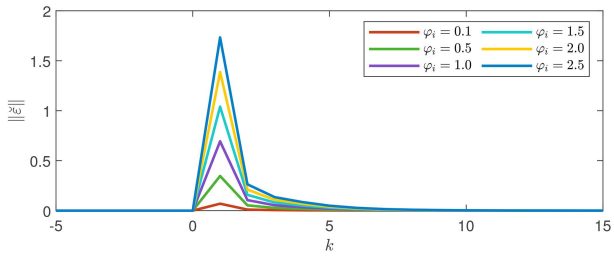

Figure 5 (Color online) Average residual deviation $\breve{\varepsilon}_k$ across different parameter values φ_i.

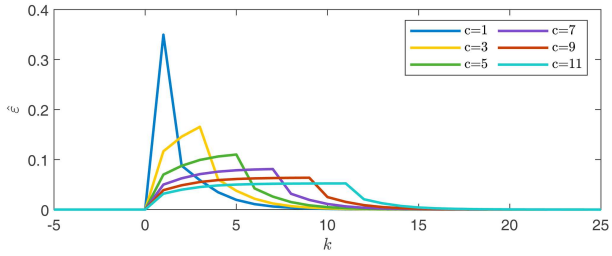

Figure 6 (Color online) Moving average residual deviation for different values of detection interval c.

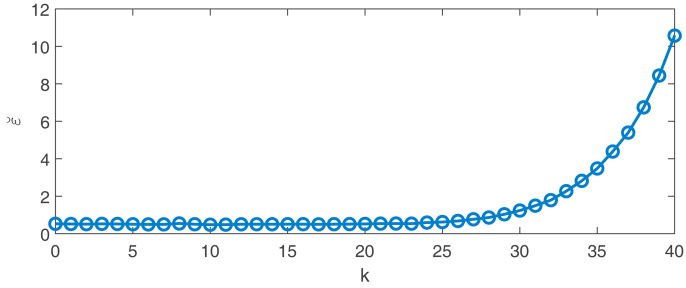

Figure 7 (Color online) Average residual deviation changes after protecting vulnerable nodes in the network.

detection interval and the stealthiness of the attack, where the moving average residual deviation $\hat{\varepsilon}_k$ is defined as follows:

$$\hat{\varepsilon}_k = \frac{1}{N+\bar{m}} \sum_{i=1}^{N} \sum_{\ell \in M_i \cup \{i\}} \Big(\frac{1}{c} \sum_{\ell=k-c+1}^{k} \|\tilde{\varepsilon}_{i\ell,h}\| \Big). \tag{57}$$

Intuitively, the moving average residual deviation caused by an attack decreases as the detection interval increases. In the domain of anomaly detection, the detection interval is commonly set to values greater than 1 to reduce the likelihood of false positives attributable to noise interference. Nevertheless, this characteristic is exploited by adversaries to augment the stealthiness of their attacks. The larger the detection interval, the less likely it is that an integrity attack will be detected. In summary, the above results reveal how an adversary can capitalize on vulnerabilities in a partially protected network to devise perfect and stealthy attack sequences to infringe on the estimation performance of the network. To withstand such integrity attacks, it is imperative to safeguard the vulnerable sensors within the network. By solving the optimization problem in (51), the sensors to be protected are determined as $y_{1,k}^{(1)}$ and $y_{5,k}^{(1)}$. Accordingly, the binary variables indicating whether the sensor is encrypted or not are given by

$$\eta_i^{(s)} = \begin{cases} 1, & i \in \{1,5\},\ s = 1, \\ 0, & \text{otherwise.} \end{cases} \tag{58}$$

As shown in Figure 7, the average residual $\bar{\varepsilon}$ under integrity attack increases dramatically when the data from vulnerable sensors are protected, which suggests that the attacker is unable to remain stealthy while launching a perfect attack. It is evident that safeguarding the data of two vulnerable sensors results in more effective security than the existing approach, which involved protecting all sensors. In conclusion, the proposed algorithm offers a cost-effective solution by strategically protecting only vulnerable nodes, thereby outperforming the indiscriminate protection of numerous nodes.

5 Conclusion

In this paper, the vulnerability of selective protection strategies in sensor networks has been investigated. The feasibility of strict stealth and non-strict stealth integrity attacks against selective protection strategies has been analyzed. The necessary conditions and sufficient conditions for an integrity attack that leads to unbounded estimation errors and evades anomaly detectors have been established. Moreover, an explicit form of integrity attacks has been designed, and an offline algorithm for generating attack sequences has been proposed. In response

to the threat of integrity attacks, a sufficient condition has been established to ensure the security of sensor networks. Additionally, a defense scheme has been provided for determining which sensors need protection to address sensor security under resource constraints. In future work, we will investigate encryption algorithms that protect data integrity for distributed state estimation.

Acknowledgements This work was supported in part by National Natural Science Foundation of China (Grant Nos. 62402130, 62202114, 62372129, U2436208), Postdoctoral Fellowship Program of China Postdoctoral Science Foundation (Grant Nos. 2024T170194, GZC20240322), Guangdong S&T Program (Grant No. 2024B0101010002), Project of Guangdong Key Laboratory of Industrial Control System Security (Grant No. 2024B1212020010), and Fundamental Research Funds for the Central Universities.

References

1 Leitao P, Karnouskos S, Ribeiro L, et al. Smart agents in industrial cyber-physical systems. Proc IEEE, 2016, 104: 1086–1101
2 Ding D R, Han Q-L, Ge X H, et al. Privacy-preserving filtering, control and optimization for industrial cyber-physical systems. Sci China Inf Sci, 2025, 68: 141201
3 Li X-M, Zou T, Lu R Q, et al. Event-based non-fragile state estimation for time-varying systems under deception attacks. Sci China Inf Sci, 2025, 68: 142204
4 Primadianto A, Lu C N. A review on distribution system state estimation. IEEE Trans Power Syst, 2017, 32: 3875–3883
5 Chen W, Wang Z, Ding D, et al. Distributed state estimation over wireless sensor networks with energy harvesting sensors. IEEE Trans Cybern, 2023, 53: 3311–3324
6 Fu C C, Zhang H, Huang C, et al. Secure cooperative output regulation for linear parameter-varying systems under DoS attacks: a resilient observer approach. Sci China Inf Sci, 2023, 66: 190203
7 Nozari E, Tallapragada P, Cortés J. Differentially private average consensus: obstructions, trade-offs, and optimal algorithm design. Automatica, 2017, 81: 221–231
8 An L W, Yang G-H. Attack detectability and stealthiness in distributed optimal coordination of cyber-physical systems. Sci China Inf Sci, 2023, 66: 199204
9 Niu M, Wen G, Lv Y, et al. Innovation-based stealthy attack against distributed state estimation over sensor networks. Automatica, 2023, 152: 110962
10 Park G, Lee C, Shim H, et al. Stealthy adversaries against uncertain cyber-physical systems: threat of robust zero-dynamics attack. IEEE Trans Automat Contr, 2019, 64: 4907–4919
11 Choraria M, Chattopadhyay A, Mitra U, et al. Design of false data injection attack on distributed process estimation. IEEE Trans Inform Forensic Secur, 2022, 17: 670–683
12 Xu H, Yang Y, Shang J, et al. Integrity attacks on remote estimation with spatial-temporal information sources. Automatica, 2023, 155: 111172
13 Hu L, Wang Z, Han Q L, et al. State estimation under false data injection attacks: security analysis and system protection. Automatica, 2018, 87: 176–183
14 Ni Y, Guo Z, Mo Y, et al. On the performance analysis of reset attack in cyber-physical systems. IEEE Trans Automat Contr, 2020, 65: 419–425
15 Jin K, Ye D. False-data injection attacks against distributed filters with improved complete-stealthiness. IEEE Trans Ind Inf, 2024, 20: 10344–10353
16 Xiao H C, Ding D R, Dong H L, et al. Adaptive event-triggered state estimation for large-scale systems subject to deception attacks. Sci China Inf Sci, 2022, 65: 122207
17 Kircher A, Bako L, Blanco É, et al. Analysis of resilience for a state estimator for linear systems. In: Proceedings of 2020 American Control Conference (ACC), 2020. 1495–1500
18 Li Y, Shi L, Chen T. Detection against linear deception attacks on multi-sensor remote state estimation. IEEE Trans Control Netw Syst, 2018, 5: 846–856
19 Zhou J, Yang W, Ding W, et al. Watermarking-based protection strategy against stealthy integrity attack on distributed state estimation. IEEE Trans Automat Contr, 2023, 68: 628–635
20 Suo Y, Chai S, Chai R, et al. Security defense of large-scale networks under false data injection attacks: an attack detection scheduling approach. IEEE Trans Inform Forensic Secur, 2024, 19: 1908–1921
21 Zhang Z, Zhang W, Qin Z, et al. A secure channel established by the PF-CL-AKA protocol with two-way ID-based authentication in advance for the 5G-based wireless mobile network. In: Proceedings of 2021 IEEE Asia Conference on Information Engineering (ACIE), 2021. 11–15
22 Liu Q, Wang Z, He X, et al. On Kalman-consensus filtering with random link failures over sensor networks. IEEE Trans Automat Contr, 2018, 63: 2701–2708
23 Zhou J, Yang W, Zhang H, et al. Security analysis and defense strategy of distributed filtering under false data injection attacks. Automatica, 2022, 138: 110151
24 Anderson B D O, Moore J B. Optimal Filtering. Englewood Cliffs: Prentice-Hall, 1979
25 Dolev D, Yao A. On the security of public key protocols. IEEE Trans Inform Theor, 1983, 29: 198–208
26 Weerakkody S, Liu X, Son S H, et al. A graph-theoretic characterization of perfect attackability for secure design of distributed control systems. IEEE Trans Control Netw Syst, 2017, 4: 60–70
27 Bhowmick C, Jagannathan S. Detection of sensor attacks in uncertain stochastic linear systems. In: Proceedings of 2019 IEEE Conference on Control Technology and Applications (CCTA), 2019. 706–711
28 Mo Y, Sinopoli B. False data injection attacks in control systems. In: Proceedings of the First Workshop on Secure Control Systems, 2010. 1–6
29 Sui T, Sun X M. The vulnerability of distributed state estimator under stealthy attacks. Automatica, 2021, 133: 109869
30 Qu F, Yang N, Fu J, et al. Stealthy attack on remote control system with local controller and its countermeasures. IEEE Trans Syst Man Cybern Syst, 2024, 54: 1147–1158