Secure State Estimation against Sparse Attacks on a Time-varying Set of Sensors [⋆]

Zishuo Li [*] **Muhammad Umar B. Niazi** [**,***] **Changxin Liu** [**]
Yilin Mo [*] **Karl H. Johansson** [**]

[*] *Department of Automation, Tsinghua University, Beijing, 100084, China (e-mail:* `lizs19@mails.tsinghua.edu.cn`, `ylmo@tsinghua.edu.cn`*)*
[**] *Division of Decision and Control Systems, Digital Futures, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm SE-100 44, Sweden (e-mail:* `changxin@kth.se`, `kallej@kth.se`*)*
[***] *Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA (e-mail:* `niazi@mit.edu`*)*

Abstract: This paper studies the problem of secure state estimation of a linear time-invariant (LTI) system with bounded noise in the presence of sparse attacks on an unknown, time-varying set of sensors. At each time, the attacker has the freedom to choose an arbitrary set of no more than p sensors and manipulate their measurements without restraint. To this end, we propose a secure state estimation scheme and guarantee a bounded estimation error irrespective of the attack signals subject to $2p$-sparse observability and a mild, technical assumption that the system matrix has no degenerate eigenvalues. The proposed scheme comprises a design of decentralized observers for each sensor based on the local observable subspace decomposition. At each time step, the local estimates of sensors are fused by a median operator to obtain a secure estimation, which is then followed by a local detection-and-resetting process of the decentralized observers. The estimation error is shown to be upper-bounded by a constant which is determined only by the system parameters and noise magnitudes. Moreover, we design the detector threshold to ensure that the benign sensors never trigger the detector. The efficacy of the proposed algorithm is demonstrated by its application on a benchmark example of IEEE 14-bus system. We show that our proposed scheme can effectively tolerate sparse attacks on an unknown set of sensors, ensuring a bounded estimation error and effectively detecting and resetting the attacked sensors.

Keywords: Estimation and filtering; Design of fault tolerant/reliable systems; Estimation and fault detection

1. INTRODUCTION

Cyber-physical systems tend to be vulnerable to intelligently devised malicious attacks because of the interconnectedness of remote sensors and unprotected cyber-infrastructures operating physical plants. This has caused significant concern in the research community (Cárdenas et al., 2009), who have highlighted the dire need for resilient/secure estimation and control algorithms.

In this paper, we consider the secure state estimation of an LTI system in the presence of attacks on an unknown subset of sensors, which may vary with time. However, the attacker is assumed to possess limited resources and, at each time instant, can corrupt at most p out of total m sensors by injecting arbitrary signals to their measurements. A milder scenario widely considered in the literature assumes that the attacked subset of sensors is fixed with respect to time. Thus, it is usually easier to detect attacks in such a scenario. The problem becomes significantly challenging when the attacked subset is time-varying, and, in this regard, the method proposed in this paper is quite fundamental.

There are three main research directions for tackling the problem of secure state estimation with a fixed subset of attacked sensors. The first is the sliding window method which formulates the state estimation task as a batch optimization problem by considering the past measurements in a finite time window. Under certain observability conditions, Fawzi et al. (2014) and Shoukry and Tabuada (2016) propose a method to minimize the norm of the residue signal for obtaining a secure state estimate. However, searching for the minimizer is of combinatorial nature because the corrupted sensor set is unknown. Shoukry et al. (2017)

employ Satisfiability Modulo Theory (SMT) to reduce the size of attacked sensor subset by pruning the optimization process. Nonetheless, the sliding window method discards history information by ignoring the measurements out of the considered window. The second method is based on estimator switching that maintains multiple parallel estimators based on measurements from sensor subsets with cardinality $m - p$. By assuming that the corrupted subset of sensors is fixed, this method guarantees the existence of at least one subset of sensors unaffected by the attack. Nakahira and Mo (2018) devise a detection scheme to select all trusted subsets of sensors and fuse the corresponding local estimates for secure estimation. However, maintaining a combinatorial number of estimators simultaneously introduces a heavy computational and storage burden. Mishra et al. (2017) replace the exhausted search of sensor subsets with SMT-based searching, which can improve the computational load to a great degree. Lu and Yang (2019) resort to the set cover approach to reduce the number of subset candidates by at least half. An and Yang (2017) adopt a sequential switching mechanism where the algorithm switches from a malicious subset to a trusted one until it converges to the benign subset of subsets. The third method is based on local decomposition-fusion that solves a combinatorial complexity problem by designing m decentralized estimators based on a centralized estimator and fusing the local estimates by a convex optimization problem (Liu et al., 2021b; Li and Mo, 2023). The optimal trade-off between security and estimation accuracy can be achieved by carefully choosing the optimization parameters.

Allowing attacks on a time-varying subset of sensors is more practical than the setup where the corrupted subset is fixed. However, the problem becomes significantly challenging in the following ways: (1) Historical knowledge on corrupted sensors becomes useless because the attacker can arbitrarily choose a totally different set at each time. (2) Computational complexity increases since the corrupted subset of sensors need to be searched at every time step. Because of the aforementioned challenges, the secure state estimation problem under time-varying attack set is not well-explored. In this regard, He et al. (2022) propose a saturation-gain filter where the gain of the innovation term is saturated if the residue signal exceeds the prescribed threshold. However, this comes at the cost of a restrictive assumption that the system is 1-step $2p$-sparse observable, i.e. [1], $\text{rank}(C_{\mathcal{I} \setminus \mathcal{C}}) = n, \forall |\mathcal{C}| = 2p$. An and Yang (2020) proposed a switching-based non-linear observer that asymptotically converges to a neighborhood of the true system state in the presence of time-varying sensor attacks. In this paper, we propose a decentralized observer scheme that has bounded worst-case estimation error in spite of time-varying attacked set, under $2p$-sparse observability. Our contribution can be summarized as follows:

- **Estimation performance:** We derive explicit estimation error bounds under p-sparse attack on an arbitrary time-varying subset of sensors.
- **Observability requirement:** Our method requires the usual $2p$-sparse observability condition. This assumption is fundamental in the literature that assumes a fixed subset of attacked sensors, and we show that no additional requirement is needed in the time-varying case.
- **Low online computational complexity:** As we employ reduced-order decentralized estimator design, we need less online computational and storage resources compared to Liu et al. (2021b).
- **Low offline computational complexity:** The $2p$-sparse observability requirement can be verified in polynomial time with respect to the number of sensors and the dimension of the system, under the assumption that the system matrix has no degenerate eigenvalues.

The rest of the paper is organized as follows. We formulate the problem of secure state estimation in Section 2 and describe the local observable subspace decomposition in Section 3. Then, in Section 4, we present the design of decentralized observers and their fusion scheme. In the same section, we present the overall algorithm and provide the performance guarantee (Theorem 4). In Section 5, a numerical example is demonstrated to corroborate the performance of the proposed method. Finally, Section 6 concludes the paper with a future outlook.

Notations: The set of positive integers is denoted by $\mathbb{Z}^+$. The n-dimensional real and complex vector spaces are denoted by $\mathbb{R}^n$ and $\mathbb{C}^n$, respectively. The notation $[x]_j$ specifies the j-th entry of a vector x. By $\text{rowspan}(A)$, we represent the linear span of the rows of matrix A. The conjugate transpose of a matrix A is A'. We use $\| \cdot \|_p$ to denote the p-norm of a vector or the induced p-norm of a matrix, which should be clear from the context. Finally, $\sigma_{\max}(\cdot)$ denotes the maximum singular value of a matrix.

2. PROBLEM FORMULATION

Consider the discrete-time LTI system described by
$$x(k + 1) = Ax(k) + w(k) \quad (1)$$
$$y(k) = Cx(k) + v(k) + a(k) \quad (2)$$
where $x(k) \in \mathbb{R}^n$ is the system's state, $w(k) \in \mathbb{R}^n, v(k) \in \mathbb{R}^m$ are the bounded process noise and measurement noise, respectively, with $\|w(k)\|_2 \leq \mathcal{B}_w$ and $\|v(k)\|_2 \leq \mathcal{B}_v$, for every $k \in \{0\} \cup \mathbb{Z}^+$. The vector $y(k) \in \mathbb{R}^m$ is the collection of measurements from all m sensors, and i-th entry $y_i(k)$ is the measurement from sensor i. The vector $a(k)$ denotes the sensor attack by an adversary and $a_i(k)$ represents the attack signal (malicious data) injected into sensor i's measurement at time k. We denote the set of sensors as $\mathcal{I} = \{1, 2, \cdots, m\}$. We consider the scenario where a time-varying subset of sensors is compromised by a malicious adversary. Let the support of vector $a \in \mathbb{R}^m$ be
$$\text{supp}(a) \triangleq \{i | 1 \leq i \leq m, a_i \neq 0\}.$$
Definition 1. (Time-varying sparse attack). The attack $a(k)$ is said to be a *time-varying p-sparse attack* if the compromised set $\mathcal{C}(k) \triangleq \text{supp}\{a(k)\}$ satisfies $|\mathcal{C}(k)| \leq p$ for all k. The set of all *admissible* time-varying p-sparse attacks is then defined as $\mathbb{A}_p$. $\diamond$

It is important to emphasize that the estimation scheme has no direct information about $\mathcal{C}(k)$, but it knows that the maximum number of corrupted sensors is p.

---

[1] The matrix $C_{\mathcal{I} \setminus \mathcal{C}}$ represents the matrix composed of rows of C with row index in $\mathcal{I} \setminus \mathcal{C}$.

Another notion that needs to be defined is that of sparse observability (Shoukry and Tabuada, 2016), which characterizes the redundancy of the system's observability.

Definition 2. (Sparse observability). The pair (A, C) is said to be *s-sparse observable* if the pair $(A, C_{\mathcal{I} \setminus \mathcal{C}})$ is observable [2] for any subset of sensors $\mathcal{C} \subset \mathcal{I}$ with cardinality $|\mathcal{C}| = s$.   ◇

The secure state estimation problem aims at recovering the system's state $x(k)$ at every time k with uniformly bounded error through sensor measurements, which might have been partly manipulated.

Definition 3. (Secure estimator). Denote $\hat{x}(k)$ as the estimate at time k. Then, an estimator is said to be *secure* against time-varying p-sparse attack if there exists a constant $\mathcal{B}_e \geq 0$ determined only by the noise magnitudes $\mathcal{B}_w, \mathcal{B}_v$ and system matrices A, C such that, for all time k and all attacks $a \in \mathbb{A}_p$,

$$\sup_{\|w\|_2 \leq \mathcal{B}_w, \|v\|_2 \leq \mathcal{B}_v} \|\hat{x}(k) - x(k)\|_2 \leq \mathcal{B}_e. \quad ◇$$

In this paper, we propose a secure state estimation scheme against time-varying p-sparse attacks, which only requires $2p$-sparse observability. This condition is less restrictive than the 1-step $2p$-sparse observability (He et al., 2022), and is necessary in the sense of exact state recovery when the system is noise-free (Shoukry and Tabuada, 2016).

3. PRELIMINARIES ON LOCAL OBSERVABLE SUBSPACE DECOMPOSITION

Before introducing our design of decentralized observers, we need some preliminaries on the decomposition of observable space into local observable subspace. In order to better represent and denote the structure of the system observable space, we consider the following assumption on the system matrix A.

Assumption 1. All the eigenvalues of matrix A have geometric multiplicity 1. Without loss of generality, we can assume that A is in the Jordan canonical form.   ◇

Remark 1. If some eigenvalues of A have geometric multiplicity greater than 1, then the observability structure becomes complicated, which makes expressing the observable space intractable, and validating the sparse observability (Definition 2) is NP-hard (Mao et al., 2022). Therefore, for secure state estimation, it is quite common to assume that A does not have degenerate eigenvalues (Lee et al., 2020).   ◇

The following assumption is needed for observability redundancy to resiliently recover the system state.

Assumption 2. The system (A, C) is $2p$-sparse observable.   ◇

Recall the index set of sensors $\mathcal{I} \triangleq \{1, 2, \cdots, m\}$. Define the index set of state entries as $\mathcal{J} \triangleq \{1, 2, \cdots, n\}$ and the observability matrix with respect to sensor i as

$$O_i \triangleq \begin{bmatrix} C_i^\top & (C_i A)^\top & \cdots & (C_i A^{n-1})^\top \end{bmatrix}^\top. \quad (3)$$

Then, the observable subspace of sensor i is

$$\mathbb{O}_i \triangleq \mathrm{rowspan}(O_i) = \mathrm{span}\left(C_i', (C_i A)', \cdots, (C_i A^{n-1})'\right).$$

---

[2] The matrix $C_{\mathcal{I} \setminus \mathcal{C}}$ represents the matrix composed of rows of C with row index in $\mathcal{I} \setminus \mathcal{C}$.

All the states that can be observed by the measurements from sensor i belong to the linear vector space $\mathbb{O}_i$. The dimension of $\mathbb{O}_i$ is denoted as n_i. The system's observable space is given by $\mathbb{O} \triangleq \bigcup_{i \in \mathcal{I}} \mathbb{O}_i$. Define $\mathbf{e}_j$ as the n-dimensional canonical basis vector with 1 on the j-th entry and 0 on all the other entries, then the following theorem characterizes the transformation between $\mathbb{O}$ and $\mathbb{O}_i$.

Theorem 1. For each sensor i, there exists a linear projection $\mathbb{O} \rightarrow \mathbb{O}_i$ represented by matrix H_i, where H_i is an $n_i \times n$ matrix, such that, for any arbitrary $x \in \mathbb{O}$, $H_i' H_i x = x$ if $x \in \mathbb{O}_i$, and $H_i' H_i x = \mathbf{0}$ if $x \in \mathbb{O} \setminus \mathbb{O}_i$.

Moreover, if Assumption 1 holds, H_i can be constructed as the following $n_i \times n$ matrix:

$$H_i = \begin{bmatrix} \mathbf{e}_{j_1} & \mathbf{e}_{j_2} & \cdots & \mathbf{e}_{j_{n_i}} \end{bmatrix}' \quad (4)$$

where $\{j_1, \cdots, j_{n_i}\} = \mathcal{Q}_i$ with

$$\mathcal{Q}_i \triangleq \{j \in \mathcal{J} \mid O_i \mathbf{e}_j \neq \mathbf{0}\}. \quad (5)$$

Proof. See Appendix A of Li et al. (2022).   □

Define

$$\tilde{A}_i \triangleq H_i A H_i' \in \mathbb{C}^{n_i \times n_i}, \quad \tilde{C}_i \triangleq C_i H_i' \in \mathbb{C}^{1 \times n_i} \quad (6)$$

where C_i is i-th row of matrix C. Theorem 1 provides an explicit formulation of the transformation H_i from $\mathbb{O}$ to $\mathbb{O}_i$. Using this model reduction defined by matrix H_i, we build a sub-system $(\tilde{A}_i, \tilde{C}_i)$ corresponding to $\mathbb{O}_i$ and local observers in the next section.

4. SECURE ESTIMATION DESIGN

In this section, we propose a design of secure estimation based on decentralized local state observers and local detectors as well as a resilient centralized fusing scheme by median operators.

4.1 Secure Estimation Algorithm

Our proposed algorithm is summarized in the following Algorithm 1 and Fig. 1. At each time instant, the algorithm is composed of three steps, i.e., (i) local observer state update; (ii) secure fusion; (iii) local detection and reset. As shown in Fig.1, step (i) and step (iii) are performed in a decentralized manner, and step (ii) fuses the local information to obtain a secure estimation by the median operator.

In step (i) (line 4 in the algorithm), the local state $\eta_i^+(k-1)$ is updated to $\eta_i(k)$ using sensor measurement $y_i(k-1)$ by (9). In step (ii) (line 5 in the algorithm), secure estimation $\hat{x}(k)$ is obtained by fusing all the local states $\eta_i(k)$ using the median operation in (10), where $\mathcal{F}_j$ is designed as the index set of sensors that can observe state j as follows:

$$\mathcal{F}_j \triangleq \{i \in \mathcal{I} \mid O_i \mathbf{e}_j \neq \mathbf{0}\}. \quad (7)$$

And the median operator is defined as

$$\mathrm{med}(z_i, i \in \mathcal{S}) \triangleq f_{(|\mathcal{S}|+1)/2}(z_i, i \in \mathcal{S}), \quad (8)$$

where operator $f_s(z_i, i \in \mathcal{S})$ equals to the s-th smallest element in the set $\{z_i | i \in \mathcal{S}\}$. In step (iii) (lines 6-10 in the algorithm), by local detection and resetting, state $\eta_i(k)$ is reset as $\eta_i^+(k)$ for the usage of next time step.

Even though we consider centralized estimation problems, the algorithm is designed in a decentralized manner to isolate the influence of corrupted sensors, i.e., compromised

Algorithm 1 Secure Estimation at sensor i

1: Offline parameters: matrices $H_i, L_i, \tilde{A}_i, \tilde{C}_i$, constants γ, n_i, sets $\mathcal{F}_j, j \in \mathcal{J}$
2: Initialize $\eta_i^+(0) := H_i \hat{x}(0)$
3: **for** every time index $k \in \mathbb{Z}^+$ **do**
4:    Update $\eta_i(k)$ by
$$\eta_i(k) := \left(\tilde{A}_i - L_i \tilde{C}_i \right) \eta_i^+(k-1) + L_i y_i(k-1) \quad (9)$$
5:    Calculate each entry of $\hat{x}(k)$ by
$$[\hat{x}(k)]_j = \mathrm{med} \left\{ \left[H_i^\top \eta_i(k) \right]_j , i \in \mathcal{F}_j \right\}, \ \forall j \in \mathcal{J} \quad (10)$$
6:    **if** $\|\eta_i(k) - H_i \hat{x}(k)\|_2 > (\sqrt{n_i} + 1)\gamma$ **then**
7:        $\eta_i^+(k) := H_i \hat{x}(k)$
8:    **else**
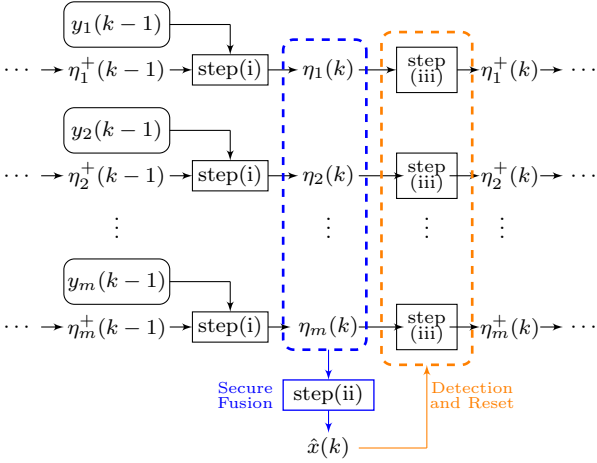9: $\eta_i^+(k) := \eta_i(k)$
10:    **end if**
11: **end for**

Fig. 1. The information flow of our proposed estimation scheme.

$y_i(k)$ will never influence the value of $\eta_j(k), j \neq i$. In this way, the combinatorial complex problem of finding the corrupted sensor set is simplified as a resilient fusion problem that can be solved in a computationally easy way, i.e., taking the median. The security of our proposed Algorithm 1 is proved in the next subsection.

4.2 Security of the Proposed Algorithm

In this subsection, we prove the security of the algorithm by mathematical induction.

Define the set of sensors whose estimation error is smaller than γ at time k as "accurate" sensor set $\mathcal{A}(k)$:
$$\mathcal{A}(k) \triangleq \{ i \in \mathcal{I} \mid \|\eta_i(k) - H_i x(k)\|_2 \leq \gamma \}.$$
Define good sensor set as those sensors who are free of attack at time k, i.e., $\mathcal{G}(k) \triangleq \mathcal{I} \setminus \mathcal{C}(k)$. We have the following theorem providing the effectiveness of local detection/resetting and resilient fusion algorithm.

Theorem 2. Suppose Assumption 1 and 2 are satisfied. Moreover, assume for each sensor i, estimator gain L_i and detector parameter γ satisfy the following inequality:
$$\sigma_{\max} \left(\tilde{A}_i - L_i \tilde{C}_i \right) \leq \frac{\gamma - \mathcal{B}_w - \|L_i\|_2 \mathcal{B}_v}{(2\sqrt{n_i} + 1)\gamma} \quad (11)$$

with $\gamma \geq \mathcal{B}_w + \|L_i\|_2 \mathcal{B}_v$. Then we have the following condition implication results:
$$|\mathcal{A}(k)| \geq m - p \Rightarrow (\text{s1}) \Rightarrow (\text{s2}) \Rightarrow (\text{s3})$$
where the three statements are:

(s1) Resilient fusion by median operator at time k is bounded:
$$\|\hat{x}(k) - x(k)\|_\infty \leq \gamma. \quad (12)$$
(s2) Accurate sensor set covers good sensor set (good sensors always provide accurate local estimate):
$$\mathcal{A}(k+1) \supset \mathcal{G}(k).$$
(s3) There will be at least the same number of accurate local estimates at time $k + 1$:
$$|\mathcal{A}(k+1)| \geq m - p. \quad (13)$$

Proof. See Appendix of Li et al. (2022). □

Remark 2. The statement (s3) means that our proposed detection algorithm has no false alarm or type-I error. Thus, no good sensor will be "wronged" and all benign local states $\eta_i(k+1), i \in \mathcal{G}(k)$, will be directly used in the fusion step at time $k+1$. This means that the detector threshold is suitable in the sense that no useful information is discarded unnecessarily. This property is crucial for maintaining observability redundancy. On the other hand, the detection algorithm may have miss detection or type-II error. Nevertheless, our proposed median-based fusion in (10) can tolerate the error introduced by undetected corrupted local states and generate secure estimation as shown in statement (s1). ◇

The intuition behind condition (11) is that, in order to achieve stable state estimate against time-varying attacked set, each local estimator i should be "stable" enough so that it immediately stops triggering the detector after the attacker has left sensor i.

We present a necessary and sufficient condition under which such γ, L_i satisfying (11) always exist.

Proposition 3. One can always find γ and L_i such that (11) holds if and only if the following holds:
$$\min_{L_i} \sigma_{\max}(\tilde{A}_i - L_i \tilde{C}_i) < \frac{1}{2\sqrt{n_i} + 1}, \quad \forall i \in \mathcal{I}.$$

By reformulating (11), one obtains that
$$\gamma \geq \frac{\mathcal{B}_w + \|L_i\|_2 \mathcal{B}_v}{1 - (2\sqrt{n_i} + 1)\|\tilde{A}_i - L_i \tilde{C}_i\|_2}. \quad (14)$$

The proof of Proposition 3 is straightforward from (14) and (11), and is omitted due to space limit. The condition required by Proposition 3 is easily satisfied because the pair $(\tilde{A}_i, \tilde{C}_i)$ is observable, which implies that the eigenvalues of $\tilde{A}_i - L_i \tilde{C}_i$ can be freely assigned.

As one may notice in Theorem 2, the result is provided in an induction-based way, i.e., $|\mathcal{A}(k)| \geq m - p \Rightarrow |\mathcal{A}(k+1)| \geq m - p$. In the following, by Theorem 2, we prove that the security of our propose estimation is guaranteed under time-varying p-sparse attacks as long as some initial condition is satisfied. This is the main result of the paper.

Theorem 4. Suppose Assumption 1-2 and the following conditions are satisfied:

(i) Inequality (11) holds.

(ii) Either the initial estimate satisfies $\|\hat{x}(0) - x(0)\|_\infty \leq \gamma$ or $|\mathcal{A}(0)| \geq m - p$.

Then, Algorithm 1 provides a secure state estimate $\hat{x}$ in the sense that

$$\|\hat{x}(k) - x(k)\|_\infty \leq \gamma, \quad \forall k \in \mathbb{Z}^+ \tag{15}$$

against any arbitrary time-varying p-sparse attack. Moreover, good sensors at time k will not trigger the detector at time $k+1$, i.e., $\forall i \in \mathcal{G}(k)$,

$$\|\eta_i(k+1) - H_i\hat{x}(k+1)\|_2 \leq (\sqrt{n_i} + 1)\gamma. \tag{16}$$

Proof. We prove (15) by mathematical induction. First, the condition (iii) makes sure $\|\hat{x}(0) - x(0)\|_\infty \leq \gamma$. By the induction structure of the result in Theorem 2, we know that estimation error bound (15) holds for all $k \in \mathbb{Z}^+$ from statement (s1), and inequality (16) holds from statement (s2). $\square$

The result in (15) can be interpreted as that the estimation error $\|\hat{x}(k) - x(k)\|_\infty$ is not worse than initial estimation error $\|\hat{x}(0) - x(0)\|_\infty$ under a time-varying set of compromised sensors. From a theoretical point of view, the result is non-trivial when the system is unstable, since the state estimate by pure prediction $\hat{x}(k) = A^k\hat{x}(0)$ has diverging error. Thus, a secure estimation algorithm must incorporate useful information from partly manipulated measurements and correct the state estimate to maintain a bounded error. From the practical perspective, the initial time $k = 0$ is seen as the time when the attack is launched and an estimator has stable estimation already. Our proposed algorithm would not worsen the estimation error upper bound compared to when there is no attack, as long as the bound does not exceed γ.

5. NUMERICAL SIMULATION

We apply our proposed estimation scheme to the IEEE 14-bus system, which is a benchmark example extensively used in the literature (Korres, 2010; Liu et al., 2021a) for illustrating the performance of secure state estimation algorithms. We adopt the continuous-time system dynamics as in the following equations (Wood et al., 2013):

$$\dot{\theta}_i(t) = \omega_i(t)$$

$$\dot{\omega}_i(t) = -\frac{1}{m_i}\left[D_i\omega_i(t) + \sum_{j \in \mathcal{N}_i} P_{tie}^{ij}(t) - P_i(t) + w_i(t)\right]$$

where $\theta_i(t)$ and $\omega_i(t)$ are the phase angle and angular frequency on bus i, respectively, m_i is the angular momentum of i, and w_i is the process disturbance. The parameter D_i is the load change sensitivity w.r.t. the frequency and is defined in Wood et al. (2013) Section 10.3. The power flow between neighboring buses i and j is given by $P_{tie}^{ij}(t) = -P_{tie}^{ji}(t) = t_{ij}(\theta_i(t) - \theta_j(t))$, where t_{ij} is the inverse of resistance between bus i and j. The power $P_i(t)$ denotes the difference of mechanical power and power demand, which is known to the system operator. The system is sampled at discrete times with sampling interval $T_s = 0.01$ s. Every bus is equipped with four sensors (one electric power sensor, one phase sensor and two angular velocity sensors). The process noise w_i and measurement noise v_i are i.i.d. uniformly distributed random vectors normalized to satisfy noise bounds $\mathcal{B}_w = 10^{-3}, \mathcal{B}_v = 10^{-2}$.

The attack is switching between all 14 buses' electrical power sensors at every time index k. The attacked bus index is switching as the following rule:

$$\text{corrupted bus index} = \begin{cases} \{1, 2, 3, 4\}, & \text{rem}(k, 3) = 1 \\ \{5, 6, 7, 8\}, & \text{rem}(k, 3) = 2 \\ \{9, 10, 11, 12, 13, 14\}, & \text{rem}(k, 3) = 0 \end{cases}$$

where $\text{rem}(k, 3)$ is the remaining of k divided by 3. For random signal attack, $a_i(k)$ is a random value uniformly distributed in interval $[-10, 10]$. For slope signal attack, $a_i(k) = k/5$, i.e., the magnitude of injected data increases by 20 for every second. The detector parameter γ is chosen to be 0.5.
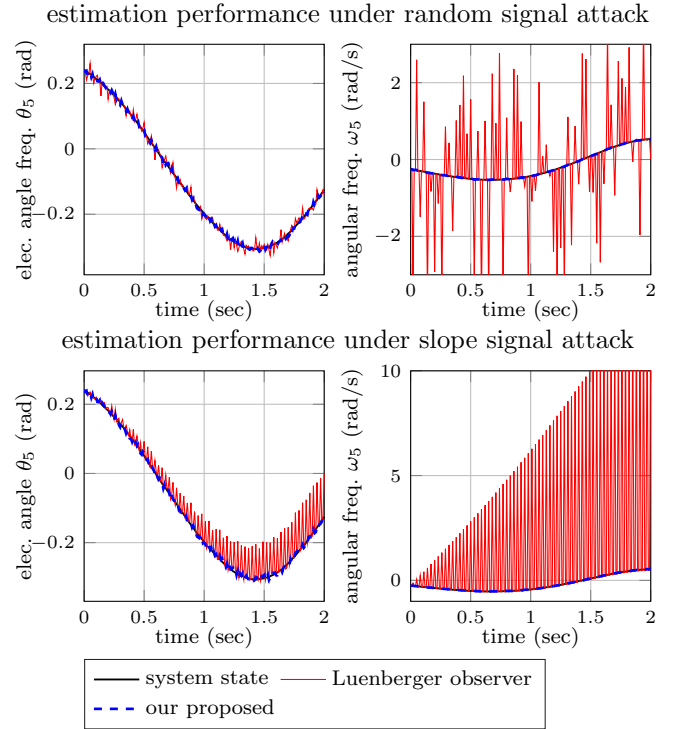

Fig. 2. Estimation of states under slope signal attack and random signal attack on IEEE 14 system. The top two sub-figures illustrate the phase angle and angular velocity of bus 5 under random signal attack. The bottom two sub-figures illustrate those under slope signal attack.

Fig. 2 demonstrates the estimation performance on bus 5 under two kinds of attacks mentioned before. The Luenberger observer used for the comparison is $\hat{x}(k+1) = A\hat{x}(k) + \sum_{i=1}^{m} H_i' L_i (y_i(k) - C_i\hat{x}(k))$. Since the electrical power measurement is directly related to angular frequency ω_i, the estimation of ω_5 is affected more significantly for insecure estimators such as Luenberger observer (shown in red line). In comparison, our proposed algorithm recovers the system state with a small error (shown in the blue dashed line).

To better illustrate the effectiveness of our proposed detection-and-reset mechanism in Algorithm 1 line 6-10, we show the residue norm $\|\eta_i(k) - H_i\hat{x}(k)\|$ before and after reset in Fig. 3. Only the first 30 samples are shown for presentation clearness. When the residue norm is larger than threshold $(\sqrt{n_i} + 1)\gamma = 2.081$ (blue triangle above the

red horizontal line), $\eta_i(k)$ is reset as $H_i\hat{x}(k)$ and the residue norm is decreased significantly (to be the blue circle). In Fig. 3, by this detection-and-reset mechanism, the residue holds stable, and the result of Theorem 4 is validated.
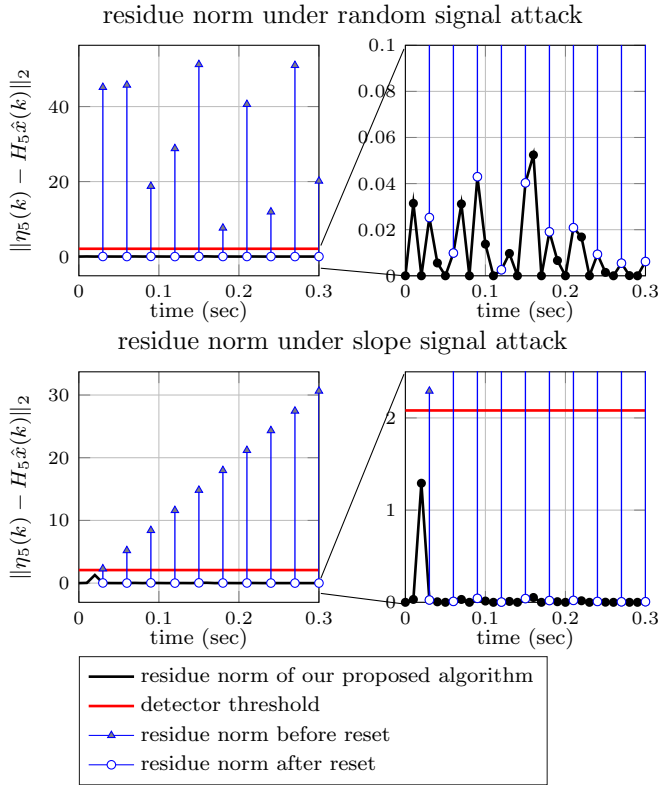
residue norm under random signal attack

residue norm under slope signal attack

Fig. 3. The residue norm $\|\eta_i(k) - H_i\hat{x}(k)\|_2$ before and after detection and resetting at bus $i = 5$. Sub-figure on the right is the zoomed-in plot of the left.

6. CONCLUDING REMARKS

By designing local estimators based on observable space decomposition and a detection-resetting mechanism, we achieved secure estimation against sparse attacks on time-varying sensor sets. The update of local states η_i as well as detection-reset operations only rely on local information. Thus, it will be straightforward to extend our method to a distributed sensor network by calculating (10) in a distributed manner. Moreover, our design only applies to system with the dynamic matrix A that has no degenerate eigenvalues. One direction to close this gap is to study observable subspace decomposition on the system where A has degenerate eigenvalues and design H_i as the canonical basis vector group of O_i, which is not necessarily composed of $\mathbf{e}_j$. Moreover, the fundamental limit of secure estimation under time-varying set attack is not well studied to our best knowledge and is left for future research.

REFERENCES

An, L. and Yang, G.H. (2017). Secure state estimation against sparse sensor attacks with adaptive switching mechanism. *IEEE Transactions on Automatic Control*, 63(8), 2596–2603.

An, L. and Yang, G.H. (2020). Supervisory nonlinear state observers for adversarial sparse attacks. *IEEE Transactions on Cybernetics*, 52(3), 1575–1587.

Cárdenas, A., Amin, S., Sinopoli, B., Giani, A., Perrig, A., Sastry, S., et al. (2009). Challenges for securing cyber physical systems. *Workshop on future directions in cyber-physical systems security*, 5(1).

Fawzi, H., Tabuada, P., and Diggavi, S. (2014). Secure estimation and control for cyber-physical systems under adversarial attacks. *IEEE Transactions on Automatic Control*, 59(6), 1454–1467.

He, X., Ren, X., Sandberg, H., and Johansson, K.H. (2022). How to secure distributed filters under sensor attacks. *IEEE Transactions on Automatic Control*, 67(6), 2843–2856.

Korres, G.N. (2010). A robust algorithm for power system state estimation with equality constraints. *IEEE Transactions on Power Systems*, 25(3), 1531–1541.

Lee, J.G., Kim, J., and Shim, H. (2020). Fully distributed resilient state estimation based on distributed median solver. *IEEE Transactions on Automatic Control*, 65(9), 3935–3942.

Li, Z. and Mo, Y. (2023). Efficient secure state estimation against sparse integrity attack for regular linear system. *International Journal of Robust and Nonlinear Control*, 33(1), 209–236.

Li, Z., Niazi, M.U.B., Liu, C., Mo, Y., and Johansson, K.H. (2022). Secure state estimation against sparse attacks on a time-varying set of sensors. *arXiv preprint arXiv:2211.05566*.

Liu, C., Liang, H., and Chen, T. (2021a). Network parameter coordinated false data injection attacks against power system ac state estimation. *IEEE Transactions on Smart Grid*, 12(2), 1626–1639.

Liu, X., Mo, Y., and Garone, E. (2021b). Local decomposition of kalman filters and its application for secure state estimation. *IEEE Transactions on Automatic Control*, 66(10), 5037–5044.

Lu, A.Y. and Yang, G.H. (2019). Secure switched observers for cyber-physical systems under sparse sensor attacks: A set cover approach. *IEEE Transactions on Automatic Control*, 64(9), 3949–3955.

Mao, Y., Mitra, A., Sundaram, S., and Tabuada, P. (2022). On the computational complexity of the secure state-reconstruction problem. *Automatica*, 136, 110083.

Mishra, S., Shoukry, Y., Karamchandani, N., Diggavi, S.N., and Tabuada, P. (2017). Secure state estimation against sensor attacks in the presence of noise. *IEEE Transactions on Control of Network Systems*, 4(1), 49–59.

Nakahira, Y. and Mo, Y. (2018). Attack-resilient $\mathcal{H}_2$, $\mathcal{H}_\infty$, and ℓ_1 state estimator. *IEEE Transactions on Automatic Control*, 63(12), 4353–4360.

Shoukry, Y. and Tabuada, P. (2016). Event-triggered state observers for sparse sensor noise/attacks. *IEEE Transactions on Automatic Control*, 61(8), 2079–2091.

Shoukry, Y., Nuzzo, P., Puggelli, A., Sangiovanni-Vincentelli, A.L., Seshia, S.A., and Tabuada, P. (2017). Secure state estimation for cyber-physical systems under sensor attacks: A satisfiability modulo theory approach. *IEEE Transactions on Automatic Control*, 62(10), 4917–4932.

Wood, A.J., Wollenberg, B.F., and Sheblé, G.B. (2013). *Power generation, operation, and control*. John Wiley & Sons.