How to Secure Distributed Filters Under Sensor Attacks

Xingkang He, *Member, IEEE*, Xiaoqiang Ren, *Member, IEEE*, Henrik Sandberg, *Member, IEEE*, and Karl Henrik Johansson, *Fellow, IEEE*

***Abstract*—In this article, we study how to secure distributed filters for linear time-invariant systems with bounded noise under false-data injection attacks. A malicious attacker is able to arbitrarily manipulate the observations for a time-varying and unknown subset of the sensors. We first propose a recursive distributed filter consisting of two steps at each update. The first step employs a saturation-like scheme, which gives a small gain if the innovation is large corresponding to a potential attack. The second step is a consensus operation of state estimates among neighboring sensors. We prove the estimation error is upper bounded if the filter parameters satisfy a condition. We further analyze the feasibility of the condition and connect it to sparse observability in the centralized case. When the attacked sensor set is known to be time-invariant, the secured filter is modified by adding an online local attack detector. The detector is able to identify the attacked sensors whose observation innovations are larger than the detection thresholds. Also, with more attacked sensors being detected, the thresholds will adaptively adjust to reduce the space of the stealthy attack signals. The resilience of the secured filter with detection is verified by an explicit relationship between the upper bound of the estimation error and the number of detected attacked sensors. Moreover, for the noise-free case, we prove that the state estimate of each sensor asymptotically converges to the system state under certain conditions. Numerical simulations are provided to illustrate the developed results.**

***Index Terms*—Attack detection, distributed state estimation, false-data injection attack, sensor attacks.**

Manuscript received November 21, 2020; accepted June 20, 2021. Date of publication June 25, 2021; date of current version May 31, 2022. This work was supported in part by National Key R&D Program of China under Grant 2018AAA0102804, in part by Knut & Alice Wallenberg Foundation, and in part by Swedish Research Council, Swedish Foundation for Strategic Research. Recommended by Associate Editor G. Gu. *(Corresponding author: Xiaoqiang Ren.)*

Xingkang He, Henrik Sandberg, and Karl Henrik Johansson are with the Division of Decision and Control Systems, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm 11428, Sweden, and also with Digital Futures, KTH Royal Institute of Technology, Stockholm 11428, Sweden (e-mail: xingkang@kth.se; hsan@kth.se; kallej@kth.se).

Xiaoqiang Ren is with the School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China (e-mail: xqren@shu.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TAC.2021.3092603.

Digital Object Identifier 10.1109/TAC.2021.3092603

I. Introduction

A CYBER-PHYSICAL SYSTEM (CPS) is a physical system controlled and monitored by computer-based algorithms. During recent years, numerous applications in sensor networks, vehicle networks, process control, smart grid, etc., have been investigated. With higher integration of large-scale computer networks and complex physical processes, these systems are confronting more security issues both in the cyber and physical layers. Thus, the research on the CPS security is attracting more and more attention.

Sensors and sensor networks are utilized to collect environmental data in a CPS. The quality of these sensors is essential for decision making. However, with the increasing number of complex tasks and the large-scale deployment of cheap and low-quality sensors, the vulnerability of system operation is inevitably increased. In this article, we consider the false-data injection (FDI) attacks in sensors networks, which is illustrated in Fig. 1, where a distributed sensor network with 30 sensors is deployed to collaboratively observe the state of a CPS. In this case, six sensors in red are under attack in the sense that their observations can be arbitrarily manipulated. We are interested to find a distributed filter to estimate the system state by employing the information provided by the sensor network in Fig. 1.

A large number of distributed filters for sensor networks have been proposed in the literature, e.g., [1] and [2]. These filters, however, would not work well in attack scenarios like Fig. 1. In order to degrade the filter performance, an attacker can strategically inject false data into the observations of attacked sensors based on its knowledge of systems. When probability distributions of the observations are affected, the filters prior-designed based on the distributions are no longer effective. For the scenario in Fig. 1, the following questions are answered in this article.

1) How to design a distributed filter such that it is resilient when the sensor network is under FDI attacks?
2) What is the maximal number of sensors under FDI attacks, such that filter stability is guaranteed?
3) How to detect which sensors are attacked and how to remove their influence on the filter performance?

A. Related Work

The security problems of CPSs have been extensively studied in the literature based on centralized frameworks, where a data center is able to collect and process the data from all sensors.

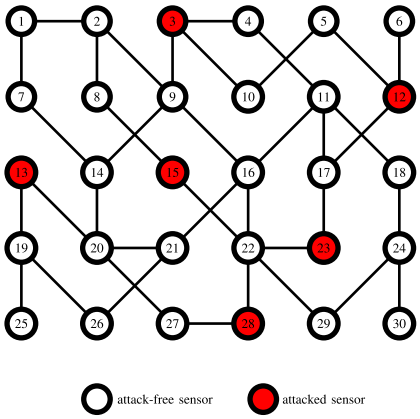


Fig. 1. Distributed sensor network under FDI attacks.

To find out whether sensors are under attack and to identify the attack signals inserted to the systems, a study on attack detection and identification for CPSs was given in [3], where the design methods and analysis techniques for centralized monitors were discussed as well. A probabilistic approach was given in [4] to estimate a static parameter in a fusion center under sparse FDI sensor attacks. To obtain attack-resilient state estimates, some centralized state estimators or observers were proposed based on optimization techniques [5]–[10], which usually face heavy computational complexity in the brute force search.

In comparison with centralized frameworks, distributed frameworks have no data center. Distributed methods rely on local computation and neighbor communication, thus they outweigh centralized methods in scalability for large networks and robustness to failures. In recent years, some investigations in the study of sensor networks under Byzantine attacks/failures have been made for the distributed state estimation of dynamical systems [11], [12], the distributed identification of a static vector parameter [13], and the distributed stochastic gradient descent [14]. Although these articles studied the worst sensor attacks (i.e., Byzantine attacks), they require complete connectivity or strong robustness of graphs, which would be quite restrictive for the systems suffering milder attacks (e.g., FDI attacks). In this direction, a distributed observer with attack detection was proposed to deal with a class of bias attacks in the observer update or sensor communication [15]. Distributed estimation for a static parameter under FDI sensor attacks was studied in [16] and [17]. In [18], a distributed optimization-based method was utilized to achieve convergence of the observer under sparse observability for linear time-invariant (LTI) systems [19] suffering FDI sensor attacks. Nevertheless, the results relied on the redesign of topology graph and infinite sensor communications between two observation updates. The authors in [20] studied the distributed dimensionality reduction fusion estimation for CPSs under denial-of-service attacks. In [21], for FDI attacks in communication networks, a distributed detection problem was studied for a group of interconnected subsystems, and an extended application to dc microgrids was given in [22]. In [23], a Bayesian framework-based joint distributed attack detection and state estimation were investigated in a cluster-based sensor network by considering FDI attacks in the communication between remote sensors and fusion nodes. However, the accurate probability distribution of attacks was required. To the knowledge of the authors, there were few results considering how to achieve the codesign of a distributed estimator and an attack detector.

B. Objectives and Contributions of this Article

In this article, we study the distributed state estimation problem for LTI systems with bounded noise over a sensor network, where the observations of a time-varying and unknown subset of sensors are arbitrarily manipulated by a malicious attacker through FDI attacks.

The objective of this article is fourfold.

1) Design a resilient distributed filter for each sensor with the potentially compromised observations and the data received from neighboring sensors.
2) Analyze the main properties of the filter, including the estimation error boundedness.
3) Design an attack detection-based filter if the compromised sensor set is known to be time invariant.
4) Analyze the main properties of the detection-based filter.

Corresponding to the four objectives, this article makes four contributions summarized in the following.

1) We design a secured distributed filter consisting of two steps (see Algorithm 1). The first step employs a saturation-like scheme, which gives a small gain if the innovation is large corresponding to a potential attack. The second step is a consensus operation of state estimates among neighboring sensors.
2) We investigate some properties of the secured filter. First, we prove that the estimation error is upper bounded if the filter parameters satisfy a condition, whose feasibility is studied by providing an easy-to-check sufficient and necessary condition (see Theorems 1 and 2). We further connect this condition to sparse observability in the centralized case (see Proposition 3). Moreover, we provide a condition such that the observations of the attack-free sensors will not be saturated after a finite time. Then, a tighter error bound is obtained (seecTheorem 3).
3) We modify the secured distributed filter by adding an attack detector (see Algorithm 2), when the set of attacked sensors is known to be time invariant. The detector is able to identify the attacked sensors whose observation innovations are larger than the detector thresholds (see Proposition 4). Moreover, with more attacked sensors being detected, the thresholds will adaptively adjust to reduce the space of the stealthy attack signals.
4) We study some properties of the secured filter with attack detection. First, the resilience of the filter is verified by an explicit relationship between the upper bound of the estimation error and the number of detected attacked sensors (see Theorem 4). Moreover, for the noise-free case, we prove that the state estimate of each sensor asymptotically converges to the system state under certain conditions (see Theorem 5).

This article designs a filter with an innovation-dependent update gain, essentially different from conventional filters with statistics-based gains (e.g., Kalman filter), in order to confine the influence of attack signals to the estimation. To handle the technical difficulties in performance analysis, a new tool inspired by bounded-input–bounded-output (BIBO) stability is provided to analyze boundedness of the estimation error. The distribution assumption on attack signals in [15] is removed in this article by allowing that the attacker can inject any attack signals. Moreover, the assumption that the attacked sensor set is fixed over time in both centralized frameworks [4]–[6], [8]–[10], [24] and distributed frameworks [11], [12], [18], [25] is extended to the time-varying case. The robustness requirement of communication graphs in [11] and [12] for a wider range of attacks and the requirement of infinite communication rate between two updates in [18] are both removed in this article. This article builds on the preliminary work presented in [26] and [27]. The main difference is fourfold. First, the set of the attacked sensors is extended from the time-invariant case to the time-varying case. Second, a new section dealing with attack detection and sensor isolation is added. Third, the results in [26] and [27] are generalized and new theoretical results with proofs are added. Fourth, more literature comparisons and simulation results are provided.

The remainder of the article is organized as follows. Section II is on the problem formulation. Section III provides the secured distributed filter and its performance analysis. The secured distributed filter with an online attack detector is studied in Section IV. After numerical simulations in Section V, Section VI concludes this article. The main proofs are given in Appendix.

Notations: $\mathbb{R}^n$ is the set of n-dimensional real vectors. $\mathbb{R}^+$ and $\mathbb{Z}^+$ are the sets of positive real scalars and integers, respectively. $\mathbb{R}^{n\times m}$ is the set of real matrices with n rows and m columns. $\mathrm{diag}\{\cdot\}$ represents the diagonalization operator. I_n stands for the n-dimensional square identity matrix. $\mathbf{1}_N$ stands for the N-dimensional vector with all elements being one. The superscript "T" represents the transpose. $A\otimes B$ is the Kronecker product of A and B. $\|x\|$ is the 2-norm of a vector x. $\|A\|$ is the induced 2-norm of matrix A, i.e., $\|A\|=\sup_{x\neq 0}\frac{\|Ax\|}{\|x\|}$. $\lambda_{\min}(A)$, $\lambda_2(A)$, and $\lambda_{\max}(A)$ are the minimum, second minimum, and maximum eigenvalues of a real-valued symmetric matrix A, respectively. $|\Gamma|$ is the cardinality of the set Γ.$\min\{a,b\}$ means the minimum between the real-valued scalars a and b. For a set $\mathcal{A}$, the indicator function $\mathbb{I}_{a\in\mathcal{A}}=1$, if $a\in\mathcal{A}$; $\mathbb{I}_{a\in\mathcal{A}}=0$, otherwise. $-plx-left-lceil\cdot\rceil$ is the ceiling function.

II. Problem Formulation

In this section, we first provide some graph preliminaries, and then, set up the problem of this article.

A. Graph Preliminaries

We model the communication topology of N sensors by an undirected graph $\mathcal{G}=(\mathcal{V},\mathcal{E})$ without self-loops, where $\mathcal{V}=\{1,2,\ldots,N\}$ stands for the set of nodes, and $\mathcal{E}\subseteq\mathcal{V}\times\mathcal{V}$ is the set of edges. If there is an edge $(j,i)\in\mathcal{E}$, the node i can exchange information with the node j, then the node j is called a neighbor of the node i, and vice versa. Let the neighbor set of the node i be $\mathcal{N}_i:=\{j\in\mathcal{V}|(j,i)\in\mathcal{E}\}$. The degree matrix of $\mathcal{G}$ is $D_{\mathcal{G}}=\mathrm{diag}\{|\mathcal{N}_1|,\ldots,|\mathcal{N}_N|\}$. The adjacency matrix is $\mathbb{A}_{\mathcal{G}}=[a_{i,j}]$, where $a_{i,j}=1$ if $(i,j)\in\mathcal{E}$, otherwise $a_{i,j}=0$. $\mathcal{L}=D_{\mathcal{G}}-\mathbb{A}_{\mathcal{G}}$ is the Laplacian matrix. Graph $\mathcal{G}$ is connected if for any pair of two different nodes i_1,i_l, there exists a path from i_1 to i_l consisting of edges $(i_1,i_2),(i_2,i_3),\ldots,(i_{l-1},i_l)$. On the connectivity of a graph, we have the following proposition.

Proposition 1 [28]: The undirected graph $\mathcal{G}$ is connected if and only if $\lambda_2(\mathcal{L})>0$.

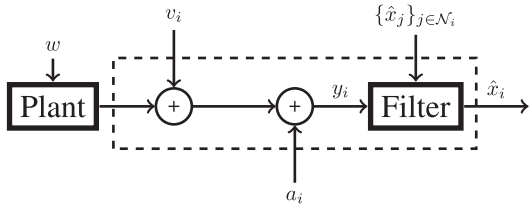


Fig. 2. Each sensor is equipped with a filter providing an estimate $\hat{x}_i$ of state x. The sensor observation y_i is potentially compromised through an attack signal a_i.

B. System Model

For a sensor network $\mathcal{G}$ under FDI attacks, we illustrate the scenario in Fig. 1, where each senor is equipped with a filter to estimate the system state (see Fig. 2 for a diagram). The state-space system model is given as follows:

$$\begin{aligned} x(t+1)&=Ax(t)+w(t)\\ y_i(t)&=C_ix(t)+v_i(t)+a_i(t), i=1,\ldots,N\end{aligned} \tag{1}$$

where $x(t)\in\mathbb{R}^n$ is the unknown system state, $w(t)\in\mathbb{R}^n$ the process noise, $v_i(t)\in\mathbb{R}$ the observation noise, $a_i(t)\in\mathbb{R}$ the attack signal inserted by some malicious attacker, and $y_i(t)\in\mathbb{R}$ the observation of sensor i, all at time t. Moreover, $A\in\mathbb{R}^{n\times n}$ is the system state transition matrix, and $C_i\in\mathbb{R}^{1\times n}$ is the observation vector of the sensor i. Both A and C_i are known to each sensor. We do not assume that (A,C_i) is observable. Without losing generality, we assume that the observation vectors are normalized, i.e., $\|C_i\|=1, i\in\mathcal{V}=\{1,\ldots,N\}$. Otherwise, we can reconstruct the observation equation of the system (1).

In this article, we consider the observation equation in (1) with scalar outputs for each sensor. This conforms with the centralized framework [5], where each row vector of the centralized observation matrix stands for the observation vector of one sensor. For the case that outputs of some sensors are not scalar, we can replace each of these sensors with a set of virtual sensors with scalar outputs, which are completely connected and connected to the neighbors of the original sensor. Then, the problem will reduce to the one studied in this article.

The following assumptions are needed.

Assumption 1: The following conditions hold:

$$\sup_t\|w(t)\|\leq b_w$$

$$\max_{i\in\mathcal{V}}\sup_t\|v_i(t)\|\leq b_v$$

$$\|\hat{x}(0) - x(0)\| \leq \eta_0$$

where $\hat{x}(0)$ is the estimate of $x(0)$ shared by all sensors, and the upper bounds are known to each sensor.

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

C. Attack Model

To deteriorate the estimation performance, a malicious attacker can compromise the observations of some targeted sensors by FDI attacks. However, due to resource limitation, the attacker can only attack a subset of all sensors at each time. Let $\mathcal{A}(t)$ and $\mathcal{A}^c(t)$ be the set of attacked sensors and the set of attack-free sensors at time t, respectively. It holds that $|\mathcal{A}(t)| + |\mathcal{A}^c(t)| = N$. We require the following assumption on the attack model.

Assumption 3: The attacker can implement the following FDI attacks to system (1): for $t = 1, 2, \ldots,$

$$\begin{aligned} &a_i(t) \in \mathbb{R}, i \in \mathcal{A}(t), |\mathcal{A}(t)| \leq s \\ &a_i(t) = 0, i \in \mathcal{A}^c(t) \end{aligned} \tag{2}$$

where sets $\mathcal{A}(t)$ and $\mathcal{A}^c(t)$ are unknown to each sensor, but s is known.

In Assumption 3, we consider the worst scenario on FDI attacks that the attacker can inject attack signals with any distribution, which is more general than results in the literature [15]. Moreover, Assumption 3 removes the requirement in [4]–[6], [8]–[10], [24], and [29] that the attacked sensor set is fixed over time.

D. Problem of Interest

Design a resilient distributed filter $\{\hat{x}_i(t)\}_{i\in\mathcal{V}}$ for the system (1) under Assumptions 1–3 by employing potentially compromised sensor observations $\{y_i(l)\}_{l=1}^t$ and the received neighbor messages over the communication graph $\mathcal{G}$, such that

$$\limsup_{t\to\infty} \|\hat{x}_i(t) - x(t)\| \leq \Delta$$

where $\Delta \geq 0$ reflects the performance of the proposed filter. Moreover, find the answers to the questions 1)–3) in the introduction.

III. Secured Distributed Filter

In this section, we first design a secured distributed filter for each sensor, and then, analyze some properties of the filter.

A. Filter Design

We consider the filter with two steps, namely, observation update and estimate consensus. In the step of observation update, by choosing $\beta > 0$, we design a saturation-like scheme to utilize observation $y_i(t)$ as follows:

$$\tilde{x}_i(t) = A\hat{x}_i(t-1) + k_i(t)C_i^T(y_i(t) - C_iA\hat{x}_i(t-1)) \tag{3}$$

where

$$k_i(t) = \begin{cases} 1, \text{ if } |y_i(t) - C_iA\hat{x}_i(t-1)| \leq \beta \\ \frac{\beta}{|y_i(t) - C_iA\hat{x}_i(t-1)|}, \text{ otherwise.} \end{cases} \tag{4}$$

Different from the gains of conventional filters or state observers (e.g., Kalman filter), the gain $k_i(t)$ is related to the estimation innovation (i.e., $y_i(t) - C_iA\hat{x}_i(t-1)$). The design of $k_i(t)$ in (4) makes sense, since if the innovation is large, observation $y_i(t)$ is more likely to be compromised. Note that $|k_i(t)(y_i(t) - C_iA\hat{x}_i(t-1))| \leq \beta$, which ensures that the attacker has limited influence to the local update. Scalar β is an observation confidence parameter reflecting the usage tradeoff between attack-free observations and attacked observations. If β is very large, almost all attack-free observations will be utilized without saturation. However, it will give much space for the attacker to deteriorate the estimation performance. If β is very small, although most attack signals $\{a_i(t)\}$ may be filtered, attack-free observations will contribute little to the estimation. The design of β will be discussed in the next subsection.

Algorithm 1: Distributed Saturation-Based Filter.

Initial setting: $(\hat{x}_i(0), \alpha, \beta, L)$
for $t = 1, 2, \ldots$ **do**
  Observation update:
  $k_i(t) = \min\{1, \frac{\beta}{|y_i(t) - C_iA\hat{x}_i(t-1)|}\}$
  $\tilde{x}_i(t) =$
  $A\hat{x}_i(t-1) + k_i(t)C_i^T(y_i(t) - C_iA\hat{x}_i(t-1))$
  Estimate consensus: Let $\hat{x}_{i,0}(t) = \tilde{x}_i(t)$
  for $l = 1, \ldots, L$ **do**
    Sensor i receives $\hat{x}_{j,l-1}(t)$ from neighbor sensor j,
    $\hat{x}_{i,l}(t) =$
    $\hat{x}_{i,l-1}(t) - \alpha\sum_{j\in\mathcal{N}_i}(\hat{x}_{i,l-1}(t) - \hat{x}_{j,l-1}(t))$
  end for
  Let $\hat{x}_i(t) = \hat{x}_{i,L}(t)$.
end for

In the step of estimate consensus, we consider a two-time-scale scheme with communication rate $L \geq 1$, i.e., each sensor can communicate with its neighbors for $L \geq 1$ times between two measurement updates. For $l = 1, 2, \ldots, L$, and $\alpha > 0$

$$\hat{x}_{i,l}(t) = \hat{x}_{i,l-1}(t) - \alpha \sum_{j\in\mathcal{N}_i} (\hat{x}_{i,l-1}(t) - \hat{x}_{j,l-1}(t)) \tag{5}$$

with $\hat{x}_{i,0}(t) = \tilde{x}_i(t)$ and $\hat{x}_i(t) = \hat{x}_{i,L}(t)$. In the lth communication, the sensor j transmits its estimate $\hat{x}_{j,l-1}(t)$ to its neighbors, $l = 1, \ldots, L$. The term $\alpha\sum_{j\in\mathcal{N}_i}(\hat{x}_{i,l-1}(t) - \hat{x}_{j,l-1}(t))$ is to make sensor estimates tend to consensus. The communication rate L is vital to guarantee the bounded estimation error especially for the case that each subsystem is not observable (i.e., (A, C_i) is not observable). It can be proven that if communicate rate L goes to infinity and parameter α is properly designed, estimates $\{\hat{x}_i(t)\}_{i=1}^N$ will converge to the same vector. However, an infinite communication rate in [16] and [30] is not necessary in this work. The design of L and α is studied in the next subsection. By (3)–(5), we provide the distributed saturation-based filter in Algorithm 1.

B. Performance Analysis

Since filtering gains $\{k_i(t)\}_{i=1}^N$ in (4) are related to the state estimates and potential compromised observations, the common stability analysis approaches, such as Lyapunov methods, may not be directly utilized. This is the main technique challenge

of this article. Inspired by the BIBO stability, we provide the following lemma to analyze boundedness of the estimation error.

Lemma 1: Consider a one-dimensional equation $x_{t+1} = F(x_t)x_t + q_0$ at time $t \geq 0$, where $x_0 \geq 0$, $q_0 \geq 0$, and $F(\cdot) \in [0, 1]$ is a monotonically nondecreasing function. If set $\Gamma = \{t \geq 1 | x_t \leq x_{t-1}\}$ is nonempty, the following conclusions hold.

1) If $q_0 \neq 0$ $\forall t_0 \in \Gamma$

$$x_t \leq F^{t-t_0}(x_{t_0})x_{t_0} + q_0 \frac{1 - F^{t-t_0}(x_{t_0})}{1 - F(x_{t_0})}, t \geq t_0.$$

2) $\sup_{t \geq t_0} x_t \leq x_{t_0}$ $\forall t_0 \in \Gamma$.

3) $\limsup_{t \to \infty} x_t \leq \inf_{t_0 \in \Gamma} x_{t_0}$.

Proof: See Appendix A. ■

If we treat x_t as an upper bound of the norm of the estimation error, based on the knowledge of x_{t_0}, we are able to use 1) to obtain a real-time upper bound of x_t, and apply 2) and 3) to obtain the uniform and asymptotic bounds of x_t, respectively. To proceed, denote

$$\lambda_0 := \min_{\mathcal{J} \subset \{1,2,\ldots,N\}: |\mathcal{J}| = N - s} \lambda_{\min} \left(\sum_{i \in \mathcal{J}} C_i^T C_i \right) \tag{6}$$

where s is the upper bound of the attacked sensor number, given in (2). Since $\sum_{i \in \mathcal{J}} C_i^T C_i$ is positive semi-definite and $s \leq N$, we have $\lambda_0 \geq 0$. Moreover, it holds that $\lambda_0 \leq (N - s)\lambda_{\max}(C_i^T C_i) = N - s$, where the equality is due to $\lambda_{\max}(C_i^T C_i) = \|C_i\|^2 = 1$ assumed after the system (1). Thus, λ_0 belongs to $[0, N - s]$. To apply Lemma 1, we construct sequence $\{\rho_t \in \mathbb{R} | \rho_t\}$ in the following:

$$\rho_{t+1} = F(\rho_t)\rho_t + q_0, \qquad \rho_0 = \eta_0 \tag{7}$$

where η_0 is given in Assumption 1 and

$$\begin{aligned} F(\rho_t) &= \|A\| \left(1 - \frac{k^*(\rho_t)}{N} \lambda_0\right) \\ k^*(\rho_t) &= \min \left\{1, \frac{\beta}{\|A\| (p_0 + \rho_t) + b_w + b_v}\right\} \\ q_0 &= \frac{N - s}{N}(b_w + b_v + \|A\| p_0) + b_w + \frac{s\beta}{N} \\ p_0 &= \frac{\sqrt{N}\beta\gamma^L}{1 - \|A\| \gamma^L} \\ \gamma &= \frac{\lambda_{\max}(\mathcal{L}) - \lambda_2(\mathcal{L})}{\lambda_{\max}(\mathcal{L}) + \lambda_2(\mathcal{L})}. \end{aligned} \tag{8}$$

Under Assumption 2, we have $\gamma \in [0, 1)$. Define $\gamma^{-1} = +\infty$ if $\gamma = 0$. The following theorem studies boundedness of the estimation error of Algorithm 1.

Theorem 1 (Bounds): Under Assumptions 1–3, consider Algorithm 1 with $\alpha = \frac{2}{\lambda_2(\mathcal{L}) + \lambda_{\max}(\mathcal{L})}$. If there exist $L > \ln \|A\| / \ln \gamma^{-1}$, $\beta, \eta_0 > 0$, such that

$$\eta_0(1 - F(\eta_0)) \geq q_0 \tag{9}$$

set $\Gamma = \{t \geq 1 | \rho_t \leq \rho_{t-1}\}$ is nonempty with $1 \in \Gamma$, where sequence $\{\rho_t\}$ is in (7). Furthermore, for $i \in \mathcal{V}$, the estimation error $e_i(t) = \hat{x}_i(t) - x(t)$ satisfies the following properties.

1) The estimation error is bounded at each time, i.e., $\forall t_0 \in \Gamma$, $t \geq t_0$

$$\|e_i(t)\| \leq R(F(\rho_{t_0}), t) + p(t)$$

where

$$\begin{aligned} R(x, t) &= x^{t-t_0}\rho_{t_0} + q_0 \frac{1 - x^{t-t_0}}{1 - x} \\ p(t) &= \sqrt{N}\beta\gamma^L \frac{1 - (\|A\| \gamma^L)^t}{1 - \|A\| \gamma^L}. \end{aligned} \tag{10}$$

2) The estimation error has a finite uniform upper bound, i.e., $\forall t_0 \in \Gamma$

$$\sup_{t \geq t_0} \|e_i(t)\| \leq \rho_{t_0} + \sup_{t \geq t_0} p(t).$$

3) The estimation error is asymptotically upper bounded, i.e.,

$$\limsup_{t \to \infty} \|e_i(t)\| \leq \inf_{t_0 \in \Gamma} \rho_{t_0} + \frac{\sqrt{N}\beta\gamma^L}{1 - \|A\| \gamma^L}.$$

Proof: See Appendix B. ■

If $t_0 = 1$, the bounds in Theorem 1 directly depend on the initial condition. With the increase of t_0, the bounds become tighter. The system designer with global system knowledge is able to examine condition (9) and calculate the error bounds in Theorem 1.

In the following theorem, we show that it is feasible to design parameters $L > \frac{\ln \|A\|}{\ln \gamma^{-1}}$, β, η_0 such that condition (9) is satisfied under the case that the system is either marginally stable or unstable, i.e., $\|A\| \geq 1$.

Theorem 2 (Feasibility): It is feasible to find positive parameters β, η_0 and $L > \frac{\ln \|A\|}{\ln \gamma^{-1}}$, and a scalar $\epsilon > 0$ such that condition (9) holds for $\|A\| \in [1, 1 + \epsilon)$, if and only if

$$\lambda_0 > s \tag{11}$$

where s and λ_0 are given in (2) and (6), respectively.

Proof: See Appendix C. ■

The condition (11) means that the maximal number of attacked sensors at each time, i.e., s, is less than scalar λ_0, which depends on the observation matrices of attack-free sensors as shown in (6). Given s, it is straightforward to check (11) with the knowledge of the system observation matrices $\{C_i\}_{i=1}^N$. For a particular system with $C_i = 1$, $i = 1, \ldots, N$, we have $\lambda_0 = N - s$. Hence, (11) is equivalent to $s \leq \lceil N/2 \rceil - 1$, which is the same maximum obtained under FDI sensor attacks in [5] and [19].

In the following theorem, by adding another condition, we show all the observations of attack-free sensors will eventually not be saturated, which contributes to tighter bounds than those in Theorem 1.

Theorem 3 (Bounds): Under the same conditions as in Theorem 1, if there is a time $t_0 \in \Gamma$ (e.g., $t_0 = 1$), such that

$$\rho_{t_0} + \sup_{t \geq t_0} p(t) < \frac{\beta - b_w - b_v}{\|A\|} \tag{12}$$

the following results hold:

1) all the observations of attack-free sensors will eventually not be saturated, i.e., $k_i(t) = 1$ $\forall i \in \mathcal{A}^c(t)$ $\forall t > t_0$;

2) compared to 1) of Theorem 1, a tighter upper bound of the estimation error is ensured, i.e., $\|e_i(t)\| \leq R(\varpi, t) + p(t) \forall t > t_0$;

3) compared to 3) of Theorem 1, a tighter asymptotic upper bound of the estimation error is ensured, i.e., $\limsup_{t\to\infty}\|e_i(t)\| \le \frac{q_0}{1-\varpi} + \frac{\sqrt{N}\beta\gamma^L}{1-\|A\|\gamma^L} < \infty$

where ρ_t is in (7), $p(t)$ and $R(\cdot,t)$ are given in (10), and

$$\varpi = \max_{\mathcal{M}\subset\{1,2,\ldots,N\}:|\mathcal{M}|=N-s} \left\| \left(I_n - \frac{1}{N}\sum_{i\in\mathcal{M}} C_i^T C_i \right) A \right\|.$$

Proof: See Appendix D. ■

The main idea to design β is to minimize two asymptotic upper bounds in Theorems 1 and 3 w.r.t. β under the constraints in (9) and (12), respectively. Since in 3) of Theorem 1, $\inf_{t_0\in\Gamma}\rho_{t_0}$ is not analytical w.r.t. β, we may choose an upper bound of $\inf_{t_0\in\Gamma}\rho_{t_0}$, e.g., $\lim_{t\to\infty}R(F(\eta_0),t)$, as the optimization loss function w.r.t. β. Note that the two optimization problems for the two cases in Theorems 1 and 3 are nonconvex, where some heuristic optimization methods can be utilized.

Algorithm resilience is on the relationship between the number of the attacked sensors and the estimation performance. Recall that s is an upper bound of the attacked sensor number given in Assumption 3, then we study the relationship between s and an upper bound of the estimation error in the following proposition.

Proposition 2: Under the same conditions as in Theorem 1, for each sensor $i\in\mathcal{V}$, the estimation error is asymptotically upper bounded, i.e.,

$$\limsup_{t\to\infty}\|e_i(t)\| \le f(s) \tag{13}$$

where $f(s) = \frac{\bar{q}_0}{1-F(\eta_0)} + \frac{\sqrt{N}\beta\gamma^L}{1-\|A\|\gamma^L}$ is a monotonically nondecreasing function w.r.t. s, in which $\bar{q}_0 = b_w + \max\{\beta, b_w + b_v + \|A\|p_0\}$ if $\|A\|<1$, otherwise $\bar{q}_0 = q_0$.

Proof: See Appendix E. ■

C. Connection With Sparse Observability

The sparse observability in the following can be used in state estimation under FDI sensor attacks.

Definition 1: The linear system defined by (1) is said to be s-sparse observable if for every set $\Gamma\subseteq\{1,\ldots,N\}$ with $|\Gamma|=s$, the pair $(A, C_{\bar{\Gamma}})$ is observable, where $C_{\bar{\Gamma}}$ is the remaining matrix by removing $C_j, j\in\Gamma$ from $[C_1^T, C_2^T, \ldots, C_N^T]^T$. Furthermore, if the pair $C_{\bar{\Gamma}}^T C_{\bar{\Gamma}} = \sum_{i=1, i\notin\Gamma}^N C_i^T C_i \succ 0$, the system is said to be one-step s-sparse observable.

In centralized frameworks, if the observations of s sensors are compromised, the system should be $2\,s$-sparse observable to guarantee the effective estimation of system state [31]. The direct relationship between (11) and the one-step s-sparse observability is given in the following.

Proposition 3 [17]: A necessary condition to guarantee $\lambda_0 > s$ is that the system (1) is one-step $2\,s$-sparse observable. If the observation vectors are orthogonal, one-step $2\,s$-sparse observability is also a sufficient condition to guarantee $\lambda_0 > s$.

If the observations of any s sensors are compromised, based on the results of Theorems 1 and 2 and Proposition 3, Algorithm 1 is able to to achieve effective estimation if the system (1) is one-step $2\,s$-sparse observable.

IV. Secured Distributed Filter With Attack Detection

In this section, we modify Algorithm 1 by adding an attack detection scheme, and then, analyze the properties of the revised algorithm. In the sequel, we need the following assumption, which is common in the literature [4]–[6], [8]–[12], [18], [24].

Assumption 4: The attacked sensor set is known to be time invariant, i.e., $\mathcal{A}(t)\equiv\mathcal{A}$, such that $|\mathcal{A}|\le s$.

Under Assumption 4, we denote $\mathcal{A}^c$ the complement of $\mathcal{A}$ in the sensor set $\mathcal{V}$, i.e., $\mathcal{A}^c\bigcup\mathcal{A}=\mathcal{V}$.

A. Detection-Based Distributed Filter

Denote $\mathcal{I}_i(t)$ the index set of detected attacked sensors known to the sensor i at time t. Moreover, we let $d_i(t)=|\mathcal{I}_i(t)|$. In order to improve the estimation performance, we aim to isolate the observations of sensors in the set $\mathcal{I}_i(t)$ in the following way: If sensor i, $i\in\mathcal{V}$, is detected to be under attack at a time T^*, i.e., $i\in\mathcal{I}_i(T^*)$, the sensor i will no longer use its observations for $t\ge T^*$. We define the following sequence $\{\bar{\rho}_{t,i}\in\mathbb{R}\}$. For each $i\in\mathcal{V}$, $t=0,1,\ldots,$ let

$$\bar{\rho}_{t+1,i} = \bar{F}(\bar{\rho}_{t,i},t)\bar{\rho}_{t,i} + \bar{q}_i(t), \qquad \bar{\rho}_{0,i}=\eta_0 \tag{14}$$

where

$$\bar{F}(\bar{\rho}_{t,i},t) = \|A\|\left(1 - \frac{k^*(\bar{\rho}_{t,i},t)}{N}\lambda_0\right)$$

$$k^*(\bar{\rho}_{t,i},t) = \min\left\{1, \frac{\beta}{\|A\|\,(p(t)+\bar{\rho}_{t,i}) + b_w + b_v}\right\}$$

$$\bar{q}_i(t) = q_0 - \frac{d_i(t)\beta}{N}$$

and q_0 is given in (8). Based on the sequence $\{\bar{\rho}_{t,i}\in\mathbb{R}\}$, we provide an attack detection condition in the following.

Detection Condition: Sensor $i\in\mathcal{V}$ is believed under attack if

$$|y_i(t) - C_iA\hat{x}_i(t-1)| > \bar{\varphi}_i(t) \tag{15}$$

where $\bar{\varphi}_i(t) = \|A\|(\bar{\rho}_{t-1,i} + p(t-1)) + b_w + b_v$, and $\bar{\rho}_{t-1,i}$ is given in (14).

Then, we propose a distributed saturation-based filter with detection in Algorithm 2, which is modified from Algorithm 1 by employing the detection condition and the observation isolation operation.

Proposition 4: Consider Algorithm 2 under the same setting as in Algorithm 1. Under the same conditions as in Theorem 1 and Assumption 4, the following results hold for every sensor i and every time $t\ge 1$.

1) The observation innovation of each attack-free sensor is upper bounded, i.e.,
$$|y_i(t) - C_iA\hat{x}_i(t-1)| \le \bar{\varphi}_i(t), i\in\mathcal{A}^c. \tag{16}$$
2) The sensors in set $\mathcal{I}_i(t)$ are under attack for sure, i.e., $\mathcal{I}_i(t)\subseteq\mathcal{A}$.
3) $\bar{\varphi}_i(t)$ is monotonically decreasing w.r.t. the number of the detected sensors (i.e., $d_i(t-1)$).

Proof: We provide the proof idea as follows. To prove (16), we just need to show $\bar{\rho}_{t-1,i}\ge\|\tilde{e}(t-1)\|$, where $\tilde{e}(t-1) = \|\frac{1}{N}\sum_{i=1}^N\hat{x}_i(t-1) - x(t-1)\|$. This is done by referring to

Algorithm 2: Distributed Saturation-Based Filter With Detection.

Initial setting: $(\hat{x}_i(0), \mathcal{I}_i(0), \alpha, \beta, L, s)$
for $t = 1, 2, \ldots$ **do**
  Update with detection: Let $\mathcal{I}_i(t) = \mathcal{I}_i(t-1)$
  if $i \in \mathcal{I}_i(t)$ **then**
    $\tilde{x}_i(t) = A\hat{x}_i(t-1)$
  else if $|\mathcal{I}_i(t)| =: d_i(t) = s$ **then**
    $\tilde{x}_i(t) = A\hat{x}_i(t-1) + C_i^T(y_i(t) - C_iA\hat{x}_i(t-1))$
  else if $|y_i(t) - C_iA\hat{x}_i(t-1)| > \bar{\varphi}_i(t)$, where $\bar{\varphi}_i(t)$ is in (15) **then**
    $\tilde{x}_i(t) = A\hat{x}_i(t-1)$, let $\mathcal{I}_i(t) = \mathcal{I}_i(t) \cup \{i\}$
  else
    $k_i(t) = \min\left\{1, \frac{\beta}{|y_i(t) - C_iA\hat{x}_i(t-1)|}\right\}$
    $\tilde{x}_i(t) = A\hat{x}_i(t-1) + k_i(t)C_i^T(y_i(t) - C_iA\hat{x}_i(t-1))$
  end if
  Estimate consensus:
  $\hat{x}_{i,0}(t) = \tilde{x}_i(t), \mathcal{I}_{i,0}(t) = \mathcal{I}_i(t)$
  for $l = 1, \ldots, L$ **do**
    Sensor i obtains $\{\hat{x}_{j,l-1}(t), \mathcal{I}_{j,l-1}(t)\}$ from sensor j,
    $\hat{x}_{i,l}(t) = \hat{x}_{i,l-1}(t) - \alpha\sum_{j\in\mathcal{N}_i}(\hat{x}_{i,l-1}(t) - \hat{x}_{j,l-1}(t))$
    $\mathcal{I}_{i,l}(t) = \bigcup_{j\in\mathcal{N}_i}\mathcal{I}_{j,l-1}(t)\bigcup\mathcal{I}_{i,l-1}(t)$
  end for
  Let $\hat{x}_i(t) = \hat{x}_{i,L}(t), \mathcal{I}_i(t) = \mathcal{I}_{i,L}(t)$.
end for

(22)–(26) and by noting that the number of attacked but undetected sensors is upper bounded by $s - d_i(t-1)$. The conclusion 2) follows from 1). The conclusion 3) is satisfied, since $\bar{\rho}_{t-1,i}$ in (14) is monotonically decreasing w.r.t. $d_i(t-1)$. ■

In our framework, an attack signal $a_i(t)$ is stealthy if the compromised observation $y_i(t)$ violates detection condition (15), i.e., a stealthy attack signal $a_i(t)$ is in set $\{a_i(t) \in \mathbb{R}|\,|\xi_i(t) + a_i(t)| \le \bar{\varphi}_i(t)\}$, where $\xi_i(t)$ is the attack-free observation innovation, i.e., $\xi_i(t) = C_iA(x(t-1) - \hat{x}(t-1)) + C_iw(t-1) + v(t)$. Since this article considers bounded noise processes, a single large attack signal (larger than $\bar{\varphi}_i(t)$) will expose the attacked sensor for sure. In other words, if there is a time t such that $a_i(t) \in \{a_i(t) \in \mathbb{R}|\,|\xi_i(t) + a_i(t)| > \bar{\varphi}_i(t)\}$, this attacked sensor i will be detected. Thus, this detection scheme differs from the methods based on constructing statistical variables for hypothesis tests on the innovation distributions (e.g., [32]). In our framework, the knowledge of the attacker on the designed detector especially on threshold $\bar{\varphi}_i(t)$ will largely influence the detected sensor number.

B. Error Bounds and Convergence

Denote $d(t)$ the maximal number of detected sensors at time t, i.e., $d(t) = \max_{i\in\mathcal{V}}\{d_i(t)\}$, with $d(0) = 0$. Given any $T \ge 0$ $\forall t \ge T$, we construct the following sequence $\{\bar{\rho}_t \in \mathbb{R}|\bar{\rho}_t, t \ge T\}$:

$$\bar{\rho}_{t+1} = F(\bar{\rho}_t)\bar{\rho}_t + \bar{q}_0, \qquad \bar{\rho}_T = \rho_T \tag{17}$$

where $\bar{q}_0 = q_0 - \frac{d(T)\beta}{N}$, and ρ_T and $F(\cdot)$ are given in (7) and (8), respectively. Then, the following theorem builds the connection between $d(T)$ and an upper bound of the estimation error of Algorithm 2.

Theorem 4: Consider Algorithm 2 under the same setting as in Algorithm 1. Under the same conditions as in Theorem 1 and Assumption 4, the estimation error is asymptotically upper bounded, i.e.,

$$\limsup_{t\to\infty}\|e_i(t)\| \le W(T)\forall T \ge 0$$

where

$$W(T) = \inf_{t_0\in\bar{\Gamma}}\rho_{t_0} + \frac{\sqrt{N}\beta\gamma^L}{1 - \|A\|\gamma^L} - \frac{d(T)\beta}{N(1-F_*)}$$

and $F_* = \inf_{t_0\in\bar{\Gamma}}F(\bar{\rho}_{t_0}) \in [0,1)$, $\bar{\Gamma} = \{t \ge T|\bar{\rho}_t \le \bar{\rho}_{t-1}\}$.

Proof: See Appendix F. ■

Algorithm 2 ensures that detected sensor number $d(T)$ is nondecreasing as T increases. As more sensors are detected (i.e., $d(T)$ is increasing), we get a tighter bound in Theorem 4. Thus, the bound in Theorem 4 is equal to or smaller than the bound in 3) of Theorem 1. In practice, the system defender can use $d(T)$ at different T to generate a sequence of nonincreasing bounds, which can be used to estimate the asymptotic error bound. In the following theorem, we provide the conditions such that the state estimate of Algorithm 2 converges to the system state asymptotically.

Theorem 5 (Convergence): Consider Algorithm 2 under the same setting as in Algorithm 1. Under the same conditions as in Theorem 1, Assumption 4, and the following conditions:

1) the system is noise-free, i.e., $w(t) \equiv 0$ and $v_i(t) \equiv 0$ for any $i \in \mathcal{V}$;
2) the attacker compromises s sensors and they are detected in finite time, i.e., there exists a finite time $\hat{t}_0$ and an i such that $d_i(\hat{t}_0) = s$;
3) matrix G is Schur stable;

then the estimate in Algorithm 2 will asymptotically converge to the state, i.e.,

$$\lim_{t\to\infty}\|\hat{x}_i(t) - x(t)\| = 0, i \in \mathcal{V}$$

where

$$G = \begin{pmatrix} 2\|A\|\gamma^L & \|A\|\gamma^L\sqrt{N-s} \\ \tau_0 & \varpi \end{pmatrix}$$

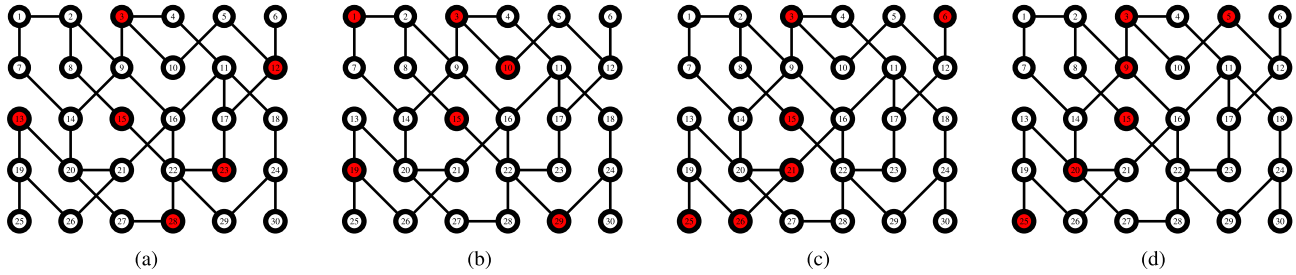
$$\tau_0 = \max_{\mathcal{M}\subset\mathcal{V},|\mathcal{M}|=N-s}\left\|\frac{1}{N}(\mathbf{1}_N^T \otimes I_n)\bar{C}^T\bar{K}_{\mathcal{M}}\bar{C}(I_N \otimes A)\right\|$$

$$\bar{K}_{\mathcal{M}} = \mathrm{diag}\{\mathbb{I}_{1\in\mathcal{M}}, \mathbb{I}_{2\in\mathcal{M}}, \ldots, \mathbb{I}_{N\in\mathcal{M}}\} \in \mathbb{R}^{N\times N}$$

where ϖ is given in Theorem 3, and $\bar{C} = \mathrm{diag}\{C_1, \ldots, C_N\}$.

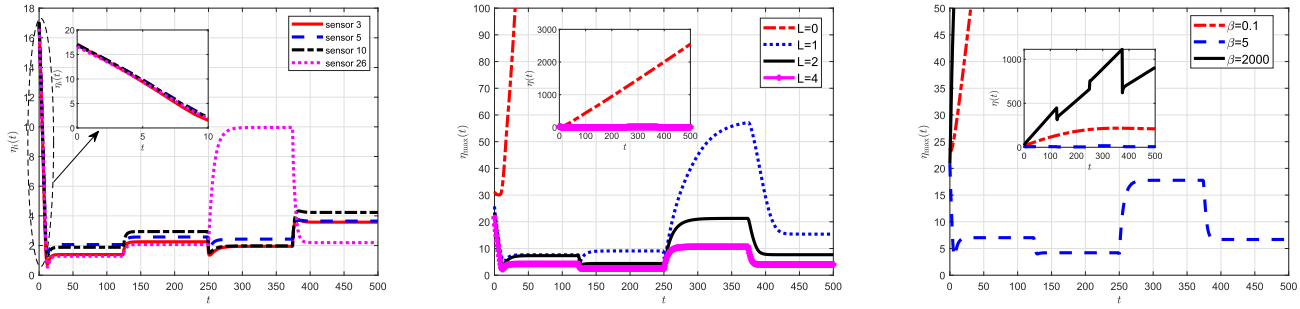
Proof: See Appendix G. ■

Condition 2) can be satisfied when the attacker compromises s sensors without using persistently stealthy attack signals. In this case, the detector is able to identify all attacked sensors in finite time and remove their influence. Otherwise, the attacker can have influence to the estimation such that the estimation error is not tending to zero, but the estimation error is still bounded as we state in Theorems 1 and 4. It is worth noting that condition 2) can be monitored by each sensor to know whether the number



(a) (b) (c) (d)

Fig. 3. Sensor network with different sets of attacked sensors over four time intervals.



(a) (b) (c)

Fig. 4. Estimation performance of Algorithm 1 for the sensor networks in Fig. 3.

of detected sensors reaches s. Condition 3) can be fulfilled if the communication rate L is relatively large, the number of the attack-free sensors (i.e., $N-s$) is relatively large, or the norm of the transition matrix (i.e., $\|A\|$) is small.

V. Simulation Results

In this section, we provide numerical simulations to show the effectiveness of the developed results.

Consider a second-order system monitored by a sensor network with 30 nodes, where $A=\begin{pmatrix}1 & 0.1\\ 0 & 1\end{pmatrix}$, $C_i=(0,1)$ for sensor $i\in\{2,3,5,9,12,16,17,21,22,25,26,29\}$, and $C_i=(1,0)$ for the rest of the sensors. All observation noise $v_i(t)$ and process noise $w_j(t)$ follow the uniform distribution in [0,0.01], where $(w_1(t),w_2(t))^T=:w(t)$. The initial state is $x(0)=(25,25)^T$. Each element of $\hat{x}(0)$ follows the uniform distribution in $[0,25]$. The bounds in Assumption 1 are assumed to be $b_v=0.01, b_w=0.02,$ and $\eta_0=50$. We consider the time interval $t=0,1,\ldots,500$, and suppose that the attacker inserts signal $a_i(t)=2(C_ix(t)+v_i(t))$ if the sensor i is under attack. We conduct a Monte Carlo experiment with 100 runs. Define the average estimation error of the sensor i and the maximum error over the whole network by

$$\eta_i(t)=\frac{1}{100}\sum_{j=1}^{100}\left\|e_i^j(t)\right\|$$

$$\eta_{\max}(t)=\frac{1}{100}\sum_{j=1}^{100}\max_{i\in\{1,\ldots,30\}}\left\|e_i^j(t)\right\|$$

respectively, where $e_i^j(t)$ is the state estimation error of the sensor i at time t in the jth run.

A. Secured Distributed Estimation

In this subsection, we verify the performance of Algorithm 1 by considering the case that the attacked sensor set is time varying. Assume that the distributed sensor networks under sensor attacks are switched in the way of Fig. 3 in each run, where a node in red means the node is under attack.

By selecting $\beta=3$ and $L=3$ for Algorithm 1, estimation errors $\eta_i(t), i=3,5,10,26$ are provided in Fig. 4(a). From this figure, we see that there are three time instants, i.e., $t=$ 125 250,375, around which the error dynamics of the plotted sensors fluctuate. The reason why the error dynamics of one sensor increase lies in two aspects. First, after a certain time instant, this sensor is under attack. If so, its observations will be compromised and the estimation performance of this sensor will be degraded. For example, in Fig. 4 (a), the errors of sensor 5 after time $t=375$, sensor 10 after time $t=125$, and sensor 26 after time $t=250$ all increase for this reason. The second aspect is that after a certain time instant, the neighbor of one senor is under attack. For this sensor, its error will also increase due to the consensus influence. For example, in Fig. 4(a), the error of sensor 3 increases after time $t=125$, since a neighbor of sensor 3, i.e., sensor 10, is under attack after time $t=125$. The reason that the estimation error of sensor 26 is large in the time interval 250–375 in Fig. 4(a) is because of the following two reasons.

1) After $t=250$, sensor 26 is persistently under attack until $t=375$.
2) The communication rate $L=1$ is small.

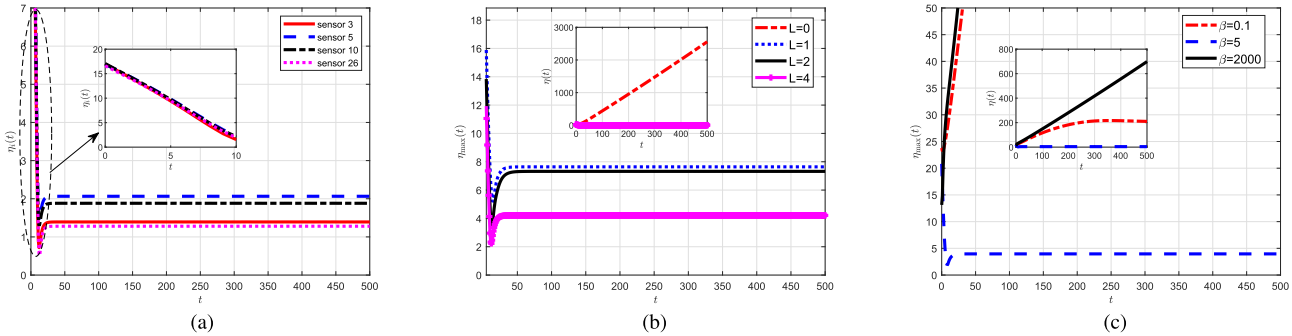


Fig. 5. Estimation performance of Algorithm 1 for the sensor network in Fig. 1.

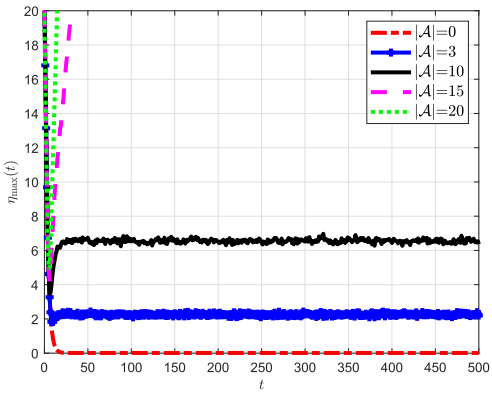


Fig. 6. Resilience of Algorithm 1.

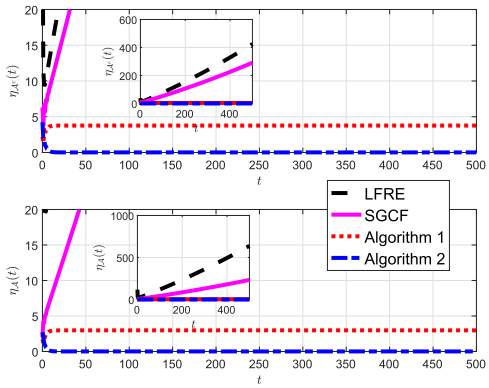


Fig. 7. Estimation error comparison of algorithms.

Since the state estimate is affected by the compromised sensor observations, the estimation error will inevitably increase if the information of neighbors is not available in time. However, Algorithm 1 provides a way to improve the estimation performance by increasing the communication rate L. In Fig. 4(b), the relationship between the estimation error $\eta(t)$ and communication rate L is studied. The result shows that with the increase of the communication rate L, estimation error $\eta(t)$ decreases. The connection between the estimation error $\eta(t)$ and parameter β is studied in Fig. 4(c) under $L = 4$. The figure shows that the estimation error with $\beta = 5$ is the smallest within the errors by, respectively, setting $\beta = 0.1, 5, 2000$. The result conforms to the discussion on the design of β, which should not be too small or too large in the algorithm setting.

To show the resilience of Algorithm 1, for the case that the attacked sensor set is time invariant, under $L = 4$ and $\beta = 5$, we study the relationship between $|\mathcal{A}|$ and $\eta(t)$ in Fig. 6 by randomly choosing a subset of the whole sensors (i.e., $\mathcal{A}$). The result of this figure shows that with the increase of $|\mathcal{A}|$, the estimation error $\eta(t)$ becomes larger. Also, when $|\mathcal{A}|$ is equal to or larger than half sensors, the estimation error is unstable.

B. Secured Distributed Estimation Under Detection

In this subsection, we consider the introduction case in Fig. 1 where the attacked sensor set is time invariant, with $\mathcal{A}(t) = \mathcal{A} = \{3, 12, 13, 15, 23, 28\}$, under which the estimation performance of Algorithms 1 and 2 is studied.

By selecting $\beta = 3$ and $L = 3$ for Algorithm 1, estimation errors $\eta_i(t), i = 3, 5, 10, 26$ are provided in Fig. 5(a). Compared to the result in Fig. 4(a), the estimation errors of these sensors in Fig. 5(a) do no fluctuate. Since the attacked sensor set is time invariant, the estimation errors become steady after transience as shown in the figure. The relationships between parameters L and β and estimation error $\eta(t)$ are shown in Fig. 5(b) and (c). Note that in Fig. 5(b), when $L = 0$, the estimation error $\eta(t)$ is divergent, since for the attacked sensors, the local observability is violated. Fig. 5(b) also shows that with the increase of the communication rate L, the estimation error of Algorithm 1 is decreasing. In Fig. 5(c), parameter $\beta = 5$ can lead to the stable estimation error, while in Fig. 4(c), the estimation error for parameter $\beta = 5$ fluctuates due to the switching of the attacked sensor set.

In order to compare with existing algorithms under the situation in Fig. 1, we define the estimation errors for the attacked sensor set and attack-free sensor set, respectively: $\eta_{\mathcal{A}}(t) = \frac{1}{100}\sum_{j=1}^{100} \max_{i\in\mathcal{A}} \|e_i^j(t)\|$, and $\eta_{\mathcal{A}^c}(t) = \frac{1}{100}\sum_{j=1}^{100} \max_{i\in\mathcal{A}^c} \|e_i^j(t)\|$. In Fig. 7, the estimation performance of Algorithms 1 and 2 with $L = 5$, the local-filtering-based resilient estimation (LFRE) [11] and the scalar-gain consensus filter[1] (SGCF) is compared. The figure shows that Algorithms 1 and 2 provide estimates with stable estimation errors for both attacked sensors and attack-free sensors, but the

[1]The filter has the same form as Algorithm 1 but $k_i(t) = 1$ for $t \geq 1$. It follows the idea in [33], which did not consider the attack scenario.

estimation errors of LFRE and SGCF are divergent. Compared to Algorithm 1, Algorithm 2 provides smaller estimation errors by successfully detecting all the attacked sensors.

VI. Conclusion and Future Work

This article studied the distributed filtering problem for LTI systems with bounded noise under false-data injection attacks in sensors networks, where a malicious attacker can compromise a time-varying and unknown subset of sensors and manipulate their observations arbitrarily. First, we proposed a distributed saturation-based filter. Then, we provided a sufficient condition to guarantee boundedness of the estimation error. By confining the attacked sensor set to be time invariant, we then modified the filter by adding an attack detection scheme. Moreover, for the noise-free case, we proved that the state estimate of each sensor asymptotically converges to the system state under certain conditions.

There are some future directions. Since this article employs a two-time-scale scheme in the filter design, it is interesting to develop algorithms that use one communication at each time. Other directions include considering more general systems (e.g., nonlinear systems) and more complex sensor networks (e.g., random or communication-delayed networks).

Appendix A

A. Proof of Lemma 1

First, we prove 2). Due to $t_0 \in \Gamma$, we have

$$F(x_{t_0-1})x_{t_0-1} + q_0 = x_{t_0} \leq x_{t_0-1}. \tag{18}$$

For $t = t_0 + 1$, by (18) and the condition that $F(\cdot) \in [0, 1]$ is a monotonically nondecreasing function, we have $x_{t_0+1} = F(x_{t_0})x_{t_0} + q_0 \leq F(x_{t_0-1})x_{t_0-1} + q_0 = x_{t_0}$. By recursively applying the aforementioned procedure, we have $\sup_{t\geq t_0} x_t \leq x_{t_0}$.

Next, we prove 1), which trivially holds for $t = t_0$. Consider the case of $t > t_0$ in the following. If $q_0 \neq 0$, it follows from (18) that $F(x_{t_0-1}) \in [0, 1)$. Due to $F(x_{t_0}) \leq F(x_{t_0-1})$, we have $F(x_{t_0}) \in [0, 1)$. By 2) and the condition that $F(\cdot) \in [0, 1]$ is a monotonically nondecreasing function, we have $F(x_{t-1}) \leq F(x_{t_0})$. Then, $x_t \leq F(x_{t_0})x_{t-1} + q_0$ with $F(x_{t_0}) \in [0, 1)$. Thus, 1) is satisfied by recursively applying the inequality for $t - t_0$ times.

Finally, we prove 3). By 1) and the definition of $\limsup$, we have $\limsup_{t\to\infty} x_t = \inf_{t\in\mathbb{Z}^+} \sup_{m\geq t} x_m \leq \inf_{t\in\Gamma} \sup_{m\geq t} x_m \leq \inf_{t\in\Gamma} x_t$.

B. Proof of Theorem 1

Let $e_i(t) = \tilde{e}(t) + \bar{e}_i(t)$, where $\bar{e}_i(t) := \hat{x}_i(t) - \hat{x}_{\text{avg}}$, and $\tilde{e}(t) = \hat{x}_{\text{avg}} - x(t)$, and $\hat{x}_{\text{avg}}(t) := \frac{1}{N}\sum_{i=1}^N \hat{x}_i(t)$. Besides, we denote

$$\begin{aligned}
X(t) &= \mathbf{1}_N \otimes x(t) \in \mathbb{R}^{Nn}\\
\bar{E}(t) &= \left(\bar{e}_1^T(t), \ldots, \bar{e}_N^T(t)\right)^T \in \mathbb{R}^{Nn}\\
Y(t) &= \left(y_1^T(t), \ldots, y_N^T(t)\right)^T \in \mathbb{R}^N\\
V(t) &= \left(v_1^T(t), \ldots, v_N^T(t)\right)^T \in \mathbb{R}^N\\
\hat{X}(t) &= \left(\hat{x}_1^T(t), \ldots, \hat{x}_N^T(t)\right)^T \in \mathbb{R}^{Nn}\\
\bar{C} &= \text{diag}\{C_1, \ldots, C_N\} \in \mathbb{R}^{N\times Nn}\\
\bar{K}(t) &= \text{diag}\{k_1(t), \ldots, k_N(t)\} \in \mathbb{R}^{N\times N}\\
P_{Nn} &= \frac{1}{N}(\mathbf{1}_N \otimes I_n)(\mathbf{1}_N \otimes I_n)^T \in \mathbb{R}^{Nn\times Nn}.
\end{aligned} \tag{19}$$

The idea for the proof is that we first show $\|\bar{e}_i(t)\|$ is upper bounded by $p(t)$, and then, we prove $\|\tilde{e}(t)\|$ is upper bounded by the quantities in 1)–3) of the theorem. The following lemma with a similar proof as in [27] ensures $\|\bar{e}_i(t)\| \leq p(t)$.

Lemma 2: Consider Algorithm 1, and let Assumptions 1–2 hold. If $\alpha = \frac{2}{\lambda_2(\mathcal{L})+\lambda_{\max}(\mathcal{L})}$, and $L > \frac{\ln\|A\|}{\ln\gamma^{-1}}$, then for $t \geq 0$

$$\left\|\bar{E}(t)\right\| \leq p(t) \tag{20}$$

where $p(t)$ and $\bar{E}(t)$ are defined in (10) and (19), respectively.

Proof of Theorem 1: It follows from (7) and (9) that $\rho_1 \leq \rho_0 = \eta_0$, which means that $1 \in \Gamma = \{t \geq 1 | \rho_t \leq \rho_{t-1}\}$. In the following, we prove 1)–3).

Under Assumption 3, there are at least $N - s$ attack-free sensors at each time. Suppose $\mathcal{J}(t)$ is the set of these $N - s$ sensors, i.e., $\mathcal{J}(t) \subseteq \mathcal{A}^c(t)$ with $|\mathcal{J}(t)| = N - s$. Denote $\mathcal{J}^c(t) = \mathcal{V} \setminus \mathcal{J}(t)$, which satisfies $|\mathcal{J}^c(t)| = s$ due to $|\mathcal{V}| = N$. By Algorithm 1 and the notations in (19), we have the dynamics of $\hat{x}_{\text{avg}}(t)$ in (21). Due to $(\mathbf{1}_N^T \otimes I_n)P_{Nn} = (\mathbf{1}_N^T \otimes I_n)(I_{Nn} - \alpha(\mathcal{L} \otimes I_n))^L = (\mathbf{1}_N^T \otimes I_n)$, we have

$$\tilde{e}(t) = M_t\tilde{e}(t-1) - \bar{m}_t + \tilde{m}_t \tag{22}$$

where M_t, $\tilde{m}_t$, and $\bar{m}_t$ are given in (21) shown at the bottom of the next page. Note that $\tilde{m}_t$ can be rewritten in the following way:

$$\tilde{m}_t = \frac{1}{N}\sum_{i\in\mathcal{J}^c(t)} C_i^T k_i(t)(y_i(t) - C_iA\hat{x}_i(t-1)).$$

Due to $k_i(t) = \min\{1, \frac{\beta}{|y_i(t)-C_iA\hat{x}_i(t-1)|}\}$, it holds that $|k_i(t)(y_i(t) - C_iA\hat{x}_i(t-1))| \leq \beta$. Since we assume $\|C_i\| = 1$ after the system model, it holds that $\|\tilde{m}_t\| \leq \frac{1}{N}\sum_{i\in\mathcal{J}^c(t)} \|C_i^T\|\beta \leq |\mathcal{J}^c(t)|\frac{\beta}{N}$. Due to $|\mathcal{J}^c(t)| = s$, we have

$$\|\tilde{m}_t\| \leq \frac{s}{N}\beta. \tag{23}$$

Regarding $\bar{m}_t$, by Assumption 3 and $k_i(t) \leq 1$, we have

$$\begin{aligned}
\|\bar{m}_t\| &\leq \|w(t-1)\| + \frac{|\mathcal{J}(t)|}{N}(\|A\|\left\|\bar{E}(t-1)\right\|\\
&\quad + \|w(t-1)\| + b_v)\\
&\leq \frac{N-s}{N}(b_w + b_v + \|A\|p_0) + b_w
\end{aligned} \tag{24}$$

where the second inequality is obtained by Lemma 2 and $\sup_{t\geq 0} p(t) \leq p_0$, where p_0 is defined in (8). Based on (22)–(24), we construct the sequence $\{\rho_t\}$ in (7). In the following, we prove that $\|\tilde{e}(t)\| \leq \rho_t$.

At the initial time, i.e., $t = 0$, by Assumption 1, we have $\|\tilde{e}(0)\| = \|\hat{x}_{\text{avg}}(0) - x(0)\| \leq \frac{1}{N}\sum_{i=1}^N \|\hat{x}_i(0) - x(0)\| \leq \eta_0$. Due to $\rho_0 = \eta_0$, $\|\tilde{e}(t)\| \leq \rho_t$ for $t = 0$. Suppose at time $t - 1$,

$\|\tilde{e}(t-1)\| \le \rho_{t-1}$. At time t, for $i \in \mathcal{J}(t)$, we consider

$$\begin{aligned} & |y_i(t) - C_i A\hat{x}_i(t-1)| \\ \le\ & \|A\|\,\|e_i(t-1)\| + b_w + b_v \\ \le\ & \|A\|\,(\|\bar{e}_i(t-1)\| + \|\tilde{e}(t-1)\|) + b_w + b_v \\ \le\ & \|A\|\,(p_0 + \rho_{t-1}) + b_w + b_v \end{aligned} \tag{25}$$

where the last inequality of (25) is obtained by noting that $\sup_{t\ge 0}\|\bar{e}_i(t)\| \le \sup_{t\ge 0} p(t) \le p_0$ and p_0 is defined in (8). Recall the form of $k_i(t)$, by (25), for $i \in \mathcal{J}(t)$, we have $k_i(t) \ge k^*(\rho_{t-1}) := \min\left\{1, \frac{\beta}{\|A\|(p_0+\rho_{t-1})+b_w+b_v}\right\} > 0$. Then

$$\begin{aligned} \|M_t\| \le\ & \|A\| \left\| \left(I_n - \frac{k^*(\rho_{t-1})}{N} \sum_{i\in\mathcal{J}(t)} C_i^T C_i \right) \right\| \\ \le\ & \|A\| \left(1 - \frac{k^*(\rho_{t-1})}{N} \lambda_0 \right). \end{aligned} \tag{26}$$

Taking norm on both sides of (22) and considering (7), (23), (24), and (26), we have $\|\tilde{e}(t)\| \le \rho_t$.

Since the defined $F(\rho_t)$ in (8) is monotonically nondecreasing function, conclusions 1)–3) of this theorem are obtained by applying the results in Lemma 1, $\|e_i(t)\| \le \|\tilde{e}(t)\| + \|\bar{e}_i(t)\|$, and (20).

C. Proof of Theorem 2

1) Sufficiency:

Case 1: For the case $s > 0$, we consider $\|A\| \in [1, 1+\epsilon)$ with $\epsilon = \frac{\lambda_0 - s}{4(N-\lambda_0)}$, which is positive due to $\lambda_0 > s$. If L is sufficiently large, $p_0 > 0$ in (8) will be sufficiently small. Thus, given noise bounds b_w and b_v, considering $N \ge s + \lambda_0$, it is feasible to choose sufficiently large β, η_0 and $L > \frac{\ln\|A\|}{\ln\gamma^{-1}}$, such that

$$\begin{aligned} & \beta \ge (1+\epsilon)(p_0+\eta_0) + b_w + b_v \\ & \beta \le \min\left\{ (1+\epsilon+\frac{\lambda_0 - s}{4\,s})\eta_0, \frac{N}{s}(\eta_0 - \frac{Q_0}{\epsilon_1}) \right\} \end{aligned} \tag{27}$$

where $\epsilon_1 = \frac{\epsilon}{1+2\epsilon} > 0$, and

$$Q_0 := \frac{N-s}{N}(b_w + b_v + \|A\|\,p_0) + b_w. \tag{28}$$

By the first inequality and second inequality of (27), we have $k_0^* = \min\{1, \frac{\beta}{\|A\|(p_0+\eta_0)+b_w+b_v}\} = 1$ and $\frac{s\beta}{N\eta_0} < 1$, respectively. Then

$$\begin{aligned} m_0 :\ & = \left(1 - \frac{s\beta}{N\eta_0}\right)\left(1 - \frac{k_0^*\lambda_0}{N}\right)^{-1} \\ & = \left(\frac{N}{s} - \frac{\beta}{\eta_0}\right)\left(\frac{N-\lambda_0}{s}\right)^{-1} \\ & = 1 + \left(\frac{\lambda_0}{s} - \frac{\beta}{\eta_0}\right)\frac{s}{N-\lambda_0} \\ & \overset{(a)}{\ge} 1 + \left(\frac{\lambda_0}{s} - 1 - \frac{\lambda_0 - s}{4\,s} - \epsilon\right)\frac{s}{N-\lambda_0} \\ & = 1 + \frac{3(\lambda_0 - s)}{4(N-\lambda_0)} - \frac{\epsilon s}{N-\lambda_0} \\ & \overset{(b)}{\ge} 1 + 2\epsilon \end{aligned} \tag{29}$$

where (a) is obtained by applying the second inequality of (27), and (b) is derived by using $\lambda_0 - s = 4\epsilon(N-\lambda_0)$ and $s \le N - \lambda_0$. From the second inequality of (27), we obtain

$$\vartheta_0 := 1 - \frac{Q_0}{\eta_0}\left(1 - \frac{s\beta}{N\eta_0}\right)^{-1} \ge 1 - \epsilon_1 = \frac{1+\epsilon}{1+2\epsilon}. \tag{30}$$

By (29) and (30), we have $\vartheta_0 m_0 \ge 1 + \epsilon \ge \|A\|$. It is easy to check that $\vartheta_0 m_0 \ge \|A\|$ is equivalent equation (9). Thus, the sufficiency is satisfied in this case with the above parameters, i.e., $\epsilon = \frac{\lambda_0 - s}{4(N-\lambda_0)}$, and β, η_0 and $L > \frac{\ln\|A\|}{\ln\gamma^{-1}}$ satisfying (27).

Case 2: For the attack-free case, i.e., $s = 0$, we consider $\|A\| \in [1, 1+\epsilon)$ with $\epsilon = \frac{\lambda_0}{4N-\lambda_0} > 0$. Similar to case 1, it is feasible to choose sufficiently large β, η_0 and $L > \frac{\ln\|A\|}{\ln\gamma^{-1}}$, such that

$$\begin{aligned} & 2\beta \ge (1+\epsilon)(p_0+\eta_0) + b_w + b_v \\ & \frac{q_0}{\eta_0} \le \frac{\lambda_0}{4\,N}(1+\epsilon). \end{aligned} \tag{31}$$

From the first inequality of (31), we see $k_0^* = \min\{1, \frac{\beta}{\|A\|(p_0+\eta_0)+b_w+b_v}\} \ge \frac{1}{2}$. With $k_0^* \ge \frac{1}{2}$, it is easy to check that (9) is satisfied if $\frac{1}{\|A\|}\frac{\eta_0 - q_0}{\eta_0} \ge 1 - \frac{\lambda_0}{2\,N}$. This inequality is satisfied due to $\|A\| \in [1, 1+\epsilon)$ and the second inequality of (31). Thus, the sufficiency is satisfied in this case with $\epsilon = \frac{\lambda_0}{4N-\lambda_0}$, and β, η_0 and $L > \frac{\ln\|A\|}{\ln\gamma^{-1}}$ satisfying (31).

$$\begin{aligned} & \hat{x}_{\text{avg}}(t) = A\hat{x}_{\text{avg}}(t-1) + \frac{1}{N}(\mathbf{1}_N^T \otimes I_n) P_{Nn} \left(I_{Nn} - \alpha(\mathcal{L}\otimes I_n)\right)^L \bar{C}^T \bar{K}(t) h(t) \\ & \bar{K}_{\mathcal{J}(t)} = \text{diag}\{k_1(t)\mathbb{I}_{1\in\mathcal{J}(t)}, \ldots, k_N(t)\mathbb{I}_{N\in\mathcal{J}(t)}\}, \quad \bar{K}_{\mathcal{J}^c(t)} = \text{diag}\{k_1(t)\mathbb{I}_{1\in\mathcal{J}^c(t)}, \ldots, k_N(t)\mathbb{I}_{N\in\mathcal{J}^c(t)}\} \\ & \bar{m}_t = w(t-1) + \frac{1}{N}(\mathbf{1}_N^T \otimes I_n)\bar{C}^T \bar{K}_{\mathcal{J}(t)}(t)\left(\bar{C}((I_N \otimes A)\bar{E}(t-1) - I_N \otimes w(t-1)) - V(t)\right) \\ & h(t) = Y(t) - \bar{C}(I_N \otimes A)\hat{X}(t-1), \quad M_t = (I_n - \frac{1}{N}\sum_{i\in\mathcal{J}(t)} k_i(t) C_i^T C_i)A, \quad \tilde{m}_t = \frac{1}{N}(\mathbf{1}_N^T \otimes I_n)\bar{C}^T \bar{K}_{\mathcal{J}^c(t)}(t) h(t) \end{aligned} \tag{21}$$

Note that for η_0 in the aforementioned two cases, we are able to make it bigger such that the initial error condition in Assumption 1 holds.

2) Necessity: We use the contradiction method. If $\lambda_0 > s$ does not hold, i.e., $\lambda_0 \leq s$. Equation (9) is equivalent to

$$\frac{\eta_0 - q_0}{\eta_0}\left(1 - \frac{k_0^*\lambda_0}{N}\right)^{-1} \geq \|A\| \geq 1. \tag{32}$$

If $\frac{s\beta}{N\eta_0} \geq 1$, from the form of q_0 in (8), we have $\eta_0 < q_0$. Then, the left-hand side of (32) is negative, which contracts with the right-hand side of (32). Thus, $\frac{s\beta}{N\eta_0} < 1$. With the same notations as case 1 of the sufficiency proof, (32) is equivalent to $\vartheta_0 m_0 \geq \|A\| \geq 1$. Due to $\frac{s\beta}{N\eta_0} < 1$, we have $\vartheta_0 < 1$. Then, m_0 has to be larger than 1, which leads to $\frac{s\beta}{N\eta_0} < \frac{k_0^*\lambda_0}{N}$. It is equivalent to $\frac{s}{\lambda_0} < \frac{k_0^*\eta_0}{\beta}$. Due to $\frac{s}{\lambda_0} \geq 1$, we have $\frac{k_0^*\eta_0}{\beta} > 1$, which however cannot be satisfied due to $k_0^* = \min\{1, \frac{\beta}{\|A\|(p_0+\eta_0)+b_w+b_v}\}$. Thus, the conjecture $\lambda_0 \leq s$ is not right, which means $\lambda_0 > s$.

D. Proof of Theorem 3

First, we prove 1). For $i \in \mathcal{A}^c(t)$, by 2) of Theorem 1, we have $\sup_{t\geq t_0}\|e_i(t)\| \leq \rho_{t_0} + \sup_{t\geq t_0} p(t)$, thus

$$\begin{aligned}
&\sup_{t\geq t_0+1} |y_i(t) - C_i A\hat{x}_i(t-1)| \\
\leq\ & \|A\| \sup_{t\geq t_0+1}\|e_i(t-1)\| + b_w + b_v \\
\leq\ & \|A\|\left(\rho_{t_0} + \sup_{t\geq t_0} p(t)\right) + b_w + b_v.
\end{aligned}$$

If (12) holds, by Algorithm 1, all the observations of the attack-free sensors will eventually not be saturated, i.e., $k_i(t) = 1 \forall i \in \mathcal{A}^c(t), t \geq t_0 + 1$.

Next, we prove 2). By 1) of this theorem, for $i \in \mathcal{A}^c(t)$, we have $k_i(t) = 1\ \forall t > t_0$, then $\|M_t\| \leq \varpi$, where M_t is defined in (21). According to the error dynamics in (22) and inequalities (23)–(24), the upper bound of $\|\tilde{e}(t)\|$ is obtained by applying Lemma 1. It follows from (26) that the bound is tighter than the one in 1) of Theorem 1. Due to $\|e_i(t)\| \leq \|\tilde{e}(t)\| + \|\bar{e}_i(t)\|$ and (20), the upper bound of $\|e_i(t)\|$ is obtained.

Finally, we prove 3). By the real-time upper bound of the estimation error, it is straightforward to have its limit superior. Next, we prove the limit superior bound is no larger than the one in 3) of Theorem 1, i.e., $\frac{q_0}{1-\varpi} \leq \inf_{t_0\in\Gamma}\rho_{t_0}$. Employing the properties $\inf x + y \geq \inf x + \inf y$ and $\inf xy \geq \inf x \inf y$ for $x, y > 0$ on $F(\rho_{t_0})\rho_{t_0} + q_0 = \rho_{t_0+1} \leq \rho_{t_0}$ yields

$$\begin{aligned}
\inf_{t_0\in\Gamma}\rho_{t_0} &\geq \frac{q_0}{1 - \inf_{t_0\in\Gamma} F(\rho_{t_0})} \\
&\overset{(a)}{\geq} \frac{q_0}{1 - \|A\|\left(1 - \frac{1}{N}\lambda_0\right)} \\
&\geq \frac{q_0}{1-\varpi}
\end{aligned}$$

where (a) holds by considering the expression of $F(\cdot)$ in (8).

E. Proof of Proposition 2

First, we consider the case of $\|A\| < 1$. By applying 1) of Theorem 1 and choosing $t_0 = 1$ and $\bar{q}_0 = b_w + \max\{\beta, b_w + b_v + \|A\|p_0\}$, we have (13). From (6) and (8), we see that $F(\eta_0)$ is a monotonically non-decreasing function w.r.t. s. Thus, $f(s)$ is a monotonically nondecreasing function w.r.t. s.

Second, we consider the case of $\|A\| \geq 1$. In the case, we have (13), by applying 1) of Theorem 1, and by choosing $t_0 = 1$ and $\bar{q}_0 = q_0$. Next, we show the $f(s)$ is a monotonically nondecreasing function w.r.t. s. As discussed previously that $F(\eta_0)$ is a monotonically nondecreasing function w.r.t. s, we just need to prove that q_0 is a monotonically nondecreasing function w.r.t. s. This is obviously ensured if $\beta > b_w + b_v + \|A\|p_0$. Next, we prove this point by contradiction. In other words, we assume $\beta \leq b_w + b_v + \|A\|p_0$. Note that (9) is equivalent to

$$1 - k^*(\eta_0)\frac{\lambda_0}{N} \leq \frac{1}{\|A\|}\left(1 - \frac{q_0}{\eta_0}\right). \tag{33}$$

Due to $\beta \leq b_w + b_v + \|A\|p_0$, we have $q_0 \geq \beta + b_w$. Then, a necessary to ensure (33) is

$$1 - k^*(\eta_0)\frac{\lambda_0}{N} \leq \frac{1}{\|A\|}\left(1 - \frac{\beta + b_w}{\eta_0}\right). \tag{34}$$

It follows from (8) that $k^*(\eta_0) = \frac{\beta}{\|A\|(p_0+\eta_0)+b_w+b_v}$. By substituting $k^*(\eta_0)$ into (34), we obtain

$$\frac{\beta}{\|A\|\,(p_0+\eta_0)+b_w+b_v}\frac{\lambda_0}{N} \geq \frac{\beta + b_w + (\|A\| - 1)\eta_0}{\|A\|\,\eta_0}$$

which cannot be satisfied due to $\lambda_0 \leq N$ and $1 \leq \|A\|$. Therefore, the assumption $\beta \leq b_w + b_v + \|A\|p_0$ does not hold.

F. Proof of Theorem 4

The proof is similar to the proofs of Theorems 1–3. In the following, we just show the main points of this proof.

Given a time $T > 0$ and the maximal number of the detected sensors at time t, i.e., $d(T)$, similar to the proof of Theorem 1 $\forall t \geq T$, we construct the following sequence $\{\bar{\rho}_t \in \mathbb{R}|\bar{\rho}_t\}$ in (17). It is straightforward to prove that $\forall t \geq T$, $\|\tilde{e}(t)\| \leq \bar{\rho}_t$, where $\tilde{e}(t) = \frac{1}{N}\sum_{i=1}^N \hat{x}_i(t) - x(t)$. Next, we study the relationship between $\bar{\rho}_t$ in (17) and ρ_t in (7). Due to $\bar{\rho}_T = \rho_T$, we have $\bar{\rho}_{T+1} = \rho_{T+1} - \frac{d(T)\beta}{N}$. Then, for $t = T + 2$, we have

$$\begin{aligned}
\bar{\rho}_{T+2} &= F(\bar{\rho}_{T+1})\rho_{T+1} + q_0 - (F(\bar{\rho}_{T+1}) + 1)\frac{d(T)\beta}{N} \\
&\leq \rho_{T+2} - (F_* + 1)\frac{d(T)\beta}{N}
\end{aligned}$$

where $F_* = \inf_{t_0\in\bar{\Gamma}} F(\bar{\rho}_{t_0}) \in [0, 1)$, and $\bar{\Gamma} = \{t \geq T|\bar{\rho}_t \leq \bar{\rho}_{t-1}\}$. By recursively applying the aforementioned operation, for $t \geq T$, we obtain

$$\bar{\rho}_t \leq \rho_t - \frac{d(T)\beta}{N}\left(\frac{1 - F_*^{t-T}}{1 - F_*}\right).$$

Then, by Lemma 1 and Theorem 1, we have $\limsup_{t\to\infty}\|\tilde{e}(t)\| \leq \inf_{t_0\in\Gamma}\rho_{t_0} - \frac{d(T)\beta}{N(1-F_*)}$. Thus, the first conclusion holds by applying Lemma 1, $\|e_i(t)\| \leq \|\tilde{e}(t)\| + \|\bar{e}_i(t)\|$, and (20).

G. Proof of Theorem 5

Under condition 1), owing to the connectivity of the network $\mathcal{G}$, there is a common time $\tilde{t} \geq \hat{t}_0$, such that $d_j(\tilde{t}) = s \, \forall j \in \mathcal{V}$. For $t \geq \tilde{t}$, all the observations of the attacked sensors are discarded. Then, we have the compact form of recursive state estimates of Algorithm 2 in the following:

$$\hat{X}(t) = (I_{Nn} - \alpha(\mathcal{L} \otimes I_n))^L \Big[(I_N \otimes A)\hat{X}(t-1) + \bar{C}^T \bar{K}_{\mathcal{A}^c}(Y(t) - \bar{C}(I_N \otimes A)\hat{X}(t-1)) \Big]. \quad (35)$$

Let $\bar{E}(t) = \hat{X}(t) - \mathbf{1}_N \otimes \hat{x}_{avg}(t)$, i.e., $\bar{E}(t) = [\bar{e}_1^T(t), \ldots, \bar{e}_N^T(t)]^T$. By referring to [27], we have

$$\begin{aligned} &\left\|\bar{E}(t+1)\right\| \\ &\leq \|(I_N \otimes A)\| \left\| (I_{Nn} - \alpha(\mathcal{L} \otimes I_n) - P_{Nn})^L \bar{E}(t) \right\| \\ &\quad + \left\| (I_{Nn} - P_{Nn}) (I_{Nn} - \alpha(\mathcal{L} \otimes I_n))^L \right. \\ &\qquad \left. \bar{C}^T \bar{K}_{\mathcal{A}^c}(Y(t) - \bar{C}(I_N \otimes A)\hat{X}(t-1)) \right\| \\ &\leq 2\|A\|\gamma^L \left\|\bar{E}(t)\right\| + \|A\|\gamma^L \sqrt{N-s} \, \|\tilde{e}(t)\|. \end{aligned} \quad (36)$$

Similar to (22), for $i \in \mathcal{A}^c, t \geq \tilde{t}, k_i(t) = 1$, we have $\tilde{e}(t+1) = M_{21}\tilde{e}(t) - M_{22}\bar{E}(t)$, where $M_{21} = (I_n - \frac{1}{N}\sum_{i \in \mathcal{A}^c} C_i^T C_i)A$ and $M_{22} = \frac{1}{N}(\mathbf{1}_N^T \otimes I_n)\bar{C}^T \bar{K}_{\mathcal{A}^c} \bar{C}(I_N \otimes A)$. Then, it holds that

$$\|\tilde{e}(t+1)\| \leq \tau_0 \left\|\bar{E}(t)\right\| + \varpi \|\tilde{e}(t)\| \quad (37)$$

where ϖ is given in Theorem 3. By (36) and (37), if the matrix $\begin{pmatrix} 2\|A\|\gamma^L & \|A\|\gamma^L\sqrt{N-s} \\ \tau_0 & \varpi \end{pmatrix}$ is Schur stable, $\|\tilde{e}(t)\|$ and $\|\bar{E}(t)\|$ go to zero asymptotically. Thus, $\|e_i(t)\|$ is convergent to zero as time goes to infinity.

Acknowledgment

The authors are grateful to the anonymous reviewers for their insightful comments and suggestions.

References

[1] G. Battistelli, L. Chisci, G. Mugnai, A. Farina, and A. Graziano, "Consensus-based linear and nonlinear filtering," *IEEE Trans. Autom. Control*, vol. 60, no. 5, pp. 1410–1415, May 2015.

[2] Q. Liu, Z. Wang, X. He, and D. Zhou, "Event-based distributed filtering with stochastic measurement fading," *IEEE Trans. Ind. Inform.*, vol. 11, no. 6, pp. 1643–1652, Dec. 2015.

[3] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Trans. Autom. Control*, vol. 58, no. 11, pp. 2715–2729, Nov. 2013.

[4] X. Ren, Y. Mo, J. Chen, and K. H. Johansson, "Secure state estimation with sensors: A probabilistic approach," *IEEE Trans. Autom. Control*, vol. 65, no. 9, pp. 3742–3757, Sep. 2020.

[5] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Trans. Autom. Control*, vol. 59, no. 6, pp. 1454–1467, Jun. 2014.

[6] M. Pajic, I. Lee, and G. J. Pappas, "Attack-resilient state estimation for noisy dynamical systems," *IEEE Control Netw. Syst.*, vol. 4, no. 1, pp. 82–92, Mar. 2017.

[7] M. Pajic, J. Weimer, N. Bezzo, O. Sokolsky, G. J. Pappas, and I. Lee, "Design and implementation of attack-resilient cyberphysical systems: With a focus on attack-resilient state estimators," *IEEE Control Syst. Mag.*, vol. 37, no. 2, pp. 66–81, Apr. 2017.

[8] Y. Shoukry, P. Nuzzo, A. Puggelli, A. L. Sangiovanni-Vincentelli, S. A. Seshia, and P. Tabuada, "Secure state estimation for cyber-physical systems under sensor attacks: A satisfiability modulo theory approach," *IEEE Trans. Autom. Control*, vol. 62, no. 10, pp. 4917–4932, Oct. 2017.

[9] Y. Shoukry *et al.*, "SMT-based observer design for cyber-physical systems under sensor attacks," *ACM Trans. Cyber-Phys. Syst.*, vol. 2, no. 1, pp. 1–27, 2018.

[10] D. Han, Y. Mo, and L. Xie, "Convex optimization based state estimation against sparse integrity attacks," *IEEE Trans. Autom. Control*, vol. 64, no. 6, pp. 2383–2395, Jun. 2019.

[11] A. Mitra and S. Sundaram, " Distributed observers for systems," *Automatica*, vol. 108, 2019, Art. no. 108487.

[12] A. Mitra, J. A. Richards, S. Bagchi, and S. Sundaram, "Resilient distributed state estimation with mobile agents: Overcoming adversaries, communication losses, and intermittent measurements," *Auton. Robots*, vol. 43, no. 3, pp. 743–768, 2019.

[13] L. Su and S. Shahrampour, "Finite-time guarantees for Byzantine-resilient distributed state estimation with noisy measurements," *IEEE Trans. Autom. Control*, vol. 65, no. 9, pp. 3758–3771, Sep. 2020.

[14] P. Blanchard, R. Guerraoui, J. Stainer, and E. M. EI Mhamdi, "Machine learning with adversaries: Tolerant gradient descent," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 119–129.

[15] M. Deghat, V. Ugrinovskii, I. Shames, and C. Langbort, "Detection and mitigation of biasing attacks on distributed estimation networks," *Automatica*, vol. 99, pp. 369–381, 2019.

[16] Y. Chen, S. Kar, and J. M. F. Moura, "Resilient distributed estimation: Sensor attacks," *IEEE Trans. Autom. Control*, vol. 64, no. 9, pp. 3772–3779, Sep. 2019.

[17] Y. Chen, S. Kar, and J. M. Moura, "Resilient distributed parameter estimation with heterogeneous data," *IEEE Trans. Signal Process.*, vol. 67, no. 19, pp. 4918–4933, Oct. 2019.

[18] L. An and G.-H. Yang, "Distributed secure state estimation for cyber-physical systems under sensor attacks," *Automatica*, vol. 107, pp. 526–538, Jan. 2019.

[19] M. S. Chong, M. Wakaiki, and J. P. Hespanha, "Observability of linear systems under adversarial attacks," in *Proc. IEEE Amer. Control Conf.*, 2015, pp. 2439–2444.

[20] B. Chen, D. W. Ho, W.-A. Zhang, and L. Yu, "Distributed dimensionality reduction fusion estimation for cyber-physical systems under attacks," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 49, no. 2, pp. 455–468, Feb. 2019.

[21] F. Boem, A. J. Gallo, G. Ferrari-Trecate, and T. Parisini, "A distributed attack detection method for multi-agent systems governed by consensus-based control," in *Proc. IEEE Conf. Decis. Control*, 2017, pp. 5961–5966.

[22] A. J. Gallo, M. S. Turan, F. Boem, T. Parisini, and G. Ferrari-Trecate, "A distributed cyber-attack detection scheme with application to microgrids," *IEEE Trans. Autom. Control*, vol. 65, no. 9, pp. 3800–3815, Sep. 2020.

[23] N. Forti, G. Battistelli, L. Chisci, S. Li, B. Wang, and B. Sinopoli, "Distributed joint attack detection and secure state estimation," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 4, no. 1, pp. 96–110, Mar. 2018.

[24] Y. Nakahira and Y. Mo, "Attack-resilient $\mathcal{H}_2$, $\mathcal{H}_\infty$, and $\mathcal{H}_1$ state estimator," *IEEE Trans. Autom. Control*, vol. 63, no. 12, pp. 4353–4360, Dec. 2018.

[25] J. G. Lee, J. Kim, and H. Shim, "Fully distributed resilient state estimation based on distributed median solver," *IEEE Trans. Autom. Control*, vol. 65, no. 9, pp. 3935–3942, Sep. 2020.

[26] X. He, X. Ren, H. Sandberg, and K. H. Johansson, "Secure distributed filtering for unstable dynamics under compromised observations," in *Proc. IEEE Conf. Decis. Control*, 2019, pp. 5344–5349.

[27] X. He, X. Ren, H. Sandberg, and K. H. Johansson, "Design of secure filters under attacked measurements: A saturation method," 2020. [Online]. Available: https://www.researchgate.net/publication/340579534.

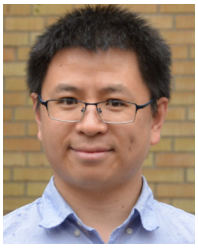
[28] M. Mesbahi and M. Egerstedt, *Graph Theoretic Methods in Multiagent Networks*. Princeton, NJ, USA: Princeton Univ. Press, 2010.

[29] A. Mitra and S. Sundaram, "Secure distributed state estimation of an system over time-varying networks and analog erasure channels," in *Proc. IEEE Amer. Control Conf.*, 2018, pp. 6578–6583.

[30] C. Zhao, J. He, and J. Chen, "Resilient consensus with mobile detectors against malicious attacks," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 4, no. 1, pp. 60–69, Mar. 2017.

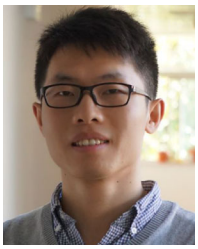
[31] Y. Shoukry and P. Tabuada, "Event-triggered state observers for sparse sensor noise/attacks," *IEEE Trans. Autom. Control*, vol. 61, no. 8, pp. 2079–2091, Aug. 2016.

[32] Z. Guo, D. Shi, K. H. Johansson, and L. Shi, "Optimal linear cyber-attack on remote state estimation," *IEEE Control Netw. Syst.*, vol. 4, no. 1, pp. 4–13, Mar. 2017.
[33] U. A. Khan and A. Jadbabaie, "Collaborative scalar-gain estimators for potentially unstable social dynamics with limited communication," *Automatica*, vol. 50, no. 7, pp. 1909–1914, 2014.



Xingkang He (Member, IEEE) received the B.S. degree from the Hefei University of Technology, Hefei, China, in 2013, and the Ph.D. degree from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China, in 2018.

He is a Postdoctoral Researcher with the Division of Decision and Control Systems, KTH Royal Institute of Technology, Stockholm, Sweden. His research interests include security and privacy of cyber-physical systems, estimation and control of networked systems, and social networks.



Xiaoqiang Ren (Member, IEEE) received the B.E. degree in automation from Zhejiang University, Hangzhou, China, in 2012, and the Ph.D. degree in control and dynamic systems from the Hong Kong University of Science and Technology, Hong Kong, in 2016.

He is a Professor with the School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, China. Prior to his current position, he was a Postdoctoral Researcher with the Hong Kong University of Science and Technology in 2016, Nanyang Technological University from 2016 to 2018, and KTH Royal Institute of Technology from 2018 to 2019. His research interests include security of cyber-physical systems, sequential decision, and networked estimation and control.

Henrik Sandberg (Member, IEEE) received the M.Sc. degree in engineering physics and the Ph.D. degree in automatic control from Lund University, Lund, Sweden, in 1999 and 2004, respectively.

He is a Professor with the Division of Decision and Control Systems, KTH Royal Institute of Technology, Stockholm, Sweden. From 2005 to 2007, he was a Postdoctoral Scholar with the California Institute of Technology, Pasadena, CA, USA. In 2013, he was a Visiting Scholar with the Laboratory for Information and Decision Systems (LIDS), Massachusetts Institute of Technology, Cambridge, MA, USA. He has also held visiting appointments with the Australian National University, Canberra, Australia, and the University of Melbourne, Melbourne, Australia. His current research interests include security of cyber-physical systems, power systems, model reduction, and fundamental limitations in control.

Dr. Sandberg was the recipient of the Best Student Paper Award from the IEEE Conference on Decision and Control in 2004, an Ingvar Carlsson Award from the Swedish Foundation for Strategic Research in 2007, and a Consolidator Grant from the Swedish Research Council in 2016. He has served on the editorial board for the IEEE TRANSACTIONS ON AUTOMATIC CONTROL and the *IFAC Journal Automatica*.

Karl Henrik Johansson (Fellow, IEEE) received the M.Sc. and Ph.D. degrees from Lund University, Lund, Sweden, in 1992 and 1997, respectively.

He is Professor with the School of Electrical Engineering and Computer Science and the Director with Digital Futures, KTH Royal Institute of Technology, Stockholm, Sweden. He has held visiting positions with UC Berkeley; California Institute of Technology; Nanyang Technological University; Institute of Advanced Studies, Hong Kong University of Science and Technology; and Norwegian University of Science and Technology. His research interests include networked control systems and cyber-physical systems with applications in transportation, energy, and automation networks.

Dr. Johansson is a member of the Swedish Research Council's Scientific Council for Natural Sciences and Engineering Sciences. He has served on the IEEE Control Systems Society Board of Governors, the International Federation of Automatic Control (IFAC) Executive Board, and is currently the Vice-President of the European Control Association. He was the recipient of best paper awards and other distinctions from the IEEE, IFAC, and Association for Computing Machinery. He has been awarded as Distinguished Professor with the Swedish Research Council and Wallenberg Scholar with the Knut and Alice Wallenberg Foundation. He was also the recipient of the Future Research Leader Award from the Swedish Foundation for Strategic Research and the triennial Young Author Prize from IFAC. He is a Fellow of the Royal Swedish Academy of Engineering Sciences, and he is an IEEE Control Systems Society Distinguished Lecturer.