2019 IEEE 58th Conference on Decision and Control (CDC)
Palais des Congrès et des Expositions Nice Acropolis
Nice, France, December 11-13, 2019

Secure Distributed Filtering for Unstable Dynamics Under Compromised Observations

Xingkang He, Xiaoqiang Ren, Henrik Sandberg, Karl H. Johansson

Abstract— **In this paper, we consider a secure distributed filtering problem for linear time-invariant systems with bounded noises and unstable dynamics under compromised observations. A malicious attacker is able to compromise a subset of the agents and manipulate the observations arbitrarily. We first propose a recursive distributed filter consisting of two parts at each time. The first part employs a saturation-like scheme, which gives a small gain if the innovation is too large. The second part is a consensus operation of state estimates among neighboring agents. A sufficient condition is then established for the boundedness of estimation error, which is with respect to network topology, system structure, and the maximal compromised agent subset. We further provide an equivalent statement, which connects to 2s-sparse observability in the centralized framework in certain scenarios, such that the sufficient condition is feasible. Numerical simulations are finally provided to illustrate the developed results.**

I. Introduction

Cyber-physical systems (CPSs) are systems controlled and monitored by computer-based algorithms. Through a CPS, physical processes and cyber components can be effectively integrated. During the recent years, numerous applications of CPSs such as sensor network, vehicle network, process control, smart grid, etc, have been well investigated in academia and industry. With higher integration of large-scale computer networks and complex physical processes, the CPSs are confronting more security issues both in software and physical layers. Thus, the research topics on CPS security are attracting more and more attention.

In a CPS, sensor observations can be utilized to obtain state estimate or to design output feedback signal to control the physical process. Due to the vulnerabilities of sensors, the malicious attacker may insert faulty data into observations of the compromised sensors. Then, the estimates or controller based on the compromised observations will be unreliable, and even bring tremendous damage to the whole system. Thus, some detection and identification schemes are considered to find out whether the sensors are under attack, and if so how to identify the attack signals inserted to the systems. A study on attack detection and identification for CPSs was given in [1], where the design methods and analysis techniques for centralized and distributed monitors were discussed as well. In [2], the joint distributed attack detection and state estimation were investigated in a Bayesian framework. To obtain attack-resilient state estimates, in the centralized framework, some state estimators or observers were proposed based on optimization techniques [3]–[8], recursive implementation [9], and probabilistic approach [10]. Compared with centralized methods, on one hand, the distributed ones have advantages in quite a few aspects, such as the structure robustness, energy saving and parallel processing. On the other hand, in the distributed framework, since each agent has limited information from local observations and neighboring communications, the distributed state estimation methods are essentially different from the centralized ones. In the distributed state estimation under compromised sensors, observer-based methods were studied for byzantine attacks, under which the compromised sensors can send faulty information to other normal sensors [11]. In [12], a distributed observer with attack detection layer was proposed to deal with a class of biasing attacks. Distributed estimation for a static parameter under compromised observations was studied in [13], where the sparse-observability condition was required to guarantee the consistency of the estimator.

In this paper, we study the secure distributed filtering or estimation problem for linear time-invariant systems with bounded noises and unstable dynamics. The main contributions of this paper are three-fold. 1) We investigate the secure distributed filtering problem under compromised observations. Unlike [12], we allow that the malicious attacker manipulates the observations arbitrarily for an unknown subset of the agents. Different from [11], [14] requiring some robustness of communication graph, we simply assume the connectivity of the graph. 2) We propose a novel secure distributed filtering framework consisting of two parts, which is essentially different from the centralized methods [3]–[10] or the distributed methods [11], [12]. The first part employs a saturation-like scheme, which gives a small gain if the innovation is too large. The second part is a consensus operation of state estimates among neighboring agents. 3) We establish a sufficient condition for the boundedness of estimation error and provide an equivalent statement, which connects to 2s-sparse observability in the centralized framework in certain scenarios, such that the sufficient condition is feasible.

The remainder of the paper is organized as follows: Section II is on preliminaries and problem formulation. Section III considers the secure distributed filter. Section IV provides

The work is supported by Knut & Alice Wallenberg foundation, and by Swedish Research Council.

X. He, H. Sandberg and K. H. Johansson are with Division of Decision and Control Systems, School of Electrical Engineering and Computer Science. KTH Royal Institute of Technology, Sweden ((xingkang,hsan,kallej)@kth.se).

X. Ren is with the Department of Automation, School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, 200444, China. He was with the Division of Decision and Control Systems, School of Electrical Engineering, Royal Institute of Technology, 114 28 Stockholm, Sweden when the manuscript was writing. (xiaoqren@kth.se)

978-1-7281-1398-2/19/$31.00 ©2019 IEEE

the performance analysis for the filter. Section V gives the numerical simulation results. The conclusions of this paper are given in Section VI.

II. Problem Formulation

A. Notations

The superscript "T" represents the transpose. $\mathbb{R}^{n\times m}$ is the set of real matrices with n rows and m columns. $\mathbb{R}^n$ is the n-dimensional Euclidean space. I_n stands for the n-dimensional square identity matrix. $\mathbf{1}_N$ stands for the N-dimensional vector with all elements being one. $\mathrm{diag}\{\cdot\}$ represents the diagonalization operator. $A\otimes B$ is the Kronecker product of A and B. $\|x\|_2$ is the 2-norm of a vector x. $\|A\|_2$ is the induced 2-norm, i.e., $\|A\|_2 = \sup\limits_{x\neq 0}\frac{\|Ax\|_2}{\|x\|_2}$. $\lambda_2(A)$ and $\lambda_{max}(A)$ are the second minimal eigenvalue and maximal eigenvalue of A, respectively. $|\Gamma|$ is the cardinality of the set Γ.

B. Graph Preliminaries

In an undirected graph $\mathcal{G} = (\mathcal{V},\mathcal{E})$, $\mathcal{V}$ stands for the set of nodes, $\mathcal{E}\subseteq\mathcal{V}\times\mathcal{V}$ is the set of edges. If there is an edge $(i,j)\in\mathcal{E}$, node i can exchange information with node j, and node j is called a neighbor of node i. Let the neighbor set of agent i be $\mathcal{N}_i := \{j\in\mathcal{V}|(i,j)\in\mathcal{E}\}$. The graph $\mathcal{G}$ is connected if for any pair of nodes (i_1,i_l), there exists a path from i_1 to i_l consisting of edges $(i_1,i_2),(i_2,i_3),\ldots,(i_{l-1},i_l)$. $\mathcal{L}$ is the Laplacian matrix whose definition is referred to [15]. On the connectivity of a graph, we have the following proposition.

Proposition 1: [15] The undirected graph $\mathcal{G}$ is connected if and only if $\lambda_2(\mathcal{L}) > 0$.

C. System model

Consider the following plant observed by N agents (e.g., sensors),

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

where $x(t)\in\mathbb{R}^n$ is the unknown system state, $w(t)\in\mathbb{R}^n$ is the process noise, $v_i(t)\in\mathbb{R}$ is the observation noise, and $a_i(t)\in\mathbb{R}$ is the attack signal inserted by some malicious attacker, all at time t. $y_i(t)\in\mathbb{R}$ is the observation of agent i. 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 agent i.

Remark 1: The essential problem is to study the influence of scalar attack signal to the estimation performance with certain number of compromised observation elements, like [13]. Thus, we consider the observation equation with scalar outputs for each agent. This conforms with the centralized framework, where each row vector of centralized observation matrix stands for the observation vector of one agent.

Definition 1: (One-step collective observability) The system (1) is called one-step collectively observable if $\sum_{i=1}^N C_i^T C_i$ is a positive definite matrix.

Remark 2: On the relation between one-step collective observability, which requires $N\geq n$, and n-step collective observability (i.e., (A,C) is observable, where $C = [C_1^T,\ldots,C_N^T]$): If A is a diagonal matrix such as $A = I_n$, the two definitions are equivalent. For general system matrices, n-step collective observability is milder than the one-step collective observability. Notice also that one-step collective observability does not mean local observability, i.e., (A,C_i) could be unobservable or undetectable, $\forall i = 1,\ldots,N$.

In this paper, the following assumptions are in need.

Assumption 1: The following conditions hold

$$\begin{aligned} &\|A\|_2 = a\geq 1, \|w(t)\|_2\leq b_w, \|v_i(t)\|_2\leq b_v,\\ &\|\hat{x}_i(0)-x(0)\|_2\leq\eta_i\leq\eta_0, i = 1,\cdots,N, \end{aligned}$$

where $\hat{x}_i(0)$ is the estimate of $x(0)$ by agent i. Besides, the bounds are known to each agent.

Assumption 2: The system (1) is one-step collectively observable, i.e., $\sum_{i=1}^N C_i^T C_i \succ 0$. The observation vector C_i is normalized, i.e., $\|C_i\|_2 = 1, i = 1,\cdots,N$.

Assumption 3: The communication graph $\mathcal{G} = (\mathcal{V},\mathcal{E})$ is undirected and connected, where $\mathcal{V} = \{1,2,\ldots,N\}$

Remark 3: To design a non-trivial filtering algorithm with guaranteed bounded estimator error, we assume $\|A\|_2\geq 1$ in Assumption 1. Otherwise, one can easily design a filter such that estimation errors keep bounded. The proposed methods and results also apply to the case where $\|A\|_2 < 1$. Assumption 2 requires a collective observability condition utilized in the existing literature on distributed estimation [16]–[18]. The normalized observation vectors can be obtained by reconstructing the system (1). Different from [11] requiring some robustness of communication graph, the connectivity of Assumption 3 is a standard condition for distributed estimation. If the graph is not connected, the problem can be studied for the connected subgraphs separately.

A typical distributed filtering problem is to design an online filter or state estimator for each agent (e.g., agent i) to estimate the system state $x(t)$ by employing the known local noisy observations $\{y_i(l)\}_{l=1}^t$ and the messages received from neighboring agents. However, if observations of some agents are compromised by a malicious attacker, the observation quality may be tremendously affected, which will bring big challenges in design and analysis of distributed filtering algorithms. In the following, we introduce the attack model.

D. Attack model

To deteriorate the estimation performance of filtering algorithms, the malicious attacker aims to persistently destroy the observation data of some targeted agents. However, due to resource limitation, the attacker has limited power to attack the set of agents. Assume that the set of compromised agents is fixed over time, and consists of no more than s agents. Since the knowledge of the attacker makes a big difference to its ability in deteriorating the estimation performance, we assume the following knowledge scope of the attacker.

Assumption 4: The attacker has full knowledge on the system (1), the network topology, and the filter of all agents. Furthermore, the observation $a_i(t)$ can be arbitrary for a compromised agent i.

Under Assumption 4, we have

$$\begin{aligned} &a_i(t) \in R, i \in \mathcal{A}, \text{ with } |\mathcal{A}| \leq s \\ &a_i(t) = 0, i \in \mathcal{N} = \mathcal{V} - \mathcal{A}, \forall t \in \mathbb{N}, \end{aligned} \tag{2}$$

where $\mathcal{A}$ is the set of agents whose observations are compromised by the malicious attacker. $\mathcal{N}$ is the set of normal agents without being affected by the attacker. Note that the sets $\mathcal{A}$ and $\mathcal{N}$ are unknown to each agent.

Remark 4: In Assumption 4, we consider the worst scenario on compromised observations that the attacker can access the full information without requiring any concrete attack models, which is more general than results in the existing literature [12].

We further require the following definitions.

Definition 2: (s-sparse observability) 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 $[C_1^T, C_2^T, \ldots, C_N^T]$.

Definition 3: (One-step s-sparse observability) The linear system defined by (1) is said to be one-step s-sparse observable if for every set $\Gamma \subseteq \{1, \ldots, N\}$ with $|\Gamma| = s$, the pair $C_{\bar{\Gamma}}^T C_{\bar{\Gamma}} = \sum_{i=1, i \notin \Gamma}^N C_i^T C_i \succ 0$, where $C_{\bar{\Gamma}}$ is the remaining matrix by removing $C_j, j \in \Gamma$ from $[C_1^T, C_2^T, \ldots, C_N^T]$.

Remark 5: Definition 2 and Definition 3 correspond to the n-step (collective) observability (i.e., (A, C) is observable) and one-step (collective) observability in Definition 1. If the system matrix A is diagonal, Definition 2 and Definition 3 are equivalent. In the centralized framework, if the observations of s agents are compromised, the system should be $2s$-sparse observable to guarantee the effective estimation of system state [19]. The connection between $2s$-sparse observability and the condition required in this work will be investigated in Theorem 2 and Lemma 1.

E. Problems of interest

We mainly consider the following problems in this paper.

1) How to design secure distributed filter for each agent by employing the local noisy observations potentially compromised by the malicious attacker?

2) What conditions can guarantee the bounded estimation error of the distributed filter in presence of the attacker (2). How can we quantify the estimation performance of the distributed filter?

III. Secure Distributed Filter

In this section, we will design a secure distributed filter for each agent.

We consider the filtering algorithm with two stages, namely, observation update and consensus. In the stage of local observation update, we design a saturation-like scheme to utilize the observation $y_i(t)$ as follows

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

where

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

Different from the gain designs of common filters or state estimators, the gain $k_i(t)$ in this work is related to the value of innovation (i.e., $y_i(t) - C_i A \hat{x}_i(t-1)$). The design of $k_i(t)$ in (4) makes sense, since if the estimation innovation is very large, the observation $y_i(t)$ is more likely to be compromised. By the designed gain $k_i(t)$, we have $|k_i(t)(y_i(t) - C_i A \hat{x}_i(t-1))| \leq \beta$, which ensures that the attacker has limited influence to the local update stage of the filter.

In the consensus stage, we suppose that each agent can communicate with its neighbors for $L \geq 1$ times between two time instants. For $l = 1, 2, \ldots, L$,

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

with $\hat{x}_{i,0}(t) = \tilde{x}_i(t)$ and we denote $\hat{x}_i(t) = \hat{x}_{i,L}(t)$. For each communication, agent j will transmit its estimate $\hat{x}_{j,l-1}(t)$ to its neighbors, $l = 1, \ldots, N$.

Remark 6: The parameter β in (4) reflects the usage tradeoff between normal observations and compromised observations. If β is very large, then almost all normal observations will be utilized without scaling. But, it will give much space that the attacker can use to deteriorate the estimation performance. If β is very small, although the most possible attack signals may be filtered by the designed gain $k_i(t)$, many normal observations will contribute little to the estimation performance. As a result, the filtering error of each agent will probably be divergent due to $\|A\| \geq 1$. The condition on β will be discussed in next section.

Remark 7: The term $\alpha \sum_{j \in \mathcal{N}_i} (\hat{x}_{i,l-1}(t) - \hat{x}_{j,l-1}(t))$ is to make the agents reach consensus. The consensus step is vital to guarantee bounded estimation error of distributed filters especially for the case that each subsystem is not observable (i.e., (A, C_i) is not observable). The parameter α can increase the consensus speed if it is well designed. It can be proven that if the consensus step L goes to infinity and the parameter α is properly designed, then the estimates $\{\hat{x}_i(t)\}_{i=1}^N$ will converge to the same vector. However, different from resilient algorithms in [20], [21] which require the states or estimates reach consensus, the consensus step L is not required to approximate infinity in this work. The requirement of the step L and the design of α, related with the system structure and performance demand, is given in next section.

By (3), (4) and (5), we obtain the secure distributed consensus filter (SDCF) in Algorithm 1.

IV. Performance analysis

In this section, we will focus on performance analysis of the proposed SDCF. Specifically, we will study the conditions to guarantee the boundedness of estimation error, and quantify the estimation performance under compromised observations.

Since the filtering gains $\{k_i(t)\}$ are related to the state estimates and potential compromised observations, the common stability analysis approaches, such as Lyapunov methods, may not be directly utilized to analyze the stability or

Algorithm 1 Secure Distributed Consensus Filter (SDCF):

```
Update: Agent $i$ uses its own observation to update the estimate
    $\tilde{x}_i(t) = A\hat{x}_i(t-1) + k_i(t)C_i^T(y_i(t) - C_iA\hat{x}_i(t-1))$
    $k_i(t) = \min\{1, \frac{\beta}{|y_i(t)-C_iA\hat{x}_i(t-1)|}\}$,
Consensus for $L$ steps: $\hat{x}_{i,0}(t) = \tilde{x}_i(t)$
    For $l$th consensus, $l = 1, \dots, L$:
        For $i$th agent, $i = 1, \dots, N$:
            Agent $i$ receives $\hat{x}_{j,l-1}(t)$ from neighbor agent $j$,
            $\hat{x}_{i,l}(t) = \hat{x}_{i,l-1}(t) - \alpha\sum_{j\in\mathcal{N}_i}(\hat{x}_{i,l-1}(t) - \hat{x}_{j,l-1}(t))$
        end
    end
Output step: $\hat{x}_i(t) = \hat{x}_{i,L}(t)$.
```

boundedness of estimation error by its dynamics. This is the main challenge for the problem of distributed recursive filter under compromised observations.

A. Boundedness of estimation error

In this subsection, we will study the conditions to guarantee the boundedness of estimation error for the SDCF in Algorithm 1. Denote $\lambda_0 := \lambda_{min}\left(\sum_{i\in\mathcal{N}^*} C_i^T C_i\right)$, where $\mathcal{N}^*$ is the agent set such that $\lambda_{min}\left(\sum_{i\in\mathcal{N}} C_i^T C_i\right)$ is minimal within all sets $\{\mathcal{N}\}$ obtained by removing any $|\mathcal{A}|$ agents from $\mathcal{V}$. Besides, for convenience, we give the following notations

$$\begin{aligned}
\gamma &= \frac{\lambda_{max}(\mathcal{L}) - \lambda_2(\mathcal{L})}{\lambda_{max}(\mathcal{L}) + \lambda_2(\mathcal{L})}, \\
p_0^* &= a\gamma^L\sqrt{N}\eta_0 + \frac{\sqrt{N}\beta\gamma^L}{1-a\gamma^L}, \\
k^* &= \min\{1, \frac{\beta}{a(p_0^*+\eta_0)+b_w+b_v}\}, \\
\mu_0 &= a\left(1 - \frac{k^*}{N}\lambda_0\right), \qquad (6) \\
Q_0 &= (1 - \frac{|\mathcal{A}|}{N})(b_w + b_v + ap_0^*) + b_w, \\
\vartheta_0 &= 1 - \frac{Q_0}{\eta_0}\left(1 - \frac{\beta|\mathcal{A}|}{N\eta_0}\right)^{-1}, \\
m_0 &= \vartheta_0\left(1 - \frac{\beta|\mathcal{A}|}{N\eta_0}\right)\left(1 - \frac{k^*\lambda_0}{N}\right)^{-1}.
\end{aligned}$$

On the boundedness of estimation error by SDCF, we have the following result.

***Theorem** 1:* Let Assumptions 1 - 3 hold and $\alpha = \frac{2}{\lambda_2(\mathcal{L})+\lambda_{max}(\mathcal{L})}$. If there exist a set of scalars $L > 0$, $\beta > 0$, $\eta_0 > 0$, such that

$$1 \le a < \min\left\{m_0, \gamma^{-L}\right\}, \qquad (7)$$

then the estimation error of SDCF(L, β), i.e., $e_i(t) = \hat{x}_i(t) - x(t)$, $\forall i \in \mathcal{V}$, satisfies

$$\lim_{t\to\infty}\|e_i(t)\|_2 \le \frac{NQ_0 + |\mathcal{A}|\beta}{N(1-\mu_0)} + \frac{\sqrt{N}\beta\gamma^L}{1-a\gamma^L} < \infty. \qquad (8)$$

Proof: See Subsection IV-C. ■

***Remark** 8:* The parameters β, L are given in the implementation of SDCF. Although η_0 is a bound of initial estimation error, we can adjust it bigger to meet the requirement. Under Assumption 3 and Proposition 1, we have $\gamma \in (0, 1)$, then we obtain $\gamma^{-L} > 1$ and $\lim\limits_{L\to\infty}\gamma^{-L} = +\infty$. Also, $m_0 > 0$ is guaranteed if $\beta < \eta_0$.

***Remark** 9:* Theorem 1 shows that by taking proper parameters L, β, η_0, the SDCF (i.e., Algorithm 1) can guarantee the boundedness of estimator error for a class of unstable dynamics. The condition (7) can be examined offline with global knowledge. To improve the estimation performance, one way is to lower the bound in (8) by designing a proper parameter β offline based on global knowledge.

B. Feasibility of condition (7)

Since the condition (7) is complex, its feasiblity needs to be testified, i.e., whether there exist a set of positive parameters L, β, η_0 such that (7) is satisfied. In this subsection, we study the feasibility of (7).

***Theorem** 2:* Condition (7) has a feasible solution on β, η_0 and L, if and only if

$$\lambda_0 > |\mathcal{A}|. \qquad (9)$$

Proof: We prove the conclusion from sufficiency and necessity. 1) Sufficiency. If $\lambda_0 > |\mathcal{A}|$, we have $\frac{\lambda_0}{|\mathcal{A}|} > 1$. For a small β and large η_0, we have $k^* = \frac{\beta}{a(p_0^*+\eta_0)+b_w+b_v} < 1$. Then we consider $\frac{\beta}{\eta_0 k^*} = \frac{a(p_0^*+\eta_0)+b_w+b_v}{\eta_0} > 1$. By choosing a sufficiently large L and η_0, let a approximate 1 sufficiently from the side larger than 1. Then, $\frac{\beta}{\eta_0 k^*}$ can approximate 1 sufficiently. As a result, we can guarantee $\frac{\lambda_0}{|\mathcal{A}|} > \frac{\beta}{\eta_0 k^*}$, which means $\left(1 - \frac{\beta|\mathcal{A}|}{N\eta_0}\right)\left(1 - \frac{k^*\lambda_0}{N}\right)^{-1} > 1$. Besides, we can choose a sufficiently large η_0, such that ϑ_0 approximates 1 from the side smaller than 1. Then, $m_0 > 1$, which means we can find feasible scalars β, η_0, L such that for any $\|A\|_2 = a \in [1, m_0)$, the condition (7) is satisfied. Therefore, the sufficiency holds. 2) Necessity. We use the contradiction method. If $\lambda_0 > |\mathcal{A}|$ does not hold, i.e., $\lambda_0 \le |\mathcal{A}|$, then we have $\left(1 - \frac{\beta|\mathcal{A}|}{N\eta_0}\right)\left(1 - \frac{k^*\lambda_0}{N}\right)^{-1} < 1$ due to $k^* < \frac{\beta}{\eta_0}$. Since $\vartheta_0 < 1$, then $m_0 < 1$, which means that there is no feasible $a \ge 1$, such that the condition (7) is satisfied. Thus, the necessity holds. ■

***Remark** 10:* Recall $\lambda_0 := \lambda_{min}\left(\sum_{i\in\mathcal{N}^*} C_i^T C_i\right)$, which reflects the one-step sparse observability of the system by removing any $|\mathcal{A}|$ agents. Since the compromised subset of agents is fixed over time, we can calculate λ_0 and compare it with $|\mathcal{A}|$.

The direct relationship between (9) and the one-step sparse observability is given in the following.

***Lemma** 1:* A necessary condition to guarantee $\lambda_0 > s := |\mathcal{A}|$ is that the system (1) is one-step $2s$-sparse observable. If the observation vectors are orthogonal and A is a diagonal matrix, then one-step $2s$-sparse observability is also a sufficient condition to guarantee $\lambda_0 > |\mathcal{A}|$.

Proof: The proof of this lemma can refer to [13]. ■

***Remark** 11:* From Lemma 1 and Theorem 2, on Algorithm 1, we have that if the observations of any s agents are under attacks, the system (1) should be one-step $2s$-sparse observable to achieve the effective estimation of system state. For the case that A is a diagonal matrix, the condition (9) conforms to the centralized framework that the system is $2s$-sparse observable [19].

C. Proof idea of Theorem 1

In this section, we provide the main idea for the proof of Theorem 1. Let $e_i(t)$ be the estimation error of agent i by Algorithm 1, i.e., $e_i(t) = \hat{x}_i(t) - x(t)$. Then we have $e_i(t) = \hat{x}_i(t) - x(t) = \tilde{e}(t) + \bar{e}_i(t)$, where $\bar{e}_i(t) := \hat{x}_i(t) - \hat{x}_{avg}$, and $\tilde{e}(t) = \hat{x}_{avg} - x(t)$, and $\hat{x}_{avg}(t) := \frac{1}{N}\sum_{i=1}^{N}\hat{x}_i(t)$. The idea for analyzing the boundedness of estimation error $e_i(t)$ is to find conditions that can guarantee the boundedness of $\bar{e}_i(t)$ and $\tilde{e}(t)$ simultaneously. As a result, the boundedness of $e_i(t)$ can be guaranteed. In the following, Lemma 2 and Lemma 3 study the boundedness of $\bar{e}_i(t)$ and $\tilde{e}(t)$, respectively.

***Lemma** 2:* Consider Algorithm 1 with $L \geq 1$, and let Assumptions 1 - 3 hold. If $\alpha = \frac{2}{\lambda_2(\mathcal{L}) + \lambda_{max}(\mathcal{L})}$, and $\|A\|_2 = a < \gamma^{-L}$, then

$$\|\bar{e}_i(t)\|_2 \leq p^*(L,t), \tag{10}$$

where $p^*(L,t) = (a\gamma^L)^t\sqrt{N}\eta_0 + \sqrt{N}\beta\gamma^L\frac{1-a^{t-1}\gamma^{L(t-1)}}{1-a\gamma^L}$. Furthermore, $\sup_{t\geq 1}\{p^*(L,t)\} \leq a\gamma^L\sqrt{N}\eta_0 + \frac{\sqrt{N}\beta\gamma^L}{1-a\gamma^L} \triangleq p_0^* < \infty$, and

$$\lim_{L\to\infty} p^*(L,t) = 0,$$
$$\lim_{t\to\infty} p^*(L,t) = \frac{\sqrt{N}\beta\gamma^L}{1-a\gamma^L} < \infty.$$

***Lemma** 3:* Consider Algorithm 1 with $L \geq 1$, and assume that Assumptions 1 - 3 hold, $\alpha = \frac{2}{\lambda_2(\mathcal{L}) + \lambda_{max}(\mathcal{L})}$, and $\|A\|_2 = a < \gamma^{-L}$. If

$$\frac{|\mathcal{A}|\beta + NQ_0}{N\eta_0} \leq 1 - \mu_0, \tag{11}$$

then

$$\lim_{t\to\infty}\|\tilde{e}(t)\|_2 \leq \frac{NQ_0 + |\mathcal{A}|\beta}{N(1-\mu_0)}. \tag{12}$$

***Remark** 12:* Lemma 2 shows that the error between each state estimate and the average estimates can be upper bounded by $p^*(L,t)$, which is uniformly upper bounded by a constant scalar p_0^* and has some asymptotic properties w.r.t. consensus step L and time t. Lemma 3 provides a sufficient condition to guarantee the boundedness of network tracking error (i.e., $\hat{x}_{avg}(t) - x(t)$). For given b_w, b_v, λ_0 and a, we can design β and L based on the condition (11) to guarantee (12).

From (7) and $a\gamma^L < 1$, we have the condition (11). By Lemma 2, Lemma 3 and the notations in (6), the conclusion of Theorem 1 holds.

V. Simulation Results

In this section, we carry out a numerical simulation to show the effectiveness of the proposed algorithm.

Regarding the system (1), we assume $A = \left[\begin{smallmatrix}1.01 & 0.1\\ 0.1 & 1.1\end{smallmatrix}\right]$ with $\|A\|_2 = 1.16$. The observation vectors are randomly selected from the set $\left\{C_1 = [1,0], C_2 = [0,1], C_3 = [\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}]\right\}$. The process noise $w(t)$ and observation noises $v_i(t), i = 1,\ldots,N$, all follow the uniform distribution between $[0,1]$. The bounds are assumed to be $b_v = 1, b_w = 1, \eta_i = 1, i = 1,\ldots,N$. We suppose the time $t = [0,100]$ with sampling interval 1. The sparse network given in Fig. 1 has $N = 100$ nodes with $\lambda_2(\mathcal{L}) = 4.1$ and $\lambda_{max}(\mathcal{L}) = 21.3$. We choose $\beta = 3$, the trails of Monto Carlo experiments are 100. Suppose that the attacker will insert the signal $a_i(t) = 2(C_i x(t) + v_i(t))$ if agent i is compromised.

For one realization with consensus step $L = 8$ and the number of compromised agents 25, we obtain the network tracking performance in Fig. 2. It shows that each element of the system state, i.e., $x_1(t)$ and $x_2(t)$, can be well estimated by agents over the network with small bounded estimation errors. The influence of consensus step L to the mean values (averaged by 100) of the maximal error norms among all agents is studied in Fig. 3 with the number of compromised agents 25, which shows that a bigger consensus step can lead to smaller estimation errors. In Fig. 4 with $L = 4$, we investigate the influence of compromised agent number to estimation errors. We see that with the increasing of compromised number, the estimation errors will become larger, and even diverge when the number is 66. The phenomena conform with former analysis, since not enough information can support an effective estimator if too many agents are compromised. Based on the above results, the utility of the proposed SDCF is validated.

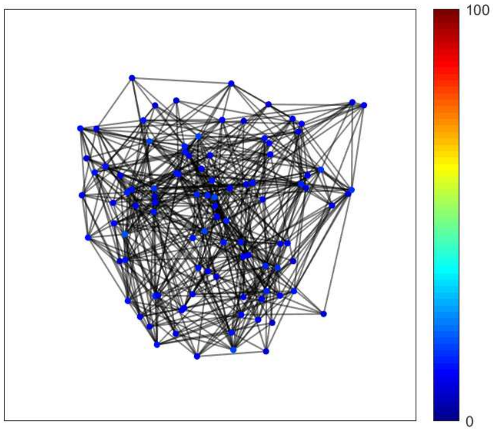

Fig. 1. A random sparse connected graph with 100 nodes.

VI. Conclusions

This paper studied the secure distributed filtering problem for linear time-invariant systems with bounded noises and unstable dynamics under compromised observations, where a malicious attacker can compromise a subset of agents and manipulate the observations arbitrarily. First, we proposed a consensus-based distributed filter. Then, we provided a sufficient condition to guarantee the boundedness of estimation error. The feasibility condition was analyzed through

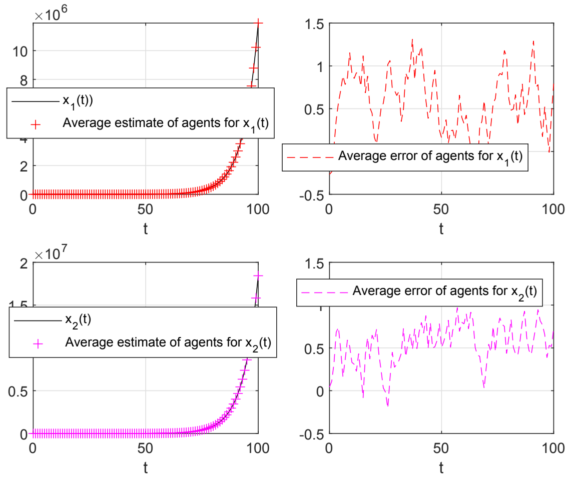

Fig. 2. Network tracking performance for each element over one realization.

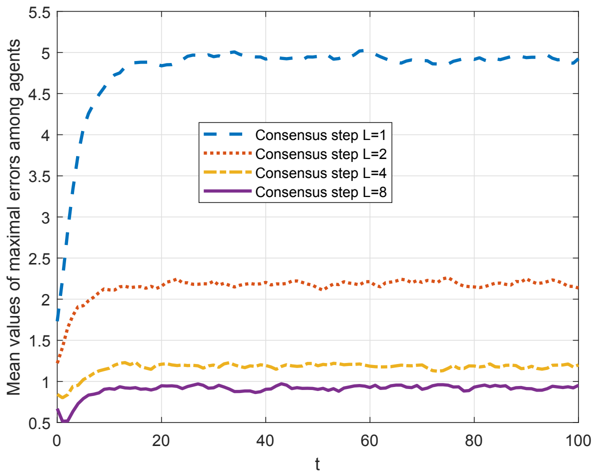

Fig. 3. The influence of consensus step to error norm dynamics.

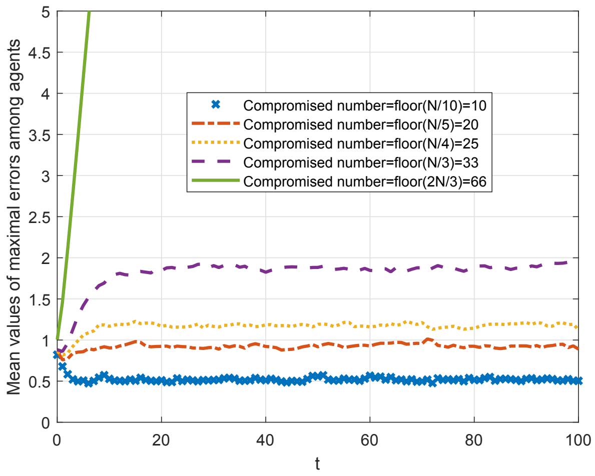

Fig. 4. The influence of compromised agent number to error norm dynamics.

an equivalent statement, which connects to 2s-sparse observability in the centralized framework in certain scenarios.

References

[1] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2715–2729, 2013.

[2] N. Forti, G. Battistelli, L. Chisci, S. Li, B. Wang, and B. Sinopoli, "Distributed joint attack detection and secure state estimation," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 4, no. 1, pp. 96–110, 2018.

[3] M. Pajic, I. Lee, and G. J. Pappas, "Attack-resilient state estimation for noisy dynamical systems," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 82–92, 2017.

[4] 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 Systems Magazine*, vol. 37, no. 2, pp. 66–81, 2017.

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

[6] 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 Transactions on Automatic Control*, vol. 62, no. 10, pp. 4917–4932, 2017.

[7] D. Han, Y. Mo, and L. Xie, "Convex optimization based state estimation against sparse integrity attacks," *IEEE Transactions on Automatic Control*, 2019.

[8] Y. Shoukry, M. Chong, M. Wakaiki, P. Nuzzo, A. Sangiovanni-Vincentelli, S. A. Seshia, J. P. Hespanha, and P. Tabuada, "SMT-based observer design for cyber-physical systems under sensor attacks," *ACM Transactions on Cyber-Physical Systems*, vol. 2, no. 1, p. 5, 2018.

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

[10] X. Ren, Y. Mo, J. Chen, and K. H. Johansson, "Secure state estimation with byzantine sensors: A probabilistic approach," *arXiv:1903.05698*, 2019.

[11] A. Mitra and S. Sundaram, "Secure distributed observers for a class of linear time invariant systems in the presence of byzantine adversaries," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, pp. 2709–2714, IEEE, 2016.

[12] 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.

[13] Y. Chen, S. Kar, and J. M. Moura, "Topology free resilient distributed estimation," *arXiv preprint arXiv:1812.08902*, 2018.

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

[15] M. Mehran and E. Magnus, *Graph theoretic methods in multiagent networks*. Princeton University Press, 2010.

[16] 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.

[17] S. Kar and J. M. Moura, "Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 674–690, 2011.

[18] S. Kar, J. M. Moura, and H. V. Poor, "Distributed linear parameter estimation: Asymptotically efficient adaptive strategies," *SIAM Journal on Control and Optimization*, vol. 51, no. 3, pp. 2200–2229, 2013.

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

[20] C. Zhao, J. He, and J. Chen, "Resilient consensus with mobile detectors against malicious attacks," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 4, no. 1, pp. 60–69, 2017.

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