Attack-Resilient $\mathcal{H}_2$, $\mathcal{H}_\infty$, and $\ell_1$ State Estimator

Yorie Nakahira, Student Member, IEEE, and Yilin Mo, Member, IEEE,

Abstract

Due to its distributed nature, a cyber-physical system is vulnerable to various faults, including sensory integrity attacks. Such faults need to be accounted for in the design of a state estimator. In this paper, we consider sparse sensor faults, in which a small unknown group of sensors can be compromised. We first show a necessary condition that allows the state to be estimated in the presence of sparse sensor faults. We then propose an estimator that is resilient to such faults assuming the necessary condition alone. The proposed estimator requires fewer sensors than existing estimators and avoids potential losses in performance due to delays. Moreover, the proposed estimator’s worst-case estimation error in the two/infinity/infinity norm is given for the two/two/infinity norm bounded input disturbance. Alternatively, this worst-case bound can be considered an extension of robust control to that takes into account a sparse-unbounded input. This extension holds potential interest for a broader audience.

Index Terms

State estimation, fault tolerance, unbounded-sparse-input to bounded-output, secure cyber-physical systems

I. INTRODUCTION

Cyber Physical Systems (CPSs) are vulnerable to corruptions in their sensors and communication channels because they operate via distributed sensing, communication, and control [1]. These corruptions can be caused by malicious agents or unforeseen accidents. Malicious agents may attempt to hijack GPS, computer, and aviation systems for financial or political gain [2], [3]. Unforeseen accidents include...
natural disasters, infrastructure wear, and human errors. Since both causes may result in severe human injuries or economic damage, CPSs should be equipped with safety critical processes that tolerate a wide range of faults.

A common method to deal with faults is the process of fault detection and isolation (FDI) [4]–[8]. The FDI process decides whether a potential fault exists, and if it does, finds the source of the fault. For a linear dynamical system with known system parameters, various methods based on the evaluation of residuals have been proposed (see [4]–[8] and references therein). These methods construct a sequence of residuals that take nonzero values only in the presence of a fault. The existence of such residuals depends on whether the fault can be separated from the disturbance (or modeling uncertainties), which in turn depends on the observability of the dynamical system. However, if we restrict our attention on sparse sensory faults, resilient state estimation can be achieved without going through the process of FDI. Such a restriction is valid when an attacker (or accident) can only change a limited number of local sensors in a large-scale system. As demonstrated later, omitting the process of FDI allows us to relax certain assumptions on observability and reduce the number of sensors, both of which are advantageous results.

Previous studies have proposed optimization problems for secure state estimations against sparse sensory attacks [9], [10]. Their algorithms solve a linear regression between the system output and the state estimate using an $\ell_1$ regularizer, which induces sparse detection of compromised sensors. These algorithms are resilient to attack if the system remains observable after removing arbitrary sensors for twice the number of compromised sensors. However, we show in this paper that a necessary condition to construct a resilient state estimator is that the system remains detectable after removing arbitrary sensors for the same number. This suggests that fundamentally it is possible to reduce the number of sensors while still achieving a resilient estimator. Furthermore, their algorithms have a delay equal to the dimension of the system state, which may cause great performance degradation, especially for systems with high amplification.

In this paper, we propose a state estimator that is resilient to sparse sensory faults while overcoming these shortcomings. The proposed state estimator employs residual-based approach and requires the above necessary condition alone to achieve resilience. This minimal assumption can reduces the number of required sensors, especially for the systems with many stable modes. Although the necessary condition alone cannot guarantee the construction of residuals that take nonzero values only in the presence of attacks, we show that a bounded state estimation error is achievable via appropriately thresholding residuals that contain mixtures of attacks and disturbances. Analytical formulas are given for the worst-case estimation error in two/infinity/infinity norm under two/two/infinity norm bounded disturbances. In
other words, the stability of an estimator can be guaranteed for \( \mathcal{H}_2, \mathcal{H}_\infty, \ell_1 \) systems, and its worst-case error is given as a function of the number of compromised sensors. Our results can also be considered as an extension of bounded-input-bounded-output (MIMO) stability to a mixture of bounded and sparse-unbounded input. This extension holds potential interest for a broader audience.

A. Related Work

Attacker’s model.

If the readings of sensor measurements can be changed by an attacker, then the attacker may alter the values of certain state variables to their financial gain [11], [12]. The study [13] sought to determine how such an attack on the electrical grid can achieve these outcomes while changing the readings of a minimal number of sensors. Complementing the results in [13], we consider achievable state estimation errors given a sparse sensor integrity attack in a linear dynamical system. While the work [13] seeks to determine how an attacker can maximize their damage to the system, our study considers how to design an estimator that mitigate such damage.

Failure detection and identification (FDI). There are a few useful books or reviews on FDI in dynamical systems [4]–[8]. A common approach in FDI is thresholding residuals that (only) reacts to faults. The construction of such residuals includes directional approach (the residual vector points to a particular direction for each fault) and structural approach (each residual are sensitive to a different subset of faults). Potential applications of FDI includes consensus networks, power grids, and wireless control networks [9], [14]–[16].

Fault tolerant estimation. Robust estimation aims to design an estimator that is resilient to sensory attack. Typical assumptions is that sensory faults are independently generated from certain distributions, or are worst-case in maximizing an estimation error. The former assumption are employed in M-estimator, L-estimator, R-estimator [17]–[19]. The latter assumption are employed in minimum mean squared error (MSE) estimator [20]. These previous studies [17]–[20] consider a stochastic setting, which is complementary to the deterministic setting of this paper.

Robust control. \( \mathcal{H}_\infty, \mathcal{H}_2, \ell_1 \) optimal control is the design and analysis tool for the controller/estimator that minimizes the induced norm from input (disturbances) to output [21], [22]. Beyond its scope on bounded input, we extend their analysis for unbounded sparse input, which models the sensor faults. Our analysis can be considered as the generalization of optimal control to take into account for unbounded sparse input.
II. Preliminary

A. Notations

The set of natural numbers is denoted $\mathbb{N}$, the set of non-negative integers is denoted $\mathbb{Z}_+$, the set of real numbers is denoted $\mathbb{R}$, the set of non-negative real numbers is denoted $\mathbb{R}_+$, and the set of complex numbers is denoted $\mathbb{C}$. The cardinality of a set $S$ is denoted $|S|$. A sequence $\{x(t)\}_{t \in \mathbb{Z}_+}$ is abbreviated by the lower case letter $x$, and the truncated sequence from $t_1$ to $t_2$ is denoted $x(t_1 : t_2)$. Let $\|x\|_0 = |\{i : \exists t \text{ s.t. } x_i(t) \neq 0\}|$ denote the number of entries in $x$ that take non-zero values for some time. The infinity-norm of a sequence $x \in \mathbb{R}^n$ is defined as $\|x\|_\infty \triangleq \sup_t \max_i |x_i(t)|$, and the two-norm of a sequence $x$ is defined as $\|x\|_2 \triangleq \left( \sum_{t=0}^{\infty} \sum_{i=1}^{n} |x_i(t)|^2 \right)^{1/2}$. Similarly, the norms of a truncated sequence $x(0 : T)$ are defined as $\|x(t_1 : t_2)\|_\infty \triangleq \max_{t_1 \leq t \leq t_2} \max_i |x_i(t)|$ and $\|x(t_1 : t_2)\|_2 \triangleq \left( \sum_{t=t_1}^{t_2} \sum_{i=1}^{n} |x_i(t)|^2 \right)^{1/2}$. Let $\ell_\infty$ be the space of sequences with bounded infinity-norm, and $\ell_2$ be the space of sequences with bounded two-norm.

B. LTI Systems and System Norms

Let $G$ be the following discrete-time linear time-invariant (LTI) system:

$$
\begin{align*}
x(t + 1) &= Ax(t) + Bw(t) & y(t) &= Cx(t) + Dw(t),
\end{align*}
$$

with the initial condition $x(0) = 0$, system state $x(t) \in \mathbb{R}^n$, system input $w(t) \in \mathbb{R}^l$, and system output $y(t) \in \mathbb{R}^m$. The transfer matrix of the system is

$$
G = \begin{bmatrix}
A & B \\
C & D
\end{bmatrix},
$$

whereas the transfer function of the system is $\hat{G}(e^{i\theta}) = B(e^{i\theta}I - A)C + D$. For $\ell_p$ system input and $\ell_q$ system output, the system norm (namely, induced-norm) is given by

$$
\|G\|_{p \rightarrow q} \triangleq \sup_{\|u\|_p \neq 0} \frac{\|y\|_q}{\|u\|_p}.
$$

In particular, the induced-norms for $(p,q) = (2,2), (2,\infty), (\infty,\infty)$ are respectively $\mathcal{H}_\infty$, $\mathcal{H}_2$, and $\mathcal{L}_1$ norms. These induced-norms are bounded when $A$ is strictly stable (i.e., all the eigenvalues of $A$ is in the open unit circle). See [22] for further details.
In particular, the induced-norms for \((p,q) = (2,2), (2,\infty), (\infty,\infty)\) are respectively \(H_\infty, H_2,\) and \(L_1\) norms, which are defined as

\[
\|G\|_{2 \rightarrow \infty} = \|G\|_2 = \int_{-\pi}^{\pi} \text{tr}(\hat{G}(e^{i\theta})\hat{G}^T(e^{-i\theta}))d\theta / 2\pi \\
\|G\|_{2 \rightarrow 2} = \|G\|_\infty = \text{ess sup}_{e^{i\theta}} \sigma_{\max}(\hat{G}(e^{i\theta})) \\
\|G\|_{\infty \rightarrow \infty} = \|G\|_1 = \max_{1 \leq i \leq n} \sum t=0^\infty \sum j=1^\infty |h_{ij}(t)|
\]

where \(h_{ij}\) is the impulse response from \(w_j(t)\) to \(y_i(t)\). These induced-norms are bounded when \(A\) is strictly stable (i.e., all the eigenvalues of \(A\) is in the open unit circle). See [22] for further details.

\[C.\text{ Detectability and State Estimation}\]

A system is detectable if all its unstable modes can be estimated using the system output. We formally state its definition and an equivalent condition below.

**Definition 1.** The pair \((A,C)\) is detectable if \(A + KC\) is stable for some matrix \(K\).

Detectability implies that, for any eigenvector \(v \in \mathbb{C}^n\) and corresponding eigenvector \(z \in \mathbb{C}\) of \(A\) (i.e., \(Av = zv\)) such that \(|z| \geq 1\), condition \(Cv \neq 0\) holds [21]. It also implies that \(A + KC\) is strictly stable for some matrix \(K\). We can use such \(K\) to construct a linear estimator

\[
\dot{x}(t + 1) = (A + KC)x(t) - K(y(t) - C\hat{x}(t)), \quad \hat{x}(0) = 0.
\]

Define its estimation error \(e\) and residual vector \(r\) as

\[
e(t) \triangleq x(t) - \hat{x}(t), \quad r(t) \triangleq y(t) - C\hat{x}(t),
\]

respectively. The signals \(e\) and \(r\) satisfy the following dynamics

\[
e(t + 1) = (A + KC)e(t) + (B + KD)w(t), \quad e(0) = 0
\]

\[
r(t) = Ce(t) + Dw(t),
\]

Therefore, the LTI system from \(w\) to \(e\), \(E(K)\), the LTI system from \(w\) to \(e\), \(F(K)\), are respectively given by

\[
E(K) = \begin{bmatrix}
A + KC & B + KD \\
I & 0
\end{bmatrix}
\]

\[
F(K) = \begin{bmatrix}
A + KC & B + KD \\
C & D
\end{bmatrix}.
\]
Both $E(K)$ and $F(K)$ have bounded induced-norms (because $A + KC$ is strictly stable) and provide upper-bounds on $\|e\|_q$ and $\|r\|_q$ as below.

Lemma 1. If $\|w\|_p \leq 1$, then the estimation error $e$ and residual vector $r$ satisfy

$$
\|e\|_q \leq \|E(K)\|_{p \rightarrow q}, \quad \|r\|_q \leq \|F(K)\|_{p \rightarrow q}.
$$

III. Problem Formulation

Consider the estimation problem in Fig. 1. The plant is a discrete-time LTI system:

$$
x(t + 1) = Ax(t) + Bw(t), \quad x(0) = 0
$$
$$
y(t) = Cx(t) + Dw(t) + a(t),
$$
where $x(t) \in \mathbb{R}^n$ is the system state, $w(t) \in \mathbb{R}^l$ is the input disturbance, $y(t) \in \mathbb{R}^m$ is the output measurement, and $a(t) \in \mathbb{R}^m$ is the bias injected by the adversary (we denote as $a(t)$ as an attack). The time horizon is non-negative integers, i.e., $t \in \mathbb{Z}_+$. The disturbance matrix $B$ has full-rank, implying that all directions in the space of state, $\mathbb{R}^n$, can be disturbed. Each sensor is indexed by $i \in \{1, \ldots, m\}$ and produces measurement $y_i(t)$, which jointly comprises the measurement vector $y(t) = [y_1(t), \ldots, y_m(t)]^T$.

A sensor $i$ is said to be compromised if $a_i(t) \neq 0$ at some time $t \in \mathbb{Z}_+$, and benign otherwise. The attack can change the measurements of at most $\rho$ sensors, i.e.,

$$
\|a\|_0 \leq \rho.
$$

We call such an attack as $\rho$-sparse. Let $S \triangleq \{1, \ldots, m\}$ denote the set of all sensors, $\mathcal{C} \subset S$ denote the set of compromised sensors, and $\mathcal{B} \triangleq S \setminus \mathcal{C}$ denote the set of benign sensors. The set $\mathcal{C}$, which is assumed to be unknown.\(^1\) A causal state estimator is an infinite sequence of functions $\{f_t\}$ where each $f_t$ maps all output measurements to a state estimate:

$$
\hat{x}(t) = f_t(y(0 : t - 1)).
$$

The estimation error of (12) is defined as the difference between the system state and the state estimate:\(^2\)

$$
e(t) \triangleq x(t) - \hat{x}(t).
$$

\(^1\)Take the setting of [10] for example. When the system is noisy, the optimization problem $\min_{x_t \in \mathbb{R}^n} \|y(t)^T, \ldots, y(t+n-1)^T - O x_t\| \quad (O$ is the observability matrix) may not give correct set of compromised sensors $\{i : \exists t, a_i(t) \neq 0\}$.

\(^2\)Although abbreviate it as $e(t)$, the estimation error is also a function of disturbance $w$, attack $a$, and the estimator $\{f_t\}$. 

Fig. 1. Diagram of the Estimation Problem in Adversarial Environment.

We characterize the performance of the estimator using the worst-case estimation error in $q$-norm subject to input disturbance with bounded $p$-norm and $\rho$-sparse attack

$$\sup_{\|w\|_p \leq 1, \|a\|_0 \leq \rho} \|e\|_q$$

for the following pairs of input/output norms:

$$(p, q) = (2, 2), (2, \infty), (\infty, \infty).$$

**Definition 2.** An causal state estimator $\{f_t\}$ is resilient to attack if it has a bounded estimation error, i.e., $\sup_{\|w\|_p \leq 1, \|a\|_0 \leq \rho} \|e\|_q < \infty$.

Our goal is to study the design problem of a resilient state estimator $\{f_t\}$. In particular, we first show a necessary condition for the construction of a resilient estimator (Section IV-A), then proposes a resilient estimator (Section IV-B), and finally analyze the estimation error and its relation to the tolerable number of compromised sensors.

**IV. NECESSARY AND SUFFICIENT CONDITIONS FOR RESILIENCE TO ATTACK**

In this section, we first provide a necessary condition for the existence of a resilient estimator and then, assuming the necessary condition, propose a resilient estimator. We first define some notation that will be used later.

For a noisy system with unknown set compromised sensors, the system may not be correctable in the sense defined in [9], [10].
Definition 3 (Projection map). Let \( e_i \) be the \( i \)th canonical basis vector of the space \( \mathbb{R}^m \) and \( \mathcal{I} = \{i_1, \ldots, i_{m'}\} \subseteq \mathcal{S} \) be an index set with carnality \( m' (\leq m) \). We define the projection map \( P_{\mathcal{I}} : \mathbb{R}^m \rightarrow \mathbb{R}^{m'} \)
as
\[
P_{\mathcal{I}} = \begin{bmatrix} e_{i_1} & \ldots & e_{i_{m'}} \end{bmatrix}^T \in \mathbb{R}^{m' \times m}. \tag{16}
\]

Using \( P_{\mathcal{I}} \) in (16), the measurements of the set of sensors \( \mathcal{I} \subset \mathcal{S} \) can be written as
\[
y_{\mathcal{I}}(t) \triangleq P_{\mathcal{I}} y(t) \in \mathbb{R}^{m'}. \]

Similarly, the measurement matrix and the sensor noise matrix corresponding to the set of sensors \( \mathcal{I} \) can be respectively written as
\[
C_{\mathcal{I}} \triangleq P_{\mathcal{I}} C, \quad D_{\mathcal{I}} \triangleq P_{\mathcal{I}} D.
\]

A. Necessary Condition for Resilience to Attack

If \((A, C_{\mathcal{K}})\) is not observable for some set of sensors \( \mathcal{K} \subset \mathcal{S} \) with cardinality \( |\mathcal{K}| = m - 2\rho \), then there exists some stealthy attack that cannot be detected [9]. However, if such a stealthy disturbance only introduces a finite error to the estate estimate, then we may still be able to design a resilient estimator. This motivates us to find a condition in which a stealthy disturbance can only introduce a finite error.

Theorem 1. Consider system (10) with \( \rho \)-sparse attack. If \((A, C_{\mathcal{K}})\) is not detectable for some set of sensors \( \mathcal{K} \subset \mathcal{S} \) with cardinality \( |\mathcal{K}| = m - 2\rho \), then there is no state estimator \( \{f_t\} \) that is resilient to attack.

We present the proof of Theorem 1 in the Appendix. Theorem 1 implies that the following (denote as Condition A) is necessary for the existence of a resilient state estimator:

A. \((A, C_{\mathcal{K}})\) is detectable for any set of sensors \( \mathcal{K} \subset \mathcal{S} \) with cardinality \( |\mathcal{K}| = m - 2\rho \).

B. The Proposed Estimator

Assuming condition A, we now propose a resilient state estimator. The proposed estimator constitutes two procedures: 1) local estimation and 2) global fusion. The local estimators are defined by groups of \( m - \rho \) sensors for all combinations
\[
\mathcal{V} \triangleq \{ \mathcal{I} \subset \mathcal{S} : |\mathcal{I}| = m - \rho \}.
\]
The number of such groups (local estimators) is \( |\mathcal{V}| = \binom{m}{\rho} \). Each local estimator \( \mathcal{I} \) generates a state estimation \( \hat{x}_{\mathcal{I}} \) separately based on the measurements of its own sensors \( y_{\mathcal{I}} \). In the global fusion process,
A global state estimation $\hat{x}$ is generated using the estimates from all local estimators $\mathcal{I} \in \mathcal{V}$. With slight overlap of notation, we use $\mathcal{I} \in \mathcal{V}$ to refer to a set of sensors as well as to the estimator that uses these sensors. Next, we outline these procedures and formally state the estimator in Algorithm 1.

1) Local Estimations: From Assumption A, for any set of sensors $\mathcal{I} \in \mathcal{V}$, there exists a matrix $K^{\mathcal{I}} \in \mathbb{R}^{(m-\rho) \times n}$ such that $A + K^{\mathcal{I}}C^{\mathcal{I}}$ is strictly stable (has all eigenvectors in the open unit circle). Using this matrix $K^{\mathcal{I}}$, we construct a local estimator that only uses measurements from the set of sensors $\mathcal{I}$ to produce a local state estimate $\hat{x}^{\mathcal{I}}$:

$$\hat{x}^{\mathcal{I}}(t+1) = A\hat{x}^{\mathcal{I}}(t) - K^{\mathcal{I}}(y^{\mathcal{I}}(t) - C^{\mathcal{I}}\hat{x}^{\mathcal{I}}(t))$$

(17)

with the initial condition $\hat{x}^{\mathcal{I}}(0) = 0$. The estimation error and residual vector of (17) is respectively defined as

$$e^{\mathcal{I}}(t) \triangleq x(t) - \hat{x}^{\mathcal{I}}(t)$$

(18)

$$r^{\mathcal{I}}(t) \triangleq y^{\mathcal{I}}(t) - C^{\mathcal{I}}\hat{x}^{\mathcal{I}}(t)$$

(19)

The LTI system from $w$ to $e^{\mathcal{I}}$ is $E^{\mathcal{I}}(K^{\mathcal{I}})$ defined in (7), whereas the LTI system from $w$ to $r^{\mathcal{I}}$ is $F^{\mathcal{I}}(K^{\mathcal{I}})$ defined in (8).

When the set $\mathcal{I}$ does not contain any compromised sensors, i.e., $a^{\mathcal{I}} = 0$, the residual vector $r^{\mathcal{I}}(t)$ is determined by disturbance $w$ alone and is bounded by

$$\|r^{\mathcal{I}}(0:t)\|_q \leq \|F^{\mathcal{I}}(K^{\mathcal{I}})\|_{p \rightarrow q}.$$ 

(20)

Condition (20) can only be violated when the set $\mathcal{I}$ contains compromised sensors, so (20) is a necessary condition for all the sensors in set $\mathcal{I}$ to be benign. The local estimator at time $t$ uses the necessary condition (20) to determine the validity of its estimate and label local estimator $\mathcal{I}$ to be invalid upon observing $\|r^{\mathcal{I}}(0:t)\|_q > \|F^{\mathcal{I}}(K^{\mathcal{I}})\|_{p \rightarrow q}$.

2) Global Fusion: From above, the set of valid local estimators $\mathcal{I} \in \mathcal{V}(t)$ is characterized as

$$\mathcal{V}(t) \triangleq \{ \mathcal{I} \in \mathcal{S} : \|r^{\mathcal{I}}(0:t)\|_q \leq \|F^{\mathcal{I}}(K^{\mathcal{I}})\|_{p \rightarrow q} \}.$$ 

(21)

4One way to find the matrix $K$ is via the Riccati equation, i.e., $K^{\mathcal{I}} = PC^{\mathcal{I}}(C^{\mathcal{I}}PC^{\mathcal{I}} + D^{\mathcal{I}}D^{\mathcal{I}})^{-1}$, where $P$ is unique stabilizing solution of the discrete-time algebraic Riccati equation $P = A(P - PC^{\mathcal{I}}(C^{\mathcal{I}}PC^{\mathcal{I}} + D^{\mathcal{I}}D^{\mathcal{I}})^{-1}C^{\mathcal{I}})A^{T} + BB^{T}$.

Sufficient conditions the existence of solution $P$ is that $(A, C^{\mathcal{I}})$ is detectable and $(A, BB^{T})$ is detectable.

5We use superscript notations for original vectors and matrices (e.g. $K^{\mathcal{I}}$ and $x^{\mathcal{I}}$, respectively) and subscript for vectors and matrices projected by (16) (e.g. $y^{\mathcal{I}}$ and $C^{\mathcal{I}}$, respectively).
Using $\mathcal{V}(t)$, we compute the global state estimate as follows: $\hat{x}(t) = [\hat{x}_1(t), \hat{x}_2(t), \ldots, \hat{x}_n(t)]$, where

$$
\hat{x}_i(t) = \begin{cases} 
\frac{1}{2} \left( \min_{\mathcal{I} \in \mathcal{V}(t)} \hat{x}^T_i(t) + \max_{\mathcal{J} \in \mathcal{V}(t)} \hat{x}^T_j(t) \right) & q = \infty \\
\frac{1}{|\mathcal{V}(t)|} \sum_{\mathcal{I}(t) \in \mathcal{V}(t)} \hat{x}^T_i(t) & q = 2.
\end{cases}
$$

(22)

**Remark 1.** Condition (20) is sufficient to construct a resilient estimator. However, when $q = \infty$, the following element-wise bound can achieve improved estimation accuracy:

$$
|r^T_i(0 : t)| \leq \sum_{j=1}^{l} \sum_{\tau=0}^{t} \{h_{F^T(K^I)}(\tau)\}_{ij},
$$

where $\{h_{F^T(K^I)}\}_{ij}$ is the impulse response from $w_j(t)$ to $r_i(t)$.

Algorithm 1 The Proposed State Estimator

```plaintext
Initialize $\mathcal{V}(0) \leftarrow \mathcal{V}$ and $\hat{x}^T(0) \leftarrow 0, \mathcal{I} \in \mathcal{V}(0)$

for $t \in \mathbb{N}$ do

    for $\mathcal{I} \in \mathcal{V}(t-1)$ do (Local Estimation)

        Initialize $\mathcal{V}(t) \leftarrow \emptyset$

        Determine $\hat{x}^T(t)$ from (17) and $r^T(t)$ from (19)

        if $\|r^T(0 : t)\|_q \leq \|F^T(K^I)\|_{p \rightarrow q}$ then

            $\mathcal{V}(t) \leftarrow \{\mathcal{V}(t), \mathcal{I}\}$

        end if

    end for

Obtain estimate $\hat{x}(t)$ from (22) (Global Fusion)

end for
```

C. Resilience of the Proposed Estimator

In this section, we show that the proposed estimator is resilient to attack, i.e., its worst-case estimation error is bounded.

**Theorem 2.** The estimator in Algorithm 1 has a bounded estimation error. In particular, the estimation error is upper-bounded by

$$
\max_{\mathcal{I} \in \mathcal{V}} \|E^T(K^I)\|_{\infty} + \max_{\mathcal{J}, \mathcal{J}' \in \mathcal{V}} \sqrt{\frac{1}{2} \log |\mathcal{V}|} \left[ \frac{D^T_{\mathcal{I}, \mathcal{J}}}{2} \right] \text{ if } (p, q) = (2, 2)
$$

$$
\max_{\mathcal{I}, \mathcal{J} \in \mathcal{V}} \left( \|E^T(K^I)\|_2 + \frac{1}{2} D^T_{\mathcal{I}, \mathcal{J}} \right) \text{ if } (p, q) = (2, \infty)
$$

$$
\max_{\mathcal{I}, \mathcal{J} \in \mathcal{V}} \left( \|E^T(K^I)\|_1 + \frac{1}{2} D^T_{\mathcal{I}, \mathcal{J}} \right) \text{ if } (p, q) = (\infty, \infty).
$$
In the above formula, the term $D_{p,q}^{I,J}$ is defined as

$$D_{p,q}^{I,J} = \alpha^{I\cap J}_{p,q} \left( \beta^{I\cap J}_{p,q} + \beta^{J\cap J}_{p,q} \right)$$

where $P_{K,I} \in \mathbb{R}^{|K|\times|I|}$ is the unique solution of $P_{K} = P_{K,I}P_{I}$, and $\| \cdot \|_{p\to q}$ is an induced norm on matrix.

The estimation error upper-bound in Theorem 2 decomposes into two terms: $\| E_I(K^I) \|_{p\to q}$ and the remaining. The first term $\| E_I(K^I) \|_{p\to q}$ characterize the error between a local estimator and the true state. That is, if the local estimator $I$ is used for a system with no attack ($a \equiv 0$), then its estimation error is bounded by $\| E_I(K^I) \|_{p\to q}$. The second term exists due to the attack in an unknown set of sensors. Moreover, it shall be noted that an increase in the tolerable number of compromised sensors $\rho$ may result in an increase in both terms, thus increasing the worst-case estimation error $\sup \| w \|_{p\leq 1, \| a \|_{0}=0} \| e \|_{q}$. An immediate consequence of Theorem 2 is that condition A is a necessary and sufficient condition for the construction of a resilient state estimator, and that Algorithm 1 resilient to attack.

**Corollary 1.** A necessary and sufficient condition for the existence of a resilient estimator is that $(A, C_K)$ is detectable for any index set $K \subset S$ with cardinality $m - 2\rho$.

**Corollary 2.** Consider system (10) with $\rho$-sparse attack. The state estimator in Algorithm 1 is resilient to attack.

**Remark 2.** When $(p,q) = (2,2)$, the error upper-bound grows at the order $o(\sqrt{\rho \log m})$ for $m \to \infty$. Hence, the error can be kept small even for systems with large $m$.

**V. PROOF OF RESILIENCE FOR PROPOSED ESTIMATOR**

We highlight important parts of the proof of Theorem 2 in this section and present the complete proof in the extended version of this paper [23]. The proof of Theorem 2 has two procedures: 1) bounding local estimation errors, and 2) bounding global fusion errors. Specifically, from the triangular inequality,
at any time \( t \in \mathbb{Z}_+ \), the estimation error satisfies

\[
\|e\|_q = \|(x - \hat{x}^T) + (\hat{x}^T - \hat{x})\|_q \\
\leq \|x - \hat{x}^T\|_q + \|\hat{x}^T - \hat{x}\|_q,
\]

where \( \mathcal{I} \in \mathcal{B} \subset \mathcal{V} \) is a set that only contains benign sensors (denote \( \mathcal{I} \) as the *benign estimator*). The benign estimator \( \mathcal{I} \) exists from assumption (11). The first term \( \|x - \hat{x}^T\|_q \) can be bounded using Lemma 1 by

\[
\|x - \hat{x}^T\|_q \leq \|E^I(K^T)\|_{p \rightarrow q}.
\]

Now it only remains to show that the second term is bounded.

To bound the second term, we first bound the difference between the estimates of any two valid local estimators \( \mathcal{J}_1, \mathcal{J}_2 \in \mathcal{V}(T) \) up to time \( T \in \mathbb{Z}_+ \), which is given in Lemma 2. We then use Lemma 2 to show that the difference between the estimates of the benign estimator \( \mathcal{I} \in \mathcal{V} \) and the global estimator is finite. This is shown in Lemma 3 for \((p, q) = (2, 2)\) and in Lemma 4 for \((p, q) = (2, \infty), (\infty, \infty)\). In these lemmas, each set of sensors in \( \mathcal{V} \) are labeled into

\[
\mathcal{J}_1, \mathcal{J}_2, \ldots, \mathcal{J}_{|\mathcal{V}|}.
\]

**Lemma 2.** Assume that Condition A holds. Let \( \mathcal{J}_1, \mathcal{J}_2 \in \mathcal{V}(T) \) be two sets of sensors that is valid at time \( T \). The divergence between the local estimator \( \mathcal{J}_1 \) and \( \mathcal{J}_2 \) up to time \( T \) satisfies

\[
\|\hat{x}^{\mathcal{J}_1}(0 : T) - \hat{x}^{\mathcal{J}_2}(0 : T)\|_q \leq \mathcal{D}_{p,q}^{\mathcal{J}_1,\mathcal{J}_2},
\]

where right hand side is finite, i.e., \( \mathcal{D}_{p,q}^{\mathcal{J}_1,\mathcal{J}_2} < \infty \).

**Lemma 3.** If condition (26) holds for \((p, q) = (2, 2)\) at all time \( T \in \mathbb{Z}_+ \), then the divergence between the benign estimator \( \mathcal{I} \) and the global estimator satisfies

\[
\|\hat{x}^T - \hat{x}\|_2 \leq \max_{\mathcal{J}_1, \mathcal{J}_2 \in \mathcal{V}} \sqrt{\log |\mathcal{V}|/2} \mathcal{D}_{p,q}^{\mathcal{J}_1,\mathcal{J}_2}.
\]

**Lemma 4.** If condition (26) holds for \((p, q) = (2, \infty), (\infty, \infty)\) at all time \( T \in \mathbb{Z}_+ \), then the divergence between the benign estimator \( \mathcal{I} \) and the global estimator satisfies

\[
\|\hat{x}^T - \hat{x}\|_\infty \leq \frac{1}{2} \max_{\mathcal{J} \in \mathcal{V}} \mathcal{D}_{\infty,\infty}^{\mathcal{I},\mathcal{J}}.
\]

We prove Lemma 2, Lemma 3, and Lemma 4 in Section V-A, Section V-B, and Section V-C, respectively. Combining all of the above, we are now ready to prove Theorem 2.
Proof (Theorem 2). Taking supremum over all \( t \in \mathbb{Z}_+ \) and maximizing over all sensor sets \( \mathcal{I} \) in Lemma 3, we obtain for \((p, q) = (2, 2)\) that \( \|x\|_q \leq \max_{I \in \mathcal{V}} \|E(I)\|_p \max_{J_1, J_2 \in \mathcal{V}} \left( \frac{D_{2,2}}{2} \right) \log |\mathcal{V}|/2 \). Applying similar argument for the case for \( q = \infty \), we obtain that \( \|e\|_\infty \leq \max_{I, J \in \mathcal{V}} \left( \|E(I)\|_p + D_{p,q} \right) \).

\( \square \)

A. Proof of Lemma 2

We first present a lemma, using which we prove Lemma 2.

**Lemma 5.** Consider system (1) where \((A, C)\) is detectable and \( \|u\|_p \leq 1 \). If \( y(t) = 0 \) for all \( t = 0, 1, \ldots, T \), then

\[
\|x(0 : T)\|_q \leq \inf_{K: A+KC \text{ strictly stable}} \|E(K)\|_{p \to q},
\]

where \( E(K) \) is given in (7).

**Proof (Lemma 5).** As \((A, C)\) is detectable, \( A + KC \) is strictly stable for some matrix \( K \). For such stabilizing \( K \), we can construct the state estimator (6). Since \( y(0 : T) = 0 \), the state estimator (6) produces zero estimate \( \hat{x}(0 : T) = 0 \). From Lemma 1, we obtain

\[
\|x(0 : T)\|_q = \|e(0 : T)\|_{q^*} \leq \|E(K)\|_{p \to q}.
\]

Taking infimum over all \( K \) such that \( A + KC \) is strictly stable, we obtain (29).

**Proof (Lemma 2).** Let \( J_1, J_2 \in \mathcal{V}(T) \). We first compute the dynamics of the local estimates \( \hat{x}^{J_i}(t) \), \( i = 1, 2 \). From (17) and (19),

\[
\hat{x}^{J_i}(t + 1) = A\hat{x}^{J_i}(t) - K^{J_i}r^{J_i}(t), \hat{x}^{J_i}(t) = 0
\]

\[
y^{J_i}(t) = C_{J_i}\hat{x}^{J_i}(t) + r^{J_i}(t)
\]

for \( t \leq T \). We define the sequences \( \phi^{J_i}(t) \) and \( \varphi^{J_i}(t) \) by

\[
\phi^{J_i}(t) \triangleq -K^{J_i}r^{J_i}(t), \quad \varphi^{J_i}(t) \triangleq P_{K_{1,2},J_i}r^{J_i}(t),
\]

where \( P_{K_{1,2},J_i} \in \mathbb{R}^{|K_{1,2}| \times |J_i|} \) is the unique solution of \( P_{K_{1,2}} = P_{K_{1,2},J_i}P_{J_i} \). Let \( K_{1,2} = J_1 \cap J_2 \) the intersection between the two sets \( J_1, J_2 \). As the measurements from subset \( K_{1,2} \subset J_i \) also satisfies

\[
y^{K_{1,2}}(t) = C^{K_{1,2}}\hat{x}^{J_i}(t) + P_{K_{1,2},J_i}r^{J_i}(t),
\]

combining with (30) yields

\[
\hat{x}^{J_i}(t + 1) = A\hat{x}^{J_i}(t) + \begin{bmatrix} I & 0 \end{bmatrix} \begin{bmatrix} \phi^{J_i}(t) \\ \varphi^{J_i}(t) \end{bmatrix}, \hat{x}^{J_i}(t) = 0
\]

\[
y^{K_{1,2}}(t) = C^{K_{1,2}}\hat{x}^{J_i}(t) + \begin{bmatrix} 0 & I \end{bmatrix} \begin{bmatrix} \phi^{J_i}(t) \\ \varphi^{J_i}(t) \end{bmatrix}.
\]
Now, let $\Delta(t)$ be the difference between the local estimator $J_1$ and local estimator $J_2$, i.e.,

$$
\Delta(t) \triangleq \hat{x}^{J_1}(t) - \hat{x}^{J_2}(t).
$$

(32)

Subtracting the equation (31) for $J_1$ from equation (31) for $J_2$, we obtain the dynamics of $\Delta$ as follows:

$$
\Delta(t + 1) = A\Delta(t) + \begin{bmatrix} I \\ 0 \end{bmatrix} \begin{bmatrix} \phi^{J_1}(t) - \phi^{J_2}(t) \\ \varphi^{J_1}(t) - \varphi^{J_2}(t) \end{bmatrix}, \quad \Delta \hat{x}(t) = 0,
$$

$$
0 = C_{K_{1,2}}\Delta(t) + \begin{bmatrix} 0 \\ I \end{bmatrix} \begin{bmatrix} \phi^{J_1}(t) - \phi^{J_2}(t) \\ \varphi^{J_1}(t) - \varphi^{J_2}(t) \end{bmatrix}.
$$

Because a valid set satisfies (21), the residual vectors of estimator $J_i$, $i = 1, 2$, are bounded by $\|r^{J_i}(0 : T)\|_q \leq \|F^{J_i}(K^{J_i})\|_{p \rightarrow q}$, which results in

$$
\left\| \begin{bmatrix} \phi^{J_1}(0 : T) \\ \varphi^{J_1}(0 : T) \end{bmatrix} \right\|_p \leq \left\| \begin{bmatrix} -K^{J_1} \\ P_{K_{1,2},J_1} \end{bmatrix} \right\|_{p \rightarrow p} \|r^{J_i}(0 : T)\|_p = \beta_{p,q}^{J_1 \cap J_2}.
$$

(33)

From the triangle inequality, we obtain

$$
\left\| \begin{bmatrix} \phi^{J_1}(0 : T) - \phi^{J_2}(0 : T) \\ \varphi^{J_1}(0 : T) - \varphi^{J_2}(0 : T) \end{bmatrix} \right\|_p \leq \beta_{p,q}^{J_1 \cap J_2} + \beta_{p,q}^{J_2 \cap J_1}.
$$

Substitute $\phi^{J_1} - \phi^{J_2}$ for $u$ in Lemma 5 and $\Delta(t)$ for $x$, we obtain

$$
\|\hat{x}^{J_1}(0 : T) - \hat{x}^{J_2}(0 : T)\|_q \leq \alpha_{p,q}^{J_1 \cap J_2} (\beta_{p,q}^{J_1 \cap J_2} + \beta_{p,q}^{J_2 \cap J_1}).
$$

B. Proof of Lemma 3

Define the following two optimization problems:

$$
P_{\delta}(n) := \max_{z_k(i) \geq 0} \sum_{i=1}^{n} \left( \frac{1}{n - i - 1} \right)^2 \left( \sum_{k=i}^{n} z_k(i) \right)^2
$$

s.t. $\sum_{i=0}^{k} (z_k(i))^2 \leq \delta, \quad k = 1, 2, \ldots, n$

$$
D_{\delta}(n) := \min_{\lambda_i > 0} \sum_{i=1}^{n} \lambda_i
$$

s.t. $\sum_{i=1}^{j} \frac{1}{\lambda_i} \leq (j + 1)^2, \quad j = 1, 2, \ldots, n.$
With slight abuse of notation, we will also denote $\mathcal{P}_\delta(n)$, $\mathcal{D}_\delta(n)$ as the optimal solutions of the optimization problem $\mathcal{P}_\delta(n)$, $\mathcal{D}_\delta(n)$, respectively. We first shown that $\mathcal{P}_\delta(N - 1)$ with

$$\delta = \max_{J_1, J_2 \in \mathcal{V}} \left( \mathcal{D}_{J_1, J_2}^{J_1, J_2} \right)^2$$

is an upper-bound of $\| \hat{x}^T - \hat{x} \|_2$ (Lemma 6). The problem $\mathcal{P}_\delta(n)$ is then converted into its dual problem $\mathcal{D}_\delta(n)$, between which the duality gap is zero (Lemma 7). The dual problem $\mathcal{D}_\delta(n)$ admits an analytical solution that can be upper-bounded by a simple formula (Lemma 10).

**Lemma 6.** If condition (26) holds for $(p, q) = (2, 2)$, then the divergence between the benign estimator $I$ and the global estimator satisfies

$$\| \hat{x}^T - \hat{x} \|_2^2 \leq \mathcal{P}_\delta(N - 1),$$

where $N = |\mathcal{V}|$ and $\delta = \max_{J_1, J_2 \in \mathcal{V}} \left( \mathcal{D}_{J_1, J_2}^{J_1, J_2} \right)^2$.

**Proof (Lemma 6).** In order to relate the infinite sequence $\hat{x}^T(t) - \hat{x}(t)$ with the finite-dimensional optimization problems $\mathcal{P}_\delta(n)$, we first divide the infinite time horizon into a finite sequence as below. Let $T_i$ be the time the set $\mathcal{I}_i$ becomes invalid and $T_0 = 0$. Without loss of generality, we assume that

$$T_1 \leq T_2 \leq \cdots \leq T_{N-1} \leq T_N = \infty.$$

The relation $T_N = \infty$ holds because $\mathcal{I}_N$ is a valid from assumption (11). We call $\mathcal{I}_N$ the benign estimator. If $T_i = \infty$, then we define $\{x^{\mathcal{I}_i}(t)\}_{t \in \mathbb{Z}^+}$ as an infinite sequence of points in $\mathbb{R}^n$. Otherwise, if $T_i$ is finite, then we define $\{x^{\mathcal{I}_i}(t)\}_{t = 0, \ldots, T_i}$ as a finite sequence of length $T_i + 1$. Recall that $\mathcal{I}_N$ is the benign estimator. Let $\Delta^{\mathcal{I}_k}(t) = \hat{x}^{\mathcal{I}_N}(t) - \hat{x}^{\mathcal{I}_k}(t)$ denote the difference between the benign estimator $\mathcal{I}_N$ and other local estimator $\mathcal{I}_k$. We first define the following variable:

$$z_k(i) = \| \Delta^{\mathcal{I}_k}(T_i + 1 : T_{i+1}) \|_2,$$

where $k = 1, \cdots, N - 1$ and $i = 0, \cdots, k$.

Using $z_k(i)$, we can bound the estimation error between the local estimator $\mathcal{I}_N$ and the global estimator as follows:
\[ \|\hat{x}_{I_N} - \hat{x}\|_2^2 = \sum_{t=0}^{\infty} \|\hat{x}_{I_N}(t) - \hat{x}(t)\|_2^2 = \sum_{i=0}^{N-2} T_{i+1} \left\| \hat{x}_{I_N}(t) - \hat{x}(t) \right\|_2^2 \]
\[ \leq \sum_{i=0}^{N-2} \left( \frac{1}{N - i} \right)^2 \left( \sum_{k=i+1}^{N-1} \|\hat{x}_{I_N}(t) - \hat{x}(t)\|_2 \right)^2 \]
\[ \leq \sum_{i=0}^{N-2} \left( \frac{1}{N - i} \right)^2 \left( \sum_{k=i+1}^{N-1} \|\hat{x}_{I_N}(t) - \hat{x}(t)\|_2 \right)^2 \]
\[ \leq \sum_{i=0}^{N-2} \left( \frac{1}{N - i} \right)^2 \left( \sum_{k=i+1}^{N-1} z_k(i) \right)^2 \]

The sum of \( i \) and \( j \) counts only up to \( N-1 \) since \( \hat{x}(t) = \hat{x}_{I_N}(t) \) for \( t > T_{N-1} + 1 \). We also used Cauchy-Schwarz inequality in the second to last line. Using the above relation, \( \|\hat{x}_{I_N} - \hat{x}\|_2^2 \) is upper-bounded by the optimal value of the following problem:

\[
\max_{z_k(i) \geq 0} \sum_{i=0}^{N-2} \left( \frac{1}{N - i} \right)^2 \left( \sum_{k=i+1}^{N-1} z_k(i) \right)^2 \quad \text{s.t.} \quad \sum_{i=0}^{k} (z_k(i))^2 \leq \max_{\mathcal{J}_1, \mathcal{J}_2 \in \mathcal{V}} \left( \mathcal{D}_{\mathcal{J}_1, \mathcal{J}_2}^{\mathcal{J}_1, \mathcal{J}_2} \right)^2, \quad k = 1, \ldots, N-1,
\]

which is the optimization problem \( \mathcal{P}_\delta(N-1) \) with \( \delta = \max_{\mathcal{J}_1, \mathcal{J}_2 \in \mathcal{V}} \left( \mathcal{D}_{\mathcal{J}_1, \mathcal{J}_2}^{\mathcal{J}_1, \mathcal{J}_2} \right)^2 \).

\[ \square \]

**Lemma 7.** The problems \( \mathcal{P}_\delta(n), \mathcal{D}_\delta(n) \) have identical optimal values, i.e., \( \mathcal{P}_\delta(n) = \mathcal{D}_\delta(n) \).

We use the following lemma to prove Lemma 7.

**Lemma 8.** The following two inequalities are equivalent

\[ \Lambda = \text{diag}(\lambda_1, \lambda_1, \ldots, \lambda_n) \geq 11^T \]
\[ \sum_{i=1}^{n} \frac{1}{\lambda_i} \leq 1, \quad \text{and} \quad \lambda_j > 0, \quad j = 1, \ldots, n \]

We will use the following lemma to prove Lemma 8.
Lemma 9 (Lemma 1.1 of [24]). If $\Lambda \in \mathbb{R}^{n \times n}$ is an invertible matrix, and $u, v \in \mathbb{R}^n$ are two column vectors, then

$$\det(\Lambda + uv^T) = (1 + v^T \Lambda^{-1} u) \det(\Lambda).$$

Proof (Lemma 8). We first show that (37) implies (38). From (37), $\lambda_i \geq 1$, so $\Lambda$ is invertible. From Lemma 9, we have

$$\left(1 - \sum_{i=1}^{n} \frac{1}{\lambda_i}\right) \prod_{i=1}^{n} \lambda_i = (1 - 1^T \Lambda^{-1} 1) \det(\Lambda) = \det(\Lambda - 11^T)$$

Since $\Lambda \geq 11^T$, the determinant of $\Lambda - 11^T$ is non-negative, which proves (38). We then show that (38) implies (37). Denote the $k$-th leading principal minor of $\Lambda - 11^T$ as $\Delta_k$. By Lemma 9, we obtain that

$$\Delta_k = \left(1 - \sum_{i=1}^{k} \frac{1}{\lambda_i}\right) \prod_{i=1}^{k} \lambda_i \geq 0,$$

which proves (37). \qed

Proof (Lemma 7). Let $v \in \mathbb{R}^{n(n-1)/2}$ be a vector that is composed of $z_{i+1:n}(j) = \{z_{i+1}(j), z_{i+2}(j), \ldots, z_n(j)\}$ for all $j = 0, 1, \ldots, p$, i.e.,

$$v \triangleq [z_{1:n}(0), z_{2:n}(1), \ldots, z_n(n)].$$

Let the following matrices be defined as

$$X = vv^T \geq 0$$

$$F_0 = \text{diag}\left(1/2^2, \ldots, 1_{n-1} 1_{n-1}^T, 1_n 1_n^T/(n+1)^2\right)$$

$$F_i = \text{diag}\vspace{0.1cm} (e_{n,i}, e_{n-1,i}, \ldots, e_{n-i+1,i}, 0_{n-i}, \ldots, 0_1)$$

where $1_k$ is a $k$-dimensional vector with all elements begin $1$; $e_{k,j}$ is a $k$-dimensional row vector with $j$-th entry being $1$ and other entries being $0$; and $0_k$ is a $k$-dimensional row vector with all elements begin $0$. Using SDP relaxation [25], the problem $P_\delta(n)$ can be converted into

$$P'_\delta(n) = \max_{X \geq 0} \text{tr}(F_0 X)$$

$$\text{s.t.} \quad \text{tr}(F_i X) \leq \delta, \forall i = 1, \ldots, n$$

Hence $P_\delta(n) \leq P'_\delta(n)$. This relaxation can be observed from the following relations:

$$\sum_{i=0}^{n-1} \left(\frac{1}{n-i+1}\right)^2 \left(\sum_{k=i+1}^{n} z_k(i)\right)^2 = v^T F_0 v = \text{tr}(F_0 vv^T)$$

$$\sum_{i=0}^{k} (z_k(i))^2 = v^T F_i v = \text{tr}(F_i vv^T).$$
We next show that the relaxation of the problem $P_\delta(n)$ to the semidefinite problem $P'_\delta(n)$ is also exact. Assume that $P_\delta(n)$ is feasible and bounded. Let $X^* = \{x^*_{ij}\}$ be the optimal solution of $P'_\delta$. Define $X$ as 

$$ X = \left[ \sqrt{x^*_{11}} \ldots \sqrt{x^*_{nm}} \right]^T \left[ \sqrt{x^*_{11}} \ldots \sqrt{x^*_{nm}} \right]. $$

From $x_{ii} = x^*_{ii}$ and (39), $X$ satisfies the constraints 

$$ \text{tr}(F_i X) = \text{tr}(F_i X^*) = \delta. \tag{40} $$

Furthermore, because $X^*$ is the optimal solution and $x_{ij} = \sqrt{x^*_{ii}x^*_{jj}} \geq x^*_{ij}$ (due to $X^* \geq 0$),

$$ \text{tr}(F_0 X^*) \geq \text{tr}(F_0 X) \geq \text{tr}(F_0 X^*). \tag{41} $$

Therefore, (40) and (41) shows that $P_\delta(n) = P'_\delta(n)$.

Next, we consider the following dual problem of $P'_\delta(n)$:

$$ D'_\delta(n) = \min_\lambda \delta \sum_{i=1}^n \lambda_i \tag{42} $$

$$ \text{s.t. } \sum_{i=1}^n \lambda_i F_i \geq F_0. $$

Because there exists an strictly positive definite matrix $X > 0$ such that $\text{tr}(F_i X) = \delta$ for all $i = 1, 2, \cdots, n$, from Slater’s condition [26], strong duality holds between $P'_\delta(n)$ and $D'_\delta(n)$. Let $\Lambda_1 = \text{diag}(\lambda_1, \ldots, \lambda_i) \in \mathbb{R}^{i \times i}$, and observe that $\sum_{i=1}^n \lambda_i F_i = \text{diag}(\Lambda_1:1, \Lambda_1:2, \ldots, \Lambda_1:n)$. Hence, the constraint $\sum_{i=1}^n \lambda_i F_i \geq F_0$ is equivalent to 

$$ \text{diag}(\lambda_1, \ldots, \lambda_j) \geq \frac{1}{(j+1)^2} \mathbf{1}\mathbf{1}^T, \forall j = 1, \ldots, n, $$

$$ \iff \sum_{i=1}^j \frac{1}{\lambda_i} \leq (j+1)^2, \quad \lambda_j > 0, \quad j = 1, \ldots, n, $$

where the second line is due to Lemma 8. Therefore, the dual problem $D'_\delta(n)$ can be reformulated into $D_\delta(n)$.

**Lemma 10.** The solution of the problem $D_\delta(n)$ satisfies

$$ D_\delta(n) = \delta \left( \frac{1}{4} + \sum_{i=2}^n \frac{1}{2i+1} \right) \leq \frac{1}{2} \delta \log(n+1). \tag{43} $$

**Proof (Lemma 10).** We first consider the case when $n \geq 2$. Let us define an auxiliary variable 

$$ s_k = \sum_{i=1}^k \frac{\delta}{\lambda_i}, \quad k = 1, \cdots, n, \tag{44} $$
where \( s_0 = 0 \), and \( s_k = 0 \) for \( k \geq n + 1 \). Using the auxiliary variable \( s_k \), we rewrite the optimization problem \( D_\delta(n) \) as

\[
\begin{align*}
\min & \quad \sum_{i=1}^{n} \frac{1}{s_i - s_{i-1}} \\
\text{s.t.} \quad & s_i \leq (i + 1)^2, \quad s_{i-1} \leq s_i, \quad i = 1, \ldots, n
\end{align*}
\] (45)

We define the Lagrangian \( L : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \) of the problem (45) as follows:

\[
L(s, \mu, \eta) = \sum_{i=1}^{n} \frac{1}{s_i - s_{i-1}} + \mu_i(s_i - (i + 1)^2) + \eta_i(s_{i-1} - s_i).
\]

Let \((s^*, \mu^*, \eta^*)\) be any optimal primal and dual variables, then \((s^*, \mu^*, \eta^*)\) satisfies the Karush-Kuhn-Tucker (KKT) conditions. Solving the KKT conditions, we obtain that

\[
\begin{align*}
\mu_i^* &= \begin{cases} 
\frac{1}{16} - \frac{1}{25} & \text{if } i = 1 \\
\frac{1}{(2i+1)^2} - \frac{1}{(2i+3)^2} & \text{if } i = 2, \ldots, n-1 \\
\frac{1}{(2n+1)^2} & \text{if } i = n
\end{cases} \\
\eta_i^* &= 0, \quad i = 1, \ldots, n
\end{align*}
\]

To see this, observe that

1) \( s_i^* - (i + 1)^2 \leq 0, \mu_i^*(s_i^* - (i + 1)^2) = 0, i = 1, \ldots, n. \)
2) \( s_{i-1}^* - s_i^* \leq 0, \eta_i^*(s_{i-1}^* - s_i^*) = 0, i = 1, \ldots, n. \)
3) \( \mu_i^* \geq 0, i = 1, \ldots, n \)
4) \( \eta_i^* \geq 0, i = 1, \ldots, n. \)
5) \( dL(s^*, \mu^*, \eta^*)/ds_i = 0 \)

Since the problem (45), which is equivalent with \( D_\delta(n) \), is a convex problem, the KKT conditions are sufficient for optimality [26]. Therefore, the optimal primal and dual variables are \((s^*, \mu^*, \eta^*)\), which result in the optimal value

\[
D_\delta(n) = \frac{1}{4} + \sum_{i=2}^{n} \frac{1}{2i + 1}.
\] (46)

Next, we upper-bound the optimal cost \( D_\delta(n) \). Since the function \( f(i) = (2i + 1)^{-1}, i \geq 1 \), is convex, from Jensen’s inequality, \( f(i) \) is upper-bounded by \( f(i) \leq \frac{1}{2} (f(i - t/2) + f(i + t/2)) \), for any \( t \in [0,1] \). Integrating this along \( t \in [0,1] \), we obtain that

\[
\begin{align*}
f(i) \leq & \frac{1}{2} \int_{t=0}^{1} f \left( i - \frac{t}{2} \right) + f \left( i + \frac{t}{2} \right) \, dt \\
= & \frac{1}{2} (\log(i + 1) - \log(i)).
\end{align*}
\]
Combining above bound with (46), we establish a lower-bound of the optimal value
\[ D_\delta(n) = \delta \left\{ \frac{1}{4} + \sum_{i=1}^{n} \frac{1}{2i+1} - \frac{1}{3} \right\} \leq \frac{1}{2} \delta \log(n+1) \]

On the other hand, when \( n = 1 \), the optimal variable is \( \lambda_1^* = \delta/4 \), which attains the optimal value \( D_\delta(1) = \delta/4 \leq \log(2)/2 \approx 0.346574 \).

Now we are ready to prove Lemma 3.

**Proof (Lemma 3).** Applying Lemma 2 and Lemma 6, Lemma 7, and then Lemma 10 consecutively, we obtain
\[ \|\hat{x}_I - \hat{x}\|_2^2 \leq P_\delta = D_\delta \leq \frac{1}{2} \log(N) \max_{J_1, J_2 \in \mathcal{V}} \left( D_{J_1, J_2} \right)^2. \]

\[ \square \]

C. Proof of Lemma 4

Lemma 4 is a trivial extension of the following Proposition. We omit its prove due to space constraints.

**Proposition 1.** Let \( z_1, \ldots, z_l \) be real numbers. Define
\[ z = \frac{1}{2} \left( \max_i z_i + \min_i z_i \right). \]

Then for any \( i \), we have
\[ |z - z_i| \leq \frac{1}{2} \max_j |z_j - z_i|. \quad (47) \]

VI. Numerical example

In this section, we demonstrate the proposed estimator through a numerical example. Consider system (10) with
\[ A = 1, \quad C = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 0 \\ 0 \end{bmatrix}, \quad D = \begin{bmatrix} 0 & 1 & 1 & 1 \end{bmatrix}, \]

and 1-sparse attack (\( \rho = 1 \)). The proposed estimator has three local estimators \( \mathcal{I} = \{1, 2\}, \{2, 3\}, \{3, 1\} \).

Let each local estimator to be defined in (17) with the parameter
\[ K^{\{1,2\}} = K^{\{2,3\}} = K^{\{3,1\}} = \begin{bmatrix} \mu & \mu \\ \mu & \mu \end{bmatrix}, \]
where $K^T$ is chosen to be symmetric (since system (10) is symmetric), and $\mu \in (-1, 0)$ for the stability of the estimator. From Theorem 2, the estimation error for $(p, q) = (\infty, \infty)$ satisfies

$$\|e\|_{\infty} \leq 2 + 3(1 - |1 + 2\mu|)^{-1}(1 + |2\mu|).$$

Minimizing the left hand side over $\mu$ to obtain $\mu = -0.5$, which results in $\|e\|_{\infty} \leq 2 + 3(1 - |1 + 2\mu|)^{-1}(1 + |2\mu|)$. The estimation error for the proposed estimator is compared with a regular estimator in Figure 2. The proposed estimator achieves bounded and small estimation error, whereas the regular estimator cannot achieve unbounded estimation error.

VII. CONCLUSION

In this paper, we propose a state estimator for a noisy linear dynamical system that is resilient to sparse sensor integrity attack. The proposed estimator does not have any delay in estimation, may reduce the required number of sensors, and works even if the attacks cannot be isolated. Its worst-case estimation error scales approximately $O(\log (\frac{m}{\rho}))$ for $\mathcal{H}_2$ system and $O(1)$ for $\mathcal{H}_\infty$ and $\ell_1$ system. While, its computational complexity scales approximately $O(\log (\frac{m}{\rho}))$.

![Fig. 2](image)

Fig. 2. The evolution of the estimation error evolution. The disturbance $w(k)$ is generated from a uniform distribution on support $[-1, 1]$ and attack $a(t) = [t * v(t), 0, 0]'$, where $v(t)$ is independent and standard normally distributed. The dynamics of the regular estimator is $\hat{x}(t + 1) = \hat{x}(t) + \frac{1}{3}[1 1 1]([y(t) - C\hat{x}(t)])$, which minimizes $\max_{\|w\|_{\infty} \leq 1} ||e||_{\infty}$ among estimators of the form (6) without the consideration of the attack.

APPENDIX A

PROOF OF THEOREM 1

An intermediate step in the proof of Theorem 1 is to consider the following condition (denote as Condition B).
B. There exist two states \((x, x')\), two disturbances \((w, w')\), and two attacks \((a, a')\) such that all of the followings are satisfied:

a) both \((x, w, a)\) and \((x', w', a')\) satisfies the dynamics (10) and assumptions \(\|a\|_0 \leq \gamma\), and \(\|w\|_p \leq 1\)

b) \(y(t) = y(t')\) at all time \(t \in \mathbb{N}\)

c) the difference between the two states is unbounded, i.e., \(\|x - x'\|_q = \infty\).

Lemma 11. If \((A, C_K)\) is not detectable for some set \(K \subset S\) with \(|K| = m - 2\gamma\), then Condition B holds.

Proof (Lemma 11). We first prove that an undetectable \((A, C_K)\) implies that the linear transformation \(O_t : \mathbb{R}^n \to \mathbb{R}^{t(m-2\gamma)}\) defined by

\[
O_t = \begin{bmatrix}
C_K \\
C_KA \\
C_KA^2 \\
\vdots \\
C_KA^{t-1}
\end{bmatrix}
\]

has a non-trivial kernel (Step 1). Form the kernel space of \(O_t\), we then find two states \((x, x')\) that satisfy condition B (Step 2).

Step 1: If \((A, C_K)\) is not detectable for some set of sensors \(K\), then at least one of the following conditions holds.

(i) For some \(z \in \mathbb{R}\), \(\text{abs}(z) \geq 1\) and \(v \in \mathbb{R}^n\), \(Av = zv\) and \(Cv = 0\).

(ii) For some complex conjugate pairs \(z, \bar{z} \in \mathbb{C}\), \(\text{abs}(z) = \text{abs}(\bar{z}) \geq 1\) and \(v, \bar{v} \in \mathbb{C}^n\), \(Av = zv, A\bar{v} = \bar{z}\bar{v}\), \(Cv = 0\), and \(C\bar{v} = 0\).

Condition (i) implies that

\[
O_t v = 0, \quad t \in \mathbb{N}
\]

and Condition (ii) implies that

\[
O_t(v + \bar{v}) = 0, \quad t \in \mathbb{N}
\]

Step2: We construct two dynamics with the same measurement. There exists two disjoint sets of sensors \(K_1, K_2\) that satisfy \(|K_1| = |K_2| = \gamma\), \(K_1 \cap K_2 \cap K = \emptyset\), and \(K_1 \cup K_2 = S\). Consider first when condition
(i) holds. Since $B$ has full-rank, there exists an impulse disturbance $w(0)$ that produces
\[ x(1) = v, \]
\[ w(t) = 0, t \geq 1 \]
\[ a_i(t) = \begin{cases} -C_i x(t) & i \in K_1 \\ 0 & \text{Otherwise} \end{cases} \]  
(51)
and
\[ x'(1) = 0 \]
\[ w'(t) = 0, t \geq 1 \]
\[ a'_i(t) = \begin{cases} C_i x(t) & i \in K_2 \\ 0 & \text{Otherwise} \end{cases} \]  
(52)
where $C_i$ denotes the $i$-th row of sensing matrix $C$. Using (49), we can show that measurement $y(t)$ under (51) and $y'(t)$ under (52) are identical. However, the state under (51) is $x(t) = z^tv$, the state under (51) is $x'(t) = 0$, and thus their difference $\|x - x'\|_q$ is unbounded. Consider next when condition (ii) holds. There exists an impulse disturbance $w(0)$ that achieves $x(1) = v + \bar{v}$, so let $x(1) = v + \bar{v}$ replace $x(1)$ in (51). Similarly, we can derive from (50) that measurements $y(t)$ and $y'(t)$ are identical, but $x(t) = z^tv + \bar{z}^k \bar{v}$ and $x'(t) = 0$, yielding unbounded $\|x - x'\|_q$. Since at least (i) or (ii) holds, we have proved Condition B.

\[ \square \]

**Lemma 12.** If Condition E holds, then a resilient estimator cannot be constructed.

**Proof (Lemma 12).** Let $\hat{x}$ be the state estimation of any estimator when measurement $y = y'$ is observed. From the the triangle inequality, the estimation error $e = x - \hat{x}$ under $(x, w, a, y)$ and error $e' = x' - \hat{x}$ under $(x', w', a', y')$ satisfies
\[ \|x - x'\|_q \leq \|e\|_q + \|e'\|_q \]  
(53)
This suggests that either $\|e\|_q$ or $\|e'\|_q$ are unbounded. Because
\[ \sup_{\|w\|_p \leq 1, \|a\|_a \leq \gamma} \|e\|_q \geq \max\{\|e\|_q, \|e'\|_q\} \]  
(54)
no estimator can achieve bounded worst-case estimation error.

\[ \square \]

**Proof (Theorem 1).** From Lemma 11, if $(A, C_K)$ being not detectable for some set $K \subset S$ with $|K| = m - 2\gamma$, then Condition B holds. However, due to Lemma 12, Condition B implies that no resilient estimator can be constructed.
ACKNOWLEDGMENT

The authors would like to thank Professor John Doyle and Professor Richard Murray for insightful discussions.

REFERENCES


