Attack-Resilient $\mathcal{H}_2$, $\mathcal{H}_\infty$, and ℓ_1 State Estimator

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

***Abstract*—This paper considers the secure state estimation problem for noisy systems in the presence of sparse sensor integrity attacks. We show a fundamental limitation: that is, 2ρ-detectability is necessary for achieving bounded estimation errors, where ρ is the number of attacks. This condition is weaker than the 2ρ-observability condition typically assumed in the literature. Conversely, we propose a real-time state estimator that achieves the fundamental limitation. The proposed state estimator is inspired by robust control and FDI: that is, it consists of local Luenberger estimators, local residual detectors, and a global fusion process. We show its performance guarantees for $\mathcal{H}_2$, $\mathcal{H}_\infty$, and ℓ_1 systems. Finally, numerical examples show that it has relatively low estimation errors among existing algorithms and average computation time for systems with a sufficiently small number of compromised sensors.**

***Index Terms*—Secure state estimation, fault tolerance, sparse sensor integrity attacks, Luenberger observer, robust control**

I. Introduction

Fault tolerance in Cyber-physical Systems (CPSs) is of great importance [1]–[5]. For example in the power system, false data injection can introduce errors in state estimation and provide financial gains for attackers [6]–[8]. In flights, autonomous vehicles, and the Internet of Things, manipulations in software and sensing can cause human injury and economic damage [9]–[11]. Motivated by these security issues, this paper studies the secure estimation problem for noisy systems in the presence of sensor integrity attacks.

For the secure estimation problem in static systems, robust estimators are extensively studied in the literature. Common robust estimators include the M-estimator, L-estimator, and R-estimator [12]–[14], and they are used to account for sensor integrity attacks in [15]. For the secure estimation problem in dynamical systems, robust control provides tools to deal with noise in estimation and control [16], [17]. Although robust control typically assumes that system disturbances are bounded or follow well-defined distributions, such assumptions may not be valid for sensor faults caused by intelligent attackers [5], [15]. Fault detection and isolation (FDI) also provides methods for identifying and pinpointing faults in sensors [18]–[21]. One common approach of FDI for linear dynamical systems under sensor integrity attacks is to construct *residuals* that take non-zero values only in the presence of faults (see [22] and references therein). The generation of such residuals is possible only when a fault is separable from normal disturbances and modeling uncertainties, which requires certain kinds of system observability.

Y. Nakahira was with the Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, 91125 USA, e-mail: ynakahir@caltech.edu, website: http://users.cms.caltech.edu/˜ynakahir/.

Y. Mo are with the Department of Electrical and Electronic Engineering, Nanyang Technological University, 639798 Singapore, e-mail: ylmo@ntu.edu.sg, website: http://yilinmo.github.io/

Manuscript received in June 2017.

When attackers can change the measurements of a limited number of sensors in large-scale systems, sensor attacks can be modeled as *sparse* but unbounded disturbances. For sparse sensor integrity attacks, recent literature has studied the fundamental limitation and achievable performance to identify the attacks and estimate the system states. Fawzi et al. show that if ρ sensors are compromised, then 2ρ-observability (*i.e.*, the system remains observable after removing any set of 2ρ sensors) is necessary to guarantee perfect attack identification and accurate state estimation for noiseless systems [23]. The authors further propose to solve a ℓ_0 problem to achieve accurate state estimation under the assumption of 2ρ-observability. This work is generalized to noisy systems by Pajic et al. [24], [25]. Shoukry et al. propose to use the Satisfiability Modulo Theory (SMT) solver to harness the complexity in secure estimation [26]. However, the worst case complexity for the ℓ_0 optimization and that of the SMT solver are combinatorial. Moreover, these estimators also have delays, which may cause performance degradation when used for real-time control. To transform the problem into a convex program, Fawzi et al. and Mo et al. propose to use optimization based methods [23], [27]. To address the estimation delays, various Luenberger-like observers are proposed [26]–[31]. It is worth noticing that the estimators proposed in [24]–[27], [29]–[31] require the assumption of 2ρ-observability or stronger to guarantee accurate attack identification and secure state estimation.

In this paper, we consider the fundamental limitation and achievable performance to achieve fault-tolerant estimation. By fault-tolerant estimation, we refer to achieving bounded estimation errors. Compared with fault identification, fault-tolerant estimation requires relaxed assumptions and accounts for potentially non-detectable and non-identifiable attacks in noisy systems. We prove that a necessary condition to achieve fault-tolerant estimation under ρ compromised sensors is that the system is 2ρ-detectable (the system needs to remain detectable after removing any set of 2ρ sensors). This necessary condition suggests that, if a system has many stable modes, then the number of sensors required to achieve fault-tolerant estimation is much smaller than that to achieve fault identification. Conversely, we propose a secure state estimator that guarantees bounded estimation error under the assumption of 2ρ-detectability. The proposed state estimator is inspired by robust control and FDI: that is, it consists of the local Luenberger estimators, the local residual detectors, and a global fusion process. A preliminary version of this paper was presented at the 2015 IEEE Conference on Decision and Control, deriving the worst-case estimation errors in the ℓ_1 system [28]. This paper extends the result of [28] to the $\mathcal{H}_2$ system and the $\mathcal{H}_\infty$ system. To the best of our knowledge, our paper is the first to show that a mixture of two-norm bounded and sparse-unbounded input can produce *two-norm* bounded

output. Finally, numerical examples show that the proposed state estimator has relatively low estimation errors among existing algorithms and average computation time for systems with a sufficiently small number of compromised sensors.

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 $\{x(t)\}_{t\in Z_+}$ 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 $\|x\|_2 \triangleq (\sum_{t=0}^{\infty}\sum_{i=1}^{n}|x_i(t)|^2)^{1/2}$. Similarly, the norms of a truncated sequence $x(0 : T)$ are defined as $\|x(t_1 : t_2)\|_\infty \triangleq \max_{t_1\le t\le t_2}\max_i |x_i(t)|$ and $\|x(t_1 : t_2)\|_2 \triangleq (\sum_{t=t_1}^{t_2}\sum_{i=1}^{n}|x_i(t)|^2)^{1/2}$. Let ℓ_∞ be the space of sequences with bounded infinity-norm, and ℓ_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:

$$x(t+1) = Ax(t) + Bw(t), \quad y(t) = Cx(t) + Dw(t), \quad (1)$$

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

$$G = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right].$$

The transfer function of the system is $\hat{G}(z) = B(zI - A)^{-1}C + D$. For ℓ_p system input and ℓ_q system output, the system norm (namely, *induced norm*) is given by

$$\|G\|_{p\to q} \triangleq \sup_{\|w\|_p\neq 0} \frac{\|y\|_q}{\|w\|_p}. \quad (2)$$

In particular, the induced-norms for $(p, q) = (2, 2), (2, \infty), (\infty, \infty)$ are given by $\mathcal{H}_\infty$, $\mathcal{H}_2$, and $\mathcal{L}_1$ norms, respectively. These induced-norms are bounded when A is strictly stable (*i.e.*, all the eigenvalues of A is in the open unit circle). See [16], [17] for further details.

To construct a linear state estimator with bounded estimation errors, the LTI system (1) is required to be detectable, *i.e.*, there exist some matrix K such that $A + KC$ is stable. Given the matrix K such that $A + KC$ is strictly stable, we can construct a linear estimator

$$\hat{x}(t+1) = A\hat{x}(t) - K(y(t) - C\hat{x}(t)),\ \hat{x}(0) = 0. \quad (3)$$

We define its estimation error e and residual vector r as

$$e(t) \triangleq x(t) - \hat{x}(t), \qquad r(t) \triangleq y(t) - C\hat{x}(t),$$

respectively. The signals e and r satisfy the following dynamics

$$e(t+1) = (A + KC)e(t) + (B + KD)w(t), \quad e(0) = 0$$
$$r(t) = Ce(t) + Dw(t),$$

Thus, the LTI system from w to e, $E(K)$, and the LTI system from w to e, $F(K)$, are respectively given by

$$E(K) = \left[\begin{array}{c|c} A + KC & B + KD \\ \hline I & 0 \end{array}\right] \quad (4)$$

$$F(K) = \left[\begin{array}{c|c} A + KC & B + KD \\ \hline C & D \end{array}\right]. \quad (5)$$

Because $A+KC$ is strictly stable, both $E(K)$ and $F(K)$ have bounded induced-norms. The induced-norms upper-bound the values of $\|e\|_q$ and $\|r\|_q$ as follows.

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

$$\|e\|_q \le \|E(K)\|_{p\to q}, \qquad \|r\|_q \le \|F(K)\|_{p\to q}. \quad (6)$$

III. Problem Formulation

We study the secure state estimation problem in the presence of sensor attacks. Consider the discrete-time LTI system:

$$\begin{aligned} x(t+1) &= Ax(t) + Bw(t), \qquad x(0) = 0 \\ y(t) &= Cx(t) + Dw(t) + a(t), \end{aligned} \quad (7)$$

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*). This system is illustrated in Fig. 1. The time indices $t \in \mathbb{Z}_+$ are non-negative integers and start from zero. The disturbance matrix B has full row rank, otherwise we can always do a Kalman decomposition and work on the controllable space of (A, B) with zero uncontrollable states. Each sensor is indexed by $i \in \{1, \cdots, m\}$ and produces measurement $y_i(t)$, which jointly comprises the measurement vector $y(t) = [y_1(t), \ldots, y_m(t)]^T$. Sensor i is said to be *compromised* if $a_i(t) \neq 0$ at some time $t \in \mathbb{Z}_+$, and *benign* otherwise. The maximum number of sensors that the attacker can compromise is ρ, *i.e.*,

$$\|a\|_0 \le \rho. \quad (8)$$

We call such an attack as *ρ-sparse*. Let $\mathcal{S} \triangleq \{1, \ldots, m\}$ denote the set of all sensors, $\mathcal{C} \subset \mathcal{S}$ denote the set of compromised sensors, and $\mathcal{B} \triangleq \mathcal{S}\backslash\mathcal{C}$ 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:

$$\hat{x}(t) = f_t(y(0 : t-1)). \quad (9)$$

[1]Take the setting of [32] for example. When the system is noisy, the optimization problem $\min_{x_t\in\mathbb{R}^n} \|[y(t)^T, \cdots, y(t+n-1)^T]^T - \mathcal{O}x_t\|$ ($\mathcal{O}$ is the observability matrix) may not give correct set of compromised sensors $\{i : \exists t, a_i(t) \neq 0\}$.

The estimation error of (9) is defined as the difference between the system state and the state estimate:[2]

$$e(t) \triangleq x(t) - \hat{x}(t). \tag{10}$$

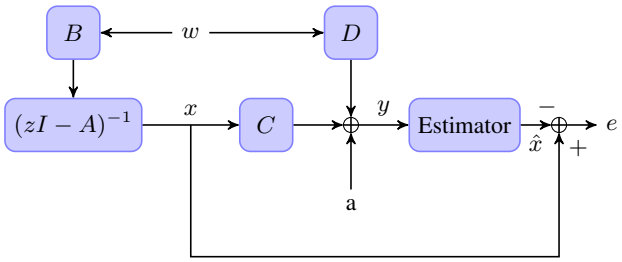

Fig. 1. Diagram of the Estimation Problem in Adversarial Environment.

We study the following worst-case estimation error in q-norm when the input contains a p-norm bounded disturbance and a ρ-sparse attack:

$$\sup_{\|w\|_p \le 1,\, \|a\|_0 \le \rho} \|e\|_q, \tag{11}$$

where the input and ouput norms are

$$(p, q) = (2, 2), (2, \infty), (\infty, \infty). \tag{12}$$

We consider (11) instead of attack isolation because the attack on a noisy system may not be correctable in the sense defined in [23], [32].

Definition 1. *An causal state estimator $\{f_t\}$ is said to be ϵ-resilient to attack if its worst-case estimation error satisfies $\sup_{\|w\|_p \le 1,\, \|a\|_0 \le \rho} \|e\|_q < \epsilon$, where ϵ is a positive and finite scalar.*

When the estimator is ϵ-resilient for some finite $\epsilon > 0$, then we say the state estimator is resilient to attack. The goal of this paper is to study the design problem of a resilient state estimator $\{f_t\}$. Towards that end, we first show a fundamental limitation for the existence of a resilient estimator (Section IV-A), and we then propose a resilient estimator (Section IV-B) and analyze the estimation errors (Section IV-C).

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.

Definition 2 (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 $m' (\le m)$. We define the projection map $P_{\mathcal{I}} : \mathbb{R}^m \to \mathbb{R}^{m'}$ as*

$$P_{\mathcal{I}} = \begin{bmatrix} e_{i_1} & \ldots & e_{i_{m'}} \end{bmatrix}^T \in \mathbb{R}^{m' \times m}. \tag{13}$$

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

Using $P_{\mathcal{I}}$ in (13), 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

$$C_{\mathcal{I}} \triangleq P_{\mathcal{I}} C, \qquad D_{\mathcal{I}} \triangleq P_{\mathcal{I}} D.$$

A. Necessary Condition for Resilience to Attack

In this section, we give a fundamental limitation for achieving bounded worst-case estimation errors.

Definition 3. *The system (7) is said to be χ-detectable if $(A, C_{\mathcal{K}})$ is detectable for any set of sensors $\mathcal{K} \subset \mathcal{S}$ with cardinality $|\mathcal{K}| = m - \chi$.*

Theorem 1. *If system (7) is not 2ρ-detectable, then there is no state estimator $\{f_t\}$ that is ϵ-resilient to attack for any finite $\epsilon > 0$.*

Proof. See the extended version of this paper [33]. □

Theorem 1 implies that the following (denote as Condition A) is necessary for the existence of a resilient state estimator:

A. The system (7) is 2ρ-detectable.

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 sensors $y_{\mathcal{I}}$. In the global fusion process, the state estimate $\hat{x}$ 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).[3] 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* $\hat{x}^{\mathcal{I}}$:[4]

$$\hat{x}^{\mathcal{I}}(t+1) = A\hat{x}^{\mathcal{I}}(t) - K^{\mathcal{I}}(y_{\mathcal{I}}(t) - C_{\mathcal{I}}\hat{x}^{\mathcal{I}}(t)) \tag{14}$$

[3]One way to find the matrix K is via the Riccati equation, *i.e.*, $K^{\mathcal{I}} = PC_{\mathcal{I}}^T(C_{\mathcal{I}}PC_{\mathcal{I}}^T + D_{\mathcal{I}}D_{\mathcal{I}}^T)^{-1}$, where P is unique stabilizing solution of the discrete-time algebraic Riccati equation $P = A(P - PC_{\mathcal{I}}^T(C_{\mathcal{I}}PC_{\mathcal{I}}^T + D_{\mathcal{I}}D_{\mathcal{I}}^T)^{-1}C_{\mathcal{I}}P)A^T + BB^T$. Sufficient conditions the existence of solution P is that $(A, C_{\mathcal{I}})$ is detectable and (A, BB^T) is detectable.

[4]We 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 (13) (e.g. $y_{\mathcal{I}}$ and $C_{\mathcal{I}}$, respectively).

with the initial condition $\hat{x}^{\mathcal{I}}(0) = 0$. The estimation error and residual vector of (14) is respectively defined as

$$e^{\mathcal{I}}(t) \triangleq x(t) - \hat{x}^{\mathcal{I}}(t) \tag{15}$$
$$r^{\mathcal{I}}(t) \triangleq y_{\mathcal{I}}(t) - C_{\mathcal{I}}\hat{x}^{\mathcal{I}}(t) \tag{16}$$

The LTI system from w to $e^{\mathcal{I}}$ is $E^{\mathcal{I}}(K^{\mathcal{I}})$ defined in (4), whereas the LTI system from w to $r^{\mathcal{I}}$ is $F^{\mathcal{I}}(K^{\mathcal{I}})$ defined in (5).

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

$$\|r^{\mathcal{I}}\|_q \leq \|F^{\mathcal{I}}(K^{\mathcal{I}})\|_{p\to q}. \tag{17}$$

Condition (17) can only be violated when the set $\mathcal{I}$ contains compromised sensors, so (17) 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 (17) to determine the validity of its estimate and label local estimator $\mathcal{I}$ to be *invalid* upon observing $\|r^{\mathcal{I}}(0:t)\|_q > \|F^{\mathcal{I}}(K^{\mathcal{I}})\|_{p\to q}$.

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

$$\mathcal{V}(t) \triangleq \left\{\mathcal{I} \in \mathcal{S} : \|r^{\mathcal{I}}(0:t)\|_q \leq \|F^{\mathcal{I}}(K^{\mathcal{I}})\|_{p\to q}\right\}. \tag{18}$$

Using $\mathcal{V}(t)$, we compute the *global state estimate* as follows: $\hat{x}(t) = [\hat{x}_1(t), \hat{x}_2(t), \cdots, \hat{x}_n(t)]$, where

$$\hat{x}_i(t) = \begin{cases} \dfrac{1}{2}\left(\min\limits_{\mathcal{I}\in\mathcal{V}(t)} \hat{x}_i^{\mathcal{I}}(t) + \max\limits_{\mathcal{J}\in\mathcal{V}(t)} \hat{x}_i^{\mathcal{J}}(t)\right) & q = \infty \\ \dfrac{1}{|\mathcal{V}(t)|}\displaystyle\sum_{\mathcal{I}(t)\in\mathcal{V}(t)} \hat{x}_i^{\mathcal{I}}(t) & q = 2. \end{cases} \tag{19}$$

Algorithm 1 The Proposed State Estimator

Initialize $\mathcal{V}(0) \leftarrow \mathcal{V}$ and $\hat{x}^{\mathcal{I}}(0) \leftarrow 0, \mathcal{I} \in \mathcal{V}(0)$
for $t \in \mathbb{N}$ **do**
  for $\mathcal{I} \in \mathcal{V}(t-1)$ **do** (Local Estimation)
    Initialize $\mathcal{V}(t) \leftarrow \emptyset$
    Determine $\hat{x}^{\mathcal{I}}(t)$ from (14) and $r^{\mathcal{I}}(t)$ from (16)
    if $\|r^{\mathcal{I}}(0:t)\|_q \leq \|F^{\mathcal{I}}(K^{\mathcal{I}})\|_{p\to q}$ **then**
      $\mathcal{V}(t) \leftarrow \{\mathcal{V}(t), \mathcal{I}\}$
    end if
  end for
  Obtain estimate $\hat{x}(t)$ from (19) (Global Fusion)
end for

C. Resilience of the Proposed Estimator

Previous works have shown that there exist estimators that can detect the attacks and recover the exact state for noiseless systems if the system is 2ρ-observable [23, Proposition 2][31, Theorem 1][29, Theorem 3.2]. In this section, we show that the proposed estimator is resilient to attack when the system is 2ρ-detectable.

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

$$\begin{cases} \max\limits_{\mathcal{I}\in\mathcal{V}} \|E^{\mathcal{I}}(K^{\mathcal{I}})\|_\infty + \max\limits_{\mathcal{I},\mathcal{J}\in\mathcal{V}} \sqrt{\frac{1}{2}\log|\mathcal{V}|}\ \mathcal{D}_{2,2}^{\mathcal{I},\mathcal{J}} & \text{if } (p,q) = (2,2) \\ \max\limits_{\mathcal{I},\mathcal{J}\in\mathcal{V}} \left(\|E^{\mathcal{I}}(K^{\mathcal{I}})\|_2 + \frac{1}{2}\mathcal{D}_{2,\infty}^{\mathcal{I},\mathcal{J}}\right) & \text{if } (p,q) = (2,\infty) \\ \max\limits_{\mathcal{I},\mathcal{J}\in\mathcal{V}} \left(\|E^{\mathcal{I}}(K^{\mathcal{I}})\|_1 + \frac{1}{2}\mathcal{D}_{\infty,\infty}^{\mathcal{I},\mathcal{J}}\right) & \text{if } (p,q) = (\infty,\infty). \end{cases}$$

In the above formula, the term $\mathcal{D}_{p,q}^{\mathcal{I},\mathcal{J}}$ is defined as

$$\mathcal{D}_{p,q}^{\mathcal{I},\mathcal{J}} = \alpha_{p,q}^{\mathcal{I}\cap\mathcal{J}}(\beta_{p,q}^{\mathcal{I},\mathcal{I}\cap\mathcal{J}} + \beta_{p,q}^{\mathcal{J},\mathcal{I}\cap\mathcal{J}})$$
$$\alpha_{p,q}^{\mathcal{K}} \triangleq \inf_{K: A+KC_{\mathcal{K}} \text{ strictly stable}} \left\| \left[\begin{array}{c|c} A+KC_{\mathcal{K}} & \begin{bmatrix} I & K \end{bmatrix} \\ \hline I & 0 \end{array}\right] \right\|_{p\to q}$$
$$\beta_{p,q}^{\mathcal{I},\mathcal{K}} \triangleq \left\| \begin{bmatrix} -K^{\mathcal{I}} \\ P_{\mathcal{K},\mathcal{I}} \end{bmatrix} \right\|_{p\to q} \|r^{\mathcal{I}}(0:T)\|_p,$$

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

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. The estimation error upper-bound in Theorem 2 decomposes into two terms: $\|E^{\mathcal{I}}(K^{\mathcal{I}})\|_{p\to q}$ and the remaining. The first term $\|E^{\mathcal{I}}(K^{\mathcal{I}})\|_{p\to q}$ characterize the error between a local estimator and the true state. That is, if the local estimator $\mathcal{I}$ is used for a system with no attack ($a \equiv 0$), then its estimation error is bounded by $\|E^{\mathcal{I}}(K^{\mathcal{I}})\|_{p\to q}$. The second term exists due to the attack in an unknown set of sensors. 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. Moreover, it shall be noted that an increase in the tolerable number of compromised sensors ρ may result in an increase in both terms, thus increasing the worst-case estimation error $\sup_{\|w\|_p\leq 1, \|a\|_0 = 0} \|e\|_q$.

Corollary 1. *A necessary and sufficient condition for the existence of an ϵ-resilient estimator for some finite $\epsilon > 0$ is that $(A, C_{\mathcal{K}})$ is detectable for any index set $\mathcal{K} \subset \mathcal{S}$ with cardinality $m - 2\rho$.*

Corollary 2. *Consider system (7) with ρ-sparse attack. The state estimator in Algorithm 1 is ϵ-resilient to attack for some finite $\epsilon > 0$.*

V. Proof for Resilience of the 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 [33]. 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

$$\begin{aligned} \|e\|_q &= \left\|\left(x - \hat{x}^{\mathcal{I}}\right) + \left(\hat{x}^{\mathcal{I}} - \hat{x}\right)\right\|_q \\ &\leq \left\|x - \hat{x}^{\mathcal{I}}\right\|_q + \left\|\hat{x}^{\mathcal{I}} - \hat{x}\right\|_q. \end{aligned} \tag{20}$$

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 (8). The first term $\|x-\hat{x}^{\mathcal{I}}\|_q$ can be bounded using Lemma 1 by

$$\|x-\hat{x}^{\mathcal{I}}\|_q \leq \|E^{\mathcal{I}}(K^{\mathcal{I}})\|_{p\to q}. \tag{21}$$

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

$$\mathcal{J}_1, \mathcal{J}_2, \cdots, \mathcal{J}_{|\mathcal{V}|}. \tag{22}$$

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

$$\|\hat{x}^{\mathcal{J}_1}(0:T)-\hat{x}^{\mathcal{J}_2}(0:T)\|_q \leq \mathcal{D}_{p,q}^{\mathcal{J}_1,\mathcal{J}_2}, \tag{23}$$

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

Lemma 3. *If condition* (23) *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*

$$\|\hat{x}^{\mathcal{I}}-\hat{x}\|_2 \leq \max_{\mathcal{J}_1,\mathcal{J}_2\in\mathcal{V}} \sqrt{\frac{1}{2}\log|\mathcal{V}|}\, \mathcal{D}_{p,q}^{\mathcal{J}_1,\mathcal{J}_2}. \tag{24}$$

Lemma 4. *If condition* (23) *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*

$$\|\hat{x}^{\mathcal{I}}-\hat{x}\|_\infty \leq \frac{1}{2}\max_{\mathcal{J}\in\mathcal{V}} \mathcal{D}_{p,\infty}^{\mathcal{I},\mathcal{J}}. \tag{25}$$

We prove Lemma 2, Lemma 3 in Section V-A, Section V-B, respectively. The proof of Lemma 4 is given in the extended version of this paper [33]. 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

$$\sup_{\substack{\|w\|_2\leq 1\\ \|a\|_0\leq\rho}} \|e\|_2 \leq \max_{\mathcal{I}\in\mathcal{V}} \|E^{\mathcal{I}}(K^{\mathcal{I}})\|_\infty + \max_{\mathcal{J}_1,\mathcal{J}_2\in\mathcal{V}} \sqrt{\frac{1}{2}\log|\mathcal{V}|}\, \mathcal{D}_{2,2}^{\mathcal{I},\mathcal{J}}.$$

Applying similar argument for the case of $q=\infty$, we obtain

$$\|e\|_\infty \leq \max_{\mathcal{I},\mathcal{J}\in\mathcal{V}} \left(\|E^{\mathcal{I}}(K^{\mathcal{I}})\|_{p\to\infty} + \frac{1}{2}\mathcal{D}_{p,\infty}^{\mathcal{I},\mathcal{J}} \right),$$

where $p=2,\infty$. □

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 $\|w\|_p \leq 1$. If $y(t)=0$ for all $t=0,1,\cdots,T$, then*

$$\|x(0:T)\|_q \leq \inf_{K:A+KC \text{ strictly stable}} \|E(K)\|_{p\to q}, \tag{26}$$

where $E(K)$ is given in (4)*.*

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 (3). Since $y(0:T)=0$, the state estimator (3) 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 (26). □

Proof (Lemma 2). Let $\mathcal{J}_1, \mathcal{J}_2 \in \mathcal{V}(T)$. We first compute the dynamics of the local estimates $\hat{x}^{\mathcal{J}_i}(t)$, $i=1,2$. From (14) and (16),

$$\begin{aligned} &\hat{x}^{\mathcal{J}_i}(t+1) = A\hat{x}^{\mathcal{J}_i}(t) - K^{\mathcal{J}_i} r^{\mathcal{J}_i}(t), \hat{x}^{\mathcal{J}_i}(0)=0 \\ &y_{\mathcal{J}_i}(t) = C_{\mathcal{J}_i}\hat{x}^{\mathcal{J}_i}(t) + r^{\mathcal{J}_i}(t) \end{aligned} \tag{27}$$

for $t \leq T$. We define the sequences $\phi^{\mathcal{J}_i}(t)$ and $\varphi^{\mathcal{J}_i}(t)$ by

$$\phi^{\mathcal{J}_i}(t) \triangleq -K^{\mathcal{J}_i} r^{\mathcal{J}_i}(t), \qquad \varphi^{\mathcal{J}_i}(t) \triangleq P_{\mathcal{K}_{1,2},\mathcal{J}_i} r^{\mathcal{J}_i}(t),$$

where $P_{\mathcal{K}_{1,2},\mathcal{J}_i} \in \mathbb{R}^{|\mathcal{K}_{1,2}|\times|\mathcal{J}_i|}$ is the unique solution of $P_{\mathcal{K}_{1,2}} = P_{\mathcal{K}_{1,2},\mathcal{J}_i} P_{\mathcal{J}_i}$. Let $\mathcal{K}_{1,2} = \mathcal{J}_1 \cap \mathcal{J}_2$ the intersection between the two sets $\mathcal{J}_1, \mathcal{J}_2$. As the measurements from subset $\mathcal{K}_{1,2} \subset \mathcal{J}_i$ also satisfies $y_{\mathcal{K}_{1,2}}(t) = C_{\mathcal{K}_{1,2}}\hat{x}^{\mathcal{J}_i}(t) + P_{\mathcal{K}_{1,2},\mathcal{J}_i} r^{\mathcal{J}_i}(t)$, combining with (27) yields

$$\begin{aligned} &\hat{x}^{\mathcal{J}_i}(t+1) = A\hat{x}^{\mathcal{J}_i}(t) + \begin{bmatrix} I & 0 \end{bmatrix} \begin{bmatrix} \phi^{\mathcal{J}_i}(t) \\ \varphi^{\mathcal{J}_i}(t) \end{bmatrix}, \hat{x}^{\mathcal{J}_i}(0)=0 \\ &y_{\mathcal{K}_{1,2}}(t) = C_{\mathcal{K}_{1,2}}\hat{x}^{\mathcal{J}_i}(t) + \begin{bmatrix} 0 & I \end{bmatrix} \begin{bmatrix} \phi^{\mathcal{J}_i}(t) \\ \varphi^{\mathcal{J}_i}(t) \end{bmatrix}. \end{aligned} \tag{28}$$

Now, let $\Delta(t)$ be the difference between the local estimator $\mathcal{J}_1$ and local estimator $\mathcal{J}_2$, *i.e.*,

$$\Delta(t) \triangleq \hat{x}^{\mathcal{J}_1}(t) - \hat{x}^{\mathcal{J}_2}(t). \tag{29}$$

Subtracting the equation (28) for $\mathcal{J}_1$ from equation (28) for $\mathcal{J}_2$, we obtain the dynamics of Δ as follows:

$$\begin{aligned} &\Delta(t+1) = A\Delta(t) + \begin{bmatrix} I & 0 \end{bmatrix} \begin{bmatrix} \phi^{\mathcal{J}_1}(t) - \phi^{\mathcal{J}_2}(t) \\ \varphi^{\mathcal{J}_1}(t) - \varphi^{\mathcal{J}_2}(t) \end{bmatrix}, \Delta(t)=0, \\ &0 = C_{\mathcal{K}_{1,2}}\Delta(t) + \begin{bmatrix} 0 & I \end{bmatrix} \begin{bmatrix} \phi^{\mathcal{J}_1}(t) - \phi^{\mathcal{J}_2}(t) \\ \varphi^{\mathcal{J}_1}(t) - \varphi^{\mathcal{J}_2}(t) \end{bmatrix}. \end{aligned}$$

Because a valid set satisfies (18), the residual vectors of estimator $\mathcal{J}_i$, $i=1,2$, are bounded by $\|r^{\mathcal{J}_i}(0:T)\|_q \leq \|F^{\mathcal{J}_i}(K^{\mathcal{J}_i})\|_{p\to q}$, which results in

$$\begin{aligned} \left\| \begin{bmatrix} \phi^{\mathcal{J}_i}(0:T) \\ \varphi^{\mathcal{J}_i}(0:T) \end{bmatrix} \right\|_p &\leq \left\| \begin{bmatrix} -K^{\mathcal{J}_i} \\ P_{\mathcal{K}_{1,2},\mathcal{J}_i} \end{bmatrix} \right\|_{p\to p} \|r^{\mathcal{J}_i}(0:T)\|_p \\ &= \beta_{p,q}^{\mathcal{J}_i,\mathcal{J}_1\cap\mathcal{J}_2}. \end{aligned} \tag{30}$$

From the triangle inequality, we obtain

$$\left\| \begin{bmatrix} \phi^{\mathcal{J}_1}(0:T) - \phi^{\mathcal{J}_2}(0:T) \\ \varphi^{\mathcal{J}_1}(0:T) - \varphi^{\mathcal{J}_2}(0:T) \end{bmatrix} \right\|_p \leq \beta_{p,q}^{\mathcal{J}_1, \mathcal{J}_1 \cap \mathcal{J}_2} + \beta_{p,q}^{\mathcal{J}_2, \mathcal{J}_1 \cap \mathcal{J}_2}.$$

Substitute $\phi^{\mathcal{J}_1} - \phi^{\mathcal{J}_2}$ for u in Lemma 5 and $\Delta(t)$ for x, we obtain

$$\|\hat{x}^{\mathcal{J}_1}(0:T) - \hat{x}^{\mathcal{J}_2}(0:T)\|_q \leq \alpha_{p,q}^{\mathcal{J}_1 \cap \mathcal{J}_2} (\beta_{p,q}^{\mathcal{J}_1, \mathcal{J}_1 \cap \mathcal{J}_2} + \beta_{p,q}^{\mathcal{J}_1, \mathcal{J}_1 \cap \mathcal{J}_2}).$$

□

B. Proof of Lemma 3

Define the following two optimization problems:

$$\begin{aligned} \mathcal{P}_\delta(n) := \max_{z_k(i) \geq 0} & \sum_{i=1}^{n} \left(\frac{1}{n-i-1} \right)^2 \left(\sum_{k=i}^{n} z_k(i) \right)^2 \\ \text{s.t.} \quad & \sum_{i=0}^{k} (z_k(i))^2 \leq \delta, \quad k = 1, 2, \ldots, n \\ \mathcal{D}_\delta(n) := \min_{\lambda_i > 0} & \sum_{i=1}^{n} \lambda_i \\ \text{s.t.} \quad & \sum_{i=1}^{j} \frac{1}{\lambda_i} \leq (j+1)^2, \quad j = 1, 2, \ldots, n. \end{aligned}$$

With the 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 show that $\mathcal{P}_\delta(N-1)$ with

$$\delta = \max_{\mathcal{J}_1, \mathcal{J}_2 \in \mathcal{V}} \left(\mathcal{D}_{2,2}^{\mathcal{J}_1, \mathcal{J}_2} \right)^2 \tag{31}$$

is an upper-bound of $\left\| \hat{x}^{\mathcal{I}} - \hat{x} \right\|_2$ (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 9).

Lemma 6. *If condition* (23) *holds for* $(p, q) = (2, 2)$*, then the divergence between the benign estimator* $\mathcal{I}$ *and the global estimator satisfies*

$$\|\hat{x}^{\mathcal{I}} - \hat{x}\|_2^2 \leq \mathcal{P}_\delta(N-1), \tag{32}$$

where $N = |\mathcal{V}|$ *and* $\delta = \max_{\mathcal{J}_1, \mathcal{J}_2 \in \mathcal{V}} \left(\mathcal{D}_{2,2}^{\mathcal{J}_1, \mathcal{J}_2} \right)^2$.

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

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

where $\mathcal{I}_N$ is the benign estimator ($\mathcal{I}_N$ only contains benign sensors). We define $\{x^{\mathcal{I}_i}(t)\}_{t \leq T_i}$ as a finite or sequence of length $T_i + 1$ for finite T_i and infinite sequence for $T_i = \infty$. Let a finite sequence $z_k(i)$, $k = 1, \cdots, N-1$, $i = 0, \cdots, k$ be defined as

$$z_k(i) = \left(\sum_{t=T_i+1}^{T_{i+1}} \|\hat{x}^{\mathcal{I}_N}(t) - \hat{x}^{\mathcal{I}_k}(t)\|_2^2 \right)^{\frac{1}{2}}.$$

Using $z_k(i)$, we obtain an upper-bound for the divergence between the benign estimator $\mathcal{I}_N$ and global estimator as follows:

$$\begin{aligned} & \|\hat{x}^{\mathcal{I}_N} - \hat{x}\|_2^2 \\ &= \sum_{i=0}^{N-2} \sum_{t=T_i+1}^{T_{i+1}} \left\| \frac{1}{N-i} \sum_{k=i+1}^{N-1} \left(\hat{x}^{\mathcal{I}_N}(t) - \hat{x}^{\mathcal{I}_k}(t) \right) \right\|_2^2 \\ &\leq \sum_{i=0}^{N-2} \left(\frac{1}{N-i} \right)^2 \left(\sum_{k=i+1}^{N-1} z_k(i) \right)^2 \end{aligned}$$

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

$$\begin{aligned} \max_{z_k(i) \geq 0} \quad & \sum_{i=0}^{N-2} \left(\frac{1}{N-i} \right)^2 \left(\sum_{k=i+1}^{N-1} z_k(i) \right)^2 \\ \text{s.t.} \quad & \sum_{i=0}^{k} (z_k(i))^2 \leq \max_{\mathcal{J}_1, \mathcal{J}_2 \in \mathcal{V}} \left(\mathcal{D}_{2,2}^{\mathcal{J}_1, \mathcal{J}_2} \right)^2, k = 1, \ldots, N-1, \end{aligned} \tag{33}$$

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

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. The proof of Lemma 8 is given in the extended version of this paper [33].

Lemma 8. *The following two inequalities are equivalent*

$$\Lambda = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) \geq \mathbf{1}\mathbf{1}^T \tag{34}$$

$$\sum_{i=1}^{n} \frac{1}{\lambda_i} \leq 1, \text{ and } \lambda_j > 0, \quad j = 1, \ldots, n \tag{35}$$

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

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

Let the following matrices be defined as

$$\begin{aligned} X &= vv^T \geq 0 \\ F_0 &= \text{diag}\left(\frac{1}{2^2}, \cdots, \frac{\mathbf{1}_{n-1}\mathbf{1}_{n-1}^T}{(n)^2}, \frac{\mathbf{1}_n \mathbf{1}_n^T}{(n+1)^2} \right) \\ F_i &= \text{diag}\left(e_{n,i}, e_{n-1,i}, \cdots, e_{n-i+1,i}, 0_{n-i}, \cdots, 0_1 \right) \end{aligned} \tag{36}$$

where 1_k is a k-dimensional vector with all elements being 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 being 0. Using SDP relaxation [34], the problem $\mathcal{P}_\delta(n)$ can be converted into

$$\begin{aligned} \mathcal{P}'_\delta(n) = \max_{X \geq 0} \quad & \text{tr}(F_0 X) \\ \text{s.t.} \quad & \text{tr}(F_i X) \leq \delta, \forall i = 1, \ldots, n \end{aligned}$$

Hence $\mathcal{P}_\delta(n) \le \mathcal{P}'_\delta(n)$. This relaxation can be observed from the following relations:

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

We next show that the relaxation of the problem $\mathcal{P}_\delta(n)$ to the semidefinite problem $\mathcal{P}'_\delta(n)$ is also exact. Assume that $\mathcal{P}_\delta(n)$ is feasible and bounded. Let $X^* = \{x^*_{ij}\}$ be the optimal solution of $\mathcal{P}'_\delta$. Define X as

$$X = \begin{bmatrix}\sqrt{x^*_{11}} & \cdots & \sqrt{x^*_{nn}}\end{bmatrix}^T \begin{bmatrix}\sqrt{x^*_{11}} & \cdots & \sqrt{x^*_{nn}}\end{bmatrix}.$$

From $x_{ii} = x^*_{ii}$ and (36), X satisfies the contraints

$$\mathrm{tr}(F_i X) = \mathrm{tr}(F_i X^*) = \delta. \tag{37}$$

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

$$\mathrm{tr}(F_0 X^*) \ge \mathrm{tr}(F_0 X) \ge \mathrm{tr}(F_0 X^*). \tag{38}$$

Therefore, (37) and (38) shows that $\mathcal{P}_\delta(n) = \mathcal{P}'_\delta(n)$.

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

$$\begin{aligned}\mathcal{D}'_\delta(n) = \min_{\lambda}\ & \delta\sum_{i=1}^{n}\lambda_i \\ \text{s.t.}\ & \sum_{i=1}^{n}\lambda_i F_i \ge F_0.\end{aligned} \tag{39}$$

Because there exists an strictly positive definite matrix $X > 0$ such that $\mathrm{tr}(F_i X) = \delta$ for all $i = 1, 2, \cdots, n$, from Slater's condition [35], strong duality holds between $\mathcal{P}'_\delta(n)$ and $\mathcal{D}'_\delta(n)$. Let $\Lambda_{1:i} = \mathrm{diag}(\lambda_1, \ldots, \lambda_i) \in \mathbb{R}^{i\times i}$, and observe that $\sum_{i=1}^{n}\lambda_i F_i = \mathrm{diag}(\Lambda_{1:1}, \Lambda_{1:2}, \ldots, \Lambda_{1:n})$. Hence, the constraint $\sum_{i=1}^{n}\lambda_i F_i \ge F_0$ is equivalent to

$$\begin{aligned}&\mathrm{diag}(\lambda_1, \ldots, \lambda_j) \ge \frac{1}{(j+1)^2}\mathbf{1}\mathbf{1}^T, \forall j = 1, \ldots, n, \\ \iff &\sum_{i=1}^{j}\frac{1}{\lambda_i} \le (j+1)^2,\ \ \lambda_j > 0,\ \ j = 1, \ldots, n,\end{aligned}$$

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

Lemma 9. *The solution of the problem $\mathcal{D}_\delta(n)$ satisfies*

$$\mathcal{D}_\delta(n) = \delta\left\{\frac{1}{4} + \sum_{i=2}^{n}\frac{1}{2i+1}\right\} \le \frac{1}{2}\delta\log(n+1). \tag{40}$$

The proof of Lemma 9 is given in the extended version of this paper [33]. Now we are ready to prove Lemma 3.

Proof (Lemma 3). Applying Lemma 2 and Lemma 6, Lemma 7, and then Lemma 9 consecutively, we obtain

$$\|\hat{x}^{\mathcal{I}} - \hat{x}\|_2^2 \le \mathcal{P}_\delta = \mathcal{D}_\delta \le \frac{1}{2}\log(N)\max_{\mathcal{J}_1, \mathcal{J}_2 \in \mathcal{V}}\left(\mathcal{D}^{\mathcal{J}_1, \mathcal{J}_2}_{2,2}\right)^2.$$

□

VI. Numerical examples

In this section, we study the proposed estimator numerically and compare it with existing algorithms from Shoukry et al. [26], Chong et al. [31], Pajic et al. [32], and Lu et al. [30]. We tested the IEEE 14-Bus system, the Unmanned Ground Vehicle (UGV), and the temperature monitor as follows.

(i) *IEEE 14-Bus system [8], [26], [36]:* The IEEE 14-Bus system is modeled as the system (7) with A and C given in [36]. We additionally add process noise and sensor noise by setting $B = \begin{bmatrix}I_{10} & O_{10,35}\end{bmatrix}$ and $D = \begin{bmatrix}O_{10,35} & I_{35}\end{bmatrix}$.

(ii) *Unmanned ground vehicle (UGV) [26], [32]:* A UGV moving in a straight line has the dynamics

$$\begin{bmatrix}\dot{p}\\ \dot{v}\end{bmatrix} = \begin{bmatrix}0 & 1\\ 0 & -b/m\end{bmatrix}\begin{bmatrix}\dot{x}\\ \dot{v}\end{bmatrix} + \begin{bmatrix}0\\ 1/m\end{bmatrix}u + \begin{bmatrix}I_2 & O_{2,3}\end{bmatrix}w,$$

where p is the position, v is the velocity, u is the force input, w is the disturbance, m is the mechanical mass, b is the translational friction coefficient, I_n is a n-dimensional identity matrix, and $O_{n,m}$ is a $m \times n$ zero matrix. We assume that the estimator can access the values of u. The UGV is equipped with a sensor measuring x and two sensors measuring v, *i.e.*,

$$y = \begin{bmatrix}1 & 0\\ 0 & 1\\ 0 & 1\end{bmatrix}\begin{bmatrix}p\\ v\end{bmatrix} + \begin{bmatrix}O_{3,2} & I_2\end{bmatrix}w.$$

The system parameters are the same as [26]: $m = 0.8kg$, $b = 1$, and sampling interval $T_s = 0.1s$.

(iii) *Temperature monitor [37]:* The heat process in a plannar closed region $(z_1, z_2) \in [0, l] \times [0, l]$ can be expressed by

$$\frac{\partial x}{\partial t} = \alpha\left(\frac{\partial^2 x}{\partial z_1^2} + \frac{\partial^2 x}{\partial z_2^2}\right), \tag{41}$$

where α is the speed of the diffusion process; and $x(z_1, z_2)$ is the temperature at position (z_1, z_2) subject to the boundary conditions

$$\left.\frac{\partial x}{\partial z_1}\right|_{t,0,z_2} = \left.\frac{\partial x}{\partial z_1}\right|_{t,l,z_2} = \left.\frac{\partial x}{\partial z_1}\right|_{t,z_1,0} = \left.\frac{\partial x}{\partial z_1}\right|_{t,z_1,l} = 0.$$

We discretize the region using a $N \times N$ grid and the continuous-time with sampling interval T_s to model (41) into (7). We additionally add process noise and sensor noise by setting w, $B = \begin{bmatrix}I_9 & O_{9,20}\end{bmatrix}$ and $D = \begin{bmatrix}O_{9,20} & I_{20}\end{bmatrix}$. We set $\alpha = 0.1m^2/s$, $l = 4m$, and $N = 5$ as in [37].

The noise w is generated from a uniform distribution between $[-1, 1]$. The time horizon is set to be $T = 100$. The number of compromised sensors is set to be 1; the compromised sensor is randomly chosen among non-critical sensors (*i.e.*, sensors that can be removed without losing observability). The attack signal is drawn from a Gaussian distribution with mean zero and variance 10^4. For the proposed algorithm, we used (17) with $(p, q) = (2, 2), (2, \infty)$, and (∞, ∞) simultaneously. We ran each example for 100 times and recorded their average estimation errors $\|e\|_2$ and computation times. The code is written in Matlab (Windows) and runs on an Intel Core i5-4690 Processor (4x3.50GHz/6MB L3

	Proposed	[26]	[31]	[24]	[30]
(i)	11.1573	17.2948	13.7927	48.9880	17.5957
(ii)	6.9108	4.7246	7.2937	22.0884	NA
(iii)	6.7833	7.9424	6.8902	8.5448	18.5803

TABLE I

ESTIMATION ERRORS IN TWO-NORM, *i.e.*, $\|e(1:100)\|_2$.

	Proposed	[26]	[31]	[24]	[30]
(i)	0.0062	0.0108	0.0045	0.0006	0.0345
(ii)	0.0001	0.0026	0.0002	0.0003	NA
(iii)	0.0023	0.0096	0.0020	0.0005	0.0048

TABLE II

AVERAGE COMPUTATION TIME FOR TIME HORIZON 100.

Cache). We summarize the estimation errors and computation times in Table I and Table II. Some entries are left as NA (not applicable) because the system (ii) does not satisfy the linear matrix inequality (LMI) assumption required by the algorithm proposed by [30]. The algorithms tested have different relative accuracies and computation times from example to example. Among all examples tested, the proposed algorithm has relatively low estimation errors and average computation times.

VII. CONCLUSION

In this paper, we propose a real-time state estimator for noisy systems that is resilient to sparse sensor integrity attacks. The proposed estimator is stable under relaxed assumptions. Its worst-case estimation errors are $O(\log \binom{m}{\rho})$ in the $\mathcal{H}_2$ system, $O(1)$ in the $\mathcal{H}_\infty$ system, and $O(1)$ in the ℓ_1 system. Its computational complexity is $O(\log \binom{m}{\rho})$.

ACKNOWLEDGMENT

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

REFERENCES

[1] A. A. Cardenas, S. Amin, and S. Sastry, "Secure control: towards survivable cyber-physical systems," in *IEEE 28th International Conference on Distributed Computing Systems Workshops (ICDCS)*. IEEE, 2008, pp. 495–500.

[2] A. Cardenas, S. Amin, B. Sinopoli, A. Giani, A. Perrig, and S. Sastry, "Challenges for securing cyber physical systems," in *Workshop on future directions in cyber-physical systems security*, vol. 5, 2009.

[3] R. R. Rajkumar, I. Lee, L. Sha, and J. Stankovic, "Cyber-physical systems: the next computing revolution," in *Proceedings of the 47th Design Automation Conference*. ACM, 2010, pp. 731–736.

[4] S. Sridhar, A. Hahn, and M. Govindarasu, "Cyber–physical system security for the electric power grid," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 210–224, 2012.

[5] A. Datta, S. Kar, B. Sinopoli, and S. Weerakkody, "Accountability in cyber-physical systems," in *Science of Security for Cyber-Physical Systems Workshop (SOSCYPS)*. IEEE, 2016, pp. 1–3.

[6] H. Sandberg, A. Teixeira, and K. H. Johansson, "On security indices for state estimators in power networks," in *First Workshop on Secure Control Systems*, 2010.

[7] L. Xie, Y. Mo, and B. Sinopoli, "Integrity data attacks in power market operations," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 659–666, 2011.

[8] Y. Liu, M. Reiter, and P. Ning, "False data injection attacks against state estimation in electric power grids," in *Proceedings of the 16th ACM conference on Computer and communications security*, 2009.

[9] J. P. Farwell and R. Rohozinski, "Stuxnet and the future of cyber war," *Survival*, vol. 53, no. 1, pp. 23–40, 2011.

[10] M. Maurer, J. C. Gerdes, B. Lenz, and H. Winner, *Autonomous driving: technical, legal and social aspects*. Springer, 2016.

[11] C. Nobles, "Cyber threats in civil aviation," *Security Solutions for Hyperconnectivity and the Internet of Things*, p. 272, 2016.

[12] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: A survey," *Proceedings of the IEEE*, vol. 73, no. 3, pp. 433–481, 1985.

[13] R. A. Maronna, D. R. Martin, and V. J. Yohai, *Robust Statistics: Theory and Methods*. Wiley, 2006.

[14] P. J. Huber and E. M. Ronchetti, *Robust Statistics*. Wiely, 2009.

[15] Y. Mo and B. Sinopoli, "Secure estimation in the presence of integrity attacks," *IEEE Transactions on Automatic Control*, vol. 60, no. 4, pp. 1145–1151, 2015.

[16] M. A. Dahleh and I. J. Diaz-Bobillo, *Control of uncertain systems: a linear programming approach*. Prentice-Hall, Inc., 1994.

[17] K. Zhou, J. C. Doyle, and K. Glover, *Robust and optimal control*. Prentice hall New Jersey, 1996, vol. 40.

[18] J. Gertler, *Fault detection and diagnosis in engineering systems*. CRC press, 1998.

[19] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri, "A review of process fault detection and diagnosis: Part i: Quantitative model-based methods," *Computers & chemical engineering*, vol. 27, no. 3, pp. 293–311, 2003.

[20] R. Isermann, *Fault-diagnosis systems: an introduction from fault detection to fault tolerance*. Springer Science & Business Media, 2006.

[21] J. Chen and R. J. Patton, *Robust model-based fault diagnosis for dynamic systems*. Springer Science & Business Media, 2012, vol. 3.

[22] R. J. Patton, P. M. Frank, and R. N. Clark, *Issues of fault diagnosis for dynamic systems*. Springer Science & Business Media, 2013.

[23] 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.

[24] 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*, vol. 37, no. 2, pp. 66–81, 2017.

[25] 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.

[26] 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*, 2017.

[27] Y. Mo and E. Garone, "Secure dynamic state estimation via local estimators," in *Decision and Control (CDC), 2016 IEEE 55th Conference on*. IEEE, 2016, pp. 5073–5078.

[28] Y. Nakahira and Y. Mo, "Dynamic state estimation in the presence of compromised sensory data," in *IEEE 54th Annual Conference on Decision and Control (CDC)*. IEEE, 2015, pp. 5808–5813.

[29] 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.

[30] A.-Y. Lu and G.-H. Yang, "Secure state estimation for cyber-physical systems under sparse sensor attacks via a switched luenberger observer," *Information sciences*, vol. 417, pp. 454–464, 2017.

[31] M. S. Chong, M. Wakaiki, and J. P. Hespanha, "Observability of linear systems under adversarial attacks," in *American Control Conference (ACC)*. IEEE, 2015, pp. 2439–2444.

[32] M. Pajic, J. Weimer, N. Bezzo, P. Tabuada, O. Sokolsky, I. Lee, and G. J. Pappas, "Robustness of attack-resilient state estimators," in *ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS)*. IEEE, 2014, pp. 163–174.

[33] Y. Nakahira and Y. Mo, "Attack-Resilient $\mathcal{H}_2$, $\mathcal{H}_\infty$, and ℓ_1 State Estimator (Extended Version)," Tech. Rep. [Online]. Available: http://users.cms.caltech.edu/~ynakahir/security.pdf

[34] S. H. Low, "Convex relaxation of optimal power flow: A tutorial," in *Bulk Power System Dynamics and Control-IX Optimization, Security and Control of the Emerging Power Grid (IREP), 2013 IREP Symposium*. IEEE, 2013, pp. 1–15.

[35] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[36] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Transactions on power systems*, vol. 26, no. 1, pp. 12–19, 2011.

[37] Y. Mo, R. Ambrosino, and B. Sinopoli, "Sensor selection strategies for state estimation in energy constrained wireless sensor networks," *Automatica*, vol. 47, no. 7, pp. 1330–1338, 2011.