Dynamic State Estimation in the Presence of Compromised Sensory Data

Yorie Nakahira, Yilin Mo

Abstract

In this article, we consider the state estimation problem of a linear time invariant system in adversarial environment. We assume that the process noise and measurement noise of the system are l_∞ functions. The adversary compromises at most γ sensors, the set of which is unknown to the estimation algorithm, and can change their measurements arbitrarily. We first prove that if after removing a set of 2γ sensors, the system is undetectable, then there exists a destabilizing noise process and attacker's input to render the estimation error unbounded. For the case that the system remains detectable after removing an arbitrary set of 2γ sensors, we construct a resilient estimator and provide an upper bound on the l_∞ norm of the estimation error. Finally, a numerical example is provided to illustrate the effectiveness of the proposed estimator design.

I. Introduction

The increasing use of networked embedded sensors to monitor and control critical infrastructures provides potential malicious agents with the opportunity to disrupt their operations by corrupting sensor measurements. Supervisory Control And Data Acquisition (SCADA) systems, for example, run a wide range of safety critical plants and processes, including manufacturing, water and gas treatment and distribution, facility control and power grids. A wide variety of motivations exist for launching an attack on the such systems, ranging from financial reasons, e.g., reducing the electricity bill, all the way to terrorism, e.g., threatening the life of possibly an entire population by controlling electricity and other life-critical resources. A successful attack

Yorie Nakahira and Yilin Mo are with the Control and Dynamical Systems Department of California Institute of Technology. Email: ynakahir@caltech.edu, yilinmo@caltech.edu

This paper is supported in part by TerraSwarm, one of six centers of STARnet, a Semiconductor Research Corporation program sponsored by MARCO and DARPA.

to such kind of systems may significantly hamper the economy, the environment, and may even lead to the loss of human life.

The first-ever SCADA system malware (called Stuxnet) was found in July 2010 and rose significant concern about SCADA system security [1], [2]. The research community has acknowledged the importance of addressing the challenge of designing secure estimation and control systems [3].

We consider a secure estimation problem inspired by security concerns that arise from the possible manipulation of sensory data. We model the underlying system as a linear time invariant system. The goal is to estimate the state of the system via the measurements collected by m sensors, with the caveat that some of these measurements can be manipulated by a malicious third party. The adversary can only manipulate at most γ sensors due to resource limitations. However, it has total control over the corrupted sensors, as it can change the measurements of the compromised sensors arbitrarily. Our goal is to construct a "resilient" estimator, whose estimation error remains bounded regardless of the noise process and the attacker's action.

We first prove that if after removing a set of 2γ sensors, the system becomes undetectable, then no resilient estimator exists, which serves as a fundamental limitation on the estimation performance. On the other hand, we provide a resilient estimator design when the system remains detectable even after removing any arbitrary set of 2γ sensors. We further derive an upper bound on the "worst-case" estimation error of the proposed estimator.

Related Work

The problem of detecting and isolating abnormalities in the systems has been extensively studied in the literature. Bad data detection and identification techniques have been widely used in large scaled systems such as power grids [4]. While such approaches are very successful in detecting and removing random sensor failures, they are not effective against intelligent attacks. Liu et al. [5] illustrate how an adversary can inject a stealthy input into the measurements to change the state estimation, without being detected by the bad data detector. Sandberg et al. [6] consider how to find a sparse stealthy input, which enables the adversary to launch an attack with a minimum number of compromised sensors. Xie et al. [7] further illustrate that the stealthy integrity attacks on state estimation can lead to a financial gain in the electricity market for the adversary. In the context of dynamical systems, a substantial amount of research efforts has

been devoted to Failure Detection and Identification (FDI) algorithms [8], [9]. Recently, FDI techniques have been applied in the security settings by Pasqualetti et al. [10], [11], Sundaram et al. [12] and Fawzi et al. [13] to detect and identify malicious behaviors in consensus networks, power grids, wireless control networks and control systems.

On the other hand, robust estimation techniques can be used to generate the state estimation which is resilient to uncertainties and abnormalities in the sensory data. For static estimation, robust estimators such as M-estimator, L-estimator, R-estimator and etc. have been proposed and widely studied [14], [15], [16]. However, these estimators usually assume that the outliers of the sensory data are generated *independently* by some other probability distribution different from the model assumptions, which may not hold in an adversarial environment. In security settings, Mo and Sinopoli [17] propose a robust estimator design, which minimizes the mean squared estimation error under the worst possible attack scenarios.

For dynamical systems, robust techniques such as $\mathcal{H}_\infty, \mathcal{H}_2, \mathcal{L}_1$ estimation and control have also been an active research area for the past decades [18], [19]. Many researches in this field assume that the noise or the disturbance of the system lies in a normed space, e.g., l_2 or l_∞, while the output of the system also belongs to a normed space. Hence, the whole system can be viewed as a linear operator that maps the disturbance to the output. The goal of a robust design is thus to minimize the induced operator norm of the system in order to minimize the effect of the noise/disturbance on the system output. In security setting, we believe that it is more reasonable to model the bias injected by the adversary on the sensory data as a *sparse* input rather than a *bounded* input, since the adversary can change the compromised sensor readings arbitrarily. Therefore, the result presented in this paper can be seen as a generalization of the robust estimation framework to include both bounded and sparse disturbances.

The problem of dynamic state estimation in the presence of compromised sensory data has also been studied in [13] for noiseless system and extended to system subject to bounded noise and modeling errors in [20]. The main difference between the estimator design discussed in this paper and the ones proposed in [13], [20] is that our estimator leverages all the sensory data collected from time 0 to perform the state estimation. On the other hand, the estimation proposed in [13], [20] only uses a finite history of the sensor measurements. As a result, in this paper, we only requires the system to be *detectable* after remove any set of 2γ sensors in order to construct a resilient estimator, while in [13], [20], the requirement is restricted to the system

being *observable* after the removal of an arbitrary set of 2γ sensors.

The rest of the paper is organized as follows: Section II describes some preliminary result on the l_1 operator norm of linear systems. Section III formulates the dynamic state estimation problem with compromised sensory data. In Section IV, we prove a condition under which no resilient estimator exists. When the condition fails to hold, we propose a resilient estimator design and provide upper bound on its worst-case estimation error in Section V. The performance of the proposed estimator is further illustrated via a numerical example in Section VI. Finally, Section VII concludes the paper.

Notations

- Let $\mathbb{N}$ be the set of non-negative integers and $\mathbb{C}$ be the set of complex numbers. For any $x \in \mathbb{C}$, denote its real part as $\text{Re}(x)$ and its absolute value as $|x|$.
- For a discrete set $\mathcal{I}$, let $|\mathcal{I}|$ be the cardinality of the set.
- For a vector $x \in \mathbb{R}^n$, denote $\{x\}_i = x_i$ to be its i_{th} element. For a matrix $M \in \mathbb{R}^{n\times m}$, denote $\{M\}_i$ to be its i_{th} row.
- We denote restriction of an infinite sequence $\{x(t)\}_{t\in\mathbb{N}}$ to its first T elements with $x(0:T)$, i.e.,
$$x(0:T) \triangleq (x(0), \ldots, x(T)).$$
The infinity norm of the finite sequence $x(0:T)$ is defined as
$$\|x(0:T)\|_\infty \triangleq \max_{0\le t\le T} \max_i |x_i(t)|.$$
The infinity norm of an infinite sequence $x = x(0:\infty)$ is defined as
$$\|x\|_\infty \triangleq \sup_{t\in\mathbb{N}} \max_i |x_i(t)|.$$
We denote l_∞^n as the space of infinite sequences of n-dimensional vectors with bounded infinity norm. We will write l_∞ when there is no confusion on dimension of the vector.
- For any matrix $A \in \mathbb{R}^{m\times n}$. We denote its induced norm as
$$\|A\|_i = \sup_{x\neq 0} \frac{\|Ax\|_\infty}{\|x\|_\infty} = \max_i \sum_j |a_{ij}|.$$
- Given a infinite sequence $\{x(t)\}_{t\in\mathbb{N}}$, let $\text{supp}(x) \triangleq \{i : \exists t \text{ s.t. } x_i(t) \neq 0\}$ and we define $||x||_0 = |\text{supp}(x)|$.
- Let $*$ be the convolution operator, i.e., $y = K * u$ is defined as $y(t) = \sum_{\tau=0}^t K(\tau)u(t-\tau)$.

II. PRELIMINARY

Consider a linear time-invariant (LTI) system

$$x(t+1) = Ax(t) + Bw(t),\ x(0) = 0 \tag{1}$$
$$y(t) = Cx(t) + Dw(t),$$

with $x(t) \in \mathbb{R}^n$, $w(t) \in \mathbb{R}^p$ and $y(t) \in \mathbb{R}^m$. The matrices A, B, C, D are real matrices with proper dimensions. Define the following function $H : \mathbb{N} \to \mathbb{R}^{m\times p}$:

$$H(t) \triangleq \begin{cases} D & t = 0 \\ CA^{t-1}B & t \geq 1 \end{cases}. \tag{2}$$

Hence, $y = H * w$. With slight abuse of notation, denote

$$H \triangleq \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]. \tag{3}$$

If A is strictly stable, H is a bounded operator with its l_∞ induced norm being [19]:

$$\|H\|_1 \triangleq \sup_{||w||_\infty \neq 0} \frac{||y||_\infty}{||w||_\infty} = \max_{1\leq i\leq n} \sum_{j=1}^{p} \sum_{t=0}^{\infty} |h_{ij}(t)|.$$

Therefore, for any $||w||_\infty \leq \varepsilon$, we have

$$||y||_\infty \leq ||H||_1 \varepsilon. \tag{4}$$

On the other hand, if (A, C) is detectable, then we know there exists an $K \in \mathbb{R}^{n\times m}$, such that $A + KC$ is strictly stable. Now consider the following linear estimator:

$$\hat{x}(t+1) = A\hat{x}(t) - K(y(t) - C\hat{x}(t)),\ \hat{x}(0) = 0. \tag{5}$$

Define the corresponding estimation error and the residue vector as

$$e(t) \triangleq x(t) - \hat{x}(t),\ r(t) \triangleq y(t) - C\hat{x}(t). \tag{6}$$

The following lemma provides bounds on the $e(t)$ and $r(t)$:

Lemma 1. *For the estimator defined in* (5) *with $A + KC$ is strictly stable, the following inequalities hold:*

$$||e||_\infty \leq ||E(K)||_1 \varepsilon, \tag{7}$$
$$||r||_\infty \leq ||G(K)||_1 \varepsilon. \tag{8}$$

where

$$E(K) = \left[\begin{array}{c|c} A+KC & B+KD \\ \hline I & 0 \end{array}\right],$$
$$G(K) = \left[\begin{array}{c|c} A+KC & B+KD \\ \hline C & D \end{array}\right].$$

Proof. Manipulating (5), we have

$$\begin{aligned} e(t+1) &= (A+KC)e(t) + (B+KD)w(t),\ e(0) = 0, \\ r(t) &= Ce(t) + Dw(t). \end{aligned}$$

Therefore, (7) and (8) can be derived from (4). □

The following lemma characterizes the infinite norm of a finite sequence of the states:

Lemma 2. *Consider system (1) with a detectable pair of* (A, C) *and* $||w||_\infty \leq \varepsilon$*. If* $y(t) = 0$ *for all* $t = 0, 1, \cdots, T$*, then*

$$||x(0:T)||_\infty \leq \inf_{K: A+KC \text{ strictly stable}} ||E(K)||_1 \varepsilon \tag{9}$$

Proof. The assumption that (A, C) is detectable implies the existence of K such that $A+KC$ is strictly stable. For such a stabilizing K, we construct a state estimator from (5). The condition $y(0:T) = 0$ implies that $\hat{x}(0:T) = 0$. Therefore, by Lemma 1 we have

$$\begin{aligned} \|x(0:T)\|_\infty &= \|x(0:T) - \hat{x}(0:T)\|_\infty = \|e(0:T)\|_\infty \\ &\leq \|e\|_\infty \leq \|E(K)\|_1 \varepsilon. \end{aligned} \tag{10}$$

Since (10) holds for all stabilizing K, we can take the infimum over all such K and get (9).

□

III. Problem Formulation

We consider the state estimation problem for the following linear time invariant system:

$$\begin{aligned} x(t+1) &= Ax(t) + Bw(t),\ x(0) = 0, \\ y(t) &= Cx(t) + Dw(t) + a(t), \end{aligned} \tag{11}$$

where the state $x(t) \in \mathbb{R}^n$. $y(t) = [y_1(t), \ldots, y_m(t)]^T \in \mathbb{R}^m$ is the sensor measurements at time t from m sensors, where $y_i(t)$ is the measurement from sensor i. We denote the set of all sensors as $\mathcal{S} \triangleq \{1, \ldots, m\}$. $w(t) \in \mathbb{R}^p$ represents the process noise and measurement noise. We assume that the matrix $\left[\begin{smallmatrix} B \\ D \end{smallmatrix}\right]$ is full row rank, which implies that the noise is exciting all states and measurements. $a(t)$ is the bias injected by the adversary.

In this paper, we make the following assumptions:

A. The noise is l_∞ bounded: $\|w\|_\infty \leq \varepsilon$.

B. The adversary can change the readings from at most γ sensors. Therefore, the bias $a(t)$ satisfies $\|a\|_0 \leq \gamma$. Let us denote the set of compromised sensors as $\mathcal{C} \subset \mathcal{S}$. The set of "good" sensors is denoted as $\mathcal{G} \triangleq \mathcal{S} \backslash \mathcal{C}$. If sensor i is a good sensor, then $a_i(t) = 0$ for all t.

We further assume that the system operator knows both ε and γ. However, it does not know the exact set $\mathcal{C}$ of the compromised sensors.

Remark 1. *One can also interpret the parameter γ as a design parameter for the system operator, in the sense that the system operator wants to design an estimator that can tolerate at most γ compromised sensors. In general, increasing γ will increase the resilience of the detector under attack. However, a large γ may result in performance degradation during normal operation when no sensor is compromised.*

A causal state estimator can be defined an infinite sequence of mappings $f \triangleq (f_0, f_1, \ldots)$, where each f_t maps past measurements $y(0 : t-1)$ to an estimate of the current state $\hat{x}(t)$, i.e., $\hat{x}(t) = f_t(y(0 : t-1))$. We define $e(t)$ to be estimation error at time t, i.e.,

$$e(t) \triangleq x(t) - \hat{x}(t) = x(t) - f_t(y(0 : t-1)). \tag{12}$$

Clearly, the sequence e depends on the noise process w, the bias a injected by the adversary and the estimator f. As a result, we can write it as $e(w, a, f, t)$. However, we will simply write $e(t)$ when there is no confusion. In this paper, we consider designing the estimator against the worst w and a. To this end, let us define the worst-case estimation performance as

$$\rho(f) \triangleq \sup_{\|w\|_\infty \leq \varepsilon, \|a\|_0 \leq \gamma, t} \|e(w, a, f, t)\|_\infty. \tag{13}$$

Definition 1. *An estimator f is called an resilient estimator if $\rho(f) < \infty$.*

The diagram of the estimation problem discussed in this paper is illustrated in Fig 1. In the next section, we provide a condition, under which there exists no resilient estimator. Later in Section V, we prove that if such a condition fails to hold, then we can construct an resilient estimator. We further provide upper bound on $\rho(f)$.

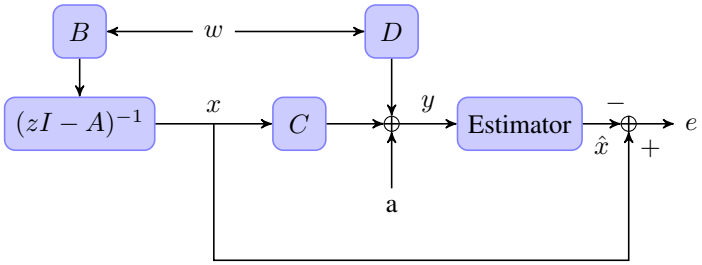

Fig. 1. Diagram of the Estimation Problem in Adversarial Environment. z^{-1} is the unit delay.

IV. Fundamental Limitation

In this section, we provide a condition under which there does not exist a resilient estimator. Before continuing on, we need the following lemma:

Lemma 3. *There does not exist a resilient estimator for system (11) if there exist infinite sequences x, x', w, w', a, a', y and y' of proper dimension, such that the follwoing conditions hold*

1) x, a, w, y *satisfy* (11)*, with* $\|w\| \leq \varepsilon$ *and* $\|a\|_0 \leq \gamma$.
2) x', a', w', y' *satisfy*

$$x'(t+1) = Ax'(t) + Bw'(t),\ x'(0) = 0,$$
$$y'(t) = Cx'(t) + Dw'(t) + a'(t),$$

   with $\|w'\| \leq \varepsilon$ *and* $\|a'\|_0 \leq \gamma$.
3) $y(t) = y'(t)$ *for all* t.
4) $\|x - x'\|_\infty = \infty$.

Proof. Let f be an arbitrary estimator. By the definition of $e(w, a, f, t)$, we have

$$e(w, a, f, t) = x(t) - f_t(y(0 : t-1)),$$
$$e(w', a', f, t) = x'(t) - f_t(y'(0 : t-1)).$$

Since $y(t) = y'(t)$ for all t, we know that $f_t(y(0 : t-1)) = f_t(y'(0 : t-1))$. Therefore,

$$e(w, a, f, t) - e(w', a', f, t) = x(t) - x'(t).$$

By triangular inequality,

$$\|e(w, a, f, t)\|_\infty + \|e(w', a', f, t)\|_\infty \geq \|x(t) - x'(t)\|_\infty.$$

Since $\|x - x'\|_\infty = \infty$, we know that at least one of the following equality holds:

$$\sup_t \|e(w, a, f, t)\|_\infty = \infty, \sup_t \|e(w', a', f, t)\|_\infty = \infty$$

Therefore, by the definition of $\rho(f)$, we know that $\rho(f) = \infty$ for all f, which implies the nonexistence of resilient estimator. □

To simplify notations, we introduce the following definition:

Definition 2. *For any index set $\mathcal{I} = \{i_1, \ldots, i_l\} \subseteq \mathcal{S}$, we define the projection matrix $P_\mathcal{I}$ to be*

$$P_\mathcal{I} = \begin{bmatrix} e_{i_1} & \ldots & e_{i_l} \end{bmatrix}^T \in \mathbb{R}^{l \times m},$$

where e_i is the ith canonical basis vector of $\mathbb{R}^m$. We further define the following vector $y_\mathcal{I}(t)$ by selecting the entries of $y(t)$ with indices in $\mathcal{I}$:

$$y_\mathcal{I}(t) \triangleq P_\mathcal{I} y(t).$$

Similarly, we define the following matrices:

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

We are now ready to state the main theorem on the nonexistence of the resilient estimator:

Theorem 1. *Consider system (11) with assumption A-B. There does not exist a resilient estimator if $(A, C_\mathcal{K})$ is not detectable for some set $\mathcal{K} \subset \mathcal{S}$ with cardinality $|\mathcal{K}| = m - 2\gamma$.*

Proof. Let $\mathcal{K}$ be a subset of $\mathcal{S}$ with cadinality $m-2\gamma$, such that $(A,\, C_{\mathcal{K}})$ is not detectable. We can find two index sets $\mathcal{K}_1,\, \mathcal{K}_2 \subset \mathcal{S}\backslash\mathcal{K}$, such that

$$|\mathcal{K}_1| = |\mathcal{K}_2| = \gamma,\, \mathcal{K} \cup \mathcal{K}_1 \cup \mathcal{K}_2 = \mathcal{S}.$$

Since (A, C) is not detectable, there exists an unstable and unobservable eigenvector $z \in \mathbb{C}^n$ and eigenvalue $\lambda \in \mathbb{C}$ of the matrix A, such that

$$Az = \lambda z,\, C_{\mathcal{K}} z = 0,\, |\lambda| \geq 1.$$

Since we assume that $\left[\begin{smallmatrix} B \\ D \end{smallmatrix}\right]$ is full row rank, we can find $w_* \in \mathbb{C}^n$, such that

$$\begin{bmatrix} B \\ D \end{bmatrix} w_* = \begin{bmatrix} z \\ 0 \end{bmatrix}.$$

Without loss of generality, we can scale z such that each entry of w_* has its absolute value to be no greater than ε, i.e., $|\{w_*\}_i| \leq \varepsilon$ for all i. Now let us consider the following noise process:

$$w(t) = \text{Re}\left(\frac{\lambda^t}{|\lambda|^t} w_0\right).$$

One can verify that

$$\|w\|_\infty \leq \sup_{t,i} \left| \frac{\lambda^t}{|\lambda|^t} \times \{w_*\}_i \right| \leq \varepsilon.$$

Since A is real, the corresponding sequence of the state x generated by w is given by

$$x(t) = \begin{cases} 0 & t = 0 \\ \text{Re}\left[\lambda^{t-1}\left(1 + \frac{1}{|\lambda|} + \cdots + \frac{1}{|\lambda|^{t-1}}\right) z\right] & t > 0 \end{cases}.$$

One can verify that $\|x\|_\infty = \infty$ for both $|\lambda| > 1$ and $|\lambda| = 1$. Now by the fact that C, D matrices are real, we have

$$C_{\mathcal{K}} x(t) + D_{\mathcal{K}} w(t) = 0.$$

Now let us construct $a(t) = \left[\begin{smallmatrix} a_1(t) & \cdots & a_m(t) \end{smallmatrix}\right]^T$, such that

$$a_i(t) = \begin{cases} 0 & i \in \mathcal{K} \cup \mathcal{K}_2 \\ -\{C\}_i x(t) & i \in \mathcal{K}_1 \end{cases}.$$

Clearly $\|a\|_0 \leq |\mathcal{K}_1| = \gamma$. The corresponding $y(t)$ satisfies

$$y_i(t) = \begin{cases} 0 & i \in \mathcal{K} \cup \mathcal{K}_1 \\ \{C\}_i x(t) & i \in \mathcal{K}_2 \end{cases}.$$

Now let us consider another sequences of noise $w' = 0$. Therefore, the corresponding $x' = 0$. Let us construct the bias a' injected by the adversary as

$$a'_i(t) = \begin{cases} 0 & i \in \mathcal{K} \cup \mathcal{K}_1 \\ \{C\}_i x(t) & i \in \mathcal{K}_2 \end{cases}.$$

One can verify that $\|a'\|_0 \leq |\mathcal{K}_2| = \gamma$. The corresponding $y'(t)$ satisfies

$$y'_i(t) = \begin{cases} 0 & i \in \mathcal{K} \cup \mathcal{K}_1 \\ \{C\}_i x(t) & i \in \mathcal{K}_2 \end{cases}.$$

Therefore, we have $y = y'$ and $\|x - x'\|_\infty = \|x\|_\infty = \infty$, which implies the nonexistence of any resilience estimator by Lemma 3. □

V. Estimator Design and Analysis

This section is devoted to the construction of a resilience estimator f and characterization of the corresponding worst-case performance $\rho(f)$. By Theorem 1, we know that the following assumption is necessary for the existence of the resillient estimator.

C. $(A, C_{\mathcal{K}})$ is detactable for any $\mathcal{K} \subset \mathcal{S}$ with cardinality $n - 2\gamma$

Therefore, we will assume that Assumption C holds throughout the section.

A. State Estimator Design

We propose the following estimator design for system (11) under the Assumption A-C.

1) Local Estimator and Detector: Let $\mathcal{I} = \{i_1, \cdots, i_{m-\gamma}\} \subset S$ be a index set with cardinality $n - \gamma$. Denote the collection of all such index sets as

$$\mathcal{L} \triangleq \{\mathcal{I} \subset \mathcal{S} : |\mathcal{I}| = m - \gamma\}.$$

For any $\mathcal{I} \in \mathcal{L}$, Assumption C implies the existence of $K^{\mathcal{I}}$ such that $A + K^{\mathcal{I}} C_{\mathcal{I}}$ is strictly stable. Therefore, we can construct a stable "local" estimator, which only uses the truncated measurement $y_{\mathcal{I}}(t)$ to compute the state estimate[1]:

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

with initial condition $\hat{x}^{\mathcal{I}}(0) = 0$.

For each local estimator, let us define the corresponding error and residue vector as

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

we define the corresponding linear operators as follows:

$$E^{\mathcal{I}}(K^{\mathcal{I}}) \triangleq \left[\begin{array}{c|c} A + K^{\mathcal{I}} C_{\mathcal{I}} & B + K^{\mathcal{I}} D_{\mathcal{I}} \\ \hline I & 0 \end{array}\right], \tag{16}$$

$$G^{\mathcal{I}}(K^{\mathcal{I}}) \triangleq \left[\begin{array}{c|c} A + K^{\mathcal{I}} C_{\mathcal{I}} & B + K^{\mathcal{I}} D_{\mathcal{I}} \\ \hline C_{\mathcal{I}} & D_{\mathcal{I}} \end{array}\right]. \tag{17}$$

By Lemma 1, we know that if $a_{\mathcal{I}}(t) = 0$ for all t, i.e., if $\mathcal{I}$ does not contain any compromised sensors, then the following inequality holds:

$$||r^{\mathcal{I}}||_\infty \leq ||G^{\mathcal{I}}(K^{\mathcal{I}})||_1 \varepsilon. \tag{18}$$

As a result, we will assign each local estimator a local detector, which checks if the following inequality holds at each time t:

$$||r^{\mathcal{I}}(0:t)||_\infty \leq ||G^{\mathcal{I}}(K^{\mathcal{I}})||_1 \varepsilon. \tag{19}$$

If (19) fails to hold, then we know the set $\mathcal{I}$ contains at least 1 compromised sensor and hence the local estimate $\hat{x}^{\mathcal{I}}(t)$ is corrupted by the adversary.

On the other hand, we call $\hat{x}^{\mathcal{I}}(t)$ a *valid* local estimate from time 0 to t if (19) holds at time t. We further define the set $\mathcal{L}(t)$ as

$$\mathcal{L}(t) \triangleq \{\mathcal{I} \in \mathcal{S} :\ (19) \text{ holds at time } t\}. \tag{20}$$

Remark 2. *Notice that* (19) *is only a sufficient condition for the index set $\mathcal{I}$ to contain compromised sensors and it is not necessarily tight. One can potentially design better local detectors*

[1]We use superscript notation for $x^{\mathcal{I}}$ and $K^{\mathcal{I}}$ in order to differentiate it from projection, which is written as subscript.

to check if there exist compromised sensors in the index set $\mathcal{I}$ to provide better performance. However, the local detector based on (19) *is suffice for us to design a resilient estimator.*

2) Global Data Fusion: We will then fuse all the *valid* local estimation $\hat{x}^{\mathcal{I}}(t)$ at time t to generate the state estimate $\hat{x}(t)$. Since we are concerned with the infinite norm of the estimation error, we will use the following equation to compute each entry of $\hat{x}(t)$:

$$\hat{x}_i(t) = \frac{1}{2}\Big(\min_{\mathcal{I}\in\mathcal{L}(t)} \hat{x}_i^{\mathcal{I}}(t) + \max_{\mathcal{I}\in\mathcal{L}(t)} \hat{x}_i^{\mathcal{I}}(t)\Big). \tag{21}$$

B. Upper Bound on the Worst-Case Estimation Error

We now provide an upper bound on the worst-case performance $\rho(f)$ for our estimator design, which is given by the following theorem:

Theorem 2. *Under Assumption A-C, the state estimator described in Section V-A is a resilient estimator for system* (11)*. Furthermore, the following inequality on $\rho(f)$ holds:*

$$\rho(f) \leq \max_{\mathcal{I},\mathcal{J}\in\mathcal{L}} \Big(\|E^{\mathcal{I}}(K^{\mathcal{I}})\|_1 + \frac{1}{2}\alpha^{\mathcal{I}\cap\mathcal{J}}[\beta^{\mathcal{I}}(K^{\mathcal{I}}) + \beta^{\mathcal{J}}(K^{\mathcal{J}})]\Big)\epsilon \tag{22}$$

where $\alpha^{\mathcal{K}}$ is defined as

$$\alpha^{\mathcal{K}} \triangleq \inf_{K: A+KC_{\mathcal{K}} \text{ strictly stable}} \left\| \left[\begin{array}{c|c} A+KC_{\mathcal{K}} & \begin{bmatrix} I & K \end{bmatrix} \\ \hline I & 0 \end{array}\right] \right\|_1,$$

and $\beta^{\mathcal{I}}(K^{\mathcal{I}})$ is defined as

$$\beta^{\mathcal{I}}(K^{\mathcal{I}}) \triangleq \max(\|K^{\mathcal{I}}\|_i, 1)\, \|G^{\mathcal{I}}(K^{\mathcal{I}})\|_1.$$

Remark 3. *It is worth noticing that if $\mathcal{I}$ does not contain compromised sensors, then minimizing the infinite norm of the local estimation error $e^{\mathcal{I}}(t)$ is equivalent to minimizing $\|E^{\mathcal{I}}(K^{\mathcal{I}})\|_1$. The second term on the RHS of* (22) *exists since the estimator does not know which local estimate can be trusted at the beginning.*

Several intermediate results are needed before proving Theorem 2. We first prove the following lemma to bound the divergence of the local estimates:

Lemma 4. *For any two index sets $\mathcal{I}$, $\mathcal{J} \in \mathcal{L}(T)$, the following inequality holds:*

$$\|\hat{x}^{\mathcal{I}}(0:T) - \hat{x}^{\mathcal{J}}(0:T)\|_\infty \leq \varepsilon\alpha^{\mathcal{I}\cap\mathcal{J}}[\beta^{\mathcal{I}}(K^{\mathcal{I}}) + \beta^{\mathcal{J}}(K^{\mathcal{J}})]. \tag{23}$$

Proof. By (14), we have

$$\hat{x}^{\mathcal{I}}(t+1) = A\hat{x}^{\mathcal{I}}(t) - K^{\mathcal{I}} r^{\mathcal{I}}(t),\ \hat{x}^{\mathcal{I}}(t) = 0,$$
$$y_{\mathcal{I}}(t) = C_{\mathcal{I}}\hat{x}^{\mathcal{I}}(t) + r^{\mathcal{I}}(t). \tag{24}$$

Let us define $\mathcal{K} = \mathcal{I} \cap \mathcal{J}$, we know that

$$y_{\mathcal{K}}(t) = C_{\mathcal{K}}\hat{x}^{\mathcal{I}}(t) + P_{\mathcal{K},\mathcal{I}} r^{\mathcal{I}}(t),$$

where $P_{\mathcal{K},\mathcal{I}} \in \mathbb{R}^{|\mathcal{K}|\times|\mathcal{I}|}$ is the unique matrix that satisfies:

$$P_{\mathcal{K}} = P_{\mathcal{K},\mathcal{I}} P_{\mathcal{I}}.$$

Now let us define $\phi^{\mathcal{I}}(t) \triangleq -K^{\mathcal{I}} r^{\mathcal{I}}(t)$ and $\varphi^{\mathcal{I}}(t) \triangleq P_{\mathcal{K},\mathcal{I}} r^{\mathcal{I}}(t)$. If $t \le T$, we know that (19) holds at time t, which implies that

$$\|r^{\mathcal{I}}(t)\|_\infty \le \|G^{\mathcal{I}}(K^{\mathcal{I}})\|_1 \varepsilon.$$

As a result,

$$\left\| \begin{bmatrix} \phi^{\mathcal{I}}(t) \\ \varphi^{\mathcal{I}}(t) \end{bmatrix} \right\|_\infty \le \left\| \begin{bmatrix} -K^{\mathcal{I}} \\ P_{\mathcal{K},\mathcal{I}} \end{bmatrix} \right\|_i \|r^{\mathcal{I}}(t)\|_\infty = \beta^{\mathcal{I}}(K^{\mathcal{I}})\varepsilon,$$

where we use the fact that each row of $P_{\mathcal{K},\mathcal{I}}$ is a canonical basis vector in $\mathbb{R}^{|\mathcal{I}|}$. Therefore, we have

$$\hat{x}^{\mathcal{I}}(t+1) = A\hat{x}^{\mathcal{I}}(t) + \begin{bmatrix} I & 0 \end{bmatrix} \begin{bmatrix} \phi^{\mathcal{I}}(t) \\ \varphi^{\mathcal{I}}(t) \end{bmatrix},\ \hat{x}^{\mathcal{I}}(t) = 0,$$
$$y_{\mathcal{K}}(t) = C_{\mathcal{K}}\hat{x}^{\mathcal{I}}(t) + \begin{bmatrix} 0 & I \end{bmatrix} \begin{bmatrix} \phi^{\mathcal{I}}(t) \\ \varphi^{\mathcal{I}}(t) \end{bmatrix}. \tag{25}$$

Similarly, we have

$$\hat{x}^{\mathcal{J}}(t+1) = A\hat{x}^{\mathcal{J}}(t) + \begin{bmatrix} I & 0 \end{bmatrix} \begin{bmatrix} \phi^{\mathcal{J}}(t) \\ \varphi^{\mathcal{J}}(t) \end{bmatrix},\ \hat{x}^{\mathcal{J}}(t) = 0,$$
$$y_{\mathcal{K}}(t) = C_{\mathcal{K}}\hat{x}^{\mathcal{J}}(t) + \begin{bmatrix} 0 & I \end{bmatrix} \begin{bmatrix} \phi^{\mathcal{J}}(t) \\ \varphi^{\mathcal{J}}(t) \end{bmatrix}, \tag{26}$$

with

$$\left\| \begin{bmatrix} \phi^{\mathcal{J}}(t) \\ \varphi^{\mathcal{J}}(t) \end{bmatrix} \right\| \leq \beta^{\mathcal{J}}(K^{\mathcal{J}})\varepsilon, \, \forall t \leq T.$$

Now let us consider $\Delta\hat{x}(t) = \hat{x}^I(t) - \hat{x}^j(t)$. By (25) and (26), we know that

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

$$0 = C_{\mathcal{K}}\Delta\hat{x}(t) + \begin{bmatrix} 0 & I \end{bmatrix} \begin{bmatrix} \phi^{\mathcal{I}}(t) - \phi^{\mathcal{J}}(t) \\ \varphi^{\mathcal{I}}(t) - \varphi^{\mathcal{J}}(t) \end{bmatrix}.$$

Hence, (23) can be proved by Lemma 2. □

Lemma 5. *Let $r_1, \ldots, r_l$ be real numbers. Define*

$$r = \frac{1}{2}\left(\max_i r_i + \min_i r_i\right).$$

Then for any i, we have

$$|r - r_i| \leq \frac{1}{2}\max_j |r_j - r_i|. \tag{27}$$

Proof. Without loss of generality, we assume that r_1 and r_2 are the largest and the smallest number among all r_is respectively. Therefore,

$$r - r_i = \frac{1}{2}(r_1 - r_i) - \frac{1}{2}(r_i - r_2).$$

Therefore, if $r_1 - r_i \geq r_i - r_2$, then

$$|r - r_i| = r - r_i \leq \frac{1}{2}(r_1 - r_i) = \frac{1}{2}\max_j |r_j - r_i|.$$

Similarly, one can prove that (27) holds when $r_1 - r_i < r_i - r_2$. □

Now we are ready to prove Theorem 2.

Proof. Let $\mathcal{G} \in \mathcal{L}$ be the set of good sensors. By Lemma 1, we know that

$$\|x - \hat{x}^{\mathcal{G}}\|_\infty \leq \|E^{\mathcal{G}}(K^{\mathcal{G}})\|_1\varepsilon.$$

Furthermore, $\mathcal{G} \in \mathcal{L}(t)$ for all t. At any given time t, assuming that the index set $\mathcal{J}$ also belongs to $\mathcal{L}(t)$. By Lemma 4, we have

$$\|\hat{x}^{\mathcal{G}}(t) - \hat{x}^{\mathcal{J}}(t)\|_\infty \leq \varepsilon\alpha^{\mathcal{G}\cap\mathcal{J}}[\beta^{\mathcal{G}}(K^{\mathcal{G}}) + \beta^{\mathcal{J}}(K^{\mathcal{J}})].$$

Therefore, by Lemma 5, we know that

$$\begin{aligned}\|\hat{x}(t)-\hat{x}^{\mathcal{G}}(t)\|_\infty &\leq \frac{1}{2}\max_{\mathcal{J}\in\mathcal{L}(t)} \varepsilon\alpha^{\mathcal{G}\cap\mathcal{J}}[\beta^{\mathcal{G}}(K^{\mathcal{G}})+\beta^{\mathcal{J}}(K^{\mathcal{J}})]\\ &\leq \frac{1}{2}\max_{\mathcal{J}\in\mathcal{L}} \varepsilon\alpha^{\mathcal{G}\cap\mathcal{J}}[\beta^{\mathcal{G}}(K^{\mathcal{G}})+\beta^{\mathcal{J}}(K^{\mathcal{J}})].\end{aligned}$$

By triangular inequality, we have

$$\begin{aligned}\|e(t)\|_\infty \leq \|E^{\mathcal{G}}(K^{\mathcal{G}})\|_1\varepsilon \\ +\frac{1}{2}\max_{\mathcal{J}\in\mathcal{L}}\alpha^{\mathcal{G}\cap\mathcal{J}}[\beta^{\mathcal{G}}(K^{\mathcal{G}})+\beta^{\mathcal{J}}(K^{\mathcal{J}})]\varepsilon.\end{aligned} \tag{28}$$

Thus, by taking the supremum over all possible "good" sensor set $\mathcal{G}$, we can prove (22). □

Combining Theorem 1 and Theorem 2, we have the following corollary:

Corollary 1. *A necessary and sufficient condition for the existence of a resilient estimator is that $(A, C_{\mathcal{K}})$ is detectable for any index set $\mathcal{K}\subset\mathcal{S}$ with cardinality 2γ.*

VI. Numerical Example

In this section, we provide a numerical example to illustrate our resilient estimator design. We choose the following parameters for the system:

$$A=1,\ C=\begin{bmatrix}1\\1\\1\end{bmatrix},\ B=\begin{bmatrix}1 & 0 & 0 & 0\end{bmatrix},\ D=\begin{bmatrix}0 & I\end{bmatrix}.$$

We further assume that $\varepsilon = 1$ and $\gamma = 1$. First, we consider designing a linear estimator described in (5) in non-adversarial settings. Due to symmetry, we assume that the estimation gain $K=\begin{bmatrix}\theta & \theta & \theta\end{bmatrix}$, where $\theta\in\mathbb{R}$. As a result, one can check that the l_1 norm of $E(K)$ satisfies:

$$\|E(K)\|_1 = \frac{1+|3\theta|}{1-|1+3\theta|},$$

where $|1+3\theta|<1$ in order to ensure the stability of the estimator. The optimal θ, which minimizes $\|E(K)\|_1$, is given by $\theta=-1/3$. The optimal $\|E(K)\|_1$ equals 2. Therefore, the estimator for the non-adversarial environment can be written as:

$$\hat{x}(t+1)=\hat{x}(t)+\frac{1}{3}\begin{bmatrix}1 & 1 & 1\end{bmatrix}(y(t)-C\hat{x}(t)),\ \hat{x}(0)=0. \tag{29}$$

We now consider our resilient estimator design in adversarial environment. To this end, we need to compute the gain $K^{\{1,2\}}$, $K^{\{2,3\}}$, $K^{\{3,1\}}$. Due to symmetry, we will only consider the gain of the following form:

$$K^{\{1,2\}} = K^{\{2,3\}} = K^{\{3,1\}} = \begin{bmatrix} \mu & \mu \end{bmatrix}.$$

One can check that $\alpha^{\mathcal{K}} = 2$ for $\mathcal{K} = \{1\}, \{2\}, \{3\}$ and

$$\|E^{\mathcal{I}}(K^{\mathcal{I}})\|_1 = \frac{1+|2\mu|}{1-|1+2\mu|}, \|G^{\mathcal{I}}(K^{\mathcal{I}})\| = 1 + \frac{1+|2\mu|}{1-|1+2\mu|}.$$

where $\mu \in (-1, 0)$ to ensure the stability of the local estimator. As a result,

$$\beta^{\mathcal{I}}(K^{\mathcal{I}}) = 1 + \frac{1+|2\mu|}{1-|1+2\mu|}.$$

Hence, the upper bound on the worst-case estimation error is

$$\rho(f) \leq 2 + 3 \times \frac{1+|2\mu|}{1-|1+2\mu|}. \tag{30}$$

The optimal μ which minimizes the RHS of (30), is $\mu = -0.5$ and the corresponding upper bound is 8.

We now compare the performance of the estimator (29) and the resilient estimator. To this end, we randomly generate $w(k)$ from a uniform distribution on the set $\|w(k)\|_\infty \leq 1$. We assume that the adversary compromise the first sensor and add an increasing bias $a(t) = \begin{bmatrix} 0.5t & 0 & 0 \end{bmatrix}^T$. The trajectory of the estimation error of the estimator (29) and the resilient estimator is plotted in Fig 2 and Fig 3 respectively.

One can see that the error for the estimator (29) grows linearly and becomes unbounded. On the other hand, our resilient estimator will detect that the index sets $\{1, 2\}$ and $\{1, 3\}$ contain the compromised sensor and hence discard the corresponding local estimates. As a result, the estimation error remains bounded.

VII. Conclusion

In this paper, we consider the problem of estimating the state of a linear time invariant system, which is driven by l_∞ noise, in the presence of compromised sensory data. We prove a necessary and sufficient condition for the existence of a resilient estimator. When such a condition holds, we propose a resilient estimator design and provide an upper bound on its worst-case estimation error. Future works include extending the noise model from bounded noise to stochastic noise and investigating the computational aspect of the resilient estimator design.

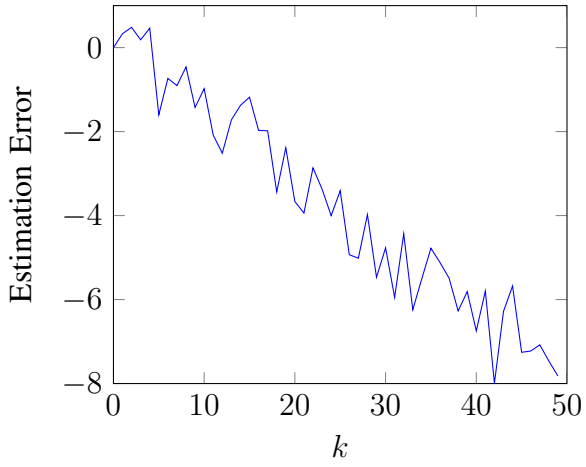

Fig. 2. The estimation error of the estimator (29).

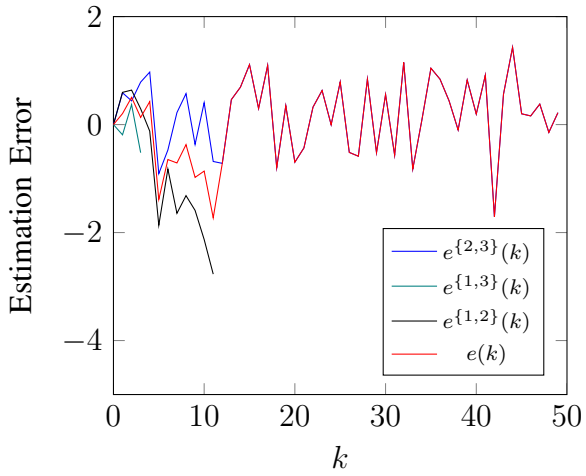

Fig. 3. The estimation error of the resilient estimator. The blue, teal and black line correspond the estimation errors for 3 local estimators, while the red line is the estimation error after fusion. The teal and black line terminate at time 4 and time 12 respectively when the corresponding local detector detects a violation of (19).

REFERENCES

[1] T. M. Chen. Stuxnet, the real start of cyber warfare? [editor's note]. *IEEE Network*, 24(6):2–3, 2010.

[2] D. P. Fidler. Was stuxnet an act of war? decoding a cyberattack. *IEEE Security & Privacy*, 9(4):56–59, 2011.

[3] Alvaro A. Cárdenas, Saurabh Amin, and Shankar Sastry. Research challenges for the security of control systems. In *HOTSEC'08: Proceedings of the 3rd conference on Hot topics in security*, pages 1–6, Berkeley, CA, USA, 2008. USENIX Association.

[4] Ali Abur and Antonio Gómez Expósito. *Power System State Estimation: Theory and Implementation*. CRC Press, 2004.

[5] Yao Liu, Michael Reiter, and Peng 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.

[6] Henrik Sandberg, Andre Teixeira, and Karl H. Johansson. On security indices for state estimators in power networks. In *First Workshop on Secure Control Systems*, 2010.

[7] Le Xie, Yilin Mo, and B. Sinopoli. Integrity data attacks in power market operations. *IEEE Transactions on Smart Grid*, 2(4):659–666, 2011.

[8] A.S. Willsky. A survey of design methods for failure detection in dynamic systems. *Automatica*, 12:601–611, Nov 1976.

[9] M. A. Massoumnia, G.C. Verghese, and A.S. Willsky. Failure detection and identification. *Automatic Control, IEEE Transactions on*, 34(3):316–321, Mar 1989.

[10] F. Pasqualetti, A. Bicchi, and F. Bullo. Consensus computation in unreliable networks: A system theoretic approach. *IEEE Transactions on Automatic Control*, 57(1):90–104, Jan 2010.

[11] F. Pasqualetti, F. Dorfler, and F. Bullo. Attack detection and identification in cyber-physical systems. *IEEE Transactions on Automatic Control*, 58(11):2715–2729, 2013.

[12] S. Sundaram, M. Pajic, C. Hadjicostis, R. Mangharam, and G. J. Pappas. The wireless control network: monitoring for malicious behavior. In *IEEE Conference on Decision and Contro*, Atlanta, GA, Dec 2010.

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

[14] S. A. Kassam and H. V. Poor. Robust techniques for signal processing: A survey. *Proceedings of the IEEE*, 73(3):433–481, 1985.

[15] Ricardo A. Maronna, Douglas R. Martin, and Victor J. Yohai. *Robust Statistics: Theory and Methods*. Wiley, 2006.

[16] Peter J. Huber and Elvezio M. Ronchetti. *Robust Statistics*. Wiely, 2009.

[17] Yilin Mo and B. Sinopoli. Robust estimation in the presence of integrity attacks. In *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*, pages 6085–6090, Dec 2013.

[18] Kemin Zhou, John Comstock Doyle, and Keith Glover. *Robust and optimal control*, volume 40. Prentice hall New Jersey, 1996.

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

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