Attack-Resilient State Estimation for Noisy Dynamical Systems

Miroslav Pajic, *Member, IEEE*, Insup Lee, *Fellow, IEEE*, and George J. Pappas, *Fellow, IEEE*

***Abstract*—Several recent incidents have clearly illustrated the susceptibility of cyberphysical systems (CPS) to attacks, raising attention to security challenges in these systems. The tight interaction between information technology and the physical world has introduced new vulnerabilities that cannot be addressed with the use of standard cryptographic security techniques. Accordingly, the problem of state estimation in the presence of sensor and actuator attacks has attracted significant attention in the past. Unlike the existing work, in this paper, we consider the problem of attack-resilient state estimation in the presence of bounded-size noise. We focus on the most general model for sensor attacks where *any* signal can be injected via compromised sensors. Specifically, we present an l_0-based state estimator that can be formulated as a mixed-integer linear program and its convex relaxation based on the l_1 norm. For both attack-resilient state estimators, we derive rigorous analytic bounds on the state-estimation errors caused by the presence of noise. Our analysis shows that the worst-case error is linear with the size of the noise and, thus, the attacker cannot exploit the noise to introduce unbounded state-estimation errors. Finally, we show how the l_0 and l_1-based attack-resilient state estimators can be used for sound attack detection and identification; we provide conditions on the size of attack vectors that ensure correct identification of compromised sensors.**

***Index Terms*—Attack-resilient state estimation, robustness of state estimators, cyberphysical systems security, linear systems.**

I. Introduction

MOST EXISTING control systems have not been built with security in mind. Even with the proliferation of different networking technologies and the use of more open control architectures, until recently, security of control systems has usually been an afterthought. Yet, with the advance of cyberphysical systems (CPS), the tight interaction between information technology and the physical world has made control components of CPS vulnerable to attack vectors well beyond the standard cyberattacks [1]. In the last few years, several incidents have clearly illustrated susceptibility of CPS to attacks and raised attention to security challenges in these systems. These include attacks on large-scale systems, such as the Maroochy Water breach [2] and the StuxNet virus attack on an industrial supervisory control and data acquisition (SCADA) system [3], [4]. In addition, attacks on modern vehicles [5]–[7] and the RQ-170 Sentinel U.S. drone that was captured in Iran [8], [9]show that even widely used, safety-critical automotive and avionics CPS can be compromised by malicious attackers.

A typical CPS contains an internal network, or multiple networks, connected by a gateway to external communications network. In many cases, such as in the automotive industry, systems rely on perimeter security where internal networks are resource constrained, mostly depending on the security of the gateway and external communication channels. However, the gateway may be compromised, becoming a threat to the system's operation [5], [6]. In addition, some of the internal components might be tampered with, allowing the attacker to access the internal communication network [7]. From the control of the CPS perspective, attacks on the internal network, where the attacker inserts messages anywhere in the sensors-to-controllers-to-actuators pathway, can be modeled as additional signals injected into the control loop via system sensors and actuators [10]. While some of these attacks can be avoided with the use of standard cryptographic tools that guarantee data integrity and authentication, this incurs a significant design, computational, and operational overhead for these usually resource-constrained systems.

On the other hand, unlike standard cybersystems, reported CPS vulnerabilities include noninvasive attacks on system sensors, where an adversarial signal is injected into the measured data by modifying a sensor's physical environment. This has been illustrated with several noninvasive attacks on global positioning system (GPS)-based navigation systems [9], [11], [12], and antilock braking systems [13] in vehicles. These attacks show that the use of standard encryption and data-authentication-based network security techniques does not guarantee secure control of CPS. In such cases, even if the stream of sensor data is properly encrypted, it would still contain incorrect values. Consequently, there is a need to focus on attack-resilient control of CPS, to ensure safety in such scenarios.

Manuscript received October 15, 2015; revised October 16, 2015 and June 25, 2016; accepted August 22, 2016. Date of publication September 8, 2016; date of current version March 16, 2017. This work was supported in part by the National Science Foundation under Grants CNS-1505701 and CNS-1505799, and in part by the Intel-NSF Partnership for Cyber-Physical Systems Security and Privacy. This material is also based on research sponsored by DARPA under agreement number FA8750-12-2-0247. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA or the U.S. Government. Recommended by Associate Editor Bruno Sinopoli.

M. Pajic is with the Department of Electrical and Computer Engineering, Duke University, Durham, NC, 27708 USA (e-mail: miroslav.pajic@duke.edu).

I. Lee is with the Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA, USA 19014 (e-mail: lee@cis.upenn.edu).

G. J. Pappas is with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA, USA 19014 (e-mail: pappasg@seas.upenn.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCNS.2016.2607420

A. Related Work

In recent years, significant efforts have been invested into the development of control techniques that exploit some knowledge of system dynamics for attack detection and attack-resilient control (e.g., [10], [14]–[20]). One line of work has focused on attack detection, [19]–[22] based on the use of standard residual probability-based detectors (e.g., the chi-square detector). For example, in [19], the authors illustrate how these detectors can be used to detect integrity attacks on SCADA systems, while in [20], the authors focus on the design of watermarked control inputs for active attack detection.

In addition to attack detection, the problem of state estimation in the presence of sensor and actuator attacks has attracted significant attention due to the fact that the CPS is capable of correctly estimating the plant's state from corrupted measurements and would be able to continue operating even under attack. For noiseless linear time-invariant (LTI) systems for which the exact plant model is known, the attack-resilient state estimation problem has been formulated as an l_0 optimization problem [15], [16]. In addition, in [23], the authors present an SMT-based state estimation technique.

However, it is unclear how robust these state estimators are to noise and modeling errors; specifically, what kind of guarantees can be provided for performance of attack-resilient state estimators for noisy dynamical systems. To the best of our knowledge, the first work on this topic was [24]. In that paper, we introduced an l_0-based attack-resilient state estimator for systems with bounded noise that can be formulated as a mixed-integer linear program (MILP). We also showed its robustness to noise and modeling errors, and provided a complex design-time procedure to bound the worst-case state estimation error in the presence of attacks.

It is worth noting that our work exploits some of the ideas initially introduced in the domain of compressed sensing (e.g., see [25] and the references within), starting from the problem considered in [26] and [27] where a sparse state was to be extracted for noisy nondynamical systems with a predefined measurement matrix and without any structured interference. These works were extended to the problem of the extraction of block-sparse signals for these systems in the presence of noise (e.g., [28]). On the other hand, error bounds for the estimation of (nonblock) sparse signals in the presence of structured interference for noisy nondynamical systems have been recently addressed in [29] and [30]. Specifically, [29] considers systems with Gaussian measurement matrices, while [30] provides a very conservative error bound due to the fact that the authors assume that state and interference are sparse.

B. Contributions of This Work

In this paper, we focus on the problem of attack-resilient state estimation for linear dynamical systems with bounded-size noise. We consider the most general model for sensor attacks where *any* signals can be injected via the compromised sensors [10]. We start from the l_0-based state estimation procedure introduced in [16], and show how it can be adapted for systems with noise. The main limitation of the l_0-based state estimators is that solving them is NP hard in general. Therefore, by exploiting properties of the l_1 norm, we provide a computationally efficient, convex optimization-based state estimation procedure for noisy dynamical systems.

Furthermore, unlike our work in [24] for the l_0 estimator, we derive rigorous analytic bounds on the state-estimation errors for l_0 and l_1-based state estimation procedures. We show that *the worst-case error is linear with the size of the noise*, and when the number of attacked sensors is not higher than a predefined number (that depends on the properties of the system's observability matrix), the attacker cannot exploit noise and modeling errors to introduce unbounded state estimation errors. In addition, we introduce a method that utilizes the presented attack-resilient state estimators for sound attack detection and identification, using the estimates of attack vectors provided by the estimators.

Preliminary versions of some of the results from the paper have been presented in [24] and [31]. This paper is significantly expanded from the conference papers, providing full proofs of all theorems from [31] as well as showing the boundedness of the state estimation error when the l_1-based estimation procedure is used. In addition, we provide a comprehensive view of the unique challenges for secure control of CPS, and unlike the conference paper, present a thorough evaluation of the state-estimation bounds introduced in this paper.

The rest of this paper is organized as follows. Section II introduces the problem formulation, while in Section III, we present attack-resilient state estimation procedures based on the l_0 and l_1 norms. In Sections IV and V, we present robustness analysis of the l_0 and l_1-based state estimators, respectfully. Finally, in Section VI, we show how the presented state estimators can be used for sound attack detection and identification, followed by evaluation of the introduced robustness bounds (Section VII) and some concluding remarks in Section VIII.

C. Notation and Terminology

For a set $\mathcal{S}$, $|\mathcal{S}|$ denotes the cardinality (i.e., size) of the set. In addition, for a set $\mathcal{K} \subset \mathcal{S}$, with $\mathcal{K}^{\complement}$ we denote the complement set of $\mathcal{K}$ with respect to $\mathcal{S}$, that is $\mathcal{K}^{\complement} = \mathcal{S} \setminus \mathcal{K}$.

We use $\mathbf{A}^T$ to indicate the transpose of matrix $\mathbf{A}$, while the ith element of a vector $\mathbf{x}_k$ is denoted by $\mathbf{x}_{k,i}$. For vector $\mathbf{x}$ and matrix $\mathbf{A}$, we denote by $|\mathbf{x}|$ and $|\mathbf{A}|$ the vector and matrix whose elements are absolute values of the initial vector and matrix, respectively. Also, for matrices $\mathbf{P}$ and $\mathbf{Q}$, by $\mathbf{P} \leq \mathbf{Q}$ we specify that the matrix $\mathbf{P}$ is *element-wise* smaller than the matrix $\mathbf{Q}$. In addition, for a symmetric matrix $\mathbf{Q}$, $\mathbf{Q} \succeq 0$ denotes that the matrix is positive semidefinite.

We use $\mathbb{R}$ to denote the set of reals. In addition, $\mathbf{I}_p$ denotes the identity matrix of size p, while $\mathbb{I}(\cdot)$ denotes the indicator function. Finally, for a vector $\mathbf{e} \in \mathbb{R}^p$, *the support* of the vector is the set

$$\mathrm{supp}(\mathbf{e}) = \{i \mid \mathbf{e}_i \neq 0\} \subseteq \{1, 2, \ldots, p\},$$

while l_0 norm of vector $\mathbf{e}$ is the cardinality of supp($\mathbf{e}$), that is $\|\mathbf{e}\|_{l_0} = |\mathrm{supp}(\mathbf{e})|$.[1]

[1] Although the l_0-norm is not formally a norm, in this paper, we will abuse the terminology and refer to it as a norm in order to use consistent terminology with the one used in previous work on this topic (e.g., [16]).

II. Problem Description

We consider LTI systems of the form

$$\begin{aligned}\mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k\\ \mathbf{y}_k &= \mathbf{C}\mathbf{x}_k + \mathbf{w}_k + \mathbf{e}_k.\end{aligned} \tag{1}$$

The plant's output vector $\mathbf{y} \in \mathbb{R}^p$ contains measurements of the plant's state $\mathbf{x} \in \mathbb{R}^n$ provided by p sensors from the set $\mathcal{S} = \{s_1, s_2, \ldots, s_p\}$. We assume the measurement noise vector $\mathbf{w} \in \mathbb{R}^p$ to be bounded; specifically, we assume that $|\mathbf{w}_k| \leq \boldsymbol{\delta}_{w_k}$, for all $k \geq 0$. Finally, the sparse vector $\mathbf{e} \in \mathbb{R}^p$ with support in the set $\mathcal{K} \subseteq \mathcal{S}$ denotes the attack vector injected by a malicious attacker using sensors from the set $\mathcal{K}$.[2]

The attack-resilient state estimation problem focuses on reconstruction of the initial system state $\mathbf{x}_0$ from a set of N output observations[3] $\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_{N-1}$. These observations are potentially corrupted by an attacker with access to the sensors from the set $\mathcal{K}$, that is

$$\mathbf{y}_k = \mathbf{C}\mathbf{A}^k\mathbf{x}_0 + \mathbf{e}_k + \mathbf{w}_k.$$

One additional goal is to provide identification of the compromised sensors (i.e., identify sensors from $\mathcal{K}$), since the actual set $\mathcal{K}$ of compromised sensors is not known before the estimation.

A. Model Motivation

The aforementioned attack-resilient state estimation problem can be also used for the general form of LTI systems

$$\begin{aligned}\mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k + \mathbf{v^p}_k\\ \mathbf{y}_k &= \mathbf{C}\mathbf{x}_k + \mathbf{v^m}_k + \mathbf{e}_k.\end{aligned} \tag{2}$$

Here, $\mathbf{A} \in \mathbb{R}^{n\times n}, \mathbf{B} \in \mathbb{R}^{n\times m}$, and $\mathbf{C} \in \mathbb{R}^{p\times n}$, while process and measurement noise $\mathbf{v^p} \in \mathbb{R}^n$ and $\mathbf{v^m} \in \mathbb{R}^p$, respectively, are bounded in size. In this general case, to obtain the plant's state at any time step t (i.e., $\mathbf{x}_t$), the goal is to utilize the previous N sensor measurement vectors $(\mathbf{y}_{t-N+1}, \ldots, \mathbf{y}_t)$ and actuator inputs $(\mathbf{u}_{t-N+1}, \ldots, \mathbf{u}_{t-1})$ to evaluate the state $\mathbf{x}_{t-N+1}$.

For dynamical systems without noise, the state can be obtained as the minimization argument of the following optimization problem [16], [24]

$$\begin{aligned}\min_{\mathbf{E}_{t,N} \in \mathbb{R}^{p\times N},\ \mathbf{x}\in\mathbb{R}^n} \quad & \|\mathbf{E}_{t,N}\|_{l_0}\\ s.\,t. \quad & \mathbf{E}_{t,N} = \mathbf{Y}_{t,N} - \Phi_N(\mathbf{x}).\end{aligned} \tag{3}$$

Here, the matrix $\mathbf{E}_{t,N} = [\mathbf{e}_{t-N+1}|\mathbf{e}_{t-N+2}|\cdots|\mathbf{e}_t]$ captures the last N attacks vectors. In addition, $\mathbf{Y}_{t,N} = [\tilde{\mathbf{y}}_{t-N+1}|\tilde{\mathbf{y}}_{t-N+2}|\cdots|\tilde{\mathbf{y}}_t]$ maintains the last N sensor measurements compensated for the impact of the inputs applied during that interval, that is

$$\begin{aligned}\tilde{\mathbf{y}}_k &= \mathbf{y}_k, \qquad\qquad\qquad\qquad k = t-N+1\\ \tilde{\mathbf{y}}_k &= \mathbf{y}_k - \sum_{i=0}^{k-t+N-2} \mathbf{C}\mathbf{A}^i\mathbf{B}\mathbf{u}_{k-1-i},\ \tilde{}k = t-N+2,\ldots,N.\end{aligned}$$

Finally, linear mapping $\Phi_N : \mathbb{R}^n \rightarrow \mathbb{R}^{p\times N}$ defined as $\Phi_N(\mathbf{x}) = \left[\mathbf{C}\mathbf{x}|\mathbf{C}\mathbf{A}\mathbf{x}|\ldots|\mathbf{C}\mathbf{A}^{N-1}\mathbf{x}\right]$ specifies the observed system evolution, due to its dynamics, from initial state $\mathbf{x}$.

Consequently, even for the general form of LTI systems as in (2), the problem of state estimation can be mapped into the state estimation for systems from (1), where control inputs are discarded. Furthermore, as shown in [24], the bounds on the size of measurement noise in (1) can be related to the bounds on the size of process and measurement noise vectors $\mathbf{v^p}$ and $\mathbf{v^m}$. It is worth noting, however, that these bounds on measurement noise in problem formulation (1), which are caused by the process noise from (2), can be very conservative in the case where the system (i.e., matrix $\mathbf{A}$) is unstable. In general, the transformation of the process and measurement noise bounds from the model (2) to only measurement noise in (1) is conservative since it has to capture the worst-case system (including noise) behavior.

III. Attack-Resilient State Estimators

We start by introducing the following notation. We use $P_\mathcal{K}$ to denote the projection from the set $\mathcal{S}$ to set $\mathcal{K}$ by keeping only rows of $\mathbf{C}$ with indices that correspond to sensors from $\mathcal{K}$. Formally, $P_\mathcal{K} = \left[\mathbf{i}_{k_1}|\cdots|\mathbf{i}_{k_{|\mathcal{K}|}}\right]^T$, where $\mathcal{K} = \{s_{k_1}, \ldots, s_{k_{|\mathcal{K}|}}\} \subseteq S$ and $k_1 < k_2 < \cdots < k_{|\mathcal{K}|}$, and $\mathbf{i}_j^T$ denotes the row vector (of the appropriate size) with a 1 in its jth position being the only nonzero element of the vector. Furthermore, for any sensor s_i and set $\mathcal{K}$ we define the matrices $\mathbf{O}_{s_i}$ and $\mathbf{O}_\mathcal{K}$ as

$$\mathbf{O}_{s_i} = \begin{bmatrix} P_{\{s_i\}}\mathbf{C}\\ P_{\{s_i\}}\mathbf{C}\mathbf{A}\\ \vdots\\ P_{\{s_i\}}\mathbf{C}\mathbf{A}^{N-1}\end{bmatrix} \quad \mathbf{O}_\mathcal{K} = \begin{bmatrix}\mathbf{O}_{s_{i_1}}\\ \mathbf{O}_{s_{i_2}}\\ \vdots\\ \mathbf{O}_{s_{i_{|\mathcal{K}|}}}\end{bmatrix}. \tag{4}$$

We will also slightly abuse the notation by using $\mathbf{O}_i$ to denote $\mathbf{O}_{s_i}$ for each sensor s_i.

In addition, we use $\tilde{\mathbf{e}}_i = \left[\mathbf{e}_{0,i}|\mathbf{e}_{1,i}|\cdots|\mathbf{e}_{N-1,i}\right]^T \in \mathbb{R}^N$ for all $i \in \{1, \ldots, p\}$ to denote the values injected via sensor s_i (i.e., attack signals on sensor s_i) at time steps $0, \ldots, N-1$.[4] From the definition, if $s_i \notin \mathcal{K}$, then $\tilde{\mathbf{e}}_i = \mathbf{0} \in \mathbb{R}^N$. Similarly, for all $i \in \{1, \ldots, p\}$, we use $\tilde{\mathbf{y}}_i = \left[\mathbf{y}_{0,i}|\mathbf{y}_{1,i}|\cdots|\mathbf{y}_{N-1,i}\right]^T \in \mathbb{R}^N$ and $\tilde{\mathbf{w}}_i = \left[\mathbf{w}_{0,i}|\mathbf{w}_{1,i}|\cdots|\mathbf{w}_{N-1,i}\right]^T \in \mathbb{R}^N$ to denote all measurements obtained by the sensor s_i and measurement noise at the sensor, respectively, at time steps $0, \ldots, N-1$. Hence, we have that for all $1 \leq i \leq p$:

$$\tilde{\mathbf{y}}_i = \mathbf{O}_i\mathbf{x}_0 + \tilde{\mathbf{e}}_i + \tilde{\mathbf{w}}_i \tag{5}$$

[2] In this work, we sometimes abuse the notation by using $\mathcal{K}$ to denote the set of compromised sensors and the set of indices of the compromised sensors.

[3] We consider the measurement history size N as an input parameter to the state-estimation procedure

[4] It is worth nothing that the vector $\tilde{\mathbf{e}}_i$ corresponds to the ith row of the matrix $\mathbf{E}$ from (3).

Finally, we define block vectors $\tilde{\mathbf{y}}, \tilde{\mathbf{e}}, \tilde{\mathbf{w}} \in \mathbb{R}^{pN}$ as $\tilde{\mathbf{y}} = \left[\tilde{\mathbf{y}}_1^T | \cdots | \tilde{\mathbf{y}}_p^T\right]^T$, $\tilde{\mathbf{e}} = \left[\tilde{\mathbf{e}}_1^T | \cdots | \tilde{\mathbf{e}}_p^T\right]^T$, and $\tilde{\mathbf{w}} = \left[\tilde{\mathbf{w}}_1^T | \cdots | \tilde{\mathbf{w}}_p^T\right]^T$, and matrix $\mathbf{O} = \left[\mathbf{O}_1^T | \cdots | \mathbf{O}_p^T\right]^T$.[5] Since each element of the measurement noise vectors $\mathbf{w}_0, \ldots, \mathbf{w}_{N-1}$ is bounded (i.e., $|\mathbf{w}_{k,i}| \leq \boldsymbol{\delta}_{w_{k,i}}, 0 \leq k \leq N-1, 1 \leq i \leq p$), we denote by $\Omega \subset \mathbb{R}^{pN}$ the feasible set of noise vectors $\tilde{\mathbf{w}}$. In addition, for any set $\mathcal{R} \subset \mathcal{S}$, we define $\tilde{\mathbf{w}}_{\mathcal{R}}$ to be the block vector obtained by concatenating $\tilde{\mathbf{w}}_{s_i}$ for all $s_i \in \mathcal{R}$ starting from the smallest i to the largest, while the corresponding $\Omega_{\mathcal{R}} \subset \mathbb{R}^{|\mathcal{R}|N}$ denotes the feasible set of vectors $\tilde{\mathbf{w}}_{\mathcal{R}}$. We similarly define the matrix $\mathbf{O}_{\mathcal{R}}$ to be obtained by concatenating matrices $\mathbf{O}_i$ for all $s_i \in \mathcal{R}$.

Now, from (5), it follows that:

$$\tilde{\mathbf{y}} = \mathbf{O}\mathbf{x}_0 + \tilde{\mathbf{e}} + \tilde{\mathbf{w}}. \tag{6}$$

For block vectors obtained by concatenating p vectors, such as $\tilde{\mathbf{e}}$ and $\tilde{\mathbf{y}}$, we also use the notation from [28]

$$\begin{aligned} \|\tilde{\mathbf{e}}\|_{l_2,l_0} &= \sum_{i=1}^{p} \mathbb{I}(\|\tilde{\mathbf{e}}_i\|_{l_2} > 0) \\ \|\tilde{\mathbf{e}}\|_{l_2,l_1} &= \sum_{i=1}^{p} \|\tilde{\mathbf{e}}_i\|_{l_2} \end{aligned} \tag{7}$$

Note that for any block vector $\tilde{\mathbf{e}}$ it holds that $\|\tilde{\mathbf{e}}\|_{l_2,l_0} = \|\tilde{\mathbf{e}}\|_{l_t,l_0}$ for any $t \geq 1$. This allows us to define block q-sparse vector $\tilde{\mathbf{e}}$ as a vector that satisfies $\|\tilde{\mathbf{e}}\|_{l_2,l_0} = q$, meaning that it has q nonzero subvectors. Hence, if the set of compromised sensors $\mathcal{K}$ has q elements (i.e., $|\mathcal{K}| = q$), then vector $\tilde{\mathbf{e}}$ is q-block sparse.

Using the above notation, the optimization problem (3) can be represented as

$$\begin{aligned} P_0: \quad & \min_{\tilde{\mathbf{e}},\mathbf{x}} \|\tilde{\mathbf{e}}\|_{l_2,l_0} \\ s.\ t. \quad & \tilde{\mathbf{y}} - \mathbf{O}\mathbf{x}_0 - \tilde{\mathbf{e}} = \mathbf{0}. \end{aligned} \tag{8}$$

Now, consider the measurement vector $\tilde{\mathbf{y}}$ for a noiseless system's (i.e., when $\Omega = \mathbf{0} \in \mathbb{R}^{pN}$) evolution due to the initial state $\mathbf{x}_0$ and attack vector $\tilde{\mathbf{e}}^*$. If the number of attacked sensors $q = |\mathcal{K}|$ is not higher than a certain number $q_{\max}$,[6] The minimization arguments of the problem P_0 are exactly the initial state $\mathbf{x}_0$ and the attack vector $\tilde{\mathbf{e}}^*$ [16]. Thus, in this case, the estimator P_0 also correctly identifies the set of attacked sensors $\mathcal{K}$. Furthermore, for noiseless systems, P_0 is optimal in the sense that if another estimator can recover the initial state (which would also result in identification of the attacked sensors), the attack-resilient state estimator based on P_0 can as well [16].

On the other hand, P_0 cannot be used when noisy sensor measurements are available (i.e., when $\Omega \neq \mathbf{0} \in \mathbb{R}^{pN}$). For instance, in this case, the point $(\mathbf{x}_0, \tilde{\mathbf{e}}^*)$ might not even be feasible. Thus, there is need to adapt problem P_0 to nonideal models that capture system noise. To achieve this, we consider the following problem that relaxes the equality constraint from (8) by including a noise allowance:

$$\begin{aligned} P_{0,\omega}: \quad & \min_{\tilde{\mathbf{e}},\mathbf{x}} \|\tilde{\mathbf{e}}\|_{l_2,l_0} \\ s.t. \quad & \tilde{\mathbf{y}} - \mathbf{O}\mathbf{x}_0 - \tilde{\mathbf{e}} = \tilde{\mathbf{w}} \\ & \tilde{\mathbf{w}} \in \Omega. \end{aligned} \tag{9}$$

The problem $P_{0,\omega}$ involves combinatorial optimization that can be solved using MILP solvers. However, solving $P_{0,\omega}$ is NP-hard in the general case, which limits its use on smaller size systems. A common approach used in compressed sensing is to replace the l_0 norm by the l_1 norm, which effectively convexifies the problem and reduces its computational requirements. Consequently, to perform the attack-resilient state estimation, we also consider the following optimization problem:

$$\begin{aligned} P_{1,\omega}: \quad & \min_{\tilde{\mathbf{e}},\mathbf{x}} \|\tilde{\mathbf{e}}\|_{l_2,l_1} \\ s.t. \quad & \tilde{\mathbf{y}} - \mathbf{O}\mathbf{x}_0 - \tilde{\mathbf{e}} = \tilde{\mathbf{w}} \\ & \tilde{\mathbf{w}} \in \Omega \end{aligned} \tag{10}$$

However, it is unclear what guarantees can be provided regarding the performance of the attack-resilient state estimators $P_{0,\omega}$ and $P_{1,\omega}$. Specifically, we are interested in obtaining worst-case bounds on the state estimation errors caused by noise and attacks on sensors, and answering the question whether the attacker can exploit the noise to introduce an unbounded state estimation error. We will also investigate conditions that ensure that the presented state estimators can be used to correctly identify the set of attacked sensors.

IV. Performance Guarantees for $P_{0,\omega}$ Estimator

In this section, we focus on the performance degradation of the $P_{0,\omega}$ state estimator due to the existence of noise. Specifically, we are interested in providing bounds on $\Delta\mathbf{x}^{l_0}$ that is defined as

$$(\mathbf{x}_{l_0,\omega}, \tilde{\mathbf{e}}^{l_0}) = \arg\min P_{0,\omega}, \quad q_{0,\omega} = \|\tilde{\mathbf{e}}^{l_0}\|_{l_2,l_0} \tag{11}$$

$$\Delta\mathbf{x}^{l_0} = \mathbf{x}_{l_0,\omega} - \mathbf{x}_0, \qquad \Delta\tilde{\mathbf{e}}^{l_0} = \tilde{\mathbf{e}}^{l_0} - \tilde{\mathbf{e}}^*. \tag{12}$$

We will also denote ith blocks of $\Delta\mathbf{x}^{l_0}, \Delta\tilde{\mathbf{e}}^{l_0}$, and $\tilde{\mathbf{e}}^{l_0}$ as $\Delta\mathbf{x}_i^{l_0}, \Delta\tilde{\mathbf{e}}_i^{l_0}$, and $\tilde{\mathbf{e}}_i^{l_0}$, respectively.

We consider systems where the number of compromised sensors $q = |\mathcal{K}|$ is not higher than $q_{\max}$—the maximal number of attacked sensors for which the system's state can be recovered in the noiseless case. Thus, before we proceed with our analysis, we first characterize conditions under which it is possible to perform the state estimation even for noiseless systems. We start with the following definition.

Definition 1 [32]: An LTI system with the form as in (1) is said to be s-sparse observable if for every set $\mathcal{K} \subset \mathcal{S}$ of size s (i.e., $|\mathcal{K}| = s$), the pair $(\mathbf{A}, P_{\mathcal{K}^\complement}\mathbf{C})$ is observable. ■

From the analysis in [16], the following holds.

Lemma 1: $q_{\max}$ is equal to the maximal s for which the system is $2s$-sparse observable.

For considered systems, the following theorem provides a bound on the maximal state estimation error caused by the existence of noise.

[5] The matrix $\mathbf{O}$ is obtained by reordering rows of the standard observability matrix for the system $(\mathbf{A}, \mathbf{C})$ and, thus, it has the same rank as the observability matrix.

[6] The number $q_{\max}$ depends on the properties of the observability matrix of the system. We will address this in more detail in Section IV.

Theorem 1: If q sensors have been attacked, where $q \leq q_{\max}$, then the error $\Delta\mathbf{x}^{l_0}$ of the state estimate obtained from optimization problem $P_{0,\omega}$ satisfies

$$\|\Delta\mathbf{x}^{l_0}\|_{l_2} \leq 2 \cdot \max_{\substack{\mathcal{R}\subset S,\\ |\mathcal{R}|=p-2q_{\max}}} \left(\|\mathbf{O}_{\mathcal{R}}^{\dagger}\|_{l_2} \cdot \max_{\tilde{\mathbf{w}}_{\mathcal{R}}\in\Omega_{\mathcal{R}}} \|\tilde{\mathbf{w}}_{\mathcal{R}}\|_{l_2} \right) \tag{13}$$

where $\mathbf{O}_{\mathcal{R}}^{\dagger}$ denotes the pseudoinverse of $\mathbf{O}_{\mathcal{R}}$ (i.e., $\mathbf{O}_{\mathcal{R}}^{\dagger} = (\mathbf{O}_{\mathcal{R}}^T\mathbf{O}_{\mathcal{R}})^{-1}\mathbf{O}_{\mathcal{R}}^T$). ■

Proof 1: From (12) and the definition of $P_{0,\omega}$, it follows that: $\|\Delta\tilde{\mathbf{e}}^{l_0} + \tilde{\mathbf{e}}^*\|_{l_2,l_0} \leq \|\tilde{\mathbf{e}}^*\|_{l_2,l_0}$. Since for all vectors $\mathbf{a},\mathbf{b}$, $\mathbb{I}(\|\mathbf{a}+\mathbf{b}\|_{l_2} > 0) \geq \mathbb{I}(\|\mathbf{a}\|_{l_2} < 0) - \mathbb{I}(\|\mathbf{b}\|_{l_2} > 0)$,[7] we have that $\|\Delta\tilde{\mathbf{e}}^{l_0} + \tilde{\mathbf{e}}^*\|_{l_2,l_0} \geq \|\Delta\tilde{\mathbf{e}}^{l_0}\|_{l_2,l_0} - \|\tilde{\mathbf{e}}^*\|_{l_2,l_0}$. Therefore

$$\|\Delta\tilde{\mathbf{e}}^{l_0}\|_{l_2,l_0} \leq 2\|\tilde{\mathbf{e}}^*\|_{l_2,l_0} \overset{r_1}{\leq} 2q_{\max}, \tag{14}$$

where r_1 holds because $\|\tilde{\mathbf{e}}^*\|_{l_2,l_0} = q$ and the number of attacked sensors q is bounded by $q_{\max}$.

From (6), we have that $\tilde{\mathbf{e}}^* = \tilde{\mathbf{y}} - \mathbf{O}\mathbf{x}_0 - \tilde{\mathbf{w}}^*$. Similarly, from the constraint (9) it follows that: $\tilde{\mathbf{e}}^{l_0} = \tilde{\mathbf{y}} - \mathbf{O}\mathbf{x}_{l_0,\omega} - \tilde{\mathbf{w}}^{l_0}$, which implies

$$\Delta\tilde{\mathbf{e}}^{l_0} = -\mathbf{O}\Delta\mathbf{x}^{l_0} - \Delta\tilde{\mathbf{w}}. \tag{15}$$

Here, $\Delta\tilde{\mathbf{w}} = \tilde{\mathbf{w}}^{l_0} - \tilde{\mathbf{w}}^*$, with $\tilde{\mathbf{w}}^{l_0}, \tilde{\mathbf{w}}^* \in \Omega$.

Therefore, from (14) and (15), there exists an, at most, $2q_{\max}$-sparse block vector $\tilde{\mathbf{z}} \in \mathbb{R}^{pN}$ ($\tilde{\mathbf{z}} = -\Delta\tilde{\mathbf{e}}^{l_0}$) with, at most, $2q_{\max}$ nonzero N-size blocks (since $\|\tilde{\mathbf{z}}\|_{l_2,l_0} \leq 2q_{\max}$)—such that

$$\mathbf{O}\Delta\mathbf{x}^{l_0} = -\Delta\tilde{\mathbf{w}} + \tilde{\mathbf{z}}.$$

This implies that at least $f = p - 2q_{\max}$ blocks of $\tilde{\mathbf{z}}$ are zero subvectors. Let us denote their indices as $i_1, \ldots i_f$, such that $i_1 < \cdots < i_f$ and the set of sensors corresponding to these indices as $\mathcal{R}$ (i.e., $\mathcal{R} = \{s_{i_1}, \ldots, s_{i_f}\}$). Hence, we have

$$\mathbf{O}_{\mathcal{R}}\Delta\mathbf{x}^{l_0} = -\Delta\tilde{\mathbf{w}}_{\mathcal{R}} \tag{16}$$

where $\Delta\tilde{\mathbf{w}}_{\mathcal{R}} = \tilde{\mathbf{w}}_{\mathcal{R}}^{l_0} - \tilde{\mathbf{w}}_{\mathcal{R}}^*$, with $\tilde{\mathbf{w}}_{\mathcal{R}}^{l_0}, \tilde{\mathbf{w}}_{\mathcal{R}}^* \in \Omega_{\mathcal{R}}$.

Note that the set $\mathcal{R}$ has $f = p - 2q_{\max}$ elements, and since the system is $2q_{\max}$-sparse observable (from Lemma 1), the pair $(\mathbf{A}, P_{\mathcal{R}}\mathbf{C})$ is observable (and $f \geq 1$). Thus, the matrix $\mathbf{O}_{\mathcal{R}}$ is full (column) rank and we can define the pseudoinverse matrix $\mathbf{O}_{\mathcal{R}}^{\dagger} = (\mathbf{O}_{\mathcal{R}}^T\mathbf{O}_{\mathcal{R}})^{-1}\mathbf{O}_{\mathcal{R}}^T$, from which it follows that:

$$\Delta\mathbf{x}^{l_0} = -\mathbf{O}_{\mathcal{R}}^{\dagger}\Delta\tilde{\mathbf{w}}_{\mathcal{R}} \Rightarrow \|\Delta\mathbf{x}^{l_0}\|_{l_2} \leq \|\mathbf{O}_{\mathcal{R}}^{\dagger}\|_{l_2} \cdot \|\Delta\tilde{\mathbf{w}}_{\mathcal{R}}\|_{l_2} \Rightarrow$$

$$\|\Delta\mathbf{x}^{l_0}\|_{l_2} \leq \max_{\substack{\mathcal{R}\subset S, |\mathcal{R}|=p-2q_{\max}\\ \tilde{\mathbf{w}}_{\mathcal{R}}^{l_0}, \tilde{\mathbf{w}}_{\mathcal{R}}^* \in \Omega_{\mathcal{R}}}} \left(\|\mathbf{O}_{\mathcal{R}}^{\dagger}\|_{l_2} \cdot \|\tilde{\mathbf{w}}_{\mathcal{R}}^* - \tilde{\mathbf{w}}_{\mathcal{R}}^{l_0}\|_{l_2} \right)$$

$$\leq \max_{\substack{\mathcal{R}\subset S,\\ |\mathcal{R}|=p-2q_{\max}}} \left(\|\mathbf{O}_{\mathcal{R}}^{\dagger}\|_{l_2} \cdot \max_{\tilde{\mathbf{w}}_{\mathcal{R}}^{l_0}, \tilde{\mathbf{w}}_{\mathcal{R}}^* \in \Omega_{\mathcal{R}}} \|\tilde{\mathbf{w}}_{\mathcal{R}}^* - \tilde{\mathbf{w}}_{\mathcal{R}}^{l_0}\|_{l_2} \right)$$

Since

$$\max_{\tilde{\mathbf{w}}_{\mathcal{R}}^{l_0}, \tilde{\mathbf{w}}_{\mathcal{R}}^* \in \Omega_{\mathcal{R}}} \|\tilde{\mathbf{w}}_{\mathcal{R}}^* - \tilde{\mathbf{w}}_{\mathcal{R}}^{l_0}\|_{l_2} \leq 2 \max_{\tilde{\mathbf{w}}_{\mathcal{R}} \in \Omega_{\mathcal{R}}} \|\tilde{\mathbf{w}}_{\mathcal{R}}\|_{l_2},$$

we have that (13) is satisfied, which concludes the proof.

It is important to highlight that the bound on the right-hand side of (13) is linear in the size of noise. Furthermore, the above theorem states that if, at most, $q_{\max}$ sensors have been compromised, the attacker cannot exploit the noise to introduce an unbounded state estimation error. Another thing to consider is the complexity of computing the term in (13). To determine the state estimation bound, we need to check $\binom{p}{p-2q_{\max}}$ different subsets $\mathcal{R}$ of the set $\mathcal{S}$, and for each $\mathcal{R}$ compute

$$\|\mathbf{O}_{\mathcal{R}}^{\dagger}\|_{l_2} \cdot \max_{\tilde{\mathbf{w}}\in\Omega_{\mathcal{R}}} \|\tilde{\mathbf{w}}\|_{l_2} = \lambda_{\max}^{\mathbf{O}_{\mathcal{R}}^{\dagger}} \cdot \max_{\tilde{\mathbf{w}}_{\mathcal{R}}\in\Omega_{\mathcal{R}}} \|\tilde{\mathbf{w}}_{\mathcal{R}}\|_{l_2},$$

where $\lambda_{\max}^{\mathbf{O}_{\mathcal{R}}^{\dagger}}$ denotes the largest singular value of $\mathbf{O}_{\mathcal{R}}^{\dagger}$, and

$$\max_{\tilde{\mathbf{w}}_{\mathcal{R}}\in\Omega_{\mathcal{R}}} \|\tilde{\mathbf{w}}_{\mathcal{R}}\|_{l_2} = \sqrt{\sum_{s_i\in\mathcal{R}} \sum_{k=0}^{N-1} (\delta_{w_{k,i}})^2}$$

for $\Omega_{\mathcal{R}}$ defined as in Section III.[8] This is significantly lower than the required computational cost for the robustness analysis from [24].

Finally, it is worth noting that for almost all systems (i.e., for almost all pairs of matrices $\mathbf{A}, \mathbf{C}$), we have that $q_{\max} = \lceil p/2 - 1 \rceil$ [16], meaning that $1 \leq p - 2q_{\max} \leq 2$. Thus, for almost all systems, to obtain the bound, we would need to evaluate the above term for either p or $p(p-1)/2$ sets $\mathcal{R}$ only.

V. Robustness of $P_{1,\omega}$ Estimator to Noise

In this section, we provide a bound on the error of the $P_{1,\omega}$ estimator due to noise. We start by introducing notation similar to the one used in the previous section:

$$(\mathbf{x}_{l_1,\omega}, \tilde{\mathbf{e}}^{l_1}) = \arg\min P_{1,\omega} \tag{17}$$

$$\Delta\mathbf{x}^{l_1} = \mathbf{x}_{l_1,\omega} - \mathbf{x}_0, \quad \Delta\tilde{\mathbf{e}}^{l_1} = \tilde{\mathbf{e}}^{l_1} - \tilde{\mathbf{e}}^* \tag{18}$$

Specifically, we are interested in obtaining a bound on $\Delta\mathbf{x}^{l_1}$.

Theorem 2: When sensors from set $\mathcal{K} \subset S$ are attacked, state estimation error $\Delta\mathbf{x}^{l_1}$ satisfies the following constraint

$$\sum_{s_i\in\mathcal{K}^{\complement}} \|\mathbf{O}_i\Delta\mathbf{x}^{l_1}\|_{l_2} \leq \sum_{s_i\in\mathcal{K}} \|\mathbf{O}_i\Delta\mathbf{x}^{l_1}\|_{l_2} + 2\sigma_{\Omega}, \tag{19}$$

where $\sigma_{\Omega} = \max_{\tilde{\mathbf{w}}\in\Omega} \|\tilde{\mathbf{w}}\|_{l_2,l_1}$. ■

Proof 2: Since $\tilde{\mathbf{e}}_{l_1}$ is a minimizer of the problem $P_{1,\omega}$, it follows that $\|\tilde{\mathbf{e}}_{l_1}\|_{l_2,l_1} \leq \|\tilde{\mathbf{e}}^*\|_{l_2,l_1}$. Thus, to find a bound on the $\Delta\tilde{\mathbf{e}}^{l_1}$ we consider a set that contains all feasible $\Delta\tilde{\mathbf{e}}^{l_1}$, which is defined as:

$$\left\{\Delta\tilde{\mathbf{e}}^{l_1} \in \mathbb{R}^{pN} \mid \|\Delta\tilde{\mathbf{e}}^{l_1} + \tilde{\mathbf{e}}^*\|_{l_2,l_1} \leq \|\tilde{\mathbf{e}}^*\|_{l_2,l_1}\right\}. \tag{20}$$

[7] Note that although l_0 is not convex (and thus not a norm), it satisfies the triangular inequality.

[8] On the other hand, if the noise bounds in Ω are defined as bounds on the l_2 norm of noise for each sensor at each time step, this term would be equal to the sum of the squared norms.

By starting from the the above feasibility condition, it follows that:

$$\sum_{i=1}^{p} \|\Delta\tilde{\mathbf{e}}_i^{l_1} + \tilde{\mathbf{e}}_i^*\|_{l_2} \le \sum_{i=1}^{p} \|\tilde{\mathbf{e}}_i^*\|_{l_2} \Rightarrow$$

$$0 \ge \sum_{i=1}^{p} \|\Delta\tilde{\mathbf{e}}_i^{l_1} + \tilde{\mathbf{e}}_i^*\|_{l_2} - \sum_{i=1}^{p} \|\tilde{\mathbf{e}}_i^*\|_{l_2} =$$

$$\overset{r_1}{=} \sum_{s_i \in \mathcal{K}} \|\Delta\tilde{\mathbf{e}}_i^{l_1} + \tilde{\mathbf{e}}_i^*\|_{l_2} + \sum_{s_i \in \mathcal{K}^{\complement}} \|\Delta\tilde{\mathbf{e}}_i^{l_1}\|_{l_2} - \sum_{s_i \in \mathcal{K}} \|\tilde{\mathbf{e}}_i^*\|_{l_2} \ge$$

$$\overset{r_2}{\ge} \sum_{s_i \in \mathcal{K}} \|\tilde{\mathbf{e}}_i^*\|_{l_2} - \sum_{s_i \in \mathcal{K}} \|\Delta\tilde{\mathbf{e}}_i^{l_1}\|_{l_2}$$

$$+ \sum_{s_i \in \mathcal{K}^{\complement}} \|\Delta\tilde{\mathbf{e}}_i^{l_1}\|_{l_2} - \sum_{s_i \in \mathcal{K}} \|\tilde{\mathbf{e}}_i^*\|_{l_2} =$$

$$= \sum_{s_i \in \mathcal{K}^{\complement}} \|\Delta\tilde{\mathbf{e}}_i^{l_1}\|_{l_2} - \sum_{s_i \in \mathcal{K}} \|\Delta\tilde{\mathbf{e}}_i^{l_1}\|_{l_2} =$$

$$= \sum_{i=1}^{p} \|\Delta\tilde{\mathbf{e}}_i^{l_1}\|_{l_2} - 2\sum_{s_i \in \mathcal{K}} \|\Delta\tilde{\mathbf{e}}_i^{l_1}\|_{l_2}.$$

Here, r_1 is satisfied by the fact that only sensors from the set $\mathcal{K}$ are attacked—i.e., all other blocks of the attack vector $\tilde{\mathbf{e}}^*$ are zero. Relation r_2 follows from the fact that $\|a+b\| \ge \|a\| - \|b\|$, for any a, b.[9]

Thus, the set from (20) can be overapproximated by the set

$$\left\{ \Delta\tilde{\mathbf{e}}^{l_1} \in \mathbb{R}^{pN} \mid \|\Delta\tilde{\mathbf{e}}^{l_1}\|_{l_2,l_1} \le 2\sum_{s_i \in \mathcal{K}} \|\Delta\tilde{\mathbf{e}}_i^{l_1}\|_{l_2} \right\} \tag{21}$$

From (5), we have that $\tilde{\mathbf{e}}_i^* = \tilde{\mathbf{y}}_i - \mathbf{O}_i\mathbf{x}_0 - \tilde{\mathbf{w}}_i^*$. Similarly, from (10) it follows that $\tilde{\mathbf{e}}_i^{l_1} = \tilde{\mathbf{y}}_i - \mathbf{O}_i\mathbf{x}_{l_1,\omega} - \tilde{\mathbf{w}}_i^{l_1}$, which implies

$$\Delta\tilde{\mathbf{e}}_i^{l_1} = -\mathbf{O}_i\Delta\mathbf{x}^{l_1} - \Delta\tilde{\mathbf{w}}_i \tag{22}$$

where $\Delta\tilde{\mathbf{w}}_i = \tilde{\mathbf{w}}^{l_1} - \tilde{\mathbf{w}}_i^*$ with $\tilde{\mathbf{w}}_i^{l_1}, \tilde{\mathbf{w}}_i^* \in \Omega_{\{s_i\}}$.

Hence, the constraint from the set definition in (21) can be represented as:

$$\sum_{i=1}^{p} \|\Delta\tilde{\mathbf{e}}_i^{l_1}\|_{l_2} \le 2\sum_{s_i \in \mathcal{K}} \|\Delta\tilde{\mathbf{e}}_i^{l_1}\|_{l_2} \quad \Leftrightarrow$$

$$\sum_{s_i \in \mathcal{K}^{\complement}} \|\mathbf{O}_i\Delta\mathbf{x}^{l_1} + \Delta\tilde{\mathbf{w}}_i\|_{l_2} \le \sum_{s_i \in \mathcal{K}} \|\mathbf{O}_i\Delta\mathbf{x}^{l_1} + \Delta\tilde{\mathbf{w}}_i\|_{l_2} \tag{23}$$

Again, using properties of norms we have

$$\|\mathbf{O}_i\Delta\mathbf{x}^{l_1} + \Delta\tilde{\mathbf{w}}_i\|_{l_2} \ge \|\mathbf{O}_i\Delta\mathbf{x}^{l_1}\|_{l_2} - \|\Delta\tilde{\mathbf{w}}_i\|_{l_2}$$

$$\|\mathbf{O}_i\Delta\mathbf{x}^{l_1} + \Delta\tilde{\mathbf{w}}_i\|_{l_2} \le \|\mathbf{O}_i\Delta\mathbf{x}^{l_1}\|_{l_2} + \|\Delta\tilde{\mathbf{w}}_i\|_{l_2},$$

and thus, from (23) it follows that the set constraint from (21) can be additionally relaxed to

$$\sum_{s_i \in \mathcal{K}^{\complement}} \|\mathbf{O}_i\Delta\mathbf{x}^{l_1}\|_{l_2} \le \sum_{s_i \in \mathcal{K}} \|\mathbf{O}_i\Delta\mathbf{x}^{l_1}\|_{l_2} + \delta_w, \tag{24}$$

where

$$\delta_w = \sum_{i=1}^{p} \|\Delta\tilde{\mathbf{w}}_i\|_{l_2} \le 2 \max_{\tilde{\mathbf{w}}_i \in \Omega_{\{s_i\}}} \sum_{i=1}^{p} \|\tilde{\mathbf{w}}_i\|_{l_2} = 2\max_{\tilde{\mathbf{w}} \in \Omega} \|\tilde{\mathbf{w}}\|_{l_2,l_1},$$

which concludes the proof.

Remark 1: Proposition 6 from [16] states that $P_{1,\omega}$ can correctly estimate the state for noiseless systems ($\Omega = \mathbf{0}$) if and only if for all $\mathcal{K}$ such that $|\mathcal{K}| = q$, it holds that:

$$\sum_{s_i \in \mathcal{K}^{\complement}} \|\mathbf{O}_i\mathbf{x}\|_{l_2} > \sum_{s_i \in \mathcal{K}} \|\mathbf{O}_i\mathbf{x}\|_{l_2}, \quad \forall \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}. \tag{25}$$

This means that (19) is tight for noiseless systems, since for $\Omega = \mathbf{0}$ (19) takes the form $\sum_{s_i \in \mathcal{K}^{\complement}} \|\mathbf{O}_i\Delta\mathbf{x}^{l_1}\|_{l_2} \le \sum_{s_i \in \mathcal{K}} \|\mathbf{O}_i\Delta\mathbf{x}^{l_1}\|_{l_2}$; this constraint when combined with (25) implies that for noiseless systems $\Delta\mathbf{x}^{l_1} = \mathbf{0}$, meaning that the state is correctly reconstructed.

Finally, if we consider systems that can deal with up to q attacks when there is no noise, from (19) and (25) it follows that the feasible set for the state estimation vector $\Delta\mathbf{x}^{l_1}$ can be described as the set where $\Delta\mathbf{x}^{l_1} = \mathbf{0}$ or it satisfies

$$\sum_{s_i \in \mathcal{K}} \|\mathbf{O}_i\Delta\mathbf{x}^{l_1}\|_{l_2} < \sum_{s_i \in \mathcal{K}^{\complement}} \|\mathbf{O}_i\Delta\mathbf{x}^{l_1}\|_{l_2}$$

$$\le \sum_{s_i \in \mathcal{K}} \|\mathbf{O}_i\Delta\mathbf{x}^{l_1}\|_{l_2} + 2\sigma_\Omega \tag{26}$$

for all $\mathcal{K} \subset \mathcal{S}$, such that $|\mathcal{K}| = q$.

Now, consider any vector $\mathbf{v} \in \mathbb{R}^n$. From (25), there exists $\epsilon > 0$ such that $\sum_{s_i \in \mathcal{K}} \|\mathbf{O}_i\mathbf{v}\|_{l_2} + \epsilon = \sum_{s_i \in \mathcal{K}^{\complement}} \|\mathbf{O}_i\mathbf{v}\|_{l_2}$, and thus all vectors of the form $\alpha\mathbf{v}$, where $\alpha \in (0, 2\sigma_\Omega/\epsilon]$ satisfy (26), while for all $\alpha > 2\sigma_\Omega/\epsilon$, $\alpha\mathbf{v}$ does not belong to the feasibility region (26). This implies that starting from $\mathbf{0}$ and moving in any direction, a point will be reached after which all new points along that direction do not satisfy (26). Therefore, when the condition (25) for correct estimation for noiseless systems is satisfied, then there exists a bounded solution of (26)—i.e., the maximal state estimation error $\Delta\mathbf{x}^{l_1}$ is bounded.

Finally, it is worth noting that the condition (25) corresponds to the *Null-Space Property* commonly used in compressed sensing literature. When (25) holds, if $\Delta\mathbf{x}^{l_1}$ is the maximization point for noise level σ_Ω, then for noise level $\tilde{\sigma}_\Omega = k\sigma_\Omega$, $k > 0$, error vector $k\Delta\mathbf{x}^{l_1}$ satisfies (19) for the new noise level $\tilde{\sigma}_\Omega$. The scaled vector is also a maximization point of the new problem (with noise level $\tilde{\sigma}_\Omega$), since otherwise there would exist a state error vector with a larger l_2 norm for the initial problem (with noise level σ_Ω). Consequently, in this case, the estimation error bound obtained by maximization of $\|\Delta\mathbf{x}^{l_1}\|_2$ over the set described by (19) depends linearly on the size of the noise. ■

From the relationship between l_2 and l_1 norms where

$$\|\alpha\|_{l_1} \ge \|\alpha\|_{l_2} \ge \frac{1}{\sqrt{n}}\|\alpha\|_{l_1}, \quad \forall\alpha \in \mathbb{R}^n, \tag{27}$$

[9] In this paper, when a norm is not clearly specified we imply that a statement is valid for any norm.

it follows that $\|\mathbf{O}_i\Delta\mathbf{x}^{l_1}\|_{l_1} \geq \|\mathbf{O}_i\Delta\mathbf{x}^{l_1}\|_{l_2} \geq \frac{1}{\sqrt{N}}\|\mathbf{O}_i\Delta\mathbf{x}^{l_1}\|_{l_1}$. Therefore,

$$\sum_{s_i\in\mathcal{K}^{\complement}} \|\mathbf{O}_i\Delta\mathbf{x}^{l_1}\|_{l_2} \geq \frac{1}{\sqrt{N}} \sum_{s_i\in\mathcal{K}^{\complement}} \|\mathbf{O}_i\Delta\mathbf{x}^{l_1}\|_{l_1} = \frac{\|\mathbf{O}_{\mathcal{K}^{\complement}}\Delta\mathbf{x}^{l_1}\|_{l_1}}{\sqrt{N}}$$

$$\sum_{s_i\in\mathcal{K}} \|\mathbf{O}_i\Delta\mathbf{x}^{l_1}\|_{l_2} \leq \sum_{s_i\in\mathcal{K}} \|\mathbf{O}_i\Delta\mathbf{x}^{l_1}\|_{l_1} = \|\mathbf{O}_{\mathcal{K}}\Delta\mathbf{x}^{l_1}\|_{l_1}.$$

The above inequalities along with Theorem 2 prove the following corollary.

Corollary 1: When sensors from set $\mathcal{K} \subset S$ are attacked, the state estimation error $\Delta\mathbf{x}^{l_1}$ satisfies

$$\|\mathbf{O}_{\mathcal{K}^{\complement}}\Delta\mathbf{x}^{l_1}\|_{l_1} \leq \sqrt{N}\|\mathbf{O}_{\mathcal{K}}\Delta\mathbf{x}^{l_1}\|_{l_1} + 2\sqrt{N}\sigma_\Omega, \tag{28}$$

where $\sigma_\Omega = \max_{\tilde{\mathbf{w}}\in\Omega}\|\tilde{\mathbf{w}}\|_{l_2,l_1}$.

Both conditions from Theorem 2 and Corollary 1 define sets that contain all feasible $\Delta\mathbf{x}^{l_1}$ when less than or equal to q sensors are attacked; the case where $q_1 < q$ sensors are attacked is covered by the scenario where $|\mathcal{K}| = q$ sensors are compromised, but $q - q_1$ sensors are inserting zero signals. However, maximization problems over these sets may be hard to solve in the general case. Thus, we introduce the following theorem that for a special class of systems provides an analytic formula for $\|\Delta\mathbf{x}^{l_1}\|_{l_2}$.

Theorem 3: Suppose that for all $\mathcal{K} \subset \mathcal{S}$ with $|\mathcal{K}| = q$ it holds

$$\mathbf{O}_{\mathcal{K}^{\complement}}^T\mathbf{O}_{\mathcal{K}^{\complement}} - qN^2\mathbf{O}_{\mathcal{K}}^T\mathbf{O}_{\mathcal{K}} \succeq \lambda\mathbf{I}_n \tag{29}$$

for some $\lambda > 0$. Then if at most q nodes are compromised the following condition holds:

$$\|\Delta\mathbf{x}^{l_1}\|_{l_2} \leq \frac{2\sqrt{N}\sigma_\Omega}{\lambda} \cdot \max_{\mathcal{K}\subset\mathcal{S},|\mathcal{K}|=q}\left(\|\mathbf{O}_{\mathcal{K}^{\complement}}\|_{l_2} + \sqrt{q}N\|\mathbf{O}_{\mathcal{K}}\|_{l_2}\right) \tag{30}$$

■

Proof 3: We start by assuming that the set of compromised sensors $\mathcal{K}$ has q elements. From (28) and (27) it follows that

$$\|\mathbf{O}_{\mathcal{K}^{\complement}}\Delta\mathbf{x}^{l_1}\|_{l_2} \leq \sqrt{q}N\|\mathbf{O}_{\mathcal{K}}\Delta\mathbf{x}^{l_1}\|_{l_2} + 2\sqrt{N}\sigma_\Omega. \tag{31}$$

Let's denote $T = \|\mathbf{O}_{\mathcal{K}^{\complement}}\Delta\mathbf{x}^{l_1}\|_{l_2} - \sqrt{q}N\|\mathbf{O}_{\mathcal{K}}\Delta\mathbf{x}^{l_1}\|_{l_2}$. Then, it holds that for $\Delta\mathbf{x}^{l_1} \neq \mathbf{0}$:

$$\begin{aligned}
T &= \frac{\|\mathbf{O}_{\mathcal{K}^{\complement}}\Delta\mathbf{x}^{l_1}\|_{l_2}^2 - qN^2\|\mathbf{O}_{\mathcal{K}}\Delta\mathbf{x}^{l_1}\|_{l_2}^2}{\|\mathbf{O}_{\mathcal{K}^{\complement}}\Delta\mathbf{x}^{l_1}\|_{l_2} + \sqrt{q}N\|\mathbf{O}_{\mathcal{K}}\Delta\mathbf{x}^{l_1}\|_{l_2}} \\
&= \frac{(\Delta\mathbf{x}^{l_1})^T\left(\mathbf{O}_{\mathcal{K}^{\complement}}^T\mathbf{O}_{\mathcal{K}^{\complement}} - qN^2\mathbf{O}_{\mathcal{K}}^T\mathbf{O}_{\mathcal{K}}\right)\Delta\mathbf{x}^{l_1}}{\|\mathbf{O}_{\mathcal{K}^{\complement}}\Delta\mathbf{x}^{l_1}\|_{l_2} + \sqrt{q}N\|\mathbf{O}_{\mathcal{K}}\Delta\mathbf{x}^{l_1}\|_{l_2}} \\
&\overset{r_1}{\geq} \frac{\lambda\|\Delta\mathbf{x}^{l_1}\|_{l_2}^2}{\|\mathbf{O}_{\mathcal{K}^{\complement}}\Delta\mathbf{x}^{l_1}\|_{l_2} + \sqrt{q}N\|\mathbf{O}_{\mathcal{K}}\Delta\mathbf{x}^{l_1}\|_{l_2}} \\
&\overset{r_2}{\geq} \frac{\lambda\|\Delta\mathbf{x}^{l_1}\|_{l_2}^2}{(\|\mathbf{O}_{\mathcal{K}^{\complement}}\|_{l_2} + \sqrt{q}N\|\mathbf{O}_{\mathcal{K}}\|_{l_2})\|\Delta\mathbf{x}^{l_1}\|_{l_2}} \\
&= \frac{\lambda}{(\|\mathbf{O}_{\mathcal{K}^{\complement}}\|_{l_2} + \sqrt{q}N\|\mathbf{O}_{\mathcal{K}}\|_{l_2})}\|\Delta\mathbf{x}^{l_1}\|_{l_2}
\end{aligned} \tag{32}$$

Here, r_1 holds due to (29), while r_2 is caused by the fact that $\|\mathbf{A}\mathbf{x}\|_{l_2} \leq \|\mathbf{A}\|_{l_2}\|\mathbf{x}\|_{l_2}$ for any matrix $\mathbf{A}$ and vector $\mathbf{x}$.

Therefore, from (31) and (32), $\Delta\mathbf{x}^{l_1}$ satisfies

$$\frac{\lambda}{(\|\mathbf{O}_{\mathcal{K}^{\complement}}\|_{l_2} + \sqrt{q}N\|\mathbf{O}_{\mathcal{K}}\|_{l_2})}\|\Delta\mathbf{x}^{l_1}\|_{l_2} \leq 2\sqrt{N}\sigma_\Omega \tag{33}$$

which implies (30).

On the other hand, consider a set $\mathcal{K}_1 \subset \mathcal{S}$ with $|\mathcal{K}_1| < q$. If (29) is satisfied for all $\mathcal{K}$ with exactly q sensors, then it also holds for $\mathcal{K}_1$. This case is also covered by the scenario where q sensors including the sensors from $\mathcal{K}_1$ are compromised, while the injected attack signals on $q - |\mathcal{K}_1|$ sensors that are not in $\mathcal{K}_1$ are equal to zero. The error signal in this case satisfies (30), although effectively only $|\mathcal{K}_1| < q$ are used to inject attacks, which concludes the proof.

Although Theorem 3 provides an analytic bound for the worst-case state estimation error obtained by $P_{1,\omega}$ for a certain class of systems, it is worth noting that it could heavily overapproximate the error due to the gains caused by the conversions between the norms (i.e., factor $\sqrt{q}N$). Still, along with Theorem 2 and Corollary 1, it provides the first analytic relation showing that the worst-case error is linear with the size of the noise, as in the case for the $P_{0,\omega}$ estimator.

Finally, when $N = 1$, the state estimation does not consider previous sensor measurements—i.e., only sensor measurements from the current time-step are taken into account. Thus, in this case, any knowledge of the system's dynamics is not utilized, and the problem if effectively mapped into the problem of estimating the state of a non-dynamical system from a single set of potentially compromised sensor measurements. This problem has drawn considerable attention in some CPS domains, mainly in the smart-grids community (e.g., [33]). In this case, the sufficient condition (29) for the analytic bound (30) from Theorem 3 can be specified as

$$(P_{\mathcal{K}^{\complement}}\mathbf{C})^T P_{\mathcal{K}^{\complement}}\mathbf{C} - q(P_{\mathcal{K}}\mathbf{C})^T P_{\mathcal{K}}\mathbf{C} \succeq \lambda\mathbf{I}_n \tag{34}$$

for some $\lambda > 0$. In addition to considerably reducing the conservativeness of (30) for $N = 1$, the condition from (34) is significantly less restrictive, which enables the use of the bound (30) (where $N = 1$) for a large class of systems.

VI. Attack Identification in Presence of Noise

In addition to computing a state estimate, the presented attack-resilient state estimation procedures also estimate attack vectors injected at time steps $k = 0, 1, \ldots, N-1$ (i.e., vectors $\tilde{\mathbf{e}}^{l_t}, t = 0, 1$). Thus, in this section we consider conditions for which the attack vectors estimates can be used for *sound identification* of compromised sensors; here, by sound identification we refer to methods that ensure that no uncompromised (i.e., valid) sensors would be identified as under attack. To simplify the notation, we will use the l_t notation (instead of l_0 or l_1) whenever we describe results that hold for both $P_{0,\omega}$ and $P_{1,\omega}$ obtained estimates.

An obvious candidate for identification procedure would be to use the policy that classifies sensor s_i as attacked if and only if $\mathbb{I}(\tilde{\mathbf{e}}_i^{l_t} \neq \mathbf{0})$. Note that, unless we can guarantee that the set of identified attacked sensors is a subset of the actual set of attacked sensors $\mathcal{K}$ (which is not known in advance), we cannot guarantee soundness of this identification procedure.[10]

On the other hand, we can use the state estimation guarantees presented in the previous two sections to provide a sound attack identification procedure. Consider the vector $\Delta\tilde{\mathbf{e}}^{l_t}$. If $\tilde{\mathbf{e}}_i^* = \mathbf{0}$ (i.e., sensor s_i is not attacked), then $\Delta\tilde{\mathbf{e}}_i^{l_t} = \tilde{\mathbf{e}}_i^{l_t}$. Consequently, if there is a bound on the values for $\Delta\tilde{\mathbf{e}}_i^{l_t}$ (i.e., error of the attack vector estimation for sensor s_i), we can guarantee that all attack vector estimates $\tilde{\mathbf{e}}_i^{l_t}$ that violate the bound effectively correspond to scenarios where sensor s_i is attacked.

To determine this bound, referred to as $D_i^{\tilde{\mathbf{e}}^{l_t}}$, we use that from (15): $\Delta\tilde{\mathbf{e}}_i^{l_t} = -\mathbf{O}_i\Delta\mathbf{x}^{l_t} - \Delta\tilde{\mathbf{w}}_i$. Thus,

$$\begin{aligned}\|\Delta\tilde{\mathbf{e}}_i^{l_t}\|_{l_2} &\leq \|\mathbf{O}_i\|_{l_2}\|\Delta\mathbf{x}^{l_t}\|_{l_2} + \|\Delta\tilde{\mathbf{w}}_i\|_{l_2}\\ &\leq \|\mathbf{O}_i\|_{l_2}\|\Delta\mathbf{x}^{l_t}\|_{l_2} + 2\max_{\tilde{\mathbf{w}}_i\in\Omega_{\{s_i\}}}\|\tilde{\mathbf{w}}_i\|_{l_2}. \qquad (35)\end{aligned}$$

Therefore, the bounds for $\|\Delta\mathbf{x}^{l_t}\|_{l_2}$, which we will refer to as $D^{\mathbf{x}^{l_t}}$, can be used to compute a bound for $\|\Delta\tilde{\mathbf{e}}_i^{l_t}\|_{l_2}$ as follows

$$D_i^{\tilde{\mathbf{e}}^{l_t}} = \|\mathbf{O}_i\|_{l_2} D^{\mathbf{x}^{l_t}} + 2\max_{\tilde{\mathbf{w}}_i\in\Omega_{\{s_i\}}}\|\tilde{\mathbf{w}}_i\|_{l_2}.$$

For instance, when $P_{0,\omega}$ is used, the bound $D_i^{\tilde{\mathbf{e}}^{l_0}}$ on $\|\Delta\tilde{\mathbf{e}}_i^{l_t}\|_{l_2}$ is

$$\begin{aligned}D_i^{\tilde{\mathbf{e}}^{l_0}} = 2\|\mathbf{O}_i\|_{l_2}\cdot \max_{\substack{\mathcal{R}\subset S,\\ |\mathcal{R}|=p-2q_{\max}}}&\left(\|\mathbf{O}_{\mathcal{R}}^{\dagger}\|_{l_2}\cdot\max_{\tilde{\mathbf{w}}_{\mathcal{R}}\in\Omega_{\mathcal{R}}}\|\tilde{\mathbf{w}}_{\mathcal{R}}\|_{l_2}\right)\\ &+ 2\max_{\tilde{\mathbf{w}}_i\in\Omega_{\{s_i\}}}\|\tilde{\mathbf{w}}_i\|_{l_2} \qquad (36)\end{aligned}$$

Now we can define a $P_{t,\omega}$-based ($t = 0, 1$) attack identification scheme as:

$$Attacked^{l_t}(s_i) = \mathbb{I}(\|\tilde{\mathbf{e}}_i^{l_t}\|_{l_2} > D_i^{\tilde{\mathbf{e}}^{l_t}}), \quad i = 1,\ldots,p. \qquad (37)$$

The following theorem shows soundness of the proposed attack identification scheme.

Theorem 4: If $Attacked^{l_t}(s_i) = 1$ then sensor $s_i \in \mathcal{K}$. Furthermore, for all attack vectors $\tilde{\mathbf{e}}^*$ for which $\|\tilde{\mathbf{e}}_i^*\|_{l_2} > 2D_i^{\tilde{\mathbf{e}}^{l_t}}$, the attack on sensor s_i will be correctly detected (i.e., $Attacked^{l_t}(s_i) = 1$). ■

Proof 4: Suppose $Attacked^{l_t}(s_i) = 1$, implying that $\|\tilde{\mathbf{e}}_i^{l_t}\|_{l_2} > D_i^{\tilde{\mathbf{e}}^{l_t}}$. Thus,

$$D_i^{\tilde{\mathbf{e}}^{l_t}} < \|\Delta\tilde{\mathbf{e}}_i^{l_t} + \tilde{\mathbf{e}}_i^*\|_{l_2} \leq \|\Delta\tilde{\mathbf{e}}_i^{l_t}\|_{l_2} + \|\tilde{\mathbf{e}}_i^*\|_{l_2} \leq D_i^{\tilde{\mathbf{e}}^{l_t}} + \|\tilde{\mathbf{e}}_i^*\|_{l_2}.$$

Thus, $\|\tilde{\mathbf{e}}_i^*\|_{l_2} > 0$ (i.e., the actual attack vector on s_i is non-zero), which means that sensor $s_1 \in \mathcal{K}$.

On the other hand, let's assume that $\|\tilde{\mathbf{e}}_i^*\|_{l_2} > 2D_i^{\tilde{\mathbf{e}}^{l_t}}$. This implies the following:

$$2D_i^{\tilde{\mathbf{e}}^{l_t}} < \|\tilde{\mathbf{e}}_i^{l_t} - \Delta\tilde{\mathbf{e}}_i^{l_t}\|_{l_2} \leq \|\tilde{\mathbf{e}}_i^{l_t}\|_{l_2} + \|\Delta\tilde{\mathbf{e}}_i^{l_t}\|_{l_2} \leq \|\tilde{\mathbf{e}}_i^{l_t}\|_{l_2} + D_i^{\tilde{\mathbf{e}}^{l_t}}.$$

Hence, $\|\tilde{\mathbf{e}}_i^{l_t}\|_{l_2} > D_i^{\tilde{\mathbf{e}}^{l_t}}$, and $Attacked^{l_t}(s_i) = 1$.

VII. Evaluation

To evaluate conservativeness of the state-estimation error bounds presented in this work, we exploit the evaluation approach from [24]. We randomly generated 100 systems with $n = 10$ states and $p = 5$ sensors, and 100 systems with $n = 20$ states and $p = 11$ sensors. Each of these systems had measurement models satisfying that the rows of the $\mathbf{C}$ matrix have unit magnitude, while all noise-bound vectors δ_{w_k}, $k \geq 0$, had elements between 0 and 2. For each of the generated 200 systems, with 10 or 20 states, we evaluated the state-estimation error $\Delta\mathbf{x}^{l_0}$ in 1000 experiments for various attack and noise realizations, where the number of attacked sensors was less than or equal to 2 for the systems with $p = 5$ sensors, and less than or equal to 5 for systems with $p = 11$ sensors, while noise realizations were bounded by the noise bounds specified by the system's δ_{w_k}.

The focus of our evaluation was the comparison between the state-estimation bounds and the observed state estimation errors due to the presence of noise. In both simulations and calculations of the error bounds we considered the case when the window size N is equal to the number of system states (i.e., $N = n$). The results of our evaluation are presented in Figs. 1 and 2. Figs. 1(a) and 2(a) present histograms of $\Delta\mathbf{x}^{l_0}$ errors for all 1000 attack scenarios for two randomly selected system with $n = 10$ and $n = 20$ states. As can be seen, the bound from Theorem 1 is an order of magnitude larger than the average state-estimation error for each system.

Furthermore, we investigated the ratio between the worst-case observed state estimation error for all 1000 simulations of each system $\mathfrak{S}$—i.e., $\max_{i=1:1000}\|\Delta\mathbf{x}_{\mathfrak{S}}^{l_0}\|_2$, and the system's error bounds $D_{\mathfrak{S}}^{\mathbf{x}^{l_0}}$ from Theorem 1

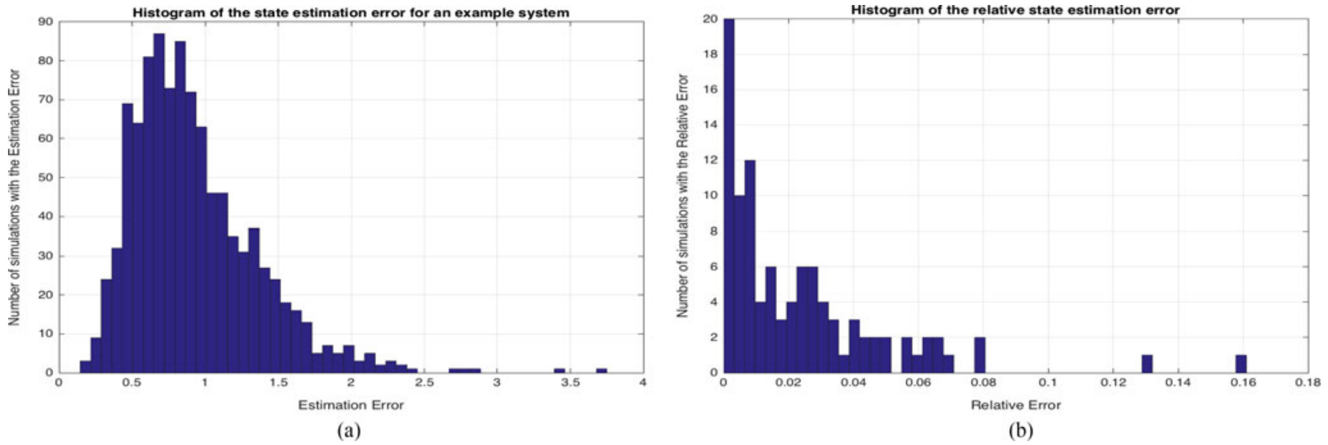
$$Rel\_error_{\mathfrak{S}} = \frac{\max_{i=1:1000}\|\Delta\mathbf{x}_{\mathfrak{S}}^{l_0}\|_2}{D_{\mathfrak{S}}^{\mathbf{x}^{l_0}}}.$$

Histograms of the relative errors for both types of systems are shown in Fig. 1(b) and 2(b). As can be observed, the maximal observed state estimation error reaches 16% of the computed bound for smaller systems ($n = 10$ states), while for larger systems (with $n = 20$ states) the maximal relative error reaches 1.5% of the computed bounds.

Conservativeness of the presented results was partially caused by the fact that we only simulated random initial points and random attack vectors, where sensor attacks are generated independently for each attacked sensor and noise profiles. As a result, the considered scenarios clearly do not capture worst-case attacks (i.e., attacks that could maximize the state estimation errors); for each system, to obtain scenarios that result in the worst-case estimation errors it is necessary to derive the corresponding attack vectors (and the initial states), which is beyond the scope of this paper.

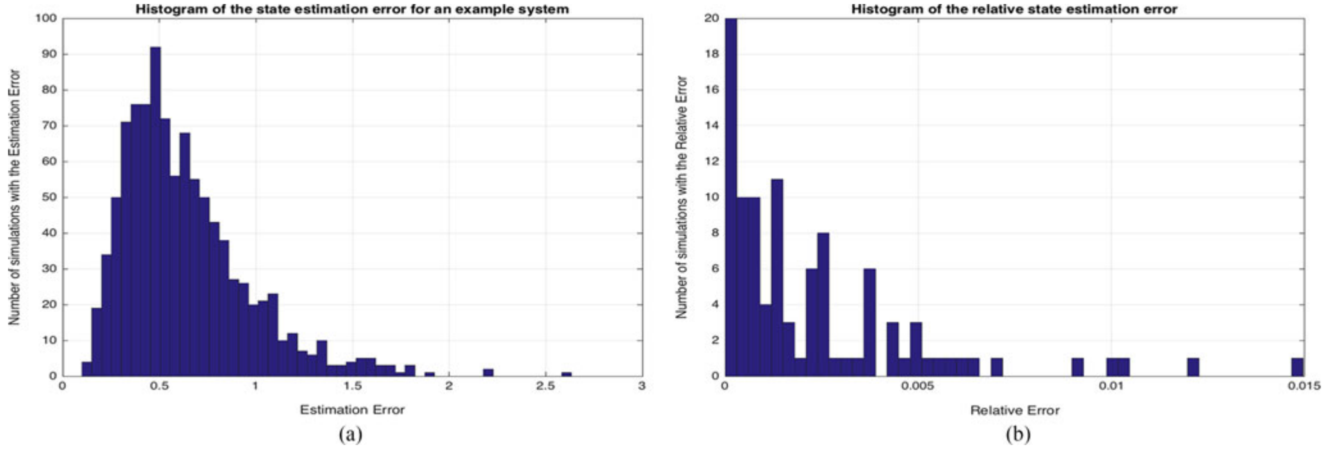
This has been highlighted in the discrepancy of the relative estimation errors for systems with different size, as illustrated in the histograms in Figs. 1(b) and 2(b). While simulating different attack and noise realizations, we observed that the obtained maximal relative estimation error reduces with an increase in

[10] To the best of our knowledge, even for a simpler problem of estimation of sparse signals α_0 from noisy measurements $\mathbf{z}$ obtained using an overcomplete dictionary $\mathbf{\Phi}$ (i.e., $\mathbf{z} = \Phi\alpha_0 + \mathbf{v}$), the l_0 based solution [26], [27] does not guarantee correct support recovery for α_0.



(a) (b)

Fig. 1. Simulation results for 1000 runs of 100 randomly selected systems with $n = 10$ states and $p = 5$ sensors. (a) Histogram for a system with the precomputed error bound equal to 28.6, (b) Histogram of the maximal relative state-estimation errors for all 100 systems.



(a) (b)

Fig. 2. Simulation results for 1000 runs of 100 randomly selected systems with $n = 20$ states and $p = 11$ sensors. (a) Histogram for a system with the precomputed error bound equal to 217.4, (b) Histogram of the maximal relative state-estimation errors for all 100 systems.

the system size n and hence the window size (since we analyzed systems for $N = n$). This can be explained by the fact that with the increase of the window size N, the number of attack vectors is also increased, meaning that due to the random attack vector selections, the probabilities to incorporate a worst-case attack is significantly reduced.

However, for small systems (e.g., $n = 1, 2$ states) we were able to generate initial states and attack vectors for which the obtained bounds were tight—i.e., the error $\|\Delta \mathbf{x}^{l_0}\|_2$ is equal to the obtained bounds. For instance, consider a dynamical system with scalar state x_0, noise vector bound $\delta_w = [\delta_{w,1}\ \delta_{w,2}\ \delta_{w,3}]^T$, sensor output vector $\mathbf{c} = [c_1\ c_2\ c_3]^T$, where $\frac{c_1}{\delta_{w,1}} \geq \frac{c_2}{\delta_{w,2}} \geq \frac{c_3}{\delta_{w,3}}$, and let's assume that either sensor s_2 or s_3 has been compromised (e.g., sensor s_2, making attack vectors $\mathbf{e}^*_{0,1} = \mathbf{e}^*_{0,3} = 0$). Note that in this case the window size is $N = 1$. For this example, if the obtained sensor measurement on sensor s_1 is $\mathbf{y}_{0,1} = c_1 \cdot x_0 - \delta_{w,1}$ and the attack signal inserted via the second sensor is $\mathbf{e}^*_{0,2} = -2\frac{c_2}{c_1}\delta_{w,1}$, then state estimate $\mathbf{x}_{l_0,\omega} = x_0 - 2\frac{\delta_{w,1}}{c_1}$ and attack vector with $\tilde{\mathbf{e}}^{l_0}_{0,1} = \tilde{\mathbf{e}}^{l_0}_{0,2} = 0$ is one of the solutions of the problem $P_{0,\omega}$ from (9) because it would satisfy that

$$\mathbf{y}_{0,1} - c_1 \cdot \mathbf{x}_{l_0,\omega} - \tilde{\mathbf{e}}^{l_0}_{0,1} = \delta_{w,1} \leq \delta_{w,1}$$

$$\mathbf{y}_{0,2} - c_2 \cdot \mathbf{x}_{l_0,\omega} - \tilde{\mathbf{e}}^{l_0}_{0,2} = \mathbf{w}_{0,2} \leq \delta_{w,2}.$$

Thus, in this case, $\Delta \mathbf{x}^{l_0} = -2\frac{\delta_{w,1}}{c_1}$ making the state estimation error bound from (13) tight (i.e., equal to the bound from (13)). Also, it is worth noting that the estimated attack signal $\tilde{\mathbf{e}}^{l_0}_{0,3}$ satisfies that

$$|\tilde{\mathbf{e}}^{l_0}_{0,3}| \leq 2\delta_{w,3} + 2\frac{c_3}{c_1}\delta_{w,1} = D_3^{\tilde{\mathbf{e}}^{l_0}},$$

for $D_3^{\tilde{\mathbf{e}}^{l_0}}$ defined as in (36). Thus, even if computed $\tilde{\mathbf{e}}^{l_0}_{0,3}$ is nonzero, the attack identification method from (37) does not misclassify sensor s_3 as compromised (i.e., $Attacked^{l_0}(s_3) = 0$). Finally, a similar approach can be used to generate examples with simple second order plant dynamics for which tight estimation bounds will be achieved.

Note that the obtained bounds from Theorem 1 are only slightly more conservative then the bounds obtained using the procedure we introduced in [24]; for instance, for the systems

TABLE I
Percentage of 1000 Randomly Generated Non-Dynamical Systems with $q = 1$ Compromised Sensor that Satisfy the Condition from (34); Here, $q_{\max} = \lfloor \frac{p-n}{2} \rfloor$ is the Maximal Number of Compromised Sensors for Which the State can be Recovered with l_0 Estimator

	Max. number of attacked sensors for which estimation is possible	% of systems satisfying (34)
$p = 9, n = 1$	$q_{\max} = 4$	100%
$p = 9, n = 2$	$q_{\max} = 3$	70.4%
$p = 9, n = 3$	$q_{\max} = 3$	17.8%
$p = 9, n = 4$	$q_{\max} = 2$	0.4%

with $n = 10$ states, the maximal relative error was 20%. On the other hand, the complexity of the error bounding algorithm from [24] limits its use to systems with smaller number of states, while the bound from Theorem 1 can be computed for systems with $n = 20$ states and $p = 11$ sensors in seconds.

Finally, we evaluate how often the condition from (34) is satisfied, which allows for the use of the very computationally efficient bounding procedure (30) for the l_1-based state estimators. We considered random systems with $p = 9$ sensors and a different number of states n. For each type of systems (i.e., n and p) we randomly generated 1000 systems and for each system we checked if there exists $\alpha > 0$ such that (34) is satisfied. Table I summarizes the results; for reference, in the table we also specify the value for $q_{\max}$—the maximal number of compromised sensors for which the state can be always recovered for systems without noise and any (optimal) state estimator (e.g., l_0-based estimator). As can be seen, with the increase in the number of states, the percentage of systems for which the bound (30) can be used significantly decreases. This has been caused by the fact that with the increase in n, the values for $q_{\max}$ decrease, meaning that even l_0-based estimator can deal with a lower number of compromised sensors. Furthermore, for these systems the worst-case recorded relative error was $2 \cdot 10^{-4}$. As described in Section V, the reason for the heavy overapproximations are the use of l_1-based convexification and the conversions between the norms. While this paper provides the first analytic bound for l_1-based state estimation and shows that the worst-case error is linear with the size of the noise, an avenue for future work will be to provide a tighter bound for the convex-optimization based approach to attack-resilient state estimation.

VIII. Conclusion

In this paper, we have considered the problem of state estimation when some of the sensors are attacked by a malicious attacker. Unlike existing work on this topic, we have investigated the case when there is bounded-size noise in the system's dynamics. We have shown how to use two estimators that incorporate noise allowance in its constraints (i.e., $P_{0,\omega}$ and $P_{1,\omega}$) and proved that the worst-case state estimation error is linear with the size of the noise present in the system. The provided bounds illustrate that l_0 based state estimation results in significantly more accurate state estimation. However, the penalty is paid in the complexity of the procedure—$P_{0,\omega}$ can be solved as a mixed integer linear program, which are NP hard in general, while $P_{1,\omega}$ can be efficiently solved using standard convex solvers and is more suited for embedded control applications.

Finally, we have derived attack identification procedures, based on these estimators, that exploit the fact that besides state estimates, estimations of attack vectors are also provided. We have shown that the proposed attack identification schemes are sound, and derived conditions on signals injected via an attacked sensor that would guarantee identification of the compromised sensor. An avenue for future work would be to determine conditions when the support of estimated attack vectors is a subset of the set of attacked vectors.

References

[1] A. Cardenas, S. Amin, and S. S. Sastry, "Research challenges for the security of control systems," in *Proc. 3rd USENIX Workshop Hot Topics in Security*, 2008, Article 6.

[2] J. Slay and M. Miller, "Lessons learned from the maroochy water breach," in *Critical Infrastructure Protection*. E. Goetz and S. Shenoi, Eds. Springer US, 2008, pp. 73–82.

[3] R. Langner, "Stuxnet: Dissecting a cyberwarfare weapon," *IEEE Security & Privacy*, vol. 9, no. 3, pp. 49–51, 2011.

[4] T. M. Chen and S. Abu-Nimeh, "Lessons from stuxnet," *Computer*, vol. 44, no. 4, pp. 91–93, 2011.

[5] K. Koscher, A. Czeskis, F. Roesner, S. Patel, T. Kohno, S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham, and S. Savage, "Experimental security analysis of a modern automobile," in *Proc. IEEE Symp. Security Privacy*, 2010, pp. 447–462.

[6] S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham, S. Savage, K. Koscher, A. Czeskis, F. Roesner, and T. Kohno, "Comprehensive experimental analyses of automotive attack surfaces," in *Proc. USENIX Security Symp.*, 2011.

[7] A. Greenberg, Hackers remotely kill a jeep on the highway. 2015. [Online]. Available: http://www.wired.com/2015/07/hackers-remotely-kill-jeep-highway/

[8] S. Peterson and P. Faramarzi, "Iran hijacked US drone, says Iranian engineer," *Christian Science Monitor*, Dec. 11, 2011.

[9] D. Shepard, J. Bhatti, and T. Humphreys, "Drone hack: Spoofing attack demonstration on a civilian unmanned aerial vehicle," *GPS World*, vol. 23, no. 8, pp. 30–33, 2012.

[10] A. Teixeira, D. Pérez, H. Sandberg, and K. H. Johansson, "Attack models and scenarios for networked control systems," in *Proc. 1st Int. Conf. High Confidence Netw. Syst., ser. HiCoNS'12*, 2012, pp. 55–64.

[11] J. S. Warner and R. G. Johnston, "A simple demonstration that the global positioning system (GPS) is vulnerable to spoofing," *J. Security Admin.*, vol. 25, no. 2, pp. 19–27, 2002.

[12] N. O. Tippenhauer, C. Pöpper, K. B. Rasmussen, and S. Capkun, "On the requirements for successful GPS spoofing attacks," in *Proc. 18th ACM Conf. Comput. Commun. Security*, 2011, pp. 75–86.

[13] Y. Shoukry, P. Martin, P. Tabuada, and M. Srivastava, "Non-invasive spoofing attacks for anti-lock braking systems," in *Cryptographic Hardware and Embedded Systems-CHES 2013*. Springer, 2013, pp. 55–72.

[14] R. Smith, "A decoupled feedback structure for covertly appropriating networked control systems," *Proc. IFAC World Congr.*, pp. 90–95, 2011.

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

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

[17] S. Sundaram, M. Pajic, C. Hadjicostis, R. Mangharam, and G. Pappas, "The wireless control network: Monitoring for malicious behavior," in *Proc. 49th IEEE Conf. Dec. Control*, Dec. 2010, pp. 5979–5984.

[18] F. Miao, M. Pajic, and G. Pappas, "Stochastic game approach for replay attack detection," in *Proc. 52nd IEEE Annu. Conf. Dec. Control*, Dec. 2013, pp. 1854–1859.

[19] Y. Mo, R. Chabukswar, and B. Sinopoli, "Detecting integrity attacks on scada systems," *IEEE Trans. Control Syst. Technol.*, vol. 22, no. 4, pp. 1396–1407, 2014.

[20] Y. Mo, S. Weerakkody, and B. Sinopoli, "Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs," *IEEE Control Syst.*, vol. 35, no. 1, pp. 93–109, 2015.

[21] Y. Mo, T.-H. Kim, K. Brancik, D. Dickinson, H. Lee, A. Perrig, and B. Sinopoli, "Cyber-physical security of a smart grid infrastructure," *Proc. IEEE*, vol. 100, no. 1, pp. 195–209, 2012.

[22] C. Kwon, W. Liu, and I. Hwang, "Security analysis for cyber-physical systems against stealthy deception attacks," in *Proc. Amer. Control Conf.* IEEE, Jul. 2013, pp. 3344–3349.

[23] Y. Shoukry, A. Puggelli, P. Nuzzo, A. L. Sangiovanni-Vincentelli, S. A. Seshia, and P. Tabuada, "Sound and complete state estimation for linear dynamical systems under sensor attacks using satisfiability modulo theory solving," in *Proc. Amer. Control Conf.*, Jul. 2015, pp. 3818–3823.

[24] M. Pajic, J. Weimer, N. Bezzo, P. Tabuada, O. Sokolsky, I. Lee, and G. Pappas, "Robustness of attack-resilient state estimators," in *Proc. ACM/IEEE Int. Conf. Cyber-Physical Syst.*, Apr. 2014, pp. 163–174.

[25] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok, "Introduction to compressed sensing," in *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012, pp. 1–68.

[26] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, 2006.

[27] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030–1051, 2006.

[28] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, 2010.

[29] R. Foygel and L. Mackey, "Corrupted sensing: Novel guarantees for separating structured signals," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 1223–1247, 2014.

[30] C. Studer and R. G. Baraniuk, "Stable restoration and separation of approximately sparse signals," *Appl. Comput. Harmon. Anal.*, vol. 37, no. 1, pp. 12–35, 2014.

[31] M. Pajic, P. Tabuada, I. Lee, and G. Pappas, "Attack-resilient state estimation in the presence of noise," in *Proc. 54th IEEE Annu. Conf. Dec. Control*, Dec. 2015, pp. 5827–5832.

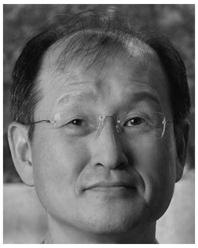
[32] Y. Shoukry and P. Tabuada, "Event-triggered state observers for sparse sensor noise/attacks," *IEEE Trans. Autom. Control*, vol. PP, no. 99, 2015.

[33] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on the smart grid," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 645–658, 2011.

Miroslav Pajic (S'06–M'13) received the Dipl. Ing. and M.S. degrees in electrical engineering from the University of Belgrade, Serbia, in 2003 and 2007, respectively, and the M.S. and Ph.D. degrees in electrical engineering from the University of Pennsylvania, Philadelphia, PA, USA, in 2010 and 2012, respectively.

Currently, he is an Assistant Professor in the Department of Electrical and Computer Engineering at Duke University, Durham, NC, USA. He also holds a secondary appointment in the Computer Science Department. Prior to joining Duke, he was a Postdoctoral Researcher in the PRECISE Center, University of Pennsylvania. His research interests focus on the design and analysis of cyberphysical systems and, in particular, real-time and embedded systems, distributed/networked control systems, and high-confidence medical devices and systems.

Dr. Pajic received various awards, including the 2011 ACM SIGBED Frank Anger Memorial Award, the Joseph and Rosaline Wolf Award for Best Electrical and Systems Engineering Dissertation from Penn, the Best Paper Award at the 2014 ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS), and the Best Student Paper award at the 2012 IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS).



Insup Lee (F'01) received the B.S. degree in mathematics (Hons.) from the University of North Carolina, Chapel Hill, NC, USA, and the Ph.D. degree in computer science from the University of Wisconsin, Madison, WI, USA.

Currently, he is Cecilia Fitler Moore Professor of Computer and Information Science and Director of PRECISE Center at the University of Pennsylvania, University Park, PA, USA. He also holds a secondary appointment in the Department of Electrical and Systems Engineering. His research interests include cyberphysical systems (CPS), real-time systems, embedded systems, high-confidence medical device systems, formal methods and tools, run-time verification, software certification, and trust management. The theme of his research activities has been to ensure and improve the correctness, safety, and timeliness of life-critical embedded systems. Recently, he has been working in medical CPS and the security of CPS.

Prof. Lee has served on many program committees, chaired many international conferences and workshops, and served on various steering and advisory committees of technical societies. He has also served on the editorial boards of several scientific journals, including Journal of ACM, ACM Transactions on Cyber-Physical Systems, IEEE Transactions on Computers, Formal Methods in System Design, and Real-Time Systems Journal. He is Chair of ACM SIGBED (2015–2017) and was Chair of IEEE TC-RTS (2003–2004). He was a member of the Technical Advisory Group (TAG) of President's Council of Advisors on Science and Technology (PCAST) Networking and Information Technology (2006–2007). He is a member of the National Research Council's committee on 21st Century Cyber-Physical Systems Education (2014–2015). He received an appreciation award from Ministry of Science, IT and Future Planning, South Korea in 2013. He received the IEEE TC-RTS Outstanding Technical Achievement and Leadership Award in 2008. His papers received the five best paper awards at conferences.

George J. Pappas (S'90–M'91–SM'04–F'09) received the Ph.D. degree in electrical engineering and computer sciences from the University of California, Berkeley, CA, USA, in 1998.

He is currently the Joseph Moore Professor and Chair of the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA, USA. He also holds a secondary appointment with the Department of Computer and Information Sciences and the Department of Mechanical Engineering and Applied Mechanics. He had previously served as the Deputy Dean for Research with the School of Engineering and Applied Science. His research interests include control theory and, in particular, hybrid systems, embedded systems, cyberphysical systems, and hierarchical and distributed control systems, with applications to unmanned aerial vehicles, distributed robotics, green buildings, and biomolecular networks.

Dr. Pappas has received various awards, such as the Antonio Ruberti Young Researcher Prize, the George S. Axelby Award, the Hugo Schuck Best Paper Award, the George H. Heilmeier Award, the National Science Foundation PECASE award, and numerous best student papers awards at ACC, CDC, and ICCPS. He is a member of the GRASP Lab and the PRECISE Center.