Multi-Model Resilient Observer under False Data Injection Attacks

Olugbenga Moses Anubi, Charalambos Konstantinou, Carlos A. Wong, Satish Vedula
Department of Electrical and Computer Engineering, FAMU-FSU College of Engineering
Center for Advanced Power Systems, Florida State University
E-mail: {oanubi, ckonstantinou, cwong, svedula}@fsu.edu

***Abstract*— In this paper, we present the concept of boosting the resiliency of optimization-based observers for cyber-physical systems (CPS) using auxiliary sources of information. Due to the tight coupling of physics, communication and computation, a malicious agent can exploit multiple inherent vulnerabilities in order to inject stealthy signals into the measurement process. The problem setting considers the scenario in which an attacker strategically corrupts portions of the data in order to force wrong state estimates which could have catastrophic consequences. The goal of the proposed observer is to compute the true states in-spite of the adversarial corruption. In the formulation, we use a measurement prior distribution generated by the auxiliary model to refine the feasible region of a traditional compressive sensing-based regression problem. A constrained optimization-based observer is developed using l_1-minimization scheme. Numerical experiments show that the solution of the resulting problem recovers the true states of the system. The developed algorithm is evaluated through a numerical simulation example of the IEEE 14-bus system.**

***Index Terms*— Resiliency, observer, cyber-physical systems, false data injection attacks.**

I. Introduction

Cyber-physical systems (CPS) are engineered systems that are built from, and depend upon, the seamless integration of cyber and physical components. Hence, CPS are tightly integrated systems at all scales and levels that leverage information, communication, and computing systems to control a physical process in an autonomous, cooperative, intelligent, and flexible manner [1]. The decreasing cost of sensing, networking, and computation tools in the era of internet-of-things (IoT) has resulted in building complex CPS with new capabilities, reducing the cost of CPS operation, and having safer and more efficient systems.

Many CPS applications are safety-critical systems in domains such as critical infrastructure (e.g., power grid systems), disaster monitoring, and healthcare environments. Therefore, it is of paramount importance to ensure overall stability of the physical process and avoid severe consequences. Towards that goal of maintaining normal operating conditions, a CPS is consistently monitored and controlled by data acquisition and control systems. CPS operators use measurements acquired from various sensors across the CPS infrastructure to estimate system state variables. These state estimates are critical since they are used to adjust the control of the physical space via management operations.

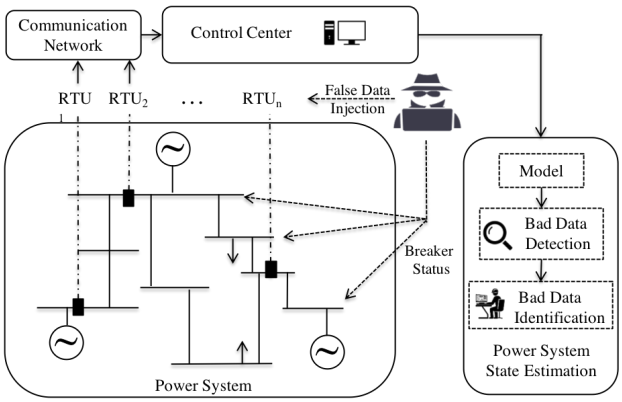

Fig. 1. State estimation under false data injection attacks (FDIAs).

In order to preserve the integrity and availability of state estimation routines in CPS-related applications, *bad data detection (BDD)* mechanisms have been traditionally used to remove faulty and erroneous measurements [2]. However, recent studies have showed that judiciously falsified data can inject errors in state variables without being detected by BDD [3]–[5]. Adversaries may launch such *false data injection attacks (FDIAs)* able to bypass BDD functions by altering the measurements sent from the field sensing devices to the central estimation station [6]. Furthermore, attackers may realize such FDIAs by hacking into sensors and meters or even infiltrate secondary channels of the supply chain in order to distort the measurements [7], [8]. Fig. 1 presents a schematic of the state estimation routine under FDIA.

Existing efforts to address the vulnerability of state estimation algorithms to FDIAs either require protection of a set of measurement sensors or verification of each state variable independently. The high computational and deployment cost, as well as significant risk involved with these approaches, have hampered their feasibility for use in practical real-time systems [4], [9]. Furthermore, existing approaches are often developed for specific system configurations [10]. As a result, it is necessary to investigate more computationally feasible, adaptive and real-time implantable resiliency methods.

In this work, we present an enhanced resilient state estimation approach for a dynamic CPS in which the data

acquired from the sensing devices are poisoned with FDIAs. Our method relies on a data-driven model with traditional compressive sensing regression. Gaussian processes (GP) are a typical candidate for building generative probabilistic regression models from historical data [11]. We demonstrate that our solution can recover the true states of the system, i.e., the system operation is able to withstand, adapt, and detect efficiently extreme adversarial FDIA settings. The developed algorithm is evaluated on a power system test case model.

The reminder of the paper is organized as follows: in Section III we provide necessary definitions and background for this work. Then, Section IV presents the formulation of the estimation problem as well as our proposed solution algorithm for the enhanced state estimator. Experimental details and simulation results are described in Section V. Our concluding remarks are discussed in Section VI.

II. Notation

The following notions and conventions are employed throughout the paper: $\mathbb{R}, \mathbb{R}^n, \mathbb{R}^{n\times m}$ denote the space of real numbers, real vectors of length n and real matrices of n rows and m columns respectively. $\mathbb{R}_+$ denotes positive real numbers. $X^\top$ denotes the transpose of the quantity X. By $Q \succeq 0$, it is meant that Q is a positive semi-definite symmetric matrix, i.e $\mathbf{x}^\top Q\mathbf{x} \geq 0\ \forall \mathbf{x} \neq 0$ and $Q \succ 0$ denotes positive definiteness which is defined with strict $>$ instead. Given $Q \succ 0$, the Q-weighted norm is defined as $\|\mathbf{x}\|_Q \triangleq \mathbf{x}^\top Q\mathbf{x}$. Normal-face lower-case letters ($x \in \mathbb{R}$) are used to represent real scalars, bold-face lower-case letter ($\mathbf{x} \in \mathbb{R}^n$) represents vectors, while normal-face upper case ($X \in \mathbb{R}^{n\times m}$) represents matrices. Let $\mathcal{T} \subseteq \{1, \ldots, n\}$ then, for a matrix $X \in \mathbb{R}^{n\times m}$, $X_\mathcal{T} \in \mathbb{R}^{|\mathcal{T}|\times m}$ is the submatrix obtained by extracting the rows of X corresponding to the indices in $\mathcal{T}$. For a vector $\mathbf{x}$, $\mathbf{x}_i$ denotes its i_{th} element. The support of a vector $\mathbf{x} \in \mathbb{R}^m$ is denoted by $\mathsf{supp}(\mathbf{x}) \triangleq \{i\ :\ \mathbf{x}_i \neq 0\}$, with $|\mathsf{supp}(\mathbf{x})| \leq m$ being the number of nonzero elements of $\mathbf{x}$. $\mathcal{S}_s^m \triangleq \{\mathbf{x} \in \mathbb{R}^m\}\ :\ 0 < |\mathsf{supp}(\mathbf{x})| \leq s\}$ denotes the set of all nonzero k-sparse vectors. Given a positive scalar $\varepsilon \in \mathbb{R}_+$, a saturation function $\mathsf{sat}_\varepsilon : \mathbb{R} \mapsto [-\varepsilon,\ \varepsilon]$ is given by

$$\mathsf{sat}_\varepsilon(x) = \begin{cases} -\varepsilon & \text{if} \quad x < -\varepsilon \\ x & \text{if} \quad |x| \leq \varepsilon \\ \varepsilon & \text{if} \quad x > \varepsilon \end{cases}$$

A best s_{th} term approximation of a vector $\mathbf{e} \in \mathbb{R}^m$ ($s \leq m$) is denoted by $\mathbf{e}[s] \triangleq \min\limits_{\|\mathbf{f}\|_0 = s} \|\mathbf{e} - \mathbf{f}\|_1$.

III. Preliminaries

In this section, for completeness of exposition and to facilitate faster comprehension of subsequent developments, we have gathered relevant results from literature that we built upon.

A. Overview of Resilient Estimators

There are numerous work in literature on the secure estimation for CPS [12]–[19]. The majority of the previous research focuses on the LTI systems ranging from a Kalman filter as predictor and estimator, unconstrained l_1-minimization to solve the error problem, and the use of machine learning paradigms for feature discovery. However, we focus only on the ones which are optimization based - since that is the approach we consider in this work. Moreover, due to sparsity assumption on the set of attacked nodes, majority of these works are based on the classical error correction problem [20]. Let $\mathbb{R}^m \ni \mathbf{y} = C\mathbf{x} + \mathbf{e}$, where $C \in \mathbb{R}^{m\times n}$ is a coding matrix $(m > n)$, be a measurement vector corrupted by an arbitrary unknown but sparse error vector $\mathbf{e}$. By sparsity, we mean that $\|\mathbf{e}\|_{l_0} \leq s < m$. The objective is to recover the input vector $\mathbf{x} \in \mathbb{R}^n$. Assuming that the coding matrix C is full rank, one can construct a matrix F such that $FC = 0$ and

$$\tilde{\mathbf{y}} = F\mathbf{y} = F(C\mathbf{x} + \mathbf{e}) = F\mathbf{e}. \tag{1}$$

Thus the decoding problem is equivalent to reconstructing a sparse vector from the observation $\tilde{\mathbf{y}} = F\mathbf{e}$ and is cast as the compressive sensing problem:

$$\underset{\mathbf{e}}{\text{Minimize:}} \|\mathbf{e}\|_{l_0} \quad \text{Subject to:} \quad \tilde{\mathbf{y}} = F\mathbf{e}. \tag{2}$$

Hayden et. al [21] obtained a sufficient condition that if all subsets of $2s$ columns of F are full rank, then any error $\|\mathbf{e}\|_{l_0} \leq s$ can be reconstructed uniquely by the solution of the optimization problem in (2). Although in some cases [22] the optimization problem in (2) is solved as is, in most cases, it does not lend itself to a solution in polynomial time due to its nonconvexity. As a result, it is often replaced with its convex neighbor:

$$\underset{\mathbf{e}}{\text{Minimize:}} \|\mathbf{e}\|_{l_1} \quad \text{Subject to:} \quad \tilde{\mathbf{y}} = F\mathbf{e}. \tag{3}$$

The two programs, however, have been shown to be equivalent under the condition that the *restricted isometric property (RIP)* holds [23]–[26].

Definition 1 (RIP [20])**.** *A matrix A has the RIP of sparsity k if there exists $0 < \delta < 1$ such that*

$$(1 - \delta)\|\mathbf{x}\|_2^2 \leq \|A\mathbf{x}\|_2^2 \leq (1 + \delta)\|\mathbf{x}\|_2^2 \tag{4}$$

for all $\mathbf{x} \in \mathcal{S}_s$. Moreover, the smallest δ for which the above inequality holds is called the restricted isometry constant*, and denoted as $\delta_s(A)$.*

The above definition essentially requires that every set of columns with cardinality less that or equal to s behaves like an orthonormal system. The following theorem lists the recovery error due to relaxed convex program above.

Theorem 1 ([20], [27])**.** *Let $\mathbf{e}$ be a sparse vector satisfying $\tilde{\mathbf{y}} = F\mathbf{e}$ and $\hat{\mathbf{e}}$ be the solution of* (3)*. If $\delta_{2s}(F) < \dfrac{1}{\sqrt{2}}$, then*

$$\|\hat{\mathbf{e}} - \mathbf{e}\|_2 \leq \frac{2}{\sqrt{s}}\left(\frac{\delta_{2s} + \sqrt{\delta_{2s}\left(\frac{1}{\sqrt{2}} - \delta_{2s}\right)}}{\sqrt{2}\left(\frac{1}{\sqrt{2}} - \delta_{2s}\right)} + 1\right)\|\mathbf{e} - \mathbf{e}[s]\|_1, \tag{5}$$

where $\mathbf{e}[s]$ *is the best* s*-term approximation of* $\mathbf{e}$*.*

Remark 1. *If* $\mathbf{e} \in \mathcal{S}_s$*, then* $\hat{\mathbf{e}} = \mathbf{e}$*. Thus, if* $\delta_{2s}(F) < \frac{1}{\sqrt{2}}$ *the relaxed program in* (3) *will recover any* s*-sparse vector* $\mathbf{e} \in \mathcal{S}_s$ *exactly!*

Now, consider the discrete LTI system

$$\mathbf{x}_{k+1} = A\mathbf{x}_k \tag{6}$$

$$\mathbf{y}_k = C\mathbf{x}_k + \mathbf{e}_k, \tag{7}$$

where $\mathbf{x}_k \in \mathbb{R}^n$ represents the state of the system at time $k \in \mathbb{N}$, $\mathbf{y}_k \in \mathbb{R}^m$ is the output of the monitoring nodes at time k and $\mathbf{e}_k \in \mathbb{R}^m$ denote the attack signals injected by malicious agents at the monitoring nodes. Let $\mathcal{K} \subset \{1, 2, \ldots, m\}$ denote the set of attacked nodes, then for all k, $|\mathsf{supp}(\mathbf{e}_k)| \subset \mathcal{K}$. The resilient estimation problem is then defined as reconstructing the initial state $\mathbf{x}_0$ from corrupt measurement $\{\mathbf{y}_k\}_{k=0}^{T}, T \in \mathbb{N}$. We look at two scenarios from literature: $\mathcal{K}$ is time-invariant [12], [28] and $\mathcal{K}$ is time-varying [13].

1) Secure estimation for fixed attacked nodes [12]: Assuming that the set $\mathcal{K}$ of attacked nodes is time-invariant:

Definition 2. s *errors are correctable after* T *steps by the decoder* $\mathcal{D} : (\mathbb{R}^m)^T \mapsto \mathbb{R}^n$ *if for any* $\mathbf{x}_0 \in \mathbb{R}^n$*, any* $\mathcal{K} \subset \{1, 2, \ldots, m\}$ *with* $|\mathcal{K}| \leq s$*, and any sequence of vectors* $\mathbf{e}_0, \ldots, \mathbf{e}_{T-1} \in \mathbf{R}^m$ *such that* $\mathsf{supp}(\mathbf{e}_k) \subset \mathcal{K}$*, we have* $\mathcal{D}(\mathbf{y}_0, \ldots, \mathbf{y}_{T-1}) = \mathbf{x}_0$*, where* $\mathbf{y}_k = CA^k\mathbf{x}_0 + \mathbf{e}_k$ *for* $k = 0, 1, \ldots, T-1$*.*

Proposition 1. *Let* $T \in \mathbb{N}\backslash\{0\}$*. The following are equivalent:*
(i) There is a decoder that can correct q *errors after* T *steps;*
(ii) For all $\mathbf{z} \in \mathbb{R}^n\backslash\{0\}$*,* $|\mathsf{supp}(C\mathbf{z}) \cup \mathsf{supp}(CA\mathbf{z}) \cup \cdots \cup \mathsf{supp}(CA^{T-1}\mathbf{z})| > 2s$*.*

Consequently, the following optimal decoder is defined for when the set of attacked nodes is fixed:

$$\mathbf{x}_0 = \arg\min_{\mathbf{x}} \|Y_T - \Phi_T(\mathbf{x})\|_{l_0} \tag{8}$$

where

$$Y_T = \left[\ \mathbf{y}_0 \ \middle|\ \mathbf{y}_1 \ \middle|\ \ldots \ \middle|\ \mathbf{y}_{T-1}\ \right] \in \mathbb{R}^{m \times T}$$

and $\Phi_T : \mathbb{R}^n \mapsto \mathbb{R}^{m \times T}$ is a linear map given by:

$$\Phi_T(\mathbf{x}) = \left[\ C\mathbf{x} \ \middle|\ CA\mathbf{x} \ \middle|\ \ldots \ \middle|\ CA^{T-1}\mathbf{x}\ \right] \in \mathbb{R}^{m \times T}.$$

2) Secure estimation for varying attacked nodes [13]: Assuming that the set $\mathcal{K}$ of attacked nodes can change with time but bounded as in $|\mathcal{K}| \leq s$:

Definition 3. q *errors are correctable after* T *steps by the decoder* $\mathcal{D} : (\mathbb{R}^m)^T \mapsto \mathbb{R}^n$ *if for any* $\mathbf{x}_0 \in \mathbb{R}^n$ *and any sequence of vectors* $\mathbf{e}_0, \ldots, \mathbf{e}_{T-1} \in \mathbf{R}^m$ *such that* $|\mathsf{supp}(\mathbf{e}_k)| \leq s$*, we have* $\mathcal{D}(\mathbf{y}_0, \ldots, \mathbf{y}_{T-1}) = \mathbf{x}_0$*, where* $\mathbf{y}_k = CA^k\mathbf{x}_0 + \mathbf{e}_k$ *for* $k = 0, 1, \ldots, T-1$*.*

Proposition 2. *Let* $T \in \mathbb{N}\backslash\{0\}$*. The following are equivalent:*
(i) There is a decoder that can correct s *errors after* T *steps;*
(ii) For all $\mathbf{z} \in \mathbb{R}^n\backslash\{0\}$ *,* $\sum_{k=0}^{T-1} \left|\mathsf{supp}(CA^k\mathbf{z})\right| > 2s$*.*

Consequently, the following optimal decoder is defined for when the set of attacked nodes is not fixed:

$$\mathbf{x}_0 = \arg\min_{\mathbf{x}} \left\|\mathbf{y}_{(T)} - \Phi_{(T)}\mathbf{x}\right\|_{l_0} \tag{9}$$

where

$$\mathbf{y}_{(T)} = \begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_{T-1} \end{bmatrix} \in \mathbb{R}^{mT},$$

$$\Phi_{(T)} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{T-1} \end{bmatrix} \in \mathbb{R}^{mT \times n}.$$

IV. Resilient Observer Development

Consider the concurrent models

$$\begin{aligned} \mathbf{x}_{k+1} &= A\mathbf{x}_k + B\mathbf{u}_k \\ \mathbf{y}_k &= C\mathbf{x}_k + \mathbf{e}_k \end{aligned} \tag{10}$$

$$\mathbf{y}_k \sim \mathcal{N}(\boldsymbol{\mu}(\mathbf{z}_k), \Sigma(\mathbf{z}_k)) \tag{11}$$

consisting of a physics-based model (10) and a data-driven prior (11) given as a function of the auxiliary variable $\mathbf{z} \in \mathbb{R}^p$. The data-driven model in (11) gives a prior distribution on the system measurements as a function of measured auxiliary variables $\mathbf{z}_k \in \mathbb{R}^p$. This provides additional layer of security by: *1)* requiring the attacker to have knowledge of the auxiliary model and the parameters, and *2)* limiting the magnitude of possible state corruption. For a more detailed explanation of the advantages of the concurrent models in (10) and (11), as well as the resulting theoretical limits on the size of feasible attacks, interested readers are referred to the references [11], [29] and the references therein.

Let $Y_k \triangleq \{\mathbf{y}_k, \mathbf{y}_{k-1}, \ldots, \mathbf{y}_{k-T+1}\} \subset \mathbb{R}^m$ and $U_{k-1} \triangleq \{\mathbf{u}_{k-1}, \mathbf{u}_{k-2}, \ldots, \mathbf{u}_{k-T+1}\} \subset \mathbb{R}^l$ be collections of the last T-samples of the system known input and output measurements respectively. The proposed resilient observer attempts to solve the following moving horizon optimization problem for all time instant $k \geq T$:

$$\begin{aligned} \text{Minimize:} \quad & \sum_{i=k-T+1}^{k} \|\mathbf{y}_i - C\mathbf{x}_i\|_{l_0} \\ \text{Subject to:} \quad & \\ & \mathbf{x}_{i+1} - A\mathbf{x}_i - B\mathbf{u}_i = 0, \\ & \qquad i = k-T+1, \ldots, k-1 \\ & C\mathbf{x}_k \in \mathcal{Y}(\mathbf{z}_k) \end{aligned} \tag{12}$$

where the convex set $\mathcal{Y}(\mathbf{z})$ has the property that:

$$p(\mathbf{y_k}^* \in \mathcal{Y} | \mathbf{z}_k, \mathcal{D}) \geq \tau. \tag{13}$$

More insight is provided in Theorem 2. The idea is essentially seeking historical and current state vectors, together with the minimum attacked channels, which completely explains the observations while satisfying the physics-based model and having a high likelihood according to the auxiliary model prior. The optimization parameter $\tau \in (0,\ 1]$ controls the likelihood threshold. It can be set to a constant value or optimized with respect to some higher-level objectives. Thus, the resilient observer optimization problem is equivalent to:

$$\begin{aligned} \text{Minimize:} \quad & \sum_{i=k-T+1}^{k} \|\mathbf{y}_i - C\mathbf{x}_i\|_{\ell_0} \\ \text{Subject to:} \quad & \mathbf{x}_{i+1} - A\mathbf{x}_i - B\mathbf{u}_i = 0, \\ & \|C\mathbf{x}_k - \boldsymbol{\mu}(\mathbf{z}_k)\|^2_{\Sigma^{-1}(\mathbf{z})} \leq \chi^2_m(\tau) \end{aligned} \tag{14}$$

where $\chi^2_m(\tau)$ is the quantile function for probability τ of the chi-squared distribution with m degrees of freedom.

However, the nonconvexity due to the index minimization objective makes the optimization problem in (14) challenging, at best, for gradient-based solution algorithms. This will make it difficult, if not impossible, to synthesize a pragmatic algorithm that can be implemented real-time for the observer. Thus, we seek convex approximation alternatives. Fortunately, as discussed in the preliminaries Section III, it is possible to approximate the index minimization objective using an ℓ_1-norm without loosing global optimality – provided the RIP condition holds. Consequently, the proposed resilient multi-model observer is given via the following convex program:

$$\begin{aligned} \text{Minimize:} \quad & \sum_{i=k-T+1}^{k} \|\mathbf{y}_i - C\mathbf{x}_i\|_{\ell_1} \\ \text{Subject to:} \quad & \mathbf{x}_{i+1} - A\mathbf{x}_i - B\mathbf{u}_i = 0, \\ & \|C\mathbf{x}_k - \boldsymbol{\mu}(\mathbf{z}_k)\|^2_{\Sigma^{-1}(\mathbf{z})} \leq \chi^2_m(\tau). \end{aligned} \tag{15}$$

After some algebraic manipulations and simplifications, the above program is equivalent to the following quadratically constrained basis pursuit problem:

$$\begin{aligned} \text{Minimize:} \quad & \left\|\mathbf{y}_{(T)} - H_{(T)}\mathbf{u}_{(T-1)} - \Phi_{(T)}\mathbf{x}\right\|_1 \\ \text{Subject to:} \quad & \left\|\Phi_T\mathbf{x} + H_T\mathbf{u}_{(T-1)} - \boldsymbol{\mu}(\mathbf{z_k})\right\|^2_{\Sigma^{-1}(\mathbf{z}_k)} \leq \chi^2_m(\tau), \end{aligned} \tag{16}$$

where

$$\mathbf{y}_{(T)} = \begin{bmatrix} \mathbf{y}_{k-T+1} \\ \mathbf{y}_{k-T+2} \\ \vdots \\ \mathbf{y}_k \end{bmatrix} \in \mathbb{R}^{mT},$$

$$\mathbf{u}_{(T-1)} = \begin{bmatrix} \mathbf{u}_{k-T+1} \\ \mathbf{u}_{k-T+2} \\ \vdots \\ \mathbf{u}_{k-1} \end{bmatrix} \in \mathbb{R}^{l(T-1)},$$

where $\Phi_{(T)}$ is defined as a results of (9),

$$H_{(T)} = \begin{bmatrix} 0 & 0 & \dots & 0 \\ CB & 0 & \dots & 0 \\ CAB & CB & \dots & 0 \\ \vdots & \vdots & & \vdots \\ CA^{T-2}B & CA^{T-3}B & \dots & CB \end{bmatrix} \in \mathbb{R}^{mT \times l(T-1)},$$

and Φ_T, H_T are the last m rows of $\Phi_{(T)}, H_{(T)}$ respectively. In the above formulation, the solution of the optimization problem gives an estimate $\hat{\mathbf{x}}_{k-T+1}$ of the state vector $\mathbf{x}_{k-T+1}$ which is then propagated forward to obtain an estimate of the current state using the physics-based dynamical model as follows

$$\hat{\mathbf{x}}_k = A^{T-1}\hat{\mathbf{x}}_{k-T+1} + G\mathbf{u}_{(T-1)}, \tag{17}$$

where

$$G = \begin{bmatrix} A^{T-2}B & A^{T-3}B & \dots & B \end{bmatrix}.$$

Suppose receding horizon T is chosen big enough (i.e $T \geq n$) and the pair (A, C) is observable, then there exists a matrix $F_{(T)}$ such that $F_{(T)}\Phi_{(T)} = 0$.[1] Consequently, the optimization problem above is equivalent to

$$\begin{aligned} \text{Minimize:} \quad & \|\mathbf{e}\|_1 \\ \text{Subject to:} \quad & \\ & \mathbf{f}_{(T)} = F_{(T)}\mathbf{e} \\ & \|\mathbf{y}_T + \mathbf{e}_T - \boldsymbol{\mu}(\mathbf{z_k})\|^2_{\Sigma^{-1}(\mathbf{z}_k)} \leq \chi^2_m(\tau), \end{aligned} \tag{18}$$

where

$$\mathbf{f}_{(T)} = F_{(T)}\left(\mathbf{y}_{(T)} - H\mathbf{u}_{(T-1)}\right),$$

and $\mathbf{e}_T, \mathbf{y}_T \in \mathbb{R}^m$ is the vector containing only the last m elements of the respective vectors $\mathbf{e}, \mathbf{y}$ in order. While the form given in (16) is more intuitive and implemented for the simulation results, the form in (18) is more suitable to proof the main result which is given next.

Theorem 2. *Given a dataset $\mathcal{D} = \{\mathbf{Z}, \mathbf{Y}\}$ containing historical auxiliary variables $\mathbf{Z} \in \mathbb{R}^{p\times T}$ and corresponding sensor measurements $\mathbf{Y} \in \mathbb{R}^{m\times T}$. Suppose that the latent sensor measurement satisfies the data-driven GPR prior given in* (11) *and that there exists $\tau \in (0,\ 1)$ such that the true measurement $\mathbf{y}_k^*$ satisfies $p(\mathbf{y_k}^*|\mathbf{z}_k, \mathcal{D}) \geq \tau$. Consider the convex optimization problem in* (16)*. Let $\hat{\mathbf{e}}$ be the solution*

[1] Let the singular value decomposition of $\Phi_{(T)}$ be given by

$$\Phi_{(T)} = \begin{bmatrix} U_1 \mid U_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \\ \hline & & 0 & \end{bmatrix} V^\top,$$

Then $F_{(T)} = U_2^\top$ is an example of such matrix.

of the equivalent form in (18)*. If* $\delta_{2s}(F_{(T)}) < \dfrac{1}{\sqrt{2}}$*, then, for any feasible sparse vector* $\mathbf{e}$*,*

$$\|\hat{\mathbf{e}}_T - \mathbf{e}_T\|_2 \leq K_1 \mathsf{sat}_1\left(K_2 \|\mathbf{e} - \mathbf{e}[s]\|_2\right), \tag{19}$$

where

$$K_1 = \sqrt{2\chi^2_m(\tau)\overline{\sigma}(\mathbf{z}_k)}$$
$$K_2 = K_3\sqrt{\frac{m-s}{2\chi^2_m(\tau)\overline{\sigma}(\mathbf{z}_k)}},$$

with

$$K_3 = \frac{2}{\sqrt{s}}\left(\frac{\delta_{2s} + \sqrt{\delta_{2s}\left(\frac{1}{\sqrt{2}} - \delta_{2s}\right)}}{\sqrt{2}\left(\frac{1}{\sqrt{2}} - \delta_{2s}\right)} + 1\right)$$

and $\overline{\sigma}(\mathbf{z}_k)$ *is the biggest singular value of* $\Sigma(\mathbf{z}_k)$*.*

Proof: Following the development in earlier part of this section, it is straightforward to see that $p(\mathbf{y_k}^*|\mathbf{z}_k, \mathcal{D}) \geq \tau$ implies that $\|\mathbf{y}_T + \mathbf{e}_T - \boldsymbol{\mu}(\mathbf{z_k})\|^2_{\Sigma^{-1}(\mathbf{z}_k)} \leq \chi^2_m(\tau)$ for any composite measurement vector $\mathbf{y}_T$ corrupted by the composite sparse signal $\mathbf{e}_T$. In order words, the inequality holds for all sparse signal $\mathbf{e}_T = \left[\mathbf{e}^\top_{k-T+1} \dots \mathbf{e}^\top_k\right]^\top$, with each $\mathbf{e}_k \in \mathcal{S}_s$ satisfying $\mathbf{y}_k = \mathbf{y}^*_k + \mathbf{e}_k$. More, there exists the *true state vector* $\mathbf{x}^*_k \in \mathbb{R}^n$ such that $\mathbf{y}^*_k = C\mathbf{x}^*_k$. This implies that

$$\begin{aligned}\mathbf{y}_{(T)} &= \mathbf{y}^*_{(T)} + \mathbf{e}_{(T)} \\ &= \Phi_{(T)}\mathbf{x}_{k-T+1} + H_{(T)}\mathbf{u}_{(T-1)} + \mathbf{e}_{(T)}.\end{aligned}$$

From which it follows that

$$\begin{aligned}\mathbf{f}_{(T)} &= F_{(T)}\left(\mathbf{y}_{(T)} - H\mathbf{u}_{(T-1)}\right) \\ &= F_{(T)}\mathbf{e}_{(T)}.\end{aligned}$$

Thus for any $\mathbf{y}_{(T)}$, the set of all $\mathbf{e}_{(T)}$ for which the quadratic inequality holds is a subset of the pre-image of $\mathbf{f}_{(T)}$ under the linear transformation $F_{(T)}$. Thus, using Lemma 1, the optimal point $\hat{\mathbf{e}}_{(T)}$ of the problem in (18) satisfies

$$\begin{aligned}\left\|\hat{\mathbf{e}}_{(T)} - \mathbf{e}_{(T)}\right\|_2 &\leq K_3\left\|\mathbf{e}_{(T)} - \mathbf{e}_{(T)}[s]\right\|_1 \\ &\leq K_3\sqrt{m-s}\left\|\mathbf{e}_{(T)} - \mathbf{e}_{(T)}[s]\right\|_2.\end{aligned}$$

Thus,

$$\begin{aligned}\|\hat{\mathbf{e}}_T - \mathbf{e}_T\|_2 &\leq \left\|\mathbf{e}_{(T)} - \mathbf{e}_{(T)}\right\|_2 \\ &\leq K_3\sqrt{m-s}\left\|\mathbf{e}_{(T)} - \mathbf{e}_{(T)}[s]\right\|_2 \end{aligned} \tag{20}$$

Moreover, adding and subtracting $\hat{\mathbf{e}}_T$ in the quadratic inequality constraint and using the left-hand-side triangular inequality yields the following sequence of inequalities;

$$\|\mathbf{y}_T + \hat{\mathbf{e}}_T - \boldsymbol{\mu}(\mathbf{z_k}) - \hat{\mathbf{e}}_T + \mathbf{e}_T\|^2_{\Sigma^{-1}(\mathbf{z}_k)} \leq \chi^2_m(\tau)$$

$$\|\hat{\mathbf{e}}_T - \mathbf{e}_T\|^2_{\Sigma^{-1}(\mathbf{z}_k)} - \|\mathbf{y}_T + \hat{\mathbf{e}}_T - \boldsymbol{\mu}(\mathbf{z_k})\|^2_{\Sigma^{-1}(\mathbf{z}_k)} \leq \chi^2_m(\tau)$$

$$\|\hat{\mathbf{e}}_T - \mathbf{e}_T\|^2_{\Sigma^{-1}(\mathbf{z}_k)} \leq 2\chi^2_m(\tau)$$

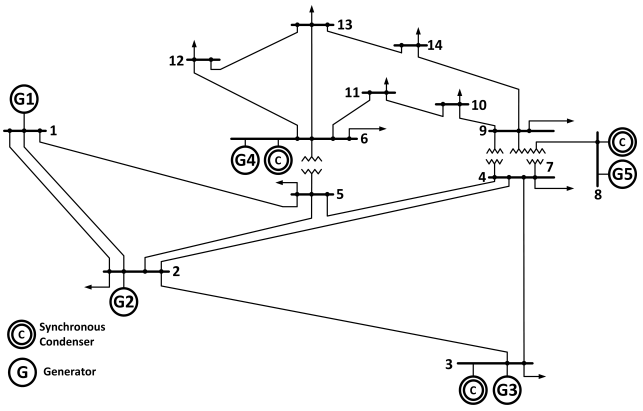

Fig. 2. IEEE 14-bus system.

which implies that

$$\|\hat{\mathbf{e}}_T - \mathbf{e}_T\|_2 \leq \sqrt{2\chi^2_m(\tau)\overline{\sigma}(\mathbf{z}_k)}. \tag{21}$$

Now, comparing (20) and (21) yields

$$\begin{aligned}\|\hat{\mathbf{e}}_T - \mathbf{e}_T\|_2 &\leq \min\left\{\begin{array}{c} K_3\sqrt{m-s}\left\|\mathbf{e}_{(T)} - \mathbf{e}_{(T)}[s]\right\|_2, \\ \\ \sqrt{2\chi^2_m(\tau)\overline{\sigma}(\mathbf{z}_k)}\end{array}\right\} \\ &\leq K_1 \min\left\{K_2\left\|\mathbf{e}_{(T)} - \mathbf{e}_{(T)}[s]\right\|_2, 1\right\} \\ &\leq K_1\mathsf{sat}_1\left(K_2\left\|\mathbf{e}_{(T)} - \mathbf{e}_{(T)}[s]\right\|_2\right)\end{aligned}$$

V. Simulation Results

The attack-resilient observer proposed in this paper is evaluated using a numerical simulation of the IEEE 14-bus system shown in Fig. 2, it has $n_b = 14$ buses and $n_g = 5$ generators. It is expected that each bus in the network has IIoT measurement devices able to provide active power injections and flow measurements.

A. Model Description

A small signal model is derived by linearizing the generator swing equations and power flow equations around the operating points under the assumption that:

- Voltage is tightly controlled at their nominal value;
- Angular difference between each bus is small;
- Conductance is negligible therefore the system is lossless.

Furthermore, the buses are ordered so that the first buses are generators, then the admittance-weighted *Laplacian matrix* is expressed as $L = \left[\begin{smallmatrix} L_{gg} & L_{lg} \\ L_{gl} & L_{ll}\end{smallmatrix}\right] \in \mathbb{R}^{N\times N}$, where $N = n_g + n_b$. Thus, allowing the system to be described by the dynamic linearized swing equations and the algebraic DC power flow equations in the following manner:

$$\begin{bmatrix} I & 0 & 0 \\ 0 & M & 0 \\ 0 & 0 & 0\end{bmatrix}\dot{x} = -\begin{bmatrix} 0 & -I & 0 \\ L_{gg} & D_g & L_{lg} \\ L_{gl} & 0 & L_{ll}\end{bmatrix}x + \begin{bmatrix} 0 & 0 \\ I & 0 \\ 0 & I\end{bmatrix}u \tag{22}$$

The state variables, $x = [\delta^\top\ \omega^\top\ \theta^\top]^\top \in \mathbb{R}^{2n_g+n_b}$, consist of $\delta \in \mathbb{R}^{n_g}$ the generator rotor angle, $\omega \in \mathbb{R}^{n_g}$ the generator

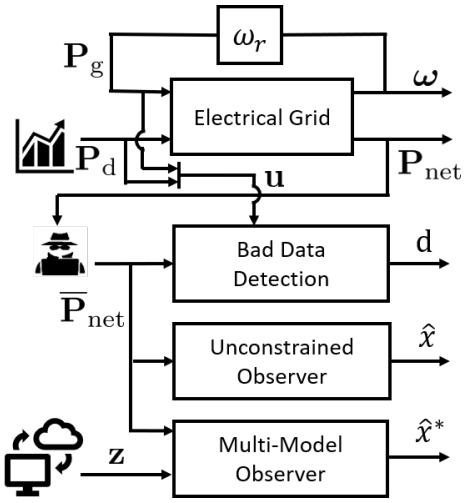

Fig. 3. Block diagram depiction of the simulation scenario.

frequency and $\theta \in \mathbb{R}^{n_b}$ the voltage bus angles. The control input $u = [P_g^\top \ P_d^\top]^\top \in \mathbb{R}^{n_g+n_b}$ consists of $P_g \in \mathbb{R}^{n_g}$ the mechanical input power from each generator, controlled in a closed-loop manner with a PI regulating the generator frequency, and $P_d \in \mathbb{R}^{n_b}$ the active power demand at each bus. Where M is a diagonal matrix of inertial constants for each generator and D_g is a diagonal matrix of damping coefficients.

For the system described in (22), the algebraic variable θ is eliminated to reduce the system dynamics to two state variables, $\tilde{x} = [\delta^\top \ \omega^\top]^\top \in \mathbb{R}^{2n_g}$, as follows:

$$\begin{bmatrix}\dot{\delta}(t)\\ \dot{\omega}(t)\end{bmatrix} = \begin{bmatrix}0 & I\\ -M^{-1}(L_{gg} - L_{gl}L_{ll}^{-1}L_{lg}) & -M^{-1}D_g\end{bmatrix}\tilde{x} + \begin{bmatrix}0 & 0\\ M^{-1} & -M^{-1}L_{gl}L_{ll}^{-1}\end{bmatrix}u,$$

$$y(t) = \begin{bmatrix}0 & I\\ -P_{\text{node}}L_{ll}^{-1}L_{lg} & 0\end{bmatrix}\tilde{x} + \begin{bmatrix}0 & 0\\ -P_{\text{node}}L_{ll}^{-1} & 0\end{bmatrix}u \tag{23}$$

Where P_{node} is a function of the system incidence and susceptance matrices, obtained by linearizing the active power injections at the buses [30]. Consequently, the bus angles vector θ(t) is given by:

$$\theta(t) = -L_{ll}^{-1}(L_{lg}\delta(t) - P_d).$$

The measurement channels $y(t) = [\omega^\top \ P_{\text{net}}^\top]^\top \in \mathbb{R}^{n_g+n_b}$ contain ω the generator frequency used in the PI feedback loop and P_{net} the net power injected at each bus.

B. Auxiliary Model

In our earlier work [11], [29], we used actual data collected from the New York independent system operator (NYISO) to build gaussian process regression (GPR) models which map from market variables (namely, `locational bus marginal prices, marginal cost loses and marginal cost congestion`) to load data (namely, `active and reactive power`). The trained GPR model is executed to give the mean $\boldsymbol{\mu}(\mathbf{z})$ and the covariance $\Sigma(\mathbf{z})$ of the data-driven auxiliary model. For the scenario, presented here, we used the obtained covariance matrix from the earlier work to locate $\boldsymbol{\mu}(\mathbf{z}_k)$ within 3 standard deviations of the true values in the simulation. This allows us to compensate for lack of data in this particular case by recreating a typical auxiliary model from known instances.

C. Simulation Process

The system in (23) is discretized for the implementation of optimization problem in (16). As depicted in the simulation scenario shown in Fig. 3, $\overline{P}_{\text{net}}$ represents a compromised measurement. The goal of the observer design is to give the correct estimate of the rotor angle $\delta(t)$ under FDIA. We also implemented a residue-based bad data detection [31] to monitor the integrity of the FDIA used in the simulation.

D. Results

We implemented three observer schemes; (i) a discretized Luenberger Observer, (ii) an Unconstrained l_1 minimization based Observer based on the results presented in the preliminary section, and (iii) a constrained Multi-model Observer implementing the optimization problem given in (16), using the auxiliary model as described above and in Fig. 3.

Fig. 4, Fig. 5 and Fig. 6 shows the results of each observer when subjected to attacks on 30% of the available measurement. The attack was triggered after 200 samples into into the simulation. This is done to show the observers performances before and after attacks.

Fig. 7 shows the integrity of the FDIA used in the simulation. As seen in the figure, the attacked measurement is able to pass a residue-based bad data detection (BDD) test. The threshold used in the simulation is 0.05 and is assumed to be known to an attacker. Different threshold values were used and the FDIA passed the BDD test for all of them. We only show one plot here due to space limitation.

To further clarify the performances depicted in the figures above, we present in Table I two metrics to give numerical comparison of the performance of each observer. The first metric is the root-mean-square (RMS) value, which quantifies the *energy* of the errors between the actual value and estimated values. The second metric is the maximum absolute value of the error. This helps us capture the worst-case performance of each of the observers compared. It is seen also that, in both metrics considered, the proposed multi-model observer outperforms the other two.

TABLE I
ERROR METRIC VALUES

	RMS metric			**Max. Abs. metric**		
	LO	**L1O**	**MMO**	**LO**	**L1O**	**MMO**
δ_1	2.8801	0.0001	0.0001	6.4274	0.0028	0.0007
δ_2	2.7967	0.0002	0.0001	6.4437	0.0022	0.0013
δ_3	3.2746	0.0018	0.0001	9.7444	0.0387	0.0013
δ_4	3.4786	0.0004	0.0004	10.7019	0.0048	0.0042
δ_5	3.329	0.0011	0.0003	9.1387	0.0121	0.0024
LO: Luenberger Observer, **L1O**: Unconstrained ℓ_1-based Observer **MMO**: Proposed Multi-Model Observer						

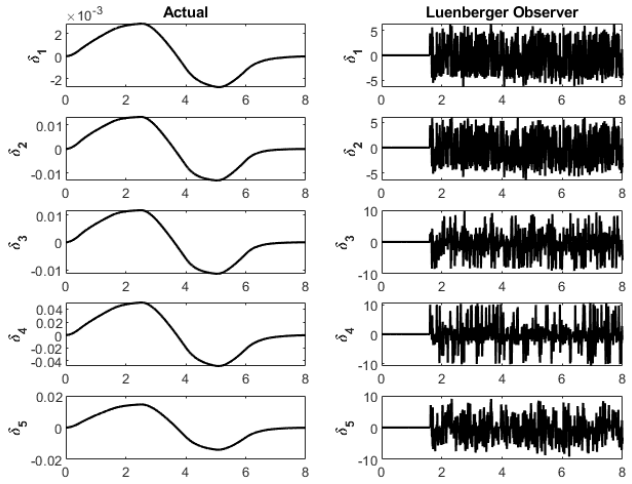

Fig. 4. A comparison of the actual $\delta(t)$ and the observed $\delta(t)$ by a discretized Luenberger Observer. It is seen that any attacks on the measurement channels leads to an inability to reconstruct the real states.

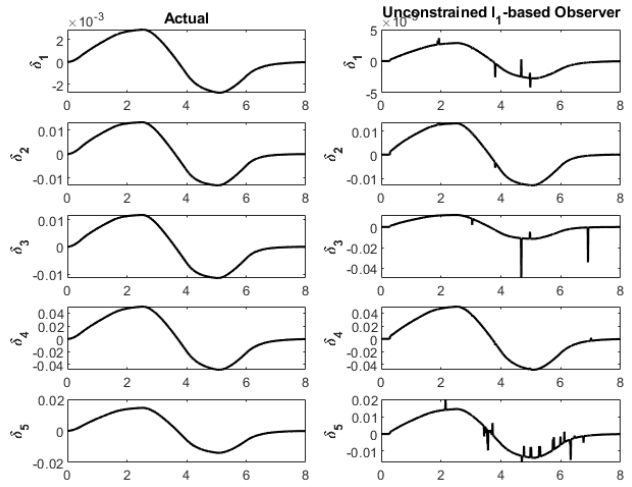

Fig. 5. A comparison of the actual $\delta(t)$ and the observed $\delta(t)$ by the Unconstrained l_1-minimization-based Observer. Although, this observer is able to reconstruct most of the signals, there are still outliers that could cause instability if used as a feedback to a controller.

VI. Conclusions

In this paper, a constrained optimization based resilient state observer is developed using l_1 minimization scheme. The novelty of the algorithm lies in its ability to take into account the machine learning model as a constraint. This constraint, the physics-based model and estimation theory is what makes this multi-model observer resilient. The developed algorithm is evaluated through a numerical example of IEEE-14 bus system. Under FDIA, state measurements differ from their true state. By incorporating the resilient observer the FDIAs can be neutralized and true states can be retrieved with further accuracy.

Some of the problems open for future work include observing the behaviour of resilient observer as filter by cascading it in closed loop with the controller. Considering more complicated constraints for reconstruction optimization problem, rather than a simple quadratic constraint. We would also study the behavior of resilient observer under FDIA and model uncertainty, and extend our approach to other CPS.

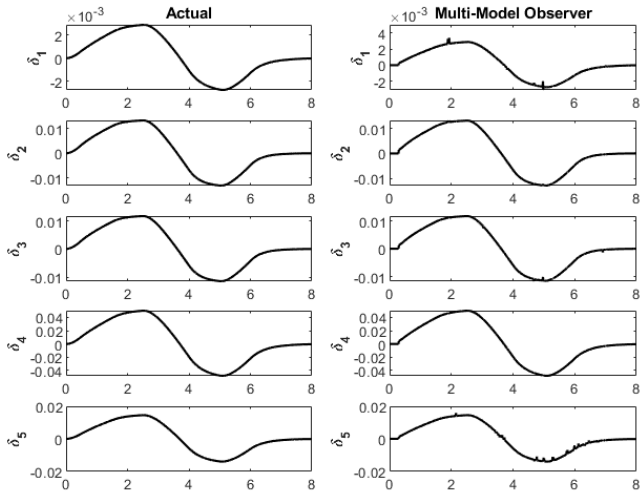

Fig. 6. A comparison of the actual $\delta(t)$ and the observed $\delta(t)$ by the proposed Multi-Model Observer. The proposed observer is able reconstruct the $\delta(t)$ to within a much more reasonable degree of accuracy. This is as a result of the additional information given through the constraint from the auxiliary model.

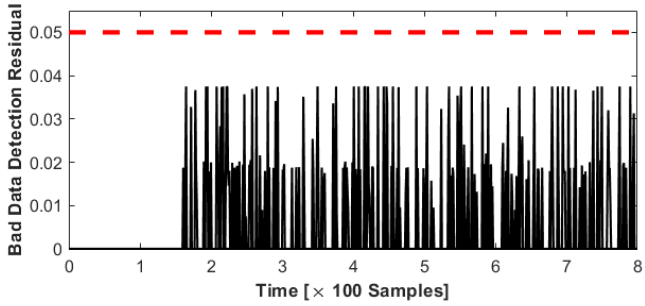

Fig. 7. A FDIA is undetectable by a Bad Data Detector that is set to a 5% threshold

References

[1] C. Konstantinou *et al.*, "Cyber-physical systems: A security perspective," in *20th IEEE European Test Symposium (ETS)*. IEEE, 2015, pp. 1–8.

[2] Y. Wu *et al.*, "Bad data detection using linear WLS and sampled values in digital substations," *IEEE Transactions on Power Delivery*, vol. 33, no. 1, pp. 150–157, 2018.

[3] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security (TISSEC)*, vol. 14, no. 1, p. 13, 2011.

[4] G. Liang *et al.*, "A review of false data injection attacks against modern power systems," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1630–1638, 2017.

[5] R. Deng *et al.*, "False data injection on state estimation in power systems – attacks, impacts, and defense: A survey," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 2, pp. 411–423, 2017.

[6] J. Hao *et al.*, "Sparse malicious false data injection attacks and defense mechanisms in smart grids," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 5, pp. 1–12, 2015.

[7] C. Konstantinou and M. Maniatakos, "A case study on implementing false data injection attacks against nonlinear state estimation," in

Proceedings of the 2nd ACM Workshop on Cyber-Physical Systems Security and Privacy. ACM, 2016, pp. 81–92.

[8] C. Konstantinou *et al.*, "Gps spoofing effect on phase angle monitoring and control in a real-time digital simulator-based hardware-in-the-loop environment," *IET Cyber-Physical Systems: Theory & Applications*, vol. 2, no. 4, pp. 180–187, 2017.

[9] A. Sayghe, O. M. Anubi, and C. Konstantinou, "Adversarial examples on power systems state estimation," in *2020 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2020, pp. 1–5.

[10] A. Ashok, M. Govindarasu, and V. Ajjarapu, "Online detection of stealthy false data injection attacks in power system state estimation," *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 1636–1646, 2018.

[11] O. M. Anubi and C. Konstantinou, "Enhanced resilient state estimation using data-driven auxiliary models," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 1, pp. 639–647, Jan 2020.

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

[13] Q. Hu *et al.*, "Secure state estimation for nonlinear power systems under cyber attacks," *arXiv preprint arXiv:1603.06894*, 2016.

[14] G. Fiore *et al.*, "Secure state estimation for cyber physical systems with sparse malicious packet drops," in *American Control Conference (ACC), 2017*. IEEE, 2017, pp. 1898–1903.

[15] C. Konstantinou and M. Maniatakos, "A data-based detection method against false data injection attacks," *IEEE Design Test*, pp. 1–1, 2019.

[16] S. Mishra *et al.*, "Secure state estimation: Optimal guarantees against sensor attacks in the presence of noise," in *Information Theory (ISIT), 2015 IEEE International Symposium on*. IEEE, 2015, pp. 2929–2933.

[17] X. Liu, Y. Mo, and E. Garone, "Secure dynamic state estimation by decomposing Kalman filter," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 7351–7356, 2017.

[18] L. K. Mestha, O. M. Anubi, and M. Abbaszadeh, "Cyber-attack detection and accommodation algorithm for energy delivery systems," in *Control Technology and Applications (CCTA), 2017 IEEE Conference on*. IEEE, 2017, pp. 1326–1331.

[19] O. M. Anubi, L. Mestha, and H. Achanta, "Robust resilient signal reconstruction under adversarial attacks," *arXiv preprint arXiv:1807.08004*, 2018.

[20] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE transactions on information theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[21] D. Hayden *et al.*, "Sparse network identifiability via compressed sensing," *Automatica*, vol. 68, pp. 9–17, 2016.

[22] M. Pajic *et al.*, "Design and implementation of attack-resilient cyber-physical systems: With a focus on attack-resilient state estimators," *IEEE Control Systems*, vol. 37, no. 2, pp. 66–81, 2017.

[23] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via 1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

[24] M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Transactions on Information Theory*, vol. 48, no. 9, pp. 2558–2567, 2002.

[25] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE transactions on Information theory*, vol. 49, no. 12, pp. 3320–3325, 2003.

[26] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Transactions on Information theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[27] T. T. Cai and A. Zhang, "Sparse representation of a polytope and recovery of sparse signals and low-rank matrices," *IEEE transactions on information theory*, vol. 60, no. 1, pp. 122–132, 2013.

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

[29] O. M. Anubi, C. Konstantinou, and R. Roberts, "Resilient optimal estimation using measurement prior," *arXiv preprint arXiv:1907.13102*, 2019.

[30] E. Scholtz, "Observer-based monitors and distributed wave controllers for electromechanical disturbances in power systems," Ph.D. dissertation, Massachusetts Institute of Technology, 2004.

[31] H.-J. Koglin, T. Neisius, G. Beiβler, and K. Schmitt, "Bad data detection and identification," *International Journal of Electrical Power & Energy Systems*, vol. 12, no. 2, pp. 94–103, 1990.