Enhanced Resilient State Estimation Using Data-driven Auxiliary Models

Olugbenga Moses Anubi, *Member, IEEE* and Charalambos Konstantinou, *Member, IEEE*

***Abstract*—This paper considers the problem of resiliency in the state estimation of a cyber-physical system (CPS) when a portion of its sensor measurements contains malicious data added by an adversarial attacker. When the attack consists of arbitrary random uncorrelated data injection, compressive sensing based regression algorithms that can recover the true states have been studied extensively in literature. However, it has been shown that it is possible to create a targeted correlated false data injection attack (FDIA) which will fool any regression based algorithm. Consequently, there have been plethora of data-driven approaches investigated to detect the occurrence of such FDIA. This paper combines a data-driven model with traditional compressive sensing regression problem. It is shown that the solution of the resulting constrained optimization problem recovers the true states of the system. The developed algorithm is evaluated through a numerical simulation example of the IEEE 14-bus system mapped to the New York Independent System Operator (NYISO) grid data.**

***Index Terms*—Resilient estimation, false data injection attacks, compressive sensing, gaussian process, regression.**

I. 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.

II. Introduction

MODERN industrial systems and applications are composed of a multitude of systems in which their cyber and physical layer are tightly integrated at all scales and levels. Such industrial cyber-physical systems (CPSs) utilize recent computing, communication, and control technologies for the realization of a more autonomous, intelligent, cooperative, and flexible industrial environment. The increased penetration level of internet-of-things (IoT)-controlled devices in CPSs enables new avenues for monitoring, control, and protection solutions by connecting smart networked embedded systems. IoT is a computing concept describing an information network in which physical objects such as sensors, vehicles, home appliances, and machines interact, exchange data, and cooperate in order to reach common goals. While the IoT affects among others transportation, healthcare, or smart homes, the industrial IoT (IIoT) refers in particular to the adoption of IoT in industrial environments [1].

In order to maintain and ensure a normal operating condition, a CPS is consistently monitored and controlled by data acquisition and control systems. Among others, CPS operators utilize measurements acquired from IIoT devices in order to estimate state variables of the CPS. These state estimates are critical as they are used to adjust the control of the physical space. In power systems, for example, once the operating state is known, estimates are utilized for energy management system (EMS) application functions such as optimal control flow, automatic generation control, and contingency analysis. For instance, contingency studies determine the ability of the grid to tolerate failures. System operators can use the contingency analysis results in order to take preventive and corrective actions and overall ensure secure operation of the system. Due to the significance of state estimation routines, it is of paramount importance that such algorithms incorporate proper mechanisms for operating resiliently in the event of malicious data attacks.

An adversary capable of obtaining access to the CPS communication network could alter – during transmission – the measurements sent from the field IIoT devices to the central estimation station [2], [3]. In addition, an adversary may launch attacks by hacking into IIoT sensors and meters or even infiltrate secondary channels of the supply chain in order to distort the measurements [4], [5]. While *bad data detection (BDD)*[1] mechanisms have been traditionally used to remove faulty and erroneous measurements [6], recent studies demonstrate that judiciously falsified information can inject errors in state variables without being detected by BDD [7]–[9]. This class of *false data injection attacks (FDIAs)* able to bypass mechanisms designed to identify bad data, enables vulnerabilities to be exploited by potential adversaries such as nation-states, terrorist organizations, malicious contractors, and disgruntled employees. It has been demonstrated that FDIAs could compromise signals in the electricity market or even mask the outage of lines [10]. Also, the impact of

The authors are with the Department of Electrical and Computer Engineering, FAMU-FSU College of Engineering and the Center of Advanced Power Systems, Florida State University, Tallahassee, FL 32310 USA (e-mail: oanubi@fsu.edu; ckonstantinou@fsu.edu).

[1]Most of these schemes are based on the largest normalized residual method, i.e., on the residual between the obtained measurements and the estimated values for these data as a function of the system states.

FDIAs could be the same as removing the attacked IIoT devices from the network [11]. Moreover, the circumstances of the Ukraine blackout in 2015 underline the plausibility of common assumptions regarding the adversarys knowledge and capabilities to mount a successful FDIA [12].

Several research efforts have highlighted the vulnerability of state estimation algorithms to FDIAs. To address the issues, existing efforts have used one of two strategies; *(i)* protect a set of measurements, and *(ii)* verify each state variable independently. However, 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 [8]. Thus, more computationally feasible, adaptive and real-time implementable resiliency strategies are needed. In addition, developed methods often perform poorly and are typically developed for specific system configurations [13]. Such techniques can be "fooled" by specially crafted FDIAs or might wrongly detect them under a different system configuration. As a result, the concept of resilient estimators in CPSs is gaining more importance in industrial environments as it is important for the system operation to withstand, adapt, and detect efficiently extreme adverse settings.

In this paper, we propose an enhanced resilient state estimation solution for a CPS in which the data acquired from the IIoT sensors and devices are contaminated with FDIAs. We combine a data-driven model with the traditional compressive sensing regression problem. The auxiliary models considered in this work use Gaussian Processes (GP) to build a generative probabilistic regression model from historical data. The resulting algorithm entails solving a convex optimization problem, thereby leveraging the computational feasibility of efficient solvers that have been robustly designed to solve such problems fast enough for real-time implementation.

While it is possible for the auxiliary measurements themselves to be under attack, our claim is that this requires a lot more resource expense by an attacker in order to successfully compromise all relevant auxiliary sources. An attacker would need to first, identify the choice used by the designer of the resilient algorithm, second, stage an attack on all possible sources of the information (this will typically be very large set e.g all sources of ambient temperature measurements and web services), and finally, understand how each source maps to the feasible sets used by the algorithm.

Paper roadmap. The remaining of the paper is organized as follows: in Section III we provide necessary definitions and background for this work. 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.

III. Preliminaries

Since this paper combines concepts from various fields, we have gathered short descriptions of the borrowed concepts in this section to give some completeness and improve readability. The following subsections include short descriptions of compressive sensing, false data injection and GP regression (GPR).

A. Compressive Sensing

There are numerous work in literature on the secure estimation for CPS [14]–[18]. However, we focus only on the ones which are optimization-based as this is the basis of the approach taken 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 [19]. Given a coding matrix $F \in \mathbb{R}^{n\times m}$ with far fewer rows than columns ($n \ll m$) and a vector of observed/measured quantities $\mathbf{y} \in \mathbb{R}^m$, the coding problem is to recover a sparse vector $\mathbf{e}, \|\mathbf{e}\|_{l_0} < m$ for which $\mathbf{y} = F\mathbf{e}$. This is cast as the compressive sensing problem:

$$\underset{\mathbf{e}}{\text{Minimize:}} \|\mathbf{e}\|_{l_0} \;\; \text{Subject to:} \;\; \mathbf{y} = F\mathbf{e}. \tag{1}$$

Hayden *et. al* [20] obtained a sufficient condition that if all subsets of $2q$ columns of F are full rank, then any error $\|\mathbf{e}\|_{l_0} \leq q$ can be reconstructed uniquely by the solution of the optimization problem in (1).

Although in some cases [21], the optimization problem in (1) is solved as is, in most instances, 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:

$$\underset{\mathbf{e}}{\text{Minimize:}} \|\mathbf{e}\|_{l_1} \;\; \text{Subject to:} \;\; \mathbf{y} = F\mathbf{e}. \tag{2}$$

The two programs, however, have been shown to be equivalent under the condition that the *restricted isometric property (RIP)* holds [22], [23]. Let $F^{\mathcal{T}} \triangleq \left(\left(F^\top\right)_{\mathcal{T}}\right)^\top \in \mathbb{R}^{n\times|\mathcal{T}|}, \mathcal{T} \triangleq \mathsf{supp}(\mathbf{e}) \subset \{1,\ldots,m\}$ be the submatrix obtained by extracting the columns of F corresponding to the indices in $\mathcal{T}$. Then the *S-restricted isometry constant* δ_S of F is defined as the smallest quantity such that

$$(1-\delta_S)\|\mathbf{v}\|^2 \leq \left\|F^{\mathcal{T}_S}\mathbf{v}\right\|^2 \leq (1+\delta_S)\|\mathbf{v}\|^2 \tag{3}$$

for all subsets $\mathcal{T}_S$ with $|\mathcal{T}_S| \leq S$ and vector $\mathbf{v} \in \mathbb{R}^{|\mathcal{T}_S|}$. This property essentially requires that every set of columns with cardinality less than S approximately behaves like an orthonormal system. Moreover, it was shown that if

$$\delta_S + \delta_{2S} + \delta_{3S} < 1, \tag{4}$$

then solving the optimization problem in (2) recovers any sparse signal $\mathbf{e}$ for which $|\mathcal{T}| \leq S$.

B. False Data Injection Attack (FDIA)

FDIA is a class of malicious data corruption which generally results in wrong deduction/inference about a system while evading particular BDD algorithms. Ever since it has been introduction in [7], FDIA has gained a lot research interests, with fairly recent comprehensive reviews appearing in [8], [9].

We start with the observation model:

$$\mathbf{y} = H\mathbf{x} + \boldsymbol{\varepsilon}, \tag{5}$$

where $H \in \mathbb{R}^{m\times n},\ n \ll m$ is a system model mapping from internal states $\mathbf{x} \in \mathbb{R}^n$ to a vector of observation $\mathbf{y} \in \mathbb{R}^m$, with measurement error $\boldsymbol{\varepsilon} \in \mathbb{R}^m$. It is assumed that H is known by the attacker, or can be estimated with a reasonable level of accuracy [24], and that $\varepsilon \sim \mathcal{N}(\mathbf{0}, \mathsf{diag}(\sigma_1^2, \ldots, \sigma_m^2))$ is the associated measurement noise, which is normally distributed with zero mean and variance σ_j^2 for the j_{th} sensor, $j = 1, \ldots m$.

Consider the regression objective function $J(\mathbf{x}|\mathbf{y})$ which, given a set of measurement vector $\mathbf{y}$, returns a function of the *regressor* $\mathbf{x}$. An example of such function is the norm-based objective $J(\mathbf{x}|\mathbf{y}) = \|\mathbf{y} - H\mathbf{x}\|_{\ell_s}, s \in \{0, 1, 2, \infty\}$. Consequently, the state estimator is defined as:

$$\hat{\mathbf{x}} = \arg\min J(\mathbf{x}|\mathbf{y}), \tag{6}$$

and a residual-based BDD scheme is given as:

$$\mathsf{BDD}(\mathbf{y}; \tau) = \begin{cases} 0 & \text{if } J(\hat{\mathbf{x}}|\mathbf{y}) \leq \tau \\ 1 & \text{otherwise.} \end{cases} \tag{7}$$

Evidently, any bad data injection $\mathbf{y}_a, \|\mathbf{y}_a\|_{\ell_0} = p \leq m$ for which

$$\min_{\mathbf{x}} J(\mathbf{x}|\mathbf{y} + \mathbf{y}_a) \leq \min_{\mathbf{x}} J(\mathbf{x}|\mathbf{y}) \tag{8}$$

will corrupt the regressor and pass the BDD test. Liu *et. al* demonstrated how easy and feasible it is to construct such FDIA by pointing attention to the fact that any $\mathbf{y}_a \in \mathsf{col}\,(H)$ will satisfy the above property [7]. Such FDIA only requires the knowledge of the system model H which is generally reasonably assumed to be known by an attacker. Moreover, the survey in [8] also chronicles researches which demonstrated the feasibility of constructing such valid FDIA from incomplete information and partial knowledge of H (e.g., [24], [25]).

C. Gaussian Process Regression (GPR)

The auxiliary models considered in this paper uses GPs to build a generative probabilistic regression model from historical data. This is a mapping from auxiliary measurements to a probability distribution of the observed measurement. This provides an added layer of redundancy to the system. An example of such auxiliary information, for a power system, could be: location marginal prices, location ambient conditions, and time of day amongst others.

GP is a collection (possibly infinite) of continuous random variables $\mathcal{G}$, any finite subset of which are jointly Gaussian. GP regression (GPR) uses GPs to encode prior distributions over functions[2]. The priors are then updated to form posterior distributions when new data is collected. For a comprehensive introduction to GP and GPR, and their applications for learning and control, the readers are directed to [26] and a recent survey in [27].

[2]In this case will be functions from auxiliary measurements to observed measurements.

Consider a dataset $\mathcal{D} = \{\mathbf{Z}, \mathbf{Y}\}$, where $\mathbf{Z} \in \mathbb{R}^{p\times N}$ is a matrix containing the values of the auxiliary variables column-wise, $\mathbf{Y} \in \mathbb{R}^{m\times N}$ are the corresponding sensor measurement values and N is the number of datapoints in the dataset. The goal is to learn an implicit mapping $f : \mathbf{R}^p \mapsto \mathbf{R}^m$ for which

$$\mathbf{y}_i = f(\mathbf{z}_i) + \boldsymbol{\varepsilon},\ \ i = 1, \ldots N, \tag{9}$$

where $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathsf{diag}(\sigma_1^2, \ldots, \sigma_m^2))$. In theory, without any further restriction, the problem is ill-defined because there are potentially many possible functions that explains the data exactly notwithstanding the measurement noise. As a means of regularization, the class of functions for consideration is refined by the restriction $f(\mathbf{z}) \sim \mathcal{GP}(m(\mathbf{z}), k(\mathbf{z}, \mathbf{z}'))$ to a GP completely specified by its mean and covariance functions[3]

$$\mu(\mathbf{z}) \triangleq \mathbb{E}[f(\mathbf{z})] \tag{10}$$

$$k(\mathbf{z}, \mathbf{z}') \triangleq \mathbb{E}[(f(\mathbf{z}) - \mu(\mathbf{z}))(f(\mathbf{z}') - \mu(\mathbf{z}'))]. \tag{11}$$

The covariance function can then be specified apriori without an explicit probability distribution. This is where the prior (possibly knowledge-based) information is encoded in the GP. While any positive definite function may pass for a covariance function, one commonly used is the squared exponential covariance function:

$$k(\mathbf{z}, \mathbf{z}') = A \exp\left(-\frac{1}{2l}\|\mathbf{z} - \mathbf{z}'\|^2\right), \tag{12}$$

where hyperparameters A and l implicitly define a smoothness-promoting prior. Given a query point $\mathbf{z} \in \mathbb{R}^p$ for the auxiliary variable, the posterior distribution for the jth sensor values is $p(y_j|\mathbf{z}, \mathcal{D}) = \mathcal{N}(\mu_j(\mathbf{z}), \Sigma_j(\mathbf{z}))$, with the mean and covariance function given by

$$\mu_j(\mathbf{z}) = \mathbf{k}(\mathbf{z})^\top \left(K + \sigma_j^2 I\right)^{-1} \mathbf{Y}_j^\top, \tag{13}$$

$$\Sigma_j(\mathbf{z}) = k(\mathbf{z}, \mathbf{z}) - \mathbf{k}(\mathbf{z})^\top \left(K + \sigma_j^2 I\right)^{-1} \mathbf{k}(\mathbf{z}),\ \ j = 1, \ldots, m \tag{14}$$

where $K \in \mathbb{R}^{N\times N}$ is a covariance matrix with entries $K_{ij} = k(\mathbf{z}_i, \mathbf{z}_j)$ and $\mathbf{k}(\mathbf{z}) \in \mathbb{R}^N$ is a vector with entries $\mathbf{k}(\mathbf{z})_i = k(\mathbf{z}, \mathbf{z}_i)$.

The overall sensor values posterior distribution is given by:

$$p(\mathbf{y}|\mathbf{z}, \mathcal{D}) = \prod_{j=1}^{m} \mathcal{N}(\mu_j(\mathbf{z}), \Sigma_j(\mathbf{z})) \tag{15}$$

$$= \mathcal{N}(\mu(\mathbf{z}), \Sigma(\mathbf{z})), \tag{16}$$

where

$$\mu(\mathbf{z}) = \begin{bmatrix} \mu_1(\mathbf{z}) \\ \vdots \\ \mu_m(\mathbf{z}) \end{bmatrix} \text{ and } \Sigma(\mathbf{z}) = \begin{bmatrix} \Sigma_1(\mathbf{z}) & & \\ & \ddots & \\ & & \Sigma_m(\mathbf{z}) \end{bmatrix}$$

[3]Also known as kernels.

IV. Problem Formulation and Solution Algorithm

We begin with concurrent models of the form:

$$\mathbf{y} = H\mathbf{x} + \boldsymbol{\varepsilon} \tag{17}$$
$$\mathbf{y} \sim \mathcal{N}(\mu(\mathbf{z}), \Sigma(\mathbf{z})) \tag{18}$$
$$\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathsf{diag}(\sigma_1^2, \ldots, \sigma_m^2)) \tag{19}$$

consisting of a physics-based model (17), a data-driven prior (18) given as a function of the auxiliary variable $\mathbf{z} \in \mathbb{R}^m$, and a knowledge-based noise prior (19). One of the main advantages of using models of this form is that the resulting blend of the generalization properties of physics-based models and the adaptive local accuracy of data-driven methods creates an additional layer of redundancy which can reveal the truth even if portions of the measurement is subject to adversarial corruption. In order to remain undetectable, any viable attack vector $\mathbf{y}_a, \|\mathbf{y}_a\|_{\ell_0} = p \leq m$ necessarily have to satisfy the likelihood dominance condition $p(\mathbf{y} + \mathbf{y}_a|\mathbf{z}, \mathcal{D}) \geq p(\mathbf{y}|\mathbf{z}, \mathcal{D})$. This provides an 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 any true state vector $\mathbf{x}^* \in \mathbb{R}^n$, the corresponding maximum possible state corruption can be obtained by solving the following noncovex optimization problem[4]:

$$\begin{aligned} \text{Maximize:} \quad & \|\mathbf{x}_a\|_{\ell_2} \\ \text{Subject to:} \quad & \\ & \|H\mathbf{x}_a\|_{\ell_0} \quad \leq p \\ & \mathsf{LHS} \quad \leq 0, \end{aligned} \tag{20}$$

$$\mathsf{LHS} = \|H\mathbf{x}_a\|^2_{\Sigma^{-1}(\mathbf{z})} - 2\left(H\mathbf{x}^* - \mu(\mathbf{z})\right)^\top \Sigma^{-1}(\mathbf{z}) H\mathbf{x}_a.$$

The second inequality is obtained from the likelihood dominance condition above, after some algebraic manipulation. It is possible to obtain some bounds on the maximum possible state corruption. This is beyond the scope of this paper and left for future work. Also, one can easily see the advantage of the auxiliary model by noticing how the covariance and the mean function drive the maximum possible state corruption. The *attack radius* ($\triangleq \max \|\mathbf{x}_a\|_{\ell_2}$) reduces with reduced GPR model uncertainty. For instance, if the auxiliary is exact (i.e., $H\mathbf{x}^* = \mu(\mathbf{z})$), then the only possible undetectable state corruption is trivial ($\mathbf{x}_a = \mathbf{0}$), rendering all attacks infeasible.

Let $\mathbf{y}^*$ be the true value of the measured variable, the enhanced resilient state estimation is cast as the optimization problem:

$$\begin{aligned} \text{Minimize:} \quad & \|\mathbf{y} - H\mathbf{x} - \boldsymbol{\varepsilon}\|_{l_0} \\ \text{Subject to:} \quad & \\ & H\mathbf{x} \in \mathcal{Y}(\mathbf{z}) \\ & \boldsymbol{\varepsilon} \in \mathcal{E}, \end{aligned} \tag{21}$$

where the convex sets $\mathcal{Y}(\mathbf{z})$ and $\mathcal{E}$ have the property that:

$$p(\mathbf{y}^* \in \mathcal{Y}|\mathbf{z}, \mathcal{D}) \geq \tau \tag{22}$$
$$p(\boldsymbol{\varepsilon}^* \in \mathcal{E}) \geq \tau. \tag{23}$$

[4]The noise term $\boldsymbol{\varepsilon}$ can be omitted without loss of generality.

The idea is essentially seeking a state vector, together with the minimum attacked channels and a highly likely noise vector, which completely explains the observations while 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. An additional layer of resiliency can be achieved by introducing a coding matrix $C \in \mathbb{R}^{n_c \times m}$ to transform the measurements used in the auxiliary model prior. Thus, the resilient state estimation optimization problem is equivalent to:

$$\begin{aligned} \text{Minimize:} \quad & \|\mathbf{y} - H\mathbf{x} - \boldsymbol{\varepsilon}\|_{l_0} \\ \text{Subject to:} \quad & \\ & \|C\left(H\mathbf{x} + \boldsymbol{\varepsilon} - \mu(\mathbf{z})\right)\|^2_{\Sigma^{-1}(\mathbf{z})} \quad \leq \chi^2_{n_c}(\tau) \\ & \|\boldsymbol{\varepsilon}\|^2_{\Sigma_\varepsilon^{-1}} \quad \leq \chi^2_m(\tau), \end{aligned} \tag{24}$$

where $\Sigma_\varepsilon = \mathsf{diag}(\sigma_1^2, \ldots, \sigma_m^2)$ and $\chi^2_{n_c}(\tau)$ is the quantile function for probability τ of the chi-squared distribution with n_c degrees of freedom.

A. Solution Algorithm

The above optimization problem is nonconvex due to the index minimization objective. This is suggestive of possible NP-hardness and associated intractability properties. 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. Moreover, under a more general situation where the RIP condition does not hold necessarily, iteratively re-weighted algorithms [28], [29] have been demonstrated to be a highly effective way of approximating the solution of the nonconvex problem with successive convex problems. In particular, for the solution of the problem in (24), the re-weighted ℓ_1 minimization scheme of [28] is employed.

Consider the operator $\mathcal{P} : \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^{m \times m} \mapsto \mathbb{R}^{n+m}$, where

$$\hat{\mathbf{x}}(W),\ \ \hat{\varepsilon}(W) = \mathcal{P}(\mathbf{y}, \mathbf{z}, W) \tag{25}$$

are given by the minimizers of the convex program:

$$\begin{aligned} \text{Minimize:} \quad & \|W\left(\mathbf{y} - H\mathbf{x} - \boldsymbol{\varepsilon}\right)\|_{l_1} \\ \text{Subject to:} \quad & \\ & \|C\left(H\mathbf{x} + \boldsymbol{\varepsilon} - \mu(\mathbf{z})\right)\|^2_{\Sigma^{-1}(\mathbf{z})} \quad \leq \chi^2_{n_c}(\tau) \\ & \|\boldsymbol{\varepsilon}\|^2_{\Sigma_\varepsilon^{-1}} \quad \leq \chi^2_m(\tau), \end{aligned} \tag{26}$$

Using this, the algorithm for the enhance state estimator is outlined in Algorithm 1.

V. Simulation Results

The proposed enhanced resilient state estimation algorithm is evaluated with simulations performed on the IEEE 14-bus test case. The IEEE 14-bus system, shown in Fig. 1a, represents a simple approximation of the American electric power

Algorithm 1 Enhanced Resilient State Estimation Algorithm

procedure OFFLINE
  $\mathcal{D} \leftarrow \{\mathbf{Z}, \mathbf{Y}\}$ ▷ Dataset sparsification
  $K \leftarrow k(\mathbf{Z}, \mathbf{Z})$ ▷ Kernel matrix
  $\Sigma_\varepsilon, A, l \leftarrow$▷ Hyperparameters initialization, see Table I
procedure ONLINE
  procedure COLLECT DATA
    $\mathbf{y} \leftarrow$ ▷ Sensor measurements at the current instant
    $\mathbf{z} \leftarrow$ ▷ Auxiliary measurements at the current instant
  procedure UPDATE MODELS
    $H \leftarrow$▷ Model-based. See Sub-section V-A for details
    for $j = 1$ to m **do** ▷ Data-driven posterior
      $\mu_j \leftarrow \mathbf{k}(\mathbf{z})^\top \left(K + \sigma_j^2 I\right)^{-1} \mathbf{Y}_j^\top$, ▷ Mean
      $\Sigma_j \leftarrow k(\mathbf{z}, \mathbf{z}) - \mathbf{k}(\mathbf{z})^\top \left(K + \sigma_j^2 I\right)^{-1} \mathbf{k}(\mathbf{z})$, ▷ Covariance
  procedure RE-WEIGHTED ℓ_1 MINIMIZATION($\mathbf{y}$,$\mathbf{z}$)
    $W \triangleq \mathsf{diag}[w_1, \ldots, w_m] \leftarrow I$
    $l \leftarrow 0$ ▷ Iteration count
    while `not converged and` $l \leq l_{max}$ **do**
      $\hat{\mathbf{x}}^l,\ \hat{\varepsilon}^l \leftarrow \mathcal{P}(\mathbf{y}, \mathbf{z}, W)$ ▷ ℓ_1 minimization
      $\mathbf{r} \leftarrow \mathbf{y} - H\hat{\mathbf{x}}^l - \hat{\varepsilon}^l$ ▷ residual
      for $j = 1$ to m **do** ▷ weights update
        $w_j \leftarrow \frac{1}{|\mathbf{r}_j| + \delta}$
      $l \leftarrow l + 1$ ▷ increment counter
    return $\hat{\mathbf{x}}^l, \hat{\varepsilon}^l$ ▷ State estimate is $\hat{\mathbf{x}}^l$

Algorithm 2 Honest Gauss-Newton Method

procedure HONEST GAUSS NEWTON
  $\mathbf{x}_0 = \begin{bmatrix}\theta_{2_0}, & \ldots, & \theta_{S_0}, & \mathbf{V}_{1_0}, & \ldots, & \mathbf{V}_{S_0}\end{bmatrix}^T$
    $= \begin{bmatrix}0, & \ldots, & 0, & 1, & \ldots, & 1\end{bmatrix}^T$
  for $\forall\ \mathbf{y}, \mathbf{x}$ **do**
    $\Delta\mathbf{x}_k = \left({J_h}^T \cdot W \cdot J_h\right)^{-1} \cdot {J_h}^T \cdot W \cdot (\mathbf{y} - h(\mathbf{x}_k))$
    $\mathbf{x}_{k+1} = \mathbf{x}_k + \Delta\mathbf{x}_k$
    $h(\mathbf{x_{k+1}})$ ▷ update: Eq. (27) – (30)
    $J_h(\mathbf{x}_{k+1})$ ▷ update: $\partial z_i / \partial x_j$ of Eq. (27) – (30)

system as of February 1962. It has 14 buses, 5 generators, and 11 loads. The system has 27 state variables which are the voltage angles and voltage magnitudes of the buses, with the first bus angle chosen as the reference one. The buses/nodes of the power grid model are assumed to be supported with IIoT measurement sensors such as remote terminal units (RTUs) able to provide bus-related measurements of active and reactive power injection and flow.

A. Setup

1) IEEE 14-bus model – NYISO data mapping: In this study, we leverage synthetic data in order to evaluate the performance of the developed estimator. Specifically, due to the lack of real-time system measurements and states, we use similarly to [4], [30], load data of New York (NY) state as provide by the NY Independent System Operator (NYISO) [31]. Five-minute load data of NYISO for 3 months (between January and March) in 2017 and 2018 are used in order to generate the states of the system. Each region of the NYISO map, shown in Fig. 1b, is mapped in an ascending order with every load bus of IEEE 14 system, i.e. using the following mapping: $[2 \rightarrow 1,\ 3 \rightarrow 2,\ 4 \rightarrow 3,\ 5 \rightarrow 4,\ 6 \rightarrow 5,\ 9 \rightarrow 6,\ 10 \rightarrow 7,\ 11 \rightarrow 8,\ 12 \rightarrow 9,\ 13 \rightarrow 10,\ 14 \rightarrow 11]$, where the first element show the load bus of IEEE 14 case the second the region of NYISO, e.g., bus 2 to region A-WEST, bus 3 to region B-GENESE, bus 4 to region C-CENTRL, etc. Each of the NYISO load vector is normalized to the initial active $\mathbf{P_0}$ and reactive $\mathbf{Q_0}$ of the IEEE 14 bus system load data based on a constant power factor (so only real power prediction is needed; this assumption can be relaxed if the historical data of reactive power is available). Doing this procedure, for every five-minute interval, we obtain $\mathbf{P_L}$ and $\mathbf{Q_L}$ load data for the 14 case system. The ratio of the derived total $\mathbf{P_L/Q_L}$ to the IEEE 14 bus initial total $\mathbf{P_0/Q_0}$ is utilized as the rate which the system generators increase their output power. In real operating conditions, the latter can be configured accordingly, based on forecasting and scheduling generation data [32].

The data obtained from the above procedure are used to compute the system state $\mathbf{x} \in \mathbb{R}^n$ via the nonlinear power flow Newton-Raphson method. The exact process is shown in Algorithm 2 with the Honest Gauss-Newton method (i.e., the weighted least square estimation in which the Jacobian matrix J_h is updated at each iteration), where W is a diagonal weight matrix (typically inverses of measurement noise variance) and the J_h of $h(\mathbf{x})$ relates the measurements to the states. Each element $J_{h_{ij}}$ represents the derivative of i_{th} measurement with respect to the j_{th} state variable based on (27) – (30).

$$\mathbf{P}_{ij} = \mathbf{g}_{ij}|\mathbf{V}_i|^2 - |\mathbf{V}_i||\mathbf{V}_j|(\mathbf{g}_{ij}\cos\theta_{ij} - \mathbf{b}_{ij}\sin\theta_{ij}) \quad (27)$$

$$\mathbf{Q}_{ij} = \mathbf{b}_{ij}|\mathbf{V}_i|^2 - |\mathbf{V}_i||\mathbf{V}_j|(\mathbf{g}_{ij}\sin\theta_{ij} + \mathbf{b}_{ij}\cos\theta_{ij}) \quad (28)$$

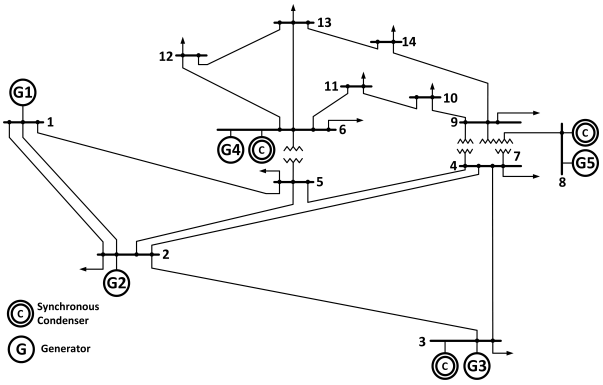
$$\mathbf{P}_i = \sum_{j \in \mathcal{S}_i} \mathbf{P}_{ij} \quad (29)$$

$$\mathbf{Q}_i = \sum_{j \in \mathcal{S}_i} \mathbf{Q}_{ij} \quad (30)$$

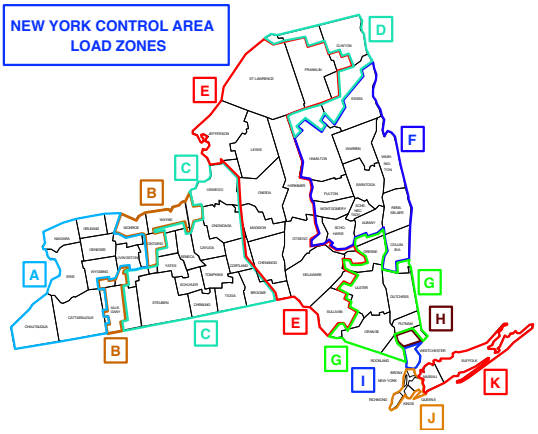
where $\mathcal{S}_i \subseteq \mathcal{S}$ indicates the set of buses connected to bus i. The argument of the sinusoidal functions, $\theta_{ij} = \theta_i - \theta_j$, denotes the voltage phase angle difference between buses i and j. Moreover, $\mathbf{g}_{ij}$ and $\mathbf{b}_{ij}$ form the line series admittance $\mathbf{y}_{ij}$ and correspond to the conductance and susceptance of the line between buses i and j, respectively.

The last step in the procedure is to compute according to J_h of the 14-bus system and for each five-minute interval, the measurement vectors $\mathbf{y} = h(\mathbf{x})$ that will be used in our simulation model, where $h : \mathbb{R}^n \mapsto \mathbb{R}^m$ represent the relationship between state variables $\mathbf{x}$ and measured values $\mathbf{y}$ based on the power flow and power injection equations formed based on the system structure.

2) Simulation Process: The enhanced resilient estimation algorithm in Algorithm 1 was implemented and ran for data collected every five minutes in a simulation environment. The simulation process is shown in Fig. 2. The process begins with the auxiliary measurements $\mathbf{z} = \begin{bmatrix} z_{\text{lbmp}} & z_{\text{mcl}} & z_{\text{mcc}} \end{bmatrix}$, which



(a) IEEE 14-bus system.



(b) NYISO map of the 11 control area load zones

Fig. 1: IEEE 14-bus system mapped into NYISO control area load zones data.

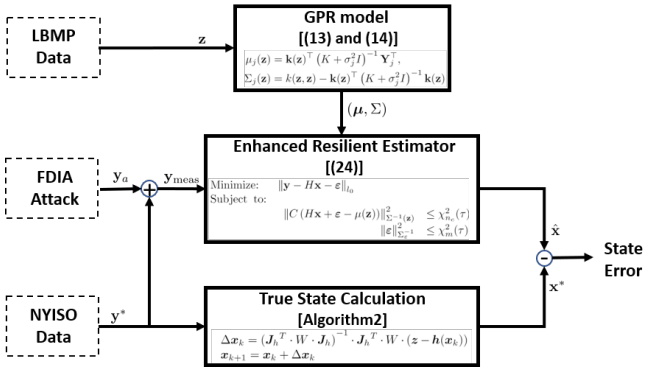


Fig. 2: Simulation process.

are actual data downloaded from the respective nodes of the NYISO transmission grid. Here, z_{lbmp} is the `locational bus marginal prices` (\$/MWh), z_{mcl} is the `marginal cost loses` (\$/MWh) and z_{mcc} is the `marginal cost congestion` (\$/MWh). Next, 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 of Section III. Two kinds of FDIA generation were used in the simulation. For the first kind,

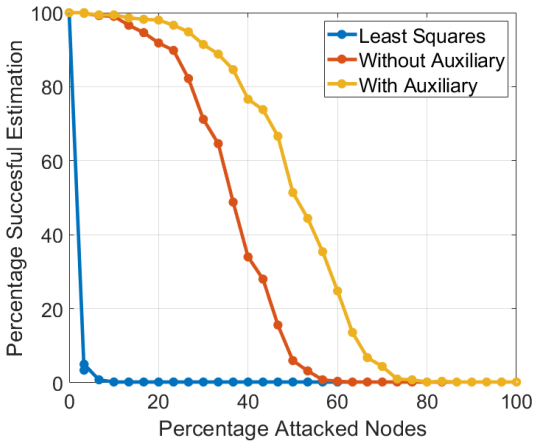


Fig. 3: Simulation results for targeted sensor measurements. Attack vectors are generated to bias select measurements locations by 500% of its true value along a randomly chosen direction. Plots is the percentage of successful estimations vs. the percentage of attacked sensor nodes.

τ	δ	C	$l_{\max}$
0.5	0.01	I	100

TABLE I: Algorithm 1 parameter set values.

attack vectors are generated to bias selected measurements locations by 500% of its true value along a randomly chosen direction. For the second kind, the attack vectors $\mathbf{y}_a$ are systematically generated to result in a specified bias in the state estimation at targeted state variables according to:

$$\mathbf{y}_a = \eta H_a \mathbf{x}^* \tag{31}$$

where $\eta > 0$ is the specified bias as a fraction of the true state, $H_a \in \mathbb{R}^{m \times n}$ is the instantaneous Jacobian matrix with the columns corresponding to the untargeted state variables set to zeros, and $\mathbf{x}^*$ is the true state vector. The *true* sensor measurement values used are the actual load data of NY state provided by the NYISO. The observed measurement inputted to the resilient estimator is the addition of the generated attack vector and the true downloaded load data. The re-weighted ℓ_1 algorithm in Algorithm 1 is then ran to produce the estimated state vector $\hat{\mathbf{x}}$. This is compared with the true state vector $\mathbf{x}^*$ obtained by executing the Honest Gauss-Newton method in Algorithm 2.

B. Results

Fig. 3 and Fig. 4 show the performance of the proposed algorithm, compared with other standard methods in literature, to the two kinds of FDIA described above. The parameter values used for the algorithm implementation that produced the results are given in Table I.

For the first set of results, three different state estimation algorithms are simulated against a FDIA directed at specific measurement locations. The three algorithms are: *1)* standard least squares ($\hat{\mathbf{x}} = \arg\min \|\mathbf{y} - H\mathbf{x}\|^2$), *2)* re-weighted ℓ_1 without the auxiliary model constraint and *3)* the proposed re-weighted ℓ_1 with auxiliary model constraint. There are 109

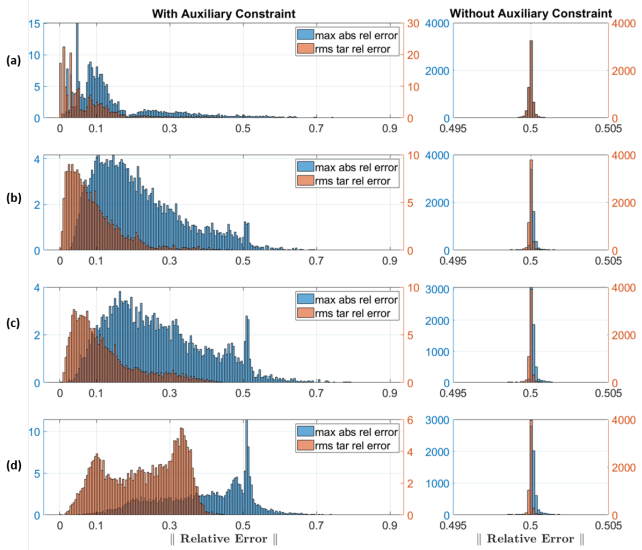


Fig. 4: Simulation results for targeted state FDIA.
Attack vectors are generated to bias particular state variables by 50%. Plotted are the distribution of the `rms` values of relative errors for targeted states and maximum absolute relative error over all state variables. Subplots: (a) 1 targeted state variable, (b) 5 targeted state variables (c) 10 targeted state variables, (d) 20 targeted state variables.

load flow measurements in the simulation. Each simulated scenario, circle points in Fig. 3, examines 200 simulations (per state estimation method) with random combinations of sensor locations having fixed percentage (`x-axis` values) of sensor nodes under attack. For each simulation and each method, a relative error performance metric is calculated as follows:

$$\texttt{RelativeError} \triangleq \max_i \frac{\hat{\mathbf{x}}_i - \mathbf{x}_i^*}{\mathbf{x}_i^*} \tag{32}$$

where $\hat{\mathbf{x}}, \mathbf{x}^* \in \mathbb{R}^{27}$ are the estimated states and true states, respectively. The estimation is successful if the relative error metric is less than that of a corresponding least square estimation with zero attacked nodes. Fig. 3 shows that the proposed approach has significant improvement over the other two. Specifically, there is over 50% more correctly estimated cases over the re-weighted ℓ_1 method when 50% of measurement nodes are compromised. This shows a lot of promise for using readily available information to boost the secure operations of critical CPS infrastructures against adversarial agents.

For the second set of results, we exclude least squares from the comparison since it under performs the re-weighted ℓ_1 methods. Also, the attacks created here will live in the range space of the system Jacobian matrices and it is obvious from Section III and from the literature (e.g., [7]) that both unconstrained methods will behave similarly under this class of attacks. Thus, to facilitate a cleaner presentation, we restrict our comparison to the re-weighted ℓ_1 algorithms – one with auxiliary constraints and the other without. Fig. 4 shows the simulations results for four different cases with different numbers of targeted state variables. For each case, we simulate an equivalent of one month operation by running the simulation process in Fig. 2 for one month, five minute interval, of actual data downloaded from the NYISO. This is equivalent to 8640 scenarios on average. For each scenario, the equation in (31) is used (with η = 0.5) to calculate the state-targeted attack for randomly selected state variables.

Fig. 4 shows two plots for each case side-by-side – one with auxiliary constraints and the other without. Each plot contains the distributions of the maximum absolute relative error in (32), as well as the root-mean-square (`rms`) values of the relative error for the targeted states. The `rms` values are indicators of the performance of the algorithms with respect to estimates of the attacked variable, while the maximum absolute relative error values indicates the overall effect the attack has on all state variables. Concentration of the distributions closer to 0 indicates good performance while concentration around η (=0.5 in this case) indicates bad performance. As can be seen from the figures, re-weighted ℓ_1 algorithms without auxiliary constraints, even though significantly outperforms least-squares based methods in general, are not resilient against state-targeted FDIA. In fact, both performance indicators have means of 0.5 and standard deviations of 0.0057 for the case with only one targeted state variable.

The proposed re-weighted ℓ_1 with auxiliary constraints shows significant improvement for both performance indicators. Noticeable effects of the state-targeted FDIA begin to appear when 10 or more states are targeted. This requires compromising more or less 85% of the system measurement, a feat that demands tremendous amount of resources from any malicious actor. Thus, by corroborating the state estimation with auxiliary model, we have demonstrated that it is possible to make it much more difficult for FDIA to succeed on CPS – even when large portions of measurements are corrupted.

VI. Conclusions

In this paper, we propose an enhanced resilient state estimation algorithm which combines a data-driven model with compressive sensing regression. Using GPR to construct an auxiliary model from historical data, we build a regression-based resilient estimator constrained by consistency with the auxiliary model. This provides a way to use secondary sources of information to strengthen the resiliency of CPSs against adversarial data corruption. The effectiveness of the presented solution is demonstrated with an application to power systems in which data acquired from various IIoT sensors and devices are poisoned with data injection attacks. The particular case tested is the IEEE 14-bus system mapped to actual load data from NYISO transmission grid. Different attack scenarios are examined for three different state estimation algorithms. The results shows remarkable resiliency boost by using locational bus marginal pricing signal as a secondary source of information to constrain existing regression-base state estimation methods. Thus, by corroborating the state estimation with auxiliary model, we have demonstrated that it is possible to make it much more difficult to attack a CPS just by corrupting portions of its sensor measurements.

Our future work will aim to incorporate additional auxiliary information in the estimation, as well as evaluate the developed

algorithms through digital real-time simulation platforms using both simulated and field data. Moreover, there are interesting theoretical questions that remain open. For instance, what is the rate of successful resilient estimation as a function of the auxiliary model uncertainty and the likelihood threshold τ? Another theoretical question of practical significance is the resulting stability assessments and margins of the resulting closed loop system when the enhanced resilient estimator is used as a dynamic filter, whereby the estimated states are fed into the underlying controller(s). An answer to these questions, and likes, will help us judge the quality of an auxiliary model required to achieve a given success rate. Finally, we aim to extend these approach to other examples of CPSs.

REFERENCES

[1] E. Sisinni *et al.*, "Industrial internet of things: Challenges, opportunities, and directions," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 11, pp. 4724–4734, 2018.

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

[3] 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.

[4] 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.

[5] S. McLaughlin *et al.*, "The cybersecurity landscape in industrial control systems," *Proceedings of the IEEE*, vol. 104, no. 5, pp. 1039–1057, 2016.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] X. Liu, Z. Li, and Z. Li, "Impacts of bad data on the PMU based line outage detection," *arXiv preprint arXiv:1502.04236*, 2015.

[11] O. Kosut *et al.*, "Limiting false data attacks on power system state estimation," in *Information Sciences and Systems (CISS), 2010 44th Annual Conference on*. IEEE, 2010, pp. 1–6.

[12] G. Liang *et al.*, "The 2015 ukraine blackout: Implications for false data injection attacks," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3317–3318, 2017.

[13] 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.

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

[15] 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.

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

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

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

[21] 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.

[22] 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.

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

[24] X. Liu and Z. Li, "False data attacks against AC state estimation with incomplete network information," *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2239–2248, 2017.

[25] M. A. Rahman and H. Mohsenian-Rad, "False data injection attacks with incomplete information against smart power grids," in *Global Communications Conference, IEEE*, 2012, pp. 3153–3158.

[26] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[27] M. Liu *et al.*, "Gaussian processes for learning and control: A tutorial with examples," *IEEE Control Systems Magazine*, vol. 38, no. 5, pp. 53–86, Oct 2018.

[28] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier analysis and applications*, vol. 14, no. 5-6, pp. 877–905, 2008.

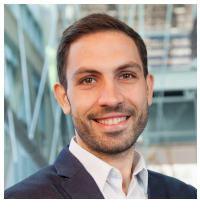
[29] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2008, pp. 3869–3872.

[30] G. Chaojun, P. Jirutitijaroen, and M. Motani, "Detecting false data injection attacks in AC state estimation," *IEEE Transactions on Smart Grid*, vol. 6, no. 5, pp. 2476–2483, 2015.

[31] NYISO. Load data. [Online]. Available: http://www.nyiso.com/public/markets_operations/market_data/load_data/index.jsp.

[32] M. Nejati, N. Amjady, and H. Zareipour, "A new stochastic search technique combined with scenario approach for dynamic state estimation of power systems," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 2093–2105, 2012.

Olugbenga Moses Anubi (M'15) received his Ph.D in Mechanical Engineering from the University of Florida. He was a Lead Control Systems Engineer at GE Global Research, Niskayuna. He is currently an Assistant Professor of Electrical and Computer Engineering at the FAMU-FSU College of Engineering, with affiliations with the Center for Advanced Power Systems (CAPS) and the Center for Intelligent Systems, Controls and Robotics (CISCOR). His research interests include robust, resilient and adaptive control systems, cyber-physical systems control, real-time optimization , robotics, and vehicle dynamics & control.



Charalambos Konstantinou (S'11-M'18) received his Ph.D. in electrical engineering from New York University (NYU), NY and the Dipl.-Ing. (M.Eng.) degree in electrical and computer engineering from National Technical University of Athens (NTUA), Greece. He is currently an assistant professor of electrical and computer engineering with Florida A&M University and Florida State University (FAMU-FSU) College of Engineering and an affiliated faculty with the Center for Advanced Power Systems (CAPS) at FSU. His research interests focus on cyber-physical, industrial control, and embedded systems security.