A Generalized False Data Injection Attacks Against Power System Nonlinear State Estimator and Countermeasures

Junbo Zhao, *Student Member, IEEE*, Lamine Mili, *Life Fellow, IEEE*, and Meng Wang, *Member, IEEE*

***Abstract*—This paper develops a generalized framework that allows us to investigate the vulnerability of the power system nonlinear state estimator to false data injection attacks (FDIAs) from the operator's perspective and to initiate some countermeasures. Unlike most existing FDIA methods, which assume a perfect knowledge of the system measurements and topology by a hacker, we derive and analyze the uncertainties for launching successful FDIAs along with their upper bounds. To effectively defend against an FDIA, we propose a robust detector that checks the measurement statistical consistency using a subset of secure PMU measurements. We first show that if these secure PMU measurements are free of bad data while making the system observable, the FDIA is detectable. We then show that detectability is also ensured if these conditions are relaxed while using alternative redundant measurements from short-term nodal synchrophasor predictions together with the robust Huber M-estimator. Numerical simulation results on the IEEE 30-bus and 118-bus systems demonstrate the effectiveness and robustness of the proposed method even the secure measurements contain noise and bad data.**

***Index Terms*—Cyber security, false data injection attacks (FDIAs), power system nonlinear state estimation, robust estimation, phasor measurement units, Neyman-Pearson detector.**

I. Introduction

DUE to a strong reliance of smart grid functions on communication networks, cyber attacks have become a major concern among power researchers. The analysis of cyber attacks on power system state estimation (SE) was pioneered by Liu *et al.* [1], where the so-called false data injection attack (FDIA) was introduced. Following this work, three types of FDIAs were pinpointed and investigated, including state attacks [1], [2], topology attacks [3], [4] and load redistribution attacks [5]. Their impacts on the electricity markets were also analyzed [6], [7].

Manuscript received April 15, 2017; revised September 2, 2017 and October 25, 2017; accepted January 12, 2018. This work was supported in part by the U.S. National Science Foundation under Grant ECCS-1711191. Paper no. TPWRS-00551-2017. *(Corresponding author: Junbo Zhao.)*

J. Zhao and L. Mili are with the Bradley Department of Electrical Computer Engineering, Virginia Polytechnic Institute and State University, Northern Virginia Center, Falls Church, VA 22043 USA (e-mail: zjunbo@vt.edu; lmili@vt.edu).

M. Wang is with the Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180 USA (e-mail: wangm7@rpi.edu).

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

Digital Object Identifier 10.1109/TPWRS.2018.2794468

To safeguard the system operation and control against cyber attacks, various detectors and mitigation methods have been proposed. They include measurement protection-based approaches [8]–[11], sparse optimization or game theory-based approaches [12], [13], innovation-based approaches [14], [15], robust estimation-based approaches [16], to name a few. *However, with the exception of [17]–[20], the bulk of the literature focused on the linear DC rather than on the AC state estimator.* In [17], the vulnerability of the nonlinear SE to FDIA was analyzed and discussed. It was shown that a hacker should know perfectly all the state variables in the attack subgraph to conduct a successful attack. In other words, she should have exactly the same information about the power system as the operators of the control center, including an exact knowledge about the measurements and the system topology. In this paper, this type of attack will be called the perfect FDIA attack. The work in [17] was later extended by Rahman *et al.* [18], Zhao *et al.* [19], and Xuan *et al.* [20] to take into account the uncertainties in the system information that may be gathered by a hacker, resulting in an imperfect FDIA. However, no analytical investigations were carried out to explain why an imperfect attack may succeed and under which conditions it will be detected by the operators of the control center. Furthermore, defense approaches have been studied extensively for a linear DC model-based FDIA. But little work has been done for a nonlinear AC model-based FDIA. Note that the practical control center uses nonlinear power system state estimator for monitoring and control. It is thus of vital importance to ensure the security and reliability of that estimator.

In this paper, an analytical framework is proposed to investigate the vulnerability of power system nonlinear state estimator to an FDIA from the operator's perspective. In particular, we propose a generalized FDIA framework against the nonlinear state estimator. In this framework, the perfect knowledge of the system information is relaxed to account for measurement, parameter and topology uncertainties. The latter may be induced by the hacker's limited real-time knowledge of the status of various grid elements or restricted access to communication channels [14]. The upper bounds of these uncertainties for launching a successful FDIA are quantified and analyzed as well. To effectively detect an FDIA, we propose a robust detector by checking the measurement statistical consistency using a subset of secure PMU measurements. It is shown that these secure measurements allow us to detect an FDIA if they are free of gross errors while making the system observable. These con-

ditions are further relaxed by using a robust Huber M-estimator together with alternative redundant measurements from short-term nodal synchrophasor predictions. Interestingly, robust state estimates provided by Huber M-estimator is shown to follow a Gaussian distribution, which enables us to derive the analytical form of the Neyman-Pearson detector.

The remainder of this paper is organized as follows: Section II introduces the existing FDIAs against nonlinear state estimator and presents the problem statement. Section III presents the proposed generalized FDIA framework, while Section IV presents the proposed robust FDIA detector. The simulation results are analyzed in Section V, and finally Section VI concludes the paper.

II. Problem Formulation

A. Power System Nonlinear State Estimation

As shown in [21], for an N-bus power system using an AC power flow model, the relationship between the vector of measurements $\boldsymbol{z} \in \mathbb{R}^m$ obtained from the supervisory control and data acquisition (SCADA) system and the state vector $\boldsymbol{x} \in \mathbb{R}^n$, which contains the nodal voltage magnitudes and phase angles, yielding $n = 2N - 1 < m$, is given by

$$\boldsymbol{z} = \boldsymbol{h}(\boldsymbol{x}) + \boldsymbol{e}, \tag{1}$$

where $\boldsymbol{h}(\cdot) : \mathbb{R}^n \to \mathbb{R}^m$ is a vector-valued nonlinear function; $\boldsymbol{e} \in \mathbb{R}^m$ is the measurement error vector that is assumed to follow a Gaussian distribution with zero mean and covariance matrix $\boldsymbol{R} \in \mathbb{R}^{m\times m}$, i.e., $\boldsymbol{e} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R})$. The state estimator is solved by minimizing the weighted least squares criterion, yielding

$$\widehat{\boldsymbol{x}} = \arg \min_{\boldsymbol{x}} \left[\boldsymbol{z} - \boldsymbol{h}\left(\boldsymbol{x}\right)\right]^T \boldsymbol{R}^{-1} \left[\boldsymbol{z} - \boldsymbol{h}\left(\boldsymbol{x}\right)\right]. \tag{2}$$

Let us apply the Gauss-Newton iterative algorithm [21] to solve for the state vector. Formally, we have

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \Delta\boldsymbol{x}^k, k = 1, 2, ..., \tag{3}$$

$$\Delta\boldsymbol{x}^k = (\boldsymbol{H}(\boldsymbol{x}^k)^T \boldsymbol{R}^{-1} \boldsymbol{H}(\boldsymbol{x}^k))^{-1} \boldsymbol{H}(\boldsymbol{x}^k)^T \boldsymbol{R}^{-1}(\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{x}^k)), \tag{4}$$

where $\boldsymbol{H}(\boldsymbol{x}^k) = \partial\boldsymbol{h}(x)/\partial\boldsymbol{x}|_{\boldsymbol{x}=\boldsymbol{x}^k} \in \mathbb{R}^{m\times n}$ is the Jacobian matrix. The algorithm converges once the norm of $\Delta\boldsymbol{x}^k$ is smaller than a pre-specified threshold. After estimation, the ℓ_2-norm detector is applied to detect the existence of bad data by checking if the following inequality holds [6], [21]:

$$\|\boldsymbol{r}\| = \|\boldsymbol{z} - \boldsymbol{h}\left(\widehat{\boldsymbol{x}}\right)\| \geq \tau, \tag{5}$$

where τ is a detection threshold of the ℓ_2-norm detector. Note that $\|\cdot\|$ is used to represent the ℓ_2-norm throughout the paper, where the subscript 2 in (5) has been dropped for simplicity.

B. Attack Model of the Nonlinear State Estimator

To perform an FDIA, we make the same assumptions as that in [6], [17], that is: i) an attacker could access the real-time measurements in a small area $\mathcal{S}$ bounded by buses, where the measurement and state indices in $\mathcal{S}$ are denoted as $\mathcal{M}_s$ and $\mathcal{I}_s$, respectively; ii) the hacker could change all the measurements in $\mathcal{S}$; iii) the hacker might have an a priori knowledge of the system topology, including the line parameters of the area $\mathcal{S}$. Thus, for the ith measurement, z_i, the attack model is

$$z_i^{(a)} = \begin{cases} z_i & \text{if } i \notin \mathcal{M}_s \\ z_i + a_i & \text{if } i \in \mathcal{M}_s \end{cases}, \tag{6}$$

where a_i is the ith element of the attack vector $\boldsymbol{a}$.

Lemma 1: Let us now assume that the hacker has obtained the same $z_i, i \in \mathcal{M}_s$ and $\widehat{x}_i$ as the operators of the control center. If the original measurement $z_i, i \in \mathcal{M}_s$, could bypass the ℓ_2-norm detector, the malicious measurement $z_i^{(a)}$ could also pass this detector under the condition $a_i = \boldsymbol{h}\left(\widehat{x}_i + c_i\right) - \boldsymbol{h}\left(\widehat{x}_i\right)$, where c_i represents the changes in the ith-attacked state variable.

Proof: Since we are interested in the area $\mathcal{S}$, the index i is omitted for simplicity. Because $\boldsymbol{z}$ can bypass the ℓ_2-norm detector, $\|\boldsymbol{r}\| = \|\boldsymbol{z} - \boldsymbol{h}\left(\widehat{\boldsymbol{x}}\right)\| \leq \tau$ holds. The ℓ_2-norm of the attacked measurement residual $\boldsymbol{r}_a$ is given by

$$\begin{aligned} \|\boldsymbol{r}_a\| &= \|\boldsymbol{z}^a - \boldsymbol{h}\left(\widehat{\boldsymbol{x}}_a\right)\| = \|\boldsymbol{z} + \boldsymbol{a} - \boldsymbol{h}\left(\widehat{\boldsymbol{x}} + \boldsymbol{c}\right)\| \\ &= \|\boldsymbol{z} + \boldsymbol{a} - \boldsymbol{h}\left(\widehat{\boldsymbol{x}} + \boldsymbol{c}\right) + \boldsymbol{h}\left(\widehat{\boldsymbol{x}}\right) - \boldsymbol{h}\left(\widehat{\boldsymbol{x}}\right)\| \\ &= \|\boldsymbol{r} + \boldsymbol{a} - \boldsymbol{h}\left(\widehat{\boldsymbol{x}} + \boldsymbol{c}\right) + \boldsymbol{h}\left(\widehat{\boldsymbol{x}}\right)\| \\ &= \|\boldsymbol{r}\| \leq \tau, \end{aligned} \tag{7}$$

which means that the attacked measurements could also avoid the detection. Note that $\widehat{\boldsymbol{x}}_a = \widehat{\boldsymbol{x}} + \boldsymbol{c}$ and $\boldsymbol{r} = \boldsymbol{z} - \boldsymbol{h}(\widehat{\boldsymbol{x}})$ is the measurement residual vector. ■

When implementing an FDIA for practical power systems, Lemma 1 intrinsically assumes that the hacker has enough computational capability to estimate the local state vector $\widehat{x}_i, i \in \mathcal{I}_s$ so that the attack vector $\boldsymbol{a} = \boldsymbol{h}\left(\widehat{x}_i + c_i\right) - \boldsymbol{h}\left(\widehat{x}_i\right)$ can be constructed. This assumption is acceptable given that the SCADA measurements are non-synchronized while the collection rates of measurements differ from one region to another one. In addition, a hacker may intentionally attack the communication system to delay the SCADA measurements for some parts of the system so that the local state x_i can be estimated by the hacker [6], [17].

C. Determining the Attack Graph of the Target Buses

Let $\mathcal{S} = \{\mathcal{B}, \Omega\}$ denote the attack graph, where $\mathcal{B}$ and Ω are the sets of buses and transmission lines, respectively; let $\mathcal{K}$ denote a set of bus indices for power injection buses, including the load and the generator buses. In [17], a topographical analysis was proposed to determine the attack graph of a single target bus. That approach is summarized below:

- *Step 1:* Let $i \in \mathcal{K}$ be the ith targeted power injection bus, the first step is to include bus i into the subgraph $\mathcal{S}_i$;
- *Step 2:* Extend $\mathcal{S}_i$ to include all the buses and branches Ω_i that are connected to bus i, where Ω_i is the set of adjacent branches connected to bus i;
- *Step 3:* If there exists any zero injection Bus j not connected to either the load or the generator on the boundary of $\mathcal{S}_i$, extend $\mathcal{S}_i$ to include Ω_j and continue Step 4; otherwise go to Step 5;

- *Step 4:* Repeat Step 3 until all buses on the boundary belong to the set $\mathcal{K}$;
- *Step 5:* Obtain the final attack subgraph as $\mathcal{S} = \underset{i\in\mathcal{K}}{\cup} \mathcal{S}_i$.

The above procedures can be simply summarized by the following lemma:

Lemma 2: For a bus provided with a power injection, its adjacent buses provided with power injections must be changed accordingly so that specified state changes can be made by the hacker. For a zero-injection bus, its adjacent power flows must sum to zero. This means that the measurements of the power injection buses adjacent to a zero power injection bus must be changed accordingly so that the equality constraints are satisfied. Consequently, buses that belong to $\mathcal{S}$ are bounded by buses in $\mathcal{K}$.

Using Lemma 2, we are able to determine the attack graph of multi-buses. The difference is that the size of the set $\mathcal{S}$ is increased with additional buses bounded by power injections.

D. Problem Statement

With the obtained attack graph $\mathcal{S}$ and all the assumptions stated in Section II-B satisfied, a perfect FDIA against a nonlinear state estimator can be achieved. However, because a hacker has typically a lack of real-time knowledge of the status of grid elements such as the position of circuit breaker switches and transformer tap changers, and also because she is restricted to access partial measurement channels, it is thus impossible for her to obtain the same state estimates as the operators of the control center in the attack graph $\mathcal{S}$. In other words, a perfect FDIA approach proposed in the literature [6], [17] seems to be impractical for realistic power systems. This is because with uncertain information of the system, $\widehat{x}_i, i \in \mathcal{I}_s$ obtained by the hacker is different from the state estimate $\widehat{x}_i^w, i \in \mathcal{I}_s$, and bias ζ_i exists, i.e., $\widehat{x}_i = \widehat{x}_i^w + \zeta_i$. Note that $\widehat{x}_i^w, i \in \mathcal{I}_s$ is the i-th state estimate calculated by the control center without an FDIA. As a result, when an FDIA occurs, the inequality constraint (7) may not hold true anymore. Interestingly, simulations carried out in [18], [20] reveal that even with some uncertain information about the system, FDIA can be successful without being detected by the control center. Furthermore, the expected changes on the target state variables are not equal to $\boldsymbol{c}$. However, no analytical investigations were carried out to explain why this imperfect attack can succeed and under which conditions it will be detected by the operators of the control center.

In this paper, an analytical investigation will be performed to show how the inequality constraint (7) can be satisfied in presence of system uncertainties to avoid the detection of an FDIA by the control center. In addition, we will quantify the maximum uncertainties a hacker can have so as to perform imperfect FDIA. The trade-off between attack magnitudes on the target state variables and the system uncertainties will be analyzed as well. Finally, to detect this type of FDIA, we will propose a measurement statistical consistency-based robust detector using a subset of secure PMU measurements.

Remark: To avoid the confusion between the bias terms $\boldsymbol{\zeta}$ and $\boldsymbol{c}$, we make the following clarifications: $\widehat{\boldsymbol{x}}$ is the estimated state vector before an FDIA and it is equal to $\widehat{\boldsymbol{x}}^w$ obtained by the control center if the hacker has the same information of the system as the control center. Otherwise, there is a difference between $\widehat{\boldsymbol{x}}$ and $\widehat{\boldsymbol{x}}^w$ caused by information uncertainties, which is the bias $\boldsymbol{\zeta}$. By contrast, $\boldsymbol{c}$ is the expected bias by the hacker when performing an FDIA.

III. Proposed Generalized FDIA Framework Against the Nonlinear State Estimator

An FDIA is in fact a type of perfect interacting and conforming bad data [22]. Therefore, the statistical tests applied to the weighted or the normalized residuals or the sum of the squared residuals (ℓ_2 detector) are unable to detect them. Without loss of generality, we consider in the sequel only the ℓ_2 detector when deriving the generalized FDIA framework.

In the developed generalized FDIA framework, we first provide a sufficient condition in Section III-A to justify theoretically how the imperfect FDIA can bypass the detector at the control center. This allows us to derive the upper bound of the uncertainties the hacker can have so as to launch a successful imperfect FDIA. Therefore, we are able to analyze the trade-off between attack magnitudes on the target state variables and the system uncertainties. To our best knowledge, this is the first attempt to provide theoretical justification to an imperfect FDIA and to quantify the tradeoff between the attack magnitude and the state bias caused by system uncertainties.

A. Sufficient Condition for an Imperfect FDIA

As clarified before, an adversary cannot obtain the same estimated state $\widehat{\boldsymbol{x}}^w$ (the subscript i is dropped for simplicity) as the operators of the control center. Here, we provide a sufficient condition for an imperfect FDIA to succeed subject to the state bias. This is shown in the following Lemma:

Lemma 3: If the true measurement residual $\|\boldsymbol{r}\| = \|\boldsymbol{z} - \boldsymbol{h}(\widehat{\boldsymbol{x}}^w)\| \leq \tau$ holds, a sufficient condition for the measurement $\boldsymbol{z}$ subject to attack $\boldsymbol{a}$ to pass ℓ_2 detector is

$$\|\boldsymbol{a} - \boldsymbol{h}(\widehat{\boldsymbol{x}}^w + \boldsymbol{c}) + \boldsymbol{h}(\widehat{\boldsymbol{x}}^w)\| \leq \gamma = \tau - \|\boldsymbol{r}\|. \tag{8}$$

Proof: When there are no bad data in the original measurements, $\|\boldsymbol{r}\| \leq \tau$ is always satisfied. The measurement residual under FDIA can be derived as

$$\begin{aligned}\|\boldsymbol{r}_a\| &= \|\boldsymbol{z}_a - \boldsymbol{h}(\widehat{\boldsymbol{x}}_a)\| = \|\boldsymbol{z} + \boldsymbol{a} - \boldsymbol{h}(\widehat{\boldsymbol{x}}^w + \boldsymbol{c})\| \\ &= \|\boldsymbol{z} + \boldsymbol{a} - \boldsymbol{h}(\widehat{\boldsymbol{x}}^w + \boldsymbol{c}) + \boldsymbol{h}(\widehat{\boldsymbol{x}}^w) - \boldsymbol{h}(\widehat{\boldsymbol{x}}^w)\| \\ &= \|\boldsymbol{r} + \boldsymbol{a} - \boldsymbol{h}(\widehat{\boldsymbol{x}}^w + \boldsymbol{c}) + \boldsymbol{h}(\widehat{\boldsymbol{x}}^w)\| \\ &\leq \|\boldsymbol{r}\| + \|\boldsymbol{a} - \boldsymbol{h}(\widehat{\boldsymbol{x}}^w + \boldsymbol{c}) + \boldsymbol{h}(\widehat{\boldsymbol{x}}^w)\| \leq \tau, \end{aligned} \tag{9}$$

which means that if the constraint (8) holds, an FDIA can not be detected by the residual statistical bad data detection test. ■

Although Lemma 3 looks straightforward, it provides a sufficient condition that the attack vector $\boldsymbol{a}$ should satisfy to avoid her detection by the operators. In addition, it serves as the foundation for the derivation of the upper bound of uncertainties the hacker can have when implementing a successful imperfect FDIA. On the other hand, it is easy to verify that a perfect attack with $\widehat{\boldsymbol{x}}^w = \widehat{\boldsymbol{x}}$ mentioned in Lemma 1 is just a special case here. Finally, since $\widehat{\boldsymbol{x}}^w$ is unknown to the hacker due to the limited knowledge of the system, conditions with the consideration of system uncertainties should be derived. That is, what is the tradeoff between system uncertainties and attack

magnitude? This question will be investigated and analyzed in Section III-B.

B. Tradeoff Between System Uncertainties and Attack Magnitude

Due to the existence of uncertain system information obtained by an attacker, the initial state vector used to construct an attack vector has uncertainties as well. This in turn yields biases on the target state variables. The larger uncertainties the hacker has, the less possibilities she can change the attack magnitudes, and vice versa. In other words, there exists a tradeoff between system uncertainties and attack magnitude. To quantify that, we define $\boldsymbol{a} = \boldsymbol{h}(\widehat{\boldsymbol{x}} + \boldsymbol{c}) - \boldsymbol{h}(\widehat{\boldsymbol{x}})$ and $\widehat{\boldsymbol{x}} = \widehat{\boldsymbol{x}}^w + \boldsymbol{\zeta}$. The ℓ_2-norm of the measurement residual becomes

$$\begin{aligned}\|\boldsymbol{r}_a\| &= \|\boldsymbol{z} + \boldsymbol{a} - \boldsymbol{h}\left(\widehat{\boldsymbol{x}}^w + \boldsymbol{c}\right)\| \\ &= \|\boldsymbol{z} + \boldsymbol{a} - \boldsymbol{h}\left(\widehat{\boldsymbol{x}}^w + \boldsymbol{c}\right) + \boldsymbol{h}\left(\widehat{\boldsymbol{x}}^w\right) - \boldsymbol{h}\left(\widehat{\boldsymbol{x}}^w\right)\| \\ &= \|\boldsymbol{z} - \boldsymbol{h}\left(\widehat{\boldsymbol{x}}^w\right) + \left(\boldsymbol{a} - \boldsymbol{h}\left(\widehat{\boldsymbol{x}}^w + \boldsymbol{c}\right) + \boldsymbol{h}\left(\widehat{\boldsymbol{x}}^w\right)\right)\| \\ &\le \|\boldsymbol{z} - \boldsymbol{h}\left(\widehat{\boldsymbol{x}}^w\right)\| + \|\boldsymbol{a} - \boldsymbol{h}\left(\widehat{\boldsymbol{x}}^w + \boldsymbol{c}\right) + \boldsymbol{h}\left(\widehat{\boldsymbol{x}}^w\right)\| \\ &= \|\boldsymbol{r}\| + \|\boldsymbol{h}\left(\widehat{\boldsymbol{x}} + \boldsymbol{c}\right) - \boldsymbol{h}\left(\widehat{\boldsymbol{x}}^w + \boldsymbol{c}\right) - \left(\boldsymbol{h}\left(\widehat{\boldsymbol{x}}\right) - \boldsymbol{h}\left(\widehat{\boldsymbol{x}}^w\right)\right)\|. \end{aligned} \tag{10}$$

Performing Taylor series expansions of $\boldsymbol{h}(\widehat{\boldsymbol{x}} + \boldsymbol{c})$ and $\boldsymbol{h}(\widehat{\boldsymbol{x}})$ at $\widehat{\boldsymbol{x}}^w + \boldsymbol{c}$ and $\widehat{\boldsymbol{x}}^w$, respectively, we obtain

$$\begin{aligned}&\boldsymbol{h}\left(\widehat{\boldsymbol{x}} + \boldsymbol{c}\right) - \boldsymbol{h}\left(\widehat{\boldsymbol{x}}^w + \boldsymbol{c}\right) \\ &\quad = \boldsymbol{h}\left(\widehat{\boldsymbol{x}}^w + \boldsymbol{c}\right) + \boldsymbol{H}_1\left(\widehat{\boldsymbol{x}} - \widehat{\boldsymbol{x}}^w - \boldsymbol{c}\right) + \boldsymbol{o}_1 - \boldsymbol{h}\left(\widehat{\boldsymbol{x}}^w + \boldsymbol{c}\right) \\ &\quad = \boldsymbol{H}_1\left(\boldsymbol{\zeta} - \boldsymbol{c}\right) + \boldsymbol{o}_1 \\ &\boldsymbol{h}\left(\widehat{\boldsymbol{x}}\right) - \boldsymbol{h}\left(\widehat{\boldsymbol{x}}^w\right) = \boldsymbol{h}\left(\widehat{\boldsymbol{x}}^w\right) + \boldsymbol{H}_2\left(\widehat{\boldsymbol{x}} - \widehat{\boldsymbol{x}}^w\right) + \boldsymbol{o}_2 - \boldsymbol{h}\left(\widehat{\boldsymbol{x}}^w\right) \\ &\quad = \boldsymbol{H}_2\boldsymbol{\zeta} + \boldsymbol{o}_2, \end{aligned} \tag{11}$$

where $\boldsymbol{H}_1 = \partial\boldsymbol{h}/\partial\boldsymbol{x}\,|_{\boldsymbol{x}=\widehat{\boldsymbol{x}}^w+\boldsymbol{c}}$ and $\boldsymbol{H}_2 = \partial\boldsymbol{h}/\partial\boldsymbol{x}\,|_{\boldsymbol{x}=\widehat{\boldsymbol{x}}^w}$ are Jacobian matrices; $\boldsymbol{o}_1$ and $\boldsymbol{o}_2$ are the higher order Taylor expansion terms. Since only the first order approximation is used in the WLS based state estimation algorithm, all the higher order terms are neglected during the iteration. In other words, $\boldsymbol{o}_1$ and $\boldsymbol{o}_2$ tend to 0 faster than the convergence of state estimation. Therefore,

$$\begin{aligned}&\|\boldsymbol{h}\left(\widehat{\boldsymbol{x}} + \boldsymbol{c}\right) - \boldsymbol{h}\left(\widehat{\boldsymbol{x}}^w + \boldsymbol{c}\right) - \left(\boldsymbol{h}\left(\widehat{\boldsymbol{x}}\right) - \boldsymbol{h}\left(\widehat{\boldsymbol{x}}^w\right)\right)\| \\ &\quad = \|\left(\boldsymbol{H}_1 - \boldsymbol{H}_2\right)\boldsymbol{\zeta} - \boldsymbol{H}_1\boldsymbol{c} + \left(\boldsymbol{o}_1 - \boldsymbol{o}_2\right)\| \\ &\quad \cong \|\left(\boldsymbol{H}_1 - \boldsymbol{H}_2\right)\boldsymbol{\zeta} - \boldsymbol{H}_1\boldsymbol{c}\| \\ &\quad \le \|\boldsymbol{H}_1 - \boldsymbol{H}_2\|\,\|\boldsymbol{\zeta}\| + \|\boldsymbol{H}_1\|\,\|\boldsymbol{c}\|. \end{aligned} \tag{12}$$

By combining (10) and (12), we get

$$\|\boldsymbol{r}_a\| \le \|\boldsymbol{r}\| + \|\boldsymbol{H}_1 - \boldsymbol{H}_2\|\,\|\boldsymbol{\zeta}\| + \|\boldsymbol{H}_1\|\,\|\boldsymbol{c}\|. \tag{13}$$

In order not to be detected by the operators of the control center, the right-hand side of (13) must be less than the detection threshold τ, that is,

$$0 \le \|\boldsymbol{H}_1 - \boldsymbol{H}_2\|\,\|\boldsymbol{\zeta}\| + \|\boldsymbol{H}_1\|\,\|\boldsymbol{c}\| \le \gamma. \tag{14}$$

The above equation shows the tradeoff between the attack magnitude and the estimation error $\|\boldsymbol{\zeta}\|$ of the state variables in the attack graph $\mathcal{S}$. Note that $\|\boldsymbol{\zeta}\|$ is caused by an uncertain information about the system state. If the estimation error is fixed, i.e., $\|\boldsymbol{\zeta}\| = \beta \neq 0$, the attack magnitude is bounded by

$$0 \le \|\boldsymbol{c}\| \le -\beta + \tfrac{\gamma + \|\boldsymbol{H}_2\|\beta}{\|\boldsymbol{H}_1\|}. \tag{15}$$

If $\|\boldsymbol{\zeta}\| = \beta = 0$, which means that the hacker can get exactly the same state estimates as the operators of the control center in the attack graph $\mathcal{S}$, the attack reduces to the perfect FDIA. The attack magnitude is bounded according to either (14) or (15) by setting $\beta = 0$. Formally, we have

$$0 \le \|\boldsymbol{c}\| \le \frac{\gamma}{\|\boldsymbol{H}_1\|}. \tag{16}$$

Remark: Note that the main scope of this paper is to investigate the vulnerability of a nonlinear WLS state estimator to imperfect FDIAs from the operator's perspective. To be more specific, given the current states estimated from the measurements and the assumed attack magnitudes, the operator knows matrices $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$. Then he can analyze how large uncertainties the hacker can have so that a success FDIA is launched under this condition. He may vary the assumed attack magnitudes to assess how the maximum uncertain information of the hacker changes if a successful FDIA is initiated. As a result, the vulnerability of the estimator can be assessed. On the other hand, due to the existence of uncertain system information and the limited access to measurements, the hacker is unable to know the exact matrices $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$. However, as long as the inequality (14) holds true, he can initiate successful FDIA with inexact $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$. This analysis can warn the operator to pay attention to the potential FDIA as the hacker is able to launch successful FDIA even with uncertain system information and limited measurements. To this end, corresponding effective countermeasures can be proposed.

IV. Proposed Robust FDIA Detector

In this section, we first present the motivations of designing a robust FDIA detector with a limited number of secure PMUs. The challenges and solution methodologies associated with the detector are discussed thoroughly. Then, the robust FDIA detector using measurement statistical consistency is proposed. To derive this detector, we enhance the data redundancy of the PMU measurements by short-term measurement forecasting, which allows us to handle noise and outliers in secure PMU measurements. We show through Theorem 1 that our robust state estimates follow a Gaussian distribution even when the PMU measurement errors are not normally distributed. This enables us to derive the Neyman-Pearson detector for an FDIA detection.

A. Motivations and Challenges

Recall that the hacker's objective is to change the estimated state vector by injecting malicious measurements. Once some measurements are compromised, the distribution of the estimated state vector will be perturbed by the attack [2], [14]. If one can find a set of measurements that will produce close approximations to the true estimated state vector and its probability distribution, then this statistical information can be further used

to double check the estimation results obtained from the remaining measurements. This strategy shares similar characteristics of the machine learning techniques, where partial data is used for learning and training while the remaining data is leveraged for validation. On the other hand, with the increasing deployment of PMUs in power systems, most of these systems are expected to be observable by PMUs. Indeed, many real systems have been observed by PMUs such as the Virginia Dominion Power [23], the 765/345/230 kV power grid in New York (NY), and the 345 kV power grid in New England (NE) [24], [25], to cite a few. In addition, the PMU observability of a given power system has been widely assumed in the literature, see [24], [29] for example. Thus, in this paper, we assume as suggested in [9] that the system is observable by a minimal set of PMUs that are made secure against cyber attacks. These PMUs that are usually installed at transmission system substations can be protected by encryptions, advanced fire walls and data package anomaly detectors, to name a few [26]–[28]. In addition, limited number of PMUs are assumed to be secure in the sense that they cannot be controlled by the hacker. On the other hand, unlike [9], we propose a robust detector based on an AC power system model that is able to handle outliers. Indeed, the authors in [9] make use of a DC not an AC state estimation model. Furthermore, they suppose that the PMU measurements are free of bad data, which is unrealistic in practice since impulsive communication noise and faulty GPS synchronization may corrupt the metered values. Note that strongly biased state estimates may result, which will mislead the operators of a control center; consequently, they may take wrong decisions based on them [30].

Remark: We assume the number of secure PMUs only guarantees the observability of the system, yielding no measurement redundancy. As a result, the PMU-based linear state estimator is not able to filter out noise as well as bad data. On the other hand, the state estimation model itself is an approximate model with uncertainties in the parameters, the topology and the measurements. Thus, a high measurement redundancy is required to reliably estimate the system state vector. This motivates us to validate and correct the remaining SCADA measurements, yielding improved state estimation results and system visualization.

B. Robust Detector Using Measurement Statistical Consistency

Similarly to the strategy proposed in [9], we assume in the proposed detector that the system is observed by a minimal set of secure PMU measurements. On the other hand, we advocate to enhance the data redundancy of PMUs by short-term measurement forecasting as proposed by Yacine *et al.* [31]. This allows our robust estimator to handle outliers in secure PMU measurements. It should be noted that it is in general challenging to forecast the future operating conditions due to many changing factors. However, we focus only on a very short-term forecast of the PMU metered variables, where the system operating conditions vary slowly, which is a reasonable assumption for practical power systems. Furthermore, system loads and renewable energy-based distributed generations change continuously from time to time, exhibiting temporal correlations. These changes in turn affect other generators and loads within the same geographic area, yielding spatial correlations. As a result, the nodal voltage and current phasors of the system exhibit similar statistical properties, which can be easily proved through the power flow equations. Thanks to these temporal and spatial correlations, we are able to use time series analysis technique to perform a short-term forecasting of the PMU metered variables. Interestingly, the temporal and spatial correlations of the nodal voltage and current phasors have been proved by [31] through field measurements. In that reference, an effective forecast of the PMU metered variables using a vector autoregressive model has been demonstrated as well. Following [31], we consider a vector autoregressive model of first order and dimension D at time instant k, i.e.,

$$\boldsymbol{y}_k = \boldsymbol{\Phi}_k \boldsymbol{y}_{k-1} + \boldsymbol{\varepsilon}_k, \tag{17}$$

where $\boldsymbol{y}_k \in \mathbb{R}^D$ is the vector of secure PMU measurements; $\boldsymbol{\Phi}_k \in \mathbb{R}^{D\times D}$ represents the transition matrix; $\boldsymbol{\varepsilon}_k \in \mathbb{R}^D$ is the Gaussian noise and $\boldsymbol{\varepsilon}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{S}_k)$, where $\boldsymbol{S}_k \in \mathbb{R}^{D\times D}$ is a non-diagonal error covariance matrix due to temporal and spatial correlations among PMU measurements. Using the Yule-Walker method, $\widehat{\boldsymbol{\Phi}}_k$ are estimated using historical measurements [31]. Then, the forecasted PMU metered values are obtained through $\boldsymbol{y}_k^f = \widehat{\boldsymbol{\Phi}}_k \boldsymbol{y}_{k-1}$, while its covariance matrix is given by $\boldsymbol{P}_k = \widehat{\boldsymbol{\Phi}}_k \boldsymbol{P}_{k-1} \widehat{\boldsymbol{\Phi}}_k{}^T + \boldsymbol{S}_k$, where $\boldsymbol{P}_{k-1}$ is the error covariance matrix of the filtered PMUs at time instant k–1. By processing the metered and the forecasted PMU values simultaneously and then performing the data prewhitening, we get the following regression form

$$\mathbf{z}_k = \boldsymbol{\mathcal{H}}_k \boldsymbol{x}_k + \boldsymbol{\eta}_k, \tag{18}$$

where $\mathbf{z}_k = \boldsymbol{L}_k^{-1} \left[(\boldsymbol{y}_k^s)^T \ \left(\boldsymbol{y}_k^f\right)^T \right]^T \in \mathbb{R}^l$ is the extended measurement vector that contains the forecasted measurements $\boldsymbol{y}_k^f$ and received PMU measurements $\boldsymbol{y}_k^s$; $l = 2D$; $\boldsymbol{L}_k$ is a matrix for prewhitening that is determined by applying a Cholesky decomposition to the augmented error covariance matrix $\boldsymbol{\Gamma} = \text{diag}[\boldsymbol{\Lambda}_k \ \boldsymbol{P}_k] = \boldsymbol{L}_k \boldsymbol{L}_k^T$. Here, $\boldsymbol{\Lambda}_k$ denotes the measurement error covariance matrix; $\boldsymbol{\mathcal{H}}_k = \boldsymbol{L}_k^{-1} [\boldsymbol{A}_k{}^T \ \boldsymbol{A}_k{}^T]^T$; $\boldsymbol{A}_k$ is a constant admittance matrix; $\boldsymbol{\eta}_k$ is the normalized error vector; and $\boldsymbol{\eta}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_k)$, where $\boldsymbol{I}_k$ is an identity matrix. Note that, only the current phasor measurement on a given line having an impedance that is very different from the rest of the lines can induce leverage point in the regression equation (18). By using the scaling technique proposed in [29], those leverage points can be eliminated and only vertical outliers are of concern.

Note that there always exists the case that system operation conditions vary abruptly, yielding unreliable predicted PMU measurements. In addition, the received PMU measurements can be corrupted with gross errors as well due to the impulsive communication noise, loss of communications, etc. To handle them while achieving good statistical efficiency, we advocate the use of the robust Huber-estimator [32]. This estimator minimizes the following objective function:

$$\mathcal{J}(\boldsymbol{x}) = \sum_{i=1}^{l} \rho(r_{S_i}), \tag{19}$$

where $\rho(\cdot)$ is the Huber convex cost function defined as:

$$\rho\left(r_{S_i}\right)=\begin{cases} r_{s_i}^2/2 & \text{for } |r_{S_i}|\le \lambda \\ \lambda\left|r_{S_i}\right|-\lambda^2/2 & \text{for } |r_{S_i}|>\lambda \end{cases}, \tag{20}$$

where the parameter λ is typically set to be between 1.5 and 3 for achieving high statistical efficiency when the measurement errors are Gaussian [32]; $r_{S_i}=r_i/s$ is the standardized residual; $r_i=\mathbf{z}_i-\boldsymbol{\alpha}_i^T\widehat{\boldsymbol{x}}$ and $\boldsymbol{\alpha}_i^T$ is the ith column vector of the matrix $\boldsymbol{\mathcal{H}}_k^T$; $s=1.4826\ c_m$ median $|r_i|$ is the robust scale estimate and c_m is a correction factor [33].

To solve (19), the following necessary condition must be satisfied:

$$\frac{\partial\mathcal{J}\left(\boldsymbol{x}\right)}{\partial\boldsymbol{x}}=\sum_{i=1}^{l}-\frac{\boldsymbol{\alpha}_i}{s}\psi\left(r_{S_i}\right)=\mathbf{0}, \tag{21}$$

where $\psi\left(r_{S_i}\right)=\partial\rho\left(r_{S_i}\right)/\partial r_{S_i}$. Multiplying and dividing r_{S_i} on both sides of (21), we obtain

$$\sum_{i=1}^{l}\boldsymbol{\alpha_i}\frac{\psi\left(r_{S_i}\right)}{r_{S_i}}\cdot\frac{r_{S_i}}{s}=\mathbf{0}, \tag{22}$$

which can be arranged in a matrix form as

$$\boldsymbol{\mathcal{H}}_k{}^T\boldsymbol{Q}\left(\mathbf{z}_k-\boldsymbol{\mathcal{H}}_k\,\widehat{\boldsymbol{x}}\right)=\mathbf{0}, \tag{23}$$

where $q\left(r_S\right)=\psi\left(r_S\right)/r_S$ and $\boldsymbol{Q}=\text{diag}\left(q\left(r_S\right)\right)$. Finally, using the iteratively re-weighted least square (IRLS) algorithm [32], the solution can be obtained through

$$\widehat{\boldsymbol{x}}_k^{(\ell+1)}=\left(\boldsymbol{\mathcal{H}}_k^T\boldsymbol{Q}^{(\ell)}\boldsymbol{\mathcal{H}}_k\right)^{-1}\boldsymbol{\mathcal{H}}_k^T\boldsymbol{Q}^{(\ell)}\mathbf{z}_k, \tag{24}$$

where ℓ is the iteration counter. The algorithm converges if

$$\left\|\widehat{\boldsymbol{x}}_k^{(\ell+1)}-\widehat{\boldsymbol{x}}_k^{(\ell)}\right\|_\infty\le\varsigma,\ e.g.,10^{-2}. \tag{25}$$

Theorem 1: The state estimation error by the Huber M-estimator above has an asymptotic normal probability distribution with zero mean and covariance matrix $\boldsymbol{V}_k$ given by

$$\boldsymbol{V}_k=\frac{\mathbb{E}\left[\psi^2(r_S)\right]}{\left(\mathbb{E}\left[\psi'(r_S)\right]\right)^2}\left(\boldsymbol{\mathcal{H}}_k^T\boldsymbol{\mathcal{H}}_k\right)^{-1}. \tag{26}$$

Proof: Let us define $T_l=T\left(F_l\right)$ as the statistic estimates of $T(F)$ and consider the ϵ-contaminated distribution $F_\epsilon=(1-\epsilon)F+\epsilon\delta_x$, where F is the true distribution while δ_x is the point mass at $\widehat{x}$ with an unknown distribution for the outliers or thick-tailed distributions. By virtue of the Glivenko-Cantelli theorem, $T\left(F_\varepsilon\right)\to T\left(F\right)$ as $\epsilon\to 0$. Taking Taylor series expansion on $\psi(r_{S_i};\widehat{x})$ of (21) about $\boldsymbol{x}$, we obtain

$$\begin{aligned}&\sqrt{l}\left\{\frac{1}{l}\sum_{i=1}^{l}\boldsymbol{\alpha}_i\cdot\psi\left(r_{S_i};\boldsymbol{x}\right)\right\}\\&+\boldsymbol{\mathcal{H}}_k^T\sqrt{l}\left(\widehat{\boldsymbol{x}}-\boldsymbol{x}\right)\left\{\frac{1}{l}\sum_{i=1}^{l}\psi'\left(r_{S_i};\boldsymbol{x}\right)\right\}\\&+\boldsymbol{\alpha}_i\cdot O_p\left(1/\sqrt{l}\right)=\mathbf{0},\end{aligned} \tag{27}$$

where $O_p(\cdot)$ represents the higher order error terms.

Using the central limit theorem, we have

$$\sqrt{l}\left\{\frac{1}{l}\sum_{i=1}^{l}\psi\left(r_{S_i};\boldsymbol{x}\right)\right\}\to_d\mathrm{Z}\sim\mathcal{N}\left(0,E_F\left[\psi(r_{S_i};\boldsymbol{x})^2\right]\right). \tag{28}$$

By the weak law of large numbers, we obtain

$$\left\{\frac{1}{l}\sum_{i=1}^{l}\psi'\left(r_{S_i};\boldsymbol{x}\right)\right\}\to_p E_F\left[\psi'\left(r_{S_i};\boldsymbol{x}\right)\right]. \tag{29}$$

Thus, by Slutsky's theorem, we get

$$\boldsymbol{\mathcal{H}}_k^T\sqrt{l}\left(\widehat{\boldsymbol{x}}-\boldsymbol{x}\right)\to_d\frac{-\mathrm{Z}}{\mathbb{E}_F\left[\psi'\left(r_{S_i};\boldsymbol{x}\right)\right]}\sim\mathcal{N}\left(0,\eta^2\right) \tag{30}$$

where $\eta^2=\frac{\mathbb{E}_F\left[\psi\left(r_{S_i};\boldsymbol{x}\right)^2\right]}{\mathbb{E}_F\left[\psi'\left(r_{S_i};\boldsymbol{x}\right)\right]^2}$. Thus, we can obtain

$$\boldsymbol{\mu}_k=\lim_{l\to\infty}\mathbb{E}\left[\sqrt{l}\left(\widehat{\boldsymbol{x}}-\boldsymbol{x}\right)\right]=\mathbf{0}, \tag{31}$$

$$\boldsymbol{V}_k=\lim_{l\to\infty}Var\left[\sqrt{l}\left(\widehat{\boldsymbol{x}}-\boldsymbol{x}\right)\right]=\frac{\mathbb{E}\left[\psi^2\left(r_S\right)\right]}{\left(\mathbb{E}\left[\psi'\left(r_S\right)\right]\right)^2}\left(\boldsymbol{\mathcal{H}}_k^T\boldsymbol{\mathcal{H}}_k\right)^{-1}, \tag{32}$$

which complete the proof. ■

After a state estimation is carried out, the estimated/interpolated SCADA measurements can be calculated through $\widehat{\boldsymbol{z}}=\boldsymbol{h}(\widehat{\boldsymbol{x}})$, where the subscript k is omitted for simplicity. Define the difference between the interpolated and the received SCADA measurements as a new residual, i.e., $\boldsymbol{v}=\boldsymbol{z}-\widehat{\boldsymbol{z}}$, we have the following theorem:

Theorem 2: With the assumption that the measurement errors of the received SCADA measurements follow a Gaussian distribution, i.e., $\boldsymbol{e}\sim\mathcal{N}(\mathbf{0},\boldsymbol{R})$ defined in (1), the new residual is normally distributed with zero mean and the covariance matrix $\boldsymbol{C}=\boldsymbol{R}+\boldsymbol{H}\boldsymbol{V}\boldsymbol{H}^T$, where $\boldsymbol{H}$ is the Jacobian matrix of the vector-valued function $\boldsymbol{h}(\cdot)$ evaluated at $\widehat{\boldsymbol{x}}$.

Proof: By taking the first order Taylor series expansion of $\boldsymbol{h}(\boldsymbol{x})$ at $\widehat{\boldsymbol{x}}$, the innovation vector $\boldsymbol{v}$ can be expressed as

$$\begin{aligned}\boldsymbol{v}&=\boldsymbol{z}-\widehat{\boldsymbol{z}}=\boldsymbol{h}\left(\boldsymbol{x}\right)+\boldsymbol{e}-\boldsymbol{h}\left(\widehat{\boldsymbol{x}}\right)\\&=\boldsymbol{h}\left(\widehat{\boldsymbol{x}}\right)+\boldsymbol{H}\left(\boldsymbol{x}-\widehat{\boldsymbol{x}}\right)+\boldsymbol{e}-\boldsymbol{h}\left(\widehat{\boldsymbol{x}}\right)\\&=\boldsymbol{H}\left(\boldsymbol{x}-\widehat{\boldsymbol{x}}\right)+\boldsymbol{e}.\end{aligned} \tag{33}$$

Thus, $\mathbb{E}\left[\boldsymbol{v}\right]=\boldsymbol{H}\mathbb{E}\left[\boldsymbol{x}-\widehat{\boldsymbol{x}}\right]+\mathbb{E}\left[\boldsymbol{e}\right]=\mathbf{0}$ and the covariance matrix is $\mathbb{E}\left[\boldsymbol{v}\boldsymbol{v}^T\right]=\boldsymbol{H}\text{cov}\left(\boldsymbol{x}-\widehat{\boldsymbol{x}}\right)\boldsymbol{H}^T+\boldsymbol{R}=\boldsymbol{H}\boldsymbol{V}\boldsymbol{H}^T+\boldsymbol{R}$. ■

Note that the asymptotic zero mean of the innovation vector $\boldsymbol{v}$ is valid under the condition that none of the received SCADA measurements is attacked. Otherwise, $\mathbb{E}\left[\boldsymbol{v}\right]=\boldsymbol{a}$, where $\boldsymbol{a}$ is the measurement bias injected by the hacker. Therefore, by checking the zero mean hypothesis of $\boldsymbol{v}$, we can determine whether an FDIA has been conducted or not. To this end, a binary hypothesis test on the measurement consistency can be developed; it is as follows:

$$\begin{cases}\mathcal{H}_0:\ \boldsymbol{v}\sim\mathcal{N}\left(\mathbf{0},\boldsymbol{C}\right)\\ \mathcal{H}_1:\ \boldsymbol{v}\sim\mathcal{N}\left(\boldsymbol{a},\boldsymbol{C}\right)\end{cases}, \tag{34}$$

where hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$ represent no FDIA and the occurrence of FDIA, respectively. By using the log-likelihood ratio test, we have

$$\boldsymbol{\xi} = \boldsymbol{v}^T \boldsymbol{C}^{-1} \boldsymbol{a} - \frac{1}{2} \boldsymbol{a}^T \boldsymbol{C}^{-1} \boldsymbol{a} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \kappa, \tag{35}$$

where κ is a decision threshold. Since $\boldsymbol{\xi}$ is a linear combination of $\boldsymbol{v}$, according to (34), it is expressed as

$$\begin{cases} \mathcal{H}_0 : \ \boldsymbol{\xi} \sim \mathcal{N}\left(\mu_0, \sigma_\xi^2\right) \\ \mathcal{H}_1 : \ \boldsymbol{\xi} \sim \mathcal{N}\left(\mu_1, \sigma_\xi^2\right); \end{cases} \tag{36}$$

and where

$$\begin{aligned} \mu_0 &= -\frac{1}{2} \boldsymbol{a}^T \boldsymbol{C}^{-1} \boldsymbol{a} \\ \mu_1 &= \frac{1}{2} \boldsymbol{a}^T \boldsymbol{C}^{-1} \boldsymbol{a} \\ \sigma_\xi^2 &= \boldsymbol{a}^T \boldsymbol{C}^{-1} \boldsymbol{a}. \end{aligned} \tag{37}$$

Therefore, given a false alarm rate P_{fa}, the relationship between the detection threshold κ and P_{fa} of the Neyman-Pearson detector is given by

$$P_{fa} = P\left(\xi \geq \kappa \,|\mathcal{H}_0\right) = \frac{1}{2} \text{erfc}\left(\frac{\kappa - \mu_0}{\sqrt{2}\sigma_\xi}\right), \tag{38}$$

where $\text{erfc}(\cdot)$ represents the complementary error function of the normal distribution. Thus, the threshold of the detector can be calculated as

$$\kappa = \sqrt{2}\sigma_\xi \text{erfc}^{-1}\left(2P_{fa}\right) + \mu_0. \tag{39}$$

Finally, the detection probability of an FDIA using the hypothesis testing is given by

$$P_d = P\left(\xi \geq \kappa \,|\mathcal{H}_1\right) = \frac{1}{2} \text{erfc}\left(\frac{\kappa - \mu_1}{\sqrt{2}\sigma_\xi}\right). \tag{40}$$

Remark: the aim of above analysis is to investigate the theoretical tradeoff between detection probability and false alarm of the proposed detector given the specified attack vector $\boldsymbol{a}$. In practice, the operator does not need to know the attack vector $\boldsymbol{a}$. Instead, the following equivalent normalized innovation vector-based statistical test is used:

$$\upsilon_{\text{Ni}} = \frac{|\upsilon_i|}{\sqrt{\boldsymbol{C}\left(i,i\right)}} \leqslant \Theta, \tag{41}$$

where υ_i and υ_{Ni} are the ith element of the innovation vector $\boldsymbol{v}$ and its normalized value, respectively; $\boldsymbol{C}(i,i)$ is the ith diagonal element of the derived covariance matrix; Θ is the detection threshold that is determined by the given confidence level of the Gaussian distribution, e.g., 3 for 99.7% confidence level. Note that we have proved that the innovation vector follows a Gaussian distribution with zero mean and the covariance matrix $\boldsymbol{C}$. The occurrence of the FDIA violates this fact and will be detected by our proposed robust detector with 99.7% confidence level.

V. Numerical Results

In this section, we use Sections V-A and V-B to demonstrate the validity of the proposed analytical FDIA framework against nonlinear state estimator, where the trade-off between attack magnitudes and information uncertainties is analyzed as well; by contrast, Sections V-C and V-D demonstrate the effectiveness of our robust FDIA detector in presence of measurement noise and bad data.

Specifically, extensive numerical simulations are carried out on the IEEE 30-bus and the 118-bus test systems. The measurement configurations of two test systems are as follows: 1) the IEEE 30-bus system is measured by 93 SCADA measurements, including 18 pairs of active and reactive power injections, 28 of pairs power flows and voltage magnitude of Bus 1; 2) the 118-bus system has 150 pairs of SCADA measurements, including 39 pairs of injection measurements and 111 pairs of flow measurements. The detailed measurement placements and topology of both test systems can be found in [34]. The meter errors of SCADA and PMU measurements follow the normal distribution with zero mean and standard deviations of 10^{-2} and 10^{-3}, respectively. The detection threshold for the normalized residual test is set to 3 with 99.7% confidence level. Two types of AC FDIAs are considered: i) *Perfect Attack:* the adversary does not have estimation errors corrupting the state variables in the attack graph $\mathcal{S} = \{\mathcal{N}, \Omega\}$; ii) *Imperfect Attack:* the adversary has estimation errors corrupting the state variables in the attack graph $\mathcal{S}$. One hundred Monte Carlo simulations are carried out to estimate the average value of the state estimation errors. The effectiveness of the proposed method will be first validated on the IEEE 30-bus system in Sections V-A– V-C, and then its scalability and robustness for larger-scale system will be tested using the 118-bus system .

A. Validation of the Imperfect State Variable Attack

A hacker is assumed to change the electricity consumption at Bus 26 through cyber attacks. To this end, she only needs to compromise the measurements P_{25-27}, Q_{25-27}, P_{25-24}, Q_{25-27}, P_{25} and Q_{25} so as to estimate V_{25} and θ_{25}. After that, the following two types of attacks are conducted without being detected by the operators of the control center.

Case 1: Perfect attack where the phase angle at Bus 26 is changed from $\theta_{26} = -0.2990$ to -0.0990 radians.

Case 2: Imperfect attack where the phase angle at Bus 26 is changed from $\theta_{26} = -0.2990$ to -0.0990 radians; here, the estimated state variables in the attack graph $\mathcal{S}$ have 5% errors, which are simulated using 100 Monte Carlo simulations. To be specific, a random sample taken from the uniform distribution $[-\Delta\tau_j + \widehat{x}_j \ \Delta\tau_j + \widehat{x}_j], j \in \Omega_i$ is used to represent the jth state variable to be estimated by the hacker; the final results are obtained by taking the average value of the 100 estimation errors.

Fig. 1 displays the results for Case 1 and Case 2. We can observe from this figure that the perfect attack has successfully changed the nonlinear SE results to the target value within 4 iterations. By contrast, the imperfect attack is not able to change the compromised state variable to the exact target value due to the uncertain knowledge of the system by the hacker. However, the difference between the hacker's target value and the

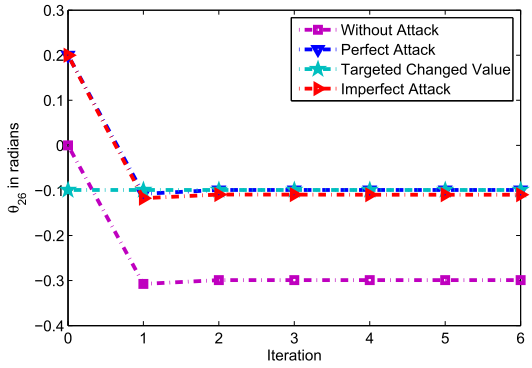

Fig. 1. Performance of perfect and imperfect FDIA over iterations of the nonlinear state estimator using weighted least squares algorithm.

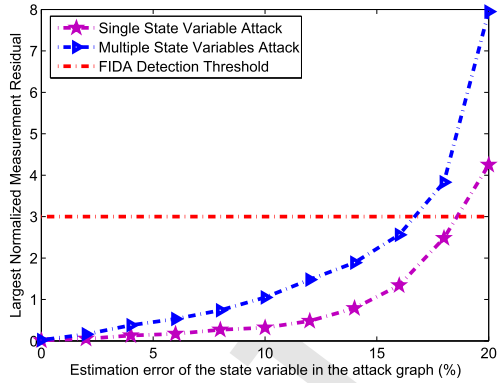

Fig. 2. Largest normalized residue vs. estimation error of the state variables in the attack graph when implementing imperfect attacks.

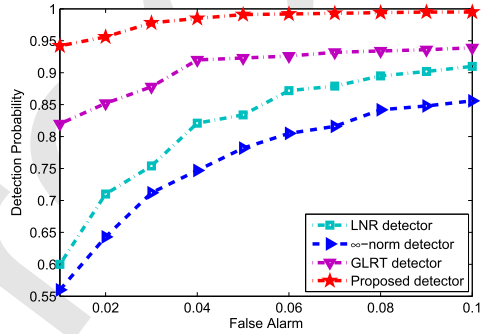

Fig. 3. Performance comparisons of different detectors for imperfect FDIA in Case 2.

achieved value is quite small. Therefore, it can be concluded that although the system information obtained by a hacker is inaccurate or contains uncertainties, she is capable of performing an imperfect attack with consequences that are close to her expectations without being detected by the traditional residual bad data detection tests.

B. Tradeoff Between Attack Magnitude and Uncertainties

It has been verified in Section V-A that a hacker is able to launch an imperfect attack to change some state variables close to her desire subject to system uncertainties. Now, what is the relationship between system uncertainties and the attack magnitude? To answer that question, let us first note that the system uncertainties will lead to estimation errors of the state variables in graph $\mathcal{S}$, which further affects the construction of the attack vector. To show this, we first resort to our theoretical results given by (14). It is clear that there is a tradeoff between the attack magnitude and the estimation errors of the state variables in graph $\mathcal{S}$. To further demonstrate this tradeoff, some simulation results are conducted and analyzed. To this end, we implement our imperfect FDIA to attack a single state variable and multiple state variables with varying errors in graph $\mathcal{S}$. The former case is similar to Case 2 while the latter one is similar to Case 3, which are imperfect attacks aimed at changing the phase angles at Buses 26 and 24 to -0.0990 and -0.1886 radians, respectively. The traditional normalized residual statistical test is used with a detection threshold of 3.

The test results are displayed in Fig. 2. It can be seen from Fig. 2 that with increased estimation errors of the state variables in the attack graph, the largest normalized residual continues to increase and finally exceeds the detection threshold. The largest estimation errors of the state variables in the attack graph when imperfect attacks are implemented successfully are 18.4% and 16.5% for single and multiple state attacks, respectively. This means that if a hacker is unable to estimate the state variables in the attack graph within a certain error tolerance, the attack will be detected. To show how this error tolerance threshold is affected by the attack magnitude, we consider changing θ_{26} and θ_{24} to -0.05 and -0.0886 radians, respectively. It is found that the error tolerance threshold decreases to 9.25% due to the increase of attack magnitude. Thus, the tradeoff between attack magnitude and system uncertainties is validated.

C. Detection of an FDIA on the Nonlinear State Estimator

As revealed in Fig. 2, if the estimation errors of the state variables in the attack graph $\mathcal{S}$ are less than 15% for Cases 2 and 3, the largest normalized residual is unable to reveal an FDIA. By contrast, our proposed detector can detect them with a high probability. To show this, we implement the proposed detector for both cases with a limited number of secure PMU measurements, say 10 PMUs, which are assumed to be installed at Buses 1, 7, 8, 10, 11, 12, 18, 23, 26 and 30. This is the minimal number of PMUs to observe the 30-bus system. All the measurements from these PMUs are assumed to be protected by the control center. The estimation errors of the state variables in the attack graph $\mathcal{S}$ vary randomly between 10% and 20%. 100 Monte Carlo simulations are carried out to assess the performance of each detector.

Figs. 3 and 4 show the receiver operating characteristic (ROC) curves for different detectors, including the generalized likelihood ratio test (GLRT)-based detector, infinite norm-based detector, and the largest normalized residual (LNR)-based detector [2].

It is observed from these two figures that the proposed detector outperforms the other three detectors. In particular, the infinite-norm, the LNR, and the GLRT detectors have more difficulties in detecting the single state variable attack than the multiple state variable attack. By contrast, our proposed detector is slightly affected. These results actually validate the fact that the more uncertainty the information a hacker obtains, the easier she will be detected by the operators of the control center. Thus, from

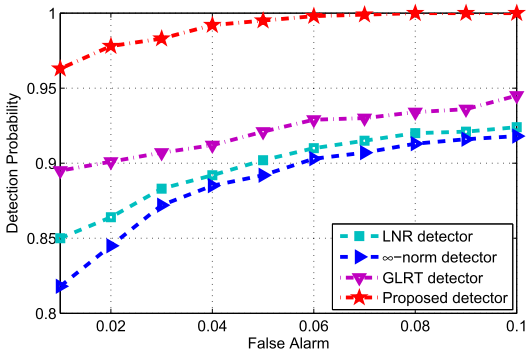

Fig. 4. Performance comparisons of different detectors for imperfect FDIA in Case 3.

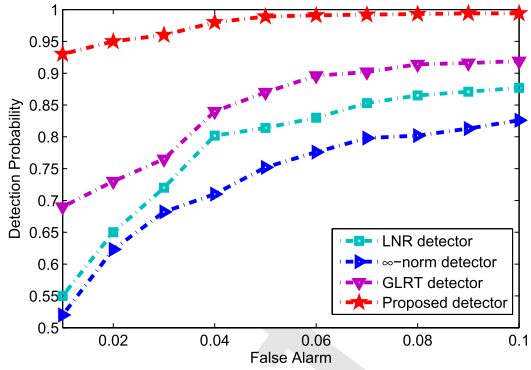

Fig. 5. Performance comparisons of different detectors for imperfect FDIA in IEEE 118-bus system.

the hacker's point of view, attacking a bus with less adjacent buses needs less information than other buses and therefore, will increase the probability of successful attack.

Note that a perfect FDIA is less likely to occur in practice, which represents certainly the worst case for the detectors at the control center. To evaluate the performance of our detector in this situation, the perfect attack of Case 1 is used. It is found that the infinite-norm, the LNR, and the GLRT detectors are unable to detect this attack. By contrast, the proposed detector is capable of detecting it with a similar performance as that of the imperfect attack of Case 2. Upon a closer look, this is not a surprising result. The reason is that by using the secure PMU measurements as well as the predicted metered values, a relative reliable set of system state information can be obtained; then the robust statistical test of measurement consistency is able to detect the compromised SCADA measurements with a high probability. Note that the secure PMU measurements and the predicted measurements are not from the same source or devices as the SCADA measurements, and thus are not affected by the attacks. In addition, as long as the attack magnitude on SCADA measurements exceed 3 times the standard deviation of the measurement error, FDIA will be detected by our detector with a probability of 95%.

D. Scalability and Robustness of the Proposed Method

To evaluate the scalability and the robustness of the proposed method, numerical tests are performed on the IEEE 118-bus system. It is assumed that 29 secure PMUs are deployed to make the system observable [9]; the adversary aims to attack the state variable θ_5 with 10 times the standard deviation error; the estimation errors of the state variables in the attack graph $\mathcal{S}$ are randomly varied between 5% and 15%. Note that only the single state attack scenario is tested. This is because from the hacker's point of view, she must try to launch a successful imperfect FDIA with as a small uncertainty as possible. In other words, she will be able to attack a small power system area with high confidence. The larger area she wants to attack, the more uncertainties about the system information are, yielding higher probability of being detected by the operators at the control center. Thus, from the operator's point of view, if the attacks on a small area with small uncertainties can be effectively detected, there is no need to worry about the risk of attacks on large areas with very high uncertainties. Fig. 5 displays the detection probability versus the

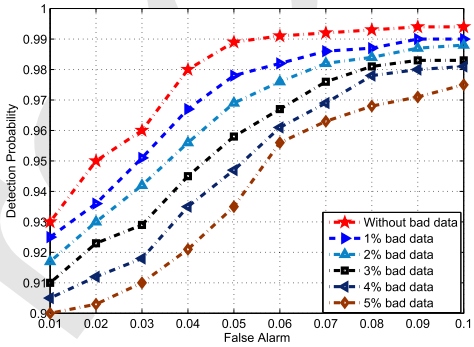

Fig. 6. Performance of the proposed detector for imperfect FDIA with various percentage of bad PMU measurements in IEEE 118-bus system.

false alarm probability for all the detectors. It can be observed that the performance of the proposed detector is superior to the other three methods. In addition, compared with the results on the 30-bus test system, our detector is slightly affected by the increased size of the test system.

To further investigate the impact of bad received and forecasted PMU measurements on the proposed detector, 1% to 5% of them are contaminated by adding errors with 10 to 15 times their standard deviations. The test results are shown in Fig. 6. Thanks to the robustness of the Huber M-estimator and the enhanced measurement redundancy with forecasted metered values, the influence of gross errors are bounded, yielding a slightly decreased performance of detecting attacks. However, the least detection probability of the proposed method is still greater than 90%. Note that for the bus with several adjacent buses, it has high local redundancy and the proposed detector can suppress several bad PMU measurements, while for those who have only one adjacent bus, it can handle fewer bad PMU measurements associated to that bus.

VI. Conclusion

The first contribution of this paper is to extend the existing perfect FDIA model by developing a generalized FDIA framework against nonlinear SE that accounts for the uncertainties in the measurements or in system topology. The upper bounds of these uncertainties for performing successful FDIA are investigated analytically. They provide the operators with a better understanding of the vulnerability of a nonlinear SE to FDIAs,

and thus may facilitate the adoption and implementation of effective defense methods. The second contribution of this paper is the development of a robust FDIA detection method that checks the measurement statistical consistency using a limited number of secure PMU measurements. Numerical results are provided to demonstrate the effectiveness and robustness of the proposed method. Future work will concentrate on the detection of topology attacks caused by the change of parameter values in the system. In addition, we will evaluate the sensitivity of the proposed approach to the accuracy of the forecasted PMU measurements and the change of system operation conditions in a short timeframe. Corresponding mitigation methods will be proposed if needed.

REFERENCES

[1] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Trans. Inf. Syst. Security*, vol. 14, no. 1, pp. 1–33, 2011.

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

[3] J. Kim and L. Tong, "On topology attack of a smart grid: Undetectable attacks and countermeasures," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1294–1305, Jul. 2013.

[4] Y. Chakhchoukh and I. Hiroyuki, "Coordinated cyber-attacks on the measurement function in hybrid state estimation," *IEEE Trans. Power Syst.*, vol. 30, no. 5, pp. 2487–2497, Sep. 2015.

[5] Y. Yuan, Z. Li, K. Ren, "Modeling load redistribution attacks in power systems," *IEEE Trans. Smart Grid*, vol. 2, no. 2, pp. 382–390, Jun. 2011.

[6] J. Liang, L. Sankar, and O. Kosut, "Vulnerability analysis and consequences of false data injection attack on power system state estimation," *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 3864–3872, Sep. 2016.

[7] S. Gao, L. Xie, A. Solar-Lezama, D. Serpanos, and H. Shrobe, "Automated vulnerability analysis of AC state estimation under constrained false data injection in electric power systems," in *Proc. IEEE 54th Conf. Decision Control*, 2015, pp. 1–8.

[8] R. B. Bobba, K. M. Rogers, Q. Wang, H. Khurana, T. J. Overbye, "Detecting false data injection attacks on DC state estimation," in *Proc. 1st Workshop Secure Control Syst.*, Stockholm, Sweden, 2010. pp. 226–231.

[9] T. T. Kim and H. V Poor, "Strategic protection against data injection attacks on power grids," *IEEE Trans. Smart Grid*, vol. 2, no. 2, pp. 326–333, Jun. 2011.

[10] J. Kim and L. Tong, "On phasor measurement unit placement against state and topology attacks," in *Proc. IEEE Int. Conf. Smart Grid Commun.*, 2013, pp. 396–401.

[11] M. Gol and A. Abur, "Effective measurement design for cyber security," in *Proc. IEEE Power Syst. Comput. Conf.*, 2014, pp. 1–8.

[12] L. C. Liu, M. Esmalifalak, Q. F. Ding, V. A. Emesih and H. Zhu, "Detecting false data injection attacks on power grid by sparse optimization," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 612–621, Mar. 2014.

[13] M. Esmalifalak, G. Shi, Z. Han, L. Y. Song, "Bad data injection attack and defense in electricity market using game theory study," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 612–621, Mar. 2013.

[14] J. B. Zhao *et al.*, "Short-term state forecasting-aided method for detection of smart grid general false data injection attacks," *IEEE Trans. Smart Grid*, vol. 8, no. 4, pp. 1580–1590, Jul. 2017.

[15] J. B. Zhao *et al.*, "Robust detection of cyber attacks on state estimators using phasor measurements", *IEEE Trans. Power Syst.*, vol. 32, no. 3, pp. 2468–2470, May 2017.

[16] Y. Chakhchoukh and H. Ishii, "Enhancing robustness to cyber-attacks in power systems through multiple least trimmed squares state estimations," *IEEE Trans. Power Syst.*, vol. 31, no. 6, pp. 4395–4405, Nov. 2016.

[17] G. Hug and J. A. Giampapa, "Vulnerability assessment of AC state estimation with respect to false data injection cyber-attacks," *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1362–1370, Sep. 2012.

[18] M. Rahman and H. Mohsenian-Rad, "False data injection attacks against nonlinear state estimation in smart power grids," in *Proc. IEEE Power Eng. Soc. General Meeting*, 2013, pp. 1–5.

[19] J. B. Zhao, G. X. Zhang, Z. Y. Dong, and K. P. Wong, "Forecasting-aided imperfect false data injection attacks against power system nonlinear state estimation," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 6–8, Jan. 2016.

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

[21] A. Abur and A. Gomez Exposito, *Power System State Estimation: Theory and Implementation*. New York, NY, USA: Marcel Dekker, 2004.

[22] L. Mili, T. Van Cutsem, and M. Pavella, "Bad data identification methods in power system state estimation—A comparative study," *IEEE Trans. Power App. Syst.*, vol. PAS-104, no. 11, pp. 3037–3049, Nov. 1985.

[23] K. D. Jones, J. S. Thorp, and R. M. Gardner, "Three-phase linear state estimation using phasor measurements," in *Proc. IEEE PES General Meeting*, 2013, pp. 1–5.

[24] E. R. Fernandes *et al.*, "Application of a phasor-only state estimator to a large power system using real PMU data," *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 411–420, Jan. 2017.

[25] S. G. Ghiocel *et al.*, "Phasor-measurement-based state estimation for synchrophasor data quality improvement and power transfer interface monitoring," *IEEE Trans. Power Syst.* vol. 29, no. 2, pp. 881–888, Mar. 2011.

[26] B. Sikdar and J. Chow, "Defending synchrophasor data networks against traffic analysis attacks," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 819–826, Dec. 2011.

[27] S. Cui, Z. Han, S. Kar, T. T. Kim, H. V. Poor, and Ali Tajer, "Coordinated data-injection attack and detection in the smart grid: A detailed look at enriching detection solutions," *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 106–115, Sep. 2012.

[28] S. Pal, B. Sikdar, and J. Chow, "Real-time detection of packet drop attacks on synchrophasor data," in *Proc. IEEE Int. Conf. Smart Grid Commun.*, 2014, pp. 896–901.

[29] M. Gol and A. Abur, "LAV based robust state estimation for systems measured by PMUs," *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 1808–1814, Jul. 2014.

[30] K. C. Sou, H. Sandberg, and K. H. Johansson, "Data attack isolation in power networks using secure voltage magnitude measurements," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 14–28, Jan. 2014.

[31] Y. Chakhchoukh, V. Vittal, and G. T. Heydt, "PMU based state estimation by integrating correlation," *IEEE Trans. Power Syst.*, vol. 29, no. 2, pp. 617–626, Mar. 2014.

[32] P. J. Huber, *Robust Statistics*. New York, NY, USA: Wiley, 1981.

[33] L. Mili, M. Cheniae, N. Vichare, and P. Rousseeuw, "Robust state estimation based on projection statistics," *IEEE Trans. Power Syst.*, vol. 11, no. 2, pp. 1118–1127, May 1996.

[34] A. Tarali and A. Abur, "Bad data detection in two-stage state estimation using phasor measurements," in *Proc. 3rd IEEE PES Innovative Smart Grid Technol. Conf.*, 2012, pp. 1–8.

Junbo Zhao (S'13) received the Bachelor's degree in electrical engineering from Southwest Jiaotong University, Chengdu, China, in 2012. He is currently working toward the Ph.D. degree in Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Falls Church, VA, USA. He did a summer internship at Pacific Northwest National Laboratory from May to August 2017. He is now the Chair of the IEEE Task Force on Power System Dynamic State and Parameter Estimation, the secretary of the IEEE Working Group on State Estimation Algorithms, and the IEEE Task Force on Synchrophasor Applications in Power System Operation and Control. His research interests include power system real-time monitoring, operations, and security.

Lamine Mili (LF'17) received the Electrical Engineering Diploma from the Swiss Federal Institute of Technology, Lausanne, Switzerland, in 1976, and the Ph.D. degree from the University of Liège, Liège, Belgium, in 1987. He is currently a Professor with the Department Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Falls Church, VA, USA. He has five years of industrial experience with the Tunisian electric utility, STEG. He was a Visiting Professor with the Swiss Federal Institute of Technology, Lausanne, the Grenoble Institute of Technology, the École Supérieure D'électricité, France, and the École Polytechnique de Tunisie, Tunisia, and did consulting work for the French Power Transmission company, RTE. His research has focused on power system planning for enhanced resiliency and sustainability, risk management of complex systems to catastrophic failures, robust estimation and control, nonlinear dynamics, and bifurcation theory.

Meng Wang (M'12) received the B.S. and M.S. degrees from Tsinghua University, Beijing, China, and the Ph.D. degree from Cornell University, Ithaca, NY, USA, in 2012. She is currently an Assistant Professor with the Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA. Her research interests include high-dimensional data analysis and their applications in power systems monitoring and network inference.