Sparse Malicious False Data Injection Attacks and Defense Mechanisms in Smart Grids

Jinping Hao, *Student Member, IEEE*, Robert J. Piechocki, *Member, IEEE*, Dritan Kaleshi, *Member, IEEE*, Woon Hau Chin, *Senior Member, IEEE*, and Zhong Fan

***Abstract*—This paper discusses malicious false data injection attacks on the wide area measurement and monitoring system in smart grids. First, methods of constructing sparse stealth attacks are developed for two typical scenarios: 1) random attacks in which arbitrary measurements can be compromised; and 2) targeted attacks in which specified state variables are modified. It is already demonstrated that stealth attacks can always exist if the number of compromised measurements exceeds a certain value. In this paper, it is found that random undetectable attacks can be accomplished by modifying only a much smaller number of measurements than this value. It is well known that protecting the system from malicious attacks can be achieved by making a certain subset of measurements immune to attacks. An efficient greedy search algorithm is then proposed to quickly find this subset of measurements to be protected to defend against stealth attacks. It is shown that this greedy algorithm has almost the same performance as the brute-force method, but without the combinatorial complexity. Third, a robust attack detection method is discussed. The detection method is designed based on the robust principal component analysis problem by introducing element-wise constraints. This method is shown to be able to identify the real measurements, as well as attacks even when only partial observations are collected. The simulations are conducted based on IEEE test systems.**

***Index Terms*—Bad data detection (BDD), malicious data attack, robust principle component analysis (PCA), smart grid security.**

I. Introduction

COMPARED with the traditional power grids, a smart grid tends to be much more reliable, efficient, and intelligent due to the remarkable advancements in sensing, monitoring, control technologies, and also the tight integration with cyber infrastructure and advanced computing and communication technologies [1]. However, this integration can lead to new vulnerabilities to cyber attacks on the power systems. Cyber attacks are reported as one of the main potential threats to the reliable operation of the power system [2], [3]. In this paper, we consider false data injection attacks (FDIAs) against the supervisory control and data acquisition (SCADA) system in smart grids.

A power grid transmission system is a sophisticated network which connects a number of electric power generators to various consumers through extensive power lines. It is extremely important to monitor the state of this complex system such that various control and planning tasks can be performed and the reliable operation of the power system is guaranteed. In power systems, state estimation [4], [5] is used to estimate system state variables through a number of power measurements and is a useful and necessary function in energy management systems (EMS).

The SCADA system obtains power status information such as transmission line power flows, bus voltages, and also circuit-braker signals through remote terminal units (RTUs). These measurements are then used for the state estimation process in EMS, which builds real-time electricity network models. In smart grids, the complex network connections as well as the Internet make SCADA systems susceptible to potential FDIAs, in which adversaries aim to contaminate the measurements collected from RTUs and bias the state estimation at the transmission level to mislead the operation of the power system. Fig. 1 presents a block diagram of the power grid, communication network, SCADA, and control center. It is critically important to understand the behavior of adversaries so that appropriate countermeasures can be designed to either protect the system from attacks beforehand or identify the malicious false data injections in the measurements.

Recently, the problem of FDIAs as well as countermeasures has attracted a lot of attention among researchers. False data in state estimation were first discussed by Schweppe *et al.* in their pioneering work about state estimation [6]. It was not well researched until Liu *et al.* [7] proposed that if adversaries possess the knowledge of power grid topology, they may inject coordinated data attacks, which could evade detection by the bad data detection (BDD) system in state estimator. Based on this strategy, plenty of efforts have been made to design effective attack algorithms and the corresponding countermeasures, such as [8]–[11].

Adversaries may launch attacks through hacking RTUs such as sensors in substations. In consideration of the accessibility of RTUs and also hacking cost, attackers always tend to control only a few RTUs to implement a successful attack [7]. Kim and Poor [8] developed a general optimization framework-based formulation for constructing sparse attack vectors when a subset of measurements is protected, while Ozay *et al.* [12]

Manuscript received September 12, 2014; revised April 17, 2015 and June 08, 2015; accepted August 07, 2015. Date of publication August 31, 2015; date of current version October 02, 2015. This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) Dorothy Hodgkin Postgraduate Award Scholarship. Paper no. TII-14-1020.

J. Hao and R. J. Piechocki are with the Department of Electrical and Electronic Engineering, University of Bristol, Bristol BS8 1UB, U.K. (e-mail: j.hao@bristol.ac.uk; r.j.piechocki@bristol.ac.uk).

D. Kaleshi is with Digital Catapult, London NW1 2RA, U.K. (e-mail: Dritan.Kaleshi@cde.catapult.org.uk).

W. H. Chin and Z. Fan are with the Telecommunications Research Laboratory, Toshiba Research Europe Ltd., Bristol BS1 4ND, U.K. (e-mail: woonhau.chin@toshiba-trel.com; zhong.fan@toshiba-trel.com).

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/TII.2015.2475695

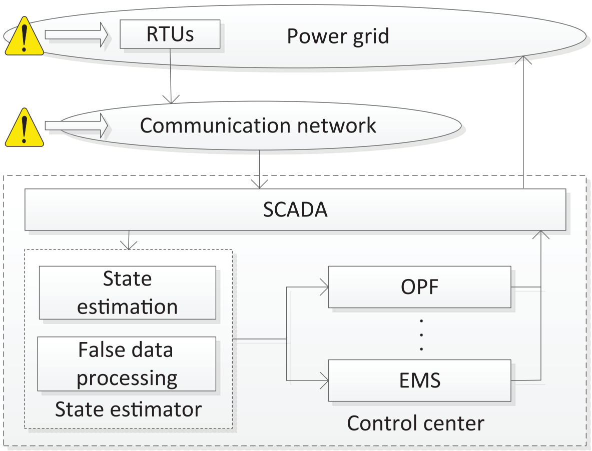


Fig. 1. Illustration of the power grid, communication network, SCADA, and control center. The warning signs indicate the vulnerabilities to FDIAs.

extended the sparse attack construction model to a distributed framework. Sandberg *et al.* [13] considered sparse attacks with injections into critical measurements, which are essential for the observability of the power grids and sensitive to attacks. In [14], methods of finding both strong stealth attacks and also optimal weak malicious data attacks (when power grid topology is unknown) with the aim of reducing the number of compromised measurements were discussed. It has been proposed that even when the power grid information is unavailable, stealth attacks can still be accomplished [15], [16]. Zhang *et al.* [17] investigated the attack strategy with consideration of communication rate constraint in cyber–physical systems.

There are two approaches to defend against the malicious data attacks. The first is to protect the system beforehand from being attacked by adversaries. This can be achieved by either protecting a number of measurements to prevent stealth attacks [18], [19], or monitoring state information directly by the deployment of phasor measurements units (PMU) [8], [20]. In practice, it is not feasible to secure all measurements to prevent attacks due to the high cost. Instead, stealth malicious attacks can be prevented by protecting a carefully selected subset of measurements. A challenge of this approach is to search the effective small measurement subset to make them immune to attacks. Bobba *et al.* [18] chose the subsets for small power test systems using brute-force search.

The second approach to deal with malicious attacks is to identify the injected false data in measurements and then either abandon the contaminated data or correct them. Traditional false data detection methods are based on residue test [6], [21]. They cannot protect state estimation from carefully designed stealth attacks. Recently, with the advancement of smart grid, new detection methods have been proposed. A survey of the existing detection methods was provided in [22]. In [14], generalized likelihood ratio test is introduced to detect weak FDIAs. The cumulative sum (CUSUM) test-based detection mechanism introduced in [16] is also designed for nonstealth attacks. Esmalifalak *et al.* [23] discussed stealth false data detection methods using machine learning. Graphical methods are used to design defending mechanisms in [24]. In [11], an effective method capable of detecting false data as well as recovering the real state information was proposed. In [25], both the attack and detection algorithms were discussed, but the subset protection method was not considered. Additionally, only preliminary results regarding the attack strategies were presented in [25]. This paper substantially discusses the methods of sparse attack construction, the strategy of system protection from attacks, and the algorithm of stealth attack detection.

This paper has made three contributions: First, methods of constructing stealth attacks are proposed for two typical scenarios. We consider a general scenario in which adversaries can access arbitrary measurements to change arbitrary state variables in state estimation. To the best of our knowledge, there is no feasible algorithm that can efficiently construct highly sparse undetectable attack vectors in this case. In [7], it is observed that the optimal undetectable attack vector to compromise the minimum number of measurements can be found using brute-force search. However, this is not practical due to the high complexity. We propose an efficient and effective attack vector construction algorithm which can quickly generate highly sparse attack vectors in this scenario.

Liu *et al.* [7] also have demonstrated that stealth attack vectors always exist when the number of measurements that can be contaminated exceeds a certain value. However, it is shown in this paper that our proposed method can launch stealth attacks by manipulating only a much smaller number of measurements with high probability. Additionally, stealth attacks in a specific scenario are also considered in this paper. An optimization-based algorithm is introduced to generate sparse targeted attack vectors to bias specified state variables with the consideration that a subset of measurements is protected.

A fast greedy search method is then proposed to quickly find a subset of measurements to be protected to defend against stealth attacks. This fast method can find a subset with the same size as brute-force search in nearly all cases. Finally, inspired by Liu *et al.* [11], we introduce a detection algorithm considering the noise case with partial observations. The proposed algorithm extends the method in [11] to address the problem of detecting stealth attacks as well as recovering true state information with only partially collected contaminated measurements. The performance of the proposed algorithms is investigated using IEEE test systems with software MATPOWER [26].

This paper is organized as follows. Section II introduces power system model and the stealth attack problem in state estimation. In Section III, proposed attack strategies for different scenarios are introduced. Section IV provides the measurement protection algorithm and Section V presents the false data detection method. Simulation results are presented in Section VI. Section VII concludes the paper.

II. Power Grid and Attack Model

In this paper, we consider a power transmission grid which consists of $n+1$ buses and l transmission lines. The network connectivity of this power grid can be described by the $(n+1)\times l$ oriented incidence matrix $\mathbf{M}$, of which each column corresponds to the power line (i,j) and consists of all

0s except the ith and jth elements having value of 1 and -1, respectively [27]. The nonsingular diagonal matrix $\mathbf{D} \in \mathbb{R}^{l \times l}$ describes the physical properties of the transmission grid with diagonal entries equal to admittance of the transmission lines. SCADA collects measurements from RTUs such as bus voltages, bus power injections, and branch power flows from the power grid and sends them to the state estimator to estimate the state of the power system in the control center.

A. State Estimation

The state estimation problem is to use power measurements to timely estimate the state of the power system. Specifically, power system state refers to bus voltage angles θ and bus voltage magnitudes V. In the linearized dc power flow model [4 Ch. 2], bus magnitudes are assumed already known and are all close to unity. Additionally, phase angle at reference bus is set to zero radians; thus, estimation of only n bus voltage angles $[\theta_1, \theta_2, \ldots, \theta_n]^T$ is required. The measurements have the following relationship with the state variables:

$$\boldsymbol{z} = \mathbf{H}\theta + \mathbf{e} \tag{1}$$

where $\boldsymbol{z} = [z_1, z_2, \ldots, z_m]^T$ denotes measurement vector and $\mathbf{H} \in \mathbb{R}^{m \times n}$ is the measurement Jacobian matrix constructed by $\mathbf{H} = \begin{bmatrix} \mathbf{DM}^T \\ \mathbf{MDM}^T \end{bmatrix}$. $\mathbf{e} = [e_1, e_2, \ldots, e_m]^T$ represents the Gaussian measurement noise and it is assumed to be zero-mean for convenience [14]. Measurements include power flows on l transmission lines and power injection at n buses. In the dc power flow model, power from bus i to j can be approximated as

$$P_{ij} = (\theta_i - \theta_j)\, d_{ij} \tag{2}$$

where d_{ij} is the admittance of the transmission line from bus i to j. Thus, power flows on transmission lines are computed by $\mathbf{MDM}^T\theta$ and power injections at buses are obtained from $\mathbf{DM}^T\theta$.

The state vector can be estimated from measurements using the weighted least-square (WLS) method [4]. In particular, system states are estimated as

$$\hat{\theta} = \left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{W}\boldsymbol{z} = \mathbf{K}\boldsymbol{z} \tag{3}$$

where diagonal weighting matrix $\mathbf{W}$ has diagonal entries equaling to the inverse of noise variances and $\mathbf{K} = \left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{W}$.

B. False Data Injection Attacks

Malicious false data can be injected by manipulating the RTU measurements to bias the estimated states. System measurements with malicious data becomes

$$\boldsymbol{z}_{\mathbf{a}} = \mathbf{H}\theta + \boldsymbol{a} + \mathbf{e} \tag{4}$$

where $\boldsymbol{a} = [a_1, a_2, \ldots, a_m]^T$ denotes the attack vector.

Bad data in measurements can lead to incorrect state estimation and cause severe outcomes. Traditional methods to detect bad data are mostly based on the residue test. The residue vector $\boldsymbol{r}$ refers to the difference between the obtained measurements and the computed value from the estimated state

$$\boldsymbol{r} = \boldsymbol{z} - \mathbf{H}\hat{\theta}. \tag{5}$$

For example, the largest normalized residue (LNR) test identifies bad data if the absolute value of the largest element in $\boldsymbol{r}$ is larger than a certain threshold τ, i.e., $\max_i |r_i| > \tau$.

However, carefully designed malicious data attacks can bypass residue-based BDD. If attackers have knowledge about the power grid topology information, or $\mathbf{H}$, they can inject stealth attacks by constructing the attack vector such that [7]

$$\boldsymbol{a} = \mathbf{H}\mathbf{c}. \tag{6}$$

The measurements can then be written as

$$\boldsymbol{z}_{\mathbf{a}} = \mathbf{H}\left(\theta + \boldsymbol{c}\right) + \mathbf{e} \tag{7}$$

where $\boldsymbol{c} \in \mathbb{R}^n$ is any arbitrary vector and denotes the errors added to the state variables introduced by $\boldsymbol{a}$. The attack is undetectable as the residue $\boldsymbol{r}$ would not change compared to that without attack $\boldsymbol{a}$ [7]. The system will regard the manipulated state $(\theta + \boldsymbol{c})$ as the real value in the state estimator.

III. Stealth Attack Strategies

In order to evade detection in the control center, attack vectors are designed to satisfy (6). Additionally, attackers would tend to compromise as few measurements as possible in effort to launch attacks with least effort. Therefore attack strategies are expected to be able to construct highly sparse attack vectors. The stealth sparse attacks were first discussed in [7], in which the authors proposed that attackers can modify state variables in state estimation without being detected by modifying a small number of carefully chosen RTU measurements. In this paper, two methods are introduced to construct sparse attack vectors for two typical scenarios: random attacks in which arbitrary measurements can be compromised and targeted attacks in which specific state variables need to be biased.

A. Random Attacks

In this scenario, it is assumed that no measurements are protected, and the changes of state variables are not specified. Attackers can hack arbitrary measurements to bias arbitrary state variables. Thus, the aim is solely to find highly sparse vector $\boldsymbol{a}$ that satisfies (6). To the best of our knowledge, there is no feasible algorithm that can efficiently construct sparse attack vectors in this case. Since $\boldsymbol{a}$ is a linear combination of the columns of $\mathbf{H}$, it is possible to generate sparse vector by column transformation of $\mathbf{H}$. However, this method cannot guarantee sparsity. It is demonstrated by Liu *et al.* [7] that a k-sparse stealth attack vector always exist if $k > m - n$. We propose a novel method that can construct a sparse attack vector with much smaller k.

Liu *et al.* [7] also proposed that a projection matrix $\mathbf{P} = \mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T$ can lead to an equivalent criterion to generate attack vector $\boldsymbol{a}$ satisfying (6)

$$\mathbf{P}\boldsymbol{a} = \mathbf{H}\left(\mathbf{H^T H}\right)^{-\mathbf{1}}\mathbf{H^T H}\boldsymbol{c} = \mathbf{H}\boldsymbol{c}$$
$$(\mathbf{P} - \mathbf{I})\boldsymbol{a} = 0.$$

Let $\mathbf{B} = \mathbf{P} - \mathbf{I}$, then undetectable attack vector $\boldsymbol{a}$ satisfies

$$\mathbf{B}\boldsymbol{a} = 0. \tag{8}$$

This criterion can be used to generate attack vectors in certain scenarios such as when a subset of measurements is protected [7]. For random attacks, this criterion can also be utilized. A straightforward way to find sparse attack vectors can be formulated using the following optimization problem:

$$\begin{aligned} &\min \ \|\boldsymbol{a}\|_0 \\ &\text{s.t. } \mathbf{B}\boldsymbol{a} = 0, \ \boldsymbol{a} \neq 0 \end{aligned} \tag{9}$$

where l_0-norm is the number of the nonzero elements in $\boldsymbol{a}$. This is a nonconvex problem and finding the solution to this problem is highly complex. However, it is obvious that a feasible vector $\boldsymbol{a}$ must be in the null space of matrix $\mathbf{B}$, which is defined as

$$\text{Null}(\mathbf{B}) = \left\{\mathbf{v} \in \mathbb{R}^m \left| \mathbf{B}\mathbf{v} = 0\right.\right\}. \tag{10}$$

Rather than solving the complex problem in (9), we propose an algorithm taking advantage of null space of $\mathbf{B}$, which can be easily computed.

Proposed algorithm. Measurements are always subject to noise due to the errors in the measuring process and the noise in the communication channels. The noise can be modeled as Gaussian distributed with variance Σ_e. The system is usually designed to be tolerant to measurement deviations within the noise level. Additionally, vectors in the null space of $\mathbf{B}$ usually comprise a small number of relatively large elements and the majority are small value elements. It provides the possibility for attackers to inject attack vectors designed based on vectors in $\text{Null}(\mathbf{B})$. Those small valued elements can be dealt with as noise if their average energy, denoted as Σ_B, is within the range of the variance of the noise Σ_e, namely $\Sigma_B \leq \Sigma_e$. Therefore, attackers only need to inject elements of large values in the column vectors of $\text{Null}(\mathbf{B})$ into the system and the number of measurements to be compromised will be greatly reduced. We define a shrinkage operation $\mathcal{S}_t$ as follows:

$$\mathcal{S}_t(x) := \begin{cases} \dfrac{x}{|x| - t}\max\left(|x| - t, \, 0\right), & |x| \neq t \\ x, & |x| = t. \end{cases} \tag{11}$$

The attack vector construction procedure can then be designed as follows. Given matrix measurement matrix $\mathbf{H}$, compute matrix $\mathbf{B}$ as well as the standard basis matrix $\mathbf{U}$ of its null space $\text{Null}(\mathbf{B})$ and choose vector $\boldsymbol{u}$ with the largest variance from all column vector in $\mathbf{U}$

$$\boldsymbol{u} = \arg\max_i\left(\text{var}\left(\mathbf{u}_i\right)\right) \tag{12}$$

where $\mathbf{u}_i$ denotes the ith column in $\mathbf{U}$. Then, scale vector $\boldsymbol{u}$ up, or down, till the maximal element reaches a designed attack value C. The last step is to force the small elements below threshold t to 0

$$\mathbf{a} = \mathcal{S}_t\left(\epsilon\boldsymbol{u}\right) \tag{13}$$

where $\epsilon = \frac{C}{\max(\boldsymbol{u})}$.

Algorithm 1 concludes the whole process of attack vector construction. It is notable that threshold t should be carefully chosen in the consideration of both sparsity and evading BDD. A higher t can generate a sparser attack vector but also increase the possibility of being detected. It is also notable that if the noise is not zero-mean, the threshold t is chosen according to the tolerable noise range. Since the measurement noise follows $\mathcal{N}(0, \sigma^2)$, it is assumed that all noise variables are within the range of $[-3\sigma, 3\sigma]$ (otherwise it will be identified as bad data). Thus, threshold t should not exceed 3σ. The following proposition can assist in choosing threshold t.

Algorithm 1. Sparse stealth attacks construction

```
Input: $\mathbf{H} \in \mathbb{R}^{m \times n}$, $C > 0, t > 0$.
Procedure:
  1) Compute $\mathbf{B} = \mathbf{H}\left(\mathbf{H^T H}\right)^{-\mathbf{1}}\mathbf{H^T} - \mathbf{I}$.
  2) Get the standard basis matrix $\mathbf{U}$ of $Null(\mathbf{B})$ so that $i$-th
     column $\mathbf{u}_i$: $\mathbf{B}\mathbf{u}_i = 0$.
  3) Find column vector $\mathbf{u}$ in $\mathbf{U}$: $\mathbf{u} = \arg\max_i(var(\mathbf{u}_i))$.
  4) Scale up/down vector $\mathbf{u}$ by $\epsilon$: $\mathbf{u}' = \epsilon\mathbf{u}$ and $\epsilon = \frac{C}{\max(\mathbf{u})}$.
  5) Shrink the vector using the threshold $t$ to obtain the sparse
     attack vector $\mathbf{a}$: $\mathbf{a} = \mathcal{S}_t\left(\mathbf{u}'\right)$.
Output: $\mathbf{a}$.
```

Proposition 1: If an attack vector $\boldsymbol{a}$ is constructed using Algorithm 1 with the shrinkage threshold t, the probability of successfully evading detection by residue-based detection algorithms in the system is at least

$$P_l(t) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{3\sigma - t}{\sigma\sqrt{2}}\right)\right] \tag{14}$$

where $\text{erf}(\cdot)$ refers to the Gauss error function and σ is the standard deviation of the Gaussian measurement noise.

Proof: Since the vector $\boldsymbol{u}$ is selected form $\text{Null}(\mathbf{B})$, it satisfies $\boldsymbol{u} = \mathbf{H}\boldsymbol{c}$, where $\boldsymbol{c} \in \mathbb{R}^n$. Let $\boldsymbol{a} = \mathcal{S}_t\left(\epsilon\boldsymbol{u}\right) = \epsilon\boldsymbol{u} - \boldsymbol{u}_t$. The residual vector $\boldsymbol{r}'$ when attack $\boldsymbol{a}$ is injected into system is calculated as

$$\begin{aligned} \boldsymbol{r}' &= \boldsymbol{z}_{\mathbf{a}} - \mathbf{H}\hat{\theta}_{\mathbf{a}} = \boldsymbol{z} + \boldsymbol{a} + \mathbf{e} - \mathbf{HK}(\boldsymbol{z} + \boldsymbol{a} + \mathbf{e}) \\ &= \mathbf{Hx} + \epsilon\mathbf{H}\boldsymbol{c} - \boldsymbol{u}_t + \mathbf{e} - \mathbf{H}(\mathbf{KHx} + \mathbf{K}\epsilon\mathbf{H}\boldsymbol{c} - \mathbf{K}\boldsymbol{u}_t + \mathbf{Ke}) \\ &= \mathbf{Hx} + \epsilon\mathbf{H}\boldsymbol{c} - \boldsymbol{u}_t + \mathbf{e} - \mathbf{Hx} - \epsilon\mathbf{H}\boldsymbol{c} + \mathbf{HK}\boldsymbol{u}_t - \mathbf{HKe} \\ &= (\mathbf{I} - \mathbf{HK})(\mathbf{e} - \mathbf{u}_t). \end{aligned} \tag{15}$$

Equation (15) indicates that when an attack generated by Algorithm 1 exists, it can be regarded that the random noise is perturbed by small amounts. A very small element comparing to σ in $\boldsymbol{u}_t$ should not impact the noise level or the residue since the shifted noise variable is still within the tolerable range. To evaluate the probability of having no impact on noise level, we consider the worst case when all elements in vector $\boldsymbol{u}_t$ equal the threshold t. In this case, it can be viewed as that the noise level is shifted down by an amount of t. The shifted noise $\mathbf{e}'$ follows $\mathcal{N}(-t,\, \sigma^2)$. Therefore, the probability when the shifted noise variables are within the normal range of $[-3\sigma,\, 3\sigma]$ can be computed by

$$P_l(t) = \int\limits_{-3\sigma}^{3\sigma} \frac{1}{\sigma 2\pi} \exp\left(-\frac{(k+t)^2}{2\sigma^2}\right) dk. \tag{16}$$

This probability can be evaluated using (14). In fact, as a large number of elements in $\boldsymbol{u}_t$ tend to be much smaller than t, nondetection probability $P(t) > P_l(t)$. Thus, proposition 1 is proved. ■

B. Targeted Attacks

In practice, adversaries may intend to modify specific state variables. In this case, the amounts in the targeted subset in the vector $\boldsymbol{c}$ are fixed. Sparse attack vector construction methods for targeted attacks have been extensively explored in the literature, e.g., [7], [8], and [12]. Additionally, certain measurements may be protected, and adversaries would not be able to compromise these secured measurements. It is notable that protecting all measurements may not be feasible due to the high cost. Therefore, sparse attack vectors need to be carefully designed to contaminate specific state variables without compromising those protected measurements.

Let $\mathcal{I}$ denote the indices of state variables that are specifically attacked. $\bar{\mathcal{I}}$ is the complementary set of $\mathcal{I}$ and denotes the indices of state variables that can be arbitrarily chosen to launch targeted attacks. Measurements Jacobian $\mathbf{H}$ is $[\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_n]$ where $\mathbf{h}_i$ denotes the ith column vector of $\mathbf{H}$. A stealth attack vector $\boldsymbol{a}$ can then be written as

$$\boldsymbol{a} = \mathbf{H}\boldsymbol{c} = \sum_{i \in \mathcal{I}} \mathbf{h}_i c_i + \sum_{j \in \bar{\mathcal{I}}} \mathbf{h}_j c_j. \tag{17}$$

In a targeted attack, the value of c_i, $i \in \mathcal{I}$ is fixed and predesigned to be injected into the state variables. Let $\boldsymbol{b} = \sum_{i \in \mathcal{I}} \mathbf{h}_i c_i$, which is predesigned by attackers. The attack vector $\boldsymbol{a}$ is then designed based on the fixed vector $\boldsymbol{b}$. As proposed in [7], (17) can be transformed using a projection matrix $\mathbf{P} = \mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T$. Since $\boldsymbol{a} - \boldsymbol{b} = \mathbf{H}_{\bar{\mathcal{I}}}\boldsymbol{c}_{\bar{\mathcal{I}}}$, where $\mathbf{H}_{\bar{\mathcal{I}}}$ denotes the submatrix of $\mathbf{H}$ containing columns with index in $\bar{\mathcal{I}}$, i.e., $\mathbf{H}_{\bar{\mathcal{I}}} = \left[\mathbf{h}_{j_1}, \mathbf{h}_{j_2}, \ldots, \mathbf{h}_{j_{|\bar{\mathcal{I}}|}}\right]$, where $j_i \in \bar{\mathcal{I}}$ for $1 \leq i \leq \left|\bar{\mathcal{I}}\right|$. By left-multiplying both sides with $\mathbf{P}_{\bar{\mathcal{I}}}$, we have

$$\begin{aligned} \mathbf{P}_{\bar{\mathcal{I}}}(\boldsymbol{a} - \boldsymbol{b}) &= \mathbf{P}_{\bar{\mathcal{I}}}\mathbf{H}_{\bar{\mathcal{I}}}\boldsymbol{c}_{\bar{\mathcal{I}}} \\ &= \mathbf{H}_{\bar{\mathcal{I}}}\left(\mathbf{H}_{\bar{\mathcal{I}}}^T\mathbf{H}_{\bar{\mathcal{I}}}\right)^{-1}\mathbf{H}_{\bar{\mathcal{I}}}^T\mathbf{H}_{\bar{\mathcal{I}}}\boldsymbol{c}_{\bar{\mathcal{I}}} \\ &= \mathbf{H}_{\bar{\mathcal{I}}}\boldsymbol{c}_{\bar{\mathcal{I}}} = \boldsymbol{a} - \boldsymbol{b} \end{aligned} \tag{18}$$

where $\mathbf{P}_{\bar{\mathcal{I}}} = \mathbf{H}_{\bar{\mathcal{I}}}\left(\mathbf{H}_{\bar{\mathcal{I}}}^T\mathbf{H}_{\bar{\mathcal{I}}}\right)^{-1}\mathbf{H}_{\bar{\mathcal{I}}}^T$.

We can then easily obtain that $(\mathbf{P}_{\bar{\mathcal{I}}} - \mathbf{I})\boldsymbol{a} = (\mathbf{P}_{\bar{\mathcal{I}}} - \mathbf{I})\boldsymbol{b}$, let $\mathbf{B} = \mathbf{P} - \mathbf{I}$, we have the following equivalent criteria for an attack vector $\boldsymbol{a}$ to be stealth

$$\mathbf{B}_{\bar{\mathcal{I}}}\boldsymbol{a} = \mathbf{B}_{\bar{\mathcal{I}}}\boldsymbol{b} \tag{19}$$

where $\mathbf{B}_{\bar{\mathcal{I}}} = \mathbf{P}_{\bar{\mathcal{I}}} - \mathbf{I}$.

Since a subset of measurements is protected, those elements in attack vector $\boldsymbol{a}$ should be restricted to 0. Let $\mathbf{y} = \mathbf{B}_{\bar{\mathcal{I}}}\boldsymbol{b}$, and assume that the pth measurement is secured, e.g., $a_p = 0$. Applying the l_1-relaxation, the sparse attack vector can be obtained by solving the following optimization problem:

$$\begin{aligned} &\min_{\mathbf{c}} \|\mathbf{H}\boldsymbol{c}\|_1 \\ &\text{s.t. } \mathbf{B}_{\bar{\mathcal{I}}}\mathbf{H}\boldsymbol{c} = \mathbf{y} \\ &\quad\ \mathbf{H}^p\boldsymbol{c} = 0 \end{aligned} \tag{20}$$

where $\mathbf{H}^p$ denotes the pth row of matrix $\mathbf{H}$ and minimizing l_1-norm of a vector $\|\boldsymbol{v}\|_1 = \sqrt{\sum_i |v_i|}$ can promote $\boldsymbol{v}$ to be sparse. This problem is well discussed in the field of compressive sensing [28] and can be quickly solved.

IV. Strategic Protection

Increasing the number of protected measurements can make the stealth attacks more difficult to be accomplished. It is obvious that stealth attacks can be completely prevented by securing all measurements. However, it is not economical or necessary to secure all measurement devices to defend against stealth attacks. Bobba *et al.* [18] explored the minimal measurement subset that is required to be protected to defend against attacks using brute-force search. This method is time-consuming and only feasible for small-sized power grids. In this section, an efficient algorithm is proposed to quickly find measurement protection subsets, which have the same sizes as that from brute-force method in nearly all cases.

Let the set $\mathcal{P} \subset \{1, 2, \ldots, m\}$ be the measurement set that are secured and the complementary set $\bar{\mathcal{P}}$ denotes the index of those measurements that can be contaminated. Similar to (20), adversaries can construct the sparse attack vector $\boldsymbol{a}$ by solving

$$\begin{aligned} &\min_{\boldsymbol{c}} \left\|\mathbf{H}^{\bar{\mathcal{P}}}\boldsymbol{c}\right\|_1 \\ &\text{s.t.}\mathbf{B}_{\bar{\mathcal{I}}}\mathbf{H}\boldsymbol{c} = \mathbf{y} \\ &\quad \mathbf{H}^{\mathcal{P}}\boldsymbol{c} = 0. \end{aligned} \tag{21}$$

If the protection set $\mathcal{P}$ is properly chosen, specific targeted attack vectors would not exist. Namely, (21) would have no solutions. Giving specified vector $\boldsymbol{c}_{\mathcal{I}}$, which is the targeted subset vector of $\boldsymbol{c}$, the straightforward method is to protect all measurements in the set corresponding to all nonzero elements in $\boldsymbol{a}$ that $\boldsymbol{a} = \mathbf{H}_{\mathcal{I}}\boldsymbol{c}_{\mathcal{I}}$. In this way, it probably requires a large number of measurements to be protected since $\boldsymbol{a}$ may not be desirably sparse. Finding or computing the smallest protection set that can prevent targeted attacks is difficult. The brute-force search method, which is discussed in [18], can guarantee finding the smallest possible sets. However, this method is extremely complex and not feasible in practice.

When a certain measurement is secured, attackers need to compromise more measurements or inject extra errors into the rest of the measurements to launch targeted attacks. From (17), we have

$$\boldsymbol{a} = \mathbf{H}\boldsymbol{c} = \boldsymbol{b} + \mathbf{H}_{\bar{\mathcal{I}}}\boldsymbol{c}_{\bar{\mathcal{I}}} \tag{22}$$

where $\boldsymbol{b}$ represents predesired injections. It is obvious that protecting certain measurements can always be more effective than others. For example, it is more important to secure the measurements corresponding to the nonzero elements in $\boldsymbol{b}$ than others. If a subset $\mathcal{P}$ of the total measurements is protected, we have

$$\boldsymbol{a}_{\mathcal{P}} = \boldsymbol{b}_{\mathcal{P}} + \mathbf{H}_{\bar{\mathcal{I}}}^{\mathcal{P}} \boldsymbol{c}_{\bar{\mathcal{I}}} = 0 \tag{23}$$

$$-\boldsymbol{b}_{\mathcal{P}} = \mathbf{H}_{\bar{\mathcal{I}}}^{\mathcal{P}} \boldsymbol{c}_{\bar{\mathcal{I}}}. \tag{24}$$

If the rank of $\mathbf{H}_{\bar{\mathcal{I}}}^{\mathcal{P}}$ is smaller than protection size $|\mathcal{P}|$, and the augmented matrix with vector $\boldsymbol{b}_{\mathcal{P}}$ can increase the rank, namely $\operatorname{rank}\left(\left[\mathbf{H}_{\bar{\mathcal{I}}}^{\mathcal{P}} | \boldsymbol{b}_{\mathcal{P}}\right]\right) = \operatorname{rank}\left(\mathbf{H}_{\bar{\mathcal{I}}}^{\mathcal{P}}\right) + 1$, then $\boldsymbol{c}_{\bar{\mathcal{I}}}$ satisfying (24) does not exist, indicating that the system is successfully protected from targeted attacks with $\boldsymbol{b}$. Otherwise, when vector $\boldsymbol{b}_{\mathcal{P}}$ cannot increase the rank of matrix $\mathbf{H}_{\bar{\mathcal{I}}}^{\mathcal{P}}$, i.e., $\operatorname{rank}\left(\left[\mathbf{H}_{\bar{\mathcal{I}}}^{\mathcal{P}} | \boldsymbol{b}_{\mathcal{P}}\right]\right) = \operatorname{rank}\left(\mathbf{H}_{\bar{\mathcal{I}}}^{\mathcal{P}}\right)$, there exist solutions of $\boldsymbol{c}_{\bar{\mathcal{I}}}$, which means that adversaries can still find attack vectors to launch targeted attacks. The problem is then to find the best solution to obtain highly sparse $\boldsymbol{a}$. It is known from (22) that making a certain subset $\mathcal{P}$ of the measurements immune to attacks can result in an attack vector $\boldsymbol{a}$ which contaminates more state variables. This makes the attacks more difficult to be accomplished. Therefore, it can be deduced that protecting certain measurement would result in a larger $\|\boldsymbol{a}\|_1$ value than that of protecting another measurement. Protecting these measurements would be more effective than others and these measurements can be regarded as *critical measurements* to targeted attacks. Based on this idea, giving specified targeted state bias vector $\boldsymbol{c}_{\mathcal{I}}$, we can design a greedy method to search a small subset of these measurements to be protected to defend from targeted attacks.

Algorithm 2 presents the greedy search method to find a small protection subset of measurements with the knowledge of existing protection set and targeted vector $\boldsymbol{c}_{\mathcal{I}}$. At each iteration, the algorithm assume that one more measurement is protected and check the feasibility of constructing attack vector $\boldsymbol{a}$. If the stealth attack vector exists when every measurement is protected one by one, the algorithm increases the protection set by selecting the most important measurement, which leads to the largest value of $\|\boldsymbol{a}\|_1$ when it is protected. The selection process continues until stealth targeted attack vector does not exist.

Algorithm 2. Greedy subset searching Algorithm

```
Input: $\mathbf{H}$, $\mathcal{I}$, $\mathbf{c}_{\mathcal{I}}$, $\mathcal{P}$.
Initialize: $\mathbf{B}_{\bar{\mathcal{I}}} = \mathbf{H}_{\bar{\mathcal{I}}}\left(\mathbf{H}_{\bar{\mathcal{I}}}^T\mathbf{H}_{\bar{\mathcal{I}}}\right)^{-1}\mathbf{H}_{\bar{\mathcal{I}}}^T - \mathbf{I}$,  $\mathbf{y} = \mathbf{B}_{\bar{\mathcal{I}}}\mathbf{H}_{\mathcal{I}}\mathbf{c}_{\mathcal{I}}$,
$\mathcal{P}' = \mathcal{P}$, $k = 1$, $\mathcal{P}_k = \mathcal{P}'$.
Iteration: At the $k$-th iteration:
Compute the complementary set $\bar{\mathcal{P}}$ of $\mathcal{P}'$.
For $i = 1 : |\bar{\mathcal{P}}|$
    Put the $i$-th entry in $\bar{\mathcal{P}}$ into $\mathcal{P}_k$: $\mathcal{P}_k = \mathcal{P}' \cup \bar{\mathcal{P}}_i$.
    Checking the feasibility of finding $\mathbf{c}$ from equation (21).
    If feasible
        Compute $\chi_i = \|\mathbf{Hc}\|_1$.
    else
        $\mathcal{P}' = \mathcal{P}_k$; Quit the iteration.
    end
end
Find index $i$ such that $\chi_i$ has the largest value.
Update set $\mathcal{P}' = \mathcal{P}' \cup \bar{\mathcal{P}}_i$.
Output: $\mathcal{P}'$.
```

For a large power grid system, it is not feasible to find the smallest protection subset to prevent any of undetectable attacks that satisfy $\boldsymbol{a} = \mathbf{H}\boldsymbol{c}$ by brute-force search. Instead, we can protect the union set of those subsets selected for protecting every single state variable. Our proposed method can quickly find a small subset that protect the whole system from any stealth attacks satisfying (6). The search procedure can be concluded in Algorithm 3.

Algorithm 3. Minimal subset selection algorithm

```
Input: $\mathbf{H}$.
Initialize: $\mathcal{P} = 0$.
For $i = 1 : n$
    Let $\mathcal{I} = \{i\}$.
    Find $\mathcal{P}_i$ using Algorithm 2.
end
$\mathcal{P} = \mathcal{P}_1 \cup \mathcal{P}_2 \cup \ldots \cup \mathcal{P}_n$.
Output: $\mathcal{P}$.
```

This method cannot guarantee the smallest subset that can be found, but it provides at least a quasi-optimal subset that contains a slightly larger number of elements. Most importantly, this method is fast and feasible in practice. In the worst case, to find a protection set with k elements, the algorithm needs to test the feasibility $\mathcal{K}_a$ times

$$\mathcal{K}_a = mk - \frac{k(k-1)}{2}. \tag{25}$$

This figure is much smaller than that using brute-force method, where it needs to test $\sum_{i=1}^{k-1}\binom{m}{i} + 1$ combinations in the best case to find the protection set with k elements. Although our proposed algorithm may not find the global optimum solution, it provides some flexibility. When it is not possible to protect certain selected measurement device in practice, the algorithm can find a suboptimal subset instead.

V. Robust Detection

Traditional residue testing-based false data detection methods cannot provide protection of state estimation from carefully designed stealth attacks. Therefore, new detection methods need to be designed to detect random errors as well as stealth attacks. It is shown that a series of measurement data exhibit low rank and sparse structure, which can be employed in anomaly detection method [11]. In practice, measurements tend to be contaminated with noise. Additionally, it may also happen that some of the measurements are lost due to the measurement device failures or disrupted communication links. In this section, these situations are addressed.

Considering a time interval T, the power system obtains a series of measurements $[\mathbf{z}_{\mathbf{a}1}, \mathbf{z}_{\mathbf{a}2}, \ldots, \mathbf{z}_{\mathbf{a}T}]$ at the time instants $t_1, t_2, \ldots, t_T$. These measurements can form a matrix $\mathbf{Z}_{\mathbf{a}} \in \mathbb{R}^{m\times T}$, which can be decomposed as

$$\mathbf{Z}_{\mathbf{a}} = \mathbf{Z} + \mathbf{A} + \mathbf{E} \tag{26}$$

where $\mathbf{Z} \in \mathbb{R}^{m\times T}$ is the block of true measurements with each column $\mathbf{z}_i$ representing true measurements at time t_i, $\mathbf{A} \in$

$\mathbb{R}^{m\times T}$ denotes the attack matrix formed by all instant sparse attacks, and $\mathbf{E}$ represents the noise.

It is known that fast system dynamics are usually well damped in the power system. This implies that the system states would change gradually in a small period T, making the matrix $\mathbf{Z}$ typically low rank. Additionally, malicious injection data matrix $\mathbf{A}$ tends to be sparse. This is due to the fact that some of the measurements may be protected and also because attackers would launch attacks with least effort. Given corrupted measurements matrix $\mathbf{Z_a}$, it is possible to recover low-rank matrix $\mathbf{Z}$ and sparse-attack matrix $\mathbf{A}$ by performing low rank and sparse decomposition, which is well discussed in the robust principle component analysis (PCA) problem [29], which solves

$$\begin{aligned}&\min \|\mathbf{Z}\|_* + \lambda \|\mathbf{A}\|_1\\ &\text{s.t.}\ \|\mathbf{Z_a} - \mathbf{Z} - \mathbf{A}\|_F \le \delta\end{aligned} \tag{27}$$

where $\|\cdot\|_*$ denotes the nuclear norm, $\|\cdot\|_F$ denotes the Frobenius norm, δ represents a small positive noise bound, and λ is the regulation parameter. This problem is also addressed in compressive sensing and matrix completion [30] literature. Thus, true measurements can be recovered and the sparse perturbations including malicious attacks and other false data can also be identified.

Unlike coordinated malicious attacks, the missing data in the measurements can result in residue changes in (5). These incomplete measurement data, as well as the measurements with errors, would be identified as bad data by traditional BDD algorithms. The proposed algorithm can not only detect the missing and inaccurate measurement data but also detect the carefully designed stealth attacks, which is undetectable to traditional methods. More importantly, the proposed algorithm can recover the true measurements from the incomplete measurements.

In order to address the problem that only noise-contaminated partial measurements are collected, the PCA problem can be extended to the following form with element-wise error constraints:

$$\begin{aligned}&\min \|\mathbf{Z}\|_* + \lambda \|\mathbf{A}\|_1\\ &\text{s.t.}\ |\mathcal{P}_\Omega(\mathbf{Z_a}) - \mathcal{P}_\Omega(\mathbf{Z} + \mathbf{A})| \preceq \epsilon\end{aligned} \tag{28}$$

where $\preceq$ represents element-wise inequality and $\mathcal{P}_\Omega(\cdot)$ denotes a projection operation, in which all elements outside the set Ω are forced to 0. ϵ is the matrix of entry-wise error bounds. It is demonstrated in [31] that this problem is equivalent to the following problem:

$$\begin{aligned}&\min \|\mathbf{Z}\|_* + \lambda \|\mathcal{T}(\mathbf{A}, \tilde{\epsilon})\|_1\\ &\text{s.t.}\ \mathbf{Z_a} = \mathbf{Z} + \mathbf{A}\end{aligned} \tag{29}$$

where $\tilde{\epsilon}$ has the same value as ϵ in the projection set Ω and infinite outside set Ω, and the soft thresholding operation $\mathcal{T}_\epsilon(a_{ij})$ is defined as

$$\mathcal{T}(a_{ij}, \epsilon) = \operatorname{sign}(a_{ij}) \cdot \max\{|a_{ij}| - \epsilon, 0\}. \tag{30}$$

A variant of the augmented Lagrangian method (ALM), which is also known as the alternating direction method of multipliers (ADMM) algorithm [32], is used to solve the problem defined by (29). The Lagrangian corresponding to this problem is

$$\mathcal{L}(\mathbf{Z}, \mathbf{A}, \mathbf{Y}, \mu) = \|\mathbf{Z}\|_* + \lambda \|\mathcal{T}(\mathbf{A}, \tilde{\epsilon})\|_1 + \langle \mathbf{Y}, \mathbf{H}\rangle + \frac{\mu}{2}\|\mathbf{H}\|_2^2 \tag{31}$$

where $\langle\cdot\rangle$ denotes the Frobenius product, $\mathbf{H} = \mathbf{Z_a} - \mathbf{Z} - \mathbf{A}$ and $\mu > 0$. λ can be set to $\sqrt{m/|\Omega|}$. We further define the singular-value thresholding operation as

$$\mathcal{D}(\mathbf{X}, \tau) = \mathbf{U}\mathcal{T}(\Sigma, \tau)\mathbf{V}^T$$

where τ is the threshold and $\mathbf{X} = \mathbf{U\Sigma V}^T$. It is notable that ADMM updates $\mathbf{Z}, \mathbf{A}, \mathbf{Y}$ separately only once in each iteration, so it is efficient. The convergence of the whole algorithm is analyzed in [32], which states that the condition for convergence requires $\sum_1^\infty \mu_k^{-1} = +\infty$ where μ_k denotes the value of μ in the kth iteration. The whole process of solving (29) is shown in Algorithm 4.

Algorithm 4. RPCA with entry wise constraints

```
Input: $\mathbf{Z}_\mathbf{a}^p = \mathcal{P}_\Omega(\mathbf{Z_a}) \in \mathbb{R}^{m\times T}$, $\tilde{\epsilon} \in \mathbb{R}^{m\times T}$, $\lambda$.
Initialize $\mathbf{Z} = \mathbf{0}$, $\mathbf{A} = \mathbf{0}$, $\mathbf{Y} = \mathbf{0}$, $\mu > 0$, $\rho > 1$, $k = 0$.
while not converged
    1) Update the value of low rank matrix $\mathbf{Z}_{k+1}$:
       $\mathbf{Z}_{k+1} = \mathcal{D}\left(\mathbf{Z}_\mathbf{a}^p - \mathbf{A}_k + \frac{\mathbf{Y}_k}{\mu_k}, \mu_k^{-1}\right)$.
    2) Compute the value of sparse matrix $\mathbf{A}_{k+1}$ by minimizing:
       $F(\mathbf{A}) = \frac{\lambda}{\mu}\|\mathcal{T}(\mathbf{A}, \tilde{\epsilon})\|_1 - \operatorname{tr}\left(\frac{\mathbf{Y}_k}{\mu_k}(\mathbf{A} - (\mathbf{Z}_\mathbf{a}^p - \mathbf{Z}_k))\right) + \frac{1}{2}\|\mathbf{A} - (\mathbf{Z}_\mathbf{a}^p - \mathbf{Z}_k)\|_F$.
    3) Update the Lagrange multiplier $\mathbf{Y}$:
       $\mathbf{Y}_{k+1} = \mathbf{Y}_k + \mu_k\left(\mathbf{Z}_\mathbf{a}^p - \mathbf{Z}_{k+1} - \mathbf{A}_{k+1}\right)$.
    4) Update $\mu_{k+1} = \rho \cdot \mu_k$.
    5) Update $k = k + 1$.
end while
Return $\mathbf{Z}$, $\mathbf{A} = \mathcal{T}(\mathbf{A}, \tilde{\epsilon})$.
Output $\mathbf{Z}$, $\mathbf{A}$.
```

It is notable that when incomplete measurements are collected, Algorithm 4 will take the missing data to be sparse anomalies and it can also recover the low-rank true measurement matrix and sparse anomaly matrix. However, the recovery accuracy would be impacted as the sparsity is changed. The recovered sparse attack matrix can ignore those injected data outside the observation set. Thus, it is more difficult to identify all malicious attacks with partial observations.

VI. Numerical Results

In this section, the algorithms introduced above are evaluated by simulations performed based on the IEEE test systems [33]. The MATLAB package MATPOWER [26] is used to simulate the power system. The convex optimization problems are solved using the convex optimization toolbox CVX [34].

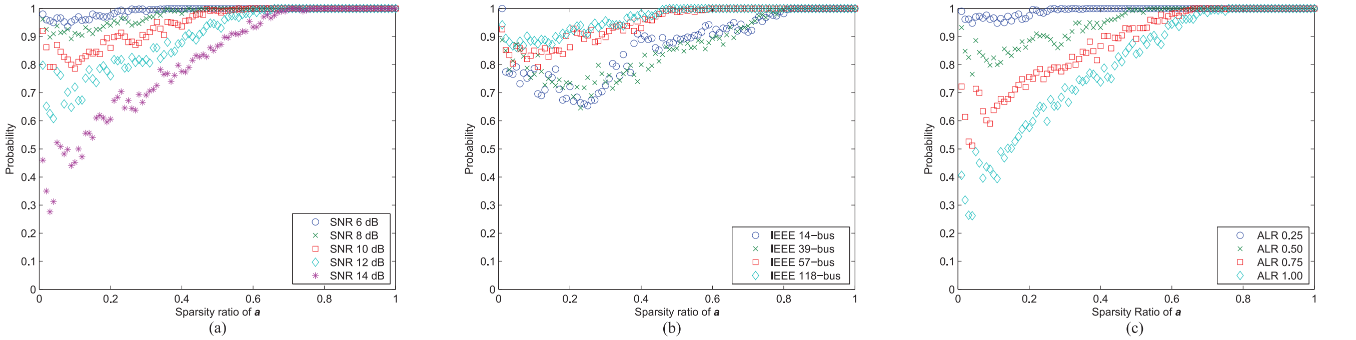


Fig. 2. Probabilities of successful attack injections (a) under different SNRs for IEEE-57 bus system, SR is 0.4; (b) for different bus systems, SR is 0.4 and $\mathrm{SNR} = 10\,\mathrm{dB}$; (c) for different SRs in IEEE-57 bus system, $\mathrm{SNR} = 10\,\mathrm{dB}$. (c) This figure utilizes random columns in $\mathrm{Null}(\mathbf{B})$ rather than that with largest variance.

A. *Performance of Stealth Attack Construction*

The performance of Algorithm 1 which generates highly sparse undetectable attack vectors is tested in different scenarios. Figs. 2 and 6 show the probabilities of successfully generating undetectable attack vectors with different levels of sparsity and different attack level ratios (ALRs), respectively. An attack is regarded as successful when the maximum value in residue vector does not exceed that without attacks. Sparsity ratio (SR) is defined as k/m, where k is the number of nonzero elements in $\boldsymbol{a}$ and m is the size of $\boldsymbol{a}$. The ALR is defined as the maximum attack value C to the mean value of the state variables: $\frac{C}{\mathrm{mean}(\theta)}$. Generally, these figures reveal that there are high probabilities for Algorithm 1 to successfully generate highly sparse undetectable attacks.

The noise in the simulation is modeled as Gaussian distributed with zero-mean. The signal-to-noise ratio (SNR) indicates the noise level compared with true measurements in the simulation. The noise may be due to measuring devices and process, or due to the communication channel noise. It is clear in Figs. 2(a) and 6(a) that in a relatively noisy case, the probability of a successful attack is extremely high (close to 1). In the low noise case, there is also high probability of injecting a successful highly sparse undetectable malicious attack. The algorithm is also assessed using different power grid system models, which is shown in Figs. 2(b) and 6(b). It is notable that in a larger bus system, Algorithm 1 can provide a better performance even for extremely sparse attacks and high ALRs. For example, the success ratio is around 90% for IEEE 118-bus system to generate stealth attacks with SR lower than 0.1, compared with 75% for IEEE 14-bus system shown in Fig. 2(b). This probability is 100% for IEEE 118-bus system to generate attacks with $\mathrm{ALR} = 1$ compared to 80% for IEEE 57-bus system shown in Fig. 6(b). Therefore, it can be anticipated that the algorithm would have a better performance in a real power system, which is much larger than the tested systems.

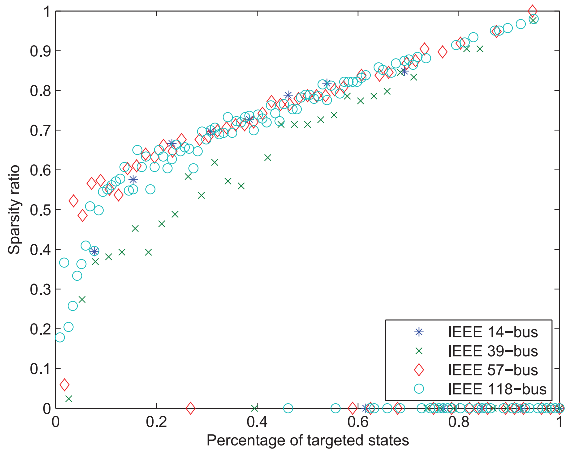


Fig. 3. Sparsity of $\boldsymbol{a}$ under different attack conditions.

TABLE I

NUMBER OF MEASUREMENTS IN PROTECTION SETS FROM TWO METHODS

Test systems	Algorithm 2	Brute-force search
IEEE 9-bus	9	8
IEEE 14-bus	15	13
IEEE 39-bus	59	–
IEEE 57-bus	84	–
IEEE 118-bus	187	–

Additionally, it can be seen from Fig. 2 that it is always harder to inject sparser attacks while Fig. 6 reveals that attacks with higher values would be more likely to be detected. Figs. 2(c) and 6(c) display the performance when injecting attacks with different SRs and ALRs. It is notable that in Fig. 6(c), the algorithm utilizes randomly selected columns in a basis matrix of $\mathrm{Null}(\mathbf{B})$ rather than that with the largest variance. The results imply that using randomly chosen columns can also successfully inject undetectable attacks with high probabilities.

It is known that stealth attacks having $m - n$ nonzero entries can always be found. In IEEE 57-bus system, this figure is 80, for which the SR is about 59%. However, by using Algorithm 1, there is still a high probability that attackers can inject undetectable attacks with SRs lower than 59%. Even for an attack with SR lower than 0.05, the success rate is still around 80% when the SNR is 10 dB and ALR is 0.5.

Targeted attack construction method in (20) is assessed under different attack conditions in which different percentage of total state variables are assumed to be modified. The targeted set is randomly selected and the protected measurement is

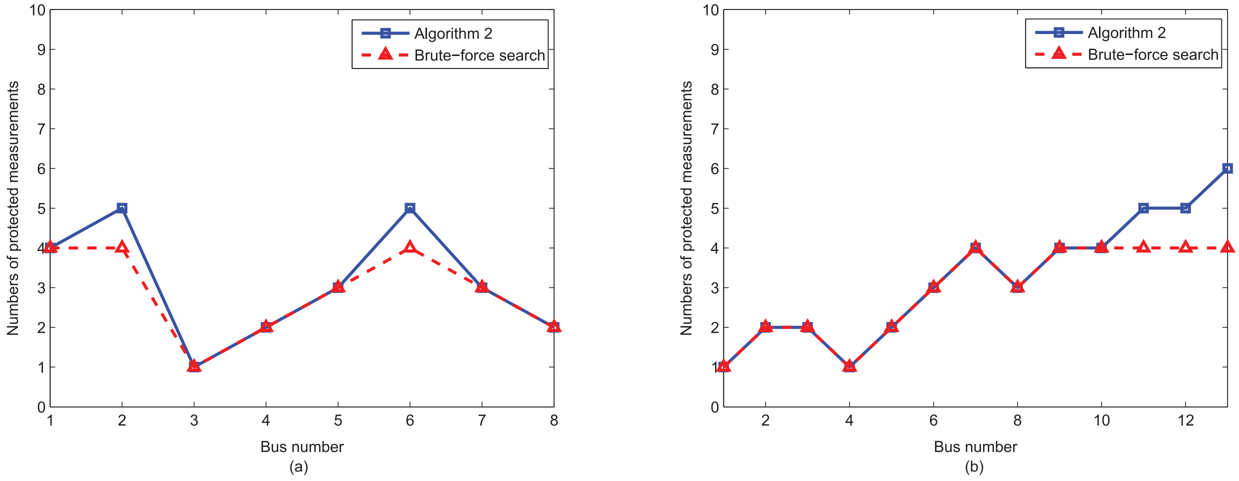


Fig. 4. Number of protected measurements to protect every single state variable from being targeted. (a) IEEE-9 bus system. (b) IEEE-14 bus system.

TABLE II
NUMBER OF TESTING TIMES FOR TWO ALGORITHMS TO FIND PROTECTION SUBSETS

Test systems	Bus number	1	2	3	4	5	6	7	8	9	10	11	12	13
IEEE-9 bus	Algorithm 2	55	70	1	18	34	75	40	25	–	–	–	–	–
	Brute-force search	1434	834	1	18	154	835	235	25	–	–	–	–	–
IEEE-14 bus	Algorithm 2	1	35	39	2	42	71	107	72	109	123	141	135	183
	Brute-force search	1	35	36	2	73	629	7108	630	7140	6235	6236	6237	7109

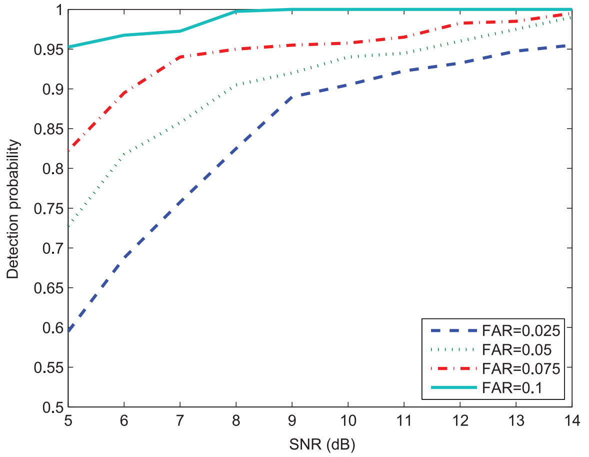


Fig. 5. Probability of successful false data detection.

TABLE III
DETECTION PROBABILITY AND MEASUREMENTS DEVIATION WITH PARTIAL OBSERVATIONS

Observations (%)	Detection probability (%)	Variable deviation (%)
100	100	0.53
95	92	2.06
90	56.3	3.7
85	44.7	5.3

also randomly chosen. It can be observed from Fig. 3 that, in order to precisely alter specified state variables, the coordinated attack vectors cannot be highly sparse. Thus, attackers need to compromise a number of measurements to launch targeted attacks. Highly sparse attacks can only be achieved when the percentage of targeted state variables is extremely low for certain test systems. For example, SR can be less than 0.1 for IEEE 39-bus system when a small number of state variables are targeted. The figure also shows that in some cases, SR of attacks are 0. They correspond to the cases that: for certain targeted set of state variables, no feasible attack vectors exist when the pth measurement is protected. Therefore, it implies that when certain carefully selected measurements are protected, attackers may not be able to inject targeted attacks.

B. Performance of Strategic Protection

This section evaluates our proposed protection algorithm. To compare the protection subset generated by the proposed algorithm with that from brute-force method, we apply IEEE-9 bus system, which contain 17 total measurements, and IEEE-14 bus system with 33 total measurements. Table I shows the number of measurements in protection subset found by two methods. The results from the proposed algorithm for other larger test systems are also provided. In the first two test systems, the smallest protection sets generated from the proposed algorithm contain only slightly more measurements than that from brute-force method. In IEEE-14 bus system, the difference of this number is quite small compared to the total number of 33 measurements. Thus, Algorithm 2 can find protection subsets with similar number of elements but spend much less time than brute-force method.

Fig. 4 displays the number of elements in the smallest protection subsets to protect every single state variable from being targeted by adversaries. The whole system protection subsets shown in Table I are the unions of the protection subsets for protecting single variables. From both figures, it can be seen that in most cases the proposed algorithm can find a protection subset having the same size as that found by brute-force method. The size differences are only 1 or 2 when the two methods find

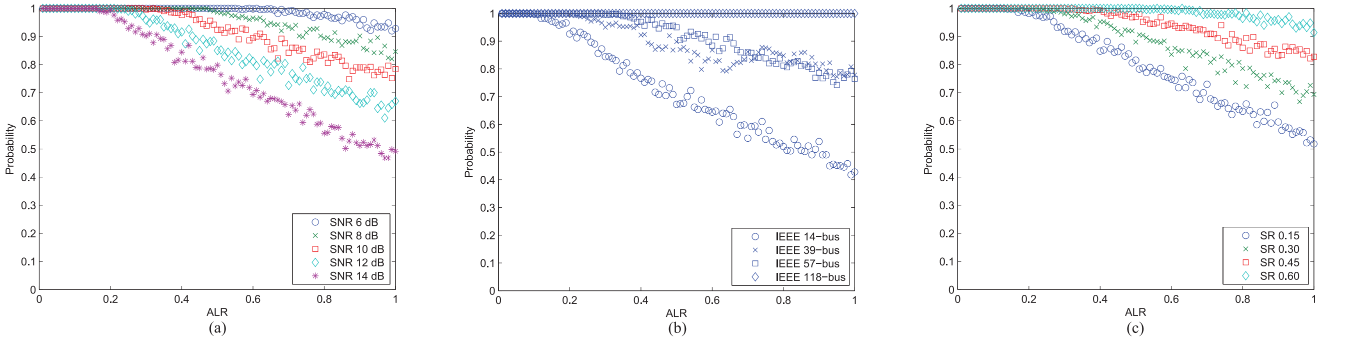


Fig. 6. Probabilities of successful attack injections (a) under different SNRs for IEEE-57 bus system, ALR is 0.5; (b) for different bus systems, ALR is 0.5 and $\mathrm{SNR} = 10\,\mathrm{dB}$; and (c) for different ALRs for IEEE-57 bus system, $\mathrm{SNR} = 10\,\mathrm{dB}$.

subsets with different number of elements. This number is quite small compared with the total number of 33 measurements in IEEE-14 bus.

Table II compares the complexities of the two algorithms in terms of the number of feasibility testing times. The results correspond with the simulation shown in Fig. 4 in which measurement protection subsets are searched for protecting every single state variable. It is obvious that when the size of protection subset exceeds 3, the difference of the two methods becomes significant. This difference is more significant when the size of protection subset is bigger as the brute-force search needs to exhaust all subset combinations with smaller sizes. It is also clear that in a larger power system, the difference is much larger for two algorithms to find a subset with same size as that in a smaller system. The testing times of the proposed algorithm will increase only slightly when the size of protection subset and the system scale grow, which is also described by (25). In a real power system, while brute-force method is infeasible because of the combinatorial complexity, the proposed method instead is fast and practical.

C. Performance of Detection

The performance of the detection algorithm is tested on IEEE 14-bus system and IEEE 57-bus system. The malicious attack vectors are constructed using our proposed Algorithm 1. In order to obtain sparsity in the rows of the attack block matrix, different column vectors in the null space in Algorithm 1 are utilized. The SR of the attacks is chosen as 15%. In Fig. 6(c), it is shown that when $\mathrm{SR} = 0.15$, traditional residual testing-based algorithms will not be able to detect those attacks. Thus, in the simulation, the algorithm is not compared with traditional methods. Additionally, recently proposed algorithms such as [23] do not deal with partial observations. These algorithms do not address the problem of error contaminated measurements as well. The detection method discussed in this paper addresses both problems. Most importantly, it can not only detect anomalies but also recover the true measurements from partial-contaminated observations.

We use the false alarm rate (FAR) which is the probability of positive alarm when there are no attacks. The noise performance of the algorithm compared to RPCA with Frobenius constraints in (27) has been extensively studied in [31]. In this paper, we focus on identifying anomalies in different scenarios when undetectable attacks are injected in power systems.

Fig. 5 shows the error tolerance performance in the IEEE 14-bus system. It is shown that when FAR exceeds 10%, the algorithm can identify attacks with high probabilities which are approaching 100%. This probability is still quite high in the presence of highly dense noise (95%). When FAR decreases, the system will absorb more noise and detection probability decreases. It can be seen that there is still a high chance of detecting anomalies: more than 90% when FAR decreases to an extremely low level (0.025) under $\mathrm{SNR} = 10\,\mathrm{dB}$.

In the case where partial measurements are collected, missing data are regarded as sparse anomalies in Algorithm 4. Additionally, nonzero entries in sparse matrix $\mathbf{A}$ can only be confirmed as attacks when they are located in the observation set. This make identifications of attacks more difficult. Algorithm 4 can circumvent this problem since it also recovers the block of true measurements. We evaluate the attack detection probabilities as well as the deviation rate of the recovered measurement variables, which is defined as $\|\mathbf{z} - \mathbf{z}'\|_2 / \|\mathbf{z}\|_2$. Table III shows the results when incomplete measurements are collected based on the IEEE 57-bus system. The FAR equals 0.05 and SNR is set to 8 dB. It can be seen that attack detection probability declines greatly with increasing missing observations. However, the recovered measurement variables experience only small deviations. Therefore, the proposed algorithm can successfully verify the true measurements, even in the situation that only partial measurements are observed (Fig. 6).

VII. Conclusion

In this paper, we looked into the problem of malicious FDIAs in power grid state estimation. We proposed stealth attack construction strategies for different scenarios and also introduced the countermeasures. It is shown that our proposed random attack construction algorithm can generate extremely sparse attack vectors. These optimal or quasi-optimal attacks can be achieved with high probability of success. The targeted undetectable attacks are obtained based on a optimization framework. The results show that attack vectors in this scenario cannot be extremely sparse, which is also discussed in literature. An efficient protection scheme is proposed in this paper to find an effective measurement protection subset to defend from the stealth attacks. The simulation results reveal that this subset searching algorithm can find a subset with almost the same size as that from the brute-force method. Additionally, a detection algorithm is introduced to detect the stealth attacks as well as other false data. This algorithm considers the case in which only partial measurements are collected in the presence of noise. The performance is demonstrated via the simulation results based on IEEE test power systems.

Acknowledgment

The authors would like to thank their colleagues at Toshiba Research Europe Ltd., for the fruitful discussions and the support of its directors.

References

[1] X. Fang, S. Misra, G. Xue, and D. Yang, "Smart grid—The new and improved power grid: A survey," *IEEE Commun. Surv. Tuts.*, vol. 14, no. 4, pp. 944–980, Oct. 2012.

[2] W. Wang and Z. Lu, "Cyber security in the smart grid: Survey and challenges," *Comput. Netw.*, vol. 57, nos. 5 and 7, pp. 1344–1371, Apr. 2013.

[3] D. B. Rawat and C. Bajracharya, "Cyber security for smart grid systems: Status, challenges and perspectives," in *Proc. SoutheastCon*, Apr. 2015, pp. 1–6.

[4] A. Abur and A. G. Expósito, *Power System State Estimation Theory and Implementation*. Boca Raton, FL, USA: CRC Press, Mar. 24, 2004.

[5] A. Monticelli, "Electric power system state estimation," *Proc. IEEE*, vol. 88, no. 2, pp. 262–282, Feb. 2000.

[6] F. C. Schweppe, J. Wildes, and D. P. Rom, "Power system static state estimation, parts I, II, III," *IEEE Trans. Power App. Syst.*, vol. PAS-89, pp. 120–135, Jan. 1970.

[7] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proc. ACM Conf. Comput. Commun. Sec.*, 2009, pp. 21–32.

[8] 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.

[9] Z. Yu and W. Chin, "Blind false data injection attack using PCA approximation method in smart grid," *IEEE Trans. Smart Grid*, vol. 6, no. 3, pp. 1219–1226, May 2015.

[10] Y. Li and Y. Wang, "State summation for detecting false data attack on smart grid," *Int. J. Elect. Power Energy Syst.*, vol. 57, pp. 156–163, May 2014.

[11] L. Liu, M. Esmalifalak, and Z. Han, "Detection of false data injection in power grid exploiting low rank and sparsity," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 9–13, 2013, pp. 4461–4465.

[12] M. Ozay, I. Esnaola, F. T. Y. Vural, S. R. Kulkarni, and H. V. Poor, "Sparse attack construction and state estimation in the smart grid: Centralized and distributed models," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1306–1318, Jul. 2013.

[13] H. Sandberg, A. Teixeira, and K. H. Johansson, "On security indices for state estimators in power networks," in *Proc. 1st Workshop Secure Control Syst. (CPSWEEK'10)*, 2010.

[14] 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.

[15] M. A. Rahman and H. Mohsenian-Rad, "False data injection attacks with incomplete information against smart power grids," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 3–7, 2012, pp. 3153–3158.

[16] Y. Huang *et al.*, "Bad data injection in smart grid: attack and defense mechanisms," *IEEE Commun. Mag.*, vol. 51, no. 1, pp. 27–33, Jan. 2013.

[17] H. Zhang, P. Cheng, J. Wu, L. Shi, and J. Chen, "Online deception attack against remote state estimation," in *Proc. World Congr. Int. Fed. Autom. Control (IFAC)*, 2014, pp. 128–133.

[18] R. B. Bobba, K. M. Rogers, Q. Wang, H. Khurana, K. Nahrstedt, and T. J. Overbye, "Detecting false data injection attacks on dc state estimation," in *Preprints 1st Workshop Secure Control Syst. (CPSWEEK'10)*, 2010, pp. 1–9.

[19] G. Dán and H. Sandberg, "Stealth attacks and protection schemes for state estimators in power systems," *Proc. 1st IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, pp. 214–219, 4–6 Oct. 2010.

[20] M. Gol and A. Abur, "PMU placement for robust state estimation," in *Proc. North Amer. Power Symp. (NAPS)*, Sep. 22–24, 2013, pp. 1–5.

[21] E. Handschin, F. C. Schweppe, J. Kohlas, and A. Fiechter, "Bad data analysis for power system state estimation," *IEEE Trans. Power App. Syst.*, vol. 94, no. 2, pp. 329–337, Mar. 1975.

[22] S. Cui, Z. Han, S. Kar, T. T. Kim, H. V. Poor, and A. 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.

[23] M. Esmalifalak, N. T. Nguyen, R. Zheng, and Z. Han, "Detecting stealthy false data injection using machine learning in smart grid," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 9–13, 2013, pp. 808–813.

[24] S. Bi and Y. J. Zhang, "Graphical methods for defense against false-data injection attacks on power system state estimation," *IEEE Trans. Smart Grid*, vol. 5, no. 3, pp. 1216–1227, May 2014.

[25] J. Hao, R. J. Piechocki, D. Kalesh, W. H. Chin, and Z. Fan, "Optimal malicious attack construction and robust detection in smart grid cyber security analysis," in *Proc. IEEE Int. Conf. Smart Grid Commun. SmartGridComm)*, Nov. 3–6, 2014, pp. 836–841.

[26] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, "Matpower: Steady-state operations, planning and analysis tools for power systems research and education," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, Feb. 2011.

[27] Z. Wang, A. Scaglione, and R. J. Thomas, "Generating statistically correct random topologies for testing smart grid communication and control networks," *IEEE Trans. Smart Grid*, vol. 1, no. 1, pp. 28–39, Jun. 2010.

[28] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[29] E. J. Candés, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 11:1–11:37, Jun. 2011.

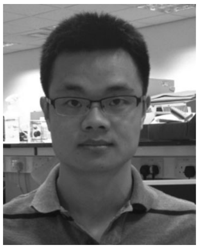
[30] E. Candés and B. Recht, "Exact matrix completion via convex optimization," *Commun. ACM*, vol. 55, no. 6, pp. 111–119, Jun. 2012.

[31] R. Paffenroth, P. du Toit, R. Nong, L. Scharf, A. P. Jayasumana, and V. Bandara, "Space-time signal processing for distributed pattern detection in sensor networks," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 1, pp. 38–49, Feb. 2013.

[32] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[33] R. Christie, "Power Systems test case archive," dept. Elect. Eng., Univ. Washington, Seattle, WA, USA, Apr. 2000 [Online]. Available: http://www.ee.washington.edu/research/pstca/

[34] M. Grant and S. Boyd. (2013, Sep.). *CVX: Matlab Software for Disciplined Convex Programming* [Online]. Available: http://cvxr.com/cvx



Jinping Hao (S'13) received the B.Eng. degree in electronic and information engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2010; the B.Eng. degree (first class Hons.) in electronic and communication engineering from the University of Birmingham, Birmingham, U.K., in 2010; and the M.Sc. degree (with Distinction) in wireless communication and signal processing from the University of Bristol, Bristol, U.K., in 2011. He is currently pursuing the Ph.D. degree in electrical and electronic engineering at the University of Bristol.

His research interests include sparse signal processing, optimization techiniques for wireless communications, and smart grids.

Robert J. Piechocki (M'06) received the M.Sc. degree (with Distinction) in wireless communications systems from the Technical University of Wroclaw, Wroclaw, Poland, in 1997, and the Ph.D. degree in wireless communications systems from the University of Bristol, Bristol, U.K., in 2002.

He is currently a Senior Lecturer with Advanced Wireless Access, and a Member of the Communications Systems and Networks Group, University of Bristol. He has authored over 100 papers in international journals and conferences, and holds 13 patents in these areas. His research interests include statistical signal processing, information and communication theory, wireless networking, body and ad hoc networks, ultra low power communications, and vehicular communications.

Dritan Kaleshi (M'00) received the Dipl.Ing. degree (with Excellence) in electronics from the Polytechnic University of Tirana, Tirana, Albania, in 1991, and the Ph.D. degree in electronic engineering from the University of Bristol, Bristol, U.K., in 2005.

He is currently a 5G Fellow with Digital Catapult, London, U.K., and a Visiting Research Fellow with the University of Bristol, where he was a Senior Lecturer in Communication Networks until 2015. He has authored over 60 papers in the field, edited two international standards, and holds three patents. He represents the U.K. in various international standardization bodies (International Organization for Standardization/International Electrotechnical Commission, European Committee for Standardization/European Committee for Electrotechnical Standardization) in areas related to Internet-of-Things (IoT), home electronic systems, and smart grid. His research interests include future networking architectures and protocols (5G and beyond), large scale loosely coupled distributed systems design, modeling and performance evaluation, and data interoperability for sensor/actuator systems (IoT).

Woon Hau Chin (S'99–M'04–SM'10) received the B.Eng. (first class Hons.) and M.Eng. degrees from the National University of Singapore, Singapore, in 1999 and 2000, respectively, and the Ph.D. degree from Imperial College London, London, U.K., in 2004, all in electrical engineering.

From 2000 to 2008, he was with the Institute for Infocomm Research, Singapore, where he was involved in the standardization of IEEE 802.11n and 3rd Generation Partnership Project Long-Term Evolution. He was also involved in several industrial projects on wireless local area networks and beyond 3G systems. He was also an Adjunct Assistant Professor with the National University of Singapore, from 2005 to 2008. Since 2008, he has been with the Telecommunications Research Laboratory, Toshiba Bristol, Bristol, U.K., where he had been involved in several EU FP7 projects and research on small cell interference mitigation, body area networks, and smart grid communications. He is currently the Research Manager with the Physical Layer Research Group, York, U.K., leading research in fundamental communication technologies.

Dr. Chin has held positions in several conference organizing committees, and is on the technical programme committee (TPC) of various international conferences. He has Co-Founded the International Symposium on Wireless Pervasive Computing (ISWPC) series of conferences, and was the TPC Co-Chair of ISWPC 2008, and a TPC Co-Chair of the Transmission Technologies Track at the IEEE Vehicular Technology Conference in 2008. He was the Founding Chair of the Vehicular Technology Chapter in Singapore, and held that position from 2006 to 2008.

Zhong Fan received the B.S. and M.S. degrees in electronic engineering from Tsinghua University, Beijing, China, in 1992 and 1994, respectively, and the Ph.D. degree in telecommunication networks from Durham University, Durham, U.K., in 1997.

He is a Chief Research Fellow with Toshiba Research Europe, Bristol, U.K. Prior to joining Toshiba, he worked as a Research Fellow with Cambridge University, Cambridge, U.K.; as a Lecturer with Birmingham University, Birmingham, U.K.; and as a Researcher with Marconi Labs Cambridge, Cambridge. He was also awarded a British Telecommunications (BT) Short-Term Fellowship to work at BT Labs, Ipswich, U.K. His research interests include wireless and IP networks, machine-to-machine, and smart grid communications.

Sparse Malicious False Data Injection Attacks and Defense Mechanisms in Smart Grids

Jinping Hao, *Student Member, IEEE*, Robert J. Piechocki, *Member, IEEE*, Dritan Kaleshi, *Member, IEEE*, Woon Hau Chin, *Senior Member, IEEE*, and Zhong Fan

***Abstract*—This paper discusses malicious false data injection attacks on the wide area measurement and monitoring system in smart grids. First, methods of constructing sparse stealth attacks are developed for two typical scenarios: 1) random attacks in which arbitrary measurements can be compromised; and 2) targeted attacks in which specified state variables are modified. It is already demonstrated that stealth attacks can always exist if the number of compromised measurements exceeds a certain value. In this paper, it is found that random undetectable attacks can be accomplished by modifying only a much smaller number of measurements than this value. It is well known that protecting the system from malicious attacks can be achieved by making a certain subset of measurements immune to attacks. An efficient greedy search algorithm is then proposed to quickly find this subset of measurements to be protected to defend against stealth attacks. It is shown that this greedy algorithm has almost the same performance as the brute-force method, but without the combinatorial complexity. Third, a robust attack detection method is discussed. The detection method is designed based on the robust principal component analysis problem by introducing element-wise constraints. This method is shown to be able to identify the real measurements, as well as attacks even when only partial observations are collected. The simulations are conducted based on IEEE test systems.**

***Index Terms*—Bad data detection (BDD), malicious data attack, robust principle component analysis (PCA), smart grid security.**

I. Introduction

COMPARED with the traditional power grids, a smart grid tends to be much more reliable, efficient, and intelligent due to the remarkable advancements in sensing, monitoring, control technologies, and also the tight integration with cyber infrastructure and advanced computing and communication technologies [1]. However, this integration can lead to new vulnerabilities to cyber attacks on the power systems. Cyber attacks are reported as one of the main potential threats to the reliable operation of the power system [2], [3]. In this paper, we consider false data injection attacks (FDIAs) against the supervisory control and data acquisition (SCADA) system in smart grids.

A power grid transmission system is a sophisticated network which connects a number of electric power generators to various consumers through extensive power lines. It is extremely important to monitor the state of this complex system such that various control and planning tasks can be performed and the reliable operation of the power system is guaranteed. In power systems, state estimation [4], [5] is used to estimate system state variables through a number of power measurements and is a useful and necessary function in energy management systems (EMS).

The SCADA system obtains power status information such as transmission line power flows, bus voltages, and also circuit-braker signals through remote terminal units (RTUs). These measurements are then used for the state estimation process in EMS, which builds real-time electricity network models. In smart grids, the complex network connections as well as the Internet make SCADA systems susceptible to potential FDIAs, in which adversaries aim to contaminate the measurements collected from RTUs and bias the state estimation at the transmission level to mislead the operation of the power system. Fig. 1 presents a block diagram of the power grid, communication network, SCADA, and control center. It is critically important to understand the behavior of adversaries so that appropriate countermeasures can be designed to either protect the system from attacks beforehand or identify the malicious false data injections in the measurements.

Recently, the problem of FDIAs as well as countermeasures has attracted a lot of attention among researchers. False data in state estimation were first discussed by Schweppe *et al.* in their pioneering work about state estimation [6]. It was not well researched until Liu *et al.* [7] proposed that if adversaries possess the knowledge of power grid topology, they may inject coordinated data attacks, which could evade detection by the bad data detection (BDD) system in state estimator. Based on this strategy, plenty of efforts have been made to design effective attack algorithms and the corresponding countermeasures, such as [8]–[11].

Adversaries may launch attacks through hacking RTUs such as sensors in substations. In consideration of the accessibility of RTUs and also hacking cost, attackers always tend to control only a few RTUs to implement a successful attack [7]. Kim and Poor [8] developed a general optimization framework-based formulation for constructing sparse attack vectors when a subset of measurements is protected, while Ozay *et al.* [12]

Manuscript received September 12, 2014; revised April 17, 2015 and June 08, 2015; accepted August 07, 2015. Date of publication August 31, 2015; date of current version October 02, 2015. This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) Dorothy Hodgkin Postgraduate Award Scholarship. Paper no. TII-14-1020.

J. Hao and R. J. Piechocki are with the Department of Electrical and Electronic Engineering, University of Bristol, Bristol BS8 1UB, U.K. (e-mail: j.hao@bristol.ac.uk; r.j.piechocki@bristol.ac.uk).

D. Kaleshi is with Digital Catapult, London NW1 2RA, U.K. (e-mail: Dritan.Kaleshi@cde.catapult.org.uk).

W. H. Chin and Z. Fan are with the Telecommunications Research Laboratory, Toshiba Research Europe Ltd., Bristol BS1 4ND, U.K. (e-mail: woonhau.chin@toshiba-trel.com; zhong.fan@toshiba-trel.com).

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/TII.2015.2475695

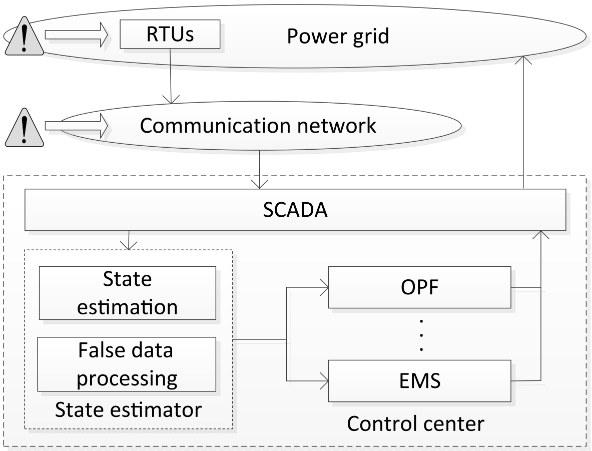


Fig. 1. Illustration of the power grid, communication network, SCADA, and control center. The warning signs indicate the vulnerabilities to FDIAs.

extended the sparse attack construction model to a distributed framework. Sandberg *et al.* [13] considered sparse attacks with injections into critical measurements, which are essential for the observability of the power grids and sensitive to attacks. In [14], methods of finding both strong stealth attacks and also optimal weak malicious data attacks (when power grid topology is unknown) with the aim of reducing the number of compromised measurements were discussed. It has been proposed that even when the power grid information is unavailable, stealth attacks can still be accomplished [15], [16]. Zhang *et al.* [17] investigated the attack strategy with consideration of communication rate constraint in cyber–physical systems.

There are two approaches to defend against the malicious data attacks. The first is to protect the system beforehand from being attacked by adversaries. This can be achieved by either protecting a number of measurements to prevent stealth attacks [18], [19], or monitoring state information directly by the deployment of phasor measurements units (PMU) [8], [20]. In practice, it is not feasible to secure all measurements to prevent attacks due to the high cost. Instead, stealth malicious attacks can be prevented by protecting a carefully selected subset of measurements. A challenge of this approach is to search the effective small measurement subset to make them immune to attacks. Bobba *et al.* [18] chose the subsets for small power test systems using brute-force search.

The second approach to deal with malicious attacks is to identify the injected false data in measurements and then either abandon the contaminated data or correct them. Traditional false data detection methods are based on residue test [6], [21]. They cannot protect state estimation from carefully designed stealth attacks. Recently, with the advancement of smart grid, new detection methods have been proposed. A survey of the existing detection methods was provided in [22]. In [14], generalized likelihood ratio test is introduced to detect weak FDIAs. The cumulative sum (CUSUM) test-based detection mechanism introduced in [16] is also designed for nonstealth attacks. Esmalifalak *et al.* [23] discussed stealth false data detection methods using machine learning. Graphical methods are used to design defending mechanisms in [24]. In [11], an effective method capable of detecting false data as well as recovering the real state information was proposed. In [25], both the attack and detection algorithms were discussed, but the subset protection method was not considered. Additionally, only preliminary results regarding the attack strategies were presented in [25]. This paper substantially discusses the methods of sparse attack construction, the strategy of system protection from attacks, and the algorithm of stealth attack detection.

This paper has made three contributions: First, methods of constructing stealth attacks are proposed for two typical scenarios. We consider a general scenario in which adversaries can access arbitrary measurements to change arbitrary state variables in state estimation. To the best of our knowledge, there is no feasible algorithm that can efficiently construct highly sparse undetectable attack vectors in this case. In [7], it is observed that the optimal undetectable attack vector to compromise the minimum number of measurements can be found using brute-force search. However, this is not practical due to the high complexity. We propose an efficient and effective attack vector construction algorithm which can quickly generate highly sparse attack vectors in this scenario.

Liu *et al.* [7] also have demonstrated that stealth attack vectors always exist when the number of measurements that can be contaminated exceeds a certain value. However, it is shown in this paper that our proposed method can launch stealth attacks by manipulating only a much smaller number of measurements with high probability. Additionally, stealth attacks in a specific scenario are also considered in this paper. An optimization-based algorithm is introduced to generate sparse targeted attack vectors to bias specified state variables with the consideration that a subset of measurements is protected.

A fast greedy search method is then proposed to quickly find a subset of measurements to be protected to defend against stealth attacks. This fast method can find a subset with the same size as brute-force search in nearly all cases. Finally, inspired by Liu *et al.* [11], we introduce a detection algorithm considering the noise case with partial observations. The proposed algorithm extends the method in [11] to address the problem of detecting stealth attacks as well as recovering true state information with only partially collected contaminated measurements. The performance of the proposed algorithms is investigated using IEEE test systems with software MATPOWER [26].

This paper is organized as follows. Section II introduces power system model and the stealth attack problem in state estimation. In Section III, proposed attack strategies for different scenarios are introduced. Section IV provides the measurement protection algorithm and Section V presents the false data detection method. Simulation results are presented in Section VI. Section VII concludes the paper.

II. Power Grid and Attack Model

In this paper, we consider a power transmission grid which consists of $n+1$ buses and l transmission lines. The network connectivity of this power grid can be described by the $(n+1) \times l$ oriented incidence matrix $\mathbf{M}$, of which each column corresponds to the power line (i, j) and consists of all

0s except the ith and jth elements having value of 1 and -1, respectively [27]. The nonsingular diagonal matrix $\mathbf{D} \in \mathbb{R}^{l\times l}$ describes the physical properties of the transmission grid with diagonal entries equal to admittance of the transmission lines. SCADA collects measurements from RTUs such as bus voltages, bus power injections, and branch power flows from the power grid and sends them to the state estimator to estimate the state of the power system in the control center.

A. State Estimation

The state estimation problem is to use power measurements to timely estimate the state of the power system. Specifically, power system state refers to bus voltage angles θ and bus voltage magnitudes V. In the linearized dc power flow model [4 Ch. 2], bus magnitudes are assumed already known and are all close to unity. Additionally, phase angle at reference bus is set to zero radians; thus, estimation of only n bus voltage angles $[\theta_1, \theta_2, \dots, \theta_n]^T$ is required. The measurements have the following relationship with the state variables:

$$\boldsymbol{z} = \mathbf{H}\theta + \mathbf{e} \tag{1}$$

where $\boldsymbol{z} = [z_1, z_2, \dots, z_m]^T$ denotes measurement vector and $\mathbf{H} \in \mathbb{R}^{m\times n}$ is the measurement Jacobian matrix constructed by $\mathbf{H} = \begin{bmatrix} \mathbf{DM}^T \\ \mathbf{MDM}^T \end{bmatrix}$. $\mathbf{e} = [e_1, e_2, \dots, e_m]^T$ represents the Gaussian measurement noise and it is assumed to be zero-mean for convenience [14]. Measurements include power flows on l transmission lines and power injection at n buses. In the dc power flow model, power from bus i to j can be approximated as

$$P_{ij} = (\theta_i - \theta_j)\, d_{ij} \tag{2}$$

where d_{ij} is the admittance of the transmission line from bus i to j. Thus, power flows on transmission lines are computed by $\mathbf{MDM}^T\theta$ and power injections at buses are obtained from $\mathbf{DM}^T\theta$.

The state vector can be estimated from measurements using the weighted least-square (WLS) method [4]. In particular, system states are estimated as

$$\hat{\theta} = \left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{W}\boldsymbol{z} = \mathbf{K}\boldsymbol{z} \tag{3}$$

where diagonal weighting matrix $\mathbf{W}$ has diagonal entries equaling to the inverse of noise variances and $\mathbf{K} = \left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{W}$.

B. False Data Injection Attacks

Malicious false data can be injected by manipulating the RTU measurements to bias the estimated states. System measurements with malicious data becomes

$$\boldsymbol{z}_{\mathbf{a}} = \mathbf{H}\theta + \boldsymbol{a} + \mathbf{e} \tag{4}$$

where $\boldsymbol{a} = [a_1, a_2, \dots, a_m]^T$ denotes the attack vector.

Bad data in measurements can lead to incorrect state estimation and cause severe outcomes. Traditional methods to detect bad data are mostly based on the residue test. The residue vector $\boldsymbol{r}$ refers to the difference between the obtained measurements and the computed value from the estimated state

$$\boldsymbol{r} = \boldsymbol{z} - \mathbf{H}\hat{\theta}. \tag{5}$$

For example, the largest normalized residue (LNR) test identifies bad data if the absolute value of the largest element in $\boldsymbol{r}$ is larger than a certain threshold τ, i.e., $\max_i |r_i| > \tau$.

However, carefully designed malicious data attacks can bypass residue-based BDD. If attackers have knowledge about the power grid topology information, or $\mathbf{H}$, they can inject stealth attacks by constructing the attack vector such that [7]

$$\boldsymbol{a} = \mathbf{H}\mathbf{c}. \tag{6}$$

The measurements can then be written as

$$\boldsymbol{z}_{\mathbf{a}} = \mathbf{H}\left(\theta + \boldsymbol{c}\right) + \mathbf{e} \tag{7}$$

where $\boldsymbol{c} \in \mathbb{R}^n$ is any arbitrary vector and denotes the errors added to the state variables introduced by $\boldsymbol{a}$. The attack is undetectable as the residue $\boldsymbol{r}$ would not change compared to that without attack $\boldsymbol{a}$ [7]. The system will regard the manipulated state $(\theta + \boldsymbol{c})$ as the real value in the state estimator.

III. Stealth Attack Strategies

In order to evade detection in the control center, attack vectors are designed to satisfy (6). Additionally, attackers would tend to compromise as few measurements as possible in effort to launch attacks with least effort. Therefore attack strategies are expected to be able to construct highly sparse attack vectors. The stealth sparse attacks were first discussed in [7], in which the authors proposed that attackers can modify state variables in state estimation without being detected by modifying a small number of carefully chosen RTU measurements. In this paper, two methods are introduced to construct sparse attack vectors for two typical scenarios: random attacks in which arbitrary measurements can be compromised and targeted attacks in which specific state variables need to be biased.

A. Random Attacks

In this scenario, it is assumed that no measurements are protected, and the changes of state variables are not specified. Attackers can hack arbitrary measurements to bias arbitrary state variables. Thus, the aim is solely to find highly sparse vector $\boldsymbol{a}$ that satisfies (6). To the best of our knowledge, there is no feasible algorithm that can efficiently construct sparse attack vectors in this case. Since $\boldsymbol{a}$ is a linear combination of the columns of $\mathbf{H}$, it is possible to generate sparse vector by column transformation of $\mathbf{H}$. However, this method cannot guarantee sparsity. It is demonstrated by Liu *et al.* [7] that a k-sparse stealth attack vector always exist if $k > m - n$. We propose a novel method that can construct a sparse attack vector with much smaller k.

Liu *et al.* [7] also proposed that a projection matrix $\mathbf{P} = \mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T$ can lead to an equivalent criterion to generate attack vector $\boldsymbol{a}$ satisfying (6)

$$\mathbf{P}\boldsymbol{a} = \mathbf{H}\left(\mathbf{H^T H}\right)^{\mathbf{-1}}\mathbf{H^T H}\boldsymbol{c} = \mathbf{H}\boldsymbol{c}$$
$$(\mathbf{P} - \mathbf{I})\boldsymbol{a} = 0.$$

Let $\mathbf{B} = \mathbf{P} - \mathbf{I}$, then undetectable attack vector $\boldsymbol{a}$ satisfies

$$\mathbf{B}\boldsymbol{a} = 0. \tag{8}$$

This criterion can be used to generate attack vectors in certain scenarios such as when a subset of measurements is protected [7]. For random attacks, this criterion can also be utilized. A straightforward way to find sparse attack vectors can be formulated using the following optimization problem:

$$\begin{aligned} &\min \ \|\boldsymbol{a}\|_0 \\ &\text{s.t. } \mathbf{B}\boldsymbol{a} = 0, \ \boldsymbol{a} \neq 0 \end{aligned} \tag{9}$$

where l_0-norm is the number of the nonzero elements in $\boldsymbol{a}$. This is a nonconvex problem and finding the solution to this problem is highly complex. However, it is obvious that a feasible vector $\boldsymbol{a}$ must be in the null space of matrix $\mathbf{B}$, which is defined as

$$\text{Null}(\mathbf{B}) = \left\{\mathbf{v} \in \mathbb{R}^m \,|\mathbf{B}\mathbf{v} = 0\right\}. \tag{10}$$

Rather than solving the complex problem in (9), we propose an algorithm taking advantage of null space of $\mathbf{B}$, which can be easily computed.

Proposed algorithm. Measurements are always subject to noise due to the errors in the measuring process and the noise in the communication channels. The noise can be modeled as Gaussian distributed with variance Σ_e. The system is usually designed to be tolerant to measurement deviations within the noise level. Additionally, vectors in the null space of $\mathbf{B}$ usually comprise a small number of relatively large elements and the majority are small value elements. It provides the possibility for attackers to inject attack vectors designed based on vectors in $\text{Null}(\mathbf{B})$. Those small valued elements can be dealt with as noise if their average energy, denoted as Σ_B, is within the range of the variance of the noise Σ_e, namely $\Sigma_B \leq \Sigma_e$. Therefore, attackers only need to inject elements of large values in the column vectors of $\text{Null}(\mathbf{B})$ into the system and the number of measurements to be compromised will be greatly reduced. We define a shrinkage operation $\mathcal{S}_t$ as follows:

$$\mathcal{S}_t\left(x\right) := \begin{cases} \dfrac{x}{|x| - t}\max\left(|x| - t, \, 0\right), & |x| \neq t \\ x, & |x| = t. \end{cases} \tag{11}$$

The attack vector construction procedure can then be designed as follows. Given matrix measurement matrix $\mathbf{H}$, compute matrix $\mathbf{B}$ as well as the standard basis matrix $\mathbf{U}$ of its null space $\text{Null}(\mathbf{B})$ and choose vector $\boldsymbol{u}$ with the largest variance from all column vector in $\mathbf{U}$

$$\boldsymbol{u} = \arg\max_i\left(\text{var}\left(\mathbf{u}_i\right)\right) \tag{12}$$

where $\mathbf{u}_i$ denotes the ith column in $\mathbf{U}$. Then, scale vector $\boldsymbol{u}$ up, or down, till the maximal element reaches a designed attack value C. The last step is to force the small elements below threshold t to 0

$$\mathbf{a} = \mathcal{S}_t\left(\epsilon\boldsymbol{u}\right) \tag{13}$$

where $\epsilon = \frac{C}{\max(\boldsymbol{u})}$.

Algorithm 1 concludes the whole process of attack vector construction. It is notable that threshold t should be carefully chosen in the consideration of both sparsity and evading BDD. A higher t can generate a sparser attack vector but also increase the possibility of being detected. It is also notable that if the noise is not zero-mean, the threshold t is chosen according to the tolerable noise range. Since the measurement noise follows $\mathcal{N}(0, \sigma^2)$, it is assumed that all noise variables are within the range of $[-3\sigma, 3\sigma]$ (otherwise it will be identified as bad data). Thus, threshold t should not exceed 3σ. The following proposition can assist in choosing threshold t.

Algorithm 1. Sparse stealth attacks construction

```
Input: $\mathbf{H} \in \mathbb{R}^{m\times n}$, $C > 0, t > 0$.
Procedure:
  1) Compute $\mathbf{B} = \mathbf{H}\left(\mathbf{H^T H}\right)^{\mathbf{-1}}\mathbf{H^T} - \mathbf{I}$.
  2) Get the standard basis matrix $\mathbf{U}$ of $Null(\mathbf{B})$ so that $i$-th
     column $\mathbf{u}_i$: $\mathbf{B}\mathbf{u}_i = 0$.
  3) Find column vector $\mathbf{u}$ in $\mathbf{U}$: $\mathbf{u} = \arg\max_i(var(\mathbf{u}_i))$.
  4) Scale up/down vector $\mathbf{u}$ by $\epsilon$: $\mathbf{u}' = \epsilon\mathbf{u}$ and $\epsilon = \frac{C}{\max(\mathbf{u})}$.
  5) Shrink the vector using the threshold $t$ to obtain the sparse
     attack vector $\mathbf{a}$: $\mathbf{a} = \mathcal{S}_t\left(\mathbf{u}'\right)$.
Output: $\mathbf{a}$.
```

Proposition 1: If an attack vector $\boldsymbol{a}$ is constructed using Algorithm 1 with the shrinkage threshold t, the probability of successfully evading detection by residue-based detection algorithms in the system is at least

$$P_l(t) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{3\sigma - t}{\sigma\sqrt{2}}\right)\right] \tag{14}$$

where $\text{erf}(\cdot)$ refers to the Gauss error function and σ is the standard deviation of the Gaussian measurement noise.

Proof: Since the vector $\boldsymbol{u}$ is selected form $\text{Null}(\mathbf{B})$, it satisfies $\boldsymbol{u} = \mathbf{H}\boldsymbol{c}$, where $\boldsymbol{c} \in \mathbb{R}^n$. Let $\boldsymbol{a} = \mathcal{S}_t\left(\epsilon\boldsymbol{u}\right) = \epsilon\boldsymbol{u} - \boldsymbol{u}_t$. The residual vector $\boldsymbol{r}'$ when attack $\boldsymbol{a}$ is injected into system is calculated as

$$\begin{aligned} \boldsymbol{r}' &= \boldsymbol{z}_{\mathbf{a}} - \mathbf{H}\hat{\theta}_{\mathbf{a}} = \boldsymbol{z} + \boldsymbol{a} + \mathrm{e} - \mathbf{HK}(\boldsymbol{z} + \boldsymbol{a} + \mathrm{e}) \\ &= \mathbf{Hx} + \epsilon\mathbf{H}\boldsymbol{c} - \boldsymbol{u}_t + \mathrm{e} - \mathbf{H}(\mathbf{KHx}+\mathbf{K}\epsilon\mathbf{H}\boldsymbol{c}-\mathbf{K}\boldsymbol{u}_t+\mathbf{K}\mathrm{e}) \\ &= \mathbf{Hx} + \epsilon\mathbf{H}\boldsymbol{c} - \boldsymbol{u}_t + \mathrm{e} - \mathbf{Hx} - \epsilon\mathbf{H}\boldsymbol{c} + \mathbf{HK}\boldsymbol{u}_t - \mathbf{HK}\mathrm{e} \\ &= (\mathbf{I} - \mathbf{HK})(\mathrm{e} - \mathbf{u}_t). \end{aligned} \tag{15}$$

Equation (15) indicates that when an attack generated by Algorithm 1 exists, it can be regarded that the random noise is perturbed by small amounts. A very small element comparing to σ in $\boldsymbol{u}_t$ should not impact the noise level or the residue since the shifted noise variable is still within the tolerable range. To evaluate the probability of having no impact on noise level, we consider the worst case when all elements in vector $\boldsymbol{u}_t$ equal the threshold t. In this case, it can be viewed as that the noise level is shifted down by an amount of t. The shifted noise $\mathbf{e}'$ follows $\mathcal{N}(-t,\, \sigma^2)$. Therefore, the probability when the shifted noise variables are within the normal range of $[-3\sigma,\, 3\sigma]$ can be computed by

$$P_l(t) = \int_{-3\sigma}^{3\sigma} \frac{1}{\sigma 2\pi} \exp\left(-\frac{(k+t)^2}{2\sigma^2}\right) dk. \tag{16}$$

This probability can be evaluated using (14). In fact, as a large number of elements in $\boldsymbol{u}_t$ tend to be much smaller than t, nondetection probability $P(t) > P_l(t)$. Thus, proposition 1 is proved. ■

B. Targeted Attacks

In practice, adversaries may intend to modify specific state variables. In this case, the amounts in the targeted subset in the vector $\boldsymbol{c}$ are fixed. Sparse attack vector construction methods for targeted attacks have been extensively explored in the literature, e.g., [7], [8], and [12]. Additionally, certain measurements may be protected, and adversaries would not be able to compromise these secured measurements. It is notable that protecting all measurements may not be feasible due to the high cost. Therefore, sparse attack vectors need to be carefully designed to contaminate specific state variables without compromising those protected measurements.

Let $\mathcal{I}$ denote the indices of state variables that are specifically attacked. $\bar{\mathcal{I}}$ is the complementary set of $\mathcal{I}$ and denotes the indices of state variables that can be arbitrarily chosen to launch targeted attacks. Measurements Jacobian $\mathbf{H}$ is $[\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_n]$ where $\mathbf{h}_i$ denotes the ith column vector of $\mathbf{H}$. A stealth attack vector $\boldsymbol{a}$ can then be written as

$$\boldsymbol{a} = \mathbf{H}\boldsymbol{c} = \sum_{i \in \mathcal{I}} \mathbf{h}_i c_i + \sum_{j \in \bar{\mathcal{I}}} \mathbf{h}_j c_j. \tag{17}$$

In a targeted attack, the value of c_i, $i \in \mathcal{I}$ is fixed and predesigned to be injected into the state variables. Let $\boldsymbol{b} = \sum_{i \in \mathcal{I}} \mathbf{h}_i c_i$, which is predesigned by attackers. The attack vector $\boldsymbol{a}$ is then designed based on the fixed vector $\boldsymbol{b}$. As proposed in [7], (17) can be transformed using a projection matrix $\mathbf{P} = \mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T$. Since $\boldsymbol{a} - \boldsymbol{b} = \mathbf{H}_{\bar{\mathcal{I}}}\boldsymbol{c}_{\bar{\mathcal{I}}}$, where $\mathbf{H}_{\bar{\mathcal{I}}}$ denotes the submatrix of $\mathbf{H}$ containing columns with index in $\bar{\mathcal{I}}$, i.e., $\mathbf{H}_{\bar{\mathcal{I}}} = \left[\mathbf{h}_{j_1}, \mathbf{h}_{j_2}, \ldots, \mathbf{h}_{j_{|\bar{\mathcal{I}}|}}\right]$, where $j_i \in \bar{\mathcal{I}}$ for $1 \leq i \leq \left|\bar{\mathcal{I}}\right|$. By left-multiplying both sides with $\mathbf{P}_{\bar{\mathcal{I}}}$, we have

$$\begin{aligned} \mathbf{P}_{\bar{\mathcal{I}}}\left(\boldsymbol{a} - \boldsymbol{b}\right) &= \mathbf{P}_{\bar{\mathcal{I}}}\mathbf{H}_{\bar{\mathcal{I}}}\boldsymbol{c}_{\bar{\mathcal{I}}} \\ &= \mathbf{H}_{\bar{\mathcal{I}}}\left(\mathbf{H}_{\bar{\mathcal{I}}}^T\mathbf{H}_{\bar{\mathcal{I}}}\right)^{-1}\mathbf{H}_{\bar{\mathcal{I}}}^T\mathbf{H}_{\bar{\mathcal{I}}}\boldsymbol{c}_{\bar{\mathcal{I}}} \\ &= \mathbf{H}_{\bar{\mathcal{I}}}\boldsymbol{c}_{\bar{\mathcal{I}}} = \boldsymbol{a} - \boldsymbol{b} \end{aligned} \tag{18}$$

where $\mathbf{P}_{\bar{\mathcal{I}}} = \mathbf{H}_{\bar{\mathcal{I}}}\left(\mathbf{H}_{\bar{\mathcal{I}}}^T\mathbf{H}_{\bar{\mathcal{I}}}\right)^{-1}\mathbf{H}_{\bar{\mathcal{I}}}^T$.

We can then easily obtain that $(\mathbf{P}_{\bar{\mathcal{I}}} - \mathbf{I})\boldsymbol{a} = (\mathbf{P}_{\bar{\mathcal{I}}} - \mathbf{I})\boldsymbol{b}$, let $\mathbf{B} = \mathbf{P} - \mathbf{I}$, we have the following equivalent criteria for an attack vector $\boldsymbol{a}$ to be stealth

$$\mathbf{B}_{\bar{\mathcal{I}}}\boldsymbol{a} = \mathbf{B}_{\bar{\mathcal{I}}}\boldsymbol{b} \tag{19}$$

where $\mathbf{B}_{\bar{\mathcal{I}}} = \mathbf{P}_{\bar{\mathcal{I}}} - \mathbf{I}$.

Since a subset of measurements is protected, those elements in attack vector $\boldsymbol{a}$ should be restricted to 0. Let $\mathbf{y} = \mathbf{B}_{\bar{\mathcal{I}}}\boldsymbol{b}$, and assume that the pth measurement is secured, e.g., $a_p = 0$. Applying the l_1-relaxation, the sparse attack vector can be obtained by solving the following optimization problem:

$$\begin{aligned} &\min_{\mathbf{c}} \|\mathbf{H}\boldsymbol{c}\|_1 \\ &\text{s.t. } \mathbf{B}_{\bar{\mathcal{I}}}\mathbf{H}\boldsymbol{c} = \mathbf{y} \\ &\mathbf{H}^p\boldsymbol{c} = 0 \end{aligned} \tag{20}$$

where $\mathbf{H}^p$ denotes the pth row of matrix $\mathbf{H}$ and minimizing l_1-norm of a vector $\|\boldsymbol{v}\|_1 = \sqrt{\sum_i |v_i|}$ can promote $\boldsymbol{v}$ to be sparse. This problem is well discussed in the field of compressive sensing [28] and can be quickly solved.

IV. Strategic Protection

Increasing the number of protected measurements can make the stealth attacks more difficult to be accomplished. It is obvious that stealth attacks can be completely prevented by securing all measurements. However, it is not economical or necessary to secure all measurement devices to defend against stealth attacks. Bobba *et al.* [18] explored the minimal measurement subset that is required to be protected to defend against attacks using brute-force search. This method is time-consuming and only feasible for small-sized power grids. In this section, an efficient algorithm is proposed to quickly find measurement protection subsets, which have the same sizes as that from brute-force method in nearly all cases.

Let the set $\mathcal{P} \subset \{1, 2, \ldots, m\}$ be the measurement set that are secured and the complementary set $\bar{\mathcal{P}}$ denotes the index of those measurements that can be contaminated. Similar to (20), adversaries can construct the sparse attack vector $\boldsymbol{a}$ by solving

$$\begin{aligned} &\min_{\boldsymbol{c}} \left\|\mathbf{H}^{\bar{\mathcal{P}}}\boldsymbol{c}\right\|_1 \\ &\text{s.t.} \mathbf{B}_{\bar{\mathcal{I}}}\mathbf{H}\boldsymbol{c} = \mathbf{y} \\ &\mathbf{H}^{\mathcal{P}}\boldsymbol{c} = 0. \end{aligned} \tag{21}$$

If the protection set $\mathcal{P}$ is properly chosen, specific targeted attack vectors would not exist. Namely, (21) would have no solutions. Giving specified vector $\boldsymbol{c}_{\mathcal{I}}$, which is the targeted subset vector of $\boldsymbol{c}$, the straightforward method is to protect all measurements in the set corresponding to all nonzero elements in $\boldsymbol{a}$ that $\boldsymbol{a} = \mathbf{H}_{\mathcal{I}}\boldsymbol{c}_{\mathcal{I}}$. In this way, it probably requires a large number of measurements to be protected since $\boldsymbol{a}$ may not be desirably sparse. Finding or computing the smallest protection set that can prevent targeted attacks is difficult. The brute-force search method, which is discussed in [18], can guarantee finding the smallest possible sets. However, this method is extremely complex and not feasible in practice.

When a certain measurement is secured, attackers need to compromise more measurements or inject extra errors into the rest of the measurements to launch targeted attacks. From (17), we have

$$\boldsymbol{a} = \mathbf{H}\boldsymbol{c} = \boldsymbol{b} + \mathbf{H}_{\bar{\mathcal{I}}}\boldsymbol{c}_{\bar{\mathcal{I}}} \tag{22}$$

where $\boldsymbol{b}$ represents predesired injections. It is obvious that protecting certain measurements can always be more effective than others. For example, it is more important to secure the measurements corresponding to the nonzero elements in $\boldsymbol{b}$ than others. If a subset $\mathcal{P}$ of the total measurements is protected, we have

$$\boldsymbol{a}_{\mathcal{P}} = \boldsymbol{b}_{\mathcal{P}} + \mathbf{H}_{\bar{\mathcal{I}}}^{\mathcal{P}} \boldsymbol{c}_{\bar{\mathcal{I}}} = 0 \tag{23}$$

$$-\boldsymbol{b}_{\mathcal{P}} = \mathbf{H}_{\bar{\mathcal{I}}}^{\mathcal{P}} \boldsymbol{c}_{\bar{\mathcal{I}}}. \tag{24}$$

If the rank of $\mathbf{H}_{\bar{\mathcal{I}}}^{\mathcal{P}}$ is smaller than protection size $|\mathcal{P}|$, and the augmented matrix with vector $\boldsymbol{b}_{\mathcal{P}}$ can increase the rank, namely $\operatorname{rank}\left(\left[\mathbf{H}_{\bar{\mathcal{I}}}^{\mathcal{P}} | \boldsymbol{b}_{\mathcal{P}}\right]\right) = \operatorname{rank}\left(\mathbf{H}_{\bar{\mathcal{I}}}^{\mathcal{P}}\right) + 1$, then $\boldsymbol{c}_{\bar{\mathcal{I}}}$ satisfying (24) does not exist, indicating that the system is successfully protected from targeted attacks with $\boldsymbol{b}$. Otherwise, when vector $\boldsymbol{b}_{\mathcal{P}}$ cannot increase the rank of matrix $\mathbf{H}_{\bar{\mathcal{I}}}^{\mathcal{P}}$, i.e., $\operatorname{rank}\left(\left[\mathbf{H}_{\bar{\mathcal{I}}}^{\mathcal{P}} | \boldsymbol{b}_{\mathcal{P}}\right]\right) = \operatorname{rank}\left(\mathbf{H}_{\bar{\mathcal{I}}}^{\mathcal{P}}\right)$, there exist solutions of $\boldsymbol{c}_{\bar{\mathcal{I}}}$, which means that adversaries can still find attack vectors to launch targeted attacks. The problem is then to find the best solution to obtain highly sparse $\boldsymbol{a}$. It is known from (22) that making a certain subset $\mathcal{P}$ of the measurements immune to attacks can result in an attack vector $\boldsymbol{a}$ which contaminates more state variables. This makes the attacks more difficult to be accomplished. Therefore, it can be deduced that protecting certain measurement would result in a larger $\|\boldsymbol{a}\|_1$ value than that of protecting another measurement. Protecting these measurements would be more effective than others and these measurements can be regarded as *critical measurements* to targeted attacks. Based on this idea, giving specified targeted state bias vector $\boldsymbol{c}_{\mathcal{I}}$, we can design a greedy method to search a small subset of these measurements to be protected to defend from targeted attacks.

Algorithm 2 presents the greedy search method to find a small protection subset of measurements with the knowledge of existing protection set and targeted vector $\boldsymbol{c}_{\mathcal{I}}$. At each iteration, the algorithm assume that one more measurement is protected and check the feasibility of constructing attack vector $\boldsymbol{a}$. If the stealth attack vector exists when every measurement is protected one by one, the algorithm increases the protection set by selecting the most important measurement, which leads to the largest value of $\|\boldsymbol{a}\|_1$ when it is protected. The selection process continues until stealth targeted attack vector does not exist.

Algorithm 2. Greedy subset searching Algorithm

```
Input: $\mathbf{H}$, $\mathcal{I}$, $\mathbf{c}_{\mathcal{I}}$, $\mathcal{P}$.
Initialize: $\mathbf{B}_{\bar{\mathcal{I}}} = \mathbf{H}_{\bar{\mathcal{I}}}\left(\mathbf{H}_{\bar{\mathcal{I}}}^T \mathbf{H}_{\bar{\mathcal{I}}}\right)^{-1} \mathbf{H}_{\bar{\mathcal{I}}}^T - \mathbf{I}$,  $\mathbf{y} = \mathbf{B}_{\bar{\mathcal{I}}} \mathbf{H}_{\mathcal{I}} \mathbf{c}_{\mathcal{I}}$,
$\mathcal{P}' = \mathcal{P}$, $k = 1$, $\mathcal{P}_k = \mathcal{P}'$.
Iteration: At the $k$-th iteration:
Compute the complementary set $\bar{\mathcal{P}}$ of $\mathcal{P}'$.
For $i = 1 : |\bar{\mathcal{P}}|$
    Put the $i$-th entry in $\bar{\mathcal{P}}$ into $\mathcal{P}_k$: $\mathcal{P}_k = \mathcal{P}' \cup \bar{\mathcal{P}}_i$.
    Checking the feasibility of finding $\mathbf{c}$ from equation (21).
    If feasible
        Compute $\chi_i = \|\mathbf{Hc}\|_1$.
    else
        $\mathcal{P}' = \mathcal{P}_k$; Quit the iteration.
    end
end
Find index $i$ such that $\chi_i$ has the largest value.
Update set $\mathcal{P}' = \mathcal{P}' \cup \bar{\mathcal{P}}_i$.
Output: $\mathcal{P}'$.
```

For a large power grid system, it is not feasible to find the smallest protection subset to prevent any of undetectable attacks that satisfy $\boldsymbol{a} = \mathbf{H}\boldsymbol{c}$ by brute-force search. Instead, we can protect the union set of those subsets selected for protecting every single state variable. Our proposed method can quickly find a small subset that protect the whole system from any stealth attacks satisfying (6). The search procedure can be concluded in Algorithm 3.

Algorithm 3. Minimal subset selection algorithm

```
Input: $\mathbf{H}$.
Initialize: $\mathcal{P} = 0$.
For $i = 1 : n$
    Let $\mathcal{I} = \{i\}$.
    Find $\mathcal{P}_i$ using Algorithm 2.
end
$\mathcal{P} = \mathcal{P}_1 \cup \mathcal{P}_2 \cup \ldots \cup \mathcal{P}_n$.
Output: $\mathcal{P}$.
```

This method cannot guarantee the smallest subset that can be found, but it provides at least a quasi-optimal subset that contains a slightly larger number of elements. Most importantly, this method is fast and feasible in practice. In the worst case, to find a protection set with k elements, the algorithm needs to test the feasibility $\mathcal{K}_a$ times

$$\mathcal{K}_a = mk - \frac{k(k-1)}{2}. \tag{25}$$

This figure is much smaller than that using brute-force method, where it needs to test $\sum_{i=1}^{k-1}\binom{m}{i} + 1$ combinations in the best case to find the protection set with k elements. Although our proposed algorithm may not find the global optimum solution, it provides some flexibility. When it is not possible to protect certain selected measurement device in practice, the algorithm can find a suboptimal subset instead.

V. Robust Detection

Traditional residue testing-based false data detection methods cannot provide protection of state estimation from carefully designed stealth attacks. Therefore, new detection methods need to be designed to detect random errors as well as stealth attacks. It is shown that a series of measurement data exhibit low rank and sparse structure, which can be employed in anomaly detection method [11]. In practice, measurements tend to be contaminated with noise. Additionally, it may also happen that some of the measurements are lost due to the measurement device failures or disrupted communication links. In this section, these situations are addressed.

Considering a time interval T, the power system obtains a series of measurements $[\mathbf{z}_{\mathbf{a}1}, \mathbf{z}_{\mathbf{a}2}, \ldots, \mathbf{z}_{\mathbf{a}T}]$ at the time instants $t_1, t_2, \ldots, t_T$. These measurements can form a matrix $\mathbf{Z}_{\mathbf{a}} \in \mathbb{R}^{m \times T}$, which can be decomposed as

$$\mathbf{Z}_{\mathbf{a}} = \mathbf{Z} + \mathbf{A} + \mathbf{E} \tag{26}$$

where $\mathbf{Z} \in \mathbb{R}^{m \times T}$ is the block of true measurements with each column $\mathbf{z}_i$ representing true measurements at time t_i, $\mathbf{A} \in$

$\mathbb{R}^{m\times T}$ denotes the attack matrix formed by all instant sparse attacks, and $\mathbf{E}$ represents the noise.

It is known that fast system dynamics are usually well damped in the power system. This implies that the system states would change gradually in a small period T, making the matrix $\mathbf{Z}$ typically low rank. Additionally, malicious injection data matrix $\mathbf{A}$ tends to be sparse. This is due to the fact that some of the measurements may be protected and also because attackers would launch attacks with least effort. Given corrupted measurements matrix $\mathbf{Z_a}$, it is possible to recover low-rank matrix $\mathbf{Z}$ and sparse-attack matrix $\mathbf{A}$ by performing low rank and sparse decomposition, which is well discussed in the robust principle component analysis (PCA) problem [29], which solves

$$\begin{aligned}&\min \|\mathbf{Z}\|_* + \lambda \|\mathbf{A}\|_1\\ &\text{s.t.}\ \|\mathbf{Z_a} - \mathbf{Z} - \mathbf{A}\|_F \le \delta\end{aligned} \tag{27}$$

where $\|\cdot\|_*$ denotes the nuclear norm, $\|\cdot\|_F$ denotes the Frobenius norm, δ represents a small positive noise bound, and λ is the regulation parameter. This problem is also addressed in compressive sensing and matrix completion [30] literature. Thus, true measurements can be recovered and the sparse perturbations including malicious attacks and other false data can also be identified.

Unlike coordinated malicious attacks, the missing data in the measurements can result in residue changes in (5). These incomplete measurement data, as well as the measurements with errors, would be identified as bad data by traditional BDD algorithms. The proposed algorithm can not only detect the missing and inaccurate measurement data but also detect the carefully designed stealth attacks, which is undetectable to traditional methods. More importantly, the proposed algorithm can recover the true measurements from the incomplete measurements.

In order to address the problem that only noise-contaminated partial measurements are collected, the PCA problem can be extended to the following form with element-wise error constraints:

$$\begin{aligned}&\min \|\mathbf{Z}\|_* + \lambda \|\mathbf{A}\|_1\\ &\text{s.t.}\ |\mathcal{P}_\Omega(\mathbf{Z_a}) - \mathcal{P}_\Omega(\mathbf{Z}+\mathbf{A})| \preceq \epsilon\end{aligned} \tag{28}$$

where $\preceq$ represents element-wise inequality and $\mathcal{P}_\Omega(\cdot)$ denotes a projection operation, in which all elements outside the set Ω are forced to 0. ϵ is the matrix of entry-wise error bounds. It is demonstrated in [31] that this problem is equivalent to the following problem:

$$\begin{aligned}&\min \|\mathbf{Z}\|_* + \lambda \|\mathcal{T}(\mathbf{A},\tilde{\epsilon})\|_1\\ &\text{s.t.}\ \mathbf{Z_a} = \mathbf{Z}+\mathbf{A}\end{aligned} \tag{29}$$

where $\tilde{\epsilon}$ has the same value as ϵ in the projection set Ω and infinite outside set Ω, and the soft thresholding operation $\mathcal{T}_\epsilon(a_{ij})$ is defined as

$$\mathcal{T}(a_{ij},\epsilon) = \text{sign}(a_{ij}) \cdot \max\{|a_{ij}| - \epsilon, 0\}. \tag{30}$$

A variant of the augmented Lagrangian method (ALM), which is also known as the alternating direction method of multipliers (ADMM) algorithm [32], is used to solve the problem defined by (29). The Lagrangian corresponding to this problem is

$$\mathcal{L}(\mathbf{Z},\mathbf{A},\mathbf{Y},\mu) = \|\mathbf{Z}\|_* + \lambda\|\mathcal{T}(\mathbf{A},\tilde{\epsilon})\|_1 + \langle \mathbf{Y},\mathbf{H}\rangle + \frac{\mu}{2}\|\mathbf{H}\|_2^2 \tag{31}$$

where $\langle\cdot\rangle$ denotes the Frobenius product, $\mathbf{H} = \mathbf{Z_a} - \mathbf{Z} - \mathbf{A}$ and $\mu > 0$. λ can be set to $\sqrt{m/|\Omega|}$. We further define the singular-value thresholding operation as

$$\mathcal{D}(\mathbf{X},\tau) = \mathbf{U}\mathcal{T}(\Sigma,\tau)\mathbf{V}^T$$

where τ is the threshold and $\mathbf{X} = \mathbf{U\Sigma V}^T$. It is notable that ADMM updates $\mathbf{Z}, \mathbf{A}, \mathbf{Y}$ separately only once in each iteration, so it is efficient. The convergence of the whole algorithm is analyzed in [32], which states that the condition for convergence requires $\sum_1^\infty \mu_k^{-1} = +\infty$ where μ_k denotes the value of μ in the kth iteration. The whole process of solving (29) is shown in Algorithm 4.

Algorithm 4. RPCA with entry wise constraints

```
Input: $\mathbf{Z_a^p} = \mathcal{P}_\Omega(\mathbf{Z_a}) \in \mathbb{R}^{m\times T}$, $\tilde{\epsilon} \in \mathbb{R}^{m\times T}$, $\lambda$.
Initialize $\mathbf{Z} = \mathbf{0}$, $\mathbf{A} = \mathbf{0}$, $\mathbf{Y} = \mathbf{0}$, $\mu > 0$, $\rho > 1$, $k = 0$.
while not converged
  1) Update the value of low rank matrix $\mathbf{Z}_{k+1}$:
     $\mathbf{Z}_{k+1} = \mathcal{D}\left(\mathbf{Z_a^p} - \mathbf{A}_k + \frac{\mathbf{Y}_k}{\mu_k}, \mu_k^{-1}\right)$.
  2) Compute the value of sparse matrix $\mathbf{A}_{k+1}$ by minimizing:
     $F(\mathbf{A}) = \frac{\lambda}{\mu}\|\mathcal{T}(\mathbf{A},\tilde{\epsilon})\|_1 - \text{tr}\left(\frac{\mathbf{Y}_k}{\mu_k}(\mathbf{A} - (\mathbf{Z_a^p} - \mathbf{Z}_k))\right) + \frac{1}{2}\|\mathbf{A} - (\mathbf{Z_a^p} - \mathbf{Z}_k)\|_F$.
  3) Update the Lagrange multiplier $\mathbf{Y}$:
     $\mathbf{Y}_{k+1} = \mathbf{Y}_k + \mu_k(\mathbf{Z_a^p} - \mathbf{Z}_{k+1} - \mathbf{A}_{k+1})$.
  4) Update $\mu_{k+1} = \rho \cdot \mu_k$.
  5) Update $k = k + 1$.
end while
Return $\mathbf{Z}$, $\mathbf{A} = \mathcal{T}(\mathbf{A},\tilde{\epsilon})$.
Output $\mathbf{Z}$, $\mathbf{A}$.
```

It is notable that when incomplete measurements are collected, Algorithm 4 will take the missing data to be sparse anomalies and it can also recover the low-rank true measurement matrix and sparse anomaly matrix. However, the recovery accuracy would be impacted as the sparsity is changed. The recovered sparse attack matrix can ignore those injected data outside the observation set. Thus, it is more difficult to identify all malicious attacks with partial observations.

VI. Numerical Results

In this section, the algorithms introduced above are evaluated by simulations performed based on the IEEE test systems [33]. The MATLAB package MATPOWER [26] is used to simulate the power system. The convex optimization problems are solved using the convex optimization toolbox CVX [34].

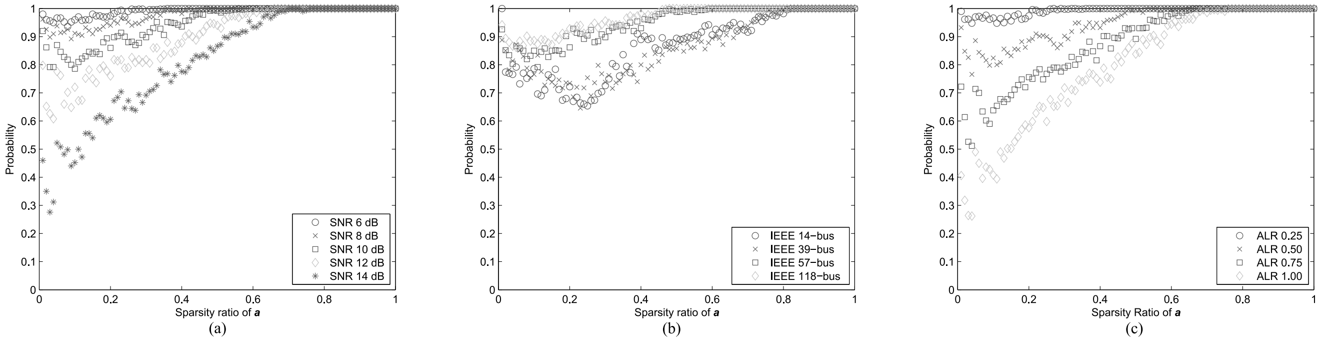


Fig. 2. Probabilities of successful attack injections (a) under different SNRs for IEEE-57 bus system, SR is 0.4; (b) for different bus systems, SR is 0.4 and $\mathrm{SNR} = 10\,\mathrm{dB}$; (c) for different SRs in IEEE-57 bus system, $\mathrm{SNR} = 10\,\mathrm{dB}$. (c) This figure utilizes random columns in $\mathrm{Null}(\mathbf{B})$ rather than that with largest variance.

A. Performance of Stealth Attack Construction

The performance of Algorithm 1 which generates highly sparse undetectable attack vectors is tested in different scenarios. Figs. 2 and 6 show the probabilities of successfully generating undetectable attack vectors with different levels of sparsity and different attack level ratios (ALRs), respectively. An attack is regarded as successful when the maximum value in residue vector does not exceed that without attacks. Sparsity ratio (SR) is defined as k/m, where k is the number of nonzero elements in $\boldsymbol{a}$ and m is the size of $\boldsymbol{a}$. The ALR is defined as the maximum attack value C to the mean value of the state variables: $\frac{C}{\mathrm{mean}(\theta)}$. Generally, these figures reveal that there are high probabilities for Algorithm 1 to successfully generate highly sparse undetectable attacks.

The noise in the simulation is modeled as Gaussian distributed with zero-mean. The signal-to-noise ratio (SNR) indicates the noise level compared with true measurements in the simulation. The noise may be due to measuring devices and process, or due to the communication channel noise. It is clear in Figs. 2(a) and 6(a) that in a relatively noisy case, the probability of a successful attack is extremely high (close to 1). In the low noise case, there is also high probability of injecting a successful highly sparse undetectable malicious attack. The algorithm is also assessed using different power grid system models, which is shown in Figs. 2(b) and 6(b). It is notable that in a larger bus system, Algorithm 1 can provide a better performance even for extremely sparse attacks and high ALRs. For example, the success ratio is around 90% for IEEE 118-bus system to generate stealth attacks with SR lower than 0.1, compared with 75% for IEEE 14-bus system shown in Fig. 2(b). This probability is 100% for IEEE 118-bus system to generate attacks with $\mathrm{ALR} = 1$ compared to 80% for IEEE 57-bus system shown in Fig. 6(b). Therefore, it can be anticipated that the algorithm would have a better performance in a real power system, which is much larger than the tested systems.

Additionally, it can be seen from Fig. 2 that it is always harder to inject sparser attacks while Fig. 6 reveals that attacks with higher values would be more likely to be detected. Figs. 2(c) and 6(c) display the performance when injecting attacks with different SRs and ALRs. It is notable that in Fig. 6(c), the algorithm utilizes randomly selected columns in a basis matrix of $\mathrm{Null}(\mathbf{B})$ rather than that with the largest variance. The results imply that using randomly chosen columns can also successfully inject undetectable attacks with high probabilities.

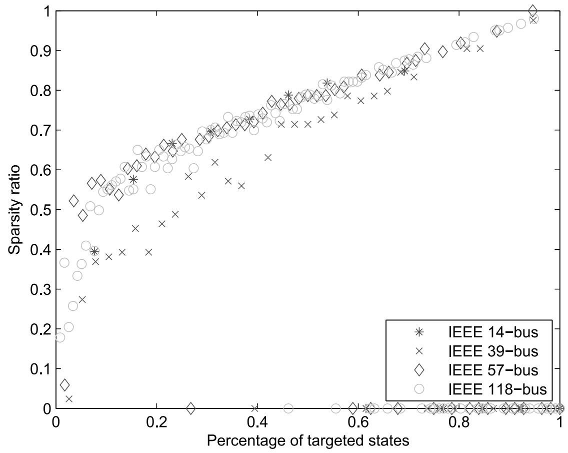


Fig. 3. Sparsity of $\boldsymbol{a}$ under different attack conditions.

TABLE I
NUMBER OF MEASUREMENTS IN PROTECTION SETS FROM TWO METHODS

Test systems	Algorithm 2	Brute-force search
IEEE 9-bus	9	8
IEEE 14-bus	15	13
IEEE 39-bus	59	–
IEEE 57-bus	84	–
IEEE 118-bus	187	–

It is known that stealth attacks having $m - n$ nonzero entries can always be found. In IEEE 57-bus system, this figure is 80, for which the SR is about 59%. However, by using Algorithm 1, there is still a high probability that attackers can inject undetectable attacks with SRs lower than 59%. Even for an attack with SR lower than 0.05, the success rate is still around 80% when the SNR is 10 dB and ALR is 0.5.

Targeted attack construction method in (20) is assessed under different attack conditions in which different percentage of total state variables are assumed to be modified. The targeted set is randomly selected and the protected measurement is

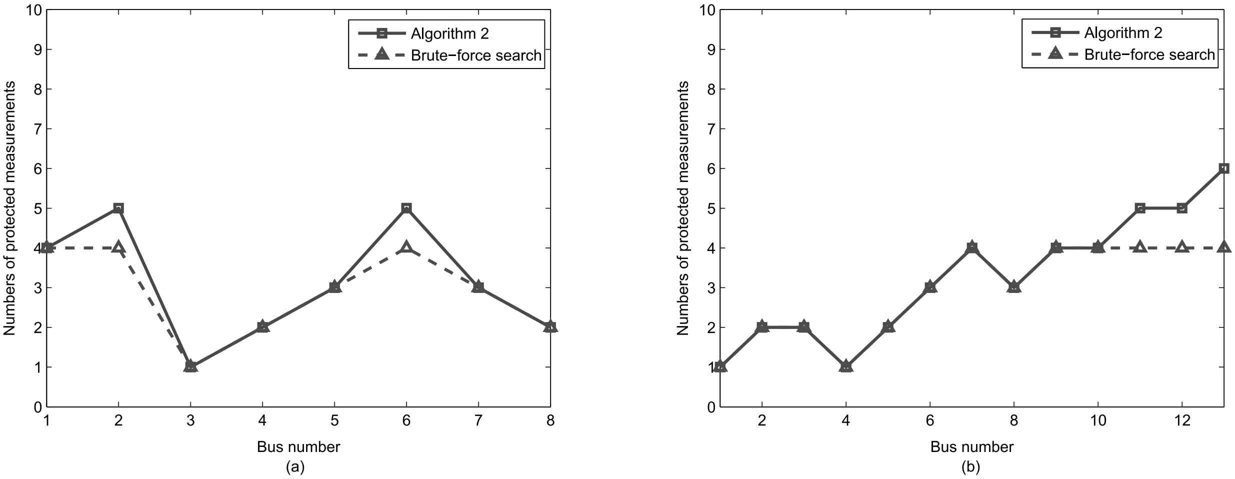


Fig. 4. Number of protected measurements to protect every single state variable from being targeted. (a) IEEE-9 bus system. (b) IEEE-14 bus system.

TABLE II
NUMBER OF TESTING TIMES FOR TWO ALGORITHMS TO FIND PROTECTION SUBSETS

Test systems	Bus number	1	2	3	4	5	6	7	8	9	10	11	12	13
IEEE-9 bus	Algorithm 2	55	70	1	18	34	75	40	25	–	–	–	–	–
	Brute-force search	1434	834	1	18	154	835	235	25	–	–	–	–	–
IEEE-14 bus	Algorithm 2	1	35	39	2	42	71	107	72	109	123	141	135	183
	Brute-force search	1	35	36	2	73	629	7108	630	7140	6235	6236	6237	7109

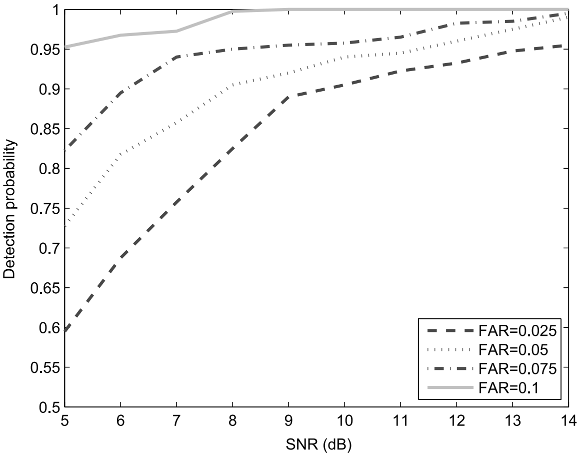


Fig. 5. Probability of successful false data detection.

TABLE III
DETECTION PROBABILITY AND MEASUREMENTS DEVIATION WITH PARTIAL OBSERVATIONS

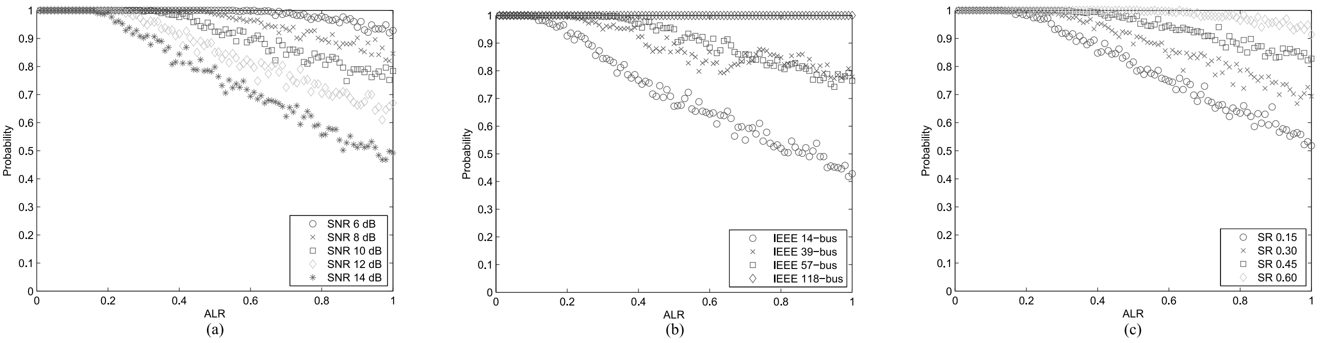
Observations (%)	Detection probability (%)	Variable deviation (%)
100	100	0.53
95	92	2.06
90	56.3	3.7
85	44.7	5.3

also randomly chosen. It can be observed from Fig. 3 that, in order to precisely alter specified state variables, the coordinated attack vectors cannot be highly sparse. Thus, attackers need to compromise a number of measurements to launch targeted attacks. Highly sparse attacks can only be achieved when the percentage of targeted state variables is extremely low for certain test systems. For example, SR can be less than 0.1 for IEEE 39-bus system when a small number of state variables are targeted. The figure also shows that in some cases, SR of attacks are 0. They correspond to the cases that: for certain targeted set of state variables, no feasible attack vectors exist when the pth measurement is protected. Therefore, it implies that when certain carefully selected measurements are protected, attackers may not be able to inject targeted attacks.

B. Performance of Strategic Protection

This section evaluates our proposed protection algorithm. To compare the protection subset generated by the proposed algorithm with that from brute-force method, we apply IEEE-9 bus system, which contain 17 total measurements, and IEEE-14 bus system with 33 total measurements. Table I shows the number of measurements in protection subset found by two methods. The results from the proposed algorithm for other larger test systems are also provided. In the first two test systems, the smallest protection sets generated from the proposed algorithm contain only slightly more measurements than that from brute-force method. In IEEE-14 bus system, the difference of this number is quite small compared to the total number of 33 measurements. Thus, Algorithm 2 can find protection subsets with similar number of elements but spend much less time than brute-force method.

Fig. 4 displays the number of elements in the smallest protection subsets to protect every single state variable from being targeted by adversaries. The whole system protection subsets shown in Table I are the unions of the protection subsets for protecting single variables. From both figures, it can be seen that in most cases the proposed algorithm can find a protection subset having the same size as that found by brute-force method. The size differences are only 1 or 2 when the two methods find



(a) (b) (c)

Fig. 6. Probabilities of successful attack injections (a) under different SNRs for IEEE-57 bus system, ALR is 0.5; (b) for different bus systems, ALR is 0.5 and $\mathrm{SNR} = 10\,\mathrm{dB}$; and (c) for different ALRs for IEEE-57 bus system, $\mathrm{SNR} = 10\,\mathrm{dB}$.

subsets with different number of elements. This number is quite small compared with the total number of 33 measurements in IEEE-14 bus.

Table II compares the complexities of the two algorithms in terms of the number of feasibility testing times. The results correspond with the simulation shown in Fig. 4 in which measurement protection subsets are searched for protecting every single state variable. It is obvious that when the size of protection subset exceeds 3, the difference of the two methods becomes significant. This difference is more significant when the size of protection subset is bigger as the brute-force search needs to exhaust all subset combinations with smaller sizes. It is also clear that in a larger power system, the difference is much larger for two algorithms to find a subset with same size as that in a smaller system. The testing times of the proposed algorithm will increase only slightly when the size of protection subset and the system scale grow, which is also described by (25). In a real power system, while brute-force method is infeasible because of the combinatorial complexity, the proposed method instead is fast and practical.

C. Performance of Detection

The performance of the detection algorithm is tested on IEEE 14-bus system and IEEE 57-bus system. The malicious attack vectors are constructed using our proposed Algorithm 1. In order to obtain sparsity in the rows of the attack block matrix, different column vectors in the null space in Algorithm 1 are utilized. The SR of the attacks is chosen as 15%. In Fig. 6(c), it is shown that when $\mathrm{SR} = 0.15$, traditional residual testing-based algorithms will not be able to detect those attacks. Thus, in the simulation, the algorithm is not compared with traditional methods. Additionally, recently proposed algorithms such as [23] do not deal with partial observations. These algorithms do not address the problem of error contaminated measurements as well. The detection method discussed in this paper addresses both problems. Most importantly, it can not only detect anomalies but also recover the true measurements from partial-contaminated observations.

We use the false alarm rate (FAR) which is the probability of positive alarm when there are no attacks. The noise performance of the algorithm compared to RPCA with Frobenius constraints in (27) has been extensively studied in [31]. In this paper, we focus on identifying anomalies in different scenarios when undetectable attacks are injected in power systems.

Fig. 5 shows the error tolerance performance in the IEEE 14-bus system. It is shown that when FAR exceeds 10%, the algorithm can identify attacks with high probabilities which are approaching 100%. This probability is still quite high in the presence of highly dense noise (95%). When FAR decreases, the system will absorb more noise and detection probability decreases. It can be seen that there is still a high chance of detecting anomalies: more than 90% when FAR decreases to an extremely low level (0.025) under $\mathrm{SNR} = 10\,\mathrm{dB}$.

In the case where partial measurements are collected, missing data are regarded as sparse anomalies in Algorithm 4. Additionally, nonzero entries in sparse matrix $\mathbf{A}$ can only be confirmed as attacks when they are located in the observation set. This make identifications of attacks more difficult. Algorithm 4 can circumvent this problem since it also recovers the block of true measurements. We evaluate the attack detection probabilities as well as the deviation rate of the recovered measurement variables, which is defined as $\|\mathbf{z} - \mathbf{z}'\|_2 \,/\, \|\mathbf{z}\|_2$. Table III shows the results when incomplete measurements are collected based on the IEEE 57-bus system. The FAR equals 0.05 and SNR is set to 8 dB. It can be seen that attack detection probability declines greatly with increasing missing observations. However, the recovered measurement variables experience only small deviations. Therefore, the proposed algorithm can successfully verify the true measurements, even in the situation that only partial measurements are observed (Fig. 6).

VII. Conclusion

In this paper, we looked into the problem of malicious FDIAs in power grid state estimation. We proposed stealth attack construction strategies for different scenarios and also introduced the countermeasures. It is shown that our proposed random attack construction algorithm can generate extremely sparse attack vectors. These optimal or quasi-optimal attacks can be achieved with high probability of success. The targeted undetectable attacks are obtained based on a optimization framework. The results show that attack vectors in this scenario cannot be extremely sparse, which is also discussed in literature. An efficient protection scheme is proposed in this paper to find an effective measurement protection subset to defend from the stealth attacks. The simulation results reveal that this subset searching algorithm can find a subset with almost the same size as that from the brute-force method. Additionally, a detection algorithm is introduced to detect the stealth attacks as well as other false data. This algorithm considers the case in which only partial measurements are collected in the presence of noise. The performance is demonstrated via the simulation results based on IEEE test power systems.

Acknowledgment

The authors would like to thank their colleagues at Toshiba Research Europe Ltd., for the fruitful discussions and the support of its directors.

References

[1] X. Fang, S. Misra, G. Xue, and D. Yang, "Smart grid—The new and improved power grid: A survey," *IEEE Commun. Surv. Tuts.*, vol. 14, no. 4, pp. 944–980, Oct. 2012.

[2] W. Wang and Z. Lu, "Cyber security in the smart grid: Survey and challenges," *Comput. Netw.*, vol. 57, nos. 5 and 7, pp. 1344–1371, Apr. 2013.

[3] D. B. Rawat and C. Bajracharya, "Cyber security for smart grid systems: Status, challenges and perspectives," in *Proc. SoutheastCon*, Apr. 2015, pp. 1–6.

[4] A. Abur and A. G. Expósito, *Power System State Estimation Theory and Implementation*. Boca Raton, FL, USA: CRC Press, Mar. 24, 2004.

[5] A. Monticelli, "Electric power system state estimation," *Proc. IEEE*, vol. 88, no. 2, pp. 262–282, Feb. 2000.

[6] F. C. Schweppe, J. Wildes, and D. P. Rom, "Power system static state estimation, parts I, II, III," *IEEE Trans. Power App. Syst.*, vol. PAS-89, pp. 120–135, Jan. 1970.

[7] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proc. ACM Conf. Comput. Commun. Sec.*, 2009, pp. 21–32.

[8] 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.

[9] Z. Yu and W. Chin, "Blind false data injection attack using PCA approximation method in smart grid," *IEEE Trans. Smart Grid*, vol. 6, no. 3, pp. 1219–1226, May 2015.

[10] Y. Li and Y. Wang, "State summation for detecting false data attack on smart grid," *Int. J. Elect. Power Energy Syst.*, vol. 57, pp. 156–163, May 2014.

[11] L. Liu, M. Esmalifalak, and Z. Han, "Detection of false data injection in power grid exploiting low rank and sparsity," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 9–13, 2013, pp. 4461–4465.

[12] M. Ozay, I. Esnaola, F. T. Y. Vural, S. R. Kulkarni, and H. V. Poor, "Sparse attack construction and state estimation in the smart grid: Centralized and distributed models," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1306–1318, Jul. 2013.

[13] H. Sandberg, A. Teixeira, and K. H. Johansson, "On security indices for state estimators in power networks," in *Proc. 1st Workshop Secure Control Syst. (CPSWEEK'10)*, 2010.

[14] 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.

[15] M. A. Rahman and H. Mohsenian-Rad, "False data injection attacks with incomplete information against smart power grids," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 3–7, 2012, pp. 3153–3158.

[16] Y. Huang *et al.*, "Bad data injection in smart grid: attack and defense mechanisms," *IEEE Commun. Mag.*, vol. 51, no. 1, pp. 27–33, Jan. 2013.

[17] H. Zhang, P. Cheng, J. Wu, L. Shi, and J. Chen, "Online deception attack against remote state estimation," in *Proc. World Congr. Int. Fed. Autom. Control (IFAC)*, 2014, pp. 128–133.

[18] R. B. Bobba, K. M. Rogers, Q. Wang, H. Khurana, K. Nahrstedt, and T. J. Overbye, "Detecting false data injection attacks on dc state estimation," in *Preprints 1st Workshop Secure Control Syst. (CPSWEEK'10)*, 2010, pp. 1–9.

[19] G. Dán and H. Sandberg, "Stealth attacks and protection schemes for state estimators in power systems," *Proc. 1st IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, pp. 214–219, 4–6 Oct. 2010.

[20] M. Gol and A. Abur, "PMU placement for robust state estimation," in *Proc. North Amer. Power Symp. (NAPS)*, Sep. 22–24, 2013, pp. 1–5.

[21] E. Handschin, F. C. Schweppe, J. Kohlas, and A. Fiechter, "Bad data analysis for power system state estimation," *IEEE Trans. Power App. Syst.*, vol. 94, no. 2, pp. 329–337, Mar. 1975.

[22] S. Cui, Z. Han, S. Kar, T. T. Kim, H. V. Poor, and A. 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.

[23] M. Esmalifalak, N. T. Nguyen, R. Zheng, and Z. Han, "Detecting stealthy false data injection using machine learning in smart grid," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 9–13, 2013, pp. 808–813.

[24] S. Bi and Y. J. Zhang, "Graphical methods for defense against false-data injection attacks on power system state estimation," *IEEE Trans. Smart Grid*, vol. 5, no. 3, pp. 1216–1227, May 2014.

[25] J. Hao, R. J. Piechocki, D. Kalesh, W. H. Chin, and Z. Fan, "Optimal malicious attack construction and robust detection in smart grid cyber security analysis," in *Proc. IEEE Int. Conf. Smart Grid Commun. SmartGridComm)*, Nov. 3–6, 2014, pp. 836–841.

[26] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, "Matpower: Steady-state operations, planning and analysis tools for power systems research and education," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, Feb. 2011.

[27] Z. Wang, A. Scaglione, and R. J. Thomas, "Generating statistically correct random topologies for testing smart grid communication and control networks," *IEEE Trans. Smart Grid*, vol. 1, no. 1, pp. 28–39, Jun. 2010.

[28] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[29] E. J. Candés, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 11:1–11:37, Jun. 2011.

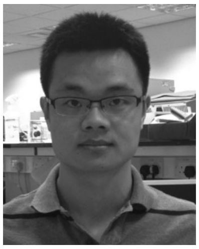
[30] E. Candés and B. Recht, "Exact matrix completion via convex optimization," *Commun. ACM*, vol. 55, no. 6, pp. 111–119, Jun. 2012.

[31] R. Paffenroth, P. du Toit, R. Nong, L. Scharf, A. P. Jayasumana, and V. Bandara, "Space-time signal processing for distributed pattern detection in sensor networks," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 1, pp. 38–49, Feb. 2013.

[32] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[33] R. Christie, "Power Systems test case archive," dept. Elect. Eng., Univ. Washington, Seattle, WA, USA, Apr. 2000 [Online]. Available: http://www.ee.washington.edu/research/pstca/

[34] M. Grant and S. Boyd. (2013, Sep.). *CVX: Matlab Software for Disciplined Convex Programming* [Online]. Available: http://cvxr.com/cvx



Jinping Hao (S'13) received the B.Eng. degree in electronic and information engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2010; the B.Eng. degree (first class Hons.) in electronic and communication engineering from the University of Birmingham, Birmingham, U.K., in 2010; and the M.Sc. degree (with Distinction) in wireless communication and signal processing from the University of Bristol, Bristol, U.K., in 2011. He is currently pursuing the Ph.D. degree in electrical and electronic engineering at the University of Bristol.

His research interests include sparse signal processing, optimization techiniques for wireless communications, and smart grids.

Robert J. Piechocki (M'06) received the M.Sc. degree (with Distinction) in wireless communications systems from the Technical University of Wroclaw, Wroclaw, Poland, in 1997, and the Ph.D. degree in wireless communications systems from the University of Bristol, Bristol, U.K., in 2002.

He is currently a Senior Lecturer with Advanced Wireless Access, and a Member of the Communications Systems and Networks Group, University of Bristol. He has authored over 100 papers in international journals and conferences, and holds 13 patents in these areas. His research interests include statistical signal processing, information and communication theory, wireless networking, body and ad hoc networks, ultra low power communications, and vehicular communications.

Dritan Kaleshi (M'00) received the Dipl.Ing. degree (with Excellence) in electronics from the Polytechnic University of Tirana, Tirana, Albania, in 1991, and the Ph.D. degree in electronic engineering from the University of Bristol, Bristol, U.K., in 2005.

He is currently a 5G Fellow with Digital Catapult, London, U.K., and a Visiting Research Fellow with the University of Bristol, where he was a Senior Lecturer in Communication Networks until 2015. He has authored over 60 papers in the field, edited two international standards, and holds three patents. He represents the U.K. in various international standardization bodies (International Organization for Standardization/International Electrotechnical Commission, European Committee for Standardization/European Committee for Electrotechnical Standardization) in areas related to Internet-of-Things (IoT), home electronic systems, and smart grid. His research interests include future networking architectures and protocols (5G and beyond), large scale loosely coupled distributed systems design, modeling and performance evaluation, and data interoperability for sensor/actuator systems (IoT).

Woon Hau Chin (S'99–M'04–SM'10) received the B.Eng. (first class Hons.) and M.Eng. degrees from the National University of Singapore, Singapore, in 1999 and 2000, respectively, and the Ph.D. degree from Imperial College London, London, U.K., in 2004, all in electrical engineering.

From 2000 to 2008, he was with the Institute for Infocomm Research, Singapore, where he was involved in the standardization of IEEE 802.11n and 3rd Generation Partnership Project Long-Term Evolution. He was also involved in several industrial projects on wireless local area networks and beyond 3G systems. He was also an Adjunct Assistant Professor with the National University of Singapore, from 2005 to 2008. Since 2008, he has been with the Telecommunications Research Laboratory, Toshiba Bristol, Bristol, U.K., where he had been involved in several EU FP7 projects and research on small cell interference mitigation, body area networks, and smart grid communications. He is currently the Research Manager with the Physical Layer Research Group, York, U.K., leading research in fundamental communication technologies.

Dr. Chin has held positions in several conference organizing committees, and is on the technical programme committee (TPC) of various international conferences. He has Co-Founded the International Symposium on Wireless Pervasive Computing (ISWPC) series of conferences, and was the TPC Co-Chair of ISWPC 2008, and a TPC Co-Chair of the Transmission Technologies Track at the IEEE Vehicular Technology Conference in 2008. He was the Founding Chair of the Vehicular Technology Chapter in Singapore, and held that position from 2006 to 2008.

Zhong Fan received the B.S. and M.S. degrees in electronic engineering from Tsinghua University, Beijing, China, in 1992 and 1994, respectively, and the Ph.D. degree in telecommunication networks from Durham University, Durham, U.K., in 1997.

He is a Chief Research Fellow with Toshiba Research Europe, Bristol, U.K. Prior to joining Toshiba, he worked as a Research Fellow with Cambridge University, Cambridge, U.K.; as a Lecturer with Birmingham University, Birmingham, U.K.; and as a Researcher with Marconi Labs Cambridge, Cambridge. He was also awarded a British Telecommunications (BT) Short-Term Fellowship to work at BT Labs, Ipswich, U.K. His research interests include wireless and IP networks, machine-to-machine, and smart grid communications.