On the Silent Perturbation of State Estimation in Smart Grid

Alireza Jolfaei*, *Senior Member, IEEE,* and Krishna Kant†, *Life Fellow, IEEE*

***Abstract*—In this paper, we address the problem of integrity of the state estimation in presence of a novel bad data injection threat. We consider persistent adversaries who inject attack vectors "silently", that is, they drift the results of the state estimation gradually in multiple steps, with each step bypassing the bad data detection step of state estimation. Prior works on the topic have shown how an adversary can bypass the bad data detection by constructing attack vectors as linear combinations of the column vectors of the measurement Jacobian matrix. We show here that the attack surface is much broader than implied by this assumption. We demonstrate our attack strategy using realistic load patterns from NYISO in the IEEE 14-bus system. We also propose a detection method that uses the expected energy of normalized residues and a paired *t*-test, and we show its effectiveness. The generality of the proposed attack and detection strategy implies that it can be used in other cyber-physical systems involving state estimation based on linearized state perturbation.**

***Index Terms*—Cyber-physical systems security, measurement Jacobian matrix, persistent adversary, perturbation bound, silent perturbation, smart grid, state estimation.**

I. Introduction

The emerging smart grid provides capabilities to continuously monitor the grid behavior based on sophisticated measurements, so that any anomalies can be detected quickly and remedial action taken. In particular, most busses in the grid use some metering capability, such as the traditional power flow meters or the newer *Phasor Measurement Units* (PMUs). The PMUs provide a high resolution data usually, 30 or 60 samples/sec, including voltage, phase, and frequency. All of the data is sent to a control center, perhaps a local control center (LCC) for each region, either directly or via intermediate data concentrators. The data is then used for a variety of smart-grid applications ranging from quick handling of emergency issues (for example, relay trip, frequency shift, and deep voltage sags) to the long-term issues such as balancing power flows and detecting slow-developing anomalies. Most of the medium and longer term applications depend on *state estimation*, a process by which the true state of the grid is assessed every few minutes. State estimation takes the raw measurements and from there determines the best estimate of the voltages and phases at various busses. Generally, the number of required bus voltages/phases is much smaller than the number of measurements thereby leading to an overdetermined system. The measurements, however, may have errors (largely due to relative time skew between various measurements), and thus a least-squares based estimation procedure is typically used to obtain the best possible estimate of the state.

Unfortunately, the use of communications exposes the grid to adversarial attacks, which are aimed at disrupting grid operations and potentially causing blackouts. A prime source of such attacks is the lack of integrity in transporting critical messages via the IEC TR 61850-90-2 protocol [1], [2]. Although the protocol provides integrity mechanisms, they are optional for critical communications that require very low latency [3], [4]. The lack of integrity protection leaves the grid data vulnerable to man-in-the-middle attacks. A potential adversary can modify the measurement data from a subset (if not all) of *Remote Terminal Units* (RTUs) and/or PMUs by modifying the bits in transit. Therefore, the *Local Control Center* (LCC) would observe corrupted measurements rather than the real measurements [7]. Such a discrepancy, if significant and undetected, may lead to catastrophic impacts, for example, blackouts in large geographic areas.

The state estimation techniques in use in power systems do detect and remove the *bad* data, or measurements that are substantially different from the expected values [8]. At a first glance, it may seem that these approaches can also defeat the malicious measurements injected by adversaries if those measurements are quite different from the expected values. However, it has been shown that the existing techniques for bad measurement detection can be bypassed if the adversary knows the topology and configuration of the power system. The reason for this failure is that the existing bypassing techniques for bad measurement detection rely on the assumption that the bad measurements make the squares of differences between the observed measurements and their corresponding estimates significant [9]. With the knowledge of the power system topology and configuration, the adversary can generate bad measurements such that the assumption above is violated; therefore, bypassing bad measurements detection [10].

In practice, the exact topology and configuration of the power system is not known precisely so the above-mentioned attacks may not work accurately. The attack mechanism considered in this paper corresponds to an adversary that poses *Advanced Persistent Threat* (APT). The goal of such an adversary is to stay in after having successfully penetrated into the system and take a series of actions, each of which itself small enough that it may not be detected but can collectively pose a real threat to the system. In our context, the adversary continuously and gradually perturbs the smart

This research was supported by NSF grant CPS-1544904.

Alireza Jolfaei is with the Department of Computing, Macquarie University, Sydney, Australia (Email: alireza.jolfaei@mq.edu.au). Krishna Kant is with the Department of Computer and Information Sciences, Temple University, Philadelphia, USA (Email: kkant@temple.edu).

grid state estimation so that the estimated state is ultimately drifted sufficiently from the normal state which forces harmful and/or wasteful actions by the control system. For example, if the estimated voltage appears to indicate a substantial voltage sag, additional generation may be kicked in or certain loads may be isolated to stabilize the voltage. In this paper, we demonstrate how to construct such gradual drifting attacks to potentially harm the system and also show how to detect them. Furthermore, we show that the space of silent attacks is much larger than previously thought. To the best of our knowledge, this is the first work on persistent attacks to drift power system state to undesirable levels.

The rest of the paper is organized as follows. Section II discusses the related work. Section III then explains the background of state estimation and related work. Section IV presents the proposed false data injection attack, and gives an approach to glide the estimated states towards desired values. Section V discusses the detection strategies. Section VI demonstrates the success of these attacks through simulation. Section VII concludes the paper.

II. Related Work

False data injection attacks against the state estimation of power grid have extensively been studied. Studies [13], [14], and [15] showed that false data injection attacks lead to increased system operation costs due to disproportionate power generation and dispatch [13] or energy routing [14], as well as economic loss due to misconduct of electricity markets [15]. A number of research efforts have been carried out to analyze, detect and handle bad data injection attacks, particularly in transmission and distribution centers. In [9], Liu et al. present a framework for a man-in-the-middle attack on the power system state estimation, through which an adversary would replace normal grid reading with malicious data. Liu et al. showed that by using a sufficient number of power meters, the adversary can perturb the estimated state without being detected by the bad data detector employed at the control center.

Following Liu et al.'s work, the link between feasibility of an attack and the observability in the state estimation of power systems was analyzed in [16], [17], and [18]. To assess the grid vulnerability against data attacks, the minimum number of adversary-controlled sensors necessary for an unobservable attack was suggested as the security index of the grid [17]. Protection against false data injection attacks on the state estimation is currently achieved by either securing a number of meter measurements physically or monitoring a number of state variables directly by PMUs. For example, in [16], Bobba et al. showed that a strategically selected set of meter measurements and state variables can help to combat false data injection attacks. When there is no verifiable state variable, it is necessary and sufficient to secure a set of basic measurements to detect attacks. To facilitate the strategic placement of secure PMUs for defense against false data injection attacks, Kim and Poor [19] suggested the use of a number of fast greedy algorithms to select a subset of meter measurements to protect.

In summary, the aforementioned works propose to secure some meter measurements and/or some state variables to make false data injection attacks impractical. They have assumed that some meter measurements can be absolutely protected, that is, the adversary cannot compromise them no matter how powerful. This assumption would be too strong for real applications. Note that even when a meter is protected from adversarial modification, it may still have a bias due to a physical malfunction or an improper parameter setting. Filtering out the measurements from such malfunctioning meters is the main objective of the legacy bad data processing and is still in practice today.

TABLE I: Notations

Notation	Description
$\boldsymbol{x}$	Vector of state variables $[x_1, x_2, \ldots, x_n]^t$ (unknown)
$\hat{\boldsymbol{x}}$	Estimated value of vector $\boldsymbol{x}$
$\boldsymbol{z}$	Original measurement vector $[z_1, z_2, \ldots, z_m]^t$, $m > n$
$\tilde{\boldsymbol{z}}$	Perturbed measurement vector $\boldsymbol{z}$
$\boldsymbol{e}$	Noise vector $[e_1, e_2, \ldots, e_m]^t$
$\boldsymbol{H}$	Measurement Jacobian matrix
$\boldsymbol{H}^{\zeta}$	Pseudo-inverse of $\boldsymbol{H}$
$\boldsymbol{W}$	Inverse of measurement error covariance matrix $cov(\boldsymbol{e})$
$\boldsymbol{a}$	Attack vector $[a_1, a_2, \ldots, a_m]^t$
τ	Bad measurement detection threshold

III. Mathematical Model

This section introduces the background of state estimation, bad data processing, and the adversary model. The notations used in this paper are described in Table I. More generally, we use boldface upper-cases letters for matrices, boldface lower-case letters for vectors, and normal font for scalars. For a matrix $\boldsymbol{X}$, $\boldsymbol{X}^{-1}$ and $\boldsymbol{X}^t$, respectively, denote the inverse and the transpose of $\boldsymbol{X}$. The $\mathbb{E}$ is the expectation operator.

A. Power Grid State Estimation and Bad Data Processing

In this Subsection, we review basic steady-state power network modeling and state-estimation techniques. More complete presentations are given in [22], [23], for example. Estimation of power grid's state (bus voltage magnitudes and phase angles) is important for monitoring the critical operational functions of power grid, such as real-time power flow monitoring, load forecasting, power dispatch, and load frequency control. The state estimator employed in power grid usually deals with a large number of states of the order of 10,000, and gives an estimate of states every few minutes [22].

B. State Estimation

We assume a centralized state estimation at the local control center or LCC, which collects measurements from RTUs and PMUs deployed throughout the power grid. PMUs provide accurate measurements of magnitude and phase angles for both bus voltage and branch current and a GPS based time-stamp. To obtain an awareness of the internal states of the power grid, state estimation is performed, that is, to estimate state variables $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]^t$ based on the PMU measurements $\boldsymbol{z} = [z_1, z_2, \ldots, z_m]^t$, under independent random measurement errors (or "noise") $\boldsymbol{e} = [e_1, e_2, \ldots, e_m]^t$. Here n and m are positive integers, $n < m$, and x_i, $z_j \in \mathbb{R}$ for i $(1 \leq i \leq n)$ and j $(1 \leq j \leq m)$. The random measurement errors are real-numbered values that are assumed

to follow a Gaussian distribution with zero mean and diagonal measurement covariance matrix $\sigma^2\boldsymbol{I}$, that is, $\boldsymbol{e} \sim \mathcal{N}(0, \sigma^2\boldsymbol{I})$.

In general, the AC power flow model [21], measurements $\boldsymbol{z}$ are related to state variables $\boldsymbol{x}$ through a nonlinear model. The computation of bidirectional (active and reactive) power flows involves complex numbers and nonlinear functions. The nonlinearities are commonly handled through a DC power flow model [24], which provides linear approximations. The linear approximation to the nonlinear relationship is expressed as

$$\boldsymbol{z} = \boldsymbol{H} \cdot \boldsymbol{x} + \boldsymbol{e}, \tag{1}$$

where the measurement Jacobian matrix $\boldsymbol{H}$ is an $m \times n$ matrix that relates the change in state variables (for example, voltage magnitude and phases) to the measured values (for example, power flows). $\boldsymbol{H}$ has full column rank when $n < m$.

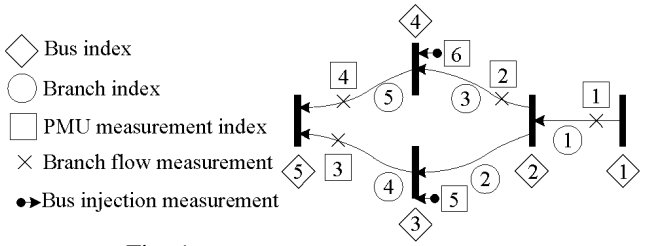

Fig. 1: A simple 5-bus power network.

As a simple example, Figure 1 shows a 5-bus power network. Assume that Bus 1 is the reference bus with $\delta_1 = 0$, and the state variables are $\boldsymbol{x} = [\delta_2, \delta_3, \delta_4, \delta_5]^t$. The meter readings are $\boldsymbol{z} = [F_{12}, F_{24}, F_{35}, F_{45}, P_3, P_4]^t$. Let b_{ij} denote the susceptance of the transmission line (i, j). Then, the DC power flow model yields the following measurement Jacobian matrix:

$$\boldsymbol{H} = \begin{bmatrix} b_{12} & 0 & 0 & 0 \\ -b_{24} & 0 & b_{24} & 0 \\ 0 & -b_{35} & 0 & b_{35} \\ 0 & 0 & -b_{45} & b_{45} \\ b_{23} & -b_{23}-b_{35} & 0 & b_{35} \\ b_{24} & 0 & -b_{24}-b_{45} & b_{45} \end{bmatrix}. \tag{2}$$

The state estimation problem is to find an estimate $\hat{\boldsymbol{x}}$ of state variables $\boldsymbol{x}$, that is, the best fit of the meter measurements $\boldsymbol{z}$ according to Equation 1, which is an over-determined linear system of equations. The measurement residual is $\boldsymbol{r} = \boldsymbol{z} - \hat{\boldsymbol{z}} = \boldsymbol{z} - \boldsymbol{H}\hat{\boldsymbol{x}}$, that is, the difference between the observed measurements $\boldsymbol{z}$ and the estimated measurements $\hat{\boldsymbol{z}}$.

The maximum likelihood estimate of $\boldsymbol{z}$ is the value that minimizes the weighted least-squares performance index [25]. The weighted least squares (WLS) criterion problem for the overdetermined system in Equation 1 is to find an estimate $\hat{\boldsymbol{x}}$ that minimizes the performance index $J(\hat{\boldsymbol{x}})$, defined as

$$J(\hat{\boldsymbol{x}}) = (\boldsymbol{z} - \boldsymbol{H}\hat{\boldsymbol{x}})^t \boldsymbol{W} (\boldsymbol{z} - \boldsymbol{H}\hat{\boldsymbol{x}}) = \sum_{i=1}^{m} \left| \frac{z_i - h_i(x_i)}{\sigma_i} \right|^2, \tag{3}$$

where the weight matrix $\boldsymbol{W}$ is the inverse of the measurement error covariance matrix $cov(\boldsymbol{e})$, that is, a diagonal matrix whose entries are reciprocals of the variances of measurement errors. In other words, $w_{ii} = \sigma_i^{-2}$ for i $(1 \leq i \leq m)$. The $J(\hat{\boldsymbol{x}})$ performance index is a widely used test, which is due to its simplicity and the fact that the test statistic has a χ^2 distribution if the data are good.

Equation 3 indicates that the weights are set as the inverse of the measurement noises. In other words, good quality measurements with a low noise level have larger weights, and vice versa. To find the minimum of the performance index, it is differentiated ($\frac{\partial J(\hat{\boldsymbol{x}})}{\partial \hat{\boldsymbol{x}}} = 0$) to obtain the first-order optimal condition; and the nonlinear WLS problem is solved by making use of iterative Gauss-Newton or Newton-Raphson methods, which gives the estimate of the states as follows:

$$\hat{\boldsymbol{x}} = (\boldsymbol{H}^t \boldsymbol{W} \boldsymbol{H})^{-1} \boldsymbol{H}^t \boldsymbol{W} \boldsymbol{z} = \boldsymbol{H}^{\boldsymbol{\zeta}} \boldsymbol{z}, \tag{4}$$

where $\boldsymbol{H}^{\boldsymbol{\zeta}}$ is the pseudo-inverse of $\boldsymbol{H}$, as $\boldsymbol{H}^{\boldsymbol{\zeta}}\boldsymbol{H} = \boldsymbol{I}$. When the measurement errors are assumed to be normally distributed with zero mean, these criteria lead to the identical optimal state estimator $\boldsymbol{H}^{\boldsymbol{\zeta}}$. Since $rank(\boldsymbol{H}^{\boldsymbol{\zeta}}) = rank(\boldsymbol{H}) = n < m$, at least n meters are needed to derive a unique state estimation.

C. Bad Measurement Detection

Bad measurements may be introduced due to various reasons, such as PMU/PDC failure and malicious attacks. Intuitively, bad measurements (outliers) may move the estimated state variables away from their true values, and therefore, bad measurements detection techniques have been developed to protect state estimation [26]. The technique iteratively checks for bad data, removes it, and tries state estimation again.

Since there is usually inconsistency among the good and the bad measurements, a common approach for detecting the presence of bad data is to examine the L_2-norm of the measurement residual against a prescribed threshold. More precisely, the existence of bad measurements is assumed if $\|\boldsymbol{z} - \boldsymbol{H}\hat{\boldsymbol{x}}\| > \tau$; otherwise, $\boldsymbol{z}$ is considered as normal measurements. τ is a constant (detection threshold) to be determined and the selection of it is a key issue. Assume that measurement errors are mutually independent and follow a normal distribution with zero mean. It can be mathematically shown that $J(\hat{\boldsymbol{x}})$ follows a χ^2_{m-n} distribution, that is, a chi-square distribution with $m-n$ degrees of freedom, because the state estimation is constrained by n independent equations. According to [27], τ is determined by a hypothesis test with a significance level (false alarm probability) α, that is, $Pr\{J(\hat{\boldsymbol{x}}) \geq \chi^2_{m-n,1-\alpha}\} = \alpha$. This means that $\|\boldsymbol{z} - \boldsymbol{H}\hat{\boldsymbol{x}}\| \geq \tau$ detects bad measurements with a false alarm probability α, where $\tau = \chi^2_{m-n,1-\alpha}$.

D. Bad Data Identification and Removal

If the bad data detector declares that the measurement data are good, that is, if the L_2-norm of the residual r is less than τ, the algorithm returns the state estimate $\hat{\boldsymbol{x}}$ and terminates. However, if the bad data detector declares that the data are bad, that is, $\|\boldsymbol{r}\|_2^2 \geq \tau^2$, bad data identification is triggered to identify and remove the bad data entry from the measurement vector.

A widely used method for identifying a bad data entry is the largest normalized residue [22], in which each $r_i^{(k)}$ is divided by its standard deviation under the hypothesis that $\boldsymbol{z}$ contains

no bad data. Therefore, each normalized residue approximately follows the standard normal distribution $\mathcal{N}(0, \boldsymbol{BW})$, where

$$\boldsymbol{B} = \boldsymbol{I} - \boldsymbol{H}\left(\boldsymbol{H}^t\boldsymbol{W}^{-1}\boldsymbol{H}\right)^{-1}\boldsymbol{H}\boldsymbol{W}^{-1}. \quad (5)$$

Accordingly, the normalized residue is calculated as $\tilde{\boldsymbol{r}} = \boldsymbol{\Omega r}$, where $\boldsymbol{\Omega}$ is a diagonal matrix whose diagonal entries are $\boldsymbol{\Omega_{ii}} = \frac{1}{\sqrt{(\boldsymbol{BW})_{ii}}}$. Once the normalized residue $\tilde{\boldsymbol{r}}$ is computed, the meter with the largest $|\tilde{r}|$ is identified as a bad meter. Accordingly, the bad data identification omits the rows of $\boldsymbol{z}$ and h that correspond to bad meters, and gives updated measurement vectors and measurement function for the next iteration.

Algorithm 1 provides the pseudo-code of the overall procedure of the iterative state estimation, bad data detection, and bad data identification.

Algorithm 1 Iterative state estimation

```
procedure STATE-ESTIMATION(z, h, W)
    while true do
        H ← ∂h(x)/∂x |_{x=0};
        x̂ ← (H^t W H)^{-1} H^t W z;
        if ‖z − H x̂‖ > χ²_{m−n,1−α} then
            Break;
        else
            (z, h) ← bad-data-removal(z, h);
        end if
    end while
    Return x̂;
end procedure
```

IV. Data Falsification Attacks

A. Problem Formulation

We consider a power system that consists of $n+1$ buses, and there are m meters that provide power flow measurements z to the state estimator via shared communication channels. It is assumed that $\boldsymbol{H}$ is static, and will not be affected by changes such as line outages, tap changes, or other topology or parameter changes. The adversary is assumed to be persistent and have the ability to change a subset of measurements needed for an attack. The change could occur by corrupting the measurement device at the substation, interfering with communication between the substation and control center, or by installation of malware at the control center. The adversary is not assumed to have the ability to observe conditions across the entire system.

The adversarial modification is mathematically modeled by

$$\overline{\boldsymbol{z}} = \boldsymbol{z} + \boldsymbol{a}, \quad (6)$$

where $\overline{\boldsymbol{z}}$ is the perturbed measurement vector, and $\boldsymbol{a}$ is a non-zero attack vector. The non-zero components of the attack vector $\boldsymbol{a}$ correspond to the compromised meters. Under the attack, the meter readings are changed by the adversary from their uncompromised values $\boldsymbol{z}$ to their post attack values $\boldsymbol{z}+\boldsymbol{a}$.

The goal of the persistent adversary is to continuously and gradually perturb states in consecutive rounds of state estimation, such that each perturbation round remains undetected and the final estimated state is drifted to a certain range that can fool the energy management system to make decisions which can harm the power system. To drift a specific state to a desired value, the adversary gradually manipulates the nonlinear optimization problem by making use of consecutive attack vectors; through which the update equation (4) is perturbed in each estimation round.

B. Bypassing Bad Data Detection

Previous works [9], [11], [12], [16], [17], and [18] have shown that malicious measurements $\overline{\boldsymbol{z}} = \boldsymbol{z}+\boldsymbol{a}$ can pass the bad measurement detection if $\boldsymbol{a}$ is a linear combination of the column vectors of $\boldsymbol{H}$, that is, $\boldsymbol{a} = \boldsymbol{Hc}$ where $\boldsymbol{c} \in \mathbb{R}^{n+1}$ is an arbitrary non-zero vector (see Liu et al.'s theorem [9] in Appendix A).

A common assumption in the literature of false data injection attacks is that the adversary has complete knowledge about the power grid topology and transmission-line admittances [28], [29]. In fact, the information about the Jacobian measurement matrix $\boldsymbol{H}$ is implicitly assumed available to the adversary in order to construct the false data injection attack vectors. Although this information is known to the operators, it is well protected and unlikely to be known to adversary [30]. Thus, an attack based on complete knowledge of $\boldsymbol{H}$ is unrealistic.

However, it is possible to construct $\boldsymbol{H}$ approximately (as discussed in section IV-E in more details). *The key contribution of this paper is to exploit the approximate $\boldsymbol{H}$ and the threshold τ along with a repeated silent attack to still be able to significantly perturb the state without being detected.*

In the following, we prove that the adversary can bypass the bad data detection using the bad data detection threshold. Knowing the detection threshold is a reasonable assumption, since it can be easily derived from the the number of system measurements and system states, which are publicly available.

Lemma 1: Let $\boldsymbol{a}$ and $\boldsymbol{b}$ denote $n\times 1$ vectors, and let τ represent a threshold. If the relations $\|\boldsymbol{a}\| \leq \tau$ and $\|\boldsymbol{a}+\boldsymbol{b}\| \leq \tau$ hold, then $\|\boldsymbol{b}\| \leq 2\tau$.

Proof: By setting $\|\boldsymbol{b}\| = \|-\boldsymbol{a}+\boldsymbol{a}+\boldsymbol{b}\|$ and using triangular inequality, the relation $\|\boldsymbol{b}\| \leq \|\boldsymbol{a}\|+\|\boldsymbol{a}+\boldsymbol{b}\|$ is obtained that concludes $\|\boldsymbol{b}\| \leq 2\tau$ following the conditions of the lemma. ■

We can now prove the following theorem using Lemma 1.

Theorem 1: Suppose that the original measurements $\boldsymbol{z}$ pass the bad measurement detection. An attack vector $\boldsymbol{a}$ is stealthy if $\|\boldsymbol{a}\| \leq 2\tau$, where τ is the detection threshold.

Proof: Since $\boldsymbol{z}$ can pass the detection, we have $\|\boldsymbol{z} - \boldsymbol{H}\hat{\boldsymbol{x}}\| < \tau$. Assume that the vector of estimated state variables obtained from $\overline{\boldsymbol{z}}$, is represented as $\overline{\hat{\boldsymbol{x}}} = \hat{\boldsymbol{x}}+\boldsymbol{c}$, where $c \in \mathbb{R}^n$ is an arbitrary non-zero vector, and the corrupted measurement is denoted by $\overline{\boldsymbol{z}} = \boldsymbol{z}+\boldsymbol{a}$. The aim of the adversary is to drift the values of state variables to specific values by making use of bad data injection, such that bad data is not detected through the detection process. In other words, the following relation

should be satisfied:

$$\|\overline{\boldsymbol{z}}-\boldsymbol{H}\overline{\hat{\boldsymbol{x}}}\| = \|\boldsymbol{z}+\boldsymbol{a}-\boldsymbol{H}(\hat{\boldsymbol{x}}+\boldsymbol{c})\| = \|\boldsymbol{z}-\boldsymbol{H}\hat{\boldsymbol{x}}+\boldsymbol{a}-\boldsymbol{H}\boldsymbol{c}\| \leq \tau \tag{7}$$

Following Lemma 1, $\|\boldsymbol{z}-\boldsymbol{H}\hat{\boldsymbol{x}}\| \leq \tau$ and $\|\boldsymbol{z}-\boldsymbol{H}\hat{\boldsymbol{x}}+\boldsymbol{a}-\boldsymbol{H}\boldsymbol{c}\| \leq \tau$ implies $\|\boldsymbol{a}-\boldsymbol{H}\boldsymbol{c}\| \leq 2\tau$.

Furthermore, using the triangle inequality,

$$\|\boldsymbol{a}\|-\|\boldsymbol{H}\boldsymbol{c}\| \leq \|\boldsymbol{a}-\boldsymbol{H}\boldsymbol{c}\| \leq 2\tau. \tag{8}$$

From the inequality above, it follows that $\|\boldsymbol{a}\| \leq 2\tau+\|\boldsymbol{H}\boldsymbol{c}\|$. Since $\|\boldsymbol{H}\boldsymbol{c}\|$ is non-negative, choosing $\|\boldsymbol{a}\| \leq 2\tau$ would preserve the inequality in all cases.

■

This theorem means to keep the perturbation from being discarded as bad data, and we need to ensure that $\|\boldsymbol{a}\| \leq 2\tau$. However, the condition is not sufficient, and some perturbations of any amount could still be flagged as bad data and discarded. However, since our goal is to persist and continue the attack, the key question is whether a reasonably small number of tries can yield a large perturbation. We show experimentally in Section VI that this is indeed the case.

C. Attack Vector Size

Realistically, the adversary may not be able to hack into a large number of PMUs in a substation, sniff and falsify all measurements. In addition, in the following, we show that the adversary would not be able to drift state variables by spoofing a large number of measurements at once if the essential condition of bypassing bad data detection is to be retained. Indeed, the adversary needs to modify only a limited number of measurements to launch a feasible attack without being detected. Assume that $I_m = \{i_1, i_2, \ldots, i_k\}$ is the set of indices corresponding to measurements, and that the adversary can modify only k number of z_{i_j}, where $i_j \in I_m$.

As explained in Subsection III-C, the sum of squared measurement residuals quantifies how well the measurement data fits a normal distribution. This is used for bad data detection, because it is expected that bad data would not fit the model well. Unlike bad data, malicious data is constructed to fit a normal distribution, decreasing the impact on measurement residues; and therefore, reducing the alarm risk. For a successful attack that remains undetected, the size of attack vector should be limited (See Appendix B). In other words, the adversary would not be able to drift state variables by spoofing a large number of measurements at once and still remain undetected.

D. Perturbation Bound

To drift a specific state to a desired value (that is, $\boldsymbol{x^{k+1}} = \boldsymbol{x^k}+\boldsymbol{\Delta x^k}$ where k indicates the iteration count), the adversary gradually manipulates the nonlinear optimization problem by making use of consecutive attack vectors; through which the update equation (4) is perturbed in each estimation round. The adversary can also make the system unobservable. In other words, following Equation (4), $\overline{\hat{\boldsymbol{x}}} = \boldsymbol{H^{\zeta}}(\boldsymbol{z}+\boldsymbol{a})$. In the following, we give an upper bound for the difference between the expected estimated states from the true state values.

Theorem 2: $\|\hat{\boldsymbol{x}}-\overline{\hat{\boldsymbol{x}}}\| \leq \frac{\|\boldsymbol{a}\|}{\xi_{min}(\boldsymbol{H})}$, where $\xi_{min}(\boldsymbol{H})$ denotes the smallest singular value of the matrix $\boldsymbol{H}$.

The proof is contained in Appendix C.

Corollary 1: $\|\hat{\boldsymbol{x}}-\overline{\hat{\boldsymbol{x}}}\| \leq \|\boldsymbol{a}\|\cdot\|\boldsymbol{H}^{-1}\|$.

Proof: Following Theorem 2, since $\|\boldsymbol{H}^{-1}\|^{-1} = \xi_{min}(\boldsymbol{H})$, the inequality $\|\hat{\boldsymbol{x}}-\overline{\hat{\boldsymbol{x}}}\| \leq \|\boldsymbol{a}\|\cdot\|\boldsymbol{H}^{-1}\|$ holds true. ■

Corollary 2: Our perturbation bound does not depend on the number of state variables, and it would not reduce as the network size increases.

Proof: This is an immediate result of Theorem 2. ■

Theorem 3: Our perturbation bound $\|\hat{\boldsymbol{x}}-\overline{\hat{\boldsymbol{x}}}\| \leq \frac{\|\boldsymbol{a}\|}{\xi_{min}(\boldsymbol{H})}$ is much larger than the bound suggested by Zhao et al. in [31].

The proof to Theorem 3 is given in Appendix D.

E. Data Falsification Attack Procedure

To implement a successful attack, the adversary needs to first retrieve the entries of the measurement Jacobian matrix within the maximum error margin of a small percentage, and then construct a non-detectable attack vector (bad data injection vector) that can bypass the bad data detection and lead the current state to a desired value. To construct the measurement Jacobian matrix, the adversary needs to know the network topology and the admittance matrix of the grid.

Generally, power utilities do not disclose details of their own grids, but there are some electricity authorities, for example, the electricity authority of the New South Wales in Australia, that keep their data open to the public [32], [33]. In the USA, an adversary can obtain the configuration of the North American power grid from the POWERmap mapping system, which contains information about every power plant, major substation, and 115-765 kV power line of the North American power grid [41]. A determined adversary can also identify the topology of the power grid using the information gathered by satellite pictures combined with existing transmission system maps [34]. As a proof of concept, Rivera et al. [34] manually computed the topology of the German transmission system in [35].

In addition to network topology, the adversary needs to determine the real values of network parameters (for example, admittance, conductance, and susceptance) using the characteristics of power lines. This information is normally provided by the cable manufacturers, and thus, it is easy to compute the approximate admittance matrix using basic power flow equations. However, the accurate values of line parameters, for example, the real length of the power line, are only known by the utility companies, and therefore, the estimated admittance values would be an approximate. In [36], Sivanagaraju et al. showed an estimation method for the entries of the admittance matrix with an error bound less than one percent. Using Sivanagaraju et al.'s method, we assume that the adversary would be able to reconstruct the entries of $\boldsymbol{H}$ within the maximum error margin of a small percentage ε.

Given a perturbed measurement Jacobian matrix $\widetilde{\boldsymbol{H}} = \boldsymbol{H}+\boldsymbol{\delta H}$, the adversary needs to determine $\boldsymbol{H}$ using $\|\boldsymbol{\delta H}\|_F \leq \varepsilon$,

where ε is a constant threshold, and $\|\boldsymbol{\delta H}\|_F$ is the Frobenius norm of $\boldsymbol{\delta H}$, that is, $\sqrt{\sum_{i=1}^{n}\xi_i^2} = \|\boldsymbol{H}\|_F$. By Weyl's inequality [37],

$$|\xi_{\min}(\widetilde{\boldsymbol{H}}) - \xi_{\min}(\boldsymbol{H})| \leq \|\widetilde{\boldsymbol{H}} - \boldsymbol{H}\|_2 \leq \|\boldsymbol{\delta H}\|_2 \leq \|\boldsymbol{\delta H}\|_F \leq \varepsilon. \quad (9)$$

Accordingly, $|\xi_{\min}(\boldsymbol{H})| \geq |\xi_{\min}(\widetilde{\boldsymbol{H}})| - \varepsilon$.

Remark 1: The singular value decomposition of an $m \times n$ matrix has computation complexity $O(\min\{mn^2, m^2n\})$ [38].

To construct the bad data injection vector, the adversary should first pass the bad data detection and then lead the current state to a desired value. To bypass the bad data detection, the adversary needs to find an attack vector $\boldsymbol{a}$ inside a closed n-ball that satisfies the relation $\|\boldsymbol{a}\| \leq \|\boldsymbol{Hc}\| + 2\chi^2_{m-n,1-\alpha}$.

The literature has considered only a special case for constructing the attack vector, that is, $\|\boldsymbol{a}\| = \|\boldsymbol{Hc}\|$. However, the space of the attack is much larger. Since the complete measurement Jacobian matrix is unknown, by Theorem 1, the adversary can use the relation $\sum_{i=1}^{n} a_i^2 \leq 2\chi^2_{m-n,1-\alpha}$ to bypass the bad data detection method. The chi-squared distribution is defined in terms of normally distributed random variables. If $Z_1, \ldots, Z_n$ are independent identically distributed standard normal variables, then $\sum_{i=1}^{n} Z_i^2 \sim \chi^2_{m-n,1-\alpha}$. In other words, to draw from a chi-squared distribution with $m-n$ degrees of freedom, the adversary can use n values drawn from standard normal. More precisely, for i $(1 \leq i \leq n)$, a_i is sampled from a normal distribution $X_1, \ldots, X_n$ with mean 0 and standard deviation $\boldsymbol{BW}$, where $\boldsymbol{B} = \boldsymbol{I} - \boldsymbol{H}\left(\boldsymbol{H}^t\boldsymbol{W}^{-1}\boldsymbol{H}\right)^{-1}\boldsymbol{HW}^{-1}$, such that the objective function $\sum_{i=1}^{n} a_i^2 \leq 2\chi^2_{m-n,1-\alpha}$ is satisfied. If the objective function value exceeds the threshold, then the bad data will be suspected; Otherwise, the measurement set is assumed to be free of bad data.

Figure 2 shows the histogram of $\sum_{i=1}^{14} a_i^2$ for randomly sampled a_i from $\mathcal{N}(0, \boldsymbol{BW})$, where $1 \leq i \leq 14$ (in an IEEE 14-bus test case). To ensure that the sample values fall within the $1-\alpha$ confidence interval of χ^2_{m-n} ($100\alpha\%$ middle values are selected), one can make multiple draws, and then discard the draws that fall beyond the $100\alpha\%$ interval. This gives a sample attack vector.

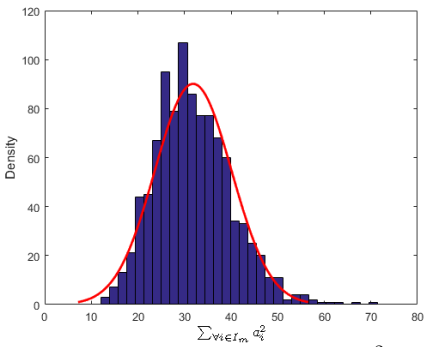

Fig. 2: Histogram of $\sum_{\forall i \in I_m} a_i^2$.

To gradually drift the state to a desired value without being noticed, the adversary iteratively injects multiple attack vectors, which are constructed by $\boldsymbol{k_i} \cdot \boldsymbol{a}$, where $1 \leq i \leq N$, $\boldsymbol{k_i} = [k_{i_1}, k_{i_2}, \ldots, k_{i_n}]^t$, and $\|\boldsymbol{k_i} \cdot \boldsymbol{a}\| \leq \|\boldsymbol{Hc}\|$. From the theoretical (Theorem 5) and practical points of view, it may not be feasible to sniff and falsify all measurement data. Therefore, the aim of the adversary is to minimize the number of meters to be compromised, that is, to find a K-sparse vector $\boldsymbol{k}$ with the minimum number of non-zero elements that satisfies $\|\boldsymbol{k} \cdot \boldsymbol{a}\| \leq 2\chi^2_{m-n,1-\alpha}$, such that the chosen elements in $\boldsymbol{k}$ have the specific values. More precisely, the adversary picks a threshold τ less than $2\chi^2_{m-n,1-\alpha}$, and computes a vector $\boldsymbol{k}$ satisfying $\|\boldsymbol{k} \cdot \boldsymbol{a}\| = \tau$ such that $\boldsymbol{k}$ has at most K non-zero elements.

A closer look at this problem reveals that it is the minimum weight solution for the linear equations problem, which is an NP-Complete problem. However, there are a number of efficient heuristic solutions, such as the matching pursuit algorithm and the gradient pursuit algorithm. The adversary can use these algorithms to find a near optimal attack vector. In our simulation, we employed the matching pursuit algorithm, which has an exponential rate of convergence for computing the sparse signal representations.

The aim of the adversary is to fool the system operator to make unnecessary and costly actions, such as generator rescheduling and load shedding. To make the system take further corrective actions, the adversary needs to continuously corrupt the estimated power flow obtained from the distribution/transmission lines. Utility equipment can only tolerate the fluctuations of voltage in a small period of time. If the voltage remains unstable for a long time, the equipment has a high probability to get damaged. The acceptable range of voltage amplitudes varies depending on the regulation compliance of different regions in different countries. In the Tasmanian region of Australia, for example, the limits of voltage variations are about $\pm 5\%$ in normal operating conditions. Although some utilities allow variations somewhat larger than $\pm 5\%$, this is a good benchmark as noted in [39].

Usually, when the system is experiencing instability, the operator would firstly reschedule the generator to solve the problem, and if this does not appear promising, load shedding will be enforced as an emergency action. To this end, the adversary needs to inject false vectors multiple times, hoping that the state is gradually drifted to a harmful range. For a successful attack, the adversary not only needs to manipulate the measurement data before the rescheduling, but also needs to manipulate the measurement data after rescheduling. However, perturbed states are bounded by the inequality $-\|\boldsymbol{a}\| \cdot \|\boldsymbol{H}^{-1}\| + \hat{\boldsymbol{x}} \leq \|\widehat{\hat{\boldsymbol{x}}}\| \leq \|\boldsymbol{a}\| \cdot \|\boldsymbol{H}^{-1}\| + \hat{\boldsymbol{x}}$, which indicates injected false vectors may correct and alleviate the impact of previous changes. Nonetheless, our simulation results in Section VI show that if are attack vectors are carefully constructed, they can drift the estimated states to an unsafe range such that it can force the LCC to repeatedly order unnecessary generator rescheduling or load shedding.

F. Discussion

Although the configuration of power systems is usually kept secret, the adversary may still be able to determine it by piecing together information from multiple sources as discussed

in Subsection IV-E. In addition to gathering information that would enable estimation of matrix $\boldsymbol{H}$, an adversary also needs to tamper measurement devices or corrupt their measurements, either at the source or during the transmission, before they are used in state estimation. This is usually made difficult by the surveillance cameras and integrity protection algorithms. Nonetheless, attacks are possible and cannot be overlooked.

In the view of above, from an adversarial perspective, it would be desirable to perturb only a few PMUs even in a large-scale power network, and the perturbation bound without detection is as large as possible. Indeed, compared to the literature [9], [11], [12], our attack surface is much larger, and previous attacks turn out to be a special case of our attack. This fact is stated as the following important observation:

Observation 1: In previous attacks, the subset of measurement devices need to be compromised and the false data injection attack vector implemented by the adversary needs to satisfy $\boldsymbol{a} = \boldsymbol{Hc}$. In contrast, our attack only requires $\sum_{\forall i \in I_m} a_i^2 \leq 2\tau$, which is a much looser condition.

V. Detection Strategy

Power systems are quasi-static systems whose state changes constantly but slowly. Indeed, the measurements taken from the grid vary slowly and their variations follow a normal distribution with an approximately zero mean. Figure 3 shows the measurement variations in electric power grid regions in New York State, in October 2012, with no false data injection attacks [40]. This figure confirms that the measurement variations are small and close to zero.

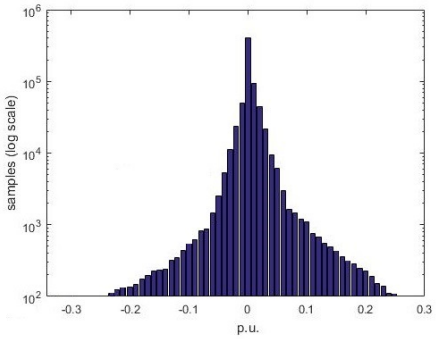

Fig. 3: Histogram of measurement variations in electric power grid regions in New York State, in October 2012 with no false data attacks.

The quasi static status of the power systems suggests the use of expected energy of normalized residues to detect silent perturbation attacks over time. We assume that we can identify a period of time, when there is no persistent attack, so that we can estimate the state under "normal conditions". In this case, the normalized residue of a true measurement vector $\boldsymbol{z}$ is given by

$$\tilde{\boldsymbol{r}} = \boldsymbol{\Omega r} = \boldsymbol{\Omega W z}, \tag{10}$$

where $\boldsymbol{\Omega}$ is a diagonal matrix whose diagonal entries are $\boldsymbol{\Omega_{ii}} = \frac{1}{\sqrt{(\boldsymbol{BW})_{ii}}}$ and $\boldsymbol{B} = \boldsymbol{I} - \boldsymbol{H}\left(\boldsymbol{H}^t\boldsymbol{W}^{-1}\boldsymbol{H}\right)^{-1}\boldsymbol{H}\boldsymbol{W}^{-1}$. Due to the normalization, each entry $\tilde{\boldsymbol{r}}_{\boldsymbol{i}}$ is distributed as $\mathcal{N}(0,1)$. If an attack vector $\boldsymbol{a}$ is added, the resulting normalized residue is

$$\tilde{\boldsymbol{r}} = \boldsymbol{\Omega W}(\boldsymbol{z}+\boldsymbol{a}) = \boldsymbol{\Omega W z} + \boldsymbol{\Omega W a}. \tag{11}$$

Normalized residues $\tilde{\boldsymbol{r}}$ are distributed as $\mathcal{N}(\boldsymbol{\Omega W a}, 1)$ and the expected energy of the normalized residues is

$$\mathbb{E}\left[\sum_{i\in I_m} \tilde{\boldsymbol{r}}_{\boldsymbol{i}}^{\,2}\right] = \sum_{i\in I_m} \mathbb{E}\left[\tilde{\boldsymbol{r}}_{\boldsymbol{i}}^{\,2}\right] = \sum_{i\in I_m} (\boldsymbol{\Omega W a})_i^2 + C, \tag{12}$$

where the constant C corresponds to the cumulative energy of residuals without attack.

Since the silent perturbations discussed here involve repeated attempts by the adversary to drift the estimated system state further and further from the true state, successful detection also must operate over multiple state estimation cycles. That is, we attempt to detect "anomaly" in state over multiple cycles to make a determination of whether the grid is under attack. Note that an anomaly detected only occasionally may be result of physical events or perturbations, which may or may not be worth further investigation.

For anomaly detection, we conduct statistical tests on the difference between the residual vector from the current measurements, $\tilde{\boldsymbol{r}}$, which may be potentially anomalous and differ from the normal measurement $\boldsymbol{r}$. To this end, we design a paired *t*-test for $\boldsymbol{r} - \tilde{\boldsymbol{r}}$. The test is specified by a confidence level α. Let μ_0 and S denote, respectively, the sample mean and sample standard deviation of the collected data. Then, the null hypothesis is that both mean and variance of the two residual vectors are identical. For this, we use the *t* as

$$t_{test} = \frac{\mu_0}{S/\sqrt{m}}, \tag{13}$$

where m is the number of measurement samples. Let t_α denote the t-value for a given confidence level α. Then, if $t_{test} < t_\alpha$, we accept the null hypothesis; otherwise, we reject the null hypothesis and detect the presence of an anomaly with confidence level α.

This process is repeated over K successive state estimation intervals, where K is a parameter. If the anomaly is detected in $k < K$ instances, we declare a persistent silent attack. Here k is also a parameter, but can be chosen as a high percentage of K (for example, 75%). The suitable choice of K depends on how frequently anomalies are seen without attacks. In most cases, such anomalies should be rare, in which case, K can be rather small (for example, 5). If an anomaly is detected, the reference residuals are kept the same; otherwise, the last block of K measurements becomes the next basis for "normal" measurements.

VI. Experimental Results

In this section, we validate the proposed false data injection attack through experiments using the IEEE 14-bus test case (Figure 4), which represents a portion of the American Electric Power System in the Midwestern US. We extracted the configuration of the IEEE 14-bus test case (particularly $\boldsymbol{H}$) from [42], and simulated attacks against state estimation using the DC power flow model. The state variables are voltage amplitudes and angles of all the buses, and the meter measurements are real power injections of all buses and real power flows of all branches. The information of state variables and measurements within various IEEE test systems is given in Table II. All experiments were simulated in MATLAB R2015b

TABLE II: Number of state variables and measurements

IEEE Bus System	14-bus
Number of voltage measurements	2
Number of real power inject measurements	7
Number of reactive power inject measurements	7
Number of real power flow measurements	8
Number of reactive power flow measurements	8
Number of total measurements	32
Number of state variables	27

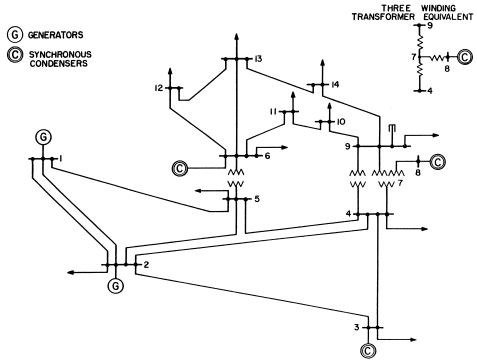

Fig. 4: IEEE 14-bus test system.

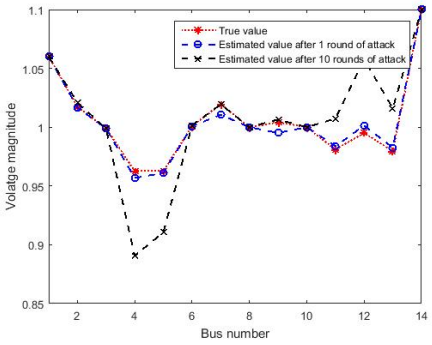

Fig. 5: Voltage magnitude comparison result.

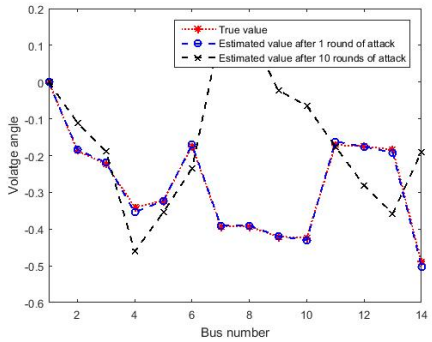

Fig. 6: Voltage angle comparison result.

on a PC with Intel Core 2 2.4 GHz processor and 8 GB of installed memory.

To simulate the behavior of compromised nodes, we added an offset (that is, chosen by the attack vector) to the measurements. The bad data offsets are added iteratively in consecutive rounds of state estimation. We ran this experiment 10^3 times. Figures 5 and 6 show a sample run of our experiments. The figures demonstrate the deviations in estimated states after several runs (1 and 10) of our attack. It is shown that an adversary is able to both perturb the bad data detection method, and also drift some of the states to a value more than $\pm 5\%$ of the nominal values. Such perturbations would compel the system operator to make unnecessary and costly actions, such as generator rescheduling and load shedding.

VII. Conclusion and Future Work

This paper proposes a false data injection attack against nonlinear state estimation in a power grid. We showed that the adversary can take the advantage of the configuration of a power system to launch such attacks to bypass the currently used bad data detection technique.

Our proposed attack does not require the knowledge of the exact Jacobian measurement matrix and can inject errors into a DC power flow system without being detected. We showed that the adversary can systematically and efficiently construct attack vectors, which not only change the results of state estimation, but also modify the results in a predicted way. The novelty of our work is the demonstration that the space of attack is much larger than assumed by previously known results, and the lack of dependence of the perturbation bound on the number of state variables. This is a significant result since our perturbation bound does not get reduced with the increase in network size. We performed simulation on IEEE test systems to demonstrate the success of the proposed attack. We proposed the use of expected energy of normalized residues and a paired *t*-test to detect silent perturbation attacks. Our results in this paper indicate that security protection of the power grid must be revisited when there are potentially malicious attacks.

Our proposed perturbation mechanism can be applied to any other similar dynamical system where the errors are small enough to allow a linearization of the state estimation. In the future, we will examine the attacks that can systematically make positive/negative fluctuations rather than linear modifications in the voltage magnitudes and phase angles. This is an interesting topic for research since impulsive and/or oscillatory transients are the most damaging type of power disturbance, as compared to unidirectional changes. We also plan to investigate new state estimation methods that use only a subset of all the available measurements. If the set of used measurements are selected strategically, such that the adversary does not know which measurements were used, it should be possible to mitigate the effect of the attacks. This is a form of moving target defense strategy.

APPENDICES

Appendix A

Theorem 4 (Liu et al. [9]): Suppose the original measurements $\boldsymbol{z}$ can pass the bad measurement detection. The malicious measurements $\overline{\boldsymbol{z}} = \boldsymbol{z} + \boldsymbol{a}$ can pass the bad measurement detection if $\boldsymbol{a}$ is a linear combination of the column vectors of $\boldsymbol{H}$, that is, $\boldsymbol{a} = \boldsymbol{H}\boldsymbol{c}$ where $\boldsymbol{c} \in \mathbb{R}^n$ is an arbitrary non-zero vector.

Proof: Follows trivially from the fact that multiplying $\boldsymbol{c}$ by $\boldsymbol{H}$ would transform it into the precise change in $\boldsymbol{z}$. Since $\boldsymbol{z}$ can pass the detection, we have $\|\boldsymbol{z} - \boldsymbol{H}\hat{\boldsymbol{x}}\| < \tau$, where τ is the detection threshold. $\overline{\hat{\boldsymbol{x}}}$, the vector of estimated state variables obtained from $\overline{\boldsymbol{z}}$, can be represented as $\overline{\hat{\boldsymbol{x}}} = \hat{\boldsymbol{x}} + \boldsymbol{c}$. If $\boldsymbol{a} = \boldsymbol{H}\boldsymbol{c}$, that is, $\boldsymbol{a}$ is a linear combination of the column

vectors $h_1, \ldots, h_n$ of $\boldsymbol{H}$, then the resulting L_2-norm of the measurement residual is

$$\|\overline{\boldsymbol{z}} - \boldsymbol{H}\overline{\hat{\boldsymbol{x}}}\| = \|\boldsymbol{z}+\boldsymbol{a}-\boldsymbol{H}(\hat{\boldsymbol{x}}+\boldsymbol{c})\| = \|\boldsymbol{z}-\boldsymbol{H}\hat{\boldsymbol{x}}+\boldsymbol{a}-\boldsymbol{H}\boldsymbol{c}\| = \|\boldsymbol{z}-\boldsymbol{H}\hat{\boldsymbol{x}}\|. \quad (14)$$

■

Appendix B

We use the following lemma to derive an estimate on the size of the attack vector (Theorem 5).

Lemma 2: If $B_i \geq 0$ are independent and identically distributed $B_i \sim \mathcal{G}$, then

$$Pr\left(\sum_{i=1}^{n} B_i \leq \tau\right) \leq \mathcal{G}^n(\tau). \quad (15)$$

Proof: If the sum of non-negative terms B_i is less than τ, then all of the terms must be less than τ. The independent and identically distributed property of the variables is then invoked to obtain the result. ■

Theorem 5: Let random variables A_i denote a_i for i $(1 \leq i \leq n)$. Let $A_i \sim \mathcal{F}$ be independent and identically distributed with any distribution other than δ_0, a point-mass at zero. Then, for each $\tau \geq 0$,

$$\lim_{n\to\infty} Pr\left(\sum_{i=1}^{n} A_i^2 \leq \tau\right) = 0. \quad (16)$$

Proof: Let $B_i = A_i^2$, and denote the distribution of B_i by function $\mathcal{G}$. If $\mathcal{F}$ is of unbounded support, then so is $\mathcal{G}$. Invoking Lemma 2, if $\mathcal{G}$ is of unbounded support, then $\mathcal{G}(\tau) < 1$ for all $\tau \geq 0$. This implies that $\mathcal{G}^n(\tau) \to 0$.

Now, we extend the problem to the case of bounded support. Suppose there exists $C > 0$ such that $\mathcal{G}(\tau) = 1$ for all $\tau \geq C$, and $\mathcal{G}(\tau) < 1$ for all $\tau < C$. The case of $\tau < C$ is already handled by the argument above. For $\tau > C$, there is a fixed $N = 1 + \frac{\tau}{C}$ such that

$$\mathcal{G}_N(\tau) = Pr\left(\sum_{i=1}^{N} B_i \leq \tau\right) < 1. \quad (17)$$

The case of bounded support is further reduced to the previous case, since $(\mathcal{G}_N(\tau))^m \to 0$, as $m \to \infty$ by considering sums over blocks of size N. ■

Appendix C: Proof of Theorem 2

Proof: Given $\overline{\hat{\boldsymbol{x}}} = \boldsymbol{H}^{\boldsymbol{\zeta}}(\boldsymbol{z}+\boldsymbol{a}) = \hat{\boldsymbol{x}} + \boldsymbol{H}^{\boldsymbol{\zeta}}\boldsymbol{a}$, we have $\|\hat{\boldsymbol{x}} - \overline{\hat{\boldsymbol{x}}}\| = \|\boldsymbol{H}^{\boldsymbol{\zeta}}\boldsymbol{a}\|$. Let the singular value decomposition be given by $\boldsymbol{H} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\boldsymbol{t}}$, where $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthogonal matrices, and $\boldsymbol{\Sigma}$ is a diagonal matrix, respectively. This gives $\boldsymbol{H}^{\boldsymbol{\zeta}} = (\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\boldsymbol{t}})^{-1} = \boldsymbol{V}\boldsymbol{\Sigma}^{-\mathbf{1}}\boldsymbol{U}^{\boldsymbol{t}}$. Thus,

$$\|\hat{\boldsymbol{x}} - \overline{\hat{\boldsymbol{x}}}\| = \|\boldsymbol{H}^{\boldsymbol{\zeta}}\boldsymbol{a}\| = \|\boldsymbol{V}\boldsymbol{\Sigma}^{-\mathbf{1}}\boldsymbol{U}^{\boldsymbol{t}}\boldsymbol{a}\| = \|\sum_{j=1}^{r} \frac{1}{\xi_j}\boldsymbol{v_j}\boldsymbol{u_j^t}\boldsymbol{a}\|. \quad (18)$$

Since $\boldsymbol{v_j}$'s are orthonormal, the Pythagorean theorem gives

$$\left\|\sum_{j=1}^{r} \frac{1}{\xi_j}\boldsymbol{v_j}\boldsymbol{u_j^t}\boldsymbol{a}\right\| = \sum_{j=1}^{r} \frac{|\boldsymbol{u_j^t}\boldsymbol{a}|^2}{\xi_j} \leq \frac{\|\boldsymbol{U}\|\cdot\|\boldsymbol{a}\|}{\xi_{\min}(\boldsymbol{H})} \leq \frac{\|\boldsymbol{a}\|}{\xi_{\min}(\boldsymbol{H})}, \quad (19)$$

where $\xi_{min}(\boldsymbol{H})$ denotes the smallest singular value of the matrix $\boldsymbol{H}$. This gives $\|\hat{\boldsymbol{x}} - \overline{\hat{\boldsymbol{x}}}\| \leq \frac{\|\boldsymbol{a}\|}{\xi_{min}(\boldsymbol{H})}$, which proves the theorem. ■

Appendix D: Proof of Theorem 3

Proof: From Theorem 2, $\|\hat{\boldsymbol{x}}\| - \|\overline{\hat{\boldsymbol{x}}}\| \leq \|\hat{\boldsymbol{x}} - \overline{\hat{\boldsymbol{x}}}\| \leq \frac{\|\boldsymbol{a}\|}{\xi_{min}(\boldsymbol{H})}$. Therefore, $\|\hat{\boldsymbol{x}}\| - \frac{\|\boldsymbol{a}\|}{\xi_{min}(\boldsymbol{H})} \leq \|\overline{\hat{\boldsymbol{x}}}\|$. This shows that small perturbations in the measurements can lead to a large drift in the state value if the smallest singular value of $\boldsymbol{H}$ is small. From $\|\hat{\boldsymbol{x}} - \overline{\hat{\boldsymbol{x}}}\| \leq \frac{\|\boldsymbol{a}\|}{\xi_{min}(\boldsymbol{H})}$, when no attack occurs, $\|\boldsymbol{a}\| = 0$ and therefore $\|\hat{\boldsymbol{x}} - \overline{\hat{\boldsymbol{x}}}\| = 0$. Following Theorem 1, the upper bound of the attack vector is $\|\boldsymbol{a}\| = 2\tau$, and therefore

$$\|\hat{\boldsymbol{x}} - \overline{\hat{\boldsymbol{x}}}\| \leq \frac{2\tau}{\xi_{min}(\boldsymbol{H})}, \quad (20)$$

that is, the upper bound of the state perturbation is based on the detection threshold τ and the smallest singular value of the matrix $\boldsymbol{H}$.

In [31], Zhao et al. have shown that

$$\|\hat{\boldsymbol{x}} - \overline{\hat{\boldsymbol{x}}}\| < \frac{\tau - \|\boldsymbol{r}\| - \|\boldsymbol{e}\|}{\|\boldsymbol{H}\|_F}, \quad (21)$$

where τ is the detection threshold, $\|\boldsymbol{r}\|$ is the measurement residual vector, $\boldsymbol{e} = \boldsymbol{H}\hat{\boldsymbol{x}} - \boldsymbol{H}\overline{\hat{\boldsymbol{x}}}$ is the measurement forecasting error, and $\|\boldsymbol{H}\|_F$ is the Frobenius norm of $\|\boldsymbol{H}\|$. Since $\xi_{min}^2 \leq \frac{1}{n}\sum_{i=1}^{n} \xi_i^2$, in the worst case scenario $\xi_{min} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \xi_i^2} = \sqrt{\frac{1}{n}}\|\boldsymbol{H}\|_F$. Accordingly,

$$\|\hat{\boldsymbol{x}} - \overline{\hat{\boldsymbol{x}}}\| \leq \frac{\tau - \|\boldsymbol{r}\| - \|\boldsymbol{e}\|}{\|\boldsymbol{H}\|_F} \leq \frac{\tau - \|\boldsymbol{r}\| - \|\boldsymbol{e}\|}{\sqrt{n}\xi_{min}(\boldsymbol{H})} \leq \frac{2\tau}{\xi_{min}(\boldsymbol{H})}. \quad (22)$$

In the ideal case where the state of the grid does not change, $\|\boldsymbol{e}\| = 0$, and also when no bad data is detected $\|\boldsymbol{r}\| < \tau$. Then our perturbation bound is much larger than Zhao et al.'s bound (Equation 22) and that in reality the state can be perturbed much larger than what was shown by Zhao et al. In addition, Zhao et al.'s bound cannot perturb a large-scale system because when $n \to \infty$ their perturbation bound tends to zero. However, our perturbation bound (Equation 20) is independent of the number of states. Furthermore, Equation 22 does not explicitly express the relationship between the attack vector entries and state perturbation, and therefore cannot be used to construct false data injection vectors. In contrast, our bound could be used to construct successful attack vectors. ■

REFERENCES

[1] IEC 61850-90-5: 2012. Communication networks and systems for power utility automation–Part 90-5: Use of IEC 61850 to transmit synchrophasor information according to IEEE C37.118, 2012.

[2] Cisco corporation, "Substation Automation Local Area Network and Security Cisco Validated Design," Cisco Book Chapter. 10.01.2020. https://www.cisco.com/c/en/us/td/docs/solutions/Verticals/Utilities/SA/2-3-2/CU-2-3-2-DIG/CU-2-3-2-DIG.html

[3] A. Jolfaei and K. Kant, "A lightweight integrity protection scheme for fast communications in smart grid," *14th International Conference on Security and Cryptography*, Madrid, Spain, pp. 31–42, 2017.

[4] A. Jolfaei and K. Kant, "Data Security in Multiparty Edge Computing Environments," *Proceedings of the GOMACTech Conference, Artificial Intelligence & Cyber Security: Challenges and Opportunities for the Government*, Albuquerque, NM, USA, pp. 17–22, 2019.

[5] M. Adamiak, M.J. Schiefen, G. Schauerman, and B. Cable, "Design of a priority-based load shed scheme and operation tests," *IEEE Transactions on Industry Applications*, vol. 50, no. 1, pp. 182–187, 2014.

[6] P. Kundu and A.K. Pradhan, "Synchrophasor-assisted zone 3 operation," *IEEE Transactions on Power Delivery*, vol. 29, no. 2, pp. 660–667, 2014.

[7] A. Jolfaei and K. Kant, "A lightweight integrity protection scheme for low latency smart grid applications," *Computers & Security*, vol. 86, pp. 471–483, 2019.

[8] Y.F. Huang, S. Werner, J. Huang, N. Kashyap, and V. Gupta, "State estimation in electric power grids: Meeting new challenges presented by the requirements of the future grid," *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 33–43, 2012.

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

[10] R. Tan et al., "Modeling and Mitigating Impact of False Data Injection Attacks on Automatic Generation Control," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 7, pp. 1609–1624, 2017.

[11] M.A, Rahman and H. Mohsenian-Rad, "False data injection attacks with incomplete information against smart power grids," *IEEE Global Communications Conference*, pp. 3153–315, 2012.

[12] A. Anwar and A.N. Mahmood, "Stealthy and blind false injection attacks on SCADA EMS in the presence of gross errors," *IEEE Power and Energy Society General Meeting*, pp. 1–5, Boston, MA, USA, 2016.

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

[14] J. Lin, W. Yu, X. Yang, G. Xu, and W. Zhao, "On false data injection attacks against distributed energy routing in smart grid," *ACM/IEEE International Conference on Cyber-Physical Systems*, 2012.

[15] L. Xie, Y. Mo, and B. Sinopoli, "Integrity data attacks in power market operations," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 659–666, 2011.

[16] 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," *First Workshop on Secure Control Systems (SCS)*, Stockholm, Sweden, pp. 1–9, 2010.

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

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

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

[20] R.D. Zimmerman, C.E. Murillo-Sánchez, and R.J. Thomas, "MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Transactions on power systems*, vol. 26, no. 1, pp. 12–19, 2011.

[21] S.H. Low, "Convex relaxation of optimal power flow: a tutorial," *IREP Symposium on Bulk Power System Dynamics and Control-IX Optimization, Security and Control of the Emerging Power Grid*, pp. 1–15, Rethymno, Greece, Aug. 2013.

[22] M. Rostami and S. Lotfifard, "Distributed dynamic state estimation of power systems," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 8, pp. 3395–3404, 2018.

[23] M. Ahmad, "Power System State Estimation," Artech House, 2013.

[24] S. Bolognani and S. Zampieri, "On the existence and linear approximation of the power flow solution in power distribution networks," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 163–172, 2016.

[25] S. Tushar, A.K. Pandey, A. Srivastava, P. Markham, and M. Patel, "Online Estimation of Steady-State Load Models Considering Data Anomalies," *IEEE Transactions on Industry Applications*, vol. 54, no. 1, pp. 712–721, 2018.

[26] O. Vuković, "Cyber-security in Smart Grid Communication and Control," Doctoral Thesis, School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm, Sweden, 2014.

[27] A.J. Wood and B.F. Wollenberg, "Power generation, operation, and control," John Wiley & Sons, 2012.

[28] A. Giani, E. Bitar, M. Garcia, M. McQueen, P. Khargonekar, and K. Poolla, "Smart grid data integrity attacks: characterizations and countermeasures π," *IEEE International Conference on Smart Grid Communications*, pp. 232–237, 2011.

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

[30] Y. Mo, T.-J. Kim, K. Brancik, D. Dickinson, H. Lee, A. Perrig, and B. Sinopoli, "Cyber-Physical Security of a Smart Grid Infrastructure," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 195–209, 2012.

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

[32] http://sydney.edu.au/engineering/electrical/courses/power/NSW_network.xml

[33] https://www.records.nsw.gov.au/series/5407

[34] J. Rivera, J., Leimhofer, and H.A. Jacobsen, "OpenGridMap: towards automatic power grid simulation model generation from crowd sourced data," *Computer Science-Research and Development*, pp. 1-11, 2016.

[35] http://vmjacobsen39.informatik.tu-muenchen.de/

[36] G. Sivanagaraju, S. Chakrabarti, and S.C. Srivastava, "Uncertainty in transmission line parameters: Estimation and impact on line current differential protection," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 6, pp. 1496–1504, 2014.

[37] R. Bhatia,"Perturbation bounds for matrix eigenvalues," Society for Industrial and Applied Mathematics, 2007.

[38] M. Holmes, A. Gray, and C. Isbell, "Fast SVD for large-scale matrices," *Workshop on Efficient Machine Learning at NIPS*, vol. 58, pp. 249–252, 2007.

[39] F.G. Montoya, R. Banos, C. Gil, A. Espin, A. Alcayde, and J. Gomez, "Minimization of voltage deviation and power losses in power networks using Pareto optimization methods," *Engineering Applications of Artificial Intelligence*, vol. 23, no. 5, 695–703, 2010.

[40] G. Chaojun, "Modelling and analysis of cyber-security and reliability of energy delivery for resilient smart grid systems," Doctoral Thesis, Department of Electrical and Computer Engineering, National University of Singapore, Singapore, 2015.

[41] P. Kinney, P. Crucitti, R. Albert, and V. Latora, "Modeling cascading failures in the North American power grid," *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 46, no. 1, pp. 101–107, 2005.

[42] S.K.M. Kodsi and C.A. Canizares, "Modeling and simulation of IEEE 14-Bus system with facts controllers", University of Waterloo, Electrical & Computer Engineering Department, Technical Report #2003-3, 2003.

Alireza Jolfaei received the Ph.D. degree in Applied Cryptography from Griffith University, Gold Coast, Australia. He is a Lecturer (Assistant Professor in North America) and a Program Leader of Cyber Security at Macquarie University, Sydney, Australia. Prior to this appointment, he worked as an Assistant Professor at Federation University Australia and Temple University in Philadelphia, USA. His current research areas include Cyber and Cyber-Physical Systems Security. He received the prestigious IEEE Australian council award for his research paper published in the IEEE Transactions on Information Forensics and Security. He is a Senior Member of the IEEE and an ACM Distinguished Speaker on the topic of Cyber-Physical Systems Security.

Krishna Kant is currently a professor in the Computer and Information Science Department at Temple University in Philadelphia, PA where he directs the IUCRC center on Intelligent Storage. Earlier he was a research professor in the Center for Secure Information Systems at George Mason University. From 2008-2013 he served as a program director at NSF where he managed the computer systems research program and was instrumental in the development and running of NSF-wide sustainability initiative named science, engineering and education for sustainability. He received his Ph.D. degree in Mathematical Sciences from University of Texas at Dallas in 1981. His research interests span a wide range including energy efficiency, robustness, and security in cyber and cyber-physical systems. He is a Fellow of IEEE.