Limiting False Data Attacks on Power System State Estimation

Oliver Kosut, Liyan Jia, Robert J. Thomas, and Lang Tong
School of Electrical and Computer Engineering
Cornell University, Ithaca, NY 14853
Email: {oek2,lj92,rjt1,lt35}@cornell.edu

***Abstract*— Malicious attacks against power system state estimation are considered. It has been recently observed that if an adversary is able to manipulate the measurements taken at several meters in a power system, it can sometimes change the state estimate at the control center in a way that will never be detected by classical bad data detectors. However, in cases when the adversary is not able to perform this attack, it was not clear what attacks might look like. An easily computable heuristic is developed to find bad adversarial attacks in all cases. This heuristic recovers the undetectable attacks, but it will also find the most damaging attack in all cases. In addition, a Bayesian formulation of the bad data problem is introduced, which captures the prior information that a control center has about the likely state of the power system. This formulation softens the impact of undetectable attacks. Finally, a new L_∞ norm detector is introduced, and it is demonstrated that it outperforms more standard L_2 norm based detectors by taking advantage of the inherent sparsity of the false data injection.**

***Index Terms*—Power system state estimation, power system security, false data attack.**

I. Introduction

A power system is composed of many interconnected generators, transmission lines, transformers and loads. To maintain reliable performance of such a system requires that operators have up to date and accurate knowledge about the state of the grid. As such, numerous meters are deployed through the network to measure bus voltage magnitudes, real and reactive power injections, and recently bus voltage and current angles. These measurements are brought together at control centers and used as the basis of state estimation, from which an estimate of the complete state of the power grid is produced.

Since the beginning of the development of state estimation [1], it has been necessary to deal with bad data. Traditionally, bad data were assumed to be caused by random errors resulting from a fault in a meter and/or its attendant communication system. These errors are modeled by a change of variance in Gaussian noise, which leads to an energy (L_2) detector (e.g. [2]–[6]). Recently, Liu, Ning, and Reiter studied the problem that several meters are seized by an adversary that is able to corrupt the measurements from those meters that are received by the control center [7]. This differs from previous investigations of the problem in that the false data at various meters can be simultaneously crafted by the adversary to defeat the state estimator, as opposed to independent errors caused by random faults. It is observed in [7] that there exist cooperative and malicious attacks on meters that all known bad data techniques will fail to detect. The authors of [7] gave a method to adjust measurements at just a few meters in the grid in such a way that bad data detector will fail to perceive the corruption of the data. In fact, this observation can be made even stronger: in a non-Bayesian framework, if an adversary has the ability to adjust the measurements from enough meters, then no algorithm at the control center will ever be able to detect that an adjustment has been made. This can be viewed as a fundamental limit on the ability of the classical formulation of state estimation to handle cooperative attacks.

Power state estimation is generally performed every few minutes, and change from one to the next is usually gradual unless a contingency has occurred that causes an abrupt change in system state. Therefore, in this paper, we take the viewpoint that the control center can use historical data to maintain and track its belief state of the system. It is therefore appropriate to exploit the knowledge of the belief state in a Bayesian formulation to detect and "correct" statistically unlikely measurements. The Bayesian framework has the advantage that there is no hard limit on the number of meters controlled by the adversary before state estimation becomes impossible. Instead, when the number of meters controlled is enough to execute the attack in [7], the estimation error will jump not to infinity, but merely to a quantity on the order of the prior variance on the state.

The highly damaging attack outlined in [7] exists only if specific sets of meters are simultaneously compromised by cooperating adversaries. It is demonstrated in [7] that if a group of adversaries randomly chooses meters to compromise, they must be capable of controlling a significant fraction of the network before they are likely to control one of these dangerous sets of meters. If they are not capable of performing this attack, [7] makes no statement how much damage the adversaries might be able to do. We develop a detectability heuristic to find the attacks to which bad data detection will be the most vulnerable given a particular set of meters controlled by the adversary. When the adversaries control a particularly damaging set of meters, the heuristic recovers the result of [7] that state estimation will result in large estimation errors; when they do not, it offers a more refined analysis that provides a metric by which we can determine which attacks might be

This work is supported in part by the National Science Foundation under and CCF-0728872.

978-1-4244-7417-2/10/$26.00 ©2010 IEEE

worse than others. Using the heuristic, it is possible to find the worst attack by calculating the singular vector associated with the smallest singular value of a particular matrix, an easily computable process.

We also study the design of the false data detector itself. Because an adversary will only be able to change measurement values at a few meters, by definition the change in measurement vector received at the control center from the true measurements taken at meters will be a sparse vector. Traditional bad data detectors are based on the L_2 norm, and therefore not well suited to detecting sparse vectors. We propose a test based on the L_∞ norm, which more accurately detects the presence of an injected sparse vector. We demonstrate that it performs better than the L_2 norm detector. Moreover, we give numerical evidence that the perfor mance of both the classical L_2 detector as well as our L_∞ detector is well approximated by our detectability heuristic.

The rest of the paper is organized as follows. Section II discusses the limit imposed on classical state estimation shown in [7], and formally presents our formulation of the bad data detection problem. Section III introduces our detectability heuristic, and explains why we believe it to work well. Section IV proposes our L_∞ Detector and presents simulation results and comparisons with the L_2 detector on the IEEE 14-bus test system. It also presents numerical evidence demonstrating that our detectability heuristic is roughly accurate. Finally, we conclude and discuss future directions in Section V.

II. Problem Formulation

Consider a DC power flow state estimation problem. This is a linearization of the AC power flow problem. The goal is to estimate the power system state variable $\mathbf{x} \in \mathbb{R}^n$ based on measurements $\mathbf{z} \in \mathbb{R}^m$. Assuming no adversary, $\mathbf{x}$ and $\mathbf{z}$ are related according to

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

where $\mathbf{H}$ is an $m \times n$ matrix, and $\mathbf{e}$ is measurement noise. We assume $\mathbf{e}$ is Gaussian with zero mean and covariance matrix Σ_e. If there is an adversary in the network injecting false data, then it is able to adjust the values of $\mathbf{z}$ that are associated with the meters to which it has access. That is,

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{e} + \mathbf{a} \tag{2}$$

where $\mathbf{a} \in \mathbb{R}^m$ represents the change in measurement values by the adversary. Observe that $\mathbf{a}$ may be nonzero only in those entries for which the adversary controls the associated meter. We assume that the adversary may seize up to k meters, but it may choose whichever meters it likes. Therefore it may choose $\mathbf{a}$ to be an arbitrary k-sparse vector.

The main observation of [7] was the following. Suppose there exists a nonzero k-sparse $\mathbf{a}$ for which $\mathbf{a} = \mathbf{H}\mathbf{c}$ for some $\mathbf{c}$. For many state estimation problems, $\mathbf{H}$ is sparse, so vectors $\mathbf{a}$ satisfying this property are not uncommon. Consider two possible state vectors $\mathbf{x}_1$ and $\mathbf{x}_2$, where $\mathbf{x}_2 = \mathbf{x}_1 + \mathbf{c}$. If $\mathbf{x}_1$ were the true state vector, and the adversary injects $\mathbf{a}$, then the measurement vector received at the control center will be

$$\mathbf{z}_1 = \mathbf{H}\mathbf{x}_1 + \mathbf{e} + \mathbf{H}\mathbf{c} = \mathbf{H}(\mathbf{x}_1 + \mathbf{c}) + \mathbf{e}. \tag{3}$$

Now suppose the true state vector is $\mathbf{x}_2$, and the adversary is not present (i.e. $\mathbf{a} = 0$). Then the measurement vector is

$$\mathbf{z}_2 = \mathbf{H}\mathbf{x}_2 + \mathbf{e} = \mathbf{H}(\mathbf{x}_1 + \mathbf{c}) + \mathbf{e}. \tag{4}$$

Observe that $\mathbf{z}_1 = \mathbf{z}_2$. Therefore, no detection algorithm will ever be able to tell that the adversary was present in the first case, because the second case could always be possible. Furthermore, the adversary can scale $\mathbf{a}$ to be arbitrarily large, and therefore move the control center's state estimate as far as it likes in the direction of $\mathbf{c}$.

Note that this analysis was decidedly non-Bayesian. That is, if there were some prior information on $\mathbf{x}$, then one of $\mathbf{x}_1$ and $\mathbf{x}_2$ may have been preferred, and the ambiguity could be resolved.

We therefore propose the following Bayesian problem formulation. Assume that $\mathbf{x}$ is jointly Gaussian with zero mean and covariance matrix Σ_x. (The zero mean assumption is for simplicity. The problem does not change if $\mathbf{x}$ has non-zero mean, but one can imagine that $\mathbf{x}$ represents the difference between the true state and the canonical voltage levels.) The control center receives $\mathbf{z}$, given by (2), where the adversary chooses $\mathbf{a}$ constrained to be k-sparse. The control center has a detector $\delta(\mathbf{z})$ which it uses to decide whether any false data may be present. We are interested only in the first step of the process: detecting whether false data is present. Potentially, if a control center detects false data, it should have an algorithm to find it, remove it, and produce a better estimate. But for now, we focus only on the design of δ.

The detector δ attempts to distinguish the following two hypotheses:

$$\begin{aligned} &H_0 : \mathbf{a} = 0 \\ &H_1 : \mathbb{E}(\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 | \mathbf{a}) - \mathbb{E}(\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 | \mathbf{a} = 0) \geq C \\ &\quad\;\; \mathbf{a} \text{ is } k\text{-sparse} \end{aligned} \tag{5}$$

where the estimate $\hat{\mathbf{x}}(\mathbf{z})$ is the minimum mean square error with no adversary present

$$\hat{\mathbf{x}}(\mathbf{z}) = \mathbb{E}(\mathbf{x}|\mathbf{z}, \mathbf{a} = 0). \tag{6}$$

That is, the control center is interested in detecting whether the adversary has injected a sufficiently large so that the resulting increase in mean square error is at least C. Thus, if it declares $\delta = 0$, it can guarantee, with a certain probability, that the estimate $\hat{\mathbf{x}}$ is within some error of the true state $\mathbf{x}$. We may write the MMSE estimate $\hat{\mathbf{x}}(\mathbf{z})$ as

$$\hat{\mathbf{x}}(\mathbf{z}) = \mathbf{K}\mathbf{z} \tag{7}$$

where

$$\mathbf{K} = \Sigma_x \mathbf{H}^T (\mathbf{H}\Sigma_x\mathbf{H}^T + \Sigma_e)^{-1}. \tag{8}$$

Given a particular injected vector $\mathbf{a}$, the mean square error can be written

$$\begin{aligned} &\mathbb{E}\|\mathbf{x} - \mathbf{K}(\mathbf{H}\mathbf{z} + \mathbf{e} + \mathbf{a})\|^2 \\ &\qquad = \mathrm{Tr}\big[(I - \mathbf{K}\mathbf{H})\Sigma_x(I - \mathbf{K}\mathbf{H})^T \\ &\qquad\qquad + \mathbf{K}\Sigma_e\mathbf{K}^T + \mathbf{K}\mathbf{a}\mathbf{a}^T\mathbf{K}^T\big]. \end{aligned} \tag{9}$$

Note that the only term in (9) dependent on $\mathbf{a}$ is the last one, meaning the H_1 condition can be written as simply

$$\|\mathbf{Ka}\|_2^2 \geq C. \tag{10}$$

For the hypothesis testing problem in (5), the false alarm probability is given by

$$\alpha = \Pr(\delta(\mathbf{z}) = 1 | \mathbf{a} = 0) \tag{11}$$

and the worst-case detection probability, determined by optimizing over the adversary's choice of $\mathbf{a}$, is given by the solution to the optimization problem

$$\begin{array}{ll} \text{minimize} & \Pr(\delta(\mathbf{z}) = 1|\mathbf{a}) \\ \text{subject to} & \|\mathbf{Ka}\|_2^2 \geq C \\ & \mathbf{a} \text{ is } k\text{-sparse.} \end{array} \tag{12}$$

Consider now this problem formulation from the adversary's perspective. It knows δ, but of course it is not constrained to keep the mean square error increase below C. Any $\mathbf{a}$ that it chooses will result in some probability of detection, and some increase in mean square error. Certainly, it can cause the most damage by choosing a point on the optimal trade-off curve between these two quantities. If it chooses a certain level of risk, i.e. a certain detection probability β, it can optimize the mean square error as follows:

$$\begin{array}{ll} \text{maximize} & \|\mathbf{Ka}\|_2^2 \\ \text{subject to} & \Pr(\delta(\mathbf{z}) = 1|\mathbf{a}) \leq \beta \\ & \mathbf{a} \text{ is } k\text{-sparse.} \end{array} \tag{13}$$

Observe that the optimizations in (12) and (13) are equivalent in that they will trace out the same trade-off curve between detection probability and mean square error, but their interpretations are slightly different: (12) represents the control center designing a detector by choosing a tolerable level of estimation error and preparing for the worst adversarial action, and (13) represents the adversary finding the best attack given a certain risk of detection.

III. Detectability Heuristic

For many detectors, solving the optimization problem (12) is difficult. In this section, we propose a heuristic for $\Pr(\delta(\mathbf{z}) = 1|\mathbf{a})$, which will allow us to rewrite (12) in a way that is easier to solve. The heuristic approximates the degree to which it is possible for the control center to detect the presence of a certain adversarial vector $\mathbf{a}$. Even though the probability in (12) depends on the detector δ, our heuristic will not. We claim that for most detectors, $\Pr(\delta(\mathbf{z}) = 1|\mathbf{a})$ will be roughly increasing in our heuristic; therefore optimizing one is as good as optimizing the other. Section IV provides some numerical evidence for this claim, but first we argue intuitively why it may be the case. In addition, we show that the most damaging attacks found using the heuristic form a generalization of the attacks found in [7].

Given a measurement vector $\mathbf{z}$, the main tool by which we may determine the implausibility of this vector as having resulted from the measurement process (1) with no adversary present is the measurement residual $\mathbf{r} = \mathbf{z} - \mathbf{H}\hat{\mathbf{x}}$ where again $\hat{\mathbf{x}} = \mathbf{Kz}$ is the MMSE estimate of $\mathbf{x}$ given $\mathbf{z}$. The measurement residual can be rewritten simply as $\mathbf{r} = \mathbf{Gz}$ where $\mathbf{G} = I - \mathbf{HK}$.

Consider a measurement z_i. The ith element of $\mathbf{H}\hat{\mathbf{x}}$ represents the control center's best estimate of the noiseless version of z_i, taking into account data from all measurements, in particular those other than z_i. Therefore if an adversary manipulates the value of z_i, we can expect that redundant measurements elsewhere in the measurement vector will hold $\mathbf{H}\hat{\mathbf{x}}$ relatively fixed, so r_i will change. If the adversary seizes the meters associated with the measurements in the set $S \subset \{1, \ldots, n\}$, then it can inject a vector $\mathbf{a}$ with sparsity pattern S. We can expect that the largest values of the measurement residual will be $\mathbf{r}_S$. In particular, the extent by which $\mathbf{r}_S$ will change because of the injection of $\mathbf{a}$ is $\mathbf{G}_{S,S}\mathbf{a}_S$, where $\mathbf{G}_{S,S}$ is the $|S| \times |S|$ matrix taken from the rows and columns of $\mathbf{G}$ corresponding to elements of S. Our proposed heuristic is given by

$$\Gamma(\mathbf{a}) = \|\mathbf{G}_{S,S}\mathbf{a}_S\|_2. \tag{14}$$

If this is small, then by injecting the vector $\mathbf{a}$, the adversary is able to move the measurement residual by only a small amount in the elements corresponding to the measurements that were manipulated. Therefore, we can expect the presence of the adversary to be detected with low probability.

For a given sparsity pattern S, the most damaging $\mathbf{a}$ will be the one minimizing $\Gamma(\mathbf{a})$. In particular, if the adversary wishes to add C to the mean square error of the control center's estimate, the worst $\mathbf{a}$ for it to use, according to the heuristic, would be the solution to the optimization problem

$$\begin{array}{ll} \text{minimize} & \|\mathbf{G}_{S,S}\mathbf{a}_S\|_2 \\ \text{subject to} & \|K\mathbf{a}\|_2^2 \geq C \\ & a_i = 0 \text{ for } i \notin S. \end{array} \tag{15}$$

This can be easily solved using singular value decomposition.

We now argue that attacks found via the optimization in (15) form a generalization of the attacks outlined in [7]. In particular, if the optimum value of (15) is 0, this corresponds exactly to the attack in [7]. Consider the case that there is no prior distribution, i.e. $\Sigma_x = \infty$. In this case $\mathbf{K} = \mathbf{H}^+$, the pseudo-inverse of H. We claim that the optimum value of (15) is 0—i.e. the matrix $\mathbf{G}_{S,S}$ is singular—exactly when there exists a nonzero vector $\mathbf{c}$ such that $\mathbf{a} = \mathbf{Hc}$ has sparsity pattern S.

The set of vectors $\mathbf{a} = \mathbf{Hc}$ form a linear space which may be equivalently written $\mathbf{Fa} = 0$ for some matrix $\mathbf{F}$. We construct one such matrix $\mathbf{F}$ as follows. Let t be the rank of $\mathbf{H}$, and let

$$\mathbf{H} = \mathbf{USV}^T \tag{16}$$

be the singular value decomposition of $\mathbf{H}$. Recall that only the first t diagonal elements of $\mathbf{S}$ are nonzero. Hence the linear space of linear combinations of the columns of $\mathbf{H}$ is equivalent to the space of linear combinations of the first t columns of $\mathbf{U}$. We denote the matrix made up of these columns of $\mathbf{U}$ by $\mathbf{U}_{1:t}$. Since $\mathbf{U}$ is unitary, this is precisely the set of vectors $\mathbf{a}$ for which $\mathbf{U}_{t+1:m}^T\mathbf{a} = 0$. Hence, we set $\mathbf{F} = \mathbf{U}_{t+1:m}^T$. In particular, if the sparsity pattern of $\mathbf{a}$ is S, then there exists a nonzero $\mathbf{a}$ such that $\mathbf{Fa} = 0$ exactly when the matrix made up of the columns of $\mathbf{F}$ from S is rank deficient. We write this matrix as $\mathbf{U}_{S,t+1:m}^T$.

We now consider the condition that the minimal heuristic is zero, and show that it is equivalent to the condition that $\mathbf{U}_{S,t+1:m}^T$ is rank deficient. We may write

$$\mathbf{G} = I - \mathbf{HK} = \mathbf{U}(I - \mathbf{SS}^+)\mathbf{U}^T. \tag{17}$$

Observe that $I - \mathbf{SS}^+$ is a diagonal $m \times m$ matrix whose first t diagonal elements are zeros and the rest are ones. Therefore

$$\mathbf{G}_{S,S} = \mathbf{U}_{S,t+1:m}\mathbf{U}_{S,t+1:m}^T. \tag{18}$$

Hence, $\mathbf{G}_{S,S}$ is singular exactly when $\mathbf{U}_{S,t+1:m}$ is rank deficient.

We can now interpret the result in [7] as stating that when the $\mathbf{a}$ optimizing (15) satisfies $\Gamma(\mathbf{a}) = 0$, part of the state becomes unobservable, so state estimation is impossible. Our claim concerning the heuristic $\Gamma(\mathbf{a})$ is a generalization of the statement: the smaller $\Gamma(\mathbf{a})$ is, the more difficult state estimation becomes.

IV. Proposed Detector

Recall that a classical bad data detector tests the 2-norm of the measurement residual. That is, it is of the form

$$\delta_2(\mathbf{z}) = \begin{cases} 1 & \text{if } \|\mathbf{z} - \mathbf{H}\hat{\mathbf{x}}\|_2 > \tau \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

There are two flaws with this detector. First, it does not take into account the inherent sparsity of the adversary's injected vector $\mathbf{a}$. A deviation in a non-sparse direction is more likely to be caused by the measurement noise than false data injection, but the 2-norm test considers all directions equivalent. A better test would be one where moving the measurement vector along a sparse direction more quickly crosses the detector's threshold than doing so in a non-sparse direction.

Second, δ_2 does not take advantage of the prior distribution on $\mathbf{x}$ in order to defeat the damaging attacks from [7]. That is, if $\mathbf{a} = \mathbf{Hc}$, then the measurement residual does not change much, but if $\mathbf{a}$ gets too large, elements of the measurement vector $\mathbf{z}$ itself may become unrealistically large. A better detector would be aware of this possibility. We next propose a simple detector that improves on both these problems.

A. The L_∞ Detector

Our detector, referred to as the L_∞ Detector, is of the following form:

$$\delta_\infty(\mathbf{z}) = \begin{cases} 1 & \text{if } \|\mathbf{z} - \mathbf{H}\hat{\mathbf{x}}\|_\infty > \tau_1 \text{ or } \|\mathbf{z}/\sigma_{\mathbf{z}}\|_\infty > \tau_2 \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

where $\mathbf{z}/\sigma_{\mathbf{z}}$ is the vector composed of each measurement normalized by its standard deviation. Note that this detector has two thresholds, τ_1 and τ_2. We will usually fix τ_2 at some level, then vary τ_1 to achieve the desired false alarm probability.

This detector has the desired properties: if one changes only a few elements of $\mathbf{z}$, then redundancy in other measurements should hold the corresponding elements of $\mathbf{H}\hat{\mathbf{x}}$ relatively fixed. Therefore, the measurement residual will grow along a sparse direction, so it will cross the threshold sooner than if it were to grow along a non-sparse direction. Moreover, the test on the weighted measurement vector itself constrains attacks of the $\mathbf{a} = \mathbf{Hc}$ type such that the size of each element z_i cannot exceed its standard deviation by more than a constant factor. Again, we use the infinity norm, because we expect that a small number of z_i will be affected.

B. Performance

Evaluating and comparing the performance of detectors for this problem is complicated by the fact that one must find the worst-case $\mathbf{a}$. We can only be certain of the performance of some detector if we can find the $\mathbf{a}$ that solves the optimization problem given by (12). The heuristic described in Section III allows us to solve this optimization problem approximately, but solving it exactly is not easy, for several reasons. First, the k-sparsity condition of $\mathbf{a}$ is difficult to deal with, as it is highly non-convex, so one may need to check all $\binom{n}{k}$ sparsity patterns to find the best $\mathbf{a}$. Moreover, even for a fixed sparsity pattern, the optimization problem is not convex, and thus difficult to solve. We now rewrite the optimization problem in (12) to show why this is so.

Minimizing $\Pr(\delta(\mathbf{z}) = 1|\mathbf{a})$ as in (12) is equivalent to maximizing $\Pr(\delta(\mathbf{z}) = 0|\mathbf{a})$. We will show that for any detector for which the rejection region (the set of $\mathbf{z}$ with $\delta(\mathbf{z}) = 0$) is convex—true for both δ_2 and δ_∞—the latter probability is log-concave in $\mathbf{a}$. Therefore the optimization problem in (12) is to maximize a concave function, but over a non-convex set given by the constraint in (10). Conversely, the equivalent optimization problem in (13) is to minimize a convex function over a non-convex set.

Proposition 1: The probability $\Pr(\mathbf{z} \in A|\mathbf{a})$ is log-concave for any convex set A.

Proof: We may write

$$\Pr(\mathbf{z} \in A|\mathbf{a}) = \int f(\mathbf{z}')\mathbf{1}(\mathbf{z}' + \mathbf{a} \in A)d\mathbf{z}' \tag{21}$$

where $\mathbf{z}' = \mathbf{Hx} + \mathbf{e}$, and $f(\mathbf{z}')$ is the probability density function of $\mathbf{z}'$. Since $\mathbf{z}'$ is Gaussian, $f(\mathbf{z}')$ is log-concave in $\mathbf{z}'$. Moreover, the function $\mathbf{1}(\mathbf{z}' + \mathbf{a} \in A)$ is log-concave in $\mathbf{z}'$ and $\mathbf{a}$ since A is convex. Therefore the integral in (21) represents the convolution of two log-concave sets, which is itself log-concave (see [8]). ■

We evaluate the two detectors on the IEEE 14-bus test system. We take bus 1 to be the reference bus, so this system has 13 state variables and 54 measurements. We assume that the prior distribution on the states are given by $\Sigma_x = \sigma_x^2 I$, and the measurement errors are given by $\Sigma_e = \sigma_e^2 I$. Throughout our simulations, we use the parameters $\sigma_x^2 = 1$ and $\sigma_e^2 = 0.1$. For the L_∞ Detector, we use $\tau_2 = 4$, and again vary τ_1 based on the desired probability of false alarm.

For the L_2 norm detector, the probability $\Pr(\|\mathbf{Gz}\|_2 \geq \tau)$ can be evaluated directly, using the techniques in [9]. Finding this probability for the L_∞ Detector is not as straightforward, so we use Monte Carlo approximation. For the case when

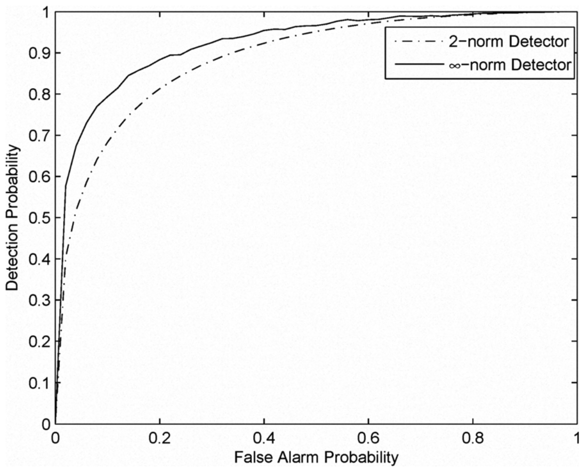

Fig. 1. ROC curves for L_∞ and L_2 residual detectors for 1-sparse false data injections.

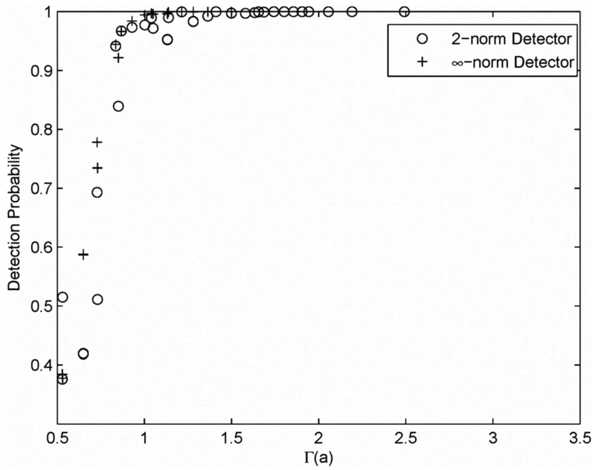

Fig. 3. Comparison of the heuristic $\Gamma(\mathbf{a})$ with the true detection probability for 1-sparse vectors for both L_2 and L_∞ detectors.

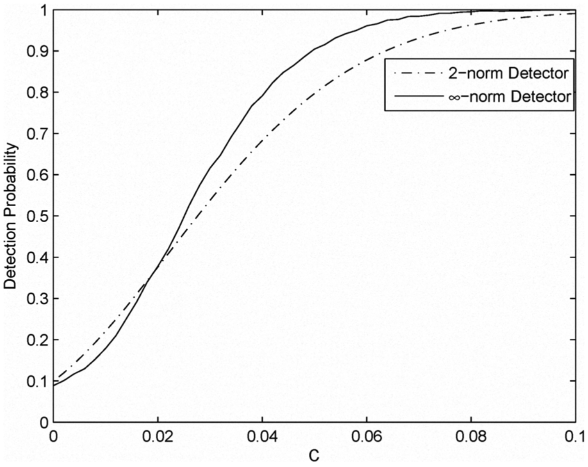

Fig. 2. Detection probabilities of L_∞ and L_2 residual detectors as functions of mean square error C for 1-sparse false data injections.

$k = 1$, i.e. the adversary controls only one meter, so $\mathbf{a}$ is 1-sparse, the optimization in (12) is easy, since it is clear that (10) should hold with equality, so for each sparsity pattern (of which there are only $n = 54$), there is only one choice for $\mathbf{a}$. For each $i = 1, \dots, n$, we evaluate the probability of detection for the $\mathbf{a}$ nonzero only in a_i satisfying equality in (10), then minimize over i. This results in Figures 1 and 2. Figure 1 compares the ROC curves for the two detectors for $C = 0.04$. For this value of C, the L_∞ Residual Detector outperforms the L_2 norm detector at all probabilities of false alarm. Figure 2 compares the detection probabilities of the two detectors at a fixed probability of false alarm of 0.1 but varying C. The L_∞ Residual Detector performs better than the L_2 norm detector at high C, but worse at low C. It seems that as the error injected by the adversary shrinks and becomes comparable to the Gaussian noise, the more standard L_2 detector becomes the better one.

Comparing the two detectors for $k > 1$ is much more difficult, for the two reasons outlined above: the number of sparsity patterns grows very large, and for each one, optimizing the probability is nontrivial. Therefore we do not do this exactly for $k > 1$, but we use the heuristic $\Gamma(\mathbf{a})$ to do some approximate analysis at higher sparsity levels.

C. Performance Approximation Using Detectability Heuristic

We first present some numerical evidence that the heuristic $\Gamma(\mathbf{a})$ described in Section III works well in that $\Pr(\delta(\mathbf{z}) = 1|\mathbf{a})$ is roughly increasing in $\Gamma(\mathbf{a})$ for both our detectors. We proceed to find approximate performance levels of these two detectors at sparsity levels above $k = 1$.

For the two detectors δ_2 and δ_∞, we consider the detection probability for all 1-sparse vectors $\mathbf{a}$ satisfying equality in (10) on the IEEE 14-bus test system. We use parameters $C = 0.01$ and $\alpha = 0.1$, and plot in Figure 3 $\Gamma(\mathbf{a})$ versus the true probability of detection for 1-sparse $\mathbf{a}$ and for both detectors. Observe that the scatter plots are roughly increasing.

Additionally, we evaluate the performance of the heuristic on 2-sparse vectors in the following way. For each pair of entries i, j of $\mathbf{a}$, we optimize (15) with $S = \{i, j\}$. This gives the $\mathbf{a}$ with sparsity S optimal according to Γ. We then evaluate the true probability of detection for the two detectors, with the same parameter values as above. The results are shown in Figure 4 for the L_2 detector and Figure 5 for the L_∞ detector. Again, the heuristic appears to track the true probabilities reasonably well.

With the heuristic, we can use brute force on $k = 3$ and $k = 4$ to find a good—if not necessarily optimal—k-sparse $\mathbf{a}$. The results of our analysis for $k = 0, \dots, 4$ are summarized in Table I. Listed there is the set of k meters that were found to be best for the adversary to control, and the resulting mean square error if the detection probability is raised from the false alarm of 0.1 to 0.5. When the sparsity reaches 4, the attack discussed in [7] becomes possible, so the mean square error radically increases. However, note that for the L_∞ Detector, it increases only to about an order of magnitude above $\sigma_x^2 = 1$.

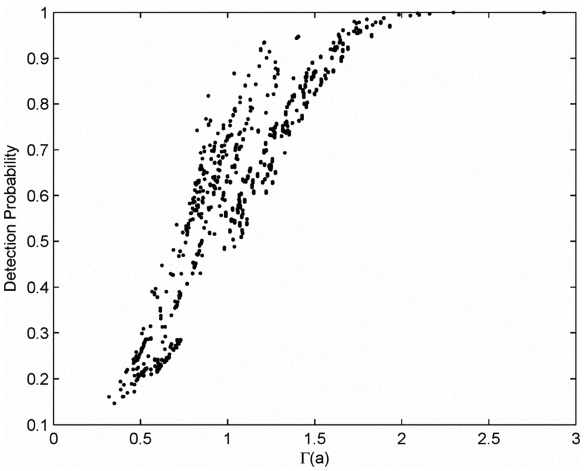

Fig. 4. Comparison of the heuristic $\Gamma(\mathbf{a})$ with the true detection probability for 2-sparse vectors for the L_2 detector.

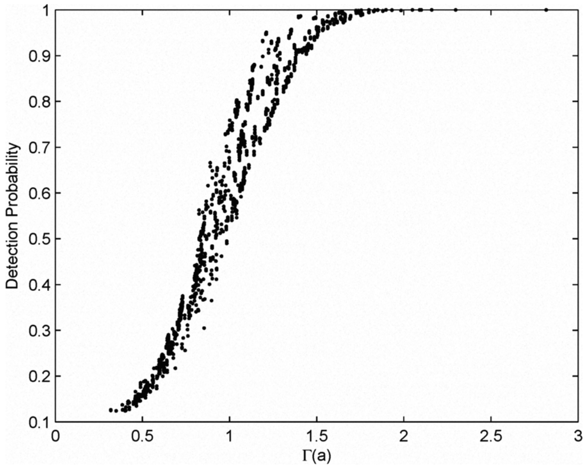

Fig. 5. Comparison of the heuristic $\Gamma(\mathbf{a})$ with the true detection probability for 2-sparse vectors for the L_∞ detector.

It would be impossible for it to be less than σ_x^2. As the L_2 detector has no ability to ascertain that this attack may be taking place, the mean square error jumps to a level several orders of magnitude higher.

V. Conclusions and Future Work

We studied the problem of adversarial false data injection in power system state estimation. We presented a novel formulation for the bad data detection problem. We introduced a heuristic for the detectability of a particular attack by the adversary, which allows particularly bad attacks to be easily computed for any set of compromised meters. Finally, we proposed a detector that can outperform the classical detector, and demonstrated that our heuristic works well for both the classical L_2 detector and our L_∞ Detector. We saw in Section IV-B that the L_∞ Detector performs better than classical L_2 detector for certain problem parameters, but not all. It would obviously be advantageous to use the one more likely to perform better in the given circumstances. In fact, one would like to design a more elaborate detector that can continuously trade off between the two, always choosing the optimal operating point. We have argued that the detector should share some of the properties of the L_∞ Residual Detector, but this is by no means a proof that it is optimal. It would be desirable to find the truly optimal detector, or if not, demonstrate rigourously that the performance of our detector is not far from the best.

TABLE I

Detector Performance with Increasing Sparsity

k	Sparsity pattern	MSE for L_2	MSE for L_∞
0	$\{\}$	0.0197	0.0197
1	$\{1\}$	0.0473	0.0450
2	$\{1, 15\}$	0.0705	0.0856
3	$\{1, 15, 35\}$	0.146	0.178
4	$\{7, 8, 28, 48\}$	20900	29.4

Furthermore, several related problems could be considered. First, real state estimation is performed over a very long period of time, in which measurements arrive asynchronously, and data is received from each meter continuously. It would be worthwhile to develop a measurement-update type state estimator with false data detection, capable of observing the long-term behavior of each meter and discerning whether it might be compromised. In such a scenario, the attacks may become more complicated as well, as an adversary would need to spread its influence over time in order to avoid being detected. Moreover, it is yet unclear exactly how much damage to the performance of the power system can be done via false data attacks such as these. For example, if it were possible to inject false data in such a way that the control center developed an incorrect belief about the topology of the network, further problems could develop as operators make control decisions based on faulty information.

References

[1] F. C. Scheppe, J. Wildes, and D. B. Rom, "Power system static state estimation, parts 1, 2, 3," *IEEE Trans. on Power Apparatus and Systems*, vol. 89, no. 1, pp. 120–135, Jan. 1970.

[2] E. Handschin, F. C. Schweppe, J. Kohlas, and A. Fiechter, "Bad data analysis for power system state estimation," *IEEE Trans. on Power Apparatus and Systems*, vol. 94, no. 2, pp. 329-337, April 1975.

[3] A. Garcia, A. Monticelli, and P. Abreu, "Fast decoupled state estimation and bad data processing," *IEEE Transactions on Power Apparatus and Systems*, vol. 98, no. 5, pp. 1645-1652, September 1979.

[4] T. Van Cutsem, M. Ribbens-Pavella, and L. Mili, "Bad Data Identification Methods In Power System State Estimation—A Comparative Study," *IEEE Trans. on Power Apparatus and Systems*, vol. PAS-104, no. 11, pp. 3037–3049, Nov. 1985.

[5] H.-J. Koglin, T. Neisius, G. Beissler, and K. D. Schmitt, "Bad data detection and identification," *Int. J. Elect. Power*, vol. 12, no. 2, pp. 94103, Apr. 1990.

[6] J. Chen and A. Abur, "Improved bad data processing via strategic placement of PMUs," *IEEE Power Engineering Society General Meeting*, pp. 509-513 Vol. 1, June 2005.

[7] Y. Liu, M. K. Reiter, and P. Ning, "False data injection attacks against state estimation in electric power grids," in *Proc. of the 16th ACM Conference on Computer and Communications Security*, 2009.

[8] A. Prékopa, "Logarithmic concave measures and related topics," in M. A. H. Dempster, editor, *Stochastic Programming*, pp. 63-82, Academic Press, 1980.

[9] J. Sheil and I. O'Muircheartaigh, "The distribution of non-negative quadratic forms in normal variables" *Journal of the Royal Statistical Society*, Series C (Applied Statistics), Vol. 26, no. 1, pp. 92–98, 1977.