Data Framing Attack on State Estimation With Unknown Network Parameters

Jinsub Kim, Lang Tong, and Robert J. Thomas
School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853
Email: {jk752, ltong, rjt1}@cornell.edu

***Abstract*—A data framing attack is presented to exploit the bad data detection and identification mechanisms at a typical ISO/RTO control center. In particular, the proposed attack frames normal meters as sources of bad data and causes the control center to remove useful measurements from the framed meters. The proposed attack uses subspace information of power system measurements; neither the network topology nor the network parameters are required for constructing the attack. It is shown that the proposed attack is capable of perturbing the power system state estimate by an arbitrary degree using only half of the critical measurements. Implications of this attack on power system operations are discussed, and the attack performance is evaluated using benchmark systems.**

***Index Terms*—Power system state estimation, framing attack, bad data test, cyber security of smart grid.**

I. Introduction

The paradigm shift to a data-driven grid control enables integration of sophisticated data processing methods for more efficient and reliable grid operations. However, it exposes the grid to possible cyber attacks that may disrupt grid operations and potentially cause a cascading failure.

Liu, Ning, and Reiter presented in [1] the first man-in-the-middle attack on the power system state estimation where an adversary replaces part of "normal" sensor data with "malicious data." They showed that, if an adversary can control a sufficient number of meter data, it can perturb the state estimate by an arbitrary amount without being detected by the bad data detector at the control center. Such undetectable attacks are referred to as *covert state attacks*.

The condition under which covert state attacks are possible is found to be equivalent to that of network observability. In particular, covert attacks are possible if and only if the network becomes unobservable when the adversary-controlled meters are removed [2]. The minimum number of meters that an adversary has to control in order to launch a covert state attack, referred to as a *security index*, is a measure of security against data attacks. It represents a fundamental limit on the capability of an adversary to disrupt grid operations covertly [2], [3].

A. Summary of results

In this paper, we show that the barrier on the capability of an attacker represented by the security index can be circumvented by constructing a data framing attack aimed at misleading the control center about the source of bad data. In particular, we show that the adversary only needs to gain control of about half of the meters required by the security index while achieving the same objective of perturbing the state estimate by an arbitrary amount without being detected.

Existing attack strategies typically assume a knowledge of network topology and network parameters. To our best knowledge, the present paper is the first to construct a data attack based on certain *subspace information* of meter measurements without network parameter and topology information.

B. Related work

There is an extensive literature on *covert state attacks* following the work of Liu, Ning, and Reiter [1]. The link between feasibility of covert state attacks and network observability was made in [2], [4]. Consequently, network observability conditions [5] can be modified for that for covert attacks and used to develop meter protection strategies [2], [4], [6]–[9]. To assess the grid vulnerability against data attacks, the minimum number of adversary meters necessary for a covert attack was suggested as the security index for the grid [2], [3].

The framing attack strategy considered here relies on bad data identification and removal techniques that have long been subjects of study (see [10]–[12] and references therein.) Typically, the residue vectors in normalized forms are widely used as statistics for the bad data test [10]. In this paper, we take the residue analysis in [10] as a representative bad data test and analyze the effect of the framing attack.

There is only limited work on attacking a network without network parameter or network topology. The use of independent component analysis in [13] is the most relevant. The authors of [13] propose to identify a mixing matrix from which to construct the attack. Generating attacks using local information has also been considered. See [14].

The rest of this paper is organized as follows. Section II introduces the measurement and adversary models with preliminaries on state attacks. Section III presents the mathematical model of state estimation and bad data processing. In Section IV, we present the main idea of the data framing attack, a theoretical justification of its efficacy, and how the estimated subspace information can be used to construct the framing attack. In Section V, the simulations with the IEEE 14-bus network and the nonlinear model demonstrate that the framing attacks designed based on the linearized system model

This work is supported in part by the National Science Foundation under Grant CNS-1135844 and the Army Research Office under Grant W911NF1010419.

can successfully perturb the state estimate. Finally, Section VI provides concluding remarks.

II. Mathematical models

A. Measurement model

For real-time estimation of the system state $\mathbf{x}$, the vector of bus voltage magnitudes and phase angles, the control center collects measurements from line flow and bus injection meters[1] deployed throughout the grid. The meter measurements are related to the system state $\mathbf{x}$ in a nonlinear fashion, and the relation is described by the AC model [12]:

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

where $h(\cdot)$ is the measurement function, and $\mathbf{e}$ is the Gaussian measurement noise.

If some of the meters malfunction or an adversary injects malicious data, the control center observes biased measurements,

$$\bar{\mathbf{z}} = h(\mathbf{x}) + \mathbf{e} + \mathbf{a}, \tag{2}$$

where $\mathbf{a}$ represents a deterministic bias. In such a case, the data are said to be *bad*, and the biased meter entries are referred to as *bad data entries*. Note that even when a meter is protected from adversarial modification, it may still have a bias due to a physical malfunction or improper parameter setting; filtering out the measurements from such malfunctioning meters was the original objective of the legacy bad data processing and is adopted in practice today [10].

In analyzing the attack effect on state estimation, we adopt the linearized DC model [12]:

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{e}, \tag{3}$$

where $\mathbf{z} \in \mathbb{R}^m$ is the measurement vector consisting of the real part of the line flow and bus injection measurements, the system state $\mathbf{x} \in \mathbb{R}^n$ is the vector of voltage phase angles at all buses except the reference bus, $\mathbf{H} \in \mathbb{R}^{m \times n}$ is the measurement matrix that relates the system state to bus injection and line flow amounts, and $\mathbf{e}$ is the Gaussian measurement noise with a diagonal covariance matrix $\mathbf{\Sigma} \triangleq \sigma^2 \bar{\mathbf{\Sigma}}$, where $\bar{\mathbf{\Sigma}}$ is a diagonal matrix representing the variation of noise variances across different meters ($\sum_{i=1}^{m} \bar{\mathbf{\Sigma}}_{ii} = 1$), and σ^2 is a scaling factor.

B. Adversary model

As described in Fig. 1, an adversary is assumed to be capable of modifying the data from a subset of meters $\mathcal{S}_{\mathrm{A}}$, referred to as *adversary meters*. The control center observes corrupted measurements $\bar{\mathbf{z}}$ instead of the real measurements $\mathbf{z}$. The adversarial modification is mathematically modeled as follows:

$$\bar{\mathbf{z}} = \mathbf{z} + \mathbf{a}, \quad \mathbf{a} \in \mathcal{A}, \tag{4}$$

where $\mathbf{a}$ is an attack vector, and $\mathcal{A}$ is the set of feasible attack vectors defined as $\mathcal{A} \triangleq \{\mathbf{c} \in \mathbb{R}^m : \ c_i = 0, \ \forall i \notin \mathcal{S}_{\mathrm{A}}\}$.

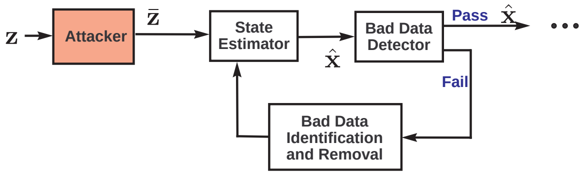

Fig. 1. Adversary model with state estimation and bad data test

The adversary is assumed to know a basis matrix $\mathbf{U}$ of the column space of $\mathbf{H}$ and the noise covariance matrix $\mathbf{\Sigma}$. In practice, a basis matrix can be inferred based on multiple measurement samples.

C. Network observability and covert state attack

A network is said to be *observable* if the DC measurement matrix $\mathbf{H}$ has full rank (*i.e.*, $\mathbf{x}$ can be uniquely determined from observing $\mathbf{Hx}$.) In practice, power network measurements should be designed to satisfy observability. Hence, we assume that the network of our interest is observable, *i.e.*, $\mathbf{H}$ has full rank.

The concept of network observability is closely related to the feasibility of a covert state attack. A covert state attack was proposed in [1] under the DC model: if there exists $\mathbf{y} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$ such that $\mathbf{Hy} \in \mathcal{A}$, then setting $\mathbf{a}$ equal to $\mathbf{Hy}$ results in

$$\bar{\mathbf{z}} = \mathbf{Hx} + \mathbf{e} + \mathbf{a} = \mathbf{H}(\mathbf{x} + \mathbf{y}) + \mathbf{e}, \tag{5}$$

and thus, $\bar{\mathbf{z}}$ cannot be distinguished from a normal noisy measurement vector with the state $\mathbf{x} + \mathbf{y}$. Furthermore, by properly scaling the attack vector (*e.g.*, $\alpha\mathbf{a}$), the adversary can perturb the state estimate by an *arbitrary* degree (*e.g.*, $\alpha\mathbf{y}$).

It was shown in [2] that a covert attack is feasible if and only if $\mathcal{S}_{\mathrm{A}}$ contains a *critical set* of meters, which is defined as a set of meters such that removing the set from the network renders the network unobservable while removing any proper subset of it does not [12]. In other words, the feasibility condition means that removing the adversary meters renders the measurement matrix rank deficient. Therefore, the *security index* of the grid [2]—the minimum number of meters an adversary needs to control to launch a covert attack—is equivalent to the cardinality of the smallest critical set.

III. State estimation and bad data processing

This section introduces a popular approach of state estimation and bad data processing [10], [12], which we assume to be employed by the control center. Fig. 1 illustrates an iterative scheme for obtaining an estimate $\hat{\mathbf{x}}$ of the system state, which consists of three function blocks: State Estimation, Bad Data Detection, and Bad Data Identification.

The iteration begins with the initial measurement vector $\mathbf{z}^{(1)} \triangleq \mathbf{z}$ and the initial measurement function $h^{(1)} \triangleq h$ where the superscript denotes the index for the current iteration.

In the kth iteration, State Estimation uses $(\mathbf{z}^{(k)}, h^{(k)})$ as an input and calculates the weighted least squares (WLS) estimate

[1] Other types of meters can also be considered, but we restrict our attention to line flow and bus injection meters for simplicity.

of the system state and the corresponding residue vector:

$$\hat{\mathbf{x}}^{(k)} \triangleq \arg\min_{\mathbf{x}} (\mathbf{z}^{(k)} - h^{(k)}(\mathbf{x}))^T (\mathbf{\Sigma}^{(k)})^{-1} (\mathbf{z}^{(k)} - h^{(k)}(\mathbf{x})),$$
$$\mathbf{r}^{(k)} \triangleq \mathbf{z}^{(k)} - h^{(k)}(\hat{\mathbf{x}}^{(k)}). \tag{6}$$

where $\mathbf{\Sigma}^{(k)}$ is the covariance matrix of the corresponding noise vector.

We assume that the $J(\hat{\mathbf{x}})$-test [10], [12] is employed for bad data detection:

$$\begin{cases} \text{bad data} & \text{if } (\mathbf{r}^{(k)})^T (\mathbf{\Sigma}^{(k)})^{-1} \mathbf{r}^{(k)} > \tau^{(k)}; \\ \text{good data} & \text{if } (\mathbf{r}^{(k)})^T (\mathbf{\Sigma}^{(k)})^{-1} \mathbf{r}^{(k)} \leq \tau^{(k)}. \end{cases} \tag{7}$$

The $J(\hat{\mathbf{x}})$-test is widely used due to its low complexity and the fact that the test statistic has a χ^2 distribution if the data are good [10]. The latter fact is used to set the threshold $\tau^{(k)}$ for a given false alarm constraint.

If Bad Data Detection (7) declares that the data are good, the algorithm returns the state estimate $\hat{\mathbf{x}}^{(k)}$ and terminates. However, if Bad Data Detection declares that the data are bad, Bad Data Identification is invoked to identify and remove one bad data entry from the measurement vector.

A widely used criterion for identifying a bad data entry is the normalized residue [10], [12]: each $r_i^{(k)}$ is divided by its standard deviation under the hypothesis that there exists no bad data entry in $\mathbf{z}^{(k)}$. Specifically,

$$\tilde{\mathbf{r}}^{(k)} \triangleq \mathbf{\Omega}^{(k)} \mathbf{r}^{(k)}, \tag{8}$$

where $\mathbf{\Omega}^{(k)}$ is a diagonal matrix with

$$\mathbf{\Omega}_{ii}^{(k)} \triangleq \begin{cases} 0 & \text{if } \{i\} \text{ is a critical set}^2, \\ \frac{1}{\sqrt{(\mathbf{W}^{(k)}\mathbf{\Sigma}^{(k)})_{ii}}} & \text{otherwise;} \end{cases} \tag{9}$$

and

$$\mathbf{W}^{(k)} \triangleq \mathbf{I} - \mathbf{H}^{(k)} ((\mathbf{H}^{(k)})^T (\mathbf{\Sigma}^{(k)})^{-1} (\mathbf{H}^{(k)}))^{-1} (\mathbf{H}^{(k)})^T (\mathbf{\Sigma}^{(k)})^{-1} \tag{10}$$

with $\mathbf{H}^{(k)}$ denoting the Jacobian of $h^{(k)}$ at $\hat{\mathbf{x}}^{(k)}$ (see Appendix of [10] for the detail.)

Once the normalized residue $\tilde{\mathbf{r}}^{(k)}$ is calculated, the meter with the largest $|\tilde{r}_i^{(k)}|$ is identified as a bad meter. Bad Data Identification removes the row of $\mathbf{z}^{(k)}$ and the row of $h^{(k)}$ that correspond to the bad meter and returns the updated measurement vector $\mathbf{z}^{(k+1)}$ and measurement function $h^{(k+1)}$, which are inputs for the next iteration.

Under the DC model (3), every step is the same with that in the AC model, except that the nonlinear measurement function $h^{(k)}(\mathbf{x})$ is replaced with the linear function $\mathbf{H}^{(k)}\mathbf{x}$ (so, the Jacobian is the same everywhere.) Note that the WLS state estimate (6) is replaced with a simple linear WLS solution:

$$\hat{\mathbf{x}}^{(k)} = ((\mathbf{H}^{(k)})^T (\mathbf{\Sigma}^{(k)})^{-1} (\mathbf{H}^{(k)}))^{-1} (\mathbf{H}^{(k)})^T (\mathbf{\Sigma}^{(k)})^{-1} \mathbf{z}^{(k)}, \tag{11}$$

and thus

$$\mathbf{r}^{(k)} = \mathbf{z}^{(k)} - \mathbf{H}^{(k)} \hat{\mathbf{x}}^{(k)} = \mathbf{W}^{(k)} \mathbf{z}^{(k)}. \tag{12}$$

IV. Data framing attack

In this section, we present the main idea of data framing attack and demonstrate that the data framing attack enables the adversary controlling only a half of a critical set of meters to perturb the state estimate by an arbitrary degree. In addition, we present the attack construction based on a basis matrix of the column space of $\mathbf{H}$.

A. Main idea and the factor-of-two result

Suppose that $\{\mathcal{S}_1, \mathcal{S}_2\}$ is a partition of a critical set, and let $\bar{\mathbf{H}}$ denote the measurement matrix after removing the meters in $\mathcal{S}_1 \cup \mathcal{S}_2$ from the grid. Since $\mathcal{S}_1 \cup \mathcal{S}_2$ is a critical set, $\bar{\mathbf{H}}$ has rank $n-1$, and the dimension of its null space $\mathcal{N}(\bar{\mathbf{H}})$ is one. Let $\Delta\mathbf{x}$ denote a unit basis vector of $\mathcal{N}(\bar{\mathbf{H}})$. Now, we consider two vectors, $\mathbf{H}_1\Delta\mathbf{x}$ and $\mathbf{H}_2\Delta\mathbf{x}$, where $\mathbf{H}_1$ is the $m \times n$ matrix obtained from $\mathbf{H}$ by replacing the rows corresponding to the meters in $\mathcal{S}_2$ with zero row vectors ($\mathbf{H}_2$ is defined in the same way by replacing the rows corresponding to $\mathcal{S}_1$.) Since $\Delta\mathbf{x} \in \mathcal{N}(\bar{\mathbf{H}})$, $\mathbf{H}_1\Delta\mathbf{x}$ has nonzero entries only at the locations corresponding to the meters in $\mathcal{S}_1$; *i.e.*, $\mathbf{H}_1\Delta\mathbf{x}$ is a feasible attack vector when $\mathcal{S}_{\text{A}} = \mathcal{S}_1$. Similarly, $\mathbf{H}_2\Delta\mathbf{x}$ is a feasible attack vector when $\mathcal{S}_{\text{A}} = \mathcal{S}_2$. Since both attack vectors are not in the column space of $\mathbf{H}$, if the noise magnitude is small (*i.e.*, $\sigma^2 \ll 1$,) any of these attacks will cause the iterative bad data processing to detect presence of bad data, identify some meters as bad, and remove them.

In the following, we will provide an intuition for why at least one of these two attack vectors—$\mathbf{H}_1\Delta\mathbf{x}$ and $\mathbf{H}_2\Delta\mathbf{x}$—will succeed in perturbing the state estimate. For the ease of presentation, we present the idea using noiseless measurements.

First, note that

$$\mathbf{H}_1\Delta\mathbf{x} + \mathbf{H}_2\Delta\mathbf{x} = \mathbf{H}\Delta\mathbf{x}, \tag{13}$$

because $\Delta\mathbf{x} \in \mathcal{N}(\bar{\mathbf{H}})$. Therefore,

$$\mathbf{H}\mathbf{x} + \eta\mathbf{H}_2\Delta\mathbf{x} = \mathbf{H}(\mathbf{x} + \eta\Delta\mathbf{x}) - \eta\mathbf{H}_1\Delta\mathbf{x}, \tag{14}$$

where $\eta \in \mathbb{R}$ is a nonzero scaling factor. Now consider the first iteration of the bad data processing. Because the part of the measurement vector that is in the column space of $\mathbf{H}$ does not affect the residues, (14) implies that setting $\mathbf{a} = \eta\mathbf{H}_2\Delta\mathbf{x}$ and setting $\mathbf{a} = -\eta\mathbf{H}_1\Delta\mathbf{x}$ result in the same residue vector and thereby removal of the same meter. On the other hand, $\mathbf{a} = -\eta\mathbf{H}_1\Delta\mathbf{x}$ and $\mathbf{a} = \eta\mathbf{H}_1\Delta\mathbf{x}$ result in the residue vectors $-\mathbf{W}^{(1)}(\eta\mathbf{H}_1\Delta\mathbf{x})$ and $\mathbf{W}^{(1)}(\eta\mathbf{H}_1\Delta\mathbf{x})$ respectively. Therefore, since bad data detection and identification exclusively relies on the *magnitudes* of residues, the two attack vectors result in removal of the same meter in the first iteration. Consequently, $\mathbf{a} = \eta\mathbf{H}_1\Delta\mathbf{x}$ and $\mathbf{a} = \eta\mathbf{H}_2\Delta\mathbf{x}$ result in removal of the same meter in the first iteration. It can be easily seen that the same logic can be applied to the subsequent iterations, and thus $\mathbf{a} = \eta\mathbf{H}_1\Delta\mathbf{x}$ and $\mathbf{a} = \eta\mathbf{H}_2\Delta\mathbf{x}$ result in removal of the same sequence of bad meters in the bad data processing, which we denote by $(i_1, \ldots, i_N)$ where i_k is the index of the meter removed in the kth iteration.

[2] If $\{i\}$ is a critical set (*i.e.*, removing the meter i makes the grid unobservable), its residue is always equal to zero [12], and the corresponding diagonal entry of $\mathbf{W}^{(k)}\mathbf{\Sigma}^{(k)}$ is zero. For such a meter, the normalizing factor is 0 such that its normalized residue is equal to 0.

Second, since the iterative bad data processing never removes an entire critical set [12], at least one meter in $\mathcal{S}_1 \cup \mathcal{S}_2$ is not contained in $\{i_1, \ldots, i_N\}$. Suppose that a meter in $\mathcal{S}_1$ is not in $\{i_1, \ldots, i_N\}$. Then, an adversary with $\mathcal{S}_{\text{A}} = \mathcal{S}_1$ can set $\mathbf{a} = \eta \mathbf{H}_1 \Delta \mathbf{x}$ such that some adversary meter remains at the end of bad data processing and perturbs the state estimate[3]. Suppose that all the meters in $\mathcal{S}_1$ are in $\{i_1, \ldots, i_N\}$. Then, it implies that at least one meter in $\mathcal{S}_2$ is not in $\{i_1, \ldots, i_N\}$, and an adversary with $\mathcal{S}_{\text{A}} = \mathcal{S}_2$ can set $\mathbf{a} = \eta \mathbf{H}_2 \Delta \mathbf{x}$ such that some adversary meter survives the bad data processing and perturbs the state estimate. Note that the adversary can adjust the perturbation level by using a proper η.

Formally, the following theorem provides a sufficient condition that guarantees that the framing attack can use one of $\mathcal{S}_1$ and $\mathcal{S}_2$ to perturb the state estimate by an arbitrary degree when the meter signal-to-noise ratios (SNRs) are high.

Theorem 4.1: Suppose that if we run the noiseless[4] version of the iterative bad data processing on $\mathbf{H}_1 \Delta \mathbf{x}$, then there exists a unique state $\mathbf{y} \in \mathbb{R}^n$ such that the final state estimate is always equal to $\mathbf{y}$ regardless of whatever decisions are made under tie[5] situations in Bad Data Identification. Under this condition, the following hold for any $\mathbf{x} \in \mathbb{R}^n$:

(1) Suppose $\mathbf{y} \neq \mathbf{0}$. If the framing attack with $\mathcal{S}_{\text{A}} = \mathcal{S}_1$ is launched, *i.e.*, $\mathbf{a} = \eta \mathbf{H}_1 \Delta \mathbf{x}$ where $\eta \in \mathbb{R}$ is a scaling factor,

$$\lim_{\sigma^2 \to 0} \Pr(\bar{\mathbf{z}}^{(N)} = \mathbf{H}^{(N)}(\mathbf{x} + \eta \mathbf{y}) + \mathbf{e}^{(N)}) = 1, \qquad (15)$$

where N is the random variable representing the total number of iterations in the bad data processing.

(2) Suppose $\mathbf{y} \neq \Delta \mathbf{x}$. If the framing attack with $\mathcal{S}_{\text{A}} = \mathcal{S}_2$ is launched, *i.e.*, $\mathbf{a} = \eta \mathbf{H}_2 \Delta \mathbf{x}$,

$$\lim_{\sigma^2 \to 0} \Pr(\bar{\mathbf{z}}^{(N)} = \mathbf{H}^{(N)}(\mathbf{x} + \eta(\Delta \mathbf{x} - \mathbf{y})) + \mathbf{e}^{(N)}) = 1. \quad (16)$$

Proof: See Appendix in [15]. ■

Theorem 4.1 implies that if the condition is met, then at least one of $\mathcal{S}_1$ and $\mathcal{S}_2$ can be used by the framing attack to perturb the state estimate by an arbitrary degree, because $\mathbf{y}$ cannot be simultaneously $\mathbf{0}$ and $\Delta \mathbf{x}$.

If the condition of Theorem 4.1 holds for a partition with $|\mathcal{S}_1| = |\mathcal{S}_2|$, then the adversary controlling only a half of the critical set can perturb the state estimate by an arbitrary degree. One important question is whether a partition $\{\mathcal{S}_1, \mathcal{S}_2\}$ with $|\mathcal{S}_1| \simeq |\mathcal{S}_2|$ that satisfies the condition can be found in general. To answer this question, we investigated critical sets associated with cuts of the network topology (*i.e.*, the set of the line meters on the cut-set lines and the injection meters on the both ends of the cut-set lines). We found 118 cuts in the IEEE 14-bus network and 290 cuts in the IEEE 118-bus network. For every critical set[6] associated with each cut, we were able to construct a parition with $\left| |\mathcal{S}_1| - \frac{|\mathcal{S}|}{2} \right| \leq 1$ satisfying the condition of Theorem 4.1 (refer to [15] for the details.)

B. Attack with unknown network parameters

Theorem 4.1 provides a way to find an adversary meter set and design the data framing attack based on $\mathbf{H}$. In fact, knowledge of a basis matrix $\mathbf{U}$ of the column space of $\mathbf{H}$ is sufficient for designing the attack. The following are the detailed steps for the attack design based on $\mathbf{U}$:

- **Step 0**. Find a critical set $\mathcal{S}$ and its partition $\{\mathcal{S}_1, \mathcal{S}_2\}$. This can be achieved by finding a set of rows of $\mathbf{U}$, removal of which makes $\mathbf{U}$ rank-deficient while removing any proper subset of it does not.
- **Step 1**. Find a nonzero vector $\Delta \mathbf{v} \in \mathcal{N}(\bar{\mathbf{U}})$ where $\bar{\mathbf{U}}$ denotes the submatrix of $\mathbf{U}$ obtained by removing the rows corresponding to $\mathcal{S}$.
- **Step 2**. Run the noiseless version of the bad data processing on $\mathbf{U}_1 \Delta \mathbf{v}$ ($\mathbf{U}_1$ is obtained from $\mathbf{U}$ by replacing the rows corresponding to $\mathcal{S}_2$ with zero row vectors, and $\mathbf{U}_2$ is similarly defined based on $\mathcal{S}_1$.) If any meter in $\mathcal{S}_1$ remains unremoved, the adversary uses $\mathcal{S}_1$ as the adversary meter set and set $\mathbf{a} = \eta \mathbf{U}_1 \Delta \mathbf{v}$ where $\eta \in \mathbb{R}$ is set according to desired perturbation amount. Otherwise, the adversary uses $\mathcal{S}_2$ as the adversary meter set and set $\mathbf{a} = \eta \mathbf{U}_2 \Delta \mathbf{v}$.

The above attack design based on $\mathbf{U}$ is possible, because all we need in designing the attack is the column or null space information of $\mathbf{H}$ (or its submatrices consisting of a subset of rows), and $\mathbf{U}$ contains all the information.

The step 0 can be omitted if the adversary already knows a critical set. In practice, when an estimate $\hat{\mathbf{U}}$ of a basis matrix is used, even small estimation errors may cause the numerical rank of a submatrix erroneous. One way to handle this problem is to compare the singular values of the submatrix with certain threshold and count the number of singular values larger than the threshold to estimate the rank.

V. Numerical results

We tested the performance of the framing attack with the IEEE 14-bus network under the AC model. As a performance metric, we used $\mathbb{E}[\|\hat{\mathbf{x}} - \mathbf{x}\|_2]$, where $\hat{\mathbf{x}}$ is the state estimate, and $\mathbf{x}$ is the true state.

In each Monte Carlo run, the true state $\mathbf{x}$ was generated by a multivariate Gaussian distribution with small variances. Its mean was set as the operating state given by the IEEE 14-bus data [16]. Based on $\mathbf{x}$, the noisy measurements were generated by the measurement model ($i.e., h(\mathbf{x}) + \mathbf{e}$). The attack vector was constructed based on an estimate[7] $\hat{\mathbf{U}}$ of a basis matrix of the column space of $\mathbf{H}$, as described in Section IV-B. Once

[3] It is not hard to see that all the entries of $\eta \mathbf{H}_1 \Delta \mathbf{x}$ corresponding to the meters in $\mathcal{S}_1$ are nonzero, because $\mathcal{S}_1 \cup \mathcal{S}_2$ is a critical set. Hence, as long as any meter in $\mathcal{S}_1$ remains, it will perturb the state estimate by nonzero amount.

[4] The noiseless version means the algorithm which the bad data processing converges to as σ^2 decays to 0. The only difference from the normal bad data processing is that in each iteration, Bad Data Detection declares that data are good if and only if State Estimation results in a zero residue vector.

[5] It is possible that a *tie* may occur in Bad Data Identification at some iteration: *i.e.*, the largest absolute normalized residue is assumed by more than one meter. In a tie situation, we assume that Bad Data Identification chooses an arbitrary meter with the largest absolute normalized residue.

[6] The average cardinality of the critical sets we considered is 15.7 for the 14-bus case and 12.7 for the 118-bus case.

[7] The estimate is obtained from the sample covariance estimate of $\mathbb{E}[\mathbf{z}\mathbf{z}^T]$ based on 1,000 independent measurements generated from the AC model.

constructed, the attack vector was added to the real part of the noisy measurements, and the iterative bad data processing[8] were executed on the corrupted measurements. Considering the linear decoupled model (see Chapter 2.7 in [12]), such addition of the attack vector is expected to modify primarily the bus voltage phase angles and have little effect on the bus voltage magnitudes. Hence, in interpreting the results, we focus on the *phase-angle* part of the state estimate error.

For comparison, we also executed the *conservative* scheme in [2], which aims to perturb the state estimate by the maximum degree while not raising any alarm in Bad Data Detection (see Problem (31) in [2] for details.)

We considered the adversary who can control $(2,3)$, $(3,4)$, and $(4,3)$: (i,j) denotes the line meter for the power flow from i to j, and (i) denotes the injection meter at bus i. The adversary meter set is a subset of a critical set consisting of $(3,4)$, $(4,3)$, $(2,3)$, $(3,2)$, (2), (3), and (4), which is associated with the cut isolating the bus 3 from the rest of the network. We tested the framing attack with three different attack magnitudes: $\|\mathbf{a}\|_1$ is 1%, 2%, or 3% of $\|\mathbf{z}\|_1$.

Fig. 2 shows the state estimate error versus the meter SNR in the AC simulations. The normal state estimate error and the state estimate error under the conservative scheme are very close, and both decays to zero as the SNR increases. For framing attacks, we tested both the attacks based on $\mathbf{H}$ matrix (solid lines) and the attacks based on a basis matrix estimate (dashed lines). Both attacks resulted in almost the same effect on the state estimate error thereby demonstrating that the estimated subspace information is sufficient for constructing an attack. The state estimate errors under the framing attacks converge to nonzero values, and the result implies that the framing attack can adjust the degree of resulting perturbation by choosing a proper attack magnitude (note that most practical meters have SNRs higher than 40 dB [17].)

VI. Conclusions

We have presented the data framing attack constructed from the subspace information of the meter measurement space. Controlling only a half of a critical set, the data framing attack can perturb the state estimate by an arbitrary degree. A theoretical justification was provided, and numerical experiments demonstrated the efficacy of the framing attack.

Our results indicate that most known countermeasures, that are aimed at merely preventing covert state attacks, are not sufficient for protection against the attacks aimed at state perturbation. In designing countermeasures, the possibility of the framing attack needs to be taken into account.

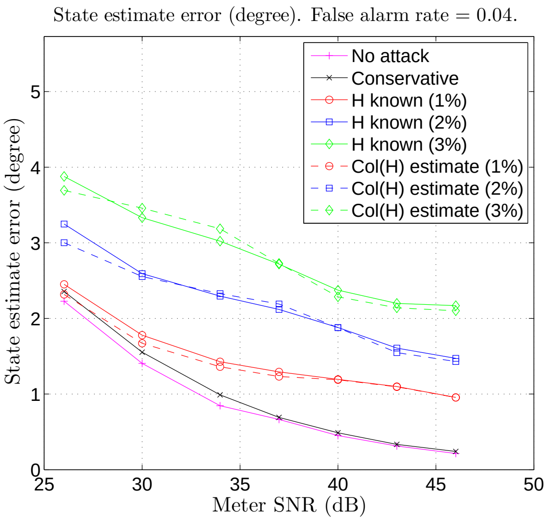

Fig. 2. AC simulations with the 14-bus network: 1,000 Monte Carlo runs. The adversary meters are $(2,3)$, $(3,4)$, and $(4,3)$.

References

[1] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proceedings of the 16th ACM conference on Computer and communications security*, 2009, pp. 21–32.

[2] 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.

[3] H. Sandberg, A. Teixeira, and K. H. Johansson, "On security indices for state estimators in power networks," in *First Workshop on Secure Control Systems,CPSWEEK 2010*, Stockholm, Sweeden, Apr 2010.

[4] R. B. Bobba, K. M. Rogers, Q. Wang, H. Khurana, K. Nahrstedt, and T. J. Overbye, "Detecting false data injection attacks on dc state estimation," in *First Workshop on Secure Control Systems,CPSWEEK 2010*, Stockholm, Sweeden, Apr 2010.

[5] G. R. Krumpholz, K. A. Clements, and P. W. Davis, "Power system observability: a practical algorithm using network topology," *IEEE Trans. Power Apparatus and Systems*, vol. 99, no. 4, pp. 1534–1542, July 1980.

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

[7] S. Bi and Y. Zhang, "Defending mechanisms against false-data injection attacks in the power system state estimation," in *2011 IEEE GLOBECOM Workshops*, Houston, TX, USA., Dec 2011.

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

[9] J. Kim and L. Tong, "On phasor measurement unit placement against state and topology attacks," in *IEEE International Conference on Smart Grid Communications*, Oct. 2013.

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

[11] T. Van Cutsem, M. Ribbens-Pavella, and L. Mili, "Bad data identification methods in power system state estimation-a comparative study," *IEEE Transactions on Power Apparatus and Systems*, vol. 104, no. 11, pp. 3037–3049, 1985.

[12] A. Abur and A. G. Expósito, *Power System State Estimation: Theory and Implementation*. CRC, 2000.

[13] M. Esmalifalak, H. Nguyen, R. Zheng, and Z. Han, "Stealth false data injection using independent component analysis in smart grid," in *IEEE International Conference on Smart Grid Communications*, Oct. 2011, pp. 244–248.

[14] M. Rahman and H. Mohsenian-Rad, "False data injection attacks with incomplete information against smart power grids," in *IEEE Global Communications Conference (GLOBECOM)*, Dec. 2012.

[15] J. Kim, L. Tong, and R. J. Thomas, "Data Framing Attack on State Estimation," *ArXiv e-prints*, `arXiv:1310.7616`, Oct. 2013.

[16] "Power Systems Test Case Archive." [Online]. Available: `http://www.ee.washington.edu/research/pstca/`

[17] "Accuracy of Digital Electricity Meters," Electric Power Research Istitute white paper, May 2010.

[8] The false alarm rate of the bad data detector is set to be 0.04 throughout all the simulations.