Impact of Data Quality on Real-Time Locational Marginal Price

Liyan Jia, Jinsub Kim, Robert J. Thomas, *Life Fellow, IEEE*, and Lang Tong, *Fellow, IEEE*

***Abstract*—The problem of characterizing impacts of data quality on real-time locational marginal price (LMP) is considered. Because the real-time LMP is computed from the estimated network topology and system state, bad data that cause errors in topology processing and state estimation affect real-time LMP. It is shown that the power system state space is partitioned into price regions of convex polytopes. Under different bad data models, the worst case impacts of bad data on real-time LMP are analyzed. Numerical simulations are used to illustrate worst case performance for IEEE-14 and IEEE-118 networks.**

***Index Terms*—Bad data detection, cyber security of smart grid, locational marginal price (LMP), power system state estimation, real-time market.**

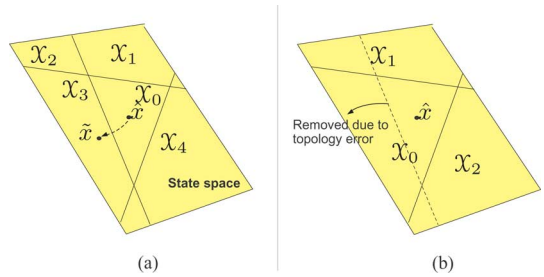


Fig. 1. Change of real-time LMPs due to bad data. (a) Bad meter data. (b) Topology error.

I. Introduction

THE deregulated electricity market has two interconnected components. The day-ahead market determines the locational marginal price (LMP) based on the dual variables of the optimal power flow (OPF) solution [1], [2], given generator offers, demand forecast, system topology, and security constraints. The calculation of LMP in the day-ahead market does not depend on the actual system operation. In the real-time market, on the other hand, an ex-post formulation is often used (e.g., by PJM and ISO-New England [3]) to calculate the real-time LMP by solving an incremental OPF problem. The LMPs in the day-ahead and the real-time markets are combined in the final clearing and settlement processes.

The real-time LMP is a function of data collected by the supervisory control and data acquisition (SCADA) system. Therefore, anomalies in data, if undetected, will affect prices in the real-time market. While the control center employs a bad data detector to "clean" the real-time measurements, miss detections and false alarms will occur inevitably. The increasing reliance on the cyber system also comes with the risk that malicious data may be injected by an adversary to affect system and real-time market operations. An intelligent adversary can carefully design a data attack to avoid detection by the bad data detector.

Manuscript received December 20, 2012; revised January 23, 2013, May 19, 2013, and August 19, 2013; accepted October 02, 2013. Date of publication November 07, 2013; date of current version February 14, 2014. This work was supported in part by a grant under the DoE CERTS program, the NSFunder Grant CNS-1135844, and a PSERC grant. Part of this work was presented at HICSS 2012 and PES General Meeting 2012. Paper no. TPWRS-01391-2012.

The authors are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: lj92@cornell.edu; jk752@cornell.edu; rjt1@cornell.edu; ltong@cornell.edu).

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

Digital Object Identifier 10.1109/TPWRS.2013.2286992

Regardless of the source of data errors, it is of significant value to assess potential impacts of data quality on the real-time market, especially when a smart grid may in the future deploy demand response based on real-time LMP. To this end, we are interested in characterizing the impact of *worst case data errors* on the real-time LMP. The focus on the worst case also reflects the lack of an accurate model of bad data and our desire to include the possibility of data attacks.

A. Summary of Results and Organization

We aim to characterize the worst effects of data corruption on real-time LMP. By "worst", we mean the maximum perturbation of real-time LMP caused by bad or malicious data, when a fixed set of data is subject to corruption. The complete characterization of worst data impact, however, is not computationally tractable. Our goal here is to develop an optimization based approach to search for *locally worst data* by restricting the network congestion to a set of lines prone to congestion. We then apply computationally tractable (greedy search) algorithms to find the worst data and evaluate the effects of worst data by simulations.

In characterizing the relation between data and real-time LMP, we first present a geometric characterization of the real-time LMP. In particular, we show that the state space of the power system is partitioned into polytope price regions, as illustrated in Fig. 1(a), where each polytope is associated with a unique real-time LMP vector, and the price region X_i is defined by a particular set of congested lines that determine the boundaries of the price region.

Two types of bad data are considered in this paper. One is the bad data associated with meter measurements such as the branch power flows in the network. Such bad data will cause errors in state estimation, possibly perturbing, as an example, the correct state estimate $\hat{x}$ in X_0 to $\tilde{x}$ in X_3 [as shown in Fig. 1(a)]. The analysis of the worst case data then corresponds to finding the

worst measurement error such that it perturbs the correct state estimation to the worst price region.

The second type of bad data, one that has not been carefully studied in the context of LMP in the literature, is error in digital measurements such as switch or breaker states. Such errors lead directly to topology errors therefore causing a change in the polytope structure as illustrated in Fig. 1(b). In this case, even if the estimated system state changes little, the prices associated with each region change, sometimes quite significantly.

Before characterizing impacts of bad meter data on LMP, we need to construct appropriate models for bad data. To this end, we propose three increasingly more powerful bad data models based on the dependencies on real-time system measurements: state independent bad data, partially adaptive bad data, and fully adaptive bad data.

In studying the worst case performance, we adopt a widely used approach that casts the problem as one involving an adversary whose goal is to make the system performance as poor as possible. The approach of finding the worst data is equivalent to finding the optimal strategy of an attacker who tries to perturb the real-time LMP and avoid being detected at the same time. By giving the adversary more information about the network state and endowing him with the ability to change data, we are able to capture the worst case performance, sometimes exactly and sometimes as bounds on performance.

Finally, we perform simulation studies using the IEEE-14 and IEEE-118 networks. We observe that bad data independent of the system state seems to have limited impact on real-time LMPs, and greater price perturbations can be achieved by state dependent bad data. The results also demonstrate that the real-time LMPs are subject to much larger perturbation if bad topology data are present in addition to bad meter data.

While substantial price changes can be realized for small networks by the worst meter data, as the size of network grows while the measurement redundancy rate remains the same, the influence of worst meter data on LMP is reduced. However, larger system actually gives more possibilities for the bad topology data to perturb the real-time LMP more significantly.

Our simulation results also show a degree of robustness provided by the *nonlinear state estimator*. While there have been many studies on data injection attacks based on DC models, very few consider the fact that the control center typically employs the nonlinear WLS state estimator under the AC model. Our simulation shows that effects of bad analog data designed based on DC model may be mitigated by the nonlinear estimator whereas bad topology data coupled with bad analog data can have greater impacts on LMP.

The rest of the paper is organized as follows. Section II briefly describes a model of real-time LMP and introduces its geometric characterization in the state space of the power system. Section III establishes the bad data models and summarizes state estimation and bad data detection procedures at the control center. In Section IV, a metric of impact on real-time LMP caused by bad meter data is introduced. We then discuss the algorithms of finding worst case bad meter data vector in terms of real-time price perturbation under the three different bad data models. Section V considers the effect of bad topology data on real-time LMP. Finally, in Section VI, simulation results are presented based on IEEE-14 and IEEE-118 networks.

B. Related Work

Effects of bad data on power system have been studied extensively in the past; see [4]–[6]. Finding the worst case bad data is naturally connected with the problem of malicious data. In this context, the results presented in this paper can be viewed as one of analyzing the impact of the worst (malicious) data attack.

In a seminal paper by Liu, Ning, and Reiter [7], the authors first illustrated the possibility that, by compromising enough number of meters, an adversary can perturb the state estimate arbitrarily in some subspace of the state space without being detected by any bad data detector. Such attacks are referred to as strong attacks. It was shown by Kosut *et al.* [8] that the condition for the existence of such undetectable attacks is equivalent to the classical notion of network observability.

When the adversary can only inject malicious data from a small number of meters, strong attacks do not exist, and any injected malicious data can be detected with some probability. Such attacks are referred to as weak attacks [8]. In order to affect the system operation in some meaningful way, the adversary has to risk being detected by the control center. The impacts of weak attack on power system are not well understood because the detection of such bad data is probabilistic. Our results are perhaps the first to quantify such impacts. Most related research works focused on DC model and linear estimator while only few have addressed the nonlinearity effect [9], [10].

It is well recognized that bad data can also cause topology errors [11], [12], and techniques have been developed to detect topology errors. For instance, the residue vector from state estimation was analyzed for topology error detection [12], [11], [13]. Monticelli [14] introduced the idea of generalized state estimation where, roughly speaking, the topology that fits the meter measurements best is chosen as the topology estimate. The impacts of topology errors on electricity market have not been reported in the literature, and this paper aims to bridge this gap.

The effect of data quality on real-time market was first considered in [15] and [16]. In [16], the authors presented the financial risks induced by the data perturbation and proposed a heuristic technique for finding a case where price change happens. While there are similarities between this paper and [16], several significant differences exist: 1) This paper focuses on finding the worst case, not only a feasible case. 2) This paper considers a more general class of bad data where bad data may depend dynamically on the actual system measurements rather than static. 3) This paper considers a broader range of bad data that also include bad topology data, and our evaluations are based on the AC network model and the presence of nonlinear state estimator.

II. Structures of Real-Time LMP

In this section, we present first a model for the computation of real-time LMP. While ISOs have somewhat different methods

of computing real-time LMP, they share the same two-settlement architecture and similar ways of using real-time measurements. In the following, we will use a simplified ex-post real-time market model, adopted by PJM, ISO New England, and other ISOs [17], [3]. We view this model as a convenient mathematical abstraction that captures the essential components of the real-time LMP calculation. For this reason, our results should be interpreted within the specified setup. Our purpose is not to include all details; we aim to capture the essential features.

In real-time, in order to monitor and operate the system, the control center will calculate the estimated system conditions (including bus voltages, branch flows, generation, and demand) based on real-time measurements. We call a branch congested if the estimated flow is larger than or equal to the security limit. The congestion pattern is defined as the set of all congested lines, denoted as $\hat{C}$. Note that we use hat (e.g., $\hat{C}$) to denote quantities or sets that are estimated based on real-time measurements. Details of state estimation and bad data detection are discussed in Section III-B.

One important usage of state estimation is calculating the real-time LMP. Given the estimated congestion pattern $\hat{C}$, the following linear program is solved to find the incremental OPF dispatch and associated real-time LMP, $\hat{\lambda} = (\hat{\lambda}_i)$ [17]:

$$\begin{aligned} \text{minimize} \quad & \sum c_i^{\mathrm{G}} \Delta p_i - \sum c_j^{\mathrm{L}} \Delta d_j \\ \text{subjcet to} \quad & \sum \Delta p_i = \sum \Delta d_j \\ & \Delta p_i^{\min} \le \Delta p_i \le \Delta p_i^{\max} \\ & \Delta d_j^{\min} \le \Delta d_j \le \Delta d_j^{\max} \\ & \sum_i A_{ki} \Delta p_i - \sum_j A_{kj} \Delta d_j \le 0, \text{ for all } k \in \hat{C} \end{aligned} \tag{1}$$

where $\Delta d = (\Delta d_j)$ is the vector of incremental dispatchable load, $\Delta p = (\Delta p_i)$ the vector of incremental generation dispatch, $c^{\mathrm{G}} = (c_i^{\mathrm{G}})$ and $c^{\mathrm{L}} = (c_j^{\mathrm{L}})$ the corresponding real-time marginal cost of generations and dispatchable loads, $\Delta p_i^{\min}$ and $\Delta p_i^{\max}$ the lower and upper bounds for incremental generation dispatch, $\Delta d_i^{\min}$ and $\Delta d_i^{\max}$ the lower and upper bounds for incremental dispatchable load, and A_{ki} the sensitivity of branch flow on branch k with respect to the power injection at bus i.

The real-time LMP at bus i is defined as the overall cost increase when one unit of extra load is added at bus i, which is calculated as

$$\hat{\lambda}_i = \eta - \sum_{k \in \hat{C}} A_{ki} \mu_k \tag{2}$$

where η is the dual variable for the load-generation equality constraint, and μ_k is the dual variable corresponding to the line flow constraint in (1).

Note that in practice, the control center may use the ex-ante congestion pattern, which is obtained by running a 5 min ahead security-constrained economic dispatch with the state estimation results and the forecasted loads (for the next five-minute interval) and choosing the lines congested at the dispatch solution [17], [3]. However, to avoid the complication due to ex-ante dispatch calculation, we assume that real-time pricing employs the estimated congestion pattern $\hat{C}$ obtained from state estimation results. By doing so, we attempt to find direct relations among bad data, the state estimate, and real-time LMPs. Notice that once the congestion pattern $\hat{C}$ is determined, the whole incremental OPF problem (1) no longer depends on the measurement data.

Under the DC model, the power system state, x, is defined as the vector of voltage phases, except the phase on the reference bus. The power flow vector f is a function of the system state x

$$f = Fx \tag{3}$$

where F is the sensitivity matrix of branch flows with respect to the system state.

Assume the system has $n+1$ buses. Then, $x \in X = [-\pi, \pi]^n$, where X represents the state space. Any system state corresponds to a unique point in X. From (3), the branch flow f is determined by the system state x. Comparing the flows with the flow limits, we obtain the congestion pattern associated with this state. Hence, each point in the state space corresponds to a particular congestion pattern.

We note that the above expression in (2) appears earlier in [1] where the role of congestion state in LMP computation was discussed. In this paper, our objective is to make explicit the connection between data and LMP. We therefore need a linkage between data and congestion. To this end, we note that the power system state, the congestion state, and LMP form a Markov chain, which led to a geometric characterization of LMP on the power system state space, as shown in the following theorem.

Theorem 1 (Price Partition of the State Space): Assume that the LMP exists for every possible congestion pattern.[1] Then, the state space X is partitioned into a set of polytopes $\{X_i\}$ where the interior of each X_i is associated with a unique congestion pattern C_i and a real-time LMP vector. Each boundary hyperplane of X_i is defined by a single transmission line.

Proof: For a particular congestion pattern C defined by a set of congested lines, the set of states that gives C is given by

$$X_i \triangleq \{x : F_i.x \ge T_i^{\max} \forall i \in C, F_j.x < T_j^{\max} \forall j \notin C\}$$

where $F_i.$ is the ith row of F [see (3)], and $T_j^{\max}$ the flow limit on branch j. Since X_i is defined by the intersection of a set of half spaces, it is a polytope.

Given an estimated congestion pattern $\hat{C}$, the envelop theorem [18] implies that for any optimal primal solution and dual solution of (1) that satisfy the KKT conditions, (2) always gives the derivative of the optimal objective value with respect to the demand at each bus, which we assume exists, i.e., each congestion pattern is associated with a unique real-time LMP vector λ. Hence, all states with the same congestion pattern share the same real-time LMP, which means each polytope X_i in X corresponds to a unique real-time LMP vector. ■

Theorem 1 characterizes succinctly the relationship between the system state and LMP. As illustrated in Fig. 1(a), if bad data are to alter the LMP in real-time, the size of the bad data has to be sufficiently large so that the state estimate at the control center is moved to a different price region from the true system state.

[1]This is equivalent to assuming that the derivative of the optimal value of (1) with respect to demand at each bus exists.

On the other hand, if some lines are erroneously removed from or added to the correct topology, as illustrated in Fig. 1(b), it affects the LMP calculation in three ways.[2] First, the state estimate is perturbed since the control center employs an incorrect topology in state estimation. Secondly, the price partition of the state space changes due to the errors in topology information. Third, the shift matrix A in (1), which is a function of topology, changes thereby altering prices attached to each price region.

III. Data Model and State Estimation

A. Bad Data Model

1) Meter Data: In order to monitor the system, various meter measurements are collected in real time, such as power injections, branch flows, voltage magnitudes, and phasors, denoted by a vector $z \in \mathbb{R}^m$.[3] If there exists bad data a among the measurements, the measurements with bad data, denoted by z_a, can be expressed as a function of the system states x

$$z_a = z + a = h(x) + w + a, \quad a \in A \tag{4}$$

where w represents the random measurement noise.

We make a distinction here between the measurement noise and bad data; the former accounts for random noise independently distributed across all meters whereas the latter represents the perturbation caused by bad or malicious data. We assume no specific pattern for bad data except that they do not happen everywhere. We assume that bad data can only happen in a subset of the measurements, S. We call S as set of suspectable meters, which means the meter readings with in S may subject to corruption. If the cardinality of S is k, the feasible set of bad data a is a k-dimensional subspace, denoted as $A = \{a : a_i = 0 \text{ for all } i \notin S\}$.

We will consider three bad data models with increasing power of affecting state estimates.

M1. *State independent bad data*: This type of bad data is independent of real-time measurements. Such bad data may be the replacement of missing measurements.

M2. *Partially adaptive bad data*: This type of bad data may arise from the so-called man in the middle (MiM) attack where an adversary intercepts the meter data and alter the data based on what he has observed. Such bad data can adapt to the system operating state.

M3. *Fully adaptive bad data*: This is the most powerful type of bad data, constructed based on the actual measurement $z = h(x) + w$.

Note that M3 is in general not realistic. Our purpose of considering this model is to use it as a conservative proxy to obtain performance bounds for the impact of worst case data.

[2]In addition to these, the change in topology will affect contingency analysis. Such effect will appear as changes in contingency constraints in real-time LMP calculation (1) [17]. However, dealing with contingency constraints will significantly complicate our analysis and possibly obscure the more direct link between bad data and real-time LMP. Hence, we consider only line congestion constraints in (1).

[3]Notice here both conventional measurements and PMU measurements can be incorporated. Although PMU data seem to have more direct impact on state estimation and real-time LMP calculation, we won't differentiate the types of measurements in the following discussion.

We assume herein a DC model in which the measurement function $h(\cdot)$ in (4) is linear. Specifically

$$z_a = Hx + w + a, \quad a \in A \tag{5}$$

where H is the measurement matrix. Such a DC model, while widely used in the literature, may only be a crude approximation of the real power system. By making such a simplifying assumption and acknowledging its weaknesses, we hope to obtain tractable solutions in searching for worst case scenarios. It is important to note that, although the worst case scenarios are derived from the DC model, we carry out simulations using the actual nonlinear system model.

2) Topology Data: Topology data are represented by a binary vector $s \in \{0,1\}^l$, where each entry of s represents the state of a line breaker (0 for open and 1 for closed). The bad topology data is modeled as

$$s_b = s + b (\mathrm{mod}\ 2), \quad b \in B \tag{6}$$

where $B \subset \{0,1\}^l$ is the set of possible bad data. When bad data are present, the topology processor will generate the topology estimate corresponding to s_b, and this incorrect topology estimate will be passed to the following operations unless detected by the bad data detector.

B. State Estimation

We assume that the control center employs the standard weighted least squares (WLS) state estimator. Under DC model

$$\hat{x} = \arg\min_x (z - Hx)^{\mathrm{T}} R^{-1} (z - Hx) = Kz \tag{7}$$

where R is the covariance matrix of measurement noise w, and $K \triangleq (H^{\mathrm{T}} R^{-1} H)^{-1} H^{\mathrm{T}} R^{-1}$.

If the noise w is Gaussian, the WLS estimator is also the maximum likelihood estimate (MLE) of state x. By the invariant property of MLE, from (3), the maximum likelihood estimate of the branch flows is calculated as

$$\hat{f} = F\hat{x} = FKz. \tag{8}$$

The congestion pattern used in real-time LMP calculation (1) is directly from state estimation and consists of all the estimated branch flows which are larger than or equal to the branch flow limits, i.e.,

$$\hat{C} = \left\{ j : \hat{f}_j \geq T_j^{\max} \right\} \tag{9}$$

where $T_j^{\max}$ is the flow limit on branch j.

In the presence of bad meter data a, the meter measurements collected by control center is actually $z_a = Hx + w + a$. By using z_a, the WLS state estimate is

$$\hat{x}_a = Kz_a = \hat{x}^* + Ka \tag{10}$$

where $\hat{x}^* = Kz$ is the "correct" state estimate without the presence of the bad data (i.e., $a = 0$).

Equation (10) shows that the effect of bad data on state estimation is linear. However, because a is confined in a k-dimensional subspace A, the perturbation on the actual system state is limited to a certain direction.

When bad data exist both in meter and topology data, the control center uses a wrong measurement matrix $\bar{H}$, corresponding to the altered topology data, and the altered meter data z_a. Then, the WLS state estimate becomes

$$\hat{x}_a = \bar{K} z_a = \bar{K} z + \bar{K} a \tag{11}$$

where $\bar{K} \triangleq (\bar{H}^{\mathrm{T}} R^{-1} \bar{H})^{-1} \bar{H}^{\mathrm{T}} R^{-1}$. Note that unlike the linear effect of bad meter data, bad topology data affects the state estimate by altering the measurement matrix H to $\bar{H}$.

C. Bad Data Detection

The control center uses bad data detection to minimize the impact of bad data. Here, we assume a standard bad data detection used in practice, the $J(\hat{x})$-detector in [5]. In particular, the $J(\hat{x})$-detector performs the test on the residue error, $r \triangleq z - H\hat{x}$, based on the state estimate $\hat{x}$. From the WLS state estimate (7), we have

$$r = (I - H(H^{\mathrm{T}} R^{-1} H)^{-1} H^{\mathrm{T}} R^{-1}) z = Uz \tag{12}$$

where $U \triangleq (I - H(H^{\mathrm{T}} R^{-1} H)^{-1} H^{\mathrm{T}} R^{-1})$.

The $J(\hat{x})$-detector is a threshold detector defined by

$$r^{\mathrm{T}} R^{-1} r = z^{\mathrm{T}} W z \underset{\text{gooddata}}{\overset{\text{baddata}}{\gtrless}} \tau \tag{13}$$

where τ is the threshold calculated from a prescribed false alarm probability, and $W \triangleq U^{\mathrm{T}} R^{-1} U$. When the measurement data fail to pass the bad data test, the control center declares the existence of bad data and takes corresponding actions to identify and remove the bad data.

In this paper, we are interested in those cases when bad data are present while the $J(\hat{x})$-detector fails to detect them.

IV. Impact of Bad Data on LMP

In this section, we examine the impact of bad data on LMP, assuming that the topology estimate of the network is correct.

One thing to notice is that in searching for the "worst" case, we take the perspective of the control center, not that of the attacker. In particular, we look for the worst congestion pattern for the LMP computation, even if this particular congestion pattern is difficult for the attacker to discover. So the focus here is not how easy it is for an attacker to find a locally worst congestion pattern; it is how much such a congestion pattern affects the LMP.

A. Average Relative Price Perturbation

In order to quantify the effect of bad data on real-time price, we need to first define the metric to measure the effect. We define the *relative price perturbation* (RPP) as the expected percentage price perturbation caused by bad data. Given that LMP varies at different buses, RPP also varies at different locations.

Let z_a be the data received at the control center and $\lambda_i(z_a)$ the LMP at bus i. The RPP at bus i is a function of bad data a, given by

$$\mathrm{RPP}_i(a) = \mathbb{E}\left(\left| \frac{\lambda_i(z_a) - \lambda_i(z)}{\lambda_i(z)} \right| \right) \tag{14}$$

where the expectation is over random state and measurement noise.

To measure the system-wide price perturbation, we define the *average relative price perturbation* (ARPP) by

$$\mathrm{ARPP}(a) = \frac{1}{n+1} \sum_i \mathrm{RPP}_i(a) \tag{15}$$

where $n+1$ is the number of buses in the system.

The worst case analysis to be followed can be used for other metrics (e.g., price increase ratios or price decrease ratios, which are closely related to the market participants' gain or loss). Similar results can be showed following the same strategies. However, the comparison among different metrics is beyond the scope of this paper.

B. Worst ARPP Under State Independent Bad Data Model

First, we consider the state independent bad data model (M1) given in Section III-A. In this model, the bad data are independent of real-time measurements.

In constructing the state independent worst data, it is useful to incorporate prior information about the state. To this end, we assume that system state follows a Gaussian distribution with mean x_0, covariance matrix Σ_x. Typically, we choose x_0 as the day-ahead dispatch since the nominal system state in real-time varies around its day-ahead projection.

In the presence of bad data a, the expected state estimate and branch flow estimate on branch i are given by

$$\mathbb{E}[\hat{x}] = x_0 + Ka \tag{16}$$

$$\mathbb{E}[f_i] = F_i . \mathbb{E}[\hat{x}] = F_i . x_0 + F_i . Ka \tag{17}$$

where $F_i.$ is the corresponding row of branch i in F.

Our strategy is to make this expected state estimate into the region with the largest price perturbation among all the possible regions, $\hat{C}^*$. From (9), this means making all the expected branch flows satisfy the boundary condition of $\hat{C}^*$

$$\begin{aligned} \mathbb{E}[f_i] &\geq T_i^{\max} \quad \text{for } i \in \hat{C}^* \\ \mathbb{E}[f_i] &\leq T_j^{\max} \quad \text{for } j \notin \hat{C}^*. \end{aligned} \tag{18}$$

However, due to the uncertainty (from both system state x and measurement noise w), the actual estimated state after attack, $\hat{x}$, may be different from $\mathbb{E}[\hat{x}]$. Therefore, we want to make $\mathbb{E}[\hat{x}]$ at the "center" of the desired price region, i.e., maximizing the shortest distance from $\mathbb{E}[\hat{x}]$ to the boundaries of the polytope price regions while still holding the boundary constraints. The shortest distance can be calculated as

$$\beta = \min\{\tilde{\beta} : |\mathbb{E}[f_i] - T^{\max}| \geq \tilde{\beta} \text{ for all i}\}. \tag{19}$$

However, the existence of bad data detector prevents the bad data vector a from being arbitrarily large. According to (12), the weighted squared residue with a is

$$r^{\mathrm{T}} R^{-1} r = z_a^{\mathrm{T}} W z_a = (w+a)^{\mathrm{T}} W (w+a) \tag{20}$$

since $WHx = 0$

Heuristically, since w has zero mean, the term $a^{\mathrm{T}} W a$ can be used to quantify the effect of data perturbation on estimation residue. Then we use $a^{\mathrm{T}} W a \leq \epsilon$ to control the detection probability in the following optimization.

Therefore, for a specific congestion pattern $\hat{C}$, the adversary will solve the following optimization problem to move the state estimate to the "center" of the price region $\hat{C}$ and keeping the detection probability low:

$$\begin{aligned} \max_{a\in A,\tilde{\beta}\geq 0} \quad & \tilde{\beta} \\ \text{subject to} \quad & \mathbb{E}[f_i] - \tilde{\beta} \geq T_i^{\max}, i \in \hat{C} \\ & \mathbb{E}[f_i] + \tilde{\beta} < T_j^{\max}, j \notin \hat{C} \\ & a^{\mathrm{T}} W a \leq \epsilon \end{aligned} \tag{21}$$

which is a convex program that can be solved easily in practice. We call a region $\hat{C}$ *feasible* if it makes problem (21) feasible.

Among all the feasible congestion patterns, the worst region $\hat{C}^*$ is chosen as the one giving the largest ARPP:

$$\hat{C}^* = \arg\max_{\hat{C}\in\Gamma} |\tilde{\lambda}_i - \lambda_i(\hat{C})| \tag{22}$$

where $\tilde{\lambda}_i$ is the LMP at bus i if the x_0 is the system state, and Γ the set of all the feasible congestion patterns. Hence, the worst case constant bad data vector is the solution to optimization problem (21) by setting the congestion pattern as $\hat{C}^*$.

C. Worst ARPP Under Partially Adaptive Bad Data

For bad data model M2, only part of the measurement values in real-time are known to the adversary, denoted as z_{o}. The adversary has to first make an estimation of the system state from the observation and prior distribution, then make the attack decision based on the estimation result.

Without the presence of bad data vector, i.e., $a = 0$, the system (5) gives

$$z_{\mathrm{o}} = H_{\mathrm{o}} x + w_{\mathrm{o}} \tag{23}$$

where H_{o} is the rows of H corresponding to the observed measurements and w_{o} the corresponding part in the measurement noise w.

The minimum mean square error (MMSE) estimate of x given z_{o} is given by the conditional mean

$$\mathbb{E}(x|z_{\mathrm{o}}) = x_0 + \Sigma_x H_{\mathrm{o}}^{\mathrm{T}} \left(H_{\mathrm{o}} \Sigma_x H_{\mathrm{o}}^{\mathrm{T}}\right)^{-1} (z_{\mathrm{o}} - H_{\mathrm{o}} x_0). \tag{24}$$

Then, the flow estimate on branch i after attack is

$$\mathbb{E}[f_i|z_{\mathrm{o}}] = F_i.\mathbb{E}[\hat{x}|z_{\mathrm{o}}]. \tag{25}$$

Still, we want to move the estimation of state to the "center". On the other hand, the expected measurement value $\mathbb{E}[z_a|z_{\mathrm{o}}] = H\mathbb{E}[\hat{z}|z_{\mathrm{o}}] + a$. Again, we need a pre-designed parameter ϵ to control the detection probability. Therefore, the solution to the following optimization problem is the best attack given congestion pattern A

$$\begin{aligned} \max_{a\in A,\tilde{\beta}\geq 0} \quad & \tilde{\beta} \\ \text{subject to} \quad & \mathbb{E}[f_i|z_{\mathrm{o}}] - \tilde{\beta} \geq T_i^{\max}, i \in \hat{C} \\ & \mathbb{E}[f_i|z_{\mathrm{o}}] + \tilde{\beta} < T_j^{\max}, j \notin \hat{C} \\ & (H\mathbb{E}[z_a|z_{\mathrm{o}}]^{\mathrm{T}}) W (H\mathbb{E}[z_a|z_{\mathrm{o}}]) \leq \epsilon. \end{aligned} \tag{26}$$

This problem is also a convex optimization problem, which can be easily solved. Among all the $\hat{C}$'s which make the above problem feasible, we choose the one with the largest price perturbation, denoted as $\hat{C}^*$. The solution to problem (26) with $\hat{C}^*$ as the congestion pattern is the worst bad data vector.

D. Worst ARPP Under Fully Adaptive Bad Data

Finally, we consider the bad data model M3, in which the whole set of measurements z is known to the adversary. The worst bad data vector depends on the value of z. Different from the previous two models, with bad data vector a, the estimated state is deterministic without uncertainty. In particular

$$\hat{x} = Kz + Ka \tag{27}$$

and the estimated flow on branch i after attack is also deterministic

$$\hat{f}_i = F_i.\hat{x} = F_i.Kz + F_i.Ka. \tag{28}$$

Similar to the previous two models, congestion pattern is called feasible if there exists some bad data vector a to make the following conditions satisfied:

$$\begin{aligned} & \hat{f}_i \geq T_i^{\max}, i \in \hat{C} \\ & \hat{f}_i < T_j^{\max}, j \notin \hat{C} \\ & (z+a)^{\mathrm{T}} W (z+a) \leq \tau, \quad a \in A. \end{aligned} \tag{29}$$

Among all the feasible congestion patterns, we choose the one with the largest price perturbation, $\hat{C}^*$. Any bad data vector a satisfying condition (29) can serve as the worst fully adaptive bad data.

E. Greedy Heuristic

The strategies presented above are based on the exhaustive search over all possible congestion patterns. Such approaches are not scalable for large networks with a large number of possible congestion patterns. We now present a greedy heuristic approach aimed at reducing computation cost. In particular, we develop a gradient like algorithm that searches among a set of likely congestion patterns.

First, we restrict ourselves to the set of lines that are close to their respective flow limits and look for bad data that will affect the congestion pattern. The intuition is that it is unlikely that bad data can drive the system state sufficiently far without being detected by the bad data detector. In practice, the cardinality of such a set is usually very small compared with the systems size.

Second, we search for the worst data locally by changing one line in the congestion pattern at a time. Specifically, suppose that a congestion pattern is the current candidate for the worst data. Given a set of candidate lines that are prone to congestions, we search locally by flipping one line at a time from the congested state to the un-congested state and vice versa. If no improvement can be made, the algorithm stops. Otherwise, the algorithm updates the current "worst congestion pattern" and continue. The effectiveness of this greedy heuristic is tested in Section VI-C.

V. Bad Topology Data on LMP

So far, we have considered bad data in the analog measurements. In this section, we include the bad *topology* data, and describe another bad data model.

We represent the network topology by a directed graph $G = (V, E)$ where each $i \in V$ denotes a bus and each $(i, j) \in E$ denotes a *connected* transmission line. For each physical transmission line (e.g., a physical line between i and j), we assign an arbitrary direction [e.g., (i, j)] for the line, and (i, j) is in E if and only if bus i and bus j are connected.

Bad data may appear in both analog measurements and digital (e.g., breaker status) data, as described in Section III-A:

$$z_a = z + a = (Hx + w) + a,\ a \in A,$$
$$s_b = s + b \ (\mathrm{mod}\ 2),\ b \in B. \tag{30}$$

As in Section IV, we employ the adversary model to describe the worst case. The adversary alters s to s_b by adding b from the set of feasible attack vectors $B \subset \{0, 1\}^l$ such that the topology processor produces the "target" topology $\bar{G}$ as the topology estimate. In addition, the adversary modifies z by adding $a \in A$ such that z_a looks consistent with $\bar{G}$.

In this section, we focus on the worst case when the adversary is able to alter the network topology without changing the state estimate.[4] We also require that such bad data are generated by an adversary causing undetectable topology change, i.e., the bad data escape the system bad data detection. For the worst case analysis, we will maximize the LMP perturbation among the attacks within this specific class. Even though this approach is suboptimal, the simulation results in Section VI demonstrate that the resulting LMP perturbation is much greater than the worst case of the bad meter data.

Suppose the adversary wants to mislead the control center with the target topology $\bar{G} = (V, \bar{E})$, a topology obtained by *removing*[5] a set of transmission lines E_Δ in G (i.e., $\bar{E} = E \setminus E_\Delta$). We assume that the system with $\bar{G}$ is observable: i.e., the corresponding measurement matrix $\bar{H}$ has full column rank.[6]

The adversarial data modification aimed at perturbing the topology estimate at the control center was studied in [19]. Suppose that the adversary changes the breaker status such that the target topology $\bar{G} = (V, \bar{E})$ is observed at the control center. Simultaneously, if the adversary introduces bad data $a = \bar{H}x - Hx$, then

$$z_a = Hx + a + w = \bar{H}x + w \tag{31}$$

which means that the meter data received at the control center are completely consistent with the model generated from $\bar{G}$. Thus, any bad data detector will not be effective.

[4]In general, the adversary can design the worst data to affect both the state estimate and network topology. It is, however, much more difficult to make such attack undetectable.

[5]Line addition by the adversary is also possible. However, compared to line removal attacks, line addition attacks require the adversary to observe a much larger set of meter measurements to design undetectable attacks. In addition, the number of necessary modifications in breaker data is also much larger: to make a line appear to be connected, the adversary should make all the breakers on the line appear to be closed. Please see [19] for the detail.

[6]Without observability, the system may not proceed to state estimation and real-time pricing. Hence, for the adversary to affect pricing, the system with the target topology has to be observable.

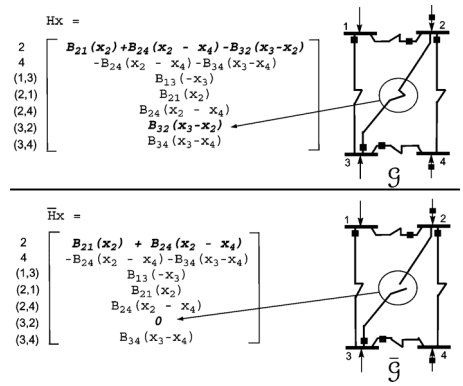


Fig. 2. Hx and $\bar{H}x$: Each row is marked by the corresponding meter (i for injection at i and (i, j) for flow from i to j).

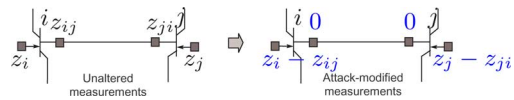


Fig. 3. Attack modifies local measurements around the line (i, j) in E_Δ.

It is of course not obvious how to produce the bad data a, especially when the adversary can only modify a limited number of measurements, and it may not have access to the entire state vector x. Fortunately, it turns out that a can be generated by observing only a few entries in z without requiring global information (such as the state vector x) [19].

A key observation is that Hx and $\bar{H}x$ differ only in a few entries corresponding to the modified topology (lines in E_Δ) as illustrated in Fig. 2. Consider first the noiseless case. Let z_{ij} denote the entry of z corresponding to the flow measurement from i to j. As hinted from Fig. 2, it can be easily seen that $\bar{H}x - Hx$ has the following sparse structure [19]:

$$\bar{H}x - Hx = -\sum_{(i,j)\in E_\Delta} \alpha_{ij} m_{(i,j)} \tag{32}$$

where $\alpha_{ij} \in \mathbb{R}$ denotes the line flow from i to j when the line is connected and the system state is x, and $m_{(i,j)}$ is the column of the measurement-to-branch incidence matrix, that corresponds to (i, j): i.e., $m_{(i,j)}$ is an m-dimensional vector with 1 at the entries corresponding to the flow from i to j and the injection at i, and -1 at the entries for the flow from j to i and the injection at j, and 0 at all other entries. Absence of noise implies that $z_{ij} = \alpha_{ij}$, which leads to

$$\bar{H}x - Hx = -\sum_{(i,j)\in E_\Delta} z_{ij} m_{(i,j)}. \tag{33}$$

With (33) in mind, one can see that setting $a = \bar{H}x - Hx$ and adding a to z is equivalent to the following simple procedure: as described in Fig. 3, for each (i, j) in E_Δ:

1) Subtract z_{ij} and z_{ji} from z_i and z_j, respectively;
2) Set z_{ij} and z_{ji} to be 0

where z_i is the entry of z corresponding to the injection measurement at bus i.

When measurement noise is present (i.e., $z = Hx + w$), the idea of the attack is still the same: to make a approximate $\bar{H}x - Hx$ so that z_a is close to $\bar{H}x + w$. Since $z_{ij} = \alpha_{ij} + w_{ij}$, z_{ij} is an unbiased estimate of α_{ij} for each $(i, j) \in E_\Delta$, and this implies that $-\sum_{(i,j)\in E_\Delta} z_{ij} m_{(i,j)}$ is an unbiased estimate

of $-\sum_{(i,j)\in E_\Delta} \alpha_{ij} m_{(i,j)} = \bar{H}x - Hx$. Hence, we set a to be $-\sum_{(i,j)\in E_\Delta} z_{ij} m_{(i,j)}$, the same as in the noiseless setting, and the attack is executed by the same steps as above.

For launching this attack to modify the topology estimate from G to $\bar{G}$, the adversary should be able to 1) set b such that the topology processor produces $\bar{G}$ instead of G and 2) observe and modify z_{ij}, z_{ji}, z_i, and z_j for all $(i,j) \in E_\Delta$. The attack is feasible if and only if A and B contain the corresponding attack vectors.

To find the worst case LMP perturbation due to undetectable, state-preserving attacks, let F denote the set of feasible $\bar{G}$s, for which the attack can be launched with A and B. Among the feasible targets in F, we consider the best target topology that results in the maximum perturbation in real-time LMPs. If ARPP is used as a metric, the best target is chosen as

$$\bar{G}^*[z] = \arg\max_{\bar{G}\in F} \sum_i \left| \frac{\lambda_i(z;\bar{G}) - \lambda_i(z;G)}{\lambda_i(z;G)} \right| \qquad (34)$$

where $\lambda_i(z;\bar{G})$ denotes the real-time LMP at bus i when the attack with the target $\bar{G}$ is launched on z, and $\lambda_i(z;G)$ is the real-time LMP under no attack.

VI. Numerical Results

In this section, we demonstrate the impact of bad data on real-time LMPs with the numerical simulations on IEEE-14 and IEEE-118 systems. We conducted simulations in two different settings: the linear model with the DC state estimator and the nonlinear model with the AC state estimator. The former is usually employed in the literature for the ease of analysis whereas the latter represents the practical state estimator used in the real-world power system. In all simulations, the meter measurements consist of real power injections at all buses and real power flows (both directions) at all branches.

A. Linear Model With DC State Estimation

We first present the simulation results for the linear model with the DC state estimator. We modeled bus voltage magnitudes and phases as Gaussian random variables with the means equal to the day-ahead dispatched values and small standard deviations. In each Monte Carlo run, we generated a state realization from the statistical model, and the meter measurements were created by the DC model with Gaussian measurement noise. Once the measurements were created, bad data were added in the manners discussed in Sections IV and V. With the corrupted measurements, the control center executed the DC state estimation and the bad data test with the false alarm probability constraint 0.1. If the data passed the bad data test, real-time LMPs were evaluated based on the state estimation results. For IEEE-14 and IEEE-118 system, the network parameters[7] are available in [20].

[7]In addition to the network parameters given in [20], we used the following line limit and real-time offer parameters. In the IEEE-14 simulation, the generators at the buses 1, 2, 3, 6, and 8 had capacities 330, 140, 100, 100, and 100 MW and the real-time offers 15, 31, 30, 10, and 20 \$/MW. Lines (2, 3), (4, 5), and (6, 11) had line capacities 50, 50, and 20 MW, and other lines had no line limit. In the IEEE-118 simulation, the generators had generation costs arbitrarily selected from $\{20, 25, 30, 35, 40\$/\text{ MW}\}$ and generation capacities arbitrarily selected from $\{200, 250, 300, 350, 400\text{ MW}\}$. Total 16 lines had the line capacities arbitrarily selected from $\{70, 90, 110\text{ MW}\}$, and other lines had no line limit. To handle possible occurrence of price spikes, we set the upper and lower price caps as 500\$ MW and $-100\$/\text{ MW}$, respectively. Total 1000 Monte Carlo runs were executed for each case.

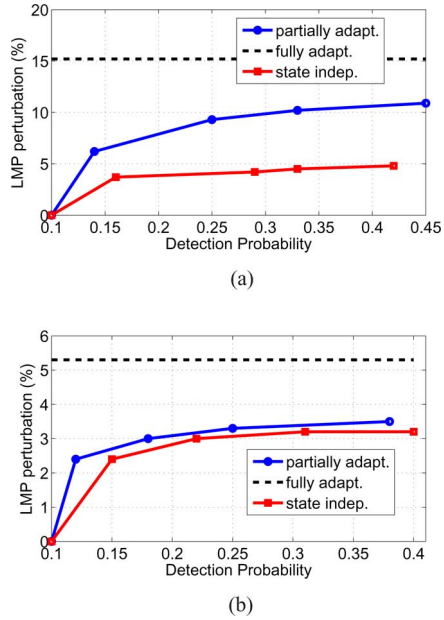


Fig. 4. Linear model: ARPP versus detection prob. (a) IEEE-14: ARPP of the worst topology data is 66.1%. (b) IEEE-118: ARPP of the worst topology data is 22.4%.

We used the number of meter data to be modified by the adversary as the metric for the attack effort. For the 14-bus system, in each Monte Carlo run, we randomly chose two lines, and the adversary was able to modify all the line flow meters on the lines and injection meters located at the ends of the lines. For the 118-bus system, we randomly chose three lines, and the adversary had control over the associated line and injection meters. Both state and topology attacks were set to control the same number of meter data[8] so that we can fairly compare their impacts on real-time LMPs. As for the meter data attack, we only considered the lines that are close to their flow limits (estimated flows under M1 and M2, or actual flows under M3) as candidates for congestion pattern search. The threshold is chosen as 10 MW in our simulation.

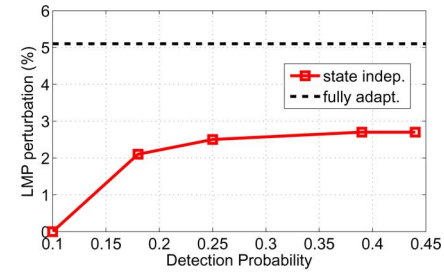
Fig. 4 is the plot of ARPPs[9] versus detection probabilities of bad data. They show that even when bad data were detected with low probability, ARPPs were large, especially for the fully adaptive bad meter data and the bad topology data.

Comparing ARPPs of the three bad meter data models, we observe that the adversary may significantly improve the perturbation amount by exploiting partial or all real-time meter data (for the partially adaptive case, the adversary observed a half of all meters.) It is worthy to point out that bad topology data result in much greater price perturbation than bad meter data.

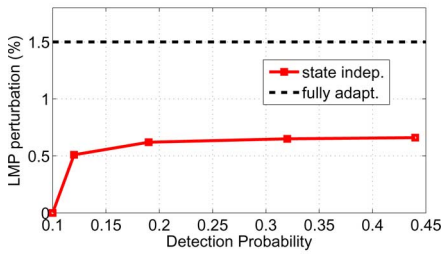
Recall the discussion in Sections II and V that bad topology data and bad meter data employ different price-perturbing mechanisms: bad topology data perturb real-time LMP by

[8]Topology attacks need to make few additional modifications on breaker state data such that the target lines appear to be disconnected to the topology processor. However, for simplicity, we do not take into account this additional effort.

[9]The detection probabilities for the fully adaptive bad meter data and the bad topology data cases were less than 0.1 in all the simulations. In the figures, we draw ARPPs of those cases as horizontal lines so that we can compare them with other cases.



(a)



(b)

Fig. 5. Nonlinear model: ARPP versus detection prob. (a) IEEE-14: ARPP of the worst topology data is 95.4%. (b) IEEE-118: ARPP of the worst topology data is 76.9%.

restructuring the price regions without perturbing the state estimate (the line-removal attack introduced in Section V does not perturb state estimate) whereas bad meter data perturb real-time LMP by simply moving the state estimate to a different price region. Therefore, the observation implies that restructuring the price regions has much greater impact on real-time LMP than merely perturbing the state estimate.

B. Nonlinear Model With AC State Estimation

The simulations with the nonlinear model intend to investigate the vulnerability of the real-world power system to the worst adversarial act, designed based on the linear model. The simulations were conducted on IEEE-14 and IEEE-118 systems in the same manner as the linear case except that we employed the nonlinear model and the AC state estimation.

Fig. 5 is the plot of ARPPs versus detection probabilities. The result shows that the proposed methodology can affect the system to some extent even when nonlinear estimator is used, especially when the bad data are present in the topology data, although the nonlinear estimator makes this effect relatively less significant compared with the linear case results.

C. Performance of the Greedy Search Heuristic

We also conducted simulation based on the proposed greedy search technique in Section IV-E. The simulation was based on 118-bus system, and all parameters were the same as those presented in Section VI-A. We compared the performance and computation time of the greedy heuristics with exhaustive search benchmark, as shown in Table I. Notice here the exhaustive search and greedy search are both over the lines that are close to their flow limits (estimated flows under M1 and M2, or actual flows under M3), the same as in Section VI-A. In Table I, the second column (average search time) is the average searching time for worst congestion pattern over 1000 Monte Carlo runs, and the third column (accuracy) is the percentage that the greedy search find the same worst congestion pattern as the exhaustive search. From the result, we can see that using greedy heuristic can give us much faster processing algorithm without losing much of the accuracy.

TABLE I
PERFORMANCE OF GREEDY SEARCH METHOD

method	average search time	accuracy
exhaustive search	1.23s	100%
greedy search	0.51s	97.3%

VII. CONCLUSION

We report in this paper a study on impacts of worst data on the real-time market operation. A key result of this paper is the geometric characterization of real-time LMP given in Theorem 1. This result provides insights into the relation between data and the real-time LMP; it serves as the basis of characterizing impacts of bad data.

Our investigation includes bad data scenarios that arise from both analog meter measurements and digital breaker state data. To this end, we have presented a systematic approach by casting the problem as one involving an adversary injecting malicious data. While such an approach often gives overly conservative analysis, it can be used as a measure of assurance when the impacts based on worst case analysis are deemed acceptable. We note that, because we use adversary attacks as a way to study the worst data, our results have direct implications when cyber-security of smart grid is considered. Given the increasing reliance on information networks, developing effective countermeasures against malicious data attack on the operations of a future smart grid is crucial. See [8], [10], [19], and [21] for discussion about countermeasures.

Although our findings are obtained from academic benchmarks involving relatively small size networks, we believe that the general trend that characterizes the effects of bad data is likely to persist in practical networks of much larger size. In particular, as the network size increases and the number of simultaneous appearance of bad data is limited, the effects of the worst meter data on LMP decrease whereas the effects of the worst topology data stay nonnegligible regardless of the network size. This observation suggests that the bad topology data are potentially more detrimental to the real-time market operation than the bad meter data.

REFERENCES

[1] F. F. Wu, P. Varaiya, P. Spiller, and S. Oren, "Folk theorems on transmission access: proofs and counterexamples," *J. Reg. Econ.*, vol. 10, pp. 5–23, Jul. 1996.

[2] E. Litvinov, T. Zheng, G. Rosenwald, and P. Shamsollahi, "Marginal loss modeling in LMP calculation," *IEEE Trans. Power Syst.*, vol. 19, no. 2, pp. 880–888, May 2004.

[3] T. Zheng and E. Litvinov, "Ex-post pricing in the co-optimized energy and reserve market," *IEEE Trans. Power Syst.*, vol. 21, no. 4, pp. 1528–1538, Nov. 2006.

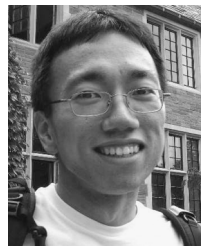
[4] A. Abur and A. G. Expósito, *Power System State Estimation: Theory and Implementation*. Boca Raton, FL, USA: CRC, 2000.

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

[6] F. C. Schweppe, J. Wildes, and D. P. Rom, "Power system static state estimation, Parts I, II, III," *IEEE Trans. Power App. Syst.*, vol. PAS-89, pp. 120–135, 1970.
[7] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proc. ACM Conf. Computer and Communications Security*, 2009.
[8] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on the smart grid," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 645–658, Dec. 2011.
[9] L. Jia, R. J. Thomas, and L. Tong, "On the nonlinearity effects on malicious data attack on power system," in *Proc. 2012 Power and Energy Society General Meeting*, Jul. 2012.
[10] G. Hug and J. A. Giampapa, "Vulnerability assessment of AC state estimation with respect to false data injection cyber-attacks," *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1362–1370, 2012.
[11] F. F. Wu and W. E. Liu, "Detection of topology errors by state estimation," *IEEE Trans. Power Syst.*, vol. 4, no. 1, pp. 176–183, Feb. 1989.
[12] K. A. Clements and P. W. Davis, "Detection and identification of topology errors in electric power systems," *IEEE Trans. Power Syst.*, vol. 3, no. 4, pp. 1748–1753, Nov. 1988.
[13] I. S. Costa and J. A. Leao, "Identification of topology errors in power system state estimation," *IEEE Trans. Power Syst.*, vol. 8, no. 4, pp. 1531–1538, Nov. 1993.
[14] A. Monticelli, "Modeling circuit breakers in weighted least squares state estimation," *IEEE Trans. Power Syst.*, vol. 8, no. 3, pp. 1143–1149, Aug. 1993.
[15] R. J. Thomas, L. Tong, L. Jia, and O. E. Kosut, "Some economic impacts of bad and malicious data," in *Proc. PSerc 2010 Workshop*, Portland, ME, USA, Jul. 2010, vol. 1.
[16] L. Xie, Y. Mo, and B. Sinopoli, "False data injection attacks in electricity markets," in *Proc. IEEE 2010 SmartGridComm*, Oct. 2010.
[17] A. L. Ott, "Experience with PJM market operation, system design, and implementation," *IEEE Trans. Power Syst.*, vol. 18, no. 2, pp. 528–534, May 2003.
[18] A. Mas-Colell and M. D. Whinston, *Microeconomics Theory*. Oxford, U.K.: Oxford Univ. Press, 1995.
[19] J. Kim and L. Tong, "On topology attack of a smart grid: Undetectable attacks and countermeasures," *IEEE J. Select. Areas Commun.*, vol. 31, no. 7, pp. 1294–1305, Jul. 2013.
[20] Power Systems Test Case Archive. [Online]. Available: http://www.ee.washington.edu/research/pstca/.
[21] T. T. Kim and H. Poor, "Strategic protection against data injection attacks on power grids," *IEEE Trans. Smart Grid*, vol. 2, no. 2, pp. 326–333, Jun. 2011.

Liyan Jia received the B.E. degree from the Department of Automation, Tsinghua University, Beijing, China, in 2009. He is currently pursuing the Ph.D. degree in the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, USA.

His current research interests are in smart grid, electricity market and demand response.



Jinsub Kim received the B.S. degree in electrical engineering from KAIST, Korea, and the Ph.D. degree in electrical and computer engineering (with minors in applied mathematics and statistics) from Cornell University, Ithaca, NY, USA.

He is a postdoctoral associate at the School of Electrical and Computer Engineering, Cornell University. He has conducted research on statistical inference for anomaly detection in communications and power networks. His graduate study was supported by a Samsung Scholarship.

Robert J. Thomas (M'73–SM'83–F'93–LF'08) is currently Professor Emeritus of Electrical and Computer Engineering at Cornell University, Ithaca, NY, USA. His technical background is broadly in the areas of systems analysis and control of large-scale electric power systems. He has published in the areas of transient control and voltage collapse problems as well as technical, economic, and institutional impacts of restructuring.

Prof. Thomas is a member of Tau Beta Pi, Eta Kappa Nu, Sigma Xi, and ASEE. He has received five teaching awards and the IEEE Centennial and Millennium medals. He has been a member of the IEEE-USA Energy Policy Committee since 1991 and was the committees Chair from 1997–1998. He is the founding Director of the 13-university-member National Science Foundation Industry/University Cooperative Research Center, PSerc and was one of 30 inaugural members of the U.S. Department of Energy Secretary's Electricity Advisory Committee (EAC). He has served as a Senior Advisor at the US DoE and as a Program Director at the NSF.

Lang Tong (S'87–M'91–SM'01–F'05) received the B.E. degree from Tsinghua University, Beijing, China, in 1985, and the M.S. and Ph.D. degrees in electrical engineering from the University of Notre Dame, Notre Dame, IN, USA, in 1987 and 1991, respectively.

He is the Irwin and Joan Jacobs Professor in Engineering at Cornell University, Ithaca, NY, USA. He was a Postdoctoral Research Affiliate at the Information Systems Laboratory, Stanford University in 1991. He was the 2001 Cor Wit Visiting Professor at the Delft University of Technology and had held visiting positions at Stanford University and the University of California at Berkeley. His research is in the general area of statistical inference, communications, and complex networks. His current research focuses on inference, optimization, and economic problems in energy and power systems.

Prof. Tong received the 1993 Outstanding Young Author Award from the IEEE Circuits and Systems Society, the 2004 best paper award from the IEEE Signal Processing Society, and the 2004 Leonard G. Abraham Prize Paper Award from the IEEE Communications Society. He is also a coauthor of seven student paper awards. He received Young Investigator Award from the Office of Naval Research. He was a Distinguished Lecturer of the IEEE Signal Processing Society.