Online Data Integrity Attacks Against Real-Time Electrical Market in Smart Grid

Song Tan, Wen-Zhan Song, *Senior Member, IEEE*, Michael Stewart, Junjie Yang, and Lang Tong, *Fellow, IEEE*

***Abstract*—The real-time electrical market operations in smart grid require reliable and accurate data from state estimation. However, state estimation is vulnerable to data integrity attacks, in which strategically manipulated meter measurements can bypass the conventional bad data detection and introduce errors. As a result, it becomes more likely for the attackers to control real-time electrical market through manipulations of meter measurements. In this paper, we first reveal the intrinsic relations between data integrity attacks and real-time electrical market operations, and explicitly characterize their complex interactions as a process simulator. Then a simulation-based global optimization problem is formulated from which attackers could maximize financial incentives through constructed data integrity attacks. More importantly, a novel systematic online attack construction strategy is proposed, such that attackers can launch the desired attacks only by the real-time data streams of meter measurements and no power network topology or parameter information is needed. A corresponding online defense strategy is also presented to detect and identify the malicious measurements without extra meter hardware investments. Finally, we evaluate the performance of the proposed attacking strategies and countermeasure through numerical simulations in IEEE test systems with both synthetic and real data from the New York Independent System Operator.**

***Index Terms*—Power system security, data attack, real-time electrical market.**

I. Introduction

THE BALANCE between supply and demand is the foundation of power system operations. In traditional power system, governments regulate all aspects market value chain (power generation, selling, transmission and distribution), in order to meet residential and business consumption [1]. With the challenge of integrating large scale intermittent renewable resources and promoting economic efficiency in Smart Grid, the electric power industry is gradually deregulated from governments, and real-time electrical market mechanisms are widely adopted by major U.S. independent system operators (ISOs) to balance supply and load, clearing market prices, and maintaining grid stability [2]. In real-time electrical market, the electricity price in the wholesale market is updated periodically (e.g., every 5 minutes) with end customers based on the real-time power generation and consumption status, which significant reduces over-provisioning and improves system efficiency. The operations of real-time electrical market are built upon data and results from state estimation. However, recent research shows that the state estimation process is vulnerable to data integrity attacks, in which the strategically manipulated meter measurements can easily bypass the conventional bad data detection and introduce errors [3]. As a result, it becomes more and more likely for attackers to control real-time electrical markets through malicious meter measurements. Therefore, investigating and systematic understanding the construction and impacts of data integrity attacks against real-time electrical market is crucial for system designers and operators to truly assess how these attacks may undermine the system's ability to provide mission-critical services.

Several existing related works have been conducted to address this issue. Xie *et al.* in [4] firstly investigate the impact of integrity attacks on power market through virtual bidding. In [5], Kosut *et al.* evaluate the proposed data attacks by their generated market revenues and the work is further studied by Jia *et al.* in pursuit of maximizing the revenues [6]. Yuan *et al.* in [7] show that the data integrity attacks can lead to increased system operating costs due to inordinate generation dispatch or energy routing. With the objective of controlling real-time Locational Marginal Prices (LMP) directly through data attacks, Tan *et al.* in [8] employ a control theory based approach to analyze the attack effect on pricing stability. Esmalifalak *et al.* in [9] novelly adopt a two-person zero-sum game approach to characterize the relations between attackers and defenders within electricity pricing. More recently, the authors in [10] and [11] have respectively proposed formal analytic frameworks to quantify the impact of data qualities on real-time LMP. However, all the above related works are based on the assumption that the attacker has full knowledge about the network information of targeted power systems, which includes network topologies and branch parameters, etc. In fact, in any given power system, the network information is huge and highly secured, and more importantly, these information are dynamic since the network topology could be reconfigured in both normal situations and contingencies.

Manuscript received November 24, 2015; revised February 19, 2016; accepted April 3, 2016. Date of publication April 5, 2016; date of current version December 21, 2017. This work was supported by the National Science Foundation under Grant NSF-1135814. Paper no. TSG-01486-2015.

S. Tan and W.-Z. Song are with the Department of Computer Science, Georgia State University, Atlanta, GA 30303 USA (e-mail: stan6@student.gsu.edu; wsong@gsu.edu).

M. Stewart is with the Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30303 USA (e-mail: mastewart@gsu.edu).

J. Yang is with the Department of Electrical and Information Engineering, Shanghai University of Electric Power, Shanghai 200090, China (e-mail: yangjj@shiep.edu.cn).

L. Tong is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: ltong@ece.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/TSG.2016.2550801

Therefore, it is rather difficult for attackers to achieve complete awareness of network information in practice. A most recent work in [12] firstly addresses this challenge by making inference through phasor observations, such that the linear structure of the power flow measurements can be acquired from independent component analysis.

In light of the above related works, through this paper, we present an online data integrity attacks against real-time electrical market in Smart Grid. Specifically, our key contributions are:

- We reveal the intrinsic relations between data integrity attacks and real-time electrical market, and explicitly characterize their complex interactions as a process simulator. Then a simulation-based global optimization problem is formulated, from which an attacker could maximize financial incentives through constructed data integrity attacks.
- We propose a novel systematic online attack construction strategy, such that attackers can launch the desired attacks only by the real-time data streams of meter measurements and no power network topology or parameter information is needed. As far as we know, our paper is the first attempt to attack real-time electrical market without network information.
- We also present a corresponding online defense strategy to detect and identify the malicious measurements. Without extra meter hardware investment, the system operator can directly employ the current collected data streams of meter measurements to detect the occurrences of attacks and identify the set of attacked measurements afterwards.
- We evaluate the performance of the proposed attacking strategies and countermeasure through numerical simulations in IEEE test systems with both synthetic and real data from the New York Independent System Operator.

II. Preliminaries and System Model

A. State Estimation and Bad Data Detection

In state estimation process, the control center collects real time measurements z from the deployed sensors and combines the network topology and parameter information to calculate the real time estimates of the unknown system variables x. Mathematically [13], let $x = (x_1, x_2, \ldots, x_n)^T$ and $z = (z_1, z_2, \ldots, z_m)^T$ denote state variables and meter measurements, respectively, where n is the number of unknown state variables, m is the number of meters, and $m \geq n$. The state variables are related to the measurements by $z = h(x) + e$, where e is the Gaussian measurement noise with zero mean and a covariance matrix $\sigma^2 I$. Under DC power flow model [13], the measurement model can be represented as:

$$z = Hx + e \tag{1}$$

where z is the bus power injection (power generation or load) and branch power flow measurements, H is an $m \times n$ full rank Jacobian matrix of the measurement model and x is the voltage phases at all buses. Then the estimated system states $\hat{x}$ and branch power flows $\hat{f}$ are given by:

$$\hat{x} = \left(H^T H\right)^{-1} H^T z, \quad \hat{f} = F\hat{x} \tag{2}$$

where F is the sensitivity matrix of branch flows with respect to the voltage phases. With DC power flow model, since there also exists a linear bijection between nodal power injections and voltage phases [14], then given a reference bus, we would have a l-by-n injection shift factor matrix S to denote the sensitivities of branch power flows with respect to the bus power injections [15], where l is the number of branches. Assume z contains the injection measurements at all buses and flow measurements across all the branches, denoted by z_{in} and z_f respectively, then we have:

$$z_f = S \cdot z_{in} + e, \quad \hat{f} = S \cdot z_{in} \tag{3}$$

Bad data detector employs residual to detect the abnormalities in measurement data. From (2),

$$\hat{z} = H\hat{x} = Kz, \quad \text{where } K = H\left(H^T H\right)^{-1} H^T \tag{4}$$

Then the measurement residual can be written as:

$$r = z - \hat{z} = (I - K)z \tag{5}$$

The detector fires an alarm when $\| r \|_2 >$ *threshold*.

B. Real-Time Electrical Market

A combined two-stage (day-ahead and real-time) market is widely adopted by major U.S. Independent System Operators (ISO) to stabilize the power system and calculate Locational Marginal Prices (LMP) [2]. In the day-ahead market, given the projected system load levels L, the ISO obtains the optimal generation dispatch P^*, the vector of predicted power generation at each bus. Then P^* are sent to all generators as generation reference, and day-ahead payments are collected from customers at all buses.

In the real-time stage, the ISO obtains the actual system response through state estimation, including the estimated power injections $\hat{P}, \hat{L}$ and branch flows $\hat{f}$. Then the following linear program [2] is solved to find the associated real-time LMP λ, a vector whose ith element λ_i is the LMP at bus i:

$$\begin{aligned} \underset{\Delta P,\ \Delta L}{\text{minimize}} \quad & \sum C_i^G \Delta P_i - \sum C_j^L \Delta L_j \\ \text{s.t.} \quad & (\tau) : \sum \Delta P_i = \sum \Delta L_j \\ & \Delta P_i^{min} \leq \Delta P_i \leq \Delta P_i^{max} \\ & (\mu_b) : \sum_i S_{bi} \Delta P_i - \sum_j S_{bj} \Delta L_j \leq 0, \text{ for } b \in \hat{C} \end{aligned} \tag{6}$$

where ΔP and ΔL are the vectors of incremental generation dispatch and load dispatch at buses, with fixed cost C^G, C^L respectively. $\Delta P_i^{min}, \Delta P_i^{max}$ are predefined lower and upper bounds, usually chosen as -2MW and 0.1MW in practice [2]. S_{bi} is element at bth row, ith column of matrix S in (3). Note that $\hat{C}$ is called *congestion pattern* [11], which denote the

sets of branches whose estimated power flow exceeds the flow limit f_b^{max},

$$\hat{C} = \left\{ b : \hat{f}_b > f_b^{max} \right\} \tag{7}$$

Then by solving (6), the real-time LMP at bus $i = 1, 2, \ldots, n$, is calculated:

$$\lambda_i = \tau - \sum_{b \in \hat{C}} S_{bi} \mu_b \tag{8}$$

where τ, μ_b are the corresponding dual variables in (6).

To clear the real-time market, the generator at bus i receives revenue $\lambda_i(\hat{P}_i - P_i^*)$, and the customer at bus j pays $\lambda_j(\hat{L}_j - L_j)$, where $\hat{P}_i$ and $\hat{L}_j$ are the estimated power generation and load at bus i and j from state estimation, respectively [2].

III. Problem Formulation

Suppose a malicious party wants to generate revenues from the real-time electrical market by compromising a subset of meters ζ_A, such that only measurements from meters in ζ_A can be modified. Note that the following strategies can also be applied to reducing customers' payments.

A. Constraints of Attacks

Firstly, since the attacker can only modify the measurements from meters in ζ_A, then the perturbed measurements has to be in the form:

$$\tilde{z} = z + a, \quad a \in \left\{ a \in \mathbb{R}^m \middle| a = \Psi c, \forall c \in R^m \right\} \tag{9}$$

where a is the attack vector, and Ψ is the diagonal matrix:

$$\Psi = diag(\psi_1, \ldots, \psi_m) \tag{10}$$

where ψ_i is a ***binary*** variable and $\psi_i = 1$ iff meter $i \in \zeta_A$.

Secondly, the attack should not be detected by the bad data detector. Based on (5), the new residual becomes $\tilde{r} = r + (I - K)a$. Based on triangular inequality,

$$\| \tilde{r} \|_2 \leq \| r \|_2 + \| (I - K)a \|_2 \tag{11}$$

Here we introduce a parameter ε, such that $\| (I - K)a \|_2^2 \leq \varepsilon$. The smaller ε is chosen, the less likely the attack will be detected. In the extreme case when $\varepsilon = 0$, the attack becomes *unobservable* [5].

B. Objective of Attacks

The objective of the attack is to maximize revenues from the real-time electrical market. From the end of Section II-B, we can see that the generator at bus i receives revenue $\lambda_i(\hat{P}_i - P_i^*)$ in normal situation. We analyze λ_i and $\hat{P}_i - P_i^*$ separately.

First, from (6) and (8), it suggests that given a shift factor matrix S, the real-time LMP λ depends only on the ISO's congestion pattern observation [11], i.e., $\hat{C}$. Meanwhile, since the ISO determines $\hat{C}$ through the estimated branch flows $\hat{f}$ as in (7), and $\hat{f}$ are solely determined by the power injection measurements within $\tilde{z}$ as in (3), therefore, we can see that $\tilde{z}$, $\hat{f}$, $\hat{C}$ and real-time LMP λ form a Markov chain, such that given a tuple of (a, z, S), there is a single corresponding λ. In other words, the LMP λ is essentially a function of (a, z, S).

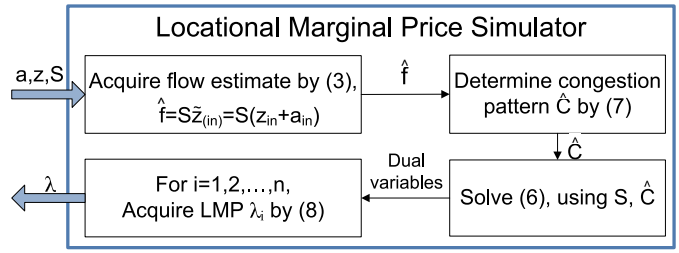


Fig. 1. Locational Marginal Price Simulator $\lambda(a, z, S)$.

So from now on, we denote LMP as $\lambda(a, z, S)$. We abstract the complex routine of $\lambda(a, z, S)$ as a simulator, and the flow chart of the simulator is shown in Figure 1.

Second, since the estimated power generation in real-time stage should match the optimal dispatch in day-ahead stage under normal situations [5], [6], then according to equation (4), for each bus i,

$$\hat{P}_i - P_i^* = K_i(z + a) - P_i^* = K_i a \tag{12}$$

where K_i is the corresponding row in matrix K to generation at bus i. Therefore, assume the attacker wants to make revenues from generations at buses within a target set $\mathbb{B}$. Then the total revenue from attack vector a is:

$$\mathcal{V}(a) = \sum_{i \in \mathbb{B}} \lambda_i(a, z, S) K_i a. \tag{13}$$

C. Construct Attacks Against Real-Time Electrical Market

From all the above, the problem of constructing data integrity attacks against real-time electrical market can be formulated as a simulation-based global optimization problem P1:

$$\text{(P1):} \quad \max_a \quad \mathcal{V}(a) = \sum_{i \in \mathbb{B}} \lambda_i(a, z, S) K_i a \tag{14}$$

$$\text{s.t.} \quad a = \Psi c, \quad \forall c \in R^m \tag{15}$$

$$\| (I - K)a \|_2^2 \leq \varepsilon. \tag{16}$$

D. Remarks

To solve P1, several facts are worth pointing out.

First, the objective function of P1 is based on complex simulator process in Figure 1. Only function values are available and there is no algebraic model to analyze differentiability and Lipschitz condition. Therefore, derivative-free optimization methods have to be employed [16]. The selection of optimization solvers is out of the scope of this paper.

Second, as in (3), (4), (6), (11), (13), the attacker will need the following knowledge to solve P1 accurately:

- 1) The meter measurement z.
- 2) The elements of matrix H, S and K.

The elements of matrix H, S and K depends on the detailed knowledge of **network information**, including network topology and branch parameters, such as the exact position of circuit breaker switches, transformer tap changers and power line admittances, etc. In fact, in any given power network, the network information is huge and highly secured, and these information could be dynamic since the

topology can be reconfigured in both normal situations and contingencies. Therefore, it is rather difficult for the attackers to achieve complete awareness of network information in practice.

IV. Online Construction of Data Integrity Attacks Without Network Information

In this section, we consider the strategy to construct data integrity attacks without network information, in other words, when matrix H, S and K are unknown to the attacker. Only real-time measurement data stream z are known to the attacker. By examining P1, we can see that the attacker need to specifically deal with $\| (I-K)a \|_2^2 \leq \varepsilon$ in the constraint, and $\lambda_i(a, z, S)$ and $K_i a$ in the objective.

A. Constraint $\| (I-K)a \|_2^2 \leq \varepsilon$

The constraint $\| (I-K)a \|_2^2 \leq \varepsilon$ determines whether or not the attack can be detected by the bad data detector. Since K is unknown, to safely launch an attack, the attacker would need to consider the extreme case of constraint $\| (I-K)a \|_2^2 \leq \varepsilon$, which is when $\varepsilon = 0$. In other words, if we could construct a vector a which has $\| (I-K)a \|_2^2 = 0$, then such an a would always satisfy the constraint and the bad data detector can never detect it. Note from (4) that when $a = Hv$, $\forall v \in R^n$, we always have $\| (I-K)a \|_2^2 = 0$. This is the so called unobservable attack [5]. Therefore, to satisfy the constraint, we just need to construct a vector a, which always lies in $\mathbb{R}(H)$, the column space of matrix H. The last question would be how to determine $\mathbb{R}(H)$ when H is unknown and even dynamic?

Inspired by [17], we can directly estimate and track the subspace $\mathbb{R}(H)$ using the measurement z. Let z_t denote the measurement vector at each time t, from (1):

$$z_t = Hx_t + e_t \tag{17}$$

To estimate $\mathbb{R}(H)$, at each time t, we aim at minimizing the following loss function,

$$J_t = \arg\min_{J \in R^{m \times n}} \sum_{i=1}^{t} \rho^{t-i} u_i(J), \tag{18}$$

where forgetting factor $0 \ll \rho \leq 1$ controls the memory and tracking ability, J is the estimated subspace with rank n since H is always full rank, and

$$u_i(J) = \min_x \|(z_i - Jx)\|_2^2, i = 1, \ldots, t \tag{19}$$

Note for all row $h = 1, 2, \ldots, m$, the objective function can be rowwise decomposed [18] as $J_t = [J_1^t, J_2^t, \ldots, J_m^t]^T$:

$$\begin{aligned} J_h^t &= \arg\min_{J_h} \sum_{i=1}^{t} \rho^{t-i} \left(z_i(h) - x_i^T J_h\right)^2 \\ &= J_h^{t-1} + \left(z_t(h) - x_t^T J_h^{t-1}\right)\left(W^t\right)^{\dagger} x_t \end{aligned} \tag{20}$$

where $W^t = \rho W^{t-1} + x_t x_t^T$ and $\dagger$ means pseudoinverse. Equation (20) is the classical formulation of Recursive Least Square (RLS) estimation with forgetting [19]. The subspace estimation and tracking process for $\mathbb{R}(H)$ is solved in Algorithm 1.

Algorithm 1 Subspace Estimation and Tracking for $\mathbb{R}(H)$

Input: A sequence of real-time measurements $z_t, t = 1, 2, \ldots$.
Initialize: An $m \times n$ random matrix J_0, and a diagonal matrix $(W^0)^{\dagger} = \delta I, \delta \gg 0$
for t=1,2,... **do**
  $x_t = (J_{t-1}^T J_{t-1})^{-1} J_{t-1}^T z_t$
  $\beta^t = 1 + \rho^{-1} x_t^T (W^{t-1})^{\dagger} x_t$,
  $\alpha^t = \rho^{-1} (W^{t-1})^{\dagger} x_t$
  $(W^t)^{\dagger} = \rho^{-1} (W^{t-1})^{\dagger} - (\beta^t)^{-1} \alpha^t (\alpha^t)^T$
  for h=1,2,...,m, in parallel **do**
    $J_h^t = J_h^{t-1} + (z_t(h) - x_t^T J_h^{t-1})(W^t)^{\dagger} x_t$
  end for
  Form J_t as $J_t = [J_1^t, J_2^t, \ldots, J_m^t]^T$
end for

Algorithm 2 Estimation for Shift Factor Matrix S

Input: Attack vector a, A sequence of real-time measurements $z_t, t = 1, 2, \ldots$.
Initialize: A diagonal matrix $(W^0)^{\dagger} = \delta I, \delta \gg 0$
for t=1,2,... **do**
  $\beta^t = 1 + \rho^{-1} (z_{in}^t)^T (W^{t-1})^{\dagger} z_{in}^t$,
  $\alpha^t = \rho^{-1} (W^{t-1})^{\dagger} z_{in}^t$
  $(W^t)^{\dagger} = \rho^{-1} (W^{t-1})^{\dagger} - (\beta^t)^{-1} \alpha^t (\alpha^t)^T$
  for j=1,2,...,l, in parallel **do**
    $S_j^t = S_j^{t-1} + (z_f^t(j) - (z_{in}^t)^T S_j^{t-1})(W^t)^{\dagger} z_{in}^t$
  end for
  Form S_t as $S_t = [S_1^t, S_2^t, \ldots, S_l^t]^T$;
end for

B. Objective $\lambda(a, z, S)$

To calculate $\lambda(a, z, S)$, the attacker should figure out the unknown matrix S. Assume z contains all the branch flow measurements and power injection measurements at all buses. Based on (3), let l denote the number of branches, for a particular branch flow measurement $z_f(j), j = 1, 2, \ldots, l$ in z_f, at each time t, we have:

$$z_f^t(j) = \left(z_{in}^t\right)^T S_j^t + e_t \tag{21}$$

where S_j^t is the jth row of matrix S at time t. Therefore, we can also estimate S_j^t through RLS with forgetting:

$$\begin{aligned} S_j^t &= \arg\min_{S_j} \sum_{i=1}^{t} \rho^{t-i} \left(z_f^i(j) - \left(z_{in}^i\right)^T S_j\right)^2 \\ &= S_j^{t-1} + \left(z_f^t(j) - \left(z_{in}^t\right)^T S_j^{t-1}\right)\left(W^t\right)^{\dagger} z_{in}^t \end{aligned} \tag{22}$$

where $W^t = \rho W^{t-1} + z_{in}^t (z_{in}^t)^T$. The process of estimating S is summarized in Algorithm 2. S_0 is initialized as a random $l \times n$ matrix. Therefore, to calculate $\lambda(a, z, S)$, at each time t, the attacker first get S_t from Algorithm 2, then invoke simulator $\lambda(a, z_t, S_t)$ as in Figure 1.

C. Objective $K_i a$

From (4), when H is unknown, K is unknown, so we cannot calculate $K_i a$ directly. However, the following Lemma 1 shed some light on the method to tackle this problem.

Lemma 1: K in (4) is an orthogonal projector onto $\mathbb{R}(H)$.

Proof: Suppose $b \in R^n$, and let $\hat{b} = H\hat{x}$ be the orthogonal projection of b onto $\mathbb{R}(H)$. Then the residual $r = b - \hat{b} = b - H\hat{x}$ is orthogonal to $\mathbb{R}(H)$, hence, it is orthogonal to each of the columns of H. As a result, we have:

$$\begin{aligned} &H^T\big(b - H\hat{x}\big) = 0 \implies H^T H\hat{x} = H^T b \\ &\implies \hat{x} = (H^T H)^{-1} H^T b \implies H\hat{x} = H(H^T H)^{-1} H^T b \\ &\implies \hat{b} = H(H^T H)^{-1} H^T b = Kb \end{aligned} \tag{23}$$

Therefore, K is an orthogonal projector onto $\mathbb{R}(H)$. ■

Based on Lemma 1, we present the following theorem to calculate $K_i a$ when K is unknown.

Theorem 1: Let matrix $K = H(H^T H)^{-1} H^T$, where $H \in R^{m\times n}$ with full rank n. Suppose there is another matrix $J \in R^{m\times n}$ also with full rank n, and $\mathbb{R}(J) = \mathbb{R}(H)$. Define matrix K' as:

$$K' = J\big(J^T J\big)^{-1} J^T, \tag{24}$$

then $K = K'$.

Proof: Since $\mathbb{R}(J) = \mathbb{R}(H)$, then:

$$\forall x \in R^n, \quad \exists y \in R^n, \quad \text{s.t.} \quad Hx = Jy. \tag{25}$$

Based on Lemma 1, matrix K' is also an orthogonal projector onto $\mathbb{R}(H)$. Therefore, for any $u \in R^m$, we can have:

$$u = Ku + r, \quad u = K'u + r' \tag{26}$$

where $Ku, K'u \in \mathbb{R}(H)$, and residual r, r' are orthogonal to $\mathbb{R}(H)$. So,

$$\big(r - r'\big)^T \big(Ku - K'u\big) = 0 \tag{27}$$

Since $r = u - Ku$, $r' = u - K'u$, we further have:

$$\big(K'u - Ku\big)^T \big(Ku - K'u\big) = 0 \tag{28}$$

which means for any $u \in R^m$, we have $Ku = K'u$. Then the orthogonal projector must be unique and $K = K'$. ■

Therefore, at each time t, the attacker can construct $K' = J_t(J_t^T J_t)^{-1} J_t^T$ using J_t generated from Algorithm 1, then use $K_i' a$ to replace $K_i a$ in the objective.

D. Summary

Based on Theorem 1, we can replace constraint $\| (I - K)a \|_2^2 \leq \varepsilon$ with $\| (I - K')a \|_2^2 \leq \varepsilon$. Therefore, the problem of attack construction without network information is formulated as P2:

$$\text{(P2):} \quad \max_a \ \mathcal{V}(a) = \sum_{i\in\mathbb{B}} \lambda_i(a, z, S) K_i' a_i \tag{29}$$

$$\text{s.t.} \ \ a = \Psi c, \quad \forall c \in R^m \tag{30}$$

$$\| \big(I - K'\big)a \|_2^2 \leq \varepsilon \tag{31}$$

Algorithm 3 Online Construction of Data Integrity Attacks Against Real-Time Electrical Market

Input: A sequence of real-time measurements $z_t, t = 1, 2, \ldots$.
Initialize: Launch Algorithm 1 and Algorithm 2;
for t=1,2,... **do**
  Get J_t from Algorithm 1 and S_t from Algorithm 2;
  if $\| J_t - J_{t-1} \|_F \leq \nu$ and $\| S_t - S_{t-1} \|_F \leq \nu$ **then**
    Construct $K' = J_t(J_t^T J_t)^{-1} J_t^T$, update P2 objective;
    Update the constraint $\| (I - K')a \|_2^2 \leq \varepsilon$ in P2;
    Solving P2 using derivative-free optimization method. (In search process, evaluations of objective function invoke simulator routine $\lambda(a, z_t, S_t)$);
    Based on solved vector a, modify the measurements of corresponding meters in ζ_A;
  end if
end for

The attack construction process at each time t is summarized in Algorithm 3, where ν is a small constant.

V. Countermeasure

In this section, we present the online defense strategy against the previously proposed attack. The defense strategy consists of two components: the attack detection component and the attack identification component. The detection component is responsible for indicating the existence of the attacks, and the identification component will be invoked afterwards to further identify the set of malicious meters. Only the real-time data streams of meter measurements are needed and no extra meter hardware investment is required.

A. Attack Detection

From the perspective of system operators, since the network topology and parameter information are highly secure in control center [13], they are the trustworthy baseline that can be employed to detect attacks. Note the shift factor matrix only depends on the network topology and parameter info, such that the matrix known to the system operator should be accurate at all times. We denote the true shift factor matrix as S_{true}. Meanwhile, as shown in Algorithm 2, the shift matrix can also be derived from the power flow measurements. Therefore, to detect the attack, the system operator could derive a corresponding shift factor matrix $\tilde{S}$ from a series of collected measurements $\tilde{z}$, then a discrepancy between the derived shift factor matrix $\tilde{S}$ and the S_{true} at hand will trigger an alarm to indicate the attack.

Specifically, suppose the control center collects a series of real time measurements $\tilde{z}$, which could be the manipulated measurements. Similarly as in (21), let l denote the number of branches, for a particular branch flow measurement $\tilde{z}_f(j), j = 1, 2, \ldots, l$ in $\tilde{z}_f$, at each time t, we have:

$$\tilde{z}_f^t(j) = \big(\tilde{z}_{in}^t\big)^T \tilde{S}_j^t + e_t \tag{32}$$

where $\tilde{S}_j^t$ is the jth row of matrix $\tilde{S}$ at time t. As a result, we can also estimate $\tilde{S}_j^t$ iteratively through RLS with forgetting:

$$\begin{aligned}\tilde{S}_j^t &= \arg\min_{\tilde{S}_j} \sum_{i=1}^{t} \rho^{t-i}\Big(\tilde{z}_f^i(j) - \big(\tilde{z}_{in}^i\big)^T \tilde{S}_j\Big)^2 \\ &= \tilde{S}_j^{t-1} + \Big(\tilde{z}_f^t(j) - \big(\tilde{z}_{in}^t\big)^T \tilde{S}_j^{t-1}\Big)\big(W^t\big)^{\dagger}\tilde{z}_{in}^t \end{aligned} \tag{33}$$

Note $W^t = \rho W^{t-1} + \tilde{z}_{in}^t(\tilde{z}_{in}^t)^T$, and its pseudoinverse can be recursively updated as:

$$\big(W^t\big)^{\dagger} = \rho^{-1}\Big(W^{t-1}\Big)^{\dagger} - \big(\beta^t\big)^{-1}\alpha^t\big(\alpha^t\big)^T \tag{34}$$

where

$$\beta^t = 1 + \rho^{-1}\big(\tilde{z}_{in}^t\big)^T\Big(W^{t-1}\Big)^{\dagger}\tilde{z}_{in}^t, \alpha^t = \rho^{-1}\Big(W^{t-1}\Big)^{\dagger}\tilde{z}_{in}^t \tag{35}$$

Then derived shift factor matrix at time t can be formed as: $\tilde{S}_t = [\tilde{S}_1^t, \tilde{S}_2^t, \ldots, \tilde{S}_l^t]^T$.

After $\tilde{S}$ is derived, the discrepancy can be calculated as:

$$\gamma = \frac{\|S_{true} - \tilde{S}\|_F}{\|S_{true}\|_F} \tag{36}$$

Then if the discrepancy γ is greater than a tuned threshold κ, an alarm would be triggered and the attack identification process will be invoked.

B. Attack Identification

To further identify which measurements have been manipulated, suppose the control center collects a series of $\mathbb{N}$ real time measurements at continuous time stamps $\tilde{z}_t, t = 1, 2, 3 \ldots, \mathbb{N}$ and construct a m-by-$\mathbb{N}$ matrix $\tilde{Z}$ by columnwise stacking these measurement vectors together. From (9), we have:

$$\tilde{Z} = Z + A \tag{37}$$

where $Z = [z_1, z_2, \ldots, z_{\mathbb{N}}]$ is the matrix containing normal measurements and $A = [a_1, a_2, \ldots, a_{\mathbb{N}}]$ is the matrix containing all the attack vectors. It is known that power system measurements change gradually in continuous time interval [20], rendering Z typically low rank. Meanwhile, since the attacker usually can only modify a limited number of meter measurements, such that matrix A tends to be sparse [21]. Therefore, the normal measurements and attack vectors can be recovered from:

$$\text{(P3):} \quad \min_{Z,A} \quad \| Z \|_* + \omega \| A \|_1 \quad \text{s.t.} \quad \tilde{Z} = Z + A \tag{38}$$

where $\| \cdot \|_*$ denotes the nuclear norm, ω is a regularization parameter. The problem P3 is the well known sparse and low-rank matrix decomposition problem [22]. As long as we can recover sparse matrix A, then based on its rows that contain nonzero elements, we can identify which measurements have been attacked. As suggested by [23], P3 can be solved by employing Alternating Direction Method of Multipliers (ADMM) algorithm. The Lagrangian function of P3 is:

$$\begin{aligned}\mathbf{L}(Z, A, Q, \mu) = &\| Z \|_* + \omega \| A \|_1 + \langle Q, \tilde{Z} - Z - A\rangle \\ &+ \frac{\mu}{2} \| \tilde{Z} - Z - A \|_2^2 \end{aligned} \tag{39}$$

where μ is positive number, Q is the Lagrangian multipliers, and $\langle\cdot\rangle$ denotes the Frobenius matrix product. For each iteration $k = 1, 2, 3, \ldots$ until convergence, Z is firstly updated as:

$$Z_{k+1} = U\mathbf{P}_{\frac{1}{\mu}}\{D\}V^T \tag{40}$$

where U, D, V^T are the singular value decomposition of matrix $(\tilde{Z} - A_k + \frac{1}{\mu}Q_k)$, and the operator $\mathbf{P}_a\{b\}$ is an elementwise applied soft thresholding function defined as:

$$\mathbf{P}_a\{b\} = \text{sign}(b) \cdot \max\{|b| - a, 0\} \tag{41}$$

Secondly, A is updated as:

$$A_{k+1} = \mathbf{P}_{\frac{\omega}{\mu}}\left\{\tilde{Z} - Z_{k+1} + \frac{1}{\mu}Q_k\right\} \tag{42}$$

Finally, the Lagrangian multiplier Q is updated as:

$$Q_{k+1} = Q_k + \mu\big(\tilde{Z} - Z_{k+1} - A_{k+1}\big). \tag{43}$$

C. Summary

From the above all, the procedure of online defense against data integrity attacks is illustrated in Algorithm 4. The attack detection process is monitoring the system using real-time measurement data stream all the time. Once an attack is detected, the attack identification process is launched to identify the potential malicious set of meters ζ_A. Then the measurements from meter set ζ_A will be removed from the measurement data stream used by the attack detection process. This procedure iterates until there is no attack indicated by the attack detection process. $\tilde{S}_0$ is initialized as a random $l \times n$ matrix, and κ is a tuned constant. $\mathbb{N}$ is the measurement buffer size, which determines how soon the system is able to identify the malicious set of meters after an attack is detected.

VI. Evaluation

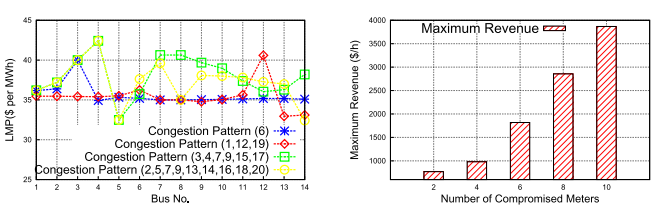
In this section, we evaluate our proposed attacking strategies through IEEE bus benchmark system [24] with both synthetic data and real load data streams from the New York Independent System Operator [25]. All the numerical simulations are conducted in Matlab platform with software packages including @MATPOWER and patternsearch solver in Global Optimization Toolbox.

A. Attack When Network Info Is Known

In this part, we examine the performance of P1 through IEEE14 bus system (14 buses and 20 branches), in which the network information is known. In this case, we use the synthetic load data that comes with MATPOWER. All power injection measurements and power flow measurements (in both directions for each line) are employed, such that $m = 54, n = 14$. We first examine the functionality of LMP simulator. Since the LMP at all buses totally depend on the congestion pattern, we directly plot the LMP under different congestion patterns in Figure 2(a). The congestion pattern includes the ID of branches whose power flows exceed the security limits. One interesting fact is that different congestion patterns could result in the same LMP at a particular bus,

Algorithm 4 Online Defense of Data Integrity Attacks Against Real-Time Electrical Market

1: M is the set of all measurement meters.
2: **Input:** A sequence of real-time measurements $\tilde{z}_t, t = 1, 2, \ldots$ from meter set M.
3: **Initialize:** A diagonal matrix $(W^0)^\dagger = \delta I, \delta \gg 0, k = 0$.
4: **for** t=1,2,... **do**
5: $\beta^t = 1 + \rho^{-1}(\tilde{z}^t_{in})^T(W^{t-1})^\dagger \tilde{z}^t_{in}$,
6: $\alpha^t = \rho^{-1}(W^{t-1})^\dagger \tilde{z}^t_{in}$
7: $(W^t)^\dagger = \rho^{-1}(W^{t-1})^\dagger - (\beta^t)^{-1}\alpha^t(\alpha^t)^T$
8: **for** j=1,2,...,l, in parallel **do**
9: $\tilde{S}^t_j = \tilde{S}^{t-1}_j + (\tilde{z}^t_f(j) - (\tilde{z}^t_{in})^T\tilde{S}^{t-1}_j)(W^t)^\dagger \tilde{z}^t_{in}$
10: **end for**
11: Form $\tilde{S}_t$ as $\tilde{S}_t = [\tilde{S}^t_1, \tilde{S}^t_2, \ldots, \tilde{S}^t_l]^T$;
12: **if** $\frac{\|\tilde{S}_t - \tilde{S}_{t-1}\|_F}{\|\tilde{S}_{t-1}\|_F} \leq \nu$ **then**
13: Calculate $\gamma = \frac{\|S_{true} - \tilde{S}_t\|_F}{\|S_{true}\|_F}$
14: **if** $\gamma > \kappa$ **then**
15: (Concurrently start a separate and independent identification process as following.)
16: Formulate matrix $\tilde{Z}$ by columnwise stacking most recent $\mathbb{N}$ measurements;
17: Initialize $Z_0 = 0$, $Q_0 = 0$, $\mu = \frac{m \cdot \mathbb{N}}{4 \cdot \|\tilde{Z}\|_1}$, and $\omega = \frac{1}{\sqrt{\max(m, \mathbb{N})}}$
18: **while** not converged **do**
19: Update Z_{k+1}, A_{k+1} and Q_{k+1} as in (40)-(43);
20: k=k+1;
21: **end while**
22: Based on the rows containing nonzero elements of recovered matrix A, assign the corresponding meters to the malicious set ζ_A.
23: Update set M as $M = M \setminus \zeta_A$.
24: Go to step 2.
25: **end if**
26: **end if**
27: **end for**



(a) LMP at buses in different congestion patterns

(b) Maximum revenues with different numbers of compromised meters

Fig. 2. Data Integrity Attacks in IEEE 14 Bus system.

e.g., the LMP at bus 4 are the same in the last two congestion patterns, both of which have branch 7, 9 congested, and are incident with bus 4.

Then we demonstrate the optimal attack vectors in P1 when different number of meters are compromised. Table I lists the optimal attack vector a against IEEE14 system under different size of ζ_A. The notation (p, q) denotes the nonzero entries of vector a, and p is the index and q is the value. The corresponding maximum revenues under optimal attack vectors are given in Figure 2(b). We can see that the number of compromised meters has a significant impact on the revenues.

TABLE I
OPTIMAL ATTACK VECTOR a AGAINST IEEE14 WITH DIFFERENT SIZES OF ζ_A

size of ζ_A	optimal attack vector a
2	(0,63.0),(2,34.4)
4	(0,65.1),(2,32.0),(14,48.5),(34,-64.0)
6	(0,69.3),(2,32.0),(3,-48.0),(14,-48.0),(15,32.0),(34,-64.0)
8	(0,79.0),(2,32.0),(3,-32.0),(4,-48.0),(5,32.0),(14,52.3),(15,32.0),(34,-64.0)
10	(0,103.0),(2,32.0),(3,-48.0),(4,-64.0),(5,32.0),(14,68.0),(15,34.0),(17,32.0),(34,-64.0),(35,-33.0)

B. Online Attack Construction: When Network Info Is Unknown

In this part, we examine the performance of P2, in which the network information is unknown. Both IEEE14 bus and IEEE118 bus system are employed.

1) Data Preparation: For IEEE14 bus system, we incorporate the real-time load data streams from the New York independent system operator (NYISO) during a 48 hour period (10/01-10/02) in 2015. In NYISO, there are 11 load regions (CAPITL, CENTRL, DUNWOD, GENESE, HUDVL, LONGIL, MHKVL, MILLWD, NYC, NORTH, WEST). The load data are recorded for each region every 5 minutes. Therefore, for each region, there are load data at $12 \times 48 = 576$ continues time instances. Meanwhile, since there are exactly 11 load buses with IEEE14 bus system, the NYISO load data at each region can be directly mapped to each load bus in IEEE14 and generate 576 corresponding real-time measurements $z_t, t = \{1, 2, \ldots, 576\}$. Specifically, the following procedures are performed [26]:

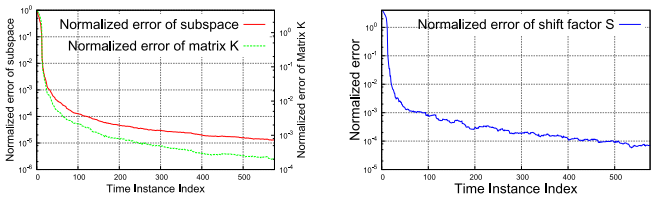
- Map each load bus of IEEE14 bus system with one region of NYISO based on Table II.
- Calculate ratio of the total loads from NYISO to the total loads of original IEEE14 buses system. Then divide each NYISO region load by the ratio and assign the resulted load to each load bus within IEEE14. The generation capacity in IEEE14 is not changed.
- Solve the system state x_t using power flow calculations based on the new loads and generate corresponding measurement z_t by the IEEE14 model.

For IEEE118 system, we leverage synthetic data generated from Monte Carlo simulation. In each Monte Carlo run, we use nonlinear state estimation model to generate measurement vector at each time instance. State vectors at different time instances are assumed to be independent and identically distributed Gaussian random vectors with the mean equal to the operating states given in 118 bus data sheet.

2) Performance of Algorithm 1 and 2: To evaluate Algorithm 1 and 2, we use normalized errors to examine the performance of estimations for subspace $\mathbb{R}(H)$, matrix K, and S, which are defined as $\frac{\|(I - J_t \cdot J_t')H\|_F}{\|H\|_F}$, $\frac{\|K' - K\|_F}{\|K\|_F}$, and $\frac{\|S_t - S\|_F}{\|S\|_F}$, respectively. Meanwhile, to evaluate the performance of proposed algorithms in dynamic topology scenarios: For IEEE14 system, we disconnect bus 4 and bus 5 at time instance 200, and reconnect them at time instance 400. For IEEE118 system, we disconnect bus 15 and bus 33 at time instance 200,

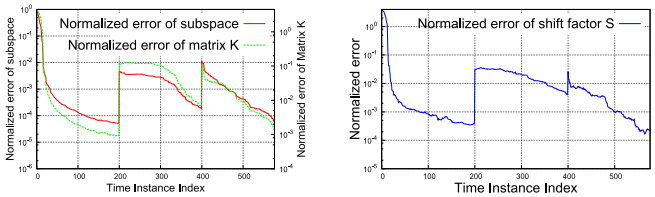
TABLE II
MAPPING BETWEEN NYISO REGIONS AND IEEE14 BUSES

Region Name	CAPITL	CENTRL	DUNWOD	GENESE	HUDVL	LONGIL	MHKVL	MILLWD	NYC	NORTH	WEST
IEEE14 Bus No.	2	3	4	5	6	9	10	11	12	13	14



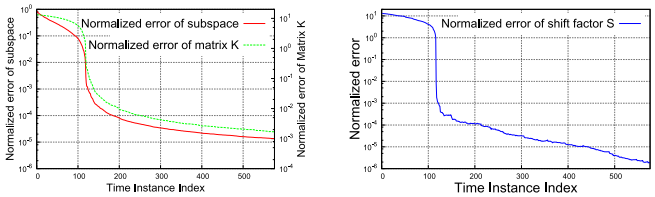
(a) Subspace and matrix K estimation error (b) Shift factor S estimation error

Fig. 3. Normalized error of estimation in IEEE 14.



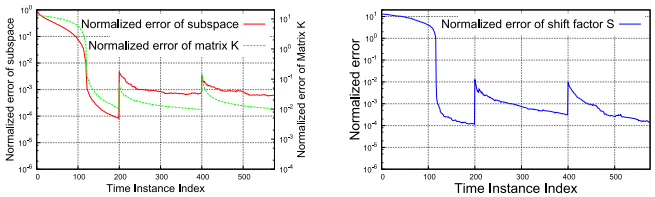
(a) Subspace and matrix K estimation error (b) Shift factor S estimation error

Fig. 4. Normalized error of estimation in IEEE 14 with dynamic topology.



(a) Subspace and matrix K estimation error (b) Shift factor S estimation error

Fig. 5. Normalized error of estimation in IEEE 118.

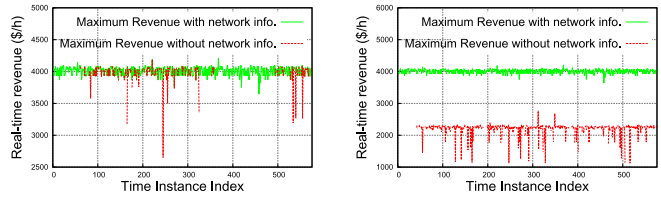


(a) Subspace and matrix K estimation error (b) Shift factor S estimation error

Fig. 6. Normalized error of estimation in IEEE 118 with dynamic topology.

and reconnect them at time instance 400. Figure 3-6 show the normalized errors as the time goes in both cases. We can see that the algorithms are more sensitive to the topology changes in a smaller size of power system, in which the peak error will be larger but can be reduced more quickly.

3) Performance of Algorithm 3: We evaluate the performance of Algorithm 3 from the perspective of attackers to see how much revenue can be generated. Since there will be errors in the constructions of $\mathbb{R}(H)$, matrix K, and S, the bad data alarm is likely to be fired when the optimal attack vector from



(a) $\varepsilon =$ threshold (b) $\varepsilon =$ threshold$/2$

Fig. 7. Real-time revenues with different ε in IEEE 14.

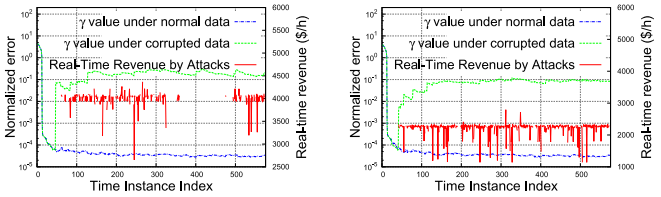
P2 is applied. Therefore, from the attacker's point of view, choosing the value of parameter ε in P2 would be critical. In IEEE14 system, the 0.05 significant-level bad data detector employs a chi-square distribution $threshold = \chi^2_{m-n,0.95} = \chi^2_{54-14,0.95}$. We present the corresponding real-time revenues under different ε in Figure 7 by employing the real load data from NYISO. The size of ζ_A is 10. In both cases, we plot the maximum revenues with known network information as a reference. When the network information is unknown, we can see that the revenues curve will start at a time point around 50 instead of 0. This is because in Algorithm 3, the normalized error of $\mathbb{R}(H)$, matrix K, and S can only become less than $\nu = 0.01$ until it collects certain amount of measurements. More importantly, the revenue curve is not continuous. The missing points in the curve are the time instances when the bad data alarm is fired due to the attack vector from P2. In that case, no revenue can be generated by the attacker. It can be seen that when ε is reduced from $threshold/2$, more time instances can generate revenues but the value of revenue is decreased correspondingly. Also note in the case when there is a load forecasting error in day-ahead market, from equation (12), since the error only adds a constant error term in final calculated real time revenues, so there will just be a vertical displacement in the profit curve compared with the one without forecasting error.

C. Countermeasure

In this subsection, we present the evaluation of countermeasure, which consists of both attack detection and attack identification in Algorithm 4.

1) Attack Detection: For the attack detection, corresponding to the two attack scenarios in Figure 7, we plot the resulted detection value γ when using both normal measurements and corrupted measurements in Figure 8. We can see that when the attack starts to generate revenue and cause data corruption, there would be a significant increase in value γ compared to its value in normal case. Therefore, using value γ can effectively indicate the existence of attacks.

2) Attack Identification: To evaluate the performance of attack identification, we examine true positive rates and false alarm rates of malicious meter identification when different



(a) $\varepsilon =$ threshold (b) $\varepsilon =$ threshold$/2$

Fig. 8. Attack detection with different ε.

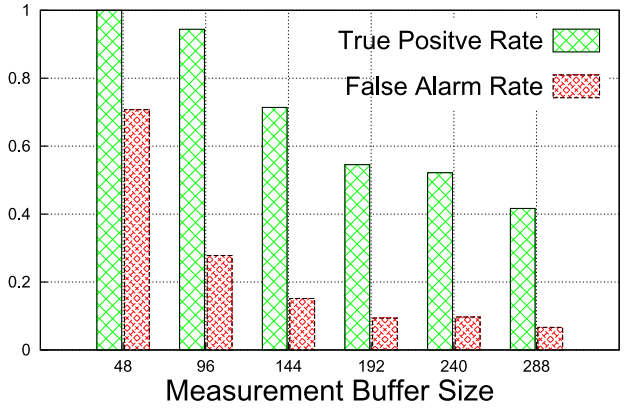


Fig. 9. Attack identification with different measurement buffer size.

measurement buffer size $\mathbb{N}$ are employed. The measurement buffer size $\mathbb{N}$ determines how quickly the system would begin to identify the malicious meter sets after an attack is detected. From Figure 9, we can see that the buffer size $\mathbb{N}$ neither can be too small nor too large, which would result in either a high false alarm rate or a poor true positive rate. A good trade off would be 96 in this case, where the formulated matrix $\tilde{Z}$ have slightly more columns than rows.

D. Online Computational Performance

The Algorithm 1, 2, 3, and 4 are computationally intensive. Since the real time measurements are usually published every few minutes (5 minutes in PJM), our algorithms should be fast and responsive enough to adapt to the data generation speed. We evaluate the these algorithms within in Matlab 2013, on our testing machine (64 bits HP desktop with Inter(R) Core(TM) i7-5500 CPU@2.40GHz and 8GB memory). The *cputime* function in Matlab is employed to track the execution time. When receiving a new measurement vector z, the average time needed (seconds) to estimate $R(H)$ (Algorithm 1), estimate shift factor matrix S (Algorithm 2), calculate attack vector a (Algorithm 3), and detect and identify corresponding attack vector a (Algorithm 4), in both IEEE14 and IEEE118 bus systems, are listed in Table III. From the table, we can see that due to the application of derivative-free optimization solver, the computation time increases significantly as the size of power network increases. However, in case IEEE118, Algorithm 3 is still responsive enough to generate the attack vector in real time. Moreover, since Algorithm 4 requires the buffering of several measurements before starting, it would be able to meet the online operation requirement.

TABLE III
COMPUTATIONAL TIME OF ALGORITHMS 1-4 IN SECONDS

Case	Algorithm 1	Algorithm 2	Algorithm 3	Algorithm 4
IEEE14 bus	0.0013	1.8989e-04	17.7542	33.4870
IEEE118 bus	0.0102	0.0027	101.546	219.772

VII. CONCLUSION

In this paper, we present online data integrity attacks against real-time electrical market. The online attack construction strategy is proposed when the attacker has no knowledge of power network information and our results show that the attacker could generate a fair amount of revenues through data integrity attacks. A corresponding online countermeasure is also presented to detect and identify the attacks. Exploring the properties of the measurement time series in state estimation gives a new perspective of security analytics for Smart Grid system. Future work would further address the situation when system topology is dynamic and the collected topology information is erroneous.

REFERENCES

[1] H. Lund, A. N. Andersen, P. A. Østergaard, B. V. Mathiesen, and D. Connolly, "From electricity smart grids to smart energy systems—A market operation based approach and understanding," *Energy*, vol. 42, no. 1, pp. 96–102, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360544212002836

[2] 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.

[3] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proc. 16th ACM Conf. Comput. Commun. Sec.*, Chicago, IL, USA, 2009, pp. 21–32.

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

[5] 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.

[6] L. Jia, R. J. Thomas, and L. Tong, "Malicious data attack on real-time electricity market," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 5952–5955.

[7] Y. Yuan, Z. Li, and K. Ren, "Quantitative analysis of load redistribution attacks in power systems," *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 9, pp. 1731–1738, Sep. 2012.

[8] R. Tan, V. B. Krishna, D. K. Yau, and Z. Kalbarczyk, "Impact of integrity attacks on real-time pricing in smart grids," in *Proc. ACM SIGSAC Conf. Comput. Commun. Sec.*, New York, NY, USA, 2013, pp. 439–450.

[9] M. Esmalifalak, G. Shi, Z. Han, and L. Song, "Bad data injection attack and defense in electricity market using game theory study," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 160–169, Mar. 2013.

[10] D.-H. Choi and L. Xie, "Sensitivity analysis of real-time locational marginal price to SCADA sensor data corruption," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1110–1120, May 2014.

[11] L. Jia, J. Kim, R. J. Thomas, and L. Tong, "Impact of data quality on real-time locational marginal price," *IEEE Trans. Power Syst.*, vol. 29, no. 2, pp. 627–636, Mar. 2014.

[12] M. Esmalifalak *et al.*, "A stealthy attack against electricity market using independent component analysis," *IEEE Syst. J.*, to be published, doi: 10.1109/JSYST.2015.2483742.

[13] A. Abur and A. G. Expósito, *Power System State Estimation: Theory and Implementation*. New York, NY, USA: Marcel Dekker, 2004.

[14] F. Wu, P. Varaiya, P. Spiller, and S. Oren, "Folk theorems on transmission access: Proofs and counterexamples," *J. Regul. Econ.*, vol. 10, no. 1, pp. 5–23, 1996.

[15] A. J. Wood and B. F. Wollenberg, *Power Generation, Operation, and Control*, 2nd ed. Hoboken, NJ, USA: Wiley, 1996.

[16] L. M. Rios and N. V. Sahinidis, "Derivative-free optimization: A review of algorithms and comparison of software implementations," *J. Glob. Optim.*, vol. 56, no. 3, pp. 1247–1293, 2013.

[17] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 95–107, Jan. 1995.
[18] Y. Chi, Y. C. Eldar, and R. Calderbank, "PETRELS: Parallel subspace estimation and tracking by recursive least squares from partial observations," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5947–5959, Dec. 2013.
[19] K. J. Åström and B. Wittenmark, *Adaptive Control*, 2nd ed. Boston, MA, USA: Addison-Wesley, 1994.
[20] J. Hao, R. J. Piechocki, D. Kaleshi, W. H. Chin, and Z. Fan, "Optimal malicious attack construction and robust detection in smart grid cyber security analysis," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Venice, Italy, Nov. 2014, pp. 836–841.
[21] L. Liu, M. Esmalifalak, Q. Ding, V. A. Emesih, and Z. Han, "Detecting false data injection attacks on power grid by sparse optimization," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 612–621, Mar. 2014.
[22] E. J. Candés and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009. [Online]. Available: http://dx.doi.org/10.1007/s10208-009-9045-5
[23] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 1–37, Jun. 2011. [Online]. Available: http://doi.acm.org/10.1145/1970392.1970395
[24] *Data Sheets for IEEE Bus Systems*. Accessed on Aug. 12, 2015. [Online]. Available: http://shodhganga.inflibnet.ac.in/bitstream/10603/5247/18/19_appendix.pdf
[25] *New York Independent System Operator Load Data*. Accessed on Aug. 12, 2015. [Online]. Available: http://www.nyiso.com/public/markets_operations/market_data/load_data/index.jsp
[26] G. Chaojun, P. Jirutitijaroen, and M. Motani, "Detecting false data injection attacks in AC state estimation," *IEEE Trans. Smart Grid*, vol. 6, no. 5, pp. 2476–2483, Sep. 2015.

Song Tan received the B.S. degree from Northeast Normal University, Changchun, China, and the M.S. degree from Georgia State University, where he is currently pursuing the Ph.D. degree with the Department of Computer Science. His current research is focused on the cyber-physical security in smart grid system, which includes bad data detection, electrical market security, and design of cyber-physical security testbed for smart grid.

Wen-Zhan Song (M'05–SM'11) is currently a Professor with the Department of Computer Science, Georgia State University. His research mainly focuses on sensor Web, smart grid, and smart environment, where sensing, computing, communication, and control play a critical role and need a transformative study. His research has received more than $6 million in research funding from NSF, NASA, USGS, and Boeing, since 2005.

Michael Stewart is an Associate Professor with the Department of Mathematics and Statistics, Georgia State University. His research is in numerical linear algebra, including the development of fast algorithms, error analysis, and applications.

Junjie Yang received the Ph.D. degree from Shanghai Jiao Tong University, China. He is currently an Associate Professor with the Department of Electric and Information Engineering, Shanghai University of Electric Power, China. His research areas are intelligent demand response in smart grid, remote and online monitoring of power substations, and wireless sensor networks.

Lang Tong (S'88–M'90–SM'01–F'05) joined Cornell University in 1998, where he is currently the Irwin and Joan Jacobs Professor of Engineering and the Cornell Site Director of the Power Systems Engineering Research Center. He was a recipient of the Best Paper Award from the IEEE Signal Processing Society and the Leonard G. Abraham Prize Paper Award from the IEEE Communications Society.