Shuguang Cui, Zhu Han, Soummya Kar,
Tùng T. Kim, H. Vincent Poor, and Ali Tajer

Coordinated Data-Injection Attack and Detection in the Smart Grid

A detailed look at enriching detection solutions

ISTOCKPHOTO.COM/© SIGAL SUHLER MORAN

A smart grid improves the efficiency of power grids via the aid of modern communication, signal processing, and control technologies. While smart grid integration enables power grid networks to be smarter, it also increases the risk of cyberattacks due to the strong dependence on the cyberinfrastructure in the overall system. In this article, the coordinated data-injection attack detection problem in the smart grid is considered. Specifically, the data-injection attack model is first introduced and a thorough survey of existing detection methods is then given. Afterward, three important efforts to enrich the detection solution are presented in detail:

1) attacker versus defender dynamics, where possible interactive attack and defense strategies are discussed in the context of secure phasor measurement unit (PMU) placement

2) distributed attack detection and state recovery, where the focus is how to achieve the optimal centralized performance with a distributed approach

3) quickest detection (QD), where the tradeoff between the detection speed and detection performance is studied.

A list of associated key open problems in this area is then presented to conclude this article.

MOTIVATION AND INTRODUCTION

The electric power industry is undergoing profound changes as our society increasingly emphasizes the importance of a smarter grid in support of clean and sustainable energy utilization. Technically, enabled by advances in sensing, communication, and actuation, power system operations are likely to involve many more distributed

Digital Object Identifier 10.1109/MSP.2012.2185911
Date of publication: 20 August 2012

1053-5888/12/$31.00©2012IEEE

real-time information gathering and processing devices. Institutionally, the increasing presence of smart metering and demand response programs may open the door to more intelligent supervisory control and data acquisition (SCADA) networks and end-user networks. Meanwhile, the deregulation of the electric power industry has unbundled generation and transmission, allowing a broad range of market participants (e.g., load-serving entities and independent power producers) to make decisions in coordination to keep a balanced power market. On the other hand, the stronger coupling between cyber- and physical operations makes power systems more vulnerable to cyberattacks. Such malicious attacks could often be coordinated across the whole network such that classic bad data detection approaches become ineffective [1]. As such, we envision that the issues arising in smart grid development demand novel information processing schemes.

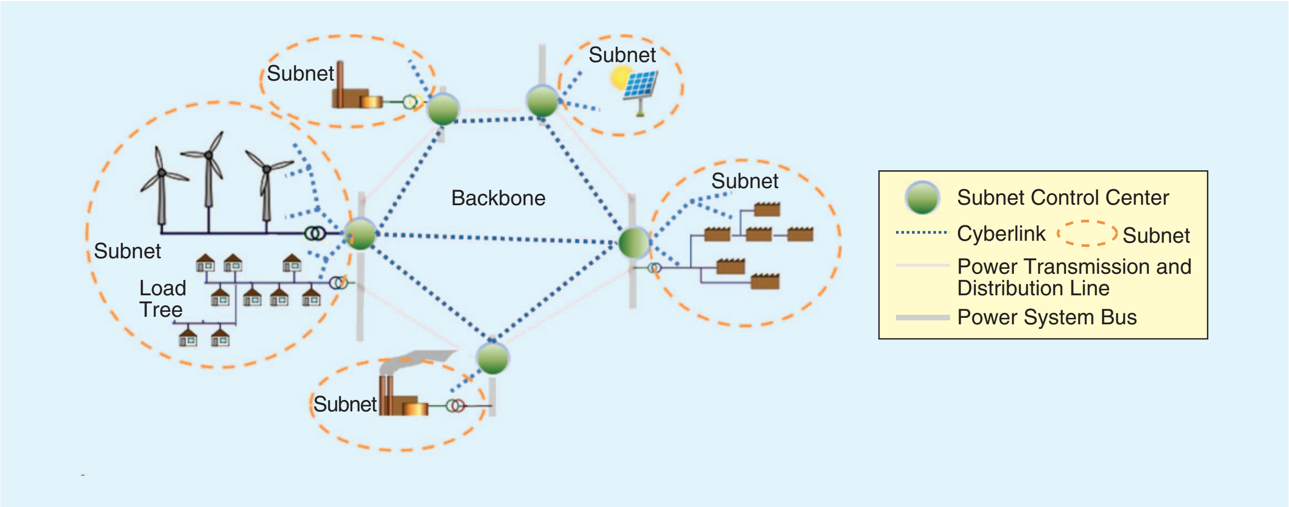
We focus on the fundamental problems in identifying and mitigating the impact of malicious cyberattacks, especially data-injection attacks during the state estimation process. In particular, we consider the case of a large interconnected power system, in which the whole grid is composed of multiple subnets as shown in Figure 1 (note the dotted-line links for the cybercontrolling part). In the system diagram, each subnet has a control center, which controls the nodes within its subnet along with the interfaces with other subnets, and coordinates with other subnet control centers to jointly detect system-wide cyberattacks. Among the subnet controllers, we assume a mesh backbone connection to exchange system information and computation results. It is clear that to protect against system-wide coordinated cyberattacks, different subnets should collaborate.

THE ELECTRIC POWER INDUSTRY IS UNDERGOING PROFOUND CHANGES AS OUR SOCIETY INCREASINGLY EMPHASIZES THE IMPORTANCE OF A SMARTER GRID IN SUPPORT OF CLEAN AND SUSTAINABLE ENERGY UTILIZATION.

State estimation was initially developed for power systems in the 1970s. Fred Schweppe, a pioneer in this area, defined power system state estimation as “a data processing algorithm for converting redundant meter readings and other available information into the estimate of the state of an electric power system” [2]. In other words, the main function of state estimation is to estimate, through processing the set of real-time redundant measurements, the electrical states of power systems, typically bus voltage magnitudes and phase angles. State estimation is a key function in building a real-time network model in the energy management system (EMS) [3], [4]. Specifically, the main tasks of a state estimator are listed below [5], for which we focus on the last one of bad data processing:

- *Observability analysis*: To determine whether a unique estimate can be found for the system state, generally prior to state estimation.
- *State estimation*: To determine the optimal estimate for the complex voltages at each bus based on the real-time analog measurements.
- *Bad data processing*: To detect measurement errors and bad data injections; to identify and eliminate them if possible.

Conventional bad data detection techniques are typically based on gross errors appearing in the measurement residuals [5]. While relatively effective against random noises, these detectors lack the ability to detect highly structured bad data that conforms to the network topology and some particular physical laws. This raises serious security concerns about intentional stealth cyberattacks that can tamper with the measurements without being detected. As



[FIG1] Distributed topology for the future smart grid (dotted lines depict the cyberpart).

more and more advanced cybertechnologies are being integrated into the EMS, such potential cyberattack threats are becoming a major security concern for regional transmission organizations (RTOs) [6].

SYSTEM MODEL AND INJECTED DATA DETECTION

In this section, we first discuss the system model for state estimation and introduce the linearized model. Then, we focus on the data-injection attack and the classic detection mechanism. Finally, we define a special type of stealthy attack.

[STATE ESTIMATION IS A KEY FUNCTION IN BUILDING A REAL-TIME NETWORK MODEL IN THE ENERGY MANAGEMENT SYSTEM.]

SYSTEM MODEL

We adopt the direct current (dc) power flow model (widely utilized to simplify the system analysis; see [1], [2], [5], and [10])

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

where $\mathbf{H} \in \mathbb{R}^{M \times N}$ is the dc power flow matrix, $\mathbf{z} \in \mathbb{R}^M$, $\mathbf{x} \in \mathbb{R}^N$, and $\mathbf{e} \in \mathbb{R}^M$ are the measurement signal vector, the system state vector, and the measurement noise vector, respectively, with the covariance matrix for $\mathbf{e}$ given as Σ_e.

BAD DATA INJECTION AND DETECTION

We now introduce the general data-injection attack problem. Based on the linearized model, the cyberdata injection attacks can be modeled as [1]

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

where $\mathbf{c}$ accounts for the bad data injected by the attacker. The attack could be launched by one single attacker, or by a group of coordinated attackers that inject the bad data based on their collective information about the network. We assume that $\mathbf{H}$ is fully known to the system operators, and may or may not be known to the attackers. The objective of the defender (system operators) is to reliably detect an injection attack (the existence of $\mathbf{c}$) in the event of an attack.

The detection dynamics can be expressed by the following composite hypothesis testing problem with $\mathcal{H}_0$ representing the no-attack hypothesis and $\mathcal{H}_1$ representing the attack hypothesis:

$$\mathcal{H}_0: \quad \mathbf{c} = \mathbf{0} \quad \text{versus} \quad \mathcal{H}_1: \quad \mathbf{c} \neq \mathbf{0}.$$

We seek to devise a mechanism that distinguishes between $\mathcal{H}_0$ and $\mathcal{H}_1$ reliably. We denote the true hypothesis and the decision of the detector by $T \in \{\mathcal{H}_0, \mathcal{H}_1\}$ and $D \in \{\mathcal{H}_0, \mathcal{H}_1\}$, respectively. Therefore, the probabilities of misdetection and false alarm are respectively given by $P_{\text{mis}} = P(D = \mathcal{H}_0 \mid T = \mathcal{H}_1)$ and $P_{\text{fa}} = P(D = \mathcal{H}_1 \mid T = \mathcal{H}_0)$.

STEALTH ATTACK

From the data-injection attack model given in the previous subsection, we observe that if the attacker has knowledge on $\mathbf{H}$, it can add $\mathbf{c} = \mathbf{H}\mathbf{b}$. As a result, we have

$$\mathbf{z} = \mathbf{H}(\mathbf{x} + \mathbf{b}) + e, \tag{2}$$

such that the control center believes that the true state is $\mathbf{x} + \mathbf{b}$. This is called stealth bad data injection [5], i.e., if the attack vector lies in the range of $\mathbf{H}$, it is not detectable by traditional statistical tests. In the section "Three New Perspectives on Future Detection Approaches," we will discuss how to mitigate this issue.

CLASSICAL APPROACHES AGAINST DATA-INJECTION ATTACKS

In this section, we give a survey of the various classical approaches against data-injection attacks. The objectives of the adversaries are not only to obtain some unauthorized information but also to paralyze the power facility by misleading the EMS with injected bad data. The EMS in the control center depends critically on estimating the system states based on data periodically collected from remote meters. If the adversaries are able to hack into the power grid and inject malicious data, the EMS may produce a false state estimate, which could potentially lead to wrong control decisions and may cause large-scale malfunction. Thus, the smart grid needs to be capable of detecting and preventing such attacks. The classical approaches are mainly based on two mechanisms: 1) adopting advanced signal processing techniques in the control center to detect data-injection attacks and 2) deploying advanced measurement units such as PMUs at various locations to reduce the chance of being subject to data-injection attacks. Such mechanisms will be reviewed next.

BAD DATA DETECTION AT CONTROL CENTER

In practice, bad data in the state estimates may be caused by two possible sources: 1) nature (e.g., some extreme weather conditions) or some faulty nodes and 2) man-made data injection. Although today's data-injection attacks usually refer to the second case, most classical approaches for detecting such attacks are based on the results developed for the first case. As such, we first review the results for the first case in the following before we review the results for the second case.

CLASSIC BAD DATA DETECTION

Classic bad data detection techniques try to detect the abnormality in the state vector estimates, which are usually caused by either nature or some faulty nodes [7]. Given the power flow measurements $\mathbf{z}$, the estimated state vector $\hat{\mathbf{x}}$ can be computed as [7]

$$\hat{\mathbf{x}} = (\mathbf{H}^T \Sigma_e^{-1} \mathbf{H})^{-1} \mathbf{H}^T \Sigma_e^{-1} \mathbf{z}. \tag{3}$$

Thus, the residue vector $\mathbf{r}$ can be computed as the difference between the measured quality and the calculated value from the estimated state

$$\mathbf{r} = \mathbf{z} - \mathbf{H}\hat{\mathbf{x}}. \tag{4}$$

Therefore, the expected value and the covariance of the residual are

$$E(\mathbf{r}) = \mathbf{0}$$

and

$$\mathrm{cov}(\mathbf{r}) = [\mathbf{I} - \mathbf{H}(\mathbf{H}^T\Sigma_e^{-1}\mathbf{H})^{-1}\mathbf{H}^T\Sigma_e^{-1}]\Sigma_e,$$

respectively.

Bad data detection due to faulty sensors and topological errors can be performed using a threshold test over $\mathbf{r}$ [7], in which the normal state hypothesis is accepted if $\max_i |r_i| \le \gamma$, where γ is the threshold and r_i is the component of $\mathbf{r}$; otherwise, the abnormal state hypothesis is accepted. Other methods such as the χ^2 method have also been proposed in the literature [5].

DATA-INJECTION ATTACK DETECTION

Man-made data-injection attacks against power grid state estimation were introduced in [1]. By leveraging the knowledge of the power network topology, it was shown [1] that a stealth data-injection attack could bypass the bad data detection scheme in today's SCADA system, which is similar to that described in the previous subsection.

To overcome this security challenge, the network designer needs to quantify the security level against the data-injection attacks. In [8] and [9], security indices were introduced for state estimators in power networks. Such indices help with locating power flows whose measurements are potentially easy to manipulate, where a larger index value indicates that a stronger coordinated attack is needed in order not to trigger an alarm in the control center. With such indices, the security bottleneck in the smart grid could be identified. Meanwhile, algorithms were proposed in [9] to place encrypted devices in the system to maximize the system security index level against stealth attacks.

On the other hand, various practical data-injection attack detection algorithms have been designed, mainly with a focus on the detection probability maximization, attack damage control, and stealth attack recovery. In [10], the adversary is assumed to use a graph theoretic approach to launch stealthy malicious data attacks. When a stealth attack vector does not exist due to meter access restrictions, attacks are constructed to minimize the residue [defined in (4)] energy while guaranteeing a certain increase in the mean square error for the state estimate. From the defender's point of view, a computationally efficient algorithm was derived to detect and localize attacks using the generalized likelihood ratio test regularized by an l_1-norm penalty on the strength of the attack.

ADVANCED MEASUREMENT UNITS

As discussed above, in addition to signal processing-based measures to detect data-injection attacks, advanced measurement units could also be deployed to reduce the chance of being attacked. Compared with traditional voltage meters, PMUs [11]–[13] are more advanced (and more expensive) units that are equipped with various security measures. In practice, PMU data is sampled from widely dispersed locations in the smart grid and synchronized from a common time source based on a global positioning system (GPS) radio clock, where a PMU can be a dedicated device or its function can be incorporated into a protective relay or other devices. As such, PMU technologies provide a tool for system operators and planners to measure the state of the electrical system and manage power quality with accurate and coherent time stamps. Consequently, synchronized comparison of the measurements across different locations is possible in real time, which bears inherent robustness against data-injection attacks. Moreover, the communication links between PMUs and data centers are usually secured and encrypted.

[PMU TECHNOLOGIES PROVIDE A TOOL FOR SYSTEM OPERATORS AND PLANNERS TO MEASURE THE STATE OF THE ELECTRICAL SYSTEM AND MANAGE POWER QUALITY WITH ACCURATE AND COHERENT TIME STAMPS.]

Many applications of PMUs in power systems have been studied in the literature. Specifically, the analysis of power system observability with PMUs was conducted in [14]. The optimal placement of PMUs was studied in [15] and [16] to maximize the measurement redundancy at the power system buses with a given number of PMUs. In [17], the nonlinear power system state estimation problem was solved with PMUs making the network completely observable. In [18], PMU linear measurement equations are directly applied and PMU data is incorporated into power flow (current) equations to accomplish the power system state estimation. The placement of PMUs in the power system was further explored in [19] and [20] to enhance power system state estimation, which in turn improves the ability to defend bad data injection.

THREE NEW PERSPECTIVES ON FUTURE DETECTION APPROACHES

In this section, three important aspects to enrich the detection results against data-injection attacks are presented in detail: 1) attacker versus defender dynamics (along the line of [10]), where possible interactive attack and detection strategies are discussed in the context of secured PMU placement; 2) distributed attack detection and state recovery, where the focus is on how to achieve the optimal centralized performance with a distributed approach; and 3) QD, where the tradeoff between the detection speed and detection performance is studied.

ATTACKER VERSUS DEFENDER DYNAMICS

From the section "System Model and Injected Data Detection," we recall that the injected attack vector $\mathbf{c}$ is not detectable if it fully lies in the range of $\mathbf{H}$. In this subsection, we discuss some possible schemes to mitigate this issue; for example, we could assume perfectly secured locations within the measurement vector by installing fully secured PMUs in the system. However, if the attackers have certain knowledge about such possible security measures, they could also optimize their strategies of choosing $\mathbf{c}$. On the other hand, the defender can take counteractions to minimize the worst-case scenario caused by the attackers. These interesting attacker versus defender dynamics need to be carefully addressed to design a smart grid that is robust against data-injection attacks.

THESE INTERESTING ATTACKER VERSUS DEFENDER DYNAMICS NEED TO BE CAREFULLY ADDRESSED TO DESIGN A SMART GRID THAT IS ROBUST AGAINST DATA-INJECTION ATTACKS.

We discuss this issue based on our results from [21], where we assume that the attacker has full knowledge of $\mathbf{H}$ as well as the defender's strategy. We also assume that it is possible to use some idealized PMUs to securely protect a subset of the measurements, preventing the attacker from changing those measurement values in such a subset. In other words, we have some trusted reference points in the measurement vector $\mathbf{z}$. However, the large number of meters in the smart grid makes it impossible to protect all of them with idealized PMUs in practice. Let $\mathcal{S}$ denote the set of indices corresponding to the measurements that are protected, and let $\bar{\mathcal{S}}$ denote the complementary set of $\mathcal{S}$. It is shown in [1] that as long as $\mathbf{c} = \mathbf{Hb}$ holds for any $\mathbf{b}$, the injection attack in (1) does not change the measurement residuals, thereby defeating conventional detection techniques based on statistical tests [5]. However, with the secured set $\mathcal{S}$, the attackers have to guarantee the following countersecurity constraint when they choose the attack vector $\mathbf{H}^{\mathcal{S}}\mathbf{b} = \mathbf{0}$, where $\mathbf{H}^{\mathcal{S}}$ is the matrix composed of the rows in $\mathbf{H}$ indexed by $\mathcal{S}$. It is clear that given an $\mathbf{H}$, as we increase $|\mathcal{S}|$ (the cardinality of $\mathcal{S}$), it is possible to make the above linear equation overdetermined such that the only feasible attack strategy is the trivial choice of $\mathbf{c} = \mathbf{0}$ and $\mathbf{b} = \mathbf{0}$, i.e., no attack.

ATTACKER'S STRATEGY

However, since the installation and maintenance cost for secured measurement points is high, $|\mathcal{S}|$ is usually small. This leads to some sparseness properties, and consequently there usually exist many feasible attack strategies for $\mathbf{c}$. Some recent work on how to find $|\mathcal{S}|$ is given in [22]. Accordingly, the attackers may choose an objective to optimize under some other extra constraints. For example, the attacker may choose to minimize the number of meters that it tampers with to reduce the probability of being detected; meanwhile, the attacker tries to guarantee a minimum distortion at at least one attack position, i.e., $\|\mathbf{b}\|_\infty \geq \tau$, where τ is a predefined threshold. As such, the optimization problem for the single attack case can be constructed as

$$\min_{\mathbf{b}} \|\mathbf{H}^{\bar{\mathcal{S}}}\mathbf{b}\|_0 \quad \text{s.t.} \;\; \mathbf{H}^{\mathcal{S}}\mathbf{b} = \mathbf{0} \text{ and } \|\mathbf{b}\|_\infty \geq \tau. \tag{5}$$

The major issue for the attacker is that finding sparse solutions for the above ℓ_0-minimization problem is in general NP-hard. It is well known, however, that optimizing the ℓ_1 norm helps with promoting sparsity. With some transformation, it has been shown in [21] that a reasonable approach for the attacker is to solve the following ℓ_1 convex relaxation problem:

$$\min_{\mathbf{b}_i} \|\mathbf{H}_i^{\bar{\mathcal{S}}}\mathbf{b}_i + \mathbf{h}_i^{\bar{\mathcal{S}}}\|_1 \quad \text{s.t.} \;\; \mathbf{H}_i^{\mathcal{S}}\mathbf{b}_i + \mathbf{h}_i^{\mathcal{S}} = \mathbf{0}, \tag{6}$$

where $\mathbf{h}_i$ denotes the ith column of $\mathbf{H}$, $\mathbf{H}_i$ denotes the $M \times (N-1)$ matrix formed by removing the ith column from $\mathbf{H}$, and $\mathbf{b}_i \in \mathbb{R}^{N-1}$ denotes the vector formed by removing the ith component b_i from $\mathbf{b}$.

We refer to (6) as a naive ℓ_1-relaxation approach. It is worth pointing out that the relaxation from ℓ_0 to ℓ_1 does not change the constraints that make the attacker successfully evade detection by adopting $\mathbf{H}_i^{\mathcal{S}}\mathbf{b}_i + \mathbf{h}_i^{\mathcal{S}} = \mathbf{0}$. The only suboptimality that the attacker may suffer is that the solution may not be the sparsest, i.e., there may exist other sets of indices to attack with a smaller cardinality.

To mitigate the above issue, as shown in [21], we could instead solve the following weighted optimization problem [23]:

$$\min_{\mathbf{b}_i} \|\mathrm{diag}(\mathbf{w}^i)(\mathbf{H}_i^{\bar{\mathcal{S}}}\mathbf{b}_i + \mathbf{h}_i^{\bar{\mathcal{S}}})\|_1 \quad \text{s.t.} \;\; \mathbf{H}_i^{\mathcal{S}}\mathbf{b}_i + \mathbf{h}_i^{\mathcal{S}} = \mathbf{0}, \tag{7}$$

where $\mathrm{diag}(\mathbf{w}^i)$ denotes a diagonal matrix whose diagonal entries are given by a weight vector $\mathbf{w}^i \in \mathbb{R}^{M-N_S} > \mathbf{0}$ with N_S the size of set $\mathcal{S}$. Note that with $\mathbf{w}^i = \mathbf{1}$, (7) is reduced to (6). The attacker may incorporate certain prior knowledge to set the weights in $\mathbf{w}^i$. It can also try to improve the performance by repeating the process for multiple iterations, with the specific strategy listed in Algorithm 1, where the fixed $\epsilon > 0$ is to regularize division by (near) zero [23].

ALGORITHM 1: AN ATTACKER'S STRATEGY

```
for $i = 1, \ldots, N$ do
    Obtain the initial $\mathbf{b}_i$;
    for a fixed number of iterations do
        Compute $\mathbf{x}^i = \mathbf{H}_i^{\bar{\mathcal{S}}}\mathbf{b}_i + \mathbf{h}_i^{\bar{\mathcal{S}}}$;
        Compute $w_k^i = \frac{1}{|x_k^i| + \epsilon}$, $k = 1, \ldots, M - N_S$;
        Solve the weighted problem (7) to obtain $\mathbf{b}_i$;
    end for
end for
```

DEFENDER'S STRATEGY

If the defender is aware of the possible strategies that the attacker chooses, the defender can search all possible subsets of $\mathcal{S}$ to protect to optimize certain detection performance. In [21], we chose to construct an optimization problem as

ALGORITHM 2: SUBSET SELECTION

```
$\mathcal{S} = \emptyset$;
repeat
    MeasureArr($\{1, \ldots, M\}$) = 0;
    for $i = 1$ to $N$ do
        Find $\mathcal{S}_i$ and $N_{Ai}$;
        if $N_{Ai} < N_A$ then
            MeasureArr($\mathcal{S}_i$) $\Leftarrow$ MeasureArr($\mathcal{S}_i$) + 1;
        end if
    end for
    $k^\star = \mathrm{argmax}_k$MeasureArr($k$);
    Add $k^\star$ to $\mathcal{S}$;
until $N_{Ai} \geq N_A$, $\forall i$
```

follows. Given an attacker strategy, let N_{Ai} be the minimum number of measurements that the attacker needs to control to inject bad data into the state of bus i without being detected. We seek to solve

$$\min_{\mathcal{S}} |\mathcal{S}| \quad \text{s.t.} \quad \min_{i\in\{1,\ldots,N\}} N_{Ai} \geq N_A, \tag{8}$$

where N_A is a positive integer as the threshold. Thus, the above formulation attempts to meet a certain level of resilience with the minimum protection cost.

To solve the above combinatorial problem, we propose a suboptimal but efficient algorithm that adds one measurement into the set $\mathcal{S}$ at a time, until all the conditions of $N_{Ai} \geq \mathrm{N_A}$ are met. The algorithm is presented in Algorithm 2.

IN OTHER WORDS, SINCE ELECTRIC GRIDS ARE TYPICALLY SPARSE, PROTECTING ONE MEASUREMENT CAN TYPICALLY IMPACT ONLY A FEW STATES.

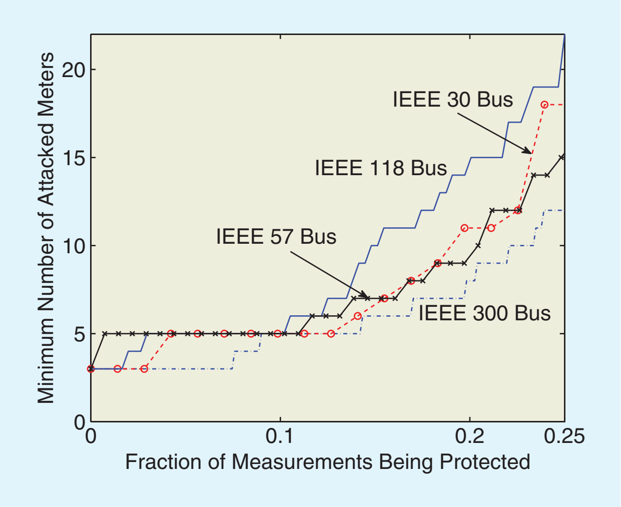
At each iteration, the algorithm emulates attacks under the current secured subset $\mathcal{S}$, which is initialized to be empty, assuming a specific attacking strategy. The key idea of the algorithm is that it maintains an array of counters, MeasureArr, counting the number of times that each measurement is manipulated by the attacker. We count only when $N_{Ai} < N_A$, i.e., when the condition on the minimum number of measurements being attacked is not met. The algorithm then determines which measurement is modified the most and moves it to the protected set $\mathcal{S}$. The maximizer may not be unique. In such cases, the algorithm chooses a random index among all the optimizers. The complexity order of the algorithm is equivalent to that of solving $N \times N_S$ linear programs, as compared to solving $\binom{M}{N_S}$ linear programs in the exhaustive search. Note that the proposed algorithm does not necessarily converge to a global optimum. In the numerical results, we run the algorithm several times with random initializations for each N_A and choose the output $\mathcal{S}$ with the smallest cardinality.

For simulation, we use MATPOWER 4.0, which is widely accepted as a valid simulation tool in power system analysis. In Figure 2, we plot the minimum number of measurements that the attacker needs to manipulate to change at least one state variable without being detected as a function of the measurement faction being protected. The sets of protected measurements are designed by Algorithm 2. For each value of N_A, we run the algorithm three times. As can be seen, all the test systems display a relatively similar behavior. At first it is fairly expensive to protect the systems, as the attacker needs to control only a few more meters to evade detection even if the designer can protect up to 10% of the measurements. Afterward, the cost of protection decreases significantly as indicated by the steeper slopes of the curves. For example, by protecting 25% of the measurements on the IEEE 57-bus system, we can force the attacker to control at least 15 m to succeed, a fivefold increase over not using any form of protection. The high initial cost of protection perhaps can be attributed to the fact that there typically is a large fraction of the states having very few associated measurements as illustrated in Figure 3. In other words, since electric grids are typically sparse, protecting one measurement can typically impact only a few states.

DISTRIBUTED DETECTION AND STATE RECOVERY

For a large grid, the attack detection and the associated state estimation problem can be computationally sophisticated since the number of the states can be large [24], which can be alleviated by distributing the computation into several processing nodes. The goal of this subsection is to introduce a distributed algorithm (denoted as $\mathcal{DA}$) from [24], in which each node converges almost surely to the optimal solution under certain conditions.

As we discussed earlier, if the attack vector **c** lies in the span of **H**, it is not detectable. Therefore, all the useful information for detection is contained in the projection of **z** onto the null space of **H**, which is denoted as **y** and given as



[FIG2] N_A versus the fraction of protected measurements.

$$\mathbf{y} = [\mathbf{I} - \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T]\mathbf{z}, \tag{9}$$

by assuming $\mathbf{H}$ is full rank, i.e., the system is observable. The overall detection performance is then dependent on the sufficient statistic $\mathbf{y}$. From now on, we focus on a distributed approach to estimate $\mathbf{y}$, and leave the other details on the joint detection and estimation problem to [24].

The distributed observation model at the ith node is give by

$$\mathbf{z}_i = \widetilde{\mathbf{H}}_i\mathbf{x} + \mathbf{e}_i + \mathbf{c}_i, \tag{10}$$

where $\mathbf{z}_i$ is the local observation vector at the ith node, $\widetilde{\mathbf{H}}_i$ corresponds to the local Jacobian matrix, with $\mathbf{e}_i$ and $\mathbf{c}_i$ being the noise and attack vector respectively influencing the measurements at node i.

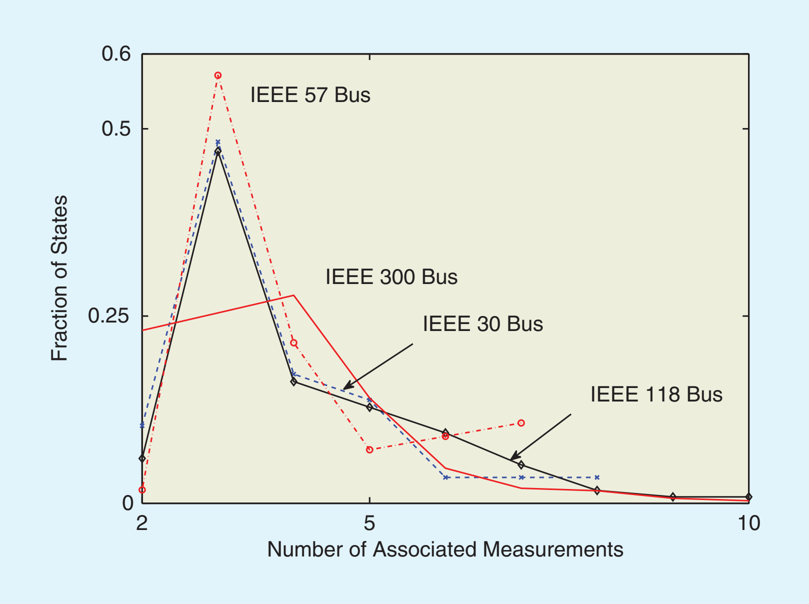
We make the following assumption on global observability.

(E.1)–Observability: The matrix $\mathbf{G} = \sum_{i=1}^{N} \widetilde{\mathbf{H}}_i^T \widetilde{\mathbf{H}}_i$ is of full-rank.

Starting from an initial deterministic estimate of $\mathbf{y}$ (the initial state may be random, but here we assume that it is deterministic for notational simplicity) at node i, denoted by $\mathbf{y}_i(0)$, each node generates by a distributed iterative algorithm a sequence of estimates, $\{\mathbf{y}_i(n)\}_{n\geq 0}$. The estimate $\mathbf{y}_i(n+1)$ of $\mathbf{y}$ in the ith node at time $n+1$ is a function of

- its previous estimate
- the communicated estimates at time n from its neighboring nodes
- the local observation $\mathbf{z}_i$.

FOR A LARGE GRID, THE ATTACK DETECTION AND THE ASSOCIATED STATE ESTIMATION PROBLEM CAN BE COMPUTATIONALLY SOPHISTICATED SINCE THE NUMBER OF THE STATES CAN BE LARGE.



[FIG3] The fraction of states as a function of the number of associated measurements for different test systems.

ALGORITHM $\mathcal{DA}$

Based on the current state $\mathbf{y}_i(n)$, the exchanged data $\{\mathbf{y}_l(n)\}_{l\in\Omega_n}$ from the communication neighborhood node set Ω_n, and the observation $\mathbf{z}_i$, we update the estimate at the ith node by the following distributed iterative algorithm:

$$\mathbf{y}_i(n+1) = \mathbf{y}_i(n) - \left\{\gamma(n)\sum_{l\in\Omega_n}[\mathbf{y}_i(n) - \mathbf{y}_l(n)]\right\}$$
$$-\gamma(n)\mathcal{P}_i^T\left[\mathbf{z}_i - \widetilde{\mathbf{H}}_i\hat{\mathbf{y}}_i(n) - \mathcal{P}_i\mathbf{y}_i(n)\right], \tag{11}$$

where $\mathcal{P}_i$ is an $M_i \times M$ selection matrix that selects the components of $\mathbf{y}_i(n)$ corresponding to the location of $\mathbf{z}_i$ in the vector $\mathbf{z}$, with $\mathbf{z}_i \in \mathbb{R}^{M_i}$ and $\mathbf{z} \in \mathbb{R}^M$, and the auxiliary state sequence $\{\hat{\mathbf{y}}_i(n)\}$ at node i is generated according to a distributed scheme,

$$\hat{\mathbf{y}}_i(n+1) = \hat{\mathbf{y}}_i(n) - \left\{\beta(n)\sum_{l\in\Omega_i}[\hat{\mathbf{y}}_i(n) - \hat{\mathbf{y}}_l(n)]\right.$$
$$\left.-\alpha(n)\widetilde{\mathbf{H}}_i^T\left[\mathbf{z}_i - \widetilde{\mathbf{H}}_i\hat{\mathbf{y}}_i(n)\right]\right\}. \tag{12}$$

In (12), $\{\gamma(n)\}$, $\{\alpha(n)\}$, and $\{\beta(n)\}$ are appropriately chosen time-varying weight sequences, which we will discuss later. Algorithm $\mathcal{DA}$ is distributed since at node i it involves only the data from the nodes in its communication neighborhood Ω_i. To implement $\mathcal{DA}$, each node stores and updates two states: $\mathbf{y}_i(n)$, the estimate of $\mathbf{z}$; and $\hat{\mathbf{y}}_i(n)$, an auxiliary state used for the update of $\mathbf{y}_i(n)$.

We note that the estimate sequence $\{\mathbf{y}_i(n)\}$ is random, due to the stochasticity of the noise. The following assumptions on the connectivity of the inter-node communication network are assumed.

(E.2)–Connectivity: The internode communication network determined by the communication neighborhoods Ω_i is connected.

(E.3)–Time varying weights: The sequences $\{\alpha(n)\}$ and $\{\beta(n)\}$ are of the forms $\alpha(n) = ((a)/(n+1)^{\tau_1})$ and $\beta(n) = ((b)/(n+1)^{\tau_2})$ respectively, where $a, b > 0$ are constants and the exponents τ_1 and τ_2 satisfy $0 < \tau_1 \leq 1$ and $0 < \tau_2 < \tau_1$. The sequence $\gamma(n)$ is of the form $\gamma(n) = ((c)/(n+1)^{\tau_3})$, where $c > 0$ is a constant and the exponent τ_3 satisfies $0 < \tau_3 \leq 1$.

A key thing to note is that, although the weights are decaying over time, i.e., $\gamma(n)$, $\alpha(n)$, $\beta(n) \to 0$ as $n \to \infty$, they are persistent, i.e., $\sum_{n\geq 0}\alpha(n) = \infty$, $\sum_{n\geq 0}\beta(n) = \infty$, and $\sum_{n\geq 0}\gamma(n) = \infty$. Whereas the decaying nature of the weight sequences guarantee convergence, the persistence condition is necessary to drive the estimators to $\mathbf{y}$ from arbitrary initial conditions.

The following result characterizes the convergence property of the proposed algorithm $\mathcal{DA}$.

THEOREM 1

Consider the $\mathcal{DA}$ under (E.1)–(E.3). Then, for each i, the estimate sequence $\{\mathbf{y}_i(n)\}$ converges almost surely to the sufficient statistic as $n \rightarrow \infty$.

The convergence rate in Theorem 1 depends on the choice of the various weight sequences. The proven convergence [24] of the above theorem allows each node to compute in a distributed way the centralized sufficient statistic needed for the construction of the optimal detector-estimator of the attack vector $\mathbf{c}$. For further details, please refer to [24].

FOR ATTACK DETECTION, IN MANY SCENARIOS THE DETECTION SPEED IS CRITICAL TO GUARANTEE A TIMELY RESPONSE IN THE GRID.

QUICKEST DETECTION

For attack detection, in many scenarios the detection speed is critical to guarantee a timely response in the grid. As such, it is important to study the tradeoff between the detection reliability and the detection speed and to operate on the optimal tradeoff curve. Such a goal can be fulfilled within the QD framework [25]. Specifically, QD attempts to determine an abnormal state change as quickly as possible based on real-time observations such that certain user-defined conditions are satisfied while maintaining a certain level of detection accuracy. The user-defined conditions are known as decision rules, which optimize the tradeoff between the stopping time and the decision accuracy. Here, we formulate the bad data detection problem as an adversary detection problem. The techniques to cover are mainly the cumulative sum test and the generalized likelihood ratio test (for others please refer to [25]).

Suppose that the control center monitors a subdivision of a smart grid with active buses. Here, we assume that the attacker does not know $\mathbf{H}$. When the system is in the normal state (no adversary), we assume a Bayesian model of the random state variable $\mathbf{z}$ with a multivariate Gaussian distribution $\mathcal{N}(\mu_z, \Sigma_z)$. Here $\mu_z = \mathbf{0}$, and $\Sigma_z = \mathbf{H}\Sigma_x\mathbf{H}^T + \Sigma_e$ with covariance Σ_x of the state $\mathbf{x}$ and covariance Σ_e of the noise $\mathbf{e}$. We use n to denote the observation time index. The adversary is assumed to be inactive initially; at a random and unknown observation time τ, it becomes active to inject the malicious data. The binary hypotheses can be formulated as

$$\begin{cases} \mathcal{H}_0: & \mathbf{z}_n \sim \mathcal{N}(0, \Sigma_z), \\ \mathcal{H}_1: & \mathbf{z}_n \sim \mathcal{N}(\mathbf{a}_n, \Sigma_z), \end{cases}$$

where $\mathbf{a}_n = [a_{n,1}, a_{n,2}, \ldots, a_{n,M}]^T \in \mathbb{R}^M$ is the unknown attack vector to inject malicious data at observation time n. In other words, we need to detect a change of the distribution from $\mathcal{N}(0, \Sigma_z)$ to $\mathcal{N}(\mathbf{a}_n, \Sigma_z)$ at an unknown time τ.

Based on Lorden's formulation [25], we minimize the worst case of detection delay T_d, which can be described as

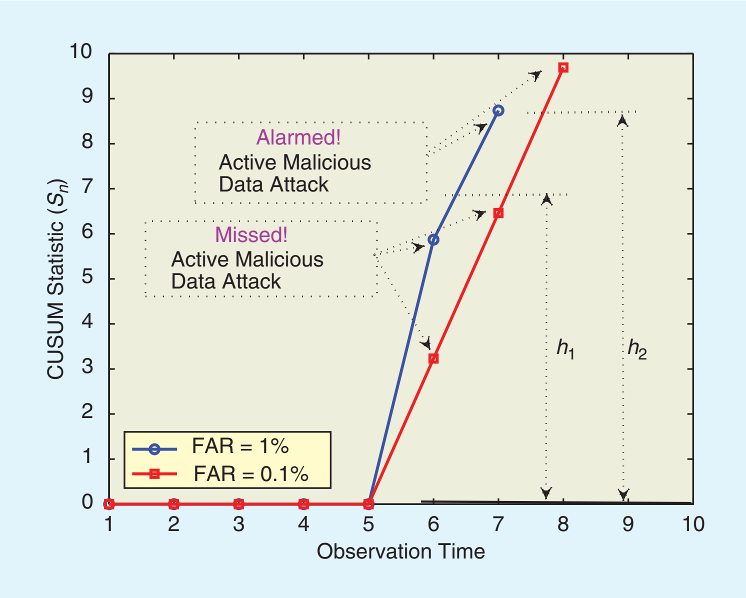
$$T_d = \sup_{\tau \geq 1} E_\tau [T_h - \tau | T_h \geq \tau], \tag{13}$$

where T_h is the decision time. To find the minimum T_d, Page's cumulative sum (CUSUM) algorithm is the best-known technique [25]. However, most CUSUM-based models assume perfect knowledge of the likelihood functions. In the scenario of attacks, the parameters of the $\mathcal{H}_1$ distribution usually cannot be completely defined. Thus, we need to deploy certain adaptive techniques to solve the unknown parameter issue in Page's CUSUM detection scheme, which we call the adaptive CUSUM test.

The proposed QD algorithm, the adaptive CUSUM test, is recursive in nature, where each recursion comprises two interleaved steps: 1) a linear unknown parameter solver and 2) a nonparametric multithread CUSUM, with the multithread CUSUM test modified from Page's CUSUM algorithm. The nonparametric multithread CUSUM test considers and incorporates the likelihood ratio terms of M measurements at observation time n to determine the stopping time T_h. The cumulative statistic at observation n is given by

$$S_n = \max_{1 \geq k \geq T_h} \sum_{n=k}^{T_h} L_n, \quad L_n(\mathbf{z}_n) = \sum_{l=1}^{M} \log \frac{f_1(\mathbf{z}_n | \mathbf{a}_n)}{f_0(\mathbf{z}_n)},$$

where L_n is the sum of the likelihood ratios for all measurements $(z_{n,l}, l \in 1, 2, \ldots, M)$ of the vector $\mathbf{z}_n$ at observation time n, $f_1(\mathbf{z}_n|\mathbf{a}_n)$ is the $\mathcal{H}_1$ multivariate normal distribution



[FIG4] **Example of QD for bad data injection.**

THE BAD DATA DETECTION PROBLEM IS A CHALLENGING ONE WITH THE GENERAL CASES STILL OPEN.

while the adversary actively injects malicious data with mean $\mathbf{a}_n$, and $f_0(\mathbf{z}_n)$ is the $\mathcal{H}_0$ multivariate normal distribution in the normal state. At observation time n, the cumulative statistic S_n can be computed recursively and described as

$$S_n = \max[S_{n-1} + L_t(\mathbf{z}_n), 0], \tag{14}$$

where $S_0 = 0$. The control center declares an alarm when the accumulation crosses a certain threshold h, i.e.,

$$\begin{cases} \text{declare } \mathcal{H}_1, & \text{if } S_n \geq h; \\ \text{declare } \mathcal{H}_0, & \text{otherwise.} \end{cases}$$

Notice that the value of h determines the accuracy of detection. If we require a higher accuracy level, h should be larger.

However, in practice we do not know the value of $\mathbf{a}_n$, i.e., how the attacker attacks. To overcome this problem, we consider the Rao test [26], which is asymptotically equivalent to the generalized likelihood ratio test. The derivation of the Rao test is similar to that of the locally most powerful test, where the Rao test involves straightforward calculation by taking derivatives with respect to the unknown parameter evaluated at zero, leading to much lower complexity than that for maximum likelihood estimation.

The CUSUM performance is illustrated in Figure 4, which shows the evolution of the CUSUM statistic versus the observation time n. Specifically, Case 1 with false alarm rate (FAR) of 1% corresponds to threshold h_1, and Case 2 with FAR of 0.1% corresponds to threshold h_2. The attack starts from Time 6. The proposed algorithm signals the alarm and then terminates the process at T_h = 7 and 8 respectively, when S_n passes the thresholds.

CONCLUSIONS

In this article, we have discussed the coordinated data-injection attack and detection problem in the smart grid. In addition to the basic problem formulation and a literature survey on existing solutions, we have presented three important aspects through which we can enrich the detection solutions. Specifically, we first discussed the possible attacker versus defender dynamics in the context of bad data injection and detection, where an optimal formulation and heuristic algorithms are given to derive various strategies. We then focused on the distributed implementation issue, where a distributed algorithm is proposed with a guaranteed convergence to the centralized solution. Finally, we introduced the QD approach, with which we could explore the optimal tradeoff between the detection speed and reliability.

The bad data detection problem is a challenging one with the general cases still open. In the following, we list several possible directions for future work:

1) In most current approaches, for the system measurement model, the statistics of the measurement noise are usually assumed to be perfectly known. The more practical cases with unknown or nonperfectly known noise statistics are of interest.

2) In the attacker versus defender dynamics, the more general cases where the attacker knows only partial information about **H** and/or where the defender knows only partial information about the attacker's strategies are worth investigation.

3) In the QD approach, methods for designing a distributed scheme to implement the proposed algorithm is another promising direction.

ACKNOWLEDGMENTS

This work was in part supported by DoD-AFOSR under grant FA9550-09-1-0107, DoD-DTRA under grant HDTRA1-07-1-0037, ARO under MURI grant W911NF-11-1-0036, ONR under grant N00014-09-1-0342, Qatar National Research Fund, and NSF under grants DMS-1118605 CNS-1117560, ECCS-1028782, CNS-0953377, and CNS-0905556.

AUTHORS

Shuguang Cui (cui@ece.tamu.edu) received his Ph.D. degree in electrical engineering from Stanford University, California, in 2005. He is now an associate professor in the Department of Electrical and Computer Engineering at Texas A&M University, College Station. His current research interests include resource allocation for cognitive networks, network information theory, statistical signal processing for complex systems, and general communication theories. He has been an associate editor for *IEEE Transactions on Signal Processing, IEEE Transactions on Wireless Communications, IEEE Communication Letters*, and *IEEE Transactions on Vehicular Technology*. He is an elected member of the IEEE Signal Processing Society's SPCOM Technical Committee (2009–2015) and is a Member of the IEEE.

Zhu Han (zhan2@mail.uh.edu) received his B.S. degree in electronic engineering from Tsinghua University in 1997 and his Ph.D. degree in electrical engineering from the University of Maryland, College Park, in 2003. Currently, he is an assistant professor in the Department of Electrical and Computer Engineering at the University of Houston. He received the 2010 NSF CAREER Award as well as the 2011 IEEE Communications Society Fred W. Ellersick Prize. He coauthored papers that won a best paper award at the 2009 IEEE International Conference on Communications and the 7th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks. He is a Senior Member of the IEEE.

Soummya Kar (soummyak@ece.cmu.edu) received the B.Tech. degree in electronics and electrical communication engineering from the Indian Institute of Technology, Kharagpur, in 2005, and the Ph.D. degree in electrical and

computer engineering from Carnegie Mellon University, in Pittsburgh, Pennsylvania, in 2010. From 2010 to 2011, he was with the electrical engineering department at Princeton University as a postdoctoral research associate. He is currently an assistant research professor of electrical and computer engineering at Carnegie Mellon University. His research interests include performance analysis and inference in large-scale networked systems, adaptive stochastic systems, stochastic approximation, and large deviations. He is a Member of the IEEE.

Tùng T. Kim (thanhkim@princeton.edu) received the B.Eng. degree in electronics and telecommunications from Hanoi University of Technology, Vietnam, in 2001, and the M.S. and Ph.D. degrees in electrical engineering from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 2004 and 2008, respectively. He held visiting positions at the University of Southern California, Los Angeles, in 2007 as well as the University of Cambridge, United Kingdom, in 2008. He is currently a postdoctoral research associate with the Department of Electrical Engineering, Princeton University, New Jersey. His research interests include information theory and signal processing with applications in wireless communications and smart grid systems. He is a Member of the IEEE.

H. Vincent Poor (poor@princeton.edu) is the dean of engineering and applied science at Princeton University, where he is also the Michael Henry Strater University Professor of Electrical Engineering. His interests include the areas of statistical signal processing, stochastic analysis and information theory, with applications in several fields including the smart grid. Among his publications is the recent book *Smart Grid Communications and Networking* (Cambridge, 2012). Recent recognition includes the 2010 Institution of Engineering and Technology Fleming Medal, the 2011 IEEE Sumner Award, and the 2011 IEEE Signal Processing Society Award. He is an IEEE Fellow and a member of the National Academy of Engineering, the National Academy of Sciences, and the Royal Academy of Engineering.

Ali Tajer (tajer@princeton.edu) is a postdoctoral research associate at Princeton University. He received the Ph.D. degree in electrical engineering and the M.A. degree in statistics from Columbia University and the M.Sc. and B.Sc. degrees in electrical engineering from Sharif University of Technology. His research interests include the areas of applied statistics, information theory, and signal processing with applications in wireless communications and smart grids. He is a Member of the IEEE.

REFERENCES

[1] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proc. 16th ACM Conf. Computer and Communications Security*, Chicago, IL, Nov. 2009, pp. 21–32.

[2] F. C. Scheweppe, J. Wildes, and A. Bose, "Power system static state estimation—Parts I, II and III," *IEEE Trans. Power App. Syst.*, vol. 89, no. 1, pp. 120–135, Jan. 1970.

[3] F. F. Wu, K. Moslehi, and A. Bose, "Power system control centers: Past, present, and future," *Proc. IEEE*, vol. 93, no. 11, pp. 1890–1908, Nov. 2005.

[4] F. Monticelli, "Electric power system state estimation," *Proc. IEEE*, vol. 88, no. 2, pp. 262–282, Feb. 2000.

[5] A. Abur and A. Gomez-Exposito, *Power System State Estimation: Theory and Implementation*. New York: Marcel Dekker, 2004.

[6] J. Weiss, *Protecting Industrial Control Systems from Electronic Threats*. New York: Momentum Press, 2010.

[7] 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.

[8] H. Sandberg, A. Teixeira, and K. H. Johansson, "On security indices for state estimators in power networks," in *Proc. 1st Workshop Secure Control Syst.*, CPS WEEK, Stockholm, Sweden, Apr. 2010, pp. 1–6.

[9] G. Dan and H. Sanberg, "Stealth attacks and protection schemes for state estimators in power systems," in *Proc. Int. IEEE Conf. Smart Grid Communications*, Gaithersburg, MD, Oct. 2010.

[10] O. Kosut, L. Jia, R. Thomas, and L. Tong, "Malicious data attacks on smart grid state estimation: Attack strategies and countermeasures," in *Proc. IEEE Int. Conf. Smart Grid Communications*, Gaithersburg, MD, Oct. 2010, pp. 220–225.

[11] A. G. Phadke, "Synchronized phasor measurements—A historical overview," in *Proc. IEEE/PES Transmission and Distribution Conf. Exposition*, Yokohama, Japan, Oct. 2002, pp. 476–479.

[12] A. G. Phadke, "Synchronized phasor measurements in power systems," *IEEE Comput. Appl. Power*, vol. 6, pp. 10–15, Apr. 1993.

[13] A. Giani, E. Bitar, M. Garcia, M. McQueen, P. Khargonekar, and K. Poolla, "Smart grid data integrity attacks: Characterizations and countermeasures," in *Proc. IEEE Int. Conf. Smart Grid Communications*, Brussels, Belgium, Oct. 2011, pp. 232–237.

[14] B. Xu and A. Abur, "Observability analysis and measurement placement for systems with PMUS," in *Proc. IEEE PES Power Systems Conf. Exposition*, New York, NY, Oct. 2004, vol. 2, pp. 943–946.

[15] S. Chakrabarti, E. Kyriakides, and D. G. Eliades, "Placement of synchronized measurements for power system observability," *IEEE Trans. Power Delivery*, vol. 24, no. 1, pp. 12–19, Jan. 2009.

[16] T. L. Baldwin, L. Mili, Jr., M. B. Boisen, and R. Adapa, "Power system observability with minimal phasor measurement placement," *IEEE Trans. Power Syst.*, vol. 8, no. 2, pp. 707–715, May 1993.

[17] C. Rakpenthai, S. Premrudeepreechacharn, S. Uatrongjit, and N. R. Watson, "PMU-based two stages state estimation for power system with nonlinear devices," in *Proc. Int. Power Engineering Conf. (IPEC 2007)*, Singapore, Dec. 2007, pp. 153–158.

[18] J.-C. Ding, "Influences of measured values from phasor measurement units on equivalent current based measurement transform algorithm in state estimation," *Power Syst. Technol.*, vol. 29, no. 5, pp. 47–51, May 2005.

[19] F. Chen, X. Han, Z. Pan, and L. Han, "State estimation model and algorithm including PMU," in *Proc. 3rd Int. Conf. Electric Utility Deregulation and Restructuring and Power Technol. (DRPT 2008)*, Nanjing, China, Apr. 2008, pp. 1097–1102.

[20] M. J. Rice and G. T. Heydt, "Power systems state estimation accuracy enhancement through the use of PMU measurements," in *Proc. PES TD 2005/2006*, Dallas, TX, May 2006, pp. 161–165.

[21] T. T. Kim and H. V. Poor, "Strategic protection against data injection attacks on power grids," *IEEE Trans. Smart Grid*, vol. 3, no. 2, pp. 326–333, June 2011.

[22] R. B. Bobba, K. M. Rogers, Q. Wang, H. Khurana, K. Nahrstedt, and T. J. Overbye, "Detecting false data injection attacks on DC state estimation," in *Proc. 1st Workshop on Secure Control Systems* (SCS'10), Stockholm, Sweden, Apr. 2010, pp. 1–9.

[23] E. J. Candes, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1-minimization," *J. Fourier Anal. Applicat.*, vol. 14, pp. 877–905, Dec. 2008.

[24] A. Tajer, S. Kar, H. V. Poor, and S. Cui, "Distributed joint cyber attack detection and state recovery in smart grids," in *Proc. IEEE Int. Conf. Smart Grid Communications*, Brussels, Belgium, Oct. 2011, pp. 202–207.

[25] H. V. Poor and O. Hadjiliadis, *Quickest Detection*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[26] A. D. Maio, "Rao test for adaptive detection in Gaussian interference with unknown covariance matrix," *IEEE Trans. Signal Processing*, vol. 55, no. 7, pp. 3577–3584, July 2007.

[SP]