Statistical Structure Learning to Ensure Data Integrity in Smart Grid

Hanie Sedghi, *Student Member, IEEE*, and Edmond Jonckheere, *Fellow, IEEE*

***Abstract*—Robust control and management of the grid relies on accurate data. Both phasor measurement units and remote terminal units are prone to false data injection attacks. Thus, it is crucial to have a mechanism for fast and accurate detection of tampered data—both for preventing attacks that may lead to blackouts, and for routine monitoring and control of current and future grids. We propose a decentralized false data injection detection scheme based on the Markov graph of the bus phase angles. We utilize the conditional covariance test CMIT to learn the structure of the grid. Using the dc power flow model, we show that, under normal circumstances, the Markov graph of the voltage angles is consistent with the power grid graph. Therefore, a discrepancy between the calculated Markov graph and learned structure should trigger the alarm. Our method can detect the most recent stealthy deception attack on the power grid that assumes knowledge of the bus-branch model of the system and is capable of deceiving the state estimator; hence damaging power network control, monitoring, demand response, and pricing scheme. Specifically, under the stealthy deception attack, the Markov graph of phase angles changes. In addition to detecting a state of attack, our method can detect the set of attacked nodes. To the best of our knowledge, our remedy is the first to comprehensively detect this sophisticated attack and it does not need additional hardware. Moreover, it is successful no matter the size of the attacked subset. Simulation of various power networks confirms our claims.**

***Index Terms*—Bus phase angles, conditional covariance test, false data injection detection, structure learning.**

I. Introduction

Among the attributes that make the grid "smart" is its ability to process a massive amount of data for monitoring, control, and maintenance purposes. In a typical transmission system operator (TSO), the substation remote terminal units (RTUs) read the status of voltages, currents, and switching states. The RTU data is redirected in data-packages to the supervisory control and data acquisition (SCADA) system via communication channels. In addition, synchronous phasor measurement units (PMUs) are being massively deployed throughout the grid. PMUs provide a higher level of detail to the SCADA system (e.g., voltage angle). The signals from the PMUs are transmitted via the RTU to the SCADA. The state estimator (SE) located at the control center aims to find the best overall snapshot solution based on all measurements.

Recent monitoring and control schemes rely primarily on PMU measurements; for example, [1] tried to increase voltage resilience to avoid voltage collapse by using synchronized PMU measurements and decision trees and [2]–[4] rely on PMUs for fault detection and localization.

The centralization of the data to the SE makes it the back door to false data injection attacks. Therefore, aforementioned methods can be deluded by false data injection attacks. Thus, it is crucial to have a mechanism for fast and accurate discovery of malicious tampering; both for preventing the attacks that may lead to blackouts, and for routine monitoring and control tasks of the smart grid. The cyber attacks have gained increasing attention over the past years. Unfortunately, there are realistic "stealthy" threats that cannot be detected with current security modules in the power network and may lead to cascading events, instability in the system, and blackouts in major areas of the network. For details on stealthy deception attack, their implementation and serious consequences (see [5]–[9]).

A. Summary of Results

We have designed a decentralized false data injection attack detection mechanism that utilizes the Markov graph of the bus phase angles. We utilize the conditional covariance threshold test CMIT [10] to learn the structure of the grid. We show that under normal circumstances, and because of the grid structure, the Markov graph of voltage angles can be determined by the power grid graph. Therefore, a discrepancy between calculated Markov graph and learned structure triggers the alarm. This paper was initiated by Sedghi and Jonckheere [11].

Because of the connection between the Markov graph of the bus angle measurements and the grid topology, our method can be implemented in a decentralized manner, i.e., at each sub-network. Currently, sub-network topology is available online and global network structure is available hourly [2]. Not only by decentralization can we increase the speed and get closer to online detection, but we also increase accuracy and stability by avoiding communication delays and synchronization problems when trying to send measurement data between locations far apart [12], [13]. Furthermore, we noticeably decrease the amount of exchanged data to address privacy concerns as much as possible.

We show that our method can detect the most recently designed attack on the power grid that remains undetected by

Manuscript received April 5, 2014; revised September 3, 2014 and January 15, 2015; accepted January 29, 2015. Date of publication March 11, 2015; date of current version June 18, 2015. Paper no. TSG-00300-2014.

The authors are with the Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089-2563 USA (e-mail: hsedghi@usc.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.2015.2403329

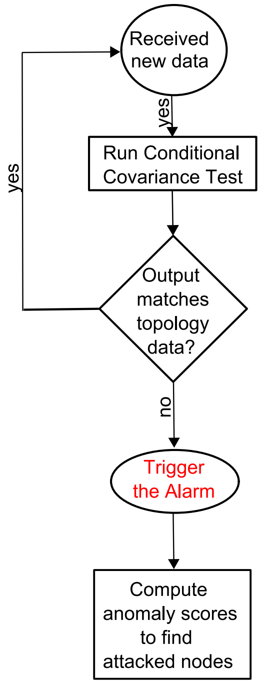
Received new data
yes
Run Conditional Covariance Test
yes
Output matches topology data?
no
Trigger the Alarm
Compute anomaly scores to find attacked nodes

Fig. 1. Flowchart of our detection algorithm.

the traditional bad data detection scheme [14] and is capable of deceiving the SE and damaging power network control, monitoring, demand response, and pricing schemes [6]. In this scenario, the attacker is equipped with vital data and has the knowledge of the bus-branch model of the grid. It should be noted that our method not only detects that the system is under attack, but also determines the particular set of nodes under the attack. The flowchart is shown in Fig. 1.

In addition, we show that our method can detect the situation where the attacker manipulates reactive power data to lead the SE to wrong estimates of the voltages. Such an attack can be designed to fake a voltage collapse or trick the operator to cause a voltage collapse. This latter detection is based on the linearization of the ac power flow around the steady state. Then using our algorithm for bus voltages and reactive power rather than bus phase angles and active power, it readily follows that this latter attack can also be detected.

B. Related Work

Although Giani *et al.* [8] suggested an algorithm for PMU placement such that the stealthy attack is observable, they report a successful algorithm only for the 2-node attack and propose empirical approaches for the 3–5-node attacks. According to [8], for cases where more than two nodes are under attack, the complexity of the approach is said to be "disheartening." Considering the fact that finding the number of needed PMUs is NP-hard and that [8] given an upper bound and uses a heuristic method for PMU placement, we need to mention for comparison purposes that our algorithm has no hardware requirements, its complexity does not depend on the number of nodes under attack, and it works for any number of attacked nodes. It is also worth mentioning that, even in the original paper presenting the attack for a relatively small network (IEEE-30), seven measurements from five nodes are manipulated. Therefore, it seems that the 2-node attack is not the most probable one.

There has been another line of work dedicated to computing the "security index" for different nodes in order to find the set of nodes that are most vulnerable to false data injection attacks [15]. Although these attempts are acknowledged, our method differs greatly from such perspectives as such methods do not detect the attack state when it happens and they cannot find the set of nodes that are under attack.

The dependency graph approach is used in [4] for topology fault detection in the grid. However, since attacks on the SE are not considered, such methods can be deceived by false data injection. Furthermore, [4] used a constrained maximum likelihood optimization for finding the information matrix, while here an advanced structure learning method is used that captures the power grid structure better. This is because in the power grid the edges are not centered but distributed all over the network. This is discussed in Section III-A.

1) Paper Outline: This paper is organized as follows. In Section II, we show that the bus phase angles form a Gaussian Markov random field (GMRF) and argue that their Markov graph is dictated by the grid structure. In Section III, we explain the conditional covariance test CMIT [10], which we use for obtaining the Markov graph among bus phase angles, and discuss how we leverage it to perform optimally for the power grid. The stealthy deception attack on the SE is introduced in Section IV. We elaborate on our detection scheme in Section V. Simulations are presented in Section VI. Section VII concludes this paper.

II. Preliminaries and Problem Formulation

A. Preliminaries

A GMRF is a family of jointly Gaussian distributions that factor according to a given graph. Given a graph $G = (V, E)$, with $V = \{1, \ldots, p\}$, consider the vector of Gaussian random variables $X = [X_1, X_2, \ldots, X_p]^\top$, where each node $i \in V$ is associated with a scalar Gaussian random variable X_i. A GMRF on G has a probability density function that can be parameterized as

$$f_X(x) \propto \exp\left[-\frac{1}{2}x^\top Jx + h^\top x\right] \qquad (1)$$

where J is a positive-definite symmetric matrix whose sparsity pattern corresponds to that of the graph G. More precisely

$$J(i,j) = 0 \Longleftrightarrow (i,j) \notin E.$$

The matrix $J = \Sigma^{-1}$ is known as the potential or information matrix, the nonzero entries $J(i,j)$ as the edge potentials, and the vector h as the vertex potential vector. In general, graph $G = (V, E)$ is called the Markov graph (graphical model) underlying the joint probability distribution $f_X(x)$, where the node set V represents random variable set $\{X_i\}$ and the edge set E is defined in order to satisfy the local Markov property. For a Markov random field, local Markov property states that $X_i \perp X_{-\{i,N(i)\}}|X_{N(i)}$, where $X_{N(i)}$ represents all random variables associated with the neighbors of i in graph G and $X_{-\{i,N(i)\}}$ denotes all variables except for X_i and $X_{N(i)}$.

B. Bus Phase Angles GMRF

We now apply the preceding to the bus phase angles. The dc power flow model [16] is often used for analysis of power systems in normal operations. When the system is stable, the phase angle differences are small, so $\sin(\theta_i - \theta_j) \sim \theta_i - \theta_j$. By the dc power flow model, the system state X can be described using bus phase angles. The active power flow on the transmission line connecting bus i to bus j is given by

$$P_{ij} = b_{ij}\left(X_i - X_j\right) \tag{2}$$

where X_i and X_j denote the phasor angles at bus i and j, respectively, and b_{ij} denotes the inverse of the line inductive reactance. The power injected at bus i equals the algebraic sum of the powers flowing away from bus i

$$P_i = \sum_{j\neq i} P_{ij} = \sum_{j\neq i} b_{ij}\left(X_i - X_j\right). \tag{3}$$

When buses i and j are not connected, $b_{ij} = 0$. Thus, it follows that the phasor angle at bus i could be represented as:

$$X_i = \sum_{j\neq i}\left\{\frac{b_{ij}}{\sum_{i\neq j} b_{ij}}\right\} X_j + \frac{1}{\sum_{j\neq i} b_{ij}} P_i. \tag{4}$$

Equation (2) can also be rewritten in matrix form as

$$P = BX \tag{5}$$

where $P = [P_1, P_2, \dots, P_p]^\top$ is the vector of injected active powers, $X = [X_1, X_2, \dots, X_p]^\top$ is the vector of bus phase angles and

$$B = \begin{cases} -b_{ij} & \text{if } i \neq j \\ \sum_{j\neq i} b_{ij} & \text{if } i = j. \end{cases} \tag{6}$$

1) Remark: Note that, because of linearity of the dc power flow model, the above equations are valid for both the phase angle X together with the injected power P and for the fluctuations of the phase angle X together with the fluctuations of the injected power P around its steady-state value. Specifically, if we let $\widetilde{P}$ refer to the vector of active power fluctuations and $\widetilde{X}$ represent the vector of phase angle fluctuations, we have $\widetilde{P} = B\widetilde{X}$. In the following, the focus is on the dc power flow model. Nevertheless, our analysis remains valid if we consider fluctuations around the steady-state values.

Because of load uncertainty, and under generation-load balance, the injected power can be modeled as a random variable [17]. The injected power is the sum of many random factors such as load fluctuations, wind turbine and photo voltaic cell output fluctuations, etc. While the independence of the constituting random variables can be justified, their identical distribution cannot. Therefore, using the Lyapunov central limit theorem (CLT) [18, Sec. 7.7.2], which does not require the random variables to be identically distributed, we can model the injected power as a Gaussian distribution.

2) Lyapunov CLT: Let $\{Y_i : i = 1, 2, \dots, n\}$ be a sequence of independent random variables each with finite expected value μ_i and variance σ_i^2. Define $s_n^2 = \sum_{i=1}^n \sigma_i^2$. If the Lyapunov condition[1] is satisfied, then $\sum_{i=1}^n (Y_i - \mu_i)/s_n$ converges in distribution to a standard normal random variable as n goes to infinity.

Considering conventional assumptions in power systems, the Lyapunov condition is met. As argued in [19], the Gaussian assumption is justified in the transmission network. The Gaussian model is also utilized in various analysis of power networks such as [20]–[23] where n is estimated to be of order 1000. To exemplify CLT, it is suggested in [24] that as few as five wind turbines would suffice to see CLT in action. Therefore, for each i, we model P_i in (3) with a Gaussian random variable. Hence the linear relationship in (5), together with the fixed phasor at the slack bus, implies that the phasor angles θ_i are Gaussian random variables [4].

The next step is to find out whether the X_is satisfy the local Markov property and, in the affirmative, to discover the neighbor sets corresponding to each node. We do this by analyzing (4). If there were only the first term, we would conclude that the set of nodes electrically connected to node i satisfies the local Markov property, but the second term makes a difference. Below, we argue that an analysis of the second term of (4) shows that this term causes some second-neighbors of X_i to have a nonzero term in matrix J. In addition, for nodes that are more than two hops apart, $J_{ij} = 0$. Therefore, as opposed to the claim in [4], a second-neighbor relationship does exist in matrix J. The second neighbor property may result in additional edges in the Markov graph between the nodes that are second neighbors in the grid graph.

As stated earlier, the powers injected at different buses have Gaussian distribution. We can assume that they are independent and without loss of generality they are zero mean. Therefore, the probability distribution function for P is $f_P(P) \propto e^{-1/2P^\top P}$. Since $P = BX$, we have $f_X(X) \propto e^{-1/2X^\top B^\top BX}$. Recalling the definition of the probability distribution function for jointly Gaussian random variables in (1), we get $J = B^T B$. Let $d(i,j)$ represent the hop distance between nodes i and j in the power grid graph G. By definition of matrix B, this leads to some nonzero J_{ij} entries for $d(i,j) = 2$. In addition, we state the following.

Proposition 1: Assume that the powers injected at the nodes are Gaussian and mutually independent. Then

$$J_{ij} = 0, \qquad \forall\ d(i,j) > 2.$$

Proof: We argue by contradiction. Assume $J_{ij} \neq 0$ for some $d(i,j) > 2$. Since $J_{ij} = \sum_k B_{ik}B_{jk}$, it follows that $\exists\ k$ s.t. $B_{ik} \neq 0, B_{jk} \neq 0$. By (6), $B_{ik} \neq 0$ implies $d(i,k) = 1$. From there on, the triangle inequality implies that $d(i,j) \leq d(i,k) + d(k,j) = 1 + 1 = 2$, which contradicts the assumption $d(i,j) > 2$. ■

It was shown in [19] that for some graphs, the second-neighbor terms are smaller than the terms corresponding to

[1] The condition requires that $\exists\delta > 0$ such that the random variables $|Y_i - \mu_i|$ have moments of order $2 + \delta$ and the rate of growth of these moments is limited in the sense that $\lim_{n\to\infty}(\sum_{i=1}^n E|Y_i - \mu_i|^{2+\delta}/s_n^{2+\delta}) = 0$.

Algorithm 1 *CMIT*(x^n; $\xi_{n,p}$, η) for Structure Learning Using Samples x^n [10]

```
Initialize $\widehat{G}_p^n = (V, \emptyset)$

For each $i, j \in V$,
if        $\min_{S \subset V \setminus \{i,j\},\ |S| \le \eta} \widehat{\Sigma}(i, j|S) > \xi_{n,p}$,
 then
    add $(i, j)$ to the edge set of $\widehat{G}_p^n$.

end if
Output:$\widehat{G}_p^n$
```

the immediate electrical neighbors of X_i. More precisely, it was shown that for lattice-structured grids, this approximation falls under the generic fact of the tapering off of Fourier coefficients [19]. Therefore, we can approximate each neighborhood with the immediate electrical neighbors. We can also proceed with the exact relationship. For simplicity, we opt for the first-neighbor analysis. We explain shortly why CMIT works with this approximation as well.

Note that our detection method relies on the graphical model of the variables. It is based on the fact that the Markov graph of bus phase angles changes under an attack. CMIT is tuned with correct data and we prove that in case of attack, the Markov graph of compromised data does not follow the Markov graph of correct data. Hence, we can tune CMIT by either the exact relationship or the approximate Markov graph. In both cases, the output in case of attack is different from the output tuned with correct data. Therefore, CMIT works for both approximate and exact neighborhoods.

III. Structure Learning

In the context of graphical models, model selection means finding the exact Markov graph underlying a group of random variables based on samples of those random variables. There are two main classes of methods for learning the structure of the underlying graphical model, convex methods and nonconvex methods. The ℓ_1-regularized maximum likelihood estimators are the main class of convex methods [25]–[28]. In these methods, the inverse covariance matrix is penalized with a convex ℓ_1-regularizer in order to encourage sparsity in the estimated Markov graph structure. The other types of methods are the nonconvex or greedy methods [10]. In this paper, we use the latter methods.

A. Conditional Covariance Test

In order to learn the structure of the power grid, we utilize the Gaussian graphical model selection method called CMIT [10]. CMIT estimates the structure of the underlying graphical model given i.i.d. samples of the random variables. CMIT is shown in Algorithm 1.

In Algorithm 1, the output is an edge set corresponding to graph G given n i.i.d. samples x^n, each of which has p variables (corresponding to vertices), a threshold $\xi_{n,p}$ (that depends on both p and n) and a constant $\eta \in \mathbb{N}$, which is related to the local vertex separation property (described later). In our case, each one of the p variables represents a bus phase angle.

The sufficient condition for output of CMIT to have structural consistency with the underlying Markov graph among variables is that the graph has to satisfy local separation property and walk-summability [10]. An ensemble of graphs has the (η, γ)-local separation property if for any $(i, j) \notin E(G)$, the maximum number of paths between i and j of length at most γ does not exceed η. A Gaussian model is said to be α-walk summable if $||\bar{\mathbf{R}}|| \le \alpha < 1$, where $\bar{\mathbf{R}} = [|r_{ij}|]$ and $||.||$ denotes the spectral or 2-norm of a matrix [10]. $\mathbf{R} = [r_{ij}]$ is the matrix of partial correlation coefficients; it vanishes on the diagonal entries and on the nondiagonal entries it is given by

$$r_{ij} \triangleq \frac{\Sigma(i, j|V \setminus \{i, j\})}{\sqrt{\Sigma(i, i|V \setminus \{i, j\})\Sigma(j, j|V \setminus \{i, j\})}} = -\frac{J(i, j)}{\sqrt{J(i, i)J(j, j)}} \qquad (7)$$

r_{ij}, the partial correlation coefficient between variables X_i and X_j for $i \neq j$, measures their conditional covariance given all other variables [29].

Regardless of whether the exact or approximate neighborhood relationship holds, the Markov graph of the bus phase angles is an example of bounded local path graphs that satisfy the local separation property. We also checked the analyzed networks for the walk-summability condition. As shown in (7) and the definition of walk-summability, this property depends only on matrix J and thus on the topology of the grid. The walk-summability does not depend on the operating point of the grid.

It is shown in [10] that, under walk-summability, the effect of faraway nodes on the covariance decays exponentially with the distance and the error in approximating the covariance by local neighboring decays exponentially with the distance. Hence by correct tuning of threshold $\xi_{n,p}$ and with enough samples, we expect the output of CMIT to follow the grid structure.

The computational complexity of CMIT is $O(p^{\eta+2})$, which is efficient for small η [10]. η is the parameter associated with local separation property described above. The sample complexity associated with CMIT is $n = \Omega(J_{\min}^{-2} \log p)$, where $J_{\min}$ is the minimum absolute edge potential in the model [10].

It is worth mentioning that since we use CMIT for structure learning of phasor data, our method is robust against measurement noise. The reason is that CMIT analyzes conditional covariance of its input data. Since input data is Gaussian, the conditional covariance can be found from covariance matrix for phasor data, i.e., $\Sigma(X, X)$ [see (8)]. Let N be the sum of the measurement noise and systematic errors. Both systematic errors and measurement noise are independent of the measured values. Also, we know that $\mathbb{E}(X) = 0$. Therefore, $\Sigma(X + N, X + N) = \Sigma(X, X) + \Sigma(N, N)$. Note that in CMIT we only look at pairs (i, j) such that $i \neq j$. Therefore as long as $\Sigma(N, N)$ has a diagonal form, this error does not influence our performance. This is the case when errors at different locations in the network are independent of each other. Measurement noise meets this criterion. Moreover, if systematic error in the network has a diagonal covariance matrix, it also does not impact our method. Even if systematic errors do not have a diagonal covariance but remain the same with time, they can be detected and compensated during an initial training phase when we are sure the system is not under the attack.

CMIT distributes the edges fairly uniformly across the nodes, while the ℓ_1 method tends to cluster all the edges together among the "dominant" variables leading to a densely connected component and several isolated points [10] and thus a disconnected graph. Therefore, the ℓ_1 method has some limitations in detecting the structure of a connected graph. The power grid transmission network is a connected graph where the edges are distributed over the network. Therefore, CMIT is more suitable for detecting the structure of the power grid.

B. Decentralization

We want to find the Markov graph of our bus phasor measurements. The connection between electrical connectivity and correlation (Proposition 1) helps us to decentralize our method to a great extent. The power network in its normal operating condition consists of different areas connected together via border nodes. A border node is any node that is also connected to a node from a different area as depicted in [30]. Therefore, we decompose our network into these sub-areas. Our method can be performed locally in the sub-networks. The sub-network connection graph is available online from the protection system at each sub-network and can be readily compared with the bus phase angle Markov graph. In addition, only for border nodes we need to consider their out-of-area neighbors as well. This can be done either by solving the power flow equations for that border link or by receiving measurements from neighbor sub-networks. Therefore, we run CMIT for each sub-graph to figure out its Markov graph. Then, we compare it with online network graph information to detect false data injection attacks.

This decentralization reduces complexity and increases speed. Our decentralized method is a substitute for considering all measurements throughout the power grid, which requires a huge amount of data exchange, computation, and overhead. In addition to having fewer nodes to analyze, this decentralization leads us to a smaller η and greatly reduces computational complexity, which makes our method capable of being executed in very large networks. Furthermore, since structure learning is performed locally, faraway relationships created by nonlinearities—ignored in Proposition 1 but intrinsic to power systems—are mitigated, hence our neighborhood assumptions are justified. Last but not least, utility companies are not willing to expose their information for economical competition reasons and there have been several attempts to make them do that [31]. Thus it is desired to reduce the amount of data exchange between different areas and our method adequately fulfills this preference.

It should be noted that the measurement vector X analyzed in this paper is a mixture of measurements from PMUs and SE output corresponding to the same time. This is achieved as follows. PMUs use GPS-sync time stamp and SE measurements in SCADA are labeled with local time stamp. Since our method is performed locally, it has two advantages. First, as discussed earlier, it avoids large delays in communication network. Second, we can use the local time stamps from SE outputs. We do not require the high rate of measurement from PMUs for our detection scheme and only consider the PMU samples at the time we have SE samples. Since both data have time stamps, we are able to form the measurement vector X with measurement data from the same time.

C. Online Calculations

For fast monitoring of the power grid, we need an on-line algorithm. As we show in this section, our algorithm can be developed as an iterative method that processes new data without the need for reprocessing earlier data. Here, we derive an iterative formulation for the sample covariance matrix. Then, we use it to calculate the conditional covariance using

$$\widehat{\Sigma}(i,j|S) := \widehat{\Sigma}(i,j) - \widehat{\Sigma}(i,S)\widehat{\Sigma}^{-1}(S,S)\widehat{\Sigma}(S,j). \tag{8}$$

As we know, in general

$$\Sigma = E\left[(X-\mu)(X-\mu)^\top\right] = E\left[XX^\top\right] - \mu\mu^\top.$$

Let $\widehat{\Sigma}^{(n)}(X)$ denote the sample covariance matrix for a vector X of p elements from n samples and let $\widehat{\mu}^{(n)}(X)$ be the corresponding sample mean. In addition, let $X^{(i)}$ be the ith sample of our vector. Then, we have

$$\widehat{\Sigma}^{(n)}(X) = \frac{1}{n-1}\left(\sum_{i=1}^{n} X^{(i)}X^{(i)\top}\right) - \widehat{\mu}^{(n)}\widehat{\mu}^{(n)\top}. \tag{9}$$

Therefore

$$\widehat{\Sigma}^{(n+1)}(X) = \frac{1}{n}\left[\sum_{i=1}^{n} X^{(i)}X^{(i)\top} + X^{(n+1)}X^{(n+1)\top}\right] - \widehat{\mu}^{(n+1)}\widehat{\mu}^{(n+1)\top} \tag{10a}$$

$$\widehat{\mu}^{(n+1)} = \frac{1}{n+1}\left[n\widehat{\mu}^{(n)} + X^{(n+1)}\right]. \tag{10b}$$

By keeping the first term in (9) and the sample mean (10b), our updating rule is (10a). Thus, we revise the sample covariance as soon as any bus phasor measurement changes and leverage it to reach the conditional covariances needed for CMIT. It goes without saying that if the system demand and structure does not change and the system is not subject to false data injection attack, the voltage angles at nodes remain the same and there is no need to run any algorithm.

IV. Stealthy Deception Attack

The most recent and most dreaded false data injection attack on the power grid was introduced in [14]. It assumes knowledge of the bus-branch model and it is capable of deceiving the SE. For a p-bus electric power network, the $l = 2p-1$ dimensional state vector x is $[\theta^\top, V^\top]^\top$, where $V = [V_1, \ldots, V_p]^\top$ is the vector of voltage bus magnitudes and $\theta = [\theta_2, \ldots, \theta_p]^\top$ the vector of phase angles. It is assumed that the nonlinear measurement model for the state estimation is $z = h(x) + \epsilon$, where $h(.)$ is the measurement function, $z = [z_P^\top, z_Q^\top]^\top$ is the measurement vector consisting of active and reactive power flow measurements and ϵ is the measurement error.

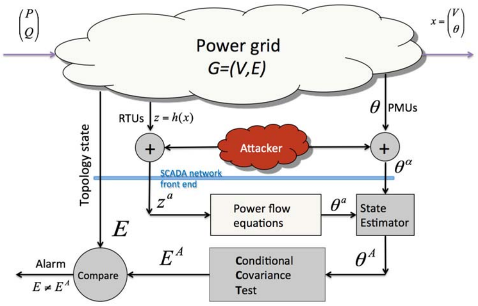

Fig. 2. Power grid under a cyber attack.

$H(x^k) := (dh(x)/dx)|_{x=x^k}$ denotes the Jacobian matrix of the measurement model $h(x)$ at x^k. The goal of the stealthy deception attacker is to compromise the measurements available to the SE as $z^a = z + a$, where z^a is the corrupted measurement vector and a is the attack vector. The vector a is designed such that the SE algorithm converges and the attack a is undetected by the bad data detection scheme. Then it is shown that, under the dc power flow model, such an attack can only be performed locally with $a \in \mathrm{Im}(H)$, where $H = H_{P\theta}$ is the matrix connecting the vector of bus injected active powers to the vector of bus phase angles, i.e., $P = H_{P\theta}\theta$. The attack is shown in Fig. 2.

V. Stealthy Deception Attack Detection

In this section, we show that our method can detect the aforementioned stealthy deception attack despite the fact that it remains undetected by the traditional bad data detection scheme. The fundamental idea behind our detection scheme is that of structure learning. Our learner, CMIT, is first tuned with correct data, which corresponds to the grid graph. Therefore, any attack that changes the structure alters the output of CMIT and this triggers the alarm. Let us consider the attack more specifically. As we are considering the dc power flow model and all voltage magnitudes are normalized to 1 p.u., the state vector introduced in [14] reduced to the vector of voltage angles, X. Since $a \in \mathrm{Im}(H)$, $\exists\ d$ such that $a = Hd$ and

$$z^a = z + a = H(X + d) = HX^a$$

where X^a represents the vector of angles when the system is under attack, z^a is the attacked measurement vector, and X is the correct phasor angle vector. Considering (3), we have $H_{ij} = -b_{ij}$ for $i \neq j$ and $H_{ii} = \sum_{i \neq j} b_{ij}$, where b_{ij} denotes the inverse of the line inductive reactance. We have

$$X^a = X + d = H^{-1}P + H^{-1}a = H^{-1}(P + a). \qquad (11)$$

As the definition of matrix H shows, it is of rank $p - 1$. Therefore, the above H^{-1} denotes the pseudo-inverse of matrix H. Another way to address this singularity is to remove the row and the column associated with the slack bus. From (11), we get

$$\begin{aligned}\Sigma\big(X^a, X^a\big) &= H^{-1}[\Sigma(P + a, P + a)]{H^{-1}}^T\\ &= H^{-1}[\Sigma(P, P) + \Sigma(a, a)]{H^{-1}}^T.\end{aligned}$$

The above calculation assumes that the attack vector is independent of the current measurement values in the network, as demonstrated in the definition of the attack [14].

An attack is considered successful if it causes the operator to make a wrong decision. For that matter, the attacker would not insert just one wrong sample. In addition, if the attack vector remains constant, it does not cause any reaction. This eliminates the case of constant attack vectors. Therefore, the attacker is expected to insert nonconstant vectors a during some samples. Thus $\Sigma(a, a) \neq 0$ and

$$\Sigma\big(X^a, X^a\big) \neq \Sigma(X, X). \qquad (12)$$

It is not difficult to show that, if we remove the assumption on independence of attack vector and the injected power, (12) still holds.

Considering (12) and the fact that matrix inverse is unique, it follows that, in case of an attack, the new Σ^{-1} will not be the same as the network information matrix in normal condition, i.e., $\Sigma^{-1}(X^a, X^a) \neq J_{\mathrm{normal}}$, and as a result, the output of CMIT will not follow the grid structure. We use this mismatch to trigger the alarm. It should be noted that acceptable load changes do not change the Markov graph and as a result do not lead to false alarms. The reason is that such changes do not falsify the dc power flow model and the Markov graph will continue to follow the defined information matrix. After the alarm is triggered, the next step is to find which nodes are under attack.

A. Detecting the Set of Attacked Nodes

We use the correlation anomaly metric [32] to find the attacked nodes. This metric quantifies the contribution of each random variable to the difference between two probability densities while considering the sparsity of the structure. The Kullback–Leibler divergence is used as the measure of the difference. As soon as an attack is detected, we use the attacked information matrix and the information matrix corresponding to the current topology of the grid to compute the anomaly score for each node. The nodes with highest anomaly scores are announced as the nodes under attack. We investigate the implementation details in the next section.

It should be noted that the attack is performed locally and because of the local Markov property, we are certain that no nodes from other sub-graphs contribute to the attack.

We should emphasize that the considered attack assumes the knowledge of the system bus-branch model. Therefore, the attacker is equipped with very critical information. Yet, we can mitigate such an "intelligent" attack.

B. Reactive Power Versus Voltage Amplitude

As mentioned before, with similar calculations, we can consider the case where the attacker manipulates reactive power data to lead the SE to wrong estimates of the voltage. Such an attack can be designed to fake a voltage collapse or trick the operator to cause a change in the normal state of the grid. For example, if the attacker fakes a decreasing trend in the voltage magnitude in some part of the grid, the operator will send more reactive power to that part and thus this could cause voltage overload/underload. At this point, the protection system would disconnect the corresponding lines. This could lead to outages in some areas and in a worse scenario to overloading in other parts of the grid that might cause blackouts and cascading events.

The detection can be done by linearization of the ac power flow and by considering the fluctuations around steady state. Then pursuing our algorithm, it readily follows that such an attack can also be detected with a similar approach to the one developed here for bus phase angles and active power.

In the rest of this section, we show how this analogy can be established. The ac power flow states that the active power and the reactive power flowing from bus i to bus j are, respectively

$$P_{ij} = G_{ij}V_i^2 - G_{ij}V_iV_j\cos\left(\theta_i-\theta_j\right) + b_{ij}V_iV_j\sin\left(\theta_i-\theta_j\right)$$
$$Q_{ij} = b_{ij}V_i^2 - b_{ij}V_iV_j\cos\left(\theta_i-\theta_j\right) - G_{ij}V_iV_j\sin\left(\theta_i-\theta_j\right)$$

where V_i and θ_i are the voltage magnitude and phase angle, respectively, at bus i and G_{ij} and b_{ij} are the conductance and susceptance, respectively, of line ij. From [33], we obtain the following approximation of the ac fluctuating power flow:

$$\widetilde{P}_{ij} = \left(b_{ij}\overline{V}_i\overline{V}_j\cos\overline{\theta}_{ij}\right)\left(\widetilde{\theta}_i-\widetilde{\theta}_i\right)$$
$$\widetilde{Q}_{ij} = \left(2b_{ij}\overline{V}_i - b_{ij}\overline{V}_j\cos\overline{\theta}_{ij}\right)\tilde{V}_i - \left(b_{ij}\overline{V}_i\cos\overline{\theta}_{ij}\right)\tilde{V}_j$$

where an overbar denotes the steady-state value, a tilde means the fluctuation around the steady-state value, and $\overline{\theta}_{ij} = \overline{\theta}_i - \overline{\theta}_j$. These fluctuating values due to renewables and variable loads justify the utilization of probabilistic methods in power grid problems.

Now assuming that for the steady-state values of the voltages we have $\overline{V}_i = \overline{V}_j \simeq 1$ p.u. (per unit) and the fluctuations in angles are about the same such that $\cos\theta_{ij} = 1$, we have

$$\widetilde{P}_{ij} = b_{ij}\left(\widetilde{\theta}_i - \widetilde{\theta}_j\right) \tag{13a}$$
$$\widetilde{Q}_{ij} = b_{ij}\left(\widetilde{V}_i - \widetilde{V}_j\right). \tag{13b}$$

It is clear from (13a) and (13b) that, we can follow the same approach we had about active power and voltage angles with reactive power and voltage magnitudes, respectively.

It can be argued that, as a result of uncertainty, the aggregate reactive power at each bus can be approximated as a Gaussian random variable and, because of (13b), the voltage fluctuations around the steady-state value can be approximated with Gaussian random variables. Therefore, the same path of approach as for phase angles can be followed to show the GMRF property for voltage amplitudes. Comparing (13b) with (2) makes it clear that the same matrix, i.e., matrix B developed in Section II-B, is playing the role of correlating the voltage amplitudes. Therefore, assuming that the statistics of the active and reactive power fluctuations are similar, the underlying graph is the same. This can readily be seen by comparing (13a) and (13b).

VI. Simulation

A. Training the System

We consider IEEE-14 bus system as well as IEEE-30 bus system. First, we feed the system with Gaussian demand and simulate the power grid. We use MATPOWER [34] for solving the dc power flow equations for various demand and use the resulting angle measurements as the input to CMIT. We leverage YALMIP [35] and SDPT3 [36] to run CMIT in MATLAB. With the right choice of parameters and threshold $\xi_{n,p}$ of CMIT, and enough measurements, the Markov graph should follow the grid structure. We use the edit distance between two graphs for tuning the threshold $\xi_{n,p}$. The edit distance between two graphs reveals the number of edges that exist in only one of the two graphs.

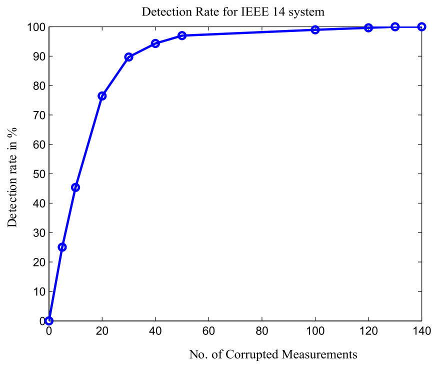

Fig. 3. Detection rate for IEEE-14 bus system.

B. Detecting Attack State

After the threshold $\xi_{n,p}$ is set, our detection algorithm works in the following manner. Each time the procedure is initiated, i.e., when any PMU angle measurement or SE output changes, it updates the conditional covariances $\hat{\Sigma}(i,j|S)$ based on new data, runs CMIT and checks the edit distance between the Markov graph of phasor data and the grid structure. A discrepancy triggers the alarm. Subsequently to an alarm, the system uses anomaly metric to find all the buses under the attack. The flowchart of our method is shown in Fig. 1.

Next, we introduce the stealthy deception attack on the system. The attack is designed according to the description in [14], i.e., it is a random vector such that $a \in \mathrm{Im}(H)$. The attack is claimed to be successful only if performed locally on connected nodes. Having this constraint in mind, for IEEE-14 test case the maximum number of attacked nodes is six and for IEEE-30 bus system this number is eight. For the IEEE-14 network, we consider the cases where 2–6 nodes are under attack. For the IEEE-30 network, we consider the cases where 2–8 nodes are under attack. For each case and for each network, we simulate all possible attack combinations. This is to make sure we have checked our detection scheme against all possible stealthy deception attacks. Each case is repeated 1000 times for different attack vector values.

When the attacker starts tampering with the data, the corrupted samples are added to the sample bin of CMIT and are therefore used in calculating the sample covariance matrix. With enough corrupted samples, our algorithm can get arbitrarily close to 100% successful in detecting all cases of attacks discussed above, for both IEEE-14 and IEEE-30 bus systems. This is shown in Fig. 3 for IEEE-14 bus system. The detection rate is averaged over all possible attack scenarios. The reason behind the trend shown in Fig. 3 is that first, for a very

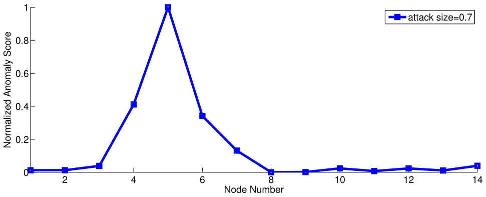

Fig. 4. Anomaly score for IEEE-14 bus system. Nodes 4–6 are under attack; attack size is 0.7.

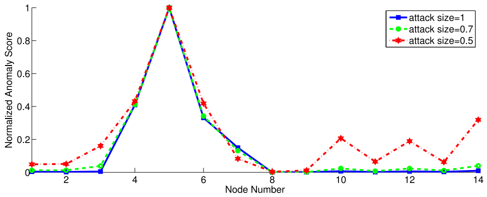

Fig. 5. Anomaly score for IEEE-14 bus system for different attack sizes. Nodes 4–6 are under attack. Attack sizes are 0.5, 0.7, and 1.

small number of corrupted measurements, the Markov graph follows the true information matrix and then, for a higher number of compromised measurements, the Markov graph follows the random relationship that the attacker is producing. When the number of compromised samples increases, they gain more weight in the sample covariance, and the chance of a change in the Markov graph increases. It can be seen that even for a small number of corrupted measurements, our method presents a good performance: the detection rate is 90% with 30 corrupted samples. The minimum number of corrupted samples to get almost 100% detection rate for IEEE-14 bus system is 130 and it is 50 for IEEE-30 bus system. Since IEEE-30 is more sparse than IEEE-14 bus system, our method performs more efficiently in the former case. Yet, for a 60 Hz system, the detection speed for IEEE-14 bus system is quite amazing as well.

C. Identifying Nodes Under Attack

The next step is to find which nodes are under attack. As stated earlier, we use anomaly score metric [32] to detect such nodes. As an example, Fig. 4 shows the anomaly score plot for the case where nodes 4–6 are under attack.[2] It means that a random vector is added to the measurements at these nodes. This attack is repeated 1000 times for different values building an attack size of 0.7. The attack size refers to the expected value of the Euclidean norm of the attack vector a.

Simulation results show that as the attack size increases, the difference between the anomaly scores of the nodes under the attack and the uncompromised nodes increases and, as a result, it becomes easier to pinpoint the attacked nodes. For example, Fig. 5 compares the cases where the attack size is 1, 0.7, and 0.5 for the attack scenario where nodes 4–6 are under attack. It should be noted that in order for an attack to be successful in misleading the TSO, the attack size should not be too small. More specifically, the attacker wants to make a change in the system state such that the change is noticeable with the hope that this would result in the wrong reaction of the TSO. If the value of the system state under the attack is close to its real value, the system is not considered under the attack as it continues its normal operation. It can be seen that, even for the smallest possible attack size that would normally not lead the operator to react, the anomaly score plot will remain reliable. For example, in the considered attack scenario, the anomaly plot performs well even for an attack size of 0.3, while it seems that a potentially successful attack under normal standards needs a bigger attack size.

D. Setting Up Anomaly Score Threshold

Setting the threshold for anomaly score is another important aspect of the detection algorithm. As discussed earlier, our scheme has two major parts. First, detection of attack state, i.e., to declare if the system is under attack. Second, the identification of the attacked nodes in case of an attack state. In Fig. 3, we analyzed the detection rate of the "attack state" versus the number of corrupted samples. In Figs. 4 and 5, we discussed how normalized anomaly score changes with different attack sizes. Now, we use this intuition to design the threshold for anomaly score. In case of attack state we calculate the normalized anomaly score for each node. For any node, if this benchmark is greater than the threshold, the node is considered to be under attack. In this context, we define the "node detection ratio (NDRo)" as the ratio of the number of attacked nodes that are correctly labeled as attacked to the total number of attacked nodes. Consequently, the "false alarm ratio (FARo)," not to be confused with the false alarm rate, refers to the number of uncompromised nodes that are mislabeled as under attack to the total number of uncompromised nodes. As in detection theory, there is a trade-off in designing this threshold value. Lower threshold values result in higher NDRo and higher FARo and vice versa. Since our goal is to detect all attacked nodes, we design the threshold such that the NDRo is approximately 100% with a very low FARo. To design the threshold, we repeat the simulation discussed for Figs. 4 and 5 for five different sets of attacked nodes, the three discussed attack sizes, and repeat each attack size 100 times. As can also be seen in the above plots, with a threshold of 0.3 for all attacked nodes, the normalized anomaly score is above the threshold. Next, we use this threshold in all possible sets of attacked nodes on IEEE-14 bus system with a attack size of 0.7 and repeat it 50 times for each set. Simulation results show that this threshold guarantees nearly 100% NDRo with a very low FARo of 3.82×10^{-5}. The reason is that anomaly score provides a precise statistical analysis of the nodes that contribute to the mismatch. Hence, we can obtain 100% detection rate with a very low FARo.

VII. Conclusion

We have proposed a decentralized false data injection attack detection scheme that is capable of detecting the most recent stealthy deception attack on power grid. To the best of our knowledge, our remedy is the first to comprehensively detect this sophisticated attack. In addition to detecting the attack state, our algorithm is capable of pinpointing the set of

[2] The numbering system employed here is the one of the published IEEE-14 system available at https://www.ee.washington.edu/research/pstca/pf14/pg_tca14bus.htm

attacked nodes. Although [8] considered the same attack on the power network, considerable progress is made in our approach versus the one in [8]. In both cases, the goal is to detect the attack. While [8] seeks a PMU placement method, our method does not require additional hardware but rather performs statistical structure learning on the measurement data. In general, both PMU placement and structure learning are NP-hard. However, the use of common knowledge of the grid structure helps us reach a polynomial time solution. The power network structure is a sparse graph that satisfies the local separation property and the walk-summability. For details on how these properties reduce the general NP-hard problem to a tractable polynomial time problem (see [10]).

As stated earlier, the computational complexity of our method is polynomial and the decentralized property makes our scheme suitable for huge networks, yet with bearable complexity and run time. In addition, our method is capable of detecting attacks that manipulate reactive power measurements to cause inaccurate voltage amplitude data. Such attack scenario can lead to, or mimic a voltage collapse.

In the conclusion, we have introduced change detection for the graphical model of a power system and showed that it can be used to detect data manipulation. Our method protects the power system against a large class of false data injection attacks, which is of paramount importance for current and future grid reliability, security, and stability.

References

[1] R. Diao *et al.*, "Decision tree-based online voltage security assessment using PMU measurements," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 832–839, May 2009.

[2] H. Zhu and G. B. Giannakis, "Sparse overcomplete representations for efficient identification of power line outages," *IEEE Trans. Power Syst.*, vol. 27, no. 4, pp. 2215–2224, Nov. 2012.

[3] C. Wei, A. Wiesel, and R. S. Blum, "Change detection in smart grids using errors in variables models," in *Proc. IEEE 7th Sensor Array Multichannel Signal Process. Workshop (SAM)*, Hoboken, NJ, USA, Jun. 2012, pp. 17–20.

[4] M. He and J. Zhang, "A dependency graph approach for fault detection and localization towards secure smart grid," *IEEE Trans. Smart Grid*, vol. 2, no. 2, pp. 342–351, Jun. 2011.

[5] C. Kwon, W. Liu, and I. Hwang, "Security analysis for cyber-physical systems against stealthy deception attacks," in *Proc. IEEE Amer. Control Conf. (ACC)*, Washington, DC, USA, 2013, pp. 3344–3349.

[6] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on smart grid state estimation: Attack strategies and countermeasures," in *Proc. IEEE Smart Grid Commun. (SmartGridComm)*, Gaithersburg, MD, USA, 2010, pp. 220–225.

[7] S. M. Amin and A. M. Giacomoni, "Smart grid—Safe, secure, self-healing," *IEEE Power Energy Mag.*, vol. 10, no. 1, pp. 33–40, Jan./Feb. 2012.

[8] A. Giani *et al.*, "Smart grid data integrity attacks," *IEEE Trans. Smart Grid*, vol. 4, no. 3, pp. 1244–1253, Sep. 2013.

[9] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Trans. Inf. Sec. Syst.*, vol. 14, no. 1, May 2011, Art. ID 13.

[10] A. Anandkumar, V. Tan, F. Huang, and A. Willsky, "High-dimensional Gaussian graphical model selection: Walk summability and local separation criterion," *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 2293–2337, Aug. 2012.

[11] H. Sedghi and E. Jonckheere, "Statistical structure learning of smart grid for detection of false data injection," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PES)*, Vancouver, BC, Canada, 2013, pp. 1–5.

[12] K. Zhu, M. Chenine, L. Nordström, S. Holmström, and G. Ericsson, "An empirical study of synchrophasor communication delay in a utility TCP/IP network," *Int. J. Emerg. Elect. Power Syst.*, vol. 14, no. 4, pp. 341–350, 2013.

[13] E. Ancillotti, R. Bruno, and M. Conti, "The role of communication systems in smart grids: Architectures, technical solutions and research challenges," *Comput. Commun.*, vol. 36, no. 17, pp. 1665–1697, 2013.

[14] A. Teixeira, G. Dan, H. Sandberg, and K. H. Johansson, "A cyber security study of a SCADA energy management system: Stealthy deception attacks on the state estimator," in *Proc. IFAC World Congr.*, Milan, Italy, Sep. 2011, pp. 11271–11277.

[15] J. M. Hendrickx, K. H. Johansson, R. M. Jungers, H. Sandberg, and K. C. Sou, "Efficient computations of a security index for false data attacks in power networks," *IEEE Trans. Autom. Control*, vol. 59, no. 12, pp. 3194–3208, Dec. 2014.

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

[17] P. Zhang and S. T. Lee, "Probabilistic load flow computation using the method of combined cumulants and Gram-Charlier expansion," *IEEE Trans. Power Syst.*, vol. 19, no. 1, pp. 676–682, Feb. 2004.

[18] B. De Finetti, *Theory of Probability*. Hoboken, NJ, USA: Wiley, 1975.

[19] H. Sedghi and E. Jonckheere, "On conditional mutual information in Gauss-Markov structured grids," in *Information and Control in Networks* (Lecture Notes in Control and Information Sciences), vol. 450, G. Como, B. Bernhardson, and A. Rantzer, Eds. Berlin, Germany: Springer-Verlag, 2014, pp. 277–297.

[20] A. Kashyap and D. Callaway, "Estimating the probability of load curtailment in power systems with responsive distributed storage," in *Proc. IEEE Int. Conf. Probab. Methods Appl. Power Syst. (PMAPS)*, Singapore, 2010, pp. 18–23.

[21] A. Schellenberg, W. Rosehart, and J. Aguado, "Cumulant-based probabilistic optimal power flow (P-OPF) with Gaussian and Gamma distributions," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 773–781, May 2005.

[22] G. Pang, G. Kesidis, and T. Konstantopoulos, "Avoiding overages by deferred aggregate demand for PEV charging on the smart grid," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Ottawa, ON, Canada, 2012, pp. 3322–3327.

[23] J. Dopazo, O. Klitin, and A. Sasson, "Stochastic load flows," *IEEE Trans. Power App. Syst.*, vol. 94, no. 2, pp. 299–309, Mar. 1975.

[24] J. Mur-Amada and J. Salln-Arasanz, "From turbine to wind farms—Technical requirements and spin-off products," in *Phase Transitions and Critical Phenomena*, vol. 18, G. Krause, Ed. Rijeka, Croatia: InTech, Apr. 2011, pp. 101–132.

[25] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical Lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, 2008.

[26] P. Ravikumar, M. Wainwright, G. Raskutti, and B. Yu, "High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence," *Electron. J. Statist.*, vol. 5 no. 4, pp. 935–980, 2011.

[27] M. Janzamin and A. Anandkumar, "High-dimensional covariance decomposition into sparse Markov and independence domains," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Edinburgh, U.K., 2012, pp. 1839–1846.

[28] M. Janzamin and A. Anandkumar, "High-dimensional covariance decomposition into sparse Markov and independence models," *J. Mach. Learn. Res. (JMLR)*, vol. 15, pp. 1549–1591, Apr. 2014.

[29] S. Lauritzen, *Graphical Models*. Oxford, U.K.: Clarendon Press, 1996.

[30] (Aug. 2014). *Pictorial Explanation of a Border Node*. [Online]. Available: https://www.dropbox.com/s/jc4jo4t26ma2mlx/border.jpg?dl=0

[31] S. R. Rajagopalan, L. Sankar, S. Mohajer, and H. V. Poor, "Smart meter privacy: A utility-privacy framework," in *Proc. 2nd Annu. IEEE Conf. Smart Grid Commun.*, Brussels, Belgium, Oct. 2011, pp. 190–195.

[32] T. Idé, A. C. Lozano, N. Abe, and Y. Liu, "Proximity-based anomaly detection using sparse structure learning," in *Proc. SIAM Int. Conf. Data Min.*, Philadelphia, PA, USA, 2009, pp. 97–108.

[33] R. Banirazi and E. Jonckheere, "Geometry of power flow in negatively curved power grids: Toward a smart transmission system," in *Proc. IEEE Conf. Decis. Control (CDC)*, 2010, pp. 6259–6264.

[34] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "MATPOWER steady-state operations, planning and analysis tools for power systems research and education," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, Feb. 2011.

[35] J. Lofberg, "YALMIP: A toolbox for modeling and optimization in MATLAB," in *Proc. IEEE Int. Symp. Comput.-Aided Control Syst. (CACSD)*, Taipei, Taiwan, Sep. 2004, pp. 284–289. [Online]. Available: http://users.isy.liu.se/johanl/yalmip/

[36] K. C. Toh, M. Todd, and R. H. Tutuncu, "SDPT3—A MATLAB software package for semidefinite programming," *Optim. Methods Softw.*, vol. 11, nos. 1–4, pp. 545–581, 1999.

Hanie Sedghi (S'09) received the B.Sc. degree in electrical engineering and the M.Sc. degree in electrical engineering, communications from the Sharif University of Technology, Tehran Iran, in 2007 and 2009, respectively. She is currently pursuing the Ph.D. degree in electrical engineering with a minor in mathematics from the University of Southern California, Los Angeles, CA, USA.

Since 2013, she has been a Visiting Researcher at MEGA Data Laboratory, University of California–Irvine, Irvine, CA, USA. Her current research interests include large-scale machine learning, high-dimensional statistics, and probabilistic models. Specifically, she has worked on designing algorithms for stochastic optimization in high dimension with tight convergence bounds, and learning discriminative models such as deep networks, mixture models, and conditional random fields.

Ms. Sedghi was the recipient of the Provost's Fellowship from the University of Southern California, in 2010. She also served as a Reviewer for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON SMART GRID, the IEEE JOURNAL OF SELECTED AREAS IN COMMUNICATIONS, and several IEEE conferences.

Edmond Jonckheere (F'91) received the electrical engineering degree from the University of Louvain, Leuven, Belgium; the Dr.-Eng. degree in aerospace engineering from Universit Paul Sabatier, Toulouse, France; and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in 1973, 1975, and 1978, respectively.

From 1973 to 1975, he was a Research Fellow of the European Space Agency. From 1975 to 1978, he was a Teaching Assistant, Research Assistant, and a Research Associate with the Department of Electrical Engineering—Systems, University of Southern California. In 1979, he was with the Philips Research Laboratory, Brussels, Belgium. In 1980, he was with the University of Southern California, where he is currently a Full Professor of Electrical Engineering and Mathematics, a Member of the Center for Applied Mathematical Sciences, and a Member of the Center for Quantum Information Science and Technology. He held short-term visiting appointments at the Max-Planck-Institute, Gttingen, Germany; Australian National University, Canberra, ACT, Australia; Cardiff University, Wales, U.K.; and Swansea University, Wales. He also held consulting affiliations with the Memorial Medical Center of Long Beach, Long Beach, CA; Lockheed-Martin, Bethesda, MD, USA; the Aerospace Corporation, El Segundo, CA; and Honeywell, Morristown, NJ, USA. His current research interests include conventional versus quantum networks, adiabatic quantum computations, and power grid.

Dr. Jonckheere is a fellow of the Institute of Electrical and Electronics Engineers for contribution to the spectral theory of linear-quadratic and H-Infinity problems.