On the Conditional Mutual Information in the Gaussian-Markov Structured Grids

Hanie Sedghi and Edmond Jonckheere

Abstract The Supervisory Control and Data Acquisition (SCADA) State Estimator (SE) and the Phasor Measurement Units (PMU's) network constitute the communication infrastructures meant to provide the "smart grid" dispatcher with wide-area bus phase angles and other data, from which the operational status of the grid can be assessed— if the measurements are not compromised somewhere along their way to the SCADA dispatch and/or the PMU concentrator. Unfortunately, this is precisely what happens under the so-called "false data injection." In this Chapter, we develop a fast test for measurement data integrity, based on the Gaussian Markov Random Field (GMRF) assumption on the PMU data. This assumption, fundamental to this Chapter, is supported by (i) the great many fluctuating generations and variable loads justifying the Gaussian distribution assumption and, as more specifically addressed in this Chapter, (ii) the DC power flow equations from which an approximate 1-neighbor property of the bus phase angles is derived. The latter topological property refers to the conditional mutual information between two random variables being non-vanishing if and only if the nodes at which they are observed are linked in the edge set of the corresponding graph. Under the Gaussian distribution assumption, the conditional mutual information is easily computable from the conditional covariance. Then it is shown that Conditional Covariance Test (CCT) together with the walk-summability and the local separation property of grid graph allows the reconstruction of the grid graph from uncompromised measurement data. On the other hand, with corrupted data, CCT reconstructs only a proper subset of the edge set of the grid graph, hence triggering the alarm.

Department of Electrical Engineering,University of Southern California, Los Angeles, CA 90089,
e-mail: {hsedghi, jonckhee}@usc.edu

1 Introduction

1.1 The smart grid and its possibly malicious events

We are concerned with fast and reliable detection of threats in the power grid. This extra capability of the grid to detect a malicious event, even when it is triggered by a sophisticated antagonistic player, is among the attributes that makes it *smart*.

Traditionally, the term *grid* is used to refer to an electricity system that supports the following four operations: electricity generation, electricity transmission, electricity distribution, and voltage stability control. In the early days, generation was co-located with distribution in what we would now call a *micro-grid* and the connections among the micro-grids were meant to transmit energy in case of such contingencies as shift in the supply/demand balance. After deregulation, however, a large-scale generation-transmission-distribution *network* became the substitute for the traditional generation-distribution *co-location.* The new network allows consumers to purchase electricity at the cheapest price across the country, as opposed to the former concept in which consumers were forced to purchase electricity from local utility companies. Other considerations calling for an overhaul of the electricity system include the reduction of carbon emission, an objective that cannot be achieved without a significant contribution from the electricity sector. This calls for a bigger share of the renewable energy resources in the generation mix and a supply/demand that must be managed more effectively. Management and control of the grid made increasingly complex by its response to electricity market conditions are, next to its ability to detect contingencies, the most fundamental attributes that make it *smart*.

Automated large scale management requires considerable exchange of information, so that the smart grid has become a two-commodity flow—electricity and information. By utilizing modern information technologies, the smart grid is capable of delivering power in a more efficient way and responding to wider ranging conditions.

Massive amount of measurements and their transmission across the grid by modern information technology, however, make the grid prone to attacks. Next to malicious events, the potential for fault events with cascading impact on the overall stability of the power grid remains. Today's power systems are not adequately equipped with fault diagnosis mechanisms against various attacks and non-malicious events such as lines sagging in trees, as it had happened right before the 2003 blackout. Thus, there is an urgent need for quick assessment of fault events so that corrective feedback control actions can be taken promptly to avoid cascading events. Fast and accurate detection of possibly malicious events is of paramount importance not only for preventing faults that may lead to blackouts, but also for routine monitoring and control tasks of the smart grid, including state estimation and optimal power flow. Fault localization in the nation's power grid networks is known to be challenging, due to the massive scale and inherent complexity.

1.2 State Estimator (SE) versus Phasor Measurement Units (PMU's)

Traditionally, the State Estimator (SE) processes the measurement data from the power meters to reconstruct the state (bus voltages and phase angles). More recently, however, synchronous Phasor Measurement Units (PMU's) with GPS time stamps have been deployed across the grid and are considered the most reliable sensing information to monitor the state of operation of the grid and, if necessary, to respond to contingencies. Even though PMU's are more reliable than SE's, for economical reasons, some parts of the grid will still use state estimators in a foreseeable future. Therefore, any attack—either tampering with the power measurement to the SE or compromising the PMU data, as shown in Figure 6—can harm the power grid.

1.3 Outline of method

In a nutshell, the conceptual foundation of our method is reconstruction of the graphical model of the phase data.

We use Conditional Covariance test for this goal. CCT algorithm can be summarized as follows: Given $i, j \in \mathcal{V}$, given a separator S, that is, a subset of $\mathcal{V} \setminus \{i, j\}$, find the correlation between X_i and X_j given the separator. If for all reasonably chosen separators this conditional covariance remains above a certain threshold, then (i, j) is declared an edge. Under some conditions, the resulting $(\mathcal{V}, \mathcal{E})$ is the Markov graph of **X**.

Next, it is shown that, under normal grid operation, and because of the grid graph structure, the Markov graph of phasors should match the power grid graph; otherwise, a discrepancy triggers the alarm.

It turns out that our method can detect the most recently contrived attack on the smart grid, which specifically fools the State Estimator, and against which no counter-measures have been suggested thus far [24]. The attack is deemed "sophisticated" in the sense that it assumes knowledge of bus-branch model of the grid.

1.4 Related work and exclusivity of approach

The line fault detection method of Reference [8] is also based on a GMRF model of the PMU data. Besides the fault versus attack detection discrepancy in motivation, the difference between [8] and the present work is two-fold. First, probably the most important contribution of this Chapter is to show that the one-neighbor property is just an approximation, and issue that was not addressed in [8]. Secondly, the

fault detection method of [8] utilizes PMU's, whereas here we utilize both PMU's and State Estimator, as shown in Fig. 6. For economical reasons, future grids will still contain state estimators in some parts, therefore, [8] and any other monitoring system that does not have a method to check for data integrity can be deluded by such an attack as false data injection.

2 Gaussian Markov Random Field (GMRF): General concept

2.1 Graphical Models

Probabilistic graphical models provide diagrammatic representation of probability distributions. This way they set up a simple way to visualize the structure of a probabilistic model and provide insight into properties of the model including conditional independence properties [4].

A graph consists of *nodes* $\mathcal{V}$ connected by *links* $\mathcal{E}$. In a probabilistic graphical model, each nodes represents a random variable or a group of random variables and the links express the probabilistic dependence relationship between random variables. The graph represents how joint probability distribution can be decomposed in factors that depend only on a subset of variables [4].

There are two major classes of graphical models: namely, *Bayesian Networks*, also known as *directed graphical models* were links are directed, and *Markov Random Fields*, also known as *undirected graphical models* where links are not directed [4].

2.2 Gaussian Markov Random Field (GMRF)

A probability distribution is said to have *global Markov property* with respect to a graph if, for any disjoint subsets of nodes I, J, S such that S separates I and J on the graph, the distribution satisfies $X_I \perp X_J | X_S$, i.e., X_I is independent of X_J conditioned on X_S. This is represented in Figure 1.

A distribution is *pairwise Markov* with respect to a given graph if, for any two nodes i and j in the graph such that there is no direct link in the graph between i and j, then X_i is independent of X_j given the states of all of the remaining nodes, i.e., $X_i \perp X_j | X_{\mathcal{V} \setminus \{i,j\}}$.

A set of random variables is said to have *local Markov property* corresponding to a graph [13] if any variable X_i is conditionally independent of all other variables X_{-i} given its neighbors $X_{N(i)}$, where $-i := \{j \in \mathcal{V} : j \neq i\}$ and $N(i) := \{j \in \mathcal{V} : (i,j) \in \mathcal{E}\}$. Local Markov property can be seen in Figure 2.

Given an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, a set of random variables $X = (X_v)_{v \in \mathcal{V}}$ form a *Markov Random Field* with respect to $\mathcal{G}$ if they have the global Markov property. It should be noted that local Markov property and pairwise Markov property are

Fig. 1 Global Markov property:
$X_I \perp X_J | X_S$

J

Separator S

I
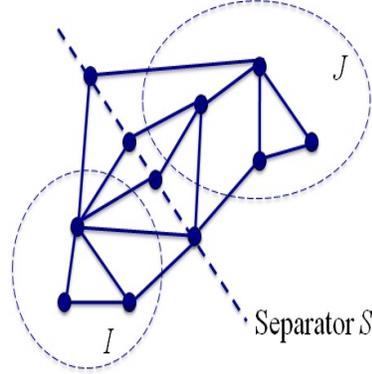
Fig. 2 Local Markov property:
$E[X_i | X_{N(i)}] = E[X_i | X_{-i}]$

$N(i)$

-1

i
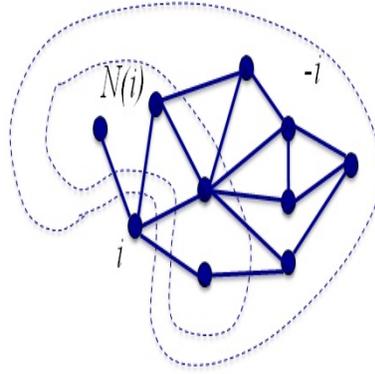
equivalent and they are a special case of global Markov property. For a strictly positive probability distribution, the properties are equivalent and it can be shown that the probability distribution can be factorized with respect to the graph [13].

One instance of this positivity condition happens in case of jointly Gaussian distributions.

A *Gaussian* Markov Random Field (GMRF) is a family of jointly Gaussian distributions that factor in accordance with a given graph. Given a graph $\mathscr{G} = (\mathscr{V}, \mathscr{E})$, with $\mathscr{V} = \{1, ..., p\}$, consider a vector of Gaussian random variables $\mathbf{X} = [X_1, X_2, ..., X_p]^T$, where each node $i \in \mathscr{V}$ is associated with a scalar Gaussian random variable X_i. A *Gaussian Markov Random Field (GMRF)* on $\mathscr{G}$ has a probability density function (pdf) that may be parametrized as

$$f_X(x) \propto exp[-\frac{1}{2}x^T J x + h^T x], \tag{1}$$

where J is a positive-definite symmetric matrix whose sparsity pattern corresponds to that of the graph $\mathscr{G}$. More precisely,

$$J(i,j) = 0 \Longleftrightarrow (i,j) \notin \mathscr{E}. \tag{2}$$

The matrix $J = \Sigma^{-1}$ is known as the *potential* or *information* matrix, the non-zero entries $J(i,j)$ as the edge potentials, and the vector $\mathbf{h}$ as the vertex potential vector [2].

In general, the graph $\mathscr{G} = (\mathscr{V}, \mathscr{E})$ is called the *Markov graph* (graphical model) underlying the joint probability distribution $f_X(x)$, where the node set $\mathscr{V}$ represents each random variable X_i, if the edge set $\mathscr{E}$ is defined in order to satisfy local Markov property. For a Markov Random Field, local Markov property states that $X_i|\mathbf{X}_{-i} = X_i|X_{N(i)}$, where $\mathbf{X}_{-i}$ denotes all variables except for X_i and $X_{N(i)}$ denotes all random variables associated with the neighbors of i.

Defining

$$r_{ij} \triangleq \frac{\Sigma(i,j|\mathscr{V} \setminus \{i,j\})}{\sqrt{\Sigma(i,i|\mathscr{V} \setminus \{i,j\})\Sigma(j,j|\mathscr{V} \setminus \{i,j\})}}, \tag{3}$$

as the *partial correlation coefficient* between variables X_i and X_j for $i \neq j$ measuring their conditional covariance given all other variables. The joint distribution of the GMRF X follows $N(\mu, (I-R)^{-1})$, with $\Sigma = (I-R)^{-1}$ is the covariance matrix and $R \triangleq [r_{ij}]$ is the matrix consisting of partial correlation coefficients off the diagonal and zeros on the diagonal entries [22], i.e.,

$$r_{ij} = -\frac{J(i,j)}{\sqrt{J(i,i)J(j,j)}}. \tag{4}$$

Therefore,

$$X_i|\mathbf{X}_{-i} \sim N(\mu_i + \sum_{j \neq i} r_{ij}(X_j - \mu_j), 1), \tag{5}$$

where the distribution is normalized to highlight the partial correlations r_{ij}.

Setting $X \sim N(\mu, J^{-1})$, the pairwise Markov property of GMRF implies that $(i,j) \notin \mathscr{E} \Leftrightarrow r_{ij} = 0$.

3 Bus phase angles as Gaussian Markov Random Field (GMRF)

This section is the "hub" of this whole Chapter. Specifically in this section, we examine the extent to which the bus phase angles of the power grid satisfy the conditions for them to qualify as a GMRF. We discuss the approximation in neighboring property between bus phase angles.

Further, in Section 4.4, we explain Conditional Covariance Test [2] as the method we have chosen for finding out the Markov graph of bus phase angles. Next, we explain why CCT method best describes this approximation and why this approxi-

mation is in fact true for a grid graph.

Finally, in Section 5, we argue a discrepancy between the output of CCT with grid graph structure means that the system is under stealthy deception attack.

3.1 AC Power Flow: review

The AC power flow states that real power and the reactive power flowing from bus i to bus j are, respectively,

$$P_{ij} = G_{ij}V_i^2 - G_{ij}V_iV_j\cos(\theta_i - \theta_j) + b_{ij}V_iV_j\sin(\theta_i - \theta_j), \tag{6}$$

$$Q_{ij} = b_{ij}V_i^2 - b_{ij}V_iV_j\cos(\theta_i - \theta_j) - G_{ij}V_iV_j\sin(\theta_i - \theta_j), \tag{7}$$

where V_i and θ_i are the voltage magnitude and phase angle, resp., at bus #i and G_{ij} and b_{ij} are the conductance and susceptance, resp., of line ij. From [3], we obtain the following approximation of the AC *fluctuating* power flow:

$$\widetilde{P}_{ij} = (b_{ij}\overline{V}_i\overline{V}_j\cos\overline{\theta}_{ij})(\widetilde{\theta}_i - \widetilde{\theta}_i), \tag{8}$$

$$\widetilde{Q}_{ij} = (2b_{ij}\overline{V}_i - b_{ij}\overline{V}_j\cos\overline{\theta}_{ij})\tilde{V}_i - (b_{ij}\overline{V}_i\cos\overline{\theta}_{ij})\widetilde{V}_j, \tag{9}$$

where bar means steady-state value, tilde means 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 voltages we have $\overline{V}_m = \overline{V}_k \simeq 1 p.u.$ (per unit), and the fluctuations in angles are about the same such that $\cos\theta_{km} = 1$, we have

$$\widetilde{P}_{ij} = b_{ij}(\widetilde{\theta}_i - \widetilde{\theta}_j), \tag{10}$$

and

$$\widetilde{Q}_{ij} = b_{ij}(\widetilde{V}_i - \widetilde{V}_j). \tag{11}$$

3.2 Gaussian distribution assumption: Transmission versus distribution network

The power flow equations can be written, conceptually, as $z = h(x)$, where $z = (P^T, Q^T)^T$ is the vector of (active and reactive) powers injected at the various buses and $x = (\theta^T, V^T)^T$ is the state, that is, the vector of voltage phase angles and voltage magnitudes at the buses.

Whether the Gaussian distribution assumption on θ is justified depends on two considerations:

1. The nature of the injected power, which could be deterministic or stochastic.

2. The linearized approximation of $z = h(x)$, the *DC power flow equations.*

Regarding the first item, in the high-voltage transmission grid, the aggregate property of the demand justifies the Gaussian distribution assumption. On the other hand, the loads in the low-voltage power distribution network do not correspond to aggregate loads but single consumers. Hence the Gaussian distribution assumption cannot be justified on the ground of the demand. The Gaussian distribution assumption can, however, be justified by the aggregation of such renewables as wind turbines and solar panels, the power output of which is inherently random. It is suggested in [18] that as few as 5 wind turbines would suffice to satisfy the Central Limit Theorem, meaning that the power generation would behave like a Gaussian random variable.

We note that the capability of detecting false data is also interesting in the distribution grid, where the high number of inexpensive sensors deployed in the grid could be hardly managed via secure communication channels[1]. For our method— which relies on the Gaussian distribution assumption—to be applicable to the distribution network, it is hence imperative to invoke the aggregation of the renewables. In the general setting where renewables need not be present, our work more realistically applies to the transmission grid, where aggregate demand is present.

Regarding the second item, if aggregate power at buses follows Gaussian distribution, by linearity of DC power flow we can reach the same conclusion for bus phase angles.

3.3 DC Power Flow: Active power versus phase angle

We now apply the preceding to bus phase angles. We would like to show that bus phase angles form a GMRF and then discuss the Markov graph associated with it. The DC power flow model [1] is often used for analysis of power systems in normal steady-state operations. When the system is stable, the phase angle differences are small. In addition, DC power flow assumes that lines are highly inductive. Therefore $\sin(\theta_i - \theta_j) \sim \theta_i - \theta_j$. Thus, the power flow on the transmission line connecting bus i to bus j is given by

$$P_{ij} = b_{ij}(X_i - X_j), \tag{12}$$

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:

$$P_i = \sum_{j \neq i} P_{ij} = \sum_{j \neq i} b_{ij}(X_i - X_j). \tag{13}$$

In the above formulation, the summation holds since $b_{ij} = 0$ is implied whenever bus i and j are not connected. Thus, it follows that the phasor angle at bus i could

---

[1] This was brought to our attention by an anonymous referee.

be represented as

$$X_i = \sum_{j \neq i} c_{ij} X_j + \frac{1}{\sum_{j \neq i} b_{ij}} P_i, \tag{14}$$

where $c_{ij} = \frac{b_{ij}}{\sum_{i \neq j} b_{ij}}$.

Because of load uncertainty in the transmission network, the injected power can be modeled as a random variable [15] and since injected power models the superposition of many independent factors (e.g. loads), it can be modeled as a Gaussian random variable, as already argued in Section 3.2. Thus, the linear relationship in (12) implies that the difference of phasor angles across a bus could be approximated by a Gaussian random variable truncated within $[0, 2\pi)$. Considering the fixed phasor at the slack bus, it is assumed that under steady-state, phasor angle measurements can be considered as Gaussian variables [8].

The next step is to find the correct neighboring relationship between the X_i's.

3.4 Local Markov property: Neighboring relationship

Here we investigate the extent to which the θ's are in a 1-neighbor relationship, by which we mean the local Markov property, $E(X_i | X_{-i}) = E(X_i | X_{N(i)})$. We look at two idealized cases: an infinite chain-structured bus system and a 2-dimensional lattice-structured bus system.

In Section 3.4.1 dealing with the idealized chain bus, we consider (13) along with independently injected powers P_i's and demonstrate a two-neighbor relationship between the X_i's, i.e., the X_i's are related to their first and second degree neighbors in the grid graph; precisely, $E(X_i | X_{-i}) = E(X_i | X_{N(i)} \cup X_{N(N(i)) \setminus \{i\}})$. This implies that the $J(i, j)$ matrix entry in Eq. (1) is nonvanishing if and only if i and j are 1-neighbors or 2-neighbors in the grid graph; in other words, $J(i, j) = 0$, $\forall d_{\text{hopt}}(i, j) \geq 3$, where $d_{\text{hop}}(\cdot, \cdot)$ denotes the hop metric defined as the distance on the graph when the link weights are normalized to 1. Furthermore, using the Toeplitz structure of the coefficient matrix of the system of equations (13) and Fourier transform techniques, Section 3.4.1 shows that $J(i, j') < J(i, j)$, for $2 = d_{\text{hop}}(i, j') > d_{\text{hop}}(i, j) = 1$, that is, the J matrix entry in Eq. (1) for second-neighbor is smaller than the J matrix entry for the first-neighbor. It is shown in Section 3.4.1 that this approximation falls under the generic fact of the tapering off of Fourier coefficients.

In Section 3.4.2, a similar result is demonstrated to hold for the idealized 2-dimensional lattice-structured grid.

Thus, we can approximate the neighboring relationship to be that of immediate neighbors in grid graph,

$$E[X_i | X_{-i}] \simeq E[X_i | X_{N(i)}]. \tag{15}$$

Therefore, we have an *approximate* local Markov property. It is conjectured that such an approximation holds whenever the grid has enough symmetry to allow for Toeplitz and related Fourier transform techniques.

Fig. 3 Infinite Line Network
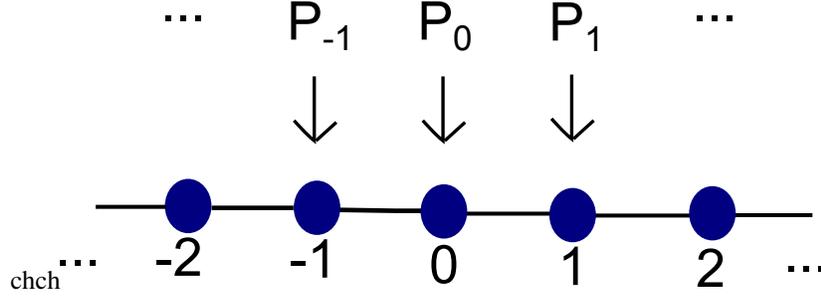
3.4.1 Independent Power Injection to an Infinite Chain

Consider a doubly infinite homogeneous chain-structured power network with $b_{i,i+1} = 1$ as shown in Figure 3.

The DC power flow equations, $P = BX$, in this specific case take the format

$$
\begin{bmatrix} \vdots \\ P_{-1} \\ P_0 \\ P_1 \\ \vdots \end{bmatrix} = \begin{bmatrix} \ddots & \ddots & & & \\ \ddots & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & -1 & 2 & \ddots \\ & & & \ddots & \ddots \end{bmatrix} \begin{bmatrix} \vdots \\ X_{-1} \\ X_0 \\ X_1 \\ \vdots \end{bmatrix}. \tag{16}
$$

Because of the symmetry of the problem, B is a doubly-infinite Toeplitz matrix, also referred to as Laurent operator. By "Toeplitz matrix," we mean a matrix whose (i, j) entry depends only on the difference of indexes, $i - j$, equivalently, a matrix with constant entries on the diagonal, constant entries on the super-diagonals, and constant entries on the sub-diagonals is a Toeplitz matrix; as can be seen from Eq. (16).

Besides the usefulness of chains as testbeds for networks with shift invariant properties, here, the most compelling justification is that the doubly infinite chain structure secures $\sum_{k=-\infty}^{+\infty} P_k = 0$, as easily seen from the cancellation of the sum of the column elements of B, but subject to some convergence issues, which are now straightened out.

To remain within the Hilbert space setup, we restrict $X \in \ell^2(-\infty, +\infty)$. Next, it can be seen that the B-operator is bounded. Indeed, taking the Fourier transform of P and X,

$$
\widehat{P}(e^{j\alpha}) = \sum_{k=-\infty}^{+\infty} P_k e^{jk\alpha}, \quad \widehat{X}(e^{j\alpha}) = \sum_{k=-\infty}^{+\infty} X_k e^{jk\alpha}, \tag{17}
$$

we get

$$
\widehat{P}(e^{j\alpha}) = (2 - e^{j\alpha} - e^{-j\alpha})\widehat{X}(e^{j\alpha}). \tag{18}
$$

From the above, we notice that B is a multiplication operator in the Fourier domain:

$$\widehat{P}(e^{j\alpha}) = \widehat{B}(e^{j\alpha})\widehat{X}(e^{j\alpha}), \quad \widehat{B}(e^{j\alpha}) := 2(1 - \cos\alpha), \tag{19}$$

where $\widehat{B}$ is referred to as the *symbol* of the operator. Clearly, the multiplication operator $\widehat{B} : L^2[0, 2\pi] \to L^2[0, 2\pi]$ is bounded and since the Fourier transform is an Hilbert space isometry the operator $B : \ell^2(-\infty, +\infty) \to \ell^2(-\infty, +\infty)$ is bounded as well. This secures $P \in \ell^2(-\infty, +\infty)$ and hence gives sense to $\sum_{k=-\infty}^{+\infty} P_k = 0$.

In order to determine the neighboring structure of a chain-generated random phase angle vector **X**, we assign a normal distribution to P. This results in X having a Gaussian distribution:

$$f_X(P) \sim e^{-\frac{1}{2}P^T \Sigma_d^{-1} P} = e^{-\frac{1}{2}X^T B^T \Sigma_d^{-1} BX}, \tag{20}$$

where the covariance Σ_d is a trace class operator, a condition necessary to secure $\int_{\ell^2(-\infty,+\infty)} e^{-\frac{1}{2}P^T \Sigma_d^{-1} P} \Pi_{k=-\infty}^{+\infty} dp_k < \infty$ along with the Gaussian property of the projection of the infinite dimensional distribution on a finite dimensional space [19, Prop. 1.8], [17]. We take $\Sigma_d = \mathrm{diag}\{\sigma_{d,k}^2 : k = ..., -1, 0, +1, ...\}$ with $\sigma_{d,k} = 1$ for $|k| \leq d$, and $\lim_{|k|\to\infty} \sigma_{d,k} = 0$ with $\sigma_{d,k} > 0$ for $|k| > d$, and such that $\sum_{k=-\infty}^{+\infty} \sigma_{d,k}^2 < \infty$. With this covariance,

$$\lim_{d\to\infty} e^{-\frac{1}{2}X^T B^T \Sigma_d^{-1} BX} = e^{-\frac{1}{2}X^T B^2 X} \tag{21}$$

where B^2 is doubly-infinite Toeplitz as well with symbol

$$\widehat{B^2}(e^{i\alpha}) = [2(1 - \cos\alpha)]^2 = 6 - 8\cos\alpha + 2\cos 2\alpha. \tag{22}$$

From the above, it follows that

$$f_X(x) \propto \exp\left(-\frac{1}{2}X^T \begin{bmatrix} \ddots & \ddots & \ddots & \ddots & \ddots \\ & 1 & -4 & 6 & -4 & 1 \\ & & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix} X\right). \tag{23}$$

According to (23), we can see a two-neighbor correlation between the X_i's. It can also be seen that the coefficients for the second-neighbors are smaller than those of the first neighbors. It should be noted that a power grid is not infinite, hence the infinite Toeplitz structure is an idealization.

3.4.2 Euclidean lattice

The preceding can be generalized to an infinite 2-dimensional Euclidean lattice. 2-dimensional Euclidean lattice is depicted in figure 4. Given a 2-dimensional lattice with vertices with integer coordinates $\{(k,l)\}_{k,l\in\mathbb{Z}}$, the neighboring relationship is $N((k,l)) = \{(k\pm 1, l), (k, l\pm 1)\}$. In other words, the susceptance $b_{(k,l),(m,n)}$ between nodes (k,l) and (m,n) is nonvanishing only if either $m = k \pm 1$ and $l = n$ or $m = k$ and $n = l \pm 1$. As in (17), we define the 2-dimensional Fourier transforms as

Fig. 4 Euclidean Lattice
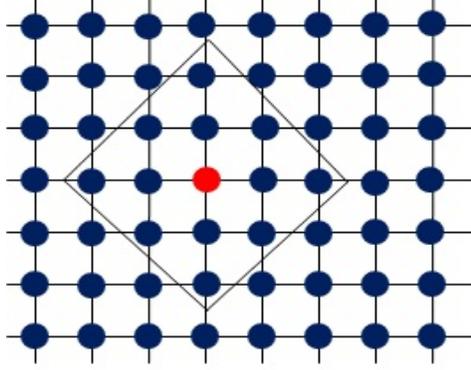
$$\widehat{P}(e^{j\alpha}, e^{j\beta}) = \sum_{k,l \in \mathbb{Z}} P_{k,l} e^{jk\alpha} e^{jl\beta}, \quad \widehat{X}(e^{j\alpha}, e^{j\beta}) = \sum_{k,l \in \mathbb{Z}} X_{k,l} e^{jk\alpha} e^{jl\beta}.$$

As in (19), the DC power flow equations can be written as

$$\widehat{P}(e^{j\alpha}, e^{j\beta}) = \widehat{B}(e^{j\alpha}, e^{j\beta}) \widehat{X}(e^{j\alpha}, e^{j\beta}), \quad \widehat{B}(e^{j\alpha}, e^{j\beta}) = \left(4 - e^{j\alpha} - e^{-j\alpha} - e^{j\beta} - e^{-j\beta}\right),$$

where $\widehat{B} : L^2([0, 2\pi)^2) \to L^2([0, 2\pi)^2)$ is the susceptance operator. In order to write the equivalent of (20) for a 2-dimensional lattice, we use Parseval's theorem as a representation of $\sum_{k,l \in \mathbb{Z}} P_{k,l}^2$ as a quadratic function of $X_{k,l}$:

$$f_X(P) \propto e^{-\frac{1}{2} \sum_{k,l \in \mathbb{Z}} P_{kl}^2} = e^{-\frac{1}{2} \frac{1}{2\pi} \oint \oint |\widehat{B}|^2 |\widehat{X}|^2 d\alpha d\beta}$$

Quadratic functions of *2-indexed* variables do not lend themselves to obvious matrix representation. The guiding idea here is to collect those $X_{k,l}$'s that are contributing to $\sum_{k,l \in \mathbb{Z}} P_{k,l}^2$. Those $X_{k,l}$'s are the coefficients of the zeroth powers of $e^{j\alpha}$ and $e^{j\beta}$ in the integrand. Given

$$
\begin{aligned}
|\widehat{B}|^2 =\, & 20 - 8e^{j\alpha} - 8e^{-j\alpha} - 8e^{j\beta} - 8e^{-j\beta} \\
& + 2e^{j(\alpha+\beta)} + 2e^{-j(\alpha+\beta)} + 2e^{j(\alpha-\beta)} + 2e^{-j(\alpha-\beta)} \\
& + e^{2j\alpha} + e^{-2j\alpha} + e^{2j\beta} + e^{-2j\beta}
\end{aligned}
$$

and

$$|\widehat{X}|^2 = \sum_{k,l,m,n \in \mathbb{Z}} X_{k,l} X_{m,n} e^{j(k-m)\alpha} e^{j(l-n)\beta}$$

it is not hard to see that

$$\sum_{k,l\in\mathbb{Z}} P_{k,l}^2 = \sum_{k\in\mathbb{Z}} (20X_{k,k}^2 - 8X_{k,k}X_{k+1,k} - 8X_{k,k}X_{k-1,k} - 8X_{k,k}X_{k,k+1} - 8X_{k,k}X_{k,k-1}$$
$$+ 2X_{k,k}X_{k+1,k+1} + 2X_{k,k}X_{k-1,k-1} + 2X_{k+1,k-1} + 2X_{k,k}X_{k-1,k+1}$$
$$+ X_{k,k}X_{k+2,k} + X_{k-2,k}X_{k,k} + X_{k,k}X_{k,k+2} + X_{k,k}X_{k,k-2})$$

Clearly, $\sum_{k,l\in\mathbb{Z}} P_{k,l}^2$ is quadratic in the $X_{k,l}$ variables, but those variables that are multiplied have their indexes within at most a 2-neighbor relationship in the lattice structure. To be somewhat more specific, what we learn over the 1-dimensional case is that the correlations decay with the ℓ^2 distance on the lattice. Indeed, for $d_{\ell^2}((k,k),(k+1,k)) = 1$, the canonical correlation $r_{(k,k),(k+1,k)} \propto 8$; for $d_{\ell^2}((k,k),(k+1,k+1)) = \sqrt{2}$, the canonical correlation $r_{(k,k),(k+1,k+1)} \propto 2$; and for $d_{\ell^2}((k,k),(k+2,k)) = 2$, the canonical correlation $r_{(k,k),(k,k+2)} \propto 1$.

As a word of technical warning, the $f_X(P)$ expression should have been written $e^{-\frac{1}{2}\sum_{k,l\in\mathbb{Z}}\Sigma_{d,(k,l)}^{-1}P_{k,l}}$, where $\sum_{k,l}\Sigma_{d,(k,l)} < \infty$ and $\Sigma_{d,(k,l)} = 1$ for $||(k,l)||_{\ell^2} \le d$ and $\Sigma_{d,(k,l)} \downarrow 0$ as $||(k,l)||_{\ell^2} \to \infty$. This brings some tempered coefficients in the correlations, which have no effect unless for $||(k,l)||_{\ell^2} \to \infty$. Working out this technicality explicitly would have, however, resulted in substantial clutter in the notation.

3.5 Reactive power versus voltage amplitude

It is clear from (10)-(11) that we can follow the same discussions we had about real power and voltage angles, with reactive power and voltage magnitudes.

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 Equation (11), voltage fluctuations around the steady-state value can be approximated as 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 (11) with (12) makes it clear that the same matrix, i.e., the B matrix developed in Section 3.4, 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 be readily seen by comparing (10) and (11).

Therefore, voltage magnitudes provide another perspective for developing a graphical model underlying the grid structure.

The dual of our approach (linear relationship between reactive power and voltage magnitude) could be generalized to include line loss by linearizing Equation (2) of [5] to produce a linear relationship between voltage, active and reactive power.

4 Model Selection

In the context of graphical models, model selection means finding the real underlying Markov graph among a group of random variables based on samples of those random variables. There are two main class of methods for learning the structure of the underlying graphical model, convex and non-convex methods. ℓ_1-regularized maximum likelihood estimators are the main class of convex methods [7, 9, 21]. 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. Other types of methods are the non-convex or greedy methods [2]. As we are faced with GMRF in our problem, it would be useful to exploit one of these structure learning methods.

We have decided to use the new Gaussian Graphical Model Selection method called *Conditional Covariance Test (CCT)* [2].

It is proven in [2] that two nodes are connected in the Markov graph iff the conditional mutual information between those measurements is greater than a threshold. For Gaussian variables, testing conditional mutual information is equivalent to Conditional Covariance Test.

In order to have structural consistency, the model should satisfy two important properties:

1. α-walk-summability,
2. (γ, η)-local separation property.

4.1 α-walk summability

A Gaussian model is said to be α-walk-summable if $||\bar{\mathbf{R}}|| \leq \alpha < 1$ where $\bar{\mathbf{R}} = [|r_{ij}|]$ and $||.||$ denotes the spectral or 2-norm of matrix, which for symmetric matrices is given by the maximum absolute eigenvalue [2]. r_{ij} is defined in (3) and (4). Considered power grids satisfy this criteria.

4.2 Local separation property

An ensemble of graphs has the (η, γ)-local separation property if for any $(i, j) \notin \mathscr{E}(\mathscr{G})$, the maximum number of paths between i, j of length at most γ does not exceed η [2]. Local separator concept is depicted in figure 5.

The power grid structure is an example of bounded local path graphs that satisfy the local separation property.

Fig. 5 Local Seperation Property: $\gamma = 3$. $N(i) = \{a, b, c\}$ is the neighborhood of i and the $\gamma-$local separator set $S(i, j; G, \gamma) = \{a, c\}$.
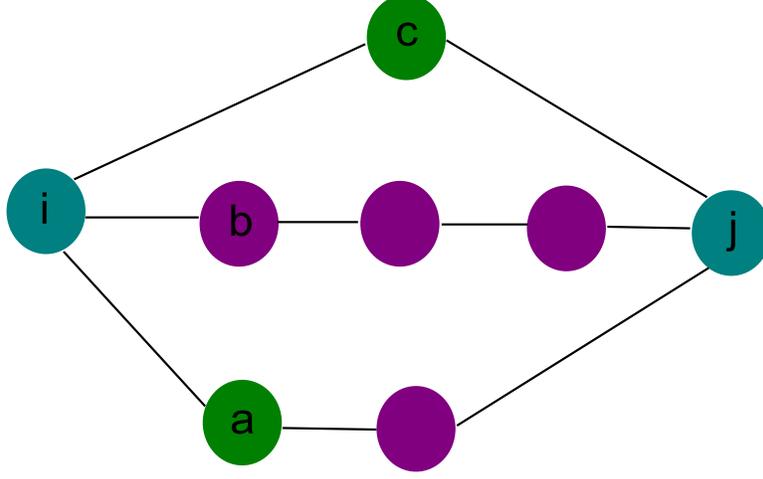
4.3 Conditional Mutual Information

Mutual information between two random variables is a quantity that measures the mutual dependence between the two random variables. In the case of continuous random variables, mutual information between random variables X and Y can be defined as

$$I(X;Y) = \int_X \int_Y f(x,y) log\left(\frac{f(x,y)}{f(x)f(y)}\right) dxdy, \tag{24}$$

where $f(x,y)$ is the joint probability density function of X and Y, and $f(x)$ and $f(y)$ are the marginal probability density functions of X and Y respectively.

Mutual information can be defined in terms of entropies as follows:

$$
\begin{aligned}
I(X;Y) =\ & H(X) - H(X|Y) \\
=\ & H(Y) - H(Y|X) \\
=\ & H(X) + H(Y) - H(X,Y) \\
=\ & H(X,Y) - H(X|Y) - H(Y|X),
\end{aligned}
\tag{25}
$$

where $H(X)$ and $H(Y)$ are the marginal entropies, $H(X|Y)$ and $H(Y|X)$ are the conditional entropies, and $H(X,Y)$ is the joint entropy of X and Y.

In the special case of Gaussian distributed random variables, for a N-dimensional Gaussian random vector $\mathbf{Z}$, we have

$$H(\mathbf{Z}) = \frac{1}{2}\log((2\pi e)^N |\Sigma|), \tag{26}$$

where Σ is the covariance matrix of $\mathbf{Z}$ [6]. This implies that the mutual information between two N-dimensional Gaussian variables is

$$I(\mathbf{X};\mathbf{Y}) = \frac{1}{2}\log\left(\frac{|\Sigma_{\mathbf{XX}}||\Sigma_{\mathbf{YY}}|}{|\Sigma_{\mathbf{XY}}|}\right), \tag{27}$$

where

$$\Sigma = \begin{bmatrix} \Sigma_{\mathbf{XX}} & \Sigma_{\mathbf{XY}} \\ \Sigma_{\mathbf{YX}} & \Sigma_{\mathbf{YY}} \end{bmatrix}. \tag{28}$$

The *conditional mutual information* is, in its most basic form, the expected value of the mutual information of two random variables given the value of a third, that is,

$$I(X;Y|Z) = E_z(I(X;Y)|Z). \tag{29}$$

This can be rewritten as [16]

$$I(X;Y|Z) = H(X,Z) + H(Y,Z) - H(X,Y,Z) - H(Z). \tag{30}$$

Conditional mutual information can also be written in terms of conditional entropies:

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z). \tag{31}$$

Therefore, considering two Gaussian random variables X_i, X_j, conditional mutual information between these two random variables conditioned on a set of random variables X_S is given by (see [2])

$$I(X_i;X_j|X_S) = -\frac{1}{2}\log\left[1 - \rho^2(i,j|S)\right], \tag{32}$$

where $\rho(i,j|S)$ is the conditional correlation coefficient, given by

$$\rho(i,j|S) := \frac{\Sigma(i,j|S)}{\sqrt{\Sigma(i,i|S)\Sigma(j,j|S)}}. \tag{33}$$

As a result, for Gaussian random variables, for testing conditional independence, testing conditional mutual information is equivalent to testing conditional covariances [2].

If the distributions deviate from Gaussian, the conditional mutual information can still be derived from (32)-(33), provided $\Sigma(i,j|S)$ is interpreted as the conditional correlation of $g_i(X_i)$ and $g_j(X_j)$, where the g_i's are nonlinear processing functions aimed at maximizing the correlation.

To be more precise, let $\rho_{g_i,g_j,g}(i,j|S)$ be the correlation coefficient between $g_i(X_i)$ and $g_j(X_j)$ conditioned upon $g(X_S)$, where g_i, g_j, and g are measurable functions. Then by nonlinear processing of X_i and X_j with the distortion functions g_i and g_j, the canonical correlation $\rho_{g_i,g_j,g}(i,j|S)$ can be made to increase towards the mutual information :

$$\sup_{g_i,g_j,g} \left(-\frac{1}{2} \log \left(1 - \rho^2_{g_i,g_j,g}(i,j|S) \right) \right) \leq I(X_i; X_j | X_S).$$

(See [10, Cor. 1].) Furthermore, the supremum can be achieved if $g_i(X_i)$ and $g_j(X_j)$ can be made jointly Gaussian conditioned upon $g(X_S)$ (see [10, Th. 3]). A computational procedure that precisely implements this idea is available in [10, Sec. 6]. A related computational implementation is the sequential selection [11, 12]. A simplified numerical procedure based on the canonical correlation between the powers of X_i and the powers of X_j is available in [23].

4.4 Conditional Covariance Test (CCT)

Conditional Covariance Test is introduced in [2]. Using CCT method, the conditional covariance is computed for each node pair $(i,j) \in \mathcal{V}^2$ and the conditioning set that achieves the minimum, over all subsets of other nodes of cardinality at most η, is found. If the minimum value exceeds the threshold $\xi_{n,p}$, then the node pair is declared as an edge.

It is shown in [2] that under walk-summability the effect of faraway nodes on covariance decays with the distance and the error in approximating the covariance by local neighboring relationship decays exponentially with the distance. Thus by correct tuning of threshold and enough number of samples, we expect the output of CCT method to follow the grid structure.

It has been shown that this method is superior to the ℓ_1 method [7, 21] as CCT distributes edges fairly uniformly across the nodes while the ℓ_1 method tends to cluster all the edges together between the "dominant" variables leading to a densely connected component and several isolated points [2]. Therefore, CCT is more suitable for constructing the structure of the power grid from measurements.

5 Stealthy Deception Attack

The most recent false data injection attack on the power grid has recently been introduced in [24]. For a P-bus electric power network, the $l = 2P - 1$ dimensional state vector $\mathbf{x}$ is $(\theta^T, V^T)^T$, where $\mathbf{V} = (V_1, ..., V_P)$ is the vector of voltage bus magnitudes and $\theta = (\theta_2, ..., \theta_P)$ the vector of phase angles disregarding the slack bus for which $\theta_1 = 0$. It is assumed that the nonlinear measurement model for the state estimation is defined by

$$z = h(x) + \varepsilon, \tag{34}$$

where $h(.)$ is the nonlinear measurement-valued function and $\mathbf{z}$ is the m-dimensional measurement vector consisting of active and reactive power measurements. $H(x^k) :=$

Fig. 6 Power grid under a cyber attack
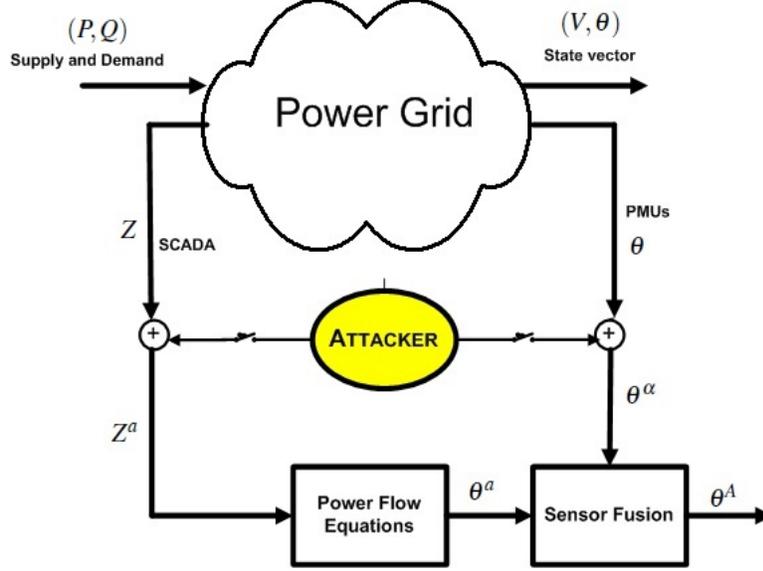
$\left\{ \left. \frac{\partial h_i(x)}{\partial x_j} \right|_{x=x^k} \right\}_{1\leq i\leq m;\, 1\leq j\leq l}$ denotes the *Jacobian matrix* of the measurement model $h(x)$ at x^k.

According to [24], the goal of a stealthy deception attacker is to compromise the measurements available to the State Estimator (SE) as

$$z^a = z + a, \tag{35}$$

where $\mathbf{z^a}$ is the corrupted measurement and $\mathbf{a}$ is the attack vector. Vector $\mathbf{a}$ is designed such that the SE algorithm converges and the attack a is undetected by the Bad Data Detection (BDD) scheme. That is, the difference between z^a and the $h(x^k)$ is less than the BDD threshold. In addition, for the targeted set of measurements, the estimated values at convergence are closest to the ones compromised by the attack. The goal of attacker is to inject some data into the state estimator such that the system does not recognize that the data is manipulated and acts upon that. Then it is shown that, subject to some limitations, such an attack can be performed with $a \in \text{Im}(H)$. The attack vector $\mathbf{a}$ is designed in such a way that the difference between z^a and z is the desired value. Figure 6 represents the attack.

It is also stated that the introduced attack is only valid if performed locally. The attack is performed under the DC flow assumption. Because of this assumption, only the $H_{P\theta}$ block of the H matrix is considered in the attack calculation and the state vector introduced in [24] reduces to the vector of voltage angles, $\mathbf{X}$. Since $a \in \text{Im}(H)$,

$$z^a = z + a = H(X + d). \tag{36}$$

Thus, we have

$$Hd = z^a - HX = a, \tag{37}$$

where $H = H_{P\theta}$, z^a is the attacker's goal and $\mathbf{X}$ is the phasor angle vector. Considering (13), 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. Clearly, H is structured as a weighted graph Laplacian. By "weighted graph Laplacian structure," we mean a symmetric matrix with its (i, j) entry that can be interpreted as the negative of the "conductance" of the (i, j) link and its (i, i) diagonal element equal to minus the sum of the other elements in row #i or column #i. This is clearly a generalization of the combinatorial graph Laplacian, where the "conductances" are normalized to 1.

Analysis of (36) and (37) shows that *the Markov graph of an attacked system changes from the grid graph.* We use this to trigger the alarm.

It should be emphasized that the attack considered here assumes the knowledge of the system's bus-branch model. Hence under this scheme the attacker is equipped with a wealth of information. Yet, we can detect such a *strong* attack with our method.

6 Simulation

We considered a 9-node grid suggested by Zimmerman et al. [20]. The structure is shown in Figure 7. First, we fed the system with Gaussian demand and simulated the power grid. We used MATPOWER [20] for solving the DC power flow equations for various demand and used the resulting angle measurements as the input to CCT algorithm. We used YALMIP [14] and SDPT3 [25] to perform CCT.

With the right choice of parameters and threshold, and enough *un-compromised* measurements, the Markov graph follows the grid structure. Table 1 shows the edit distance between the Markov graph and the grid graph that is used to lead us to the correct threshold.

Table 1: Normalized edit distance under CCT for Figure 7, measurement size=400

Threshold	No. of Links of Markov graph	Edit Distance
0.0037	10	1
0.0038	**9**	**0**
0.0039	7	2

Next we introduced the stealthy deception attack to the system. We considered the cases where 2, 3 or 4 nodes are under attack. For each case, we simulated all possible attack combinations. In all attack scenarios, the Markov graph of tampered

Fig. 7 Evaluated 9-node network
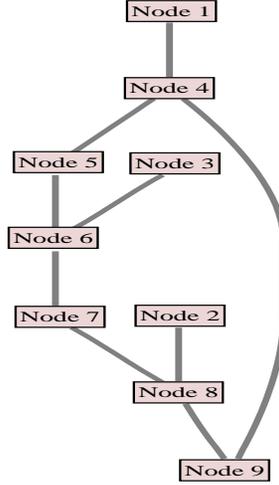
measurements lacked at least one link that was present in grid graph, a discrepancy that triggered the alarm. Thus we successfully detected the attack. Table 2 summarizes different attack scenarios and the corresponding detection ratio.

It should be noted that table 1 shows the required samples for tuning the method to a specific network structure. So, it shows the initialization step for any network which is enough to do once at the beginning as far as the network topology remains the same. Simulation results show that in this network even if only 1 of the samples is a corrupted sample it is enough to have 100% detection rate.

It should be noted that Since we have made connections between phase angle measurements Markov graph and power grid graph, the method can be performed in a decentralized manner. In addition, as stated in [2], the complexity of CCT method is polynomial.

Table 2: Stealthy deception attack on the grid shown in Fig. 7

No. of attacked nodes	Detection Ratio
2	100
3	100
4	100

7 Conclusion

We have shown that such statistical learning techniques as the Conditional Co-variance Test—which is equivalent to test on conditional mutual information in the Gaussian case—allows us to reconstruct the topology of the power grid as the Markov graph of the phase angle measurements (or the voltage magnitudes) at the buses. One of the main points of this Chapter is that phase angle data only approximately satisfies the Markov property relative to the grid, a fact that was overlooked in [8]. As shown in Section 3.4, correlations indeed extend beyond the one-neighbor relationship. Nevertheless, since the farther away neighboring relationship is weaker and less significant than the one-neighbor relationship, as shown here on the 9-bus system, this difficulty can be overcome by correctly choosing the threshold.

Finally, if the phase angle data is compromised, the reconstructed Markov graph will be different from the grid interconnection, even when the attack is "stealthy" and launched with the knowledge of the bus-branch model.

In further work, we would like to demonstrate the same concepts and results on more realistic bus systems.

References

1. A. Abur and A.G Exposito. *Power System State Estimation, Theory and Implementation*. Marcel Dekker, 2004.

2. A. Anandkumar, V. Tan, F. Huang, and A.S. Willsky. High-dimensional gaussian graphical model selection:walk summability and local separation criterion. *Journal of Machine Learning*, June 2012. accepted.

3. R. Banirazi and E. Jonckheere. Geometry of power flow in negatively curved power grid. In *49th IEEE Conference on Decision and Control*, pages 6259–6264, Dec 17-20 2010.

4. C. M. Bishop. *Pattern recognition and machine learning*. Springer, New York, USA, 2006.

5. S. Bolognani and S. Zampieri. A distributed control strategy for reactive power compensation in smart microgrids. Preprint. ArXiv:1106.5626,Oct. 2012.

6. M. Borga. Learning Multidimensional Signal Processing. Technical Report SE-581 83, Linkping University, Sweden, 1998.

7. J. Friedman, T. Hastie, and R. Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 2007.

8. M. He and J. Zhang. A dependency graph approach for fault detection and localization towards secure smart grid. *IEEE Transactions on Smart Grid*, 2:342–351, June 2011.

9. M. Janzamin and A. Anandkumar. High-dimensional covariance decomposition into sparse markov and independence models. Preprint. ArXiv:1211.0919, Nov. 2012.

10. E. Jonckheere and B.-F. Wu. Mutual Kolmogorov-Sinai entropy approach to nonlinear estimation. In *IEEE Conference on Decision and Control*, pages 2226–2232, Tucson, Arizona, Dec. 1992. Available at `http://eudoxus2.usc.edu/CHAOS/Wu_Jonckheere_Mutual Kolmogorov-Sinai.pdf`.

11. W. E. Larimore. Identification and filtering of nonlinear systems using canonical variate analysis. In M. Casdagli and S. Eubank, editors, *Nonlinear Modeling and Forecasting, SFI Studies in the Sciences of Complexity*, volume 12. Addison-Wesley, 1992.

12. W. E. Larimore and J. Baillieul. Identification and filtering of nonlinear systems using canonical variate analysis. In *29th IEEE Conference on Decision and Control*, pages 635–640, Honolulu, HI, December 1990.

13. S.L. Lauritzen. *Graphical models: Clarendon Press*. Clarendon Press, 1996.

14. J. Lofberg. YALMIP: A Toolbox for Modeling and Optimization in MATLAB. In *IEEE international symposium on Computer Aided Control Systems Design (CACSD)*, September 2004. Available from http://users.isy.liu.se/johanl/yalmip/.

15. M. Luettgen, W. Karl, A. Willsky, and R. Tenney. Multiscale representations of markov random fields. *IEEE Transaction on Signal Processing*, 41:33773396, Dec 1993.

16. K. Makarychev, Y. Makarychev, A. Romashchenko, and N. Vereshchagin. A new class of non-Shannon-type inequalities for entropies. *Communications in Information and Systems*, 2(2):147–166, December 2002.

17. Stefania Maniglia and Abdelaziz Rhandi. Gaussian measures on separable hilbert spaces and applications. Technical report, Lecture Notes of the University of Lecce, Italy, Quaderno 1/2004, 2004. ISBN: 88-8305-010-X.

18. J. Mur-Amada and J. Salln-Arasanz. From turbine to wind farms - technical requirements and spin-off products. In Gesche Krause, editor, *Phase Transitions and Critical Phenomena*, volume 18, pages 101–132. InTech, April 2011.

19. G. Da Prato. *An Introduction to Infinite-Dimensional Analysis*. Springer, Berlin, Heidelberg, 2000.

20. C. E. Murillo-Snchez R. D. Zimmerman and R. J. Thomas. Matpower steady-state operations, planning and analysis tools for power systems research and education. *Power Systems, IEEE Transactions on*, 26(1):12–19, Feb. 2011.

21. P. Ravikumar, M.J. Wainwright, G. Raskutti, and B. Yu. High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electronic Journal of Statistics*, (4):935–980, 2011.

22. H. Rue and L. Held. *Gaussian Markov random fields : theory and applications*. CRC, 2005.

23. Khushboo Shah, Edmond Jonckheere, and Stephan Bohacek. Dynamic modeling of Internet traffic for intrusion detection. *EURASIP Journal on Advances in Signal Processing*, 2007:Article ID 90312, 14 pages, 2007. doi:10.1155/2007/90312.

24. 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 *IFAC World Congress*, September 2011.

25. K. C. Toh, M.J. Todd, and R. H. Tutuncu. SDPT3 - a MATLAB software package for semidefinite programming. *Optimization Methods and Software*, 11:545–581, 1999.