IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON DIGITAL FORENSICS THROUGH MULTIMEDIA SOURCE INFERENCE

Received April 27, 2019, accepted May 17, 2019, date of publication May 31, 2019, date of current version July 2, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2920326

A Deep and Scalable Unsupervised Machine Learning System for Cyber-Attack Detection in Large-Scale Smart Grids

HADIS KARIMIPOUR[1], (Member, IEEE), ALI DEHGHANTANHA[1], (Senior Member, IEEE), REZA M. PARIZI[2], (Member, IEEE), KIM-KWANG RAYMOND CHOO[3], (Senior Member, IEEE), AND HENRY LEUNG[4], (Fellow, IEEE)

[1]University of Guelph, Guelph, ON N1G 2W1, Canada
[2]Kennesaw State University, GA 30060, USA
[3]University of Texas at San Antonio, TX 78249, USA
[4]University of Calgary, Calgary, AB T2N 1N4, Canada

Corresponding author: Ali Dehghantanha (adehghan@uoguelph.ca)

ABSTRACT Smart grid technology increases reliability, security, and efficiency of the electrical grids. However, its strong dependencies on digital communication technology bring up new vulnerabilities that need to be considered for efficient and reliable power distribution. In this paper, an unsupervised anomaly detection based on statistical correlation between measurements is proposed. The goal is to design a scalable anomaly detection engine suitable for large-scale smart grids, which can differentiate an actual fault from a disturbance and an intelligent cyber-attack. The proposed method applies feature extraction utilizing symbolic dynamic filtering (SDF) to reduce computational burden while discovering causal interactions between the subsystems. The simulation results on IEEE 39, 118, and 2848 bus systems verify the performance of the proposed method under different operation conditions. The results show an accuracy of 99%, true positive rate of 98%, and false positive rate of less than 2%

INDEX TERMS Anomaly, cyber-attack, smart grid, statistical property, machine learning, unsupervised learning.

I. INTRODUCTION

Today's power systems consist of a network of sensors and generators that allow two way communication within the system's infrastructure as well as reliable energy production through integration of Distributed Energy Resources (DERs) and Advanced Metering Infrastructure (AMI). While this complex communication system has tremendous advantage, by improving energy efficiency, reliability, and manageability, it increases the system's vulnerabilities to cyber-attacks due to the tremendous number of devices and access points that operate outside the traditional administrative domain. Since failures in the power grid may lead to catastrophic events, it is highly important to investigate the effects of cyber-attacks in a power system.

As reported in [1], lack of system awareness is the main reason in the North American blackouts, which highlight the importance of cyber-attack analysis to maintain a stable and reliable operation of the power supply. A cyber-attack can result in overload that will damage the equipment, or false demand request which can result in lots of energy generated [2]–[4]. Besides, a malicious attack can also cause false negatives, i.e., false overload condition in a power system. Other disruptions in different parts of the smart grid, electric vehicle infrastructure, is also possible. It is shown in [5], [6] that malicious attacks by blocking communications with a device can stop services in substation computers. Therefore, real time cyber-attack detection is paramount for the reliable performance of the critical infrastructure including smart grids. Online and continuous system monitoring is a requirement to detect targeted cyber-attacks and achieve attack resilience [7].

In general, individual sensors in a large-scale network are the main target of security compromises. A compromised insider can easily access information stored in a compromised node. In theory, key revocation of any compromised

The associate editor coordinating the review of this manuscript and approving it for publication was Zhen Qin.

node is possible by applying an authentication mechanism to sensor networks. However, authentication approaches based on cryptography or security gateway design, such as the one described in [8], [9], are infeasible due to the computation and storage constraints of the system. The existing studies within the smart power grid context mainly focus on the networking security of the cyber elements [10]–[12], advanced anomaly detection techniques [13], [14], and secure control theories based on different state estimation techniques [15]. A detailed analysis about presence of cyberattack in a power system is described in [16].

Although the above mentioned solutions are capable of immunizing the power systems, majority of them are mathematically too expensive, physically impractical and not scalable for large-scale complex network. Nowadays, huge amount of data is generated all over the grids which increase accessibility for real-time system monitoring. Exploring these data greatly enhances the performance monitoring, diagnosis, and prognosis of anomaly in complex systems. Historical data describing the system's operation can help identify anomalies and potential attacks. However, traditional Bad Data Detection (BDD) techniques are not prepared for real time computational and storage issues due to the large-volume of data generated in the smart grid. These challenges opens up the possibility of using data analytical techniques, such as Machine Learning (ML), to tackle complex structure data sets with AI to detect and prevent cyber-attacks. ML algorithms can be used to analyze various combinations of measurements through AMI, states, and control actions by learning their patterns [17], [18]. It can detect False Data Injection (FDI) attack by learning the non-linear, complex relationship between measurements. This can be done in a similar fashion to successful techniques applied to other power system problems as seen in the research literature [19].

There are limited studies on the application of ML on cyber-security of the smart grids. Several ML algorithms are tested and compared in [20] for detection of FDI attacks. General conclusions was made about the success of machine learning in classifying FDI attacks. [21] proposed a hybrid intrusion detection method based on common path mining method to detect abnormal power system events from PMU data, relays, and energy management system (EMS) logs. A cyber-attack detection techniques based on the correlation between two PMU parameters using Pearson correlation coefficient was used in [22]. This method analyzed the change of correlation between two PMU parameters using Pearson correlation coefficient. Authors in [20] utilized Gaussian process combined with ML to model the attack strategy for anomaly detection. In [23] a supervised ML–based scheme is proposed to detect a cyber-deception assault in the state estimation process. A deep learning method which recognize important features of FDI attacks in real-time is also proposed in [24].

Performance of the existing, data-driven attack detection techniques can be improved using Probabilistic Graphical Models (PGM) to model complex system behavior. Among PGMs, Dynamic Bayesian Networks (DBN) are useful tools which can represent complex systems evolving in time using the causal relationships between system components [25]. Moreover, new techniques should be developed to handle the complex and high dimensional data to maintain the robustness, scalability and accuracy of the attack detection mechanisms. To reduce the computational burden in large data sets, feature extraction can be used to transform the original features into a more meaningful representation by reconstructing its inputs and it involves reducing the amount of resources required [26], [27]. Detection techniques that do not rely on pre-classified training data are essential, as there exists anomalies which cannot be measured or simulated.

In this work, we propose a smart grid anomaly detection method to extract the patterns of changes in FDI attacks. The revealed features are employed to detect the attacks in real-time. Symbolic Dynamic Filtering (SDF) is used to build a computationally efficient feature extraction scheme to discover causal interactions between the smart grids sub-systems through DBN. Mutual Information (MI), DBN and learning algorithms are used to detect unobservable cyber-attacks based on free energy as the anomaly index. Our goal is to capture dependencies between variables through associating of a scalar energy to each variables, which serves as a measure of compatibility. The scalability of the proposed technique is examined on various IEEE test systems which was modeled on PSS/E software. The results show high accuracy and low false alarm under different operation conditions. It should be mentioned that the proposed method does not only relies on the pattern in the training data sets but It also uses the concept of free energy to differentiate between the energy level in the attacked and normal data sets. Therefore, even new and unseen attacked can be detected.

The main contributions of this work are as follows:

- Formulation of an unsupervised approach to detect an anomaly in smart grids without labeling data sets.
- Proposing a scalable method by reducing computational burden through data reduction by SDF.
- Developing a strong learning model based on DBN.
- Proposing a model-free approach, which can be employed in hierarchical and topological networks for different attack scenarios.

The rest of the paper is organized as follows. Mathematical formulations are described in Section II. Proposed cyber-attack detection method is presented in Section III. Section IV discusses the case studies and simulation results followed by the conclusion in Section V.

II. MATHEMATICAL MODELING

A. GENERATOR'S MODEL

In this work, smart grid is modeled as a multi-agent, cyber-physical system where each of these agents include a generator, a measurement device, a distributed control agent, and an energy storage system that can inject or absorb real power in the system [28]. The dynamic and static state of the system

are described as follows:

$$\dot{x} = f(x, u, \eta)$$
$$z = h(x, u, \varepsilon) \quad (1)$$

where x is the system state including the dynamic state of the generator (e.g. rotor speed and rotor angle) and the static state of the network (voltage magnitude and phase angle). f (.) describes the non-linear, dynamic behavior of the generators and h (.) is the measurements non-linear function. u and z represent the output and measurements vector, respectively.

The 4-th order (two-axis) model of generator i's can be described as [29]:

$$\begin{aligned}
\dot{\delta}_i &= \Omega_s \Delta\omega_i \\
\dot{\omega}_i &= \frac{\omega_s}{2H_i}(P_{Mi} - P_{Ei} - D_i\Delta\omega_i) \\
\dot{E}'_{qi} &= \frac{1}{T'_{di}}\left(-E'_{qi} - \left(X_{di} - X'_{di}\right) I_{di} + V_{fi}\right) \\
\dot{E}'_{di} &= \frac{1}{T'_{qi}}\left(-E'_{di} + \left(X_{qi} - X'_{qi}\right) I_{qi}\right) \\
E'_{qi} &= V_{qi} + R_{ai}I_{qi} + X'_{di}I_{di} \\
E'_{di} &= V_{di} + R_{ai}I_{di} - X'_{qi}I_{qi}
\end{aligned} \quad (2)$$

where $\dot{()}$ denotes the time derivative. Generator parameters are described using Table 1.

TABLE 1. Generator parameter description.

Parameter	Description
δ	rotor angle
$\Delta\omega$	rotor speed
Ω_s	system frequency
D	coefficient of damping
E'_d	transient electromotive force in d-axis
E'_q	transient electromotive force in q-axis
V_f	field voltage
H	machine inertia constant per unit
I_d, I_q	stator current in d-axis and q-axis
R_a	armature resistance
X_d, X_q	d-axis and q-axis reactance
X'_d, X'_q	d-axis and q-axis transient reactance
T'_d, T'_q	d-axis and q-axis open loop time constant
P_E	electrical output torque
P_M	mechanical input torque

For synchronous generator i, excitation system controls the field voltage, the mechanical torque is controlled by the associated speed governor, and the electrical output can be calculated as follows:

$$P_{Ei} = E'_{di}I_{di} + E'_{qi}I_{qi} + \left(X'_{qi} - X'_{di}\right) I_{di}I_{qi}. \quad (3)$$

Let E_i denote the internal voltage of generator i, then P_{Ei} can be expressed as [30]:

$$P_{Ei} = \sum_{k=1}^{N} |E_i|\,|E_k|\,(G_{ik}\cos(\delta_i - \delta_k) + B_{ik}\sin(\delta_i - \delta_k)) \quad (4)$$

where $G_{ik} = G_{ki}$ and $B_{ik} = B_{ki}$ are the conductance and susceptance between generators i and k, respectively.

In this work, the goal is to learn and predict the dynamic behavior of the smart power grid (where generators are modeled as explained in this section) to detect anomaly/cyber-attacks. SDF, DBN, and RBM are used to develop a computationally efficient tool for discovering the interactions between the subsystems.

B. ATTACK REPRESENTATION

Traditionally, the integrity of the state estimation process is verified through BDD method by computing the L-norm of measurement residual [31]. The presence of bad data is determined if

$$\left\|z - H\hat{x}\right\| > T_r \quad (5)$$

where $z \in R^N$ is the measurement vector, $\hat{x} \in R^D$ is the estimated state vector, and $H \in R^{N\times D}$ is the Jacobian matrix.

A threshold T_r is pre-defined to maintain the accuracy of the state estimation. Aside from the fact that cyber-attacks bypass the existing BDD technique, measurement redundancy required for BDD approaches makes them impractical for smart grid technology. In intelligent cyber-attacks, specifically FDI attacks, the goal of the adversary is to control a subset of the measurements and manipulate the state variables arbitrarily. It can be done by injecting a false data vector $z_a \in R^N$ which by pass traditional BDD techniques. Suppose the malicious attack intentionally manipulates the meter readings by z_a. Accordingly, the attack- incurred measurement change can be written as:

$$z = H\hat{x} + z_a + \epsilon = H\left(\underbrace{\hat{x} + c_a}_{\hat{x}_a}\right) + q_a + \epsilon \quad (6)$$

where ϵ is the measurement noise, and $\hat{x}_a$ is the faulty estimated state.

The injected false data (z_a) can be decomposed into two parts $a = Hc_a$ and q_a, where $c_a \in R^D$ is an injected vector of data which bypass BDD tests since it lies in the column space of H, and q_a is the only detectable part that lies in the complementary space where $H\left(H^TH\right)^{-1}H^Tq_a = 0$. In other words, the stealth attack vectors (z_a) always exists even if the adversary can get partial access to the network topology and line parameters to construct malicious attacks that completely lie in (H), i.e., $q_a = 0$, thereby bypassing the existing BDD methods [32].

The following assumptions are considered in the model of the attack:

- In this work, the assumption is that the attacker has limited resources and could only manipulate limited number of measurement readings. This could be either power injection or power flow data, for a time period $T_a \subseteq T$. This is a realistic assumption because, in the context of power networks it is not realistic to assume that all sensors report faulty measurements at the same time.

Moreover, in reality, compromising all measurements results in huge cost and effort for attackers.

- Complete knowledge of the system is literally impossible for an outsider. Therefore, the attacker has partial knowledge of the system topology and security mechanisms. Such knowledge can be obtained by statistical analysis of data sent from the remote terminal units (RTUs) to the control center or by physically capturing the security information embedded in a node.

In this work, strategic sparse FDI attack with least absolute shrinkage and selection operator (LASSO) is considered. Jacobian matrix (H) is decomposed based on a row-wise approach. A sub Matrix $H^S = \left(H_{j_i,:}, H_{j_{N-|S|,:}}\right)$, of H is created to represent the secure measurements, where $H_{j_i,:}$ is the j_i-th row of H, such that $H^S c_a = 0$. Likewise, sub-matrix H^A is constructed for attacked measurements. Finally, the attacker's strategy is defined in a way to find a solution c_a which optimize following objective function:

$$\begin{aligned} &\text{Minimize } \left\|H^A c_a\right\|_0 \\ &\text{Subject to } H^S c_a = 0, \\ &\qquad \|c_a\|_\infty \geq \tau, \end{aligned} \tag{7}$$

where $\tau \geq 0$ is a given constant. The optimization problems is solved using LASSO and Regressor Selection algorithms. More details about the attack construction is available in [33].

The goal of the attacker is to manipulate rotor speed and angle through FDI attack by hacking into the communication network. Hence, $\forall t \in T_a$, for generator i, the effect of FDI attacks on the system state can be written as:

$$x_i^a(t) = x_i(t) + \gamma_i x_i(t) + C_i \tag{8}$$

where the γ_i is a constant coefficient and C_i represents a constant bias in the attacked states. In other word, the attacker is interested to alter the system state by γ_i and C_i. Considering that, the attacker will design z_a in a way that the attack vector remains unobservable for the operator and traditional BDD methods. In the experiments, we assume that the attacker has access to λ measurements, which are randomly chosen to generate a λ-sparse attack vector.

III. PROPOSED ENERGY-BASED CYBER-ATTACK DETECTION

In this section, a cyber-attack detection framework is proposed which utilize DBN modeling, feature extraction through MI and RBM for data training. DBN and MI are applied to smart grid test systems with extensive measurements, and the RBM is used to capture the patterns in system behaviour that are extracted by the unsupervised DBN model (data are not labeled).

The proposed data driven framework for anomaly detection is depicted in Fig. 2. At first, the system is partitioned into several sub-systems. Then causal dependency between nominal characteristics of subsystems are learned using SDF. The proposed method is a computationally efficient tool, which reduce the computational burden by: 1) selecting a subset of measurements through feature selection and SDF, and 2) by domain decomposition and data processing on several subsystems in parallel, rather than dealing with whole system at once.

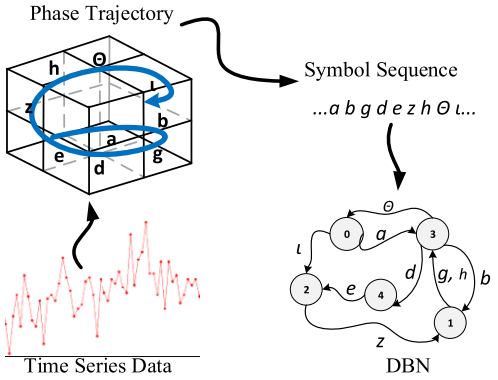


FIGURE 1. Illustration of the steps to generate DBN using SDF-based feature extraction.

A. SYMBOLIC DYNAMIC FILTERING

In the proposed feature extraction method based on SDF, the time series data are first converted into symbol sequences, and then DBN are defined from these sequences to compress the information into low-dimensional statistical patterns. The phase space of the system in Eq. (1) is divided into a finite number of cells. A compact region Ω is identified by introducing a partition $B \equiv \{B_0, \ldots, B_{m-1}\}$ consisting of m mutually exclusive (i.e., $B_j \bigcap B_k = \emptyset \ \forall j \neq k$) and exhaustive ($\bigcup_{j=0}^{m-1} B_j = \Omega$) cells. The dynamic system describes the time-series data as $O \equiv \{\beta_0, \ldots, \beta_{m-1}\}, \beta_i \in \Omega$, which passes through the cells of the partition B [34], [35]. To understand the concepts of partitioning and mapping into the symbol alphabet, consider the system shown in Fig. 1 [34].

Consider the cell visited by a trajectory as a random variable S with symbol value $s \in \mathcal{A}$. Symbol alphabet is the set $\mathcal{A}$ of m different symbols that mark the elements in the partition. Every initial state $\beta_0 \in \Omega$ produce a series of symbols which can be defined by mapping from the phase space into the symbol space as follows:

$$\beta_0 \rightarrow \ s_{i0}\, s_{i1} \ldots s_{ik} \tag{9}$$

Eq. (8) is called symbolic dynamics. The symbolization process converts multi-dimensional space into a symbol sequence, and then into a DBN.

B. DYNAMIC BAYESIAN NETWORKS

DBNs are probabilistic graphical models that can demonstrate system's state as a set of variables, and model the probabilistic dependencies of the variables between time steps. In this work, a high order DBN on ξ variables $x_t = \{x_{1,t}, \ldots, x_{\xi,t}\}$ at different time points $t = 1, \ldots, T$ is considered. Each $x_{i,t}$ represents the expression of state i at

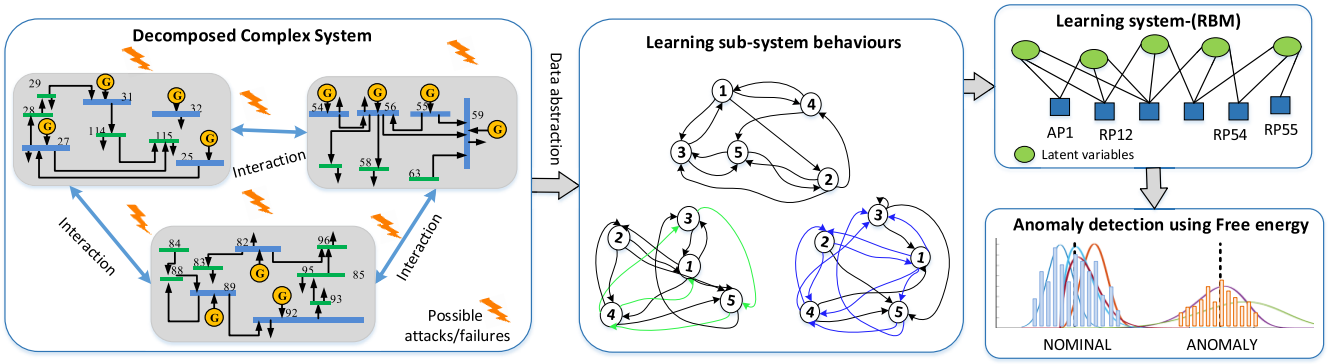


FIGURE 2. Proposed framework for cyber-attack detection using unsupervised learning.

time t. Symbol sequence is extracted from the variables set by SDF. To find the occurrence probability for a new symbol s_n, we assume that the DBN satisfies the $\mathcal{L}$-th order Markov property:

$$P\left(s_n|s_{n-1}\dots s_{n-\mathcal{L}}\dots s_0\right) = P\left(s_n|s_{n-1}\dots s_{n-\mathcal{L}}\right) \quad (10)$$

Thus, a state transition matrix Π which describes the $\mathcal{L}$-th order Markov chain can be defined based on the training data. The order of the model is set based on trial an error. Let the state at time instant k be denoted as q_k. The ij-th element of Π can be defined as follows:

$$\Pi_{ij} \triangleq P\left(q_{k+1} = s_i|q_k = s_j\right) \quad (11)$$

In this work, since we are dealing with several time series, we use a modified version of Markov chain ($x\mathcal{L}$-th order Markov chain) [36] to predict the occurrence probability for a new symbol in a series $\mathbb{A}$ using the last $\mathcal{L}$ symbol for another series $\mathbb{B}$. $\Pi^{\mathbb{A}}$ and $\Pi^{\mathbb{B}}$ are defined for $\mathcal{L}$-th order Markov representing sub-systems $\mathbb{A}$ and $\mathbb{B}$, respectively. The same way, causal dependencies of $\mathbb{A}$ on $\mathbb{B}$ and $\mathbb{B}$ on $\mathbb{A}$ can be represented by cross state transition matrices $\Pi^{\mathbb{AB}}$ and $\Pi^{\mathbb{BA}}$, respectively.

Features from $\mathcal{L}$-th order Markov chain are known as the atomic patterns (APs) and the one for $x\mathcal{L}$-th order Markov chain are referred as the relational patterns (RPs). State-transition matrices $\Pi^{\mathbb{AB}}$ and $\Pi^{\mathbb{BA}}$, can be described as:

$$\begin{aligned}\pi_{kl}^{\mathbb{AB}} &\triangleq P\left(q_{n+1}^{\mathbb{B}} = l|q_n^{\mathbb{A}} = k\right)\forall_n \\ \pi_{ij}^{\mathbb{BA}} &\triangleq P\left(q_{n+1}^{\mathbb{A}} = j|q_n^{\mathbb{B}} = i\right)\forall_n\end{aligned} \quad (12)$$

where $j, k \epsilon Q^{\mathbb{A}}$ and $i, l \epsilon Q^{\mathbb{B}}$, $Q^{\mathbb{A}}$ and $Q^{\mathbb{B}}$ are the state vector related to sequence $\mathbb{A}$ and $\mathbb{B}$, respectively.

Given a multivariate time series, the symbol sequences S is generated with partitioning. After that, a high order DBN is used to define the subsequent states and transition probabilities between the vertices. We use MI criteria to extract important feature of an AP or an RP. MI develops a generalized linear correlation coefficient that measures the relationship between two random variables. A non-zero value in MI means the two variables are independent towards each other. MI between state sequences q^A and q^B can be written as Importance metric $I^{\mathbb{AB}}$ as follows:

$$I^{\mathbb{AB}} = I\left(q_{k+1}^{\mathbb{B}}; q_{k+1}^{\mathbb{A}}\right) = H\left(q_{k+1}^{\mathbb{B}}\right) - H\left(q_{k+1}^{\mathbb{B}}|q_k^{\mathbb{A}}\right) \quad (13)$$

where,

$$\begin{aligned}\mathrm{H}\left(q_{k+1}^{\mathbb{B}}\right) &= -\sum_{i=1}^{Q^{\mathbb{B}}} P\left(q_{k+1}^{\mathbb{B}} = i\right)\log_2 P\left(q_k^{\mathbb{B}} = i\right) \\ \mathrm{H}\left(q_{k+1}^{\mathbb{B}}|q_k^{\mathbb{A}}\right) &= -\sum_{i=1}^{Q^{\mathbb{A}}} P\left(q_k^{\mathbb{A}} = i\right)\mathrm{H}\left(q_{k+1}^{\mathbb{B}}|q_k^{\mathbb{A}} = i\right) \\ \mathrm{H}\left(q_{k+1}^{B}|q_k^{a} = i\right) &= -\sum_{j=1}^{Q^{\mathbb{B}}} P\left(q_{k+1}^{\mathbb{B}} = j|q_k^{\mathbb{A}} = i\right) \\ &\quad \times \log_2 P\left(q_{k+1}^{\mathbb{B}} = j|q_k^{\mathbb{A}} = i\right)\end{aligned}$$

More details about the MI-based causality can be found in [32]. The variation of the MI matrix ($I^{\mathbb{AB}}$) between two-time periods can be driven as:

$$\delta\left(I\right) = I_{t_1}^{\mathbb{AB}} - I_{t_2}^{\mathbb{AB}} \quad (14)$$

Large δ means a strong predictive and informative link in AP or RP that can be used to distinguish the two kinds of end uses.

Once the models are ready, patterns of system's behaviour are learned by the RBM. Test data are used to compute the likelihood of the learned features. In this work, we used Restricted Boltzmann Machine (RBM) for this purpose.

C. RESTRICTED BOLTZMANN MACHINE

Boltzmann Machine is a generative method to model the unknown distribution of data. Unlike most of the Machin Learning techniques that only discriminate some data vectors in favor of others, Boltzmann Machine can also generate new data with given joined distribution, as well as pattern completion in case of missing inputs. It is also considered more feature-rich and flexible. RBM belongs to the class of stochastic Energy-based Models (EM) [38]. In EM, each state of the system is associated to an specific energy level. Such a system can be described by a network of stochastic binary neurons (a set of visible variables $v = \{v_1, \dots, v_N\}$) which are connected a set of hidden variables $h = \{h_1, \dots, h_K\}$.

System's state can be described based on joint configurations of the visible and hidden variables. It is proved that model estimation in RBM amounts to maximize the likelihood of the training data with low-energy state. As a result, an anomaly will appear as a configuration with low probability or high-energy [39]. Given binary variables v and hidden variables h, the joint probability of a state ($Pr\,(v, h)$) can be described based on the energy of that state ($En\,(v, h)$), with a Boltzmann distribution function:

$$Pr\,(v, h) = \frac{\exp(-En(v, h))}{\sum_{v,h} \exp(-En(v, h))} \tag{15}$$

where

$$En\,(v, h) = -\sum_{i=1}^{N} a_i v_i + \sum_{k=1}^{K} \left(b_k + \sum_{i=1}^{N} w_{ik} v_i \right) h_k \tag{16}$$

where a, b, and w are model parameters which are calculated through maximization of the probability of the training data with low-energy state.

Data density can be rewritten as:

$$Pr\,(v) \propto \sum_{h} \exp(-En(v, h)) = \exp(-F(v)) \tag{17}$$

where $F\,(v)$ is known as free-energy and can be rewritten as:

$$F\,(v) = -\log\,(\Pr\,(v)) + constant \tag{18}$$

Therefore, free energy can be used as the anomaly index to rank data instances in linear time. The trained RBM is employed to identify cyber-attack based on the probability and energy level of event. Anomaly is represented by an event with high energy or low probability. The assumption is that cyber-attacks change the interaction among the sub-systems and results in different patterns in DBN. For simplicity of training, $I^{\mathbb{AB}}$ can be normalized into binary states (0 and 1 for low and high values, respectively) for APs and RPs. Finally, changes in the parameters related to the accepted patterns are used to identify cyber-attacks. A distribution of free energy is used to detect low probability events or cyber-attacks based on distance metric. For the normal operation condition, free energy will have similar distribution to that of the training data. The assumption is that the training data are mostly collected from normal operation condition. Therefore, the learnt RBM can effectively capture the normal operation of the system.

To quantify the difference between the energy distributions in training and test data, Relative Entropy (RE) metric is used. The relative entropy between two probability distributions is a measure of the distance between them. RE for two probability distributions P and Q on a finite set X, can be described as [35], [36],

$$RE(P\|Q) = \sum_{X} P(x) \log \frac{P(x)}{Q(x)} \tag{19}$$

where P and Q refer to the distribution of free energies in the normal situation and under cyber-attack, respectively. Free energies in the normal operation condition ($F\,(v^n)$) and under cyber-attack condition ($F\,(v^{ca})$), can be calculated using Eq. (14). A symmetric RE distance can be defined as [38],

$$RE_d(P\|Q) = RE(P\|Q) + RE(Q\|P) \tag{20}$$

which can be used as an index for cyber-attack/anomaly detection. This index will be compared with a Detection Threshold (DT) to detect the cyber-attack. Too low thresholds may results in many false attack detection, while too high thresholds may lead to unidentified attack. In this work, most of the RE values calculated through training are assumed to be normal, while a few of them are outliers. To find the DT, the normal distribution is used as the baseline. The assumption is that 95% of the data are within two standard deviations of the mean. $\forall\ \widetilde{DT}$ satisfying $\left|\{RE_i{:}\widetilde{DT} \geq RE_i\}\right| = 0.95|\{RE_i\}|$, $i = 1, 2, \ldots,$ n that, $DT = min\{\widetilde{DT}\}$ where RE_i is the i-th RE in the training data. Then, anomaly is detected when $RE\,(t) \geq DT$. The steps can be summarized as follows:

- Transform time series data to symbolic sequence.
- Model the subsystems and their interactions using DBN.
- Evaluate the information based metric values using MI (I^{ij}).
- Generate a binary vector of length L using I^{ij}, and assign a state 0 or 1 to each I^{ij}.
- Use RBM with visible nodes corresponding to APs and RPs to learn the behaviour pattern.
- Detect anomaly by calculating the occurrence probability of the current observation based on trained RBM.

The anomaly detection process algorithm is described in Fig.3.

Algorithm 1: Anomaly Detection Algorithm

1 Collect normal and test data
2 Calculate the normalized pattern ($I^{\mathbb{AB}}(t), I^{\mathbb{AB}}_{norm}(t), I^{\mathbb{AB}}_{test}(t)$)
3 Select RBM structure, visible and hidden units (v, h)
4 Input $I^{\mathbb{AB}}_{norm}(t)$ into RBM
5 Obtain biases and weights for trained RBM
6 Compute the free energy $F(v)$of the normal data
7 Calculate RE and RE_d
8 Determine the threshold DT
9 Update inputs $I^{\mathbb{AB}}_{test}(t)$
10 Obtain free energy of updated $I^{\mathbb{AB}}_{test}(t)$
11 Compute RE and RE_d based on the current and previous inputs
12 Compare current RE_d with DT to detect anomaly

FIGURE 3. Proposed algorithm for anomaly detection.

IV. CASE STUDIES AND SIMULATION RESULTS

In this section a case studies under different operation condition are simulated to validate performance of the proposed

method. Case 1 is modeled as a multi-agent cyber-physical system based on IEEE-39 bus model where each agent includes a generator as described in Section II, a measurement device, a distributed control agent, and an energy storage system as shown in Fig.4. Energy storage represent the energy that can be fed into the system by different micro grid or renewable sources. The same analysis is performed for all case studies, however, for the sake of space only the results of Case 1 are included in this section.

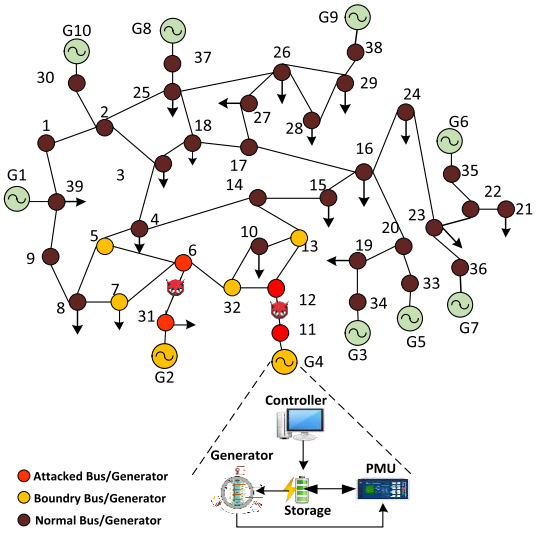


FIGURE 4. IEEE 39 bus system under cyber-attack in line 6-31 and 11-12.

A. TEST SYSTEM

Details of the case studies are listed in Table 2 and adapted from Matpower [42]. All case studies are assumed to be fully observable. To make sure about the accuracy of the historical data a level of security is added to the measurement model. Large-scale power grids contain thousands of meters which makes the protection of measurements highly expensive. In order to reduce the cost, we identify the critical meters to protect them based on optimal PMU placement [31]. We also assume that the system topologies remain unchanged over the typical days. Case studies are implemented in Matlab R2017a and carried out on a PC with a Core(TM) i7-7700 CPU, 3.6 GHz, and a RAM of 32.00 GB.

TABLE 2. Units for magnetic properties.

Case No.	No. of Buses	Num. of Generator	Num. of Lines	Num. of State	Num. of Measurements
1	39	10	46	40	171
2	118	54	186	216	609
3	2848	547	3776	2188	12673

By exploring the MI index, dependency between a subset of variables that influence each other in the normal condition is used for anomaly detection. The model generated by RBM represent the normal system since most of the collected data are collected are from the normal conditions. It should be mentioned that collected data are labeled as normal or anomalous. Training data are used to obtain the baseline for the normal condition which will be used for selecting the threshold for the anomaly. A moving window in a subset of the training data (with distribution P) is used to compute the distribution Q representing the dynamic behavior of the system. In order to measure the distance between Q and P, the RE metric is applied in each subset. Similar setting is used for the testing data. Finally, the two RE are compared to detect anomalous condition (cyber-attack in our case).

The attack strategy is designed to overload lines 6-31 and 11-12. The attack region is shown in Fig. 4. Normalized measurement residual under normal operation condition, due to fault, and due to cyber-attacks are presented in Fig.5 for Case 1. It can be seen that all the measurements residuals due to cyber-attacks have almost the same magnitude as the measurement residual under normal operation condition which implies that conventional residual test cannot detect the stealthy cyber-attacks. It should be noted that faults will results in significant residual in the measurement residual as shown in Fig. 5. In case of a fault in the system, the operator will be notified and clear the fault. Therefore, the fault will not affect the states of the system.

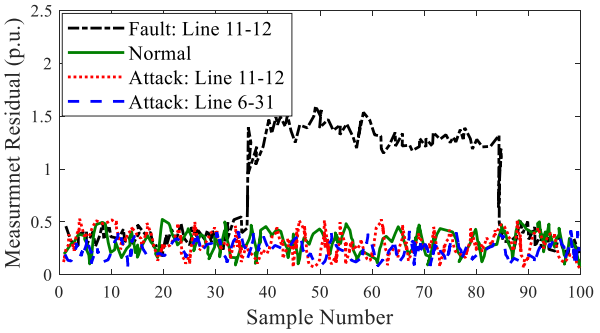


FIGURE 5. Measurement residual before and after cyber-attack on Case 1.

In Fig. 6, the variation in the lower plot is in an acceptable zone. However, in the top plot, the variation significantly increases during the attack between 35-65 samples. This indicates that there is a potential case of cyber-attack that has gone unnoticed in bad data detection. Therefore, estimated states with high error could be fed into the rest of the system, which may result in irreparable damages.

B. ACCURACY, FALSE POSITIVE AND TRUE POSITIVE

In the smart grid analysis, the major concern is not only the detection of cyber-attacks, but also the ability to avoid false alarms. Therefore, performance of the proposed method is analyzed based on the True Positives (TP), the True Negatives (TN), the False Positives (FP), and the False Negatives (FN), which are defined in Table 3.

The learning abilities and memorization properties of the algorithms are measured by the False Positive Rate (FPR), True Positive Rate (TPR), and Accuracy (Acc) values, which

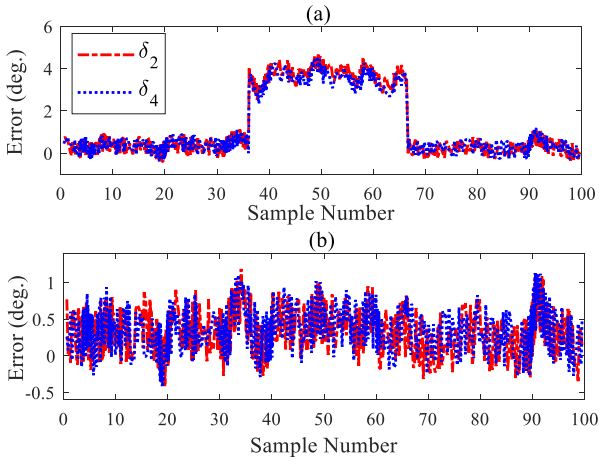


FIGURE 6. Variation on state estimation error, a) without cyber-attack, b) with cyber-attack on Case 1.

TABLE 3. Units for magnetic properties.

	Attacked	Secure
Classified as Attacked	TP	FP
Classified as Secure	FN	TN

are defined as [43]:

$$FPR = \frac{FP}{TN + FP}$$
$$TPR = \frac{TP}{TP + FN}$$
$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (21)$$

Low FPR of 0% means that none of the secure measurements are misclassified as attacked. TPR of 100% clarifies that none of the attacked measurements are misclassified as secure. Accuracy of 100% means that each measurement classified as attacked is an attacked measurement, and each measurement classified as secure is a secure measurement.

1) *Effect of Threshold on FPR*- Fig. 7 shows the variation of FPR as a function of detection threshold for single attack (SA)

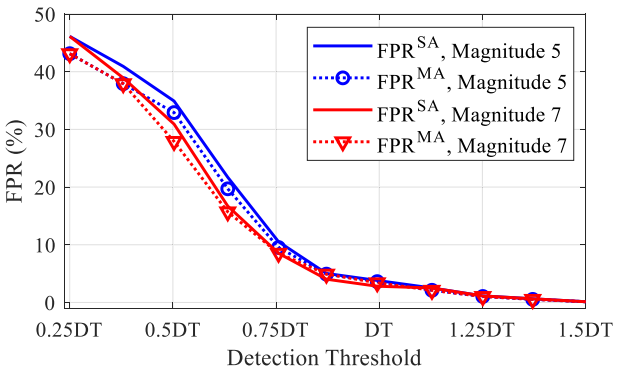


FIGURE 7. FPR under single and multiple cyber-attack for two different attack magnitudes on Case 1.

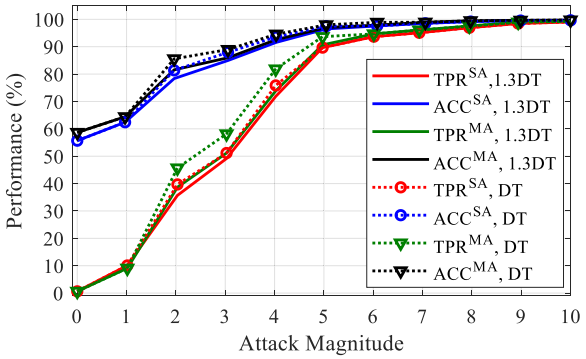


FIGURE 8. TPR and ACC under single and multiple cyber-attack for two different detection thresholds on Case 1.

and multiple attack (MA) on state variables δ_2, δ_4. For each case DT was varied from 0.25DT to 1.5 DT, where DT is the threshold defined in Section III. As can be seen from the figure, FPR decreases sharply with increase in detection threshold. This indicates that, when the threshold is too low, the algorithm becomes too aggressive in attack detection, thus suffering from high false-alarm rate.

In addition, as the figure shows, magnitude of the attack and number of attacks does not affect the FPR significantly. Moreover, it can be seen that for threshold larger than DT, FPR becomes negligible (i.e., under 2%). Therefore, DT is used as the threshold for the proposed method. Similar trend was observed in the trend of changes in FPR vs. the threshold for other states.

2) *Effect of Attack Magnitude on TPR and ACC*- Fig. 8 shows the variation of TPR and ACC as a function of attack magnitude for two attack scenarios on state variables δ_2, δ_4. 1 (1% of the original measurement) and 10 (10% of the original measurement) indicate low and high attack magnitudes, respectively. Medium magnitude (here indicated by 5) is the regular type of attack on the literature. To verify the effect of detection threshold on TPR and ACC, the results are plotted for two different thresholds.

As shown in Fig.8, by increasing the attack magnitude, TPR and ACC quickly approached 100%. In addition, it can be seen that a very high threshold adversely affects the TPR and impacts the minimum size detectable attack. The results show that DT defined in Section III can effectively detect an attack with medium and higher strength with almost 99% accuracy and 98% TPR. A similar trend was observed in the changes of TPR and ACC vs. the attack magnitude for all states. Summary of the results for different case studies are reported in Table 4.

3) *Effect of Attack Sparsity on TPR and ACC*- to analyze the effect of attack sparsity, attacks with different sparsity $\lambda/N \in [0, 1]$ are generated. N represents the total number of measurements in the system. As shown in Fig. 9, both TPR and ACC increase as the number of contaminated measurements increases. Here, sparsity 1 means all measurements are manipulated by the attacker. The figure shows that proposed

TABLE 4. Summary of results for single attack with medium attack magnitude (average).

Case	TPR^{SA}	TPR^{MA}	FPR^{SA}	FPR^{MA}	ACC^{SA}	ACC^{MA}
1	98.0%	98.1%	2.01%	1.98%	98.9%	99.1%
2	98.1%	98%	1.96%	1.96%	99.1%	99.0%
3	98.1%	98.1%	1.98%	1.97%	99.0%	99.1%

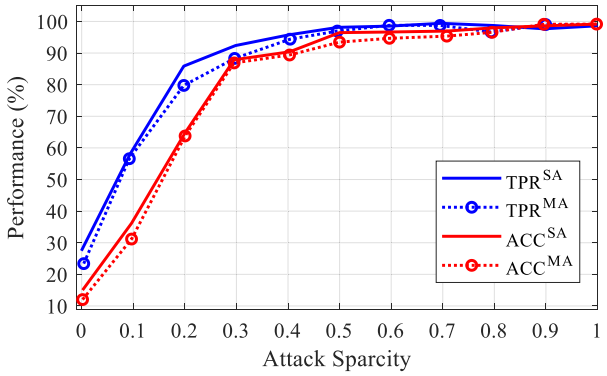


FIGURE 9. TPR and ACC under single and multiple cyber-attack for different attack sparsity on Case 1.

algorithm has very high TPR (94%) and ACC (90%) when only 35% of the measurements are manipulated. Once half of the measurements are attacked, which is a realistic assumption for successful attack implementation from the attacker's perspective, the algorithm is highly effective with 99% TPR and 98% ACC.

C. PERFORMANCE ANALYSIS UNDER DIFFERENT OPERATION CONDITION

To validate efficieny of the proposed method, four different scenarios are considered:1) normal condition without attack, 2) random attack, 3) single FDI attack on 6-31, 4) multiple, simultaneous FDI attacks on lines 6-31 and 11-12. Proposed method is compared with the two most popular BDD approaches; LNR test and Chi-Square test. The threshold is set to 3σ while σ is the standard deviation, to minimize the false positives due to the noise, thus FPR due to noise is less than 1% [44]. For accurate and detailed comparison, the threshold is normalized for all detectors. The same criterion is considered for setting threshold in LNR test. For more information about LNR and Chi-Square test refer to [20]. Detector's output are depicted in Fig. 10.

As shown in Fig. 10 (a), in normal operation condition, the output of all detectors is under the threshold which specifies that there is no trace of bad data or cyber-attack in the system. Fig. 10 (b) shows that all methods are able to detect the random attack. Since the attack is unintelligent, it will leave its trace in the data sets and the operator will be informed of an attack presence. The random bad data, which was injected to the measurement set, results in significant changes in the measurement residual vector, which leads to the increase in

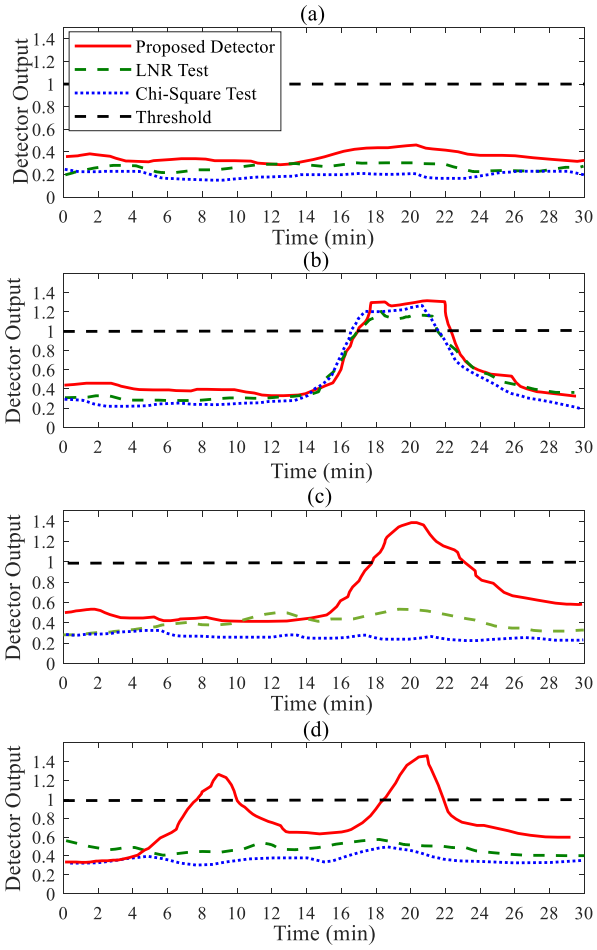


FIGURE 10. Detector out put under a) normal condition, b) random attack, c) single cyber-attack, d) multiple cyber-attack on Case 1.

cost function. In an optimal state estimation, we evaluate the cost function based on the residual of the measurements. In the normal operation condition, without bad data in the system, the cost function follows a normal distribution with zero mean. Under a random attack, the cost function will pass the threshold for optimal state estimation. Therefore, both LNR and chi-square tests will trigger the alarm successfully.

In case of single or multiple FDI attacks, as can be seen in Fig. 10 (c) and (d), the cost function for both LNR and Chi-Square detector stayed in the true range of predefined thresholds. Both approaches resulted in their normalized residue values below the specified threshold and thus they were unable to detect the attack in the system. However, in the same setup, output of the proposed detector is above the given threshold and can trigger the alarm. The main reason is that the LNR test and Chi-Square test are based on residual of the measurement vector while cyber-attacks are carefully crafted to bypass the statistical detector with no trace in residual vector. Similar results were observed for all case studies. Average detection time for all case studies was 1ms with 0.2ms deviations.

In general, any type of FDI attack in line or system topology results in the same changes in the network with minor modification. Therefore, the proposed method can successfully detect various FDI attacks from different sources. Furthermore, since the proposed scheme analyze the patterns between the compromised data and the normal data, its success rate does not depends on the attack scenarios.

V. CONCLUSION

In the context of smart grid anomaly detection, the solutions proposed in the literature are mainly offline approaches with restriction to deal with dynamically evolving cyber threats. This paper propose a real time and computationally efficient tool for anomaly detection that utilizing feature extraction scheme and time series partitioning to discover causal interactions between the subsystems. DBN concept and learning algorithms based on Boltzmann Machine are used to detect unobservable attacks based on free energy as the anomaly index. Performance of the proposed algorithm was evaluated on different IEEE test systems and under different operation conditions for several measures (TPR, FPR, and ACC). The results demonstrated that the system achieves an accuracy of 99%, TPR of 98% and FPR of less than 2%.

REFERENCES

[1] J. E. Dagle, "Postmortem analysis of power grid blackouts—The role of measurement systems," *IEEE Power Energy Mag.*, vol. 4, no. 5, pp. 30–35, Sep./Oct. 2006.

[2] Z. Huang, C. Wang, T. Zhu, and A. Nayak, "Cascading failures in smart grid: Joint effect of load propagation and interdependence," *IEEE Access*, vol. 3, pp. 2520–2530, 2015.

[3] Y. Cai, Y. Li, Y. Cao, W. Li, and X. Zeng, "Modeling and impact analysis of interdependent characteristics on cascading failures in smart grids," *Int. J. Elect. Power Energy Syst.*, vol. 89, pp. 106–114, Jul. 2017.

[4] H. Karimipour and V. Dinavahi, "On false data injection attack against dynamic state estimation on smart power grids," in *Proc. IEEE Int. Conf. Smart Energy Grid Eng. (SEGE)*, Aug. 2017, pp. 388–393.

[5] G. Dondossola, J. Szanto, M. Masera, and I. N. Fovino, "Effects of intentional threats to power substation control systems," *Int. J. Crit. Infrastruct.*, vol. 4, nos. 1–2, pp. 129–143, 2008.

[6] T. Morris, S. Pan, J. Lewis, J. Moorhead, N. Younan, R. King, M. Freund, and V. Madani, "Cybersecurity risk testing of substation phasor measurement units and phasor data concentrators," in *Proc. CSIIRW*, Oct. 2011, Art. no. 24.

[7] A. Ameli, A. Hooshyar, E. El-Saadany, and A. M. Youssef, "Attack detection and identification for automatic generation control systems," *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 4760–4774, Sep. 2018.

[8] H. H. Pajouh, R. Javidan, R. Khayami, D. Ali, and K.-K. R. Choo, "A two-layer dimension reduction and two-tier classification model for anomaly-based intrusion detection in IoT backbone networks," *IEEE Trans. Emerg. Topics Comput.*, to be published.

[9] R. Khan, K. Mclaughlin, D. Laverty, and S. Sezer, "Design and implementation of security gateway for synchrophasor based real-time control and monitoring in smart grid," *IEEE Access*, vol. 5, pp. 11626–11644, Jun. 2017.

[10] R. Xu, R. Wang, Z. Guan, L. Wu, J. Wu, and X. Du, "Achieving efficient detection against false data injection attacks in smart grid," *IEEE Access*, vol. 5, pp. 13787–13798, Jul. 2017.

[11] C.-C. Sun, A. Hahn, and C.-C. Liu, "Cyber security of a power grid: State-of-the-art," *Int. J. Elect. Power Energy Syst.*, vol. 99, pp. 45–56, Jul. 2018.

[12] X. Liu, N. Xiong, N. Zhang, A. Liu, H. Shen, and C. Huang, "A trust with abstract information verified routing scheme for cyber-physical network," *IEEE Access*, vol. 6, pp. 3882–3898, 2018.

[13] C. Alcaraz, C. Fernandez-Gago, and J. Lopez, "An early warning system based on reputation for energy control systems," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 827–834, Dec. 2011.

[14] X. He, L. Chu, R. C. Qiu, Q. Ai, and Z. Ling, "A novel data-driven situation awareness approach for future grids—Using large random matrices for big data modeling," *IEEE Access*, vol. 6, pp. 13855–13865, 2018.

[15] A. Farraj, E. Hammad, and D. Kundur, "A distributed control paradigm for smart grid to address attacks on data integrity and availability," *IEEE Trans. Signal Inf. Process. Over Netw.*, vol. 4, no. 1, pp. 70–81, Mar. 2018.

[16] I. Friedberg, X. Hong, K. Mclaughlin, P. Smith, and P. C. Miller, "Evidential network modeling for cyber-physical system state inference," *IEEE Access*, vol. 5, pp. 17149–17164, 2017.

[17] A. Azmoodeh, A. Dehghantanha, M. Conti, and K.-K. R. Choo, "Detecting crypto-ransomware in IoT networks based on energy consumption footprint," *J. Ambient Intell. Humanized Comput.*, vol. 9, no. 4, pp. 1141–1152, 2018.

[18] N. Milosevic, A. Dehghantanha, and K.-K. R. Choo, "Machine learning aided Android malware classification," *Comput. Elect. Eng.*, vol. 61, pp. 266–274, Jul. 2017.

[19] H. Karimipour and V. Dinavahi, "Robust massively parallel dynamic state estimation of power systems against cyber-attack," *IEEE Access*, vol. 6, pp. 2984–2995, Dec. 2017.

[20] M. Ozay, I. Esnaola, F. T. Y. Vural, S. R. Kulkarni, and H. V. Poor, "Machine learning methods for attack detection in the smart grid," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 8, pp. 1773–1786, Aug. 2016.

[21] S. Pan, T. Morris, and U. Adhikari, "Developing a hybrid intrusion detection system using data mining for power systems," *IEEE Trans. Smart Grid*, vol. 6, no. 6, pp. 3104–3113, Nov. 2015.

[22] J. Landford, R. Meier, R. Barella, X. Zhao, E. Cotilla-Sanchez, R. B. Bass, and S. Wallace, "Fast sequence component analysis for attack detection in synchrophasor networks," in *Proc. 5th Int. Conf. Smart Cities Green ICT Syst. (SmartGreens)*, Rome, Italy, 2016.

[23] S. Ahmed, Y. Lee, S.-H. Hyun, and I. Koo, "Feature selection–based detection of covert cyber deception assaults in smart grid communications networks using machine learning," *IEEE Access*, vol. 6, pp. 27518–27529, 2018.

[24] Y. He, G. J. Mendis, and J. Wei, "Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism," *IEEE Trans. Smart Grid.*, vol. 8, no. 5, pp. 2505–2516, Sep. 2017.

[25] D. Codetta-Raiteri and L. Portinale, "Dynamic Bayesian networks for fault detection, identification, and recovery in autonomous spacecraft," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 45, no. 1, pp. 13–24, Jan. 2015.

[26] S. Mohammadi, H. Mirvaziri, M. Ghazizadeh-Ahsaee, and H. Karimipour, "Cyber intrusion detection by combined feature selection algorithm," *J. Inf. Secur. Appl.*, vol. 44, pp. 80–88, Feb. 2019.

[27] C. A. Murthy, "Bridging feature selection and extraction: Compound feature generation," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 4, pp. 757–770, Apr. 2017.

[28] H. Karimipour and V. Dinavahi, "Extended Kalman filter-based parallel dynamic state estimation," *IEEE Trans. Smart Grid*, vol. 6, no. 3, pp. 1539–1549, May 2015.

[29] J. D. Glover, M. Sarma, and T. Overbye, *Power System Analysis and Design*, 5th ed. Boston, MA, USA: Cengage, 2011.

[30] A. R. Bergen and V. Vittal, *Power Systems Analysis*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2000.

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

[32] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Trans. Inf. Syst. Secur.*, vol. 14, no. 13, pp. 1–33, May 2011.

[33] M. Ozay, I. Esnaola, F. T. Yarman Vural, S. R. Kulkarni, and H. V. Poor, "Sparse attack construction and state estimation in the smart grid: Centralized and distributed models," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1306–1318, Jul. 2013.

[34] A. Ray, "Symbolic dynamic analysis of complex systems for anomaly detection," *Signal Process.*, vol. 84, no. 7, pp. 1115–1130, 2004.

[35] C. Rao, A. Ray, S. Sarkar, and M. Yasar, "Review and comparative evaluation of symbolic dynamic filtering for detection of anomaly patterns," *Signal, Image Video Process.*, vol. 3, no. 2, pp. 101–114, 2009.

[36] S. Sarkar, S. Sarkar, K. Mukherjee, A. Ray, and A. Srivastav, "Multi-sensor information fusion for fault detection in aircraft gas turbine engines," *J. Aerosp. Eng.*, vol. 227, no. 12, pp. 1988–2001, Dec. 2013.

[37] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.

[38] C. Liu, A. Akintayo, Z. Jiang, G. P. Henze, S. Sarkar, "Multivariate exploration of non-intrusive load monitoring via spatiotemporal pattern network," *Appl. Energy*, vol. 211, pp. 1106–1122, Feb. 2018.

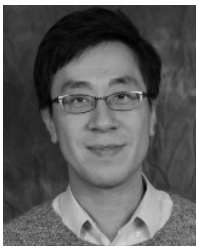
[39] C. Liu, S. Ghosal, Z. Jiang, and S. Sarkar, "An unsupervised anomaly detection approach using energy-based spatiotemporal graphical modeling," *Cyber-Phys. Syst.*, vol. 3, nos. 1–4, pp. 66–102, 2017.
[40] B. J. Frey and N. Jojic, "A comparison of algorithms for inference and learning in probabilistic graphical models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 9, pp. 1392–1416, Sep. 2005.
[41] D. J. C. MacKay, *Information Theory, Inference and Learning Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
[42] 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.
[43] W. Dixon and F. Massey, *Introduction to Statistical Analysis*, vol. 344. New York, NY, USA: McGraw-Hill, 1969.
[44] A. Abur and A. Gómez-Expósito, *Power System State Estimation: Theory and Implementation*. New York, NY, USA: Marcel Dekker, 2004.

HADIS KARIMIPOUR received the Ph.D. degree in energy system from the Department of Electrical and Computer Engineering, University of Alberta, in 2016. Before joining the University of Guelph, she was a Postdoctoral Fellow with the University of Calgary, working on cyber security of the smart power grids. She is currently an Assistant Professor with the School of Engineering, Engineering Systems and Computing Group, University of Guelph, Ontario. Her research interests include large-scale power system state estimation, cyber-physical modeling, cyber-security of the smart grids, and parallel and distributed computing. She serves as the Chair of the IEEE Women in Engineering (WIE) and chapter Chair of the IEEE Information Theory in Kitchener-Waterloo section.

ALI DEHGHANTANHA received the Ph.D. degree in security in computing, and has a number of professional certifications including CISSP and CISM. He is the Director of Cyber Science Lab with the University of Guelph, Ontario, Canada. His lab is focused on building AI-powered solutions to support cyber threat attribution, cyber threat hunting, and digital forensics tasks in the Internet of Things (IoT), industrial IoT, and Internet of Military of Things (IoMT) environments. He has served for more than a decade in a variety of industrial and academic positions with leading players in cyber-security and artificial intelligence. Prior to joining the University of Guelph, he has served as a Senior Lecturer with the University of Sheffield, U.K., and as an EU Marie-Curie International Incoming Fellow with the University of Salford, U.K.

REZA M. PARIZI received the B.Sc. and M.Sc. degrees in computer science in 2005 and 2008, respectively, and the Ph.D. degree in software engineering in 2012. He is a Faculty with the College of Computing and Software Engineering, Kennesaw State University (KSU), GA, USA. He is a Consummate Technologist and Cybersecurity Researcher with an entrepreneurial spirit. He is the member of the IEEE Blockchain Community, the IEEE Computer Society, and ACM. Prior to joining KSU, he was an Associate Professor with New York Institute of Technology. His research interests include R&D in decentralized computing, the IoT, and emerging issues in the practice of secure software-run world applications.

KIM-KWANG RAYMOND CHOO received the Ph.D. degree in information security from Queensland University of Technology, Australia, in 2006. He currently holds the Cloud Technology Endowed Professorship at The University of Texas at San Antonio (UTSA). In 2016, he was named the Cybersecurity Educator of the Year –APAC (Cybersecurity Excellence Awards are produced in cooperation with the Information Security Community on LinkedIn), and in 2015, he and his team won the Digital Forensics Research Challenge organized by Germany's University of Erlangen-Nuremberg. He is the recipient of the 2018 UTSA College of Business Col. Jean Piccione and Lt. Col. Philip Piccione Endowed Research Award for Tenured Faculty, Outstanding Associate Editor of 2018 for IEEE Access, British Computer Society's 2019 Wilkes Award Runner-up for his paper published in the 2018 volume of *The Computer* Journal (Oxford University Press, 2019), the *EURASIP Journal on Wireless Communications and Networking* (JWCN) Best Paper Award, the IEEE TrustCom 2018 Best Paper Award, ESORICS 2015 Best Research Paper Award, 2014 Highly Commended Award by the Australia New Zealand Policing Advisory Agency, Fulbright Scholarship in 2009, 2008 Australia Day Achievement Medallion, and British Computer Society's Wilkes Award in 2008. He is also a Fellow of the Australian Computer Society and Co-Chair of the IEEE Multimedia Communications Technical Committee's Digital Rights Management for Multimedia Interest Group.



HENRY LEUNG is a Professor with the Department of Electrical and Computer Engineering, University of Calgary. Before joining this university, he was with the Department of National Defense (DND) of Canada as a Defense Scientist. He conducted research and development of automated surveillance systems, which can perform detection, tracking, identification, and data fusion automatically as a decision aid for military operators. His current research interests include big data analytic, chaos and nonlinear dynamics, information fusion, machine learning, signal and image processing, robotics, and the Internet of Things. He has published extensively in the open literature on these topics. He has over 200 journal papers and 200 refereed conference papers. He has been the Associate Editor of various journals such as the IEEE Circuits and Systems Magazine, *International Journal on Information Fusion*, the IEEE Signal Processing Letters, and the IEEE Transactions on Circuits and Systems. He has also served as Guest Editors for the special issue Intelligent Transportation Systems for the International Journal on Information Fusion and Cognitive Sensor Networks for the *IEEE Sensors* Journal. He is the Topic Editor on Robotic Sensors of the *International Journal of Advanced Robotic Systems*. He is the Editor of the Springer book series on Information Fusion and Data Science.