Real-Time Detection of False Data Injection Attacks in Smart Grid: A Deep Learning-Based Intelligent Mechanism

Youbiao He, *Student Member, IEEE*, Gihan J. Mendis, *Student Member, IEEE*, and Jin Wei, *Member, IEEE*

***Abstract*—Application of computing and communications intelligence effectively improves the quality of monitoring and control of smart grids. However, the dependence on information technology also increases vulnerability to malicious attacks. False data injection (FDI), that attack on the integrity of data, is emerging as a severe threat to the supervisory control and data acquisition system. In this paper, we exploit deep learning techniques to recognize the behavior features of FDI attacks with the historical measurement data and employ the captured features to detect the FDI attacks in real-time. By doing so, our proposed detection mechanism effectively relaxes the assumptions on the potential attack scenarios and achieves high accuracy. Furthermore, we propose an optimization model to characterize the behavior of one type of FDI attack that compromises the limited number of state measurements of the power system for electricity theft. We illustrate the performance of the proposed strategy through the simulation by using IEEE 118-bus test system. We also evaluate the scalability of our proposed detection mechanism by using IEEE 300-bus test system.**

***Index Terms*—False data injection (FDI) attacks, deep learning, state vector estimator (SVE), deep learning based identification (DLBI) scheme, supervisory control and data acquisition (SCADA).**

I. Introduction

THE SMART grid, as a cyber-physical critical infrastructure, boasts higher reliability, efficiency, and consumer-centricity in an environment of increasing power demand. The inter-communicating devices, such as Phasor Measurement Units (PMUs) and smart meters have opened exciting opportunities, in which the acquisition, transmission, and consumption of high-granularity real-time power system data are facilitated through the integration of communications, computing, and advanced control technologies. Such dependence on information technology naturally raises questions as to the consequences of cyber attacks on power system operation. To address this issue, this paper explores one typical data integrity attack, False Data Injection (FDI) attack, that is approved to be an emerging severe threat to the Supervisory Control and Data Acquisition (SCADA) system [1]. Furthermore, we consider that the attack is launched for stealing electricity. As stated in [2]–[4], electricity theft has become a serious concern for utility companies in recent years. Recently, the real-time state measurement data from SCADA system have been leveraged to detect the electricity theft [5]. In our paper, we consider the FDI attack model targeting at increasing the success rate of electricity theft by compromising the real-time state measurements such as the real-time load profiles. We propose a deep-learning-based intelligent mechanism to detect this type of FDI attack in real-time. We would like to clarify that, although we consider one specific FDI attack, our proposed mechanism can be generalized to detect different FDI attacks.

Various research has been developed on proposing various FDI attack scenarios and developing the corresponding detection strategies. Yang *et al.* [6] formulated the least-effort attack strategy and proposed a protection-based defense scheme and a detection-based defense scheme. Ozay *et al.* [7] modelled the attack strategy using Gaussian process and used machine learning methods for attack detection. Liu *et al.* [8] developed an FDI detection mechanism by using the properties of the low dimensionality of measurements and sparsity of attacks. In [9], the equivalent measurement transformation and the largest weighted residual method were integrated for detecting the FDI attacks. In the research field of electricity theft, some research has been established model the behavior of electricity thieves and utilities by using game-theoretic methods. Cárdenas *et al.* [2] proposed a game theory model in which the thieves target at minimizing their probability to be detected with a fixed amount of energy to steal while utilities aim at achieving the tradeoff between the cost of theft detection mechanism and loss energy. There are also some works developed to detect the electricity theft by analyzing the readings of smart meters [10]–[13]. As far as we know, we are the first to model the FDI attacks for electricity theft and propose a real-time detection mechanism.

In our proposed real-time detection mechanism, we apply deep learning techniques to recognize the behavior patterns of FDI attacks using the historical measurement data and employ the revealed features to detect the FDI attacks in real-time. Deep learning techniques were recently proposed to capture the higher-order statistical structure of the complex data by arranging the feature detectors in layers [14]. Deep Belief Network (DBN) is one of the fundamental and

Manuscript received August 31, 2016; revised December 22, 2016 and April 4, 2017; accepted May 3, 2017. Date of publication May 11, 2017; date of current version August 21, 2017. Paper no. TSG-01187-2016. *(Corresponding author: Jin Wei.)*

The authors are with the Department of Electrical and Computer Engineering, University of Akron, Akron, OH 44325 USA (e-mail: jinwei06@gmail.com).

Digital Object Identifier 10.1109/TSG.2017.2703842

1949-3053 © 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

Authorized licensed use limited to: SICHUAN UNIVERSITY. Downloaded on February 09,2025 at 08:24:48 UTC from IEEE Xplore. Restrictions apply.

widely used deep learning techniques that is constructed with a stack of Restricted Boltzmann Machines (RBMs) [15]–[17]. In our mechanism, we propose an extended DBN architecture, called Conditional Deep Belief Network (CDBN), that exploits Conditional Gaussian-Bernoulli RBM (CGBRBM) to extract high-dimensional temporal features [18]–[20]. Different from the CDBN proposed in [18] and [19] that was designed for modeling human motion, our CDBN architecture is designed for analyzing the temporal attack patterns that are presented by the real-time measurement data from the geographically distributed sensors/meters [20]. To be clear, we would like to emphasize that there are two main differences between our proposed CDBN and the one in [18] and [19]. Firstly, our proposed CDBN is designed as a classifier while the one in [18] and [19] is a time-series generative model to predict the future motion. Secondly, our CDBN architecture only employs CGBRBM technique for the first hidden layer and uses the conventional RBM technique for all the other hidden layers. By doing so, we effectively reduce the complexity of training and execution time of the CDBN architecture.

Furthermore, because the essential mechanism of our proposed detection scheme is based on recognizing the difference in the patterns between the compromised data with FDI attacks and the normal data without any cyber attacks, the success of our proposed detection scheme is dependent on the sensitivity of the pattern recognition rather than the assumption of the scenarios of the FDI schemes. Our proposed electricity theft model that is presented in Section III is developed to characterize the behaviors of one type of FDI attacks. However, the application of our proposed detection scheme is not limited to detecting this type of FDI attacks.

Next section illustrates the problem settings for our work. In Section III, we describe our proposed real-time detection mechanism for FDI attacks Mechanism. We also propose an optimization model to characterize one type of FDI attack for electricity theft. Section IV introduces the deep-learning based scheme that is the central part of our real-time detection mechanism. Simulation results and conclusions are presented in Sections V and VI, respectively.

II. Problem Settings

System state estimation is the crucial mechanism for maintaining the stability and efficiency of the modern power grids [21]. As shown in Fig. 1, the measurement data, such as bus voltage, bus power flow, branch power flow, and load profiles, from the sensors or meters, are typically sent to a control center using an industrial control system known as SCADA system that analyzes the received measurement data, estimates the states of the power system, detects the potentials of contingency, and sends the corresponding control signals to the Remote Terminal Units (RTUs) to ensure the reliable operation.

A. DC State Estimation and False Data Injection Attack

The state estimator in the control center uses the measurement data reported from the SCADA system and the steady system model to estimate the system state over time.

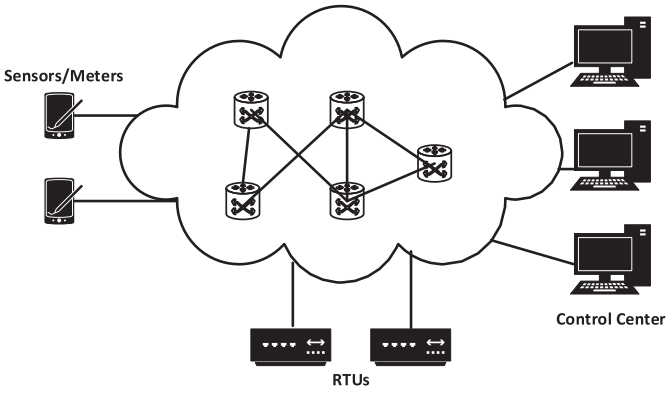

Fig. 1. One example structure of communication and control system of the model power grids.

Let $\mathbf{z} = [z_1, z_2, \ldots, z_m]^T \in \mathbb{R}^m$ be measurement vector, $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T \in \mathbb{R}^n$ be state vector, and $\mathbf{e} = [e_1, e_2, \ldots, e_m]^T \in \mathbb{R}^m$ denote the measurement error vector. We can describe the observation model as follows:

$$\mathbf{z} = \mathbf{h}(\mathbf{x}) + \mathbf{e}, \tag{1}$$

The observation model can be further described by Eq. (2) in the DC power flow model:

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

where $\mathbf{H} \in \mathbb{R}^{m \times n}$ is the Jacobian matrix that represents the topology of the power system, and $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2)$ denotes the environment noise modelled by Additive White Gaussian Noise (AWGN) with standard deviation σ.

By applying the statistical estimation criteria, such as Minimum Mean-Square Error (MMSE), the estimated system state $\mathbf{x}$ can be formulated by the following equation:

$$\hat{\mathbf{x}} = \left(\mathbf{H}^\mathbf{T} \Lambda \mathbf{H}\right)^{-1} \mathbf{H}^\mathbf{T} \Lambda \mathbf{z}, \tag{3}$$

where Λ is a diagonal matrix whose diagonal elements are $\Lambda_{ii} = \sigma^{-2}$.

In our work, the attackers have the knowledge of the system topology that is represented by the Jacobian matrix $\mathbf{H}$ and possess the capability of compromising a limited number of state estimation data such as load profiles to launch the FDI attacks for electricity theft. Under these assumptions, the observation model in the presence of the FDI attacks can be described in the following:

$$\hat{\mathbf{z}_a} = \mathbf{H}\mathbf{x} + \mathbf{a} + \mathbf{e}, \tag{4}$$

where $\mathbf{a}$ is the attack vector and $\mathbf{a} \neq \mathbf{0}$ when FDI attacks occur.

III. Real-Time Mechanism for Detecting False Data Injection Attacks

A. Overview of the Proposed Detection Mechanism

Figure 2 provides an overview of our proposed real-time mechanism for detecting FDI attacks. Our proposed detection mechanism mainly consists of a State Vector Estimator (SVE) and a Deep-Learning Based Identification (DLBI) scheme. As

Authorized licensed use limited to: SICHUAN UNIVERSITY. Downloaded on February 09,2025 at 08:24:48 UTC from IEEE Xplore. Restrictions apply.

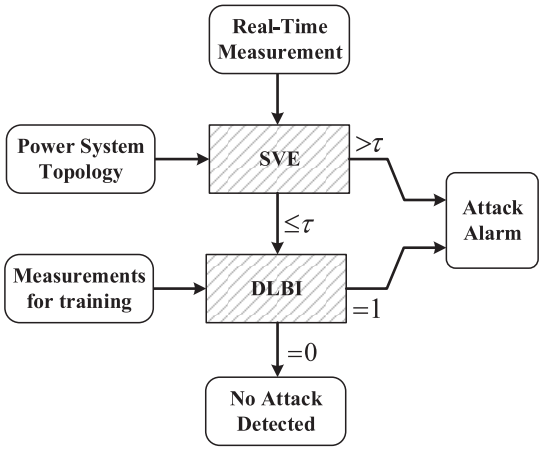

Fig. 2. Our proposed deep learning based real-time mechanism for detecting FDI attacks.

described in Eq. (5), SVE evaluates the quality of the real-time measurement data by calculating the ℓ_2-norm of measurement residual and comparing the calculation result, η, with a predetermined threshold τ. SVE reports attack alarm if $\eta > \tau$ and the measurement is considered as compromised data.

$$\begin{cases} \eta = \left\|\hat{\mathbf{z}} - \mathbf{H}\hat{\mathbf{x}}\right\|_2 > \tau, & \text{Attack alarm is reported;} \\ \eta = \left\|\hat{\mathbf{z}} - \mathbf{H}\hat{\mathbf{x}}\right\|_2 \leq \tau, & \text{No attack alarm is reported.} \end{cases} \tag{5}$$

where $\|\cdot\|_2$ denotes ℓ_2-norm operation. The measurement data with $\eta \leq \tau$ will be passed to DLBI scheme for further evaluation. In this paper, the selection of the value of the threshold τ is numerically studied as shown in the simulation section. Theoretically, the value of τ should be in an appropriate range. If τ is too small, the robustness of the SVE system to the environment noise will be reduced, potentially resulting in high positive false alarms on the FDI detection. Additionally, if τ is too large, the effectiveness of the SVE system can be compromised, which may increase the processing load of the subsequent DLBI scheme.

We would like to clarify that our deep learning based real-time mechanism is designed under the assumption that the physical topology of the power system does not change dramatically within a short time [22]. Practically, the supervisory control and data acquisition (SCADA) system can effectively detect this situation [23], [24]. Our work focus on addressing the cyber attacks that are launched by the attackers who can access to and change the measurement data without causing the severe change in the physical status of the system. Additionally, our work also applies to the power system with renewable energy, such as Photovoltaics (PV) and wind turbine, considering the extensive research and industrial work established to mitigate the uncertainty of the renewable energy generation [25]–[29].

Based on Eq. (5), the corrupted observation $\hat{\mathbf{z}}_a$ that cannot be detected by SVE can be characterized as follows:

$$\begin{aligned} \left\|\hat{\mathbf{z}}_a - \mathbf{H}\hat{\mathbf{x}}_{\mathbf{bad}}\right\|_2 &= \left\|\mathbf{z} + \mathbf{a} - \mathbf{H}\left(\hat{\mathbf{x}} + \mathbf{c}\right)\right\|_2 \\ &= \left\|\mathbf{z} - \mathbf{H}\hat{\mathbf{x}} + (\mathbf{a} - \mathbf{H}\mathbf{c})\right\|_2 \\ &\leq \left\|\mathbf{z} - \mathbf{H}\hat{\mathbf{x}}\right\|_2 + \|(\mathbf{a} - \mathbf{H}\mathbf{c})\|_2 \leq \tau \end{aligned}$$

Let $\tau_a = \tau - \|\mathbf{z} - \mathbf{H}\hat{\mathbf{x}}\|_2$. We have that if $\|\mathbf{a} - \mathbf{H}\mathbf{c}\| \leq \tau_a$, the FDI attacks can bypass the SVE mechanism. Therefore, we can describe the sufficient condition in which the FDI attack is able to pass the SVE system as follows [1]:

$$\mathbf{a} = \mathbf{H}\mathbf{c} + \mathbf{t}, \tag{6}$$

where $\mathbf{H}$ is the Jacobian matrix available to the attackers, $\mathbf{c}$ and $\mathbf{t}$ are vectors designed by the attackers, and $\|\mathbf{t}\|_2 \leq \tau_a$. We call the FDI attacks that can be detected by SVE as observable FDI attacks and the ones that cannot be detected by SVE as unobservable FDI attacks. Therefore, in our proposed detection mechanism, the function of DLBI scheme is to detect the unobservable FDI attacks.

Our DLBI scheme employs the CDBN to recognize the high-dimensional temporal features of the FDI attacks. To achieve this goal, our DLBI scheme consists of two essential mechanisms: (1) updating the CDBN architecture via the training procedure with certain predetermined rate and (2) detecting the potential FDI attacks by using the currently updated CDBN in real-time. The authors would like to clarify that these two essential mechanisms are implemented in parallel. Therefore, the computing time of the training procedure does not lead to any delay in the detection procedure. Additionally, the CDBN architecture is trained using the historical measurements that were collected from the previous experiences and processed through SVE. The ones that can activate the attack alarm of SVE were excluded from the collection. Furthermore, if the historical measurement was collected in the situation where the existence of FDI attacks was verified, it is assigned with label $l = 1$. If the measurement was collected in the situation where a thorough security check was deployed and no FDI attacks were detected, it is assigned with label $l = 0$. Otherwise, this measurement vector is kept unlabeled. Furthermore, considering that the labeled compromised data are very limited in the real application, we extend this type of data by the following two steps: (1) artificially generating the compromised data that have the similar patterns as the ones collected from the real world, and (2) generating the compromised data based on our proposed FDI attack model for electricity theft that is detailed in Section III-B. The authors would like to mention that by including the labeled compromised data generated by using our proposed electricity theft model, the capability of DLBI scheme can be effectively extended to detect the FDI attacks whose pattern have not been appeared in the past. The details of our DLBI scheme are introduced in Section IV.

B. FDI Attack Model for Electricity Theft

In our work, we model one type of FDI attack that is launched for electricity theft by specifying the following threat model for the attackers: 1) the attackers have knowledge of the power system topology, 2) the attackers are capable of corrupting a limited number of state measurement data which are the load profiles in our work, 3) the attackers have the basic understanding of the SVE mechanism described in Eq. (5) without knowing the threshold τ, and 4) the attackers aim to maximize the benefit of stealing electricity while minimizing the risk of being detected when manipulating the measurement data. We

Authorized licensed use limited to: SICHUAN UNIVERSITY. Downloaded on February 09,2025 at 08:24:48 UTC from IEEE Xplore. Restrictions apply.

would like to claim that the third assumption in the threat model is reasonable since the function of SVE is fundamental for detecting FDI attacks in general.

We define $\mathcal{B}$ as the collection of the indices of all the state measurements of the system and $\mathcal{A}$ as the set of the indices of the load profiles that are associated with the attackers, and thus $\mathcal{A} \subseteq \mathcal{B}$. Letting $\mathcal{C}$ be the cardinality of the set $\mathcal{B}$, we define $\mathbf{n}$ as a $\mathcal{C}$-dimension vector whose elements are as follows:

$$\begin{cases} n_i = 1, & \text{if } i \in \mathcal{A}; \\ n_i = 0, & \text{if } i \notin \mathcal{A} \text{ and } i \in \mathcal{B}. \end{cases} \tag{7}$$

Then we define the objective function of the attackers as follows:

$$U = p\mathbf{n}^T\mathbf{a} + q\frac{2}{\exp\left(\mathbf{r}^T\mathbf{D}\mathbf{r}\right)/\lambda + 1}, \tag{8}$$

where $\mathbf{a}$ is the attack vector at a certain time step, p and q are the preference parameters, and λ is the scale parameter. $\mathbf{n}^T\mathbf{a}$ measures the total difference between the actual values and the compromised measurements of the load profiles associated with the attacks and thus models the benefit of launching the FDI attacks for power theft. $\mathbf{D}$ is a $\mathcal{C} \times \mathcal{C}$ diagonal matrix in which the inverse of ith diagonal element D_{ii} indicates the vulnerability of the ith state measurements. Furthermore, $\mathbf{r}$ is the ratio between the injected bias $\mathbf{a}$ and the accurate values of the state measurements $\mathbf{L}$, which is calculated as follows:

$$\mathbf{r} = \mathbf{a}./\mathbf{L}, \tag{9}$$

where the operator ./ denotes the element division of the two vectors with the same dimension. Therefore, the second term in Eq. (8) evaluates the probability of not being detected when manipulating the measurement data.

By using Eq. (9), we can characterize the FDI attacks by using the following optimization models:

$$\begin{aligned} \underset{\mathbf{a}}{\text{maximize}} \quad & U \\ \text{subject to} \quad & \|\mathbf{a} - \mathbf{H}\mathbf{c}\| \le \hat{\tau}_a, \\ & \mathbf{N}(\mathbf{a} + \mathbf{L}) \preceq 0, \\ & \mathbf{a}^T\mathbf{M}\mathbf{a} = 0, \end{aligned} \tag{10}$$

where $\hat{\tau}_a$ is the threshold for the SVE mechanism estimated by the attackers,

$$\begin{cases} \mathbf{M} = diag\{\mathbf{1} - \mathbf{n}\}, \\ \mathbf{N} = diag\{\mathbf{n}\}, \end{cases} \tag{11}$$

and $diag\{\cdot\}$ is an operator that transform a vector to an associated diagonal matrix described as follows:

$$diag\left\{\begin{pmatrix} a_{1,1} \\ a_{2,2} \\ \vdots \\ a_{n,n} \end{pmatrix}\right\} = \begin{pmatrix} a_{1,1} & 0 & \dots & 0 \\ 0 & a_{2,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{n,n} \end{pmatrix}$$

Moreover, the third constraint in the optimization model is defined to model the fact that the sign of the power measurements has to be the same after FDI attacks occur. This is because that whether the power measurements evaluate power generation (positive sign) or load consumption (negative sign) are predetermined by the system topology and the FDI attacks causing the changes in the signs can be easily detected when manipulating the data. Therefore, it is reasonable to assume that attackers can realize this fact and avoid this situation by leveraging the topology information.

C. Solving the Proposed FDI Attack Model

To solve our proposed optimization model described in Eq. (10), we exploit the Sequential Quadratic Programming (SQP) algorithm [30]. To implement SQP algorithm, we transform the original problem to the QP subproblem and achieve the solution via numerical iteration. To construct the QP subproblem, we first define the Lagrangian function as follows:

$$\mathcal{L}(\mathbf{a}, \lambda_1, \lambda_2, \lambda_3) = U(\mathbf{a}) + \lambda_1^T h_1(\mathbf{a}) + \lambda_2^T h_2(\mathbf{a}) + \lambda_3^T g(\mathbf{a}), \tag{12}$$

where

$$\begin{cases} h_1(\mathbf{a}) = \|\mathbf{a} - \mathbf{H}\mathbf{c}\| - \hat{\tau}_a \\ h_2(\mathbf{a}) = \mathbf{N}(\mathbf{a} + \mathbf{L}) \\ g(\mathbf{a}) = \mathbf{a}^T\mathbf{M}\mathbf{a} \end{cases}$$

Using Eq. (12), the QP subproblem can be modelled in the following:

$$\begin{aligned} \max_{\mathbf{d}_k} \quad & \nabla U(\mathbf{a}_k)^T\mathbf{d}_k + \frac{1}{2}\mathbf{d}_k^T\mathbf{H}_k\mathbf{d}_k \\ \text{subject to} \quad & h_1(\mathbf{a}_k) + \nabla h_1(\mathbf{a}_k)^T\mathbf{d}_k \le 0, \\ & h_2(\mathbf{a}_k) + \nabla h_2(\mathbf{a}_k)^T\mathbf{d}_k \le 0, \\ & g(\mathbf{a}_k) + \nabla g(\mathbf{a}_k)^T\mathbf{d}_k = 0. \end{aligned} \tag{13}$$

where $\mathbf{H}_k$ denotes the Hessian of the Lagrangian as shown in Eq. (14).

$$\mathbf{H}_k = \nabla^2_{\mathbf{aa}}\mathcal{L}(\mathbf{a}_k, \lambda_k), \tag{14}$$

where $\lambda_k = \{\lambda_1^k, \lambda_2^k, \lambda_3^k\}$ where $\lambda_1^k, \lambda_2^k, \lambda_3^k$ are the values of $\lambda_1, \lambda_2, \lambda_3$ at the kth iteration, respectively. By solving the subproblem, we can get the value of $\mathbf{d}_k$ which is used for the iterative update. Letting k denote the index of the current iteration, we can get the new $\mathbf{a}_{k+1}$ by adding $\mathbf{d}$ calculated during the kth iteration as follows:

$$\mathbf{a}_{k+1} = \mathbf{a}_k + \alpha_k\mathbf{d}_\mathbf{k}$$

where α_k is the step length which is determined to make sure that merit function can obtain a sufficient decrease, which can be described using the following equation:

$$\phi(\mathbf{a}_k + \alpha_k\mathbf{d}_k) \le \phi(\mathbf{a}_k) + \eta\alpha_k D_{\mathbf{d}_k}\phi(\mathbf{a}_k) \tag{15}$$

where η is a chosen parameter, $D_{\mathbf{d}_k}\phi(\mathbf{a}_k)$ denotes the directional derivative of $\phi(\mathbf{a}_k)$ in the direction $\mathbf{d}_k$ and $\phi(\mathbf{a}_k)$ is merit function used to control the size of the steps, which is defined as follows:

$$\phi(\mathbf{a}_k) = U(\mathbf{a}_k) + r_1 h_1(\mathbf{a}_k) + r_2 h_2(\mathbf{a}_k) + \max(0, r_3 h_3(\mathbf{a}_k)), \tag{16}$$

where r_1, r_2, and r_3 are penalty parameter set as follows:

$$r_i = \max_i\left\{\lambda_i, \frac{\lambda_i + (r_k)_i}{2}\right\}, \ i = 1, 2, 3$$

To sum up, we present the algorithm for solving our nonlinear programming problem with equality and inequality

Authorized licensed use limited to: SICHUAN UNIVERSITY. Downloaded on February 09,2025 at 08:24:48 UTC from IEEE Xplore. Restrictions apply.

TABLE I
ALGORITHM OF SQP FOR NONLINEAR CONSTRAINED OPTIMIZATION

Start:
Step 1: Initialize the pair (a_0, λ_0) and calculate $n \times n$ symmetric positive definite Hessian approximation H_0, accordingly; Select the parameters $\eta \in (0, 0.5)$, $\tau \in (0, 1)$;
Step 2: Check the stop condition:
If it is satisfied, stop the algorithm,
otherwise, go to Step 3 with $k = 0, 1, 2, ...$
Step 3: Compute $\mathbf{d}_k$ by solving the subproblem defined in Eqn. (13)
Step 4: Choose r_k such that $\mathbf{d}_k$ is a descent direction for the merit function $\phi(\mathbf{a})$ defined in Eq. (16) by initializing $\alpha_k = 1$;
Step 5: Check the condition defined in Eqn. (15):
If it is satisfied, go to Step 6,
otherwise, update $\alpha_k = \tau_\alpha \alpha_k$, where $\tau_\alpha \in (0, \tau)$ is a predetermined parameter.
Step 6: Calculate $\mathbf{a}_{k+1} = \mathbf{a}_k + \alpha_k \mathbf{d}_k$
Step 7: Compute λ_{k+1} by solving
$\lambda_{k+1} = -[\mathbf{A}_{k+1}\mathbf{A}_{k+1}^T]^{-1}\mathbf{A}_{k+1}\nabla f_{k+1}$
Step 8: Compute $\mathbf{H}_{k+1}$ by updating $\mathbf{H}_k$ using Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [31], [32].
End

constraints in Table I [31], [32]. In the table, $\mathbf{A}$ is defined as follows:

$$\mathbf{A}(\mathbf{a}_k) = \left[\nabla h_1(\mathbf{a}_k), \nabla h_2(\mathbf{a}_k), \nabla g(\mathbf{a}_k)\right]. \tag{17}$$

IV. DEEP LEARNING-BASED IDENTIFICATION SCHEME

As stated in Section III-A, in our detection method, a DLBI scheme is developed to detect the compromised data that bypass the SVE mechanism. In our DLBI scheme, we propose a CDBN by integrating the standard deep belief network structure with CGBRBM that is capable of addressing real-valued input and modeling the impact of the historical observations on the current behavior feature extractions. Using the proposed CDBN, our identification scheme is able to detect the unobservable FDI attacks in real-time by learning the temporal behavior features of the FDI attacks. Fig. 3 illustrates the structure of our proposed CDBN architecture. As shown in Fig. 3, the CDBN employs the CGBRBM technique for the first hidden layer and stacks $K-1$ conventional RBMs on the top of the CGBRBM, where K is the number of hidden layers of our CDBN architecture. To accomplish the detection functionality, our CDBN is designed as a binary classifier by using a single output unit to indicate whether the evaluated measurements are compromised by FDI attacks.

A. Training Procedure for CDBN

Conventional RBM: Energy function of a conventional RBM is as follows:

$$E(\mathbf{v}, \mathbf{h}) = -\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}h_i v_j - \sum_{j=1}^{m} c_j v_j - \sum_{i=1}^{n} d_i h_i \tag{18}$$

where v_j is the jth element of the visible vector, h_i is the ith element of the vector consisting of the hidden units, w_{ij} is the ijth element of the weight matrix between the visible and hidden units, d_i and c_j denote the ith and jth elements of the bias vectors for the hidden layer and visible layer, respectively, n represents the number of the hidden units, and m is the number of the visible units.

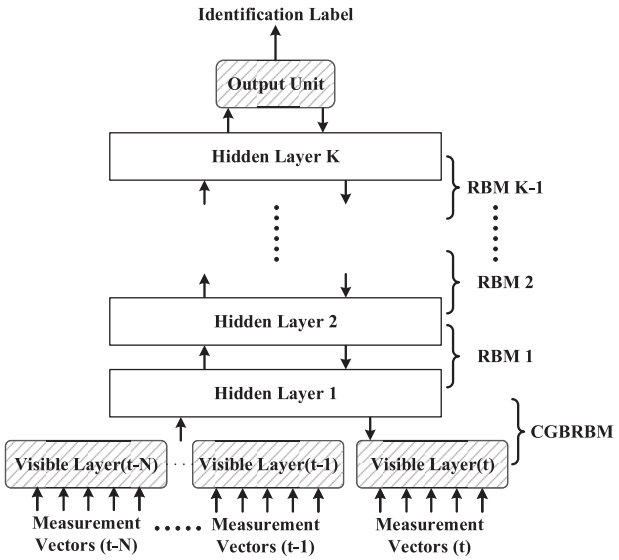

Fig. 3. The structure of our CDBN architecture designed in the DLBI scheme: K is the number of the hidden layers of the CDBN, t indicates the current time index, and N represents the size of the previous-time observation window.

Using Eq. (18), the activation conditional probability distributions of hidden and visible units, given the adjacent layer unit values, can be calculated by using the sigmoid activation function, which is shown in the following:

$$\begin{cases} p(h_i = 1|\mathbf{v}) = sigm\Big(d_i + \sum_{j=1}^{m} w_{ij}v_j\Big), \\ p\big(v_j = 1|\mathbf{h}\big) = sigm\big(c_j + \sum_{i=1}^{n} w_{ij}h_i\big). \end{cases} \tag{19}$$

where $sigm(x) = 1/(1+e^{-x})$ is sigmoid function. By using the gradient-based Contrastive Divergence (CD) technique [33], the weights and biases of the conventional RBMs are updated as follows:

$$\begin{cases} w_{ij} = w_{ij} - \alpha\big(\langle v_j h_i\rangle_m - \langle v_j h_i\rangle_l\big), \\ d_i = d_i - \alpha(\langle h_i\rangle_m - \langle h_i\rangle_l), \\ c_j = c_j - \alpha\big(\langle v_j\rangle_m - \langle v_j\rangle_l\big), \end{cases} \tag{20}$$

where α denotes the learning rate, and $\langle\cdot\rangle_l$ and $\langle\cdot\rangle_m$ are the expectations computed over the data and model distributions, respectively.

Conditional Gaussian-Bernoulli RBM: Figure 4 illustrates the structure of the CGBRBM designed for our CDBN architecture. As shown in Fig. 4, CGBRBM contains $N+1$ visible layers and one hidden layer.

Energy function of the CGBRBM is defined in the following:

$$E(\mathbf{v}_t, \ldots, \mathbf{v}_{t-N},\ \mathbf{h}) = -\sum_{i}^{n}\sum_{j}^{m} \frac{v_j}{\hat{\sigma}_j^2} h_i w_{ij} - \sum_{i}^{n} d_{i,t}h_i + \sum_{j}^{m} \frac{(v_{j,t} - c_{j,t})^2}{2\hat{\sigma}_j^2} \tag{21}$$

Authorized licensed use limited to: SICHUAN UNIVERSITY. Downloaded on February 09,2025 at 08:24:48 UTC from IEEE Xplore. Restrictions apply.

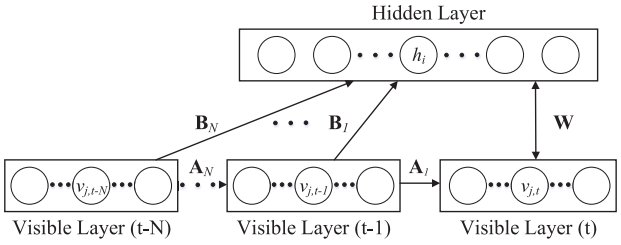

Fig. 4. The structure of CGBRBM designed for our proposed CDBN: N is the size of the previous-time observation window, and $\mathbf{A}_k$ and $\mathbf{B}_k$, where $k = 1, \ldots, N$, are the weight matrices that model the impact of the kth previous visible vector on the current visible vector and the hidden layer, respectively.

where N is the size of the previous-time observation window, v_j is the jth element of the visible vector, h_i is the ith element of the vector consisting of the hidden unit values, w_{ij} is the ijth element of the weight matrix between the visible and hidden units, $\hat{\sigma}_j$ is the standard deviation of the jth element of the visible vector, n and m are the number of the hidden units and number of the visible units, respectively, and $\mathbf{b}_t$ and $\mathbf{c}_t$ are calculated by using the following equations:

$$\begin{cases} \mathbf{d}_t = \mathbf{d} + \sum_{k=1}^{N} \mathbf{v}_{t-k}\mathbf{B}_k, \\ \mathbf{c}_t = \mathbf{c} + \sum_{k=1}^{N} \mathbf{v}_{t-k}\mathbf{A}_k, \end{cases} \tag{22}$$

where $\mathbf{b}$ and $\mathbf{c}$ denote the bias vectors for the hidden and visible layers, respectively, and v_{t-k} is the kth previous visible vector.

Based on Eq. (21), the conditional probability distributions of the hidden and visible units can be calculated as follows:

$$\begin{cases} p(h_i = 1|\mathbf{v_t} \ldots \mathbf{v_{t-N}}) = sigm\left(d_{i,t} + \sum_{j=1}^{m} \frac{w_{ij}v_{j,t}}{\hat{\sigma}_j^2}\right), \\ p\left(v_{j,t} = v|\mathbf{h}\right) = N\left(c_{j,t} + \sum_{i=1}^{n} w_{ij}h_i, \hat{\sigma}_j^2\right). \end{cases}$$

where $\mathcal{N}(\mu, \hat{\sigma}^2)$ is the normal distribution with the mean μ and the standard deviation $\hat{\sigma}$. By applying the gradient-based CD technique, the structure of the CGBRBM architecture can be updated as follows:

$$\begin{cases} w_{ij} = w_{ij} - \alpha\left(\left\langle \frac{v_{j,t}}{\hat{\sigma}_j^2}h_i \right\rangle_m - \left\langle \frac{v_{j,t}}{\hat{\sigma}_j^2}h_i \right\rangle_l\right), \\ a_{ijk} = a_{ijk} - \alpha\left(\left\langle \frac{v_{j,t-k}}{\hat{\sigma}_j^2}v_{i,t} \right\rangle_m - \left\langle \frac{v_{j,t-k}}{\hat{\sigma}_j^2}v_{i,t} \right\rangle_l\right), \\ b_{ijk} = b_{ijk} - \alpha\left(\left\langle \frac{v_{j,t-k}}{\hat{\sigma}_j^2}h_i \right\rangle_m - \left\langle \frac{v_{j,t-k}}{\hat{\sigma}_j^2}h_i \right\rangle_l\right), \\ d_i = d_i - \alpha(\langle h_i\rangle_m - \langle h_i\rangle_l), \\ c_{j,t} = c_{j,t} - \alpha\left(\left\langle \frac{v_{j,t}}{\hat{\sigma}_j^2} \right\rangle_m - \left\langle \frac{v_{j,t}}{\hat{\sigma}_j^2} \right\rangle_l\right). \end{cases} \tag{23}$$

where w_{ij}, a_{ijk}, and b_{ijk} are the elements of the weight matrices $\mathbf{W}$, $\mathbf{A}_k$, and $\mathbf{B}_k$, respectively, α denotes the learning rate, and $\langle\cdot\rangle_l$ and $\langle\cdot\rangle_m$ are the expectations computed over the data and model distributions, respectively.

The unsupervised training of RBMs and CGBRBM characterizes the hierarchical temporal features presented by the compromised and normal measurements. After the unsupervised training, a fully connected output node is added on top of the model as shown in Fig. 3. The output node is a binary node with sigmoid activation function defined in Eq. (19), which can be used to represent two labels indicating the compromised measurement and the normal one. Additionally, the fully trained structure of the neural network, as shown in Fig. 3, is fine-tuned by using backpropagation supervised training with the available labeled data [34].

In the backpropagation fine-tuning procedure, the weight matrix and the bias vector of the hth hidden layer are updated as follows:

$$\begin{cases} \Delta W_{h,i,j} = -\eta\delta_{h,i}p_{h-1,j} \\ \Delta d_{h,i} = -\eta\delta_{h,i} \end{cases} \tag{24}$$

where $\Delta W_{h,i,j}$ is the update value for the ijth element of the weight matrix, $\Delta d_{h,j}$ is the update value for the jth element of the bias vector, $p_{h-1,j}$ is the activation probability of the jth element of the $(h-1)$th hidden layer, η is the learning rate, and

$$\delta_{h,j} = p_{h,j}(1 - p_{h,j}) \sum_{k}^{K} \delta_{h+1,k}W_{h+1,j,k}, \tag{25}$$

where $p_{h,j}$ is the activation probability of the jth element of the hth hidden layer, $W_{h+1,j,k}$ is the jkth element of the weight matrix of the $(h+1)$ hidden layer, and K is the number of elements in the $(h+1)$th hidden layer. Similarly, the weight vector and the bias value of the single-unit output layer are updated as follows:

$$\begin{cases} \Delta W_{o,j} = -\eta\delta_o p_{H,j} \\ \Delta d_o = -\eta\delta_o \end{cases} \tag{26}$$

where $\Delta W_{o,j}$ is the update value for the jth element of the weight vector, Δd_o is the update value for the bias, $p_{H,j}$ is the activation probability of the jth element of the last hidden layer whose index is $h = H$, and

$$\delta_o = p_o(1 - p_o)(l_o - L), \tag{27}$$

where p_o is the activation probability of single output unit, l_o is the predicted output label, and L is the actual value of the output label.

V. Simulations

In this section, we illustrate the performance of our proposed detection mechanism using IEEE 118-bus power test system as illustrated in Fig. 5 and detailed in [35]–[37]. Additionally, we also evaluate the scalability of our work by using a larger-scale test system, the IEEE 300-bus system as detailed in [38]. In 118-bus power system, the state vector $\mathbf{x} \in \mathbb{R}^{118}$ is composed of the voltage phase angles of the individual buses and the measurement vector $\mathbf{z} \in \mathbb{R}^{490}$ consists of the measurements of the real power injection of the individual buses and branches. Furthermore, in the IEEE 300-bus

Authorized licensed use limited to: SICHUAN UNIVERSITY. Downloaded on February 09,2025 at 08:24:48 UTC from IEEE Xplore. Restrictions apply.

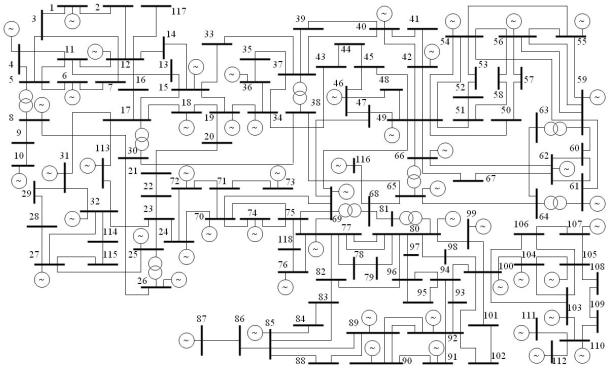

Fig. 5. IEEE 118-bus power test system.

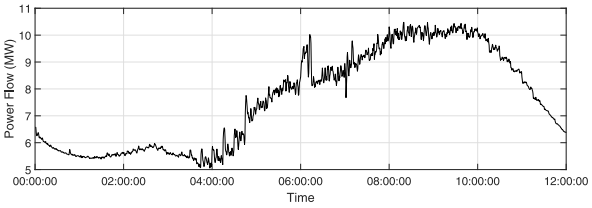

Fig. 6. One example normal load profile.

system, the state vector is $\mathbf{x} \in \mathbb{R}^{300}$ and the measurement vector is $\mathbf{z} \in \mathbb{R}^{1122}$, which are defined in the same way as those vectors in the IEEE 118-bus system. MATPOWER toolbox [39] is implemented to generate the Jacobian matrix $\mathbf{H}$ and achieve the state vectors and normal measurement vectors by performing DC optimal power flow analysis. In our simulations, we use the load profiles collected from the real world of which only a few are verified compromised data. In order to have adequate labeled compromised data for training the CDBN architecture in our DLBI scheme, we extend this type of data by analyzing the patterns of the verified compromised data from the real world by using the Fourier transform and principal components analysis techniques [40] and artificially generate more compromised data having the similar patterns with those from the real world. Moreover, we also generate additional synthetic compromised data based on our proposed FDI attack model for electricity theft described in Eq. (10). By combining the real-world data with the ones generated artificially, we obtain sufficient labeled compromised load profiles for implementing our DLBI scheme effectively.

We evaluate the performance of our real-time detection mechanism by considering four case studies. Considering the limited resources of the attackers in the practical situation, it is reasonable to assume that the FDI attacks can only corrupt limited number of load profiles. Therefore, as stated in the threat model in Section III-B, we assume that the FDI attacks can corrupt up to 64 load profiles that are the power measurements of the 64 buses in IEEE 118-bus power test system and up to 231 load profiles in IEEE 300-bus system. Figure 6 shows one example normal load profile within one day [41].

TABLE II
COMPARISON OF THE ACCURACY OF OUR SCHEME ON DETECTING THE UNOBSERVABLE FDI ATTACKS BY USING THE CDBN WITH DIFFERENT NUMBER OF HIDDEN LAYERS

Number of Compromised Buses	Accuracy (%)		
	3 Hidden Layers	4 Hidden Layers	5 Hidden Layers
32	93.90	94.59	95.58
40	96.20	96.59	96.76
48	96.78	96.93	96.98
56	97.94	97.97	97.98
64	97.69	98.06	98.10

A. Case Study I

In this case, we evaluate the performance of our proposed real-time mechanism for detecting the FDI attacks. We evaluate the impact of the following three factors on the detection accuracy: (1) the number of compromised load profiles, (2) the environment noise level, and (3) the threshold value of SVE τ. To illustrate the effectiveness of our CDBN architecture, we also compare the performance of our detection scheme, that employs the DLBI scheme, with the detection schemes that are achieved by replacing the CDBN in our detection scheme with Artificial Neural Network (ANN) and Support Vector Machine (SVM), respectively. In the simulation, the ANN is composed of one hidden layer with 25 units and the SVM is implemented with Gaussian kernel. To guarantee a fair comparison, we use the same number of labeled data for the training procedure of all these three methods.

1) The Design of Our CDBN Architecture: We first study the impact of the selection of the critical parameters for the CDBN architecture on the effectiveness of our proposed scheme in detecting the unobservable FDI attacks. The critical parameters include the number of hidden layers and the size of the previous-time observation window.

The accuracy of our scheme on detecting the unobservable FDI attacks by using the CDBN architecture with different numbers of hidden layers is shown in Table II. Each value of the detection accuracy presented in Table II is achieved by calculating the average among 30 independent trials. As shown in Table II, we consider different numbers of attacked meters for the evaluation. From Table II, we can observe the detection accuracy of our proposed scheme increases as the number of hidden layers of the CDBN architecture increases. However, the computational complexity also increases as the number of hidden layers increases. Therefore, to achieve a good balance between detection accuracy and the computational complexity, we design our CDBN architecture to have 5 hidden layers.

We continue to numerically investigate the impact of the size of the previous-time observation window of the CDBN on the performance of our scheme on detecting the unobservable FDI attacks. The simulation results are presented by using the confusion matrices in Table III. As shown in Table III, we study the detection accuracy of our proposed scheme by considering the size of the previous-time observation window increases from 3 to 6. In the simulation, we use the CDBN having 5 hidden layers and 50 units for each hidden layer. As shown in Table III, we can get that our proposed

Authorized licensed use limited to: SICHUAN UNIVERSITY. Downloaded on February 09,2025 at 08:24:48 UTC from IEEE Xplore. Restrictions apply.

TABLE III
CONFUSION MATRICES OF THE PERFORMANCE OF OUR PROPOSED SCHEME ON DETECTING THE UNOBSERVABLE FDI ATTACKS BY USING THE CDBN WITH DIFFERENT SIZES OF THE PREVIOUS-TIME OBSERVATION WINDOW

Observation Window Size	Real Label	Number of Testing Data	Identified to be Compromised	Identified to be Normal	Detection Accuracy (%)	Total Detection Accuracy (%)
3	Compromised	1290	1209	81	93.72	94.63
	Normal	1280	57	1223	95.55	
4	Compromised	1301	1244	57	95.62	95.59
	Normal	1259	56	1203	95.55	
5	Compromised	1290	1237	53	95.89	96.16
	Normal	1260	45	1215	96.43	
6	Compromised	1264	1207	57	95.49	94.41
	Normal	1276	85	1191	93.33	

TABLE IV
CONFUSION MATRIX FOR OUR DETECTION SCHEME

Actual Label	Number of Testing Data	Identified to be Compromised	Identified to be Normal	Detection Accuracy (%)
Compromised	1290	1237	53	95.89
Normal	1260	45	1215	96.43

scheme achieves the best performance in detecting the unobservable FDI attacks when the observation window has the size of 5. We believe this simulation result is reasonable. Generally speaking, more temporal information is included when the previous-time observation window size increases, and thus the accuracy of our scheme on detecting the unobservable FDI attacks increases. However, this assertion fails when the size of the observation window is too large and the structure of the CDBN, such as the number of hidden layers and the number of units in each hidden layer, is not sufficient to effectively characterize the temporal features embodied by data collected within the observation window. Considering the tradeoff between computation complexity and accuracy, we decide not to enlarge the CDBN structure and select the size of the previous-time observation window as 5.

2) Evaluation of Our Detection Scheme: To study the impact of the number of compromised load profiles on the performance of our detection scheme, we model the environment noise as the Additive White Gaussian Noise (AWGN) $\sim \mathcal{N}(0, 0.5)$, assign the threshold τ as 10, and consider the different numbers of compromised load profiles $k = 32, 40, 48, 56, 64$. We obtain each load profile consisting of 360 samples by sampling the power measurements of the load buses at every 4 minutes for one day. From each load profile consisting of 360 samples, we obtain 50 labeled samples and 200 unlabeled data for training the CDBN used in our DLBI scheme. Fig. 7 compares the detection accuracy achieved by our detection scheme with the ANN-based and SVM-based detection schemes. From Fig. 7, we can observe that our detection method can achieve the highest detection accuracy among the three different methods. Furthermore, as the number of the compromised load profiles k increases from 32 to 64, our proposed detection method can achieve the accuracy of detection above 95%.

Furthermore, the average performance of our detection scheme is presented using the confusion matrix in Tables IV. As discussed in Section V-A1, the CDBN designed in our detection scheme has 5 hidden layers each of which has 50 units and selects the size of the previous-time observation window as 5. From Tables IV, it is clear that the average accuracy of detecting an existing FDI attack is 95.89% and the false alarm ratio is 3.57%. Therefore, the overall detection accuracy is 96.16%. Figures 8 shows the Receiver Operating Characteristic (ROC) curve for our detection scheme. The ROC is achieved by plotting the False Positive Rate (FPR) versus the True Positive Rate (TPR). In our work, FPR is used to measure the specificity of our detection scheme and defined as the probability that the normal data is identified to be compromised. TPR is used to evaluate the sensitivity of our scheme and defined as the probability that the compromised data is identified to be compromised. From Fig. 8, we can observe that the area under the curve is approximately equal to 1. This indicates that the detection accuracy of our scheme is 1, which validates the excellent performance of our detection scheme [42].

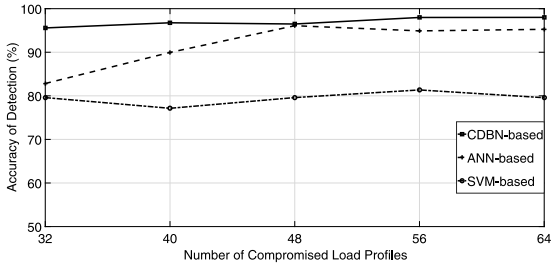

Fig. 7. The accuracy of detecting the unobservable FDI attacks when the number of the compromised load profiles changes from 32 to 64.

Additionally, our simulation is implemented by using MATLAB on a 64-bit computer with the processor of 3.3 GHz clock speed. The simulation time of our proposed scheme to detect the unobservable FDI attacks is 1.01 ms, which can satisfy the requirement of the real-time detection.

Next, we continue to numerically evaluate the robustness of our proposed detection mechanism to the noise in the data acquisition environment. In the simulation, we fix the number of the corrupted load profiles $k = 64$, the threshold $\tau = 10$, and the environment noise $\sim \mathcal{N}(0, \sigma)$ with the standard deviation σ changing from 0.5 to 2.5. As shown in Fig. 9, compared

Authorized licensed use limited to: SICHUAN UNIVERSITY. Downloaded on February 09,2025 at 08:24:48 UTC from IEEE Xplore. Restrictions apply.

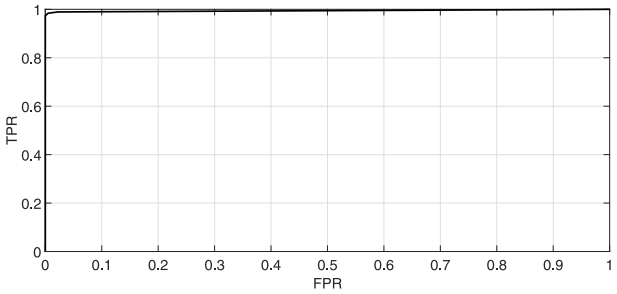

Fig. 8. ROC curve for our detection scheme.

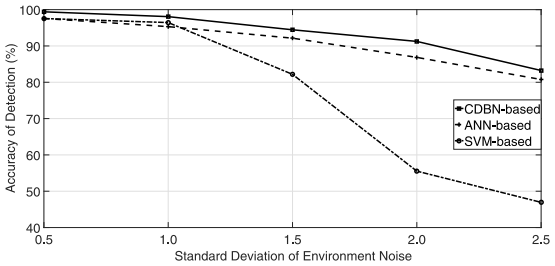

Fig. 9. Accuracy of detecting the unobservable FDI attack when the standard deviation of environment noise changes from 0.5 to 2.5.

with the ANN-based and SVM-based methods, our proposed detection scheme achieves higher detection accuracy. From Fig. 9, we can also get that as the noise level increases, the detection accuracy of all of the three detection schemes decrease. This is because that the patterns of the normal data and the compromised data are less distinguishable when the noise level increases. As shown in Fig. 9, it is clear that our proposed detection mechanism can achieve the accuracy of detection above 90% when the standard deviation $\sigma \leq 2.0$, which illustrates the robustness of our detection scheme to the environment noise. In other words, as long as the difference caused by the injected attack vector $\mathbf{a}$ is larger than that is caused by the environment noise, our proposed detection scheme is able to achieve a high detection accuracy.

At last, we explore the interaction between the SVE and DLBI scheme that are two essential components of our proposed real-time detection mechanism. To achieve this goal, we analyze the impact of the SVE detection threshold on the accuracy of our DLBI scheme on detecting the unobservable FDI attacks. We set $k = 64$, $\sigma = 0.5$, and the SVE detection threshold τ changing from 5 to 25. For each value of τ, we sample the power flow measurements at load buses every four minutes, resulting in 360 data sets for one day's time interval. As shown in Fig. 10, we can observe that compared with ANN-based and SVM-based detection schemes, our method that exploits CDBN architecture can detect the attack with the accuracy above 90 % even when the threshold is less than 10, which validates that our method outperforms the ANN-based and SVM-based schemes. In our proposed FDI attack model for electricity theft, we assume that the attackers have some knowledge of the detection threshold of SVE. Therefore, when the detection threshold of SVE increases, the attacks will also increase the value of their injected false data which

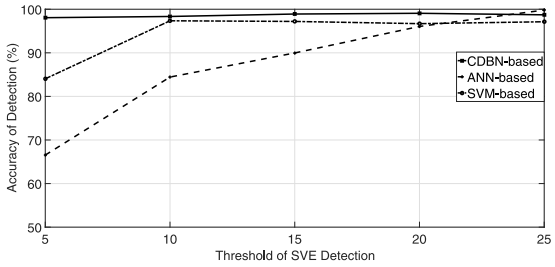

Fig. 10. Accuracy of detecting the unobservable FDI attacks when the detection threshold of SVE changes from 5 to 25.

may cause a bigger difference between the patterns of the real measurements and compromised ones. Accordingly, the ANN-based detection scheme can realize a satisfactory accuracy for detecting the attacks when the SVE threshold $\tau \geq 10$ and the SVM-based detection scheme can detect the attack with the relatively high accuracy when $\tau \geq 20$.

B. Case Study II

In this case, we consider a scenario where there exist occasional operation faults potentially resulting in the deviation on the state measurements. We will evaluate the performance of our proposed detection scheme in the presence of the occational operation faults. In the simulation, we model the occasional operation faults by assuming that they cause the readings of the meters to be held at a constant value within certain time interval. We study the impact of the following three factors on the performance of our detection scheme: (1) the number of the buses having occasional operation faults, (2) the environment noise level, and (3) the threshold value of SVE τ.

We first evaluate the performance of our proposed scheme on detecting the unobservable FDI attacks by considering the different number of simultaneous operation faults. We generate the measurement vectors with the occasional operation fault lasting for two hours within one day. Due to the level of the power flow on the buses is about tens megawatts, we set the constant value of -10 megawatts as the occasional operation fault. We can conclude from our both numerical and theoretical analysis that this operation fault is able to bypass the SVE mechanism with the threshold $\tau = 10$. Furthermore, we set the standard deviation of environment noise $\sigma = 0.5$ and consider the number of the buses having the operation faults simultaneously changing from 1 to 5. The simulation result is presented in Fig. 11, which illustrates that the detection accuracy achieved by our proposed scheme is above 97% when the number of the simultaneous operation faults $k \geq 3$. This is because, when $k \geq 3$, the operation faults lead to the patterns of compromised data that are distinguishable from the normal patterns by using our detection scheme.

To study the robustness of our detection method to the environment noise, we consider different noise levels characterized by the standard deviation σ changing from 0.5 to 2.5. We also set the number of meters having the operation faults as 2 and the detection threshold of as $\tau = 10$. The simulation result is shown in Fig. 12. From Fig. 12, we can observe that the detection accuracy of our scheme decreases as the noise level of

Authorized licensed use limited to: SICHUAN UNIVERSITY. Downloaded on February 09,2025 at 08:24:48 UTC from IEEE Xplore. Restrictions apply.

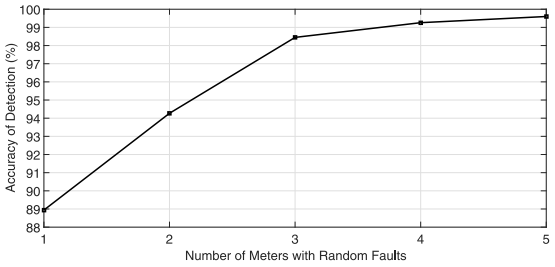

Fig. 11. Accuracy of detecting the unobservable FDI attack when the number of meters having occasional operation faults changing from 1 to 5.

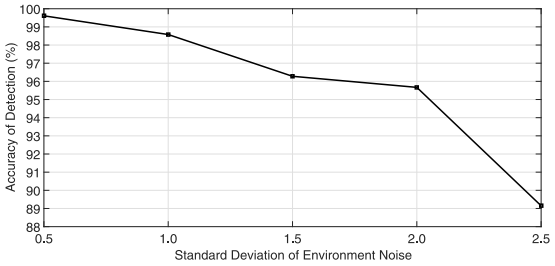

Fig. 12. Accuracy of detecting unobservable FDI attack when the standard deviation of the environment noise changes from 0.5 to 2.5.

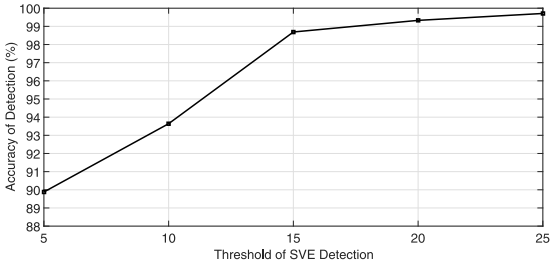

Fig. 13. Accuracy of detecting the unobservable FDI attack when the detection threshold of the SVE changes from 5 to 25.

the environment increases, which is similar to the conclusion achieved in Case Study I.

At last, we study the interaction between the SVE and the DLBI scheme by numerically analyzing the impact of the detection threshold of the SVE on the performance of the DLBI scheme. In the simulation, we set the standard deviation of the environment AWGN as $\sigma = 0.5$, the number of the meters compromised by the occasional operation fault as 2, and the detection threshold of the SVE τ increasing from 5 to 25 with the changing step 5. As shown in Fig. 13, the detection accuracy of our detections scheme is higher than 98.5% when the detection threshold of the SVE is $\tau \geq 15$. We believe it is reasonable. When the detection threshold of the SVE increases, more FDI attacks bypass the SVE system. This results in the bigger difference between the patterns of the compromised data and those of the normal data, which potentially increases the detection accuracy of the subsequent DBLI scheme.

C. Case Study III

In this case, we evaluate the scalability of our detection scheme by using IEEE 300-bus power test system. In this test

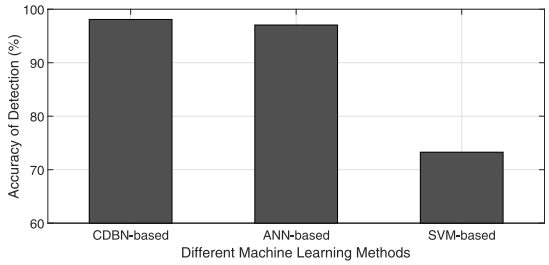

Fig. 14. Accuracy of detecting the unobservable FDI attack in IEEE 300-bus power test system.

system, we assign 231 buses as the load buses and assume that the attackers can have compromise up to 165 load buses by launching FDI attacks. Consider that practically some buses may not have the meters/sensors, in our simulation, we assume that only 100 load buses have meters/sensors to provide real-time measurements. We evaluate the performance of our detection scheme by comparing with the ANN-based and the SVM-based scheme stated in Case Study I. In the simulation, we set the detection threshold of the SVE as $\tau = 30$ and the standard deviation of the AGWN as $\sigma = 0.5$. The authors would like to mention that we select a larger value of τ for IEEE 300-bus system to reduce the potentially increased false positive by considering that the larger system is more sensitive to the environment noise according to the higher structure complexity. As illustrated in the simulation results in Fig. 14, among the three schemes for comparison, our scheme achieves the highest accuracy of detection and SVM-based scheme has the worst performance with the detection accuracy below 75 %. The simulation result in this case study also implies that the full observation of the power system is not required for the success of our detection scheme. In our ongoing work, we analyze the minimum number of the load buses having the sensing capability required by our proposed detection scheme.

D. Case Study IV

In this case, we take the initial step towards studying the real-world FDI attacks for electricity theft. In the practical situation, the attackers launch the FDI attacks to realize electricity theft while reducing the probability of being detected, and thus it is reasonable to assume that the attackers are prone to limit the attack ratio, which is defined as the ratio between the values of the injected negative data and those of the associated actual measures of the load consumption, to be $\leq \zeta$. Therefore, we revise the third constraint in the attack model proposed in Eq. (10) as follows:

$$\mathbf{N}(\mathbf{a} + \zeta \mathbf{L}) \preceq 0, \tag{28}$$

where the selection of ζ is dependent on the personality of the attackers and $\zeta = \frac{1}{2}$ is selected for our simulation in this case study.

Then we solve the revised FDI attack model to generate the synthetic labeled compromised data for training the CDBN architecture of our detection scheme. The performance of our proposed detection scheme is also evaluated by using the

Authorized licensed use limited to: SICHUAN UNIVERSITY. Downloaded on February 09,2025 at 08:24:48 UTC from IEEE Xplore. Restrictions apply.

TABLE V
CONFUSION MATRIX FOR OUR DETECTION SCHEME

Actual Label	Number of Testing Data	Identified to be Compromised	Identified to be Normal	Detection Accuracy (%)
Compromised	1276	1196	80	93.73
Normal	1196	19	1225	98.51

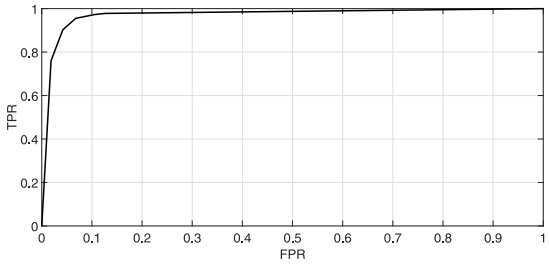

Fig. 15. The ROC curve for our detection scheme.

revised attack model and the evaluation results are presented by using the confusion matrix and ROC curve which are in Table V and Fig. 15, respectively. From Table V and Fig. 15, we can observe that our proposed scheme can detect the unobservable FDI attacks with detection accuracy above 93%, which illustrates the effectiveness of our proposed detection scheme on detecting the more practical FDI attacks.

VI. CONCLUSION

In this paper, we propose a deep learning-based scheme for detecting the FDI attacks in real-time. We propose one type of FDI attack for the purpose of power theft by instructing an optimization model. Our proposed scheme employs Conditional Deep Belief Network (CDBN) to efficiently reveal the high-dimensional temporal behavior features of the unobservable FDI attacks that bypass the SVE mechanism. The captured behavior features are then exploited to detect the potential FDI attacks on real-time measurements. In the simulations, we illustrate our work by four cases using IEEE 118-bus power test system and IEEE 300-bus system. In the first two cases, we study the impacts of the number of the compromised measurements, the noise level of the environment, and detection threshold for our SVE on the performance of our detection scheme. To investigate the scalability of our scheme, we use IEEE 300-bus system in the third case. Furthermore, towards modeling the real-world FDI attacks, we revise our FDI attack model proposed in Section III-B by considering more practical constraint of the attack.In the simulations, we evaluate the performance of our detection method by comparing with those of the ANN-based and SVM-based methods. The simulation results illustrate that our detection scheme is resilient to the different numbers of the attacked measurements, the different detection thresholds of SVE, and some levels of environment noise levels. The simulation results also show that our detection method can achieve the high detection accuracy in the presence of the occasional operation faults. In the future, we will extend our work by modeling more practical behaviors of the FDI attacks and analyze the minimum number of the sensing units required for the success of our proposed detection scheme.

REFERENCES

[1] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Trans. Inf. Syst. Security*, vol. 14, no. 1, pp. 1–33, May 2011.
[2] A. A. Cárdenas, S. Amin, G. Schwartz, R. Dong, and S. Sastry, "A game theory model for electricity theft detection and privacy-aware control in AMI systems," in *Proc. 50th Annu. Allerton Conf. Commun. Control Comput. (Allerton)*, pp. 1830–1837, Oct. 2012.
[3] V. Gaur and E. Gupta, "The determinants of electricity theft: An empirical analysis of Indian states," *Energy Policy*, vol. 93, pp. 127–136, 2016.
[4] Ç. Yurtseven, "The causes of electricity theft: An econometric analysis of the case of Turkey," *Utilities Policy*, vol. 37, pp. 70–78, Dec. 2015.
[5] S. Mathankumar and P. Loganathan, "Detection of power theft in HT consumer using SCADA interfacing with GIS system," *Int. J. Res. Eng. Appl. Sci.*, vol. 5, no. 7, pp. 1–12, Jul. 2015.
[6] Q. Yang *et al.*, "On false data-injection attacks against power system state estimation: Modeling and countermeasures," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 3, pp. 717–729, Mar. 2014.
[7] 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.
[8] L. Liu, M. Esmalifalak, Q. Ding, V. A. Emesih, and Z. Han, "Detecting false data injection attacks on power grid by sparse optimization," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 612–621, Mar. 2014.
[9] Z. Hu *et al.*, "False data injection attacks identification for smart grids," in *Proc. 3rd Int. Conf. Technol. Adv. Elect. Electron. Comput. Eng. (TAEECE)*, Beirut, Lebanon, pp. 139–143, Apr./May 2015.
[10] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and A. M. Mohammad, "Detection of abnormalities and electricity theft using genetic support vector machines," in *Proc. IEEE Region Conf. TENCON*, Nov. 2008, pp. 1–6.
[11] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and M. Mohamad, "Nontechnical loss detection for metered customers in power utility using support vector machines," *IEEE Trans. Power Del.*, vol. 25, no. 2, pp. 1162–1171, Apr. 2010.
[12] S. S. S. R. Depuru, L. Wang, and V. Devabhaktuni, "Support vector machine based data classification for detection of electricity theft," in *Proc. IEEE/PES Power Syst. Conf. Expo.*, Phoenix, AZ, USA, Mar. 2011, pp. 1–8.
[13] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and F. Nagi, "Improving SVM-based nontechnical loss detection in power utility using the fuzzy inference system," *IEEE Trans. Power Del.*, vol. 26, no. 2, pp. 1284–1285, Apr. 2011.
[14] J. Heaton, *Artificial Intelligence for Humans, Volume 3: Deep Learning and Neural Networks*. North Charleston, SC, USA: CreateSpace Independent Publ. Platform, Oct. 2015.
[15] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.
[16] S. Sukhbaatar, T. Makino, K. Aihara, and T. Chikayama, "Robust generation of dynamical patterns in human motion by a deep belief nets," *J. Mach. Learn. Res.*, vol. 20, pp. 231–246, Nov. 2011.
[17] A. Fischer and C. Igel, "An introduction to restricted Boltzmann machines," in *Proc. CIARP*, vol. LNCS 7441. Buenos Aires, Argentina, 2012, pp. 14–36.
[18] G. W. Taylor, G. E. Hinton, and S. T. Roweis, "Modeling human motion using binary latent variables," in *Proc. Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2006, pp. 1345–1352.
[19] G. Taylor, "Composable, distributed-state models for high-dimensional time series," Ph.D. dissertation, Dept. Comput. Sci., Univ. at Toronto, Toronto, ON, Canada, 2009.

Authorized licensed use limited to: SICHUAN UNIVERSITY. Downloaded on February 09,2025 at 08:24:48 UTC from IEEE Xplore. Restrictions apply.

[20] J. Wei and G. J. Mendis, "A deep learning-based cyber-physical strategy to mitigate false data injection attack in smart grids," in *Proc. Cyber Phys. Syst. Week (CPS)*, Vienna, Austria, 2016, pp. 1–6.
[21] P. Kundur, N. J. Balu, and M. G. Lauby, *Power System Stability and Control*, vol. 7. New York, NY, USA: McGraw-Hill, 1994.
[22] D. Van Hertem, J. Verboomen, K. Purchala, R. Belmans, and W. L. Kling, "Usefulness of DC power flow for active power flow analysis with flow controlling devices," in *Proc. 8th IEE Int. Conf. AC DC Power Transm. (ACDC)*, London, U.K., Mar. 2006, pp. 58–68.
[23] M. Thomas and J. McDonald, *Power System SCADA and Smart Grids*. Boca Raton, FL, USA: CRC Press, 2015.
[24] R. Krutz, *Securing SCADA Systems*. New York, NY, USA: Wiley, 2005.
[25] J. Gupta and A. Kumar, "Fixed pitch wind turbine-based permanent magnet synchronous machine model for wind energy conversion systems," *J. Eng. Technol.*, vol. 2, no. 1, pp. 52–62, 2012.
[26] M. A. G. de Brito, L. P. Sampaio, G. Luigi, G. A. e Melo, and C. A. Canesin, "Comparative analysis of MPPT techniques for PV applications," in *Proc. Int. Conf. Clean Elect. Power (ICCEP)*, Jun. 2011, pp. 99–104.
[27] N. Cao, Y. J. Cao, and J. Y. Liu, "Modeling and analysis of grid-connected inverter for PV generation," in *Advanced Materials Research*, vol. 760. Dürnten, Switzerland: Trans Tech, 2013, pp. 451–456.
[28] D. Haribabu, A. Vangari, and J. N. Sakamuri, "Dynamics of voltage source converter in a grid connected solar photovoltaic system," in *Proc. Int. Conf. Ind. Instrum. Control (ICIC)*, May 2015, pp. 360–365.
[29] Y. Lei, A. Mullane, G. Lightbody, and R. Yacamini, "Modeling of the wind turbine with a doubly fed induction generator for grid integration studies," *IEEE Trans. Energy Convers.*, vol. 21, no. 1, pp. 257–264, Mar. 2006.
[30] J. Nocedal and S. Wright, *Numerical Optimization*. New York, NY, USA: Springer, 2006.
[31] M. Avriel, *Nonlinear Programming: Analysis and Methods*. Mineola, NY, USA: Dover Books Comput. Sci., Sep. 2003.
[32] S. Wright and J. Nocedal, *Numerical Optimization*, vol. 35. New York, NY, USA: Springer, 1999, pp. 67–68.
[33] M. A. Carreira-Perpignan and G. Hinton, "On contrastive divergence learning," in *Proc. Artif. Intell. Stat.*, 2005, pp. 1–8.
[34] J. Li, J.-H. Cheng, J.-Y. Shi, and F. Huang, "Brief introduction of back propagation (BP) neural network algorithm and its improvement," in *Advances in Computer Science and Information Engineering*, Heidelberg, Germany: Springer, 2012, pp. 553–558.
[35] S. Blumsack, *Network Topologies and Transmission Investment Under Electric-Industry Restructuring*, Ph.D. dissertation, Eng. Public Policy, Carnegie Mellon Univ., Pittsburgh, PA, USA, 2006.
[36] P. Anderson and A. Fouad, *Power System Control and Stability*. New York, NY, USA: IEEE Press, 2003.
[37] P. Rao, M. L. Crow, and Z. Yang, "STATCOM control for power system voltage control applications," *IEEE Trans. Power Del.*, vol. 15, no. 4, pp. 1311–1317, Oct. 2000.
[38] P. Hines, S. Blumsack, E. C. Sanchez, and C. Barrows, "The topological and electrical structure of power grids," in *Proc. 43rd Hawaii Int. Conf. Syst. Sci.*, Jan. 2010, pp. 1–10.
[39] R. D. Zimmerman and C. Murillo-Sánchez. (Dec. 2016). *MATPOWER 6.0 User's Manual*. [Online]. Available: http://www.pserc.cornell.edu/matpower/manual.pdf
[40] I. Jolliffe, *Principal Component Analysis*. New York, NY, USA: Springer, 2002.
[41] C. Grigg *et al.*, "The IEEE reliability test system-1996. A report prepared by the reliability test system task force of the application of probability methods subcommittee," *IEEE Trans. Power Syst.*, vol. 14, no. 3, pp. 1010–1020, Aug. 1999.
[42] M. Gonen, *Analyzing Receiver Operating Characteristic Curves With SAS*. Cary, NC, USA: SAS Inst., 2007.

Youbiao He received the B.E. degree in electrical engineering with the Dalian University of Technology, China, in 2013. He is currently pursuing the M.S. degree in electrical and computer engineering with the University of Akron. His research interests include the cyber-physical energy system, cyber-physical systems security, renewable energy integration, game theoretic modeling, and machine learning. Furthermore, he has authored the paper entitled *Towards Smarter Cities: A Self-Healing Resilient Microgrid Social Network* for which he was a recipient of the Best Paper Award in the 2016 IEEE PES General Meeting.

Gihan J. Mendis received the B.Sc. Eng. (Hons.) degree in electronics and telecommunication engineering from the University of Moratuwa, Sri Lanka, in 2012, and the M.Sc. degree in electrical engineering from the University of Akron, in 2016, where he is currently pursuing the Ph.D. degree in electrical and computer engineering. He was with Ocius Technologies LLC, as the Principal Scientist, in 2016, for four months. His research interests include the deep learning algorithms, decentralize deep learning, cyber security, cognitive radar sensors, and cognitive radio. Furthermore, he has co-authored the paper entitled *Towards Smarter Cities: A Self-Healing Resilient Microgrid Social Network* for which he was a recipient of the Best Paper Award in the 2016 IEEE PES General Meeting.

Jin Wei received the B.E. degree from the Beijing University of Aeronautics and Astronautics, China, in 2004, the M.S. degree in electrical engineering from the University of Hawaii at Manoa in 2008, and the Ph.D. degree in electrical and computer engineering from the University of Toronto, Canada, in 2014. She is an Assistant Professor of Electrical and Computer Engineering with the University of Akron. She was a Post-Doctoral Fellow with National Renewable Energy Laboratory in 2014, for 3 months. Her research interests include the cyber-physical energy system, cyber-physical systems security, renewable energy integration, security and privacy for large-scale complex systems, deep-learning algorithm development, social sensing, game-theoretical analysis, opportunistic hybrid communication infrastructure design, and cognitive wired/wireless communication networks. Her research group was a recipient of the Best Paper Award in the 2016 IEEE PES General Meeting for the paper entitled *Towards Smarter Cities: A Self-Healing Resilient Microgrid Social Network*.

Authorized licensed use limited to: SICHUAN UNIVERSITY. Downloaded on February 09,2025 at 08:24:48 UTC from IEEE Xplore. Restrictions apply.