IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal
SPECIAL SECTION ON SECURITY ANALYTICS AND INTELLIGENCE FOR CYBER PHYSICAL SYSTEMS

Received September 7, 2017, accepted October 15, 2017, date of publication November 2, 2017, date of current version December 5, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2769099

A Novel Data Analytical Approach for False Data Injection Cyber-Physical Attack Mitigation in Smart Grids

YI WANG[1], (Member, IEEE), MAHMOUD M. AMIN[1,2], (Senior Member, IEEE), JIAN FU[3], (Senior Member, IEEE), AND HEBA B. MOUSSA[4], (Student Member, IEEE)

[1]Department of Electrical and Computer Engineering, Manhattan College, Riverdale, NY 10471, USA
[2]Power Electronics and Energy Conversion Department, Electronic Research Institute, Cairo 12611, Egypt
[3]Department of Electrical Engineering and Computer Science, Alabama A&M University, Normal, AL 35762, USA
[4]Department of Electrical and Computer Engineering, City College of New York, New York, NY 10031, USA

Corresponding author: Yi Wang (yi.wang@manhattan.edu)

ABSTRACT False data injection cyber-physical threat is a typical integrity attack in modern smart grids. These days, data analytical methods have been employed to mitigate false data injection attacks (FDIAs), especially when large scale smart grids generate huge amounts of data. In this paper, a novel data analytical method is proposed to detect FDIAs based on data-centric paradigm employing the margin setting algorithm (MSA). The performance of the proposed method is demonstrated through simulation using the six-bus power network in a wide area measurement system environment, as well as experimental data sets. Two FDIA scenarios, playback attack and time attack, are investigated. Experimental results are compared with the support vector machine (SVM) and artificial neural network (ANN). The results indicate that MSA yields better results in terms of detection accuracy than both the SVM and ANN when applied to FDIA detection.

INDEX TERMS Data analytical, false data injection, cyber-physical attack, smart grid.

I. INTRODUCTION

Internet of Things (IoTs) and Cyber-physical systems (CPSs) have emerged to represent the next generation of engineering systems that will drive the fourth industrial revolution. Smart grid systems have evolved to follow this IoTs/CPSs trend, becoming a critical societal infrastructure as it involves vital elements in our day-to-day life. Meanwhile, the modern smart grid system brings new security challenge, i.e., cyber-physical attack or threat, due to the deep integration of the physical space (traditional power network infrastructure) with the cyber space (information sensing, processing and control) for efficient energy consumption and transmission. Therefore, cyber-physical attack exploits the vulnerabilities in the cyber space that will have an adverse impact on the physical space of smart grids. Consequently, it can undermine or even totally disrupt the control systems underlying electric power grids. Cyber-physical attacks have resulted in many security problems, and have become a critical concern for both industrial control system users and vendors. In 2010 the Stuxnet worm attacked Iran's Natanz nuclear fuel enrichment facility and infected computers in other countries, including India, Indonesia, China, Azerbaijan, South Korea, Malaysia, the United States, the United Kingdom, Australia, Finland and Germany [1]. More recently, Ukraine's power grid was successfully disrupted by a Trojan called ''BlackEnergy'' on Dec 23, 2015, resulting in several power outages that affected approximately 225,000 customers [2].

False data injection attacks (FDIAs), or bad data injection attack, are the most studied cyber-physical attacks in smart grid security from recent primary studies [3]. In this paper, FDIA is broadly defined including data integrity attacks in both cyber-space and physical space. Cyber-space FDIAs contain cyber-attacks on the communication messages from sensors and meters, etc., that result in invalid operations. Physical space FDIA is introduced in the realm of electrical power grid by Deng *et al.* in 2009 [4]. This attack violates data integrity, which aims to inject malicious measurements and to modify the state estimation results. FDIAs are stealth attacks that can bypass the existing detection scheme. There are several consequences of FDIA. For example, the dynamic pricing signal can be manipulated by the malicious attacker [5]. Therefore, the adversary injects false measurements to

disrupt the smart grid operations and cause economic loss.

One key sensing component is the source of FDIAs: synchronized phasor measurement units (PMUs) [6]. PMUs provide a solution for time-synchronizing the phase and sequence measurements from nodes that are geographically dispersed, so as to monitor, control, evaluate, and protect the smart grid system. However, traditional defense approaches for FDIAs have not prepared for the data challenges caused by the large-scale deployment of PMU in the future smart grid CPS. Large volumes of data produced from PMU presents real time computational and storage challenges [7]. However, this challenge also represents an opportunity for data analytical techniques, such as machine learning, to detect and prevent FDIAs. Currently, machine learning algorithms have been applied to cybersecurity fields of sensor networks, vehicular networks and smart grid [8]–[11]. This trend comes from the fact that cybersecurity has become more sophisticated and complex than before, and traditional manual and signature-based approaches are no longer effective [12]. The feature of machine learning is that it attempts to process large volumes of data by learning well from patterns at a level that can be beyond human comprehension. Besides, machine learning is able to learn the non-linear, complex relationship between measurements to detect false PMU data injection. Therefore, machine learning is a very attractive data analytical approach for PMU data analytics under FDIAs.

Several analytical approaches have been proposed for FDIAs mitigation in recent years, including the Particle Swarm Optimization, Bayesian framework, Random Forests, Adaboost, and the common path mining method to mitigate FDIAs [12]–[15]. A most recent work proposed a FDIA detection method based on support vector machine (SVM) [16]. Inspired by recent work, a novel data analytical approach based on the margin setting algorithm (MSA), is proposed to mitigate cyber-physical attacks. MSA is a relatively new machine learning algorithm that has been used in image processing fields [17], [18]. This is the first work to employ MSA to detect false data injection in smart grids. The test results demonstrate the validity of the proposed MSA, which outperforms the other existing machine learning approaches, such as SVM and artificial neural networks with respect to FDIAs detection accuracy.

II. RELATED WORK

Recently, research has been conducted to defend against FDIA as a cyber-physical threat in CPS. The central part of CPS is the control system. From a modeling point of view from the control theory community, most of the research is based on model-based paradigm [3]. However, with the growing trend of large scale CPSs, huge amounts of data are produced. The enormous data, i.e., ''Big Data'' in CPS demands more cost-effective solutions: data-based paradigms, to handle the big data challenge. Accordingly, the FDIAs defense mechanisms can be classified as theoretical, application-based approaches and data analytical approaches using machine learning, artificial intelligence, data mining, statistics, etc.

A. THEORETICAL-BASED APPROACHES

Some methods have been proposed to defend FDIAs theoretically. The attacker builds attack vectors that are against the state estimations [19]. The pioneer work of FDIAs presented this attack with the assumption that power grid configuration, such as topological information and transmission line admittance are known to the adversary. If the topology information is even altered by the adversary, the state estimations will be falsified. Kim et al. presented an FDIA under the incorrect network topology that deceives control center by altering certain meters and network switches [20]. Jia *et al.* discussed that FDIAs can cause errors in topology and state estimation, which results in bad data injection to the real-time location margin price [21]. Similarly, another topology poisoning attack scenario showed that it could compromise optimal power flow routine [22]. Esmalifalak *et al.* proposed an inference algorithm based on a linear independent component analysis (ICA). Grid topology and power states could be inferred from phasor observations by an attacker. Attacker could launch a low detectable FDIA [23]. Besides, Liu *et al.* propose nuclear norm minimization method and low rank matrix factorization approach to solve matrix separation and detect FDIAs [24]. These FDIAs are based on the DC power flow model. Besides, FDIAs under the AC models are discussed by some researchers as well. Rahman and Mohsenian-Rad presented a method for constructing a non-linear FDIA under AC model [25]. Gu *et al.* proposed a new detection method by tracking the dynamics of measurement variations calculated using metrics of Kullback-Leibler distance (KLD) of two distributions. When FDIAs are injected, KLD becomes larger as current distribution of measurement variations deviates from old data [26].

B. APPLICATION-BASED APPROACHES

Other methods are proposed on the application domain to FIDAs. Lin et al. proposed an approach on FDIAs against distributed energy distribution. This approach led to the consequence of disrupting the effectiveness of the distribution process [27]. Besides, some research work indicates that large scale PMU deployment can defend against FIDAs, but the cost is very high [28]. To address this drawback, a PMU placement algorithm is proposed by modeling an optimization problem. Solving this problem will place minimum number of PMUs at optimal locations with least cost [29]. Another similar work proposed greed algorithms that can be utilized to place secure PMUs at appropriate locations [28]. However, the PMU deployment still suffers from the time stamp attack that sends falsified GPS signals [30]. Besides, FDIAs can inject a set of significant measurements, which masks the transmission line outages [31]. The other application-based methods take into account of the structures of networks. Ozay *et al.* considered both centralized and distributed structures under sparse FDIAs, which compromises

a modest number of meter readings [32]. Liu *et al.* proposed an efficient method to find the optimal FDIA local region with less network information, i.e., parameter information of a limited number of power lines [33].

C. DATA ANALYTICS APPROACHES

Only a small number of FDIA research are based on data analytical methods, including statistics, data mining, machine learning, and artificial intelligence. Especially, machine learning and deep learning used by data intensive applications, i.e., big data collected from sensors and meters, provide a way to tackle complex structure data sets with artificial intelligence.

For statistical methods, a regularized maximum likelihood estimator (MLE) is proposed to recover the grid topology from public available market data. This market data is the real-time locational marginal prices that are computed based on Lagrange multipliers of the network-constrained economic dispatch. The grid topology Laplacian matrix is estimated through an algorithm based on iterative direction method of multipliers (ADMM) [35]. Another recent work proposed a new Cumulative Sum (CUSUM) algorithm based on the generalized likelihood ratio (GLR) for FDIAs online sequential detection [13]. This approach outperforms the existing first-order cumulative sums detectors with respect to the average detection delay. Valenzuela et al. developed an algorithm to detect anomalies from real-time power flow results based on principle component analysis (PCA). PCA is used to differentiate regular power flow variability from irregular ones [35]. Another method utilized PCA approximation to conduct roughly stealthy FDIAs without the knowledge of grid topology [36].

For machine learning, data mining and artificial intelligence approaches, most of them are used for anomaly detection for FDIAs. Kosut *et al.* proposed an approach that utilized a Bayesian framework to formulate the FDIAs problems. They introduced a heuristic to detect FDIAs with low computational overhead [12]. Hink *et al.* detected the power system faults and cyber-attacks using several batch processing-based machine learning and data mining algorithms, including Random Forests, Naïve Bayes, SVM, Adaboost, etc. [14]. Pan *et al.* proposed a hybrid intrusion detection system using common path mining method to detect abnormal power system events from data with a fusion of PMU data, information from relay, network security logs and energy management system (EMS) logs [15]. Landford *et al.* proposed a machine learning approach to detect FDIAs using a two-class SVM. This method analyzed the change of correlation between two PMU parameters using Pearson correlation coefficient [16].

III. MACHINE LEARNING ALGORITHMS

Supervised learning approach leverages the prior knowledge of the training labels and features in the training set to make classifications or predications for the testing data sets. MSA is able to learn from the large volumes of historical time series PMU data to detect FDI anomalies and make predications. Another two popular classification algorithms are support vector machines and artificial neural networks. They are used for comparison in our experiment.

A. ARTIFICIAL NEURAL NETWORKS

Artificial neural networks (ANN) is a computational architecture that mimics the biological neural structure of the brain and form interconnected groups of artificial neurons. Each neuron in ANNs is a set of input values (x^i) and associated weights (w_i). The neurons are organized into layers. ANN starts with the input layer. The next layer contains at least one hidden layer. The final layer is the output layer. Among the various architecture of ANN, the feedforward, back propagation (BP) neural network is the most popular, effective model to recognize patterns.

Suppose the input of ANN is $x = [x_1, x_2, \ldots, x_n]^T$ and output $y(x) = [y_1, y_2, \ldots, y_n]^T$. The exists a mapping M from the input space $X : \{x \in X | x$ is the input to the system$\}$ to output space $Y : \{y \in Y | y$ is the output of the system for given input $x\}$. The mapping M is presented as [37]:

$$M : X \rightarrow Y \tag{1}$$

The BP learning process can be considered a process to gradually adjust the network internal parameters, i.e., weight w in the weight space ω, i.e., $w \in \omega$, so that the difference between the expected outputs $\hat{y}(x, w)$ and real outputs $y(x)$ of the network is minimal:

$$\min \| \hat{y}(x, w) - y(x) \|_2 \tag{2}$$

This process contains two phases: forward propagation and weight update. During first phase, the input value is propagated from input layer, via the hidden layer to the output layer using the weight value and offset value of the network. Then the output of the network is compared with the expected output. The difference between the real output and expected output is the error. The second phase, the weight is continuously updated and modified to minimize the error.

B. SUPPORT VECTOR MACHINE

SVM is a supervised machine learning algorithm that uses training data sets to make predictions. The goal of SVM is to separate a given set of binary labeled training data with a hyperplane that is maximally distant from them, i.e. with maximized margin. However, a hyperplane cannot separate the training data if they are non-linearly separable. Hence, "kernels trick" is introduced to map the training data from its original input space to a high dimensional space where a linear mapping can be achieved. In this case, the hyperplane found by the SVM in the feature space corresponding to a non-linear decision boundary in the original input space. Several common kernel functions are: linear kernel, Gaussian radial basis kernel and Sigmoid kernel, etc [38].

Given a training data set with n data samples as (x_i, y_i), where $x_i \in R^d, y_i \in \{-1, 1\}, i = 1, 2, \ldots, n.$, x_i is the

feature vector and y_i is the classification label. The decision boundary of SVM is a hyerplane H: (w, b), where w is a normal vector, or a weight vector, perpendicular to the hyperplane with initial value $w_0 = 0$. It is adjusted iteratively each time when training examples are misclassified by current w. b is the bias. The hyperplane equation is defined as

$$w^T x_i + b = 0. \tag{3}$$

To assign class labels to each class for test data, another two hyperplane H1 and H2 are used to determine their classification labels:

$$\begin{cases} H1 : w^T x_i + b \geq 1, & if\ y_i = +1 \\ H2 : w^T x_i + b \leq -1, & if\ y_i = -1 \end{cases} \tag{4}$$

Therefore, since the final goal is to find the hyperlane with the largest margin, it should satisfy the equation (4) and minimize weight vector $||w||_2$, where $||\cdot||_2$ is Euclidian norm function. The points on H1 and H2 are called support vectors. To solve the minimization problem, Lagrange multiplier method and Karush-Kuhn-Tucker (KTT) condition are used to transform this problem to its dual problem. Therefore, an equivalent dual problem of minimizing $||w||_2$ is a maximization problem solving by QP (Quadratic Programming) below:

$$\begin{aligned} &maximize\ W(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j x_i \cdot x_j \\ &subject\ to \sum_{i=1}^{m} y_i \alpha_i = 0 \\ &\qquad W = \sum_{i=1}^{m} \alpha_i y_j x_i \\ &\qquad 0 \leq \alpha_i \leq C, \quad i = 1, \ldots, m. \end{aligned} \tag{5}$$

Where $\alpha_1, \ldots \alpha_m$ is the Lagrangian multiplier associated with each training example (x_i, y_i). The Lagrangian multipliers are bounded by C, called a box constraint. α_i is the lagrangian multipliers for the support vectors.

IV. FALSE DATA INJECTION ATTACKS

A. FALSE DATA INJECTION ATTACK

False data can be injected into the physical model of the smart grid, so that the state estimation is corrupted. Let us assume the physical model of smart grid with N buses in AC power model can be viewed as follows [40]:

$$z = h(x) + e \tag{6}$$

where x is an n-dimensional state vector $x\{x_1, x_2, \ldots, x_n\}^T$ $(x_i \in R)$ for n state variables. z is an m-dimensional state vector $z\{z_1, z_2, \ldots, z_m\}^T$ $(z_i \in R)$ for m measurements, including the injected active or reactive power flow for each bus, transmission lines, etc. e is a m–dimensional error vector. This error vector assumes a Gaussian noise with mean value of 0 and covariance R. In the DC power model,

$$z = Hx + e \tag{7}$$

where H is an invariable Jacobi matrix of h(x) denoted as:

$$H = \frac{\partial h(x)}{\partial x}|_{x=x_0} \tag{8}$$

After FDIA, the measurement z will be:

$$z = h(x) + e + \alpha \tag{9}$$

where α is the attack vector. The bad data detector (BDD) module examines difference between real value z and estimated $\hat{z}$. If difference exceeds threshold value τ, i.e.,

$$|\hat{z} - z| > \tau \tag{10}$$

the false data will be detected.

B. FDIAs TAXONOMY

Only limited number of studies mentioned on FDIAs taxonomy. Most current survey research only broadly classifies the cyber-physical threats into several categories by the target of the attacks: software, hardware, communication stack, implementation of protocols. Ashok *et al.* classified threats into timing-based attacks that flood the communication network with packets; integrity attacks that corrupt the data; reply attacks that hijack the packets in transit of PMU and PDC (power distribution center) [41]. Other research work classifies the FDIAs in defense mechanism and attack strategy, or in cyber-side and physical side. However, their taxonomy is not adequate for all types of FDIAs.

Moreover, many researches focus on model-based paradigm that concentrates on modelling the intelligent, coordinated attacks in the past few years. However, with the large volumes of data derived from more and more complicated CPS, the cost of modelling tends to be higher comparing to the efficiency and performance that get improved [3]. Therefore, cybersecurity is intertwined with big data, thus cyber-physical threats (CPTs), especially FDIAs are proposed to study and to be classified in a data-based model for the future CPS. This aim is achieved through comprehensively studying all current FDIAs in both cyber and physical layer, and classifying FDI-CPTs in a data centric paradigm.

V. DATA CENTRIC PARADIGM OF FDIAs

Data centric paradigm is the trend to analyze the FDIAs with respect to attacks on data. Nowadays, smart grid CPS have been collecting massive amounts of domain-specific information, such as PMUs. Given large volumes of data generated in real time, it is necessary to shift from traditional model-based paradigm that only considers the attack models to the data centric paradigm for data analytics. This comes from the fact that model-based methods have limitation of performance and efficiency improvement when processing big data from CPS. Besides, fast development of hardware, such as graphic processing unit (GPU) provides itself as a high-performance tool to accelerate big data analytical tasks.

Data centric paradigm aims to analyze the FDIAs in cyber-physical layers, and the integration of the two layers.

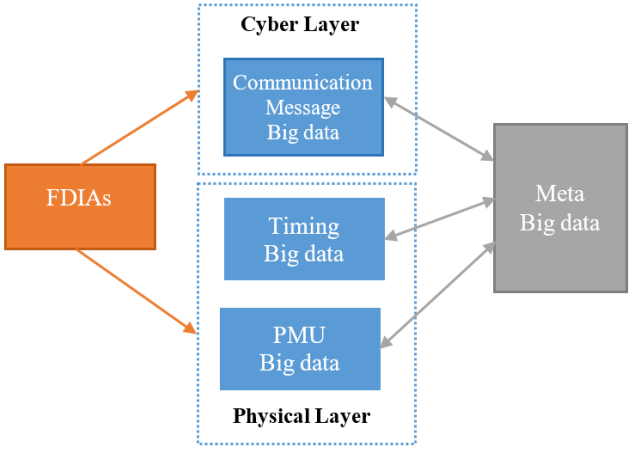

FIGURE 1. Proposed Data Centric Paradigm FDIAs taxonomy (Delete AMI Data).

Data attacks attempt to insert, alter, or delete data or control commands in both layers. One consequence is misleading the smart grid to make wrong decisions. Figure 1 shows the proposed big data centric paradigm to classify FDIAs. This taxonomy contains three categories: 1) big data in cyber layer, including communication message big data. They are the data transported through network protocols, such as Modbus/TCP, DNP3/TCP, IEC 61850, etc.; 2) big data in physical layer, including PMU and Timing big data. They are the data that affect the physical model of smart grids; 3) big data across both layers. They are Meta data for sophisticated FDIAs, which integrate two or more diverse data sets.

A. NETWORK COMMUNICATION MESSAGE BIG DATA

FDIAs in the cyber-layer are attacks that intrude into the communication network. They include several attacks by jamming the network communication by false data, flooding packets in the network. One result is password and authentication failure. Another result of FDIAs is the Denial of Service(DoS) attack on both sensor data and control data, so that availability of the devices is lost. In home area networks, the number of connections for smart meters is limited. The FDIAs spoof this identity, and flood large amounts of bad data to reach the connection limit to result in denial of service. Then smart meters are out of network, FDIAs can launch the next steps of attack to inject false data to control center.

B. PMU AND TIMING BIG DATA

FDIAs in the physical-layer are attacks that intrude into the smart grid with the knowledge of configuration of power system. The mainly source comes from stealthy false data injection from PMU. After false data is injected into the physical model of the smart grid, the state estimation is corrupted. It is proposed to classify FDIAs according to the data source: PMU and time stamp data. One example of timing stamp attack (TSA) happens when the adversary injects the false data into GPS signal and disrupt PMUs time synchronization.

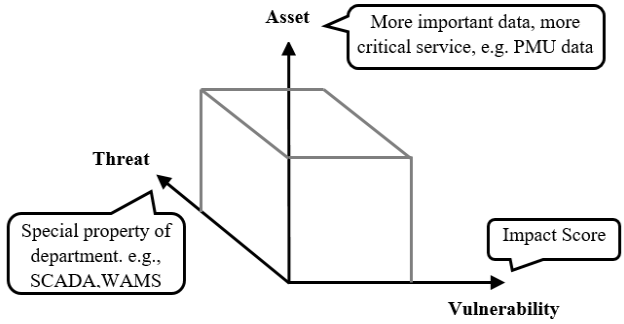

FIGURE 2. Quantitative FDIAs Evaluation.

The consequence is the transmission line fault, voltage instabilities and event location disturbance.

C. META BIG DATA

FDIAs in a more sophisticated type integrate cyber-layer and physical layer. It may require an automated mechanism to mingle multiple highly diverse datasets. In big data approach, we can use ontology-based semantic analysis to find the scheme of relationships between concepts for FDIAs [42]. This type of FDIAs can be analyzed by a hybrid approach using attack graphs [43]. An attack graph is a succinct representation of all paths in a system. In this graph, the final node is the attacker's goal. Some FDIAs are sophisticated through a chain of steps to achieve adversary's goal across cyber-physical layers.

D. QUANTITATIVE EVALUATION

There are many FDI cyber-physical threats, how to quantitatively evaluate them can give us a guidance about the severity of FDIAs. Through this approach, we can find the most critical FDIAs in smart grids. Current risk analysis model contains attack trees, CRAMM [44]. It is a method that can quantitatively evaluate risks through an analysis model. Evaluation of FDIAs can be proposed to use the risk analysis model in CRAMM shown in Figure 2. This risk analysis involves the identification and assessment of three aspects. They are values of assets, levels of threats and vulnerabilities. Final evaluation can be determined as a product of threat, vulnerability, and asset values:

$$\text{Risk} = \text{Asset} \times \text{Threat} \times \text{Vulnerability} \tag{11}$$

Values of assets will evaluate the importance of the data. A scoring system with metrics is proposed here, which is inspired from Common Vulnerability Scoring System (CVSS). This score considers of several metrics for FDIAs integrity attack as shown in Table 1.

VI. PROPOSED MARGIN SETTING ALGORITHM

Margin setting algorithm is a novel data analytical approach based on machine learning. It is the first work to apply MSA to mitigate FDIAs. MSA can learn from the data and recognize patterns, and perform anomaly detection on data

TABLE 1. FDIA vulnerability impact score metrics.

Metrics	Values
Attack vector	network, adjacent, local and physical
Attack complexity	low or high
Privileges required	none, low or high
Integrity	low or high

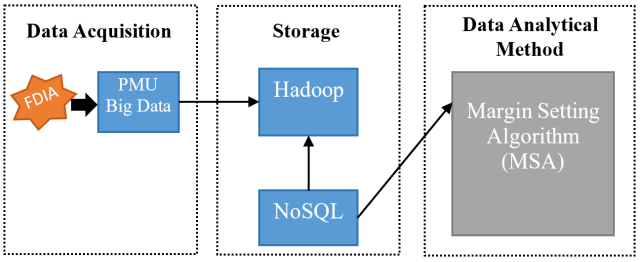

FIGURE 3. Proposed Data Centric Paradigm FDIAs taxonomy.

in CPSs, such as smart grids. Decision boundaries of MSA are hyperspheres called prototypes. It is defined as a center-radius form,

$$\mathrm{G} = \left\{\left(\omega_i, R_i, C_p\right), (i = 1, 2, \ldots, N)\right\} \tag{12}$$

where ω_i is the center of G, R_i is the radius of G, $\mathrm{C_p}$ is the class label. p is the total number of classes for classification. N is the number of prototypes belonging to class C_p.

Figure 3 shows the data flow using MSA data analytical method to detect FDIAs: (a) Data acquisition: large volumes of PMU data is collected. FDIAs are injected adversary into PMU data. (b) Data storage: in this stage, big data are loaded using the Hadoop approach [45]. Hadoop can provide resilient storage and big data sets processing in a distributed computing environment. NoSQL may be used to provide mechanisms for the retrieval of data that is not in the traditional relational database format [46]. (c) Data analytics: in this stage, data are retrieved from database server. There are mainly two tasks for this phase: 1) FDIA detection. FDIAs can be detected by MSA data analytical method based on the time series historical PMU data; 2) FDIA can be predicted in the future if they match the similar pattern that we have learned from MSA. The MSA algorithm is explained below using the flowchart in Figure 4. Online PMU data is gathered from PDC as the input of the MSA algorithm. MSA build initial classification boundaries called prototypes. Then the prototypes are trained by MSA to generate the optimal prototypes as the output. The output can classify the abnormal data and normal data. Abnormal data are results from FDIAs.

VII. TEST RESULTS

In this section, the performance of the proposed MSA is demonstrated by comparing it with another two state of the art machine learning data analytical methods - SVM and ANN. Extensive experiments are conducted on both the simulation

Algorithm 1 MSA Cyber-Physical Attack Detector

Notation: Unif [0,1]: random numbers from [0, 1] space with uniform distribution.
Prototype: $G_i = \left(\omega_i, R_i, C_p\right)(i = 1, 2, \ldots, N)$, ω_i is the center, R_i is the radius, C_p is the class label. N is the number of prototypes belonging to class C_p.
$MF^n_{G_i}$: maximum fitness of the n$^{\text{th}}$ mutation.
θ: θ-percent margin $\theta = 0$;
MQ: maximum number of generation MQ $= 20$;
MW: maximum number of mutation MW $= 20$;
M : Number of mutation M $= 0$;
Q : Number of generation. Q $= 0$;
Input: PMU time series data including n features to build a training set S with k samples, where $S = \{(x_1, \ldots, x_m)$. $x_i(1 \leq k \leq m)$ is n-dimension $(n \geq 2)$ vector. Training sets labels $C_p(p = 1, 2)$ is associated with each training data sample x_k indicating two classes, abnormal $C_1 = -1$ and $C_2 = 1$ (normal).
Initialize:
Set $\theta \leftarrow 0$, MQ $\leftarrow 20$, MW $\leftarrow 20$, M $\leftarrow 0$, Q $\leftarrow 0$;
While $S \neq \emptyset$ or Q $<$ MQ
Normalize S into [0,1] space. Randomly generate N n-dimensional data sample $\omega_i \in$ Unif[0, 1].
While $MF^n_{G_i} > MF^{n+1}_{G_i}$ **or** $M < MW$

1) Build prototypes for training set including abnormal and normal class. $G_i = \left(\omega_i, R_i, C_p\right)(i = 1, 2, \ldots, h < N)$ for each class C_p. Suppose there are h prototypes for abnormal class C_1, (N-h) prototypes for normal class C_2.
2) Center ω_i with class C_p is decided by the minimum Euclidean distance d_i from each ω_i to x_k,
$$d_i = min|\,|\omega_i - x_k|\,|. \tag{13}$$
If x_k is with label C_p, then ω_i is the center of class C_p.
3) Radius R_i with class C_p for each ω_i:
$$R_i = min|\,|\omega_i - x_k|\,|. \tag{14}$$
If $p = 1$, x_k is with a class label C_p when $p = 2$.
If $p = 2$, x_k is with a class label C_p when $p = 1$.
4) Calculate fitness for G_i for each class C_p separately. The fitness of G_i is denoted as F_{G_i}. Its value is the number of data samples falling inside of G_i (a hypersphere) geometrically. The largest F_{G_i} is detonated as MF_{G_i}.
5) Mutation. Select a center ω'_i of prototype G_i of class C_p to mutate. Calculate f_p:
$$f_p = \frac{F_{G_i}}{\sum_1^h F_{G_i}}. \tag{15}$$
If i in ω'_i satisfy the following and $\zeta \in$ Unif[0, 1]:
$$\sum\nolimits_{\xi=1}^{i-1} f_\xi < \zeta \leq \sum\nolimits_{\xi=1}^{i} f_\xi. \tag{16}$$
Mutate ω'_i to its neighbor area. The mutated centers are:
$$\omega'_i + \delta. \tag{17}$$

Algorithm 1 *(Continued.)* MSA Cyber-Physical Attack Detector

Where $\delta = \varepsilon\alpha U$. ε is random sign symbol $\{-1,1\}$. $\alpha \in$ Unif[0, 1]. U is the maximum perturbation:

$$U = \begin{cases} \omega_k & \omega_k < 0.5 \\ 1 - \omega_k & otherwise. \end{cases} \quad (18)$$

6) M ← M + 1
end while
7) Update training set. Store the optimal prototype G_i^o with $MF_{G_i}^n$, and radius $R_{i,o}$. Remove all data samples falling inside of all prototypes of the current generation n, denoted as G_i^n geometrically. The radius of G_i^n is $R_{i,n}$:

$$R_{i,n} = (1 - 0.01\theta)R_{i,n} \quad (19)$$

Store the reduced training set S'. Where $0 \leq \theta < 100$.
8) S ← S′
9) Q ← Q + 1
end while
Output: Total optimal prototypes G_i^o for each generation for each class label C_p:

$$G_{C_p} = \cup_{t=1}^{Q} G_i^o|_{t,C_p}. \quad (20)$$

The prototypes G_{C_p} geometrically classify the abnormal data, i.e., spoofed data, and normal data.

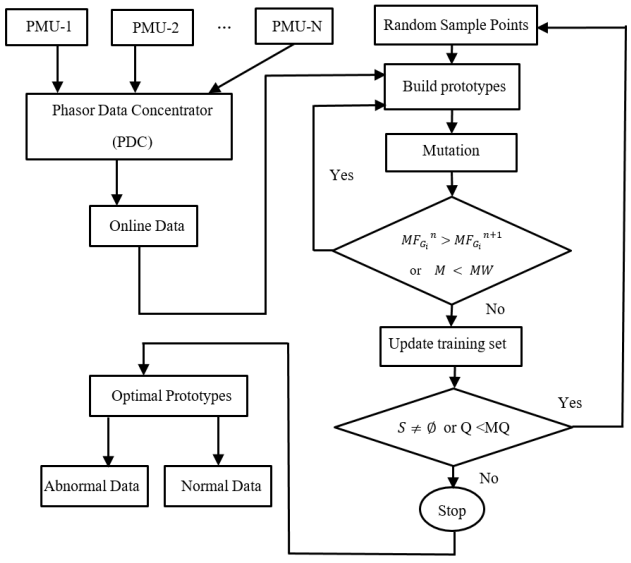

FIGURE 4. MSA Cyber-Physical Attacker Detector Flowchart.

PMU data sets and real-world PMU data sets. Simulation data sets are generated from a six-bus power system employing MATLAB/Simulink in the multi-area WAMS network based on synchrophasors data.

A. SYSTEM DESCRIPTION

The principle of a multi-area WAMS network scenario based on synchrophasors data with the aid of a broadband communication network is described in this section. The system consists mainly of two layers (physical power layer and cyber communication layer) as shown in Figure 5. First, the electrical power system layer, which consists of line-line 208V generating station with 50-kW output rated power, a wind-based power renewable source of 24-kW rated power, 3-power transformers (T1-T3) linking the different parts of the electrical system, 3-short transmission lines (T.L.1-T.L.3), 6-buses (B1-B6), 11- circuit breakers (CB1-CB11) and 2-loads each of 30-kW. Secondly, the WAMS communication layer consists of 3-PMUs, locating at generation and load buses, and one phasor data concentrator (1-PDC) which collects the data received from the remote PMUs. The PDC performs protocol conversion from IEEE C37.118 to several common power system protocols suitable for analysis and control actions in the control center. A satellite GPS synchronization device is used to enable all PMU measurements and date collection for all buses at same instant.

TABLE 2. Simulation parameters.

Parameter	Specification
Rated voltage and frequency	208 Vrms (ph-to-ph), 60 Hz
Generation station rated power	50 kW
Wind generator rated power	24kW
Total load rated power	60 kW
No. of loads	2 (at B2, B3)
No. of transformers (Y/Y)	3 (208V/11 kV)
No. of transmission lines	3 (each as short T.L.)
No. of buses	6
PMU message report rate	60 msg./sec.
Sampling rate	10 kHz

B. VALIDATION SETUP

A software-based validation setup was constructed to validate the proposed approach in WAMS for smart grid applications. The simulation is performed in MATLAB/Simulink 2016a programming environment on a desktop computer with Intel Core i7-4790, 3.6-GHz CPU, 64-bit Windows 7 Enterprise operating system. Figure 6 shows the Simulink model of a six-bus power system incorporated in WAMS environment.

The simulation parameters are shown in Tables 2, 3. As shown in Table 2, bus 1, to which a 50 kW generator is connected, is the swing bus. Bus 2, to which a 24 kW wind power generation and local load are connected, is a voltage-controlled bus. Buses 3-6 are load busses. With current PMU technology, the sampling rate of PMUs can reach as fast as 48 samples per cycle, i.e., 2880 samples per second. Such a high sampling speed, however, is not feasible in this study as the proposed estimators need to process the transmitted data and perform the computation within the time frame. PMU measurements are sampled at 10 kHz in this study. A total of 100,000 data sample measurements are collected from each PMU to investigate the proposed MSA performance.

TABLE 3. Simplified π model line parameters in the software simulation.

Bus-to-Bus	Length km	R (Ω)	X (Ω)
4-5	10	0.1273	3.5200
4-6	15	0.1910	5.2799
5-6	20	0.2546	7.0399

TABLE 4. FDIA detection performance of simulated data sets for playback attack.

Accuracy (in %)	SVM	ANN	MSA
PMU1	99.679	99.694	99.704
PMU2	99.789	99.799	99.805
PMU3	99.824	99.830	99.833

The performance evaluation of MSA is compared with SVM and ANN. The default parameters are chosen. In MSA, the maximum generation MW = 20 and Maximum mutation MQ = 20. In ANN, a feed forward back-propagation network is implemented. The maximum number of epochs to train is set to 10, performance goal is set to 0, and learning rate is 0.01. In SVM, radial basis function is used for kernel type. Other parameters are default values in LibSVM [47]. All results are averaged from ten repeated experiments.

C. ATTACK MODEL

Data points can be dropped for loss of internet connection. Some large disturbance may be caused by short circuits on transmission lines, large or sudden loss of generation or load, transmission line trip and reclosing actions. All zero magnitude, angles and frequencies are pre-processed so that this case will not affect the performance of the proposed method. In a large cyber-physical system, differentiating normal and abnormal data can be overwhelming since cyber-physical attacks can be very complicated and may lead system to behave naturally. To model FDI scenario and evaluate the performance of our proposed algorithm, it is assumed that the attacker can take control of a subset of PMU readings. We consider two cases when attackers have limited knowledge of the power network. They are spoof playback and time attack [49], [50]. For some knowledgeable attackers, they may launch complicated attack by injecting data that matches the patterns of the normal events. This situation is beyond the scope of this paper, since it is rare for large-scale power networks. Suppose the total time we collected the data is T, two FDI scenarios are constructed as follows:

1) PLAYBACK ATTACK

The initial T/2 time of the data is played back in reverse to produce the latter T/2 time data. Data is collected based on time. The three attributes: magnitude, angle and frequency are recorded. Time t and t + 1 denote the time that two data samples collected sequentially.

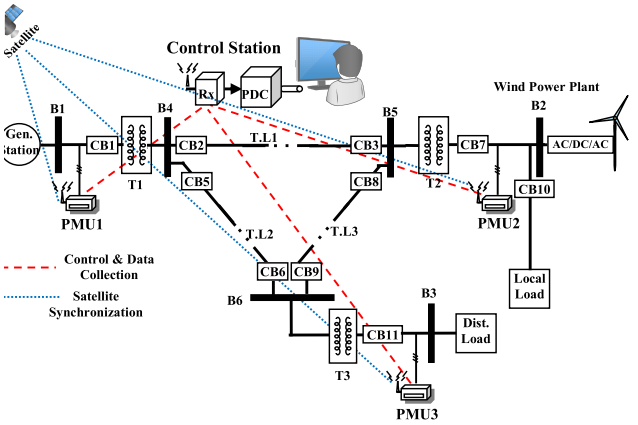

FIGURE 5. The proposed WAMS network involving wind renewable power plant.

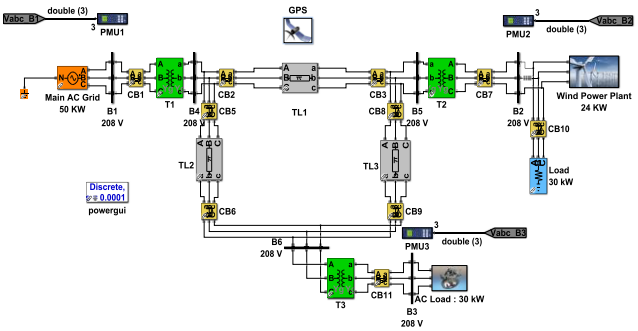

FIGURE 6. Simulink model schematic of the proposed system.

2) TIME ATTACK

The final T/2 of data is re-sampled using different rates of time. The rates of time vary from very slow (a factor of 4 slower than real-time) to near real time (a factor of 7/6 slower than real-time).

D. SIMULATION DATA SETS

The PMU data collected from the Simulink model of WAMS network. The false data ratio for PMU1, PMU2 and PMU3 under playback attack are: 0.334%, 0.196% and 0.086% for totally 100,000 samples.

The experimental results for FDI playback attack is reported in Table 4. It can be seen that the proposed MSA outperforms ANN and SVM in terms of detection accuracy for all three PMU stations. Moreover, when the false data ratio increases, the performance of all three algorithms, presenting a decreasing trend as shown in Figure 7. Besides, when the false data ratio increases, MSA tends to yield better performance comparing to ANN and SVM. For example, when false data ratio is 0.086% in PMU3 station, MSA has a slightly better performance, i.e., 0.003% better than ANN and 0.009% than SVM. However, in PMU1 station case with 0.334% false data ratio, MSA is 0.01% and 0.025% better than ANN and SVM, respectively.

The results for FDI time attack are reported in Table 5. It is observed that MSA gives higher detection accuracy in

TABLE 5. FDIA detection performance of simulated data sets for time attack.

Accuracy (in %)	Factor of 7/6 slower			Factor of 3/2 slower			Factor of 2 slower			Factor of 4 slower		
	SVM	ANN	MSA	SVM	ANN	MSA	SVM	ANN	MSA	SVM	ANN	MSA
PMU1	97.634	97.649	97.659	97.831	97.846	97.856	98.926	98.941	98.952	99.532	99.547	99.556
PMU2	97.724	97.734	97.741	97.922	97.933	97.940	99.022	99.032	99.039	99.689	99.699	99.705
PMU3	97.897	97.903	97.905	98.097	98.102	98.106	99.135	99.141	99.145	99.749	99.755	99.758

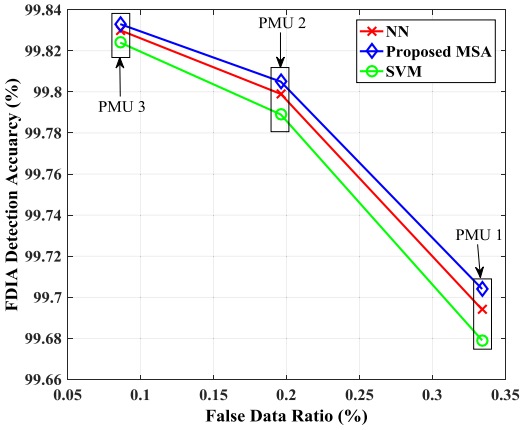

FIGURE 7. FDIA Detection Performance in terms of false data ratio among proposed MSA, ANN, and SVM using simulated PMU data from the Simulink model of WAMS network under playback attack.

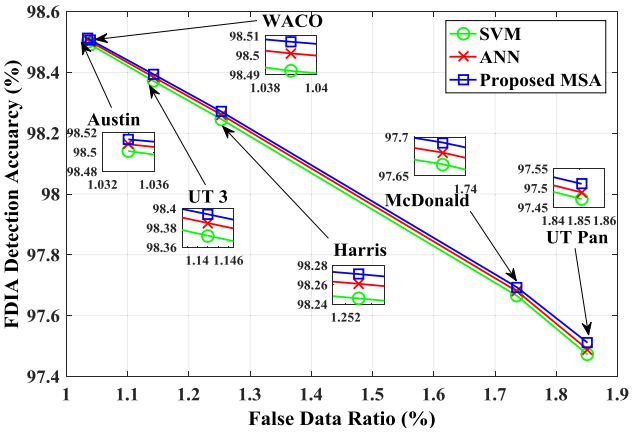

FIGURE 8. FDIA Detection Performance in terms of false data ratio among proposed MSA, ANN, and SVM using experimental data sets under playback attack.

TABLE 6. FDIA detection performance of experimental data sets for playback attack.

PMU Stations	False Data Ratio	Accuracy (in %)		
		SVM	ANN	MSA
McDonald	1.736%	97.665	97.680	97.693
Harris	1.253%	98.246	98.261	98.271
UT Pan	1.851%	97.472	97.489	97.511
UT 3	1.142%	98.372	98.385	98.394
Austin	1.034%	98.492	98.501	98.507
WACO	1.039%	98.501	98.508	98.513

TABLE 7. Experimental data sets false data ratio for time attack.

PMU Stations	Time Attack			
	Factor of 7/6 slower	Factor of 3/2 slower	Factor of 2 slower	Factor of 4 slower
McDonald	1.784%	1.718%	1.678%	1.521%
Harris	1.631%	1.585%	1.545%	1.457%
UT Pan	1.350%	1.311%	1.271%	1.233%
UT 3	1.209%	1.166%	1.126%	1.015%
Austin	1.051%	1.003%	0.963%	1.013%
WACO	0.663%	0.623%	0.583%	0.658%

all scenarios we consider in this experiment. When the re-sampling rates gets slower, ranging from a factor of 7/6 to 4 slower than the real time, the detection accuracy performance increases. For example, MSA performance raises from 97.634% to 99.557% when the re-sample rate is changed from 7/6 (near real time) to a factor of 4 slower than real time. In addition, note that when the false data ratio is the lowest for PMU3 station, the MSA can only yield a slightly better performance than ANN, i.e., 0.002%, 0.004%, 0.004% and 0.003% better in the four time-attack scenarios. However, when the false data ratio is highest in PMU1 station, the MSA yields an average 0.01% better than ANN. This indicates that the larger the false data ratio, the better performance MSA compared with SVM and ANN.

E. EXPERIMENTAL DATA SETS

PMU data collected from Texas Synchrophasor Network are chosen for our experiment to detect FDIAs using a real-PMU dataset [48]. The sampling rate for this data is 30-Hz, so only low frequency (<15Hz) oscillations can be analyzed. An hourly PMU data consisting of 108,000 data points are used for analysis. Each data point includes three signals: voltage magnitude, angle and frequency. All measurements are taken under the customer-level (120-V). There are six PMU stations operating in the network. The name labels of these stations are: McDonald, Harris, UT Pan, UT 3, Austin, WACO. Experiments are conducted under playback attack and time attack.

The performances of FDI playback attack are presented in Table 6 and Figure 8. It is observed that the proposed

TABLE 8. FDIA detection performance of experimental data sets for time attack.

Accuracy (in %)	Factor of 7/6 slower			Factor of 3/2 slower			Factor of 2 slower			Factor of 4 slower		
	SVM	ANN	MSA	SVM	ANN	MSA	SVM	ANN	MSA	SVM	ANN	MSA
McDonald	96.543	96.559	96.574	97.473	97.489	97.505	98.042	98.057	98.072	98.765	98.782	98.791
Harris	96.729	96.744	96.757	97.661	97.675	97.688	98.228	98.243	98.255	98.951	98.965	98.976
UT Pan	97.031	97.045	97.055	97.962	97.976	97.986	98.531	98.543	98.555	99.253	99.267	99.275
UT 3	97.156	97.168	97.177	98.087	98.101	98.108	98.655	98.666	98.676	99.378	99.388	99.398
Austin	97.288	97.301	97.306	98.219	98.231	98.237	98.787	98.799	98.805	99.512	99.518	99.528
WACO	97.522	97.531	97.535	98.453	98.462	98.466	99.021	99.031	99.035	99.744	99.753	99.758

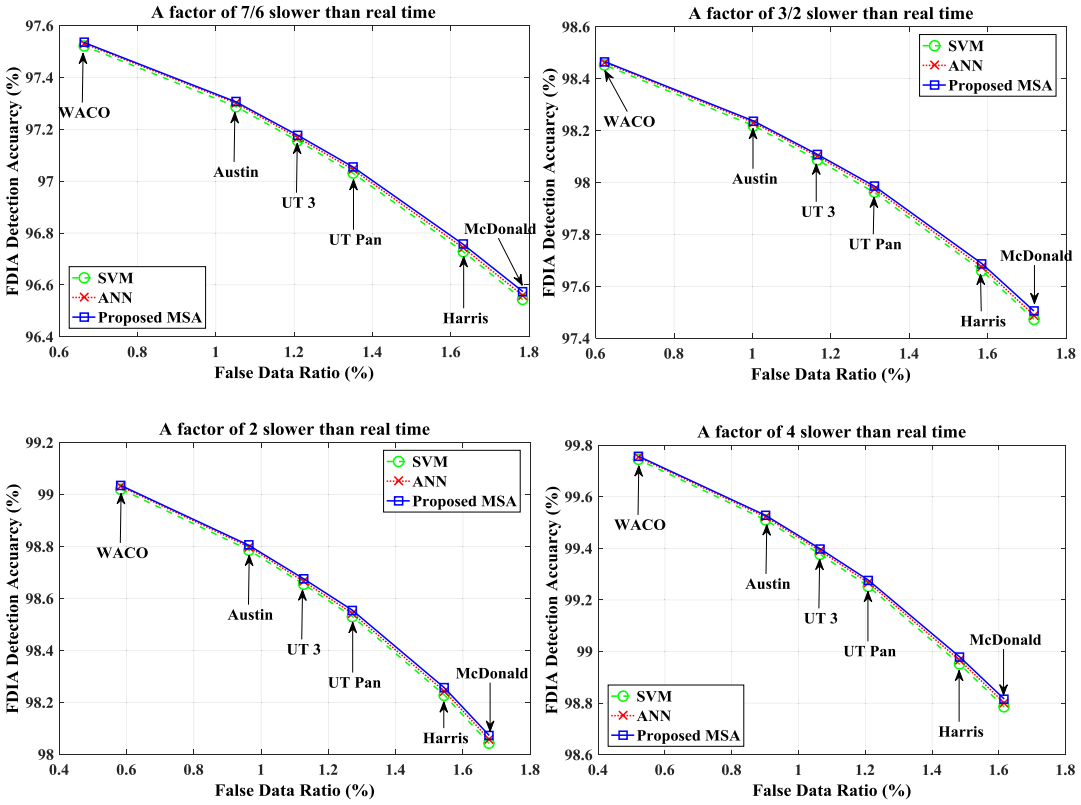

FIGURE 9. FDIA Detection Performance in terms of false data ratio among proposed MSA, ANN, and SVM using experimental data sets under time attack for a factor of (7/6, 3/2, 2, 4) slower than real time sample rate.

MSA achieves better performance than SVM and ANN in all six different PMU stations. The FDIA detection performance presents the same trend as the experiment with simulation data sets. When the false data ratio increases, all the algorithms, SVM, ANN and the proposed MSA experience a linear decline with respect to their performances to detect FDIAs. Besides, MSA has much better performance when the false data ratio is higher. For instance, MSA achieves 0.02% and 0.06% higher detection accuracy than ANN and SVM in UT Pan station with the highest false data ratio of 1.851% among six PMU stations.

Table 8 and Figure 9 illustrates the performances of MSA with ANN and SVM for four FDI time attack scenario using the experimental sets. The false data ratios for time attack scenarios are presented in Table 7. The four scenarios are simulated by changing the different re-sampling rate, i.e. factor of 7/6, 3/2, 2 and 4 slower than real time in the final 30 minutes of the time series data sets. Overall, the MSA outperforms NN and SVM in all four different scenarios for all six PMU stations. When the false data ratio decreases, the performance tends to increase. Take an example in the factor of 7/6 scenario, the McDonald PMU yields 96.543% accuracy when false data ratio is 1.784%. However, the performance goes up to 97.522% for 0.663% false data ratio in WACO PMU. Figures 9 presents a decline trend for FDI time attack detection accuracy when the false data ratio increases.

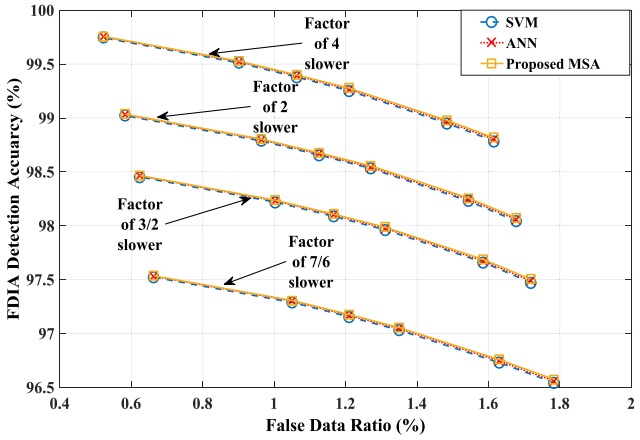

FIGURE 10. FDIA Detection Performance in terms of false data ratio among proposed MSA, ANN, and SVM using experimental data sets under time attack for a factor of (7/6, 3/2, 2, 4) slower than real time sample rate.

Although the false data ratio varies in six PMU stations for four different cases, it is worth to note that the re-sampling rate gets slower, the performance gets higher in detection accuracy. For example, it can be seen from Figure 10 that when false data ratio is 0.8%, factor of 7/6 scenario outputs about 97.45% detection accuracy, while factor of 3/2, 2 and 4 scenarios present higher detection accuracies of around 98.36%, 98.85% and 99.6%.

VIII. CONCLUSION

In this paper, a novel data analytical method employing MSA was demonstrated to defend against false data injection cyber-physical attack in smart grids. This is the first work to use MSA to detect FDIAs. Besides, a data centric paradigm is proposed to analyze the FDIAs in a big data scenario. The performance of the proposed MSA method through both simulation and experimental data sets has been investigated. Simulation data sets were generated from a MATLAB/Simulink model employing a six-bus power system in WAMS network. The real-world data sets were entered based on hourly data from Texas Synchrophasor Network. The experimental results demonstrated that the proposed MSA achieved better accuracy with minimum error than traditional SVM and ANN algorithms to detect FDIAs. In the future, more sophisticated FIDAs cyber-physical attacks on PMU data will be investigated based on the proposed method. Additionally, the proposed MSA will be employed to process various enormous amount of data in real time, handling the big data challenge in the future smart grid cyber-physical systems.

REFERENCES

[1] J. P. Farwell and R. Rohozinski, "Stuxnet and the future of cyber war," *Survival*, vol. 53, no. 1, pp. 23–40, 2011.
[2] *Analysis of the Cyber Attack on the Ukrainian Power Grid*, document, (E-ISAC), Mar. 2016.
[3] Y. Z. Lun *et al.* (2016). "Cyber-physical systems security: A systematic mapping study." [Online]. Available: https://arxiv.org/abs/1605.09641
[4] R. Deng, G. Xiao, R. Lu, H. Liang, and A. V. Vasilakos, "False data injection on state estimation in power systems—Attacks, impacts, and defense: A survey," *IEEE Trans. Ind. Inf.*, vol. 13, no. 2, pp. 411–423, Apr. 2017, doi: 10.1109/TII.2016.2614396.
[5] J. Hao, R. J. Piechocki, D. Kaleshi, W. H. Chin, and Z. Fan, "Sparse malicious false data injection attacks and defense mechanisms in smart grids," *IEEE Trans. Ind. Inf.*, vol. 11, no. 5, pp. 1–12, Oct. 2015.
[6] D. Mah, P. Hills, V. O. K. Li, and R. Balme, *Smart Grid Applications and Developments*. London, U.K.: Springer, 2014.
[7] S. Wallace *et al.*, "Big data analytics on a smart grid: Mining PMU data for event and anomaly detection," in *Big Data, Principles and Paradigms*. San Mateo, CA, USA: Morgan Kaufmann, 2016.
[8] J. Wei and G. J. Mendis, "A deep learning-based cyber-physical strategy to mitigate false data injection attack in smart grids," in *Proc. Joint Workshop Cyber-Phys. Secur. Resilience Smart Grids (CPSR-SG)*, Apr. 2016, pp. 1–6.
[9] T. Ma, F. Wang, J. Cheng, Y. Yu, and X. Chen, "A hybrid spectral clustering and deep neural network ensemble algorithm for intrusion detection in sensor networks," *Sensors*, vol. 16, no. 10, p. 1701, 2016.
[10] W. Hardy, L. Chen, S. Hou, Y. Ye, and X. Li, "DL4MD: A deep learning framework for intelligent malware detection," in *Proc. Int. Conf. Data Mining (DMIN)*, 2016, p. 61.
[11] M. J. Kang and J. W. Kang, "Intrusion detection system using DNN for in-vehicle network security," *PLoS ONE*, vol. 11, no. 6, p. e0155781, 2016.
[12] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on the smart grid," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 645–658, Dec. 2011.
[13] S. Li, Y. Yilmaz, and X. Wang, "Quickest detection of false data injection attack in wide-area smart grids," *IEEE Trans. Smart Grid*, vol. 6, no. 6, pp. 2725–2735, Nov. 2015.
[14] R. C. B. Hink, J. M. Beaver, M. A. Buckner, T. Morris, U. Adhikari, and S. Pan, "Machine learning for power system disturbance and cyber-attack discrimination," in *Proc. 7th Int. Symp. Resilient Control Syst. (ISRCS)*, 2014, pp. 1–8.
[15] 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.
[16] J. Landford *et al.*, "Fast sequence component analysis for attack detection in synchrophasor networks," in *Proc. 5th Int. Conf. Smart Cities Green ICT Syst. (SmartGreens)*, Rome, Italy, 2016, p. 268.
[17] J. Fu, H. Caulfield, and C. Glenn, "Primitive attempt to turn images into percepts," *Int. J. Mach. Learn.*, vol. 5, no. 6, pp. 963–970, 2014.
[18] Y. Wang, R. Adhmai, J. Fu, and H. Al-Ghaib, "A novel supervised learning algorithm for salt-and-pepper noise detection," *Int. J. Mach. Learn.*, vol. 6, no. 4, pp. 687–697, 2015.
[19] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proc. 16th ACM Conf. Comput. Commun. Secur.*, Chicago, IL, USA, Nov. 2009, pp. 21–32.
[20] J. Kim and L. Tong, "On topology attack of a smart grid: Undetectable attacks and countermeasures," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1294–1305, Jul. 2013.
[21] L. Jia, J. Kim, R. J. Thomas, and L. Tong, "Impact of data quality on real-time locational marginal price," *IEEE Trans. Power Syst.*, vol. 29, no. 2, pp. 627–636, Mar. 2014.
[22] M. A. Rahman, E. Al-Shaer, and R. Kavasseri, "Impact analysis of topology poisoning attacks on economic operation of the smart power grid," in *Proc. IEEE 34th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Jul. 2014, pp. 649–659.
[23] M. Esmalifalak, H. Nguyen, R. Zheng, and Z. Han, "Stealth false data injection using independent component analysis in smart grid," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Oct. 2011, pp. 244–248.
[24] 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.
[25] M. A. Rahman and H. Mohsenian-Rad, "False data injection attacks against nonlinear state estimation in smart power grids," in *Proc. Power Energy Soc. Gen. Meet. (PES)*, 2013, pp. 1–5.
[26] G. Chaojun, P. Jirutitijaroen, and M. Motani, "Detecting False Data Injection Attacks in AC State Estimation," *IEEE Trans. Smart Grid*, vol. 6, no. 5, pp. 2476–2483, Sep. 2015.
[27] J. Lin, W. Yu, X. Yang, G. Xu, and W. Zhao, "On false data injection attacks against distributed energy routing in smart grid," in *Proc. IEEE/ACM 3rd Int. Conf. Cyber-Phys. Syst.*, Apr. 2012, pp. 183–192.

[28] J. Chen and A. Abur, "Placement of PMUs to enable bad data detection in state estimation," *IEEE Trans. Power Syst.*, vol. 21, no. 4, pp. 1608–1615, Nov. 2006.

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

[30] S. Gong, Z. Zhang, H. Li, and A. D. Dimitrovski. (Jan. 2012). "Time stamp attack in smart grid: Physical mechanism and damage analysis." [Online]. Available: https://arxiv.org/abs/1201.2578

[31] X. Liu, Z. Li, X. Liu, and Z. Li, "Masking transmission line outages via false data injection attacks," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 7, pp. 1592–1602, Jul. 2016.

[32] 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.

[33] X. Liu, Z. Bao, D. Lu, and Z. Li, "Modeling of local false data injection attacks with reduced network information," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 1686–1696, Jul. 2015.

[34] V. Kekatos, G. B. Giannakis, and R. Baldick, "Grid topology identification using electricity prices," in *Proc. PES Gen. Meeting Conf. Expo.*, 2014, pp. 1–5.

[35] J. Valenzuela, J. Wang, and N. Bissinger, "Real-time intrusion detection in power system operations," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1052–1062, May 2013.

[36] Z. H. Yu and W. L. Chin, "Blind false data injection attack using PCA approximation method in smart grid," *IEEE Trans. Smart Grid*, vol. 6, no. 3, pp. 1219–1226, May 2015.

[37] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[38] B. Ayhan, M.-Y. Chow, and M.-H. Song, "Multiple discriminant analysis and neural-network-based monolith and partition fault-detection schemes for broken rotor bar in induction motors," *IEEE Trans. Ind. Electron.*, vol. 53, no. 4, pp. 1298–1308, Jan. 2006.

[39] Z. Guan, N. Sun, Y. Xu, and T. Yang, "A comprehensive survey of false data injection in smart grid," *Int. J. Wireless Mobile Comput.*, vol. 8, no. 1, pp. 27–33, 2015.

[40] A. Ashok, A. Hahn, and M. Govindarasu, "Cyber-physical security of wide-area monitoring, protection and control in a smart grid environment," *J. Adv. Res.*, vol. 5, no. 4, pp. 481–489, 2014.

[41] T. Stepanova, A. Pechenkin, and D. Lavrova, "Ontology-based big data approach to automated penetration testing of large-scale heterogeneous systems," in *Proc. 8th Int. Conf. Secur. Inf. Netw.*, 2015, pp. 142–149.

[42] P. J. Hawrylak, M. Haney, M. Papa, and J. Hale, "Using hybrid attack graphs to model cyber-physical attacks in the smart grid," in *Proc. 5th Int. Symp. Resilient Control Syst. (ISRCS)*, 2012, pp. 161–164.

[43] I. E. Fray, "A comparative study of risk assessment methods, MEHARI & CRAMM with a new formal model of risk assessment (FoMRA) in information systems," in *Proc. CISIM*, 2012, pp. 428–442.

[44] M. Sivakumar, C. Sadagopan, and M. Baskaran, "Wireless sensor network to cyber physical systems: Addressing mobility challenges for energy efficient data aggregation using dynamic nodes," *Sensor Lett.*, vol. 14, no. 8, pp. 852–857, 2016.

[45] Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, "Health-CPS: Healthcare cyber-physical system assisted by cloud and big data," *IEEE Syst. J.*, vol. 11, no. 1, pp. 88–95, Mar. 2017.

[46] *LibSVM*. Accessed: Jan. 15, 2017. [Online]. Available: https://www.csie.ntu.edu.tw/~cjlin/libsvm/

[47] A. Allen, M. Singh, E. Muljadi, and S. Santoso, "PMU data event detection: A user guide for power engineers," Nat. Renew. Energy Lab. , Golden, CO, USA, Tech. Rep. NREL/TP-5D00-61664, 2014.

[48] Z. Zhang, S. Gong, A. D. Dimitrovski, and H. Li, "Time synchronization attack in smart grid: Impact and analysis," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 87–98, Mar. 2013.

[49] C. Bonebrake and L. O. Neil, "Attacks on GPS time reliability," *IEEE Security Privacy*, vol. 12, no. 3, pp. 82–84, Mar. 2014.

YI WANG (S'14–M'15) received the B.S. degree in information management and information system and the M.S. degree in computer science from the Wuhan University of Science and Technology, Wuhan, China, in 2006 and 2009, respectively, and the Ph.D. degree in computer engineering from The University of Alabama in Huntsville, Huntsville, AL, USA, in 2015. Upon completing of the Ph.D. degree, he joined the Department of Electrical and Computer Engineering, Manhattan College, Riverdale, NY, USA, as an Assistant Professor. His research interests include image processing, pattern recognition, and cybersecurity and cyber-physical systems.

MAHMOUD M. AMIN (S'09–M'12–SM'16) received the B.Sc. and M.Sc. degrees in electrical engineering from Helwan University, Egypt, in 2003 and 2008, respectively, and the Ph.D. degree from Florida International University (FIU), Miami, FL, USA, in 2012. From 2003 to 2008, he was a teaching assistant with several academic institutions in Egypt. Since 2007, he joined the Electronic Research Institute (ERI), Egypt, as a research assistant (RA). From 2009 to 2012 he was RA with FIU. Since 2012, he has been an Assistant Professor with Manhattan College, Riverdale, NY, USA, and a Researcher with ERI. He has published over 40 articles in professional journals and conference proceedings. His current research interests include power electronics in sustainable energy systems and smart grid security. Prof. Amin was a recipient of the Graduate Student Association Scholarly Forum Prize Paper Award 2010 as well as the IEEE PES GM 2010 Graduate Student Poster Contest Award.

JIAN FU (M'05–SM'15) received the Ph.D. degree in computer science and engineering from The University of Alabama in Huntsville in 2005, and the M.S. degrees in computer sciences and physics from Alabama A&M University in 1998 and 1996, respectively. He is currently a Professor with the Department of Electrical Engineering and Computer Science, Alabama A&M University. His research interests are pattern recognition, image processing, and computer vision. Prof. Fu was a recipient of the Jim Zimmerman Award from AAMURI in 2008. His research papers were selected in SPIE Milestones Series in 1997 and listed as the most widely read article in SPIE library in remote sensing in 2010.

HEBA B. MOUSSA (S'10) received the B.Sc. degree in electrical engineering from Helwan University, Cairo, Egypt, in 2007, and the M.Sc. degree in electrical engineering from Manhattan College, Riverdale, NY, USA, in 2014. She is currently pursuing the Ph.D. degree in electrical engineering with the City College of New York, New York, NY, USA. Since 2014, she has been an Adjunct Instructor with Manhattan College. Her current research interests include wireless communication, computer networks, cyber-physical systems, and smart grid security.