Detection and Identification of Cyber and Physical Attacks on Distribution Power Grids with PVs: An Online High-Dimensional Data-driven Approach

Fangyu Li, Rui Xie, Bowen Yang, Lulu Guo, Ping Ma, Jianjun Shi, Jin Ye, WenZhan Song

***Abstract*—Cyber and physical attacks threaten the security of distribution power grids. The emerging renewable energy sources such as photovoltaics (PVs) introduce new potential vulnerabilities. Based on the electric waveform data measured by waveform sensors in the distribution power networks, in this paper, we propose a novel high-dimensional data-driven cyber physical attack detection and identification approach (HCADI). Firstly, we analyze the cyber and physical attack impacts (including cyber attacks on the solar inverter causing unusual harmonics) on electric waveforms in distribution power grids. Then, we construct a high dimensional streaming data feature matrix based on signal analysis of multiple sensors in the network. Next, we propose a novel mechanism including leverage score based attack detection and binary matrix factorization based attack diagnosis. By leveraging the data structure and binary coding, our HCADI approach does not need the training stage for both detection and the root cause diagnosis, which is needed for machine learning/deep learning-based methods. To the best of our knowledge, it is the first attempt to use raw electrical waveform data to detect and identify the power electronics cyber/physical attacks in distribution power grids with PVs.**

***Index Terms*—Attack Diagnosis, Distribution Power Grids, Solar Inverter, Leverage Score, Binary Matrix Factorization.**

I. Introduction

POWER electronics converters are becoming more vulnerable to cyber/physical attacks due to their growing penetration in Internet of Things (IoT) enabled applications including the smart grids [1]. Due to the lack of cyber awareness in power electronics community [1], it becomes more urgent to develop cyber/physical attack detection and identification strategies for power electronics converters in many safety-critical applications since these malicious attacks can lead to a catastrophic failure and substantial economic loss if not detected in the early stage.

Manuscript received XXX, 2019; revised XXX, 2019; accepted XXX, 2019; online XXX, 2019. (*Corresponding author: Fangyu Li.)*

The research is partially supported by NSF-1663709, NSF DMS-1222718, NIH R01GM113242, NIH R01GM122080, NSF DMS-1438957, NSF DMS-1440038, NSF DMS-1925066, NSF ECCS-1946057 and Southern Company.

F. Li, B. Yang, L. Guo, J. Ye and W.Z. Song are with Center for Cyber-Physical Systems, University of Georgia, Athens, GA 30602, USA (e-mail: fangyu.li@uga.edu, bowen.yang@uga.edu, lulu.guo@uga.edu, jin.ye@uga.edu, wsong@uga.edu).

R. Xie is with Department of Statistics and Data Science, University of Central Florida, Orlando, FL 32816 USA (e-mail: rui.xie@ucf.edu).

P. Ma is with Department of Statistics, University of Georgia, Athens, GA 30602, USA (e-mail: pingma@uga.edu).

J. Shi is with H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA (e-mail: jianjun.shi@isye.gatech.edu).

Attacks are studied in applications which are intensively dependent on power electronics converters, including power grids with voltage support devices [2], distribution systems with solar farms [3], with power electronics driven HVAC (Heating, ventilation, and air conditioning) systems [4], and microgrids [5], [6]. However, they mostly focus on either analyzing or detecting cyber attacks affecting grid level stability, functionality and operational costs. In [7], a model-based method was developed to detect data integrity attacks on automation generation control of transmission systems. In [3], a physical-law based detection was developed to detect false data attacks which attempt to reduce the output power of solar energy in distribution systems. In [4], a secure information flow framework was developed for 118-bus distribution network with power electronics driven HVAC system. In [8], a physics-based, cooperative mechanism was developed to detect stealthy attacks in DC microgrids with a number of DC-DC converters, which can bypass most of observer-based detection methods. In [9], a physics-based framework to detect false-data injection attacks in DC microgrids with a number of DC-DC converters. While power electronics converters are included in their cyber security monitoring frameworks, they are designed to detect one particular type of grid-level cyber attacks but those on the devices (power electronics converters) are not studied. Thus, their cyber security framework is not applied to (1) cyber attack detection on power electronics converters, which might affect the performance of power electronics converters; and (2) the root cause identification when a variety of attacks occur.

As smart grids are evolving to complex cyber-physical systems, there might be a variety of cyber and physical attacks including coordinated attacks. Data-driven approaches are gaining increased attention in recent years due to the advancements in sensing and computing technologies [10]–[13]. They show great potentials in detecting and identifying complicated cyber and physical attacks. The data sources for these purposes include solar power plants, wind turbines, hydroelectric plants, marine turbines, phasor measurement unit (PMU), microgrids, fault detectors, smart meters, smart appliances and electric vehicles [14]. In [15], A data-driven time-frequency analysis was proposed to detect the dynamic load altering attacks. In [16], a data-driven hidden structure semi-supervised machine was proposed to implement the power distribution network attack detection. In [17], multistream data flow was employed to build effective and efficient attack-resilient solutions against the cyber threats targeting electric grids. In [18], a data-driven and low-sparsity false data injection

attack strategy against smart grid was investigated. In [19], a machine learning solution was proposed to identify the false data injection attacks on transmission lines of smart grids. Existing data-driven approaches, however, have not yet been used to detect cyber and physical attacks in the device level (power electronics converters). Thus, a data-driven methodology is needed to detect and identify a variety of cyber and physical attacks, that negatively affect both the power electronics converter (such as solar inverter) and other critical components (such as relays and generators) in power grids.

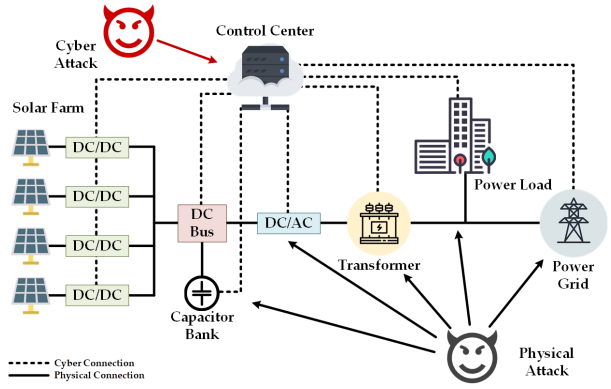


Fig. 1: Cyber and physical attacks threaten the security of the distribution power grid with a solar farm.

Fig. 1 shows the diagram of distribution power grid with solar farms. The solar farm is physically connected to the distribution grid through the DC/DC, DC/AC converters and the grid connected transformers. Then the major components and control center are connected through cyber networks. The attacks in red are the potential cyber attacks on the control center (such as data integrity attacks on inverter feedback/control signals or some abnormal delay injected to the control signal), which will compromise the performance of the grid and power electronics converters; the attacks in black are the physical attacks to the power grid facilities (such as single and multiple phase short circuit faults of transformers/generators, abnormal load/capacitor bank cut-off). We need to detect and diagnose both cyber and physical attacks to the distribution power grids with PV systems. Compared with the traditional hardware protection, for example relays, we aim to develop a comprehensive data-driven solution to adaptively, efficiently and accurately monitor the power grid with more and more various power electronics devices, protecting the system from cyber and physical attacks, even subtle ones.

In this paper, we propose to develop a data-driven methodology to detect and identify the cyber and physical attacks on distribution power grid with solar farms. We firstly analyze and simulate the impacts of cyber and physical attacks on electrical waveforms in distribution power grid with solar farms. Then we propose a high-dimensional data-driven cyber physical attack detection and identification approach (HCADI) based on feature extraction, anomaly detection and matrix factorization. Finally, we test and evaluate our HCADI approach in a MATLAB model of distribution power grid with solar farms in different cyber and physical attack scenarios. The contributions and innovations of our work are:

1) We develop a novel HCADI framework that effectively detects and identifies both cyber and physical attacks on the grid level and device level (power electronics converters) in distribution power grid with solar farms.
2) We propose an innovative waveform data based signal processing and online statistics associated method to detect the cyber and physical attacks. The proposed data-driven method detects attacks based on the dependence structure of multi-dimensional streaming sensor data.
3) We propose to use the feature distribution of latent variables based on matrix factorization to diagnose the cyber attack types. The proposed attack diagnosis approach doesn't require a training stage, which is superior to machine learning/deep learning based methods in terms of computational efficiency.

II. Cyber Physical Modeling and Control of PVs

In general, solar farms include four major components: solar panels, first stage DC/DC converter, second stage DC/AC inverter, and the grid connected transformer. Here, we analyze, detect, and identify cyber attacks on the solar inverter, causing the unusual harmonics and then poor power quality in distribution systems.

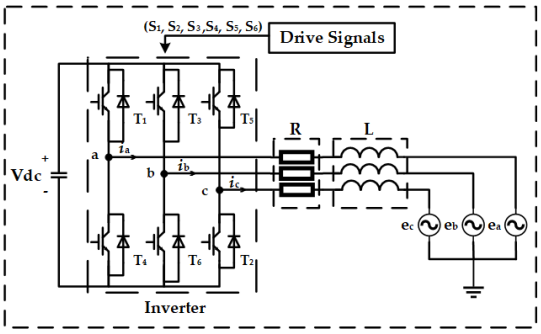


Fig. 2: Main circuit topology of the inverter.$S_1 \sim S_6$ denote the switching signals.

The main topology of the solar inverter is shown in Fig. 2, and the generalized physical model of DC/AC solar inverter is derived as follows:

$$\begin{cases} \frac{di_a}{dt} = -\frac{R}{L}i_a - \frac{e_a}{L} + \frac{V_{dc}}{3L}(2s_a - s_b - s_c), \\ \frac{di_b}{dt} = -\frac{R}{L}i_b - \frac{e_b}{L} + \frac{V_{dc}}{3L}(-s_a + 2s_b - s_c), \\ \frac{di_c}{dt} = -\frac{R}{L}i_c - \frac{e_c}{L} + \frac{V_{dc}}{3L}(-s_a - s_b + 2s_c), \end{cases} \tag{1}$$

where the control signals s_a, s_b, s_c will be sent from the cyber system and are defined as:

$$s_a = \begin{cases} 1 \ (S_1 = 1, S_4 = 0) \\ 0 \ (S_1 = 0, S_4 = 1) \end{cases}, \quad s_b = \begin{cases} 1 \ (S_3 = 1, S_6 = 0) \\ 0 \ (S_3 = 0, S_6 = 1) \end{cases}, \quad s_c = \begin{cases} 1 \ (S_5 = 1, S_2 = 0) \\ 0 \ (S_5 = 0, S_2 = 1) \end{cases}, \tag{2}$$

i_a, i_b, i_c are the currents of each phase, e_a, e_b, e_c are the phase voltages of the power grid and L and R are the inverter inductance and resistance, V_{dc} is the DC bus voltage after the

first stage DC/DC converter. To simplify the analysis process, direct-quadrature-zero (DQZ) transformation is adopted [20]:

$$\begin{cases} \frac{di_d}{dt} = -\frac{1}{L}e_d + \frac{1}{L}V_{dc}S_d + \omega i_q - \frac{R}{L}i_d, \\ \frac{di_q}{dt} = -\frac{1}{L}e_q + \frac{1}{L}V_{dc}S_q - \omega i_d - \frac{R}{L}i_q, \end{cases} \tag{3}$$

where ω is the electric angular frequency, and the control input is transformed as S_d and S_q, and other variables are all corresponding to the $d-$ and $q-$ axis components.

Fig. 3 shows the control diagram of the solar farm system, and the cyber attack on the solar inverter is denoted red, which injects a long time delay to the solar inverter control signals.

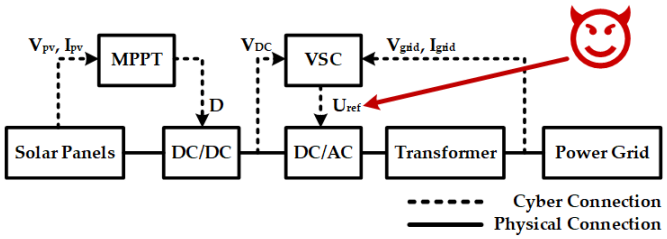


Fig. 3: Control diagram of the solar farm system.

III. Methodology

A. Problem setup

Suppose we have sequential observations at k sensors, $x_1(t), x_2(t), \ldots, x_k(t)$. Before the emergence of the attack, the observations are normal conditions following the electronic model $\eta(\cdot)$ described in Section II with a random noise, i.e., $\epsilon(t) \sim N(0, \sigma^2)$. When an attack occurs, the observations at different sensors will capture it but with different responses. We assume the attack signal is causal, i.e., $\eta(t) = 0, \forall t < 0$.

For the ith sensor, the observed data can be expressed as:

$$\begin{aligned} x_i(t) &= \eta(t) + \epsilon_i(t), && t = 1, 2, \ldots, \tau, \\ x_i(t) &= \alpha_i \eta^*(t - \tau_i) + \epsilon_i(t), && t = \tau + 1, \tau + 2, \ldots, \end{aligned} \tag{4}$$

where α_i is the unknown amplitude of the change at the ith sensor. A sensor data matrix X can be constructed, $X(t) = [x_1(t), \ldots, x_k(t)], X \in \mathbb{R}^{k \times n}$, n is the data sample number.

B. Feature Extraction

The measured normal waveform data are typically sinusoidal functions for AC power grids. In order to extract the waveform information with impacts from different attacks, we need to extract signal features first, such as the health index in [21] and signal quality measurements in [22].

1) Instantaneous Features: The waveforms of voltage and current signals $\mathbf{V} = [V_1, V_2, \ldots, V_N]^T$, $\mathbf{I} = [I_1, I_2, \ldots, I_N]^T$ are measured from a network with size N the nodal, where depending on the number of phases at node i, V_i and I_i can be row vectors of size 1, 2 or 3. In order to characterize the waveform properties, we adopt instantaneous properties from:

$$s_c(t) = s(t) + j\mathcal{H}\{s(t)\} = A(t)e^{j\psi(t)}, \tag{5}$$

where $s(t)$ is the real signal, $s_c(t)$ is the complex expression, $A(t)$ is the instantaneous amplitude (IA) (envelope), $\psi(t)$ is the instantaneous phase(IP), $\mathcal{H}$ is the Hilbert transform as:

$$\mathcal{H}\{s(t)\} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{s(\tau)}{t - \tau} \mathrm{d}\tau. \tag{6}$$

Thus, for a three phase current $I_n = [I_{nA}, I_{nB}, I_{nC}]^T$, where $I_{nA} = A_{I_{nA}} e^{j\psi_{I_{nA}}(t)}$. Similarly, V_n can be expressed as $V_n = [V_{nA}, V_{nB}, V_{nC}]^T$, where $V_{nA} = A_{V_{nA}} e^{j\psi_{V_{nA}}(t)}$.

2) Differences: The changes of the nodal DC voltages and branch currents can be expressed as:

$$\Delta V_n = V_n(t) - V_n(t - w), \tag{7}$$

$$\Delta I_{np} = I_{np}(t) - I_{np}(t - w), \tag{8}$$

where, w is the analysis window size, n and p denote two arbitrary neighboring nodes.

For the AC voltages and currents, considering the instantaneous features in Section III-B1, the differences can be expressed as:

$$\Delta V_{nA} = A_{V_{nA}}(t) - A_{V_{nA}}(t - w), \tag{9}$$

$$\Delta I_{npA} = A_{I_{npA}}(t) - A_{I_{npA}}(t - w), \tag{10}$$

where only Phase A is showed, Phases B and C have the similar expressions. In the normal distribution power grids, the voltages and currents should be stable. If abnormal changes happen to ΔV_n and ΔI_{np}, an event can be detected based on certain thresholding methods [23], [24]. Here, instead of directly using the difference, we treat it as one dimension of the high-dimensional detection metrics matrix.

3) Unbalance: In the AC power grids, single, two or even three phase issues could exist. The waveforms of Phases A,B, and C allow a relatively straightforward phase unbalance characterization based on direct comparisons of phase signal attributes. Based on the IA defined in Eq. (5), we define the current unbalance characterization functions I_α, I_β, and I_γ as:

$$I_{n\alpha} = \frac{1}{3} \sum_{i \neq j}^{i,j \in \{A,B,C\}} (A_{I_{ni}} - A_{I_{nj}})^2. \tag{11}$$

$$I_{n\beta} = \frac{I_{max} - I_{min}}{I_{max}}, \tag{12}$$

$$I_{n\gamma} = \sum_{i \neq j}^{i,j \in \{A,B,C\}} \Gamma(A_{I_{ni}}, A_{I_{nj}}), \tag{13}$$

where, $I_{n,max} = \max\{A_{I_{nA}}, A_{I_{nB}}, A_{I_{nC}}\}$ and $I_{n,min} = \min\{A_{I_{nA}}, A_{I_{nB}}, A_{I_{nC}}\}$, Γ denotes a thresholding function to measure the difference. If I_β is not zero, there exists unbalance among the three phases. Then we use I_γ to determine the how many phases are affected, and I_α to measure the absolute changes. Similarly, we can also get V_α, V_β and V_γ.

C. High-dimensional Data Matrix Construction

In Section III-A, we build a data matrix X in general, and $X \in \mathbb{R}^{k\times n}$ with n being the number of data samples and k being the number of sensors. Because of the feature extraction in Section III-B, the streaming data from one node on an AC distributed power grid become high dimensional instead of just one. For a DC node the feature matrix is $[V, I, \Delta V, \Delta I]^T$, while for an AC node the feature matrix is $[A_{V_A}, A_{V_B}, A_{V_C}, A_{I_A}, A_{I_B}, A_{I_C}, \Delta V_A, \Delta V_B, \Delta V_C, \Delta I_A, \Delta I_B, \Delta I_C, V_\alpha, V_\beta, V_\gamma, I_\alpha, I_\beta, I_\gamma]^T$. Note that for a node, the current measurements could be more than one as the connections with other nodes can be multiple. So the listed matrices are still general formats. In reality, the feature matrices will have even larger dimensions. In short, the detection data matrix combines all the feature matrices from all the nodes in the networks, and will be used for attack detection and root cause diagnosis.Thanks to the recent growth in wireless communication, the monitoring data even over a large area can be efficiently collected [25].

D. Statistical Leverage Score for Attack Detection

After constructing the high-dimensional data matrix in Section III-C, we apply a novel data-driven anomaly detection method based on the feature matrix $Y \in \mathbb{R}^{n\times m}$ with n time sample points and m features to detect the attack emergence. Since the observed signal are recorded along time and has multiple dimensions, the multidimensional time series model will be a natural choice for modeling such data. To the best of our knowledge, traditional attack detection and identification methods, including the distributed attack detection and the adaptive fault detection methods, did not fully utilize the feature contained in the multidimensional time series model [26]–[28], while ignoring those temporal or cross-correlated features may lead to biased detection results. The *vector autoregressive* (VAR) model as a fundamental model in the study of multivariate time series are considered to capture the dynamics of the signals [29], [30]. On time domain, each data point is correlated its previous values; on spatial domain, each sensor records one dimensional data and data from different dimension are correlated to each other spatially, i.e., cross-correlated. The temporal dependency in the time domain of the signal calls for the time series modeling, where the autoregressive model can effective capture such features. The autoregressive (AR) model is a time series model that the observations are specified to be depended on its own previous values and a stochastic term. On the other hand, the similarity and dissimilarity among spatial features that we extract makes the VAR model, which is an extension of AR model and incorporates the multidimensional cross-correlation, suitable for the analyzing the high-dimensional data. In general, we consider the parameter estimation of a m-dimensional vector autoregressive model of order p, i.e., VAR(p),

$$\begin{aligned}\mathbf{y}_t' &= \mathbf{y}_{t-1}'\Phi_1' + \mathbf{y}_{t-2}'\Phi_2' + \cdots \mathbf{y}_{t-p}'\Phi_p' + \mathbf{e}_t' \\ &= \mathbf{x}_t'\mathbf{B} + \mathbf{e}_t',\end{aligned} \tag{14}$$

where $\mathbf{y}_t$ is the k-dimensional response vector observed at time point $t \in \{1, \ldots, n\}$, $\mathbf{B} = [\Phi_1', \Phi_2', \cdots, \Phi_p']'$ is the $mp \times m$ model parameter matrix, $\mathbf{x}_t = (\mathbf{y}_{t-1}', \mathbf{y}_{t-2}', \cdots, \mathbf{y}_{t-p}')'$ is a column vector of previous values of length mp, and $\mathbf{e}_t$ is a sequence of independent and identically distributed (i.i.d.) stochastic random vectors with mean zero and and finite non-singular covariance matrix $\mathbb{E}[\mathbf{e}_t\mathbf{e}_t'] = \mathbf{\Psi}$. The unknown parameter $\mathbf{B}_t$ at time sample t can be estimated as

$$\begin{aligned}\mathbf{B}_t &= \arg\min_{\mathbf{B}} \sum_{t\in\{1,2,\ldots\}} ||\mathbf{y}_t' - \mathbf{x}_t'\mathbf{B}||_2^2, \\ &= \left(\sum_t \mathbf{x}_t\mathbf{x}_t'\right)^{-1}\left(\sum_t \mathbf{x}_t\mathbf{y}_t'\right).\end{aligned} \tag{15}$$

If we define a *Hat Matrix* $P_H = \mathbf{x}_t'\left(\sum_t \mathbf{x}_t\mathbf{x}_t'\right)^{-1}\mathbf{x}_t$, the predicted value can be expressed as $\hat{\mathbf{y}}_t = P_H\mathbf{y}_t$. And, the i-th diagonal element of P_H,

$$\ell_{ii} = \frac{\partial\hat{\mathbf{y}}_i}{\partial\mathbf{y}_i} = \mathbf{x}_i'\left(\sum_t \mathbf{x}_t\mathbf{x}_t'\right)^{-1}\mathbf{x}_i, \tag{16}$$

is the *statistical leverage score* of the i-th observation, which has been used to regression diagnostics to quantify the influential observations, and data dependent subsampling [30]–[32]. Alternatively, the leverage score can be expressed as

$$\ell_{ii} = ||\mathbf{u}_i||_2^2, \tag{17}$$

where $\mathbf{u}_i'$ comes from the rows of the orthogonal matrix U, which can be calculated from the left singular matrix of the singular value decomposition (SVD) on matrix $[\mathbf{x}_1 \ \cdots \ \mathbf{x}_t]$ [32], [33]. By calculating the leverage score based on the VAR model, we can identify the highly influential data points that change the system status rather than the random noise, which can effectively reduce the false alarms in the attack detection. High-leverage score data points have the extreme or outlying behaviours such that they can effectively identify the anomaly values of the underlying observations.

For streaming feature signals $Y(t)$, finding the orthogonal basis to calculate the leverage score can be implemented in an online fashion through streaming principal component analysis (PCA) [34], [35]. The implementation of streaming leverage score calculation is discussed in [30].

E. Binary Matrix Factorization to Diagnose Attack Root Causes

After detecting the influential data points as possible attacks using the statistical leverage score, we propose a data-driven matrix factorization method for attack root causes diagnosis. Matrix factorization techniques such as Non-negative Matrix Factorization (NMF) or SVD consist of an important family of data analysis tools that yield a compact representation of signals as linear combinations of a small number of 'basis' referred to as latent variables or states [36]. Attack detection based on result of matrix factorization can be adopted to diagnose the cyber attack root causes [37]. The construction of traditional process monitoring methods based on multivariate statistics neglects the temporal correlation and spacial dependency of latent variables at different sampling times, and those methods also assume latent variables satisfying a particular distribution.

Here we consider to decompose signals into the binary basis and its corresponding weights. The binary basis reveal a unique

binary coding as the latent states to indicate the fault types or the root causes of the attack. By examining the combination of the binary coding, we can effectively and efficiently diagnose the root causes of the attacks. Specifically, if the input signal is a real-valued matrix $Y \in \mathbb{R}^{n\times m}$, we aim to decompose Y into a product of a binary matrix H and a weight matrix W, i.e., $Y \approx HW$. The binary matrix factorization (BMF) method is free from the input signal distribution assumptions, which leads to a data-driven method for attack root cause analysis without a training process.

We implement the following BMF algorithm to exam the attack diagnosis. Given $Y \in \mathbb{R}^{n\times m}$ as the input data matrix, we formulate the BMF as an optimization problem: find $H_1 \in \{0,1\}^{n\times r}$ and $W_1 \in \mathbb{R}^{r\times m}$, such that $Y \approx H_1 W_1$ with $r < m$. Using a metric of the F-norm (Frobenius norm), the general BMF problem takes the form:

$$\begin{aligned} &\min_{H_1,W_1} \ \frac{1}{2}\|Y - H_1W_1\|_F^2, \\ &\text{subject to } H_1 \in \{0,1\}^{n\times r},\ W_1 \in \mathbb{R}^{r\times m}, \end{aligned} \tag{18}$$

which can be solved by enumerating all vertices of the n-dimensional cubic $[0,1]^n$ contained in affine subspace of Y and selecting a maximal independent subset. In summary, the scalable speed up algorithm to find the vertices is:

1) Randomly selecting from candidate vertices, which yields candidate matrices $\{H_1^{(l)}\}_{l=1}^{s}$;
2) Subsequently solving $H_1 = \arg\min_{H_1^{(l)}} \min_{W_1} ||Y - H_1^{(l)}W_1||_F^2$ given the current estimate of W_1;
3) Update the weight estimate by $W_1 = \arg\min_{W_1} ||Y - H_1^{(l)}W_1||_F^2$ given the current estimate of $H_1^{(l)}$;
4) Alternate Steps (2) and (3) until converge.

The convergence analysis of the algorithm can be found in [37].

We introduce the multilayer binary matrix factorization for detailed root cause diagnosis. After the fitst layer BMF, we denote the recovered signals as $\hat{Y} \coloneqq H_1W_1$, and the residuals as $R_1 \coloneqq Y - \hat{Y} = Y - H_1W_1$. Now we can perform the second layer binary matrix factorization as

$$\begin{aligned} &\min_{H_2,W_2} \ \frac{1}{2}\|R_1 - H_2W_2\|_F^2, \\ &\text{subject to } H_2 \in \{0,1\}^{n\times r_2},\ W_2 \in \mathbb{R}^{r_2\times m}. \end{aligned} \tag{19}$$

The rows $H_{1(t,m)}, t \in \{1,\dots,n\}$ of the binary matrix H_1 form the basis elements that indicate the binary coding of the latent states of the signal. The rows $H_{2(t,m)}, t \in \{1,\dots,n\}$ of the binary matrix H_2 contains the detailed elements that indicate the binary coding of the pattern change of the signal. By jointly examining the binary coding of both the H_1 and H_2, we can determine the root causes of the attack through the one-to-one mapping of binary coding and root causes.

IV. Algorithm

Based on the theories introduced in Section III, we propose a online high dimensional data-driven cyber-physcial attack detection and diagnosis algorithm called HCADI, whose workflow is shown in Fig. 4. First, electric waveform data are obtained continuously to construct streaming data. As the streaming data are measured from the sensors in the distribution power networks, the streaming data matrix has high dimensions with AC and DC voltages and currents. Before the features extraction, a typical pre-processing operation filters out the noise interferences and conditions the data if data samples are missing or time stamps are not stable. Using the Eqs. (5) to (13), from the high dimensional data matrix, we build a high dimensional feature matrix, whose dimension is even higher. Based on the leverage score, the abnormal changes in the feature matrix can be detected. Otherwise, if there is no anomaly, the whole system will analyze the next streaming data segmentation. Once an anomaly is detected, we apply the BMF method to identify the attack types based on the binary coding results. The advantage of using an attack detection step before the attack diagnosis is the efficiency, as the diagnosis is more time and computation consuming than the detection.

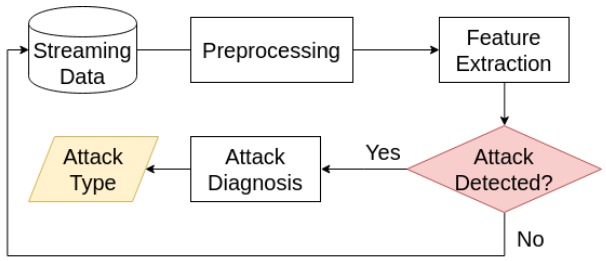


Fig. 4: Workflow of the proposed HCADI system. The attack detection is highlighted with red shadow, and the attack diagnosis result is in yellow.

V. Simulation

A simulation based on a MATLAB Simulink Demo, 400kW Grid-Connected PV Farm Network, is conducted to generate waveforms of some typical fault in small scale power network. The power network topology is shown in Fig. 5.

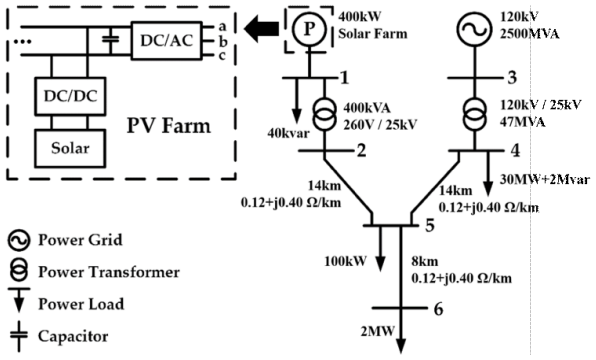


Fig. 5: Simulation topology of a 400 kW Grid-Connected PV Farm Network.

The power grid is modeled as an ideal voltage source with a rate voltage of 120 kV, and connected to the sub-transmission network with a rate voltage of 25 kV through a 47 MVA power transformer. The PV farm includes four solar blocks, each of them connected to the DC bus through a DC/DC converter. And a three phase inverter is adopted to transfer the DC power to the AC. And to match the voltage level of the sub-transmission system, a 400 kVA power transformer is

used to connect the PV farm and the sub-transmission system. Moreover, four linear loads are modeled in the system: 30 MW on Bus 4, denoted the power grid load, 100 kW and 2 MW on Bus 5 and Bus 6, denoted the sub-transmission system loads, and 40 kvar reactive power compensation on Bus 1 as well as a 2 Mvar reactive power compensation on Bus 4, modeled as capacitive power loads. Under normal operation condition, the voltage and current waveforms of Bus 2 are shown in Fig. 6. The sampling frequency is 50k Hz, and 0.5 seconds (s) data are simulated for each scenario, which have 25001 samples. Note that, to clearly illustrate details, we only plot 0.1 s data around the event time in Figs. 6∼11.

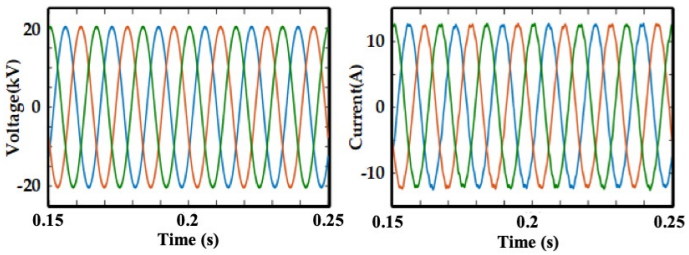


Fig. 6: Normal operation condition waveforms of (left) the voltage and (right) current on Bus 2.

Using the simulation system described above, we simulate typical fault conditions, each of which has featured waveforms:

1) **Physical Attacks:** Short circuit fault is one of the most common physical faults in power systems, which could be caused by human behaviors and natural hazards, such as misoperations, cyber-attacks, storm and lighting. And the outcomes of short circuit fault depend on many factors such as fault location, short fault type and damage severe degree. So four different short circuit faults are simulated.
   a) **Main grid grounded short circuit fault:** A single phase grounded short circuit fault of Bus 4 results in distortion of the voltage and the current. The waveform of Bus 4 is shown in Fig. 7, it is easy to note that this fault causes transient impacts on currents and spike voltage and steady state asymmetric components.

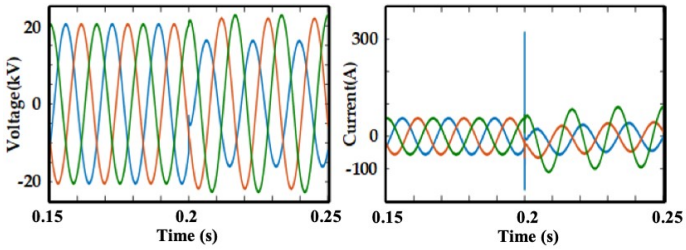


Fig. 7: Main grid single phase grounded short circuit fault waveforms of (left) the voltage and (right) current on Bus 4.

   b) **Solar transformer grounded short circuit fault:** The short circuit faults happen on Bus 2, which can be single phase or double phases. A double phases (phase a and phase b) grounded short circuit fault waveforms of Bus 4 are shown in Fig. 8. Note that the fault current is even more severe than that from the main grid fault described above.

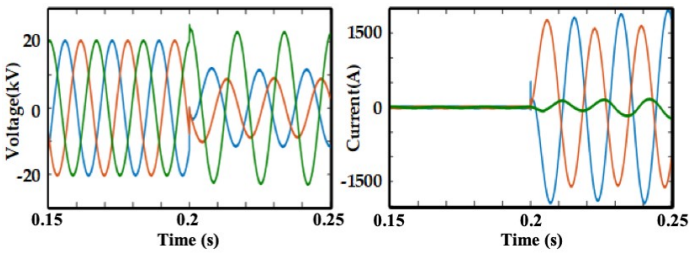


Fig. 8: Solar transformer double phases (phase a and phase b) grounded short circuit fault waveforms of (left) the voltage and (right) current on Bus 2.

2) **Cyber Attacks:**
   a) **Extra reactive power compensation in solar system:** Fig. 9 shows the waveforms of Bus 1 when the PV farm is injected extra reactive power compensation, which is possibly caused by false data injection in the control center. In the simulation model, extra reactive power is modeled as capacitive power load and injected to Bus 1, which could be caused by misoperations and purposeful attacks.

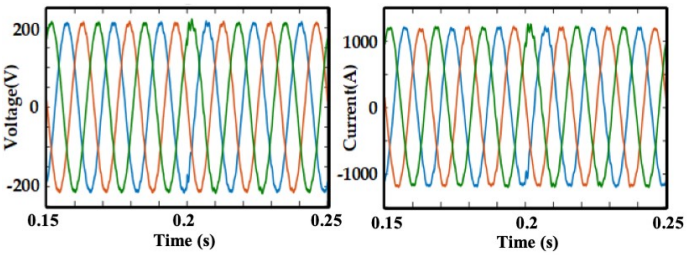


Fig. 9: Extra reactive power compensation in solar system waveforms of (left) the voltage and (right) current on Bus 1.

   b) **PV farm inverter attacked:** The solar inverter hacked situation is simulated. A 1 ms delay is added to the inverter controller signal to simulate the "data integrity" attack [22]. The waveforms of Bus 1 are shown in Fig. 10.

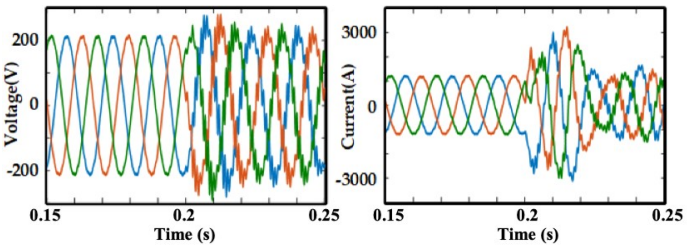


Fig. 10: PV farm inverter attacked waveforms of (left) the voltage and (right) current on Bus 1.

   c) **30MW linear load cut off:** Heavy load cutting off is another common fault in power system which could be caused by the integrity attack to the control center. When heavy load is cut off in a short period,

the power system will generate sever oscillations. The waveforms of Bus 4 are shown in Fig. 11.

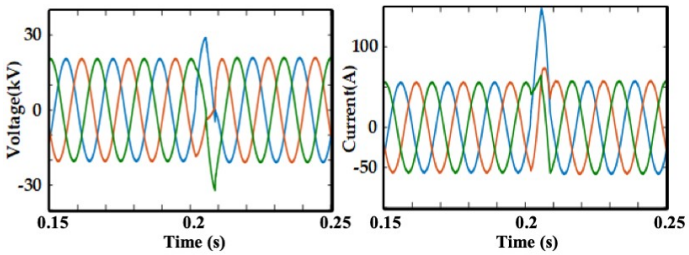


Fig. 11: 30 MW linear load cut off waveforms of (left) the voltage and (right) current on Bus 4.

VI. Evaluation

A. Pre-processing and Feature Extraction

The first step of the proposed algorithm is the normalization. Because our approach is based on matrix structure analysis, the unbalanced amplitudes among different observations will influence the following statistical analysis. Thus, we normalize the data matrix before the feature extraction, and one example of the main grid grounded short circuit fault in Fig. 7 is shown in Fig. 12. Note that, the AC components are normalized according to their IAs, while DC components are based on their maximum and minimum values in the segments. There are 6 nodes (5 AC nodes and 1 DC node) in Fig. 5, so the vectors in data matrix are aligned following the node number.

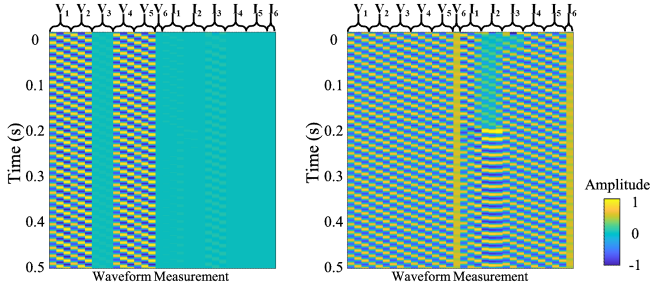


Fig. 12: Data matrix normalization in the situation of main grid grounded short circuit fault. (Left) Raw waveform matrix; (Right) Normalized waveform matrix. Each vector corresponds to one voltage or current waveform, which is either one phase of AC components or one DC component. As there are 5 AC nodes and 1 DC node, the data matrix dimension is 32.

Based on the normalized data matrix, we extract the feature matrix according to Section III-B. Since AC components generate instantaneous features, differences and unbalances, while DC components do not have the unbalance features, the dimension of feature matrix is $32 + 32 + 30 = 94$, shown in Fig. 13. With the sophisticated feature extraction, the latent data structure information is better characterized, and the attack detection robustness can also be improved. Comparing Fig. 12 and Fig 13, it is clear that the feature matrix exhibits more information of the data anomaly than the original data matrix, which is valuable for attack detection and diagnosis.

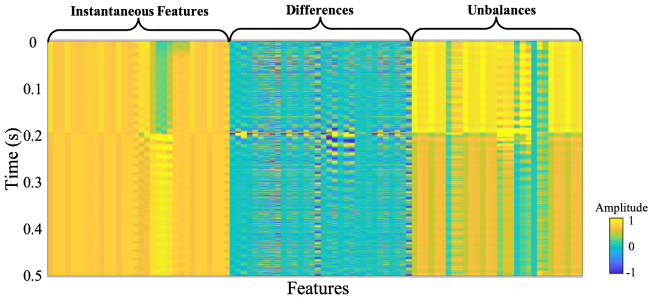


Fig. 13: Feature matrix extracted from the normalized waveform matrix shown in Fig. 12. The total dimension is 94, including 32 columns of instantaneous features, 32 columns of differences, and 30 columns of unbalances.

B. Attack Detection Using Leverage Score

Using the statistical leverage score introduced in Section III-D, we can detect the abnormal changes in the matrix structure. Fig. 14(a) shows the leverage scores extracted from raw data matrix and feature matrix, respectively. As the attack happens at $t = 0.2$ s, both raw data based and feature based leverage scores can highlight the attack appearance. However, the leverage score extracted from the raw data is not robust. Fig. 14(b) shows the leverage scores extracted with 10 dB noises. The attack can still be clearly detected by feature matrix based leverage score, but not by the one based on the raw data matrix. Thus, it is necessary to use the feature matrix as the robustness must be considered.

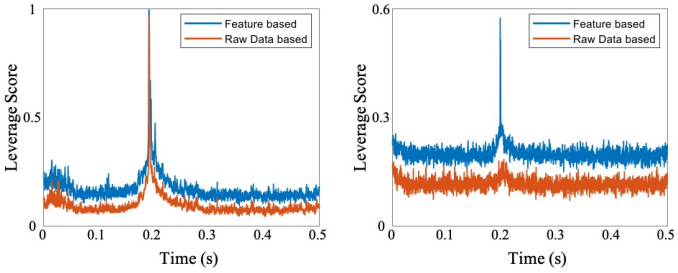


Fig. 14: Attack detection based leverage score on raw data matrix and feature matrix in (a) noise free situation and (b) with 10 dB random noise .

C. Attack Diagnosis Using BMF

As discussed in Section III-E, $H_{1(t,k)}$ and $H_{2(t,k)}$ of different situations can be obtained by BMF. Figs. 15 and 16 demonstrate the first and second layer binary coding results, where black color denotes 1 while white denotes 0. The decomposed binary bases $H_1 s$ and $H_2 s$ illustrate the observed data structures of different distribution power grid operation scenarios. Normal condition shows a different performance compared with the attacked situations, that H_1 is continuous and H_2 has no residues. However, it is difficult to directly distinguish different attacks using original $H_1 s$ and $H_2 s$.

For attack diagnosis, we use both $H_{1(t,k)}$ and $H_{2(t,k)}$ distributions as stated above. In order to visualize the high-dimensional matrices shown in Figs. 15 and 16, a visualization

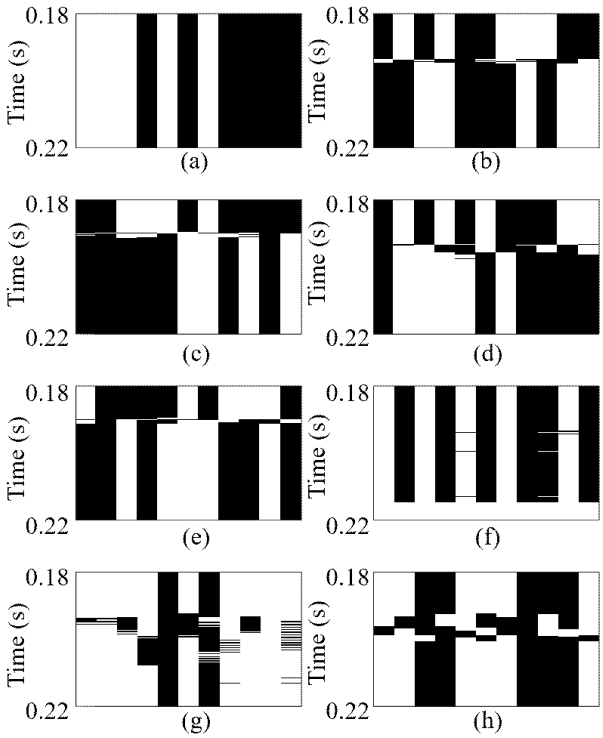


Fig. 15: $H_{1(t,k)}$ of (a) normal condition, main grid (b) single phase (A1) and (c) double phases (A2) grounded short circuit fault, Solar transformer (d) single phase (A3) and (e) double phases (A4) grounded short circuit fault, (f) extra reactive power compensation in solar system (A5), (g) PV farm inverter attacked (A6) and (h) 30MW linear load cut off (A7). Black denotes 1, while white denotes 0.

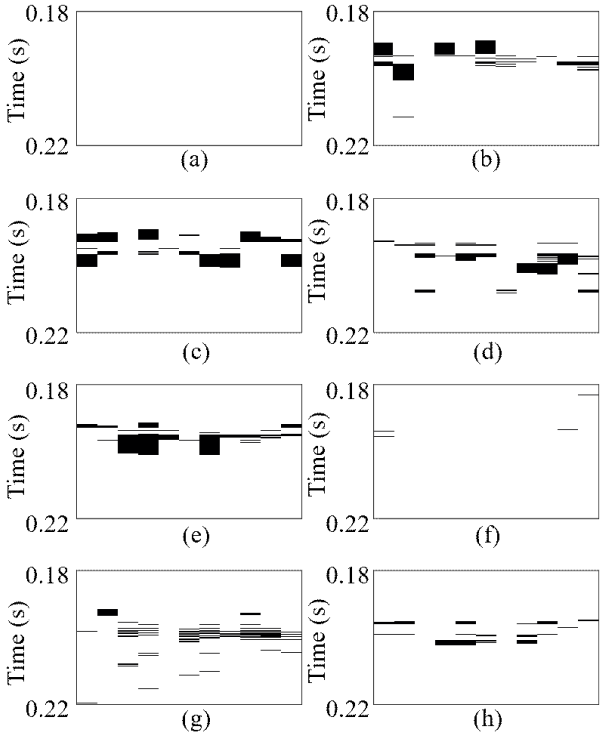


Fig. 16: $H_{2(t,k)}$ corresponding to the $H_{1(t,k)}$ in Fig. 15, respectively. Black denotes 1, while white denotes 0.

method called t-Distributed Stochastic Neighbor Embedding (t-SNE) [38] is adopted. It is a nonlinear dimensionality reduction technique for visualizing high-dimensional data in a low-dimensional space, in our study, two dimensions. The advantage of t-SNE is the "distance-preserving" property [39], which means the Kullback–Leibler divergence and the corresponding Euclidean distance between two clusters are appropriately preserved during the dimensionality reduction process. Fig. 17 shows the 2D visualization results of $H_{1(t,k)}$ and $H_{2(t,k)}$. In Fig. 17(a), most attacks are clustered at different locations, but A6 doesn't have a dense distribution. Thus, the visualization of $H_{2(t,k)}$ in Fig. 17(b) is an important complement to the attack diagnosis. Thanks to the "distance-preserving" property of t-SNE, the well separated clusters in the 2D space are also well separated in the original high dimensions. Therefore, the proposed double layered BMF is promising for attack diagnosis.

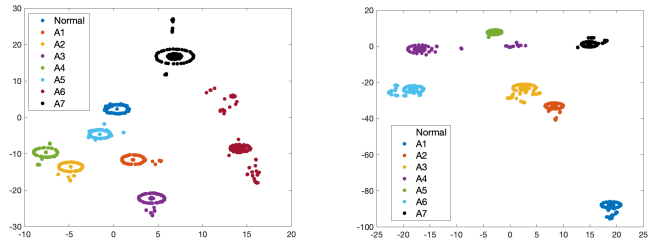


Fig. 17: Visualization of (a) $H_{1(t,k)}$ and (b) $H_{2(t,k)}$ using t-SNE. It is clear that different attack types can be identified. Note that, as the two dimensional space is a latent space, the two axes do not have physical meanings. However, the relative distances among different clusters can be used to measure the similarities of different clusters, indicating the potential hidden relations among different attacks.

VII. Conclusion

Solar farms and other renewable energy sources bring potential attack vulnerabilities to distribution power networks. In this paper, we propose a novel cyber-physical attack detection and diagnosis approach called HCADI based on high dimensional data-driven methods. Features of the streaming waveform data are constructed to be an analysis matrix, which has the inherent data structure. Therefore, the leverage score method can identify the anomaly brought by the attacks. Then based on the binary coding results from BMF, the attack types can be identified. The proposed approach is a data-driven statistical structure analysis without a training stage, making it efficient and implementable in an online real-time style.

References

[1] J. C. Balda, A. Mantooth, R. Blum, and P. Tenti, "Cybersecurity and power electronics: Addressing the security vulnerabilities of the internet of things," *IEEE Power Electronics Magazine*, vol. 4, no. 4, pp. 37–43, 2017.

[2] B. Chen, S. Mashayekh, K. L. Butler-Purry, and D. Kundur, "Impact of cyber attacks on transient stability of smart grids with voltage support devices," in *2013 IEEE Power & Energy Society General Meeting*. IEEE, 2013, pp. 1–5.

[3] Y. Isozaki, S. Yoshizawa, Y. Fujimoto, H. Ishii, I. Ono, T. Onoda, and Y. Hayashi, "Detection of cyber attacks against voltage control in distribution power grids with pvs," *IEEE Transactions on Smart Grid*, vol. 7, no. 4, pp. 1824–1835, 2016.

[4] Y. Cao, K. Davis, and S. Zonouz, "A framework of smart and secure power electronics driven hvac thermal inertia in distributed power systems," in *2018 IEEE Green Technologies Conference (GreenTech)*. IEEE, 2018, pp. 127–132.

[5] X. Liu, M. Shahidehpour, Y. Cao, L. Wu, W. Wei, and X. Liu, "Microgrid risk analysis considering the impact of cyber attacks on solar pv and ess control systems," *IEEE Transactions on Smart Grid*, vol. 8, no. 3, pp. 1330–1339, 2017.

[6] H. Zhang, W. Meng, J. Qi, X. Wang, and W. X. Zheng, "Distributed load sharing under false data injection attack in an inverter-based microgrid," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 2, pp. 1543–1551, 2019.

[7] S. Sridhar and M. Govindarasu, "Model-based attack detection and mitigation for automatic generation control," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 580–591, 2014.

[8] S. Sahoo, S. Mishra, J. C.-H. Peng, and T. Dragicevic, "A stealth cyber attack detection strategy for dc microgrids," *IEEE Transactions on Power Electronics*, 2018.

[9] O. A. Beg, T. T. Johnson, and A. Davoudi, "Detection of false-data injection attacks in cyber-physical dc microgrids," *IEEE Transactions on industrial informatics*, vol. 13, no. 5, pp. 2693–2703, 2017.

[10] H. Liu, F. Hussain, Y. Shen, S. Arif, A. Nazir, and M. Abubakar, "Complex power quality disturbances classification via curvelet transform and deep learning," *Electric Power Systems Research*, vol. 163, pp. 1–9, 2018.

[11] D. D. Ferreira, J. M. de Seixas, A. S. Cerqueira, C. A. Duque, M. H. J. Bollen, and P. F. Ribeiro, "A new power quality deviation index based on principal curves," *Electric Power Systems Research*, vol. 125, pp. 8–14, 2015.

[12] O. P. Mahela, A. G. Shaik, and N. Gupta, "A critical review of detection and classification of power quality events," *Renewable and Sustainable Energy Reviews*, vol. 41, pp. 495–505, 2015.

[13] Y. Shi, F. Li, W. Song, X.-Y. Li, and J. Ye, "Energy audition based cyber-physical attack detection system in iot," in *ACM Turing Celebration Conference - China (TURC)*, 2019, pp. 1–5.

[14] S. Tan, D. De, W. Song, J. Yang, and S. Das, "Survey of Security Advances in Smart Grid: A Data Driven Approach," *IEEE Communications Surveys and Tutorials*, vol. 18, no. 1, pp. 397–422, 2017. [Online]. Available: http://dx.doi.org/10.1080/17445760.2017.1294690

[15] S. Amini, F. Pasqualetti, and H. Mohsenian-Rad, "Detecting dynamic load altering attacks: A data-driven time-frequency analysis," in *2015 IEEE International Conference on Smart Grid Communications (SmartGridComm)*. IEEE, 2015, pp. 503–508.

[16] Y. Zhou, R. Arghandeh, and C. J. Spanos, "Partial knowledge data-driven event detection for power distribution networks," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5152–5162, 2018.

[17] X. Lu, B. Chen, C. Chen, and J. Wang, "Coupled cyber and physical systems: Embracing smart cities with multistream data flow," *IEEE Electrification Magazine*, vol. 6, no. 2, pp. 73–83, 2018.

[18] J. Tian, B. Wang, and X. Li, "Data-driven and low-sparsity false data injection attacks in smart grid," *Security and Communication Networks*, vol. 2018, 2018.

[19] P. Xun, P. Zhu, Z. Zhang, P. Cui, and Y. Xiong, "Detectors on edge nodes against false data injection on transmission lines of smart grid," *Electronics*, vol. 7, no. 6, p. 89, 2018.

[20] J. Ye, X. Yang, H. Ye, and X. Hao, "Full discrete sliding mode controller for three phase pwm rectifier based on load current estimation," in *2010 IEEE Energy Conversion Congress and Exposition*. IEEE, 2010, pp. 2349–2356.

[21] K. Liu, N. Z. Gebraeel, and J. Shi, "A data-level fusion model for developing composite health indices for degradation modeling and prognostic analysis," *IEEE Transactions on Automation Science and Engineering*, vol. 10, no. 3, pp. 652–664, 2013.

[22] B. Yang, F. Li, J. Ye, and W. Song, "Condition Monitoring and Fault Diagnosis of Generators in Power Networks," in *IEEE Power & Energy Society General Meeting*, 2019.

[23] F. Li, A. Shinde, Y. Shi, J. Ye, X.-Y. Li, and W.-Z. Song, "System statistics learning-based iot security: Feasibility and suitability," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 6396–6403, 2019.

[24] F. Li, Y. Shi, A. Shinde, J. Ye, and W.-Z. Song, "Enhanced cyber-physical security in internet of things through energy auditing," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 5224–5231, 2019.

[25] P. P. Parikh, M. G. Kanabar, and T. S. Sidhu, "Opportunities and challenges of wireless communication technologies for smart grid applications," in *IEEE PES General Meeting*. IEEE, 2010, pp. 1–7.

[26] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE transactions on automatic control*, vol. 58, no. 11, pp. 2715–2729, 2013.

[27] Y. Shoukry and P. Tabuada, "Event-triggered state observers for sparse sensor noise/attacks," *IEEE Transactions on Automatic Control*, vol. 61, no. 8, pp. 2079–2091, 2015.

[28] Y. Lu, R. Xie, and S. Y. Liang, "Adaptive online dictionary learning for bearing fault diagnosis," *The International Journal of Advanced Manufacturing Technology*, vol. 101, no. 1-4, pp. 195–202, 2019.

[29] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[30] R. Xie, Z. Wang, S. Bai, P. Ma, and W. Zhong, "Online decentralized leverage score sampling for streaming multidimensional time series," in *International Conference on Artificial Intelligence and Statistics*, 2019.

[31] X. Zhang, R. Xie, and P. Ma, *Statistical Leveraging Methods in Big Data*. Cham: Springer International Publishing, 2018, pp. 51–74. [Online]. Available: https://doi.org/10.1007/978-3-319-18284-1_3

[32] P. Ma, M. W. Mahoney, and B. Yu, "A statistical perspective on algorithmic leveraging," *The Journal of Machine Learning Research*, vol. 16, no. 1, pp. 861–911, 2015.

[33] P. Drineas, M. Magdon-Ismail, M. W. Mahoney, and D. P. Woodruff, "Fast approximation of matrix coherence and statistical leverage," *Journal of Machine Learning Research*, vol. 13, no. Dec, pp. 3475–3506, 2012.

[34] I. Mitliagkas, C. Caramanis, and P. Jain, "Memory limited, streaming pca," in *Advances in Neural Information Processing Systems*, 2013, pp. 2886–2894.

[35] P. Jain, C. Jin, S. M. Kakade, P. Netrapalli, and A. Sidford, "Streaming pca: Matching matrix bernstein and near-optimal finite sample guarantees for oja's algorithm," in *Conference on Learning Theory*, 2016, pp. 1147–1164.

[36] Z. Yang, P. Wang, X. Ye, and S. Wang, "Fault detection method based on margin statistics of generalized non-negative matrix factorization," in *2017 Chinese Automation Congress (CAC)*. IEEE, 2017, pp. 4723–4728.

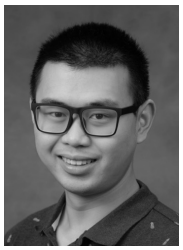
[37] M. Slawski, M. Hein, and P. Lutsik, "Matrix factorization with binary components," in *Advances in Neural Information Processing Systems*, 2013, pp. 3210–3218.

[38] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.

[39] T. Zhao, J. Zhang, F. Li, and K. J. Marfurt, "Characterizing a turbidite system in canterbury basin, new zealand, using seismic attributes and distance-preserving self-organizing maps," *Interpretation*, vol. 4, no. 1, pp. SB79–SB89, 2016.

Fangyu Li is a postdoctoral fellow with the College of Engineering, University of Georgia. He received his PhD in Geophysics from University of Oklahoma in 2017. His Master and Bachelor degrees were both in Electrical Engineering, obtained from Tsinghua University and Beihang University, respectively. His research interests include signal processing, seismic imaging, machine learning, deep learning, distributed computing, Internet of Things (IoT), and cyber-physical systems (CPS).

Rui Xie received the Ph.D. degree in statistics from the University of Georgia, Athens in 2019. He is currently an Assitant Professor in the Department of Statistics and Data Science at the University of Central Florida. His research interests include the development of statistical sketching and sampling methods for large-scale streaming dependent data, and the applications in different fields ranging from streaming online learning and sampling, spatial pattern reconstruction with sketching, to decentralized computing.



Bowen Yang received the B.S. degree in electrical engineering from Huazhong University of Science and Technology, Wuhan, China, in 2018. He is currently a Ph.D. student and Research Assistant with the University of Georgia, USA. His current research interests include advanced control for power electronics and electric machines, energy management system, and cyber-physical security for intelligent electric drives.

Lulu Guo received the B.S. degree in vehicle engineering from Jilin University, Changchun, China, in 2014, and the Ph.D. degree in control engineering from Jilin University in 2019. He is currently a Postdoctoral Research Associate with the University of Georgia, USA. His current research interests include advanced vehicle control, energy management, and vehicle cybersecurity.

Ping Ma received the Ph.D. degree in statistics from Purdue University, USA in 2003. Currently, he is the Professor of Statistics and co-directs the big data analytics lab at the University of Georgia, USA. His research interests include the development of new statistical theory and methods in analyzing vast and complex data to solve scientific and engineering problems, and applications in modern genomic, epigenetic, geophysics, chemical sensing and brain imaging researches. He is a Fellow of the American Statistical Association (ASA). He was Beckman Fellow at the Center for Advanced Study at the University of Illinois at Urbana-Champaign, Faculty Fellow at the US National Center for Supercomputing Applications, and a recipient of the US National Science Foundation CAREER Award.

Jianjun Shi received the B.S. and M.S. degrees in electrical engineering from the Beijing Institute of Technology, Beijing, China, in 1984 and 1987, respectively, and the Ph.D. degree in mechanical engineering from the University of Michigan, Ann Arbor, in 1992. Currently, he is the Carolyn J. Stewart Chair Professor in the H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta. His research interests include the fusion of advanced statistical and domain knowledge to develop methodologies for modeling, monitoring, diagnosis, and control for complex manufacturing systems. Dr. Shi is a Fellow of IISE, ASME, and INFORMS, an elected member of International Statistics Institute, an Academician of the International Academy for Quality (IAQ), and a member of ASQ and ASA.

Jin Ye (IEEE S'13-M'14-SM'16) received the B.S. and M.S. degrees in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2008 and 2011, respectively. She also received her Ph.D. degree in electrical engineering from McMaster University, Hamilton, Ontario, Canada in 2014. She is currently an assistant professor of electrical engineering and the director of the intelligent power electronics and electric machines laboratory at the University of Georgia. She is a general chair of 2019 IEEE Transportation Electrification Conference and Expo (ITEC), a publication chair and women in engineering chair of 2019 IEEE Energy Conversion Congress and Expo (ECCE). She is an associate editor for IEEE Transactions on Transportation Electrification and IEEE Transactions on Vehicular Technology. Her main research areas include power electronics, electric machines, energy management systems, smart grids, electrified transportation, and cyber-physical systems.

WenZhan Song received his Ph.D. in Computer Science from Illinois Institute of Technology (2005), B.S. and M.S. degrees from Nanjing University of Science and Technology (1997 and 1999). He is a Chair Professor of Electrical and Computer Engineering in the University of Georgia. Dr. Song's research focuses on cyber-physical systems and their applications in energy, environment, food and health sectors. He received NSF CAREER award in 2010.