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, Senior Member, IEEE, and 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 article, we propose a novel high-dimensional data-driven cyber physical attack detection and identification (HCADI) approach. First, we analyze the cyber and physical attack impacts (including cyber attacks on the solar inverter causing unusual harmonics) on electric waveforms in the 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, binary matrix factorization (BMF), distribution power grids, leverage score, solar inverter.

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]. Owing 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.

Attacks are studied in applications that are intensively dependent on power electronics converters, including power grids with voltage support devices [2], distribution systems with solar farms [3] and with power electronics-driven heating, ventilation, and air conditioning (HVAC) systems [4], and microgrids [5], [6]. However, they mostly focus on either analyzing or detecting cyberattacks affecting grid-level stability, functionality, and operational costs. In [7], a model-based method was developed to detect the data integrity attacks on automation generation control of transmission systems. In [3], a physical-law-based detection was developed to detect false data attacks that attempt to reduce the output power of solar energy in distribution systems. In [4], a secure information flow framework was developed for a 118-bus distribution network with the 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 the observer-based detection methods. In [9], a physics-based framework was proposed 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 cyberattacks, but those on the devices (power electronics converters) are not studied. Thus, their cyber security framework is not applied to: 1) cyberattack 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 occurs.

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 semisupervised 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 grids was investigated. In [19], a machine learning solution was proposed to identify the false data injection attacks on the 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 a solar inverter) and other critical components (such as relays and generators) in power grids.

Fig. 1 shows the diagram of the distribution power grid with solar farms. The solar farm is physically connected to the distribution grid through the dc/dc and dc/ac converters and the grid-connected transformers. Then, the major components and the control center are connected through cyber networks. The attacks in red are the potential cyberattacks 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 and abnormal load/capacitor bank cutoff). We need to detect and diagnose both the cyber and physical attacks to the distribution power grids with photovoltaic (PV) systems. Compared with the traditional hardware protection, for example, relays, we aim to develop a comprehensive data-driven solution to monitor the power grid adaptively, efficiently, and accurately with more and more various power electronics devices, protecting the system from cyber and physical attacks, even subtle ones.

In this article, 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 first 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 as follows.

1) We develop a novel HCADI framework that effectively detects and identifies both cyber and physical attacks on the grid level and the 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 cyberattack types. The proposed attack diagnosis approach does not 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, the first-stage dc/dc converter, the second-stage dc/ac inverter, and the grid-connected transformer. Here, we analyze, detect, and identify cyberattacks on the solar inverter, causing the unusual harmonics and then poor power quality in distribution systems.

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

\[
\begin{align*}
\frac{di_a}{dt} &= -\frac{R}{L} i_a - \frac{e_a}{L} + \frac{V_{dc}}{L}(2s_a - s_b - s_c) \\
\frac{di_b}{dt} &= -\frac{R}{L} i_b - \frac{e_b}{L} + \frac{V_{dc}}{L}(-s_a + 2s_b - s_c) \\
\frac{di_c}{dt} &= -\frac{R}{L} i_c - \frac{e_c}{L} + \frac{V_{dc}}{L}(-s_a - s_b + 2s_c)
\end{align*}
\]  

(1)

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

\[
\begin{align*}
s_a &= \begin{cases} 1 & (S_1 = 1, S_2 = 0) \\
0 & (S_1 = 0, S_2 = 1) 
\end{cases} \\
s_b &= \begin{cases} 1 & (S_3 = 1, S_4 = 0) \\
0 & (S_3 = 0, S_4 = 1) 
\end{cases} \\
s_c &= \begin{cases} 1 & (S_5 = 1, S_6 = 0) \\
0 & (S_5 = 0, S_6 = 1) 
\end{cases}
\end{align*}
\]  

(2)

where \(i_a, i_b, i_c\) are the currents of each phase, \(e_a, e_b, e_c\) are the phase voltages of the power grid, \(L\) and \(R\) are the inverter inductance and resistance, and \(V_{dc}\) is the dc bus voltage after the first-stage dc/dc converter. To simplify the analysis process, the direct-quadrature-zero (DQZ) transformation is adopted [20] to parameterize the phase voltages of the power grid, all corresponding to the \(d\)-axis components.

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

### 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 \(\eta(t)\) 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 \(i\)th sensor, the observed data can be expressed as

\[
\begin{align*}
x_i(t) &= \eta(t) + \epsilon_i(t), & t = 1, 2, \ldots, \tau \\
x_i(t) &= \alpha_i \eta^*(t - \tau) + \epsilon_i(t), & t = \tau + 1, \tau + 2, \ldots
\end{align*}
\]  

(4)

where \(\alpha_i\) is the unknown amplitude of the change at the \(i\)th sensor. A sensor data matrix \(X\) can be constructed, \(X(t) = [x_1(t), \ldots, x_k(t)]\), \(X \in \mathbb{R}^{k \times n}\), where \(n\) 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 the voltage and current signals \(V = [V_1, V_2, \ldots, V_N]^T, I = [I_1, I_2, \ldots, I_N]^T\) are measured from a network with size \(N\). The nodal waveforms, where depending on the number of phases at node \(i\), \(V_i\) and \(I_i\) can be the row vectors of size 1, 2, or 3. In order to characterize the waveform properties, we adopt instantaneous properties from

\[
s_i(t) = s(t) + j\mathcal{H}(s(t)) = A(t)e^{j\theta(t)}
\]  

(5)

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

\[
\mathcal{H}(s(t)) = \frac{1}{\pi} \int_{-\infty}^{t} \frac{s(t)}{t - \tau} d\tau.
\]  

(6)

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

2) **Differences:** The changes in the nodal dc voltages and branch currents can be expressed as

\[
\begin{align*}
\Delta V_n &= V_n(t) - V_n(t - w) \\
\Delta I_{np} &= I_{np}(t) - I_{np}(t - w)
\end{align*}
\]  

(7)

(8)

where \(w\) is the analysis window size, and \(n\) 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

\[
\begin{align*}
\Delta V_{nA} &= A_{nA}(t) - A_{nA}(t - w) \\
\Delta I_{npA} &= A_{npA}(t) - A_{npA}(t - w)
\end{align*}
\]  

(9)

(10)

where only Phase A is showed, and 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 \(\Delta V_n\) and \(\Delta I_{np}\), 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 the direct comparisons of phase signal attributes. Based on the IA defined in (5), we define
the current unbalance characterization functions $I_a$, $I_b$, and $I_c$ as

$$I_{na} = \frac{1}{3} \sum_{i \neq j} (A_{i=1} - A_{i=0})^2.$$  (11)

$$I_{nb} = \frac{I_{max} - I_{min}}{I_{max}}.$$  (12)

$$I_{nc} = \sum_{i \neq j} (A_{i=0} - A_{i=1})^2.$$  (13)

where $I_{a,max} = \max\{A_{i=1}, A_{i=0}, A_{i=0}\}$ and $I_{a,min} = \min\{A_{i=1}, A_{i=0}, A_{i=0}\}$, and $\Gamma$ denotes a thresholding function to measure the difference. If $I_b$ is not zero, there exists an unbalance among the three phases. Then, we use $I_c$ to determine the number of affected phases, and $I_a$ to measure the absolute changes. Similarly, we can also get $V_a$, $V_b$, and $V_c$.

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 $[V, I, \Delta I, \Delta V]^T$, while for an ac node, the feature matrix is $[V_A, V_B, V_C, I_A, I_B, I_C, \Delta V_A, \Delta V_B, \Delta V_C, \Delta I_A, \Delta I_B, \Delta I_C, I_A, I_B, I_C]^T$. Note that, for a node, the current measurements could be more than 1 as the connections with other nodes can be multiple. Therefore, 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 the nodes in the network 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 is 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 use 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 is considered to capture the dynamics of the signals [29], [30]. On time domain, each data point is correlated with its previous values; on the spatial domain, each sensor records 1-D data and data from different dimensions are correlated with each other spatially, i.e., cross correlated. The temporal dependence in the time domain of the signal calls for time series modeling, where the autoregressive (AR) model can effectively capture such features. The AR model is a time series model that the observations are specified to be dependent on its own previous values and a stochastic term. On the other hand, the similarity and dissimilarity among spatial features that we extract make the VAR model, which is an extension of the AR model and incorporates the multidimensional cross correlation, suitable for analyzing the high-dimensional data. In general, we consider the parameter estimation of an $m$-dimensional VAR model of order $p$, i.e., $\text{VAR}(p)$

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \cdots + \Phi_p y_{t-p} + e_t$$

$$= x_t^T B + e_t$$  (14)

where $y_t$ is the $k$-dimensional response vector observed at time point $t \in \{1, \ldots, n\}$, $B = [\Phi_1, \Phi_2, \ldots, \Phi_p]^T$ is the $mp \times m$ parameter matrix, $x_t = [y_{t-1}, y_{t-2}, \ldots, y_{t-p}]^T$ is a column vector of previous values of length $mp$, and $e_t$ is a sequence of independent and identically distributed (i.i.d.) stochastic random vectors with mean zero and finite nonsingular covariance matrix $\mathbb{E}[e_t e_t^T] = \Psi$. The unknown parameter $B$, at time sample $t$ can be estimated as

$$B_t = \text{arg min}_B \sum_{i=1}^n ||y_{t,i} - x_{i}^T B||_2^2.$$  (15)

If we define a Hat Matrix $P_H = \sum_{i} (x_{i}^T x_{i})^{-1} x_{i}$, the predicted value can be expressed as $\hat{y}_t = P_H y_t$. In addition, the $i$th diagonal element of $P_H$

$$\ell_{ii} = \frac{\partial \bar{y}_i}{\partial y_i} = x_{i}^T \left( \sum_{i} x_{i} x_{i}^T \right)^{-1} x_i$$  (16)

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

$$\ell_{ii} = ||u_i||_2^2$$  (17)

where $u_i$ 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 $[x_1 \cdots x_n]$ [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 behaviors 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 cause diagnosis. Matrix factorization techniques such as nonnegative 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 signals as linear combinations of a small number of “basis” matrix factorization method for attack root cause diagnosis. The construction of traditional process monitoring methods based on multivariate statistics neglects the temporal correlation and spatial dependence of latent variables at different sampling times, and those methods also assume latent variables satisfying a particular distribution.

Here, we consider decomposing signals into the binary basis and its corresponding weights. The binary basis reveals unique latent matrices 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 efficiently 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 for attack root cause analysis without a training process.

We implement the following BMF algorithm to examine 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 $Y \approx H_1W_1$ with $r < m$. Using a metric of the F-norm (Frobenius norm), the general BMF problem takes the form

$$\min_{H_1, W_1} \frac{1}{2} ||Y - H_1W_1||_F^2,$$

$$s.t. \ H_1 \in \{0, 1\}^{n \times r}, \ W_1 \in \mathbb{R}^{r \times m} \ (18)$$

which can be solved by enumerating all the 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 as follows.

1) Randomly selecting from candidate vertices, which yields candidate matrices $\{H_1^{(0)}\}_{i=1}^t$;
2) Subsequently solving $H_1 = \arg \min_{H_1} \min_{W_1} ||Y - H_1^{(0)}W_1||_F^2$ given the current estimate of $W_1$;
3) Update the weight estimate by $W_1 = \arg \min_{W_1} ||Y - H_1^{(0)}W_1||_F^2$ given the current estimate of $H_1^{(0)}$;
4) Alternate Steps (2) and (3) until converge.

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

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

$$\min_{H_2, W_2} \frac{1}{2} ||R_1 - H_2W_2||_F^2$$

$$s.t. \ H_2 \in \{0, 1\}^{r \times s}, \ W_2 \in \mathbb{R}^{s \times m}. \ (19)$$

The rows $H_{1(t,m)}, t \in \{1, \ldots, n\}$ of the binary matrix $H_1$ form the basis elements that indicate the binary coding of the latent states of the signal. The rows $H_{2(t,m)}, t \in \{1, \ldots, n\}$ 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 an online high-dimensional data-driven cyber-physical 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 feature extraction, a typical preprocessing operation filters out the noise interferences and conditions the data if data samples are missing or time stamps are not stable. Using (5)–(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.

V. Simulation

A simulation based on an MATLAB Simulink Demo, 400-kW grid-connected PV farm network, is conducted to generate waveforms of some typical fault in a small-scale power network. The power network topology is shown in Fig. 5.
The power grid is modeled as an ideal voltage source with a rate voltage of 120 kV and connected to the subtransmission 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. In addition, a three-phase inverter is adopted to transfer the dc power to the ac. Moreover, to match the voltage level of the subtransmission system, a 400-kV A power transformer is used to connect the PV farm and the subtransmission system. Moreover, four linear loads are modeled in the system: 30 MW on Bus 4, denoted as the power grid load; 100 kW and 2 MW on Bus 5 and Bus 6, denoted as the subtransmission 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 the normal operation condition, the voltage and current waveforms of Bus 2 are shown in Fig. 6. The sampling frequency is 50 kHz, and 0.5-s data are simulated for each scenario, which have 25 001 samples. Note that, to clearly illustrate details, we only plot 0.1-s data around the event time in Figs. 6–11.

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, cyberattacks, storm, and lighting. In addition, the outcomes of short-circuit fault depend on many factors such as fault location, short fault type, and damage severe degree. Therefore, 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 the distortion of the voltage and the current. The waveform of Bus 4 is shown in Fig. 7, and it is easy to note that this fault causes transient impacts on currents and spike voltage and steady-state asymmetric components.

b) Solar Transformer Grounded Short-Circuit Fault: The short-circuit faults happen on Bus 2, which can be single phase or double phases. Double-phase (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.
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.

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.

c) 30-MW Linear Load Cutoff: Heavy-load cutting off is another common fault in power systems, which could be caused by the integrity attack to the control center. When heavy load is cutoff in a short period, the power system will generate severe oscillations. The waveforms of Bus 4 are shown in Fig. 11.

VI. EVALUATION

A. Preprocessing 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 six nodes (five ac nodes and one dc node) in Fig. 5, so the vectors in the data matrix are aligned following the node number.

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 and dc components do not have the unbalance features, the dimension of the feature matrix is 32 + 32 + 30 = 94, as 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 Figs. 12 and 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.

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 the raw data matrix and the feature matrix, respectively. As the attack happens at $t = 0.2$ s, both the 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 the 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.
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 \( H_{1S} \) and \( H_{2S} \) 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 \( H_{1S} \) and \( H_{2S} \).

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 method called t-distributed stochastic neighbor embedding (t-SNE) [38] is adopted. It is a non-linear 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 is appropriately preserved during the dimensionality-reduction process. Fig. 17 shows the 2-D 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 does not 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 2-D space are also well separated in the original high dimensions. Therefore, the proposed double-layered BMF is promising for attack diagnosis.

VII. CONCLUSION

Solar farms and other renewable energy sources bring potential attack vulnerabilities to distribution power networks. In this article, 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


Fangyu Li received the B.S. degree in electrical engineering from Beihang University, Beijing, China, in 2009, the M.S. degree in electrical engineering from Tsinghua University, Beijing, in 2013, and the Ph.D. degree in geophysics from The University of Oklahoma, Norman, OK, USA, in 2017. He is currently a Post-Doctoral Fellow with the College of Engineering, University of Georgia, Athens, GA, USA. His current 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, GA, USA, in 2019. He is currently an Assistant Professor with the Department of Statistics and Data Science, University of Central Florida, Orlando, FL, USA. His current research interests include the development of statistical sketching and sampling methods for large-scale streaming dependent data, the applications in different fields ranging from streaming online learning and sampling, and spatial pattern reconstruction with sketching, to decentralized computing.

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

Lulu Guo received the B.S. degree in vehicle engineering and the Ph.D. degree in control engineering from Jilin University, Changchun, China, in 2014 and 2019, respectively. He is currently a Post-Doctoral Research Associate with the University of Georgia, Athens, GA, 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, West Lafayette, IN, USA, in 2003. He is currently a Professor of statistics and co-directs the Big Data Analytics Lab, University of Georgia, Athens, GA, USA. His current 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 genomics, epigenetics, geophysics, chemical sensing, and brain imaging researches. Dr. Ma is a fellow of the American Statistical Association (ASA). He was a Beckman Fellow at the Center for Advanced Study, University of Illinois at Urbana-Champaign, and a Faculty Fellow at the US National Center for Supercomputing Applications. He was 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, MI, USA, in 1992. He is currently the Carolyn J. Stewart Chair Professor with the H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA. His current research interests include the fusion of advanced statistical and domain knowledge to develop methodologies for modeling, monitoring, diagnosis, and control for complex manufacturing systems.

Jin Ye (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, and the Ph.D. degree in electrical engineering from McMaster University, Hamilton, ON, Canada, in 2014. She is currently an Assistant Professor of electrical engineering and the Director of the Intelligent Power Electronics and Electric Machines Laboratory, University of Georgia, Athens, GA, USA. Her current research interests include power electronics, electric machines, energy management systems, smart grids, electrified transportation, and cyber-physical systems.

WenZhan Song received the B.S. and M.S. degrees from the Nanjing University of Science and Technology, Nanjing, China, in 1997 and 1999, respectively, and the Ph.D. degree in computer science from the Illinois Institute of Technology, Chicago, IL, USA, in 2005. He is currently the Chair Professor of electrical and computer engineering with the University of Georgia, Athens, GA, USA. His current research interests include cyber-physical systems and their applications in the energy, environment, food and health sectors. Dr. Song received the NSF CAREER Award in 2010.