Oct 23rd, 1:00 PM - 1:30 PM

Cybersecurity Strategy against Cyber Attacks towards Smart Grids with PVs

Fangyu Li
Kennesaw State University, fli6@kennesaw.edu

Maria Valero
Kennesaw State University, mvalero2@kennesaw.edu

Liang Zhao
Kennesaw State University, lzhao10@kennesaw.edu

Yousef Mahmoud
Kennesaw State University, ymahmoud@kennesaw.edu

Abstract

Cyber attacks threaten the security of distribution power grids, such as smart grids. The emerging renewable energy sources such as photovoltaics (PVs) with power electronics controllers introduce new potential vulnerabilities. Based on the electric waveform data measured by waveform sensors in the smart grids, we propose a novel cyber attack detection and identification approach. Firstly, we analyze the cyber attack impacts (including cyber attacks on the solar inverter causing unusual harmonics) on electric waveforms in distribution power grids. Then, we propose a novel deep learning based mechanism including attack detection and attack diagnosis. By leveraging the electric waveform sensor data structure, our 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. For comparison, we have evaluated classic data-driven methods, including -nearest neighbor (KNN), decision tree (DT), support vector machine (SVM), artificial neural network (ANN), and convolutional neural network (CNN). Comparison results verify the performance of the proposed method for detection and diagnosis of various cyber attacks on PV systems.

Location

Zoom Session 1 (Main Papers Track)

Disciplines

Electrical and Electronics | Information Security | Power and Energy | Signal Processing

Comments

key words:

Attack diagnosis, smart grids, data integrity attack, machine learning, deep learning

1. Introduction

Power grids have become more vulnerable to cyber threats than before (Sarangan et al., 2018). Power electronics converters are becoming more vulnerable to cyber-attacks due to their growing penetration in Internet of Things (IoT) enabled applications, including the smart grids (Balda et al., 2017). In response to this emerging concern, developing cyber-secure power electronics converters has received increased attention from the IEEE power electronics society (PELS) that recently launched a cyber-physical-security initiative. There are two main reasons: First, to improve the operation efficiency and eliminate human intervention, the power grid has been more and more connected, resulting in increasing challenges in reliability, security, and stability. Second, a significantly increased amount of distributed energy resources (DERs), such as solar photovoltaic (PV) (Liu et al., 2016) that are typically power electronics converters, are being incorporated into smart grids. Due to the lack of cyber awareness in power electronics community (Balda et al., 2017), it becomes more urgent to develop cyber-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 which are intensively dependent on power electronics converters, including power grids with voltage support devices (Chen et al., 2013), distribution systems with solar farms (Isozaki et al., 2016), with power electronics driven HVAC (Heating, ventilation, and air conditioning) systems (Cao et al., 2018), and microgrids (Liu et al., 2017; Zhang et al.,2019). However, they mostly focus on either analyzing or detecting cyber-attacks affecting grid-level stability, functionality, and operational costs. In Sridhar and Govindarasu (2014), a model-based method was developed to detect data integrity attacks on automation generation control of transmission systems. In Isozaki et al. (2016), 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 Cao et al. (2018), a secure information flow framework was developed for 118-bus distribution network with power electronics driven HVAC system. In Sahoo et al. (2018), a physics-based, cooperative mechanism was developed to detect stealthy attacks in DC microgrids with multiple of DC-DC converters, which can bypass most of observer-based detection methods. In Beg et al. (2017), a physics-based framework to detect false-data injection attacks in DC microgrids with multiple 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 (CPS), there might be a variety of cyber-attacks including coordinated attacks. To mitigate the vulnerability, model-based and data-driven methods have been proposed (Esmalifalak et al., 2014). However, model-based methods that rely on the accurate mathematical models of the healthy systems are hard to be used in real applications because of an unavoidable model-reality mismatch for the complexity of power electronics-based smart grids. Data-driven methods, on the other hand, employing measured data without an explicit mathematical model, are currently receiving attention (Li et al., 2019b; 2019c). To date, the grid security heavily focuses on the system-level and almost neglects the device-level security, particularly power electronics converters, which has not been well addressed (Balda et al., 2017). In our previous work (Li et al., 2019d), we detected and diagnosed a variety of cyber-physical threats for distribution systems with PV farms, including cyber-attacks on the solar inverter controller, cyber-attacks on relays/switches, and other faults (e.g., short circuit faults). Data-driven approaches are gaining increased attention in recent years due to the advancements in sensing and computing technologies (Liu et al., 2018; Ferreira et al., 2015; Mahela et al., 2015; Shi et al., 2019). They show great potentials in detecting and identifying complicated cyber-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 (Tan et al., 2017). In Amini et al. (2015), A data-driven time-frequency analysis was proposed to detect the dynamic load altering attacks. In Zhou et al. (2018), a data-driven hidden structure semi-supervised machine was proposed to implement the power distribution network attack detection. In Lu et al. (2018), multistream data flow was employed to build effective and efficient attack-resilient solutions against the cyber threats targeting electric grids. In Tian et al. (2018), a data-driven and low-sparsity false data injection attack strategy against the smart grid was investigated. In Xun et al. (2018), 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-attacks at the device level (power electronics converters). Thus, a data-driven methodology is needed to detect and identify a variety of cyber-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, 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; cyber-attacks can also target the power grid facilities (such as single/multiple phase short circuit faults of transformers/generators, abnormal load/capacitor bank cut-off). We need to detect and diagnose cyber-attacks to the distribution power grids with PV systems. Compared with the traditional hardware protections, for example, relays, we develop a comprehensive data-driven solution to adaptively, efficiently, and accurately monitor the power grid with various power electronics devices, protecting the system from cyber-attacks, even subtle ones.

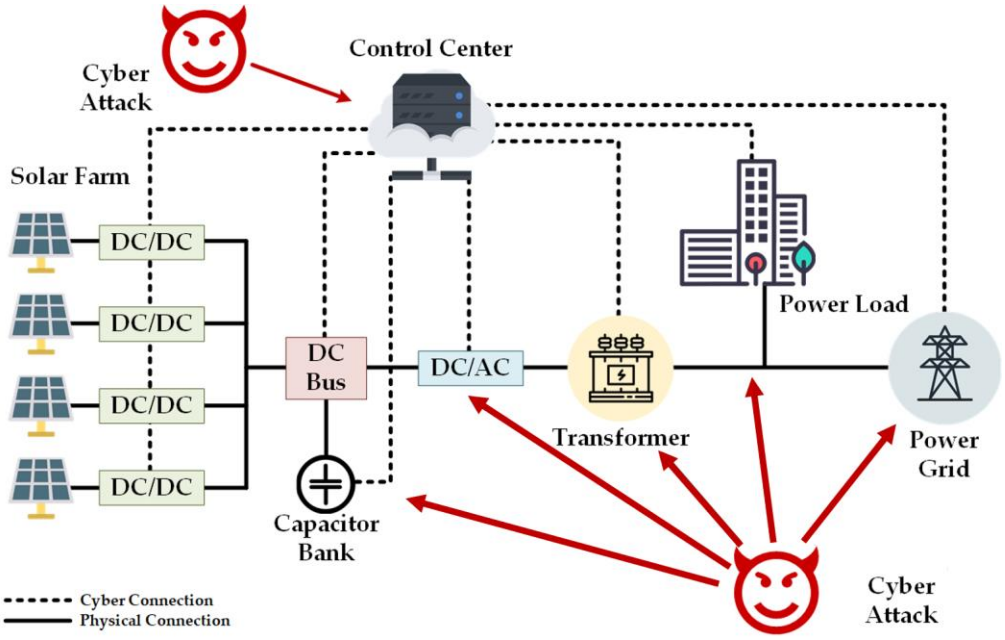

Figure 1. Cyber-attacks threaten the security of the distribution power grid with a solar farm.

In this paper, we propose to develop a data-driven methodology to detect and identify the cyber-attacks on the distribution power grid with solar farms. We first analyze and simulate the impacts of cyber-attacks on electrical waveforms in the distribution power grid with solar farms. Here, we propose a data-driven deep sequence learning method for automatic cyber-attack diagnosis of smart grids with PVs based on feature extraction, anomaly detection, and feature characterization. Unlike our previous approach, we propose to use only one voltage sensor and one current sensor at the point of common coupling for PV systems to detect and diagnose cyber-attacks on DC/DC and DC/AC converters. We test and evaluate our approach in a MATLAB model of the distribution power grid with solar farms in different cyber-attack scenarios (more than 3000 cases). Here, we assume that the waveform sensor at the point of common coupling (PCC) is secure and trustworthy. In real applications, its

communication channel can be encrypted to ensure the security of waveform data. We propose to use multilayer long short-term memory (MLSTM) networks (Gers et al., 1999) to handle intrinsic sequential characteristics of streaming sensor data. Five data-driven methods are engaged as comparison methods, which are K-nearest neighbor (KNN), decision tree (DT), support vector machine (SVM), artificial neural network (ANN), and convolutional neural network (CNN). Finally, the contributions and innovations of our work are:

1. We develop a novel framework that effectively detects and identifies both cyber-attacks on the grid level and device level (power electronics converters) in the 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-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 does not require a training stage, which is superior to machine learning/deep learning-based methods in terms of computational efficiency.

2. 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.

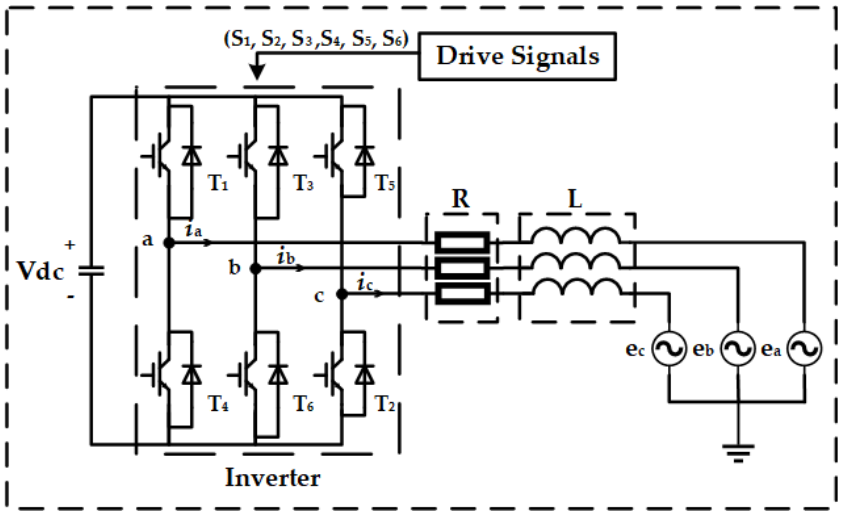

Figure 2. Main circuit topology of the inverter. $S_1 \sim S_6$ *denote the switching signals.*

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:

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

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

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

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 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 (Ye et al., 2010):

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

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 false signal to the solar inverter control signals. Cyber-attacks disrupt the system by manipulating data or introducing corruption. Attacks are assumed to happen between the end devices (or sensors) and the control center, e.g., smart grid measurement data can be attacked in conjunction with the solar panel measurement data. Cyber-attacks are usually defined as mixing the original data/measurements vector with a malicious vector (Beg et al., 2017):

$$\boldsymbol{Z} = \alpha * \boldsymbol{W} + \boldsymbol{Z}_0, \quad (4)$$

where $\boldsymbol{Z}$ is the compromised data vector that is eventually used by the system, $\boldsymbol{Z}_0$ is the true measurement, $\boldsymbol{W}$ is a general compromised data vector which can be independent or determined by $\boldsymbol{Z}_0$, α is a multiplicative factor that defines the weight of the attack vector.

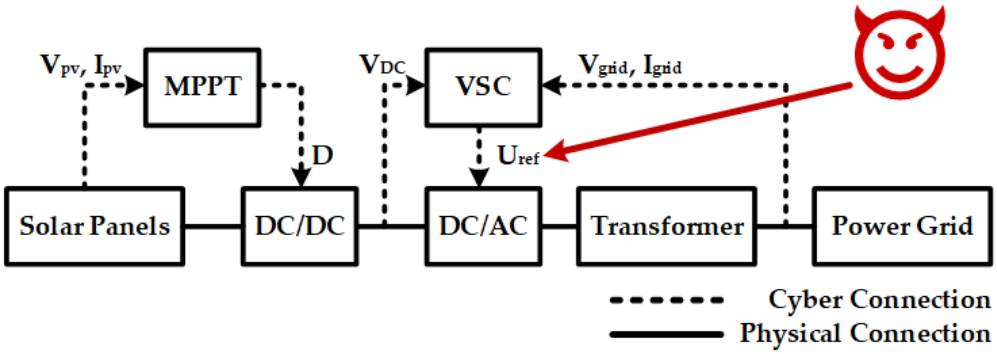

Figure 3. Control diagram of the solar farm system.

3. Methodology

3.1. Problem setup

Suppose we have sequential observations at k sensors, $x_1(t), x_2(t), \dots, x_k(t)$. Before the emergence of the attack, the observations are normal conditions following the electronic model $\eta(\cdot)$ described in Section 2 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:

$$\begin{aligned} x_i(t) &= \eta(t) + \epsilon_i(t), && t = 1,2, \dots, \tau, \\ x_i(t) &= \alpha_i \eta^*(t - \tau_i) + \epsilon_i(t), && t = \tau + 1, \tau + 2, \dots, \end{aligned} \tag{5}$$

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

3.2. 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 Liu et al. (2013) and signal quality measurements in Yang et al. (2019).

3.2.1. Instantaneous Features

The waveforms of voltage and current signals $\boldsymbol{V} = [V_1, V_2, \dots, V_N]^T$, $\boldsymbol{I} = [I_1, I_2, \dots, I_N]^T$ 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:

$$s_c(t) = s(t) + j\mathcal{H}\{s(t)\} = A(t)e^{j\psi(t)}, \quad (6)$$

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:

$$\mathcal{H}\{s(t)\} = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{s(\tau)}{t-\tau}\mathrm{d}\tau. \quad (7)$$

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 $V_{nA} = A_{V_{nA}}e^{j\psi_{V_{nA}}(t)}$.

3.2.2. Differences

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

$$\Delta V_n = V_n(t) - V_n(t-w), \quad (8)$$

$$\Delta I_{np} = I_{np}(t) - I_{np}(t-w), \quad (9)$$

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 3.2.1, the differences can be expressed as:

$$\Delta V_{nA} = A_{V_{nA}}(t) - A_{V_{nA}}(t-w), \quad (10)$$

$$\Delta I_{npA} = A_{I_{npA}}(t) - A_{I_{npA}}(t-w), \quad (11)$$

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 (Li et al., 2019b; 2019c). Here, instead of directly using the difference, we treat it as one dimension of the high-dimensional detection metrics matrix.

3.2.3. 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. (6), we define the current unbalance characterization functions I_α, I_β, and I_γ as:

$$I_{n\alpha} = \frac{1}{3} \sum_{i \neq j}^{i,j \in \{A,B,C\}} (A_{I_{ni}} - A_{I_{nj}})^2. \quad (12)$$

$$I_{n\beta} = \frac{I_{max} - I_{min}}{I_{max}}, \quad (13)$$

$$I_{n\gamma} = \sum_{i \neq j}^{i,j \in \{A,B,C\}} \Gamma (A_{I_{ni}}, A_{I_{nj}}), \quad (14)$$

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 an unbalance among the three phases. Then we use I_γ to determine 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.

3.3. High-dimensional Data Matrix Construction

In Section 3.1, 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 3.2, 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 an AC node has the matrix $[A_{V_A}, A_{V_B}, A_{V_C}, A_{I_A}, A_{I_B}, A_{I_C}, \Delta V_A, \Delta V_B, \Delta V_C, \Delta I_A, \Delta I_B, \Delta I_C, V_\alpha, V_\beta, V_\gamma, I_\alpha, I_\beta, I_\gamma]^T$. 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, monitoring data, even over a large area can be efficiently collected (Parikh et al., 2010).

3.4. Attack Detection Model

Without loss of generality, we assume that there are various states of PV systems, including the normal state and under-attack states with various attack types. Because it is difficult to accurately detect and identify various types of attacks simultaneously, we propose to first focus on detecting whether the PV system is under attack or not. We apply the one-class detection as the attack detection model, which has been widely applied for outlier detection to accurately classify the normal and under-attack states (Maglaras and Jiang, 2014). Training a one-class detection model only requires normal data, which is an advantage for a potentially large number of attacks.

Our proposed detection model is expressed as

$$g(\boldsymbol{x}(t)) = \text{sgn}\,(\mathcal{G}^*(\boldsymbol{x}(t)) - \rho), \tag{15}$$

where $\boldsymbol{x}(t)$ denotes a vector of time series of smart grid sensor data from $t - L$ to t. $\mathcal{G}^*$ is the trained one-class model. ρ is the detection error threshold (DET), so if the prediction error is larger than DET, it may indicate an anomaly. A sign function is defined as

$$\text{sgn}\,(\alpha) := \begin{cases} 1 & \text{if } \alpha \geq 0, \\ -1 & \text{if } \alpha < 0. \end{cases} \tag{16}$$

3.5. Attack Diagnosis Model

The attack identification is actually a classification model based on a multi-classification model to identify attack types. Nevertheless, the seriousness of the same type of attack is also important but has not been well explored. Also, the cross-entropy loss function often in practice means a cross-entropy loss function for classification problems and a mean squared error loss function for regression problems (Goodfellow et al., 2016). Therefore, to analyze not only the attack types but also the seriousness, we propose a cross-entropy loss between the empirical distribution defined by the training set and the probability distribution defined by the model, following

$$J(\theta) = -\mathbb{E}_{x,y\sim\hat{p}_{\text{data}}} \log\, p_{\text{model}}\,(\boldsymbol{y}|\boldsymbol{x}). \tag{17}$$

3.6. Multilayer LSTM based Deep Sequence Learning

Since we try to model electric waveform data which have complicated non-linear temporal characteristics, we leverage the LSTM model. The structure of the recurrent neural network (RNN) utilizes the information memory at the previous time to apply to the current sequence data prediction. However, RNN training long sequences in a multilayer network will generate gradient disappearance and explosion (Bengio et al., 1994). While LSTM uses the concept of the gate structure to control the state of the unit layer at each time to retain the data information. The benefits of LSTM cells are in using the guided gates for selectivity, remembering both short and long-term behaviors across many time series, which effectively solves the problem of gradient diffusion and explosion. Fig. 4 shows the proposed MLSTM architecture, which not only remembers sequential information but also carries out more rigorous screening of time information. So, we can generalize the behavior complexity of the PV system without a huge dataset. Specifically, hyperparameters for MLSTM models are batch size = 128, learning rate = 0.001, hidden size = 32, optimizer = *Adam*, number of layers = 2 (detection) / 5 (diagnosis), which. The parameters are obtained through experiments and trials. Note that CNN shares most of the hyperparameters of MLSTM in our study.

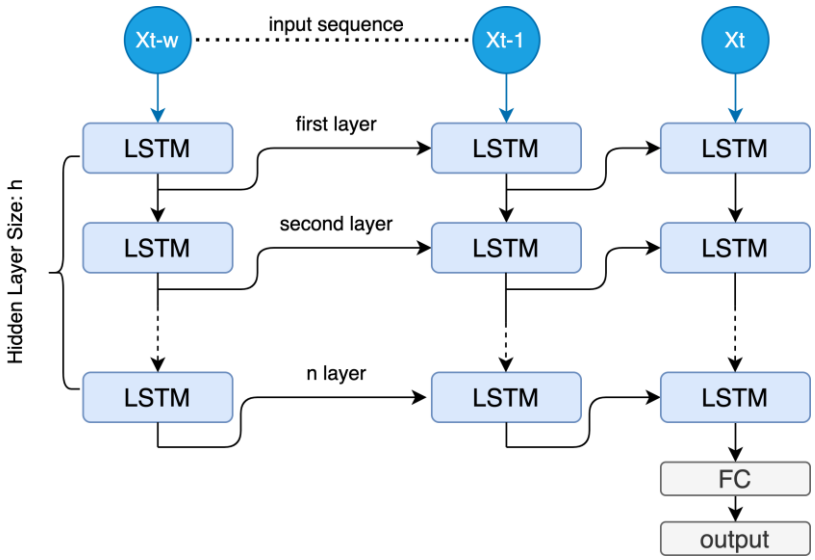

Figure 4. Proposed multilayer LSTM architecture.

4. Algorithm

Based on the theories introduced in Section 3, we propose an online high dimensional data-driven cyber-attack detection and diagnosis algorithm, whose workflow is shown in Fig. 5.

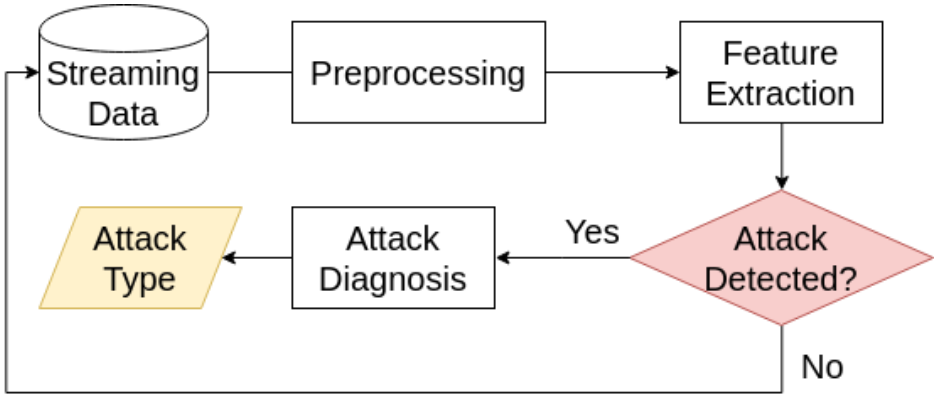

Figure 5. Workflow of the proposed approach. The attack detection is highlighted with red shadow, and the attack diagnosis result is in yellow.

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 pre-processing operation filters out the noise interferences and conditions the data if data samples are missing or timestamps are not stable. Using the Eqs. (6) to (14), from the high dimensional data matrix, we build a high dimensional feature matrix, whose dimension is even higher. Based on the MLSTM attack

detection model, 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 diagnosis model to identify the attack types. 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.

5. Simulation

A simulation-based on MATLAB Simulink, 400kW Grid-Connected PV Farm Network, is conducted to generate waveforms of some typical fault in small scale power network. The main power grid is modeled as an ideal voltage source, and the load is linear. One rate voltage of $260V/25kV$, 400kVA, transformer connects the PV farm, which includes four DC/DC converters and one DC/AC inverter, to the power grid. The power network topology is shown in Fig. 6.

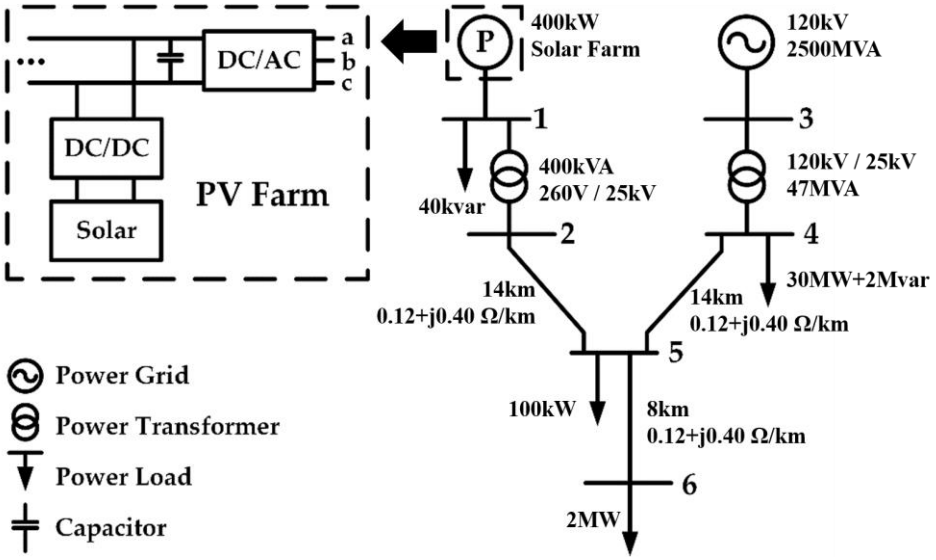

Figure 6. Simulation topology of a 400 kW Grid-Connected PV Farm Network.

The power grid is modeled as an ideal voltage source with a rated voltage of 120 kV and connected to the sub-transmission network with a rated 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. A three-phase inverter is adopted to transfer the DC power to the AC. 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 conditions, the voltage and current waveforms of Bus 2 are shown in Fig. 7.

Here, cyber-attacks on the DC/DC controller sensor only change the current and voltage of the PV panel. Following the cyber-attack model in Eq. (4), α_V and α_I represent the fake measurement coefficient of voltage and current in the PV panel. $(\alpha_V, \alpha_I) \in [(0,0), (2,3), (2,0.3), (0.5,3), (0.5,0.3)]$. For the DC/AC controller, the cyber-attacks inject a time delay into sensor feedback, $t_{delay} \in [0, 4ms, 6ms, 8ms, 10ms, 12ms, 14ms]$.

Considering the uncertainty of cyber-attacks, the attacks happened at different time are simulated in our model, such as phase angles $0°, 30°, 60°, 90°, 120°, 150°, 180°$. To test the robustness of the proposed method towards different conditions, we also consider the irradiation impact on power generation. The irradiation on the PV panel varies in the ranges of 900, 941, 967, 988, 1000 w/m^2. Thus, more than 3,900 training samples are simulated. The waveform at the point of common coupling is obtained to verify our proposed method. The sampling frequency is 50k Hz, and 0.5 seconds (s) data are simulated for each scenario, which has 25001 samples. Note that, to clearly illustrate details, we only plot 0.1 s data around the event time in Figs. 7~12.

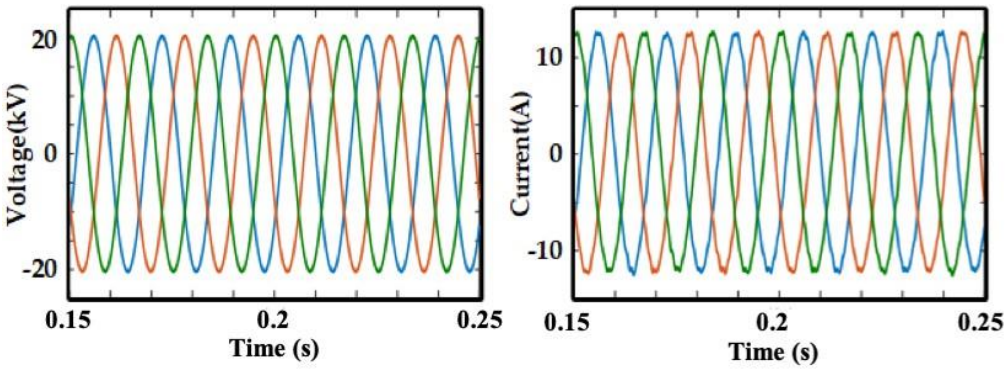

Figure 7. Normal operation condition waveforms of (left) the voltage and (right) current on Bus 2.

Using the simulation system described above, we simulate typical cyber-attack conditions, each of which has featured waveforms. 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 maloperations, cyber-attacks, storms, and lighting. The outcomes of short circuit faults depend on many factors such as fault location, short fault type, and severe degree damage. So, four different short circuit faults are simulated.

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. 8, it is easy to note that this fault causes transient impacts on currents and spike voltage and steady-state asymmetric components.

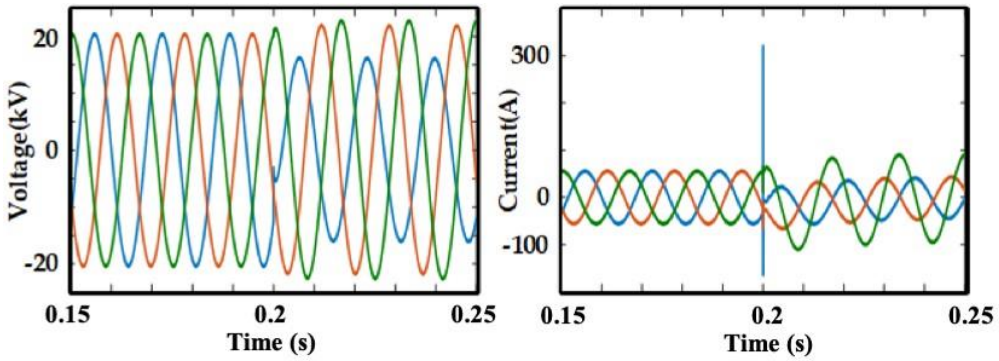

Figure 8. Main grid single phase grounded short circuit fault waveforms of (left) the voltage and (right) current on Bus 4.

Solar transformer grounded short circuit fault: The short circuit faults happen on Bus 2, which can be single-phase or double-phases. A double-phase (phase a and phase b) grounded short circuit fault waveforms of Bus 4 are shown in Fig. 9. Note that the fault current is even more severe than that from the main grid fault described above.

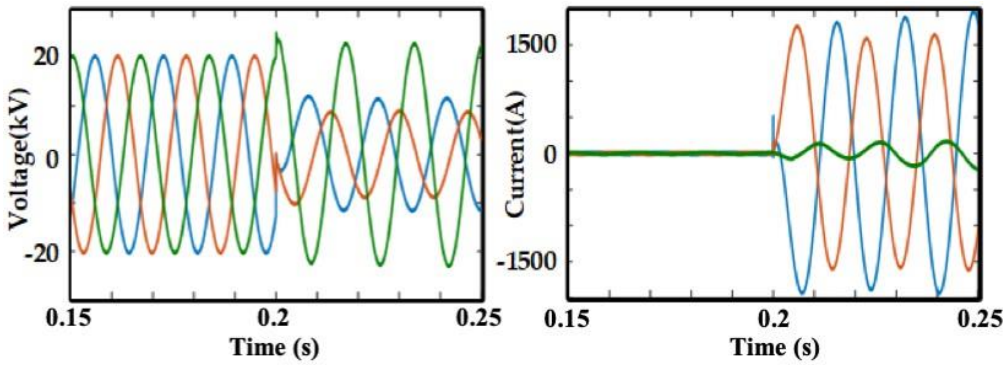

Figure 9. Solar transformer double phases (phase a and phase b) grounded short circuit fault waveforms of (left) the voltage and (right) current on Bus 2.

Extra reactive power compensation in solar system: Fig. 10 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 a capacitive power load and injected to Bus 1, which could be caused by maloperations and purposeful attacks.

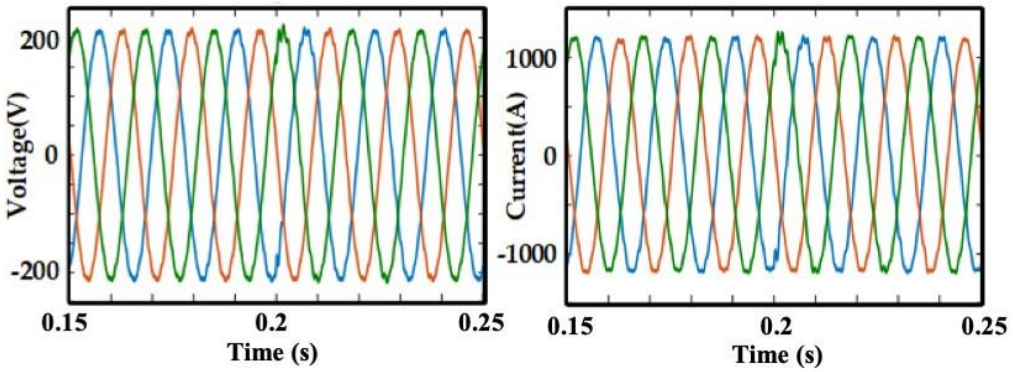

Figure 10. Extra reactive power compensation in solar system waveforms of (left) the voltage and (right) current on Bus 1.

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 (Yang et al., 2019). The waveforms of Bus 1 are shown in Fig. 11.

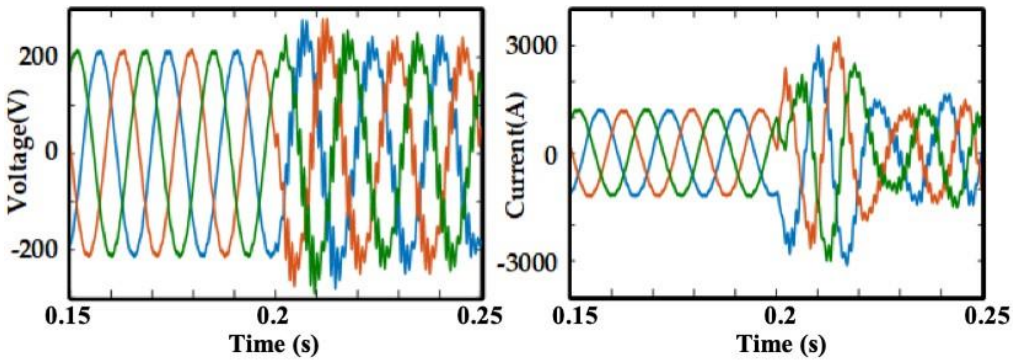

Figure 11. PV farm inverter attacked waveforms of (left) the voltage and (right) current on Bus 1.

30MW linear load cut off: Heavy load cutting off is another common fault in the power system which, could be caused by the integrity attack to the control center. When a heavy load is cut off in a short period, the power system will generate severe oscillations. The waveforms of Bus 4 are shown in Fig. 12.

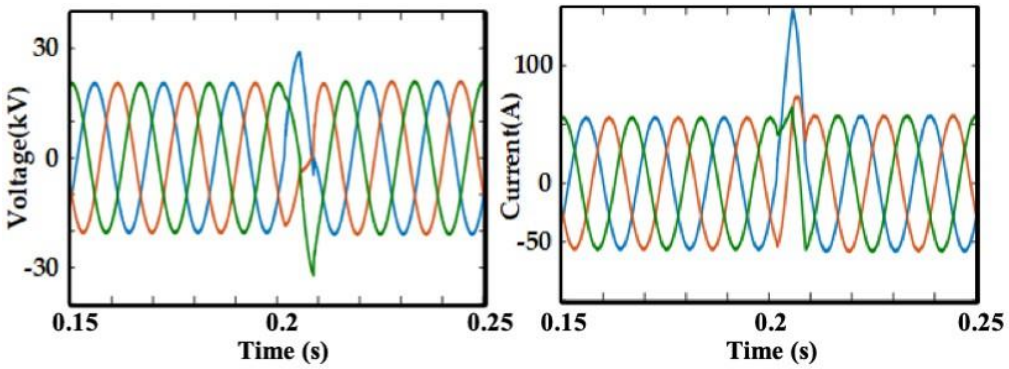

Figure 12. 30 MW linear load cut off waveforms of (left) the voltage and (right) current on Bus 4.

6. Evaluation

6.1. 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. 8 is shown in Fig. 13. 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 (5 AC nodes and 1 DC node) in Fig. 6, so the vectors in the data matrix are aligned following the node number.

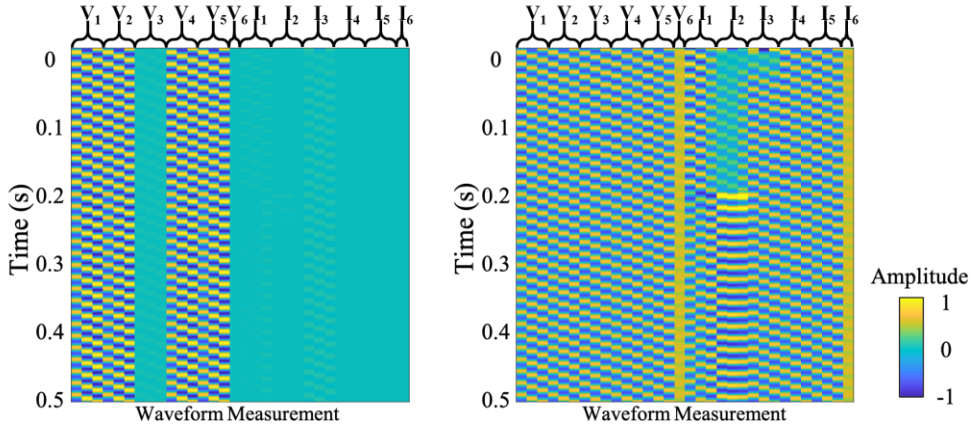

Figure 13. Data matrix normalization in the situation of main grid grounded short circuit fault. (Left) Raw waveform matrix; (Right) Normalized waveform matrix. Each vector corresponds to one voltage or current waveform, which is either one phase of AC components or one DC component. As there are 5 AC 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 3.2. 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. 14. With the sophisticated the feature extraction, the latent data structure information is better characterized, and the attack detection robustness can also be improved. Comparing Fig. 13 and Fig 14, 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.

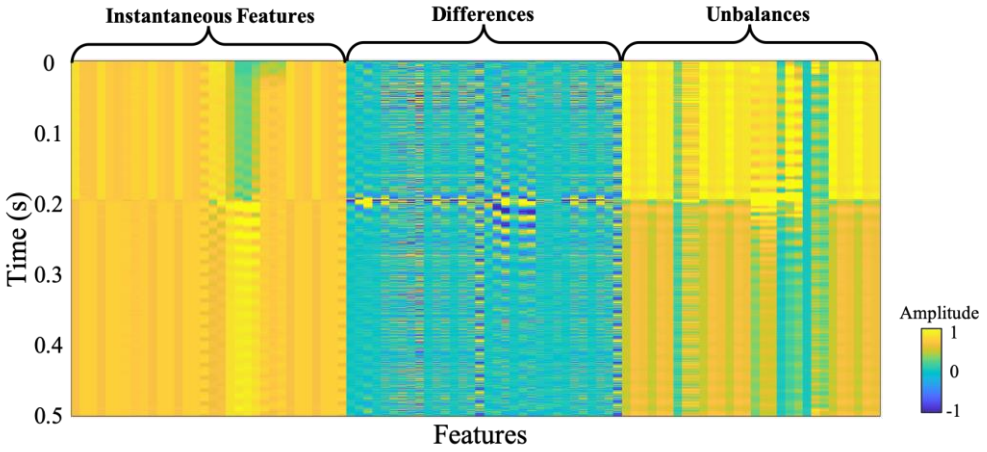

Figure 14. Feature matrix extracted from the normalized waveform matrix shown in Fig. 13. The total dimension is 94, including 32 columns of instantaneous features, 32 columns of differences, and 30 columns of unbalances.

6.2. Comparison Models

To validate the performances of the proposed MLSTM method, classic machine learning and deep learning models, such as KNN, SVM, DT, ANN, and CNN, are compared, which are powerful data-driven methods with a wide range of applications (Goodfellow et al., 2016). For the machine learning models, data features, such as frequency, amplitude, phase angle (because of AC waveform), spectrum properties, are extracted. For deep learning models, data streams are managed to be fed into models. We implemented them through Pytorch (1.3.1) (Paszke et al., 2017) and Sklearn (0.22.1) (Pedregosa et al., 2011) on an Ubuntu 16.04 server (CPU: i7-6850K, 3.60 GHz, RAM 64GB) armed with GPU (GeForce GTX 1080 Ti). For the validation purpose, we utilize ten-fold randomized cross-validation with 80% training data and 20% testing data for the model training. To quantitatively evaluate method performances, we employ accuracy, precision, recall, and F1 score, which are obtained from the confusion matrix for detection and classification evaluation (Li et al., 2019a). We adopt an offline training and online testing strategy.

6.3. Metrics

To quantitatively evaluate method performances, we employ accuracy, precision, recall, and F1 score, which are obtained from the confusion matrix for detection and classification evaluation (Li et al., 2019a). The confusion matrix has indexes of True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN). Precision (TP/(TP+FP)) that represents the true fault detection rate is expected to be as high as possible, because the higher precision is, the less false alarm. Recall (TP/(TP+FN)) represents the ability to find all data points of interest. In our case, the higher recall is, the more true attacks are detected. Similarly, F1 (2TP/(2TP+FP+FN)=2 precision recall/(precision+recall)) that represents the combination property of precision and recall is expected to be as high as possible.

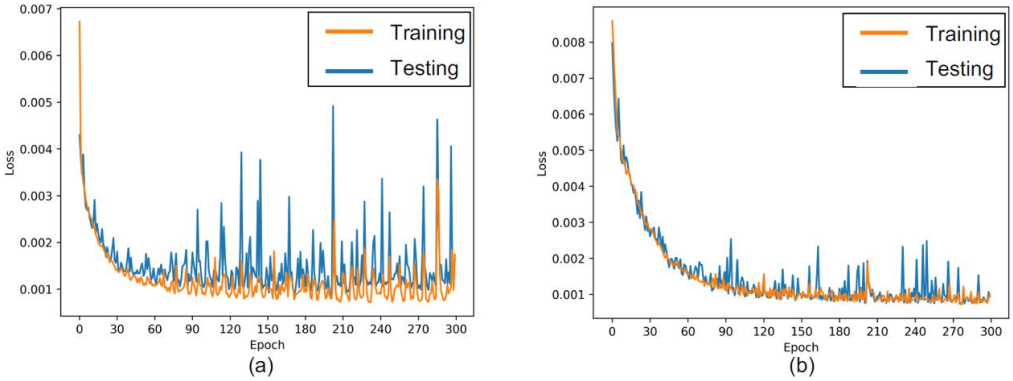

Figure 15. (a) CNN and (b) MLSTM loss curves in the attack diagnosis with window length 100 (0.1 s).

6.4. Attack Detection Performance Evaluation

In the attack detection stage, all data-driven models are trained under the one-class model structure, which is simple with efficient computations. So, the attack detection model has ensured its applicability in practice and thus achieves a real-time manner. Table 1 shows the evaluation metrics: accuracy, recall, precision, and F1 score. In order to further characterize the model sensitivity, we also test the analysis window with different window lengths. It is clear that the proposed MLSTM achieves the best performances in terms of all metrics, with only two layers. SVM cannot achieve good performance, maybe because the data structure is too complicated. KNN and DT show acceptable performances, but not as good as CNN and MLSTM. Due to the shallow model depth, ANN does not show ideal performances, while CNN achieves very good performances with only two layers. Compared with CNN, MLSTM achieves high detection accuracy even when the window size is 50 (0.05 s), and with longer analysis window length, MLSTM can even do better.

6.5. Attack Diagnosis Performance Evaluation

Different from attack detection where only normal and abnormal data are labeled, attack diagnosis requires more detailed data analysis. Because of the data unbalance that normal condition has a large amount of available data. At the same time, each attack scenario only has limited available data, the accuracy of all data-driven models is high, but some have really bad recall, precision and F1 scores, as listed in Table 2. However, MLSTM and CNN still show the advantages of deep learning models even with five layers. Besides the slightly better performances in terms of metrics compared with CNN, MLSTM has another advantage. Fig. 15 displays the training and testing performances of CNN and MLSTM with the same analysis window length. MLSTM shows a smoother loss curve, which means it potentially has better model robustness and stable performances. Notice that MLSTM demonstrates the best performances when the analysis window size is 80 or 100. Although the metrics achieved other peaks with window size 200, that would be overfitting on interferences.

Table 1. Detection performance evaluation using metrics (Accuracy, F1, recall and precision).

Window Size	50	80	100
SVM	0.79/0.47/0.31/0.96	0.77/0.43/0.28/0.97	0.75/0.42/0.27/0.96
KNN	0.90/0.83/0.83/0.84	0.91/0.85/0.86/0.85	0.91/0.87/0.87/0.87
DT	0.92/0.86/0.81/0.92	0.92/0.86/0.82/0.92	0.91/0.87/0.86/0.88
ANN	0.85/0.85/0.81/0.85	0.91/0.91/0.90/0.91	0.91/0.91/0.90/0.91
CNN	0.93/0.93/0.91/0.93	0.97/0.97/0.97/0.97	0.97/0.97/0.97/0.97
MLSTM	**0.97/0.97/0.96/0.97**	**0.98/0.98/0.97/0.98**	**0.98/0.98/0.97/0.98**
Window Size	140	160	200
SVM	0.71/0.36/0.22/0.96	0.69/0.36/0.22/0.97	0.67/0.34/0.21/0.98
KNN	0.90/0.87/0.87/0.87	0.89/0.86/0.87/0.86	0.88/0.86/0.85/0.87
DT	0.91/0.89/0.91/0.87	0.93/0.91/0.92/0.91	0.93/0.92/0.94/0.89
ANN	0.85/0.85/0.85/0.86	0.82/0.82/0.80/0.82	0.75/0.73/0.70/0.78
CNN	0.94/0.94/0.93/0.94	0.95/0.95/0.95/0.95	0.97/0.97/0.97/0.97
MLSTM	**0.97/0.97/0.97/0.97**	**0.97/0.97/0.96/0.97**	**0.98/0.98/0.98/0.98**

Table 2. Diagnosis performance evaluation using metrics (Accuracy, F1, recall and precision).

Window Size	50	80	100
SVM	0.95/0.12/0.11/0.12	0.94/0.03/0.02/0.09	0.95/0.11/0.11/0.12
KNN	0.95/0.02/0.02/0.02	0.94/0.01/0.01/0.01	0.95/0.02/0.01/0.02
DT	0.95/0.12/0.12/0.12	0.95/0.06/0.05/0.06	0.95/0.12/0.12/0.12
ANN	0.95/0.10/0.09/0.10	0.95/0.09/0.08/0.10	0.96/0.11/0.11/0.11
CNN	0.91/0.83/0.83/0.84	0.95/0.90/0.87/0.93	0.95/0.94/0.91/0.97
MLSTM	**0.97/0.93/0.90/0.96**	**0.97/0.94/0.93/0.96**	**0.98/0.95/0.92/0.97**
Window Size	140	160	200
SVM	0.93/0.01/0.01/0.11	0.93/0.01/0.01/0.14	0.93/0.08/0.08/0.08

KNN	0.93/0.01/0.01/0.02	0.93/0.01/0.01/0.01	0.92/0.01/0.01/0.01
DT	0.93/0.04/0.03/0.04	0.93/0.04/0.03/0.05	0.93/0.06/0.06/0.06
ANN	0.94/0.06/0.03/0.13	0.94/0.06/0.05/0.08	0.93/0.12/0.12/0.12
CNN	0.95/0.92/0.90/**0.95**	0.96/0.93/0.90/0.96	0.97/0.96/0.96/0.96
MLSTM	**0.96/0.92/0.91**/0.94	**0.96/0.93/0.90/0.96**	**0.98/0.97/0.96/0.97**

7. CONCLUSION

Solar farms and other renewable energy sources bring potential attack vulnerabilities to distribution power networks. We propose a cyber security mechanism by combining a one-class detection model and an attack diagnosis model, which are tailored for electric waveform profiles of a solar PV smart grid for real-time attack detection and identification. First, an analysis was conducted on cyber-attacks on the smart grid with solar PV farm embedded. Features of the streaming waveform data are constructed to be an analysis matrix, which has the inherent data structure. Then, an MLSTM based comprehensive approach was developed. We apply the one-class detection model to detect whether a PV farm is under attack or not. When it is detected to be under attack, we identify the attack type by leveraging the attack diagnosis model. The proposed mechanism has been evaluated using a MATLAB Simulink solar farm model and achieves much-improved attack detection and diagnosis performances.

CITATIONS

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

Balda, J. C., Mantooth, A., Blum, R., and Tenti, P. (2017). Cybersecurity and power electronics: Addressing the security vulnerabilities of the internet of things. IEEE Power Electronics Magazine, 4(4):37–43.

Beg, O. A., Johnson, T. T., and Davoudi, A. (2017). Detection of false-data injection attacks in cyber-physical dc microgrids. IEEE Transactions on industrial informatics, 13(5):2693–2703.

Bengio, Y., Simard, P., and Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. IEEE transactions on neural networks, 5(2):157–166.

Cao, Y., Davis, K., and Zonouz, S. (2018). A framework of smart and secure power electronics driven HVAC thermal inertia in distributed power systems. In 2018 IEEE Green Technologies Conference (GreenTech), pages 127–132. IEEE.

Chen, B., Mashayekh, S., Butler-Purry, K. L., and Kundur, D. (2013). Impact of cyber-attacks on transient stability of smart grids with voltage support devices. In 2013 IEEE Power & Energy Society General Meeting, pages 1–5. IEEE.

Esmalifalak, M., Liu, L., Nguyen, N., Zheng, R., and Han, Z. (2014). Detecting stealthy false data injection using machine learning in smart grid. IEEE Systems Journal, 11(3):1644–1652.

Ferreira, D. D., de Seixas, J. M., Cerqueira, A. S., Duque, C. A., Bollen, M. H. J., and Ribeiro, P. F. (2015). A new power quality deviation index based on principal curves. Electric Power Systems Research, 125:8–14.

Gers, F. A., Schmidhuber, J., and Cummins, F. (1999). Learning to forget: Continual prediction with lstm.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). Deep learning. MIT press.

Isozaki, Y., Yoshizawa, S., Fujimoto, Y., Ishii, H., Ono, I., Onoda, T., and Hayashi, Y. (2016). Detection of cyber-attacks against voltage control in distribution power grids with PVs. IEEE Transactions on Smart Grid, 7(4):1824–1835.

Li, F., Clemente, J., Valero, M., Tse, Z., Li, S., and Song, W. (2019a). Smart home monitoring system via footstep-induced vibrations. IEEE Systems Journal, pages 1–7. Early access.

Li, F., Shi, Y., Shinde, A., Ye, J., and Song, W.-Z. (2019b). Enhanced cyber-physical security in internet of things through energy auditing. IEEE Internet of Things Journal, 6(3):5224–5231.

Li, F., Shinde, A., Shi, Y., Ye, J., Li, X.-Y., and Song, W.-Z. (2019c). System statistics learning-based IoT security: Feasibility and suitability. IEEE Internet of Things Journal, 6(4):6396–6403.

Li, F., Xie, R., Yang, B., Guo, L., Ma, P., Shi, J., Ye, J., and Song, W. (2019d). Detection and identification of cyber and physical attacks on distribution power grids with PVs: An online high-dimensional data-driven approach. IEEE Journal of Emerging and Selected Topics in Power Electronics, pages 1–10. Early Access.

Liu, H., Hussain, F., Shen, Y., Arif, S., Nazir, A., and Abubakar, M. (2018). Complex power quality disturbances classification via curvelet transform and deep learning. Electric Power Systems Research, 163:1–9.

Liu, K., Gebraeel, N. Z., and Shi, J. (2013). A data-level fusion model for developing composite health indices for degradation modeling and prognostic analysis. IEEE Transactions on Automation Science and Engineering, 10(3):652–664.

Liu, X., Shahidehpour, M., Cao, Y., Wu, L., Wei, W., and Liu, X. (2016). Microgrid risk analysis considering the impact of cyber-attacks on solar pv and ess control systems. IEEE transactions on smart grid, 8(3):1330–1339.

Liu, X., Shahidehpour, M., Cao, Y., Wu, L., Wei, W., and Liu, X. (2017). Microgrid risk analysis considering the impact of cyber-attacks on solar pv and ess control systems. IEEE Transactions on Smart Grid, 8(3):1330–1339.

Lu, X., Chen, B., Chen, C., and Wang, J. (2018). Coupled cyber and physical systems: Embracing smart cities with multistream data flow. IEEE Electrification Magazine, 6(2):73–83.

Maglaras, L. A. and Jiang, J. (2014). A real time ocsvm intrusion detection module with low overhead for SCADA systems. International Journal of Advanced Research in Artificial Intelligence, 3(10).

Mahela, O. P., Shaik, A. G., and Gupta, N. (2015). A critical review of detection and classification of power quality events. Renewable and Sustainable Energy Reviews, 41:495–505.

Parikh, P. P., Kanabar, M. G., and Sidhu, T. S. (2010). Opportunities and challenges of wireless communication technologies for smart grid applications. In IEEE PES General Meeting, pages 1–7. IEEE.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A. (2017). Automatic differentiation in pytorch.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12:2825–2830.

Sahoo, S., Mishra, S., Peng, J. C.-H., and Dragicevic, T. (2018). A stealth cyber-attack detection strategy for dc microgrids. IEEE Transactions on Power Electronics, 34(8), 8162-8174.

Sarangan, S., Singh, V. K., and Govindarasu, M. (2018). Cyber-attack-defense analysis for automatic generation control with renewable energy sources. In 2018 North American Power Symposium (NAPS), pages 1–6. IEEE.

Shi, Y., Li, F., Song, W., Li, X.-Y., and Ye, J. (2019). Energy audition based cyberphysical attack detection system in iot. In ACM Turing Celebration Conference - China (TURC), pages 1–5.

Sridhar, S. and Govindarasu, M. (2014). Model-based attack detection and mitigation for automatic generation control. IEEE Transactions on Smart Grid, 5(2):580–591.

Tan, S., De, D., Song, W., Yang, J., and Das, S. (2017). Survey of Security Advances in Smart Grid: A Data Driven Approach. IEEE Communications Surveys and Tutorials, 18(1):397–422.

Tian, J., Wang, B., and Li, X. (2018). Data-driven and low-sparsity false data injection attacks in smart grid. Security and Communication Networks, 2018.

Xun, P., Zhu, P., Zhang, Z., Cui, P., and Xiong, Y. (2018). Detectors on edge nodes against false data injection on transmission lines of smart grid. Electronics, 7(6):89.

Yang, B., Li, F., Ye, J., and Song, W. (2019). Condition Monitoring and Fault Diagnosis of Generators in Power Networks. In IEEE Power & Energy Society General Meeting.

Ye, J., Yang, X., Ye, H., and Hao, X. (2010). Full discrete sliding mode controller for three phase pwm rectifier based on load current estimation. In 2010 IEEE Energy Conversion Congress and Exposition, pages 2349–2356. IEEE.

Zhang, H., Meng, W., Qi, J., Wang, X., and Zheng, W. X. (2019). Distributed load sharing under false data injection attack in an inverter-based microgrid. IEEE Transactions on Industrial Electronics, 66(2):1543–1551.

Zhou, Y., Arghandeh, R., and Spanos, C. J. (2018). Partial knowledge data-driven event detection for power distribution networks. IEEE Transactions on Smart Grid, 9(5):5152–5162.