Real-Time Detection of Cyber-Attacks in Modern Power Grids with Uncertainty using Deep Learning

Mostafa Mohammadpourfard*, Fateme Ghanatpishe†, Yang Weng ‡, Istemihan Genc*, Mehmet Tahir Sandikkaya*
*Istanbul Technical University, Istanbul, Turkey
†University of Tehran, Tehran, Iran
‡Arizona State University, Tempe, USA

Abstract—The smart grid, which is critical for developing smart cities, has a tool called state estimation (SE), which enables operators to monitor the system’s stability. While the SE result is significant for future control operations, its reliability is strongly dependent on the data integrity of the information obtained from the dispersed measuring devices. However, the dependence on communication technology renders smart grids vulnerable to advanced data integrity attacks, presenting significant concerns to the overall reliability of SE. Among these attacks, the false data-injection attack (FDIA) is gaining popularity owing to its potential to disrupt network operations without being discovered by bad data detection (BDD) methods. Existing countermeasures are limited in their ability to cope with sudden physical changes in the smart grid, such as line outages, due to their development for a certain system specifications. Therefore, the purpose of this paper is to develop an attack detection scheme to find cyber-attacks in smart grids that are influenced by contingencies. In particular, a detection framework based on long short-term memory (LSTM) is proposed to discern electrical topology change in smart grids from real-time FDIA. Results show that the developed framework surpasses the present techniques.

Index Terms—Cybersecurity, deep learning, smart grid, topology changes.

I. INTRODUCTION

Smart power grids are cyber-physical systems (CPS) defined by an integrated structure of a physical power transmission and distribution system with the communication and cyber infrastructure [1]. Although smart grid requirements significantly increase the power system reliability and operation demands, its dependency on Internet of Things (IoT)-based structures and data communication systems makes it vulnerable to cyber-attacks [2], [3]. Attacks on power systems can be broadly classified into two types: physical and cyber attacks. Each of these types of attacks pursues specific goals [4]. The first type of attack is physical attacks. Physical attack targets various physical components of the power grid in order to change or damage the network topology. Direct power outages and sometimes cascading events are considered outcomes of this attack. The second type of attack often targets the supervisory control and data acquisition (SCADA) system, remote terminal units (RTUs), online measuring equipment connected to the network, and data communications infrastructure. This type of attack endangers network security by tricking the network operators [5], [6]. As a result, the necessity to defend future smart grid systems from different cyber-attacks is on the rise. There has been a lot of attention paid to cyber-attacks against state estimation (SE) among the known cyber-attacks [4], [5].

SE’s results may be tainted by using incorrect observations thus it’s essential to identify and delete them. This is performed through the application of techniques referred to as bad data detection (BDD) [5]. Similarly, false data injection attacks (FDIAs) may damage the accuracy of SE results, threatening the reliability of smart grid operations. They must be found and rejected promptly to escape dramatic impacts. [6]. It is feasible for an adversary to modify the state of the grid by manipulating the outputs of several devices in a manner that the BDD cannot detect [7]. FDIA could induce arbitrary errors into SE results, potentially causing large blackouts and economic losses [8]. To safeguard SE solutions against FDIA, several data-driven methods have been developed as provided in Table I.

<table>
<thead>
<tr>
<th>Ref</th>
<th>Main idea</th>
</tr>
</thead>
<tbody>
<tr>
<td>[9]–[12]</td>
<td>Classification methods including MLP, decision trees, SVM, KNN, etc. are used to classify data into normal and attacked classes.</td>
</tr>
<tr>
<td>[13]</td>
<td>Combination of Neural networks (NNs) and decision trees (DT) are used to classify measurements.</td>
</tr>
<tr>
<td>[14], [15]</td>
<td>Deep learning-based RNNs are utilized to find FDIA.</td>
</tr>
<tr>
<td>[16]</td>
<td>GANs and Autoencoder are used to identify FDIA.</td>
</tr>
<tr>
<td>[17]</td>
<td>A classifier is proposed which combines randomized trees with kernel PCA.</td>
</tr>
</tbody>
</table>

A. Motivations

Although there is a growing amount of work on FDIA, the developed detection algorithms are often tailored to particular network architecture and practically are unable to adjust to constantly time-evolving smart grids, making them useful for stationary data. This means that the impact of contingencies such as a line outage in the present systems has not been examined. This assumption would be invalid in a smart grid where frequent topology changes might enhance unpredictability, resulting in substantial state fluctuations in the network.
As a result, the distribution of measurements will alter after topology changes. A technique optimized for a particular network topology may be unable to detect an attack in a separate network configuration since they are created without regard for evolving operating conditions in future smart grids, resulting in a shift in data distribution. In other words, current countermeasures have been made based on the historical data from the system without any changes to its topology. This is while after a line outage, the distribution of new incoming test measurements could change randomly. This is shown in Figure 1. An algorithm trained on the right (blue) side distribution will not be able to correctly predict new incoming test samples from the left (red) distribution. Consequently, models trained on such historical data will no longer be able to accurately forecast the future since data distributions are shifting and past discoveries are no longer connected to the new ones.

This work designs a detection tool that can handle the uncertainty introduced by topology changes to overcome these constraints. In particular, this research carefully develops an LSTM-based methodology for quantifying such uncertainties and detecting threats. The following is an overview of our significant contributions:

1) The paper explores the impacts of topology changes such due to a line outage on cybersecurity, i.e., it demonstrates the distinction between normal variations caused by contingencies and FDIA and proves that a physical change in the power system may result in a reduction in the accuracy of currently available FDIA detection systems.

2) The behavior of existing supervised learning approaches in spotting FDIA under topology changes is thoroughly investigated.

3) A LSTM-based framework is built to describe the complex character of power networks caused by topology changes. The results show that our method can differentiate between intrinsic uncertainties and an FDIA in real-time.

The rest of the paper is organized as follows. Section II explains FDIA. Section III explains some essential concepts of LSTM. The numerical findings are presented in Section IV. Section V contains the paper’s conclusion.

II. SYSTEM MODEL AND BACKGROUND

A. State Estimation

The state estimator in the control center processes the collected measurement data, such as bus voltage, real and reactive power injections, and real and reactive power flow in the branches from remote terminal units (RTUs) and voltage and current phasors from the phasor measurement units, to estimate the most probable states of a power system. The power systems SE models typically employ the recorded measurements to estimate the unknown state variables. The SE model can be shown as follows [18]:

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

where \( \mathbf{z} = [z_1, z_2, \ldots, z_m]^T \in \mathbb{R}^{m \times 1} \) and \( \mathbf{x} = [x_1, \ldots, x_n; \theta_1, \ldots, \theta_n]^T \in \mathbb{R}^{2n \times 1} \) denote the measurements and the state variables, respectively; \( \mathbf{e} = [e_1, e_2, \ldots, e_m]^T \in \mathbb{R}^{m \times 1} \) is the assumed Gaussian noise with zero mean and covariance \( \mathbf{R} \); and \( \mathbf{h}() \) is the Jacobian matrix that represents the topology of the physical grid. According to minimizing the weighted least squares criterion, the estimated system state can be further formulated by:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} [\mathbf{z} - \mathbf{h}(\mathbf{x})]^TK^{-1}[\mathbf{z} - \mathbf{h}(\mathbf{x})]$$

A common approach for detecting the presence of bad data is looking at \( \ell_2 \)-norm detector as follows [19]:

$$\|r\| = \|\mathbf{z} - \mathbf{h}(\hat{\mathbf{x}})\| \geq \tau,$$

where \( \hat{\mathbf{x}} \) is the estimated state variable. Consequently, this residual would be high in presence of inaccurate measurements. Therefore, if the expression’s value in Eq. (3) is greater than a specific threshold, it verifies the presence of bad measurements. The commonly accepted DC SE is employed in this work [10], [11], [13]. In DC SE, phase angles are used to indicate system states \( \mathbf{\theta} = [\theta_1, \theta_2, \ldots, \theta_n] \). In this model, the relation between \( \mathbf{z} \) and \( \mathbf{x} \) is given by: [18]:

$$\mathbf{z} = \mathbf{Hx + e},$$

where \( \mathbf{H} \) demonstrates the connection between \( \mathbf{z} \) and \( \mathbf{x} \). \( \mathbf{e} \) represents measurement error vector. System states \( \hat{\mathbf{x}} \) often obtained via a weighted least squares method as follows estimation as:

$$\hat{\mathbf{x}} = (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{z}.$$
B. FDIA Model

The SE could be affected by FDIA, if the attacker has adequate knowledge of the system topology presented by the Jacobian matrix. In this case, an attacker manipulates the sensor measurements and makes an arbitrary change in the estimated state variables without being detected by the above-mentioned BDD test. The principle of FDIA could explained as following: suppose \( a = [a_1, a_2, \cdots, a_m]^{T} \in \mathbb{R}^{n \times 1} \) is the attack vector that adds to the clean measurement vector. Therefore, \( z_a = z + a \) represents the manipulated measurement vector and consequently \( \hat{x}_{bad} \) would be the estimated state when applying the manipulated measurements \( z_a \). Then

\[
\|r_a\|_2 = \|z_a - H\hat{x}_a\|_2 = \|z + a - H(\hat{x} + c)\|_2 = \|z - H\hat{x} + (a - Hc)\|_2 = \|z - H\hat{x}\|_2 = \|r\|_2
\]

III. Deep Learning Methods

Nowadays, deep learning has emerged as a scientific breakthrough in the field of computer science and artificial intelligence. Thus, this science has a special place in all sciences and all kinds of industrial applications. The development of deep learning algorithms has significantly contributed to the evolution of science based on artificial intelligence and solved most of the problems related to artificial neural network and machine learning techniques. So that today, the use of deep learning applications has been able to solve many problems related to energy and power systems and can be used mainly for forecasting applications, fault diagnosis, and detection of various types of cyber-attacks in power and energy systems. Nonlinear data processing in multiple layers and feature extraction from data in several continuous steps are two critical factors in the performance of deep learning techniques [20]. So far, many algorithms have been introduced for deep learning applications that have supervised and unsupervised learning. In this paper, the detection of FDIA in power systems is done by developing an accurate multilayer LSTM-based framework which captures uncertainties of modern power grids and successfully detects cyber-anomalies.

A. Long Short-Term Memory (LSTM)

LSTM is one of the deep learning applications that was first mentioned in [21]. Solving the problems corresponding to the vanishing gradient when training recurrent neural networks with the gradient-based back-propagation through time technique is the most important feature of LSTM, which makes the LSTM superior to other networks [22]. In addition, the LSTM is a time-series method and can perform well in processing the time series and big data [22]. Each LSTM unit comprises a forget gate, input gate and output gate, as depicted in Fig. 2. In this structure, the cell state is updated by the results of the forget gate and input gate. The mathematical structure of the LSTM is as follows [23]:

\[
f_t = \sigma(\sum_{u=1}^{U} W_{uf} x^t_u + \sum_{k=1}^{H} W_{hf} h^t_{v-1} + b_f) \tag{7}
\]

\[
i_t = \sigma(\sum_{u=1}^{U} W_{ui} x^t_u + \sum_{k=1}^{V} W_{hi} h^t_{v-1} + b_i) \tag{8}
\]

\[
o_t = \sigma(\sum_{u=1}^{U} W_{uo} x^t_u + \sum_{k=1}^{V} W_{ho} h^t_{v-1} + b_o) \tag{9}
\]

\[
a_t = \tanh(\sum_{u=1}^{U} W_{ua} x^t_u + \sum_{k=1}^{V} W_{ha} h^t_{v-1} + b_a) \tag{10}
\]

\[
c_t = c_{t-1} \Phi f_t + i_t \Phi a_t \tag{11}
\]

\[
m_t = o_t \Phi \tanh c_t \tag{12}
\]
TABLE II
LSTM MODEL PARAMETERS.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input layers</td>
<td>1</td>
</tr>
<tr>
<td>Hidden layers</td>
<td>3</td>
</tr>
<tr>
<td>Hidden units</td>
<td>100</td>
</tr>
<tr>
<td>Dropout</td>
<td>0.25</td>
</tr>
<tr>
<td>Sequence length</td>
<td>5</td>
</tr>
<tr>
<td>Batch size</td>
<td>50</td>
</tr>
<tr>
<td>Learning rate</td>
<td>0.0005</td>
</tr>
<tr>
<td>Optimizer</td>
<td>Adam</td>
</tr>
<tr>
<td>FC layers</td>
<td>1</td>
</tr>
<tr>
<td>Output layers</td>
<td>1</td>
</tr>
</tbody>
</table>

4 hidden layers are adequate for simulating power system uncertainty. After executing several trials on the training dataset (see section IV), it was determined that LSTM with three hidden layers is appropriate for modeling the IEEE 14-bus test system, and that the average training loss is acceptable. For a model with 1 or 2 layers, the method was unable to converge. With 4 hidden layers, LSTM demonstrated increased training loss and increased training duration. In addition, a more sophisticated LSTM design will increase the likelihood of overfitting. Consequently, the simpler LSTM with 3 hidden layers was chosen for this study. Figure 3 shows the training loss of the model with various layers.

Finally, the activation function SoftMax is used in this design. For the LSTM network to operate well, dropout and sequence length must be taken into account [24]. Network performance is affected by each of these variables in a unique and specific manner. Dropout is seen as a regulatory solution since it is used to prevent the overfitting process. The sequence length provides the number of samples that preceded the current one. These parameter’s value may be determined by a number of tests. To this end, data was divided into a training set (80%) and a test set (20%) to achieve. The parameters for the carefully designed LSTM model in this paper are listed in Table II. The final architecture of the developed LSTM network for detection of FDIA considering uncertainties raised by topology changes is shown in Fig. 4.

IV. Numerical Results

Various tests are designed to evaluate our methodology’s performance. A system with no topology reconfigurations is used to analyze FDIA in Case I. Case II shows how the proposed framework can manage contingencies. The simulations are performed in MATPOWER and IEEE 14-bus test system is utilized as the test system as shown in Fig. 5. The load data used in the simulations is given by the New York Independent System Operator (NYISO) which helps to ensure that the simulations are more accurate.

Each time a system is breached, we inject two new FDIAs into the system, each worth 90% and 110% of their true value. Each system state and numerous states $\theta_{2,7}$ and $\theta_{5,9}$ are injected with fake data. Attack detection is conducted by labeling and categorizing normal and attacked system states and active powers. The proposed method is evaluated by performance evaluation indicators such as F1-score which is calculated as follows [25]:

$$\ F1 = \frac{2pr}{p + r} \ (13)$$
where $p$ and $r$ denote the precision and recall, respectively, and calculated as:

$$p = \frac{tp}{tp + fp} \quad (14)$$

$$r = \frac{tp}{tp + fn} \quad (15)$$

where $tp$ shows the true positive and true negative, respectively, $fp$ and $fn$ represents the false positive and false negative, respectively.

**A. Case I: FDIA Detection without Contingencies**

In this case, the network is expected to operate regularly until the seventh day. In other words, regular observations continue for six days until the attacker injects FDIA into the measurements of the last day. Existing well-established machine learning-based FDIA detectors such as SVM, KNN, and C4.5 [25] are also applied to this dataset for comparison purposes. Table III summarizes the findings in this case.

<table>
<thead>
<tr>
<th>Method</th>
<th>$F_1$</th>
<th>FP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed method</td>
<td>0.996</td>
<td>0%</td>
</tr>
<tr>
<td>SVM</td>
<td>1</td>
<td>0%</td>
</tr>
<tr>
<td>C4.5</td>
<td>0.993</td>
<td>0.97%</td>
</tr>
<tr>
<td>KNN</td>
<td>1</td>
<td>0%</td>
</tr>
</tbody>
</table>

**B. Case II: FDIA Detection in System with Contingencies**

Scheduled maintenance and cyber-attacks may all lead to a change in the network configuration [2]. However, system reconfiguration raises uncertainty, which changes sample distribution and reduces FDIA detection accuracy. After a line interruption, false data is inserted to test the established framework’s resilience. To this end, it is assumed that the network works normally until day seven and a line outage at 10 AM in this day and the system will continue to work under that contingency till the day is over. However, the adversary injects false data into measurements between 6 PM and 12 PM. This means FDIA is conducted after system reconfiguration where the data distribution of the new system differs from the old one.

Table IV summarizes the results of executing the pre-built and stored machine learning models on the dataset of Case I.

<table>
<thead>
<tr>
<th>Method</th>
<th>$F_1$</th>
<th>FP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed method</td>
<td>0.91</td>
<td>0.39%</td>
</tr>
<tr>
<td>SVM</td>
<td>0.478</td>
<td>8.16%</td>
</tr>
<tr>
<td>C4.5</td>
<td>0.482</td>
<td>8.45%</td>
</tr>
<tr>
<td>KNN</td>
<td>0.736</td>
<td>8.16%</td>
</tr>
</tbody>
</table>

**C. discussion**

Pre-trained machine learning models that are presently used to identify FDIA lose their capacity to consistently classify samples and offer primary performance when a contingency occurs in the system, as can be seen in the results. The underlying reason for this reduction is that topology changes create natural jumps in the data, enhancing variance without adding attacks that fool existing FDIA recognition systems. This is in contrast to the carefully designed FDAI detection tool, which identifies FDIA and manages contingencies. This is because the proposed approach is specifically intended to accurately describe the underlying distribution of power grid data and to eliminate the uncertainties generated by the network reconfiguration. As a consequence, current detectors are incapable of managing the dynamic character of future power grids, resulting in inaccuracy in classifying observed values.

Fig. 6 summarizes the findings in numerical results. As it is clear, the current detector’s F-measure dropped dramatically following contingencies and also suffer from a high false-positive rate. This is because the sampling distribution of incoming data drifted, and previous observations are unrelated to the new ones. This is despite the fact that the developed framework has a higher F-measure and a lower percentage of
FDIAs threaten the security of power grids. These attacks can bypass bad measurement detection methods. Present FDIA detection techniques are targeted for a certain network configuration and ignore the problems of an ever-changing and developing power grid as a consequence of topology changes. Contingencies increase uncertainties impairing the performance of existing FDIA detectors, resulting in incorrect labeling of new samples. This paper discussed strategies for defending against FDIA in smart grids that have been influenced by network reconfigurations. In particular, a framework based on the LSTM algorithm is carefully built, which accurately captures the dynamic character of intelligent grids. Results demonstrated that the carefully designed model is resilient to changes in smart grids and exceeds conventional FDIA detection methodologies. In the future, we want to use our newly developed detection framework to spot anomalies on future intelligent distribution systems. Because of the randomness of the distribution grid, it is difficult to identify possible attacks there. In addition, it will be explored if the suggested technology can be used and expanded to detect and defend vulnerable components of integrated energy systems from cyber-attacks in order to provide a resilient functioning.

V. CONCLUSION

FDIAs threaten the security of power grids. These attacks can bypass bad measurement detection methods. Present FDIA detection techniques are targeted for a certain network configuration and ignore the problems of an ever-changing and developing power grid as a consequence of topology changes. Contingencies increase uncertainties impairing the performance of existing FDIA detectors, resulting in incorrect labeling of new samples. This paper discussed strategies for defending against FDIA in smart grids that have been influenced by network reconfigurations. In particular, a framework based on the LSTM algorithm is carefully built, which accurately captures the dynamic character of intelligent grids. Results demonstrated that the carefully designed model is resilient to changes in smart grids and exceeds conventional FDIA detection methodologies. In the future, we want to use our newly developed detection framework to spot anomalies on future intelligent distribution systems. Because of the randomness of the distribution grid, it is difficult to identify possible attacks there. In addition, it will be explored if the suggested technology can be used and expanded to detect and defend vulnerable components of integrated energy systems from cyber-attacks in order to provide a resilient functioning.

REFERENCES


