An Accurate False Data Injection Attack (FDIA) Detection in Renewable-Rich Power Grids

Mostafa Mohammadpourfard*, Yang Weng†, Istemihan Genc*, Taesic Kim‡
*Istanbul Technical University, Istanbul, Turkey
†Arizona State University, Tempe, USA
‡Texas A&M University-Kingsville, Texas, USA

Abstract—An accurate state estimation (SE) considering increased uncertainty by the high penetration of renewable energy systems (RESs) is more and more important to enhance situational awareness, and the optimal and resilient operation of the renewable-rich power grids. However, it is anticipated that adversaries who plan to manipulate the target power grid will generate attacks that inject inaccurate data to the SE using the vulnerabilities of the devices and networks. Among potential attack types, false data injection attack (FDIA) is gaining popularity since this can bypass bad data detection (BDD) methods implemented in the SE systems. Although numerous FDIA detection methods have been recently proposed, the uncertainty of system configuration that arises by the continuously increasing penetration of RESs has been given less consideration in the FDIA algorithms. To address this issue, this paper proposes a new FDIA detection scheme that is applicable to renewable energy-rich power grids. A deep learning framework is developed in particular by synergistically constructing a Bidirectional Long Short-Term Memory (Bi-LSTM) with modern smart grid characteristics. The developed framework is evaluated on the IEEE 11-bus system integrating several RESs by using several attack scenarios. A comparison of the numerical results shows that the proposed FDIA detection mechanism outperforms the existing deep learning-based approaches in a renewable energy-rich grid environment.

I. INTRODUCTION

Computing, telecommunication, and automation technologies will all be critical components of smart grids. However, this high reliance exposes the network to a wide spectrum of cyber-attacks that may degrade the reliability of smart grids and eventually destabilize critical national infrastructure sectors, resulting in substantial market failures, civic unrest, and significant financial loss [1], [2]. For example, power outages in three Ukrainian areas in 2015 was due to compromising communication infrastructures. US-CERT also published a study in 2018 detailing the hack of US energy networks [3]. Consequently, the need to protect smart grid systems from various cyber-attacks is an increasing concern. Among the established ones, attacks on state estimation (SE) have received a lot of attention [3]. SE is critical to the sustainability of the future power grid since the information it provides is employed in many other energy management system tasks (EMS) [5].

The accuracy of SE relies on measurement devices and communication in the grid. Thus, it is important to evaluate the trustworthiness of data used in SE. Commonly used evaluation methods are bad data detection (BDD) that finds severely corrupted measurements which are normally contaminated by noise, or sensor faults, or communication errors [6].

False data injection attacks (FDIAs) typically manipulate the integrity of sensor data in a stealthier manner to bypass BDD. Therefore, FDIAs may pose a substantial threat to the credibility of grid operations by damaging the accuracy of state estimate findings and must be detected and rejected quickly in order to prevent severe negative consequences [6]. As a result, FDIA has the capacity to introduce random biases into SE outputs, possibly leading to catastrophic effects such as massive blackouts and financial damages [6]. To protect SE solutions against FDIA, many data-driven detectors have been proposed. An overview of key contributions from available studies is provided in Table I.

Although many FDIA detection methods have been recently proposed, the existing methods are often tailored to a certain grid architecture, which is unable to adapt constantly evolving power grids. Specifically, rapidly increasing penetration of renewable energy systems (RESs) will significantly raises unpredictability of the grid architecture, resulting in degrading the accuracy of the existing FDIA detection methods.

To address the challenges, this paper proposes a new FIDA detection system capable of coping the uncertainty caused by renewable-rich power grids. A methodology based on Bidirectional Long Short-Term Memory (Bi-LSTM) is proposed to detect FDIAs while quantifying the uncertainties introduced

<table>
<thead>
<tr>
<th>Ref</th>
<th>Main idea</th>
</tr>
</thead>
<tbody>
<tr>
<td>[9]–[12]</td>
<td>Attacked samples are classified as either “normal” or “false” using classification techniques such as MLP, decision trees, SVM, KNN, etc.</td>
</tr>
<tr>
<td>[1]</td>
<td>Utilizing convolutional neural networks and image-based approaches like Recurrence Plots, a two-tiered CNN-based system is developed.</td>
</tr>
<tr>
<td>[13]</td>
<td>A framework utilizing neural networks and decision trees to classify attacks from non-attack data is proposed.</td>
</tr>
<tr>
<td>[14]</td>
<td>Recurrent neural networks are used to detect FDIA.</td>
</tr>
<tr>
<td>[15]</td>
<td>Generative adversarial networks incorporated with Autoencoder is utilized to develop a framework to detect FDIA.</td>
</tr>
<tr>
<td>[17]</td>
<td>A classifier combining the randomized trees and kernel principal component analysis is designed.</td>
</tr>
</tbody>
</table>
by integrating RESs and validated by simulation results. The contributions of this paper include:

1) This paper thoroughly examines the influence of RESs in terms of cybersecurity perspective, i.e., it illustrates the difference between natural fluctuations caused by uncertainty and FDIA detection accuracy. It is observed that including RES into the reference system could degrade the accuracy of conventional FDIA detection methods.

2) The behavior of commonly used machine learning (ML)-based approaches for FDIA detection in the renewable-rich power grid is thoroughly analyzed.

3) The proposed deep learning-based framework can identify the complex character of the renewable-rich power grid. The results indicate that the developed methodology can differentiate between intrinsic uncertainties and an FDIA.

The remainder of the paper is organized as follows. Section II explains FDIAs. Section III explains critical concepts of Bi-LSTM. Section IV presents the numerical results. Section V contains the paper’s conclusion.

II. PROBLEM SETTING

To evaluate the FDIA, this paper uses the widely used DC SE [10, 11, 13]. In DC SE, system states are represented by bus phase angles $\theta = [\theta_1, \theta_2, \cdots, \theta_n]$. In this model, the relation between the collected measurements from the power network $z = [z_1, z_2, \ldots, z_m]^T$ and system state vector $x = [\theta]^T$ is provided by: [18]:

$$z = Hx + e,$$  \hspace{1cm} (1)

where $H$ represents relation between $z$ and $x$ obtained from the grid’s physical structure. $e$ is the vector of measurement errors. System states $\hat{x}$ is generally achieved by the use of the weighted least squares method as follows estimation as:

$$\hat{x} = (H^TH)^{-1}H^Tz,$$  \hspace{1cm} (2)

where $W$ is the error covariance matrix. 2-Norm residual of measurement $\|r\|_2 = \|z - H\hat{x}\|_2$ is utilized in smart grids to discover faulty data. If $\|r\|_2$ is less than a predetermined value $\tau$, the measurement would be considered as a normal one. However, FDIA can manipulate the system states while bypassing residual analysis-based BDD [7]. If an attacker who is familiar with $H$, he/she can methodically build an attack vector $a = Hc$ to inject into the initial normal measurement. $c$ is a random error vector inserted into the systems’ normal estimations $\hat{x}$. By using the newly altered measurement $z_a = z + a$, inaccurate estimations $\hat{x}_a = \hat{x} + a$ will be produced. FDIA will escape the BDD because it does not affect $\|r\|_2$:

$$\|r_a\|_2 = \|z_a - H\hat{x}_a\|_2 = \|z + a - H(\hat{x} + c)\|_2 = \|z - H\hat{x} + (a - Hc)\|_2 = \|z - H\hat{x}\|_2 = \|r\|_2$$  \hspace{1cm} (3)

III. DEVELOPED FRAMEWORK

As noted previously, most current work on FDIAs uses time-invariant past data, rendering them more appropriate for stationary data. This is while data tends to fluctuate with time, and the data distribution is not static in smart grids in the event of integration of RESs. To address the problem and detect FDIAs in smart grids with RESs, a learning-based framework based on Bi-LSTM is proposed to differentiate between inherent uncertainties and an FDIA on each bus. This section summarizes some critical foundational principles on LSTM, Bi-LSTM and also utilized hyperparameters in the proposed framework.

A. Long Short-Term Memory (LSTM)

Recurrent neural networks (RNNs) are one of the well-known deep learning algorithms used to solve problems with sequential inputs. The main purpose of these techniques is to learn the long-term correlation between input data and to model them for future predictions. The modeling of large amounts of data over a long timeframe has problems that plague other conventional methods. However, despite all the benefits of RNNs, these methods suffer from some major problems such as gradient vanishing and explosion. The RNN-related problems were solved by introducing a deep architecture called long short-term memory (LSTM) [19].

LSTM is one of the powerful applications of deep learning, first introduced in 1997 by Hockeiter et al [20]. This network was able to improve the problems and limitations associated with RNNs as well. With the ability to retain information about previous situations in the training process, LSTM was able to demonstrate its superiority in processing data over a long period of time. Fig. 1 shows a structural schematic of an LSTM unit. As can be seen, the adjustment of this structure is based on input, forget, and output gates, and the signal flow in this architecture is performed by memory cell state blocks [20]. Each of these gates has unique functions, the process of calculating each variable at time $t$ is shown as follows [20]:

$$f_t = \sigma(W_{if}x_t + W_{mf}m_{t-1} + b_f)$$  \hspace{1cm} (4)

$$i_t = \sigma(W_{ii}x_t + W_{mi}m_{t-1} + b_i)$$  \hspace{1cm} (5)

$$o_t = \sigma(W_{io}x_t + W_{mo}m_{t-1} + b_o)$$  \hspace{1cm} (6)

$$a_t = \tanh(W_{ia}x_t + W_{ma}m_{t-1} + b_a)$$  \hspace{1cm} (7)

$$c_t = c_{t-1} + f_t a_t$$  \hspace{1cm} (8)

$$m_t = o_t \tanh c_t$$  \hspace{1cm} (9)
where $\sigma$ is the logistic sigmoid function, $f_t, i_t, o_t, c_t$, and $a_t$ demonstrates forget gate, input gate, memory cell, and hidden vector respectively. $W_{t_f} = W_{tf}, W_{t_o}, W_{t_i}$, and $W_{m_f} = W_{mf}, W_{mi}, W_{mo}$ represents trainable weights of the respective gates. $b_f, b_i, b_o$, and $b_h$ are output biases. Operator $\Phi$ shows the Hadamard product [20].

B. Bidirectional Long Short-Term Memory (Bi-LSTM)

Bi-LSTM is an improved version of the LSTM network that is used extensively to model large volumes of data and extract features from time-series data. Figure 3 shows the structural schematic of the Bi-LSTM, and unlike the traditional LSTM, the Bi-LSTM training process is such that the transfer of information in two directions is the feed forward (black line in Fig. 2) and vice versa (blue line in Fig. 2) is finished using the hidden state [21]. In this structure, at time $t$, the calculation of the hidden layer and the output layer in two directions is as follows:

\[
\begin{align*}
\overrightarrow{h_t} &= \sigma(\overrightarrow{W_t} \times x_t + \overrightarrow{V_t} \times \overrightarrow{h_{t-1}} + \overrightarrow{b}) \\
\overleftarrow{h_t} &= \sigma(\overleftarrow{W_t} \times x_t + \overleftarrow{V_t} \times \overleftarrow{h_{t+1}} + \overleftarrow{b}) \\
y_t &= \sigma(U[\overrightarrow{h_t}; \overleftarrow{h_t}] + c)
\end{align*}
\]

Fig. 2: The structural architecture of the Bi-LSTM network.

Table II

<table>
<thead>
<tr>
<th>Layer (parameter type)</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input (InputLayer)</td>
<td>33</td>
</tr>
<tr>
<td>Reshape (Reshape)</td>
<td>(1, 33)</td>
</tr>
<tr>
<td>Dropout (Dropout)</td>
<td>0.4</td>
</tr>
<tr>
<td>Flatten (Flatten)</td>
<td>44</td>
</tr>
<tr>
<td>Dense (Dense)</td>
<td>14</td>
</tr>
<tr>
<td>Sequence length</td>
<td>1</td>
</tr>
<tr>
<td>Fully Connected headen layer</td>
<td>1</td>
</tr>
<tr>
<td>Hidden unit</td>
<td>100</td>
</tr>
<tr>
<td>Output (Dense)</td>
<td>16</td>
</tr>
</tbody>
</table>

IV. NUMERICAL RESULTS

Two tests are developed and modeled to confirm the effectiveness of our methodology. Case I analyzes FDIA on measurements in a system without any RESs. Case II illustrates how the developed method can handle the integration of RESs into the network through injection of FDIA into system states after integration of wind farms.

A. Data Preparation

IEEE 14-bus test system is employed to accomplish the experiments utilizing MATPOWER [22] as shown in Fig. 3. To provide more accurate simulations, the load data utilized in the simulations is provided by NYISO [23]. Load profiles for 11 areas are available online every five minutes. To produce data for SE, load buses of the test system are associate with NYISO areas and then accommodate the pre-processed data into the MATPOWER test case. It is noteworthy that the dataset contains $T = 2045$ normal samples.

B. Attack Simulation

We introduce two unique FDAs into each attacked system state that are 90 percent and 110 percent of their real val-
ues, respectively. Specifically, false data is injected into each system state and multiple states $\theta_{2.7}$ and $\theta_{8.9}$ simultaneously. Active powers and system states are selected as features of the proposed method. Table III summarizes the characteristics of the generated dataset.

### C. Performance Evaluation Metrics

The evaluation of the proposed method is performed using $F_1$ score which its mathematical calculation of each of these indicators is as follows:

$$ F_1 = \left( \frac{2 \times p \times r}{p + r} \right), \quad (13) $$

where $p$ and $r$ denotes the precision and recall, respectively, and are calculated as:

$$ p = \left( \frac{TP}{TP + FP} \right), \quad r = \left( \frac{TP}{TP + FN} \right), \quad (14) $$

where true positive is represented as TP, FP shows the false positive and FN denote the false negative.

### D. Case I: FDIA Detection without RESs

In Case I, the network is imagined to function routinely until the 7th day. Therefore, there are six days of normal observations and the final day is fully replaced by attacks. To compare the proposed BiLSTM-based detection framework with current well-established FDIA detection algorithms, machine learning-based methods such as SVM, KNN and C4.5 are also trained on the generated dataset. Table IV shows the evaluation results for this case. As one can see, most methods were capable of successfully detecting all attack samples in this case when there is no integration of RES.

#### TABLE III

<table>
<thead>
<tr>
<th>Case</th>
<th>No. of features</th>
<th>No. of normal samples</th>
<th>No. of attack samples</th>
<th>No. of all samples</th>
<th>No. of class labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>33</td>
<td>2045</td>
<td>576</td>
<td>10686</td>
<td>16</td>
</tr>
<tr>
<td>II</td>
<td>33</td>
<td>2045</td>
<td>576</td>
<td>10686</td>
<td>16</td>
</tr>
</tbody>
</table>

### E. Case II: FDIA Detection in system with integration of RESs

The rising energy consumption needs the integration of RESs into the current electric grids [25]. However, the dramatic growth in incorporation of RESs increases the uncertainty leading to a transformation in the underlying distribution of collected samples and a reduction in accuracy of the current FDIA detection methods. To test the developed framework’s resilience in coping with incorporation of RESs into the system by conduction of FDIA after integration of wind farms as source of RESs. To this purpose, it is assumed that the network will continue to function properly through day four. Two wind farms and a total of 36 turbines are incorporated on the fourth day in order to supply at least 18 MW of the load of buses 3 and 4 as presented in Fig. 3.

Used wind turbines in the test system are identical in design and have a capacity of 1 MW, with a rated speed of 14 m/s, a cut-in speed of 4 m/s, a rated power of 1000 kW, and a cut-out speed of 25 m/s. It is noteworthy that data of wind speeds to complete the simulations is obtained from the Wunderground website for the Niagara Falls, New York [26]. A wind turbine’s output energy is measured as follows:

$$ v_f \leq v_{ci} \leq v_{co}, \quad v_f \leq v_{r}, \quad (15) $$

where $v_f$ represents the projected wind speed, $P_{rated}$ denotes the wind turbine’s rated power, $v_{ci}$ is the cut-in speed, $v_{r}$ represents the rated speed, and $v_{co}$ is the wind turbine’s cut-off speed.

For this case, we imagine that the attacker thoroughly targets the seventh day’s observations. This indicates the FDIA occurs 3 days after the wind turbines are installed. The results of running pre-built and saved ML models in Case I, on the dataset of this Case, are summarized in Table V. As one can see, when new RESs are integrated into the system, the pre-trained ML models which currently are utilized to detect FDIA, lose their ability to reliably classify samples and provide primary efficiency. The primary explanation for this decline is because the incorporation of RESs introduces natural oscillations into the data, raising the variance without introducing attacks that confuse current FDIA identification tools. This contrasts with the carefully developed FDIA detection framework, which recognizes FDIA and manages RESs integration effectively. This happens because the developed technique is precisely designed to properly characterize the underlying distribution of power grid measurements and overcome the uncertainties introduced by incorporation of distributed generations into
network. As a result, conventional approaches are unable to manage the dynamic nature of future power grids leading to inaccurate classification of observed values.

V. CONCLUSION

FDIAs can result in inaccurate state estimations and consequently wrong monitoring actions. Existing FDI detection techniques have been optimized for a particular system setup and overlook the reality of a constantly evolving and changing power grid due to the increasing renewable energy integration. Such enhanced uncertainties negatively affect the performance of existing FDI classifiers, resulting in erroneous labeling of fresh samples. This paper examined techniques to defend against FDI in power grids that RESs have impacted. Specifically, a framework based on Bi-LSTM algorithm is systematically built, which accurately captures the dynamic character of intelligent grids. The results show that the proposed model is robust to changes in smart grids and surpasses existing FDI detection strategies.

REFERENCES


[2] A. Shefaei, M. Mohammadpourfard, B. Mohammad-Ivatloo and Y. Weng, “Revealing a New Vulnerability of Distributed State Estima-
tion: A Data Integrity Attack and an Unsupervised Detection Al-


Autoencoder in Smart Grid,” International Journal of Critical Infra-


machine learning algorithms on capturing future distribution network 
anomalies,” IET Generation, Transmission & Distribution, vol. 13, no. 8, pp. 1441-1455Available


integrity attacks under concept drift,” International Journal of Electrical 

disturbances and cyber-attacks based on machine learning,” Journal of 


of false data injection attacks in smart grids using Recurrent Neural 
Networks,” IEEE Power & Energy Society Innovative Smart Grid 

data injection attack in industrial control system using long short term 
memory recurrent neural network,” 13th IEEE Conference on Industrial 

in Smart Grids: A Semi-Supervised Deep Learning Approach,” IEEE 

Randomized Trees-Based Scheme for Stealthy Cyber-Attack Detection 


and architectures,” IEEE Access, vol. 7, pp. 53040-53065, 2019, doi: 
10.1109/ACCESS.2019.2912200.

Networks Based on LSTM,” IEEE Transactions on Neural Networks 

[21] A. Moradzadeh, H. Moayyed, S. Zakeri, B. Mohammad-Ivatloo, and 
A. P. Aguiar, “Deep Learning-Assisted Short-Term Load Forecasting 

[22] R. D. Zimmerman and C. E. Murillo-Sanchez, “Matpower, a matlab 
power system simulation package,” http://www.pserc.cornell.edu/ mat-

[Accessed Jan 2021].

[24] H. Jiawei, J. Pei and M. Kamber, “Data mining: concepts and tech-

[25] T. Ahmad, H. Zhang, and B. Yan, “A review on renewable energy and-
electricity requirement forecasting models for smart grid and buildings,” 

Range & Local,” The Weather Company, LLC, [Online]. Available: 

TABLE V

<table>
<thead>
<tr>
<th>Technique</th>
<th>$F_1$</th>
<th>FP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bi-LSTM</td>
<td>0.99</td>
<td>1%</td>
</tr>
<tr>
<td>SVM</td>
<td>0.914</td>
<td>18.6%</td>
</tr>
<tr>
<td>C4.5</td>
<td>0.843</td>
<td>17.18%</td>
</tr>
<tr>
<td>kNN</td>
<td>0.911</td>
<td>2.7%</td>
</tr>
</tbody>
</table>

- FP: False Positive