Research on False Data Injection Attack Detection of Smart Grid Based on Machine Learning

Zhen Tao¹, *, Weishu Zhan²

¹University of bath, bath, BA2 7PD, United Kingdom
²University of Northern Iowa, Cedar Falls, 50618, United States

*Corresponding Author’s E-mail: zt374@bath.ac.uk

Abstract. The safe and reliable operation of power systems is an important guarantee for the healthy development of the national economy. Industry and people's lives are inseparable from electricity, so the safety and reliability of electricity supply is very important. The sudden interruption of power supply will not only bring serious economic losses, but also seriously affect people's normal lives and even endanger social stability. False data injection attack (FDIAs) are a new type of power system network attack method. FDIAs are a new type of power system network attack method. It can successfully bypass the bad data detection mechanism, offset the power measurement data, and mislead the control center under extremely subtle conditions. Therefore, it poses a very serious threat to the stable operation of the power system. Therefore, this article first analyzes the principle of false data injection attacks, in order to provide a theoretical basis for subsequent attack detection. Then this paper constructs a detection method based on extreme learning from the perspective of optimizing learning efficiency. Based on the IEEE14-bus standard test system, the method is verified through simulation, which shows the feasibility of this method, which provides a direction for building a safe and stable smart grid.

1. Introduction

In normal production and life, the power system is very important. The safe operation of the power system plays a crucial role in maintaining social stability and the normal development of the national economy. The development of information technology makes the power industry face many new problems and challenges. The new demand has caused a major change in the power system structure. The smart grid is a tremendous change, and various factors have prompted the research and development of smart grids in various countries. The smart grid system is a complex coupled network system. It is composed of a physical power system and an information communication system. Therefore, a smart grid system is a cyber-physical system (CPS). FDIAs are the attack method against EMS/SCADA. FDIAs are the new type of attack method in recent years. It exploits the loopholes of traditional bad data detection in the power system, and injects false data into the metering devices distributed in the power grid, resulting in a deviation of the state estimation results from the non-attack state. Therefore, the control center makes erroneous decisions based on the acquired erroneous system operating status. Therefore, FDIAs not only endanger the safe operation of the power system, but even cause the collapse of the entire power system. In the current power system, such attacks cannot be effectively detected, which is extremely harmful to the safe and stable operation of the power system. Therefore, the research on the detection mechanism of FDIAs is of great significance.
2. Analysis of the principle of FDIAs

In reality, the equipment may fail. Non-human factors such as incorrect connection or interference in the communication system, the measurement value collected by the SCADA system will have a certain degree of measurement noise, which will deviate from the actual situation and cause the state estimation result to be wrong. Ensure the reliability of the state estimation results and eliminate the errors caused by non-human factors, the largest normalized residual (LNR) method is widely used in power systems for processing.

\[ r = z - H(x) \]

The main basis for detecting bad data is: \( |r| < \tau \), \( \tau \) is the set threshold of bad data. If \( |r| < \tau \), the collected measurement value is normal data, otherwise it means that the collected data contains bad data. Therefore, the data needs to be further processed to eliminate bad data until it can pass the detection of bad data again.

FDIAs use the defects of this detection method. If \( a = [a_1, a_2, \cdots, a_m]^T \) represents the false data vector injected by the attacker in the measured value, the actual measurement data is \( z_{bad} = z + a \). The estimated state variable at this time is \( x_{bad} = \hat{x} + c \), \( c = [c_1, c_2, \cdots, c_n]^T \) represents the error vector introduced in the state variable due to the injection of false data. In this case, the residual expression is as follow.

\[ \|r\| = \|z_{bad} - Hx_{bad}\| = \|z + a - H\hat{x} + a - Hc\| \]

When \( a = Hc \), the following formula holds as follow.

\[ \|r\| = \|z_{bad} - Hx_{bad}\| = \|z - H\hat{x}\| \]

The traditional method of detecting bad data based on the residual equation cannot effectively find the tampered measurement data. In this way, an attacker can make arbitrary changes to the measured value, which affects the state assessment of the power system and causes serious losses.

3. Research on false data injection attack detection based on machine learning

3.1 Attack detection mechanism of machine learning

Because the new type of FDIAs are highly deceptive, the primary task of attack detection is to ensure detection accuracy. The traditional FDIAs detection idea is to analyze the residual of the objective function value through state estimation and analyze whether there is bad data in the system. This step is also an indispensable step of the SCADA system. Through operation in practice, it is proved that this detection method has strong feasibility, and most of the bad data caused by the failure of remote collection equipment or network attacks can be detected. But this detection method needs to set the threshold manually, and judge whether the system is safe by comparing the relationship between the residual and the threshold. If the threshold is too large, a large number of recall rates will be sacrificed. If the threshold is too small, many normal data will be removed, resulting in an increase in the false detection rate. No matter how small the threshold is set, FDIAs can always escape the detection of bad data detection modules, and many traditional detection methods fail in this case.

The classification method of machine learning has a unique advantage in dealing with such problems. According to the attack mechanism of false data injection, the data samples are divided into normal measurement data and the measurement data after being attacked, constructing labeled positive and negative data samples, and using machine learning classification methods to train the attack detection model. This method avoids the artificial selection of thresholds, and can effectively improve the detection accuracy of FDIAs by mining the relationships between data. The amount of measurement data in the power system is growing rapidly. With the increase in the amount of training data, the accuracy of the detection model can be improved. The mechanism of attack detection for machine learning is as follows.
Suppose a given power measurement data set containing positive and negative samples before and after the attack as follow.

\[ X = \{x_j, j = (1,2,\ldots,m) \} \quad (4) \]

There are the following classification tag values as follow.

\[ Y = \{y_j, j = (1,2,\ldots,m), y_j = \{-1,1\} \} \quad (5) \]

Then there is the training data set as follows.

\[ TraX = (x_j, y_j) \in X \times Y \quad (6) \]

Assuming that the test data to be judged is \( x'_j \) and the classification result is \( c_j \), it has the following relationship with the prediction function after training.

\[ c_j = f (x'_j) \quad (7) \]

The problem of detecting FDIAs can be transformed into the following relationship.

\[ c_j = \begin{cases} 
-1 & \text{if } a \neq 0 \\
1 & \text{if } a = 0 
\end{cases} \quad (8) \]

In formula (8), \( a \) is the attack vector. If \( a = 1 \), it means that the \( j \)th measurement vector has not been attacked. Otherwise, the \( j \)th measurement vector has been attacked. Figure 1 summarizes the attack detection process based on machine learning.

---

3.2 Detection technology based on ELM

There are many methods of machine learning, and neural networks are one of them. It is a calculation model, and its calculation function is realized by imitating the function and structure of biological neural network. Neural networks are calculated by connecting many artificial neurons. It has learning function and modeling function, and is nonlinear modeling. Neural networks usually consist of three parts, including structure, activation function and learning rules. Its structure defines the variables and their topological relations in the network, including the weights and excitation values of neurons. The neural network structure usually consists of three parts, namely: input layer, output layer and hidden layer.

The node and line power values are obtained through state estimation and used as input values, and the output values are calculated after the network. According to the numerical results, it is divided into attacked and non-attacked samples. The training is based on the samples. The purpose of training is to obtain network parameters, including weights, biases, and activation functions. The back-propagation algorithm is usually used to train the neural network. The training process is mainly divided into two
stages. The first stage is the forward propagation of information, which passes through the hidden layer from the input layer and finally reaches the output layer. The second stage is the back propagation of errors, from the output layer to the hidden layer, and finally to the input layer. Adjust the weight and offset of the hidden layer to the output layer in turn, and the weight and offset of the input layer to the hidden layer.

Extreme learning machine (ELM) is also called over-learning machine. It belongs to machine learning. It is an artificial neural network model. It is a learning algorithm for solving a single hidden layer feedforward neural network. When training a neural network, the back-propagation algorithm is generally used for training. The training process consists of two stages. The forward propagation of information is the first stage, starting from the input layer, passing the hidden layer in the middle, and then reaching the output layer. The error propagation in the opposite direction is the second stage, starting from the output layer, passing the hidden layer in the middle, and then reaching the input layer. Adjust the offset and weight, first adjust the hidden layer to the output layer, and then adjust the input layer to the hidden layer.

ELM randomly generates the input layer and hidden layer weights and offsets, and then obtains the output weights through analysis and calculation, which is generally obtained by the generalized inverse solution of the matrix. According to the neural network as follow:

\[
HV = Y
\]  
(9)

In formula (9), \( v \) is the output weight vector, \( H \) is the hidden layer output matrix combined with the input value, and \( Y \) is the label matrix used to set the classification label value. \( H \) and \( v \) can be expressed as follow.

\[
H = \begin{bmatrix}
  f(\omega_{11} \times x_1 + a_1)  & \cdots  & f(\omega_{1L} \times x_1 + a_L) \\
  \vdots  & \ddots  & \vdots \\
  f(\omega_{N1} \times x_N + a_1)  & \cdots  & f(\omega_{NL} \times x_N + a_L)
\end{bmatrix}
\]  
(10)

\[
v = \begin{bmatrix}
v_1^T \\
  \vdots \\
  v_L^T
\end{bmatrix}
\]  
(11)

The training step is divided into three steps. First, the weights and offsets from the input layer to the hidden layer are randomly assigned. Second, the training data is used as an input value to solve the output matrix of the hidden layer. Third, the weights from the hidden layer to the output layer are obtained according to the generalized inverse solution of the output matrix. After obtaining the output weights, the network parameters of the entire network can be obtained, so that it can be judged whether there is a false data attack based on the new input data.

### 3.3 Simulation analysis

The ELM algorithm is tested in the power flow data. The experimental data used in this article comes from the IEEE 14 node system. IEEE 14 is an open and widely used power test system. Assume that the load on each bus node in the power system is evenly distributed between 50% and 150% relative to the base load. After the power state estimation data is collected, a certain proportion of bad data and false data are injected into it according to the attack method shown in [11].
Each detection accuracy value in Table 1 is derived from the average value of 20 independent repeated tests. As shown in Table 1, we have considered the influence of the number of measured values that an attacker can tamper on the detection accuracy. With the increase of the number of hidden layers, the real-time false data detection mechanism proposed in this paper gradually increases the recognition rate of false data attacks. Therefore, this article sets the number of hidden layers in the ELM structure to 5 layers. Its purpose is to achieve a better balance between operation complexity and detection accuracy.

Table 1. The detection accuracy of ELM algorithm with different hidden layer levels under different number of disturbed buses

| Number of disturbed buses | 3 hidden layers | 4 hidden layers | 5 hidden layers |
|---------------------------|----------------|----------------|----------------|
| 32 | 93.91 | 94.58 | 95.57 |
| 41 | 96.19 | 96.57 | 96.75 |
| 48 | 96.77 | 96.93 | 96.97 |
| 57 | 97.95 | 97.97 | 97.97 |
| 32 | 97.68 | 98.05 | 98.11 |

From the measurement results presented in Table 2, when the value of the observation window is 5, the ELM mechanism can effectively detect false data attacks in the power system. The data obtained through simulation experiments are reasonable. When the observation window increases, more temporal information will be obtained, so the ELM mechanism can more effectively detect false data in the power system. Taking into account the trade-off between calculation accuracy and calculation complexity, we decided not to expand the ELM structure and maintain the size of the observation wound to 5.

Table 2. The detection accuracy of the false data and normal data by the ELM mechanism and the total detection accuracy

| Window size | True label | Data scale | False data | Normal data | Accuracy (%) | Total accuracy (%) |
|-------------|------------|------------|------------|-------------|--------------|-------------------|
| 3           | False      | 1291       | 1209       | 82          | 93.65        | 94.60             |
|             | True       | 1281       | 56         | 1224        | 95.55        |                   |
| 4           | False      | 1303       | 1245       | 58          | 95.55        | 95.51             |
|             | True       | 1258       | 57         | 1201        | 95.47        |                   |
| 5           | False      | 1292       | 1238       | 54          | 95.82        | 96.09             |
|             | True       | 1262       | 46         | 1216        | 96.35        |                   |
| 6           | False      | 1266       | 1206       | 60          | 95.26        | 94.27             |
|             | True       | 1278       | 86         | 1192        | 93.27        |                   |
This paper compares the detection accuracy of the three detection mechanisms, which are ELM and ANN detection and SVM detection in this paper. The test results are shown in Figure 3. According to the detection results in the figure, it can be seen that among the three different detection mechanisms, the ELM detection mechanism has obtained the highest detection accuracy. In addition, as the load curve k injected with false data increases from 32 to 64, the detection accuracy of the ELM detection mechanism exceeds 95%.

![Figure 3 The impact of the number of false data injected by the attacker on the load curve on the accuracy of the three detection methods](image)

4. Conclusion
The degree of intelligence of the power system continues to increase, and the possibility of being attacked by the network will be higher. Network attacks pose a greater threat to power systems. This paper has conducted some research on the detection of FDIAs. The attack detection method of machine learning has good generalization ability. This paper constructs the detection method of the station limiter uniform machine from the perspective of optimizing the learning efficiency. The experiment uses the IEEE 14-node test system for testing, and conducts various tests on the proposed attack detection method. The experimental results show that the two detection methods are effective and provide a direction for building a safe and stable smart grid.

Reference
[1] Tang Y, Chen Q, Li M.Y., et al. (2016) A Summary of Research on Network Attack in the Environment of Power Information Physics Fusion System[J]. Automation of Electric Power Systems, 40(17): 59-69.
[2] Zhu Jie, Zhang Gexiang, Wang Tao, et al. A Summary of Fraudulent Data Attack and Defense in Power System State Estimation[J]. Power System Technology, 2016, 40(08): 2406-2415.
[3] Zhao Junhua, Liang Gaoqi, Wen Fushuan, et al. Lessons from the Ukrainian incident: protection against false data injection attacks against the power grid[J]. Automation of Electric Power Systems, 2016, 40(7): 149-151.
[4] Patel S C, Sanyal P. Securing SCADA systems[J]. Information Management & Computer Security, 2008, 16(4):398-414.
[5] Wang Yiliang. Research on Detection and Prevention of False Data Attack in Smart Grid[D]. Beijing: Tsinghua University, North China Electric Power University, 2014.
[6] Li Min, Wang Gang, Shi Lei, et al. Risk analysis of smart grid information security[J]. North China Electric Power, 2017,01:62-65.
[7] Li Yuanchong, Yue Rixuan, Ceng Jing. Smart Grid Blind Online False Data Injection Attack Based on Nuclear Principal Component Analysis[J]. Power System Technology, 2018,42(07): 2270-2278.
[8] Liu X, Li Z. Local Load Redistribution Attacks in Power Systems with Incomplete Network Information[J]. IEEE Transactions on Smart Grid, 2014, 54(4): 1665-1676.
[9] Li S, Yilmaz Y, Wang X. Quickest Detection of False Data Injection Attack in Wide-Area Smart Grids[J]. IEEE Transactions on Smart Grid. 2017, 6(6): 2725-2735.

[10] Guan Y, Ge X. Distributed Attack Detection and Secure Estimation of Networked Cyber-Physical Systems against False Data Injection Attacks and Jamming Attacks[J]. IEEE Transactions on Signal & Information Processing Over Networks, 2018, 4(1): 48-59.

[11] KOUSUT O, JIA L, THOMAS R. J., et al. Malicious data attacks on the smart grid[J]. IEEE Transaction on Smart Grid, 2011, 2(4): 645-658.