Research on Network Anomaly Detection Based on Coordinated Multiple Algorithms

Yezhao Chen¹, Yingyan Jiang¹, Zhihua Yang¹, Yuanfeng Wang¹ and Peng Lin²

¹Guangdong Power Grid Co., Ltd. Power Dispatching Control Center, Guangzhou, China
²Beijing Vectinfo Technologies Co., Ltd, Beijing, China
²E-mail: linpeng@vectinfo.com

Abstract. With the development of power communication networks, network anomaly detection has become an important problem that must be solved to ensure its normal operation. In this paper, aiming at the one-sidedness of the anomaly detection algorithm, a network anomaly detection method based on coordinated multi-algorithm is proposed. Five modules are designed to detect abnormal flow, and the proposed method is verified experimentally. The experimental results show the effectiveness of the proposed method.

1. Introduction
Guangdong power grid has completed the integrated management and control system for transmission network in the general dispatch. Establishing a network-wide intelligent operation and maintenance system is one of the main problems to be solved.

This paper studies the anomaly detection method [8] of power communication network, and proposes a network anomaly detection method based on multi-algorithm coordination.

The structure of this paper is as follows: Chapter 2 analyzes the current research, Chapter 3 proposes an anomaly detection method based on coordinated multi-algorithm, Chapter 4 experimentally verifies the proposed method, Chapter 5 summarizes the full text.

2. Research Status
The existing research [3] collects flow statistics information, constructs the traffic matrix of the whole network. The method [4] includes three different BPCA methods.

Literature [5-7] extracts the network traffic characteristics of historical network traffic data, standardizes the network traffic data characteristics, and obtains a reduced feature subset, obtains SVM classifier.

Problems of existing methods are as follows:
1) How to combine the metrics of each algorithm?
2) How to set the threshold for determining anomaly?
3) How to reduce time complexity and space complexity?

3. Network Anomaly Detection Method Based on Coordinated Multiple Algorithm

3.1. General Framework
The general structure and flow of the network anomaly detection method based on coordinated multiple algorithm proposed in this paper is shown in Figure 1. The general framework is divided into five modules. The output of all GBDT modules is shown in Table 1.
Network traffic data

Flow rate measurement module

Hurst value measurement module

Entropy measurement module

Exponential smoothing module

GBDT module

Network Anomaly handling system

Figure 1. System operation flow chart

Table 1. GBDT module output value and meaning correspondence table

| Output of GBDT module | Network Status               |
|-----------------------|-----------------------------|
| 0                     | No anomaly                  |
| 1                     | DoS attack or DDoS attack   |
| 2                     | Network scanning attack     |
| 3                     | Port scanning attack        |

3.2. Flow Rate Measurement Module
The calculation method for this module is:

\[ v = \frac{l}{t} \]  \hspace{1cm} (1)

Where \( v \) is the network flow rate, \( l \) is the network traffic data size, and \( t \) is the time to pass the number of network traffic.

3.3. Hurst Value Measurement Module
Using self-similar method to detect DoS attacks has the advantage of detecting sudden attacks during the detection period.

3.4. Entropy Measurement Module
This module is mainly responsible for detecting attacks on network scanning and port scanning, and assisting in detecting DoS attacks. It can also detect worm attacks.

3.5. Exponential Smoothing Module
This module uses the exponential smoothing model to predict network traffic. The degree of network anomaly can be determined by comparing the predicted value and the actual value.

The output value of this module is the predicted fluctuation value and the predicted judgment value. The formula for the predicted fluctuation value is:

\[ v = \frac{(p - l)}{l} \]  \hspace{1cm} (2)

Among them, \( v \) is the predicted fluctuation value, the closer to 0, the smaller the fluctuation, \( p \) is
the predicted value of the current flow, and $l$ is the actual flow value. The formula for the predicted judgment value is:

$$j = \begin{cases} 
0, & v < \varepsilon \\
1, & v \geq \varepsilon 
\end{cases}$$

(3)

Among them, $J$ is the predicted judgment value, and $\varepsilon$ is the boundary of the judgment value. According to the data, this experiment takes it as 0.3.

This module cooperates with the Hurst value detection module. Formula (2) and Formula (3) show that when the larger data gap between the cycle before and after the attack, the larger value of $v$. When $v$ exceeds $j$, $\varepsilon$ becomes 1, and both outputs $v$ and $j$ have significant changes, which is a higher probability for judging it as a DoS attack.

3.6. The GDBT Module
This module uses the GDBT algorithm to train or detect whether an anomaly has occurred.

4. Experimental Verification
The experimental data is the data of China Southern Power Grid from July 18, 2017 to July 24, 2017. The data fields are separated by ",", the first line is about the meaning of these fields, as shown in Table 2, the rest of the data corresponds to the first line of fields.

| SourceIP | UsersID | uName | uID | uName | sPort | DestID | dPort |
|----------|---------|-------|-----|-------|-------|--------|-------|
| NetPro   | TranPRO | Appid | appName | AppID | AppName | RecoID | RecPort |

Table 2. The first line field of network data

The experimental results of each module are as follows.

4.1. Flow Rate Measurement Module
As shown in Figure 2, some of the time points in the picture (a) are higher, some of the time points are slightly higher, and most of the time points are lower. The picture (b) is the interception of part of the data on picture (a) for the better analysis of the data. (x-axis stands for time and y-axis stands for the flow)

Figure 2 (a). The analysis diagram of flow passes through the module

Figure 2 (b). Part of the flow analysis chart

4.2. Hurst value Measurement Module
As shown in Figure 3, some of the time points in the picture (a) are higher, and most of the time points are medium. The picture (b) is the interception of the data on the picture (a). (x-axis stands for time and y-axis stands for the hurst value)
4.3. Entropy Measurement Module
In the picture (a), the entropy values of all abnormal time points are significantly reduced. In the picture (b), the entropy value of the DoS attack time point is smaller than that of the network scan and port scan time points. In the picture (c), the amount of change cannot be accurately measured. As shown in picture (d), it is impossible to accurately determine whether an anomaly occurs. (x-axis stands for time and y-axis stands for the source IP entropy)
4.4. Exponential Smoothing Module
The output of the exponential smoothing module is the predicted fluctuation value and the predicted
decision value. As shown in Figure 5, the abnormal time point is the same as the flow rate
measurement module. The time point has a lower value at 300. The formula (2) that the predicted
value fluctuation at this time will be low. (x-axis stands for time and y-axis stands for exponential
smoothing model predictive value in figure 5a, exponential smoothing model decision value in figure
5b respectively.)

![Figure 5 (a). Exponential smoothing model predictive value map](image)

![Figure 5 (b). Exponential smoothing model decision value map](image)

4.5. GDBT Module and Conclusion
Table 3 shows that each indicator performs well and the model has a good effect. This model can be
used to detect network traffic anomaly.

Table 4 shows that the amount of data of the test data is large. At this time, the problem that the
model did not detect during the K-fold cross-validation was exposed.

Table 3. K-fold cross-validation verifies the reliability of each indicator protection.

|                | No attack | Dos | Port scan | Network scan |
|----------------|-----------|-----|-----------|--------------|
| **Precision Rate** | 0.99444444 | 1   | 0.9       | 0.98         |
| **Recall Rate**   | 1         | 1   | 0.85      | 1            |
| **F1 Score**      | 0.99714286| 1   | 0.86666667| 0.98888889   |
| **Confusion Matrix** | [343, 0, 0, 0] | [1, 28, 0, 18] | [0, 35] |
Table 4. Indicators of test data

|                      | [No attack] | Dos | Port scan | Network scan |
|----------------------|-------------|-----|-----------|--------------|
| Precision rate       | [0.98981052] | 0.99708333 | 0.9945 | 0.98840093  |
| Recall Rate F1 Score | [1. 1]      | 0.8484697 | 0.86666667 | 0.86666667 | 0.98049051 |
|                      | [0.99714286] | 1   | 1         | 0.98888889  |
|                      | [10351]     | 0   | 0         | 0            |
|                      | [0]         | 846 | 0         | 0            |
|                      | [88]        | 2   | 530       | 13           |
|                      | [19]        | 0   | 3         | 1048         |

5. Conclusion
Most of the anomaly detection methods of power communication network in our country use single algorithm. This paper proposes a network anomaly detection method based on coordinated multi-algorithm. The experimental results show that the proposed method can effectively detect the anomaly of the power communication network.

6. Acknowledgements
This work was financially supported by Research and Application on Intelligent Operation Management Technology in Voice Exchange Network (036000KK52160009).

7. References
[1] Bing F and Liangrui Tang. Vulnerability Analysis of Power Communication Network 2014 J. Chin. Soc. for. Elec. Eng. 34(7):1191-97.
[2] Xin M and Xi C. Security Architecture of Electric Power Communication Network 2012 J. Tele. For. Elec. Pow. Sys. 33(1):34-38.
[3] Dengyin Z and Jianfei L. Network Traffic Anomaly Detection Based on Relative Entropy 2012 J. Jour. of. Nan. Uni. of. Posts. And. Tele. 32(5):26-31.
[4] Yekui Q, Ming C, and Lixin Y. Network-Wide Anomaly Detection Method Based on Multiscale Principal Component Analysis 2012 J. Jour. Of. Soft. 23(2):361-377.
[5] Jiajia Z and Jia C. An Anomaly Detection Method Based on Entropy and SVM Multi-class Classifier 2016 J. Com. Tec. And. Dev. 26(3):31-35.
[6] Jie C. Research of Feature Reduction and Traffic Classification Method based on SVM 2017 D. Jil. Uni.
[7] Lin Y. Research on Network Traffic Anomaly Detection System Based on SVM 2010 D. Yun. Uni.
[8] Jie M and Jiaxin W. Research on anomaly detection algorithm of multi layer, data center network traffic 2017 J. Jour. of. He. Uni. Of. Eng. 29(1):62-66.