Security assessment of power data link based on dynamic weight change

Hongbin Zhu¹, ShuoCui¹, Ming Jin¹, Lina Jiang¹, Shasha Huang¹, Yifan Zhang²* and Wenting Zhang²

¹Big Data Center of State, Grid Corporation of China, Beijing, 100000, China
²Computer science and technology, North China Electric Power University, Beijing, 102200, China
*Corresponding author’s e-mail: 120192227091@ncepu.edu.cn.

Abstract. The Data link security assessment is to ensure the security of data transmission, data storage and data usage in the link. It is an important standard to reflect the safety of data link in time and accurately. We proposed a data link security assessment model based on dynamic weights. The model is evaluated from four aspects: intrusion detection, data content detection, data storage index detection, and data usage permission detection. For each aspect, we used the method of setting dynamic weights to score and evaluate. By setting dynamic weights, we can promptly reflect the potential safety hazards and remind the staff. At present, the security evaluation system for data links is not perfect. The model we proposed can evaluate and score from multiple aspects to reflect the security status of the link.

1. Introduction

The evaluation model is widely used in many fields. The evaluation model mainly reflects the problems of the data to be evaluated in time through scoring. It can be used in many fields such as quality assessment, status assessment and safety assessment. With the rapid development of mobile Internet technology and the arrival of the Internet and the information age in the 21st century, data has become the most important strategic core resource. Enterprise competition has entered the era of big data competition, requiring companies to efficiently integrate, centralize and manage data and operate scientifically to create huge value for the enterprise, promote the transformation of the enterprise, and win the competitive advantage in the market. With the rapid development of the ubiquitous power Internet of Things and the continuous advancement of the construction of data centers, the data resources possessed by the power system are constantly enriched. However, in the data link monitoring, there are still many problems in data management. By establishing a power data link security assessment system, better security protection of power system data is given.

This paper proposes a new power data link security assessment model based on dynamic variable weights, which can reflect the security issues in the data link in time. First of all, we divide this model into four modules for testing by data type classification. The data of different modules have different types of characteristics. We have set a corresponding variable weight scoring method for the modules. This method of scoring can strengthen the attention to important indicators, and the deduction table and the error show a proportional relationship. After obtaining the evaluation scores of the four modules, we will again use the dynamic weighting method to obtain the overall score of the link security. This method focuses on key issues as the main goal. The lower the evaluation value, the
greater the weight, and serious problems will be given serious warning. The experimental analysis shows that our method can better remind the staff than the currently used evaluation model.

2. Related work
The currently used evaluation model algorithms generally use analytic hierarchy process, order relationship analysis and entropy method to set the weight and score. Xiao[1] proposed the use of multiple correlation coefficient weighting method for weight assignment. Jia[2] et al. proposed the fuzzy analytic hierarchy process for safety assessment. Ma[5] uses the sequence relationship analysis method to evaluate the energy saving of power plants. The weight values of these methods cannot be changed at any time according to the actual situation.

Currently, in terms of intrusion detection, more methods are used based on traditional machine learning and deep learning, and the results of using these methods for detection are not bad. Alom[3] and others used DBN to conduct experiments on the intrusion detection data NSL-KDD, and they obtained 97.5% accuracy using only 40% of the training data. Fiore[4] and others proposed the use of a semi-supervised restricted Boltzmann machine for anomaly detection, and they obtained 96% accuracy. The accuracy of these methods can reach a good standard, but there is still room for improvement[6-7].

3. Sections, subsections and subsubsections
In this part, we will introduce in detail the data link security assessment model based on dynamic variable weight. In the first part, we respectively introduce the content and indicators of the four detection modules. In the second part, we focus on the variable weight scoring system. Finally, we introduce the security assessment system that integrates the four modules in the data link.

3.1. Module detection basis
In the detection part of the data link security assessment model, we divided it into 4 parts, namely intrusion detection, data content detection, data storage indicator detection, and data usage permission detection.

For the intrusion detection part, we selected the auto-encoder for intrusion detection after experimental comparison.

For the data content detection part, we use a string matching algorithm to detect the data in the data link. The purpose of string matching is to detect whether the sent data are a set data type, and to detect the existence of illegal data. The intruder may implement attack behaviors by inserting some attack data into the transmitted data, and performing string data inspection on the transmitted data can provide good security protection.

For the data storage indicator detection part, we set some important security indicators according to the data storage specification requirements, which include the compliance of the data storage path, the compliance of the data storage backup situation and more. These indicators are tested by data storage systems and databases, and compliance is "1" and non-compliance is "0". Through this detection of data storage behavior is reasonable, to prevent attackers from intruding through the data storage stage.

For the data usage permission detection part, we mainly detect user ID and user permissions, and detect whether there are illegal users using data and whether users are performing illegal operations.

The above is the detection basis for the four parts of the data link security evaluation model. For the detection results, we will set weights and score accordingly.

3.2. Module scoring system based on variable weight
The module scoring system scores the corresponding modules, and we have established different scoring methods according to the particularity of the data characteristics. We divide data content detection and data storage index detection into one group and use one scoring standard. We divide intrusion detection and data use permission detection into one group and use one scoring standard.
First, let's introduce the scoring standards used in data content detection and data storage index detection. Divide the two parts into one group and use one scoring standard because they detect more indicators, and each indicator will output a test result. An error in one indicator will not cause particularly serious harm like the other two parts. For this part of the scoring, we use the weighted arithmetic increase method.

The weighted arithmetic increase method is simply to set a unique weight for each error. The more the number of errors, the greater the weight, the more points will be deducted. We first set the upper and lower limits of the scoring system. For the upper limit, we set a full score of 100 points, and the first error will be deducted 40 points. From the second error, we will use the weighted arithmetic increase method to deduct points. The reason why we set the first error deduction to a fixed value is that we want to make the difference between right and wrong more obvious, so as to better remind the staff to correct the error. For the lower limit, we set the total number of errors not to exceed 50% of the total number of indicators. When there are too many errors, we will directly score zero to alert the staff. Then remove the first error and the error exceeding the specified number, and we will use the weighted arithmetic increase method to score the remaining number of errors.

We take the data storage indicator as an example to make a detailed introduction. First, we set a basic deduction weight \( W_2 \), and set the value of \( W_2 \) to 1. This weight is the weight of the deduction for the second error. When the 3rd, 4th,..., nth error occurs, \( W_n = n-1 \), the weight of the deduction for each error increases with the increase of the error number. Let the deduction base corresponding to the deduction weight be \( S_2 \). The base deduction algorithm is shown in formula 1. \( S_2 \) and \( W_2 \) correspond to each other, and \( S_2 \) as the base for deduction also represents the deduction for the second error. The data storage index has 10 detection indexes, the maximum number of errors is 5, and if there are more than 5 errors, it is zero. Except for the first error, the weights and deducted points for other errors are shown in Table 1.

\[
S_2 = (100 - 40)/(W_2 + W_3 + \ldots + W_n)
\]

### Table 1. Deduction weight and deduction.

| Deduction weight | Deduction |
|------------------|-----------|
| Second error     | 1         | 6         |
| Third error      | 2         | 12        |
| The fourth error | 3         | 18        |
| The fifth error  | 4         | 24        |

Next, we will introduce the scoring standards used in intrusion detection and data use permission detection. Since these two parts are very important to data link security, their scoring system is also very strict, and their scores are only 0 or 100. If problems occur, they must be discovered and resolved in time. If it is normal during intrusion detection, give 100 points, and directly give 0 points when an intrusion occurs. If there is no violation of the data usage permission test, 100 points will be given, and 0 points will be directly given if there is a violation.

### 3.3. Data link security assessment model based on dynamic variable weight

Next, the scoring of the four parts needs to be integrated to form the final scoring of our evaluation system. This scoring system uses a dynamic variable weight method for scoring. We set a weight value for the scores of the four detection modules for data content detection, data storage index detection, intrusion detection, and data usage permission detection, respectively, \( a_1 \), \( a_2 \), \( a_3 \), and \( a_4 \). We set the scores obtained by the four modules as \( t_1 \), \( t_2 \), \( t_3 \), and \( t_4 \). We set the final score as \( T \). We hope that the weight values of the first two modules can change with the changes in the scores obtained by the modules, so that problems can be reflected to the staff in time. So we set the weight value of each part to 100 minus the score of each part, so that the lower the score, the higher the weight, and the problem can be reflected in time. The latter two modules are given higher importance. When there is a problem, the total score is directly 0. If there is no problem, it will not participate in the final score. We set the state of the latter two modules when there is a problem, in this case the total
score $T=0$. When there is no problem with the latter two modules, the algorithm we use is shown in formula 2.

$$T = \left[ a_1 \ast (100 - t_1) + a_2 \ast (100 - t_2) \right] / \left[ 100 \ast (200 - t_1 - t_2) \right]$$  \hspace{1cm} (2)

## 4. Experiment

In this section we will show the experimental details in detail. First, we conduct experiments on the KDD99-based intrusion detection data set, and select the best detection algorithm through experiments. Then we input the relevant data collected from the electric power big data center into our model for evaluation, and compare it with the commonly used analytic hierarchy process, order relationship analysis method and entropy weight method.

### 4.1. Intrusion detection experiment

In recent years, machine learning methods have been used a lot in intrusion detection, and the effect is very good. In order to select better machine learning methods for intrusion detection, we compare them through experiments. We chose logistic regression, naive Bayes, KNN, decision tree, random forest, BP neural network and auto-encoder for experimental comparison. The algorithm selection is made by comparing accuracy, precision, recall and f1 score.

### 4.2. Data link security assessment model experiment based on dynamic variable weight

We collected horizontal and vertical data from the electric power big data center for security assessment and scoring. For the detection of data content, we collected 21 data detection indicators such as MID, CLUSTER\_AREA, UPDATED\_TIME, etc. For data storage indicator detection, we collected the 10 indicators mentioned in Section 3 as the detection indicators. For intrusion detection and data usage permission detection, we also collected relevant indicators for security assessment. We will use these indicators to compare the final safety assessment scores through the analytic hierarchy process, order relationship analysis, entropy method and our proposed model.

| Experimental algorithm | Accuracy (%) | precision (%) | recall (%) | f1 score (%) |
|------------------------|-------------|---------------|------------|-------------|
| Logistic regression    | 84.8        | 97.5          | 82.1       | 88.3        |
| Naive Bayes            | 92.9        | 96.5          | 94.3       | 94.5        |
| KNN                    | 92.1        | 97.4          | 91.3       | 95.4        |
| Decision tree          | 93.0        | 98.1          | 91.4       | 95.5        |
| Random forest          | 92.7        | 96.9          | 91.0       | 95.2        |
| BP neural network      | 98.6        | 97.9          | 95.4       | 96.3        |
| Auto-encoder           | 98.8        | 98.3          | 96.8       | 97.8        |

### 4.3. Experimental results

In this part, we first show the experimental comparison results based on the KDD99 intrusion detection data set, and then show the comparison results between our proposed model and other evaluation and scoring methods.

#### 4.3.1. Intrusion detection experiment results

We use the KDD99 data set as an experimental data set for experiments. Through the use of logistic regression, naive Bayes, KNN, decision tree, random forest, BP neural network and auto-encoder, the experimental results in Table 2 are obtained.

By comparing the results of accuracy, precision, recall and f1 score, we found that the accuracy of the detection results obtained by the auto-encoder algorithm is significantly better than other algorithms, so we use the auto-encoder algorithm for the intrusion detection part of our proposed model.
4.3.2. Experimental results of data link security assessment model based on dynamic variable weight.

We input the data collected from the electric power big data center to generate data sets into the analytic hierarchy process, order relationship analysis method, entropy weight method and our proposed model for comparison. We set data content detection, data storage index detection, intrusion detection, and data use permission detection as part one, part two, part three, and part four, respectively. Since the analytic hierarchy process, the order relationship analysis method and the entropy weight method adopt a unified weight setting and are not divided into four parts for weight setting and scoring like our proposed model, we first use Table 3 to carry out these three methods.

| Evaluation algorithm                        | The first part of the number of errors | The second part of the number of errors | The third part of the situation | The fourth part of the situation | Evaluation score |
|---------------------------------------------|----------------------------------------|----------------------------------------|---------------------------------|---------------------------------|-----------------|
| Analytic Hierarchy Process                  | 2                                      | 3                                      | normal                          | normal                          | 62              |
| Analytic Hierarchy Process                  | 0                                      | 4                                      | normal                          | normal                          | 71              |
| Analytic Hierarchy Process                  | 0                                      | 1                                      | normal                          | unnormal                        | 20              |
| Order relationship analysis                 | 2                                      | 3                                      | normal                          | normal                          | 52              |
| Order relationship analysis                 | 0                                      | 4                                      | normal                          | normal                          | 64              |
| Order relationship analysis                 | 0                                      | 1                                      | normal                          | unnormal                        | 22              |
| Entropy method                              | 2                                      | 3                                      | normal                          | normal                          | 89              |
| Entropy method                              | 0                                      | 4                                      | normal                          | normal                          | 98              |
| Entropy method                              | 0                                      | 1                                      | normal                          | unnormal                        | 23              |

We will show the error situation in each of the four parts. Since these three methods are not divided into parts for scoring, we only make the final score based on the algorithm, and we will round down the final score. In Table 4 we show the results of scoring using our proposed model. Comparing the results in Table 3 and Table 4, it can be found that when data usage permissions and intrusion detection are flawed, the lower level skips access to the upper level content, and the data are illegally invaded and modified, it can be directly judged as 0 points, which better reminds the staff to find errors. The staff's slack is caused by the high score. The method can clearly show the problem.
Table 4. Experimental results of dynamic variable weight method.

| Evaluation algorithm | The first part errors | The first part score | The second part errors | The second part score | The third part situation | The third part score | The fourth part situation | The fourth part score |
|----------------------|-----------------------|----------------------|------------------------|-----------------------|--------------------------|---------------------|--------------------------|-----------------------|
| 1                    | 2                     | 58                   | 3                      | 42                    | normal                   | 100                 | normal                   | 100                   |
| 2                    | 0                     | 100                  | 4                      | 24                    | normal                   | 100                 | normal                   | 100                   |
| 3                    | 1                     | 60                   | 6                      | 0                     | unnormal                 | 0                   | normal                   | 100                   |
| 4                    | 0                     | 100                  | 1                      | 60                    | normal                   | 100                 | unnormal                 | 0                     |

5. Summary
How to detect and discover existing security problems in time during the transmission of power data through the data link is extremely critical. In order to model and solve this problem, we propose a power data link security assessment model based on dynamic variable weights. Compared with the regular evaluation and scoring model, this model can feedback the problems to the staff in a more timely manner. We divide the model into four parts for scoring, and finally we will do total scoring. In the final scoring, we use a dynamic variable weight method to set the weight. The part with a low block score will occupy a large weight. The lower the score, the greater the weight, which reflects the safety problem in time. The experimental results verify the advantages of our proposed power data link security assessment model based on dynamic variable weights compared with common models.

References
[1] Xiao, X.(2014)Urban rail transit network operation risk and safety assessment. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CDFD&dbname=CDFD1214&filename=1014369298.nh?v=Zkw%25mmd2Blj4iL8Hx6ccFStATM96lMd0E4ziQT8R0b5En9uHRbg0EFxhYX65HNe5VKokr.
[2] Jia, C.Q., F, D.Q. (2016)Safety assessment of industrial control system based on fuzzy analytic hierarchy process. J. Journal of Zhejiang University., 50(4): 759-765.
[3] Alom, M.Z.,Bontupalli, V.R.,Taha, T.M. (2015) Intrusion detection using deep belief networks. C. 2015 National Aerospace and Electronics Conference (NAECON). IEEE, pp.339-344.
[4] Fiore, U.,Palmieri, F., Castiglione, A., et al. (2013) Network anomaly detection with the restricted Boltzmann machine. J. Neurocomputing, pp.122:13-23.
[5] M, Y. (2014)Energy-saving evaluation of Mawan Power Plant based on order relational analysis method. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CMFD&dbname=CMFD201501&filename=1014366272.nh?v=Zkw%25mmd2Fmnd2F%25mmd2FchzWSzi1c4nzuMfU9NGVT96H6wm%25mmd2FV4RFJo9o4nT8SocvU1ppicvMIC.
[6] Bo, D., Wang, X.. (2016) Comparison deep learning method to traditional methods using for network intrusion detection, In Proceedings of the 2016 8th IEEE International Conference on Communication Software and Networks (ICCSN), IEEE. Beijing. pp.581–585.
[7] Alves, T., Das, R., Morris, T. (2018) Embedding Encryption and Machine Learning Intrusion Prevention Systems on Programmable Logic Controllers. J. IEEE Embed. Syst. Lett., 10:99–102.