Computer Vulnerability Detection and Early Warning Method Based on Data Mining Technology

Tianyu Ren, Xiaohu Wang, Jiahan Dong, Guangxin Guo, Chao WANG

State Grid Beijing Electricpower Company Electric Power Research Institute, Beijing, China

Keywords: Data mining technology, Computer vulnerabilities, Vulnerability screening and early warning, Apriori algorithm

Abstract: Computer security is a problem that has been attached great importance to in the process of computer development. Only by developing a new detection vulnerability screening and early warning technology can we ensure the security of computer use. Therefore, this paper studies the computer vulnerability detection and early warning method based on data mining technology. Based on the Apriori algorithm of data mining technology, a hybrid detection method is proposed in this paper, which is based on data mining technology. In order to verify the method in this paper, we carry out the simulation experiment through the experimental simulation platform, and select the other three methods as the control, and analyze the detection time, detection rate, false alarm rate and false negative rate. The results show that the detection accuracy of static detection technology is 63.45%, false positive rate is 36.55%, false positive rate is 15.23%, dynamic detection technology detection accuracy rate is 71.64%, false positive rate is 28.36%, false positive rate is 14.32%, binary comparison technology detection accuracy rate is 82.36%, false positive rate is 17.64%, false positive rate is 11.63%, detection accuracy of this method is 95.72%, false positive rate is 95.72% 28% and 7. 32% respectively. From the simulation test results, we can see that the computer vulnerability detection and early warning technology based on data mining technology can ensure the running time and detection rate, and control the false detection rate strictly and accurately detect the results, so as to achieve the best detection effect.

1. Introduction

The rapid development of computer technology has greatly changed people's life, making people's life and work more efficient [1-2]. However, any kind of thing has two sides. While the computer is convenient for people's life and improving work efficiency, there are also computer security problems. Computer security has always attracted people's attention [3-4].

With the continuous emergence of computer vulnerabilities, computer vulnerabilities become an important factor endangering computer security issues, creating opportunities for illegal computer attackers [5-6]. Many attackers take a variety of attack methods to solve the problem of computer security vulnerabilities, and destroy computer security. In view of this situation, it is necessary to accurately grasp the performance and characteristics of computer security vulnerabilities, and take corresponding vulnerability screening and early warning technology to eliminate the security problems caused by computer security vulnerabilities. It is of great significance for computer security to study a method that can quickly check and warn computer vulnerabilities.

Apriori algorithm is the basis of quantifying association rules, which is often used in computer vulnerability ranking. Therefore, in the research, this paper first describes the Apriori algorithm, which is commonly used in computer vulnerability screening and early warning, and introduces the connotation of computer vulnerability. In addition, for computer vulnerabilities, this paper compares a variety of different computer vulnerability detection and early warning methods, combined with its advantages, proposes a computer vulnerability detection and early warning method based on data mining technology, and carries out simulation experiments through the experimental simulation platform, and analyzes the detection time, detection rate, false alarm rate and false negative rate. The results show that the hybrid detection technology proposed in this paper has better performance.

  DOI: 10.25236/mmmce.2021.008

2. Apriori Algorithm and Computer Vulnerability

2.1 Priori Algorithm

At present, big data mining technology contains a variety of different types of mining algorithms, including classification algorithm, clustering algorithm, association rule algorithm and so on. In practical application, the selection and use of specific algorithms are mainly determined by the object target, so as to achieve the predetermined data analysis and mining results. Clustering is the basis of Apriori algorithm for quantifying association rules, which is often used in computer vulnerability ranking. Apriori algorithm, on the one hand, is to cluster one or a pair of quantitative attributes to form a qualified cluster or interval; on the other hand, it searches frequent clusters to obtain quantitative association rules based on distance. Therefore, the related definitions include: the first step of the cluster feature (CF), association cluster feature (ACF) and other definitions, and the second step of the combination of cluster formation rule definition.

Definition 1: clustering feature (CF)

CF is used to describe the information aggregation of object sub clustering, and includes the cluster information projected on other attribute sets. The formula is as follows:

$$CF(C_X)=\left(N,\sum_{i=1}^{N}t_i[X],\sum_{i=1}^{N}t_i[X]^2\right) \quad (1)$$

In the above formula, N represents the number of all tuples in the sub cluster.

Definition 2: associated cluster feature (ACF)

Assuming $C_X=t_1,t_2,...t_n$, the ACF formula is as follows:

$$ACF(C_X)=\left(N,\sum_{i=1}^{N}t_i[Y],\sum_{i=1}^{N}t_i[Y]^2\right) \quad (2)$$

Definition 3: cluster density evaluation

Let S[X] be the set of N data sets $t_1,t_2,...t_N$ projected to attribute set X. Then the distance measurement formula of S[X] is:

$$d(s[X])=\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\delta_X(t_i[X],t_j[X])}{N(N-1)} \quad (3)$$

In the above formula, δ represents the distance measure between tuples. The larger the distance degree of S[x], the greater the deviation of projection of data set to attribute set X.

Definition 4: partition of interval rules

A cluster C on attribute set X should be less than or equal to the density threshold, and d_0^X should be greater than or equal to the frequency value S_0.

$$d(C_X[X])\le d_0^X \quad (4)$$

$$C_X\ge S_0 \quad (5)$$

In the above formula, S_0 defines the minimum number of tuples in the cluster. When the data clusters satisfying the density distribution are more dense, the data clusters meeting the frequency threshold will have enough support.

2.2 Computer Vulnerabilities

Computer vulnerability refers to the operation security problems caused by unsafe factors in the process of software development by computer software developers. The unsafe factors are diverse, including the factors at the hardware level. Since the hardware development does not reach the corresponding technical level and is related to the software itself, any software development can not be perfect, In addition, the protocol is also an important factor of computer vulnerability.

3. Simulation Experiment Design

3.1 Simulation Platform Environment

The simulation platform uses VC + + programming language and m icrosoftVisualStudio.NET2008VC. net platform.

3.1.1 The Configuration of the Computer Used in the Experimental Simulation is as Follows:

Operating system: Microsoft Windows XP SP2;

CPU:IntelPentiumDualCPUE2200@2.20GHz;

Memory: 1037420kb;
Hard disk: 640g;

3.1.2 The Running Environment of the Instance Program is Configured as Follows:

Client program:
Operating system: Microsoft Windows XP SP2;

CPU:IntelPentiumDualCPUE2200@2.20GHz;

Memory: 1037420kb;
Hard disk: 640gb;
Server program:
Operating system: Microsoft Windows 2003 SP2;

CPU:IntelPentiumDualCPUE2200@2.20GHz;

Memory: 2GB;
Hard disk: 1TB;

3.2 Design of Simulation Experiment

In this paper, three simulation experiments are carried out on the simulation platform. The static detection technology, dynamic detection technology, binary comparison technology and the hybrid detection technology proposed in this paper are analyzed from the detection time, detection rate, false alarm rate and false alarm rate.

4. Computer Vulnerability Detection and Early Warning Method Based on Data Mining Technology

4.1 Analysis of Computer Vulnerability Detection and Early Warning Method Based on Data Mining Technology

4.1.1 Static Detection Technology

Static detection technology is mainly carried out by manual detection, which is completed by inspectors. If necessary, static detection can be implemented with the help of source code analysis device. However, this link needs to obtain the support of source code, which can be realized by analyzing the source code. Therefore, in order to use the static detection technology in this link, we need to have two conditions: one is to establish the source code feature database, the other is to establish the rule base. Therefore, it is necessary to improve the effectiveness of static detection technology according to the essential basic conditions of static detection technology.

4.1.2 Dynamic Detection Technology

Static detection technology for computer security vulnerabilities detection and early warning needs to have a prerequisite, that is, the object system source code acquisition, the realization of this premise is very difficult. Through the analysis of the computer running environment to detect computer security vulnerabilities. Compared with static detection technology, dynamic detection technology is more convenient and accurate, but the detection efficiency is not high. Because the computer system is not single, this results in the dynamic detection technology can not be unified scanning, resulting in the limited scope of application of dynamic detection technology, the application of large software security vulnerability detection is limited.

4.1.3 Binary Comparison Technology

Binary comparison technology is very different from other vulnerability mining technologies. In particular, by specifying the differences between the previous binary code and installing the updated version, the binary comparison technology can be used to analyze the vulnerability This technology is also called code matching technology.

4.2 Analysis of Simulation Experiment Results

In this paper, three simulation experiments are carried out on the simulation platform, and the detection time, detection rate, false alarm rate and false alarm rate are analyzed.

4.2.1 Analysis of Computer Vulnerability Detection Time Based on Data Mining Technology

This paper first analyzes the detection time of the three tests, and the results are shown in Table 1.

Table 1 Analysis of The Time of Early Warning and Detection of Computer Vulnerability Detection Based on Data Mining Technology

Detection time	First test	Second test	The third test
Static detection technology	34s	36s	37s
Dynamic detection technology	30s	32s	34s
Binary comparison technology	25s	26s	25s
Hybrid detection technology	11s	10s	12s

In this paper, three simulation experiments are carried out on the simulation platform, and then the detection time is compared. It can be seen from Table 1 and Figure 1 that the detection time difference of each computer vulnerability screening and early warning method is small in the three tests, and there are obvious differences among different methods. The detection time of static detection technology on the first detection is 34s, the second detection time is 36s, the third detection time is 37s, and the detection time of dynamic detection technology on the first detection is 30s, and the second detection time is 30s The detection time is 32s, the third detection time is 34s, the detection time of binary comparison technology on the first detection is 25s, the second detection time is 26s, and the third detection time is 25s. The detection time of the hybrid detection technology proposed in this paper is 11S on the first detection, 10s for the second detection, 12s for the third detection In view of the above, the hybrid detection technology proposed in this paper consumes less time in computer vulnerability screening and early warning, and can carry out computer vulnerability early warning more quickly.

4.2.2 Analysis of Early Warning Detection Rate, False Positive Rate and False Negative Rate of Computer Vulnerability Investigation Based on Data Mining Technology

In addition to analyzing the detection time of computer vulnerability detection based on data mining technology, this paper also analyzes the detection rate, false alarm rate and false negative rate of the four methods. The results are shown in Table 2 and Figure 2.

Table 2 Analysis of Early Warning Detection Rate, False Positive Rate and False Negative Rate of Computer Vulnerability Investigation Based on Data Mining Technology

Computer vulnerability investigation and early warning method	Detection rate	False positive rate	Omission rate
Static detection technology	63.45%	36.55%	15.23%
Dynamic detection technology	71.64%	28.36%	14.32%
Binary comparison technology	82.36%	17.64%	11.63%
Hybrid detection technology	95.72%	4.28%	7.32%

In conclusion, it can be seen that the hybrid detection technology proposed in this paper can obtain a high detection rate, and the detection time is relatively short, which can guarantee the detection effect of computer vulnerability screening and early warning is satisfactory.

5. Conclusions

Computer security vulnerability is an inevitable problem, so it is very necessary to research on computer vulnerability detection and early warning technology. Computer security vulnerability detection and early warning is the basic premise to ensure computer security and play computer functions. Therefore, this paper proposes a computer vulnerability detection and early warning method based on data mining technology, and studies it. In addition, this paper believes that the computer vulnerability detection and early warning technology is not immutable. It is necessary to comprehensively consider various causes of vulnerabilities according to the basic types and characteristics of vulnerabilities. Therefore, it is necessary to further improve the computer security vulnerability detection technology, and adopt scientific security vulnerability investigation and early warning technology for different research objects and different situations, Static detection technology, dynamic detection technology and hybrid detection technology are flexibly applied to computer security vulnerability detection and early warning.

References

[1] Quach A, Wang Z, Qian Z. Investigation of the 2016 Linux TCP Stack Vulnerability at Scale. ACM SIGMETRICS Performance Evaluation Review, Vol.44, No.1, pp.8,.2017.

[2] Bochkov A V. Vulnerability assessment methodology and some methodical aspects of critical infrastructure protection. International journal of systems assurance engineering and management, Nov.Supplpll1, No.10, pp.s45-s57, 2019,

[3] Adewoyin O O , Joshua E O , Akinyemi M L , et al. Investigation to determine the vulnerability of reclaimed land to building collapse using near surface geophysical method. Journal of Physics: Conference Series, Vol.852, No.1, pp.6, 2017,.

[4] Malik H , Gjomemo R, Venkatakrishnan V N, et al. Remote Check Truncation Systems: Vulnerability Analysis and Countermeasures. IEEE Access, Vol.1, No.1, pp.99, 2020

[5] Zhao W, Cai Y, Li Z, et al. Injury prediction and vulnerability assessment using strain and susceptibility measures of the deep white matter. Biomech Model Mechanobiol, Vol.16, No.1115/1, pp.1709-1727. 2017,

[6] Gerin M I, Gerin V B, Blair J, et al. A neurocomputational investigation of reinforcement-based decision making as a candidate latent vulnerability mechanism in maltreated children. Development & Psychopathology, Vol.29, No.10, pp.1689. 2017,