Computer Network Security under the Background of Big Data

Fei Chen
School of Big Data, Chongqing Vocational College of Transportation, Chongqing, China

*Corresponding author e-mail: qiangjiahang@cqjy.edu.cn

Abstract. With the advent of the era of big data, data has exploded. This paper mainly studies the analysis of computer network security under the background of big data. In order to make these data intuitively reflect the network security situation, this paper uses the method of weight allocation to normalize the five groups of data, and allocates a larger weight to the security problems with high network security threat and a small weight to the security problems with low network security threat. In this paper, 3603 sample data are randomly selected as evaluation data according to the proportion of the type of data set unchanged. After setting the parameters, you can train and test. According to the risk level, take corresponding control measures to control the system risk at an acceptable level. The data shows that a certain number of attacks are carried out on the protected network in the 5th, 12th, 20th, 28th, 36th, 44th and 52nd minutes, and the scale of each attack is twice that of the last attack. The results show that the construction and deduction method of network security event scenarios can provide effective support for the research of network security incident coping strategies.

Keywords: Big Data; Computer Network Security; Model Analysis; Risk Assessment

1. Introduction
Big data technology can comprehensively utilize multi-source and massive data resources to mine the time and space rules of passenger travel and obtain travel information. In addition; deep learning has the advantages of continuously learning a large number of sample data and extracting non-linear; unstable and highly uncertain data characteristics. At the same time; it can comprehensively consider the impact of social and economic factors; temporal and spatial factors; and various emergencies on passenger flow changes; and provide advanced technology and advanced technology for improving the accuracy of passenger flow prediction under different environments in urban rail transit passenger flow prediction data support.

With the gradual development of network attacks; especially the recent encryption tunnel VPN technology; attack concealment tips; and the use of mail attacks; web viruses and other ways; so that the network internal penetration ability has been enhanced [1]. In this case; although the boundary is still secure; it does not mean that the internal network is also secure [2]. The security situation of the whole network will not be completely consistent with the security situation of the network [3-4].
the whole; the existing solutions to network security threats can not fully meet the needs of the public [5-6]. In addition; the false reports and missing reports of network security equipment make it difficult for network administrators to timely and accurately grasp the security situation of the network [7]; so they can not immediately make operations to reduce the losses caused by threats [8]. Network attackers use the complex dependency between vulnerabilities in the target network to implement multi-step network combination attacks against them; gradually destroy the security strategy of the target network; and constantly generate new network attack capabilities [9-10].

The network security analysis model describes the related attributes of network security; and establishes the attack rules that cause the transition of network security state. It is the basis of attack graph generation; and the attack graph generated by real target network is generally large and complex. In particular; the possible loop paths in the attack graph will make it very complex to calculate the probability of successful arrival of nodes by attackers. The data mining system under the big data platform collects the latest research results of big data and data mining.

2. Big Data and Network Security

2.1. Big Data
According to the different storage media of massive raw data or intermediate data in the process of big data processing; it can be divided into memory-level big data calculation and disk-level big data calculation. Therefore; from the perspective of the industry; the construction of data warehouse is a project and a process. In terms of data volume; big data includes all data sets of the enterprise. However; in terms of its scale and complexity; the existing data acquisition; data management; and data calculation technologies cannot be processed and analyzed well; and it needs to be processed and analyzed accordingly. Big data storage computing technology and solutions. In the era of big data; traditional display technology has been difficult to meet the visualization needs of large-scale; multi-dimensional and unstructured mass data. Visualization technology and information display face many challenges in the way of information presentation. For two elements \( x \) and \( y \); if the sample values are set to \( x_i \) and \( y_i \) (\( i = 1, 2, \ldots, n \)); the correlation coefficient calculation formula between them is as follows:

\[
R = \frac{\text{cov}(x, y)}{\sqrt{D(x)} \sqrt{D(y)}}
\]

(1)

In the formula; \( R \) is the correlation coefficient and \( \text{cov}(x; y) \) is the covariance of \( x \) and \( y \).

Let \( f(x; p) \) and \( f(x; q) \) be the equivalent probability distribution on \( x \); then the estimate of \( e \) is:

\[
\hat{e} = \frac{1}{N} \sum_{i=1}^{N} I_{[x(x_i)\in \chi]} f(x, p) \frac{f(x, p)}{f(x, q)}
\]

(2)

Using the Hausdorff distance formula; the optimal parameter solution can be derived as:

\[
\nu^* = \arg \max_{\nu} E_{\nu} I_{[x(X; \nu)]} \ln f(X; \nu)
\]

(3)

2.2. Cybersecurity
Many network businesses need the support of a network security platform in the process of handling. In view of the hidden security risks on these networks; we must first analyze the hidden dangers; then implement targeted network security technology defenses; and finally use anti-virus and defense technologies to detect and kill to ensure the safe transmission and operation of company and individual information. As time changes; the matter-element sequence can be obtained from the dynamic matter-element; and the change of certain characteristics can be observed through the matter-element sequence. For example; the abnormal transfer of user permissions can be found through the user permission sequence.
3. Network Security Simulation Experiment

3.1. Experimental Environment
In order to enable these data to intuitively reflect the network security situation; this paper uses the method of assigning weights to normalize these five sets of data; and assign a larger weight to the security issues with high network security threats. Low security issues are assigned a smaller weight. In this paper; while keeping the proportion of types in the data set unchanged; 3603 samples were randomly selected as evaluation data according to the proportion. The results of data division are shown in Table 1.

| Table 1. Data division results |
|-------------------------------|
| Normal Dos U2R R2L Probe |
| Training samples 498 1563 46 89 327 |
| Test sample 213 669 19 38 141 |
| Total sample set 711 2232 65 127 468 |

3.2. Establishment of Network Security Assessment Model
After setting the parameters; you can start training and testing. First click "Open" to browse and import the training file storing the training data to form the training.txt file; and the model file model.txt that needs to store the training results. Therefore; corresponding control measures are taken according to the risk level to control the system risk at an acceptable level.

4. Discussion

4.1. Impact of Big Data on Network Security
The maintenance of network security should be completed by professional management personnel. The maintenance means should be simple and practical; and the measures that are not required to be taken are too complicated. If the measures taken are too complicated; the operation speed of the system itself will be reduced. It will also cause the occurrence of hidden danger of the system itself. In the early network; the general line network is the main channel; which is expropriated; so it is easy to produce signal conflict; which leads to the decline of network performance. Because the transmission mechanism of Ethernet is CSMA / CD; the conflict is inevitable. The conflict domain refers to the occupation of a port; and the conflict will directly lead to network disconnection. The results of network security risk assessment are shown in Table 2. The comprehensive risk impact of the campus network is not big; the system risk degree is acceptable; so it is not necessary to prepare for the next assessment. Among them; equipment and facilities assets are relatively serious; so we should strengthen security defense to ensure the safe operation of the network. Therefore; through the above analysis; the method proposed in this paper can effectively evaluate the security risk level of network system. It can be concluded that if both the attacker and the defender attach importance to the long-term interests and both sides adopt the optimal strategy at each stage; the total expected return of the attacker is 72.2 and that of the defense system is -72.2 when the probability of the game continues is 0.8. Here; if the defense system does not adopt the optimal strategy in the repeated game between the two sides; it will leave a security risk to the protected host system; and the total loss of the host system may increase significantly. At the same time; the attacker will frequently attack the system; resulting in more serious loss of the host system.

| Table 2. Cybersecurity risk assessment results |
|---------------------------------------------|
| Evaluation vector | Risk level | Risk assessment value |
| (0.200;0.200;0.200;0.300;0.100) | Lower | 3.720 |
| (0.000;0.400;0.200;0.300;0.100) | Medium | 4.060 |
| (0.400;0.400;0.100;0.100;0.000) | Higher | 7.380 |
4.2. Algorithm Test Results
In the process of testing; three placement algorithms are used to place virtual machines. The number of virtual machines in 5; 10; 15; 20 and 30 cases are tested respectively. The specifications and types of each test virtual machine are randomly generated. The virtual machine placement test data is shown in Figure 1. From the experimental data; we can see that the imbalance degree of physical machine resource utilization is kept in a small range; and the degree of uneven resource utilization is smaller than that of using simple and chance algorithms. The virtual machine placement strategy studied in this paper is more balanced than simple and chance algorithms in resource utilization; which can better support the network security event scenario construction and deduction method studied in this paper. At the 5th; 12th; 20th; 28th; 36th; 44th and 52nd minutes; a certain number of attacks are carried out on the protected network. The scale of each attack is twice that of the last attack. Therefore; it can be expected that the network security situation value should be increased in the form of throwing objects. Therefore; the method proposed in this chapter can reflect the change of attack more accurately than the contrast experiment.

| (0.100;0.200;0.400;0.200;0.100) | Lower | 3.000 |
| (0.100;0.300;0.300;0.200;0.100) | Medium | 4.960 |
| (0.400;0.400;0.100;0.100;0.000) | Medium | 4.920 |

![Figure 1. Virtual machine placement test data](image-url)
accuracy; the longer the time consuming; the smaller the error; and the more accurate the simulation of network security defense system model; but the cost is also greater; so it is necessary to weigh the simulation times and errors in practical application.

5. Conclusions
In this paper; the network security technology is studied systematically; and the firewall; intrusion detection and anti-virus are elaborated; so that it can be used in the management process of network security events; which plays a good role in the actual work.

This paper uses the self-learning and nonlinear function fitting ability of artificial neural network to predict the network security situation. It can be seen that the prediction model proposed in this paper has better learning ability; the prediction accuracy is greatly improved compared with the other three models.

In the era of big data; with the rapid development of Internet technology; network security management has become an important guarantee work. There are no inherent rules for the operation of entities; so it is impossible to implement them completely according to the fixed rules.

References
[1] Buchler N ; Rajivan P ; Marusich L R ; et al. Sociometrics and observational assessment of teaming and leadership in a cyber security defense competition[J]. Computers & Security; 2018; 73(3):114-136.
[2] Qabajeh I ; Thabtah F ; Chiclana F . A recent review of conventional vs. automated cybersecurity anti-phishing techniques[J]. Computer ence Review; 2018; 29(8):44-55.
[3] Lakhno V ; Kasatkin D ; Misiura M ; et al. DESIGN OF KNOWLEDGE BASE FOR CYBER SECURITY SYSTEMS ON THE BASIS OF SUBJECT IDENTIFICATION METHOD[J]. Cybersecurity Education ence Technique; 2020; 4(8):135-148.
[4] Cruickshank I J ; Carley K M . SPECIAL SECTION ON EMERGING APPROACHES TO CYBER SECURITY: Analysis of Malware Communities Using Multi-Modal Features[J]. IEEE Access; 2020; 8(1):77435-77448.
[5] Oluwafunmilayo G . An Assessment of Cybersecurity Technologies in the Selected Universities in Southwestern Nigeria[J]. International Journal of Computer Applications; 2019; 178(50):11-18.
[6] Jaiyen S ; Sornsuwit P . A New Incremental Decision Tree Learning for Cyber Security based on ILDA and Mahalanobis Distance[J]. Engineering Journal; 2019; 23(5):71-88.
[7] Xiao-Li M . Statistical paradoxes and paradoxes in big data (I): Law of large populations; big data paradox; and the 2016 US presidential election[J]. The Annals of Applied Statistics; 2018; 12(2):685-726.
[8] Andre J C; Antoniu G; Asch M; et al. Big Data and Extreme-Scale Computing: Pathways to Convergence - Toward a Shaping Strategy for a Future Software and Data Ecosystem for Scientific Inquiry[J]. International Journal of High Performance Computing Applications; 2018; 32(4):435-479.

[9] Wang Y; Kung L A; Byrd T A. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations[J]. Technological Forecasting & Social Change; 2018; 126(1):3-13.

[10] Wang X; Zhang Y; Leung V C M; et al. D2D Big Data: Content Deliveries over Wireless Device-to-Device Sharing in Large Scale Mobile Networks[J]. IEEE Wireless Communications; 2018; 25(1):32-38.