Computer Network Security Analysis Based on Security Fault Tree Method

Meilian Li

School of Electronic and Electrical Engineering, Anhui Sanlian University, Hefei, Anhui, 230601, China

lianslim@163.com

Keywords: Security fault tree, Computer, Network security

Abstract: This paper is a network security situation assessment method based on fault tree analysis. First, the largest multi-step attack sequence excavated is evidence, and then multiple pieces of evidence are fused through a synthetic formula to accurately assess the host's situation risk value. Then, the threat value and importance of each host in the network are used to calculate the risk value of the entire network situation. Finally, experiments prove that this method can accurately evaluate the network situation. Combined with computer visualization technology, a visual analysis system for the conveyor belt run-out fault tree is established quickly and accurately, and corresponding qualitative and quantitative analysis is performed. To strengthen the mastery of the network security situation, and constantly adjust the plan, make the network management more targeted, and can improve the complete diagnosis efficiency of the network.

1. Introduction

The emergence and popularization of computer networks have indeed brought convenience and a profound impact on people's lives, but at the same time, hidden security risks have also been exposed. There are inherent security vulnerabilities in the design of the computer network itself. At the same time, with the popularization of network technology and the expansion of the scale of the network, many criminals use the vulnerabilities in the network for virus propagation, network intrusion, and extortion for self-interest. These hazards have not only affected the security of people's use of the network, but also caused important information leakage of government agencies, military aircraft agencies, schools, and financial institutions, and even caused harm to national security. With the expansion of the scale of the Internet and the increasingly complex network environment, various types of network security incidents are emerging.

Frizzo-Barker and others proposed a three-level model of network situational awareness [1]. He summarized situational awareness into three steps of feature extraction, state awareness, and situation prediction. Wang and others proposed a network situation fusion perception and risk assessment model [2]. Grover and others abandoned the single network security system framework and proposed the concept of an integrated network security framework [3]. Yaqoob and others believe that cognitive computing is a new mechanism for solving network security situational awareness [4]. At the same time, they believe that applying an OODA-like design to cognitive computing can improve the cognitive ability of the system [5]. It contains a variety of security strategies, mainly to control network communications, allow normal access, and prevent illegal access from the outside world. From the early packet filtering technology attached to the router to the current adaptive firewall technology, the firewall has developed into one of the indispensable security protection technologies in the network security system [6].

This book presents the necessity of using security failure methods to achieve situational awareness in a large-scale network environment. The feasibility and practical application value of the research method in this paper are determined. The main work and innovations of the paper are described, including the use of feature similarity method to achieve the redundancy elimination of multi-source alarm data, the implementation of the improved Apriori algorithm to mine frequent multi-step attack sequences, and the use of DS evidence theory to achieve Evaluation of the network situation.

DOI: 10.25236/iwmecs.2020.063

2. Safety Fault Tree Method Safety Network Analysis and Design

2.1 Rebuilding the Safety Fault Tree Model

The failure rate of each event in the failure model is generally obtained through the statistics of the failure data, and can also be obtained through the experience of the maintenance management personnel [7]. The failure rate of the component to be analyzed is the product of the failure rate of the next-order event. According to the logical relationship in the fault tree model, the failure rate of the smallest cut set is analyzed in order, and finally, the failure rate of the top event can be obtained.

The probability of the output event of the fault tree and gate structure is shown in formula (1).

$$M(A) = \cap M(B_i) = \prod^{n} M(B_i) \tag{1}$$

The probability of occurrence of the upper event of the fault tree or gate structure is shown in formula (2).

$$N(A) = \cap H(B_i) = \coprod_{i-1} n \prod^{n} M(B_i) \tag{2}$$

When performing fault tree analysis, different components, different workflows, and even different people play different roles in different environments. Therefore, the impact degree of each event in the fault tree model will also change. The concept of “importance” was introduced. Importance refers to the impact of the occurrence of each bottom event or cut set on the top event in the fault tree model. The greater the value of importance, the greater the impact of the event or cut set on the entire system, and it is also the object that needs to be focused on when designing and maintaining the system.

$$Q_i(\partial) = \frac{1}{2}\left(\sum \phi(x_1, x_2, ..., x_n)\right) \tag{3}$$

The degree of change in the probability of the occurrence of the basic event caused by the change in the probability of the basic event is called the probability importance, which is expressed by X. Since the Q function of the top event occurrence probability is a multi-linear function, as long as the partial derivative of the independent variable K_q is obtained, the probability importance coefficient of the basic event can be obtained:

$$K_q(x) = \frac{Q_i(\partial)}{\partial q_i} \tag{4}$$

During the establishment of the fault tree model, the top event of the fault tree is generally the main cause of system failure, then the second-level event that causes the top event, and then the third-level event that causes the second event. Finally, logical events are used to connect each level of events with logical symbols to form a tree structure until all the correct logical relationships are established with the top events.

When performing system fault tree analysis, data recording is also very important. Through daily statistics on the number of occurrences of each failure factor, the probability of occurrence of each failure is obtained, and the probability of occurrence of system failure (ie, top event) is obtained through logical calculation. At the same time, according to the probability of occurrence of each fault, the degree of its influence on the occurrence of the system fault is obtained, and the importance degree of each basic event relative to the top event is obtained using the calculation formula of the importance degree, and then arranged.

2.2 Network Security Analysis and Design

As shown in Figure 1, the fault tree analysis and visualization system mainly include menus for event definition, logical relationship selection, event information statistical recording, graphical interface, data calculation, information output, and interface settings. There are also some common icons below the menu bar for users to choose from. The middle part of the system operation interface is the editing area of the system, and a common fault tree model can be established through the common icons in the common menu bar. On the left side of the interface are the operation key settings of the system, including model creation key, system information storage key, output key, qualitative, quantitative analysis key, setting key, etc., which is convenient for quickly establishing and calculating fault tree model.

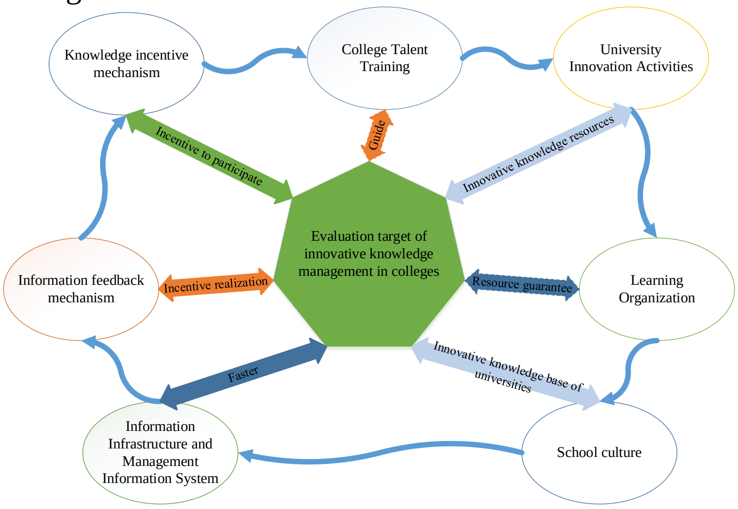
Knowledge incentive mechanism
College Talent Training
University Innovation Activities
Incentive to participate
Guide
Innovative knowledge resources
Evaluation target of innovative knowledge management in colleges
Information feedback mechanism
Incentive realization
Resource guarantee
Learning Organization
Faster
Innovative knowledge base of universities
Information Infrastructure and Management Information System
School culture

Fig.1 Model of Equivalent Electron Ls Coupled Atomic State

Through the calculation of the minimum cut set, find out all the reasons that may lead to the deviation of the conveyor belt deviation, so that equipment users and managers can fully understand the dangers of the system, and strive to minimize the minimum cut set of the system at the design stage, and the probability of occurrence of the smallest cut set is the lowest so that the system can ensure the lowest risk in subsequent operations.

Quantitative analysis is the core function of the fault tree visualization system design. Through the definition of events and the entry of relevant information in the drawing function module, including the statistics of the occurrence probability of each basic event, the visualization system can automatically solve the top according to the logical relationship between events. After entering all the belt conveyor belt deviation fault events and fault probability into the system, the probability of occurrence of the top event is calculated to be 15.02%, which is about 0.12% of the calculation error from the previous calculation of the probability of 13.97%. Because the calculation accuracy of the visualization software is much higher than the formula calculation, it can be seen the probability of deviation of the belt conveyor belt is about 15.25%.

After normalization, the alarm consists of an eleven-tuple (Id, Analyzer_Id, Src_ip, Dst_ip, Src_port, Dst_port, Create_Time, Classification, Type, Threat_level, Data). The meaning of each field is listed in Table 1.

Table 1 Conversion Rules between Score Levels and Scores

name	Type	meaning
Id	Int	Alarm number
Analyzer_Id	Int	Identified ID number
Src_ip	Int	Attack source address
Dst_ip	Int	Attack destination address
Src_port	Float	Attack source port
Dst_port	Float	Attack destination port
Create_time	Float	When the alarm occurred
Score grade	Corresponding score	Corresponding score range

After pre-processing the data in the data source, the data in the source database need to be loaded into the data warehouse. Because the fact table references the primary key in the dimension table, the historical dimension table data is loaded before the historical data fact table data.

3. Results Analysis

3.1 Network Security Alarm Characteristics of Similarity Calculation

Calculate the similarity of the extracted attributes and fuse alarms with similarity greater than the threshold. Different threshold values will get different alarm fusion effects. In this paper, we set different thresholds to perform fusion experiments on the generated Snort alarms. The results are shown in Figure 2.

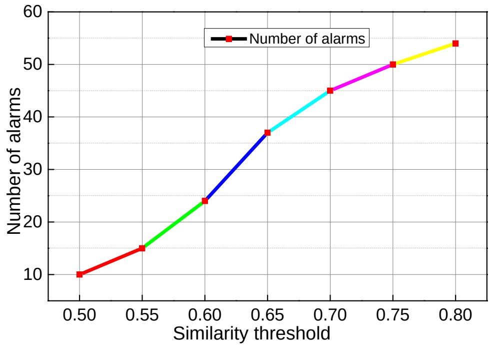

Fig.2 The Effect of Alarm Fusion under Different Thresholds

It can be seen that the larger the similarity threshold is set, the more base alarm types will be generated after the alarm fusion, which may result in insufficient fusion, and the alarms belonging to the same attack behaviour will be divided into different attack events. On the contrary, if the similarity threshold is set too small, it will not work. If the threshold is set too small, there will be fewer base alarm categories. Each type of base alarm contains too many original alarm information, which may merge original alarms that do not belong to the same attack behaviour. According to the experience of experts, it is found that the fusion effect is the best when the similarity threshold is set between 0.651 and 0.724. Within this range, various detection tools generate alarms for the same data source. The base alarm category generated by the fusion is very good. Similarly, this paper sets the similarity threshold to 0.747 and the time threshold to 25 min.

Whether a multi-step attack can occur successfully depends on the success rate of each single-step attack it contains. Only when all single-step attacks have successfully occurred can the multi-step attack successfully progress to the final attack stage. We already know that the multi-step attack process defined in this article is mainly divided into five steps: information detection, vulnerability scanning, vulnerability exploitation, privilege escalation, and launching attacks. The complexity of these five attacks is increasing step by step. The higher the complexity of the attack, the greater the difficulty of the attack.

3.2 Performance Analysis

In the case of using the same data set, the minimum support is set to 0.04, and the time required to mine frequent attack sequences under different transaction numbers between the Apriori algorithm and the N-Apriori algorithm is compared. The results are shown in Figure 3.

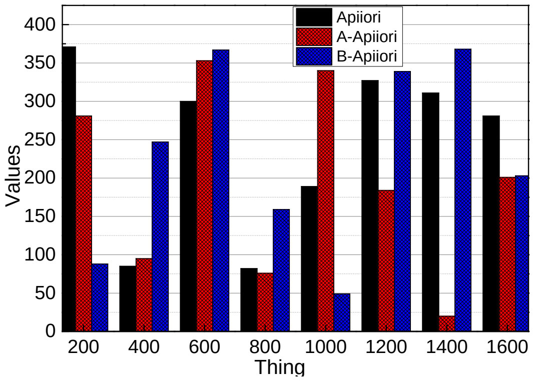

Fig.3 Comparison of the Running Time of the Two Methods under the Same Number of Transactions

From the results in Figure 3, it can be shown that the improved algorithm running time is less than the original Apriori algorithm running time when the database contains a different number of transactions. This is because the N-Apriori algorithm only needs to traverse the candidate set once during the pruning process to cut off the non-frequent item set in the candidate set, and once the item set is found to be infrequent item set during the scanning immediately stop scanning the item set, reducing the number of database scans, so the improved algorithm is more efficient than the original Apriori algorithm in mining frequent multi-step attack sequences. Given the different minimum support levels, the running time of the improved algorithm is lower than the running time of the original Apriori algorithm. First, because the N-Apriori algorithm only needs to scan the frequent Lk-1 item set during the pruning process. You can complete the deletion of infrequent itemsets. Also, in the process of determining the frequent itemset Lk, once it is found that the non-frequent item set count of a certain item set reaches the infrequent item set threshold, it immediately stops scanning the database, reducing the running time of the algorithm. Therefore, the improved algorithm is more efficient than the original Apriori algorithm in mining frequent multi-step attack sequences.

This paper studies the association analysis of alarm data and proposes a mining method for frequent multi-step attack sequences based on the improved Apriori algorithm. By mining the connection between alarm data, the multi-step attack pattern embedded in it is obtained, and a higher level is obtained. The attack information provided a basis for the subsequent situation assessment. This chapter first briefly introduces the definition and steps of association analysis, then describes the background and specific process of the classic frequent item mining algorithm Apriori and analyzes the shortcomings of the algorithm, and proposes improvements to the Apriori algorithm.

4. Conclusion

This paper proposes a multi-step attack sequence network security analysis and mining method based on the improved fault tree method. The multi-step attack pattern existing in the network is mainly mined through the frequent item mining algorithm in data mining, and the causal relationship between the individual attack steps is found to obtain a higher level of attack information. This article first discusses the fault tree algorithm and its shortcomings and then puts forward its ideas for improvement. Finally, the improved fault tree algorithm is used to mine the multi-step attack sequence. This paper proves through experiments that the improved algorithm improves the mining efficiency of frequent multi-step attack sequences.

Acknowledgement

This research was financially supported by Department of Education Anhui Province Natural Science Foundation of China:” Fault Diagnosis Method and Application for Complicated Dynamic Network Systems with Variable Topologies on Micro-grid” (KJ2018A0601), and Quality engineering project of Anhui Sanlian University: “teaching team of signal processing and control series courses” (No.: 18zlgc007)

References

[1] Frizzo-Barker, Julie, et al. “An empirical study of the rise of big data in business scholarship.” International Journal of Information Management, vol.36, no.3, pp. 403-413, 2016

[2] Wang, Yichuan, et al. “An integrated big data analytics-enabled transformation model: Application to health care.” Information & Management, vol.55, no.1, pp. 64-79, 2018

[3] Grover, Purva, and Arpan Kumar Kar. “Big data analytics: A review on theoretical contributions and tools used in literature.” Global Journal of Flexible Systems Management, vol.18, no.3, pp. 203-229, 2017

[4] Yaqoob, Ibrar, et al. “Big data: From beginning to future.” International Journal of Information Management, vol.36, no.6, pp. 1231-1247, 2016

[5] Alaei, Ali Reza, Susanne Becken, and Bela Stantic. “Sentiment analysis in tourism: capitalizing on big data.” Journal of Travel Research, vol.58, no.2, pp. 175-191, 2019

[6] Saggi, Mandeep Kaur, and Sushma Jain. “A survey towards an integration of big data analytics to big insights for value-creation.” Information Processing & Management, vol.54, no.5, pp. 758-790, 2018

[7] Muhammad, Syed Sardar, Bidit Lal Dey, and Vishanth Weerakkody. “Analysis of factors that influence customers’ willingness to leave big data digital footprints on social media: A systematic review of literature.” Information Systems Frontiers, vol.20, no.3, pp. 559-576, 2018