Research Article

Design of a Network Security Audit System Based on Log Data Mining

Yanbo Xing

International College of Zhengzhou University, Zhengzhou, 450000 Henan, China

Correspondence should be addressed to Yanbo Xing; xyb18738607008@stu.zzu.edu.cn

Received 13 December 2021; Revised 8 February 2022; Accepted 12 February 2022; Published 25 February 2022

Academic Editor: Zhiguo Qu

Copyright © 2022 Yanbo Xing. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The purpose of network security auditing is to safeguard network and information security through the assessment of network security vulnerabilities. Data mining is mainly used to mine potential information from massive amounts of log data, which can provide both accurate and valuable auxiliary data for network security auditing and relevant information for monitoring and management of terminals in different network environments. Firstly, the theory of data mining technology is explained, secondly, some data mining algorithms applied to network security audit are discussed, and finally, the design difficulties and functions of the network security audit system based on data mining are studied, aiming to provide reference for securing network and identifying hidden risks.

1. Introduction

As the basis for achieving network security objectives, network security auditing not only enables timely detection of the harmful behavior in the system but also secures the network by testing, evaluating, and analyzing the vulnerability of specific network security and enabling the identification of attacks and the tracking of violations of security laws. Its three main components include network behavior auditing, database network access auditing, and application behavior [1–3]. However, with the dramatic increase in network data traffic, the important information hidden behind it is also increasing, and there is an urgent need to accurately analyze and extract the hidden predictive information; so, the choice of which technology to analyze becomes the focus of research [4].

As an advanced intelligent data analysis technology, data mining is the process of extracting potentially useful information from a large amount of fuzzy, random data as a basis for decision-making [5]. The core of the application of this technology in network security auditing is to select the appropriate data analysis algorithm by specifying the task and purpose of mining and then compare and judge the nature of current network behavior based on the information implied by the logged behavior data [6].

Data mining technology provides data support for network security protection by revealing the inherent connections and characteristics of log data [7]. However, it is important to note that data mining algorithms are not applicable to all network log data and must be combined with the purpose and task of network security auditing to select the appropriate algorithm.

Classification and prediction algorithms are the most fully developed area of data mining. In summary, data classification can be expressed in the form of rule sets or decision trees, and in the case of security audit applications, the behavior of new audit data is predicted by classifying user-related data [8]. However, it is important to note that the basis for classification needs to be based on experience and experimental results, and that appropriate thresholds need to be selected as a means of screening system characteristics. In general, if there is some association between two or more variables in a database, then we can assume that they are correlated. It can be either a causal or a simple correlation; so, it can be said that data correlation is a fundamental and important feature of databases. The application of this technology in network security auditing is to analyze the correlation between security events by discovering the correlation of data sets, thus providing a basis for improving the

FIGURE 1: Network security audit system framework.

accuracy of auditing risk alert events. For example, network log data is an external manifestation of user behavior, and by studying the correlation characteristics of data over a period of time and comparing and analyzing past historical normal behavior patterns, potential abnormal behavior can be identified, and security risk alerts can be made to safeguard users' network security [9–11].

Sequence analysis, when applied in network security auditing, is primarily designed to uncover connections between data, i.e., to uncover the causal links between the data being audited. In contrast to association rules, sequence analysis incorporates the concept of time and aims to capture the relationships between database records in a window of time. By analyzing the regularity of event sequence patterns, it can help in the selection of effective statistical features when constructing intrusion detection models [12]. Clustering is a fundamental technique in data mining and is mostly used to analyze user characteristics. And clustering analysis finds groups of objects that are strongly correlated, while anomaly detection finds objects that are not strongly correlated with other objects; so, clustering has a clear advantage in anomaly detection. In network security auditing, network data are first categorized according to its characteristics, with high similarity in standard characteristics of similar data and low similarity between different class groups. Clustering enables effective analysis of access to network resources, the order of access, and secure identification of similar browsing patterns between users.

The contributions of this paper are as follows:

We expound the relevant theories of data mining technology, discuss the data mining algorithm applied to network security audit, and study the design difficulties and functions of network security audit system based on data mining.

We also design storage threads for system logs, network data, and firewall logs and extract their own logs to the buffer through each thread function. The log reading mode

FIGURE 2: The full process of the log information collection module.

of the provider is different from the log reading mode of the provider.

Through the experimental analysis of data mining technology, we can know that the application of data mining in network security audit is mainly to mine useful data for security audit from massive data.

2. Current Developments in Audit Technology

With the rapid development of the network, the issue of network security has become particularly important. Firewalls and IDS are effective in preventing unauthorized access to systems, but they only serve a defensive function [13]. It is not enough to have these tools, but we need to be proactive in our efforts to combat crime. At the same time, intrusions and sabotage from within the network are becoming more and more serious, and companies and departments urgently need appropriate monitoring systems to supervise unscrupulous employees and ensure corporate security.

The idea of using system log information for security auditing was first introduced by James P. Anderson in 1980 [14], and it was not until 1995 that the first practical network security vulnerability auditing software, SATAN, was released, but SATAN is technically demanding and very inconvenient to use [15]. In recent years, many tools have emerged in this area, but most are system user-based audit tools, such as SCE (Security Configuration Editor) in Windows NT 5.0 or the audit tools that come with Unix, which has limited ability to audit security events in the entire network. A number of foreign companies and

FIGURE 3: Audit subsystem flow chart.

research institutes have already launched many corresponding tools, for example, the inspect distributed network security audit system in Germany and the HASHIS system at Purdue University are experimental systems for auditing [16].

Currently, there is active international research in this direction, most of which is supported by the US Department of Defense Advanced Programs and the National Science Foundation. At the initial stage, intensive research in data mining-based intrusion detection has been conducted by Wenke Lee's group at Columbia University [17, 18] and Stephanie Forrest's group at University of New Mexico (UNM). The data mining-based audit analysis system implemented by Wenke Lee's group is significantly better than other systems in detecting denial-of-service attacks and scanning, but not enough research has been done to detect anomalous internal user actions [19]. However, Professor Wenke Lee pointed out that data mining techniques can be used to extract feature patterns from a large number of audit logs of hosts and then use these rules to build an analyzer that can analyze abnormal user activities [20].

Stephanie Forrest's research group used short sequence matching algorithms [21] to perform detailed analysis of host security events and sequences of system calls generated by user behavior, but there has been insufficient research in detecting abnormal operations by internal users, and audit log analysis of user behavior such as file access behavior has not been addressed [22]. [23] used association rule mining algorithms typical of data mining to establish user behavior patterns and analyze user command sequences through association rules, but the paper pointed out that

FIGURE 4: Flow chart of linear matching of rules.

the mining algorithm requires a large amount of memory space and reducing the space overhead is an issue for further research. Applying data mining techniques to log analysis of user behavior is theoretically feasible and technically possible [24]. The difficulty lies in finding an efficient data mining algorithm that takes up less memory and extracts frequent behavior patterns that reflect the characteristics of user behavior in a large volume environment based on the characteristics of user log behavior records [25].

3. Implementation of Network Security Audit Based on Data Mining Technology

3.1. Overall Design. Based on the relevant key technology research, we have designed the following model for the network security audit system specifically as shown in Figure 1.

The main functions include the following:

- (1) Collecting system logs, security logs, application logs, firewall logs, and network data from the operating system and processing the collected data in a uniform format [26]
- (2) Real-time processing and analysis of the logs collected into the audit system, followed by the generation of corresponding alarms according to the rule base and the possibility of alerting the administrator via email or pop-up dialogs
- (3) Postprocessing analysis of logs collected into the audit system and generation of reports
- (4) To store the logs collected in the audit system as regular or quantitative backups and to support the import of various logs from outside
- (5) Provides a management interface for users and supports efficient combination of conditions for querying inventory logs
- (6) Audit results can be recorded for intrusions and violations and can be reproduced at any time for

This function is necessary for accountability and data recovery; finally, the system can be used to extract unknown or undetected patterns of intrusive behavior

- (7) Log tracking of audit findings

FIGURE 5: Flowchart for matching time reasonableness rules.

FIGURE 6: Flowchart for adding rules.

3.2. Logging Module Design. Create a separate thread for the extraction of system logs, network data, and firewall logs and implement the extraction of the respective logs via each of the each thread function performs the function of extracting the respective logs into a buffer. We use the provider pattern: the main functions to be implemented are the insert() and get() methods for reading and inserting logs for the different log collection objects. An abstract class log provider is used to declare the two methods insert() and get() that different providers must implement, and multiple provider classes are written to implement the two methods of the abstract class. Multiple providers are placed in a collection of log provider collection, which is used to handle the multiple providers we need. We use a configuration class log provider configuration section to declare which providers exist and which providers are available by default and then encapsulate these providers so that we can call the insert() and

FIGURE 7: Effect of different data mining.

get() methods [27]. Figure 2 shows the flow of the log collection implementation.

3.3. Audit Module Functional Design. The audit subsystem process is shown in Figure 3.

From the framework structure diagram designed above, it is clear that the audit function mainly consists of the following: completion of collected data, real-time response, update rules, and audit report. The auditing of data in the system is divided into historical log data and real-time log data analysis by matching the existing.

The linear matching process of rules is as follows: the system security log data generated by the relevant event actions are analyzed and summarized to form a linear sequence based on rules. Real-time data analysis is when the system obtains the event logs generated by the operation behavior and extracts the event ID (InstanceId) and audit type [28]. Information characteristics of the event description are used as priority parameters in the rules, i.e., the more specific the event characteristics of the corresponding database in the same order of event patterns, the more priority the event behavior pattern has the matching.

Real-time data auditing is the analysis and matching of data after acquisition, and then the corresponding response is made according to the result of the matching (the system is implemented as an exact match) and the danger level of the event in the rules, as shown in Figure 4.

Time reasonableness rule matching process is as follows: since the chance of leaking information by internal employees in the network is much greater than that of external personnel, limiting the normal use time of internal employees from violating the operation, if an employee uses

FIGURE 8: Effectiveness of different network security audits.

FIGURE 9: Different audit effects.

the machine without permission outside of the time limit, it will be considered a violation, and an alarm will be generated and sent to the designated. The time reasonableness rule matching process is shown in Figure 5.

The alarm response at this point consists of two main components: the legality of the user in question and the reasonableness of the use of the system by the legitimate user for a limited period of time. If one of the two does not meet the requirements, it is considered to be an illegal use or a legal user who has violated the reasonableness of the time and is given an immediate alarm for this event behavior pattern.

3.4. Audit Rule Development, Automatic Generation, and Maintenance of Updates. The audit function is based on audit rules and audit ideas, of which audit rules are the basis and premise of the whole audit system. Audit rules and data acquisition

are equivalent to the hardware and software parts of a computer system, which are indispensable and at the same time mutually fundamental. In the system, data acquisition provides a place for the implementation of audit rules, while audit rules provide direction and support for data acquisition [12].

The rules' update and maintenance in the system mainly include adding, deleting, and browsing of rules. The design and implementation of adding rules are as follows: when the administrator or user finds a new problem pattern, the relevant characteristics of this behavior pattern can be collated and extracted at any time to form a collective of attributes required in the rule base, and then the pattern can be saved in the database and updated for application. This makes it possible to use the new rules immediately after they have been added to the system, making up for the fact that the original design idea had to be reenacted before it could be applied, and also reducing the possibility of data omissions and misinterpretations. This is shown in Figure 6.

The design and implementation of the added rules are as follows: in the process of use, some of the rules lose the need to exist as time passes, and the administrator improves the security maintenance of the system. Rule data stored in the database not only occupies storage space but also occupies system memory when the system is running, which increases the overhead of the system when running and affects the overall performance of the system, which is not conducive to the performance of the system; so, the system provides the integration of this part of the rules for deletion [29, 30].

4. Experimental Analysis

Through the above analysis of data mining technology, it can be seen that the application of data mining in network security auditing is mainly to dig out the data useful for security auditing

from the mass data. The implementation process of data mining technology in network security audit application is shown in Figure 7. Log format compatibility is as follows: log format compatibility is the biggest problem encountered in the process of centralized analysis of network security events. Generally speaking, different hardware devices are not compatible with each other; so, the conversion of log formats needs to be considered in the system design to unify the format of log data.

Management of log data is as follows: network log data has been showing a rapid growth; so, a perfect backup, recovery, and processing mechanism should be established in the process of data management. Centralized analysis of log data is as follows: an attacker may sometimes attack multiple network servers at the same time; so, how to carry out correlation analysis of logs from multiple servers is a topic that needs further research in the network security audit system. Automatic generation of analytical reports and statistical reports is showed in Figure 8: the large amount of log information makes it difficult for administrators to view and analyze; so, the system is required to provide an intuitive analysis report or statistical report. The administrator can detect abnormal network conditions in a timely and effective manner through the automatically generated reports.

For the event audit effect to be achieved by the system, it is important to do both preprocessing of data and query and management of data logs when designing data mining-based network security audits. Only when both aspects are carried out simultaneously can be real-time and historical auditing of events be achieved, ensuring the security of account remote login and predicting the inherent information implied under the comprehensive audit data. The system needs to be able to manage and maintain rules, which require the audit system to be expandable, and to continuously improve the management of data rules by adding and deleting relevant rules. And to facilitate better management, the system also needs to achieve automatic report generation functions under specific query conditions, as showed in Figure 9 for different audit effects.

The system should realize the query of audit results, and when designing the functions, it should include the classification query of audit results and the saving and printing of query results. These are not only beneficial to the manager's efficiency in managing abnormal data but also enable the handling of security hazards that exist in the network system, thus safeguarding network information security. Commonly used queries are audit time, danger level, audit strategy, etc. The network security audit system should also do a good job of summarizing the audit data and be able to find the loopholes and problems in the network system by comparing the original audit data and the result data and further explore the relevant solution measures.

5. Conclusions

The importance of network security auditing in safeguarding network information security is becoming increasingly prominent, especially with the continuous development of network information technology, and the complexity and diversity of attack methods are also increasingly obvious, thus bringing great challenges to network security. The

introduction and application of data mining technology can not only effectively improve the detection efficiency of security auditing but also effectively improve the speed and attractiveness of network security auditing, which greatly improves network security. However, it should be noted that the various algorithms of data mining all have their unique advantages; so, the complementary strengths and effective combination between the methods should be a subject of in-depth research, and more comprehensive research work is yet to be carried out.

Data Availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

References

- [1] R. G. Rajan and L. Zingales, "The firm as a dedicated hierarchy: a theory of the origins and growth of firms," *Quarterly Journal of Economics*, vol. 116, no. 3, pp. 805–851, 2001.
- [2] M. Bould, "The dreadful credibility of absurd things: a tendency in fantasy theory," *Historical Materialism*, vol. 10, no. 4, pp. 51–88, 2002.
- [3] C. H. Cao, Y. N. Tang, D. Y. Huang, W. M. Gan, and C. Zhang, "IIBE: an improved identity-based encryption algorithm for WSN security," *Security and Communication Networks*, vol. 2021, 8 pages, 2021.
- [4] D. Wu, C. Zhang, L. Ji, R. Ran, H. Wu, and Y. Xu, "Forest fire recognition based on feature extraction from multi-view Images," *Traitement du Signal*, vol. 38, no. 3, pp. 775–783, 2021.
- [5] D. Zhao, Q. Xu, and Z. Feng, "Analysis and design for intrusion detection system based on data mining," in *In 2010 second international workshop on education technology and computer science*, vol. 2, pp. 339–342, Wuhan, China, March 2010.
- [6] S. Yi and F. Deng, "Research of network intrusion-detection system based on data mining," in *Recent Progress in Data Engineering and Internet Technology*, Berlin, Heidelberg, 2012.
- [7] L. Wang, C. Zhang, Q. Chen et al., "A communication strategy of proactive nodes based on loop theorem in wireless sensor networks," in *2018 Ninth International Conference on Intelligent Control and Information Processing (ICICIP)*, pp. 160–167, Wanzhou, China, Nov 2018.
- [8] X. Qing, "Log-based network security audit system research and design," in *Advanced materials research*, pp. 1426–1431, Trans tech publications Ltd, 2010.
- [9] Z. M. Wang, M. X. Liu, H. Hu, and W. X. Jiang, "Study on network security audit system based on agent technology and log mining," in *Advanced materials research*, pp. 534–537, Trans tech publications Ltd, 2013.
- [10] M. A. Aydın, A. H. Zaim, and K. G. Ceylan, "A hybrid intrusion detection system design for computer network security," *Computers & Electrical Engineering*, vol. 35, no. 3, pp. 517–526, 2009.

- [11] H. Li, D. Zeng, L. Chen, Q. Chen, M. Wang, and C. Zhang, "Immune Multipath Reliable Transmission with Fault Tolerance in Wireless Sensor Networks," in *International Conference on Bio-Inspired Computing: Theories and Applications*, pp. 513–517, Springer, Singapore, 2016.
- [12] L. Zhuo and Z. Wang, "Research and implementation of log-based network security audit system," *Energy Procedia*, vol. 11, pp. 2021–2026, 2011.
- [13] X. W. Ning and P. Y. Liu, "Log system design in support of linkage analysis of security audit and computer forensics," *Computer Engineering and Design*, vol. 30, no. 24, pp. 5580–5583, 2009.
- [14] Y. Dong, R. Wang, and J. He, "Real-time network intrusion detection system based on deep learning," in *2019 IEEE 10th international conference on software engineering and service science*, pp. 1–4, Beijing, China, October 2019.
- [15] Z. Jia, C. Shen, X. Yi, Y. Chen, T. Yu, and X. Guan, "Big-data analysis of multi-source logs for anomaly detection on network-based system," in *2017 13th IEEE conference on automation science and engineering*, pp. 1136–1141, Xi'an, China, August 2017.
- [16] L. Deng, D. Li, X. Yao, D. Cox, and H. Wang, "RETRACTED ARTICLE: Mobile network intrusion detection for IoT system based on transfer learning algorithm," *Cluster Computing*, vol. 22, no. 4, pp. 9889–9904, 2019.
- [17] M. F. Kabir and S. Hartmann, "Cyber security challenges: an efficient intrusion detection system design," in *2018 international young engineers forum*, pp. 19–24, Costa da Caparica, Portugal, May 2018.
- [18] F. Shu, M. Li, S. Chen, X. Wang, and F. Li, "Research on network security protection system based on dynamic modeling," in *2017 IEEE 2nd information technology, networking, electronic and automation control conference*, pp. 1602–1605, Chengdu, China, December 2017.
- [19] L. Zhang, "Design of a sports culture data fusion system based on a data mining algorithm," *Personal and Ubiquitous Computing*, vol. 24, no. 1, pp. 75–86, 2020.
- [20] S. Agrawal and G. Jain, "A review on intrusion detection system based data mining techniques," *International Research Journal of Engineering and Technology (IRJET)*, vol. 4, no. 9, pp. 402–407, 2017.
- [21] P. Anitha and B. Kaarthick, "Oppositional based Laplacian grey wolf optimization algorithm with SVM for data mining in intrusion detection system," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 3, pp. 3589–3600, 2021.
- [22] P. An, Z. Wang, and C. Zhang, "Ensemble unsupervised auto-encoders and Gaussian mixture model for cyberattack detection," *Information Processing & Management*, vol. 59, no. 2, article 102844, 2022.
- [23] W. Wang and Y. Xu, "Research on awareness method of cloud user abnormal behavior based on log audit," in *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, pp. 1944–1950, Chengdu, China, December 2018.
- [24] T. T. Teoh, Y. Y. Nguwi, Y. Elovici, W. L. Ng, and S. Y. Thiag, "Analyst intuition inspired neural network based cyber security anomaly detection," *International Journal of Innovative Computing, Information and Control*, vol. 14, no. 1, pp. 379–386, 2018.
- [25] K. A. Tarnowska and A. Patel, "Log-based malicious activity detection using machine and deep learning," in *Malware Analysis Using Artificial Intelligence and Deep Learning*, Springer, Cham, 2021.
- [26] M. Amar and B. E. Ouahidi, "Weighted LSTM for intrusion detection and data mining to prevent attacks," *International Journal of Data Mining, Modelling and Management*, vol. 12, no. 3, pp. 308–329, 2020.
- [27] N. Afzaliseresht, Y. Miao, S. Michalska, Q. Liu, and H. Wang, "From logs to stories: human-centred data mining for cyber threat intelligence," *IEEE Access*, vol. 8, pp. 19089–19099, 2020.
- [28] H. Azwar, M. Murtaz, M. Siddique, and S. Rehman, "Intrusion detection in secure network for cybersecurity systems using machine learning and data mining," in *2018 IEEE 5th international conference on engineering technologies and applied sciences (ICETAS)*, pp. 1–9, Bangkok, Thailand, November 2018.
- [29] P. Rutravigneshwaran, "A study of intrusion detection system using efficient data mining techniques," *International Journal of Scientific Research in Network Security and Communication*, vol. 5, no. 6, pp. 5–8, 2017.
- [30] N. I. Ghali, "Feature selection for effective anomaly-based intrusion detection," *International Journal of Computer Science and Network Security*, vol. 9, no. 3, pp. 285–289, 2009.