Research on Web Access Information Safety Mining Method in Distributed Environment

Yafang Li

Suzhou Chien-Shiung Institute of Technology, Taicang, Jiangsu, 215411, China

Keywords: Distributed environment; web access; information safety; mining

Abstract: With the rapid development of network and communication technology, especially the development of computer network technology, the security of web access is becoming more and more important, and the protection of data security and integrity has attracted more and more attention. Information safety is becoming a comprehensive and multidimensional problem. It means to prevent the information property from being intentionally or accidentally disclosed, changed, damaged or the information from being identified and controlled by the illegal system This paper analyzes the user behavior that may lead to information safety problems in people's daily web access behavior, and seeks a safe, civilized access behavior and information safety mining method.

1. Introduction

With the continuous development and progress of Internet technology in China, the network scale of both enterprises and government agencies is constantly expanding, the utilization rate of computers is constantly increasing, and the network equipment is constantly being updated and upgraded. At the same time, the problem of network information safety is constantly increasing, but there are many unsafe factors in network applications, which are mainly manifested in information leakage, information tampering, illegal use of network resources, illegal information infiltration, fake information and so on. The information safety problem brought by web access is particularly prominent[1]. Therefore, based on the development of network environment, this paper studies a new method of network information safety mining-distributed network information safety management and control system, in order to ensure the security of network information and bring a safe and stable network environment to users.

2. The hidden danger of computer network information safety

At present, the factors that affect the construction of network information safety mainly include people, things and things. In general, these factors threaten network security through direct or indirect means, resulting in network data damage or destruction, or even unexpected consequences. Through the investigation of Internet access behavior among people of different ages, and the statistical analysis of the questionnaire, the following analysis results of Internet access behavior are obtained. According to the attention of different groups on computer and network security (see Figure 1), the computer and network security status of the elderly group is worrying. Among these measures, the proportion of relying on third-party security management software for computer and network security management is large [2], which shows that Internet users have weak knowledge of computer and network. What is more worrisome is that Internet users generally do not understand the corresponding security settings of computers and networks, and there is no set computer and network software and hardware environment, which means that the first line of defense of information safety is broken, let alone information safety in web access behavior. Common web vulnerabilities in network information safety include: SQL injection vulnerability, cross-site scripting vulnerability, weak password vulnerability, HTTP header tracking vulnerability, Struts2 remote command execution vulnerability, file upload vulnerability, private IP address disclosure vulnerability, unencrypted login request and sensitive Information disclosure vulnerability.

  DOI: 10.25236/icceme.2022.005

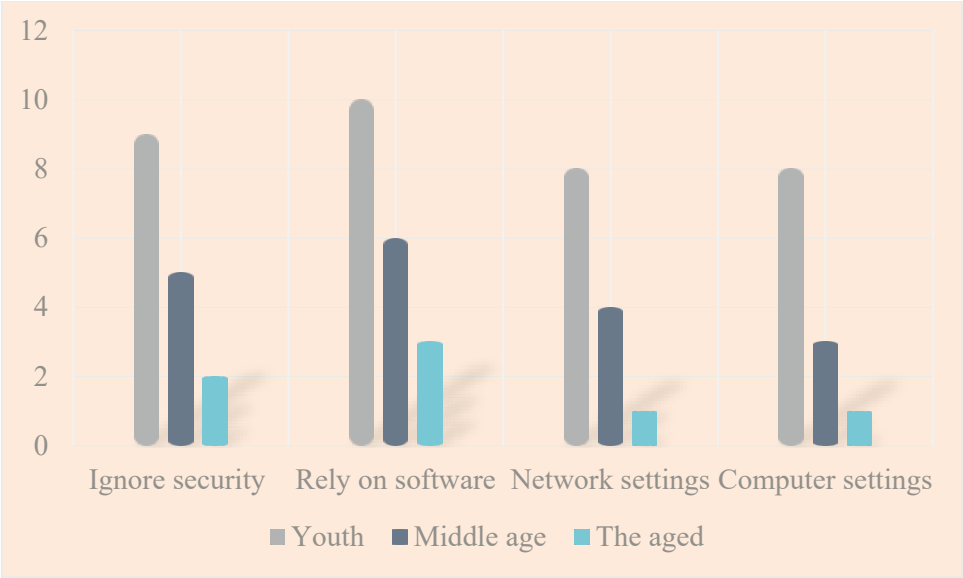

Figure 1 Proportion of different groups' attention to network information safety

There are many ways to prevent information safety from being threatened during network access. First of all, we must have a safe and reliable computer and network environment. In people's daily network access behavior, we should learn to make necessary computer and network-related security settings, download system and software patches in time, install antivirus software and firewall and other security management software, and create a good network access environment [3]; Secondly, when visiting the Internet, it is necessary to have the ability to identify true and false information, and the amount of information on the Internet is huge. Learning to distinguish right from wrong is an effective measure for information safety and civilized internet access. Finally, for users, it is more important to develop good network access habits. When downloading software from regular websites, you should read and understand the webpage prompts carefully, and don't casually publish personal sensitive information in forums and social networking sites, so as to establish a firm awareness of information safety and protect personal privacy.

3. User network information safety in a distributed environment

3.1. Classification of Web Mining

Web mining refers to the discovery of implicit, unknown, potentially applied, and non-trivial patterns from a large collection of documents. The objects it deals with include static web pages (text, multimedia information, etc.), web databases, internal structures of web pages, web structures, and user usage records. Through the mining of these information, information that cannot be obtained only through text retrieval can be obtained. Web mining is divided into three categories, as shown in Figure 2: content mining, structure mining, application mining [4]. Among them, structure mining is used to extract network topology information—link information between pages. Such as which pages are linked by other pages, which pages point to other pages, etc. Application mining is used to extract information on how customers use their browsers to browse and use links, which pages customers visit, how long they stay on each page, and what they click next. Content mining is used to extract text, pictures or other information that make up the content of web pages. Search engines, intelligent agents, and some recommendation engines all use content mining to help customers find what they need in the vast cyberspace.

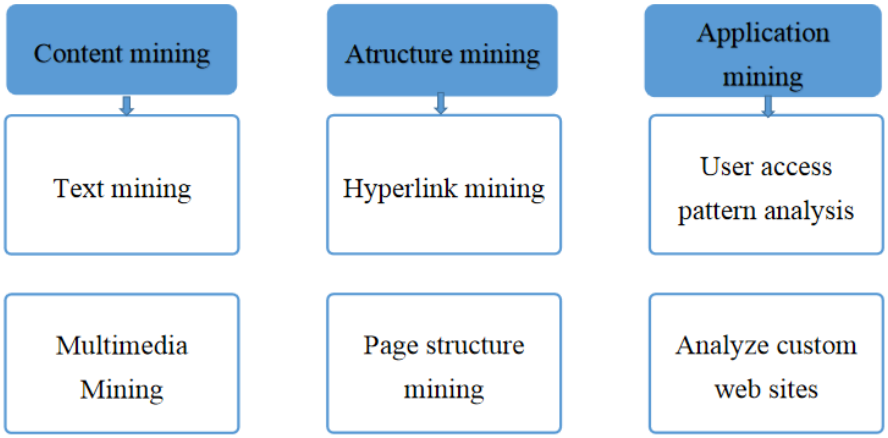

Figure 2 Web mining classification

Why is web data mining (DM) a key technology to improve the performance of network information safety?

All kinds of network servers and proxy servers keep access logs, which record all kinds of important events in the network system. Through them, we can know the running status of the system, monitor user behavior, audit security events, diagnose errors and abnormalities, etc. Although these data provide the possibility of security application, it is often necessary to analyze them at a higher level of abstraction, that is, to identify the relationship between multiple events and events within a period of time, that is, to identify the event patterns, so as to make better use of these data [5]. However, Web DM, through access path analysis, association rule discovery, sequence pattern analysis, classification rule discovery, cluster analysis and other technologies, extracts system characteristic attributes related to security from the acquired resource data, and automatically generates a detection model of security incidents according to the system characteristic attributes, which is used to automatically identify security incidents, so as to avoid manual analysis and coding detection model, greatly improve the efficiency of pattern recognition and rule construction, and make it possible to improve the performance of network information safety.

In recent years, a certain number of academic research achievements have emerged one after another. The serial pattern DM technology based on data security and information protection has been explored, and a DM algorithm based on important serial attribute hiding has been designed, which has achieved effective privacy protection in DM [6]. Privacy protection of big DM should not only protect the relevant privacy data of each site in the process of DM, but also ensure the overall expected effect of DM. In the typical algorithm of DM, the privacy protection technology of data encryption is adopted, and the original data is encrypted by homomorphic encryption technology. In DM, the encrypted ciphertext is directly processed, which not only ensures the security of private data, but also does not affect the effect of DM. Homomorphic encryption technology does not decrypt the original data, but directly performs complex calculation operations on the encrypted data by using big DM algorithms, and can get the same results as the original data operations before data encryption [7].

3.2. Web Access Information Safety Detection Simulation

The principle of user network information safety detection in distributed environment is that, in the process of detecting user network information safety in distributed environment, the main structure sample characteristic information of user network information in distributed environment is obtained first, and the PSO identification tree is established by using the hierarchical structure of access information, and the information entropy corresponding to each layer of the identification tree is obtained. According to the information entropy, the corresponding user network information characteristics in distributed environment are divided, thus realizing the accurate detection of user network information safety in distributed environment. The specific process is as follows:

Assuming that there are m access points $r_1, r_2, \ldots, r_n$ is described, and nodes with the same characteristics of user network information between different access points have probabilistic correlation.Then the probability of all access points is described by $P(r_1), P(r_2), \ldots, P(r_n)$. When $P(r)=1$, the information entropy of web access points is given by formula (1):

$$G(R)=P_1W(r_1)+P_2W(r_2)+\ldots+P_{\mathrm{m}}W(r_n)=P(r)\log_2 P(r) \tag{1}$$

Combining with the information gain theory, the user network information in distributed environment is discretized and extended to the corresponding PSO identification tree, which ensures that the shortest distance between the access point and the identification tree can be obtained by analyzing the correlation between the access point and the identification tree through the running state of the access point in the identification tree [8].

By extracting user demand directional features from key information in user network information in a distributed environment, it is assumed that user network information in different time periods is stable. In the piecewise stationary linear mapping space, TLX and TLY represent their respective judgment users. Network information short-time domain window, using formula (2) to obtain the key information feature extraction decision formula of user network information is:

$$TL_X(X,Y)=\begin{Bmatrix} Test, if\left(GD_X(x,y)\rangle T_X\right) \\ NonText, Otherwise \end{Bmatrix} \tag{2}$$

According to the above analysis, because the user network information in the distributed environment has strong self coupling and complex characteristics of disconnection, it is difficult to detect the user network information safety under the interference of the distributed network platform. A user network information safety detection method based on fuzzy support vector machine (FSVM) theory in distributed environment is proposed [9].

3.3. Construction of distributed network information safety management and control system

In the past, we usually used compound control measures for network information safety problems, but this control measure not only has many levels of control, but also has low control efficiency. Distributed network information safety management and control system is a new management and control system for network information safety. Its management and control servers and clients are designed in different modes, which can effectively guarantee the safe, effective and stable operation of the network [10]. The equipment organization of the distributed network information safety management and control system constructed in this paper mainly includes three levels of equipment organization, namely virtual view, physical view and user view. As shown in Table 1, Table 1 is the equipment organization table of distributed network information safety management and control system.

Table 1 Device organization table of distributed network information safety management and control system

Serial number	Device Organization Name	Main effect
1	Virtual view	Link organization devices to ensure device security
2	Physical view	The basis of various functional calculations
3	User view	Correction to virtual view

The distributed network information safety management and control system requires a variety of technologies, mainly including server virus detection technology, intrusion prevention technology and the construction of intrusion monitoring system technology. Supported by the above technologies, the distributed network information safety management and control system, as a terminal for network information monitoring and management, is not affected by different user IP addresses and IDS. It does not need cross regional interaction and connection, and it does not need to consume too much traffic to work easily and conveniently [11]. As a positive and active intrusion

prevention monitoring means, intrusion prevention technology is mainly installed at the entrance and exit of network equipment. In its detection process, once the intrusion behavior is detected, it will automatically delete the network attack package. The application of intrusion prevention technology not only reduces the possibility of virus intrusion at the source of network information intrusion, but also realizes the intelligent deletion of network information attack packets, which avoids the possibility of hacker intrusion to a certain extent.

4. Conclusions

In a word, the research shows that compared with the traditional network information safety management and control system, the network information safety management and control system in the distributed environment is more convenient and efficient, and the management and control technology is also more advanced, which can effectively make up for the shortcomings of the traditional network information safety management and control system and ensure the security of network information.

References

[1] Liu Ruijun. User Network Information Safety Detection Simulation in Distributed Environment [J]. Computer Simulation, 2017, 034(008):421-424.

[2] Hu Fei, Chen Shuo, Liu Wei, et al. Construction of distributed network information safety management and control system [J]. Information and Communication, 2019(4):2.

[3] Cheng Xu, Huang Taigui, Cheng Qi, et al. Research and practice of distributed power information safety acquisition technology [J]. Journal of Anhui Electrical Engineering Vocational and Technical College, 2018, 23(1):6.

[4] Fu Zhuqiang, Li Xuejiao. Analysis of distributed information safety operation and maintenance management platform based on big data [J]. Vigor, 2019(23):1.

[5] Pan Jinfeng, Hu Xiaoqin. Research on Big Data Information Safety Evaluation Platform Based on Distributed Data Stream [J]. Journal of Bengbu University, 2021(2):84-87.

[6] Jia Xiaoli, Wu Rui, Wu Siying. Parallel and Distributed Bi-level Clustering of Web Access Patterns [J]. Computer Engineering and Applications, 2019, 55(23):7.

[7] Chai Wenguang, Zhou Ning. Research on the Integration of Network Information Safety Prevention and Web DM Technology [J]. Intelligence Theory and Practice, 2009(3):5.

[8] Wang Yan. Research based on the integration of network information safety and Web DM technology [J]. Information Safety and Technology, 2014, 5(12):3.

[9] Zhao Yuepin. Design and Implementation of Network Information Safety Prevention and Web DM System [J]. Modern Electronic Technology, 2017, 40(4):5.

[10] Wang Xiaojun. Design and Research of Network Information Safety Prevention and Web DM System [J]. Electronic Design Engineering, 2018, 26(12):5.

[11] Liu Xiang. The Integrated Application of Network Information Safety Prevention and Web DM Technology [J]. China Management Informatization, 2017, 20(22):2.