A New Information Leakage Defendable Model

Shufen Liu, Zhengxiang Zhang, Yaorui Cui, lintao Wu
College of Computer Science and Technology ,Jilin University, Changchun 130012,China
liusf@jlu.edu.cn;zhangzhengxiang@gmail.com;yrwu@jlu.edu.cn;cuilintaostone@163.com

Abstract

In enterprise's informationization construction, a highly efficient information system inevitably emphasize on the sharing of information, paperless office etc to provide the enterprise information-sharing environment, simultaneously, all data almost can be completely copied in a few minutes, which puts the enterprise information system in high risk, therefore the establishment of an effective and reliable security system has been a top issue to solve the problem of business informationization .But unfortunately, there is not a reasonable solution at present ,which can keep the efficiency of the system while protecting the confidentiality of sensitive information. In view of the above-mentioned fact, an intranet information disclosure defendable security model based on both crypt-isolation and log-audit is proposed. This model with both access-audit and encryption method will reasonably control the enterprise staff's behavior and effectively prevent the enterprise sensitive information from being leaked, intentionally or unintentionally.

Keywords: *information leakage, trusted routing domain, pattern match, identity-based authentication*

1. Introduction

At present, the enterprise information security systems mainly focus on both how to guard against the invasion and to resolve system collapse. Installing a firewall and intrusion detection systems and running antivirus software are the technical means to prevent, and data backup, remote disaster recovery, backup power , system recovery and system cluster etc are used to face the risk of system collapse. How to protect the enterprise information system security has been a weak link of building the entire information security system. Through a large number of cases, it is found that the enterprise internal staff are often operators of divulging the enterprise sensitive information, simultaneously it is also information leakage channel of the most difficult to prevent. This mainly lies in the staff to be in enterprise's intranet, which lets them have the greater opportunity to gain enterprise's sensitive information than the outside net's hackers. The enterprise staff may be through the computer storage media (such as USB storage devices, compact disks and floppy disks, etc.), printers and network intentionally or unintentionally to divulge the enterprise sensitive information to outside the enterprise application environment.

In order to resolve these potential safety problems which be mentioned above, this paper studies the existing solutions and based on these proposes an intranet information disclosure defendable security model based on both crypt-isolation and log-audit, which uses both access-audit and encryption method to build a safeguard system of the enterprise information security. Consequently under the premise of keeping the efficiency of the system, this model resolves the problem that the enterprise sensitive information is leaked by above-mentioned methods.

2. Research for current technologies

To avoid escaping, internet connections, mobile storage devices, floppy, CD-ROM and printers are forbidden by many enterprises. Although these forcible methods stop the staff from escaping the sensitive information out of the enterprise, it also reduces usability of the system, and it is not convenient with the stuff.

Reference to [4], a new security model based on password isolation is supplied. The model separates principle part(processes) and object part(data files) into two classes: high security class and low security class. The sensitive information that is protected is stored in high security data files. The core theory is: when a process reads sensitive information, it will be increased to high security class; when high security processes write data to external devices or network, it will encrypt data and match the created data with high security. As a result, sensitive information of application environment always exists with cryptograph, which forms a virtual classified network of password isolation. Although the model can effectively prevent the enterprise sensitive information from being leaked and to a large extent improve the usability of the system, there are also some problems. Owing to data files which are accessed by high security processes are not always enterprise sensitive information, the model will encrypt some insensitive information, which undoubtedly reduces the efficiency of the system. In addition when high security data files are printed, the model imports a credible

978-1-4244-3291-2/08/$25.00 ©2008 IEEE

terminal which answers for decrypting sensitive information and accomplishing the print processing. It must bring on a system bottleneck. When a staff want to print files, he must first send files to the credible terminal and wait for the completion of former print operations in order to he can start to print. This will bring very big inconvenient to staff's work. In view of the above-mentioned problems, on the existing basis this paper proposes a new solution of protecting intranet information security.

3. Improved intranet security model

In order to solve the problems of the existing solutions and improve the system's usability and efficiency, the model brings out a concept in the field of network division, which is trusted routing domain [3]. This concept divides those hosts which are closely connected and trusted into one trusted routing domain. Those hosts within the same trusted routing domain communicate with each other using proclaimed letters, while hosts separate in different trusted routing domains talk to each other using cryptograph. Otherwise, when write information to external devices or networks, the model use selective encryption techniques. The core idea of the model is : identify the hosts to insure its legal identity. Check on the information flow within hosts, if it needs to be transport by network, then we should make sure whether the destination is in the same trusted routing domain as the host in order to decide whether use cryptograph or not; If there is a need for printing, we should check on the information waiting for printing out; If the information is being transported to mobile storage devices, we should check the security level of the information, and information including any sensitive word needs to be encrypted first and then stored, otherwise, it could be stored directly. The intranet security model brought out by this paper could be described as the following graph:

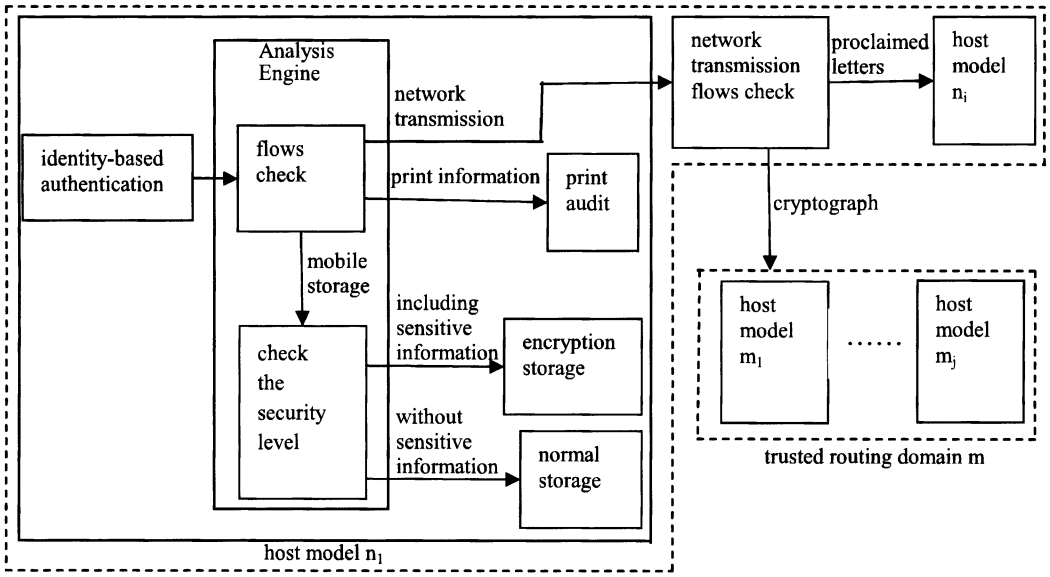

Figure 1. An intranet security model based on both crypt-isolation and log-audit

3.1. Selective encryption

In traditional encryption systems , some insensitive information is also encrypted. In order to solve this problem, the model brings out a concept of selective encryption, which is : the intranet administrator set the match conditions of sensitive information. If the information matches one of these conditions, it will be encrypted first and sent to external devices or networks, otherwise, the information will be sent by proclaimed letters. The searching match algorithm designed by the model consists of the following two parts:

1, pretreatment: using weighted Euclidean distance optimized model based on file statistic eigenvalue weight, and the K-sphere model based on that mentioned by paper[7] and considering the magic number of files, this algorithm can identify the file type correctly and translate document files(word, excel, pdf etc.) into text files. Processing methods of other media files can be defined by intranet security administrator.

2, pattern match: there must be plenty of match conditions of sensitive information in real applications. If using single pattern match algorithm, we shall run the algorithm once again for each condition, which causes low efficiency. However, multiple pattern match algorithm only needs to scan the text flow once and could verify all the match conditions of sensitive information, which extremely improved the efficiency.

As we know that most of the parts need to be processed during the match searching procedure are Chinese documents, and Chinese belongs to large character set language. During the pattern match procedure, there would be a lot of unused characters, and this information can be used to accelerate the speed of matching, which is in accord with the idea of Quick Search Algorithm[2]. So the pattern match algorithm designed by this model is based on the classic AC multiple pattern match algorithm[1], and takes advices from paper[5] about how to absorb the ideas of QS algorithm based on AC algorithm and how to implement error allowed semblable multiple pattern match algorithm for both English and Chinese mentioned by paper[6], so that this algorithm can use the information of matching failures to omit as many characters as it can to accelerate the match speed, and complete the match for most of the match conditions precisely, and for some extremely sensitive match conditions, it will perform the semblable matching.

3.2. Print Audit

In printing data file processing, this model uses the idea of a strong audit. Strong audits the most vital role is "the deterrent". With strong audit mechanism, criminal or negligent acts will be recorded. The original psychological often feel lucky to convergence. Through print time, host IP address, print content, print file name, staff name, subordinate department, print pages, print shares, and other information detailed records, effectively prevent information leakage and illegal print. The concrete implementation of the idea is: first the print bitmap of print content is created. This step is realized through revising print processing in printing course to be possible to print two times, one of which is true print directly using real printer driver print, and the other uses virtual printer driver with which we can generate a print bitmap that we need. Print-generated map is BMP format, occupying a large space, therefore, we need to compress this print bitmap. After compressing the size, the bitmap is original several dozens 1, which can both save the space and raise the efficiency. The print bitmap of being converted size will be stored in the local temporary folder. This is because taking into account after audit server's linkage is interrupted, this bitmap cannot be uploaded directly to the server. At this time first the print bitmap is preserved in local. When detecting usable connection, it is uploaded again. Finally the local temporary folder is deleted in order to not be found by users. For obtaining print pages and printing shares in printing course, the real print during printing processing need be revised, in order to let it acquire shares and pages and preserve them in the temporary file. The idea exhaustively records print information and related print content, at the same time it ensures the true print speed without any impact.

3.3. Identity-based authentication

In order to access the intranet resources with other people's identity, some intention of illegal staff use other IP address and MAC. The model has been improved to prevent that. First of all, get the equipment handle of network Interface Card with the function hMAC = CreateFile(szMACFileName, GENERIC_READ, FILE_SHARE_READ |FILE_SHARE_WRITE, NULL, OPEN_EXISTING, 0, INVALID_HANDLE_VALUE).The parameter szMACFileName is the equipment driver's "filename". It can be found in the register list under the item HKEY_LOCAL_MACHINE \ SOFTWARE \ Microsoft \ Windows NT \ CurrentVersion \ NetworkCards. When the hMAC is specific, the MAC can be obtained with the function DeviceIoControl(hMAC , IOCTL_NDIS_QUERY_GLOBAL_STATS , &OidCode , sizeof(OidCode) , OidData , sizeof(OidData), &ReturnedCount, NULL) (OidData is array of char,OidCode is a var of type NDIS_OID). If OidCod = OID_802_3_PERMANENT_ADDRESS, it returns the MAC set to the NIC when it came out of the factory. And it returns the MAC set to the NIC currently if OidCode = OID_802_3_CURRENT_ADDRESS. It is easy to see whether the user change the MAC arbitrarily by comparing the two MACs.

4 .Experiments analysis

Experiments are accomplished under the local area network where four trusted routing domains are composed of twenty nodes and a server which is used to store audit logs.

4.1. Selective encryption test

During the actual test, a group of test sets are designed in terms of the match conditions of sensitive information, and they are tested separately in 20 nodes, with the result shown in Fig.2.

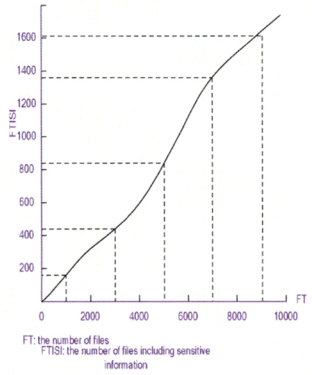

Figure 2. Selective encryption test

It can be seen from the figure that the selective encryption idea used in the secure model can accurately examine files containing sensitive information and the accuracy is above 99.9%. Thereby without any influence in system efficiency, the problem that sensitive information of the enterprise is leaked by output devices or network has been solved.

4.2. Print audit test

A computer is selected in every one of the four trusted routing domains to complete the print audit test, and testing results can be viewed in the server, as shown in Fig.3 and Fig.4.

IP address	File name	Time	Pages	Shares	User name	Department
10.80.187.139	project document.doc	2008-03-17 19:15:56	1~12	1	david	one department
10.80.197.138	project log.doc	2008-03-17 20:15:25	1~2	1	john	two department
10.80.207.126	log document.xls	2008-03-20 15:26:35	2~16	1	jane	planning department
10.80.217.122	conference record.doc	2008-03-24 09:06:34	1~1	6	edward	finance department

Figure 3. Print basic information monitoring results

Figure 4. Print content monitoring result

It can be seen from the above results that the secure model depends on strong audit idea to be able to accurately record all kinds of printing information. As a result, the strong precaution and deterrent forces prohibit the internal staff from illegal operations.

5. Conclusion

In view of the security problems in the process of enterprise's informationization , an intranet information disclosure defendable security model based on both crypt-isolation and log-audit is proposed. This model with both access-audit and encryption method effectively ensures the security and integrity of enterprise sensitive information. The solution doesn't only show the advantages of the traditional solution including goal, robustness and security, but also supply better availability and efficiency, which keeps the efficiency of the system while protecting the confidentiality of sensitive information. How to effectively combine with both the trusted computing platform of the whole network and IDS systems is the further work.

References

[1] Aho A V, and Corasick M J. Efficient string matching:an aid to bibliographic search. Communications of the ACM , 1975, 18(6): 333~340

[2] Sunday M Daniel. A very fast substring search algorithm. Communications of the ACM, 1990, 33(8): 132~142

[3] Huang D J, Cao Q, Sinha A, and et al. New architecture for intra-domain network security issues. Communications of the ACM, 2006, 49(11): 64~72

[4] Zhao Yong, Liu Jiqiang, Han Zhen, and Shen Changxiang. The application of information leakage defendable model in enterprise intranet security. Journal of Computer Research and Development, 2007, 44(5): 761~767

[5] Wang Yongcheng, Shen Zhou, and Xu Yizhen. Improved algorithms for matching multiple patterns. Journal of Computer Research and Development, 2007, 44(5): 761~767

[6] Gao Peng, Zhang Deyun, Sun Qindong, Zhai Yahui, and Lu Wuchun. A multiple approximate string matching algorithm of network information audit system. Journal of Software, 2004, 15(7): 1074~1080

[7] Zheng Jie, Luo Junyong, and Lu Bin. Documents type identification based on statistical characteristic. Computer Engineering, 2007, 33(1): 142~144