Acta Technica *62* No. 2B/2017, 53–62

Exploration and implementation of user behavior forensics analysis system of computer network based on system log[1]

Wenzhe Lu[2, 3, 4]

Abstract. In order to make the forensic analysis of crime occurred in the computer system and computer network, and access to the electronic evidence of invasion, based on the background of "research on security guarantee system for large-scale network", the forensic analysis of computer network user behavior based on log system is made. First of all, the computer network intrusion forensics technology is analyzed, and then the functions of forensic analysis system are designed. In addition, the working flow and function of two subsystems of log collection agent and forensics server are focused on and realized. Moreover, the performance of the system is tested and analyzed from the system log collection performance, data transmission performance and other aspects. The test results show that the system has the basic function of network forensics, and meets the design requirements. In a word, the system designed basically meets the expected purpose.

Key words. Computer crime, forensic analysis, system log, log collection.

1. Introduction

With the development and popularization of computer and network technology, the current computer crime shows a growing trend. The loss brought about by the computer crime is incalculable, which brought great damage to the national life and social stability and development. But when the criminals made use of computer for crime, there was no body description nor fingerprints, let alone others to be used as circumstantial evidence. While the evidence is always in the form of electronic information, easily broken and changed, so the research on computer

[1]The author acknowledges the National Natural Science Foundation of China (Grant No. 51578109), the National Natural Science Foundation of China (Grant No. 51121005).

[2]Computer Network Information Center, Chinese Academy of Sciences, Beijing, 100080, China
[3]University of Chinese Academy of Sciences, Beijing, 100080, China
[4]North Long net (Beijing) science and Technology Co., Ltd., Beijing, 100190, China

forensics technology is particularly important [1]. As a result, this paper intends to take the system log records of criminal behavior as the foundation, to study the methods of computer network forensics. A LAN oriented distributed computer network user behavior forensics based on system log is designed. It is expected to make statistics and correlation analysis of the system log through the system, and to supply it for emergency response reference or as the court evidence. As a result, criminals should pay the price for their crimes and consequences, so as to purify the network environment, renovate the network order, and safeguard the national security and social stability.

2. Experimental procedure

2.1. Computer network intrusion forensics based on system logs

The implementation model of computer network forensics based on system logs: Based on the feasibility and legal requirements of the system log as electronic evidence, this paper first of all proposes a computer network user behavior forensics analysis system model based on system log. The main idea of designing computer network user behavior forensics analysis system is to collect the system logs of the key servers in a LAN to a special forensic server for collective storage and forensics analysis. The theoretical model of the system is shown in Fig. 1.

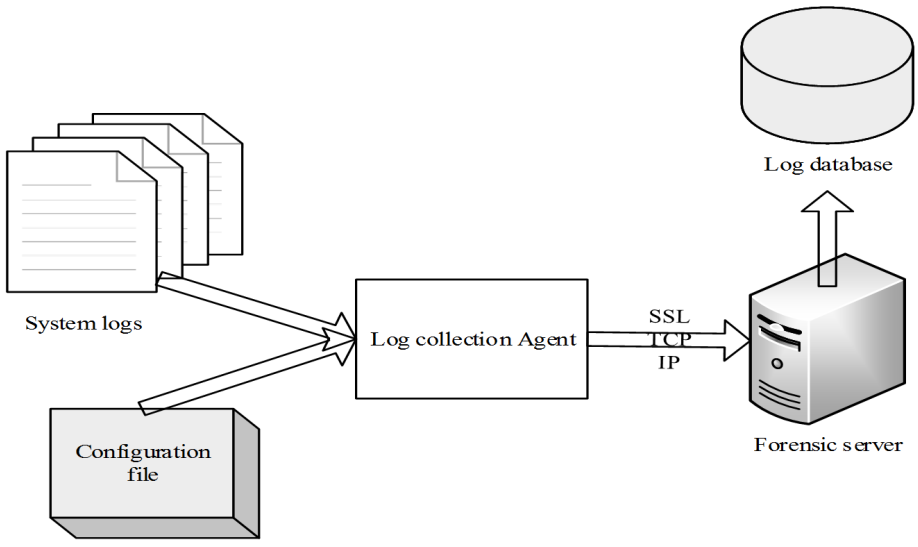

Fig. 1. The computer network user behavior forensics analysis system model based on system log

Figure 1 shows that the system consists of two parts: log collection agent and forensics server. The system uses client/server structure, and the log collection Agent is the client side, deployed in the protected servers in the LAN, responsible

for the integrity verification, acquisition and sending of system log information on the target machine. The configuration file defines the parameters of collection agent; forensics server is the server terminal of system, which is responsible for the receiving and preservation of log data sent by Agent. In addition, it is also necessary to make the integrity protection and verification of the original log data [2]. At the same time, the forensics server is responsible for preprocessing and management of log data. The forensics analysis of log data is the most important function of forensics server. According to the instructions of forensic personnel, statistics analysis and association analysis is needed for log data in the database.

Log forensics analysis:

Computer forensics analysis is confronted with massive data and it cannot be analyzed manually one by one. It is necessary to use computer system to screen evidence materials related to computer crimes. The log statistics analysis process is shown in Fig. 2.

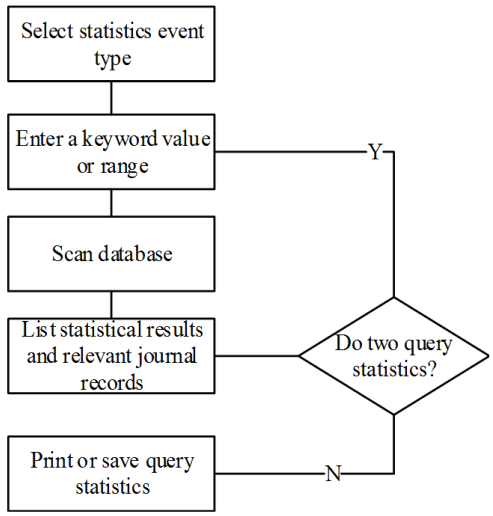

Fig. 2. Statistical analysis flow chart

2.2. Functional module design of computer network user behavior forensics analysis system

The system is a distributed computer network forensics system based on agent. It takes a LAN as the application background, and the system log of server in LAN for forensics object. The overall goal is to make a forensic analysis of the intrusion occurred in the important LAN server, and to dig out the illegal behavior evidence of network users in the system log [3]. In this paper, for the computer network user behavior forensics analysis system based on the system logs, the specific module is shown in Fig. 3.

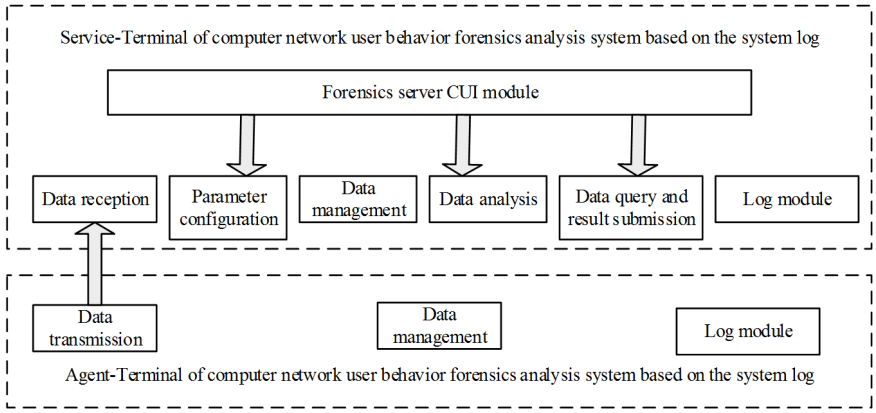

Fig. 3. System overall function module diagram

2.3. Major functions implementation of computer network user behavior forensics analysis system

Log collection agent sub-function implementation:

The log collection agent subsystem consists of three modules. Among them, the data management module is responsible for the data storage format conversion, data integrity verification and calling action and the corresponding time stamp in the log module recording of data management process, and the abnormal changes of system log. The data transmission module is responsible for collecting the initialization parameters and dynamic adjustment, the establishment and management of the SSL channel, regular acquisition of data log according to the configuration parameters. And through the SSL channel, it is transmitted to the server, and at the same time, it completed the recording of data transmission process and history. The log module is mainly responsible for accepting the calling of other modules, recording the process of collection agent end forensics operating, abnormal cases of system log files, and the corresponding time stamp information, so as to maintain the integrity of evidence chain. The work-flow of the log collection agent is shown in Fig. 4.

Implementation of log forensics server sub-functions:

The log forensics server module is one of the core modules of the system, which is mainly responsible for the acquisition of log information on the server. Among them, the data receiving module mainly receives the log data from the log collection agent entity, which can complete the establishment of SSL channel, receive data and synchronization, and record the receiving process. The data management module can store all kinds of logs in layers according to the source IP address, log type, and receiving time. For log files in the same type and the same source, they are merged to a paper file. At the same time, this part functions, for the log data received, can carry out data integration, data cleaning, data reduction and so on pretreatment [4]. And the CES algorithm is adopted to make signature encryption of original files received, to conduct ACL control of the log file, so as to ensure that only those who have the administrator privileges can access to. The data analysis module is mainly

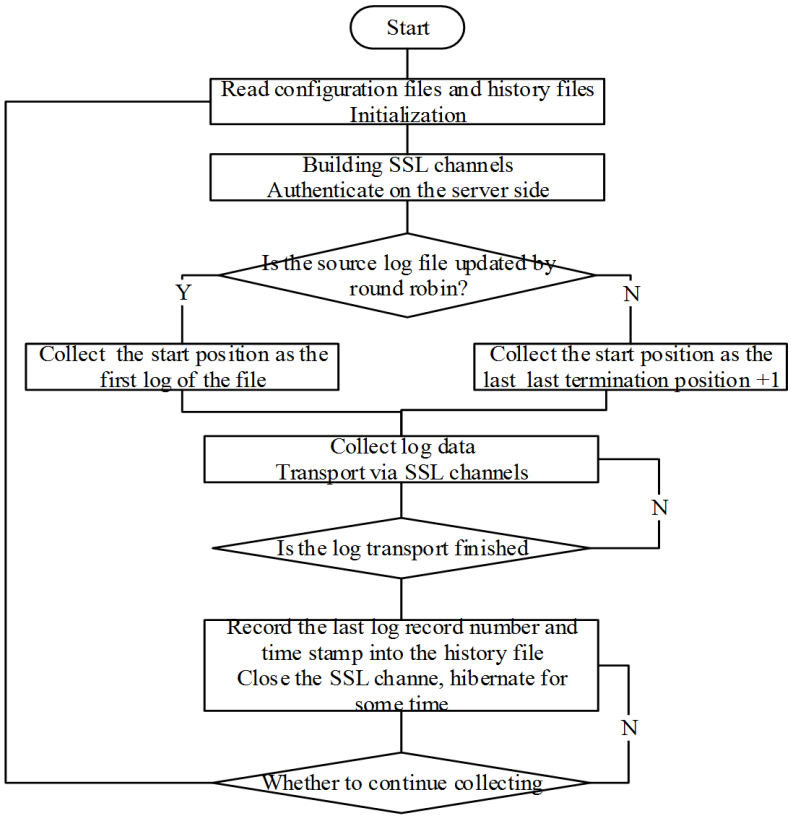

Fig. 4. Work-flow chart of log collection agent

based on the instructions of forensics personnel for statistical analysis, clustering and correlation analysis of log data, to form the forensic analysis report, and to send the forensic analysis results to the GUI module and display in an appropriate way. The data query and submit module can make all kinds of queries of the original log files according to forensics analysis results, extract the query result, and perform the signature lock using CES algorithm, so as to form the log evidence file. This module should also have the function to verify the integrity of the extracted log file. The parameter configuration module is responsible for receiving the instructions sent by the GUI module about modifying server parameters configuration. Modify the corresponding parameter in the configuration file can control the corresponding parameters configuration in the data receiving module and data processing module. The log module is designed for being called by the other modules and recording the whole process of data processing and abnormal situations in the forensics process [5]. Gu I module is mainly responsible for modifying the server configuration parameters, accepting statistical analysis, association analysis, query extraction and so on request operations on the interface of users, and transmitting the request to other modules. In the meanwhile, it calls the data of log module and displays the data processing

record. The work-flow of the forensics server is shown in Fig. 5.

The users interact with the forensics server through the GUI interface, and start the log receiving module, parameter configuration module, data analysis module, and query module through the start menu items. In addition, uses complete the server configuration, receiving and management of log data, forensic analysis of log data, query and log data extraction and submit, forensic results integrity verification and so on. The data management module and forensics log module are called by other modules in the process of forensics.

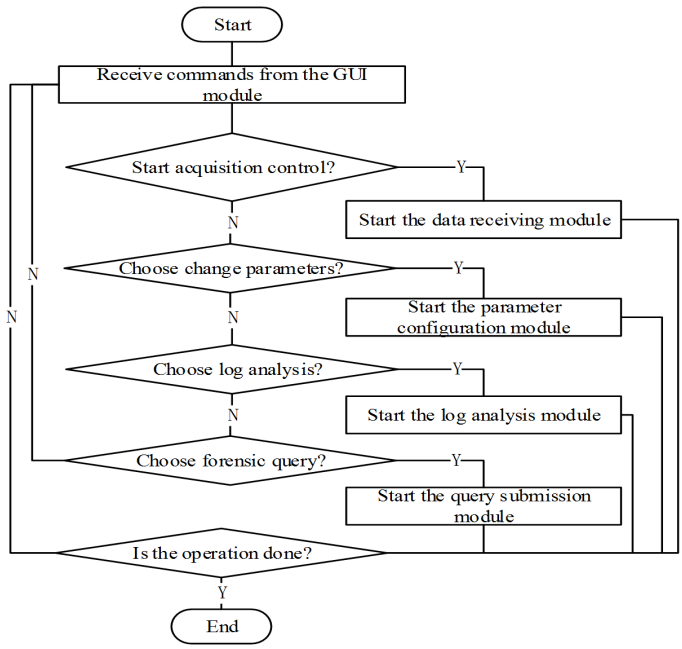

Fig. 5. Work-flow chart of log forensics server

Interface implementation:

The log collection agent interface is shown in Fig. 6. The client configuration file ssl_client.conf provides the relevant parameters setting of client, including client certificates and keys, evidence server IP address and port, acquisition time interval and acquisition object, client log file name and storage path and so on information. By directly editing the file, we can change the parameters of client; history record file send_history.txt records the history information that each log transfers, for the reference for the next transmission, so as to achieve the integrity verification of incremental transmission and log files. The client log file records the process, log management, abnormal cases and so on information in the process of each log transfer, and it is filled in and inquired by the client log module. The interface relationship of the forensics server subsystem is shown in Fig. 7.

Design and implementation of forensic database:

The forensic database is used to save the processed log data, and the database can

provide us with quick queries. Relational databases can also realize the association query of multiple database tables. Database queries can be simply implemented only by using SQL statements, which are incomparable for queries bade on file. Therefore, this system uses the database based scheme in statistical analysis and correlation analysis of logs.

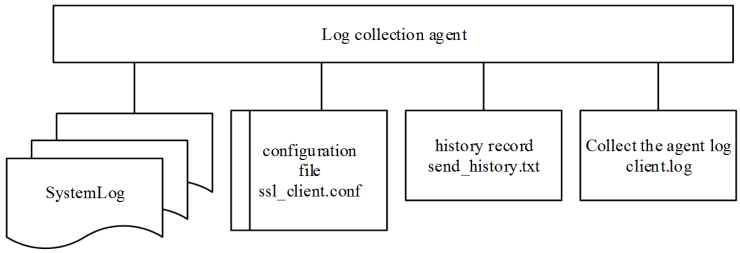

Fig. 6. Interface to a forensic server subsystem

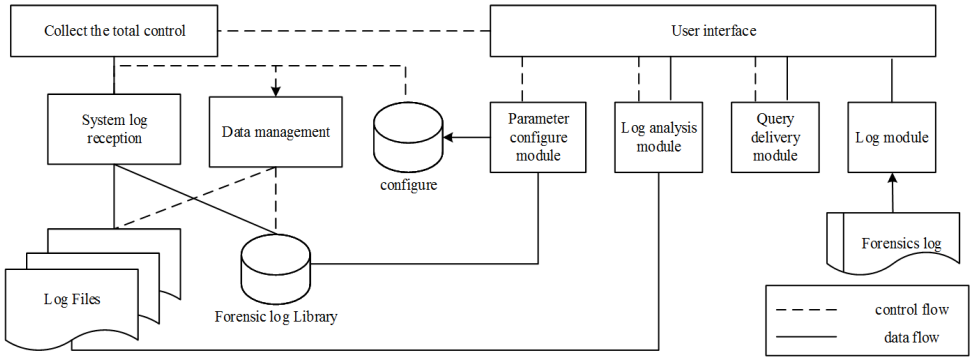

Fig. 7. Interface to a forensic server subsystem

The system uses relational database, using open source database software MySQL to build the database, MySQL has cross platform, simple management and operation, able to handle large data, quick query and so on advantages. The database is defined based on a log, a data table. The field of table is basically corresponding to attribute domain of various log record tables, just removing some irrelevant attribute domains and adding Logsource (log source, the source host IP address), Result (possible outcome of events) and Condnion (event condition) [6]. The fields are set to facilitate the subsequent forensic analysis, and to fill in the value of these fields in the data preprocessing of log data written to the database.

3. Results and discussion

After the completion of construction of computer network user behavior forensics analysis system based on the system log, it is necessary to test various functions of the system and determine the performance of the system, so as to verify the feasibility of

the system. For the design of the computer network user behavior forensics analysis system, part of the test results are as follows.

3.1. Acquisition performance test

It mainly tests the adaptability of collecting agents to load changes and the rationality of acquisition model. The test method is to change the generation speed of system log through the conscious file, directory operations and process operations.

During the test process, when the speed of the log data changes rapidly, the acquisition agent can automatically adjust the acquisition time interval, and timely collect and transfer the log data, which executively protect the log security. When the speed of log data generation varies little, the acquisition time interval remains stable. Experiments show that the system has a certain self-adaptability. However, there is a certain lag in the adjustment action, and the choice of time interval increments affects the sensitivity of adjustment. The calculation of the initial time interval will also affect the subsequent dynamic adjustment. If it is too large or too small, it will lead to frequent adjustment of the time interval during the acquisition process. The choice of the minimum time interval should be determined by the maximum load of the system.

3.2. Transmission performance test

Through the SSL channel packet transmission, the speed and accuracy between the acquisition agent and forensic agent are mainly tested. Table 1 gives the time taken for different size log data transmissions with different lengths of the RSA key (the data transmission buffer takes 4096 bytes). As can be seen from the table, the length of the RSA key has little effect on the transmission time. This is because in the SSL protocol, the RSA public key algorithm is used to exchange the session key, the data transmitted is encrypted with a symmetric key algorithm DES and CBC3 [7]. The changes in key length RSA only affect DES session key encryption speed. DES key is only 64 bits long, and a data acquisition only exchanges for a session key. In the case of software implementation, the DES algorithm is 10 times faster than the RSA algorithm, and 100 times faster than the hardware implementation. As can be seen from the table, the data transmission time is directly proportional to the data size. The SSL channel transmission data can reach an average of 600 KByte/second, which fully meets the needs of log transmission of server in a medium-sized and small-sized LAN. As long as the suitable acquisition time interval is selected, it will not cause the evidence server congestion and denial of service.

Another factor affecting the transmission speed of the acquisition is the log buffer size, which is the size of log data for one time transmission of a client and server. Table 2 gives the effect of buffer on the log transmission speed (RSA key length is 4096 bits). As can be seen from the table, the buffer size is directly proportional to the data transmission speed. However, as discussed in [8], when a transmission data is less than the buffer size, the capacity of the buffer has little effect on the transmission speed. Of course, the buffer cannot be too large. On the one hand, it is

because the sampling interval is not too large, and the amount of log data generated will not be too great. On the other hand, the buffer is too large so that it requires fragmentation and reassembly in the TCP message, and large buffer requires large memory that it will affect the performance of the host server. In combination with the above factors, the system buffer is set to 4096 bytes in size.

Table 1. Transmission performance test

	RSA key length			
Data size	1024 bits	2048 bits	3072 bits	4096 bits
1000 kB	1.613	1.583	1.604	1.617
2000 kB	3.195	3.275	3.149	3.249
4000 kB	6.309	6.449	6.374	6.509
8000 kB	12.528	12.613	12.463	12.614
12000 kB	18.797	18.867	18.591	18.572

Table 2. Effect of buffer size on the speed of log transfer (time unit: seconds)

	Buffer size		
Data size	1024 bytes	2048 bytes	4096 bytes
8000 kB	47.874	24.419	12.614
20 kB	0.150	0.095	0.080
10 kB	0.050	0.050	0.047

3.3. Log analysis test

Log analysis test mainly tests the accuracy and reliability of system's analysis of log data. After the test, the system can count on the log data according to the type of event, establish a network and user behavior, and find the abnormal log records. The two times statistics provided by the system can further narrow the scope for analysis. The number of data in the database table is an important factor affecting the performance of statistical analysis. The greater the amount of data, the slower the query and statistics, and the greater the demand for memory at the same time. Therefore, the database must be backed up regularly, with only a week's data left in the database and reimported into the database when needed.

In association analysis, the system can find the relevant log records of security events according to the time stamp and relevant features of log records. Factors affecting the accuracy of correlation and analysis velocity are the main selection of $\triangle t$ value, correlation property, and relevant log table, and the definition of condition and result field value in the pretreatment. What is more, the accuracy of clock synchronization network is also an important factor.

The memory size of the forensic server also affects the performance of log analysis, and large memory support is required for database query, log query and extraction.

4. Conclusion

Through in-depth research and analysis of computer forensics technology, system log forensics analysis and other related technologies, computer network user behavior analysis system based on the system log is designed. In the design process, this paper chooses client/server structure as a whole system structure. At the same time, it makes specific design and implementation of two sub-systems log acquisition agent and forensics server constituting the whole system. The work flow is given, and the system database and interface are designed. Finally, the performance test of the designed system is carried out. The analysis results show that the system has basically reached the expected goal, to achieve the requirements of computer forensics.

References

[1] H. Alipour, Y. B. Al-Nashif, P. Satam, S. Hariri: *Wireless anomaly detection based on IEEE 802.11 behavior analysis.* IEEE Transactions on Information Forensics and Securityy *10* (2015), No. 10, 2158–2170.

[2] J. Yan, H. He, X. Zhong, Y. Tang: *Q-learning-based vulnerability analysis of smart grid against sequential topology attacks.* IEEE Transactions on Information Forensics and Security *12* (2017), No. 1, 200–210.

[3] S. Oh, M. Pandey, I. Kim, A. Hoogs: *Image-oriented economic perspective on user behavior in multimedia social forums: An analysis on supply, consumption, and saliency.* Pattern Recognition Letters *72* (2016), 33–40.

[4] R. Liu: *Research on IPV6-based computer crime evidence dynamic forensics technology.* IEEE International Conference on Communication Systems and Network Technologies, 4–6 April 2015, Gwalior, India, IEEE Conference Publications (2015), 720 to 724.

[5] K. C. Seigfried-Spellar, N. Villacís-Vukadinovič, D. R. Lynam: *Computer criminal behavior is related to psychopathy and other antisocial behavior.* Journal of Criminal Justice *51* (2017), 67–73.

[6] S. He, J. Zhu, P. He, M. R. Lyu: *Experience report: System log analysis for anomaly detection.* IEEE International Symposium on Software Reliability Engineering (ISSRE), 23–27 October 2016, Ottawa, ON, Canada, IEEE Conference Publications (2016), 207–218.

[7] Y. Shi: *Research of social network log analysis system based on MongoDB.* Open Automation and Control Systems Journal *7* (2015), No. 1, 1621–1628.

[8] T. Grance, S. Chevalier, K. K. Scarfone, H. Dang : *Guide to integrating forensic techniques into incident response.* NIST Special Publication (2006), Report No. 800–86.

Received August 7, 2017