Journal of Information Hiding and Privacy Protection
DOI:10.32604/jihpp.2021.017497
Article

A Fast Detection Method of Network Crime Based on User Portrait

Yabin Xu[1,2,*], Meishu Zhang[2] and Xiaowei Xu[3]

[1]Beijing Key Laboratory of Internet Culture and Digital Dissemination Research, Beijing, 100101, China
[2]Computer School, Beijing Information Science and Technology University, Beijing, 100101, China
[3]Department of Information Science, University of Arkansas at Little Rock, Little Rock, Arkansas, 72204, USA
[*]Corresponding Author: Yabin Xu. Email: xyb@bistu.edu.cn
Received: 01 February 2021; Accepted: 29 March 2021

Abstract: In order to quickly and accurately find the implementer of the network crime, based on the user portrait technology, a rapid detection method for users with abnormal behaviors is proposed. This method needs to construct the abnormal behavior rule base on various kinds of abnormal behaviors in advance, and construct the user portrait including basic attribute tags, behavior attribute tags and abnormal behavior similarity tags for network users who have abnormal behaviors. When a network crime occurs, firstly get the corresponding tag values in all user portraits according to the category of the network crime. Then, use the Naive Bayesian method matching each user portrait, to quickly locate the most likely network criminal suspects. In the case that no suspect is found, all users are audited comprehensively through matching abnormal behavior rule base. The experimental results show that, the accuracy rate of using this method for fast detection of network crimes is 95.9%, and the audit time is shortened to 1/35 of that of the conventional behavior audit method.

Keywords: User portrait; abnormal behavior audit; network crime; abnormal behavior rule base

1 Introduction

With the rapid development of network technology, a large number of e-government and e-commerce systems based on traditional network platforms or cloud platforms are increasingly put into use, aiming to provide users with convenient remote services. However, some systems have low maturity, insufficient security considerations, and insufficient precautionary measures. There are often some operation bugs or security holes, which provide some network hackers or network criminals with opportunities. By pretending to be normal users, abusing and maliciously using resources, they wait for opportunities to change or steal important data, carry out vulnerability scanning and port sniffing, even implant Trojans or launch attacks, which pose a serious threat to some e-government systems or e-commerce systems built on traditional network platforms or cloud platforms.

Although these platforms are usually deployed with firewall, intrusion detection system, intrusion prevention system and other security protection measures, they are often powerless to disguise users who are legally registered or uninvited guests who across the wall. Although these platforms also usually have a logging system that can be audited after the fact, due to the large number of users in the platform, the amount of real-time recorded operational behavior data is so huge that it is difficult to handle with conventional storage and computing resources. Therefore, the audit efficiency is very low, and the accuracy is even more difficult to guarantee.

Therefore, researching and designing an efficient auditing method for network crime can help find suspicious users and detect network crime cases timely and accurately. It is of great significance to effectively guarantee the security of network platforms or cloud platforms.

Data shows that repeated crime is common [1,2], and there are many researches on criminal record [3,4]. At present, relevant theories and methods have been widely used in police case solving. In view of this, this paper applies the idea to the audit of network crime, and proposes a method of detection of network crime based on user portrait. The main research ideas are as follows: In advance, construct the abnormal behavior rule base for all kinds of abnormal behaviors, and construct the user portrait including basic attribute tags, behavior attribute tags and abnormal behavior similarity tags for the network users who have ever had abnormal behaviors. When a network crime occurs, firstly, according to the category of network crime, get the corresponding tag values of all user portraits; then quickly audit the online users who have had abnormal behaviors by matching user portraits to lock the network criminal suspect approximately. If no suspect is found, the method of matching abnormal behavior rule base is used to audit all users.

The innovations of this paper are as follows:

(1) Construct user portrait for network users who have ever had abnormal behaviors. Through constructing basic attribute tags, behavior attribute tags and abnormal behavior similarity tags, the behavior characteristics of users are accurately depicted; according to the statistics and analysis results of repeated crime rate, the key audit is carried out for these users.

(2) An efficient detection method of network crime is proposed. When the network crime occurs, we can find the most suspicious first k users by quickly matching user portraits, and then audit them, which can greatly improve the efficiency of behavior audit.

2 Related Work

In order to detect network intrusion effectively and accurately, some scholars put forward the network intrusion detection model, and research and design the corresponding intrusion detection methods. Denning [5] proposed the intrusion detection model for the first time by monitoring the audit records of the system to find abnormal patterns used by the system, thereby detecting violations. Jia et al. [6] established a convolutional neural network model and applied it to network intrusion detection, effectively improved the classification accuracy of intrusion detection. Wei et al. [7] proposed a new intrusion detection method based on deep belief network (DBN) and extreme learning machine (ELM), combining the ability of automatic feature extraction of DBN and the advantages of rapid learning and good generalization of ELM, which improved the recognition rate and operation efficiency of intrusion detection. Creech et al. [8] proposed a new host-based intrusion detection method, which applied semantic structure to kernel-level system calls, improved detection rate and reduced false alarm rate. Xie et al. [9] combined the intrusion detection with single class SVM algorithm and proposed an anomaly detection system of ADFA-LD data set. The proposed technology can achieve acceptable performance while keeping the computing cost at a low level.

The above intrusion detection methods can detect network intrusion behavior to a certain extent, but it can only be used to distinguish normal and abnormal behavior, and cannot audit specific users.

The log records the process of all activities related to the security operation of the user, which is an important basis for security event tracing and forensics analysis [10]. By mining the log data, we can find the abnormal behavior of malicious users. For the log data set containing UNIX commands, Schonlau et al. [11] used the method of mathematical statistics to extract the number of users, the proportion of total times of command occurrence and the proportion of times of single user command occurrence, and used these characteristics to analyze whether there is abnormal behavior in the data set. Duan et al. [12] summarized the characteristics of 22 common security threat events and wrote them into the rule base, and detected abnormal behaviors by matching logs with the rule base.

Wang et al. [13] established a log analysis model that used signature-based methods and related analysis algorithms to extract security events from collected logs with acceptable false positives. Zhao et al. [14] designed and implemented a pattern mining algorithm for the host operating system log based on the idea of

association rules, which transformed the process-level log data describing the system operation into the log describing the user behavior, and provided more appropriate information for the cloud data security audit. Tian et al. [15] proposed a block-based log recording method and a public audit method based on binary audit tree, which can realize error location and support selective verification of multiple log blocks.

The above log audit methods can analyze the user's behavior through the user's log, but the accuracy of judging whether the user's behavior is abnormal needs to be improved.

Giuseppe et al. [16] put forward the concept of user portrait, which is widely used in precision marketing, recommendation system and other scenarios [17,18]. Zhu et al. [19] used user portrait for abnormal behavior detection, constructed user portrait of normal users, used the method of machine learning to learn the behavior of normal users, and judged whether the tested behavior was abnormal through Mahalanobis distance and isolated forest algorithm. G. Zhao et al. [20] used basic tag, function tag, level tag, and behavior tag to establish user portrait of normal users, and then used pattern matching algorithm to detect anomalies and achieve precise positioning. Wang et al. [21] defined a variety of tags, established user portrait for normal users, obtained user behavior characteristics through various means such as log mining, and matching with user portraits, to find abnormal behaviors of users.

The above abnormal behavior detection methods based on user portrait are all to judge whether the behavior is normal by constructing user portrait for normal users and analyzing the behavior of normal users, but they cannot distinguish the types of abnormal behavior, and the description of user behavior is not comprehensive.

3 Construction of Abnormal Behavior Rule Base

In order to accurately judge whether the user's behavior is abnormal, this paper constructs an abnormal behavior rule base for each abnormal behavior, and judges whether the user's behavior is abnormal by matching the user's behavior sequence with the abnormal behavior rule base.

Each abnormal behavior rule base contains all the rules of the abnormal behavior. Each rule contains an abnormal behavior number and an operating frequency. This paper uses the behavior sequence of all abnormal users in the database to construct an abnormal behavior rule base.

The steps to construct the abnormal behavior rule base are as follows:

Step 1: Count the behavior and frequency of each abnormal user.

Step 2: Use the hierarchical clustering algorithm to cluster users according to the distance of the behavior sequence. The distance J (A, B) between the two sequences A and B is calculated as follows:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A|+|B|-|A \cap B|} \quad (1)$$

where, $0 \leq J(A,B) \leq 1$.

Step 3: Get the common operations and operation frequency of each class cluster.

Step 4: Calculate the average frequency of each common operation as the frequency in the abnormal behavior rule, and get the abnormal behavior rule: {operation 1: average frequency; operation 2: average frequency Operation n: average frequency}. (Assume there are n common operations).

4 Construction and Update of User Portrait

4.1 Construction of User Portrait

4.1.1 User Portrait and Its Composition

User portrait is a highly refined version of the user, which transforms the user's attribute characteristics into regular and computer-readable data format. It is mainly used in precision marketing, recommendation system and other scenarios. To characterize the user's behavior more accurately, this paper constructs a user portrait for each user who has ever had abnormal behavior. The composition of the user portrait is shown in Tab. 1.

Table 1: Composition of user portrait

first-level tag	second-level tag
basic attribute	user category
behavior attribute	historical abnormal behavior statistics
	behavior statistics
abnormal behavior similarity	abnormal behavior similarity

4.1.2 Construction of Basic Attribute Tags

Basic attributes are static attributes of the user, such as: user category, permission level, etc. Since the data set used in this paper only contains Linux commands, the basic attribute tag in this paper only include user category. The user category includes three categories: ordinary user, super user, and administrator.

The basic attribute tag can be directly obtained by the information on the user's data.

4.1.3 Construction of Behavior Attribute Tags

Behavior attribute tag include two sub-tags: history abnormal behavior statistics and behavior statistics. Among them, the historical abnormal behavior statistics tag records the number of abnormal behaviors of the user, which can intuitively display the historical abnormal behaviors of the user; the behavior statistics tag calculates all the operations and frequency of the user in a period of time.

The historical abnormal behavior statistics tag is expressed in the form of a dictionary: {abnormal behavior 1: number 1, abnormal behavior 2: number 2 abnormal behavior n: number n}. During initialization, the number of abnormal behaviors is 0. When abnormal behaviors occur, the corresponding number of abnormal behaviors is updated.

The steps for constructing behavioral statistics tags are as follows:

Step 1: read the behavior sequence of the current user;

Step 2: get the operation number contained in the sequence;

Step 3: count the number of each operation;

Step 4: get the behavior statistics tag, which looks like: {operation 1: number 1, operation 2: number 2, ..., operation n: number n}.

4.1.4 Construction of Abnormal Behavior Similarity Tags

The abnormal behavior similarity is the similarity between the user's behavior sequence and various abnormal behavior rule bases. For each abnormal behavior, the similarity between the user behavior sequence and the rule base is calculated by matching the user's behavior sequence with the rules in the abnormal behavior rule base.

For each abnormal behavior, the construction steps for abnormal behavior similarity tags are as follows:

Step 1: Get the user's behavior sequence F.

Step 2: Calculate the similarity between the behavior sequence F and each rule in the abnormal behavior rule base by matching F with each rule in the abnormal behavior rule base. The calculation formula of the similarity S_i between the behavior sequence F and the rule *i* is as follows:

$$S_i = \frac{|m_1|+|m_2|+\cdots+|m_n|}{|s_1|+|s_2|+\cdots+|s_n|} \quad (2)$$

where, $|s_1|$, $|s_2|$, …$|s_n|$, respectively represent the average frequency corresponding to operation 1, operation 2, ..., operation n in the rule; and $|m_1|$, $|m_2|$, ... $|m_n|$ respectively represent the frequency of corresponding operations in the behavior sequence F.

In order to more accurately calculate the similarity between user behavior sequence and rules, this paper designs some rules for the value:

$$|m_i| = \begin{cases} |m_i| & if\ |m_i| \le |s_i| \\ |s_i| & if\ |m_i| > |s_i| \end{cases} \quad (3)$$

Step 3: Calculate the abnormal behavior similarity, that is, the similarity between the user behavior sequence and the rule base. Since the user behavior sequence matches any rule in the rule base, it means that the user's behavior belongs to the abnormal behavior. In this paper, the maximum value of the similarity between the user behavior sequence and the rule in the rule base is used as the similarity of the abnormal behavior.

The calculation formula of the abnormal behavior similarity SIM_t of the user behavior sequence to the abnormal behavior t is as follows:

$$SIM_t = \max_{0 \le i \le m} \{S_i\} \quad (4)$$

where, m is the number of rules in the rule base of abnormal behavior t.

After three kinds of tags are constructed, a complete user portrait is formed. Fig. 1 is an example of user portrait. Since the data set contains six types of abnormal behavior.

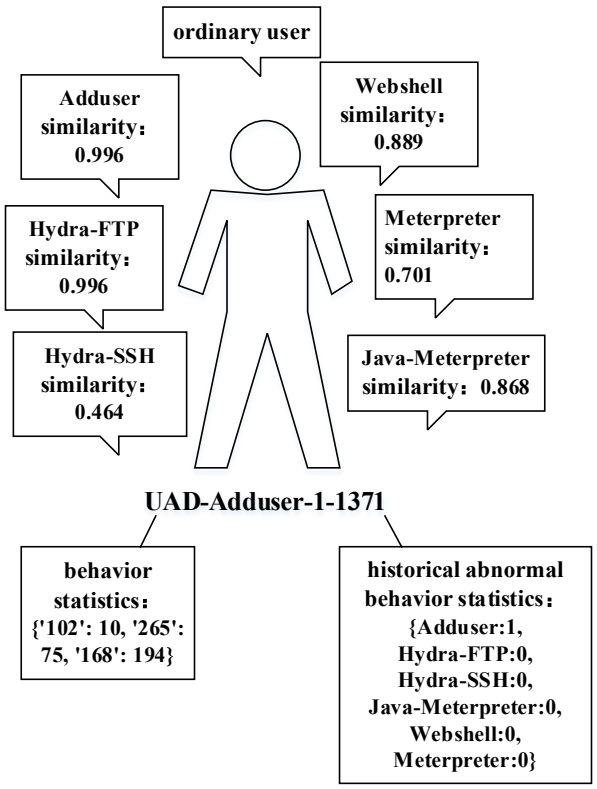

Figure 1: An example of user portrait

Adduser, Hydra-FTP, Hydra-SSH, Java-Meterpreter, Meterpreter, and Webshell, the abnormal behavior similarity tags contains 6 tags.

4.2 Update of User Portrait

To find the abnormal behavior of uses in time and deal with it as soon as possible, the update frequency of the user portrait can be set according to the actual situation, and the update time can be selected as the least busy time.

The updating process of user portrait is as follows:

Firstly, judge whether the user's behavior is abnormal according to the user's behavior attribute tags and abnormal behavior similarity tags: if the user's behavior is abnormal behavior, add 1 to the corresponding number of historical abnormal behavior.

Then, the user's behavior sequence is updated, and the user's operation and frequency are counted to construct the behavior statistics tags.

Finally, the current user's behavior sequence is matched with the abnormal behavior rule base, and the abnormal behavior similarity is calculated. So that the abnormal behavior similarity tags are obtained.

5 Detection of Network Crime

5.1 Design Principle of Network Crime Detection

The basic principle of network crime detection is shown in Fig. 2. When a network crime occurs, firstly, according to the category of the crime, use the historical abnormal behavior statistical tag and the abnormal behavior similarity tag to audit the users in the key user portrait database; If no suspicious users are found in the fast detection based on the user portrait, the abnormal behavior audit method based on the abnormal behavior rule base is used to audit all user logs, and the user portrait with abnormal behavior is constructed.

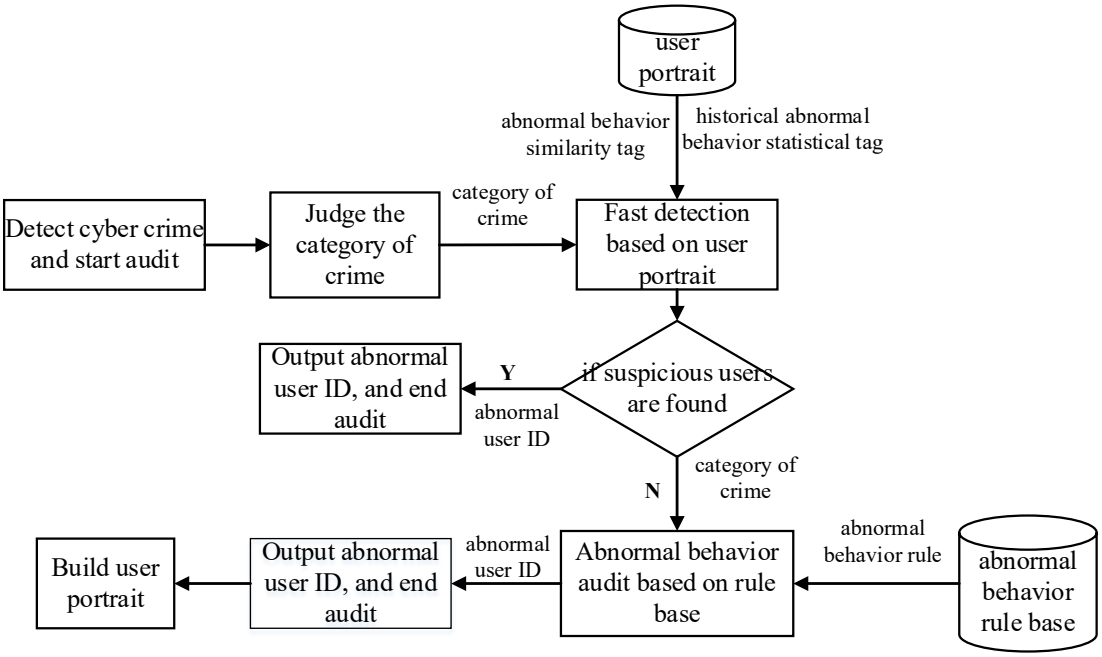

Figure 2: Detection process of network crime

5.2 Fast Detection Method Based on User Portrait

The fast detection process based on user portraits is as follows: Firstly, determine the correlation value with various abnormal behaviors according to the category of network crime. Secondly, the historical abnormal behavior statistics tags and abnormal behavior similarity tags of all key users in the user portrait database are obtained. Then, the Naive Bayes method is used to calculate the crime probability of each key user, and the top k users with the highest crime probability are found. Finally, judge whether there are suspicious users in the k users: if there are suspicious users, output the ID of the user, and the detection is over.

The calculation formula of crime probability is as follows:

$$P_{err} = \sum_{i=1}^{t} P_i * C_i \tag{5}$$

where, P_i is the abnormal behavior probability of the class i abnormal behavior, C_i is the correlation value between the current criminal behavior and the class i abnormal behavior, and t is the number of types of abnormal behavior.

The abnormal behavior probability is the probability of the abnormal behavior of the current user. The abnormal behavior probability P_i is determined by the abnormal behavior similarity S_i and the number of historical abnormal behaviors N_i. The value range of S_i is [0, 1], the value range of N_i is [0, + ∞] ($1 \le i \le m$, m is the number of users with user portrait). In order to eliminate the impact of different indicators due to different dimensions, we first normalize the indicators S_i and N_i: $Z_i = S_i / \sqrt{\sum (S_i)^2}$, $Y_i = N_i / \sqrt{\sum (N_i)^2}$. Then the linear weighting method is used to calculate the probability of abnormal behavior, and the calculation formula is as follows:

$$P_i = w_1 * Z_i + w_2 * Y_i \tag{6}$$

where, w_1 and w_2 are the weights of the two indicators, which are obtained by using the analytic hierarchy process. Here, due to the indicator Z_i is more important than the indicator Y_i, we can obtain the weight through qualitative and quantitative analysis: $w_1 = 0.875$, $w_2 = 0.125$.

The algorithm is as follows:

Algorithm 1. Fast detection algorithm based on user portrait

Input: number of users m, k, category of network crime, upper limit of normal behavior T_{normal}

Output: ID of abnormal users

1. According to the category of network crime, get the correlation value Ci between the category and various abnormal behaviors;
2. Get the number of historical abnormal behaviors Ni and the abnormal behavior similarity Si in the user portrait($1 \le i \le m$);
3. for(i=1;i≤m;j++)
4. Calculate the crime probability;
5. end for
6. Find the top k users with the highest crime probability;
7. for(j=1;j≤k;j++)
8. Get the user's criminal behavior probability P;
9. if(P＞T_{normal})
10. Output the ID of the user;
11. end for;

The complexity of Algorithm 1 mainly depends on the first for loop of 3-5 lines and the second for loop of 7-11 lines. The complexity is O (m) and O (k) respectively. So the complexity of algorithm 1 is O (m) + O (k).

5.3 Audit Method Based on Abnormal Behavior Rule Base

The audit process based on the abnormal behavior rule base is as follows: Match the current user's behavior sequence in the user's operation log with the most relevant abnormal behavior rule base, and use formula (2) to calculate the similarity between its behavior sequence and each rule. Then, select the maximum similarity value S_{max}, and judge whether the value is greater than the threshold value D_{normal} of the user's abnormal behavior similarity: If $S_{max} > D_{normal}$, the user's behavior is abnormal, the user's ID is recorded, and the user's portrait is constructed. Continue to audit the remaining users until all users have completed the audit.

The algorithm is as follows:

Algorithm 2. Audit algorithm based on abnormal behavior rule base

Input: the number of users n, category of network crime, the threshold of the user's abnormal behavior similarity D_{normal}

Output: ID of abnormal users

1. According to the category of network crime, the number t of rules in the corresponding abnormal behavior rule base is obtained;
2. for(i=1;i≤n;i++)
3. for(j=1;j≤t;j++)
4. Calculate the similarity S_{ij} between the behavior sequence of user i and rule j;
5. end for
6. Get the maximum SIM of S_{ij};
7. if(SIM>D_{normal})
8. Output the ID of the user;
9. Construct a user portrait for the user;
10. else
11. continue;
12. end for

The complexity of Algorithm 2 mainly depends on the double-layer for loop of 2–12 lines. The time complexity of the double-layer for loop is O (n * t). Since t represents the number of rules in the exception rule base, the amount of data is small, and n can be ignored relative to the number of users, so the complexity is: O (n * t) = O (n).

6 Experiment and Analysis

6.1 Data Sets and Evaluation Indicators

This paper uses the ADFA-LD data set for experiments. The ADFA-LD data set is a host-level intrusion detection data set released by the Australian Defense Academy. It is a data set containing system call syscall sequences of intrusion events. ADFA-LD data has characterized various types of system calls and marked the types of attacks. Tab. 2 shows the types of attacks.

Since the data set contains 6 types of attacks, this paper analyzes the 6 types of attacks to obtain 6 abnormal behavior rule bases, and gives the similarity between the user behavior sequence and the 6 abnormal behavior similarity tags of the user portrait.

To verify the effect of abnormal behavior audit, this paper uses accuracy, recall, precision, F value and audit time as evaluation indicators. TP indicates the number of abnormal records correctly classified, FN indicates the number of abnormal records classified into normal records, TN indicates the number of normal records classified correctly, and FP indicates the number of normal records classified into abnormal records.

Table 2: ADFA-LD experimental data set

Type of attack	Number	Marked type
Training	833	Normal
Validation	4372	Normal
Hydra-FTP	162	Attack

Hydra-SSH	148	Attack
Adduser	91	Attack
Java-Meterpreter	125	Attack
Meterpreter	75	Attack
Webshell	118	Attack

Definition 1. Accuracy represents the probability that all records will be accurately classified. The calculation formula is as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{7}$$

Definition 2. Recall (R) represents the probability of abnormal records being correctly classified. The calculation formula is as follows:

$$R = \frac{TP}{TP + FN} \tag{8}$$

Definition 3. Precision (P) represents the correct proportion of detected abnormal records. The calculation formula is as follows:

$$P = \frac{TP}{TP + FP} \tag{9}$$

Definition 4. F value is the weighted harmonic average of recall and precision. The calculation formula is as follows:

$$F = \frac{2 * P * R}{P + R} \tag{10}$$

6.2 Experimental Results and Analysis

6.2.1 Contrast Experiment of Abnormal Behavior Detection Effect

To verify the stability of the algorithm, the following experiments are carried out: 20%, 40%, 60%, 80% and 100% of the data in the data set are respectively selected for abnormal behavior detection experiments. The accuracy, recall, precision and F value are calculated. The results are shown in Fig. 3.

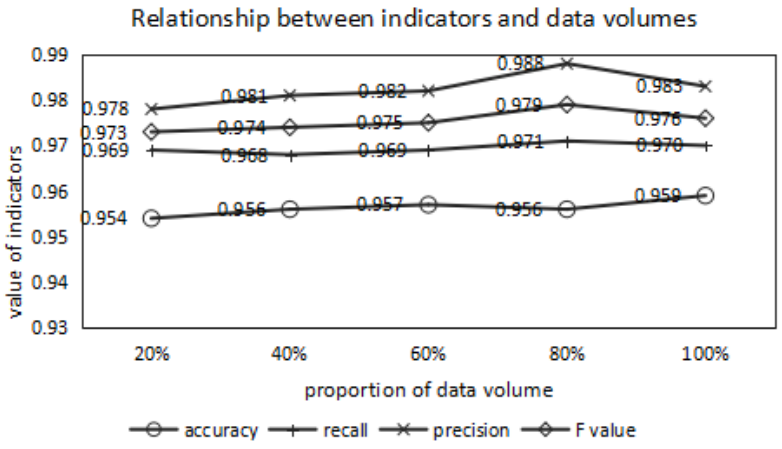

Figure 3: The values of the four indicators of the algorithm under different data volumes

It can be seen from Fig. 3 that with the change of data amount, the accuracy remains between 0.95–0.97, the recall between 0.96–0.98, the precision between 0.97–0.98, and the F value between 0.97–0.99, almost no change. The results show that the detection effect of this algorithm is basically stable.

The reasons are as follows: when constructing the abnormal behavior rule base, the algorithm uses all the abnormal records in the data set. Therefore, no matter how the amount of data changes, the detection effect basically remains unchanged.

References [19,20] both use user portrait to detect abnormal behavior, but both have certain shortcomings. Based on their research, this paper proposes a comprehensive user portrait which can be used to detect abnormal behavior. To verify the validity of the user portrait proposed in this paper, all the data in the ADFA-LD data set are used for experiments. In the same environment, the three methods perform abnormal behavior detection on all records in the ADFA-LD data set, and calculate their accuracy, recall, precision, and F value. The results are shown in Tab. 3.

Table 3: Comparison of abnormal behavior detection results of the three algorithms

	Accuracy	Recall	Precision	F value
References 19	0.865	0.906	0.938	0.943
References 20	0.868	0.871	0.975	0.92
The algorithm	0.959	0.970	0.983	0.976

The comparison results of abnormal behavior detection show that compared with the algorithms in [19] and [20], this algorithm has higher accuracy, recall, precision and F value.

The reasons are as follows: The algorithms in [19] and [20] both analyze normal behavior and construct rules for normal behavior. A successful match with the rules of normal behavior is normal behavior, otherwise it is abnormal behavior. The two methods do not analyze the rules of abnormal behavior, so the effect is not very good. And this algorithm focuses on the rules of abnormal behavior, which can not only correctly distinguish between normal and abnormal behavior, but also know exactly the type of abnormal behavior, so the effect is better.

6.2.2 Audit Efficiency Comparison Experiment

The audit algorithm proposed in this paper is divided into two parts: the fast detection based on user portrait of the first m users and the audit based on the abnormal behavior rule base. The conventional audit method will audit the behavior of all n users. Therefore, in the audit effect comparison experiment, this paper designs two groups of experiments: the first group is the abnormal behavior implementer in the first m users; the second group is the abnormal behavior implementer in the last n-m users. The audit algorithm proposed in this paper is compared with the conventional audit algorithm, and the audit time is compared.

The data set selected in this paper only contains the short-term behavior sequence of each user, it is not a continuous long-term sequence, so the abnormal behavior audit cannot be carried out. Therefore, this experiment adds a period of behavior sequence to each user randomly. In order to audit, only one user's behavior sequence is abnormal in the experiment. In this experiment, 1,000 users are selected for audit, including 60 users who had abnormal behavior. This experiment used two methods to audit all six kinds of abnormal behaviors. When there is implementer with abnormal behavior among the current k users, the experimental results are shown in Fig. 4.

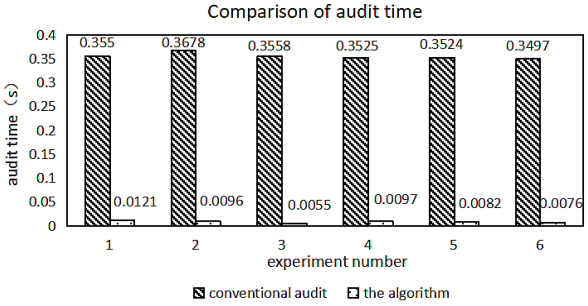

Figure 4: Comparison of audit time

It can be known from Fig. 4 that when there is implementer with abnormal behavior among the current k users, the audit time of the conventional audit method is about 0.35 s, and the audit time of this algorithm is about 0.01 s. It can be seen that the audit time of this algorithm is significantly better than the conventional algorithm.

The reasons are as follows: this algorithm gives priority to audit users with user portrait. The matching of user portraits reduces the audit scope and improves the audit speed. But conventional audit algorithms require auditing of all users, and the audit time is longer.

When there is implementer with abnormal behavior among the last n-k users, the experimental results are shown in Fig. 5.

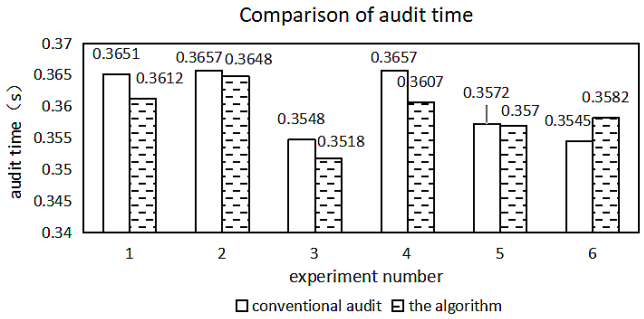

Figure 5: Comparison of audit time

It can be known from Fig. 5 that when there is implementer with abnormal behavior among the last n-k users, the audit time of the two algorithms is between 0.34–0.37 s, indicating that the two methods are equivalent.

The reasons are as follows: The implementer of the abnormal behavior is in the last n-k users, and both methods need to audit all users. So the audit time of the two algorithms is almost the same.

The experimental results of 5.2.1 and 5.2.2 show that the method proposed in this paper has high accuracy, recall, precision, F value and short audit time.

7 Conclusion

Based on the analysis of the existing methods, this paper studies the high incidence of network crime, applies the idea of repeated crime to the audit of abnormal behavior, and puts forward a method of network crime detection based on user portrait. This method constructs an abnormal behavior rule base for each abnormal behavior, and constructs a user portrait including basic attribute tag, behavior attribute tag, and abnormal behavior similarity tag for each user who has ever had abnormal behavior. When a network crime occurs, the key users are subject to fast detection based on user portrait firstly; if the fast detection does not achieve the expected effect, then the audit based on the abnormal behavior rule base is conducted to audit the behavior of all users until the implementer of the network crime is found.

Experimental results show that compared with the existing abnormal behavior detection methods, the algorithm has higher accuracy, recall, precision and F value; compared with the conventional audit method, the algorithm has a shorter audit time. Therefore, the detection method of network crime based on user portrait proposed in this paper has obviously improved the efficiency of network crime audit.

Funding Statement: This research is supported by The National Natural Science Foundation of China under Grant (No. 61672101), Beijing Key Laboratory of Internet Culture and Digital Dissemination Research (No. ICDDXN004) and Key Lab of Information Network Security of Ministry of Public Security (No. C18601).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] Y. Zheng, "Youth repeat crime rate is over 30%," [Online]. Available: http://news.sohu.com/20080123/n254827196.shtml. 2008.

[2] Y. Wei and Z. Hong, "Difficult integration into society, frequent repeated Crimes," [Online]. Available: http://epaper.southcn.com/nfdaily/html/2013-03/20/content_7174511.htm. 2013.

[3] P. Chen, Z. Zeng and X. Hu, "Predictive analysis and identification of habitual offender identity based on machine learning," *Journal of Criminal detection Police University of China*, vol. 145, no. 5, pp. 124–128, 2018.

[4] L. Zhang, H. Niu, Z. Wang and X. Liu, "Research on the construction of early warning model of criminals based on big data," *Netinfo Security*, vol. 220, no. 4, pp. 82–89, 2019.

[5] D. E. Denning, "An intrusion-detection model," *IEEE Transactions on Software Engineering*, vol. 13, no. 2, pp. 222–232, 1987.

[6] F. Jia and L. Kong, "Intrusion detection algorithm based on convolutional neural network," *Transaction of Beijing Institute of Technology*, vol. 37, no. 12, no. 1271–1275, 2017.

[7] S. Wei, H. Liu and Z. Zhao, "Research on intrusion detection based on DBN-ELM," *Computer Engineering*, vol. 44, no. 9, pp. 153–158, 2018.

[8] G. Creech and J. Hu, "A semantic approach to host-based intrusion detection systems using contiguous and discontinuous system call patterns," *IEEE Transactions on Computers*, vol. 63, no. 4, pp. 807–819, 2014.

[9] M. Xie, J. Hu, X. Yu and E. Chang, "Evaluating host-based anomaly detection systems: application of the frequency-based algorithms to ADFA-LD," in *Int. Conf. on Network and System Security*, pp. 1711–1716, 2014.

[10] W. Wang, X. Du, N. Wang and D. B. Shan, "Review on security audit technology for cloud computing," *Computer Science*, vol. 44, no. 7, pp. 16–20, 2017.

[11] M. Schonlau and M. Theus, "Detecting masquerades in intrusion detection based on unpopular commands," *Information Processing Letters*, vol. 76, no. 1, pp. 33–38, 2000.

[12] J. Duan, Y. Xin and Y. Ma, "Research and design of security audit log system based on web application," *Netinfo Security*, vol. 14, no. 10, pp. 76–82, 2014.

[13] X. Wang, J. Zhang, M. Wang, L. J. Zu, Z. H. Lu *et al.,* "CDCAS: a novel cloud data center security auditing system," in *IEEE Int. Conf. on Services Computing (SCC)*, pp. 605–612, 2014.

[14] C. Zhao, S. Tu, H. Chen and Y. Huang, "Log-based analysis on users' behavior in cloud security auditing," *Modern Electronics Technique*, vol. 40, no. 2, pp. 1–5, 2017.

[15] H. Tian, Z. Chen, C. Chang, Y. Huang, T. Wang *et al.,* "Public audit for operation behavior logs with error locating in cloud storage," *Soft Computing*, vol. 23, no. 12, pp. 1–14, 2018.

[16] A. Giuseppe and S. Umberto, "User profile modeling and applications to digital libraries," in *3rd European Conf. on Research and Advanced Technology for Digital Libraries*, pp. 184–197, 1999.

[17] W. Husain and L. Y. Dih, "A framework of a personalized location-based traveler recommendation system in mobile application," *International Journal of Multimedia & Ubiquitous Engineering*, vol. 7, no. 3, pp. 11–18, 2012.

[18] H. Zeng and S. Wu, "User image and precision marketing on account of big data in Weibo," *Modern Economic Information*, vol. 6, no. 16, pp. 306–308, 2016.

[19] J. Zhu, G. Chen, Y. Shi and Z. Xue, "Abnormal behavior detection based on user profile," *Communications Technology*, vol. 50, no. 10, pp. 180–185, 2017.

[20] G. Zhao and X. Yao, "Anomaly detection model based on user portrait," *Netinfo Security*, vol. 17, no. 7, pp. 18–24, 2017.

[21] W. Wang and Y. Xu, "Research on awareness method of cloud user abnormal behavior based on log audit," in *ICCC*, pp. 1945–1950, 2018.