Computer Real-Time Location Forensics Method for Network Intrusion Crimes

Yingsu Qi

(Corresponding author: Yingsu Qi)

Beijing Police College, Beijing, China

No. 11, Nanjian Road, Nankou Town, Changping district, Beijing 102202, China

(yingsq_ys@yeah.net)

(Received May 31, 2018; Revised and Accepted Oct. 13, 2018; First Online Mar. 6, 2019)

Abstract

In recent years, with the development of Internet technology, people's life has been more convenient with the help of the Internet. But it has also given rise to a new form of crime, Internet crimes. Computer forensics is thus born to deal with new kinds of crimes. In this study, the improved K-means clustering algorithm was adopted to obtain computer real-time evidence of network intrusion crimes. The detection performance of the improved algorithm for the four types of characteristic data was analyzed by MATLAB. Moreover, the detection performance of traditional clustering algorithm under different intrusion attack modes was compared. The results demonstrated that the improved algorithm is more suitable for detecting the first kind of characteristic data; and compared with the traditional clustering algorithm, under the three flood attack modes of User Datagram Protocol (UDP), Internet Control Messages Protocol (ICMP) and Transmission Control Protocol (TCP), the improved algorithm is better and has faster speed of data processing. In conclusion, the improved K-means clustering algorithm can be applied to the computer real-time location and forensics of network intrusion crimes.

Keywords: Clustering Algorithm; Computer Forensics; Network Intrusion; Network Security

1 Introduction

With the globalization of information, the Internet has developed rapidly. Enterprises, governments and individuals cooperate with each other on the Internet, which improves work efficiency as well as convenience and enriches people's spare time life [7]. The Internet has the advantages of openness and sharing. On the one hand, with these two advantages, the network has a variety of resources and facilitates people's life. On the other hand, since these two advantages are not only for the general public, cybercriminals can also take advantage of these two advantages to launch an invasion against enterprises, governments or individuals through the Internet.

In recent years, the popularity of mobile terminals that can be connected to the Internet has also diversified the ways of network crimes. Due to the convenience brought by the network, it brings greater security risks to users [6]. Therefore, in order to protect the network security of users and combat the illegal criminal acts with the help of the network, computer forensics technology emerges at the historic moment. Computer forensics technology [11] intercepts and analyzes network intrusion data to obtain relevant criminal evidence. The early computer forensics technology is usually used to reverse and analyze the intrusion scene after the event to obtain relevant evidence. However, with the rapid development of network technology, the corresponding traces of the perpetrators are often eliminated after the invasion, making it difficult to obtain effective evidence for the early static forensics [10]. Therefore, the real-time dynamic forensics technology that can collect effective evidence in large-scale networks is required.

Here are the relevant studies. Sun *et al.* [12] proposed a modeling method for Evolved Packet Core (EPC) network intrusion system in order to obtain appropriate defense strategies. Firstly, the attack behavior in the network is described, and the intrusion attack is introduced into the model based on time. Finally, UPPAAL is used to verify the correctness of the proposed model. Compared with other typical attack models, the proposed model is comprehensive and integrated. Malialis *et al.* [8] proposed a method combining task decomposition, team rewards and punishments and forms of rewards and punishments, which is called differential rewards and punishments system. The system learns quickly. The extensibility of the algorithm is demonstrated in experiments involving 1000 learning agents. Compared with baseline technology and throttling technology, this method is more effective. It can adapt to complex attack rates and dynamics, and it is robust to agent failure.

Hodo proposed the use of Artificial Neural Network (ANN) to deal with malicious attacks in the Internet [3].

The classification of normal mode and threat mode in network is studied. The experimental results show that this method has 99.4% accuracy and can detect all kinds of DDoS attacks successfully. In this paper, the improved k-means clustering algorithm is adopted to obtain computer real-time evidence of network intrusion crimes. The detection performance of the improved algorithm for the four types of characteristic data was analyzed by simulation on MATLAB. At the same time, the detection performance of traditional clustering algorithms under different intrusion attack modes was compared.

2 Computer Forensics Technology

With the rapid development of network technology, people's life style has changed dramatically. The emergence of the Internet not only provides convenience, but also provides a new criminal channel which is different from the traditional one. In view of the new network crime, the computer forensics technology arises at the historic moment. Computer forensics is also called network forensics. The evidence of network crime provided by it is recognized by law and can be used as an effective and reliable basis in judgment. The key principle of computer forensics [13] is to use the relevant network data analysis technology to judge the motive of the crime and to collect the network crime data of the suspect specifically, which is taken as the evidence of legal trial.

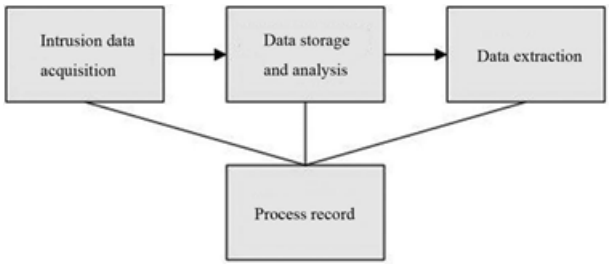

Figure 1: Technical process of computer forensics

As shown in Figure 1, the process of computer forensics technology is divided into three steps [4]. First, data for an intrusion is acquired. Then the collected data is preserved and analyzed. Finally, the evidence and data obtained from the analysis are extracted for a summary report. In the three processes of evidence collection, based on the continuity principle of effective evidence, it is necessary to record the three processes.

In terms of network intrusion data, network intrusion is usually an attack on a server with a large amount of garbage data, and the actual purpose of the operation is hidden in it for an invasion purpose. The common network attack mode is a flood attack that is simply rough and difficult to defend. The common flood attack modes are User Datagram Protocol (UDP) flood attack, Transmission Control Protocol (TCP) flood attack and Internet Control Messages Protocol (ICMP) flood attack. The above flood attacks are all DDoS flood attacks, which flood the server with a large amount of garbage data, causing the server to run out of resources and go down. The difference between the three DDoS flooding attacks mentioned above lies in the different message data used for the attack. UDP message mainly attacks target IP address. TCP mainly uses the TCP protocol to suspend a service. Information for requesting confirmation is continuously sent to consume server resource. And access to other users is blocked . ICMP is a traffic attack that consumes resources by sending more than 65,535 bytes of data to the server.

In order to ensure the effectiveness of computer forensics, the following three principles should be followed [2]:

1) The crime scene is made of a timely blockade to ensure the preservation of evidence maximization;

2) The collection of evidence network integrity and continuity must be guaranteed. The final evidence submitted shall be the same as the final evidence investigated;

3) In the whole process of evidence collection, there must be a third party to supervise so to ensure the impartiality of evidence.

3 Cluster Detection Algorithm Based on Network Packet Analysis

When the intrusion data is analyzed, the corresponding intrusion detection algorithm should be applied. In this paper, clustering detection algorithm based on network packet is selected. The working principle of clustering algorithms is simply to classify the collected data and distinguish normal data from abnormal data. In practical work, the collected data are first converted into vectors, and then the vector data are classified by clustering algorithms. There are different clustering algorithms for different data types. In this paper, k-means clustering algorithm [1] is selected and its principle is as follows:

$$\begin{aligned} d(a,b) &= d_{continuous}(a,b) + d_{discrete}(a,b) \\ a &= \{a_1, a_2, \cdots, a_i, a_{i+1}, a_{i+2}, \cdots, a_n\} \\ b &= \{b_1, b_2, \cdots, b_i, b_{i+1}, b_{i+2}, \cdots, b_n\} \end{aligned} \quad (1)$$

where $d(a,b)$ is the distance between eigenvectors a and b; $d_{continuous}(a,b)$ is the distance of continuous eigenvectors; $d_{discrete}(a,b)$ is the distance of discrete eigenvectors, and $a_1, a_2, \cdots, a_i$ and $b_{i+1}, b_{i+2}, \cdots, b_n$ are continuous eigenvectors; $a_i, a_{i+1}, a_{i+2}, \cdots, a_n$ and $b_1, b_2, \cdots, b_i$ are

discrete eigenvectors. Moreover,

$$\begin{aligned} d_{continuous}(a,b) &= \sqrt{\sum_{j=1}^{i}(a_j - b_j)^2} \\ d_{discrete}(a,b) &= \sum_{j=i+1}^{n} d(a_j, b_j) \\ d(a_i, b_i) &= 0 \text{ if } a_i = b_i \\ d(a_i, b_i) &= 1 \text{ if } a_i \neq b_i. \end{aligned} \quad (2)$$

Then according to the calculated distance between the two eigenvectors of a and b, the similarity of the two features is determined, and the classification is carried out accordingly.

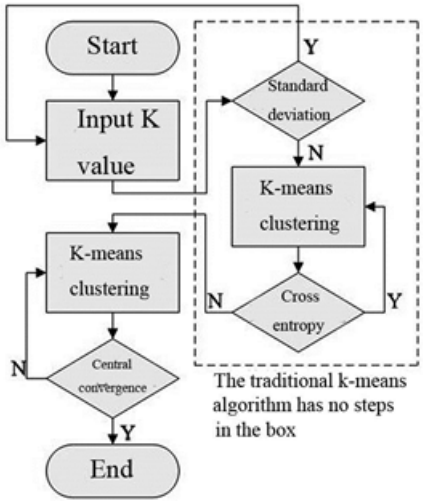

Figure 2: The process of the improved k-means algorithm

As shown in Figure 2, K represents the number of categories. In the traditional k-means algorithm process, there are no steps in the dotted box. The clustering effect of data depends on whether the choice of K value is reasonable. When the detected data is discrete, the defect exposure is more obvious. In this paper, the data of computer forensics for network intrusion is discrete. Therefore, in order to improve the accuracy and efficiency of forensics, combined with the theory of hierarchical clustering, the traditional algorithm is improved by increasing standard deviation and cross entropy [9]. The formula of its increase steps are:

Standard deviation is:

$$\sigma = \sqrt{\frac{d(a_i, C)}{m}} \quad (3)$$

where C is the threshold; M is the number of categories.

Cross entropy is:

$$D_R = -\log(\frac{1}{mn}\sum_{i=1}^{n}\sum_{j=1}^{m} e^{\frac{-||a-b||^2}{2\sigma^2}}), \quad (4)$$

where n is the number of another category.

The improved K-means algorithm is shown in Figure 2. First, K value and eigenvector data are input to calculate its center point. Then K is divided into the right number of species by the standard deviation. Next, it is aggregated by cross entropy. After completion, the classification is carried out through the k-means algorithm until the convergence of K value is stable, and the algorithm is completed.

4 Simulation Test

4.1 Data Preparation

KDD99 data set was used [14]. Data in the data set can be divided into four categories according to characteristics. The first category is a description of the characteristics of the server status. The second category is a description of the characteristics of the operations received by the server. The third category is a description of the characteristics of network traffic in the past 2 s. The fourth category is a description of the characteristics of network traffic over the first 100 connections. There are totally 40 features. Each data in the data set has 42 dimensions. The first 41 dimension is the characteristic attribute of the data, and the last one is the decision attribute, which indicates whether the data is abnormal and is used to detect the performance of the algorithm. The data set including data and normal data of the known network intrusion type is used to simulate the real network environment. In this paper, 100,000 pieces of data were selected in the data set, of which 30,000 pieces were normal data, and the rest were network intrusion data.

4.2 Data Preprocessing

If the weight of eigenvalues in the data is too large, it will affect the accuracy of the detection algorithm. Therefore, it is necessary to normalize the characteristic data. The formula [5] is:

$$x' = \frac{x - N_{mn}}{N_{max} - N_{min}}, \quad (5)$$

where x is the characteristic data that needs to be normalized; N_{min} is the minimum value in the characteristic data; and N_{max} is the maximum value in the characteristic data. When the maximum and minimum are equal, the normalized eigenvalue is denoted as 0.

4.3 Experimental Environment

In this study, the algorithm model is written by Matlab simulation platform, and the experiment is carried out on the laboratory server. The server is configured as: Windows 7 system, I7 processor, 16G memory.

4.4 Experimental Settings

1) The improved k-means algorithm is used to detect the data set where the initial K value is set as 6; the initial standard deviation is set as 0.6; the splitting parameter c_1 is set as 2.1, and the aggregation parameter c_2 is set as 0.4.

2) Packet uploads in the network are simulated as attacks. The traditional k-means algorithm and the improved k-means algorithm are used to detect the intrusion data, and each attack mode is simulated 4,000 times.

4.5 Evaluation Criteria

The performance evaluation of intrusion detection algorithm is usually represented by three data which are accuracy rate B_C, false alarm rate E_A and failure rate d. For the performance of intrusion detection algorithm, the higher the accuracy, the better; and the lower the false alarm rate and failure rate, the better. The calculation formula is as follows:

$$B_C = \frac{C_P + C_N}{C_P + C_N + M_P + M_N}, \quad (6)$$

$$E_A = \frac{M_N}{C_N + M_N}, \quad (7)$$

$$N_A = \frac{M_P}{C_P + M_P}, \quad (8)$$

where C_P is the attack data correctly classified; C_N is normal data with correct classification; M_N is the normal data with misclassification, and M_P is a misclassified attack data.

4.6 Experimental Results

4.7 Detection Performance of the Improved K-means Algorithm

As shown in Figure 3, the detection accuracy rate of the improved k-means algorithm for the intrusion of Data I is 97%. The false alarm rate is 11.1%, and the missing alarm rate is 5.6%. The detection accuracy rate of the intrusion of Data II is 94%. The false alarm rate is 12.3%, and the missing alarm rate is 6.2%. The detection accuracy rate of the intrusion of Data III is 93.1%. The false alarm rate is 13.5%, and the missing alarm rate is 7.1%. The detection accuracy rate of the intrusion of Data IV is 89%. The false alarm rate is 13.9%, and the missing alarm rate is 8.2%. It can be intuitively seen from the Figure 3 where the improved K-means algorithm has the highest detection accuracy and the lowest false alarm rate and omission rate on Data I. Compared with the other three types of data, the improved K-means algorithm is more suitable for detecting the network intrusion attacks of Data I.

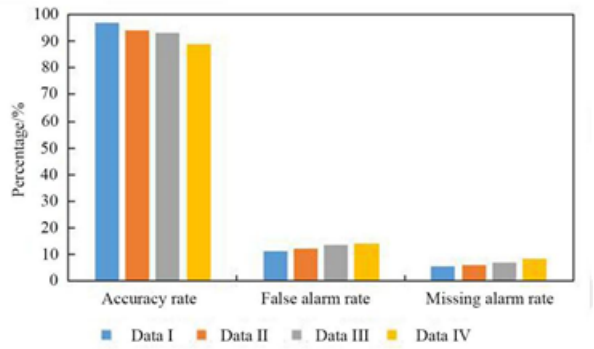

Figure 3: Intrusion detection results of the improved K-means algorithm

4.8 Performance Comparison with Traditional K-means Algorithm

As shown in Figure 4, in the UDP simulated flood attack mode of Data I, the detection accuracy rate of the traditional K-means algorithm is 91.4%. The false alarm rate is 9.8%, and the missing alarm rate is 7.8%. The detection accuracy rate of the improved k-means algorithm is 93.2%. The false alarm rate is 4.5%, and the missing alarm rate is 3.1%.

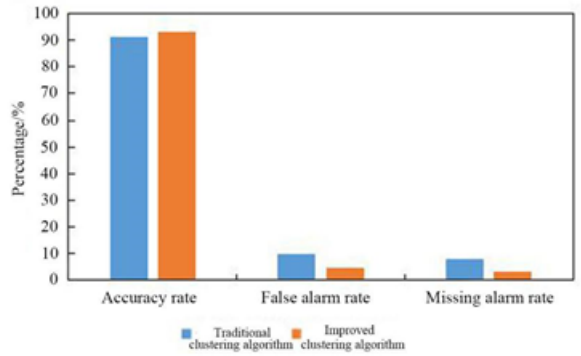

Figure 4: Detection performance of the two algorithms in UDP flood attack mode

As shown in Figure 5, in the ICMP simulated flood attack mode of Data I, the detection accuracy rate of the traditional K-means algorithm is 91.3%. The false alarm rate is 9.6%, and the missing alarm rate is 7.5%. The detection accuracy rate of the improved k-means algorithm is 93.5%. False alarm rate is 4.6%, and missing rate is 2.9%. It can be seen that the improved clustering algorithm has better detection performance for ICMP flood attack mode.

As shown in Figure 6, in the TCP simulated flood attack mode of Data I, the detection accuracy rate of the traditional K-means algorithm is 91.2%. The false alarm rate is 9.9%, and the missing alarm rate is 7.9%. The detection accuracy rate of the improved k-means algorithm

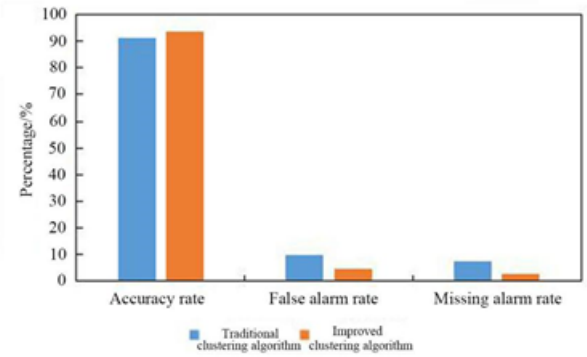

Figure 5: Detection performance of the two algorithms in ICMP flood attack mode

is 94.1%. The false alarm rate is 4.8%, and the missing alarm rate is 3.3%. It can be seen that the improved clustering algorithm has better detection performance for TCP flood attack mode.

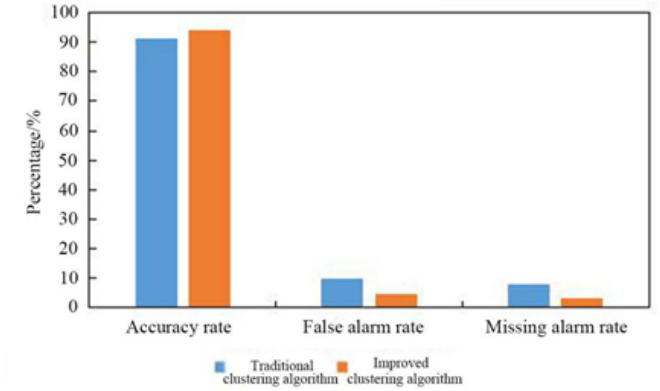

Figure 6: Detection performance of the two algorithms in TCP flood attack mode

As shown in Table 1, the processing speed of traditional clustering algorithm for UDP flood attack is 331.5 piece/s. The processing speed of ICMP flood attack is 321.6 piece/s. The processing speed of TCP flood attack is 335.4 piece/s. The processing speed of the improved clustering algorithm for UDP flood attack is 523.6 piece/s. The processing speed of ICMP flood attack is 528.4 piece/s. The processing speed of TCP flood attack is 531.3 piece/s. It can be seen from the statistical results that the improved clustering algorithm is significantly faster than the traditional clustering algorithm in processing the intrusion data. Moreover, according to the comparison of the amount and time of the evidence of the intrusion data obtained by computer forensics in the past network intrusion crimes, it can be seen that the processing performance of the improved clustering algorithm for the intrusion data exceeds the actual performance requirements, and the algorithm has been able to be used for the real-time forensics of network intrusion crimes.

5 Conclusion

In this study, the improved K-means clustering algorithm was adopted to obtain computer real-time evidence of network intrusion crimes. In addition, the detection performance of the improved algorithm for the four types of characteristic data was simulated and analyzed in MATLAB. Moreover, the detection performance of traditional clustering algorithm under different intrusion attack modes was compared. The performance of the improved k-means clustering algorithm in detecting the first type of feature data is better than the other three. Under the flood attack modes of UDP, ICMP and TCP, the improved k-means clustering algorithm is superior to the traditional clustering algorithm in the detection performance of intrusion data. The two algorithms were compared in the processing speed of the intrusion data under the three intrusion attack modes, and the improved algorithm obviously exceeds the traditional clustering algorithm.

References

[1] L. Duan, F. Yu, L. Zhan, "Improved fuzzy c-means clustering algorithm," in *International Conference on Natural Computation, fuzzy Systems and Knowledge Discovery*, IEEE, pp. 44–46, 2016.

[2] D. Gugelmann, F. Gasser, B. Ager, *et al.*, "Hviz: HTTP(S) traffic aggregation and visualization for network forensics," *Digital Investigation*, vol. 12, pp. s1-s11, 2015.

[3] E. Hodo, X. Bellekens, A. Hamilton, *et al.*, "Threat analysis of IoT networks using artificial neural network intrusion detection system," in *International Symposium on networks*, IEEE, pp. 6865–6867, 2016.

[4] F. Karpisek, I. Baggili, F. Breitinger, "WhatsApp network forensics: Decrypting and understanding the WhatsApp call signaling messages," *Digital Investigation*, vol. 15, pp. 110-118, 2015.

[5] N. Khamphakdee, N. Benjamas, S. Saiyod, "Improving intrusion detection system based on snort rules for network probe attacks detection with association rules technique of data mining," *Journal of ICT Research & Applications*, vol. 8, no. 3, pp. 234–250, 2015.

[6] S. Khan, A. Gani, A. W. Wahab, *et al.*, "Network forensics: Review, taxonomy, and open challenges," *Journal of Network & Computer Applications*, vol. 66, pp. 214–235, 2016.

[7] M. Korczynski, A. Hamieh, J. H. Huh, *et al.*, "Hive oversight for network intrusion early warning using DIAMoND: A bee-inspired method for fully distributed cyber defense," *IEEE Communications Magazine*, vol. 54, no. 46, pp. 60–67, 2016.

[8] K. Malialis, S. Devlin, D. Kudenko, "Distributed reinforcement learning for adaptive and robust network intrusion response," *Connection Science*, vol. 27, no. 3, pp. 19, 2015.

Table 1: Processing speed of the two algorithms in different attack modes

Detection algorithm	Intrusion attack mode	Number of invasions	Total detection time/s	Processing speed (piece/s)
Traditional clustering algorithm	UDP flood attack	5216	15.73	331.5
	ICMP flood attack	4215	13.11	321.6
	TCP flood attack	4856	14.48	335.4
Improved clustering algorithm	UDP flood attack	5745	10.97	523.6
	ICMP flood attack	4351	8.23	528.4
	TCP flood attack	4951	9.32	531.3

[9] P. Nayak, A. Devulapalli, "A fuzzy logic-based clustering algorithm for WSN to extend the network lifetime," *IEEE Sensors Journal*, vol. 16, no. 1, pp. 137–144, 2015.

[10] A. Singhal, C. Liu, D. Wijesekara, "POSTER: A logic based network forensics model for evidence analysis," in *ACM Sigsac Conference on Computer and Communications Security*, pp. 1677–1677, 2015.

[11] G. Singhchhabra, P. Singh, "Distributed network forensics framework: A systematic review," *International Journal of Computer Applications*, vol. 119, no. 19, pp. 31–35, 2015.

[12] Y. Sun, T. Y. Wu, X. Q. Ma, H. C. Chao, "Modeling and verifying the EPC network intrusion system - based on timed automata," *Journal of Pervasive and Mobile Computing*, vol. 24, pp. S1574119215001145, 2015.

[13] M. Vallentin, R. Sommer, R. Sommer, "VAST: A unified platform for interactive network forensics," in *Usenix Conference on Networked Systems Design and Implementation*, pp. 345–362, 2016.

[14] T. Yoshioka, S. Karita, T. Nakatani, "Far-field researched and recognition using CNN-within DNN-HMM with convolution in time," in *IEEE International Conference on Acoustics, Researched and Signal Processing*, pp. 4360–4364, 2015.

Biography

Yingsu Qi, female, born in July 1979, is a master of engineering and lecturer. Her research interests include computer forensics and information security. She has presided over the project of Beijing Education Committee's Talent Program and published articles including Computer Forensics Technology, Thoughts on Network Quality Training from the Perspective of Personal Information Security and Research on Enterprise Information Sharing Platform Based on Real-time Database and XML and participated in the compilation of a ministerial textbook Handbook of Public Security Information Application Law.