Hindawi
Security and Communication Networks
Volume 2023, Article ID 9869160, 1 page
https://doi.org/10.1155/2023/9869160

Retraction

Retracted: Anomaly Intrusion Detection of Wireless Communication Network-Based on Markov Chain Model

Security and Communication Networks

Received 5 December 2023; Accepted 5 December 2023; Published 6 December 2023

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] H. Zhang, W. Lan, and D. Zhang, "Anomaly Intrusion Detection of Wireless Communication Network-Based on Markov Chain Model," *Security and Communication Networks*, vol. 2022, Article ID 3255006, 11 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 3255006, 11 pages
https://doi.org/10.1155/2022/3255006

Research Article

Anomaly Intrusion Detection of Wireless Communication Network-Based on Markov Chain Model

Huifang Zhang,[1] **Wangsen Lan**,[1] **and Desheng Zhang**[2]

[1]*Department of Mathematics, Xinzhou Teachers University, Xinzhou, Shanxi 034000, China*
[2]*School of Science, Xi'an University of Technology, Xi'an, Shanxi 710054, China*

Correspondence should be addressed to Huifang Zhang; zhf150481@xztu.edu.cn

Received 30 March 2022; Revised 30 May 2022; Accepted 4 June 2022; Published 5 July 2022

Academic Editor: Mohammad Ayoub Khan

In order to solve the increasingly serious security problems of wireless networks, research on abnormal intrusion detection methods of wireless communication networks based on Markov chain model is proposed. What is usually observed is not the known intrusion behavior but the abnormal phenomenon in the communication process studied, which is completed by detecting the change of system behavior or usage. In this paper, the Markov chain model is used to detect the abnormal intrusion of wireless communication networks. Through the analysis and selection of parameters, the experimental results are ideal, and a variety of judgment methods are compared and analyzed. First, this method can easily distinguish between normal and abnormal data, which reduces the time by about 50% compared with the previous method; Second, the detection result of analysis method 2 is better than that of analysis method 1, and the accuracy is about 20%. The new method proposed in this paper has the characteristics of simple calculation, low algorithm complexity, and easy online detection. This method overcomes the disadvantage that the single-step Markov chain analysis and detection method cannot be strictly established in the nature of the Markov chain, has lower algorithm complexity than the multistep Markov chain analysis and detection method, and is simpler than the parameter calculation of hidden Markov chain model.

1. Introduction

In recent years, the popularity of e-commerce and e-government applications has further improved the status of the network in the national economy and people's livelihood. The network has been directly related to the operation and development of modern society [1]. In addition, the security problems in the network are endangering various applications of the network, which seriously affects the further development of the network [2]. With the construction of network infrastructure and the proliferation of Internet users, the problems of network and information security are becoming more and more serious, and the losses caused by criminals invading public networks are becoming more and more huge [3]. Mainly because

(1) Computer viruses emerge one after another, wreak havoc all over the world, and gradually show new transmission trends and characteristics. Its main performance is that it spreads rapidly, and more and more "hybrid viruses" and "mutated viruses" are formed by combining with hacker technology. This kind of virus can replicate itself and enhance the ability of active attack and active infection [4].

(2) The momentum of malicious attacks by criminals on the global network is increasing year by year. Statistics from the US computer emergency response team (Cert) show that cyberattacks show a trend of combining the destructive behavior of criminals with the spread of viruses. This means that network security has encountered new challenges, that is, the threat of viruses, Trojans, mites, and network attacks, which may cause rapid and large-scale infection, paralyze the host or server, and lose data and information, and the loss is immeasurable [5].

(3) Due to the incompleteness of technology and design, the system has defects or security vulnerabilities.

These vulnerabilities or defects mainly exist in computer operating systems and network software.

(4) Military forces around the world are stepping up their research on information warfare. In recent years, the attack means against the network is no longer limited to the usual tricks, such as modifying and deleting data, but grandly stepped on the stage of information warfare and became a means of information warfare [6].

At present, the common forms of network attacks include password theft, spoofing attack, defect and backdoor attack, authentication failure, protocol defect, information disclosure, exponential attack, virus and worm, denial of service attack, and so on. Therefore, in order to protect the security and reliability of the information, in addition to using legal and management means, it also needs to rely on technical methods. Traditional security technologies are divided into two categories: static security technology and dynamic security technology. Static security technology refers to the use of some peripheral devices through manual methods, which are mainly used to protect the system from external attacks. Its representative product is our common firewall. At present, many popular security products in the market belong to the category of static security technology. These static security devices can only provide certain defense capabilities against attacks from external systems. Once external intruders enter the system, cheat the authentication system, and become internal members, the static security technology will lose its defense capability. The disadvantage of static security technology is that it requires a lot of manual operation to implement and maintain, and the system administrator needs to have high network security technology ability and cannot actively track network intrusion [7, 8]. A tool capable of detecting network data of predetermined security violations; Tools that can detect network viruses; It is a tool that can automatically report or alarm when safety problems are found. Specifically, dynamic security technology includes network intrusion, security vulnerability scanning, dynamic response, and audit analysis. The biggest advantage of dynamic security technology is "initiative." By combining real-time data capture, real-time data analysis, and network monitoring system, according to the data in the specific security database, through analysis, we can quickly find the characteristics of dangerous attacks and then give an alarm. At the same time, we also provide certain protective measures [9, 10]. Intrusion detection technology is a very important dynamic security technology. If it is used together with the traditional static security technology, it will greatly improve the security protection level of the system. Intrusion detection refers to the process of identifying and dealing with the malicious use of computer and network resources. It can not only detect external attacks but also monitor the unauthorized behavior of internal users. Intrusion detection system (IDS) is a preventive measure of the security system. As the name implies, it attempts to detect, identify, and isolate intrusion attempts or unauthorized use of computers [11]. With the continuous discovery of system vulnerabilities and attacks, the status of the intrusion detection system in the whole security system continues to improve, and its role is becoming greater and greater [12]. Figure 1 shows the research on the design method of intrusion detection system for the wireless network.

2. Literature Review

IDS is a new generation of security products developed in recent years. It attempts to detect, identify, and isolate intrusion attempts or unauthorized use of computers. It can not only monitor the online access activities but also alarm for the ongoing attacks and even take corresponding measures, such as blocking or shutting down the equipment. Different from other security products, intrusion detection system needs more intelligence. It must analyze the obtained data and draw useful results [13–15]. As a security technology, the main purpose of an intrusion detection system is to: first, identify intruders; second, identify intrusion behavior; third, detect and monitor successful security breakthroughs; fourth, provide important information in time to resist invasion and prevent the occurrence and expansion of events [16–18]. Intrusion detection system needs to judge the current state of the network or host it monitors, which is not speculation. It needs to make a judgment based on the information contained in the original data. According to the source of original data, the intrusion detection system can be divided into host-based intrusion detection system (HIDS), network-based intrusion detection system (NIDS), and application-based intrusion detection system.

2.1. Host-Based Intrusion Detection System. A HIDS is a kind of detector system which is loaded on various servers of the organization and controlled by the central manager. The detector can find various events, take actions on a specific server, or send notification information. HIDS detector can also judge whether an attack is successful because the attack occurs on the platform where the detector is located. Different types of HIDS detectors can achieve different types of IDS targets. Not all detectors are applicable to various types of institutions or even cannot be applicable to all servers in the same institution. Therefore, the most suitable detector room must be confirmed for each server [19]. Although HIDS is not as fast as NIDS, it does have incomparable advantages. These advantages include being able to determine whether the attack is successful, finer monitoring granularity, flexible configuration, an environment that can be used for encryption and exchange, insensitive to network traffic, no need for additional hardware, etc. The disadvantages of HIDS are the cost of a HIDS system may be higher than NIDS because each server must have a detector license (the price of a detection unit is low, but a large number of detectors will lead to a high overall cost of the system). The processor capacity on the server and the detector process running on the server will occupy 5%–10% of the CPU capacity. If the detector is located on a system with a large load, this will affect the verification performance, and

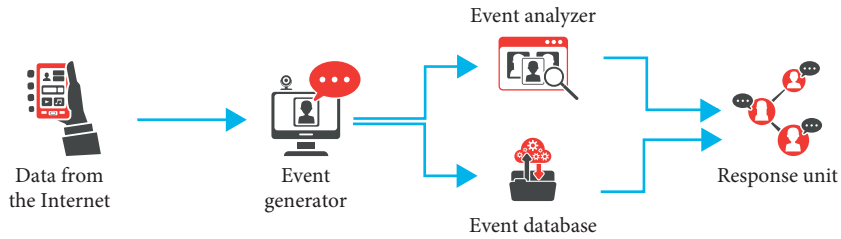

Figure 1: Research on design method of intrusion detection system for the wireless network.

it is necessary to purchase a terminal equipment room with higher performance [20].

2.2. Network-Based Intrusion Detection System. A NIDS exists as a software process on a special hardware system [21]. NIDS has the advantages of fast detection speed, good concealment, wider field of vision, fewer detectors, it is difficult for attackers to transfer evidence, and it is independent of the operating system. It can be configured on a special machine and will not occupy any resources on the protected device. Disadvantages include the following: if the communication data matches the preconfigured rules or features, the NIDS system can only alarm; NIDS may miss useful communication data due to high bandwidth occupation or route change; NIDS cannot determine whether the attack is successful; NIDS cannot view the encrypted communication data; Switched networks (as opposed to shared media networks) require special configuration to see all communication data.

2.3. Application-Based Intrusion Detection System. The application-based intrusion detection system can be said to be a special subset of HIDS or further refinement of the implementation of HIDS, so its characteristics, advantages, and disadvantages are basically the same as those of HIDS. The main feature is the use of monitoring sensors to collect information at the application layer. Because this technology can more accurately monitor the behavior of a user's application, this technology has attracted more and more attention in the increasingly popular e-commerce. It monitors the activities in a software application, and the information source is mainly the application log. Its monitoring content is more specific, and the monitoring object is narrower accordingly. These three intrusion detection systems are complementary. Network-based intrusion detection can objectively reflect network activities; it especially can monitor the blind area of system audit. Host-based and application-based intrusion detection can more accurately monitor various activities in the system. Most of the actual systems are a mixture of these three systems [22].

IDS usually uses two basic detection principles (analysis methods) to analyze events and detect intrusion behaviors, namely misuse detection and anomaly detection. The goal of misuse detection is to find known intrusion patterns where the analysis method adopted by most commercial IDs. The anomaly analysis method attempts to detect the abnormal pattern of system behavior, which is less applied in practical IDs. There are two most effective methods to detect misuse of IDS.

2.4. Misuse Detection. Misuse detection is to model abnormal behaviors which are previously recorded and confirmed as misuse or attack. The misuse detector analyzes the activities of the system and finds those events or event sets that match the predefined attack characteristics. Because the pattern corresponding to the attack is a feature, misuse detection is often also called feature-based detection. The current approach is to set the pattern of each attack event as an independent feature, so the detection of intrusion behavior becomes the matching search of features. If it matches the known intrusion features, it is considered an attack. If it does not match, it is not considered an attack. Of course, the actual situation is much more complicated.

2.5. Anomaly Detection. Anomaly detection is to model normal behavior, and all events that do not conform to this model are suspected to be attacks. Anomaly detection first collects the historical data of normal operation activities for a period of time and establishes the normal behavior profile representing users, hosts, or network connections. Event information is then collected, and various methods are used to determine whether the activity of the detected event deviates from the normal behavioral pattern. These methods mainly include threshold detection, statistical methods, and some new technologies, such as neural networks [23].

The intrusion detection methods currently in use are shown in Figure 2.

Some frame vulnerabilities in MAC layer directly or indirectly threaten the wide application of wireless networks. However, the traditional IDS for wired networks uses the upper layer information, such as the network layer, transmission layer, and application layer, for detection, but the intrusion detection system specially designed for the wired network cannot analyze and detect the MAC layer information of the wireless network. Therefore, wired network intrusion detection cannot detect specific attacks in wireless networks [24]. Therefore, the research on wireless network intrusion detection systems lags behind. Some existing wireless network commercial software can only detect some specific types of attacks. The traditional pattern matching wireless network intrusion detection is difficult to adapt to the new wireless network intrusion detection, and there are great defects for the future distributed and orderly cooperative attacks [25]. While the wireless network monitor

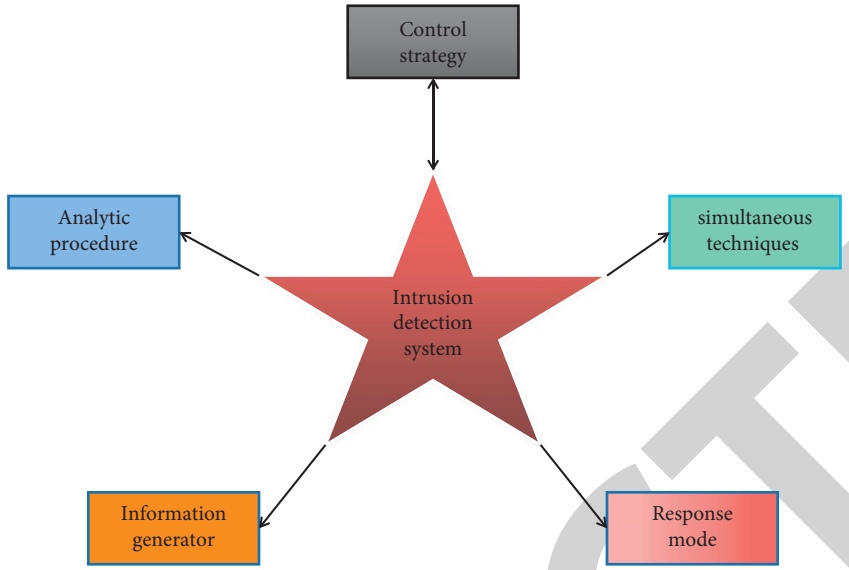

Figure 2: Classification of intrusion detection.

provides fixed monitoring capability, it also needs to realize the intrusion detection function for different types of wireless networks.

3. Research Methods

3.1. Overview of Markov chain. Markov chain is a special case of the Markov random process. It is a Markov process with discrete state and time parameters.

Suppose that the parameter set t of the Markov process $\{X, n \in T\}$ is a discrete-time set, $T = 0.1,2,\ldots,\}$. The corresponding X_n, the state space composed of all possible values, is the discrete state set $I = i_1, i_2, i_3,\ldots,\}$. Relevant definitions are as follows.

Definition 1. There is a random process $\{X_n, n \in T\}$, if for any integer $n \in T$ and any $i_0, i_1,\ldots,i_{n+1} \in I$, the conditional probability satisfies the following equation:

$$P\{X_{n+1} = i_{n+1} | X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n\} = P\{X_{n+1} = i_{n+1} | X_n = i_n\}. \quad (1)$$

Then $\{X_n, n \in T\}$ is called the Markov chain.

Definition 2. The intuitive meaning of conditional probability $P\{X_{n+1} = j | X_n = i\}$ is the probability that the system is in state I at time n and in state j at time $n + 1$. Note that the conditional probability is $p_{ij}(n)$, as shown in the following formula:

$$p_{ij}(n) = P\{X_{n+1} = j | X_n = i\}. \quad (2)$$

It is called the one-step transition probability of the Markov chain $\{X_n, n \in T\}$ at time n, where $i, j \in I$, is simply referred to as the transition probability.

Generally, the transition probability $p_{ij}(n)$ is related not only to states i and j, but also to time n. When $p_{ij}(n)$ does not depend on time n, it means that the Markov chain has a stationary transition probability.

Definition 3. If the transition probability $p_{ij}(n)$ of the Markov chain $\{X_n, n \in T\}$ is independent of n for any i,j ∈ I, then the Markov chain is said to be homogeneous, and $p_{ij}(n)$ is recorded as p_{ij}.

Let P represent the matrix composed of one-step transition probability p_{ij}, and the state space $I = \{1, 2, \ldots,\}$ Then, as shown in the following formula:

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} & \cdots \\ p_{21} & p_{22} & \cdots & p_{2n} & \cdots \\ \vdots & \vdots & & \vdots & \end{bmatrix}. \quad (3)$$

It is called the one-step transition probability matrix of the system state, which has properties as follows:

$$p_{ij} \geq 0, i, j \in I, \quad (4)$$

$$\sum_{j \in I} p_{ij} = 1, i \in I. \quad (5)$$

In equation (5), the summation of j is the summation of all possible states of state space I. This property shows that the sum of any element in the one-step transition probability matrix is 1.

Definition 4. The conditional probability of weighing is shown as follows:

$$p_{ij}^{(n)} = P\{X_{m+n} = j \mid X_m = i\} (i, j \in I, m \geq 0, n \geq 1). \quad (6)$$

Equation (6) is the n-step transition probability of the Markov chain $\{X_n, n \in T\}$, and P $(n) = (p_{ij}^{(n)})$ is the n-step transition matrix of the Markov chain, where $p_{ij}^{(n)} \geq 0, \sum_{j \in I} p_{ij}^{(n)} = 1$, P (n) is also a random matrix.

Definition 5. Let $\{X, n \in T\}$ be a Markov chain and $P_j = P\{X0, = j\}$ be the initial probability of $\{X_n, n \in T\}$. Let $\{p_j, j \in I\}$ be the initial distribution of $\{X_n, n \in T\}$, Pt (0) = (P_P, P2,. . .) is the initial probability vector.

Theorem 1. *Let $\{X_n, n \in T\}$ be a Markov chain, then for any integer $n \geq 0$, the n-step transfer matrix P (n) has the following properties:*

$$(1)\ P^{(n)} = PP^{(n-1)}, \tag{7}$$

$$(2)\ P^{(n)} = P^n. \tag{8}$$

3.2. Application of Markov chain in Intrusion Detection. The intrusion is detected by the method of real-time detection and analysis of the system call of the privileged process in the monitoring system. The reason is that the process behavior is described by the statistical characteristics of the system call sequence issued by it. The statistical characteristics of the system call sequence corresponding to normal behavior and abnormal behavior are different. If the statistical characteristics of the sequence sent by a process are different from those of normal behavior, it can be determined that the process has security problems. In addition, an observable short system call sequence is established to distinguish normal behavior and abnormal behavior. First, a table of all short normal system call sequences is established in the system, and then the system call sequences of the monitored process are matched with each short sequence in the table. If the matching proportion is large, the process can be considered normal behavior; otherwise, it is considered abnormal behavior. If the normal system call sequence is as follows: open, read, mmap, mmap, open, getrlimit, mmap, close, set the length of the short sequence to 3, as shown in Table 1:

If the observed system call sequence is as follows: open, read, mmap, open, open, getrlimit, mmap, close, by looking up the table, we can know that there are 4 mismatches in this system call sequence. Mismatch ratio V is the percentage of the number of mismatches in the total, as as follows:

$$V = \frac{m}{k(L-(k+1)/2)}, \tag{9}$$

where L is the length of the observed system call sequence, k is the number of short sequence system calls contained in the selected sliding window, and M is the number of mismatches. For example, in the above example, $L = 8$, $k = 3$, $m = 4$, then V is 22%. Then you can judge whether the process is normal by comparing it with a preset threshold. This method has a small amount of calculation and high detection accuracy. However, this method has many disadvantages: (1) It only records the system call sequence of normal behavior in the normal behavior database. As long as it appears in the system call sequence of the legal process, it will be added to the table, and it is not related to the statistical characteristics of these system call sequences, which has a certain impact on the accuracy of detection; (2)

Table 1: Normal system call sequence.

Call	Position 1	Position 2	Position 3
Open	Read	Mmap	Mmap
	Getrlimit		Close
Read	Mmap	Mmap	Open
Mmap	Mmap	Open	Getrlimit
	Open	Getrlimit	Mmap
Getrlimit	Close		
Close	Mmap	Close	

Establishing a list of short system call sequences with normal behavior in the system and querying this table will occupy more system resources; (3) Intruders may avoid detection by inserting some irrelevant system calls into their process.

3.3. Improvement of Anomaly Intrusion Detection Method Based on Markov chain Model. Shell is an important layer in the UNIX system. It is the interface between the user and the system. The most common use of a shell is to appear as a command interpreter; that is, it receives the commands entered by the user, analyzes them, creates subroutines, realizes the functions specified by the command by the subprocess, and sends out prompt symbols after the subprocess terminates its work. In addition, the shell is also a high-level programming language, which can write programs with strong functions but simple coding. In particular, it organically combines relevant UNIX commands, which can greatly improve the efficiency of programming and make full use of the open performance of the UNIX system to design commands suitable for their own requirements. Shell commands and system calls have some commonalities in data form, so the Markov chain model is improved, and the system call sequence of privileged process is used as training and test data. The specific implementation steps are as follows:

(1) Firstly, the training data of normal behavior of intrusion detection system is obtained. As the training data of normal behavior, it should contain almost all relevant system calls. This paper uses the system call sequence of sendmail daemon as the training data of normal behavior used to establish the model. Secondly, the one-step state transition probability matrix P of sendmail daemon is calculated by statistical method, and the element $P_{ij} = N_{ij}/N_i$, where N_{ij} represents the number of adjacent occurrences of system call i and system call j, and Ni represents the number of occurrences of system call i. Finally, calculate the initial state probability vector of sendmail daemon, as follows:

$$PT(0) = (p1, p2, p3, \ldots, pn). \tag{10}$$

(i) Pi = Ni/N, 1<puncsp> </puncsp>≤<puncsp> </puncsp> $i \leq n$, where Ni represents the number of system calls i, and N represents the total number of system calls.

(2) Analysis and detection stage

Firstly, the system call sequence of normal and abnormal behavior of the intrusion detection system in the monitored time is obtained as the test data. Then, the sliding window sequence analysis method is used to segment the tested system call sequence. For each short sequence (X1, X2, X3,. . ., Xi) after segmentation, two analysis and detection methods of system call sequence using a one-step Markov chain, namely:

$$p(X_1, X_2, \ldots, X_i) = p_{X_1} \prod_{T=2}^{t} p_{X_{T-i} X_T}. \quad (11)$$

$$p(X_1, X_2, \ldots, X_i) = p_{X_1} \times \sqrt[i-1]{p_{X_1 X_2} \times \ldots \times p_{X_{i-1} X_i}}. \quad (12)$$

The probabilities are calculated respectively to produce two probability sequences (q1, q2,. . ., qr) with length r (if i is different, the r of the two probability sequences is also different). At the same time, the system call of the tested system call sequence is matched with the system call of the normal system call sequence used to establish the model, and then the unknown system call is detected and reported to the system.

Secondly, because the user's behavior may deviate from its historical behavior in a short time, we use the sliding window sequence analysis method again for the above two probability sequences (q1, q2,. . ., qr). The judgment value HK is calculated for each short probability sequence generated by sliding window segmentation. The calculation formula is as follows:

$$H_k = -\frac{1}{w} \sum_{i=k}^{w+k-1} \lg[P(q_i)]. \quad (13)$$

where $1 \le k \le$ r−w+1,w is the length of the sliding window.

Assuming that the system call sequence is (4, 2, 66, 66, 4, 138) and the sliding window length is 4, the state and transition diagram of the Markov chain model is shown in Figure 3:

3.4. Decision Stage. In the judgment stage, this paper will use two methods to judge the detection results, that is, to distinguish between normal behavior and abnormal behavior:

First, draw two kinds of decision value curves: normal test data and abnormal test data. If there is a big difference between the decision value curves of normal and abnormal test data, it shows that the improvement of this method is effective in distinguishing between normal system calls and abnormal system calls and comparing the detection results under the two analysis methods. Secondly, the traditional anomaly detection method is used to select the appropriate threshold. In order to detect normal and abnormal system calls, by analyzing the judgment value curves of normal test data and abnormal test data, after selecting different thresholds, define the proportion of abnormal probability sequence and compare the detection results under the two analysis methods.

4. Analysis of Experimental Results

4.1. Introduction of Experimental Data. This paper mainly uses the system call trace of sendmail privileged process running in UNIX system as the experimental data. First, the privileged process is selected because of the following.

First, privilege process is the main channel for intruders to obtain certain authority; Second, the user behavior space is very wide, and the activity mode often changes greatly with time, so it is difficult to fully describe its behavior mode, while the behavior space of privilege process is limited, so it is easy to capture its behavior characteristics and internal relations. Based on the above two points, this paper selects the system call of a privileged process. Secondly, sendmail, as one of the most complex privileged processes in the UNIX system, is a program used to receive and send e-mail. At present, there are intrusion detection tests against sendmail attacks. Therefore, this paper can collect normal data and intrusion data. The data acquisition environment: the machine environment is Sun SPARC stations, the operating system is Sun OS4.1.1 and 4.1.4 without patches, with sendmail built-in, and the data acquisition software is March package version 3.0. Some normal data are synthesized and others are real-time. Synthetic tracking data is collected by running a preconfigured script in the actual environment. This script is designed to achieve training, not to meet the needs of users. Real-time normal data is collected when normal users use the computer system.

4.2. Model Establishment

4.2.1. Calculation of Initial State Probability Vector and One-step State Transition Probability Matrix of Markov Chain Model. In the experiment, this paper uses the system call of sendmail daemon as the normal behavior training data. This training data is used to determine the state of the Markov chain and calculate the probability vector of the initial state of the Markov chain. The sum of the Markov chain initial state probability vector is shown in Figure 4.

There are 53 system calls in sendmail decision-making process, and the system call number is as follows:1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 14, 17, 18, 19, 23, 27, 32, 38, 40, 4l, 45, 50, 54, 56, 59, 61, 66, 74, 75, 78, 83, 85, 88, 89, 93, 94, 95, 100, 101, 102, 104, 105, 106, 108, 112, 121, 122, 123, 124, 128, 138, 155, 167. Therefore, in Figure 4, the probability greater than 0 is the initial probability used by the above 53 systems.

4.2.2. Unknown System Call Detection. In the analysis and detection phase, unknown system calls in the test data can be detected. The test results of normal data are shown in Table 2; The detection results of five kinds of abnormal data are shown in Table 3. It can be seen from Tables 2 and 3 that only system calls that have not appeared in normal training data are detected in syslog attack. It also shows that the sendmail daemon with 53 system calls selected in this paper basically includes the system calls in the sendmail privileged process, and its system call sequence can be well used as the training data of the normal behavior of the model.

4.2.3. Parameter Analysis and Selection. There are two important parameters in this model: namely, the sliding window length L and W. L is the sliding window length when

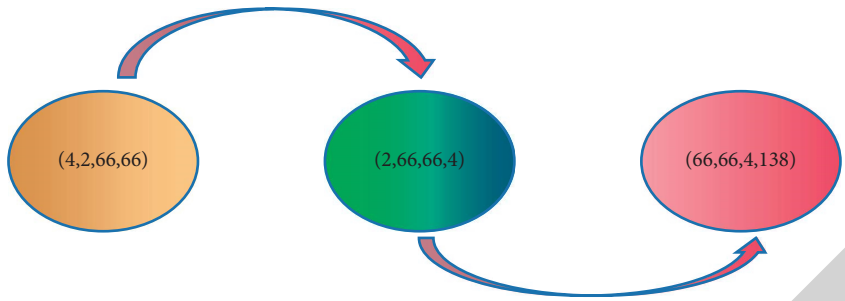

Figure 3: State transition diagram of Markov chain model.

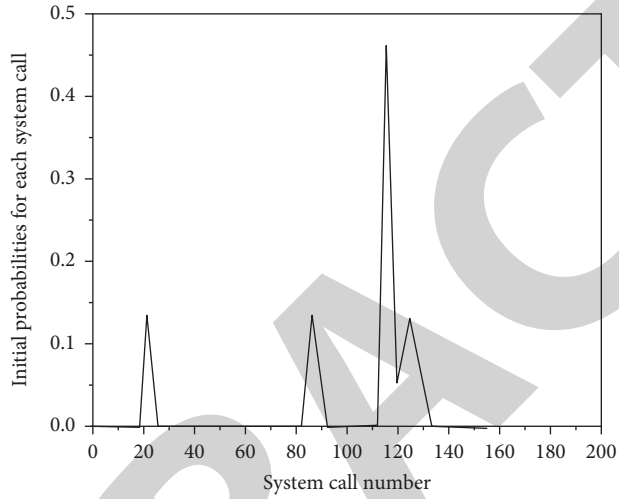

Figure 4: Probability vector diagram of the initial state.

Table 2: Test results of normal data.

File name (∗ .int)	Total number of system calls	Number of system call numbers	Location system call number
Plus	98194	56	Without
Queue	96342	50	Without
Sendmail	19531	45	Without

Table 3: Syslog attack detection results.

File name (∗ .int)	Total number of system calls	Number of system call numbers	Location system call number
Sysiog-local-1	1520	45	84, 52, 36
Sysiog-local-2	1583	52	84, 52, 36
Sysiog-remote-1	1867	48	84, 52, 36
Sysiog-remote-2	1558	43	84, 52

dividing the system call sequence, and W is the sliding window length when dividing the probability sequence. The values of L and w will determine the effect of the detection model. Therefore, this paper takes sendmail system call tracking record as an example to study the parameters L and W. The reason why sendmail is used for research is that the judgment value of sendmail in normal test data changes smoothly. As shown in Figure 5, a part can be intercepted for research.

In the case of $l = 2$ and $w = 90$, the probability sequence values calculated by the two one-step Markov chain analysis and detection methods are the same. After calculation, the length of the probability sequence is 18925. Here, we only intercept the first 1000 for research. After calculation, the value of the probability sequence is small, so it is calculated after sliding window processing so as to facilitate the comparison of normal data and abnormal data. Under the two analysis methods of the one-step Markov chain, after taking different values for L and W, the calculated decision value curve is shown in Figures 6 and 7. In Figure 6, the above group of curves: the solid line is $l = 3$ and $w = 90$, and the dotted line is $l = 3$ and $w = 100$; The following group of curves: the solid line is $l = 2$ and $w = 90$, and the dotted line is $l = 2$ and $w = 100$. In Figure 7, there are four groups of curves:

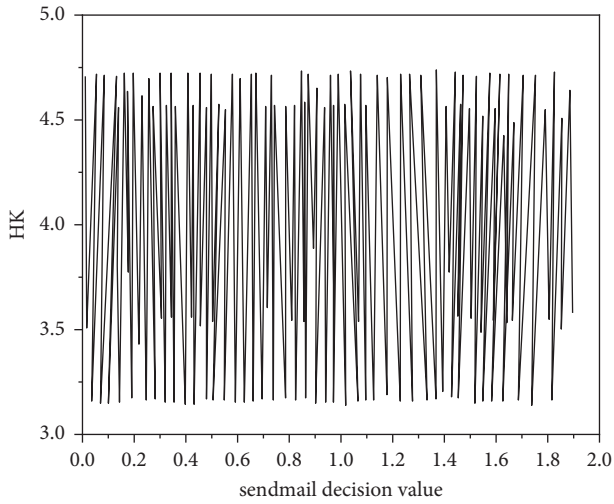

Figure 5: Sendmail decision value curve.

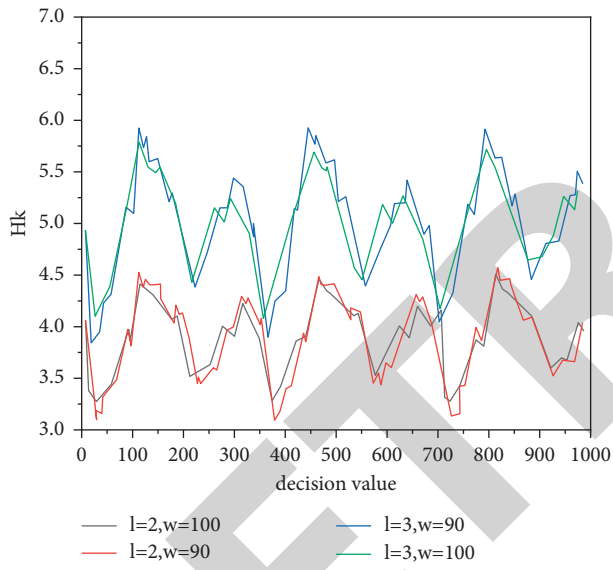

Figure 6: Curve of judgment value when the values of l and w are different (analysis method 1).

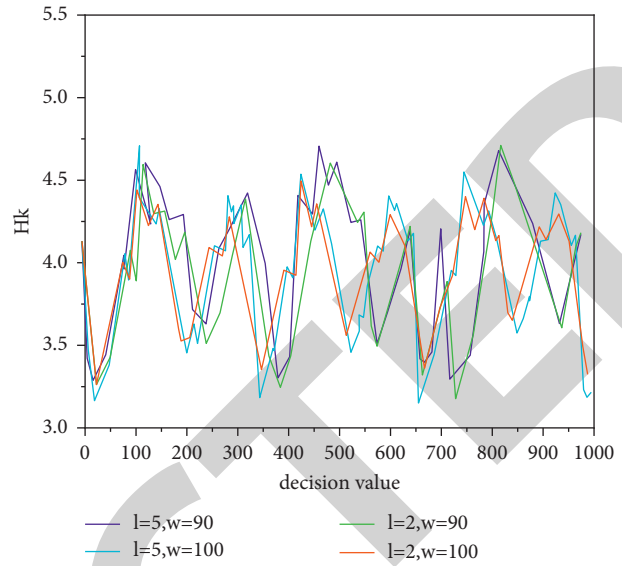

Figure 7: Curve of judgment value when the values of l and w are different (analysis method 2).

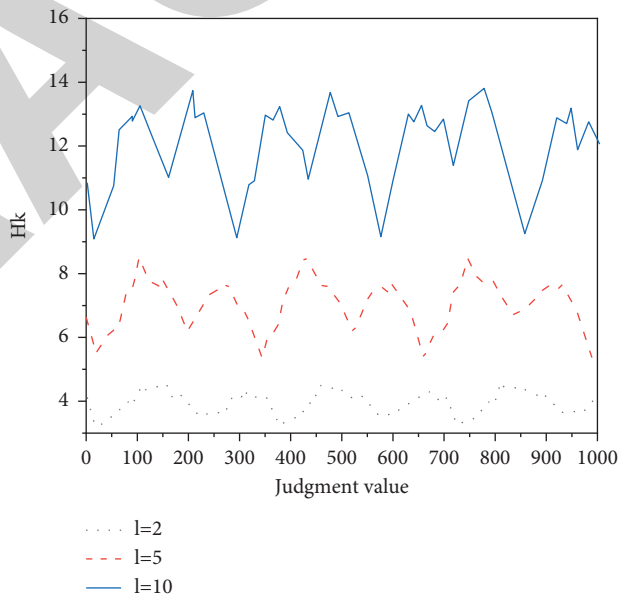

Figure 8: Sendmail judgment curve ($w = 90$, analysis method 1).

solid line $1 = 5$ and $w = 90$, dotted line $l = 5$ and $w = 100$, dotted line $l = 2$ and $w = 90$, and double-crossed line $l = 2$ and $w = 100$.

As can be seen from the above two figures, the variation range of the judgment value calculated in the case of analysis method l is determined by parameter l, and the variation amplitude is determined by parameter w, among which the dotted lines with $l = 2$ and $w = 100$ change most smoothly; In the case of analysis method 2, the change of amplitude and range is determined by l and w, and the dotted line with $1 = 5$ and $w = 100$ changes the most smoothly; Therefore, this paper selects $l = 2$, $w = 100$ (analysis method 1), $l = 5$, $w = 100$ (analysis method 2).

After parameters L and W are selected, we study the decision value curve again by changing the value of another parameter on the premise of setting one parameter to verify whether the above-selected parameter values are reasonable. Case 1: under the two analysis methods, l takes 2, 5, 10, $w = 90$; In Case 2, under the premise of $l = 2$ and $l = 5$, respectively, W is taken as 30, 45, 90, and 100. Respectively, calculate the judgment value and draw its curve, as shown in Figure 8 and 9. Because the decision value curve changes smoothly, we only intercept the first 1000 decision values Hk for research.

As can be seen from Figure 8, when w is certain, the larger L is, the more obvious the change of the judgment value curve is. Therefore, when calculating the judgment value in the case of analysis method 1, the $1 = 2$ (solid line) selected in this paper is feasible. As can be seen from

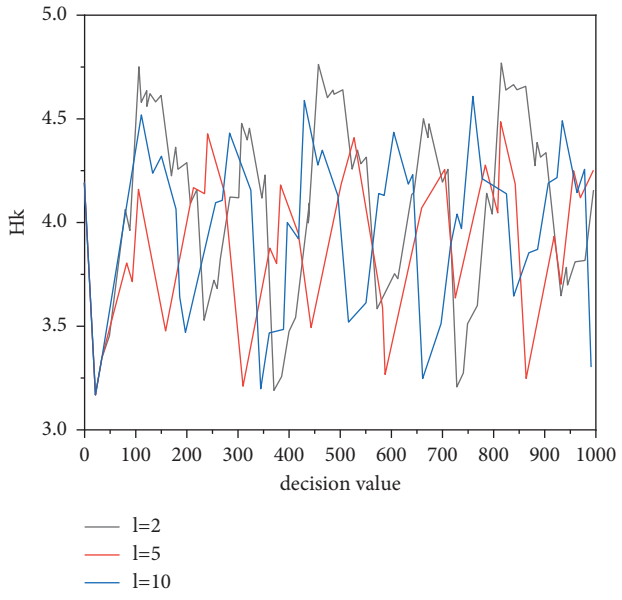

Figure 9: Curve of sendmail judgment value ($w = 90$, analysis method 2).

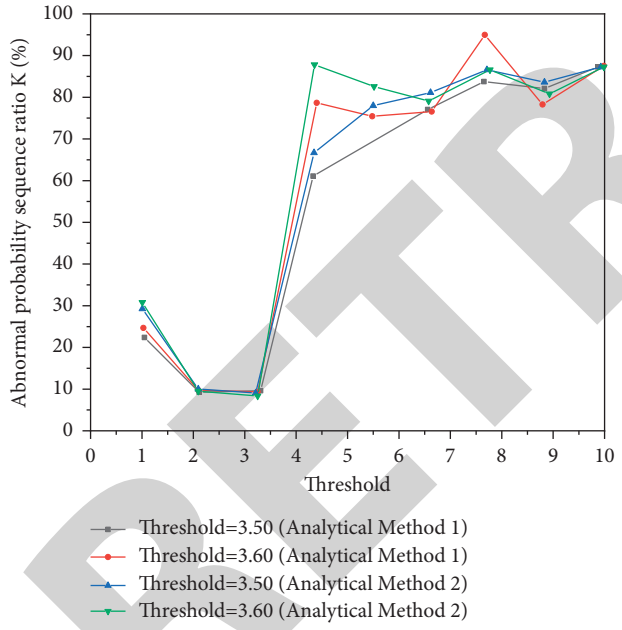

Figure 10: Ratio of threshold value to abnormal probability sequence K (%).

Figure 9, the change of L does not have a great impact on the decision value curve. Therefore, when calculating the judgment value in the case of analysis method 2, it is also feasible to select $l = 5$ (dotted line) (although the curve with $l = 10$ changes most smoothly, if 1 is too large, the properties of the Markov chain will be affected).

4.3. Threshold Selection and Analysis. In many works of literature, threshold parameters are used in the research of the Markov chain. By giving a threshold, the detected sequence is calculated. If the calculation result is greater than the set threshold, it indicates that the sequence is normal. The selection of the general threshold is based on the experience of the system administrator and the characteristics of the data. In this paper, we also study the threshold according to the characteristics of this paper. In this paper, by observing the decision value curves of normal test data and abnormal test data, we select different thresholds and then calculate the proportion of abnormal probability series, respectively. In order to detect anomalies, we will reasonably select the threshold. The formula for calculating the proportion of abnormal probability series:

Abnormal probability sequence proportion K (%) = number of probability sequences between the upper and lower thresholds/total number of probability sequences.

As can be seen from Figure 10, k of abnormal test data is larger, K of normal test data is smaller, and K increases with the increase of threshold. For normal test data, the K calculated by analysis method 1 is better than that calculated by analysis method 2, but the gap between the two is relatively small. For abnormal test data, most of the K generated by using analysis method 2 is better than that calculated by using analysis method 1. Therefore, in general, first, this method can easily distinguish between normal and abnormal data; Second, the detection result in the case of analysis method 2 is better than that in the case of analysis method 1.

5. Conclusion

Markov chain model is a commonly used statistical analysis model of intrusion detection, which has been continuously studied and developed in recent years. This method mainly improves the model or combines it with other methods to achieve the purpose of intrusion detection. However, each method has its own advantages and disadvantages. Based on the Markov chain model, this paper improves the analysis and detection method; that is, the detection data are analyzed and detected by using two analysis and detection methods of the one-step Markov chain and the special formula in the references. The details are as follows: firstly, the Markov chain model is established; Secondly, the detected data is detected based on the model: two different sequence analysis and detection methods of one-step Markov chain are used to calculate the detected data to generate probability sequence, and then the sliding window analysis method is used to analyze the probability sequence, and each segmented probability sequence is calculated by special formula to generate two judgment values; Third, analyze two kinds of decision value curves to see whether normal data and abnormal data are effectively detected, and compare the two kinds of decision values; Finally, in order to achieve the ideal experimental results, through the analysis of the judgment value, set a reasonable threshold, define the proportion of abnormal probability sequence, and use the proportion of abnormal probability sequence to detect the normal data and abnormal data again.

In this paper, the sendmail privileged process system call tracking record is used to experiment and verify the model. Two methods of decision value curve and threshold are used

to judge the normal and abnormal data, and a good detection effect is obtained. Therefore, it shows the effectiveness of the new method proposed in this paper. The new method in this paper has the characteristics of simple calculation, low algorithm complexity, and easy online detection. It overcomes the disadvantage that the single-step Markov chain analysis and detection method cannot be strictly established in the nature of the Markov chain, which has lower algorithm complexity than the multistep Markov chain analysis and detection method and is simpler than the parameter calculation of hidden Markov chain model. Although the model has many advantages, we still need to do a lot of work in the future to make it work: first, study the properties of various normal data and abnormal data, and further study the parameters of the method according to the properties of different data; Second, experiment with the data of various privileged processes to make the method universal. Third, build a practical platform for online real-time detection.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] V. de Paula, J. J. Curto, and T. Sole, "Application of the Markov chain model to sunspots and solar plages for the period 1910 to 1937 using data from ebro catalogues," *Solar Physics*, vol. 296, no. 6, p. 92, 2021.

[2] Y. Huang, B. Yang, M. Wang, B. Liu, and X. Yang, "Analysis of the future land cover change in Beijing using CA-Markov chain model," *Environmental Earth Sciences*, vol. 79, no. 2, p. 60, 2020.

[3] P. Di Sanzo, A. Pellegrini, M. SaNnIcandro, B. Ciciani, and F. Quaglia, "Adaptive model-based scheduling in software transactional memory," *IEEE Transactions on Computers*, vol. 69, no. 5, pp. 621–632, 2020.

[4] A. Theorell and K. Nöh, "Reversible jump mcmc for multimodel inference in metabolic flux analysis," *Bioinformatics*, vol. 36, no. 1, pp. 232–240, 2020.

[5] P. K. Quoika, M. L. Fernández-Quintero, M. Podewitz, F. Hofer, and K. R. Liedl, "Implementation of the freely jointed chain model to assess kinetics and thermodynamics of t coil-globule transition by markov states," *The Journal of Physical Chemistry B*, vol. 125, no. 18, pp. 4898–4909, 2021.

[6] M. K. Elmezughi, T. J. Afullo, and N. O. Oyie, "Performance study of path loss models at 14, 18, and 22 ghz in an indoor corridor environment for wireless communications," *SAIEE Africa Research Journal*, vol. 112, no. 1, pp. 32–45, 2021.

[7] C. Hu, L. Dai, S. Han, and X. Wang, "Two-timescale channel estimation for reconfigurable intelligent surface aided wireless communications," *IEEE Transactions on Communications*, vol. 69, no. 11, pp. 7736–7747, 2021.

[8] S. Sugiura, Y. Kawai, T. Matsui, T. Lee, and H. Iizuka, "Joint beam and polarization forming of intelligent reflecting surfaces for wireless communications," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 2, pp. 1648–1657, 2021.

[9] W. Dai, Q. Yan, Z. Hong, Z. Li, M. Wang, and C. Yang, "Signal demodulation method for underwater optical wireless communication by measuring the time interval between adjacent photons," *Photonic Network Communications*, vol. 41, no. 2, pp. 202–210, 2021.

[10] I. N'Doye, D. Zhang, M. S. Alouini, and T. M. Laleg-Kirati, "Establishing and Maintaining a Reliable Optical Wireless Communication in Underwater Environment," *IEEE Access*, vol. 9, no. 99, p. 1, 2021.

[11] J. Huang, C.-X. Wang, H. Chang, J. Sun, and X. Gao, "Multi-frequency multi-scenario millimeter wave mimo channel measurements and modeling for b5g wireless communication systems," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 9, pp. 2010–2025, 2020.

[12] K. Seong, D.-K. Jung, D.-H. Yoon et al., "Time-interleaved SAR ADC with background timing-skew calibration for UWB wireless communication in IoT systems," *Sensors*, vol. 20, no. 8, p. 2430, 2020.

[13] H. Hao, D. Hui, and D. Lau, "Material advancement in technological development for the 5g wireless communications," *Nanotechnology Reviews*, vol. 9, no. 1, pp. 683–699, 2020.

[14] J. B. Padhy, B. Patnaik, and I. Bhubaneswar, "Co-ofdm and dp-qpsk based dwdm optical wireless communication system," *Journal of Optical Communications*, vol. 42, no. 2, pp. 311–323, 2021.

[15] O. F. Rashid, Z. A. Othman, S. Zainudin, and N. A. Samsudin, "Dna Encoding and Str Extraction for Anomaly Intrusion Detection Systems," *IEEE Access*, vol. 9, no. 99, p. 1, 2021.

[16] S. Sarvari, N. Sani, Z. M. Hanapi, and M. T. Abdullah, "An Efficient Anomaly Intrusion Detection Method with Feature Selection and Evolutionary Neural Network," *IEEE Access*, vol. 8, no. 99, p. 1, 2020.

[17] R. E. Alonso, J. E. Diaz-Verdejo, A. E. Alonso, and G. Madinabeitia, "How Much Training Data Is Enough? a Case Study for Http Anomaly-Based Intrusion Detection," *IEEE Access*, vol. 8, no. 99, p. 1, 2020.

[18] M. Hosseinzadeh, A. M. Rahmani, B. Vo, M. Bidaki, M. Masdari, and M. Zangakani, "Improving security using svm-based anomaly detection: issues and challenges," *Soft Computing*, vol. 25, no. 4, pp. 3195–3223, 2021.

[19] W. Song, M. Beshley, K. Przystupa et al., "A software deep packet inspection system for network traffic analysis and anomaly detection," *Sensors*, vol. 20, no. 6, p. 1637, 2020.

[20] H. Gonzalez, L. Bloise, F. J. Maza, V. A. Molina, and A. Delorenzi, "Memory built in conjunction with a stressor is privileged: reconsolidation-resistant memories in the crab neohelice," *Brain Research Bulletin*, vol. 157, pp. 108–118, 2020.

[21] M. A. Werder, M. Huss, F. Paul, A. Dehecq, and D. Farinotti, "A bayesian ice thickness estimation model for large-scale applications," *Journal of Glaciology*, vol. 66, no. 255, pp. 137–152, 2020.

[22] T. Qin, R. Du, A. Kusiak, H. Tao, and Y. Zhong, "Designing a resilient production system with reconfigurable machines and movable buffers," *International Journal of Production Research*, vol. 2021, no. 6, pp. 1–16, 2021.

[23] B. Cai, L. Zhang, and Y. Shi, "Control synthesis of hidden semi-markov uncertain fuzzy systems via observations of

hidden modes," *IEEE Transactions on Cybernetics*, vol. 50, no. 8, pp. 3709–3718, 2020.

[24] Z. Wang, N. Guo, S. Wang, and Y Xu, "Prediction of highway asphalt pavement performance based on Markov chain and artificial neural network approach," *The Journal of Supercomputing*, vol. 77, no. 2, pp. 1354–1376, 2021.

[25] A. Koutsioubas, "anaklasis: a compact software package for model-based analysis of specular neutron and X-ray reflectometry data sets," *Journal of Applied Crystallography*, vol. 54, no. 6, pp. 1857–1866, 2021.