Research Article

Application of Computer Data Mining Technology Based on AKN Algorithm in Denial of Service Attack Defense Detection

Xiang Huang

Hunan Mass Media Vocational and Technical College, Changsha 410100, Hunan, China

Correspondence should be addressed to Xiang Huang; 332374935@qq.com

Received 28 December 2021; Revised 14 January 2022; Accepted 19 January 2022; Published 12 February 2022

Academic Editor: Alireza Souri

Copyright © 2022 Xiang Huang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Denial of service attacks have become one of the most difficult network security problems because they are easy to implement, difficult to prevent, and difficult to track, and they have brought great harm to the network society. Denial of service (Dos) is a phenomenon in which a large number of useless data packets or obstructive content are maliciously transmitted to the target server, which makes the target server unable to provide users with normal services. Denial of service attack (Dos attack) is a very typical network attack method, and the main harm of Dos attack is to exhaust service resources, making the computer or network unable to provide normal services. And AKN (adaptive Kohonen network) is an adaptive neural network proposed in recent years, and an algorithm summarized by using the characteristics of the neural network is called the AKN algorithm. This algorithm can realize fast, low-consumption, and high-precision denial of service detection in complex networks. In the era of big data, network security is becoming more and more important, and in order to maintain the security of network data, this article studies the common forms and principles of Dos attacks, as well as the current corresponding defense detection methods. It also investigates several commonly used algorithms of computer data mining technology, such as clustering algorithms, classification algorithms, neural network algorithms, regression algorithms, website data mining, and association algorithms, and proposes a computer data mining model based on the AKN algorithm. In addition, the computer data mining technology based on the AKN algorithm is used to conduct defensive detection experiments under Dos attacks and compares with classic algorithms. Experimental results show that experiments based on the AKN algorithm have better defensive detection effects than classic algorithms, with a detection accuracy rate of more than 97% and a detection efficiency improvement of more than 20%.

1. Introduction

1.1. Background and Significance. The paralysis of a major network may endanger the information security of individuals, enterprises, and even the country; therefore, maintaining the security of network information is a major responsibility and mission. Dos attack is an important hacking method, and its purpose is to exhaust all available network resources based on the large amount of traffic reaching the network. However, the target computer or network usually cannot provide normal services and ultimately cannot pass legitimate user requests. Data mining technology can realize the capture of data information, the inspection of data information, and the analysis of abnormal problems in the computer network system. It analyzes whether there is abnormal data and then takes a series of

defensive measures to completely eliminate the abnormal data, thereby maintaining a safe network environment. The AKN algorithm can accurately and efficiently detect complex denial of service attacks. It uses the above characteristics to study and explore the theoretical feasibility and practical applicability of the computer data mining technology using the AKN algorithm for defensive detection under Dos attacks. The research results will be able to achieve a great improvement in the efficiency of network security detection and provide better protection for network security.

1.2. Related Work. In recent years, the rapid development of the Internet, big data, and electronic communications has brought great convenience to people, but there are also more and more network insecure factors, such as virus threats,

privacy leaks, and telecommunication fraud. Some hackers who specifically attack other network ports for the purpose of stealing data and destroying servers are even more terrifying than network viruses. In order to maintain the security of network data, many scholars have conducted indepth research on this. Adi et al. used legitimate traffic or fast crowds that may have the high traffic characteristics seen in DDoS attacks (distributed denial of service attack is a common type of Dos attack) to test in four different protocol-related attack scenarios, and they proved that legitimate HTTP/2 flash crowds and traffic can be activated to cause rejection service [1]. Chang and Hu built a security MAC (media access control) system to defend against destructive denial of service (DDoS) attacks and network control threats while retaining the benefits of coordination among cooperative users. The theoretical analysis and implementation evaluation of Chang and Hu proved that its MAC system is superior to other solutions by 76–159% [2]. Douglas et al. checked whether there are ethical justifications for using or operating websites that provide users with targeted distributed denial of service (DDoS) attacks (called “bootstrap programs”). Douglas et al. identified the parties related to the bootstrap website and the way the bootstrap program operates, and they also studied the potential reasons for checking the use and operation of the bootstrap website and criticized citizens who use DDoS attacks in terms of morality [3]. Pan et al. analyzed the law of selecting points for diabetic peripheral neuropathy (DPN) based on data mining technology. Pan et al. analyzed the acupuncture prescriptions by applying data mining methods such as rule analysis and modified mutual information, calculated the frequency of each acupoint, and obtained 19 acupoint combinations. And it is concluded that acupuncture DPN is mainly based on the replenishment method, promotes qi and blood circulation, and chooses the main points of Yangming meridian and Backshu [4]. Zhao-Yi et al. used computer data mining technology to extract and analyze the regular information of the prescription of antiwind medicine, which promotes its use, research, and development in modern clinical medicine. Zhao-Yi et al. used a series of data processing software for data mining and comprehensive analysis. He consulted the works of dispelling wind medicine, counted the number of prescriptions of dispelling wind medicine, and determined the frequency of use of dispelling wind medicine. Through association rule analysis, factor analysis, and core drug network analysis, they revealed the association between syndrome symptoms and wind-dispelling drug treatment and revealed the law of the use of these drugs [5]. Hu et al. proposed a new method for data stream clustering with the help of adaptive neuro-fuzzy system integration. He found that the proposed ensemble is formed by adaptive neuro-fuzzy self-organizing feature maps in the parallel processing mode, and the final result is selected by the best neuro-fuzzy self-organizing feature maps [6]. Denial of service is also one of the most pressing security threats to in-vehicle network systems. In order to provide perfect detection, Durrani et al. used a decision tree (an intuitive graphical decision analysis method for evaluating project risks and judging its feasibility) classification

to identify changes in traffic behavior in a timely manner with a low error rate as a detection solution. Durrani et al. described the characteristics of VANETs, aiming to determine the most critical types of Dos attacks on vehicle networks. In addition, he also comprehensively reviewed the available solutions in the current literature to mitigate this attack on in-vehicle communications [7]. Distributed Denial of Service (DDoS) attacks are also one of the most prominent attacks in the cloud-assisted wireless body area network (WBAN), and it not only interrupts communications but also reduces network bandwidth and capacity. Abbas et al. proposed an enhanced very fast decision tree (EVFDT) that can successfully detect DDoS attacks. The experimental results show that the EVFDT algorithm achieves high detection accuracy with a low false alarm rate. However, these detection methods either cannot be used in complex networks or cannot be accurately detected, or the detection speed is too slow. The AKN algorithm can perform accurate and efficient detection in a complex network environment [8].

1.3. Innovation. The AKN algorithm can be used for accurate and efficient detection in a complex network environment. Using the superdata search and analysis capabilities of data mining technology, it can effectively screen virus programs in computer networks. The common types and principles of denial of service attacks are analyzed, which can better deal with Dos attacks. Experimental comparison with traditional data mining technology can better reflect the advantages of the AKN algorithm.

2. Data Mining Scheme Based on AKN Algorithm

2.1. Dos Attack Theory.

- (1) The emergence of Dos attacks: there will be some system and software vulnerabilities in the process of system design and production. Attackers use these vulnerabilities to carry out malicious attacks in order to obtain a certain benefit or achieve a certain purpose [9].
- (2) Characteristics of Dos attacks: Dos attacks are easy to implement, with high attack intensity, wide attack range, strong distribution of attack sources, and strong attack concealment, and attacks are difficult to defend. Because Dos attack tools can be seen everywhere, anyone can use these attacks, even if they do not have much understanding of network security. Even people who do not have much understanding of computers and networks can download readymade tools from the Internet as long as they have access to the Internet and then launch attacks on selected victims at will. In addition, offensive data packets, especially storm attack data packets that rely on a large number of data packets to attack, are indistinguishable from normal data packets. Therefore, it is extremely difficult to prevent denial of service attacks from the 'victim's side.

FIGURE 1: Schematic diagram of a distributed denial of service attack.

- (3) Types of Dos attacks: according to the principle of attack, it can be divided into resource exhaustion type, service interruption type, and physical damage type. The resource exhausted type means that the attacker uses some means to send a large amount of data to the target server to exhaust its service resources, such as CPU, bandwidth, and RAM. Disrupted service refers to the attacker's use of defects or bugs in the target server itself to cause the server to crash. Physical damage type refers to a denial of service attack that destroys the target server by physical means, such as using water or fire to short-circuit or disconnect the relevant network cable or directly destroy the router or network cable related to the network cable. According to the attack method, the first is called a virus packet attack, which is to send a large amount of abnormal data based on the flaws and vulnerabilities in the design of the target server. In the process, it makes the victim's system abnormal and causes the victim's system to crash. The second one is called a storm attack which sends a large number of data packets to the target system, and when the maximum number of data packets is accumulated, the target server will be compromised, and the typical distributed Dos attack is a storm attack. The third is called a redirect attack, which does not use virus packets or storms to attack the victim but modifies some parameters in the network. For example, the cache of the domain name system allows the data packets sent by the victim or

sent to the victim to be redirected to other places [10].

- (4) Principle of Dos attack: taking a distributed denial of service attack as an example, in order to hide themselves, the attacker first finds machines with a low management level, low security, and easy scanning, and then destroys and controls these puppet machines. The attack process is relatively simple, and the attacker manipulates the computer to send a series of control commands to the puppet computer and can transmit a huge amount of offensive data to the target server immediately or at a specific point in time. Or the attacker sets the puppet computer within a specific range of time, and when the time runs out, the puppet computer starts to attack the target server. The details are shown in Figure 1.
- (5) Defense detection of Dos attacks: at present, there is no complete solution for the defense and detection of Dos attacks; generally, the system is patched or upgraded. The defense methods can be summarized into four types. The first is attack detection, which is to scan and detect Dos attacks in the network, generate alerts, and trigger an attack response mechanism. The second is attack response; that is, after the server is subjected to a Dos attack, the attack traffic is reduced through message filtering, speed limit, etc., to reduce the impact of the Dos attack and to restore the server to normal. This function needs to be implemented as soon as possible to ensure the availability of the service. The third type is the source trace, which is to track the attacker behind the Dos or the final real host. The fourth is antiattack; that is, according to the formation conditions and principles of Dos attacks, protective measures such as patches and encryption must be taken in advance and moral criticism to reduce the number of attackers [11].
- (6) Dos attack status: malicious attacks from the network layer may interrupt the transmission of data packets in the channel. For example, a compromised router in the network may refuse to send or receive data packets, and Dos attacks can also cause packet loss. Assuming that the total length of the attack time is restricted in proportion to the total time, the attack state of Dos can be recorded as shown in formula (1) and meets the requirements of formula (2).

$$\{\theta_s(t) \in \{0, 1\}\}_{t \in Z_0}, \quad (1)$$

$$Q \left[\sum_{t=0}^{n-1} \theta_s(t) \leq n + \frac{n}{\omega} \right] = 1, \quad n \in Z. \quad (2)$$

Among them, $t \in Z_0$ is the moment when the data packet is about to be sent between the sensor and the controlled object. $\theta(t) = 0$, which means that the data packet transmission in the channel at time t is successful, and $\theta(t) = 1$, which means that the data packet transmission at time t has

failed, which means that it has suffered a Dos attack at this time. And $n \geq 0$, $\omega > 1$, for the number of data packets transmitted is n , at most $n + \omega n$ data packets are affected by the Dos attack.

Noting that when $n = 0$, it means that there is no Dos attack at the initial moment, that is, $\theta(t) = 0$. Due to the limitation of its own energy, the attacker cannot cause all data packets to be lost, and $1/\omega$ represents the average value of data packets subjected to Dos attacks [12].

2.2. Data Mining Technology. Data mining technology is one of the most important technologies for social development in the context of big data. Through the generation and use of computer mining technology, the data information obtained is more complete and effective, the processing accuracy is also high, and more complete results can be obtained. From the analysis of real-life applications, big data applications can develop visual data processing technology, and its data computing and processing capabilities are the most important. At present, the algorithms of data mining technology are as follows: clustering algorithm, classification algorithm, neural network algorithm, regression algorithm, website data mining, and association algorithm. The AKN algorithm studied in this article is a neural network algorithm that simulates the learning and thinking process of the human brain. After training a few samples, we generated a similar algorithm that can distinguish different sample data. Data mining technology extracts data rules that people cannot easily reflect intuitively from huge, noisy, and large-scale data because all these data are a useful potential rule. Traditional data mining techniques usually start with limited traditional data mining techniques. With the widespread use of the Internet, network big data analysis technologies are constantly upgrading. The data analysis technology that combines traditional data mining with the basic concepts of the Internet is widely used in today's world. Network data mining technology quickly and accurately identifies potentially useful information to users and various behavioral information hidden in the network, and provides users with various services with different needs [13].

The basic process of computer-based data mining technology is shown in Figure 2, and determining the target of data mining is a major task of data analysis. Firstly, the main way to determine goals is to conduct user interviews, investigate and analyze key executives in different departments, determine the final business goals, and conduct business understanding analysis. Secondly, data mining engineers need to perform "data understanding" tasks based on business understanding to realize data collection and management. The data preparation process is more cumbersome because the process is to check and modify data, use network information to collect and organize data, and convert and modify data. After the change is completed, it provides a basis for establishing a data model. Once the modeling is completed, the model needs to be optimized multiple times to complete all the work in the data mining cycle. With the continuous deepening of user applications, data mining strategies and models are also continuously

optimized, and this mining technology can only operate stably after the business rules are stable [14].

2.3. AKN Algorithm.

- (1) The source of the AKN algorithm: the Kohonen network (KN for short) proposed by Professor Teuvo Kohonen of the University of Helsinki in Finland is a self-organizing competitive neural network. Self-organization means that the network is unsupervised and can learn independently, and according to the characteristics of the environment, the element value can be adjusted through the self-organizing function relationship so that the neural network can automatically distinguish and aggregate classification. In this form of expression, neurons will match a specific input form and enhance the impression, resulting in sensitivity. Therefore, the input form can be divided into different clusters through self-organized training and learning. Each cluster has different response characteristics to the input form, so the neuron can become a detector in a certain environment [15].
- (2) The principle of the AKN algorithm: the operating mechanism of KN is that when data are input to the neural network, the Euclidean distance (Euclidean distance, also called meta value in this article) between the input node of the neuron node of the output layer and the neuron node of the output layer is calculated. The victory cell is the neuron with the smallest Euclidean distance, and the cell value coefficients of the victory cell and neighboring neurons can be adjusted to make the cell values of the generated neuron and surrounding neurons closer to the input sample. After multiple exercises, the finally connected element value distribution of each neuron has a specific relationship. This distribution aggregates similar patterns between input values to be divided into different types of neurons, so neurons of the same type have similar element value factors, but different types of neurons have great differences in value coefficients. During practice, the efficiency of the modified cell value and the cell value of neighboring neurons continue to decrease; that is, the neurons of the same kind are gradually gathered. The traditional algorithm that uses the characteristics of the neural network to perform calculations is called the KN algorithm. The traditional unsupervised KN algorithm has limitations in its classification. Unsupervised classification of unknown category data is feasible, but the results of the same type of data will correspond to different network nodes. For one-to-one corresponding node categories, the KN classification categories are more than the actual data categories. In order to improve the classification effect, the improved AKN algorithm proposed in this paper adds a supervised learning process after the first stage of unsupervised learning to allow the system to adapt to the correct classification results. The principle of this algorithm is shown in Figure 3.

FIGURE 2: Flowchart of data mining.

FIGURE 3: Schematic diagram of the AKN algorithm.

As shown in Figure 3, if there are S KN output nodes, denoted by f , and the input node is at the bottom, the node is represented by k , and the calculation is represented by the symbol u . If the input vector has n elements, there are n nodes in total at the input. C_{Kf} represents the size of the element value from the input node to the output node, and the output nodes on the same plane can also be connected to each other. The function of the AKN algorithm is to increase the supervised learning process, which uses a large number of samples for training through independent learning, continuously adjusts the network's element value, and finally obtains the clustering distribution of the data [16].

- (3) AKN algorithm content: the algorithm steps are as follows:

Input: training and test samples;

Output: the sample element value coefficient matrix after training and the clustering category and test accuracy of the test sample;

- (1) t_0 is the start, t_1 is a constant, representing time, and t_2 is the total number of training; the value of the connection element from the input node to the output node is assigned with a random number, combined with the expression of the sum of the neighborhood $A_{f,f(x)}(s)$ and the training rate $\eta(s)$, and the number of output neurons is S .

- (2) Network input mode

$$Z = [z_1, z_2, \dots, z_K]^T. \quad (3)$$

Among them, K is the dimension of the input vector.

- (3) Initializing the weight:

$$C_f = [C_{1f}, C_{2f}, \dots, C_{kf}]^T. \quad (4)$$

Among them, $f = 1, 2, \dots, S$.

- (4) Calculation of sample vector and element value:

$$d_f = \|Z - C_f\| = \sqrt{\sum_{u=1}^S (x_u - C_{uf}(t))^2}. \quad (5)$$

Then, the element value expression of the victory element f^* is

$$d_{f^*} = \min\{d_f\}. \quad (6)$$

- (5) Adjusting the element value vector of nodes connected in the output and geometric neighborhood:

$$C_f(n+1) = C_f(n) + \eta(s)A_{f,f(x)}(n)(Z - C_f(n)). \quad (7)$$

Among them, $A_{f,f(x)}(s)$ is the neighborhood adjustment function.

- (6) If there is still a training sample dataset, go back to step (2);

- (7) If all samples are trained and $t < t_2$, $t = t + 1$, return to step (2); otherwise, return to step (8).

- (8) Training is over [17].

As shown in (8), $\eta(s)$ is the training rate at s , and as the number of training continues to increase, the value of s continues to increase, and then the inverse proportional relationship according to (8) will continue to decrease.

$$\eta(s) = \eta_1 \exp\left(1 - \frac{s}{t_2}\right). \quad (8)$$

Among them, t_2 is the total number of training; the initial training rate is η_1 , $s = 0, 1, 2, \dots$

The adjustment of the neighborhood is done by

$$A_{f,f(x)}(s) = \exp\left(1 - \frac{d_{f,u}^2}{2\varepsilon(s)^2}\right). \quad (9)$$

Among them, the distance from each neuron on the same plane to the victory element is represented by $d_{f,u}$, and

the parameter that controls the data distribution is represented by $\varepsilon(s)$. The adjustment formula is σ

$$\varepsilon(s) = \varepsilon_1 \exp\left(1 - \frac{n}{t_1}\right). \quad (10)$$

And, the training and test data are normalized, as shown in formulas (11)–(13):

$$x = \frac{(x_{\max} - x_{\min}) \times (y - y_{\min})}{y_{\max} - y_{\min}} + x_{\min}, \quad (11)$$

$$x'_{kf} = \frac{x_{uf}}{\sqrt{x_{1f}^2 + x_{2f}^2 + \dots + x_{kf}^2}}, \quad (12)$$

$$C'_{kf} = \frac{C_{uf}}{\sqrt{C_{1f}^2 + C_{2f}^2 + \dots + C_{kf}^2}}. \quad (13)$$

According to the Euclidean distance d_f of the above formula (5), the degree of clustering between the input sample and the output neuron is expressed as

$$\begin{aligned} \rho(C_f(Z)) &= \{\rho(d_f(Z)), 0, \sqrt{0.2}\}, \\ \rho(C_f(Z)) &= e^{-d_f(Z)^2/0.4}. \end{aligned} \quad (14)$$

When $\rho > 0.5$, the output neuron is defined as the cluster to which the input sample belongs.

According to the category, the average value of each category is calculated, as shown in the following formula:

$$\bar{d}_x = \frac{x'_{kf}}{x_{1f} + x_{2f} + \dots + x_{kf}}, \quad (15)$$

$$\bar{d}_C = \frac{C'_{kf}}{C_{1f} + C_{2f} + \dots + C_{kf}}. \quad (16)$$

Using the cluster data calculation model, formula (17) can be used to obtain the neuron center value of the cluster [18].

$$d(f, u) = \sqrt{C_1|x_{1u} - x_{1f}|^2 + C_2|x_{2u} - x_{2f}|^2 + \dots + C_k|x_{ku} - x_{kf}|^2}. \quad (17)$$

3. Experiment and Analysis

3.1. Model Design of AKN Algorithm. Denial of service attack detection based on the AKN (adaptive Kohonen network) algorithm includes four aspects: data collection, collection of characteristic element values, cluster analysis, and detection and judgment. The data detection method of the AKN algorithm used in this paper is to divide the data sample into multiple detection paths per unit time and then collect the special element values of the detection paths for cluster analysis according to the impact of Dos attacks on the degree of clustering. In the pretrained nonattack sample data, the range value is extracted, and the cluster is analyzed according to the principle of Dos attack to determine

whether a Dos attack occurs. The denial of service attack detection model based on the AKN algorithm can realize high-speed, low-consumption, and high-precision detection of complex network denial of service attacks. According to the previous content, we know the principle of Dos attack, the principle of AKN algorithm, and the process of data mining technology, from which we establish an algorithm model [19].

To select data from the learning database, we used the traditional KN algorithm and the AKN algorithm to analyze 300, 500, 1,000, 2,000, 3,000, 4,000, and 5,000 datasets with three-dimensional features for small-class clustering and multiclass clustering. The result of the comparison is shown in Figure 4 [20].

In Figure 4, it can be seen that as the sample points increase, the squared errors of the two algorithms increase. And the AKN algorithm shown in the left picture is calculated with 2,500 sample points; the square of error is 165, while the square of error of the traditional KN algorithm on the right is 255. Under the calculation of 5,000 sample points, the error is 310, while the traditional KN algorithm on the right is 430. Through the comparison of the sum of square error function values of the two algorithms through the above numerical experiments, it can be found that the performance of the AKN algorithm is better than that of the KN algorithm when the number of clusters is certain, and the more points, the more obvious the effect; in the case of a certain number of points and more clusters, the better the performance of the AKN algorithm. From the following experiments, it can be seen that the average performance of the new algorithm when clustering multiple classes is about 30% higher than the average performance of the KN algorithm, while the average performance of the AKN algorithm when clustering fewer classes is about 20% higher than the average performance of the Kohonen algorithm. In data mining, cluster analysis with a large amount of data and a large number of categories is often encountered, and the detection effect of the AKN algorithm will be good.

To study the actual utility of the AKN algorithm, the high-configuration Intel Core i9 processing computer used in this paper is equipped with a 2.6 GHz frequency graphics card and uses 6GB of memory; however, the actual use environment uses a complex multiclustering AKN network model. Random 22,000 data records are selected; 6,000 records are used for AKN neural network training, and the other 16,000 records will be used for testing. Table 1 shows the identification types of test data, which are divided into five categories: Dos attack, Probing, Normal, R2L, and U2R [21].

The output neuron node is located in the 8×8 square matrix. The victory element node is used as the sample node, and the competing neuron layer node in the same plane with the shortest distance from the input sample is calculated according to formulas (5) and (6). The meta value of the winning meta node is adjusted according to formula (7), which allows the domain radius and learning rate to gradually decrease with the progress of the process, thereby allowing the input data to gradually concentrate on multiple nodes, and the neural network can complete the clustering

FIGURE 4: Error square graph of AKN algorithm and traditional KN algorithm.

TABLE 1: Identification types of test data.

Type	Meaning	Classification ID
Normal	Normal record	Normal
DoS	Denial of service attack	Back, land, Neptune, pod, smurf, teardrop
Probing	Surveillance and other detection activities	psweep, nmap, portsweep, satan
R2L	From remote machine unauthorized access	ftp_write, guess_passwd, imap, mul-tihop, phf, spy, warezclient, warez-master
U2R	Illegal access by ordinary users to local superuser privileges	Buffer_overflow, loadmodule, perl, rootkit

function. In this paper, the maximum field is 1.6, the minimum field is 0.3, the maximum learning probability is 0.15, and the minimum learning probability is 0.02, including a total of 32,000 network learning adjustments [22].

3.2. Data Mining Model Based on AKN Algorithm. According to the process of data mining technology, the data are preprocessed first. Since each connection in the dataset is described by multiple characteristics, there are numbers and character types to describe different attacks with multiple characteristics. For text data, the dataset must be preprocessed because it must be converted to a digital format before it can be recognized by the Kohonen network. Table 2 shows the numeric codes corresponding to the two-dimensional character data, and Table 3 shows the numeric codes corresponding to the preprocessing results of the three-dimensional character data [23].

In order to compare the performance of the algorithm in this paper, two methods are selected for experimental testing. In this paper, the false alarm rate, the correct rate, and the time used are used to evaluate the detection performance of the network data anomaly detection method, as shown in formulas (18) and (19):

$$E_r = \frac{E}{S}, \quad (18)$$

$$Y_r = \frac{T}{S}. \quad (19)$$

Among them, E_r is the false alarm rate, Y_r is the correct rate, E is the number of falsely reported test samples, T is the number of detected test samples, and S is the total number of test samples.

The processed training data are used as input data for AKN and traditional KN calculations to perform training. When the network training is completed, the processed test data are sent to the AKN and KN network environment for calculation. The resulting Dos test set and DoS detection are shown in Figure 5 [24].

It can be seen from Figure 5 that among the 30 test samples in the statistical test, the KN algorithm did not match the test results in the 18th and 23rd types of tests, and there were two errors. However, the AKN algorithm has only one error in the 23rd category, indicating that the prediction error of AKN is smaller than that of KN, and the prediction is more accurate. The AKN algorithm prediction accuracy rate is about 97%, while the KN algorithm is about 91%. Next, we conduct 10 sets of model tests on the two algorithms to calculate their calculation time and accuracy. The test results are shown in Figures 6 and 7 [25].

It can be seen from Figure 6 that the test accuracy of the AKN algorithm is generally higher than that of the KN algorithm, and the lowest and highest accuracy of the AKN algorithm are both higher than the KN algorithm, indicating that its detection stability and accuracy are better than the KN algorithm. It can be seen from Figure 7 that in these 10 sets of test experiments, the running time of the AKN

TABLE 2: Numerical coding for second dimensional character data.

Agreement type	Icmp	Tcp	Udp	Other agreements
Coding	1	2	3	4

TABLE 3: Numerical coding for third dimensional character data.

Internet service	Login	Hostnames	Ftp	Finger	ecr_i	Time	Systat
Coding	0	2	5	8	10	11	15
Internet service	Private	Netstat	domain_u	telnet	http	ftp_data	Other
Coding	18	20	21	23	25	28	30

FIGURE 5: Comparison of prediction accuracy between AKN and KN algorithms.

algorithm is also better than that of the KN algorithm. In general, the AKN algorithm is superior to the traditional KN algorithm in terms of detection stability, accuracy, and time efficiency [26].

3.3. Investigation of the Research Status of Data Mining Technology. At present, data mining technology is widely used in medicine, finance, network security, and

telecommunications. In terms of Dos attacks, which are relatively common in network security, data mining techniques have also been studied and applied. According to the survey of this technology, classification and statistics are respectively carried out from the researched literature, as shown in Table 4.

As can be seen from Table 4, among the 6,389 articles on data mining technology surveyed, 2,910 articles are about network security research, of which 1,235 articles are about

FIGURE 6: Comparison of AKN and KN algorithm test accuracy.

FIGURE 7: Comparison of AKN and KN algorithm test running time consumption.

TABLE 4: Survey of articles on data mining technology.

Total articles	6,389		
Number of articles on cyber security	2,910		
Number of articles on Dos attack	1,235		
Types of data mining technology	Classification algorithm	Regression analysis	Cluster analysis
Number of articles	243	50	430
Types of data mining technology	Association rules	Neural network algorithm	Web data mining
Number of articles	103	313	96

Dos attacks. In addition, 430 papers on clustering analysis, 313 papers on neural network algorithms, and 243 papers on classification algorithms, respectively, account for the top three in the number of documents. It shows that these three types of data mining technologies have relatively more research applications in network security. In fact, the

application of big data mining technology is not the application of a single algorithm. It is often a combination of multiple algorithms before data mining. It abstracts the process of data mining to form a universal data mining method, and compared with the traditional single method, it has higher availability, reproducibility, and versatility [27].

FIGURE 8: Dos test results.

3.4. Data Mining Technology and Dos Attack Test Based on AKN Algorithm. The data mining technology based on the AKN algorithm not only uses the self-organizing learning ability of the neural network but also combines the characteristics of cluster analysis. Conducive to the Dos attack test based on the data mining technology of the AKN algorithm, the experiment collected 16,000 sample sets and performed data statistics on different detection categories. The obtained test set and detection results are shown in Figure 8 [28].

It can be seen from Figure 8 that after self-organizing training and detection of the AKN algorithm, the data samples are gathered into 4 clusters. The data samples are clustered into 5 clusters. The five clusters are distributed in a specific detection range; cluster 1 is distributed in the range of codes 100 to 150, cluster 2 is in codes 50 to 100, cluster 3 is in codes 150 to 200, cluster 4 is within the code 150 to 200, and cluster five is within the code 200 to 250; however, each cluster has a small amount that has not been calculated. In this regard, the difference between the predicted category and the calculated result category is also counted, as shown in Figure 9 [29].

It can be seen from Figure 10 that in the predicted two clustering categories, 300 detection data were predicted, respectively, and the detection error of category one was 9, and the accuracy rate reached 97%, and the detection error

FIGURE 9: Comparison of AKN algorithm prediction category and actual category.

of category 2 is 8, and the accuracy rate reaches 97.33%. It can be seen from Figure 10 that the detection accuracy of the four clusters of the AKN algorithm is higher than that of the KN algorithm, which is about 6% higher than the traditional KN algorithm [30–32]. This shows that the detection accuracy of the AKN algorithm is higher, and it can more effectively defend against Dos attacks [33–36].

4. Discussion

This article first studies the generation, characteristics, types, and principles of Dos attacks, and then studies the current research status of data mining technology on denial of service attacks, and understands the characteristics, processes, and principles of data mining technology. Then, the source, principle, content, and steps of the AKN algorithm are studied, and the AKN algorithm model and related experiments are established and analyzed.

FIGURE 10: Comparison of accuracy rates of cluster detection results.

In the research process, this article also draws on a lot of relevant documents and understands that more and more scholars are doing research on network security, and research on typical denial of service attacks is in full swing. As a neural network algorithm that can automatically adapt and organize learning, it is especially suitable for this kind of data and network complex environment, and it is also more systematic, and the superiority of the algorithm can also be explained by the experimental results.

In the experimental part, this article collects data samples, uses the AKN algorithm steps and formulas, illustrates the diagrams, and compares the experimental data with the traditional neural network algorithm to obtain the superiority of the AKN algorithm. As a result, the detection accuracy of the AKN algorithm is higher, and it can defend against Dos attacks more quickly and effectively.

5. Conclusions

After establishing the AKN algorithm model, this paper detects and analyzes the data samples of the Dos attack, compares it with the traditional KN algorithm, and finally concludes that the computer data mining technology based on the AKN algorithm is used in the defense and detection of denial of service attacks. The detection accuracy rate is as high as 97%, while the traditional KN algorithm is only 91%, and the detection accuracy rate is increased by at least 6 percentage points. And its detection running time is less than the traditional KN algorithm, and the average detection performance is increased by about 30% when detecting multiple clusters, and the average detection performance is increased by about 20% when detecting a few clusters. Therefore, we can be sure that the use of computer data mining technology based on the AKN algorithm can perform faster, more efficient, and accurate defense detection under Dos attacks.

Data Availability

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Conflicts of Interest

The author declares no conflicts of interest.

References

- [1] E. Adi, Z. A. Baig, P. Hingston, and C.-P. Lam, "Distributed denial-of-service attacks against HTTP/2 services," *Cluster Computing*, vol. 19, no. 1, pp. 79–86, 2016.
- [2] S. Y. Chang and Y. C. Hu, "SecureMAC: securing wireless medium access control against insider denial-of-service attacks," *IEEE Transactions on Mobile Computing*, vol. 16, no. 12, 2017.
- [3] D. Douglas, J. J. Santanna, D. Ricardo, and L. Z. Granville, "Booters: can anything justify distributed denial-of-service (DDoS) attacks for hire?" *Journal of Information, Communication and Ethics in Society*, vol. 15, no. 1, pp. 90–104, 2017.
- [4] H. Pan, H. Wang, Y. Wang, and H. Huang, "Rules of acupoint selection for diabetic peripheral neuropathy based on data mining technology," *Zhongguo Zhen Jiu=Chinese Acupuncture & Moxibustion*, vol. 36, no. 10, pp. 1111–1114, 2016.
- [5] W. Zhao-Yi, H. Zheng-De, Y. Ping, R. Ting, and L. Xin-Hui, "Regularity of wind-dispelling medication prescribed by li dong-yuan: a data mining technology-based study," *Digital Chinese Medicine*, vol. 3, no. 1, pp. 20–33, 2020.
- [6] Z. Hu, Y. V. Bodyanskiy, Y. Bodyanskiy, O. K. Tyshchenko, and O. Boiko, "An ensemble of adaptive neuro-fuzzy kohonen networks for online data stream fuzzy clustering," *International Journal of Modern Education and Computer Science*, vol. 8, no. 5, pp. 12–18, 2016.
- [7] A. Durrani, S. Latif, R. Latif, and H. Abbas, "Detection of denial of service (DoS) attack in vehicular ad hoc networks: a systematic literature review," *Ad Hoc & Sensor Wireless Networks*, vol. 42, no. 1-2, pp. 35–61, 2018.
- [8] H. Abbas, R. Latif, S. Latif, and A. Masood, "Performance evaluation of enhanced very fast decision tree (EVFDT) mechanism for distributed denial-of-service attack detection in health care systems," *Annals of Telecommunications*, vol. 71, no. 9, pp. 477–487, 2016.
- [9] H. Lu, R. Setiono, and H. Liu, "Effective data mining using neural networks," *IEEE Transactions on Knowledge & Data Engineering*, vol. 8, no. 6, pp. 957–961, 2016.
- [10] C. Helma, T. Cramer, S. Kramer, and L. Raedt, "Data mining and machine learning techniques for the identification of mutagenicity inducing substructures and structure activity relationships of noncongeneric compounds," *Journal of Chemical Information and Computer Sciences*, vol. 35, no. 4, pp. 1402–1411, 2018.
- [11] A. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153–1176, 2017.
- [12] D. A. Adeniyi, Z. Wei, and Y. Yongquan, "Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method," *Applied Computing and Informatics*, vol. 12, no. 1, pp. 90–108, 2016.
- [13] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, "Machine learning and data

- mining methods in diabetes research,” *Computational and Structural Biotechnology Journal*, vol. 15, pp. 104–116, 2017.
- [14] X. Yan and L. Zheng, “Fundamental analysis and the cross-section of stock returns: a data-mining approach,” *Review of Financial Studies*, vol. 30, no. 4, pp. 1382–1423, 2017.
 - [15] L. Xu, C. Jiang, J. Wang, J. Yuan, and Y. Ren, “Information security in big data: privacy and data mining,” *IEEE Access*, vol. 2, no. 2, pp. 1149–1176, 2017.
 - [16] S. García, J. Luengo, and F. Herrera, “Tutorial on practical tips of the most influential data preprocessing algorithms in data mining,” *Knowledge-Based Systems*, vol. 98, pp. 1–29, 2016.
 - [17] H. Hong, H. R. Pourghasemi, and Z. S. Pourtaghi, “Landslide susceptibility assessment in Lianhua County (China): a comparison between a random forest data mining technique and bivariate and multivariate statistical models,” *Geomorphology*, vol. 259, pp. 105–118, 2016.
 - [18] P. A. Flach, “Encyclopedia of machine learning and data mining,” *Journal of Shijiazhuang Vocational Technology Institute*, vol. 2, no. 2, pp. 110–114, 2016.
 - [19] Z. Ge, Z. Song, S. X. Ding, and B. Huang, “Data mining and analytics in the process industry: the role of machine learning,” *IEEE Access*, vol. 5, 2017.
 - [20] T. Emoto, T. Yamashita, T. Kobayashi et al., “Characterization of gut microbiota profiles in coronary artery disease patients using data mining analysis of terminal restriction fragment length polymorphism: gut microbiota could be a diagnostic marker of coronary artery disease,” *Heart and Vessels*, vol. 32, no. 1, pp. 39–46, 2017.
 - [21] H. Hong, P. Tsangaratos, I. Ilia, J. Liu, A.-X. Zhu, and W. Chen, “Application of fuzzy weight of evidence and data mining techniques in construction of flood susceptibility map of Poyang County, China,” *The Science of the Total Environment*, vol. 625, pp. 575–588, 2018.
 - [22] M. Rich, “Denial of service,” *CQ Amateur Radio*, vol. 72, no. 12, pp. 8–48, 2016.
 - [23] V. S. Dolk, P. Tesi, C. D. Persis, and W. P. M. H. Heemels, “Event-triggered control systems under denial-of-service attacks,” *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, 2016.
 - [24] H. Luo, Z. Chen, J. Li, and A. V. Vasilakos, “Preventing distributed denial-of-service flooding attacks with dynamic path identifiers,” *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 8, pp. 1801–1815, 2017.
 - [25] X. Zhong, I. Jayawardene, G. K. Venayagamoorthy, and R. Brooks, “Denial of service attack on tie-line bias control in a power system with PV plant,” *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 1, no. 5, pp. 375–390, 2017.
 - [26] S. Khan, A. Gani, A. Wahab, and P. Singh, “Feature selection of denial-of-service attacks using entropy and granular computing,” *Arabian Journal for Science and Engineering*, vol. 43, no. 9, pp. 1–10, 2017.
 - [27] T. Marwala, V. M. Kuthadi, and R. Selvaraj, “Ant-based distributed denial of service detection technique using roaming virtual honeypots,” *IET Communications*, vol. 10, no. 8, pp. 929–935, 2016.
 - [28] Y. Wang, Y. Zhang, X. Hei, W. Ji, and W. Ma, “Game strategies for distributed denial of service defense in the cloud of things,” *Journal of Communications and Information Networks*, vol. 1, no. 4, pp. 143–155, 2016.
 - [29] T. Boraten and A. Kodi, “Mitigation of hardware trojan based denial-of-service attack for secure NoCs,” *Journal of Parallel and Distributed Computing*, vol. 111, pp. 24–38, 2017.
 - [30] Y. Zeng, G. Chen, K. Li, Y. Zhou, X. Zhou, and K. Li, “M-skyline: taking sunk cost and alternative recommendation in consideration for skyline query on uncertain data,” *Knowledge-Based Systems*, vol. 163, pp. 204–213, 2019.
 - [31] Y. Chen, Y. Ping, Z. Zhang, B. Wang, and S. He, “Privacy-preserving image multi-classification deep learning model in robot system of industrial IoT,” *Neural Computing and Applications*, vol. 33, no. 10, pp. 4677–4694, 2021.
 - [32] B. Bo Rong, Y. Yi Qian, K. Kejie Lu, H. H. Hsiao-Hwa Chen, and M. Guizani, “Call admission control optimization in WiMAX networks,” *IEEE Transactions on Vehicular Technology*, vol. 57, no. 4, pp. 2509–2522, 2008.
 - [33] L. Li, H. Zhang, Y. Xia, and H. Yang, “Security estimation under denial-of-service attack with energy constraint,” *Neurocomputing*, vol. 292, pp. 111–120, 2018.
 - [34] J. Chen and C. Yu, “A WebGIS-based system for urban stormwater risk analysis using a cloud matter-element model,” *International Journal of Intelligent Information Technologies*, vol. 16, no. 3, pp. 80–99, 2020.
 - [35] Z. Xu, G. Zhu, N. Metawa, and Q. Zhou, “Machine learning based customer meta-combination brand equity analysis for marketing behavior evaluation,” *Information Processing & Management*, vol. 59, no. 1, Article ID 102800, 2022.
 - [36] L. Yin, X. Li, L. Gao, C. Lu, and Z. Zhang, “A novel mathematical model and multi-objective method for the low-carbon flexible job shop scheduling problem,” *Sustainable Computing: Informatics and Systems*, vol. 13, no. 3, pp. 15–30, 2017.