Hindawi
Journal of Sensors
Volume 2021, Article ID 5558860, 11 pages
https://doi.org/10.1155/2021/5558860

Research Article

An Exhaustive Research on the Application of Intrusion Detection Technology in Computer Network Security in Sensor Networks

Yajing Wang,[1] **Juan Ma**,[1] **Ashutosh Sharma**,[2] **Pradeep Kumar Singh**,[3] **Gurjot Singh Gaba**,[4] **Mehedi Masud**,[5] **and Mohammed Baz**[6]

[1]*Internet of Things Technology Department, Shanxi Vocational &Technical College of Finance & Trade, Taiyuan, 030031 Shanxi, China*
[2]*Institute of Computer Technology and Information Security, Southern Federal University, Russia*
[3]*Department of CSE, ABES Engineering College, Ghaziabad, Uttar Pradesh, India*
[4]*School of Electronics and Electrical Engineering, Lovely Professional University, Phagwara, Punjab 144411, India*
[5]*Department of Computer Science, College of Computers and Information Technology, Taif University, P. O. Box 11099, Taif 21944, Saudi Arabia*
[6]*Department of Computer Engineering, College of Computer and Information Technology, Taif University, PO Box. 11099, Taif 21994, Saudi Arabia*

Correspondence should be addressed to Yajing Wang; yajingwang4@gmail.com and Mehedi Masud; mmasud@tu.edu.sa

Received 17 February 2021; Revised 7 May 2021; Accepted 13 May 2021; Published 29 May 2021

Academic Editor: Omprakash Kaiwartya

Intrusion detection is crucial in computer network security issues; therefore, this work is aimed at maximizing network security protection and its improvement by proposing various preventive techniques. Outlier detection and semisupervised clustering algorithms based on shared nearest neighbors are proposed in this work to address intrusion detection by converting it into a problem of mining outliers using the network behavior dataset. The algorithm uses shared nearest neighbors as similarity, judges whether it is an outlier according to the number of nearest neighbors of a data point, and performs semisupervised clustering on the dataset where outliers are deleted. In the process of semisupervised clustering, vast prior knowledge is added, and the dataset is clustered according to the principle of graph segmentation. The novelty of the proposed algorithm lies in outlier detection while effectively avoiding the dependence on parameters, thus eliminating the influence of outliers on clustering. This article uses real datasets: lypmphography and glass for simulation purposes. The simulation results show that the algorithm proposed in this paper can effectively detect outliers and has a good clustering effect. Furthermore, the experimentation reveals that the outlier detection-based SCA-SNN algorithm has the best practical effect on the dataset without outliers, clearly validating the clustering performance of the outlier detection-based SCA-SNN algorithm. Furthermore, compared to the other state-of-the-art anomaly detection method, it was revealed that the anomaly detection technology based on outlier mining does not require a training process. Thus, they overcome the current anomaly detection problems caused due to incomplete normal patterns in training samples.

1. Introduction

With the widespread advancement in the Internet and online platforms, network security requirements have also become inevitable [1, 2]. Various threats related to computer network security can be seen nowadays, like software bugs and intrusions. These bugs appear due to the large functionality and large size of the software or the operating system. The intruders who do not have access to this data may steal useful private information against the consent of the network users. However, the firewalls are placed in between two or more computers dedicated to isolating these networks based on determining rules or policies. But these firewalls are not enough to be secured from such types of attacks. This is the scenario where intrusion detection systems play a vital role in stopping the cyber attacks and analyze the security problems at the time of such intrusions so that these situations can be tackled in the future [3–5]. The intrusion detection systems collect the computer network information to track the possibility of attacks or misuses against ethical concerns [6, 7]. There are several types of network data concerns that fall into the category to be protected by intrusion detection, like network traffic data, system status files, and system-level test data [8–10]. There exist various applications of network intrusion detection systems which are depicted in Figure 1.

The network traffic processing application can convert the traffic into various network parameter patterns, helpful in management. The prevention system is liable to detect the threats, and threat classification is done utilizing signature matching that is designated to match the input against the already present pattern. The other applications include threat reporting and anomaly detection that detects the traffic signatures [11, 12].

With the rapid development and application of computer network technology and the increasing number of computer network users, ensuring the security of information on the network has become a key technology of computer networks [13–15]. However, various security mechanisms have been developed to protect computer networks, such as user authorization and authentication, access control, data encryption, and data backup. But the above security mechanisms can no longer meet the current network security needs [16]. Network intrusions and attacks are still not uncommon. Therefore, intrusion detection is one of the key technologies that emerged in information and network security assurance. Introducing intrusion detection technology is equivalent to introducing a closed-loop security strategy [17, 18] into the computer system.

This article addresses intrusion detection by converting it into a problem of mining outliers using the network behavior dataset. A preventive technique for intrusion protection of computer network security is proposed to detect the outliers using the semisupervised clustering algorithms based on shared nearest neighbors. The nearest neighbor similarity criteria are used in this work to judge the outlier according to the number of nearest neighbors of a data point, and on this basis, semisupervised clustering is performed for deleting the outliers. The novelty of the proposed algorithm lies in outlier detection while effectively avoiding the dependence on parameters, thus eliminating the influence of outliers on clustering. This work used the real dataset for simulation and compared it with the other anomaly detection technologies. It was revealed that the anomaly detection technology based on outlier mining does not require a training process. This overcomes the current anomaly detection problems caused due to incomplete normal patterns in training samples. Furthermore, the proposed algorithm effectively detects outliers and provides good clustering outcomes based on the similarity.

The rest of this article is arranged as follows: Section 2 presents the state-of-the-art literature review followed by the research methods depicted in Section 3. Section 4 provides the results and discussion part of the experimental analysis done for the two datasets, followed by the concluding remarks in Section 5.

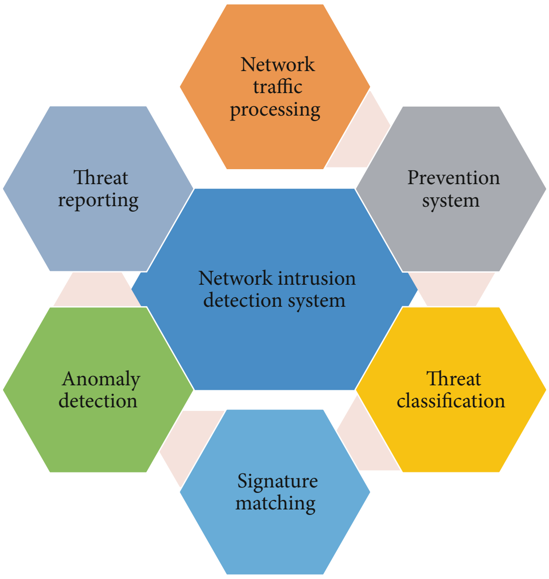

Figure 1: Applications of network intrusion detection system.

2. Literature Review

Domestic research on intrusion detection technology and methods started relatively late, but with the in-depth exploration of universities, scientific research institutes, and enterprises, the development is very rapid, and many new detection theories and results have been produced. The current research on intrusion detection technology mainly covers neural networks, data mining, support vector machines, artificial immunity, etc., involving smart grids, industrial infrastructure, industrial networks, and other related fields [19–22].

Sun et al. proposed an improved method of cascading transmission edges. Using the character interval, the character interval can be used to represent several consecutive characters, which can effectively reduce the number of transmission edges. In addition, the two methods before and after the improvement were compared through comparative experiments. The results show that the number of transmission edges can be reduced to 10% of that before the improvement, thereby increasing the efficiency of deep packet inspection [23].

Haojie et al. analyzed the potential security threats of 5G in-vehicle networks and focused on intrusion detection methods for in-vehicle networks. Four experimental scenarios were selected from potential attacks on the vehicle network, and real car data were collected to compile various attack databases for the first time. In order to find the appropriate method to identify different attacks, four lightweight intrusion detection methods are proposed to identify the abnormal behavior of the vehicle network. In addition, the research carried out a comparison of the detection performance between the four detection methods with the consideration of comprehensive evaluation indicators. The evaluation results provide the best lightweight detection solution for the vehicle network. This article helps to understand the advantages of test methods in the detection performance of in-vehicle networks. Furthermore, it promotes the application of detection technologies to safety issues in the automotive industry [24].

Zhang et al. took intrusion detection system (IDS) as the research object, established an IDS model based on data mining, obtained experimental results, and drew relevant experimental conclusions. At the same time, it was compared with traditional IDS, and six experiments were carried out. As a result, the detection rate, false-negative rate, and false-positive rate of two different methods in six experiments were obtained. The experiment concludes that the intrusion detection system using data mining has better network protection and security performance, and the detection ability of network vulnerability intrusion is stronger. Thus, this research provides a new way to detect and research network protection security loopholes [25].

Kumar et al. proposed a model in which a set of training examples obtained by using a network analyzer (i.e., Wireshark) can be used to construct an HMM. Since it is not an intrusion detection system, the obtained file trace can be used as a training example to test the HMM model. It also predicts the probability value of each test sequence and indicates whether the sequence is abnormal. This article also shows a numerical example; the example calculates the best observation sequence for the HMM and state sequence probability [26].

The innovation of this paper is that the problem of intrusion detection can be converted into the problem of mining outliers in the network behavior dataset. Compared with other anomaly detection technologies, the anomaly detection technology based on outlier mining does not require a training process, which overcomes the current anomaly detection faced with the problem of high false alarm rate caused by incomplete normal patterns in training samples. This paper describes the outlier mining algorithm based on the similarity.

3. Research Methods

3.1. Classification of Intrusion Detection. Through the research of existing intrusion detection technology methods, intrusion detection technology can be classified from different angles:

(1) According to the source of detection data, there are three categories: host-based intrusion detection technology, network-based intrusion detection technology, and host- and network-based intrusion detection technology. The above three intrusion detection technologies all have their own advantages and disadvantages and can complement each other. However, a complete intrusion detection system must be distributed based on both the host and the network

(2) According to the detection technology: divided into anomaly detection technology and misuse detection technology. Anomaly detection technology can also be called behavior-based intrusion detection technology, which assumes that all intrusions have abnormal characteristics. On the other hand, misuse detection technology, also known as knowledge-based intrusion detection technology, expresses intrusion behavior in attack mode and attack signature

(3) According to the working method: it can be divided into offline detection and online detection. Offline detection: it is a non-real-time system that analyzes audit events after the event and checks for intrusions. Online detection: real-time online detection system, which includes real-time network data packet analysis and real-time host audit analysis

(4) The system network architecture is divided into centralized detection technology, distributed detection technology, and layered detection technology. The analysis result is transmitted to the adjacent upper layer, and the detection system of the higher layer only analyzes the analysis result of the next layer. In addition, the hierarchical detection system makes the system more scalable by analyzing the hierarchical data [27–30].

3.2. Intrusion Detection System and Working Principle. An intrusion detection system refers to the system used to detect various intrusion behaviors. It is an important part of the network security system. By monitoring the operation status of the network and computer system, various attack attempts, attack behaviors, or attack results are found. And then promptly issue an alarm or make a corresponding response to ensure the confidentiality, integrity, and availability of system resources. Intrusion detection systems have been widely used and researched as an important means to resist network intrusion attacks [31, 32]. The basic intrusion detection system for computer network security is depicted in Figure 2.

The intrusion detection system is a typical “snooping device.” It does not bridge multiple physical network segments (usually only one listening port). It does not need to forward any traffic, but only needs to passively and silently collect the messages it cares about on the network. Based on the collected messages, the intrusion detection system extracts the corresponding traffic statistical characteristic values. It uses the built-in intrusion knowledge base to perform intelligent analysis and comparison with these traffic characteristics [33, 34]. According to the preset threshold, the message traffic with higher matching coupling will be considered an offense. The intrusion detection system will wake up and alarm or carry out a limited counterattack according to the corresponding configuration. The principle of intrusion detection is shown in Figure 3.

The workflow of an intrusion detection system is roughly divided into the following steps:

(1) Information collection. The first of intrusion detection is information collection, which includes the content of network traffic, the status, and behavior of user connection activities

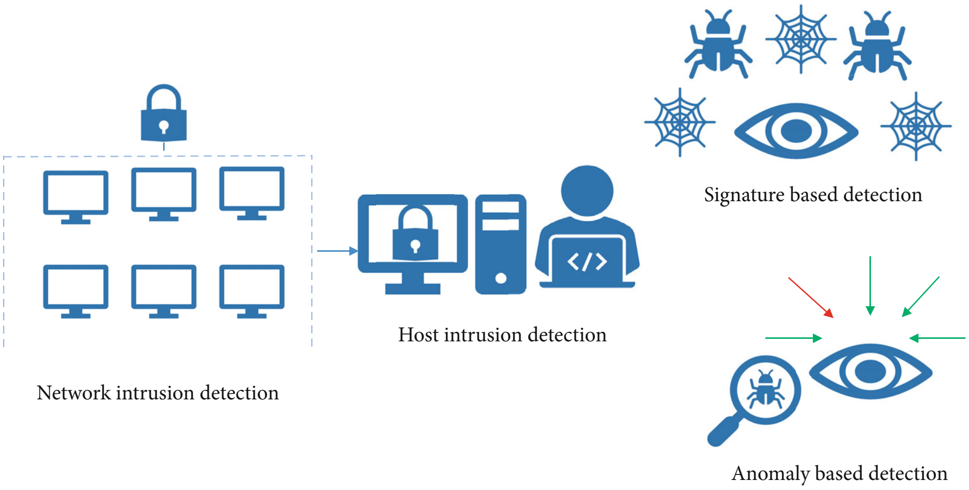

Figure 2: Basic intrusion detection system for computer network security.

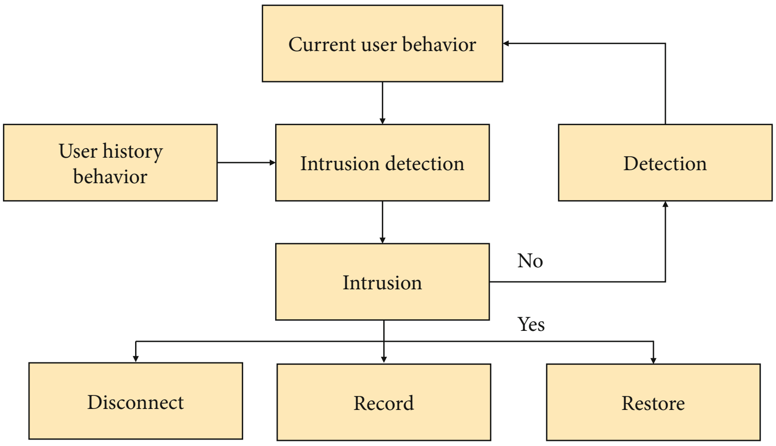

Figure 3: Intrusion detection principle.

(2) Signal analysis. The information collected above is generally analyzed by three technical means: pattern matching, statistical analysis, and completeness analysis. The first two methods are used for real-time intrusion detection, while integrity analysis is used for postmortem analysis

(3) Real-time recording, alarm, or limited counterattack. The fundamental task of IDS is to make appropriate responses to intrusions. These responses include detailed log records, real-time alarms, and limited counterattack sources. The only technical methods to identify intrusions are user characteristics, intruder characteristics, and activity-based. The structure of the intrusion detection system is shown in Figure 4

3.3. Intrusion Detection Technology Methods. At present, there are many standard intrusion detection technology methods, and a few are listed below for explanation.

(1) *Neural network anomaly detection.* This method can be self-learning and self-adaptable to user behavior and can effectively process and judge the possibility of intrusion according to the actual monitored information. The prediction of the error rate of the next event reflects the abnormal degree of user behavior to a certain extent. At present, this method is widely used, but the method is not yet mature, and there is no more complete product [35–38]

(2) *Probabilistic statistical anomaly detection.* This method is based on the modeling of historical user behavior, and based on early evidence or models, the audit system detects the user's use of the system in real time, according to the user behavior probability stored in the system. The statistical model is used to detect, and when suspicious user behavior is found, it keeps track and monitors and records the user's behavior

(3) *Expert system misuse detection.* Aiming at characteristic intrusion behaviors, expert systems are often used for detection. In the realization of the expert detection system, the knowledge of the safety expert is expressed through the rules of the If-Then structure (or a compound structure). Therefore,

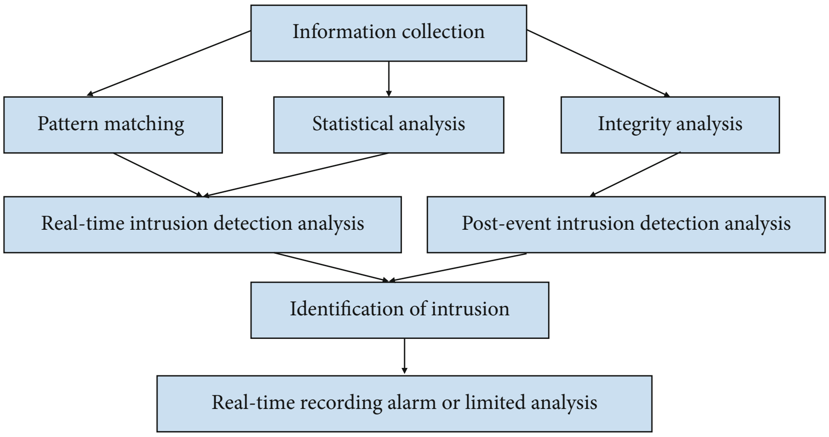
Information collection
Pattern matching
Statistical analysis
Integrity analysis
Real-time intrusion detection analysis
Post-event intrusion detection analysis
Identification of intrusion
Real-time recording alarm or limited analysis

Figure 4: Intrusion detection system structure.

establishing an expert system depends on the completeness of the knowledge base, which depends on the completeness and real-time nature of audit records

(4) *Model-based intrusion detection.* Intruders often use specific behavioral procedures when attacking a system, such as the behavioral sequence of guessing passwords. This behavioral sequence constitutes a model with specific behavioral characteristics. According to the attack represented by this model, the behavioral characteristics of intention can detect malicious attack attempts in real time

Invasion technology has undergone rapid changes in scale and method, and intrusion methods and techniques have also progressed and developed. Outlier mining is an important direction of research on intrusion detection technology. Outlier mining is to mine a small part of abnormal data from a large amount of complex data, which is novel and significantly different from conventional data patterns. Outlier mining is often anomalous data mixed in a large amount of high-dimensional data, and these anomalous data will bring serious consequences. Currently, in the field of intrusion detection research, many scholars apply cluster analysis to anomaly detection. But through the analysis of the characteristics of intrusions, it can be considered that outlier mining technology is more suitable for anomaly-based intrusion detection than clustering technology. Because there is a clear difference between normal behavior and abnormal behavior, and in real applications, the number of abnormal behaviors is much lower than the number of normal behaviors [39–44]. Compared with the entire network behavior, the intrusion behavior is a small number of abnormal data, which can be treated as an isolated point in the dataset, which can better reflect the nature of the intrusion. Therefore, intrusion detection can be converted into the problem of mining outliers in the network behavior dataset. Compared with other anomaly detection technologies, the anomaly detection technology based on outlier mining does not require a training process. Therefore, it overcomes the current anomaly detection problems. They are faced with the problem of a high false alarm rate caused by incomplete normal patterns in training samples.

3.4. Steps Involved in Proposed Intrusion Detection Algorithm.

The outlier mining algorithm proposed in this article is based on the similarity index described in the following steps.

Step 1. Enter the dataset: A matrix with n rows and m columns indicates that each record in the original network record set of n intrusion detection has m characteristic attributes. Suppose the domain $X = \{x_1, x_2, \cdots, x_n\}$ is the object to be detected, and each object has m indicators, namely, $x_i = \{x_{i1}, x_{i2}, \cdots, x_{im}\}$, $i = (1, 2, \cdots, n)$, expressed as a data matrix:

$$X = \begin{pmatrix} x_{11} & K & x_{1m} \\ M & O & M \\ x_{n1} & L & x_{nm} \end{pmatrix}. \tag{1}$$

Step 2. Find the set of isolated points in n objects: in order to judge the degree of dispersion of each object in x, first calculate the similarity coefficient r_{ij} between each object pair and form a similarity coefficient matrix, namely,

$$R = \begin{pmatrix} r_{11} & K & r_{1n} \\ M & O & M \\ r_{n1} & L & r_{nn} \end{pmatrix}, \tag{2}$$

$$r = 1 - \sqrt{\frac{1}{n} \sum_{k=1}^{m} \left(x_{ik} - x_{jk}\right)^2}, \tag{3}$$

$$p_i = \sum_{j=1}^{n} r_{ij}. \tag{4}$$

Among them, p_i is the sum of the i^{th} row of the relative coefficient matrix. The smaller the value, the farther the

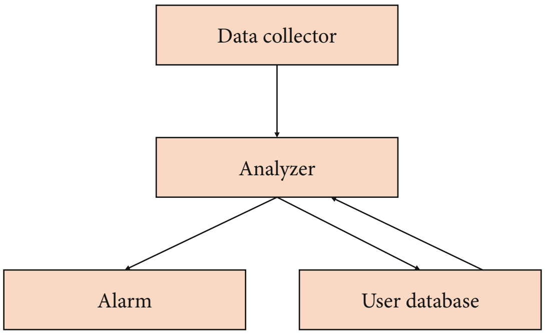

Figure 5: System structure of anomaly intrusion detection based on similarity and isolated point analysis method.

Table 1: Data object distribution of lypmphography dataset.

Classification	Classification	Percentage
General category	Categories 2 and 3	95.9%
Outlier class	Categories 1 and 4	4.1%

object i is from other objects, which is the postoption of the isolated point set.

$$\lambda_i = \frac{p_{\max} - p_i}{p_{\max}} \times 100\%. \quad (5)$$

Among them, λ is the threshold, and objects with $\lambda_i \geq \lambda$ are considered as outliers.

(1) Anomaly intrusion detection system based on similarity and outlier analysis method: the abnormal intrusion detection system is to obtain the audit data of the system and use them as the original data value of the intrusion detection data source and user behavior characteristics, and then use the similarity and outlier analysis method to divide the original data values into normal datasets and isolated datasets. Finally, point the dataset to determine whether it is under attack

(2) The system structure of abnormal intrusion detection is based on similarity and outlier analysis methods. It is composed of collectors, analyzers, alarms, and user databases. The structure diagram is shown in Figure 5

(3) Working principle of the proposed algorithm:

  (i) The data collector mainly collects the original audit data and then transmits the data to the data analyzer

  (ii) The data analyzer has the functions of data transmission and data analysis. On the one hand, it receives the alarm information from the qualitative data and transmits it to the alarm. On the other hand, the data analyzer transmits the quantitative data to the user database

Table 2: Data object distribution of the glass dataset.

Classification	Classification	Percentage
General category	Categories 1, 2, 3, and 7	89.8%
Outlier class	Categories 5 and 6	10.2%

Table 3: Outlier detection results on the lypmphography dataset.

K value	Direct isolation	Derivative outliers	Correct isolation points	Accuracy
8	3	12	4	66.7% (4/6)
12	5	10	4	66.7% (4/6)
16	8	15	6	66.7% (4/6)

Table 4: Outlier detection results on the glass dataset.

K value	Direct isolation	Derivative outliers	Correct isolation points	Accuracy
8	10	24	16	66.7% (16/22)
10	12	28	18	73.7% (18/22)
16	16	33	22	100% (22/22)

  (iii) The user database uses an outlier mining algorithm based on similarity sum to divide quantitative data into the normal dataset and outlier dataset. Then, transfer these two datasets to the analyzer

  (iv) The analyzer receives the alarm information from the outlier dataset, transmits the information to the alarm, and then transmits the normal dataset to the user database

  (v) The user database updates the normally transmitted dataset so that the user database in the intrusion detection system can accurately describe user behavior characteristics. The description of algorithm for intrusion detection

Step 1. Get the original data x_i of the current user's resource usage at a certain moment.

Step 2. Calculate the degree of dispersion of each object in X; that is, calculate the similarity coefficient r_{ij} between each object.

Step 3. Calculate P_i and λ_i in the i^{th} row of the similarity coefficient matrix.

Step 4. If the object with $\lambda i \geq \lambda$ is considered an outlier set, there is abnormal behavior and alarms; otherwise, it belongs to normal user behavior. The user database is updated.

From the perspective of time consumption, it is mainly the comparison of distance. Although the anomaly detection

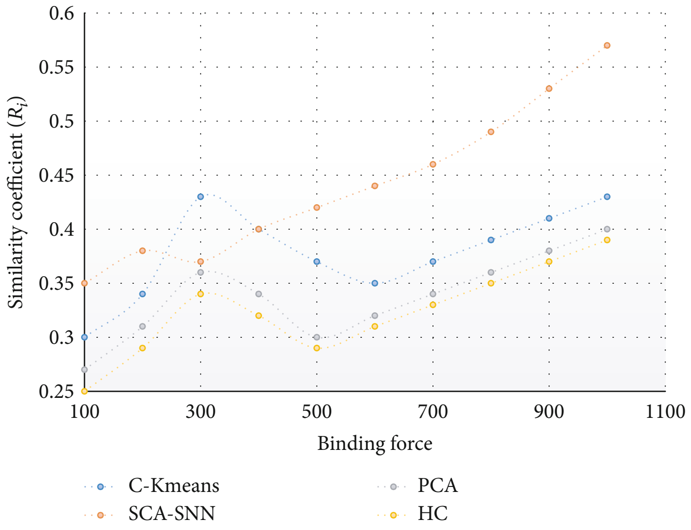

Figure 6: Experimental results on the lypmphography dataset.

technology of outlier mining adds extra time and space consumption than the cluster-based anomaly detection technology, it also improves the algorithm's performance and improves the performance of intrusion attacks and detection rate.

4. Results and Discussion

The experimental datasets in this article are all from UCI real datasets, and the experimental results are the average of data obtained from multiple experiments. The performance judgment of outlier detection is mainly based on analyzing the proportion of correct outliers detected in all outliers, and the evaluation function of semisupervised clustering algorithm is used.

The known number of paired constraints is the initial set of constraints randomly generated by the system. The known constraints are subtracted from the evaluation index because, in the semisupervised clustering algorithm, the known supervision information cannot reflect the effect of the clustering algorithm. The experiment uses the lypmphography dataset and the glass dataset for comparison experiments. The object distribution of the dataset is shown in Tables 1 and 2.

The experimental results of outlier detection are shown in the table. The first column in the table is the K value. The second column indicates the number of isolated points obtained by analyzing the nearest neighbors of the data points; that is, the data points with very few nearest neighbors are direct. It is judged as an isolated point. The second column indicates the number of isolated points obtained from the nearest neighbor set of the isolated point called a derived isolated point. The fourth column refers to the true isolated point among the isolated points obtained in the second column. Finally, the last column is the correct rate of outlier detection.

The experimental results of the lypmphography dataset are shown in Table 3. Since the number of real categories in the dataset is 4, the experiment starts training from $K = 4$. When $K = 8$, we get that the nearest neighbor set of a data point contains very few objects, so we determine it as an outlier and analyze the data points in the nearest neighbor set of the outlier. So, when the K value is 8, we get 12 outliers, including 4 correct outliers. When $K = 12$, although the accuracy rate of outlier detection is not improved from the table above, the number of outliers obtained from analyzing the characteristics of the classes decreases, and some data points that are judged incorrectly are removed.

From this perspective, it is clear that the detection rate has increased by 7%. When $K = 16$, all 6 outliers were detected, and the detection rate reached 100%. The algorithm also has apparent effects on the glass dataset (as shown in Table 4).

The two semisupervised clustering algorithms: C-Kmeans and Sine Cosine Algorithm-based sharing nearest neighbor (SCA-SNN), are evaluated in this study for outlier detection for both the lypmphography and glass dataset. Furthermore, the semisupervised clustering is performed on the "denoising" dataset after detecting the outliers. The experimental results obtained from these methods are also compared with other state-of-the-art methods like hierarchical clustering (HC) and principle component analysis (PCA) to determine the effectiveness of semisupervised clustering. The experimental results are shown in Figure 6–9.

Figure 6 presents four different algorithms for the lypmphography dataset experimental outcomes before finding the outliers and without performing the denoising step. The experimental dataset utilized in Figure 7 is the "denoising" lypmphography dataset, which only contains the second and third types of the original lypmphography dataset. For experimental comparison on this dataset (done in Figure 7), it can be seen that as the number of paired constraints increases, the effect of the SCA-SNN algorithm is steadily increasing among all other algorithms. However, after removing the outliers, the C-Kmeans algorithm also provides relatively stable performance, and there is no significant fluctuation of the clustering results. But from the overall

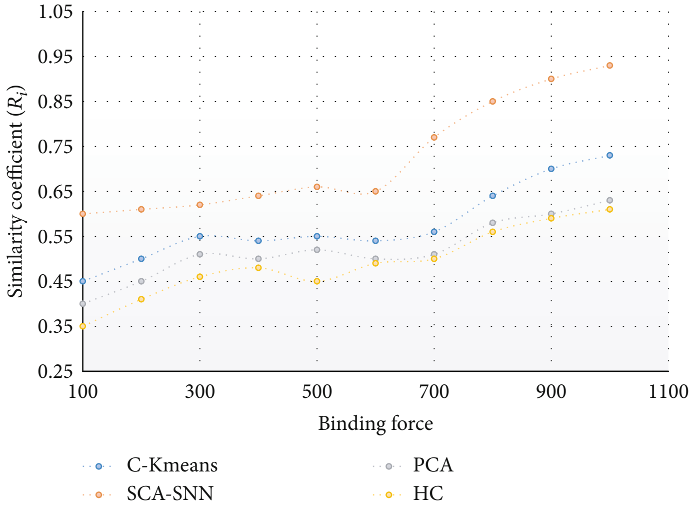

Figure 7: Experimental results of the "denoising" lypmphography dataset.

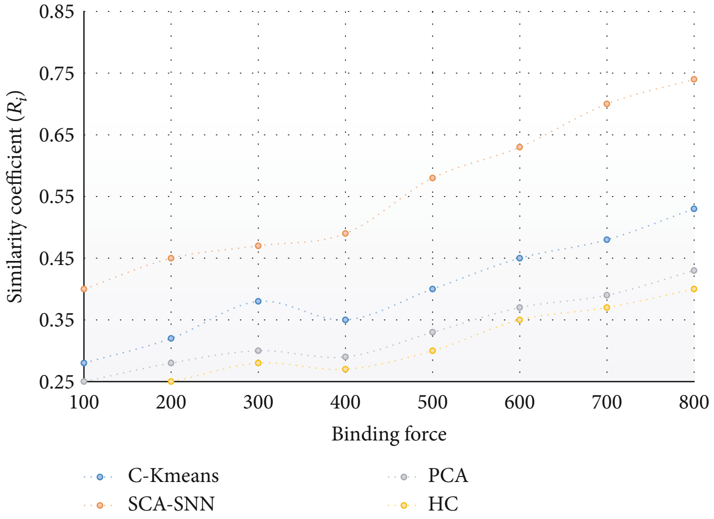

Figure 8: Experimental results on the glass dataset.

clustering results, the performance of the SCA-SNN algorithm is better than the C-Kmeans, PCA, and HC algorithm.

All the algorithms do not have noticeable results on the original lypmphography dataset. Although the experimental results have improved as the number of paired constraints increases when the number of constraints reaches 1000, the correct judgment rate of the C-Kmeans algorithm is only 0.48, and the SCA-SNN algorithm only reaches 0.58, which indicates that the data is concentrated. Furthermore, the outlier data caused a great impact on the clustering results and weakened the guiding role of the paired constraints, resulting in the entire clustering algorithm without good results.

Figures 8 and 9 are the experimental results on the glass dataset. It can be found from Figure 8 that the C-Kmeans algorithm exhibits its instability due to the existence of "noise" data. From the overall perspective of the clustering results, the clustering performance of the SCA-SNN algorithm is always better than that of the C-Kmeans, PCA, and HC algorithm.

Regardless of whether there are outliers in the dataset, the clustering effect of the SCA-SNN algorithm is better than that of the C-Kmeans algorithm and the other state-of-the-art algorithms, especially after removing the outliers. On the set, the SCA-SNN algorithm has better experimental results.

From the above four experimental results, the outlier detection-based SCA-SNN algorithm has the best experimental effect on the dataset without outliers, which shows that the detection of outliers is a crucial process and fully validates the clustering performance of the outlier detection-based SCA-SNN algorithm. In many practical applications, the dataset often contains some outliers. These outliers may contain potentially valuable information. Therefore, mining outliers can effectively improve the performance of clustering

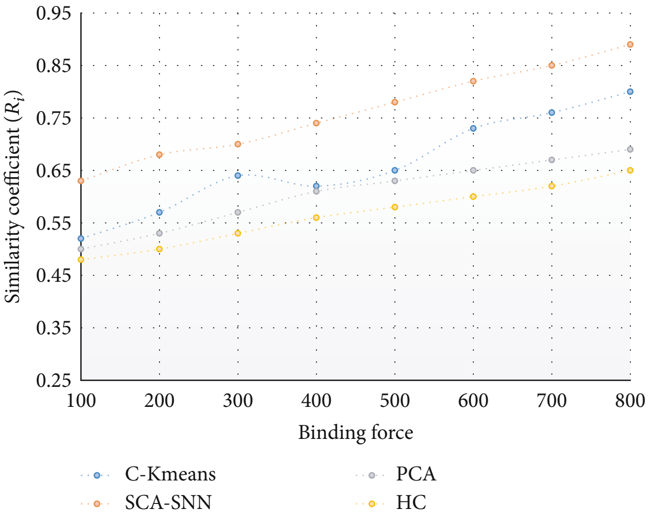

Figure 9: Experimental results of the "denoising" glass dataset.

and get the correct classification. It can also help people obtain more valuable information.

5. Conclusion

This paper proposes an outlier detection and semisupervised clustering algorithm based on nearest neighbor similarity. The wood algorithm uses the C-Kmeans algorithm to train the dataset, which can obtain a reasonable and accurate data sharing nearest neighbor set, and quickly and accurately detect global outliers based on the obtained results, which also has a significant effect on local outliers. The algorithm effectively avoids the insufficient preprocessing of noise points and the influence of inaccurate input parameters on the results. Also, it overcomes the problem of large calculations such as the Jarvis-Patrick algorithm. In the process of semisupervised clustering, the acquired paired prior knowledge is expanded to maximize the guiding effect of prior knowledge. The algorithm detects outliers and effectively avoids the dependence on parameters and eliminates the influence of outliers on clustering. The algorithm combines prior knowledge and expands, making the clustering process "rules to follow." Experiments on real datasets show that the outlier detection algorithm combined with semisupervised clustering results in the best clustering results. Furthermore, the experimentation reveals that the outlier detection-based SCA-SNN algorithm has the best experimental effect on the dataset without outliers. This approach shows that the detection of outliers is crucial and fully validates the clustering performance of the outlier detection-based SCA-SNN algorithm.

With the increasingly prominent network security issues, the research of intrusion detection technology has attracted more and more attention. An intrusion detection algorithm based on outlier data mining is given based on the in-depth study of data mining intrusion detection technology. Outlier mining technology can complete anomaly detection work. When the abnormal data is much smaller than the normal data, the detection result is better than anomaly detection technology based on clustering. In general, the statistical distribution of abnormal and normal behavior in-network data meets the conditions of use of outlier mining. Network security has always been a concern of people. However, with the further development of the network and the diversification of hacker attacks, there is still much research and challenging issues to be solved urgently.

Data Availability

All data has been shared within the manuscript.

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding this study.

Acknowledgments

The authors would like to acknowledge the support of Taif University Researchers Supporting Project number (TURSP-2020/239), Taif University, Taif, Saudi Arabia.

References

[1] M. Masud, G. S. Gaba, S. Alqahtani et al., "A lightweight and robust secure key establishment protocol for Internet of medical things in COVID-19 patients care," *IEEE Internet of Things Journal*, 2021.

[2] M. Masud, M. Alazab, K. Choudhary, and G. S. Gaba, "3P-SAKE: privacy-preserving and physically secured authenticated key establishment protocol for wireless industrial networks," *Computer Communications*, vol. 175, pp. 82–90, 2021.

[3] R. G. Bace, *Intrusion detection*, Sams Publishing, 2000.

[4] K. Scarfone and P. Mell, *Guide to intrusion detection and prevention systems (idps)*, vol. 800, no. 2007, 2007NIST Special Publication, 2007.

[5] G. Rathee, A. Sharma, R. Kumar, F. Ahmad, and R. Iqbal, "A trust management scheme to secure mobile information

centric networks," *Computer Communications*, vol. 151, pp. 66–75, 2020.

[6] M. Poongodi, A. Sharma, V. Vijayakumar et al., "Prediction of the price of Ethereum blockchain cryptocurrency in an industrial finance system," *Computers & Electrical Engineering*, vol. 81, article 106527, 2020.

[7] B. Dayıoğlu, *Use of Passive Network Mapping to Enhange Network Intrusion Detection, [M.S. thesis]*, University Library, Middle East Technical University, Turkey, 2001.

[8] T. Lappas and K. Pelechrinis, *Data Mining Techniques for (Network) Intrusion Detection Systems*, vol. 92521, Department of Computer Science and Engineering UC, Riverside, Riverside CA, 2007.

[9] G. Dhiman, K. K. Singh, M. Soni et al., "MOSOA: a new multi-objective seagull optimization algorithm," *Expert Systems with Applications*, vol. 167, article 114150, 2021.

[10] G. Rathee, A. Sharma, H. Saini, R. Kumar, and R. Iqbal, "A hybrid framework for multimedia data processing in IoT-healthcare using blockchain technology," *Multimedia Tools and Applications*, vol. 79, no. 15-16, article 7835, pp. 9711–9733, 2020.

[11] M. A. Aydın, A. H. Zaim, and K. G. Ceylan, "A hybrid intrusion detection system design for computer network security," *Computers & Electrical Engineering*, vol. 35, no. 3, pp. 517–526, 2009.

[12] V. Singh and S. Puthran, "Intrusion detection system using data mining a review," in *2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC)*, pp. 587–592, Jalgaon, India, 2016.

[13] D. Rathore and A. Jain, "Design hybrid method for intrusion detection using ensemble cluster classification and som network," *International Journal of Advanced Computer Research*, vol. 2, no. 3, pp. 181–186, 2019.

[14] M. Masud, G. S. Gaba, K. Choudhary, R. Alroobaea, and M. S. Hossain, "A robust and lightweight secure access scheme for cloud based E-healthcare services," *Peer-to-Peer Networking and Applications*, pp. 1–15, 2021.

[15] M. Masud, G. S. Gaba, K. Choudhary, M. S. Hossain, M. F. Alhamid, and G. Muhammad, "Lightweight and anonymity-preserving user authentication scheme for IoT-based healthcare," *IEEE Internet of Things Journal*, 2021.

[16] W. Meng, E. Tischhauser, Q. Wang, Y. Wang, and J. Han, "When intrusion detection meets blockchain technology: a review," *IEEE Access*, vol. 6, no. 1, pp. 10179–10188, 2018.

[17] F. Farahnakian and J. Heikkonen, "Anomaly-based intrusion detection using deep neural networks," *International Journal of Digital Content Technology and its Applications*, vol. 12, pp. 70–118, 2018.

[18] T. Qian, Y. Wang, M. Zhang, and J. Liu, "Intrusion detection method based on deep neural network," *Huazhong Keji Daxue Xuebao*, vol. 46, no. 1, pp. 6–10, 2018.

[19] R. Priyadharshini and E. J. Leavline, "Cuckoo optimisation based intrusion detection system for cloud computing," *International Journal of Computer Network and Information Security*, vol. 10, no. 11, pp. 42–49, 2018.

[20] A. U. Makarfi, K. M. Rabie, O. Kaiwartya, X. Li, and R. Kharel, "Physical layer security in vehicular networks with reconfigurable intelligent surfaces," in *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, pp. 1–6, Antwerp, Belgium, 2020.

[21] A. Jayaswal and R. Nahar, "Detecting network intrusion through a deep learning approach," *International Journal of Computer Applications*, vol. 180, no. 14, pp. 15–19, 2018.

[22] S. Kumar, K. Singh, S. Kumar, O. Kaiwartya, Y. Cao, and H. Zhou, "Delimitated anti jammer scheme for Internet of vehicle: machine learning based security approach," *IEEE Access*, vol. 7, pp. 113311–113323, 2019.

[23] R. Sun, L. Shi, C. Yin, and J. Wang, "An improved method in deep packet inspection based on regular expression," *Journal of Supercomputing*, vol. 75, no. 6, pp. 3317–3333, 2019.

[24] H. Ji, Y. Wang, H. Qin, Y. Wang, and H. Li, "Comparative performance evaluation of intrusion detection methods for in-vehicle networks," *IEEE Access*, vol. 6, pp. 37523–37532, 2018.

[25] J. Zhang, "Detection of network protection security vulnerability intrusion based on data mining," *International Journal of Network Security*, vol. 21, no. 6, pp. 979–984, 2019.

[26] P. Narwal, D. Kumar, and S. N. Singh, "A hidden markov model combined with markov games for intrusion detection in cloud," *Journal of Cases on Information Technology*, vol. 21, no. 4, pp. 14–26, 2019.

[27] H. Yao, Q. Wang, L. Wang, P. Zhang, M. Li, and Y. Liu, "An intrusion detection framework based on hybrid multi-level data mining," *International Journal of Parallel Programming*, vol. 47, no. 4, pp. 740–758, 2019.

[28] A. Yang, Y. Zhuansun, C. Liu, J. Li, and C. Zhang, "Design of intrusion detection system for internet of things based on improved bp neural network," *IEEE Access*, vol. 7, pp. 106043–106052, 2019.

[29] S. Pundir, M. Wazid, D. P. Singh, A. K. Das, J. J. P. C. Rodrigues, and Y. Park, "Intrusion detection protocols in wireless sensor networks integrated to Internet of things deployment: survey and future challenges," *IEEE Access*, vol. 8, pp. 3343–3363, 2020.

[30] S. Naseer, Y. Saleem, S. Khalid et al., "Enhanced network anomaly detection based on deep neural networks," *IEEE access*, vol. 6, pp. 48231–48246, 2018.

[31] X. Li, M. Xu, P. Vijayakumar, N. Kumar, and X. Liu, "Detection of low-frequency and multi-stage attacks in industrial Internet of things," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 8820–8831, 2020.

[32] Y. Xun, J. Liu, and Y. Zhang, "Side-channel analysis for intelligent and connected vehicle security: a new perspective," *IEEE Network*, vol. 34, no. 2, pp. 150–157, 2020.

[33] A. Gupta, R. K. Jha, P. Gandotra, and S. Jain, "Bandwidth spoofing and intrusion detection system for multistage 5g wireless communication network," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 1, pp. 618–632, 2018.

[34] H. Yang and F. Wang, "Wireless network intrusion detection based on improved convolutional neural network," *IEEE Access*, vol. 7, pp. 64366–64374, 2019.

[35] M. Poongodi, A. Sharma, M. Hamdi, M. Maode, and N. Chilamkurti, "Smart healthcare in smart cities: wireless patient monitoring system using IoT," *The Journal of Supercomputing*, no. article 3765, pp. 1–26, 2021.

[36] X. Xu, L. Li, and A. Sharma, "Controlling messy errors in virtual reconstruction of random sports image capture points for complex systems," *International journal of system assurance engineering and management*, pp. 1–8, 2021.

[37] G. K. Sodhi, S. Kaur, G. S. Gaba, L. Kansal, A. Sharma, and G. Dhiman, "COVID-19: role of robotics, artificial intelligence,

and machine learning during pandemic," *Current Medical Imaging*, vol. 17, 2021.

[38] Y. Liu, Q. Sun, A. Sharma, A. Sharma, and G. Dhiman, "Line monitoring and identification based on roadmap towards edge computing," *Wireless personal communications*, no. article 8272, pp. 1–24, 2021.

[39] M. Fan and A. Sharma, "Design and implementation of construction cost prediction model based on SVM and LSSVM in industries 4.0," *International Journal of Intelligent Computing and Cybernetics*, vol. 14, no. 2, pp. 145–157, 2021.

[40] H. Sun, M. Fan, and A. Sharma, "Design and implementation of construction prediction and management platform based on building information modelling and three-dimensional simulation technology in industry 4.0," *IET collaborative intelligent manufacturing*, 2021.

[41] X. Ren, C. Li, X. Ma et al., "Design of multi-information fusion based intelligent electrical fire detection system for green buildings," *Sustainability*, vol. 13, no. 6, p. 3405, 2021.

[42] A. Sharma and R. Kumar, "A framework for pre-computated multi-constrained quickest QoS path algorithm," *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, vol. 9, no. 3-6, pp. 73–77, 2017.

[43] M. Poongodi, M. Hamdi, A. Sharma, M. Ma, and P. K. Singh, "DDoS detection mechanism using trust-based evaluation system in VANET," *IEEE Access*, vol. 7, pp. 183532–183544, 2019.

[44] D. Kumar, A. Sharma, R. Kumar, and N. Sharma, "A holistic survey on disaster and disruption in optical communication network," *Recent Advances in Electrical & Electronic Engineering (Formerly Recent Patents on Electrical & Electronic Engineering)*, vol. 13, no. 2, pp. 130–135, 2020.