Eng Int Syst (2023) 5: 337–345
© 2023 CRL Publishing Ltd

Engineering
Intelligent Systems

A Fast Detection System for WSN Node Intrusion on the Basis of Association Rules

Zaiying Liu[1,*] and Younes Menni[2]

[1] *School of General Courses, Shanghai Civil Aviation College, Shanghai 200232, China*
[2] *Department of Engineering and Architecture, University of Parma, Parma 43124, Italy*

In order to ensure the healthy development of a network, information security issues need to be addressed from the outset. The establishment of a rapid intrusion detection system is of great significance for medical security and military protection. In this study, a software-defined WSN (ESD-WSN) algorithm based on association rules is proposed for an intrusion detection system to solve the security problem. Firstly, the CHC-based fuzzy association rule mining algorithm is optimized, and then the association rules-based algorithm is optimized using a software-defined WSN algorithm. The optimization results for the five association rules show that the lift of the data mining algorithm reaches 2.96505, 2.18890, 2.03460, 1.95000 and 1.94337 respectively; when the minimum support and confidence values are about 0.4 and 0.8, the mining performance of the ESD-WSN algorithm is better; when the minimum support is low, the time-consumption of the FPL1 algorithm and the ESD-WSN data mining algorithm in is low; when the number of iterations is 400, the fitness function value of the ESD-WSN algorithm is low. The proposed software-defined WSN algorithm based on association rules performs well when applied to intrusion detection systems, the detection rate of the algorithm is improved, and the number of false alarms is reduced.

Keywords: association rules; WSN; intrusion detection system

1. INTRODUCTION

Currently, Internet-based information processing systems are subject to various threats and challenges. For instance, information about a targeted object can be collected locally through the wireless sensor network (WSN); then, the collected data is processed and, finally, the data is transmitted to the network owner. The data is mainly used for agricultural monitoring, environmental monitoring and smart city development, and in defense domains, which make WSN attractive to security attackers [1]. Internet security is very important for efficient communication, and the security level must be improved to ensure the normal transmission of information. An intrusion detection system (IDS) is an important security tool in network security, which can be used to protect the wireless sensor network services and protect the infrastructure from invisible and unpredictable attacks. However, its detection rate is low, the false positive rate is high, and the computational overhead is high. In response to these problems, researchers have optimized and improved the intrusion detection system.

Al-Daweri et al. [2] demonstrated that adaptive intrusion detection systems are crucial for protecting computer networks, and have proposed an adaptive intrusion detection system that can perform updates to deal with new attacks. The homegenetic ensemble approach was adopted to construct the model.Pacheco et al. [3] introduced an artificial neural network-based method to detect and analyse anomalous behavior. to the researchers implement an adaptive IDS that can confirm the damage time of the fog node and ensure smooth communication through preset processing methods. The test results indicated that this model is relatively reliable, and there are few false reports. Also, the confirmation of the fog node damage time is accurate. Machine learning and related technologies have been widely applied in network

[*] Email of corresponding author: liuzaiying@mjc-edu.cn

and host-level attack defence and detection. Research has explored a deep learning model, deep neural network (DNN), to establish an intrusion detection system to timely discover unpredictable network attacks and determine their categories [4]. The aim of this current study is to address the problem of the high false alarm rate of intrusion detection system by optimizing the intrusion detection system, thereby improving its performance and detection capability.

2. RELATED WORK

The wireless sensor network (WSN) is an important part of a network's physical system. Static or mobile sensors can self-organize to form WSNs. Within a short time, typical attacks such as black hole attack, gray hole attack, flood attack and scheduling attack can destroy a WSN system. To solve these problems, Jiang s et al. created an intrusion detection method, SLGBM, to detect attacks on wireless sensor networks using A WSN-DS data set. The research shows that the detection rate of the SLGBM is 99.8% for normal attacks, 99.4% for black hole attacks, 99.1% for black hole attacks, 96.5% for flood attacks, and 96.1% for scheduling attacks. This method is obviously superior to the current typical detection methods [5]. In different application fields, the optimization of WSN algorithm can improve its application performance and practicability. For example, Yarinezhad and Hashemi [6] proposed a distributed cellular learning automata-based scheme that classifies and manages the WSN's fault node, and detects and reboots faulty nodes according to the fault state, and improves the service quality of WSNs. Li et al. [7] proposed a quantum ant colony multi-objective routing algorithm (QACMOR) for detecting problems in manufacturing environments, which introduces quantum computing and multi-objective fitness function. Zeng et al. [8] proposed a new mobile wireless sensor network (MWSN), which achieves communication and positioning functions by integrating Ad-Hoc and ultra-wideband (UWB) technologies. The direct cause of an accident is of great significance. Studies have shown that the time synchronization process is integrated into the routing and data transmission in the network, reducing the overhead and transmission exchange, which will be beneficial to large-scale WSNs [9].

With the rapid growth of today's social information data, the emergence of data mining technology can solve the difficulty of extracting information from big data by applying association rules that can detect strong links between frequent and complex itemsets. With the traditional association rule algorithm, the minimum support and confidence need to be preset before the algorithm is applied. However, people's subjectivity can have a significant impact on these two indicators. In response to these problems, researchers have optimized the association rule algorithm. For example, some scholars proposed an improved association rule algorithm, PSOFP, which can be used to find the global optimal solution, and introduced it into the Optimized Intelligent Particle Swarm Optimization Algorithm. Then, to address the problem of mining association rules, the FP growth algorithm is proposed. In addition, information entropy is adopted to determine the effectiveness of mining association rules. Some studies have shown that the correlation of social security events can be analyzed by using the improved algorithm [10]. One research has constructed an association rule mining algorithm on top of the particle swarm optimization algorithm, called the PSO-GES algorithm. Experiments on real transaction databases show that the quality of association rules of PSO-GES is better than existing algorithms for association rules mining [11]. Some studies have proposed a quantitative association rule extraction module that combines multiple algorithms. Indicators such as the extraction rule numbers, high confidence and support in the dataset have been greatly improved compared to the basic algorithm [12]. Zhang et al. [13] combined the multi-dimensional multi-objective bigroup discrete firefly algorithm (MODGDFA), and proposed a new association rule mining method that uses the MODGDFA algorithm to extract the association rules of the scheme design from a historical database containing a great amount of information [13]. Letrache et al. [14] proposed a dynamic partitioning strategy of OLAP cube based on association rule algorithm, which analyzes user queries in a specific time period to find frequent predicate itemsets, and then divides the data cube to improve query and processing performance. By combining network and association rules technology to discover moving paths, the researchers developed an association rule mining algorithm that models the edges as items and paths within each time span as transactions. Experiment results show that this method can effectively mine the motion patterns in the sliding trajectory data [15]. In order to modify the existing clustering-based methods, Chen CH et al. proposed a membership function adjustment mechanism for a fuzzy temporal association rule mining algorithm to generate a unique membership function customized for each item in the dataset. The feasibility of the algorithm has been proved in multiple data sets including real data sets [16].

The aforementioned literature indicates that many optimization studies have been conducted on association rules and WSN methods, but few studies have combined the two. Given the vulnerability of WSNs to attacks, and the false alarm rate in association rule algorithms, this current study optimizes the association rule algorithm and WSN algorithm. The proposed WSN intrusion rapid detection system based on association rules is intended to improve the detection of intrusions.

3. SOFTWARE-DEFINED WSN METHOD OPTIMIZATION ON THE BASIS OF ASSOCIATION RULES

3.1 CHC-Based Fuzzy Association Rules Mining algorithm

In the era of Internet information and data explosion, data mining technology is conducive to the development of the Internet and the application of data. Association rules are important data mining algorithms that can effectively extract and filter the information hidden in the data to obtain valuable

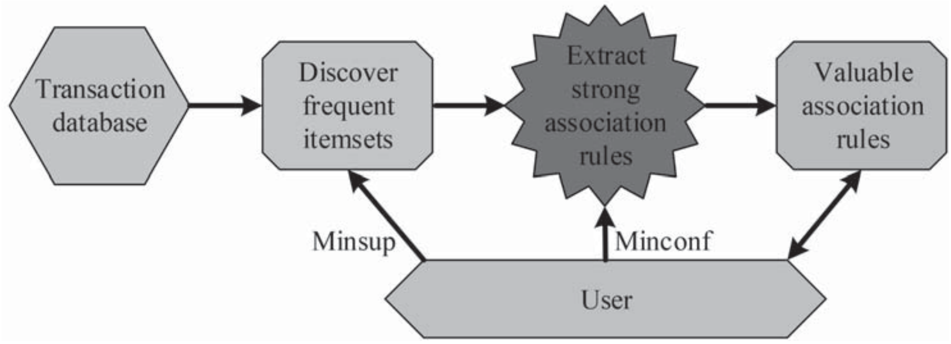

Transaction database
Discover frequent itemsets
Extract strong association rules
Valuable association rules
Minsup
Minconf
User

Figure 1 Basic process of association rules.

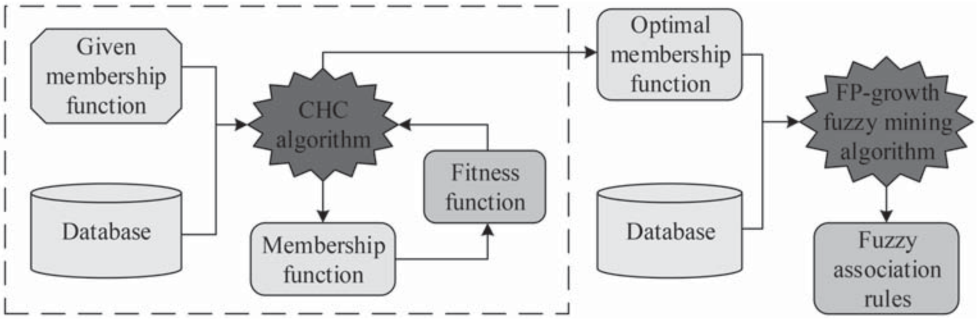


Figure 2 Flow chart of fuzzy association rule mining algorithm on the basis of CHC.

information [17]. As an unsupervised machine learning algorithm, the association rule mining algorithm can be classified according to the type, level and dimension of the data to be mined. The basic process for obtaining association rules consists of two steps. First, all frequent itemsets must be found, and then association rules should be obtained. The flow chart for this process is shown in Figure 1.

The triple semantic model can be used to express the membership function. The CHC algorithm is adopted to optimize the membership function in order to improve the accuracy of the function and obtain the optimal membership function. Then, on the basis of the optimal membership function, the FP-growth algorithm is used to obtain fuzzy association rules, which can effectively mine frequent itemsets and reduce the number of scans of the database. The basic flowchart is shown in Figure 2.

The best membership function can be obtained by the CHC algorithm. On this basis, FP-growth algorithm is used to build a tree structure to obtain the fuzzy association rules. The specific process of the CHC-based fuzzy association rule mining algorithm is as follows. First, input the chromosome $A\{a_1, a_2, a_3, \ldots, a_i, \ldots, a_n\}$, the membership function of chromosome $a_i = \{a_{i1}, a_{i2}, a_{i3}, a_{i4}\}$ is $u_i = \{a_{i1}, a_{i2}, a_{i3}, a_{i4}\}$; $L_{initial}$ is used to represent the initialization threshold; min*Sup* is used to represent the minimum support; and min *Conf* is used to represent the minimum confidence. Output fuzzy association rules, the initial population contains N chromosomes, and initialize them. Then, the fitness value of each chromosome is calculated, the data in the database is preprocessed, and the membership function is used to convert the data into the corresponding fuzzy interval. Then, calculate the chromosome x_i support according to Formulas (1) to (3).

$$count(x) = \left[\sum_{t \in T} \min \ \mu_x(t), \ \sum_{t \in T} \max \ \mu_x(t) \right] \quad (1)$$

$$\sup port(x) = \frac{count(x)}{|T|} \quad (2)$$

$$\begin{aligned}[\sup port_T(x)]_\alpha = \Bigg[& \frac{1}{|T|} \sum_{t \in T} \min\{\mu_x(t) | t \in [t]_\alpha\}, \\ & \frac{1}{|T|} \sum_{t \in T} \max\{\mu_x(t) | t \in [t]_\alpha\} \Bigg] \end{aligned} \quad (3)$$

In Formulas (1) to (3), t represents transactions in data set, T is used to represent the data set, $|T|$ is used to represent the number of data sets, the fuzzy partitions of *low, middle* and *high* in the membership function are represented by α, and the membership degree of the itemset is represented by μ_x, and fuzzy partitions of itemsets are usually represented by sets $[t]_\alpha$. Determining the relationship between the support of x_i and min *Sup*, if the support of x_i is greater than min *Sup*, then put it into the frequent itemset L_1. Next, use Formulas (2) to (5) to calculate the fitness value of each chromosome in the initial population.

$$GM3M = (\delta \gamma \rho)^{\frac{1}{3}} \quad (4)$$

$$fitness(c) = \sum \sup port(x_i) \cdot GM3M \quad (5)$$

In Formulas (4) to (5), δ is used to represent the displacement amount, γ is used to represent the lateral amplitude rate, ρ represents the area similarity; its value range is [0, 1]. The parental chromosomes are two chromosomes randomly selected from the initial population. The Hamming distance of the parental chromosomes is calculated and

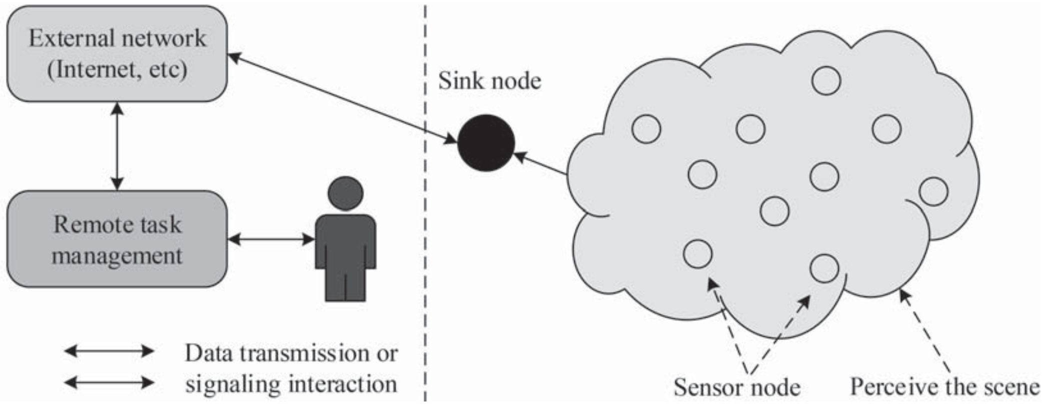


Figure 3 Wireless sensor network architecture.

compared with $L_{initial}$. If the Hamming distance is greater than the initialization threshold, the chromosome crossover operation cannot be performed. Using the parent center BLX operator method to cross chromosomes, the offspring chromosomes can be obtained, and their fitness values can be calculated. The parental chromosome of the next generation is selected from the parental and progeny chromosomes, and the N chromosomes with the highest fitness value are adopted as the parental chromosome, and this step is repeated for iterative optimization. Select the best chromosome in parent and offspring; if there is no change, decrease the $L_{initial}$ value to find the best chromosome. If the best membership function is not selected, repeat the above steps to obtain the best function. The descending sorting of frequent 1-itemsets is generated according to the upper limit of support, and then the number of frequent patterns is obtained taking the maximum support into account. Find out the association rules that exist in all frequent itemsets, calculate the confidence of the rules, and generate strong association rules.

3.2 Research and Optimization of Software-Defined WSN Based on Association Rules

A WSN has several advantages: low cost, large fault tolerance, and wide application in the market [18]. It consists of three nodes: sensing, aggregation and management. The sensing node has low computing power and limited transmission distance, and the communication mode used by the sensor node and the sink node usually chooses the multi-hop mode, which plays the role of the acquisition terminal and the routing relay. The aggregation node transmits the data obtained from the external network such as the Internet to the management node, and acts as a gateway between the wireless sensor network and the external network. The management node can achieve direct interaction with users, and can control the access to wireless sensor data resources. Figure 3 shows the architecture of the WSN.

However, nowadays, the disadvantages of wireless sensor networks have gradually emerged, such as the large number of nodes and a relatively discrete distribution. To improve the quality of WSN's data communication, and to carry out dynamic network management conveniently and flexibly, the traditional wireless sensor technology is optimized by introducing the software-defined network (SDN) technology. However, the existing SD-WSN technology is vulnerable to attacks, including attacks on forwarding devices, control planes, and management stations, and may also cause problems such as data leakage, data modification, and denial of service. In order to solve these problems, based on fuzzy association rules, this study introduces the artificial bee colony algorithm to optimize K-means clustering, and applies the optimization algorithm to SD-WSN technology to ensure the efficient and stable operation of the network. This new software-defined WSN is named ESD-WSN. ESD-WSN consists of a base station, gateway device, proxy node and sensor node. The architecture is shown in Figure 4.

Cluster analysis can analyze multivariate data and classify it according to the data's internal characteristics to form similar data categories. Most of the network data comes from an unknown source, so the security of the data cannot be determined. Cluster analysis can distinguish normal behaviors from attack behaviors according to the different characteristics of the data, so as to detect network intrusions. The clustered data includes data matrix and dissimilarity matrix. The data matrix is also known as a 'bimodal matrix', which uses ρ attributes such as name or age to form n object person, the calculation formula is shown in Formula (6).

$$N \text{ objects} \left\{ \begin{bmatrix} x_{1l} & \dots & x_{1f} & \dots & x_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{f1} & \dots & x_{ff} & \dots & x_{fn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & \dots & x_{nf} & \dots & x_{nn} \end{bmatrix} \right. \tag{6}$$

The dissimilarity matrix is also called a 'single-mode matrix', which is used to show the similarity between objects. The rows and columns in the matrix are used to represent the same object, the difference between objects i and j is expressed as $x(ij)$. $x(ij)$ is a non-negative number that is proportional to the degree of difference. Formula (7) is used to calculate the dissimilarity matrix.

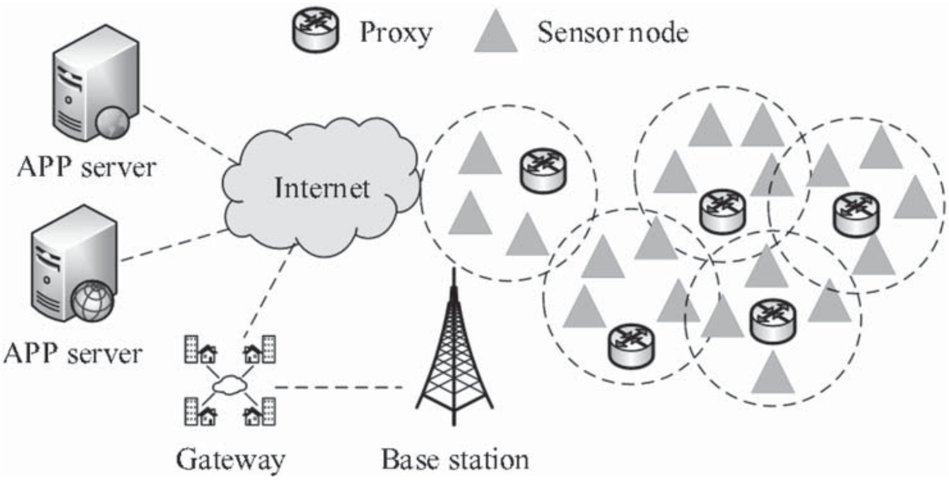


Figure 4 Overall architecture of ESD-WSN.

$$\begin{bmatrix} 0 & & & & \\ x(2,1) & 0 & & & \\ x(2,1) & x(2,1) & 0 & & \\ \vdots & \vdots & \vdots & 0 & \\ x(2,1) & x(2,1) & \dots & \dots & 0 \end{bmatrix} \tag{7}$$

Suppose two different objects and C, and E, $C = \{c_1, c_2, \dots, c_n\}$, $E = \{e_1, e_2, \dots, e_n\}$, each object contains n characteristic attributes that can be measured. The degree of dissimilarity of C and E is defined as $d(C, E) = f(C, E) \to R$, and the real number field is represented by R. The dissimilarity can be obtained by calculating the Euclidean distance between undirected scalar data, and can also be expressed by the Manhattan distance and the Minkowski distance. The Formulas (8)–(10) are used for the calculations.

$$d(C, E) = \sqrt{(c_1 - e_1)^2 + (c_2 - e_2)^2 + \dots + (c_n - e_n)^2} \tag{8}$$

$$d(C, E) = |c_1 - e_1| + |c_2 - e_2| + \dots + |c_n - e_n| \tag{9}$$

$$d(C, E) = \sqrt[p]{|c_1 - e_1|^p + |c_2 - e_2|^p + \dots + |c_n - e_n|^p} \tag{10}$$

In Formula (10), when $p = 1$, the Minkowski distance can be derived, and when $p = 2$, the Euclidean distance can be derived. For binary variables that can only be defined as 0 or 1, the dissimilarity can be obtained by calculating the proportion of the same-valued attributes of elements in the same order. As an extension of binary variables, the dissimilarity of categorical variables can be expressed by the mismatch rate, as shown in Formula (11).

$$d(c, e) = \frac{p - m}{p} \tag{11}$$

In Formula (11), p is used to represent the total number of variables, under the same state value, the number of variables of objects c and e is represented by m. For a vector with magnitude and direction, the cosine of the vector can be calculated to obtain its dissimilarity, as shown in Formula (12).

$$s(C, E) = \frac{C'E}{\|C\|\|E\|} \tag{12}$$

In Formula (12), $\|C\|$ and $\|E\|$ represent the Euclidean norm of C and E, respectively. In the K-means clustering algorithm, the square error criterion function is used to converge the cluster centers obtained after two clusterings, as shown in Formula (13).

$$J(h,\ \mu) = \sum_{i=1}^{k} \|x^i - \mu_{h^{(i)}}\| \tag{13}$$

In Formula (13), the mean value of cluster i is expressed by $\mu_{h^{(i)}}$. The smaller the average error of the samples in the cluster, the more similar the samples are, and the distance between the samples can be calculated using the Euclidean formula. In the traditional K-means clustering algorithm, the K value is selected according to experience, and the selection of the K value and the initial cluster center has a great influence on the experimental results. When data quantity is large, K-means clustering needs more time. To improve the clustering effect and performance of K-means, this study adopts the improved artificial bee colony algorithm to optimize the clustering algorithm. Assuming there is N nectar source $\{x_1, x_2, \dots, x_N\}$, the solution of a set of problems represented by d-dimensional vectors can be represented by a honey source. Assuming that bees have carried out M cycles, each honey source can be mined L time. In the initialization stage, the d-dimensional vector is used to represent the N group solutions of the randomly generated N problems, calculate the fitness values of each solution, and arrange them according to the size, and divide them into two parts, which are the leading bee and the following bee. Around the existing honey sources, the leading bee generates new honey sources by using Formula (14) through domain search. Compare the fitness values of the two honey sources, and save the honey source with the higher value.

$$X_{ij} = x_{ij} + r_{ij}(x_{ij} - x_{kj}) \tag{14}$$

In Formula (14), the new position near x_{ij} is represented by X_{ij}, and k and j are random numbers, where $k = 1, 2, 3, \dots, N$ and $k \neq i$. $r_{ij} \in [-1, 1]$, X_{ij} can be constrained nearby x_{ij}. In the following phase, use Formula (15) to calculate the probability to find the next leader bee.

Table 1 Results for lifting degrees.

Rule	$a(l) \to b(l)$	$b(l) \to a(l)$	$c(l) \to a(l)$	$a(l) \to c(l)$
Lifting degree	1.35452	1.67786	1.95000	2.03460
Rule	$d(l) \to c(l)$	$c(l) \to d(l)$	$d(m) \to c(l)$	$c(l) \to d(m)$
Lifting degree	2.18890	0.78760	1.94337	2.96505

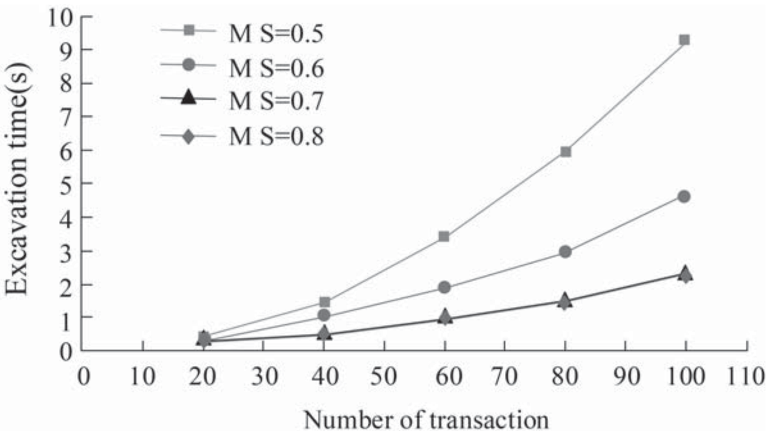


Figure 5 Relationship between transaction number and excavation time.

$$P_i = \frac{fitness_i}{\sum\limits_{i=1}^{N} fitness_i}; \; i = 1, 2, 3, \ldots, N \tag{15}$$

In Formula (15), the fitness value of the nectar source i is expressed by $fitness_i$, the probability of selecting leading bees is expressed by P_i. After the leading bee is selected, a new nectar source is generated according to Formula (14), the fitness value is compared, and the nectar source with higher value is saved. Finally, after L cycles, when the nectar source has not changed, the lead bee will become a scout bee to find a new nectar source. The fitness function can have an impact on the evolutionary direction and evolutionary behavior of the population. Suppose that k samples in the data set $X = \{x_1, x_2, \ldots, x_n\}$ are divided into classes S_i, $S_i = \{x_1, x_2, \ldots, x_k\}$. In class S_i, the in class distance of the sum of squares of data x and y is shown in Formula (16), the distance between class S_i and other classes is expressed as the distance between classes, as shown in Formula (17).

$$d_{det}(S_i) = \sum_{x,y \in S_i} \|x - y\|^2 \tag{16}$$

$$d_{wit}(S_i, S_j) = \sum_{j=1,\ j \neq i}^{k} \frac{1}{kp} \sum_{x \in S,\ y \in S_j} \|x - y\|^2 \tag{17}$$

In Formula (17), x represents the data object in class S_i, y represents the data object in class S_j, k represents the number of objects in class S_j, k represents the number of objects in class S_j. The fitness function is shown in Formula (18).

$$fitness_i = \sum_{i=1,\ j \in K}^{k} \left[d_{bet}(S_i) + \frac{1}{d_{wit}(S_i, S_j)} \right] \tag{18}$$

4. SIMULATION ANALYSIS OF SOFTWARE-DEFINED WSN ALGORITHM ON THE BASIS OF ASSOCIATION RULES

In this experiment, a fuzzy association rule mining algorithm is proposed on the basis of CHC. To verify the superiority of this algorithm, the CHC-based fuzzy association rule mining algorithm and the A priori algorithm are compared using randomly-collected data. The optimal membership function of chromosomes can be calculated by the CHC-based fuzzy association rule mining algorithm, and the confidence and lift are calculated after fuzzy processing of the data. The association rules with high confidence are retained and combined into a dataset database. First, select association rules with higher confidence in the two libraries, then calculate the lift according to formula $lift(X \to Y) = \frac{P(X,Y)}{P(X) \cdot (Y)}$. The experimental results are compared and analyzed in the C++ programming language. The improved results of the new algorithm are shown in Table 1.

In Table 1, l represents low and m represents middle. As can be seen from Table 1, the lift in $c(l) \to d(m)$, $d(l) \to c(l)$, $a(l) \to c(l)$, $c(l) \to a(l)$, $d(m) \to c(l)$ are 2.96505, 2.18890, 2.03460, 1.95000 and 1.94337, respectively, indicating that these five rules have high reliability and accuracy in the database, which proves that the fuzzy association rule mining algorithm based on CHC proposed in this study is more optimized.

The algorithm mining efficiency is affected by the data structure, database size and setting parameters. In the experiment, the relationship between things' quantity, mining time, minimum support, minimum confidence, and association rules' quantity were investigated.

According to Figure 5, M S represents minimum support. Increasing the number of things in the database can prolong

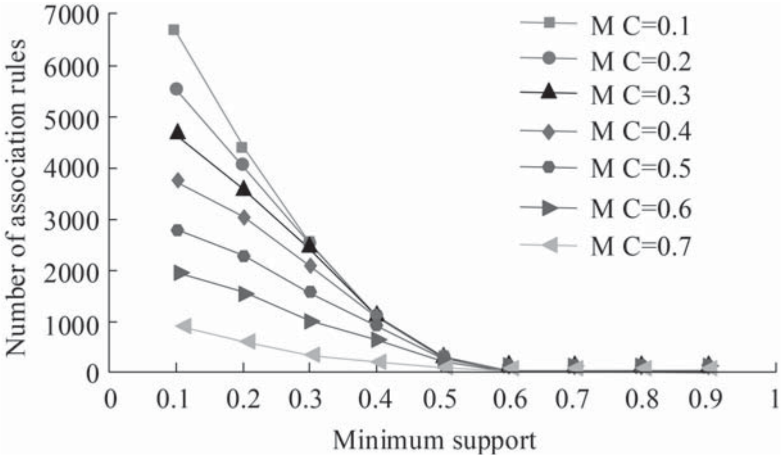


Figure 6 Relationship between rule number and minimum support.

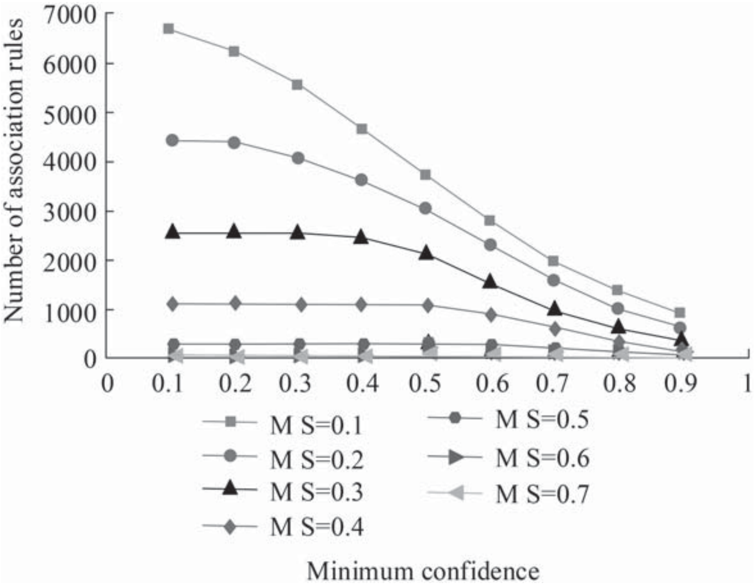


Figure 7 Relationship between rule number and minimum confidence.

the mining time, and the relationship increases approximately linearly; the mining efficiency is inversely proportional to the minimum support. The minimum support has no significant impact on the mining efficiency at the threshold. It shows that the optimization model proposed in this study has good scalability.

In Figure 6, M C represents minimum confidence. When the value of the minimum confidence is a fixed value, the number of association rules gradually decreases with the increase of the minimum support. The change occurs in two stages. In the first stage, when the minimum support increases, the number of association rules decreases greatly; in the second stage, as the minimum support increases, the number of association rules gradually decreases, and finally arrives at 0.

In Figure 7, when the value of the minimum support is fixed and higher than the minimum confidence, the number of association rules remains unchanged while minimum confidence increases; when the value of the minimum support is fixed and less than the minimum confidence, the number of association rules decreases while minimum confidence increases. The results of the analysis show that the algorithm's mining performance is better when the M S is 0.4 and the M C is about 0.8. Therefore, , this study uses the processed fuzzy database in the network topology to compare different algorithms to test new algorithm's performance. Using the mining time and minimum support as indicators, the algorithm proposed in this study is compared with the FDMA algorithm and FPLI algorithm. The results are shown in Figure 8.

As seen in Figure 8, when the value of the minimum support is small, the time required by the FPL1 algorithm and the algorithm to mine frequent itemsets is much lower than that of FDMA algorithm. When the degree of minimum support increases, the time required by three algorithms gradually tends to be consistent, which may be due to the gradual decrease of the highest order in the used database that can satisfy the degree of minimum support.

In order to obtain better algorithm performance, this research proposes a software-defined WSN algorithm on the basis of fuzzy association rules, and uses CHC algorithm and FP-growth algorithm to obtain fuzzy association rules for data mining. On this basis, the artificial bee colony algorithm is

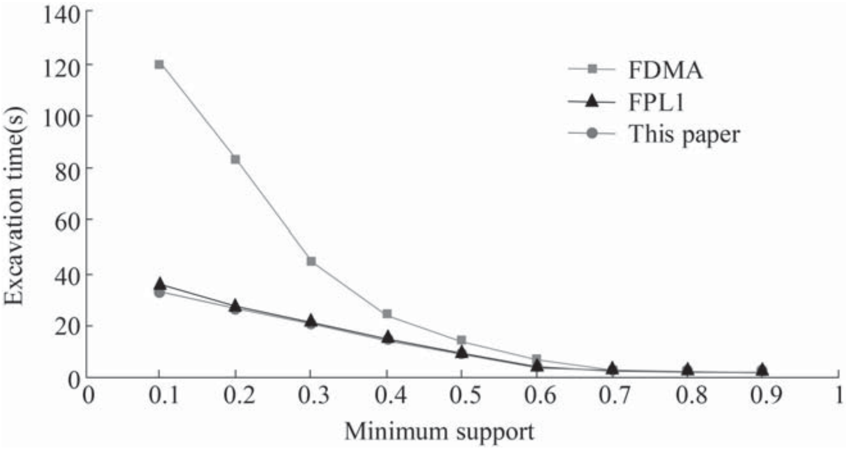
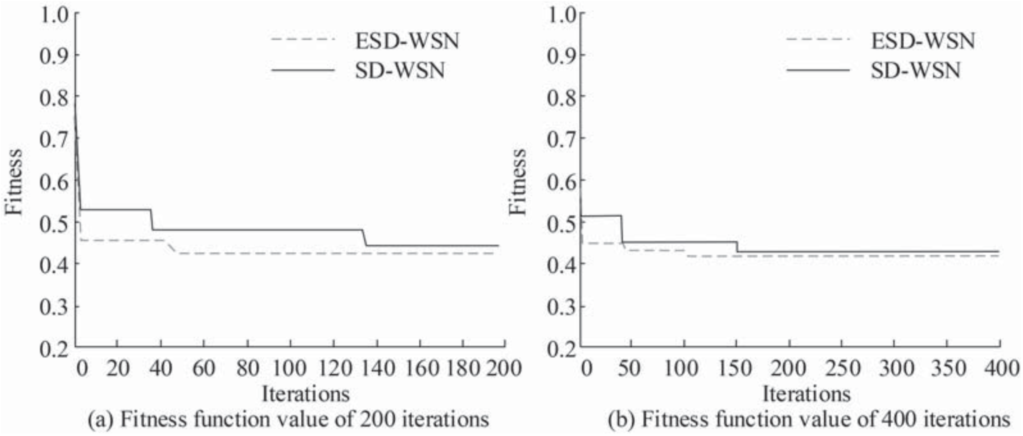


Figure 8 Comparison of mining times under different supports.



(a) Fitness function value of 200 iterations

(b) Fitness function value of 400 iterations

Figure 9 Iterative fitness function value.

introduced to make the K-means clustering optimal. Then, the new method is applied to SD-WSN technology to ensure the efficient and stable operation of the intrusion detection system. In this experiment, the KDDCUP99 dataset, which can be used in intrusion detection experiments, is selected to test the accuracy and availability of ESD-WSN, and the detection and false alarm rate were used as indicators for verification. After data cleaning, format conversion and normalization of the dataset, ESD-WSN and SD-WSN algorithms are trained in the data set. After 200 and 400 iterations of the data using the two algorithms, the corresponding fitness function value results can be obtained, as shown in Figure 9.

As can be seen from Figure 9, after 200 and 400 iterations of the data, the fitness function value of the SD-WSN is higher than ESD-WSN, which indicates that ESD-WSN has a lower fitness function value than SD-WSN. The ESD-WSN' clustering effect has been significantly improved. When the number of iterations is 400, the fitness function value of ESD-WSN algorithm is lower than after 200 iterations. Therefore, the number of iterations is set to 400 to conduct a comparative experiment on the detection rate and false alarm rate of K-means clustering algorithm, SD-WSN and ESD-WSN. Figure 10 shows the results for the detection rate and false positive rate.

In Figure 10, the improved ESD-WSN algorithm proposed in this study has significantly improved the detection rate and greatly reduced the false positive rate. The optimized ESD-WSN algorithm can obtain good clustering and detection outcomes. As indicated by the experimental results above, the improved ESD-WSN has better performance and is a feasible and effective approach for the establishment of an intrusion detection system.

5. CONCLUSION

For the software-defined WSN intrusion detection system based on association rules proposed in this study, the algorithm is optimized. According to the optimization results, the algorithm lift of $c(l) \rightarrow d(m)$, $d(l) \rightarrow c(l)$, $a(l) \rightarrow c(l)$, $c(l) \rightarrow a(l)$, $d(m) \rightarrow c(l)$ are 2.96505, 2.18890, 2.03460, 1.95000 and 1.94337 respectively. When the M S and M C are about 0.4 and 0.8 respectively, the mining performance of ESD-WSN algorithm is better; When the minimum support is small, the time required by FPL1 algorithm and ESD-WSN algorithm to mine frequent itemsets is less than that of FDMA algorithm; when the minimum support increases, the time required by the three algorithms gradually tends to be consistent; when the number of iterations is 400, the fitness function value of the ESD-WSN algorithm is low. It can be concluded that the software defined WSN algorithm based on association rules proposed in this study

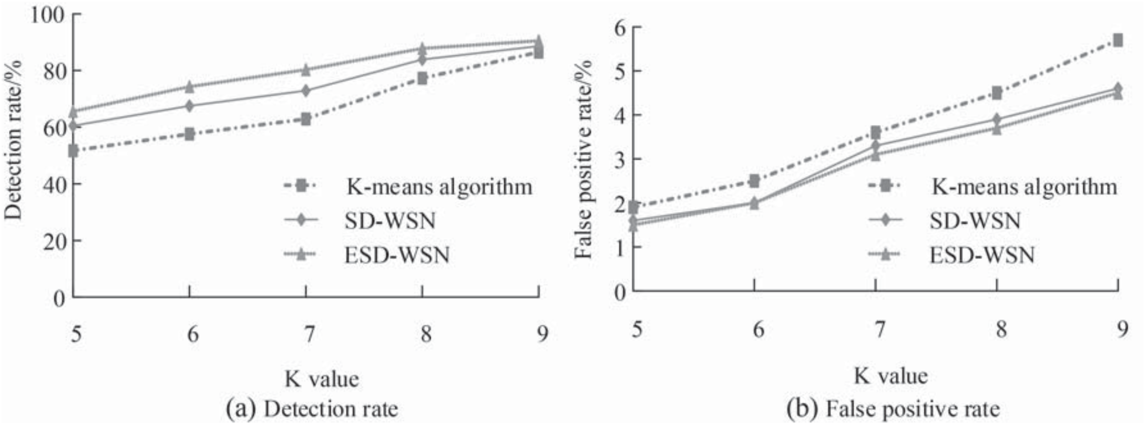


Figure 10 Change curve of detection rate and false positive rate of different algorithms.

performs well when applied to an intrusion detection system, the detection rate of the algorithm is improved, and the false positive rate is reduced. This research proposes an innovative software-defined WSN algorithm based on fuzzy association rules, which improves the performance and detection rate of the traditional algorithm and can be used for the optimization of an intrusion detection system. However, the study has limitations. The improved algorithm does not distinguish between active intrusion and passive intrusion. In the next study, we will improve the specific content and establish a further improved rapid intrusion detection system.

FUNDING

The research is supported by the 2019 Shanghai Education Science Research Project (C19093).

REFERENCES

1. Almomani I, Alromi A. 2020. Integrating software engineering processes in the development of efficient intrusion detection systems in wireless sensor networks. *Sensors, 20*(5):1375–1402.
2. Al-Daweri MS, Abdullah S, Ariffin KAZ. 2021. An adaptive method and a new dataset, UKM-IDS20, for the network intrusion detection system. *Computer Communications, 180*(1):57–76.
3. Pacheco J, Benitez VH, Filix-Herran LC, et al. 2020. Artificial neural networks-based intrusion detection system for Internet of Things fog nodes. *IEEE Access, 8*(1):73907–73918.
4. Vinayakumar R, Alazab M, Kp S, et al. 2019. Deep learning approach for intelligent intrusion detection system. *IEEE Access, 8*(1):41525–41550.
5. Jiang S, Zhao J, Xu X. 2020. SLGBM: An intrusion detection mechanism for wireless sensor networks in smart environments. *IEEE Access, 8*(1):169548–169558.
6. Yarinezhad R, Hashemi S N. 2019. Distributed faulty node detection and recovery scheme for wireless sensor networks using cellular learning automata. *Wireless Networks, 25*(5):2901–2917.
7. Li F, Liu M, Xu G. 2019. A quantum ant colony multi-objective routing algorithm in WSN and its application in a manufacturing environment. *Sensors, 19*(15):3334–3347.
8. Zeng P, He J, Gao B. 2021. Reliable robot-flock-based monitoring system design via a mobile wireless sensor network. *IEEE Access, 9*(1):47125–47135.
9. Su T, Xu H, Zhou X. 2019. Particle swarm optimization-based association rule mining in big data environment. *IEEE Access, 7*(1):161008–161016.
10. Baró GB, Martínez-Trinidad JF, Rosas R, et al. 2020. A PSO-based algorithm for mining association rules using a guided exploration strategy. *Pattern Recognition Letters, 138*(1):8–15.
11. Moslehi F, Haeri A, F Martínez-Lvarez. 2020. A novel hybrid GA–PSO framework for mining quantitative association rules. *Soft Computing, 24*(6):4645–4666.
12. Zhang Z, Chai N, Ostrosi E, et al. 2019. Extraction of association rules in the schematic design of product service system based on pareto-MODGDFA. *Computers & Industrial Engineering, 129*(3):392–403.
13. Letrache K, Beggar OE, Ramdani M. 2019. OLAP cube partitioning based on association rules method. *Applied Intelligence, 49*(2):420–434.
14. Yu WH. 2019. Discovering frequent movement paths from taxi trajectory data using spatially embedded networks and association rules. *IEEE Transactions on Intelligent Transportation Systems, 20*(3):855–866.
15. Chen CH, Chou H, Hong TP, et al. 2020. Cluster-based membership function acquisition approaches for mining fuzzy temporal association rules. IEEE *Access, 8*(1):123996–124006.
16. Safara F, Souri A, Serrizadeh M. 2020. Improved intrusion detection method for communication networks using association rule mining and artificial neural networks. *IET Communications, 14*(7):1192–1197.
17. Abella CS, Bonina S, Cucuccio A, et al. 2019. Autonomous energy-efficient wireless sensor network platform for home/office automation. *IEEE Sensors Journal, 19*(9):3501–3512.