Review on ACO-based Intrusion Detection Method in Computer Networks using Fuzzy Association Rules

Abhi B C[1], Mr Pradeep Nayak[2], Abhilash C M[3], Abhishek M S[4], Adarsh[5]

Faculty, Department of Information Science and Engineering[2] Students, Department of Information Science and Engineering[1,3,4,5]
Alva's Institute of Engineering and Technology, Mijar, Mangalore, Karnataka, India.

Email: 18842adarsh@gmail.com

Abstract The identification of harmful activity and infiltration into systems and computer networks is a fundamental difficulty in this field due to the growing usage of computer networks. Nine fundamental steps make up the study's suggested approach for intrusion detection using fuzzy association rules and ant colony optimization. Every node in the graph that Ant Colony operates on gives seven parameters for the suggested approach. The Apriori method creates the initial fuzzy rules based on the values in the ant colony's most populous city in the graph. To increase the precision of intrusion detection in computer networks, the main rules are refined through a process known as conventional weighted aggregation. Using decision trees using NSL-KDD data, the efficiency of the optimized rules is assessed. Assessments show that the suggested approach has a high degree of intrusion detection accuracy and decrease in error risk when compared to other approaches.

Keywords: computer network, data mining, ant colony optimization, fuzzy association rule, and intrusion detection.

I. INTRODUCTION

In actuality, intrusion is any behavior or action that compromises a resource's availability, confidentiality, or integration [1]. Intrusion detection systems notify and regulate these actions, and firewalls are made to stop intruders from accessing security resources. Although it is impossible to totally eliminate intrusions, steps must be taken to automatically and instantly control user behavior and stop invasive activities.

IT also employs intrusion detection systems, which analyze user data to assess network system risk [2]. Differentiating between primary and spam emails is one of its uses. By doing this, businesses who offer these services may improve the quality of their offerings and stop spam from entering the inbox. Because of the rise in network connectivity, computer Systems can be attacked. These scripts frequently run in the programs or operating system. The main objective of such Attacks are interfering with systems' security measures and operating the system without permission. The findings of the intrusion detection study demonstrate that these Studies were considerably successful in addressing security requirements in computer networks. We'll talk about a few of these studies below.

A study titled "artificial immune system of fuzzy cooperative for intrusion detection in wireless sensor networks" was carried out by Shahabuddin Shamshirband et al. [3]. The nature of distributed denial of service attacks makes it extremely difficult to identify such malicious activity in wireless sensor networks using conventional intrusion detection systems. The paper suggests an artificial immune system based on fuzzy cooperation, a method inspired by nature. This approach is a modular defensive technique that is based on the notion of risk associated with the human immune system in order to update the fuzzy activation threshold for security response and determine abnormalities in sensor activity.

Selma Elhagh et al. [4] conducted a study entitled "the combination of fuzzy genetic and paired learning algorithms to enhance detection rates on intrusion detection systems". In this article, approaches that are based on computational intelligence are employed to robustly implement intrusion detection systems and accurately. In this paper, fuzzy

genetic system in the paired learning framework has been applied. The study of Hung-Jen is one of the recent initiatives in this field. "Intrusion detection system: a study" by Liao et al. thorough analysis," who suggested a new classification to ascertain contemporary detecting techniques and have also produced a improved comprehension of intrusion detection systems by the use of tables and pictures [5]. This article's suggested approach attempts to provide the best rules for attack detection by combining the Apriori method with the ant colony algorithm. More specifically, the ant colony is to give the Apriori algorithm the actual level of support and confidence it needs to produce appropriate rules. Then, by weighting the rules that are obtained, some of the rules are eliminated. With repetitions of the ant colony method, optimized rules are generated.

ll. SUGGESTED METHOD

The nine steps of the suggested approach are described in specifics.

A. Step 1: Ant colony method initialization Initialization is done for the ants on each point. Initialization is carried out at random and taken into consideration to choose the best answer [6]. The number of cities will be Ncity, and the number of ants will be Nant. The type of initialization will be such that cities are regarded as seven. Seven values are represented by each city. Ants attempt to Go between these cities and select the one that best reflects your ideals. utilized because of its structure are more suitable for the suggested method. These ants can all be employed in Apriori methods. function [7].

B. Fuzzy segmentation in step two the triangle membership function is used to segment the entire dataset into linguistic variables [8]. Every variable turns into a k linguistic variable. A and μ represent each linguistic variable and its corresponding membership function. In this study, we take k to be equal to 3. This implies that each variable will be transformed into three linguistic variables, each of which has three possible values: low, medium, and high. It is possible to convert all values to the three specified levels. Equation (1) is used to determine each feature's membership degree:

$$\mu_{i_1}^{x} = \begin{cases} 0 & \text{value} \le a \\ \frac{\text{value}-a}{b-a} & a \le \text{value} \le b \\ \frac{c-\text{value}}{c-b} & b \le \text{value} \le c \\ 0 & c \le \text{value} \end{cases} \quad (1)$$

Where:

$$a = \min_x \quad (2)$$

$$b = \min_x + \frac{\max_x - \min_x}{2} \quad (3)$$

$$c = \max_x \quad (4)$$

More specifically, equation (1) is used to calculate the values of μ (low value), μ (mid value), and μ (high value) for each value of dataset fields. The triangle membership function is connected to this equation. The field's value is the argument of value. Equations (2) through (4) should be used to determine the values of a, b, and c in order to determine the value in equation (1). Figure 1 [9] depicts the structure of ants during the ant colony optimization process.

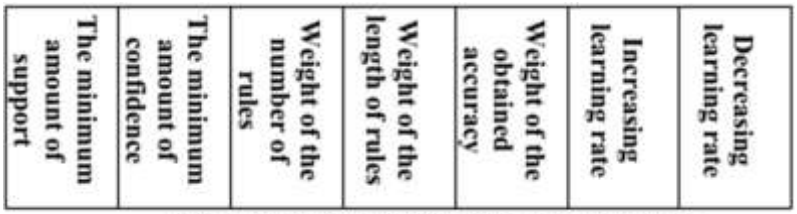

Figure 1: Structure of ants in the ant colony.

C. Step three: creating a group of frequently occurring fuzzy things The Apriori approach is used to produce the set of

often occurring objects. Attack training data is subjected to the Apriori approach. The f-member item set is one of the most widely used items in associate fuzzy induction. To put it another way, this section aims to use the Apriori approach to extract f-member common patterns associated with assaults for rule development, which can then be applied to test data using decision trees. F denotes how many fields, or features, there are in the dataset. As a result, the Apriori algorithm should be used to generate all common patterns off length, and the level of support for each pattern should be calculated. Equation (5), which uses the fuzzy value of a given membership function, is used to determine fuzzy support [10].

$$\text{FuzzySupport}\left(A_{i_k}^{x_1} \times A_{i_k}^{x_2} \times \ldots \times A_{i_k}^{x_{f-1}} \times A_{i_k}^{x_f}\right) = \frac{\sum_{p=1}^{n}\left(\mu_{i_k}^{x_1} \times \mu_{i_k}^{x_2} \times \ldots \times \mu_{i_k}^{x_{f-1}} \times \mu_{i_k}^{x_f}\right)}{n} \quad (5)$$

D. Generation of fuzzy rules in step four It's time to extract rules from the set of frequently occurring items. Rules are induced based on the degree of confidence [10]. In a fuzzy induction, confidence is equivalent to:

$$\text{FuzzyConfidence}(R) = \frac{\text{FuzzySupport}\left(A_{i_k}^{x_1} \times A_{i_k}^{x_2} \times \ldots \times A_{i_k}^{x_{f-1}} \times A_{i_k}^{x_f} \times A_{i_k}^{x_\gamma}\right)}{\text{FuzzySupport}\left(A_{i_k}^{x_1} \times A_{i_k}^{x_2} \times \ldots \times A_{i_k}^{x_{f-1}} \times A_{i_k}^{x_f}\right)} \quad (6)$$

Fuzzy confidence, whose confidence is equal to or more than the minimum quantity of FC, can be used to design an efficient set of rules. We refer to the outcome of the rule as
A. For instance, the support and confidence of the rule A –> B are determined using equations (7) and (8) [11]:

$$\text{FuzzySupport}(A \rightarrow B) = \frac{\sum_{x_p \in T} \mu_{AB}(x_p)}{|N|} \quad (7)$$

$$\text{FuzzyConfidence}(A \rightarrow B) = \frac{\sum_{x_p \in T} \mu_{AB}(x_p)}{\sum_{x_p \in T} \mu_{A}(x_p)} \quad (8)$$

where N is the total number of transactions.

- μ (x): represents the extent to which the xp transaction complies with the antecedent portion of the rule (A).
- μ (x): represents the extent to which the xp transaction complies with the antecedent and consequent of rule (B).

E. Step five: eliminating a few regulations It is evident that there is a tendency to minimize the number of regulations. Assume that the result (right) part of n distinct rules is the same. R1 through Rn are the rules. All of the rules from R2 to Rn are regarded as extra rules and must be eliminated if the precedent (left) portion of the rules is such that equation (9) is satisfied for their precedent section.

$$R_1 \epsilon R_2 \epsilon \ldots \epsilon R_n \quad (9)$$

F. The sixth step involves applying a traditional weighted aggregation function to calculate fit. In the earlier stages, rules were created. Equation (10) [12] is used to calculate f(ri), which is the fit of the ri set of rules.

$$f(r_i) = w_{Ac} \times \text{Accuracy}(r_i) - w_g \times n_{ri} - w_{Ar} \times Ar_{ri} \quad (10)$$

Where:

- Accuracy: indicates how well the rules produced in the it algorithmic iteration were used for classification. The 10-fold approach is used for the evaluation. The weight of categorization accuracy is known as WAc.
- nri: is the number of rules associated with associate rules produced in the suggested algorithm's ith iteration.
- wg: is the weight taken into account for the quantity of regulations.
- Arri: is the average number of rules in the algorithm's ith iteration.
- wAr: is the weight of the rule set's average length.

The number of requirements posed for a rule is its length. Given that accuracy is maintained, it is evident that rules with the shortest length are preferred. The most populous city with ants on the graph is where the weights used in this stage are chosen.

G. Step seven: rules to determine the weight of fuzzy rules It is evident that trustworthy rules improve system efficiency. Since rules are created based on attacks, the weight of a rule that is displayed as wa is increased if an attack in the training section is correctly identified; otherwise, it is decreased. As a result, rules pertaining to every attack sample found in the training dataset are chosen. Equation (11) increases the rule's weight if the intended sample is correctly classified by the rule; otherwise, equation (12) decreases it.

$$w_\alpha = w_\alpha + \vartheta_i(1 - w_\alpha) \quad (11)$$

$$w_\alpha = w_\alpha - \vartheta_d w_\alpha \quad (12)$$

where, in relation to weight change, ω and fi represent rising and falling learning rates, respectively. The most populated city with ants on the graph determines these criteria. Hey. Eighth step: finishing the ant colony process Every ant in the algorithmic process has seven dimensions. For the Apriori algorithm to proceed, it requires the least amount of confidence and support. In the initial step of the suggested procedure, these values were randomly created by ants. The process of creating rules is then carried out using the values found in the most populated city with ants. The values pertaining to ants are updated in this stage, and the earlier procedures are repeated.

III. RESULTS OF THE EXPERIMENT

Following an examination of the specifics and generalities of the suggested approach in section 2, it's time to assess its effectiveness and contrast it with earlier techniques. The reliable and well-liked DKK99 dataset information about issues with intrusion detection. But, intrinsic statistical issues with the KDD99 dataset Measurement standards have been made public for researchers in current research [14]. As a result, scholars in this area have attempted to address the issues with this dataset and suggested a more NSL- KDD is a suitable version of this dataset. There are 41 features in each dataset entry, and there is a label that establishes the normalcy of each record or intrusion's intricacy.

Table 1: values of the parameters in step one of the ant colony algorithm

Parameter	Value
N_{ant}	30
N_{city}	30
α	1
β	1
ρ	0.1

Table 2: parameters of the step six of the proposed method

Parameter	Value
w_{Ac}	18.5
w_g	1.1
w_{Ar}	1.1

Weighting the regulations is the focus of step eight. More specifically, the proposed approach's ultimate objective is to use the seven values associated with the most populous city at the conclusion of the ant colony method's iterations for the Apriori method. Additionally, each rule's significance is stated. If a particular sample is correctly identified, the weight of a rule represented by way is increased; otherwise, it is dropped. As a result, rules pertaining to every sample in the dataset are chosen. To put it another way, it is preferable to have a rule that is only used once and is not very exclusive. Pheromones and ants are updated in stages eight and nine so they can travel toward a different city. As a result, values that are suitable for the Apriori algorithm's implementation are obtained at the conclusion of the ant algorithms' iterations. It is possible to determine which rule is more significant than the others based on the weights assigned to each rule at step seven.

Table 3: values related to parameters of step seven

Parameter	Value
ϑ_i	0.1010
ϑ_d	0.0015

The findings are examined and contrasted with alternative approaches after examining the values of the parameters set in the suggested strategy. It should be mentioned that the findings of rival approaches are taken from the research of [15, 16]. It is necessary to first choose an appropriate criterion for comparison. Confusion matrix is one of the most crucial factors used to assess classification algorithms. The dimensions of this square matrix match the number of classes in the categorization. True detections are associated with elements of the confusion matrix's main diagonal, while incorrect detections are associated with parts of the secondary diagonal. The system's efficiency is assessed using three reliable criteria: precision, recall, and f-measure.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (13)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (14)$$

$$\text{F-measure} = \frac{2\times\text{Precison}\times\text{Recall}}{\text{Precison}+\text{Recall}} \quad (15)$$

Table 4 displays the precision, recall, and f-measure results of the suggested approach along with a comparison with rival approaches.

Table 4: comparison of the results of the proposed method with other methods

Method	Precision	Recall	F-measure
Fuzzy rules with C4.5 tree [15]	53.22	58.56	55.76
Learning fuzzy rules with naïve Bayse [15]	49.22	60.86	54.42
Learning fuzzy rules with 5NN [15]	60.24	59.24	59.73
Learning fuzzy rules with SVM [15]	73.32	59.02	65.39
Javaid et al. [16]	72.11	63.09	67.29
Proposed method	74.44	89.1	81.11

The cause of the higher recall improvement is one of the things in Table 4 that requires explanation. An increase in recall indicates that the suggested strategy has extracted more relevant results. Table 3 should be used to determine the cause of this. Table 3 makes it evident that a rule's weight increases more quickly than it decreases. This is due to the endeavor to maintain rules rather than break them. As a result, the regulations are either eliminated or given less weight. However, as Table 4 shows, this results in a decline in precision.

Considering the results shown in Table 4, the superiority of the three criteria show the suggested approach. In addition to criteria given in equations (13) through (15), the ROC diagram is utilized to Compare the approaches. This graphic compares two categorization techniques. TPR (true positive) is its vertical axis. ratio), with FPR (false positive ratio) serving as its horizontal axis. The ROC diagram is depicted in Figure 2. As is evident, It is assumed that the suggested approach is better than alternative approaches. because of the diagram's higher level in relation to the suggested approach. Table 5 shows the confusion matrix, which consists of binary attack class confusion matrix.

Table 5: values related to parameters of step seven

		Predicted	
		Yes	No
Actual	Yes	69.35	30.65
	No	3.23	96.77

Lastly, it should be mentioned that the Apriori technique is solved using the minimal support value of 0.22 and the minimum confidence level of 0.6 chosen by the suggested method.

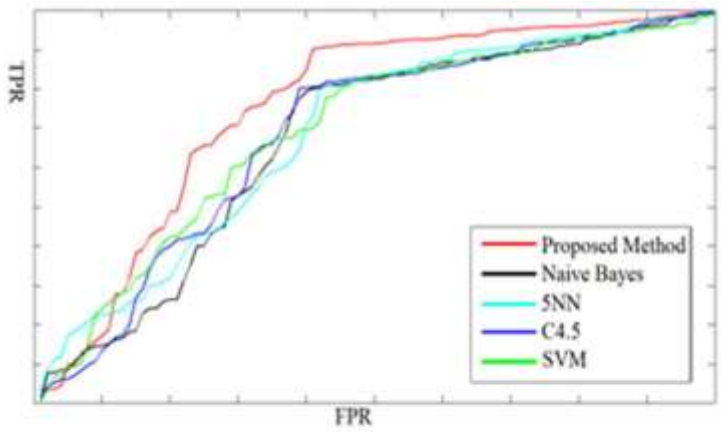

Figure 2: comparison of the proposed method with competitor methods using ROC

According to the ROC diagram, the suggested approach has a larger area below the graph. We can therefore infer that the suggested approach is better than comparable approaches based on this criterion. Lastly, it should be mentioned that the suggested approach and necessary item sets were implemented using MATLAB Version 7.

IV. SUMMARY AND UPCOMING ACTIONS

The goal of the current work is to create rules that are ideal in terms of length, quantity, and accuracy by combining the Apriori algorithm with ant colonies. There are nine fundamental steps in the suggested method. Each node in the graph that Ant Colony operates on gives seven parameters for the suggested approach. The Apriori algorithm creates the initial fuzzy rules based on the values in the most populated city in the graph associated with the ant colony. In order to increase the accuracy of intrusion detection in computer networks, the primary rules are adjusted through a process known as conventional weighted aggregation. The suggested approach is competitive with fuzzy rule-based approaches, as demonstrated by the ROC diagram and comparison in Table
4. The Apriori algorithm uses a lot of memory and storage. Due to the frequency of this algorithm and the requirement to assess all of the current approaches, it is also discovered to be quite slow. The suggested method is typically too slow when the Apriori algorithm is combined with novel techniques. Thus, the goal of this effort is to identify alternative algorithms to replace Apriori.

REFERENCES

[1] A.L. Buczak, E. Guven, “A Survey Of Data Mining and Machine Learning Methods For Cyber Security Intrusion Detection”, IEEE Communications Surveys C Tutorials, 2015.

[2] W. Meng, E. W. Tischhauser, Q. Wang, Y. Wang, J. Han, “When Intrusion Detection Meets Blockchain Technology: A Review”, IEEE Access, 2018.

[3] S. Shamshirband, N. B. Anuar, M. L. M. Kiah, V. A. Rohani, D. Petković, S. Misra, A. N. Khan, “Co-Fais: Cooperative Fuzzy Artificial Immune System for Detecting Intrusion in Wireless Sensor Networks”, Journal Of Network and Computer Applications, 2014.

[4] S. Elhag, A. Fernández, A. Bawakid, S. Alshomrani, F. Herrera, "On The Combination Of Genetic Fuzzy Systems and Pairwise Learning for Improving Detection Rates on Intrusion Detection Systems”, Expert Systems With Applications, 2015.

[5] H. J. Liao, C. H. R. Lin, Y. C. Lin, K. Y. Tung, “Intrusion Detection System: A Comprehensive Review”, Journal Of Network And Computer Applications, 2013.

[6] M. Dorigo, T. Stützle, “Ant Colony Optimization: Overview and Recent Advances”, In Handbook Of Metaheuristics, 2019.

[7] A. Inokuchi, T. Washio, H. Motoda, “An Apriori-Based Algorithm For Mining Frequent Substructures From Graph Data”, In European Conference on Principles Of Data Mining and Knowledge Discovery, 2000.

[8] W. Pedrycz, “Why Triangular Membership Functions?”, Fuzzy Sets And Systems, 1994.

[9] T. Stützle, M. López-Ibánez, P. Pellegrini, M. Maur, M. M. De Oca, M. Birattari, M. Dorigo, “Parameter Adaptation in Ant Colony Optimization”, In Autonomous Search, 2011.

[10] E. Dasseni, V. S. Verykios, A. K. Elmagarmid, E. Bertino, “Hiding Association Rules by Using Confidence and Support”, In International Workshop on Information Hiding, 2001.

[11] J. Alcala-Fdez, R. Alcala, F. Herrera, "A Fuzzy Association Rule-Based Classification Model for High-Dimensional Problems with Genetic Rule Selection and Lateral Tuning", IEEE Transactions on Fuzzy Systems, 2011.

[12] M. R. Gholamian, S. M. Sadatrasoul, Z. Hajimohammadi, "Fuzzy Apriori Rule Extraction Using Multi-

Objective Particle Swarm Optimization: The Case of Credit Scoring", Scientific Information Database, 2012.

[13] J. R. Quinlan, "C4. 5: Programs For Machine Learning", Elsevier, 2014.

[14] L. Dhanabal, S. P. Shantharajah, "A Study On NSL-Kdd Dataset for Intrusion Detection System Based on Classification Algorithms", International Journal of Advanced Research in Computer and Communication Engineering, 2015.

[15] M. S. Abadeh, J. Habibi, "Computer Intrusion Detection Using an Iterative Fuzzy Rule Learning Approach", IEEE International Fuzzy Systems Conference, 2007.

[16] A. Javaid, Q. Niyaz, W. Sun, M. Alam, "A Deep Learning Approach for Network Intrusion Detection System", In Proceedings of the 9th EAI International Conference on Bio-Inspired Information and Communications Technologies, 2016.