International

Congress on Multidisciplinary Natural Sciences

Uluslararası Multidisipliner Doğa Bilimleri Kongresi

ISBN:NNNN-NNNN

# Artificial Bee Colony Algorithm for Anomaly Based Intrusion Detection

*Murat CELIK*[1]
*Rifat KURBAN*[2]
*Tuba KURBAN*[3]

**Abstract**

Computer networks are facing an increasing number of threats. Therefore, establishing and maintaining a secure computing environment is very important. Researchers use variety of methods to ensure the security of networked systems with anomaly-based intrusion detection systems (IDS). Data classification is one of the main problems of these anomaly-based detection. Artificial bee colony algorithm is an effective optimization algorithm that models foraging behavior of bees in nature. In this paper, an artificial bee colony algorithm based, semi-supervised intrusion detection method is proposed to optimize the cluster centers and identify the best clustering solutions. Experimental studies are carried out on different sub-sets of KDD Cup 99 database to evaluate the performance of the proposed method. Test results show that the proposed algorithm can be used as a model for anomaly-based intrusion detection system.

**Keywords:** intrusion detection; artificial bee colony algorithm; data clustering

## 1. Introduction

With the widespread usage of information systems, the internet has become more popular in our daily lives. With the importance of information technology day by day, attackers who want to

[1] Lisansüstü Öğrencisi, Erciyes Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Müh. Anabilim Dalı
[2] Dr. Öğr. Üy., Kayseri Üniversitesi, Teknik Bilimler MYO, Bilgisayar Teknolojileri Bölümü, ORCID: 0000-0002-0277-2210
[3] Dr. Öğr. Üy., Gazi Üniversitesi, Mühendislik Fak., Bilgisayar Müh. Bölümü, ORCID: 0000-0001-5590-5307

benefit themselves by using the weak sides of this technology have emerged. The goal of these attackers is to infiltrate information systems and capture confidential information or render systems inoperable. Many of the attacks affects the devices such as personal computers, mobile devices and servers that connected to the internet or intranet. As a result of attacks, credit card passwords are stolen, server systems are infiltrated, and information services are rendered inoperable. In 2015, a hacker group infiltrated the ATG system used in gasoline pumps in the USA and changed the prices of gasoline pumps and caused financial damage to the company [1]. In recent history, many similar infiltration events have been encountered. Cisco, the world's leading manufacturer of network and network security solutions, detected approximately 50,000 daily network attacks in its 2014 security report [2]. The annual number of attacks reported through Cisco security systems reaches 18 million. Firewall systems have been used for many years to prevent threats from computer networks. However, these systems that only perform packet filtering are not sufficient to prevent all threats. Existing intrusion detection systems (IDS) are widely used to prevent network attacks. These systems are positioned in front of internal, networked devices that need to be protected. Thus, it is aimed to prevent harmful connections by analyzing the input and output traffic.

## 2. Intrusion Detection Systems

Intrusion detection systems can be classified into 2 groups:

### 2.1 Misuse based IDS

Signature-based intrusion detection systems, they incorporate pre-defined signatures of known attacks. Main advantage of these systems is that they can detect malicious traffic with high accuracy. Despite the low false alarm rate, a significant disadvantage is that they are vulnerable to new attacks. Because their databases do not have the signatures of new attacks. In addition, the signature databases of these systems need to be updated periodically.

### 2.2 Anomaly based IDS

Anomaly based IDS's do not have signature databases. The main advantage of these systems is easily able to detect new types of attacks. But high false alarm rate than signature-based systems, is disadvantage of these systems [3].

Anomaly-based intrusion detection systems must be able to detect malicious network flows in real time and in an optimum time. Merely, if the system can detect malicious content at the right time can ensure the security of networked systems.

## 3. Artificial Bee Colony Algorithm (ABC)

Many swarm intelligence-based algorithms that model the behavior of animals in nature are developed. Ant colony algorithm [4] and particle swarm optimization [5] are well known examples of swarm intelligence. Karaboga [6] has proposed artificial bee colony (ABC) optimization algorithm that

models foraging behavior of honeybees. These intelligent techniques are highly preferred for intrusion detection. In this paper, ABC algorithms is used to improve the data clustering process for intrusion detection.

## 4. Related Work

Researchers have used various computational intelligence [7] methods, machine learning methods and artificial intelligence optimization algorithms in the design of these systems to classify network flows as normal or abnormal. These methods are used to improve the computation time and decrease false alarm rate. Immune algorithm is used to improve the classification accuracy and to reduce the computational cost to train the classification model in intrusion detection [8]. A weighted objective function that considers trade-off between the maximizing the detection rate and minimizing the false alarm rate is proposed with particle swarm optimization algorithm [9]. An efficient hybrid clustering algorithm for intrusion detection is proposed that combines the ant lion optimization and k-means, to improve the performance of k-means and to converge to the global optimum [10]. Multiple IDS approaches based on various deep neural networks have been published in the literature to address privacy concerns and security threats [3, 11].

After creating initial values with a small percentage of labeled train data, true alarm rate can be increased by updating classifier values periodically with a metaheuristic algorithm while continuing network flow the accuracy of the Tsai's work [12] is a good example of this method.

In this study, an artificial bee colony based semi-supervised intrusion detection algorithm is proposed. Artificial Bee Colony algorithm was used in data clustering to optimize cluster centers and identify the best clustering solutions to improve true alarm rate of an intrusion detection algorithm.

## 5. Proposed Algorithm

As shown in Figure 1, the proposed algorithm is composed of three phases.

### 5.1 Creating Initial Classifiers from Labelled Data

Firstly, to classify unknown network traffic, initial clusters are created with train dataset that contains labelled patterns and the initial cluster centroids are calculated.

### 5.2 Clustering Unlabeled Data

In the second phase, a test dataset that contains unlabeled patterns are read sequentially by the proposed algorithm like a simulation of an intrusion detection system. Each pattern is saved to nearest cluster according to the distance to the cluster center. If that pattern is a new type of traffic. a new cluster is created This process continues until certain number of patterns. After reaching enough number of patterns, cluster centers and solutions are given to the optimization algorithm as input. This process will be repeated as number of patterns.

### 5.3 Optimizing Classifiers by ABC algorithm

Artificial Bee Colony algorithm produces candidate solutions based on a cost function. Best clustering solution selected as a result of a certain iteration and optimized cluster centroids is sent back to the clustering phase. Afterwards, clustering is continued with the solutions that optimized by ABC.

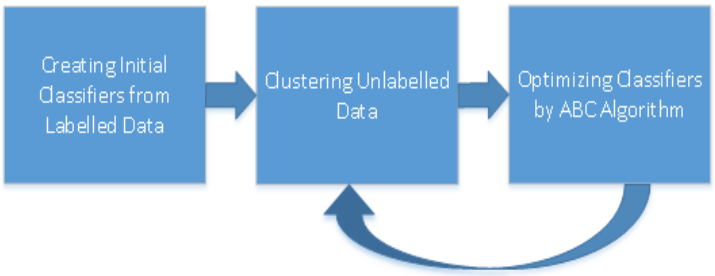


**Figure 1.** Phases of the proposed IDS algorithm.

### 5.4 Flow Chart of the Proposed Algorithm

a) Preprocessing

Each pattern of KDD 99 dataset has 41 attributes. In order to reduce computational complexity, nine attributes such as, duration, protocol_type, service, count, srv_count, serror_rate, srv_serror_rate, rerror_rate, srv_rerror_rate are selected. Tsai achieved good results with these attributes in his intrusion detection work that based on particle swarm optimization [12]. Therefore, same attributes are used in the study. Some attributes expressed as alphanumeric in the KDD Cup 99 dataset, these attributes are converted to numerical values for numerical optimization.

b) Initial Classsifiers and Cluster Centroids

Labelled patterns are *normal* or considered as an *attack* in the training data set. Afterwards, patterns are saved to related clusters according to their labels, and cluster centroids ***(m)*** are calculated. Also, average distance $(d_m)$ between cluster centroid and patterns that belong to that cluster is calculated using the Mahalanobis distance method.

c) Food Sources from Initial Classifiers for ABC

Unlike the classic ABC algorithm, initial food sources are not produced completely random. However, to increase the diversity of the population, each initial parameter $(D_i)$ of the classifiers, are added with randomly generated numbers in ***[1,1]*** based on a uniform distribution. There are $\boldsymbol{K}$ number of cluster centers with $\boldsymbol{D}$ parameters. Thus, each food source represented as $\boldsymbol{K} * . \boldsymbol{D}$ dimension of vectors.

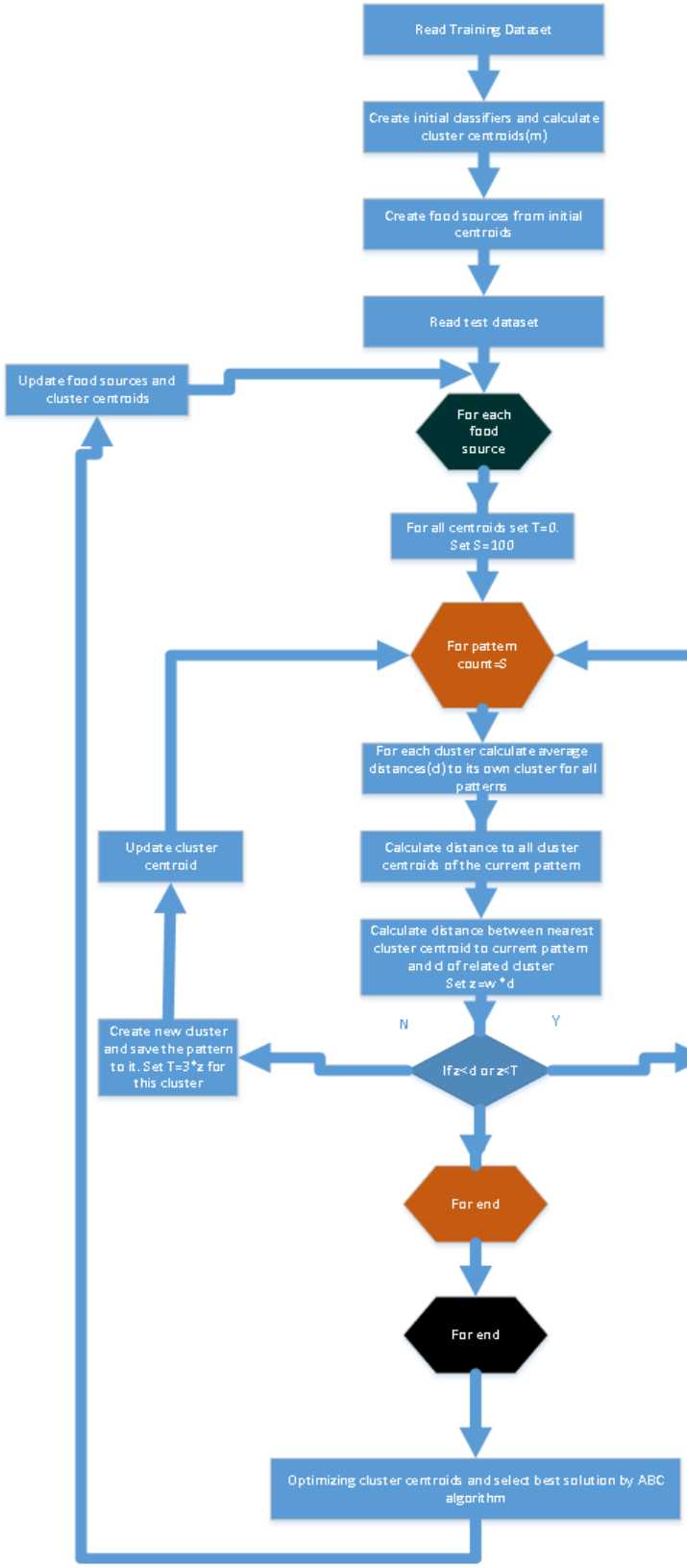


**Figure 2.** Flowchart of the proposed IDS algorithm.

d) Clustering Patterns of the Test Dataset

Each unsigned pattern that enters the intrusion detection system is saved to an existing cluster or saved a new cluster is created. Two threshold values are used for clustering process. First threshold value, z, is calculated with Equation (1). And the second threshold value is T that used to create a new cluster. To calculate membership of a pattern to the cluster, $d_m$ parameter that weighted with a coefficient $(\omega)$ is used for similarity measure. Unreported several experiments show that $\omega = \mathbf{10}$ produces best results. Initial value of $T = 0$.

$$z_m = \omega * d_m \qquad (1)$$

Nearest cluster center distance $(d_x)$ is calculated for each pattern $(x)$ that enters the system. If $d_x \leq z_m$ then the pattern saved to cluster $m$. Otherwise a new cluster is created and pattern $x$ saved to that new cluster. Also the distance threshold value $(T_m)$ of the new cluster is set to $T_m = 3 * z_m$. Then centroid and $(d_m)$ of the related cluster is updated. This process continues until the read sample number S = 100. The operations made up to sample S is repeated for each food source that initially produced.

e) Optimizing Food Sources

Initial solutions and updated solutions based on the test samples are used as the initial solutions for ABC. Quality of food sources (candidate solutions) are calculated with Equation (2) where $i$ is the food index, $k_i$ is cluster number, $\alpha_i^j\ j$ is the cluster, $m_i^j\ j$ is the cluster center. The objective is to minimize the inter-distance of clusters.

$$f_i = \sum_{j=1}^{k_i} \sum_{\forall x \in \alpha_i^j} ||x - m_i^j||^2 \qquad (2)$$

Table 1. shows control parameters of ABC algorithm.

**Table 1.** Control parameters of ABC algorithm

| Number of Food Sources | Swarm Size | Number of Iterations | Limit | Number of Parameters to be Optimized |
|---|---|---|---|---|
| 10 | 20 | 100 | 25 | 9 |

## 6. Experimental Results

KDD CUP 99 dataset [13] was created by Massachusetts Institute of Technology. This set has 494.000 patterns of network connection vector. Each pattern has 41 attribute that expressing a normal or an abnormal connection. Dataset has 20% normal and %80 attack patterns [14]. There are 21 different types of attacks in the dataset. These are named in the dataset; teardrop, satan, nmap, rootkit, neptune, guess_passwd, land, bufferoverflow, warezmaster, ipsweep, multihop, warezclient, perl, back, ftp_write, loadmodule, spy, imap, smurf and pod. Five subsets are prepared for experimental studies that include variety of patterns from KDD CUP 99 to test the proposed algorithm. Firstly, the algorithm is trained with 10% part of dataset described in [12]. Also, K-Means clustering method performance is tested on sub datasets to compare with the performance of the proposed algorithm. For each test, with K-Means clustering; $k_{max}$ is set to proposed algorithms maximum cluster number. Additionally, in fine tuning stage of clustering process, to compare the performance of two metaheuristic algorithm, PSO and ABC algorithms are used. For PSO particle number is set to 10, iteration number is set 100, balance factors is set $c_1 = c_2 = 1.49$ and inertia weight is set to $\omega = 0.72$[15]. The results are average values obtained with run 30 times for all algorithms. Simulation results of all algorithms are shown in Table 2.

**Table 2.** Simulation results.

| Test | Number of Attack Types | | Number of Patterns | Accuracy Rate (%) | | |
|---|---|---|---|---|---|---|
| | Train | Test | | ABC | PSO | K-Means |
| Test 1 | 2 | 4 | 150 | **63.44** | 57.66 | 62.74 |
| Test 2 | 2 | 3 | 300 | 48.50 | 62.16 | **89.15** |
| Test 3 | 1 | 2 | 450 | 85.62 | 61.48 | **91.47** |
| Test 4 | 5 | 7 | 900 | **79.95** | 55.29 | 71.22 |
| Test 5 | 2 | 6 | 1800 | **80.76** | 39.60 | 78.63 |

## 7. Conclusions

In this paper, we tested a semi-supervised algorithm based on Artificial Bee Colony Algorithm for intrusion detection system. According to five tests, true clustering of the proposed algorithm has achieved approximate success against K-Means.

PSO algorithm is used with proposed method as a metaheuristic in some experiments, instead of ABC algorithm. Experimental results show that ABC has better performance than PSO. The number of patterns used for training is important to increase the accuracy of the method. In this study, proposed method that dynamically update classifiers with the help of ABC algorithm, detect the unknown attacks, is performed effectively. Experimental results shows that proposed algorithm can be used as an anomaly-based intrusion detection system.

## References


[1] News of The Register **http://www.theregister.co.uk/2015/02/11/anonymous_hacks_fuel_station_monitoring_system/** (Access Date:2015 Dec )

[2] *Cisco 2014 Annual Security Report,* **http://www.cisco.com/web/offer/gist_ty2_asset/Cisco_2014_ASR.pdf,** (Access Date: 2015 Dec)

[3] Lee, S.-W., Mohammed sidqi, H., Mohammadi, M., Rashidi, S., Rahmani, A.M., Masdari, M., and Hosseinzadeh, M., 2021, Towards secure intrusion detection systems using deep learning techniques: Comprehensive analysis and review, **Journal of Network and Computer Applications**, 187, pp. 103111.

[4] Shokouhifar, M., 2021, FH-ACO: Fuzzy heuristic-based ant colony optimization for joint virtual network function placement and routing, **Applied Soft Computing**, vol. 107, p. 107401.

[5] Kan, X., Fan, Y., Fang, Z., Cao, L., Xiong, N.N., Yang, D., and Li, X., 2021, A novel IoT network intrusion detection approach based on Adaptive Particle Swarm Optimization Convolutional Neural Network, **Information Sciences**, 568, pp. 147-162.

[6] Hajimirzaei, B., Navimipour, N.J., 2019, Intrusion detection for cloud computing using neural networks and artificial bee colony optimization algorithm, **ICT Express** 5(1): 56-59.

[7] Shamshirband, S., Fathi, M., Chronopoulos, A.T., Montieri, A., Palumbo, F., and Pescapè, A., 2020, Computational intelligence intrusion detection techniques in mobile cloud computing environments: Review, taxonomy, and open research issues, **Journal of Information Security and Applications**, 55, pp. 102582

[8] Wei, W., Chen, S., Lin, Q., Ji, J., and Chen, J., 2020, A multi-objective immune algorithm for intrusion feature selection, **Applied Soft Computing**, 95, pp. 106522.

[9] Hosseini Bamakan, S.M., Wang, H., Yingjie, T., and Shi, Y., 2016, An effective intrusion detection framework based on MCLP/SVM optimized by time-varying chaos particle swarm optimization, **Neurocomputing**, 199, pp. 90-102.

[10] Chen, J., Qi, X., Chen, L., Chen, F., and Cheng, G., 2020, Quantum-inspired ant lion optimized hybrid k-means for cluster analysis and intrusion detection, **Knowledge-Based Systems**, 203, pp. 106167

[11] Kunang, Y.N., Nurmaini, S., Stiawan, D., and Suprapto, B.Y., 2021, Attack classification of an intrusion detection system using deep learning and hyperparameter optimization, **Journal of Information Security and Applications**, 58, pp. 102804

[12] Tsai, C.,Incremental particle swarm optimisation for intrusion detection, **IET Networks, 2(**3): 124-130, 2012

[13]KDD Cup 99 Dataset **http://kdd.ics.uci.edu/databases/kddcup99/,** (Access Date: 2015 May)

[14]Tavallaee, M.,Bagheri, E., Lu, W., Ghorbani, A.A., 2009, A detailed analysis of the KDD CUP 99 data set. **Computational Intelligence for Security and Defense Applications**, pp. 1-6

[15]Merwe, D.,Engelbrecht A., 2003, Data clustering using particle swarm optimisation, **Congress on Evolutionary Computation,1**: 215-220