Using A Hybrid Algorithm and Feature Selection for Network Anomaly Intrusion Detection

Ali Hussein Shamman Al-Safi†, Zaid Ibrahim Rasool Hani†, Musaddak M. Abdul Zahra‡,*

†Computer Techniques Engineering Department, Al-Mustaqbal University College, Hilla, Babil, Iraq
‡Electrical Engineering Department, College of Engineering, University of Babylon, Hilla, Babil, Iraq.

*Corresponding Author Email: musaddaqmahir@mustaqbal-college.edu.iq

ABSTRACT

Today's, networks security of has become the important problem in each distributed system. A lot of attacks are becoming less able to detect with software of antivirus and firewall. For improving the security, intrusion detection systems (IDSs) are utilized for detecting the anomalies in traffic of network. Network anomaly detection issue is determining, if incoming traffic of network is anomalous/legitimate. The automated system of detection schemed for identifying the incoming anomalous patterns of traffic usually apply widely utilized techniques of machine learning. In the article, we have utilized the Information Gain-based algorithm. The algorithm chooses the features optimal number from dataset of NSL-KDD. Additionally, we have integrated selection of feature with the technique of machine learning namely as Support Vector Machine (SVM) by utilizing the algorithm of artificial bee colony as well as Optimization-Cuckoo Search Algorithm for optimizing SVM hyper parameters for dataset effective classification. Proposed method performance has been assessed on the modern intrusion dataset as NSLKDD. Experimental results show that the proposed method outperforms also achieves high accuracy in comparison to the other modern techniques in NSLKDD.

KEYWORDS

Intrusion detection systems (IDS), Anomaly intrusion detection, Cuckoo Search Algorithm (CSA), Feature Selection (FS), artificial bee colony algorithm (ABC), Support Vector Machine (SVM), NSL-KDD Dataset.

INTRODUCTION

Intrusion detection systems (IDSs) are widely utilized elements in infrastructure of network security. They protect the networks with detecting signature and anomaly-based intrusions. On the basis of methods of attack detection, Intrusion detection systems are able to be placed into three groups: specification, anomaly and misuse-based [1]. Detection based on Misuse has ability for detecting known attacks patterns as datasets of IDS has their signatures record. System main problem which relies on the matching of pattern is the fact that they are not able to detect the unknown intrusions since their signatures do not exist in database of IDS [2]. Premise behind IDS detection based on anomaly is that process of attack is able to create the behavior which is various from usual behavior of user. Studying the pattern of attack is able to aid the IDS for detecting new attacks. Detection based on anomaly came to conquer the problem as this is able to detect each suspicious activities which are deviated from usual system operations. Algorithms of machine learning are one of such feasible techniques where created system can learn for making decisions with no explicitly programmed and with minimal intervention of human involved in process.

Algorithms of machine learning are able to be grouped for unsupervised, supervised, as well as the algorithm of hybrid learning [3]. Between the algorithms of supervised learning, support vector machine (SVM) is according to statistical theory of learning and it is the most effective algorithm to predict performance for the nonlinear issues. This has high capabilities of generalization. Support vector machine is the powerful technique to overcome the problems of classification, processing of image, detection of intrusion with great performance. In the article, we have utilized the algorithm according to technique of Information Gain for the selection of feature. The algorithm chooses the features optimal number from dataset of NSL-KDD. Additionally, we have
combined selection of feature with technique of machine learning as SVM. Setting of parameter is critical for SVM solution accuracy and efficiency. We utilize the algorithm of artificial bee colony [4] as well as the algorithm of Optimization-Cuckoo Search [5] for optimizing SVM parameters for the effective NSL-KDD dataset classification. This paper is organized as below. In next part, studies performed on systems of Intrusion Detection utilizing intelligent approaches of swarm as well as Machine Learning in literature are briefed. In third part, the proposed method is presented. Experimental methodology as well as results is provided in fourth part. Final remarks are presented in final part.

RELATED WORK

Recent related literature is rich in works of research coping with the intelligent IDSs by the algorithms of machine learning in network traffic classification/intrusion detection. Historically, the most challenging technological obstacle while coping with IDSs deployed on gateways of network is represented by restricted computational power that makes it hard for running full-fledged IDS. Most of the information asymmetric circulation as well as distinguishing boundaries among usual and unusual issues of parameters are cleared by the approach of feature selection, also this is usually utilized before algorithms of ML for increasing training efficiency. In this part we study works performed in Intrusion Detection Systems field by utilizing approaches of swarm intelligent as well as techniques of Machine Learning.

In [2], schemed the detection based on anomaly, with adopting modified Cuckoo Search Algorithm (CSA) named Mutation Cuckoo Fuzzy (MCF) for the selection of feature and Evolutionary Neural Network (ENN) for classification. The proposed search algorithm utilizes the mutation for more properly examine space of search for letting candidates for escaping the local minima. In addition, solution value is assessed according to objective function as well as clustering method of Fuzzy C Means (FCM) utilized for providing the best results for overlapping dataset as well as creating search domain of fuzzy membership that contains the whole feasible compromise solutions. The proposed model has been utilized practically to intrusion detection issue and been validated by utilizing dataset of NSL-KDD. Experimental results show that decreasing the features with using and choosing the most relevant features are able to improve the time of execution and simultaneously increase IDS performance and efficiency.

In [6], the new hybrid IDS has introduced with the Adaptive Grasshopper Optimization Algorithm (AGOA) and Ensemble of Feature Selection (EFS) combination for anomalies classification and detection, resulted from the attack in network of computer. At first, the supervised feature selection method based on filter is described as ensemble FS that integrates features ranking created by various filters that is able to aid for eliminating irrelevant features. But, GOA is quickly trapped in local optima also premature convergence seems while applied to the sophisticated issues. For overcoming the problem, they have introduced GOA adaptive behavior named AGOA which is able to contribute for accurately predicting behavior of networks traffic. Additionally, technique of AGOA is applied for selecting the accurate parameters of SVM that escape SVM over-fitting concern. The proposed model has presented the low false alarm and high rate of detection.

In [7], presents the algorithm of wrapper feature selection for system of intrusion detection. The algorithm uses Dynamic differential annealed optimizer (DDAO) for controlling selected method. The proposed algorithm is standardized to 51 functions of test. Algorithm of Dynamic differential annealed optimizer is correlated with high quotes optimization algorithms count. The proposed method has outdistanced these methods; some cases with great performance have been illustrated. There are issues of spring designs and guarded path planning practical engineering optimization is selected as an issue. DDAO became global lesser efficiency of problem, DDAO has found the best possible solution for issue of spring design observed by different mechanisms.

In [8], presents the novel binary model of classification for intrusion detection according to Artificial Bee Colony algorithm (ABC) and Dragonfly algorithm (DA) hybridization to train the artificial neural network (ANN) for increasing accuracy rate of classification for the non-malicious and malicious traffic in networks. At first model chooses appropriate weights and biases by using hybrid (ABC) and (DA). After that, neural network is retrained by utilizing such ideal values for model of intrusion detection to be able to find novel attacks. The
other ten metaheuristic algorithms were adapted for training neural network as well as their performances were compared with that of the proposed model. Additionally, four kinds of datasets of intrusion detection evaluation were applied for assessing the proposed model in comparison to others. The experiments results have shown the important development in inefficient network intrusion detection over the other methods of classification.

In [9], schemed intelligent technique of intrusion detection by using algorithm of opposition self-adaptive grasshopper optimization according to perceptive and mutation strategy for identifying the novel bad features. Reinforcement learning is applied in support vector machine as gain actor critic with the machine of support vector as the function of fitness in proposed method for obtaining the low false positive and negative rates, high detection rate. Technique based on gradient is applied in dynamic environment so that agent learns rapidly more specific policies of action selection. For proving proposed methodology efficacy, simulation results have estimated on three various datasets like CIC-IDS 2017, AWID and NSL-KDD concerning rate of detection as well as the rate of false-positive. The proposed model has presented high rate of detection, accuracy and low rate of false-positive in NSL-KDD.

In [10] perform the algorithm of feature selection for intrusion data of network for detecting intrusions on traffic of real time network by utilizing real time and high accuracy speed. They apply Cuckoo Search (CS) as the algorithm of feature selection into three datasets of intrusion: Botnet ISCX 2017, NSL-KDD, KDD Cup 99. They compare Cuckoo Search (CS) algorithm performance to the other two Evolutionary Algorithms: Particle Swarm Optimization (PSO) and Genetic Algorithm (GA). Experiments illustrate that in dataset of NSL-KDD, algorithm of CS decreases attributes number from 41-9. In the case of performance classification, PSO is superior to CS and GA in intrusion datasets of NSL-KDD.

THE PROPOSED METHOD

One of the essential issues faced in a lot of algorithms utilized for the selection of feature is high computational complexity of time. Because of this, in the paper, Information Gain was utilized for selecting features because of basic IG benefit in comparison to the other algorithms. SVM is the technique of machine learning, utilized for classification of data. Basic purpose is finding hyperplane which well separates points of data with the maximum margin hyper-plane [11]. Parameters in item of regularization as well as function of kernel are known as hyper-parameters in SVMs that play the important role in performance of algorithm. Existing techniques to adjust hyper-parameters in SVMs are able to be briefed into two types: one is according to analytical techniques; another is according to the heuristic searches. First type of techniques determines hyper-parameters with some generalized error measures gradients [12]. Second type of techniques determines hyper-parameters with state-of-the-art heuristic algorithms containing the genetic algorithms, simulated annealing algorithms as well as the other evolutionary strategies. This article utilizes the algorithm of artificial bee colony [4] and algorithm of Optimization-Cuckoo Search [5] for optimizing SVM parameters. The proposed IDS model contains three stages:

• **Pre-processing stage**: In this part, dataset of NSL KDD is preprocessed with transforming symbolic valued attributes for numeric also applying algorithm of discretization.

• **Feature selection stage**: In this part, Information Gain is employed for the selection of feature.

• **Post-Processing stage**: Classification/optimization stage, now, SVM is utilized for the classification. Parameters of SVM are chosen by algorithm of Cuckoo Search (CS) and artificial bee colony (ABC). The algorithm of Hybrid Particle Swarm Optimization-Cuckoo Search. Fig. 1 illustrates the proposed system diagram.
Selection of Feature

In the study, we utilized techniques of Information Gain (IG) for the selection of feature. The method depends on approach based on filter. Approach based on filter is applied currently, containing Information Gain (IG), Gain Ratio (GR), Correlation Feature Selection (CFS). Though rapid and simple, such methods do not always propose the improved rate of detection/detection stage accuracy but aid classifier for seeking the better accuracy.

Artificial bee colony algorithm -Cuckoo Search Algorithm with Lévy Flights

In algorithm of ABC, artificial bees colony includes three bees groups: scouts, onlookers and employed bees. First half of colony includes employed artificial bees and second half includes onlookers. For each source of food, there is just one employed bee. On the other hand, employed bees number is equal with food sources number. Abandoned food source employed bee becomes scout. Search performed by artificial bees is able to be briefed as below:

- Employed bees determine the source of food in food source neighbourhood in their memory.
- Employed bees distribute their information with the onlookers in hive and onlookers choose one sources of food.
- Onlookers choose the source of food in food sources neighbourhood selected by themselves.
the employed bee that source has been abandoned becomes scout also begins to randomly search novel source of food

Every search cycle includes three stages: moving onlooker and employed bees onto sources of food as well as computing their nectar amounts as well as determining scout bees after that randomly moving them onto feasible sources of food. The source of food represents feasible solution to an issue in order to be optimized. Food source nectar amount corresponds to solution quality which is represented by that source of food. Onlookers are placed on foods utilizing the method of “selection of roulette wheel”. Each colony of bee has scouts which are explorers of colony. Explorers do not have any guidance when seeking for food. They are basically concerned with finding each type of source of food. As a result of this behavior, scouts are characterized by the low costs of search also the low average in quality of food source. Occasionally, scouts are able to discover entirely unknown and rich food sources accidentally. In artificial bees case, artificial scouts could have feasible solutions group fast discovery as the task. In algorithm of ABC, one of the employed bees is chosen and grouped as scout bee. Classification is handled by parameter of control named “limit”. If the solution representing the source of food is not developed by trials predetermined number after that source of food is abandoned by the employed bee as well as employed bee related to that source of food becomes scout. Trials number to release the source of food is equal to “limit” value that is the essential ABC algorithm control parameter [4].

The ABC creates the randomly distributed SN solutions initial population (sources of food), that SN shows size of swarm. Allow \( X_i = \{x_{i1}, x_{i2}, \ldots, x_{iD}\} \) represent solution of \( i \)th in swarm, that \( D \) is size of dimension. Every employed bee \( X_i \) creates the novel solution of candidate \( V_i \) in its present position neighborhood as below [13]:

\[
v_{i,j} = x_{i,j} + \phi_{i,j}(x_{i,j} - x_{k,j}) \tag{1}
\]

After that utilizing nectar information, employed bees calculate value of fitness for every source of food by utilizing Equation (2):

\[
f_{it_j} = \frac{1}{1 + f_j} \tag{2}
\]

That \( f_j \) is subset \( S_j \) objective value. \( f_j \) is according to subset indiscernibility relation to classes. After subsets fitness values are known, onlookers gain information from employed bees also chooses the subset for exploitation. Onlooker bee pointing to the specific subset chooses the subset in neighborhood according to probability that is computed by utilizing Equation (3)

\[
P_j = \frac{f_{it_j}}{\sum_{n=1}^{m} f_{it_n}} \tag{3}
\]

In Equation (3), \( f_{it_j} \) is neighborhood feature subset \( S_j \) fitness that onlooker has chosen. While the subset of feature is chosen, onlooker creates the novel solution \( V_{ij} \) with utilizing Equation (4):

\[
V_{ij} = x_{ij} + \phi_{ij}(x_{ij} - x_{kj}). \text{ Lévy} \tag{4}
\]

Random walk through flight of Lévy is more effective in exploring space of search as the length of its step is much longer in long run. Due to that the distribution of power-law is often linked to several features which are scale-free, flight of Lévy is able to illustrate fractal behavior and self-similarity in patterns of flight. Naturally, with ABC performance improving purpose, flight of Lévy is taken to replace traditional ABC algorithm random searching method into account. So, modified algorithm of ABC is known as ABC-CS.

Lévy Flights and Cuckoo Search

Cuckoo Search main stage. Cuckoo Search is the type of nature inspired algorithms of metaheuristic; special cuckoo species aggressive reproduction strategy stimulates CS proposal. Three idealized laws [are delimited, last law means to introduce several novel solutions in algorithm. This is able to be approximated by \( n \) host nests friction \( pa \) for producing novel nests [5].
Problem of optimization for being solved is defined as the function of objective ($X$), $X = \{x_1, \ldots, x_D\}$ in D-dimensional space. For elaborating ABC-CS algorithm constructional details, we normalized CS with ABC variables. There are $N$ host nests $\{X_i, i = 1, \ldots, N\}$ in specified space of search. Every nest $X_i = \{x_{i1}, \ldots, x_{id}\}$ (representations are similar as onlooker $X_i$ in ABC) presents the feasible optimization problem solution to be solved. One CS main states is seeking for novel nests $X(t+1)$. population. Moreover, novel nests are achieved by utilizing flight of Lévy as below:

**Levy Flights.** flight of Lévy is one type of random walk with length of stage that has distribution of Lévy. This is defined to algorithm of CS for getting the scale-free search pattern of Lévy -flight-style intermittent. In [9], this is illustrated that flights of Lévy can increase resource searches efficiency in uncertain environment. Distribution of Lévy is described as

$$L(s, \gamma, \mu) = \left\{ \begin{array}{ll} \frac{\gamma}{2\pi} \exp\left[-\frac{\gamma}{2(s-\mu)}\right] \frac{1}{(s-\mu)^{2\gamma}} & 0 < \mu < s < 0 \\ 0 & \text{otherwise} \end{array} \right. \quad (5)$$

That $\mu > 0$ is the minimum stage and $\gamma$ is the parameter of scale. In algorithm of Mantegna, length of stage $s$ is able to be computed by

$$s = \frac{u}{|v|^\frac{1}{\beta}} \quad (6)$$

where $u$ and $V$ are derived from usual distribution. Which is,

$$u \sim N(0, \sigma_u^2) \quad (7)$$

$$v \sim N(0, \sigma_v^2) \quad (8)$$

where $\Gamma(z)$ is function of Gamma $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$. In a case when $z = n$ is the integer, $\Gamma(n) = (n - 1)!$.

So, formed algorithm of ABC-CS is detailed as below.

**Stage 1.** Initialize parameters of Bee Colony

**Stage 2.** Every employed bee creates the subset of hyper parameter (binary bit string), exploits it

**Stage 3.** Every Onlooker bee chooses subsets of hyper parameter, assesses the fitness by passing them to classifier, creates novel subsets of hyper parameter also exploit them

**Stage 4.** Determine subsets of hyper parameter for being neglected and determine the employed bee as scout to create novel sunset of hyper parameter

**Stage 5.** Memorize the best subset of hyper parameter subset that is produced

**Stage 6.** Repeat stages 2-5 for the pre-determined iterations number

Support vector machine

Support vector machine were presented by Vapnik (1995) initially to solve classification and regression analysis problems. Support vector machine is the technique of supervised learning which is trained for classifying various data groups from different disciplines. This has been utilized for problems of two-class classification and are applicable on both non-linear and linear data tasks of classification. Support vector machine produces the multiple hyperplane/hyperplanes in high-dimensional space, the best hyperplane in them is one which shares data into various classes optimally with the largest separation among classes. The non-linear classifier utilizes different functions of kernel for estimating margins. Basic kernel function’s objective is maximizing margins among hyper-planes. Currently, a lot of highly promising applications have been improved by researchers due to
maximizing interest in SVMs. Support vector machine has been widely utilized in processing of image and applications of pattern recognition [14]–[18].

PLATFORM OF EVALUATION

Assumptions of Problem contain criteria of evaluation, utilized datasets, intended methods for comparison. In this part, several tests are carried out on dataset of NSL-KDD and important results are extracted from achieved outputs. Whole tests utilized the system with processor of Intel (R) Core (TM) i7-2500K with GB12 memory, additionally, Matlab2018 is applied for the software of simulator that is run on Windows 10 64bit.

Initialization of Parameters

For evaluating the proposed algorithm quality, we compared parameters based on parameters paper [9]. In this way, for maximum repetition, we’ve taken maxIter parameter to be equals to 100 and Population number to be equal to 10. In Table 1., settings according to parameters are illustrated (ABC-CS).

**Table 1.** Initial values for parameters in ABC-CS

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial values</th>
</tr>
</thead>
<tbody>
<tr>
<td>numPopulation</td>
<td>10</td>
</tr>
<tr>
<td>maxIter</td>
<td>100</td>
</tr>
<tr>
<td>limit</td>
<td>10</td>
</tr>
<tr>
<td>φ</td>
<td>0.1</td>
</tr>
<tr>
<td>β</td>
<td>1.5</td>
</tr>
<tr>
<td>σV, σu</td>
<td>1</td>
</tr>
</tbody>
</table>

Criteria of Evaluation

Criteria achieved from two-classes confusion matrix

Matrix of confusion has information on predicted and real classification. Efficiency of model is able to be assessed according to this matrix data. If data are two classes of normal and abnormal, this data confusion matrix is as below which is illustrated in Table 2.

**Table 1.** Matrix of confusion

<table>
<thead>
<tr>
<th>The detected class</th>
<th>Abnormal sample</th>
<th>Normal sample</th>
<th>Real class</th>
</tr>
</thead>
<tbody>
<tr>
<td>FN</td>
<td>TP</td>
<td>Normal sample</td>
<td></td>
</tr>
<tr>
<td>TN</td>
<td>FP</td>
<td>Abnormal samples</td>
<td></td>
</tr>
</tbody>
</table>

**TN (true negative):** abnormal samples number which were detected correctly.

**FN (false negative):** normal samples number which were detected as abnormal.

**TP (true positive):** normal samples number which were detected correctly.

**FP (false positive):** abnormal samples number which were detected as normal.

1. **Accuracy Correct Classification Rate (ACC):** samples rate which are predicted correctly.
Using A Hybrid Algorithm and Feature Selection for Network Anomaly Intrusion Detection

\[ \text{ACC} = \frac{TN + TP}{FP + FN + TP + TN} \]  
\[ (9) \]

2. Reliability Percentage to output (precision): criterion is described as below:

\[ \text{precision} = \frac{TP}{FP+TP} \times 100 \]  
\[ (10) \]

3. Recall: this illustrates algorithm success percentage in detecting samples according to every class.

\[ \text{Recall} = \frac{TP}{FP+TP} \times 100 \]  
\[ (11) \]

Dataset of Intrusion detection

We utilized dataset of NSL KDD to evaluate the proposed method in intrusion detection field into networks of computer. Dataset is presented for solving some KDD'99 dataset inherent problems. Below, the summary of dataset description is given. 41 features are accessible for every sample in database of NSL-KDD. Four various attacks classes, like PRB (proving), DOS (Denial-Of-Service), U2R (User-to-Root), R2L (Remote-to-Local) have been detected that should be distinguished from normal samples.

- **U2R class**: attacks class that first an attacker achieves the user ID of system after that takes vulnerable parts advantage for gaining access to root of system.

- **R2L class**: In this state, logs of attacker into system as the user by sending packets of data through network and gains local access to this.

- **DOS class**: In this attack kind, an attacker disturbs system performance completely by making different resources of system like memory busy and computing resources. Therefore, system cannot respond to the legal requests.

- **PRB class**: In this attack kind, an attacker explores network for collecting information and detecting the vulnerabilities. After that, by having information according to machines and services in network, this is able to detect behaviors of network easily.

Generally, there are 22 known attacks in dataset of train that are grouped into these four class. Intrusion detection datasets samples number is illustrated in Table 3.

**Table 2. Intrusion detection dataset samples number**

<table>
<thead>
<tr>
<th>Sum</th>
<th>DOS</th>
<th>U2R</th>
<th>R2L</th>
<th>Probe</th>
<th>Normal</th>
<th>classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>125973</td>
<td>45297</td>
<td>52</td>
<td>995</td>
<td>11656</td>
<td>67343</td>
<td>NSL-KDD Train+</td>
</tr>
<tr>
<td>22544</td>
<td>7458</td>
<td>67</td>
<td>2887</td>
<td>2422</td>
<td>9710</td>
<td>NSL-KDD Test+</td>
</tr>
</tbody>
</table>

Test results Analysis

In the article, a train and test data set is utilized that includes two target values like abnormal ad normal. Each known attacks kinds are grouped as abnormal traffic when rest of traffic of network is grouped as normal traffic. Basic dataset like NSL-KDD has 41 features and one label. Stage of preprocessing is carried out on NSL-KDD. As you can see in Table 4, results demonstrate that the proposed method accuracy equals to 89.71% (Fig. 1) that is improved 13.49% in comparison to the basic method of paper. Reason for the improvement is able to be taken utilizing algorithm of ABC-CS. In the research, we attempted for using ABC-CS algorithm procedure for hyper parameter optimization in SVM.
Using A Hybrid Algorithm and Feature Selection for Network Anomaly Intrusion Detection

Table 4. The proposed method and the paper [9] method Comparison results

<table>
<thead>
<tr>
<th></th>
<th>paper [9]</th>
<th>Proposed method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>91.7</td>
<td>94.21</td>
</tr>
<tr>
<td>Recall</td>
<td>94.00</td>
<td>98.25</td>
</tr>
<tr>
<td>Precision</td>
<td>92.00</td>
<td>94.99</td>
</tr>
</tbody>
</table>

CONCLUSION

In the article, the method is presented by combining the algorithm of artificial bee colony algorithm as well as the algorithm of Optimization-Cuckoo Search for optimizing SVM parameters for effective dataset classification. Experimental results illustrate that this model is able to improve intrusion detection accuracy and increase intrusion detection system efficiency. Additionally, this has been seen that the proposed model has better efficiency in comparison with basic model of paper. Results illustrate that the proposed method accuracy equal to 94.21%, that is improved 2.51 % in comparison with basic method. So, future work should concentrate on reducing selected features number as well as proposed model application for developing the effective IDS.

REFERENCES


Using A Hybrid Algorithm and Feature Selection for Network Anomaly Intrusion Detection


