

A New Method for Intrusion Detection Using Genetic Algorithm and Neural Network

M. Hosseinzadeh Moghaddam¹; S. Javad Mirabedini²; T. banirostam³

Received (2017-09-12) Accepted (2017-10-28)

Abstract — In order to provide complete security in a computer system and to prevent intrusion, intrusion detection systems (IDS) are required to detect if an attacker crosses the firewall, antivirus, and other security devices. Data and options to deal with it. In this paper, we are trying to provide a model for combining types of attacks on public data using combined methods of genetic algorithm and neural network. The goal is to make the designed model act as a measure of system attack and combine optimization algorithms to create the ultimate accuracy and reliability for the proposed model and reduce the error rate. To do this, we used a feedback neural network, and by examining the worker, it can be argued that this research with the new approach reduces errors in the classification.

Index Terms — Intrusion Detection System, Neural Network, Genetic Algorithm, Clustring and firewall.

1- Faculty of Engineering, Islamic Azad University, Central Tehran Branch.

(moh.hosseinzadehmoghadam.eng@iauctb.ac.ir)

2,3- Assistant Professor of Islamic Azad University, Central Tehran Branch.

1. INTRODUCTION

With the rapid development of communication and information technology and its applications, especially in computer networks, there is a new competition in information security and network security. Despite the various tools and software installed on the computer, a number of unauthorized people are also infected with computers that are connected to the network. To establish security, determining appropriate access levels and identifying vulnerabilities to gain access to computer networks has become more important than the past.

Even a computer system is exposed to high risk when connected to a network, including threats to viruses and infiltrations for a computer system. Viruses can be controlled extensively by installing antivirus software and updating regularly.

Also, any unauthorized access to a computer's resource is said to be infiltration; in defense of these threats, many security techniques have been developed over the past decades. One of these techniques is cryptography or firewall. The main purpose of Intrusion Detection Systems is not only to prevent attacks, but their task is to detect and possibly detect attacks and security problems in a system or computer network that is important to knowing for the administrator.

This proposed technique is a hybrid method of two genetic algorithms and a neural network, which is used for intelligence algorithms. These methods in the intrusion detection system and the flexibility of these systems against a variety of attacks, as well as the intelligent methods used in this article, find the power of discovery and intrusions that were previously incapable of using existing methods. But intelligent intrusion detection systems have the ability to learn and can analyze packets entered into the analysis network and detect ordinary and abusive users. According to our studies, it has been shown that intelligent methods have been able to detect, in an acceptable manner, attacks and infiltrations carried out on a network. The most important advantage of smart methods is the ability of these methods to attack or influence, which previously had no information about them and their type of behavior.

In section 2, a detailed overview of the related literature is given. is reviewed, in section 3, the proposed method, and its components are detaild along with working principles of this genetic algorithm and neural network.in section 4, we evaluate experiments the research methodology and We conclude this paper in section 5.

2. LITERATURE OVERVIEW

So far, a number of researchers have been using genetic and neural network algorithms to detect intrusions. Some of these studies seek to provide better results in achieving useful patterns in intrusion detection systems.

Alshush et al. [1] and [10] provided a fuzzy approach to reducing the rate of non-recognition correctly using the correlation method of events, which is a complete avoidance of computer attacks, the IDS to reduce the damage caused by computer attacks, There is a very important role to play here. There are two methods of intrusion detection:

- Abuse method
- Abnormal behavior behaviors

Alshush in his article on Intrusion Detection System introduced the CITDS 1 intelligent method for the two methods of intrusion detection. The result of this paper is that the unique motorcycle performance in this method is rarely satisfactory, and also in current CIDS Two outstanding points are raised: 1) The CIDS system architecture 2) There is a communication alert algorithm that explains and compares the system for use of CIDS and increases the multidimensional security system in discussion and competition in alert communication.

There are several different techniques for alert communication that these smart calculations use in the application program on IDS2. This method completely provides soft and soft solutions for the IDS problem in a soft calculation, as well as the proposed method to reduce the rate of false alarms and increase the discovery rate.

Lee and colleagues presented a method based on a support vector machine and removal techniques that combines clustering methods, ants community algorithm and supporting vector machines, which is the result of this paper. , A high penetration detection rate for intrusion detection is presented. [2].

Pendah et al. [3] used a combination of classifiers instead of a classifier in the intrusion detection system. This article focuses on the recognition of computer and network security with regard to the growth of the Internet in everyday life. Often, a simple classification algorithm is used in networks that can detect normal and abnormal behavior from data traffic in the network. This algorithm does not work correctly in detecting attacks at a false alert rate. Therefore, in order to solve the problem, the intelligence classification method is used to detect network intrusion. In addition to increasing the efficiency, this algorithm can be used in the algorithm. Oversight or learning filters data by using clustering of seniors' data sets These results have been tested based on the NSL-KDD dataset.

Sinvasa and Udani [4] have proposed a method for extracting weights in the neural network for an intrusion detection system using genetic algorithm technique, in which neural networks and genetic algorithms are two techniques for learning and optimization, both of which The technique has strengths and these two techniques are introduced in two separate ways. In Sinvasa and Yudani's paper, they describe the genetic algorithm in which neural networks are used which, while effectively nerve networks They identify the effective process of neural networks using a successful genetic algorithm.

Hurang et al. [5] A hybrid method based on hierarchical clustering algorithms as well as a function that selects a number of important and simple features and ultimately combines them with a variety of vector-based techniques. Slow down. This method could also reduce the training time and efficiency. Also, in this study, KDD99 data 1 was used to verify the validity of the function. The best performance of this method has been to detect DOS and Prob on the network. Then, using the simple Bayesian algorithm for node classification, the result obtained for the attack type from the KDD99 data set is used for evaluation. However, this solution is not practical for real networks, since the K-Means algorithm requires more time to process large amounts of data in real networks, which can lead to bottlenecks and collisions in the system.

In another paper, we propose a new growing decision tree algorithm based on the theory of large grain aggregates. Given that learning in the real world and in different dimensions is interactive, increasing and dynamic. The data also appears at anytime, anywhere. So growing learning in data mining is important. Also, decision trees can include this issue. In this study, the theory of large grain collection was used to eliminate waste properties and eventually reduce features. In order to improve the diagnostic accuracy, IDS uses an increasingly intrusive detection method. Due to the fact that decision trees and large grain theory with discrete features can work better, in this paper, the discretization method 2 has been used. The discretization method is split into two nonmonitoring discrete and discrete-4-discretization methods. In this paper, two algorithms show that the time of model construction in the new algorithm is compared with the three algorithms ID3, DTRS and C4.5 is lower, but the accuracy of classification of these algorithms is not like the first algorithm.

Ozge Cepheli [6], Distributed Denial of Service attacks are detected, the most threats that occur today on the Internet. This method acts as a hybrid method for intrusion detection to use abnormal and signature-based behavioral methods. This method can be used by all devices whose ports are open on the Internet, and also for high traffic precision where network traffic is high on the target side.

A number of methods have been presented to prevent distributed service attacks, although there are some deficiencies in these methods, but because these attacks are the most serious attacks on network security. This paper uses a novel framework, called the Intrusion Detection Combination System, to accurately detect intrusion detection. In the education section, the Goose method uses an abnormal behavior and the snort method for the detection of influenza. This method improves the efficiency of Distributed Service Prohibition attacks and shortens detection latency by using both abnormal and signaturebased behavioral techniques.

In a paper presented by Buczak, he explores data mining and learning techniques for cybercrime detection. As stated in this article, cyber security is a collection of techniques and processes designed to protect computers, networks, applications and data interfaces with these attacks.

This article describes the cyber security system as a combination of the host security system, which describes this in three ways. Abuse Technique, Non-Abnormal Behavior Technique, and Combination of the Two Techniques. In this article, we have designed techniques for exploiting known attacks and abnormal behavior techniques for unknown attacks. Intrusion Detection Review In this paper, using a data mining and learning machine method, some of which have been able to reduce the latency of attack detection to zero, as well as the normal activity occurring on the network for each system classified. What has been proposed by Buczak in the field of computer science and the learning machine has been focused on other intrusion detection papers focused more on the release of specific indicators, although it is effective for cybercriminals, but these methods It does not work without displaying data, and it does not suit the complexity of the network and the time consumed, and has performed incremental testing algorithms to improve it. That's why we need to make updates every day to detect abnormal behavior [7].

Gong et al. [8] presented a methodology based on associative rules and genetic algorithm programming to detect network intrusion by combining abusive and abusive methods. This method has been a development of genetic programming that could distinguish between normal data, known attacks and unknown attacks.

Yang et al. [9] presented a hybrid penetration detection system using protocol and data analysis techniques in the malware detection model using Chaid, Ques and C&R decision tree algorithms on data Samples were tested.

Zhang and Chen [11], have introduced a model of model dependency dependency law algorithm. In this method, several dependencies of the attack algorithm were identified and the use of the large-grain aggregate theory of algorithm1 for intrusion detection systems as well as the overall performance of the intrusion detection system were investigated.

Moda et al. [12] proposed an intrusion detection solution in the network by combining supervised learning and non-monitoring learning methods. They used the K-Means algorithm for non-monitoring education and the Naive Bayes algorithm for using supervised learning. The first step of the algorithm is to use the K-Means algorithm to group data into a normal state or attack type.

There are two basic reasons for using Nilam Divodi and Apranga Trapathy in the technique of solidarity events in the network intrusion detection system in 2015. First, the detection of network attacks is usually based on information or data from distributed sensors in the intrusion detection system, which during this The attack is generalized in the events of the network and therefore difficult to assess in the attack situation for a wide network also, the correlation between the results of the SNORT intrusion detection system and the associated simple event associated with the advancement of the project through the DARPA data set has been taken into account, which implies the relevance of alerts based on the same name due to different IP addresses and eliminates alerts It repeats itself and thus reduces downloaded information from the network administrator. In this article, the relevance of events is a condition that can be combined with

the process of unifying things that increase the quality of information and, in turn, reduce the amount of events. The purpose of the sensors is to collect network activities. These activities include network traffic, abnormal behavior of the user.

This means that the sensors detect two network and host activities. Network activities in this type of architecture include:

• Classification of low-level layer protocols according to the ISO standard in the link layer, network layer, and transmission layer.

• Service and Application Level Protocols (such as: SMTP, HTTP, FTP)

• Email content or web pages.

Host activity includes client, server, and routers. In this activity, the user identifies the operation on the host to enter the network or systems that includes the hardware and operating system and its applications including email, web pages.

Also, the true positive rate for the data set is 92.1%. However, this method is only used for situations where network traffic and high flexibility are available, and this training data may not be available on a real and functional network, which may reduce the system upgrade efficiency currently in place. A grid that involves increasing these limits, and the only benefit of this method is to reduce the dependency on the abnormal detector. [13]

At the end of this section, the comparison with the methods of intrusion detection, each of which has its own advantages and disadvantages. Which is shown in Table 1.

Intrusion Detection	Advantages	Disadvantages
System Techniques		
Signature-based detection technique	1. Low computational cost	1. No new known attacks can be detected.
	2-High detection rate for previously known attacks	2. The false alert rate has increased.
	 Intrusion detection based on pattern matching or knowledge-based pre- configuration 	
anomaly-based detection technique	1. Reduce false alert rates for unknown attacks	 Requires most scheduling to detect an attack.
	2. Use the probability test by collecting the behavior of the nodes for intrusion detection	2. Accuracy detection based on the collection of behavior or features in the network.
Fuzzy logic technique	1-Use of quantitative features 2. Provides flexibility for uncertainties in the network	Detecting the accuracy of the artificial neural network is lower.
Associative rules technique	1-Use of signature of known attacks or appropriate attacks in detecting misuse	1. Can not detect all unknown attacks.
		2. Requires a number of database scans to create rules.
Svm technique	 Properly categorized infiltration if the data sample size limitation is given to the network The number of large and large features is used. 	Categorization is done in separate attributes, therefore, the need for preprocessing of the features is required.

Table1: Comparison of the methods used to detect intrusion

3. PROPOSED METHOD

Regarding Table 1, it can be concluded that commonly all of the current intrusion detection systems use the extracted features of network traffic to evaluate intrusive patterns in the network or to look for a fragment of features and patterns of the case Comment. The main thing here is, some of these features are unrelated in the data set used in the proposed method, which leads to redundancy. In this paper, we use the combination of neural network intrusion detection and genetic algorithms to benefit from the effectiveness of the networks. As a result, it combines (abnormalybased detection and abuse based detection) to discover infiltration. We increase detection integrated method for intrusion detection. The proposed method, by eliminating unnecessary and unused features, also makes it possible to simplify the problem of detecting faster penetration and higher precision. In Figure 1, a proposed method of work is presented as a combination of genetic algorithm and neural network.



Figure 1. A proposed method of work that is presented in a combination of genetic algorithm and neural network.

As shown in Fig. 1, we initially applied a genetic algorithm to reduce the dimensions of the features in the standard data set so that we can minimize the weight of the neurons versus the maximum weight in the network that we have done the preprocessing in two stages. First, using Matlab's main component analysis technique, we did the job that low performance accuracy in intrusion detection, which caused the error to increase, this would reduce the confidentiality of information, which, in order to solve this problem, minimize the weight of the operation Neurons were done using genetics only and without diminishing the dimensions by analyzing the main components. This method has increased the accuracy of the operation compared to the previous method and reduced the error.

We first used the NSL-KDDCup99 dataset, which uses these 41 features, the 9 main attributes that share the detection of attacks on the network.

So, characterizing the characteristics of the chromosome length and the number of genes in the chromosome and then the function of competence and the crossover method. Each chromosome consists of 4 genes that are:

• Mode: Specifies the type of infiltration (00 -Our source has a Register Id)

• Opcode: Specifies the authorized factors on the network.

• Target: Specifies the record of the destination or destination node.

• Source: Specifies the source or port entry for the network.

Then, as shown in Fig. 1, after the preprocessing operation using the genetic algorithm and minimizing the weight of the neurons, it is time to classify the data by the neural network. In previous articles, perceptron neural networks were used. In this paper, the feedback neural network has been used. Because it is supposed to be activated as a firewall so that if one percent of our network does not detect the influence it can return after some time if the malicious operation occurs in the network and re-check for intrusion detection, so the nerve grid of the type of multilayer networks is. This will increase the accuracy and reduce the error that we carry out all tests in the Matlab software.

There are two different sections in this implementation.

• Clustering for 4 network attacks.

• Neural network interacting with genetic algorithm.

In order to increase the accuracy and reduce the error and noise in the grid, we use the algorithms of genetic algorithm and feedback grid.

The reason for using the feedback neural network in this paper is that the feedback neural network is a dynamic neural network, and in modeling a system using this dynamic neural network, it defines an appropriate architecture and defines different optimal weights in the network. In This neural network architecture provides each input pattern with important information about structured data. And, using the function function that specifies the relationship between the input and output neurons in a network, This function is of a sigmoid function (logistic) and linear type. The hidden nodes used in this grid calculate the weights, which establishes a complex nonlinear mapping between input and output variables. Dynamic neural networks means minimizing the forecast error using the MSE and RMSE metrics that are obtained from the following relationships(1),(2).

$$MSE = (\overline{Y}_t - Y_t)^2 / n \qquad (1)$$

$$\mathsf{RMSE} = \sqrt{(\overline{Y}_t - Y_t)^2 - n} \tag{2}$$

 Y_t and Y_t express the actual trend and predicted trend in the feedback neural network. This neural network architecture defines the best network structure over time and receives feedback from the network. In Figure 2, we show the feedback of the neural network architecture.



Figure 2. Neural network feedback architecture

As shown in Fig. 2. Our network inputs are input signals and target values. If the network receives feedback from the network, the neural network determines whether the network has the ability to generalize the relationships learned in the model and has been able to identify the influence. Therefore, we use proper optimization algorithms for training that have a suitable velocity and converge to the rest of the algorithms. We use the genetic algorithm to select the attribute and weigh the attributes. In this section, the algorithm chooses the optimal properties in the training section among all the extracted features, which also increases the accuracy in addition to reducing the number of features and computational costs.

In all the algorithms that are based on the evolutionary algorithm to determine the characteristics, the ability of the algorithm to search all the parts of the search space in the network and its ability to exploit the best solutions. This algorithm applies feedback classification to classify offensive attacks in the education section. This algorithm finds a combination of weights and attributes that cause minimal error for the neural network. Then, through a dynamic neural network, our network can be carefully chosen to determine the optimal network architecture by receiving feedback from the network and the process of testing and error and time spent. That is why, in Figure 1, we use the genetic algorithm to search, optimize and teach the machine, the framework and rules governing the genetic algorithm so that we can execute things in parallel, and also the genetic algorithm first begins with the optimal values of the various parameters of the neural network Gains. In this genetic algorithm, we begin the initial population with random answers. This can lead to multiple searches on a population of variables at the same time. In this paper, the optimal network optimization structure and weights are determined using the merit function shown in equations 3 and 4, which determines the optimal performance of the generated response.

Fitness=
$$1/2$$
 (x+ p) (3)

$$P = \frac{100}{n} \sum_{i=1}^{n} w_i \tag{4}$$

In Equation 3, the suitability function in this proposed model is shown on the basis of the mean error x and the percentage of total weights (connections) in the network. In Equation 4, p is a parameter that shows the sum of different weighing values through different network architectures Let N denote the number of our generations in the genetic algorithm, w is the weight of the inputs of the network, and also the number 100 is the maximum number of generations in this genetic algorithm.

In this proposed model, considering the minimum value of the function of merit during the implementation of the network algorithm with the least complexity and error is obtained. In this paper, we introduced a new method for detecting intrusion in the network so that we can act as a firewall to prevent intrusion from entering the network. We did all this in Matlab. Because the negative feedback grid reduces the computational costs and increases the accuracy of the classification. The other reason is that the new method acts on the network, if it does not recognize one percent of the penetration and passes through the network, then we realize the penetration, so we can go back and do a search to prevent the penetration from entering the network. In the next section, we describe the results of the simulation of the proposed method, and then we make a comparison with the previous methods.

4. SIMULATION RESULTS

In this paper, we test and test in two stages. First, we perform the test with all nine main characteristics of the NSLKDDCUP99 dataset, which is related to the four types of attacks that can occur in the network, and then, in the second step, we performed the test using the reduction tool and the genetic algorithm. This is to say that the same 9 features that we did in the first step of the test were now done using the Dimension Testing Tool. Due to the results obtained, the error rate is increased and the accuracy rate in classifying and detecting attacks in the network is reduced and minimized. It can be said that in the initial model, the use of all 9 characteristics in terms of the rate of classification of attacks in the network and the error rate show a better situation for the proposed model.

The output of our simulation in this paper, using the genetic algorithm and the feedback grid, is shown in Table 1-1 below the minimum error in training time.

2	2	n
4	4	υ

Precision rate	Recall rate	Accuracy rate
97.5% %80		98.5%

Table 1-1 shows the minimum error during training, and after obtaining the learning parameters, we give the test data to the Matlab software by the neural network, and the output of the simulation with accuracy of 97.5% and accuracy of 98.5% and with the Recall rate of 80% in Equation 5 is calculated.

This call rate is due to our ability to detect the extent of the system attack on the network, and we calculate the function of accuracy and accuracy in equation 6 and equation 7.

$$Recall = 100 * (tp(i) / (tp(i) + fn(i)))$$
(6)

A C C U R E N C Y (i) = 1 0 0 * (T P (i) + F N (i)) / (TP(i)+FP(i)+TN(i)+FN(i)) (7)

And therefore the output obtained from the proposed algorithm is shown in Figure 1.



Figure 3- Minimum error in training time

Figure 3 shows that with this population, our accuracy and accuracy in the detection of network intrusion has increased, and this is done using the genetic algorithm to normalize the characteristics and use the main component analysis tool to reduce the dimension for network training and its testing. We make feedback with neural networks. If we use the main component analysis to reduce the dimensions, that is, to choose the best of these 9 features, we will not give a better result but also increase the error rate, but if with the same 9 main attributes in the intrusion detection training will reduce the error rate and increase the accuracy rate, so we performed the feedback from the neural network. In Table 2 compares this proposed method with previous methods that with the k-mean algorithm, SVM algorithm and snort and bro methods.

Table2 - Comparison of the proposed method with the method of intrusion detection

Algorithms and methods	K-mean	SVM	snort	bro	proposed algorithm
Detection	82.3%	57.6%	88.88%	77.78%	92%
rate					
Accuracy	77.25%	73.24%	80%	70%	94%
rate					
Recall	52%	18.14%	20%	30%	80%
rate					

We also show this comparison on Figure 4.



Figure4 - Comparison of the proposed method with the method of intrusion detection

As shown in Table 2 and Figure 4, the methods used to detect intrusions are not appropriate and for this reason, we have used the feedback neural network method so that we can infiltrate the network with a higher accuracy and less error.

5. CONCLUSION

Given the simulation of the proposed method in Matlab software, we can conclude that if we reduce the dimensions for training using the genetic algorithm, the number of errors increases and, therefore, the accuracy of the intrusion detection is low, so we have the features The corresponding 80% is used to train and then the remaining 20% to test the feedback from the neural network, which reduces the error, making the penetration better detected and increasing accuracy. This is because today's attacks on computer systems reduce security processes and access to information, which can damage network security and data, or system and user files to destroy. For this reason, we used neural network feedback techniques and genetic algorithms for intrusion detection and used it to simulate Matlab software. We have been able to provide an effective way to prevent intrusion of the unauthorized person. We will use these techniques because of the higher accuracy that we have with other methods. The neural network presented in this paper is a novel approach to intrusion detection using genetic algorithm, and also feedback neural networks is one of the inseparable properties of dynamic neural network in the ability to learn in the network, and in general, The process of learning in the neural network means setting and categorizing attacks on the network, and to reduce the energy consumption in the network, we use the combination of genetic algorithm and neural network, Usually the network can not be used with real data because data is not related to a particular application and environment. They should be consistent across the entire environment. Therefore, we used the NSL-KDDCup99 dataset and, by selecting key parameters, we were able to penetrate the network We categorize the detection and type of attacks with high precision, and this algorithm can be used to optimize the accuracy and speed of the two-dimensional algorithm in parallel.

6. REFERENCES

[1] H.T. Elshoush, I.M. Osman ,2011 . "Alert correlation in collaborative intelligent intrusion detection systems : A survey", applied soft computing, Elsevier, pp. 221-238

[2] Li, Yinhui, et al , 2012. "An efficient intrusion detection system based on support vector machines and gradually feature removal method." Expert Systems with Applications , pp. 424-430.

[3] Panda, Mrutyunjaya, Ajith Abraham, and Manas Ranjan Patra , 2012. "A Hybrid Intelligent Approach for Network Intrusion Detection." Procedia Engineering,pp.1-91

[4] P.Srinivasu, and S. Avadhani , 2012. "Genetic Algorithm based Weight Extraction Algorithm for Artificial Neural Network Classifier in Intrusion Detection."Procedia¬ Engineering¬,¬pp.144-153]

[5] Horng, Shi-Jinn, et al , 2011. "A novel intrusion detection system based on hierarchical clustering and support vector machines." Expert systems with Applications , pp. 306-313.

[6] Ozge Cepheli,Saliha Buyukcorak and Gunes Karabulut Kurt, 2016 ." Hybrid Intrusion Detection System for DDOS Attacks. " Journal of Electrical and Computer Engineering,Article ID1075648,8 pages.

[7] M. Li, S. Pan, Y. Zhang, and X. Cai ,2016. "Classifying networked text data with positive and unlabeled examples," Pattern Recognition Letters, vol. 77, pp. 1–7.

[8] Y. Gong, S. Mabo, C. Chen , 2009 . "Intrusion detection system combining misuse detection and anomaly detection using genetic network programming", ICROS-SICE international joint conference, pp. 3463-3467.

[9] J. Yang, X. Chen, X._Xiang, J. Wan, 2010 ." HIDS-DT: An effective hybrid intrusion detection system based on decision tree", IEEE international conference on communications and mobile computing, pp. 70-75.

[10] H.T. Elshoush, I.M. Osman, 2011 ." Reducing false positives through fuzzy alert correlation in collaborative intelligent intrusion detection systems : A review", IEEE international conference on fuzzy systems , pp. 1-8

[11] J.Zhang, and X.Chen, 2012."Research on Intrusion Detection of Database based on Rough Set"¬, International Conference on Solid State Devices and Materials Science, Physics Procedia, pp.1637-1641.

[12] Z.Muda, et al, 2011."Intrusion detection based on K-Means clustering and Naïve Bayes classification." Information Technology in Asia (CITA 11), 7th International Conference on. IEEE.

[13] Neelam Dwivedi and Aprna Tripathi , 2015." Event Correlation for Intrusion Detection Systems. "IEEE International Conference on Computation & Communication Technology.