Journal of Academic and Applied Studies
Vol. 2(12) December 2012, pp. 9- 18
Available online @ www.academians.org
Special Issue on Computer Science

ISSN1925-931X

Combination of Three Machine Learning Algorithms for Intrusion Detection Systems in Computer Networks

Ali Reza Zebarjad[1] , Mohmmad Mehdi Lotfinejad[2]

[1]*Dapartment of Computer, Andimeshk Branch, Islamic Azad University, Andimeshk,Iran*
[2]*Dapartment of Computer, Mahshahr Branch, Islamic Azad University, Mahshahr,Iran*

Abstract

Expansion of computer network causes the network to be the target of attacks. In this study, using the idea of mixture experts efficient and effective intrusion detection system is presented. In this method of KDD 99 data set for training and test data and using experts algorithms to reduce the dimension attributes that are less important in identifying attacks and eliminating the overhead of data input has reduced. In order to make different errors in the output of the three types of classification, support vector machines, decision trees and neural networks, multi-layer perceptron is used. Makes different errors, experts can identify weaknesses in the other attacks to different causes, including lack of training goes back, can identify. Also in this method of learning the weaknesses of each experts, to provide experts management is a possible mediator. experts mediator network to allow such actions in the output shows, that the new sample should decide otherwise with relevant experts to filter the output of the impact on the final result. The results of this study indicate optimal performance of this system is acceptable.

Keywords: intrusion detection systems, support vector machines, decision trees, neural networks, multi-layer perception, mixture expert

1- Introduction

With the expansion of Internet use and exposure to various software tools available to hackers infiltrated the computer network attack is widely increased. The first line of defense and defense firewall as is used for security, but a firewall in a limited capacity to detect attacks. The use of intrusion detection systems required to deal with suspicious activity inside and output the network. This system analyze data packets on the network and eliminating non-essential information on regular or decision packages are being attacked. Intrusion detection systems according to various methods for data mining large packets of information they use, but due to a variety of attacks using a technique alone does not result in an acceptable figure. Recent studies show that using a combination of methods usually used to achieve optimal results. In this paper an intrusion detection system based on the idea of mixed experts is presented. Instead of using a classification in the mixture experts, the combination of several identical or different classification is used for final decision. Mixture multiple classification will be effective if the classification error are different have to cover each other's errors.

The various categories, using the different sets of learning is achieved. For this purpose, methods of randomization feature space between the base classifiers [1], adding noise to the input data [2], the nonlinear transformation on the feature vectors [3], Bagging Method making [4] and the algorithm to Bagging half & half [5] have been proposed. Operational methods to prepare samples for different training based classifiers do. Operational methods to prepare samples for different training based classifiers do. With different learning sets are expected to different based classifiers for complex systems is provided.

In reference [6] a combination of hierarchical decision tree methods, Support Vector Machine and Hybrid Decision Tree-SVM is used. In this way related samples within each model output is passed through to decide about the final output are used. Experimental results show that better methods for detecting attacks, U2R, Probe the decision tree, Support Vector Machine and the Hybrid Decision Tree-SVM. The combination of the ideas used in the majority and the final result is determined. In reference [7] for intrusion detection in a three-layer hierarchical structure of the combined methods of experts and a majority voting, Average Rule out, Dempster-Shafer Combination and simple Bayes is used to combine results. The survey results show that simple Bayes methods for combining experts is far better than other options.

In reference [8] is compared with four multi-stage hybrid structure. Under each of these structures based on a multi-layer neural networks classification algorithms, support vector machines, Navi Bayes and decision tree is built. So that for the first stage of passing classified Dataset attacks separated from the normal packages, and then later as a group are identified by passing the remaining Dataset attacks. At each stage, the attacks were classified as poor. Are classified separately passed through it again until it goes to all the attacks to be detected. Also, given that the first stage of a multi-layer neural networks to isolate 100% of the normal modes, So in other multi-stage manufacturing systems based on Bayes Navi Support Vector Machine and are designed to separate the normal modes in the first stage of a multi-layer neural networks are used. The disadvantages of these methods are high cost, complexity and training time may be excessive.

2- KDD Cup 99

KDD Cup 99 data is used in this research. **A** copy of the data from the MIT Lincoln Laboratory as Dataset by DARPA 1998 and with the goal of research in network intrusion detection is provided. It includes many packages Dataset normal and simulated environmental attack on the United States Air Force from a local network in America has been obtained for 9 weeks [9]. This is Dataset contains about five million records. And each record includes 41 feature.

The records of 10% due to Dataset usually it is used for training and testing. Dataset training that included 494 014 and The test is comprised of 311 021 records. It is worth mentioning that these two sets have an additional feature. Specifies the packet type. In Tables (1) and (2) the type and name for each sets of attacks has been inserted.

Table 1: Type and attacks on training Dataset

Attacks Classification of	**Attack Name**
Probing	Port-sweep, IP-sweep, Nmap, Satan
Denial of Service(Dos)	Neptune, Smurf, Pod, Teardrop, Land, Back, Apache2
User to Root (U2R)	Buffer-overflow, Load-module, Perl, Rootkit, spy
Remote to Local(R2L)	Guess-password, Ftp-write, Imap, Phf, Multihop, Warezmaster, Warezclient

Table 2: Type and attacks in test Dataset

Attacks Classification of	**Attack Name**
Probing	Port-sweep, IP-sweep, Nmap, Satan Saint, Mscan
Denial of Service(Dos)	Neptune, Smurf, Pod, Teardrop, Land, Back, Apache2, Udpstorm, Process-table, Mail-bomb
User to Root (U2R)	Buffer-overflow, Load-module, Perl, Rootkit, spy, Xterm, Ps, Http-tunnel, Sql-attack, Worm, Snmp-guess
Remote to Local(R2L)	Guess-password, Ftp-write, Imap, Phf, Multihop, Warezmaster, Warezclient,Snmpget-attack, Named, Xlock, Xsnoop, Send-mail

3-The proposed method

The system of multi-layer perception neural network classifiers, support vector machines and decision tree is used as follows (Figure 1). Ten blocks in the network of experts that included three expert groups SVM, DT and MLP is composed of three experts in the field of mediation, each MLP has 10 neurons in the output layer is used. This data structure is divided into four stages and aspects of the training set, each block of training expert, expert training of mediators and the results are combined. The following commentary will explain each of them.

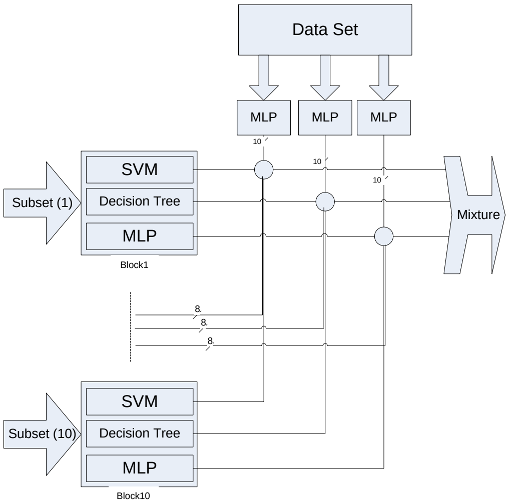

Figure 1: The proposed method

3-1- Dimension reduction and data sets into a training set

Some features of KDD data sets do not have much impact in attack identified, Therefore, in order to minimize data overhead, Relief feature selection algorithm using a number of features that are less than 10 percent effective in detecting attacks will be deleted. After eleven feature of this algorithm is due to low participation rates, identify patterns in the data sets were excluded. The remaining attributes and also identify the extent of their participation in the attack after the Relief algorithm in Table 3 is shown.

Combining experts in this case is that no classification alone is not able to identify any patterns, So we are trying to make is that we are experts at identifying patterns that produce the minimum error, And are also different from each other in order to compensate for their shortcomings. In other words, if the classification is quite correct to detect R2L attacks U2R attacks, but fails to properly identify, The Next classification does not need to be able to properly detect R2L attacks, but the convention should be trained so that they will have the ability to detect U2R attacks, The first classification is covered so weakness. In order to achieve the above goal, This classification of each block to identify only two types of train attack. Certainly for a class classification detect two types of attacks is far easier to identify five types of classes will vary. Also, given that each of the three-class blocks classification different base is used. Despite the similar Dataset different results are achieved. This classification method is based on trying to block certain attacks are expert and professional. And finally the MLP network mediator according to their expertise is recognized each classification will determine just how much the input pattern.

Table 3: The remaining features of the algorithm Azamal Relief

Ranked	Attribute
0.7004	3
0.6932	4
0.678	32
0.6713	26
0.6517	25
0.6492	19
0.5967	12
0.5358	6
0.4484	1
0.4331	27
0.4211	16
0.3753	5
0.3644	28
0.3482	22
0.3463	31
0.345	17
0.3169	14
0.2568	41
0.2479	23
0.2462	40
0.2333	2
0.2087	36
0.1796	24
0.1713	37
0.1478	35
0.1212	30
0.1173	38
0.1128	11
0.1088	33
0.1011	34

. Experts and specialized training as given in Dataset five different classes, It is necessary to separate the two classes together into a block can be imposed. Until after the training, expert corresponding block is able to identify these two classes. Therefore Dataset that its dimensions have been reduced in level based on five existing classes and the dual training is divided into 10 subsets, So that each set of records only two classes remain (Table 4). In principle, any type of attack is different.

Table 4: Subset based on type classes

No.	Class	Class name
1	1,2	Normal, Probe
2	1,3	Normal, Dos
3	1,4	Normal,U2R
4	1,5	Normal, R2L
5	2,3	Probe, Dos
6	2,4	Probe,U2R
7	2,5	Probe, R2L
8	3,4	Dos, U2R
9	3,5	Dos, R2L
10	4,5	U2R, R2L

3-2- Expert training for each block

Create and implement a mediator network and experts each block has been read by the software Matlab. Expert testing to try and error for SVM with polynomial kernel and previous studies have shown that in most cases is associated with better results [10]. Therefore, this study used the kernel polynomial convention.

The decision tree consists of nodes, and edges are leaves. Each node is an attribute that defines the separation given by it. Each node of the edge is labeled with values by the next node or leaf is connected to the edges, And each of these labels in terms of attributes, the data leads to the desired class. For discontinuous features, and each branch is formed and for each attribute value associated with a specified threshold and is divided into two branches. This neural network is trained, the network output computed using the input vector and the corresponding target vector (corresponding to each set of classes) is compared to its And the difference between the calculated output and desired output is called the error to be released back across the network. Gradient Descent Algorithm is weighted according to the desire of change and is adjusted to minimize error. Running training samples are used, Errors are computed and the weights are adjusted every vector until the total error for the training samples to a small amount would be acceptable. It is worth mentioning that the error rate is not determined by us, In this case the number of iterations or Epoch (providing any information to the network weights are modified. When all data collection was given to the network, say an Epoch has been completed). For the desired end would be.

After training, the expert all Dataset been fully applied to reduce the dimensions and results are obtained. Certainly any expert who attacks them is specialized to detect. It is worth mentioning that only the expert opinions about the attacks for which they trained is considered valid and the rest zero. For example, if an expert is trained to attack two and three and a sample output for a type or class of four to identify these comments are not acceptable in order to be zero. Thus, for every expert there is a new vector. Input or output for each sample is in the order of zero or one the sign of correct diagnosis is the expert. Given that there are different three expert in each block. Three is the array that each array has 10 columns and row number is 494 014. The answer in each row of the validity of the SVM, DT, or a neural network is an input. (One means expert answer valid and zero means that the expert could not identify this sample). For example if the table 5 is the first step, Table 6 as the matrix is considered an expert mediator.

Table 5: Sample of SVM diagnosis with respect to the number of classes that can identify

Sample Class	2-1	3-1	4-1	5-1	3-2	4-2	5-2	4-3	5-3	5-4
1	0	1	3	5	0	3	0	3	0	5
2	5	0	4	0	2	4	5	0	2	0
3	1	1	0	0	4	2	0	3	0	0

Table 6: training matrix for mediator based on the table 5

Sample Class	2-1	3-1	4-1	5-1	3-2	4-2	5-2	4-3	5-3	5-4
1	0	1	0	1	0	0	0	1	0	1
2	0	0	1	0	1	1	1	0	0	0
3	1	1	0	0	0	1	0	1	0	0

3-3- Expert training for mediator

Three matrix obtained in the previous step, in principle, the three networks and the mediator must train to achieve these goals. So for three neural network with four of the middle layer is the middle layer has 25 neurons and each neuron has 30 inputs and 10 columns of the array is given in the output neurons are 10. As the neural network training phase before the networks are trained with the specifications mentioned.

Network interface learn how to weight each expert opinion with respect to the input pattern is assigned to it. Learning in this structure, the type of competitive learning, means fewer errors when learning on the learning sample may be more encouraging [11].

3-4- Combined results

At this stage, each new data sample, both the input and the network mediator is applied to each block of three expert. results obtained from experts and the network mediator desired filter based on previous training for the new sample, which acts on the classification results, expert point out that actions have to be determined. If the answer valid, but some expert also diagnosed, the majority is used. In the case of a tie vote, the class is chosen that is most representative. Each expert is because they actually represent two classes. In other words, if one is expert to produce a corresponding mediator, means that the effective output is in recognition classification its. But if a zero and the output filter does not apply expert opinion. results all classification that the network mediator are allowed to apply to the output, Compared with the same class as the final result is selected and the highest vote is taken.

4- Results

The composition of the reference methods, SVM, decision tree, Hybrid decision tree-SVM and combining these three methods have been used [6]. Attacks Probe will identify good practice. But in a good report card does not detect U2R attacks. Also in reference to a three-layer hierarchical structure and composition results were used [7]. With improved results in detecting U2R attacks than the previous reference, but the ability to detect R2L attacks are not necessary. The results of the proposed method with the results of reference [6] in Table 7 is inserted. And in Figure 2 are compared with the reference results. The hybrid method presented in reference [6] Probe into attacks only be fully identified. But performance is not good for detecting U2R attacks.

Method presented in this study to identify high ability classes are normal and Dos. Almost all existing methods lack the proper training of U2R and R2L attacks are detected, however, that the proposed method accurately detects these attacks has increased. In this method, because each expert trained to attack only two samples, With lack of them can decide to sample properly. Also, given that the two classes to a block with three different expert have different error are applied, Even if an expert fails to correctly identify other classification to compensate for this defect. The mediator for the expert networks that have no correct answer Considered a zero weight and are prohibited from applying in the exits. It is worth mentioning that every expert in the areas where they can be trained to express opinions, Otherwise, the answer does not apply in the output, so the lack of a specific class does not have much impact on the result.

Because this method eliminates the features that identify the classes do not have much impact, Overhead of additional data entry is eliminated. The proposed method has good performance was due to the use classification three different classification, as well as the partition main Dataset 10 different collections, output error causes a variety of experts that this error range specified by the network mediator and the management. Using this range, the experts have been weaknesses in the face covered with new samples, and to achieve acceptable performance.

Figure 2: Diagram of comparing the convention proposed method with other methods

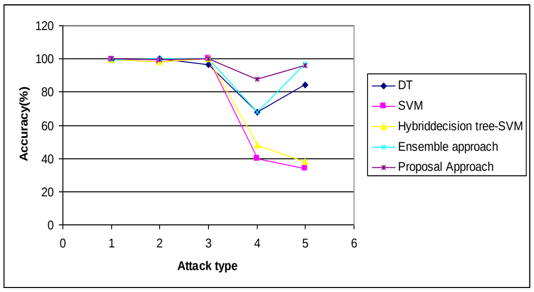

Table 7: Results of the proposed method compared with the reference [6]

Attack type	Accuracy(%)				
	DT	SVM	Hybrid decision tree-SVM	Ensemble approach	Proposal Approach
Normal	99.64	99.64	99.7	99.7	99.95
Probe	99.86	98.57	98.57	100	99.8
DOS	96.83	99.92	99.92	99.92	99.96
U2R	68	40	48	68	87.56
R2L	84.19	33.92	37.8	97.16	96.12

References

[1] Skurichina, M., Duin, R. P. W., "Bagging, Boosting and the Random Subspace Method for Linear Classifiers", Pattern Anal. and Applications, 2002, vol. 2, pp. 121-135.

[2] Raviv, Y., Intrator, N., "Bootstrapping With Noise: An Effective Regularization Technique", Connection Science, 1996, vol. 8, pp. 355-372.

[3] Sharkey, A., Sharkey, N., Chandroth, G., "Diverse Neural Net Solutions to a Fault Diagnosis Problem", Neural Computing and Applications, 1996, vol. 4, pp. 218-227.

[4] Breiman, L., "Bagging Predictors", Machine Learning, 1996, vol. 24, no. 2, pp. 123-140.

[5] Breiman, L., "Half & Half Bagging and Hard Boundary Points", Technical Report, Statistics Department, University of California, Berkeley, 1998.

[6] Peddabachigari, S., Abraham, A., Grosan, C., Thomas, J., "Modeling Intrusion Detection System Using Hybrid Intelligent Systems", Journal of Network and Computer Applications, 2007, vol. 30, pp. 114-132.

[7] Chou, T. S., Fan, J., Fan, S., Makki, K., " Ensemble of Machine Learning Algorithms for Intrusion Detection", IEEE International Conference on Systems, Man and Cybernetics, San Antonio, TX, USA, 2009, pp. 3976 – 3980.

[8] Mohamed, M. A., "Development of Hybrid-Multi-Stages Intrusion Detection Systems", IJCSNS International Journal of Computer Science and Network Security, 2010, vol. 10, no. 3.

[9] KDD’99 Archive: The Fifth International Conference on Knowledge Discovery and Data Mining.
URL: *http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html*

[10] ABMS, A., Abraham, A., "An Empirical Comparison of Kernel Selection for Support Vector Machines", In: Proceedings of The Second International Conference on Hybrid Intelligent Systems: Design, Management and Applications, The Netherlands: IOS Press, 2002, pp. 321-330.

[11] Jordan, M., Jacobs, R., "Modular and Hierarchical Learning Systems", The Handbook of Brain Theory and Neural Networks, MIT Press, Cambridge, MA, 1995.