A HYBRID INTRUSION DETECTION SYSTEM USING PARTICLE SWARM OPTIMIZATION FOR FEATURE SELECTION

[1]SEDIGHEH KHAJOUEI NEJAD, [2]SAM JABBEHDARI, [3]MOHAMMAD HOSSEIN MOATTAR

[1]Computer Engineering Department, Sirjan Branch, Islamic Azad University, Sirjan, Iran
[2]Computer Engineering Department, North Tehran Branch, Islamic Azad University, Tehran, Iran
[3]Computer Engineering Department, Mashhad Branch, Islamic Azad University, Mashhad, Iran
E-mail: [1]se_khajouei_nejad@yahoo.com, [2]s_jabbehdari@iau-tnb.ac.ir, [3]moattar@mshdiau.ac.ir

Abstract- The ultimate goal of this paper is to develop systems for intrusion detection in computer networks to achieve the best accuracy performance. This study suggests the idea of classifier combination. This hybrid approach is based on optimization and feature selection using a combination of two-step approach for the classification. In the first step, using Accelerated Particle Swarm Optimization (APSO) algorithm, a set of best discriminating features is selected. Then using a combination of three classifiers namely KNN, Decision Tree and Neural Network, intermediate data is generated. Finally these data is fed to the AdaboostM2 classifier for final decision. The performance of the proposed approach is evaluated with criteria such as F measure, Accuracy and False Alarm. Experiments on KDD-CUP99 dataset show the effectiveness of the proposed approach compared to the most recent approaches in this context.

Index Terms- Intrusion Detection, Hybrid Classifier, Particle Swarm Optimization, Adaboost, Feature selection

I. INTRODUCTION

Systems called Intrusion Detection System (IDS) is required to make proper decisions if the attacker passes fire wall, antivirus and other equipments and enter the security system. In the development of intrusion detection systems, the ultimate goal is to achieve the best possible accuracy. However, the most important issue is the problem of false alarm.

Currently a number of intrusion detection systems are proposed in the field of machine learning, including some works done using support vector machine can work [1]. In [1] neural network and support vector machine method are used separately for network intrusion detection. In [2] SVM is used as a classification method. Sarasamma [3] used similar approach while evaluated different subsets of features for the detection of various attacks.

Assembly-based approaches have also been used to increase the efficiency of the individual hierarchies. The hybrid approaches compose of several categories with simple learning algorithms. The methods of combining the results of the classification are also important issues. Using hybridization in general leads to improved accuracy. Large amount of recent studies have focused on the use of hybrid intrusion detection. In [4] the combination of KMeans and naive Bayes is used to improve intrusion detection system. In [5] kernel PCA in combination with particle swarm optimization (PSO) and radial basis functions (RBF) is used to improve the classification. In [6] out of 41 features in the KDD-CUP 99 dataset, six are selected using entropy. In [7] a hybrid approach based on anomaly detection is proposed for intrusion detection in wireless sensor networks. In [8] the combination of decision tree algorithm, K-Means and C4.5 improved the classification results. Reference [9] introduced a hybrid framework. In [10], Multilayer Perceptron and Radial Basis Functions and the combination of these two methods is used.

This paper presents a multi-stage hybrid methodology for intrusion detection in computer networks. After this introduction, the proposed method is discussed in Section 2. In Section 3 we discuss the evaluation of the proposed method. Finally, conclusions will be presented in Section 4.

II. PROPOSED APPROACH

The proposed method includes a feature selection step using Accelerated Particle Swarm Intelligence (APSO). Then, the basic classifiers are created. Finally, using the results obtained, the intermediate data are formed and sent to final layer for classification.

A. APSO Feature selection

A novel and modified Particle Swarm Optimization (PSO) by the name of Accelerated Particle Swarm Optimization (APSO) is utilized for feature selection, which benefits from high rate of convergence to the global optimum. In APSO model the velocity vector is updated by equation (1).

$$v_i^{t+1} = v_i^t + \alpha r(t) + \beta(g^* - x_i^t) \qquad (1)$$

In which r(t) is the standard Gaussian distribution N(0.1), v_i^t is velocity of ith particle in tth iteration, g^* is the best position within the swarm and x_i^t is the position of the ith particle in tth iteration. To increase the convergence rate of PSO and enhance it, the velocity parameter is removed and the equation to update the position of each particle in APSO is defined as Eq. (2).

$$x_i^{t+1} = (1-\beta)x_i^t + \beta g^* + \alpha r \qquad (2)$$

In Eq. (2), αr is the parameter to escape from local optimum. r and α values are determined based on standard Gaussian distribution.

The goal of the proposed approach is to find the best features for IDS classification. Therefore the variables of a solution have binary values which are either 1 if the feature is selected for classification and 0 is the feature is not appropriate. In the proposed encoding each solution is demonstrated by a vector of length N where N is the total number of primary features.

B. Initial classification and intermediate data formation

After normalization and selecting the appropriate features for each class, the training data set with selected features of each class are fed into three basic classifiers. Each of these classifiers are binary and determine whether or not an input sample belongs to a particular class.

The analyzed data-sets includes 5 different classes: (1) denial of service (DOS), (2) port scan, (3) remote to user, (4) user to root, and (5) normal behavior attacks. Three different classifiers are used in order to create intermediate results. The classifiers were decision tree, k-nearest neighbors, and neural network. Therefore, the output of this component that is used for creating intermediate points consists of 15 values.

C. Final classification component

In this stage, having collected intermediate results from the previous stage and having made a new data-set with 15 features, a series of ensemble classifiers are exploited for learning. Ensemble classifiers are based on decision trees with adaptive boosting (AdaBoost). This algorithm exploits the whole data-set for training each classifier.

III. EVALUATIONS

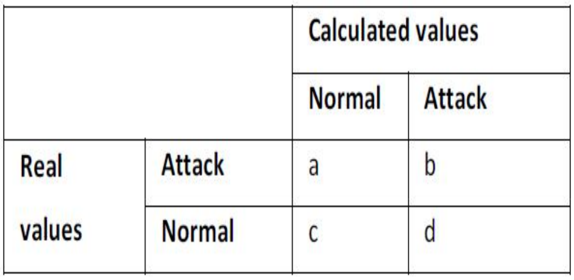
A. KDD-CUP99 Dataset

KDD99 data collection is a standard for intrusion detection evaluation. The data set divides into two main classes of normal and attack. Attack class consists of the four classes of DOS, R2L, U2R and Probe. KDD99 dataset includes 41 features.

B. Evaluation criteria

If system states are divided into two general states of normal and attack in IDS, there will be four conditions. They are shown in Table 1.

Table 1. Different conditions in an IDS



		Calculated values	
		Normal	Attack
Real values	Attack	a	b
	Normal	c	d

The calculation of each factor used in the study is based on Eq. (3) to Eq. (6).

$$Accuracy = \frac{a+b}{a+b+c+d} \quad (3)$$

$$\text{Recall} = \frac{d}{c+d} \quad (4)$$

$$Precision = \frac{d}{b+d} \quad (5)$$

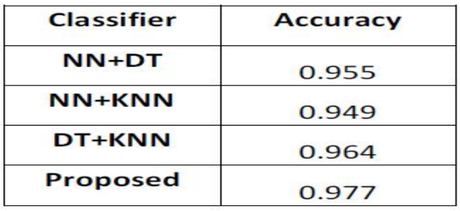
$$F-Value = \frac{2\times \text{Recall}\times Percision}{\text{Recall}+ Percision} \quad (6)$$

C. Experimental results

The final results obtained by the proposed hybrid approach is evaluated and compared with other approaches. Also, the effectiveness each step of the hybrid approach is evaluated separately.

Hybrid approach evaluation: In further experiments the effectiveness of the hybrid classification method to produce intermediate results is discussed. In these experiments which are shown in Table 2 and Figure 1, at first the stand-alone classifier are combined in a two-classifier mode and then evaluated with the condition as when all three approaches are assembled in the unified framework.

Table 2. Accuracy of the proposed hybrid approach compared with the case when two classifiers are combined



Classifier	Accuracy
NN+DT	0.955
NN+KNN	0.949
DT+KNN	0.964
Proposed	0.977

Figure 1 shows that the proposed hybrid 3-classifier approach outperforms all the combination of 2-classifiers. On the other hand, the combination of the decision tree and K nearest neighbor classifiers is better than other combinations that explain the greater generalization of these approaches compared with neural network.

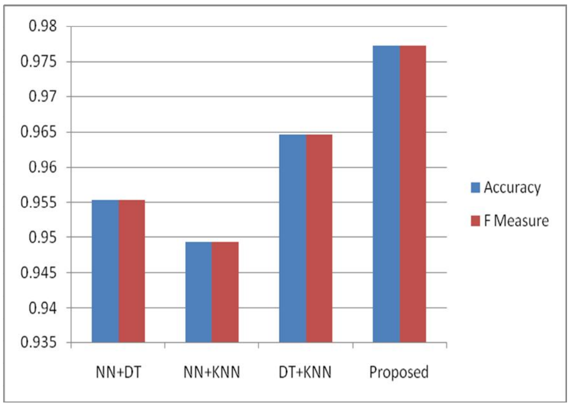


Figure 1. Accuracy and F measure for the proposed hybrid approach compared with the case when two classifiers are combined

The Same reults using Error rate and FAR are demonstrated in Table 3 which denotes the significance of the proposed approach.

Table 3. Error rate and FAR of the proposed hybrid approach compared with the case when two classifiers are combined

Classifier	Error rate	FAR
NN+DT	0.0446	0.0035
NN+KNN	0.0507	0.0052
DT+KNN	0.0354	0.0046
Proposed	0.0227	0.0020

Comparison with recent works: This section compares the proposed IDS with two other similar recent approaches. Multi-class pattern classification method using single classifiers and feature extraction using genetic multi-objective multi-dimensional feature space (MMOG) is proposed in [13]. This method tries to map the data to a new multi-dimensional decision space for better discrimination. To achieve this objective genetic method is used for feature extraction. Finally, this approach uses a simple multi-class classification procedure [13].

Another method is a hybrid approach to intrusion detection systems based on the total distance (DSSVM) [14]. In this paper a combination of a three-stage method is presented. In the training set, the data set is divided into K clusters. In the second stage, for each cluster, the distance between the centers of the clusters is calculated for each sample. The third step is to learn the training set to the SVM [14]. Figure 2 shows a comparison between the proposed methods with these two approaches. This comparison is shown in Figure 2 for a two-class attack-normal case. The results show that the proposed method is more efficient compared to the other approaches from both F measure and accuracy point of view.

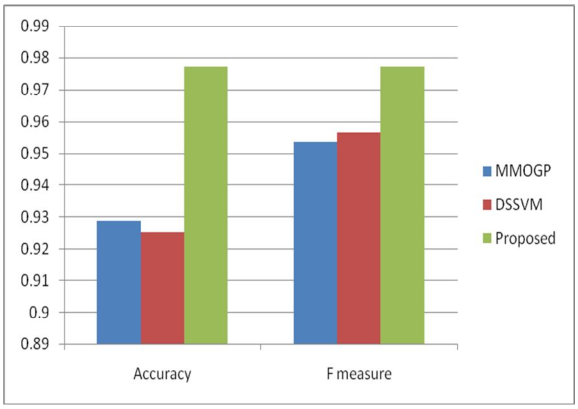


Figure 2. Accuracy and F measure of the proposed approach and approaches in [13] and [14] for 2 class case

These reults are also demonstrated in Table 4 for the 2 class case.

Table 4. Accuracy and F measure of the proposed approach and approaches in [13] and [14] for 2 class case

Approach	Accuracy	F measure
MMOGP [13]	0.938	0.953
DSSVM [14]	0.925	0.956
Proposed	0.977	0.977

Then the proposed method is compared with the two methods mentioned earlier for 5 class point of view. Figure 3 shows the F measure of the proposed approach for each of the 5 classes in comparison with the results obtained from [13] and [14]. The experiments show that the proposed approach has a superior performance on all the class except for probe class. Even in the case in U2R class which suffers from data scarcity, the results of the proposed approach are satisfactory.

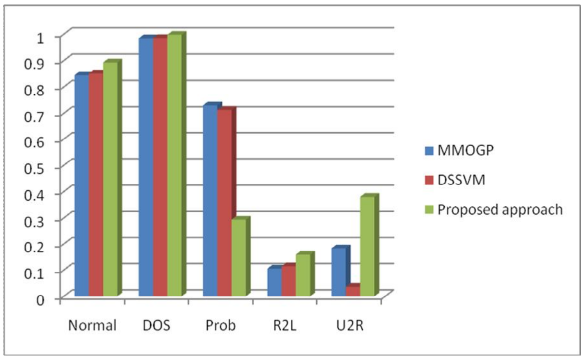


Figure 3. Accuracy and F measure of the proposed approach and approaches in [13] and [14] for 5 class case

CONCLUSIONS

This paper proposed a hybrid system with three levels. The first level extracted features Accelerated PSO. In the second level, the generated data-sets are fed into decision tree classifiers, k-nearest neighbor and neural network. Then, median results were obtained. Finally, in the third level, a classifier carried out the classification based on AdaBoost algorithm. The proposed hybrid approach was evaluated with respect to accuracy, recall, precision and F value. Compared to other approaches, the proposed approach shows a higher performance.

REFERENCES

[1] W.H. Chen, Sh-H. Hsu, H-P. Shen, ”Application of SVM and ANN for intrusion detection,” Journal Computers and Operation Research, Vol. 32 Issue 10, 2005, pp. 2617–2634.

[2] Z. Zonghua, S. Hong, “Application of online-training SVMs for real-time intrusion detection with different considerations,” Computer Communications journal (28), 2005, pp. 1428–1442.

[3] S. T. Sarasamma, Q. A. Zhu, J. Huff, “Hierarchical kohonenen net for anomaly detection in network security,” IEEE Transactions on Systems, Man and Cybernetics - Part B, 35 (2), 2005, pp. 302–312.

[4] Z. Muda, W. Yassin, M. N. Sulaiman, N. I. Udzir, “A Kmeans and Naïve Bayes learning approch for better intrusion detection”, Information Technology Journal 10(3), 2011, pp. 648-655.

[5] R. Xu, R. An, X.F.Geng, “Research intrusion detection based PSO-RBF classifier”, 2nd International Conference on Software Engineering and Service Science (ICSESS), 2011, pp. 104-107.

[6] B. Agarwal, N. Mittal, “Hybrid approach for detection of anomaly network traffic using data mining techniques”. Procedia Technology 6, 2012, pp. 996–1003.

[7] D. I. Curiac, C. Volosencu, “Ensemble based sensing anomaly detection in wireless sensor networks,” Expert Systems with Applications 39(10), 2012, pp. 9087–9096.

[8] A.P. Muniyandi, R. Rajeswari, R. Rajaram, R. “Network anomaly detection by cascading k-means clustering and c4. 5 decision tree algorithm,” Procedia Engineering 30, 2011, pp. 174–182.

[9] M. Panda, A. Abraham, M.R. Patra, “A hybrid intelligent approach for network intrusion detection,” Procedia Engineering 30, 2011, pp. 1–9.

[10] M. Govindarajan, R.M. Chandrasekaran, ”Intrusion detection using neural based hybrid classification methods,” Journal Computer Networks 55 ,2011, pp. 1662-1671.

[11] H.H. Hsu, C.W. Hsieh, M.D. Lu, “Hybrid feature selection by combining filters and wrappers,” Expert Systems with Applications 38, 2011, pp. 8144–8150.

[12] S.L. Humpherys, K.C. Moffitt, M.B. Burns, J.K. Burgoon, W.F. Felix, “Identification of fraudulent financial statements using linguistic credibility analysis,” Decision Support Systems, 2011, pp. 585-594.

[13] K. Badran, P. Rockett, “Multiclass pattern classification using single, multi-dimensional feature-space feature extraction evolved by multi-objective genetic programming and its application to network intrusion detection,” Springer US Genetic Programming and Evolvable Machines Journal 13(1), 2012, pp. 33-63.

[14] G. Chun, Z. Yajian, Y. Ping, Z. Zhang, G. Lio, Y. Yang, “A distance sum-based hybrid method for intrusion detection,” Springer US Applied Intelligence Journal, 2014, pp. 178-188.

★★★