Global Journal of Computer Science and Technology
Network, Web & Security
Volume 13 Issue 13 Version 1.0 Year 2013
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Ensemble of Soft Computing Techniques for Intrusion Detection

By Deepika Veerwal, Naveen Choudhary & Dharm Singh
Maharana Pratap University of Agriculture and Technology, India

Abstract - In the present world scenario network-based computer systems have started to play progressively more vital roles. As a result they have become the main targets of our adversaries. To apply high security against intrusions and attacks, a number of software tools are being currently developed. To solve the problem of intrusion detection a number of pattern recognition and machine learning algorithms has been proposed. The paper states the problem of classifier fusion with soft labels for Intrusion Detection. Performance of Artificial Neural Networks (ANN) and Support Vector Machines (SVM) is presented here. The performance of fusing these classifiers using approaches based on Dempster- Shafer Theory, Average Bayes Combination and Neural Network is proposed. As shown through the experimental results combined classifiers perform better than the individual classifiers.

Keywords : *intrusion detection, pattern classification, multiple classifier fusion, decision fusion, artificial neural network.*

GJCST-E Classification : *C.2.0*

Strictly as per the compliance and regulations of:

Ensemble of Soft Computing Techniques for Intrusion Detection

Deepika Veerwal [α], Naveen Choudhary [σ] & Dharm Singh [ρ]

Abstract - In the present world scenario network-based computer systems have started to play progressively more vital roles. As a result they have become the main targets of our adversaries. To apply high security against intrusions and attacks, a number of software tools are being currently developed. To solve the problem of intrusion detection a number of pattern recognition and machine learning algorithms has been proposed. The paper states the problem of classifier fusion with soft labels for Intrusion Detection. Performance of Artificial Neural Networks (ANN) and Support Vector Machines (SVM) is presented here. The performance of fusing these classifiers using approaches based on Dempster-Shafer Theory, Average Bayes Combination and Neural Network is proposed. As shown through the experimental results combined classifiers perform better than the individual classifiers.

Keywords : intrusion detection, pattern classification, multiple classifier fusion, decision fusion, artificial neural network.

I. Introduction

As Internet is evolving rapidly, dependency on computer networks has been increased. The threat of computer crimes is increasing as computer technology is evolving and the detection and preemption of such infringement become more and more intricate. A set of actions that tries to break the availability, confidentiality or integrity of the resource is termed as Intrusion (Debar, Dacier & Wespi, 2000). Intrusions and attacks violates the security policies of a computer system illegally, malicious break-in into a computer system or representing a system unusable or unreliable. Most system security mechanisms are intended to prevent unauthorized access to system data as well as the resources.

An intrusion detection system actively monitors the functioning of the system, and decides whether these functions are indication of an attack or constitute a genuine and valid use of the system (Lee & Stolfo, 1998).

In pursuance of detecting attacks, intrusion detection system is classed into two types. First is misuse detection, which is based on the signatures of attacks. The main objective of misuse detection is to signify attacks in the form of signatures so that if the same attack appears in future it can be easily detected and prevented. It is typically linked to a large database of attack signatures. A different way to deal with this difficulty is to follow the model proposed by Denning (Denning, 1987). Anomaly detection is rooted on defining the network behavior. It works on hypothesis that abnormal behavior is infrequent and different from normal behavior. For this reason, it creates models for normal behavior. If the system's behavior deviates from the normal behavior, then it is considered as intrusion or attack.

In rest of the paper, a concise introduction to the similar work in the area of intrusion detection is presented in section 2. A brief introduction to the proposed work is presented in section 3. In section 4 we present the experimental results of ANNs, SVMs, and their ensemble as well as the comparison from the past work. In section 5, we conclude our results.

II. Related Work

Various soft computing techniques have been applied to anomaly detection because of its benefit of finding important learning that tells about the user's behavior from large audit datasets (Lee & Stolfo, 1998; Lee, Stolfo & Mok, 1998). The main data mining techniques used are Statistics (Anderson, Lunt, Javitz, Tamaru & Valdes, 1995), Artificial Neural Network (ANN) (Lippmann, 2000; Fox, Henning, Reed & Simonian, 1990; Debar, Becker & Siboni, 1992; Cannady, 1998; Mukkamala, Sung & Abraham, 2005) and Support Vector Machines (SVM) (Mukkamala, Sung & Abraham, 2005) for misuse and anomaly detections. In particular we reviewed techniques which used ANN, SVM and Multi Classifier Systems. (Mukkamala, Sung & Abraham, 2005; Giacinto, Roli & Didaci, 2003; Cordella, Limongiello & Sansone, 2004; Giacinto, Perdisci, Delrio & Roli, 2008) proposes the Multiple Classifier Systems (MCS) for intrusion detection. In the above mentioned works, classifiers have been trained on different network services (for example ftp, mail, etc.). (Giacinto, Roli & Didaci, 2003) was trained to detect different and new types of attacks (Giacinto, Perdisci, Delrio & Roli, 2008). A new approach of serial combination of classifiers was proposed in (Cordella, Limongiello & Sansone, 2004). Different classifiers process network traffic serially. Classifier has to decide on every stage whether the observed pattern is from one of the attack class or not. If

Authors α σ ρ : Department of Computer Science & Engineering, College of Technology and Engineering, Maharana Pratap University of Agriculture and Technology, Udaipur, Rajasthan, India.
E-mails : veerwaldeepika@gmail.com, naveenc121@yahoo.com, dharm@mpuat.com

it is not then it has to be further forwarded to the next stage of the classifier trained on different cases. Reported results exhibit that MCS improves the efficiency and performance of IDS which were based on statistical pattern recognition techniques.

III. Proposed Work

Some or more errors are always produced by the above mentioned individual methods, inspite of complete or trustworthy input information. As some techniques gives better result on some set of data, we can assume that different methods performing on different set of data will result out in different errors. Based on this assumption we can ensemble individual techniques performing well on different set of data so that the system's efficiency can be increased. By combining multiple expert techniques, it will surely detract overall classification error and as a result correct outputs will be highlighted.

Information (data) can be fused out on three levels of abstraction: data level, feature level and lastly at classifier level.

Various methods have been developed for ensemble of individual classifiers, also referred as classifier fusion or ensemble of experts or decision fusion. Classifier fusion techniques are broadly classified in two general groups. Objective of methods that comes under first group is to emphasize on the development of structure of classifier. Classifier's outputs is of no use till the combination process find out the single best classifier or a group of classifiers whose output improves the performance of the system and after the selection only their outputs are taken for consideration to make a final decision or for the next stage of processing (Xu, Krzyzak & Suen 1992; Shafer, 1976). The other group of methods operate mainly on classifier's outputs, and effectively the combination of classifier's outputs is calculated (Ruta & Gabrys, 2000; Rogova & Menon, 1998; Zhang, 2002).

As it is known that the traffic pattern is either normal or malicious. It can be obtained from various sources such as the content information, the traffic statistics and other basic connection information. The proposed work take the benefit of this fact, and endeavor to build a classifier which will able to combine information from various sources to make the decision more informed.

As we know that the traffic patterns changes with time and a classifier that is able to adapt these changes is desirable. The proposed work also tries to incorporate this adaptability in the classifier. So if a new type of attack appears, the classifier would sense it as a previously unseen pattern, and try to update the functioning of the classifier so that it can detect this type of attack in future.

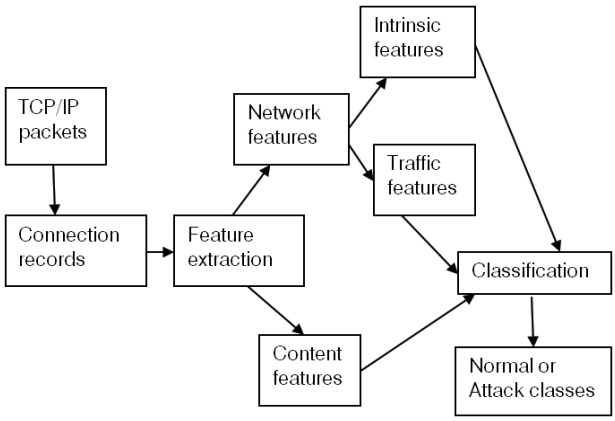

Figure 1 : Pattern Recognition in IDS

Fig. 1 shows that how the classifier is trained to classify attacks. Depending on the algorithm and the classifier model used, the ability to train and predict different attacks would be different. To solve the problem of intrusion detection a number of pattern recognition and machine learning algorithms has been proposed. We extend these models by fusing outputs from different models and reaching at a consensual decision amongst all classifiers. This approach is also known as classifier ensemble or consensus aggregation.

Our aim is to model a classifier with multiple fused algorithms which can result in better accuracy for all types of attacks. To achieve better level of accuracy from individual algorithms, first step would be designing of different connectional models in order to attain the improved performance for classifiers.

In the next step, test data is passed through the individual connectional models and then the related outputs are recorded. Assume that the classification performance given by ANN (RP), ANN (SCG) and SVM are pn; qn; and rn respectively, and the corresponding desired value is sn: Then, our task will be to combine pn; qn; and rn in order to get the best output value which will maximize the classification accuracy. In the proposed ensemble intrusion detection approach we used the Dempster Shafer Theory, Baye's theory and Neural Network Combination.

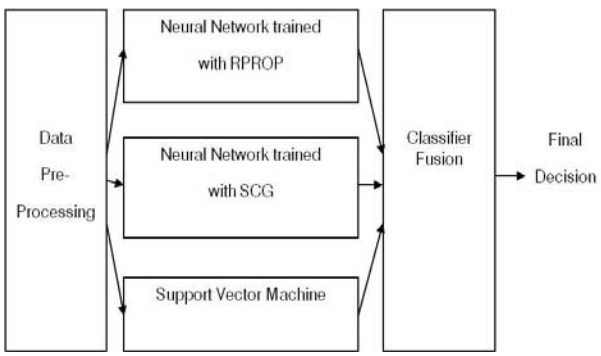

Figure 2 : Classifier fusion for ids

Fig. 2 shows the proposed system architecture. In the experiment we have fused two different Multi Layer Feed Forward Neural Networks trained using two different training algorithms and one Support Vector Machine. We analyzed three fusion strategies, namely Dempster Shafer Theory based Fusion, Bayesian Fusion and Neural Network Combination.

Dataset has been trained and tested on individual classifier as well as on the multiple fused classifiers.

IV. Experimental Results

In our experiments KDD99 dataset is used from the DARPA98 network traffic dataset. Individual TCP packets are assembled by TCP connections. Here, we perform five class classifications. Class 1 represents normal data, Class 2 represents DoS, Class 3 represents Probe, Class 4 represents remote to local, Class 5 represents user to root.

The whole dataset is divided in two parts: Train Set and Test Set with 494021 and 311029 records respectively. For reasons of time and computational complexity we have only taken into account any 30000 records for training and 20000 records for testing the classifiers. The selection of these records was random and stratified selection was done for all classes except for U2R and R2L. We used bootstrap method (Giacinto, Perdisci, Delrio & Roli, 2008) to increase the number of samples from the U2R class and R2L class since they are poorly represented in the dataset.

Status of a TCP connection being normal or some specific type of attack can be predicted on the basis of a set of 41 features. Out of 41 features, 38 are numeric features and rests 3 are symbolic features. From the 41 features, 11 features are discarded as they represent a constant value for all the connections. These discarded features are used for other services. Remaining 30 features are further divided into 3 classes as 4 intrinsic features, 7 content feature and 19 traffic features.

IDS is more efficient if it can make correct predictions. To show the performance of IDS, a confusion matrix is created. It consists of four outcomes namely TP, TN, FP, and FN. TP and TN are the correct prediction that the data is normal and attack respectively. FP and FN are the incorrect prediction that the data is normal and attack respectively. Based on this, we use the measures shown in Table 2 to quantify the performance of IDSs. A good IDS will have low FPR and high TPR.

Table 1 : Confusion Matrix

Predicted Positive	True Positive (TP)	False Positive (FP)
Predicted Negative	False Negative (FN)	True Negative (TN)
	Actually Positive	**Actually Negative**

Table 2 : Performance Measures

Measure	Formula
True Positive Rate (TPR)	TP/(TP+FN)
False Positive Rate (FPR)	FP/(FP+TN)
Accuracy (AC)	(TP+TN)/(TP+TN+FN+FP)

- TPR is the ratio of positive cases correctly identified to the actually positive cases.
- FPR is the ratio of negatives cases incorrectly classified as positive to the actually negative cases.
- AC is the proportion of total number of correct predictions.

From the confusion matrix generated by the Neural Network Resilient Back Propagation (RP) technique, Neural Network Scaled Conjugate Gradient (SCG) technique, Support Vector Machine (SVM) technique, Fusion using Dempster Shafer Theory (DST), Fusion using Average Bayes Combination, and Fusion using Neural Network Combination, we have calculated AC, FPR and TPR.

Table 3 : Average Global Performance of classifiers

	AC	FPR	TPR
RP	0.9277	0.0723	0.85378
SCG	0.95266	0.04734	0.84522
SVM	0.92102	0.07898	0.88712
DST	0.96304	0.03696	0.868
Bayes	0.93724	0.06276	0.87578
NN_comb	0.95234	0.04766	0.84516

Table 3 shows the average AC, TPR and FPR of all classifiers. From this table we can show that DST has the highest AC and lowest FPR. Bayes fusion has the good TPR but the worst AC and poor FPR. This is not acceptable. Hence among the six classifiers presented here DST performs the best as Dempster Shafer Theory works on the principle of combining evidence from different sources and based on that a belief is made which is useful for making the decision.

a) Comparative Study with other Methods

To compare proposed methodology with other existing popular techniques we selected two methods from (Mukkamala, Sung & Abraham, 2005) and (Giacinto, Roli & Didaci, 2003). (Mukkamala, Sung &

Abraham, 2005) fuses the decision from three neural networks and SVM. They follow a majority voting strategy to fuse the classifier decisions wherein the class which gets the most votes is selected. (Giacinto, Roli & Didaci, 2003) uses decisions from multiple neural networks trained on different feature subsets of the data and fuses their decision. There are five fusion strategies that were described in the work out of which the linear combination of classifiers performed the best. We too follow the same strategy and compare with our results in table 4 and fig. 3.

Table 4 : Comparison with other Approaches

	AC	FPR	TPR
DST	0.96304	0.03696	0.868
Bayes	0.93724	0.06276	0.87578
NN_comb	0.95234	0.04766	0.84516
Mukkamala, Sung & Abhraham (2005)	0.9545	0.04921	0.8591
Giacinto, Roli & Didaci (2003)	0.94808	0.04418	0.85907

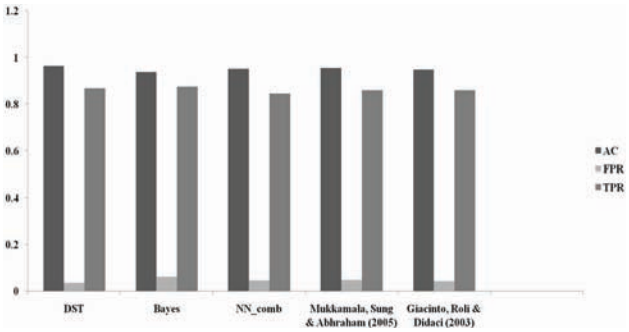

Figure 4 : Graphical representation of Table 4

V. Conclusion

The proposed methodologies evidently shown the significance of using ensemble approach based on distinct feature representation for modeling IDSs. In this paper, we have experimented ensemble IDSs with three different fusion techniques. Intrusion Detection can be analyzed as a pattern recognition (classification) task. From the experiments we carried out we can say that the MCF approach provides better accuracy with low false alarm generation than that provided by an individual classifier trained on the training data set. Multi classifier paradigms do not always give better performance. In some cases when the evidence is highly conflicting some fusion strategies fail. Out of the three fusion techniques studied, Dempster Shafer Theory based fusion performs the best. So instead of developing an accurate classifier we can develop many weak classifiers and combine them using Dempster Shafer Theory to get a good result in terms of accuracy and low false alarm rate among the Neural Network combination, Bayesian Fusion as well as from the past work of (Mukkamala, Sung & Abraham, 2005) and (Giacinto, Roli & Didaci, 2003).

References Références Referencias

1. Anderson, D., Lunt, T.F., Javitz, H., Tamaru, A., Valdes, A. (1995). Detecting unusual program behavior using the statistical component of the next-generation intrusion detection expert system (nides). SRI-CSL, 95(06).
2. Cannady, J. (1998). Artificial neural networks for misuse detection. NATIONAL INFORMATION SYSTEMS SECURITY CONFERENCE, 7(6).
3. Cordella, L.P., Limongiello, R., Sansone, C. (2004). Network intrusion detection by a multi stage classification system. Multiple Classifier Systems, LNCS 3077. Springer, 324-333.
4. Debar, H., Becker, M., Siboni, D. (1992). A neural network component for an intrusion detection system. Research in Computer Security and Privacy, 240-250.
5. Debar, H., Dacier, M., Wespi, A. (2000). A revised taxonomy for intrusion-detection systems. Annals of Telecommunications, 55(7), 361-378.
6. Denning, D.E. (1987). An intrusion-detection model. IEEE Transactions on Software Engineering, 13(2), 222-232.
7. Fox, K.L., Henning, R.R., Reed, J.H., Simonian, R. (1990). A Neural Network Approach Towards Intrusion Detection. Proceedings of the 13th National Computer Security Conference, 125-134
8. Giacinto, G., Roli, F., Didaci, L. (2003). Fusion of multiple classifiers for intrusion detection in computer networks. Pattern Recognition Letters, 24(12), 1795-1803.
9. Giacinto, G., Perdisci, R., Delrio, M., Roli, F. (2008). Intrusion detection in computer networks by a modular ensemble of one-class classifiers. Information Fusion, 9(1), pp. 69-82.
10. Lee, W., Stolfo, S.J. (1998). Data mining approaches for intrusion detection. Computer, 7(6).
11. Lee, W., Stolfo, S.J., Mok, K.W. (1998). Mining audit data to build intrusion detection models. In Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining, 66-72.
12. Lippmann, R. (2000). Improving intrusion detection performance using keyword selection and neural networks. Computer Networks, 34(4), 597-603.
13. Mukkamala, S., Sung, A.H., Abraham, A. (2005). Intrusion detection using ensemble of soft computing paradigms. Neural Networks, 28(3), 233-248.
14. Rogova, G.L., Menon, R. (1998). Learning in distributed systems for decision making. Center for multisource information fusion.
15. Ruta, D., Gabrys, B. (2000). An overview of classifier fusion methods. Computing and Information Systems, 7(1).

16. Shafer, G. (1976). A mathematical theory of evidence. Princeton university press.
17. Xu, L., Krzyzak, A., Suen, C.Y. (1992). Methods of combining multiple classifiers and their applications to handwriting recognition. IEEE Transactions on Systems, Man, and Cybernetics, 22(3), 418-435.
18. Zhang, B. (2002). Class-wise multi-classifier combination based on dempster-shafer theory. In Proceedings the Seventh International Conference on Control, Automation, Robotics and Vision (Icarv 2002).

This page is intentionally left blank