International Journal of Security and Its Applications
Vol.8, No.5 (2014), pp.311-322
http://dx.doi.org/10.14257/ijsia.2014.8.5.28

An Effective Intrusion Detection System Based onMulti-layers Mining Methods

Ming Yao

Baotou Vocational & Technical College, Baotou, China
20858242@qq.com

Abstract

In this paper, we propose a multi-layer selection and mining methods for effective intrusion detection, which utilize feature selection, classification, clustering and evidence theory for decision making. In the experiments, DARPA KDD-99 intrusion detection data set is used for evaluation. It shows that our proposed classifier not only classifies and separates the normal and abnormal data, but also reduces false positive and false negative besides detecting all four kinds of attacks.

Keywords: *Network security, Intrusion detection, Feature selection, Classification, Clustering, Dempster-Shafertheory*

1. Introduction

The security of computer networks is one of the major concerns. It has further become a challenge because of rapid developments in computing and communication technologies. Such challenge of system security will always be there because the techniques which were once supposed to be very powerful, become either obsolete or of little use as technological developments take place. Intrusion detection aims at detecting the intruders in computer networks and providing security to them. Development of intrusion detection systems was started from the work of Anderson: Computer Security Threat Monitoring and Surveillance [1]. Since then several researchers have been working on IDSs. Denning developed the Intrusion Detection Expert System (IDES), a very important step in the journey of intrusion detection systems, based on user behavior that was used to analyze the audit trails of mainframe computers and to build profiles of user activities [2]. In past, ensemble based intrusion detection techniques have been investigated. Ensemble means the combinations of some or all of the base classifiers in such a way that the resultant combination can classify the input data more precisely [3, 4]. Combining several identical classifiers however does not provide any better results. In [5], it is reported that there should be diversity among the base classifiers to generate an ensemble classifier, which can be obtained by maximizing coverage of the data, the percentage of data that can be classified correctly by at least one base classifier.

2. Related Works

In last couple of years, there have been various approaches for designing intrusion detection systems, which primarily require data collection that are used for design purpose. Some of the data may have redundant features and also have ambiguity, which lead to reduction in performance. One of the first techniques consists of a set of base features selecting classifier, *e.g.*, the work of Giacinto and Roli [7]. In that work, the problem domain

ISSN: 1738-9976 IJSIA

was the ftp service of the DARPA KDD-99 data set [8] that selected 30 out of 41 features from the dataset. They built three neural networks using 4 intrinsic features, 19 traffic features, and 7 content features. They also built one neural network using all of 30 selected features for the sake of comparison. All the networks had 5 output neurons (one normal and four attack classes), and a number of input neurons equal to the number of features. In this work, it is reported that the ensemble based techniques improve the detection rate as compared to the individual base feature selecting classifier. Their experiments consisted of only on ftp instead of all the services provided by the KDD-99 dataset. The papers [9, 10] discuss different soft computing techniques in every individual base classifier by using KDD-99 training dataset in both training and testing. In this paper, five classes were discussed, which are normal, DoS, Probe, U2R, and R2L. In these works, four base classifiers: neural networks, SVM, k-nearest neighbour (k-NN [11], and decision trees were used to improve classification individually and then fuse their inferences using three combination strategies: majority voting, average rule and belief function. The ensemble model had overall performance 99.68% detection rate (DR) and 0.87% false positive rate (FPR). The paper [12] discusses IDS based on artificial neural networks and support vector machine and reports the detection rate 43.6% to 100% and false positive rate 0.27% to 8.53% using different thresholds for 250 attacks and 41,426 normal sessions on KDD'97. The paper [13] discusses an experimental framework for comparative analysis of both supervised and unsupervised learning techniques and reports 95% DR and 1% FPR using C.45 algorithm [14]. It is well known that principal component analysis (PCA) is one of the most popular feature reduction and data compression methods that have also been applied to design IDS [15]. In [16], neural network principal component analysis (NNPCA) and nonlinear component analysis (NLCA) are discussed to reduce the dimension of network traffic patterns by comparing information of the compressed data with that of the original data. In [17], PCA has been used to detect selected denial-of-service and network probe attacks. The loading values of the various feature vector components with respect to the principal components have been analyzed. In [18], an ensemble method for intrusion detection is discussed by considering two types of classifiers; ANN and SVM. Another ensemble method is proposed in [19] that consider every individual classifier as independent by a diverse soft computing technique as well as different feature subset for each classifier and then the results of the ensemble members are combined. In this paper we propose a new combining classifier approach to intrusion detection by considering a set of heterogeneous classifiers. Four different base classifiers perform classification over an input pattern. Results are then combined using three combing methodologies: basic features, content features and special features classifiers. The work of Mukkamala, *et al.,* [20] also used KDD-99 training dataset and performed classification in five classes: normal, DoS, Probe, U2R, and R2L. Their ensemble model consists of one of three multilayer feed forward neural network, Support Vector Machine (SVM) and multivariate adaptive regression splines (MARS). They have used the majority voting technique, which combines the outcomes of the individual base classifier. In that work also, the ensemble approach has performed better than each individual base classifier. Hansen and Salamon [21] have proved that multi-classifiers work only when it is possible to build individual classifiers which are more than 50% accurate. All these works have focused only in detection rate in known and unknown intrusions but they do not consider reducing the false alarm rate.

3. Our Proposed Approach

We first use the K-Means algorithm for clustering and the intrusion detection benchmark dataset KDD-99 for training and testing purposes. The data is first clustered

using K-means and the records of the selected features may belong to one of the five clusters: normal, U2R, R2L, probe, DoS. The records are labeled with the cluster indexes. In real life dataset, generally there is no clear boundary between normal and abnormal activity of a user because some patterns of attacks are similar to the normal activities. Therefore, we select a variety of supervised learning techniques to deal with the vagueness in real life such as fuzzy K-NN classifier, Naïve Bayes Classifier, decision tree based classifier. All these techniques can provide a dynamic decision boundary of network traffic to the network connections. For the ensemble based classifiers, 3 partial feature subsets, 9 basic features (1-9), 9 content features (10-18), and 23 traffic features(19-41), of the original KDD-99 41 features are applied to the three base feature selecting classifiers. After the base feature selecting classifier's classification is done, the results from different classifiers and the hybrid module are combined with the output of the data mining based module. We carry out the following fusion techniques: majority voting, average rule, Bayesian Combination method. In Figure 1, we have used four different types of classifiers: Fuzzy Belief k-NN, Decision Tree classifier, Naïve Bayes classifier, and Data mining based classifier. The first three classifiers are used in groups while the last one is data mining based classifier for detecting patterns of attacks in the incoming connections. First we cluster the data to obtain the class labels of the connections, then we use the hybrid module which is made up of a hybrid algorithm combining the k-nearest neighbor and Naïve Bayes for anomaly detection and the last one signature based part reduces the false alarm rate of the whole system (see Figure 1).

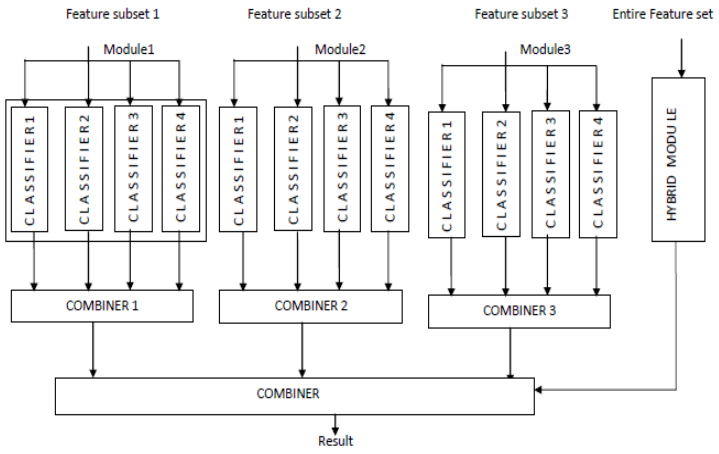

Figure 1. Framework of Proposed Approach

We now discuss different modules of the ensemble.

3.1. Classifier

We have used classifier, a model that assigns a class label to a data item described by a set of attributes. A classifier is first trained by a set of training patterns. A classifier is useful in pattern classification problems. Following are the classifiers that are used in our work.

3.1.1. Fuzzy Belief k-NN: Detecting intrusions can be treated as a classification task, *i.e.,* to classify the network traffic into normal or attack category. The k-NN method is effective in many pattern classification problems. For an input to be classified, a number of k-nearest

training patterns are obtained based on the Euclidean distance between the input and every training pattern. The input is assigned to the class that has got maximum number of votes (majority voting principle), *i.e.,* the input is classified to the most frequent class label among the k-nearest training patterns. The major drawback of this algorithm is that the precision of the classification may decrease if all the selected k- nearest training patterns are equally important (not considering the distances). Some intrusions may be similar to that of the normal activities, that is, the boundary between them (normal behavior and attack) is very unclear. To remove this drawback, fuzzy k-NN is discussed that assigns multiple membership grades to classes rather than a single class by using distance differences from the k- nearest patterns. The confidence values are in proportion with the membership grade. Fuzzy Belief kNN [22, 23] is different from fuzzy kNN in the sense that it uses fuzzy belief function to resolve conflicts.

3.1.2. Decision Tree based Classifier: the ID3 builds a decision tree in which each internal node denotes a test on attribute, each branch represents an outcome of the test, and the leaf node represents a class or class distributions. The top most nodes are the root node. For classifying an unknown sample, the attribute values of the sample are tested against the decision tree. A path is traced from the root to a leaf node that holds the class prediction for that sample. The ID3 algorithm builds a decision tree from the root node by choosing one remaining attribute with the highest information gain as the test for the current node. Here we have used C4.5 algorithm, a modified version of the ID3 algorithm for building the decision trees for classification.

3.1.3. Naive Bayes Classifier: This classifier is based on the conditional probability used in the classification problems. It uses Bayes' theorem with independent assumptions, *i.e.,* the set of features are assumed to be conditionally independent of each other. If there is a set of classes and a connection is to be assigned to one of the classes, the connection is assigned to the class with the highest probability. By applying the Naïve Bayes classifier to an intrusion detection system, the set of traffic data is used to find the prior probabilities for normal or attack. When unseen network traffic arrives, the classifier uses Bayes's theorem to decide whether the new traffic belongs to normal or attack class.

3.2. Clustering and Classification Method

We use feature selection, clustering, and classification. We have applied a hybrid algorithm, a combination of k-Nearest and Naïve Bayes Classifier, to the incoming network traffic data. Before applying the hybrid algorithm, we use a reduced set of attributes by using a feature selection algorithm because many attributes in the network traffic data are irrelevant with the context in which we are interested to work with. We have used entropy based feature selection method for selecting the attributes and removing the redundant and irrelevant ones. The algorithm consists of two parts. It first removes the irrelevant features with poor prediction ability and collects the relevant attributes to the target class, calculates the mutual information between the features and class. The algorithm arranges the features in descending order of their degrees of association to the target class. The features with information measure equals to zero are removed. In second part, it removes the redundant features that are intercorrelated with one or more features. We calculate the intercorrelation. In clustering, we have used the k-means algorithm that has the capability of clustering large amount of data. The main idea of this algorithm is to first start with k clusters, each consisting of only centroid. Then, associate each point in the given dataset to the nearest cluster using the Euclidean distance and recalculate the positions of centroids in all the clusters. We repeat this

process until the centroids become stable. The Euclidean distance d_{ij} between the objects *i* and *j* (x_{ik} is k-th attribute of ith object, x_{jk} is k-th attribute of jth object) is defined as

$$d_{ij} = \sqrt{\sum_{k=1}^{n}(x_{ik} - x_{jk})^2} \quad (1)$$

In the hybrid module, we have used a combination of k-NN and Naïve Bayes described above. The algorithm for hybrid classification is given below.

Algorithm

a. *Feature Selection:Removal of irrelevant features from original input dataset D that includes features X and target class T.*

i. For each feature f_i in X, calculate mutual information MU(T, f_i) and sort it in descending order.

ii. Put f_j, whose MU(T, f_i)>0, into relevant feature set Rxy and remove remaining features.

iii. For each feature f_j in set Rxy, calculate pairwise mutual information MU(f_i, f_j)

iv**.** Select those features having MU(f_i, f_j) >T, a predefined threshold, and put those features into set B, where mutual information MUxx=∑ MU(f_i, f_j)

v. Calculate means μ_x and μ_y of Rxx and R_{yy} of autocorrelation coefficients and then W=Rxx/Ryy, R=W*R_{yy} - R_{xx}

vi. Select f_j from set B, whose R>0, and put into final set F

b. *Clustering: application of K-Means Algorithm K-Means(Dataset, k), k is number of clusters=5.*

i. Obtain cluster indexes to append them to connection records and update a separate copy of dataset file.

ii. Take a part of connection records in the modified dataset table and apply those records to hybrid classification algorithm to build training normal dataset D.

iii. Take a part of dataset, say, D_j**.** For each record x in D_j do

If x is in database (of signatures),

Then x is anomalous

Else

Find scores of dist(x, y), for all x, y ЄD_j, for other record y.

iv. Arrange distances in ascending order.

v. Find first k shortest distances and pick up the first shortest k nearest neighbors

If(voting(x, N) < voting(x, A)) thenx is Normal

Elseif(voting(x, N)>voting(x, A)) then x is abnormal

Else

Calculate class conditional and prior probabilities for Naïve Bayes' classifier.

vi.Estimation of sample,for example, x belongs to cluster $K_j$if posterior probability $P(k_j/x)$ is minimum for all j=1,2,3….n (n=5).

The posterior probability is given by

$$P(k_j/x) = \frac{P(x/k_j)P(k_j)}{P(x)}, P(x/k_j) = P(k_j)\prod_{i=1\to n} P(x_i/k_j),$$

Prior probability $P(k_j) = \frac{\sum_{i=1}^{n} t_{i\to K_j}}{\sum_{i=1}^{n} t_i}$, and conditional probability

$$P(x_i/k_j) = \frac{\sum_{i=1}^{n} x_{i\to k_j}}{\sum_{i=1}^{n} t_{i\to k_j}}.$$

3.3. Data Mining based Classifier

After the process of ensemble classification, the main concern is to reduce false alarm rate. The most common problem of anomaly intrusion detection is high false alarm rate that occurs when the given pattern deviates from the normal behavior. Thus, a two level model is used that includes the anomaly and data mining based part. The data mining based part acts as rule or misuse based part. Its function is to find the required information from the data and deduce inferences. Its function is to find the rules to extract the patterns of normal behavior from the training data. We have used some signature based algorithms for extracting the patterns of normal behavior from the given training data. Thus, we first train with normal behavior of traffic stream so that in the testing phase the new pattern can be checked if it has seen that pattern in training phase. If yes, it will assume the incoming connection as normal, otherwise as an attack.

We can write the decision rules as

Rule: *IF condition of Feature1 <value> then,*
The traffic is a normal behavior
Otherwise it is an attack

In each rule, the first part consists of a number of conditions that are satisfied by a number of features. The consequent part of the rule decides what actions should be taken. The data mining classifier compares the result of ensemble features selecting classifier with the well defined normal patterns that were used for training. The data mining classifier will only check the data when there is disagreement between the result of ensemble feature selecting classifier and itself.

3.4. Evidence Theory based Combiner

Appropriate choice of combination method has an effect on the performance. There are a variety of combination methods that can be classified as linear and nonlinear methods, statistical-based methods, and some computationally intelligent methods. The linear combination method includes summation and average and the nonlinear combination method includes majority voting. The computationally intelligent methods include fuzzy logic, neural networks, and genetic algorithms. We have used different combinations of the methods to carry out some fusion techniques, e.g. majority voting, average rule, Dempster-Shafer technique method to combine the outputs together. Majority voting rule assigns the incoming network connection to be assigned to the majority class. The average rule assigns the network traffic to the maximum value of summation of the posterior probability divided by the number of network classifiers.

Dempster-Shafer and Bayesian combination methods assign the input network traffic to the class with highest belief value. But the Bayesian combination method involves computation of the prior probability of each class, whereas the Dempster-Shafer computes the probability to indicate the attack or normal classes. It has been shown that the integration of the different base classifiers can produce better result.

In this paper, different base classifiers created by clustering, Naive Bayes, Decision trees are integrated and an ensemble based approach that combines the results of different classifiers by using different techniques described above is used. All classifiers have been assigned weights and the results of the classifiers having false positive rate less than T are combined and taken into account. The threshold T, user defined threshold, has been incorporated for reducing the false alarm rate. There are n classifiers, each having five different classes: Normal, DoS, U2R, R2L, Probe. We

denote wj as jth classifier weight, pj as false positive rate of the first j classifiers, mi as sum of first i classifiers, c as sum of first j classifier sj.

$$m_i = \sum_{j=i}^{i} p_j(x) w_j c_i^j, \text{where } w_j \varepsilon(0,1) \quad (2)$$

$$p_j(x) = \begin{cases} 1, \text{if false positive rate} < 0.6 \\ 0, \text{otherwise} \end{cases} \quad (3)$$

By integrating the results from all classifiers, the false positive rate can also be reduced.

Decision Algorithm

a. Given training dataset $T=\{(x_1,y_1), (x_2,y_2),\ldots,(x_n,y_n)\}$, where n is number of training samples, (x_i,y_i) denotes ith instance of training samples. For $B=\{b_1, b_2,...,b_m\}$, m refers to number of base classifiers, each base classifier was obtained by its corresponding algorithm.
b. After preprocessing dataset T, randomly select a set of samples $S = \{s_1, s_2,..,s_k\}$, a subset of T. Build the classifiers $b_1, b_2,\ldots,b_m$ by their corresponding classification algorithm
c. Calculate $C_j = \sum_{i=1}^{n} c_j^i$, where c_j^i is priority value of ith class in jth classifier.
d. Results of all classifiers by weighted majority voting rule, and all classifiers having a false positive rate greater than a threshold are combined and their combined classification result is the output.

4. Experimental Evaluation

For evaluating the performance of our proposed approach, we have used DARPA KDD-99 benchmark dataset. First, we describe the content of the dataset.

4.1. Data Set Introduction

The data set used for our experiments is DARPA KDD-99 benchmark dataset, which is also known as 'DARPA Intrusion Detection Evaluation dataset'. It includes three independent sets: whole KDD, 10% KDD, and corrected KDD. Here, we have taken 10% KDD and corrected KDD as training set and testing set, respectively. The training set contains a total of 22 training attack types, with an additional 17 types in the testing set only. Total 39 attack types are included, which fall into four main classes: denial of service (DoS), probe, user to root (U2R), and remote to local (R2L).

Feature selection is used only in the hybrid module part. It is not used in the first three classifiers because they all work upon all 41 features of the DARPA KDD CUP-99 dataset. So, we do not need feature selection. It is used to remove the redundant and irrelevant attributes from the feature set and gives only necessary attributes. Since we have to use all 41 features of the dataset for clustering and classifiers, if we do not use feature selection, it will take more time. So, we first do feature selection, followed by clustering and classification in order to operate the algorithm faster.

4.2. Experimental Results

In our experiments, we use standard measurements such as detection rate (DR), false positive rate (FPR) and overall classification rates (CR) to calculate the performance of the system. The terms True Positive

Table 1. Performance of These Classifiers in Four Groups

		Group 1			Group 2			Group 3		
		DR	FPR	CR	DR	FPR	CR	DR	FPR	CR
Layer 1	Classifier 1	80.01	20.34	80.45	68.93	3.89	81.56	89.92	6.45	80.56
	Classifier 2	90.54	45.27	77.45	71.67	12.41	76.45	71.01	1.45	56.34
	Classifier 3	76.56	17.02	70.16	87.40	11.02	69.33	76.94	20.97	81.34

Table 2. Performance of Combiners of Layers 2 and 3

		Majority voting			Average Rule			Dampster-Shafer		
		DR	FPR	CR	DR	FPR	CR	DR	FPR	CR
Layer 2	Combiner 1	81.90	21.87	82.71	90.01	31.95	78.54	86.82	4.10	78.87
	Combiner 2	78.90	16.71	86.76	53.65	8.75	65.89	78.93	0.05	69.45
	Combiner 3	71.67	7.39	85.92	79.93	4.84	85.76	77.56	8.68	81.56
Layer 3	Final Result	79.51	6.81	80.56	80.04	4.17	81.89	79.94	5.66	73.67

Table 3. Performance of Three Classifier using Full Feature Set

	DR(%)	FPR(%)	CR(%)
Classifier 1	71.76	8.93	69.03
Classifier 2	68.72	9.45	75.45
Classifier 3	79.45	1.56	80.05

Table 4. Detection Rate on Four Attack Groups using Ensemble based Method

Attack	DR(%)
DoS	95.69
Probe	49.19
U2R	2.79
R2L	4.04
Normal	97.01

Table 5. Classification Performance of Proposed Intrusion Detection System

	Metric	Proposed System	
		Training	Testing
Probe	Precision	0.891417	0.891417
	Recall	0.4419	0.4419
	Accuracy	0.89564	0.89128
DoS	Precision	0.902151	0.902371
	Recall	0.85697	0.85871
	Accuracy	0.9167	0.917618
U2R	Precision	0.06781	0.06781
	Recall	0.210105	0.210105
	Accuracy	0.952613	0.953623
R2L	Precision	0.0843	0.0843
	Recall	0.29917	0.29917

	Accuracy	0.938437	0.938751
NORMAL	Precision	0.868476	0.868417
	Recall	0.970177645	0.971754
	Accuracy	0.83456	0.82785

4.3. Discussions

The experimental results have been recorded in the form of a table that contains FPR (false positive rate), DR (detection rate), CR (classification rate) for three classifiers at different levels. In one of the classifiers we have use Fuzzy Belief *k*NN classification algorithm, which is run at least 20 times for each k to minimize the inaccuracy. The results show that this proposed feature selecting classifier outperforms the base classifiers. For this ensemble based classifier, we measure FPR (false positive rate), DR (detection rate), CR (classification rate) and the corresponding results for each of the classifiers in each of the layers are taken and then the final result is obtained. For the proposed ensemble based system, the overall detection rate of different types of attacks: DoS, Probe, U2R, R2L are 95.69%, 49.19%, 4.04%, 97.01%, respectively and the overall classification rate, accuracy, precision ,recall are given in Tables 1-5.

The detection rate means detection rate for the individual attacks U2R, R2L, DoS, and probe. We measure the detection rate with 1 Module, 2 Module, 3 Module. After getting the performance of individual base classifier, we then generate the final result by combining the outputs of the base classifiers. In each of the base modules, we have used different classifiers containing different classification algorithms. So this model has a combination of ensemble feature selecting classifiers and a data mining based classifier. With the combination of these, we get the final result. It is shown that the detection rate is initially low in comparison to the other base level modules, but as we increase the number of modules (number of classes) the detection rate slowly increases. There may be inaccuracies for using less modules, because the classifiers initially show to have very low classification performance, but as the time increases the classifiers learn more and start classifying more accurately. Since the classifiers are based on supervised learning, first we need to train all classifiers with proper labeled training dataset containing all types of intrusions. The reason behind this is that when the number of modules is less, their classification result are combined, which may not have been trained with the dataset containing the patterns of attack. Thus, it can detect only those kind of attacks for which it has been trained. But as we increase the number of modules, the number of classifiers will also increase. Since different classifiers may learn differently at the time of their learning, their classification performance can be different while classifying the data. However, if we combine the modules, all classifiers also are combined and their inaccuracies like misclassification rate get reduced. Thus, combining the outputs of different classifiers gives some improvement in the detection rate and by using some combination method in the combiners; the false alarm rate gets reduced because in the ensemble based approach we take the output of that classifier which has the least false alarm rate.

5. Conclusions

In this paper, we propose a multi-layer selection and mining methods for effective intrusion detection, which utilize feature selection, classification, clustering and evidence theory for decision making. In the experiments, DARPA KDD-99 intrusion detection data set is used for evaluation. It shows that our proposed classifier not only

classifies and separates the normal and abnormal data, but also reduces false positive and false negative besides detecting all four kinds of attacks.

References

[1] P. J. Anderson, "Computer Security Threat Monitoring and Surveillance", James P. Anderson Co., Fort Washington, PA, **(1980)**.
[2] T. G. Dietterich, "Ensemble Learning Methods", In: M.A. Arbib (ed.), Handbook of Brain Theory and Neural Networks, 2nd ed., MIT Press, **(2001).**
[3] L. Hansen and P. Salamon, "Neural network ensembles," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, **(1990),** pp. 993-1001.
[4] C. Brodley and T. Lane, "Creating and exploiting coverage and diversity," In: Proc. AAAI-96 Workshop on Integrating Multiple Learned Models, **(1996),** pp. 8-14, 1996.
[5] D. E. Denning, "An intrusion-detection model." IEEE Transactions on Software Engineering, vol. SE-13, no. 2, **(1987)** February, pp. 222-232.
[6] P. Innella, "The Evolution of Intrusion Detection Systems", Tetrad Digital Integrity LLC, **(2011)** June 3, http://www.symantec.com/connect/articles/evolution-intrusion-detection-systems.
[7] G. Giacinto and F. Roli, "Intrusion Detection in Computer Networks by Multiple Classifier Systems," 16th International Conference on Pattern Recognition, vol. 2, **(2002),** pp. 390-393.
[8] "KDD'99 archive: The Fifth International Conference on Knowledge Discovery and Data Mining", URL: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.
[9] A. Borji, "Combining Heterogeneous Classifiers for Network Intrusion Detection," Lecture Notes in Computer Science, Springer, vol. 4846, **(2008),** pp. 254-260.
[10] L.L. DeLooze, "Attack Characterization and Intrusion Detection using an Ensemble of Self-Organizing maps," 2006 International Joint Conference on Neural Networks, **(2006)** July**,** pp. 2121-2128, Vancouver, BC, Canada.
[11] T. Denoeux, "A k-Nearest Neighbor Classification Rule Based on Dempster-Shafer Theory," IEEE Transactions on Systems, Man and Cybernetics, vol. 25, no. 5, **(2002),** May, pp. 804-813.
[12] C. W. Hua, H. S. Hsun and S. H. Pin, "Application of SVM and ANN for intrusion detection," Comput. Oper. Res., vol. 32, no. 10, **(2005),** pp. 2617–2634.
[13] L. Pavel, D. Patrick, S. Christin and K. Rieck, "Learning Intrusion Detection: Supervised or Unsupervised", In: Roli, F., Vitulano, S. (eds.) ICIAP 2005. LNCS, Springer, Heidelberg, vol. 3617, **(2005),** pp. 50–57.
[14] J. R. Quinlan, "C4.5: Programs for Machine Learning", Morgan Kaufmann, **(1993)**.
[15] "Oja.: Principal components, minor components, and linear neural networks", Neural Networks, vol. 5, no. 6, **(1972),** pp. 927–935.
[16] G. K. Kuchimanchi, V. V. Phoha, K. S. Balagami and S. R. Gaddam, "Dimension reduction using feature extraction methods for Real-time misuse detection systems," Proc. of 2004 IEEE Workshop on Information Assurance and Security, West Point, NY, **(2004),** pp. 195–202.
[17] K. Labib, V. R. Vemuri, "Detecting and visualizing denial-of-service and network probe attacks using principal component analysis", Proc. of third Conf. on Security and Network Architectures, La Londe, France, **(2004)**.
[18] T. S Chou and T. N. Chou, "Hybrid Classifier Systems for Intrusion Detection", Seventh Annual Communications Networks and Services Research Conference, **(2009)** Ma7 11-13, pp. 286 – 291.
[19] L. Didaci, G. Giacinto and F. Roli, "Ensemble Learning for Intrusion Detection in Computer Networks", Proc. of Workshop su apprendimento automatico: metodi ed applicazioni, **(2006)**.
[20] S. Mukkamala, A. H. Sung and A. Abraham, "Intrusion Detection Using an Ensemble of Intelligent Paradigms," Journal of Network and Computer Applications, vol. 28, no. 2, **(2005),** pp. 167-182.
[21] L. K. Hansen and P. Salamon, "Neural Network Ensembles," IEEE Transactions on Pattern Analysis Machine Intelligence, vol. 12, no. 10, **(1990),** pp. 993-1001.
[22] T. G. Dietterich, "Ensemble Methods in Machine Learning", Proc. of the 1st Intern. Workshop on Multiple Classifier Systems, Cagliari, Italy, LNCS 1857, Springer, **(2000)** June, pp. 1-15.
[23] T. S Chou, J. Fan, S. Fan and K. Makki, "Ensemble of Machine Learning Algorithms for Intrusion Detection", IEEE International Conf. on Systems, Man and Cybernetics, **(2009)** October 11-14, pp. 3976 – 3980.

Author

Ming Yao, he associate professor,obtained his GCT at Inner Mongolia University,now working in Baotou Vocational&Technical College,the author is mainly engaged in the research of computer network and applications.