Intrusion Detection using an Ensemble of Classification Methods

M.Govindarajan and RM.Chandrasekaran

Abstract—**one of the major developments in machine learning in the past decade is the ensemble method, which finds highly accurate classifier by combining many moderately accurate component classifiers. This paper addresses using an ensemble of classification methods for intrusion detection. Due to increasing incidents of cyber attacks, building effective intrusion detection systems are essential for protecting information systems security, and yet it remains an elusive goal and a great challenge. In this research work, new hybrid classification method is proposed using classifiers in a heterogeneous environment using arcing classifier and their performances are analyzed in terms of accuracy. A Classifier ensemble is designed using a Radial Basis Function (RBF) and Support Vector Machine (SVM). Here, modified training sets are formed by resampling from original training set; classifiers constructed using these training sets and then combined by voting. Empirical results illustrate that the proposed hybrid systems provide more accurate intrusion detection systems.**

***Index Terms*— Classification Accuracy, Ensemble, Intrusion Detection, Radial Basis Function, Support Vector Machine.**

I. Introduction

Traditional protection techniques such as user authentication, data encryption, avoiding programming errors and firewalls are used as the first line of defense for computer security. If a password is weak and is compromised, user authentication cannot prevent unauthorized use, firewalls are vulnerable to errors in configuration and suspect to ambiguous or undefined security policies (Summers, 1997). They are generally unable to protect against malicious mobile code, insider attacks and unsecured modems. Programming errors cannot be avoided as the complexity of the system and application software is evolving rapidly leaving behind some exploitable weaknesses. Consequently, computer systems are likely to remain unsecured for the foreseeable future. Therefore, intrusion detection is required as an additional wall for protecting systems despite the prevention techniques. Intrusion detection is useful not only in detecting successful intrusions, but also in monitoring attempts to break security, which provides important information for timely countermeasures (Heady et al., 1990; Sundaram, 1996). Intrusion detection is classified into two types: misuse intrusion detection and anomaly intrusion detection.

Manuscript received April 26, 2012. Dr.M.Govindarajan is with the Annamalai University, Annamalai Nagar, Tamil Nadu, India (phone: 91-4144-221946; e-mail: govind_aucse@yahoo.com)
Dr.RM.Chandrasekaran is with Annamalai University, Annamalai Nagar, Tamil Nadu, India (phone: 91-4144-238444; e-mail: aurmc@sify.com)

Misuse intrusion detection uses well-defined patterns of the attack that exploit weaknesses in system and application software to identify the intrusions (Kumar and Spafford, 1995). These patterns are encoded in advance and used to match against user behavior to detect intrusions. Anomaly intrusion detection identifies deviations from the normal usage behavior patterns to identify the intrusion. The normal usage patterns are constructed from the statistical measures of the system features, for example, the CPU and I/O activities by a particular user or program. The behavior of the user is observed and any deviation from the constructed normal behavior is detected as intrusion.

Several machine-learning paradigms including neural networks (Mukkamala et al.,2003), linear genetic programming (LGP) (Mukkamala et al., 2004a), support vector machines (SVM), Bayesian networks, multivariate adaptive regression splines (MARS) (Mukkamala et al., 2004b) fuzzy inference systems (FISs) (Shah et al.,2004), etc. have been investigated for the design of IDS. In this paper, we investigate and evaluate the performance of decision trees (DT), SVM, hybrid DT–SVM and an ensemble approach. The motivation for using the hybrid approach is to improve the accuracy of the intrusion detection system when compared to using individual approaches. The primary objective of this paper is ensemble of radial basis function and Support Vector Machine is superior to individual approach for intrusion detection in terms of classification accuracy.

The rest of this paper is organized as follows: Section 2 describes the related work. Section 3 presents hybrid Intelligent Intrusion Detection System and Section 4 explains the performance evaluation measures. Section 5 focuses on the experimental results and discussion. Finally, results are summarized and concluded in section 6.

II. RELATED WORK

The Internet and online procedures is an essential tool of our daily life today. They have been used as an important component of business operation (T. Shon and J. Moon, 2007). Therefore, network security needs to be carefully concerned to provide secure information channels. Intrusion detection (ID) is a major research problem in network security, where the concept of ID was proposed by Anderson in 1980 (J.P. Anderson, 1980). ID is based on the assumption that the behavior of intruders is different from a legal user (W. Stallings, 2006). The goal of intrusion detection systems (IDS) is to identify unusual access or attacks to secure internal networks (C. Tsai , et al., 2009) Network-based IDS is a valuable tool for the defense-in-

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)
(revised on 3 October 2012)

depth of computer networks. It looks for known or potential malicious activities in network traffic and raises an alarm whenever a suspicious activity is detected. In general, IDSs can be divided into two techniques: misuse detection and anomaly detection (E. Biermann et al.,2001; T. Verwoerd, et al., 2002)

Misuse intrusion detection (signature-based detection) uses well-defined patterns of the malicious activity to identify intrusions (K. Ilgun et al., 1995; D. Marchette, 1999) However, it may not be able to alert the system administrator in case of a new attack. Anomaly detection attempts to model normal behavior profile. It identifies malicious traffic based on the deviations from the normal patterns, where the normal patterns are constructed from the statistical measures of the system features (S. Mukkamala, et al., 2002). The anomaly detection techniques have the advantage of detecting unknown attacks over the misuse detection technique (E. Lundin and E. Jonsson, 2002). Several machine learning techniques including neural networks, fuzzy logic (S. Wu and W. Banzhaf, 2010), support vector machines (SVM) (S. Mukkamala, et al., 2002; S. Wu and W. Banzhaf, 2010) have been studied for the design of IDS. In particular, these techniques are developed as classifiers, which are used to classify whether the incoming network traffics are normal or an attack.

Irrespective of whether good anomaly detection methods are used, the problems such as high false alarm rates, difficulty in finding proper features, and high performance requirements still exist. Therefore, if we are able to mix the advantages of both learning schemes in machine learning methods, according to their characteristics in the problem domain, then the combined approach can be used as an efficient means for detecting anomalous attacks. In this paper, we have chosen to focus on the Support Vector Machine (SVM) and Radial Basis Function (RBF) among various machine learning algorithms.

The most significant reason we chose the SVM is because it can be used for either supervised or unsupervised learning. Another positive aspect of SVM is that it is useful for finding a global minimum of the actual risk using structural risk minimization, since it can generalize well with kernel tricks even in high-dimensional spaces under little training sample conditions.

In Ghosh and Schwartzbard (1999), it is shown how neural networks can be employed for the anomaly and misuse detection. The works present an application of neural network to learn previous behavior since it can be utilized to detection of the future intrusions against systems. Experimental results indicate that neural networks are ‘‘suited to perform intrusion state of art detection and can generalize from previously observed behavior’’ according to the authors.
Freund and Schapire (1995,1996) propose an algorithm the basis of which is to **a**daptively **r**esample and **c**ombine (hence the acronym--arcing) so that the weights in the resampling are increased for those cases most often misclassified and the combining is done by weighted voting.

In this paper, we propose an anomaly intrusion detection system using radial basis function and support vector machine and evaluate the effectiveness of the proposed RBF-SVM hybrid system by conducting several experiments on NSL-KDD dataset. We examine the performance of the RBF-SVM hybrid classifier in comparison with standalone RBF and standalone SVM classifier.

III. HYBRID INTELLIGENT INTRUSION DETECTION SYSTEM

This section shows the proposed RBF-SVM hybrid system which involves Radial Basis Function (RBF) and Support Vector Machine (SVM) as base classifiers.

A. RBF-SVM Hybrid System

The proposed hybrid intelligent intrusion detection network system is composed of four main phases; Preprocessing phase, feature reduction phase, classification phase and Combining Phase. Figure 1 describes the structure of the hybrid intelligent intrusion detection network system.

1) NSL-KDD Dataset Preprocessing

Pre-processing of NSL-KDD dataset contains three processes; (1) Mapping symbolic features to numeric value, (2) Data scaling, since the data have significantly varying resolution and ranges. The attribute data are scaled to fall within the range [0, 1]. and (3) Assigning attack names to one of the five classes, 0 for normal, 1 for DoS (Denial of Service), 2 for U2R (User to Root), 3 for R2L (Remote to Local) , and 4 for Probe.

2) Dimensionality Reduction

Dimension Reduction techniques are proposed as a data pre-processing step. This process identifies a suitable low-dimensional representation of original data. Reducing the dimensionality improves the computational efficiency and accuracy of the data analysis.

Steps:

- ✓ Select the dataset.
- ✓ Perform discretization for pre-processing the data.
- ✓ Apply Best First Search algorithm to filter out redundant & super flows attributes.
- ✓ Using the redundant attributes apply classification algorithm and compare their performance.
- ✓ Identify the Best One.

a) Best first Search

Best First Search (BFS) uses classifier evaluation model to estimate the merits of attributes. The attributes with high merit value is considered as potential attributes and used for classification. It Searches the space of attribute subsets by augmenting with a backtracking facility. Best first may start with the empty set of attributes and search forward, or start with the full set of attributes and search backward, or start at any point and search in both directions.

3) Classification Methods

1) Radial basis Function Neural Network

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

The RBF (Oliver Buchtala, et al., 2005) design involves deciding on their centers and the sharpness (standard deviation) of their Gaussians. Generally, the centres and SD (standard deviations) are decided first by examining the vectors in the training data. RBF networks are trained in a similar way as MLP. The output layer weights are trained using the delta rule. The RBF networks used here may be defined as follows.

- ✓ RBF networks have three layers of nodes: input layer, hidden layer, and output layer.
- ✓ Feed-forward connections exist between input and hidden layers, between input and output layers (shortcut connections), and between hidden and output layers. Additionally, there are connections between a bias node and each output node. A scalar weight is associated with the connection between nodes.
- ✓ The activation of each input node (fanout) is equal to its external input where is the th element of the external input vector (pattern) of the network (denotes the number of the pattern).
- ✓ Each hidden node (neuron) determines the Euclidean distance between "its own" weight vector and the activations of the input nodes, i.e., the external input vector the distance is used as an input of a radial basis function in order to determine the activation of node. Here, Gaussian functions are employed. The parameter of node is the radius of the basis function; the vector is its center.
- ✓ Each output node (neuron) computes its activation as a weighted sum The external output vector of the network, consists of the activations of output nodes, i.e., The activation of a hidden node is high if the current input vector of the network is "similar" (depending on the value of the radius) to the center of its basis function. The center of a basis function can, therefore, be regarded as a prototype of a hyper spherical cluster in the input space of the network. The radius of the cluster is given by the value of the radius parameter.

2) Support Vector Machine

The support vector machine (SVM) is a recently developed technique for multi dimensional function approximation. The objective of support vector machines is to determine a classifier or regression function which minimizes the empirical risk (that is the training set error) and the confidence interval (which corresponds to the generalization or test set error) (Vapnik, V, 1998).

Given a set of N linearly separable training examples $S = \{x_i \in R^n | i = 1,2,...,N\}$, where each example belongs to one of the two classes, represented by $y_i \in \{\div 1,-1\}$, the SVM learning method seeks the optimal hyperplane **w . x** +b = 0, as the decision surface, which separates the positive and negative examples with the largest margins. The decision function for classifying linearly separable data is:

$$f(\mathrm{X}) = sign(W.X + b) \qquad (3.1)$$

Where **w** and b are found from the training set by solving a constrained quadratic optimization problem. The final decision function is

$$f(x) = sign\left(\sum_{i=1}^{N} a_i y_i (x_i..x) + b\right) \qquad (3.2)$$

The function depends on the training examples for which a_i s is non-zero. These examples are called support vectors. Often the number of support vectors is only a small fraction of the original data set. The basic SVM formulation can be extended to the non linear case by using the nonlinear kernels that maps the input space to a high dimensional feature space. In this high dimensional feature space, linear classification can be performed. The SVM classifier has become very popular due to its high performances in practical applications such as text classification and pattern recognition.

The support vector regression differs from SVM used in classification problem by introducing an alternative loss function that is modified to include a distance measure. Moreover, the parameters that control the regression quality are the cost of error C, the width of tube ε and the mapping function ϕ. In this research work, the values for polynomial degree will be in the range of 0 to 5. In this work, best kernel to make the prediction is polynomial kernel with epsilon = 1.0E-12, parameter d=4 and parameter c=1.0.

A hybrid scheme based on coupling two base classifiers using arcing classifier adapted to data mining problem is defined in order to get better results. The main originality of proposed approach relies on associating two techniques: extracting more information bits via specific linguistic techniques, space reduction mechanisms, and moreover a arcing classifier to aggregate the best classification results.

4) Combining Classifiers

Breiman introduced Arcing ('Adaptive Resampling and Combining') as a generalization of Bagging and Boosting. In Arcing, as Breiman puts it, "modified training sets are formed by resampling from the original training set, classifiers constructed using these training sets and then combined by voting.

Arcing is a more complex procedure. Again, multiple classifiers are constructed and vote for classes. But the construction is sequential, with the construction of the $(k+1)^{st}$ classifier depending on the performance of the k previously constructed classifiers.

At the start of each construction, there is a probability distribution {p(n)} on the cases in the training set. A training set T' is constructed by sampling N times from this distribution. Then the probabilities are updated depending on how the cases in T are classified by C(x,T'). A factor β>1 is defined which depends on the misclassification rate. If the nth case in T is misclassified by C(x,T'), then put weight βp(n) on that case. Otherwise define the weight to be p(n). Now divide each weight by the sum of the weights to get the updated probabilities for the next round of sampling. After a fixed number of classifiers have been constructed, voting is done for the class.

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

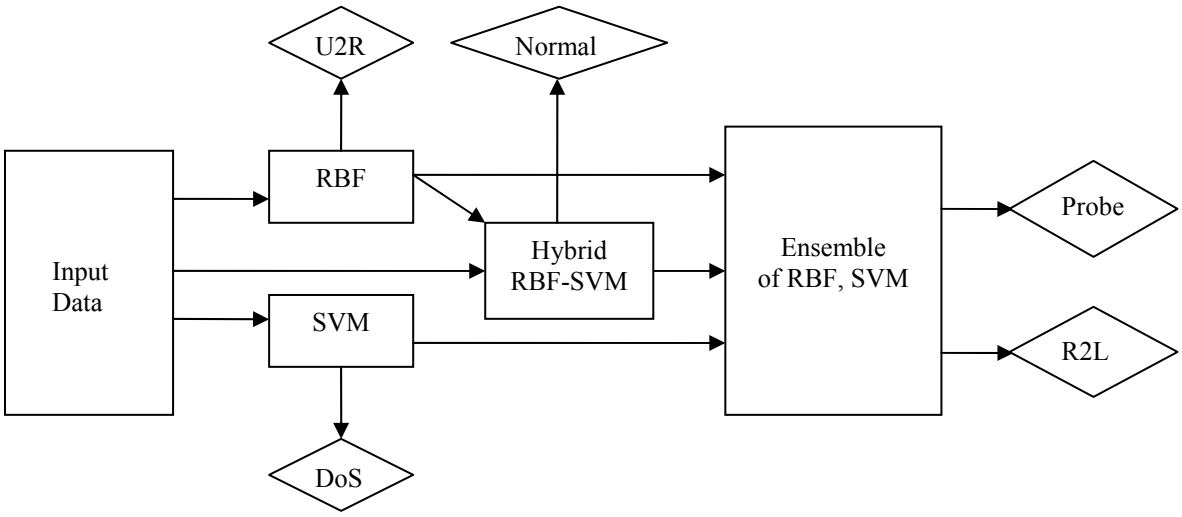

Figure 1. Architecture of hybrid intelligent intrusion detection network system.

Arcing Algorithm

1. At the *k*th step, using the current probabilities{p(n)}, sample with replacement from T to get the training set $T^{(k)}$ and construct classifier C_k using $T^{(k)}$.
2. Run T down the classifier C_k and let m(n) be the number of misclassifications of the nth case by $C_1, \ldots, C_k$.
3. The updated k+1 step probabilities are defined by

$$p(n) = (1+m(n)^4) / \Sigma (1+m(n)^4)$$

After K steps the $C_1, \ldots, C_k$ are combined by unweighted voting.

The implementation of Arcing begins with each training example having equal probability of being sampled; as a sequence of classifiers and training sets are constructed, these probabilities are increased for examples that have been misclassified. That is, the training set is resampled to obtain a training set S_1, and then repeatedly resample training sets Tt (t = 1,….T) with increasing probability for difficult examples. The contribution relies on the association of all the techniques used in proposed method. The data pre-processing allows getting more efficient and accurate computations, and then the voting system enhance the results of each classifier. The overall process comes to be very competitive.

IV. PERFORMNACE EVALUATION MEASURES

A. Cross Validation Technique

Cross-validation (Jiawei Han and Micheline Kamber**,** 2003) sometimes called rotation estimation, is a technique for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. 10-fold cross validation is commonly used. In stratified K-fold cross-validation, the folds are selected so that the mean response value is approximately equal in all the folds.

B. Criteria for Evaluation

The primary metric for evaluating classifier performance is classification Accuracy: the percentage of test samples that are correctly classified. The accuracy of a classifier refers to the ability of a given classifier to correctly predict the label of new or previously unseen data (i.e. tuples without class label information). Similarly, the accuracy of a predictor refers to how well a given predictor can guess the value of the predicted attribute for new or previously unseen data.

V. EXPERIMENTAL RESULTS AND DISCUSSION

A. Dataset Description

The data used in classification is NSL-KDD, which is a new dataset for the evaluation of researches in network intrusion detection system. NSL-KDD consists of selected records of the complete KDD'99 dataset (Ira Cohen, et al., 2007). NSL-KDD dataset solve the issues of KDD'99 benchmark [KDD'99 dataset]. Each NSL-KDD connection record contains 41 features (e.g., protocol type, service, and ag) and is labeled as either normal or an attack, with one specific attack type. The attacks fall into four classes:

- DoS e.g Neptune, Smurf, Pod and Teardrop.
- R2L: unauthorized access to local from a remote machine e.g Guess-password, Ftp-write, Imap and Phf.
- U2R: unauthorized access to root privileges e.g Bu_er-overow, Load-module, Perl and Spy.
- Probing eg. Port-sweep, IP-sweep, Nmap and Satan.

B. Experiments and Analysis

The NSL- KDD dataset are taken to evaluate the proposed RBF-SVM intrusion detection system. All experiments have been performed using Intel Core 2 Duo 2.26 GHz processor with 2 GB of RAM and weka software (Weka: Data Mining Software in java).

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)
(revised on 3 October 2012)

Table 1: The Performance of Base and Hybrid Classifiers

Dataset	Classifiers	Classification Accuracy
NSL- KDD dataset	RBF	83.57 %
	SVM	83.58 %
	RBF-SVM	85.19 %

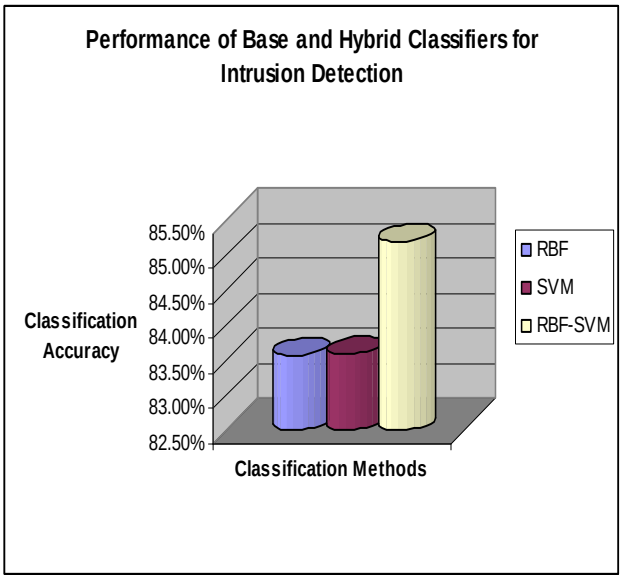

Figure 2. Classification Accuracy

The data set described in section 5 is being used to test the performance of base classifiers and hybrid classifier. Classification accuracy was evaluated using 10-fold cross validation. In the proposed approach, first the base classifiers RBF and SVM are constructed individually to obtain a very good generalization performance. Secondly, the ensemble of RBF and SVM is designed. In the ensemble approach, the final output is decided as follows: base classifier's output is given a weight (0–1 scale) depending on the generalization performance as given in Table 1. According to Table 1, the proposed hybrid model shows significantly larger improvement of classification accuracy than the base classifiers and the results are found to be statistically significant. We show that proposed ensemble of RBF and SVM is superior to individual approaches for intrusion detection problem in terms of Classification accuracy

VI. Conclusion

In this research, we have investigated some new techniques for intrusion detection and evaluated their performance based on the 42-dimensional of NSL-KDD dataset to approximately 20% of its original size and then classifying the reduced data by RBF and SVM. We have explored RBF and SVM as intrusion detection models. Next we designed a hybrid RBF-SVM model and RBF, SVM models as base classifiers. Finally, we propose a hybrid intelligent intrusion detection network system to make optimum use of the best performances delivered by the individual base classifiers and the hybrid approach. The hybrid RBF-SVM shows higher percentage of classification accuracy than the base classifiers and enhances the testing time due to data dimensions reduction.

Based on our experiment results, we have the following observations.

- SVM exhibits better performance than RBF in the important respects of accuracy.
- Comparison between the individual classifier and the combination classifier: it is clear that the combination classifiers show the significant improvement over the single classifiers.

Acknowledgment

Author gratefully acknowledges the authorities of Annamalai University for the facilities offered and encouragement to carry out this work.

References

[1] J.P. Anderson, (1980), "Computer security threat monitoring and surveillance",Technical Report, James P. Anderson Co., Fort Washington, PA.

[2] E. Biermann, E. Cloete and L.M. Venter, (2001), "A comparison of intrusion detection Systems", Computer and Security, vol. 20, pp. 676-683.

[3] Freund, Y. and Schapire, R. (1995) A decision-theoretic generalization of on-line learning and an application to boosting. In proceedings of the Second European Conference on Computational Learning Theory, pp 23-37.

[4] Freund, Y. and Schapire, R. (1996) Experiments with a new boosting algorithm, In Proceedings of the Thirteenth International Conference on Machine Learning, 148-156 Bari, Italy.

[5] Ghosh AK, Schwartzbard A. (1999), A study in using neural networks for anomaly and misuse detection. In: The proceeding on the 8th USENIX security symposium, <http://citeseer.ist.psu.edu/context/1170861/0>; [accessed August 2006].

[6] Heady R, Luger G, Maccabe A, Servilla M. (1990), The architecture of a network level intrusion detection system. Technical Report, Department of Computer Science, University of New Mexico.

[7] K. Ilgun, R.A. Kemmerer and P.A. Porras, (1995), "State transition analysis:A rule-based intrusion detection approach" IEEE Trans. Software Eng. vol. 21, pp. 181-199.

[8] Ira Cohen, Qi Tian, Xiang Sean Zhou and Thoms S.Huang, (2007), "Feature Selection Using Principal Feature Analysis", In Proceedings of the 15th international conference on Multimedia, Augsburg, Germany, September, pp. 25-29.

[9] Jiawei Han , Micheline Kamber, (2003), " Data Mining – Concepts and Techniques" Elsevier Publications.

[10] KDD'99 dataset, (2010), http://kdd.ics.uci.edu/databases, Irvine, CA, USA.

[11] Kumar S, Spafford EH. (1995), A software architecture to support misuse intrusion detection. In: Proceedings of the 18th national information security conference, p. 194–204.

[12] E. Lundin and E. Jonsson, (2002), "Anomaly-based intrusion detection: privacy concerns and other problems", Computer Networks, vol. 34, pp. 623-640.

[13] D. Marchette, (1999), "A statistical method for profiling network traffic". In proceedings of the First USENIX Workshop on Intrusion Detection and Network Monitoring (Santa Clara), CA. pp. 119-128.

[14] Mukkamala S, Sung AH, Abraham A. (2003), Intrusion detection using ensemble of soft computing paradigms, third international conference on intelligent systems design and applications, intelligent systems design and applications, advances in soft computing. Germany: Springer; p. 239–48.

[15] Mukkamala S, Sung AH, Abraham A. (2004a), Modeling intrusion detection systems using linear genetic programming approach, The 17th international conference on industrial & engineering applications of artificial intelligence and expert systems, innovations in applied artificial intelligence. In: Robert O., Chunsheng Y., Moonis A., editors. Lecture Notes in Computer Science, vol. 3029. Germany: Springer; p. 633–42.

[16] Mukkamala S, Sung AH, Abraham A, Ramos V. (2004b), Intrusion detection systems using adaptive regression splines. In: Seruca I, Filipe J, Hammoudi S, Cordeiro J, editors. Proceedings of the 6th

international conference on enterprise information systems, ICEIS'04, vol. 3, Portugal, p. 26–33 [ISBN:972-8865-00-7].
[17] S. Mukkamala, G. Janoski and A.Sung, (2002), "Intrusion detection: support vector machines and neural networks" In proceedings of the IEEE International Joint Conference on Neural Networks (ANNIE), St. Louis, MO, pp. 1702-1707.
[18] Oliver Buchtala, Manuel Klimek, and Bernhard Sick, Member, IEEE**,** (2005) **"**Evolutionary Optimization of Radial Basis Function Classifiers for Data Mining Applications", IEEE Transactions on systems, man, and cybernetics—part b: cybernetics, vol. 35, no. 5.
[19] Shah K, Dave N, Chavan S, Mukherjee S, Abraham A, Sanyal S. (2004), Adaptive neuro-fuzzy intrusion detection system. IEEE International Conference on Information Technology: Coding and Computing (ITCC'04), vol. 1. USA: IEEE Computer Society; p. 70–74.
[20] T. Shon and J. Moon, (2007), "A hybrid machine learning approach to network anomaly detection", Information Sciences, vol.177, pp. 3799-3821.
[21] Summers RC. (1997) Secure computing: threats and safeguards. New York: McGraw-Hill.
[22] Sundaram A. (1996), An introduction to intrusion detection. ACM Cross Roads; 2(4).
[23] W. Stallings, (2006), "Cryptography and network security principles and practices", USA: Prentice Hall.
[24] C. Tsai , Y. Hsu, C. Lin and W. Lin, (2009), "Intrusion detection by machine learning: A review", Expert Systems with Applications, vol. 36, pp.11994-12000.
[25] Vapnik, V. (1998). Statistical learning theory, New York, John Wiley & Sons.
[26] T. Verwoerd and R. Hunt, (2002), "Intrusion detection techniques and approaches", Computer Communications, vol. 25, pp.1356-1365.
[27] Weka: Data Mining Software in java http://www.cs.waikato.ac.nz/ml/weka/
[28] S. Wu and W. Banzhaf, (2010), "The use of computational intelligence in intrusion detection systems: A review", Applied Soft Computing, vol.10, pp. 1-35.

Date of modification: 03-10-2012
Brief description of the changes: Inclusion of Table 1 entitled The Performance of Base and Hybrid Classifiers in Section 5, the experimental results and discussion under the topic "B. Experiments and Analysis"