A comparative analysis of SVM and its stacking with other classification algorithm for intrusion detection

Conference Paper · April 2016

DOI: 10.1109/ICACCA.2016.7578859

CITATIONS
123

READS
2,624

5 authors, including:

Nanak Chand
National Institute of Technical Teachers Training and Research
3 PUBLICATIONS **163** CITATIONS
SEE PROFILE

Preeti Mishra
Malaviya National Institute of Technology Jaipur
31 PUBLICATIONS **1,469** CITATIONS
SEE PROFILE

Rama Krishna Challa
National Institute of Technical Teachers Training and Research,Chandigarh
208 PUBLICATIONS **1,835** CITATIONS
SEE PROFILE

Emmanuel S Pilli
Malaviya National Institute of Technology Jaipur
151 PUBLICATIONS **3,210** CITATIONS
SEE PROFILE

A Comparative Analysis of SVM and its Stacking with other Classification Algorithm for Intrusion Detection

Nanak Chand*, Preeti Mishra†, C. Rama Krishna*, Emmanuel Shubhakar Pilli† and Mahesh Chandra Govil†
Department of Computer Science & Engineering
*National Institute of Technical Teachers′ Training & Research
Chandigarh, India
Email: nanak.cse@nitttrchd.ac.in, rkc_97@yahoo.com
†Malaviya National Institute of Technology
Jaipur, India
Email: scholar.preeti@gmail.com, espilli.cse@mnit.ac.in, govilmc@yahoo.com

***Abstract*—Network attacks have become more pervasive in the cyber world. There are various attacks such as denial of service, scanning, privilege escalation that is increasing day by day leading towards the requirement of a more robust and adaptable security techniques. Anomaly detection is the main focus of our paper. Support Vector Machine (SVM) is one of the good classification algorithm applied specially for intrusion detection. However, its performance can be significantly improved when it is applied in integration with other classifiers. In this paper, we have performed a comparative analysis of SVM classifier's performance when it is stacked with other classifiers like BayesNet, AdaBoost, Logistic, IBK, J48, RandomForest, JRip, OneR and SimpleCart. Multi-Classifier algorithm have better classification power when compared to a single classifier algorithm specially for detecting low frequency attacks such as guess password, rootkits, spyware etc. Our preliminary analysis over NSL-KDD′99 dataset shows that stacking of SVM and Random Forest provides the best performance with accuracy of around 97.50% which apparently better than SVM (91.81%).**

Keywords*—Machine Learning, Multi-Classifier, Intrusion Detection.

I. Introduction

Internet is a globally accessed medium which provide connectivity throughout the world. However, growth in information technology also leads to comfort for attackers due to which number of cyber-attacks are increasing day by day. Examples of some recent attacks on Internet are DDoS attack faced by Rutgers University on Sept 28, 2015 which caused their website and Internet access to slow down [1], Several Thai Government websites have been hit by a suspected distributed-denial-of-service (DDos) attack making them impossible to access on Sept 30, 2015 [2]. The main threat to Internet is intrusion which means entrance to the system by force or without anyone′s permission. Some examples of attacks on internet are DOS, Probe, virus, rootkit, scanning, rootkit, cross site scripting, SQL injection and malware etc. DOS is an attack that floods the network with unnecessary traffic to restrict the legitimate users to access the network. Probing is used to monitor and collect data about network. It is a type of passive attack. In R2L (Remote-To-Local) attack, attacker tries to gain unauthorized access in victim′s machine. In U2R (User-To-Root) attack, attacker already has an account on victim′s machine but he try to get the root access. Rootkit is a malicious software that provides unauthorized access to a computer. Malware is a program code used to disrupt a system by gaining unauthorized access, propagating infected codes to the system etc.

Intrusion detection is the technique used to defend our systems from kinds of attacks. It can be further classified into two categories anomaly detection and misuse detection. Anomaly detection technique is used to identify the outliers i.e. events or observations which are not matching with other items in a dataset. In misuse detection techniques data set is compared with predefined signatures and these signatures are derived from some set of rules to avoid attacks. It is also called as signature-based technique. The main problem with misuse based IDS is that they fail to detect new attacks whose signatures are not present while the anomaly based techniques are adaptive in nature as they can identify novel attacks.

Data Mining is an analytical process designed to extract useful patterns from any dataset. These patterns can be used for building user profiles, detecting anomalous data and misuse behavior. Hence, data mining algorithms are useful in building classifiers to detect attacks. Machine learning is one of the best approaches for data mining. Machine learning algorithms are set of algorithms which can learn patterns from data and make the predictions accordingly. They are used to discover valuable knowledge from large dataset. It brings together computer science and statistics to increase the prediction power. They can be further classified into two categories supervised and unsupervised algorithms. Supervised machine learning builds classification model on the basis of already classified data called training data while in unsupervised learning algorithms data is classified into different classes with high intra-cluster similarity and inter-cluster dissimilarity. We are taking Support

Vector Machine as a basis for detailed analysis of comparative study. SVM is a set of supervised algorithms which can be used for classification, regression and outlier detection. Support vector machine is one of the most accurate and robust algorithms for classification. Nowadays, these machine learning algorithms are used very widely for IDS as they provide high security and take less time to detect attacks.

Single classifier is the classification model build on the basis of one classification algorithm. While, Multi-Classifier [3] is a type of hybrid intelligent system that combines two or more classification algorithms to produce the final model. On basis of the topology, Multi-classifier can be further divided into 2 categories Serial Multi-Classifier and Parallel Multi-Classifier. Multi-Classifiers are used to produce the best form of any classification model. Stacking is used to implement Serial Multi-Classifier. Stacking combines multiple classifiers generated by different machine learning algorithms. It is a two-phase process, which generates a set of base classifiers in the first phase and then combines these base classifiers to generate a meta-classifier in the second phase.

We are considering Support Vector Machine (SVM) as a base model for comparing in this paper, because it is one of the most efficient algorithms for data mining. Motaz et.al [5] have compared the performance of SVM with other machine learning algorithms but we are trying to integrate SVM with other classifiers and then compare them to evaluate best SVM based Multi-Classifier.

Major contribution of the paper can be summarized as follows:

- To carry out performance analysis of SVM with 9 other machine learning algorithm namely BayesNet, AdaBoost, Logistic, IBK, J48, RandomForest, JRip, OneR and SimpleCart and measure their performance in different parameters like Accuracy, Specificity, Sensitivity, Precision and Recall.
- To find the best SVM based multi-classifier algorithm outperforming SVM with critical observation of simulation.

Rest of paper is organized as follow: section 2 describes related work. In section 3, we have provided a brief description of machine learning and its algorithms. Section 4 is describing the tool and data set used for the experiments, and the results analysis of different classifiers. In last section conclusion and future work is described.

II. Related Work

In this section, we will give a brief description about the previous work related to SVM and comparative analysis of various classification algorithms. Kim et al [6] proposed Intrusion detection model based on the SVM. They have considered intrusion detection as two-class classification or multi-class classification problem. They have used KDD′99 dataset for training and testing. After preprocessing the dataset, SVM classifier model is built using KDD′99 training dataset. They have tested the SVM intrusion detection model for three different conditions such as Feature deletion, five class classification and comparison of training and testing of results to KDD′99 contest. In result set, they have successfully demonstrated that SVM Intrusion detection model works efficiently in all the three environments.

Bhavsar et al [7] effectively proposed an Intrusion Detection System using SVM with less training time. They have performed proper data preprocessing so that training time of SVM can be reduced. Experiments were performed using NSL-KDD′99 dataset. They also compared variation in performance of SVM for different kernel function. For every experiment, data was preprocessed by applying a transformation, normalization and discretization. Then, their performance was measured for three different kernel modes namely Gaussian, Polynomial and Sigmoid kernel. Evaluation result shows that Gausian kernel can reduce the learning time of SVM if the data is properly preprocessed.

Mulay et al [8] proposed an IDS using support vector machine and decision tree. They have used KDD′99 data set for training and testing. Initially KDD′99 data set was provided as an input, then SMO was applied to generate five different SVMs. After that, they have prepared a decision tree from these five trees. The trained decision tree model was used for intrusion detection. They have measured the performance of this decision tree based SVM on the basis of accuracy, training time and testing time. The proposed IDS should be faster than other algorithms.

Mortaz et.al [5] presented comparative analysis for support vector machine and other machine learning algorithms such as standard classification methods, hybrid algorithms, ensemble method and hybrid ensemble method. They used terrorist attack dataset for training and testing. After training of different models, these models were used to predict which terrorist groups are responsible for terrorist′s attacks. Initial data set values were reduced by taking the most informative attribute into consideration from terrorism data set. They used removal imputation and special coding for handling missing data elements. They have performed the experiments for 3 different conditions such as whole data set is used as test data, 66% is used for training and 34% data is used for testing the model, using 10 Fold cross validation as testing option. Evaluation results describe that SVM algorithm is best in experiment 3 and in other two SVM′s performance is average.

Himadri et al [4] gave a comparative study of classification technique for intrusion detection. They used NSL KDD data set for test data set. They ran test data for 22 classification algorithms on Weka tool. They checked performance of each algorithm on the basis of parameters namely accuracy sensitivity, specificity and time. After this, top 10 algorithms were selected on the basis of above parameters for comparison. Top ten algorithms were J48, BayesNet, Logistic, SGD, IBK, JRip, PART, Random forest, Random tree and REP Tree. The performance of each classification model is measured using 10 fold cross validation. Analysis of this paper shows that Random forest is the best technique for intrusion detection. Mukkamala et al[21] described the approaches to detect intrusions based on neural network and SVM. DARPA dataset is used for

the experiments. They have compared the performance of intrusion detection system using neural networks and support vector machine both delivers high performance of around 99%.

III. Machine Learning Approaches

Machine Learning is the process to detect the knowledgeable patterns from data and to learn from these patterns without any explicit instruction. Machine learning algorithm overlaps with the interface between statistics and computers. All the machine learning algorithms are mathematically optimized. Machine learning is also used for building classification based IDS which are more adaptive in nature. We have used some of these algorithms to give a comparison.

SVM is a supervised type of machine learning algorithms which can be used for regression, classification and outlier detection. SVM Classifier uses the concept of hyperplanes to classify data. A hyperplane is a subspace of vector space that has one less than the dimensions of vector space. An optimal hyperplane finds maximum margin between two different planes. Support vectors are most important data points in hyper planes. A hyperplane H in R^n can expressed as [10]:

$$H = \{x : a^x = b\} \tag{1}$$

where a is an element of R^n, a!=0 and b is an element of R are given.
SVM perform well with both linear and nonlinear data sets. SVM requires extensive training time [7]. BayesNet i.e Bayesian Network is a structured network graph with set of probabilities. BayesNet classifiers are used to handle missing values dataset. They predict the missing values on the basis of Bayes Theorem which can be expressed as [11]:

$$P(H|E,c) = \frac{P(H|c).P(E|H,c)}{P(E|c)}$$

where H is hypothesis, E is evidence and c is the background information. This classifier works very well with rough data sets. AdaBoost [9] is a type boosting algorithms which is used with other algorithms to improve their performance. AdaBoost is used to solve many practical problems by boosting power of any weak classifier. AdaBoost produces an effective classifier by combining rough and inaccurate rules. Logistic [12] is regression analysis technique used to estimate a relationship between the variable. It is a type of non-linear regression in which the curve is constructed using logarithmic operations. Logistic classifiers can also handle noisy data. They are efficient classifiers but require an extensive training time like SVM. IBK [13] is k-NN based classifier where k is the number of neighbors. It comes under the category of Lazy Classifiers in which the target function is approximated locally. IBK uses k-fold cross validation to provide an efficient classifier with high accuracy rate. IBK can also be used for distance weighting.

J48 [9] is a decision tree based classifiers. J48 is one of the Top 10 algorithms in data mining [9]. J48 uses information gain and default gain to rank tests. This classifier is used to generate decision tree based on the input dataset. The generated tree is used as a classifier. It has very high performance for small set of data. While for rough and large set of data their performance is not so well. Random Forest [14] develops number of trees from the training dataset. Each dataset will pass through this forest of trees for classification, and the result is calculated by averaging prediction from all the trees. It provides excellent performance in terms of accuracy, sensitivity etc. Random Forest performs exceptionally well with large data sets. Cart [15] is based on Classification And Regression Tree in which decision tree is constructed by splitting every node into two child repeatedly. The splitting process starts from root and ends at leafs covering the whole training dataset. Cart uses the Gini diversity index to split nodes.

JRip [16] and OneR [17] are association rule based classifiers. They develop the classification models on the basis of association rules mined from the data. JRip implements propositional rule learner classifier which works in two phase first grow and then prune to avoid over-fitting. OneR i.e. One Rule makes one rule for each attribute and then select most informative attribute with less error to build the classifier. It generates 1-level tree. All the attributes in one contribute independently with equal importance.

IV. Experiments And Results

We have performed these experiments on Waikato Environment for Knowledge Analysis (WEKA) [22] by using 20 percent of NSL-KDD-Test dataset [18]. In every experiment, SVM is taken as Meta classifier along with one another classification algorithm. We have selected top 9 classification algorithms of data mining to integrate with SVM. NSL-KDD dataset consists two kinds of traffic records normal and anomaly. Anomaly data keeps the records of all four kinds of data namely DOS, Probe, R2L and U2R. Multi-Classifiers are used to classify data record into two classes namely normal and anomaly. Our main goal is to find which classification algorithm will give the best performance for detecting the intrusions when it is integrated with SVM.

A. Data Set

The KDD Cup′99 data set is very large data set for intrusion detection and suffers from some problems like redundancy so we have used NSL-KDD [18] data set which has 11,850 records and 42 attributes. NSL-KDD data set is also publicly available.

B. Performance Metrics

We have compared the performance of all the classifiers on the basis of Accuracy, Sensitivity, Specificity, Precision, Recall, and ROC Curve. The value of accuracy, sensitivity and specificity, is estimated by True Positive, True Negative, False Positive and False Negative. These values are obtained from confusion matrix. Confusion matrix is a table that describes the performance of classifiers. The values of precision and recall are obtained from the result set of WEKA.

TABLE I
PERFORMANCE TABLE FOR ALL THE CLASSIFIERS

SVM With Other Classifiers	Accuracy	Sensitivity	Specificity	Precision	Recall
SVM	91.81	89.82	92.09	91.70	91.80
SVM & BayesNet	92.14	73.45	97.40	93.10	92.10
SVM & AdaBoost	90.53	80.1	92.28	90.10	90.50
SVM & Logistic Regression	92.58	89.65	93.04	92.40	92.60
SVM & IBK	95.79	89.38	97.17	95.80	95.80
SVM & J48	97.12	93.22	97.96	97.10	97.10
SVM & Random Forest	97.50	93.49	98.38	97.60	97.60
SVM & JRip	97.21	92.33	98.29	97.20	97.20
SVM & OneR	91.77	79.14	94.36	91.60	91.80
SVM & Simple Cart	97.49	93.48	98.38	97.50	97.50

Accuracy is one of the basic measures for describing the performance of any algorithm. It describes the degree to which an algorithm can correctly predict the positive and negative instances and is calculated by the formula:

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP}$$

Sensitivity measures the proportion of positives that are correctly identified by a learning method and are calculated by the formula:

$$Sensitivity = \frac{TP}{TP + FN}$$

Specificity is the measure of the proportion of negatives that are correctly identified by a learning algorithm and is calculated by the formula:

$$Specificity = \frac{TN}{TN + FP}$$

Precision is the fact of being accurate and correct. Precision gives the idea of correctly predicted instances. It is measured as proportion of true positive from all positives and is calculated as:

$$Precision = \frac{TP}{TP + FP}$$

Recall measure how much relevant data is retrieved from any machine learning algorithm. It focuses on the valuable information.

ROC Curve [19] is a graph that describes the performance of classifiers. It is plotted between False Positive (i.e. 1-specificity) and True Positive (sensitivity). In RoC curve, AuC i.e. area under the curve is directly proportional to its performance. Classifiers having more area under the curve will have high performance and vice-versa.

C. Result Analysis

We have used 10-cross validation for testing and evaluating each algorithm. In 10-cross fold, input data set is divided into 10 subsets. Each time one subset is used as test set and the other 9 subsets form the training set. Table 1 summarizes the result of all classifiers. RoC Curve is drawn by taking false positive rate value i.e. 1 - specificity on x-axis and true positive value which is also called sensitivity on y-axis. Area under the curve for RoC Curve is specified by a term called AuC as shown in Figure 1.

AuC of Multi-Classifiers of SVM with AdaBoost, Logistic, OneR and Single SVM is lesser than other classifiers which mean there performance is comparatively low as shown in Figure 1. SVM requires extensive training time and performs well with properly preprocessed dataset [7]. As NSL-KDD is a preprocessed dataset, so it gives high performance in detecting intrusions. BayesNet is based on Bayesian theorem. BayesNet is a highly scalable classifier and performs well for classifying rough dataset like medical data. For NSL-KDD which is a preprocessed data set, BayesNet is providing approximately the same results for accuracy, precision and recall. While its sensitivity i.e. prediction about normal class is only 73.45% which is very low as compared to other algorithms as shown in Table 1. RoC Curve of SVM with BayesNet has average area AuC which indicates that it has average performance for detecting intrusions as shown in Figure 1. AdaBoost i.e. Adaptive Boosting is a machine learning algorithm which is used in integration with another classifier to boost its prediction power but when integrated with SVM its performance degrades. Because SVM is a strong classifier, and as described by Li et al [20] AdaBoost with strong classifier is not viable. And the same is found to be true in our evaluation result as accuracy of AdaBoost with SVM is 90.53% while accuracy of only SVM is 91.81%. Logistic is a strong classifier but it is a weak learner i.e. it requires large data set to train classifier. Because of this stacking, both these algorithms does not give good performance. Logistic gives better performance than SVM as its accuracy, specificity, precision and recall is more than SVM. IBK is lazy trainer means it stores the training instances and do real work only at the time of classification and in this classifier the main advantage is that the learning system concurrently solves many problems. As a result combining this algorithm with SVM gives very high rate of accuracy i.e. 95.79%. Both JRip and OneR are based on association rule mining but JRip is a fast and ripper algorithm. OneR creates one rule for each attribute and then picks up rule with least error. Hence combining SVM with JRip gives high accuracy and sensitivity values i.e. 97.21% and 92.33% while with OneR, it provides accuracy and sensitivity values as 91.77% and 79.14%. As the NSL-kDD is well known data

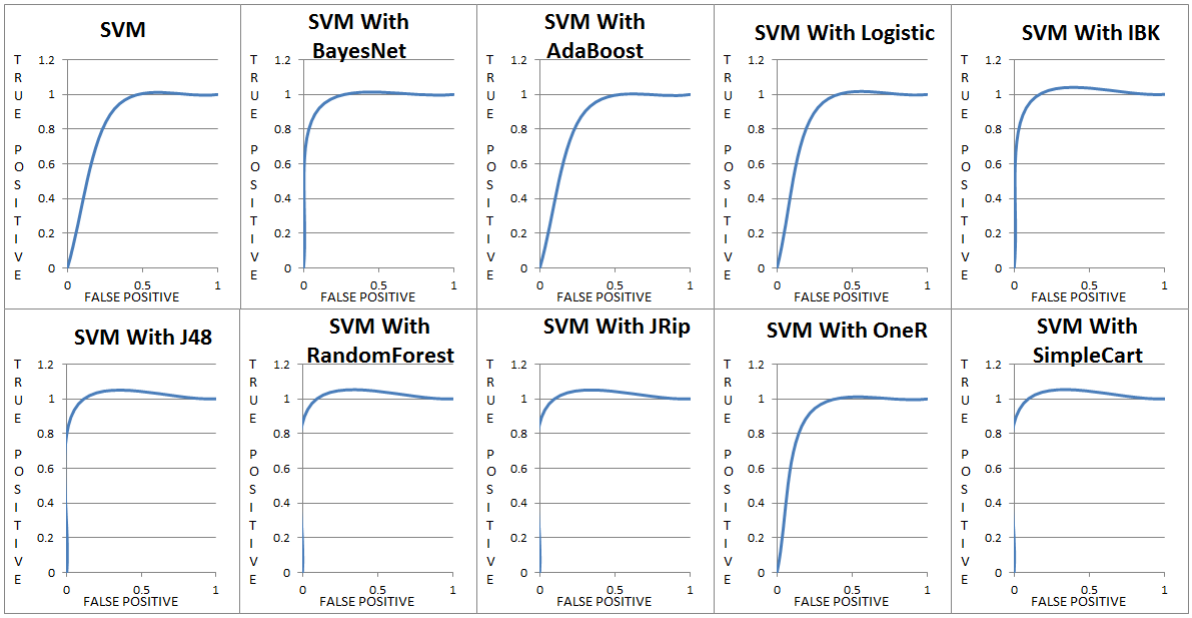

Fig. 1. ROC Curve For All The Classifiers

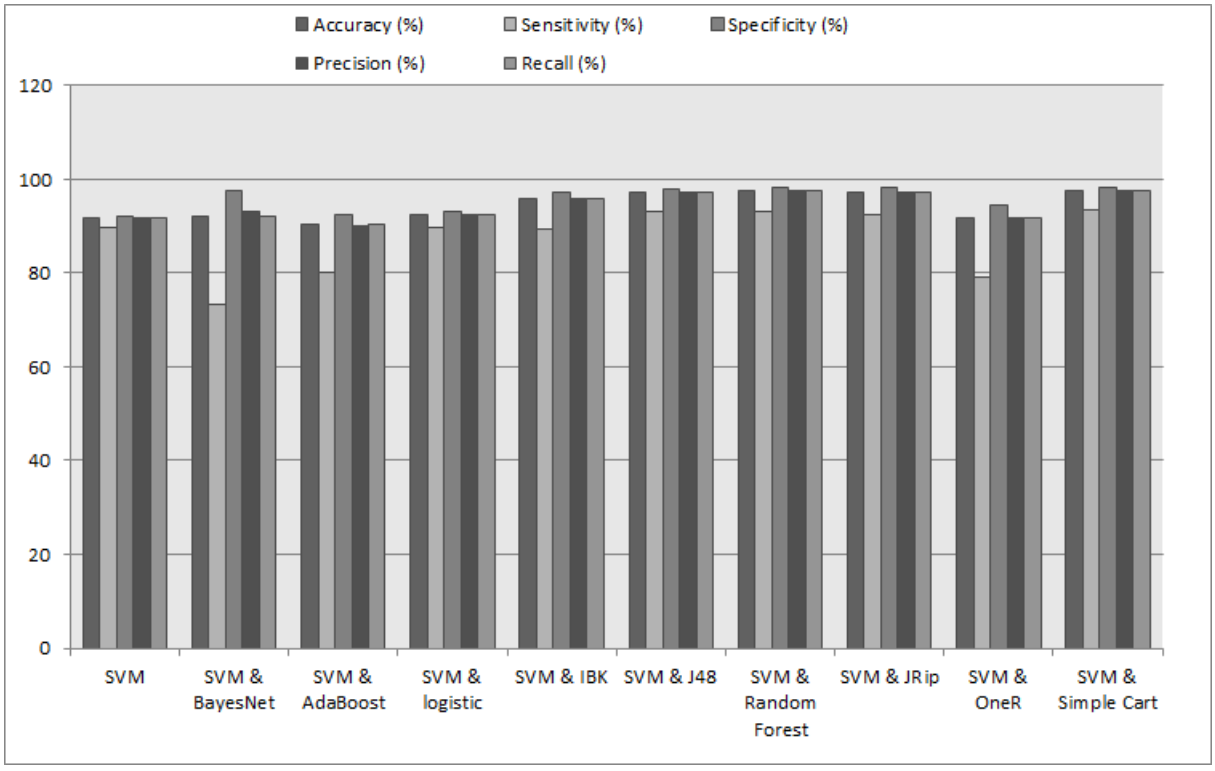

Fig. 2. Performance Graph For All The Classifiers

set without any problems like redundancy, it is very easy for a classifier to make its decision tree therefore it will increase the performance of SVM for intrusion detection. Therefore J48, Simple Cart and Random Forest are providing accuracy rate of around 97%. Out of these decision trees Random Forest is giving the best performance in all evaluating parameters. Multi-classifiers like SVM with J48, Random Forest, IBK and SimpleCart has high AuC as shown in Figure 1, so there performance is also very high. The comparative graph of all the classifiers is shown in Figure 2.

V. Conclusion

Anomaly detection using machine learning techniques is one the growing research area. The adaptability and learning power of machine learning algorithms have fascinated the eyes of researchers working in different areas. With view of network security, we intend to apply machine learning specially using SVM as base classifier for detecting intrusions in network. We have measured performance of SVM and its stacking with 9 other machine learning algorithms. We have used NSL-KDD′99 data set to analyze performance of all the classifiers. The purpose of this study is to find out the best Multi-Classifier algorithm outperforming SVM classifier. We found that not all classifiers increase the performances of SVM. Stacking of SVM and Random Forest is providing the best performance because Random Forest is a very good classifier to perform complex classification task.

In future, we plan to provide the optimization algorithm together with our Multi-Classifier algorithm (SVM & Random Forrest) to improve the performance of classification. We will also try to apply the technique in security critical applications such as Cloud Security and Forensics.

References

[1] M. Waterhouse. [2015]. *Rutgers University's computer network under DDoS attack; website, Internet access down on campus* [Online]. Available: http://abc7ny.com/technology/rutgerscomputer-network-under-attack;-website-internet-access-down-oncampus/1006255/.

[2] BBC News. [2015]. *Thai government websites hit by denial-of-service attack - BBC News* [Online]. Available: http://www.bbc.com/news/world-asia-34409343.

[3] M. Wozniak, M. Graa and E. Corchado, "A survey of multiple classifier systems as hybrid systems", *in Information Fusion*, vol. 16, pp. 3-17, 2014.

[4] M. Khorshid, T. Abou-El-Enien and G. Soliman, "A comparison among support vector machine and other machine learning classification algorithms", *in IPASJ International Journal of Computer Science*, vol. 3, no. 5, pp. 26-35, 2015.

[5] D. Kim and J. Park, "Network-Based Intrusion Detection with Support Vector Machines", *in Information Networking*, pp. 747-756, 2003.

[6] Y. Bhavsar and K. Waghmare, "Intrusion Detection System Using Data Mining Technique: Support Vector Machine", *in International Journal of Emerging Technology and Advanced Engineering*, vol. 3, no. 3, pp. 581-586, 2013.

[7] S. Mulay, P. Devale and G. Garje, "Intrusion Detection System Using Support Vector Machine and Decision Tree", *in International Journal of Computer Applications*, vol. 3, no. 3, pp. 40-43, 2010.

[8] H. Chauhan, V. Kumar, S. Pundir and E. Pilli, "A Comparative Study of Classification Techniques for Intrusion Detection", *in International Symposium on Computational and Business Intelligence*, 2013 pp. 1-6.

[9] S. Mukkamala, G. Janoski and A. Sung, "Intrusion detection using neural networks and support vector machines", *in International Joint Conference on Neural Networks. IJCNN*, 2002.

[10] K. Lipkowitz, *Reviews in computational chemistry*. Chichester: Wiley, 2007.

[11] D. Heckerman, *Data Mining and Knowledge Discovery*, vol. 1, no. 1, pp. 79-119, 1997.

[12] X. Wu, V. Kumar, J. Ross Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. McLachlan, A. Ng, B. Liu, P. Yu, Z. Zhou, M. Steinbach, D. Hand and D. Steinberg, "Top 10 algorithms in data mining", *in Knowledge and Information Systems*, vol. 14, no. 1, pp. 1-37, 2007.

[13] C. Gates, J. McNutt, J. Kadane and M. Kellner, "Scan Detection on Very Large Networks Using Logistic Regression Modeling", *in 11th IEEE Symposium on Computers and Communications (ISCC)*, 2006.

[14] S. Vijayarani and M. Muthulakshmi, "Comparative Analysis of Bayes and Lazy Classification Algorithms", *in International Journal of Advanced Research in Computer and Communication Engineering*, vol. 2, no. 8, pp. 3118-3124, 2013.

[15] A. Liaw and M. Wiener. "Classification and Regression by randomForest".*in R News* 2(3), 1822, 2002.

[16] Salford Systems. (2015). *CART Classification And Regression Trees - Data Mining and Predictive Analysis Software* [Online]. Available: http://www.salford-systems.com/products/cart.

[17] S.Sayad. *OneR*[Online]. Available: http://www.saedsayad.com/oner.htm.

[18] A. Rajput, R. Aharwal, M. Dubey, S. Saxena and M. Raghuvanshi, "J48 and JRIP Rules for E-Governance Data", *in International Journal of Computer Science and Security (IJCSS)*, vol. 5, no. 2, pp. 201 - 207, 2011.

[19] Machine Learning Group. [2015] *Weka 3 - Data Mining with Open Source Machine Learning Software in Java* [Online]. Available: http://www.cs.waikato.ac.nz/ml/weka/.

[20] KDD 99. [Online]. Available: http://kdd.ics.uci.edu/databases/kddcup1999.html/.

[21] T. Fawcett, "An introduction to ROC analysis", *in Pattern Recognition Letters*, vol. 27, no. 8, pp. 861-874, 2006.

[22] X. Li, L. Wang and E. Sung, "AdaBoost with SVM-based component classifiers",*in Engineering Applications of Artificial Intelligence*, vol. 21, no. 5, pp. 785-795, 2008.