© 2016 IJSRSET | Volume 2 | Issue 6 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

Ant Colony Optimization for Intrusion Detection System Based on KNN and KNN-DS with detection of U2R, R2L attack for Network Probe Attack Detection

Akarshika Rawat, Prof. Ankita Choubey

Department of Computer Science, Shri Ram Institute of Science and Technology, Jabalpur, Madhya Pradesh, India

ABSTRACT

The k-nearest neighbor (k-NN) is one of the most popular algorithms used for classification in various fields of pattern recognition & data mining problems. In k-nearest neighbor classification, the result of a new instance query is classified based on the majority of k-nearest neighbours. Recently researchers have begun paying attention to combining a set of individual k-NN classifiers, each using a different subset of features, with the hope of improving the overall classification accuracy. This paper describes a hybrid design for intrusion detection that combines anomaly detection with misuse detection. The proposed method includes an ensemble feature selecting classifier and a data mining classifier. The former consists of four classifiers using different sets of features and each of them employs a machine learning algorithm named fuzzy belief k-NN classification algorithm. The latter applies data mining technique to automatically extract computer users' normal behavior from training network traffic data. The outputs of ensemble feature selecting classifier and data mining classifier are then fused together to get the final decision. The experimental results indicate that hybrid approach effectively generates a more accurate intrusion detection model on detecting both normal usages and malicious activities.

Keywords : Intrusion Detection; Machine Learning; Data Mining, KNN

I. INTRODUCTION

With the rapid progression of computer technology, computer violations are increasing at a fast pace. Such malevolent activities become more and more sophisticated and can easily cause millions of dollar in damage to an organization. Detecting those intrusions becomes an important issue of computer security. Generally, there exist two main intrusion detection techniques: anomaly detection and misuse detection. Misuse detection involves the comparison of observed traffic data with a set of well-defined rules that describe signatures of intrusions. If the signature of observed network traffic is not matched with any of predefined rules, it is declared as an attack. This approach can detect the recognized attacks in an efficient way with high level of accuracy. However, it suffers from its inability of identifying attacks which differ from those predefined patterns. A minor variation of an attack may not be identified during the whole detection procedure. Anomaly detection searches for intrusive activities by comparing network traffic to those established acceptable normal usage patterns learned from training data. If the pattern of observed data is different from those learned normal ones, the data is classified as an attack. This approach can successfully detect novel and unseen malicious occurrences from computer users. However, it suffers from a high volume of false alarms. During the past years, a variety of approaches have been proposed by using either anomaly or misuse detection techniques. However, most of approaches only focus on reducing *false positive rate* (*FPR*) [1], [2] or improving *detection rate* (*DR*) [3], [4] in known and unknown intrusions. Therefore, in this research we propose an approach that combines the merits of anomaly detection and misuse detection techniques. Our goal is not only to achieve high *DR* on malicious activities but also to reduce *FPR* on normal computer usages from network traffic. We develop a hybrid model that includes an ensemble feature selecting classifier with four base

classifiers. Each base classifier uses a subset of features to derive independent decision about a network traffic, then all the decisions from multiple ones are combined into a fused result to obtain a better *DR*. In the kernel of each base feature selection classifier, a developed machine learning algorithm, fuzzy belief *k*-NN classification algorithm, is used to detect both known and novel attacks. We focus on detecting attacks that attackers use illegal approach to gain access to the target host and thus further to exploit the system's vulnerabilities. Additionally, data mining technique is applied to automatically extract decision rules for normal behavior in order to construct a filter to reduce the *FPR*. We afterward combine the output of this classifier with the result from ensemble feature selecting classifier to derive an output, which is the final decision of the input network traffic. In favor of verifying our proposed model is successful, the intrusion detection benchmark data set *DARPA KDD99* [5] is used.

II. METHODS AND MATERIAL

A. Hybrid Intrusion Detection Model

Ensemble is to combine the outputs of a set of base classifiers together in a proper way when classifying input data. The fused result is expected to perform a better outcome than that of any individual base classifier within the ensemble. In the past, approaches to intrusion detection based on ensemble techniques have been investigated with the use of different feature subsets [6] or soft computing techniques [7] in every individual classifier. However, they only focused on improving the detection rate in known and unknown intrusions but did not consider reducing the number of false alarms. Therefore, we propose a hybrid intrusion detection model that includes an ensemble feature selecting classifier and a data mining classifier to act as anomaly detection and misuse detection to improve both *DR* and *FPR*, respectively. The former consists of a set of base feature selecting classifiers and each uses partial feature space. The latter applies data mining technique to look for patterns of normal activities. We believe that the overall performance of this hybrid architecture is better than that of each individual classifier. Also, *DR* and *FPR* are more accurate than those of using full feature set. Fig. 1 depicts our design.

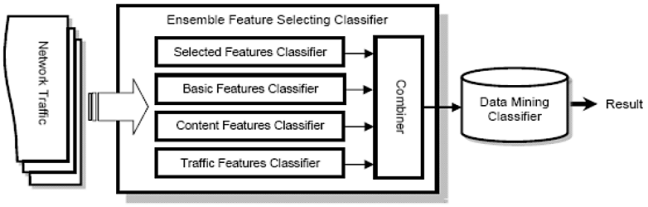

Figure 1. Hybrid Intrusion Detection Model

B. Proposed Work

Author of [7] proposed a security solution based on KNN method. For better evaluation of unknown attacks and authors method our proposed method uses the same concept incremented one step further. In proposed work detection of suspicious traffic using clustering well be tested integrating the SVM filter on them.

Following attractive points is interesting in proposed method

1. Classifying network traffic using SVM (support Vector Machine)
2. Then applying, Clustering based detection and prevention of intrusion on real time traffic instead of KNN.
3. Visualizing the network traffic.

Clustering Method:

As mentioned in the article [8], The proposed clustering based detection will be work as follow

Suppose there are N records in the flow file, which has to be converted into a group of K clusters. Medoid is a record in the traffic file which is intended to best represent the cluster.

1. Select initial Medoids: Randomly select few records from traffic file as initial K medoids.
2. Form clusters: Assign each record from traffic file to the nearest medoid.
1. Select new Medoids: Calculate the sum of distance from all records to their medoids.
3. Optimize each cluster: For each cluster, select an object as a new medoid such that the sum of distances of all objects to their medoids is reduced.
4. Form new clusters: Assign each object from traffic file to the nearest medoid.
5. Repeat step 4 and 5 arbitrary number of times (or) if the new medoids obtained in step 4 are same as in the previous iteration.

The intrusion detection systems (IDSs) are usually classified into two main categories: signature based and anomaly based [1]. The signature based systems are designed to identify attacks that follow patterns previously recognized and reported by security experts, where each signature identifies a specific attack. In anomaly based IDSs, the normal behavior of the system or network traffic are represented and, for any behavior that varies over a pre-defined threshold, an anomaly activity is identified.

A weakness of signature based intrusion detection systems is the incapability of identifying new types of attacks or variations of known attacks. By the other side, in anomaly based IDSs, the number of false positives generated is higher than on those based on signatures. An important issue in anomaly based IDSs is how these systems should be trained, i.e., how to define what is a normal behavior of a system or network environment (which features are relevant) and how to represent this behavior computationally. The training process requires a large amount of data and many artificial intelligence techniques can be employed, such as ANNs (Artificial Neural Networks). Artificial intelligence techniques have been used for both signature based and anomaly based intrusion detection. Among these techniques, we can cite the use of expert systems [2]. These systems employ a set of rules that represent patterns of known attacks or vulnerabilities to detect intrusions. Some data mining techniques have been used to identify normal patterns of behavior [3], [4]. Artificial neural networks had already been applied in IDSs [5], [6]. Most of these neural networks are composed of a set of input, some intermediate layers and one output. These networks have the capacity to identify patterns and variations of these patterns (variations of the same attack). The main contribution of the author [7] is an intelligent intrusion detection system (IIDS) which uses two artificial intelligence techniques in sequence to better identify anomalies and to reduce the false positive rate present in related works. Another contribution is the analysis of the importance of each feature for each class of attack (DoS, Probe, R2L and U2R) present in the KDD Cup 1999 Data (The Third International Knowledge Discovery and Data Mining Tools Competition) in order to define which of these features are relevant to each class of attack. Figure 2. Shows the Authors proposed architecture.

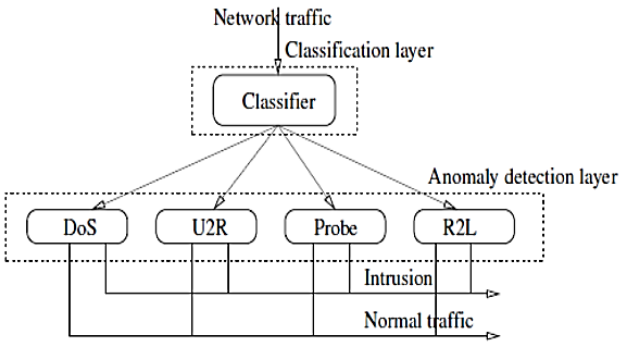

Figure 2. Authors [7] Proposed Octopus IDS Architecture

In this synopsis, the main contribution is to develop an intelligent intrusion detection system (IIDS) using machine learning techniques (KNN) in sequence to better identify anoma-lies and to reduce the false positive rate present in related works. Another contribution is the analysis of the importance of each feature for each class of attack (DoS, Probe, R2L and U2R) present in the KDD Cup 1999 Data (The Third International Knowledge Discovery and Data Mining Tools Competition) in order to define which of these features are relevant to each class of attack.
For better approximation and evaluation of proposed solution we will also applied real network traffic to better evaluate the proposed system

III. RESULTS AND DISCUSSION

A Result of different attack encountered in network in classified form

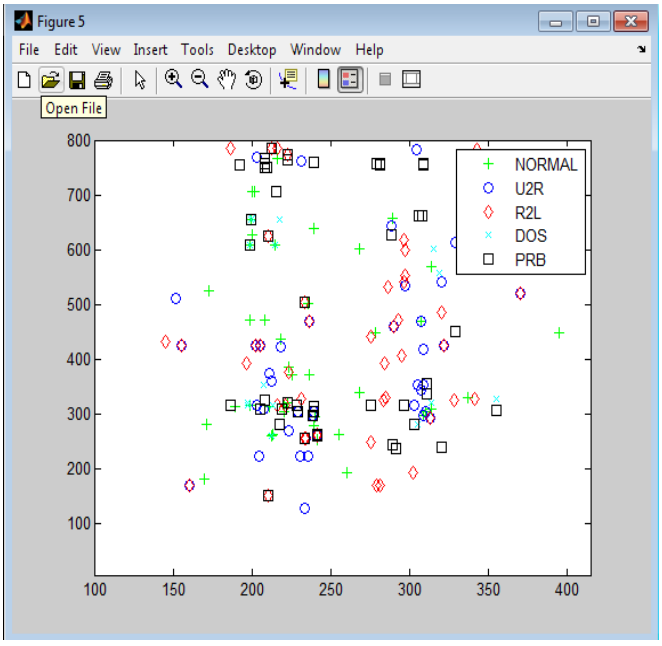

Figure 3

IV. CONCLUSION

In this paper, we develop a hybrid intrusion detection model. The model consists of a set of base feature selecting classifiers and each uses partial original feature space as well as a data mining classifier. The basic idea is using ensemble feature selection technique to promote the detection rate and data mining technique to reduce the number of false alarms. It is a combination of both anomaly detection and misuse detection. The experimental results shows the hybrid model has a better detection performance with both low FPR on normal computer usages and high DR on malicious activities than those classifiers using full feature set. It also shows the overall performance of this Hybrid architecture is better than that of each individual base feature selecting classifier. This result demonstrates that our approach is effective, which generates a more accurate intrusion detection model by combining diverse base classifiers with different feature subsets.

V. REFERENCES

[1] J. Allen, A. Christie, W. Fithian, J. McHugh, and J. Pickel,"State of the practice of intrusion detection technologies," in CMU/SEI-99-TR-028, 2014.

[2] T. Lunt, "Detecting intruders in computer systems," in Conference on Auditing and Computer Technology, 2013.

[3] S. M. Bridges and R. B. Vaughn, "Fuzzy data mining and genetic algorithms applied to intrusion detection," in National Information Systems Security Conference, October 2014.

[4] R. Mukkamala, J. Gagnon, and S. Jajodia, "Integrating data mining techniques with intrusion detection methods," in Advances in Database and Information Systems Security, 2014.

[5] G. Giacinto, F. Roli, and L. Didaci, Fusion of multiple classifiers for intrusion detection in computer networks, 2013.

[6] H. D. Lee, "Training a neural-network based intrusion detector to recognize novel attacks, systems, man and cybernetics," in IEEE Transactions on Computer, 2012, pp. 294–299.

[7] Paulo M. Mafra, Vinicius Moll, Joni da Silva Fraga and Altair Olivo Santin "Octopus-IIDS: An Anomaly Based Intelligent Intrusion Detection System", IEEE, 2010.

[8] Raghavan Muthuregunathan, Siddharth S, Srivathsan R and Rajesh SR "Efficient Snort Rule Generation using Evolutionary computing for Network Intrusion Detection", IEEE, First International Conference on Computational Intelligence, Communication Systems and Networks, 2009.

[9] L. Xu, A. Krzyzak and C.Y. Suen, "Several Methods for Combining Multiple Classifiers and Their Applications in Handwritten Character Recognition," IEEE Transactions on System, Man and Cybernetics, SMC-22(3), pp. 418-435, 1992.

[10] S. Raudys and F. Roli, "The Behavior Knowledge Space Fusion Method: Analysis of Generalization Error and Strategies for Performance Improvement," Proceedings of International Workshop on Multiple Classifier Systems, pp. 55-64, Guildford, Surrey, 2003.

[11] L. K. Hansen and P. Salamon, "Neural Network Ensembles," IEEE Transactions on Pattern Analysis Machine Intelligence, 12(10), pp. 993-1001, 1990.

[12] S. A. Dudani, "The Distance-Weighted k-NN Rule," IEEETransactions on Systems, Man and Cybernetics, vol. 6, no. 4, pp. 325-327, 1976.

[13] J. R. Quinlan, C4.5: Programs for Machine Learning, MorganKaufmann, 1993.

[14] K. Jones and R. S. Sielken, Computer System Intrusion Detection: ASurvey, Technical Report, Computer University of Virginia, 2000.

[15] A. N. Toosi and M. Kahani, "A Novel Soft Computing Model UsingAdaptive Neuro-Fuzzy Inference System for Intrusion Detection,"Networking, Sensing and Control, 2007 IEEE International Conference, pp. 834-839, London, UK, April 2007.

[16] D. Song, M. I. Heywood, and A. N. Zincir-Heywood, "Training Genetic Programming on Half a Million Patterns: An Example from Anomaly Detection," IEEE Transactions on Evolutionary Computation, 9(3), pp. 225-240, 2005.

[17] M. Keller, M. R. Gray, and J. A. Givens Jr., "A Fuzzy k-Nearest Neighbor Algorithms," Transactions on Systems, Man and Cybernetics, vol. SMC-15(4), pp. 580-585, 1985.

[18] T. Denoeux, "A k-Nearest Neighbor Classification Rule Based on Dempster-Shafer Theory," IEEE Transactions on System.