International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 4 Issue: 7 152 - 155

Intrusion Detection System by using SVM and Ant Colony

Ms. Priti V. Thorat
Dept. of Computer Engineering
SRES College of Engineering,
Kopargaon- 423601,(M.S.), India
e-mail: pritithorat55@rediffmail.com

Prof. P. N. Kalavadekar
Dept. of Computer Engineering
SRES College of Engineering,
Kopargaon- 423601,(M.S.), India
e-mail: kprak3004@gmail.com

***Abstract*—** In today's world, it is very difficult to protect our system from different types of attacks or viruses . Intrusion Detection System (IDS) are basically designed to defend computer system from various types of attacks. In this paper a new approach is proposed for network intrusion detection. This new approach will make the combination of SVM algorithm with Self-Organized Ant Colony Network (CSOACN) algorithm to develop new IDS. It will take advantages of both while avoiding their limitations. The basic task of this approach is to classify network packet as normal or abnormal while minimizing misclassification. For experimental setup we had used standard benchmark NSLKDD dataset which is advanced version of KDDCUP99 dataset. The system shows best result as we compare it with different types of IDS which was developed by using different algorithm.

Keywords- *— Network security, network attack, Intrusion Detection Systems (IDS), Support Vector Machine (SVM), Clustering based on Self-Organized Ant Colony Network (CSOACN).*

*****

I. INTRODUCTION

Intrusion detection is the detection of actions that attempt to compromise the integrity, confidentiality or availability of a resource in the network. In case of an intrusion, an IDS (Intrusion Detection System) detects it as soon as possible and takes appropriate action on it. The combination of facts, such as the rapid growth of the Internet, the vast financial possibilities opening up in electronic trade, and the lack of truly secure systems, makes IDSs an important front edge research orientation of network security.

Although many different IDSs have been developed, their detection types generally fall into one of two categories: anomaly detection or misuse detection. In Anomaly detection, it looks for behavior that deviates from normal use of system whereas in misuse detection, it looks for behavior that matches a known attack scenario. The normal behavior is based on lots of factors and highly variable. Therefore, the selection of features is the main issue in anomaly detection. The approach of misuse detection is to model abnormal system behaviors at first and define any other behaviors as normal behavior. Mainly, known intrusion attacks are represented in the form of pattern or signature. In network intrusion detection, packets that match those attack scenarios can be detected and different suitable actions will be taken for different types of intrusions. There were a great deal of time and effect has been invested in IDSs. Various approaches have been used to classify normal or attacks behavior. For example, statistical approaches, predictive pattern generation, neural networks, expert systems, keystroke monitoring, model based intrusion detection, NSM (Network Security Monitor), autonomous agents, fuzzy logic network and data mining. However, it is observed that neither the anomaly detection nor the misuse detection can detect all kinds of intrusion attempts on their own. So to provide advanced level of security an hybrid of anomaly and misuse detection has been developed.

SVM (Support Vector Machine) and CSOACN (Clustering based on Self-Organized Ant Colony Network) are approaches in data mining to generate classifiers. Both methods have been applied in intrusion detection to separate the normal and abnormal network connecting records. From the machine learning point of view, the process of SVM comes under supervised learning whereas the CSOACN algorithm is the unsupervised learning. In this paper, we present a framework that aims to combine the two approaches in intrusion detection and therefore, to reach a better system performance.

II. SUPPORT VECTOR MACHINE AND ANT COLONY

A. Support Vector Machine

SVM is supervised learning algorithm which is used for pattern recognition and regression. Every network connecting record is described by several features of the connection and a record is represented as a data point. In the case of support vector machines, the data points are written as high dimensional vectors. Using some training vectors (collected

history data), the SVMs separate the data points by generating a hyperplane between the normal and abnormal data points. Since there might have more than one hyperplanes separate the two classes of data, the question for SVMs is how to select the one that have the maximum margin classifier. Therefore, SVMs are actually a quadratic optimization problem. The traditional SVM algorithm is operated over the entire training data set. The number of training data points determines the dimension of the matrix for computing the kernel function, which influences the time of solving the QP problems. However, SVMs have the property that the points that do not lie on the margin are not necessary to be involved in the computation. Same decision function is found if some of the training data points, excepting the support vectors, are removed. Hence, for SVM, the number of training data points can be reduced without losing accuracy. In order to reduce the number of training data, an active learning into SVM is introduced. Initially, a SVM classifier was trained by using only small amount of data points from the whole training data set. The SVM classifier was then gradually modified by adding new data points for SVM training. After each training process, the output classifier is used to separate the entire data.

The recurrence of training a new SVM classifier can stop when a required correct classification rate is obtained.

B. *Ant Colony*

In the real world of self-organized ant colony network, a population of ant-like agents move objects on the 2-D grid to cluster similar objects into same regions. Object and ant are the two basic entries in the program. As an object is described by several of its features, each object can be denoted by an vector O_i and each feature of the object can be denoted by v_{ij}. All objects on the ant colony network for clustering thus can be denoted as d dimensional vectors as follows:

$$\{O_1, O_2, O_3, \ldots\ldots\ldots\ldots, O_N\}$$

$$O_i = \{V_{i1}, V_{i2}, V_{i3}, \ldots\ldots\ldots, V_{id}\}$$

where, N is the number of objects and d is the dimension of features. Network connecting records described by several features can be viewed as objects in CSOACN. These objects belong to different classes (i.e., normal and different kinds of intrusions). As the profiles of both abnormal and normal data are defined as different clusters, intrusion detection classifiers with both anomaly and misuse detection pattern can be constructed by applying clustering.

As an ant colony network possesses properties such as flexibility, robustness, decentralization and self organization, it can suggest very interesting heuristics. Optimization and control algorithms based on swarm intelligence, including Ant Colony Optimization and Ant Colony Routing, are well known. In the area of data mining, particularly for clustering purposes, there are also many studies using the metaphor of ant colonies.

III. SYSTEM ARCHITECTURE

The architecture of the proposed system is as shown in Fig. 1.

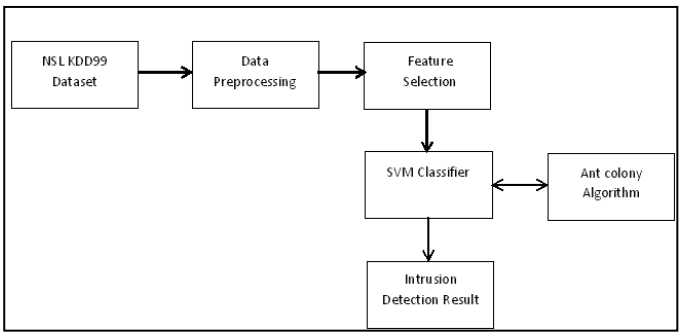

Figure 1. System Architecture

The system consists of four functional components:

1. Data Preprocessing
2. Feature Selection
3. SVM Classifier
4. Ant colony Algorithm

We will discuss each of the module in detailed:

A. *Data Preprocessing and Feature Selection*

In data preprocessing essential features are get extracted from the dataset. As NSL KDD consists of 41 different features with respect to packet but from them only some features will get select. Those are as follows:

1. Attribute
2. Duration
3. Protocol
4. Flag
5. Service
6. Source byte
7. Destination byte
8. Class

First, a randomly generated population of potential solutions is created. Then crossover, mutation and selection are applied to each generation until an acceptable solution is found or some time limit is exceeded. Crossover is where two individuals swap sequences of bits to form two new individuals. Crossover takes two rules and creates new rules by swapping the bits of the old rules. Mutation is where random bits in an individual, or possible solution, are randomly changed. The fitness of an individual is specified by the fitness function, which determines the quality of a particular individual.

B. *SVM Classifier*

Support Vector Machine is generally used for the classification of the given objects. These packets then

 ISSN: 2321-8169
152 - 155

distinguished so that they can be used for clearly distinguishing the packets about their nature whether they are anomalous in nature or perfectly fines so for that purpose the classification is done by the SVM.

C. Ant Colony Algorithm

This algorithm is mainly used for enhancing throughput of the given system. The ant colony clustering takes the input from the SVM which has classified the given packets in to different category based on the nature of packets. Then these packets will be given to the ant colony algorithm where they get matched with some predefined format. Both algorithms will used in following way [1]:

Input: Training set with each data point labeled as positive or negative (class labels).

Output: A classifier.

i. *Begin*
ii. *Randomly select data points from each class.*
iii. *Generate a SVM classifier.*
iv. *While more points to add to training set do*
v. *Find support vectors among the selected points;*
vi. *Apply CSOACN clustering around the support vectors;*
vii. *Add the points in the clusters to the training set;*
viii. *Retrain the SVM classifier using the updated training set;*
ix. *End*
x. *End*

IV. EXPERIMENTAL RESULT ANALYSIS

Dataset used

To evaluate the effectiveness of proposed system, KDDCUP99 and NSLKDD99 dataset is used as standard dataset. The details of this dataset are given in following section.

KDDCUP99

For the implementation of this approach, the system has used the KDD 99 datasets which are based on the 1998 DARPA dataset. It has 41 features for each packet or network connection. It has dataset for four major types of attack i.e. User to Root, Remote to Local, Denial of Service and Probe. It consists of different components. We have used "kddcup.data_20_percent" as training. In this case the training set consists of 494,021 records among which 97,280 are normal connection records, while the test set contains 311,029 records among which 60,593 are normal connection records.

NSL KDD99

The NSL-KDD data set is a refined version of KDD"99 data set. It contains all essential records of the KDDCup data set. Redundant records are removed so that classifier does not produce un-biased result. It has sufficient number of records of train as well as test data sets. The number of selected records from each difficult level group is inversely proportional to the percentage of records in the original KDD data set [12]. Each record has 41 features and a label assigned to each either as an attack type or as normal. NSLKDD is used by different data mining based machine learning algorithms like Support Vector Machine (SVM), Decision Tree, K-nearest neighbor, K-Means and Fuzzy C-Mean clustering algorithms.

Initially we loaded train dataset and among the 41 parameters we had extracted 6 parameters to form the rules. The features are extracted by using genetic algorithm. The result we got after the final generation that was used as input for SVM algorithm. SVM algorithm in combine with ant colony algorithm will form the two cluster i.e. one has the rules for normal connection and the second has rules for abnormal connection. In this way the classifier gets trained.

We tested our system in two phase.1) testing on test dataset 2) testing on remote connection. In both of phases we achieve good result. The result of first phase testing is shown in fig.2.

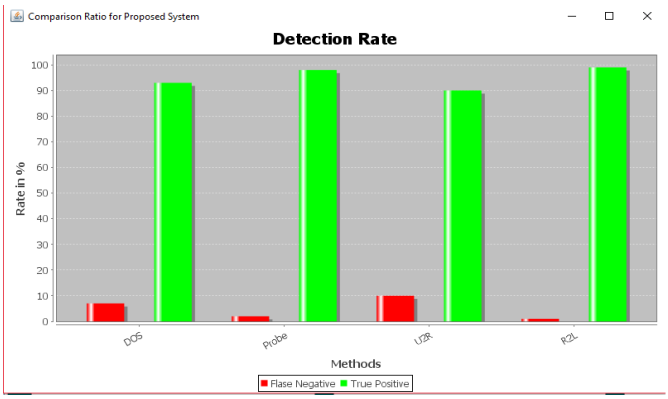

Figure 2. Result of phase1 testing

The result first phase was compared with existing system which uses Fuzzy Genetic Algorithm. From the experimental results, it is observed that our system achieve good detection rate, low false negative rate and high true positive rate. It is shown in Fig.3.

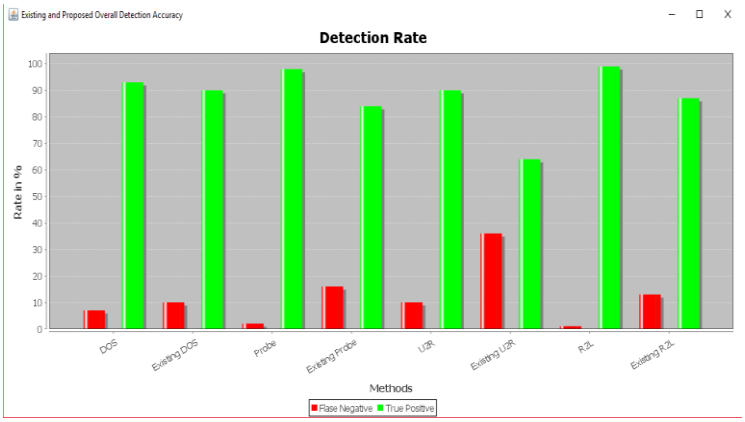

Figure 3. Comparison of proposed and existing system

In second phase of testing, we done the classification on live packet which is received through remote connection. In this case our classifier also gives good result with respect to detection rate of various attacks. The result of second phase testing is shown in Fig.4.

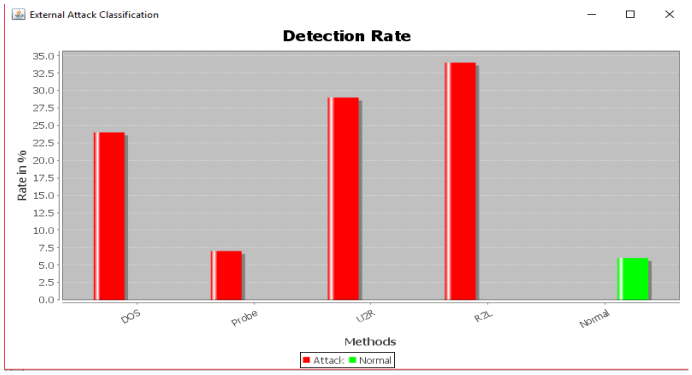

Figure 4. Result of online classification

V. CONCLUSION

We had implemented hybrid intrusion detection system which was designed by combining two algorithm i.e. SVM and Ant colony. This approach gives best result as compare to other approaches. We had done the combination supervised learning algorithm i.e. SVM with unsupervised learning algorithm i.e. Ant colony. We get good detection rate and high true positive rate n low false negative rate. The algorithm may be enhanced in some aspects. For example, the training and testing speeds may be improved by applying the dimension reduction on the input data.

ACKNOWLEDGMENT

I take this opportunity to thank my project guide and PG Coordinator Prof. P. N. Kalavadekar and Head of the Department Prof. D. B. Kshirsagar for their valuable guidance and for providing all the necessary facilities. I also thankful to all the staff members of the Computer Engineering of S.R.E.S's College of Engineering, Kopargaon. I would also like to thank the Institute for providing the required facilities, Internet access, e-resources and important books. I would like to thank my parents and my friends who have constantly bolstered my confidence and without whose moral support and encouragement, this work would have been impossible.

REFERENCES

[1] Wenying Feng, Qinglei Zhang, Gongzhu Hu,Jimmy Xiangji Huang, "Mining network data for intrusion detection through combining SVMs with ant colony networks," Future Generation Computer Systems, pp.127-140,2014.

[2] M. Dave, "Intrusion Detection System Using Genetic Algorithm," Journal Of Information, Knowledge And Research In Computer Engineering, Vol.02, Issue 02,Oct 2013.

[3] Mostaque Hassan, "Network Intrusion Detection System with Genetic Algorithms and Fuzzy Logic," International Journal of Innovative Research in Computer and Communication Engineering, Vol. 1, Issue 7, September 2013.

[4] Rupesh B. Jadhav, Mr. Balasaheb B. Gite, "Real Time Intrusion Detection With Fuzzy, Genetic and Apriori Algorithm,"International Journal of Advance Foundation and Research in Computer (IJAFRC) Volume 1, Issue 11, November 2014.

[5] S. Selvakani and R.S. Rajesh, "Genetic Algorithm for Framing Rules for Intrusion Detection," International Journal of Computer Science and Network Security, Vol. 7 No.11, November 2007.

[6] Jungwon Kim, King"s Coll., Bentley, P.J.,"Towards an artificial immune system for network intrusion detection: an investigation of dynamic clonal selection Evolutionary Computation,"CEC '02, 2002.

[7] Namita Shrivastava,Vineet Richariya, "Ant Colony Optimization with Classification Algorithms used for Intrusion Detection," International Journal of Computational Engineering & Management, Vol. 15 Issue 1, January 2012.

[8] Chuan Cai,Liang Yuan, "Intrusion Detection System based on Ant Colony System,"Journal Of Networks, Vol. 8, No. 4, April 2013.

[9] Yogita B. Bhavsar, Kalyani C.Waghmare, "Intrusion Detection System Using Data Mining Technique: Support Vector Machine," International Journal of Emerging Technology and Advanced Engineering, Volume 3, Issue 3, March 2013.

[10] R. Shanmugavadivu, Dr.N.Nagarajan,"Network Intrusion Detection System Using Fuzzy Logic," Indian Journal of Computer Science and Engineering, Vol. 2 No. 1,2007.

[11] UCI KDD Archive, KDD Cup 1999 data, 1999. http://kdd.ics.uci.edu/databases/ kddcup99/kddcup99.html.

[12] http://nsl.cs.unb.ca/NSL-KDD/