Building of Intrusion Detection System by using SVM and Ant colony Algorithm

Priti Thorat, Prof. P. N. Kalawadekar

[1] PG Student, Computer Engineering, SRES' COE Kopargaon, Maharashtra, India
[2] Associate Professor, Computer Engineering, SRES' COE Kopargaon, Maharashtra, India

ABSTRACT

Nowadays large scale data clustering and its classification have become challenging area. For that purpose, Intrusion Detection System (IDS) are designed to defend computer system. In this paper a new approach is applied for network intrusion detection. This new approach will make the combination of SVM algorithm with Self-Organized Ant Colony Network (CSOACN) algorithm to 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. In this work we are going to use standard benchmark NSLKDD dataset which is advanced version of KDDCUP99 dataset.

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

1. INTRODUCTION

INTERNET is the source of open and trusted communications. This openness at the same time results in harmful actions such as financial losses, damage to confidential information, maintaining availability of services. Intrusion Detection System has become the basic need for the successful transmission of contents on network. IDS provide two basic features: Intrusion Detection System is the combination of two features i.e. Visibility and Control. This makes possible to enforce a security policy to make computer network secure. Visibility is the feature that allows observing the network traffic and Control is the feature that controls network traffic. The detection scheme of IDS is classified into two categories: anomaly detection and misuse detection. In Anomaly detection the behavior is get consider that change from normal behavior. In misuse detection the behavior is get consider that matches a known attack condition. The normal behavior is based on lots of inoffensive factors and highly variable. Therefore, the main issue in anomaly detection is of selection of features to monitor. The way of misuse detection is to model abnormal system behaviors at first and define other behaviors as normal behavior. The intrusion attacks are represented in the form of patterns, activities that match those attack scenarios can be detected. The main issue in misuse detection systems is the pattern identification and signature depiction of the attack, which should involve all possible variations but should not match normal activities. There are various approaches to identify normal or attack behavior of the system. For example, statistical approaches, fuzzy logic, data mining, etc. However, it is observed that neither anomaly nor misuse detection are able to detect all types of attacks by their own. Therefore it is needed to go beyond of it. It can be done with the help of hybrid approach of both the detection models. By using SVM (Support Vector Machine) and CSOACN (Clustering based on Self-Organized Ant Colony Network) the classifiers are get generated for data mining. Both of these methods have been applied in intrusion detection to classify the normal and abnormal (attack) connecting record. The SVM algorithm is the type of supervising learning whereas the CSOACN algorithm is the type of unsupervised learning. This approach has significant advantages in terms of overhead, scalability and flexibility.

1.1 Problem Statement

This system presents a framework for intrusion detection by combining two existing machine learning methods (i.e. SVM and CSOACN). It is an anomaly based intrusion detection system. The IDS based on the new algorithm can be applied as pure SVM, pure CSOACN or their combination by constructing the detection classifier under three different training modes respectively. The basic task is to classify network activities (in the network log) as normal or abnormal while minimizing misclassification. This new approach combines the SVM method with CSOACNs to take the advantages of both while avoiding their weaknesses.

1.2 Objective

General objectives of this system are:

- To classify network activities (in the network log) as normal or abnormal while minimizing misclassification.
- To defend computer systems from various cyber attacks and computer viruses.
- To get higher average detection rate, minimum training time.
- To balance the performance of IDS in terms of efficiency and accuracy.

2. LITERATURE SURVEY

Most of the researchers concentrate on genetic algorithm for creating the rules with the help of KDDCUP 99 dataset. The training and testing datasets gives the better consequences for all existing systems. Most of earlier detection system has used Genetic Algorithms (GA). The GA is used to generate classification rules. M. Dave [2] has described the genetic algorithm for intrusion detection method. The proposed method used Genetic Algorithm and implemented it in SNORT using the DARPA datasets for testing. Using SNORT with genetic algorithms creates a unique situation. By using GA they can reduce the rule set of SNORT which leads to, better utilization of SNORT by means of time and also not compromising security risk as it reduce the SNORT rule set based on similarities. Hassan [3] has proposed Intrusion Detection System by using Genetic Algorithm and Fuzzy Logic. It efficiently detects the intrusive activities in network. R. B. Jadhav [4] represents the approach of intrusion detection in network using Genetic, Fuzzy and Apriori Algorithm. In this approach author has proposed a method which is combination of genetic rule, fuzzy rule and association rule for better outcome of intrusion detection. Fuzzy rule can classify network attack data for being a machine learning algorithm whereas genetic algorithm provides best optimal solution by finding proper fuzzy rule and apriori algorithm provides best association rule to detect attack. This approach can be implemented on both well known dataset i.e. KDD cup 1999 as well as on own network dataset. IDS has been evaluated in terms of detection rate, detection speed, false alarm rate and attack types. S. Selvakani [5] has used Genetic Algorithm for network intrusion detection. He has used DARPA dataset, which is standard benchmark for testing intrusion detection system. According to experimental results it is observed that the given approach is effective and more flexible to detect attacks. Jungwon Kim [6] has proposed a system for network intrusion detection. This system has used increment al learning approach on converged data. The result shows the effect of three important parameters such as tolerisation period, activation threshold, and life span. N. Shrivastava [7] has used Support Vector Machine algorithm and Nave Bayes algorithm with Ant Colony Optimization technique. He has used two parameters for performance evaluation: detection rates and false alarm rate. In this system, fuzzy if- then rule is used to increase the accuracy of intrusion detection. C. Cai,L. Yuan [8], has proposed a system in which the network packet is get analyzed and then the ant colony clustering algorithm is used to classify the packet. Finally ,the new algorithm is designed to remove the repetitive computation and increase the detection rate. Y. B. Bhavsar [9], proposed the system in which classification has done by using SVM. The effectiveness of proposed system has identified by conducting some experiments on NSLKDD Cup99 dataset. It is the improved version of KDD Cup99 dataset. The SVM is the most eminent algorithm but it has some drawback, its training time is more. It can be reduced by choosing proper SVM Kernel. R. Shanmugavadivu [10] proposed an anomaly based intrusion detection system. He has used fuzzy rules for attack detection. The set of fuzzy rules are automatically identified by using fuzzy rule learning method which is more efficient for attack detection. Firstly, the rules were generated from attack data as well as normal data. He has used frequent items from both types of data. After that the fuzzy rules were identified automatically and these rules are used to classify test data as normal or attack data. They have used KDDCUP99 dataset for experimental purpose. It has been proved from the results that the proposed system is very effective in detecting intrusions in the computer networks

3. PROPOSED SYSTEM ARCHITECTURE

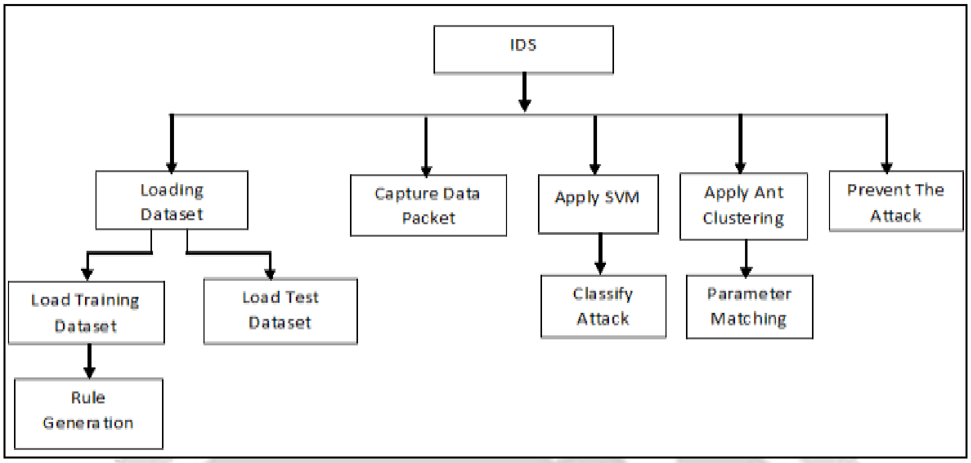

Fig -1: System Overview

3.1 Load Dataset

In this phase training dataset get loaded. There are many dataset but we are going to use KDDCup99 and NSLKDD. It will be given as input to the system and then rules are get generated and this rules will dene the packet formation parameter.KDD99 is standard dataset for evaluation of data-mining based IDSs. In KDD99, the data records of attacks fall into four main categories: Denial of Service (DoS), Remote to Local (R2L), User to Root (U2R), and Probing. In KDD99, each record is described with the help of 41 features.

3.2 Rule Generation (Preprocessing)

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 attributes:

a) Duration
b) Protocol
c) Flag
d) Service
e) Source byte
f) Destination byte
g) 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.

3.3 Capture Data Packet

The packet generation may do in various ways i.e. packet can be generated from the Command prompt or by using any tool such as wireshark. After capturing the packets the features are forwarded towards the IDS then IDS will go to test these packets.

3.4 Apply SVM

Support Vector Machine is generally used for the classification of the given objects. These packets then distinguished so that they can be used for clearly distinguishing the packets about their nature whether they are anomalous in nature or perfectly nes so for that purpose the classification is done by the SVM.

3.5 Apply Ant Colony

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.
1) Begin
2) Randomly select data points from each class.
3) Generate a SVM classifier.
4) While more points to add to training set do
5) Find support vectors among the selected points;
6) Apply CSOACN clustering around the support vectors;
7) Add the points in the clusters to the training set;
8) Retrain the SVM classifier using the updated training set;
9) End
10) End
```

4. CONCLUSION

A new machine learning based data classification algorithm Combined Support vector and ant colony will be develop for the intrusion detection problem. This system will detect the normal and abnormal packet by using SVM and Ant colony. In order to achieve superior performance two offered machine learning algorithms SVM and AC are combined to enhance accuracy rate and faster running time.

5. ACKNOWLEDGEMENT

I take this opportunity to thank our 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, which were indispensable in the completion of this work. 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.

6. 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.