Survey on Intrusion Detection System using Support Vector Machine

Pooja Champaneria¹, Prof. Bhavin Shah², Asst. Prof. Krunal Panchal³

¹Department of Computer Engineering, L. J Institute of Engineering & Technology, Ahmedabad, India.
²M.C.A. Programme, L. J Institute of Management Studies, Ahmedabad, India.
³PG Department, L. J Institute of Engineering & Technology, Ahmedabad, India.

Abstract— Various transaction activities like air ticket reservation, online banking, distance learning, group discussion and so on are carried out using Internet. Carrying out such activities include exchange of useful data which is to be protected from malicious attacks. If some malicious activities are taking place in the network, it is essential to alert the user about it. Detection of malicious activities is of prime importance to protect our data. To protect the data that are exchanged on the network, we need to implement a system that realizes faster attack detection and response accordingly. For detecting malicious activities, we use Intrusion Detection System (IDS) for security purpose. This paper focuses on IDS using Support Vector Machine (SVM). During the survey, the major challenges found are low detection rate and low response time. To overcome these challenges, parallel implementation of SVM classifier and appropriate feature reduction technique should be applied to achieve high detection rate and low response time.

Keywords — Classification, Intrusion Detection System, Multiclass SVM, Neural Network, Support Vector Machine.

I. INTRODUCTION

Due to explosive growth of information exchange and electronic commerce in recent decade, we are using Internet. We use Internet to exchange our useful assets. So we need to implement some security mechanism in order to protect our useful assets. For sharing of confidential, private, public or business assets using Internet connection, protection of these assets against intrusive attacks is of prime importance. Detection of the intrusive behavior is the most important part in the network to prevent the intrusive attacks. For detecting this intrusive behavior in the network, intrusion detection systems are being incorporated in the network. Intrusion detection mechanism can be defined as the process of monitoring the events occurring in a computer system or network and analyzing the same for the signs of intrusion [8].

Various emerging field like genetic algorithm, swarm intelligence, fuzzy logic, neural network are used for incorporating security mechanism in the network.

This paper focus on one of the technique of neural network namely SVM for incorporating security mechanism in the network. Support Vector Machine (SVM) is a supervised learning approach. SVM algorithm, invented by Vladimir N Vapnik, performs the function of classification on the input data.

Section II gives brief overview about Intrusion Detection System. Section III discusses about Support Vector Machine. Section IV includes discussion about literatures reviewed. Section V includes comparison table and comparison analysis of the literatures surveyed. Section VI includes conclusion of the literatures surveyed.

II. INTRUSION DETECTION SYSTEM

We carry out various activities like online shopping, bill payment, movie ticket reservation, air ticket reservation, form filling for examinations and so on. While carrying out activities mentioned above we share important assets so we need some security mechanism for protecting our useful assets from damage caused by intruder. For protecting and preventing our assets from intrusive attacks, Intrusion Detection System is implemented to detect the intrusive behavior in the network. Intrusion detection system is classified [7] mainly as (1) host based intrusion detection system which analyzes activity of one single host in the network and (2) network based intrusion detection system which analyzes activities performed on the network. Technique to detect intrusion in the network is categorized [7] as misuse detection and anomaly detection. Misuse detection performs match of current activities with the attack signatures stored and if match occurs, attack is detected. In anomaly detection if any deviation is found is daily profile of user activity, system is under attack and alarm is triggered. Four main type of network attack [18] are categorized as:

- DoS: Denial of Service, e.g. sync flood
- R2L: Unauthorized access form remote machine, e.g. guessing password
International Journal of Emerging Technology and Advanced Engineering

- U2R: Unauthorized access to local super user(root) privileges, e.g. various buffer overflow attack
- Probe: probing and surveillance, e.g. port

III. SUPPORT VECTOR MACHINE

The original SVM algorithm was invented by Vladimir N. Vapnik and the current standard incarnation i.e. soft margin was proposed by Corinna Corted and Vapnik in 1993 and published in 1995[13]. SVM is a kind of statistical learning based on structural risk minimization principle [11,12]. SVM do not require feature reduction in order to avoid over fitting as the number of parameter depends on the margin that separated the data point [9]. SVM takes input as $S = \{(x_1, y_1), (x_2, y_2), ..., (x_i, y_i)\}$ where $x_i \in \mathbb{R}^n$ and $i=1,2,...,l$ is set of inputs and classifies them into $y_i \in \{\pm 1\}$ which is the target class. SVM uses hyper plane for separating the data points where the data point may belong to one of the positive or negative target class.

The optimization problem can be defined as:

$$\min \frac{1}{2} ||w||^2 + C \sum_{i=1}^{l} \xi_i \quad (1)$$

Subject to,

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i; \quad \xi_i \geq 0$$

Where $C$ denoted the regularization parameter. The decision function i.e. hyper plane which performs classification of the SVM is defined as:

$$f(x) = sign(\sum_{i=1}^{l} \alpha_i y_i k(x, x_i) + b) \quad (2)$$

Where $\alpha_i$ is the Langrange multiplier. Figure below shows the optimal hyper plane for binary classification of data points. An optimal hyper plane should be selected which maximizes the distance between the nearest points in the class [16]. This distance is known as margin.

As SVM performs binary classification, we need some mechanism that classifies the data intelligently when more than one target class is present. In order to perform classification of data into more than one class, multiclass classification approach is adopted. SVM performs multiclass classification using one the following approach 1) one against rest approach 2) one against one approach 3) directed acyclic graph approach[17].

IV. LITERATURE REVIEW

A. Network Intrusion Detection based on Agent and SVM[1]

This paper introduces the network intrusion detection using agent and SVM to improve the detection precision of intrusive attacks. The network intrusion detection system consists of data acquisition module, intrusion detection agent and the management agent. The model in this paper uses four SVM classifiers which classifies the network data into five classes: DoS, probe, U2R, R2L and normal. Experiment is performed on KDDCUP99 dataset to detect the accuracy of attack by applying the SVM model. Same dataset is used to measure the attack detection precision by applying the back propagation approach. By comparing the results, we can know that detection precision of SVM is 0.9457 which is more compared to back propagation which is 0.8771. From the experiment conducted, applying the technique of SVM and agent is better than applying artificial neural network (back propagation).


This paper describes the use of a hybrid method of rough set and SVM. Hybrid method is used in this paper where rough set approach is used for data reduction and SVM approach is used for classification of data and detection of intrusion. Here output may belong to one of the five classes: probe, U2R, R2L and DoS attack data and normal data. Before supplying the data to SVM classifier, reduction of data is performed by applying rough set so that training data are simplified. After that, the data passes through the SVM classifier.
Lastly, experiment is performed on the dataset selected and from the results of the experiment conclusion is made that detection accuracy of R5-SVM is more(93.64%) as compared to SVM(86.62%).

C. An Approach towards Intrusion Detection using PCA Feature Subsets and SVM[3]

PCA is used for feature transformation into high dimension for finding the feature subset after which their performance can be determined during testing process in terms of detection rate and false alarms. PCA explores the correlation between each of the features and determine the important axis to express the scattering of data. Reduced features will be used in training SVM classifier which speeds up the learning of normal and intrusive pattern and to detect these patterns in cross validation. Different feature subsets are selected from transformed feature to evaluate the best subset by applying the steps of PCA. Experiments are performed and various parameters like sensitivity, specificity, false positive rate and false negative rate are evaluated for different feature subsets. From the results of the experiments, we can conclude that improvement in accuracy and false alarm is seen for subset of 10 features where accuracy is 99.465% and false alarm rate is 0.526%.

D. Fuzzy multi-class support vector machine based on binary tree in network intrusion detection[4]

Combining the fuzzy support vector machine and multi-class support vector machine improves the detection accuracy and reduces the training time. One of the important things of FSVM is choosing an appropriate fuzzy membership function [15]. Different membership represents different contributions to the learning of decision surface. A fuzzy membership is assigned during training so it can reduce the influence of outliers and reduce the time of interaction. To construct a binary tree using fuzzy multi class support vector machine, each data class is assigned a fuzzy membership after removing it from the dataset and then train a SVM classifier. This is repeated until all the data in the dataset are trained. Here KDDCUP99 dataset is used for experiment. From the results we conclude that the approach used in this paper is better than Back Propagation, one-versus-rest and Multi class Support Vector Machine.

E. Fuzzy ESVDF approach for intrusion detection system[5]

For feature selection Enhanced Support Vector Decision Function (ESVDF) approach integrated with fuzzy inferencing model is used in this paper.

Contribution of each feature is calculated using a decision function which is weighted sum of training pattern plus a bias. Fuzzy inference model works on three basic components: 1)database 2)knowledge base 3)inference engine. ESVDF is based on SVDF and forward selection ranking (FSR) or backward elimination ranking (BER) to select the most effective features. Experiment is performed on KDDCUP99 dataset. Fuzzy ESVDF approach uses 7 features. Experiment is performed using SVM and neural network approach and from the results, the conclusion can be made easily that using fuzzy ESVDF, improvement in parameters like training time, testing time, accuracy and false positive rate are seen. Execution time of fuzzy ESVDF is short compared to ESVDF/FSR and ESVDF/BER.

F. Detection of DDoS attacks using Enhanced Support Vector Machines with real time generated dataset[6]

Aim of the paper is to generate Distributed Denial of Service dataset and detect them using the Enhanced Support Vector Machine. These dataset contains 10 type of DDoS attack with 14 attributes. EMCSVM is used to classify attack data and normal data. The dataset used focuses on both application layer and network layer DDoS attack.

For training three kernel functions: linear, polynomial and radial basis kernel function are used. Testing result shows that classification rate using radial basis function is more efficient. Also classification using ESVDF is more efficient than SVM. The classification accuracy using the generated dataset is more compared to using the KDDCUP99 dataset using EMCSVM approach.

G. A Hybrid Artificial Immune System for IDS based on SVM and belief function[7]

This paper uses hybrid approach where intrusive behaviour is detected using Dendritic cell algorithm and Dempster belief theory, and classification of data is performed using SVM. The concept of AIS and Dempster belief theory is used to minimize the false alarm generation and detect the intrusive data. DCA performs the classification of the data into normal or abnormal category and Dempster belief theory computes the probability of evidence that supports the class of normal and abnormal data. Experimental result shows that the detection rate for SVM and DCA hardly reaches 92% where as if we apply the hybrid method of this paper, we can see improvement in the detection rate which is 96%. The results shows that by applying the proposed method we can see improvement in the false positive rate, false negative rate alarm generation and detection rate of the attack data.
The comparison of all the literatures discussed above is shown in TABLE I. In the table, almost all approach used the standard benchmark KDDCUP99 dataset for experiment basis which potentially makes the task of comparison easy. The main aim of using SVM in real time application is due to its fast learning and classification. The main focus of incorporating IDS system in network is to increase the rate of attack detection and minimize the generation of false alarms. As KDDCUP99 dataset contains total 41 features, we need to train our data using all of them. The experimental results show that by reducing the feature set before performing the experiment, we can observe improvement in the detection rate of the system. Also more time is required to process more features. So we can extract the useful feature which helps in improving the performance of our system.

Objective of this paper is to find the current challenges of IDS using SVM technique. From the comparison table shown below for the literatures studied, we can find response time [1][7] and detection rate [1][2][4][5][7] as two major challenges.

1. **Response time:**
   Reason for more response time is mentioned as following. First reason is the use of SVM classifier in sequential manner [1]. Second reason is use of all features [2][4].
   Solution to the first problem is implementing SVM classifier in parallel. But parallel implementation requires more hardware resources.
   As a solution to second problem, we can use PCA feature reduction technique to minimize feature set [3] which decreases the computational time and hence can improve response time.

2. **Detection rate:**
   Reason for low detection rate is selection of improper method for preprocessing the data [4][7].
   As a solution to the problem stated above, results in the literatures show that improvement in detection rate is observed when PCA feature subset[3] is used instead of using fuzzy membership function[4] for preprocessing the dataset.

3. **False alarm:**
   Reason for false alarm generation in the system is either due to over training or lack of training provided the data points during preprocessing step.

In paper [3] by applying feature selection process, reduction in false alarm rate is achieved as training process is improved. By applying DCA and Dempster belief theory of paper [7] reduction in false alarm rate is observed as shown in experimental results.

VI. CONCLUSION

Various new types of attacks are generated everyday to harm network data. We need to protect and prevent our important assets that we exchange on the network. To detect these abnormal or attack data, we implement Intrusion Detection System. SVM have properties of high scalability and high speed of classification which proves it efficient for Intrusion Detection System. The current problem observed during the survey, are low detection rate and more response time for IDS using SVM. To achieve higher detection rate and low response time we should apply PCA feature reduction technique and parallel implementation of SVM classifier.

VII. DECLARATION

All the content of the current paper is purely written by Pooja Champaneria under guidance of Bhavin Shah and Krunal Panchal. The content of this paper is written by Author1 (Pooja Champaneria) while Author2 (Prof. Bhavin Shah) had guided the work and Author3 (Asst. Prof. Krunal Panchal) had reviewed this paper. Hence Author1 is responsible for the content and issues related with plagiarism.

REFERENCES


Ginny, Mak, G. Ratzer, H. Vangheluwe. "The implementation of support vector machine using the sequential minimal optimization approach".


Mahesh, Pal. "Multiclass Approaches for Support Vector Machine Based Land Cover Classification".

KDD Cup 1999 DATA, the UCI KDD Archive Information and Computer Science, University of California, Irvine, http://kdd.ics.uci.edu/databases/kddcup99/task.html, last edited on November 30, 1999 at 10:46am.
<table>
<thead>
<tr>
<th>Model</th>
<th>Dataset</th>
<th>Detection rate (%)</th>
<th>Training time(sec), False alarm(%) , no of feature</th>
<th>Methodology</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobile Agent Based SVM [1]</td>
<td>Kddcup99</td>
<td>94.57</td>
<td>-</td>
<td>SVM with Mobile Agent technology</td>
<td>Detection precision of SVM is more than BPNN</td>
<td>SVM classifier are applied in chain which increases response time</td>
</tr>
<tr>
<td>RS-SVM [2]</td>
<td>-</td>
<td>93.64</td>
<td>-</td>
<td>Data reduction by RS &amp; Intrusion Detection by SVM</td>
<td>Accuracy of RS-SVM is more than SVM</td>
<td>Hardware requirement increases if all classifier run in parallel</td>
</tr>
<tr>
<td>PCA feature subset [3]</td>
<td>Kddcup99</td>
<td>99.465</td>
<td>0.52,5%,10</td>
<td>Adatron algorithm by SVM</td>
<td>Accuracy increases using feature subset</td>
<td>Measure of detecting intrusive pattern &amp; FNR is accurate for 30 feature subset</td>
</tr>
<tr>
<td>Fuzzy MCSVM [4]</td>
<td>Kddcup99</td>
<td>Dos-95 Probe-92 U2R-51 R2L-40</td>
<td>9.11sec</td>
<td>Combination of Fuzzy SVM &amp; Multiclass SVM based on binary tree</td>
<td>Training time &amp; accuracy of Fuzzy Multi class SVM is better than MSVM, one-versus-rest, BP</td>
<td>Inferior than MSVM &amp; 1-v-r to detect R2L data</td>
</tr>
<tr>
<td>Fuzzy ESVDF [5]</td>
<td>Kddcup99</td>
<td>99.66</td>
<td>2.351sec, 0.24%,7</td>
<td>ESVDF integrated with fuzzy inferencing model</td>
<td>Improvement in execution time</td>
<td>Accuracy and FPR using 41 features is more accurate</td>
</tr>
<tr>
<td>ESVM [6]</td>
<td>Own dataset</td>
<td>-</td>
<td>14</td>
<td>DDoS dataset is created using script and EMCSVM to classify data</td>
<td>Application &amp; network layer attacks are detected</td>
<td>Linear &amp; polynomial function do not give good accuracy</td>
</tr>
<tr>
<td>AIS [7]</td>
<td>Kddcup99</td>
<td>96.00</td>
<td>-</td>
<td>DCA classification and belief function for evidence</td>
<td>Improvement in accuracy, FPR &amp; FNR</td>
<td>Feature reduction process is time consuming</td>
</tr>
</tbody>
</table>