International Journal of Engineering Research and General Science Volume 3, Issue 2, Part 2, March-April, 2015
ISSN 2091-2730

Intrusion Detection System Based on Genetic-SVM for DoS Attacks

A.Naveena Devi [#1], K.Pradeep Mohan Kumar [*2]

[#]*M.Tech & CSE & Periyar Maniammai University*
[*]*Research Scholar & CSE & Periyar Maniammai University*
Vallam,Thanjavur, India
[1] naveena1310@gmail.com
[2] pradeep_nv2004@yahoo.co.in

***Abstract*--**Nowadays IDS have become a necessary component for protecting interconnection of computer resources and network environment very effectively. Denial-of-service (DoS) is an attack that attacks on public web servers have been recently become a tedious problem in computer society. A denial of service (DoS) attack is a malicious attempt to compromise a server or a network resource unavailable to legitimate users, usually by temporarily interrupting or blocking the services of a requested normal host in the Internet. An intrusion detection system (IDS) was used for detecting malicious traffic, blocking and reporting to the authorized person to take necessary action. So far, many different approaches like encryption techniques, firewall, and access control have been followed in to increase the detection accuracy of DoS attacks. But still it is not sufficient capacity to protect our computer resources very much effectively. so, In this paper, we are focusing on developing new hybrid based IDS model based on genetic algorithm (GA) and support vector machine(SVM) for DoS attack Detection. In the proposed Hybrid IDS model, attacks are identified by training the SVMclassifiers after extracting features from PMU 2014 datasets using geneticalgorithm. SVM classifier deals with large volume of data, make it easy to detect suspicious behaviors, causing speed training and testing process.Genetic-SVM based on wrapper feature selection which is superior then filter based feature selection. The proposed work was implemented in Mat lab7.2. The result shows that the proposed hybrid ids has high detection accuracy (99.5%) and fewer false alarms compared to the existing available models.

***Keywords*—**Intrusion Detection System (IDS),Denial of Service (DoS), GA-SVM

I. INTRODUCTION

An intrusion detection system (IDSs) is the process of identifying, blocking and responding unauthorized activity to the system administrator to take necessary action. Based on the data collection mechanism IDS can be classified in to three types (i) HIDS, (ii)NIDS,(iii)Hybrid IDS. HIDS resides on a particular host and looks for the indications of attacks on that host system. NIDS is located on a separate system and monitor the network traffic for finding attacks based on rule set. Hybrid IDS perform both the functionality of Network based and Host based intrusion detection system. Based on the attack detection techniques IDS can be classified into (i)Anomaly Detection, (ii) misuse Detection. In anomaly based detection, captured network traffic data is used to differentiate attack traffic data compared with predefine normal pattern.

On the other hand, misuse detection system, also called as signature based IDS, uses patterns of well known attacks to match with captured traffic to find out attack pattern easily A lot of computational intelligence approaches have been proposed by the researchers for example artificial neural network, fuzzy sets, evolutionary computation, expert system approach, rule based approach, artificial immune systems etc [2].The existing datasets KDDCUP99 include the Neptune, smurf, Pod and Teardrop are the types of DoS attacks. The DoS attacks is a emerging attacks for creates threat to business and Internet service providers around the world. Computational mechanism is needed to encounter this type attacks and extend the support provided to environment security that increase the hopeful of the users to do Internet based business. so, IDS is the more power full tool to detect the various types of DoS attacks with higher detection accuracy, reducing false alarms. The working of IDS shown in fig.1.IDS continuously captured the network activity and gives the report to the system administrator. Finally the alarm report will be generated. The both function of monitor system and security administrator response to the intrusion.

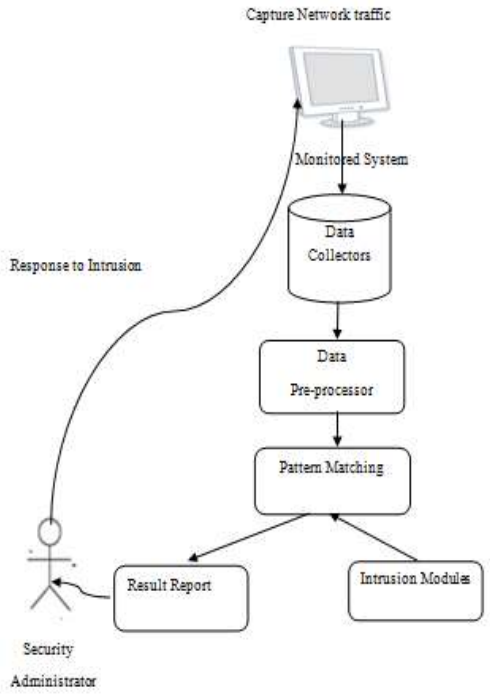

fig 1.Working of IDS

IDS is tested with large amount of dataset that the aim is slow training, testing process and low detection rate. So, feature extraction is the challenging task in developing IDS [8].Generally, the implementation of IDS consists of three phases such as data preprocessing, features extraction and classifier. The tasks that are carried out in preprocessing phases are (i) identifies the attributes and their value (ii) Convert categorical to numerical data (iii) Data normalization and (iv) compute redundancy check and handle about null value. Feature extraction process is a preprocessing step when constructing IDS, used to reduce the dimensionality of the dataset by removing irrelevant, redundant features and improving the prediction accuracy of the classifier using selected features from the dataset. Classifier module finds the conditions of the traffics are either legitimate or malicious attack. Classifier is faced with a problem when it has to generate rules with many attributes or features.

Obviously, the time required to generate rules is proportional to the number of features. In addition, irrelevant and redundant features can reduce both the predictive accuracy and comprehensibility of the induced rule and degrade the classifier speed. Thus, selecting the most relevant features is necessary, this strategy is implemented to simplify the rules and reduce its computational time while retaining the quality of classifier, as it represents the original features set. Support Vector Machines are known as maximum-margin classifiers since they find the optimal hyper plane between two classes, defined by a number of support vectors. The feature of the technique is mainly due to the introduction of calculation of pattern weight that allows us to prevent the effects of outliers by permitting a certain amount of misclassification errors. Although this technique was able to provide only linear classification and also handle non-linear problems. The Objective function is used to implicitly map the data points into a higher-dimensional feature space. The rest of this paper is organized as follows. In section II discusses the related works about existing algorithm to detect dos attack, In section III Illustrates the proposed genetic-SVM abased IDS model. Section IV describes the implementation and performance of the proposed algorithm using PMU dataset. In the last section, deals with conclusion and future work.

II. EXISTING METHOD

In order to detect the SYN flood attacks, There are many methods and frameworks. A few of them are given in this section. The authors detected the SYN flooding attacks at leaf routers which connect end hosts to the Internet, that utilizes the normalized difference between the number of SYNs packets and the number of FIN (RST) packets in a time interval. The router recognizes that some attacking traffic is mixed into the current traffic, . If the rate of SYNs packets is much higher than that of FIN (RST) packets by a non-parametric cumulative sum algorithm. Similar works have been presented, where the fast and effective method was proposed for detecting SYN flood attacks. Moreover, a linear prediction analysis was proposed for DoS SYN flood attack detection. This mechanism makes use of the exponential back off property of TCP used during timeouts. it is shown that this approach is able to

detect an attack within short delays, By modelling the difference of SYN and SYN&ACK packets. Again this method is used at leaf routers to detect the attack without the need of maintaining any state.

However, considering the fact that the sources of attack can be distributed in different networks, there is a lack of analysis for the traffic near the sources and also the detection of the source of SYN flooding attack in TCP based low intensity attacks is missing. Moreover, a quite similar approach has been used , which also considers a non-parametric cumulative sum algorithm; then apply it to measure the number of SYN packets, and by using an exponential weighted moving average for obtaining a recent estimate of the mean rate after the change of SYN packets. Three counters algorithms for SYN flooding defence attacks was proposed and included detection and mitigation. The detection scheme utilizes the inherent TCP valid SYN–FIN pairs behaviour, which is capable of detecting various SYN flooding attacks with high accuracy and short response time. The mitigation scheme works in high reliable manner for victim to detect the SYN packets of SYN flooding attack. Although the given schemes are stateless and required low computation overhead, making itself immune to SYN flooding attacks, and the attackers may retransmit every SYN packet more than one time to destroy the mitigation function.

The authors have built a standard model generated by observations from the characteristic between the SYN packet and the SYN+ACK response packet from the server. The author have proposed a method to detect the flooding agents by considering all the possible kinds of IP spoofing, which is based on the SYN/SYN-ACK protocol pair with the consideration of packet header information. The Counting Bloom Filter is used to classify all the incoming SYN-ACK packets to the sub network into two streams, and a nonparametric cumulative sum algorithm is applied to make the detection decision by the two normalized differences, with one difference between the number of SYN packets, the number of the first SYN-ACK packets, another difference between the number of the first SYN-ACK packets and the number of the retransmission SYN-ACK. There are also some other related studies such as SYN cookies, SYN filtering mechanisms , SYN cache, SYN proxy (firewall), SYN kill and D-SAT .The ESDM is a simple but effective method to detect SYN flooding attacks at the early stage. The ESDM achieves shorter detection time and small storage space. However, these exiting methods or defence mechanisms which oppose to the SYN flooding attack are effective only at the later stages, when attacking signatures are obvious.

II. PROPOSED GENETIC-SVM BASED IDS

The architecture of the proposed GA-SVM model is shown in Fig.1. The architecture contains two phases (i) Training phase (ii) Testing phase. In training phase, KDDCUP 99 dataset undergoes data pre-processing and the pre-processed data is then fed to the feature selection block where feature selection is done using genetic algorithm. The selected features are then given as input to the classifier where DoS attack patterns are classified using PSO. The other stage is the testing stage where the captured traffic is pre processed as in training phase and the identified patterns are matched with the stored DoS patterns in database thereby taking a decision. If any new patterns were found during the analysis of traffic behaviour and if it was found against the legitimate traffic, the new pattern will be captured and updated in the database. The implementation of Genetic-SVM based IDS has three phases which includes 3.1. Preprocessing, 3.2Feature Selection and 3.3. Classifier.

A. Data Preprocessing

Data Preprocessing is an important step in the machine learning computing that eliminates out of range values, impossible data combinations, missing values etc. Generally data preprocessing includes learning, normalization, transformation, feature extraction and selection. The output of the data preprocessing is the final training set that extracts knowledge for the testing phase. The following steps are involved in data preprocessing.

- Identifying features and its related values.
- Converting original feature data value in to numerical data value.
- Applying data normalization based on min-max normalization.
- Performing similarity checks and removing null values.

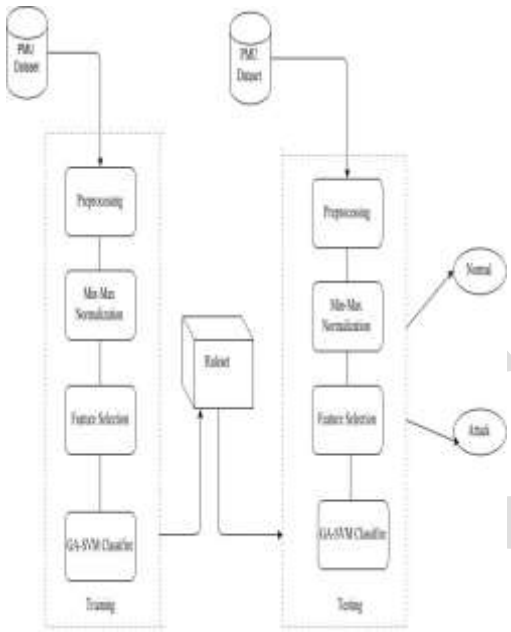

Fig. 2. Architecture of Genetic-SVM based IDS

B. Feature selection based on Genetic algorithm

Accuracy of the classifier depends on the selection of optimum feature subset. Feature selection method is mainly used for selecting the subset of features from the original data set. Two feature selection methods namely filter method and wrapper method were already proposed. Filter method is mainly based on the general characteristics of data features without involving machine language. These features are ranked based on certain criteria, where features with highest rank values are selected as optimal. The main advantages of filter method are low computational cost without involving any machine language algorithm for feature selection. Frequently used filter method is the information gain method. Wrapper method is mainly used for feature subset selection from the data set based on objective function and analysis of the performance of feature subset. In this paper, Genetic Algorithm (GA) is used to select optimal feature subset from the dataset. GA reduces the PMU features from 41 attributes to 6 attributes those are related to the characteristics of DoS attack thereby reducing 85% of the space of features. The six attributes that are considered by the GA are 1.Protocol, 2.src_bytes, 3.dst_bytes, 4.count (No of connections to the sameDest),5.srv_count,6.same_srvrate,7.Diff_srvrate,8.logged_in,9.diff_host_same_src_portrate,10.service.

The existing KDDCUP'99 dataset contains huge number of redundant records. 10% of the full dataset contains two types of DoS attacks (Smurf and Neptune). These two types constitute over 71% of the testing dataset which completely affects the evaluation of IDS. The steps involved in GA where features are selected from the dataset are presented below in Alg.1.

1) Initialize pre-processed data as population.
2) Calculate objective function based on derived rules for DoS attack for each individual pre processed data.
3) Select individual solution.
4) Perform mating of pair of individuals.
5) Perform mutation operation.
6) Calculate objective function for newly created population.
7) If (6) is satisfied, stop the operation.
8) If (6) is not satisfied, repeat step 3.
9) Return the best features from PMU dataset that reflects the properties of DoS attacks.

Alg.1. *Genetic Algorithm based feature selection.*

C.SVM classifier

GA generates a set of enhanced population of chromosomes, i.e. a group of individuals with different chromosomes. Each individual chromosome consists of ten different parameters namely protocol_type, service, src_bytes, dst_bytes, count, srv_count, Ssrv_rate, Dssrv_rate, logged in, Dst_host_same src port rate. The pattern weight of the individual chromosome should be determined properly by using the training dataset to include as many solutions as possible. Calculate the fitness value of each individual in the

initial population using Eq. (2) and rank them according to their fitness value. In Eqn.2, X indicates the training dataset and Y indicates the enhanced chromosome subset.

$$F = \sum_{k=1}^{n} X * Y$$

Equ 2

To calculate the fitness value of an individual or a chromosome, the training record is compared with each gene of the chromosome in the normal population. So each and every record generates different pattern value for different feature values. Similarly training record is compared with each gene of the chromosome in the attack population. So each and every record generates different pattern value for different feature values. Finally we will calculate support vector values for normal pattern and attack pattern with the help of pattern weight for normal and attack population. Now, we will get two SVM values i.e 1 and 0. F>=1 indicates normal record and F<1 indicates attack record. SVM classifies our dataset based on newly generated hyper plane values. Now each and every testing record is compared with each and every gene of the normal and attack population. This will generate a pattern weight {0, 1, 2, 3, 4, 5} based on which we will identify whether our testing record belongs to normal or attack model.

Rule set for Dos Attacks:

Normal Rule set

protocol=tcp,sourceIp=172.20.62.33,DestIp=172.20.62.255,178>src_byte<322,10>Dst_byte<224,SYNcount=1or2,Ack=1or2,FIN_count=1or2,RSTbit=0,Outofseqpacket=0,0>dst_,host_same_src_port_rate<1,Src_ data_packet=55.

Neptune Rule Set

protocol=tcp,sourceIp=172.20.62.33,DestIp=172.20.62.255, src_byte= 0, Dst_byte=0, SYN bit=3 to 160,ACK=0,FIN=0,RSTbit=0,Outseqpacket=0,0>dst_host_same_src_port_rate<1,Src_data_packet=0.)

Smurf Rule Set

protocol = UDP, source Ip=172.20.62.33 ,Dest Ip=172.20.62.25, src_byte= 221120,Dst_byte=0,SYN bit=238to512,ACK=0,FINcount=0,RST bit=0,Out seq packet=0,0<dst_,host_same_src_port_rate<2,Src data packet=560.

IV. SIMULATION RESULT AND DISCUSSIONS

The simulation of the proposed IDS model was implemented in MAT Lab 7.1 environment. Using PMU Dataset 2014, feature selection has been done using genetic search filter method. Genetic search has reduced the dimensionality of the feature from 113 to 12. Genetic search has reduced 89% of the features.2000000 lakhs records are used for testing.These records mainly focused on TCP,UDP traffic around 100% records belongs to DoS attack traffic.So, it is necessary for training the machine language for DoS attack. if done, that will increase the detection accuracy and reducing false alarm rate. Only 0.5%, 0.5% and 0.5% of Normal, Smurf and Neptune attack instances has been misclassified respectively.

Hybrid based IDS(GA-SVM) is evaluated based on how correctly intrusion is predicted. Given event is compared with predefined knowledge of IDS and it produces four types of outcomes.

-True Positive.

-True Negative.

-False Negative.

-False Positive.

Detection rate=TP/(TP+TN)*100.

TABLE I. GA-SVM EXPERIMENTRESULTS

Test Data	**Trainig Data**	**Test data**	**Detection accuracy(%)**
			Enhanced GA-SVM (Anomaly and Misuse Detection)
Normal	8671	5460	100
Smurf	40018	20456	99.5
Neptune	152065	62345	99.5

V. CONCLUSION

In this thesis, new hybrid based computational techniques were proposed for extracting the attacking patterns available in the datasets. The result shows that enhanced GA reducing false alarm rate incorporates with SVM. In this model irrelevant and redundant features are not recognized that brings down the processing speed of evaluating the known patterns. An efficient features selection model eliminates dimension of data, reduce redundancy and ambiguity caused by none important attributes. Hence, the performances of the proposed hybrid models are better than existing models.The proposed methods performs the classification task and extract the recovered knowledge using GA-SVM.These systems are highly reliable, adequate interpretability and compare with several well known algorithms such as SVM,snort based hybrid system, Teacher Learning based Optimization IDS, Group Teacher Learning based Optimization IDS and Fuzzy logic IDS. The experiment results emphasized that the proposed hybrid models are suitable technique and produced better accuracy compared to the existing model. In future work, the octopus activities will be studied, use as a detection technique to find the patterns of the attacks and evaluate the performance with existing IDS.

REFERENCES:

[1] Yogendra Kumar Jain and Upendra, "An Efficient Intrusion Detection based on Decision Tree Classifier Using Feature Reduction," International Journal of Scientific and Research Publication, Volume 2, Issue 1, pp. 1-6, January 2012.

[2] J. Gómez1, C. Gil2, N. Padilla1, R. Baños2, and C. Jiménez1," Design of a Snort-Based Hybrid Intrusion Detection System".

[3] D. M. Divakaran, H. A. Murthy and T. A. Gonsalves, "Detection of SYN Flooding Attacks Using Linear Prediction Analysis", 14th IEEE International Conference on Networks, ICON 2006, pp. 218-223, Sep. 2006.

[4] D. Nashat,X. Jiang and S. Horiguchi, "Detecting SYN Flooding Agents under Any Type of IP Spoofing", IEEE International Conference on eBusiness Engineering table of contents, 2008.

[5] W. Chen and D.-Y. Yeung, "Defending Against SYN Flooding Attacks Under Different Types of IP Spoofing", ICN/ICONS/MCL '06, IEEE Computer Society, pp. 38-44, April 2006.

[6] G. Wei, Y. Gu and Y. Ling, "An Early Stage Detecting Method against SYN Flooding Attack", International Symposium on Computer Science and its Applications, pp.263-268, 2008.

[7] CERT Advisory CA-1996-21 "TCP SYN flooding and IP spoofing attacks".[Online].Available: http://www.cert.org/advisories/CA-1996-21.html

[8] BoonPing Lim Md. Safi Uddin, 2005,"Statistical-based SYN-flooding Detection Using Programmable Network Processor".(ICITA'05)

[9] Tamas Abraham, "IDDM: Intrusion Detection using Data Mining Techniques", Information Technology Division,Electronics and Surveillance Research Laboratory, DSTOGD-0286.

[10] Wei Li, "Using Genetic Algorithm for Network Intrusion Detection", Department of Computer Science and Engineering, Mississippi, State University, Mississippi State,Ms 39762.

[11] J. Haggerty, T. Berry, Q. Shi and M. Merabti, 2004, "A System for Early Detection of TCP SYN Flood Attacks",

IEEE Communications Society Globecom