ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

A Review paper on Identifying Intrusion Detection System using Hybrid technique with Support Vector Machine

Swapna Yendole, Prof. Sujata Tuppad

M.E Student, Dept. of CSE., MSSCET, Jalna, India

Professor, Dept. of CSE., MSSCET, Jalna, India

ABSTRACT: Various transaction activities such as booking airline tickets, online banking, distance learning, group discussion and so on are performed using the Internet. The realization of these activities involves the exchange of useful data that must be protected against malicious attacks. If certain malicious activities occur in the network, it is essential to alert the user to this. Detecting malicious activity is critical to protecting our data. To protect the data that is exchanged on the network, we need to implement a system that achieves faster attack detection and response accordingly. To detect malicious activity, we use the intrusion detection system (IDS) for security reasons. This article focuses on the IDS using the Vector Support Machine (SVM). During the survey, the main challenges identified were the low detection rate and the response time. To overcome these challenges, the parallel implementation of the SVM classifier and the appropriate technique of characteristic reduction must be applied to achieve a high detection rate and a low response time.

KEYWORDS: Classification, Intrusion Detection System, Multiclass SVM, Neural Network, Support Vector Machine.

I. INTRODUCTION

Due to the explosive growth of information exchange and e-commerce over the last decade, we are using the Internet. We use the Internet to exchange our useful assets. So we have to put in place a security mechanism to protect our valuable assets. For the sharing of confidential, private, public or commercial property by means of an Internet connection, the protection of these assets against intrusive attacks is of paramount importance. Detection of intrusive behavior is the most important part of the network to prevent intrusive attacks. In order to detect this intrusive behavior in the network, intrusion detection systems are incorporated into the network. The intrusion detection mechanism can be defined as the process of monitoring events occurring in a computer system or network and analyzing the same for intrusion signs [8].Different emerging areas such as genetic algorithm, swarm intelligence, fuzzy logic, neural network are used to integratethe security mechanism into the network.

We carry out various activities such as online shopping, payment of invoices, reservation of movie tickets, booking of air tickets, filling out forms for exams and so on. As part of the activities mentioned above, we share important assets, we need a security mechanism to protect our valuable assets from damage caused by the intruder. To protect and prevent our assets against intrusive attacks, the intrusion detection system is implemented to detect the intrusive behavior of the network. The intrusion detection system is classified [7] mainly as (1) host-based intrusion detection system which analyzes the activity of a single host in the network, and (2) the detection system of Network intrusion that analyzes the activities carried out on the network. The technique of intrusion detection in the network is categorized [7] as detection of misuse and detection of anomalies. The misuse detection performs the matching of current activities with the stored attack signatures and if a match occurs, an attack is detected. In the anomaly detection if a deviation is found is the daily profile of the user's activity, the system is attacked and the alarm is triggered.Four main type of network attack [18] are categorized as:

DOI: 10.15680/IJIRCCE.2016. 0412181

DoS: Denial of Service, e.g. sync flood

R2L: Unauthorized access form remote machine, e.g. guessing password
U2R: Unauthorized access to local super user(root) privileges, e.g. various buffer overflow attack
Probe: probing and surveillance, e.g. port

II. RELATED WORK

In [2] Author presents network intrusion detection using agent and SVM to improve the accuracy of detection of intrusive attacks. The network intrusion detection system consists of a data acquisition module, an intrusion detection agent and the management agent. The model used in this paper uses four SVM classifiers that classify network data into five classes: DoS, probe, U2R, R2L, and normal. The experiment is performed on the KDDCUP99 dataset to detect the precision of the attack by applying the SVM model. The same dataset is used to measure the attack detection accuracy by applying the back propagation approach. By comparing the results, one can know that the SVM detection accuracy is 0.9457 which is more compared to the back propagation which is 0.8771. From the experiment performed, the application of the technique of SVM and agent is preferable to the artificial neural network application (back propagation).

In[3]It also shows the optimal hyper plane for binary classification of data points. An optimal hyper plane should be selected such that it should maximize the distance between its nearest points that belong to the class [16]. This distance is known as margin.As SVM performs binary classification, we need a mechanism that intelligently classifies data when more than one target class is present. In order to classify data in more than one class, a multiclass classification approach is adopted. SVM performs the multi-class classification.

This paper describes the use of a raw ensemble hybrid method and SVM. Hybrid method is used in this paper where the approximate approach is used for data reduction and SVM approach is used for classification of data and detection of intrusion. Here, the output can belong to one of five classes: probe, U2R, R2L and the DoS attack data and the normal data. Before providing the data to the SVM classifier, data reduction is accomplished by applying a rough set to simplify the training data. After that, the data passes through the SVM classifier Finally, the experiment was carried out on the selected data set and the results of the experiment concluded that the detection accuracy of RS-SVM was higher (93.64%) than that of SVM (86.62%).

III. PROPOSED ALGORITHM

We focus on an automated learning model using a modified vector support machine (SVM) that combines the benefits of supervised and unsupervised learning. In addition, we propose a preliminary feature selection process using GA to select more appropriate packet fields. Now we discuss our hybrid algorithm steps which are as follows:

Step 1 - Load the kdd dataset first.

Step 2 - Preprocessing the Data

Here, process all the data in the database. The KDD CUP database "99 contains 41 functions such as dst_bytes, src_bytes, etc. Since the SVM classification uses only numerical data for testing and training, it is necessary to convert the textual data into numerical values. , We assumed some numeric values for different text functionality, such as "protocol_type" functionality "tcp" as 3, "udp" as 7, and "icmp" as 9 etc as shown in the table.

Transformation Table for translating the Text data to numeric data in KDD cup'99 Data Set

TYPE	CLASS	NO.
Attack/ Normal	Attack	1
	Normal	0
Protocol Type	TCP	3
	ICMP	9

	UDP	7
Flag	OTH	1
	REJ	2
	RSTO	3
	RSTOS0	4
	RSTR	5
	S0	6
	S1	7
	S2	8
	S3	9
	SF	10
	SH	11
Services	Auth	1
	Bgp	2
	Courier	3
	csnet_ns	4
	Ctf	5
	Daytime	6
	Discard	7
	Domain	8
	domain_u	9

Step 3- Feature Selection Algorithm (Weka SVM Decision Tree Stump)

In this work, the genetic algorithm-based approach is proposed to select the optimal characteristics of the 41 global characteristics. The selected characteristics discriminate in the predictive class when classifying for abnormality.

The steps of the algorithm are as follows:

1. Generate a random population of n chromosomes (data set of appropriate solutions for the problem)

2. Evaluate the physical condition f (x) = k (x) / sqrt (k + k (k-1) x) where k is a random number and x represents the chromosome of the population

3. Create a new population by repeating the following steps until the new population is complete,

A) Choose two relative chromosomes of a population according to their fitness (better physical condition, greater chance of being selected).

B) With a probability of crossing, the parents form a new offspring (children). If no crossing has been done, the offspring is an exact copy of the parents.

C) With a mutation probability, mutate new offspring at each locus (position in the chromosome).

D) Place new descendants in a new population.

4. Use the new generated population for a new algorithm sequence

5. If the final condition is satisfied, stop and return the best solution in the current population

6. Go to step 2.

Step 4- Selected feature

The main reason for selecting the KDD Cup 99 dataset is that it is currently the commonly used dataset that is shared by many researchers. In this dataset, 41 attributes are used in each record to characterize the behavior of network traffic. Of these 41 attributes, 38 are numeric and 3 symbolic. The characteristics present in the KDD dataset are grouped into three categories and are discussed below.

A. Basic Functions: Basic features include all attributes extracted from a TCP / IP connection. These functions are extracted from the packet header and include bytes src, dst_ bytes, protocol etc.

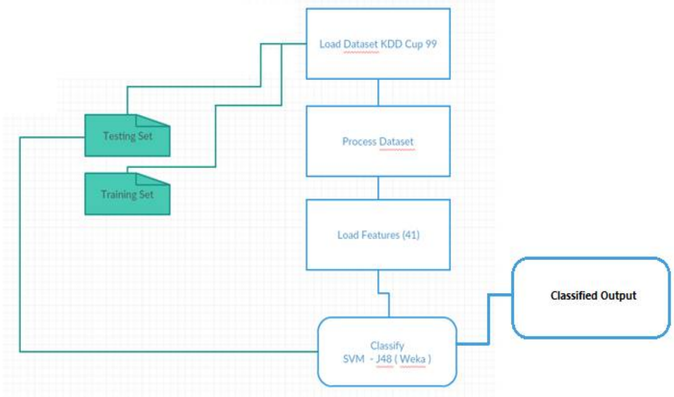

Figure 1 Architectural Diagram

B. Content features: These functions are used to evaluate the payload of the original TCP packet and to look for suspicious behaviors in the payload part. This includes features such as the number of failed logon attempts, the number of file creation operations, and so on. In addition, most R2L and U2R attacks do not have frequent sequential sequences. This is because DoS and Probing attacks involve many connections to certain hosts in very short time, but R2L and U2R attacks are embedded in the data portions of the packets and usually involve a single connection. To detect these types of attacks, content-based features are used.

C. Traffic functions: These functions include functions calculated according to a window interval and are divided into two categories
I) "Same host" functions: These functions are derived only by examining the connections of the last 2 seconds which have the same destination host as the current connection and calculating the statistics relating to the behavior of the protocol, the service and so on.
B. Content features: These functions are used to evaluate the payload of the original TCP packet and to look for suspicious behaviors in the payload part. This includes features such as the number of failed logon attempts, the number of file creation operations, and so on. In addition, most R2L and U2R attacks do not have frequent sequential sequences. This is because DoS and Probing attacks involve many connections to certain hosts in very short time, but R2L and U2R attacks are embedded in the data portions of the packets and usually involve a single connection. To detect these types of attacks, content-based features are used.

C. Traffic functions: These functions include functions calculated according to a window interval and are divided into two categories
II) "Same host" functions: These functions are derived only by examining the connections of the last 2 seconds which have the same destination host as the current connection and calculating the statistics relating to the behavior of the protocol, the service and so on.

Step 5- Classification algorithm

We have divided the user's behavior into two classes, namely attack and normal, where the user's behavior is the collection of different attacks belonging to the five classes such as
1 Normal-- Normal
2 DoS - apache2, back, earth, mailbomb, neptune, pod, processtable, smurf, tear, udpstrom

3 Probe - ipsweep, mscan, nmap, portsweep, saint, satan

4 R2L - ftp_write, guess_passwd, imap, multihop, named, phf, sendmail, spy, snmpgetattack, snmpguess, warezclient, warezmaster, worm, xlock, xsnoop
5 U2R - buffer_overflow, httptunnel, loadmodule, perl, ps, rootkit, sqlattack, xtern

The purpose of our SVM experiment is to differentiate the normal behavior and attack behavior of the user. In our experiments, normal data is classified as -1 and all attacks are categorized as +1.

The design of the basic input data and the output data areas are given as follows:

(Xi, yi), ..., (xn, yn), x Rm, y {+ 1, -1}
Where xi, yi, ..., (xn, yn) are train data, n is the number of samples, m is the input vector and y is inserted into the category +1 or -1 respectively. On the linear problem, a hyperplane can be divided into two categories. The formula of the hyperplane is:

$(W.x) + b = 0$

The categories are:
$(W.x) + b \geq$ if $yi = +1$
$(W.x) + b \leq$ if $yi = -1$

Step-6 classification result

Step 7- Anomaly detect.

IV. Pseudo Code

```
Step 1:Load the kdd dataset first
Step 2:  Preprocessing the Data.
Step 3:  Feature Selection Algorithm ( Weka SVM  Decision Tree Stump )
Step 4:  Selected feature
Step 5: Classification algorithm
Step 6:  classification result
Step 7: Anomaly detect.
 Step 8: End
```

V. Conclusion and Future Work

Different new types of attacks are generated every day to harm network data. We must protect and prevent our important assets that we exchange on the network. To detect this abnormal or attack data, we implement the intrusion detection system. SVMs have high scalability and high-speed classification properties, which proves its effectiveness for the intrusion detection system. The current problem observed during the survey is a low detection rate and more response time for ID using SVM. To obtain a higher detection rate and a lower response time, we must apply the technique of reduction of the PCA characteristics and the parallel implementation of the SVM classifier.

REFERENCES

[1] A hybrid method based on Genetic Algorithm, Self-Organised Feature Map, and Support Vector Machine for better Network Anomaly Detection, 4th ICCCNT 2013 July 4-6, 2013, Tiruchengode, India
[2]Guan Xiaoqing,GuoHebin,ChenLuyi,“Network intrusion detection method based on Agent and SVM” ,The 2nd International Conference on Information Management and Engineering (ICIME), IEEE 2010.

[3]Liu Zhiguo, Kang Jincui, Li Yuan "A hybrid method of rough set and support vector machine in network intrusion detection" 2ndInternational Conference on Signal Processing System(ICSPS),IEEE 2010.
[4]Noreen Kausar, BrahimBelhaouari Samir, SuziahSulaiman, IftikharAhmad, Muhammad Hussain "An Approach towards Intrusion Detection using PCA Feature Subsets and SVM"InternationalConference on Computer & Information Science (ICCIS),IEEE 2012.
[5]Lei Li, Zhi-ping Gao, Wen-yanDing ,"Fuzzy multi-class support vector machine based on binary tree in network intrusion detection",International Conference on Electrical and Control Engineering(ICECE),IEEE 2010.
[6]Zaman S, Karray F "Fuzzy ESVDF approach for intrusion detection system", International Conference on Advanced Information Networking and Applications(AINA), IEEE 2009.
[7]Subbulakshmi T, Shalinie S. M, Ganapathi Subramanian V,BalaKrishnan K, AnandKumar D, Kannathal K "Detection of DDoS attacks using Enhanced Support Vector Machine with real time generated dataset" Third International Conference on AdvancedComputing(ICoAC), IEEE 2011.
[8]Singh S, Singh J P, Shrivastva G "A hybrid artificial immune system for IDS based on SVM and belief function", Fourth International Conference on Computing, Communications and Networking Technologies(ICCCNT),IEEE 2013.
[9]D.S Bauer, M.E Koblentz "NIDX- an expert system for real-time network intrusion detection",IEEE, Proceedings of the Computer Networking Symposium, 1988. pp. 98-106.
[10]SrinivasMukkamala, Guadalupe Janoski, Andrew Sung "Intrusion Detection Using Neural Networks and Support Vector Machines",IEEE, International Joint Conference on neural network,IEEE 2002.
[11]SafaaZaman and FakhriKarray "Features Selection for Intrusion Detection Systems Based on Support Vector Machines", 6th IEEE Consumer Communications and Networking Conference, IEEE 2009.
[12]V.N. Vapnik, "An Overview of Statistical Learning Theory", IEEE Transactions on Neural Networks, 10(5):988-999, 1999.
[13]Xuehua Li, LanShu "Fuzzy Theory Based Support Vector Machine Classifier", Fifth International Conference on Fuzzy Systems and Knowledge Discovery, IEEE 2008.
[14]Cortes C Vapnik V "Support-vector networks", 1995 Machine Learning**20**: 273.
[15]Ginny Mak, G. Ratzer, H. Vangheluwe "The implementation of support vector machine using the sequential minimal optimization approach".
[16]A. Zadeh "Fuzzy Sets", Information and Control 8,338-353, 1965.
[17]Colin Campbell "Simple Learning Algorithms for Training Support Vector Machines".
[18]Mahesh Pal "Multiclass Approaches for Support Vector Machine Based Land Cover Classification".
[19] P.Garc a-Teodoroa,, J. Daz-Verdejo, G. Macia-Fernandez, E. Vazquez, , "Anomaly-based network intrusion detection: Techniques, systems and challenges,"computes& security 28,(2009) 1 8-2 8.
[20] TaeshikShon', YongdueKim, Cheolwon Lee', and JongsubMoon , "A Machine Learning Framework for Network Anomaly Detection using SVM and GA," Proceedings of the 2005 IEEEWorkshop on Information Assurance and Security, 2005.
[21] VarunChandola, "Anomaly detection: A Survey,"A modified version of this technical report will appear in ACM computing Surveys,(2009).
[22] MarinaThottan and ChuanyiJi, "Anomaly Detection in IP Networks,"IEEE Transactions On Signal Processing,Vol. 51, No. 8,August 2003
[23] Lung Huang, Chieh-Jen Wang, "A GA-based feature selection and parameters optimization for support v
[24]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

BIOGRAPHY

Swapna R. Yendole is a Student of Master of Engineering in the Computer Science and Engineering Department, Matsyodari College of Engineering & Technology, Babasaheb Ambedkar Marathwada University, She Received Bachelor of Engineering (BE) in 2013 from BAMU ,Aurangabad, Maharashtra, India

DOI: 10.15680/IJIRCCE.2016. 0412181