ISSN:0975-9646

Denial of Service Attack and Classification Techniques for Attack Detection

Prajakta Solankar[1], Prof. Subhash Pingale[2], Prof. Ranjeetsingh Parihar[3]

[1,2,3]*Dept. Of Computer Science & Engineering,*
Solapur University
SKN Sinhgad College of Engineering Pandharpur,
Maharashtra, India

***Abstract*—Denial of service attack is critical threat to the internet and affects computing systems such as web server, database server etc. An internet user performs activities like emailing, online banking and news, general browsing etc. Denial of service (DoS) attack prevents the user from accessing these services.The attackers's main aim while performing denial of service attack is to prevent legitimate users from using the system resources. That's why there is need to detect DoS attacks. This paper shows various techniques for classification of attack. K-Nearest Neighbor (K-NN), support vector machine (SVM), decision tree and naïve bayes are described and experimental results by using weka tool are determined.**

***Keywords*—Denial of Service, classification, SVM, K-NN, decision tree and naive bayes**

I. INTRODUCTION

In recent years internet has been affected by Denial of service attack. Denial of service attack is caused due to availability issue. The availability of data is important while running a business and huge losses may occur if information is not available for longer time because of attack against computers. Such attacks are known as Denial of Service. The purpose of denial of service attack is to prevent or disrupt the services provided by victim. The DoS attack stops legitimate user from accessing services provided by internet, victim which can be any host or router on network or stops access to network resource.

When DoS attack occurs then there is no way to fix the problem or victim not able to provide its services on the internet [1].DoS attack tools are easily available on the internet, so any computer user can perform DoS attack by using tools like Trinoo. Some companies also perform DoS attack because of competition in market.There are various forms of DoS attacks in the literature according to several parameters: Smurf attack, back, land, Neptune, teardrop attack etc are DoS type attacks. Some types of DoS attack are explained as follow:

In smurf attack, the attacker uses ICMP echo request packets to perform an attack. Attacker sends these packets to intermediate device. ICMP packets contain source address same as victim's IP address and intermediate devices address as destination address. Then all the intermediate hosts receive ICMP requests and sends replies to victim's address. Because of this victim address gets flooded with these replies and consumes bandwidth of network. Mostly router and OS are affected by this type of attack. In Land attack; the attacker sends TCP SYN packets to the victim's address. These packets contain source and destination address same as victim. So victim machine go into loop because it replies to itself [2].

In Teardrop attack attacker sends packets in fragments with overlapping offset value. When teardrop attack runs against a machine, it will crash or reboot. On windows machine user will experience blue screen of death.

The back attack is launched against an apache web server, which is flooded with request containing a large number of front slash characters (/) in URL descriptions. When server tries to process all these requests then it becomes unable to process other legitimate requests and hence it denies a service to its customers.

In February 2000, Yahoo has experienced DDoS attack. Citi and other U.S. financial institutions also experienced DDoS attacks which were intended to disrupt consumer online banking services in 2012. HSBC servers affected a number of HSBC websites around the world because of Denial of service attack on October 2012.Even in early 2013 and 2014 also DDoS has happened.DoS attack degrades the performance of server by flooding the ICMP traffic as discussed above in the smurf attack. So detecting denial of service attack is necessary.

II. RELATED WORK

The author proposed a triangle area based nearest neighbor approach for detecting intrusions. In this paper they presented k-means is used as clustering method to find the center clusters for each category. They developed triangle areas. Finally they used k-NN classifier for classification [3].

The machine learning techniques radial basis function neural network (RBFNN) and support vector machine (SVM) proposed to solve the problem of denial of service attacks which is serious threat to the internet. They evaluated the performance of both the methods for binary class and multi class classification [4]. In [5] Srinivas Mukkamala et. al. compared performances of SVM and neural network. They observed that both SVM and neural network have accurate results (greater than 99% for testing set).

Venkata et al. said that feature selection also necessary to improve the efficiency. In this paper they proposed Information Gain (IG) and Triangle area based KNN for

selecting feature. They combined Greedy K-means clustering algorithm and support vector machine (SVM) classifier for detecting network attacks [6].Pingjie Tang et. al.[7] has shown comparison of TANN, TALR and TASVM. TASVM has greater accuracy than TANN and TALR.

This work analyzed different methods of attribute normalization for preprocessing the data during anomaly intrusion detection. They evaluated performance with anomaly detection algorithms, PCA (principle component analysis), k-NN (K-Nearest Neighbor) and one class SVM [8]. M.Govindarajan et. al. used k-nearest neighbor for intrusion detection [9].

III. CLASSIFICATION TECHNIQUES

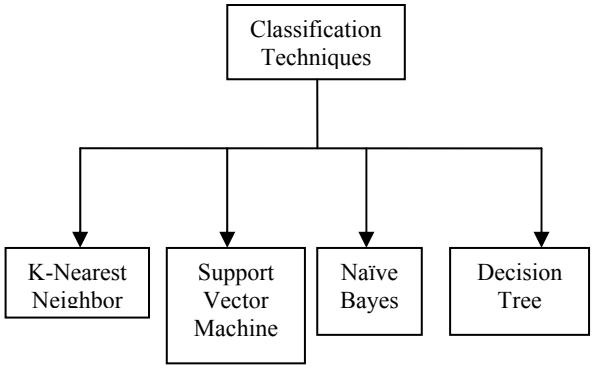

Fig.1 Classification Techniques

A. K-NN (k-Nearest Neighbor):

K-nearest neighbor is one of the simple classification techniques. It determines the distance between data points on input vectors and unlabeled data points are assigned to its nearest neighbor. In this technique k is important parameter. When the value of k is 1, then the object is assigned to class of its nearest neighbor. If value of k is large, then it takes more time for prediction and affects the accuracy by reducing effect of noise. It requires large storage for larger dataset [10].

The advantages are simple, analytically tractable. Disadvantages are large storage requirements, lazy learner. In the k-NN underlying structure of the data is not identified.

B. SVM (support vector machine):

The support vector machine is used for classification and prediction. It is supervised learning algorithm. In this method hyper plane is used to separate data points. It is binary classifier so it separates data points into two classes +1 and -1.The normal data and suspicious data is represented by +1 and -1 respectively[11]. The line which separates dataset into two classes is separating hyperplane and points closest to the hyperplane are support vectors. The following fig.2 shows how two classes are separated.

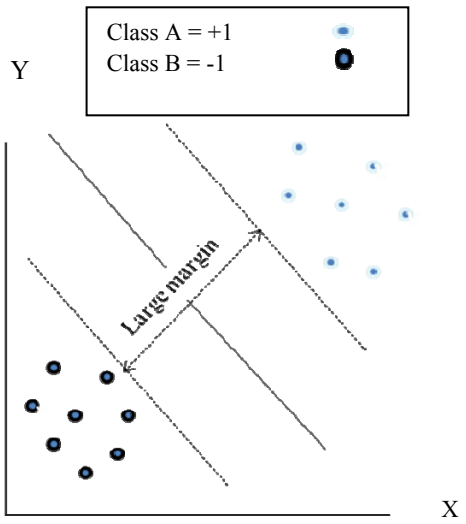

Fig.2. Two possible separating hyperplanes

The advantages are easy to interpret results, efficient. In case of two object classification, it gets superior performance. It handles only binary classification is the drawback of support vector machine.

C. Naïve Bayes:

Naïve bayes is based on probability. According to membership probability it predicts the class. It analyzes the relation between dependent and independent variable to derive the conditional probability.

The bayes theorem is:

$$P(H/X) = P(X/H).P(H)/P(X)$$

Where, X represents the data record and H is hypothesis. The prior probability is represented by P (H).The P(X/H) and P (H/X) are the posterior probability of X and H respectively, conditioned on H and X respectively [10].

The advantage is easy to implement. It may be employed on larger data points but increases the time complexity. The assumptions of conditional independence, need of probability data are disadvantages of this technique.

D. Decision Tree:

It is one of the commonly used classification techniques. Binary classification tree is created in decision tree. Each node in tree represents binary predicate on one attribute. One branch in the tree corresponds to positive instance of predicate and other corresponds to negative instance of the predicate. There are various decision tree algorithms in the literature.

Domain knowledge is not required for tree construction is advantage of decision tree. The data mining expert with little knowledge of networking can construct decision tree models. Another advantage is learned results are easy for understanding.

Disadvantages are decision tree models in case of numeric dataset are complex, decision tree algorithms are unstable [12].

IV. DATASET FEATURES

We used Kddcup99 dataset to check performance of various classification techniques. Dataset contains smurf, Neptune, teardrop, pod etc records of dos attack. The kdd dataset is modified with 110627 instances and 13 attributes. We used weka tool to compare various classification techniques. For this purpose dataset is used in ARFF format. The ARFF means attribure relation file format. Table I and II contain features and associated values. Table III. shows various features used for evaluation [13].

TABLE I
KDDCUP99 DATASET FEATURES

1	duration	22	is guest login
2	protocol_type	23	count
3	service	24	Srv count
4	flag	25	Serror rate
5	src_bytes	26	Srv serror_rate
6	dest_bytes	27	rerror rate
7	land	28	srv rerror rate
8	wrong fragments	29	same srv rate
9	urgent	30	diff srv rate
10	hot	31	srv diff host rate
11	no. of failed logins	32	dst host count
12	logged_in	33	dst host srv count
13	#compromised	34	dst host same srv rate
14	root shell	35	dst host diff srv rate
15	su attempted	36	dst host same srv port rate
16	#root	37	dst host srv diff host rate
17	#file creations	38	dst host serror rate
18	#shells	39	dst host srv serror rate
19	#access files	40	dst host rerror rate
20	#outbound cmds	41	dst host srv rerror rate
21	is hot login	42	label

TABLE II
DATASET FEATURE VALUES

0,icmp,ecr_i,SF,1032,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,53,53,0,
0,0,0,1,0,0,171,77,0.45,0.04,0.45,0,0.01,0,0.37,0,smurf
0,icmp,ecr_i,SF,1032,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,57,57,0,0,0,0,
1,0,0,255,57,0.22,0.02,0.22,0,0,0,0,0,smurf
0,tcp,http,SF,54540,8314,0,0,0,2,0,1,1,0,0,0,0,0,0,0,0,0,1,1,0,0,0,0,
1,0,0,255,255,1,0,0,0,0,0,0.05,0.05,back
0,tcp,http,SF,54540,8314,0,0,0,2,0,1,1,0,0,0,0,0,0,0,0,0,1,1,0,0,0,0,
1,0,0,255,255,1,0,0,0,0,0,0.03,0.03,back
0,tcp,http,SF,54540,8314,0,0,0,2,0,1,1,0,0,0,0,0,0,0,0,0,1,1,0,0,0,0,
1,0,0,165,165,1,0,0.01,0,0,0,0.01,0.01,back
0,tcp,private,S0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,90,15,1,1,0,0,0.17,
0.06,0,255,10,0.04,0.07,0,0,1,1,0,0,neptune
0,tcp,private,S0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,90,17,1,1,0,0,0.19,
0.06,0,255,12,0.05,0.08,0,0,1,1,0,0,neptune
0,tcp,private,S0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,90,17,1,1,0,0,0.19,
0.06,0,255,5,0.02,0.07,0,0,1,1,0,0,Neptune
0,udp,private,SF,28,0,0,3,0,0,0,0,0,0,0,0,0,0,0,0,0,0,58,58,0,0,0,0,1,
0,0,130,58,0.45,0.03,0.45,0,0,0,0,0,teardrop
0,udp,private,SF,28,0,0,3,0,0,0,0,0,0,0,0,0,0,0,0,0,0,58,58,0,0,0,0,1,
0,0,255,158,0.62,0.38,0.62,0,0.09,0,0.29,0,teardrop
0,icmp,ecr_i,SF,1480,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,4,0,0,0,0,
1,0,0.5,4,151,1,0,1,0.14,0,0,0,0,pod
.....
.....

TABLE III
SELECTED FEATURES

features	1,2,5,6,7,8,23,30,32,34,35,38,42

V. RESULTS AND DISCUSSION

Table IV Shows various classification techniques result in weka tool. For this purpose above selected features are used. All the technique have better results greater than 98%.The SVM and K-NN having higher result than naïve bayes and decision tree. Problem with naïve bayes is need of probability data.K-NN takes more time than SVM.K-NN is lazy learner. The fig.3 & 4 shows comparison of correctly and incorrectly classified instances of different classification techniques.

TABLE IV
CLASSIFICATION TECHNIQUES RESULTS IN WEKA TOOL

Technique	**function**	**Correctly Classified Instances**	**Incorrectly Classified Instances**
Support Vector machine	functions.SMO	110626 99.9991%	1 0.0009%
Naïve Bayes	bayes.NaiveBayes	110563 99.9421%	64 0.0579%
Decision Tree	Trees.Decision Stump	118617 98.1831%	2010 1.8169%
K-nearest neighbor	Lazy.IBK	110626 99.9991 %	1 0.0009 %

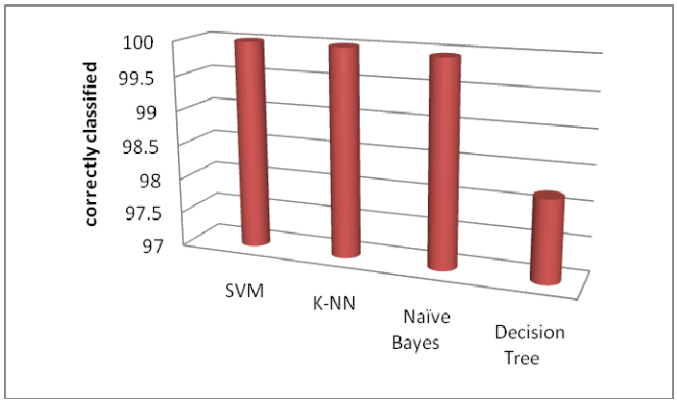

Fig. 3 Comparison of correctly classified instances of different classifiers

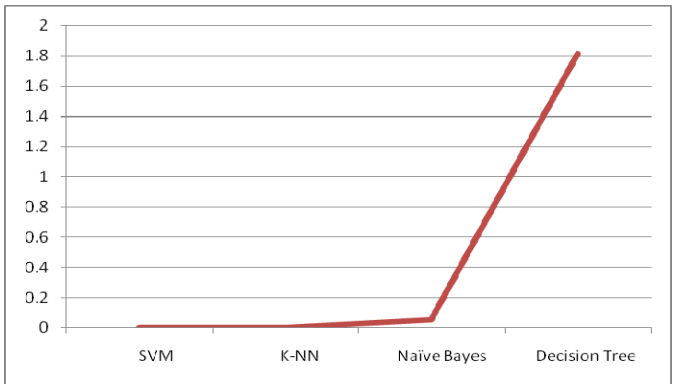

Fig.4 Comparison of incorrectly classified instances of different classifiers

VI. Conclusion

In this paper various denial of service attack types and review of various classification techniques like support vector machine, k-NN, naïve bayes and decision tree are given. From weka tool, we analysed that support vector machine and k-NN having more accuracy than all other but k-NN requires more time.

Acknowledgment

I would like to express thanks to my guide S.V Pingale and co-guide R.B. Parihar.They helped and supported for this work and family for moral support.

References

[1] Sílvia Farraposo, Laurent Gallon, Philippe Owezarski, "Network Security and DoS Attacks".

[2] "Denial of service (DoS) Attack Prevention" Edition Software Version 2.9.1 C613-03127 00 REV B.

[3] Chih-Fong Tsai and Chia-Ying Lin, "A triangle area based nearest neighbors approach to intrusion detection" Pattern Recognition, vol.43, pp. 222 – 229, (2010).

[4] Gloria C.Y. Tsang, Patrick P.K. Chan, Daneil S. Yeung and Eric C.C. Tsang "Denial of service detection by support vector machines and radial-basis function neural network" 2004 IEEE.

[5] Srinivas Mukkamala, Guadalupe Janoski and Andrew Sung, "Intrusion Detection Using Neural Networks and Support Vector Machines" 0-7803-7278-6/02/$10.00 02002 IEEE.

[6] Venkata Suneetha Takkellapati and G.V.S.N.R.V Prasad," Network Intrusion Detection system based on Feature Selection and Triangle area Support Vector Machine" International Journal of Engineering Trends and Technology-Volume3Issue4- 2012.

[7] Pingjie Tang, Rang-an Jiang and Mingwei Zhao, "Feature selection and design of intrusion detection system based on k-means and triangle area support vector machine" 978-0-7695-3940-9/10 $26.00 © 2010 IEEE.

[8] Wei Wang , Xiangliang Zhang , Sylvain Gombault , and Svein J. Knapskog , "Attribute Normalization in Network Intrusion Detection" 2009 10th International Symposium on Pervasive Systems, Algorithms, and Networks.

[9] M.Govindarajan and Rlvl.Chandrasekaran, "Intrusion Detection Using k-Nearest Neighbor" 978-1-4244-4787-9/09/$25.00 ©2009 IEEE.

[10] Abhaya, Kaushal Kumar, Ranjeeta Jha and Sumaiya Afroz, "Data Mining Techniques for Intrusion Detection: A Review" International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 6, June 2014.

[11] Mohammadreza Ektefa, Sara Memar, Fatimah Sidi and Lilly Suriani Affendey, "Intrusion Detection Using Data Mining Techniques" 978-1-4244-5651-2/10/$26.00 ©2010 IEEE.

[12] Ajayi Adebowale, Idowu S.A and Anyaehie Amarachi A." Comparative Study of Selected Data Mining Algorithms Used For Intrusion Detection" International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-3, Issue-3, July 2013.

[13] H. Güneş Kayacık, A. Nur Zincir-Heywood and Malcolm I. Heywood," Selecting Features for Intrusion Detection:A Feature Relevance Analysis on KDD 99 Intrusion Detection Datasets" Appendix 1. Description of KDD 99 Intrusion Detection Dataset Features.