IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-3, Issue-8 E-ISSN: 2347-2693

Detection of DDoS Attack Using UCLA Dataset on Different Classifiers

Aakriti Aggarwal[1*], Ankur Gupta[2]

[1*,2] Dept. of Computer Science and Engineering,
Geeta Institute Of Technology and Management, Kurukshetra, India

www.ijcseonline.org

Received: Jul /18/2015 Revised: Jul/27/2015 Accepted: Aug/20/2015 Published: Aug/30/ 2015

Abstract- Distributed denial of service attack have strong Impact on security of internet because these attacks affects the normal functioning causing loss of billions of dollars. DDoS is very harmful to network as it delays the legitimate users from excessing the server. However these networks were well equipped in security yet they were damaged by DDoS attack. In this paper, the proposed system presents both detecting and classifying schemes of DDoS attack using K-NN, SVM and Naïve Bayesian. The algorithms are developed by using various features of attack packets. By studying the incoming and outgoing network traffic and different classifiers are used to analyze these features. The main objective of this paper is to study classifiers on one dataset for DDoS attack.

Keywords: DDoS attack, Internet Securities, Attack Packet

I. INTRODUCTION

Internet is widely used by many people all over the world. As the increase in number of internet users with new and developed services, many securities attack threats have become popular. Due to this all computer systems have to suffer from securitiesdamage which are both economically costly and technologically difficult to be sold by the manufacturers [1].

The DDoS attacks usually do not exploit of security vulnerabilities of network-connected systems, but instead they aim to distort victim's services by processing the huge capacity of system or by flooding the bandwidth of the target or by scanning vulnerable hosts, such as SYN Flooding, SYN Scanning and so on. Thus intrusion detection system (IDS) plays an important role in detection of anomalities and attacks in network [2].

There are two separated steps for DDoS attack the first step is to compromise innocent systems that are accessible in the internet and install attack tools in these compromise systems. This is known as turning of computers into "zombies". In the second step attacker sends the attack command to the "zombies" through a secure channel for a launching a bandwidth attack against the target victim. In this paper we proposed the DDoS detection and classification system by using different classifiers. We evaluate the proposed scheme using UCLA dataset, which is largely used as one of the publically available datasets for network base anomaly detection system.

This paper is organized as follows: section II presents some previous related work concern DDoS attack detection. Section III describes system architecture of proposed scheme. Section IV gives the experimental evaluation and finally paper is concluded with the conclusion.

II. LITRETURE SURVEY

Karimazad and Faraahi[7] proposed and anomaly based method for detection of DDoS attack which is implemented on features of attack packets, analyzing them using Radial Bases Function(RBF) neural networks. Vectors with seven features are used for the activation of RBF neural network and classify traffic into legitimate and attack traffic classes. They used UCLA dataset for evaluation of approach.

In [3], for detection of DDoS attack, the correlation between the incoming and outgoing traffic of a network is surveyed and the changes in correlation are used. Fuzzy classifier is used in their method in order to assure the accuracy. DARPA dataset is used in this method. In [4] for the classification of traffic pattern to normal and diverse attacks combined data mining approach is used. This approach uses decision tree algorithm to select important attributes and neural networks are utilized to analyze selected attributes. In [5] proposed to determine DDoS attack signatures by analyzing the TCP/IP packet header against rules and

conditions and distinguishing the difference between attack and non-attack traffic. ICMP, TCP and UDP flooding attacks were the prime focus of the author in this paper.

In [6] uses cluster analysis method for detecting DDoS attack. A technique is proposed for proactive detection of DDoS attack by taking advantage from its architecture. Procedures on which DDoS attacks are based are mainly focused and then select variables are built on these features. After the cluster analysis detection of attack is achieved. This paper exercises with DARPA dataset.

III. SYSTEM ARCHETECTURE

The system architecture model is shown in fig.1. It can be divided into five main modules, they are Collections of packets, Preprocessing unit, Feature Extraction, Train/Test splitter, classifiers and then evaluator.

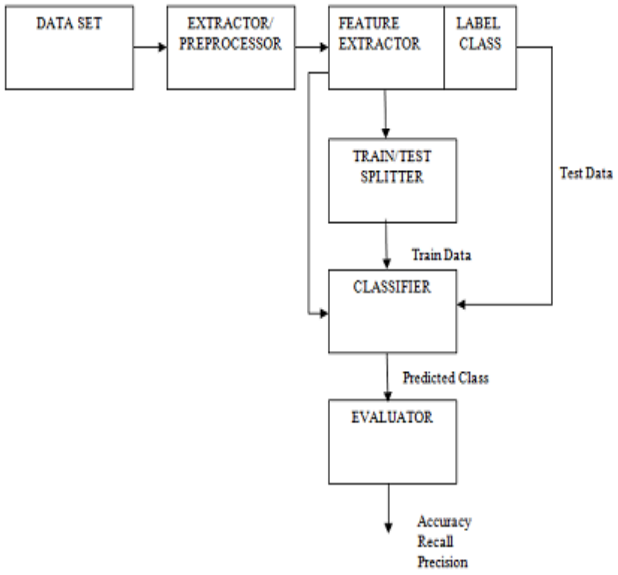

Figure 1: System Model

3.1 Packet Collection Module

The system collects the incoming and outgoing networks packets. We evaluate the proposed scheme using UCLA dataset [9], which is widely used as one of the openly available datasets for network based anomaly detection system.

3.2 Preprocessing Unit

This section preprocessed the packets and pairing of these packets are done. In DDoS attack out of huge amounts of packets communication takes place between very few packets. So, through this unit we extract those packets which are actually communicating.

3.3 Feature Extraction Module

This feature extraction module calculate the various features for DDoS detection. These features are very apt to distinguish abnormal behavior from variation of normal behavior. These features are shown below:

- Number of packets: Total number of packets from source IP to destination IP. In case of attack, the attacker sends a huge number of packets to victim system.
- Number of bytes: Total number of bytes sent from source IP to destination IP. During launch of DDoS attack number of bytes increases.
- Average packet size: It is the ratio of number of bytes to number of packets. During attack time average packet size increases.
- Packet rate: Rate of packet per second. For calculating the packet rate:
  Packet rate per second = $n_p \times 1/ (t_e - t_s)$
  n_p = number of packets
  t_e = end packet sent time
  t_s = start packets sent time
- Byte rate: Rate of packets byte per second. For calculating byte rate:
  Byte rate per second = $b_t \times 1/ (t_e - t_s)$
  b_t = total number of bytes
  t_e = end packet sent time
  t_s = start packets sent time
- Time-Interval Variance: The attacker sends attack packets at the same time span while launching of DDoS attack. So time interval variance will be closer to zero.
  For calculation of time interval variance:

(i) First, calculate the mean:

$$\bar{t} = \frac{\sum t_i}{i} \quad , \; i = 1,2,3,\ldots$$

(ii) Second, square deviation of the mean:

$$t_c^2 = \frac{\sum \left(t_n - \bar{t}\right)^2}{n}$$

c = collection number(c1,c2,c3,…)

(iii) Third, calculate time-interval variance:

$$t_c = \sqrt{\frac{\sum (t_n - \bar{t})^2}{n}}$$

- Packet-size Variance: Normal packets have different packet sizes, attack packet's size are the same. So packet size variance will be close to zero. For calculating packet size:

(i) First, calculate the mean:

$$\bar{p} = \frac{\sum p_i}{i} \quad , \quad i = 1,2,3,\ldots$$

(ii) Second, square deviation of the mean:

$$p_c^2 = \frac{\sum (p_n - \bar{p})^2}{n}$$

(iii) Third, calculate packet-size variance:

$$p_c = \sqrt{\frac{\sum (p_n - \bar{p})^2}{n}}$$

Now, according to these extracted features of our packets we will predict our classes. Here we have only two classes of packets i.e. attack packet and non-attack packet and we will label these classes.

3.4 Train/ Test Splitter

Packets with extracted features and label class is given to train test splitter. Here, percentage wise separation of train and test class is done. For example we can split hold data in 60-40 ratio i.e. 60% of train class and 40 % of test class.

3.5 Classifiers

A. Support Vector Machines:

Support Vector Machines are supervised learning method use for classification. It is usually deals with pattern classification that means this algorithms is used for classifying the different types of patterns [12]. The basic support vector machine takes a set of input data and for each given input, it predicts which of the two possible classes from the output making it non probabilistic binary linear classifier[11]. Advantage of SVM is we can apply linear classification technique to nonlinear data. Its prediction accuracy is generally high and has long training time.

B. Naïve Bayes

Naïve Bayes is a simple probabilistic classifier. It assumes that the effect of an attribute value on given class is independent of the values of the other attributes. This type of assumptions is called class conditional independence [10].The probabilities applied in Naïve Bayes algorithm are calculated using Bayes rule. The probability of Hypothesis can be calculated on the basis of Hypothesis and evidence about the Hypothesis

P(H|X) = P(X|H)P(H)/ P(X)

C. K- Nearest Neighbor Algorithm

K-nearest neighbor is a supervised learning algorithm where the result of new instant query is classified based on majority of K-nearest neighbor category. The main aim of this algorithm is to classify a new object based on attributes and training samples. A Euclidean distance major is used to calculate how close each member of the training set is to the test clash i.e. being examine [8]

Table 1 Classification Scheme

No: of pkts	Avg pkt size	Time-Interval variance	Packet size variance	Number of bytes	Packet rate per sec	Bytes Rate per second	No: of Flag packets	Class
L	L	>0	L	L	<α	L	L	Normal
H	H	<0	<0	H	>λ	H	H	Attack

Where α = minimum packet rate

λ = maximum packet rate

L = LOW, H = HIGH

3.6 Evaluator

In evaluator unit output from classifier i.e. predicted class is given as input and performance evaluation of propose system using UCLA dataset is evaluated using confusion matrix. Confusion matrix is given in the table given below.

Table 2: Confusion Matrix

	Predicted Class	
Actual Class	Positive	Negative
Positive	TP	FN
Negative	FP	TN

- True Positive (TP): When the outcome is correctly classified as positive when it is positive.
- True Negative (TN): When the outcome is correctly classified as negative when it is negative.
- False Positive (FP): When the outcome is incorrectly classified as positive when it is negative.
- False Negative (FN): When the outcome is incorrectly classified as negative when it is positive.

IV. EXPERIMENTAL EVALUATION

1. Accuracy - It is the effectiveness of the classifier by its percentage of correct prediction. Fig 2 shows the accuracy.

Accuracy = (TP+TN)/TP+TN+FP+FN

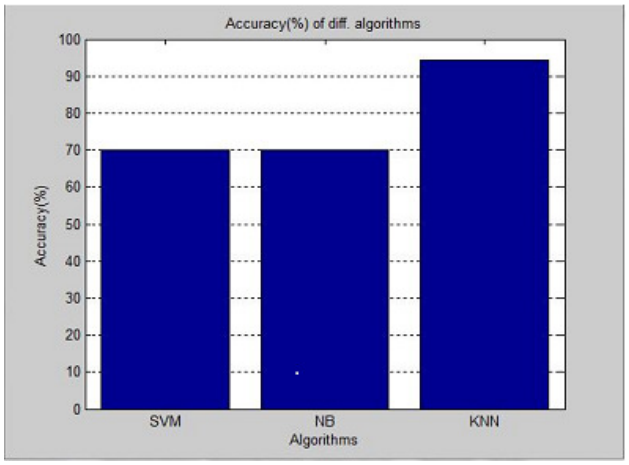

Figure 2: Accuracy

2. Sensitivity/Recall – It is the proportion of actual positive cases which are correctly identified.

Sensitivity = TP/TP+FN

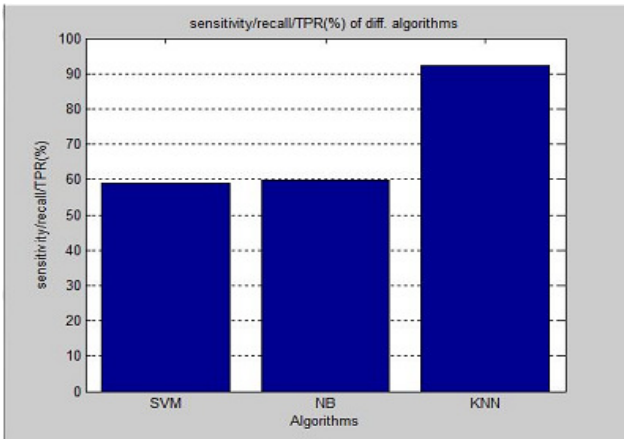

Figure 3: Sensitivity

3. Specificity – It is the proportion of actual negative cases which are correctly identified.

Specificity = TN/TN+FP

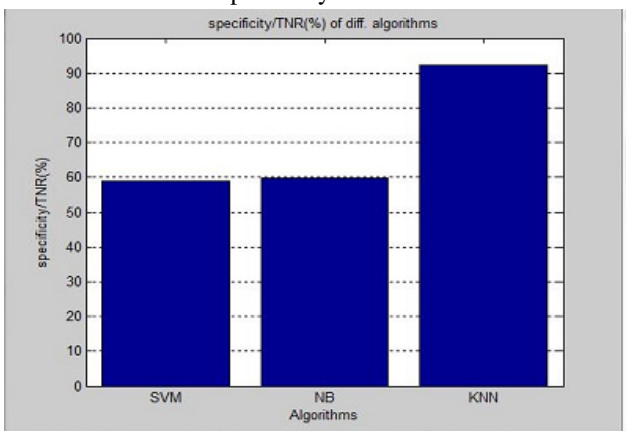

Figure 4: Specificity

4. Precision – it is the proportion of positive cases that were correctly identified.

Precision = TP/TP+FP

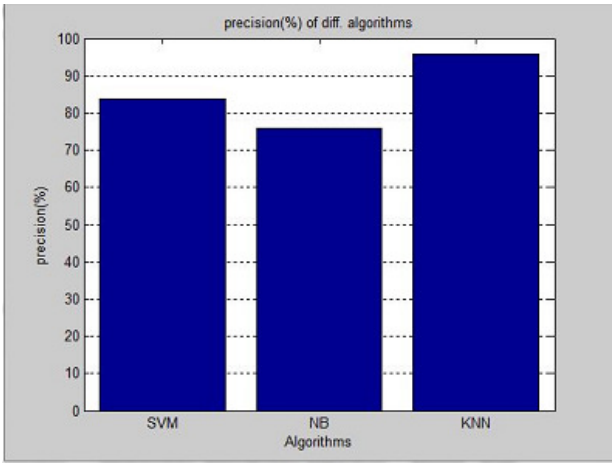

Figure 5: Precision

5. F- measure – It is the harmonic mean of precision and sensitivity.

FM = 2*(Precision*Recall)/(Precision + Recall)

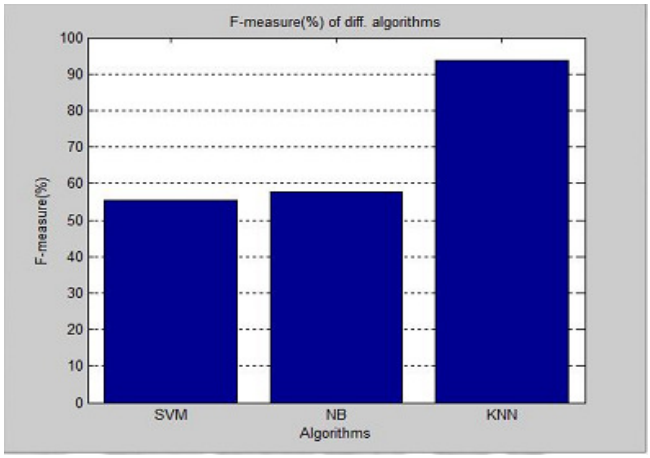

Figure6: F- Measure

6. Time complexity – It is the time taken by different algorithms to process the data.

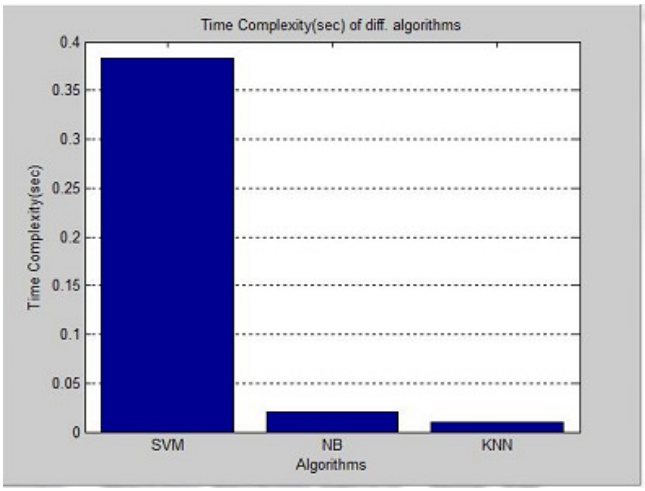

Figure 7: Time Complexity

CONCLUSION

This paper concludes a system that analyzes the network traffic and classifies the network traffic packet as normal and attack packet. Particularly we proposed a combined approach to detect normal and DDoS attack packets in traffic and then different data mining algorithms are applied to it. All this method comprises of two main steps firstly pairing of packets is performed and then different features of these packets are extracted. These features are examine to distinguish abnormal behavior of packets from variation of normal behavior of packets. Lastly different classifier algorithms like Naïve Bayes, SVM and KNN are applied on the data to study which algorithm is giving the best result, KNN gives the best results in terms of accuracy, time taken and many other features. The proposed technique is evaluated using UCLA dataset which is widely used and one of the few publically available datasets for network based anomaly detection system. The result shows that out of three data mining algorithm KNN gives best result with 94% accuracy and 96% precision in classifying the attack packets from non-attack packets.

REFERENCES

[1]. H. F. Lipson, "Tracking and Tracing" Cyber Attacks: Technical Challenges and Global Policy Issues", CERT Coordination Centre, Special Report: CMU/SEI-2002-SR-009, **2002**

[2]. N. Stephen and N. Judy, Network Intrusion Detection, 3nd ed., New Riders Publishing, United States of America, **2002**.

[3]. A. D. BasheerNayef, "Mitigation and traceback countermeasures for DDoS attacks", Iowa State University, **2005**. .

[4]. Chen, Y. Hwang, K., W. S. Ku, "Distributed change-point detection of DDoS attacks over multiple network domains." Proceedings of the IEEE International Symposium on Collaborative Technologies and Systems, Las Vegas, NV, 14-17 May. IEEE CS, **2006**, pp. 543–550.

[5]. L. Limwiwatkul, A. Rungsawang, " Distributed denial of service detection using TCP/IP header and traffic measurement analysis", Proceedings of the IEEE International Symposium Communications and Information Technology, Sapporo, Japan, 26-29 October, IEEE CS, **2006**, , pp. 605–610.

[6]. Lee, Juhyun Kim, Ki Hoon Kwon, Younggoo Han, Sehun Kim, "DDoS attack detection method using Cluster analysis", Expert System with Applications 34, **2008**, pp.1659-1665.

[7]. K. Reyhaneh, F. Ahmad, "An Anomaly-Based Method for DDoS Attacks Detection using RBF Neural Networks", International Conference on Network and Electronics Engineering IPCST vol.11, **2011**, IACSIT Press, Singapore.

[8]. Cristóbal Romero, Sebastián Ventura, Pedro G. Espejo and César Hervás," Data Mining Algorithms to Classify Students".

[9]. UCLA CSD packet traces.http://www.lasr.cs.ucla.edu/ddos/traces/public/usc.

[10]. N. Abirami, T. Kamalakannan and Dr. A. Muthukumaravel ," A Study on Analysis of Various Data mining Classification Techniques on Healthcare Data" International Journal of Emerging Technology and Advanced Engineering, Volume 3, Issue 7, July **2013**,pp.604-607.

[11]. K. Wisaeng,"A Comparison of Different Classification Techniques for Bank Direct Marketing", International Journal of Soft Computing and Engineering, Volume-3, Issue-4, September **2013**, pp. 116-119.

[12]. S. Archana and Dr. K. Elangovan, "Survey of Classification Techniques in Data Mining", International Journal of Computer Science and Mobile Applications, Vol.2 Issue. 2, February-**2014**, pg. 65-71.