Volume 6, Issue 6, June 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

DDOS Attack Detection and Prevention Using Ensemble Classifier (Random Forest)

Alpna
Mtech Student,
Computer Science Department,
University Institute of Engineering and Technology,
Kurukshetra University, Haryana, India

Dr. Sona Malhotra
Assistant Professor,
Computer Science Department,
University Institute of Engineering and Technology,
Kurukshetra University, Haryana, India

Abstract: Distributed denials of service attack (DDOS) have strong impact on the cyber world. This cyber attack halts the normal functioning of the organization by IP spoofing, bandwidth overflow, consuming memory resources etc and causes a huge loss. In this paper we are using the classification scheme based on extraction features using the UCLA dataset. Then we are using random forest technique for classification to detect and prevent the attack. The result shows the highest accuracy with less error rate as compared to other classifiers.

Keywords: DDOS, proposed algorithm, random forest, experimental results.

I. INTRODUCTION

Now days the internet become the basic need of the society. As the use of internet increases the need of security also increases. DDOS is the main threat attack that causes interruption in the user's network. The intrusion detection system (IDS) is used to detect the anomalies and threats in the network.

Denial-of-service (DoS) attack is an attempt which makes a machine or network resource unavailable to its intended users, such as to temporarily or indefinitely interrupt or suspend services of a host connected to the Internet. Denial of service attacks are designed to consume resources so that other users are unable to use the resources and are therefore "denied service". In a computer network environment, the key resources are CPU, memory, and bandwidth. The combination of multiple machines to launch a Denial-of-Service attack, this becomes a Distributed Denial of Service (DDoS) attack. DDoS means when the source of the attack is not coming from a single source, but multiple source.

DDoS attacks can be divided into three types:

Volume Based Attacks**:** these attacks include UDP floods, ICMP floods, and other spoofed-packet floods. The attack's goal is to saturate the bandwidth of the attacked site, and magnitude is measured in bits per second (Bps).

Protocol Attacks: this Includes SYN floods, fragmented packet attacks, Ping of Death, Smurf DDoS. This attack consumes actual server resources, or those of intermediate communication equipment, such as firewalls and load balancers, and is measured in Packets per second.

Application Layer Attacks: this includes low-and-slow attacks, GET/POST floods, attacks that target Apache, Windows or OpenBSD vulnerabilities and more. Comprised of seemingly legitimate and innocent requests, the goal of these attacks is to crash the web server, and the magnitude is measured in Requests per second.

II. RELATED WORK

This section describes the previous work to detect the attacks.

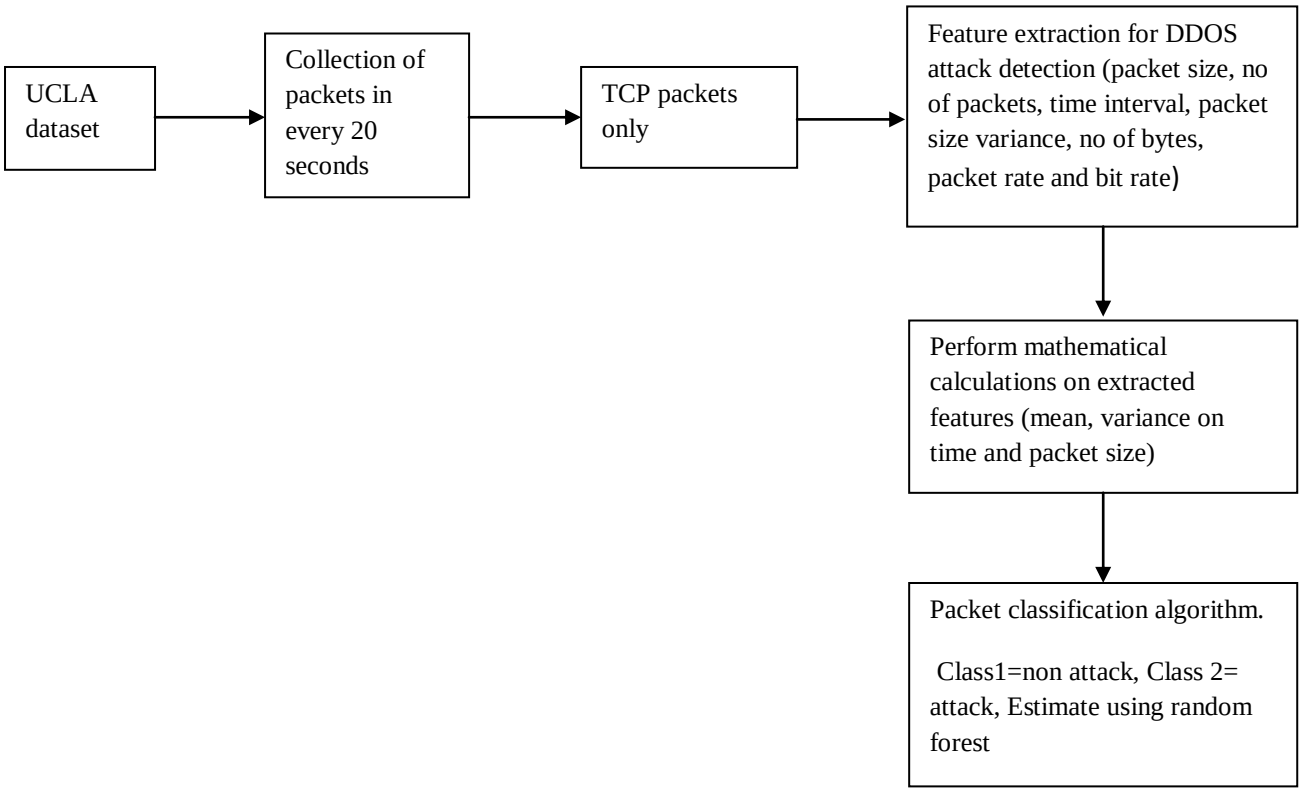
An ant based framework is used that exploits the significance of state full and stateless signatures and preserving the legitimate packets only, thereby discarding the contaminated packets [6]. A.Prathap, R. Sailaja proposed a system that presents the trend detection methodology to detect a DDOS at its early propagation stage, they propose a domain based approach the mechanism that combines stateful and stateless signature to provide early detection[3]. Mitko bogdanoski presents the two anomaly detection algorithms as an effective mechanism against DDOS. Adaptive threshold algorithm measures the network traffic and compares it with previously defined threshold. CUSUM cumulative sum algorithm based on change point detection techniques based on hypothesis testing, i.e. The rising and falling of edges at anomalous period [4].An efficient method to detecting and mitigation against TCP SYN flooding attacks using Three Counters Algorithm, which detects spoofed IP packets up to 80%. They use three counting filters to record related information: a) C-1: to record the first SYN packets of each connection b) C-2: to record the SYN packets, whose connections have completed the three-way handshake? c) C-3: to record the other SYN packets [1]. Optimized HCF Filtering and window matrix techniques are to detect Distributed Denial of Service attack. The algorithm says that the packets are legitimate IP packets and spoofed IP packets along with their IP addresses. Based on this result, they conclude to accept or discard the packets [7]. Karimazad and Faraahi [2] proposed an anomaly based DDoS detection method based on features of attack packets, obtained from study the incoming network traffic and analyzing them using Radial Basis Function (RBF) neural networks. Vectors with seven features are used to activate an RBF neural network and classify traffic into normal and

DDoS attack traffic classes. They evaluated the approach using UCLA Datasets. Their system can be classified either normal or attack, but that can't be classified and identified what types of attacks. Thew Thew Oo and Thandar Phyu [5] proposed algorithms are developed based on various features of attack packets from incoming and outgoing network traffic using K-Nearest neighbor to analyze these features. System is a combined data mining approach to detect protocol anomaly against DDoS attack. In this paper, traffic features are extracted from network traffic and then it is clustered into normal and attack traffic by using a data mining classification algorithm. This paper only shows detection phrase using only proposed algorithm.

Now, in this paper, the proposed system is statistical detection of attacks using the two proposed algorithms and random forest. The results show that our system can detect and classify types of DDoS attack (attack/ non attack) .

III. PROPOSED WORK

System architecture:



A.) Feature Extraction Module

In this section we extract the various features that can be used to detect the DDOS attack. These features are reliable to differentiate between the normal (non attack) and abnormal (attack) packets. We choose to use these features as he suggested because the analysis of traffic based on these features can recognize the attack in packets. These features are shown in the following:

1) Average Packet Size: DDoS attacks floods traffic to the victim's system to consume system resources, thus causes the increase in the average packet size at attack time. So this feature can use to identify DDoS attacks.
2) Number of Packets: DDoS attack send the large number of packets to the victim network. Therefore, the number of packets increases as compared to normal case.
3) Time Interval Variance: The experiments show when DDoS attack launches, agents send attack packets in the same time span. Then we can detect this attacks using Time Interval Variance. Whenever packet sending time spans are more similar together, Time Interval

Variance will be closer to zero. Variance can be calculated through (1):

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

4) Packet Size Variance: According to our studies, we found that attack packets sizes are the same. However, normal packets have different sizes even when they belong to the same file. DDoS packets can be identified by using the Packet Size Variance.
5) Number of Bytes: as the no of bytes increases the chances to launch DDOS attack increases.
6) Packet Rate: This feature shows the packet rate sent from a source address to a destination in a specific time span. Packet rate increases significantly in attack time.

$$n_p \times \frac{1}{(t_e - t_s)}$$

Where *np* stands for number of packets, *te* is last packet sent time and *ts* is first packet sent time.

7) Bit rate: as the bit rate is very high the chances of DDoS attack also high.

$$b_t \times \frac{1}{(t_e - t_s)}$$

Where *bt* is total number of bytes, *te* is last packet sent time and *ts* is first packet sent time

Along with the above mentioned features, the system also uses the number of SYN, ACK and FIN flags packets to classify attack types such as SYN Flooding, ACK Scanning.

B.) Packet classification algorithm:

After the feature extraction the classification algorithm is applied to classify the normal packets and the attacked packets. Our aim is to check the flow and classify the packets as attack or non attack on the basis of threshold. If the value is more than threshold then we say that the packet is abnormal and we classify it as attack packet, but if the value is less than the threshold then we say the packet is normal. In this system we observe only TCP packets. We have classification table also:

No of packets	Low	High
Average packet size	Low	High
Time interval variance	>0	<0
Packet size variance	low	<0
Packet rate per second	<α	>α
Byte rate per second	low	High
No of flag packets	low	High
Class	Normal(non attack)	Attack

X=low, y= high
α = threshold value.

C.) Random forest algorithm:

Random forest is an ensemble learning method for classification and regression operate by constructing a magnitude of decision tree. It consists of many decision trees and develops lot of decision tree based on random selection of data and selection of variables. We assumes that the user knows about the construction of single classification trees. Random Forests grows many classification trees. To classify a new object from an input vector, put the input vector down each of the trees in the forest. Each tree gives a classification, and we say the tree "votes" for that class. The forest chooses the classification having the most votes (over all the trees in the forest).

Each tree is grown as follows:

1. If the number of cases in the training set is N, sample N cases at random - but *with replacement*, from the original data. This sample will be the training set for growing the tree.
2. If there are M input variables, a number m<<M is specified such that at each node, m variables are selected at random out of the M and the best split on these m is used to split the node. The value of m is held constant during the forest growing.
3. Each tree is grown to the largest extent possible. There is no pruning.

Here we compare our algorithm with KNN (K Nearest node) [5].

IV. EVALUATION OF ATTACK DETECTION

The performance evaluation of proposed system using UCLA dataset is evaluated using the classification accuracy rate, recall, precision and F-measure with matrix confusion,

Here we have TP= True positive TN= True negative FP= False positive FN= False negative

Following formulas are used to evaluate system by recall precision and F- measure.

I) Recall means how many selected items are relevant.

$$\text{Recall} = \frac{TP}{TP + FN}$$

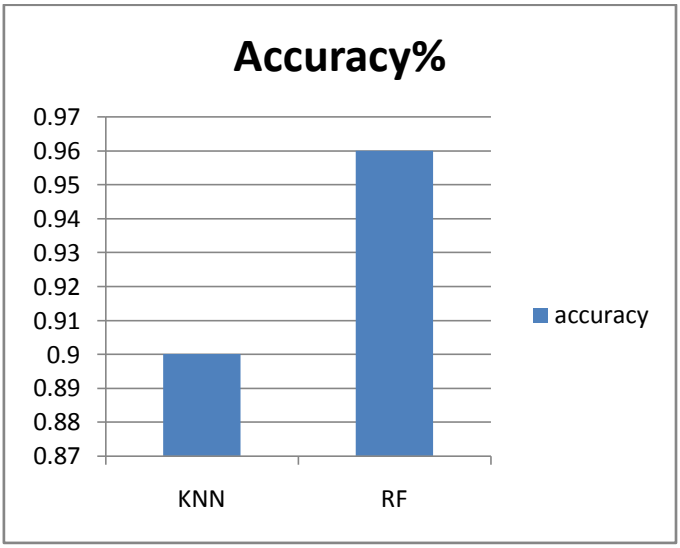
II) Precision means how many relevant items are selected.

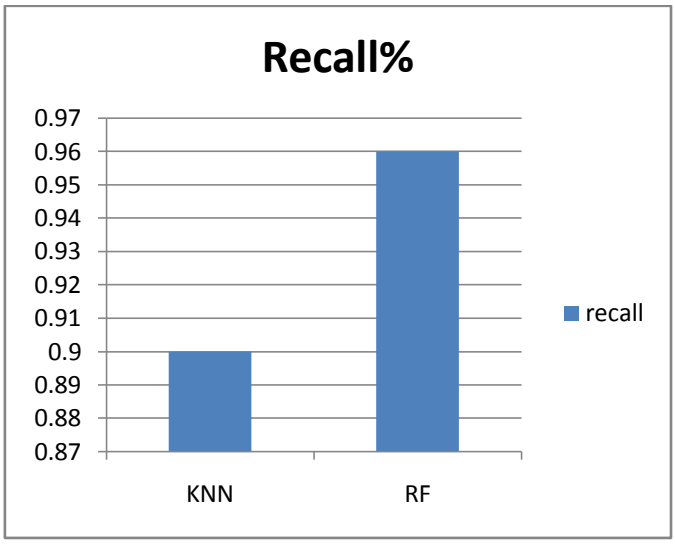
$$\text{Precision} = \frac{TP}{TP + FP}$$

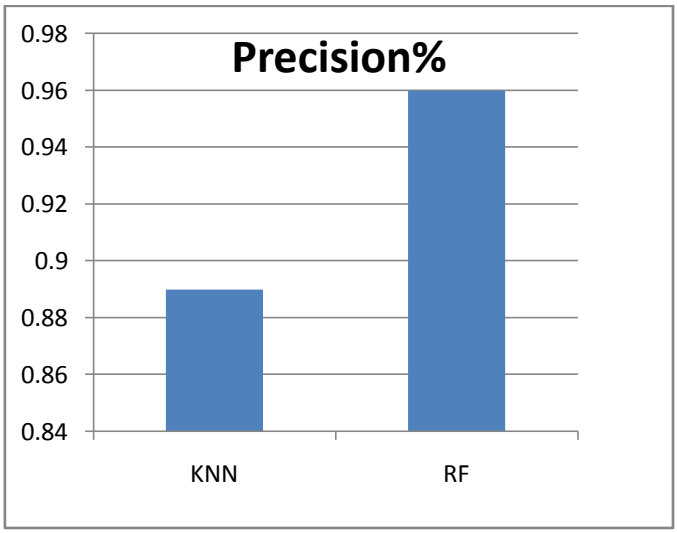
III) F –measure is combination of recall and precision.

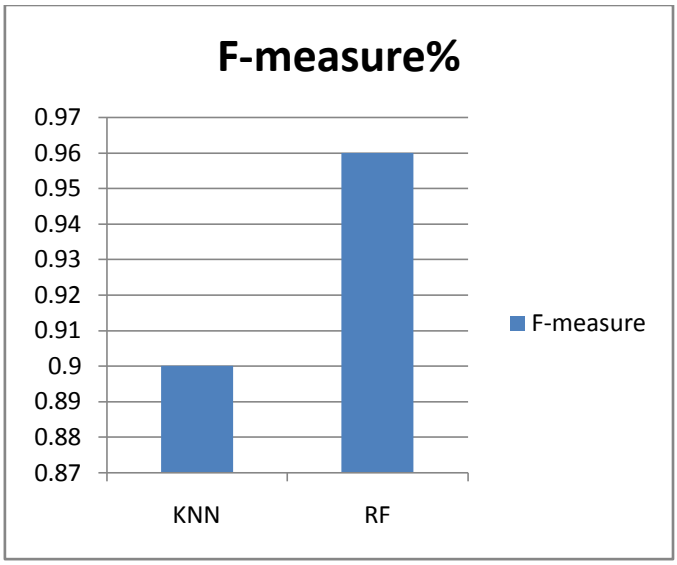
$$\text{F-measure} = \frac{2*\text{recall}*\text{precision}}{\text{Recall}+\text{precision}}$$

Algorithm	Accuracy	precision	recall	F-measure
KNN	0.92	0.89	0.90	0.9
Random forest	0.97	0.96	0.96	0.96









V. CONCLUSION

DDOS is the main threat of cyber world and cannot be detected easily. In this paper we have shown the architecture how to detect the DDOS attack with the help of our proposed algorithm and random forest algorithm for classification of packets i.e. normal or attack. We compare the KNN algorithm with our random forest algorithm and we find the higher accuracy and less error.

ACKNOWLEDGEMENT

This task would be incomplete without the mention of people whose cooperation made it possible, whose constant guidance and encouragement crown all efforts with success. I own my regards *to Dr. Sona Malhotra (ASTT PROFFESOR,CSE DEPT, UIET KURUKSHETRA,KURUKSHETRA UNIVERSITY)* my Guide, for reviewing, advising, suggestion, motivation and extended keen interest

REFERENCES

[1] S.Gavaskar, R.Surendiran, Dr.E.Ramaraj, *Three Counter Defense Mechanism for TCP SYN Flooding Attacks*, International Journal of Computer Applications (0975 – 8887) Volume 6– No.6, September 2010

[2] Reyhaneh Karimazad, Ahmad Faraahi, *An Anomaly-Based Method for DDoS Attacks Detection using RBF Neural Networks*, 2011 International Conference on Network and Electronics Engineering IPCSIT vol.11 (2011) © (2011) IACSIT Press, Singapore

[3] A.Prathap, R.Sailija," *detection and prevention of denial of service attack using distributed denial of service detection mechanism*", international journal of computer science and information technologies, 2012, ISSN: 09759646.

[4] Mitko Bogdanoski, Tomislav Shuminoski, Aleksandar Risteski, *Analysis of SYN Flood DOS attack,* I.J. Computer network and information security, June 2013, 8, 1-11

[5] Thwe Thwe Oo, Thandar Phyu, *Statistical Anomaly Detection of DDoS Attacks UsingK-Nearest Neighbor*, International Journal of Computer & Communication Engineering Research (IJCCER) Volume2 - Issue 1 January 2014.

[6] Dimple Juneja, Neha Arora, *An Ant Based Framework for Preventing DDoS Attack in Wireless Sensor Networks*, International Journal of Advancements in Technology, ISSN 0976-4860.

[7] G. Usha Devi*, M. K. Priyan, E. Vishnu Balan, C. Gokul Nath and M. Chandrasekhar, *Detection of DDoS Attack using Optimized Hop Count Filtering Technique*, Indian Journal of Science and Technology, Vol 8(26), DOI: 10.17485/ijst/2015/v8i26/83981, October 2015.

[8] https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm

