International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 12 | Dec 2018 www.irjet.net p-ISSN: 2395-0072

# DDOS DETECTION SYSTEM USING C4.5 DECISION TREE ALGORITHM

**Santosh Kumar Pydipalli[1], Srikanth Kasthuri[1], Jinu S[1]**

[1] *Jr.Telecom Officer, Bharath Sanchar Nigam Limited, Bangalore*

---

**Abstract** - In today's digital world, cyber attacks became much severe and intelligent, causing enormous damage. With the evolution of IoT, securing the digital infrastructure is much essential to avoid any system collapse. Due to enhanced bandwidth availability and with advent of new technologies, hackers became more capable to throw new challenges. DDoS (Distributed Denial of Service) attack is much recurrent nowadays, that inflicts serious damage and affects the network /application performance. As hackers are finding new ways to attack the systems, zero day vulnerabilities can cost dearly. By the time, patch applied for security exploit, the systems may get compromised by attackers .So it's essential to develop intelligence to detect and mitigate complex security attacks. In the present paper, we proposed a model that can combine the strengths of both signature based DDoS detection and anomaly based DDoS detection. We designed a machine learning model that can learn from attack patterns and alert the security analyst. In this model, DDoS attack sample is extracted from Intrusion Detection Evaluation Dataset (CICIDS2017) and divided the sample into training and test data. We used Weka 3.8.2, a java based data mining and machine learning tool for building the model. Pre-processing is done with Weka supervised attribute filter and classification of training set is performed using J48 / C4.5 decision tree algorithm, with 10-fold cross-validation. The accuracy of Machine learning model is validated using test dataset. This model, coupled with signature detection techniques, can perform automatic and effective differentiation between benign traffic and DDoS flooding.

**Key Words: DDoS, Weka, C 4.5, Decision Tree.**

## 1. INTRODUCTION

There are numerous attacks in present day world on the network and server resources. Distributed Denial of Service (DDoS) attacks are attacks targeting the availability of network and Server resources there by causing service obstruction to the legitimate users and performance degradation of resources. These attacks can easily be launched on machines present within the network or on the machines present in the different network. Based on the attacker and victim present in a network they can be classified into 2 types.

**Inbound /Outbound attacks**: These are the attacks which originate in a network are targeting a particular user/users present in a completely different network.

**Cross bound attacks:** These are the attacks which originate in a network and are targeting a user/users present in the same network.

### Distributed DoS attacks (DDoS):

Denial of service attacks (DOS) will prevent the legitimate user from accessing resources. Basically these DoS attacks are classified mainly into 2 types Network level attacks and Application level attacks. Network Level attacks will stop the genuine user/users from accessing network resources and Application level attacks will disable the users from accessing the services by consuming the server resources. These attacks are wide spread in present day digital world. In DDoS attacks instead of using a single machine the attacker uses multiple systems to flood the bandwidth or resources of a targeted system, generally one or more web servers.

During a DDoS attack, Hosts or Bots coming from distributed sources overwhelm the target with illegitimate traffic so that the available resources (of Server or any Network) cannot respond to genuine clients. These DDoS attacks mainly classified as Volume Based attacks, Protocol attacks, application layer attacks, computational attacks, Vulnerability attacks etc.

### 1.1 Volumetric DDoS attacks:

Volumetric DDoS attacks are designed to saturate and overwhelm network resources, circuits etc by brute force. According to M/s. Arbor Networks, 65% of DDoS attacks are volumetric in nature.

Common attacks: DNS Amplification, NTP Amplification, UDP Flood.

### 1.1.1 DNS Amplification:

A DNS amplification attack is a most advanced type of DDoS attack. In this type of attack, attacker will send large amounts of incoming data to a server by exploiting the vulnerabilities of DNS server there by making the server and its resources inaccessible.

In this attack a large number of DNS request are send by spoofing the IP address to one or more DNS servers. Depending on the configuration these DNS servers will start responding to those IP addresses from which the request is originated. This will ultimately load the victim's servers.

### 1.1.2 NTP Amplification:

These attacks work in similar to DNS attacks. In this the attacker will target the NTP servers and exploit the vulnerabilities in them to generate huge amounts of traffic. NTP is one of the network protocols, this is used by connected machines in the internet for synchronization of their clocks.

In most basic type of NTP amplification attack, the attacker will send "get monlist" request repeatedly to a NTP server while spoofing the IP address of machine. The NTP server will respond to this request by sending the list to the spoofed IP address. This response is much larger in size than the request, generating huge amounts of traffic towards the target machine and ultimately leading to degradation of services to genuine requests.

### 1.1.3 UDP Flood:

It's a type of denial of service attack in which huge number of UDP packets are send with the aim of shattering that device's ability to respond and process there by causing denial of service to genuine traffic.

### 1.2 Protocol attacks:

These type of attacks that causes a target to be inaccessible by exploiting vulnerabilities in the Layer 3 and Layer 4 protocol stack. Common types of Protocol attacks are SYN flood and Ping of Death.

### 1.2.1 SYN Flood:

This is a type of denial of service attack. In this attacker will send SYN (Synchronization) packets to every port on the server using fake IP addresses. The server will respond with SYN/ACK packets from each port. This will consume the resources of server and making it unresponsive to legitimate traffic.

### 1.2.2 Ping of Death:

It's a type of denial of service attack in which an attacker sends an ICMP packet larger than the maximum allowable size, this can crash the machine, or freeze the services offered by the machine. It's used to make the device unstable by intentionally sending large sized packets to the target device over an IPV4 network.

### 1.3 Application Layer DDoS attacks:

These attacks that exploit vulnerability in Layer 7 Protocol stack. These are the most sophisticated ones and very difficult to identify and mitigate.

Common types of Application layer attacks are HTTP Flood attacks and attack on DNS services.

### 1.3.1 HTTP Flood:

HTTP flood attack is a volumetric DDoS attack basically destined to shatter a target server with HTTP requests. This will consume the available server resources making it unresponsive/unreachable to actual genuine requests.
This is a layer7 DDoS attack in which the attacker uses typical HTTP GET/POST methods to fetch information from a server or application. Unlike other DDoS attacks these attacks does not uses malformed packets or IP spoofing. This makes it difficult to detect these attacks by advanced detection systems. This is because this traffic characteristic resembles genuine traffic pattern.

### 1.3.2 DNS Flood:

The attacker targets multiple Domain Name System (DNS) servers in a particular area, thereby hampering the resolution of resource records of that area or sub-areas. DNS flood attacks normally uses the high bandwidth connectivity of bots to make the DNS servers respond to a large number of DNS requests. In a multi level DNS flood attack, the Source IP address will be a spoofed one, so that the large number of DNS replies will be sent to that spoofed IP address, which is the intended target.

### 1.4 Conventional Intrusion Detection Techniques:

### 1.4.1 Signature Based Detection Technique:

This detection is known as misuse technique. This will compare the "known patterns" of detrimental activity with the already present signatures stored in the database. This technique is very accurate and it's only work with the Known attacks. This technique is very fast in comparison to other detection techniques because all it has to do is to look up for list of known signatures of attacks and if it finds a match it will report this. A commonly used tool in signature detection is SNORT tool. On the negative side if someone starts a new attack there will not be any protection because it will not find a match with the signatures already present. In this case similar to Anti-virus, once a new attack is recorded the database needs to be updated.

### 1.4.2. Anomaly Based Detection Technique:

Unlike, Signature based detection, Anomaly based detection doesn't have a information regarding well known attacks. This technique make use of the current event pattern & determine the anomalies in that pattern. Shannon-Wiener's index theory analyzes random data patterns and point outs uncertainty in the current pattern. Entropy is the

measure of abnormality and randomness. If the samples taken to analyze are from a single group, then entropy is observed to be lesser in comparison with patterns taken from multiple groups. Headers present in the sampled data are analyzed to determine the IP and ports before computing their entropy. The Detection systems are made in such a way that it detects DDOS attack, in case the pattern under observation exceeds the entropy by a fixed threshold.

## 2. RELATED WORKS

Machine learning techniques are frequently used for anomaly detection. They have received considerable attention among the intrusion detection researchers to address the weaknesses of knowledge based detection techniques. Another experiment developed on three intrusion detection models based on Multi-Layer Perceptron (MLP), C 4.5, and SVM classifiers showed that C 4.5 is the best method in terms of detection accuracy and minimum training time [2]; it achieved the accuracy rate of (99.05%). For this reason, we choose the C4.5 algorithm to detect the DDoS attacks in our proposed model. We compared the output accuracy with Naivebayes algorithm.

The model proposed using C 4.5 decision tree algorithm can provide encouraging results, when combined with signature based DDOS detection systems.

## C 4.5 Algorithm:

C 4.5 is extension of ID3 algorithm. It tries to find simple and small decision trees. C 4.5 builds decision trees from a set of training data on basis of information entropy.

$$\text{Entropy } (\vec{x}) = -\sum_{k=1}^{n} \frac{|x_k|}{|\vec{x}|} \log \frac{|x_k|}{|\vec{x}|}$$

Iterating over all possible values of $(\vec{x})$, the conditional probability:

$$Entropy\ (k/\vec{x}) = \frac{|x_k|}{|\vec{x}|} \log \frac{|x_k|}{|\vec{x}|}$$

It uses the fact that each attribute of the data can be used to make a decision that splits the data into smaller subsets. C4.5 examines the normalized information gain ratio that results from choosing an attribute for splitting the data.

The attribute with the highest information gain ratio is the one used to make the decision.[4]

Given a learning set S and a non class attribute X, the Gain Ratio is defined as:

$$IGR(S|X) = \frac{IG(S|X)}{-\sum_{i'} \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|}}$$

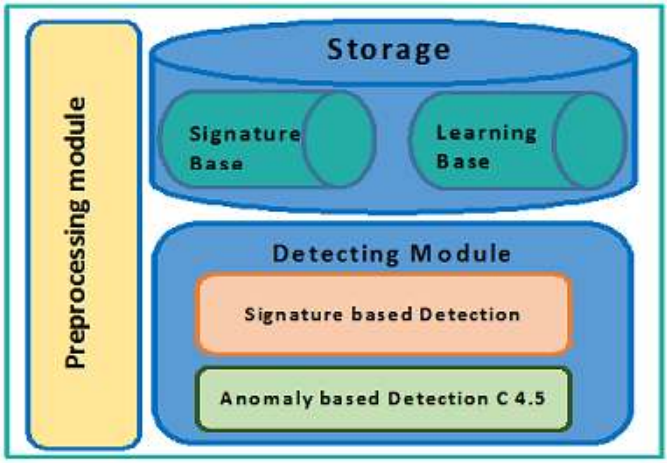



Figure1: Combined model for DDOS Detection

### 2.1. Input DDoS attack Data

The input DDoS sample is taken from Intrusion Detection Evaluation Dataset (CICIDS2017)[1]. It contains benign and the most up-to-date common attacks, which resembles the true real-world data. The data is captured over a period of five days with diversified attacks. Low Orbit Ion Canon (LOIC) tool is used for DDoS attack generation. The incidents are more than 2 million, making the dataset comprehensive.

As this dataset is huge, we choose random subset from the available data. The experiment is repeated with random input, so that the results are not biased.

### 2.2. Pre-processing and supervised filtering

The input dataset contains a total of 78 attributes. Pre-processing is done in Weka with supervised attribute filter. This will ensure greater accuracy and less processing time for building the model.

Key Chosen attributes of dataset:

- Maximum Packet length in Forward Direction
- Total Length of packet in Forward direction
- Total Length of packet in Backward direction
- Average packet size
- Destination port

As a standard practice, training and test datasets are divided in 70:30 ratio. This dataset is taken randomly from original data and experiment is repeated to validate the consistency.

### 2.3. Classification Using Model

We have used C4.5 algorithm for building this model, as it proved to be more efficient in detection of DDoS attacks. C4.5 algorithm is used to generate a decision tree developed by Ross Quinlan.

J48 is an open source Java implementation of the C4.5 algorithm in Weka. 10-fold Cross validation is done and the trained model is validated with test dataset.

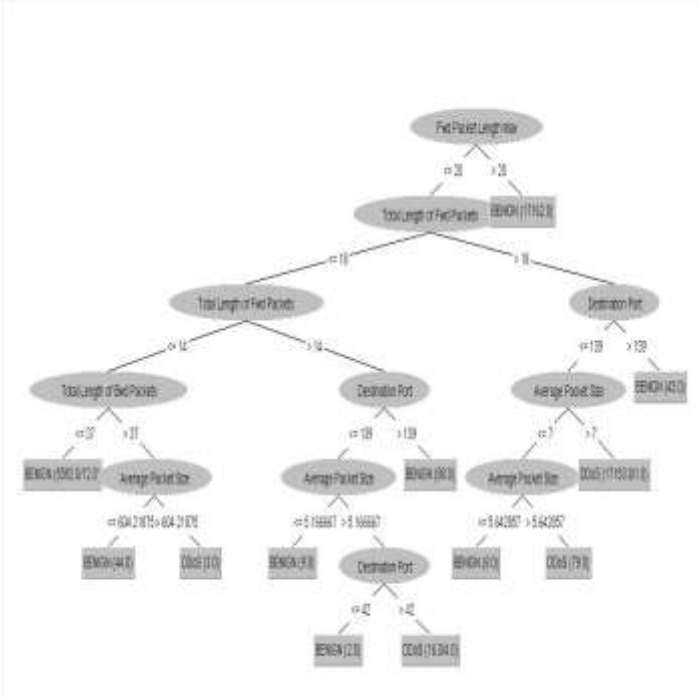


Figure 2. Decision Tree using C 4.5 Algorithm

```
=== Summary ===

Correctly Classified Instances       29979               99.9333 %
Incorrectly Classified Instances        20                0.0667 %
Kappa statistic                          0.9986
Mean absolute error                      0.0009
Root mean squared error                  0.0256
Relative absolute error                  0.1793 %
Root relative squared error              5.1723 %
Total Number of Instances            29999

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   MCC      ROC Area  PRC Area  Class
                 0.999    0.001     0.999      0.999     0.999     0.999     1.000     1.000    DDoS
                 0.999    0.001     0.999      0.999     0.999     0.999     1.000     1.000    BENIGN
Weighted Avg.    0.999    0.001     0.999      0.999     0.999     0.999     1.000     1.000

=== Confusion Matrix ===

     a     b   <-- classified as
 17064    11 |     a = DDoS
     9 12915 |     b = BENIGN
```

Figure3. Weka Output for C4.5 model training dataset

```
=== Summary ===

Correctly Classified Instances       12990               99.9308 %
Incorrectly Classified Instances         9                0.0692 %
Kappa statistic                          0.9986
Mean absolute error                      0.001
Root mean squared error                  0.0262
Total Number of Instances            12999

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   MCC      ROC Area  PRC Area  Class
                 1.000    0.001     0.999      1.000     0.999     0.999     0.999     0.999    DDoS
                 0.999    0.000     0.999      0.999     0.999     0.999     0.999     0.999    BENIGN
Weighted Avg.    0.999    0.001     0.999      0.999     0.999     0.999     0.999     0.999

=== Confusion Matrix ===

     a     b   <-- classified as
  7386     3 |     a = DDoS
     6  5604 |     b = BENIGN
```

Figure4. Weka Output for C4.5 model test dataset

### Confusion Matrix:

A confusion matrix is the summary of prediction results on a classification problem. The number of correct and incorrect predictions are summarized with count values and broken down by each class.

Confusion matrix is shown in Table 1 which is the basis for checking the accuracy and credibility of the proposed model.

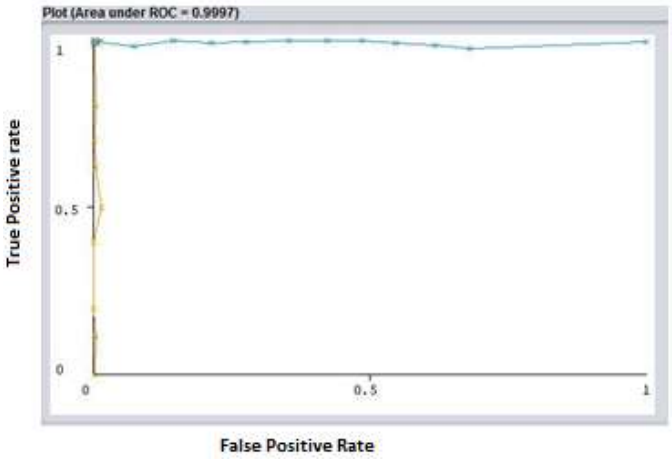
Tabel1 .Confusion matrix for C 4.5 Algorithm Model

| Type of Dataset | DDOS Packets | BENIGN Packets |
|---|---|---|
| Training Dataset | 17064 | 11 |
| | 9 | 12915 |
| Test Dataset | 7386 | 3 |
| | 6 | 5604 |

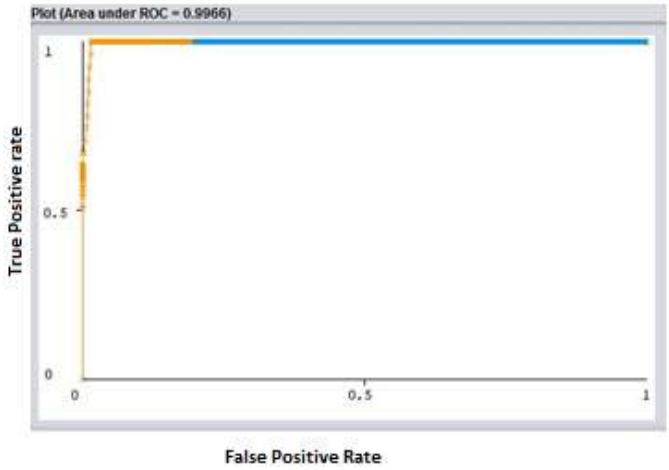
To confirm the accuracy, we conducted the same experiment with NaiveBayes classification algorithm. A comparison chart is made as follows:

Tabel2. Comparison Chart between NavieBayes and C 4.5

| Parameter | C 4.5 | NaiveBayes |
|---|---|---|
| Accuracy | 99.93% | 91.64% |
| Kappa Statistic | 0.9988 | 0.8255 |
| Mean Absolute Error | 0.009 | 0.081 |
| Root Mean Square Error | 0.0228 | 0.2813 |
| Time taken to build model | 0.4 sec | 0.06 sec |




Area under ROC, C.4.5 Model




Area under ROC, Naviebayes Model

## 3. CONCLUSION AND FUTURE SCOPE

C 4.5 Decision Tree algorithm provided 99.93% accuracy in differentiating DDoS attack and benign traffic. Even though, the build time is more compared to Naviebayes, the accuracy level is improved significantly with C4.5 decision tree algorithm. The attribute filtration plays a major role for high accuracy and increased build speed. Combination of signature plus anomaly based model can increase the reliability in DDoS detection. This model may be used for identification of bot participating in DDoS attack so that those IP addresses could be blocked.

Finally, as a future work, we are planning to develop a prototype system based on deep learning so that, better detection capabilities can be inherited by processing huge real-time streaming data.

## REFERENCES

[1]. Iman Sharafaldin, Arash Habibi Lashkari and Ali A. Ghorbani., "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization", 4th International Conference on Information Systems Security and Privacy (ICISSP), Purtogal, January 2018

[2] Ismanto, Heru & Wardoyo, Retantyo., Comparison of running time between C4.5 and k-nearest neighbor (k-NN) algorithm on deciding mainstay area clustering. International Journal of Advances in Intelligent Informatics. 2. 1. 10.26555/ijain.v2i1.49.

[3] Marwane Zekri, Said El Kafhali, Noureddine Aboutabit and Youssef Saadi., "DDoS Attack Detection using Machine Learning Techniques in Cloud Computing Environments", 3rd International Conference of Cloud Computing Technologies and Applications (CloudTech), October 2017

[4]. Wei Wang, 2, Sylvain Gombault, Thomas Guyet., "Towards fast detecting intrusions: using key attributes of network traffic", Third International Conference on Internet Monitoring and Protection, pp.86-91, 2008

[5]. Agrawal, S. and Rajput, R. S., "Denial of Services Attack Detection using Random Forest Classifier with Information Gain", International Journal of Engineering Development and Research, September 2017