I.J. Wireless and Microwave Technologies, **2019, 1, 52-59**
Published Online January 2019 in MECS(http://www.mecs-press.net)
DOI: 10.5815/ijwmt.2019.01.06

***Available online at** http://www.mecs-press.net/ijwmt*

DDoS Attack Detection Using C5.0 Machine Learning Algorithm

[a]Hariharan. M, [b]Abhishek H. K and [c]B. G. Prasad

[a]Department of CSE, BMS College of Engineering, Bengaluru, Karnataka - 560019, India
[bc]Department of CSE, BMS College of Engineering, Bengaluru, Karnataka - 560019, India

Received: 29 July 2018; Accepted: 22 October 2018; Published: 08 January 2019

Abstract

Distributed Denial of Service has always been an issue while dealing with network security. The potential of DDoS attacks is not limited by any security measures. This type of attack does not attempt to breach a security perimeter but aims to make the service unavailable to legitimate users. This is particularly an issue in private clouds as public clouds have sophisticated systems to prevent DDoS attacks. DDoS attacks can be used as a shield for other malicious activities. Open resource access model of the Internet is exploited by Distributed Denial of Service attackers. The main objective of this paper is to detect DDoS attacks using C5.0 machine learning algorithm and compare the results with other state of the art classifiers like Naïve Bayes classifier and C4.5 decision tree classifier. The focus is on an offline detection model.

Index Terms: Denial of Service, DDoS, Machine Learning, Decision Tree, C5.0

1. Introduction

Machine Learning, Big Data, and Cloud Computing are three technologies that are shaping up the field of computer science. They are, as well, interlinked to one another. Machine learning applications require large amounts of data and cloud-based setups are required to store and process this huge amount of data. In all the service-based applications, it is very important to ensure availability, reliability, and consistency to the customers. To ensure the above, it is necessary to have the application very secured. Different types of attacks prevent the environment from providing a good service. One such attack is the Denial of Service attack, where the attacker aims at making the resource unavailable to the legitimate users by disrupting the services of the host.

* Corresponding author. Tel.:
E-mail address: mhariharan695@gmail.com

Distributed Denial of Service (DDoS), is relatively a very powerful, yet simple means to target Internet resources [1]. They have a major impact on the computing environment because these attacks add the many-to-one dimension to the ordinary problem, in turn, making mitigation and prevention tasks difficult. Open resource access model is the greatest advantage of Internet system architecture which also turns out to be an inherent weakness. DDoS exploits this weakness.

Applications of Machine Learning (ML) are spread out in a wide range. One of the most significant applications is Classification. There are several ML classification techniques like Bayesian classification [2], Support Vector Machines [3], Decision Tree algorithm (ID3, C4.5, C5.0) [4, 5], and K-Nearest Neighbor Classifier [6]. Each of these techniques considers different aspects of the dataset and provides accurate classifications.

This paper aims at giving a solution to Distributed Denial of Service attack by detecting them using C5.0 machine learning algorithm and thereby preventing the denial of service scenarios.

Rest of the paper is organized as follows, Section II contains the Literature Survey related to DDoS Attack detection and some Machine Learning techniques, Section III describes the proposed system, Section IV gives the results and section V concludes the research work with the future scope.

2. Related Works

Zekri et al [7] focus on an efficient way in which the DDoS attack can be detected. There are various other machine learning techniques which either take more time or are less accurate. This work considers the C4.5 algorithm and it is shown that it is less accurate and takes more time to generate the decision tree.

The latest algorithm which is proven to be efficient is the C5.0 algorithm which takes less time and memory compared to C4.5. In [8] the authors have focused on the classification of the network traffic, C5.0 algorithm has been implemented and the result was better than that of the C4.5 algorithm. Another work which aimed at generating an improved decision tree with feature selection and reduced error pruning [9], uses several algorithms like ID3, C4.5 and C5.0. The results proved that C5.0 performs better than the other techniques in terms of both accuracy and memory usage. The work by [10] which focused on improving the crop yield by making predictions using crop pest data by comparing the performance of C4.5 and C5.0 algorithms. The results showed that C5.0 achieves higher accuracy taking lesser time.

In [11] application layer DDoS attack detection is considered. It shows the importance of feature extraction and proves that fewer features should be considered to avoid the curse of dimensionality. In [12], a work which aimed at detecting DDoS attacks against datacentres with correlation analysis showed that a lot of overhead is involved in K-Nearest Neighbour classifier.

Though a majority of the companies have their network secured, a study by [13] shows that big companies like Yahoo, CNN, eBay, and Amazon.com were damaged by DDoS attacks in 2000. To keep the network stable, the service providers must always monitor their systems. Once the systems appear abnormal, suitable security measures should be initiated immediately [14].

3. Proposed System

The Proposed system consists of a setup for simulating a DDoS attack, capturing the packets from the network to create the dataset for training the model. A pre-processing step is essential to ensure that the data is consistent. Finally the system contains an implementation of the C5.0 Machine learning algorithm which is used for the classification.

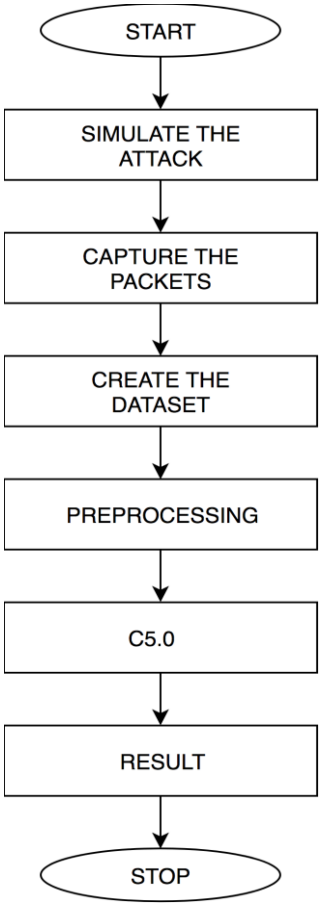

Fig.1. System Design

As shown in Fig. 1, the DDoS attack is simulated in a virtual environment comprising of several virtual machines. The packets belonging to both normal traffic and attack traffic are captured. These captured packets are used to create the dataset. Preprocessing of the created dataset is done to remove redundant features and following that, the C5.0 Machine Learning algorithm is applied on the preprocessed dataset. The results of the classification are obtained in the final step.

A. Working Setup

Simulation of the attack is done using virtual machines created on VMWare Workstation. A single server/victim machine and multiple zombie/attacker machines are created on VMWare. The zombie systems attack the server and this attack is performed using a popular tool called hping3. TCP SYN Flooding and UDP flooding are used for the attack. Wireshark, the free and open source packet analyzer is used for packet capturing. The captured packets are exported as comma separated values (csv) formatted files to create the dataset. Irrelevant features are removed from the dataset during preprocessing. C5.0 Machine learning algorithm is implemented in R programming language. The dataset is read and classification algorithm is applied to get the results.

B. Packet Format

Table 1. Sample Packet Format

Time	Source IP	Destination IP	Src Port	Dst Port	Len
1.55	192.168.1.10	192.168.1.13	4105	5895	1001
1.57	192.168.1.11	192.168.1.13	4104	5896	1001
1.58	192.168.1.12	192.168.1.13	4099	5899	1001
1.59	192.168.1.10	192.168.1.13	4094	5904	0

Table 1 shows the sample format of preprocessed packets in the dataset. Src port is the source UDP port number and Dst port is the destination UDP port number.

C. Goals of the Model

In this subsection, the goals to be achieved by this model are listed:

- High accuracy
- Faster detection rate
- Low computational cost
- Detection of cloud-based DDoS attacks

D. Classifiers

The results obtained from C5.0 classifier is compared with the existing state of the art classifiers like Naïve Bayes classifier and C4.5 Decision tree classifier.

Naïve Bayes is a simple probabilistic classifier which works based on Bayes theorem/rule. It assumes class conditional independence which considers the independent nature of different attribute values. According to Bayes theorem, the probability of Hypothesis can be calculated based on the Hypothesis and the evidence about the Hypothesis

$$P(H|X)=P(X|H)P(H)/P(X) \quad (1)$$

C4.5 is a state of the art statistical classifier. It builds a decision tree from a set of training data. Entropy based gain ratio is used as the splitting criterion for choosing the attributes. This helps to avoid the overfitting problem and choose the attribute that most effectively splits the data. The Gain ratio is calculated from Information gain which is the measure of the information that is gained by partitioning based on a certain attribute.

C5.0 offers significant improvements over C4.5. It is extremely fast and more memory efficient than its predecessor. Information gain (Entropy) is used as the splitting criterion. C5.0 supports boosting option in which the classifier creates multiple decision trees and combines them to build a unified model. Winnowing is another feature of C5.0 by which it automatically removes those attributes that may not be helpful in classification.

E. Evaluation

The proposed model is evaluated against other classification methods by using Confusion Matrix as shown in Table 2.

Table 2. Confusion Matrix

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

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

The accuracy of the classifier is its effectiveness by the percentage of correct prediction

$$\text{Accuracy} = (TP+TN)/TP+TN+FP+FN$$

4. Results and Discussion

The experimental evaluation is done on the basis of classifier accuracy and the time taken by each of the classifiers. Fig. 2. Shows the accuracy of the three classifiers for different cases. In all the cases, Naïve Bayes classifier has the least accuracy whereas C4.5 classifier performs with a very high accuracy. But C5.0 classifier outperforms the other two in all the cases with 100% accuracy.

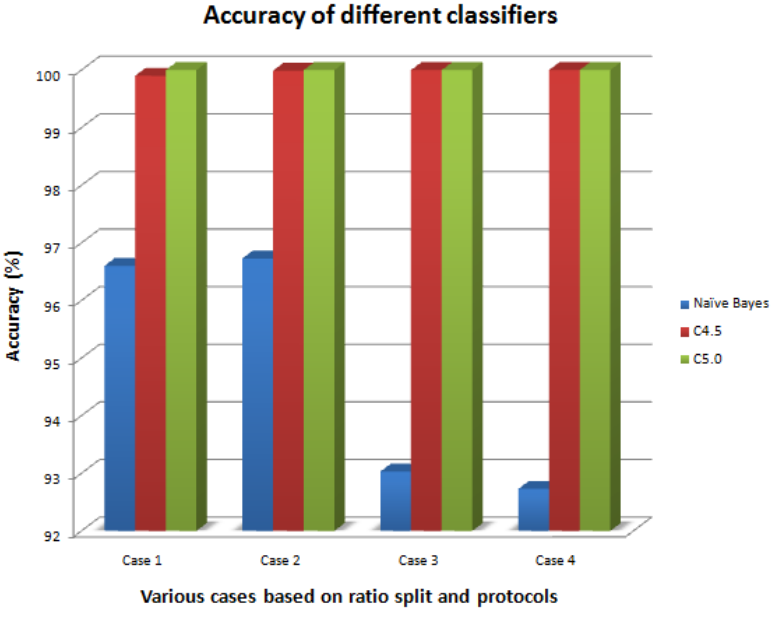

Fig.2. Accuracy

Fig. 3. Shows the time taken by the three different classifiers for creating the model and making the predictions. The C4.5 decision tree classifier takes the longest time in all the cases whereas Naïve Bayes classifier takes the least time. C5.0 classifier takes very little time compared to C4.5 for classification.

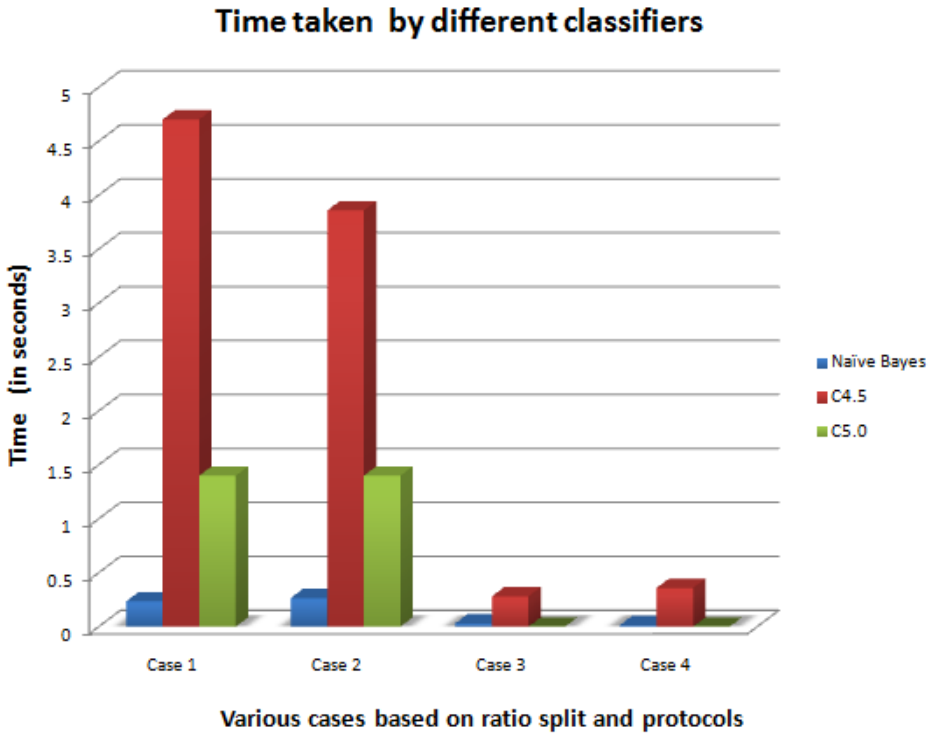

Fig.3. Time taken

Table 3. Comprehensive Result

	Accuracy (%)	**Time(seconds)**
Naïve Bayes	94.77	0.284
C4.5	99.99	2.30
C5.0	100	0.70

Table. 3, gives the comprehensive results of the different classifiers based on their accuracy in percentage and the time they took in seconds. The results are the average of the values taken over several test cases. The different test cases are created by taking different test-train data split like 30-70, 20-80 and type of traffic like TCP or UDP.

100,000 packets were used to train and test the algorithm. The results show that C5.0 finds out the attack packets with 100% accuracy. The time taken to detect is negligible compared to C4.5 algorithm and is considerably less than Naive Bayes algorithm. In all the parameters C5.0 algorithm outperforms the other two algorithms and proves to be the best algorithm to detect the attack packets in a DDoS attack.

5. Conclusion and Future Scope

In this paper, a DDoS attack detection system using C5.0 machine learning algorithm is proposed. The system was compared with other state of the art classifiers and the results showed that the performance was better than the state of the art methods. C5.0 classifier produces shorter decision trees taking very less time when compared to the C4.5 classifier and yet gives better accuracy in classification than both Naïve Bayes as well as C4.5 classifiers. This performance difference becomes more and more considerable as the dataset size increases.

Even though the results are exceptional, the model is still more of a damage limitation model rather than a preventive model. The detection takes place after the damage has been done. So, as a future extension of this project, a real-time DDoS attack detection model may be developed. A model that could use the best features of the C5.0 classifier and is able to detect an attack as and when it happens would be a perfect solution for the threats posed by Distributed Denial of Service attacks.

Acknowledgements

This project is supported by TEQIP – III of BMS College of Engineering Bengaluru, Karnataka, India.

References

[1] Douligeris, C., & Mitrokotsa, A. (2004). DDoS attacks and defense mechanisms: classification and state-of-the-art. *Computer Networks*, *44*(5), 643-666.

[2] Cheeseman, P. C., Self, M., Kelly, J., Taylor, W., Freeman, D., & Stutz, J. C. (1988, August). Bayesian Classification. In *AAAI*(Vol. 88, pp. 607-611).

[3] Suthaharan, S. (2016). Support vector machine. In *Machine learning models and algorithms for big data classification* (pp. 207-235). Springer, Boston, MA.

[4] Quinlan, J. R. (2014). *C4. 5: programs for machine learning*. Elsevier.

[5] Quinlan, R. (2004). Data mining tools See5 and C5. 0.

[6] Peterson, L. E. (2009). K-nearest neighbor. *Scholarpedia*, *4*(2), 1883.

[7] Zekri, M., El Kafhali, S., Aboutabit, N., & Saadi, Y. (2017, October). DDoS attack detection using machine learning techniques in cloud computing environments. In *Cloud Computing Technologies and Applications (CloudTech), 2017 3rd International Conference of* (pp. 1-7). IEEE.

[8] Bujlow, T., Riaz, T., & Pedersen, J. M. (2012, January). A method for classification of network traffic based on C5. 0 Machine Learning Algorithm. In *Computing, Networking and Communications (ICNC), 2012 International Conference on*(pp. 237-241). IEEE.

[9] Pandya, R., & Pandya, J. (2015). C5. 0 algorithm to improved decision tree with feature selection and reduced error pruning. *International Journal of Computer Applications*, *117*(16).

[10] R. Revathy, and R. Lawrance. (2017). "Comparative Analysis of C4.5 and C5.0 Algorithms on Crop Pest Data," *International Journal of Innovative Research in Computer and Communication Engineering(IJIRCCE)* Vol. 5, Special issue 1, March 2017

[11] Liao, Q., Li, H., Kang, S., & Liu, C. (2015). Application layer DDoS attack detection using cluster with label based on sparse vector decomposition and rhythm matching. *Security and Communication Networks*, *8*(17), 3111-3120.

[12] Xiao, P., Qu, W., Qi, H., & Li, Z. (2015). Detecting DDoS attacks against data center with correlation analysis. *Computer Communications*, *67*, 66-74.

[13] Karimazad, R., & Faraahi, A. (2011, September). An anomaly-based method for DDoS attacks detection using RBF neural networks. In *Proceedings of the international conference on network and electronics*

Engineering (Vol. 11, pp. 44-48).

[14] Zhong, R., & Yue, G. (2010, April). DDoS detection system based on data mining. In *Proceedings of the 2nd International Symposium on Networking and Network Security, Jinggangshan, China* (pp. 2-4).

Authors' Profiles

Hariharan M Obtained Bachelor of Technology in Computer Science and Engineering from NSS College of Engineering Palakkad, Kerala in 2017. He is currently pursuing his Masters in Computer Science and Engineering from BMS College of Engineering, Bengaluru, Karnataka.

Abhishek H K pursued Bachelor of Engineering in Computer Science and Engineering from SJCE Mysore, Karnataka in 2015. He is currently pursuing his Masters in Computer Science and Engineering from BMS College of Engineering, Bengaluru, Karnataka.

Dr. B G Prasad pursued Bachelor of Engineering in Computer Science and Engineering from PES College of Engineering, Mandya, Karnataka in 1987. He pursued his Masters in Computer Science and Engineering from Indian Institute of Technology, New Delhi in 1992 and Ph.D. from IIT, New Delhi in 2003. He is currently working as a Professor and Head of the Department of Computer Science and Engineering, BMS College of Engineering, Bengaluru, Karnataka.

How to cite this paper: Hariharan. M, Abhishek H. K, B. G. Prasad, "DDoS Attack Detection Using C5.0 Machine Learning Algorithm", International Journal of Wireless and Microwave Technologies(IJWMT), Vol.9, No.1, pp. 52-59, 2019.DOI: 10.5815/ijwmt.2019.01.06