Performance Analysis of Machine Learning Algorithms for Detection of DDoS Attack

Dr S Veeramani[1], Nandhini Venkatesan[2], T Abinesh[3], A Priyadharshini[4]

[1]Dr S Veeramani Professor Computer Science department Dr Mahalingam College of Engineering and Technology
[2]Nandhini Venkatesan Computer Science department Dr Mahalingam College of Engineering and Technology
[3]T Abinesh Computer Science department Dr Mahalingam College of Engineering and Technology
[4]APriyadarshini Computer Science department Dr Mahalingam college of Engineering and technology

---

Abstract -WithCloud computing refers to a type of Internet-based computing that provides shared pool of resources such as network bandwidth, memory, computer processing and user applications. These resources can be rapidly provisioned on demand over Internet to End Users with less maintenance and less infrastructure cost. The cloud computing services can be categorized into three models. Software-as-a-service (SaaS), Platform-as a Service (PaaS) and Infrastructure-as-a-Service (IaaS). DDoS is a type of aggressive attack that causes serious troubles on cloud servers. The aim is to detect DDoS attack using machine learning algorithms and compare them in terms of their performance.

Key Words:DDoS attack, Cloud computing, Denial of Service, Intrusion, Machine learning.

1. INTRODUCTION

Tthe term Denial of Service (DoS) was originally coined by Gligor in an operating system context, but since became widely adopted [5]. A DoS attack involving more than one computer to target a victim in a coordinated manner is called a Distributed Denial of Service (DDoS) attack. This work focuses on using machine-learning techniques for detecting DDoS attacks. The problem of attack detection using machine-learning techniques is not new to literature. While signature detection techniques can detect attacks based on signatures of already learnt attacks, anomaly detection techniques learn network traffic from a baseline profile and detect anomalies as ones that deviate significantly from the baseline profile [2]. Signature detection techniques are effective against known attacks while anomaly detection has the ability to detect unknown and new attacks (zero-day) [3]. Data flow generated by attack presents irregular status. This makes DDoS attacks launched easily, prevented and tracked difficultly and so forth. Furthermore, DDoS attacks have become one of the essential threats to network security. Following, we analyze machine learning in order to pinpoint our system and have some grasp of what machine learning is and how it is evolving. Machine Learning is a sub-set of artificial intelligence where computer algorithms can be used to autonomously learn from data.

Machine learning severely affects most industries and the jobs within them. To arrive at this point advancement in the field of machine learning get passed by many major milestones As the amount of data we produce continue to grow exponentially, the need for computers to analyze this large amount of data and to draw more efficient and accurate conclusions becomes a serious requirement [1]. To this end, enhancing computers ability to process, analyze and learn from growing large amounts of data also requires more attention. As data grows and expands, decision tools need to evolve. In other hands we focus in this project on detection schemes we use machine learning techniques to detect and mitigate the DDoS attacks.

2. Methodology

In this section, we briefly describe the various machine learning algorithms and the problem domains they are frequently used in. Many decision tree and rule induction algorithms have already been suggested in the literature. The Naive Bayes algorithm is a probabilistic classifier, it assumes that the effect of a variable values on a given class is independent of the values of other variables. This assumption is called class conditional independence[4]. Decision tree is one of the most well-known and used classification algorithms. C4.5 algorithm which was developed by Ross Quinlan is the most popular tree classifier. This algorithm is based on ID3 (Iterative Dichotomiser 3) algorithm that tries to find a small decision tree. The decision tree generated by C4.5 can be used for classification, and it often referred to as a statistical classifier [7].

2.1 DATASET PRE-PROCESSING

The intrusion detector learning task is to build a predictive model (i.e. a classifier) capable of distinguishing between ``bad'' connections, called intrusions or attacks, and ``good'' normal connections [6]. The 1999 KDD intrusion detection contest uses a version of this dataset. A standard set of data to be audited, which includes a wide variety of intrusions simulated in a military network environment, was provided. The complete dataset can be downloaded

from this link: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html [8]. The dataset contains 494021 records. Each record contains 42 fields 4 of which have string values. The fields which have string values are: Field 2 - protocol type, Field 3 – service, Field 4 – flag and Field 42 - attack type. To reduce the complexity and speed up the process, the dataset is split up into subsets that do not have any shared records. The resultant data after reading them and all the data is saved into the SQL Server Database.

2.2 DDOS Attack detection using Decision Tree (C 4.5) Algorithm

C 4.5 is extension of ID3 algorithm. It tries to find simple and small decision trees. C4.5 builds decision trees from a set of training data on basis of information entropy by using the below formula.

$$Entropy(S) = \sum_{i=1}^{C} P_i log_2 P_i$$

Equation 1 Entropy calculation

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. Given a learning set A and a non-class attribute E, the Gain Ratio is defined as:

$$Gain(A) = E(Current\ set) - \sum E(all\ child\ sets)$$

Equation 2 Gain Ratio formula

As a standard practice, training and test datasets are divided in 90:10 ratio. This dataset is taken randomly from original data and experiment is repeated to validate the consistency.

2.3 DDOS Attack detection using Naive Bayes Algorithm

The main purpose of this module implementation IDNB (Intrusion Detection by Naive Bayes) is to detect intrusion packets or data for increasing the performance of processing model. The proposed scheme is able to incorporate flow correlation information in to the classification process. IDNB (Intrusion Detection by Naive Bayes) demonstrates malicious behavior detection rates in certain circumstances while does not greatly affect the network performances. NB is one of the earliest classification methods applied in intrusion detection system which is an effective probabilistic classifier employing the Bayes' theorem with naive feature independence assumptions.
A probabilistic classifier is a classifier that is able to predict, given a sample input, a probability distribution over a set of classes, rather than only predicting a class for the sample [9]. Naive Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem. Maximum-likelihood training can be done by evaluating a closed-form expression. Which takes linear time, rather than by expensive iterative approximation as used for many other types of classifiers. NB classifier requires a small amount of training data to estimate the parameters of a classification model.

2.4 Summary of Results

Performance Evaluation for decision tree

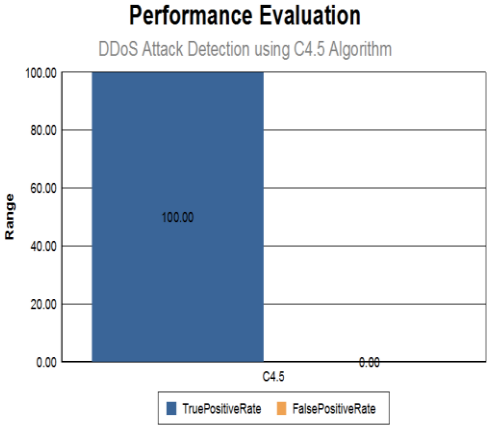

Figure 1 graph for decision tree

Algorithm	TruePositiveRate	FalsePositiveRate
C4.5	100.00	0.00

Figure 2- Efficiency measure for decision tree

Performance Evaluation for Naïve Bayes

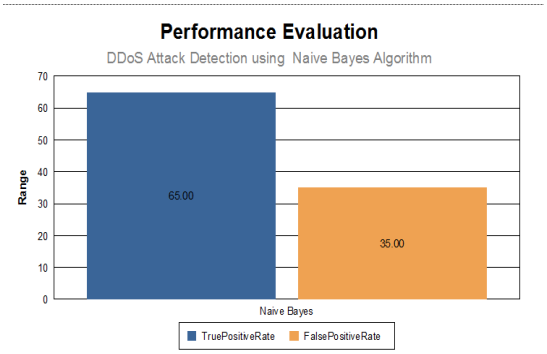

Figure 3 graph for Naïve Bayes

Algorithm	TruePositiveRate	FalsePsotiveRate
Naïve Bayes	65.00	35.00

Figure 4 Efficiency measures for Naive Bayes

3. CONCLUSIONS

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 Naive Bayes, 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 inDDoS detection [10].

REFERENCES

1. P. Xiao, W. Qu, H. Qi, and Z. Li, “Detecting ddos attacks against data center with correlation analysis,” Computer Communications, vol. 67, pp. 66–74, 2015.
2. R. Karimazad and A. Faraahi, “An anomaly-based method for ddos attacks detection using rbf neural networks,” in Proceedings of the International Conference on Network and Electronics Engineering, 2011, pp. 16–18.
3. R. Zhong and G. Yue, “Ddos detection system based on data mining,”in Proceedings of the 2nd International Symposium on Networking and Network Security, Jinggangshan, China, 2010, pp. 2-4.
4. Y.-C. Wu, H.-R. Tseng, W. Yang, and R.-H. Jan, “DDoS detection and traceback with decision tree and grey relational analysis,” International Journal of Ad Hoc and Ubiquitous Computing, vol. 7, no. 2, pp. 121– 136, 2011.
5. J. Li, Y. Liu, and L. Gu, “DDoS attack detection based on neural network,” in 2nd International Symposium on Aware Computing (ISAC),. IEEE, 2010, pp. 196–199.
6. Zekri, Marwane, Said El Kafhali, Noureddine Aboutabit, and Youssef Saadi. "DDoS attack detection using machine learning techniques in cloud computing environments." In 2017 3rd International Conference of Cloud Computing Technologies and Applications (CloudTech), pp. 1-7. IEEE, 2017.
7. Barati, Mehdi, Azizol Abdullah, Nur Izura Udzir, Ramlan Mahmod, and Norwati Mustapha. "Distributed Denial of Service detection using hybrid machine learning technique." In 2014 International Symposium on Biometrics and Security Technologies (ISBAST), pp. 268-273. IEEE, 2014.
8. Zhou, Baojun, Jie Li, Jinsong Wu, Song Guo, Yu Gu, and Zhetao Li. "Machine-Learning-Based Online Distributed Denial-of-Service Attack Detection Using Spark Streaming." In 2018 IEEE International Conference on Communications (ICC), pp. 1-6. IEEE, 2018.
9. Kim, Donghoon, and Ki Young Lee. "Detection of DDoS Attack on the Client Side Using Support Vector Machine." International Journal of Applied Engineering Research 12, no. 20 (2017): 9909-9913.
10. Chen, Liguo, Yuedong Zhang, Qi Zhao, Guanggang Geng, and ZhiWei Yan. "Detection of DNS DDoS Attacks with Random Forest Algorithm on Spark." Procedia computer science 134 (2018): 310-315.