JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

DDoS ATTACK BASED ON SEMI SUPERVISED MACHINE LEARNING APPROACH

[1]K. Arun Babu, [2]G. Tejaswini Ratna Tulasi, [3]K. Sravani,[4]Ch. Akhila,[5]G. Surya Prakash

[1]Asst. Professor, Department of Computer Science, Bapatla Engineering College, Bapatla, India
[2]CSE, Department of Computer Science, Bapatla Engineering College, Bapatla, India
[3]CSE, Department of Computer Science, Bapatla Engineering College, Bapatla, India
[4]CSE, Department of Computer Science, Bapatla Engineering College, Bapatla, India
[5]CSE, Department of Computer Science, Bapatla Engineering College, Bapatla, India

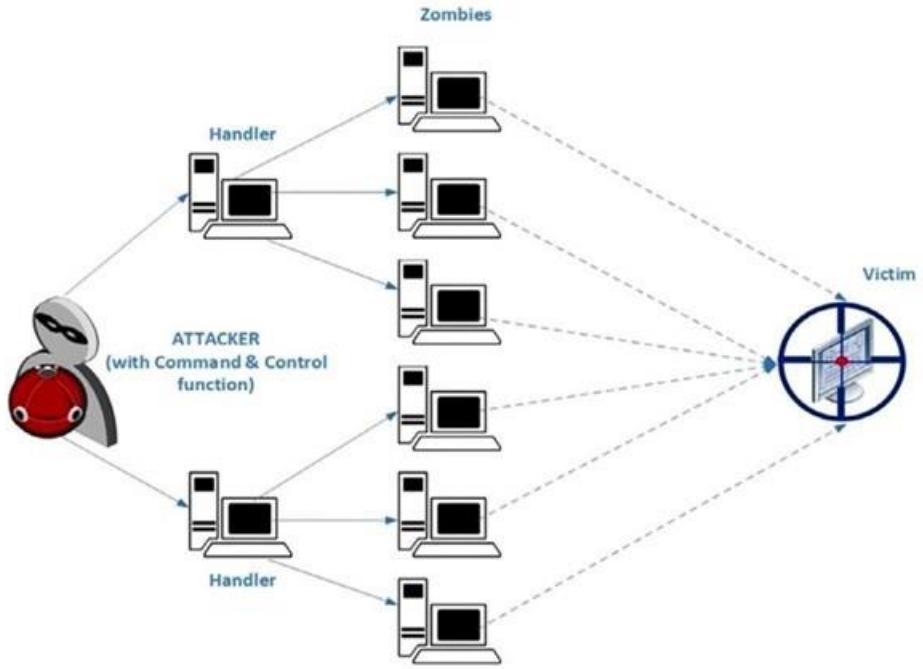
ABSTRACT: Distributed Denial of Service refers to a cyber-attack resulting in victims being unable to access systems and network resources, essentially disrupting internet services. Many machine learning techniques are there to detect the DDoS attack, but the attack remains major issue of the internet. The existing system have low detection accuracy and high false positive rates. In order to increase the accuracy and reduce the false positive rates, this paper presents a semi-supervised approach to detect the DDoS attack. Semi Supervised Approach takes the advantage of both supervised and unsupervised approaches by the ability to work on labelled and unlabeled datasets. Unsupervised part of our approach is K-Means clustering algorithm used to obtain classes to distinguish attacks from normal traffic. After that supervised algorithms of Support Vector Machine and Random Forest are applied for classification purpose. By using this approach, we can increase accuracy and reduce the false positive rates.

INDEX TERMS - DDoS attacks, K-Means, Support Vector Machine, Random Forest.

1.INTRODUCTION

In computing, a denial-of-service attack is a cyber-attack in which the perpetrator seeks to make a machine or network resource unavailable to its intended users by temporarily distributing services of a host connected to internet. A distributed denial-of-service (DDoS) attack occurs when multiple systems flood the bandwidth or resources of a targeted system, usually one or more web servers. Such an attack is often the result of multiple compromised systems (for example, a botnet) flooding the targeted system with traffic. A botnet is a network of zombie computers programmed to receive commands without the owner's knowledge. When a server is overloaded with connections, new connections can no longer accepted.

The main advantages to an attacker of using a distributed denial-of-service attack are that multiple machines can generate more attack traffic than one machine, multiple attack machines are harder to turn off than one attack machine, and that the behavior of each attack machine can be stealthier, making it harder to track and shut down. These attacker advantages cause challenges for defense mechanisms. For example, merely purchasing more incoming bandwidth than the current volume of the attack might not help, because the attacker might be able to simply add more attack machines. This, after all, will end up completely crashing a website for periods of time. The main purpose of DDoS attack is to make an online service or website unavailable by flooding it with unwanted traffic from multiple computers. For a DDOS attack to be successful, an attacker will spread malicious software to vulnerable computers, mainly through infected emails and attachments. This will create a network of infected machines which is called botnet. The attacker can then instruct and control the botnet, commanding it to flood a certain site with traffic. So that the network ceases to work, taking the site offline.



1.1 Common Flow of DDoS Attack

In this paper, a semi-supervised machine learning technique is applied. Our approach mainly consists of two methods Supervised and unsupervised learning methods. The unsupervised method comprises K means algorithm and in supervised learning method includes classification algorithms like Random Forest and Support Vector Machine are Used. At first an unsupervised learning method is applied to get the clusters of network traffic data and then it should be divided into desired number of classes. After labeling the data points as normal traffic or abnormal traffic, finally supervised learning algorithms Support Vector Machine and Random Forest are used for classification of DDoS attack.

2. LITERATURE SURVEY

Literature survey is the most important step in software development process. Before developing the tool, it is necessary to determine the time factor, economy, and company strength. Once these things are satisfied, then next step is to determine which operating system and language can be used for developing the tool. Before building the system, the above consideration is taken into account for developing the proposed system. As part of Literature survey, a lot of information is gathered and gained.

A defense system was presented by Boro D. et al. referred to as DyProSD that combines both the merits of feature-based and statistical approach to handle DDoS flooding attack. The statistical module marks the suspicious traffic and forwards to an ensemble of classifiers for ascertaining the traffic as malicious or normal.

Recently, Van Loi C. proposed a novel one class learning approach for network anomaly detection based on combining auto-encoders and density estimation. Authors have tested their method on the NSL-KDD dataset and obtained satisfactory results.

Mohamed I. et al. have proposed a supervised DoS detection method based on a feed-forward neural network. Mustapha B. et al. have presented a two-stage classifier based on RepTree algorithm and protocols subset for network intrusion detection system.

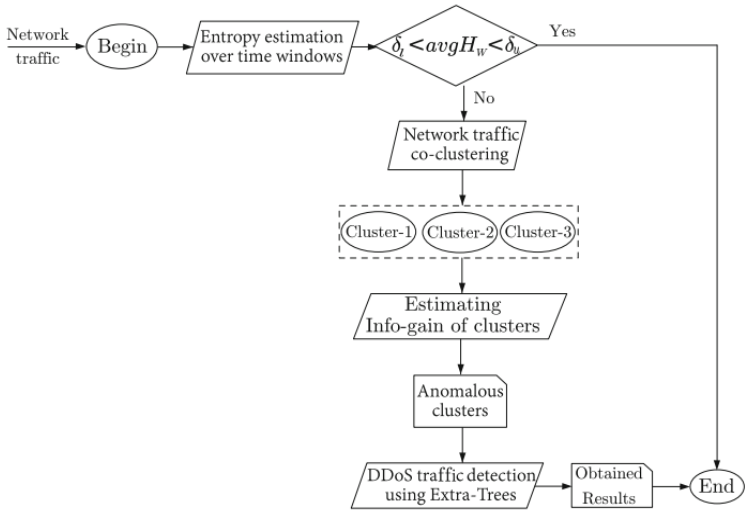
Akilandeswari V. et al. have used a probabilistic Neural Network to discriminate flash crowd events from DDoS attacks. The method achieves high DDoS detection accuracy with low false positive rates. Similarly, Ali S.B. et al. have proposed an innovative ensemble of Sugeno type adaptive neuro-fuzzy classifiers for DDoS detection using an effective boosting technique named Marliboost.

Mohiuddin A. and abdun Naser M have proposed an unsupervised approach for DDoS detection based on the co-clustering algorithm. The authors have extended the co-clustering algorithm to handle categorical attributes.

The detection of the Distributed Denial of service attack approaches that are mentioned in literature are mainly consists of two parts unsupervised method and supervised method. Based on the datasets used in the previous approaches, unsupervised method often faces combines both the merits of feature-based and statistical approach to handle DDoS flooding attack. The statistical module marks the suspicious traffic and forwards to an ensemble of classifiers for ascertaining the traffic as malicious or normal.

3.EXISTING SYSTEM

In existing system, they introduce a methodology to detect the attack of DDoS. Itis a five-fold steps data mining technique process. The approach starts with entropy estimation based on the time-based sliding window. Whenever the time windowaverage entropy exceeded its lower or upperthresholds, the co-clustering algorithmdivides the network traffic data into different clusters.



3.1 Flowchart of the Existing system

Estimation of entropy based on the time sliding windows allows to identify the sudden changes in the incoming network data those are generally caused by the DDoS attacks. Incoming network traffic within the same windows having anomalous values of entropy isconsider as containing DDoS traffic. To determine the normal cluster, they evaluate the information gain ratio over the average entropy of the features of FSD between the obtained clusters and received network traffic data in the current time window. Hence, the cluster is considered as a normal cluster if it produces lower information gain ratio else other clusters are considered as abnormal. Those abnormal clusters are taken for pre-processing and classification techniques using supervised learning algorithm Extra-Trees. These Extra Trees is a classification algorithm used to determine whether it is normal traffic or anomalous traffic.

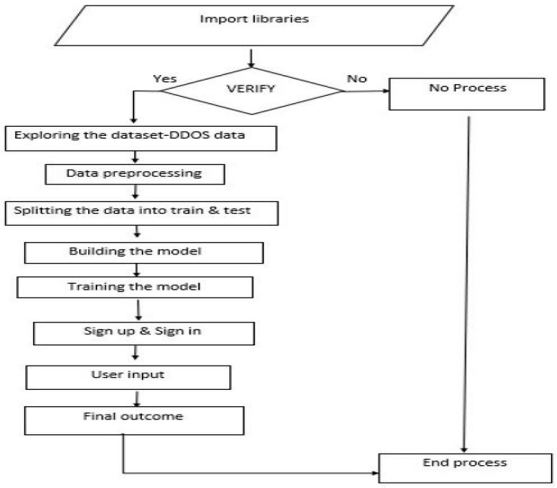
4. PROPOSED SYSTEM

In this paper we propose a DDoS attack detection based on semi supervised machine learning approach. Semi-supervised is a type of Machine Learning algorithm that represents the middle ground between supervised and unsupervised machine learning methods. This semi supervised learning is an important category that lies between the supervised and unsupervised machine learning. As a part of the unsupervised machine learning methods, K means clustering algorithm is used and Support Vector Machine (SVM) and Random Forest (RF) are algorithms of the supervised learning methods that are used for the classification of the DDoS attack accurately.

At first, we import all the required libraries after that the DDoS attack Dataset is explored. Here we use the NSL KDD dataset to evaluate the performance of our proposed approach. The next step is to preprocess the data. Data preprocessing is an important step for cleaning the data, removing noisy data and make it more suitable for a machine learning method model and also increases the accuracy and efficiency of a model. And then we split the data into train and test. Here the main part of our approach starts.

An unsupervised K-means algorithm is built. This algorithm is iterative method that divides the dataset into K clusters. It assigns data points to a cluster such that the sum of the squared distance between the data points and the cluster's centroid is at the minimum. The less variation we have within clusters, the more homogeneous the data points are within the same cluster. Therefore, we labeled the data points as normal traffic or abnormal traffic using this algorithm.

After that supervised algorithms are applied on the labeled data. Here we use the Random Forest and Support Vector Machine algorithms. Random Forest is an ensemble learning method for classification and regression. It grows multiple decision trees which are merged together for a more accurate and prediction. The logic behind the Random Forest is that the individual's decision trees perform much better as a group than do alone. While using Random Forest for classification, each tree gives a classification or a vote. The forest chooses the classification with most of the votes.



4.1 Flowchart of a Proposed system

Other classification algorithm is Support Vector Machine. It is also used for both classification and regression. Support vector Machine model basically a representation of multiple classes in a hyperplane in multidimensional space. Here a hyperplane will be generated in an iterative manner by SVM so that the error can be minimized. The goal of SVM is to divide the datasets into classes to find a maximum marginal hyperplane. These both proposed classification algorithms are classifying the data and detect whether the traffic is a normal traffic or containing the DDoS traffic. For Support Vector Machine algorithm, we get an accuracy of 98.7% and for Random Forest algorithm we get 99% accuracy. Therefore, we increase the accuracy and reduce the false positive rates compared to the existing approach.

5.CONCLUSION

In this paper, we introduced DDoS attack detection based on semi supervised machine learning approach. The algorithms K means, Random Forest and SVM algorithm are used to create a useful malware detection model. Our evaluation demonstrates the efficiency of this solution, and our trained model greatly improves existing approaches.

6.REFERENCES

[1] Bhuyan MH, Bhattacharyya DK, Kalita JK (2015) Low-rate and high-rate ddos attack detection of an empirical evaluation of information metrics. Pattern Recogn Lett 51:1–7
[2] Yu S (2014) Distributed denial of service attack and defense, Springer, Berlin
[3] Saied A, Overill RE, Radzik T (2016) DDOS attack detection of known and unknown using artificial neural networks.
[4] Papalexakis EE, Beutel A, Steenkiste P (2014) Network anomaly detection using co clustering. In: Encyclopedia of social network analysis and mining.
[5] Boroujerdi AS, Ayat S (2013) A robust ensemble of neurofuzzy classifiers for ddos attack detection. In: 2013 3rd international conference on computer science and network technology.
[6] Ahmed M, Mahmood AN (2014) Network traffic pattern analysis using improved information theoretic co-clustering based collective anomaly detection. In: International conference on security and privacy in communication systems. Springer, Berlin, pp 204–219
[7] Wikipedia (2016) 2016 dyn cyber attack