IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

DDOS Attack Detection Using Machine Learning For Network Performance Improvement

Dhairya Lunkad[1], Govind Singh[2]

M.Tech Student[1], Dept. Of Computer Science, Shri Shankaracharya Engineering College, C.G.
dhairyalunkad@gmail.com

Assistant Professor[2], Dept. Of Computer Science, Shri Shankaracharya Engineering College, C.G.
govind.singh@sstc.ac.in

Abstract

DDoS attacks are a critical issue for companies that have been integrating their technology to public networks, allowing multiple attackers to access data or render services to large companies or countries The Distributed Denial of Service (DDoS) attacks affect the availability of Web services for an indeterminate period of time, flooding the company's servers with fraudulent requests and denying requests from legitimate users, generating economic losses by unavailable rendered services. Therefore, the aim of this paper is to show the process of detection prototype DDoS attacks using a supervised learning model by Support Vector Machines (SVM), which captures network traffic, filters HTTP headers, normalizes the data on the basis of the operational variables: rate of false positives, rate of false negatives, rate of classification and then sends the information to corresponding training and testing sets With the selected attributes, various machine learning models, like Navies Bayes ,SVM, and proposed methods based on Navies Bayes are developed for efficient detection of DDoS attacks. Then our experimental results show that Fuzzy c-means clustering gives better accuracy in identifying the attacks.

Keywords: Navies Bayes , Support Vector Machines , Machine Learning , DDOS Attack Detection

INTRODUCTION

A DDoS attack consists in to throw tens or hundreds of thousands of requests per second to a server from different locations or IPs; the concept of "Distributor" is concerning that these requests are made from hundreds of thousands of infected machines (commonly called "zombies") which are governed by "botnets" in a coordinated way at the same time, i.e. SYN Flood, Smurf attacks, which are a sum of bandwidth, memory usage and target's processing, usually no servers could handle ending in a collapse of service because it cannot answer every request; therefore it's necessary the development of new techniques and prototypes to detect fraudulent attacks of concurrent requests in an effective and efficient way also it's necessary in order to avoid the unavailability of service and economic losses. Machine learning using SVM have been used with great success in the field of information security and pattern recognition research in different processes of classification, prediction and regression. The application of techniques with a SVM supervised model has large advantages over rule-based techniques, since the generation of the model is based on a statistical model that changes its behaviour according to the input parameters defined and based on a training rule that requires human interaction; in the prototype evaluation it was found that the correct classification rate of normal or abnormal requests in the training phase is directly related to standardization and proper selection of the input parameters, allowing the output variables are generated with minimum percentage of misclassification, generating confidence in the generated model and the detection of these behaviours. The paper describes: The contextual reference and some relevant work in section 2. The Section 3 shows the proposed model of the development and application of machine learning prototype. In section 4 some results are evident and in the end, finally in Section 5 the conclusions are presented.

Related work

Wani, A. R., et al (2019). The DDoS attacks on the cloud computing environment are mainly application layer which sends out requests following the communication protocol which are then hard to distinguish in the network layer because their pattern matches the legitimate requests thus making the traditional defence systems not applicable. DDoS flooding attacks on cloud can be of various categories like session and request flooding attacks, slow response and asymmetric attack. All these flooding attacks generate traffic which resembles that of a legitimate user which becomes tougher for the target to distinguish between attack and legitimate traffic thus blocking the services for the legitimate user.

WU ZHIJUN, XU QING wt al (220) In recent years, as a new type of network architecture, software-defined networks have attracted great interest from researchers. They are gradually being widely applied to various fields of the network. However, low-rate DDoS attacks against the data layer have not yet become research hotspots, and related research results are also less. This paper first studies how to launch such attacks and verifies the effectiveness of such attacks. Then, by extracting the four features related to the flow rules, the feature data set for detecting such attacks is established, and the FM-based detection method for low-rate DDoS attack in SDN is proposed.

Yadav, S., & Subramanian, S. (2016). The features are extracted from the web server log (Attack and Normal) to build an AL-DDoS attack dataset. Pre-processing is performed on the extracted features so that all features are in numeric form. Then the pre-processed features are fed as input to the feature learning module. In the feature learning, more abstract features are learned by using deep learning methods such as Stacked Auto Encoder and a deep architecture is constructed. The AL-DDoS attack dataset is split into two datasets, viz., training dataset and testing dataset. This is done for cross validation in order to avoid over fitting of results. In this paper, 66% split technique is used for cross validation. In the training phase, the algorithm is trained to learn high levels of features of AL-DDoS attack dataset and in testing phase the algorithm is tested for incoming traffic.

Bakker, J. et (2018) This paper has shown how statistical classification can be deployed using SDN to detect DDoS attacks. Three classifiers were selected in an off-line environment to be integrated with nmeta2. These were then evaluated on a physical network testbed by replaying a DDoS attack scenario. While statistical classification can be deployed using SDN to classify traffic, careful consideration must be made to pick classifiers that result in the smallest possible packet processing overhead. Although the classifiers did not demonstrate a high DR, results did suggest that particular statistical classification methods can classify network traffic under normal conditions using the nmeta2 architecture. Under a DDoS attack scenario however, nothing is safe.

Khuphiran, P et al (2018) the application of machine learning algorithm for the problem of DDoS attack detection has been addressed. Two algorithms, Support Vector Machine (SVM) and Deep Feed Forward (DFF) have been evaluated to demonstrate the feasibility of applying these algorithms. The experiments have been conducted to compare the performance of these two algorithms. It has been found that DFF can classify the data with a higher accuracy. Therefore, deep learning is a useful choice for the classification of DDoS attack packets in terms of accuracy. However, SVM is an appropriate choice for faster classification method. In our future work, we plan to examine both two algorithms with the real time data to develop a useful method that is available with the real networks.

Yuan, X., Li, C., & Li, X. (2017). It is hard to detect low rate attack because it looks similar to the legitimate network traffic from the victim-end. Meanwhile, DDoS attacks toward victim systems must be generated over time. Otherwise, it won't be malicious to the network/system resources. This suggests the importance of historical information in DDoS detection. Single-packet based detection method can't improve the performance due to the missing historical pattern in the learning model.

Subbulakshmi, T et al (2011) According to the widespread of using computer networks, the number of attackers is increasingly rapidly. So, intrusion detection system as a defence tool is very important because firewall cannot provide protection against attacks for inside an organization. Moreover the defence tool has to detect new attacks inside. The datasets provide a way to train the IDS in an effective manner. In this paper a DDoS dataset which is derived from various other parameters of the DDoS attacks with latest types of DDoS attack data is generated. A new type of SVM called EMCSVM is proposed with added weights for the dataset records for detection of the DDoS attacks into various classes.

DDOS ATTACKS AND DETECTION METHODOLOGIES

Despite the fast increasing popularity of cloud services, ensuring the security and availability of data, resources and services remains an ongoing research challenge. Distributed denial of service (DDoS) attacks are not a new threat, It is major security issue and a wide topic of ongoing research interest. In this section we discuss the various DDoS intend and Launch methods that could be used to conduct or facilitate DDoS attacks, as well as reviewing intrusion Detection Methodologies and defence strategies.

A. DDoS Attack

1) DDoS intend and Launch methods: DoS attacks are intended attempts to stop legitimate users from accessing a specific network resource. The Open Systems Interconnection Model (OSI model), is useful in understanding the types of DDoS attacks we are dealing with .DDoS attacks target specific layers of a network connection(application layer attacks target layer 7, protocol layer attacks target layers 3 and 4). The first DDoS attack incident was reported [30].

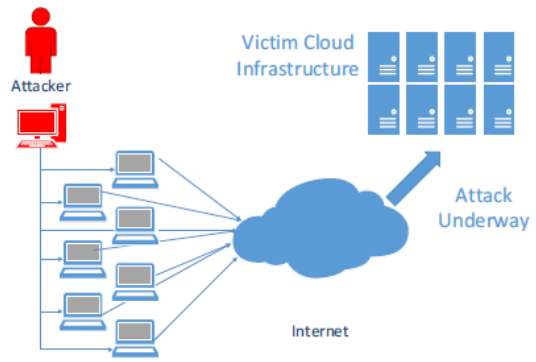

Fig. 1. Typical Architecture of DDoS Attacks

Fig. 1. Typical Architecture of DDoS Attacks

Currently, there are two main methods to trigger a DDoS attacks in the Internet. The first is to send some malformed packets to the victim (i.e., vulnerability attack). The second method, involves an attacker trying to do one or both of the following:

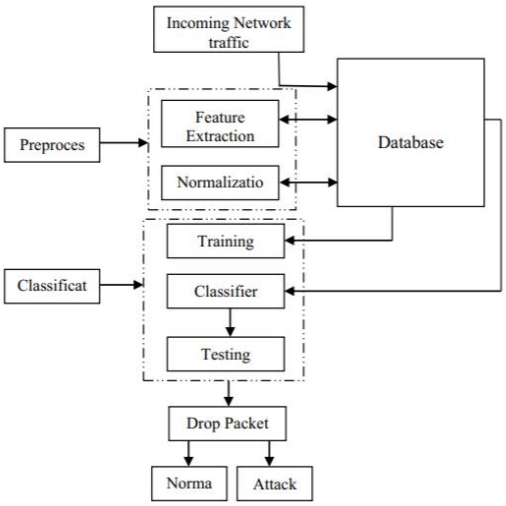

Figure 2: Existing model (Sumathi, S., & Karthikeyan, N. (2020)

Disrupt a legitimate users connectivity by exhausting bandwidth, router processing capacity or network resources.

These are essentially network/transport-level flooding attacks . (i.e., flooding attacks)

• Disrupt legitimate users services by exhausting the server resources (e.g., sockets, CPU, memory, disk/database bandwidth, and I/O bandwidth) These essentially include application-level flooding attacks.

2) The attacker's incentives: DDoS attackers are usually motivated by various justifications. Analyzing the attackers incentives help to stop and respond to these attacks .

• Economical/Financial gain: A major concern of corporations generally performed by frustrated individuals, possibly with lower technical skills.

• Intellectual Challenge: The attacks are usually young hacking enthusiasts who want to show off their capabilities to experiment and learn how to launch various attacks.

• Cyber warfare: This category of Attackers are usually politically motivated to attack a wide range of critical sections of another country.

Machine Learning Algorithms In this section, we briefly describe the various machine learning algorithms employed in the proposed framework.

Naive Bayes the Naïve Bayes is a simple probabilistic classifier [13]. 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.

SVM In classification and regression, Support Vector Machines are the most common and popular method for machine learning tasks . In this method, a set of training examples is given with which each example is marked belonging into one of two categories. Then, by using the Support Vector Machines algorithm, a model that can predict whether a new example falls into one categories or other is built.

Proposed methodology

DDoS Attack Distributed Denial of Service (DDoS) attack is the simple and a robust technique to attack Internet and system resources. The side effect seriously affects real networks together with insect viruses. Many researches for detection mechanism have performed, because the DDoS attack increases. The existing protection mechanisms have defence capability that is exclusively limited to set of DDoS attacks. There are many applications where data mining procedures can be situated in the detection of DDoS attacks. The following study discusses the extraction of a feature set from two different sources of datasets of Internet traffic. These are the public-domain CAIDA Dataset and traffic collected on the smart and secure environment (SSE) Network. Various types of DDoS attacks are studied to select the traffic parameters that change unusually during such attacks. Twenty-three features are collected and ranking the twenty-three features is done with Information Gain and Chi-Square statistic which reduces the number of features to eight. All the features used in this paper are calculated at an interval of 1 second. Since these classes are well divided as attack and normal, it is possible to apply various machine learning algorithms for the detection. The approach considered is to use the feature selection mechanism discussed previously and build the classifier using various machine learning algorithms such as SVM, K-NN, Naive Bayesian,. This phase of the study is an evaluation of the performance of the selected set of machine learning algorithms in detecting DDoS attacks. The performance measures are the receiver operating characteristic (ROC) curve and F-measure. An important

Training In training phase each data from the given dataset are given into pre-processing, feature extraction. In this stage, features for each individual data from dataset are trained and assigned classes for trained features. There are two classes in this algorithm are Normal and DDoS attack.

Testing In testing, the test samples are fed into DNN classifier to classify the test instance with given class j using training features. If the classifier classify given class correctly then the process provides better result. If the decision of DNN classifier is not final then the decision is made by cost minimization algorithm approach.

Proposed methods based on Bayesian algorithm is one of the classifiers in machine learning approach, which classifies the data by assigning the class label to the problem instance [10, 11], where the class label is from the dataset. Main concept it assumes is that all the features are independent of each other. This assumption is known as class conditional independence It requires a small amount of training data for classification of the attack. It follows a Bayesian probabilistic model which has been given below:

$$p(C|S_0, S_1, \ldots.., S_n) = \frac{p(C).p(S_0, S_1, \ldots.., S_n|C)}{p(C|S_0, S_1, \ldots.., S_n)}$$

WhereC is the class of dataset, S0, S1,…,Sn is the set of features in the dataset, and p() is a probability function

Our proposed methods bsed on naive Bayes classifier emerged out to be the best classifier for detecting DDoS attacks. most of the authors have considered approaches and also suggested that naïve Bayes approach will give better accuracy. So, naive Bayes, procedure in data mining were experimented and also compared to the accuracy and error. The 10-fold cross-validation was employed in the experimentation

Table 1. Classification results

Method Used	Correct Classification %	Detection Time (in seconds)
Naive Bayesian	96.2	0.52
SVM	93.4	0.23
KNN	92.6	0.25
proposed methods	98.7	0.21

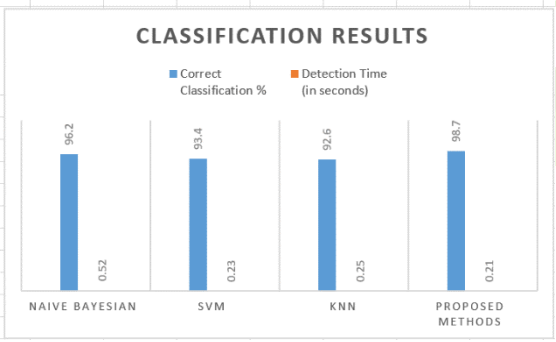

Figure 2: Classification results

Conclusion

The training process and tuning the Machine Learning from the standard data set are the basis for the generated model and it has an acceptable percentage of classification at the time of the evaluation of the prototype in a production environment with real information. The selection of metrics in the intrusion detection problem: false positive rate, false negative rate, rate classification, ROC curves, allow having a standard of comparison against other models. The application of techniques with supervised training as SVM model, has large advantages over the technique based on rules, since the generation of the model is based on a statistical model that changes its behavior according to the input parameters defined in the training and based on rules it requires human interaction. In the prototype evaluation was found a better classification rate for normal and anomalous requests in the training phase, is directly related to standardization and proper selection of input parameters, allowing output variables to be generated with minimum percentage of misclassification, generating reliability in the generated model and the detection of these behaviours.

REFERENCES

[1]. [1] Wani, A. R., Rana, Q. P., Saxena, U., & Pandey, N. (2019). Analysis and Detection of DDoS Attacks on Cloud Computing Environment using Machine Learning Techniques. 2019 Amity International Conference on Artificial Intelligence (AICAI). doi:10.1109/aicai.2019.8701238

[2]. [2] WU ZHIJUN , XU QING , WANG JINGJIE , YUE MENG (2020) Low-Rate DDoS Attack Detection Based on Factorization Machine in Software Defined Network, Digital Object Identifier 10.1109/ACCESS.2020.2967478, VOLUME 8, 2020

[3]. [3] Yadav, S., & Subramanian, S. (2016). *Detection of Application Layer DDoS attack by feature learning using Stacked Auto Encoder. 2016 International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT).* doi:10.1109/icctict.2016.7514608

[4]. [4] Bakker, J. N., Ng, B., & Seah, W. K. G. (2018). Can Machine Learning Techniques Be Effectively Used in Real Networks against DDoS Attacks? 2018 27th International Conference on Computer Communication and Networks (ICCCN). doi:10.1109/icccn.2018.8487445

[5]. [5] Khuphiran, P., Leelaprute, P., Uthayopas, P., Ichikawa, K., & Watanakeesuntorn, W. (2018). *Performance Comparison of Machine Learning Models for DDoS Attacks Detection. 2018 22nd International Computer Science and Engineering Conference (ICSEC).* doi:10.1109/icsec.2018.8712757

[6]. [6] Yuan, X., Li, C., & Li, X. (2017). *DeepDefense: Identifying DDoS Attack via Deep Learning. 2017 IEEE International Conference on Smart Computing*

(SMARTCOMP). doi:10.1109/smartcomp.2017.7946998

[7]. [7] Subbulakshmi, T., BalaKrishnan, K., Shalinie, S. M., AnandKumar, D., GanapathiSubramanian, V., & Kannathal, K. (2011). *Detection of DDoS attacks using Enhanced Support Vector Machines with real time generated dataset. 2011 Third International Conference on Advanced Computing*. doi:10.1109/icoac.2011.6165212

[8]. Hoyos Ll, M. S., Isaza E, G. A., Vélez, J. I., & Castillo O, L. (2016). *Distributed Denial of Service (DDoS) Attacks Detection Using Machine Learning Prototype. Advances in Intelligent Systems and Computing, 33–41*. doi:10.1007/978-3-319-40162-1_4

[9]. Suresh, M., & Anitha, R. (2011). *Evaluating Machine Learning Algorithms for Detecting DDoS Attacks. Communications in Computer and Information Science, 441–452*. doi:10.1007/978-3-642-22540-6_42

[10]. Rajawat A.S., Upadhyay P., Upadhyay A. (2021) Novel Deep Learning Model for Uncertainty Prediction in Mobile Computing. In: Arai K., Kapoor S., Bhatia R. (eds) Intelligent Systems and Applications. IntelliSys 2020. Advances in Intelligent Systems and Computing, vol 1250. Springer, Cham. https://doi.org/10.1007/978-3-030-55180-3_49

[11]. Narasimha Mallikarjunan, K., Bhuvaneshwaran, A., Sundarakantham, K., & Mercy Shalinie, S. (2018). *DDAM: Detecting DDoS Attacks Using Machine Learning Approach. Advances in Intelligent Systems and Computing, 261–273*. doi:10.1007/978-981-13-1132-1_21

[12]. Abid, K.: An efficient intrusion detection using J48 decision tree in KDDCUP99 dataset. Int. J. Emerging Technol. Adv. Eng. 6(2), (2016)

[13]. Sumathi, S., & Karthikeyan, N. (2020). *Detection of distributed denial of service using deep learning neural network. Journal of Ambient Intelligence and Humanized Computing*. doi:10.1007/s12652-020-02144-2

[14]. A. Singh Rajawat and S. Jain, "Fusion Deep Learning Based on Back Propagation Neural Network for Personalization," 2nd International Conference on Data, Engineering and Applications (IDEA), Bhopal, India, 2020, pp. 1-7, doi: 10.1109/IDEA49133.2020.9170693.