Detection of DDoS Attacks Using Supervised Learning Technique

M A Prriyadarshini, S Renuka Devi
School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, India
E-mail: renukevi.s@vit.ac.in

Abstract. The development in digital technologies has been amplified by the drastic growth in the field of digital media applications and it changes the entire life style of the people. The electronic devices and digital contents are becoming more notable and popular among the users. As an emerging research area, security has attracted a lot of attention towards security professionals, experts and practitioners. In spite of this increased interest, this field still faces manifold challenges and issues. Such a fast development in digitization contains a block with attacks from an indirect security attack called Distributed Denial of service. An attack that damages the targets, services or network. The main objective of this paper is to improve the detection rate of the DDoS attacks and to overcome these threats in order to provide a safe network. So, a novel method is proposed in this paper to detect the DDoS attacks using support vector machine. A simulation of a network traffic is created to track and capture the patterns from any type of DDoS attacks with the help of rule creation and blacklisting. Further, using rule creation and SVM model, the simulation result based on different instances of time shows the observations of the real time requests that improves the efficiency of detecting the DDoS attacks.

1. Introduction

Everything in this world needs a protection today. As we are developing the digital network for easy communications, we have to key in the suitable security measures. Distributed Denial of Service Attack is one such attack that attacks the victim’s resources and cause severe damages by flooding their networks, sending more messages and requests. The motivation behind DDoS attacks have major impact in the society. The victim has to undergo several consequences after the attack has been made. To overcome this issue, we have created a simulation to examine the attacks through a simple technique with the help of machine learning model. In this paper, we focus on detecting the DDoS attacks using a supervised learning technique to provide a safe network.

2. Distributed Denial of Service Attacks

Security and Digital world have to be the co-workers in the upcoming technology hunt. We need to develop and seek the capabilities of network security to secure the network from various DDoS attacks. When multiple systems attack the victim by taking their host networks or server down either
by sending more requests which set to be valid, or flooding the sites with more data packets. The main aim of this attack is to render all the services and websites unavailable. This can be done either by sending large number of messages, requests, data packets to website, or even a to flood a site with more network traffic. The OSI model consists of seven layers, and DDoS attack takes place in three major layers. Namely Layer: 3 Network layers, Layer: 7 Application layer and Layer: 4 Transport layers. The DDoS attack acts like a traffic that is clogged and prevents the normal traffics to reach their destination. It gains control over the victim’s system and infects it with malware. This technique is used since the beginning of the attacks through Botnet. It connects all the internetworked services into one or more bots and turns all the attacks into Zombies. Then, these Zombies establishes a call that communicates through a remote control affecting the victim’s server via multiple requests eventually that ends up in overflowing capacity. This will result in Distributed Denial of Service Attack from a normal network traffic.

2.1. Motive Behind DDoS Attacks
There are several ways in which DDoS attacks can happen. The main motivation differs from hacker to hacker. It can be because of Business feuds which results in affecting one’s companies’ growth, Cyber vandalism that leads to an absurd defamation of an individual. Extortion/Shakedown forcing a hacker to perform the attack in return giving them lumpsum amount. Indulging in Cyber warfare activities, Ideology/persuasion for their pleasure. Stealing victim’s data. With all these motivations behind the attack in the foreseeable future DDoS attacks will happen for money. It will be the root of all the evil that help hackers achieve their goal.

2.2. Types of DDoS attacks
There are three major types of DDoS attacks. They are,

- **Volume based Attacks**
  It is one of the most common attacks even today. This attack sends enormous amount of data from the available network’s bandwidth. This leads to a congestion leading to an amplification that requests to send massive amount of traffic to the victim’s services which will affect the responses and results in clogging of the server.

- **Protocol Attacks**
  Protocol attacks are also known as State exhaustion attack that affects the network layer of the victim’s system. This disrupts the tablespaces capacity such as the firewalls, and load balancers that will forward the requests to the target users’ system. This way the victim’s system gets affected for use.

- **Application Layer Attacks**
  Disruption of victim’s website or service with huge number of requests until it reaches the exhaustive stage of the request sent. The requests might require downloading large data or requesting the data in database. If the target is attacked with million requests at a time, it will slow down the process of the system and can even get locked immediately.

2.3. Challenges with DDoS Attacks
DoS attacks are Dynamic in nature; hence we cannot detect the actual IP address from where the attack has been originated. Its source can be spoofed, and It uses more than one system(node) as a source from different places over the internet. It’s source sometimes is compromised to launch attacks too. The type of attack is very complex since it is from different sources and requires human
intervention [3]. In 2017, there is a lack of agile granular responses that may increase financial problems for processing power when any system is affected to DDoS attack.

3. Machine Learning
Ability of a system to learn itself, and improve based on the learning without comprehensive programming is known as Machine learning. There are four different types of machine learning methods adopted. They are Supervised, Unsupervised, Semi-supervised and Reinforcement methods. These methods help in solving various problems in the IT sector effectively.

3.1. Classification of Machine Learning Algorithms

- **Supervised Learning Techniques**
  One among the classifications is SUPERVISED LEARNING TECHNIQUES. The aim is to collect data and learn from it using various Machine learning based supervised techniques. Some commonly used algorithms are Nearest Neighbor, Decision tree, Support vector machines, Naïve Bayes, linear regression [16].

- **Unsupervised Learning Techniques**
  In this method, the data are unlabeled, the system trains itself and produces the output. This helps in picking the important data that is required for the analyses [18]. There are various clustering algorithms like association rules, k-means clustering.

- **Semi-supervised learning Techniques**
  It is a combination of supervised and unsupervised learning methods where we will use both labelled and unlabeled data [17].

- **Reinforcement Learning Techniques**
  Reinforcement training is based on trial and error method for a particular decision. It gains experiences from the previous trainings and gives accurate knowledge based on the response received.

4. Detecting Distributed Denial of Service Attacks using Support Vector Machine
Supervised method is popular and used widely in solving today’s real time problems. This learning method collects an input and output data. It creates a rule that it will help map the input to the output via a modelled algorithm. The training will continue until there is an accurate result with high performance. There are two types of supervised learning methods. They are Classification, that allows us to separate the data and Regression, that will fit the data. This will give us fast and accurate results.

With a rapid change in the internet development, there are high chances of cyber-attacks especially DDoS attacks will happen that may affect an entire network. This type of attack uses various fake packets, IP addresses globally and targets the victim’s network, website or services. In order to reduce this type of attack, it should be detected in an early stage to avoid any damage to the system. Hence, supervised learning method will help detect the DDoS attacks accurately because of the nature of supervised learning methods. The related works are mentioned in the literature review section that gives an overview of various research works associated with machine learning methods that helped with the detection process providing a result that shows the performance, training and accuracy of the machine learning model. Various techniques and their advantages and disadvantages are observed that
gives the overall view of detecting the Distributed Denial-of-Service Attacks through machine learning algorithms. In this paper, we use a supervised learning technique to detect the DDoS attacks. Support Vector Machine helps to map data to a high dimensional space to categorize it even if it could not be separated linearly.

4.1. SVM (Supported Vector Machine)
Support vector machine is used for both classifications as well as regressions. In this method the data is plot in an N-D space. It computes the number of features that will evaluate each element in the training model that has a specific range. Further, the classification takes place in two distinct classes to perform hyper-plane activity. In Support Vector Machine, it is difficult to contain a direct hyper plane activity between the features model class. This method has a model called KERNEL. Its job is to convert low-dimensional information into high dimensional information. That is, it will become a distant model. This separation process is known as kernel. It is helpful in classifying non-dimensional issues. Hence, the nearest point to the margin distance is chosen from the hyper plane. This helps in viewing the data with broader scope. The sunset used when training adds more efficiency and that becomes one of the advantages to have better accuracy.

Fig 4.1 Detection of DDoS attacks methodology
Fig 4.1 describes how the DDoS attack will be collected and processed for training to produce a trained model output to detect the attack in an early stage. The existing model uses NAIDA method to detect the abnormal network traffic of the system using the network sniffer. This is executed using the behaviour of the client and it filters the packets and detects the DDoS attacks. Whereas, the proposed system uses the trained dataset of support vector machine using the weka jar, and based on the behaviour of the type of attacks. Thus, this system improves efficiency to detect the DDoS attacks by classifying the type of attacks and blocks it and reports the type of worm detected.

5. Related Work
Umarani and Sharmil [1] has proposed a new model to evaluate the flow of the network by creating an access matrix from the network traces that they have gathered. This paper uses two important aspects that helps in evaluating the performances. They are using the False Positive Rate to highlight their work. Two classification algorithms are used in this paper. They are K-nearest neighbor and Naïve Bayes. The Key features of this method includes selection of core attributes using PCA, increase in Detection rate and average FPR. The constraints are getting the actual attack signatures especially in the newer attacks. Hence, the flow characteristics in the low layers are not adequate to differentiate the
Application layer DDoS attacks from the normal flash crowd event. This is justified in the paper that helps solve the attacks with a solution.

Robinson and Thomas [2] gave a brief explanation on the machine learning algorithms with an aim to produce error free detections using precision and recall to find the accuracy. The authors also evaluate all the trained set using promethee, an aided software that helps in evaluation. The main advantage that is computed in this paper is usage of header parser to trace the networks and this is modelled to extract the required features for the analysis. This helps in finding the accuracy over stream of packets into flow. Only constraint is the normalization of dataset is not required, and the module does not require any distinguish attributes between flooding attacks from normal traffic attacks.

Alkasassbeh et al., [3] proposed a data mining based technique. Network simulator (NS2) is used in this work as it has the capability of producing effective results which reflect a real environment. The collected knowledge has been recorded for various forms of attack that focuses on the network and application layers. MLP, Naïve Bayes and Random forest are applied using the collected dataset to classify the types of DDoS attacks. The advantages are collection of all network traffics and auditing the NIDS before training the datasets helps to produce high accuracy level, and precisions for minority classes to detect the types of attacks easily. The constraint is inconsistent accuracy rate especially when the data are uneven. Hence, the precision and recall were calculated for each class types of attacks separately.

Khuphiran et al., [4] trains the model to find the best fit from the trained algorithms. This model results in providing the best suitable algorithm to detect the DDoS attacks based on the speed by which the model gets trained. They use two methods, one using SVM and with Deep Feed Forward (DFF) which is a deep learning classifier. The advantage of this paper is its analyses method on the packers and IP addresses. But, only few pcap files are processed that shows the traffic features with separates into two time periods for distributions. This shows every time we have to run the attack in a traffic to make sure it is fit for the model.

Ahmed et al., [5] uses Machine learning algorithms for more effective training and to apply ML techniques in many classification problems. The exciting features of SDN can be augmented with the deep learning algorithms for intelligent mitigation of DDoS attacks. The advantages is that the amplification attacks which are capable of paralyzing the entire network can be efficiently detected by using the traffic features surveyed through this work. [15] The constraint is that the SVM Count of flow entries per unit time, Count of distinct Source IP and Source Port for unit time, Ratio of pair wise flow entries to total entries, Standard Deviation calculated for count of packets and bits per period does not give accurate results when compared to Deep learning techniques.

Sreeram et al., [6] has used and Bio inspired Bat algorithm. In this technique, a novel approach is proposed to detect DDoS attack in the Application layer by invoking the HTTP floods. Bio-inspired bat algorithm is used to achieve efficiency in performance the advantage is that the customization of the BAT algorithm to train the data and test it to check if the attacks have happened or not, or it is a normal attack is a great advantage for this method. This help in reducing unwanted data and takes only selected features to detect the DDoS attack to process with minimum complexity. The constraints are the specificity and comparison of F-measure test resulted after evaluation shows only the positive labels of one group. That is before training the data, and after training the data the resulted set shows deviation in the statistical metrics.

Singh and De [7] deals with blacklisting the IP addresses from multiple HTTP requests that triggers various attacking services in the targeted system. From the attack that has affected the target’s system
it checks whether the data is requesting from suspect mode or By-pass mode. In order to find who sent the request (Human or BOT) a concept called CAPTCHA is introduced to verify the system. The advantages are that the backlisting of IP addresses, and HTTP counter filters are done to differentiate between normal attack and suspected attack in the IP addresses. This helps in backlisting the addresses using CAPTCHA technique to differentiate from human intervention and BOTs. The constraint is that there is a restriction of access to many data and only less data is taken into account for backlisting of IP address. Hence the accuracy rate will be less when compared to large datasets.

Ramamoorthi et al., [8] proposed a method using Enhanced Support Vector Machines with string kernels. The author tested the Application layer and Network layer Attacks with classification technique and ensured to provide 99% accurate results. Subbulakshmi et al., [9] proposed a unified approach using enhanced support vector machines and filtering mechanisms. The authors explain the automatic detection of DDoS through mechanisms. They use spoofed and non-spoofed data for detecting the attacks. For this, they make sure of the enhanced SVM and HCF mechanisms. The strength of the attack is calculated using the defense mechanism called as Lanchester law. This online approach of the system helps in detecting the DDoS attack and automatically drops the filtered packets. The advantage of this proposed model is usage of Hop Count Filtering (HCF) to detect the spoofed IP addresses. This mechanism dives into the clustered addresses in the mapped table.

Khare and Oak [10] proposed a model using the decision tree classifier. The advantage of this model is that it detects accurately with the live system in real conditions. They analyses the data mining approach to detect the DDoS attacks.

6. Proposed Model

DoS attacks are Dynamic in nature; hence we cannot detect the actual IP address from where the attack has been originated. Its source can be spoofed, and it uses more than one system(node) as a source from different places over the internet. [3] It’s source sometimes is compromised to launch attacks too. The type of attack is very complex since it is from different sources and requires human intervention. In 2017, there is a lack of agile granular responses that may increase financial problems for processing power when any system is affected to DDoS attack. It is pretty hard to overcome these attacks after the damage has been occurred. But, in this paper, we undergo a thorough analysis of classification and create a DDoS attack detector using Support Vector Machine Classifier Model to detect the malicious attacks that affects the users. Fig. 6.1 shows the architecture of the proposed model.

For this, we develop a **HoneyPot detector** that acts as a real system to probe and attack the users with malicious attacks.

**Connection Establishment**

i) Create a **client** side

ii) Create a **server** side

iii) **Router** to connect with the IP Address

iv) Send the file to the Server and Router, **Check** for DDoS Attacks

v) **Detection** using SVM model

A proxy server is created, it contains the preprocessed data. For this, **Support Vector Machine** (SVM) supervised learning technique is used. To detect the DDoS attacks, the trained dataset is fed into the Proxy server created. Feature extraction takes place. The proxy server will compare the live data with the trained data which is fed into the Proxy server and it gives the results based on the efficiency and accuracy to detect both active and passive attacks in the network layer. It classifies the
SVM Algorithm in Weka Tool and the dataset contains selected features requires for the processing of detection, that trained dataset will be tested in the detector code created using Java by importing weka.jar file in Java. It checks whether there are any malicious attacks present in the file. If any MALICIOUS ATTACKS is detected SVM classifier detects and prints “Malicious Attack detected” else it prints “Malicious Attack not detected” in the HoneyPot Application.

Fig 6.1 Architecture of the proposed method

In the existing system, the honey pot detector’s hybrid framework does not support any rule set because there is no process for identifying the attacks in the network.[26] This problem raises day by day as the data is sent over to the server, due to this many companies could not find what type of attack it is. Whereas, in the proposed system the attacks are collected and pre-processes the network traffic and predicts the attacks. The rule creation of the detector in the simulator finds the rate efficiently using the help of the rule set of SVM model. Thus, The SVM identifies this attack and increases the reliability and security of the network.

The process of forwarding the attack requests to the balancer detects for the traffic in the server. The experiment results show the various parameters uses and displays the work of the simulator. [19] By using this proxy in the simulator, it is easy to identify the attack present in the information which will be sent by the user. This will be blacklisted in future usage. This identification of attacks makes sure that the user will not be allowed to send any other data further to others and ensures security.

This is then routed to the router for further filtration. Later based on the nature of the attack it creates a dynamic schema using rule creation method. This technique blacklists the IP addresses and blocks them from reaching the server.

7. Experiment Results

As proposed in the novel method, the SVM model is used along with the rule detection algorithm using javaML. The network proliferation is done for all the type of attack that’s been detected. This results in a method to solve the detection issues associated with the DDoS attacks.

We propose a HoneyPot based SigFree using SVM, a real- time blocker for preventing DDoS attacks. As we mentioned earlier in the proposed model, the SVM model will be trained to detect the malicious attacks based on the condition of network features for 1-time window and based on previous window information packets after connecting to the Router (Preprocessing in Weka Tool) After entering the
server and router IP addresses, the setup file is selected. While sending it to the router, the proxy server blocks the path and the protocol, the detected time of the virus and time is displayed and the router does not get any information further. By, this way we can detect and find what type of DDoS attack is sent by the intruder and at what time they have sent the attacks is displayed.

![Detection of DDoS Attacks using Supervised Learning Technique](image)

Fig 7.1 Client-Side Simulation

Figure 7.1 shows the client-side simulation on how the system gets connected to the server. The below figure 7.2 shows whether an attack is detected or not. Before an attack is detected the original file details are displayed. After scanning, if the file contains any attack the malicious attack it trains and executes what threat it contains and displays that the attack is displayed. The network administrator then collects all the relevant information related to the threat. The proposed system creates a framework which prevents the vulnerable and hostile situation even before the attacker attacks the network. Blacklisting takes place, and blocks the user from receiving the cyber-attack. The network sniffer generates the results.
Fig 7.2 The Detection of DDoS Attacks in the simulation

Fig. 7.3 shows the server end of the simulation. If the message is sent without any attack the receiver gets the exact same copy of the file.
If there is any attack sent along with the message that is passed to the server, the proposed model for detection of DDoS attacks dataset that is trained using SVM classifier works here. Each time the detector checks for the attack it detects whether the trained dataset contains any malicious attack or not. When it detects that there exists an attack, the proxy server blocks and displays the following
information about the attack. The details of the attack are shown in Fig. 7.4.

![Fig 7.4 Details of the Attack](image)

While improving the efficiency of the simulation, the trade-off stands for certain to stay, and selecting the proper balance will likely still be a difficult task for the client. It's the users' opinion to choose this kind of system which can provide an additional security tool, however, a decent deal of streamlining work for detecting the DDoS attacks remains before it may be widely deployed.

8. Conclusion
As technology is experiencing rapid changes every day, the world has to face different security issues and cyber-attacks. An increase in DDoS attacks is measured and it is reported that 8.4 million cases have been recorded in the year 2019. In this research work, in significance of the rule detection algorithm generated it is compared with the SVM model thus it helps in detection of the DDoS attacks using the proposed model. Hence, the novel method brings out the key factors to detect the attacks with SVM classifier simulator to detect the DDoS attacks efficiently by improving the standards using machine learning technique.

9. References

[1] Umarani S and Sharmila D 2014 Predicting application layer DDoS attacks using machine learning algorithms. *International Journal of Computer, control Quantum and information Engineering* 8.10

[2] Robinson RR and Thomas C 2015 Ranking of machine learning algorithms based on the performance in classifying ddos attacks *Recent Advances in Intelligent Computational Systems IEEE* pp. 185-190

[3] Alkasassbeh M, Al-Naymat G, Hassanat A B and Almseidin M 2016 Detecting distributed denial of service attacks using data mining techniques *International Journal of Advanced Computer Science and Applications* 7.1 pp.436- 445

[4] Khuphiran P, Leelaprute P, Uthayopas P, Ichikawa K and Watanakesuntorn W 2018 Performance Comparison of Machine Learning Models for DDoS Attacks Detection *Inf.*
Conf. on Computer Science and Engineering Conference IEEE pp. 1-4

[5] Ahmed M E, Kim H and Park M 2017 Mitigating DNS query-based DDoS attacks with machine learning on software-defined networking. IEEE Military Communications Conference pp. 11-16

[6] Sreeram I and Vuppala V P K 2017 HTTP flood attack detection in application layer using machine learning metrics and bio inspired bat algorithm Applied computing and informatics

[7] Singh K J and De T 2015. DDOS attack detection and mitigation technique based on Http count and verification using CAPTCHA Inf. Conf. on Computational Intelligence and Networks IEE pp. 196-197

[8] Ramamoorthi A, Subbulakshmi T. and Shalinie S M 2011 Real time detection and classification of DDoS attacks using enhanced SVM with string kernels Inf. Conf. on Recent Trends in Information Technology IEEE pp. 91-96

[9] Subbulakshmi T, Parameswaran P, Parthiban C, Mariselvi M., Anusha JA and Mahalakshmi, G 2013 A unified approach for detection and prevention of DDoS attacks using enhanced support vector machines and filtering mechanisms ICTACT Journal on Communication Technology

[10] Khare M and Oak R 2020 Real-Time Distributed Denial-of-Service (DDoS) Attack Detection Using Decision Trees for Server Performance Maintenance In Performance Management of Integrated Systems and its Applications in Software Engineering Springer, pp 1-9

[11] Cheng J, Xu R, Tang X, Sheng V S. and Cai C, 2018 An abnormal network flow feature sequence prediction approach for DDoS attacks detection in big data environment. Computers Materials & continua, 55.1 pp.95-119

[12] Sachdeva M, Kumar K and Singh G 2016 A comprehensive approach to discriminate DDoS attacks from flash events Journal of information security and applications 26 pp.8-22

[13] Tang X, Cao R, Cheng J, Fan D. and Tu W 2019 DDoS Attack Detection Method Based on V-Support Vector Machine International Symposium on Cyberspace Safety and Security pp. 42-56.

[14] Myint Oo M, Kamolphiwong S, Kamolphiwong T and Vasupongayya S 2019 Advanced Support Vector Machine-(ASVM-) Based Detection for Distributed Denial of Service (DDoS) Attack on Software Defined Networking (SDN) Journal of Computer Networks and Communications

[15] Rivas P, DeCosutis C, Oakley M, Antaki A., Blaskey N, LaFalce S. and Stone S 2019. Machine Learning for DDoS Attack Classification Using Hive Plots IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference pp. 0401-0407.

[16] Mirjalili S, Faris H and Aljarah I 2020 Introduction to Evolutionary Machine LearningTechniques. In Evolutionary Machine LearningTechniques Springer pp. 1-7

[17] Yu F R and He Y 2019 Introduction to Machine Learning In Deep Reinforcement Learning for Wireless Networks Springer pp. 1-13

[18] Wang L, Zhu Y and Pan H 2019 Unsupervised reinforcement learning for video summarization reward function Proceedings of the 2019 International Conference on Image, Video and Signal Processing pp. 40-44

[19] Feinstein L., Schnackenberg D, Balupari R and Kindred D 2003. Statistical approaches to DDoS attack detection and response In Proceedings DARPA information survivability conference and exposition IEEE 1 pp. 303-314
[20] Lewis D D 1998 Naive (Bayes) at forty: The independence assumption in information retrieval. *In European conference on machine learning* Springer pp. 4-15

[21] Kalmegh S 2015 Analysis of weka data mining algorithm reptree, simple cart and randomtree for classification of indian news *International Journal of Innovative Science, Engineering & Technology* 2.2 pp.438-446

[22] Mahajan A and Ganpati A 2014 Performance evaluation of rule-based classification algorithms. *International Journal of Advanced Research in Computer Engineering & Technology* 3.10 pp.3546-3550

[23] Dogra A K and Wala T 2015 A comparative study of selected classification algorithms of data mining *International Journal of Computer Science and Mobile Computing* 4.6, pp.220-229

[24] Bhargava N, Sharma G., Bhargava R. and Mathuria M 2013. Decision tree analysis on j48 algorithm for data mining *Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering* 3.6

[25] McClish D K 1989 Analyzing a portion of the ROC curve. *Medical Decision Making* 9.3 pp.190-195

[26] Arivanantham P Attack Identification in Peer to Peer Network based on Chaotic Analysis.