IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

Detection of DDoS Attack using Machine Learning: A Review

Chiti Gupta[1], Dr. Aditya Vidyarthi[2], Dr. Jitendra Singh Kushwah[3], Dr. Ashish Gupta[4]
Department of Information Technology[1,2,3,4]
Institute of Technology & Management, Gwalior, India

Abstract: *Distributed Denial of Service, or DDOS attacks. A DDoS attack is a very disruptive form of attack, enabling the service to be rendered unsafe in some organizations that provide them with a connection to the internet community. This kind of threat keeps getting complicated and will most likely increase in number from one moment to the next. Thus, it becomes quite difficult to detect or devise a larger defense against this kind of threat. Hence, sophisticated intrusion detection systems (IDS) need to be developed, which will classify and identify abnormal internet traffic behavior.*
More complicated and personal, however, they are advancing their number from one day to another, and detecting and defending against such threats is difficult. Thus, it demands decent intrusion detection system programming to classify and identify abnormal internet traffic behaviors.

Keywords: Cyber Attack Detection, DDoS Attack Prevention, Intrusion Detection Systems, Machine Learning, Network Security,

I. INTRODUCTION

Internet services are become a component of both individuals' and businesses' daily need. The increasing demand for network services has led to a rise in attacks by network intruders who aim to prevent genuine users from using the services. DDOS attacks [1] are those that interrupt or cause network programs to operate more slowly. Attackers use the millions of publicly accessible computer systems on the internet to launch a denial-of-service assault. As a result, servers frequently ignore responses from authorized users and remain occupied with requests brought about by assaults. The most well-known websites are those of banks, social media, and academic institutions. To protect the important data and services from those attackers, it is therefore vital to deploy one or more security solutions in a computer network, such as firewall intrusion detection systems (IDS) or anti-virus software.

Figure 1: DDoS Attack

DOI: 10.48175/IJARSCT-24969

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

One common way to deal with DDoS attacks is to use an intrusion detection system (IDS) [2]. An IDS is a component that guarantees the availability, confidentiality, and integrity of Web services and computer network resources. A system that uses machine learning algorithms is able to recognize and categories DDoS attacks, hence removing the incursion. It is difficult to attain complete performance accuracy in terms of attack detection and categorization, nevertheless.

II. TYPES OF DDOS ATTACK

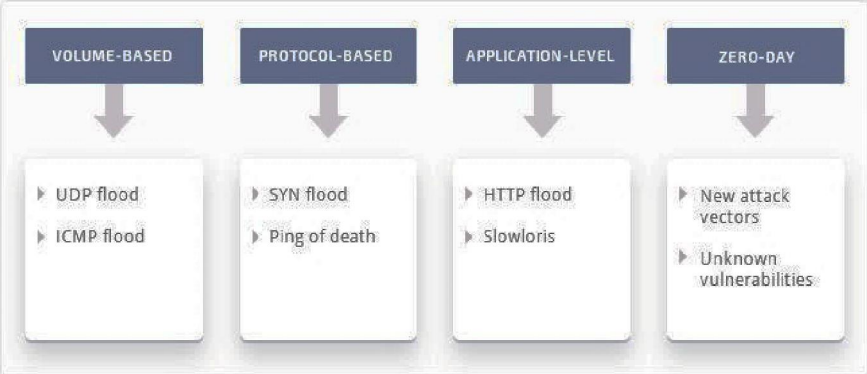

Figure 2: Types of DDoS Attack

UDP Flood:

A UDP flood [3] is a type of DoS volume attack in which the attacker uses IP packets containing UDP packets to assault and share the host's random ports. The way the hosts look for apps linked to certain datagrams during this attack is explained in Figure 3 below. The host returns an "Unreachable Destination" packet to the sender if nothing is discovered. The network will become completely unresponsive to the much-needed legitimate traffic as a result of this torrent bombardment.

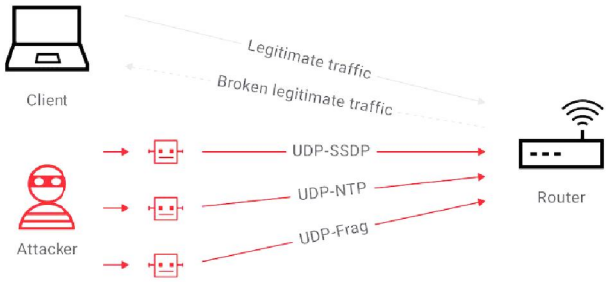

Figure 3: UDP Flood

ICMP(PING) FLOOD:

The most frequent type of assault is called a "ping flood," or "ICMP flood," in which the attacker bombards the victim's equipment with ICMP echo requests, or "pings," forcing it to shut down. The ICMP flood attack, which involves flooding the victim's network with request packets while knowing that the system will respond with as many reply

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

packets as it can, will be described in the figure. Hopeful and fearful are the file kinds used to capture a target down for ICMP [7] and requests utilizing proprietary software or code

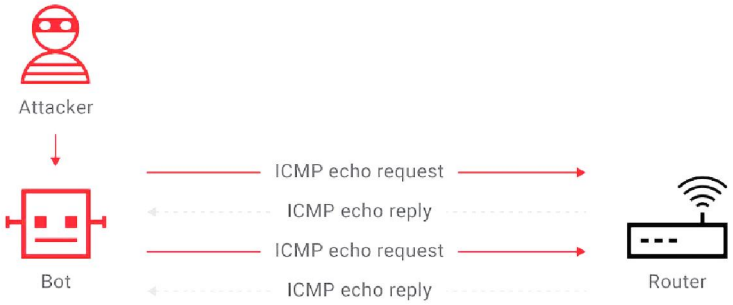

Figure 4: ICMP Flood

SMURF ATTACK:

This is another type of DDoS [9] assault in which a large number of Internet Control Message Protocol (ICMP) packets are transmitted to a computer network via an IP broadcast address, primarily utilizing the victim's spoof source IP. The majority of devices on a network will automatically reply by responding to the source IP address, as seen in the picture below. Traffic can overwhelm the attacker's computer if there are more systems than the size of the packet being received and responding to it.

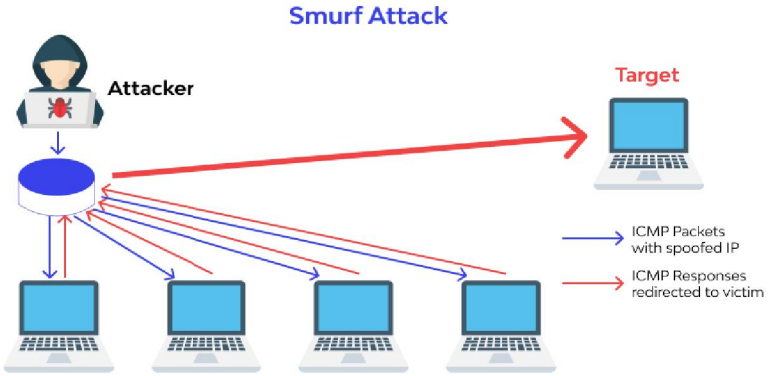

Figure 5: SMURF Attack

HTTP FLOOD ATTACK:

As seen in Figure 6, an HTTP Flood is a type of distributed volumetric Denial-of-Service attack. Its purpose is to overload a server with HTTP requests. Specific requests from actual users would experience denial-of-service whenever the target also becomes overloaded with queries and is unable to respond to normal traffic.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

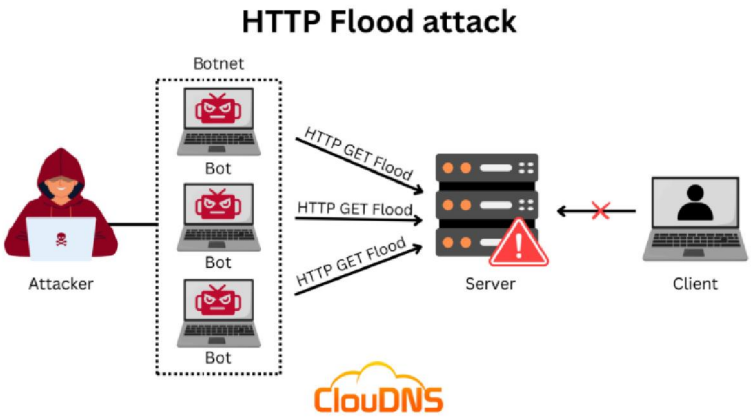

Figure 6: HTTP Flood Attack

III. MACHINE LEARNING METHODS RELATED TO DDOS ATTACK DETECTION

Signature-based IDS is typically a human operation that would take long hours of testing developing and deploying a signature and creating another signature for unknown attacks too. It becomes very much human-less. Anomaly-based IDS evolved from Machine Learning languages and solves the problem by providing such a framework that it will learn from data and predict unknown states of information on what has been learned data.

Naïve Bayes

Naive Bayes is centered around the Bayesian classification framework. It is a straightforward and efficient method for establishing classifiers, which assign class labels to instances represented as feature value vectors. These class labels are derived from a specific finite set. Hidden Naive Bayes (HNB) [11] approach yields more reliable outcomes compared to the Standard Naïve Bayes model. The HNB technique can be effectively employed to predict intrusion issues, such as Denial of Service (DOS) attacks, by leveraging strongly correlated dynamic characteristics and extensive network data stream capabilities.

The Naive Bayes algorithm is based on the principles of Bayesian classification. It is recognized as a simple and effective method for creating classifiers that assign class labels to cases represented as vectors of feature values, with these labels being derived from a limited set. The Hidden Naive Bayes (HNB) model demonstrates greater reliability than the Standard Naïve Bayes model. This technique can be utilized to forecast intrusion threats, including DOS attacks, by taking advantage of highly correlated dynamic features and robust network data stream capabilities.

Naive Bayes operates on the Bayesian classification model, providing a straightforward and efficient approach to classifier development. This method assigns class labels to instances, which are defined as vectors of feature values, with labels derived from a finite set. Hidden Naive Bayes (HNB) model offers more dependable results than the Standard Naïve Bayes model. The HNB technique is particularly useful for predicting intrusion issues, such as Denial of Service (DOS) attacks, by utilizing strongly linked dynamic characteristics and extensive capabilities in network data streaming.

Decision Tree:

One of the principal techniques that machine learning and data mining apply is the decision tree. Moreover, it may be used as a model of prediction in which findings about an object are mapped to inferences about the preferred value of the object. A decision tree can be applied to the investigation of decision data to draw out, in a very explicit and visual way, how decisions are made. In this method, a data set that is under study is developed. Therefore, if the new data element is given for classification the former dataset will classify it properly.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

The Decision Tree Algorithm was used to detect denial-of-service attacks. Using classification techniques, data mining is used to detect Denial of Service assaults. There are two types in this particular approach: "normal" traffic and "anomalous" traffic. In addition to discussing the J48 decision tree algorithm's efficacy in Denial of Service assaults, the article will compare it to other rule-based algorithms like Decision Table. Md. Dewan. Farid et al. [3] presented an anomaly-based network intrusion detection learning method in their study that uses a decision tree approach to identify various types of intrusions and stops attacks from routine activities. The KDD99 dataset, which is primarily designed for network intrusion detection, is employed in this study.

Artificial Neural Network:

ChandrikaPalagiri demonstrated that a modelling network may recover a believable result for illustrating a neural network [12], specifically for a given attack. Researchers are also concentrating on a neural network that can identify choices quickly and in real time.

One notable choice for the base classifier in a publication by Madhav Kale et al. is Resilient Back-Propagation, or RBP. For the actual classification decision, the paper's focus was on improving the RBP classifier's efficiency by combining the outputs of classifiers using Neyman Pearson's cost reduction strategy. Detection accuracy and cost per sample were the two metrics used to gauge the effectiveness of the RBP Boost classification system.

The purpose of this research by Md Salem et al. [15] was to investigate if a firewall could actually track its traffic patterns in order to identify targeted denial of service attacks. This paper used statistical studies of firewall logs for a large network to analyse a baseline of the network. This study used statistical analysis of firewall logs for a broad network to determine a baseline.

Using the Holt-Winter and linear regression techniques, estimated traffic rates were computed to compare with the baseline. The analysis's findings were positive, indicating a widespread campaign in the network based on the departure from the expected rejected packet rates.

Author Mohammad Masoud Javidi et al [17] Proposed IDS that employs the supervised neuralnetwork for malicious DDOS in the NSLKDD database. The researcher used a signature-based methodology in the proposed IDS. IDSs are designed with a neural network that can detect various forms of DoS attacks and have a different IDS for each one to identify the particular attack.

Support Vector Machine:

T. Subbulakshmi [9] first proposed the concept of an SVM, which attracted a lot of interest from the machine learning research community. This approach uses examples with target values to classify and regress supervised learning techniques. The SVM algorithm uses a group of trained examples to generate a design that divides a technique into two classes based on which new instances are expected to closely fit into one of the two categories.

In DOS attacks; Packets are first grabbed from the network, and data is instantly fed into the RST. The RST-chosen feature sets would then be used in training and testing with the SVM model, etc. The findings were subsequently tested.and showed to function through PCA). It shows that RST and SMV are well-capable of this task,and it increases the false positive ratefor improving upward efficiency.

Subbulakshmi et al[9] have published a paper towards

Online network monitoring and activation of a security approach at a timewhenever doubtful activity is there. This approach helps identify both the non-spoofed IPs and spoofed IPs. The ESVM-based method has been employed in thework for spoof detection mechanisms throughspoofs identification in IP Hop Count Filtering has also been taken into consideration by the author to identify the spoofs in IP. These IPs are used to initiate the defense. The Lanchester Rule is applied to determine the attack force applied to trigger the defense mechanism.

K-means clustering:

K-Means clustering is a subset of unsupervised machine learning approaches [12] that enable partitioning of samples into K sets in a manner that maximizes the similarities within the samples in the set.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

How K-means Clustering Works:

- **Initialization**: Choose K centroid either randomly or based on the average from each sample in the K groups.
- **Assignment:** It consists of grouping every sample to the K groups as per their closeness or distance to the K centroids.
- **Update:** Figure the average for every sample that was assigned to the K sections to enable determining the K centroids.
- **Repeat:** Observe step 2 and 3 continuously till the centroids stabilize or stop as per an alternative threshold.
- Key Characteristics of K-means Clustering:
- **Unsupervised learning:** Labeled data is not essential for its application [13].
- **Non-hierarchical:** Clusters don't conform to be children in other clusters.
- **Non-deterministic**: When set K is changed the clusters divined differ.
- **Advantages of K-means Clustering:**
- **Simple and efficient:** K-means is known for being quite simple and fast.
- **Effective for spherical clusters:** Essentially, K-means works for those clusters that are pinpointed by spherical shapes.
- **Disadvantages of K-means Clustering:**
- **Not suitable for non-spherical clusters**: Complex shapes tend to be hard for k to manage.
- **Requires choosing K**: When working with clusters, it is essential to pre-determine the cluster number K.

Real-World Applications of K-m

Figure7: Algorithm Overview

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

IV. CONCLUSION

Finally, conclusion in the following comprehensive analysis, it is evident that web attacks pose serious risks, and IDS and IPS may not be sufficient to handle these new attacks that infect networks. Moreover, there is a necessity for machine learning approaches so that the severity of these attacks can be understood correctly and so that enterprises implement measures to prevent specific attacks.

V. FUTURE ENHANCEMENT

Recent trends in computer network attacks particularly the more advanced ones i.e., XXX, DDoS, Smurf and UDP flooding, as well as the newest types in as HTTP ontology should be adequate at this point. Thus, it will be possible to increase the level of protection provided by the firewall configuration.

REFERENCES

[1] M. Alkasassbeh, G. Al-Naymat et.al,"Detecting Distributed Denial of Service Attacks Using Data Mining Technique", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, pp. 436-445, 2016. Science and Information Technologies, Vol. 6 (2), pp. 1096-1099, 2015.

[2] HodaWaguih, "A Data Mining Approach for the Detection of Denial-of-Service Attack", International Journal of Artificial Intelligence, vol. 2 pp. 99106(2013).

[3] Dewan Md. Farid, Nouria Harbi, EmnaBahri, Mohammad ZahidurRahman, ChowdhuryMofizurRahman, "Attacks Classification in Adaptive Intrusion Detection using Decision Tree", International Journal of Computer, Electrical, Automation, Control and Information Engineering, Vol:4, No:3, 2010.

[4] KiriWagsta,ClaireCardie,Seth Rogers,Stefan Schroedl, "Constrained K-means Clustering with Background Knowledge" Proceedings of the Eighteenth International Conference on Machine Learning, 2001, p. 577-584.

[5] Singh, S.K., Gupta, A.K. "Application of support vector regression in predicting thickness strains in hydro-mechanical deep drawing and comparison with ANN and FEM" (2010) CIRP Journal of Manufacturing Science and Technology, 3(1), pp. 66-72

[6] Ramesh.G, Madhavi, K. "Summarizing Product Reviews using NLP based Text Summarization", International Journal of Scientific & Technology Research, September 2019. (Scopus).

[7] Mangesh Salunke, RuhiKabra, Ashish Kumar. "Layered architecture for DoS attack detection system by combine approach of Naive Bayes and Improved K-means Clustering Algorithm", International Research Journal of Engineering and Technology (IRJET), Volume: 02 Issue: 03, June-2015.

[8] Ramesh G, Madhavi K., "Best keyword set recommendations for building service-based systems" International Journal of Scientific and Technology Research, October, 2019.

[9] T. Subbulakshmi et.al, "A Unified Approach for Detection and Prevention of DDoS Attacks Using Enhanced Support Vector Machine and Filtering Mechanisms", ICTACT Journal on Communication Technology, June 2013.

[10] T. Subbulakshmi, K. Balakrishnan; S.M.Shalinie,D.Anand Kumar,V.Ganapathi Subramanian, K. Kannathal."Detection of DDoS attacks using Enhanced Support Vector Machines with real time generated dataset", ICTACT Journal on Communication Technology, Volume: 04, Issue: 02 , June 2013.

[11]Kanagalakshmi.R, V. NaveenantonyRaj, "Network Intrusion Detection Using Hidden Naïve Bayes Multiclass Classifier Model," International Journal of Science, Technology &Management, Volume No.03, Issue No. 12, December 2014.

[12] A. Bivens, C. Palagiri, R. Smith, B. Szymanski, M. Embrechts, et al, "Network based intrusion detection using neural networks," Intelligent Engineering Systems through Artificial Neural Networks, vol. 12, no. 1, pp. 579–584, 2002.

[13] Ch. Mallikarjuna Rao, G. Ramesh, Madhavi, K., "Feature Selection Based Supervised Learning Method for Network Intrusion Detection", International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-8, Issue-1, May 2019.

[14] R Vijayasarathy, BalaramanRavindran, S.V Raghavan,"A System Approach to Network Modeling for DDoS Detection using a Naive Bayesian Classifier", Department of Computer Science and Engineering IIT Madras, India.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

[15] Mohammed Salem, Helen Armstrong, "Identifying DOS Attacks Using Data Pattern Analysis", Australian Information Security Management Conference Security Research Institute Conferences,2008.
[16] Thirupathi, N., Madhavi K., Ramesh G., Sowmya Priya, K. "Data Storage in Cloud Using Key-Policy Attribute-Based Temporary Keyword Search Scheme" (KP-ABTKS), Lecture Notes in Networks and Systems, 2020.
[17] Mohammad Masoud Javidi, Mohammad Hassan Nattaj, "Journal of mathematics and computer Science 6 (2013), 85-96.
[18] Madhavi.K., G. Ramesh, G. Lavanya" Load effectiveness on coverage-technique for test case prioritization in regression testing", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-7 May, 2019.