Realtime Traceback Mechanism for Multiple Attacks Using Machine Learning

Shubham Rastogi
Department of Computer Science and Engineering
Galgotias University
Greater Noida,India

Saurabh Prakash
Department of Computer Science and Engineering
Galgotias University
Greater Noida, India

Deepak Yadav
Department of Computer Science and Engineering
Galgotias University
Greater Noida, India

***Abstract*—: With the rise of organization based registering innovations like Cloud Computing, Fog Computing and IoT (Internet of Things), the setting of digitizing the classified information over the organization is being embraced by different associations where the security of that delicate information is considered as a significant concern. Longer than 10 years there is a monstrous development in the utilization of the web alongside the innovative progressions that request the requirement for the improvement of productive security calculations that could withstand different examples of the security breach. The DDoS attack is the main organization-based assault in the area of PC security that disturbs the web traffic of the objective worker. This examination mostly centers to distinguish the improvement and exploration holes in the advancement of proficient security calculations tending to DDoS attacks in different omnipresent organization conditions.**

Keywords*—DOS, DDOS, MALWARE, REALTIME TRACEBACK MECHANISM, MACHINE LEARNING*

I. INTRODUCTION

Presently with the coming of 4G, 5G organizations and financial shrewd devices there is a huge development in the utilization of the internet that has become a piece of day-by-day life. An immense scope of administrations gave over the internet in assorted application areas, for example, business, entertainment, and training, and so forth made it an indispensable segment in outlining different plans of action. This setting made security over remote organizations as the most significant factor while utilizing the internet from unstable associations. Many security frameworks and algorithms are created to empower assurance from Internet-based attacks while conceiving elite IDS (Intrusion detection systems) which go about as a cautious divider while defying the attacks over internet-baseddevices. Conveyed engineering-based figuring conditions like distributed computing and IoT are more inclined towards DDoS attacks in which various gadgets are facilitated to dispatch attacks over disseminated targets. Denial of Service attack is a purposeful attempt by malicious clients to totally upset or corrupt the availability of services to genuine users. DDoS is an extended type of DoS attack. It is a malicious attempt which disrupts the traffic of a network or a server by sending a large amount of traffic to the network or server from multiple resources. Now-a-days, the effect of DDoS attacks on the internet is rising unnecessarily. Such types of attacks are generally committed by a network of compromised computers known as Botnet. The goal of these types of attacks is to exhaust network services so that the network services become unavailable for the legitimate users. The marvel of spreading registering depends on the one-to-many measurement where these sorts of attacks may make a potential measure of harm to the server assets. It is seen from the past examination that the harm limit, just as the upsetting idea of the DDoS attacks, is continuously expanded with the pace of internet utilization [1,2].

II FEASIBILITY ANALYSIS

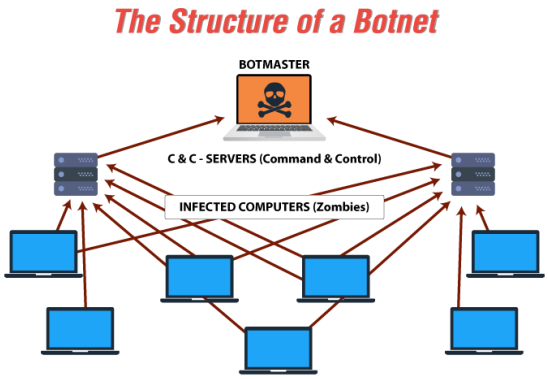
Machine learning is a technique in which computers use various algorithms to determine relationships, patterns and trends. The main objective of machine learning is to make computers learn how to make predictions about the possible trends, patterns and outcomes using the examples and given information datasets. Machine Learning Algorithms can be used to train, detect and prevent such DoS/DDoS attacks. When the server/network is under such attack, a report can be sent to the administrator. Machine learning classification algorithms can categorize if the occurring event is a DoS/DDoS attack or not. Machine learning models can

be trained using data sets of a variety of security events. As DOS/DDOS attacks not only makes the network services unavailable but can also lead to compromising sensitive and confidential information of the enterprise or the user, therefore early detection of such malicious attacks can be possible using machine learning models [3-5].

III. PROBLEM FORMULATION

A) DDoS Attacks

DDoS stands for Distributed Denial of Service. It is a type of cyber-attack. It is a malicious process to disrupt the usual traffic of a particular server or network with the help of Botnets. The attacker remotely controls multiple devices with the help of malware and these devices are called Bots and they are collectively known as Botnets. These Botnets send multiple requests and flood the server/network which ultimately makes the server busy and makes the traffic high. It is very similar to a forced traffic jam. As a result of the whole process, it makes the server unresponsive and this is called DDoS (Distributed Denial of Service) attack [6,7,8].



B) Malware

It is a type of cyberattack through which malicious softwares can lead to a structure through different types of techniques. E-mail attachments, software downloads from unrecognized websites, and operating system weakness are the most widely recognized sources of malware. Once established, malware covers up itself by appending itself to the real code and spreads to different parts of the system. The main purpose of malware is to concede unauthorized admission to a computer or system. Ransomware, a type of malicious software which restricts the client's access to the system by encrypting the confidential data until a fee demand is paid. Though, a variety of malwares like Trojans, worms etc. are continuously developing to challenge associations as well as people the same.

1. C) Phishing Scams

2. Phishing is a cyber-crime which uses E-mail to reach the host system and represents itself as a reputable/ genuine organization from where the host can relate itself like banks, University, Social Networking Sites and so on. The mail lures the host by various ways to tap on the certain link and fill the personal details such as id, passwords, account number, credit card details, phone number and other sensitive details. As the host enters the data it gets stored in the server of the attacker.

IV. REQUIRED TOOLS

A) Predictive Analysis

To successfully resist cyber threats, the people in the IT sector need to know how these threats resemble, when and from where it is coming from. This process is done through software which works on Machine learning and can predict the source as well as the reason for the malfunctioning. These software are very effective for DDoS attacks as it traces the similar IP Address of Botnets [9].

3. B) Backup Critical Data

Keeping in the mind of the DDoS and ransomware attacks it would be beneficiary to the organization/institution to have a backup plan.The basic goal behind backing up the crucial data is to restore the system and services quickly without any major problem. The organization should maintain a copy of the database and other essential stuffs time to time, as they would never be in condition of “All or Nothing” even after the attacks. This method would also beneficiary if the system collapses, Hard drive failure or the operating system failure [10].

C) SLA Assurances

The term SLA Assurance stands for service level agreement assurance. It is a type of agreement between two or among multiple parties where one is a customer and the rest are the service providers. This process can have formal legal contracts or any informal contract. These parties may belong to the same or different organization. There are three types of Service Level Agreements [11]:

Customer based SLA: This type of agreement is signed between a customer and a service provider agreeing all the relevant services the customer can use. It contains detailed information of the quantity and quality of the service [12].

Service based SLA: This type of agreement is provided by the service provider for all its customers.

Multi-Level SLA: This type of agreement is changed time to time to meet the requirement of end users of the company

D) Cyber Insurance

In today's digital world where hacking is turned into an acknowledged danger for almost all organizations. Any company can't afford the loss of data. So, as to ensure security one can buy Cyber Insurance. It is a type of insurance which is designed for business protection. Cyber Insurance companies provide a range of services like protection against identity theft, cyber stalking, malware attack, IT theft loss, Phishing, E-mail spoofing etc. The cyber protection market is required to develop to $20 billion by 2025 [13-14].

4. E) Training and Awareness

5. Cyber security awareness means how much the user at the end level knows about the threats and their consequences, risks which are going to be used world-wide. There are countless personal information leaks out due to a Phishing technique which inserts malware in the network system. Training & teaching the officials and workers about these cyber threats would decrease the probability that they would tap on any random or fake mail and tap any link and get trapped in any malicious act like Phishing. The organization ensures that all the employees are aware of the current threats and their respective methods/techniques to avoid them [15-17].

V.METHODS OF CYBER ATTACK DETECTION

A) SIGNATURE BASED DETECTION

Signature-based Detection means detection of attacks on network just by recognizing some known patterns like sequences of network traffic or some specific malicious instructions of some malwares. Its working is same as anti-virus software which determines these detected malicious patterns as signatures. Signature-based detection works on a pre-programmed list which contains recognized indicators of compromise (IOCs). Only limitation of signature-based detection is that it can only recognize those malicious attacks whose signatures exists in the framework or which are already known. It fails for any new malicious attack whose behavior changes or whose signatures are not there in the framework.

Intrusion Detection System

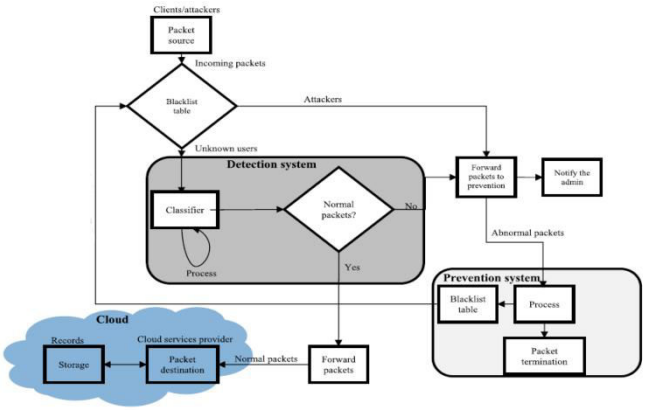
Intrusion Detection System is a security system that monitors network traffic and detects any violation of policy or any potential threats for the network or a malicious attack. When it detects any malicious activity, it generates alerts and reports to the network administrator. IDS can be hardware devices or software. A Security Information and Event Management system known as SIEM system, collects or reports all the malicious activities or policy violations on the network. Only limitation of IDS is that it can only detect malicious activities but can't respond to them. Some IDS are capable of responding to suspicious activities. In general IDS can be classified as Network Intrusion Detection System (NIDS) which usually monitors incoming network traffic and another one is Host Intrusion Detection System (HIDS) which is used to monitor important files of the operating system.

Intrusion Prevention System

An Intrusion Prevention System (IPS) is a network security tool, which monitors the security threats on the network, and after detecting threats it takes necessary action to prevent any malicious activity. Intrusion Prevention System can be a hardware or a software. When it detects any potential threat it prevents it, reports it to the network administrator. IPS is placed inline, in the flow of network traffic between the source and destination and it sits just behind the firewall [18]. There are different types of intrusion prevention system:

1) Network intrusion prevention system.
2)Host intrusion prevention system.
3)Network behavior analysis.
4)Wireless intrusion prevention system.

Intrusion prevention systems may also use a honeypot also known as decoy high-value data, which attracts malicious users and stops them from reaching their targets and in turn restricting them from performing malicious activities.



B) Anomaly-based Method:

Signature based methods were unable to detect new attacks or attacks which were not made known to signature based IDS. On the other hand, Anomaly-based detection is used to detect new attacks and threats whose

behavior changes. In anomaly-based IDS there is use of machine learning to make a trustful program model. Anomaly based detection involves training the model with a normalized baseline thencomparing activity against that baseline. and anything that is against will be considered as a threat. Machine learning based IDS are more powerful and beneficial than signature-based IDS.

C) Policy-Based methods:
Policy-Based methods are not so commonly used as signature-based methods and anomaly-based methods. In this method, security policies are employed by the administrator which are defined by an enterprise. Activities which go against these policies are blocked and dropped.

VI. CYBER ATTACK DETECTION USING MACHINE LEARNING

Using Machine learning algorithms, we can detect and traceback cyberattacks and react to them before they produce results in real-time. Most of the enterprises across are using cloud services to construct and handle software applications. Microservices is one of the most software development techniques and Application Program Interface (API) is one of the types of microservice used in numerous sectors such as banking sector, storage sector and healthcare sector. Several instances of microservices are automatically initiated whenever required. Under many circumstances, it becomes impossible for humans to monitor and detect all the disingenuous instances which in turn leads to a bigger cyber-security risk. A system of APIs works with an assumption that each of the routines will be called only a limited number of times in a day and this can be a necessary solution to these types of attacks. In case of API fails, troubleshooting or debug procedures are performed, number of API calls might increase but they should not go outside a range. We can make a basic assumption, if calls are invoked more than a limited number, then in such situations DOS/DDOS attacks may be possible. The Machine learning algorithm can be trained by utilizing log data to decide if the system is under attack or not based on some attributes. Log information of different microservices can be monitored using several log monitoring tools. Also, Multiple machines can be used to attack multiple APIs which are exposed by a target. Enterprises using API's, their sensitive and crucial information, can be compromised by DoS/DDoS attacks. DOS/DDOS attacks not only makes the services unavailable for legitimate customers but can also lead to data breach. Machine Learning Algorithms can be used to train, detect and prevent such DoS/DDoS attacks. When the server/network is under such attack, a report can be sent to the administrator. Machine learning classification algorithms can categorize if the occurring event is a DoS/DDoS attack or not

.VII. PROCEDURE AND STEPS (ALGORITHM):

STEP 1: EXPLORATION OF DATA SETS
The first step of the project will be exploration of data sets.In this section there will be an investigative analysis of the big data sets of security events i.e. DOS/DDOS attacks. Data sets can be acquired by using log data like Client's IP address,API request, date and time from the log data can be used as attributes which can be used by preprocessors in real-time, which in turn calculates the number of hits on a specific API for a specific date and time, and specific IP address A system of APIs works with an assumption that each of the routines will be called only a limited number of times in a day and this can be a necessary solution to these types of attacks. In case of API fails, troubleshooting or debug procedures are performed, number of API calls might increase but they should not go outside a range. We can make a basic assumption, if calls are invoked more than a limited number, then in such situations DOS/DDOS attacks may be possible.

STEP 2: REQUIRED MODEL DEVELOPMENT
In the second step we will use the Grid Search CV algorithm which is a pipelining algorithm for developing our required model. In this algorithm we will provide a list of algorithms like KNN, K-means,SVM, Naive Bayes. The parameters for the respective algorithms will be given in the form of a dictionary. So, using the above procedure we will develop our required machine learning model.

STEP 3: PERFORMANCE EVALUATION OF MODEL
In this last step, we will evaluate the model's performance. There are two parameters in the Research algorithm to evaluate a model's performance, first is 'FIT' other one is 'SCORE'.Using these two, first we will fit all the algorithms one by one that we are passing and we will store their scores in an array and in the last we will print the score values of all these algorithms that we have used for our model development. Algorithm which gives the highest score value will be selected for our dataset.Highest score means that the algorithm is performing well for a given dataset.

VIII. CONCLUSION

Throughout this research paper we made a study about all the possible methods of tracing different cyber security attacks and various methods that can be used to detect and block them. We also learnt and obtained several insights about machine learning models and how they are developed. The actual working model will be developed according to the above-mentioned sections.

IX. REFERENCES

1. John, A., and T. Sivakumar. "Ddos: Survey of traceback methods." *International Journal of Recent Trends in Engineering* 1.2 (2009): 241.
2. Yu, Shui, et al. "Traceback of DDoS attacks using entropy variations." IEEE transactions on parallel and distributed systems 22.3 (2010): 412-425.
3. Suresh, Manjula, and R. Anitha. "Evaluating machine learning algorithms for detecting DDoS attacks." *International Conference on Network Security and Applications*. Springer, Berlin, Heidelberg, 2011.
4. Kumar, P. Arun Raj, and S. Selvakumar. "Distributed denial of service attack detection using an ensemble of neural classifier." *Computer Communications* 34.11 (2011): 1328-1341..
5. Gurulakshmi, K., and A. Nesarani. "Analysis of IoT Bots against DDOS attack using Machine learning algorithm." 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI). IEEE, 2018..
6. J. Fan, W. Xu, Y. Wu, and Y. Gong (2010). Human tracking using convolutional neural networks. *IEEE Transactions on Neural Networks*, vol. 21, no. 10, pp. 1610–1623.
7. Jia, Yizhen, et al. "Flowguard: an intelligent edge defense mechanism against IoT DDoS attacks." *IEEE Internet of Things Journal* 7.10 (2020): 9552-9562.
8. A. John, M. Sugumaran and R. S. Rajesh, "Performance analysis of the past, present and future indexing methods for spatio-temporal data," *2017 2nd International Conference on Communication and Electronics Systems (ICCES)*, Coimbatore, India, 2017, pp. 645-649, doi: 10.1109/CESYS.2017.8321157.
9. Yusof, MohdAzahariMohd, Fakariah Hani Mohd Ali, and Mohamad Yusof Darus. "Detection and defense algorithms of different types of DDoS attacks using machine learning." *International Conference on Computational Science and Technology*. Springer, Singapore, 2017.
10. Li, L. I., and Su-bin SHEN. "Packet track and traceback mechanism against denial of service attacks." *The Journal of China Universities of Posts and Telecommunications* 15.3 (2008): 51-58.
11. Loukas, Georgios, and GülayÖke. "Protection against denial of service attacks: A survey." *The Computer Journal* 53.7 (2010): 1020-1037.
12. Sharma, Vishal, Vinay Verma, and Anand Sharma. "Detection of DDoS Attacks Using Machine Learning in Cloud Computing." *International Conference on Advanced Informatics for Computing Research*. Springer, Singapore, 2019.
13. Liu, Lei, et al. "Real-time diagnosis of network anomaly based on statistical traffic analysis." *2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications*. IEEE, 2012.
14. Singh, Karanpreet, Paramvir Singh, and Krishan Kumar. "A systematic review of IP traceback schemes for denial of service attacks." *Computers & Security* 56 (2016): 111-139.
15. Kale, Madhav, and D. M. Choudhari. "DDOS attack detection based on an ensemble of neural classifier." *International Journal of Computer Science and Network Security (IJCSNS)* 14.7 (2014): 122.
16. Abubakar, Rana, et al. "An effective mechanism to mitigate real-time DDoS attack." *IEEE Access* 8 (2020): 126215-126227..
17. Chen, Wen, et al. "A DDoS attacks traceback scheme for SDN-based smart city." *Computers & Electrical Engineering* 81 (2020): 106503.
18. Saied, Alan, Richard E. Overill, and Tomasz Radzik. "Detection of known and unknown DDoS attacks using Artificial Neural Networks." *Neurocomputing* 172 (2016): 385-393.).