Monitoring and Filtering Traffic Data using Big Data Analysis for DDoS

Harmeet Kaur Bawa[1]
GGSIPU, BVCOE,
New Delhi, India
Email id: 19harmeet96@gmail.com

Abhind A[2]
GGSIPU, BVCOE,
New Delhi, India
Email id: abhind.ambujakshan@gmail.com

Sagar[3]
GGSIPU, BVCOE,
New Delhi, India
Email id: er.sagarkoli@gmail.com

Shilpa Gupta[4]
GGSIPU, BVCOE,
New Delhi, India
Email id: shilpa.gupta@bharatividyapeeth.edu

Abstract- **As technology continues to advance rapidly, the graph of cybercrime has seen an exponential rise. To counter cyber-attacks such as Distributed Denial of Service, the chosen technique is based on Big Data. There is a rising trend of using prediction analysis since they are generally more accurate. This technique helps in reducing the amount of illegitimate data flowing to the server. Data traffic will be monitored using Big Data analysis at an early stage when data reaches the destination source. This algorithm has been designed to strive towards providing a solution against DDoS attacks as it provides a real time data analysis and maps the incoming data to previously machine-learnt traffic data for identifying attacks and preventing them from damaging the system using Big Data Analysis. This technique named as DPROS is provided for preventing UDP floods, Syn Floods, Ping of Death, HTTP flooding, NTP Amplification floods and Zero Day Floods.**

Keywords: Algorithm, Analysis, Big Data, DDoS Attack, Mapping, Traffic Analysis

NOMENCLATURE

1. PKT_SIZE: Packet size
2. NUMBER_OF_PKT: Number of packets
3. NUMBER_OF_BYTE: Number of bytes
4. PKT_RATE: Packet rate
5. PKT_AVG_SIZE: Packet average size
6. PKT_DELAY: Packet delay
7. FIRST_PKT_SENT: First packet sent
8. LAST_PKT_RECEIVED: Last packet received

I. INTRODUCTION

Machine and man continue to learn simultaneously, being reliant on each other as one creates technology and the other works at instigating losses. DDoS stems from its parent, the DoS attacks. **D**enial **o**f **S**ervice (DoS) is a cyber-attack where the perpetrator tries to disable or block the machine or network resource, which makes it unavailable to the users it is intended for.

These attacks are achieved by flooding the host machine or resource with exorbitant requests leading to overloading of the host machine or network resource due to which legitimate data is left unprocessed.

Distributed **D**enial **o**f **S**ervice (DDoS) attacks are analogous to overcrowding of a train by people without tickets due to which people with tickets are not able to enter it. Similarly DDoS attacks are distributed in nature i.e. these attacks are caused by use of more than one, rather millions of spoofed IP Addresses which attack on the target machine in order to exhaust its bandwidth or its resources like the processor or memory.[9]

The earliest Distributed Denial of Service attack can be credited to Creeper created by Bob Thomas (1971). [7] Creeper is a self-replicating program that moved through the TENEX operating systems which was facilitated with entry via the ARPANET. Since then, thousands of attacks have affected networks and machines. The largest recorded attack was of 602Gbps data flow to the main servers of BBC's website in January 2016.

A. The Vectors:

DDoS attacks are accomplished using botnets or IP spoofing. The idea is to make the machines or network resources unavailable to the users. The perpetrators, in order to not get exposed, spoof IP addresses and hence flood the target with data arising from several legal IP addresses which are being used without the knowledge of the actual user. Botnets are generally referred to as computers or machines that are infected by a malicious software and controlled remotely by the perpetrator without the knowledge of the user. [1, 2, 9]

B. The Cause:

DDoS attacks can be instantiated using a small flood or a large flood of data. Usually, large data pool is sent in Denial of Service (DoS) attacks. Hence, smaller chunks of illegitimate data are sent to the target machine or resource with the aim to flood the computer or network resources. Smaller chunks are tough to identify because they appear as legitimate data to the system and hence is allowed to pass through. Later, the data accumulates and floods the system resources. [10]

C. The Types:

DDoS attacks are of several types, a few popular ones are enumerated below:

1. *UDP Flood*
   UDP i.e. User Datagram Protocol is a networking protocol which is session-less. A large number of UDP packets are sent on random ports as a part of this attack on the remote host. Hence, the system is enforced to send UCMP packets. This renders the system unusable to the legitimate users, resulting in 'slow death'. [9]

2. *SYN Flood*
   A SYN flood is initiated when an attacker sends repeated SYN requests to the server and the attacker receives a syn-ack i.e. syn acknowledgement but the attacker doesn't acknowledge the system as required by the TCP three-way handshake. Eventually, the system resources get consumed and thus, unable to accept legitimate requests. [4, 6, 8]

3. *Ping of Death*
   Ping of death is an attack named especially after its action due to sending malicious pings to the host system. Multiple IP packets are created from a large IP packet. These are pinged to the host system resulting in an IP packet of approximately 65,535 bytes. [3, 11]

4. *HTTP Flood*
   In HTTP flood attacks, neither spoofing occurs nor are malformed packets sent. Several botnets are employed which get activated during the time of attack and send requests to the host machine or network resource which results in major loss as all the resources get utilized trying to process all the illegitimate data. [1, 9]

5. *NTP Amplification*
   Network Time Protocol Amplification attack involves the attacker gaining access to publically available NTP servers to allow flooding of the network resources. [8]

6. *Zero Day Flood*
   The most recent and preferred attacks are executed by exploiting vulnerabilities and inefficiencies of systems, also called as backdoors which are used for launching attacks and viruses. [8]

The final executed attacks are a combination of two or more flooding techniques generally.

D. The Effects:

1. Network Disruption
2. Exhaustion of network resources and machines physically
3. Mal-alignment or destruction of system configuration
4. Bandwidth consumption
5. Consumption of limited resources

E. Prevention of DDoS:

Since every DDoS attack is aimed at a different resource, it is impossible to provide prevention techniques that can satisfy every machine or network requirement. The prevention techniques present are:

1. *Null route or Black Hole Technique*
   Since every packet of information passes through several gateways before reaching its destination, the gateways contain a list of IP addresses which are untrusted and hence, are sent to the Null Route. Due to this, the server does not reply with a syn-ack request and hence prevents the attack from causing damage. [4, 9]

2. *Filtering Techniques*
   Several filtering techniques such as SYN proxy, Anomaly recognition, Granular rate limiting, dark address prevention, white and black lists, etc. are present. [9]

3. *Domain Name System Blacklist (DNSBL)*
   IP addresses considered as potentially harmful are listed by programs for filtering purposes. Hence, flagging or rejecting of messages is done and listed. [8]

4. *Others*
   Security patches, IP being changed rapidly (hopping), IP broadcast, etc. are techniques which are very commonly used. [8]

F. *The Challenge: Big Data Analysis*

The algorithm in this paper has been created using big data analysis. Big data analysis has been chosen because the probe implemented is done on large data sets. The data sets referenced have been chosen to provide the precis of the system c*urrently* facing a DDoS attack. Hence, real time simulated data has been evaluated to understand how the systems get flooded by illegitimate data and how effective measures can be taken to prevent it. Since the technique involves prior and present data sets to be scrutinized, big data has been used to make the technique faster and accurate. [3]

II. RELATED WORK

Big data analysis has been a prevalent and effective technique in preventing DDoS attacks. Big data works on the strategy of preventing DDoS attacks by an early analysis of the system. The system is put on probe to detect incoming data traffic on the system by observing the following vital attributes: [3]

1. PKT_SIZE
2. NUMBER_OF_PKT
3. NUMBER_OF_BYTE
4. PKT_RATE
5. PKT_AVG_SIZE
6. PKT_DELAY
7. FIRST_PKT_SENT
8. LAST_PKT_RECEIVED

The DPROS Algorithm works against volume based attacks, SYN flood, UDP flood, Ping of Death, NTP amplification, HTTP flood and Zero Day flood attacks by using the attributes as described:

1. Number of Packets:
   The number of packets received are analyzed for plotting the graph by big data analysis.

2. Number of bytes:
   The amount of bytes that are present in a packet are observed to plot graph.

3. Incoming packet size:
   Incoming packet size is required to observe peaks of the incoming data as well as to analyze it for ping of death.

4. Outgoing packet size:
   Outgoing packet size is observed for syn-ack, sent by the system during syn flooding.

5. Packet average size:
   Average packet size is used as a reference from where peak size is noted.

6. Packet delay:
   Packet delay is used for preventing UDP and Syn flooding.

7. First packet sent:
   First packet sent is stored for checking packets for accumulation of data.

III. THEORY

DPROS algorithm works on the concept of Big Data Analysis. According to this algorithm, the system for which this will be implemented will be previously analyzed over a period of a week to map the incoming data packet rates. The data packets are observed during peak time and average time, and a graph is plotted for this prior analysis.

Subsequently, when data is now being received by the host system, the data is stored in a high speed buffer memory and is then plotted immediately as further data gets stored and plotted. Next, the algorithm proposes to record the peaks of the graph and compare the data with the existing recorded data and implement the conditions provided.

Our proposed system is dynamic in nature because it is based on continuous learning of network traffic by the machine.

Algorithm part 0 (pre requisites):

Initially the system is provided with a regular day network traffic data set for a period of more than a week, as initial threshold value for comparisons is noted. Post initializing, the system will reconfigure its threshold value depending upon its graph at peak and moderate period (reference 0).

Algorithm part 1:

Incoming data packets are temporarily stored in a fast buffer memory for plotting real time data at a sampling rate of 10 milliseconds.

1. Data will be referred from the buffer and will be plotted in a scatter chart where:
   X-axis represents time
   Y-axis represents data

2. High peaks are analyzed from graphs i.e. high peaks in the real time graph will be compared with the existing graphs.

3. If the value of peak is greater than thrice the threshold value of peaks in existing data then it is illegitimate and will be sent to firewall, if value is within limit then it will be considered legitimate for the present state and is examined further.

4. If for any TCP protocol the time for SYN-ACK exceeds the threshold time, then the data will be sent to firewall.

5. If no application responds for a given UDP packets within a threshold time then it will be sent to firewall.

6. If reassembled IP fragments from different sources add up to form a large IP fragment that can overflow the memory buffer then it will be sent to firewall.
7. If all the above conditions (4-6) are not satisfied then the data would be considered legitimate and will be send to the host computer for processing.

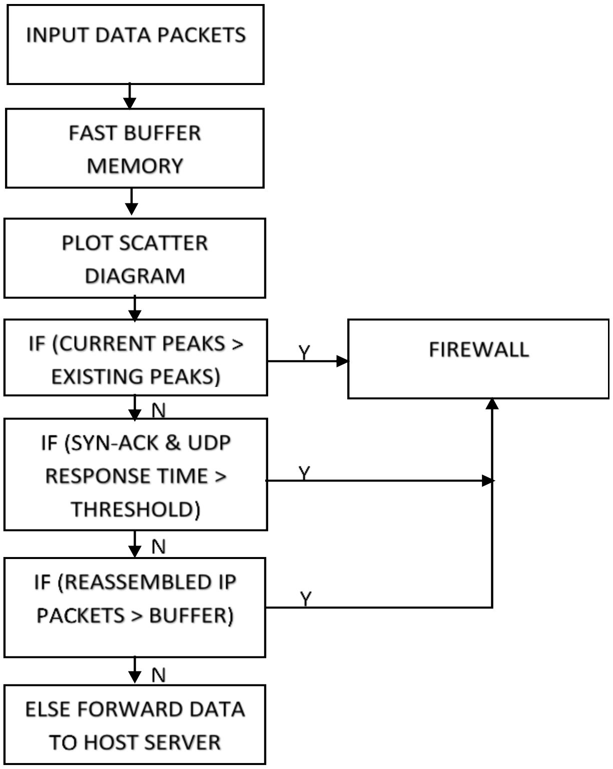
INPUT DATA PACKETS
FAST BUFFER MEMORY
PLOT SCATTER DIAGRAM
IF (CURRENT PEAKS > EXISTING PEAKS)
Y
FIREWALL
N
IF (SYN-ACK & UDP RESPONSE TIME > THRESHOLD)
Y
N
IF (REASSEMBLED IP PACKETS > BUFFER)
Y
N
ELSE FORWARD DATA TO HOST SERVER

Fig. 1: Block Diagram DPROS Algorithm

This algorithm is based on big data analysis. A graph is plotted to analyze data rate to predict illegitimate traffic.

The inbound data on that day will then be logged and kept as a reference (reference 1) for the analysis that will take place the next day. Hence, the machine is made to learn reference 0 and reference 1, where reference 1 is updated per diem.
Through this technique, the system will remain updated with the recent traffic flow and hence prevent future attacks as well as increase network security.

IV. RESULTS AND DISCUSSION

The algorithm is based on big data analysis which helps making it robust than other preventive techniques since this is centered on predictions. Prediction analysis helps in foretelling the result which helps in latching on to untrusted sources and hence, preventing probable attacks from stressing the system or network resources.
The algorithm can be implemented for complex systems since it offers several protection layers to the system due to a firewall provided to catch exceptions if the conditions mentioned are satisfied and hence reducing the chance of hacking due to several types of floods. It also reduces the complexity of security provided to protect the system against DDoS attacks.
The outcome is a system well protected from DDoS attacks through the process of big data analysis. Coded execution is proposed for empirical results to be obtained, which will in turn strengthen a priori work.

V. CONCLUSION

The proposed hypothetical system is able to protect the host server from Distributed Denial of Service attack. The algorithm consists of multiple stages of filtration processes for ensuring proper functioning of the system. The first stage includes a graphical comparison between present and threshold sample of data packets which are plotted using this algorithm, which implements big data analysis to plot and compare incoming data packet rates. The second stage consists of checking the integrity of TCP i.e. whether SYN-ACK responds in threshold time. Further, it compares the response time for any application of UDP with the threshold time. Finally, it checks the reassembled IP packets for overflow of memory buffer. After this processing, only legitimate data is able to pass through to the system, preventing any attacks.

VI. FUTURE SCOPE

As the attack rates are increasing day in - day out, not only by intensity, but also in diversity. The proposed system is flexible enough to ensure that with the update of attacking methods, the algorithm can be adjusted according to the requirements during that time. The system has enough capacity for such transformations without seizing the integrity of the algorithm.

REFERENCES

[1] Alomari, E., Manickam, S., B. Gupta, B., Karuppayah, S. and Alfaris, R. (2012). Botnet-based Distributed Denial of Service (DDoS) Attacks on Web Servers: Classification

and Art. *International Journal of Computer Applications*, 49(7), pp.24-32.

[2] A. Yaar, A. Perrig, D. Song, Stackpi: New packet marking and filtering mechanisms for ddos and ip spoofing defense, Tech. Rep., Carnegie Mellon University, 2003.

[3] C. Cheng, H. Kung, and K. Tan, "Use of spectral analysis in defense against DoS attacks," in Proceedings of 2002 IEEE GLOBECOM, Taipei, China, 2002

[4] C. Inc.-1 SYNDefender. Available at http://www.checkpoint.com/products/firewall-1/syndefender.html, 1997.

[5] C. Inc. TCP SYN Flooding Attack and the FireWall-1 SYNDefender.Available at http://www.checkpoint.com/products/firewall-1/syndefender.html, 1997.

[6] [6] C. Schuba et al. Analysis of a Denial of Service Attack on TCP. In *Proceedings of the 1997 IEEE Symposium on Security and Privacy*. IEEE Computer Society Press, 1997

[7]

[8] [7] History-computer.com. (2016). *History of Computers and Computing, Internet, Birth, First computer virus of Bob Thomas*. [Online] Available at: http://history-computer.com/Internet/Maturing/Thomas.html [Accessed 20 Oct. 2016].

[9]

[10] [8] Mirkovic, J. and Reiher, P. (2004). A taxonomy of DDoS attack and DDoS defense mechanisms. *ACM SIGCOMM Computer Communication Review*, 34(2), p.39.

[11]

[12] [9] P. Ferguson, D. Senie, Network ingress filtering: Defeating denial of service attacks which employ ip source address spoofing, IETF, RFC 2267.

[13]

[14] [10] Renuka Devi, s. (2012). A Hybrid Approach to Counter Application Layer DDOS Attacks. *International Journal on Cryptography and Information Security*, 2(2), pp.45-52.

[15]

[16] VIII. WEB REFERENCES

[17] [11]Chow, S. (2016). *Performance Analysis of a System During a DDoS Attack*. [online] Available at: http://www.sfu.ca/~spc12/ [Accessed 17 Oct. 2016].