A COMPARATIVE STUDY OF VARIOUS DDOS ATTACKS THERE DETECTION AND DISCRIMINATION APPROACHES AGAINST A FLOW CORRELATION APPROACH

Pradeep.P [*]

S.Aji Kumar[**]

Abstract—Distributed Denial of Service (DDoS) attack is a serious threat to the Internet Community.Eventhough the main intension of DDoS attacks is to access the web resources for illegitimate purpose, they posses a serious threat to the legitimate users. Various techniques are available to detect a DDoS attack. But the main challenge is how to discriminate a DDoS attack from a flash crowd. In this paper we had studied about the various type of DDoS attacking tools there behavior and the various DDoS detection method .The attacks we studied spanned from DoS attacks to the attacks from sophisticated Botnets(Hybrid Botnets).As each attack detection method will be broken by a new detection overcoming methods all the attacks and the detection approaches have been studied in a sequence. In many cases it was found that the attackers can mimic or imitate the characteristics of a legitimate traffic by the illegitimate traffic. So a novel approach is needed to discriminate this mimicking and discriminate the attacks from a legitimate access. We were able to compare the available techniques with the most comprehensive and updated technique using a flow correlation approach. This technology was found to be a

[*] P.G Student, Department of Computer Science & Engineering, P.S.N College of Engineering & Technology, Tirunelvelli

[**] Associate Professor,Department of Computer Science & Engineering, P.S.N College of Engineering & Technology, Tirunelvelli

promising technique for discriminating DDoS attacks from the legal crowd called as Flash Crowds

Index Terms –DDoS attacks, flash crowds, similarity discrimination, Traffic Mimicking ,Hybrid Botnets

1 INTRODUCTION

Internet architecture is designed in such a way that the basic technology available will provide equal priority for both legitimate and illegitimate traffic. This feature of Internet is misused by the attackers of cyber community with the motivation of money ,black mailing and activities based on revenge. One of the studies done when the DoS attacks began as a serious threat shown that Denial of service(DoS)attack frequency found 12,000 attacks within a three week period in year 2001[1].The main concept used in DDoS attacks is to reduce system availability to block legal oe legitimate users. Some of the most popular attacks methods include DoS flooding attacks,botnets and by a zombie(client)-attacker(master).Thus the attacks will be done on the common architecture like network based system and distributed based systems

Anyway in all these cases the attacks will consume the network buffer capacity,CPU processing cycles and bandwidth. Hence a bottleneck will be formed by the attacks[2] which will degrade the system performance and will gradually keeps the legitimate users away from the service. In this paper we studied about the various types of attacks initiated from a distributed environment whose intension is to attack a web location. Several attack detection approaches have been considered and it was found that the attack detection methods was promising but they were not able to discriminate between an attack and a legal web access(Flash crowd).The most promising attack detection method uses a flow correlation based approach which is an entirely changed concept from a similarity based approach. This method became successful because it made easy to discriminate the attack and the application can be easily launched along with the routers along the server side.

2.RELATED STUDIES

For the proper summarization of the advanced techniques to detect and discriminate DDoS attacks it is necessary to study about the various techniques for attacks and the basic approaches to detect the attacks. So our studies started from the attack methods ,there methods,behaviour and some of the popular attack detection methods

2.1 Attack Types

The basic of all attacks is to exploit weakness and to generate vulnerability based attacks and flooding attacks

a)Vulnerability attacks

According to this approach network packets will exploit the weakness in a network protocol and will interact maliciously with the applications using this protocol. The attacks once explored the vulnerability can cause more memory consumption ,extra CPU processing ,make forced system reboot and sometimes system performance degrading .Examples of these attacks are transmission control protocol Synchronization(TCP SYN) Flag and targa3 attacks

b)Flooding Attacks

This attack works on the principle that the victim should be send with a large and occasionally continuous amount of network Table-1 DDoS attack propagation mechanisms

Mechanism	Design Complexity	Detect ability	Propagation Speed
Operating System	Medium	High	Low
Services	Medium	Medium	Medium
Applications	High	Low	High
Social Engineering	Low	Medium	Low

traffic overload .By doing this the legitimate workloads can also make congestion and thus a bottleneck will be created for the victim. These types of attacks will not utilize the software vulnerability .The main reason for the attackers to select these type of attacks is that this type of attacks can be initiated from any type of network protocols including TCP/IP ,Internet Control Message protocol(ICMP),User datagram Protocol (UDP).Some example tools used for these type

of attacks are Synhose,hping2.Thus the user will be provided wiath a scenario such that these workload is not processed

c)Botnet attacks

Botnets are comprised of collection of internet worked computers which are otherwise known as Botnets.[4].These computers can be remotely controlled by attackers .The attackers can control it from there offices, home or while in a journey. The main challenge in this approach is how to propagate the attacks to the user computers. If the user or victim computer is equipped with the network overloading avoiding tools or flooding control tool then it will be difficult for the attackers

The various propagation mechanisms as shown in table-1 makes clear that the propagation speed of the attacks through applications are high. The botnets can be organized basically on three approaches ,centralized, Peer-to-peer botnets are much difficult to be detected .Botnets can be operated on various topologies which is mentioned in table-3.Thus it can be concluded that identity theft based attacks are the most difficult to be detected .At any time the botnets have to communicate each other instances of bot.Botnets will usually communicate each other in such a way that the communication should be hidden .The overall procedure for controlling the botnets is called command and control approach The various types of command and control are

Centralized

Table-2 Command and Control Topologies

Topology	Design Complexity	Detectablity	Message Latency
Centralized	Low	Medium	Low
Peer-to Peer	Medium	Low	Medium
Unstructured	Low	High	High

As per this method in a network of attacking computers there will be a central communication controlling computer. This computer is known as a server or central control. All the communication between computer will be controlled by the central computer. The main problem in this type of architecture is that the central computer should be of high configuration and

should be active at all time. If the central computer is not working then the entire architecture will be not working

Table-3 Attack Classes

Topology	Detectablity	Design Complexity	Attack Value
Single host DDoS	High	Low	Low
Multiple host DDoS	Medium	Medium	Medium
Identity Theft	Low	High	Medium
Spam	Medium	Medium	High
Phishing	Medium	High	Medium

d)Hybrid Peer-to Peer(H-P2P)

This type of communication have several advantages over the centralized approach .In this type of command and control mechanism each computer or attacking host can work as an independent unit .Thus if a computer or a group of computers is hacked by the attacker detectors or not working then also the architecture will work in such a way that the attacks will be continuing from the non hacked computers. The most promising one is a hybrid peer to peer botnet[5].The major advantages of this architecture is that

1) The botnets can be in a working state and controlled state even after a major portion of the computers in the network are not functioning or compromised

2) The topology of the network will be hidden even after some hosts in the botnets are captured

3) The attack detection methods should be failed when trying to detect bots via there common traffic pattern

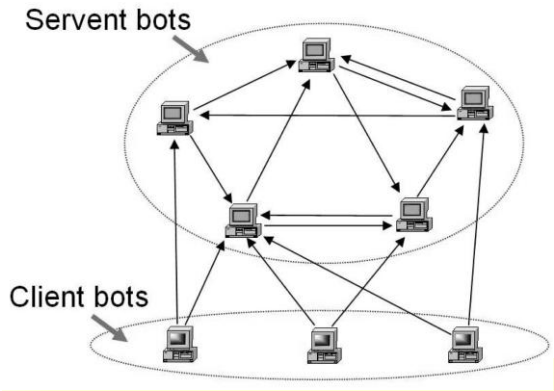

Fig-1 Hybrid Peer to peer Botnets

The basic architecture of Hybrid peer to peer botnets is shown in fig-1.The botnets are divided into 2 peer list ,Servent bots and client bots. Each member in a servent and client will interact each other. Thus even a set of computers is not functioning some other computer will control this computer. Also each time when a bot receives a new command it forwards this new command to all hosts in its peer list

Attack Detection Methods

1) Basic Approaches

The major goal of all detection approaches is to detect and distinguish between legitimate user traffic and illegitimate user traffic.Attack detection can be done using various procedures. One most common approach is to detect the attack on the victim side. These type of attack detection methods can be installed as a part of firewall software, router utility that controls a subnetwork.Some approaches used are

1.1) Activity Profiling

Activity profiling deals with monitoring the packet header information. The time gap between the consecutive web request will be considered as an attribute. The overall network traffic is the overall sum of packet rates including both the incoming and outgoing flows during a unit time. Activity profiling will make use of protocols like TCP,ICMP and SNMP

1.2)Sequential Change-point Detection Method

Change point detection algorithm isolate a traffic statistics change because of attacks. This type of algorithm can initially filter traffic data by adding, ports and protocols and store the result flow as a time series. The flow information during the time series can be seen as subjected to a sudden change. One common type of sequential change point detection algorithm is a cumulative

sum(cusum) algorithm. The cusum approach identifies deviation in the actual versus expected local average in the time series. This difference will be analysed against a threshold. In such a situation cusums recursive statistic value will be increased. When the analysis is done during normal traffic then the statistics value will go down.

1.3)Wavelet Analysis

Wavelet analysis makes use of the spectral aspects of a traffic. Wavelet can provide both time and frequency description and it can be used to determine the time at which a frequency change occur. Wavelets can be used to separate anomalous signals from background noise separate analysis is required for both the signal components and noise components to determine the presence of anamlolies.An appreciable change that can be done to this wavelet analysis for improving the efficiency is that the overall time can be represented as a time series resulting a time localized high and mid band spectral energies. The cusum change point detection methods can be applied here also by using discrete wavelet analysis to post process the cusum statistics response

2)Advanced Methods

2.1) A Non Gaussian and long memory statistical Method

This detection procedure makes use of a non gaussian model to model the network traffic and apply analysis of the traffic against time window. Once the traffic is modeled two values are needed for the analysis[6]. One is a reference time window. The model during the reference time window is analyzed against the prior or previous model during the same reference time window. The difference in the model of the network traffic is called quadratic distances. This quadratic distances are compared against a threshold to confirm the attacks. The overall performance of the techniques relies on the sample network flow used to analyses if the attack characteristics is out of the range of the calculated models the the approach will be unpredictable. The various models that should be made for this methods include the non Gaussian models of a regular traffic and a process used to simulate the feature of a possible attack. Example Gamma Arfima models. This process can model the first and second order statistics of aggregated computer network traffic time series.

2.2)A Hidden Markov model used for detecting Anomalies behavior in User Browsing Behaviors

In almost all of the DDoS detecting methods available the main approach is to analyze the TCP or protocol layer behavior characteristics. But this technology uses application layer analysis to detect a DDoS attack. The browsing behaviors of the users can be modeled in a better manner with the help of a hidden Markov Model. Once the user behaviors is modeled the entropy of the web surfing behavior will acyt as a main criteria for detecting DDoS attacks[7].The browsing behavior of an user can be modeled with the help of two values. The web site structure(The various web pages, Hyperlinks etc) and the way how the users will access the web site.

The user web request consists of two states HTTP-ON and HTTP-OFF.HTTP-On will be initiated when the user have started a web request with the help of a hyperlink or some other form. Once a page is provided for the user then the user will start using he web page. The later access of the web page is called as HTTP-OFF state. This model cant be modeled well when the user is accessing the web page by directly typing the URL.Thus with the help of a state diagram the various states of the web access by the users can be model. This model is the basic method for modeling a user behaviour.The various state changes of this state diagram can be analyzed for detecting the anomalies from a user. A detector and filter method can be placed in between the user and the internet. The Markov model can be trained with the usual behavior of an user. The entropy between a behavior ant the trained behavior will be the key method used by the detector. The more the entropy then the more the anomaly will be. Based on the arte on anomaly the method will either forward the traffic or report it as a suspected threat. A major draw back of following this approach is that the total number of paths that should be analyzed when the network traffic is high will be of large number. In such a situation an approach called M-Algorithm will be used. At any time only M out of Total N paths will be selected. These paths are selected on the basis of metric values. Thus the total number of paths to be analyzed can be reduced based on the entropy

3)PROPOSED WORK

A Flow Correlation approach to detect and Discriminate flash crowds from DDoS Attacks

During the years 2001-2010 various cases of DDoS attacks was reported. As a result more researches have been done in this field and lot of techniques has developed for detecting DDoS attacks. But the major problem with these techniques is that they were unable to discriminate a

large network traffic as either a DDoS attack or a Flash Crowd. Flash Crowd is a large network traffic which is legitimate. One of the main practical occurrence of flash crowds is on news web site ,University web site, lottery result web site etc.The scenario that happens here is that during a peak period lot of network traffic will occur. An example is mentioned in Fig-2.In this figure the traffic which is represented in red dots are attacks coming from a botnet and others are from legitimate users. It can be found that all the attacks requests are coming from almost related locations. The number of web request will be large and it may be considered as a DDoS attacks bt the known technologies. But the traffic is legitimate. The main technology used by almost all methods for detecting DDoS attacks is a similarity based approach. That is these a methods will verify the address from which the heavy traffic is coming and will conform an attack if the number of request coming from a location is much large compared to the normal users. In case of a news web site if a flash news o\r hot news is published then lot of people will be accessing the website form a common web browser or a particular IP address. The similarity based approach will fail to detect these scenario as a legitimate access. So ths proposed system provides a novel approach to detect and discriminate a flash crowd from a DDoS attack abusing a flow correlation approach[8]

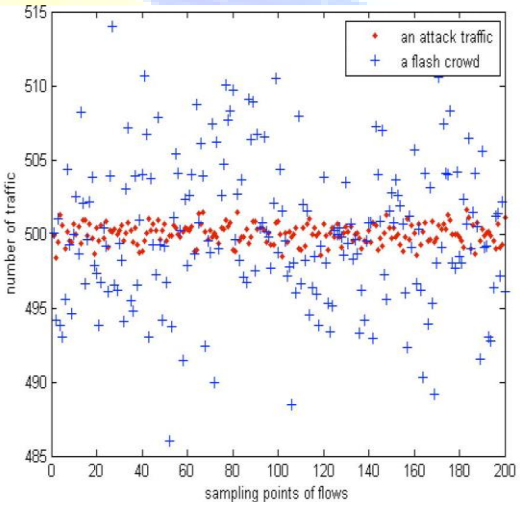

Fig -2 Flash Crowd and DDoS attacks

Definitions used

Network Flow

For a given network the network flow is the total number of traffic intended to a particular location

Flow Strength

For a given Network flow the maximum amount of flow that can occur during a unit time is called as a Flow Strength

Flow Fingerprint

For any Network flow and its Flow Strength an equivalent network flow can be formed and this is called as a Flow fingerprint

Flow Correlation Coefficient

If Xi and Xj are two network flows then a Flow correlation coefficient can be formed from these values as

$$\rho_{x_i,x_j}[k] = \frac{r_{x_i,x_j}[k]}{\frac{1}{N}\left[\sum_{n=1}^{N-1} x_i^2[n] \sum_{n=1}^{N-1} x_j^2[n]\right]^{1/2}}.$$

This Flow correlation coefficient forms the key in this method

Analysis

Once the flow related values are found the next step is to analyze the values to detect and discriminate the flows between a DDoS and Flash crowds. For the analysis various supporting theorems are there. These theorems will include the behavior of the flows and can be found to be effective

Theorem 1

If there are two flows called as Xi and Xj if there exists a standard deviation variable, the this variable value will be inversely proportional

Thus this theorem can be used in a way that if we can find the standard deviation between two flows then the value will not be same for the flash crowd and the DDoS attack even though the network traffic is same.

Theorem 2

For a network flow the flow correlation value will approach to 0 as the network flow exists for long time for a flash crowd

Thus even if the time of existence of two flows are same for the flash crowd it can be seen that the correlation value will approach 0

Theorem 3

In the absence of any noise or in an ideal situation the flow coreelation values will be different for same flows originating from a same network.

Thus this theaorem can be applied in a host operating from a botnet such that even the request are coming from a same network and is managed by a same master there is a correlation among the flows.

Thus by finding the various network parameters and calculating the flow correlation co efficient values from these values we can analyse two flows and can apply the theorems to confiem whether it is an attack or Flash crowd

4) CONCLUSION

DDoS attacks are a main threat to the internet community. The various methods used to detect the DDoS attacks will be broken when a new technology for attacking comes. Also in many cases the legitimate attacks will be considered as an attack and will be blocked .So deriving a new technology to efficiently discriminate DDoS and a flash crowd is necessary. The various approach we found in Section – have been successful in detecting the anomalies in user behavior [7] and the threshold in web requests[6] but they all failed to discriminate the traffic.Also the methods which are considered as the basic approaches[3] were applied in the network layer.The flow correlation approach mentioned in our study as a novel approach is found to be effective[8].This method operates on the application layer and can easily discriminate the attacks and flash crowds also the storage space required for the network traffic parameters is less in this case. But the main challenge that should be met by this technology is that in the case of super botnets.Super botnets operate in such a way that the number of requests for a web location from these botnets will be same as that of a legitimate access. Thus the network parameter values will be same for both the legitimate access and illegitimate access. Current researches are going on to modify the technologies available to defend super botnets.

REFERENCES

1) D. Moore, G.M. Voelker, and S. Savage, "Inferring Internet Denial-of-Service Activity," Proc. Usenix Security Symp.,Usenix Assoc., 2001; http://citeseer.ist.psu.edu/moore01inferring.html.

2) V.L.L. Thing, M. Sloman, and N. Dulay, "A Survey of Bots Used for Distributed Denial of Service Attacks," Proc. SEC, pp. 229-240,2007.

3) G. Carl, G. Kesidis, R. Brooks, and S. Rai, "Denial-of-ServiceAttack-Detection Techniques," IEEE Internet Computing, vol. 10,no. 1, pp. 82-89, Jan./Feb. 2006.

4)V.L.L. Thing, M. Sloman, and N. Dulay, "A Survey of Bots Used for Distributed Denial of Service Attacks," Proc. SEC, pp. 229-240, 2007.

5 P. Wang, S. Sparks, and C.C. Zou, "An Advanced Hybrid Peer-to-Peer Botnet," IEEE Trans. dependable and Secure Computing, vol. 7,
no. 2, pp. 113-127, Apr.-June 2010.

6. A. Scherrer, N. Larrieu, P. Owezarski, P. Borgnat, and P. Abry, "Non-Gaussian and Long Memory Statistical Characterizations for Internet Traffic with Anomalies," IEEE Trans. Dependable Secure
Computing, vol. 4, no. 1, pp. 56-70, Jan.-Mar. 2007.

7)Y. Xie and S.-Z. Yu, "A Large-Scale Hidden Semi-Markov Model for Anomaly Detection on User Browsing Behaviors," IEEE/ACM Trans. Networking, vol. 17, no. 1, pp. 54-65, Feb. 2009.

8) Shui Yu, Wanlei Zhou, Weijia Jia, , Song Guo,Yong Xiang, and Feilong Tang "Discriminating DDoS Attacks from Flash Crowds Using Flow Correlation Coefficient" IEEE Trans.Parallel and Distribued Systems,Vol 23,No 6,June 2012

.