Protection from Application Layer DDoS Attacks for Popular Websites

Sreeja Rajesh

***Abstract*—One of the major threats for the Internet's reliability and stability is Distributed Denial-of-Service (DDoS) attack. The attackers are becoming more sophisticated and organized, also several high-profile attacks targeted prominent Websites. These are the prime reasons that gained importance for the study of DDoS attack detection and prevention. It becomes more undetectable if the legitimate HTTP requests are utilized by Application-layer based DDoS attacks to overwhelm the victim resource. Whenever such an attack occur or mimics the normal flash crowd event of a popular website, then it leads to serious problems. Flash crowd is a situation when a large number of web users are simultaneously accessing a popular Website, which results in a sudden increase of traffic to the website and this may cause the site to be virtually unreachable[1]. Here a mechanism to capture the normal flash crowd event pattern is introduced and the App-DDoS attack monitoring, detection and then blocking of further attack is implemented. An effective method is introduced to identify whether the surge in traffic is caused by normal Web surfers or by App-DDoS attackers. Access Matrix (AM) is defined to detect App-DDoS attacks based on userlogs and threshold value. Hidden Markov model is used to detect App-DDoS attack based on user behavior.**

***Index Terms*—Application-layer, distributed denial of service (DDoS), popular website.**

I. Introduction

A Denial of Service (DoS) is an intentional attempt by malicious users/hackers to completely disrupt or degrade availability of services/resources to legitimate/authorized users [2]. Some DoS attacks are SYN Flood, teardrop, smurf, black holes, octopus. DoS attacks exploit weaknesses in Internet protocols, operating systems, applications, and protocol implementation in operating systems. The services to the legitimate users may be disrupted completely by the Distributed Denial of Service (DDoS) attacks. Every member involved in the DDoS attack generates relatively small amount of traffic. But the combined result overwhelms the target system and it either responds so slowly as to be unusable or crashes completely. Mirkovic *et al*. [3] described DDoS attacks as amplified form of DoS attacks, where attackers direct hundreds or even thousands of compromised hosts called zombies against a single target. These zombie hosts are innocent computers who are unaware that they have been recruited for attacking by the attackers, from the millions of computers that are accessing the Internet through high-bandwidth and always available connections. The servers may be severely damaged due to the DDoS attacks and they are also greater threats to the development of new Internet services. Conventionally, DDoS attacks are carried out at the network layer, such attacks are known as Network-DDoS attacks. ICMP flooding, SYN flooding and UDP flooding are examples for Net-DDoS attack. Net-DDoS attack is an approach of attackers to exhaust entire network bandwidth so that the targeted host will either provide limited services, or provide services to only some users, or will not provide any services to its authorized users. It is not as easy as in the past for attackers to launch the DDoS attacks based on network layer because many researches have been done in this area and various methods to detect and block the Net-DDoS attacks are identified. When the simple Net-DDoS attacks fail, attackers shift their offensive strategies to application-layer attacks and establish a more sophisticated type of DDoS attacks [1]. To circumvent detection, the victim Web servers are attacked by HTTP GET requests (e.g., HTTP Flooding) and large image files are requested from the victim server in overwhelming numbers. In another case, attackers run a massive number of queries through the victim's search engine or database query to bring the server down [2]. Such attacks are called application-layer DDoS (App-DDoS) attacks. The MyDoom worm [4] and the CyberSlam [5] are all instances of this type attack Surfers. The contributions in this paper are three fold: 1) The Access Matrix (AM) is defined which detects the application-layer DDOS based on user logs and threshold value; 2) Hidden Markov model (HMM) [6] is used to capture the patterns of normal flash crowd and to monitor App-DDoS attacks based on user behaviour; 3) design monitoring system and validate it by a real flash crowd traffic and emulated App-DDoS attacks; The main aim of DDoS Protection System is to block the malicious traffic that overwhelms the remote system and allow only normal traffic to the server so that sever is not affected by application-layer DDOS attack.

II. Literature Survey

Various researches have been done during the past with regard to the detection of DDoS attacks from three layers of OSI namely layer 3-*Network layer*, layer 4-*Transmission layer* and layer 7-*Application layer*. The attacks done on the application layer were very fewer in the past and hence the researches done on Application-Layer protection are also few in number. Techniques to detect Application-layer DDOS attacks are highlighted below:

A. Client Puzzle Protocol

Client Puzzle Protocol (CPP) is an algorithm that will not allow any abuse of the server resources. According to this algorithm, any client who needs to establish a connection

Manuscript received March 9, 2013; revised June 30, 2013.

Sreeja Rajesh is with Department of Computer Science and Engineering, Jyothi Engineering College, Cheruthuruthy, Thrissur-679 531 Kerala, India (e-mail: m.sreeja79@gmail.com).

with the server has to first correctly solve a mathematical puzzle. After solving the mathematical puzzle, the client returns the solution to the server and the server will either quickly confirm, reject or drop the connection based on the solution of the client. The client needs to perform only a minimum amount of computation as the puzzle is made simple and easily solvable. Only negligible computational cost would be experienced by genuine used. Any client who tries to simultaneously establish a large number of connections will be unable to do so because of the computational cost (time delay).

B. Intrusion Detection System

A software that is used to automate the intrusion detection process is known as intrusion detection system (IDS).

C. Ingress Filtering

In computer networks a technique used to make sure that packets coming into the networks are actually from the one that they claim to be is known as Ingress filtering.

D. Threshold Value

The number of requests that a server can handle without straining its resources is called as threshold value. It is defined as a predetermined percentage of the maximum number of requests that a server can handle.

The IEEE papers which helped in the literature survey for the project "Protection from Application Layer DDOS Attack for Popular Websites" are mentioned below:

1) In the paper "Protection from Distributed Denial of Service Attacks Using History-based IP Filtering", IEEE 2003, Tao Peng , Christopher Leckie and Kotagiri Ramamohanarao [7], proposed a mechanism called History-based IP Filtering (HIF) for the edge router to admit the incoming packets according to a pre-built IP address database. The IP address database is based on the edge router's previous connection history. Also when the database size happens to be large enough search time increases and hence the delay.
2) Srikanth Kandula, Dina Katabi, Matthias Jacob and Arthur Berger in their IEEE paper "Botz-4-Sale: Surviving Organized DDOS Attacks That Mimic Flash Crowds" [5], proposed the design and implementation of Kill-Bots, a kernel extension to protect Web servers against DDOS attacks that masquerade as flash crowds. The author suggests the use of CAPTCHAs to distinguish the IP addresses of the legitimate clients from those of attack machines. In contrast to prior work on CAPTCHAs, this system allows legitimate users to access the attacked server even if they are unable or unwilling to solve graphical tests. The design is implemented in the Linux kernel and evaluated it in Planetlab. Kandula et al. design a system to protect a web cluster from DDOS attacks by designing a probabilistic authentication mechanism using CAPTCHAs (acronym for "Completely Automated Public Turing test to tell Computers and Humans Apart"). Unfortunately, requiring all users to solve graph puzzles may result in the possibility of annoying users and introducing additional service delays for legitimate users. This paper may not serve the purpose if any automated techniques are being used by attackers to solve the graphical puzzles. Fig. 1. shows the Kill-Bots Overview

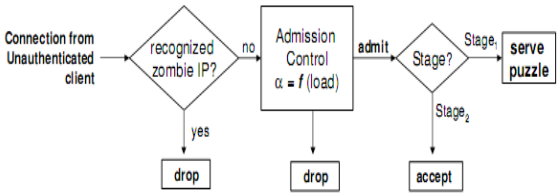

Fig. 1. Kill-Bots overview (Note that graphical puzzles are only served during stage).

3) "Monitoring the Application-Layer DDOS Attacks for Popular Websites" IEEE 2009, proposed by Yi Xie and Shun-Zheng Yu [1]. This method involves high mathematical computations.

III. Research Elaboration

Web user behavior is mainly influenced by the structure of website (e.g., the Web documents and hyperlink) and the way users access web pages. In this paper, the Application Denial of Service attack is considered as anomaly browsing behavior. Characteristics of Web access behavior shown in Fig. 2, plots the HTTP request number (average user hits) per 5 sec during the burst Web workload on the popular-website that is requesting for protection. It is observed that the normal flash crowd is mainly caused by the sudden increment of user request rate.

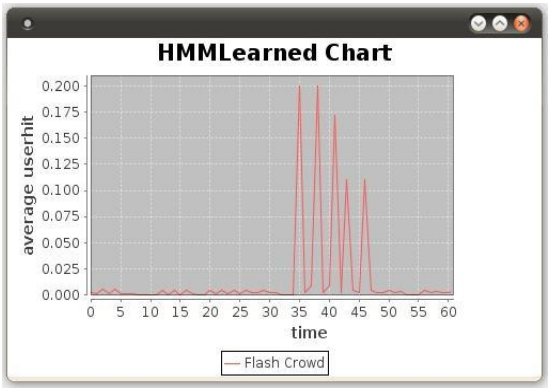

Fig. 2. Normal flash crowd.

Characteristics of Web access behavior whenever there is any attempt of an application-layer DDOS attack by increasing the traffic to a popular-website is shown in Fig. 3.

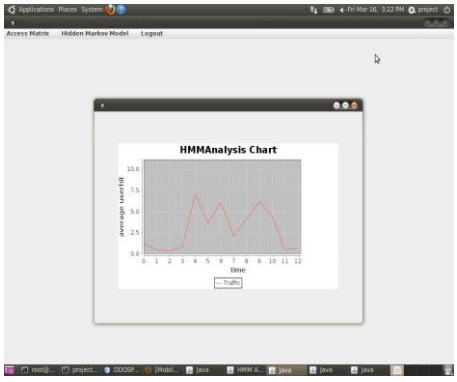

Fig. 3. Application-Layer DDOS attack.

These results show that the user's access behavior profile can be used to detect the abnormal varieties of user's browsing process during the flash crowd. The document popularity has been widely used to characterize the user behavior and improve the performance of Web server and Internet cache.

In this paper the defensive mechanism against Application Layer DDoS attack requires the following

A. Access Matrix

Information about the users accessing the website are stored in a matrix called Access Matrix [1]. The access matrix model is the policy for user authentication. It is used to describe which users have access to what objects.

B. Hidden Markov Model

If a system being modeled is having unobserved state and is assumed to be a Markov process then such a statistical model is known as Hidden Markov Model (HMM). In a regular Markov model, the state is directly visible to the observer, and hence the state transition probabilities are the only parameters. In a Hidden Markov model, the state is not directly visible, but output dependent on the state is visible. Each state has a probability distribution over the possible output tokens. Therefore the sequence of tokens generated by an HMM gives some information about the sequence of states. The adjective 'hidden' refers to the state sequence through which the model passes, not to the parameters of the model [7]. Some of the applications of Hidden Markov models are in temporal pattern recognition such as speech, handwriting, etc.

IV. System Analysis

A. Existing System

Existing detection mechanisms operate at the network level to detect DDOS floods in the network. For example, the anomaly detection system assigns every packet a score based on the probability of it being a legitimate packet given the attribute values it carries. In contrast, there are other mechanisms which detect anomalies in the traffic distribution instead of traffic volumes. However, the attacks considered in this paper cannot be detected by such tools as the attacks may not necessarily deviate the network statistics in either volume or distribution. Other detection mechanisms attempt to catch intrusions both at the network and the host level. Distinguishing a DDOS attack from a flash crowd has also proven difficult. Internet DDoS attack is real threat on websites such as Yahoo, CNN, Amazon, eBay, etc i.e. services were unavailable for several hours due to Lack of defense mechanism on current Internet and also for individual Systems. The on hand feature for user behaviors can be summarized as the following ways. The first is based on probabilistic model, a double Pareto distribution for the linkchoice, and a log-normal distribution for the revisiting, etc. The second is based on click-streams and web content, e.g., data mining to capture a web user's usage patterns from the click-streams dataset and page content. The third is based on the Markov model, e.g. Markov chains to model the URL access patterns that are observed in navigation logs based on the previous state.

B. Proposed System

The objective of proposed system is to protect the popular websites from Application-Layer DDOS attack. The proposed system monitors the traffic and detects the application layer DDOS attack based on user logs and user behavior. Once the attack is identified the attacker system will be blocked from further overwhelming the traffic. Techniques used are Access Matrix and Hidden Markov Model. The Access Matrix defines access privilege of various users of a particular Websites. The administrator of popular websites has to register themselves with the DDOS software for protecting their site from application layer DDoS attack. After registration process the administrator of popular website has to generate the access matrix. The data provided to access matrix are as follows:

1) Information about their legitimate users.
2) Url's which their legitimate users are not allowed to access.
3) The access levels of the legitimate users etc.

Whenever any user is requesting a page from the administrator of the popular website, the packets are captured and the following informations are identified:

1) The IP address of the requesting system.
2) Userid of the user who is requesting the service.
3) Access level of the user.
4) The url which he is trying to access.
5) The services (webpages) he is requesting.

These information will be crosschecked with access matrix and if there is any discrepancy (ie. If a user is trying to access the webpage which is restricted to him) then an attack is detected and the requests coming from that IP will be blocked. A Threshold value is set and if a user is requesting a webpage greater than the threshold value is also considered as an attempt of attack. Hence this technique is used to detect application DDOS attack using the logs of the web server and threshold value.

The Hidden Markov Model is used to detect application DDoS attack based on the user behaviour. The input to the Hidden Markov model is the average hit and average page count of the user and the output of the model will be the decision made (whether he is a normal user or an intruder) as shown in Fig.4. Whenever input is given we get the output and the state transitions within the system is not visible, hence the name Hidden Markov Model

Fig. 4. Hidden markov model

There are two phases in Hidden Markov Model. They are Training phase and Comparison phase.

Training phase (Learning phase): The Training model is generated periodically during this phase. Training period is set to 1 minute. The legitimate users of the popular website (ie. Website that has registered with the DDOS Protection system for protection) are allowed to access the website until the training period expires. All the requests coming to

that popular website will be captured by the DDOS Protection system and the information about the IP address of the system requesting service, the user-id of the user who is requesting the service, access level of the user, which url he is trying to access and what are the services (webpages) he is requesting can be identified. For every 5 second the average hits of the user accessing the webpage is computed and average hit for 1 minute is computed based on user-types. The Training Model contains information about the average hits for all user-types accessing the website and this represents the normal flash crowd. Characteristics of Web access behavior as shown in Fig. 3, plots the HTTP request number (average user hits) per 5 sec during the burst Web workload on the popular-website that is requesting for protection.

Comparison phase (Analysis phase): whenever any user, requests service from the popular website, the DDOS Protection system will capture the requests. The user behavioral patterns are computed and are compared with the trained model which is developed during the Learning phase. If any discrepancies are found then App-DDOS attack is detected and all requests coming from that IP will be blocked from further accessing the system.

C. Advantages

1) One can make these systems to take into account the user's series of operations information.
2) Suitable for on-line detection as there is an intensive computation for page content processing.
3) The effectiveness of packet filter is the best.

V. Discussions

The conventional security technologies such as firewalls [8] Intrusion Detection Systems (IDSs) [9] and access control lists in routers are unable to defend networks from App-DDoS attacks. The main reason is that, it is almost impossible to differentiate between legitimate and attack packets since the potency of flooding Distributed Denial of Service attacks depends only on the volume of attack traffic and does not depend upon the exploitation of software bugs or protocol vulnerabilities. Consequently, flooding DDoS packets do not need to be malformed, such as invalid fragmentation field or a malicious packet payload. As a result, the flooding DDoS traffic looks very similar to legitimate traffic [10]. It is a real challenge to defend against these attacks as flooding DDoS attacks are very dynamic to elude existing defense systems. Due to the seriousness of Distributed Denial of Service attacks and the growth of sophistication of the attackers led to development of numerous defense mechanisms. But still, the tremendous growth in the number of Distributed Denial of Service attacks and their financial implications press the need of a comprehensive solution. The comprehensive solution against Distributed Denial of Service attacks can be devised only if the Internet community incorporates better ways to accumulate details of attack.

VI. Conclusion

In order to create defenses for attacks it is necessary to obtain timely and significant information by monitoring dynamic network activities. Most of the current efforts and researches focuses on detecting network layer DDoS attack also called Net-DDoS attacks with stable background traffic. This paper aims at signaling the Application Layer DDoS attacks during flash crowd event. This is done by revealing the dynamic shifts in normal burst traffic and thus monitoring Web traffic. The proposed method is based on Access Matrix and Hidden Markov Model. This method reveals early attacks merely depending on the threshold specified, user logs, user behavior and gives all the privilege for administrator who can effectively identify and block the connections for specified attacking host. Measures can be devised to check for IP spoofing as an additional detection process. Further this scheme can be applied to client-server architecture thus providing double protection.

References

[1] Y. Xie and S. Z. Yu, "Monitoring the Application-Layer DDoS Attacks for Popular Websites," in *Proc. Networking, IEEE/ACM Transactions*, Feb. 2009, pp. 15-25.

[2] G. Coulouris and J. Dollimore, *DISTRIBUTED SYSTEMS*, 4th *EDITION*, pearson education 2005.

[3] C. Chang, "Defending Against Flooding-Based Distributed Denial of Service Attacks: A Tutorial," *Computer Journal of EEE Communication Magazine*, vol. 40, no. 10, pp. 42-51, 2002.

[4] Incident Note IN-2004-01 W32/Novarg. (2004). A Virus. CERT. [Online]. Available: http://www.cert.org/incident_notes/ IN-2004-01.html

[5] S. Kandula, D. Katabi, M. Jacob, and A. W. Berger. (2004). *Botz-4-Sale: Surviving Organized DDoS Attacks that Mimic Flash Crowds*. MIT,Tech.Rep.TR-969. [Online]. Available: http://www.usenix.org/events/nsdi05/tech/kandula/ kandula.pdf

[6] S. Z. Yu and H. Kobayashi, "An efficient forward-backward algorithm for an explicit duration hidden Markov model," *IEEE Signal Process. Lett.*, vol. 10, no. 1, pp. 11–14, Jan. 2003.

[7] T. Peng ,C. Leckie, and K. Ramamohanarao, "Protection from Distributed Denial of Service Attacks Using History- based IP Filtering," June 2003, vol. 1, pp. 482 - 486

[8] J. Mirkovic and P. Reiher, "A Taxonomy of DDoS Attack and DDoS Defense Mechanisms," *Computer Journal of ACM IGCOMM*, vol. 4, no. 2, pp. 39-53, 2004.

[9] C. Douligeris and A. Mitrokotsa, "DDoS attacks and defense mechanisms: Classification and state-of-the-art," *Telecommunications Networking*, vol. 44, no. 5, pp. 643– 666, Apr. 2004.

[10] J. Mirkovic, G. Prier, and P. L. Reiher, "Attacking DDoS at the source," in *Proc. 10th IEEE Int. Conf. Network Protocols*, Sep. 2002, pp. 312–321.

Sreeja Rajesh is with Department of Computer Science & Engineering, Jyothi Engineering College, Cheruthuruthy. She completed B.E in Computer Science from Mangalore University in 2000 and completed M.Tech. in Computer Science [specialization in Data Security] from Cochin University of Science & Technology in 2012. She has 13 years of teaching experience in various Self Financing and Regional Engineering Colleges. Her professional career outside academia includes creating awareness about environment protection among common people by conducting environment related rallies, conferences and programs.