Survey on Defending Web-Proxy Based Attacks
Akhila K Raju 1, Misha Ravi 2
Department of Computer science and Engineering
Sree Buddha College of Engineering, Elavumthitta, Pathanamthitta, India

Abstract:
Different server side defending mechanisms exists to resist web-proxy based distributed denial of service attacks (DDoS). DDoS is one of the most vulnerable threats that affect the client and server communication. The client can access services from server through different proxy server. Attackers can use this web proxy as an attacking tool by sending malicious requests to server through proxy. Defending such attacks by Web proxies is a tedious task. In proxy-to-server traffic, there is no foreseeing difference between the legitimate traffic and the attack traffic. In such cases, a server denies services for a proxy server as a whole. Thus along with an attacking client the legitimate users also need to suffer with DoS. This survey paper discusses various methods to identify web-proxy based attacks.

Keywords: web-proxy, network traffic, distributed denial of service, attack detection.

I. INTRODUCTION
Distributed denial of service has become one of the most vulnerable hazards to network system. Comparing with DoS attacks, the DDoS attack are more complicated and is a tedious task to detect them. The typical method is a network traffic flood DDoS attack to Web servers, in which multiple sources attack the particular target at the same time. Usually botnets are used to run these attacks. A DDoS attack can be throw in two steps. First, an attacker creates an attacking network that is distributed and includes a large number of compromised computers (zombies). Then, flood a huge amount of traffic towards victims. Second, through automated malicious codes, such as worms and allow the attacker to install DoS attacking programs that can scan other hosts, install flood packets. The reason for the increase of DDoS attacks is due to the basic design of internet. Attackers can abuse the large amount of resources available in the internet such as routers to deliver flood attacks to the target system.

The proxy server is the mediator between the client and server connection. Requests from the client are received by the proxy server and decide to forward it to the server. A proxy server can be used as an attacking tool by the following ways: the attacker sends malicious requests to the proxy and forces it to send to the server; attacker drops the connection between itself and proxy server. The web proxy-based HTTP attacks are not easy to detect, the reason is that: real attacking hosts are hiding behind the proxy server; server only knows the information from the proxy server. According to the server there is no foreseeing difference between the normal and vulnerable traffic in proxy to server traffic. Thus, it is difficult for the server to monitor and detect the attack requests.

II. LITERATURE SURVEY
A. Anomaly-based network intrusion detection
Intrusion Detection System (IDS) is a security tool that is used to improve the security of data and network communication. IDS are mainly categorized into two, signature-based or anomaly-based [2]. Both systems are similar, main difference between them are in the concept of attack and anomaly. An attack is simply a sequence of task that makes the system security vulnerable. And, an anomaly is “an event that is suspicious from the perspective of security”. A-NIDS (anomaly-based network intrusion detection systems) are emerging concept to protect systems and networks from vulnerabilities. This can be mainly grouped into three statistical based, knowledge-based, and machine learning-based.
In statistical-based A-NIDS method, a profile is created which includes the network traffic behavior. This is based on the IP address, traffic rate, number of packets, etc. In the process of anomaly detection, current behavior is compared with the previously trained behaviour and checks whether it cross certain threshold value. The Knowledge-based techniques are one of the most commonly used approaches, in which the data is classified based on certain rules. Initially it consists of a training data, from which different classes and attributes are extracted. Then set of procedures and classification rules are deduced. Finally the data is processed. The machine learning scheme can be applied in different areas such as, Bayesian networks, neural networks, Markov models, Genetic algorithms, etc. It is based on explicit or implicit model that helps to categorize the pattern analysed.

B. Sequence-order-independent network profiling
This system mainly concentrates on the application-layer DDoS (App-DDoS) attacks [3]. The difference of App-DDoS with other conventional DDoS is that, it uses only trustful methods for attacks. An app-DDoS attack sends small packets only through trustful TCP connections such as HTTP and HTTPS and the real IP addresses are used for attacks. Since these attacks mimic the legitimate users, normal requests are indistinguishable from legitimate users. In order to detect App-DDoS this system uses the concept profiling the web browsing behaviour, thus the sequence order of web page can be used to
detect such attacks. But the sequence order may vary according to individuals and browsing behaviours, here proposed a sequence order independent technique to profile the traffic behaviour.

C. Low-Rate DDoS Attacks Detection and Trace-back by Using New Information Metrics

Two new information metrics are used here to detect low-rate DDoS attacks [4]. They are generalized entropy metric and the information distance metric. Information metric helps to identify difference in traffic with different probability distributions. The limitations of existing includes: First, need to train the traffic behavior gradually. Second, they have high false-positive rate. Third, extracting features of normal behavior from vulnerable one is a tedious task. Shannon’s entropy and Kullback–Leibler’s divergence methods are legitimate methods to overcome these limitations. Also an IP trace-back analysis is proposed here. It is the ability to find the source of an IP packet without relying on the source IP field in the packet.

D. Discriminating DDoS Attacks from Flash Crowds Using Flow Correlation Coefficient

Flash crowd is an unexpected surge in visitors to a particular Web site, which is typically because of some newsworthy event that just took place. Here proposes a discriminating algorithm to differentiate DDoS attacks from flash crowd with the help of flow correlation method [5]. A flow correlation coefficient is used as a metric to measure the similarity between suspicious flows to differentiate DDoS attacks from genuine flash crowds. The proposed discriminating algorithm works independently for specific DDoS flooding attacks. Thus it is effective against unknown forthcoming flooding attacks.

E. Trace-back of DDoS Attacks Using Entropy Variations

A trace-back technique is used here to detect DDoS attacks with the help of Entropy variations [6]. This entropy variation is based on the difference between the normal traffic and the DDoS attack traffic. IP trace-back policy is used here, two major methods for IP trace-back are: probabilistic packet-marking (PPM) and the deterministic packet marking (DPM). Both of them require routers to mark each packet and they are vulnerable to hacking. Since IP trace-back must be free from packet pollution, the proposed method does not include packet marking. The packets that are moving through routers can be classified as flows, which are interpreted by the upstream router where a packet came from, and the destination address of the packet. If a DDoS is recognized, the victim identifies the position of zombies by initially find which of its upstream routers are in the attack based on the flow entropy variations. Then it forwards the requests to their immediate upstream routers to identify the attacker sources. This process continues until it reaches the attack source. Advantages of this method include: it is free from the limitations of packet marking mechanism; it is efficient to detect packet flooding DDoS attacks.

F. A Large-Scale Hidden Semi-Markov Model for anomaly Detection on User Browsing Behaviours

Hidden Semi-Markov Model (HSM) is an extension of Hidden Markov Model (HMM) is proposed here to describe the web browsing behaviours [7]. The underlying process of the HSMM forms the semi-markov chain. Each state in the markov chain has a variable duration that corresponds to the number of observations produced while in the state. HSMM is used in different fields such as, speech recognition, anomaly detection for network security, recognition of human genes in DNA, language identification, brain functional MRI sequence analysis, channel modelling, Internet traffic modelling, speech synthesis, image segmentation, etc. Here HSMM is used for anomaly detection on user browsing behavior; it depends on the structure of a website, which includes a large amount of web documents, hyperlinks, and the way the user accesses the webpages. A new efficient algorithm (M-algorithm) is also proposed here for implementation of the forward process of HSMM and detecting on-line App-DDoS attacks.

G. Monitoring the Application-Layer DDoS for Popular Websites

In the case of detecting App-DDoS for popular websites, a new technique is proposed based on document popularity [8]. The concept of spatial-temporal pattern is used for detection of DDoS attacks. That is, an Access Matrix (AM) is defined to access the spatial-temporal patterns of normal traffic, and then HSMM is used to find the variations in AM and apply two analysis techniques for multidimensional data for HSMM. They are principal component analysis (PCA) and independent component analysis (ICA). PCA is based on basing on converting large amount of variables into a smaller number of uncorrelated variables by detecting small number orthogonal linear combinations of the real variables which have the highest difference.

H. Resisting Web Proxy-Based HTTP Attacks By Temporal and Spatial Locality Behavior

The temporal and spatial locality is used to access the network traffic to defend the web proxy based DDoS attacks [1]. Temporal locality means the reuse of particular data or resources, within a little time duration. That is, if a specific memory location is referenced at a particular point then there is a probability that the same location will be referenced in the near future. Spatial locality is defined as the use of data within close storage locations; if a specific storage position is referenced at a particular time, then a probability exists to reference the adjacent storage positions in the near future. The HSM is used here to access the traffic behaviour of web proxies. Also, a soft control scheme is included to make the detection process better; it transforms the vulnerable request to normal by reshaping rather than roughly dispose it. The advantages of this method includes: it provides good detecting capability, it is independent from gradually fluctuating web contents, and it provide fast attack detection than existing systems. When detecting web proxy based DDoS attacks, instead of denying services to the proxy server, this system provide services to maximum legitimate users. That is, it
deliver high false positive rate (FPR) compared to other techniques.

III. CONCLUSION
The web proxy based DDoS attacks are becoming more vulnerable to network security. Thus strong defending methods needed to resist such kinds of attacks. This survey paper studies different techniques that are used to detect the DDoS attacks. Also, the advantages and limitations of each method are analyzed here.

REFERENCES


