NOVEL TRENDS AND TECHNIQUES USABLE FOR SOPHISTICATED APPLICATION LAYER DENIAL OF SERVICE ATTACKS DETECTION

1,2. UNIVERSITY OF ŽILINA, FACULTY OF ELECTRICAL ENGINEERING, DEPARTMENT OF TELECOMMUNICATION & MULTIMEDIA, ŽILINA, SLOVAKIA
3. UNIVERSITY OF LINKÖPING, DEPARTMENT OF COMPUTER AND INFORMATION SCIENCE, DIVISION FOR DATABASES AND INFORMATION TECHNIQUES, LINKÖPING, SWEDEN

ABSTRACT: As increasing number of security threats and attacks continuously appear and security in the network has become a basic requirement, the need of developing flexible, reliable and automated security mechanisms that can detect and respond to threats in real time has posed a big challenge for researches. This paper focuses on description of Application layer Denial of Service (DoS) and Distributed Denial of Service (DDoS) attacks, which present a continuous critical threat to the Internet services. Over some period of time, researchers proposed many solutions to prevent the DoS/DDoS attacks from different OSI layers, but there has been done only a very small research on application layer. In this paper, we consider sophisticated attacks that utilize legitimate application layer requests from legitimately connected network machines to overwhelm Web server. In this paper we propose several known mechanisms to combat application layer DoS/DDoS attacks continuing with proposing most recent approaches and trends which are concurrently under the development.

KEYWORDS: denial of Service, application layer, Intrusion Detection System

INTRODUCTION

Network security and security policy is a frequently discussed topic in state of the art computer world. The obligation of protecting sensitive information, data and services placed on computer networks and Internet is a obvious question today. There are lots of new threads appearing quite often and countermeasures against them have to be taken. Usually, computer threads can be categorized into four main classes, like: reconnaissance attacks, password attacks, denial of service attacks and malware. This paper will talk more about the third mentioned attack type, Denial of Service (DoS) attacks and problems associated with the appearing defense mechanisms to counter this kind of attack.

A. Denial of Service Attack Description

Several years ago when Denial of Service attacks started to appear quite frequently, the need of defending networks and servers against this kind of security threat became serious. This need to protect servers and other network systems is an important aspect in network security as it requires only a little effort to execute DoS attack. Today plenty of application servers and network facilities may suffer from DoS and Distributed Denial of Service (DDoS) attacks and that is why it is needed to wide inform on what mechanisms these attacks work, in what manner are these mechanisms evolving and how to defend servers and network systems against this malicious activity. [1]

The main goal or purpose of DoS and DDoS attacks is to prevent authorized hosts from using a service. The service can be either for free or it can be paid, the attacker doesn’t differentiate due to the service fee. It is important to notice that DoS and DDoS attacks differ from other classes of computer and security attacks with the purpose of the attack. The goal is not to steal or misuse the sensitive data, DoS/DDoS attacks aim at creating network congestion or overloading the application server by generating a large amount of traffic addressed to the victim. Usually, a malicious user blocks legitimate users from accessing network services by exhausting or depleting the resources of the victim’s server. [2]

DoS and DDoS attacks are aimed at any network device but most often at application layer servers, like DNS servers, electronic mail servers or web servers to make the most popular services unavailable for users. This can be done by several approaches, but most usually either by consuming the network bandwidth, the CPU cycles or by consuming the RAM memory of the victim device. Due to the attack performance, DoS attacks can be categorized in the manner of what damage they cause into three main categories:

- Destructive DoS attacks
- Resource consumption DoS attacks
- Bandwidth consumption DoS attacks

It is evident that in the first case the device stops to work normally. In this case, the attacker can cause
power interruption or destruction of some configuration information. It can be said that this is the simplest way how to interrupt the accessibility to the service but on the other hand it can have most serious consequences. Second example of DoS attack impact is resource consumption where the principle of attack is to overuse resources of the victim’s hardware. In the same manner bandwidth consumption attacks simply consume bandwidth capacity of a network by sending bogus requests to the victim server thus clogging the subnetwork of victim’s origin with fake traffic. [2]

B. Distributed Denial of Service and Distributed Reflector Denial of Service Attacks

Distributed Denial of Service is a special kind of DoS which goal is to increase the attacks intensity by using a number of computers. DDoS attacks are considerably more effective than DoS because they allow increasing the attack intensity by simultaneous use of number of computers. DDoS attacks represent a frequent disturbance to services hosted on high-profile web servers such as banks, credit card payment gateways, insurance companies and others.[3] DDoS occurs when multiple systems flood the bandwidth or resources of a targeted system, what makes the attack more efficient and complicates searching out the originator of the attack. [4] Distributed Denial of Service is usually performed within a logical structure, which can be seen on Figure 1.

![Figure 1. Structure of Distributed Denial of Service Attack](image1)

The layered structure consists of a client, who is an attacker and who connects to several compromised system called handlers. These handlers indicate commands to multiple zombie agents which in turn facilitate the DDoS attack on the target host or system. By the way, each handler can control up to thousand agents. The attacker usually uses in Internet Relay Channel for communication with the agents. Many systems can be compromised by an attacker using a variety of methods because operating systems and network protocols were developed without applying security engineering. This results in providing hackers a lot of vulnerable machines on Internet which can be misused by an attacker for building an army of attackers. The attacker usually misuses known vulnerability of a computer system and implements a malicious code into the victim. Created zombie botnet then simply attacks the victim with a large amount of traffic. Distributed Denial of Service attacks can be divided into two main categories:

- Flooding attacks
- Vulnerability attacks

Flooding DDoS attacks consume resources such as network bandwidth by overwhelming bottleneck link with a high volume of traffic. DDoS flooding is basically a resource overloading problem. By the term of resource can be understood bandwidth, memory, CPU cycles, file descriptors and buffers etc. [5] Thus, service is denied to legitimate users due to limited bottleneck bandwidth. However, resources of distribution networks are not a problem in case of commercial servers if these are situated quite close to their backbone network with high bandwidth access links. But hardware resources of the server such as processing capacity, buffer limit etc., are put under stress by flood of seemingly legitimate requests generated by DDoS attack zombies. Each request consumes some CPU cycles and once the total request rate overlaps the service rate of a server, the requests start to be placed in a buffer of the server and after some time the buffer gets overfilled. Due to buffer over run, other incoming requests are dropped. The congestion and flow control signals then try to force legitimate clients to decrease their rate of sending requests, whereas attacking packets keep coming. Finally, there comes a state when there is only the attacking traffic reaching at the server. Thus, service is denied to legitimate clients. Moreover, in [6] Robinson highlights that as attack strength grow by using multiple sources, the computational requirements of even filtering traffic of malicious flows become an additional burden at the target. Vulnerability attacks use the expected behavior of the protocols such as TCP and HTTP to the attacker’s advantage. The computational resources of the server are then tied up by seemingly legitimate requests of the attackers and thus prevent the server from processing requests from authorized users. Almost all of DoS and DDoS attacks are targeted at TCP based services.

DDoS attacks can result in a great damage to network services. Since they can rapidly degrade the network performance and are difficult to detect, they have become one of the most serious security challenges to the current Intrusion detection systems. However if DDoS attacks are detected in sufficient short time, the loss caused by this attack can be reduced to minimum. So far, effective and complex solutions to defeat all features of DDoS attacks haven’t been found yet. Therefore, DDoS attack detection is still an attractive area for researches. [7]

An important goal for an attacker is to stay anonymous or to hide true source of the attack traffic. For these purposes attackers found out a novel methodology of attack called Reflector Denial of Service (RDoS). On the Figure 2 can be seen a principle of Reflective Denial of Service attack.
Most DDoS detection related research was focused on the IP layer, because there are a lot of parameters, which can be monitored. These mechanisms attempt to detect attacks by analyzing specific features like SYN/FIN packets for SYN flooding attacks detection. In [15], DDoS attacks were discovered by analyzing the TCP packet header against the predefined rules and conditions and afterwards the difference between normal and abnormal traffic was distinguished. Noh [16] attempted to detect attacks by computing the ratio of TCP flags (including FIN, SYN, RST, PSH, ACK, and URG) to TCP packets received at a Web server. However, only a little work has been done on application layer for DoS/DDoS attack detection because Application DDoS attacks appeared only a few times in the past. Ranjan [17] used statistical methods to detect characteristics of HTTP sessions and employed rate-limiting as the primary defense mechanism. Other researchers combatted the App-DDoS attacks for example by puzzle [18]. In [19], there are used two properties to distinguish the DoS/DDoS and normal flash crowd:

1. a DoS event is due to an increase in the request rates for a small group of clients while flash crowds are due to increase in the number of clients,
2. DoS clients originate from new client clusters as compared to flash crowd clients which originate from clusters that had been seen before the flash event.

But none of these approaches can be thought as a complex and effective solution for Application DDoS attacks detection, which makes this area still attractive for researchers.

APPLICATION DOS/DDOS ATTACKS

DoS and DDoS attacks have caused severe damage to network devices and services. In the past it was most common to execute these attacks at network and transport layer such as ICMP flooding, SYN flooding, or UDP flooding. The intent of these attacks is to consume the network bandwidth or overload the number of possible parallel connections and deny the service to legitimate users. Many studies have explored this type of attack and many different schemes how to protect the network have been proposed. Because of that, it is not as easy as in the past for attackers to launch the DDoS attacks based on network layer. This has led to uncovering a new, sophisticated strategy on how to overload network
devices, such as application servers. The main difference between Application DDoS and network layer oriented DDoS is in exploiting vulnerabilities of application protocols. This means that connections on network and transport layer have to be estimated correctly. This makes the detection of the attack much more complicated because it is difficult to differentiate attacking traffic from the legitimate one. [20] This performance of attack requires less number of attacking traffic because the goal is to reach resource limits of a concrete service, which is always lower than the total amount of possible TCP or UDP connections. This way can the attacker exhaust all the possible connections on application layer regardless to hardware limitations of the victim’s device. Most times are targeted frequently used application servers like HTTP server, FTP server, SMTP server etc. A lot of application DDoS attacks target HTTP, in which case they aim to exhaust the resource limits of Web services.

If we are interested in HTTP or HTTPS service then there are two protocol weaknesses which are usually misused:

HTTP GET
HTTP POST

HTTP GET flood attack is performed with the misuse of the first mentioned weakness of HTTP protocol. In this type of attack, attackers send a large number of malicious HTTP GET requests to a target server [21, 22]. Since these packets have legitimate HTTP payloads, victim servers cannot distinguish normal HTTP GET requests from the malicious requests. Thus servers have to serve all requests as normal requests, and they exhaust their resources finally.

Another attacking approach is used when the attacker performs Slowloris attack. It is also based on HTTP GET request weakness, but the victim is not flooded with spoofed requests but it uses time delayed HTTP GET headers. In principle, the attacker doesn’t send HTTP GET request header on one time, but he separates the lines of the header and sends each line separately. The web server creates the connection with the attacker and simply waits until the end of the request header and this could take a long time. This way is the connection reserved for the malicious request for a long period of time. There is a default threshold, which indicates the maximum timeout when has the next line of header arrive, otherwise will be the connection closed. On Apache web servers it is usually 300s. This time is then set as a break time for sending next line of request header on the attacker side. An attacker can than exhaust web server resources with multiple connections created in this manner. [23]

Last most used attack strategy is to misuse weakness of HTTP POST request. A POST request includes a message body, which can use any encoding. The field Content-Length in the HTTP Header tells the web server how large the message body is. The attacker then completes HTTP Header portion and sends it in full to the web server. The attacker then sends HTTP message body in sequences for example 1 Byte per 110 seconds. Web servers will just obey the Content-Length written in the header field and wait for the remaining message body to be sent. By waiting for the complete message body to be sent, web servers can support users with slow or intermittent connections. When there is multiplied such type of connections, the web server gets under DDoS attack.

APPLICATION LAYER DOS/DDOS ATTACK DETECTION POSSIBILITIES

There are several reasons for attack detection. First, if a target can detect an attack before the actual damage occurs, the target can get more time to implement attack reaction and protect legitimate users. Second, attack detection can help to identify the attackers so that legal actions can be taken. Third, if attacks can be detected close to attack sources, attack traffic can be filtered before it wastes any network bandwidth. However, there is generally insufficient attack traffic in the early stage of an attack and in the links close to the attack sources. Consequently, it is easy to mistake legitimate traffic as attack traffic. Therefore, it is challenging to accurately detect attacks quickly and close to the attack sources.[24]

There has been done only a little research in the past about detecting application layer Dos/DDoS attacks because this type of attack is quite new and it wasn’t executed that often in the past. Here are highlighted some older techniques used for application layer DDoS detection:

Client Puzzle Protocol
Client Puzzle Protocol (CPP) is an algorithm for use in Internet communication, whose goal is to prevent abuse of server resources. The idea of the CPP is to necessitate from all clients connecting to a server to correctly solve a mathematical puzzle before establishing a connection, during the time, when the server is probably under attack. After solving the puzzle, the client would return the solution to the server, which the server would quickly confirm or reject and drop the connection. The puzzle is made simple and easily solvable but requires at least a minimal amount of computation on the client side. Non malicious users would experience just a negligible computational cost but attacking clients that try to simultaneously establish a large numbers of connections would be unable to do so because of the computational cost (time delay). This method holds promise in fighting some types of spam as well as other attacks like Denial of Service.

Ingress Filtering
In computer networks, ingress filtering is a technique used to make sure that incoming packets don’t have spoofed source IP addresses in their headers. Generally networks receive packets from other networks. Normally a packet will contain the IP address of the computer that originally sent it. This allows other computers in the network to know where it came from, which is needed for things like sending a packet back to the sending computer. In certain cases, the sending IP address will be spoofed. This is typically done as part of an attack, so that the attacked computer does not know where the attack is really coming from. In ingress filtering, packets coming into the network are filtered based on previous gained information from the originating
Anomaly-based detection has become a major focus of research, due to its ability to detect new attacks, including DoS attacks. [27] Building a normal profile is most times the first step for all anomaly-based detection techniques. Since there is no clear definition of what is normal, statistical modeling plays a crucial role in constructing the normal profile. Statistical anomaly detection includes two major parts. First part is to find effective parameters to generate similarity measures. The parameters can be IP packet length, IP packet rate, and so on. The second part is to calculate the similarity between the normal profile and new traffic. If the distance between the monitored traffic and the normal traffic profile is larger than a given threshold, a DoS/DDoS attack is detected. [28] The common challenge for all anomaly-based intrusion detection systems is that it is difficult or almost impossible for the training data to provide all types of normal traffic behavior. As a result, legitimate traffic can be classified as attack traffic, causing a false positive. To minimize the false positive rate, a larger number of parameters are used to provide more accurate normal profiles. This on other hand degrades the detection speed, which is actually very important.

Thus current research activities in the field of network intrusion detection of application oriented attacks focuses mostly on anomaly based intrusion detection and at the present, most approaches and techniques applied in the detection process are related to machine learning.

### CURRENT TRENDS IN APPLICATION LAYER DOS/DDOS ATTACK DETECTION

A typical complex detection tool that also uses anomaly-based detection approach is Anomaly Based Intrusion Detection System (AB IDS). This system can be later classified into two different groups based on whether it analyses the features of each packet separately or if it analyses the whole connection. Concerning this feature, there is a distinction of IDS between

- Packet-oriented and
- Connection-oriented systems.

A packet-oriented system uses a single packet as minimal information source, while a connection-oriented system considers features of the whole communication before establishing whether it is anomalous or not. Theoretically, a connection-oriented system could use as input the content (payload) of a whole communication, which would allow a more precise analysis. But this would require a longer computational time, which could limit the throughput of the system by introducing extra latency time.

In practice, a connection-oriented system typically takes into account the number of sent or received bytes, the duration of the connection and transport layer protocol used. As written in [8], most AB IDSs in practice are packet-oriented.

Based on [29], there are two adequate measures for anomaly-based intrusion detection at the application layer: payload length and payload histograms. Main conclusion of their experimental work is that the payload length should not be used as an isolated feature for distinguishing between normal and anomalous traffic. However, its use in conjunction
with other features is shown to be a good choice, since it contributes information related to the normality of the payload. The normality degree of a given payload could be evaluated with the use of conditional probabilities. Those probabilities can be achieved by using a Markov chain. In this context, [20] proposes an anomaly-based scheme to detect attacks against the HTTP service that follows the basic idea of modeling the payload as a Markov process. The HTTP specification defines a common structure for every payload, which is composed of several sections each containing different information units. Since each section has its own set of allowed values according to its purpose and semantics, it is natural to suppose that the probability of occurrence of certain strings within each section of the payload is not uniform throughout the request. Some proposals how to detect anomalies in application-layer traffic have been already made. As written in [30], features, or monitored parameters of an application query, that are usually considered as relevant for detecting application layer malicious activity are:
- Attribute length
- Attribute character distribution
- Attribute presence or absence
- Attribute order

Current research activities in the field of anomaly-based network intrusion detection of application layer DoS attacks, as stated in [30], are focused on three different directions:
- Statistical-based AB IDS
- Knowledge-based AB IDS
- Machine learning-based AB IDS.
In statistical-based techniques, the network traffic activity is captured and a profile representing its stochastic behavior is created. This profile is based on metrics such as the traffic rate, the number of packets for each protocol, the rate of connections, the number of different IP addresses, etc. The desired model for knowledge-based IDS is constructed manually by a human expert, in terms of a set of rules that seek to determine legitimate system behavior. If the specifications are complete enough, the model will be able to detect illegitimate behavioral patterns. Machine learning techniques are based on establishing an explicit or implicit model that enables the analyzed patterns to be categorized. A singular characteristic of these schemes is the need for labeled data to train the behavioral reference model. At present, most approaches and techniques applied in the detection process are related to machine learning. Most important machine-learning schemes are:
- Bayesian networks
- Markov models
- Neural networks
- Fuzzy logic techniques
These are most recent approaches and trends concurrently under the development of Intrusion Detection Systems with the focus on application layer DoS attacks detection.

CONCLUSIONS

In this paper, we focused on DoS/DDoS attack description and consequently aimed the attention at the detection of Application Layer DoS/DDoS attacks and presented methodologies used for attack detection. While most current effort focuses on detecting DDoS attacks performed at network and transport layer with stable background traffic, we proposed two main detection architectures aiming at monitoring Web traffic on application layer in order to discover dynamic changes in normal burst traffic. Signature based DoS attack detection techniques generally use one or more features of DoS attacks, and can identify attack traffic effectively. However, all these techniques are based on one or more assumptions, which are not always reliable. On the other hand anomaly based detection techniques are facing a dilemma of how to choose a tradeoff between processing speed and detection accuracy. Beside this we presented also most current methodologies used by anomaly-based Intrusion Detection System for application layer DoS attack detection.

Recent intrusion detection techniques combine and correlate information from different detectors while individual detectors are designed to monitor only a specific protocol or behavior. At present, most approaches are related to machine learning by Markov models, anagrams and others.

REFERENCES


