International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, New York, USA
Volume 4– No.6, December 2012 – www.ijais.org

Detection of DDoS Attack in Semantic Web

Prashant Chauhan
BISAG
Gandhinagar
Gujarat, India

Abdul Jhummarwala
BISAG
Gandhinagar
Gujarat, India

Manoj Pandya
BISAG
Gandhinagar
Gujarat, India

ABSTRACT

We are currently living in an era where information is very crucial and critical measure of success. Various forms of computing are also available and now a days it is an era of distributed computing where all the task and information are shared among participated nodes. This type of computing can have homogeneous or heterogeneous platform and hardware. The concept of cloud computing and virtualization has gained much momentum and has become a more popular phrase in information technology. Many organizations have started implementing these new technologies to further reduce costs through improved machine utilization, reduced administration time and infrastructure costs.

Cloud computing also faces challenges. One of such problem is DDoS attack so in this paper we will focus on DDoS attack and how to overcome from it using honeypot. For this here open source tools and software are used.

Typical DDoS solution mechanism is single host oriented and in this paper focused on distributed host oriented solution that meets scalability.

General Terms

Cloud Security, Network Monitoring, Distributed Log Processing.

Keywords

Cloud Security, DDoS, Honeypot, Hadoop.

1. INTRODUCTION

In cloud computing hardware, software etc. are available to the end user in form of services. The user can subscribe for the required service and he will be charged on the go. Means he needs not to pay while resources are not in use. In cloud security we have basic three characteristics confidentiality, integrity and availability.

Availability is one of important aspect in cloud security stating availability of service while user want to consume but due to some security threats availability can not be achieved. DDoS attack is one such threat which is distributed form of Denial of Service attack in which service is consumed by attacker and legitimate user can not use the service.

We can find solution against DDoS attack but they are based on single host and lacks performance so here Hadoop system for distributed processing is used.

2. Cloud Computing

In cloud computing every resource is available to end user in form of service and user pay only for what he has used. This kind of freedom for the user leads cloud services more popular. Cloud computing services are separated by layer shown in fig 1.

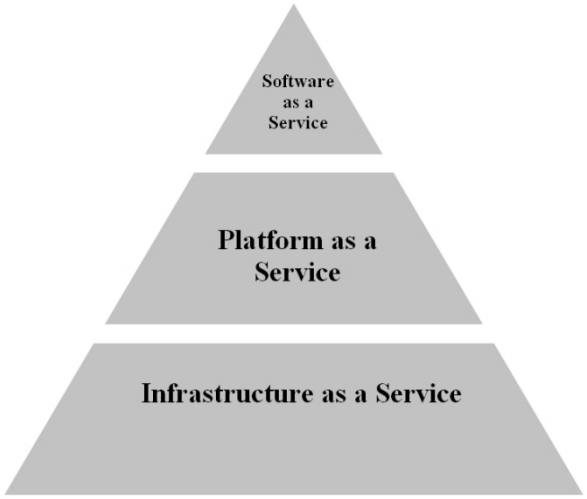

Fig 1 Layers of Cloud Computing

1. Application Layer

This layer provides Software as a Service (SaaS). Applications are provided on demand basis in this layer. Highly scalable internet based applications are hosted on the cloud and offered as a service to the end user. Google Docs, SalesForce.com, Acrobat.com are some popular SaaS.

2. Platform Layer

This layer provides Platform as a Service (PaaS). In this layer platforms used to design, develop, build and test application are provided by the cloud infrastructure. Some popular PaaS providers are Google App Engine, Azure Service Platform.

3. Infrastructure Layer

This layer provides Infrastructure as a Service (IaaS). In this layer hardware and network equipment are provided as service. It also provides storage and database management and computing capabilities on demand. Some well known IaaS providers are Amazon Web Services, Go Grid and 3 Tera.

3. Cloud Security

Cloud security has three basic aspects called confidentiality, Integrity and Availability. This are also called basic security goals. Fig 2 shows cloud security goals and availability is focused in this research paper. DDoS attack is an issue of availability of any cloud service.

A. Confidentiality

Confidentiality is the prevention of the intentional or unintentional unauthorized disclosure of contents. Sometime it is referred as Secrecy or Privacy in other terms. It means that Confidentiality is the prevention of the intentional or unintentional unauthorized disclosure of contents. Only authorized person can read, write, view, print and know about the content is available.

To ensure confidentiality network security protocols, Data Encryption services, Authentication services are used.

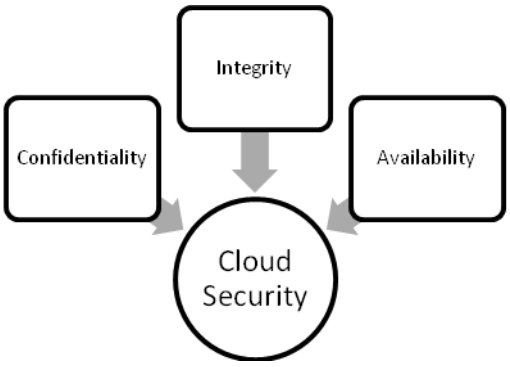

Fig 2 Cloud Security Goals

B. Integrity

Integrity is the guarantee that message sent is received without changing and the message is not intentionally or unintentionally altered.

To ensure integrity Firewalls, Intrusion Detection Systems etc. are used.

C. Availability

Availability is the element that creates reliability and stability in the system. It means that the data is available whenever asked.

To ensure Availability redundant disks and network system and Backup utility are used.

4. Distributed Denial of Service Attack

4.1 Definition and Purpose

A denial-of-service attack (DoS attack) is an attempt to make a computer resource unavailable to its intended users. Distributed denial-of-service attack (DDoS attack) is a kind of DoS attack where attackers are distributed and targeting a victim. It generally consists of the concerted efforts of a person, or multiple people to prevent an Internet site or service from functioning efficiently or at all, temporarily or indefinitely.

Motives of attackers to perform Denial of Service attack could be of different nature; an attacker might want to show his power by attacking a large, popular Web site that will permit him or her to be recognised in the underground community. Another possibility of motive is a political one, for example, a political party can ask an attacker to do a DoS attack on the Web site of a concurrent political party. Most common motive is the commercial one; a company can ask an attacker to attack a concurrent; during the time where the commercial website is unavailable it will lose lots of money and worst, the trust of user and they might want to go shopping on another website.

4.2 How DDoS Attack works

DDoS attack is performed by infected machines called bots and group of bots is called botnet. This bots (Zombie) are controlled by an attacker by installing malicious code or software which acts as per command passed by attacker. Bots are ready to attack anytime upon receiving command from the attacker. Many types of agents have scanning capability that permit to identify open port of a range of machines. When the scanning is finished, the agent takes the list of machines with open port and launches vulnerability-specific scanning to detect machines with un-patched vulnerability. If the agent found a machine with vulnerability, it could launch an attack to install another agent in the machine.

Fig 3 shows DDoS attack architecture explaining its working mechanism.

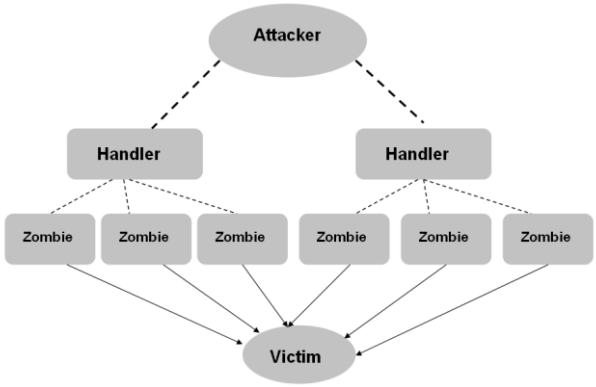

Fig 3 DDoS attack Architecture

For controlling agents, the attackers use Command and Control (C&C) server, currently, the majority of botnet use Internet Relay Chat (IRC). Other possibilities exist as Peer-to-Peer (P2P) C&C, Instant Messaging (IM) C&C or even Web-based C&C server. When the agent is connected to the C&C it can ask for updates as IP address for other C&Cs, software update or exploit software. The agent could also ask for orders, if the agent is newly installed it could ask order for protect himself, this could be by asking the C&C the location of the latest anti-virus and prevent it to detect the agent by stopping the service.

A DoS attack's main characteristic is that an attacker attempts to prevent one or more legitimate users of a service from the use of the required resources. There are two general forms of DoS attacks: those that crash services and those that flood services. Therefore, he attempts (1) to inhibit legitimate network traffic by flooding the network with useless traffic. (2) to deny access to a service by disrupting connections between two parties, (3) to block the access of a particular individual to a service, or (4) to disrupt the specific system or service itself. Here http flooding as a DDoS attack is considered and its solution is proposed.

5. Hadoop MapReduce

5.1 Introduction

MapReduce, developed by Google, is a software paradigm for processing a large data set in a distributed parallel way. Since

Google's MapReduce and Google file system (GFS) are proprietary, an open-source MapReduce software project, Hadoop, was launched to provide similar capabilities of the Google's MapReduce platform by using thousands of cluster nodes. Hadoop distributed file system (HDFS) is also an important component of Hadoop, that corresponds to GFS. Hadoop consists of two core components: the job management framework that handles the map and reduces tasks and the Hadoop Distributed File System (HDFS). Hadoop's job management framework is highly reliable and available, using techniques such as replication and automated restart of failed tasks. The framework has optimization for heterogeneous environments and workloads, e.g., speculative (redundant) execution that reduces delays due to stragglers.

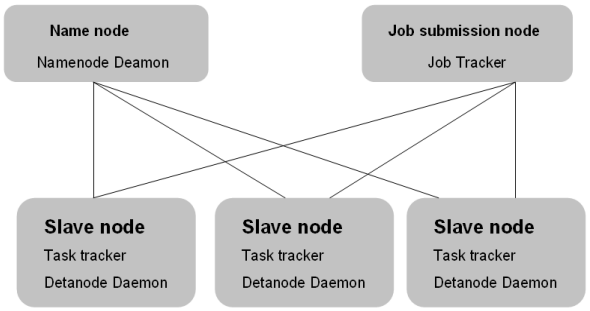

Fig 4 Hadoop Multinode Cluster Architecture

Fig 4 shows Hadoop multinode cluster architecture which works in distributed manner for MapReduce problem.

5.2 DDoS Attack and Hadoop

DDoS attack is critical and find easily but conventional solution are single host oriented. We know that for that we can use honeypot system to overcome from it. Some network monitoring tools like wireshark, nmap are available but they just sniff the network and generate log files in huge size. So to handle this gigantic file we need a special file system like HDFS and MapReduce framework for processing log effectively and efficiently. So Hadoop is a well suited solution here.

6. MapReduce for Counter-based DDoS Detection method

In this method simple MapReduce algorithm to detect DDoS with URL counting is implemented. This algorithm needs three input parameters of time interval, threshold and unbalanced ratio, which can be loaded through the configuration property or the distributed cache mechanism of MapReduce. Time interval limits monitoring duration of the page request. Threshold indicates the permitted frequency of the page request to the server against the previous normal status, which determines whether the server should be alarmed or not. The unbalanced ratio variable denotes the anomaly ratio of response per page request between a specific client and a server. This value is used for picking out attackers from the clients.

The map function filters non-HTTP GET packets and generates key values of server IP address, masked timestamp, and client IP address. The masked timestamp with time interval is used for counting the number of requests from a specific client to the specific URL within the same time duration. The reduce function summarizes the number of URL requests, page requests, and server responses between a client and a server. Finally, the algorithm aggregates values per server. When total requests for a specific server exceeds the threshold, the MapReduce job emits records whose response ratio against requests is greater than unbalanced ratio, marking them as attackers. While this algorithm has the low computational complexity and could be easily converted to the MapReduce implementation, it needs a prerequisite to know the threshold value from historical monitoring data in advance.

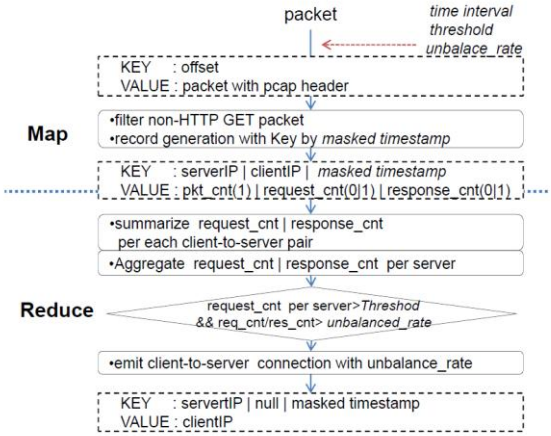

Fig 5 MapReduce for counter-based DDoS detection

7. Result

As far as scalability is concerned, the performance of the counter based DDoS detection method is used by varying cluster nodes. Figure 6 shows statistical graph of the detection job with ten worker nodes was completed within 25 minutes for 500GB and 47 minutes for 1TB, which is over 8 times faster than one node's and 2.9 times faster than 3 nodes' respectively. From evaluation results the increased performance enhancement is observed as the volume of input traffic becomes large.

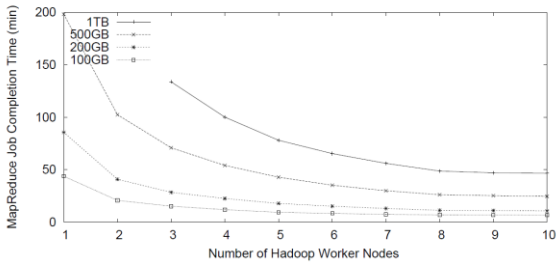

Fig 6 Completion time of a counter-based DDoS detection job regarding various # of Hadoop datanodes and traffic sizes

8. Conclusion

Scalability of the detection of DDoS attack is focused in this paper and focused a Hadoop based DDoS detection model. The approach of Hadoop based solution solves scalability issues by parallel data processing for DDoS attack detection for a huge volume of traffic. Other approaches are single host based and suggesting memory increment for efficiency but in

this paper suggested a Hadoop based solution which solves scalability issue by parallel data processing.

This kind of MapReduce job is preferable for Hadoop as it is Computation oriented framework.

9. REFERENCES

[1] NathalieWeiler, Eleventh IEEE International Workshops on Enabling Technologies,2002, "Honeypots for Distributed Denial of Service Attacks"

[2] Yuqing Mai, Radhika Upadrashta and Xiao Su, ITCC'04 "J-Honeypot: A Java-Based Network Deception Tool with Monitoring and Intrusion Detection"

[3] Yeonhee Lee, Wonchul Kang, and Youngseok Lee, 51-63, TMA'11"A Hadoop-based Packet Trace Processing Tool"

[4] Yun Yang, Hongli Yang and JiaMi, IEEE, 2011, "Design of Distributed Honeypot System Based on Intrusion Tracking"

[5] Anup Suresh Talwalkar, 2011, "HadoopT - Breaking the Scalability Limits of Hadoop"

[6] Flavien Flandrin, 2010, "A Methodology To Evaluate Rate-Based Intrusion Prevention System Against Distributed Denial Of Service"