Distributed Denial of Service Attack on Big Data Analytics

Miss. Kalyani R. Gawande [1], Prof. P. V. Kale [2]
[1]Dept of CSE
[2]Assistant Professor, Dept of CSE
[1, 2] PRPCEM, Amravati

Abstract- *Everything now a days become electronic and connected to internet to transmit and store data. Thereforerate of data flow over the internet has been increased. The volume, variety and velocity of data flow has introduced the new term called BigData. There is no concrete definition of Big data, mostly it defines by its 3-Vs properties (Velocity, Volume and Variety). Later on 2 V's are added by IBM and Oracle in Big Data's definition. Traditional tools and techniques are not sufficient to handle Big Data operations like Big Data management, sorting, storing etc. For this purpose distributed file system has been introduced, Hadoop (HDFS, Mapreduce) is mostly used for BigData's operations. Most companies adopted and shifted their business to BigData (Hadoop), but still some companies like Govt. organizations and other security firms are hesitating to shift on it due to its security lags.*

Keywords- DDoS, Hadoop, MapReduce

I. INTRODUCTION

Distributed denial of service (DDoS) attacks continues to grow as a threat to organizations worldwide. From the first known attack in 1999 to the highly publicized Operation Ababil, the DDoS attacks have a history of flooding the vic- tim network with an enormous number of packets, hence exhausting the resources and preventing the legitimate users to access them. After having standard DDoS defense mechanism, still attackers are able to launch an attack. These inadequate defense mechanisms need to be improved and integrated with other solutions. The purpose of this paper is to study the characteristics of DDoS attacks, various models involved in attacks and to provide a timeline of defense mechanism with their improvements to combat DDoS attacks. In addition to this, a novel scheme is proposed to detect DDoS attack efficiently by using Map Reduce programming model.
By increasing the usage of electronic media and shiftedeverything on internet, the rate of data flow over internet has been increased; this rate of data production which is structured and unstructured can't be processed and managed via traditional methods [1]. Structured data is the well sorted data which can be stored in traditional databases like library catalogues, census etc. whereas Unstructured data is the data which cant be manage, store by the traditional databases like videos, pictures, media files, etc [1]. This structured and unstructured data having 3Vs features (Velocity, Volume and Variety) and known as Big Data, later on 2 more Vs has been added to bigdata definition which is (Value and Veracity). To handle big data the distributed framework has been introduced, Hadoop is one of famous framework used for Bigdata management, sort, store and distribution [2].

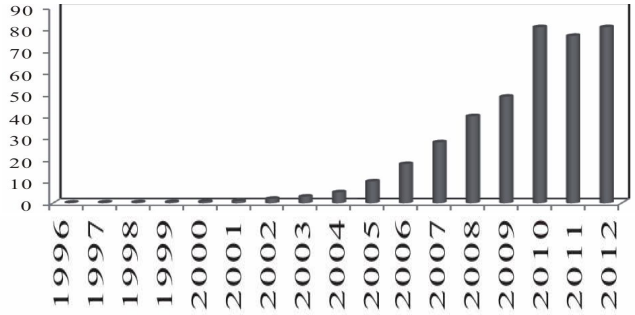
DDoS attack is a distributed, large scale coordinated attempt of flooding the network with an enormous amount of packets which is difficult for victim network to handle, and hence the victim becomes unable to provide the services to its legitimate user and also the network performance is greatly deteriorated [3]. This attack exhausts the resources of the victim network such as bandwidth, memory, computing power etc. The system which suffers from attacked or whose services are attacked is called as "primary victim" and on other hand "secondary victims" is the system that is used to originate the attack. These secondary victims provide the attacker, the ability to wage a more powerful DDoS attack as it is difficult to track down the real attacker [4].

Denial of Service (DoS) attacks is used to consume all the resources of the target machine (victim's services) and becomes a known issue in 1980's. But, in 1990's these attacks have been noticed as it becomes a serious problem to the Internet society gradually [3-6]. DDoS attack is a distributed, large scale coordinated attempt of exhausting the network with an enormous amount of request, which overload the victim's machine and the victim's machine becomes unable to provide the services to its legitimate user and hence the network performance will be greatly deteriorated.

In DDoS attack, the attacker selects the compromised machine (i.e. those machines which have loopholes) and network of the compromised machines are called botnet. These botnets are further instructed to execute commands in order to consume all the resources available on vic- tim's system. Currently attacks are being launched by using two approaches. The first approach is to send mali- cious packet

injected with virus, worms as a running ap- plication, is called as vulnerability attack. The other very common method is to debilitate the victim's system, by exhausting the resources such as input-output bandwidth, database bandwidth, CPU, memory, etc. [7].

A group called "Izz ad-Din al-Qassam Cyber Fighters" [8] has launched DDoS attack against many US Banks such as Bank of America, Citi Group, HSBC and Capital One. As a result, these online banking sites have de-graded. Figure 1 shows various attacks over the years. From the figure, we can see that total number of attacks increases gradually every year. Table 1 shows some se- riousDDoS attack incidents in past years. It is noted that attacks incidents are increasing gradually specially in financial market.



Elementary Principle

1. HDFS
2. MAPREDUCE

Hadoop Distributed File System

It is a subproject of hadoop. It is designed to provide a file system to run on commodity hardware. The objective of HDFS is to store data in the presence of failure including Name node failure, data node failure and network partitions. HDFS uses master /slave architecture which contains master that controls the slaves. It consists of a single name node and master node manages the file system.

Mapreduce

It is a programming model and implementation for processing and generating data sets with parallel algorithm. It consists of a Map ( ) function that performs filtering and sorting and Reduce ( ) function which performs a summary operation. Mapreduce libraries have been written in many languages. But the optimized implementation of map reduce is hadoop. Mapreduce is used to mitigate the ddos attack

FAIR SCHEDULER

Fair scheduler has been developed byfacebook.Fair scheduler is a method which is used to assign resources to jobs. When a single job is running then it uses whole cluster when more than one job is running then cluster is divided into the task slots. This organizes job into pools and divides these pools. This scheduler follows first-in –first-out (FIFO). It can limit number of jobs per pool. Fair scheduler supports hierarchical queues. Jobs are put in the pools and according to the job, user get his own pool of task. Fair scheduler works on the memory by default. It gives priorities to the weights. Basically, it provides fast response time for jobs and its production.

II. DISRIBUTED DENIAL-OF-SERVICE DdOS

DDOS attack is an attempt to make a network unavailable. In this, attackers are distributed and targeting a client. It is unable to access the resource such as e-mail, social sites and other internet application. This attack typically targets on sites or services which are hosted by high profile web services for example banks and credit card payment gateway.

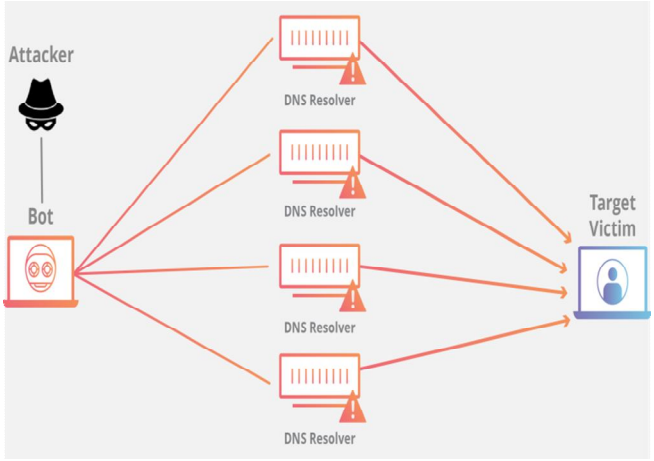


Fig : Architecture of a Ddos Attack

HOW DOES DDOS ATTACK WORK?

DDOS attack attacks on a company's network by flooding the network with the IP or information. Internet can only handle a finite amount of traffic but the flooding targets the system. DDOS attack works by infected machines which is called zombie. These machines are becoming zombie by installing code which is passed by attacker to make it infected machine or victim. Now if the attackers found a zombie, it can install a code for another machine.

III. DDOS ATTACK DETECTION WITH HADOOP

A. Data Generation

Flume's architecture is simple, robust, and flexible. In our system we are using flume tool for data generation or we can say log file generation.The below figure 2 shows a typical deployment of Flume that collects log data from a set of application servers. The deployment consists of a number of logical nodes, arranged into three tiers. The first tier is the agent tier. Agent nodes are typically installed on the machines that generate the logs and are your data's initial point of contact with Flume. They forward data to the next tier of collector nodes, which aggregate the separate data flows and forward them to the final storage tier.

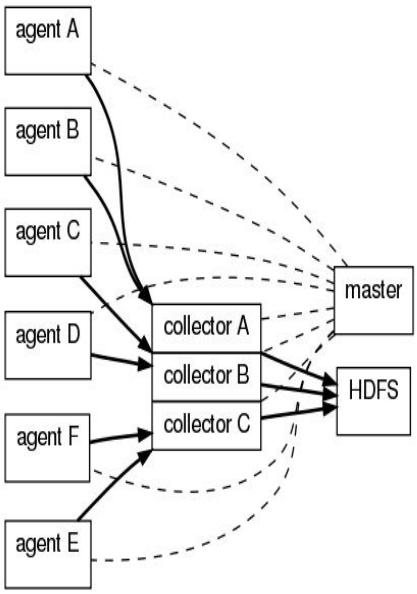
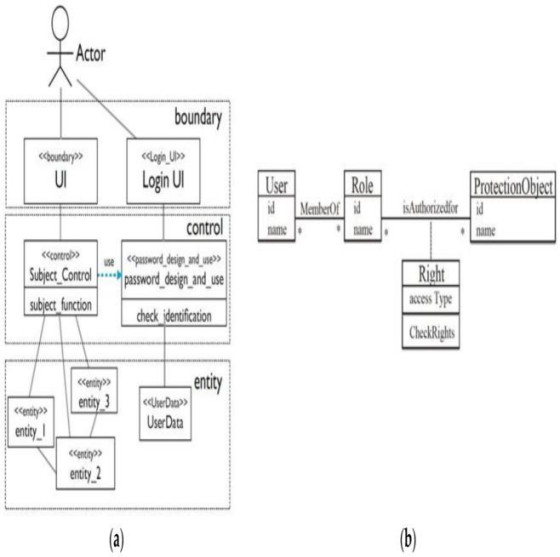


Figure : Flume tool work

B. Counter Based Algorithm

According to algorithms proposed by Yeonhee Lee and Youngseok Lee we are implementing this algorithms in hadoopframework[9]. Map Reduce algorithm to detect DDoS with URL counting. To lower the false positive rate, we adopted response rate against page requests. This algorithm needs three input parameters of time interval, threshold and unbalance ratio, which can be loaded through the configuration property or the distributed cache mechanism of Map Reduce. 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 unbalance 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. In Map Reduce algorithm, 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 Map Reduce job emits records whose response ratio against requests is greater than unbalance ratio, marking them as attackers. While this algorithm has the low computational complexity and could be easily converted to the Map Reduce implementation, it needs a prerequisite to know the threshold value from historical monitoring data in advance.We can use Java for implementation of counter based algorithm but it requires the thousand lines of code.So we can reduce this by using Pig Latin in which large java code is replaced by minimum lines of code. Pig provides a light-weight method for approximately measuring how much time is spent in different user-defined functions (UDFs) and Loaders. Approximate microseconds measures the approximate amount of time spent in a UDF and approximate invocations measures the approximate number of times the UDF was invoked



Characteristics of DDoS Attack

Following are the different ways to characterize the distributed denial of service attack:

1) Disruptive/Degrade Impact

After being a part of attack, the victim either to stop providing services to the client or the services are de- graded that means some of the services are still being provided to the client even the victim's system is under the attack.

2) Exploiting Vulnerability

Network of machines which follows the instructions of master attacker to send request for a service on a victim's machine to consume its all the resources.

3) Dynamic Attack Rate

Sometime attacker make down the websites very quickly by sending large no of request more than its capacity, is known as constant attack rate. While some- times attacker takes time to make it down by sending packets in variable length of request that is not constant, known as variable attack rate.

4) Automated Tools:

Attackers can be classified by automated tools also and their skills. Attack can be performed manually; semi automated or fully automated tools

IV. CONCLUSION

The paper includes latest techniques such as Hadoop along with other available techniques for detection of distributed denial of service attacks so that a comprehensive solution can be developed with several detection layers to trap the intrusion keeping in mind the limitations of these detection techniques. Counter based, Access pattern and Data generation algorithm by using Map Reduce in Hadoop.

REFERENCES

[1] ShwetaTripathi, Brij Gupta, AmmarAlmomani, Anupama Mishra, Suresh VeluruHadoop Based Defense Solution to Handle Distributed Denial of Service (DDoS) Attacks Journal of Information Security (JIS), 2013 DOI: 10.4236/jis.2013.43018.

[2] Priya P. Sharma, Chandrakant P. Navdeti Securing Big Data Hadoop: A Review of Security Issues, Threats and Solution - International Journal of Computer Science and Information Technologies (IJCSIT), Issue 2, Volume 5, 2014, ISSN: 0975-9646.

[3] B. M. Leiner, V. G. Cerf, D. D. Clark, R. E. Kahn, L. Kleinrock, D. C. Lynch, J. Postel, L. G. Roberts and S. Wolff, "A Brief History of the Internet," 2000. http://www.isoc.org/internet/history/brief.shtml

[4] B. B. Gupta, R. C. Joshi and M. Misra, "Defending against Distributed Denial of Service Attacks: Issues and Challenges," Information Security Journal: A Global Perspective, Vol. 18, No. 5, 2009, pp. 224-247.

[5] C. Douligeris and A. Mitrokotsa "DDoS Attacks and Defense Mechanisms: Classification and State of the Art," Elsevier Science Direct Computer Networks, Vol 44, No. 5, 2004, pp. 643-666. doi:10.1016/j.comnet.2003.10.003

[6] S. M. Specht and R. B. Lee, "Distributed Denial of Service: Taxonomies of Attacks, Tools, and Countermeasures," Proceedings of the International Workshop on Security in Parallel and Distributed Systems, San Francisco, 15-17 September 2004, pp. 543-550.

[7] A. Mishra, B. B. Gupta and R. C. Joshi, "A Comparative Study of Distributed Denial of Service Attacks, Intrusion Tolerance and Mitigation Techniques," European Intelligence and Security Informatics Conference, EISIC 2011, 12-14 September 2011, pp. 286, 289.

[8] T. Kitten, "DDoS: Lessons from Phase 2 Attacks," 2013 http://www.bankinfosecurity.com/ddos-attacks-lessons-from-phase-2-a-5420/op-1.

[9] Yeonhee Lee and Youngseok Lee "Detecting ddos attacks with hadoop", ACM Student Workshop, December 6 2011, Tokyo, Japan.

[10] SavneetKaurAhuja Review On MapReduce to Mitigate DDoS Attack- International Journal of Application or Innovation in Engineering & Management (IJAIEM) Volume 3, Issue 11, November 2014 – ISSN 2319 4847.