REVIEW ON MAPREDUCE TO MITIGATE DDOS ATTACK

SAVNEET KAUR AHUJA

Savneet kaur ahuja[1],mtech student, electronics and communication Jmit radaur

ABSTRACT

Denial-of-service (DOS) is one of the most unbounded problems for the internet. As the traffic increases day by day on internet, Flooding problem takes place which results DDOS attack. To handle this attack, mapreduce, a popular tool of big data is used which is based on the principle of parallelism and hadoop which is the open source implementation for mapreduce. The purpose of this paper is to study the DDOS attack using mapreduce by hadoop.

Keywords:- Mapreduce, Ddos attack, Hadoop, fair scheduler

1. INTRODUCTION

Mapreduce is the famous framework for cloud computing. It follows the parallel principle in the distributed computing environment. The most important feature of mapreduce is to allow user to process a large amount of data in a short time. There are various fields where mapreduce is used for example bioinformation, scientific analysis, machine learning, internet data analysis and security. Mapreduce can be implemented by hadoop which is an open source framework. Today's Internet is the best way to communicate but because of so many traffic, attack increases. One of the attacks in the internet is DDOS attack. DDOS attack is one attack which makes network unavailable and target on the client. A solution of DDOS attack is mapreduce using hadoop. Fairer scheduler is used to detect the DDOS attack

2. HADOOP

Firstly, Google has invented the Google Mapreduce so that they could collect the data from the users. Then further it has invented hadoop and hadoop Mapreduce which is an open source project and has similar capabilities to Google Mapreduce. In hadoop, Thousands of cluster nodes are used.

2.1 History

It has been created by Dong Cutting and Mike Cafarella in 2005. It has developed to support distribution for search engine project. . It is licensed under APACHE LICENSE 2.0. This is written in Java Runtime Environment (JRE) 1.6 or higher version. The operating system is cross-platform. Its developer is APACHE SOFTWARE FOUNDATION.

2.2 USED BY

Hadoop is used by Amazon, Facebook, Google, New York Times, Veoh, Yahoo etc

2.3 HADOOP CLUSTER ARCHITECTURE

It includes single cluster and multiple slave nodes or worker nodes. A single cluster or master node consists of Job Tracker, Task Tracker, Namenode and Datanode and slave node consists of Datanode and Task tracker.

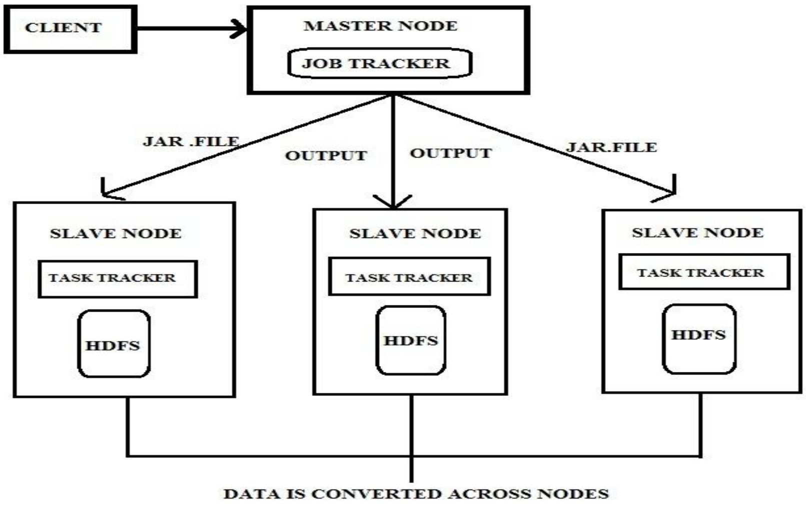

Fig. 1 Hadoop Cluster Architecture

2.4 Elementary Principle

1. HDFS
2. MAPREDUCE

2.4.1 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.

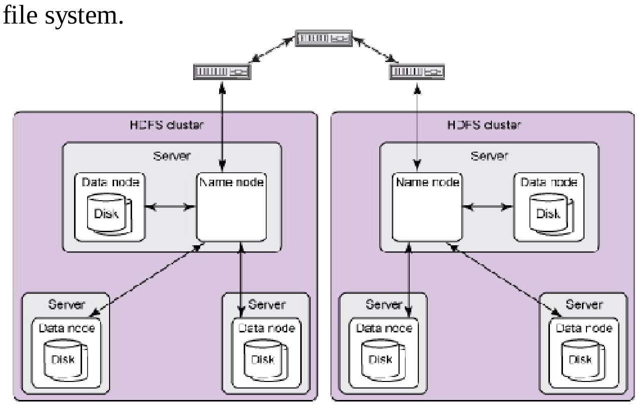

Fig. 2 Hadoop Distributed File System

2.4.2 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 mapreduce is hadoop. Mapreduce is used to mitigate the ddos attack

2.5 FAIR SCHEDULER

Fair scheduler has been developed by facebook.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.

3 DISRIBUTED DENIAL-OF-SERVICE

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.

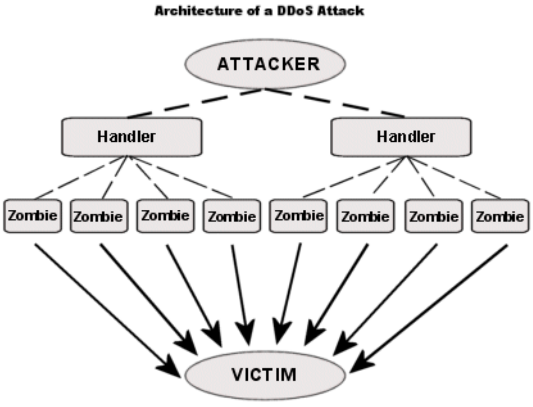

Fig. 3 Architecture of a Ddos Attack

3.1 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.

3.2 DDOS GOALS

Goal of DDOS is to damage the client for personal reasons, damage opposition or for popularity.

3.3 Impacts

- DDOS attack impacts on the security of the internet as it is independent so with the help of coding, any machine can easily become victim.
- No of resources handle by the internet is limited so flooding (DDOS attack) is also having great impact on internet.

4 REVIEW ON THE FOLLOWING PAPERS

YEOHEE LEE AND YOUNGSEOK LEE This paper proposes the counter based detection method which counts the number of web pages request. In this paper, more effective factor is frequency of page requests which will come from clients. It simply counts the URL and through this, it detects the DDOS attack. This is the mapreduce algorithm. It follows the response rate against page request and traffic volume which is given by H.LIU, Y SUN and M.KIM. This paper includes three parameters which is time interval, threshold and unbalance ratio .In this, function generates key value of server and client. Then, Reduce function reduces the number of URL requests, page request and server responses. If value of requests exceeds the threshold then response ratio is greater than unbalance ratio and it acts as attacker. It also uses the access pattern based method. It differentiates the attackers from normal client. It has two mapreduce jobs. First it gives sequence number to the jobs between clients and server and calculates the spending time and counts the byte. Second job is to knock out the infected hosts by comparing the sequence number and spending time. This paper also discusses about the drawbacks of this method as it is highly complex for detecting DDOS attack.

G.S.NAVALE, VIVEK KASBEKAR, VIJAY GANJEPATIL, SHRAVANTI BUGADE This paper describes the detection of DDOS by using two algorithms. Firstly, it describes the data generation in which flume tool is used for log file generation which collects the log data from application server. Then it implements the algorithm which is proposed by Y.LEE and Y.LEE for reducing the false proportion rate, the algorithm is counter based algorithm. It includes the parameters which are time interval, threshold and unbalanced ratio and then it uses pig Latin which reduces the thousand lines of codes which is written in java and gives a light weight method for reducing the code and reduces the time which is spent on it. Then this paper uses the counter based algorithm as the input of Access pattern algorithm. It differentiates the attackers from normal client. This paper also uses various algorithms in 'R' such as regression algorithm, aggregation algorithm for detecting the attackers. Dashboard is also used for better user interface. This is the extended version of Y.LEE and Y.LEE paper.

RANA KHATTAK, SHEHAR BANO, SHUJAAT HUSSAIN, ZAHID ANWAR This paper tells about the DDOS lifecycle. In this paper, binary format dataset is provided is converted into ext format. Then it is forwarded to HDFS for hadoop cluster. It is provided by Lincon lab. An experiment of mapreduce cluster is carried out which is containing a large set of attack data. A connection table is populated by analyzing the packets one by one. I, If IP address crosses the 'horizontal threshold' or 'vertical threshold', then it label as attacker and forwards to the next level. After phase 1 and phase 2 it passes to phase 3. If a lot of IP comes from single internet host then such packets are malicious. For implementation, java language is used and karmaphere is used for hadoop. Traffic analyzing technique is used. Results of this paper are following:- firstly it compares the performance of java program. Secondly, the experimental results are size of data is 911 MB and 3 malicious hosts are present. 6.98 GB data which does not contain any attackers.

DONGJIN YOO, KWANG MONG SIM In this paper, comparison of different scheduling methods takes place. It discusses the two categories which are asynchronous processing and speculative execution. Discussion of delay scheduling in hadoop and Quincy scheduler in dryad takes place for fairness constraints. There are following implementing systems are hadoop, dryad and sphere. Dryad is different from hadoop. Hadoop is mapreduce model and dryad is graphically model like DAG (direct acyclic graph). There are scheduling issues which are discussed in this paper is locality , synchronization, fairness criterion and the scheduling methods are synchronous overhead which contains asynchronous processing and speculative execution and methods for fairness with locality improvement contains delay scheduling with fair scheduler and quincy scheduler.

Description of following methods are :-

Asynchronous processing – it allows asynchronous processing which contains two level local or global
map and reduce. Speculative execution –it monitors whether a task is stragglers.

JELENA MIRKOVIC, JANICE MARTIN AND PETER REIHER This paper tells us about the characteristics of DDOS Defense mechanisms. This paper discusses about the DDOS attack strategy, goals and classification on the basis of degree of automation, exploited vulnerability, attack rate dynamics and its impact. It also tells about the DDOS defense mechanism. This is classified by activity level and by development location. This paper concludes that the DDOS is a serious attack and this paper concludes that the DDOS is a serious attack and this paper achieves the clear idea of problem and its solution.

5 CONCLUSION

The techniques which have discussed in this paper are
i. Counter based algorithm
ii. Access pattern algorithm
iii. Extended algorithm of counter based algorithm
iv. Traffic analyzing technique
Comparison of different scheduling methods has been discussed. DDOS attacks have been studied in this paper.

REFERENCES

[1] Hadoop Distributed File System. http://hadoop.apache.org/ common/docs/current/hdfs design.html.
[2] Fair Scheduler. http://hadoop.apache.org/common/docs/r0.20. 2/fair scheduler.html.
[3] Capacity Scheduler. http://hadoop.apache.org/common/docs/ r0.20.2/capacity scheduler.html.
[4] Mirkivic and P. Reiher, A Taxonomy of DDoS Attack and DDoS Defense Mechanisms, ACM SIGCOMM CCR, 2004
[5] Yeonhee Lee and Youngseok Lee "Detecting ddos attacks with hadoop", ACM Student Workshop,December 6 2011, Tokyo, Japan
[6] G.S. Navale,Vivek Kasbekar,Vijay Ganjepatil,Shravanti Bugade, Detecting and analyzing ddos attack using mapreduce in hadoop, International Journal of Industrial Electronics and Electrical Engineering, ISSN: 2347-6982, Volume- 2, Issue- 2, Feb.-2014
[7] http://en.wikipedia.org/wiki/Apache_Hadoop
[8] Rana Khattak, Shehar Bano, Shujaat Hussain, Zahid Anwar, DOFUR: DDoS Forensics Using mapreduce, 2011 frontiers of information technology
[9] Dongjin yoo, kwang mong sim, a comparative review of job scheduling for mapreduce. Proceedings of ieee ccis2011

AUTHOR

Savneet kaur ahuja Persuing M.Tech (ECE), JMIT, Radaur, Haryana, India . Her research interest areas are Big data, wireless network, Security.