Detection and Avoidance of Live DDoS Attack on Cloud Environment

Kiran Sunil Salve[1] Rahul Desai[2] Asif Yunus Sayyad[3] Maruti Suresh Surwase[4] Akash M. Kunbithop[5]
[1,2,3,4,5]JSPMs BSIOTR wagholi, Pune, India

Abstract— Many systems use servers to manage and store their data, sometimes the servers are slowed down because of multiple user requests. Most of which are attackers or unauthorized users and some are genuine users. In computing, a denial-of-service (DoS) attack is an attempt to make a machine or network resource unavailable to its intended users. A distributed denial-of-service (DDoS) is where the attack source is more than one, often thousands of unique IP addresses. Flooding is one of the typical DDoS attacks that exploit normal TCP connections between a client and a target web server. In this project we are trying to devise a DDoS anomaly detection method on Hadoop that implements a Map Reduce-based detection algorithm against the Flooding attacks.

Key words: Attack, Hadoop, Security, DDos, Mapreduce

I. INTRODUCTION

In computing, a denial-of-service (DoS) attack is an attempt to make a machine or network resource unavailable to its intended users, such as to temporarily or indefinitely interrupt or suspend services of a host connected to the Internet. Denial of service is typically accomplished by flooding the targeted machine or resource with superfluous requests in an attempt to overload systems and prevent some or all legitimate requests from being fulfilled.[1]

A distributed denial-of-service (DDoS) is where the attack source is more than one, often thousands of, unique IP addresses. It is analogous to a group of people crowding the entry door or gate to a shop or business, and not letting legitimate parties enter into the shop or business, disrupting normal operations. The scale of DDoS attacks has continued to rise over recent years, even reaching over 400Gbit/s.[2]

Criminal perpetrators of DoS and DDoS attacks often target sites or services hosted on high-profile web servers such as banks, credit card payment gateways. Motives of revenge, blackmail or activism can be behind other attacks.

1) The United States Computer Emergency Readiness Team (US-CERT) defines symptoms of denial-of-service attacks to include:

- Unusually slow network performance (opening files or accessing web sites)
- Unavailability of a particular web site
- Inability to access any web site

Dramatic increase in the number of spam emails received (this type of DoS attack is considered an e-mail bomb)

2) Addition symptoms may include:

- Disconnection of a wireless or wired internet connection
- Long-term denial of access to the web or any internet services

If the attack is conducted on a sufficiently large scale, entire geographical regions of Internet connectivity can be compromised without the attacker's knowledge or intent by incorrectly configured or flimsy network infrastructure equipment HTTP GET flooding is one of the typical DDoS attacks that exploit normal TCP connections between a client and a target web server. As the volume of Internet traffic increases explosively year after year, the Intrusion Detection Systems (IDSes) have faced the issue on how to assure both scalability and accuracy of analyzing the DDoS attack from these huge volume of data.

Identify applicable sponsor/s here. If no sponsors, delete this text box (sponsors*).*

Hadoop is an open-source distributed cluster platform that includes a distributed file system, HDFS and the programming model, MapReduce. In this project we are trying to devise a DDoS anomaly detection method on Hadoop that implements a MapReduce-based detection algorithm against the HTTP GET flooding attack.

Information security has been a field of increasing importance in the information society. DDoS attack has been a threat to web services since long time and has become more serious threat recently with the advancements in internet technology and tools. DoS attacks came into popularity in the year 2000 when websites such as Yahoo, Amazon, and CNN were crippled using these attacks. The first security attack faced by the Internet is a worm occurrence in 1988 (Ren Rochlis and Eichin 1989). The curve showing dramatic shoot up in cyber terrorism in the recent past due to broadband explosion, technology globalization and e-commerce growth can be studied from various security news sites. This opens up more security holes, and the heavy damage and loss incurred was periodically reported (Garber 2000).

Methods to address this security problem have been initialised based on monitoring, measuring and modeling traffic characteristics that vary in flow rate, flow volume and flow behavior. Designing an effective defense mechanism require knowledge on related issues like the type of attack, the 24 mode of launching the attacks, and impact made by the attacks on the network and the target systems. Different modes through which DoS attacks are commonly launched was studied by Mirkovic and Reiher (2004). Attacks could be launched manually or in automated way. With the sophisticated set of attack tools springing up, the launching becomes easier and rapid. The attacks are direct attacks when launched from a single source or indirect when launched via some agents or reflector nodes that multiply the attack received and uses different paths to instigate the attack and to hide the identity of the source. Distributed denial of service attack is perpetrated by a collection of nodes called zombies or bots that are compromised by a master node so that the magnitude of the impact is larger. Attackers normally use BotNet using internet relay channel to carry out DDoS attacks because of which the identity of the true attacker becomes harder to trace.

It is not always possible. DDoS attack by its nature pose several challenges that every defense system should face upon. The primary demands are summarized as follows.

1) DoS attacks can be launched using packets resembling legitimate traffic at a higher rate and hence distinct characterization of attack by per packet analysis
2) Attackers may be intelligent enough to conceal themselves from being detected by throwing low volume or low rate traffic from multiple sources and handling such sophisticated attacks is again a challenge.
3) Defense systems need to compromise between protecting the network resources earlier from being depleted on one hand, 28 while handling enormous traffic volume for accurate attack identification on the other hand. Hence deciding upon where to deploy the defense system in the network is a challenging task.

II. LITERATURE SURVEY

Information security has been a field of increasing importance in the information society. DDoS attack has been a threat to web services since long time and has become more serious threat recently with the advancements in internet technology and tools. Dos attacks came into popularity in the year 2000 when websites such as yahoo, amazon, and cnn were crippled using these attacks. The first security attack faced by the Internet is a worm occurrence in 1988 (Ren Rochlis and Eichin 1989). The curve showing dramatic shoot up in cyber terrorism in the recent past due to broadband explosion, technology globalization and e-commerce growth can be studied from various security news sites. This opens up more security holes, and the heavy damage and loss incurred was periodically reported (Garber 2000). Methods to address this security problem have been initialized based on monitoring, measuring and modeling traffic characteristics that vary in flow rate, flow volume and flow behavior. Designing an effective defense mechanism require knowledge on related issues like the type of attack, the 24 mode of launching the attacks, and impact made by the attacks on the network and the target systems. Different modes through which DoS attacks are commonly launched was studied by Mirkovic and Reiher (2004). Attacks could be launched manually or in automated way. With the sophisticated set of attack tools springing up, the launching becomes easier and rapid. The attacks are direct attacks when launched from a single source or indirect when launched via some agents or reflector nodes that multiply the attack received and uses different paths to instigate the attack and to hide the identity of the source. Distributed denial of service attack is perpetrated by a collection of nodes called zombies or bots that are compromised by a master node so that the magnitude of the impact is larger. Attackers normally use BotNet using internet relay channel to carry out DDoS attacks because of which the identity of the true attacker becomes harder to trace. DDoS attack by its nature pose several challenges that every defense system should face upon. The primary demands are summarized as follows.

a) DDoS attacks can be launched using packets resembling legitimate traffic at a higher rate and hence distinct characterization of attack by per packet analysis is not always possible.
b) Attackers may be intelligent enough to conceal themselves from being detected by throwing low volume or low rate traffic from multiple sources and handling such sophisticated attacks is again a challenge.
c) Defense systems need to compromise between protecting the network resources earlier from being depleted on one hand, 28 while handling enormous traffic volume for accurate attack identification on the other hand. Hence deciding upon where to deploy the defense system in the network is a challenging task.

III. PROPOSED SYSTEM

The DDoS attack on a Hadoop environment can be a major setback in the performance of the Hadoop cluster, the challenge is not only to prevent the DDoS attack, but also distinctively identify an attack for multiple requests from a genuine user.

Our project aims at identification of such DDoS attack on the Hadoop environment and avoiding the same in real time in order to ensure the optimal performance from the Hadoop clusters.

The outcome of the project will be in real time in order to ensure the uninterrupted functioning of the Hadoop cluster, which will provide better results in its operation.

IV. EXPECTED RESULTS

Proposal of an efficient way to detect the live DDoS attack and prevent it, the DDoS attack on the Hadoop cluster gets detected and avoided using an optimally suited method.

V. CONCLUSION

The DDoS attacks are a high risk factors in hadoop, particularly flooding attack, which is one of the easiest to implement but one of the most effective type of attack as well. The Report reviews the implementation of an efficient algorithm which will not only detect the live DDoS attack on Hadoop cluster, but also avoid it, thus ensuring the smooth functioning of the system, as well as optimal performance without performance overhead.

REFERENCES

[1] Hadoop. https://hadoop.apache.org/.
[2] Hadoop yarn. http://hortonworks.com/hadoop/yarn/.
[3] Mapreduce. http://wiki.apache.org/hadoop/MapReduce.
[4] Mausezahn. http://www.perihel.at/sec/mz/.
[5] Secure copy. linux.die.net/man/1/scp.
[6] Tshark: Network analyzer. www.wireshark.org/docs/manpages/ tshark.html.
[7] Paul J Criscuolo. Distributed denial of service: Trin00, tribe flood network, tribe flood network 2000, and stacheldraht ciac-2319. Technical report, DTIC Document, 2000.
[8] Yeonhee Lee, Wonchul Kang, and Youngseok Lee. A hadoopbased packet trace processing tool. In Jordi Domingo-Pascual, Yuval Shavitt, and Steve Uhlig, editors, Traffic Monitoring and Analysis, volume 6613 of Lecture Notes in Computer Science, pages 5163. Springer Berlin Heidelberg, 2011.
[9] Yeonhee Lee and Youngseok Lee. Detecting ddos attacks with hadoop. In Proceedings of The ACM CoNEXT Student Workshop, page 7. ACM, 2011.

[10] Vern Paxson. Bro: A system for detecting network intruders in realtime. Comput. Netw. 31(23-24):24352463, December 1999.
[11] Martin Roesch. Snort - lightweight intrusion detection for networks. In Proceedings of the 13th USENIX Conference on System Administration, LISA 99, pages 229238, Berkeley, CA, USA, 1999. USENIX Association. 37 References
[12] Saman Taghavi Zargar, James Joshi, and David Tipper. A survey of defense mechanisms against distributed denial of service (ddos) flooding attacks. Communications Surveys Tutorials, IEEE, 15(4):20462069, 2013.

