Literature Review

Sri Rachana Achyuthuni (sa9547) Akhil Karrothu (ak8367)

ABSTRACT

For our term project, we are planning to implement Round Robin Approach for Load Balancing in Cloud Computing. We would also be comparing this Load Balancing approach with two other known Load Balancing Approaches - AWS Classic Load Balancer [1] and HAProxy Load Balancer [3]. In this paper, we have examined Load Balancing, Round Robin method, AWS Classic Load Balancing method and HAProxy. With this knowledge, we have gained a better insight on how to proceed with the project.

1. INTRODUCTION

For any web application on the internet, it is important to ensure that each request made to the application is responded to in an appropriate manner. Hence, it is important to ensure that each server hosting the web application is not over used.

Load Balancing is the technique introduced commonly for Cloud Computing to ensure that the load on each server in a cluster is distributed equally. Load Balancing ensures high performance and throughput. In cloud computing, it facilitates distribution amongst servers placed anywhere in the world, thus, decreasing the response time. In case of failures, further requests are transmitted to other alive servers in the cluster maintaining reliability of the application.

Some challenges faced during Load Balancing are:

- **Single Point Failure**: Since the Load Balancer is the first point to which requests are sent to, if there's a failure in the Load Balancer then, the entire application might become unresponsive.
- **Multiple Nodes of different types**: Sometimes the Load Balancer might have to forward requests to servers which are of different configurations. Hence, it has to be designed in such a way that it can transmit requests to servers of any configuration.
- **Multiple Load Balancers**: Having multiple Load Balancers would ensure more reliability. However, implementign multiple Load Balancers is a major challenge.

There have been many surveys conducted to understand the need and the challenges of Load Balancing in Cloud Computing. We have referred to the survey paper by Kumar et. al. [6] which discusses Load Balancing and it's challenges and the research paper by Alakeel [4] that throws light on Load Balancing in Distributed Systems. In section 2, we will discuss some of the algorithms commonly used for Load Balancing today. We will also look into some existing Load Balancers in section 3.

2. METHODS OF LOAD BALANCING

There are several policies or algorithms that can be used to implement Load Balancing in Cloud Computing.

2.1 Round Robin Load Balancing

In this method, the requests are forwarded to each server in the cluster in a sequential manner. This method ensures that each link is utilized equally. It also ensures that each server is utilized. However, it does not stop to see whether any of the other servers have lesser load. It has been observed that using this algorithm, there is a higher chance of packets being dropped. [5] This is the algorithm that we would be implementing for our project.

2.2 Number of Connections Load Balancing

In this method, only the number of connections that each server has with the Load Balancer is taken into consideration. The Load Balancer picks the server with the least connections to transmit the request to. This method is highly efficient when the given cluster of servers are all identical in terms of the CPU and Memory configurations. In that case, it ensures fast and reliable performance. [7]

2.3 Weighted Fair Queuing Load Balancing

This method, in simpler terms, is a weighted Round Robin method. Here, each link is assigned a particular weight based on several factors such as the capacity of the server, the number of requests it is currently processing, etc. The Load Balancer picks the next server with the least weight ensuring that a fair Load Balancing. This algorithm has proven to decrease delays and ensure high performance. It is particularly beneficial in cases where the servers present in the cluster all vary in terms of CPU and Memory configurations [7]

3. RELATED WORK

In this section, we will be looking closely at two popular existing Load Balancers to understand and compare them with the Load Balancer that we would be implementing in our project.

3.1 AWS Classic Loadbalancer

This is one of the types of Load Balancing service provided by AWS, one of the leading cloud providers. A *listener* is

configured for a particular port. The connection requests for the protocol and port configured for the listener are monitored and then, forwarded to whichever port specified. The Classic Load Balancer can also be scaled. The instances behind the Load Balancer can also be dynamically scaled based on the load. It mainly works on Layer 7. [2]

3.2 HAProxy

It is a fast and reliable load balancer available for free. It can be configured at the network level, the server level or the link level. It can work on both Layer 7 and Layer 4. It can be used for content-switching, TCP normalizing, HTTP Regulation, etc. [3]

4. CONCLUSION & FUTURE WORK

Load Balancing is highly important and necessary in Cloud Computing. However, there are some challenges that need to be worked upon when implementing Load Balancers. There are several methods that can be used for implementing the Load Balancing algorithm such as Round Robin, Number of Connections and Weighted Fair Queuing. AWS Classic Load Balancer and HAProxy Load Balancer are two highly in demand and popular Load Balancers. AWS Classic Load Balancer is priced while HAProxy is free. Classic Load Balancer works on Layer 7 which HAProxy works on both Layer 4 and Layer 7.

The next steps in our project is to implement the Round Robin Algorithm in one of the AWS EC2 Servers and compare it with AWS Classic Load Balancer and HAProxy Load Balancer.

5. REFERENCES

[1] Aws classic load balancer. `https://aws.amazon.com/elasticloadbalancing/`. Accessed: 2020-09-25.

[2] Aws classic load balancer. `https://docs.aws.amazon.com/elasticloadbalancing/latest/classic/introduction.html`. Accessed: 2020-09-25.

[3] Haproxy. `http://www.haproxy.org`. Accessed: 2020-09-25.

[4] A. Alakeel. A guide to dynamic load balancing in distributed computer systems. *International Journal of Computer Science and Network Security (IJCSNS)*, 10, 11 2009.

[5] S. Kaur, K. Kumar, J. Singh, and Navtej Singh Ghumman. Round-robin based load balancing in software defined networking. In *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, pages 2136–2139, 2015.

[6] P. Kumar and R. Kumar. Issues and challenges of load balancing techniques in cloud computing: A survey. *ACM Comput. Surv.*, 51(6), Feb. 2019.

[7] J. P. Ponciano and N. Anani. Load balancing in modern network infrastructures — a simulation model. In *2014 9th International Symposium on Communication Systems, Networks Digital Sign (CSNDSP)*, pages 841–846, 2014.