

# A Comprehensive Analysis Of Load Balancer In Virtualization

*Miss. Chandni B. Patel*  
*Parul Institute Of Engineering And*  
*Technology, Vadodara, Gujarat, India*

*Prof. Jwalant Baria*  
*Parul Institute Of Engineering And*  
*Technology, Vadodara, Gujarat, India.*

## Abstract

*Load balancer have a decisive role in every enterprise network as they serve as an entry point and have major impact on the performance and the availability of the network. We develop a load balancer service to handle the load of server without the necessity for a specific piece of hardware*

## 1. Introduction

Now a day's, everything is going online. As a result of the popularity of the web, providers of web sites and storage space providers want to ensure the availability of access to information for their users and the guarantee that requests are processed as quickly as possible. If a server gets more requests than it can handle, this can be attack by using multiple hosts to provide the same service. A web farm of multiple hosts grouped together will share the total load of the client requests, thus reduce the response time and increase the QoS, ultimately resulting in satisfied customers. Another important issue amongst service providers is their degree of uptime of the servers. This is also refereed as server availability.[1]

Virtualization is commonly defined as a technology that introduces a software abstraction layer between the hardware and the operating system and applications running on top of it. Its main advantages include isolation, consolidation and multiplexing of resources. Other benefits of virtualization include saving on power by consolidation of different virtual machines on a single physical machine, migration of virtual machine for load balancing etc. Virtualization provides full control of resource allocation to

administrator, resulting in optimum use of resources. More recently, another advantage of virtualization - live migration of virtual machine is increasingly used to better handle workload balancing across physical machines in data center, especially when the available resources in physical machine are not sufficient for VMs.[2]

Load balancer is usually a software program that is listening on the port where external clients connect to access services. The load balancer forwards requests to one of the "backend" servers, which usually replies to the load balancer which may have security benefits by hiding the structure of the internal network and preventing attacks on the kernel's network stack or unrelated services running on other ports. If you are load balancing across several servers and one of the servers fails, your service will still be available to your users, as the traffic will be diverted to the other servers in your server farm.

## 2. Advantages of load balancer

Scalability  
Efficient use of equipment  
Easy integration  
Low overhead  
High availability  
Increased performance

## 3. Applications of load balancer

E-commerce  
Healthcare website  
Education (E-learning)  
VoIP  
Etc

## 4. Load balancing technologies

### 4.1 NAT-based Load Balancing

Network Address Translation is an important function of the Internet gateway, it can convert the internal address and external addresses in order to make the computer with internal address have access to the external network. Similarly, when the external network visits an external address of a gateway, the address-translating gateway can forward it to the address mapping the internal address of the computer. Thus the request is distributed to the various servers to achieve load balancing. This approach can fulfil

the task of load balancing, but work of central load balancing equipment is very heavy, incoming service request and the response packets sent by the server have to undergo treatment, and thus the device is likely to become a bottleneck in the system.[4]

### 4.2 Load balancing through direct routing

For this technology, the central load balancer dynamically selects a server according to the load status of each server, transforms the MAC address of the data frame to the MAC address of the server selected, and then sends this data frame to the server via the LAN. This method requires that the central load-balancing device and server group must have a physical network card connected through the LAN. VIP address (the user can see the IP address) is shared by the central device and the server group. All the VIP addresses of the server are configured in the respective Non-ARP network equipment. The server processes message and then it directly back to the client.[4]

### 4.3 Load-balancing based on the IP tunnelling technology

When one server of the cluster found itself heavily loaded, it can forward the request to other server to process by the IP tunnelling technology. This requires that all the servers support IP tunnelling technology, but also brings additional packing / unpacking. When the whole cluster system is heavily loaded, packet loop may occur. Therefore, its performance and reliability are not high.[4]

## 5. Load balancing Algorithms

Server load balancing is a useful and indispensable technique for providing better performance for many applications. In this section, we recall well-known server load balancing mechanisms.

### 5.1 Round-Robin Algorithm

The round-robin scheduling algorithm sends each incoming request to the next server in it's list. Thus in a three server cluster (servers A, B and C) request 1 would go to server A, request 2 would go to server B, request 3 would go to server C, and request 4 would go to server A, thus completing the cycling or 'round-robin' of servers. It treats all real servers as equals regardless of the number of incoming connections or response time each server is experiencing. Virtual Server provides a few advantages over traditional round-robin DNS. Round-robin DNS resolves a single domain to the different IP addresses, the scheduling granularity is host-based, and the caching of DNS queries hinders the basic algorithm, these factors lead to significant dynamic load imbalances among the real servers. The scheduling granularity of Virtual Server is network connection-based, and it is much superior to round-robin DNS due to the fine scheduling granularity.[5]

### 5.2 Weighted round-robin Algorithm

The weighted round-robin scheduling is designed to better handle servers with different processing capacities. Each server can be assigned a weight, an integer value that indicates the processing capacity. Servers with higher weights receive new connections first than those with less weights, and servers with higher weights get more connections than those with less weights and servers with equal weights get equal connections. In the implementation of the weighted round-robin scheduling, a scheduling sequence will be generated according to the server weights after the rules of Virtual Server are modified. The network connections are directed to the different real servers based on the scheduling sequence in a round-robin manner. The weighted round-robin scheduling is better than the round-robin scheduling, when the processing capacity of real servers are different. However, it may lead to dynamic load imbalance among the real servers if the load of the requests vary highly.[5]

### 5.3 Least-connection Algorithm

The least-connection scheduling algorithm directs network connections to the server with the least number of established connections. This is one of the dynamic scheduling algorithms; because it needs to count live connections for each server dynamically. For a Virtual Server that is managing a collection of servers with similar performance, least-connection scheduling is good to smooth distribution when the load of requests vary a lot. Virtual Server will direct requests to the real server with the fewest active connections[5]

**Table 1: Comparison of scheduling algorithms**

	Round-robin Algorithm	Weighted round-robin Algorithm	Least-connection algorithm
<b>Definition</b>	Simplest scheduling algorithm directs the network connections to the different servers in the cluster	The weighted round robin scheduling can treat the real servers of different processing capacities	Least-connection scheduling algorithm directs network connections with the least number of active connections
<b>Work flow</b>	The round-robin DNS resolves the single domain to the different IP addresses, the scheduling granularity is per host, and the caching of DNS hinder the	Servers with higher weights receives new connections first and get more connections than servers with lower weights, and servers with equal weights can get an equal distributions	Load balancer with least-connection schedule will distributes more requests to real server with fewer active connections by keeping track of live connections

	algorithm take effect. Round-robin scheduling is based on first-come first-served queuing	of new equations	to the real servers
<b>Scheduling</b>	Static	Dynamic	Dynamic
<b>Server capacity (each node)</b>	Each node should have equal processing capacity because it would lead to load imbalance	Vary processing capacity for each node because we could assign weight that indicates its processing capacity	Each node should have equal processing capacity because it would lead to load imbalance

## 6. Conclusion

load balancing of a server is a very important research area in network. In our proposed idea & concept we are looking to improvise the load balancer functionality for multiple protocol support with benchmark analysis of the load balancer software. To provide more scalability and reliability load balancer can be configured with clustering and failover support for real time environments. In addition to this load balancer testing and implementation on virtualized and private cloud.

## 7. References

- [1] Ankush P. Deshmukh , Kumarswamy Pamu , "Applying Load Balancing: A Dynamic Approach", *International Journal of Advanced Research in Computer Science and Software Engineering*, Volume 2, Issue 6, June 2012
- [2] Akshay Chandak, Akshay Kanfode, Amit Joshi, "Dynamic Load Balancing of Virtual Machines using QEMU-KVM" *International Journal of Computer Applications (0975 – 8887) Volume 46– No.6, May 2012*
- [3] Sameena Naaz, Afshar Alam, Ranjit Biswas, "Load Balancing Algorithms for Peer to Peer and Client

*Server Distributed Environments” International Journal of Computer Applications (0975 – 888)*

*Volume 47– No.8, June 2012*

[4] Yang Jiao, Wei Wang, “*Design and Implementation of Load Balancing of Distributed-system-based Web Server*” , IEEE 2010, Third International Symposium on Electronic Commerce and Security.

[5]

<http://www.linuxvirtualserver.org/docs/scheduling.html> [online].

[6] Mariam Lahami, Moez Krichen, Rochdi Abid, Mohamed Jmaiel “*Distributed Test Architecture for Load Balancing System*” Signals and Telecommunication Journal Volume 1 Number 1 March 2012

IJERT