Efficient Load Balancing for Bursty Demand in Web based Application Services via Domain Name Services

Mei Lu Chin[1]

Faculty of Computer Science
& Information Technology,
Universiti Malaysia Sarawak,
94300 Kota Samarahan, Malaysia.
meilu850422.chin@gmail.com[1]

Chong Eng Tan[2]

Faculty of Computer Science
& Information Technology,
Universiti Malaysia Sarawak,
94300 Kota Samarahan, Malaysia.
cetan@ieee.org[2]

Mohd Imran Bandan[3]

Faculty of Computer Science
& Information Technology,
Universiti Malaysia Sarawak,
94300 Kota Samarahan, Malaysia.
bnimran@fit.unimas.my[3]

Abstract-In this paper, we propose an enhancement to the conventional Round Robin DNS load balancing technique which allows a single domain name to be associated with several web server IP addresses in a rotated order in a server cluster. Once the DNS server resolves the domain name to one of the web server IP address, the subsequent requests from the same client will be sent to the same server regardless of the current condition of the server. However, this conventional load balancing technique has a few disadvantages; Firstly, the DNS server is unaware of the status of web servers whether their services are available or not. Secondly, making changes to the DNS records can take days to propagate across the local DNS caches because the conventional DNS server will not delete their cached DNS records until the time to live (TTL) expires. Lastly, DNS server assumes all web application servers to have equal capability to offer their services without taking into account of the server's load and performance level. In this paper, a new approach is presented to enhance the DNS load balancing services to be more intelligent in load distribution. The DNS server is taking consideration of the status of the servers and will be distributing the services requests based on the performance matrix of the servers in the cluster.

Keywords-load balancing; Round Robin; DNS

I. INTRODUCTION

As the increasing demand for speed and processing power grows, load balancing is developed to solve limitations in large data warehousing and transactions processing that are involved in businesses, education and government. Apart from that, web sites which are popular will also experience overwhelming workload and their service can be overloaded quickly in a short period of time.

Load balancing is a common technique to distribute the works or processes among two or more computers, hard drives, network traffic links, CPUs and other resources. The technique is to achieve the minimize response time, optimal resource utilization and maximizing the throughput. Besides, the reliability through redundancy can also be increased with the used of multiple components with load balancing instead of a single component.

Owing to that, researchers have been trying to utilize parallel computers architecture to reduce the difficulty in implementing load balancing and allowing the users to make full use of the computing power available. Therefore, load balancing technique has been introduced widely to automate equalizing of workloads of the processors through the execution of parallel programs. The main purpose of load balancing is to improve the performance of parallel computers. In addition, load balancing is also used for sharing computational workloads or functions through virtualization. In virtualization, the user views multiple machines as a single virtual machine.

In this paper, we propose an enhancement to the conventional Round Robin DNS Load Balancing technique. The conventional Round Robin DNS Load Balancing technique allows a single domain name to be associated with several server IP addresses in a rotated order. This technique redirects the client request to either one of the server in the server cluster. After the DNS server resolves the requested domain name to one of the server IP address, the subsequent requests from the same client will be sent to the same server. Unfortunately, this conventional technique does not taking into account the current server condition for the request distribution.

Firstly, the DNS server is unaware of the operation status of the web servers which means the DNS server does not know the servers in the cluster whether their services are available or not [1]. Secondly, DNS server will assume all of the web application servers to have equal capability to offer the services regardless the server are overloaded or underutilized.

Our proposal on an enhanced Round Robin DNS load balancing services offers more intelligence in load distribution by the DNS server. Under the new scheme, the DNS server is now aware of the operation status of the servers by creating a reporting mechanism for reporting server failure in the cluster to the DNS server. Moreover, the service requests are now to be distributed based on the performance matrix of the servers in the cluster. The performance matrix can be formed based on