WEB SERVER LOAD BALANCING

MUHAMAD SAIFUL AZWAN BIN ISMAIL

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

BORANG PENGESAHAN STATUS TESIS*

JUDUL: **WEB SERVER LOAD BALANCING**

SESI PENGAJIAN: **2010/2011**

Saya **MUHAMAD SAIFUL AZWAN BIN ISMAIL**
(HURUF BESAR)

mengaku membenarkan tesis (PSM/Sarjana/Doktor Falsafah) ini disimpan di Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dengan syarat-syarat kegunaan seperti berikut:

1. Tesis adalah hakmilik Universiti Teknikal Malaysia Melaka
2. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan untuk tujuan pengajian sahaja
3. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. ** Sila tandakan (/)

_____	SULIT	(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)
_____	TERHAD	(Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/ badan di mana penyelidikan dijalankan)
___/___	TIDAK TERHAD	

(TANDATANGAN PENULIS)

Alamat tetap: KM 14.5 Jln Berangan 6,
Umbai, 77300 Merlimau,
Melaka.

Tarikh : 8-7-2011

(TANDATANGAN PENYELIA)

Mr. Othman bin Mohd

Tarikh : 08-07-2011

CATATAN: * Tesis dimaksudkan sebagai Laporan Akhir Projek Sarjana Muda (PSM)
** Jika tesis ini SULIT atau TERHAD, sila lampirkan surat daripada pihak berkuasa.

WEB SERVER LOAD BALANCING

MUHAMAD SAIFUL AZWAN BIN ISMAIL

This report is submitted in partial fulfillment of the requirements for the

Bachelor of Computer Science (Networking)

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2011

DECLARATION

I hereby declare this project report entitled

WEB SERVER LOAD BALANCING

is written by me and is my own effort and that no part has been plagiarized without citations.

STUDENT : _____ Date: 8-7-2011

(MUHAMAD SAIFUL AZWAN BIN ISMAIL)

SUPERVISOR : _____ Date: 08-07-2011

(MR. OTHMAN BIN MOHD)

DEDICATION

This work is dedicated to my beloved family and siblings, who passed on a love of reading and respect for education.

To my supportive friends and my supervisor, thank you so much for assist and help.

ACKNOWLEDGEMENTS

First and for most, I would to take this chance to express my gratitude to UTEM for provided chance enrolling this course and compulsory taken by all students. Many skills and experience I had obtain that never be taught in normal lecture and laboratory session by involving hand-on discipline in project management. These skills and experience only can be get through practical experience by working on large project. All problems arise from this project and guide and maturing my decision and way of thinking. I am feeling very content and grateful to my PSM supervisor EN. OTHMAN BIN MOHD able to spare his time even he is a packed with schedule and sharing his wisdom and guidance when I needed most. He stand beside and whole way giving support and encouragement. In addition of boosting my morale, he was the one who gave brilliant idea and guided solutions in process of completing this project. I would to gave my highest gratitude and thanks to him as I ,for surely will remember all teachings and guide that he had gave.

Not forgetting my mother and father who gave me all support I need. Thank you for your guidance love and care. You shall never be forgotten. I also indebt to all those individuals involved in this project that gave critiques and comment to improve and push me to produce project that have best quality and satisfy and requirement need. Finally, I also would like to acknowledge the contributions to all my friend that gave me advise, moral support, useful reference notes and guidance during this project occur. Last but not least for those that I had not mention in here but have directly or indirectly helping and guiding me towards completing PSM . Your efforts and time are much appreciate.

ABSTRACT

Load balancing is especially important for networks where it is difficult to predict the number of requests that will be issued to a web server. In order to reply many requests, load balancer typically employs two or more web server in a load balancing scheme. The main problem is unbalance data because of the loaded website crawls when more requests avalanche in. If one server starts to get swamped, requests are forwarded to another server with more capacity. Load balancing can also refer to the communications channels themselves. The main purpose of load balancer is to balance the loaded websites website crawls when more request avalanche in. The concept of the load balancing project is when the load balancer goes down; automatically the hot-standby (backup) will take over. By using Piranha software provide the ability of checking a server pool states. When one of the servers of the server pool is down, it will be remove it from the network. In addition, Piranha implements are to handle director failover. For solution, implement the Load Balancer is a good way. Load Balancer is distributing processing and communications activity evenly across a computer network so that no single device is overwhelmed.

ABSTRAK

Adalah sukar untuk meramalkan bilangan permintaan ke atas laman sesawang merupakan faktor utama 'load balancing' sangat penting di dalam rangkaian komputer. Untuk membolehkan menjawab atau membalas permintaan yang banyak daripada pengguna 'load balancer' menggunakan dua atau lebih pelayan sesawang di dalam implementasikan 'load balancer' . Masalah utama yang wujud adalah, capaian ke laman sesawang menjadi perlahan apabila terlalu banyak permintaan oleh pengguna pada masa tertentu. Di dalam implementasikan 'load balancer', apabila sesuatu pelayan dibanjiri dengan permintaan pengguna yang banyak, permintaan tersebut akan dialihkan ke pelayan yang lain. Tujuan utama 'load balancer' adalah untuk seimbangkan capaian ke laman sesawang apabila terdapat permintaan daripada pengguna yang terlalu ramai pada masa tertentu. Konsep 'load balancing' adalah apabila 'load balancer' gagal berfungsi 'hot-standby' akan mengambil alih tugas 'load balancer' asal. Perisian Piranha mempunyai aplikasi yang dapat memeriksa keadaan kumpulan pelayan. Apabila salah satu pelayan gagal berfungsi, Piranha akan mengeluarkan pelayan daripada rangkaian. Bagi mengatasi masalah ini, implimentasikan 'load balancing' di dalam rangkaian merupakan penyelesaian terbaik. 'Load balancing' adalah proses pengagihan dan komunikasi di dalam rangkaian komputer supaya tidak ada satu peralatan pun akan terbeban.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATONS

TERMS	DESCRIPTION
CPU	Central Processing Unit
UTeM	Universiti Teknikal Malaysia Melaka
VRRP	Virtual Router Redundancy Protocol
VRRPv2	Virtual Router Redundancy Protocol version 2
URL	Uniform Resource Locator
No	Number
RP	Research Problem
RQ	Research Question
ICT	Information and Communications Technology
LVS	Linux Virtual Server
BSD	Berkeley Software Distribution
GUI	Graphical user interface
SSL	Secure Sockets Layer
PDIOO	Planning, Design, Implementation, Operation, and Optimization
OSI	Open Systems Interconnection
PERT	Program Evaluation and Review Technique
KTCPVS	Kernel TCP Virtual Server
IPVS	IP Virtual Server
UDP	User Datagram Protocol
PC	Personel Computer
NIC	Network Interface Card
lb1	Load balancer 1
lb2	Load Balancer 2
rs1	Real Server 1

rs2	Real Server 2
NAT	Network Address Translation
rr	Round Robin
DR	Direct Routing
RHEL	Red Had Enterprise Linux
HDD	Hard Disk Drive
RAM	Random Access Memory
LAN	Local Area Network
LVS-DR	Linux Virtual Server –Direct Routing
RIP	Real IP Address

CHAPTER I

INTRODUCTION

1.1 Project Background

Load balancing is a technique to distribute workload evenly across two or more computers, network links, CPUs, hard drives, or other resources, in order to get optimal resource utilization, maximize throughput, minimize response time, and avoid overload[1]. This chapter will describe about the web server load balancing. Load balancing helps make networks more efficient. The main function is to divide the amount of work that a server has to do between two or more server so that more work gets done in the short amount of time and make client get served faster. It distributes traffic efficiently among network servers so that no individual server is overburdened. Networks that receive high amounts of traffic may even have one or more servers dedicated to balancing the load among the other servers and devices in the network[2] . Load balancing can be implemented with hardware, software, or a combination of both.

The purpose of load balancer scheme implementation in UTeM's lab environment is to maintain the equilibrium of web client request. Typically, two or more web servers are employed in a load balancing scheme. In case one of the servers begins to get overloaded, the requests are forwarded to another server. Load balancing brings down the service time by allowing multiple servers to handle the requests. This service time is reduced by using a load balancer to identify which server has the appropriate availability to receive the traffic[3] .

When a service is providing through, a webpage request is sent to the load balancer, which forwards the request to one of the servers. Then, that server responds back to the load balancer, which in turn sends the request on to the end user. Load balancing allows the service to continue even in the face of server down time due to server failure or server maintenance. If there are load balancing across several servers and one of the servers fails, the service will still be available to end users, as the traffic will be diverted to the other servers in your server farm. It is a project started to create a full-featured virtual router for Linux, which include load balancing through Linux Virtual Server, failover via VRRP and health checks to monitor real servers.

1.2 Problem Statements

Normally, the surfing time to the website increase when the request from client increase. The requested website by client not balance in network. In order to perform a better performance can implement load balancing, because it helps enhances the performance of the servers, leads to their optimal utilization and ensures that no single server is overwhelmed.

Server load balancer allows the addition and removal of servers to a site at any time, and the effect is immediate. Among other advantages, this allows for the maintenance of any machine, even during peak hours with little or no impact to the site. A load balancer can also intelligently direct traffic using cookies, URL parsing, static and dynamic algorithms, and much more[4] . Table 1.1 and Table 1.2 shows the research problems and research question are occur.

Table 1.1 : Shows The Research Problems In This Project

No	Research Problem
RP 1	There are many request from users to access the internet services. It will effected the limited bandwidth. The main problem is to determine best way to balance data arise.
RP 2	There are many software use for load balancing in the market currently. Which software can give better solution to users.

Table 1.2 : Shows The Research Problems and Research Question

RP	RQ	Research Question
1	1	What is the best way to balance data when web server received too much request?
2	2	How to determine the best software to use in load balancing implementation?

1.3 Objective

Objective that be achieved at the end of this project are describe below and Table 1.3 summarize the objective that be achieve at the end of this project :

i. To determine best way to balance data when too much request in web server : Load balancing can be implement by using hardware, software or combination hardware and software. But at this project only focusing at software implementation. There are three packet-forwarding methods NAT, tunneling, and direct routing that can be used for implement load balancing.

ii. To figure out best software to use in load balancing implementation : There are a few software that can be use in order to implement load balancing such as Piranha, Keepalived, Kernel TCP Virtual Server and so forth.

Table 1.3 : Shows The Research Problems, Research Questions, and Research Objectives In This Project.

RP	RQ	RO	Objective
1	1	1	To determine best way to balance data when too much request in web server.
2	2	2	To figure out best software to use in load balancing implementation

1.4 Scope

Scope of PSM project that is going to be conducted as follows:

i. The project will be conducted in lab environment .

ii. There is two high availability component involve in this project. One is call the director and another one is a real server. The load balancer and high availability service will be provided by the director. All request from client need to go through the director which decided which real server to forward the request base on algorithms configured. The real server is a actual server that providing services such as web server, mail and so forth.

1.5 Project significance

To verify whether the Piranha is the best software to implement server load balancing and help enhances the performance of the servers.

1.6 Expected Output

The expected output to overcome when browse the virtual server ip from client computer client can see the web server interface. It is to check it running. When web server interface appears, that means this project is successful.

1.7 Conclusion

This chapter help to understand what the project background, scope of the project and problem statement clearly before started this project. Load balancer is use to balance the request make by the client and make the system better performance. In the next chapter, will explain about literature review and project methodology.

CHAPTER II

LITERATURE REVIEW AND PROJECT METHODOLOGY

2.1 Introduction

A literature review is a body of text that aims to review the critical points of current knowledge on a particular topic. It can be published and unpublished work from secondary sources of data in the areas of specific interest to the researches. One of the purposes of do the literature review is to make sure that no important techniques or an aspect that has been found in the past will be ignored on this project. Background information and justification of the research undertaken will be write in this chapter by the researcher. This chapter describe about fact and finding that researcher get from journal, book or website and the project requirement including the software, hardware and network requirement.

This chapter also describe about project methodology that used in this project. Project methodology is a combination of step-by-step methods and techniques for successful planning of projects. For develop this project, it combines many of disciplines, project management, analysis, specification, design, testing in order to complete this project.

2.2 Literature Review

Literature review is about the investigating and analyzing the current systems that are similar and applied same technology with the system that will be implementing.

2.2.1 Domain

The domain for this project is ICT in Advance manufacturing Technology based on Networking and Distributed Computing. The Internet is ever increasing popularity demands the web sites to handle large amount of requests. This has created an urgent need for a more powerful web server architecture to handle this problem. One of the solution is implement load balancing. A promising solution to load balancing is a distributed architecture that can route incoming requests transparently among several server nodes. This approach can improve throughput performance and web server system scalability.

2.2.2 Keyword

There are several terms being used in this project and served as the keywords in doing research, they are described as below:

i. Load balancer : Popular Web sites cannot rely on a single powerful server nor on independent mirrored-servers to support the ever-increasing request load. Distributed Web server architectures that transparently schedule client requests offer a way to meet dynamic scalability and availability requirements[5] . With the rapid growth of both information and users, how to effectively improve the quality of network service becomes an urgent problem to be addressed. Load balancing is to overcome this problem in an effective way[6]. Load balancer is the distributing processing and communication activity evenly across a computer network so that no single device is overwhelmed. There are several method of balancing load like round