MASTER'S THESIS

Mechanisms for provisioning quality of service in web servers

Chan, Ka Ho

Date of Award:
2008

Mechanisms for Provisioning Quality of Service in Web Servers

CHAN Ka Ho

A thesis submitted in partial fulfillment of the requirements
for the degree of
Master of Philosophy

Principal Supervisor: Dr. Xiaowen Chu

Hong Kong Baptist University

May 2008

Abstract

Since the number of web users is ever-increasing and the types of web services keep on growing, the ways to boost web server performance and server capacity are very important in nowadays commercial world. As the hardware cost keeps dropping, users can purchase a computer server with high computation power at a low price. Therefore, the price limitation of the server machine is eased out. It is obvious that even the most powerful server may still fail to provide timely services for every user. There are two ways to solve the problem: (1) the introduction of service differentiation techniques to provide better services for the premium users; (2) the application of cluster-based web servers.

In order to provide service differentiation, Proportional-Integral (PI) controllers and fuzzy controllers are attractive candidates. However, PI controllers require an accurate model and assume a linear characteristic. With such limitations, PI controllers cannot work well for the web server application. On the other hand, fuzzy controllers, based on complicated fuzzy sets, can work well under bursty environment. However, there are large amount of parameters needed to be tuned which are very difficult to make initial approximate adjustment since there are no cookery book to do it. Moreover, these parameters settings usually suffer from nonzero steady state error. In order to combine the advantages of the PI controllers and the fuzzy controllers, we have proposed the fuzzy PI controller. There are only simple fuzzy set definitions. With the fuzzy controllers' characteristics, the settling time is highly reduced. At the same time, it performs well in steady state just like the PI controllers.

Some people may argue that employing a Content-Delivery Network (CDN) can

be a good way to handle the loading problem of the web servers. However, large amount of investment is required. For most of the enterprises, this is not a reasonable and applicable method. Therefore, we would like to employ our proposed fuzzy PI controllers on top of the cluster-based web servers. In order to handle the system uncertainties and load balancing problem, we would like to dispatch the connections in a weighted round robin manner with a preemptive mechanism. It is manifest that the cluster environment is more chaotic. Therefore, a fuzzy PI controller with preemptive nature is a good way to handle the uncertainties. Thus, better services can be provided for the premium users. At the same time, the total number of clients is governed by the number of the back-end server nodes. In short, the fuzzy PI controllers can be scaled up and the investment is only limited which are beneficial for small to medium scale enterprises.

On the other hand, the computation power of the server machines is increasing conspicuously. However, the server capacity has not been enhanced by much. A number of researchers have proposed the application of high performance computation methods to boost the server capacity. However, one of the well known web servers, Apache, cannot support large amount of concurrent clients. The advantages of the Apache web server application are (1) the large amount of built-in modules; (2) the ease of building new supporting modules. In order to support high performance computing by the Apache web server, a high performance thin client proxy server, which serves as the middle-tiered between the clients and the Apache web server, is developed. The proxy server can support a large amount of client connections. Therefore, the overall server capacity can be enhanced.

Table of Contents