異質網頁伺服器負載平衡系統

Load Balancing Strategy for Heterogeneous Web Server System

包蒼龍 Tsang-Long Pao
大同大學資訊工程學系副教授
Associate Professor
Dept. of Information Science & Engineering, Tatung University
大同大學資訊工程學系系所主任
Director
Dept. of Information Science & Engineering, Tatung University

陳建伯 Jian-Bo Chen
銘傳大學電腦與通訊工程學系講師
Lecturer
Dept. of Information & Telecommunications Engineering, Ming Chuan University
銘傳大學資訊網路處網路技術組組長
Section Chief, Network Technology Section
Information and Network Division, Ming Chuan University

(Received November 28, 2005; Revised February 18, 2006; Accepted March 9, 2006)

摘要：負載平衡系統可以改善熱門網站的效能。大部分的負載平衡系統使用同質性的網頁伺服器來當成後端的伺服器。在這種情況之下，工作派遣的演算法可以很容易的在這些網頁伺服器中達成負載分配。然而，當這些網頁伺服器的服務能力是不同的時候，我們必須使用某些特定的因素，比方說是剩餘能力等，來當成做決定的條件。在我們所提出的系統架構中，我們是使用剩餘能力當成是派遣演算法的參數，同時並與輪詢演算法及最小連線數演算法做效能上的比較，比較了命中率、使用率以及丟棄率。我們同時也測試當系統過載時，需要增加哪一種等級的伺服器以及其數量，才能使系統可以負荷。

關鍵字：負載平衡、異質網頁伺服器、輪詢、最小連線數、剩餘能力

Abstract: Load balancing system can improve the performance of a popular Web server. Most of the load balancing systems use homogeneous Web servers as their back-end servers. In this situation, the dispatching algorithm can obtain load sharing among the Web servers quite easily. However, when the Web server system contains servers with different capacities, we must use factors such as the remaining capacity of these Web servers as the criteria for decision making. In our proposed system, we use the remaining capacity as the parameter of the dispatching algorithm and compare the performance with systems using round robin and least connection algorithms. The hit rate, utilization, and drop rate are compared. We also derived a formula to calculate how many servers are needed when the system is overloaded and what kind of capacity the server is needed.

Key words: load balance, heterogeneous web server, round robin, least connection, remaining capacity