

Chapter 4

Energy–Efficient Server Selection Algorithms for Distributed Applications

Tomoya Enokido
Rissho University, Japan

Ailixier Aikebaier
Seikei University, Japan

Makoto Takizawa
Seikei University, Japan

ABSTRACT

A wide range of distributed applications are developed on various types of scalable information systems like Peer-to-Peer (P2P) and cloud computing systems. Various types of distributed algorithms are so far developed to minimize the computation time and response time and to maximize the throughput. On the other hand, it is getting more significant to realize energy-aware information systems for reducing the consumption of natural resources. In order to realize the energy-aware information systems, it is critical to reduce the total electrical power consumption of computers and networks. In this chapter, the authors classify applications on distributed systems into computation, communication, and general types of applications. Then, they define power consumption models of a server to perform each type of application process. Based on the power consumption models, the authors propose algorithms to select one of the servers for each type of application so that the total power consumption of servers can be reduced and also evaluate the algorithms in terms of not only performance but also energy consumption.

1. INTRODUCTION

Information systems are now becoming scalable like Peer-to-Peer (P2P) (Buford, Yu, & Lua, 2009;

Perera, 2008) and cloud systems (Grossman, 2009; Zhang & Zhou, 2009). Types of computational devices like server computers and sensor nodes are interconnected in networks like wireless and wired networks. There are various kinds of distributed applications like Web applications

DOI: 10.4018/978-1-4666-1839-8.ch004

(Bianchini & Carrera, 2000), database applications (Ferraiolo, Kuhn, & Chandramouli, 2007), and Google file systems (Ghemawat, Gobioff, & Leung, 2003) on distributed systems. In order to realize distributed applications, various types of distributed algorithms are so far developed. For example, algorithms for allocating computation resources like CPU and memory to processes and synchronizing multiple conflicting processes are discussed to minimize the computation time and response time, maximize the throughput, and minimize the memory space (Rahimi & Haug, 2010; Enokido & Takizawa, 2010). On the other hand, it is getting more significant to realize energy-aware information systems for reducing the consumption of natural resources and resolve air pollution on the Earth as discussed in Green IT (Hemmert, 2010; Murphy, Sterling, & Dekate, 2010). It is critical to reduce the total electrical power consumption of information systems in order to realize energy-aware information systems. Various kinds of hardware technologies like low-power consumption CPUs are now being developed. We do not discuss the hardware specification of each server like CPU and memory at the micro level. We rather consider how much electronic power a whole server consumes at a macro level. We measure how much electric power a server consumes to perform each type of application by using the power meter Watts up?.Net (Watts up? Products, 2011) where the power consumption of each server can be measured every one second.

In this chapter, we consider a cluster of server computers and discuss how to reduce the total power consumption of the server computers. In the cluster system, a client issues a request to a load balancer. Each time the load balancer receives a request from a client, the load balancer selects a server in the cluster and forwards the request to the server. On receipt of the request, the server performs the request as an application process and sends a response to the client. We classify applications, which are performed on the cluster system into three types: computation, communi-

cation, and general types of applications. In the computation type applications, a client issues a request to a server and the server mainly consumes computation resources to process the request. Database applications are typical examples of the computation type applications. In the communication type applications, a server mainly consumes network resources to transmit a large volume of data to a client like File Transfer Protocol (FTP) applications. In the general type of applications, applications are composed of both the computation and communication modules. Apache Web server (Apache HTTP Server Version 2.0 Documentation, 2011) using the Deflate function (Apache Module mod_deflate, 2011) is a typical example of the general type of applications. Suppose a client issues a request to a Web server, which is implemented by Apache 2.0 to obtain a Web page. On receipt of the request, the Web server obtains data in requested Web pages and compresses the data into a file by using the Deflate module. Then, the Web server transmits the compressed file to the requesting client.

First, we discuss a simple computation model and a power consumption model of a server for a computation type application. In the simple computation model, any application process is performed on a server with maximum clock frequency and the more number of application processes are concurrently performed, the larger computation resources is consumed. According to the simple computation model, we discuss how to estimate when current application processes terminate on a server. In the simple power consumption model, if at least one application process is performed on a server, the electric power is maximally consumed on the server. Based on the simple computation model and the power consumption model, we propose two types of server selection algorithms to reduce the total power consumption of servers for computation type applications: *Computation Latency Based* (CLB) and *Power Consumption Latency Based* (PCLB) algorithms. In the CLB algorithm, a server which can most early finish

35 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/energy-efficient-server-selection-algorithms/67380

Related Content

Strategic Management of Innovation Focusing on Confluence of Continuity and Change

J. S. A. Bhat, Sushil and P. K. Jain (2011). *International Journal of Green Computing* (pp. 106-126).

www.irma-international.org/article/strategic-management-innovation-focusing-confluence/61378

Augmenting the Potential of Ambient Energy With Special Reference to Solar Energy for Bringing a Paradigm Shift in the Indian Economy

Abhishek Srivastava and Saurajit Mallick (2022). *International Journal of Social Ecology and Sustainable Development* (pp. 1-14).

www.irma-international.org/article/augmenting-the-potential-of-ambient-energy-with-special-reference-to-solar-energy-for-bringing-a-paradigm-shift-in-the-indian-economy/288539

Comparison of Garbage Classification Frameworks Using Transfer Learning and CNN

Mahendra Kumar Gourisaria, Rakshit Agrawal, Vinayak Singh, Manoj Sahni and Linesh Raja (2022). *International Journal of Social Ecology and Sustainable Development* (pp. 1-23).

www.irma-international.org/article/comparison-of-garbage-classification-frameworks-using-transfer-learning-and-cnn/313973

An Intelligent Approval System for City Construction based on Cloud Computing and Big Data

Guanlin Chen, Erpeng Wang, Xinxin Sun and Yizhe Lu (2020). *Sustainable Infrastructure: Breakthroughs in Research and Practice* (pp. 248-263).

www.irma-international.org/chapter/an-intelligent-approval-system-for-city-construction-based-on-cloud-computing-and-big-data/240843

Strategic Behaviors in Sustainable Development: The Case of Tunisian Banks

Ines Shili Heraghi and Ezzeddine Zouari (2016). *International Journal of Sustainable Economies Management* (pp. 25-38).

www.irma-international.org/article/strategic-behaviors-in-sustainable-development/176621