Chapter 18

Distributed Dynamic Load Balancing in P2P Grid Systems

You-Fu Yu
National Taichung University, Taiwan, ROC

Po-Jung Huang
National Taichung University, Taiwan, ROC

Kuan-Chou Lai
National Taichung University, Taiwan, ROC

ABSTRACT

P2P Grids could solve large-scale scientific problems by using geographically distributed heterogeneous resources. However, a number of major technical obstacles must be overcome before this potential can be realized. One critical problem to improve the effective utilization of P2P Grids is the efficient load balancing. This chapter addresses the above-mentioned problem by using a distributed load balancing policy. In this chapter, we propose a P2P communication mechanism, which is built to deliver varied information across heterogeneous Grid systems. Basing on this P2P communication mechanism, we develop a load balancing policy for improving the utilization of distributed computing resources. We also develop a P2P resource monitoring system to capture the dynamic resource information for the decision making of load balancing. Moreover, experimental results show that the proposed load balancing policy indeed improves the utilization and achieves effective load balancing.

INTRODUCTION

Recently, grid computing is one of attractive architectures for high-performance computing. The grid computing system is an Internet-scale distributed computing system for sharing distributed resources across the traditional organization boundary. In grid systems, the most important issues include how to integrate the dynamically heterogeneous distributed resources, and how to improve the utilization of these integrated resources (Dandamudi, 1995). Although these various grid projects aim at sharing distributed resources from different virtual organizations (VOs), it is still difficult to share distributed
Distributed Dynamic Load Balancing in P2P Grid Systems

resources due to the different goals in building different VOs.

The peer-to-peer (P2P) computing system is another Internet-scale computing model where computers share distributed resources via exchanges among the participating computers (An- droutsellis-Theotokis et al., 2004; Li et al., 2006). The widespread deployment of P2P computing systems offers great potential for resource sharing. The P2P system has the similar objective of the grid system to coordinate large sets of distributed resources. Therefore, many projects attempt to integrate these two complementary technologies to form an ideal distributed computing system (Amoretti et al., 2005; Shan et al., 2002; Shudo et al., 2005).

In this chapter, we propose a P2P-based mechanism to form a P2P Grid platform for achieving load balancing of distributed computing resources. In general, the job submission in grid systems is carried out by a global resource broker to distribute load. Here, we propose a campus-to-campus Uni-P2P communication model to integrate the Taiwan UniGrid (Taiwan UniGrid, 2009) and the Taiwan TIGER system (Yang et al., 2005) by using a P2P communication mechanism which builds the communication pipes among sites in different grid systems. This campus-to-campus Uni-P2P communication model also supports a P2P resource monitoring system that captures the dynamic resource usage. In the P2P Grid platform, super peers are employed to manage grid sites. The concept of super peers, which exhibit more powerful computing ability, bandwidth and hardware capacity, is also considered in this Uni-P2P communication model to improve the efficiency of searching distributed resources. Moreover, we propose a dynamic distributed load balancing policy to improve the idle resource utilization in the P2P Grid platform.

The rest of this chapter is structured as follows: related works are discussed in section 2 followed by the discussion of the system architecture in sections 3. Experimental results are shown in section 4. Section 5 describes conclusions and future research directions.

RELATED WORKS

There are many middlewares (e.g., Globus Toolkit, Unicore, gLite, etc.) which have been developed for grid systems. Most of them focus on providing the core middleware services for supporting the development functionality of high-level applications. However, they usually depend on specialized servers to maintain the distributed resource information. On the other hand, P2P systems adopt decentralized resource discovery approaches and thus do not rely on any specialized servers to capture distributed resource information. In this section, we present the related works of grid information systems and load balancing policies.

Resource Monitoring Systems

There are resource monitoring software for capturing the resource information, such as Ganglia, Gstat (LCG), MDS, NWS and REMOS. Ganglia is a distributed resource monitoring system; it monitors system performance and system information such as CPU load, memory usage, hard disk usage, I/O load, and network bandwidth. Gstat is the resource monitoring tool developed by ASGC in order to support the members of EGEE in handling global grid resources. Gstat supports information such as the number of CPUs and their load, the number of waiting/running jobs, and the response time from GIIS. MDS (Monitor and Discovery System) is one of the Globus Toolkits; it supports information services and monitors/searches grid resources. NWS (Network Weather Services) is also a distributed resource monitoring system. It monitors the performance of networks and computing resources periodically, and then predicts future system performance by real time information. REMOS (REsource MOntoring System) allows the application to capture the