Chapter II
Resource-Aware Load Balancing of Parallel Applications

Eric Aubanel
University of New Brunswick, Faculty of Computer Science, Canada

ABSTRACT

The problem of load balancing parallel applications is particularly challenging on computational grids, since the characteristics of both the application and the platform must be taken into account. This chapter reviews the wide range of solutions that have been proposed. It considers tightly coupled parallel applications that can be described by an undirected graph representing concurrent execution of tasks and communication of tasks, executing on computational grids with static and dynamic network and processor performance. While a rich set of solution techniques have been proposed, there has not been of yet any performance comparisons between them. Such comparisons will require parallel benchmarks and computational grid emulators and simulators.

INTRODUCTION

Distributed high performance computing (HPC) applications have formed an important class of grid applications from the early days of the I-Way (DeFanti et al., 1996; Foster and Kesselman, 1999) to the TeraGrid (http://www.teragrid.org) of today. The main reason is that the aggregation of multiple parallel computers permits problem solutions that require more resources than are available in a single system. Many of these applications, such as partial differential equation (PDE) solvers, can be described by an undirected graph representing concurrent execution of tasks and communication between tasks, as in Foster’s PCAM design methodology (Foster, 1995). Parallel execution requires partitioning of the application’s graph in such a way that communication between the resulting
Resource-Aware Load Balancing of Parallel Applications

Recent work has begun to address the problem of resource-aware load balancing. This includes DRUM, JOSTLE, MinEX, PaGrid, PART, and SCOTCH, among others. Our goal is to review and contrast these recent efforts and discuss the many avenues for future work. A comprehensive review of this area has not been published to date (Li and Lan, 2004; Devine et al., 2005).

BACKGROUND

Since heterogeneous computing is a broad research area and there are many definitions of computational grids, we define the characteristics of the applications and platforms that we are considering.

Applications

We consider tightly-coupled parallel applications that can be described by an undirected graph representing concurrent execution of tasks and communication between tasks. The most important class concerns the numerical solution of PDEs (Schloegel, Karypis and Kumar, 2003), but such applications also include molecular dynamics problems (Koenig and Kalé, 2007) and cellular automata (Cappuccio et al., 2001). Parallel execution requires partitioning of the application’s graph in such a way that communication between the resulting subgraphs is minimized and the load is roughly balanced. These applications are typically iterative, with alternating communication and computation phases (Botadra et al., 2007). Although partial overlap is possible, the entire communication overhead cannot be overlapped with computation because of the dependency between subsequent iterations. Scheduling of operations on dense matrices on heterogeneous systems has been studied (Dongarra and Lastovetsky, 2006), but is not considered here. We are also not concerned here with scheduling of independent tasks or scheduling of workflows represented by directed...
Related Content

Managing Inconsistencies in Data Grid Environments: A Practical Approach
[www.igi-global.com/article/managing-inconsistencies-data-grid-environments/47211?camid=4v1a](www.igi-global.com/article/managing-inconsistencies-data-grid-environments/47211?camid=4v1a)

Implementation of a DES Environment
[www.igi-global.com/chapter/implementation-des-environment/38266?camid=4v1a](www.igi-global.com/chapter/implementation-des-environment/38266?camid=4v1a)

Sustainable Consumerism via Context-Aware Shopping
[www.igi-global.com/article/sustainable-consumerism-via-context-aware-shopping/188859?camid=4v1a](www.igi-global.com/article/sustainable-consumerism-via-context-aware-shopping/188859?camid=4v1a)

Gpop: A Global File Popularity Measurement for Unstructured P2P Networks
[www.igi-global.com/article/gpop/127082?camid=4v1a](www.igi-global.com/article/gpop/127082?camid=4v1a)