

## Chapter 5.4

# The Cost-Based Resource Management in Combination with QoS for Grid Computing

**Chuliang Weng**

*Shanghai Jiao Tong University, China*

**Jian Cao**

*Shanghai Jiao Tong University, China*

**Minglu Li**

*Shanghai Jiao Tong University, China*

### ABSTRACT

*In the grid context, the scheduling can be grouped into two categories: offline scheduling and online scheduling. In the offline scheduling scenario, the sequence of jobs is known in advance, scheduling is based on information about all jobs in the sequence. While, in the online scheduling scenario a job is known only after all predecessors have been scheduled, and a job is scheduled only according to information of its predecessors in the sequence. This chapter focuses on resource management issue in the grid context, and introduces the two cost-based scheduling algorithms for offline job assignment and online job assignment on the computational grid, respectively.*

### 1. INTRODUCTION

As a new infrastructure for next generation computing, computational grid enables the sharing, selection, and aggregation of geographically distributed heterogeneous resources for solving large-scale problems in science, engineering and commerce (Foster, I. & Kesselman, C.,

1999). Many studies have focused on providing middleware and software programming layers to facilitate grid computing. There are a number of projects such as Globus, ChinaGrid, EcoGRID that deal with a variety of problems such as resource specification, information service, resource allocation and security issues in a grid environment involving different administrative domains. However, a crucial issue for the efficient deployment of distributed applications on the grid

DOI: 10.4018/978-1-4666-0879-5.ch5.4

is that of scheduling (Foster, I. & Kesselman, C. (Ed.), 1996).

A computational grid consists of geographically distributed heterogeneous resources, such as CPU with different speed, and network with different bandwidth. These resources in the grid context are independent, while they are not even directly comparable because they are measured in unrelated units. This will make it difficult to choose the optimal machine from the list of available machines to which to schedule a given job.

The concept of cost is adopted for the scheduling problem based on economic principles. The key of the method is to convert the total usage of different kinds of resources, such as CPU, memory and bandwidth into a homogeneous cost. According to the goal of minimizing the total cost, arrival jobs are scheduled in the computational grid. This method could facilitate the determination on scheduling, while considering variety of aspects which have influenced on the scheduling performance. However, the open issue of the cost-based scheduling strategy is how to determine the cost based on the usage factor of the variety of resources.

In the grid context, the scheduling can be grouped into two categories: offline scheduling and online scheduling. In the offline scheduling scenario, the sequence of jobs is known in advance, scheduling is based on information about all jobs in the sequence. While, in the online scheduling scenario a job is known only after all predecessors have been scheduled, and a job is scheduled only according to information of its predecessors in the sequence. In this chapter, we focus on resource management issue in the grid context, and introduce the two cost-based scheduling algorithms for offline job assignment and online job assignment on the computational grid, respectively.

Firstly, the cost-based resource management and scheduling methodology is introduced for the computational grid, which borrows the idea from economic principles. Then, a cost-based offline scheduling algorithm Qsufferage is presented for

the offline scheduling mode in the grid environment. The algorithm considers the location of each task's input data, while makespan and response ratio are chosen as metrics for performance evaluation. The Qsufferage algorithm determines scheduling policy with minimizing the makespan of the whole application and minimizing the waiting time for executing as QoS for an individual task in the application. The performance of algorithm Qsufferage, Xsufferage, Sufferage, Min-min and Max-min is tested by simulation that is based on the SimGrid Toolkit.

Thirdly, a cost-based online scheduling algorithm is presented for the online scheduling mode in the grid environment. Compared to other online scheduling algorithms, the presented algorithm can provide the lower limit of performance with theoretical guarantee. For validating the effectiveness of the presented algorithm, we compare the performance of the presented algorithm with the greedy algorithm. Experimental result shows that the presented algorithm can outperform the greedy algorithm.

The work is inspired by the economic mechanism, through which the total usage of different kinds of resources, such as CPU, memory and bandwidth can be converted into a single cost. It is convenient for scheduler to assign job in the computational grid according to minimizing cost of all resources, achieving the goal of achieving the maximal performance of the grid system.

## **2. BACKGROUND**

Ensuring that the variety of applications would achieve good performance in the grid environment is not a trivial task, and a number of issues make scheduling such applications challenging. Resources on the grid are typically shared so that the contention created by multiple applications results in dynamically fluctuating delays and qualities of service. Moreover, due to dynamic nature of grids, information about the whole

13 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/cost-based-resource-management-combination/64530](http://www.igi-global.com/chapter/cost-based-resource-management-combination/64530)

## Related Content

---

### Performance Aware Planning Algorithms for Cloud Environments

Jyoti Thamanand Kamal Kumar (2018). *International Journal of Distributed Systems and Technologies* (pp. 1-15).

[www.irma-international.org/article/performance-aware-planning-algorithms-for-cloud-environments/196264](http://www.irma-international.org/article/performance-aware-planning-algorithms-for-cloud-environments/196264)

### Task-Based Crowd Simulation for Heterogeneous Architectures

Hugo Perez, Benjamin Hernandez, Isaac Rudominand Eduard Ayguade (2016). *Innovative Research and Applications in Next-Generation High Performance Computing* (pp. 194-219).

[www.irma-international.org/chapter/task-based-crowd-simulation-for-heterogeneous-architectures/159045](http://www.irma-international.org/chapter/task-based-crowd-simulation-for-heterogeneous-architectures/159045)

### Sustainable and Interoperable e-Infrastructures for Research and Business

Giuseppe Andronico, Roberto Barbera, Marco Fargetta, Emidio Giorgio, Salvatore Marcoand Diego Scardaci (2009). *Grid Technology for Maximizing Collaborative Decision Management and Support: Advancing Effective Virtual Organizations* (pp. 23-39).

[www.irma-international.org/chapter/sustainable-interoperable-infrastructures-research-business/19337](http://www.irma-international.org/chapter/sustainable-interoperable-infrastructures-research-business/19337)

### User Interface Distribution Method Based on Pi-Calculus

Sergey Sakulin, Alexander Alifimtsev, Evgeny Tipsin, Vladimir Devyatkovand Dmitry Sokolov (2019). *International Journal of Distributed Systems and Technologies* (pp. 1-20).

[www.irma-international.org/article/user-interface-distribution-method-based-on-pi-calculus/232303](http://www.irma-international.org/article/user-interface-distribution-method-based-on-pi-calculus/232303)

### Maintaining Redundancy in Peer-to-Peer Storage Systems

Anwitaman Datta, Di Wu, Liu Xinand Adam Wierzbicki (2010). *Handbook of Research on P2P and Grid Systems for Service-Oriented Computing: Models, Methodologies and Applications* (pp. 616-634).

[www.irma-international.org/chapter/maintaining-redundancy-peer-peer-storage/40820](http://www.irma-international.org/chapter/maintaining-redundancy-peer-peer-storage/40820)