

Chapter 1.6

Reliability and Performance Models for Grid Computing

Yuan-Shun Dai

University of Electronics Science Technology of China, China & University of Tennessee, USA

Jack Dongarra

University of Tennessee, USA; Oak Ridge National Laboratory, USA; & University of Manchester, UK

ABSTRACT

Grid computing is a newly developed technology for complex systems with large-scale resource sharing, wide-area communication, and multi-institutional collaboration. It is hard to analyze and model the Grid reliability because of its largeness, complexity and stiffness. Therefore, this chapter introduces the Grid computing technology, presents different types of failures in grid system, models the grid reliability with star structure and tree structure, and finally studies optimization problems for grid task partitioning and allocation. The chapter then presents models for star-topology considering data dependence and tree-structure considering failure correlation. Evaluation tools and algorithms are developed, evolved from Universal generating function and Graph Theory. Then, the failure correlation and data dependence are considered in the model. Numerical examples are illustrated to show the modeling and analysis.

INTRODUCTION

Grid computing (Foster & Kesselman, 2003) is a newly developed technology for complex systems with large-scale resource sharing, wide-area communication, and multi-institutional collaboration etc, see e.g. Kumar (2000), Das *et al.* (2001), Foster *et al.* (2001, 2002) and Berman *et al.* (2003).

Many experts believe that the grid technologies will offer a second chance to fulfill the promises of the Internet.

The real and specific problem that underlies the Grid concept is coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations (Foster *et al.*, 2001). The sharing that we are concerned with is not primarily file exchange but rather direct access to computers, software, data, and other resources. This is required

DOI: 10.4018/978-1-4666-0879-5.ch1.6

by a range of collaborative problem-solving and resource-brokering strategies emerging in industry, science, and engineering. This sharing is highly controlled by the resource management system (Livny & Raman, 1998), with resource providers and consumers defining what is shared, who is allowed to share, and the conditions under which the sharing occurs.

Recently, the Open Grid Service Architecture (Foster *et al.*, 2002) enables the integration of services and resources across distributed, heterogeneous, dynamic, virtual organizations. A grid service is desired to complete a set of programs under the circumstances of grid computing. The programs may require using remote resources that are distributed. However, the programs initially do not know the site information of those remote resources in such a large-scale computing environment, so the resource management system (the brain of the grid) plays an important role in managing the pool of shared resources, in matching the programs to their requested resources, and in controlling them to reach and use the resources through wide-area network.

The structure and functions of the resource management system (RMS) in the grid have been introduced in details by Livny & Raman (1998), Cao *et al.* (2002), Krauter *et al.* (2002) and Nabrzyski *et al.* (2003). Briefly stated, the programs in a grid service send their requests for resources to the RMS. The RMS adds these requests into the request queue (Livny & Raman, 1998). Then, the requests are waiting in the queue for the matching service of the RMS for a period of time (called waiting time), see e.g. Abramson *et al.* (2002). In the matching service, the RMS matches the requests to the shared resources in the grid (Ding *et al.*, 2002) and then builds the connection between the programs and their required resources. Thereafter, the programs can obtain access to the remote resources and exchange information with them through the channels. The grid security mechanism then operates to control the resource access through the Certification, Au-

thorization and Authentication, which constitute various logical connections that causes dynamicity in the network topology.

Although the developmental tools and infrastructures for the grid have been widely studied (Foster & Kesselman, 2003), grid reliability analysis and evaluation are not easy because of its complexity, largeness and stiffness. The grid computing contains different types of failures that can make a service unreliable, such as blocking failures, time-out failures, matching failures, network failures, program failures and resource failures. This chapter thoroughly analyzes these failures.

Usually the grid performance measure is defined as the task execution time (service time). This index can be significantly improved by using the RMS that divides a task into a set of subtasks which can be executed in parallel by multiple online resources. Many complicated and time-consuming tasks that could not be implemented before are working well under the grid environment now.

It is observed in many grid projects that the service time experienced by the users is a random variable. Finding the distribution of this variable is important for evaluating the grid performance and improving the RMS functioning. The service time is affected by many factors. First, various available resources usually have different task processing speeds online. Thus, the task execution time can vary depending on which resource is assigned to execute the task/subtasks. Second, some resources can fail when running the subtasks, so the execution time is also affected by the resource reliability. Similarly, the communication links in grid service can be disconnected during the data transmission. Thus, the communication reliability influences the service time as well as data transmission speed through the communication channels. Moreover, the service requested by a user may be delayed due to the queue of earlier requests submitted from others. Finally, the data dependence imposes constraints on the sequence

20 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/reliability-performance-models-grid-computing/64481

Related Content

A Traitor Tracking Method Towards Deep Learning Models in Cloud Environments

Yu Zhang, Linfeng Wei, Hailiang Li, Hexin Cai and Yingxin Wu (2022). *International Journal of Grid and High Performance Computing* (pp. 1-17).

www.irma-international.org/article/a-traitor-tracking-method-towards-deep-learning-models-in-cloud-environments/301588

Empirical Evaluation of Map Reduce Based Hybrid Approach for Problem of Imbalanced Classification in Big Data

Khyati Ahlawat, Anuradha Chug and Amit Prakash Singh (2019). *International Journal of Grid and High Performance Computing* (pp. 23-45).

www.irma-international.org/article/empirical-evaluation-of-map-reduce-based-hybrid-approach-for-problem-of-imbalanced-classification-in-big-data/232211

Improving Distributed Renewable Energy Power Planning Through Particle Swarm Algorithm

Honglei Yao, Xin Liu, Yingxian Chang, Donglan Liu and Rui Wang (2024). *International Journal of Distributed Systems and Technologies* (pp. 1-14).

www.irma-international.org/article/improving-distributed-renewable-energy-power-planning-through-particle-swarm-algorithm/349743

Overview of Global Supercomputing

Richard S. Segall and Neha Gupta (2015). *Research and Applications in Global Supercomputing* (pp. 1-32).

www.irma-international.org/chapter/overview-of-global-supercomputing/124335

QoS-Based Job Scheduling and Resource Management Strategies for Grid Computing

Kuo-Chan Huang, Po-Chi Shih and Yeh-Ching Chung (2012). *Grid and Cloud Computing: Concepts, Methodologies, Tools and Applications* (pp. 1315-1331).

www.irma-international.org/chapter/qos-based-job-scheduling-resource/64541