Chapter 2
A Computational Grid Scheduling Model to Maximize Reliability Using Modified GA

Zahid Raza
Jawaharlal Nehru University, India

Deo Prakash Vidyarthi
Jawaharlal Nehru University, India

ABSTRACT

This paper presents a grid scheduling model to schedule a job on the grid with the objective of ensuring maximum reliability to the job under the current grid state. The model schedules a modular job to those resources that suit the job requirements in terms of resources while offering the most reliable environment. The reliability estimates depict true grid picture and considers the contribution of the computational resources, network links and the application awaiting allocation. The scheduling executes the interactive jobs while considering the looping structure. As scheduling on the grid is an NP hard problem, soft computing tools are often applied. This paper applies Modified Genetic Algorithm (MGA), which is an elitist selection method based on the two threshold values, to improve the solution. The MGA works on the basis of partitioning the current population in three categories: the fittest chromosomes, average fit chromosomes and the ones with worst fitness. The worst are dropped, while the fittest chromosomes of the current generation are mated with the average fit chromosomes of the previous generation to produce off-spring. The simulation results are compared with other similar grid scheduling models to study the performance of the proposed model under various grid conditions.

INTRODUCTION

Last few decades have witnessed an enormous development in the area of computing. These advancements have led to availability and use of superior hardware and powerful efficient software. The developments coupled with the growth in the web technologies have enabled the users at different geographic locations to fetch the information of their need, share their work and most importantly collaborating with others towards a problem solving.

DOI: 10.4018/978-1-4666-2065-0.ch002

Many a times it happens that research groups desire for a high end computational facility which is not available with them either due to the excessive cost or the investment not justifying the cost with respect to its contribution towards the whole research work. Contrary to this, a scenario could be in which a research group owning a high end computational facility but may not be able to exploit it to its full extent. The solution, in such cases, is to have an environment of cooperative engineering in which people can share their information and computational facilities. This need is addressed by a grid which is an aggregation of heterogeneous resources spread over multiple administrative domains. When a user becomes a part of the grid, it enables him to view the grid as a supercomputing computational resource with almost every computational capability available anywhere and allowing him to use them irrespective of his own computational capabilities and physical location. A grid could be of many types depending on the objectives of its usage. A computational grid is dedicated towards compute intense jobs; a data grid manages the storage retrieval and maintenance of data; a bio grid for biological applications and so on (Prabhu, 2008; Tarricone & Esposito, 2005; Taylor & Harrison, 2009).

Grid computing system is different from other computing systems like Distributed Computing Systems (DCS) and Cluster computing. The scale of operation of the grid in terms of the number of nodes involved and the heterogeneity of resources, their management policies and geographical locations is very large as compared to its peers with no Single System Image (SSI). The resources of various computers in a network are used concurrently to solve a single problem in a collaborative fashion. The grid which involves a participation of people and resources at a heterogeneous scale, presents huge challenges due to the varying system wide performance matrix and a common acceptable policy. For a computational grid, these issues may range from handling jobs of varying nature and requirements, services offered, load balancing, interactive job handling, grid topology, robustness, job allocation, security, reliability, maintenance and commercial feasibility (Vidyarthi, Tripathi, & Sarker, 2001). As scheduling is an NP hard problem, a number of approaches have been proposed in the literature to schedule a job on the grid considering one or the other objectives. Soft computing tools are quite useful for such class of problems where sub-optimal solution is acceptable. Genetic algorithm (GA) is one such tool that is widely used for scheduling problem. GA is different from the conventional optimization techniques as it explores various peaks by considering a set of points evaluated on the basis of an objective function rather than climbing one peak. GA works on the Darwin's theory of "survival of the fittest" by reproducing the fittest members of the past generation to generate the next population. Since the new population is resulted from the best fit members of the past generation it is expected that the new members will be more fit. Thus, over the generations, the fitness of the newer generations keep on improving and the process is repeated till the result saturates. Mutation is used after every few generations to randomly alter the properties of the population, enabling it to come out of the local optima.

Since the grid involves heterogeneity at every level i.e. the computational resources, applications demanding execution, network characteristics enabling communication, the system is always prone to failures. The failure may range from hardware to software. Although many proactive and reactive approaches to deal with these failures are available, it is always desirable that the job should be scheduled with better reliable execution at that moment of time.

A number of job schedulers for the grid are available in the literature with various approaches towards scheduling the jobs. Scheduler to minimize turnaround (Raza & Vidyarthi, 2008a) has been reported. A batch scheduling scheme to minimize makespan and flowtime using Cellular Mementic Algorithms is proposed in Xhafa,

Related Content

High-Throughput GRID Computing for Life Sciences
Giulia De Sario, Angelica Tulipano, Giacinto Donvito, Giorgio Maggiand Andreas Gisel (2012). *Grid and Cloud Computing: Concepts, Methodologies, Tools and Applications (pp. 821-840).*
www.irma-international.org/chapter/high-throughput-grid-computing-life/64517

Fuzzy based Data Fusion for Energy Efficient Internet of Things
Madan Mohan Agarwal, Mahesh Chandra Govil, Madhavi Sinhaand Saurabh Gupta (2019). *International Journal of Grid and High Performance Computing (pp. 46-58).*
www.irma-international.org/article/fuzzy-based-data-fusion-for-energy-efficient-internet-of-things/232212

Scenarios of Next Generation Grid Applications in Collaborative EnvironmentsA Business-Technical Analysis
Vassilikil Andronikou, Dimosthenis Kyriazis, Magdalini Kardara, Dimitrios Halkosand Theodora Varvarigou (2009). *Grid Technology for Maximizing Collaborative Decision Management and Support: Advancing Effective Virtual Organizations (pp. 40-60).*
www.irma-international.org/chapter/scenarios-next-generation-grid-applications/19338

Bioinformatics and Patient Survival Analysis of Digestive Tract Tumor Marker NCAPG Based on Public Medical Databases
Jiahang Wang, Mingqiang Linand Fang Ouyang (2022). *International Journal of Distributed Systems and Technologies (pp. 1-12).*
www.irma-international.org/article/bioinformatics-and-patient-survival-analysis-of-digestive-tract-tumor-marker-ncapg-based-on-public-medical-databases/307946

Energy-Efficiency in Cloud Data Centers
Burak Kantarciand Hussein T. Mouftah (2014). *Communication Infrastructures for Cloud Computing (pp. 241-263).*
www.irma-international.org/chapter/energy-efficiency-in-cloud-data-centers/82540