GA Based Scheduling Model for Computational Grid to Minimize Turnaround Time

Zahid Raza, Jawaharlal Nehru University, India
Deo P. Vidyarthi, Jawaharlal Nehru University, India

ABSTRACT

Scheduling on distributed systems is an NP hard problem and grid being a wide heterogeneous expandable system makes scheduling even a tougher job. Genetic algorithm, based on natural selection and evolution has gained popularity in recent times because of its effectiveness in handling optimization problems. In this article, a job-scheduling model for a computational grid with the objective of minimizing the turnaround time using genetic algorithm is proposed. The model evaluates various clusters in the grid to find the most suitable one with minimum turnaround time for the job-scheduling. Simulation studies compare the performance of this model with other similar models.

Keywords: Computational Grid, Genetic Algorithm, Hard Problems, Job Precedence Graph, NP, Scheduling

INTRODUCTION

The demand for computational power by compute intensive jobs requires better utilization of the available computing resources as provided by the grid platform. A grid is an aggregation of heterogeneous resources spreading over many administrative domains, capable of supporting varied applications on a grand scale. Some of the many possible uses of the grid could be collaborative engineering, cooperative engineering, data exploration, high throughput computing, meta application and high performance computing. Grid enables the user to view it as a single unified tool with an enormous computational power suited especially for compute intensive applications which are otherwise difficult to solve. Some of these applications could be weather forecasting, DNA computing, numerical data processing, signal processing and particle physics.

Grid can be further classified by the objectives of the application. A data grid specializes in abstraction, storage, data access, replica management, controlled sharing and management of large amount of distributed data, while a bio grid is dedicated to the study of the biological data desired by the rules of bioinformatics. A computational grid is mainly concerned with the allocation and execution
aspects of the job, demanding high computational power, in order to maximize the system throughput while minimizing the turnaround time of the job submitted as well as balancing load amongst the system resources. A computational grid, therefore, enables its constituents with the computational power irrespective of its own computational capabilities (Tarricone and Esposito, 2005; Ernemann, Hamscher and Yahyapour, 2002; Hamscher, Schwiegelshohn, Streit and Yahyapour, 2005; Grid Computing Info centre, 2008).

A job is considered to be a collection of various modules submitted along with its Job Precedence Graph (JPG) in which the position of a module indicates its precedence level of execution. JPG further defines the degree of concurrency for the job execution. Thus, a job is a single set of concurrent modules to be executed on a set of resources. The resources could be computers, workstations, PDA’s, storage devices and network links, constituting the grid. Scheduling is the problem of mapping of these job modules on the grid resources keeping in mind their precedence as defined in the JPG (Vidyarthi, Tripathi and Sarkar, 2001; Vidyarthi, Tripathi and Sarkar, 2001).

Genetic algorithm is an effective tool often used to solve NP class of problems. GA is better than other conventional optimization techniques, as it works on a set of points rather than on a single point. Further, due to working over a set of data simultaneously the probability of finding a false peak is reduced. Genetic Algorithm is a random search method but different from the other similar methods in the sense that it uses the information generated in the past and as well as for the present estimates (Goldberg, 2007; Mitchell, 1999).

The proposed model uses genetic algorithm to schedule the jobs on the grid with the objective of minimizing the turnaround time of the given jobs. A grid can be considered to consist of many specialized clusters. If the specialization of the cluster matches with the nature of the job, the cluster is evaluated for the execution cost that it can offer to the job. Here, nature of the job refers to the specifications and requirements of the job in terms of resource characteristics. For example, a graphics application may need computational resources specialized for graphic specific jobs or a database application may demand resources which handles faster storage, data movement and retrieval for efficient execution of the job.

The drawback with most of the task schedulers is that they have considered allocation of only one task over the nodes of the grid. This does not account for the previous workload on the nodes and therefore makes an unrealistic schedule. Moreover, the single entry point for the job becomes a bottleneck in the system’s performance, which is the case with many grid schedulers. Further, they do not consider the job’s requirements in terms of resources and assume to have control over the scheduling policy of the independent nodes, which is also unrealistic (Afzal, Darlington and McGough, 2006; Aggarwal and Aggarwal, 2006; Dalheimer, Pfrendt and Merz, 2005; Ranjan, Harwood and Buyya, 2006; Shan, Hongzhang, Oliker and Biswas, 2003; Fidanova and Durchova, 2006; Spooner, Jarvis, Cao, Saini and Nudd, 2003; Wiriyaprasit, Muangsin and Veera, 2004).

The use of genetic algorithm for dynamic load balancing has been observed and reported in (Zomaya, Yee-Hwei, 2001), where genetic algorithm has been used to design and implement a scheduler to minimize the turnaround time, make-span and the idle time of the resources of the job following the real time constraints (Aggarwal, Kent and Ngom, 2005). The Pareto algorithm using GA has been used for the wireless grids (Benedict and Vasudevan, 2008). Since GA itself leaves room for improvements, an improved GA scheme for faster optimization has been proposed in (Yin, Wu and Zhou, 2007). A threshold selection operator has been proposed in (Khanbary, Mohammed and Vidyarthi, 2009) for improved results.

The rest of the article is organized as follows. GA is briefly discussed in the next section followed by the proposed GA based scheduler along with the assumptions, data structures and the notation used. The fitness function (cost estimation) of the job allocation is derived...
Time Domain Replica Signal Based Interference Compensation for SP-MIMO/OFDM with Large Delay Spread Channel

The Impact of Virtualization on High Performance Computing Clustering in the Cloud