A Scheduling Model with Multi-Objective Optimization for Computational Grids using NSGA-II

Zahid Raza, Jawaharlal Nehru University, India
Deo Prakash Vidyarthi, Jawaharlal Nehru University, India

ABSTRACT

Scheduling a job on the grid is an NP Hard problem, and hence a number of models on optimizing one or other characteristic parameters have been proposed in the literature. It is expected from a computational grid to complete the job quickly in most reliable grid environment owing to the number of participants in the grid and the scarcity of the resources available. Genetic algorithm is an effective tool in solving problems that requires sub-optimal solutions and finds uses in multi-objective optimization problems. This paper addresses a multi-objective optimization problem by introducing a scheduling model for a modular job on a computational grid with a dual objective, minimizing the turnaround time and maximizing the reliability of the job execution using NSGA – II, a GA variant. The cost of execution on a node is measured on the basis of the node characteristics, the job attributes and the network properties. Simulation study and a comparison of the results with other similar models reveal the effectiveness of the model.

Keywords: Computational Grid, Multi-Objective Optimization, NSGA-II, Pareto Fronts, Scheduling

INTRODUCTION

The effort to attain a cooperative, collaborative, and high performance computing environment with ubiquitous presence has led to grid formation. Aided with the enormous reach of the World Wide Web, the grid promises to provide a future computing platform with almost every computing capability. The grid is an aggregation of heterogeneous resources spread in various administrative domains working together to solve a problem. Being a part of the grid enables the user to compute anything anywhere irrespective of his own computational capabilities or presence (Prabhu, 2008; Tarricone & Esposito, 2005; Taylor & Harrison, 2009).

A grid, being composed of a number of heterogeneous participants with their own administrative domains and policies, poses a number of challenges for job execution. These challenges
range from the topology of the grid affecting the job submission to discovering the suitable resources for the job and conveying the results back to the user. Since the resources constituting the grid are scarce and the access could be under some economic policy, it is always desired to get the job executed in the most reliable manner with minimum turnaround incurred for better utilization of the grid. Reliability is the ability of a system to perform and maintain its functions in routine circumstances, as well as in hostile or unexpected circumstances. As the composition of the grid is dynamic the resources are prone to failures. Thus, under the dynamic and unreliable grid conditions it becomes essential that the job execution takes place in a reliable manner. This reliable execution further adds strength to the objective in which turnaround is minimized. The catch here is that it is not always necessary that scheduling the job on the resources with the objective of turnaround minimization may result in most reliable execution of the job and vice versa. Since optimization of one objective may not result in the optimized solution for the other objective, the two possibly conflicting objectives need to be met simultaneously. The best way to attain this is to have the solutions explored and accept the ones with better results corresponding to both the objectives. Going this way, solutions are obtained which may not be optimum for any of the objectives but certainly better when seen as a trade-off between the two objectives. This is the approach adopted in the multi-objective optimization problems (MOP). Finally, with the requirements in hand, the user or the system zeroes in on the best solutions as per the need and the domain knowledge.

A number of job schedulers for the grid are available in the literature with various approaches towards scheduling the jobs. Schedulers to minimize turnaround (Raza & Vidyarthi, 2008) and maximizing reliability of job execution (Raza, & Vidyarthi, 2009b) have been reported. A batch scheduling scheme to minimize makespan and flowtime using Cellular Mementic Algorithms is proposed in (Xhafa, Alba, & Dorronsoro, 2007). A high level timed Petri net to model the workflow of grid tasks is proposed in (Yaojun, Changjun, & Xuemei, 2005) whereas a Game Theoretic approach is used in (Rzadca, Tryatram, & Wierzbicki, 2007). A scheduling model considering the availability of the computational resources, communication delays and the resource reservation is reported in (Aggrawal & Aggrawal, 2006).

Genetic algorithm is an effective tool for solving an optimization problem that involves navigation through a big search space. It is based on the Darwin’s theory of evolution with “survival of the fittest”. GA works with exploration of various solutions simultaneously making use of the past information for the future. This, coupled with mutation, avoids the false peaks to move towards the global optima. With suitable modifications, GA finds use for multi-objective optimization problems employing specialized fitness functions along with the diversity preserving mechanisms. The set of solutions at any time in a MOP is known as Pareto-optimal solutions which are accessed from time to time. The solutions are said to be acceptable if they satisfy the given objectives without being dominated by others. The use of evolutionary algorithms like GA for MOP is much better in comparison to the classical optimization techniques which stress on the conversion of the problem to a single optimization problem by considering one Pareto-optimal solution at a time. Further, the decision on the precise use of weights itself creates a bottleneck in the performance and effectiveness for the classical methods.

The rest of the paper is organized as follows. The use of the GA in solving multi-objective optimization problems. The next section describes the model in terms of the scheduler, notation used, fitness function for turnaround time and reliability estimates and the algorithm for job scheduling using NSGA-II. The simulation study and performance comparison of the model with the simple GA and non GA based similar models is then presented.
Related Content

Riddle of the Sphinx: Paradox Revealed and Reveiled
www.irma-international.org/chapter/riddle-sphinx-paradox-revealed-reveiled/28369/

Classical Sliding and Generalized Variable Structure Controls for a Manipulator Robot Arm with Pneumatic Artificial Muscles

A Qualitative Systems Thinking Approach in Understanding the Implementation of Innovation on Construction Projects
www.irma-international.org/article/a-qualitative-systems-thinking-approach-in-understanding-the-implementation-of-innovation-on-construction-projects/122113/

Theory and Practice of Ant-Based Routing in Dynamic Telecommunication Networks
www.irma-international.org/chapter/theory-practice-ant-based-routing/28379/
A Conceptual Framework for Effective Knowledge Management Using Information and Communication Technologies
www.irma-international.org/chapter/conceptual-framework-effective-knowledge-management/68214/