سومین کنفرانس فناوری اطلاعات و دانش
۶ تا ۸ آذرماه ۱۳۸۶

# Distributed Resource Scheduling in Grid Computing Using Fuzzy Approach

Hamed Vahdat-Nejad, Reza Monsefi, Hossein Deldari

*Computer Engineering Department, Ferdowsi University of Mashhad, Iran*
{Hamed.vh@gmail.com, rmonsefi@um.ac.ir, hdeldari@yahoo.com }

**Abstract**
*Scheduling is a fundamental issue in achieving high performance in multiclusters and computational grids. In this paper, we investigate the scheduling problem using the multilayer scheduling model. For attaining scalability, the proposed scheduler uses a distributed approach with the capability of considering the local clusters advantages for executing the jobs. On the other hand, the distributed scheduler exploits the capabilities of fuzzy logic to qualitatively deal with different parameters available in the scheduling decision. Simulation results show the effectiveness of the algorithm in terms of job completion time.*

*Keywords*: Fuzzy theory, Grid computing, Job scheduling, Distributed scheduling

## 1. Introduction

A grid computing infrastructure is a collection of resources connected by a network, in which, by means of appropriate software, resource discovery and sharing is made possible [1]. One way to create a platform for grid computing is to interconnect existing separate clusters. These clusters may be located within a single organization or across different administrative organizations [2]. Scheduling is an important issue in grid computing, and parallel jobs constitute a typical workload in the scheduling scenario. Parallel jobs can be classified into two categories: Rigid and moldable [2]. Rigid parallel jobs run on a user specified number of resources, while moldable jobs can run on different number of computational resources. In this paper we consider rigid parallel jobs.

One class of applications that typically run on a grid superstructure is the class of single-program-multiple-data (SPMD) applications, also known as data-parallel applications. Data-parallel applications are partitioned into tasks which perform computations on separate pieces of a data set. The tasks work together to process the entire data set, and are collectively called a job. The data-parallel program model is often used to solve scientific computing problems. These jobs may run for many hours or days, and can consume a large amount of system resources. The job may perform a large amount of computation, communication between tasks, or both [1].

One category of rigid jobs that we have investigated is data parallel applications. A grid scheduler uses the information of grid system and jobs to produce an assignment of tasks to machines for the given grid job. The general problem of mapping tasks to machines has been shown to be NP-complete [3]. The scheduling of parallel jobs has been extensively studied in a single cluster environment [5, 6]. Several heuristic algorithms have been developed to schedule tasks to machines on heterogeneous computing systems. Eleven such scheduling algorithms have been evaluated in [4]. These algorithms are developed for heterogeneous computing systems. Some heuristic scheduling algorithms for grid environments are developed in [7, 8, 9]. They deal with tasks and machines in terms of assigning tasks to machines, and have the deficiency of not being scalable, when applied to a large scale grid. For attaining scalability, we use a distributed approach, in a way that the arrival job can be submitted to any of the