CHAPTER 3

PROBLEM DEFINITION AND PROBLEM SOLVING METHODOLOGY

Scheduling is a challenging task in grid environment. The scheduling system must optimize user's objectives such as cost, response time and deadline satisfaction. To achieve the promising potentials of tremendous distributed heterogeneous resources, effective and efficient scheduling algorithms are fundamentally important. However, there are still some unaddressed issues that need attention in scheduling algorithms for grid environment. This chapter describes the problem that has been identified in this work and outlines the methodology used in the research.

3.1 PROBLEM DEFINITION

Grid scheduler provides an interface between consumers and the underlying resources. In a grid system, computational resources are connected by network and a set of users submit applications for execution on the resources. The scheduling system of grid computing environment is responsible for managing the resources and also dealing with the set of applications. In the event of a set of jobs waiting for execution, the scheduling system should be able to allocate suitable resources to jobs and achieve required performance goals.

In existing scheduling algorithms, the nature of job is not considered while allocating them to resources. Hence, if a heavy job is
allocated to a resource with low processing capability, it will decrease the overall efficiency of the scheduling. The task of the scheduler is not only searching for suitable resources to run tasks but to consider user-dependent Quality of Service (QoS) requirements too.

Since grid resources are dynamic, an effective load balancing algorithm is required to maintain load at the resources. Most of the existing scheduling algorithms do not consider user deadline to complete the jobs. To improve user satisfaction, the deadline of the jobs needs to be considered. In order to reduce the completion time and job execution failure, failure rate of the resources has to be considered.

In this work, a computational grid system which consists of heterogeneous grid resources and user applications with different size and requirements is considered. The problem of scheduling \( n \) jobs on \( m \) computing resources with the objective of minimizing the processing time and utilizing the resources effectively is discussed. In this model, static heuristic mapping is considered for mapping meta-tasks. Each machine executes a single task at a time. In heterogeneous computing environment, the size of meta-tasks and the number of machines are known prior.

If the number of jobs is less when compared to available resources in the grid, then the jobs can easily be scheduled to suitable resources that satisfy user’s requirements. If the number of jobs is more than the number of available resources, then an efficient scheduling algorithm is required for allocation of jobs. In this research, number of jobs is more than the computing resources. A job is not assigned to different resources and job migration is not allowed.

To formulate the problem, the set of heterogeneous resources are considered as \( R = \{ R_1, \ldots, R_m \} \) and the set of independent jobs are
represented as $J = \{J_1, \ldots, J_n\}$. The problem of finding the choice of the best pairs of resources and jobs is a NP complete problem. The jobs submitted are considered as independent and can be executed in parallel with other available tasks. Also the jobs are considered static i.e. the number of tasks submitted is fixed and they do not change with time. The users submit jobs with requisition time i.e. within which the job needs to be completed, which can also be called demanded deadline of the user for the submitted jobs.

The scheduling function $S$ is defined as mapping user submitted jobs to the resources in the resource set. It is defined as

$$S = J \rightarrow R$$  \hspace{1cm} (3.1)

### 3.2 PROBLEM SOLVING METHODOLOGY

This research proposes various heuristic scheduling algorithms. The proposed scheduling algorithms are both resource centric and application centric. These algorithms consider heterogeneity of resources and jobs. The resources are characterized by varying the capability of the resources expressed in MIPS, baud rate of the resources (bps) and failure rate of the resources. Heterogeneity of the jobs includes length of the jobs expressed in Million Instructions (MI), input file size, output file size and deadline given by the user.

The proposed algorithms have the following assumptions in prior:

1. The number of meta-tasks that are to be scheduled
2. The number of resources available in the grid for processing the meta-tasks
3. The processing capacity of each resource in MIPS
4. Job length expressed in Million Instructions
User deadline of the jobs (user demand)

Baud rate of the resources

Input and Output file size

3.3 CONCLUSION

This chapter thus describes the problem that has been considered for the research. It also gives the problem solving methodology adopted to achieve the objective. The next chapter describes PUD algorithm and UDDA algorithm which mainly focus on the deadlines of the available static jobs. It also considers the expected completion time of the set of jobs.