An Overview of Task Scheduling and Performance Metrics in Grid Computing

Ashish Chandak
Department of Computer Science and Engineering, National Institute of Technology, Rourkela, India

Bibhudatta Sahoo
Department of Computer Science and Engineering, National Institute of Technology, Rourkela, India

Ashok Kumar Turuk
Department of Computer Science and Engineering, National Institute of Technology, Rourkela, India

ABSTRACT
Grid computing is form of distributed computing and task scheduling is heart of grid computing. In this paper, we studied task scheduling and various types of system model used for task scheduling. Various possible task states and performance metrics for task scheduling are identified.

Keywords
Task Scheduling, Performance Metrics, Grid Computing, Task States.

1. INTRODUCTION
The Scheduling is the creation of a schedule: a (partially) ordered list specifying how contending accesses to one or more sequentially reusable resources will be granted. Such resources may be hardware such as processors, communication paths, storage devices or they may be software, such as locks and data objects. Task scheduling is the assignment of a set of tasks to some certain resources by means of starting and ending time of tasks, subject to certain constraints [1]. Task scheduling is integrated part of grid computing. The purpose of task scheduling is to allocate resources for executing task. Task scheduling guides resource allocation. As there are many nodes where a task can be executed, the first question to be answered is how to assign the tasks to them. This assignment is known as task allocation, or global scheduling. Once tasks have been allocated, the problem becomes one of defining a feasible local schedule for each node. In this paper, we look at techniques for allocating and scheduling tasks on processors.

2. TASK SCHEDULING POLICIES
In A scheduling algorithm can be classified into clairvoyant or non clairvoyant, with regard to knowledge about characteristics of tasks. A clairvoyant scheduling algorithm may use information of tasks characteristics such as service demand, whereas a non clairvoyant algorithm assumes nothing about the characteristics of the jobs [2].

2.1 Task Assignment
Depending upon task requirement, task can be executed on single site or multiple sites.

- Single Site Scheduling: - In single site scheduling, tasks are not shared among sites. It is executed only on single site.
- Multi Site Scheduling. : - Sometimes task scheduling is not possible at single site so in this case tasks are distributed to multiple sites. Grid scheduler dispatches task to multiple sites. In this case, task is divided into subtask and it is distributed over multiple administrative domains. This allows the execution of large task requiring more nodes than available on a single site.

2.2 Types of Grid Model for Task Scheduling
System modelling is a technique to express, visualize, analyst and transform the architecture of a system. A system model is then a skeletal model of the system. It indicates how each component of system interacts within application. System modelling eases the technical transformation of systems.

a) Set of Sites

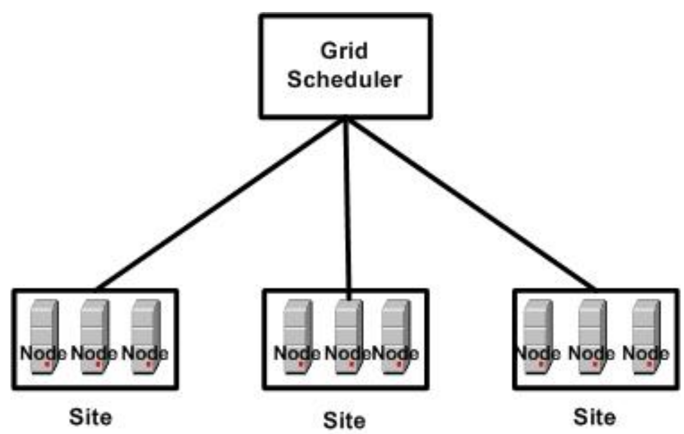

Fig. 1 Set of Sites

Site consists of number of computing nodes connected in a distributed manner. These nodes can be used to solve computational problem. Fig. 1 shows set of sites. [3, 4] consider set of sites in grid computing for task scheduling.

b) Set of Cluster

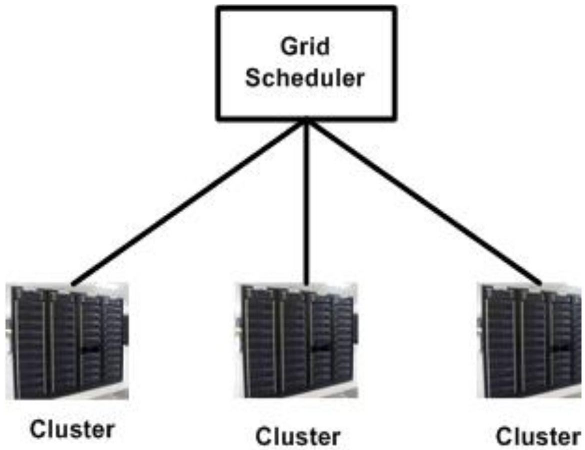

Fig. 2 Set of Clusters

In cluster computers are combined into a unified system through software and networking. This aggregation power of computer can be used for computational purpose. Fig. 2 shows set of clusters for task allocation [5].

c) Set of Mobile Nodes

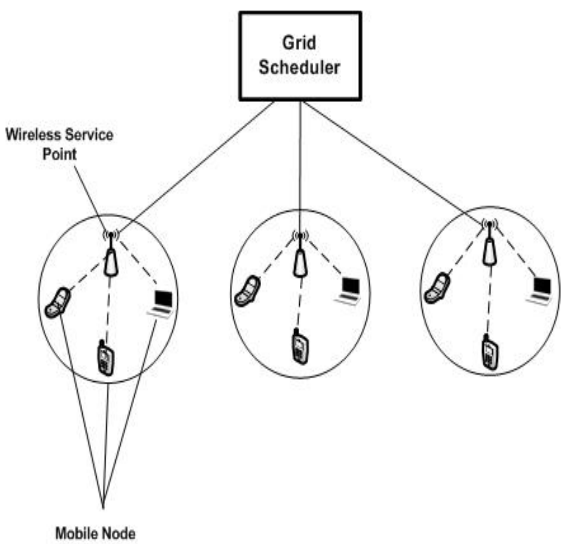

Fig. 3 Set of Mobile Nodes

d) Set of Site and Mobile Node

Sometimes computing nodes of site are busy so task execution gets delayed. In this scenario, mobile nodes can be used in combination with sites. If the sites are busy then tasks are passed to mobile nodes to computation. Fig. 4 shows site and mobile nodes used for grid model. [7] considers site plus mobile node in grid application for task scheduling.

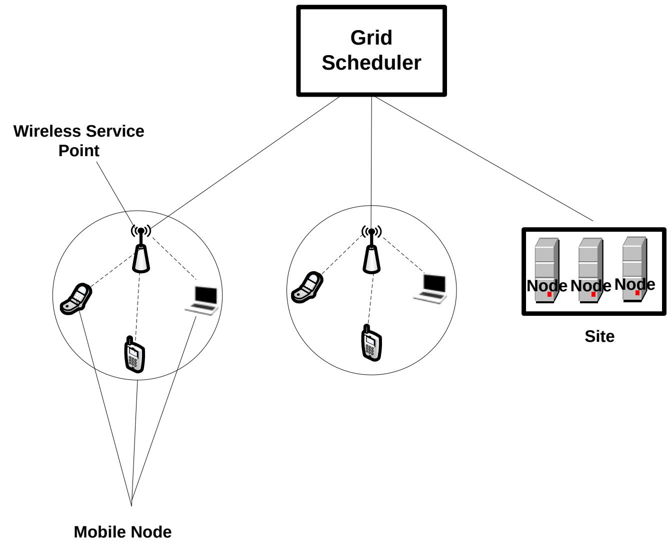

Fig. 4 Set of Mobile Nodes and Site

3. TASK STATES IN GRID COMPUTING

Task in grid application can be any one of the states as shown in Fig. 5. Basic fundamental states are:-

- User: - Tasks are generated by the user.
- Grid user interface: - User submit task through grid user interface.
- Grid scheduler: - Task arriving at a grid scheduler is placed in a global task queue and is served on the order of their arrival i.e. in first come first served (FCFS) manner. Grid scheduler checks workload and resource availability of each site and forward task to site that satisfy task requirement. Thus, it will decompose and distribute tasks to remote sites. If scheduler is busy or no sites satisfy task requirement then task is terminated and appropriate message is sent to user.
- Local scheduler: - It contains task which are assigned by grid. scheduler and are waiting to get share of resources. It is responsibility of local scheduler to schedule the task for execution which is assigned by grid scheduler.
- Executing: - Task which is actually using CPU cycles belongs to this state.
- Blocked: - Task for which required resource is currently not available is put in the blocked state.
- Finished: - Task after complete execution goes into this state. Result of the completed task returned to user through grid scheduler.

4. PERFORMENCE METRICS IN GRID COMPUTING

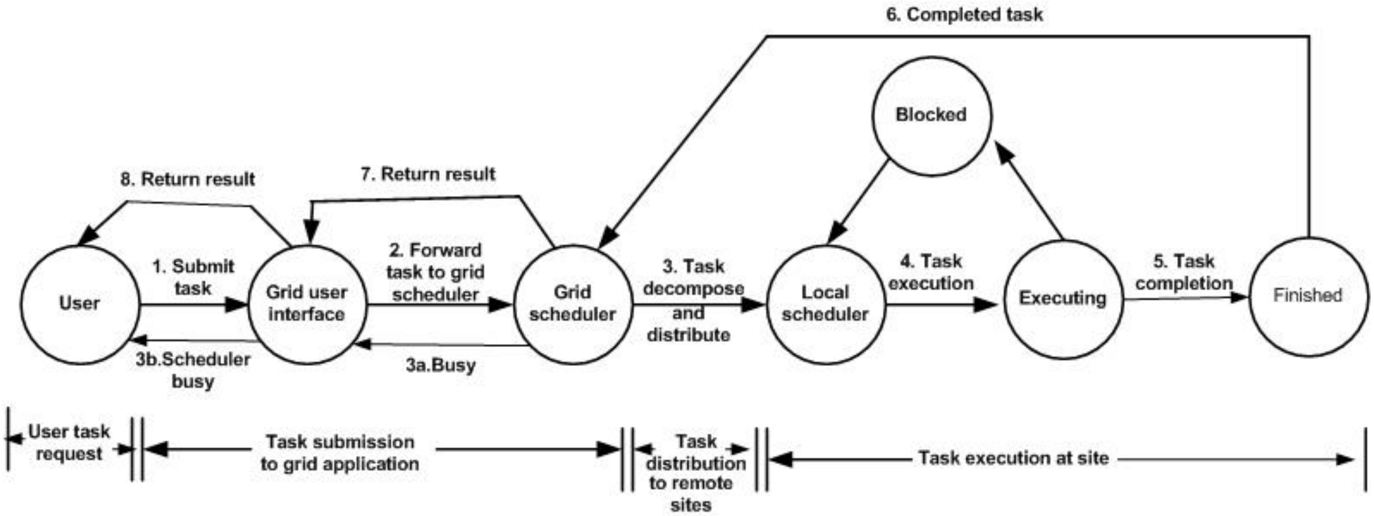

Fig. 5 Task States in Grid Computing

SI.	Parameter	Reference	Description
1.	Makespan	[8]	Makespan is maximum of completion time. makespan=max(CT_i)
2.	Resource Utilization	[9]	The resource utilization is defined as amount of resource is busy in executing tasks.
3.	Service Reliability	[10]	It is defined as a) Accessibility: - Service is available when desired i.e. when consumer want service. b) Continuity - Consumer has uninterrupted service over desired duration. c) Performance - Meets the consumer expectation
4.	Fairness Deviation	[9]	The fairness of the market means that each resource owner has an equal opportunity to offer its resource and it can obtain a fair profit according to its capability.
5.	Success Task Execution	[9]	Success Task Execution means task is executed within its deadline.
6.	Response Time	[10]	The response time of a task is the time interval between this task is arrive into the system until it is completed. Response time includes the waiting time in the queues and the service time.

Table 1 Performance Metrics in Grid Computing

Grid system performance has been measured by various performance metrics. The most commonly used performance metrics are i) Makespan ii) Resource Utilization iii) Service Reliability iv) Fairness Deviation v) Success Task Execution (vi) Response Time. Table I shows description of various performance metrics.

4.1 Uses of Performance Metrics

- Performance metrics helps in measuring significant work.
- Unmeasured work can be minimized or eliminated.
- Desired outcomes are necessary for work evaluation.
- It helps in timely corrective action.
- Work that is not measured or assessed cannot be managed because there is no objective information to determine its value.

5. CONCLUSION

In this paper, we have studied the task scheduling methodology and reviewed various types of task scheduling algorithm, nature of task and computing type of tasks is overviews. Various types of system model has been studied given in literature and grid system model has been presented. Different possible states of task and task execution cycle in grid computing are presented. We identified various performance metrics used in grid computing. Uses of performance metrics are discussed.

6. REFERENCES

[1] C. Stein D. Karger and J. Wein. Scheduling algorithm. in Algorithms and Theory of Computation Handbook. CRC Press, 1999.

[2] C.E. Leiserson Y. He, W.J. Hsu. Provably efficient online nonclairvoyant adaptive scheduling. IEEE Trans. Parallel Distrib. Syst., 19:1263-1279, 2008.

[3] Albert Y. Zomaya S. M. MeRiky Subrata and Bjorn Landfeldt. Cooperative power-aware scheduling in grid computing environments. Journal of Parallel and Distributed Computing,70: 84-91, 2010.

[4] Marek Wieczorek, Stefan Podlipnig, Radu Prodan, and Thomas Fahringer. Applying double auctions for scheduling of workows on the grid. In Proceedings of the 2008 ACM/IEEE conference on Supercomputing, pages 1 - 11, 2008.

[5] Mohammad K. Akbari Bahman Javadi and Jemal H. Abawajy. Analytical communication networks model for enterprise grid computing. Future Generation Computer Systems, 23:737 - 747, 2007.

[6] Preetam Ghosh and Sajal K. Mobility-aware cost-e_cient job scheduling for single-class grid jobs in a generic mobile grid architecture. Future Generation Computer Systems, 26:1356 - 1367, 2010.

[7] Chunlin Li and LaYuan Li. A multi-agent-based model for service-oriented interaction in a mobile grid computing environment. Pervasive and Mobile Computing, 7:270-284, 2011.

[8] M. Naghibzadeh Kobra Etminani. A min-min max-min selective algorihtm for grid task scheduling. In ICI 2007. 3rd IEEE/IFIP International Conference in Central Asia, 2007.

[9] Hesam Izakian, Ajith Abraham, and Behrouz Tork Ladani. An auction method for resource allocation in computational grids. Future Generation Computer Systems, 26(2):228 -235, 2010.

[10] Yuan-Shun Dai and Xiao-Long Wang. Optimal resource allocation on grid systems for maximizing service reliability using a genetic algorithm. Reliability Engineering and System Safety, 91(9):1071-1082, 2006.