# Optimization of Task Scheduling Algorithm in Distributed System: A Review

**Sonia Sharma**

*M.Tech Student, Dept. of CSE, KITM Kurukshetra, Haryana, India, sharmasonia246@gmail.com*

**Abstract:** *A Network Computing System is a virtual computer formed by a networked set of heterogeneous machines that agree to share their local resources with each other. A grid is a very large scale network computing system that scales to internet size environments with machines distributed across multiple organizations and administrative domains. There source management system is the central component of grid computing system. Resources in the grid are distributed, heterogeneous, autonomous and unpredictable. A resource management system matches requests to resources, schedules the matched resources, and executes the requests using scheduled resources. Scheduling in the grid environment depends upon the characteristics of the tasks, machines and network connectivity. The paper provides a brief overview of resource management in grid computing considering important factors such as types of resource management in grid computing, resource management models and comparison of various scheduling algorithm in resource management in grid computing.*

**Keywords:** *grid computing, resource management, scheduling algorithms.*

## 1. INTRODUCTION

Grid Computing is coordinated resource sharing and problem solving in dynamic, Multi institutional virtual organizations [1].Grid Computing is the flexible, secure, and coordinated sharing of resources among dynamic collections of individuals, institutions, and resources [2].The resource management is the central to the operation of a grid. The basic function of resource management is to accept requests for resources from machines with in the grid and assign specific machine resources to a request from the overall pool of grid resource for which the user has access permission. A resource management system matches requests to resources, schedules the matched resources, and executes the requests using the scheduled resources. Grid resources are the entities such as processor, disk space, memory provider and aims at a more automatic

and autonomous resource configuration through the introduction of a Conversion Factory [9]. Service Level Agreements with the aid of formalized business goals (Business Level Objectives), complexity analyses, and knowledge of previous configurations.

### Agent-Based Grid Resource Management

Agent-based approach for resource management [10] is flexible, robust, and scalable. Resource Managements match requests to available resources and arbitrate between requests with respect to Business Values. Each agent implements one or more allocation semantics, and advertises its capabilities to the Naming/Discovery service. Their behavior depends on the allocation semantic they implement and extends the basic behavior consisting in matching a requested resource profile with an available resource within its set of primary resources.

### An Enterprise-Based Grid Resource Management System:

An Enterprise based resource management system [11] is highly scalable resource management system built using COTS (Commercial Off-The-Shelf) components used in web server infrastructures. The web services and enterprise components constituted the core of the system that provides fault tolerance as well as high performance and scalability.

### Agreement-Based Resource Management

Agreement based resource management is on the basis of standards activities and next generation resource management services [12]. Agreements abstract local management policy by representing an underlying resource strictly in terms of policy terms which it is willing to assert, and in doing so provides the basis for building a variety of alternative grid resource management strategies. An agreement provides a powerful mechanism for virtualizing or abstracting a resource. A resource may have complex internal policy for deciding permission, relative priorities, and other scheduling and management procedures.

## Grid Resource Management Based On P2P Technology

A resource management approach in grid computing environment based upon P2P technology [13] can manage dynamic grid computing resources efficiently. Different kinds of grid computing resources are organized into a P2P overlay network and available information is published in type of grid service. With this approach, the computational resources of a grid system can be scheduled dynamically according to the real-time workload on each peer. The application of this approach is introduced into mapping of users to RMS's and resources to space, network bandwidth etc. that are managed by the resource management system. The grid resource scheduling process can be defined as the process of matching a query for resources, described in terms of required characteristics, to a set of resources that meet the expressed requirements. Job scheduling is the mapping of jobs to specific physical resources, trying to minimize some cost function specified by the user.

## 2. RELATED WORK

There have been some studies on survey and taxonomy of grid resource management. The most popular is done by Klaus Krauter and Raj Kumar Buyya[3], it described an abstract model and defined various concepts and terminologies for describing resource management architecture for grid computing. The work is also done by [4] for grid resource management system, it described issues and functions for GRMS and survey various grid resource management system. In 2010, Bo Wang, Gang Chen [5] presented various grid resource management models and context for grid resource management system. In 2010, [6] surveyed the various job scheduling and resource scheduling in grid computing. Scheduling experiments are performed with the help of GridSim simulator by setting the values to the number of jobs and processing time is recorded to analyze the feasibility of algorithm. In 2010, [7] surveyed and simulate various job scheduling strategies in grid computing environment. Also compare various jobs scheduling algorithm considering various parameters like type of scheduling model, type of resources that particular job scheduling algorithm favors(i.e. heterogeneous or homogenous), response time, load balancing, etc. In 2012, Isah Abdul Azeez and SafwanaHaque [8] have given a review for resource management in grid computing and described resource management concepts.

## 3. TYPES OF GRID RESOURCE MANAGEMENT

### SLA-Supported Resource Management

This architecture reflects the business needs of service DDGRID (Drug Discovery GRID), a grid system for drug discovery and design, to evaluate the performance.

### Failure-Aware Grid Resource Management System

Novel load-based failure recovery strategy used in the Virtual Resource Management framework [14]. The mechanism is applicable in any environment where distributed resources must be managed and failures of the system are critical, e.g. Service Level Agreements are given for the correct and complete execution of a job. Given approach adapts to the actual load situation and determines a remapping interval accordingly, which diminishes the danger of underestimating failure durations as any job is remapped before it is actually endangered of being terminated.

### A Scalable Wide-Area Grid Resource Management Framework

A scalable wide-area grid resource management framework [15] is a hierarchical framework that permits wide-area grid resource management. The framework uses a hierarchical organization of resource managers. The hierarchy is composed of multiple levels of resource managers, similar to business and government organizations. Resources are assigned to jobs through decentralized inter and intra organizational collaborations between resource managers.

### Semantically Enhanced Grid Resource Management

Semantic based Grid Resource Management(S-GRM) [16] system utilizes semantic metadata to describe and discover both logical and physical resources. The system integrates the component search into existing Grid application development environment and discovery of nodes in resource manager in a comprehensive perspective.

### HRMF-G: A Grid Based Hierarchical RM Framework

Hierarchical Resource Management framework [17] effectively resolved the bottleneck problem in the centralized resource management mode and non-reliability problem in the distributed resource

management mode, and also resolved the load imbalance problem and other problems in the centralized resource schedule mode and the distributed resource schedule mode.

## Performance Based Grid Resource Management

A Generalized Stochastic PetriNet (GSPN) model is for performance evaluation of grid [18]. In this model, three roles are defined: Users (U), Resource Management System (RMS), and RMS's are assumed to be both n to 1, which means that each users always chooses a preferable designated RMS to submit their tasks to, and each resource is directly managed by only one RMS.

## 4. RESOURCE MANAGEMENT MODELS

Model of resource management gives information regarding the relationship between users and resources provider, how resources are chosen for users, how resources are allocated. There are several different models for grid resource management system and these several different models may be used in the same system.

## A) Economic Model for GRM

It provides a decentralized resource management capability and is adaptable to changes in the environment and user requirements. It is scalable, controllable, measurable and easily understandable policy for management of resources. There are [19] several market-based economic models such as a commodity market, tenders and auctions along with the architecture and algorithms for their implementation in grid computing systems.

## B) Architectural Models for GRM

For the architecture of grid resource management three different models are [20]: Hierarchical Model, Abstract Owner Model and Market Model. The hierarchical model exhibits the approach followed in many contemporary grid systems. The abstract owner model follows an order and delivery approach in job submission and result gathering. The market/economy model follow economic model in resource discovery and scheduling that can coexist or work with contemporary systems and captures the essence of both hierarchical and abstract owner models.

## C) Pricing Model for GRM

This price model for grid resources [21] is based upon the theory of financial option. The use of option value computation is for grid resources usage and to select the best point of exercise of the option to utilize any of the grid resources. This helps the user as well as the grid resources provider to optimize resources for profitability. A price varying function which controls the price of resources and ensure the grid users get the maximum at best prices and the resources provider also make reasonable revenue at the current base price setting. A price variant function in price model is to adjust the charges for resources at various times so that the grid remains busy. This function helps the resource provider in keeping the grid busy and recovering the investment on the infrastructure in a Turnaround time is a task oriented characteristics Resources(R). All resource requests by users are to be submitted to some resource management system, instead of directly to resources. The predetermined period of time.

## D) Resource Management Model of Electric Power

System Based on Grid Technology. A resource management model of power system based on grid technology [22] can fully use computer hardware infrastructure and provide a favorable support for large scale computing environment in electric power system. This model consists of two independent parts: management node and computing node. As a center node, management node is responsible for the maintenance of grid users and resource information, safety management, job scheduling, the job monitor, task decomposition and transmission computing result collection and so on. Computing node receives computation tasks from the management node, and generates intermediate results.

## E) Resource Management Model Based On P2P Technology

Web services grid resource management and discovery model based on P2P technology [23] appliesthe semantic web technology and P2P technology to the grid resource search, improves the retrieving utilization, search speed and precision rate of grid resources and better solves the system expandability and reliability. This model is divided into two layers. The under layer is composed of Web Service Resource Providers (WSRP) nodes, called WSRP network. The upper layer is structural P2P networks, composed of Web Service Resource Register and Inquiry Broker (WSRRAIB) nodes with high reliability, long online time and less variation, called WSRRAIB network.

## 5. SCHEDULING

In grid environment, to schedule resources and jobs, certain scheduling components are required which decides which resources are assigned to which jobs. Assignment of resources to job can be explicitly and implicitly. Explicitly assignment can be specified by external rules base or programmable interface. Implicitly assignment can be implemented by choice of state estimation algorithms and rescheduling approaches.

### A) Scheduler

The function of the scheduler as task allocation or task mapping in distributed computing system. It should be noted that multiple tasks from various users are submitted on any nodes of the distributed computing system from time to time. The scheduler must also accommodate this multiplicity of the disjoint that considers the time duration between submission and completion of a task. Obviously, it will be desirable to minimize this time. Throughput of the system is a characteristic that measures the number of tasks successfully executed in unit time. This quantity must be maximized. Interactive response is about the requirement of interactive users during the execution of the tasks. Scheduler is responsible for resource discovery, resource trading, resource selection and job assignment.

### B) Scheduling Model

Scheduling models describes structure of resource management system [4] and scalability of system.Scheduling models can be classified ascentralized, hierarchical or decentralized.

### I. Centralized Scheduling Model

In centralized scheduling model, all jobs are submitted to a single scheduler who is responsible for scheduling them on the available resources. Since all the scheduling

### II. Decentralized Scheduling Model

In decentralized scheduling model, there is no central scheduler, scheduling is done by the resource requestor and owners independently. This approach is scalable and suits grid systems. But individual schedulers should cooperate with each other in making scheduling decisions and the schedule generated may not be the optimal schedule.

### III. Hierarchical Scheduling Model

In this model, the schedulers are organized in a hierarchy. High level resource entities are scheduled at higher levels and lower level smaller sub- entities are scheduled at lower levels of the scheduler hierarchy. This model is a combination of above two models.

### C) Scheduling Algorithms

i. Capacity Planning and Stochastic Scheduling in Large-Scale Grids (CPSS) Capacity Planning and Stochastic Scheduling information is available at one single position the scheduling decision are optimal but this approach is not very scalable in a grid system.[24] is very efficient and scalable approach that lead to increased resource utilization, lower cost per workflow execution and the ability to solve huge problem sizes. Capacity Planning and Stochastic Scheduling is used for large scale grid computing infrastructures that aims to minimize the cost of application execution while ensuring that Quality of Service constraints are satisfied with desired confidence levels. This Scheduling formulation is based on queuing theory to specify the stochastic nature of the grid. In modeling the grid as a queuing network make able to use more accurate .scheduling. This is especially important in terms of task. Characteristic parameters for any scheduler are: turnaround time, throughput and interactive response time. Forecasting mechanisms to determine workload and its behavior which improves our ability to negotiate for advance reservations and grid futures. In this scheduling, the job of the scheduler is to select a set of resources onto which to schedule the tasks and manage an application, coordinate the execution of the tasks and manage the data distributions and communication between the tasks. Here, problem formulation implicitly performs capacity planning in that it assigns just enough resources to be able to handle the total service workload, while minimizing the cost of resources

ii. Stochastic Model and Evolutionary Optimization Algorithm for Grid Scheduling (SMEOA) Stochastic Model and Evolutionary Optimization Algorithm for grid scheduling [25] address the issues of uncertainty or randomness of sources. Uncertainty of sources has major impact include machine breakdowns, unexpected releases of high priority jobs, network status and soon. A new stochastic model for grid scheduling and a novel evolutionary scheduling algorithm is based on MSE (Minimum Stochastic Expectation) model. A resource Use Pattern Analysis (UPA) method to predict resources and grid QoS policy bring resources to improve grid efficiency.

iii. Research on Novel Dynamic Resource Management and job scheduling in grid computing (RNDRM) This scheduling model is based on Heap Sort Tree (HST)

[26] for computing the available computational power of the nodes (resource) as well as whole grid system. Here the resource with largest available computational ability among the whole grid system is selected to be the root node of the HST and it is ready for the scheduler to submit a job. The algorithm design for job scheduling is well suitable for the complex grids environment and it is based on agents.

iv. Virtual Computing Grid using Resource Pooling (VCGRP)

The System is based on loosely coupled concept. Virtual Computing Grid means the system can choose a resource and allocate tasks to it. Here, it is a single point web based access known as Virtual Computing Grid Portal and the Virtual Computing Grid Monitor is a central resource manager for the System [27].

v. A Grid-distance Based Scheduling for Grid Resource Management (GDBS) establishing lasting and numerous relationships between consumer and providers, for it helps form more effective VOs (Virtual Organizations).In resource scheduling, choosing near and low-cost resources using grid-distance as a key criterion improves the performance of grid resource scheduling and has better characteristics in executing tasks, which achieves the aim of scheduling optimization. Therefore, the strategies based on the grid distance scheduling are:

1. Achieving the aim of choosing the physically closer resource and show the advantage of closer resource in term of completing time communication reliability and cost;

2. The closer resource is obviously more stable in communication;

3. Optimizing the whole grid resource scheduling based on grid distance scheduling so as to improve the capacity of grid in dealing with job in unit time.

vi. Resource-cost-based Multi-agent Systems Scheduling for Grid Resource Management (RCBMAS)

A multi-agent systems resource management method in Grid environment fully utilizes the agent's autonomy, heterogeneity and distribution etc. to cooperate jobs in the Grid environment. The method utilizes resource-cost-based Grid resource scheduling algorithm [29] to find the most appropriate resource for each Grid task. Resource-cost is a triple, represented as physical distance, bandwidth available of communication network and the use cost .The first is a constant and the others vary with both the communication load and the relationship between supply and requirement.

vii. Resource Discovery Algorithm Based on Small-World Cluster in Hierarchical Grid Computing Environment (RDA)

Small-world cluster into hierarchical grid in which intra-cluster adopts centralized management and cluster center nodes form small-world network. The architecture strikes a balance between high efficiency of total centralized management and good scalability of absolute distributed disposal. In the process of constructing small-world network, a new construction method on the basis of NW (Newman-Watts) model is short range contacts and long range contacts are represented by routing tables in a logic and dynamic way not by Manhattan distance. Hierarchical architecture based on small-world cluster and SWRD algorithm used in education resource grid have high search success rate and low query cost, where small world and resource discovery make perfect combination.There exists no bottleneck and Resource scheduling algorithm model based on grid distance [28], which can optimize scheduling by using close resources in the grid resource central database, and system is scalable and reliable. Discovering resources according to different situations is in fact another expression of distribution [33].

## 6. CONCLUSION

In this paper, a review of task scheduling system in grid computing is presented. Various types of task scheduling in grid computing and various models for task scheduling in grids computing are discussed. Various scheduling algorithm in grid computing have been analyzed. A comparison on various parameters like architecture type, environment type (Heterogeneous or Homogeneous), response time, load balancing, and resource utilization is done on different types of job scheduling.

## REFERENCES

[1] Ian Foster, "What is the Grid? A Three Point Checklist", July 20, 2012.

[2] Alan Blatecky, "Grid Computing", EDUCAUSE Center for Applied Research, Research Bulletin, Volume 2002, Issue 17, September 3, 2012.

[3] Klaus Krauter, Raj kumar Buyya, and Muthucumaru Maheswaran, "A Taxonomy and Survey of Grid Resource Management Systems",

[4] Chaitanya Kandagatla, "Survey and Taxonomy of Grid Resource Management Systems"

[5] Bo Wang, Gang Chen, Hao Luo, Weimin Zheng, "A Taxonomy and Survey of Resource Management in Grid", Sixth International Conference on Semantics, Knowledge and Grids, 2010

[6] Raksha Sharma, Vishnu Kant Soni, Manoj Kumar Mishra, PrachetBhuyan, "A Survey of Job Scheduling and Resource Management in Grid Computing", World Academy of Science, Engineering and Technology 64, 2010.

[7] Vishnu Kant Soni, Raksha Sharma, Manoj Kumar Mishra, "An Analysis of Various Job Scheduling Strategies in Grid Computing", 2nd international Conference on Signal Processing systems(ICSPS), 2010.

[8] Isah Abdul Azzez and SafwanaHaque, "Resource Management in Grid Computing", Greener Journal of Science and Technological Research, ISSN: 2276-7835, Vol. 2(1), pp.024-031, April, 2012.

[9] Peer Hasse lmeyer, Bastian Koller, Lutz Schubert, Philopp Wieder, "Towards SLA-supported Resource Management"

[10] A. Lenica, F.Ogel, F. Peshanski, J.-P.Briot, "Agentbased Grid Resource Management".