Efficient Resource Management with Increase in Global Grid Systems Quality-of –Service (QOS)

G. Michael, R. Kavitha

Received: 01 April 2018 ▪ Revised: 21 April 2018 ▪ Accepted: 03 June 2018

Abstract: Resource allocation is a challenging problem in grid computing. In this paper an algorithm for efficient resource management with increase in global grid systems Quality-of –Service (QoS) is proposed by combining the advantages of the existing algorithms Application-centric and System-centric. The problem is formulated into constrained optimization problem. Application layer and fabric layer is communicated to increase the grid system’s efficiency. Comparison is performed to show the efficiency of the proposed joint application and fabric layer algorithm against the Application and system-centric algorithm.

Keywords: Grid Computing, Grid Resource Management, Scheduling.

INTRODUCTION

Resource management includes monitoring, identification of available resources, allocation of resource to jobs and scheduling. To support a variety of applications, a grid system should be capable of managing the resources. Since grid resources are dynamic, autonomous and heterogeneous in nature it is difficult to manage the physical and logical resources. Each resource has their owner called as resource providers who charge for using their resources. And the resource consumers are who uses the resources by paying the cost. Both resource providers and users have different goals. Resource provider tries to improve its benefit by increasing the cost of the resource. On the other hand, Resource users concentrate on getting the resource in a limited time and budget. Methods for upholding strict resource scheduling has to be designed to allow grid to provide resources based in the users request and the resource providers preferences. As economic models are introduced in Grid computing, the cost that an application needs to pay for utilizing the resource becomes a concern for some of Grid users. Many applications use compound objective functions, for example, some want both shorter execution time and lower economic costs. Based on the objective function, the scheduling system is differentiated into two types: Application-centric scheduling and System-centric scheduling. Application-centric scheduling aims at improving the efficiency of each application. For example, the time taken by a job to complete its task. Whereas, the System-centric scheduling looks forward to improve the benefits of the resource provider such as, utilization of the resource, revenue of the provider. Since the objective function of both the algorithms are highly conflicting, there is a lack of global grid System Quality of service.

In this paper, Joint optimization algorithm is proposed which aims at improving the grid system QoS. The advantages of application-centric and system-centric are combined. Cross layer is introduced to provide a interlayer coupling between the application and fabric layer. Sum utility maximization problem is used. Exchange of information’s occurs between the application and fabric layer. Application layer is included in the joint optimization since the end to end quality noted by the user depends on the application. Fabric layer is also included because the major challenge of the nature of grid resource is that both the layers cooperate with each other. Application layer adjust the demand according to the availability of resource and the fabric layer allocates the resources requested by the user.

BACKGROUND

Li Chunlin, Li Layuan [1] has proposed a System-centric scheduling by combining Application-centric and Resource-centric scheduling policy. Utility functions were used to express grid user’s Quality of

G. Michael, Associate Professor, Department of Computer Science and Engineering, BIST, BIHER, Bharath Institute of Higher Education & Research, Selaiyur, Chennai. E. mail: michaelcse@gmail.com

R. Kavitha, Associate Professor, Department of Computer Science and Engineering, BIST, BIHER, Bharath Institute of Higher Education & Research, Selaiyur, Chennai.
Service requirement, resource provider's benefit function and system's objectives. In order to verify the efficiency of the proposed scheduling algorithm, the performance of application optimisation scheduling, resource optimisation scheduling, SOS were compared with a traditional Round-Robin algorithm. Performance of the proposed algorithm was analysed under different system parameters. There are six performance metrics used: resource utilization, deadline miss ratio, throughput ratio, allocation efficiency, payment ratio and user satisfaction. Daniel Colin Vanderster[2] have mentioned resource allocation and task scheduling strategies on computational grid. Number of techniques for resource allocation on the grid has been described such as, first come first served, backfilling. Choi et al [3] proposed a cross-layer optimization for streaming video delivery in wireless environment between application layer, data link layer, and physical layer. Chiang [4] proposed that network utility maximization problem can be used to design the overall communication network. And each layer defines the decomposed sub problem. He presents that in case of transport and a physical layer, link congestion prices is the better layering price. Rajkumar Buyya, Manzur Murshed, and David Abramson have proposed a computational economy framework for regulating the supply and demand for resources and allocating them for applications based on the users' quality of services requirements. The framework requires economy driven deadline and budget constrained (DBC) scheduling algorithms for allocating resources to application jobs in such a way that the users' requirements are met. A new scheduling algorithm, called DBC cost time Optimisation has been proposed, which extends the DBC cost optimisation algorithm to optimise for time, keeping the cost of computation at the minimum. DBC extends the cost-optimisation algorithm to optimise the time without incurring additional processing expenses. This is accomplished by applying the time-optimisation algorithm to schedule task farming or parameter-sweep application jobs on distributed resources having the same processing cost.

**JOINT OPTIMIZATION OF RESOURCE ALLOCATION**

**A. Formal Description**

Two main entities involved in grid computing are resource consumers, the one who submits the job and request for the resource and the resource providers, who provide the resources to the user according to the request and availability of the resource. Both the resource provider and the users have a varying objective function, which leads to inferior performance of the grid system. The aim of the global grid system is to maximize the sum utility. Joint application and fabric layer optimization servers to maximize the sum utility by allowing the application layer and the fabric layer to interact. Cross-layer is used to exchange needed information's between layers. The entire framework is decomposed into individual sub problems. Calculation of payment and the resource demand is done at the application layer and the allocation of resource is done in fabric layer. Each sub problems are solved independently. The grid system is said to be in optimum, if the user's chosen cost and the network's allocated rates are equal.

**B. System Design**

![Figure 1: Joint optimization and resource allocation](image_url)


C. Mathematical Model

A grid system with n grid resource provider and m grid users are considered. The set of resource users and providers are denoted as GU and GR. It is assumed that the resource provider GR, 1<j<n, and grid user GU, 1<i<m that uses the resources of the resource provider. Xij denotes the resource of grid resource provider j allotted to the grid user i. Utility for a grid user i is denoted as Ui. Each grid user has an individual utility function Ui. The purpose is to allocate grid resource optimally to the users in order to maximize the global utility.

\[
\text{Max } \sum_i U_i \\
\text{such that, } C_j \geq \sum_i X_{ij}, T_i \geq \sum_i t_i \\
E_i \geq \sum_j P_{ij} X_{ij} > 0
\]

There are some constraints used to maximize the utility of the grid system. The constraints are as follows. (1) The fraction resource unit already allotted should not be greater than the capacity of the resource provider; (2) Grid user should finish the job within time T_i and the budget E_i. Global grid optimization problem is solved by using lagrangian method. Global grid optimization is denoted as

\[
L(\lambda, \beta, \gamma, X_{ij}, P_{ij}) = \sum_i U_i - \lambda \left( \sum_j X_{ij} - C_j \right) - \beta \left( \sum_i t_i - T_i \right) - \gamma \left( \sum_j P_{ij} - E_i \right)
\]

Where \( \lambda, \beta, \gamma \) are the lagrangian multipliers.

Table 1: Notations

<table>
<thead>
<tr>
<th>Notations</th>
<th>Meanings</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>Number of resource provider</td>
</tr>
<tr>
<td>m</td>
<td>Number of user</td>
</tr>
<tr>
<td>i</td>
<td>ith User</td>
</tr>
<tr>
<td>j</td>
<td>jth Resource provider</td>
</tr>
<tr>
<td>C_j</td>
<td>Capacity of the resource provider j</td>
</tr>
<tr>
<td>E_i</td>
<td>Budget of the user i</td>
</tr>
<tr>
<td>P_{ij}</td>
<td>Payment of the grid user i to the resource provider j</td>
</tr>
<tr>
<td>X_{ij}</td>
<td>Resource allocation obtained by the user from resource provider</td>
</tr>
<tr>
<td>Pr_j</td>
<td>Unit price of the resource provider j</td>
</tr>
<tr>
<td>P_{ij}^*</td>
<td>Unique optimal solution at the application layer</td>
</tr>
</tbody>
</table>

There are two maximization problems namely, User Quality of service satisfaction problem at the application layer and resource allocation problem at the fabric layer.

\[
AF = \text{Max} \{ \left( E_i - \sum_j P_{ij} \right) + (T_i - \sum q_{in}/X_{ij}) \} \\
\text{for } n = 1 \text{ to } N
\]

\[
RF = \text{Max} \sum_j P_{ij} (\log X_{ij}+1)
\]

AF is conducted in the Application layer and the problem RF is solved in the fabric layer with the constraints Ti ≥ \( \sum q_{in} \) and C_j ≥ \( \sum X_{ij}q_{in} \) is the total resource requirement of the user m’s nth job. Log function is used since it increases as the resource allocation increase. Resource provider aims to increase P_{ij} (log X_{ij}+1) to increase its revenue. The variable \( \lambda \) controls the interaction between the layers and coordinates the used demand and the resource supply at the fabric layer.

ALGORITHM AND DESCRIPTION

Joint application-fabric layer algorithm aims at improving the global grid utility. In the application layer a unique payment to the resource and the new demand is calculated. In the fabric layer unique optimal resource allocation is done. The entire problem is divided into two sub problems.

//Grid user at application layer
Receives from the resource provider j the price pr_j^n;

\[
P_{ij}^* = \text{Max} \{ AP \left( P_{ij} \right) \} \\
\text{If } E_i \geq \sum P_{ij}
\]

Then X_{ij}(n+1) = P_{ij}^*/Pr_j^n;

Return to resource provider j ;
Else Return Null;


//Resource provider at fabric layer
Receives $X_{ij}^n$ from Grid user $i$

If $C_j \geq \sum X_{ij}$

Then $Pr_j^{(n+1)} = \max \{\varepsilon, Pr_j^n + \eta(\sum X_{ij} - C_j)\}$

Return new price to all Grid users;
Else Return Null;

a. Application Layer SubProblem

Once the user at the application layer receives a price $pr_j^n$ from the resource provider $j$, the user calculates the unique optimal payment to the grid resource provider. Since Network Utility framework is used maximum of the cost is chosen as the optimal payment of the application layer.

$$P_{ij}^* = \max \{AP(P_{ij})\}$$

Joint application-fabric layer algorithm uses network utility maximization to maximize the global grid sum utility.

Then the budget of the user is compared with the total cost the user has allotted for the jobs.

If $E_i \geq \sum P_{ij}$

If it is greater than the budget, no new resource demand can be made. Else

$$X_{ij}^{(n+1)} = \frac{P_{ij}^*}{Pr_j^n}$$

The new resource demand is sent to the resource provider for allocation of resource.

b. Fabric Layer SubProblem

$X_{ij}^*$ is the unique optimal solution to the fabric layer which is calculated as

$$X_{ij}^* = \frac{P_i((C_i+n) / \sum P_j)^{-1}}{1}$$

$K=1$ to $n$

Once the grid resource provider $j$ receives the new demand $X_{ij}^n$ from the user $i$, it calculates whether the demand exceeds the capacity $C_j$

If this condition is satisfied then the new price for the resource is calculated to increase the revenue of the provider.

$$Pr_j^{(n+1)} = \max \{\varepsilon, Pr_j^n + \eta(\sum X_{ij} - C_j)\}$$

And finally returns the new price $Pr_j^{(n+1)}$ to all grid users. This process repeats for every iteration until no new demands are received by the resource provider.

SIMULATIONS

In simulation, three entities are created namely resource provider, grid user and the scheduler where the joint optimization is implemented.
Joint application-fabric layer algorithm is compared with the existing application-centric and system-centric scheduling algorithms to prove the performance of the proposed algorithm. Application-centric concentrates on improving the benefit of the user. It attempts to complete the job as early as possible under a budget constraint. System-centric scheduling aims at improving the revenue of the provider. Two metrics such as resource utilization and user utility are used to compare the performance of joint application-fabric layer algorithm with application-centric and system-centric scheduling algorithms.

Under light load, the resource utilization of joint application-fabric layer scheduling algorithm is greater when compared to the application-centric and system-centric scheduling. For different load factor the resource utilization and the user utility varies. Lower the load factor higher the user utility.

CONCLUSION

Resource management is a challenging issue in grid computing. Improper management of resources leads to an inferior performance of grid system with poor QoS in terms of both resource provider and resource consumers. Most of the existing algorithms are either concerned about the user satisfaction or resource provider’s revenue, but not both. The proposed algorithm improves the global grid utility. Entire problem is divided into two sub problems and solved individually. Coordination between the demand and the resource allocation is managed by using a variable. In future joint optimization of other layers such as collective layer, connectivity layer and resource layer can be included.
REFERENCES


