# Processor Allocation in Grid using Binary Linear Programming

M.D. Samrajesh[1] Joice John[2]

ABSTRACT :

The essence of Grid Computing is federation, process acceleration and virtualization of computing resources; it creates a dynamic computing environment for sharing resources.

Grid resource management is the process of locating various types of resources, finding their capacity and their current status and allocating them and monitoring their state. One of the main challenges in Grid resource management is the optimal allocation of the various resources available in the grid system to the tasks submitted by users. An optimal grid allocation strategy will make the grid system work efficiently by completing the tasks effectively utilizing the available resources in the grid system to its optimum. This paper discusses about using **Binary Integer Linear Programming (Binary ILP)** technique for optimally assigning the requested tasks to various available resources in the grid system. The various tasks at time t are quantified as their sizes in bytes. The primary memory, in bytes of each resource in the grid system is considered as its capacity. The allocation of various tasks to various resources are optimally done considering the various parameters influencing the grid allocation viz Speed of the CPU, Average waiting time, Band width, Latency time, Data transfer rate, Physical storage devices etc...

[1,2] Asst. Professor,
Rajagiri School of Computer Science,
Rajagri, Kalamassery, Kochi-683104
E-mail : samrajesh@rajagiri.edu
E-mail : joice_john@yahoo.com

KEYWORDS : Grid Computing, Processor, Binary Integer Linear Programming, Grid, Task, Task Size, Capacity.

## 1. INTRODUCTION

Grid computing is becoming a popular way of providing high performance computing for many process intensive, scientific & business applications. The territory of grid computing is beyond parallel or distributed computing, requiring the management of a large number of heterogeneous resources with varying, distinct policies and controlled by multiple organizations. Grid computing allows a number of competitive and/or collaborative organizations to share mutual resources, including documents, software, computers, data and sensors, to seamlessly process data and computationally intensive applications [9].

Grid computing requires an effective allocator for the better utilization of the dynamic resources. The execution of user processes must simultaneously satisfy both job execution constraints and system usage policies. Although many scheduling techniques for various computing systems exist, traditional scheduling systems are inappropriate for scheduling tasks onto grid resources. First, grid resources are geographically distributed and heterogeneous in nature. One of the central concepts of a grid is that of a Virtual Organization (VO) , which is a group of consumers and providers united in their secure use of distributed high-end computational resources towards a common goal. Actual organizations, distributed nationwide or worldwide, participate in one or more VOs by sharing some or all of their resources. Second, these grid resources have decentralized ownership and different local scheduling policies dependent on their VO. Third, the dynamic load and availability of the resources require mechanisms for discovering and characterizing their status continually [9]. Although the above referenced systems address one or more of these characteristics, they do not address all of them in a cohesive manner for grids. This paper develops a framework for resource allocation in grid computing subject to a set of constraints. Here the allocation strategy can control the task assignment to grid resource.

Resource providers and task submitters who participate within a VO share resources by defining *how* resource usage takes place in terms of *where, what, who*, and *when* it is owed. Accordingly, the policies may be represented in a three dimensional space consisting of *resource provider, task submitter*, and various parameter of the grid resource. Currently, we further assume that all tasks have equal priority to use a resource.

### 1.1 Resource Provider

A resource provider is defined as an entity, which shares some particular physical resource within the context of a grid. Physical resources participating in a grid are often organized into hierarchical groups [6].

**Grid->Cluster->Network->Processor/ Resource**

A Typical grid consists of many Clusters/domains each containing one or many networks, which in turn contains machines. *Capacity vector* consists of the Processing capacity of a resource-measured primarily from the available physical main memory in the resource location.[8]

### 1.2 Task Submitter

A task submitter is defined as an entity, which consumes resources within the context of a grid. In general, a task for processing the task in any grid resources may be submitted by any user with a registration of participation. *Task vector* consists of size of the tasks submitted to the Grid allocator in bytes.

### 1.3 Parameters influencing the allocation of the Grid Resource

Since the resources in grid are heterogeneous specific information regarding each resource has a vital role in the allocation of the resource following are the identified parameters which has influence on the allocation.[4]

**Speed of the CPU( $\overline{Si}$ )**

In a Grid various Types of CPU with various capacity exists (Itanium, Xeon, Pentium, AMD, Motorola etc) .The general unit used to measure the speed of CPU is Hz, i.e., number of operations can be done per second, and here it is represented as Si the speed in million cycles per second of the computing resource [1].

**The Average Waiting Time (Wi)**

The average time a process waits for reallocation at the resource location due to contention with other local processes in a multiprogramming system is termed as Wi.

Wi(Xi)= Ti(Xi)- Xi where $T_i$- is the average Turn around time of resource i and Xi is the average service time at resource.[10]

**Bandwidth ($B_{ij}$)**

Since in a grid the resources and the tasks are at remote locations the network connectivity bandwidth plays an important role, Bandwidth is the amount of data that can be transmitted in a fixed amount of time. For digital devices, the bandwidth is usually expressed in bits per second (bps) or bytes per second. , Here it's represented in Mbps of the network resource [2]

**Latency time (Lij)**

It is the time required for a signal to travel from one point on a network to another measured in seconds.[5]

**Data Transfer rate ($X_i$)**

The movement of information from one location to another, here its from a local disk drive to memory or from the memory to the local disk drive , measured in Mbytes/sec of Physical data storage at resource i.

**Physical Storage Devices (Di)**

When a remote process is scheduled some I/O operation may be required to be performed so the free space of the Local Physical Storage Devices is also very much important, its measured in bytes [3].

**Cost of usage($C^s_i$)**

The cost measured in Indian Rupees (INR) /Sec of usage of the resource.

**Cost of Bandwidth ($C^b_{ij}$)**

The cost of Bandwidth in Rs/Mbps provided by the network resource

**Cost of Transfer ($C^d_i$)**

Cost of each Mbyte transferred to/from physical Data Storage in Rs/Mbyte

## 2. Grid Resource Allocation using IPP

At time $t$, $n$ tasks of tasks come to the allocator and $m$ resources available for allocation of which some of them may be idle.

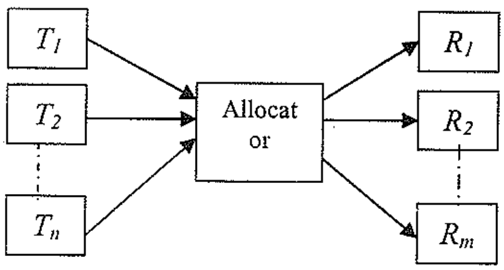


**Figure1-Task and the Resources**

$S_1, S_2, \ldots S_n$ be the sizes of the requested tasks made by $T_1, T_2, \ldots T_n$ in bytes.

$C_1, C_2, \ldots C_m$ be the capacites of resourses $R_1, R_2, \ldots R_m$ in bytes

At time $t$ if one machine say $R_k$ is not idle then the coreesponding capacity $C_k$ will be 0.

$a_{ij}$ can be a defined as coefficients which can have a user defined proposition of weights based on the parameters $\overline{S}i$ , $W_i$, $B_{ij}$, $L_{ij}$, $X_i$, $D_i$, $C^s_i$, $C^b_{ij}$, and $C^d_i$ so that allocations are made optimally according the grid allocation policy. A grid allocation policy should be formulated based on the above-mentioned parameters and their weightage for which resource allocation should be optimized.

The above defined can be represented in a matrix as follows.

| Task / Resource | $T_1$ | $T_2$ | ... | $T_n$ | Capacity |
|---|---|---|---|---|---|
| $R_1$ | $a_{11}$ | $a_{12}$ | ... | $a_{1n}$ | $C_1$ |
| $R_2$ | $a_{21}$ | $a_{22}$ | ... | $a_{2n}$ | $C_2$ |
| $R_3$ | ... | ... | ... | ... | $C_3$ |
| ... | .. | ... | ... | ... | ... |
| $R_m$ | $a_{m1}$ | $a_{m2}$ | ... | $a_{mn}$ | Cm |
| Task Size | $S_1$ | $S_2$ | .. | $S_n$ | |

**Table1 – Allocation Matrix**

The decision variable for the allocation problem can be defined as

$x_{ij}$ = 1 if $i^{th}$ resourse $R_i$ is given $j^{th}$ task $T_j$

= 0 otherwise.

for $i \in \{1,2,3....m\}$ , $j \in \{1,2....n\}$

Then the Binary Integer Linear Programming problem can be used to make the optimal allocation.[7]

Maximise $\sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij}a_{ij}$

Subject to the constraints

$\sum_{i=1}^{m} x_{ij} \leq 1$ $\forall j$ (one task will be given to atmost one resource)

$\sum_{j=1}^{n} x_{ij} \leq 1$ $\forall i$ (one resourse will get atmost one task.)

$\sum_{j=1}^{n} x_{ij}S_j \leq C_i$ $\forall i$ (all resources will be allocated only up to its capacity)

$x_{ij} \in \{0,1\}$

*Corollary 1*

Every allocation will be given equal weightage ie. $a_{ij} = 1$ $\forall i, j$

*Corollary 2*

Allocations are made so as to minimise the total capacity wastetage in the utelised resourses.

$$a_{ij} = Max(b_{ij}) - b_{ij} \quad \forall b_{ij} \geq 0 \text{ where}$$
$$b_{ij} = C_i - S_j \quad \forall i, j$$

### 2.1 Experimental Results

Let a grid sysytem consists of 4 resourses $R_1, R_2, R_3, R_4$ having capacites 20, 40 30 and15 megabytes repectivily, of which $M_4$ is not idle. 2 requests for performance of the tasks $T_1, T_2$ of capacites 25 and 18 mega bytes are made. These tasks are to be allocated to resourses so as to minimise the total capacity wastage in utelised reourses.

Then the $a_{ij}$ values using corollary 2 can be obtained as

| Tasks / Resourses | $T_1$ | $T_2$ | **Capacity** |
|---|---|---|---|
| **$R_1$** | -5 | 20 | **20** |
| **$R_2$** | 7 | 0 | **40** |
| **$R_3$** | 17 | 10 | **30** |
| **$R_4$** | -25 | -18 | **0** |
| **Task Size** | **25** | **18** | |

**Table 2- Allocation Table for Experimental Data**

Making the optimal allocation using IPP , the solution is $x_{12} = 1$ and $x_{31} = 1$. This implies that $T_1$ should be given to $R_3$ and $T_2$ to $R_1$, with minimum total capacity wastage in the utelised resourses as 7 megabytes.

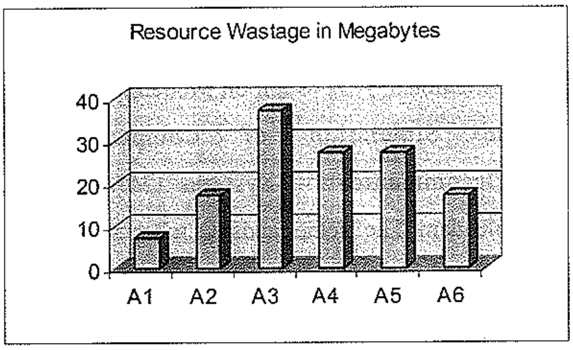


**Figure2-Comparism Graph**

A1 repersents the capacity wastage when allocation of resourses are done using Binary ILP. A2 to A6 are the capacity wasteage for other posible allocation of resourses.

## 3. Conclusion

Initial experimental results demonstrate the usefulness of this strategy. The efficiency of this allocation strategy depends on the efficiency of the allocation policy set. Estimation of appropriate weights to the decision parameters in $a_{ij}$ would provide the most efficient allocation.

## 4. References


[1] Albeaus Bayucan, *"Globus-enabled PBS: The PBS-Globus Interface"*, Globus retreat 2000, Pittsburgh, PA, August 2000

[2] Baker, M., Buyya, R., & Laforenza, D. *Grids and "Grid technologies for wide area distributed computing". Software Practice and Experience* 32, 15 (2002), 1437–1466.

[3] Buyya, R., Abramson, D., Giddy, J., And Stockinger, H. *"Economic models for resource management and scheduling in grid computing"*. The Journal of Concurrency and Computation: Practice and Experience 14, 13-15 (2002), 1507–1542.

[4] D. Menascé and E. Casalicchio, *"A Framework for Resource Allocation in Grid Computing"*, Proc. 12th Annual Meeting of the IEEE/ACM International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS), Volendam, The Netherlands, October 5-7, 2004.

[5] D. Menascé, *"Mapping Service Level Agreements in Distributed Applications"*, IEEE Internet Computing, Vol.8, No. 5, September/October 2004.

[6] Dumitrescu, C., Foster, I., Wilde, M., *Policy-based "Resource Allocation for Virtual Organizations*, http://www.griphyn.org/ documents document_ server/uploaded_documents/doc—683—vopolicy.v16.pdf

[7] Handy A,Taha, *"Operations Research an Introduction"*, 3rd Ed., Pearson Education, 2002

[8] Milan Milenkovic, *Operating System Concepts*, 2nd Ed., Tata McGraw Hill, 1997

[9] Paul Avery, Ian Foster, *"The GriPhyN Project: Towards Petascale Virtual-Data Grids"*, GriPhyN Technical Report 2000, http://www.griphyn.org/ documents/document_server/ uploaded_documents/doc—501—proposal_all.pdf

[10] Silberschatz , Galvin, Gagne, *"Operating System Concepts"*, 6the Ed., John Wiley & sons Inc., 2003


**Author's Biography :**


***Mr. Sam Rajesh*** M D has a Masters degree in Philosophy in Computer Science from MS University , He has around 7 years of teaching/Industrial experience , He has presented 2 papers in Grid Computing in National Seminars, also written 2 articles in Science magazines. He is currently the Secretary of Computer Society of India (CSI)- Cochin chapter.

***Mr. Joice John*** holds a Masters degree in Technology in Engineering Statistics form Cochin University of Science and Technology. He also holds a masters degree in Statistics. He has 3 years of teaching/industrial experience He completed one research project funded by UNDP and was the statistical consultant for a project funded by ADB. He has one national publication and presented papers/case studies in 2 national and 1 state level seminar.