大阪産業大学論集　自然科学編　第119号　2008

Modeling for the Optimal Co-scheduling Problem of Data Replication and Job Execution in Data Grids

Akiko Nakaniwa*, Srikumar Venugopal** and Rajkumar Buyya**

* Osaka Sangyo University, Japan

akiko@ise.osaka-sandai.ac.jp

** The University of Melbourne, Australia

srikumar@csse.unimelb.edu.au

raj@csse.unimelb.edu.au

Abstract

There are many challenges in Data Grids, and especially the data replication and the job scheduling are significant problems. There have been many works on the data replication and the job scheduling in Data Grids separately. However, there are still only a few works on solving those two significant problems together in Data Grids.

In this work, we propose the optimization model for the co-scheduling problem of the data replication and the job execution in Data Grids, which is more realistic and widely adaptable to real systems. We use 0-1 integer programming model to formulate the co-scheduling problem of the data replication and the job scheduling. Our final goal in this work is to find the optimal solution, that is, the data replication and job assignment, such that it minimizes the response time. The proposed optimization model will lead us this goal.

1. Introduction

Data Grids primarily deal with providing services and infrastructure for distributed data-intensive applications that need to access, transfer, and modify massive datasets stored in distributed storage resources. There are many challenges in Data Grids, and especially the data replication and the job scheduling are significant problems. How to replicate data files in distributed storage resources determines their quality of service, such as response time.

Received October 31, 2007

*Department of Information Systems Engineering, Osaka Sangyo University

**Computer Science and Software Engineering Department

To replicate date files closer to compute sites is one solution to decrease response time. However, due to the capacity of storage resources, it is impossible to replicate all massive datasets to all storage resources. Therefore, we need to figure out how we can best allocate replication of data files to distributed storage resources. In addition, it is important to consider the data replication together with the job scheduling, since we need to know where (which compute sites) to execute jobs in order to exactly know where to replicate data files.

There have been many works on the data replication and the job scheduling in Data Grids separately. However, there are still only a few works on solving those two significant problems together in Data Grids. Even in those works, the assumptions are very simple and it is hard to apply to the real Data Grid systems.

In this paper, we propose the optimization model for the co-scheduling problem of the data replication and the job execution in Data Grids, which is more realistic and widely adaptable to real systems. We use 0-1 integer programming model to formulate the co-scheduling problem of the data replication and the job scheduling. Two kinds of 0-1 variables are defined, which are on the data file allocation to storage resources and the job assignment to compute sites, respectively. We formulate the response time using these 0-1 variables. The response time consists of the data file transfer time, waiting time and job execution time.

Our final goal in this work is to find the optimal solution, that is, the data replication and job assignment, such that it minimize the response time. The proposed optimization model will lead us this goal.

2. Related Works

The data replica allocation and the task allocation in distributed systems have been worked for many years [1-7]. Recently, there also have been many works on the file allocation and the task allocation in Data Grids [8][9][10]. In these works, the file allocation and the task (job) allocation have been considered separately. However, in Data Grids, it is important to consider the file allocation and the task allocation simultaneously.

In the work [11], they challenge the co-scheduling problem of the data replication and job scheduling. In this work, three kind of schedulers, i.e. External Scheduler and Local Scheduler for the job scheduling and Dataset Scheduler for the data scheduling have been proposed. For Local Scheduler, they use FIFO (First In First Out) and they define five different algorithm for External Scheduler and four algorithms for Dataset Scheduler. In total, they have 20 algorithm combinations to evaluate by simulation experiments. However, the assumptions in these algorithms are very simple and the simulation results are difficult to apply to the real Data Grid systems.

3. Optimization Model

In this paper, we propose the optimization model for the co-scheduling problem of the data replication and the job execution in Data Grids, which is more realistic and widely adaptable to real systems. We use 0-1 integer programming model to formulate the co-scheduling problem of the data replication and the job scheduling.

3.1 System Model

We introduce a system model of Data Grids(Fig.1).

There exist C computing hosts and each computing host is denoted by $c_i\left(1 \leq i \leq C\right)$. Also, there exist D data hosts and each data host is denoted by $d_i\left(1 \leq i \leq D\right)$. Computing hosts and data hosts can be on the same machines. The service rate of a computing host of c_i is defined as μ_i. The capacity of a data host d_i is defined as β_i.

The number of distinct data files is F and each file is denoted by $f_i\left(1 \leq i \leq F\right)$. The size of a file f_i is defined as s_i. Also, the number of jobs submitted to computing hosts is J and each job is denoted by $j_i\left(1 \leq i \leq J\right)$. The arrival rate of a job j_i is defined as λ_i.

Each job needs one or multiple data files. Here, we define the identifier of the combination of jobs and files Z_{ik} as follows.

$$Z_{ik} = \begin{cases} 1 & (\text{the job } j_i \text{ needs file } f_k) \\ 0 & (\text{otherwise}) \end{cases}$$

Moreover, the communication rate (bandwidth) between hosts is defined as $b\left(c_i,\ d_j\right)$. We assume that if a computing host c_i and a data host d_j is in the same machine, $b\left(c_i,\ d_j\right)$, is equal to 0. We define the execution length of a job j_i on a computing host c_j as e_{ij}.

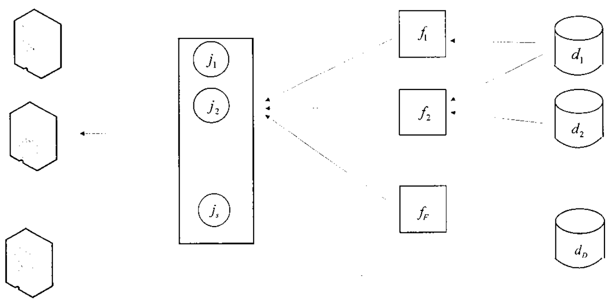

Fig.1 System Model of Data Grids

3.2 Definition of 0-1 Variables

Since we propose the optimization model for the co-scheduling problem of the data replication and the job execution in Data Grids by 0-1 integer programming model, we define two kinds of 0-1 variables, that is, on the file allocation (data replication) and the job assignment (job scheduling).

1. File allocation

$$X_{ik} = \begin{cases} 1 & \text{(the file } f_k \text{ exists on the data host} d_i) \\ 0 & \text{(otherwise)} \end{cases}$$

2. Job assignment

$$Y_{ik} = \begin{cases} 1 & \text{(the job } j_k \text{ is scheduled on the computing host } c_i) \\ 0 & \text{(otherwise)} \end{cases}$$

3.3 Formulation

In this optimization model, we aim to minimize the response time, that is, the time from submitting jobs to finishing jobs.

1. Initial Condition

As the initial condition, the following three equations have to be satisfied.

1) File allocation

All distinct files have to be stored at least one data host.

$$\sum_k X_{ik} \geq 1$$

2) Job assignment

All jobs have to be scheduled on at least one computing host.

$$\sum_k Y_{ik} \geq 1$$

3) Combination of jobs and files

All jobs have to need at least one data file.

$$\sum_i Z_{ik} \geq 1$$

2. Objective Function

We formulate the response time T for the job j_i as the objective function. Here, we assume the response time consists of the waiting time T_w, the file transfer time T_t and the job execution time T_e.

$$T = T_w + T_t + T_e$$

We formulate the waiting time T_w, the file transfer time T_t and the job execution time T_e by using pre-defined 0-1 variables.

The waiting time T_w can be formulated by applying the famous queueing theory [12]. We may assume this waiting time behavior as M/G/1, and here we take M/D/1 for analytical simplicity. We can formulate the waiting time T_w by the Pollaczek-Khinchin mean-value formula and Little's result.

$$T_w = \sum_j Y_{ji} \frac{\lambda_i / \mu_j^2}{2(1 - \lambda_i / \mu_j)}$$

The file transfer time is to transfer data files from data hosts to computing hosts where a job is executed. It becomes the maximum transfer time between data files that needs the job execution.

$$T_t = \max_k \left[\sum_j \sum_l \sum_i \left(X_{jk} Y_{li} Z_{ik} \frac{s_k}{b(c_l, d_j)} \right) \right]$$

The job execution time can be formulated simply by using the definition of the execution length of a job on a computing host.

$$T_e = \sum_j Y_{ji} e_{ij}$$

3. Constraint

There is the restriction to store data files in data hosts since data hosts have their capacity. This should be the constraint of this optimization problem.

$$\sum_k X_{ik} f_k \leq \beta_i$$

4. Considerations

In this paper, we have proposed the optimization model for the co-scheduling problem of the data replication and the job execution in Data Grids, which is more realistic and widely adaptable to real systems. We use 0-1 integer programming model for it.

Our final goal in this work is to find the optimal solution, that is, the data replication and job assignment, such that it minimize the response time. The proposed optimization model will lead us this goal. As future works, we will propose the heuristic approach and carry out simulation experiments to evaluate this optimization model. Moreover, we will develop the real Data Grid systems between Japan and Australia and carry out actual experiments.

References

[1] W. W.Chu, "Optimal file allocation in a multiple computer system", IEEE Transactions on Computers, Vol.C-18, No. 10, pp.885-890, October 1969.

[2] S. Ceri, G. Martella and G. Pelagatti, "Optimal File Allocation for a Distributed Database on a Network of Minicomputers", Proceedings of International Conference on Databases, July 1980.

[3] R. Blankinship, A.R.Hevner and S.B.Yao, "An Iterative Method for Distributed Database Optimization", Data & Knowledge Engineering 21, Elsevier Science, 1997.

[4] K. Kalpakis, K.Dasgupta and O.Wolfson, "Optimal Placement of Replicas in Trees with Read, Write, and Storage Costs", IEEE Transactions on Parallel and Distributed Systems, Vol.12, No.6, pp.628-637, June 2001.

[5] Wesley W. Chu and Lance M-T. Lan, "Task Allocation and Precedence Relations for Distributed Real-Time Systems", IEEE Transactions on Computers, Vol.C-36, No.6, pp.667-679, June 1987.

[6] Muthucumaru Maheswaran, Shoukat Ali, Howard Jay Siegel, Debra Hensgen and Richard F. Freund, "Dynamic Mapping of a Class of Independent Tasks onto Heterogeneous Computing Systems", Journal of Parallel and Distributed Computing, Vol. 59, pp.107-131, 1999.

[7] A. K.Verma, T.Tamhankar, "Reliability-Based Optimal Task-Allocation in Distributed-Database Management Systems," IEEE Transactions on Reliability, Vol.46, No.4, pp.452-459, December 1997.

[8] Rashdur M. Rahman, Ken Barker and Reda Alhajj, "Study of Different Replica Placement and Maintenance Strategies in Data Grid", 7th IEEE International Symposium on Cluster Computing and the Grid (CCGRID07), July 2007.

[9] Ali Elghirani, Riky Subrata and Albert Y. Zoyama, "Intelligent Scheduling and Repliation in Datagrids: a Synergetic Approach", 7th IEEE International Symposium on Cluster Computing and the Grid (CCGRID07), July 2007.

[10] Srikumar Venugopal and Rajkumar Buyya, Mapping Heuristics for Scheduling Distributed Data-Intensive Applications on Global Grids, Journal of Parallel and Distributed Computing, Elsevier Press, 2007 (in press, accepted on July 12, 2007).

[11] Kavitha Ranganathan and Ian Foster, "Simulation Studies of Computation and Data Scheduling Algorithms for Data Grids", Journal of Grid Computing No.1, pp.53-62, 2003.

[12] Leonard Kleinrock, "QUEUEING SYSTEMS", A Wiley-Interscience publication, 1975.