Asian Journal of Scientific Research

ISSN 1992-1454

Asian Journal of Scientific Research 6 (4): 789-796, 2013
ISSN 1992-1454 / DOI: 10.3923/ajsr.2013.789.796

Economic-based ACO Algorithm for Data Intensive Grid Scheduling

S. Aranganathan and K.M. Mehata
Department of Computer Science Engineering, BSA Crescent Engineering College, Chennai, India

Corresponding Author: S. Aranganathan, Department of Computer Science Engineering, BSA Crescent Engineering College, Chennai, Tamilnadu, India

ABSTRACT

The scope of grid computing is rapidly growing in distributed heterogeneous environments for the need to utilize and share large-scale resources to solve complex scientific problems. Economic models are effective in collaborating large-scale heterogeneous data and computational resources that are typically owned by different organizations with diverse interests. Scheduling is the most crucial task to achieve high performance in both computation and data grids. To utilize the grid efficiently for both resource providers and consumers, an efficient job scheduling algorithm is required. The proposed algorithm allows resource providers and consumers to take autonomous scheduling decisions and that both parties can derive sufficient incentives based on their economic interests. It is based on the general adaptive scheduling heuristic which employs a Quality of Service (QoS) guided component that emphasizes more on reliability. The algorithm was successfully tested in simulation environment. Experiments showed that the proposed economic and ant heuristic method was able to significantly improve performance by 10-25% even in unreliable network conditions.

Key words: Data-intensive scheduling, ant colony optimization algorithm, pheromone intensity

INTRODUCTION

The analysis of high-energy physics, molecular modeling and earth sciences datasets and their dissemination over a wide geographic area require high-capacity resources, such as supercomputers, high-bandwidth networks and mass storage systems (Aggarwal and Kent, 2005). Grid computing technology unites geographically distributed and heterogeneous computing, storage and network resources and enables pooling of resources to achieve the common goal.

The fact that grid computing can support both computation and data-intensive applications is widely acknowledged (Anjum *et al*., 2006; McCltchey *et al*., 2007). Based on the types of applications that grids support, they could be classified as computational and data grids. However, the efficient use of grids requires, chiefly, overcoming several challenges in security, resource management, scheduling and data management (Foster and Kesselman, 2003).

The users and resource providers who join a grid have different motivations. The objective functions may be either application-centric or system-centric. The conventional grid applications deal with two important parameters: makespan and cost. Makespan is time consumed from beginning of a job to the end of the last task in the job. Cost is the payment made for resource utilization. Because an end user may not wish to pay much price, they would like to negotiate price based on demand, value and available budget. In order to safeguard the interest of grid users as well as achieve fairness of resource sharing, a flexible, decentralized and economic scheduling is required (Haque *et al*., 2011).

A grid could be considered as similar to a commodity market which is decentralized, competitive and dynamic where consumers and providers have their own specific objectives. Due to this observation, the application of economic models is suggested to optimize resource management and to solve grid scheduling problems (Buyya *et al*., 2005).

Economic models differ from one another in terms of their strengths and weaknesses. They are evaluated based on such criteria as admission control, broadcasting overhead, decentralization, evaluation of market price, capability to handle a large number of users and economic efficiency. Economic models in grid scheduling involve several market behaviors, such as bargain, bid, auction and so on (Xiao *et al*., 2008). Yu *et al*. (2005) and Buyya *et al*. (2005) have described a few economic models, such as auction model, commodity market model and tender/contract-net model, for use in a grid scenario. They also experimentally evaluated these models in computational and data grid environments.

Haque *et al*. (2011) reviewed the English auction and double auction models. In the former model, the auctioneer anticipates to obtain the true market value for the resource being auctioned. Consumers are free to increase their bids for the resource they are competing. When no bidder is willing to increase the bid, the auction ends. The model is found to be suitable for increasing the revenues since it supports competition among users but it causes high communication overhead. This also helps to identify the demand for a particular resource(Xiao *et al*., 2008).

The ACO algorithm introduced by Dorigo *et al*. (1996) is based on the cooperative behavior of ant colonies. When blind insects like ants go in search for food, they lay a trail of pheromone along the path. As the number of ants following the trail grows, the attraction for that trail will increase (Xu *et al*., 2003). The main objective of proposed scheduling heuristic is to allow the resource consumers and resource providers to take autonomous scheduling decisions so that both parties get sufficient incentives based on their economic interest. With the adaptive QoS-guided component such as resource computation, resource communication and reputation of the resource, the algorithm emphasizes more on these components.

ECONOMIC-BASED QoS-GUIDED ANT ALGORITHM

A data-intensive grid S has several portals. Every participant in grid S is autonomous and plays individually in the decentralized scheduling framework. A service provider can join the grid via this portal. Also, a consumer can submit an announcement to grid S via the same portal. On receiving the job announcement, a provider may bid for the job. The true value of the provider can be evaluated through Job Success Rate (JSR). JSR is calculated dynamically based on service price stability, service reputation, service reliability and flexibility of the service provider.

Problem description: Data Intensive Grid S consisting of n resource providers is denoted by R = {r1,r2... .rn} and a set of m resource consumers by C = {c1,c2... .cm}. If consumers can submit up to k jobs/announcements, it is denoted by Y = {y1,y2...yk}. Each announcement includes deadline for a particular job, budget, QoS and JSR. The objective of the framework is to maximize consumer objectives, such as makespan and lesser payment and provider objectives, such as higher profit and parallel/efficient utilization of resources. The pattern of interactions between the resource provider and consumer is given in the following steps:

- **Step 1:** A consumer submits a job announcement to the grid, which is broadcasted to all the providers
- **Step 2:** On receipt of job announcement with JSR, each provider estimates whether it could be able to meet the deadline. If yes, the provider sends a bid containing the price for the job directly to the consumer; and if no, the provider ignores the job announcement
- **Step 3:** After receiving all the bids, the consumer chooses the provider who charges the least and sends the job
- **Step 4:** (a) To calculate the average price for resources:

$$P = \sum_{i=1}^{n} \frac{p_i}{n}$$

where, p_i denotes the price of ith resource.
(b) Adjust all the providers for above prices
- **Step 5:** A pool of resources is selected using the ACO heuristic method

ACO algorithm: The ants build their solution with both information encoded in the pheromone trail and problem-specific information in the form of a heuristic (Yan *et al.*, 2005).

Initialization of algorithm: All the pheromone values and parameters are initialized in the beginning of the algorithm.

Solution construction: N artificial ants are used in the algorithm. They set out to build N solutions to the problem based on pheromone and heuristic values using the selection rule.

Pheromone updating: After all ants complete their solution by the end of each iteration, the pheromone values are updated.

The following notations are used in the mathematical model:

T_i = Deadline given by user i
b_i = Budget of user i
P_j = Unit price of resource j
W_j = Total workload of resource j
C = Communication bandwidth
Tr_i = Time required completing the job at resource j
JSRj = Job success rate of resource j
m = number of CPUs
p = processing power (Micro Instruction Per Second)

Based on the equation given by Zhao *et al.* (2006), the following equation has constructed for calculating the innate performance of the resource:

$$\tau_j(0) = I1*(m*p)+I2*C+I3*\ JSRj \quad (1)$$

where, $\tau_j(0)$ is initial innate performance of resource j, I1, I2, I3 is intensive weightage factors.

The pheromone value $\tau_j(t)$ represents the favorability of scheduling a particular job i onto a particular resource j at time t.

The probability that the task is allocated to resource j within a job is computed using the formula:

$$P_j(t) = \frac{[\tau_j(t)]^\alpha [\tau_j(0)]^\beta}{\sum_\mu [\tau_\mu(t)]^\alpha][\tau_\mu(0)]^\beta} \quad (2)$$

where, $\tau_j(t)$ is pheromone intensity on the path from scheduler to resource j at time t, $\tau_j(0)$ is innate performance of resource j, α is importance of pheromone, $\boldsymbol{\beta}$ is resource-innate attribute, μ is resource available for the job.

The below formula is used to update the pheromone intensity on the path from schedule to corresponding resource:

$$\tau_j^{new} = \rho\tau_j^{new}+\Delta\tau_j \quad (3)$$

where, τ_j^{new} is the change of pheromone on path from the scheduler to resource j. ρ is evaporation of pheromone ($0\le\rho\le1$). When a task is allocated to resource j, $\Delta\tau_j$ = -K; K is quality of the resource the task consumed, which is calculated using Eq. 1. When a task is canceled and the resource is still in service, $\Delta\tau_j$ = K. This will restore resource quality. When a task successfully returns from resource j, $\Delta\tau_j = C_e \bullet K$, $C_e = 1$; C_e is the encouraging factor. Otherwise, if the task fails in returning from resource j, $\Delta\tau_j = Cp \bullet K$; C_p is the punishing factor and $0\le C_p\le1$.

The following steps are used to find optimal values for these parameters. These steps are added in the algorithm to adjust automatically based on the resource status. Environment variable B_j is used to make scheduling decisions according to the resource status and network performance. Note that the assignment of 0.9 or above to B_j should lead to a stable environment.

B_j is calculated as follows:

$$B_j = \frac{\frac{\text{Resource failure rate(\%)} + \text{Network stability(\%)}}{2}}{100}$$

Where:

$$\text{Network stability} = \frac{\text{Current bandwidth}}{\text{Re quired bandwidth}}$$

Procedure for adjusting the parameter:

```
Procedure ParaAdj
Begin
Compute Bj
If (Bj> = 0.9)
Assign β = 0.5, α = 1-β
Else If (Bj > = 0.8 and Bj<0.9)
Assign β = β+0.1, α = 1-β
Else If (Bj> = 0.6 and Bj<0.8)
Assign β = β+0.2,α = 1-β
Else If (Bj<0.6)
Assign β = β+0.3,α = 1-β
End if
End if
End if
End if
End ParaAdj
```

EXPERIMENTAL TESTING

When evaluated using GridSim (Sulistio *et al*., 2005) simulation toolkit, the algorithm was found to be working efficiently. In addition, a discrete event simulator interface was developed using Java to simulate the pattern of interaction between consumers and service providers, who were being modelled as two different entities in the simulation system. Diverse sets of jobs are submitted at different intervals to the grid evaluate the performance of the algorithm under various loads. Results of the experiments are shown here and a comparison is made with previous algorithms. The simulation parameters used to evaluate the algorithm are shown in Table 1.

Equation 2 takes both innate and real-time attributes of the resources into consideration when a task chooses resources appropriate to it. In this way, the selection can effectively avoid being influenced by the fluctuation of resource performance. A major advantage of this algorithm is that the parameters adjust automatically based on network behavior.

Pheromone is critical for the success of the algorithm. It could be set to a higher value in high-steady environments or a smaller value in other environments. The β determine the importance of heuristic information. Again, different values could be tested. A high β value is necessary to achieve a good solution in dynamic environment. Heuristic intensive weightage factors, parameters α and β and how their variations affect the performance of the algorithm are

Table 1: Simulation parameters and economic attributes of the experiments

Parameter	Value
Number of grid sites	150
Number of nodes in each site	30
Bandwidth capacity (Mbps to 2 Gbps)	200
Processing capacity (Micro Instruction Per Second)	512 to 1024
Total number of providers	300
Total number of users	800
Budget of users (units)	100-1000
Price of resources (units)	100-2000
Number of files	100-15000
Size of each file (MB-4 GB)	100

Table 2: Performance of proposed economic based algorithm in terms of cost

No. of jobs	No. of resources used	Adaptive QoS algorithm (Units)	Economic-based adaptive QoS algorithm (Units)	Improvement(%)
50	14	502	451	10.1
100	34	772	676	12.4
500	84	1196	999	16.5
1000	104	1486	1164	21.7
2000	134	1679	1249	25.6

Table 3: Performance of proposed economic-based algorithm in terms of revenue

No. of jobs	No. of resources used	Adaptive QoS algorithm (Units)	Economic-based adaptive QoS algorithm (Units)	Improvement(%)
50	14	16208	18153	12
100	34	21502	24942	16
500	84	25256	29802	18
1000	104	26208	31711	21
2000	134	28510	35638	25

Table 4: Performance of proposed economic-based algorithm in terms of resource utilization

No. of jobs	No. of resources used	Adaptive QoS algorithm (%)	Economic-based Adaptive QoS algorithm (%)	Improvement (%)
50	14	67	73	10
100	34	71	80	12
500	84	74	84	14
1000	104	78	89	15
2000	134	82	96	17

studied in detail. Their values are decided by the stability of the network and resource condition. Quality and consistency of the network are among the most important factors evaluated. By adjusting the parameters across trials, the performance of the algorithm appears to improve.

Intensive weightage factors, I1, I2 and I3, are added to Equation 1 for making good scheduling decisions and prioritizing the QoS parameters. The heuristic value is driven by these intensive weightage factors. Initially, I1, I2 and I3 were assigned 0.25, 0.5 and 0.25, respectively which yielded optimum results when tested under various scenarios. Results of the experiments and the improvement in performance of the economic model in terms of cost by 10-25 % are shown in Table 2. The economic costs of the resources are mentioned in unit price. When number of job is increased, the performance in terms of cost also increased comparatively with existing adaptive QoS algorithm. In the result of second experiment shown in Table 3 express the performance of proposed economic-based algorithm in terms of revenue by 12-25% comparatively with existing algorithm. The Table 4 expresses performance of proposed economic-based algorithm in terms of resource utilization by 10-17% with existing algorithm. The utilization of the resource mentioned in percentage.

The results of experiments and the improvement in performance of the economic model in terms of cost with existing algorithm are shown in Fig. 1. Figure 2 shows the performance of proposed economic-based algorithm in terms of revenue. The performance of proposed economic-based algorithm in terms of resource utilization is shown in Fig. 3.

Three metrics, namely economic cost, resource utilization and revenue for providers, are used to compare the performance of proposed economic algorithm (Aranganathan and Mehata, 2011a, b). From the graphs, the proposed scheduling algorithm seemed to considerably effective in reducing cost for consumers and maximizing resource utilization and revenue for resource

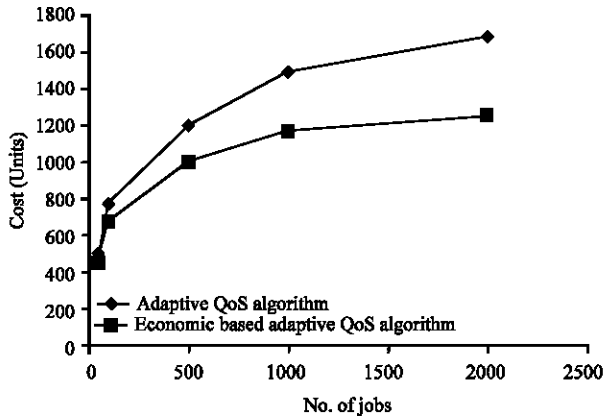

Fig. 1: Performance of proposed economic based algorithm in terms of cost

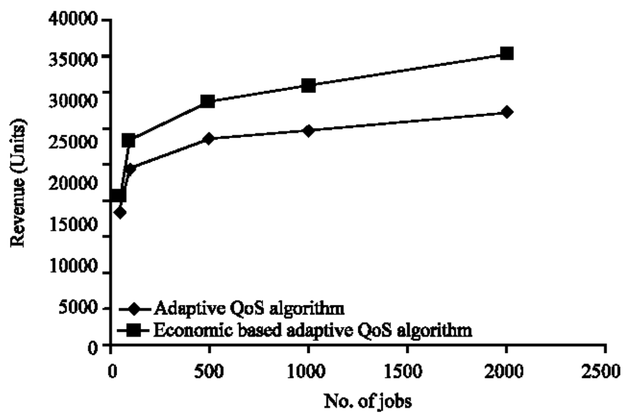

Fig. 2: Performance of proposed economic-based algorithm in terms of revenue

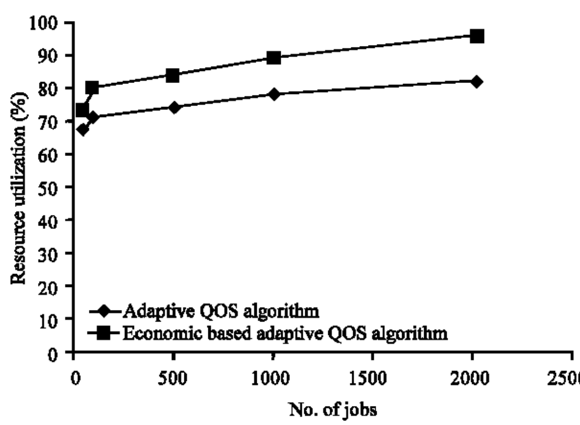

Fig. 3: Performance of proposed economic-based algorithm in terms of resource utilization

providers. For instance, the proposed algorithm manages to decrease cost by 10-25%. The parameters α, β and intensive weightage factors I1, I2 and I3 play a significant role. Experimental results showed that the algorithm is workable even under unreliable resource and network conditions.

CONCLUSION

A description of the proposed economy-based scheduling algorithm is presented in this study. Through extensive simulations, it is established that the proposed scheduling approach could significantly minimize the makespan and cost of grid user even under unreliable network conditions. Intensive weightage factors, the parameters of heuristic value and pheromone intensity are the key elements optimized by this scheduling algorithm. Further study needs to be carried out for the algorithm described here under different domain-dependent applications and newer models are expected to be generated.

REFERENCES

Aggarwal, A.K. and R.D. Kent, 2005. An adaptive generalized scheduler for grid applications. Proceedings of the 19th International Symposium on High Performance Computing Systems and Applications, May 15-18, 2005, Guelph, Ontario, Canada, pp: 188-194.

Anjum, A., R. McClatchey, A. Ali and I. Willers, 2006. Bulk scheduling with the DIANA scheduler. IEEE. Trans. Nuclear Sci., 53: 3818-3829.

Aranganathan, S. and K.M. Mehata, 2011a. Adaptive QoS guided ant algorithm for data intensive grid scheduling. Eur. J. Sci. Res., 58: 133-139.

Aranganathan, S. and K.M. Mehata, 2011b. An ACO algorithm for scheduling data intensive application with various QoS requirements. Int. J. Comput. Appl., Vol. 27, No. 10. 10.5120/3340-4598

Buyya, R., D. Abramson and S. Venugopal, 2005. The grid economy. Proc. IEEE, 93: 698-714.

Dorigo, M., V. Maniezzo and A. Colorni, 1996. Ant system: Optimization by a colony of cooperating agents. IEEE Trans. Syst. Man Cybern. Part B: Cybern., 26: 29-41.

Foster, I. and C. Kesselman, 2003. The Grid 2: Blueprint for a New Computing Infrastructure. Morgan Kaufmann.

Haque, A., S.M. Alhashmi and R. Parthiban, 2011. A survey of economic models in grid computing. Future Generation Comput. Syst., 27: 1056-1069.

McCltchey, R., A. Anjum, H. Stockinger, A. Ali, I. Willers and M. Thomas, 2007. Data Intensive and Network Aware (DIANA) grid scheduling. J. Grid Comput., 5: 43-64.

Sulistio, A., U. Cibej, B. Robic and R. Buyya, 2005. A toolkit for modelling and simulation of data grids with integration of data storage, replication and analysis. Technical Report No. GRIDS-TR-2005-13, Grid Computing and Distributed Systems Laboratory, University of Melbourne, Australia.

Xiao, L., Y. Zhu, L.M. Ni and Z. Xu, 2008. Incentive-based scheduling for market-like computational grids. IEEE Trans. Parallel Distributed Syst., 19: 903-913.

Xu, Z., X. Hou and J. Sun, 2003. Ant algorithm-based task scheduling in grid computing. Proceedings of the Canadian Conference on Electrical and Computer Engineering, Volume 2, May 4-7, 2003, Montreal, Canada, pp: 1107-1110.

Yan, H., X.Q. Shen, X. Li and M.H. Wu, 2005. An improved ant algorithm for job scheduling in grid computing. Proceeding of the International Conference of Machine Learning and Cybernetics, Volume 5, August 18-21, 2005, Guangzhou, China, pp: 2957-2961.

Yu, J., R. Buyya and C.K. Tham, 2005. QoS-based scheduling of workflow applications on service grids. Proceedings of the 1st IEEE International Conference on e-Science and Grid Computing, December 5-8, 2005, Melbourne, Australia, pp: 1-8.

Zhao, X., B. Wang, N. Du, C. Zhao and L. Xu, 2006. QoS-based algorithm for job allocation and scheduling in data grid. Proceedings of the 5th IEEE International Conference on Grid and Cooperative Computing Workshops, October 21-23, 2006, Hunan, China, pp: 20-26.