ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 17, March 2017

Scheduling of Relative Tasks in Grid Computing to provide QoS to Grid users

M.DHARANI. [1], M.MOHANAPRIYA [2]

1. P.G. Student, Dept. of MCA, VSB Engineering College, Karur, Tamilnadu, India
2. HoD, Dept. of MCA, VSB Engineering College, Karur, Tamilnadu, India

Abstract

Grid Computing is a new infrastructure of 21st century that provides access to computational and data intensive resources to its users. The core purpose behind the evolution of Grid computing is to utilize the idle capacity of the resources without acquiring the resources physically. The grid users are resource providers and resource consumers. The resource providers are looking for optimum utilization of their resources and profitability by lending their resources to grid users. The resource consumers have keen interest in processing their jobs within minimum cost and time. The FastPGA based proposed mechanism is multi-objective approach that optimizes total schedule length to minimize the overall execution time, penalty cost and job tardiness time to provide QoS to grid users. The proposed scheduling mechanism enables the grid users to execute their applications within provided budget and deadline. The users have to pay minimum penalty cost if the resource usage cost goes beyond the predetermined budget of the user. The task dependency among grid workflows is devised in directed acyclic graph(DAG) form.

1. INTRODUCTION

Grid is kind of large scale distributed computing of heterogeneous resources that belongs to different virtual organizations. The resources are shared, selected and aggregated on the basis of their availability, capability, usage cost and other user's QoS (quality of services) parameters. The computational grid has been evolved from the power grid where users can utilize power without knowing the place of the origin of power. Similarly, in computational grid, the end users can utilize the resources that are distributed geographically for execution of their applications. The resource management is done by Grid scheduler .The Grid scheduler is responsible for allocating appropriate resources to the user applications. The efficient scheduling of dependent jobs in a heterogeneous computing environment is a challenging problem in Grid Computing. Grid scheduling of dependent tasks can be defined as mapping of tasks on resources that preserves the dependency and fulfils the Grid user's QoS parameters. Several

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 17, March 2017

scheduling algorithms have been proposed by researchers to address the workflow scheduling issues in Grids. Grid Scheduling for dependent tasks or workflows require to map and manage successful execution of dependent tasks on Grid resources. Task and resource mapping in Grid environment is known as NP-hard problem. The workflow problem in this project is formulated in such a way that the proposed technique should take care of the inter-dependency of tasks and their execution order to optimize the entire schedule according to the defined objective functions. This project presents an extensive approach to schedule the dependent tasks in Grid environment using **Fast Pareto Genetic Algorithm** (FastPGA) based nature inspired algorithm.

2. LITERATURE SURVEY

There are various algorithm was developed in the field of grid computing. Some of the algorithms are:

1. Maria Arsuaga-Rios and Miguel A. Vega-Rodriguez presented a Multiobjective Brain Strom Algorithm (MOBSA) based on the brain storming in which humans processed the job in order to optimize the job scheduling problem in grid. MOBSA is based on two objectives: Execution time and execution cost.

2. Mansoure Yaghoobi et al. proposed a game theory based approach for minimizing the time and cost. The brokers act as a players and players compete to maximize the profits.

3. P. Mathiyalagan et al. proposed an intelligent water drop algorithm along with ACO algorithm Intelligent water drop is used to find out the resources according to job requirement and routing information.

4. Xi Li et al. defined a concept of Deadline Satisfaction Degree of workflow which is used to represent the probability of workflow to be completed before its deadline.

5. Sunita Bansal and Chittarajan Hota proposed an Efficient Algorithm on Heterogeneous Computing System (EAHCS) which manage the load across the machines and reduce the makespan.

6. Chaokun Yan et al. proposed a Reliability Enhanced Grid Workflow Scheduling Algorithm with budget constraint which can maximize the reliability.

7. Fabio Coutinho et.al defined an energy efficient model and HGreen heuristic which assign the heaviest workflow tasks to energy efficient resources.

8. Dengpan Yin and Tevfik Kosar proposed A-star based data-aware workflow scheduling algorithm.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 17, March 2017

Algorithms allow overlap of data placement and task execution and due to this turnaround time and time complexity are decreased. This algorithm extended to the co-scheduling problem.

9. Zhang Wenpeng and Liu Hongzhao proposed algorithm with combination of advantages of genetic algorithm, clonal algorithm and simulated annealing.

10. Wei-Neng Chen and Jun Zhang proposed Ant Colony Optimization algorithm for scheduling the tasks with various parameters. Seven new heuristic was developed for the ACO. Adaptive scheme also developed to enable artificial ants to one of heuristic on the basis pheromone values.

11. Long Hao et al. proposed a Deadlock Segment Leveling (DSL) a novel heuristic which divide the workflow application into segments and further segments are partitioned into groups.

12. Kassian Plankensteiner et al. proposed resubmission impact a fault tolerance heuristic for scientific workflows. Heuristic is based on task replication and task resubmission.

2.1 EXISTING SYSTEM

A GA based approach is developed to seek out time and cost constrained scheduling problem. The fitness function addresses budget and deadline constraints. For costfitness, the solution satisfies the budget-constraint and provides individuals with minimum cost and for the timefitness, the solution provides the individuals with minimum time. The LOSS and GAIN approach adjusts time and cost in an optimal manner that meets user's budget requirements and process the tasks in less time. If the schedule provides cost less than the pre-defined budget by the user then the schedule is considered as final assignment. If the cost of schedule goes beyond the pre-defined cost then LOSS mechanism is applied that minimizes the cost.

2.1.1 Disadvantages

- Waiting time can be large if short requests wait behind the long ones.
- It is not suitable for time sharing systems where it is important that each user should get the CPU for an equal amount of time interval.
- A proper mix of jobs is needed to achieve good results

2.2. PROPOSED SYSTEM

Grid scheduling for dependent tasks is a kind of NP-hard problem. Using FastPGA algorithm to solve the dependent task scheduling problem requires the suitable representation of individuals in the given population. In proposed work, four

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 17, March 2017

scheduling criteria are considered. The problem definition consists of the following:

- A task can start only after the completion of its predecessor.
- Many tasks are scheduled on one resource (cluster).
- Heterogeneous resources(clusters) are available to plan the schedule
- Each task appears one time only in the given schedule.
- Each task is scheduled on a resource at one time slot.
- The capacity of each resource.
- The expected execution time (EET) of each task is known in advance.

The aim of the proposed work is to minimize three scheduling criteria of dependent tasks workflow in grid environment.

2.2.1 Advantage of Proposed System:

- It supports dependent task workflow scheduling in grid environment.
- The results obtained from FastPGA based approach are very promising.
- This technique is to address multi-objective scheduling problem. T
- It minimizes the total schedule length, penalty cost that user has to bear if the job exceeds its budget given by grid user and the job tardiness that is excess time taken by a job after meeting its deadline time.

3. SYSTEM MODEL AND PROBLEM DEFINITION

Grid scheduling of dependent tasks can be defined as mapping of tasks on resources that preserves the dependency and fulfils the Grid user's QoS parameters. Several scheduling algorithms have been proposed by researchers to address the workflow scheduling issues in Grids. Grid Scheduling for dependent tasks or workflows require to map and manage successful execution of dependent tasks on Grid resources. Task and resource mapping in Grid environment is known as NP-hard problem. The workflow problem in this project is formulated in such a way that the proposed technique should take care of the inter-dependency of tasks and their execution order to optimize the entire schedule according to the defined objective functions. This project presents an extensive approach to schedule the dependent tasks in Grid environment using Fast Pareto Genetic Algorithm (FastPGA) based nature inspired algorithm. [6] proposed a secure hash message authentication code. A secure hash message authentication code to avoid certificate revocation list checking is

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 17, March 2017

proposed for vehicular ad hoc networks (VANETs). The group signature scheme is widely used in VANETs for secure communication, the existing systems based on group signature scheme provides verification delay in certificate revocation list checking. In order to overcome this delay this paper uses a Hash message authentication code (HMAC).

Job scheduling is based on the necessity of a user who has a set of jobs to execute. Since the user's machine is not able to process the jobs either because of resource or hardware constraints, the user can use the grid system for running the job. The user submits the set of jobs to the job scheduler and the job scheduler splits the job depending on certain factors and gives it to the machines having available resources on the grid. The machines will complete the task and final result will be given to the user. In order to achieve this job scheduling strategies has to be followed. Job scheduling and resource scheduling are the two main necessities in grid computing. In job scheduling, the job scheduler has to find the appropriate resource for the job that the user submits. It has to find the best machine in grid to process the user job. Grid has two main schedulers such as local schedulers and grid schedulers. The local schedulers work in local computational environment and hence it is reliably, fast connection, works in uniform environment and also takes full control of the homogeneous resources. Grid Schedulers also called as meta-schedulers are the top level schedulers. They are responsible for orchestrating resources that are managed by different local schedulers.

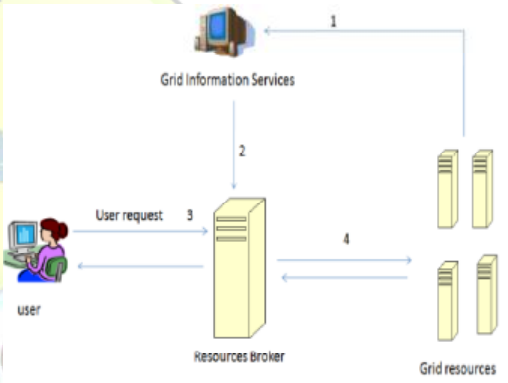

Fig.1 Architecture of the Grid System

3.1 OVERVIEW OF THE SYSTEM

Tasks are represented using DAG (Directed Acyclic Graph). It is defined as G = (V, E), where V is a set of v nodes/vertices and E is a set of e directed edges. Here nodes of DAG represent tasks and edges represent the dependencies between the tasks with directions. The source node of an edge is called the parent node while the destination node is called the child node. A node with no parent is called an entry node and a node with no

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 17, March 2017

child is called an exit. DAG starts with a root node and ends with an end node. In Fig. 2, node1 is root node and node 10 is end node. Fig. 2 also shows data which is transferred from the parent to their child. Using size of data and bandwidth between the machines it can calculate the communication cost. From Fig.2, it can be said that tasks 5 cannot start its execution until task 2 and task 6 completes their execution.

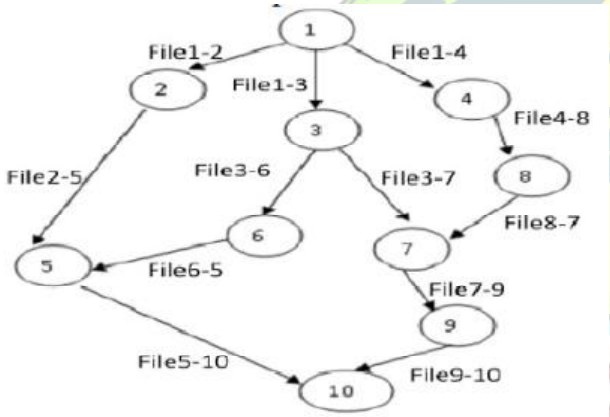

Fig.2 DAG with data transfer

Scheduling can also be classified into static and dynamic scheduling. In static scheduling, before execution the jobs are assigned to the suitable machines and those machines will continue executing those jobs without interruption. In dynamic scheduling, the rescheduling of jobs is allowed. The jobs executing can be migrated based on the dynamic information about the workload of the resources. In grid, there may be lots of resources to run a job. The main focus is to find the appropriate resource for the job that is to schedule the job. The methods for job scheduling are centralized, hierarchical and decentralized. In centralized scheduling, there will be a centralized scheduler and it is responsible for scheduling the jobs. It is very useful when all the resources have same objective. In hierarchical, there will be central scheduler. All jobs will be submitted to the central scheduler. The central scheduler redirects the jobs to the global scheduler. In decentralized, there is no central scheduler. Distributed schedulers coordinate with each other to schedule jobs.

The Fast PGA algorithm for job scheduling in grid aims at submitted jobs to resources based on the processing ability of jobs as well as the characteristics of the jobs. Fast PGA algorithm is the bio-inspired heuristic algorithm, which is derived from the social behavior of ants. Ants work together to find the shortest path between their nest and food source. When the ants move, each ant will deposit a chemical substance called pheromone. Using this pheromone, the shortest path is found. The same concept is used to assign jobs in grid computing. When a resource is assigning a job and completes, its

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 17, March 2017

pheromone value will be added each time. If a resource fails to finish a job, it will be punished by adding less pheromone value. The issue here is the stagnation, where there is a possibility of jobs being submitted to same resources having high pheromone value. The algorithm is as follows.

(i) The user will send request to process a job

(ii) The grid resource broker will find a resource for the job

(iii) The resource broker will select the resource based on the largest value in the pheromone value matrix

(iv) The local pheromone update is done when a job is assigned to a resource.

(v) The global pheromone update is done when a resource completes a job

(vi) The execution result will be sent to the user When the resource broker select a particular resource for a job j, jth column of the Pheromone Value matrix will be removed and jobs will be assigned to other resources. Thus the load balancing is achieved.

3.2 MODULE DESCRIPTION

3.2.1 GRID INFORMATION SERVICE

In grid environment, information about the available resources and about their attributes is sent to GIS (Grid Information Service).It manages and indexes all the resources available from multiple grid resource providers, and obtains information from each particular grid resource, acting as pricing interface for users, and updating the database when new information is available.

3.2.2 RESOURCE BROKER

Resource broker retrieves the information from GSI. In general, each grid manager registers its service resources through the grid resource broker. The service user negotiates with the service broker; According to the contract, the broker selects, and then presents highly trusted resources to users from the trusted resource pool. Resource broker will find the best suitable resource or resources for the requests and allocate it to the user.

3.3.3 USER REQUEST

The user request for the service to the broker with the policies so that client needs the specific services for production usability, reliability and security to the consumers. The broker can dynamically sort high-performance resources by analyzing the historic resource information in terms of providing highly trusted resources.

3.3.4 WORKFLOW SCHEDULING MODULE

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 17, March 2017

The scheduling process of workflow having 3 phases

Matching Phase: Resources are selected which satisfy the requirement of tasks. The minimal requirement of tasks defined in terms of static information like memory, architecture etc.

Scheduling Phase: In this phase, the resources are selected for the sequence of task by considering the constraints and rules imposed by the users. Heuristics are used for getting the optimal solution due to NP problem.

FastPGA is a scheduling algorithm in which scheduling is done according to the deadline of jobs and remaining time. The job with the earliest deadline will get resource first while the job with large value of deadline will have to wait irrespective of their execution time. In FastPGA scheduling algorithm first execution time along with the deadline of the job is entered. Then according to this scheduling algorithm, first six jobs are directly assigned to the processor P1, P2, P3, P4, P5 and P6 respectively. For the next jobs; processors are assigned according to the deadline of the jobs. If the current job on the processor has large value of deadline than the arriving job, then processor will stop executing the job and will execute the new coming job as its value of deadline is small. But if the current job has small deadline value then processor will keep executing the current job. Then in this case the next processor will be checked for the job and this process will continue until the job gets the processor for its execution. This process is repeated for each job. In the case if no processor satisfied this deadline condition then job will get the processor with shortest remaining time.

Execution Phase: The task is assigned to selected resources and is executed. Some management and administration are considered.

4. CONCLUSION

A new FastPGA approach implementation is coupled with .Net to support dependent task workflow scheduling in grid environment. The results obtained from FastPGA based approach are very promising. The FastPGA based approach adapts a Pareto-based multi-objective optimization technique to address multi-objective scheduling problem. The proposed approach minimizes the total schedule length, penalty cost that user has to bear if the job exceeds its budget given by grid user and the job tardiness that is excess time taken by a job after meeting its deadline time.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 17, March 2017

Algorithm research of task scheduling is one of the key techniques in grid computing. The system has analyzed scheduling problem that composed of numbers of dependent tasks. There is in-depth analysis for the typical task scheduling algorithm GS in heterogeneous systems and the algorithm considered the cost of communication. The proposed system is an improved Genetic based algorithm that is suitable for computing grid. The corresponding experiments show that improved algorithm can improve performance of execution.

FUTURE SCOPE

Grid computing can solve complex tasks in shorter time and utilizes the hardware efficiently. To make the grid work efficiently, best job scheduling strategies have to be employed. Job scheduling is the foremost step in grid computing where the users' jobs are scheduled to different machines.

- In scheduling, can be considered reliability, performance of grid resources.
- Explore new heuristics and introduce new execution scenarios. Try to add more heuristic for more improvement in the algorithm.
- GCAA can apply with different features. Implementation of method on real grid.

REFERENCES

[1] R. Buyya and S. Venugopal, "A Gentle Introduction to Grid Computing and Technologies," Computer Society India, no. July, pp. 9–19, 2005.

[2] I. Foster, C. Kesselman, and M. Rey, "Computational Grids Reasons for Computational Grids," 1998.

[3] J. M. Schopf, "Ten actions when Grid scheduling: The User as a Grid Scheduler," Grid Resource Management state of the art and Future trends, pp. 15–23, 2004.

[4] M. Bilal, Q. Maryam, M. Dehnavi, M. Shuaib, Q. Hameed, H. Ilias, R. Nikos, T. Loukopoulos, S. U. Khan, and C. X. Albert, "Survey on Grid Resource Allocation Mechanisms," Journal of Grid Compuing., vol. 9, 2014.

[5] H. Eskandari, C. D. Geiger, and G. B. Lamont, "FastPGA: A Dynamic Population Sizing Approach for Solving Expensive Multiobjective Optimization Problems," pp. 141–155, 2007.

[6] Christo Ananth, M.Danya Priyadharshini, "A Secure Hash Message Authentication Code to avoid Certificate Revocation list Checking in Vehicular Adhoc networks", International Journal of

Applied Engineering Research (IJAER), Volume 10, Special Issue 2, 2015,(1250-1254)

[7] R. Buyya and M. Murshed, "GridSim: a toolkit for the modeling and simulation of distributed resource management and scheduling for Grid computing," Concurrency and Computation: Practice and Experience, vol. 14, no. February, pp. 1175–1220, 2002.

[8] A. Intelligence, "Fundamentals of Genetic Algorithms Artificial Intelligence Fundamentals of Genetic Algorithms Artificial Intelligence," Thesis GA 42, no. x, 2010.

[9] J. Carretero, F. Xhafa, and A. Abraham, "Genetic Algorithm Based Schedulers for Grid Computing Systems," International Journal of Innovative Computing, Information and Control, vol. 3, no. 6, pp. 1–19, 2007.

[10] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," IEEE Transactions on Evolutionary Computation, vol. 6, no. 2, pp. 182–197, 2002.

[11]http://jmetal.sourceforge.net/ algorithms.html.