Comparison of Efficient Job Scheduling Mechanisms Used in Grid Computing: A Review

Bharti Arora[1], Sami Anand[2]
Student, Computer Science Department[1]
Asst.Prof. Computer Science Department[2]
Lovely Professional University, Jalandhar India

ABSTRACT

Availability of fast and reliable internet, high speed networks and powerful computers led to the emergence of a new paradigm known as Grid Computing. Grid computing is a form of distributed computing which enables the users to share the heterogeneousresources which are geographically dispersed to solve a complex problem.The heterogeneous nature and tremendous amount of resources makes Grid scheduling a challenging task.Heuristic scheduling is widely used in grid environment due to its advantage over time and cost.We have performed comparative study on Simulated Annealing algorithms (SA), Taboo search algorithms (TS) are categorized under Local search based .Ant Colony scheduling algorithms (ACS), Genetic algorithms (GA),Particle Swarm Optimization Algorithm (POS) are categorized under Population Based. Motivation of this memorandum is to encourage the amateur researcher in the field of grid computing, so they understand easily the concept of scheduling and can contribute in the development more scheduling algorithm. This study provides a comparative analysis on various scheduling algorithm potentials**.**

Keywords
Grid computing, job scheduling, resource scheduling, Algorithm, performance

1. INTRODUCTION
Grid computing has emerged from distributed computing where the resources of many computers in a network are used to solve a single problem at the same time. The difference between the conventional high performance computing such as cluster computing and grid is that grid tends to be loosely coupled, heterogeneous and geographically distributed [1].If large numbers of tasks are computed on the geographically distributed resources, a reasonable scheduling algorithm must be adopted in order to get the minimum completion time. To achieve the promising potentials of tremendous distributed resources,
effective and proficient scheduling algorithms are vitally important [12].Job scheduling is the mapping of jobs to specific physical resources, demanding to reduce some cost for the execution and the time required for the execution. When the number of tasks is less than the number of resources, then tasks can be allocated on resources on the first come first serve principle otherwise when the number of tasks is more than the number of available resources, allocation of task is to be made by some scheduling schemes.
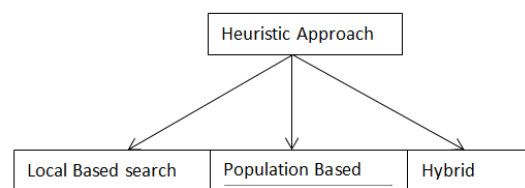

Figure 1Criterian for various approaches

Job scheduling is the mapping of jobs to specific physical resources, demanding to reduce some cost for the execution and the time required for the execution. When the number of tasks is less than the number of resources, then taskscan be allocated on resources on the first come first serve principle otherwise whenthe number of tasks is more than the number of available resources, allocation of task is to be made by some scheduling schemes. The grid schedulers deal with organizing the information providers in such a way that the users can have an easy access to the data. They can also identify the file system or whether any type of resource is cached or which resource is rapidly available. In the grid system, an end user submits the job that has to be executed with some constraints like job execution deadline, cost for the execution and the time required for the execution[2]. Grid resource manager estimates the resource requirements and provides the functionality for discovery and publishing of resources as well as scheduling, submission and monitoring of jobs.

In this review paper we are focusing on Heuristic approaches for optimization. Heuristic scheduling algorithms which are sub optimal are faster than cost and time intensive algorithms. Heuristic scheduling is widely used in grid environment due to its advantage over time and cost [1].Heuristic approach is subdivided into Local search based, Population based and Hybrid heuristics. Simulated Annealing algorithms (SA), Taboo search algorithms (TS) are categorized under Local search based.
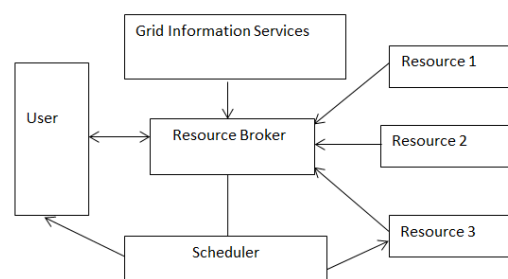

Figure 2 Functioning of resource broker

Ant Colony scheduling algorithms (ACS), Genetic algorithms (GA),Particle Swarm Optimization Algorithm (POS) are categorized under Population Based. Hybrid heuristics are combination of local search based and population based. These algorithms are summarized as follows:

2. A COMPARATIVE ANALYSIS AMONG VARIOUS SCHEDULING ALGORITHMS

2.1 Ant Colony SchedulingAlgorithms (ACS):This algorithm reveals the behavior of the real ants. When the ants hunt for food,they deposit some amount of pheromone on the path and connect to each other by the pheromone laid on the trails travelled. The amountof pheromone on every path will affect the possibility of other ants to select the path. With time, as the amount of pheromone in the shortest path between the nest and food source increases, the number of ants attracted to the shortest path also increases At last all the ants will select the shortest path. In the context of Grid Scheduling working components are the ants, computational power is the food, and the mapping is the path[4][5]. The amount of information deposited on components may depend on the quality of the solution found. ACS has been applied to number of combinatorial optimization (CO) problems such as traveling salesman problem (TSP), Job-shop Scheduling Problem (JSP), Vehicle Routing Problem (VRP), Quadratic Assignment Problem (QAP), etc.

Advantages

- Less affected by initial poor solutions due to selection of random path.

- Can be used in dynamic applications such as Vehicle Routing Problem (VRP), Quadratic Assignment Problem (QAP).

Disadvantages

- ACO algorithm performs much slower than the other scheduling algorithms. The reason behind that ACO computes several times for searching the optimal resource that is allocated to process the job. Hence, the calculation time of scheduling consumes more resources.

- Time to convergence is uncertain but convergence is guaranteed.

2.2Genetic Algorithm
Genetic Algorithm is an array of searching algorithms it searches a solution space of optimal solution to a problem. The key characteristic of the GA is how the searching is done. Actually, the algorithm creates a *population* of possible solutions to the problem and lets them *evolve* over multiple generations to find better and better solutions.[7], [8].The better solution of generation is evaluated according tothe fitness value and the candidates with better fitness values are used to create further solutions through crossover and mutation processes.In this paper [6] author proposed a modified genetic algorithm to schedule the tasks within the grid environment and to find most suitable match between the tasks and grid resources.The author considered two parameters the total make span and throughput of the environment.The goal of this algorithm is to minimize the makespan of tasks and thereby increase the throughput of the environmentMoreover, the proposed GA converges to the suitable solution for a low number of iterations.

Disadvantages

- The main disadvantages of genetic algorithms are the disruption of good sub-solutions by crossover and mutation operations.

- And undesired population diversity loss by selection operations, which constantly decreases the variety of its specimens.

2.3 Simulated Annealing algorithms
This concept is motivated by a thermal process "annealing in metallurgy", a process involves heating and controlled cooling of a material to increase the size of material and minimize their defects.SA theory states that if the temperature is reduced sufficiently slowly the solid will reach thermal equilibrium which is an ideal state.[13]In the context of Grid Scheduling thermal equilibrium is an ideal task-resource mapping, temperature is total completion time of mapping and change of temperature is the process of mapping change [7].It aims to find simplyacceptably good solution in fixedchunk of time rather than the best optimal solution.

Advantages

- Itstatistically guarantees finding an acceptably good solution.

Disadvantages

- This method is incapableto tell whether it has reached to good solution some other complimentary techniques are required to perform this.

- The selection of the initial parameters is challenging task.

2.4 Tabu Search Algorithm
Tabu exploits neighborhood search approach in which it moves from one solution to another solution in the neighborhood until it reaches some stopping criteria. It employs the concept of memory .It maintains the list of recently visited solutions as well as stores information gathered during the whole complete exploration process.[7] [9].So Tabu search is also a solution space search. Possible mapping of tasks with machine is compared with currently mapped tasks with machine. If the new makespan,flowtime and completion time of a task is better than the current one, the new solution is saved replacing the current one. The final mapping is saved from the solution space in to the tabu list. The next mapping of task with machine is done by considering this updated tabu list.

Advantages

- In terms of implementation, Tabu is conceptually much simpler and computationally less demanding than other techniques.
- The memory structures employed in the Tabu search serves as a guide forsearching a new possible solution without reverting back to previously encountered solutions.

Disadvantages

- The memory spaceallocation needed to provide the design space with enough memory tofind an optimal solution.

2.5 Particle Swarm Optimization Algorithm
It is population based heuristic approach for optimization. It imitates the behavior of flying birds in the search space.

Initially a swarm of particles is generated randomly. Each particle represents an independent feasible solution in problem space about resource scheduling. Each particle has an updating position vector and updating velocity vector[10]. Changing velocity enables the particle to search around it individual best position and updating global best position .The flying particle in space adjusts its position according to its own experience and the experience of its neighbors[11].

Advantages
- The optimization method is not generation based.Each particle represents an independent solution

- This allows an easy parallelization and hence more computations per time unit are possible

Disadvantages
- The method suffers from the partial optimism, which causes the less exact at the regulation of its speed and the direction.
- Then the method cannot work out the problems of scattering and optimization and the method cannot work out the problems of non-coordinate system, such as the solution to the energy field and the moving rules of the particles in the energy field.

Table 1: Comparison between Local Search Based algorithms

S.No	Simulated Annealing	Tabu Search Algorithm
1	SA is a search technique based on a thermal process "annealing in metallurgy"	Tabu Search is a mathematical method for local search space
2	In fixed amount of time it finds acceptably good solution.	It moves from one solution to another solution in the neighborhood until it reaches some stopping criteria
3	It does not make use of historical memory	It make use of historical memory

Table 2: Comparison between Population BasedAlgorithms

S. No	Ant Colony Scheduling Algorithm	Genetic Algorithm	Particle Swarm Optimization Algorithm
1	It imitates the behavior of the ants	It is based on the evolutionary process for large search space	It imitates the behavior of flying birds in large search space
2	The better solution is evaluated on the basis of amount of pheromone deposited on the path.	The better solution of generation is evaluated according tothe fitness value	The better solution is evaluated according to the position of particle in the space.
3	The scheduler attempts to find optimal distribution of wok unit to processing nodes.	The candidates with better fitness values are used to create further solutions through crossover and mutation processes.	The flying particle in n dimension space adjusts its position according to its own experience and the experience of its neighbors.
4	The ants represent the work units,food represents the computational power and the path represents the mapping.	Each chromosomes represent the feasible solution, a fitness value represents the completion time of task	Each dimension represents a task and each particle represents a feasible solution

3. CONCLUSION
The various heuristic scheduling algorithmshave been studied and classified .Their merits and demerits have also been discussed .The goal of this study is to provide a basis for comparison and provide insight into characteristics of each algorithm in order to select appropriate algorithms according to the characteristics of jobs ,machines and internet connectivity.This study will help concerned researchers to generate more proficient and competent scheduling algorithms.The future work will be concentrated on eradicating the demerits of algorithms to make them more efficient in heterogeneous and dynamic environment.

4. REFERENCES
[1] G. Jaspher W. Kathrine , MansoorIlaghi U, "Job Scheduling Algorithms in Grid Computing – Survey" ,

International Journal of Engineering Research & Technology (IJERT), Vol. 1, 2012

[2] Dr.K.Vivekanandan, D.Ramyachitra, "A Study on Scheduling in Grid Environment", International Journal on Computer Science and Engineering ,Vol. 3 ,No. 2, Feb 2011

[3] Raksha Sharma, Vishnu Kant Soni, Manoj Kumar Mishra, PrachetBhuyan,"A Survey of Job Scheduling and Resource Management in Grid Computing" World Academy of Science, Engineering and Technology 40 2010

[4] SiriluckLorpunmanee, Mohd Noor Sap, Abdul Hanan Abdullah, and Chai Chompoo-inwai ,"An Ant Colony Optimization for Dynamic Job Scheduling in Grid Environment", International Journal of Computer and Information Engineering 1:8 2007

[5] Nada M. A. Al Salami,"Ant Colony Optimization Algorithm",UbiCC Journal, Volume 4, Number 3, August 2009

[6] Reza Entezari-Maleki, Ali Movaghar, "A Genetic Algorithm to Increase the Throughput of the Computational Grids",International Journal of Grid and Distributed Computing Vol. 4, No. 2, June, 2011

[7] Tinghuai Ma, QiaoqiaoYan,Wenjie Liu, Dinghai Guan and SunYoung Lee, "Grid Task SchedulinAlgorithm Review" IETE Technical Review, Vol. 28, 2011.

[8] Yang Gao, HongqiangRongb, Joshua Zhexue Huang, "Adaptive grid job scheduling with genetic algorithms", Future Generation Computer Systems, Vol. 21,2005

[9] FatosXhafa, Javier Carretero, "A Tabu Search Algorithm For Scheduling Independent Jobs in Computational Grids", Computing and Informatics, Vol. 28, 2009.

[10] Lei Zhang, Yuehui Chen, Runyuan Sun, Shan Jing and Bo Yang, "A Task Scheduling Algorithm Based on PSO for Grid Computing", International Journal of Computational Intelligence Research ISSN 0973-1873 Vol.4, No.1 ,2008.

[11] Qinghai Bai, "Analysis of Particle Swarm Optimization Algorithm", Computer and Information Science, Vol. 3,No. 1, 2010

[12] G. Jaspher W. Kathrine ,MansoorIlaghi "Survey OnGridScheduling", Journal of Computer Applications, Vol. 3, No.3, 2010

[13] R. Joshua Samuel Raj,Dr. V. Vasudevan, "Beyond Simulated Annealing in Grid Scheduling", International Journal on Computer Science and Engineering, Vol. 3, No. 3, Mar 2011