EFFECTIVE WORKFLOW MANAGEMENT ALGORITHMS FOR OPTIMIZATION THE SCHEDULING IN GRID COMPUTING

Baljeet Kaur
Sarpreet Singh

1Department of Computer Engineering, Sri Guru Granth Sahib World University, Fatehgarh sahib, India
2Department of Computer Engineering, Sri Guru Granth Sahib World University, Fatehgarh sahib, India

ABSTRACT

Grid computing is a high performance computing environment to solve larger scale computational demands. Grid computing contains resource management, scheduling, security problems, information management and so on. Scheduling is the use to achieve high performance in grid environments. This paper explains one of the large growing optimization algorithms, the ant colony algorithm and genetic algorithm in Grid computing. It uses different heterogeneous systems for sharing large-scale resources to solve very large scientific problems. In this paper, two optimization algorithms are used for scheduling techniques in grid computing. It aims to find a suitable allocation of resources for each job with the comparison of hybrid ACO and GA. We compared our proposed hybrid results with previous results in terms of execution time & execution cost which proves that our new hybrid approach has better results.

Keywords: Grid Computing, Job Scheduling, genetic Algorithms, and Ant colony optimization algorithm.

1. INTRODUCTION

Computing is an activity which is used for creating and designing computers. Computing includes all the software and hardware designing, structuring and developing processes and manages them properly. Parallel and distributed are the types of computing.

In parallel computing, memory is shared between all the processors to exchange information. Bit-level, instruction level and data level are its type.

In distributed computing, each processor has its own private memory. Information is exchanged by passing messages between the processors. Distributed computing is divided into various techniques.

- Cloud computing
- Grid Computing
- Cluster computing
a) **Cloud Computing** is one of the types of distributed computing. Cloud computing is a branch of computing that comes under sharing of data instead of local servers. This policy of distributed computing is finished through pooling of all computer resources together. It is under the control of software not under any external users. Cloud computing is scalable process which runs over the virtual system with the help of service of internet. It is considered as a next generation of highly scalable distributed computing system. Cloud computing incorporates on-demand deployment, virtualization, open source software and Internet delivery of services. It also allowed users and its customers to access their personal files at any computer with the help of internet without installation.

b) **Grid computing** is hardware and software consisted infrastructure that gives a dependable, consistent and high capable computations at the end. It is different from cloud computing as it deals with homogenous system but grid computing is deal with heterogeneous type system which includes processors and workstations and clusters and gives different computational results. The resources are geographically distributed at different computer in grid computing. Grid computing allowing their computing power to be shared. Grids can merge the resources of thousands of different computers to make an especially powerful computing resource, so that it can be easily accessible from the comfort of a personal computer and can be easily useable. In grid every node has its own resource manager and works as a single entity [1].

b) **Cluster computing** is the third category that comes under distributed computing. Cluster computing is a form of computing in which a group of computers are linked together so that they can act like a single entity. Clusters are generally deployed to improve performance and availability over that provided by a single computer. It is usually much more cost-effective than single computers of comparable speed or availability. Components of all the clusters are mainly common [8]. As grid computing is heterogeneous in nature and it has no control over the locally available resource because its resources are distributed geographically. So information becomes limited due to no control over local resources. Current scheduling system in grids is time-dependent and cannot meet the necessary requirements like network allocation etc.

**1.1 SCHEDULING**

Scheduling is choosing the most suitable resource for a job to complete its execution either in terms of waiting time, turnaround time or cost. Scheduler is used to manage the jobs and resources. Scheduler performs two main functions; First scheduler selects the appropriate computational resource for the job and then assigns the resource to the jobs. The main objective of scheduling is to reduce the completion time of an application by properly allocating the jobs to the processors. In this paper we describe properties and services characterizing the Grid Resource Management System (RMS) that provides services for the management and exploitation of the resources (scheduling services are part of the RMS). Then we tackle the problem of scheduling applications in a Grid environment and give a brief summary of current projects dealing with Grid scheduling systems.

**1.2 SCHEDULING ALGORITHMS**

**Genetic algorithm**

Genetic algorithm starts from a population of randomly generated individuals, and is an iterative process, with the population in each iteration called a generation. In each generation, the fitness of every individual in the population is evaluated; the fitness is usually the value of the objective function in the optimization problem being solved. The new generation of candidate solutions is then used in the next
iteration of the algorithm. Commonly, the algorithm terminates when either a maximum number of generations has been produced, or a satisfactory fitness level has been reached for the population. [4]

**Ant Colony Optimization algorithm**

Ant Colony Optimization (ACO) is one of the metaheuristics. It can be applied not only to solve discrete optimization problems but also to solve both static and dynamic combinational optimization problems. This behavior enables ants to find the shortest paths between food sources and their nest. In fact, they deposit a chemical pheromone tail on the ground after they walk from the nest to food sources and vice versa. [10]

II. RELATED WORK

Hassan [1] proposed workflow management based on a multi-objective Genetic Algorithm (GA) to improve grid computing performance. Task runtime is an important factor in grid computing. The proposed method considers a collection of levels as workflow so that to reduce the need to check workflow dependencies after a solution is obtained for the next population. As a result of this both time of scheduling and solution quality are improved. Results are presented which show that the proposed method has better performance compared to similar techniques.

Sandip et al. [2] proposed an Enhanced Genetic Algorithm (EGA) for achieving task scheduling with load balancing. The Enhanced Genetic Algorithm (EGA) is designed based on the standard Genetic Algorithm (GA). The method requires an encoding scheme which can represent all legal solutions to the optimization problem. The simulation results show that proposed algorithm yields better performance when compared with other traditional heuristic approaches.

Reza et.al [3] proposed to assign the tasks to the grid resources with goal of minimizing the total make span of the environment. The algorithm uses genetic approach to find the most suitable match between the tasks and grid resources. The simulation results obtained after applying the proposed algorithm is to schedule independent and sequential tasks to the grid resources so that it displays the applicability of the algorithm in grid environments.

Pooranian [4] et al. combine the genetic algorithm and Gravitational Emulation Local Search (GELS) as a method to solve scheduling problem by which simultaneously pay attention to two factors of time and number of missed tasks. Results shows that the proposed algorithm can decrease make span while minimizing the number of missed tasks compared with the traditional methods. The purpose of Grid computing is to utilize computational power of idle resources which are distributed in different areas. Since task scheduling includes in the NP-hard problems various researches have focused on invented algorithms especially the genetic ones. But since genetic is an inherent algorithm which searches the problem space globally and does not have the efficiency required for local searching.

Amin et.al [5] proposed approach gravitational attraction search as a local search algorithm has been associated to genetic algorithm to enhance its capability to search more intelligent in problem search space and achieve accurate response in less time. Comparing Hybrid Genetic Algorithm-Gravitational Attraction Search algorithm (HYGAGA) and genetic algorithm results asserts significant enhancement in the performance of search algorithm. In addition, Hybrid Genetic Algorithm-Gravitational Attraction Search algorithm (HYGAGA) could attain more appropriate solutions with comparing to other genetic algorithms like Genetic Algorithm (GA) and Hill Climbing (HC), Genetic Algorithm (GA) and Simulated Annealing (SA) and Genetic Algorithm (GA) and Tabu Search (TS) implemented and analyzed to compare their ability in search in problem search spaces. The Hybrid Genetic Algorithm-Gravitational Attraction Search, HYGAGA, proposed as a memetic algorithm to solve grid task scheduling problem.
III. PROPOSED METHODOLOGY

Work flow scheduler is that works on top of multiple local schedulers. The grid scheduling process makes resource allocation decisions involving resources over multiple administrative domains, which introduce several challenges for work flow scheduler. This paper explains one of the large growing optimization techniques, the ant colony algorithm and genetic algorithm in Grid computing. The Genetic algorithm starts from a population of randomly generated individuals, and is an iterative process, with the population in each iteration called a generation. In each generation, the fitness of every individual in the population is evaluated; the fitness is usually the value of the objective function in the optimization problem being solved The GA can solve every optimization problem which can be described with the chromosome encoding. Genetic algorithm is a method which is very easy to understand and it practically does not demand the knowledge of mathematics. Genetic algorithms are easily transferred to existing simulations and models. ACO is inspired by a colony of ants that work together in foraging behavior. This behavior encouraged ants to find the shortest path between their nest and food source. Every ant will deposit a chemical substance called pheromone on the ground after they move from the nest to food sources and vice versa. Therefore, they will choose an optimal path based on the pheromone value. The path with high pheromone value is shorter than the path with low pheromone value. We are comparing our proposed hybrid result with previous in terms execution time and execution cost both these factors are better in case of hybrid.

3.1 Proposed Model

The proposed model focuses on following objectives which are helpful to reduce the burden of the processor.

- To provide High quality workflow scheduling solutions in grid computing.
- To implement hybrid algorithm using GA & ACO.
- To improve Grid Computing performance.

3.2 Basic Block Design

In this proposed work, Hybrid algorithm is used which is a combination of both Genetic Algorithm & Ant colony optimization. GA starts with a set of randomly selected chromosomes as the initial population that encodes a set of possible solutions then the chromosomes are evaluated according to their fitness values using some measures of profit or utility that we want to optimize. Ant Colony Optimization (ACO) is one of the metaheuristics. It can be applied not only to solve discrete optimization problems but also to solve both static and dynamic combinational optimization problems. ACO is inspired by a colony of artificial ants.

3.3 Hybrid Algorithm

Step1:- Select no. of resources and tasks.
Step2:- Apply Hybrid algorithm.
Step3:- Find Fitness value of tasks and ability to handle it resources.
Step4:- Acc. to crossover function assigns resource to tasks.
Step5:- Acc. To tasks information will be shared with other task.
Step6:- Find instance best value for every task.
Step7:- Repeat step 2&7.
Step8:- Find gross best value for whole scheduling.
Step 9: compare execution time with individual GA & ACO.

Step 10: End.

IV. RESULTS AND FUTURE WORK

This proposed hybrid approaches are compared with previous approaches that give us effectiveness. Effectiveness of the system is calculated in terms of execution time & execution cost. The execution cost and time of each task for each service able to execute it can be estimated from the length of the tasks and the service speeds. In future we can use in load balancing algorithm. Figure 4.1 shows the scheduling algorithms in which select each algorithm. Figure 4.2 show the results for proposed hybrid method which is combination of both Genetic Algorithm and Ant Colony Optimization. Which is calculated by make span time to jobs and machines. Figure 4.3 show the results of proposed hybrid algorithm shows which are combination of both Genetic Algorithm and Ant Colony Optimization. Execution time & execution cost values are calculated.
Grid computing is a computing environment with high performance to solve larger scale computational demands. Allocation of resources, security, scheduling are the key challenges of grid computing. The two approach genetic algorithm and ant colony optimization has been implemented to obtain the optimal solution. In this we also compared our proposed hybrid results with previous results in terms of execution time and execution cost.

Table 1: Execution time and execution cost Results.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Execution Time</th>
<th>Execution cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA</td>
<td>11</td>
<td>30</td>
</tr>
<tr>
<td>ACO</td>
<td>10</td>
<td>27</td>
</tr>
<tr>
<td>Hybrid</td>
<td>9</td>
<td>25</td>
</tr>
</tbody>
</table>
time & execution cost which proves that our new hybrid approach has better results. The proposed Hybrid technique which is composition of genetic and ant colony optimization gives better performance so that scheduling becomes fast. In this approach, we can use in load balancing algorithms.

REFERENCES


