Workflow Scheduling Algorithms in Grid Computing

Neha Bhardwaj
CSE Department UIET
Kurukshetra University, Kurukshetra
Haryana, INDIA
bhardwaj.mylife@gmail.com

Abstract—Grid computing is a process of aggregate the functionality of different geographically resources and provide services to the user. Scheduling is most popular research area in grid computing for achieving high performance. In scheduling, tasks are assigning to resources and maintain the execution. Dependencies constraints are need to preserve for workflow scheduling in grid computing. In this paper, surveyed various workflow scheduling algorithms. Lot of research has been done in the area workflow scheduling but still there is a requirement of optimization techniques like bat optimization, water drop optimization to the workflow scheduling.

Keywords—Grid Scheduling; QoS; Workflow Scheduling; Intelligent Water Drop Algorithm; Bat Algorithm.

I. INTRODUCTION

Grid computing is like a distributed system with non-interactive workloads having large number of files. Grid computing is highly dispersed with heterogeneous resources than cluster computing. Unlike cluster computing, nodes of grid computing are having different task or application to perform. Grid is an environment for solving the problem, which is submitted by the one more user without knowing the location of resources[1].

Grid scheduling is process of assigning the task to resources grid and achieving the high performance. Grid workflow is set of tasks which are distributed on the resources for achieving the goal. Data intensive application such e-science, high energy physics, chemistry etc. uses the grids to manage, share and process large sets of data [1].

Grid having various challenges in security [2], implementation models [3], access control [4], resource management and workflow management. In workflow, the tasks are set of sub tasks and having the dependency among them. Hence there are various challenges faced in grid workflows. Some of the challenges are [5]:

- Grid having various challenges like accessibility, availability etc.
- Workflow tasks distributed on various resources for completion.
- Some tasks of workflow need to be executed in parallel and concurrent manner.

Grid workflows create, manage and execute various grid applications with high efficiency. Grid workflow is highly dynamic, heterogeneous and distribution in nature therefore it is very difficult to solve the problem in grid environment. Grid workflow scheduling algorithms based on DAG (Directed Acyclic Graph). Scheduler assigns appropriate resources to workflow tasks for execution to be completed and satisfy objective functions applied by the users. The problem of mapping the workflow tasks to the distributed and heterogeneous resources is belongs to class of problems which is known as NP-Hard problems [6]. For such workflow problems, there is no algorithm which provides the optimal solutions within the polynomial time. Many heuristic and meta-heuristic based algorithms have been proposed for the implementation of workflow in grid environment.is very difficult to solve the problem in grid environment. Grid workflow scheduling algorithms based on DAG (Directed Acyclic Graph) [6].

A. Grid workflow Tasks Type

Grid workflow have two types of task [7]:

- Metatasks: The result of metatask is not affected by the sequence of metatask because they are dependent but there is no dependent relation among metatasks. The objective of scheduling algorithm for metatask is makespan.
- Dependent tasks: The tasks having the same relation like data, temporal relation. So, the order of the task cannot be changed.

B. Scheduling Phase

The scheduling process of workflow having 3 phases [7]:

- 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.

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

II. WORKFLOW SCHEDULING

Workflow is set of task, subtasks having dependency among them. Workflow scheduling algorithm takes input of workflow model which provide the set of task (i.e workflow). There are two types of workflow model [6]:

1. Deterministic model: The input tasks and dependency among them is known in advance.
2. Non-deterministic model: The input task and dependency is known at run time.

The deterministic type of algorithm is represented in the form DAG. Define number jobs in workflow and edges between jobs such as \((T_i, T_j)\) edge. \(T_i\) is a parent of \(T_j\) hence \(T_i\) must be executed before execution of \(T_j\) [8].

![Figure 1. DAG Representation of Workflow.](image)

In a workflow, task execution is decided on the basis of dependencies among them and executed after completion of parent node. As shown in fig. 1 \(T_3\) is executed after completion of \(T_0\) and \(T_1\). The related task like \(T_2\) and \(T_3\) are assigned to same service node. The priorities are used for solving the conflict among tasks. The two dimensional string to represent the scheduling result of fig. 1 is shown in fig. 2 [8].

![Figure 2. Schedule Plan](image)

III. EXISTING WORKFLOW SCHEDULING ALGORITHM

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 [9] 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. [10] 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. [11] 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. [12] 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 Chittaranjan Hota [13] 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. [14] proposed a Reliability Enhanced Grid Workflow Scheduling Algorithm with budget constraint which can maximize the reliability.
7. Fabio Coutinho et.al [15] defined an energy efficient model and HGreen heuristic which assign the heaviest workflow tasks to energy efficient resources.
8. Dengpan Yin and Tevfik Kosar [16] proposed A-star based data-aware workflow scheduling algorithm. 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.
10. Wei-Neng Chen and Jun Zhang [18] 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. [19] proposed a Deadlock Segment Leveling (DSL) a novel heuristic which divide the workflow application into segments and further segments are partitioned into groups.

<table>
<thead>
<tr>
<th>Scheduling Approach</th>
<th>Algorithm</th>
<th>Year of Publication</th>
<th>Type of Scheduling</th>
<th>Objectives</th>
<th>Future Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain Storm (Swarm algorithm) [9]</td>
<td>Multiobjective Brain Storm Algorithm(MOBS A)</td>
<td>2014</td>
<td>Meta-Heuristic</td>
<td>Reduce Execution time and Economic Cost</td>
<td>Try to add more objectives.</td>
</tr>
<tr>
<td>Game theory Approach [10]</td>
<td>Genetic and Random Algorithm</td>
<td>2013</td>
<td>Heuristic or Meta-Heuristic</td>
<td>Reduce Time and cost</td>
<td>Different choice of algorithm can be used in the proposed Approach.</td>
</tr>
<tr>
<td>Deadline Satisfaction Degree of workflow(DSDW) [12]</td>
<td>Deadline satisfaction Enhanced Scheduling Algorithm</td>
<td>2011</td>
<td>Heuristic</td>
<td>Reduce Total Execution Time</td>
<td>In scheduling, can be considered reliability, performance of grid resources.</td>
</tr>
<tr>
<td>Task with maximum completion time assign to fastest machine [13]</td>
<td>Efficient Algorithm on Heterogeneous Computing System(EAHCS)</td>
<td>2011</td>
<td>Heuristic</td>
<td>improved Makespan, System Utilization and Load balancing</td>
<td>Communication cost can considered for further work.</td>
</tr>
<tr>
<td>A-star algorithm [16]</td>
<td>Data-Aware Workflow Scheduling Algorithm based on A-star</td>
<td>2011</td>
<td>Heuristic</td>
<td>Minimized turnaround time</td>
<td>Try to add more heuristic for more improvement in the algorithm.</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>------------------------------------------</td>
<td>------</td>
<td>----------------</td>
<td>----------------</td>
<td>------------------------------------------------------------------</td>
</tr>
<tr>
<td>Segment Leveling [19]</td>
<td>Deadline Segment leveling Algorithm</td>
<td>2009</td>
<td>Heuristic</td>
<td>Cost Optimization and Synchronization</td>
<td>Try to add more heuristic so that it become more synchronized and cost optimized.</td>
</tr>
<tr>
<td>Resubmission Impact for Fault tolerance combination of Task Replication and Task Resubmission, [20]</td>
<td>HEFT with Task Replication</td>
<td>2009</td>
<td>Heuristic</td>
<td>Without information of about resource failure, Resubmission impact wastes on average 42% less than other approaches without changing in task rate and overall performance.</td>
<td>Try reduce waste of resources by introducing new methods.</td>
</tr>
</tbody>
</table>

### IV. PROBLEM FORMULATION

Intelligent drop optimization technique is applied with ant colony algorithm to reduce makespan and improve load balance [11]. There is need to apply this optimization technique in such way other constraints like availability, budget constraints etc. are also improved. Bat optimization [21] is also need to apply in scheduling problem of grid environment. Bat optimization algorithm is applied in cloud computing [22] which is an meta-heuristic algorithm.

### V. CONCLUSION

In this paper, surveyed various workflow scheduling algorithm. The algorithm was explained and tabulated them on the basis of scheduling approach, type of scheduling, year, objectives and future work. Lots of research carried out but there are still requirement of improvement or need to be optimized. As concluded that there is need to explore intelligent drop optimization and bat optimization in the scheduling problem of grid environment.
REFERENCES

The data-intensive workflow in scientific and enterprise grids has gained popularity in recent times. Data-intensive workflow needs to access, process and transfer large datasets that may each be replicated on different data hosts. Because of the large data sets, the execution time is bounded by the cost of data transfer. @article{Xu2009ADW, title={A Data-Intensive Workflow Scheduling Algorithm for Grid Computing}, author={Meng Xu and Li-zhen Cui and Haiyang Wang and Yanbing Bi and Ji Bian}, journal={2009 Fourth ChinaGrid Annual Conference}, year={2009}, pages={110-115} }. Meng Xu, Li-zhen Cui, +2 authors Ji Bian. Published in. These scheduling algorithms focus on resource management, response time, load balancing and performance. Index Terms: Cloud computing, Make Span, Priority, Quality of Service, Resource Allocation, Task Scheduling, Task Completion Time. â€œuâ€œ. 1 1. â€œTowards improving QoS-guided scheduling in gridâ€œ Third ChinaGrid Annual Conference(CHINAGRID). Dunhuang, Gansu, China, 2008, p.89-95. â€œA survey of various workflow scheduling algorithms in cloud environmentâ€œ. International Journal of Computer Application (IJCA).2011,p.26-30. IJSTR©2013 www.ijstr.org. Abstractâ€œScheduling in Grid computing has been an active area of research since its beginning. However, beginners find very difficult to understand related concepts due to a large learning curve of Grid computing. Thus, there is a need of concise understanding of scheduling in Grid computing area. This paper strives to present concise understanding of scheduling and related understanding of Grid computing system. Section IV covers scheduling algorithms related to Grid computing falling into two main types: resource scheduling and application scheduling. Section V highlights methodology of evaluation of scheduling algorithms using both real system and simulation tool based approaches. Finally, Section VI provides conclusion. II.