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 metatasks 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 some 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].

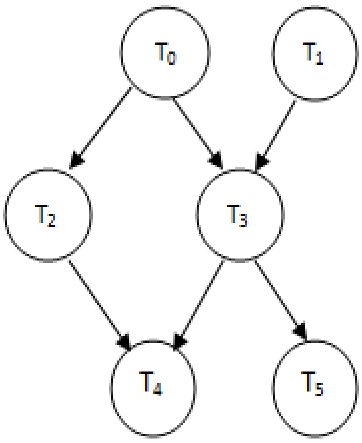

Figure 1. DAG Representation of Workflow.

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].

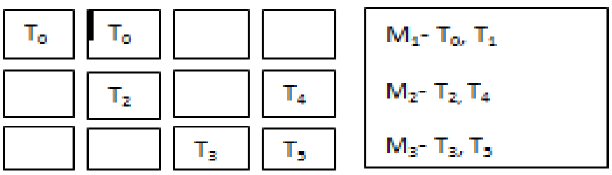

Figure 2. Schedule Plan

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 Chittarajan 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.
9. Zhang Wenpeng and Liu Hongzhao [17] proposed algorithm with combination of advantages of genetic algorithm, clonal algorithm and simulated annealing.
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.
12. Kassian Plankensteiner et al. [20] proposed resubmission impact a fault tolerance heuristic for scientific workflows. Heuristic is based on task replication and task resubmission.

Table 1. Survey of Workflow Scheduling Algorithm

Scheduling Approach	Algorithm	Year of Publication	Type of Scheduling	Objectives	Future Work
Brain Storm (Swarm algorithm) [9]	Multiobjective Brain Storm Algorithm(MOBSA)	2014	Meta-Heuristic	Reduce Execution time and Economic Cost	Try to add more objectives.
Game theory Approach [10]	Genetic and Random Algorithm	2013	Heuristic or Meta-Heuristic	Reduce Time and cost	Different choice of algorithm can be used in the proposed Approach.
Ant Colony with intelligent water drop [11]	Ant Colony Algorithm and Intelligent Water Drop Algorithm	2013	Meta-Heuristic	Reduce makespan, improve system utilization and load balancing	Try optimized more by using different approaches.
Deadline Satisfaction Degree of workflow(DSDW) [12]	Deadline satisfaction Enhanced Scheduling Algorithm	2011	Heuristic	Reduce Total Execution Time	In scheduling, can be considered reliability, performance of grid resources.
Task with maximum completion time assign to fastest machine [13]	Efficient Algorithm on Heterogeneous Computing System(EAHCS)	2011	Heuristic	improved Makespan, System Utilization and Load balancing	Communication cost can considered for further work.
Budget Constraints [14]	Reliability Enhanced Grid Workflow Scheduling Algorithm with Budget Constraint	2011	QoS based heuristic	Better Reliability and Adaptability	Performance impact of budget distribution and define improved budget distribution strategy.
HGreen Heuristic [15]	HGreen heuristic Algorithm and Energy Efficient Model	2011	Heuristic	Reduced Power Consumption	Explore new heuristics and introduce new execution scenarios.
A-star algorithm [16]	Data-Aware Workflow Scheduling Algorithm based on A-star	2011	Heuristic	Minimized turnaround time	Try to add more heuristic for more improvement in the algorithm.

Genetic, Clonal Selection, Simulated Annealing [17]	Genetic Clonal Annealing Algorithm(GCAA)	2010	Meta-Heuristic	Load Balancing	GCAA can apply with different features. Implementation of method on real grid.
7 Heuristics : Reliability greedy, Time greedy, Suggested Deadline, Suggest Budget, Time/Cost, Overall Performance [18]	Ant Colony System (ACS) Algorithm	2009	QoS based Heuristic	Reliability, Time, Cost	Try to add more heuristic and make it more optimized.
Segment Leveling [19]	Deadline Segment leveling Algorithm	2009	Heuristic	Cost Optimization and Synchronization	Try to add more heuristic so that it become more synchronized and cost optimized.
Resubmission Imapct for Fault tolerance combination of Task Replication and Task Resubmission.[20]	HEFT with Task Replication	2009	Heuristic	Without information of about resource failure, Resubmission impact wastes on average 42% less than other approaches without changing in task rate and overall performance.	Try reduce waste of resources by introducing new methods.

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

[1] J. Yu, and R. Buyya. "A taxonomy of scientific workflow systems for grid computing." ACM Sigmod Record vol. 34,no. 3, pp. 44-49, 2005.

[2] E. Cody, R. Sharman, R. H. Rao, and Shambhu Upadhyaya. "Security in grid computing: a review and synthesis." Decision Support Systems, ScienceDirect, vol. 44, no. 4, pp. 749-764, 2005.

[3] M. Parashar, J. C.Browne, "Conceptual and implementation models for the grids." in Proceedings of the IEEE, vol. 93, No. 3, 2005.

[4] G. Zhang and M. Parashar. "Dynamic context-aware access control for grid applications." in IEEE 4th International Workshop on Grid Computing, pp. 101-108, 2003.

[5] S. Benedict and V. Vasudevan. "Scheduling of scientific workflows using niched pareto ga for grids." in IEEE International Conference on Service Operations and Logistics, and Informatics, pp. 908-912, 2006.

[6] J. Yu, R. Buyya, and K. Ramamohanarao. "Workflow scheduling algorithms for grid computing." Metaheuristics For Scheduling in Distributed Computing Environments, Springer Berlin Heidelberg, pp. 173-214, 2008.

[7] H. Jin, ed. Grid and Cooperative Computing-GCC 2004 Workshops: GCC 2004 International Workshops, IGKG, SGT, GISS, AAC-GEVO, and VVS, Wuhan, China, October 21-24, 2004. Vol. 3252. Springer Science & Business Media, 2004.

[8] L. Xiang-ning, Y. Guang-ping, and G. Lin-ying. "Design of a dynamic scheduling engine of grid workflow." in IEEE 3rd International Conference on Communication Software and Networks, pp. 258-262, 2011.

[9] M. Arsuaga-Rios, and M. A. Vega-Rodriguez. "Cost optimization based on brain storming for grid scheduling." in IEEE 4th International Conference on Innovative Computing Technology, pp. 31-36, 2014.

[10] M. Yaghoobi, A. Fanian, H. Khajemohammadi, and T. Aaron Gulliver. "A non-cooperative game theory approach to optimize workflow scheduling in grid computing." in 2013 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, pp. 108-113, 2013.

[11] P. Mathiyalagan, S. N. Sivanandam, and K. S. Saranya. "Hybridization of modified ant colony optimization and intelligent water drops algorithm for job scheduling in computational grid." ICTACT Journal on Soft Computing, vol. 4, no. 1, 2013.

[12] X. Li, Z. Hu, and C. Yan. "A deadline satisfaction enhanced workflow scheduling algorithm." in 19th Euromicro International Conference on Parallel, Distributed and Network-Based Processing, pp. 55-61, 2011.

[13] S. Bansal, and C. Hota. "Efficient algorithm on heterogeneous computing system." in International Conference on Recent Trends in Information Systems, pp. 57-61, 2011.

[14] C. Yan, Z. Hu, X. Li, Z. Hu, and P. Xiao. "Reliability enhanced grid workflow scheduling algorithm with budget constraint." in IEEE International Conference on Information and Automation, pp. 629-633, 2011.

[15] F. Coutinho, L. Alfredo V. de Carvalho, and R. Santana. "A workflow scheduling algorithm for optimizing energy-efficient grid resources usage." in IEEE 9th International Conference on Dependable, Autonomic and Secure Computing, pp. 642-649, 2011.

[16] D. Yin, and T. Kosar. "A data-aware workflow scheduling algorithm for heterogeneous distributed systems." in International Conference on High Performance Computing and Simulation, pp. 114-120, 2011.

[17] Z. Wenpeng, and L. Hongzhao. "A load balancing method based on genetic clonal annealing strategy in grid environments." in International Conference on Educational and Network Technology, pp. 549-552, 2010.

[18] W.-N. Chen, and J. Zhang. "An ant colony optimization approach to a grid workflow scheduling problem with various qos requirements." in IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 39, no. 1, 29-43, 2009.

[19] H. Long, D. Rui-Hua, and H. Ming-He. "Cost-effective heuristics for grid workflow scheduling based on segment leveling." in IEEE WASE International Conference on Information Engineering, vol. 1, pp. 496-501, 2009.

[20] K. Plankensteiner, R. Prodan, and T. Fahringer. "A new fault tolerance heuristic for scientific workflows in highly distributed environments based on resubmission impact." in 5th IEEE International Conference on e-Science, pp. 313-320, 2009.

[21] J. Krause, J. Cordeiro, R. S. Parpinelli, and H. S. Lopes. "A survey of swarm algorithms applied to discrete optimization problems." Swarm Intelligence and Bio-inspired Computation: Theory and Applications, Elsevier Science & Technology Books, pp. 169-191, 2013.

[22] S. Raghavan, C. Marimuthu, P. Sarwesh, and K. Chandrasekaran. "Bat algorithm for scheduling workflow applications in cloud." in International Conference on Electronic Design, Computer Networks & Automated Verification, pp. 139-144, 2015.