Certain Improvements in Optimization Techniques for Grid Scheduling

N. Krishnamoorthy[1] and Dr. R. Asokan[2]

[1.] Assistant Professor (SG), Department of CSE, Kongu Engineering College, Perundurai, Erode-638052,
[2.] Principal, Kongunadu College of Engineering and Technology, Tholurpatti, Thottiam-621215, INDIA
krishnamoorthyer@gmail.com

Abstract: Grid computing is a computing framework based on large-scale resource sharing to run grid enabled applications. The grid system's efficiency and quality of service depends upon core functions such as resources discovery and scheduling. The system attempts to optimize scheduling to enhance system performance and also aims to use resources efficiently. This work proposes implementation of a hybrid optimization algorithm based on Memetic and Fish School optimization module, for optimal grid scheduling in a network grid. The proposed Memetic - Fish Swarm Optimization (MFSO) scheme incorporates local search techniques in the standard Fish Swarm Optimization algorithm, resulting in an efficient and effective optimization method. The proposed approach aims to dynamically create an optimal schedule to finish jobs within minimum time duration.

[N. Krishnamoorthy and R. Asokan. **Certain Improvements in Optimization Techniques for Grid Scheduling.** *Life Sci J* 2013;10(7s):568-572]. (ISSN: 1097-8135). http://www.lifesciencesite.com. 89

Keywords: Grid; Resource Selection; Grid Scheduling; Memetic optimization; Fish Swarm Optimization (FSO)

1. Introduction

With the increase in intensive computational tasks, high computational requirements resources are required. The capacity of a stand-alone machine is insufficient for intensive computation such as scientific experiments, problem solving, and simulations [1, 3]. Dedicated resources like supercomputers and mainframes are utilized by organizations at high costs. Grid Computing are applicable in varied domains where distributed computing, on-demand computing, data-intensive computing and collaborative computing are required. Dynamic resources are utilized for differing applications by the Grid technology. At present, it is a technology with high prospective for good resource utilization [1, 2]. Scheduling is one grid computing implementation issue that is widely researched for optimizing the grid system usage.

A Grid framework consists of both hardware and software infrastructure that is capable of providing reliable, steady, and inexpensive way to high-end computational capabilities [4]. Resources from distributed communities are shared in the grid environment. Terms used frequently in grid computing are defined:

Task: A task is the smallest unit that can be scheduled by the scheduler and allotted to a resource. Task has parameters such as priority, memory requirement, deadline, etc.

Job: Jobs are also termed as metatask, or application in the grid terminology and is used interchangeably. Job consists of a set of tasks that require a set of resources for completion. Jobs are at times composed of sub-jobs and/or tasks, and sub-jobs made up of tasks.

Resource: Resources perform an operation, such as memory for data storage, processor for data processing, bandwidth for data transfer and so on.

Site: Sites are autonomous unit composed of one or multiple resources. It is also termed as node.

Task scheduling: Scheduling refers to assigning of tasks to a group of resources that are distributed in multiple domains. Schedulers in the grid system are responsible for scheduling, and it manages various tasks – resource discovery and resource selection - required for effective application execution.

The aim of grid scheduling is to find an optimal resource for the task at hand to improve system performance in terms of resource utilization and optimal time for completion of the tasks. The scheduling strategies are classified as: centralized approach, distributed approach, hierarchical approach and multi-agent scheduling [5]. In Centralized approach, one (centralized) scheduling center is responsible for all the resources in the grid. However, in distributed approach, multiple scheduling centers are responsible for scheduling resources in the grid. Both centralized and distributed approaches are incorporated in a Hierarchical approach. Mini agent program dominates all network scheduling functions including resource discovery and management in the Multi-Agent approach. All these approaches are effective in scheduling grid resources - for different environment to a certain extent [6].

The functions of scheduling systems assign required resources for tasks, maintain resource load balance, and provide secure, reliable and fault-tolerant functioning. On receipt of scheduling request for applications from a host, the scheduler assigns the application to the host by choosing from the pool of applications. The resources are generally selected based on the prediction of the computing power of the host [7]. When multiple jobs are to be processed, scheduling tries to ensure task completion by resources simultaneously. Thus, when *n* users have submitted tasks, they are queued awaiting scheduling. The scheduler on receipt of the job, reads and removes the job from the queue when idle, assigning proper resources required for job execution. On completion of the assigned job, the system updates resource node to the resource queue for further assignment.

Highly sophisticated resource management systems are required for proper computing resources to compute scientific problem in a Grid. This is possible only when appropriate approaches and technologies master the large-scale grid networks and computing complexities. This paper proposes to optimize grid scheduling through hybrid optimization. The hybrid optimization is based on Memetic and Fish Swarm Optimization (FSO). The reminder of the paper is as follows: Section 2 relates works in literature with regard to optimization in grid scheduling. Section 3 deals with methodology, section 4 provide results, and a follow up discussion with section 5 concluding the paper.

2. Related Works

Zhang et al [8] proposed a heuristic approach based on particle swarm optimization (PSO) algorithm to solve task scheduling in a grid environment. Each particle represents a possible solution and the position vectors are transformed from a continuous variable to the discrete variable. The proposed approach optimizes schedule to ensure minimum completion time when completing tasks. Simulated experiments results show that PSO algorithm ensures better schedule than genetic algorithm.

Abraham et al [9] introduced a PSO based novel approach to schedule jobs on computational grids. Representations of particles position and velocity in a conventional PSO are extended from real vectors to fuzzy matrices. The proposed approach dynamically generates an optimal schedule to complete tasks in minimum time and also utilize resources efficiently. The proposed method is evaluated/compared with a direct Genetic Algorithm (GA) and Simulated Annealing (SA) approach thereby proving that this algorithm was applicable in a computational grid environment for job scheduling.

Izakian et al [10] represented a discrete Particle Swarm Optimization (DPSO) approach to schedule jobs on the grid. PSO, a population-based search algorithm is based on the simulation of birds and fish (flocking and schooling) social behavior. Particles fly in problem search space to locate optimal or near-optimal solutions. The proposed method ensures that the scheduler minimizes makespan and flowtime simultaneously in a grid. Experiments prove the proposed method to be more efficient, even surpassing those of reported meta-heuristic algorithms for this issue.

Kashyap et al [11] suggested security driven scheduling using genetic algorithm (SDSG) aiming at maximizing security while restricting security overhead to a certain limit. Extensive simulation over dynamically created heterogeneous grid environment revealed that SDSG achieved better security with reduced security overhead and makespan when compared to other algorithms like Min-MinMax-Min, SPMinMin and SPMaxMin.

Pinel et al [12] proposed a new parallel asynchronous cellular genetic algorithm for multi-core processors which is applied to scheduling independent tasks in a grid. Finding optimal schedules is an NPhard problem, for which evolutionary algorithms do find near-optimal solutions. The parallelism of the algorithm, and also different recombination and new local search operators are analyzed. This algorithm improved earlier schedules on benchmark problems. This algorithm’s parallelism makes it suitable for the bigger problem instances.

Xue et al [13] presented an optimization model for task scheduling and developed a hybrid clonal selection genetic algorithm (HCSGA) to solve it effectively. HCSGA first cloned a new group of individuals, and then used crossover/selection independently on all generated individuals respectively. The analysis and experiment’s result lead one to the conclusion that HCSGA has rapid convergence, good global search capacity, and is superior to other algorithms simultaneously.

3. Methodology

The Memetic - Fish Swarm Optimization (MFSO) is a hybrid algorithm that combines FSO with local search techniques. MFSO consists of two main mechanisms, a global mechanism in charge for global search of the search space, and a local one performing more refined search around potential solutions for problems on hand. In the following, the Memetic Algorithms, Fish Swarm Optimization, as well as the proposed MPSO scheme are described.

Memetic algorithms (MA)

Memetic algorithms (MA) consist of a family of population-based, heuristic search algorithms, which performs global optimization [14]. The inspiration behind its development was

Dawkins' "meme" [15], representing a unit of cultural evolution that exhibits refinement, as well as adaptation models in natural systems mixing evolutionary individuals adaptation with individual learning within a lifetime. A stage of individual optimization usually in the form of local search forms a part of MA search operation.

Although MAs appear similar to Genetic Algorithms (GAs) [16], the MA mimics the cultural and not the biological evolution. GAs ensure solving of problems by mimicking similar processes used by nature. To evolve a solution to a problem, GAs employs the same combination of selection, recombination and mutation as applicable to genes. Genes are usually not modified during an individual's lifetime - in nature - whereas memes do modify during an individual lifetime. Hence, most MAs are inferred as a cooperative–competitive algorithm of optimizing agents.

Generally, an MA is described as follows:

```
Begin
        Population Initialization
        LocalSearch
        Evaluation
   Repeat
        Recombination
        Mutation
        LocalSearch
        Evaluation
        Selection
     Until termination criterion is satisfied
        Return best solution
End
```

During initialization, a population is created within the search space. On taking an individual as input, the Local Search function executes a local search. Evaluation function in MA plays the role of the objective function. After an initial population has been created, a recombination process ensues for selected individuals. Recombining of selected individuals based on the Recombination function to create new individuals. The Mutation function performs mutation on some individuals in the population. Selection function chooses individuals who will survive in the next population. Termination condition can include various criteria including time-expiration and/or generation-expiration [17].

Fish Swarm Optimization (FSO)

FSO's idea is to simulate the process of fish swarm foraging [18]. Fish swarm typically find nutrient-rich areas quickly in oceans and lakes. When the behavioral characteristics of fish swarms is observed, fish activities feature clustering, following and foraging behaviors. The following are typical behaviors of a fish swarm:

(i) Foraging behavior. Generally, fish swim freely and randomly in water. When they find food, they quickly swim to the way where food is gradually increased

(ii) Clustering behavior. When the fish swarm moves, to survive and to avoid harm they naturally cluster as a group. In clustering, fishes follow three rules:

a. Separation rule. They try to avoid overcrowding and keep distance from each other.

b. Alignment rules. They try to move in the same direction with partners as much as possible.

c. Cohesion rule. They try to move to the center of the partners.

(iii) Following behavior. When one/several fish find food, the other fish near them swim in the direction of the food.

FSO algorithm uses the model similar to most swarm intelligence algorithms like genetic algorithm (GA). It first analyzes the problem and determines coding for optimization of individuals (artificial fishes). It then generates and initializes the original population of individuals and then based on the behavioral characteristics of the fish in the swarm; perform optimizing iteration of the population using algorithm operations. Finally, it decodes the best individual obtained as a layout pattern solution. The flowchart of the FSO is given in Fig.1.

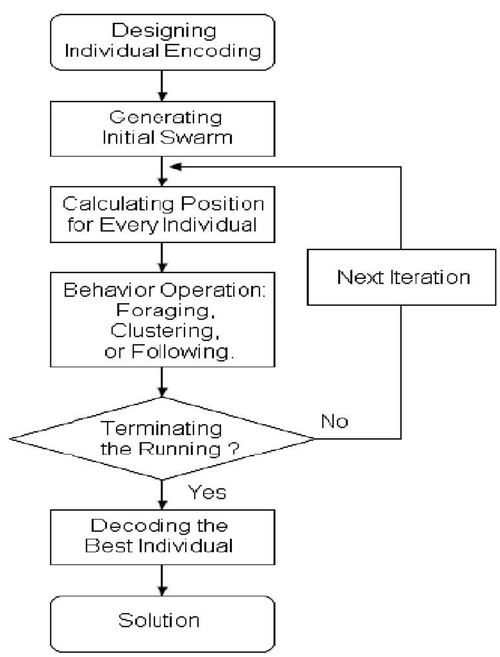

Fig 1: Flowchart of the FSO

FSO is based on simulating a fish swarm foraging, similar to GA which is also based on the iterative process. FSO is applicable to a wide range of practical project optimizations. Compared to traditional optimization algorithms and other evolutionary computations, FSO's characteristic can be summarized as follows:

(i) FSO has a faster convergence rate; it can solve problems with requirement of real-time control.

(ii) To some optimization problems that do not need solutions but only precision, FSO could provide a feasible solution.

The proposed hybrid optimization flowchart is shown in Figure 2.

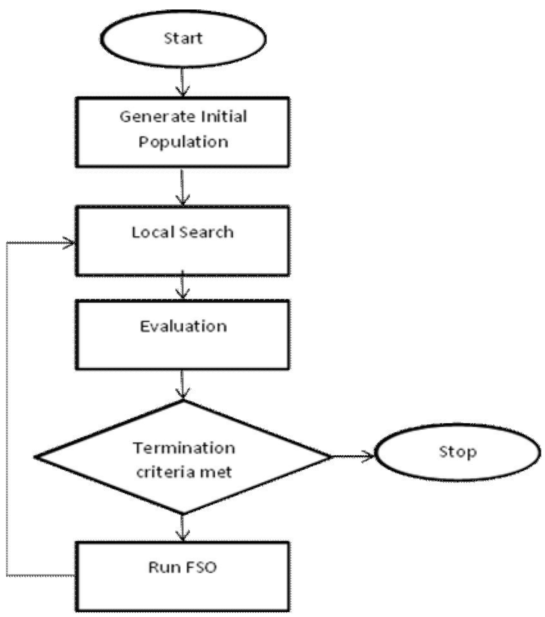

Figure 2: Flowchart of Hybrid Optimization

4. Results

Scheduling methods effectiveness is evaluated using evaluation metrics like makespan and flowtime. Makespan is the time the grid requires for finishing the latest task. The flowtime is defined as the total execution times for all tasks presented to the grid. The effectiveness of the proposed scheduling method is assessed and evaluated using makespan. The simulations were conducted using 40 clusters of resource clusters and 300 jobs. Simulations were conducted using the hybrid MFSO scheme of grid scheduling. Figure 3 shows the Makespan time vs. number of iterations achieved by the proposed MFSO.

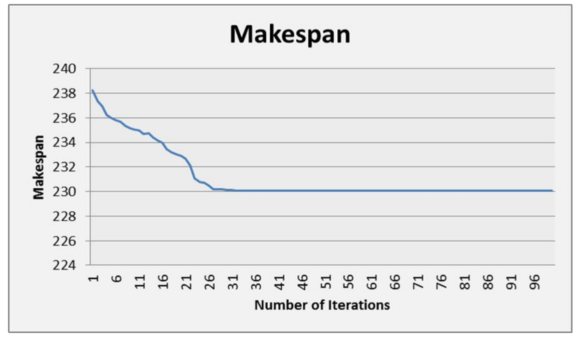

Figure 3: Makespan vs. Number of Iterations

It is observed for the graph that the makespan converges at 33rd iteration. The proposed MFSO achieves optimal scheduling efficiently.

5. Conclusion

This paper presents Memetic - Fish Swarm Optimization (MFSO) a hybrid algorithm that combines FSO with local search techniques. MFSO consists of two main mechanisms, a global mechanism in charge for global search of the search space, and a local one performing more refined search around potential solutions for problems on hand.The proposed MFSO is designed to achieve optimized scheduling in a grid environment. The effectiveness of the proposed scheduling method is assessed and evaluated using makespan. The simulations were conducted using 40 clusters of resource clusters and 300 jobs. Simulation results demonstrate the effectiveness of the proposed MFSO.

References

1. Berman, F., Fox, G., & Hey, A. J. (Eds.). (2003). Grid computing: making the global infrastructure a reality (Vol. 2). Wiley.
2. Rahman, M., Ranjan, R., Buyya, R., & Benatallah, B. (2011). A taxonomy and survey on autonomic management of applications in grid computing environments. Concurrency and Computation: Practice and Experience, 23(16), 1990-2019.
3. Foster, I., Kesselman, C., Nick, J. and Tuecke, S. "The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration", Technical Report, Open Grid Service Infrastructure WG, Global Grid Forum, 2002.
4. I. Foster and C. Kesselman (editors), The Grid: Blueprint for a Future Computing Infrastructure, Morgan Kaufmann Publishers, USA, 1999.
5. Iosup, A., & Epema, D. (2011). Grid computing workloads. Internet Computing, IEEE, 15(2), 19-26.

6. Hamscher, U. Schwiegelshohn, A. Streit, R. Yahyapour, *Evaluation of Job-Scheduling Strategies for Grid Computing*, in Proc. of GRID 2000 GRID 2000, First IEEE/ACM International Workshop, pp. 191-202, Bangalore, India, December 2000.
7. M, Arora, S.K. Das, R. Biswas, A Decentralized Scheduling and Load Balancing Algorithm for Heterogeneous Grid Environments, in Proc. of International Conference on Parallel Processing Workshops (ICPPW'02), pp.:499 – 505, Vancouver, British Columbia Canada, August 2002,
8. Zhang, L., Chen, Y., Sun, R., Jing, S., & Yang, B. (2008). A task scheduling algorithm based on pso for grid computing. International Journal of Computational Intelligence Research, 4(1), 37-43.
9. Abraham, A., Liu, H., Zhang, W., & Chang, T. G. (2006). Scheduling jobs on computational grids using fuzzy particle swarm algorithm. In Knowledge-Based Intelligent Information and Engineering Systems (pp. 500-507). Springer Berlin/Heidelberg.
10. Izakian, H., Ladani, B. T., Abraham, A., &Snasel, V. (2010). A discrete particle swarm optimization approach for grid job scheduling. International Journal of Innovative Computing, Information and Control, 6(9), 4219-4233.
11. Kashyap, R., & Vidyarthi, D. P. (2011). Security-Driven Scheduling Model for Computational Grid using Genetic Algorithm. In Proceedings of the World Congress on Engineering and Computer Science (Vol. 1).
12. Pinel, F., Dorronsoro, B., &Bouvry, P. (2010, April). A new parallel asynchronous cellular genetic algorithm for scheduling in grids. In Parallel & Distributed Processing, Workshops and Phd Forum (IPDPSW), 2010 IEEE International Symposium on (pp. 1-8). IEEE.
13. Xue, X., & Gu, Y. (2010). Global optimization based on hybrid clonal selection genetic algorithm for task scheduling. Journal of Computational Information Systems, 6(1), 253-261.
14. Schutze, O., Talbi, E. G., Coello, C. C., Santana-Quintero, L. V., &Pulido, G. T. (2007, April). A memetic PSO algorithm for scalar optimization problems. In Swarm Intelligence Symposium, 2007. SIS 2007. IEEE (pp. 128-134). IEEE.
15. Dawkins R. (1976): The Selfish Gene, (Oxford University Press, New York).
16. Goldberg, D. E., & Holland, J. H. (1988). Genetic algorithms and machine learning. Machine Learning, 3(2), 95-99.
17. Petalas, Y. G., Parsopoulos, K. E., &Vrahatis, M. N. (2007). Memetic particle swarm optimization. Annals of Operations Research, 156(1), 99-127.
18. X.L. Li, Z.J. Shao, J.X. Qian, ―An Optimizing Method Based on Autonomous Animats: Fish-swarm Algorithm‖, Systems Engineering - Theory & Practice, vol.006, no.11, pp.32-38, 2002.

1/8/2013