IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727,
PP 61-65
www.iosrjournals.org

A Genetic Based Task Scheduling Strategy for Distributed Environment

Sunil Kumar[1], Sumit Mittal[2], Manpreet Singh[3]
[1](Department of Computer Science& Engineering, Guru Nanak College for Girls, Sri Muktsar Sahib, India)
[2] (Department of Computer Science & Engineering, M. M. University, Mullana, India)
[3](Department of Computer Science & Engineering, M. M. University, Sadopur, India)

ABSTRACT: *With the emergence of Cloud computing and Grid Computing, Distributed Scheduling (DS) problems have attracted attention by researchers in recent years. Distributed scheduling requires an uneven distribution of tasks on individual processors. Different heuristic based algorithms to perform the task scheduling have been proposed by the various researchers. This paper offers a new strategy for task scheduling which is based on genetic algorithm. The final results show that proposed algorithm is an improvisation over Standard Genetic Algorithm in terms of makespan and resource utilization.*
Keywords – *Distributed Scheduling, Genetic Algorithm, Meta- Heuristic, Makespan.*

I. INTRODUCTION

The last two decades have witnessed great breakthrough in the use of Distributed system in a wide variety of applications consequently with improvements in different technologies, theory and software tools [1]. A distributed framework is a collection of autonomous hosts which are associated through a computer network. Distributed systems have various advantages which make them useful such as shareability, expandability, improved performance, reliability and availability. However, there are also some major issues as well like scheduling, security and multiple point of failure [2]. Scheduling is responsible for efficient utilization of resources in terms of minimum execution time (make-span) and load balancing. Many algorithms have been proposed to cater the need of task scheduling on distributed resources. Some of the new challenges in distributed systems for scheduling applications are mentioned below [3]:

- All the existing resources are in shared state.
- Shared data must be protected from various attacks.
- Synchronization must be done correctly to prevent data corruption.
- The network is heterogeneous because of different network infrastructures. Hence, the given resources may not perform same function for any given task.

To overcome the challenges of such distributed environment, an effective task scheduling algorithms is needed. There are two major categories of task scheduling approaches: Heuristic and Meta-Heuristic approaches.

- Heuristic Approach

In the heuristic scheduling, there are various well-known scheduling algorithms such as Min-Min, Max-Min, sufferage and Max-Min<>Min-Min. The Min-Min heuristic [4] schedule a task having minimum completion time is assigned to the faster machine first followed by task with longer time to get the minimum make-span. The only dissimilarity which exists between the two algorithms is that the Max-Min heuristic [4] approach works on the concept of priority. It assigns the level of priority to the task according to its execution time. The task having maximum completion time is assigned to the fastest machine. On the other hand, Max-Min<>Min-Min algorithm combines the Max-min and Min-min efficiencies. Along with this, it considers the property of grid resources during task scheduling [5].

- Meta- Heuristic Approach

This approach is used to generate a heuristic approach which provides a good solution for a problem. There are various meta- heuristic approaches, but well known approaches are Genetic Algorithm (GA) and Simulated Annealing. Genetic Algorithm is selecting the best solution from the randomly generated initial population [6]. Simulated Annealing is a single point search meta-heuristic. It gets the solution in the nature of GA by using point to point method [7].

This paper works in the area of optimizing task scheduling in distributed environment. For this, a modified Genetic Algorithm is used that works on improving the minimum execution time of tasks and balanced load on every machine.

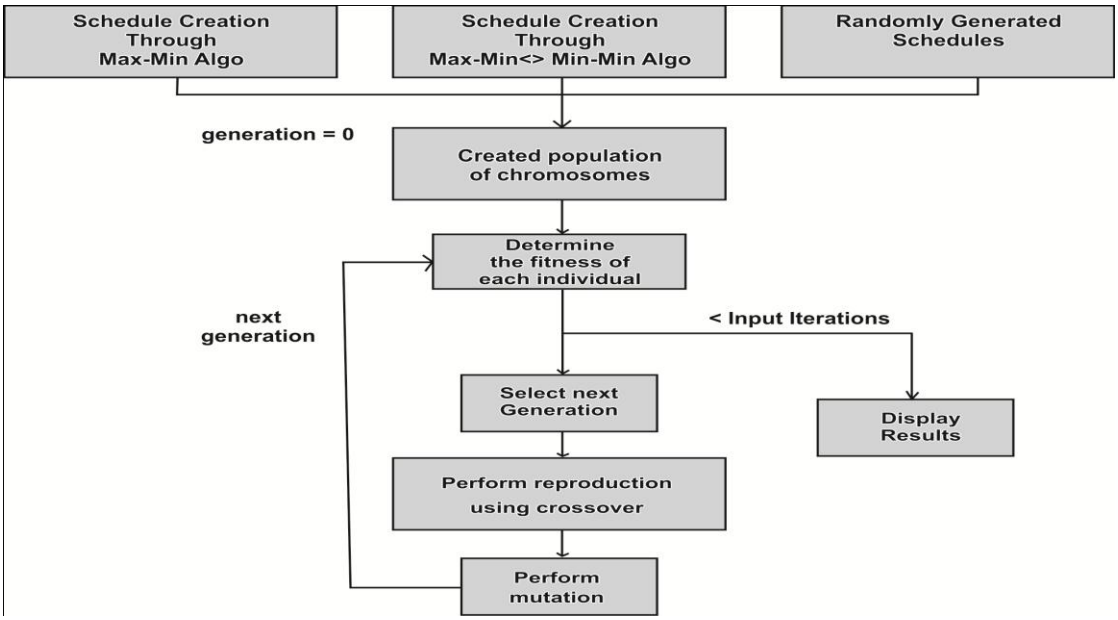
II. RELATED WORK

As it is well known, the general problem of scheduling is NP-complete. A large number of algorithms have been designed to schedule tasks in distributed environment. There are number of scheduling algorithms which works well and are very effective but due to some reasons such as high cost of communication they are not been used in some distributed environments commonly known as grid and cloud environment [8] [9]. QoS is a basic parameter for calculating the success of any algorithm. In [10], authors proposed a solution for QoS based scheduling in Grid-JQA architecture. To evaluate QoS parameter, an aggregation method is applied by the authors. The aggregation formula is a combination of parameters with weighting factors. Although this algorithm performance is effective in various environment but still it is counted as an unpractical solution. Cloud computing is a fast growing technology and needs some good task scheduling algorithms. A scheduling based on the concept of Genetic Algorithm for cloud environment has been proposed in [11]. In this, in every cycle, task scheduler calls the GA scheduling function which then generates a set of task and then evaluates it. Another task scheduling algorithm for the cloud environment has been proposed in [12]. This algorithm is based on ABC (Activity Based Costing). In this approach, every job is provided a priority as per the importance of its activity and a queue is designed. After placing tasks in queues, the task from high priority queue is selected and when this task gets finished, the first task of the mid priority queue is shifted to high priority queue and in this way tasks from all the queues starts execution. In similar to this, authors of [13] proposed a new algorithm named as QoS priority grouping scheduling. Few factors which are considered in this scheduling algorithm are deadline and task acceptance rate. In [14], authors proposed a new algorithm which is works on the concept of Load Balancing Ant Colony Optimization (LBACO) algorithm for cloud environment.

III. PROPOSED ALGORITHM

The purpose is to design such a schedule of tasks in which execution time should be minimum and simultaneously it should balance the load of resources. For this purpose, the merits of genetic algorithm is merged with two heuristic algorithms (Max-Min and Max-Min<>Min-Min). A genetic algorithm (GA) is a well known search technique. It is used in computing to find an approximate solution to optimization and search problems. Genetic algorithm is a set of instructions that is repeated to solve a problem. It follows steps inspired by the biological processes of evolution. It is considered as a particular class of evolutionary algorithms that use techniques inspired by evolutionary biology such as inheritance, mutation, selection, and crossover as shown in Fig1.

Fig.1. Structure of Proposed Algorithm



The basic steps of proposed algorithm are creation of initial population, crossover, mutation and fitness criterion to get the optimum results.

- Creation of Initial Population

For the creation of initial population of meta-heuristic, one schedule from Max-Min, another from Max-Min<>Min-Min and remaining on random basis are selected.

- Crossover

A two point crossover is used to generate a child chromosomes having better quality.

- Mutation

Mutation is used to ensure that the all obtained individuals are not same. In this algorithm, rebalancing technique is used as a mutation operator, as the aim is also to balance the load between machines.

- Fitness Criterion

The minimum make-span is the key fitness criterion while the load at resources is also considered for schedule selection.

- Termination Condition

In the proposed algorithm, when the algorithm satisfies the predefined number of iterations, it gets terminated.

IV. IMPLEMENTATION & RESULTS

A best scheduling algorithm is one that gives the better result for various tasks with constraint of resources. The simulation parameters used to check the performance of proposed algorithm is shown in Table 1.

Table.1. Simulation Parameters

Parameter	Value
Total Tasks	10-60
Machines used	10
Defined Iterations	1-150
Crossover Type	Two Point Crossover
Mutation Type	Rebalancing
Termination Condition	Number of Iteration

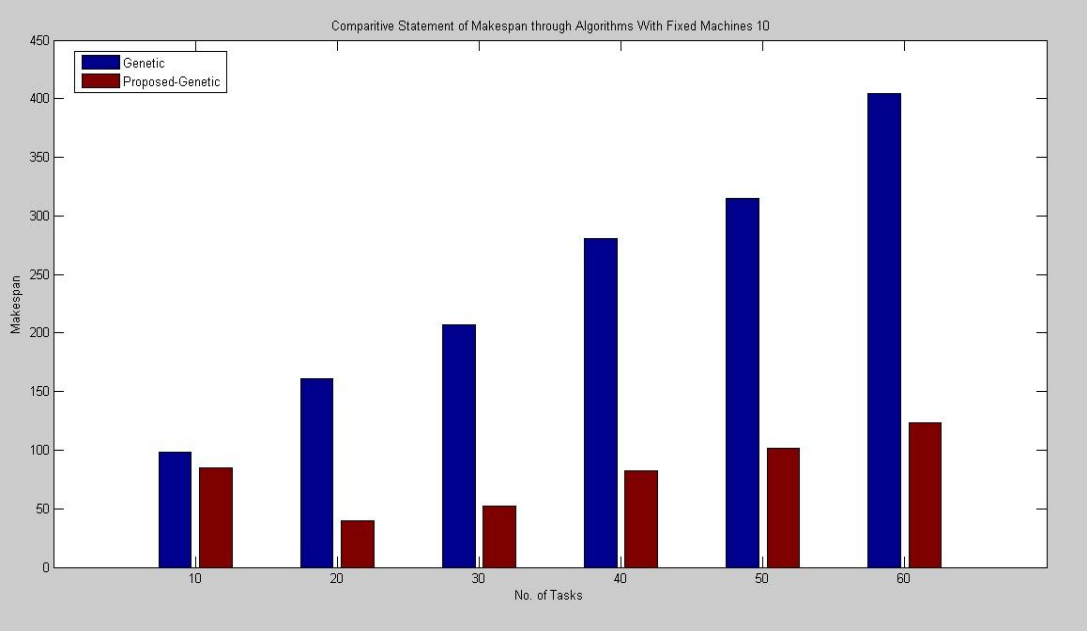


Fig.2. Performance comparison in terms of makespan

Fig 2 shows the makespan calculated with fixed number of machines 10 and variant number of tasks (10-60) for standard GA and proposed algorithm.

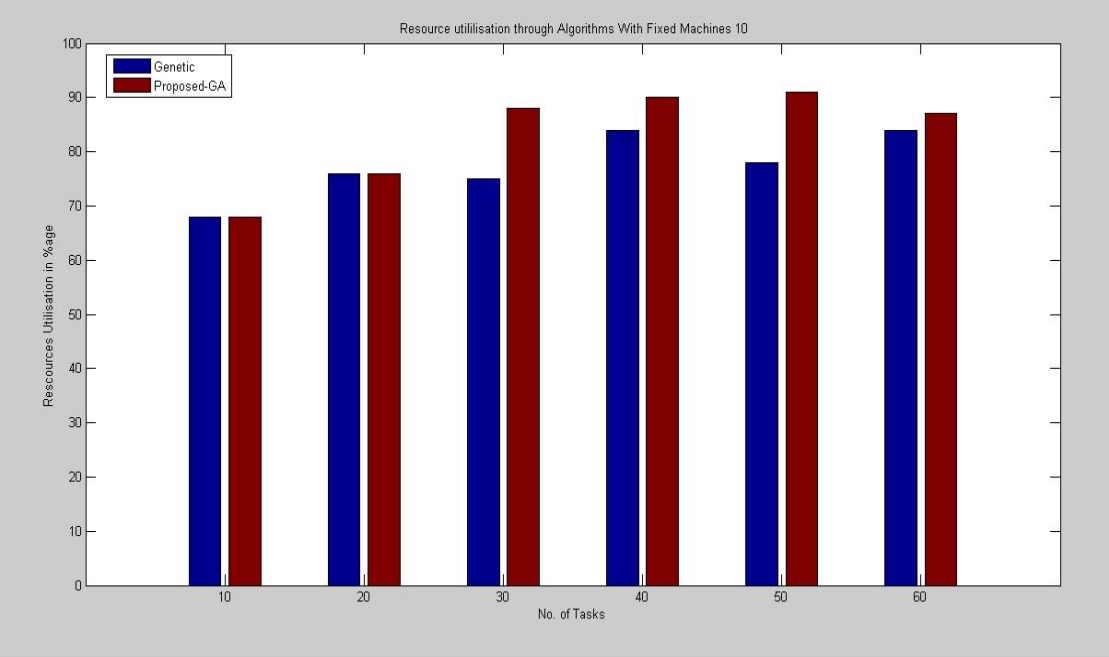


Fig.3. Performance comparison in terms of resource utilization

Fig 3 shows the resource utilization (Load Balance) for the same. The results highlights that the proposed algorithm maintains good performance under heavy loads also (increase in number of tasks).

V. CONCLUSION

It is concluded that proposed algorithm outperforms the standard genetic algorithm mainly with respect to makespan. In present era, security is an important issue so further research will be oriented towards the implementation of efficient and secure task scheduling algorithm for distributed environment.

REFERENCES

[1] A. Y. Zomaya, Parallel and distributed computing (ed. New York, McGraw-Hill, 1996).
[2] J. Naik k, V. Kumar and N. Satyanarayana, Scheduling Tasks Most Suitable Fault tolerant Resource for Execution in Computational Grid, International Journal of Grid and Distributed Computing, 5, 2012, 121-131.
[3] J. Blythe, S. Jain, E. Deelman, Y. Gil, K. Vahi, A. Mandal and K. Kennedy, Task Scheduling Strategies for Workflow-based Applications in Grids, International Symposium on Cluster Computing and Grid (CCGrid), IEEE, 2005, 759-767.
[4] M. Maheswaran, S. Ali, H.J. Soegel, D. Hensgen, and R Freund, Dynamic Matching and Scheduling of a class of Independent Tasks onto Heterogeneous Computing Systems, 8th Heterogeneous Computing Workshop, San Juan 1999,30-44.
[5] M. Singh and P. K. Suri, QPS Max-Min<>Min-Min: A QoS Based Predictive Max-Min, Min-Min Switcher Algorithm for Job Scheduling in a Grid, Information Technology Journal, 7(8) , 2008 1176-1181.
[6] S.H. Woo, S. B. Yang, S. D. Kim and T.D. Han, Task scheduling in distributed computing systems with a genetic algorithm, High performance computing on the information highway, IEEE, Seoul,1997,301-305
[7] S. Zheng, W. Shu and L. Gao, Task scheduling using Parallel Genetic simulated Annealing Algorithm, International conference on Service operations and Logistics and Informatics, IEEE, Shanghai, 2006, 46-50.
[8] H.Topcuoglu, S. Hariri, and M.Y. Wu, Performance-Effective and Low-Complexity Task Scheduling for Heterogeneous Computing, IEEE Trans. Parallel and Distributed Systems, 13(3), 2002. 260- 274.

[9] D. Amalarethinam and F. Selvi, A minimum makespan grid workflow scheduling algorithm, International Conference on Computer Communication and Informatics, IEEE, Coimbatore, 2012,1-6.
[10] L. Khanli and M. Analoui, Grid_JQA: A QoS Guided Scheduling Algorithm for Grid Computing, 6th International Symposium on Parallel and Distributed Computing (ISPD'07), IEEE, Hegenberg, 2007,34-36.
[11] S. Jang, T. Kim, J. Kim and J. Lee, The Study of Genetic Algorithm-based Task Scheduling for Cloud Computing, International Journal of Control and Automation, 5(4), 2012,157-162.
[12] Qi Cao, Z. Wei and W. Gong, An Optimized Algorithm for Task Scheduling based on Activity based Costing in Cloud Computing, 3rd International Conference on Bioinformatics and biomedical Engineering, IEEE, Beijing, 2009, 1-3.
[13] F. Dong, J. Luo, L. Gao and L. GE, A Grid Task Scheduling Algorithm Based on QoS Priority Grouping, 5th International Conference on Grid and Cooperative Computing (GCC'06), IEEE,Hunan, 2006, 58-61.
[14] Li Kun, Xu Gaochao, Zhao Guangyu, Dong Yushuang and D. Wang, Cloud Task Scheduling Based on Load Balancing Ant Colony Optimization, 6th Annual China grid Conference, IEEE, Liaoning, 2011, 3-9.