CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 CONCLUSION

In this thesis, the problem of task scheduling on Heterogeneous Distributed Computing System (HDCS) is studied. This problem is considered to be a multiobjective optimization problem. The two objectives namely Makespan and reliability are considered in this work. Since static task scheduling problem is a NP-hard problem, several heuristics and metaheuristics have been proposed in the literature. In this research work, two different methods for solving the task scheduling problem, one based on heuristic approach and the other based on metaheuristics is applied. The heuristic based scheduling algorithms are efficient since they narrow the search down to a very small portion of the solution space by means of greedy heuristics. However, due to the greedy nature, heuristics based approaches are not likely to produce consistent results on a wide range of problems. In order to obtain schedules of better quality, many well known metaheuristics, including Genetic Algorithm(GA), Ant Colony Optimization(ACO), Simulated Annealing (SA), Particle Swarm optimization(PSO) and others have been adopted.

To perform a comparison of different methods to solve the task scheduling problem, a task graph generator tool is developed. The tool can be used to generate random task graphs, resource graphs and real application task graph with various characteristics.
A Preliminary version of the tool is deployed in the internet (http://sourceforge.net/projects/taskgraphgen/). The statistics shows a convincing number of downloads of the tool by the researcher community. The scheduling algorithms along with the performance metrics are also available in the tool.

In this research, the traditional list scheduling heuristic, Heterogeneous Earliest Finish Time (HEFT) and Reliable Dynamic Level Scheduling (RDLS) algorithm are combined to propose a new algorithm, Improved Reliable HEFT (IRHEFT). This improved algorithm extends the traditional list scheduling heuristics to optimize both the objectives simultaneously. In addition, IRHEFT includes the load difference factor to improve the reliability. The performance of the algorithm is compared with the existing bi-objective RHEFT and RDLS algorithms. The comparison shows that there is a significant increase in the reliability factor due to the introduction of load difference factor in the processor selection phase.

Multiobjective evolutionary algorithms (MOEAs) are well suited for solving multiobjective optimization problems. The Multiobjective Genetic Algorithm (MOGA) and Multiobjective Evolutionary Programming (MOEP) both based on non-dominated sorting are used to solve the task scheduling problem considering the two objectives of makespan and reliability. The main advantages of the proposed MOEAs include (i) obtaining a pareto optimal set of solutions compared to the single solution with the weighted-sum method for solving the multiobjective optimization problem (ii) the objectives can be considered as such without normalization (iii) obtaining a set of non-dominated solutions at reduced computational times. Further, the convergence and diversity of the obtained non-dominated solutions are computed and it shows that MOGA is best suited for the task scheduling problem.
The performance of MOEAs can be improved by hybridization with local search. Hybridization of MOEAs improves the convergence speed to Pareto front. The two well-known Pareto-dominance based MOEAs: the Non-dominated Sorting Genetic Algorithm (NSGA-II) and improved Strength Pareto Evolutionary Algorithm (SPEA2) are implemented and hybridized with local search procedure. Simple Neighborhood Search (SNS) algorithm is used as the local search algorithm. The weighted-sum and Pareto dominance based MOEA for the task scheduling problems are compared with the hybrid version of these approaches. A simple triangular membership function is used to select the best weight combination for multiple objectives. The non-dominated solutions of NSGA-II, SPEA2 to TSPHS for the same real time task graph is obtained. The simulation results show that better results were obtained by the HNSGA-II especially in terms of diversity of obtained non-dominated solutions.

6.2 FUTURE WORK

The current trend in designing scheduling algorithms is with respect to user’s demands, that is, provide Quality of Service (QoS) based scheduling and this research study can be considered to be a step in that direction. There are other objectives such as deadline, robustness, energy consumption which can be considered for multiobjective task scheduling. Although individual solutions to each of these problems have been proposed, growing size of heterogeneous computing system makes it difficult to find efficient solutions to these problems that can be implemented in a practical system considering multiple objectives. The performance of the approaches studied in this research can be tested using the data obtained from real distributed computing environment. Incorporating task duplication heuristic in the metaheuristics can be considered to improve the schedule quality.
In implementing a scheduler, other challenging problems, including task profiling for a given application, analytical benchmarking of the machines in the system, resource management, etc., need to be addressed. The predictability of the performance of an application under a scheduling algorithm depends on the quality of the information that is provided to the scheduler about the application and the heterogeneous system.