CHAPTER 2

LITERATURE SURVEY

2.1 INTRODUCTION

Heterogeneous distributed computing systems (HDCS) are the emerging platforms for executing computationally intensive applications with diverse computing needs. They are computing platforms where hardware or software components located at networked computers communicate and coordinate their actions by passing messages (Coulouris et al 2005). It enables user to access services and execute applications over a heterogeneous collection of computers and networks.

In recent years, HDCS are well suited to meet the computational demands of large, diverse groups of tasks. The problem of mapping (including matching and scheduling) tasks and communications is a very important issue since an appropriate mapping can truly exploit the parallelism of the system thus achieving large speedup and high efficiency (Braun et al 2001). It deals with assigning (matching) each task to a machine and ordering (scheduling) the execution of the tasks on each machine in order to minimize some cost function.

Makespan or the schedule length is the most common cost function. The general problem of optimally mapping tasks to machines in HDCS has been shown to be NP-complete (Garey and Johnson 1979). Even problems constructed from the original mapping problem by making simplified
exploration of the search space. The algorithm was evaluated for set of benchmark problems with varying CCR and computation costs.

A hybrid Genetic algorithm was proposed by (Varghes et al 2009) for scheduling real-time tasks on multiprocessor system. The algorithm combined the Earliest Deadline First scheduling with GA. The algorithm shows schedules with higher processor utilization can be generated by the hybrid algorithm.

A heuristics-based bee colony algorithm for the multiprocessor scheduling problem was proposed by (Rebreyend et al 2010). The algorithm combines the bee-algorithm with a heuristic in order to produce quickly good solutions. The chosen heuristic uses a greedy approach and hybridization was done using the indirect representation. The heuristic is a list heuristic and the bee algorithm finds the best order for the ordered list of tasks used by the heuristic.

A heuristic based hybrid genetic algorithm for heterogeneous multiprocessor scheduling was proposed by (Wen et al 2010). The algorithm extends the traditional GA-based approaches in three aspects. First, it incorporates GA with Variable neighborhood Search (VNS), to enhance the balance between global exploration and local exploitation of search space. Second, two novel neighborhood structures, in which problem-specific knowledge concerned with load balancing and communication reduction is utilized. Third, the use of GA with upward ranking heuristics is presented. The results prove that hybrid algorithm outperforms the pure version.

Hence, these observations have motivated this research work to apply heuristic and metaheuristics approaches to the task scheduling problem. The need for a parametric task graph generator was essential to apply and compare the algorithms. To do the comparison of the algorithms a task graph generation tool is also developed.