ABSTRACT

The need for faster computers is driven by the demands of both computation-intensive applications in science and engineering and data-intensive applications in commerce. The main motivation of parallelization of any sequential algorithm is the desire to reduce the total execution time of the algorithm. Parallel computing provides a way to reduce the time taken to execute a task by dividing the task between multiple computers that can work simultaneously to complete the job. As the cost of dedicated parallel machines is high, computing on a cluster of workstations is a viable alternative for a wide range of engineering applications. The thesis aims at finding cost effective parallel implementation of algorithms on a cluster to solve complex applications.

The first part of the thesis addresses the issues that arise in the parallel implementations. The major issues of concern are load balancing, a widely used technique to balance the workload among multiple processors and fault tolerance, the ability to tolerate failures. The proposed load balancing approach is based on a central scheduler with hybrid genetic algorithms and genetic local search methods. Replication approach has been proposed to improve the performance and fault tolerance of a multiprocessor system.

The second part of the thesis deals with the cost effective parallel formulation of the implicitly parallel algorithms like genetic algorithms, data mining algorithms and neural networks to solve the complex computation and data intensive applications.

Coarse grained parallel genetic algorithm with migration has been used to solve the graph coloring and job shop scheduling problems. The performance of the parallel genetic algorithm has been evaluated using three crossovers for various benchmark graphs. Parallel genetic algorithm has been employed to solve both independent and dependent jobs from the standard benchmark task graphs.

A new parallel Buddy Prima algorithm has been proposed for maximal frequent itemset mining of very large databases. Parallel itemset tree has been implemented for targeted querying of large dynamic databases. The proposed parallel formulation of the enhanced backpropagation network based on training set and training session parallelism has been employed to solve the classification problems. Parallel Self organizing maps are used for image compression of industry standard images.

The performance of the parallel algorithms has been evaluated in terms of speedup, execution time and the degree of parallelism. Experimental results indicate the effectiveness of the proposed parallel algorithms over the sequential ones.