ABSTRACT

This research investigates certain load balancing algorithms in a distributed computer system to distribute jobs among processors in such a way that each processor has approximately the same workload. The algorithms are based on certain new approaches for migration of jobs. A routing algorithm and a directed network algorithm are introduced for migration of jobs with the aim to reduce the communication overhead. In addition to these algorithms, the thesis presents a partitioning approach and a partnership approach to handle the migration problem.

The partitioning approach basically splits the original system into a number of components of a specified size and balances the load first within each component and then among all components. Three efficient partitioning techniques are suggested: hierarchical, flat and mixed. The hierarchical approach partitions the original system into a number of hierarchical levels. The policy for splitting the system is based on the previous status of the processors. By the flat partitioning approach, the structure of the original network is split into different blocks. The mixed partitioning approach is implemented with a combination of hierarchical and flat approaches. These techniques are useful for a large system.

In the partnership approach each processor is considered to be a member of a system of families and is associated with partner processors belonging to the same family system. The load balancing algorithm is based on the idea that each processor may be made to cooperate with its own partner or partners in the family for balancing the load. Various approaches are suggested for making partners of a processor. Based on these approaches different load balancing algorithms: sequential, parallel and heuristic, are suggested for balancing the load among members of the system of families.

The effectiveness of the proposed load balancing algorithms is studied by simulation for which a general-purpose distributed system simulator is developed. While investigating the performance of the proposed load balancing algorithms by simulation, both the communication cost and the processing cost are considered. The thesis also presents a static load balancing approach for a real time distributed system which can reduce the loss of jobs considerably.