Improved Load Balancing Strategies for Distributed Computing Systems

Author
S. SELVAM

Supervisor
Prof. Moinuddin
Department of Electrical Engineering
Jamia Millia Islamia

The main problem addressed in this thesis is performance improvement of load balancing (LB) strategies in distributed computing systems (DCS). A distributed computing system consists of more than two dispersed autonomous processor nodes located geographically over an area. These processors are interconnected by communication links between them. A distributed process is defined as a set of tasks which together work towards a common goal. Methodologies have been developed for the assignment of tasks onto the processor nodes of distributed computing systems.

The thesis proposes improved load balancing strategies for static and dynamic load balancing in distributed computing system. The following heuristics are proposed for static load balancing in task allocations , estimation of communication overhead and transfer delay.

- A heuristic is proposed for faster task allocation in state space search. In this , the allocation of more than one task is checked first in the search tree. The time spent on the additional complexity is relatively less to the time gain in achieving the final search node of task allocation.

- In a redundant distributed computed system, the completion time of the metatask can be minimised by multiprocessing all the tasks allotted to a processor node. The processor node contains ‘n’ number of redundant elements depending upon the level of redundancy. By multiprocessing all the tasks of a node, there can be some considerable gain in the completion time during the peak hours. However, the negligible deviation in the total probability of the system is to be traded off for the gain in completion time.

- The estimation of communication overhead and the transfer delay are the critical factors in load balancing. The maximum communication overhead and the transfer delay are estimated by the proposed method. The trends of the estimated values with the number of processor are studied using different job arrival patterns.

The thesis also proposes improved heuristics in dynamic load balancing through diffusion of workloads, hydrodynamic sharing, preemptive load balancing using process lifetime and preemptive load balancing using prediction techniques as below;

- In distributed computing systems, the strategic nature of the load change forces the application of dynamic load balancing and that leads to the preference of diffusion schemes. In the present work, the applications of different schemes on distributed computing system are studied. The variation in the load balancing effectiveness of gradient method (GM), sender initiated diffusion (SID) and receiver initiated diffusion (RID) are studied against the granule size.

- The hydrodynamic diffusion flow principle may be suitable for a distributed computing system where lesser computation and message overhead are evident. The application of hydrodynamic load balancing strategy on distributed computing system is applied in networks with different processors and the corresponding communication overhead and delay are studied.

- The possibilities of increasing the throughput and minimisation of completion time are also achieved in distributed computing system by preemptive load balancing where all prolonging tasks on slow processors are transferred to the faster ones. The performance gains are compared with a non-preemptive dynamic load balancing strategy.

- Adaptation of process lifetime as a measure of process longevity is also considered in a predication based dynamic load balancing heuristics to attain better performance. The performance variation of the proposed heuristic is studied against a non-preemptive migration strategy.