Abstract

Proportional fair scheduling (Proportionate Fair (Pfair) [22], Early Release Fair (ERfair) [7], etc.) has a number of salient features which make it an effective resource management strategy for the integrated scheduling of complex real-time applications in systems including embedded infotainment, automotive control, avionics systems, network link allocation servers, etc. These features include: i. ability to provide temporal isolation and a minimum guaranteed Quality of Service (QoS) to each client task, ii. ability to fully utilize the resource capacity, iii. graceful degradation for all client tasks in times of overload, etc. However, most of the proportional fair scheduling approaches possess high scheduling overheads (unrestricted migrations and context switches together with relatively high computational complexity) making it infeasible in many practical scenarios especially in real-time systems with stringent resource and performance constraints.

This dissertation presents a few novel ideas towards low-overhead and accurate proportional fair scheduling of dynamic recurrent tasks for real-time multiprocessor systems. The work also endeavors to develop proportional fair algorithms that can work effectively under practical situations like limited power, transient overloads, etc. We have presented some theoretical and experimental analysis of the methods to show their efficacy. These should provide scope for developing application-specific schedulers under various fairness-speed-power-overload requirements.

Our first contribution has been the development of algorithms that produce significant reduction in the number of inter-processor task migrations and context switches but still retain the optimal fairness and schedulability properties of ERfair [7]. A second significant contribution has been the conceptualization and formalization of the frame based (periodic partitioning followed by global resynchronization) proportional fair multiprocessor scheduling that achieves high reductions in task migrations albeit at the cost of occasional slight deviations from perfect fairness. We have also developed flexible partition oriented frame based schedulers that are able to reduce the number of migrations in current ERfair schedulers by upto 100 times while simultaneously providing low and bounded fairness distortion along with typically lower scheduling complexities than ERfair. Our final two contributions include design and development of a procrastination based energy-aware ERfair scheduling algorithm with large energy savings and an *a priori* overload handling strategy using a low-overhead admission control technique for ERfair. Experimental results have demonstrated the versatility and efficacy of the proposed approaches.

Keywords: Proportional fairness, ERfair scheduling, Task migration, Context switches, Job procrastination, Overload handling