Optimizing Job Scheduling on Multicore Computers

Supervisor: Dr. Ligang He
Department of Computer Science, University of Warwick

Reserved for Natalya Bondarenko

Research objectives: the main aim of this project is design and develop the GPU-based algorithm to find the optimal co-scheduling solutions

Summary of the background: It is common nowadays that multiple cores reside on the same chip and share the on-chip cache. Resource sharing may cause performance degradation of the co-running jobs. Job co-scheduling is a technique that can effectively alleviate the contention. Many co-schedulers have been developed in the literature, but most of them do not aim to find the optimal co-scheduling solution. Being able to determine the optimal solution is critical for evaluating co-scheduling systems. Moreover, most co-schedulers only consider serial jobs. However, there often exist both parallel and serial jobs in some situations. In paper [1], a graph-based method is developed to find the optimal co-scheduling solution for serial jobs, and then the method is extended to incorporate parallel jobs. The extensive experiments have been conducted to evaluate the effectiveness and efficiency of the proposed co-scheduling algorithms. The results show that the proposed algorithms can find the optimal co-scheduling solution for both serial and parallel jobs. It is possible to parallelize the proposed co-scheduling algorithms to speedup the scheduling process using CUDA, which is a programming model designed for GPUs. GPU's offer a finer level of data parallelism than a CPU, with which large speedups are possible.

Prospective deliverables: GPU-based algorithm of optimal co-scheduling solutions for multicore computers

References:

[1] Huanzhou Zhu, Ligang He and S. A. Jarvis, "Optimizing Job Scheduling on Multi-core computers", in IEEE 22nd International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems