Session 6: Performance Issues

Session Chair: K. Waldschmidt

To what extent is “Computer Science” an empirical science? Unlike physics or biology, in Computer Science we study creatures of our own design, artifacts of own making — be these circuits, computer architectures, software structures or world-wide webs. However, as everyone active in this field knows, the observable characteristics of these creatures always retain the ability to surprise, and often to shock, their creators. We are not free to propose any designs we fancy. We are constrained by realities which we cannot understand without the techniques common to every empirical science: carefully constructed experimentation, measurement and analysis.

The five papers in this session analyze the performance of systems ranging from the microscopic to the planetary in scale. The first two papers delve into the realm of ILP (Instruction-Level Parallelism) in Multiprocessor and Multithreaded processors, and examine two areas which have received much attention recently, prefetching and branch prediction.

The first paper, “Hardware Prefetching in Bus-Based Multiprocessors: Pattern Characterization and Cost-Effective Hardware”, by M. Garzarán, J. Briz, P. Ibáñez and V. Viñals, considers how the number of processors and the memory access patterns in a program influence the relative performance of various prefetching mechanisms in a bus-based SMP (Symmetric MultiProcessor). The authors propose a cost-effective hardware prefetching solution for implementation on modest-sized multiprocessors.

The second paper, “Evaluating the Effects of Branch Prediction Accuracy on the Performance of SMT Architectures”, by R. Goncalves, M. Pilla, G. Pizzol, T. Santos, R. Santos and P. Navaux, studies the impact of varying the accuracy of branch prediction on both superscalar and SMT (Simultaneous MultiThreaded) architectures. They were able to determine that, while increasing branch prediction accuracy always helps the performance of superscalar architectures, it often does not help the performance of SMTs.

The third paper, “Predicting the Time of Oblivious Programs”, by J. Gonzalez, C. Leon, F. Piccoli, M. Printista, C. Rodriguez, moves from the microscopic world of ILP into the more abstract realm of the BSP (Bulk Synchronous Programming) model. The accuracy of performance predictions made with this model, extended to incorporate “oblivious synchronization” and “machine partitions”, is analyzed for two algorithms. The performance of one of the algorithms could be predicted with less than 3 percent error, but the other algorithm’s observed performance was over 20% less than that predicted by the model.

The remaining two papers focus on large, distributed systems. The fourth paper, “Static and Dynamic Scheduling Algorithms for Scalable Web Server Farm”, by E. Casalicchio and S. Tucci, analyzes both the demands placed on popular web sites and the ability of different scheduling techniques to meet these demands. Their analysis shows that, depending on the content of the web services provided, dynamic scheduling may be completely sufficient to meet even heavy demand peaks.

The fifth and final paper, “Implementing On-line Techniques to Allocate File Resources in Large Distributed Systems”, by E. Pagani and G. Rossi, studies the behavior of algorithms which dynamically relocate files nearer to the set of users which are currently accessing them. The authors characterize the conditions under which such a dynamic allocation algorithm is advantageous.