Load Balancing Networks*

Sarantos Kapidakis†
Department of Computer Science
University of Crete
Heraklion 71110, Greece
and
Institute of Computer Science
Foundation for Research and Technology – Hellas
Heraklion 71110, Greece

Marios Mavronicolas‡
Department of Computer Science
University of Cyprus
Nicosia 1678, Cyprus

Abstract

Designing a good *task allocation* algorithm faces the challenge of allowing high levels of *throughput,* so that tasks are executed fast and processor parallelism is exploited, while still guaranteeing a low level of *memory contention,* so that performance does not suffer because of limitations on processor-to-memory bandwidth. In this work, we present a comparative study of throughput and contention guarantees provided by *load balancing networks,* a new class of distributed, asynchronous algorithms for real-time task allocation in shared memory multiprocessors. Load balancing networks generalize *balancing networks* [1] to accomodate tasks with varying completion times.

On the theoretical side, we formulate precise and crisp definitions for capturing the quality of load balancing provided by general task allocation algorithms; we use these definitions for formally evaluating the throughput performance of specific constructions of load balancing networks that we propose. Furthermore, we introduce a formal, complexity-theoretic measure of contention incurred by tasks with varying completion times, and use it to analyze the contention performance of these constructions. This measure formally generalizes one proposed earlier by Dwork *et al.* [2]. Our theoretical results display precise and subtle trade-offs between throughput and contention performances for load balancing networks.

On the practical side, we propose an experimental platform for evaluating the actual performance of load balancing networks through a series of carefully designed experiments that simulate these networks on real shared memory multiprocessor machines. Our experimental approach encompasses a rigorous methodology for randomly generating tasks that are not merely "random", but rather belong to common classes of tasks such as *periodic* and *sporadic.* Our experimental results reveal that load balancing networks substantially outperform in performance classical, centralized methods for task allocation in both throughput and memory contention.

*This work has been supported by ESPRIT III Basic Research Project # 8144 (LYDIA – Load Balancing on High-Performance Parallel and Distributed Systems).

†E-mail address: sarantos@csi.forth.gr

‡Part of the work of this author was performed while visiting Institute of Computer Science, Foundation for Research and Technology – Hellas, Greece. Partially supported by funds for the promotion of research at University of Cyprus (research project "Load Balancing Problems on Shared Memory Multiprocessor Architectures"). E-mail address: mavronic@csi.forth.gr

References

[1] J. Aspnes, M. Herlihy and N. Shavit, "Counting Networks and Multi-Processor Coordination," *Proceedings of the 23rd Annual ACM Symposium on Theory of Computing,* pp. 348–358, May 1991.

[2] C. Dwork, M. Herlihy and O. Waarts, "Contention in Shared Memory Algorithms," *Proceedings of the 25th Annual ACM Symposium on Theory of Computing,* pp. 174–183, May 1993.