Approximation Algorithms for Scheduling Parallel I/O Operations *

Ravi Jain†
Applied Research
Bellcore

John Werth
Dept. of Computer Sciences
Univ. of Texas at Austin

Extended Abstract

The performance of parallel computers for many interesting classes of applications is often limited by the speed of data transfers rather than the speed of computation. This observation has motivated the argument that the performance of a parallel computer system should be measured in terms of its data transfer rates, both within the system and across a network, instead of the peak floating-point computation rate [9, 7]. This data transfer bottleneck is especially apparent in the case of I/O data transfers. Technology and application trends indicate that this I/O bottleneck is likely to become increasingly important in the future [1, 2, and references therein],

While the I/O bottleneck has been receiving increasing attention in the past few years, most solutions have focused on improving the performance of a few components of the parallel computer system. Typically, attention has been paid to improving the performance of I/O devices using fairly *low-level* parallelism in techniques such as disk striping and interleaving [8, 1, and references therein]. Widely applicable solutions, however, will require an integrated approach which addresses the problem at multiple system levels, including applications, systems software, and architecture. We have proposed that within the context of such an integrated approach, scheduling parallel I/O operations will become increasingly attractive and can potentially provide substantial performance benefits [5, 4]. We have developed a family of algorithms for scheduling parallel I/O operations under various architectural and logical constraints in the context of a general model for specifying scheduling problems [6, 3, 2].

In this talk we describe a simple parallel I/O scheduling problem and show that it can be represented as the problem of edge-coloring a bipartite graph, i.e., to color the edges so that adjacent edges receive different colors. We also describe a related problem, the constrained edge-coloring problem, in which each color may be used to color at most k edges. Optimal algorithms for these problems use the minimum number of colors to color the edges. We consider approximation algorithms for constrained and unconstrained edge-coloring of bipartite graphs. For the unconstrained problem, bounds on the number of colors used by the approximation algorithms to color the graph are known; we show that these bounds are tight. For the constrained problem we obtain bounds on the number of colors used by the approximation algorithms, and their time complexities.

*This research was performed while the first author was at the University of Texas, and was partially supported by the IBM Corporation through grant 61653 and by the State of Texas through TATP Project 003658-237.

†Address correspondence to Ravi Jain, Applied Research, Bellcore, 445 South St, Morristown, NJ 07960. e-mail: rjain@thumper.bellcore.com

We summarize some experimental simulation results which indicate that, in exchange for a small execution time overhead, these approximate scheduling algorithms can provide substantial improvements in I/O completion times.

Index terms: Data transfer scheduling, graph edge coloring, I/O bottleneck, multiprocessor operating systems, parallel I/O, scheduling heuristics, simultaneous resource scheduling.

References

[1] G. A. Gibson. *Redundant disk arrays: Reliable, parallel secondary storage*. PhD thesis, Univ. of Calif., Berkeley, Comp. Sci. Div, 1990. Also available as Tech. Rep. UCB/CSD 91/613.

[2] Ravi Jain. Scheduling data transfers in parallel computers and communications systems. Technical Report TR-93-03, Univ. Texas at Austin, Dept. of Comp. Sci., Feb. 1993.

[3] Ravi Jain, Kiran Somalwar, John Werth, and J. C. Browne. Scheduling parallel I/O operations in multiple-bus systems. *J. Par. and Distrib. Comp.*, Dec. 1992. Special Issue on Scheduling and Load Balancing.

[4] Ravi Jain, Kiran Somalwar, John Werth, and J. C. Browne. Requirements and heuristics for scheduling parallel I/O operations. May 1993. Submitted for publication.

[5] Ravi Jain, Kiran Somalwar, John Werth, and J. C. Browne. Scheduling parallel I/O operations. In *Comp. Arch. News*, Dec. 1993.

[6] Ravi Jain, John Werth, J. C. Browne, and G. Sasaki. A graph-theoretic model for the scheduling problem and its application to simultaneous resource scheduling. In *ORSA Conf. on Computer Science and Operations Research: New Developments in their Interfaces*, Jan. 1992. Available from Pergamon Press.

[7] H. Jordan. Scalability of data transport. In *Proc. Scalable High Perf. Computing Conf.*, pages 1–8, 1992.

[8] D. A. Patterson, G. A. Gibson, and R. H. Katz. A case for redundant arrays of inexpensive disks (RAID). In *Proc. SIGMOD*, 1988.

[9] J. E. Smith, W. C. Hsu, and C.Hsuing. Future general purpose supercomputer architectures. In *Proc. Supercomp. '90*, pages 796–804, 1990.