NASA/CR- 1998 208059

Heterogeneous Distributed Computing for Computational Aerosciences

Vaidy Sunderam

Dept of Math & Computer Science
Emory University
Atlanta, GA 30322

NASA Grant NAG 2-828

Summary of Research and Final Report

Heterogeneous Distributed Computing for Computational Aerosciences

Summary of Research and Final Report

V. S. Sunderam
Emory University

1. Project Description

The research supported under this award focuses on heterogeneous distributed computing for high-performance applications, with particular emphasis on computational aerosciences. The overall goal of this project was to and investigate issues in, and develop solutions to, efficient execution of computational aeroscience codes in heterogeneous concurrent computing environments. In particular, we worked in the context of the PVM[1] system and, subsequent to detailed conversion efforts and performance benchmarking, devising novel techniques to increase the efficacy of heterogeneous networked environments for computational aerosciences. Our work has been based upon the NAS Parallel Benchmark suite, but has also recently expanded in scope to include the NAS I/O benchmarks as specified in the NHT-1 document. In this report we summarize our research accomplishments under the auspices of the grant.

2. Research Accomplishments

The first phase of the project involved converting and coding the five NAS NPB kernels to the PVM system and experimentally investigating their performance in three different cluster environments. These results were very well received and have been detailed in publications [3], [4], and [5]. In addition, the benchmarks were made publicly available to the research community and over 175 requests for the software were received and honored. In this period, the NAS NPB benchmark specifications evolved with the addition of a new class of problem size and a few minor modifications. In the next phase, we focused on three major activities, namely: (a) Revision of the PVM versions of the NAS codes, and inclusion of the three simulated applications to construct a full complement of PVM NAS benchmarks; (b) Development of a multithreaded framework for PVM -- a new research effort oriented towards enhancing the efficiency of the NAS benchmarks and other scientific applications by enabling overlapped communications and computation, and by enabling smaller granularity without loss of performance; and (c) Development of a parallel I/O interface for PVM to support large parallel applications that have high I/O bandwidth. Towards the end of the project, we further revised and refined the PVM substrate, and also conducted a new suite of performance benchmarks on high-end workstations interconnected by high speed ATM and 100-Mb Ethernet networks, and have obtained contemporary results for NPB codes on cluster computing systems.

Brief description of work

Our tasks in this project included the conversion, enhancement, and optimization of the NPB kernels to execute under PVM, to experiment with the PVM versions of the NPB codes, to develop a full benchmark suite by adding the BT, SP, and LU benchmarks, and to analyze performance and efficiency aspects. This work was performed by the PI (Vaidy Sunderam) and by Steve Moyer and Soeren Olesen (postdoctoral fellows) and by Adam Ferrari, Nancy Sedora, and Xiaowu Lu, Emory graduate students partially supported under this award. These results are described in papers listed below as well as graduate theses [6], [7]. The software for the NPB benchmarks and PIOUS for the NHT benchmarks have been placed in the public domain, and in the time since release, over 175 requests for the NPB software and over 120 requests for the PIOUS software have been received. Work on these conversions and experiments have continued over the past several months, and recently, benchmark results have been consolidated and published in a technical report which will be sent to a journal or conference for public dissemination.

To experimentally investigate novel methods of improving the efficacy of network based concurrent computing, a threads-based system for PVM was designed and developed. The TPVM (Threads-oriented PVM) system is an experimental auxiliary subsystem for the PVM distributed system, which supports the use of lightweight processes or ''threads'' as the basic unit of parallelism and scheduling. TPVM provides a library interface which presents both a traditional, task based, explicit message passing model, as well as a data-driven scheduling model that enables straightforward specification of computation based on data dependencies. The TPVM system comprises three basic modules: a library interface that provides access to thread-based distributed concurrent computing facilities, a portable thread interface module which abstracts the required threads-related services, and a thread server module which performs scheduling and system data management. Our design is still under development, but a prototype implementation has allowed us to perform a number of preliminary experiments. These have provided strong evidence that TPVM can offer improved performance, processor utilization, and load balance to several application categories. The research results of the TPVM system are included in [7], recently submitted for publication.

In a related effort aimed at addressing the data handling and I/O requirements of large parallel applications, we have also been working on the PIOUS parallel I/O system under the auspices of this award. This work is motivated by the fact that most

[5] V. S. Sunderam, ''Recent Initiatives in Heterogeneous Parallel Computing'', *Parallel Computing: Technology and Practice*, Eds. J. P. Gray and F. Naghdy, pp. 1-16, IOS Press, Amsterdam, November 1994.

[6] Xiaowu Lu, ''Implementation and Performance Issues of NAS Parallel Benchmarks on PVM 3.3'', M. S. Thesis, Emory University, May 1995.

[7] A. Ferrari and V. S. Sunderam, ''TPVM: Distributed Concurrent Computing with Lightweight Processes'', HPDC-4, February 1995.

[8] S. Moyer and V. S. Sunderam, ''Parallel I/O as a Parallel Application'', *International Journal of Supercomputer Applications*, Vol. 9, No. 2, summer 1995.

[9] S. Moyer and V. S. Sunderam, ''Scalable Concurrency Control for Parallel File Systems'', Workshop on I/O for Parallel and Distributed Systems, *Intl. Parallel Processing Symposium*, April 1995.

[10] Nancy Sedora, ''Implementation and Analysis of the NHT-1 I/O Benchmarks on PIOUS for PVM'', M. S. Thesis, September 1995.

[11] I. Zoraja, H. Hellwagner, and V. S. Sunderam, ''SCIPVM: Parallel Distributed Computing on SCI Workstation Clusters'', to appear, *Journal of Concurrency: Practice and Experience*, 1998.

[12] A. Ferrari and V. S. Sunderam, ''Multi-Paradigm Distributed Computing with TPVM'', *Journal of Concurrency: Practice and Experience*, Vol. 10, No. 3, pp. 199-228, March 1998.

[13] C. Carothers, B. Topol, R. Fujimoto, J. Stasko, V. Sunderam, ''Visualizing Parallel Simulations in Network Computing Environments'', *Proc. 1997 Winter Simulation Conference*, pp. 110-117, Orlando, December 1997.

[14] S. Olesen and V. S. Sunderam, ''Performance of Parallel and I/O Benchmarks on PVM/PIOUS-Based Network of Workstations'', preprint, Emory University Math & CS, April, 1998.

metacomputing environments either provide no I/O facilities, or serialize all I/O requests and thus limit application performance and scalability. PIOUS is a parallel file system designed to incorporate true parallel I/O into existing metacomputing environments. Because PIOUS is itself a parallel application, the functionality of a given metacomputing environment is extended without modification. Two papers describing the PIOUS software architecture and programming model as well as preliminary results from a prototype PIOUS implementation have been written[8],[9]. Performance results are encouraging, demonstrating the potential of the PIOUS architecture for achieving scalable file system bandwidth in a network computing environment.

The two subsystems outlined above were worked on to improve the performance and functionality of the PVM network computing system as part of our ongoing research in heterogeneous computing. Several of the ideas and implementation techniques were influenced by the requirements of computational aeroscience applications, as exemplified by the NPB kernels and application benchmarks. Simultaneously with these efforts, we also continued to fine tune and optimize the NPB implementations for PVM as well as the NHT-1 I/O benchmarks. We have conducted systematic measurements of these at several stages during the course of the project, on machines and networks that varied in capabilities and speed. Two such sets of results have been reported in technical reports, the second one just recently. These efforts indicate the viability of network and cluster computing for high performance computational science applications and for other scientific computing codes, and highlight the constraints and conditions under which sub-optimal as well as near-optimal performance is likely to be achieved.

Deliverables produced

At the end of the project, our research supported by this award has produced as deliverables:

- A complete suite of the 8 NAS parallel benchmarks for the PVM system, including the 5 kernels and 3 simulated applications, involving over 23,000 lines of code. Both versions together have been distributed to over 175 requestors since the beginning of this project.

- A preliminary version of the TPVM system for threads-based concurrent computing in PVM has been designed and developed. Initial results are very encouraging and several external research groups have been given the software to facilitate their own research.

- A beta version of the PIOUS parallel I/O system has been completed and has been released. Over 120 copies of the distribution have been given out at the time of writing of this report.

Research Output

In terms of research and education, the following have been accomplished under the auspices of this award:

- Experimental research data concerning the use of cluster computing environments for computational aeroscience applications have been measured and published.
- Novel systems level enhancements to network computing systems have been developed and incorporated into the PVM heterogeneous computing system.
- Two postdoctoral fellows, Steven Moyer and Soeren Olesen, have had their postdoctoral training in aspects of this project.
- Three graduate students, Adam Ferrari, Nancy Sedora, and Xiaowu Lu have been partially supported by this grant and have completed Masters' theses.
- Eleven papers have been written as a result of work performed under this grant; ten have been accepted, and ones just been completed and will be submitted for publication.
- This work has resulted in four invited lectures by the PI, and three contributed talks at conferences.

Publications

[1] V. S. Sunderam, G. A. Geist, J. J. Dongarra, and R. Manchek, ''The PVM Concurrent Computing System: Evolution, Experiences, and Trends'', *Journal of Parallel Computing*, 20(4), pp. 531-546, March 1994.

[2] S. White, A. Alund, and V. S. Sunderam, ''Performance of the NAS Parallel Benchmarks on PVM Based Networks'', *Journal of Parallel and Distributed Computing*, Vol. 26, No. 1, pp. 61-71, April 1995.

[3] Scott White, ''The NAS Parallel Benchmarks on Virtual Parallel Machines'', M. S. Thesis, Emory University, May 1993.

[4] A. Beguelin, J. Dongarra, A. Geist, R. Manchek, and V. S. Sunderam, ''Recent Enhancements to PVM'', *International Journal of Supercomputer Applications*, Vol. 9, No. 2, pp. 108-127, summer 1995.