Lecture Notes in Computer Science 1199

Dhabaleswar K. Panda   Craig B. Stunkel (Eds.)

Communication and Architectural Support for Network-Based Parallel Computing

First International Workshop, CANPC'97
San Antonio, Texas, USA, February 1-2, 1997
Proceedings

Springer

Volume Editors

Dhabaleswar K. Panda
The Ohio State University
Department of Computer and Information Science
Columbus, OH 43210-1277, USA
E-mail: panda@cis.ohio-state.edu

Craig B. Stunkel
IBM T.J. Watson Research Center
P.O. Box 218, Yorktown Heights, NY 10598, USA
E-mail: stunkel@watson.ibm.com

Cataloging-in-Publication data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Communication and architectural support for network based parallel computing : **first international workshop ; proceedings / CANPC '97, San Antonio, Texas, USA, February 1997. Dhabaleswar K. Panda ; Craig B. Stunkel (ed.). - Berlin ; Heidelberg ; New York ; Barcelona ; Budapest ; Hong Kong ; London ; Milan ; Paris ; Santa Clara ; Singapore ; Tokyo : Springer, 1997**
(Lecture notes in computer science ; Vol. 1199)
ISBN 3-540-62573-9

NE: Panda, Dhabaleswar K. [Hrsg.]; CANPD <1, 1997, San Antonio, Tex.>; GT

CR Subject Classification (1991): C.2, D.1.3,F.1.2, D.4.4

ISSN 0302-9743
ISBN 3-540-62573-9 Springer-Verlag Berlin Heidelberg New York

Typesetting: Camera-ready by author
SPIN 10548953 06/3142 - 5 4 3 2 1 0 Printed on acid-free paper

Preface

As the performance gap between commodity microprocessors and exotic high-end processors continues to close, microprocessor-based massively parallel processors (MPPs) are becoming commonplace for achieving supercomputer performance levels. Similarly, clusters of workstations connected by local area networks (LANs) are increasingly being employed as cost-effective parallel processing systems. Such configurations are often termed Networks of Workstations (NOWs) [1], Clusters of Workstations (COWs), or simply clusters [2].

Efficient and scalable parallel processing implies efficient communication and synchronization. Although the use of workstation technology can be relatively inexpensive, commodity workstation hardware and software components have not typically provided low latency, high bandwidth inter-node communication. Strategies for improving communication and synchronization fall into several categories, some of which are:

- Better interfaces between the processor and the network
- More efficient implementations of existing end-to-end protocols (e.g., TCP/IP)
- Light-weight end-to-end communication protocols
- High-performance interconnect technology and protocols
- Operating system and architectural support for communication and synchronization
- Architectural support for distributed shared memory
- Load balancing techniques
- Collective communication support

Unlike most MPP systems, NOW systems may operate in a "shared" environment and might consist of heterogeneous workstations and networks. Such systems may be also integrated with an existing computing environment like a department or a lab, all of which makes it more difficult to achieve optimal performance.

CANPC '97—the Workshop on Communication and Architectural Support for Network-based Parallel Computing—addresses these and other issues which have an impact on the effectiveness of clusters used as parallel systems. Potential authors submitted 10-page extended abstracts which were typically reviewed by 4 referees, including at least two program committee members. We were able to accept 19 papers out of a total of 36 submissions. We believe that the resulting selections comprise an important compilation of state-of-the-art solutions for network-based parallel computing systems. This CANPC workshop was sponsored by the IEEE Computer Society, and was held in conjunction with HPCA-3, the 3rd International Symposium on High-Performance Computer Architecture, held in San Antonio on Feb. 1-5, 1997. The workshop itself took place on Feb. 1-2.

We would like to thank all of the authors who submitted papers to this workshop. Special thanks go to the program committee and the other referees

for providing us with high-quality reviews under tight deadlines. We thank Lionel Ni for his support of this workshop, including the use of his web-based review software which made our jobs considerably easier. Thanks to Rajeev Sivaram for porting and installing this software to our web server at Ohio State and maintaining it. Lastly, we thank Springer-Verlag for agreeing to an extremely tight publication schedule in order to provide the workshop attendees with these proceedings as they registered.

February 1997 Dhabaleswar K. Panda and Craig B. Stunkel

References

1. T. Anderson, D. Culler, and Dave Patterson. A Case for Networks of Workstations (NOW). *IEEE Micro*, pages 54–64, Feb 1995.
2. G. F. Pfister. *In Search of Clusters.* Prentice Hall, 1995.

CANPC'97 Program Committee

Dhabaleswar K. Panda, *Ohio State University* (co-chair)
Craig B. Stunkel, *IBM T.J. Watson Research Center* (co-chair)

Tilak Agerwala, *IBM, USA*
Henri Bal, *Vrije University, The Netherlands*
Adam Beguelin, *Carnegie Mellon University, USA*
Jehoshua Bruck, *Caltech, USA*
Al Davis, *University of Utah, USA*
David Du, *University of Minnesota, USA*
Jose Duato, *University of Politécnica de Valencia, Spain*
Sandhya Dwarkadas, *University of Rochester, USA*
Ian Foster, *Argonne National Lab, USA*
Michael Foster, *National Science Foundation, USA*
Ching-Tien Ho, *IBM Almaden Research Center, USA*
Lionel Ni, *Michigan State University, USA*
Steve Scott, *Cray Research, USA*
Marc Snir, *IBM T.J. Watson Research Center, USA*
Per Stenstrom, *Chalmers University, Sweden*
Vaidy Sunderam, *Emory University, USA*
Anand Tripathi, *NSF/Univ. of Minnesota, USA*
Thorsten von Eicken, *Cornell University, USA*
David Wood, *University of Wisconsin, USA*
Sudhakar Yalamanchili, *Georgia Tech, USA*

Referees

B. Abali
T. Agerwala
H. Bal
A. Beguelin
J. Bonney
J. Bruck
X. Chen
A. Davis
B. Dimitrov
D. C. DiNucci
J. M. Draper
D. Du
J. Duato
S. Dwarkadas
I. Foster
M. Foster
J. C. Gomez
W. J. Hahn
P. J. Hatcher
C.-T. Ho
M. A. Iverson
J. Kim
C.-T. King
M. Kaddoura
I. Kodukula
P. Leung
P. Marenzoni
E. Markatos
W. Meira Jr.
N. Mekhiel
R. G. Minnich
L. Ni
N. Nupairoj
K. Omang
K. Pingali
K. A. Robbins
S. Scott
R. Sivaram
H. Sivaraman
M. Snir
P. Stenstrom
X.-H. Sun
V. Sunderam
P. Sundstrom
A. Tripathi
J. S. Turner
T. von Eicken
D. Wood
S. Yalamanchili
H. Yamashita
X. Zhang

Table of Contents

Efficient Communication Mechanisms for Cluster Based Parallel Computing 1
A. Davis, M. Swanson, and M. Parker
University of Utah, USA

Stream Sockets on SHRIMP 16
S. N. Damianakis, C. Dubnicki, and E. W. Felten
Princeton University, USA

A Simple and Efficient Process and Communication Abstraction for Network Operating Systems 31
D. C. DiNucci
NASA Ames Research Center, USA

Efficient Adaptive Routing in Networks of Workstations with Irregular Topology 46
F. Silla, M. P. Malumbres, A. Robles, P. López, and J. Duato
Universidad Politéchnica de Valencia, Spain

A Deadlock Avoidance Method for Computer Networks 61
B. Abali
IBM T.J. Watson Research Center, USA

Extending ATM Networks for Efficient Reliable Multicast 73
J. S. Turner
Washington University, USA

ATLAS-I: A Single-Chip ATM Switch for NOWs 88
M. G. H. Katevenis, P. Vatsolaki, D. Serpanos, and E. Markatos
Institute of Computer Science, FORTH, Greece

Arachne: A Portable Threads Library Supporting Migrant Threads on Heterogeneous Network Farms 102
B. Dimitrov and V. Rego
Purdue University, USA

Transparent Treatment of Remote Pointers Using IPC Primitive in RPC Systems 115
S. Kim, *Semyung University, Korea*
M. Hyun, *Daewon Junior College, Korea*
S. Lee, *Kyungpook National University, Korea*

An Operating System Support to Low-Overhead Communications in NOW Clusters 130
P. Marenzoni, G. Rimassa, M. Vignali, M. Bertozzi, G. Conte, and P. Rossi, *Università di Parma, Italy*

Distributed Hardware Support for Process Synchronization in NSM Workstation Clusters 144
J. Bonney, R. Ramanujan, A. Ahamad, S. Takkella, and K. Thurber
Architecture Technology Corporation, USA

Synchronization Support in I/O Adapter Based SCI Clusters 158
K. Omang
University of Oslo, Norway

Load Balancing for Regular Data-Parallel Applications on Workstation Network 173
M. Kaddoura
Architecture Technology Corporation, USA

A Comparison of Three High Speed Networks for Parallel Cluster Computing 184
H. Bal, R. Hofman, and K. Verstoep
Vrije Universiteit, The Netherlands

Understanding the Performance of DSM Applications 198
W. Meira Jr., T. J. LeBlanc, N. Hardavellas
University of Rochester, USA
C. Amorim
UFRJ, Rio de Janeiro, Brazil

Performance Metrics and Measurement Techniques of Collective Communication Services 212
N. Nupairoj and L. M. Ni
Michigan State University, USA

CLAM: Connection-less, Lightweight, and Multiway Communication Support for Distributed Computing 227
J. C. Gomez and V. Rego, *Purdue University, USA*
V. S. Sunderam, *Emory University, USA*

Network-wide Cooperative Computing Architecture (NCCA) 241
H. Yamashita, T. Suguri, S. Kinoshita
NTT Information and Communication Systems Laboratories, Japan
Y. Okada, *NTT Network Strategy Planning Department, Japan*

Data Movement and Control Substrate for Parallel Scientific Computing 256
N. Chrisochoides, *University of Notre Dame, USA*
I. Kodukula and K. Pingali, *Cornell University, USA*

Author Index 269