Agent-based High-Performance Computing

Omer Rana, Kate Stout and David Kotz

o.f.rana@cs.cf.ac.uk, kate.stout@sun.com, dfk@cs.dartmouth.edu

1 Introduction

Rapid advances in technologies such as networking, parallel computing and information management, have led to the development of intelligent software components that act autonomously on the behalf of users, can analyse and access a diverse range of information, can react to changes in their environment, and can cooperate and coordinate their activities to complete a task or goal. Such components may be distributed across a network, and may work seamlessly to perform this goal, and do so without the direct intervention of a user. Such technology involves the integration of ideas from many different disciplines such as artificial intelligence, parallel processing, knowledge sharing, object-oriented design, information retrieval, distributed workflows, and databases. The objectives of developing such systems are to provide an emergent functionality that allows a robust, flexible, and scalable approach to solving problems and providing services in various application domains.

In addition to advances leading to the development of such components— *agents*— a similar change of emphasis is occurring in parallel and high-performance computing. The move from "Big Iron" to "commodity computing," whereby clusters of commodity workstations are used to create a parallel virtual machine, is converting the field of parallel computing from a niche discipline to a wider research and development community. To enable this change, workstation clusters need to support a range of different programming environments, architectures and applications— hence the development of specialised libraries such as MPI and PVM, amongst others. Recent interest in large-scale distributed-computing environments, or *computational grids*, that provide dependable, consistent and pervasive access to high-end computational resources is a focus for the high-performance computing community. These environments have the potential to change fundamentally the way we think about computing, as our ability to compute will no longer be limited to the resources we currently have on hand. The ability to integrate supercomputing resources, on demand, will enable integration of sophisticated data analysis, image processing and real-time control to be utilised within scientific instruments and simulations. Or, resources of a nation-wide, or a continent-wide strategic-computing reserve may be used to perform time-critical computational tasks in times of crisis, for example to perform diverse simulations as we plan responses to an oil spill. Agent-based computing is therefore the next obvious step. The use of techniques such as code mobility and speech-acts (through KQML and FIPA) open up new research challenges when used with approaches and themes familiar to high-performance computing. The combined use of commodity computing ideas and agent technologies could lead to new applications, in areas like resource management, data mining, data warehousing, and electronic commerce.

Many issues remain in bringing about this change, however, such as understanding how agents can be deployed on a large scale, and how to build effective "agent communities". In agent systems with a large number of agents, or where the agent environment is highly dynamic or heterogeneous, special challenges arise in managing and controlling agents. Furthermore, agent deployment must be sufficiently robust and reliable, so that scientists and commercial organisations will entrust agents with mission-critical applications.

2 The workshop

The workshop was organised around two main themes:

- The use of "agent technology" in problem-solving applications, such as scientific-computing research or large-scale business problems.
- The investigation of practical issues that arise when deploying agents in a large-scale environment.

The workshop was particularly aimed at fostering interaction between researchers and practitioners in the high-performance computing and agents communities.

3 Session 1: Practical deployment

To employ a large number of agents effectively within an application, a robust and scalable management and control approach is required. *Session 1* of the workshop covered the issues involved in practical deployment of agent systems, and contains three papers from active players in industry. Craig Thompson co-founded "Object Services and and Consulting, Inc.", and served as principal investigator on the DARPA project "Scaling Object Services Architectures to the Internet". He is organiser and co-chair of the OMG Internet Platform Special Interest Group, chartered to merge the OMG OMA architecture with Internet and Web standards to enable large-scale Internet-enabled object-based distributed computing, and also its Agent Working Group. His paper on "Agents for the Masses" discusses practical issues in using component technologies and services for deploying agents in various application domains. The next paper, by Dana Moore et al., describes a decentralised agent control and management strategy that would prevent network flooding, and maintain good overall system throughput, for workflow systems based on mobile agents. Various issues are discussed with reference to the "Autopilot" text-processing workflow system at AT&T. Amy Unruh from MCC presents the last paper in this session of lessons learned from the "InfoSleuth" project, which supports knowledge integration, information retrieval, and support for agent security and multimedia information.

4 Session 2: Managing distributed multi-agent systems

Session 2 covered special issues arising in managing distributed multi-agent systems. The first paper by R. Ghanea-Hercock et al. describes a two-tier management system for mobile-agent systems used for task distribution on a number of processors, at BT Labs, UK. The authors describe three types of agents for achieving this objective: “Hive” agents which interact with the user, “Queen” agents which monitor machine usage, and “Scout” agents, which gather statistics on computing resources. Muthukrishnan and Suresh then describe a dynamic approach to load balancing and task allocation based on agents, for a network of workstations. Bivens et al. concluded the session with a paper on decentralised network management, which accounts for traffic generated as a result of network management in addition to computational tasks. Their paper describes an agent-based framework for facilitating the management of large networks in pro-active network management, where historical data is used to build models of network traffic to support management decisions.

5 Session 3: Problem-solving applications

Session 3 focused on the use of agents within particular applications, with the first paper from Shirun Ho et al. at Hitachi Research Laboratory in Japan. They describe a middleware system that enables heterogeneous scientific simulations to be combined, with agents providing high-level semantic transformations and automatic data communication between various physical and mathematical models. Their approach also accounts for multiple calculation engines, execution times, physical model accuracy and simulation scales. Their system has been demonstrated on a Hitachi SR2201 parallel machine, as well as workstation clusters. Two applications are described in their paper, a mesh-mesh heterogeneous simulation for analysing quantum transport characteristics in a microelectronics device, and a particle-mesh heterogeneous simulation to calculate thermal diffusion from current flow of a high-frequency device. The second paper in this session comes from a leading group in multi-disciplinary “Problem Solving Environments” (M-PSEs) at Purdue University. The group has been actively employing agent-based technologies in high-performance computing applications, from their earlier work on recommender agents based on machine-learning approaches (neural networks and fuzzy logic), primarily for match-making purposes, to their current work on agent based M-PDEs for linking together various partial differential equation (PDE) solvers. The final paper of the session comes from W-S Chen et al. and describes a mobile-agent infrastructure, called “Pathfinder” for managing mobile-agent based applications. The infrastructure is demonstrated with a parallel ray-tracing application.

6 Panel sessions

The panel sessions in this workshop were suggested by Kate Stout, and form an integral and important part of the workshop. Best practise and suggestions for implementing agent-based systems will be discussed. The results of the session will be published on the workshop Web site and may form a working document for future deployment of agent systems in high-performance computing.

Workshop Web site:

`http://www.cs.cf.ac.uk/User/O.F.Rana/agents99/`

Workshop organisers

David Kotz
Department of Computer Science
Dartmouth College
Hanover, NH 03755-3510, USA
dfk@cs.dartmouth.edu

Omer Rana
Parallel and Scientific Computing Group
Department of Computer Science
Cardiff University
PO Box 916
Cardiff CF2 3XF, UK
omer@cs.cf.ac.uk

Workshop chairs

Kate Stout
Agent Research Team
Sun Laboratories
Sun Microsystems, Inc.
One Network Drive,
Burlington, MA 01803-0902, USA
Kate.Stout@sun.com

David Walker
Parallel and Scientific Computing Group
Department of Computer Science
Cardiff University
PO Box 916
Cardiff CF2 3XF, UK
david@cs.cf.ac.uk

Program committee

David Walker	Cardiff University, UK
David Kotz	Dartmouth College, USA
Omer Rana	Cardiff University, UK
Kate Stout	Sun Microsystems, Massachusetts, USA
Philippe De Wilde	Imperial College, London, UK
Mark Baker	Portsmouth University, UK
Siamek Hasanzadeh	Sun Microsystems, Palo Alto, USA
Geofferey Fox	Syracuse University, USA
Vladimir Getov	University of Westminster, UK
Lyndon Lee	BT Labs, UK
Anupam Joshi	University of Maryland, USA
Jean-Louis Pazat	EuroTools and IRISA, France
Elias Houstis	Purdue University, USA
Michael Fisher	Manchester Metropolitan University, UK
Giovanna Serugendo	Teleinformatics Group, University of Geneva, Switzerland
Jeremy Baxter	DERA, UK
Paolo Petta	Austrian Research Institute for Artificial Intelligence
Erann Gat	JPL, Caltech, USA
Katia Sycara	Carnegie Mellon, USA
Danny Lange	General Magic, USA
Kurt Rothermel	IPVR, Universitaet Stuttgart, Germany
Hyacinth Nwana	BT Labs, UK
Piyush Mehrotra	ICASE, NASA Langley, USA
Hillol Kargupta	Washington State University, USA
Jeffrey Bradshaw	Intelligent Agent technology, Boeing, USA
Bent Thomsen	ICL, UK
Denis Caromel	INRIA, Sophia Antipolis, France
Maria Gini	University of Minnesota, USA
Giacomo Piccinelli	Hewlett Packard Labs, Bristol, UK