The Anatomy of the Grid: Enabling Scalable Virtual Organizations

Ian Foster

Mathematics and Computer Science Division, Argonne National Laboratory
Department of Computer Science, The University of Chicago
`foster@{mcs.anl.gov,cs.uchicago.edu}`
`http://www.mcs.anl.gov/~foster`

Extended Abstract

The term "the Grid" was coined in the mid-1990s to denote a proposed distributed computing infrastructure for advanced science and engineering [4]. Considerable progress has since been made on the construction of such an infrastructure (e.g., [1,6,7]) but the term "Grid" has also been conflated, at least in popular perception, to embrace everything from advanced networking to artificial intelligence. One might wonder whether the term has any real substance and meaning. Is there really a distinct "Grid problem" and hence a need for new "Grid technologies"? If so, what is the nature of these technologies, and what is their domain of applicability? While numerous groups have interest in Grid concepts and share, to a significant extent, a common vision of Grid architecture, we do not see consensus on the answers to these questions.

My purpose in this talk is to argue that the Grid concept is indeed motivated by a real and specific problem and that there is an emerging, well-defined Grid technology base that solves this problem. In the process, I develop a detailed architecture and roadmap for current and future Grid technologies. I also argue that while Grid technologies are currently distinct from other major technology trends, such as Internet, enterprise, distributed, and peer-to-peer computing, these other trends can benefit significantly from growing into the problem space addressed by Grid technologies.

The real and specific problem that underlies the Grid concept is *coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations*. The sharing that we are concerned with is not primarily file exchange but rather direct access to computers, software, data, and other resources, as is required by a range of collaborative problem-solving and resource-brokering strategies emerging in industry, science, and engineering. This sharing is, necessarily, highly controlled, with resource providers and consumers defining clearly and carefully just what is shared, who is allowed to share, and the conditions under which sharing occurs. A set of individuals and/or institutions defined by such sharing rules form what we call a *virtual organization* (VO).

The following are examples of VOs: the application service providers, storage service providers, cycle providers, and consultants engaged by a car manufacturer to perform scenario evaluation during planning for a new factory; members

R. Sakellariou et al. (Eds.): Euro-Par 2001, LNCS 2150, pp. 1–4, 2001.

of an industrial consortium bidding on a new aircraft; a crisis management team and the databases and simulation systems that they use to plan a response to an emergency situation; and members of a large, international, multiyear high-energy physics collaboration. Each represents an approach to computing and problem solving based on collaboration in computation- and data-rich environments.

As these examples show, VOs vary tremendously in their purpose, scope, size, duration, structure, community, and sociology. Nevertheless, careful study of underlying technology requirements leads us to identify a broad set of common concerns and requirements. In particular, we see a need for highly flexible sharing relationships, ranging from client-server to peer-to-peer and brokered; for complex and high levels of control over how shared resources are used, including fine-grained access control, delegation, and application of local and global policies; for sharing of varied resources, ranging from programs, files, and data to computers, sensors, and networks; and for diverse usage modes, ranging from single user to multi-user and from performance sensitive to cost-sensitive and hence embracing issues of quality of service, scheduling, co-allocation, and accounting.

Current distributed computing technologies do not address the concerns and requirements just listed. For example, current Internet technologies address communication and information exchange among computers but not the coordinated use of resources at multiple sites for computation. Business-to-business exchanges focus on information sharing (often via centralized servers). So do virtual enterprise technologies, although here sharing may eventually extend to applications and physical devices. Enterprise distributed computing technologies such as CORBA and Enterprise Java focus on enabling resource sharing within a single organization. Storage service providers (SSPs) and application service providers allow organizations to outsource storage and computing requirements to other parties, but only in constrained ways: for example, SSP resources are typically linked with a customer via a virtual private network. Emerging "Internet computing" companies seek to harness idle computers on an international scale [3] but, to date, support only centralized access to those resources. In summary, current technology either does not accommodate the range of resource types or does not provide the flexibility and control on sharing relationships needed to establish VOs.

It is here that Grid technologies enter the picture. Over the past five years, research and development efforts within the Grid community have produced protocols, services, and tools that address precisely the challenges that arise when we seek to build scalable VOs. These technologies include security solutions that support management of credentials and policies when computations span multiple institutions; resource management protocols and services that support secure remote access to computing and data resources and the co-allocation of multiple resources; information query protocols and services that provide configuration, monitoring, status information about resources, organizations, and services [2]; and data management services that locate and transport datasets between storage systems and applications.

Because of their focus on dynamic, cross-organizational sharing, Grid technologies complement rather than compete with existing distributed computing technologies. For example, enterprise distributed computing systems can use Grid technologies to achieve resource sharing across institutional boundaries; in the ASP/SSP space, Grid technologies can be used to establish dynamic markets for computing and storage resources, hence overcoming the limitations of current static configurations.

In my talk, I will expand upon each of these points in turn. My objectives are to (1) clarify the nature of VOs and Grid computing for those unfamiliar with the area; (2) contribute to the emergence of Grid computing as a discipline by establishing a standard vocabulary and defining an overall architectural framework; and (3) define clearly how Grid technologies relate to other technologies, explaining both why various emerging technologies are not yet the Grid and how these technologies can benefit from Grid technologies.

It is my belief that VOs have the potential to change dramatically the way we use computers to solve problems, much as the web has changed how we exchange information. As the examples presented here illustrate, the need to engage in collaborative processes is fundamental to many diverse disciplines and activities: it is not limited to science, engineering and business activities. It is because of this broad applicability of VO concepts that Grid technology is important.

Acknowledgments

This text is based on the introductory section of an article [5] that addresses these issues at length. I thank my co-authors, Carl Kesselman and Steven Tuecke, for their contributions, as well as numerous colleagues with whom we have discussed these ideas.

References

1. J. Beiriger, W. Johnson, H. Bivens, S. Humphreys, R. Rhea. Constructing the ASCI Grid. *Proc. 9th IEEE Symposium on High Performance Distributed Computing*, 2000, IEEE Press. 1
2. K. Czajkowski, S. Fitzgerald, I. Foster, C. Kesselman. Grid Information Services for Distributed Resource Sharing. *Proc. 10th IEEE Intl. Symp. on High Performance Distributed Computing*, IEEE Press, 2001. www.globus.org/research/papers/MDS-HPDC.pdf 2
3. I. Foster. Internet Computing and the Emerging Grid. *Nature Web Matters*, 2000. www.nature.com/nature/webmatters/grid/grid.html. 2
4. I. Foster, C. Kesselman (eds.). *The Grid: Blueprint for a New Computing Infrastructure*, Morgan Kaufmann, 1999. 1
5. I. Foster, C. Kesselman, S. Tuecke. The Anatomy of the Grid: Enabling Scalable Virtual Organizations. *Intl. Journal Supercomputer Applications*, 2001. www.globus.org/research/papers/anatomy.pdf. 3

6. W. Johnston, D. Gannon, W. Nitzberg. Grids as Production Computing Environments: The Engineering Aspects of NASA's Information Power Grid. *Proc. 8th IEEE Symposium on High Performance Distributed Computing*, 1999, IEEE Press. 1
7. R. Stevens, P. Woodward, T. DeFanti, C. Catlett. From the I-WAY to the National Technology Grid. *Communications of the ACM*, 40(11):50-61. 1997. 1