

GeoMiddleware to Support Interoperability for Grid Computing

Marc P. Armstrong
Department of Geography and Program in
Applied Mathematical and Computational
Sciences
The University of Iowa
marc-armstrong@uiowa.edu

Shaowen Wang
Research Services of Information
Technology Services, and Department of
Geography
The University of Iowa
shaowen-wang@uiowa.edu

Introduction

Middleware is software that is designed to support the interoperability of computer applications that use (and/or produce) different types of information. Middleware for geographic applications is particularly important because required information may have scale dependent relationships, which may cascade to cause problems related to level of generalization, dimension change, and categorical precision. The challenge of supporting interoperability is especially vexing when multiple types of data (from various sources and with variable error characteristics) must be meaningfully integrated for use in distributed applications. In this position paper we set out a general geographic information processing problem as a way to motivate discussion about the use of middleware to support analyses in distributed, heterogeneous Grid computing environments.

Grid Computing and Middleware

The Grid refers to an infrastructure that enables the integrated, collaborative, and coordinated use of distributed heterogeneous computing resources, such as computers, networks, databases, and scientific instruments owned and managed by multiple organizations (Foster and Kesselman, 1999). Since the very early stages of the evolution of the Grid, middleware has been a primary focus of software development and research effort (Foster *et al.*, 2001). Middleware supports applications in distributed computing environments by providing services that enable the interconnectivity and interoperability of applications, systems and devices.

GeoMiddleware

Several disciplines have turned to Grid computing to find solutions to computationally intensive problems; this has required the development of domain-specific middleware (see GEON¹ and iVDGL²). Such middleware exploits the characteristics of problems and helps application scientists manage and use the Grid effectively. Geographic information science (GIScience) must develop middleware that captures important geographic characteristics of problems. This GIScience-specific Grid middleware may provide interoperable geographic analysis services that are able to reconcile different data and metadata regarding formats and semantics. This “GeoMiddleware” helps GIScientists locate, allocate, and use Grid resources effectively and efficiently (Wang *et al.*, 2002).

A Motivating Problem

The use of grid computing to explore computationally complex network optimization problems by multiple members of a decision making group requires a considerable amount of data integration, multi-user interaction, computational decomposition and reconstitution, and

¹ <http://www.geogrid.org/>

² <http://www.ivdgl.org/>

distributed visualization. Each step of this complex chain of events requires access to metadata that can be used by middleware to orchestrate activities in an appropriate sequence to achieve a goal of multi-participant interaction with the problem-solving process. To accomplish this goal, data must be assembled to describe:

- A street network, possibly from different sources and with different scales; this may cause problems with distances calculations (Armstrong and Dalziel, 1989);
- Distributed demand for services can be estimated from census tabulations for small areas that are assigned to locations in the network, also introducing errors.
- Supply of existing services (if any);
- Number, type, and capacity of needed services;
- Criteria used by one or more decision-makers;
- Metadata and data that describes solutions and their characteristics;
- Maps, graphs and statistical information and metadata about individual solutions and collections of them;
- The set of available computational resources and their current status;
- Partitions of problems that can be allocated to available resources according to a scheduling plan;
- Reassembled problem sub-solutions from the partitions;
- Preferences for visualization options; and
- Information about choices, or modifications to criteria and solutions generated.

Research Challenges

Though a markup language, such as GML, can be used to describe some of the semantic characteristics of needed data, the use of disparately sourced data in analyses remains a problem if it is used in a distributed, heterogeneous environment. Data also may have markedly different granularities, ranging from a parameter (maybe a byte) to gigabytes of road network information. It is difficult to rationally assemble data for analysis when it is derived from different sources and with different geographical error characteristics that arise as a consequence of different compilation scales and collection purposes. For example, if road networks are used to calculate distances between demand and supply locations, differences in the level of generalization in the data for a study area can significantly affect the results obtained. As data are generalized, categories may get collapsed as well. The interaction between generalization in the geometric domain and the level of categories maintained is not always well-specified and loss of meaning can occur.

When road network data are used to represent large numbers of supply and demand locations in optimization, problems can quickly become computationally intensive. Interoperability problems may be further exacerbated by the imposition of additional (parallel) Grid computing procedures that are put into place to improve computing performance. Oftentimes, these procedures include strategies such as domain decomposition and task scheduling. Subsequent to the computation of results, additional procedures must be specified to integrate and represent solutions. GeoMiddleware must be designed to manage the integration of computational procedures as well as data fusion involving multiple sources and scales with different characteristics of errors and levels of generalization. GeoMiddleware must be able to reassemble partitions and handle multiple types of interactions of distributed users. As a consequence of these and other problems, the development of GeoMiddleware to support interoperable Grid applications is a difficult challenge.

References

- Armstrong, M.P. and Dalziel, B. 1989. Digitizing and distance estimation errors in spatial analysis: scale and network configuration effects. In: *Proceedings of the International Geographic Information Systems (IGIS) Symposium '89*. Washington, DC: Association of American Geographers, pp. 75-84.
- Armstrong, M.P., Cowles, M.K., Wang, S., 2005. Using a computational grid for geographic information analysis. Forthcoming in *The Professional Geographer*.
- Foster, I., Kesselman, C., and Tuecke, S., 2001. The anatomy of the grid: enabling scalable virtual organizations. *International Journal Supercomputer Applications*, **15** (3), available at: <http://www.globus.org/research/papers.html>, accessed on October 1, 2004.
- Foster, I., and Kesselman C., 1999. *The Grid: Blue Print for a New Computing Infrastructure*. Morgan Kaufmann Publishers, Inc. San Francisco, CA.
- Wang, S., Armstrong, M. P., and Bennett D. A. 2002. Conceptual basics of middleware design to support Grid computing of geographic information. In: *Proceedings of 2nd International Conference on Geographic Information Science*, September, 25 - 28, Boulder, CO, USA, 197-200.
- Wang, S. and Armstrong, M.P. 2003. A quadtree approach to domain decomposition for spatial interpolation in grid computing environments. *Parallel Computing*, **29** (10): 1481-1504.