Pure appl. geophys. 163 (2006) 2467–2483
0033–4553/06/122467–17
DOI 10.1007/s00024-006-0146-7

Pure and Applied Geophysics

A Grid Framework for Visualization Services in the Earth Sciences

G. ERLEBACHER,[1] D. A. YUEN,[2] Z. LU,[1] E. F. BOLLIG,[2]
M. PIERCE,[3] and S. PALLICKARA[3]

Abstract—Scientific visualization is an ingredient essential to understanding the large amounts of data generated from large-scale numerical simulations, laboratory experiments and geological surveys. Visualization forms an integral component of any complete framework, together with services to handle mathematical and statistical analysis, storage, feature extraction, and other functions. To support rapid and seamless collaborations and communication between researchers across geographically disparate regions necessitates a distributed infrastructure that supports redundancy, fault tolerance, and most importantly, ease of use. We describe herein an architecture based on Naradabrokering, a publish/subscribe framework that supports the above requirements. We have implemented an initial version of this architecture and describe some initial experiments.

Key words: Middleware, collaboration, visualization, web-services, grids.

1. *Introduction*

In the past fifteen years, scientific visualization (SV) has evolved from a tool used by a few researchers to understand their data into an indispensible approach towards improving understanding of the extremely large data sets currently being produced. SV is used in many areas of applied research, ranging from meteorology and oceanography, hydrology, material sciences, earth physics, etc., with various degrees of sophistication. Techniques used to enhance data understanding, beyond looking at rows of numbers on printouts, range from simplistic line plots to time-dependent volumetric renderings of data sets at resolutions of 1000^3 and beyond. There is also a need for higher-dimensional representations such as complex higher-dimensional phase diagrams.

---

[1] School of Computational Science and Information Technology, Department of Mathematics/Computer Science, Florida State University, Tallahassee, FL.
E-mail: zhenyulu@cs.fsu.edu; erlebach@csit.fsu.edu

[2] Minnesota Supercomputer Institute, Department of Geophysics, University of Minnesota, Minneapolis, MN. E-mail: davey@msi.umn.edu; bollig@scs.fsu.edu

[3] Community Grids Lab, Indiana University, Bloomington, IN.
E-mail: mpierce@cs.indiana.edu; spallick@indiana.edu

Comparing the current hardware requirements to those of the early 1990s, we find much in common. In 1990, high-end supercomputers cost in the range of $20–30 million while desktop workstations cost upwards of $30,000. Today the discrepancy is even more severe. Our largest supercomputers cost $100,000,000 while desktop computers with high-end graphics processors cost in the range of $3,000. While the ratio between high-end and low-end computer costs and performance have increased over the years, the network bandwidth has only increased moderately fast, although storage requirements are increasing exponentially fast. Over the past 15 years, there has been an increasing emphasis on collaborative efforts within a given institution, across institutions, even across continents. For example, the ACES program (http://www.uq.edu.au/ACES) studies earthquake dynamics, and involves researchers in Australia, China, and the United States. There is a pressing need to not only share data, but to facilitate sharing of the visual experience, both static and animated. The fast-paced evolution of technology makes it all the more difficult to achieve real collaboration in the midst of changing standards, evolving tools, and the lack of a common software infrastructure (COHEN, 2005).

Large numbers of researchers wish to collaborate with one-another, but are restricted to data visualization on local hardware resources (workstations, large displays, etc.). There is very little support for *easily* sharing their visualizations with one another. Researchers are increasingly mobile, own a wide variety of end display devices, ranging from laptops to hand-held devices, and have access to commodity (i.e., cheap) graphics cards and graphics processors of incredible power (ERLEBACHER and YUEN, 2004). Yet there is very little support to provide a seamless integration between these various technologies. For example, if a researcher develops a tool to visualize earthquake clusters, environments should exist to simplify the integration of this tool within themselves, making it available to all. This vision is addressed by various researchers working on grid technology (FOSTER and KESSELMAN, 1999; FOSTER *et al.*, 2002; UYAR *et al.*, 2003; FOX *et al.*, 2003), a loose collection of hardware and software resources made available to anybody around the world with adequate permissions. Catastrophic events, such as the recent tsunami that produced wide-spread devastation across multiple countries, calls for collaborative systems capable of reaching wide audiences across multiple channels of communication (GADGIL *et al.*, 2005).

2. *Dataset Glut*

In all areas of the earth sciences, datasets are growing in size at what appears to be exponential rates. For example, when a 3D dataset expands by a factor of two in each coordinate direction, the size of one snapshot expands by a factor of eight, or approximately an order of magnitude.

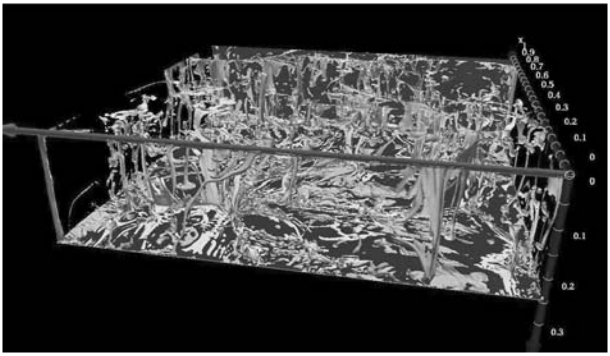

Figure 1
An example of strongly time-dependent high Rayleigh number convection in the Earth's mantle. A Rayleigh number of 10^8 is employed here for a purely basally-heated configuration with an aspect-ratio of $4 \times 4 \times 1$. A grid configuration of $400 \times 400 \times 400$ points has been employed. The temperature fields cast in volumetric rendering are displayed here for both the upwelling (pink isosurface)and downwelling (blue isosurface). Data courtesy Fabien W. Dubuffet.

A 512^3 numerical simulation that stores five variables at the finest resolution in single precision requires $2^{29} \approx 540$ Mbytes of disk space. Storing 200 time steps for future analysis, for example to track the evolution of vortices and other features, requires 100 Gbytes of disk space. The same data set at a resolution of 1024^3 would require 800 Gbytes of storage. As computer power continues to follow the predictions of Moore's law, storage requirements will become increasingly severe.

Satellite systems such as Erbes [NASA] currently collect a Terabyte of data daily. The resolving power of field equipment is also increasing. For example, high resolution spectral channel equipment can scan terrain over 100 simultaneous channels. Resolution is expected to increase further in the immediate future.

3. New Science

Mantle convection, currently recognized as the driving mechanism of plate tectonics, governs the thermal and chemical evolution of the earth. Mantle convection involves both thermal and compositional transport and is a phenomenon with a high range of temporal and spatial scales due to the strong nonlinearities in the governing equations, ranging from rheology to sharp compositional gradients. Figure 1 shows temperature isosurfaces in the earth's mantle using numerical simulation data at $Ra = 10^8$. The complexity of the isosurfaces and the richness of the plume structures suggest the development of plume extraction techniques and interactive navigation (i.e., exploration) capabilities that could be shared between multiple users.

Based on the results of 2-D simulations, convective transitions are expected to occur at Rayleigh numbers within the range $Ra = 10^{10}$ and $Ra = 10^{12}$ (VINCENT and YUEN, 2000). Such high Rayleigh numbers would require a grid of at least 1250^3 points spaced uniformly. It would be very impractical to process and visualize the huge time-dependent data sets from these simulations at the desktop. Rather, a large number of tools will be available to the community, some open source, some commercial, to help datamine the data, extract relevant spatial-temporal subsets (themselves very large) and display them in a manner that clarifies the physics. To this end, users will leverage existing parallel graphics APIs (not useful at the desktop), and other existing (or yet to exist) analysis tools. We have used visualization techniques to examine the interrelationship between the temperature and velocity fields at $Ra = 10^{10}$ at a grid resolution of 512^3. We have found that at this Ra, the convective cells do not extend fully between upper and lower boundaries, but rather, break up into one set close to the cold slab, and one near the heated slab, thus cutting off the heat transport between the two boundaries (Fig. 2). This finding required the computation of a large number of streamlines (Fig. 3). Although we downloaded one full data set to the desktop to accomplish this task, it was an inefficient procedure due to lack of memory and insufficient computing resources. It would be more efficient to extract the streamlines on a supercomputer using parallel algorithms and display them either locally (if the number of streamlines is moderate) or remotely. Given the difficulties encountered with a single data set, it is clear that the analysis of a collection of hundreds of such files at the desktop is all but impossible, requiring multiple Terabytes of disk space, and upwards of a Terabyte of local fast memory.

Recently, the role of water in geodynamics is receiving increased attention, driven by laboratory experiments, seismic analysis and numerical modeling.

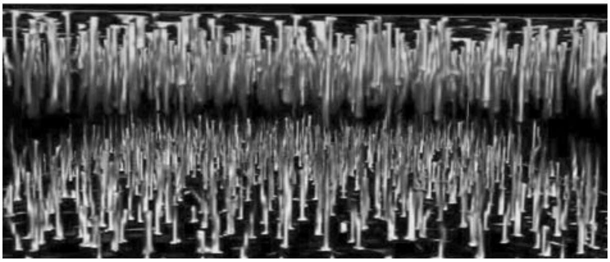

Figure 2
An example of extremely high Rayleigh number convection at $Ra = 10^{10}$. A grid configuration of $600 \times 600 \times 600$ points has been employed. We notice that the system is partially layered from the increasingly large Rayleigh number, which is in agreement with earlier results in 2-D convection by (VINCENT and YUEN, 2000). The upwellings and downwellings are denoted by the yellow and blue colors, respectively. The morphologies are bounded by a certain magnitude of the dimensionless temperature.

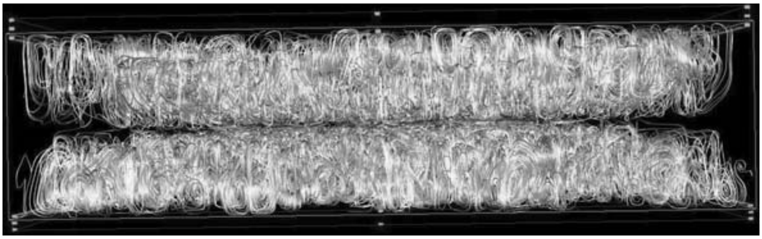

Figure 3
Streamlines corresponding to the data in Figure 2 at $Ra = 10^{10}$. They clearly indicate a layered configuration.

Introducing water into the dynamical models changes the nature of the equations to a complex system characterized by multiple components with many different intrinsic scales. We have modeled the dynamics of hydrous plume development above subducting slabs. Our 2-D model incorporates the effects of thermal-chemical buoyancy as well as the combined effects of both hydration and melting. We have used up to *one billion* tracers to delineate the evolution of the lithological, rheological and deformation structure of the subduction zone with a resolution between 10 and 50 meters (GERYA and YUEN, 2003b). Tracers are integrated simultaneously with a Lagrangian energy equation and the solution is interpolated back to an Eulerian system for the solution of the thermal-chemical momentum equation. The rheologies of the different components are explicitly included. Using visualization, we have found that, contrary to prevailing wisdom, hot rising mantle flows prevail in the mantle wedge above the subducting slab, while the partially molten hydrated upwellings (wet cold plumes) in the mantle wedge are characterized by a colder thermal anomaly of 300 to 400 degrees (GERYA and YUEN, 2003b).

Visualizing one billion tracers is indeed a daunting task and a grand challenge problem. Currently there exists no single display device available for unveiling all of the minute details simultaneously. In order to address this serious obstacle in visualization, we promote the use of remote visualization. Our web-based remote solution is an image service with zooming capability that requires minimal bandwidth, allows the user to explore the data across the temporal dimension, across several thermo-physical properties, and at different spatial scales.

A 3-D configuration requires at least 100 billion tracers to achieve a resolution of several hundred meters. (Tackling this problem will also help address other problems in volcanology.) It should be noted that while it is possible to display the 2-D data in its entirety or in planar subsets without ambiguity (Fig. 4), the same cannot be said for 3-D. New techniques will be developed to create a multiscale representation of the 3-D data to automatically detect and extract relevant flow features, and to display them in a meaningful form. Unfortunately, it is highly unlikely that the networks will

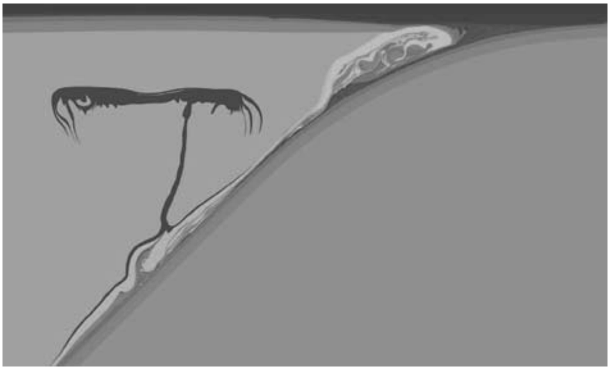

Figure 4
Compositional fields developed from subduction dynamics modeled with the tracer method (GERYA and YUEN, 2003a). The different components of the oceanic crust and lithosphere are portrayed by the various colors. Twelve components have been employed and one billion tracers have been used (RUDOLPH *et al.*, 2004). Further information can be found in (GERYA and YUEN, 2003a) and a web-interface can be found on the **WebIS** site at http://webis.msi.umn.edu.

ever become fast enough to transmit the raw data, or meaningful subsets thereof, to local sites. Rather, it will be more expedient to maintain the data at one or more storage archives and use local computational resources to reduce and display the data, returning only the 2-D bitmaps to the desktop.

4. Middleware

For the foreseeable future, there is an urgent need for *easy to use* tools to help earth scientists cull information from the onslaught of data accumulating at daily rates in the range of hundreds of Gigabytes to Terabytes. For example, the Earthscope project (http://www.earthscope.org), in particular, recognizes the value of data to facilitate better understanding of our planet, judging from the range of new experimental and computational techniques the project supports to expand data collection in the geosciences. Unfortunately, with the exception of a few sites such as the Southern California Earthquake Center, JPL and the Scripps Institution of Oceanography, geoscientists generally lack the necessary technical knowledge to tackle the challenges of remote visualization and analysis over the Internet, instead capitalizing on the burgeoning developments in grid technology. Moreover, easy to use software and transparent middleware able to handle large data sets in a *collaborative mode are still not available*. Although Fox and his group (PIERCE *et al.*, 2003; LEE *et al.*, 2003) have been actively involved in the architectural design of (SERVO 2002) through the creation of sophisticated portlet technology and XML schemas describing a variety of geoscience processes, their systems remain point-to-point, noncollaborative, and visualization remains at best a painful process of generating single images of two-dimensional data. While many projects promise visualization tools, they remain low on their priority lists behind code access, code and metadata management, and security. Recent work has begun to investigate the use of Narada messaging for certain problems within SERVO (GADGIL *et al.*, 2005) with the objective, among several, of removing the point-to-point restriction.

Substantial analysis of the huge data sets is impeded due to insufficient network bandwidth. The promise of higher speed networks (Internet III) only postpones these problems temporarily since the rate of increase in network bandwidth appears to be substantially slower than the rate of increase of the average data set. Currently, transmitting a high-resolution microscopy image (at a resolution of $20,000 \times 20,000$, which is as dense as a high-resolution Schmidt plate image used in astronomy, over the current Internet would take at least one to two hours. It would be more efficient to break up the data into multiple files and at several resolutions, and access the subsets as needed (BEYNON *et al.*, 2000). Very often the scientists must travel to another site to optimize their productivity.

Visualization is but one of several components necessary to promote effective collaboration between distributed teams of researchers. As these teams keep growing in size and diversity, the grand challenge problems that can be addressed become increasingly complex and difficult to solve. Grids were developed to address this concern, although they are still difficult to use and often do not present the user with an easy to use, extensible interface. There are several efforts underway to develop web portals (PIERCE *et al.*, 2002) to present users with a "simple" mechanism to access backend resources (supercomputers, storage areas, visualization and analysis servers).

When a user accesses a grid resource, its identity must often be known in advance, following the client/server model. There is an urgent need for a more general mechanism to link client devices (hand held devices, laptop and desktop computers, and large screen displays) with servers that support a variety of services (or functionalities) necessary to the researcher. Figure 5 illustrates the basic difference between point-to-point communication and a publish/subscribe paradigm. In the latter, a client *publishes* a message, including a topic tag, to the middleware fabric, in our case Naradabrokering (http://www.naradabrokering.org). A destination point (a server, service, or other generic utility) only processes those messages tagged with a topic it has subscribed to.

We propose a middleware architecture that enables users to enter requests into the system without consideration for how the request will be executed or by whom. One or more schedulers are responsible for analyzing the completed requests and submitting them to appropriate resources for execution. The results of the completed request are eventually (faster is better) returned to the user. A successful middleware framework should support fault-tolerance, service redundancy, journaling systems,

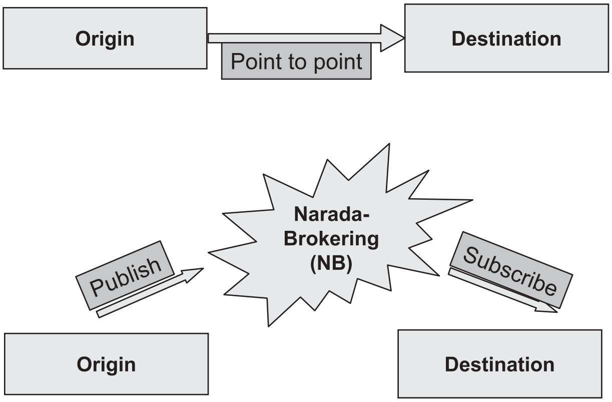

Figure 5
Top: The client (origin) must explicity define the IP address or hostname of the destination server. Bottom: The origin and destination machines are decoupled through the publish/subscribe middleware (see text).

interactivity, and most importantly, collaboration. In the area of visualization, tools are required for users to share displays. Although bitmap sharing can already be accomplished through the Virtual Network Computing API (http://www.realvnc.com), there are no generally accepted tools for two-way interaction via direct manipulation of these bitmaps. Alternative approaches are desirable, such as permitting multiple users to share an exploration session through a spatio-temporal data set. The underpinning of our architecture is the NaradaBrokering system described next.

4.1. NaradaBrokering

NaradaBrokering (NB) (http://www.naradabrokering.org) is a distributed messaging infrastructure that can be used to intelligently route data between the originators (publishers) and registered consumers (Fox and Pallickara, 2003; Fox and Pallickara, 2002a,b, 2001). Its strength lies in the complete decoupling between message source and destination. Rather than specify the precise origin and destination of a message with an IP address or any other tag that uniquely identifies the actual hardware, messages are tagged by topic. Thus, NB is a ‘‘black box’’: publishers send (i.e., publish) input messages, with an associated topic, while subscribers receive (i.e., consume) messages that correspond to any topic they subscribed to. NB manages messages, tracks topics, and routes message content to the appropriate subscribers. As a result, a given message can be recevied by several subscribers, opening the door to effective collaborative systems. NaradaBrokering has already been used to build an audio/video conferencing system (Bulut *et al.*, 2003), and is being investigated for certain aspects of the iSERVO program (GADGIL *et al.*, 2005).

The smallest unit of the messaging infrastructure that would provide a backbone for routing these messages must be able to intelligently process and route messages while working with multiple underlying network communication protocols. We refer to this unit as a *broker* to distinguish it clearly from the application *servers* that would be among the sources/links to messages processed within the system. Entities within the system use the broker network to effectively communicate and exchange data with each other. These interacting entities could be any combination of users, resources, services and proxies. They are also referred to as clients.

NaradaBrokering has a cluster-based architecture that allows the system to scale to support large client concentrations by adding new broker nodes, thus sustaining a high volume of messages between interacting entities. NaradaBrokering manages all communication complexities and optimizations pertaining to the deployment of specific network transports or to the communication across domains separated by firewall boundaries. The main challenge is to develop a topic naming scheme and standardized approaches to connect/disconnect clients, schedulers and other types of entities.

4.2. Our Solution: publish/subscribe

Our architecture will be fully web-service compliant when completed. Any task, such as file storage/retrieval, file filtering, visualization, video production, etc. can be recast as a web service with its own resources, such as a single cluster to visualize large data sets.

Our ultimate goal is to provide an infrastructure aimed at providing visualization-related services to the geophysics community with the following attributes:

1. Complete decoupling between the users who *request* services and the resources that *execute* these services.
2. Availability of multiple machines to perform visualization and computational services.
3. Support for data files stored on one of several computational platforms.
4. No single point of failure.
5. Automatic resource discovery transparent to the users.
6. User interfaces built into the leading web browsers.
7. Support of several versions of client software to accommodate users without the latest java plug-in technology.
8. Automatic scalability, a result of the built-in scalability of the NaradaBrokering middleware.
9. Language independence for clients and services.

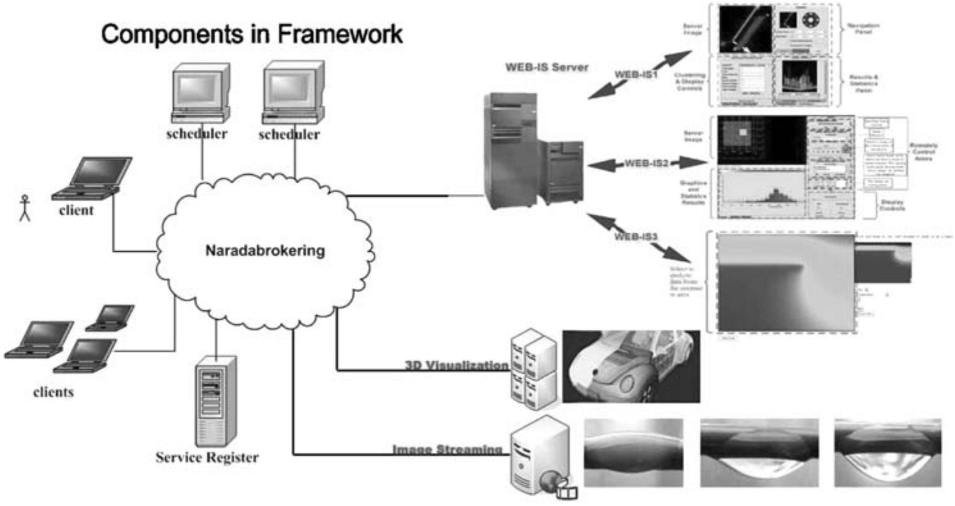

Figure 6
Example web-services connected to our proposed middleware system.

We have developed the framework for such an architecture, described in the next section.

4.3. System Components

Our system is constructed from five principal components (Fig. 6): the broker network, the clients, the services, the project managers, and the service registry/topic managers. In what follows, *entity* refers to any client/service/manager connected to NB.

1. *The Broker Network* is set up by a set of cooperating broker nodes (they define the middleware fabric) responsible for effectively organizing entity interests and routing the right message to the right set of entities.
2. *Client* entities make task requests and often display the task results through graphical user interfaces (GUIs). The client publishes a high level task description to NB along with any additional metadata necessary to complete the task (i.e., data set descriptions, commands). One consequence of the client proxy is that a SOAP message can be received by multiple subscribers, which will form the basis of future collaborative services.
3. *Resources/Services* within the system execute the tasks (or subtasks) requested by the clients. Resources subscribe to themselves, simplifying resource discovery. Services are typically accessed through a set of procedures described within a WSDL file, stored in a public database along with descriptive metadata. This will enable existing geoinformatics projects, such as GEON (http://www.geongrid.org), CHRONOS (http://www.chronos.org), CIG (http://www.geodynamics.org),

iSERVO (http://www.iservo.edu.au/), and ACCESS (http://www.access.edu.au), to develop their own interfaces to our services.

4. *Project Managers/Task Schedulers* offload the clients and are responsible for successful task completion. A task is a unit of work specified by a client. Sometimes, a specified task can be decomposed into several subtasks. For the most part, the client is solely interested in the correct, complete, and expedient execution of the task. For example, a volume rendering task could be executed on multiple servers and the various images composited before returning the final image to the client. The task scheduler is responsible for taking a task and splitting it up into a set of smaller subtasks. The task scheduler then constructs a dependency chain for these subtasks and decides the sequence in which these tasks need to be performed, taking concurrency into account where possible.
5. *Topic Managers and Service Registries* are fundamental to the user transparency of our architecture. Clients must discover a scheduler to manage requests on their behalf; project managers must discover resource entities to accomplish particular tasks. In a future version, the resource entities may have their own discovery requirements. We are investigating two complementary approaches, the first for predefined topics, the second for dynamically generated topics.

Entities willing to share resources subscribe to predefined topics, one for each resource type. Entities interested in locating these resources simply publish to these topics, together with credentials and other relevant information. The broker network routes the discovery request to all entities that expressed their interest through subscription. Depending on the policy maintained, on the receiving entities, and on the credentials (or the lack thereof) in the request, the sharing entity may choose to respond with the resource or information pertaining to the resource, request additional credentials from the requesting entity or not respond at all.

4.4. Flow of Information through NaradaBrokering

Figure 7 illustrates the flow of information within NB given a client who wishes to compute the factorial of some integer. This client, having no knowledge of available services, solicits the help of a scheduler that identifies one or more services capable of accomplishing the requested task, and chooses one among them based on computational load.

1. The client publishes a message to the topic “Sch”, subscribed to by all schedulers.
2. All schedulers publish a response including their own ID (used as a topic name) and information about their current load.

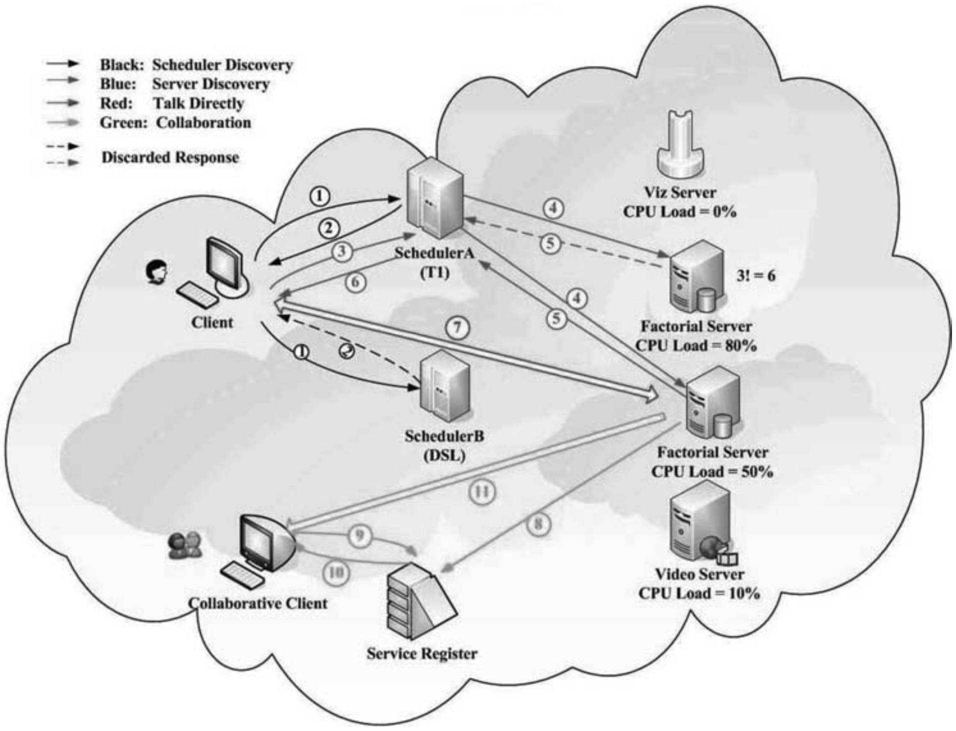

Figure 7
Flow of information between clients and web services through the control of a scheduler.

3. The client chooses one scheduler (SchedulerA) and publishes the service name (Factorial) to its ID.
4. SchedulerA publishes a message to topic ‘‘Factorial’’ subscribed to by all service providers that can provide the Factorial service.
5. Each service provider that provides a factorial service publishes a message that includes a unique topic ID and load information.
6. The scheduler chooses one service provider and sends its topic ID to the client.
7. The client retrieves the topic ID of the factorial service, should it be necessary for client and services to establish a direct link for reasons of efficiency (e.g., for the transfer of bitmaps). (The following steps pertain to collaborative clients):
8. The chosen service provider will register the current session to a service register and unregister itself at the end of the session.
9. The collaborative client sends a request to the service register to get the IDs of current running sessions.
10. The service register returns the list of current running sessions to the collaborative client.
11. New clients can now join the session using the server and client IDs. All clients that collaborate will, under the current setup, see identical displays.

The architecture of Figure 7 has already been implemented. We are currently working on strategies to fold web services into the framework.

4.5. Advantages of the Proposed Scheme

A system like NaradaBrokering, based on publish/subscribe has many advantages once it is coupled with the naming and discovery scheme just described:

1. No constraints are imposed on the number or location of clients, resources, task schedulers, or topic managers connected to NB.
2. The system allows for asynchronous discovery of task schedulers, available resources, resource metadata, file locations, etc.
3. Both task schedulers and resources can accept tasks based on an estimate of load.
4. The system has the ability to maintain audit trails to track system usage, task execution, and task execution failures.
5. In the event of failures, it will be possible to return partial results and the status of the computation to the client.
6. The middleware supports built-in collaboration mechanisms: Task results can be viewed by multiple users sharing one or more task IDs. Additional filtering may be required to allow for different client devices (PDA, laptop, etc.) (LEE *et al.*, 2003).
7. Task updates can be published to a "Task update" topic, perhaps with an associated task ID, that clients subscribe to separately. This would allow clients to keep track of task execution, similar to monitoring a FedEx package through a tracking number.
8. NB provides the option to cache results to improve performance in the presence of repetitive tasks.

5. Portals, Visualization Web Services

Recent efforts have focused on the creation of a web portal to enable users to interrogate their data sets stored on a remote server (GARBOW *et al.*, 2001; 2005a,b). These efforts have led to the **WebIS** system (Fig. 8). The existing framework promotes user interactivity and portability across wired and wireless networks. The challenge was to simultaneously provide users with the flexibility to explore their data while minimizing the effects of network latency. Currently, the users choose one of several precomputed slices through a data set that is downloaded to the client. After an initial delay, the user then has the ability to zoom, translate, perrform statistical analysis over a subregion and display histograms. Some magic lens technology is also implemented. A magic lens is a localized spatial region through which alternate information can be viewed. For example, thermal conductivity can be viewed through a lens that overlays the temperature field. These concepts are extended in YUEN *et al.* (2004) to include the remote visualization of earthquake clusters over the

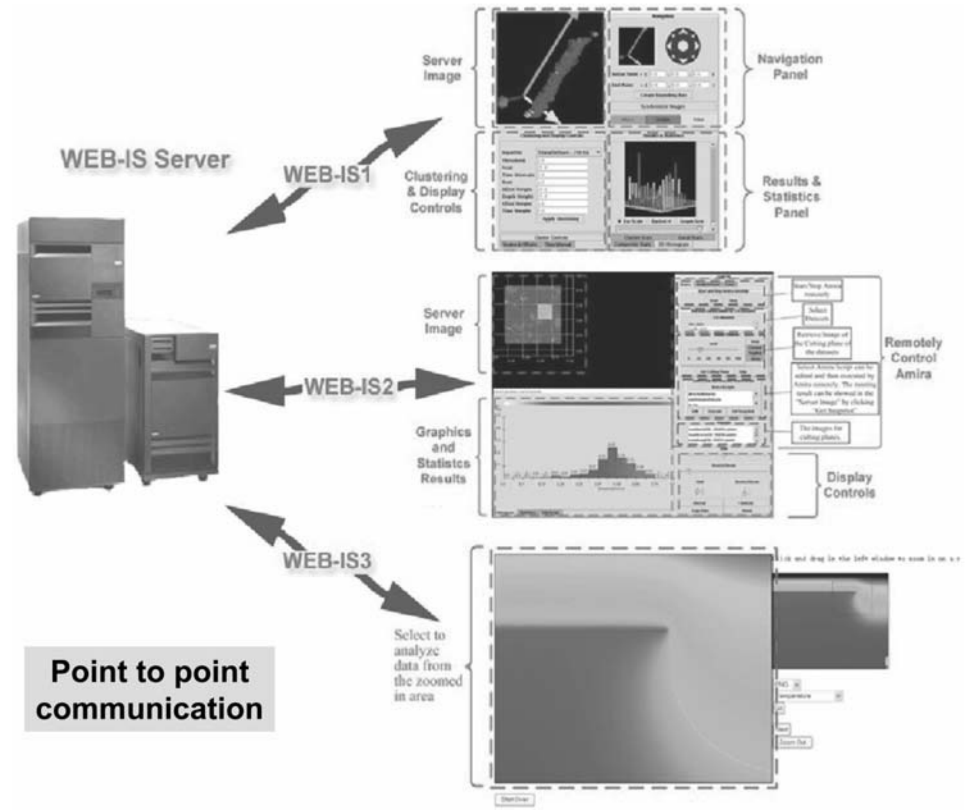

Figure 8

WebIS is a software system that allows the visualization and analysis of large data sets that reside on a server or a client. At this time, the communication between client and server is point-to-point.

network combined with off-screen rendering. This suggests the use of server farms dedicated to visualization via Web services.

Service nodes on the Naradabrokering framework described above will be populated with the server side of the **WebIS** system, along with several other visualization-related web services, such as image/video processing and movie creation. On the client side, user interfaces will be crafted and adapted to the variety of devices in use. These interfaces will be decoupled from the server side and communicate with them through web service specifications (WSDL files).

One of the most important challenges faced by these types of portals is compatibility across a range of devices. The largest displays have on the order of 10^6–10^7 pixels, while current (IPAQ, Jornada, OQO) handheld devices have much smaller resolutions, on the order of 10^4–10^5 pixels. This extreme range makes it virtually impossible to design a consistent user interface. An interface designed for a large screen will be very tedious to interact with on a small screen. On the other hand, a user interface crafted for a handheld device wastes a lot of space on the larger screen of a workstation.

The range of networking speeds is also an issue. Our middleware will use the built-in protocols of NaradaBrokering, supplemented by compression algorithms suitable to the transmission of encoded imagery and thresholded wavelet transforms. Furthermore, while command messages will travel through the middleware system, be logged, discover available services and execute commands on them, resulting images *may* be transmitted directly to the client for reasons of efficiency.

6. Future Trends and Perspectives

There has been insufficient attention paid to the efficient analysis and facile means of visualization and display of large-scale data sets in the earth sciences. Initiatives in computing, such as iServo (AKTAS *et al.*, 2005) (http://quakesim.jpl. nasa.gov), CIG (http://www.geodynamics.org) in geodynamics, and the Vlab in computational mineral physics (http://www.vlab.msi.umn.edu) will have to take a leadership role in promoting remote visualization and the practical use of allied services, such as **WebIS**.

Education is an important aspect of this program and workshops are sorely needed in areas that span the disciplines of visualization and standards-based web technologies (Java, XML, web services). By hosting several meetings, ACES has already played an important role in laying the foundation in these areas. Inroads at the national level have also been visible through several special symposiums held at the Fall American Geophysical Union meeting since 2001.

In this paper we have described a workable framework suitable for carrying out remote visualization by geographically distributed researchers. To this framework, we will connect a variety of visualization-related web services and address the issue of efficiency: Minimize communication and increase user interactivity. To move ahead, we must develop regional visualization centers, such as present at NCSA at the University of Illinois, which has recently installed large shared-memory (on the order of Terabytes) visualization facilities to enable entirely new classes of data analysis and visualization. We hope to use such powerful machines to host web services made available to the Earth Science community. These services would go a long way towards enhancing multi-way communication and exchange of data, which would be of great utility in times of disaster, such as the recent earthquake and tsunami in Asia in late 2004.

Acknowledgments

This research has been supported by the Geophysics, Advanced Computational Research, Math-Geo, Information Technology Research (VLab project: ITR-0428774) and the Advanced Computer and Infrastructure and Research programs of the National Science Foundation, the Dept. of Energy and the Environmental Protection Agency. We would also like to thank Geoffrey Fox for useful discussions on the project.

REFERENCES

AKTAS, M., AYDIN, G., DONNELLAN, A., FOX, G., GRANAT, R., GRANT, L., LYZENGA, G., MCLEOD, D., PALLICKARA, S., PARKER, J., PIERCE, M., RUNDLE, J., SAYAR, A., and TULLIS, T. (2005), *iSERVO: Implementing the International Solid Earth Research Virtual Observatory by integrating computational grid and geographical information web services*, Pure Appl. Geophys. *163*, 2281–2296.

BEYNON, M., FERREIRA, R., KURC, T. M., SUSSMAN, A., and SALTZ, J. H. (2000), *Datacutter: Middleware for filtering very large scientific data sets on archival storage systems.* In *IEEE Symposium on Mass Storage Systems*, 119–134, http://citeseer.nj.nec.com/beynon00datacutter.html.

BULUT, H., FOX, G., PALLICKARA, S., UYAR, A., and WU, W. (2003), *Integration of Naradabrokering and audio/video conferencing as a web service.* In *IASTED International Conference on Communications, Internet, and Information Technology 2002.*

COHEN, RONALD, E. E. (2005), *High performance computing requirements for the computational earth sciences*, Tech. Rep., Carnegie Institution of Washington, http://www.geo-prose.com/computational_SES.html.

ERLEBACHER, G. and YUEN, D. A. (2004), *A wavelet toolkit for visualization and analysis of large data sets in earthquake research*, Pure Appl. Geophys. 2215–2229.

FOSTER, I. and KESSELMAN, eds., *The Grid. Blueprint for a New Computing Infrastructure* (Morgan Kaufmann Publishers, Inc. (1999)).

FOSTER, I., KESSELMAN, C., NICK, J. M., and TUECKE, S. (2002), *The physiology of the grid, an open grid architecture for distributed systems integration*, Tech. Rep., Global Grid Forum, http://www.globus.org/research/papers.html.

FOX, G. and PALLICKARA, S. (2001), *An approach to high performance distributed web brokering*, Ubiquity *2*(38).

FOX, G. and PALLICKARA, S. (2002a), *An event service to support grid computational environments*, Concurrency and Computation: Practice and Experience, *1*, 1–34.

FOX, G. and PALLICKARA, S., *The Narada event brokering system: Overview and extensions.* In *Proceedings of the 2002 International Conference on Parallel and Distributed Processing Techniques and Applications* (PDPTA'02) (ed. H. Arabnia, *I*), pp. 353–359 (CSREA Press 2002b).

FOX, G. and PALLICKARA, S., *Naradabrokering: An event based infrastructure for building scaleable durable peer-to-peer grids.* In *Grid Computing: Making the Global Infrastructure a Reality* chap. 22, 579–600 (John Wiley, West Sussex, England (2003)).

FOX, G., GANNON, D., KO, S.-H., LEE, S., PALLICKARA, S., PIERCE, M., QIU, X., RAO, X., UYAR, A., WANG, M., and WU, W. (2003), *Peer-to-peer grids.* In *Making the Global Infrastructure a Reality* (ed. F. Berman, G. C. Fox, and A. J. G. Hey), pp. 471–489 (John Wiley, West Sussex, England 2003).

GADGIL, H., FOX, G., PALLICKARA, S., PIERCE, M., and GRANAT, R. (2005), *A scripting based architecture for management of streams and services in real-time grid applications.* In *Proceedings of the IEEE/ACM Cluster Computing and Grid 2005 Conference (CCGrid 2005)*, to appear, http://grids.ucs.indiana.edu/ptliupages/publications/HPSearch-mgmtArch. pdf.

GARBOW, Z. A., OLSON, N. R., YUEN, D. A., and BOGGS, J. M. (2001), *Interactive web-based map: Applications to large data sets in the geosciences*, Visual Geosciences *6*, 3, 1–14.

GARBOW, Z. A., ERLEBACHER, G., YUEN, D. A., BOGGS, J. M., and DUBUFFET, F. W. (2003), *Web-based interrogation of large-Scale geophysical data sets from handheld devices*, Visual Computing. *8*, 1, 1–20.

GARBOW, Z. A., ERLEBACHER, G., YUEN, D. A., SEVRE, E. O. D., NAGLE, A. R., and KANEKO, J. Y. (2002), *Web-based interrogation of large-scale geophysical data sets and clustering analysis of many earthquake events from desktop and handheld computers*, in *EOS, TRANS, AGU abstract* .

GERYA, T. V. and YUEN, D. A. (2003a), *Characteristics-based marker-in-cell method with conservative finite-differences schemes for modelling geological flows with strongly variable transport properties*, Phys. Earth Planet. Int. *140*, 295–320.

Gerya, T. V. and Yuen, D. A. (2003b), *Rayleigh Taylor instabilities from hydration and melting propel 'cold plumes' at subduction zones*, Earth and Planetary Sci. lett. *212*, 47–62.

Lee, S., Ko, S., Fox, G., Kim, K., and Oh, S. (2003), *A web service approach to universal accessibility in collaboration services*. In *Proceedings of 1st International Conference on Web Services, Las Vegas*.

Pierce, M., Fox, G., Youn, C., Mock, S., Mueller, K., and Balsoy, O. (2002), *Interoperable web services for computational portals*. In *Proceedings of High Performance Networking and Computing SC'02*.

Pierce, M., Youn, C., and Fox, G. (2003), *Interacting data services for distributed earthquake modeling*. In *International Conference on Computational Science*, pp. 863–872.

Rudolph, M., Gerya, T. V., Yuen, D. A., and De Rosier S. (2004), *Visualization of complex multiscale phenomena at subduction zones*, Visual Geosciences. SERVO (2002), Solid Earth Research Virtual Observatory, http://www.servogrid.org.

Uyar, A., Pallickara, S., and Fox, G. (2003), *Towards an architecture for audio video conferencing in distributed brokering systems*, Communications in Computing , 17–23.

Vincent, A. P. and Yuen, D. A. (2000), *Transition to turbulent thermal convection beyond* $Ra = 10^{10}$ *detected in numerical simulations*, Phys. Rev. E *61*(5), 5241–5247.

Yuen, D. A., Garbow, Z. A., and Erlebacher, G. (2004), *Remote Data Analysis, Visualization and Problem Solving Environment (PSE) Based on Wavelets in the Geosciences*, Visual Geosciences, Online First.

(Received December 30, 2004, revised August 10, 2005, accepted August 12, 2005)