BUILDING A NATIONAL COLLABORATORY

Stewart C. Loken
Lawrence Berkeley National Laboratory
Berkeley, CA, 94720

Abstract:

The concept of a collaboratory was introduced in a study for the National Academy of Sciences. Bringing together collaboration and laboratory, a collaboratory is intended to provide a mechanism for scientists to work together and to share unique research facilities like instruments, computers and stored data, often more than one at a time, in ways that would not be possible without high-speed networks. The Department of Energy, through its DOE2000 program, has initiated an effort to develop and enhance the tools needed to make all of the DOE facilities accessible to scientists in the United States and to improve the ability of those scientists to work together. This paper will describe recent progress in the development of tools to support collaboratories including video conferencing, electronic notebooks and shared virtual spaces. It will also review progress on the underlying technologies needed to make collaboratories work such as collaboration management, interoperability frameworks and network protocols. The paper will also describe efforts to incorporate these technologies into ongoing science programs and summarize the lessons learned from working with these scientists.

Background

The concept of a collaboratory was introduced in a study for the National Academy of Sciences[1]. That study described the concept of a collaboratory as a

> *"... 'center without walls', in which the nation's researchers can perform their research without regard to geographical location - interacting with colleagues, accessing instrumentation, sharing data and computational resources, [and] accessing information in digital libraries."*

Bringing together collaboration and laboratory, a collaboratory is intended to provide a mechanism for scientists to work together and to share unique research facilities such as instruments, computers and stored data, often more than one at a time, in ways that would not be possible without high-speed networks. The Department of Energy, through its DOE2000 program, has initiated an effort to develop and enhance the tools needed to make all of the DOE facilities accessible to scientists in the United States and to improve the ability of those scientists to work together.

The collaboratory effort has three major goals:

- To provide access to unique DOE resources by other labs, universities and industry
- To enable and promote collaboration between labs, universities and industry
- To open DOE facilities and expertise to educators and to industry

The plan for the Collaboratory Program was developed in series of workshops involving both computer scientists and potential users. The focus was on developing new capabilities and enhancing existing tools to make them more useful to researchers. The initial selection of projects covered the full range from network protocols and services to collaborative applications. The projects are:

- Electronic Notebooks
- Floor Control for Video Conferencing

- Shared Virtual Reality
- Software Infrastructure or Framework
- Collaboration Management
- Quality of Service
- Security Infrastructure

It was recognized that the technology projects needed to have close interactions with existing and future users. In addition, it was clear that the program needed to develop plans to support the tools as they mature. All of the R&D Projects have made significant progress since the DOE2000 program began and the tools are now being deployed into a number of collaboratory projects. A repository has been established at LBNL to distribute emerging tools.

Collaboration Tools

The tools to support collaborative work can be separated into those that have persistent information and those that require synchronous interaction among the collaborators. Persistent information may be accessed by collaborators at any time and includes the following:

- Email
- News groups
- Papers and reports
- Mail
- Electronic Notebook

Synchronous tools include:

- Telephone
- Video Conference
- Chat/White board
- Shared authoring & applications
- Shared VR space
- Instrument control

Electronic Notebooks

An important new technology for the Collaboratory is the electronic notebook; it brings together all aspects of the persistent information from the collaboration. This technology has the advantage that it can be shared by remote collaborators and it is always available for input or reading. It can contain rich media types such as text, images, files, 3D structures, voice, animations and video and it can take input directly from computers and instrument. It permits easy transfer of information from one notebook to another, has a simplified notarization process, allows complex querying/ searching and can include hyperlinks to other data and references. The goals of the DOE2000 Notebook [2] project were to:

- Design a common, open Notebook Architecture
  - extensible as technology advances
  - interoperable with other notebook viewers
  - customizable for unique inputs of a given project

- Develop prototype implementations
  - make them available to DOE collaboratories
  - general research community
  - education
  - industry

Video Conferencing

The video conferencing tools bring together most of the synchronous information-sharing described above. As part of the DOE2000 program, work has continued on the MBone Tools developed at LBL [3] and used by many HEP groups including the CERN-based Virtual Room [4]. To enhance these tools, a new Conference Controller has been developed that enables remote control of conference tools and cameras. This is extremely important for remote seminars and conferences as it is the remote users who are best able to judge what is wrong with a transmission and who have the biggest stake in a good quality broadcast. It has also become clear that the MBone tools do not meet all needs, especially on platforms other than Unix, and as a result, other tools are being used for specific projects. These include:

- NetMeeting
- PictureTalk
- Streaming JPEG
- CUSeeMe

All of these have features that make them useful to collaborations but none currently provides the full set of desired features on all platforms in use by the scientific community.

A feature that has been missing from all the conferencing tools is Floor Management, the ability to recognize remote participants who wish to comment or ask a question. This problem is being addressed in one of the DOE2000 R&D projects. This tool will provide floor control and mediation for MBone conferencing tools and will plug into the existing protocol support. It will support two coordination models, moderated meetings and consensus meeting.

Information or software for conferencing is available at a number of sites [5]

Shared VR

The concept of Shared Virtual Reality takes the collaboration beyond video conferencing into a virtual world where scientists from remote locations share space as if they were in the same room. The DOE2000 project is designed to explore the potential usefulness of this concept for scientific collaboration. Its goals are the following:

- Investigate the **integration** of collaboration technology with immersive virtual environments
- Develop **tele-immersion concepts** and demonstrate them on interesting DOE applications areas
- Explore the **notion of persistence** in shared collaborative spaces
- Investigate alternative **software architectures** for developing tele-immersion applications
- Develop new concepts for supporting wide area collaboration and **man-machine interactions**
- Investigate large-scale VR for **collaborative design**

As part of the project, the group at Argonne National Laboratory has developed Many Worlds [6] which is an Object-Oriented core software environment for developing tele-immersive applications. It enables synchronous sharing of VR applications and provides multiple users, multiple worlds, multiple environments.

It has navigation tools, avatars, control, user-interaction, communications and session management services. It provides persistent world servers and it works on the CAVE family of display environments. It will be some time before these techniques permit the implementation of a fully immersive control room for a HEP experiment but this work is demonstrating some interesting ways of working together.

Integration Framework

Collaboratories are currently built of a set of tools that do not plug together into a coherent framework. To ensure widespread acceptance of the tools, it is necessary to produce a distributed computing architecture that will support the development of collaborations. This infrastructure will include a common communication library including multicast and unicast with various reliability levels. The objectives of the DOE2000 [7] project are the following:

- Facilitate development and interoperability of collaboratory components by providing:
  - Convenient access to unicast and multicast messaging
  - Common communication API for unicast and multicast communication
  - Reliable multicast communication
  - CORBA evaluation and integration
  - Directory services
  - Integration of security
  - Access from multiple languages (Java, C++ and C)

The framework must support various communication protocols including unreliable, unordered (UDP) and reliable source ordered (TCP) Unicast as well as many types of Multicast, unreliable, unordered (IP Multicast, RTP), unreliable source ordered (filtered IP Multicast), reliable source ordered (XTP), and reliable totally ordered delivery (Totem, InterGroup).

The project has created a common communication API that is described in a document available from WWW pages. The prototype implementation of a previous version of API was completed in Java and a prototype implementation over Nexus nearing completion in C++. In addition, prototype application uses are under development including the DOE 2000 electronic notebook.

Security and Authentication Architecture

The widespread use of collaboration tools and especially the remote use of experimental facilities requires distributed security architectures that are flexible, effective and easily deployed, administered and used for:

- Authentication
- Authorization
- Access control
- Confidentiality
- Infrastructure protection
- Distributed enterprise

Fortunately the requirements of the scientific community are very similar to financial services industry. Both can use Public Key Infrastructure and general certificate content interpretation. The DOE2000 project [8]is evolving a use-condition centered model and architecture that provides verifiable use conditions and ensures secure and verifiable satisfaction of those conditions. The project has designed and implemented the "baseline" digital signature software that uses X.509 identity certificates, multiple stakeholder use-condition

certificates, use-attribute certificates. It has implemented a "policy engine" for collecting, validating, and matching the certificates.

Internet Quality of Service

The ability to differentiate types of service is crucial for the success of collaboratory efforts. The Internet today is based on a “best effort” model that treats all traffic as having the same priority and transmits it when space is available. This model is not acceptable when a collaboration needs real-time interactions such as a video conference or the control of an experiment. To address this issue, the DOE program has introduced a differentiated services model that labels high-priority packets so they can move to the front of each queue at each router. Because the priority information is attached to each packet, this model does not require advance setup in the network. Processing within the backbone of the network is reduced to the lowest possible level.

The determination of which packets get priority labels is handled on the outer edges of the network by system known as a “bandwidth broker”. This broker looks at a request for priority service, validates the request through the security and authentication service, and permits the priority if the requested priority bandwith fits within the total priority traffic permitted for the institution requesting it.

This model has been designed and implemented on the ESnet link between LBNL and Argonne National Lab using early versions of Cisco router software. The developers are now working with the Internet Engineering Task Force (IETF) to standardize the approach and both ESnet and Internet2 are planning to deploy the service in the near future.

Collaboratory Pilot Projects

The pilot projects were designed to test the emerging technology and give feedback to technology developers. As part of the DOE2000 program, the managers identified two major projects: the Diesel Combustion Collaboration and the Materials Microcharacterization Collaboratory. Within the program, some other efforts are going ahead with other (limited) funds. In the Diesel Collaboratory [10], the focus is on the design of next generation of engines which must meet very tight emissions limits. This program uses collaborative computation as well as experiments to validate models. It has strong connection to the diesel engine manufacturers, DOE laboratories and university researchers. Some of the features of the Diesel Collaboratory are shared combustion models with the ability to steer the computation as preliminary results emerge and a library of combustion images from models and experiment. The collaborators make extensive use of conferencing tools and notebook technologies. The features of the Materials Microcharacterization Collaboratory [10] include the definition of a common interface to instruments at all sites, remote control of instruments, and *In-situ* experiments using computer control that monitors the progress of the experiment and makes adjustments automatically. The collaborators also use video conferencing, streaming video and electronic notebooks.

Many of the issues that have emerged from the pilot projects are common to other disciplines:

- Security, especially for proprietary data
- Network Infrastructure at partners’ sites
- Concerns about connections to Internet
- Security to protect instruments
- Avoiding “least common denominator” for instrument features
- Diversity of platforms

Signs of Progress

A number of lessons can be learned by observing the pilots. The pilots have both used a mix of commercial software and custom applications. None of the tools is a perfect fit to needs and the collaboratory tools are not well integrated with each other or with other packages already in use.

Despite problems and rough edges, however, the tools are being used in a number of other scientific collaborations. Industry partners do seem to be joining in the collaborations although less quickly than university and laboratory scientists. New collaboratory projects are starting now that some of the pioneers have shown the usefulness of these ideas. More information on the DOE2000 program and the Collaboratory Tools is available [11]

References

1. National Collaboratories: Applying Information Technology for Scientific Research, National Academy Press, 1993.
2. www.epm.ornl.gov/enote
3. www-itg.lbl.gov/mbone
4. P. Galver, H. Newman, G. Denis, Ch. Isnard, paper 103, these proceedings
5. DOE2000.lbl.gov/doe2k
6. www-fp.mcs.anl.gov/division/research/vr_summary.htm
7. www-itg.lbl.gov/CIF
8. www-itg.lbl.gov/security arch
9. DCC www-collab.ca.sandia.gov/Diesel/ui.new
10. MCC tpm.amc.anl.gov/mmc
11. DOE2000. www-c.mcs.anl.gov/DOE2000