Pervasive Collaborative Computing Environment (PCCE)

http://www-itg.lbl.gov/Collaboratories/pcce.html
PI: Deb Agarwal (DAAgarwal@lbl.gov)

Executive Summary
September 2001

Vision:

Collaboration is not accomplished only through a single interaction mode, such as face-to-face conversation. Within a collaboration communication will normally progress through several interaction modes including email, telephone calls, videoconferencing, meetings, etc. Many of the scientific collaboratory tools available today focus on highly interactive tools such as videoconferencing, which are useful for meetings and presentations. But the rest of the time, collaborators need less intrusive and more flexible ways to stay in touch and thus work together. Time-ordered asynchronous communication is a particularly pressing need for collaborations with participants located around the globe.

The goal of this project is to research, develop, and integrate the tools required to support a flexible, secure, seamless collaboration environment that supports the continuum of interactions between collaborators. This environment is envisioned as a persistent space -- the Pervasive Collaborative Computing Environment (PCCE) -- which will allow participants to locate each other, use asynchronous and synchronous messaging, share documents, share progress and results, share applications, and hold videoconferences. It will leverage existing and recently proposed tools such as Grid services, Internet Relay Chat (IRC), Web Distributed Authoring and Versioning (WebDAV), electronic notebooks, and videoconferencing capabilities. Furthermore, we recognize shared development and management of computing activities as a common collaboration task present in many disciplines. We feel that shared collaborative control and monitoring of these tasks constitutes an important aspect of the work performed in these environments. In addition to facilitating interactions between collaborators, we will implement workflow management tools that allow shared management of this important aspect of a collaborations daily activities. By basing the PCCE tools on the DOE Science Grid computing environment, we hope to maximize their applicability to a wide range of collaborative research efforts.

Major Goals and Technical Challenges:

Much of the real work of collaboration is accomplished through informal and often asynchronous mechanisms and the need for such informal communications is increasing. Collaborators currently revert to email and telephone to support these interactions because present tools lack the asynchronous and ad hoc communications mechanisms they need. Furthermore, with the advent of grids, in particular the DOE Science Grid, the need for effective collaboration tools to support collaborative computation is increasing. One of the significant new tool requirements introduced by these computation-based collaboratories is the need for workflow tools to coordinate computational work and human processing. The problem of asynchronous communications centers, in brief, on the desire to carry on meaningful, time-ordered exchanges with parties that are not able to provide immediate responses to incoming queries. A conversation with a collaborator should be able to flow naturally from synchronous to asynchronous and back again. The ease of use, privacy and security play a major role in acceptance of the technologies and thus must be addressed in the fundamental system.

Recent research has shown that messaging systems and presence awareness mechanisms help to provide some of the types of interactions that make these collaborations successful. Presence and messaging systems provide a basic role in connecting the collaborators in a shared context that provides very basic connectivity capabilities. By increasing a collaborators' ability to hold spontaneous meetings, these capabilities increase the level of social interactions among collaborators. A key challenge will be to improve the support of asynchronous communication in these messaging and presence awareness systems. In the area of web-based, shared authoring environments, the

WebDAV effort (Web Distributed Authoring and Versioning) sponsored by the IETF has made great progress in extending HTML to support shared authoring and controlled access to web documents. The increased availability of robust WebDAV document servers will provide the needed infrastructure for ubiquitous file and document sharing. However, the user interface and usage semantics that mediate the use of these tools remains in a fairly immature state. An important challenge will be to develop use case strategies and interface methods that provide a natural match between collaborators needs and the capabilities of these standards.

In the area of scientific workflow management, we have examined available offerings in the commercial world. Although there seems to be no shortage of activity in this area, current toolsets all appear to suffer from similar shortcomings. First, they lack the ability to specify and effectively manage the execution of simple programs on most unix-like computing platforms. Since these workflow systems are primarily targeted at tasks such as "perform a data base query" and "wait for web form submission", they lack the necessary semantics to invoke simple programs and pass input data and parameter file names. Secondly, business workflow models typically have very limited models of task parallelization. Since many scientific computing problems are designed to take full advantage of available computing resources for execution of “trivially parallel” analysis and simulation tasks, this limitation is a major barrier to their use. This project will therefore concentrate on a persistent set of services that will allow submission, control, and sequencing of computing tasks that, together, constitute a large, collaboration-wide computing task. By targeting these services for execution within the DOE Science Grid computing environment, we will be able to leverage existing Grid security, authorization, and information services to serve both the collaborative communications environment and scientific workflow management tools. The challenge faced by this project is the design, implementation, and integration of the tools required to create a collaboration environment that supports as many of the daily collaboration activities as possible.

Major Milestones and Activities:

Year 1

- Evaluate existing collaboration tools for possible use in the environment
- Prototype integration of available Grid security and directory services
- Install Condor-G and develop a program based interface to the Condor scheduler
- Develop a prototype workflow model based on the Supernova Factory data processing

Year 2

- Integrate Grid security and directory services and COG toolkits
- Integrate WebDAV-based services and SOAP interface for workflow
- Add reliable and recoverable event reporting to Condor’s DAGman

Year 3

- Integrate collaboration and workflow interfaces
- Develop workflow exception handling
- Integration of an electronic notebook and terminal sharing capability into the environment
- Evaluate collaboration environment

Current Connections with Other SciDAC Projects:

DOE Science Grid, Reliable and Secure Group Communication