

PET White Paper

# Network-Based Remote Collaboration and Training: Progress and Plans

David E. Bernholdt, Nancy McCracken, Marek  
Podgorny and Geoffrey C. Fox

*{bernhold,njm,marek,gcf}@npac.syr.edu*

Northeast Parallel Architectures Center, Syracuse  
University

10 May 1999

## 1. Introduction

Since its inception, the PET program has heard calls from users for access to training "anytime, anywhere", and for assistance in taking advantage of modern networking and tools to facilitate both training and research collaboration among geographically dispersed groups. These very logical requests pose a challenge to the current state of the art in these areas, to which the PET program has responded with an effort to develop tools, techniques, and content to support remote training and collaboration. We have now reached the point where certain capabilities are being deployed for more routine use (especially in the training area), and plans are in place for the staged deployment of other capabilities (primarily in the collaboration area). This White Paper briefly describes the progress that has been made to date, the deployment plans, and the rationale behind them. This represents a snapshot of a rapidly moving field taken from our vantage point both as academic researchers in the area, and as participants in the PET program.

## 2. Terminology

When speaking of both training and collaborative activities, it is useful to divide the area into "synchronous" and "asynchronous"

modes of interaction. Synchronous interactions require all parties to be available at the same time. Traditional classroom teaching and telephone calls are synchronous interactions. Asynchronous interactions don't require everyone to be available in real time, and include voice-mail, e-mail, the web, and other familiar forms of communication. Asynchronous training/education includes, for example, self-study of a textbook, or of materials published on the web.

Our belief in the PET program is that both synchronous and asynchronous training is appropriate in the DoD HPC community depending on the circumstances. In some cases, "students" will be motivated enough, and have sufficient background to successfully self-study available material, while at other times they will need the structure, educational support, and interactivity provided by the synchronous mode of operation.

### **3. Asynchronous Collaboration**

Simple asynchronous collaboration is already so pervasive that it is rarely thought of in these terms. E-mail, mailing lists, and newsgroups are all based on the sending and receiving of messages. The world-wide web is probably the other principal async. collab. tool. We have found that these basic tools can be usefully augmented in some fairly straightforward ways:

- (web-accessible) archiving of mailing lists
- use of page-design tools to help provide a consistent look and feel in a web site, to aid navigation
- use of web-linked databases and other back-end tools to facilitate the management and presentation of large volumes and/or frequently changing data
- web search engines focused on a particular knowledge domain or problem area

It is possible to deploy far more complex and sophisticated tools for asynchronous collaboration, for example a Lotus Notes-based intranet, however this must be done with caution. We have observed cases where complex tools, requiring a significant effort

to develop, have essentially been ignored by the user community because they are harder to use than basic web/e-mail without providing sufficient benefit, or because they lack a high level and high quality of connectivity with the basic tools. (For example, many find Lotus Notes URLs hard to remember and navigate, making it hard to jump quickly to the information you want. Within the PET program, we do plan to investigate async. collab. tools further, but at present it appears that the mildly augmented set of basic tools are quite effective.

One particular area that does seem to be under-exploited is the use of web-linked databases. This is not because the technology isn't available, but rather because the tools and techniques are often unfamiliar to computational scientists compounded by the fact that they are tools rather than a finished products. A few examples of this technology are already available or under development within the PET program: the domain-specific search engine mentioned above, as well as several examples in the "Training Infrastructure" area, described briefly below. But users (MSRC users, PET partners, and others) need to be encouraged to think more about applications of this technology to problems in their CTA.

#### **4. Asynchronous Training**

For practical purposes, async. training at the moment amounts to making training courseware available via the web in some fashion. However a great deal of variability is possible in the quality and sophistication of the courseware -- it can range from material that is designed to accompany a lecture, which may not be useful without the lecture, to material specially crafted for stand-alone use. Since high-quality stand alone material can be expensive and time-consuming to develop, there can be some attraction in the simpler approach. The PET program has, and continues to make use of both approaches, since both ends of the spectrum can be useful and practical under different circumstances.

An asynchronous training resource can be created relatively easily by capturing a live presentation of the material. Simply recording

the presentation on videotape is an obvious approach that has been used at most of the MSRCs at least to some extent. Some experiments have also been done with videotaping the instructor without a class present, and with editing the video to produce a more refined presentation. Typically, however, the management and duplication of video tapes has required too much of the Training staff's time to consider this an efficient approach. And in practice such videos, are not often used. One site, with a video library of roughly 250 different events (trainings, workshops, seminars), reports around one use or outside request per month. By digitizing the a/v stream, it can be distributed easily on the web. This makes the captured lecture a more cost effective and easier training resource to produce and distribute. This has been done in a number of cases using RealNetworks' proprietary RealPlayer system, usually from a videotape recording, and requiring manual synchronization between the a/v stream and the slide changes. Under the PET program, Syracuse University has developed an integrated hardware/software system which goes a step further in the automation of the process. The "LecCorder" is a PC system with a commercial MPEG-1 encoder board. It can take a live a/v feed, i.e. from a class, and produce an archival-quality MPEG-1 digital audio/video stream, as well as capturing the instructor's slide changes during the presentation. Off-line, the high-quality, high-bandwidth a/v stream can be "down-converted" into more network-friendly formats such as H.261 and H.263 (videoconferencing standards) or RealPlayer. Combined with the "clickstream" and the lecture slides themselves, an asynchronous training resource can be produced and published on the web quickly and with almost no human intervention. Lectures captured using LecCorder and other methods are available or planned at all of the MSRCs; there is no data yet on the extent of their use. It is worth noting that these approaches require high-quality recordings to work from -- s-video or better; standard VHS recordings will generally not produce acceptable results.

With a greater investment of time, it is possible to develop online

material of greater depth, as an alternative to augmenting it with recorded lectures. This is the approach taken in the Cornell Theory Center's (CTC) Virtual Workshop series, which has also been used by the PET program. Besides having material which is meant to truly stand on its own, the Virtual Workshops are run in specific time frames, during which on-line and telephone-based consulting regarding the class are available. The PET program has sponsored a number of Virtual Workshops, for which CTC has charged \$20,000 each for up to 75-100 participants. Although relatively expensive, this approach has proven fairly popular, and more than 170 DoD users have taken advantage of them. CTC has about 40 modules (each equivalent to a single lecture or lab) from which a course can be designed. To our knowledge no one from the PET program has discussed with CTC the possibility of developing new modules, but this would presumably be a fairly expensive proposition given the depth and quality of the material.

Another asynchronous training tool of note is the CD-ROM of educational material on High Performance Computing which was compiled with support from the CEWES MSRC PET program and distributed at the 1998 DoD HPC Users Group Conference at Rice University. More than 300 of these discs have been distributed so far, and anecdotal evidence indicates that they have been well received, including many requests for discs from individuals who did not pick one up at the Conference or didn't attend. The second edition of this resource, which through additional and refined material has expanded to two discs, will be distributed at the 1999 DoD HPC UGC.

## **5. Synchronous Collaboration and Training**

Underlying both synchronous collaboration and training is some kind of toolset that handles the necessary interactions among participants in the session. A variety of tools are available with different capabilities and degrees of sophistication. Examples include mbone and CUSeeMe (audio/video conferencing, shared whiteboard), Microsoft NetMeeting (sharing Microsoft Office

tools), NCSA's Habanero and Syracuse University's Tango Interactive (both of which go well beyond simple a/v conferencing and shared whiteboards). All of these tools have some place in a general "collaboration toolbox". The PET program has, over the last several years, focused primarily on two systems: mbone and Tango Interactive.

In the beginning of the PET program, mbone tools were used to extend trainings to remote sites. This system offers audio/video conferencing and a whiteboard. The system suffered from a number of problems which eventually lead to its abandonment as a distance training tool. Many people found that they did not have access to the multicast networking capability, the quality of the a/v transmission (dependent on the quality of service of the network) often left much to be desired, and there were practical obstacles to using the whiteboard to present lecture slides from both the viewpoint of the instructor and of the students. Mbone tools were originally developed for unix platforms, and little if any development of this system is still going on. As a result, Windows and Macintosh platforms are not well supported, and even for more recent unix systems it can be hard to obtain the necessary drivers. More recently, PET activities have centered around the Tango Interactive system as a general, extensible framework for both education and collaboration. Development of Tango was begun under a DARPA initiative in 1996 as a C4I tool, and it has been refined and enhanced over the last several years, with support from various sources including the PET programs. It offers a much broader range of collaborative tools than mbone, including a number that were specifically developed for educational use. It supports Windows PCs (95, 98, NT, etc.) as well as several unix platforms and although it's actual performance is subject to the quality of service of the underlying network, most observers agree that it provides better quality audio and video than mbone in poor network environments. It is worth noting that the user perception of any synchronous collaboration/training system, including Tango, can be strongly affected by the network quality of service

of the underlying connections.

The initial application of Tango in the PET program was in distance education. This is a fairly well structured type of collaborative interaction, in which the instructor needs an understanding of the use and limits of the collaborative framework, while the students generally do not need as much. With the students located together in an electronic classroom, an appropriately trained support person can provide the required Tango expertise on the receiving end. Using this approach, Syracuse University in New York is now in its fourth semester of delivering regular, semester-long academic credit classes to Jackson State University in Mississippi. And in the current semester, the recipient base has expanded to include Clark-Atlanta University and Mississippi State University as well as an individual at the Waterways Experiment Station. Jackson State University has also begun using the same tools to deliver a course to Morgan State University in Maryland.

The experience gained in these experiments has been critical in guiding work on the Tango system, especially in pointing up where robustness needed improvement. It is also important to realize that Tango is merely a tool for education and collaboration, and that in addition to insuring that the tool functions as intended, it must also be used effectively. The on-going educational work involving Tango has also allowed us to explore some of the sociological factors which make this form of instruction different from traditional face-to-face educational settings. With this experience we have been able to modify and improve our methods to provide a better educational experience, and we are developing enough experience to begin looking forward to issues particular to the training situation, and to less structured environments.

Training is similar to the traditional academic environment in terms of being structured, but because of the compressed time frame, is less forgiving of the occasional problem, and provides less time for students to become comfortable with the tools. In conjunction with the July 1998 release of version 1.0 of Tango

Interactive, considered to be the first one to be appropriate for general deployment, the technology was transitioned into PET training activities.

In a collaborative effort involving the Ohio Supercomputer Center (OSC), Syracuse University, and the CEWES MSRC PET program, two prototype distance training classes were taught using Tango Interactive to deliver them to remote sites. In September 1998, a day-long training on Fortran90 was taught at the CEWES MSRC training room and delivered to the training facilities at the ARL MSRC (the HEAT Center) and OSC. In January 1999, a two-day class in OpenMP was presented at the CEWES MSRC training room and delivered to all three other MSRCs, the NRL Distributed Center, and OSC. In both cases, Syracuse University also monitored the class, but did not participate as students. As shown in Table 1, the January class reached more than 30 students in one of the most geographically distributed uses of Tango to date. OSC is also increasingly using Tango to deliver trainings it offers for its own users, and for trainings sponsored by the National Computational Science Alliance (NCSA). Once such class in February on "Java for Scientific Computing" was delivered from Ohio University to the Ohio Supercomputer Center and the Alliance ACCESS Center.

We are also extending this technology to other events similar in structure to training, such as academic-style seminars: in April, one of the authors (Fox) recently used Tango to present a seminar simultaneously to all four MSRCs without leaving Syracuse. The presentation was also recorded with our LecCorder system and has already been requested by several people unable to attend the original event. Integration of recording capabilities directly into Tango is also on the drawing board.

**Table 1.** Number of participants by site in the prototype Tango-based distance training/seminar events. The Java training was offered by the same OSC/NPAC collaboration that presented the two PET trainings, but was in this case sponsored by the NCSA



## Alliance.

<b>Site</b>		<b>Fortran90 28 Sep '98</b>
ACCESS	-	-
ARL	5	10
ASC	-	1
CEWES	13	10
NAVO	-	2
NRL-DC	-	2
OSC	6	9
Ohio U	-	-
<b>TOTAL</b>		<b>24</b>

### **6. Synchronous Training Dissemination Plan**

Based on these experiences, we have begun to understand some of the issues around broader deployment of Tango both for training and for collaboration. The two primary issues, which are related, are the fact that collaboration tools in general are not yet part of the standard computing environment, so that they are not familiar and natural to many users, and that as a consequence it is especially important to provide a high level of technical support as users are developing more experience with the tools. Moreover, the less structured the environment, the better the understanding of the tools which is required for their use to succeed -- or to even have a good understanding of how they might be used in a given collaborative situation. This has lead us to develop an overall plan for the deployment of collaborative tools into the Modernization Program which works from the most structured situations to the least, and in all cases begins with a small number of focused "experiments" in order to better understand the sociological and technical issues around each type of application.

1 Tutorials offered in use and support of Tango Interactive

2 Education & Training delivered to centralized facilities with support staff

3 Selected seminars and well-structured meetings (formal briefings, User Group Meetings, etc.) delivered to centralized facilities with support staff

4 Work with selected groups on collaborative applications

5 Education & Training delivered to selected desktops as base of experienced users and support staff increases

Tutorials on Tango itself have already been offered in a number of venues, and these offerings will increase as part of our effort to introduce more people to these tools and their capabilities. They are meant to compliment opportunities to learn about Tango through actually using it (items 2-5), and we expect that both formal trainings and hands-on experience will prove beneficial in producing interested users. As described above, item (2) has passed the experimental stage and is being deployed as a more routine aspect of PET training activities. Following from this, we are now beginning the experimental stage of items (3-5). While the importance of item (5) has been stressed by a variety of people, inside both the PET and user communities, it appears last in this plan because we believe that the experience gained from work on the other items will contribute significantly to our understanding of how to insure that direct-to-the-desktop delivery will be successful. Equally beneficial will be the increased base of instructors, support staff, and users familiar with Tango who can be called upon as part of a (formal or informal) support network.

## **7. Training Infrastructure**

Two current activities, supported by the ASC MSRC PET program, also bear mention. WebWisdomDB is a tool to manage and present large collections of (web-based) curriculum material. It supports both synchronous and asynchronous delivery of course materials to students, but more importantly it provides support to (one or more) instructors to manage the courseware. The system is built around the use of XML to specify access to a web-linked database. It exemplifies an important approach where XML is used to specify domain specific information in HTML-like tags. One

can view this system as an early prototype of a portal to education and training; an approach of increasing interest in PET with for instance the ASC "Gateway" activity (not discussed in this paper) viewed as a portal to computing and using XML to define web-based computing services.

The ASC Training Database will help simplify the currently labor-intensive tasks around training courses: registration, reminders, account setup and removal, and assessment. It integrates with existing MSRC procedures and data structures, and automates many tasks that are currently done manually. Both of these tools are designed to assist instructors and PET/MSRC staff, and have only a small exposure to the actual user community, but represent additional ways that web/network technology can be used to support and enhance the PET program's ability to provide training to the DoD user community. It is also worth noting that both of these projects are, at their core, web-linked databases, and thereby provide additional examples of how web-linked database technology might be used to support communication and collaboration among DoD users.

## **8. Related Activities**

There are of course many important national projects in the area of web-based training. DoD's ADL (or Advanced Distributed Learning) activity has made impressive use of web-linked databases to host asynchronous training material. The IMS (Instructional Management System) project from EDUCAUSE has taken the lead in building industry consensus on technical standards for an internet architecture for learning, and proposed standards for metadata in the area. The popular WebCT system provides automated authoring support and important tools such as glossaries and quizzes. The focus of these systems has been technology for authoring and asynchronous delivery of training material. As such they are synergistic with most of the PET initiatives for Tango Interactive is designed to support "any" web authoring system whether it be stored in a database or not. In fact,

the computational science community uses a variety of authoring tools with perhaps PowerPoint as the most popular and this has guided the development of Tango Interactive. The National Science Foundation, through the NCSA Alliance, has played a significant role in supporting and influencing the design of Tango Interactive.

One important activity was a study of the implications of universal access. The shared event model of collaboration used by Tango Interactive naturally allows the curriculum material to be rendered separately on each client. This could allow one to deliver classes with one set of clients emphasizing graphical display (with perhaps variable resolution reflecting available network bandwidth) and another set sonification of the material for visually impaired users. Tango Interactive from NPAC and Habanero from NCSA are probably the leading collaboration systems in academia built around the so called shared event model. It is worth noting that since 1 January 1999, NPAC has averaged 50 distinct downloads per week (with more than 20% of those downloads being Tango servers) and this software is part of Netscape's list of "approved" plug-ins available from their site.

## **9. Summary**

We have presented what amounts to a current snapshot of the rapidly moving area of network-based collaboration and training taken with the needs of the PET program firmly in mind. We have endeavored to chart a course which offers the PET program the biggest return on investment in these cutting-edge technologies while at the same time doing our best to avoid technological cul de sacs, which are all too common in such a fast-paced field.

A small set of basic and very familiar tools seem adequate to support asynchronous remote collaboration and training (especially e-mail and the web). There are more technically sophisticated tools available, but their value in practice is not clear -- there is strong evidence to suggest that the level and quality of integration with basic tools like e-mail and the web is more important than new

capabilities introduced by the sophisticated tools.

Not surprisingly then, for asynchronous training, it is the educational content that requires the bulk of the effort. The PET program has examined a number of approaches to facilitate the rapid, low-cost development of asynchronous training content by capturing live presentations and publishing them on the web. At the other end of the spectrum, it is also possible to invest quite a bit of effort in developing courseware that can be accessed at varying levels of depth. The PET program has been experimenting with this approach by contracting the opportunity to participate in several of the Cornell Theory Center Virtual Workshops. Internally, several other training classes of greater depth are under development or on the drawing board.

For synchronous training and collaboration, it is more the required tools, and their general unfamiliarity to the user community, which has required us to go slowly with their introduction. Our initial experience has led us to a five point plan for deployment of these tools into the DoD HPC community, working from more to less structured environments. Implementation of the plan is well underway, with basic education and training activities now positioned to become a routine part of PET training. The plan will ultimately lead to both routine collaborative use of Tango and direct to the desktop delivery of trainings.