

Grid computing and collaboration technology in support of fusion energy sciences

D.P. Schissel

General Atomics, P.O. Box 85608, San Diego, California 92186-5608

Contact Author: D.P. Schissel, General Atomics, P.O. Box 85608, San Diego, California 92186-5608, Phone: (858) 455-3387, Fax: (858) 455-3586 email: schissel@fusion.gat.com

Received (

Abstract. Science research in general and magnetic fusion research in particular continue to grow in size and complexity resulting in a concurrent growth in collaborations between experimental sites and laboratories worldwide. The simultaneous increase in wide area network speeds has made it practical to envision distributed working environments that are as productive as traditionally collocated work. In computing power, it has become reasonable to decouple production and consumption resulting in the ability to construct computing grids in a similar manner as the electrical power grid. Grid computing, the secure integration of computer systems over high speed networks to provide on-demand access to data analysis capabilities and related functions, is being deployed as an alternative to traditional resource sharing among institutions. For human interaction, advanced collaborative environments are being researched and deployed to have distributed group work that is as productive as traditional meetings. The DOE SciDAC initiative has sponsored several Collaboratory Projects, including the National Fusion Collaboratory Project, to utilize recent advances in grid computing and advanced collaborative environments to further research in several specific scientific domains. For

fusion, the collaborative technology being deployed is being used in present day research and is also scalable to future research, in particular to the ITER experiment that will require extensive collaboration capability worldwide. This paper briefly reviews the concepts of grid computing and advanced collaborative environments and gives specific examples of how these technologies are being used in fusion research today.

I. INTRODUCTION

Science today is as much based on large teams of scientists as on the efforts of individual experimentalists and theorists. This shift over time to science as a large team enterprise is the result of both increasingly complex problems and the availability of increasingly powerful technology. Whether one considers the Compact Muon Solenoid experiment at the Large Hadron Collider in CERN, the design of ITER, or the Human Genome Project, large teams of geographically distributed scientists are working jointly on experimental and theoretical problems to advance their science. Certainly, the increase in computing power over time (doubling every 18 months on average) has helped to propel science forward. But during this same period, computer networks have increased in speed approximately twice as fast increasing two orders of magnitude in five years. If this trend continues, computer connectivity over wide area networks will be essentially unlimited and the need to solve computational problems locally disappears.

Although not new [1,2], the concept of sharing distributed computing power has been the focus of intense computer science research over the past decade. Termed the “Grid,” it offers the potential for providing secure access to remote services, to allow scientific collaborators to share resources on an unprecedented scale, and for geographically distributed groups to work together in ways not previously possible [3].

This paper reviews the general concepts of Grid computing and Advanced Collaborative Environments (ACE) as they apply to scientific research. With those concepts in hand, the paper gives an overview of the National Fusion Collaboratory Project that is applying, both separately and together, Grid and ACE technologies to advance magnetic fusion science. Combined together, this new infrastructure is being molded to create the collaborative control room for tokamak operations. Although being used in today’s research, these new capabilities need to be significantly enhanced before

their usage is routine. Finally, the paper concludes with a discussion of how Grids and ACE can be applied to future fusion research including ITER.

II. GRID COMPUTING

The decoupling of production and consumption of food, water, and power has played a major role in the modernization of society over the past several centuries. During the time of the founding fathers each home typically had its own wood stove for heat; production and consumption were directly linked. A little over one hundred years ago, Thomas Edison established a central electrical generation station at Pearl Street in lower Manhattan. Although confined to a radius of approximately a half-mile from the station, homes and streets in that area all had the ability to use this centrally generated electricity. The consumption and production of electricity had started to be decoupled. Today, electrical consumption is completely decoupled from production and it has enabled on-demand access, the economics of scale, consumer flexibility, and new devices.

It is this analogy of the electrical power grid that has given rise to the term “grid” for computing [3]; the large-scale integration of computer systems via high-speed networks to provide on-demand access to computational resources (data, codes, visualization) that are not available to an individual at one location. To users, the highly integrated networks that embody grid systems are transparent so that services furnished from afar appear to be provided by local computers. Returning to the analogy of the electrical power grid, when a user plugs a device into an electrical wall socket, the how and where of electrical generation are immaterial. Useful functions (electricity) are hidden by an interface (plug) that conceals the details (power plant) of how they are implemented allowing the individual to concentrate on using the function.

In a traditional computing environment, software users typically install and run programs on a local machine (CPU cycle production and consumption are directly linked). This requires that developers create and maintain versions of their software for the different platforms used by their users. This also requires users to update their local installations as the software is updated. In a grid computing environment applications,

systems, and other computing resources are abstracted into services [4]. This abstraction allows users to invoke services on local or remote hosts without concerning themselves with the details of how such services are implemented. Utilizing remote hosts implies the user has transitioned out of their own administrative domain into that of another organization. To be effective, this type of sharing needs to be governed by a set of rules including what is shared, who is allowed to share, and under what conditions sharing can occur. When physically different administrative domains work together in such a coordinated fashion with a set of clearly defined sharing rules they form what is called a Virtual Organization (VO) [3]. It is the VO that allows geographically separated groups to share computer resources in a controlled fashion to work toward a common goal (Fig. 1).

A. Security

The Internet is an open system, where the identity of the communicating partners is not easy to assure. Furthermore, the communication path traverses an indeterminate set of routing hosts and may include any number of eavesdropping and active interference possibilities. Thus, Internet communication is much like anonymous postcards, which are answered by anonymous recipients. The ability to remove local administrative boundaries to form VOs implies the ability to have a sufficient security infrastructure to ensure that the sharing rules that define the VO can be enforced. Thus a major component for the successful implementation of computer grids is security [5] including authentication, authorization, data encryption and so on. For this paper and understanding of basic grid security, we will cover only the first two topics.

Authentication is the process in a computerized transaction that gives assurance that a person or computer acting on a person's behalf is not an impostor. Authorization is the process of determining, by evaluating applicable access control information, whether a

subject is allowed to have the specified types of access to particular resource. Once a subject is authenticated, it may be authorized to perform different types of access. To use an analogy from everyday life, when boarding a commercial aircraft flight, a drivers license provides authentication that you are who you say you are while the airplane boarding pass determines that you have the right to use the requested resource (airplane).

Unlike the physical interaction that occurs in the above analogy with the driver's license, there is no physical interaction in a computer transaction and therefore identity verification is harder. Two parties can communicate securely by using symmetric cryptography [6]. In this system, two parties (Bob and Patrick) agree on a cryptosystem and also agree on a key. For Bob to send a message to Patrick, he first encrypts the message using the key, sends this ciphertext message to Patrick, who then uses the same key to decrypt the message so it can be read. The security for this system rests in the key, divulging the key means that anyone can encrypt or decrypt messages. The analogy is placing a message in a safe where the key is the combination and anyone with the combination can open the safe. Assuming a separate key is used for each pair of users in a network, the total number of keys increases rapidly as the number of users increases; for n users there are $n(n-1)/2$ keys. For a worldwide encryption system, securing this number of keys can be a daunting task.

Public-key cryptography [7] presents an easier to manage solution for key distribution that scales to large groups. This system uses two different keys, one public and one private, where it is computationally hard to deduce the private key from the public key. Anyone with the public key can encrypt a message but not decrypt it and only the person with the private key can decrypt the message. Mathematically the process is based on the trap-door one-way function [8], which are relatively easy to compute but significantly harder to reverse unless the secret is known. That is, given x it is easy to compute $f(x)$, but given $f(x)$ it is hard to compute x . However, there is some secret information y , such that given $f(x)$ and y it is easy to compute x . Using this system, Bob sends a

message to Patrick encrypting it using Patrick's public-key. Patrick then uses his private-key to decrypt the message so it can be read. If Patrick places his public-key in a publicly available database, Bob can send a secure message with no prior communication between the two (Fig. 2). Therefore, public-key cryptography solves the key-management problem of symmetric cryptosystems.

Public Key Infrastructure (PKI) is a technology to distribute and use asymmetrical keys. PKI gives trust that the public key being used truly belongs to the person or machine with whom/which they wish to communicate [9]. In the previous example, trust needs to be established for Bob to believe that the public-key he uses really does belong to Patrick instead of an impostor. This trust is established through the usage of certificate authorities (CAs) who issue X.509 certificates where a unique identity name and the public key of an entity are bound together through the digital signature of that CA (certificate = trust + public-key). Typically, a registration authority (RA) is responsible for the identification and authentication of certificate subscribers before the CA issues certificates.

Once a user's identity has been validated they are still not given open access to all resources (codes, computers, visualization tools or data). These are made available only to those users who have the proper authorization. Presently there exists no standard for authorization on grids. However it is often broken down into a policy decision point (PDP) and a policy enforcement point (PEP). The PEP is a software component, wrapped around a resource, that either allows or denies access to a resource. The PDP, usually one central authority accessed by all PEP requests, decides what resource access is allowed (Fig. 3).

The implementation for establishing the identity of a consumer of a resource (authentication) and for determining whether an operation is consistent with agreed upon sharing rules (authorization) frames and ultimately defines the virtual organization. We

will now turn to how grids can be used by VOs by examining data management, computing, and visualization resources.

B. Data management

Although data has been stored in different methodologies throughout human history [10], it is the relational model for computer data storage where both entities and relationships are represented in a uniform way [11] that is most commonly used today. The relational database with its Structured Query Language (SQL) is well suited to the client-server model with a graphical user interface. In its simplest form, data management consists of one administrative domain and one physical data location. As data collections have grown, data under one administrative domain has become distributed over numerous physical locations. Today, data management on grids involves data spread over many administrative domains and many physical locations that taken together comprise the virtual organization discussed previously.

Data management on grids has many challenges including diverse usage scenarios, heterogeneity at all system levels, and performance demands associated with access, manipulation, and analysis of large quantities of data. To be useful to the scientific community, users must be able to discover desired data based on metadata attributes. Metadata is the information about data describing, for example, the content, quality, condition, and other characteristics of the data. Once found, data needs to be efficiently moved between storage locations or between programs and storage. To be efficient, data movement includes replication, caching, and bulk data access. Finally, the coupling of computations with operations on data resources introduces new optimization problems but this functionality becomes critical as data repositories grow in size.

An example of efficient data access in a grid environment is the file oriented access supported by GridFTP [12] that provides a uniform interface to various storage systems. GridFTP supports parallel data transfers as well as mechanisms for reliable and

restartable data transfer that is critical for very large data collections. An example of a large-scale science project utilizing GridFTP along with a number of other technologies is the Earth Systems Grid project [13] that is serving multi-terabits of data to climate researchers worldwide.

C. Computing

For computing to occur in a grid environment, mechanisms must be provided for starting programs and monitoring and controlling the execution of the resulting processes. Useful also are management mechanisms that allow for control over allocated resources as well advanced reservation capabilities. Informational discovery is required to obtain information about the structure and state of a computational resource (e.g. load, memory).

The Globus toolkit [14] is an example of a software package that is open source and contains a set of services and software libraries that support grids and grid applications. It has become the foundation for many grid projects worldwide in both academia and industry. An example of such a project was EUROGRID [15] that connected major academic and commercial centers in Europe with an emphasis on high-performance applications using specialized architectures. This project demonstrated distributed simulation codes from different application areas including biomolecular, weather prediction, structural analysis, and real time data processing.

D. Visualization

There are a number of approaches that can be undertaken to support visualization of very large data sets in a grid environment. At one end, parallel rendering [16] can be performed on nodes spread throughout the virtual organization to support graphical discovery at one location. The other end of the spectrum would be to move the data to the scientist and do everything locally. Where in the spectrum of solutions the answer lies is

typically very dependent on the problem being solved and the computational constraints imposed by the existing infrastructure (data set size, interactivity rates, local visualization capability, etc.). Further, a grid based visualization system will most likely need to be dynamic as a user might need to interact with the system, for example to trade visualization accuracy for frame rate.

One example of such a visualization system is ParaView [17], being developed as an open source initiative to develop a multi-platform visualization application to support distributed computational models to process large datasets. The TeraGrid Project [18], is deploying a distributed infrastructure for open scientific research including 20 teraflops of computing power distributed at nine sites, 1 petabyte of data storage, grid computing toolkits, and high-resolution visualization environments. ParaView is being deployed on TeraGrid in an attempt to solve the problem of grid based visualization of very large datasets.

III. ADVANCED COLLABORATIVE ENVIRONMENTS

Researchers often want to aggregate not only data and computing power, but also human expertise. Collaborative problem formulation, data analysis, and the like are important grid applications. The goals of the advanced collaborative environment (ACE) is to use computer mediated communications techniques to enhance work environments, to enable increased productivity for collaborative work, and to exploit the use of high-performance computing technologies to improve the effectiveness of large-scale collaborative work environments. To be effective, collaboration environments should provide lightweight and ubiquitous components that support a wide variety of interaction modes. Such remote work can range from the highly structured (e.g. formal presentations) to more informal, spontaneous collaborations (e.g. cooperative software development).

Many of today's scientific collaborative tools such as videoconferencing tools are highly interactive and only support formal meetings well. Although videoconferencing is an important part of a collaborative environment, much of the work of scientific collaboration requires more informal and asynchronous mechanisms (important for world-wide collaborations). Engaging in informal interactions and sharing documents and data have been shown to be an important part of an effective collaboration [19]. The most fundamental characteristic of a collaborative environment is ubiquity. Collaborators should be able to enter and work within the environment from their desktop machine, their laptop, another user's computer, or any other digital device, or from any location. The collaboration tools need to provide a real benefit to all the users instead of being only one-sided, which is counterproductive [20].

Many tools have been developed to facilitate remote collaboration and a complete review is beyond the scope of this paper. Text-based messaging systems such as America On-line Instant Messaging (AIM), ICQ, Yahoo Messenger, and Jabber are primarily

intended for one-on-one conversation but they have been recently extended to include group-chat as well as file transfer. Several systems such as the Virtual Room Videoconferencing System (VRVS) [21] and WebEx offer web-oriented, low-cost, bandwidth—efficient, extensible means of videoconferencing and remote collaboration over IP networks. VRVS is used extensively in the High Energy Physics field and creates a virtual meeting room where participants “gather” as if they were together in the same physical room. VRVS transmits all active video and audio channels to all participants via a network of “reflectors.” Integrated into VRVS is a “chat” capability that allows for back-channel communication during the virtual meeting.

The Mbone videoconferencing tools (vic, vat, rat, wb, and sdr) provide multicast, multi-way videoconferencing over IP networks that allows all users to be seen and heard as equal participants. The Access Grid software [22] extends the Mbone work to include the ability to utilize for scientific research a complex multi-site visual and collaborative experience integrated with high-end visualization environments (Fig. 4). AG nodes range in size from the individual desktop similar to VRVS to a very large meeting room. Integrated into the AG system is a modified VNC (Virtual Network Computing) that allows for more efficient interactive sharing of complex visualizations. VNC [23] allows Internet sharing of a computer desktop with one or more remote clients. Within the AG environment, VNC is being used to interactively share applications during working meeting and to broadcast electronic slide presentation. Since its inception, VRVS has been extended to also interact with Mbone and Access Grid systems as well as the integration of VNC to create a more interactive virtual meeting.

Tiled display walls [24] are being investigated to enhance the collaborative work environment of large groups of collocated individuals. A tiled display wall utilizes multiple projectors tiled together to build a bright, high-resolution, seamless display with 16 ft x 8 ft, 20 million pixel displays not uncommon (Fig. 5). With the increased speeds of computers, networks, and graphics cards, the ability to deploy tiled displays at low

cost with commodity components is being realized. Such a display offers a large-format environment for presenting high-resolution visualizations or multi-source smaller visualizations to a collaborative group than would be possible on standard displays. For collocated individuals, this interactive shared visualization takes the place of “passing around” a graphical printout or “calling over” scientists to collaboratively view a normal desktop display. Given the ability to share visualizations from remote locations to tiled displays, off-site scientists can interactively share visualizations and participate in large-group discussions, something previously not possible. The computer science research being conducted within the area of tiled displays includes parallel rendering, user interfaces, image blending, computational alignment, and color balancing [25].

IV. THE NATIONAL FUSION COLLABORATORY PROJECT

Historically, efforts to improve collaboration within the U.S. fusion community have included sharing of resources and co-development of tools mostly carried out on an *ad hoc* basis. The community has considerable experience in placing remote collaboration tools into the hands of real users [26]. The ability to remotely view operations and to control selected instrumentation and analysis tasks was demonstrated as early as 1992 [27]. Full remote operation of an entire tokamak experiment was tested in 1996 [28,29].

The National Fusion Collaboratory Project [30] is funded by the United States Department of Energy (DOE) under the Scientific Discovery through Advanced Computing Program (SciDAC) to develop a persistent infrastructure to enable scientific collaboration for all aspects of magnetic fusion research. Initiated in late 2001, this project builds on the past collaborative work performed within the U.S. fusion community and adds the component of computer science research done within the USDOE Office of Science, Office of Advanced Scientific Computer Research. The project is a collaboration itself uniting fusion scientists and computer scientists from seven institutions to form a coordinated team. This group is leveraging existing computer science technology where possible and extending and/or creating new capabilities where required.

The vision for FusionGrid is that experimental and simulation data, computer codes, analysis routines, visualization tools, and remote collaboration tools are to be thought of as network services which represents a fundamental paradigm shift for the fusion community. In this model, an application service provider (ASP) provides and maintains software resources as well as the necessary hardware resources. The project is creating a robust, user-friendly collaborative software environment and making it available to the more than one thousand fusion scientists in forty institutions who perform magnetic fusion research in the United States. In particular, the project is developing and deploying

a national Fusion Energy Sciences Grid (FusionGrid) that is a system for secure sharing of computation, visualization, and data resources over the Internet. The FusionGrid goal is to allow scientists at remote sites to fully participate in experimental and computational activities as if they were working at a common site thereby creating a virtual organization of the U.S. Fusion community. This Grid's resources are protected by a shared security infrastructure including strong authentication to identify users and fine-grain authorization to allow stakeholders to control their own resources. FusionGrid will shield the users from software implementation details and allow a sharper focus on the physics with transparency and ease-of-use being the crucial elements. In this environment, access to services is stressed rather than data or software portability. FusionGrid is not focused on computer cycle scavenging (e.g. SETI@home) or distributed supercomputing that are typical justifications for Grid computing, but simply on making the ASP paradigm effective for Grids.

Accomplishing the Project's goals will advance scientific understanding and innovation in magnetic fusion research by enabling more efficient use of existing experimental facilities and more effective integration of experiment, theory, and modeling. Physics productivity will be increased by (1) allowing more transparent and uniform access to analysis and simulation codes, to data, and to visualization tools resulting in more researchers having access to more resources, (2) creating a standard tool set for remote data access, security, and visualization allowing more researchers to build these into their own services, (3) enabling more efficient utilization of experimental time through more powerful between pulse data analysis and through enhanced human participation (both remote and collocated) resulting in faster experimental progress at less cost, (4) facilitating the comparison of theory and experiment through transparent remote data access with appropriate security, and (5) facilitating multi-institution collaborations through the creation of the standard toolset. The Project will also increase the productivity of code and tool developers by (1) supporting more users with fewer

installations at reduced cost (e.g. see Section IV.B TRANSP), (2) facilitating shared code development projects resulting in more rapid code creation through enhanced interaction with remote staff, code sources, data and visualizations (e.g. shared code debugging), and (3) creating a standard tool set for remote data access, security, and visualization allowing these services to be easily built into new code systems.

A. FusionGrid security

FusionGrid security employs Public Key Infrastructure (PKI) to secure communication on the Internet through the use of a public and private cryptographic key pair that is obtained and shared through a trusted authority as was discussed previously. FusionGrid uses the X.509 certificate standard and the FusionGrid CA to implement PKI for secure communication. A scientist who desires to join FusionGrid will generate a public/private key pair and apply to the FusionGrid CA for an X.509 certificate as discussed in Section II.A. That request goes to the RA who will verify their identity and the validity of their request (i.e. determine they are known members of the community and they have a reason to join FusionGrid).

FusionGrid certificates are managed on behalf of the user by a myProxy online certificate repository [31] securely installed at LBNL. In this system, the user's long-term certificate (private key+trust) is securely stored on the myProxy server and access to FusionGrid is accomplished by having the user submit their username and password to the myProxy server. The submittal process is typically done behind the scene on behalf of the user; they only have to type their username and password. With this FusionGrid login, the myProxy server issues a short-term certificate that is then used for authentication. By storing the users' long-term certificates on the myProxy server, users no longer have to manage their certificate, but instead delegate that task to a FusionGrid administrator.

The secure authenticated connections are accomplished using the Globus Toolkit [14]. All a user needs to do is submit their username and password to the myProxy server once per day. This single sign-on is accomplished behind the scenes by the use of a short-lived proxy certificate that is derived from the user's long-term X.509 certificate. The proxy certificate uses its own unencrypted private key, so that it can make frequent authenticated connections on behalf of the user to multiple services without requiring additional password interactions with the user. The benefit to the user is that they need only log-on once no matter how many different services they desire to use.

Centralized authorization of FusionGrid resources is accomplished through the Resource Oriented Authorization Management System (ROAM). This system allows a resource provider to implement either a simple or complex authorization policy using a web browser interface. System flexibility is maintained since the resource provider is allowed to either use existing permission levels or define their own as required. The system is implemented using an Apache-based web server and a single PHP (a recursive acronym for PHP: Hypertext Preprocessor) script. PHP is an open-source, server-side HTML embedded scripting language used to create dynamic Web pages (e.g. search results from a database). A dynamic Web page is a page that interacts with the user, so that each user visiting the page sees customized information. A PostgreSQL database is used to manage all the authorization information. Access to this information is done via secure HyperText Transport Protocol (HTTPS) using either the user's certificate, if present, or by a myProxy login. Resources also check for authorization using HTTPS communication.

B. FusionGrid data and computing

Data access on FusionGrid has been made available using the MDSPlus data acquisition and data management system [32] combined with the relational database

Microsoft SQL server. MDSplus, developed jointly by MIT, LANL, and the IGI in Padua, Italy, is by far the most widely used data system in the international fusion program. Based on a client/server model, MDSplus provides a hierarchical, self-descriptive structure for simple and complex data types [33,34] and is currently installed and used in a variety of ways by about 30 experiments, spread over 4 continents. It is deployed as a complete data acquisition and analysis systems for C-Mod (MIT); RFX (IGI, Padua); TCV (EPFL, Switzerland); NSTX (PPPL); Helic (ANU, Australia); MST (U. Wisconsin); HIT (U. Washington); CHS (NIFS, Japan); and LDX (MIT). It is used to store processed data for DIII-D, for the collaborative data archives assembled by the ITPA, and for the inputs and outputs of several widely used codes including EFIT, TRANSP, NIMROD and GS2. JET and ASDEX-Upgrade are using MDSplus as a remote interface to existing data stores and KSTAR has adopted it as a data acquisition engine for data stored in other formats. The result is a *de facto* standard that greatly facilitates data sharing and collaborations across institutions.

MDSplus and the Globus Toolkit have been combined to create secure X.509 certificate based client/server data access on FusionGrid using the standard MDSplus interface without any loss in speed or functionality. SQL Server is securely accessible via MDSplus since a production of release of Globus for Windows is not available. Presently, the three main MDSplus experimental data repositories at Alcator C-Mod, DIII-D, and NSTX are securely available on FusionGrid. Data management by MDSplus of large datasets generated by simulation codes is presently being tested using results from NIMROD simulations. NIMROD is a 3D MHD simulation code that runs on very large parallel computers. Using the MDSplus server at DIII-D, output from NIMROD runs up to 100 GB have been stored and served to users for further data analysis and visualization. Although successful, this storage methodology proved to be inefficient. The installation of an MDSplus server on the NERSC LAN along with the high-performance computational servers has been undertaken to investigate increased

throughput capability. Parallel network data transport are also being investigated in order to overcome TCP/IP flow control limits for high bandwidth, high latency connections.

The code TRANSP, used for time dependent analysis and simulation of tokamak plasmas, was released as a service on FusionGrid late in 2002 [35] along with supporting infrastructure development (data storage, monitoring, user GUI) [36]. This FusionGrid service has been so successful that it has become the production system for TRANSP usage in the United States and is starting to be adopted internationally. Running on a Linux cluster at PPPL, over 4600 TRANSP runs from ten different experimental machines have been completed within the FusionGrid infrastructure (Fig. 6). European scientists use TRANSP on FusionGrid with approximately 40% of the runs performing analysis on data from European machines. This approach has drastically reduced the efforts to support and maintain the code which were previously required of the developers and by users' sites.

When users request TRANSP Grid services, their proxy certificate is used to verify their identity through the Globus GSI. Once authenticated, users are authorized to run TRANSP via ROAM. To use the TRANSP FusionGrid service, the inputs and outputs are stored in MDSplus trees. Sites without their own MDSplus server can also receive TRANSP output in the traditional NetCDF file format via GridFTP; a PPPL-provided script makes this GridFTP task very simple. The IDL-based PreTRANSP application is one technique to simplify TRANSP usage by assisting the user in preparing TRANSP inputs, managing code runs, and launching TRANSP. For now at least, PreTRANSP is only used for the preparation of TRANSP runs.

Recently the GATO ideal MHD stability code was released as a FusionGrid computational service running on a Linux computer at General Atomics (GA). Following the same design as the TRANSP service, the time required to deploy GATO on FusionGrid was minimal. This result has given confidence that the design of FusionGrid will scale to the deployment of many services.

It is important to note that users executing a code run on either the PPPL or GA Linux systems need not deal with local computer accounts. For example, when a collaborator runs TRANSP on FusionGrid, connections made by their proxy are mapped to PPPL-assigned “run production” accounts created specially for the TRANSP service. These run production accounts are implemented as local UNIX accounts on the PPPL cluster, and are used to ensure data privacy. Users never need to learn a new set of passwords or host names as this account mapping happens behind the scenes. This greatly simplifies the task of account administration.

With multiple applications distributed throughout a Grid infrastructure, it becomes a challenge to monitor the progress and state of each application. Users of a Grid environment need to know the specific state of code runs, when their data results are available, or if the requested application is even available. To track and monitor applications on the FusionGrid, the FusionGrid Monitor (FGM) has been developed as a Java Servlet which can accept and monitor information from remote and distributed applications [36]. Currently, FGM tracks TRANSP and GATO analysis runs on the National Fusion Grid, and provides updated information on each individual run, including: current state, cpu time, wall time, comments, and access to log files that have been produced by the analysis. The Fusion Grid Monitoring system has built to provide user output through HTML, utilizing both server push and client pull capabilities. This allows multiple users to connect to FGM, view their code runs by using a web browser, and obtain updated information without excessive user input or client software. Designed in the Java language, the monitoring system is portable, and with the inclusion of the Java Expert System Shell (JESS), the system is also expandable and customizable. Online access to log files is available through FGM, utilizing anonymous FTP.

FGM has been recently extended to include an Internet-accessible Java-based graphical monitoring tool, EIVis, to display results from remote simulations as they are computed. The EIVis monitoring not only shows that the remote computational service is

operating; it also allows select results to be made available in the control room or at collaborator sites even before the run is completed.

B FusionGrid advanced collaborative environment

The goals of FusionGrid's advanced collaborative environment service is to use computer mediated communications techniques to enhance work environments, to enable increased productivity for collaborative work, and to exploit the use of high-performance computing technologies to improve the effectiveness of large-scale collaborative work environments. Examples of such collaboration include off-site support of experimental operations, large group collaborations in a tokamak control room, simulation/experimental data analysis meetings, and shared code debugging.

Tiled display walls are being used to enhance the collaborative work environment of the tokamak control room. Such a display offers a large-format environment for presenting high-resolution visualizations or multi-source smaller visualizations to a collaborative group than would be possible on standard displays.

As a prototype FusionGrid service, tiled display walls have been tested in a variety of usage modalities ranging from two tiled walls geographically separated being tied together by software to form shared collaborative displays to a single tiled wall used for collocated group sharing and discussion. Based on the success of these tests, a 2-tile front projection system has been installed in the NSTX control room and a 3-tile rear projection system in the DIII-D control rooms (Fig. 7). Based on VNC, this service allows any researcher either in the control room or off-site, with proper authentication and authorization, to share any X-windows based visualization, with the entire control room. For scientists within the control room, this interactive shared visualization takes the place of "passing around" a graphical printout or "calling over" scientists to collaboratively view a normal desktop display. For scientists off-site, this service gives

them the capability to interactively share visualizations and participate in experiments, something previously not possible. The software has been designed so that the remote scientist need not purchase any special hardware and they can therefore share pieces of the larger control room display wall on their single desktop display.

The Access Grid is used by FusionGrid to create a service that enables group-to-group interaction and collaboration that improves the user experience significantly beyond teleconferencing. The Access Grid includes the ability to utilize for scientific research a complex multi-site visual and collaborative experience integrated with high-end visualization environments. Developed exclusively for a FusionGrid service, the personal interface to the Access Grid (PIG) has been developed as a low cost alternative to a full-blown conference room size Access Grid node (Fig. 8).

V. THE COLLABORATIVE CONTROL ROOM

The combination of Grid computing with collaboration technologies such as the Access Grid (AG) with application sharing has the potential to dramatically improve the efficiency of experimental sciences. The combination of these technologies into a unified scientific research environment called the collaborative control room poses unique challenges but creates the possibilities of high reward in the form of increased efficiency of experiments.

Magnetic fusion experiments operate in a pulsed mode. In any given day, 25-35 plasma pulses are taken with approximately 10 to 20 minutes between each ~10-second pulse. For every plasma pulse, up to 10,000 separate measurements versus time are acquired at sample rates from kHz to MHz, representing about a gigabyte of data. Throughout the experimental session, hardware/software plasma control adjustments are made as required by the experimental science. These adjustments are debated and discussed among the experimental team. Decisions for changes to the next pulse are informed by data analysis conducted within the roughly 20-minute between-pulse interval.

Data analysis to support experimental operations includes between pulse analysis of raw acquired data as well as the merging of numerous data sources for whole-device simulation of the experimental plasma. Results of more detailed, computationally demanding predictive simulations, carried out during the planning phase prior to the experiment, are made available for comparison to the actual experimental results in real time.

This mode of operation places a large premium on rapid data analysis that can be assimilated in near-real time. The experimental science can be made more efficient by pushing the boundaries in two directions. First, by running codes on geographically dispersed resources the amount and detail of both analysis and simulation results can be

increased. Second, by bringing in expertise from geographically remote teams of experts, the depth of interpretation can be increased leading to improved assimilation of those results. In order to be fully functional, the collaborative control room requires (1) secured computational services that can be scheduled as required, (2) the ability to rapidly compare experimental data with simulation results, (3) a means to easily share individual results with the group by moving application windows to a shared display, and (4) the ability for remote scientists to be fully engaged in experimental operations through shared audio, video, and applications.

A. Prototype implementation

A prototype implementation of the collaborative control room was developed and a simulation was demonstrated at the SC2003 meeting. The demonstration involved remote codes, resources, and scientific teams in the experiment. Offsite collaborators (SC2003 show floor, Phoenix) joined in a mockup of a DIII-D experiment located in San Diego. AG technology allowed for shared audio and video as well as shared applications. The Access Grid was used to give the remote scientist the feeling of being part of the control room at a distance by allowing the remote scientist to talk to and see the DIII-D control room, enabling the remote scientist to ask questions of the operators there as well as see what was going on in the control room as it was happening. This could never be achieved with just a telephone call. As the number of offsite collaborators grows, this can be achieved only in a limited fashion with present day videoconferencing technology. Additionally, since AG technology is open source, it can be easily expanded to add services and tools specific to a tokamak control room something that would be much more difficult, if not impossible, with commercial videoconferencing equipment. The offsite collaborators could hear DIII-D announcements from both the scientist and engineer in charge, as well as see via a Web interface the state of the pulse cycle and the

status of data acquisition and between pulse data analysis, and how much time was left before the next pulse.

As the data was gathered into MDSPlus, the remote scientist was able to open standard data processing and viewing applications such as ReviewPlus and EFITViewer to start the analyzing process. Once the remote scientist identified data points of interest, they were able to move the application to a region that was shared between the Access Grid node and the control room. This area could be seen and interacted with by both parties. This kind of interaction is a significant step forward from the present situation. A modified VNC was used to handle the remote desktop sharing.

Between-pulse data analysis of the plasma shape (EFIT running at PPPL) was conducted on FusionGrid through a computational reservation system that guaranteed a specific analysis to be completed within a set time window (Sec. V.B). Additionally, the TRANSP service was run at PPPL for the first time between pulses, giving the scientists data that was previously available only after the experimental day had ended. The offsite team members were able to collaborate more efficiently by being able to share their personal display with the room's shared display. This capability allowed visualizations to be efficiently compared for debate before reporting results back to the DIII-D control room. The results of this demonstration and the feedback from fusion scientists has helped sharpen the requirements for a truly collaborative control room for fusion experiments.

B. Computational reservations for between-pulse analysis

For the SC03 demonstration, agreement-based interactions utilizing the Globus Toolkit 3 (GT3) enabled fusion scientists to negotiate end-to-end guarantees on execution of remote codes between the experimental pulses [37]. This mode of interaction has high potential for resolving problems of provisioning in Grid computing with specifications

being drafted by the Grid Resource Allocation Agreement Protocol (GRAAP) working group of the GGF.

In the demonstration, the FusionGrid TRANSP service was run on the PPPL cluster to support DIII-D operations and the results were simultaneously visualized in Phoenix and San Diego. In preparation for the demonstration, significant work was done to reduce TRANSP run production time, through both software and hardware changes, to about 6 minutes, which was found to be acceptable for an experimental run. The actual TRANSP run execution time was slightly over 3 minutes; the balance of the time was due to network data transfers. These data transfer delays will be reduced through further optimization of the software.

In this first between pulse data analysis using TRANSP, only one timeslice of the experimental data was run. In principle it would be best to run a fully time-dependent TRANSP simulation. At some point in the future if the TRANSP code is parallelized, this reservation system will scale to large runs on multi-node computational clusters.

C. Usage during experimental operations

With the success of the prototype demonstration at the SC2003 meeting, the first usage of the collaborative control room tools during actual experimental operations was in early 2004. In this case, Dr. J.S. deGrassie was the Scientific Coordinator for an experiment on the EFDA-JET fusion experiment in England from his home laboratory in San Diego (Fig. 9). Access Grid technology was used for multiple video images and a unified audio stream between San Diego and the JET control room. Web based displays gave Dr. deGrassie “real-time” tokamak information as well as pulse cycle and general data analysis status. Although considered a success for a first time usage in actual experimental operations, the experience nevertheless indicated areas of where improvements are needed. The biggest area needing improvement was the ability to more efficiently share visualizations of data analysis. This would allow each site to “look over

each others shoulder” and discuss their analysis in real time. Instead, a site would verbally explain their analysis, wait for the other site to catch up by creating their own visualization, and then have the data discussion. In the time demanding environment of a tokamak control room this type of interaction is too inefficient.

Subsequent to the remote collaboration with JET, similar collaborations have been undertaken with Germany (IPP) and Japan (JAERI). At times, these have been as successful as the original JET experience and at other times scientists have reverted to the telephone when the quality of the IP based audio/video service is too varied. These experiences all reinforce how low the tolerance is for error in a tokamak control room. If a remote meeting experiences audio or video difficulty, delaying or rescheduling is inconvenient but an option. In the control room, every ~20 minutes wasted means a lost opportunity for new data, and with experimental time so precious, it is unacceptable. In recognition of this situation, the NFC project has created the basic tiled wall and computing infrastructure of a control room in a separate laboratory. This setup will allow software to be further tested in a mock control room situation as well as allow for training of scientific staff before used in actual experimental operation.

VI. CONCLUSIONS AND FUTURE PROSPECTS

There is a significant worldwide effort to develop and deploy Grid computing and advanced collaborative environments in support of numerous scientific disciplines including fusion energy research. For fusion energy sciences, although substantial progress has been made, more work is clearly required to reach the point where off-site participation is as rewarding as on-site participation. With the worldwide focus on ITER as the next generation machine, its success requires advanced remote collaboration capability. This capability and success needs to include more than just the experimental physics program. The final design, engineering, and construction phases will be worldwide collaborations as well and although they will not need the collaborative control room they will need the ability to richly interact with their distant colleagues. The ability to interactively share engineering drawings, conduct design reviews, and view 3-D mockups of machine components are all clearly required. Imagine an electronic tabletop display (Hitachi recently announced such a product), one can envision a drawing “rolled-out” electronically on a table. The designer picks up her pen and begins “pointing” to different areas of the drawing as the ad-hoc meeting begins during ITER construction to solve an unforeseen problem. Her counterparts spread around the world have a similar table, see the drawing and her “pointing,” and when they look up into the accompanying display device they see images of their colleagues and hear their words.

When ITER operation begins, the collaborative control room will allow scientists to share data and knowledge as readily as their engineering counterparts. Large tiled displays will present information to the assembled team in the control room. Off-site scientists will be sharing the results of their analysis to the large display as well as to individual small displays of their colleagues. ITER’s integrated data acquisition and data management system not only allows for simultaneous data availability worldwide, but also the automatic starting of many data analysis tasks. Utilizing computational Grids,

these tasks are dispatched to computer systems located worldwide and their results are rapidly integrated back into the ITER data management system. Some of these tasks are run on very large state-of-the-art supercomputers requiring network quality of service and CPU scheduling so as to support the time critical data analysis environment of the control room. Utilized in this way, the worlds most power computer systems run data analysis and simulations in concert with the largest fusion device to advance fusion science in ways not previously envisioned.

Due to the complexity of these problems and the importance of reaching a satisfactory solution, design work and testing needs to start early in the ITER construction phase. Present day fusion energy science research provides an excellent proving ground for research to support the needs of ITER. Although ITER's first plasma is over a decade away, the intervening time should be used to continue to develop the technology outlined in this paper.

Acknowledgment

This work was supported in part by the U.S. Department of Energy Office of Advanced Scientific Computing under No. DE-FC02-01ER25455, the Office of Fusion Energy Sciences under DE-FC02-04ER54698, and the SciDAC Program. The author acknowledges valuable contributions from the entire National Fusion Collaboratory Project Team and from the staff at the DIII-D National Fusion Facility.

References

- [1] V.A. Vyssotsky, F.J. Corbato, R.M. Graham, Fall Joint Computer Conference, AFIPS Conf. Proc. **27**, 203 (1965). <http://www.multicians.org/fjcc3.html>.
- [2] L. Kleinrock, UCLA Press Release, July 3, 1969.
- [3] I. Foster, C. Kesselman, Eds, *The Grid2: Blueprint for a New Computing Infrastructure*, Morgan Kaufman, San Francisco, California (2004).
- [4] I. Foster, C. Kesselman, J. Nick, S. Tuecke, *The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration*, Open Grid Service Infrastructure WG, Global Grid Forum, June 22, 2002.
- [5] F. Berman, G. Fox, A. Hey, Eds, *Grid Computing: Making the Global Infrastructure a Reality*, John Wiley and Sons (2003).
- [6] B. Schneier, *Applied Cryptography*, John Wiley and Sons (1996).
- [7] W. Diffie, M.E. Hellman, "New Directions in Cryptography," IEEE Transactions on Information Theory, Vol. IT-22 (1976) p. 644.
- [8] J. Grollman, A.L. Selman, "Complexity Measures for Public-Key Cryptosystems," Proceedings of the 25th IEEE Symposium on the Foundation of Computer Science (1984) p. 495.
- [9] ITU-T X.509 (03/00): Information Technology – Open systems interconnection, the directory: Authentication framework. ITU, Recommendation, 2002.
- [10] J. Gray, "Data Management: Past, Present, and Future," IEEE Computer **29**, 38 (1996).
- [11] *An Introduction to Database Systems*, 8th Edition, C.J. Date, Addison Wesley, (2003).
- [12] W. Allcock, J. Bester, J. Bresnahan, *et al.*, "Data Management and Transfer in High-Performance Computational Grid Environments," Parallel Computing **28**, 749 (2002).

- [13] I. Foster, E. Alpert, A. Chervenak, *et al.*, “The Earth Systems Grid II: Turning Climate Datasets into Community Resources,” Annual Meeting of the American Meteorological Society (2002).
- [14] I. Foster, C. Kesselman, “Globus: A Metacomputing Infrastructure Toolkit,” *International Journal of Supercomputing Applications* **11**, 115 (1997).
- [15] J. Brooke, M. Foster, S. Pickles, K. Taylor, T. Hewitt, “Mini-Grids: Effective Test-Beds for GRID Applications,” *Proceedings of IEEE Workshop Grid2000*, R. Buyya, M. Baker (Eds.), Springer LNCS 1971, 158-169 Bangalore, December 2000. <http://www.eurogrid.org/>
- [16] T.W. Crockett, “An Introduction to Parallel Rendering,” *Parallel Computing* **23**, 819 (1997).
- [17] <http://www.paraview.org/>
- [18] <http://www.teragrid.org/>
- [19]. D.A. Agarwal, S.R. Sachs, W.E. Johnston, “The Reality of Collaboratories,” *Computer Physics Communications* **10** (1998).
- [20]. J. Grudin, “Groupware and Social Dynamics: Eight Challenges for Developers,” *Communications of the ACM*, Vol 34 (1994).
- [21] D. Adarczyk, D. Collados, G. Deris, *et al.*, “Global Platform for Rich Media Conferencing and Collaboration,” to be published in the *Proceedings of the Conference for Computing in High Energy and Nuclear Physics*, La Jolla, California (2003).
- [22]. L. Childers, T. Disz, R. Olson, M.E. Papka, R. Stevens, and T. Udeshi, “Access Grid: Immersive Group-to-Group Collaborative Visualization,” *Proceedings of*

- the 4th Int. Immersive Projection Technology Workshop, Ames, Iowa (2000).
<http://www.accessgrid.org/>.
- [23] Tristan Richardson, Quentin Stafford-Fraser, Kenneth R. Wood, and Andy Hopper, "Virtual Network Computing," IEEE Internet Computing, Vol. 2, 33 (1998). <http://www.realvnc.com/>.
- [24] K. Li, H. Chen, Y. Chen, *et al.*, "Early Experiences and Challenges in Building and using a Scalable Display Wall System," IEEE Computer Graphics and Applications, Vol 20, 671 (2000).
- [25] G. Humphreys, P. Hanrahan, "A Distributed Graphics System for Large Tiled Displays," Proceedings of the IEEE Conference on Visualization 1999, 215 (1999).
- [26] T. Fredian, J. Stillerman, "MDSplus Remote Collaboration Support—Internet and World Wide Web," Fusion Engineering and Design **43**, 327 (1999).
- [27] R. Fonck, *et al.*, "Remote Operation of the TFTR BES Experiment From an Off-Site Location," Rev. Sci. Instrum. **63**, 4803 (1992).
- [28] S. Horne, M. Greenwald, T. Fredian, I. Hutchinson, B. LaBombard, J. Stillerman, Y. Takase, S. Wolfe, T. Casper, D. Butner, W. Meyer, and J. Moller, "Remote Control of Alcator C—Mod from LLNL," Fusion Technology **32**, 52 (1997).
- [29] B.B. McHarg, T.A. Casper, S. Davis, D. Greenwood, "Tools for Remote Collaboration on the DIII—D National Fusion Facility," Fusion Eng. Design **43**, 343 (1999).
- [30] D.P. Schissel, J.R. Burruss, S.M. Flanagan, *et al.*, "Building the US National Fusion Grid: Results from the National Fusion Collaboratory Project," Fusion Eng. Design **71** 245 (2004). <http://www.fusiongrid.org/>
- [31] J. Novotny, S. Tuecke, V. Welch, "An Online Credential Repository for the Grid: MyProxy," Proceedings of the 10th IEEE International Symposium on High

- Performance Distributed Computing, San Francisco, California, IEEE Computer Society Press, Los Alamitos, California (2001).
- [32] T.W. Fredian, J. Stillerman, "MDSplus, Current Developments and Future Directions," *Fusion Eng. Design* **60**, 229 (2002). <http://www.mdsplus.org/>
- [33] J.A. Stillerman, T.W. Fredian, K.A. Klare, G. Manduchi, *Rev. Sci. Instrum.* **68**, 939 (1997).
- [34] J. Stillerman, T.W. Fredian, "The MDSplus Data Acquisition System, Current Status and Future Directions," *Fusion Eng. Design* **43**, 301 (1999).
- [35] J.R. Burruss, "Remote Computing Using the National Fusion Grid," *Fusion Eng. Design* **71** (2004).
- [36] S. Flanagan, J.R. Burruss, C. Ludescher, *et al.*, "A General Purpose Data Analysis Monitoring System with Case Studies from the National Fusion Grid and the DIII-D MDSplus Between Shot Analysis System," *Fusion Eng. Design* **71**, 263 (2004).
- [37] K. Keahey, T. Fredian, *et al.*, "Computational Grids in Action: the National Fusion Collaboratory." *Future Generation Computing Systems*, Vol. 18 (2002) p. 1005.

Figure Captions

Fig. 1. In grid computing, Virtual Organizations are formed when different administrative domains agree to share computer resources to work towards a common goal. In a grid computing environment applications, systems, and other computing resources are abstracted into services freeing the researcher from the need to know the implementation details and allowing them to instead concentrate on the scientific problem at hand.

Fig. 2. Public-key cryptography uses two different keys, one public and one private, where it is computationally hard to deduce the private key from the public key. Public-keys can be kept in a publicly available database and individuals hold their private key. This solves the key management problem of symmetric cryptosystems and therefore presents the capability of scaling to organizations with many members.

Fig. 3. Authorization is the process of determining, by evaluating applicable access control information, whether a subject is allowed to have specific types of access to a particular resource. The policy enforcement point is usually wrapped around each resource and asks the central policy decision point whether access should be granted.

Fig. 4. The Access Grid is an IP based collaboration environment that includes audio, video, and shared applications. Ranging from large room installations to individual desktops, the Access Grid is being used to benefit scientific research through a complex multi-site visual and collaboration experience integrated with high-end visualization environments.

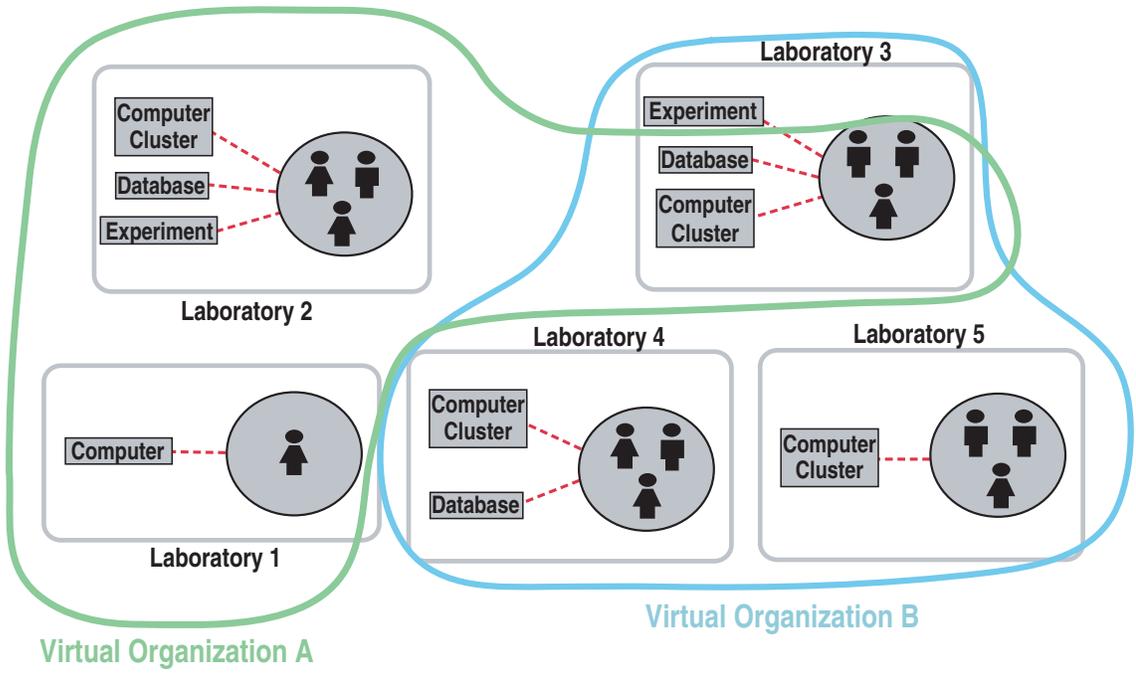
Fig. 5. Tiled displays present the illusion of one large unified display and can be used to enhance collaboration between large groups of collocated individuals.

Fig. 6. TRANSP is installed on a Linux cluster at PPPL and is made securely available to authorized members of the FusionGrid virtual organization. It can be run from anywhere in the world and utilizes secure MDSplus for both input and output data.

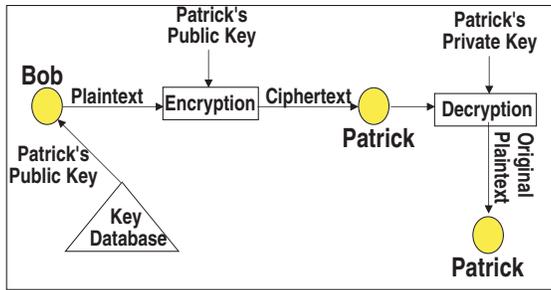
Fig. 7. Tiled display walls have been installed in both the DIII-D and NSTX control rooms to facilitate collaboration and shared data analysis amongst the collocated scientists in the control room as well as between off-site scientists and the control room staff.

Fig. 8. Access Grid nodes are being used by FusionGrid to create a service that enables group-to-group interaction and collaboration and improves the user experience significantly beyond teleconferencing.

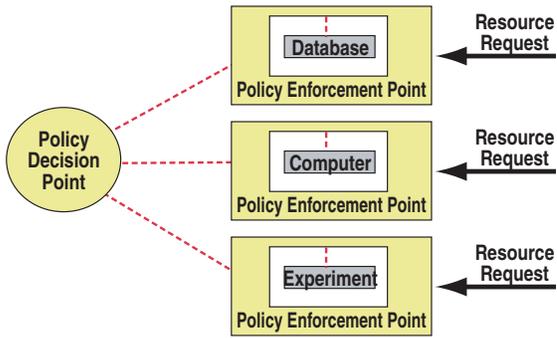
Fig. 9. (a) A scientist from his home institution in San Diego utilized FusionGrid services (AG, VRVS, and MDSplus) to be the Scientific Coordinator for the EFDA-JET experiment in England. Audio, video, shared data, and shared applications create the beginnings of the collaborative control room. (b) Enlargement of the video from the JET control room. (c) Enlargement of the data traces from JET.



D.P. Schissel Figure 1



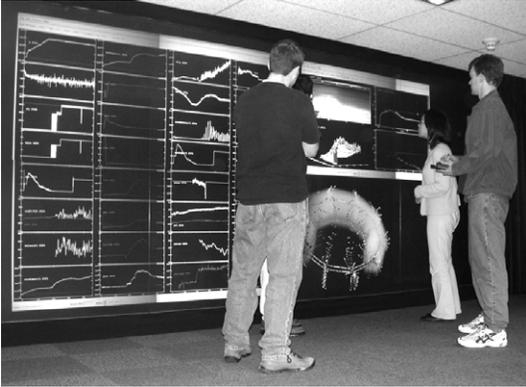
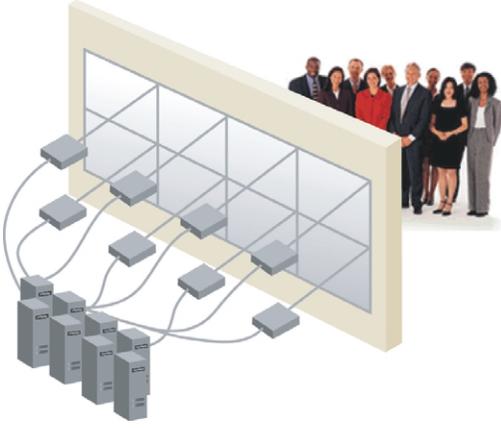
D.P. Schissel Figure 2



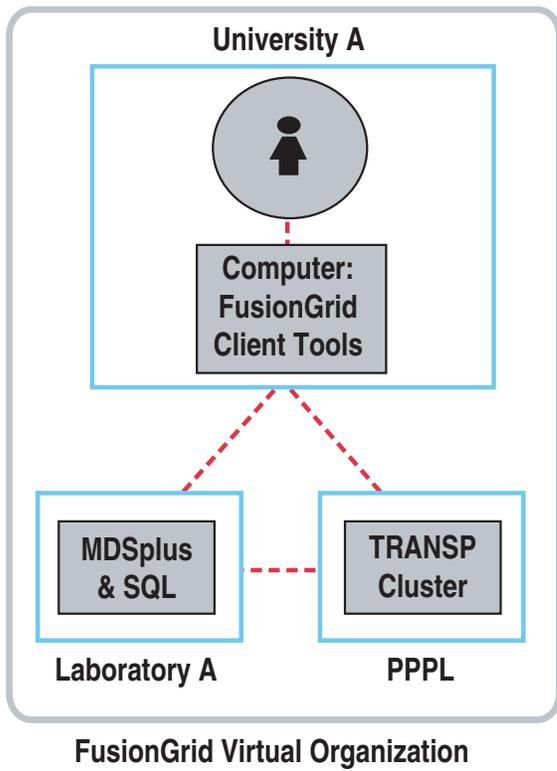
D.P. Schissel Figure 3



D.P. Schissel Figure 4



D.P. Schissel Figure 5



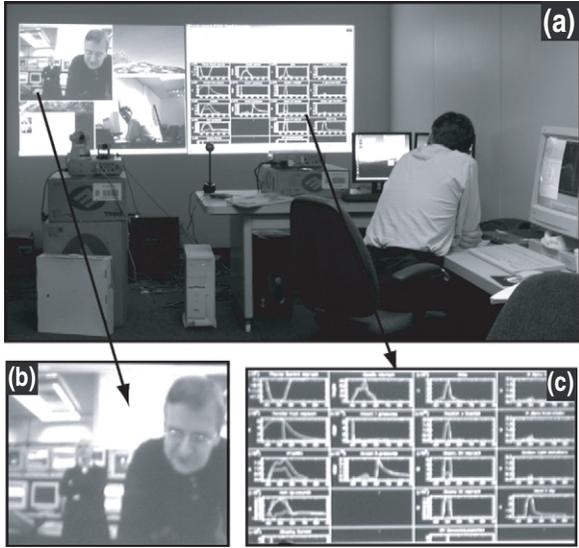
D.P. Schissel Figure 6



D.P. Schissel Figure 7



D.P. Schissel Figure 8



D.P. Schissel Figure 9