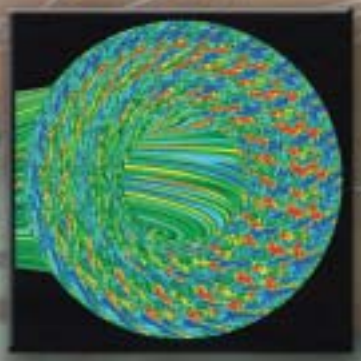
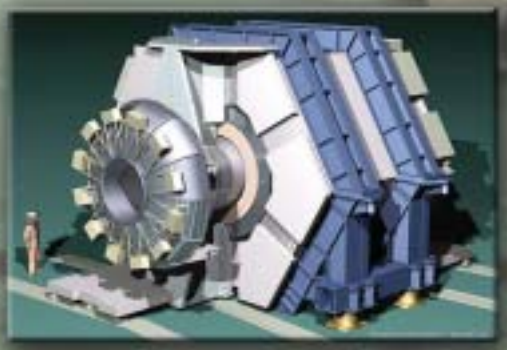
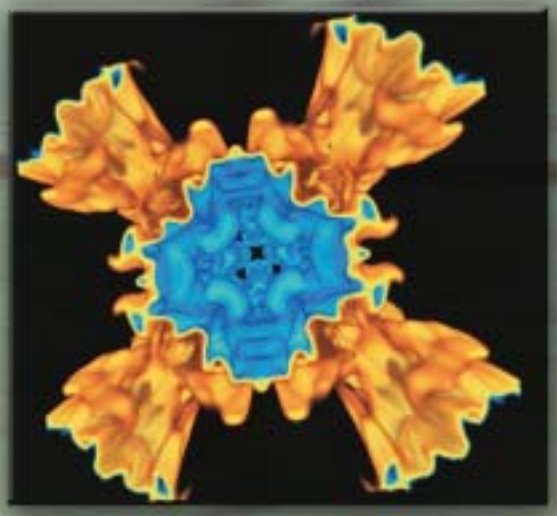
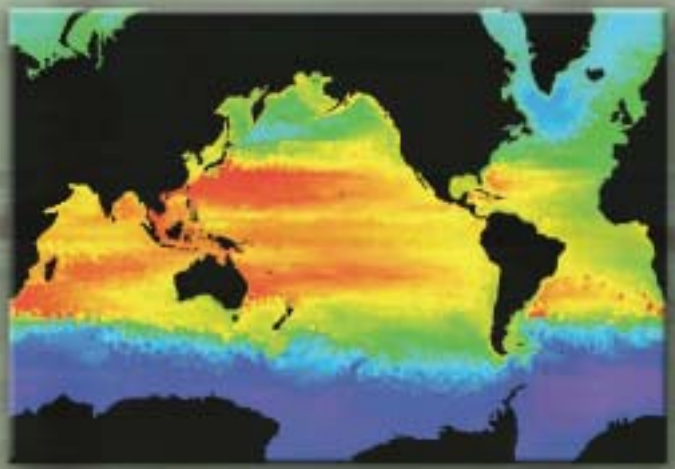
DOE Science Networking Challenge: Roadmap to 2008









Office of Science
U.S. Department of Energy

DOE Science Networking Challenge: Roadmap to 2008

Report of the June 3-5, 2003, DOE Science Networking Workshop
Conducted by the Energy Sciences Network Steering Committee at the request of the Office of Advanced Scientific Computing Research of the U.S. Department of Energy Office of Science

Workshop Chair

Roy Whitney

Report Editors

Roy Whitney
Larry Price

Energy Sciences Network Steering Committee

Larry Price, Chair
Charles Catlett
Greg Chartrand
Al Geist
Martin Greenwald
James Leighton
Raymond McCord
Richard Mount
Jeff Nichols
T.P. Straatsma
Alan Turnbull
Chip Watson
William Wing
Nestor Zaluzec

Working Group Chairs

Wu-chun Feng
William Johnston
Nagi Rao
David Schissel
Vicky White
Dean Williams

Workshop Support

Sandra Klepec
Edward May

Argonne National Laboratory, with facilities in the states of Illinois and Idaho, is owned by the United States Government and operated by The University of Chicago under the provisions of a contract with the Department of Energy.

DISCLAIMER

Available electronically at http://www.doe.gov/bridge and at http://www.es.net/hypertext/welcome/pr/aboutesnet.html #programs

Available for processing fee to U.S. Department of Energy and its contractors, in paper, from:

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
Phone: (865) 576-8401
Fax: (865) 576-5728
Email: reports@adonis.osti.gov

We are grateful for the editorial assistance of Steve Corneliussen at Jefferson Lab. Argonne's Information and Publishing Division prepared the document for publication, and we are thankful for the services of Vic Comello who served as editor, Judi Damm who developed the document's graphical design, and Gary Weidner who managed print production.

Cover design by Judi Damm, Information and Publishing Division, Argonne National Laboratory

Table of Contents

Tables

Figures

DOE Science Networking Challenge: Roadmap to 2008

U.S. Department of Energy (DOE) Science Networking resources are crucial for achieving DOE's 21st century science mission. This report puts forth a roadmap for the networks and collaborative tools that the Science Networking and Services environment requires for DOE science fields including astronomy/astrophysics, chemistry, climate, environmental and molecular sciences, fusion, materials science, nuclear physics, and particle physics. Success in these fields depends on scientists' ability to move large amounts of data, access experimental and computing resources via the network, and collaborate in real time from multiple locations across the country and around the world. Implementation of the roadmap presented in this report will be a critical element in keeping DOE a leader in world-class scientific discoveries.

Across the globe, new networking capabilities are emerging and being enthusiastically incorporated — examples include computational and data grids (large numbers of computers and data resources working together across networks), high-speed wireless networking, super-high-speed metro-scale networks for relatively nearby sites, all-optical routers and switches, and inexpensive connections to local computers. Each new capability enables substantial new collaborative functions and efficiencies. However, sophisticated structures and services can be used effectively only if the network infrastructure itself provides the necessary environment. Increasingly, the network must become a collaborative medium for exchanging information, with a core of higher-level services supported by the network providers, in addition to meeting the basic requirements of bandwidth and connectivity. Thus, this report, the result of workshop input from 66 of the nation's leading scientists and their collaborators, proposes initiatives in three areas:

- **Production and high-impact networking.** The operational "production" services that defined the early generations of scientific networking must continue to evolve. Also needed are "high-impact" network services for high-rate transfers of increasingly enormous volumes of data — the terabytes (millions of megabytes) and even petabytes (billions of megabytes) that at present can be handled and analyzed only at originating locations.
- **Technology, services, and collaboratory tools.** Emerging from R&D programs are new, higher-level capabilities in the areas of collaboratory tools and middleware, the software that makes disparate software applications "interoperable," much as the World Wide Web does, and "manageable" as a system of facilities spread nationally, and globally. These emerging capabilities need to be operationally supported more broadly, with systematic progression from R&D to pilot programs to long-term production use.
- **Network research.** A separate, dedicated, R&D network is needed to allow the testing of new protocols while permitting science to proceed in parallel without interruptions caused by network failures and by test requirements for extremely high bandwidth.

The goal of updating DOE Science Networking aligns directly with national priorities as articulated in June 2003 by the directors of the Office of Science and Technology Policy and the Office of Management and Budget.[1] Without an enriched information infrastructure supporting DOE science, fewer breakthroughs would be accomplished and fewer answers to research questions would be obtained with the available funds. For DOE to achieve the goals of its investments in new scientific directions, DOE networking and services must match or exceed the worldwide pace of development.

New costs for the proposed effort start at an estimated $15.5M in Year 1 and grow, as more capabilities are incorporated to $21.5M in Year 5. Since the FY 2003 budget for ESnet, middleware, collaboratory pilot programs, and network research is $39M, the increased funding for the new DOE Science Networking and Services capabilities amount to a 55% growth by the end of the 5-year period.

[1] *Memo M-03-15 from John H. Marburger III and Mitchell E. Daniels, Jr., for the Heads of Executive Departments and Agencies, dated June 5, 2003.*

1. Introduction

This report establishes a roadmap for a new approach to the DOE Science Networking and Services needed for science in the U.S. Department of Energy in the 21st century. It has become increasingly clear[2] that the network provided for DOE science in the past will not be adequate to keep that science competitive in the future. This roadmap, if implemented and followed during the next five years, will solve that problem. The past 5 years have seen a broad and general movement toward the assumption of and reliance on networked systems in all of the large new initiatives for DOE science. It is clear that the success of science depends increasingly on the ability of scientists to move large amounts of data, access computing and data resources, and collaborate in real time from multiple remote locations. It is also abundantly clear that business-as-usual in the network and information services that underpin the scientific collaborations will fall woefully short of what is needed. New capabilities such as computational and data grids, high-speed wireless networking, super-high-speed metro-scale networks, and cheap gigabit Ethernet have arrived in turn and have been enthusiastically incorporated into the arsenal of science, each permitting substantial new collaborative abilities and efficiencies. However, sophisticated structures and services using basic network connections can be used effectively only if the network infrastructure itself provides the necessary environment. Increasingly, the network must become a collaborative information exchange, with a core of higher-level services supported by network providers in addition to basic bandwidth and connectivity.

The August 2002 workshop, High-Performance Networks for High Impact Science, and its report[2] studied in some detail the network requirements of the coming generation of science programs and facilities in the DOE Office of Science (SC), using scenarios submitted by investigators in each of the SC programs. Analysis of these scenarios led to these conclusions (quoting from the report):

- Increasingly, science depends critically on high-performance network infrastructure, where much of science already is a distributed endeavor or rapidly is becoming so.
- We can define a common "infrastructure" with advanced network and middleware capabilities needed for distributed science.
- Paradigm shifts resulting from increasing the scale and productivity of science depend on an integrated advanced infrastructure that is substantially beyond what we have today.

  These paradigm shifts are not speculative. Several areas of DOE science already push the existing infrastructure to its limits as they implement elements of these approaches. Examples include high-energy physics with its worldwide collaborations distributing and analyzing petabytes of data; systems biology access to hundreds of sequencing, annotation, proteome, and imaging databases that are growing rapidly in size and number; and the astronomy and astrophysics community that is federating huge observation databases so it can, for the first time, look at all of its observations simultaneously. The clear message from the science application areas is that *the revolutionary shifts in the variety and effectiveness of how science is done can only arise from a well integrated, widely deployed, and highly capable distributed computing and data infrastructure, and not just any one element of it.*

It is no accident that these observations and the urgent need to update the science information infrastructure fit remarkably with the national priorities for science and technology articulated by the directors of the Office of Science and Technology Policy and the Office of Management and Budget in their memo of June 5, 2003[3] with the subject "FY 2005 Interagency Research and Development Priorities." That memo says, in part, "In general, the Administration will favor investments in Federal R&D programs that sustain and nurture America's science and technology enterprise through the pursuit of … critical research fields *and their enabling infrastructure*"

[2] *http://doecollaboratory.pnl.gov/meetings/hpnpw/finalreport/*

(emphasis added). Of the five Interagency Priorities for R&D Budgets listed in the memo, one is *Networking and Information Technology R&D*.

This memo's focus on networking recognizes advanced networking infrastructure as a basic enabler of present-day science, and as an area that presents great opportunities for the future empowerment of modern science and technology. The rapid pace of advances in the world of networks and network services, as well as the specific interest of the broader federal government, present both a challenge and an opportunity for DOE science. It is time to make a concerted effort to systematically embrace the rush of new capabilities and to formulate a detailed plan to use them to keep DOE science at the forefront of a new generation of scientific discoveries.

The August 2002 High-Performance Networks for High Impact Science workshop report called for the development of a roadmap for an integrated infrastructure that would include:

- A new network provisioning model supporting an integrated three-element network with *production-level networking* in support of traditional program requirements; *network resources for high-impact DOE science programs*, including science application and grid research; and *network resources for network research* that enable experimentation with new concepts.
- Enabling middleware research.
- Enabling network research and accelerating the deployment of the fruits of this research in the service of science.
- A network governance model appropriate to these integrated functions.

The June 3–5, 2003, workshop that led to the present report brought together a broad group of 66 experts, including active investigators from DOE science programs and experts on network operations and emerging network capabilities. They represented universities, national and international laboratories, Internet 2, National Lambda Rail, USAWaves, and three major U.S. telecommunications vendors, plus the DOE Office of Science itself. These participants with their various backgrounds and in some cases competing interests agreed with the following key points:

- The roadmap for production, high-impact, and research networks being presented in this report is the most effective and efficient path for the Office of Science to achieve its networking-related scientific goals,
- The Office of Science networking requirements differ significantly from standard commercial IP requirements and university requirements,
- The production and high-impact network boundary is at the 10 Gbps (i.e., lambda) level for the foreseeable future,
- Only the Office of Science would do much of the network research necessary to meet the Office of Science requirements in a useful time frame,
- Collaborating with university, international, and commercial partners where possible would be very beneficial,
- Central management of the production and high-impact networks with a centrally managed collaboration for the research network would prove to be the most cost-effective and efficient way to achieve the Office of Science networking requirements, and
- Doing the R&D and then providing the core services to support collaboratory tools including grid technologies is critical to the ongoing efficient and effective infrastructure support of DOE science.

This workshop was one of a series of workshops orchestrated by several agencies with goals associated with advancing science. Appendix H lists and describes several of these influential earlier and June 2003 related workshops and conferences. This workshop started from the requirements of the August 2002 High-Performance

[3] *Memo M-03-15 from John H. Marburger III and Mitchell E. Daniels, Jr., for the Heads of Executive Departments and Agencies, dated June 5, 2003.*

Networks for High-Impact Science workshop report and developed a detailed roadmap listing milestones and estimated costs for providing the DOE Science Networking needs to keep DOE science competitive during the next five years. The infrastructure for science was planned in the following categories (which are detailed in the indicated later sections of the present report):

- Section 3 provides a plan for the production and high-impact network pieces of the three-part provisioning model.
- Section 4 identifies 13 middleware technologies and services in priority order and provides a detailed plan for deployment. The top five technologies were judged to be essential, and the next 3 as very important for the support of DOE science. Appendix D provides detailed descriptions and roadmaps for these 8 technologies and services.
- Section 5 outlines a coordinated plan of network research and the network resources needed for developing and testing the new capabilities in a way that can be coordinated with the production and high-impact network functions for efficiency.
- Sections 6 and 7 map structures for management and governance issues identified by the previous workshop.

Section 2 of the present report describes the rapidly evolving overall context for this roadmap. The result of implementing the roadmap during the next 5 years will be the substantially more capable, flexible, and cost-effective DOE Science Networking that will enable DOE science programs to make the most productive use of their research funding. If DOE does not take advantage of this opportunity to support its science with an enriched information infrastructure, less science — in other words, fewer breakthroughs and fewer questions answered — will be accomplished with the available funds. Since the European Union and individual European countries, including the UK and The Netherlands, are making plans for a substantial expansion of networking in support of research and education, we can expect that corresponding support in the U.S. will be needed to maintain the strong record of U.S. leadership in science. DOE runs the risk of negating its investment in new scientific directions if it does not provide correspondingly sophisticated infrastructure.

The new capabilities needed to meet the challenge posed by DOE science programs require a somewhat higher level of investment in the information exchange infrastructure. This investment is needed, however, to enable the effective use of the much larger investments being made directly in the science programs themselves. The costs of the additional capabilities are summarized in Section 8.

As indicated in the Executive Summary, new costs start at an estimated $15.5M in Year 1 and grow, as more capabilities are incorporated into the operating and supported networks, to $21.5M in Year 5, the last year considered in this 5-year roadmap. Since the FY 2003 budget for ESnet, middleware, collaboratory pilot programs and network research is $39M, the increased funding for the new DOE Science Networking capabilities amount to a 40% growth in the first year and a 55% growth by the end of the 5-year period. These increases are justified, considering how the enhanced networking and services infrastructure would be beneficial to the potential for scientific discovery across the Office of Science.

2. Achieving DOE Science Yesterday, Today, and Tomorrow

The mission of the DOE Office of Science is to achieve scientific discoveries by the most effective means possible with an optimal use of resources. Advanced supercomputers and experimental facilities play a vital role in pushing the frontiers of scientific discovery. Accordingly, the Office of Science funds 10 world-class laboratories and a large number of research groups at universities, and collaborating institutions. In this system, the three most valuable resources are:

- Highly trained collaborative groups of scientists having a wide spectrum of talents and backgrounds;
- World-class scientific tools, many of which are at the billion dollar scale of investment of federal resources; and
- Infrastructure and management systems that enable the scientists to make effective use of the tools.

The system is inherently large and complex. Scientists and engineers with diverse backgrounds frequently form both small and large collaborations to make scientific discoveries by taking advantage of various resources and adapting the tools and systems so as to make them an integral part of their daily working lives. They continuously work to improve their scientific tools and systems so that they can advance science.

One of the most useful advancements for science over the last half century has been the rapid evolution of integrated circuit technology. For the past several decades, the density of components on an integrated circuit has doubled every 18 months, and this trend is expected to continue unabated into the next decade. This growth rate, known as Moore's Law [1], has been incorporated in the technology roadmap of the global semiconductor industry [2]. For science, the impact of this increasing capability in processing power lies in increasingly more evolved and complex experiments performed faster and at much larger scales. Two corollaries are (1) that the amount of data that is produced is also rapidly increasing, and (2) the scientific environment is becoming more collaborative and complex. The first challenge has been dealt with by the rapid evolution of computing and networking infrastructures. In fact, networking capabilities have increased faster than Moore's Law for two decades. The second challenge has been dealt with by the evolution of collaboratory/middleware tools, such as the World Wide Web [3], which was invented in a high-energy physics laboratory to improve sharing of experimental data and information.

Science-driven networking requirements for achieving discoveries derive from three factors:

- The volume of data, both experimental and from simulations;
- The collaborative tools used for analyzing and understanding the data; and
- The visualization, computational steering, and other desktop computing tools used by scientists.

Advances in all three of these areas have resulted in the growth of traffic on the Office of Science's Energy Sciences Network (ESnet), which has doubled every year since 1992. To fully appreciate this, understand that on any single day today, ESnet transports more bits of information than it did for the entire years of 1992 and 1993 combined! To help in understanding the scientific drivers, the following table provides some specific examples of DOE scientific goals and the associated experimental, simulation, and analysis data going to media that are involved in achieving the goals. Much of this information is from the August 13-15, 2002, workshop report, *High-Performance Networks for High-Impact Science*.

Table 2-1 Science Data Network and Collaboratory Drivers

1995 – 1999	2002 – 2004	2007 – 2009
Climate In 1998, there were about 5 TB/year of experimental and simulation climate data going to media. About this time, the DOE and other agencies launched a long-range program to acquire experimental data and support simulations.	Climate experimental data and modeling data at the three largest U.S. facilities currently totals 100 TB (NERSC – 40 TB, ORNL – 40 TB, and NCAR [non-DOE] – 20 TB) and is being added to at a rate of 20 TB/year.	By 2008, network-assessable climate experimental and simulation data in the U.S. will be increasing at rate of 3 PB/year. This is due to greatly enhanced experimental measurements and simulations.
Fusion Energy Plasma physics/fusion research at DOE's three main experimental facilities — General Atomics, MIT, and PPPL — and numerical simulations generated 2 TB of data in 1998 (mostly from experiments).	Present plasma physics/fusion experiments and simulations are generating 20 TB/year of data (each contributing roughly half).	Driven mainly by large-scale advanced simulations and preparation for a burning plasma experiment, fusion researchers will be generating 1 PB/year of data by 2008. They also need the necessary collaborative tools to be full partners in the international program.
Hadron Structure Investigation of the quark-gluon structure of the nucleon and nuclei resulted in 50 TB of data and analysis the first full year of operation of all of the experimental facilities of CEBAF at JLab in 1998.	Currently CEBAF experiments and analysis, including those associated with the discovery of the pentaquark, produce 300 TB/year of data.	CEBAF's upgrade to 12 GeV to investigate quark confinement and detailed quark distributions will produce several PB/year.
Quark-Gluon Plasma The goal for the RHIC at BNL is discovering the quark-gluon plasma thought to exist at the edge of the Big Bang. RHIC began operations in 2000.	RHIC has early results that indicate that it may have discovered the quark-gluon plasma and is currently putting 600 TB/year to media.	By 2008, RHIC will increase the amount of data going to media to 5 PB/year as it details its information on the quark-gluon plasma.
Materials Science – Neutrons Neutron Science is critical for investigating the properties of materials by neutron scattering.	The SNS is currently under construction at ORNL. It will increase the U.S.'s neutron science capabilities by more than an order of magnitude.	The SNS will turn on in late 2006 and achieve full operation in 2008, at which time it will produce 200 TB/year of data and analysis.
Materials Science – Photons The four DOE-funded light sources (ALS, APS, NLS and SSRL) are used to investigate the properties of materials and the structure of biological molecules, such as proteins. In 1998, they accumulated 3 TB of data.	Currently the four light sources are acquiring and sending data at the rate of 30 TB/year over ESnet.	The drive to understand the dynamics as well as the structure of materials and biological molecules using greatly enhanced detectors will result in at least a 5-fold increase in the acquisition of data at the light sources by 2008 to 150 TB/year.

1995 – 1999	2002 – 2004	2007 – 2009
Chemistry – Combustion Simulations for combustion are critical to improve our use of energy. The simulations were generating 100 GB/year in 1998.	Construction of a Web-based archive for collaborative sharing and annotation of a broad range of chemical science data is now under way. Combustion is currently generating 3 TB/year and is storing annotated feature and data subsets to this archive.	In 2007, combustion simulations will produce several PB/year of data to be collaboratively visualized, mined, and analyzed. In addition, there will be several 100s of TB/year of experimental data generated, plus publication and annotation in Web-accessible archives of 100s TB/year for collaborative research.
Chemistry – Environmental EMSL at PNNL came on-line in 1997 with the mission of understanding and controlling the molecular processes that underlie our environmental problems. In 1998, it put 250 GB to media.	EMSL's unique combination of simulations, high-field magnetic resonance instruments, high-performance mass spectrometers, optical imaging instruments, and more generate 100 TB/year to media.	As high rate proteomic and nanoscale facilities and high-end supercomputers come on-line, EMSL's rate of putting data to media will increase to 2 PB/year by 2008.
Genomes to Life In the area of proteomics and metabolomics for Genomes to Life (GTL), there was less then 10 GB of data on-line in the world in 1998.	Proteomics and metabolomics currently are capable of generating 400 TB/year. Note, GTL information for a single microbe generates 20 PB of proteomic data and 16 PB of metabolite data.	Proteomics and metabolomics data generation has the potential to increase to the level of tens of PB/year by 2008.
Particle Physics In the search for the fundamental building blocks of the universe, the discovery of the top quark at the FNAL in 1995 required 70 TB of data and analysis from 1992 to 1995.	For the search for the Higgs boson at FNAL, 500 TB/year of data and analysis are currently being put to media.	Investigation of the properties of the Higgs boson will result in CERN Large Hadron Collider experiments acquiring 10 PB/year of data. 3-4 PB/year of the data will be moved to BNL and FNAL, and then onto U.S. universities, beginning in 2007. Processing this data will generate several additional PB/year.
Universe Asymmetry BaBar's mission at SLAC is to discover why our universe has an asymmetric distribution of matter and anti-matter. It went on-line in 1999.	BaBar currently has 200 TB/year of data and analysis going to media. To date, over a PB has been moved to partners in Europe for analysis.	Upgrades to the PEP-II accelerator will result in a quadrupling of BaBar's 2003 rate to close to 1 PB/year going to media as it searches for a deep understanding of processes at the origin of our universe.

As seen in the table above, on average from 1998 to 2008, there will be a 500- to 1,000-fold increase in the amount of data going to media at many DOE Office of Science facilities. As systems become more distributed and more integrated the amount of data transported on demand (as well as in an organized fashion) increases more rapidly than the amount of data acquired and processed at the central laboratories. Hence, 1,000 times per decade may be an underestimate, especially as effective data-intensive grid systems are built. These estimates roughly match the doubling seen every year in the amount of traffic moving across ESnet. What follows is a summary of the key factors driving this increase:

- The most important factor is that for many experiments, more data results in the increased potential for scientific discovery. In addition, the faster the data can be acquired, analyzed, and simulated, the faster the pace of scientific discovery. Scientists are very motivated to get data as rapidly as possible.
- Moore's Law of doubling the density of electronic circuits every 18 months applies to detectors as well as computers. Scientists have been very aggressive in increasing the spatial resolutions of their detectors. This corresponds to greatly increased channel density and consequently substantial increases in their data rates.
- For many scientific instruments, there are two additional dimensions that can increase data rates even faster than Moore's Law. The 100 megahertz clock speeds of the early 1990s have been replaced by gigahertz speeds in 2003 and will increase by close to a factor of 10 by 2008. This means that the ever higher density detectors are also pumping out data faster and faster. The second additional dimension is that for some experiments, the instruments can be layered in the physical third dimension. As the instruments shrink and their component costs decreases, multilayer instruments will become more common. Again, the result is data going to storage media at higher rates.
- Simulations have matured to the level that they are now considered to be the third leg of science, complementing theory and experiment. High-end computers have been growing in capabilities even faster than desktop computers. In terms of producing data from simulations, the software environments for high-end computers have advanced in capabilities at rates matching or exceeding Moore's Law. For many areas of science, high-end computers now generate and store simulation data to media at rates comparable to experiments, and in some cases exceed them.
- Experimental and/or simulation data stored in media (raw experimental data, analyzed data, simulated data, etc.) is typically analyzed by multiple scientists using multiple tools. Sometimes these tasks are carried out on very high-end visualization systems, but more often on a scientist's desktop. The capabilities of these desktop computers have been doubling roughly every 18 months. Since 1996, the disks for desktop computers have been increasing in storage density even faster, at rates over 100% per year. This rate is projected to return to the Moore's Law rate of 60% per year for the next 5 years. As seen in the table above, scientists have vast stores of data that they frequently move to and from their desktops and through multiple computational systems.

2.1 Science-Driven Collaboratories

A number of DOE large-scale science projects critically depend on collaborations of multidisciplinary researchers who collectively require capabilities that are unavailable at any single national laboratory or university. These projects span a wide spectrum of disciplines, including high-energy physics, climate simulation, fusion energy, genomics, and astrophysics, among others. In addition, the new experimental facilities coming

on-line, such as ITER, LHC, and SNS, as well as the currently active facilities, such as ALS, APS, CEBAF, EMSL, FNAL Tevatron (Run II of CDF and D0), NLS, RHIC, and the SLAC PEP-II accelerator (BaBar), SSRL, and others, present unprecedented requirements for distributed, collaborative data analysis. These collaborations invariably involve geographically distributed resources such as supercomputers and clusters that offer massive computational speeds, user facilities that offer unique experimental capabilities, and repositories of experimental and computational data. These teams of researchers could be dispersed across the country or around the globe. Compounding the problem in some cases, access to these facilities must be tightly coordinated and controlled over wide-area networks. Indeed, seamless access to these distributed resources by researchers is essential to carrying out DOE missions, and the "network" and the associated collaboratory or grid tools have become critical components of the modern scientific infrastructure, much like the supercomputers or experimental facilities.

The DOE Office of Science envisions a seamless, high-performance network infrastructure to facilitate collaborations among researchers and their access to remote experimental and computational resources. Such an infrastructure can eliminate resource isolation, discourage redundancy, and promote rapid scientific progress through the interplay of theory, simulation, and experiment. For example, timely distribution of multi-petabytes of LHC data produced at CERN, in Switzerland, can eliminate the bottleneck experienced by U.S. physicists today due to inadequate bandwidth in the trans-Atlantic and U.S. networks. Also, the ability to remotely access complex scientific instruments in real time will enable interactive collaborations among geographically dispersed researchers, without the need for coordinated travel and duplications of specialized experimental instruments. An example is ITER, where it is envisaged that the new facility will be operated remotely by teams of geographically dispersed researchers from across the world.

In the August 2002 workshop, representatives of a range of DOE science disciplines were asked to provide information on how they currently use networking and network-associated services and what they saw as the future process of their science that would require, or be enabled by, adequate high-performance computers, high-speed networks, and advanced middleware support. Climate modeling has been picked as one of four examples from the August 2002 workshop to illustrate the importance of networks with enhanced services as part of an integrated cyber infrastructure for science.

Better climate modeling [4] is essential to understanding phenomena such as hurricanes, droughts and precipitation pattern changes, heat waves and cold snaps, and other potential changes that, e.g., promote disease-producing organisms or impact crop productivity. Better climate modeling requires very high-performance computing to permit simulation of realistic spatial and temporal resolution — it makes a huge difference in our ability to accommodate the impact of a sustained drought if we know the county-level geographic extent of the drought ten or twenty years in advance, rather than only that a drought is likely in this century and that it will affect the Midwest.

"Climate model" is a bit of a misnomer because the climate is determined by a complex interplay of physical and biological phenomena (See Figure 2-1). There are dozens of models connected by feedback loops that must be included in a realistic simulation of climate that will result in the accuracy needed to inform policy and advance planning issues that are critical for the well being of our society. The complexity of climate is typical of most macro-scale phenomena from cosmology to cellular function, so the issues raised by climate modeling are characteristic of much of science.

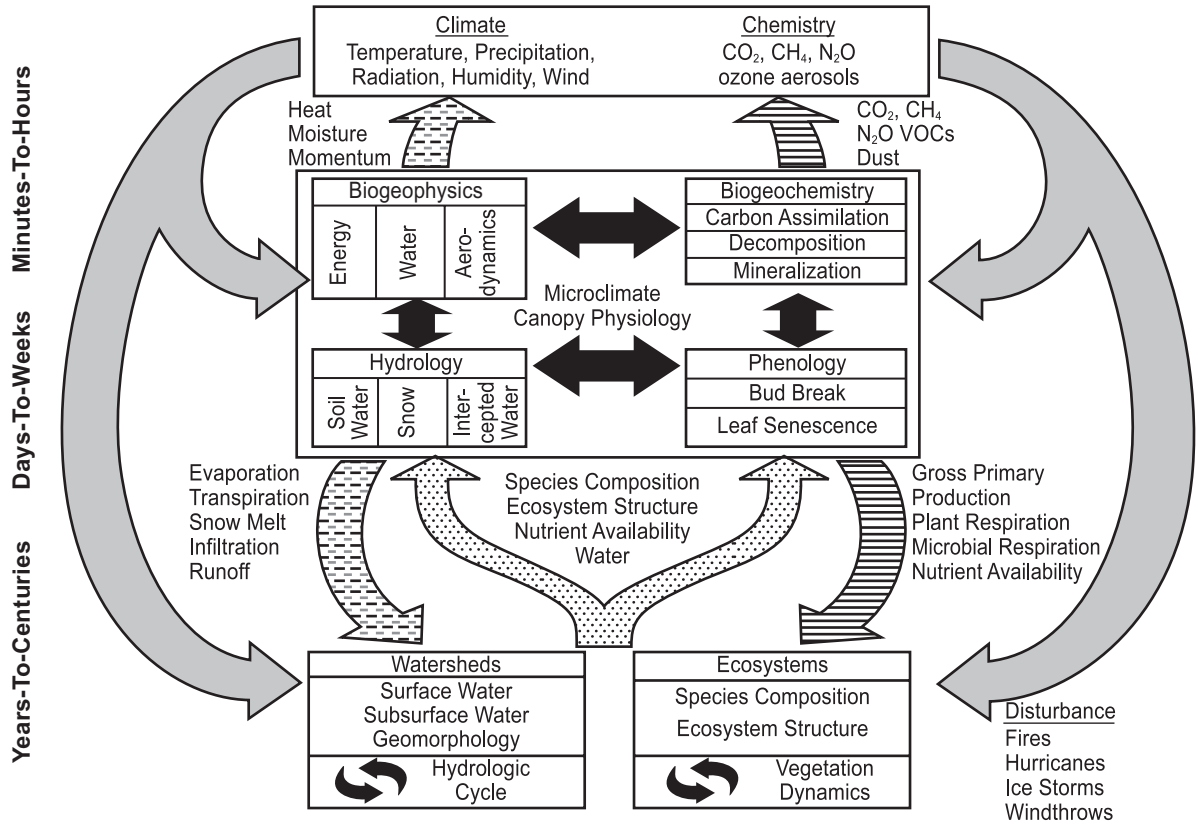


Figure 2-1 The Complexity of Climate Simulation [5]

Since the climate is an extremely complex phenomenon that involves modeling many separate elements in order to obtain the required accuracy, each of these elements is a discipline in its own right, and is studied by a different group of specialists.

Better climate modeling requires that the many institutions working on various aspects of climate be able to easily describe, catalogue, and seamlessly share the knowledge and the vast amounts of data that underlay the knowledge in order to facilitate the required interdisciplinary collaboration. Nonetheless, all of these sub-models must interoperate in the same way that all of the elements that make up the climate interact. This multidisciplinary simulation produces an inherently distributed computing environment as the models of the discipline's specialists are accessed and combined into an overall model of the climate via the collaboratory or grid environment.

Further, the many specialized scientific groups that work on the different components that go into a comprehensive model build specialized software and data environments that will probably never be homogenized and combined on a single computing system. Almost all multidisciplinary simulations are inherently distributed, with the overall simulation consisting of software and data on many different systems combined into a virtual system by using collaboratory tools and facilities for building distributed systems. This, then, represents the vision of the future process of science in the climate community — to have the necessary computing power, access to annotated data, and interoperation of diverse sub-models, i.e., a collaboratory such that a realistic model of climate can make predictions that have great value to human society.

2.2 Science-Driven Evolution of Common Network Services

Over the last two decades, DOE science managers took several key steps called for by the rapid expansion of data and collaborations needed to achieve DOE science. In the mid-1980s, the utility of improved networking between DOE laboratories and their university collaborators was recognized, and several networks, including the High Energy Physics Network (HEPnet) and the Magnetic Fusion Energy Network (MFEnet), that ran different protocols were combined to form the ESnet. Although ESnet started as a multi-protocol network, the Internet Protocol (IP) is now used throughout because of its compatibility with the university communities and commercial vendor tools. Beginning in the early 1990s, the development of collaboratory tools began in earnest. Initially they were focused on distributed computing, file sharing, and instrument control. Three 1990s DOE/SC/ASCR/MICS programs in this area were the Distributed Informatics, Computing, & Collaborative Environment (DICCE); the Distributed Computing Experimental Environment (DCEE); and DOE2000 Collaboratories. The DOE2000 program is now expanding to include grid technologies, and currently the Scientific Discovery through Advanced Computing (SciDAC) program and MICS are supporting the R&D and implementation of grid-style collaboratory and computational tools for DOE science. The following table gives some examples of how the collaboratory tools are advancing in the DOE science environment.

While almost everyone connected to the Internet use tools, such as e-mail, it is largely the geographically distributed science community with its petabytes of data that is driving the usage of computational grids, remote instrument control, and collaborative visualizations, and DOE scientists with their vast research facilities are among those who are leading the way. To keep DOE science on track for the coming five years and longer, the networking and collaborative tools will need to match both the explosive growth of scientific data and the collaborative resources needed to produce, analyze, and understand the data. The R&D going into collaboratory tools and grid technologies will need to move into production services as long-term infrastructure available to support the mission of the Office of Science.

Table 2-2 DOE Science Community Use in Percent of Middleware Services

Middleware Service	1998	2003	2008
IP based audio-video/Access Grid/VRVS	<5	10	80
IP based telephone	<1	5	30
ISDN based video conferencing	10	50	5
Global directories – finding people	50	80	95
Global directories – finding services	<1	20	80
Computational grids for science	<1	20	80
Remote instrument control	<5	10	50
Collaboratively shared visualization	<1	10	50
Web services/portals	<2	20	80
Security infrastructure – PKI/certificates	<1	20	80
Security infrastructure – secure protocols	10	80	99

Due to the importance of the collaboratory tools, serious consideration has been given to changing the name of the Energy Sciences Network to something more inclusive. The argument for doing this is that a new name would symbolically capture the broader impact of ESnet and the collaboratory tools of DOE science. The argument against would be that in the networking community and across the Office of Science and in Congressional committees that support DOE, ESnet is recognized as one of the best (if not the best) networks in the world for support of science. In this report, we propose that the networking portion of the larger enterprise remain known as ESnet and that a new umbrella name such as Science Networking and Services be deployed to include both ESnet and the collaboratory/grid environment for DOE science.

In summary, as seen by the science drivers presented above, it is projected that Office of Science networking and services requirements will continue to double (or more) every year for the next five years (as they have since 1992). Meeting these networking requirements will require research and development specifically targeted at Office of Science networking issues. In addition, grid-style collaboratory tools will need the projected enhancements to be able to be used in efficiently and effectively managing the data and achieving the scientific discoveries that are the mission of the Office of Science.

2.3 References and Notes

1. Moore, G.E., *Electronics* 38(8), April 19, 1965. Available at: ftp://download.intel.com/research/silicon/moorespaper.pdf and http://www.intel.com/research/silicon/mooreslaw.htm.
2. International Technology Roadmap for Semiconductors 2002 Update. Available at: http://public.itrs.net.
3. See: http://public.web.cern.ch/public/about/achievements/www/www.html.
4. This scenario is due to Al Kellie, Gary Strand, and Tim Killeen of the National Center for Atmospheric Research (NCAR).
5. Figure courtesy of Gordon Bonan, NCAR. It is taken from Bonan, G., *Ecological Climatology: Concepts and Applications*, Cambridge: Cambridge University Press (2002).

Substantial changes in almost all aspects of the networking and collaboration environments are taking place within DOE and within the larger communities of DOE's collaborators, both at universities and internationally. Consequently, to successfully provide networking and collaborative services for scientific discovery, some basic assumptions (including the provisioning model) need to be changed to take advantage of the changing technologies, scientific resources, and opportunities.

ESnet today is provisioned as a production network providing base IP services plus a number of special services that are significantly beyond standard commercial IP services but are critical for achieving Office of Science mission goals (see Appendix B on ESnet today). Production means that the services are intended to be available +99.9% of the time. Note that the provisioning model for the current production services is that many of the services are provided by commercial vendors under the management of a central group at LBNL that manages the integration of the standard commercial fare with the additional services needed by the Office of Science.

3.1 Provisioning Models

The August 2002 workshop, High-Performance Networks for High-Impact Science, looked at alternative models for meeting the networking and middleware/collaboratory requirements to achieve scientific discovery. The critical issue, of course, is how to provide networking when the demand grows 100% every year and when network technologies go through rapid generational changes to meet (and fuel) this growth in demand, as is now occurring in going beyond 10 Gbps, and when environments need to be set up to support collaboratory tools across both R&D and production grids. With respect to networking, the workshop arrived at the following network provisioning model:

Network Provisioning Model. To be responsive to applications' needs in a timely and cost-effective manner, the programs of the Office of Science would benefit from the formation of an integrated three-element network provisioning model that provides:

1. ***Production level networking*** in support of traditional program requirements.
2. ***Network resources for high-impact DOE science programs*** including science application and Grid research—This element provides a platform for deploying prototypes to those programs that require high-capability networking or advanced services that are not satisfied by production level networking.
3. ***Network resources for network research*** needed to support high-impact DOE science programs. These resources enable experimentation with new concepts and technologies that are relevant to the science-program needs.

An integrated network provisioning strategy is considered optimal for the three-element network. Networking for DOE science would benefit from a process of planning, coordination, funding, and implementation that encompasses all three elements. Factors that should be taken into consideration include the following:

- *A shared vision of success must be motivated, where some measures of success are across all three elements.*
- *As new services are moved into production, some production support costs likely will increase.*
- *The network program must position itself to be agile and not rooted too firmly in any one provisioning model.*

The June 2003 workshop reaffirmed the above provisioning strategy and proposed a provisioning roadmap through 2008 that is presented in this section. A high-level description of the model follows:

- IP production services with +99.9% reliability will continue to be critical for achieving DOE

science. Collaboratory activities, such as remotely running experiments and steering supercomputer applications, are especially dependent on the reliability of these production services. In addition, the scientific work depends on managed production services that are currently provided in addition to commercial IP services.

- High-impact network services refer to services required by a subset of production network sites, whose demands may be extremely high at times but for which a considerably less reliable and potentially schedulable service will nevertheless meet the needs. Provision of such services at greater than 99.9% reliability with the ability to match potential peak demands would otherwise be prohibitively expensive. These high-impact services will be designed to move the massive amounts of data from experiments and simulations with multi-terabyte to multi-petabyte datasets. Other network services with a major impact on science, such as high-end remote visualizations, that cannot be provided cost-effectively by the production network will be candidate high-impact network services.
- The research network services will exist in a network test bed environment in which R&D specific to Office of Science requirements for networking can be performed in partnership with other network test beds. Commercial IP providers must deal with the problem of "many small data flows" — moving small/modest amounts of data around for tens or hundreds of millions of customers. Research network providers such as ESnet face the problem of "a few very large data flows" — scientists need to move terabyte/petabyte-scale datasets around to tens or hundreds of scientists collaborating on scientific discovery. The problems are quite different, and significant research is required to provide efficient and effective solutions.
- The current ESnet management and governance elements will be expanded to orchestrate the movement of R&D results from the research network to the high-impact and production networks. In addition, this will take place for the implementation of those central services that support R&D and the deployment of collaboratory tools. ESnet management will continue its relations and partnerships with the broad international science community that are critical to achieving modern goals for scientific discovery.

Several technical factors affect our understanding of where the break-point between production and high-impact networks will be. The underlying physical mechanism for networks is that the information is carried by modulation of light in a fiber cable. Multiple wavelengths of light can be carried within one fiber. For current and projected electro-optical technologies,[1] the maximum rate of information that a can be carried on a single wavelength is 10 gigabits/second (Gbps), also known as OC192. Carrying information at rates beyond 10 Gbps requires the use of multiple wavelengths, also known as multiple lambdas (for the Greek symbol used for referring to wavelengths). In addition, it is uncertain that the current transmission control protocol, TCP, can be advanced adequately to efficiently transport information faster then 10 Gbps, and it is uncertain how to control multiple data streams with different priorities at these speeds. These are two critical areas of R&D planned for the research network.

ESnet today already has a portion of its backbone running at 10 Gbps (OC192) and plans to get most of its major links to 10 Gbps by 2005. The problem is that with the 100% growth in demand

[1] *Transmission using single wavelengths at 40 Gbps and above is technically possible, but is currently difficult and expensive. The consensus of experts at the June 2003 workshop was that commercial offerings would remain at 10 Gbps for the foreseeable future. The broader expert community is split on this issue, and major equipment vendors are preparing to support 40 Gbps in the marketplace, if needed.*

every year, by 2005-2006 depending on the link, 10 Gbps will not be adequate to meet the demands coming from Office of Science experiments, simulations, and visualizations. The projection is that by 2008 the ESnet backbone will need to be at 40 Gbps or higher to meet the demands. Looking at what technology can do and what the demands will be leads to a natural break-point between production and high-impact networking:

- Through the 2008 time frame, the production networking requirements should be able to be met by a 10 Gbps backbone.
- The high-impact network will come in lambda increments of 10 Gbps and be capable of meeting 40 Gbps by 2008 or sooner.
- The combination of the production and high-impact networks will be required to meet the full requirements for network capacity.

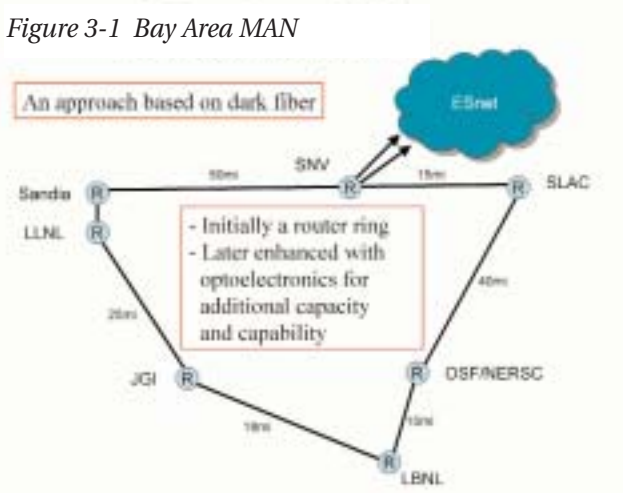
A major challenge is that the technologies do not exist today to take data from a single source and move it to a single remote destination beyond 10 Gbps. In fact, doing this at this rate from data sources to data destinations even in the same computer center is far from routine today. This challenge is known as the end-to-end (E2E) challenge. The network techniques being considered for meeting the challenge of greater-than 10-Gbps data transport include lambda circuit switching and optical packet switching, both of which are on the leading edge of R&D. The roadmap for meeting this challenge is the subject of the Section 5 of this report on research networking.

There are alternative ways to consider provisioning the networking to meeting the 40 Gbps requirements of 2008:

- The Office of Science could procure dark fiber and then light it itself. This has the advantage of system control, possibly leading to lower costs in some cases. It has the disadvantage of substantial up-front costs. This trade-off has been investigated and appears to be a net advantage in the near-term for the Bay Area and the Chicago area for Metropolitan Area Networks (MANs) where there is close proximity of multiple Office of Science laboratories and user interconnects. Once the fibers have been procured, adding more wavelengths only requires upgrades or additions to the equipment at the ends of the fibers.
- Multiple lambdas can be procured from a commercial vendor. This has the advantage that the vendor is providing the operation and maintenance of the lambda service. The issue may be finding a vendor that can support the lambda circuit switching or optical packet switching in the time frame necessary to meet Office of Science requirements. Due to the current and likely near-term status of telecom vendors, this issue may not be constraining. The disadvantage of this option is that the costs over time may be higher for some portions of the network.

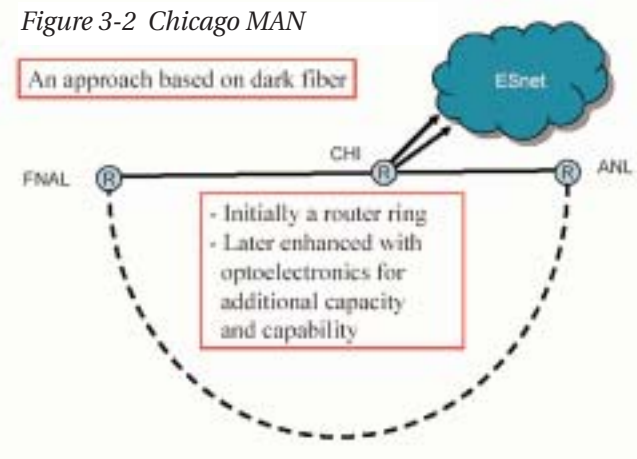
The Bay Area and Chicago area proposed MANs are shown in the following two figures:

Figure 3-1 Bay Area MAN



[2] *In a MAN environment, no fiber path is long enough to require amplification or regeneration equipment along the path.*

Figure 3-2 Chicago MAN



Additional advantages of the two MANs are that they include the current and potential long-haul telecom vendor's points-of-presence (POPs) so that if/when the vendors change, it will be relatively inexpensive to switch vendors. There also may be other somewhat larger MANs to be explored that would include Massachusetts to Virginia and possibly Tennessee.

3.2 Production and High-Impact Roadmap

The June 2003 workshop explored the above alternatives and laid out a roadmap that appears to take advantage of the best of each. The following table shows the goals and deliverables for the roadmap. Note that the schedule to deploy lambda switching may need to be advanced, depending on the schedule of deployment of lambda switching by DOE science collaborators. The funding estimates have been provided by ESnet management and include estimates based on vendor projections.

There are a number of key elements in the roadmap:

- A phased flow occurs for moving capabilities from the research network to the high-impact and production network services.
- Both MANs and WANs are critical for optimal deployment.
- E2E is a major theme. The data flows and collaboratories/grids must function all the way from the data sources such as DOE lab experiments, supercomputer centers, data centers, etc., to the collaborators at universities and other labs including international institutions.

3.3 Production and High-Impact Network Management

Alternatives were considered at the workshop for the management of the provisioning of the roadmap. A clear consensus was that to be both effective and efficient, there needs to be an agile central management of services. As one international participant commented, if the Office of Science did not have the centrally managed ESnet, it would have to create it. Additionally, buying networking piecemeal is very expensive, and such a plan would not allow for providing the additional services required for science that are beyond the typical interest of commercial IP vendors.

Central management is also effective for working relationships with the Office of Science's multiple university and international collaborators. Most of advanced networking for DOE science is global, and collaborating will also tend to keep costs under control and avoid duplication of efforts. Other leading academic and international research networks currently driving network development include Abilene, SuperSINET, SURFnet, Translight, UKLight, and USLIC.

Also as an example, at the June 2003 workshop and elsewhere in this report, it was noted that when the National Science Foundation ended NSFnet in 1995, the plan was that commercial vendors would provide the services. The university community served by NSFnet could not get the desired services directly from the commercial sector, so the university community set up Internet 2, also known as Abilene, to provide their required services. While a commercial vendor provides most of the underlying services to

Table 3-1 DOE Science Networking Roadmap

Year	Goal	Deliverable	Comments	Initial Funding ($M)	Annual Funding ($M)
0.5	A strategy for implementation of the roadmap.	ESnet management provides a strategy for review for going beyond OC-192 and achieving 40 Gbps by 2008. The strategy will include interfacing ESnet with other similar science, research, and educational networks.	The strategy will address production services, high-impact services, and networking research facilities.		
1	Establish local fiber communications infrastructure to support roadmap requirements.	Establish MAN dark fiber ring in the Bay Area. The ring will link DOE facilities and appropriate communications sites in the area.	The MAN shall be the basis for providing production services, high-impact services, and test bed facilities for networking research.	3	0.3
2	As above.	Establish a MAN dark fiber ring in the Chicago area.	As above.	1.5	0.2
2.5	Initiate lambda-switching capability.	Deploy lambda-switching equipment in Bay Area MAN.	Enable high-impact services in Bay Area.	2.2	0.2
3	As above.	Deploy lambda-switching equipment in Chicago area.	As above in Chicago area.	0.9	0.1
3	(1) Extend the MAN capabilities on a nationwide basis. (2) Gain experience in a switched lambda architecture.	Provide a switched-lambda capability between the two MANs at a minimum of 10 Gbps. Extend to labs outside MAN areas.	Approach to allow switching of lambda between production, high-impact, and research uses.	1.6	0.7
3.5	Gain experience at 20 Gbps.	Support R&D for E2E services at 20 Gbps.		0.8	0.5
4	Extend the new capabilities on a national basis.	Establish E2E test beds involving a mixture of DOE labs and university collaborators that can routinely fully utilize 2×10 Gbps bandwidth.		1.2	0.8
4.5	Gain experience at 40 Gbps.	Support R&D for E2E services beyond 40 Gbps.		0.8	0.2
5	40 Gbps capability.	Provide E2E high-impact services at 40 Gbps.		2.0	2.0

Internet 2, the special requirements of the university community can only be met by special requirements in their contract and their central management ofnon-typical services. Note, the Office of Science's networking and collaboratory requirements are more similar to those of Internet 2 than those of standard commercial IP services.

Consequently, this roadmap proposes that there is central management of the production and high-impact services and that there is central coordination of the research network services. The reason that the research network requires coordination rather than line management alone is that a number of elements needed for the research network will likely not necessarily be under the control of the Office of Science. For example, portions of the university community are beginning the National Lambda Rail (NLR) (www.getlight.net), which is envisioned to provide a mixture of dark fiber and lambda services. The Southeastern Universities Research Association (SURA) also is arranging access to similar resources (USAWaves) that may be joined to the NLR. The European community has several projects at the lambda level, and they are working with the Office of Science and university communities. In addition, there are several Office of Science-funded network elements for research into networking that have been kept separate from the current production ESnet so as not to interfere with ESnet. To achieve the networking R&D for the Office of Science goals for E2E connectivity, it is crucial that network research involve all of these communities and elements. So, portions of the research network will be managed by the central group and portions of the research network will come from these other groups. Coordination through the central group will be key for the success of the enterprise, including the critical step of moving results from the R&D into the production and high-impact networks as E2E services.

Funding for the networking R&D and the collaboratory/middleware components of the roadmap are provided in their respective sections and in the budget summary. The funding for production and high-impact networking is included in the table above. As is pointed out in Appendix B of this report, networking bandwidth requirements have been growing at a rate of 100% per year and, as is presented in Section 2 of the report, the networking bandwidth requirements are projected to continue to grow at this same rate or faster for the time period of this roadmap through 2008. Appendix B also notes that the cost for networking bandwidth is decreasing at a rate of 40% per year and is currently volatile. These effects are taken into account in considering the funding increments shown in the table. Consequently, the budget for production and high-impact networking will need to increase as indicated in the table so that the production and high-impact networks can keep pace with the demands for achieving the scientific discoveries of the Office of Science. The table below summarizes the incremental funding requirements from the more detailed table below:

Table 3-2 Incremental Funding Requirements for Production and High-Impact Networks

FY	Incremental Funding ($M)
1	3
2	4
3	4
4	4
5	5

The strategy for separating the current production network into production and high-impact components plus research networking was described above as being driven by the underlying technology determining how information is carried by light on fiber optic cables. There are additional motivations. In this model, less funding is required to achieve a given level of performance. This is due to two factors. Achieving

99% reliability is less expensive than achieving +99.9% reliability, so having a portion of the network at 99% reliability reduces costs. The second reason is that not every DOE lab and user site needs high-impact services. The switched-lambda services can be carefully crafted to place the networking resources at exactly those locations where they are most needed. This will also be true for optical packet-switching services as they become available. While this careful provisioning is also done for production services, having both production and high-impact services available allows for even finer budgetary tuning.

There is an additional feature of this model that assists with funding. When a given program has very specific networking requirements, the model is very adaptable so the program can directly fund these requirements yet achieve the cost savings of being part of the larger project. For example, it will often be relatively inexpensive to add an additional lambda to an existing given circuit or set of circuits, where it would be prohibitively expensive for the program to procure the services of a one-of-a-kind item.

The workshop participants agreed that the high-impact network should be centrally managed by the same group managing the production network. As discussed above, the DOE science research networking activities should in part managed by the central group and in general coordinated with the central group. This will be the most cost effective management model, and this is the only way to efficiently move R&D results in to the high-impact and production networks and to achieve the end-to-end performance required to achieve the scientific goals. Information on ESnet management and governance is presented in Appendix B and additional relevant information for the roadmap is provided in Section 6 on Management and Section 7 on Governance.

4. Technology and Services

4.1 Overview

Network technologies and services/middleware are needed to translate the potential of fast, functional networks into actual scientific breakthroughs by enabling easier and faster access to and integration of remote information, computers, software, visualization, and/or experimental devices, as well as interpersonal communication. Middleware makes it possible for an individual scientist or scientific community to address its application requirements by:

- Facilitating the discovery and utilization of scientific data, computers, software, and instruments over the network in a controlled fashion.
- Integrating remote resources and collaboration capabilities into local experimental, computational, and visualization environments.
- Averting (or diagnosing the cause of and mitigating) failures or poor performance in the distributed systems.
- Managing, in a community setting, the authoring, publication, curation, and evolution of scientific data, programs, computations, and other products.

Network technologies make it possible to establish and manage the required communication that is the foundation of the distributed science environment.

The vision is of a science environment in which integrated scientific theory, experiment, and simulation can fully interact and be integrated, thereby leading to faster convergence on producing scientific knowledge. Major scientific instrumentation systems, such as DOE Office of Science's synchrotron X-ray sources at ANL, BNL, LBNL, and SLAC, the gigahertz NMR systems at PNNL, the high-energy and nuclear physics particle accelerators at multiple DOE labs (BNL, FNAL, JLab, and SLAC), the neutron sources at ORNL, and numerous smaller facilities are all national user facilities in that they are available to scientific collaborators throughout the country and internationally. The Office of Science also collaborates in the development and use of major international facilities such as CERN's Large Hadron Collider. These are all sources of massive amounts of data, generated at high rates (hundreds of megabits/sec and more). All of this data requires complex analysis by scientific collaborations at the DOE labs and at hundreds of universities [1]. Ultimately the results of all of this experimental science must be compared and contrasted with theory, usually through the mechanism of simulation. Although we are just beginning to have the necessary infrastructure available, we have hints of the power of this vision in examples such as model-driven magnetic fusion experiments, accelerator controls, and simulations that play a key role in the real-time process of evaluating supernova candidates for observation and cosmology parameter determination, etc.

In order for all of the required resources to be used effectively in an integrated science cycle, where simulation, experiment, and theory interact, middleware for scheduling, access, and management mechanisms for all of the distributed resources involved are essential. For on-line experiments to be integrated with computing resources in general, and supercomputers in particular, the computing resources, storage resources, and interconnecting network performance must all be available simultaneously and reliably. This requires various new technologies in both networks and in the computing and storage resources to support the new demands for building virtual systems, such as ultrascale protocols, quality of service, reservation, co-scheduling, etc.

Furthermore, there must be comprehensive and easily used middleware that provides security, resource coordination, discovery, uniform access, etc. The rapidly emerging consensus is that the approach of computing and data grids [2] is the common approach to this sort of middleware, and the hundreds of people from around the

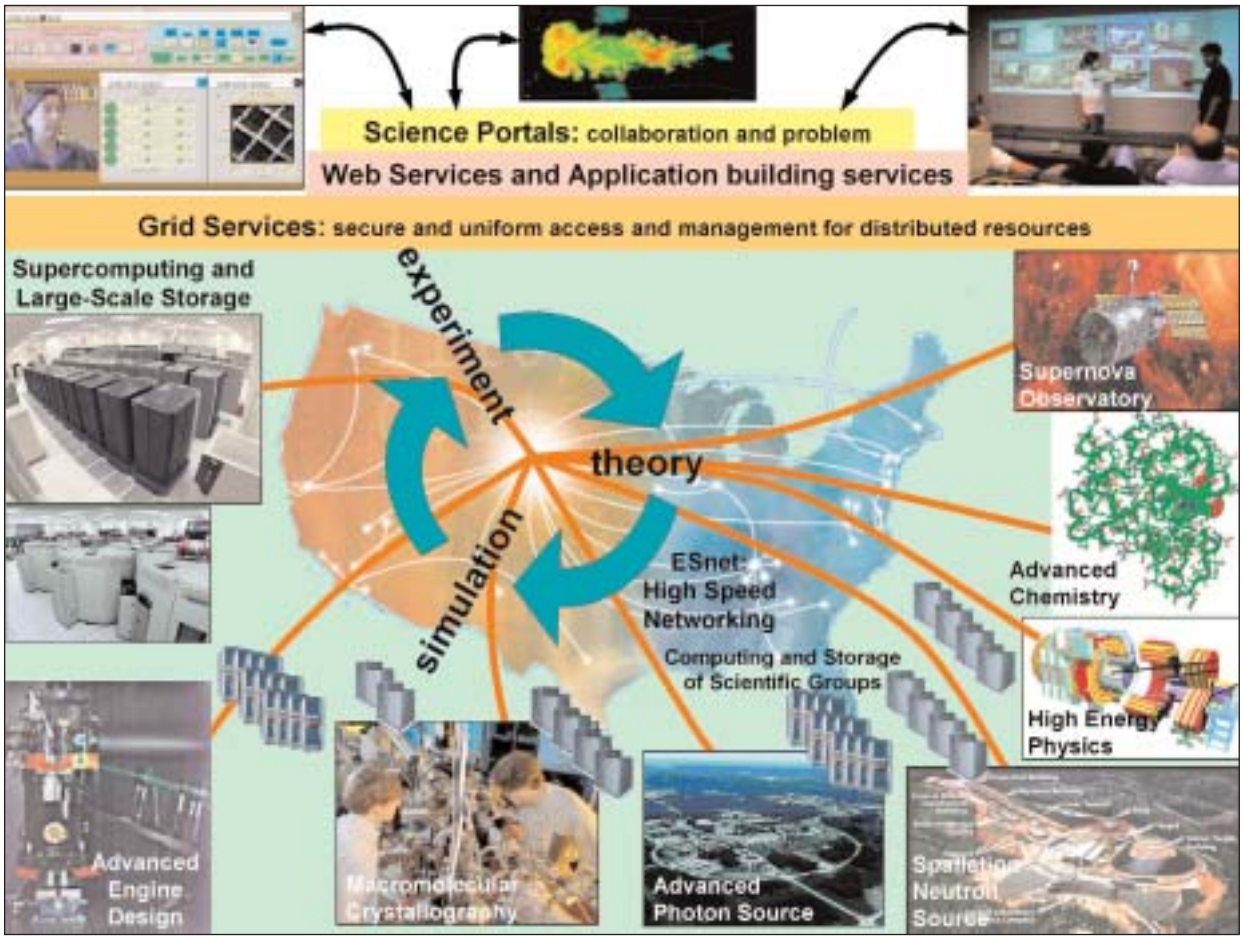


Figure 4-1 Integrated Cyber Infrastructure for Science

world who are working on grid standards and best practice at the Global Grid Forum [3] bear this out. This section of the report assumes that grids are being deployed as the standard middleware for distributed science, and looks at some of the issues that are important for making grids and distributed science environments successful in the DOE science environment.

4.2 Middleware Requirements for Science

Taking as an example of the future process of science in the climate community (as described in Section 2), the vision is to have the necessary computing power, access to annotated data, and interoperation of diverse sub-models, so that a realistic model of climate can make predictions that have great value to human society.

Considering this, together with the other science case studies from the August 2002 workshop (and summarized in Appendix A of this report), this future process of science generally is enabled by:

- Remote access to distributed computing, data, and collaboration tools in order to accommodate the fundamentally distributed nature of the science community.
- Computing capacity adequate for a task at the time the task is needed by the science — in particular, supercomputers incorporated into "virtual" systems, so that the simulations whose components run on supercomputers may integrate with the many different computing systems of the science community.
- Data capacity sufficient for the science task is provided independently of location, and

managed by information systems needed for building and maintaining knowledge bases and sharing them among disciplines.

- Software services providing a rich environment that gives scientists the ability to build multidisciplinary simulations in ways that are natural to the scientific process, rather than having to focus on the details of managing the underlying computing, data, and communication resources.
- Communication capacity and capability sufficient to support all of the aforementioned transparently to both systems and users.

The computing capacity requirements of science are absolutely essential components of such simulations and are being documented in other venues such as the Science Case for Large-Scale Simulation (SCaLeS) Workshop [4]. Data capacity is typically provided by the large managed archives associated with the country's supercomputer centers or by specialized scientific data centers (such as the multi-petabyte deep archive system at DOE's National Energy Research Scientific Computing Center – NERSC [5]) or by specialized scientific data centers.

Summarizing the general observations and conclusions from the August 2002 High-Performance Network Planning Workshop provides useful insights in the range and immediacy of the technologies and middleware services that are needed by the applications. (see also Appendix A)

The first and perhaps most significant observation was that a lot of science already is, or rapidly is becoming, an inherently distributed endeavor. Large-scale science experiments in the DOE environment involve collections of collaborators who frequently are multi-institutional, and the required data and computing resources are frequently even more widely distributed than the collaborators. Further, as scientific instruments become more and more complex (and therefore more expensive), they are increasingly being used as shared facilities with remote users. Even numerical simulation — an endeavor previously centered on one or a few supercomputers — is becoming a distributed endeavor. Such simulations are increasingly producing data of sufficient fidelity that the data is used in post-simulation situations — as input to other simulations, to guide laboratory experiments or to validate or calibrate other approaches to the same problem. This sort of science depends critically on an infrastructure that supports the process of distributed science.

The second observation is that every one of the science areas need high-speed networks and advanced middleware to couple, manage, and access resources such as the widely distributed high-performance computing systems, the many medium-scale systems of scientific collaborations, high data-rate instruments, and the massive data archives. Taken together, these elements and the networks that interconnect them are critical to next-generation science and to the formation and operation of highly interactive, large-scale scientific collaborations. That is, all of these elements are required to produce an advanced distributed computing, data, and collaboratory infrastructure for science that will enable breakthroughs at an accelerated rate, as a result of triggering paradigm shifts in how science itself is conducted. Paradigm shifts resulting from increasing the scale and productivity of science depend completely on such an integrated advanced infrastructure that is substantially beyond what we have today. Further, these paradigm shifts are not speculative; several areas of DOE science already are pushing the existing infrastructure to its limits while trying to move to the next generation of science. Examples include high-energy physics with its worldwide collaborations analyzing petabytes of data (described in the August 2002 High-Performance Network Planning Workshop) and the data-driven astronomy and astrophysics community that is federating the huge databases being generated by a new generation of observing instruments so that entirely new science can be done by looking at

all of the observations simultaneously (e.g., the National Virtual Observatory [6] illustrates this point very well. Specifically, see "New Science: Rare Object Searches" [7].)

A third observation was that there is considerable commonality in the services needed by the various science disciplines. This means that we can define a common infrastructure for distributed science, and this is the role of computing and data grids (see, e.g., Reference 8).

There are a set of underlying network technologies and services that are required for all of this, and these technologies and services are the topic of this section.

4.3 Technology and Service Requirements

To be effective, the grid middleware must be deployed widely and must have a set of services and infrastructure that supports it. The type of grid middleware described thus far provides the essential and basic functions for resource access and management. As we deploy these services and gain experience with them, it becomes clear that higher-level services also are required to make effective use of distributed resources; for example, generalized data management such as the virtualized data use in high-energy physics data analysis [9].

One example of such a higher-level service is the brokering functionality to automate building application-specific virtual systems from large pools of resources. Another example is collective scheduling of resources so that they may operate in a coordinated fashion. This is needed to allow, for example, a scientist to use a high-performance computing system to do real-time data analysis while interacting with experiments involving on-line instruments. It can also allow simulations from several different disciplines to be run concurrently, exchange data, and cooperate to complete a composite system simulation, as is increasingly needed to study complex physical and biological systems. Collective scheduling depends on the ability to co-schedule computing resources, usually through advance reservation capabilities in batch queuing systems and the ability to reserve bandwidth along the network paths connecting the resources. Advance network bandwidth is discussed in some detail below. Higher-level services also provide functionality that aids in componentizing and composing different software functions so that complex software systems may be built in a plug-and-play fashion. The current approach to these services leverages large industry efforts in Web services based on extensible markup language (XML) to integrate Web and grid services. This will allow the use of commercial and public domain tools such as Web interface builders and problem-solving environment framework builders to build the complex application systems that provide the rich functionality needed for maximizing human productivity in the practice of science. Much work remains, but the potential payoff for science is considerable.

However, there is also a collection of capabilities that are needed to support/enable grids. For example, it must be possible for the grid-based systems of collaborators to communicate as needed through site security. In addition, these communications will often require network quality of service (QoS) for bandwidth reservation or in support of real-time services. In both cases, grids can provide only the higher-level service that coordinates resource-level reservation. These capabilities are required in common by essentially all distributed science applications.

4.4 Technologies and Services

To identify the capabilities needed to support grids and grid-enabled distributed applications, the Technologies and Services Working Group started with the application requirements of the August, 2002 High-Performance Network Planning Workshop [10] and looked at what currently inhibits the use of grid middleware to provide a

complete and robust middleware environment.

Thirteen issues were identified and rank ordered by the working group as to their impact on the success of distributed science. The five top ranked issues were judged to be *essential* in order to establish and maintain the basic middleware infrastructure. The next three were rated as *very important*, and the remaining five *as very important* or *important*. The identified issues, in ranked order, are:

1. *Integrated grid and site security*. Neither grids nor any distributed science infrastructure will exist unless we solve the problem of authenticating and authorizing users and user processes with respect to collaborating site firewalls or other cyber security non-firewall solutions so access may be gained to resources.
2. *Collaboration services and shared environments*. Examples include shared desktops and applications, on-demand video conferencing that is easily used within the user's environment, remote control rooms, and shared file systems.
3. *Performance monitoring and problem diagnosis (cross domain)*. Debugging and performance tuning of distributed applications are essential and very hard. Effective tools will provide enhanced productivity gains and increased capabilities for users.
4. *Guaranteed network performance and delivery – network quality of service (QoS)*. To reliably connect remote instruments to supercomputers so that experiment and simulation can interact, or to efficiently and effectively interconnect application components that must run simultaneously, the ability to specify a required network performance (throughput, response time, etc.) at a given time is essential.
5. *Authorization*. We must have a robust way to specify who is trusted to do what across domains. This is essential for sharing remote data and resources.
6. *Namespace management*. A global managed namespace for scientific objects — data, text, programs, etc. — is essential in a widely distributed environment.
7. *Publish/subscribe portal (papers + data)*. A generalized facility that is reliability managed is needed as a repository of long-term data.
8. *Self-defensive networks and intrusion detection*. It will never be possible to both share distributed resources and keep out all unauthorized entities (hackers, etc.), so we must have more sophisticated intrusion detection systems that can mount an active defense against malicious intruders.
9. *Caching and computing in the network*. When large-scale datasets are analyzed by distributed communities, data must be free to gravitate to the most useful locations in the network to minimize data movement and/or maximize the ability to rapidly process the data and to maximize ease of access. This will require caching services in the network. Similarly, computing services in the network can help minimize data movement by analyzing and reducing data at its cached location.
10. *Multicast IP (the group communication protocol of the Internet)*. IP multicast is an important capability for creating scalable collaboration services. However, after several years of deployment, this service is still not stable, and there are competing alternatives. The service needs comprehensive diagnostic tools that allow problems to be easily analyzed and corrected.
11. *Allocation management*. Any time a limited or quality service is deployed (e.g., network-based caches and bandwidth reservation), an allocation management scheme is needed. This should be solved in a general way and provided as a grid-based service for whatever services require allocation management.
12. *Replica management*. As soon as network data caches are available, the most basic

service for managing such caches is a replica management service that keeps track of where data replicates are cached, which is a "master" copy, etc.

13. *Metadata catalogue*. A fundamental aspect of distributed data and services is to have robust catalogues that describe the data and services in sufficient detail that they may be located and characterized. Like most of the catalogue services in this list, there does not need to be a centralized catalogue — most scientific collaborations will maintain their own data — however, there needs to be a centralized catalogue of catalogues to tie everything together. This is analogous to the role of the root name servers in the Internet Domain Name System (DNS).

Detailed descriptions of the first 8 issues are provided in Appendix E along with their associated roadmap table.

Several critical issues are not addressed here because they were neither middleware nor network technology issues. In particular, developing technology for advance reservation scheduling of network resources, data sources, and computing systems, especially supercomputers, is a critical issue for building distributed virtual systems that include time-critical components, such as instruments, or require co-scheduling to work, such as distributed simulations.

4.5 Milestones and Metrics, Business and Operation Model, and Budget

In the Appendix D tables, we have tried to capture a realistic development and deployment schedule, the ongoing operational costs for the service, and a budget for both. Each of the tables lays out a roadmap from where we are today with the technology, through R&D, to an operational production capability, and finally the estimated on-going cost of operating the service to provide middleware infrastructure for DOE science. The metrics are:

- Meeting the roadmap milestones for the indicated costs, and
- Achieving utilization of the technologies and services by the DOE science community.

It is envisioned that R&D for these services will be mostly funded via the DOE MICS office and receive oversight from the field from the Science Networking and Services Committee (SNSC) discussed in Section 7 on governance. It is anticipated that as projects move into their pilot phase, they will have joint funding from the MICS office and the other Office of Science program offices participating in the pilots. When the services move into the production environment, it is envisioned that they will receive their ongoing funding from the MICS for the core services used by multiple collaboratory or grid activities. Individual collaboratories or grids will be funded by the program funding the specific scientific endeavor.

While the core services will need oversight by the SNSC and centralized management, the core services themselves may be provisioned either centrally or in a distributed model. For example, today ESnet provides centralized video conferencing and centralized PKI support. However, distributed support for some of the core services will be considered as the services evolve.

A summary of the roadmap for the critical five services is in the following table. Details of the capability development areas for the first 8 issues along with task plans and detailed budgets are given in Appendix D. The table below provides both the R&D budget required to produce each required service and the on-going operations budget required to sustain the service for the DOE science community.

Table 4-1 Milestone and Cost Summary for Technologies and Services

Technology/ Service	Goal	Major Milestones by Year from Start			
		R&D for Pilot Service	Pilot Service Rollout	Prototype Production Service	Deploy Production Service
Integrated Grid & Site Security	Authenticate and authorize users at site firewalls so that the many legitimate communication and data flows of scientific collaboration may easily gain access to collaborating sites.	3	3.5	4	4.5
Collaborative Services	Provide capable and easily used tools to facilitate widely dispersed scientific collaboration.	2	2.5	3.5	4
Performance Modeling	Greatly simplify finding and correcting the many performance problems that can occur in widely distributed applications in order to provide the high bandwidth, end-to-end communication needed by data-intensive science applications.	3	5	5.5	6
Network QoS	Mechanisms for specifying and reserving network bandwidth in order to enable connection of remote instruments to supercomputers so that experiment and simulation may interact, and to enable building distributed applications whose components must run simultaneously.	2.5	3	3.5	4
Authorization	Allow resource stakeholders to specify and enforce use conditions so that remote sharing of resources and data may happen transparently among authorized users.	2.5	3.5	4.5	5.5

Cost, $K			
R&D	Operation Cost to Full Production	Total to Deploy	Annual Operation Cost in Production
5,600	800	***6,400***	800
4,400	1,000	***5,400***	800
6,200	1,400	***7,600***	800
4,800	1,100	***5,900***	1,500
5,100	2,000	***7,100***	500

4.6 References

1. "Real-Time Widely Distributed Instrumentation Systems," In I. Foster and C. Kesselman, eds., *The Grid: Blueprint for a New Computing Infrastructure*, Morgan Kaufmann (1998).
2. See, e.g., Johnston, W.E., "The Computing and Data Grid Approach: Infrastructure for Distributed Science Applications," Computing and Informatics – Special Issue on Grid Computing (2002). Available at: http://www.itg.lbl.gov/~johnston/Grids/homepage.html#CI2002.
3. See: http://www.gridforum.org.
4. June 24-25, 2003, Washington, DC, and Seattle, WA. Available at: http://www.pnl.gov/scales.
5. See: http://hpcf.nersc.gov/storage/hpss/#whatishpss.
6. See: http://www.us-vo.org.
7. See: http://www.roe.ac.uk/wfau/nvo/sld006.htm
8. Foster, I., and Kesselman, C., eds., The Grid, 2nd edition, Morgan Kaufman (2003); and Berman, F., Fox, G., and Hey, T., eds., *Grid Computing: Making the Global Infrastructure a Reality*, John Wiley & Sons (2003).
9. See: GriPhyN – Grid Physics Network. Available at: http://www.griphyn.org.
10. DOE Office of Science, High-Performance Network Planning Workshop, Reston, VA, August 13-15, 2002. Available at: http://doecollaboratory.pnl.gov/meetings/hpnpw.

Historically, science communities have expected network capabilities to be readily available when needed, and generally they have been. However, in recent years the gap between ready-out-of-the-box, end-to-end (E2E) performance and theoretical performance (represented as exemplified by WAN speed records) has been widening. As was pointed out at the recent DOE Science Computing Conference [1], that gap has now reached three orders of magnitude, as shown in Figure 5-1. Coupled with the unprecedented requirements of large-science applications, this gap will be a severe bottleneck to the execution of the above-mentioned DOE projects. For instance, the leading operational link speeds are currently at the 10-Gbps level (OC192) but are available only at the backbone. At the applications end-hosts, the transport throughputs typically reach only a few tens of megabits, and routinely reach only a small number of hundreds of megabits with considerable effort and constant attention. Multiple Gbps can be reached with Herculean efforts from teams of network and application experts; these bandwidths, although widely touted, are ephemeral. Furthermore, bandwidth is not the only measure of needed performance. There are no technologies available in the current operational wide-area networks that will provide either the guaranteed stability needed for real-time steering and control over long-haul Internet connections or the agility needed for instantly redirecting massive visualization flows.

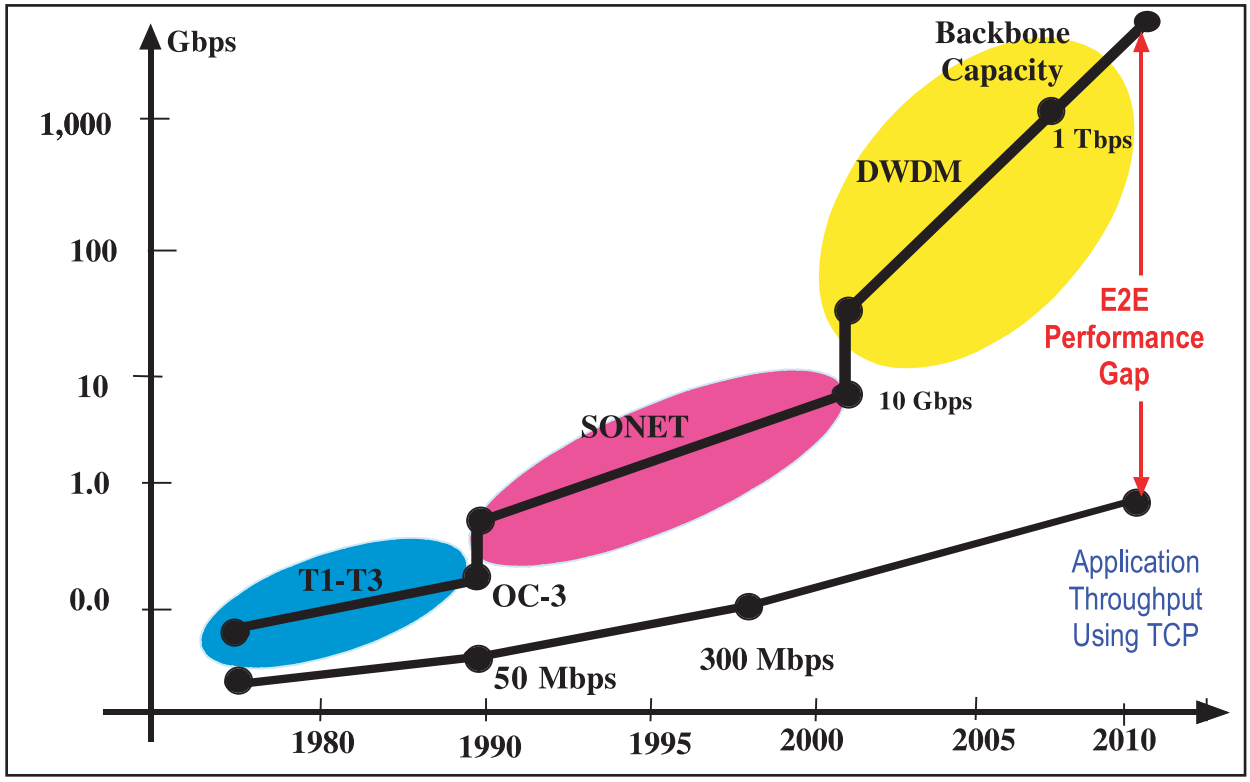


Figure 5-1 Performance Gap between Application Throughput and Optical Network Speeds

Generically, the needed network capabilities are quite specific to the above DOE projects both in terms of high throughputs and advanced network capabilities. While there are some overlaps with projects from other federal agencies, the totality of these requirements is unique to DOE. Internet2 and the commercial Internet are optimized for the academic and mass markets, respectively, to address the needs of a new generation of networked applications. Their focus areas include spatially dense and large-scale sensor networks, pervasive computing networks, wireless networks, application-level overlay networks, extensible networks, and emergency-response networks [2-4]. These topics do not address the acute requirements of DOE's science applications, which are not among the top priorities of industry and other agencies and are unattractive to industries.

ESnet has successfully provided leading-edge networking services for thousands of DOE scientists and their collaborators worldwide. However, the extreme requirements of the emerging science applications well exceed its current capabilities. Therefore, revolutionary advances in ESnet are required to support DOE's science mission and to take advantage of worldwide scientific resources such as the European data grid.

The overall goal of the research network efforts is to provide an integrated and stable environment to perform the development and testing of the required network technologies for real high-performance networks and science applications. The technologies supported by the research network include reliable transfer of terabyte-scale datasets at gigabits to terabits per second rates, interactive remote visualization of petabyte-scale datasets at 10-30 frames per second, interactive steering of remote computation and instruction in real time, remote real-time operation of large experimental facilities, and cyber security mechanisms with minimal impact on application performance. Each of these activities were discussed and developed at this workshop. *As noted above, these requirements in total are quite specific to DOE's science applications, and are not among the top priorities of other agencies and are less attractive to industries.*

5.1 Capabilities Developed under Research Network

The overall network requirements of DOE in general and the Office of Science in particular range from routine to extreme, and thus require an infrastructure that consists of a production and a high-impact network. In addition, a research network is essential for the research and development activities needed to meet the challenges of end-to-end network performance of large-science applications. In this section, we discuss the research topics that are to be supported by the research network.

The network capabilities to be developed using the research network to address DOE large-science needs include the following main categories:

1. Reliable and sustained transfers of terabyte-scale data at Gbps to Tbps rates,
2. Remote interactive and collaborative visualization of large datasets of petabyte scale,
3. Steering computations on supercomputers and experiments at user facilities,
4. Interactive collaborative steering of computations visualized through multiple perspectives, and
5. Securing scientific cyber environments with minimal impact on applications.

In particular, it is essential that these capabilities be transparently available to application scientists with little or no additional demands on their time and effort to utilize them. In particular, it is not very effective if these capabilities require sustained efforts from teams of network and application experts just to use them.

5.2 Research Network Technologies

To realize the above capabilities, several network research tasks will be carried out in conjunction with the existing technologies within the context of applications.

Ultra high-throughput transport protocols. The current dominant transport protocol, namely TCP, was originally designed and optimized for low-speed data transfers over the Internet. It lacks the performance and scalability to meet the above challenges. Over the years, DOE has been a leader in successfully enhancing TCP to meet the performance objective of science applications. Achieving high-throughput data transfers involves two distinct approaches. At one extreme, TCP methods on shared IP networks can be adapted and scaled to Gbps-Tbps rates. The challenges here include investigating various parts of TCP, such as sustained slow-start and robust congestion avoidance, to sustain the required throughputs. At the other extreme, one could provide dedicated high-bandwidth circuits from source to destination nodes wherein suitable rate-control methods can be used for transport. This approach avoids the complicated problem of optimizing TCP by avoiding the congestion. Particularly in the future when experiments generate petabytes of data, the transport protocols and network provisioning that can sustain Tbps throughputs will be needed. For such throughputs, dedicated circuit-switched lambdas that are dynamically provisioned require very innovative non-TCP transport protocol mechanisms to achieve Tbps throughput. Optical packet-switched networks are anticipated to follow the deployment of switched-lambda networks and will require R&D to be integrated into the Office of Science environment, starting in 2005-2006 so that they will be deployable beginning in 2008-2009.

Stable end-to-end control channels. For supporting interactive visualizations over wide-area networks, two channels are needed: a *visual channel* that transfers the image data from source to destination, and a *control channel* that transfers the control information from the user to the visualization server. The former channel must provide appropriate sustained data rates to present adequate visualization quality to the user, whereas the latter should provide a low jitter rate to avoid destabilizing the control loop. There are several possibilities for implementing the visual channels, ranging from transporting the geometry (for example, as OpenGL codes) to be rendered at the user locations to rendering at the host and just forwarding the visuals (for example, using xforwarding). In a particular application, a combination might be required based on the bandwidth needed for data and visualization pipelines. In either case, the throughput should be sustained to maintain adequate visual quality. From the network transport viewpoint, both of these channels require stable throughput, which can only be partially achieved over IP shared networks, typically in a probabilistic sense. On the other hand, they are easier to achieve if two dedicated channels can be provided on demand. Advances both in transport and provisioning methods would be required to achieve these capabilities.

Real-time collaborative control and data streams. Agile transport protocols are needed to control remote computations. Typically a computation is monitored remotely, perhaps by visualizing a partial parameter space, and steered into regions of interest. It is very important that the steering operations be supported on a robust channel to place the computation in an appropriate parameter region. Note that an inadequate control channel can result in undershoot or overshoot problems, which waste valuable computational resources, particularly on supercomputers. The control problem is more acute in remote control of experimental devices, where delays in control commands can result in severe damage. In an extreme case, jitter in high-frequency components can result in resonance, which could lead to a complete loss of control. Furthermore, when the steering or control operations are to be coordinated between multi-

ple users at geographically dispersed locations, the control channels must be suitably coordinated. Except for very simple steering and control operations, TCP on IP networks does not provide the desired stability levels. The approach based on dedicated channels together with associated transport methods would be investigated for this class of capabilities.

Dynamic provisioning and on-demand bandwidth assignment. The concept of dynamic provisioning is promising to enable many desirable networking capabilities for high-impact science applications, such as:

a) Dynamic allocation of coarse-grain bandwidth at the lambda level.

b) Sharing of optical paths, which has been traditionally been done at the packet level.

c) Dedicated optical paths with transport protocols multiplexing, i.e., the transport protocol is decoupled from the underlying core network.

d) Efficient utilization of expensive core network by high-impact application.

The requirement of providing an on-demand, dedicated optical channel requires allocation policies and implementations that are absent in packet-switched IP networks. The requests for dedicated optical channels (lambdas) will be sent by the science applications to central bandwidth *allocation servers*, which maintain the "state" of the network. Once the request is accepted, *implementation servers* will step up the optical channels, maintain them for the allocated duration, and then tear them down. During the allocation, the end systems can use transport protocols that optimize application performance as well as the dedicated optical path. Such a capability does not exist over IP networks, and is unlikely to be developed elsewhere, so it must be developed for this class of DOE large-scale applications. Note that the allocation servers must be capable of implementing higher-level policies for granting the requests as well as scheduling the circuits by maintaining the state of available bandwidth levels of various network links. In addition, suitable routing and switching hardware and software must be in place to enable on-demand set up, maintenance, and tear down of the various circuits.

Architecture and infrastructure issues. Due to the exploratory nature of the research network, it is very important to provide a ubiquitous monitoring and measurement infrastructure to assist in diagnosis, debugging, and performance optimization. Also, the research network architecture must have provisions for operating system (OS) bypass, remote data memory access (RDMA), and other non-conventional implementations for network technologies.

Cyber security issues. The DOE science environments consist of very complex and expensive distributed computing resources and science instruments that must be protected from cyber attacks. As recently evidenced, the proliferation of strict firewalls, particularly at DOE sites, rendered several network-based applications inoperable. In particular, several legacy applications that relied on open-socket communications simply stopped working, since firewalls by default denied the communications on general ports. While this problem can be temporarily fixed by port exceptions or moving hosts into open portions of the networks, it leaves them vulnerable to attacks. More systematic efforts are needed to provide graceful interoperation of science applications under secured network environments. Today's crude packet filters and firewalls have limiting effects on the data transmission rates, which in turn limit application throughputs.

Traditionally, the DOE science community has relied on commercial cyber security products to meet its security needs. In ultra high-speed network infrastructures, the capabilities of low-speed security systems (10 Mbps – 655 Mbps) such as firewalls and intrusion detection systems

are being seriously questioned. This problem calls for two major issues to be addressed:

1. *Scalable technologies for cyber security systems.* Today's cyber security systems are designed to protect networks running at low speed. Considering that the data rates of the emerging science environment will exceed several gigabits/sec and beyond, it is of particular importance to pay special attention to cyber security issues that may present major obstacles to scientific discoveries. It is of particular importance to develop and deploy intrusion detection, firewall, or alternative site security systems capable of operating at multi-gigabits/sec and higher. Furthermore, most intrusion detection and filtering methods have been developed for packet-switched networks. A new class of methods might be needed to secure the networks that provide on-demand, end-to-end dedicated paths. Optical packet-switched networks will require a further range of considerations.
2. *Cyber security policies for open scientific environment.* It is important to develop methods that are compatible with and conducive to overall DOE security policies in terms of expression and interpretation.

To summarize from a network research viewpoint, the technical areas to be investigated include:

- TCP and non-TCP transport methods capable of sustained throughput at 40 Gbps or higher rates under shared as well as dedicated long-haul network connections,
- Dynamic provisioning of end-to-end circuits with dedicated bandwidth for steering and control operations, and
- Cyber security methods capable of operating at extremely high speeds and also under dedicated paths.

In these areas, the tasks typically include leveraging existing methods and designing novel methods specifically optimized to these high-performance networks, as well as the tools that bring these capabilities readily to application users. These are highly specific to DOE large-scale applications; they need to be available within the next five years to fulfill various critical elements of DOE's mission. This is not the time scale being envisaged by either industry or other government agencies. The research will need to be done by DOE, and a research network with adequate components and development environments is essential to the design, implementation, testing, and transitioning of these capabilities to operational networks.

5.3 Research Network Environments: Transition to Production

A research network that incorporates the required state-of-the-art software and hardware networking components, including routers/switches, high-bandwidth long-haul links, protocols, and application interface modules, is essential to meeting the challenges of DOE needs. This network will enable the development of various network technologies based on long-haul high-bandwidth paths supported by powerful next-generation components in configurations unavailable in current networks and test beds.

The network technologies for high-throughput control, dynamic provisioning, interactive visualization and steering, and high-performance cyber security will be developed in gradual evolutionary stages through close interaction with application users. The applications can be executed directly on the research network by the users, and the mature technologies will be smoothly transitioned to applications and the production network. The preferred research network would be an adjunct to the production and high-impact networks so that controlled experiments can be conducted by switching the traffic into and from them for testing purposes. The activities include the following:

- Transfer network technologies to science applications through joint projects involving

network research and applications users;

- Transfer complete application solutions to production networks through projects that utilize the network research in combination with the production portion of ESnet;
- Develop technology for end-to-end solutions for applications using teams of joint network researchers, ESnet personnel, and application users; and
- Experimental network testing activities involving researchers from across the country and the continents.

5.4 Essential Office of Science Network R&D Summary List

In addition to the overall research areas, the following items must be given high priority:

1. 40+ Gbps transport solution.
2. Dynamic bandwidth allocation (scheduling, prioritization, sharing, etc.).
3. Ultra high-speed cyber security components and policies.
4. Ultra high-speed transport protocol engineering.
5. Network measurement, monitoring, and analysis at ultra high speed.
6. Ultra high-speed network modeling, simulations, and traffic engineering.

5.5 Milestones and Metrics

A progression of milestones in the order of increasing time frames can be identified:

1. *Immediate time frame (1-2 years)*. Utilizing existing network transport and provisioning technologies to deliver the best possible performance for various network tasks.
2. *Short-term research (2-3 years)*. Enhancing the existing network technology (but maintaining backward compatibility) to deliver ten times the current performance levels.
3. *Long-term research, development, and engineering (4-5 years)*. Developing radically new technologies to deliver capabilities orders of magnitude better than current performance.

The following metrics will be paid special attention in assessing the technologies developed on the research network.

1. End-to-end scaling.
2. Elimination of geographical barriers.
3. Enabling new modalities and technologies of research.
4. Ease of use by application users.
5. Level and impact of cyber security.

5.6 Roadmap

The roadmap of research network activities includes two classes of activities that are highly interdependent: (1) research network infrastructure and provisioning, and (2) network transport and application support. We list the roadmaps in each area separately for conceptual clarity, but note that the individual tasks are intimately tied. The DOE MICS office started funding a test bed that provides OC192 connectivity from ORNL to Sunnyvale, CA, with a lambda-switching capability. This test bed will be leveraged as a part of the infrastructure for the research network.

5.6.1 Infrastructure and Provisioning Activities

There are two different types of network services that will be supported by the research network infrastructure. First type is based on IP, which resembles the production network but with additional support for research activities, including access to intermediate routers and cyber security equipment. The second type provides end-to-end dedicated circuits by utilizing optical routers, switches, and provisioning platforms that are not found in current wide-area networks. Various

Table 5-1 Infrastructure and Provisioning Summary

Year	Tasks	Cost ($M)
1	Establish IP and lambda-switching infrastructures at 1-10 Gbps rates.	5
2	Establish IP and lambda-switching infrastructures at 10-40 Gbps rates.	5
3	End-to-end provisioning of circuits using a combination of WAN and LAN environments.	5
4	Multi-resolution provisioning of pools of IP and end-to-end circuits to application users' desktops.	5
5	On-demand provisioning of multi-resolution interacting IP and end-to-end circuits to users' desktops.	5

yearly tasks together with cost estimates is provided in Appendix E.

In first two years, the basic IP and lambda-switching backbone infrastructure will be established using multiple OC192 links to provide peak bandwidths of 40 Gbps. In the next year, this access will be extended through LANs to application end hosts. In the next two years, the sophistication of the provisioning will be enhanced to provide on-demand interacting pools of end-to-end circuits with multiple resolutions to users' desktops.

5.6.2 Network Transport and Application Support

The transport and application activities are designed to match the progression of infrastructure capabilities. Since the provisioned circuit infrastructure is planned approximately a year behind its IP counterpart, the activities of the former are suitably delayed in the roadmap. The yearly tasks for these activities are listed below, and a more detailed breakdown of each of these tasks together with cost estimates is provided in Appendix E.

Under each IP and circuit-switched provisioning, the tasks gradually progress from high-throughput transport methods to those involving collaborative interactive channels with multiple resolutions. The task sequence for IP networks precedes the other by approximately a year. While the tasks seem quite similar under both provisioning modalities, the resultant capabilities could be quite different; for example, the precision of control achieved on a dedicated circuit cannot be matched by that implemented over a shared IP connection. Furthermore, the technical aspects of various tasks vary significantly between the provisioning modes. For example, the transport on IP networks must deal with congestion which is a non-issue for dedicated circuits. On the other hand, the dedicated circuits must be requested, granted, and set up before the transport modules can be invoked, unlike the IP networks where the transport may start any time. From a user perspective, however, the functionality must be transparent of the provisioning mode. The application support efforts will concentrate on achieving such levels of transparency.

Table 5-2 Network Transport and Application Support Summary

Year	Tasks	Cost ($M)
1	• Develop TCP and non-TCP protocols for IP networks for Gbps throughputs using RDMA, OS bypass, and striping methods; • Execute and demonstrate multi-Gbps transfers under real application environments; and • Assess the effect of IP sniffers and firewalls on transport throughputs.	4 3 1
2	• Develop high-throughput transport protocols and associated RDMA, OS bypass, and striping technologies for switched-lambda circuits; • Develop and demonstrate transport protocols for remote visualizations over IP and switched sub-lambda circuits; and • Assess the effect of cyber security measures on protocols for both IP and provisioned circuits.	4 3 1
3	• Develop and demonstrate protocols for computational steering and instrument control over IP networks and switched-lambda circuits, and • Assess the impact of firewalls and packet sniffers on the applications requiring visualizations and remote control.	7 1
4	• Develop and test unified APIs for operating and optimizing collaborative pools of IP and provisioned circuits of multiple resolutions, and • Assess the impact of firewalls and packet sniffers on the applications requiring pools of IP and provisioned circuits for visualization and control.	7 1
5	• Develop and test modules to support interacting IP and dedicated channels for distributed interactive collaborations; and • Develop and demonstrate a complete solution suite for a large-science application that requires large data transfers, collaborative visualizations and steering, and interactive real-time control across the country.	3 5

5.7 Business and Operational Models

Several business models were considered for the research network at the workshop.

1. The first option is a single network logically partitioned into production, high-impact, and research networks. This offers the advantages of low deployment cost, flexible resource allocation, and seamless technology transition from development to deployment. Furthermore, by suitably provisioning the network, one can test research components under real production traffic at least for brief periods. The major disadvantages include high operational complexity and potential interference between various types of network traffic.
2. The second option is other extreme, to build and operate three separate physical networks through commercial contract(s). While this option has the advantages of low operational complexity and complete service isolation, it makes them non-conducive to transitioning new capabilities from research to production environments.
3. The third option is a hybrid approach consisting of a logically separate production and high-impact networks under a single infrastructure, which also provides flexible

provisioning of a research network. The advantages of this method are an uninterrupted production network and closer collaboration between network research and high-impact applications. Essentially, the research network becomes an adjunct to the production network but coexists with the production network so that high-end applications can be easily executed on the research networks. This option has the disadvantage of high initial investment.

Upon consideration, the third option best meets the operational needs of the research network.

The overall management of the research network will be governed by the Science Networking and Services Committee in terms of allocation of resources at the lambda level to various projects and institutions. In addition, there are additional lower-level allocation issues due to the on-demand provisioning aspects as well as the possibility of the network becoming unavailable as a result of experimentation and testing. Note that applications could request dedicated circuits or stable but shared connections at certain times. On the other hand, certain network research projects could push the network limits, possibly crashing the routers and/or hosts. These tasks will be scheduled on a demand basis. These lower-level, on-demand allocations on the research network will be governed by the allocation and scheduling committee, which decides the access policies for various network research and application communities both at a high-level and on a daily operational level. The policies at the daily operational level will be integrated into the scheduling and provisioning modules that will be developed in the second year. These policies will be periodically examined and updated as per the requirements of the active projects.

5.8 Research Network Budget Summary

The annual program budget for the research network and the associated activities is $13.0M. This budget includes two separate components along the lines of the roadmap. The first part pays for various high-bandwidth links and the associated routers and switches as well as the personnel at various institutions to support the research network. The second component provides funding for (a) network research projects that target various network technologies that address specific application needs, and (b) pilot applications involving teams of application scientists and network researchers.

1. *Ultra-scale research network infrastructure.* The yearly cost is $5.0M. This cost includes the link costs of $3.0M, the equipment cost (routers, switches, and hosts) of $1.0M, and personnel cost of $1.0M. See Appendix E for details on various options.
2. *Network transport and applications support.* The yearly cost is $8.0 M. About 10-15 network research projects will be supported, each at an annual budget of $400-500K each, for a total of $5.0M. There are expected to be 3-4 collaborative pilot application projects (each with yearly budget of approximately $1.0M) for a total cost of $3.0M.

5.9 References

1. DOE Science Computing Conference: The Future of High Performance Computing and Communications, June 19-20, 2003; http://www.doe-sci-comp.info.
2. NSF Workshop on Network Research Testbeds, October 17-18, 2002; http://gaia.cs.umass.edu/testbed_workshop. Dealt with developing networks with capabilities beyond the current ones. This workshop focused on broad issues not specific enough to encompass DOE large-science needs.
3. NSF ANIR Workshop on Experimental Infrastructure Networks, May 20-21, 2002; http://www.calit2.net/events/2002/nsf/index.html.
4. NSF CISE Grand Challenges in e-Science Work, December 5-6, 2001; http://www.evl.uic.edu/activity/NSF/index.html. Identified the cyber-infrastructure requirements, including networking technologies, to address the nation's science and engineering needs.

6. Management

Appropriate and effective management models will be essential to the success of the Science Networking and Services project as outlined and envisioned in this document. The success and longevity of the current ESnet program establish an unparalleled track record that attests to the importance of this aspect of the project. We propose to build and expand on this demonstrated approach. At the highest level, the governance of the project will be broadened and modified to encompass the additional stakeholders in this expanded project/program scope and is addressed in Section 7. Below that level, there will need to be additional consideration of how the various components will be managed. This is discussed in this section.

6.1 Management Models and Alternatives

The management model used today for ESnet is one of central funding, management, and operations. ESnet is a production backbone network, providing connectivity between DOE sites and also between those sites and the global Internet. ESnet also provides some initial services to support collaboratories and grids. National laboratories and other sites that directly connect to ESnet have complete management responsibility for the site's local area network (LAN) that connects the site's users and systems to ESnet, and this particular approach has proven to work very effectively over the years. There is probably no viable alternative to the approach of independent management of each site LAN by site personnel. However, there have been suggestions that alternatives to a centrally organized backbone network should be considered, and the trade-offs of two alternatives are briefly discussed below along with a review of the benefits of the current approach.

An analysis reveals a number of significant advantages to the current approach:

1. *Cost savings.* The common approach and volume purchasing allowed by a centralized approach gives very significant overall cost savings through the additional leveraging it provides in dealing with vendors for both hardware and communications services. The centralized approach allows networking resources to be efficiently allocated on an overall basis in support of the DOE mission, and also to be reallocated more easily as requirements change, for example, in support of a new initiative.
2. *Labor savings.* The overall manpower needed to support the 24×7 (24 hours per day, seven days per week) requirements of a production networking environment can be minimized, as a much smaller central staff is needed in comparison to each site having to separately meet this requirement. In addition, there are numerous operational WAN issues that are resolved on behalf of the entire ESnet community, rather than requiring each site to dedicate personnel to working on the same issues. In addition, as there is a resulting great commonality in technology, approach, hardware, etc., inherent in the common approach for all ESnet sites, there is resulting appreciable leveraging and saving of effort.
3. *Critical networking services for the DOE science mission.* Many current ESnet network functions and many of the additional services put forth in this report are not available from commercial ISP vendors. The current approach facilitates provisioning of these services in an optimal fashion for achieving the mission.
4. *Common technology.* The DOE science community is aggressive in its use of networking and is, in many cases, well served by the incorporation of leading-edge technology. A common backbone infrastructure makes such technology available to all users and sites, without the issues of incompatibilities between competing vendors or service providers.
5. *Security.* In the existing model, each site maintains responsibility for its internal cyber

security. However, a centrally managed backbone such as ESnet can provide additional protection. For example, during the recent SQL worldwide cyber attack, ESnet personnel worked through the night and well into the weekend to restrict access from infected sites to both the DOE community as well as the rest of the global Internet, in some cases shutting down access until on-site staff were able to respond.

6. *Corporate identity.* ESnet is well known and respected worldwide in the research and education networking community. A connection to ESnet by counterpart R&E networks around the world is understood to constitute a connection to the DOE scientific enterprise.

Two alternatives to the current ESnet approach are briefly considered below:

1. *Commercial ISP services.* In this model, each site on ESnet would become responsible for contracting with a commercial Internet service provider for network support, in place of the services of ESnet. This approach has a number of drawbacks:
   - Higher overall cost to DOE for networking support would result, since the saving through leveraging costs and effort as described above would be reduced or eliminated. Furthermore, while a centrally managed project such as ESnet is heavily motivated to continue finding new ways to reduce costs, a commercial provider is much less likely to be so motivated.
   - Network resources would no longer be dedicated to the DOE science community, but would be shared with and subject to impact from general public traffic. This would have a significant impact on performance and predictability. Commercial ISP networks do not set up systems to deal with the large data flows coming from and going to terabyte/petabyte-scale data sources. They are configured for millions of relatively low-demand network users. Current ESnet services such as multicast and policy routing would not be available.
   - It would be nearly impossible to manage, consolidate, or redirect networking resources on a DOE-wide basis; for example, to respond to an emerging initiative or a significant new direction in scientific endeavor.
   - Introduction of leading-edge technology would be much more difficult, as it would require an interoperable roll out of a desired technology among multiple ISPs in order to be broadly available to the DOE science community. Also, vendors are motivated to roll out technology on the basis on its profit potential, rather than its importance to the support of scientific research, which would potentially restrict or eliminate the introduction of technology of specific interest to DOE.
   - The user community currently served by ESnet enjoys the support of an operational staff that is both of excellent quality and dedicated solely to the DOE research community. It is clear that this would not be the case with a mixture of commercial ISP providers.
   - Many of the R&E interconnects of vital importance to DOE science are not available to commercial networks, a prominent example being Abilene, which serves the U.S. academic community.
   - It is also clear that some programs would opt to establish their own network along the lines of the ESnet model, rather than convert to commercial networking services, resulting in a further balkanization of the networking activities within DOE.
   - This approach would also essentially eliminate the existence of a central 24×7 operational staff that would form the foundation and home for many of the

services and technologies that will be needed in the future (see Section 4).

For the above reasons, this approach is not considered to be a preferred model for providing network services to the DOE science community.

2. *Dark fiber-based services.* In this model, ESnet would procure dark fiber on a nationwide basis and be responsible for "lighting" it (i.e., providing the optoelectronic equipment needed to make the fiber functional) and for long-term maintenance and operation. This is in place of the current model whereby the nationwide communications services needed by ESnet are procured from a "long-haul carrier."
   - This approach is basically a variation of the provisioning model and would not necessarily have an impact on the management model, but would clearly fit within a central management model (and perhaps demonstrate another advantage of central management — the ability to roll over to new technology on a community-wide basis with relative ease).
   - The major factor in favor of this approach is that once the initial investment has paid for the dark fiber and for lighting it, the cost for incremental upgrades (i.e., an additional wavelength) is relatively inexpensive. However, the initial investment can be very substantial. Also of concern is the fact that this is an untried approach and, given the very substantial up-front investment, could seriously jeopardize the conduct of scientific research, should it prove to be an erroneous approach, given that the cost of recovery would be substantial and probably very time consuming.
   - For these reasons, the approach will start on a regional scale with respect to production services to validate its viability. Note that the roadmap will probably include an assessment of this approach on the national scale via the research network component, where the risk can be much more readily tolerated.

6.2 Management of the Other Network Components

1. *High-impact.* This component of the new networking environment is intended to (a) provide state-of-the-art support for very high-end or high-impact applications and (b) provide a testing/hardening ground for the introduction of the next generation of networking and/or communications technology. The workshop suggested that it would almost surely not be cost effective to build a separate network to support high-impact activities and, furthermore, it would be very difficult, and perhaps unwise, to isolate those applications (see additional discussion in Section 5.7 under "Business and Operational Models"). In addition, the production networking staff must be integrally involved with the high-impact networking component to ensure a smooth transition to production status. The high-impact requirements should be met by phasing the next-generation technologies into ESnet in a manner compatible with and leading to full production quality of support. Accordingly, high-impact networking support would fall under the same management model as production networking.

2. *Research.* This component is intended to (a) support the network research activities within DOE/Office of Science and (b) explore the viability of emerging technologies. During workshop discussions, it was recommended that an example of a key service from the centrally managed organization (i.e., the equivalent of the current ESnet) would be that of "lambda provider;" i.e., the wavelength should be made available to the research community on a demand basis. Again, this component of service (i.e., bandwidth provisioning on demand) of the

research network activity would also fall under the same management model as above. However, the overall model for the research network would be one of distributed management, as the infrastructure and research would be under the control of the research staff(s). Note that the research network is anticipated to have significant interfaces with external research networking activities.

6.3 Transition from Research to Production Status

It is generally agreed that an ultimate goal of the research networking activity is to move the capabilities into the production and high-impact networking environments. Successfully doing so will require an integration of the efforts in the three categories and a smooth transition from one phase to the next. In some instances, it could be expected that the lines of separation between the three activities will be blurred, particularly as a given area matures and gradually moves to the next (more robust) category. This is discussed in somewhat more detail in Section 5.3 under "Research Network Environments: Transition to Production."

6.4 Management of Technology and Services

ESnet Management currently provides video conferencing and PKI central services for the Office of Science scientific researchers. As the full suite of technologies and services becomes available to researchers discussed in this report for support of collaboratories and grids, it is envisioned that there will need to be central management of the support for core services. Note that central management does not necessarily imply central location of all of these services. The central management will orchestrate the movement of technology and services from the R&D phase into the production phase for long-term support of the operational collaboratories and grids used by the Office of Science.

7. Governance for DOE Science Networking and Services

Computer networking by its nature has many stakeholders who need to be heard with regard to the organization and operation of the networks. Thus, the issue of governance is an important one when the mission and scope of the networks are being reconsidered.

7.1 Current Governance Model

Of the three kinds of network functionality discussed in this report, DOE presently provides only a production network, namely ESnet, and some independent elements of a research network. On the DOE side, the program is administered by a program manager within the MICS division of the ASCR program in the Office of Science. The network is procured and operated by a contractor, presently LBNL, with funding from MICS. Although this does not provide all of the functionality for research testing or even for the high-impact uses discussed in this report, it is generally felt to have been an outstanding program that has met even the advanced production requirements of the Office of Science to a very good approximation with limited funding.

The ESnet contractor is assisted by a committee representing the network's user programs, primarily the programs of the Office of Science. The ESnet Steering Committee (ESSC) is charged to:

- Document, review, and prioritize network requirements for all Office of Science programs.
- Review the ESnet budget as presented by MICS, evaluating that budget with respect to the prioritized network requirements.
- Identify network requirements that require further research.
- Establish performance objectives for ESnet.
- Propose innovative techniques for enhancing ESnet's capabilities.
- Advise ESnet management personnel at LBNL.

Members of the ESSC represent specific DOE program offices, and are appointed by those offices, at whose pleasure they serve. Length of service of an ESSC members varies and experience has shown that the committee benefits equally from the consistency of purpose and direction provided by long-term members and the innovations fostered by new representatives. MICS nominates an ESSC member to serve as committee chairperson, and the selection becomes final upon approval by a vote of the entire ESSC.

7.2 Discussion and Recommendations

A central question is whether the three network functionalities and associated services are funded and organized as multiple separate entities or in a more centralized way. In other parts of this report, technical and organizational arguments are made favoring joint operation of the network parts so that technology can flow smoothly from more advanced to more production-oriented networks. For the same reasons, we favor a centralized funding model, with one organization being responsible for three types of networks. Although most of the present funding comes from one office in DOE, we believe that the funding model within DOE should permit funds to come from multiple program offices when appropriate to fund a specific functionality and give program offices a sense of responsibility for the network(s) and the associated services.

The expanded functionality, and therefore expanded constituencies, of the new Science Networking and Services should lead to increased coordination within DOE. This will be true at one level within the ASCR/MICS organization, where programs in network research and the development of middleware and collaboratories will have an increased reliance on the DOE network as a place to test and demonstrate new capabilities. But it is also true that the other program offices of the Office of Science should have a means of coordination and input within the DOE organization, especially when they are explicitly providing some of the funding for the network or services to meet their own special needs. Thus, we recommend a Headquarters Working Group of people

from the program offices to coordinate the process of providing funding from multiple program offices and the allocation of the resulting resources. A headquarters committee like this is involved in the SciDAC program, in which funding is associated with all of the Office of Science programs. At least one member of this working group should attend Science Networking and Steering Services Committee meetings (see below).

Finally, we consider the process of providing community input and guidance into the operation of the network and associated services. Although the network and services, and therefore the stakeholder communities, are more varied if the network evolves as envisioned here, we see a system of direct representation of the user communities of the different Office of Science programs, as has been provided by the ESSC, as being required to extend the record of responsiveness to user communities into the new period. Although we think the new committee should be closely modeled on the ESSC, the model of representation at least would have to be changed somewhat. The ESSC representation comes in a fairly uniform way across the programs of the Office of Science. The system of functions outlined here will not draw its user community in such a uniform way from across Office of Science programs. Although the production and high-impact networks can be seen as a continuation of the present ESnet, with broad usage from all parts of the Office of Science, the research network function will have a narrower group of users, primarily but not exclusively drawn from the network research program in the MICS division. The high-impact network function, still different, will be of interest to specialized parts of several programs that have a need to transfer data in large quantities and/or quickly respond to real-time or quasi-real-time requirements. The users of central services in support of the deployment of collaboratories and grids have their own unique distribution.

To deal with these specialized requirements, we propose a committee structure based on the Science Networking and Services Committee (SNSC), which would be chosen to give broad representation across the Office of Science, but augmented by specialized representatives cognizant of the needs in the research, high-impact, and technology and service areas. The SNSC would form subcommittees to provide guidance specifically for each of the three network and service functions, with subcommittee chairs and perhaps vice chairs, but not the remainder of the membership, drawn from the SNSC.

The Headquarters Working Group would be charged to guide policy and coordinate DOE Science Networking and Services for the Office of Science, including specifying requirements for the performance and services of the different network and service components and interaction with the network providing organization(s) on appropriate strategies and trade-offs to make in seeking the optimum mix of network capabilities. The composition of the committee would be as follows:

- Two representatives from each Office of Science program office, with one chosen by the program office and one chosen by (hopefully from) the relevant federal advisory committee.
- One or two representatives from the network provider(s).
- Two or three additional representatives chosen by ASCR to provide input from the network research and middleware programs.

The SNSC could could have up to four subcommittees, including the three discussed above and a technical/site-oriented subcommittee like the present ESCC.

We note that the Science Networking and Services Committee will need expertise on the ways in which the programs will use the network and also on opportunities of working cooperatively with vendors of advanced networking infrastructure.

8. Budget Summary

The following table presents the total budget required to implement this roadmap. As discussed, some of the proposed budgets would be covered by existing budgets.

The FY03 MICS budget for ESnet, middleware, collaboratory pilots, and network research is $39M. In the totals in Table 8-1, $16.5M would come from the current ESnet budget and approximately half of the Technology and Services and Research Network R&D budgets, i.e., $3.5M and $4M, respectively, would come from the existing programs in these areas. A number of the current activities for middleware, collaboratory pilots, and network research lie outside of the activities described in this report. This means that current resources would cover approximately $16.5M + $3.5M + $4M = $24M of the $39.5M requirements in the first year. The additional $15.5M in new funding needed to support this roadmap would represent a 40% increase in the current budget of $39M for the first year. The increase for the fifth year would be 55% over current funding for these programs.

While a 40% increase in these programs for the first year ramping to a 55% increase in the fifth year would represent a substantial increase in these specific programs, these increases would represent a very small increase to the total Office of Science budget. Without these investments, the DOE science program would risk falling behind in scientific accomplishment because of not providing the scientific infrastructure which is becoming available to the rest of the world. With these infrastructure investments for networking and services, DOE will be equipped to keep its leadership position in world-class scientific discoveries. These infrastructure investments for networking and services will be critical elements in keeping DOE a leader in world-class scientific discoveries.

Table 8-1 Total Budget Required to Implement Roadmap ($M)

Year	Production & High-Impact	Technology and Services		Research Network		
	Operations*	R&D**	Operations	R&D***	Operations	Total
1	19.5	7		8	5	39.5
2	20.5	7	1	8	5	41.5
3	20.5	7	2	8	5	42.5
4	20.5	7	3	8	5	43.5
5	21.5	7	4	8	5	45.5

* This includes the current ESnet budget of $16.5M.

** Approximately half of the proposed R&D budget, i.e., $3.5M for Technologies and Services, would be covered by the current budget.

*** Approximately half of the proposed R&D budget, i.e., $4M for Research Network, would be covered by the current budget.

The following sections summarize the requirements of seven major programs or facilities in the Office of Science in order to provide some of the scientific motivation for high-performance networks and the middleware services that facilitate the use of those networks. This is a summary of the science driver case studies in the August 2002 workshop, and additional material is available in the report of that workshop. Note that most of these science problems require an integrated infrastructure that includes large-scale computing and data storage, high-speed networks, and middleware to integrate all of these.

A.1 Climate Modeling Requirements

We need better climate models to better understand climate change. Climate models today are too low in resolution to get some important features of the climate right. We need better analysis to determine things like climate extremes (hurricanes, droughts and precipitation pattern changes, heat waves and cold snaps) and other potential changes as a result of climate change. Over the next five years, climate models will see an even greater increase in complexity than that seen in the last ten years. The North American Carbon Project (NACP), which endeavors to fully simulate the carbon cycle, is an example. Increases in resolution, both spatially and temporally, are in the plans for the next two to three years. A plan is being finalized for model simulations that will create about 30 terabytes of data in the next 18 months.

These studies are driven by the need to determine future climate at both local and regional scales as well as changes in climate extremes — droughts, floods, severe storm events, and other phenomena. Over the next five years, climate models also will incorporate the vastly increased volume of observational data now available (and even more in the future), both for hindcasting and intercomparison purposes. The end result is that instead of tens of terabytes of data per model instantiation, hundreds of terabytes to a few petabytes (10^{15}) of data will be stored at multiple computing sites, to be analyzed by climate scientists worldwide. Middleware systems like the Earth System Grid and its descendents must be fully utilized for scientific analysis and to disseminate model data. Additionally, these more sophisticated analyses and collaborations will demand much greater bandwidth and robustness from computer networks than are now available.

As climate models become more multidisciplinary, scientists from fields outside of climate studies, oceanography, and the atmospheric sciences will collaborate on the development and examination of climate models. Biologists, hydrologists, economists, and others will assist in the creation of additional components that represent important but as yet poorly known influences on climate. These models, sophisticated in themselves, will likely be run at computing sites other than where the parent climate model was executed. To maintain efficiency, data flow to and from these collaboratory efforts will demand extremely robust and fast networks.

An example of results obtainable from complex climate models is shown in the Figure A-1. Models from multiple groups, including atmospheric, terrestrial, and oceanographic researchers, were linked to achieve the results. Climate data at both NCAR and NERSC were involved, along with researchers from multiple laboratories and universities.

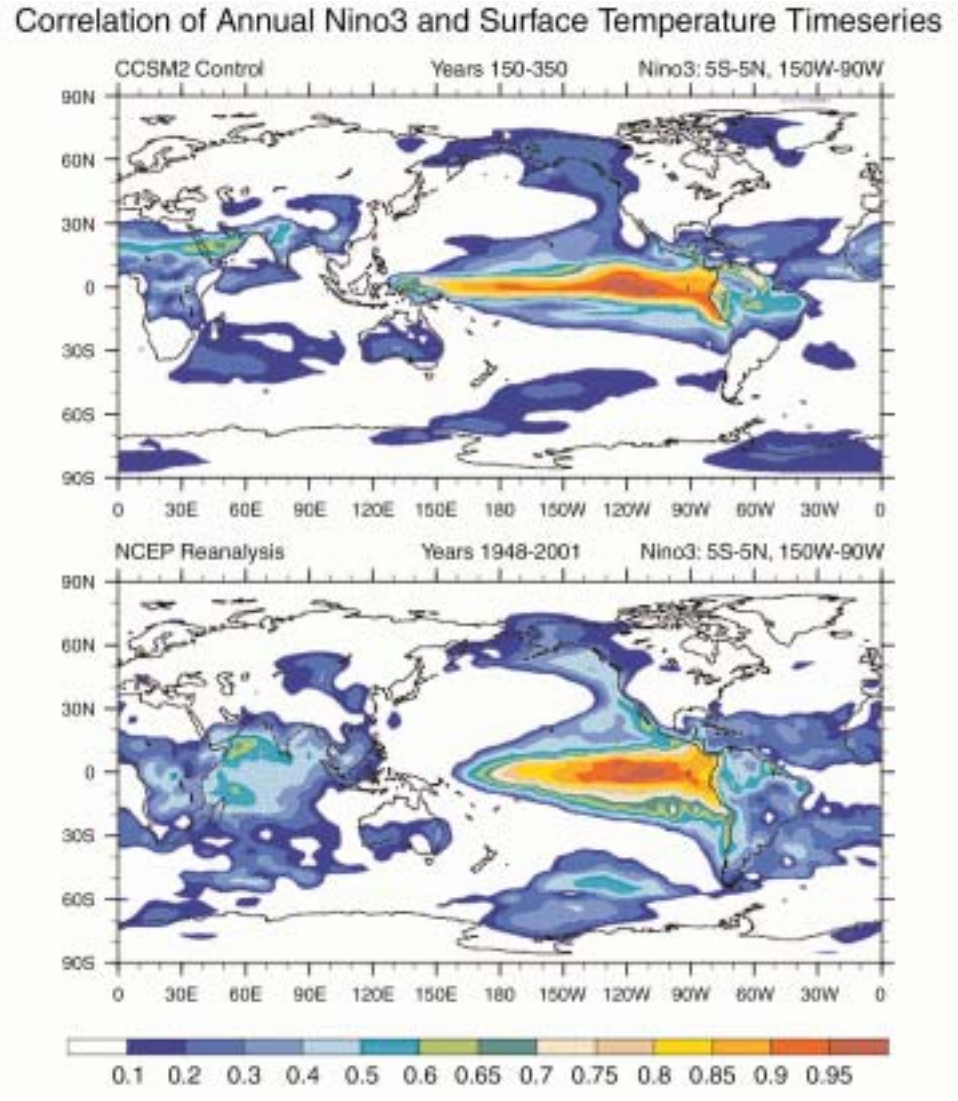


Figure A-1 Two hundred years of modeling El Niño events and surface temperatures on the Community Climate System Model (CCSM2) required more than 200 uninterrupted days on a supercomputer at NERSC. CCSM2 tightly couples four complex models, including atmosphere and land modeling codes.

In the period five to ten years out, climate models will again increase in resolution, and many more fully interactive components will be integrated. At this time, the atmospheric component may become nearly mesoscale (commonly used for weather forecasting) in resolution. Climate models will be used to drive regional-scale climate and weather models, which require resolutions in the tens to hundreds of meters range, instead of the hundreds of kilometers resolution of today's Community Climate System Model (CCSM) and Parallel Climate Model (PCM). There will be a true carbon-cycle component, wherein models of biological processes will be used, for example, to simulate marine biochemistry and fully dynamic vegetation. These scenarios will include human population change, growth, and econometric models to simulate potential changes in natural resource usage and efficiency. Additionally, models representing solar processes will be integrated, to better simulate incoming solar radiation. Climate models at this level of sophistication will likely be run at more than one computing center in distributed

Figure A-2 Spallation Neutron Source Facility at Oak Ridge National Laboratory.

fashion, which will demand extremely high-speed and tremendously robust computer networks to interconnect them. Model data volumes could reach several petabytes, which is a conservative estimate, and this must be distributed to an even larger number of collaborators.

A.2 Spallation Neutron Source Requirements

Six DOE laboratories are partners in the design and construction of the Spallation Neutron Source (SNS), a one-of-a-kind facility at Oak Ridge, Tennessee, that will provide the most intense pulsed neutron beams in the world for scientific research and industrial development. When completed in early 2006, the SNS will enable new levels of investigation into the properties of materials of interest to chemists, condensed matter physicists, biologists, pharmacologists, materials scientists, and engineers, in an ever-increasing range of applications.

The SNS (Figure A-2) supports multiple instruments that will offer users at least an order of magnitude performance enhancement over any of today's pulsed spallation neutron source instruments. This great increase in instrument performance is mirrored by an increase in data output from each instrument.

In fact, the use of high-resolution detector arrays and supermirror neutron guides in SNS instruments means that the data output rate for each instrument is likely to be close to two orders greater than a comparable U.S. instrument in use today. This, combined with increased collaboration among the several related U.S. facilities, will require a new approach to data handling, analysis, and sharing.

The high data rates and volumes from the new instruments will call for significant data analysis to be completed off-site on high-performance computing systems. High-performance network and distributed computer systems will handle all aspects of post-experiment data analysis and the approximate analysis that can be used to support near real-time interactions of scientists with their experiments.

Each user is given a specific amount of time (0.5 to 2 days) on an instrument. The close to real-time visualization and partial-analysis capabilities, therefore, allow a user to refine the experiment during the allotted time. For the majority of SNS user experiments, the material or property being studied is novel, and this capability is essential for the experimentalist to focus in on the area of interest and maximize the science accomplished in the limited amount of beam time.

In this scenario, the combined data transfer between the 12 SNS instruments and a distributed computer network for real-time data mapping is estimated to be a constant 1 Gbps, including visualization.

It is anticipated that analysis of experimental data in the future may be achieved by incorporating a scattering law model within the iterative response function extraction procedure. These advanced analysis methods are expected to require the use of powerful off-site computing systems, and the data may transit the network several times as the experiment/experimenter/simulation interaction converges to an accurate representation.

A.3 Macromolecular Crystallography Requirements

Macromolecular crystallography is an experimental technique that is used to solve structures of large biological molecules (such as proteins) and complexes of these molecules. The current state-of-the-art implementation of this technique requires the use of a source of very intense, tunable, X-rays that are produced only at DOE's large synchrotron radiation facilities at ANL, BNL, LBNL, and SSRL. In the United States, 36 crystallography stations are distributed among the synchrotron facilities and dedicated to macromolecular crystallography. The high operating cost of these facilities, coupled with the heavy demand for their use, has led to an emphasis on increased productivity and data quality that will need to be accompanied by increased network performance and functionality.

The data acquisition process involves several interactive on-line components, data archiving and storage components, and a compute-intensive off-line component. Each component has associated networking requirements. On-line process control and data analysis are real-time, interactive activities that monitor and coordinate data collection. They require high-bandwidth access to images as they are acquired from the detector. On-line data analysis now is limited primarily to sample quality assurance and to the data collection strategy. There is increasing emphasis on expanding this role to include improved crystal scoring methods and real-time data processing to monitor sample degradation and data quality. On-line access to the image datasets is collocated and could make good use of intelligent caching schemes as a network service.

High-performance networking can play several roles in on-line control and data processing. For example, at the Brookhaven National Laboratory National Synchrotron Light Source, several approaches to remote, networked, collaboratory operation are anticipated. The datasets most often are transferred to private institutional storage, and this requirement places a large burden on the data archiving process that transfers the data between on-line and off-line storage units. Current requirements for the average data transfer rate are 10 to 200 Mbps per station. It is

expected that in five to ten years, this will increase by an order of magnitude to 100 Mbps to 2 Gbps per station. This is further exacerbated by the fact that most research facilities have from four to eight stations. This places a future requirement of 320 Mbps to 16 Gbps per facility. Advanced data compression schemes may be able to reduce this somewhat, but the requirement is still for Gbps of network bandwidth.

In addition to increased raw network bandwidth, the next-generation high-performance networking infrastructure will need to provide tools and services that facilitate object discovery, security, and reliability. These tools are needed for low-latency applications such as remote control as well as high-throughput data-transfer applications such as data replication or virtual storage systems.

A.4 High-Energy Physics Requirements

The major high-energy physics experiments of the next twenty years will break new ground in our understanding of the fundamental interactions, structures, and symmetries that govern the nature of matter and space-time. The largest collaborations today, such as the CMS and ATLAS collaboration, are building experiments for CERN's Large Hadron Collider (LHC) program and encompass 2,000 physicists from 150 institutions in more than 30 countries. These collaborations represent the major U.S. involvement in high-energy physics experiments for the next decade, and include 300 to 400 physicists in the United States, from more than 30 universities, as well as the major U.S. high-energy physics laboratories.

The high-energy physics problems are the most data-intensive known. The current generation of operational experiments at SLAC (BaBar) and FNAL (D0 and CDF), as well as the experiments at the Relativistic Heavy Ion Collider (RHIC) program at Brookhaven National Laboratory, face many data and collaboration challenges. BaBar in particular already has accumulated datasets approaching a petabyte (10^{15} bytes). These datasets will increase in size from petabytes to exabytes (1 exabyte = 1000 petabytes = 10^{18} bytes) within the next decade. Hundreds to thousands of scientist-developers around the world continually develop software to better select candidate physics signals, better calibrate detectors, and better reconstruct the quantities of interest. The globally distributed ensemble of facilities, while large by any standard, is less than the physicists require to do work in a fully creative way. There is thus a need to solve the problem of managing global resources in an optimal way to maximize the potential of the major experiments for breakthrough discoveries.

Collaborations on this global scale would not have been attempted if the physicists could not plan on excellent networks to interconnect the physics groups throughout the life-cycle of the experiment and to make possible the construction of grid middleware with data-intensive services capable of providing access, processing, and analysis of massive datasets. The physicists also must be able to count on excellent middleware to facilitate the management of worldwide computing and data resources that must all be brought to bear on the data analysis problem of high-energy physics.

To meet these technical goals, priorities have to be set, the system has to be managed and monitored globally end-to-end, and a new mode of "human-grid" interactions has to be developed and deployed so that the physicists, as well as the grid system itself, can learn to operate optimally to maximize the workflow through the system. Developing an effective set of trade-offs between high levels of resource utilization and rapid turnaround time, plus matching resource usage profiles to the policies of each scientific collaboration over the long term, present new challenges (new in scale and complexity) for distributed systems.

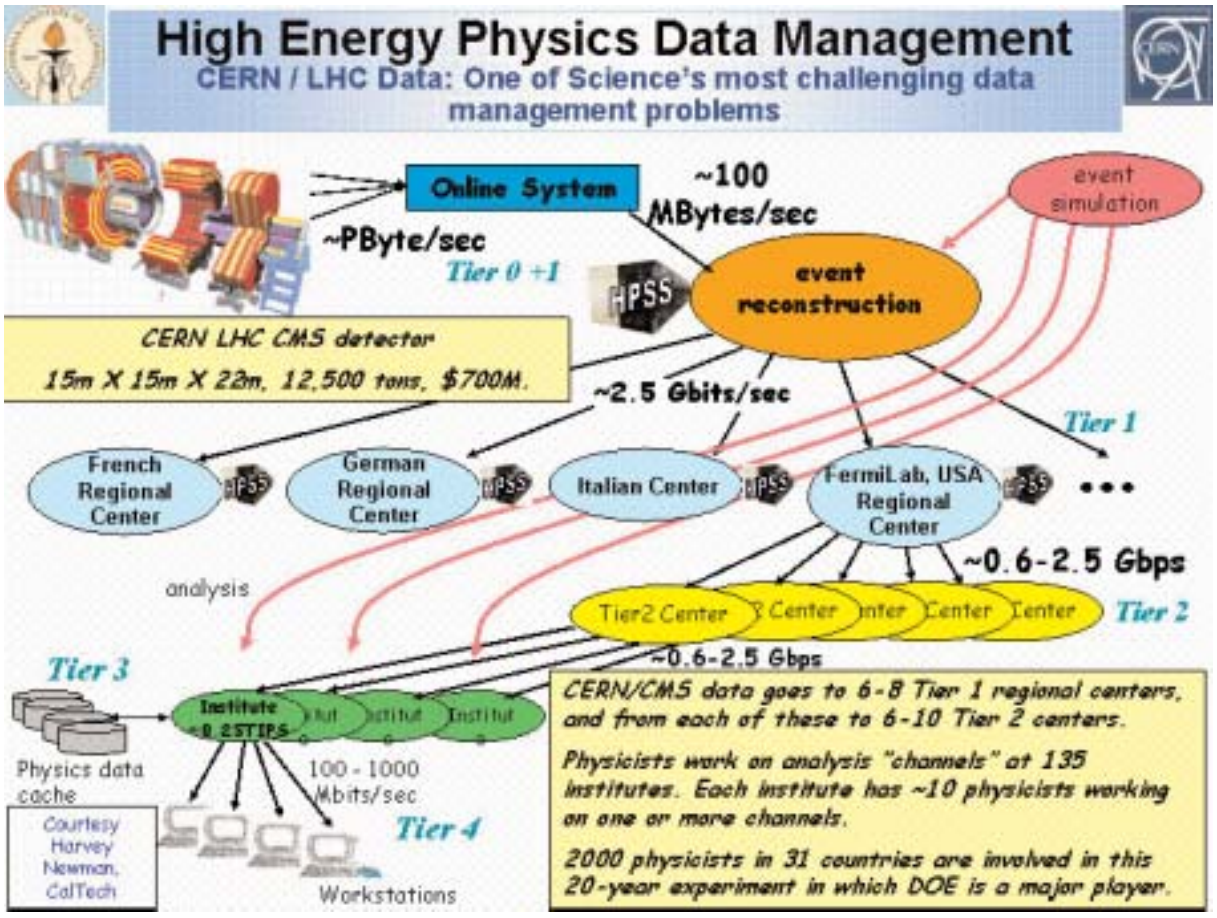


Figure A-3 A Hierarchal Data Grid as Envisioned for the Compact Muon Solenoid Collaboration. The grid features generation, storage, computing, and network facilities, together with grid tools for scheduling, management, and security.

A.5 Magnetic Fusion Energy Sciences Requirements

The long-term goal of magnetic fusion research is to develop a reliable energy system that is environmentally and economically sustainable. To achieve this goal, it has been necessary to develop the science of plasma physics, a field with close links to fluid mechanics, electromagnetism, and nonequilibrium statistical mechanics. The highly collaborative nature of the fusion energy sciences (FES) is due to a small number of unique experimental facilities and a computationally intensive theoretical program that are creating new and unique challenges for computer networking and middleware.

In the United States, experimental magnetic fusion research is centered at three large facilities (Alcator C–Mod, DIII, and NSTX) with a present-day replacement value of over $1 billion. Magnetic fusion experiments at these facilities operate in a pulsed mode, producing plasmas of up to 10 seconds duration every 10 to 20 minutes, with 25 to 35 pulses per day. For each plasma pulse, up to 10,000 separate measurements versus time are acquired, representing several hundreds of megabytes of data. Throughout the experimental session, hardware/software plasma control adjustments are discussed by the experimental team and made as required to optimize the use of the available experiment time. The experimental team is typically 20 to 40 people, with many par-

ticipating from remote locations. Decisions for changes to the next plasma pulse are informed by data analysis conducted within the roughly 15-minute-between-pulse interval. This mode of operation requires rapid data analysis that can be assimilated in near real time by a geographically dispersed research team.

The computational emphasis in the experimental science area is to perform more, and more complex, data analysis between plasma pulses. Five years from now, analysis that is today performed overnight should be completed between pulses. It is anticipated that the data available between pulses will exceed the gigabyte level within the next five years. During an experimental day, anywhere from 5 to 10 remote sites can be participating. Datasets generated by these simulation codes will exceed the terabyte level within the next three to five years. Additionally, these datasets will be analyzed in a manner analogous to experimental plasmas to which comprehensive comparisons will need to be made.

Enhanced visualization tools now being developed will allow this order of magnitude increase to be effectively used for decision making by the experimental team. Clearly, the movement of this quantity of data in a 15- to 20-minute time window to computational clusters, to data servers, and to visualization tools used by an experimental team distributed across the United States and the sharing of remote visualizations back into the control room will require much more network bandwidth and middleware than is available today.

In fusion, the need for real-time interactions among large experimental teams and the requirement for interactive visualization and processing of very large simulation datasets are particularly challenging. Some important components that will help to make this possible include easy-to-use and easy-to-manage user authentication and authorization framework, global directory and naming services, distributed computing services for queuing and monitoring, and network quality of service (QoS) in order to provide guaranteed bandwidth at particular times or with particular characteristics in support of analysis during experiments.

A.6 Chemical Sciences Requirements

The chemistry community is extensive and incorporates a wide range of experimental, computational, and theoretical approaches in the study of problems, including advanced, efficient engine design; cleanup of the environment in the ground, water, and atmosphere; the development of new green processes for the manufacture of products that improve the quality of life; and biochemistry for biotechnology applications including improving human health.

To overcome current barriers to collaboration and knowledge transfer among researchers working at different scales, a number of enhancements must be made to the information technology infrastructure of the community:

- A collaboration infrastructure is required to enable real-time and asynchronous collaborative development of data and publication standards, formation and communication of interscale scientific collaborations, geographically distributed disciplinary collaboration, and project management.
- Advanced features of network middleware are needed to enable management of metadata, user-friendly work flow for Web-enabled applications, high levels of security especially with respect to the integrity of the data with minimal barriers to new users, customizable notification, and Web publication services.
- Repositories are required to store chemical sciences data and metadata in a way that preserves data integrity and enables Web access to data and information across scales and disciplines.
- Either tools now used to generate and analyze

data at each scale must be modified or new translation/metadata tools must be created to enable the generation and storage of the required metadata in a format that allows an interoperable workflow with other tools and Web-based functions. These tools also must be made available for use by geographically distributed collaborators.

- New tools are required to search and query metadata in a timely fashion and to retrieve data across all scales, disciplines, and locations.

Advanced computing infrastructure that is being developed will revolutionize the practice of chemistry by allowing scientists to link high-throughput experiments with the most advanced simulations. Chemical simulations taking advantage of the soon-to-come petaflop architectures will enable chemists to guide the choice of expensive experiments and reliably extend the experimental data into other regimes of interest. The simulations will enable researchers to bridge the temporal and spatial scales from the molecular up to the macroscopic and to gain novel insights into the behavior of complex systems at the most fundamental level. For this to happen, they will need to have an integrated infrastructure including high-speed networks, vast amounts of data storage, new tools for data mining and visualization, modern problem-solving environments to enable a broad range of scientists to use these tools, and, of course, the highest-speed computers with software that runs efficiently on such architectures at the highest percentages of peak performance possible.

A.7 Bioinformatics Requirements

The field of computational biology, particularly that of bioinformatics, has undergone explosive growth since the first gene-sequencing work emerged in the mid 1980s. The understanding of biological processes, the ability to model them, and the ability to organize information and develop algorithms also have progressed rapidly. The field is now transitioning to a stage where algorithmic progress has outpaced computing capabilities in terms of raw compute cycles, storage, and especially fast, secure, and usable information discovery and sharing techniques. These factors limit progress in the field.

Applications that dominate today's computing requirements in bioinformatics include genome sequence analysis, pairwise alignment, computational phylogenetics, coupling of multiple model levels to determine metabolic pathways, and secondary database searching. On the more distant research horizon, research areas include sequence structure-function prediction, computation of the genotype-phenotype map, protein folding, molecular computing, genetic algorithms, and artificial intelligence solutions that will require real-time harnessing of grid resources for large-scale parallel computation.

Although the networking requirements of computational biology have much in common with other areas of computational science, they differ substantially in the aspects described in the remainder of this section. It is to be noted that some of these differences are of a quantitative nature, while others are qualitatively unique to the characteristics of the information bases and algorithms that make up the field.

The growth of the number of researchers involved in computational biology is outpacing that of almost any other biomedical science. This necessitates highly effective solutions for authentication and authorization of grid access; policy-based control and sharing of grid-based resources; and automated management of individual logins at large numbers of grid sites. National and international research communities will also need to construct virtual organizations, resource allocation policies, and charging mechanisms that span grid providers, because bioinformatics grids have different funding sources (ranging from state funds in North Carolina and Michigan, to federal R&D programs, to foreign

funds in the European Union and Japan).

While genomic databases of the past decade were sized in gigabytes, today's databases are pushing terabytes and growing roughly according to Moore's Law — doubling approximately every 18 months, with petabyte applications well within view. Performing grid computation on relational data will require the integration of heterogeneous databases to form worldwide federations of unprecedented scale. In addition, database replicas will need to be maintained accurately and synchronized with high integrity as huge amounts of data are exchanged. Significant research will be required in distributed database replication and grid-wide database mining applications to meet the federation and performance requirements of bioinformatics.

One of the most important collaborative activities in bioinformatics today is that of annotation, which would be greatly enhanced by the integration of multiparty messaging technologies with database versioning techniques, possibly augmented by being multicast with closely integrated file transport and visualization. This requires enhancements to network data transport protocols and QoS mechanisms. Collaborative imaging systems for use in the life sciences will involve both shared exploration and annotation of ultra-high-resolution images by multiple distant collaborators, coupled with computational-intensive pattern recognition, that require real-time transport of large image data.

A.8 Summary

These examples represent an important, if not major, set of DOE Office of Science science discipline areas. All of these case studies indicate, many individually and certainly taken together, the critical need for very high-speed networking and advanced middleware to create, together with high-performance computing, an integrated cyber infrastructure for DOE science networking. This is the motivation of this roadmap.

B.1 A Success to Date, a Challenge for the Future

This appendix gives an overview of DOE networking to date, with a particular focus on the ESnet project. It will show that the project has been quite successful since its inception in 1985 in meeting the networking and collaboration services support required by DOE science. However, this report is intended to also show that a substantial change in almost all components of the networking and collaboration environment in DOE will be required if this record of success is to continue and agency scientific research is to continue unimpeded. The approach is to build upon past success where appropriate and initiate changes where required to meet the future requirements. The changes envisioned will impact virtually all the areas discussed below, including funding, technology, architecture, services, and governance.

B.2 Project Overview

ESnet is a wide area network (WAN) infrastructure project that supports the scientific research mission of the U.S. Department of Energy. The ESnet project/investment supports the agency's mission, strategic goals, and objectives by providing DOE with an interoperable, effective, and reliable communications infrastructure and leading-edge network services.

ESnet is engineered and operated by ESnet networking staff located at Lawrence Berkeley National Laboratory (LBNL), in Berkeley, California. ESnet activity is guided by the ESnet Steering Committee (ESSC), with one or more representatives appointed by each of the five Office of Science programs and with additional representation from the DOE Defense Programs (DP) and Human Resources (HR). The ESnet

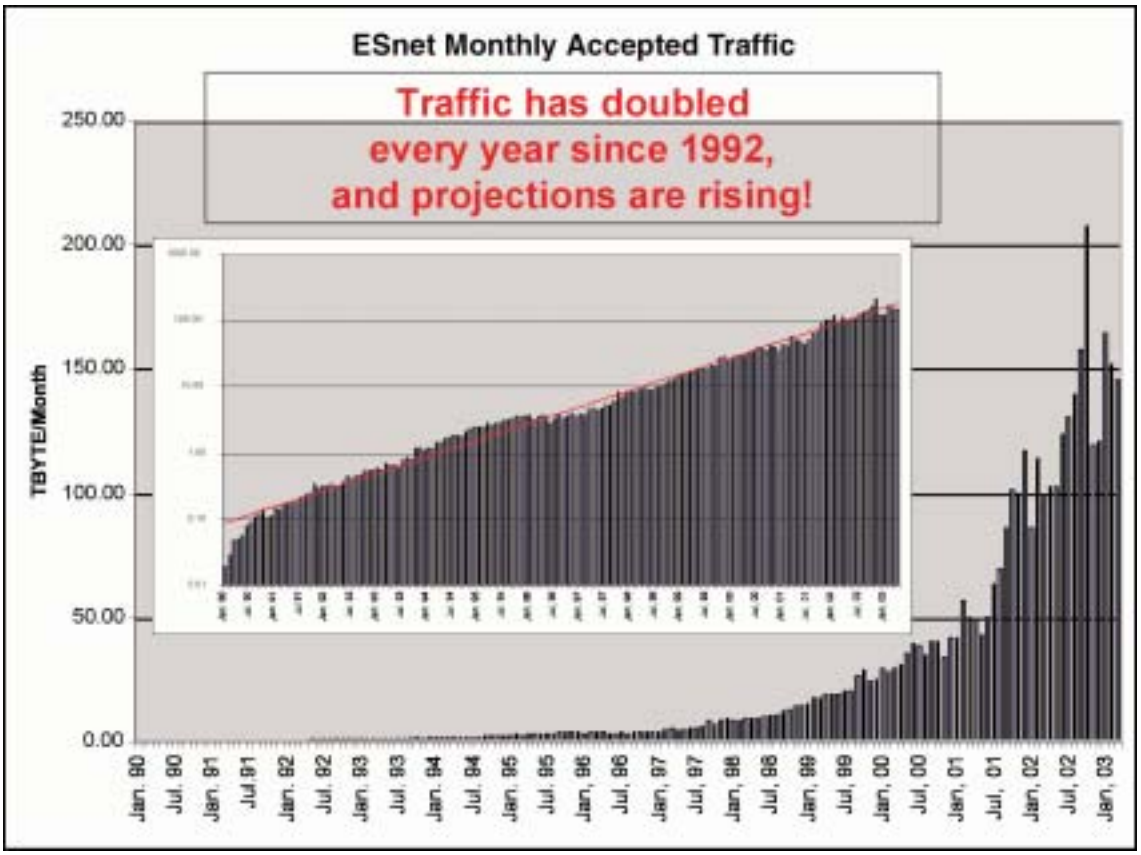


Figure B-1 ESnet Total Monthly Network Traffic

Coordinating Committee (ESCC) coordinates both the participation of and information dissemination to the individual institutions that benefit from the use of ESnet. The current ESnet Program Plan, dated March 1, 2001, prepared by the Steering Committee, is available on-line on the ESnet Web page (www.es.net).

Networking has been an essential component of infrastructure support for DOE science since the mid-1970s. ESnet was chartered in 1985 when it was recognized that the various research programs in DOE were beginning to develop independent approaches to networking support. ESnet was initiated as a consolidated, centrally funded program that would optimize costs, avoid duplication of effort, and provide for centralized management within this growing area of need. The program has been extremely effective in consolidating the networking needs of the Office of Science over a substantial period of years, including 1992 to the present, during which traffic has grown at a consistent annual rate of 100%, i.e., it has doubled every year, with no sign of abating (see Figure B-1). The ESnet project has managed to meet these ever growing network requirements and yet stay within budget since its inception.

The current Program Plan may appear conservative relative to the roadmap addressed in this report. This is a result of more recent requirements and projections expressed over a longer (and later) timeframe that will demand a more aggressive approach than outlined in the 2001 Program Plan.

Figure B-2 below gives a pictorial overview of ESnet as of early 2003. It may be noted that an additional upgrade to the backbone is currently planned for the last quarter of 2004 (Q1FY05) to upgrade the "southern route" from OC48 to OC192. Meanwhile, there will be emphasis on bringing up the access speed of various sites on

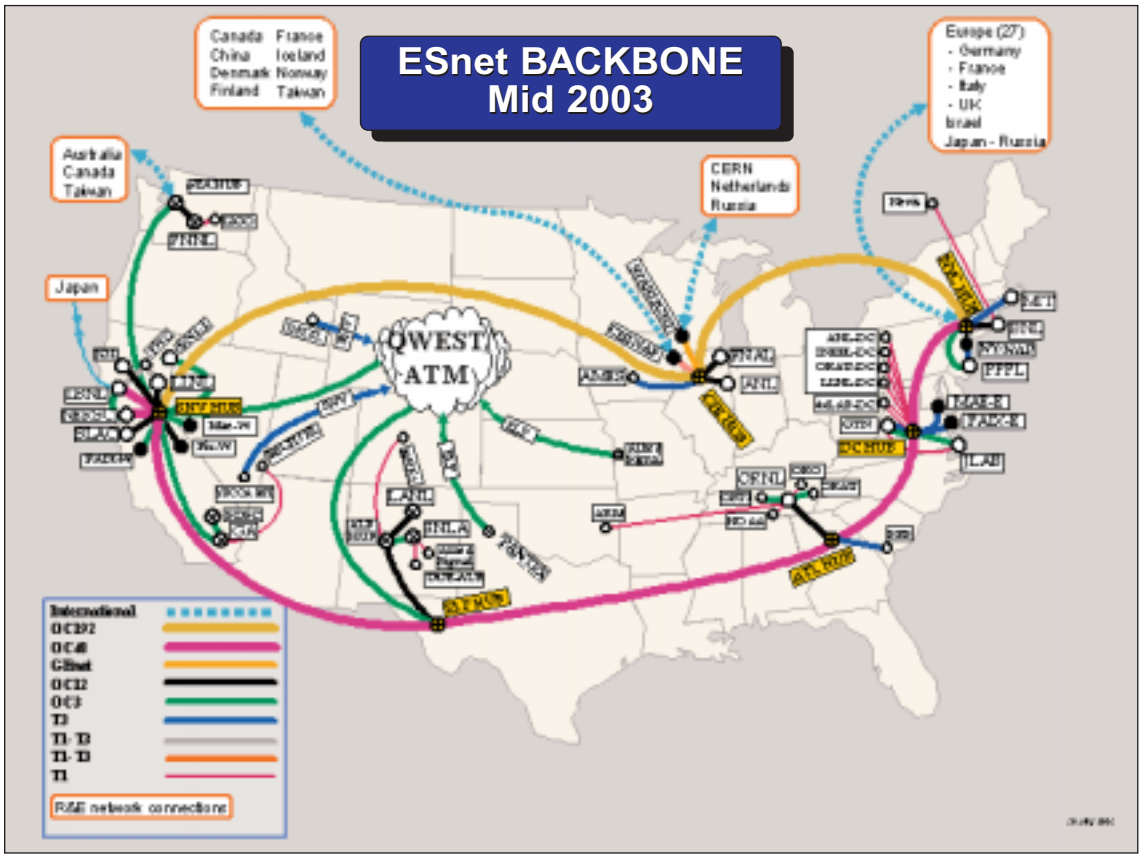


Figure B-2 ESnet Topology FY 2003

ESnet to allow them to take advantage of the enhanced backbone speeds.

B.3 Enabling DOE Scientific Discovery through Networking

The scale of DOE science is demanding in many respects, including the size of the massive research facilities required, the many academic disciplines involved, the national and international extent of the network, the size of the databases being transmitted, and the performance levels of the computing requirements. The support demands for this unique research environment are also exceptional.

The pace of scientific discovery in DOE depends strongly on an excellent research environment, which includes not only one-of-a-kind instruments such as accelerators, reactors, and detectors, but also leading-edge infrastructure such as networking and collaboration services. An appropriate set of tools, sized and structured specifically to meet the challenging requirements of the DOE research mission, will prove to be a well leveraged investment by optimizing the effectiveness of the much more expensive components of the mission such as its facilities and research staff.

Several areas of DOE science already push the existing infrastructure to its limits as they implement new elements. Examples include high-energy physics with its worldwide collaborations distributing and analyzing petabytes of data; access by systems biology to hundreds of sequencing, annotation, proteome, and imaging databases that are growing rapidly in size and number; and the astronomy and astrophysics community that is federating huge observational databases so it can look at all of the observations simultaneously for the first time. The clear message from the science application areas is that revolutionary shifts in the variety and effectiveness of how science is done can only arise from a well integrated, widely deployed, and highly capable distributed computing and data infrastructure.

It is projected that in the next five years ESnet must provide a broader spectrum of capabilities and services to support a more diverse portfolio of science. Robust and reliable production services are essential, as well as a commitment to network research for next-generation capabilities. Bandwidth usage will rise considerably faster than Moore's Law (i.e., the doubling of the number of circuit elements of integrated circuits every 18 months that has resulted in doubling the level of computer performance for a given cost), driven by a host of extremely challenging research problems, the broader penetration of terascale computing, terascale/petascale databases, and the integration of collaboration and remote access technologies into broad practice. Seamless integration of capabilities and services from many sources will be needed, incorporating both community and commodity components and application interfaces. Such network capabilities are at the heart of enabling the future vision of ESnet, offering an entire spectrum of capabilities, from palmtop to grid computing.

Although the deployment strategy employed to date has proven effective, it will not suffice to meet the demands of the DOE scientific research mission by 2008. This report proposes to accelerate the rate of deployment of network services by incorporating a more tightly integrated cycle of network research, leading to early deployment in the high-impact network, and finally incorporation into the production environment, as an ongoing cycle. In addition, renewed recognition must be given to the fact that DOE science is global in nature, and, to be effective, the approach must be well integrated with the national academic networking efforts such as Internet2 and international R&D networking projects such as GEANT in Europe and SINET in Japan.

B.4 Unique Capabilities of ESnet

The needs of the DOE research community are different from those of the general public served by the commercial Internet service providers

(ISPs) of today. Accordingly, ESnet has many characteristics and capabilities that differ from those available from commercial ISPs.

One of the most important aspects is that the ESnet project is solely dedicated to the DOE research mission, and therefore can be structured as both a cost-efficient and effective means of meeting the demands, current and projected, of that mission — rather than being driven by the necessity of commercial profit. Staffing is similarly focused on a single community, and as a consequence can offer unparalleled and directed support to its customer set. For example, during a recent cyber-attack, the ESnet staff were among the first "on the scene" to deal with the issue and worked over 24 hours straight, in shifts, to contain the worm and protect the community from additional infection, in some cases many hours before site staff were able to respond. The level of trust established over the years has allowed major sites in DOE to scale back staffing in favor of letting ESnet deal with most WAN issues on their behalf.

The requirements between the commercial/residential networking community and the DOE science research community differ dramatically in scale. DOE researchers now routinely talk about the need to move terabyte (a million megabyte) and even petabyte (a thousand terabyte) datasets across the network, whereas in the commercial sector a gigabyte (a thousand megabyte) file is considered quite large. This requires the bandwidth available between individual researchers and facilities to be several hundreds or even thousands of times faster than that typically provided commercially for a residential user.

High-performance external interconnects to other research and education communities are absolutely vital to the DOE research community, as many of its collaborations involve researchers from academic institutions and/or foreign locations. Commercial networks typically do not (and cannot in some cases) interconnect with the R&E networks serving these remote communities. As an example, the Abilene network, which serves over 200 major academic institutions in the U.S., has no commercial network interconnects. On the other hand, ESnet maintains high-speed interconnects with all major R&E networks on a global basis, including three (West Coast, East Coast, Central U.S.) with Abilene.

ESnet has typically provided a number of technical capabilities that are not available (at least at the time of initial deployment) on a commercial basis. Examples include:

1. Jumbo frames – Much larger frames (packets) support the fast transfers required by DOE science by minimizing the amount of end-system overhead required to move a given amount of data. ESnet supports 9000+ byte frames, whereas nearly all commercial ISPs are limited to 1500 bytes. The commercial ISP limitations are projected to continue for the foreseeable future. Office of Science data volumes cannot be moved via commercial ISPs.
2. Multicast – used by all Office of Science (SC) programs in applications where data needs to be sent from one source to many destinations simultaneously. A prominent example is video in support of video conferencing via an advanced Access Grid™ node.
3. Class of service – used by the HEP and LIGO project to provide various types of traffic with priority that may be higher or lower than standard traffic. For example, HEP has requested the lower priority capability to mark large data transfers during working hours and therefore avoid blocking normal traffic, while allowing any unused bandwidth to be utilized.
4. CCC/MPLS – Circuit Cross-Connect and MultiProtocol Label Switching are a vendor-specific and general capability, respectively, which when combined allow ESnet to carry traffic that cannot be carried by an Internet

Protocol-only network. Specifically, this capability allows ESnet to support the ATM (Asynchronous Transfer Mode) traffic needed by the NNSA SecureNet project, which also makes use of ESnet services.

5. IPv6 – The current Internet Protocol (IP) version is v4. IPv6 is the next version, offering advantages such a very much expanded address space (64 bits vs. 32 bits) and security enhancements. ESnet has been a catalyst in the early roll out of IPv6, with the world's first official IPv6 address assignment, and has been a prime mover in areas including the 6Tap. Recently ESnet and Abilene (the backbone network of Internet2) established native IPv6 "peering." These activities will help to smoothen the eventual transition to this new network level protocol.

B.5 ESnet Status

ESnet is a high-speed data communications network serving thousands of Department of Energy researchers and collaborators worldwide. Managed and operated by the ESnet staff at Lawrence Berkeley National Laboratory, ESnet provides direct connections to more than 30 major DOE sites at bandwidths up to 2.5 Gbps (billion bits per second) and with backbone bandwidths up to 10 Gbps. Connectivity to the global Internet is maintained through "peering" arrangements with more than 100 separate ISPs and research and education networks. Funded principally through DOE's Office of Science, ESnet allows DOE scientists to use unique DOE research facilities and computing resources independent of time and location with state-of-the-art performance levels.

ESnet today represents one of DOE's areas of excellence. Networking in support of DOE scientific research is now so integrated into the day-to-day operations of the Department's programs and laboratories that it would be impossible to effectively accomplish the science mission of the Department and the national laboratories without it. High-performance computing, scientific and engineering research, computational science, and a broad spectrum of "at-a-distance" interactions between people and resources, all at widely dispersed sites, are critical to the success of the Department's science mission. ESnet represents a consolidated and cost-effective approach to providing the requisite high-performance, leading-edge networking support across the agency. As such, ESnet is the largest multi-program collaborative activity within the U.S. Department of Energy.

As the computing and communications requirements of the Department's programs continue to evolve, improved ESnet capabilities are continuously enhanced and upgraded to provide higher levels of network performance, to support wider network availability and to enable use of more sophisticated applications. Accordingly, ESnet has a history of continuous improvement in its support of the science within DOE. For example, ESnet now carries more research traffic in a single day than it did in total for the years 1992 and 1993 combined.

Recent enhancements have upgraded the backbone (major interconnect links of the network) from OC12 (622 million bits/sec) to a combination of OC48 (2.5 billion bits/sec) and OC192 (10 billion bits/sec) — growth factors of four and sixteen over 2002 capabilities. Work is now under way to increase the connection speed of the sites connected to ESnet by equivalent factors.

Other Esnet-managed services have been recently implemented or augmented as well, including (1) a Public Key Infrastructure Certificate Authority (PKI CA) service that provides public key-based certificates to researchers and systems, allowing them to be authenticated (identity verified) within the grid activities of DOE and its collaborators; (2) an H323 (network) based video conferencing hub, allowing researchers and collaborators to meet and work "at a distance" using the network to carry the

video traffic; and (3) performance centers that aid in problem resolution (particularly performance-related issues) by providing midway points where performance can be verified as a means of breaking an end-to-end problem into manageable sections.

B.6 ESnet Budget

The budget for ESnet has enjoyed modest growth since 1990, although current projections to FY 2005 show continuation of a flattening trend over the years (see Figure B-3 below). Note, however, that the graph and trend line do not account for inflation, indicating that actual budget growth is much flatter than shown. At the same time, bandwidth demands (which comprise a major portion of ESnet project costs) continue to grow at a relentless rate of 100% per year (refer to Figure B-1).

The ESnet project has managed costs over the years to meet ever increasing traffic demands within a budget that has been growing at a decreasing rate. Although the market cost of a unit of bandwidth has also been decreasing over the years, the ESnet project has many costs that do not decrease over time, e.g., person-power and maintenance. This complex set of both upward and downward pressures on cost has been aided by deploying several unique approaches (such as establishing "hubs" at vendor point-of-presence [POP] locations) for cost containment. This has resulted in an ongoing overall annual reduction of approximately 40% per year in the monthly project cost to accept (and deliver) a unit of traffic (terabyte per month) demand, as shown below in Figure B-4. However, it should be clear that a 40% annual reduction in unit cost is not sufficient to sustain an annual growth of 100% in traffic within a fixed budget.

B-7 Management

The ESnet project is centrally funded by the DOE Office of Science through the MICS program manager and managed by staff located at Lawrence Berkeley National Laboratory. The responsibilities of the ESnet project staff span all aspects of project management including both long-term and short-term planning, budget, installation, and operations (on a 24×7 basis). ESnet has a well established reputation for excellence on a national and international basis and is

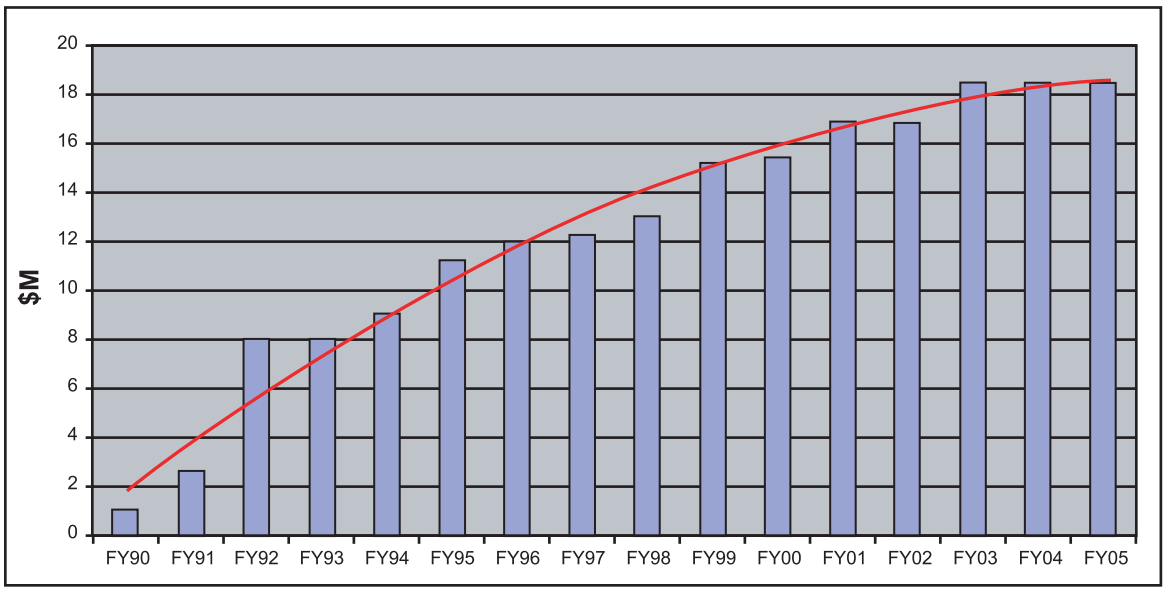


Figure B-3 ESnet Annual Budget (FY90-FY05)

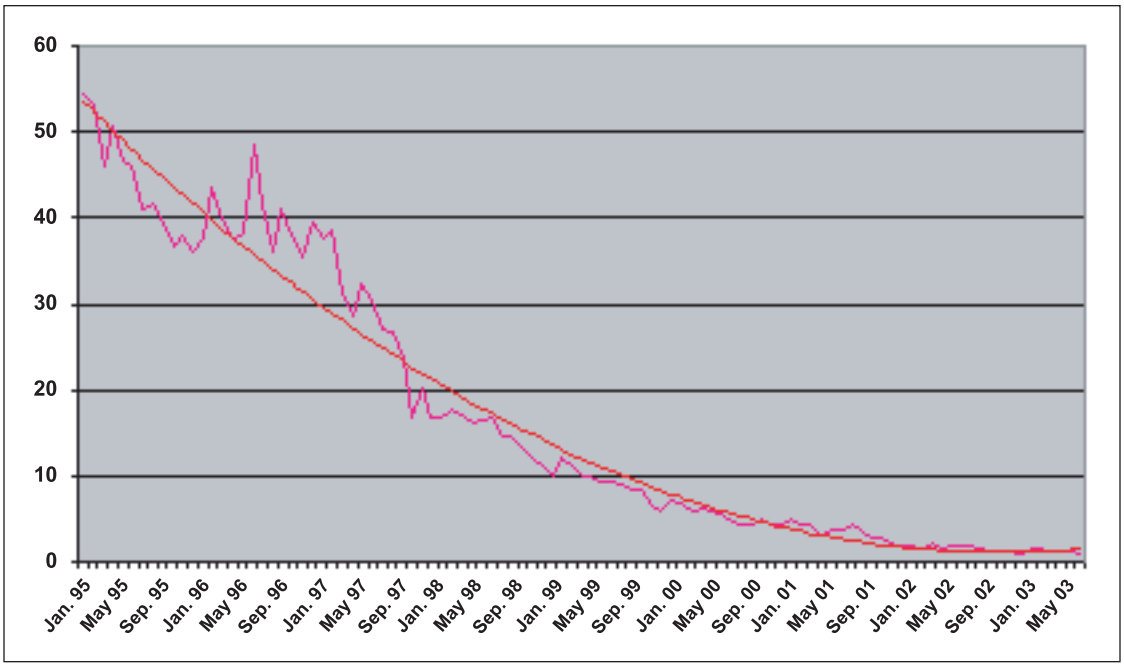


Figure B-4 ESnet Monthly Cost of a Unit of Delivered Traffic

currently one of the longest continuously operational IP networks in the world.

B.8 Governance

ESnet is a broad, multi-site DOE project that can be successful only with the broad support, cooperation, and involvement of both the SC programs it supports as well as the personnel at each connected site. It was recognized very early that a successful project would depend upon a governance model that would include providing a role for those all those involved. Figure B-5 below shows the current structure of the governance organization for the ESnet program. As the figure indicates, the ESnet Steering Committee (ESSC) serves as the "hub," with both SC and some non-SC program representatives serving on the committee having direct interaction with the ESnet project manager. Subcommittees, including the ESnet Coordinating Committee (ESCC), deal with specific issues.

The ability of ESnet to respond quickly and effectively to the evolving needs of DOE mission requirements is strengthened by the strong cooperation between ESnet management, program representatives, site networking staff, and end users. The ESnet Steering Committee (ESSC) has been a vital component of the ESnet process from the very beginning of the project. The ESSC is composed primarily of Office of Science program representatives in addition to providers of ESnet services. The committee is charged with identifying and prioritizing network requirements as well as reviewing implementation plans and resource allocation. Standing subcommittees, flexible task forces, and working groups ensure that members of the user community are strongly involved with the evolution of ESnet.

Although appearing complex, the governing committee structure as shown below has proven effective. However, the changes anticipated by the Roadmap to 2008 will necessitate changes, if the structure is to remain effective. Possible changes are discussed in Section 7.

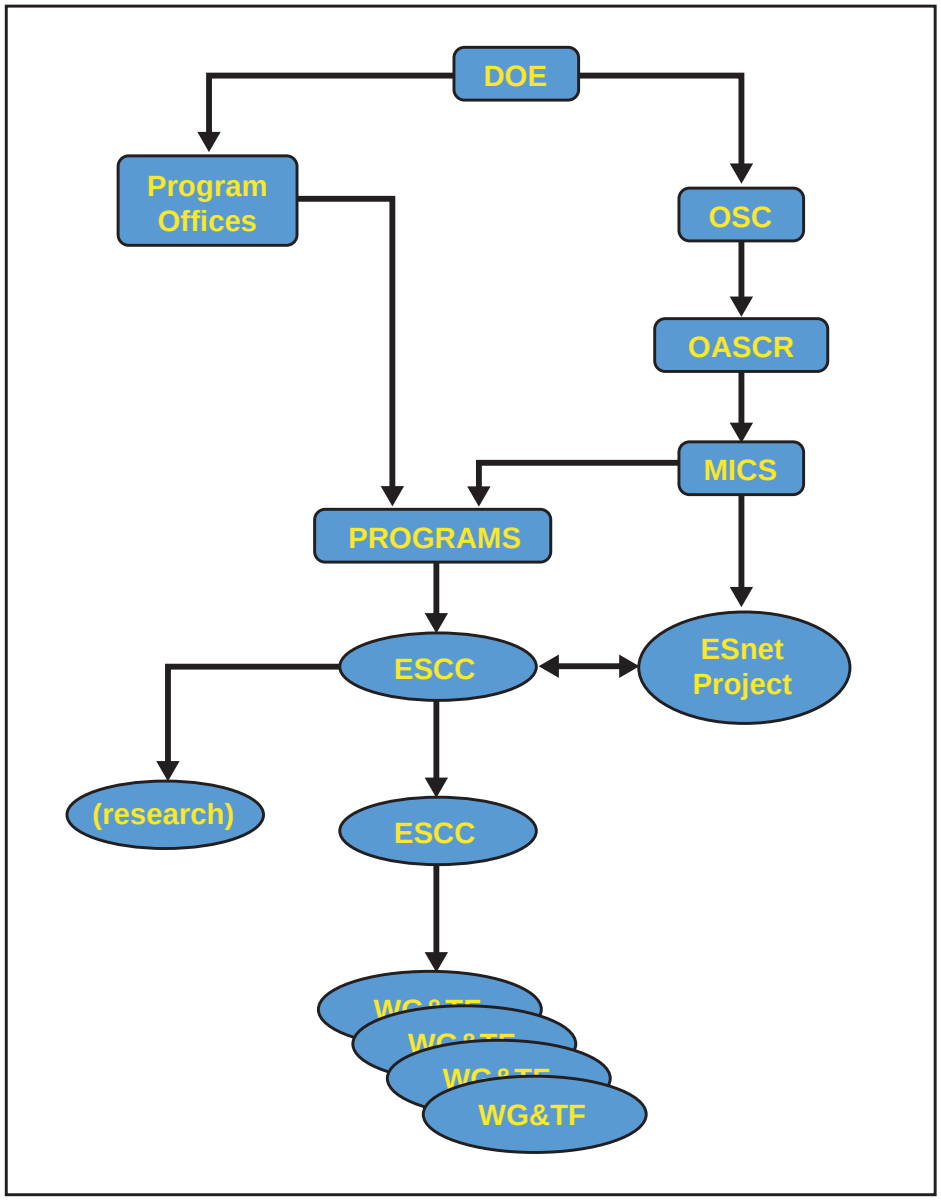


Figure B-5 ESnet Committee Structure

The last several years have seen dramatic changes in the global environment supporting and impacting networking activities. Changes have been widespread and rapid, with impact in the areas of (a) rapid introduction of new technology with associated downward pressure on pricing, (b) the bursting of the "dot-com bubble," resulting in both the availability of excess asset capacity and many communications vendors either in or on the edge of bankruptcy, (c) the self-formulation of a major new R&E network in the U.S. serving "high-end" applications for major academic institutions, (d) deregulation in the international arena resulting in drastic cuts in pricing and enabling the growth of significant new R&E networks in Europe, Japan, and Canada with international connections to the U.S. of a size unthinkable 10 years ago, and (e) the emergence of grids that promise to make distributed resources available on a basis approaching the ease of accessing electricity on the power grid.

These changes create both future challenges and opportunities for ESnet. This section briefly discusses some of these changes, how ESnet has taken advantage of some of the resulting opportunities, and some of the challenges the future may hold.

C.1 Technology

A very rapid pace of technology innovation has been the norm in networking for many years. However, the deployment of optical-based communications and more recently the move to and subsequent enhancements of DWDM (dense wavelength division multiplexing) have dramatically and systematically reduced the cost of a unit of communications bandwidth over the past several years. ESnet has taken advantage of this by implementing a deployment of DWDM in the network backbone during early 2003, moving the maximum backbone capacity from OC12 to a maximum of OC192 — a performance jump factor of sixteen that will help to meet future projected traffic growth.

ESnet and similar R&E networks have now all incorporated 10 Gbps links into their network backbones and, in some cases, even are offering access at such rates. 10 Gbps offers the maximum single channel capacity currently available. The next generation will be undoubtedly be 40 Gbps, although timing for commercial availability is at question, due in large part to the investment required at time when vendor capital funds are very sparse for new deployment at a time when a large oversupply of existing assets already exist. The roadmap to 2008 will need to address the issue of focus on single channel 40 Gbps links vs. multiple 10 Gbps links as requirements grow to press the capability of single channel 10 Gbps links.

A related issue for the future is the requirement for "on-demand" switched capacity between endpoints at bandwidths of 10 Gbps and possibly beyond. This requirement here will face the same dilemma of roll-out schedule as above. In addition, related issues such as establishing end-system connectivity, costing, scheduling, and sharing with the production environment will make this a very difficult requirement to meet and will need to be initially addressed in the research component of the roadmap on an early basis.

C.2 Economy

One of the major drivers of recent changes was the collapse of the dot-com-driven economy. The boom years of the dot-com era led many communications and networking vendors to dramatically expand infrastructure investment (without necessarily a good business plan) to capture market share. With the subsequent collapse of this economy, many vendors found themselves with significant unused infrastructure and/or capacity on hand and a slowing of growth in demand. A closely related fallout has been the Chapter 11 filings and/or subsequent questionable solvency of many of the vendors. A second related fallout has been the availability of "dark fiber" (fiber without the associated optoelectronic

equipment to light the fiber for operation) to end customers on both a regional and national basis. Although dark fiber typically requires a large initial investment to acquire and light (i.e., deploy the optical and electronic equipment necessary to made the fiber functional), subsequent incremental deployment of additional communication channels on the same fiber can often be had at relatively very low cost.

A small number of cooperative efforts are being formed to take advantage of this situation, including entities known as the National Lambda Rail and USAWaves, as two prominent examples. NLR has acquired dark fiber from Level-3 and optoelectronics from Cisco, while USAWaves appears to have established a working relationship with AT&T to acquire services on an incremental cost basis. Both will potentially have capabilities on a nationwide basis and at costs that could be substantially below market.

ESnet is now evaluating dark fiber for use on both a regional and national basis. The most promising area currently appears to be to implement a means of providing access circuits to ESnet member sites on a regional basis, but with additional capacity to support other applications, e.g., network research activities.

C.3 Academic

Although dating back to the mid-90s, there have been significant environmental changes that have led to appreciable differences and enhancements in access for DOE researchers to the U.S. domestic academic community. NSF chose to terminate support for networking for this community, a network with many similarities to ESnet known as NSFnet during the mid-1990s. The expectation was for the academic community to turn to the commercial networking market for support. However, following a few years of totally inadequate service, the academic community consolidated its requirements and formulated a collaborative effort headed by an organization known as Internet2 (I2) and began deployment of a private high-performance network called Abilene, which now has over 200 members, including most major academic institutions in the U.S.

ESnet has been a collaborator with the I2 community since its early formative years and remains so today. ESnet currently has three interconnects in place with the Abilene network to provide low-latency and high-performance networking between DOE researchers and their U.S.-based academic collaborators.

The academic community is exploring opportunities for multiple lambda-level services, i.e., services above OC192/10 Gb/sec. The community is setting up the National Lambda Rail (NLR) as a platform for provisioning these services. NLR may also integrate with USAWaves. As NLR evolves, it will be important for ESnet to interface with the services running on NLR, as many of the users of Office of Science facilities are located at universities, and these users will use these high-end services as their method of moving terabyte/petabyte-scale volumes of data to and from ESnet.

C.4 International

Deregulation of pricing, particularly in the European market, has precipitated dramatic price reductions for both national and international communications. In particular, this has allowed international research and education networks to provision international links to the U.S. (generally either to Chicago or New York) with capacities that would have been unthinkable only a few years ago.

One direct impact of deregulation, particularly in Europe, has been the dramatic growth of foreign research and education networks. One of the largest is the GEANT network, managed by the DANTE organization; it is a pan-European backbone network with major links at 10-Gbps capacity interconnecting more than 30 European

countries and with international connectivity to the U.S. via dual 2.5-Gbps links. Networks of similar scale and performance have emerged in Canada (CANARIE) and in Japan (Super SINET).

ESnet has been taking advantage of the resulting opportunities to establish peering interconnects with networks that support DOE international collaborators, again at enormously improved performance levels, with interconnects at both Chicago and New York. This directly supports the demands for high-performance access to international collaborators and research facilities from the U.S.-based DOE scientific community.

Like the academic community, the European research community is exploring provisioning services beyond 10 Gbps. They desire to connect both to ESnet and the NLR at these speeds. Like DOE science, European science is involved in multiple projects with petabyte-scale datasets that are driving the networking and collaboratory requirements. The earliest and largest international project is the LHC at CERN that, beginning in the 2007 time frame, will routinely move petabyte-scale datasets to multiple locations in the world, including two in the U.S., Brookhaven National Laboratory (BNL) and Fermi National Accelerator Laboratory (FNAL). LHC will use grid technologies to manage the metadata, data flows, and other complexities of the processes. Note, the U.S. portion of this process also involves a collaboration with NSF to fund the U.S. universities that will receive their LHC data flows from the two U.S. LHC data centers at BNL and FNAL.

C.5 Grids

Grids provide the services, resources, infrastructure, and people needed to enable the dynamic construction and use of collaborative problem-solving environments using geographically and organizationally dispersed high-performance computing, networking, information, and data handling resources. This distributed infrastructure, made available through grid technologies, reduces or eliminates barriers to the coordinated use of scientific resources so that the physical locations of data, computers, sensors, and users could (if adequately provisioned and managed) become largely irrelevant. By so doing, this new infrastructure enables a far closer coupling than currently possible of scientists with each other and with science resources of all kinds. Grids, in turn, are dependent upon both high-performance networks as the foundation of their capabilities as well as other services upon which the successful implementation of the grids concept depends.

ESnet provides the necessary networking support to the grids implementation activities within the DOE research community. In addition, the ESnet project is providing some of the services essential to the grids implementation, including (1) a Public Key Infrastructure Certificate Authority (PKI CA) as the basis for personnel and system authentication and personnel collaboration services including audio, video, and data conferencing; and (2) collaboration services for personnel including audio, video, and data conferencing centers with a Web-based reservation capability.

The ESnet PKI CA is currently actively used by the DOE Science Grid, the SciDAC Particle Physics Data Grid, the Earth Systems Grid, and the Fusion Grid projects, and it also facilitates scientific collaboration between the U.S. and European high-energy physics projects.

There is need for a number of additional support services to facilitate the successful deployment of the many emerging grids. This topic is addressed much more fully in Section 4, "Technology and Services."

C.6 Security

The widespread availability of global networking to the general public has brought with it a rapid growth in both the numbers and sophistication of cyber-security attacks. Furthermore, the events

of Sept. 11, 2001, have significantly crystallized the potential for even more serious threats. Accordingly, the need for increased cyber-security protection has risen in concert with the threat growth. However, commercial implementations of security capabilities, such as firewalls and intrusion detection systems, typically lack the ability of operate at the leading-edge performance levels dictated by the demands of the DOE research mission.

Therefore, security will be one of the areas that must be addressed by the networking research community to ensure that DOE unclassified science can continue to be effective, open, and competitive, yet remain secure. Future network security will need to address increasing intelligence in the network, intelligence made necessary by the need to reduce complexity (and the level of expert manpower to operate and manage) next-generation systems such as grids based on global networks. This topic is also addressed further in Section 4, "Technology and Services."

ESnet has made progress in meeting these divergent demands through a new backbone architecture that will separate external access from internal access and furthermore will allow the external access to be monitored and controlled in the event of a large-scale attack.

Appendix D: Technology and Services Roadmap Tables

The Technologies and Services Working Group identified 13 issues, which were rank ordered by the working group as to their impact on the success of distributed science. The five top-ranked issues were judged to be essential to establishing and maintaining the basic middleware infrastructure. The next three were rated as very important, and the remaining five as very important or important. This appendix provides in ranked order both the descriptive details and the roadmap details of the five essential and first three very important issues.

In these tables, we have tried to capture a realistic development and deployment schedule for each service, the ongoing operational costs for the service, and a budget for both. Each of the tables lays out a roadmap from where we are today with the technology, through R&D, to an operational production capability, and finally the estimated ongoing cost of operating the service to provide middleware infrastructure for DOE science.

D.1 Integrated Grid and Site Security

The grid computing concept depends heavily upon the ability of virtual organizations (VOs) to utilize services that may exist at many different research institutions independent of geography or network domain. One of the most significant performance impediments of present-day grid computing are the security/protective systems (firewalls and related technologies) that are required to be in place at some research institutions. All firewalls and related technologies that a grid application must pass through must typically be manually configured to support a specific grid application. If the grid application configuration changes, it may not have all of the necessary firewall permissions necessary to function properly. This operational paradigm is inefficient of labor resources, does not scale to larger and larger grids, and introduces delays into research programs.

What is required is the development of a security service that would help overcome these problems by providing a standard interface for VOs to perform firewall administration that allows for fine-grained access control policy and lifetime management of ports. Such a service would allow a VO to securely and reliably authenticate itself to firewall(s) or a secure service that can itself authenticate to the firewall(s), and thereby obtain the necessary firewall conduits for VOs' applications to function properly. Desirable functionality would include: (1) support for the service-level agreements (SLAs) and managed lifetime in the Global Grid Forum (GGF), (2) authenticate to the security perimeter in order to authorize access, (3) local policy enforcement, (4) auditing, (5) NAT support, and (6) high performance.

While we believe this delegation of firewall and related technology maintenance to the VO provides a good balance between risk and functionality, there are issues associated with implementation, such as: (1) the creation of a trusted authentication server, (2) new tools to integrate authentication/authorization processes and firewalls, and (3) potential modification to current grid functionality for compatibility. Further, research institutions that choose to participate in this process must accept this new security paradigm. It will be necessary to define the policies to be used for the purpose of modifying firewall behavior by VOs. The process to develop these policies must include cyber security and technical staff from ESnet, research institutions, and the Global Grid Forum.

Table D-1 Integrated Grid & Site Security

Milestones and Metrics	FTEs Working on Task	Completion (mo. from start)	Task Duration (mo.)	Task Cost ($K)
Establish the process - Identify existing, unclassified research - ESNet/Site collaboration (ESCC subcommittee) - Identify natural, existing communities (DOE Science Grid, collaboratory projects) - Coordinate with Global Grid Forum - Coordinate with DOE Cyber Security policy makers - Coordinate with research institution cyber security officers	4	3	3	250
Establish the use cases that define the requirements and scope - All grid functions, ad hoc collaboratory services need to be addressed, and off-site access need to be addressed	6	9	6	750
Initial technology choices - Authentication compatible with grid security - Investigate commercial and public on-going research - Identify candidate technologies to pursue based upon detailed research of each candidate	6	12	3	375
Research and development - Vendor partnerships - Grid developer partnerships - Authorization approach	6	36	24	3,000
Test and evaluation, initial pilot service rollout - Implement on representative test network - Include representatives of target communities and research organizations - Perform extensive usability testing - Perform extensive validation of security models	6	42	6	750
Technology decision - Perform Cost-benefit analysis of approach(es) - Coordinate with LSN/MAGIC and/or other as appropriate	4	44	2	167
Implement the prototype production service, continue pilot service - Coordinate with sites and user communities - Develop detailed plan of rollout - Determine deployment costs - Determine support and sustainment costs - Establish community advisory committee - Document usability aspects - Implement and initiate the prototype production service	6	50	6	750
Deploy first production service	4	54	4	333
Total				**6,375**
Ongoing operation, anual cost E.g. What is the additional cost over the current network services to provide this service?				**750**

D.2 Collaboration Services

Future distributed scientific research implies the need for geographically separated groups to work together remotely as a coherent whole. For example, it is reasonable to assume that not all members of an experimental team will be on-site for all experiments. In fact, it is probably desirable and practical to carry out a significant amount of scientific work remotely. Effective remote participation in experiments on the scale envisioned will demand significantly enhanced and coordinated resource sharing and problem solving in a dynamic, multi-institutional international environment. For example, the future magnetic fusion control room for a burning plasma experiment will require key adjustments to be made to hardware/software controls only after a vast amount of data have been assimilated in near real time. Successful operation in this mode will require the movement of large quantities of data between pulses to computational clusters, to data servers, and to visualization tools used by an experimental and theoretical team distributed across the world and the sharing of remote visualizations and decision-making back into the control room.

Beyond experimental operations, such activities as group data analysis, writing and debugging of simulation codes, workshops, and seminars can all be more efficiently conducted with the ability to effectively participate remotely. Collaboration on this scale is a technically demanding problem since it requires the presentation of a working environment to off-site personnel that is every bit as productive and engaging as when they are physically on-site.

Provided services must cover the full range of high-quality group access, high-quality desktop access (for those with the bandwidth), and limited-quality desktop access. Limited-quality desktop stations need to be able to participate in higher-quality meetings, albeit with reduced sound and picture quality. Access Grid has started to provide the full immersion presence required for advanced collaborative environments, including continuous presence and persistent scenarios like remote experimental control room participation.

Overall ease of use and reliability must be built into these capabilities. An important complement to real-time interactive remote participation is a good recording and on-demand replay service for important meetings, including any presented materials (e.g., slides). Application- or desktop-sharing technology is a basic building block for remote meetings. It allows for a range of usage scenarios, including discussing scientific analysis results over a distance, sharing electronic presentations in meetings, and the operation of remote computers with specialized software.

Table D-2 Collaboration Services

Milestones and Metrics	FTEs Working on Task	Completion (mo. from start)	Task Duration (mo.)	Task Cost ($K)
Establish the process - Identify existing research - ESNet/Site collaboration (ESCC subcommittee) - Identify natural, existing communities (DOE Science Grid, collaboratory projects) - Coordinate with Global Grid Forum	4	3	3	250
Establish the use cases that define the requirements and scope - Persistent and shared workspaces and applications - Session management for AV services - Production video teleconferencing (scheduling, switch/multipoint, venue server) - Lightweight/casual collaboration - Code repositories and code project management	6	6	3	375
Initial technology choices - MUD, DAV+Portal - VRVS, VNC, DMX - H323 - Access Grid	2	7	1	42
Research and development - Define initial testbed - Vendor partnerships - Packaged services	6	21	14	1,750
Test and evaluation, initial pilot service rollout - Sample infrastructure and target community - Usability and functionality testing beyond initial testbed	8	30	9	1,500
Technology decision - Cost benefit analysis	4	33	3	250
Implement the prototype production service, continue pilot service - Coordinate with sites and user communities - Develop detailed plan of rollout - Determine deployment, support, and sustainment costs - Establish community advisory committee - Document usability aspects - Implement the Advanced Collaborative Environment	6	42	9	1,125
Deploy first production service	2	45	3	125
Total				**5,417**
Ongoing operation, anual cost - Maintain present capabilities - New site training and support				750

D.3 Performance Monitoring

Monitoring is the measurement and publication of the state of a computing/networking component at a particular point in time. To be effective, monitoring must be "end-to-end," meaning that all components between the application endpoints must be monitored. This includes software (e.g., applications, services, middleware, operating systems), end-host hardware (e.g., CPUs, disks, memory, network interface), and networks (e.g., routers, switches, or end-to-end paths).

Monitoring is required for a number of purposes, including status checking, troubleshooting, performance tuning, debugging, application steering, characterizing usage, planning, setting expectations, developing and auditing service-level agreements. For example, assume a grid job has been submitted to a resource broker, which uses a reliable file transfer service to copy several files to the site where the job will run, and then runs the job. This particular process should normally take 15 minutes to complete, but two hours have passed and the job has not yet completed. Determining what, if anything, is wrong is difficult and requires a great deal of monitoring data. Is the job still running or did one of the software components crash? Is the network particularly congested? Is the CPU particularly loaded? Is there a disk problem? Was a software library containing a bug installed somewhere? Monitoring provides the information to help track down the current status of the job and locate problems.

Table D-3 Performance Monitoring

Milestones and Metrics	FTEs Working on Task	Completion (mo. from start)	Task Duration (mo.)	Task Cost ($K)
Establish the process E.g. Coordinate with Global Grid Forum, Internet2, set up site coordination through ESnet Coordinating Committee	4	3	3	250
Establish the use cases that define the requirements and scope E.g. Clarify the use cases that identify services required. These include: - Tools for performance analysis and distributed application debugging, tools and services for Failure analysis - Input for path optimization / scheduling (e.g. replica selection) - Flow and service auditing (SLA enforcement, capacity planning, service usage patterns) - Expert consulting	6	9	6	750
Initial technology choices E.g. Identify candidate technologies to pursue based upon detailed analysis of each candidate technology	4	12	3	250
Research and development - Define initial monitoring testbed - Develop and/or integrate monitoring mechanisms (hardware and software) - Publish, archive, and analyze the data (define data models, common schemas, etc.) - Cross domain monitoring data usage - Monitors for network – passive and active, applications, hosts, and protocols - User/host tools (data gathering, host participation in testing, analysis) - Define and implement security model for all monitoring components	6	36	24	2,500
Test and evaluation, initial pilot service rollout - Functional testing of Monitoring Services on 'test network' - Technology down select via cost-benefit analysis	6	58	22	2,500
Implement the prototype production service, continue pilot service - Implement and initiate the monitoring service	7	64	6	875
Deploy first production service	3	70	6	500
Total				**7,625**
Ongoing operation, anual cost E.g. What is the additional cost over the current network services to provide this service?				750

D.4 Network QoS

Supporting the routine creation of robust, high data-rate, distributed applications that support critical but transient network uses, such as scientific instruments that produce data only during experiments and specific data analysis exercises, necessitates developing various network functionality and middleware services, including the ability to establish high-speed, end-to-end network bandwidth reservations among the elements of distributed systems.

Every aspect of this environment is dynamic and requires the ability to move data among geographically dispersed sites at very high rates based on prescheduled "contracts" for the diverse resources that make up the data collection, storage, and analysis systems and the network resources that connect them together.

Bandwidth reservation is one aspect of the general resource reservation problem, but one that is central to the environment, and one that involves unique network issues.

Bandwidth reservation that can be used in the context of a general resource reservation scheme involves bilateral end-node agreements that "reserve" bandwidth in the sense that a site actively manages allocation against one or more classes of service. The overall limits on a class of service are established in the corresponding service-level agreement between the institution of the end nodes and the ISP, but the allocation of flows to this class is closely managed by the end node institutions at the site egress.

Further, the resource allocation should be policy-based in a way that allows automated reservation, and it should also be possible to proxy one's policy-based authority to another site, so that the bilateral agreements necessary for inter-site application operation happen automatically.

Table D-4 Network QoS

Milestones and Metrics	FTEs Working on Task	Completion (mo. from start)	Task Duration (mo.)	Task Cost ($K)
Establish the process E.g. Coordinate with Global Grid Forum, Internet2, set up site coordination through ESnet Coordinating Committee	2	3	3	125
Establish the use cases that define the requirements and scope E.g. Specified bandwidth (e.g. Instrument data streams, data transfer), instrument control (low jitter), interacting simulations, human response (e.g. anything interactive), human collaboration tools, scheduled bulk data.	2	6	3	125
Initial technology choices E.g. Flow level mechanism, allocation mechanism, security.	6	15	9	1,125
Research and development E.g. Testbed for deployment testing, access control, network management tools, bandwidth brokers for co-scheduling.	8	27	12	2,000
Test and evaluation, initial pilot service rollout E.g. Enlarge the testbed to include a useful set of DOE sites, do cross-domain testing (e.g. w/ Internet2)	8	33	6	1,000
Implement the prototype production service, continue pilot service E.g. Implement prototype production service, establish community allocation committee, deal with usability, resolve site issues.	8	39	6	1,000
Deploy first production service E.g. Train operations staff, deploy production equipment, set up allocation management system, integrate w/ Network Operations Center	2	45	6	500
Total				**5,875**
Ongoing operation, anual cost E.g. What is the additional cost over the current network services to provide this service?				1,500

D.5 Authorization

Successful distributed science requires providing more or less transparent access to resources and stakeholders, and owners of resources need to be able to protect resources from inappropriate use at the same time. Authorization middleware is needed to provide a set of scalable tools by which this protection is defined and enforced. Note the distinction between the authentication process, which verifies the identity of users, and authorization, which defines what an authenticated user can do.

Resources that need protection include both real and intellectual property. Those responsible for scientific instruments must ensure to a very high degree that remote control does not jeopardize their safe and productive operation. Data from experiments is ordinarily made available only to those researchers actively collaborating on it. Cycles on powerful supercomputers are a valuable resource along with software and data that must be protected from misuse, modification, or theft. Sophisticated physics and engineering codes represent important intellectual property, and their developers need to verify the integrity of the software, to guarantee that it is used correctly and to ensure that they are given proper credit for their work. For all these cases, transparent on-line access is predicated on the ability of the resource owners to control use.

Flexibility and scalability must be built into any authorization scheme if it is to be widely used. Given the scope and highly decentralized nature of our scientific enterprise, it is clear that stakeholders need to be able to set policy for their own resources without mediation by a central authority. Systems that allow policy for each resource to be set by its owners solve one aspect of the scalability problem. Another challenging aspect of providing scalability is on the enforcement side of the problem. A goal for authorization middleware is to enable the "single sign-on" paradigm, where a user authenticates himself once then has transparent access to all resources whose owners authorize his use.

Overall ease of use should be built into authorization middleware. The interface must be able to span resources on different time scales, include preemptive scheduling, and be linked to historical databases to provide rapid access to current and previous usage. Users should be able to view their own entries in authorization databases to help them in assembling the resources they need for their own work. It is critical that users receive informative feedback when they are refused access. To expedite problem resolution, users should get information on which resource is being denied and why. At the same time, the system should provide resource owners with the capability for logging and auditing.

Table D-5 Authorization

Milestones and Metrics	FTEs Working on Task	Completion (mo. from start)	Task Duration (mo.)	Task Cost ($K)
Establish the process E.g. Coordinate with Global Grid Forum, Internet2, set up site coordination through ESnet Coordinating Committee	2	3	3	125
Establish the use cases that define the requirements and scope - Supercomputer access (especially NERSC) - Computationally intensive application services - Instrument or experiment access (Tokamak control room, synchrotron X-ray sources) - Data access - Personal collaboration sessions, including Access Grid	2	9	6	250
Initial technology choices - Consider SAML. XACML, X.509 attribute certificates - Evaluation authentication choices: X.509, Kerberos, user/password pairs - Evaluate enforcement mechanisms - dynamic account creation, sandboxing techniques	6	15	6	750
Research and development - Formulate authorization as an OGSI service - Define points in the infrastructure to make authorization decisions - Consider integration with firewalls - Address enforcement issues	6	27	12	1,500
Test and evaluation, initial pilot service rollout - Implement and deploy several prototype authorization services supporting different authorization and policy models, in support of different use cases - Develop a set of evaluation criteria - e.g. ease of use for stakeholders, site administrators and users, clarity and appropriateness of policy definitions, robustness in the case of network failures - Have users of the services evaluate them according to the criteria - Evaluate enforcement mechanisms	6	39	12	1,500
Implement the prototype production service, continue pilot service - Chose a restricted number of authorization services to fully implement probably based on the different requirements of the use cases - Provide tools for implementing enforcement mechanisms if necessary - Set up security/authority infrastructure - Help application developer's to use new services	10	55	16	2,500
Deploy first production service - Fully document the use and risk assessment for site administrators and stakeholders - Make authenticated versions of the standard authorization services available - Support an infrastructure to find authenticated security information: e.g. public keys of trusted severs and authorities, revocation support for authentication and authorization tokens. Probably organized per VO., but developed and supported by ESnet	6	67	12	500
Total				**7,125**
Ongoing operational, anual cost - Keeping the security information infrastructure secure and current - Account creation and destruction - Maintaining authorization policy				500

D.6 Namespace Management

As scientific enterprises such as large-scale simulations or experimental science data analysis collaborations become more highly distributed, the need emerges for information services that exist at the enterprise level instead of at the local computer center level. Key among these information services is a multipurpose name service.

Part of the success of the current Internet can be attributed to the use of user-friendly hierarchical names to refer to computers, such as www.my company.com, and a corresponding distributed Domain Name System (DNS) to provide the mapping between the computer name and its address. This naming service also has a powerful scalability feature in that it is fully distributed; a particular DNS server need only know about the part of the name space for which it is responsible. Thus, there are top-level DNS servers that operate on the rightmost fields in the name (the domain), and these services can redirect inquiries to domain-specific DNS servers to look up a particular sub-domain or host.

Many similar needs for translation between names and addresses, or between names and some other property, exist within highly dispersed enterprises.

Among the many applications of naming (lookup) services within a large-scale science enterprise, naming and location of datasets is probably the most significant. For some fields, data is produced (simulated or acquired) at multiple sites. For other fields, it is acquired at one site, transferred to one or more other sites, and at these multiple sites derived data sets are produced. It would be inefficient for a client application to be required to search all sites in order to find a particular dataset. One scheme being widely adopted today is to give each dataset a globally unique name, and then maintain the relationship between that name and the dataset location within some type of naming service.

A root-level name service, similar to the root domain name servers, would enable a powerful set of location-independent tools and applications to be built up. As ESnet and other research networks push the capabilities of the science network infrastructure to include more than just transporting bytes from one place to another, a key component of these higher level services will be such a name service.

Table D-6 Namespace Management

Milestones and Metrics	FTEs Working on Task	Completion (mo. from start)	Task Duration (mo.)	Task Cost ($K)
Establish the process - Do natural communities already exist (DOE Science Grid, collaboratory projects) - Coordinate with Global Grid Forum	2	3	3	125
Establish the use cases that define the requirements and scope - Like domain name resolution – opaque string translation - Functions to be provided	4	6	3	250
Initial technology choices - Pursue LDAP/LDAP and XQuery/SOAP in parallel - Security	6	9	3	375
Research and development - Define schema - Define testbed	6	21	12	1,500
Test and evaluation, initial pilot service rollout - Sample infrastructure - Target community - Usability testing	6	24	3	375
Test and evaluation, technology down select - Cost-benefit analysis - Coordination with LSN/MAGIC and/or other as appropriate	4	36	12	1,000
Implement the prototype production service, continue pilot service - Establish community advisory committee - Usability aspects	4	43	7	583
Deploy first production service	2	48	5	208
Total				**4,417**
Ongoing operation, anual cost E.g. What is the additional cost over the current network services to provide this service?				500

D.7 Publish/Subscribe Portals

As noted earlier, science is rapidly becoming an inherently distributed endeavor, with an increasing dependence on electronic access to growing volumes of distributed data. At the same time, access to the associated metadata, models, and annotation is also especially critical to the progress of interdisciplinary science in fields such as combustion, biology, and nanoscience. This need is often most evident as research results appear in peer-reviewed journals, when much of the underlying data, detailed model descriptions, and numerical results remain unavailable. Public data portals are already critical to fields such as biology and astronomy; there is a rapidly growing need for improved standards, more features, and for support across many more DOE science communities.

The ESnet Publish/Subscribe Data Portal/Service will support DOE science by publishing and providing access to data that complement peer-reviewed journal publications. The permanence, unique identifiers, and easy accessibility provided by this service will be a significant step toward the formation of the scientific knowledge bases and communities that will revolutionize science in the future.

This collaboration infrastructure provides a natural channel for the association of metadata and other information with data to allow expression of the data in the language of other disciplines, making it most appropriate for publication.

Data policies should require minimal rights to be transferred to the service and allow content providers to otherwise document their intellectual property and licensing terms. At the same time, open sharing of data can be fostered by implementation of new DOE open-sourcing policies similar to those now being pursued for software. Public read access to data should not require authentication of the user. Negotiations of policy with other science and technology agencies, publishing houses, and industrial users will be carried out to optimize integration of services and data availability.

Table D-7 Publish-Subscribe Portal

Milestones and Metrics	FTEs Working on Task	Completion (mo. from start)	Task Duration (mo.)	Task Cost ($K)
Establish the process E.g. Coordinate with Global Grid Forum, Internet2, set up site coordination through ESnet Coordinating Committee	2	3	3	125
Establish the use cases that define the requirements and scope E.g. Clarify the use cases that identify services required, define ESnet role	4	9	6	500
Initial technology choices Detail requirements and dependencies and associated time-lines of other services including namespace management, collaboration services, etc.	4	12	3	250
Research and development - Focused development of key underlying technologies and services	8	24	12	2,000
Test and evaluation, initial pilot service rollout - Pilot service with 1 or 2 disciplines to refine goals, technologies, and policies	8	48	24	2,000
Technology decision Decide on scope and strategy of implementation by analyzing costs and benefits. Define growth/coverage strategy. Coordinate with other agencies and entities.	4	52	4	333
Implement the prototype production service, continue pilot service Implementation, including deployment and support plans, community advisory committees, staffing	6	60	8	1,500
Deploy first production service Production service includes hardware, management of user communities, and enforcement of editorial and other policies	6	72	12	1,000
Total				**7,708**
Ongoing operation, anual cost E.g. What is the additional cost over the current network services to provide this service?				500

D.8 Self-Defensive Networks

Many scientific disciplines are adopting grid and other similar models of research that depend upon widely distributed but cooperative computing systems and databases. Examples include high-energy physics with its worldwide collaborations distributing and analyzing petabytes of data; systems biology access to hundreds of sequencing, annotation, proteome, and imaging databases that are growing rapidly in size and number; and the astronomy and astrophysics community that is federating huge observation databases so it can, for the first time, look at all of the observations simultaneously.

The grid concept depends heavily upon the ability to move large amounts of information among multiple systems that may exist at many different research institutions independent of geography or network domain. The Energy Sciences Network (ESnet) today provides the infrastructure to support this model.

This research model proposes that network performance for scientific collaboration, distributed data analysis, and grid computing can be significantly improved by designing systems with selected elements of detection and protection on the edges of ESnet, thus creating a protected environment for at least selected protocols, addresses, and/or types of services within ESnet. The current approach of firewalls and intrusion detection are implemented primarily at the individual research laboratories that are interconnected by ESnet. This proposal would add intrusion detection with an automated response capability as well as selected firewall policies to the locations where ESnet and ESnet member labs interconnect with external networks.

Secure authentication and dynamic configuration of the firewall will greatly ease the burden involved with getting grid applications to work by greatly reducing manual configuration, and will be more secure since access will be available only when it is actually needed, rather than having a range of ports sitting open at all times. However, this does not address compromised credentials or authorized persons who act maliciously (intentionally or not), etc. This self-defensive networks work would address this by developing technologies to monitor the traffic, detect malicious and otherwise inappropriate behavior, and potentially respond to it without human intervention.

This research model would be deployed in a manner as to be self-defensive in that the systems would be designed to continuously monitor traffic and automatically identify actual attacks, patterns of potential attacks, or unauthorized intrusions and to block the sources of such traffic. This architecture would provide a more open environment within ESnet. For example, denial of service, Web defacement, and other malicious attacks would be identified and stopped at ESnet borders before reaching the research labs. It is anticipated that this detection technology will be capable of identifying and blocking even technically sophisticated attacks that may occur over long periods of time.

Table D-8 Self-Defensive Networks

Milestones and Metrics	FTEs Working on Task	Completion (mo. from start)	Task Duration (mo.)	Task Cost ($K)
Establish the process - Identify existing, unclassified research - ESnet/Site collaboration (ESCC subcommittee) - Coordinate with Global Grid Forum - Coordinate with DOE Cyber Security policy makers - Coordinate with adjacent Non-ESnet security organizations - Identify and coordinate with relevant non-ESnet researchers	4	3	3	250
Establish the use cases that define the requirements and scope - ESnet perimeter firewall: Identify candidate 'trusted' protocols and services; Define Physical scope of 'trusted' portion of ESnet - High Speed Firewalls: Identify commercial or publicly funded research to leverage; Identify the best technology candidates - Self Healing Firewalls: Identify commercial or publicly funded research to leverage; Identify the best technology candidates; Define scope of research consistent with DOE requirements	8	9	6	1,000
Initial technology choices E.g. Identify candidate technologies to pursue based upon detailed analysis of each candidate technology	6	12	3	375
Research and development	6	24	12	1,500
Test and evaluation, initial pilot service rollout - Functional testing on 'test network' - Continue coordination as in 1)	6	30	6	750
Implement the prototype production service, continue pilot service - Coordinate with sites - Continue coordination as in 1) - Develop policies and plans for incident response (DOE sites, ESnet, Adjacent Network Security Staff) - Provide 'user' education as necessary - Develop detailed plan of roll-out - Implement prototype production service	9	36	6	1,125
Deploy first production service	5	42	6	625
Total				**5,625**
Ongoing operation, anual cost E.g. What is the additional cost over the current network services to provide this service?				500

The tables in this section specify the details, including the costs and durations, for the research network roadmap for years 1 through 5. Following the outline of Section 8, we separately list the activities of the two classes: (1) infrastructure and provisioning, and (2) network transport and application support. Note, however, that the transport and application activities critically depend on the provisioning capabilities available at various stages of the roadmap.

E.1 Infrastructure and Provisioning Activities

The following three different options for the research network are considered by the group, and appropriate decisions will be made at different stages in the roadmap.

- Option 1: SONET links: $5.0 M

  Link costs: $3.0M (20 long-haul OC192 links from NLR; 5 long-haul OC192 from other carriers)

  Routers, switches, hosts: $1.0M

  Personnel: $1.0M

- Option 2: DWDM links: $5.0 M

  DWDM links, regen. equipment, and contracts: $3.0 (NLR+others)

  Routers, switches, hosts: $1.0M

  Personnel: $1.0M

- Option 3: Dark fiber: $5.0M

  Fiber links, carrier equipment, and contracts: $3.0 (NLR+others)

  Routers, switches, hosts: $1.0M

  Personnel: $1.0M

In terms of costs and management options, these options are similar. In either case, the research network controls certain routers and switches to implement various provisioning modes. The network services below will be subcontracted, and the details depend on the specifics of the link capabilities (such as dark fiber, DWDM, SONET, etc.)

The following tables are expanded versions of the roadmap outlined in Section 8 for infrastructure and provisioning activities.

Table E-1 Year 1: Establish IP and Lambda Switching Infrastructures at 1-10 Gbps Rates

Year 1 Total Cost: $5M	FTE	Completion Time	Duration	Cost ($M) Equip+FTE
Establish wide-area OC192 links between various DOE sites to span thousands of miles across the country.	0.5	6 months	6 months	2.0+0.125
Install IP routers at the end points and meet points of OC192 links.	0.5	6 months	3 months	1.0+0.125
Install and test packet sniffers at multiple gigabit speeds at end points.	0.5	9 months	3 months	0.5+0.125
Install and test firewalls that operate at Gbps speeds.	0.5	12 months	6 months	0.4+0.125
Install optical switches with the capability of lambda switching at the meet points of OC192 links; establish signaling links to the switches; develop signaling modules for lambda switching.	1.5	12 months	9 months	1.0+0.375
Set up hosts and/or clusters with multiple Gigabit Ethernet Network Interface Cards.	0.5	6 months	6 months	0.1+0.125

Table E-2 Year 2: Establish IP and Lambda Switching Infrastructures at Tens of Gbps

Year 2 Total Cost: $5M	FTE	Completion	Duration	Cost ($M) Equip+FTE
Add additional OC192 and DWDM links between various DOE sites to achieve tens of Gbps bandwidths.	0.5	6 months	6 months	3.0+0.125
Install Multi-Service Provisioning Platform (MSPP) at the meet points of links.	0.5	9 months	3 months	0.5+0.125
Install MSPPs with the capability of sub-lambda switching at end points of links; establish signaling network connections between optical switches, MSPPs, and hosts.	0.5	9 months	3 months	0.4+0.125
Develop scheduler for on-demand sub-lambda provisioning of end-to-end circuits; develop user modules to request the paths; develop modules for path set up and tear down.	1.5	12 months	12 months	0+0.375
Set up hosts and/or clusters with 10 GigE NICs.	0.5	12 months	3 months	0.1+0.125
Install and test packet sniffers and firewalls for IP networks at tens of Gbps.	0.5	12 months	6 months	0+0.125

Table E-3 Year 3: End-to-End Provisioning of Circuits Using Combination of WAN and LAN Environments

Year 3 Total Cost: $5M	FTE	Completion	Duration	Cost ($M) Equip+FTE
Continued cost of lambdas, equipment, and maintenance.	1.5	12 months	12 months	3.0+0.375
Install end equipment at dark fiber LANs to support IP and dedicated circuits.	0.25	6 months	6 months	0.1+0.065
Integrate dark fiber LANs into long-haul IP infrastructure.	0.25	12 months	6 months	0.2+0.065
Integrate dark fiber LANs into sub-lambda by connecting.	0.5	12 months	6 months	0.2+0.125
Integrate the circuit scheduler and the associated setup modules over the signaling network using MSPPs for on-demand provisioning of end-to-end circuits.	1.0	12 months	12 months	0+0.25
Install all optical network switches at the meet points of long-haul links.	0.25	12 months	6 months	0.5+0.065
Install sniffers and firewalls for end-to-end circuits.	0.25	12 months	6 months	0+0.065

Table E-4 Year 4: Multi-resolution Provisioning of Pools of IP and End-to-End Circuits from Desktop to Desktop of Application Users

Year 4 Total Cost: $5M	FTE	Completion	Duration	Cost ($M) Equip+FTE
Continued cost of lambdas, equipment, and maintenance.	1.5	12 months	12 months	4.0+0.375
Enhance the scheduler and set up modules for on-demand provisioning of collection of IP and end-to-end circuits on per application basis.	1.5	6 months	6 months	0+0.375
Update the signaling methods and network to support on-demand provisioning of pools of circuits.	1.0	12 months	6 months	0+0.25

Table E-5 Year 5: On-Demand Provisioning of Multi-resolution Interacting IP and End-to-End Circuits

Year 5 Total Cost: $5M	FTE	Completion	Duration	Cost ($M) Equip+FTE
Continued cost of lambdas, equipment, and maintenance.	1.5	12 months	12 months	4.0+0.375
Enhance the scheduler and set up modules for on-demand provisioning of interacting of IP and end-to-end circuits on per application basis.	1.5	6 months	6 months	0+0.375
Update the signaling methods and network to support on-demand provisioning of interacting circuits for on-line collaborative applications.	1.0	12 months	6 months	0+0.25

E.2 Network Transport and Application Support

The following tables are expanded versions of the roadmap outlined in Section 5 for network transport and application support activities.

Table E-6 Year 1 Network Transport and Application Support Activities

Year 1 Total Cost: $8M	FTE	Completion	Duration	Cost ($M)
Development of high-throughput transport TCP and non-TCP protocols for IP networks to achieve multi-Gbps throughputs.	8.0	6 months	6 months	2.0
Investigation of RDMA, OS bypass, and striping methods for high throughputs for IP connections.	4.0	12 months	6 months	1.0
Comparative analysis and decision making about various high-throughput transport protocols and support technologies for IP networks.	4.0	12 months	6 months	1.0
Execute and demonstrate multi-Gbps transfers under real application environments — pilots.	12.0	12 months	12 months	3.0
Assess the effect of IP sniffers and firewalls on transport throughputs.	4.0	12 months	6 months	1.0

Table E-7 Year 2 Network Transport and Application Support Activities

Year 2 Total Cost: $8M	FTE	Completion	Duration	Cost ($M)
Develop high-throughput transport protocols and associated RDMA, OS bypass, and striping technologies for switched lambda circuits.	8.0	6 months	6 months	2.0
Develop transport protocols for remote visualizations over IP networks.	4.0	12 months	6 months	1.0
Develop transport protocols for remote visualizations over switched sub-lambda circuits.	4.0	12 months	6 months	1.0
Execute and demonstrate visualization under real application environment — pilots.	12.0	12 months	12 months	3.0
Assess the effect of cyber security measures on protocols for both IP and provisioned circuits.	4.0	12 months	6 months	1.0

Table E-8 Year 3 Network Transport and Application Support Activities

Year 3 Total Cost: $8M	FTE	Completion	Duration	Cost ($M)
Develop modules and protocols for computational steering over IP networks and switched lambda circuits.	8.0	12 months	12 months	2.0
Develop modules and protocols for remote instrument control over switched sub-lambda circuits.	8.0	12 months	12 months	2.0
Demonstrate applications requiring computational steering of programs running on supercomputers both on IP and provisioned circuits — pilots.	6.0	12 months	12 months	1.5
Demonstrate applications requiring remote instrument control over provisioned circuits — pilots.	6.0	12 months	12 months	1.5
Assess the impact of firewalls and packet sniffers on the applications requiring visualizations and remote control.	4.0	12 months	6 months	1.0

Table E-9 Year 4 Network Transport and Application Support Activities

Year 4 Total Cost: $8M	FTE	Completion	Duration	Cost ($M)
Develop unified APIs for obtaining and operating pools of IP and provisioned circuits of multiple resolutions.	4.0	6 months	6 months	1.0
Develop optimization modules that match the protocols with the channels on per application basis.	4.0	12 months	12 months	2.0
Test applications requiring pools of circuits to support high-throughput, remote visualization and computational steering — pilots.	12.0	12 months	12 months	3.0
Develop protocols for implementing collaborative channels of various combinations.	4.0	12 months	12 months	1.0
Assess the impact of firewalls and packet sniffers on applications requiring pools of IP and provisioned circuits for visualization and control.	4.0	12 months	6 months	1.0

Table E-10 Year 5 Network Transport and Application Support Activities

Year 5 Total Cost: $8M	FTE	Completion	Duration	Cost ($M)
Develop optimization modules to support interacting IP and dedicated channels for distributed interactive collaborations.	6.0	12 months	12 months	1.5
Test applications requiring interacting circuits to support high-throughput, remote visualization, and computational steering.	6.0	12 months	12 months	1.5
Develop and demonstrate a complete solution suite for a large-science application that requires large data transfers, collaborative visualizations, and steering from participants distributed across the country — research + pilot.	4.0+ 6.0	12 months	12 months	1.0+1.5
Develop and demonstrate a complete solution suite for an application requiring large data transfers and interactive real-time control of an experimental facility from participants distributed across the country — research + pilot.	4.0+ 6.0	12 months	12 months	1.0+1.5

Appendix F: Agenda of the June 2003 DOE Science Networking Workshop

Tuesday, June 3

8:00 Registration and Continental Breakfast

8:30 Welcome, Introductions, Charge, Goals – Gary Johnson (DOE), George Seweryniak (DOE), Larry Price (ANL), Roy Whitney (JLab)

9:00 Driving Science, Including the Office of Science's 20-Year New Facilities Plan – Dan Hitchcock (DOE)

9:30 Report from the High-Performance Network Planning Workshop – Ray Bair (PNNL)

10:00 Report from the Networking Research Workshop – Wu-chun Feng (LANL)

10:30 Break

11:00 Technologies and Services Possibilities – Bill Johnston (LBNL)

11:30 Networking and High-End Computing – Juan Meza (LBNL)

12:00 Working Lunch

12:30 ESnet Today and Future Scenarios – Jim Leighton (LBNL)

1:30 Working Groups – Parallel Sessions

1. Production and High-Impact Network – Vicky White (FNAL) and Dean Williams (LLNL)
2. Research Network – Nagi Rao (ORNL) and Wu-chun Feng (LANL)
3. Technology and Services – Bill Johnston (LBNL) and David Schissel (GA)

3:30 Break

4:00 Parallel Sessions Continue

5:30 Networking and the Current Context – Joel Perriott (OMB)

6:00 Reception

Wednesday, June 4

8:00 Continental Breakfast

8:30 Other Scientific and Educational Networks – Rick Summerhill (Internet2)
9:00 Other Research Networks – Ron Johnson (U. Wash.)

9:30 Non-U.S. Networks – Olivier Martin (CERN)

10:00 Break

10:15 Working Groups – Parallel Sessions

12:00 Working Lunch

12:30 Network Donations – Don Riley (UMd)

1:00 Working Groups - 10 Minute Status Reports

3:30 Break

4:00 Parallel Sessions Continue

5:00 One-Page Summaries/Outlines to Chair

Thursday, June 5

8:00 Continental Breakfast

Working Groups – Reports

8:30 Production and High-Impact Network – Vicky White (FNAL)
and Dean Williams (LLNL)

9:15 Research Network – Wu-chun Feng (LANL) and Nagi Rao (ORNL)

10:00 Break

10:30 Technology and Services – Bill Johnston (LBNL) and David Schissel (GA)

11:15 Wrap-up – Roy Whitney (JLab)

12:00 Adjourn

Appendix G: DOE Science Networking Workshop Participants

1.	Bill Allcock	ANL	allcock@mcs.anl.gov
2.	Guy Almes	Internet2	almes@internet2.edu
3.	Ray Bair	PNNL	raybair@pnl.gov
4.	David Bernholdt	ORNL	bernholdtde@ornl.gov
5.	Scott Bradley	BNL	bradley@bnl.gov
6.	Javad Boroumand	Cisco Systems	javadb@cisco.com
7.	Jeff Candy	General Atomics	candy@fusion.gat.com
8.	Charles Catlett	ANL	catlett@mcs.anl.gov
9.	Helen Chen	SNL	hycsw@ca.sandia.gov
10.	Tim Clifford	Level 3 Communications	tim.clifford@level3.com
11.	Robert Collet	AT&T Govt. Solutions	bobcollet@att.com
12.	Roger Cottrell	SLAC	cottrell@slac.stanford.edu
13.	Robert Cowles	SLAC	bob.cowles@stanford.edu
14.	Gary Crane	SURA	gcrane@sura.org
15.	Glen Crawford	DOE	glen.crawford@science.doe.gov
16.	Pat Dreher	MIT Lab. for Nucl. Sci.	dreher@mit.edu
17.	Richard Egan	ANL	dick.eagan@anl.gov
18.	Wanda Ferrell	DOE	wanda.ferrell@science.doe.gov
19.	Wu Feng	LANL	feng@lanl.gov
20.	Irwin Gaines	Fermilab/DOE	irwin.gaines@science.doe.gov
21.	Bruce Gibbard	BNL	gibbard@bnl.gov
22.	Martin Greenwald	MIT–Plasma Science	g@psfc.mit.edu
23.	Dan Hitchcock	DOE	daniel.hitchcock@science.doe.gov
24.	Gary Johnson	DOE	garyj@er.doe.gov
25.	Ron Johnson	Univ. of Washington	ronj@cac.washington.edu
26.	William Johnston	LBNL	wejohnston@lbl.gov
27.	Kevin Jones	NASA	kevin.m.jones.1@gsfc.nasa.gov
28.	Wesley Kaplow	Qwest Communications	wesley.kaplow@qwest.com
29.	Scott Klasky	PPPL	sklasky@pppl.gov
30.	James Leighton	LBNL	jfl@es.net
31.	Paul Love	Internet2	epl@internet2.edu
32.	Olivier Martin	CERN	olivier.martin@cern.ch
33.	Ed May	ANL	may@anl.gov
34.	Juan Meza	LBNL	jcmeza@lbl.gov
35.	George Michaels	PNNL	george.michaels@pnl.gov
36.	Richard Mount	SLAC	richard.mount@stanford.edu
37.	Shawn McKee	Univ. of Michigan	smckee@umich.edu
38.	Thomas Ndousse	DOE	tndousse@er.doe.gov
39.	Harvey Newmann	Caltech	harvey.newman@cern.ch
40.	Jeff Nichols	ORNL	nicholsja@ornl.gov
41.	Joel Parriott	OMB	jparriot@omb.eop.gov
42.	Donald Petravick	Fermilab	petravick@fnal.gov
43.	Lawrence Price	ANL	lprice@anl.gov

44.	Larry Rahn	SNL	rahn@sandia.gov
45.	Nagi Rao	ORNL	raons@ornl.gov
46.	Anne Richeson	Qwest Communications	anne.richeson@qwest.com
47.	Donald Riley	Univ. of Maryland	drriley@umd.edu
48.	Samtaney Ravi	PPPL	samtaney@pppl.gov
49.	David Schissel	General Atomics	schissel@fusion.gat.com
50.	Volker Schmidt	EFDA	volker.schmidt@igi.cnr.it
51.	Mary Anne Scott	DOE	scott@er.doe.gov
52.	George Seweryniak	DOE	seweryni@er.doe.gov
53.	T.P. Straatsma	PNNL	tps@pnl.gov
54.	Raymond Struble	Level 3 Communications	raymond.struble@level3.com
55.	Rick Summerhill	Internet2	rrsum@greatplains.net
56.	Frank Tapsell	AT&T Govt. Solutions	tapsell@att.com
57.	Albert Thomas	Fermilab	thomas@fnal.gov
58.	Brian Tierney	LBNL	bltierney@lbl.gov
59.	Alan Turnbull	General Atomics	turnbull@fusion.gat.com
60.	Chip Watson	Jefferson Lab	watson@jlab.org
61.	Victoria White	Fermilab	white@fnal.gov
62.	Dean Williams	LLNL	williams13@llnl.gov
63.	Roy Whitney	Jefferson Lab	whitney@jlab.org
64.	Tom Worlton	ANL	tworlton@anl.gov
65.	Mary Fran Yafchak	SURA	maryfran@sura.org
66.	Charles Young	SLAC	young@slac.stanford.edu

Appendix H: Related Workshops and Conferences

This report grew out of the DOE Science Networking Workshop held June 3-5, 2003, by the Energy Sciences Network (ESnet) Steering Committee. The workshop was associated with the following nine other recent and June 2003 workshops, and this report has been influenced by them in varying degrees:

1. DOE Office of Science workshop: Science-Based Case for Large-Scale Simulation (SCaLeS), June 24-24, 2003; http://www.pnl.gov/scales.

2. DOE Science Computing Conference: The Future of High-Performance Computing and Communications, June 19-20, 2003; http://www.doe-sci-comp.info.

3. DOE Workshop on Ultra High-Speed Transport Protocols and Network Provisioning for Large-Science Applications, April 10-11, 2003; http://www.csm.ornl.gov/ghpn/wk2003. Focused on the specific areas of network provisioning and transport that address DOE large-science networking needs.

4. High-Performance Network Planning Workshop, August 13-15, 2002; Report: High-Performance Networks for High-Impact Science; available at: http://DOECollaboratory.pnl.gov/meetings/hpnpw. Identified a number of science areas having high-performance networking and collaboratory needs.

5. NSF Workshop on Ultra-High Capacity Optical Communications and Networking, October 21-22, 2002. Several high-performance network capabilities could be enabled by optical networking technologies; this NSF workshop on that topic was narrow in terms of the technologies considered, but broad in terms of network capabilities.

6. NSF Workshop on Network Research Testbeds, October 17-18, 2002; http://gaia.cs.umass.edu/testbed_workshop. Dealt with developing networks with capabilities that surpass current networks. Both this workshop and the one listed as item 7 focused on broad issues not specific enough to encompass DOE large-science needs.

7. NSF ANIR Workshop on Experimental Infrastructure Networks, May 20-21, 2002; http://www.calit2.net/events/2002/nsf/index.html. (See description for item 5.)

8. NSF CISE Grand Challenges in e-Science Workshop, December 5-6, 2001; http://www.evl.uic.edu/activity/NSF/index.html. Identified cyber infrastructure requirements, including networking technologies, to address the nation's science and engineering needs.

9. Network Modeling and Simulation Program, DARPA; http://www.darpa.mil/ipto/research/nms. Addressed simulation and emulation technologies of large-scale wireless and wireline networks; focused on general aspects of networking for DoD; not specific to scientific needs.

Appendix I: Acronym List

ALS	Advanced Light Source at LBNL
ANL	Argonne National Laboratory
API	Application Programming Interface
APS	Advanced Photon Source at ANL
ASCR	Advanced Scientific Computing Research
ATLAS	A Toroidal LHC ApparatuS at LHC/CERN
ATM	Asynchronous Transfer Mode
BaBar	particle physics experiment at SLAC
BNL	Brookhaven National Laboratory
CANARIE	Canada's Advanced Internet Development Organization
CCC	Circuit Cross Connect
CCSM	Community Climate System Model
CDF	Collider Detector at Fermilab
CEBAF	Continuous Electron Beam Accelerator Facility at JLab
CERN	European Organization for Nuclear Research
CMS	Compact Muon Solenoid detector at LHC/CERN
D0	D-Zero detector at FNAL
DANTE	European science network
DCEE	Distributed Computing Experimental Environment
DICCE	Distributed Informatics, Computing, & Collaborative Environment
DIII	Doublet version III (a General Atomics fusion experiment)
DMX	data management exploration
DOE	U.S. Department of Energy
DOEMICS	DOE Mathematical, Information, and Computational Sciences
DP	DOE Defense Programs
DWDM	dense wavelength division multiplexing
E2E	end to end
EMSL	Environmental Molecular Sciences Laboratory at PNNL
ESCC	ESnet Coordinating Committee
ESnet	Energy Sciences Network
ESSC	ESnet Steering Committee
FES	Fusion Energy Sciences
FNAL	Fermi National Accelerator Laboratory
FTE	full-time equivalent
GB	gigabyte
GEANT	European science network
GeV	billon electron volts
GGF	Global Grid Forum

GTL	Genomes to Life
HEP	high-energy physics
HEPnet	High Energy Physics Network
HR	human resources
IP	Internet Protocol
ISDN	Integrated Services Digital Network
ISP	Internet service provider
ITER	International Thermonuclear Experimental Reactor
JLab	Jefferson Lab, Thomas Jefferson National Accelerator Facility
LBNL	Lawrence Berkeley National Laboratory
LDAP	Lightweight Directory Access Protocol
LHC	Large Hadron Collider at CERN
LIGO	Laser Interferometer Gravitational-Wave Observatory
LSN/MAGIC	Large Scale Networking/Middleware And Grid Infrastructure Coordination
MAN	Metropolitan Area Network
Mbps	Mega bits per second
MFEnet	Magnetic Fusion Energy Network
MICS	Mathematical, Information, and Computational Sciences
MIT	Massachusetts Institute of Technology
MPLS	MultiProtocol Label Switching
MSPP	Multi-Service Provisioning Platform
MUD	Multi-User Domain
NACP	North American Carbon Project
NAT	Network Address Translators
NCAR	National Center for Atmospheric Research
NERSC	National Energy Research Scientific Computing Center at LBNL
NIC	Network Interface Card
NLR	National Lambda Rail
NLS	National Light Source at BNL
NMR	Nuclear Magnetic Resonance
NNSA	National Nuclear Security Administration
NSF	National Science Foundation
NSTX	National Spherical Tokamak Experiment
OGSI	Open Grid Service Infrastructure
ORNL	Oak Ridge National Laboratory
OS	operating system
PB	petabyte
PEP	electron-positron storage rings at SLAC

PKI	Public Key Infrastructure
PKICA	Public Key Infrastructure Certificate Authority
PNNL	Pacific Northwest National Laboratory
POPs	points-of-presence
PPPL	Princeton Plasma Physics Laboratory
QoS	quality of service
R&D	research and development
R&E	research and education
RDMA	Remote Direct Memory Access
RHIC	Relativistic Heavy Ion Collider at BNL
SAML	Security Assertion Markup Language
SC	DOE Office of Science
SCaLeS	Science Case for Large-scale Simulation
SciDAC	Scientific Discovery through Advanced Computing
sec	second
SINET	Science Information Network
SLA	service-level agreement
SLAC	Stanford Linear Accelerator Center
SNS	Spallation Neutron Source at ORNL
SNSC	Science Networking and Services Committee
SOAP	Simple Object Access Protocol
SONET	Synchronous Optical Network
SQL	Structured Query Language
SSRL	Stanford Synchrotron Radiation Laboratory at SLAC
TB	terabyte
TCP	Transport Control Protocol
U.S.	United States
UK	United Kingdom
VNC	Virtual Network Computing
VO	virtual organization
VRVS	Virtual Room Videoconferencing System
WAN	wide-area network
XACML	Extensible Access Control Markup Language
XML	Extensible Markup Language

Prepared for the Office of Advanced Scientific Computing Research
of the U.S. Department of Energy Office of Science

http://www.sc.doe.gov/ascr/