ESnet4: Advanced Networking and Services Supporting the Science Mission of DOE's Office of Science

William E. Johnston
ESnet Dept. Head and Senior Scientist
Lawrence Berkeley National Laboratory
May, 2007

1 Introduction

In many ways, the dramatic achievements in scientific discovery through advanced computing and the discoveries of the increasingly large-scale instruments with their enormous data handling and remote collaboration requirements, have been made possible by accompanying accomplishments in high performance networking. As increasingly advanced supercomputers and experimental research facilities have provided researchers powerful tools with unprecedented capabilities, advancements in networks connecting scientists to these tools have made these research facilities available to broader communities and helped build greater collaboration within these communities. To meet the networking demands of its researchers, the U.S. Department of Energy's Office of Science operates the Energy Sciences Network, or ESnet. Established in 1985, ESnet is managed by Lawrence Berkeley National Laboratory and currently connects tens of thousands of researchers at 27 major DOE research facilities to universities and other research institutions in the US and around the world.

As the single largest supporter of basic research in the physical sciences in the United States, the Office of Science (SC) is the Federal Government's largest single funder of materials and chemical sciences, and it supports unique and vital parts of U.S. research in climate change, geophysics, genomics, life sciences, and science education. In FY2008 SC will support 25,500 PhDs, PostDocs, and Graduate students and 21,500 users of SC facilities, half of which come from universities. [1] To ensure that ESnet continues to meet the requirements of the major science disciplines supported by the Office of Science, a series of workshops were held over the past several years to examine these networking and middleware requirements. Participants concluded that modern large-scale science requires networking that is global in extent, extremely reliable, adaptable to changing requirements, capable of providing bandwidth bounded only by the latest technology, and able to support large volumes of sustained traffic. These requirements have resulted in a new approach and architecture for ESnet. This new architecture includes elements supporting multiple, high-speed national backbones with different characteristics, redundancy, quality of service and circuit oriented services, all the while allowing interoperation of these elements with the other major national and international networks supporting science. The approach is similar to, and designed to be compatible with, other research and education networks such as Internet2 in the United States and DANTE/GÉANT in Europe.

ESnet's mission is to provide an interoperable, effective, reliable, high performance network communications infrastructure, along with selected leading-edge Grid-related and collaboration services in support of SC's large-scale, collaborative science. ESnet must provide services that enable the SC science programs that depend on:

- Sharing of massive amounts of data
- Supporting thousands of collaborators world-wide
- Distributed data processing
- Distributed data management
- Distributed simulation, visualization, and computational steering
- Collaboration with the U.S. and international research and education community

To this end, ESnet provides network and collaboration services to SC laboratories, and also serves programs in most other parts of DOE.

What Is ESnet

ESnet is:

- A large-scale IP network built on a national circuit infrastructure with high-speed connections to all major US and international research and education (R&E) networks
- An organization of 30 professionals structured for the service
- An operating entity with an FY06 budget of $26.6M
- A tier 1 ISP providing direct peerings will all major networks – commercial, government, and research and education (R&E)
- The primary DOE network providing production Internet service to almost all of the DOE labs and most other DOE sites. This results in ESnet providing an estimated 50,000 - 100,000 DOE users and more than 18,000 non-DOE researchers from universities, other government agencies, and private industry that use SC facilities with global Internet access.

ESnet's Place in U. S. and International Science

A large fraction of all of the national data traffic supporting U.S. science is carried by three networks – ESnet and Internet2, and National Lambda Rail. These three entities represent the architectural scope of science oriented networks.

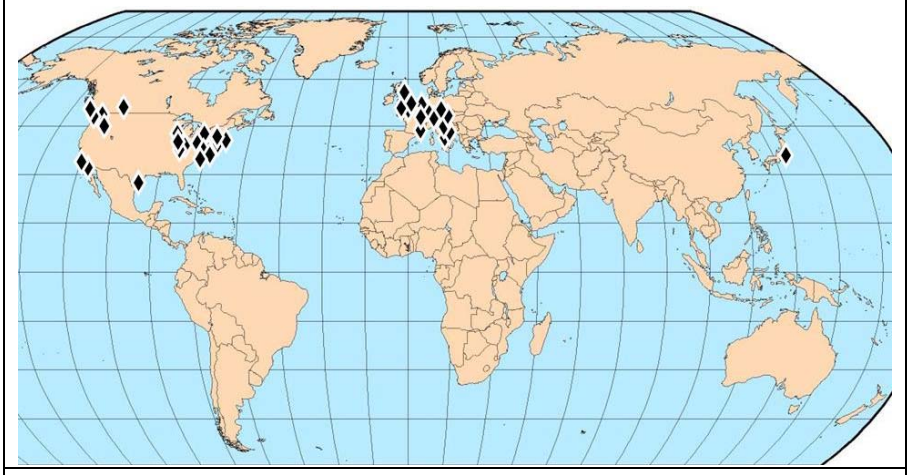

Figure 1. The large-scale data flows in ESnet reflect the scope of Office of Science collaborations

ESnet's top 100 data flows generate 50% of all ESnet traffic (ESnet handles about $3x10^9$ flows/mo.) 91 of the top 100 flows are from the DOE Labs (not shown) to other R&E institutions (shown on the map) (CY2005 data)

ESnet is a network in the traditional sense of the word. It connects end user sites to various other networks. Internet2 is a backbone network connecting U.S. regional networks to each other and international networks. NLR is a collection of light paths or lambda channels that are used to construct specialized R&E networks.

ESnet serves a community of directly connected campuses – the Office of Science labs; in essence ESnet interconnects the LANs of all of the labs to the outside world. ESnet also provides the peering and routing needed for the labs to have access to the global Internet. Internet2 serves a community of regional networks that connect university campuses. These regional networks – NYSERNet (U.S. northeast), SURAnet (U.S. southeast), CENIC (California), etc., – have regional aggregation points called GigaPoPs and Internet2 interconnects the GigaPoPs. Internet2 is mostly a transit network – the universities and/or the regional networks provide the peering and routing for end-user Internet access. This is also very similar to the situation in Europe where GÉANT (like Internet2) interconnects the European National Research and Education Networks (NRENs) that in turn connect to the LANs of the European science and education institutions. (The NRENs are like the US regionals, but organized around the European nation-states).

The top level networks – ESnet, Internet2, GÉANT, etc. – work closely together to ensure that they have adequate connectivity with each other so that all of the connected institutions have high-speed end-to-end connectivity to support their science and education missions. ESnet and Internet2 have had joint engineering meetings for several years and ESnet, Internet2, GÉANT, and CERN have also formed an international engineering team that meets several times a year.

An ESnet goal is that connectivity from a DOE lab to US and European R&E institutions should be as good as lab-to-lab and university-to-university connectivity. The key to ensuring this is engineering, operations, and constant monitoring. ESnet has worked with the Internet2 and the international R&E community to establish a suite of monitors that can be use to provide a full mesh of paths that continuously checks all of the major interconnection points.

2 Evolving Science Environments Drive the Design of the Next Generation ESnet

Large-scale collaborative science – big facilities, massive amount of data, thousands of collaborators – is a key element of DOE's Office of Science. The science community that participates in DOE's large collaborations and facilities is almost equally split between SC labs and universities, and has a significant international component. Very large international facilities (e.g., the LHC particle accelerator at CERN in Switzerland and the ITER experimental fusion reactor being built in France) and international collaborators participating in U.S.-based experiments are now also key elements of SC science, requiring the movement of massive amounts of data between the SC labs and these international facilities and collaborators. Distributed computing and storage systems for data analysis, simulations, instrument operation, etc., are becoming common; and for data analysis in particular, Grid-style distributed systems predominate. (See, e.g., the Open Science Grid – an SC led distributed Grid computing project – http://www.opensciencegrid.org/)

This science environment is very different from that of a few years ago and places substantial new demands on the network. High-speed, highly reliable connectivity between labs and U.S. and international R&E institutions is required to support the inherently collaborative, global nature of large-scale science. Increased capacity is needed to accommodate a large and steadily increasing amount of data that must traverse the network to get from instruments to scientists and to analysis, simulation, and storage facilities. High network reliability is required for interconnecting components of distributed large-scale science computing and data systems and to support various modes of remote instrument operation. New network services are needed to provide bandwidth guarantees for data transfer deadlines, remote data analysis, real-time interaction with instruments, coupled computational simulations, etc.

There are many stakeholders for ESnet. Foremost are the science program offices of the Office of Science (Advanced Scientific Computing Research, Basic Energy Sciences, Biological and Environmental Research, Fusion Energy Sciences, High Energy Physics, and Nuclear Physics – see http://www.science.doe.gov/). ESnet also serves labs and facilities of other DOE offices (e.g., Energy Efficiency and Renewable Energy, Environmental Management, National Nuclear Security Administration, and Nuclear Energy, Science and Technology). Other ESnet stakeholders include SC-supported scientists and collaborators at non-DOE R&E institutions (85% of all ESnet traffic comes from or goes out to non-DOE R&E organizations), and from the networking organizations that provide networking for these non-DOE institutions.

Requirements of the ESnet stakeholders are primarily determined by three approaches. Instruments and facilities that will be coming on-line over the next 5–10 years and will connect to ESnet (or deliver data to ESnet sites in the case of LHC and IETR) are characterized by considering the nature of the data that will be generated and how and where it will be stored, analyzed, and used. The process of science in the disciplines of direct interest to SC is examined to determine how the process of that science will change over the next 5-10 years and how these changes will drive demand for new network capacity, connectivity, and services. Finally, ESnet traffic patterns are analyzed based on the use of the network in

the past 2-5 years to determine the trends, and then projecting how the network must change to accommodate the future traffic patterns implied by these trends.

(SIDEBAR 1) A Case Study: The Data Analysis for the Large Hadron Collider

The major high energy physics (HEP) experiments of the next 20 years will break new ground in our understanding of the fundamental interactions, structures and symmetries that govern the nature of matter and space-time. Among the principal goals are to find the mechanism responsible for mass in the universe, and the "Higgs" particles associated with mass generation, as well as the fundamental mechanism that led to the predominance of matter over antimatter in the observable cosmos.

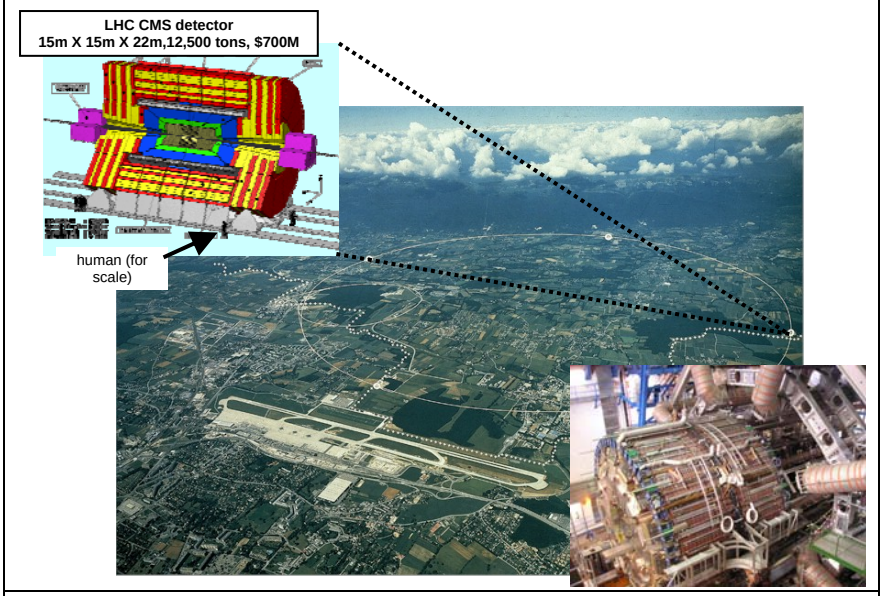

Figure 2. The Large Hadron Collider at CERN

An aerial vire of CERN and a graphic showing one of the two large experiments (the CMS detector). The LHC ring is 27 km circumference (8.6 km diameter) and provides two counter-rotating, 7 TeV proton beams collide in the middle of the detectors. (Images courtesy CERN.)

The largest collaborations today, such as CMS [11] and ATLAS [12] that are building experiments for CERN's Large Hadron Collider program (LHC [13]), each encompass some 2,000 physicists from 150 institutions in more than 30 countries. The current generation of operational experiments at Stanford Linear Accelerator Center (SLAC) (BaBar [14]) and Fermilab (D0 [15] and CDF [15]), as well as the experiments at the Relativistic Heavy Ion Collider (RHIC, [17]) program at Brookhaven National Lab, face similar challenges. BaBar, for example, has already accumulated datasets approaching a petabyte.

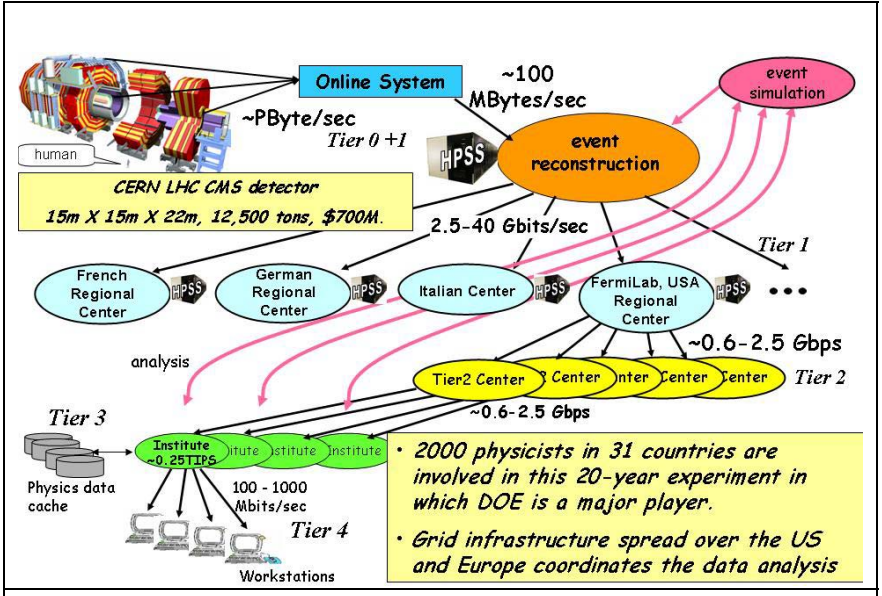

Figure 3. High Energy Physics Data Analysis

This science application epitomizes the need for collaboratories supported by Grid computing infrastructure in order to enable new directions in scientific research and discovery. The CMS situation depicted here is very similar to Atlas and other HEP experiments. These experiments will each be collecting, cataloguing, and analyzing several petabytes/year by 2008-2009. (Adapted from original graphic courtesy Harvey B. Newman, Caltech.)

The HEP (or HENP, for high energy and nuclear physics) problems are among the most data-intensive known. Hundreds to thousands of scientist-developers around the world continually develop software to better

select candidate physics signals from particle accelerator experiments such as CMS, better calibrate the detector and better reconstruct the quantities of interest (energies and decay vertices of particles such as electrons, photons and muons, as well as jets of particles from quarks and gluons). These are the basic experimental results that are used to compare theory and experiment. The globally distributed ensemble of computing and data facilities (e.g., see Figure 1), while large by any standard, is less than the physicists require to do their work in an unbridled way. There is thus a need, and a drive, to solve the problem of managing global resources in an optimal way in order to maximize the potential of the major experiments to produce breakthrough discoveries.

Collaborations on this global scale would not have been attempted if the physicists could not plan on high capacity networks: to interconnect the physics groups throughout the lifecycle of the experiment, and to make possible the construction of Data Grids capable of providing access, processing and analysis of massive datasets. These datasets will increase in size from petabytes to exabytes (10^{18} bytes) within the next decade. Equally as important is highly capable middleware (the Grid data management and underlying resource access and management services) to facilitate the management of world wide computing and data resources that must all be brought to bear on the data analysis problem of HEP [6].

(END OF SIDEBAR)

Requirements from Data and Collaboration Characteristics of Instruments, Facilities, and Science Practice

There are some 20 major instruments and facilities currently operated or being built by SC [1], plus the LHC at CERN and ITER in France. To date, ESnet has characterized 14 of these for their future requirements. DOE facilities such as the big accelerators, RHIC at Brookhaven and the Spallation Neutron Source at Oak Ridge National Laboratory, and the SC supercomputer centers (NERSC at Lawrence Berkeley, NLCF at Oak Ridge, and ALCF at Argonne), as well as the LHC at CERN, are typical of the hardware infrastructure of SC science. These facilities generate four types of network requirements: bandwidth, connectivity and geographic footprint, reliability, and network services.

In order to determine the requirements of SC science based on how the process of conducting scientific research will change, a set of case studies were developed in which the science communities were asked to describe how they expected to have to be doing their science in five and ten years in order to make significant progress. Computer scientists then worked with the scientists to translate the new processes into network requirements – in particular those related to collaboration, data sharing and remote analysis, remote instrument control, and large-scale simulations coupled with each other and/or with external sources of data (e.g., operating instruments).

Bandwidth needs are determined by the quantity of data produced and the need to move the data for remote analysis. Connectivity and geographic footprint are determined by the location of the instruments and facilities, and the locations of the associated collaborative community, including remote and/or distributed computing and storage used in the analysis systems. These locations also establish requirements for connectivity to the network infrastructure that supports the collaborators (e.g., ESnet connectivity to Internet2 and the US regional R&E networks, and GÉANT and the European national R&E networks – the NRENs).

The reliability requirements are driven by how closely coupled the facility is with remote resources. For example, off-line data analysis – where an experiment runs and generates data and the data is analyzed after the fact – may be tolerant of some level of network outages. On the other hand, when remote operation or analysis must occur within the operating cycle time of an experiment ("on-line" analysis, e.g., in magnetic fusion experiments), or when other critical components depend on the connection (e.g., a distributed file system between supercomputer centers), then very little network downtime is acceptable (see Table 1). The reliability issue is critical and drives much of the design of the network. Many scientific facilities in which DOE has invested hundreds of millions to billions of dollars, together with their large associated science communities, are heavily dependent on networking. Not surprisingly, when

the experiments of these facilities depend on the network, then these facilities and scientists demand that the network provide very high availability (99.99+%), in addition to very high bandwidth.

The fourth requirement is in the area of types of service. In the past, networks typically provided a single network service – best-effort delivery of data packets[a] – on which are built all of today's higher-level applications (FTP, email, Web, socket libraries for application-to-application communication, etc.), and best-effort IP multicast (where a single outgoing packet is, sometimes unreliably, delivered to multiple receivers). In considering future uses of the network by the science community, several other network services have been identified as requirements, including bandwidth guarantees[b], traffic isolation[c], and reliable multicast.

Bandwidth guarantees are typically needed for on-line analysis, which always involves time constraints. Another type of application requiring bandwidth guarantees is distributed workflow systems such as those used by high energy physics data analysis. The inability of one element (computer) in the workflow system to adequately communicate data to another will ripple through the entire workflow environment, slowing down other participating systems as they wait for required intermediate results, thus reducing the overall effectiveness of the entire system.

Traffic isolation is required because today's primary transport mechanism – TCP – is not ideal for transporting large amounts of data across large (e.g., intercontinental) distances. There are protocols better suited to this task, but these protocols are not compatible with the fair-sharing of TCP transport in a best-effort network, and are thus typically penalized by the network in ways that reduce their effectiveness. A service that can isolate the bulk data transport protocols from best-effort traffic is needed to address this problem.

Reliable multicast is a service that, while not entirely new, must be enhanced to increase its effectiveness. Multicast provides for delivering a single data stream to multiple destinations without having to replicate the entire stream at the source, as is the case, e.g., when using a separate TCP-based connection from the source to each receiver. This is important when the data to be delivered to multiple sites is too voluminous to be replicated at the source and sent to each receiving site individually. Today, IP multicast provides this capability in a fragile and limited way (IP multicast does not provide reliable delivery as TCP-based transport does). New services may be required to support reliable and robust multicast.

---

[a] Packet management by IP networks is not deterministic, but rather statistical. That is, IP packets are injected into the network from many computers that are all connected to a single router – e.g. a typical large SC Lab will have many internal "subnets" all of which connect through different interfaces to a single site gateway router that provides connectivity to the outside world. The packets are queued in the router in whatever order they reach the routing processor (also called the forwarding processor). The packets in the queue waiting to be forwarded to their next-hop destination are intermixed indiscriminately by virtue of being queued immediately from several different input connections. As long as the queue does not overflow this is not an issue (in fact is the norm) since every packet is routed through the network independently of every other packet. If the packets come into a router through several interfaces and they are all processed out through a single interface – as is typical, e.g., for a site gateway router that has several connections on the site side and a single connection on the Wide Area Network side – then it is possible for the forwarding processor to fall behind. This can happen either because the forwarding processor is not fast enough to keep up with the routing (which is rare in modern routers) or because the aggregate input traffic bandwidth exceeds the bandwidth of the single output interface (a circumstance that, in principle, is easily realized). When this happens the input queue for the forwarding engine will fill and "overflow" – this is called network congestion. The overflow process is a random discard of the incoming packets, and the overall effect is that there is no guarantee that a packet sent to a router is forwarded on to its next hop toward its destination – packet forwarding is a "best-effort" process. (Users typically see congestion as a slowdown in the network – they do not see the packet loss directly because most applications use TCP as a reliable transport protocol. TCP uses IP packets to move data through the network and it detects packet loss and automatically resends the lost IP packets in order to ensure reliable data delivery.)

[b] Bandwidth guarantees are provided in IP networks by doing two things: First, the packets in a bandwidth-guaranteed connection are marked as high priority and are forwarded ahead of any waiting best-effort packet. Second, the bandwidth-guaranteed connections are managed so that they can never exceed the available bandwidth anywhere in the path to their destination. This entails limiting the input bandwidth of a bandwidth-guaranteed connection to an agreed upon value, and then by limiting the number of such connections so as not to exceed the available bandwidth along the path.

[c] Traffic isolation is provided in a way similar to bandwidth guarantees in that the packets are queued and forwarded in such a way that they do not interact with other classes of traffic such as best-effort.

In the case studies that have been done to date [5], one or more major SC facilities have identified a requirement for each of these network capabilities.

The case studies of [2], [4], and [5] were picked both to get a good cross-section of SC science and to provide realistic predictions based on highly probable changes in the scientific process in the future. The case studies were conducted over several years and included the following Office of Science programs and associated facilities: Magnetic Fusion Energy, NERSC, ACLF, NLCF, Nuclear Physics (RHIC), Spallation Neutron Source, Advanced Light Source, Bioinformatics, Chemistry / Combustion, Climate Science, and High Energy Physics (LHC).

Summary of the conclusions of the case studies

There is a high level of correlation between network requirements for large and small scale science – the primary difference being bandwidth – and so meeting the requirements of the large-scale stakeholders will generally provide for the requirements of the smaller ones, provided the required services set is the same.

Some of the non-bandwidth findings from the case studies included:

- The geographic extent and size of the user base of scientific collaboration is continuously expanding. As noted, DOE US and international collaborators rely on ESnet to reach DOE facilities, and DOE scientists rely on ESnet to reach non-DOE facilities nationally and internationally (e.g., LHC, ITER). Therefore, close collaboration with other networks is essential in order to provide high-quality end-to-end service, diagnostic transparency, etc.
- Robustness and stability (network reliability) are essential. Large-scale investment in science facilities and experiments makes network failure unacceptable when the experiments depend on the network.
- Science requires several advanced network services for different purposes. There are requirements for predictable latency and quality of service guarantees to support remote real-time instrument control, computational steering, and interactive visualization. Bandwidth guarantees and traffic isolation are needed for large data transfers (potentially using TCP-unfriendly protocols), and network support for deadline scheduling of data transfers.

The aggregation of requirements from the 14 case studies (see [5]) results in:

- Reliability
  - The Fusion requirements of 1 minute of down time during an experiment that runs 8–16 hours a day, 5–7 days a week, implies a network availability of 99.999%. LHC data transfers can only tolerate a small number of hours of outage in streams that operate continuously for 9 months per year, otherwise the analysis of the data coming from the LHC will fall too far behind to ever catch up. This implies a network availability of 99.95%.
  - These needs result in a requirement for redundancy (which is the only practical way to achieve this level of reliability) both for site connectivity and within ESnet.
- Connectivity
  - The geographic reach of the network must be equivalent to that of the scientific collaboration. Multiple peerings with the other major R&E networks are needed to add reliability and bandwidth for inter-domain connectivity. This is critical both within the US and internationally.
- Bandwidth
  - A bandwidth of 10 Gb/s site-to-site connectivity is needed now, and 100 Gb/s will be needed by 2010. Multiple 10 Gb/s peerings (interconnections) with the major R&E networks will be needed for data transfers. The network must have the ability to easily deploy additional 10 Gb/s circuits and peerings as needed by new science projects.

- Bandwidth and service guarantees are needed end-to-end, so all R&E networks must interoperate as one seamless fabric. Flexible rate bandwidth guarantees are needed – that is, a project must be able to ask for the amount of bandwidth that it needs and not be forced to use more or less.

The case studies include both quantitative and qualitative requirements.

Requirements from Observing Traffic Patterns

From the analysis of historical traffic patterns, several clear trends emerge that result in requirements for the evolution of the network so it can deal with the projected impact of the trends.

The first and most obvious pattern is the exponential growth of the total traffic handled by ESnet (Figure 4 and Figure 5). This traffic trend represents a 10x increase every 47 months on average since 1990 (Figure 5). ESnet traffic just passed the 1 petabyte per month level with about 1.5 Gb/s average, steady-state load on the New York-Chicago-San Francisco path. If this trend continues (and all indications are that it will accelerate), the network must be provisioned to handle an average of 15 Gb/s in four years. This implies a minimum backbone bandwidth of 20 Gb/s, because the network peak capacity must be at least 40% higher than the average load in order for today's protocols to function properly with bursty traffic (which is the norm). In addition, the current traffic trend suggests that 200 Gb/s of core network bandwidth will be required in eight years. This can only be achieved within a reasonable budget by using a network architecture and implementation approach that allows for cost-effective scaling of hub to hub circuit bandwidth.

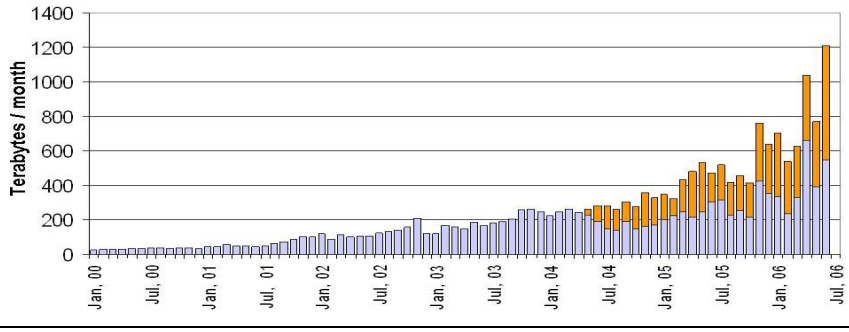

Figure 4. Total ESnet traffic by month, 2000–2006.

The segmented bars from mid-2004 on show that fraction of the total traffic in the top 1000 data flows (which are from large-scale science facilities). (There are typically several billion flows per month in total, most of which are minuscule compared to the top 1000 flows.)

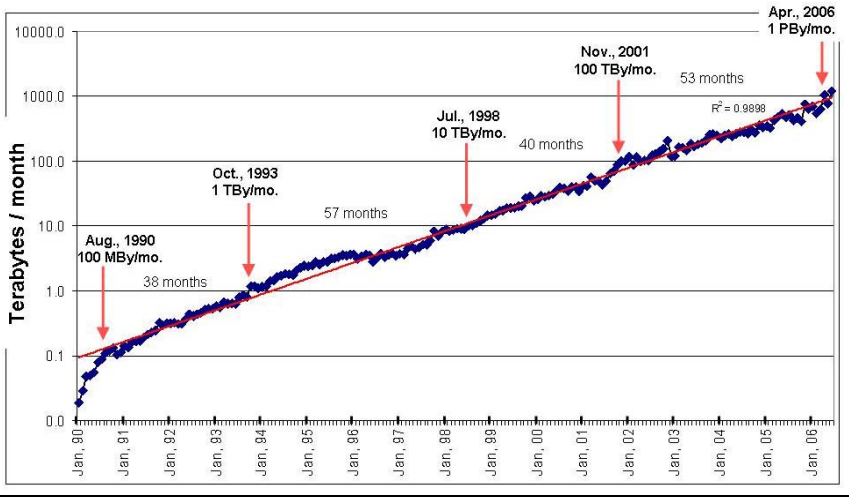

Figure 5. Log plot of ESnet traffic since 1990.

The second major change in traffic is the result of a dramatic increase in the use of parallel file mover applications (e.g., GridFTP). This has resulted in the most profound change in traffic patterns in the history of ESnet. Over the past 18 months, this has resulted in a change from the historical trend where the peak system-to-system ("workflow") bandwidth of the largest network users increased along with the increases in total network traffic, to a situation where the peak bandwidth of the largest user systems is coming down, and the number of flows that they generate is going up, while the total traffic continues to increase exponentially. This reduction in peak workflow bandwidth, together with an overall increase in bandwidth, is the result of the decomposition of single large flows into many smaller parallel flows. In

other words, the same types of changes that happened in computational algorithms as parallel computing systems became prevalent are now happening in data movement – that is, parallel I/O channels operating across the network. This is illustrated in Figure 6, where the top 100 host-to-host data transfers, in one month averages, for a sampling of months over the past 18 months, are represented in the bar charts labeled "Host to Host Traffic." Next to these graphs is the total network traffic for that month, segmented as in Figure 4.

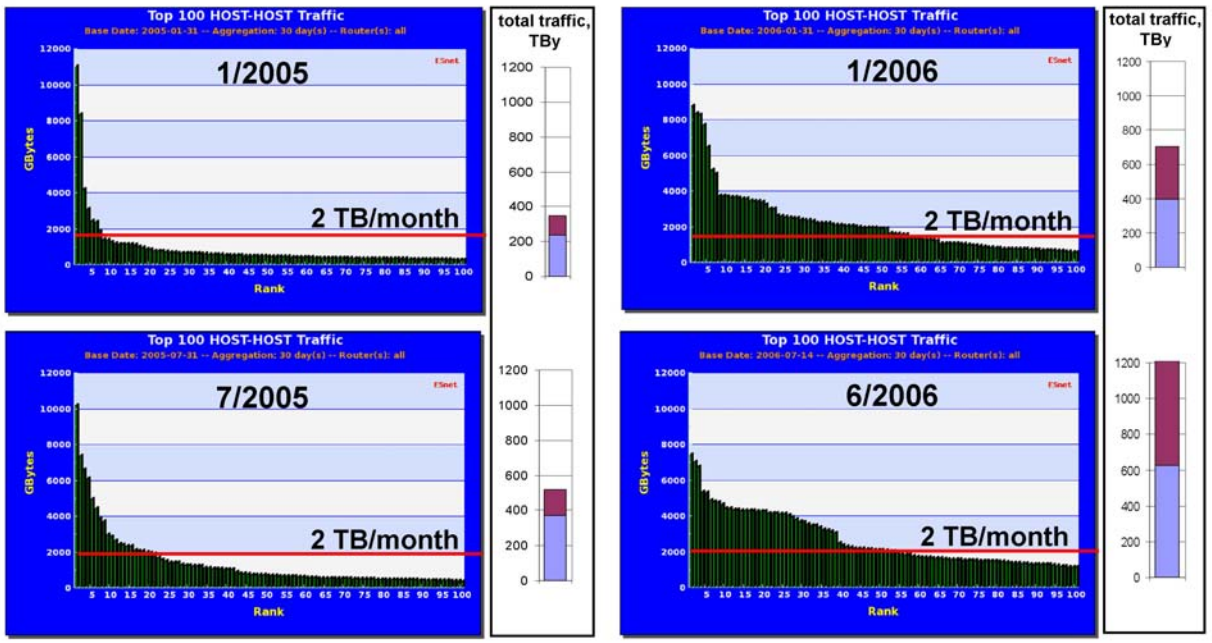

Figure 6. ESnet's traffic patterns are evolving due to increasing use of parallel file movers.

The third clear traffic trend is that over the past two years the impact of the top few hundred workflows – there are of order 6×10^9 flows per month in total – has grown from negligible before mid-2004 to more than 50% of all traffic in ESnet by mid-2006! This is illustrated in Figure 4, where the top part of the traffic bars shows the portion of the total generated by the top 100 hosts.

The fourth significant pattern comes from looking at the source and destination locations of the top data transfer systems – an examination that shows two things. First is that the vast majority of the transfers can easily be identified as science traffic since the transfers are between two scientific institutions with systems that are named in ways that reflect the name of the science group. Second, for the past several years the majority of the large data transfers have been between institutions in the US and Europe and Japan, reflecting the strongly international character of large science collaborations organized around large scientific instruments (Figure 7).

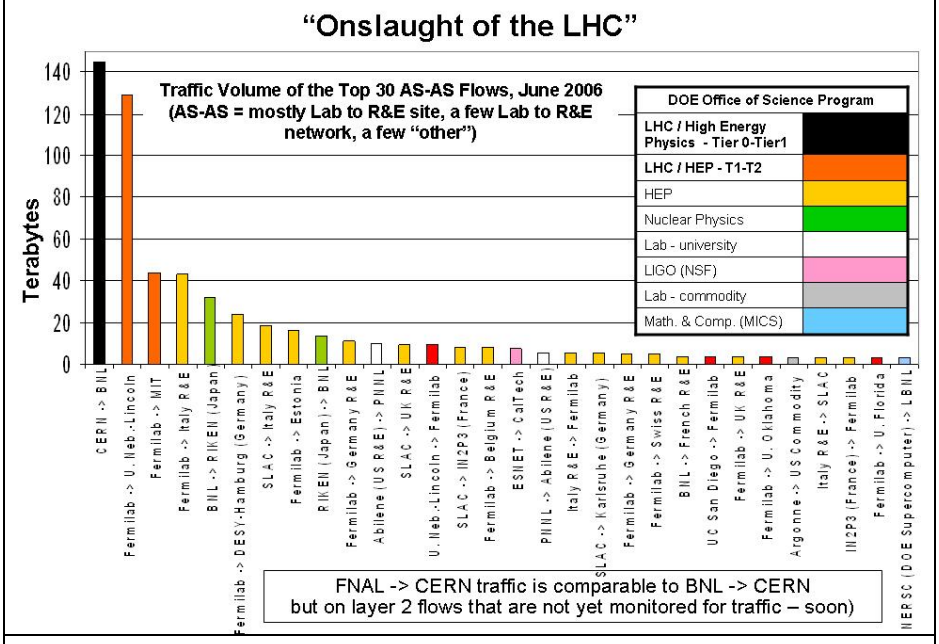

Figure 7. Traffic patterns due to new uses of the network by the LHC. LHC to BNL is the No. 1 traffic generator; FNAL to and from US universities accounts for Nos. 2, 3, 13, 23, 24, and 28.

Finally, Figure 7– only somewhat jokingly entitled "Onslaught of the LHC" – also illustrates the limitation of using traffic trends alone to predict the future network needs of science. No traffic observations could have predicted the upsurge in LHC data movement, both from CERN to the SC labs and from the SC labs to US universities. Obviously traffic trend analysis cannot predict the start of new science projects.

Network Requirements Summary

The combination of the case studies and the traffic pattern trends adds quantitative aspects to the general requirements that were identified early in this paper.

The aggregate network capacity must reach 100–200 Gb/s in the five- to seven-year time frame. Network reliability must increase from the historical 99.9% to 99.99% to something more like 99.99% to 99.999% availability to the end site. The peerings – external network interconnections between national R&E and international R&E networks and ESnet – must increase both in bandwidth and reliability in a similar fashion.

Several specific new network services related to bandwidth guarantees must be introduced into the production network.

A general requirement is that there must be flexibility in provisioning the network capacity. The location of the greatest need for bandwidth within the network will change over time, and the budgetary resources available for the network may also change. It must be possible add and move hub-to-hub capacity as needed and to deploy new capacity on a schedule determined by science needs and funding availability.

3 What Does ESnet Provide?

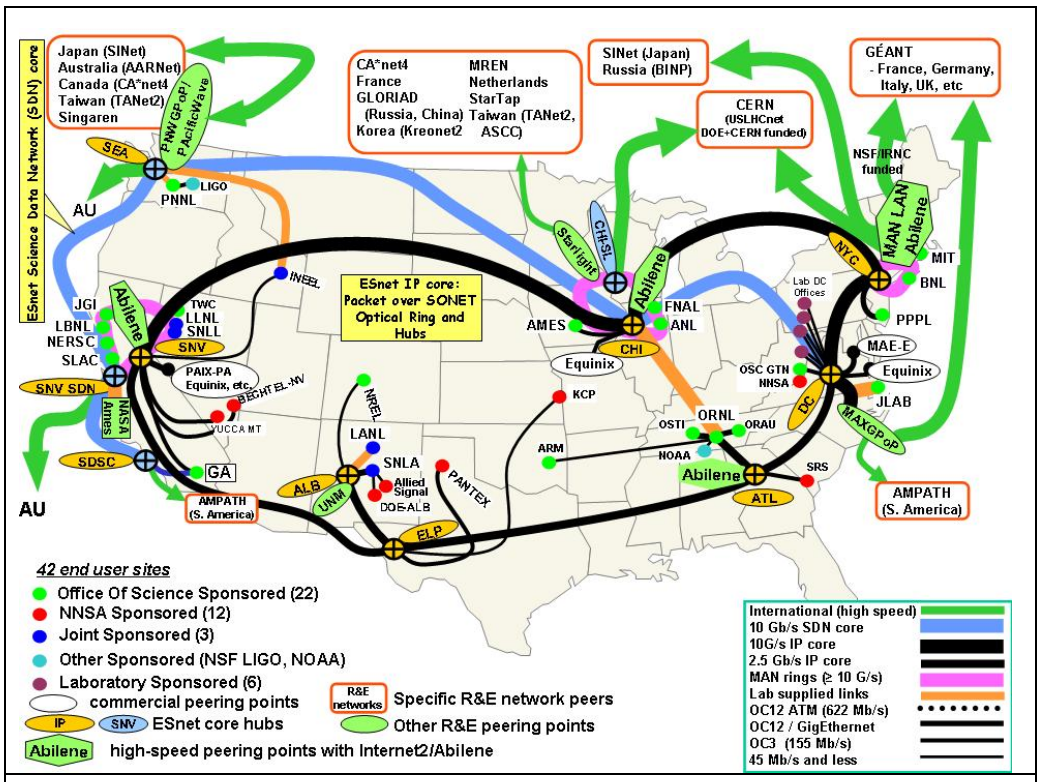

Figure 8. ESnet3

Today's networks provide global high-speed Internet connectivity for DOE facilities and collaborators (ESnet in early 2007).

All three of the research and education communities mentioned here – those associated with ESnet, Internet2 and the U.S. regional nets, and GÉANT and the NRENS – serve institutions that have requirements for, and access to, a common set of Internet services. However, as mentioned, exactly which of the several networks associated with a given institution provides the services varies. ESnet, while it has a substantial user base (50,000-100,000 users in the 42 served sites), is small compared to the U.S. higher education community served by Internet2 or the overall European R&E community served by GÉANT. However, ESnet provides a convenient microcosm in which to describe the services provided by various network organizations to the scientific community. (While not excluding the education community, ESnet does not serve that community and the services ESnet provides to its customer base may not map one to one with services offered to higher education institutes.)

One of the characteristics of science oriented networks is that they must provide a relatively small number of sites with very large amount of bandwidth. (As opposed to commodity ISPs like AOL or EarthLink which are tailored to provide a huge number of users a relatively small amount of bandwidth.) In particular, ESnet must provide high bandwidth access to DOE sites and to DOE's primary science collaborators in the science community. This is accomplished by a combination of high-speed dedicated circuits that connect the end sites and by high-speed peerings with the major R&E network partners.

ESnet builds and operates a comprehensive IP network infrastructure (IPv4, IP multicast, and IPv6, peering, routing, and address space management) based on commercial and R&E community circuits. The current physical architecture is shown in Figure 8 which illustrates the extent and diversity of the circuits.

ESnet provides full and carefully optimized access to the global Internet. This is essential, as mentioned above, for the best possible access to the sites where collaborators are located. In order to accomplish this ESnet has peering agreements with many commercial and non-commercial networking. Those agreements result in routes (reachability information) being exchanged between all of the networks needed to provide comprehensive (global) Internet site access.

As noted above, in order to provide DOE scientists access to all Internet sites, ESnet manages the full complement of Global Internet routes. This requires about 160,000 IPv4 routes from 180 peers. (The peering policy mentioned above selects these 160,000 routes from about 400,000 that are offered at all of the peering points.) These peers are connected at 40 general peering points that include commercial, research and education, and international networks. With a few of ESnet's most important partner networks (notably Internet2 and GÉANT), direct peering (core router to core router) is done to provide high performance.

How the Network is Operated

The normal operation of a network like ESnet involves monitoring both the state of the logical network (connectivity to the rest of the Internet) and the state of the physical network (the operational status of the network links, switches, routers, etc.).

Managing the logical network entails ensuring that there are paths from the systems at the DOE labs to every other system connected to the Internet. This is accomplished by having a comprehensive set of routes to all of the active IP address space through the peering process described above. Managing these routes in order to provide high quality access to the global Internet is a never ending task because ISPs come and go and change their relationship to other ISPs, etc. Automated tools to control which routes are accepted from specific peers help keep this process maintainable.

The physical network is managed largely through extensive, continuous monitoring. The eleven hubs and 42 end sites are monitored minute by minute at more than 4500 physical and logical interfaces. This includes every aspect of the operating state of the equipment and the traffic flowing over every interface. All of this information is simultaneously analyzed by a network monitoring system and entered into a database that is accessible to the ESnet engineers at various locations around the country.

Scalable Operation is Essential

R&E networks like ESnet are typically operated with a small staff. The key to this is that everything related to the operation of the network and related services must be scalable. The question of how to manage a huge infrastructure with a small number of people dominates all other issues when looking at whether to support new services (e.g. Grid middleware): Can the service be structured so that its operational aspects do not scale as a function of the user population? If not, then the service cannot be offered.

In the case of the network itself, automated, real-time monitoring of traffic levels and operating state of some 4500 network entities is the primary network operational and diagnosis tool. Much of the analysis of this information (generated in real-time with sample intervals as short as minutes or generated asynchronously as alarms) is automatically analyzed and catalogued as to normal or abnormal, urgent or not. Urgent abnormal events filter up through a hierarchy of operational and engineering staff. The entire ESnet network is operated 24x7x365 by about 16 people.

What Does the Network Actually Look Like?

The ESnet3 core consists of 11 hubs and sub-hubs. A typical ESnet hub is illustrated in Figure 9, though this is actually the new ESnet4 hub in Washington, DC – "WDC." The core routers have the primary job of high-speed forwarding of packets. They have the high-speed interfaces for the 2.5 and 10 Gb/s cross-country circuits (at WDC there are several 10 Gb/s circuits to New York and one 2.5 Gb/s circuit to Atlanta).

There are also several direct connections for high-speed peerings on the core router. At WDC this includes a 10Gb/s connection to MAX (Mid-Atlantic Exchange) that in turn provides to several R&E networks and universities and to GÉANT; a 10Gb/s connection to Internet2, and a 10 Gb/s circuit to Jefferson Lab, an ESnet site. Most ESnet hubs also have a peering router that connects to the core router and to the commercial peers that happen to have presence in that hub (the 'IP peering router" in Figure 9). The separation of the peering function from the core routing function simplifies management and allow for a more effective cyber security stance.

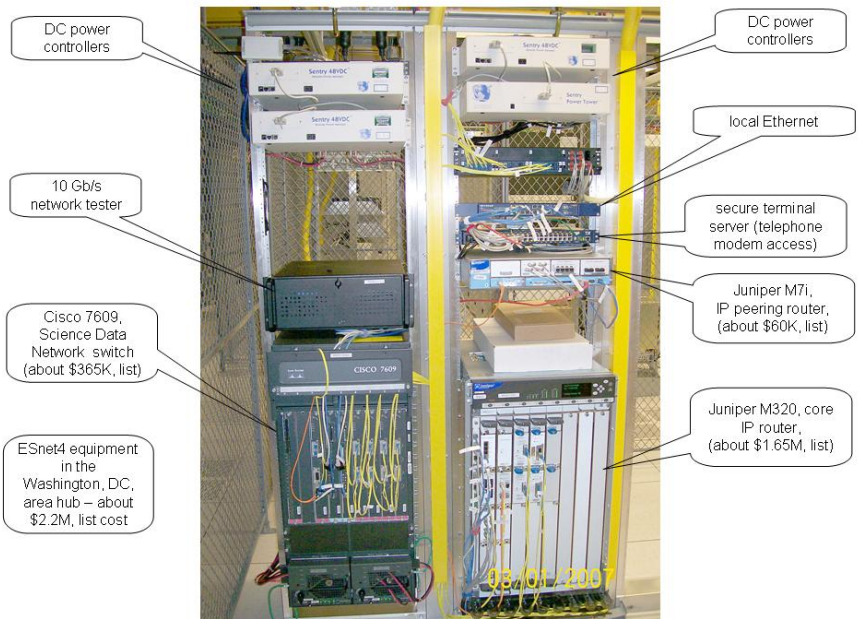

Figure 9. The ESnet4 network equipment of a typical, medium-size, ESnet hub (Washington, DC – "WDC")

Supporting and auxiliary equipment consists of a secure terminal server that provides access of last resort by telephone modem, a power controller that allows for remote power cycling of all of the other equipment, and one or more performance testing systems. At WDC ESnet has one type of performance testing systems: The Performance Center systems provide for interactive diagnostics and are available to ESnet engineers, to site network engineers, and to end users. The Performance Center platform supports various perfSONAR (see section 0, below) services for user and network operator performance testing. There is also a local management network.

4 Enabling Future Science: ESnet's Evolution over the Next 10 Years

Based both on the projections of the science programs and the changes in observed network traffic and patterns over the past few years, it is clear that the network must evolve substantially in order to meet the needs of DOE's Office of Science mission needs.

The current trend in traffic patterns – the large-scale science projects giving rise to the top 100 data flows that represent about 1/2 of all network traffic – will continue to evolve. As the LHC experiments ramp up in 2006-07, the data to the Tier-1 centers (FNAL and BNL) will increase 200-2000 times. A comparable amount of data will flow out of the Tier-1 centers to the Tier-2 centers (U.S. universities) for data analysis. The DOE National Leadership Class Facility supercomputer at ORNL anticipates a new model of computing in which simulation tasks are distributed between the central facility and a collection of remote "end stations" that will generate substantial network traffic. As climate models achieve the sophistication and accuracy anticipated in the next few years, the amount of climate data that will move into and out of the NERSC center will increase dramatically (they are already in the top 100 flows) Similarly, the experiment facilities at the new Spallation Neutron Source and Magnetic Fusion Energy facilities will start using the network in ways that require fairly high bandwidth with guaranteed quality of service.

This evolution in traffic patterns and volume will result in the top 100 - 1000 flows accounting for a very large fraction of the traffic in the network, even as total ESnet traffic volume grows: The large-scale science data flows will overwhelm everything else on the network.

The current, few gigabits/sec of average traffic on the backbone will increase to 40 Gb/s (LHC traffic) and then increase to probably double that amount as the other science disciplines move into a collaborative production simulation and data analysis mode on a scale similar to the LHC. This will get the backbone traffic to 100 Gb/s as predicted by the science requirements analysis three years ago.

The old hub and spoke architecture (through 2004) would not let ESnet meet these new requirements. The current core ring cannot be scaled to handle the anticipated large science data flows at affordable cost. Point-to-point, commercial telecom tail circuits to sites are neither reliable nor scalable to the required bandwidth.

ESnet's Evolution – The Requirements

In order to accommodate this growth, and the change in the types of traffic, the architecture of the network must change. The general requirements for the new architecture are that it provide:

1) High-speed, scalable, and reliable production IP networking, connectivity for University and international collaboration, highly reliable site connectivity to support Lab operations as well as science, and Global Internet connectivity

2) Support for the high bandwidth data flows of large-scale science including scalable, reliable, and very high-speed network connectivity to DOE Labs

3) Dynamically provisioned, virtual circuits with guaranteed quality of service (e.g. for dedicated bandwidth and for traffic isolation)

In order to meet these requirements, the capacity and connectivity of the network must increase to include fully redundant connectivity for every site, high-speed access to the core for every site (at least 20 Gb/s, generally, and 40-100 Gb/s for some sites) and a 100 Gb/s national core/backbone bandwidth by 2008 in two independent backbones.

ESnet4: A New Architecture to Meet the Science Requirements

The strategy for the next-generation ESnet is based on a set of architectural principles that lead to four major network elements and a new network service for managing large data flows.

The architectural principles are:
A) Use ring topologies for path redundancy in every part of the network – not just in the core.
B) Provide multiple, independent connections everywhere to guard against hardware and fiber failures.
C) Provision one core network – the IP network – specialized for handling the huge number (3×10^9/mo.) of small data flows (hundreds to thousands of bytes each) of the general IP traffic.
D) Provision a second core network – the Science Data Network (SDN) – specialized for the relatively small number (hundreds to thousands) of massive data flows (gigabytes to terabytes each) of large-scale science (which by volume already accounts for 50% of all ESnet traffic and will completely dominate it in the near future).

These architecture principles lead to four major elements for building the new network:
1) A high-reliability IP core network based on high-speed, highly capable IP routers to support:
   - Internet access for both science and lab operational traffic, and some backup for the science data carried by SDN
   - science collaboration services
   - peering with all of the networks needed for reliable access to the global Internet.

2) A Science Data Network core network based on layer 2[a] (Ethernet) and/or layer 1 (optical) switches for:
   - o multiple 10 Gb/s circuits with a rich topology for very high total bandwidth to support large-scale science traffic and for the redundancy needed to high reliability
   - o dynamically provisioned, guaranteed bandwidth circuits to manage large, high-speed science data flows
   - o dynamic sharing of some optical paths with the R&E community for managing peak traffic situations and for providing specialized services such as all-optical, end-to-end paths for uses that do not yet have encapsulation interfaces (e.g. Infiniband)
   - o an alternate path for production IP traffic.
3) Metropolitan Area Network (MAN) rings connecting labs to the core(s) to provide:
   - o more reliable (ring) and higher bandwidth (multiple 10 Gb/s circuits) site-to-core connectivity
   - o support for both production IP and large-scale science traffic
   - o multiple connections between the Science Data Network core, the IP core, and the sites.
4) Loops off the core rings to provide for dual connections to remote sites where MANs are not practical

These elements are structured to provide a network with fully redundant paths for all of the SC Labs. The IP and SDN cores are independent of each other and both are ring-structured for resiliency. These two national cores are interconnected at several locations with ring-structured metropolitan area networks that also incorporate the DOE Labs into the ring. This will eliminate all single points of failure except where multiple fibers may be in the same conduit (as is frequently the case between metropolitan area points of presence and the physical sites). In the places where metropolitan rings are not practical (e.g. the geographically isolated Labs) resiliency is obtained with dual connections to one of the core rings. (See Figure 10.)

The theoretical advantages of this architecture are clear but it must also be practical to realize in an implementation. That is, how does ESnet get to the 100 Gb/s multiple backbones and the 20-40 Gb/s redundant site connectivity that is needed by the SC community in the 3-5 yr time frame?

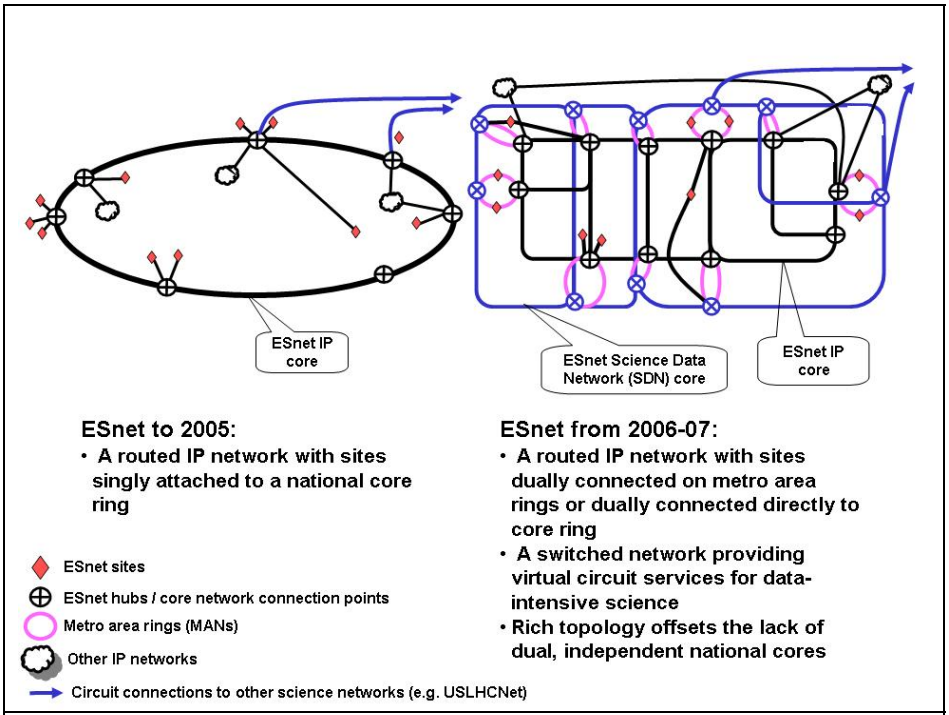

Figure 10. The evolution of the ESnet architecture.

Building ESnet4

Internet2 – the network that serves the US R&E community – has partnered with Level 3 Communications Co. and Infinera Corp. for a

[a] The "layer" term refers to the Open Systems Interconnect (OSI) standard model. Very briefly, layer 1 refers to the sending and receiving bits at the optical or electrical interface. Layer 2 refers to how a computer gets access to a network – e.g. via an Ethernet interface. Layer 3 refers to routing and switching (e.g. IP routers) and layer 4 refers to data transport (e.g. TCP). The OSI model does not map perfectly onto the IP model, but the terms are used anyway. Likewise referring to an Ethernet switch as a "layer 2" device and an IP router as a "layer 3" is not strictly accurate since almost all modern Ethernet switches can do some IP routing and almost all IP routers can so some Ethernet switching. Again, however, the terms are used anyway.

dedicated optical fiber infrastructure with a national footprint and a rich topology - the "Internet2 Network." The fiber will be provisioned with Infinera Dense Wave Division Multiplexing equipment that uses an advanced, integrated optical-electrical design. Level 3 will maintain the fiber and the DWDM equipment as part of its commercial network – a very important consideration for reliability. The DWDM equipment will initially be provisioned to provide10 optical circuits (lambdas or waves) across the entire fiber footprint (40-80 is max number.)

ESnet has partnered with Internet2 to:

- Share the optical infrastructure
- Develop new circuit-oriented network services
- Explore mechanisms that could be used for the ESnet Network Operations Center (NOC) and the Internet2/Indiana University NOC to back each other up for disaster recovery purposes

ESnet will build its next generation IP network and its new circuit-oriented Science Data Network primarily on Internet2 circuits that are dedicated to ESnet, together with a few National Lambda Rail and other circuits. ESnet will provision and operate its own routing and switching hardware that is installed in various commercial telecom hubs around the country, as it has done for the past 20 years. ESnet's peering relationships with the commercial Internet, various US research and education networks, and numerous international networks will continue and evolve as they have for the past 20 years.

ESnet4 will also involve an expansion of the multi-10Gb/s Metropolitan Area Rings in the San Francisco Bay Area, Chicago, Long Island, Newport News (VA/Washington, DC area), and Atlanta to provide multiple, independent connections for ESnet sites to the ESnet core network. (Building the Metropolitan

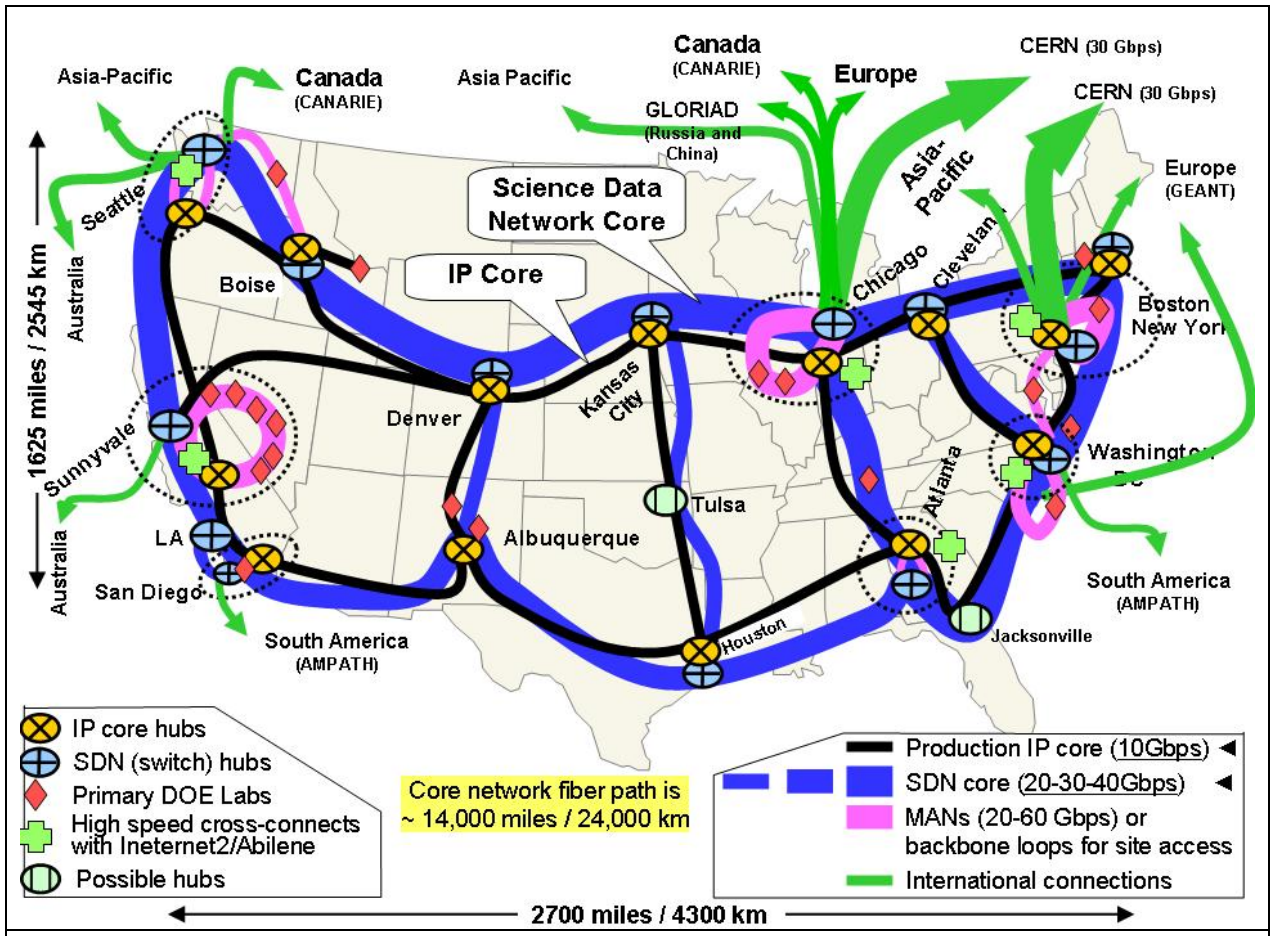

Figure 11. ESnet4 2012 configuration.

The next generation of optical DWDM equipment and network switches and routers is expected to be in place by 2010-2011 to provide 10X over the current per-circuit bandwidth – that is 100 Gb/s per circuit. The core networks will grow to 40-50 Gbps in 2009-2010 and, with new technology, to 400-500 Gbps in 2011-2012.

Area Networks that get the Labs to the ESnet cores is a mixed bag and somewhat opportunistic – a combination of R&E networks, dark fiber networks, and commercial managed lambda circuits are used.)

5 New Network Services

New network services are also critical for ESnet to meet the needs of large-scale science.

One of the most important new network services identified by the Roadmap workshop [5] is dynamically provisioned virtual circuits that provide traffic isolation that will enable the use of non-standard transport mechanisms that cannot co-exist with TCP based transport and provide guaranteed bandwidth.

Guaranteed bandwidth was identified as important in three specific situations.

The first situation is that it is the only way that we currently have to address deadline scheduling – e.g. where fixed amounts of data have to reach sites on a fixed schedule in order that the processing does not fall so far behind that it could never catch up. This is very important for experiment data analysis

The second situation is where remote computing elements are involved in control of real-time experiments. Two examples of this were cited in the applications requirements workshop [2] – one from magnetic fusion experiments and the other from the Spallation Neutron Source. The magnetic fusion situation is that theories are tested with experiments in Tokamak fusion reactors. The experiments involve changing the many parameters by which the reactor can operate and then triggering plasma generation. The "shot" (experiment) lasts a few 10s of milliseconds and generates hundreds of megabytes of data. The device takes about 20 minutes to cycle for the next shot. In that 20 minutes the data must be distributed to the remote collaborators, analyzed, and the results of the analysis fed back to the reactor in order to set up the next experiment (shot). In order to have enough time to analyze the data and use the parameters to set up the next experiment, 200-500 Mb/s of bandwidth must be guaranteed for 2-5 minutes to transmit the data and leave enough time to do that analysis. The situation with the SNS is similar.

The third situation is when Grid-based analysis systems consist of hundreds of clusters at dozens of universities that must operate under the control of a workflow manager that choreographs complex workflow. This requires quality of service to ensure a steady flow of data and intermediate results among the systems. Without this, systems with many dependencies and with others dependent on them would stop and start with the interruptions propagating throughout the whole collection of systems creating unstable and inefficient production of analysis results that would reduce the overall throughput necessary to keep up with the steady generation of data by the experiment. (This is of particular concern with the huge amount of data coming out of the LHC experiments.)

OSCARS: Guaranteed Bandwidth Service

DOE has funded the OSCARS (On-demand Secure Circuits and Advance Reservation System) project to develop and deploy the various technologies that provide dynamically provisioned circuits and various qualities of service (QoS) that can be integrated into a production net environment. Such "circuits" are called "virtual circuits" (VCs) because that are defined in software and thus are mutable (as opposed to hardware established circuits).

The end-to-end provisioning will initially be provided by a combination of Ethernet switch management of λ (optical channel) paths in the MANs and Ethernet VLANs and/or MPLS paths (Multi-Protocol Label Switching and Label Switched Paths - LSPs) in the ESnet IP and SDN cores.

There are two realms in which OSCARS must operate: 1) intra-domain – that is, to establish a schedulable, guaranteed bandwidth circuit service within the boundary of the ESnet network; 2) inter-domain – e.g. to provide end-to-end QoS between DOE labs and US and European universities.

The OSCARS architecture is based on Web/Grid Services and consists of modules for:

- User/application request and claiming;

- Authentication, Authorization, and Auditing (AAAS) to handle access control, enforce policy, and generate usage records;
- A Bandwidth Scheduler Subsystem (BSS) that tracks reservations and maps the state of the network (present and future) for managing the allocation of virtual circuits;
- A Path Setup Subsystem (PSS) that setups and tears down the on-demand virtual circuits.

OSCARS, and the similar systems in the other networks, is well into development and prototype deployment. A number of OSCARS circuits are currently being tested between DOE Labs and European R&E institutions. For further information see [18].

Network Monitoring

The next generation networks that are currently being built by ESnet, Internet2, GÉANT, and others, are much higher speed and much more complex than the current networks. Further, as the science community demands higher and higher performance from the applications that use the network, a user service to perform network monitoring has become an important goal.

perfSONAR is a global collaboration to design, implement and deploy a network measurement framework that is accessible by the user community. It is also being developed to provide the monitoring of the LHC Optical Private Network (OPN - http://lhcopn.web.cern.ch/lhcopn/) that provides the main data feeds from CERN to the National LHC Data Centers. perfSONAR is a Web Services based Framework whose main components are:

- Measurement Archives (MA)
- Measurement Points (MP)
- Lookup Service (LS)
- Topology Service (TS)
- Authentication Service (AS)

Some of the currently deployed services are:

- Utilization MA
- Circuit Status MA & MP
- Latency MA & MP
- Bandwidth MA & MP
- Looking Glass MP
- Topology MA

perfSONAR is an active collaboration. The basic framework is complete, the protocols are being documented, and new services are being developed and deployed. The collaboration currently involves about 25 organizations in the US and Europe. For more information see [20] and [21].

ESnet Grid, Middleware, and Collaboration Services Supporting Science

The key requirements studies whose results are guiding the evolution of ESnet ([2], [4], and [5]) have identified various middleware services that need to be in place, in addition to the network and its services, in order to provide an effective distributed science environment.

These services are called “science services” – services that support the practice of science. Examples of these services include:

- Trust management for collaborative science
- Cross site trust policies negotiation
- Long-term PKI key and proxy credential management
- Human collaboration communication
- End-to-end monitoring for Grid / distributed application debugging and tuning
- Persistent hierarchy roots for metadata and knowledge management systems

There are a number of such services for which an organization like ESnet has characteristics that make it the natural provider. For example, ESnet is trusted, persistent, and has a large (almost comprehensive

within DOE) user base. ESnet also has the facilities to provide reliable access and high availability of services through assured network access to replicated services at geographically diverse locations.

However, given the small staff of an organization like ESnet, a constraint on the scope of such services is that they must be scalable in the sense that as the service user base grows, ESnet interaction with the users does not grow.

There are three such services that ESnet provides to the DOE and/or its collaborators.

- Federated trust
  - policy is established by the international science collaboration community to meet its needs
- Public Key Infrastructure certificates for remote, multi-institutional, identity authentication
- Human collaboration services
  - video, audio, and data conferencing

Authentication and Trust Federation Services

Cross-site identity authentication and identity federation is critical for distributed, collaborative science in order to enable routine sharing computing and data resources, and other Grid services. ESnet provides a comprehensive service to support secure authentication.

Managing cross-site trust agreements among many organizations is crucial for authorization in collaborative environments. ESnet assists in negotiating and managing the cross-site, cross-organization, and international trust relationships to provide policies that are tailored to collaborative science.

ESnet Public Key Infrastructure

Grid computing and data analysis systems rely on Public Key Infrastructure (PKI) for their security. ESnet provides PKI and X.509 identity certificates that are the basis of secure, cross-site authentication of people and Grid systems. The ESnet root Certification Authority (CA) service supports several CAs with different uses and policies that issue X.509 identity certificates after validating the user request against the policy of the CA. For example, the DOEGrids CA has a policy tailored to accommodate international science collaboration, the NERSC (DOE Office of Science supercomputer center) CA policy integrates CA and certificate issuance with NERSC user accounts management services, and the FusionGrid CA supports the policy of the DOE magnetic fusion program's FusionGrid roaming authentication and authorization services, providing complete key lifecycle management.

The ESnet PKI is focused entirely on enabling science community resource sharing and its policies are driven entirely by the science communities that it serves. That is, the trust requirements of the science communities are formally negotiated and encoded in the Certification Policy and Certification Practice Statement of the CAs.

The DOEGrids CA (www.doegrids.org) was the basis of the first routine sharing of HEP Grid computing resources between United States and Europe.

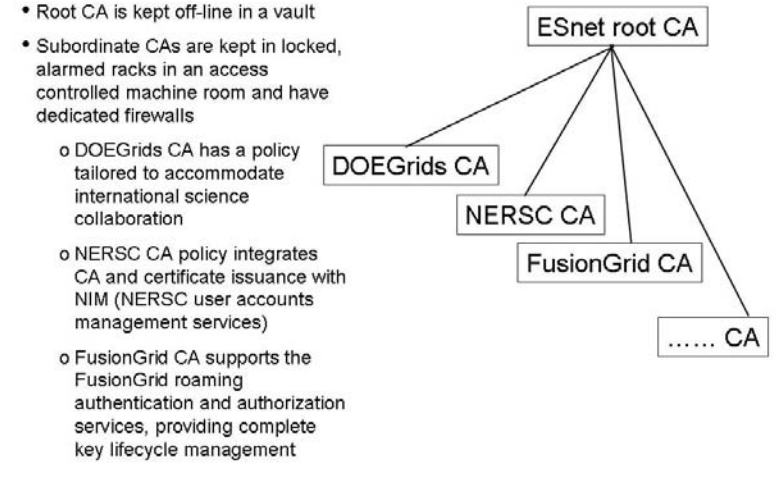

Figure 12. ESnet Public Key Infrastructure

Federation and Trust management

ESnet has been working with the international Grid community develop policies and processes that facilitate the establishment of multi-institutional and cross-site trust relationships. This effort led to the development of two key documents used by the community and published by the Global Grid Forum (GGF): CA Policy Management Authority guidelines, and a reference Certificate Policy and Certification Practices Statement (CP/CPS). Policy Management Authorities (PMAs) encode, manage, and enforce the policy representing the trust agreements that are worked out by negotiation. The PMA guidelines outline how to establish a PMA. The CP/CPS guidelines were written to outline issues of trust that must be addressed when setting up a CA.

These documents are used by the regional CA providers to organize their management and to specify their policies. The European, EU Grid PMA, and the Asia Pacific, AP PMA, both use these documents for their communities.

ESnet represents the DOE and NSF Grid user community by participation as a full member on the EUGrid PMA. This is a requirement because of the need to collaborate between the two user communities. To better serve the Americas Grid community, and to help off load certification by the EU Grid PMA, ESnet has helped establish The Americas Grid PMA (TAGPMA).

The formation of the EU, AP PMA, and Americas Grid PMAs has created a need to coordinate these regional efforts to insure a common, global trust based federation. The International Grid Trust Federation (IGTF) was fostered by ESnet to help coordinate the global efforts of trust management. In March 2003, ESnet met in Tokyo with a number of international PMAs, this led to the establishment of the IGTF (www.GridPMA.org). The IGTF has grown to include the three major regional PMAs: www.EUGridPMA.org, www.APGridPMA and the new www.TAGPMA.org (Americas). It will be the publishing point for various policies and official points of contact.

Voice, Video, and Data Tele-Collaboration Service

The human communication aspect of collaboration, especially in geographically dispersed scientific collaborations, represents an important and highly successful ESnet Science Service. This service provides audio, video, and data teleconferencing with the central scheduling essential for global collaborations.

The ESnet collaboration service supports more than a thousand DOE researchers and collaborators worldwide with H.323 (IP) videoconferences (4000 port hours per month and rising), audio conferencing (2500 port hours per month) (constant), and data conferencing (150 port hours per month). Web-based, automated registration and scheduling is provided for all of these services.

6 Conclusions

ESnet is an infrastructure that is critical to DOE's science mission, both directly and in supporting collaborators. It is focused on the Office of Science Labs, but serves many other parts of DOE.

ESnet is implementing a new network architecture in order to meet the science networking requirements of DOE's Office of Science. This architecture is intended to provide high reliability and very high bandwidth.

Grid middleware services for large numbers of users are hard – but they can be provided if careful attention is paid to scaling. ESnet provides PKI authentication services and world-wide video and audio conferencing to DOE scientists and their collaborators.

Acknowledgements

The ESnet senior network engineering staff that are responsible for the evolution of ESnet consists of Joseph H. Burrescia, Michael S. Collins, Eli Dart, James V. Gagliardi, Chin P. Guok, Yvonne Y. Hines, Joe Metzger, Kevin Oberman and Michael P. O'Connor. The staff responsible for Federated Trust includes Michael W. Helm and Dhivakaran Muruganantham (Dhiva). The staff responsible for the Tele-Collaboration services includes Stan M. Kluz and Clint Wadsworth. This group of people contributed to this paper.

ESnet is funded by the US Dept. of Energy, Office of Science, Advanced Scientific Computing Research (ASCR) program, Mathematical, Information, and Computational Sciences (MICS) program. Dan Hitchcock is the ESnet Program Manager and Thomas Ndousse-Fetter is the Program Manager for the network research program that funds the OSCARS project.

ESnet is operated by Lawrence Berkeley National Laboratory, which is operated by the University of California for the US Dept. of Energy under contract DE-AC03-76SF00098.

Notes and References

[1] http://www.energy.gov/, Science and Technology tab.

[2] High Performance Network Planning Workshop, August 2002 http://www.doecollaboratory.org/meetings/hpnpw

[3] DOE Workshop on Ultra High-Speed Transport Protocols and Network Provisioning for Large-Scale Science Applications, April 2003 http://www.csm.ornl.gov/ghpn/wk2003

[4] DOE Science Networking Roadmap Meeting, June 2003 http://www.es.net/hypertext/welcome/pr/Roadmap/index.html

[5] Science Requirements for ESnet Networking, http://www.es.net/hypertext/requirements.html

[6] LHC Computing Grid Project http://lcg.web.cern.ch/LCG/

[7] http://www.sc.doe.gov/ascr/20040510_hecrtf.pdf (public report)

[8] ASCR Strategic Planning Workshop, July 2003 http://www.fp-mcs.anl.gov/ascr-july03spw

[9] Planning Workshops-Office of Science Data-Management Strategy, March & May 2004 http://www-user.slac.stanford.edu/rmount/dm-workshop-04/Final-report.pdf

[10] ESG - Earth System Grid. http://www.earthsystemgrid.org/ ESG - Earth System Grid. http://www.earthsystemgrid.org/

[11] CMS - The Compact Muon Solenoid Technical Proposal. http://cmsdoc.cern.ch/

[12] The ATLAS Technical Proposal. http://atlasinfo.cern.ch/ATLAS/TP/NEW/HTML/tp9new/tp9.html

[13] LHC - The Large Hadron Collider Project. http://lhc.web.cern.ch/lhc/general/gen_info.htm

[14] The BaBar Experiment at SLAC. http://www-public.slac.stanford.edu/babar/

[15] The D0 Experiment at Fermilab. http://www-d0.fnal.gov/

[16] The CDF Experiment at Fermilab. http://www-cdf.fnal.gov/

[17] The Relativistic Heavy Ion Collider at BNL. http://www.bnl.gov/RHIC/
[18] http://www.es.net/oscars
[19] http://csrc.nist.gov/piv-project/
[20] "Measuring Circuit Based Networks," Joe Metzger, ESnet. http://www.es.net/pub/esnet-doc/index.html#JM021307
[21] "ESnet Network Measurements," Joe Metzger, ESnet. http://www.es.net/pub/esnet-doc/index.html#JM021307