ESnet
Energy Sciences Network

April-June 2006 Volume 1, Issue 2

Berkeley Lab

www.es.net

news

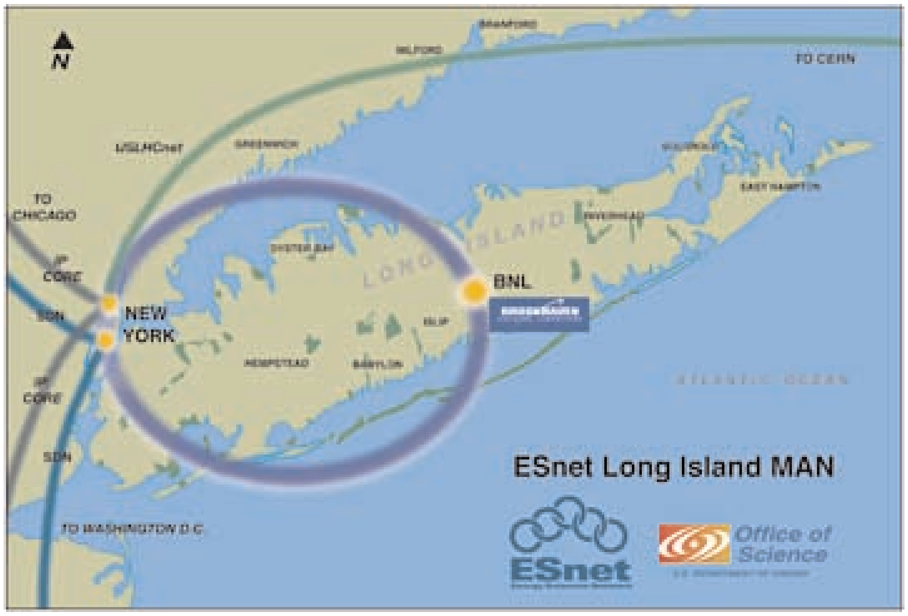
LONG ISLAND METROPOLITAN AREA NETWORK BOOSTS BNL CONNECTIVITY

By Mike O´Connor

ESnet continues to roll out its next-generation architecture on schedule with the March 14 completion of the Long Island Metropolitan Area Network, connecting Brookhaven National Laboratory to the ESnet point of presence (POP) 60 miles away in New York City. The new design is built on a diversely routed dual 10 gigabits per second (Gb/s) dense wave division multiplexed (DWDM) ring that connects the DOE Office of Science laboratory to both the ESnet IP core and the new Science Data Network (SDN) for high-throughput science data and collaboration.

This service upgrade provides BNL with a fully usable aggregate bandwidth of 20 Gb/s, an eightfold increase over the lab´s previous 2.4 Gb/s link to Manhattan. This architecture is designed to be easily expanded to eight waves supplying 80 Gb/s to the lab. BNL is the largest Tier 1

continued on page 4



ESnet NEWS

This is the second issue of *ESnet News*, highlighting accomplishments and services of the U.S. Department of Energy´s ESnet (Energy Sciences Network), managed by Lawrence Berkeley National Laboratory. ESnet News will be distributed four times a year via email and may be freely distributed. ESnet News is edited by Jon Bashor, JBashor@lbl.gov or 510-486-5849.

Lehman: ESnet Well Managed, Responsive

On February 21, 2006, 25 world-class experts representing the DOE Office of Science (SC) and SC programs, along with advanced networking specialists and industry luminaries began an intense three-day Integrated Baseline Review of ESnet. The committee examined various aspects of ESnet including requirements gathering, technical approach, cost, schedule and funding. LBNL and ESnet staff spent one day giving presentations to the full committee, followed by a day of interactive participation with the various subcommittees.

The review, known as a "Lehman Review," was led by Daniel R. Lehman,

continued on page 2

ESnet, Internet2 Take Key Step toward End-to-End On-Demand Bandwidth System

Two of the nation´s leading networks, ESnet and Internet2, have demonstrated an automated system for providing on-demand end-to-end bandwidth service to support large-scale research.

ESnet developed the system in response to scientists who need to move large amounts of data within a specific time interval. The system was initially developed to work within ESnet´s internal backbone which connects national laboratories. "But for on-demand bandwidth service to truly work, you need to go end-to-end and not just serve one network," said Chin Guok, an ESnet network engineer and principal investigator for the project. "In order to do this, you need to dynamically coordinate with the other networks to allocate the same resources at the same time."

In April, ESnet and Internet2, a network consortium led by more than 200 U.S. universities working with industry and government, conducted a test of the system by configuring a path between an Internet2 test site in Indianapolis and an ESnet hub in Sunnyvale, Calif. A guaranteed bandwidth of 25 Mb/s was requested, and the April 6 test achieved the requested throughput of 25 Mb/s.

ESnet developed the base code which allows users to request dedicated

continued on page 2

ESnet-Internet2 Test

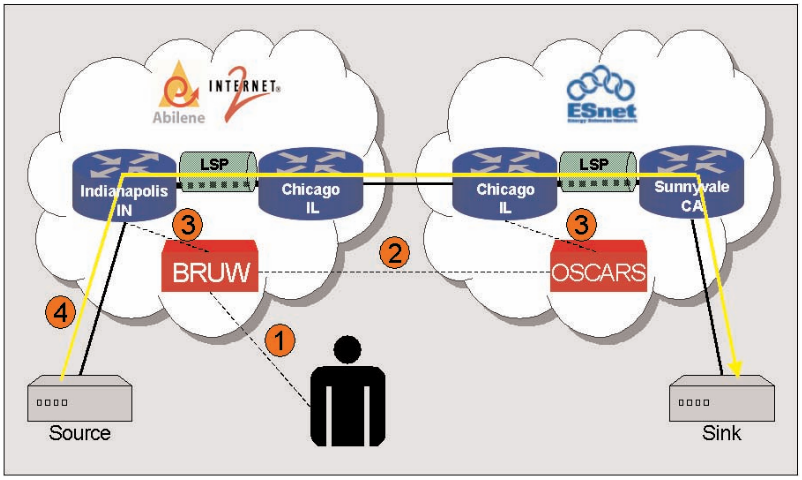
continued from page 1

bandwidth at a specified time between specified locations. The code, which is being further developed by both ESnet and Internet2 staff, also authenticates the requestor and ensures that the person is authorized to use the network service at the requested level.

The system is an extension of ESnet's On-Demand Secure Circuits and Advance Reservation System (OSCARS) project to develop and deploy a prototype service that enables on-demand provisioning of guaranteed bandwidth secure circuits within ESnet.

While other organizations have demonstrated interdomain on-demand bandwidth services, it's typically done at the lower network layers such as the optical layer, by provisioning lambdas, Guok said. The ESnet-Internet2 collaboration is the first step toward implementing a production service providing on-demand bandwidth on a shared IP network.

"The test also helped us identify other issues, such as vendor implementations, which we need to address," Guok said. "This is just a first step in the journey to production."



Path setup for end-to-end bandwidth reservation.

In the past, it was possible to reserve bandwidth, according to Guok, but this typically involved lots of phone calls and emails involving requests and specifications. Such a process could take days or weeks to complete. With the new system, the request can be completed in a matter of minutes. "It's a huge savings in both time and cost," Guok said.

The challenges to providing dedicated, reserved bandwidth occur at multiple levels. First there is the issue of how to reserve network resources. This requires that traffic be isolated and the bandwidth guaranteed. Then, there is a need for authentication and authorization when the traffic crosses domains. Finally, "virtual circuits" need to be set up through each network and then stitched together between the various networks. This requires that the peering points be identified in advance.

The new system allows ESnet to engineer the network traffic, avoiding congested points. It also provides flexibility to meet user requirements. For example, videoconferencing may not require as much bandwidth as data transfer, but it does need lower latency and the system can build a path that is shorter.

Lehman Review

continued from page 1

director of the Office of Science's Office of Project Assessment. Michael Strayer, associate director of the Office of Advanced Scientific Computing Research, requested that Lehman lead an SC review to evaluate ESnet's plan for the next five years for providing high performance network support to SC researchers at the national laboratories and universities, and its plan for interacting with international networks that are critical to SC missions.

The final report of the committee sent a strong positive message in regard to the current and future directions of ESnet. An excerpt from the Executive Summary states: "The Committee found that ESnet is effectively managed and responsive to its customers and DOE Program Management. However, an increase in network capability would be needed to meet the future demands of scientific collaborators (existing and new customers). ESnet presented plans to upgrade its current basic Internet Protocol (IP) production services and create the Science Data Network (SDN) to address the future needs (with an emphasis on the next five years). Overall, the gathering of requirements from various scientific domains was well accomplished; the technical approach for upgrading the network was found to be sensible and consistent with other network approaches; … successful transition of the 'old' ESnet to the 'new' ESnet can be a key element that demonstrates that DOE is fully engaged in leading the advancement of science for the country."

According to Bill Johnston, head of ESnet at LBNL, the review resulted from ESnet proposing to build the next-generation SC network in response to a series of workshops that gathered requirements from the science community.

"All of the input from the science community indicated that the future of large-scale science that is at the core of the Office of Science mission will depend on very high-speed networking that is integrated into a modern, service-oriented, computing environment," said Johnston. "The networks of the not-so-distant future will not only have to be an order of magnitude faster than today's networks, but will have to offer service guarantees that allow widely distributed data processing and computing systems to operate reliability and smoothly.

"ESnet developed a plan that will provide these characteristics over the next five years, and the successful Lehman Review is an important step toward making the plan reality," Johnston concluded.

ESCC's Role Becoming More Formalized

By Scott Bradley, Chairman, ESCC

It has been a very busy and productive few months for the ESnet Site Coordinators Committee (ESCC). Since our last meeting in Albuquerque, Kevin Oberman has been leading the effort to create an ESCC position paper on Secure DNS (DNSSEC) implementation, in the hopes that DOE will be able to influence OMB guidance on the technology prior to it becoming final. In the next couple of months, we will be similarly examining OMB direction on IPv6.

At the May meeting of the Systems of Labs' Computing Coordinating Committee (SLCCC), the committee moved to formally sponsor the ESCC as a formal subcommittee, with the following provisions:

- The ESCC will continue forward in the role that it has held in the past, that is, as a technical coordinating committee for the primary stakeholders in ESnet, including DOE laboratories and other ESnet sites.
- The ESCC will revise and update its charter to reflect this changed relationship with SLCCC, will follow SLCCC subcommittee requirements and will adopt other changes as required by SLCCC in the future.

This is a logical and welcome evolution of our role within DOE, given the natural relationship between the strategic IT planning role of the SLCCC and our operational role within the ESCC.

Our next meeting will be held July 19-20 in Madison, Wis., in conjunction with Joint Techs. At that meeting we will be kicking off our planning process for the deliverables described above, and your participation is most welcome!

ESnet Provides Tech Support, Expertise for Moving Bulk Data

The movement of bulk data over long-haul networks has become integral to modern computational science. Raw instrument data, computational data sets, and visualization data must be moved between institutions to take advantage of powerful computational and other resources.

Moving large (tens of gigabytes or more) data sets across the network is difficult using the default TCP settings on most hosts. Hosts are configured by default to allow them to function well for most common uses, e.g., reasonable performance on an Ethernet LAN, and reasonable performance when connecting to Web servers on the commodity Internet. In most cases, successful high-speed bulk data transfers rely on users reconfiguring their hosts with non-default TCP parameters. There are several resources describing TCP performance tuning – two examples are http://dsd.lbl.gov/TCP-tuning/ and http://www.psc.edu/networking/projects/tcptune/.

The following table illustrates the data rate required to move a data set of a particular size in a given amount of time. The red shaded areas indicate single-stream throughput rates greater than 10 Gb/s, while the green shaded areas indicate data rates that should be straightforward to achieve. These speeds are easily achievable using Fast Ethernet interfaces.

This table shows that while the default TCP settings of most hosts may not be appropriate, significant throughput can be achieved in a few hours with modest resources. Two good approximations are that one can move about 1 terabyte per 100 Mb/s of throughput per day, and one can move 10 gigabytes per 1 Mb/s of throughput per day. The first requires slightly greater than Fast Ethernet speeds due to protocol overhead, Fast Ethernet is capable of maximum throughput rates in the 85 to 90 Mb/s range. The second is easily achievable with a laptop.

If ESnet users are unable to move their data over the network, there are several resources available. The ESnet performance testers (https://performance.es.net/) provide a means for testing pieces of the ESnet network to determine the location of the problem. There is an ESnet white paper on bulk data transfer available here (http://www.es.net/pub/esnet-doc/TCP-Bulk-Data-Transfer.pdf). Users are also encouraged to contact the ESnet network operations center at trouble@es.net to address performance issues.

10 PB	300,240 Gb/s	25,020 Gb/s	3,128 Gb/s	1,043 Gb/s	149 Gb/s	35 Gb/s
1 PB	30,024 Gb/s	2,502 Gb/s	313 Gb/s	104 Gb/s	15 Gb/s	4 Gb/s
100 TB	2,932 Gb/s	244 Gb/s	31 Gb/s	10 Gb/s	2 Gb/s	340 Mb/s
10 TB	293 Gb/s	24 Gb/s	3 Gb/s	1 Gb/s	145 Mb/s	34 Mb/s
1 TB	29 Gb/s	2 Gb/s	305 Mb/s	102 Mb/s	15 Mb/s	3 Mb/s
100 GB	3 Gb/s	239 Mb/s	30 Mb/s	10 Mb/s	1 Mb/s	331 Kb/s
10 GB	286 Mb/s	24 Mb/s	3 Mb/s	994 Kb/s	142 Kb/s	33 Kb/s
1 GB	29 Mb/s	2 Mb/s	298 Kb/s	99 Kb/s	14 Kb/s	3 Kb/s
100 MB	3 Mb/s	233 Kb/s	29 Kb/s	10 Kb/s	1 Kb/s	<1 Kb/s
	5 Minutes	1 Hour	8 Hours	24 Hours	7 Days	30 Days

LI MAN Boosts BNL Bandwidth

continued from page 1

data distribution site for the ATLAS experiment at CERN and requires this next-generation network to support its ATLAS mission.

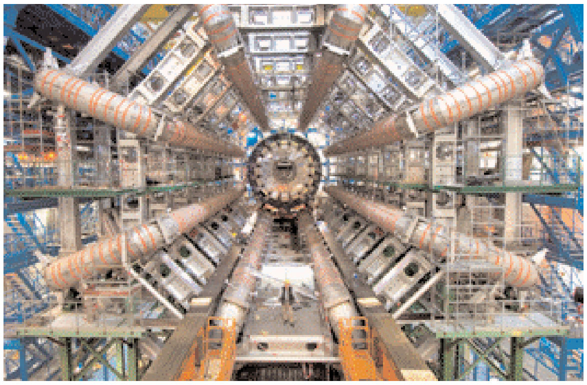
One of these two 10 Gb/s lambdas in place at BNL will access the ESnet SDN and be provisioned primarily for dedicated circuits scaling to support the demands of petabyte data transfers, such as those required by ATLAS. The data will be generated by the ATLAS experiment to be conducted on the Large Hadron Collider (LHC) currently under construction at CERN. This data will be carried from CERN to New York City by USLHCnet (the American LHC Network).

ATLAS (A Toroidal LHC ApparatuS), is one of four detectors located at the Large Hadron Collider (LHC), now under construction near Geneva, Switzerland.

In preparation for the experiment going on line and to ensure the network and data centers can accommodate the flood of data, a series of data challenges with simulation data have been carried out. Following the completion of the recent fourth and final LHC Data Challenge prior to production, ESnet contacted BNL to find out how this new network is helping them support the mission of BNL within the ATLAS collaboration. ESnet News sat down with BNL's CIO Thomas Schlagel; Bruce Gibbard, the head of ATLAS computing at BNL; and Scott Bradley, manager of data networks and voice services.

Question: Tom, how important is the success of the ATLAS program at BNL to the mission of the laboratory and how will BNL benefit from its long-term success?

Tom Schlagel: The success of the ATLAS experiment and the Tier-1 US-ATLAS center is extremely important to the mission of the laboratory in conducting world-class research. In addition to the science aspects, the ATLAS and CMS experiments at the LHC are at the forefront of global collaborative computing using Grid technologies. As has happened frequently in the past, the particle physics community is driving computing innovation which will have long-term consequences outside of the scientific research arena. BNL will benefit directly by participating in these activities. The hope is that the computing and networking capabilities put in place to support ATLAS, along with the experience gained by putting them in place, will be beneficial to other research programs in the future.



The ATLAS experiment at CERN, which will result in 5 PB of data being processed at BNL in 2008.

Q: Bruce, as the head the ATLAS Tier 1 Computing Facility at BNL, how would you describe the role BNL plays in the collaboration?

Bruce Gibbard: The ATLAS collaboration will process and distribute its physics data using a hierarchy of computing centers. The sole Tier 0 center is located at CERN in Switzerland, while the next level of 10 Tier 1 centers is globally distributed in Europe, Taiwan, Canada, and the U.S. The Tier 1 sites are each responsible for archiving an allotted portion of the "raw" detector data, as well as providing the compute cycles required to process, and reprocess as necessary, that portion to produced the reduced data sets required for detailed analysis.

BNL is the largest of the ATLAS Tier 1 centers and the only one in the U.S, and so is responsible for archiving and processing approximately 20 percent of the ATLAS raw data as well as serving the processed data out to individual ATLAS users and sites in the U.S. Once a new processing pass on the data is completed, based on the latest version of the reconstruction software and calibration data, the Tier 1 sites exchange copies of the portions they processed. These consist of ESD (Event Summary Data) and AOD (Analysis Object Data). This results in each Tier 1 receiving, archiving and then serving as a distribution point for the ATLAS ESD and AOD data sets.

The next level of the hierarchy, the Tier 2 centers located at universities and research institutions, is where the majority of the researchers are expected to work. For ATLAS in the U.S., there are currently three Tier 2 centers, and two more will be selected for funding this year. The Tier 2 sites draw on the data sets made available at the Tier 1 sites.

Q: With the tremendous amount of raw LHC data acquired, then processed into additional data sets, exchanged between Tier 1 sites and distributed to Tier 2 sites, resulting in a total data volume of 5 PB at BNL in 2008, is the new LI MAN up to the task?

Bruce Gibbard: Our wide area bandwidth projections indicate that ATLAS should be adequately provisioned for bandwidth through 2008 and we have gained substantial additional confidence from participation in the LHC Data Challenge 4 exercise. During this multi-week exercise, we were able to sustain an average transfer rate from CERN to our disk arrays of 191 Mb/s (~1.5 Gb/s) compared to a target rate of 200 Mb/s. This was actually a higher average rate than that achieved by any other Tier 1 site. However, our ATLAS bandwidth projections do not account for bandwidth contention between ATLAS and other BNL projects, such as RHIC. I was pleased that during the data challenge exercise we did not have to take any special action to restrict ATLAS use of network resources. We did have some concerns going into it about the RHIC project moving very large data sets from BNL to the Riken Center in Japan during the data challenge, but thanks to the recent upgrade of our network to a fully duplex

continued on page 5

ESnet Plays Strong Role in Extending Grid Trust to Central and South America

ESnet staff, who have helped lead international efforts to develop and deploy common tools for authenticating Grid users among sites in North America, Europe and Asia, have now helped bring Latin American nations on board.

ESnet's DOEGrids Certificate Authority (CA) played a significant role in helping organize the first face-to-face meeting of the Americas Grid Policy Management Authority (TAGPMA), held March 27-29 in Rio de Janeiro. This meeting was hosted by RNP (the Brazilian national research and education network) and CBPF (the Brazilian Physics Research Institute). ESnet managed the CA evaluation team and process, and invited participation from EUGridPMA members.

And to enable the participation of Grid CA operators and other stakeholders who weren't able to attend in person, Alan Sill of the Texas High Energy Grid used his expertise in ESnet Collaboration Service (ECS) to provide video and audio links.

TAGPMA, the newest member of the International Grid Trust Federation (www.GridPMA.org), is the means by which U.S., Canadian, and Latin American Grid CAs can be evaluated against Grid standards for CA operations, and then accepted into the master list of trusted Grid CAs.

The meeting coincided with the appearance of EELA (E-Infrastructure shared between Europe and Latin America) in the Grid community. EELA is a joint EU and Latin American initiative to enable e-infrastructure in Latin America, in particular developing Grid and network-related projects. EELA joined TAGPMA as a relying party member of the PMA.

Particularly important at this moment is enabling Brazilian and other Latin American participation in projects related to the Large Hadron Collider at CERN. The first TAGPMA meeting began the compliance review process for five CAs: Brazil, Latin American Catch-all (hosted by Brazil), Chile, Argentina and Mexico. Other Latin American countries such as Venezuela sent observers, and organizers expect that several other countries and organizations will participate in EELA and TAGPMA in the near future.

The meeting featured strong participation from the U.S. academic and national lab communities. Dartmouth College, the lead for the Higher Education Bridge Certificate Authority, continued to work on building trust relationships between the Grid and academic PKI efforts. This effort will help parallel efforts in Latin American academic and research communities.

Additionally, the chairman of the EUGridPMA and several technical experts from that community made the long trip to Rio to help TAGPMA and EELA understand global trust efforts and technical specifications. "We identified several policy and technical issues that need further refinement, particularly how we express and certify 'levels of assurance' and other certificate quality attributes," said ESnet's Tony Genovese. "These issues will be developed in upcoming meetings in GGF and the regional IGTF PMAs."

In addition to DOEGrids (ESnet), other founding members of the TAGPMA are CANARIE, Open Science Grid, TeraGrid, Texas High Energy Grid, SDSC, FNAL and HEBCA/USHER (Dartmouth).

LI MAN Boosts BNL Bandwidth

continued from page 4

10 Gb/s path, it all just worked.

Q: Scott, as the manger of BNL's data networking and voice services, you find out pretty quickly when the network is not performing. Would you please describe how this new network is performing and a little bit about the process that put it in place.

Scott Bradley: The performance of the new Long Island Metropolitan Area Network has been outstanding and we're very pleased with it. It's a high performance network that was the product of an extremely successful coordinated effort between BNL and ESnet. BNL took advantage of the technical expertise, project management and financial backing of ESnet that enabled us to exploit the detailed knowledge we've acquired through our many years of dealing with the local carriers in the Long Island and New York metropolitan area. Working with ESnet in this way has saved countless man hours and budget dollars otherwise spent on WAN circuit and dark fiber leases we would have needed to reach New York and Chicago. BNL was in effect freed to concentrate effort on our core competencies. Did I mention how well we scored in (LHC) Service Challenge 4 (smile)? I believe that this partnership has become much stronger as a result of this process and as we move forward we will continue to rely on ESnet for connecting to LHC ATLAS and its global collaboration.

ESnet thanks Tom Schlagel, Bruce Gibbard, Scott Bradley, Dantong Yu and Frank Burstein for their contributions to this article.

SERVICES

ESnet Provides Free DNS Backup Services

For most of its existence, ESnet has provided domain name system (DNS) services for its own domains as well as those of its customers. Due to the great importance of functional DNS service, even when a site is not connected to the network, having your DNS data available from geographically diverse locations with excellent connectivity to the rest of the Internet cannot be overstated. Internet best correct practices have long stated that at least two geographically diverse servers should be required for any domain.

ESnet provides backup service with servers across the country. If your site does not have one or more geographically diverse backups, we can solve the problem. If your site could use this service, send a request to hostmaster@es.net.

Many people assume that there is no real problem with not having DNS running if the site is unreachable, but this is not true. Services such as email treat systems that do not respond very differently from systems that are not found in DNS. Lack of connectivity is treated as a temporary failure. Lack of DNS is usually seen as evidence that the destination does not exist, and no continued efforts to connect are made. Email is bounced and users annoyed.

As to the future, there are changes coming to greatly enhance the robustness of DNS.

First, ESnet is working toward implementing "anycast" DNS service for ESnet and any site for which ESnet provides backup service. In a nutshell, anycast DNS service means that the actual service is provided by any of the available servers based on the "closest" route. If a server fails, DNS queries are simply routed to the next closest server in the anycast group.

For this to work, all ESnet servers must have identical authoritative data, so all sites for which ESnet provides backup service must allow all ESnet DNS servers to do zone transfers. This means that firewalls must allow TCP connections to server port 53, and the servers must allow the transfers.

Second, OMB is currently working on a requirement for all government-owned DNS servers to implement Secure DNS (DNSSEC). This will results in several changes in how we all operate. Zones must be signed and all servers for a domain will have to set up shared secrets to control all communications between them.

All DNS active personnel should read NIST 800-81. It provides the relevant procedures for implementing both signed responses and transaction signing. This may also require more processing power for servers.

If you are interested in discussions of these issues, please drop a note to hostmaster@es.net and ask to be added to the DNSSEC mailing list.

DISCLAIMER This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California. Ernest Orlando Lawrence Berkeley National Laboratory is an equal opportunity employer.

ESnet Provides Network Time Synchronization Services

ESnet is now supporting Network Time Protocol (NTP) servers to provide very accurate time synchronization sources for ESnet sites. The intent of this service is to support two Stratum 1 time sources that ESnet sites can use to synchronize their time servers and redistribute within their own sites.

The ESnet IPv4 NTP servers are chronos.es.net and saturn.es.net. The IPv6 service names are chronos6.es.net and saturn6.es.net. The East Coast server is named after Chronos, the ancient Greek god of time, while the West Coast server is named after Saturn, the Roman god of time.

"Although ESnet has provided time services informally to the ESnet site community for many years, during a recent review of NTP on our backbone systems, we discovered several cases where sites were using less-than-ideal ESnet time services," said Joe Metzger of the ESnet staff. "We decided to leverage a couple of the clocks we are using in our network measurement infrastructure to provide significantly more accurate time sources to our community. This also allows us to eliminate an unnecessary security exposure on other systems."

The ESnet servers are using Praecis Ct CDMA (code division multiple access) clocks from Endrun Technologies. The vendor claims that the CDMA signal should be within 10 microseconds of GPS time. ESnet has found the CDMA time sources to be very accurate in the network measurement systems.

CDMA clocks rely on a time signal embedded in certain cellular telephone signals. CDMA cell towers use GPS receivers to get the time and then relay it into the CDMA signal.