National Science Foundation, Directorate for Computer Information Science and Engineering
Division of Advanced Networking Infrastructure & Research (ANIR)

NSF Cooperative Agreement No. ANI-9730202
September 2002 Quarterly Status Report

December 3, 2002

Tom DeFanti, Maxine Brown, Andy Johnson, Dan Sandin, Jason Leigh, Alan Verlo, Laura Wolf
Electronic Visualization Laboratory
University of Illinois at Chicago

Linda Winkler, Bill Nickless, Caren Litvanyi
Argonne National Laboratory

Jim Williams, Stephen Peck
Indiana University

Table of Contents

A. Summary of Technical Activities

A.1. Euro-Link Network Status and Institutions

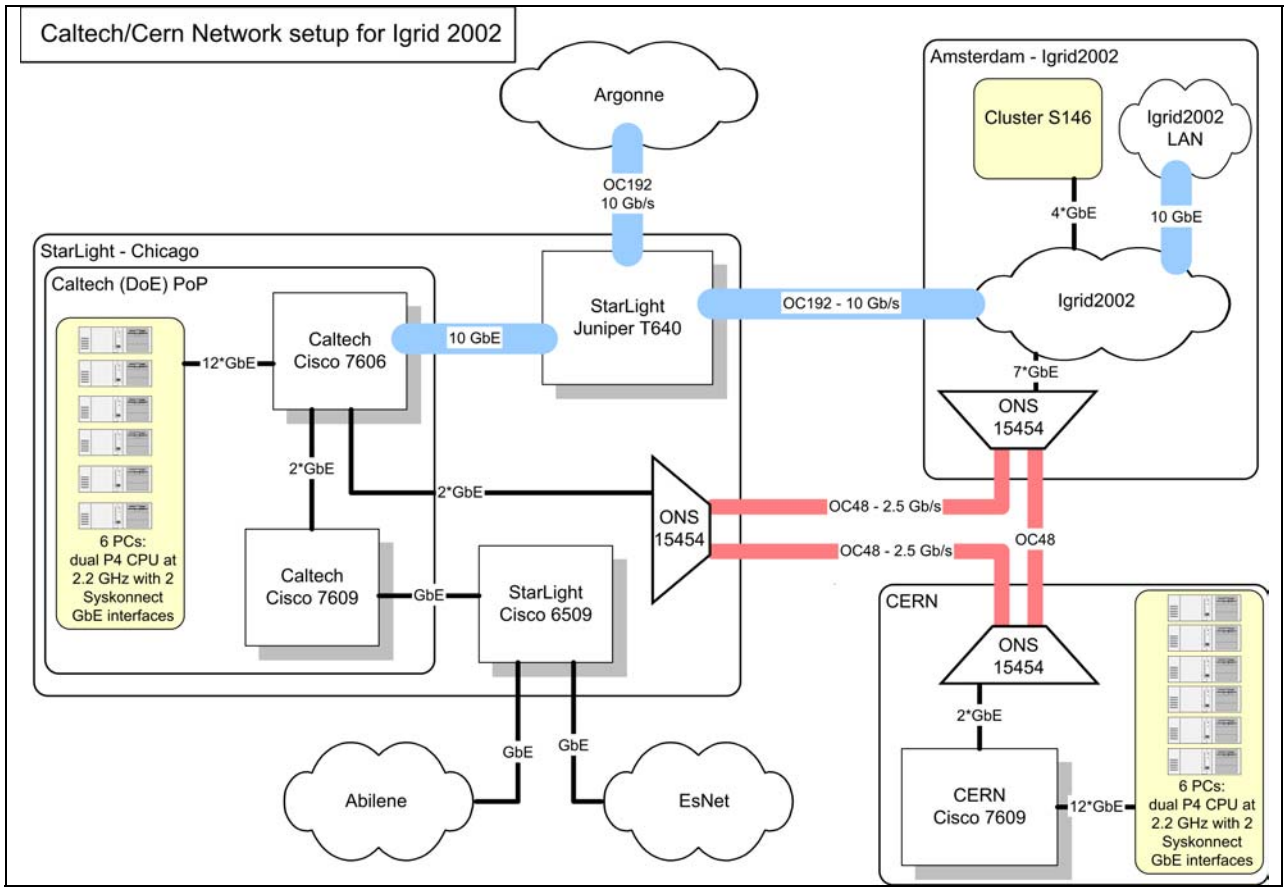
A.1.a. CERN

CERN connected to the StarLight facility in late March 2002 with an OC-12 (622Mb). The provider, KPNQwest, went out of business in May, forcing CERN to temporarily transit traffic via Géant. In September 2002, CERN began using a new 622Mb link from Chicago to Geneva/CERN provisioned by Deutsche Telekom/T-Systems. CERN established connectivity with StarLight's IPv6 infrastructure.

CERN's OC-12c circuit supports research production traffic between IN2P3 and CERN to vBNS, ESnet, Abilene, etc. (IN2P3 is the National Institute of Nuclear and Particle Physics, a CNRS [French National Center for Scientific Research] institute.) After 2002, the link will be upgraded by a factor of ~2 each year, to reach 10Gb by 2006.

The European Union (EU) DataTAG 2.5Gb circuit between CERN and Starlight was successfully delivered by T-Systems on August 20. The full DataTAG testbed could not be deployed immediately because of late ordering and delayed delivery, so the circuit was initially connected to Cisco 7600 routers. The full layer2/layer3 testbed was to be ready by the end of October. CERN plans to upgrade the EU DataTAG link to 10Gb, and to connect to the TeraGrid at 2.5Gb (and then upgrade to 10Gb), in 2003.

Below is the Caltech/CERN/DataTAG network configuration, developed in time for the *iGrid 2002* event in Amsterdam at the end of September.



Equipment for the EU DataTAG 2.5Gb circuit between StarLight (Chicago) and CERN (Geneva), including servers, switches and routers, was provided by Caltech (with USA Department of Energy funding) and StarLight/University of Illinois at Chicago (with USA National Science Foundation funding). Equipment in Europe was provided by the European Union in the framework of the EU DataTAG Project. To take advantage of the transatlantic OC-192 donation between Chicago and Amsterdam that was provided by Level 3, Cisco loaned Caltech a 10GigE module and a 16-port GigE module.

A.1.b. IUCC

Israel disconnected its STAR TAP connection in May 2002 and no longer participates in Euro-Link funding.

A.1.c. NORDUnet

NORDUnet's 622Mb transoceanic link to Abilene's New York POP was provisioned by Deutsche Telekom/T-Systems. From New York, it connects to StarLight via a 155Mb Qwest link. NORDUnet wants to maintain this connection until August 2003, when it hopes to bring a 2.5Gb wavelength to the US.

Update: NORDUnet's agreement to temporarily pass NaukaNet routes has terminated. NaukaNet, the Russian research and education network, had been using Teleglobe as its transoceanic provider, which went out of business in May 2002. The new NaukaNet OC3 circuit provisioned by Telia was connected to StarLight in September.

A.1.d. Renater2

RENATER2 maintains an OC-3 link to STAR TAP. We have received no information on possible future upgrades.

A.1.e. SURFnet

In early September, Level 3 donated a temporary 10Gb transoceanic wavelength from Chicago (StarLight) to Amsterdam (SARA) in support of iGrid 2002 (in addition to the 2.5Gb link from Chicago to Amsterdam paid for by SURFnet). Beginning October 1, SURFnet negotiated with Level 3 to keep the 10Gb link, and turn off the 2.5Gb link, through the end of the year. T-Systems will provide the 10Gb lambda as of January 2003.

SURFnet plans to terminate its remaining OC-12 connection to StarLight on December 31, 2002. It will continue its OC-12 to New York donated by IEEAF/Tyco. (Note: The OC-12 at StarLight connects SURFnet's Cisco 12008 border router to its backbone in The Netherlands. In January, SURFnet will use 1Gb of its 10Gb lambda for border router connectivity.)

SURFnet's four major peers in the US are Abilene, CA*net, Esnet and vBNS. Abilene, ESnet and CA*net4 are currently connected to the StarLight router, with the vBNS planning to connect soon. Until then, SURFnet continues to use the shared OC-12 between StarLight and STAR TAP (at Ameritech) in order to reach the vBNS.

SURFnet plans to upgrade its border router in the first quarter of 2003. Also, SURFnet has a Cisco ONS 15454 (optical transport platform) at StarLight, which others (e.g., CERN, MREN) are welcomed to use.

The UIC National Center for Data Mining (NCDM) has clusters at both StarLight and NetherLight in Amsterdam. EVL and NCDM will soon begin testing applications between Chicago and Amsterdam at 2xGigE rates.

A.2. Engineering Services

A.2.a. StarLight/Abilene Connectivity

In September, Abilene established a 10Gb connection between Abilene and StarLight to enable its participating institutions to use the StarLight/NetherLight link. Its Juniper T640 core router, located at the Qwest PoP, with two OC-192 circuits to other Abilene backbone sites, connects to StarLight via two I-WIRE dark fiber connections. One 10GigE connection is for StarLight National Research Network traffic (e.g., SURFNet) and a 1GigE connection is for MREN traffic.

Abilene will maintain its direct connection to the STAR TAP NAP (currently OC-12, but may be reduced to OC-3 based on decreasing demand) until at least the fourth quarter of 2002 (April 2003 at the latest). After that, Abilene will decide whether to continue its direct connection based on the number of peer networks still connecting there.

Note: In addition to its StarLight connection, Abilene has a 10Gb connection between its New York City POP and NetherLight that was provisioned by Tyco Telecom through the Internet Educational Equal Access Foundation (IEEAF) in September 2002.

A.2.b. StarLight/STAR TAP Connectivity

An OC-12 connection exists between the StarLight facility and STAR TAP (at the Ameritech NAP) that is available for shared use by StarLight customers. Funding for this link ends May 1, 2003, with a one-year extension likely.

A.2.c. STAR TAP Router

In our distributed STAR TAP/StarLight environment, there is a STAR TAP Router at the Ameritech NAP (Cisco 7505), a STAR TAP Router at 710 N. Lake Shore Drive (Juniper M10), and an MREN Router at 710 (Juniper M5). A Cisco 6509 serves as an additional MREN Router as well as the StarLight Gigabit Ethernet Exchange Point switch.

A.2.d. STAR TAP/StarLight NLANR Web Cache

For several years, NAP.NET (bought later by Genuity) donated 1Mb of Internet commercial transit service to STAR TAP/StarLight, which was used to support the NLANR web cache for international STAR TAP participants. Recently Genuity notified us that it would begin to charge for its service. We determined that due to the *much* better Internet connectivity in Europe now, the need for a web cache in Chicago has decreased substantially. Hence, we terminated Genuity's service and we are returning the web cache to NLANR.

A.2.e. 6TAP

Both STAR TAP and StarLight host IPv6 routers, provided by ESnet.

A.2.f. DiffServ

Oliver Yu and Jason Leigh are setting up the EMERGE 2 DiffServ testbed between EVL, Northwestern University, and KISTI (Korea Institute of Science and Technology Information) over KREONET/STARTAP. In EMERGE 1, priority queuing was tested between EVL and Argonne National Lab. EMERGE 2 will attempt to employ Weighted Fair Queuing. CERN has expressed interest in participating in DiffServ experiments.

Students working under Oliver Yu are designing schemes for comparing QoS performances for different DiffServ types, investigating TCP tuning over DiffServ networks, and comparing per-flow policing versus per-class policing.

A.2.g. StarLight/STAR TAP Documentation

In the past year, the networking engineers were so busy moving equipment and links between STAR TAP and StarLight that it was difficult for us to keep web documentation current. We are now in the process of updating both the StarLight and STAR TAP web sites.

A.2.h. International Transit Network (ITN)

Rather than bring circuits to STAR TAP or StarLight, some international research networks peer with Abilene at one of the US coasts, where they can pass traffic to US universities (via Abilene) and to other international research networks (via Abilene and CA*net ITNs). Further information is available on the Abilene [www.ucaid.edu/abilene/html/itnservice.html] and CA*net web sites [www.canet3.net/optical/documentation.html], as well as STAR TAP's [www.startap.net/CONNECT] page.

A.3. NOC Services

In August, the NOC completed work on a new internal database of Global network contact information.
The Global NOC moved to the NetSaint network monitoring system; however, the program is not currently available on the web page due to technical difficulties with security authentication. The NOC is working on a corresponding reporting tool linked to NetSaint that will make the information available on the pages in the future. NetSaint replaced Whatsup Gold and the NOC-developed BGP session monitor <http://www.netsaint.org/>.

MRTG graphs depicting StarLight traffic appear on the STAR TAP NOC web page under Network Monitoring <http://noc.startap.net/noc.html>. Traffic monitors for the StarLight M10, 6509 and LS1010 appear at: <http://loadrunner.uits.iu.edu/mrtg-monitors/starlight/>.

B. Euro-Link Performance Analysis Tools

B.1. Network Monitoring Tools

Bandwidth Utilization Radar Map

A graphical radar map for StarLight routers will soon be available on the StarLight website.

UCAN: Unified Collaboratory for Analyzing Networks

UCAN is an integrated, collaborative, extendable workbench tool for network engineers, researchers and application programmers using networks. UCAN runs collaborative network tests between multiple sites, thus helping analyze network performance in group environments. The first prototype was deployed at KISTI (Korea Institute of Science and Technology Information). Ongoing work includes the integration of UCAN with Quanta and the development of a shared widget library independent of UCAN.

For iGrid 2002, UCAN was used to perform network tests over the Level 3-donated link and to help debug network performance for some of the iGrid applications.

B.2. High-Bandwidth Transmission Over Long Distance Networks

Quality of Service Adaptive Networking Toolkit (Quanta)

In July, EVL repackaged and distributed its robust CAVERNsoft toolkit as two packages: CAVERNsoft for user collaboration and tele-immersion, and Quanta, for network communications. Quanta, which is specifically targeted for optical networks, is an applications-level API that translates high-level data distribution requirements into low-level optimized networking protocols and parameter settings. Whereas CAVERNsoft requires Quanta in order to run, the reverse is not true, enabling Quanta to support many more applications.

Quanta provides a rich set of networking tools and data distribution mechanisms for high-performance applications, including: message passing, distributed shared memory, remote procedure calls, remote file I/O, Forward Error Corrected UDP, Parallel TCP for bulk data transfer, and collaborative performance monitoring. A new data transfer scheme Reliable Blast UDP (RBUDP) was recently added to Quanta (see below).

Work is underway to provide Quanta with mechanisms to signal DiffServ-capable routers and make dedicated lightwave reservations on optically switched networks. Quanta's DiffServ testbed consists of a set of DiffServ routers at EVL, KISTI (Korea Institute of Science and Technology Information) and Northwestern University. Quanta's optically switched testbed consists of StarLight and OMNInet.

Reliable Blast UDP (RBUDP)

RBUDP is a technique to accelerate reliable data transmission over fat networks. In RBUDP, the sender blasts all the data (each packet is identified by a sequence number) to the receiver. Upon receipt, the sequence numbers are checked and any lost packets are identified. The receiver then sends a lost packet report back to the sender through TCP. Upon receipt of the report, the sender retransmits the lost packets. This procedure continues until the receiver receives all packets. This technique is believed to be most effective when used in conjunction with QoS, since guaranteed bandwidth may minimize transmission errors. The RBUDP scheme exploits low transmission errors to maximize throughput.

An analytical model was developed to predict the performance of RBUDP, and experiments were performed on the testbed between Chicago and Amsterdam. A related paper was published in IEEE Cluster 2002. During iGrid 2002, Vrije University used RBUDP to run their parallel video streaming application.

The current RBUDP is used primarily for bulk data transfer, although it could be used to stream data with some tradeoff in latency. Next steps are to design and implement a variation of RBUDP specially optimized for real-time streaming applications.

Forward Error Corrected RBUDP

Applying FEC to RBUDP will commence next quarter.

B.3. Ultra-High-Bandwidth Transmission Over Long Distance Networks (StarLight)

Interrupt Coalescing and Jumbo Frames

Two 16-node Linux clusters arrived in August; one for the StarLight facility and one for EVL. In early September, one partially assembled system was shipped to iGrid 2002 to support tiled-display applications and a second partially assembled system was installed at StarLight. The fully assembled clusters will hopefully be operational by December.

The StarLight cluster will augment the four-node PC cluster currently used by EVL to run tests to Northwestern University. The new cluster will have three times the bus bandwidth (400MHz x 64bits=2.98GBytes/s) of the current four-node cluster. In the past, EVL was able to only achieve ~500Mb (local area) PC transmission rate out of the box. With appropriate tuning using *Interrupt Coalescing* and *Jumbo Frames*, EVL can achieve 850Mbps.

EVL is working with Cees de Laat at University of Amsterdam and Paul Wielinga of SARA to perform tests over the 2.5Gb (now 10Gb) NetherLight/StarLight link. Tests include using AT&T's Virtual Network Collaboration (VNC) software to stream desktop graphics, as well as EVL's TeraVision hardware/software to stream mono/stereo graphics.

TeraVision: Ultra-Resolution Visualization Streaming

TeraVision is a hardware- and software-independent solution for real-time image distribution in advanced collaborative environments. The goal of the TeraVision project is to send high-resolution video streams between clusters using distributed servers and clients. The software enables multiple streams of synchronized video to be streamed between clusters, thus making it possible to send information to tiled or stereo displays. Future versions of TeraVision will incorporate multicasting and collaboration controls, enabling disperse groups to make collaborative presentations, particularly when used in conjunction with the Access Grid.

In July, code for UDP streaming was written, integrated and tested. Comprehensive testing of UDP and TCP streaming were done between Chicago and Amsterdam to identify performance bottlenecks. (Results will be published in a forthcoming paper.) Tests were run to see TeraVision's ability to stream stereoscopic video. Some coding-related synchronization problems were identified and fixed.

In August, code was written to perform video compression in real time. Considerable time was spent getting the code to run fast under Windows. We were ultimately able to push single streams at 400Mb (or 20 frames-per-second) using UDP between Amsterdam and Chicago. SGI Onyx machines were tested with TeraVision boxes. Video compression code will be integrated in the coming quarter. Also written was a *floor control* feature, which will enable multiple users to collaborate by acting as either servers (send video) or clients (receive video). This too will be integrated in the coming quarter.

In September, TeraVision boxes were shipped to Amsterdam and to Greece to support several iGrid 2002 applications. For more information, see <http://www.evl.uic.edu/cavern/teranode/teravision>

LambdaRAM: Optically Connected Wide Area Network Memory

LambdaRAM is an application being developed to address long-haul latency in optical networks. This technique collects memory in a compute cluster and then allocates it as a cache to minimize the effects of latency over long-distance, high-speed networks. LambdaRAM takes advantage of multiple-gigabit networks (available on the StarLight and OMNInet testbeds) to pre-fetch information before an application is likely to need it (similar to how RAM caches work in computers today). LambdaRAM extends this concept over high-speed networks. We have started testing LambdaRAM (UDP blasting) between the UIC National Center for Data Mining's cluster and EVL's cluster. For more information, see <http://www.evl.uic.edu/cavern/teranode/gridram>

C. Accomplishments

C.1. Meetings

September 27, 2002. A small closed Lambda Workshop, a follow-up to the TERENA-organized workshop held in 2001, was held at University of Amsterdam. Representatives from NSF (Tom Greene), StarLight (Tom DeFanti, Maxine Brown, Joe Mambretti, Linda Winkler), I-Light/Indiana, TeraGrid, NORDUnet, SURFnet, Japan, TERENA, UK, Internet2, IEEAF, CERN, the Czech Republic and CANARIE attended. Minutes of the meeting are available upon request.

September 23-26, 2002. iGrid 2002 <www.igrid2002.org> took place. iGrid 2002 was organized by Dutch and USA organizations. Institutions in The Netherlands were: Amsterdam Science & Technology Centre, GigaPort Project, SARA Computing and Networking Services, SURFnet and Universiteit van Amsterdam/ Science Faculty. Institutions in the USA were: Argonne National Laboratory/ Mathematics and Computer Science Division, Indiana University/ Office of the Vice President for Information Technology, Northwestern University/ International Center for Advanced Internet Research, and University of Illinois at Chicago/ Electronic Visualization Laboratory. Major funding for iGrid 2002 was provided by the GigaPort Project, the Amsterdam Science & Technology Centre and the USA National Science Foundation, with in-kind support by SARA Computing and Networking Services (with funding from the NWO/NCF) and the Universiteit van Amsterdam.

September 17, 2002. Alan Verlo participated (via teleconference) in a JET meeting held at NSF in Arlington, VA.

September 15, 2002. Tom DeFanti and Maxine Brown met with Cees de Laat, a research collaborator from University of Amsterdam, who was in Chicago for a brief time to attend a meeting of the Global Grid Forum Executive Committee.

August 26-28, 2002. Tom DeFanti attended the MAGIC (Middleware And Grid Infrastructure Coordination Committee) Workshop sponsored by the Large Scale Networking (LSN) Coordinating Group of the Interagency Working Group for Information Technology Research and Development. The goal of the Workshop is to produce a report that describes a vision for middleware and Grid services that will enable scientific applications and cooperation (5 to 10 years out) and to identify needed Federal networking research to implement that vision. This *Blueprint for Future Science Middleware and Grid Research and Infrastructure* will also be used by LSN agency members to develop their research plans and programs. It will also be used to inform decision makers and researchers of the research needed in middleware and grids, and the need for a persistent infrastructure to support their use. <http://www.nsf-middleware.org/MAGIC/>

August 20, 2002. Alan Verlo participated (via teleconference) in a JET meeting held at NSF in Arlington, VA.

August 9, 2002. Level 3's Geoff Jordan, Kevin O'Hara, John Verduzco and Sara Bleau visited EVL to meet with Maxine Brown, Jason Leigh and Joe Membretti to learn about EVL and NU optical networking activities. Subsequent to the meeting, Level 3 donated a temporary 10Gb link from StarLight to SARA in support of iGrid 2002.

August 8, 2002. Level 3's John Verduzco and Sara Bleau visited EVL to meet with Maxine Brown, Jason Leigh and Laura Wolf.

August 7, 2002. Gert Svensson, KTH/PDC (Royal Institute of Technology, Center for Parallel Computers), visited EVL to meet with Jason Leigh to learn about current VR hardware and software research, including Linux clusters and to get an overview of EVL's computing and networking environments and Grid activities.

July 28-August 1, 2002. Alan Verlo, Linda Winkler, Caren Litvanyi and Bill Nickless attended the NLANR/Internet 2 Joint Techs Workshop in Boulder, Colorado. On July 31, they participated in a JET (Joint Engineering Taskforce) meeting.

C.2. Activities

iGrid 2002, Amsterdam, September 23-26, 2002.

Last quarter, considerable time and effort was spent preparing for the third biennial international Grid (iGrid) 2002 event at WTCW in Amsterdam. EVL, along with University of Amsterdam and Gigaport, co-organized the applications-driven testbed event, which showcased 28 scientific and art applications from 16 countries. StarLight served as the optical exchange point for many of the applications and a temporary Level 3-donated link augmented the SURFnet 2.5Gb link between Chicago and Amsterdam in support of the applications. The event was tremendously successful, attended by over 250 people and providing the world's research community an opportunity to work together to advance the state of the art in LambdaGrid-intensive computing. Notable achievements include:

- SURFnet/SARA stress-tested the LAN and WAN networks and routers by generating and sustaining 8Gb of data in and out of the facility for one hour on the last day of demonstrations. Bandwidth usage statistics can be found on the iGrid website <www.igrid2002.org> (scroll to "A page with bandwidth usage statistics is available") or see <http://www.igrid2002.org/bandwidth/index.html>.

- A Terabyte of research data was transferred in less than a DVD/minute over a newly established "light path" extending 12,000 kilometers from TRIUMF in Vancouver to CERN. The demonstration required dedicated portions of fiber-optic networks, spanning one provincial (British Columbia's BCnet) and two national research and education networks (CA*net4 and SURFnet) to establish the on-demand private network. Experiments were also conducted on the Chicago to CERN DataTAG link. The project culminated in establishing the first large-scale end-to-end "light path". ***Researchers were able to transfer a***

Terabyte of research data from disk-to-disk in under 3 hours – this is equivalent to transferring a full CD in less than 8 seconds (or a full-length DVD movie in less than 1 minute). Using *bbftp,* they could transfer 60GB files in 10 minutes; 1TB is about 17 of these files, or about 170 minutes transfer time. Using *Tsunami,* they achieved slightly better rates. Peak transfer rates in excess of 1Gb were achieved, twice the previous known record for this distance. This is the first establishment of an inter-domain end-to-end "light path" dedicated for a research application. The "light path" directly connecting TRIUMF and CERN is the longest known single hop network spanning the distance from Vancouver to Geneva via StarLight, NetherLight and DataTAG. (Note: Performance was affected by using 10GigE alpha-test cards from Intel. Subsequent testing with beta-test cards has shown that they perform roughly 4 times better.)

- UIC's Laboratory for Advanced Computing/National Center for Data Mining (LAC/NCDM) and the International Center for Advanced Internet Research (iCAIR) at Northwestern University set a transatlantic data-transfer record using Photonic Data Services (PDS), a technique they jointly developed. Using PDS, data was transmitted at 2.8Gb as part of a data-mining application. The demonstration was the first for PDS and shows the potential for data-mining applications to drive the use of available bandwidth.

- The first long-term, transatlantic 10Gb wavelength circuit dedicated to research and education was inaugurated at iGrid 2002 with a demonstration of uncompressed HDTV over IP sponsored by the Pacific Northwest GigaPOP and the Research Channel. The circuit, provided by Tyco Telecommunications to the Internet Educational Equal Access Foundation (IEEAF), was provisioned from the US to Europe through the efforts of Internet2 and SURFnet.

- Artist Jackie Matisse's application "Kites Flying In and Out of Space" used a Grid model for real-time steering of calculations on computers distributed worldwide over high-speed networks. Each of the 12 kites appearing in the piece utilized up to 15Mbps. The application used distributed servers in Chicago, Canada, Japan, Singapore and Virginia to compute each kite's physical dynamic properties. Each server sent a single kite's motions to iGrid where it was visualized in a CAVE. The application was scalable both computationally and geographically. It proved a good test of high-speed networking because the application required a multicast-enabled network to accomplish communications. As a side benefit, the kites become a visual metaphor for network performance as the kite motions (e.g., fast, slow) responded to latency of the network data.

C.3. Publications

E. He, J. Leigh, O. Yu, T.A. DeFanti, "Reliable Blast UDP: Predictable High Performance Bulk Data Transfer," 4th IEEE International Conference on Cluster Computing, Chicago, Illinois, September 2002. <http://www.evl.uic.edu/cavern/papers/cluster2002.pdf>

M. Thorson, J. Leigh, G. Maajid, K. Park, A. Nayak, P. Salva, S. Berry, "AccessGrid-to-Go: Providing AccessGrid Access on Personal Digital Assistants," in Proceedings of Access Grid Retreat, La Jolla, California, 2002.

J. Leigh, J. Girado, R. Singh, A. Johnson, K. Park, T.A. DeFanti, "TeraVision: a Platform and Software Independent Solution for Real Time Display Distribution in Advanced Collaborative Environments," in Proceedings of Access Grid Retreat, La Jolla, California, 2002.

C.4. Software Releases

In July, the new CAVERNsoft and Quanta software systems were released. <http://www.openchannelsoftware.org/projects/CAVERNsoft_G2>.

D. Collaboration Activities

Working with SARA to do network performance studies over long, fat networks using various transmission techniques (TCP, UDP, FEC, RUDP).

Working with CERN on RUDP and DiffServ tests.

Working with University of Amsterdam to investigate integrating their AAA (Authentication, Authorization & Accounting) management-level middleware with UIC/NU's intelligent-signaling control-plane middleware; this will be applied to optical networks, both intra-domain (OMNInet) and inter-domain (StarLight).

E. Problems

No significant problems were encountered this quarter.

F. Any Proposed Changes in Future Plans

No changes to date.

G. Summary of Award Expenditures (July-September)

Available upon request.