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

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

February 25, 2003

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/DataTAG

The advanced layer2/layer3 multi-vendor 2.5Gb DataTAG testbed is in place except for the Alcatel 1670 layer2 multiplexer. Installation has been delayed until February 2003 due to its need for a channelized rather than concatenated circuit in order to transparently map Gigabit Ethernet frames into SONET/SDH frames. 10GigE tests with Intel NICs between CERN, StarLight, Caltech and Sunnyvale (SLAC) will begin soon, after which the layer2/layer1 Alcatel multiplexers will be deployed between CERN and StarLight.

CERN is now in the final negotiation stages with Deutsche Telekom/T-Systems for the consolidation of its two circuits to StarLight (622Mb for production traffic and 2.5Gb for R&D/DataTAG traffic) onto a single 10Gb transparent optical lambda on September 1, 2003, thus providing the missing part of the 10Gb triangle between Amsterdam, Chicago and Geneva.

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.

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.

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 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.

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 October, as per an agreement negotiated by SURFnet after iGrid 2002, Level 3 maintained its 10Gb Chicago to Amsterdam transoceanic link and turned off the 2.5Gb link. On December 17, SURFnet engineers began testing a new 10Gb lambda delivered by T-Systems. T-Systems replaced the Level 3 connection in January 2003. SURFnet also has access to a 10Gb link from Amsterdam to New York donated by IEEAF/Tyco.

A.2. Engineering Services

A.2.a. StarLight/Abilene Connectivity

Abilene's Juniper T640 core router, located at the Qwest PoP, along 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 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) until April 2003. The number of peer networks still connected there will determine whether Abilene continues its direct connection.

A.2.b. StarLight/STAR TAP Connectivity

An OC-12 connection exists between StarLight and STAR TAP (at the Ameritech NAP) for shared use by StarLight customers. This OC-12 StarLight/STAR TAP link will be maintained until April 30, 2004.

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

Last quarter, Genuity's no-cost agreement to provide 1Mb of Virtual Transit Service ended. The service had been used since March 2000 to support a NLANR web cache for international STAR TAP participants. Due to great improvements in connectivity in Europe, STAR TAP engineers decided not to procure a new contract. The service ended October 25. On November 1, the web cache was removed by STAR TAP engineers and returned to NLANR.

A.2.e. 6TAP

Both STAR TAP and StarLight host IPv6 routers, provided by ESnet. For information, see <http://www.6tap.net>.

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, premium service provisioning for tele-immersive applications across a single domain was tested between EVL and Argonne National Lab. In EMERGE-2, assured service provisioning for multimedia applications across multiple domains will be deployed. The results of these experiments will be compared to the results of using the over-provisioned (uncongested) 10Gb SURFnet and 2.5Gb DataTAG links.

NLANR (National Laboratory for Applied Network Research) is providing EVL with an OC3-MON monitoring tool for the EMERGE-2 testbed. Students working under Oliver Yu are designing schemes for GARA/Globus-based resource reservation, TCP and RBUDP (Reliable Blast UDP) transmissions, differentiated transmissions of MPEG-2 video applications, and adaptive QoS control over this multi-domain DiffServ network.

A.2.g. StarLight/STAR TAP Documentation

An updated list of StarLight providers appears at:
<http://www.startap.net/starlight/CONNECT/connectCarrierInfo.html>

The updated STAR TAP Networks list appears at: <http://www.startap.net/NETWORKS/>

The iGrid 2002 applications have been added to the StarLight web site:
<http://www.startap.net/starlight/igrid2002/index.html>

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

Some international research networks opt to 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

The Global NOC has moved to Nagios (an updated version of Netsaint) network monitoring system and is now writing scripts to present the individual network up/down status on the Euro-Link NOC home page in a secure fashion. Links should be added in the next quarter.

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

New software is being developed that will be capable of graphically depicting StarLight network traffic.

UCAN: Unified Collaboratory for Analyzing Networks

UCAN is a network workbench tool for collaborative performance monitoring, network testing and management. The first prototype of UCAN has undergone miscellaneous bug fixes and improved to use the Quanta networking toolkit. The new additions also include automated installation scripts for easy configuration of the libraries.

Microsoft recently expressed interest in using UCAN for application performance monitoring. The new version of UCAN has been packed and sent for internal use. Future plans for UCAN include using it for further DiffServ tests on the EMERGE-2 testbed between EVL and KISTI.

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

Quanta: Applications-Centric Communications Middleware

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.

Quanta is an applications-level API that translates high-level data distribution requirements into low-level optimized networking protocols and parameter settings, and is specifically targeted for optical networks. 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, Reliable Blast UDP (RBUDP), and collaborative performance monitoring.

Reliable Blast UDP (RBUDP)

Last quarter, we recreated the RBUDP experiments performed during iGrid 2002 and achieved a 725 Mb read data transfer rate between a pair of Linux PCs located in Chicago and Amsterdam.

During iGrid 2002, new EVL Senior Research Scientist Luc Renambot applied Quanta's RBUDP to a parallel graphics streaming application called Griz. Using our analytical model and parameters, we were able to predict the number of animation frames that Griz had to package into a single payload to achieve full utilization of the Amsterdam-Chicago StarLight link. In terms of real-time streaming optimization, we completed the analytical modeling and prediction part; implementation is planned in the coming quarter.

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. The current RBUDP is used primarily for bulk data transfer, although it could be used to stream data with some tradeoff in latency.

Forward Error Corrected RBUDP

Applying FEC to RBUDP will commence next quarter.

CERN/DataTAG and Caltech Experiments

CERN reported interesting results with respect to high-performance single stream TCP/IP streams using a variety of conventional and innovative new stacks (e.g., HSTCP, FAST, Scalable TCP, Grid DT). The results were achieved over the DataTAG circuit and its extensions to Amsterdam, Sunnyvale and Baltimore (SC'02 conference) with DataTAG project partners and associated teams at Caltech and SLAC. Although CERN is not part of TeraGrid, agreements were made to connect it to Sunnyvale and Baltimore via the StarLight router to allow DataTAG to participate to SC'02.

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

Resources

Two new 16-node Linux clusters, one for the StarLight facility and one for EVL, were purchased. The EVL cluster, connected to a 3x5 tiled display, became operational in November. A partially assembled system was installed at StarLight in September to support iGrid 2002 (6 nodes of 16 are currently operational); facility power issues will delay the full StarLight installation until the end of summer 2003.

UIC's National Center for Data Mining (NCDM) expanded its 4-node cluster to 7 nodes at StarLight to support its iGrid 2002 application. NCDM's Terra Wide Data Mining (TWDM) testbed consists of a 3-node cluster (each node has a 1GigE connection, for a total of 3GigE) in Amsterdam, a 4-node cluster with 4 OC-12 in Ottawa, and a 7-node cluster with 7 GigE at StarLight. NCDM added a temporary 4-node cluster with 4 GigE at Baltimore for SC'02.

Interrupt Coalescing and Jumbo Frames

The new 16-node cluster at EVL augments a four-node cluster currently used by EVL to run tests to Northwestern University. The new cluster has 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 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 tile or stereo displays. See: <http://www.evl.uic.edu/cavern/optiputer/teravision>.

During October and November, more experiments were performed to collect performance figures. The video compression algorithm was rewritten. To ensure maximum efficiency of other similar modules, different compilers were also evaluated. A module for multicasting was written, tested on Linux and ported to Windows. The team commenced an investigation of Linux-based frame-grabbers that would allow us to run TeraVision servers entirely on Linux. In December, the decision was reached that TeraVision v2.0 would be written from scratch to provide more flexibility. The design for v2.0 is now complete and will have the following advantages over v1.x:

- New networking protocols, compression modules, frame grabbers and other video manipulation modules using a quick plug-in approach. The new design loads the appropriate module at run-time, making the design cleaner and easier to maintain.
- An RLE compression module. Newer compression algorithms will be integrated later.
- Multicasting capabilities (with scatter-gather calls) and RBUDP based on student Eric He's RBUDP code.
- A protocol for an intelligent server-client communication mechanism that will make it easier to control multiple clients and servers on different machines.
- An intuitive User Interface (UI) to allow the user to easily control more parameters in each of the modules (i.e., compression quality or network packet sizes). The UI will contain visualization tools for analyzing performance of different system components.
- A console application for remote login and control, and a solution for X x Y to M x N display problem.
- File streaming using different file formats (.gif, .bmp, .jpeg, .ppm, .xpm, etc.)

TeraVision multicasting and collaboration controls will enable disperse groups to make collaborative presentations, particularly when used in conjunction with the Access Grid.

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 (NCDM) cluster and EVL's cluster. For more information, see <http://www.evl.uic.edu/cavern/optiputer/lambdaram>.

In October, LambdaRAM was tailored to support TeraScope and demonstrated at a GGF6-affiliated event hosted at EVL. For the demo, we implemented LambdaRAM as a memory pool to pre-fetch data from NCDM's remote database server for display on an EVL cluster display. LambdaRAM is designed to maximize network bandwidth utilization in order to match the computational capability of the cluster. In November, we gave a similar demonstration in the Project DataSpace booth at SC'02 in Baltimore. In December, we worked on benchmarks for

LambdaRAM and began modifying LamdaRAM to be generic middleware that will support other applications besides TeraScope.

C. Accomplishments

C.1. Meetings

December 17, 2002. Tom DeFanti, Alan Verlo and other STAR TAP engineers teleconferenced into the Joint Engineering Team (JET) meeting held at NSF in Arlington, Virginia. Tom gave a presentation about the future of STAR TAP and StarLight. Discussion took place about deploying IPv6 at various GigaPoPs, including StarLight.

November 18-22, 2003. Tom DeFanti, Maxine Brown, Jason Leigh, Greg Dawe and several students attended SC'2002. Jason Leigh and his students demonstrated TeraScope in the "Project DataSpace" research booth organized by Bob Grossman of the UIC National Center for Data Mining (NCDM). Grossman demoed a collaborative project with researchers from Chicago, Ottawa and Amsterdam, which was awarded the SC'02 High Performance Bandwidth Challenge Award for Innovative, High Speed, Data Correlation – Best Use of Emerging Infrastructure. The group included researchers from NCDM, CANARIE, and SARA, who have been working together over the past year to produce real-time merging of data over lambda networks. At SC02, they presented the first demonstration of the technology, with impressive results. A stream of data was moved over SURFnet connecting a cluster of computers at SARA to a cluster of computers at StarLight at over 2.8Gb. At the same time, a stream of data was moved over Canada's CA*net4 network connecting a computer cluster at CANARIE in Ottawa and a UIC computer cluster at StarLight in Chicago at over 2Gbps. Both streams used a new. NCDM-developed protocol called SABUL designed for high-performance data transport. The two streams of data were merged at over 500Mbps per node on StarLight's three-node cluster. These so called "lambda joins" are an important component for distributed data mining applications

November 20, 2003. Alan Verlo teleconferenced into the Joint Engineering Team (JET) meeting held at the SC'2002 conference in Baltimore. Linda Winkler and other STAR TAP engineers were in attendance.

October 18, 2002. Amsterdam/Chicago Optical Research Collaboration held at UIC/EVL. Attendees included: Cees de Laat and Leon Gommans (UvA), Peter Clarke (UCL), Valerie Taylor, Joe Mambretti, Jim Chen, Elizabeth Bacon, Fei Yeh, David Lillethun (NU), Tom DeFanti, Maxine Brown, Jason Leigh, Oliver Yu, Mitch Theys, Bob Grossman (UIC). Agenda and PowerPoints at: http://www.startap.net/starlight/ABOUT/meetOpResCollab02.html

October 15, 2003. Alan Verlo teleconferenced into the Joint Engineering Team (JET) meeting held at the ACCESS Center in Arlington, Virginia. He reported on the success of iGrid, and answered questions from other Fednet representatives on StarLight/STAR TAP connectivity. JET is a standing committee of the Large Scale Networking (LSN) Coordinating Group (CG) of the Interagency Working Group (IWG) on Information Technology Research and Development (ITRD). The IWG reports to the National Science and Technology Council (NSTC) and its Committee on Technology (CT). The LSN and its teams serve as the forum for coordinating activities of the LSN Program Component Area.

October 14, 2002. In conjunction with Global Grid Forum 6 (GGF6) being held in Chicago, EVL hosted an open house on Monday, October 14, from 4-6 PM <http://www.gridforum.org/Meetings/ggf6/default.htm>. We showcased several iGrid demonstrations from UIC/EVL, UIC/NCDM, NU/ iCAIR and NCSA ACCESS Center at EVL as part of a self-guided tour of the lab to educate the global Grid community about StarLight as an enabling facility to large-scale global Grid applications. Featured applications included: TeraScope, Kites Flying In and Out of Space, High Performance Data Webs, and Photonic TeraStream. Approximately 200 visitors came, including Euro-Link partners from Amsterdam, CERN and the UK; photos from the event are documented at <ftp://ftp.evl.uic.edu/pub/INcoming/raj/ggf%20pics/>.

October 11, 2002. Tom DeFanti, Maxine Brown, Jason Leigh and Oliver Yu (EVL), and Joe Mambretti and Jeremy Weinberger (NU/iCAIR), met with Jack Waters (CTO), John Verduzco, Sarah Bleau and Jason Booma of Level 3 to discuss OMNInet and OptIPuter projects. Level 3 presented and demonstrated its onTAP product.

C.2. Activities

GGF6 Open House, October 14, 2002.

EVL hosted an open house for attendees of the Sixth Global Grid Forum (GGF6) Workshop in Chicago. The event featured demonstrations of TeraScope: Visual Tera Mining (EVL), Kites Flying In and Out of Space (NCSA/UIUC and EVL), High Performance Data Webs (LAC/UIC) and Photonic TeraStream (iCAIR and Northwestern University). The demonstrations were selected to promote StarLight as an enabling facility to large-scale global Grid applications.

C.3. Publications

R. Singh, J. Leigh, T. A. DeFanti, "TeraVision: A High Resolution Graphics Streaming Device for Amplified Collaboration Environments," Journal of Future Generation Computer Systems (FGCS), Elsevier Science Press, to appear.

E. He, J. Alimohideen, J. Eliason, N. Krishnaprasad, J. Leigh, O. Yu, T. A. DeFanti, "QUANTA: A Tooklit for High Performance Data Delivery over Photonic Networks," Journal of Future Generation Computer Systems (FGCS), Elsevier Science Press, to appear.

S. Venkataraman, J. Leigh, T. Coffin, "Kites Flying In and Out of Space - Distributed Physically-based Art on the GRID," Journal of Future Generation Computer Systems (FGCS), Elsevier Science Press, to appear.

C. Zhang, J. Leigh, T. DeFanti, M. Mazzucco, R. Grossman, "TeraScope: Distributed Visual Data Mining of Terascale Data Sets Over Photonic Networks," Journal of Future Generation Computer Systems (FGCS), Elsevier Science Press, to appear.

Thomas A. DeFanti, Jason Leigh, Maxine D. Brown, Daniel J. Sandin, Oliver Yu, Chong Zhang, Rajvikram Singh, Eric He, Javid Alimohideen, Naveen K. Krishnaprasad, Robert Grossman, Marco Mazzucco, Larry Smarr, Mark Ellisman, Phil Papadopoulos, Andrew Chien, John Orcutt, "Teleimmersion and Visualization with the OptIPuter," Proceedings of the 12th International Conference on Artificial Reality and Telexistence (ICAT 2002), The University of Tokyo, Japan, December 3-6, 2002, to be published by Ohmsha/IOS Press, <www.ic-at.org>.

Jason Leigh, Andrew Johnson, Kyoung Park, Atul Nayak, Rajvikram Singh, Vikas Chowdhry, Thomas A. DeFanti, "Amplified Collaboration Environments," VizGrid Symposium, Tokyo, November 2002.

C.4. Software Releases

CAVERNsoft G2 is available at: <http://www.openchannelsoftware.org/projects/CAVERNsoft_G2>.

Quanta 0.1, released in August 2002, is available at <www.evl.uic.edu/cavern/quanta>.
Quanta 0.2 is scheduled for release in March 2003.

D. Collaboration Activities

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

Working with CERN on RBUDP 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 (October-December)

Available upon request.