Riding the light towards new science

LARRY SMARR

is at the Department of Computer Science and Engineering at the University of California, San Diego (UCSD), La Jolla, California 92093, USA and the Institute for Telecommunications and Information Technology at UCSD.
e-mail: lsmarr@soe.ucsd.edu

A new Internet is emerging. One in which dedicated optical circuits allow researchers to connect to computers all over the world and interactively work with massive datasets in real time. The technology is opening up new avenues in science, and this is just the beginning.

In the United States, broadband Internet to the home offers consumers a few megabits per second of bandwidth, yet gigabit-per-second Ethernet, with hundreds of times the bandwidth, is now the standard on personal and desktop computers. Therefore we have a 'broadband gap', which effectively throttles anything but nearby computing and storage devices from sending large amounts of data to each other.

The current shared Internet works well for the familiar applications of e-mail, instant messaging or web browsing, because the data objects involved are typically kilobytes to megabytes in size (thus matched to the available bandwidth), and they are delivered quickly enough for modest interactive visualization. However, frontier efforts of scientific research generate gigabytes or even terabytes of data — a thousand to a million times the size of large web objects — and it often takes many hours or days to transfer such information over the conventional Internet. This is far too long to support interactive use, so a new hybrid Internet has emerged. This Internet offers researchers optional access to dedicated 'light paths', 1 to 10-Gbit s^{-1} optical circuits that researchers can use to connect to remote computers and interactively access massive datasets on local, regional, national and international scales.

BUST AND BOOM

Until the late 1990s, such personal optical circuits would have been prohibitively expensive. But the massive investment of the 'dot-com boom' era drove a widespread buildout of fibre optics, and technology developed the ability to run 40 or more different wavelengths of light (lambdas) along a single fibre. When the post-2000 telecom 'bust' came, university researchers, cities and whole states snapped up access to these lambdas.

Illinois and Indiana were the first states to set up such networks around 2000. Now, over two dozen state and regional optical networks (RONs) have been established, many of which have come together to establish a national optical network called the National LambdaRail (NLR)[1]. The NLR provides up to 40 lambdas of 10 Gbit s^{-1} that connect various RONs. One of these lambdas, called CAVEwave, is intensively used as part of experiments discussed below. Another hybrid national network, known as Internet2[2], connects over two hundred research universities and has recently added an optical overlay network called NewNet. In 2001, research groups around the world established the Global Lambda Integrated Facility (GLIF) to facilitate the development of applications, middleware and large-scale distributed systems. GLIF is built from a number of lambdas contributed by its member participants, who own or lease them.

GET CONNECTED

One large-scale project using this cyber infrastructure is the National Science Foundation-funded 'OptIPuter' (www.optiputer.net). Led by the California Institute for Telecommunications and Information Technology (Calit2, a partnership of UCSD and the University of California, Irvine) and the University of Illinois at Chicago's Electronic Visualization Laboratory (EVL), OptIPuter ties together the efforts of researchers from over a dozen campuses. It uses lambdas to form end-to-end uncongested Internet protocol (IP) networks between computers, data-storage facilities and visualization resources resulting in a so-called LambdaGrid, that is, a grid consisting of both computational resources and networks. Just the same as using the World Wide Web on your PC, the illusion is that all the words, images and videos from around the planet are on your computer. But in this case, the data objects are orders of magnitude larger than those manipulated by the regular web.

TELESCIENCE

Parallel clusters of PCs are the appropriate termination devices for lambdas, because they can be scaled up in computing and visualization power, storage capacity and bandwidth to match the greater bandwidth offered by lambda connections. The OptIPuter project combines PC clusters configured by Rocks software[3], together with tiled display walls (Fig. 1) to create (OptIPortals), displays containing tens to hundreds of megapixels that are

OPTIPORTAL

Figure 1 An OptIPortal tiled display wall. Such lambda-connected displays enable researchers to view local high-resolution images at the same time as high-definition images from remote labs or computers in real time. Photo courtesy of David Lee, National Centre for Microscopy and Imaging Research (NCMIR) at UCSD.

connected to the GLIF through lambdas. Jason Leigh and his collaborators at EVL have developed a window-management system for OptIPortals, the Scalable Adaptive Graphics Environment (SAGE), which makes it possible to display several large-screen images at once. As Fig. 1 illustrates, researchers can stand at one of these OptIPortals and view high-definition video conferences, web browsers, large two-dimensional bitmaps or three-dimensional visualizations, connecting people to information in powerful ways.

DIVERSIFYING SCIENCE

In September 2005, the scientific potential of the GLIF was explored at the iGrid 2005 workshop[4] at Calit2 UCSD. The meeting attracted 450 attendees from 24 countries, with 49 real-time demonstrations from 20 of those countries, most using 1- or 10-Gbit s^{-1} lambdas. As became apparent, a number of innovative scientific applications are quickly moving to this new global optical fabric.

Very Long Baseline Interferometry (VLBI) is used by astronomers to take detailed images of distant radio-emitting objects in the universe. Traditionally, VLBI data, recorded using antennas scattered over the surface of the Earth, are simultaneously recorded to magnetic tapes and then physically shipped to a correlator that searches for common signals in the data. A VLBI team at iGrid 2005, led by the Massachusetts Institute of Technology (MIT), used dedicated optical networks to transfer data in real time from antennas in Massachusetts and Maryland (in the USA) and Sweden, at 0.5 Gbit s^{-1} per station, directly to the MIT Haystack correlator. Such information transfer allows scientists to view the correlated signals very soon after they are recorded and adjust their observations accordingly.

These emerging fast optical circuits can also advance science a little closer to home. The National Science Foundation's Ocean Research Interactive Observatory Networks (ORION) program is creating a new generation of ocean observatories, including some that use commercial submarine optical-fibre cables to connect remote users to a diverse set of scientific instruments located on the ocean floor. Using dedicated fibre technology, the University of Washington's LOOKING project[5], led by John Delaney, successfully generated a live high-definition video transmission that showed hydrothermal vent marine life on the seafloor at a depth of 2.5 kilometres. This streaming video was transmitted from the ship by satellite to Washington and then over CAVEwave lambda connections to the Calit2 centre. Such real-time observations will become routine over the next decade as more high-resolution underwater cameras are attached to optical fibres.

Also at iGrid 2005, Mark Ellisman's National Centre for Microscopy and Imaging Research[6] (NCMIR) at UCSD successfully conducted a collaborative microscopy experiment with one of the world's most powerful electron microscopes in Osaka, Japan. Using a dedicated lambda linking Osaka and San Diego, high-definition video images from both a remote microscope and a television camera in the microscope control room were brought together (creating a scene similar to that in Fig. 1). Telescience, such as this, enables remote scientific instruments to deliver their output directly into the lab using dedicated optical networks.

DATA MINING

Based on the successes of these and many other demonstrations, the first set of production facilities are being built for data-intensive science. Over 10,000 particle physicists worldwide will be connected to the Large Hadron Collider (LHC) particle accelerator in Geneva, Switzerland, which will turn on this year and become one of the leading drivers of advanced networking. According to Harvey Newman at Caltech, particle physicists will need 10-Gbit s^{-1} lightpaths this year, and within five years their bandwidth appetite will be 100 times that amount!

The Gordon and Betty Moore Foundation has funded Calit2 to create a cyber infrastructure for Advanced Marine Microbial Research and Analysis (CAMERA)[7]. CAMERA features lambda-connected computers and storage servers capable of holding very large amounts of environmental metagenomic data — starting with the global ocean survey carried out by Craig Venter. Metagenomics also requires a global view of data and the ability to zoom into detail interactively, so CAMERA is installing OptIPortal displays in a number of genomic laboratories across the United States, creating a national data-driven OptIPuter. CAMERA enables fundamental biological research, as well as applications in emerging areas such as genetically engineered microbes that generate biofuels.

THE BIG SCREEN

Commercial applications of these LambdaGrids are closer than one might think. The movie industry is rapidly emerging from a hundred years of using chemical media (film) into the 21st century in which everything is digital, and involves a massive flow of bits. Very-high-resolution digital technology offers a comparable visual experience to that obtained with conventional 35-mm film. A consortium of the major Hollywood studios, called the Digital Cinema Initiatives, have standardized on 2K images (which have 2,048 × 1,080 pixels) and 4K images (which have roughly 4,096 × 2,160 pixels). A 4K image is four times the resolution offered by high-definition television, and 24 times that of a standard television image.

With this in mind, Calit2 embarked upon a project called CineGrid, which aims to apply OptIPuter technology to the transmission, storage and image processing of high-resolution digital media. For example, 4K video displays an 8-megapixel image 24 times a second, requiring an uncompressed transfer rate of 6 Gbit s^{-1} and storage of terabytes per hour. As part of our iGrid2005 conference, we installed a Sony 4K projection system at Calit2 and connected it to compression/decompression computers through the GLIF. We then demonstrated a 4K trans-Pacific teleconference between Calit2 and the Research Institute for Digital Media and Content at Keio University in Tokyo (Fig. 2). Olympus 4K cameras transmitted images of the Keio University President discussing the future of digital cinema with the Chancellor of UCSD. In 2006, CineGrid further engaged researchers in Hollywood and on the West Coast of the United States to conduct several successful field trials[8].

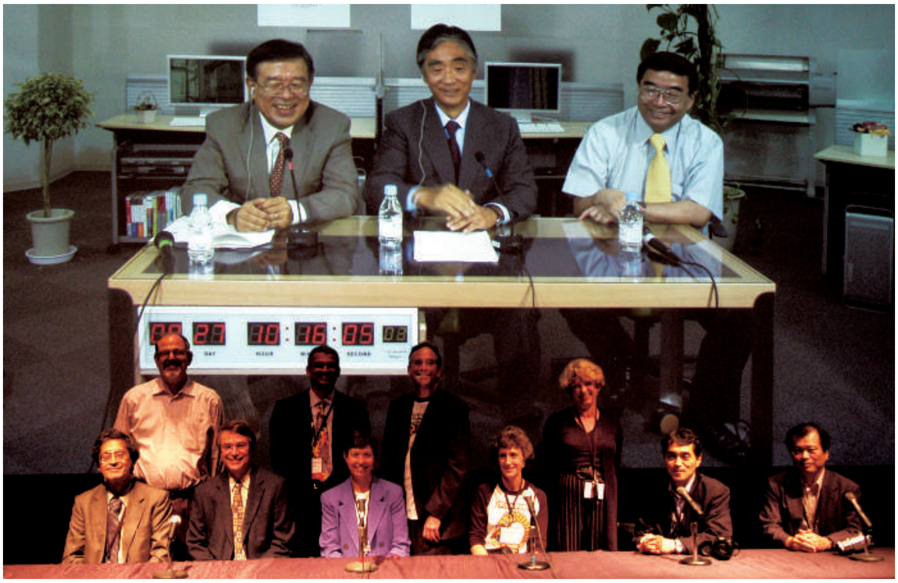

Figure 2 Global video conferencing by means of an OptIPortal. Photo courtesy of Richard Edlund.

A PLANETARY NETWORK

It is clear that an Internet-technology leap has occurred over the past decade, bringing dedicated light paths over optical fibres. As we are already seeing, this technology is enabling a new generation of scientific research and commercial applications. The global LambdaGrid testbed is now an ideal place for public–private partnerships to explore innovative optical equipment that could enhance the capabilities of this developing planetary infrastructure.

References

1. http://www.nlr.net
2. http://www.internet2.edu
3. http://www.rockclusters.org
4. http://www.igrid2005.org
5. http://www.lookingtosea.ucsd.edu
6. http://www.ncmir.ucsd.edu
7. http://www.camera.calit2.net
8. http://www.cinegrid.org