Dynamic Provisioning of LightPath Services for Radio Astronomy Applications

Jerry Sobieski (Mid-Atlantic Crossroads), Tom Lehman (University of Southern California / Information Sciences Institute), Chester Ruszczyk (MIT Haystack Observatory), Rick Summerhill (Internet2), Alan Whitney (MIT Haystack Observatory)

Abstract

A demonstration at iGRID 2005 used dynamic, deterministic, and dedicated LightPath network services to link radio telescopes from around the world with computational facilities at the MIT Haystack Observatory to create a single coherent instrument for real-time astronomical and geodetic research. The "electronic Very Long Baseline Interferometry" (e-VLBI) application provides ultra-high resolution images of very faint and very distant objects in the universe. The application specific network topology carried 2Gbps of VLBI data from radio telescopes in Europe, North America, and Japan to Haystack for real-time correlation processing. This paper describes the application, the network technologies employed for the demonstration, the results, challenges and future work.

1. Introduction

A new paradigm for connection-oriented services has emerged within the advanced networking community in recent years. The term "LightPath", which was originally coined to describe connection-oriented services anticipated for photonic networks, has entered the vernacular to describe a much broader set of network services where ***dedicated*** and ***deterministic*** network resources are provisioned between globally distributed application end points. At the same time within the broader science community there is a set of advanced "eScience" applications emerging which seek to seamlessly integrate globally distributed science teams and facilities. These new eScience applications require network service capabilities and performance beyond that which is generally available to them via existing infrastructures. Development of high-performance networks which can rapidly provision LightPaths in direct support of eScience applications is an anticipated solution for these network power users. This paper reports on the results of such an integration of a large eScience application and dedicated LightPath services provisioned across multiple domains on a global basis.

In this demonstration at iGRID 2005, an international team of network researchers, engineers, and scientists have collaborated to demonstrate how a fundamental radio astronomy technique called Very Long Baseline Interferometry (VLBI) is able to benefit from LightPath services. Radio telescopes in the United States, Great Britain, Sweden, the Netherlands, and Japan were linked to computational facilities at the MIT Haystack Observatory in Westford, Massachusetts using LightPaths. These LightPaths allowed each station to send 512Mbps of astronomical VLBI data over high-speed networks (hence 'e-VLBI') to Haystack for real time correlation. While the dedicated and deterministic performance of these LightPaths enables the correlation process to proceed in real-time, it is the automated and dynamic flexibility used to establish the LightPaths that has the potential to substantially change the way VLBI is done. These capabilities will provide a number of benefits to the VLBI community, such as rapid feedback from results to help optimize observation strategies, as well as the future promise of multi-Gbps data rates which will directly improve the sensitivity of the observations without building larger telescopes.

Specifically, the demonstration plan was to link the Haystack correlator with up to four radio telescopes to create a single real-time application spanning three continents. The participating telescopes were 1) Westford Observatory (1 km from Haystack), 2) the Goddard Geophysical and Atmospheric Observatory (GGAO – 800 km away in Greenbelt, MD), 3) the Onsala Space Observatory (southwestern Sweden), 4) the Jodrell Bank Observatory (outside Manchester, UK), 5) the Westerbork Observatory (in northeast Netherlands), and 6) the Kashima Space Observatory (north of Tokyo, Japan). Ultimately, the team successfully linked three sites together with actual correlations running in real-time: Westford and GGAO in the United States, and Onsala Space Observatory in Sweden, each providing 512 Mbps to the correlator.

This was the first time such real-time correlation had been done across such distances and at such data rates. This demonstration involved significant development and testing of applications code, and involved coordinating station availability among ongoing science experiments, appropriate sensor characteristics, continual network verification testing of global LightPaths, and operational debugging and support. All of these six facilities contributed extensively to the success of the demonstration.

This demonstration at iGRID 2005 has shown that such real-time e-VLBI correlation can be effectively applied on a global basis and in a highly dynamic and flexible fashion. Emerging advanced network services can be established dynamically, carry tens of Gbps of traffic, and have a global reach. This has opened the door to more effective use of expensive and scarce science instruments. Real-time e-VLBI also encourages the development of new application architectures for e-VLBI computational resources and enables the opportunity to set up, configure and receive immediate feedback on observations using very expensive and scarce equipment.

This paper is organized as follows. Section 1 is this introduction. Section 2 describes the radio astronomy application, what it is and how it works, some of its history and the nature of the network requirements. Section 3describes LightPath services and the DRAGON project GMPLS protocol capabilities. Section 3 addresses the network architecture for the iGRID demo and the results of the effort. Section 4 details the future directions for the network research and further deployment, and the future of e-VLBI technologies within astronomy community and advanced networks.

2. The Application: Radio Astronomy and e-VLBI

Very-Long-Baseline Interferometry (VLBI) has been used by radio astronomers for more than 30 years as one of the most powerful techniques for studying objects in the universe at ultra-high resolution and for measuring earth's motion with ultra-high precision. VLBI combines simultaneously acquired data from a global array of up to ~20 radio telescopes to create a single coherent instrument, as illustrated in Figure 1 for a simple 2-element VLBI array. In brief, the VLBI process involves pointing several very sensitive radio telescope antennae at a single object in deep space. Each of these stations simultaneously record the radio frequency noise received at the station. The vector defined by the geographic separation between pairs of receiving stations is called a "baseline". By correlating the data streams received at the stations, scientists are able to study in exquisite detail very distant and very faint objects in the universe and to measure the motions of the Earth in space with exquisite accuracy. By nature, VLBI data are digital representations of the analog signal arriving at a radio telescope. Almost always this appears as white Gaussian noise with the implication that each sample is independent and the data are fundamentally uncompressible**.**

2.1 The VLBI Process

The VLBI process is tied to a reference frame defined by very distant radio objects. These distant objects have been studied in great detail over the years, and have been positioned very accurately in the 3-space of the cosmos. It is because of this well defined and highly precise reference frame that VLBI is the only technique which can fundamentally measure the orientation of the Earth in inertial space. For Earth science studies, VLBI provides direct measurement of the vector between globally-separated telescopes with an accuracy of a few millimeters and, by measuring the changes in baseline length over time, yields motions of the tectonic plates with a precision of ~0.1 mm/yr. Paradoxically, VLBI's value in understanding the detailed motions of the Earth in an inertial reference frame make it currently the only way to study the shape of the Earth's liquid iron core and to determine the magnetic connections of the liquid core to the mantle above it and to the solid core at the center of the Earth. VLBI's usefulness for astronomical studies is no less striking: VLBI allows images of distant radio sources to be made with resolutions as high as tens of *micro*arcseconds, much better than any optical telescope. (This is roughly equivalent to discerning individual dimples on a golf ball in San Francisco from New York City!)

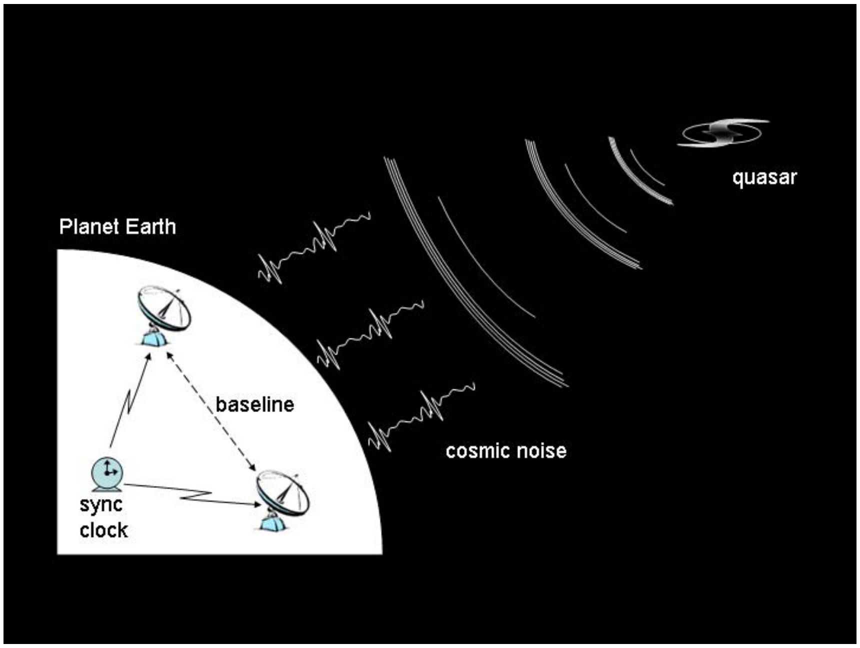

Figure 1 Very Long Baseline Interferometry

Traditionally, VLBI data streams are recorded on magnetic tapes or disks that are then shipped to a central site for correlation processing. With the emergence of high-performance networks and infrastructure, this laborious and expensive data-collection and transport process now has the possibility of being superseded by modern integrated sensor-computation models, data transfer mechanisms, potentially opening important new capabilities and scientific possibilities.

The transmission of VLBI data via high-speed networks is dubbed 'electronic-VLBI' (e-VLBI), the development of which is now underway worldwide. e-VLBI can be exercised in either of two ways: 1) real-time transfer via a direct connection from telescope to correlator or 2) quasi-real-time transmission by first buffering the data through a buffer memory (semiconductor or magnetic media) capable of storing seconds to hours of data before transmission. The latter is generally preferred for a host of practical reasons, but both are workable and both provide many of the same advantages.

The buffering of data either at the acquisition station or at the correlator is the responsibility of the Mark 5 data system. The Mark 5 is a specialized PC-based data-acquisition system that collects the digital data directly from the antenna and either stores it to disk or encapsulates it into IP packets and forwards it toward the correlator. The Mark 5 system has been adopted by the international VLBI community as the standard acquisition system for nearly every VLBI telescope worldwide (Japan being a notable exception which uses a slightly different capture format called K5.) At the correlator end, another Mark 5 receives the data from the network and writes it to disk for later correlation, or feeds the data stream directly into the correlator interface units. One can conceptualize the e-VLBI application as a star topology, with the correlator in the center, and each remote station streaming data to the correlator. Due to the size of the electronic buffering available at the correlator, the data streams from all stations must arrive within about a 2-3 second window.

The correlators themselves are hardware optimized digital signal processors designed specifically for the VLBI application. The algorithms used to process the VLBI data can be highly optimized in application specific integrated circuits. While these systems are extremely efficient and remain the fastest means of processing current VLBI data, they are expensive and fall quickly behind state-of-the-art. New architectures and methods are being investigated for the next generation of correlators that will be more scalable and able to keep up with the rapidly increasing sampling rates/station and greater numbers of observations anticipated over the next 5-10 years.

The advantages for scientific productivity and technical operations of e-VLBI over traditional VLBI are several:
Faster turnaround of results: Typical time to completion of processing of traditional VLBI usually measured in at least days and sometimes several weeks, largely limited by shipping of physical media. This makes it almost impossible to use 'fast-response' observations to guide quick follow-on observations.
Higher sensitivity**:** The sensitivity of most VLBI observations increases as the square root of the data rate. The potential to extend e-VLBI to multi-Gbps data rates in the future will allow an increase in the sensitivity of observations beyond those possible with traditional recorded media, and without having to build larger or more sensitive antennas.
Lower costs**:** e-VLBI will reduce and potentially eliminate the need for expensive tape or disk pools while at the same time allowing full automation of observations, all towards the goal of lowering cost.
Quick diagnostics and tests**:** Some aspects of VLBI equipment setup are very difficult to test and diagnose without actually correlating data from another station. With e-VLBI, the proper operation of the equipment at each station can be determined at the time of the experiment.

2.2 History of e-VLBI

Electronic transmission of VLBI data from antenna to correlator has been an obvious but difficult goal of VLBI practitioners since the origin of the VLBI technique in the 1960's. A pioneering experiment in 1977 linked the signals from two antennas over a real-time satellite link [12]. Transmission of small amounts of data (~1 Mb/station) over ordinary telephone lines to a software correlator was successfully accomplished at Haystack in 1979 [16]. Since that time, the dominant activity in e-VLBI has been in Japan, with the Keystone project in ~1995 [13] linking four antennas in real-time at 256 Mbps and, more recently, dedicated Gbps networks. Furthermore, high-speed radio links have been used to transmit data from orbiting antennas, notably the TDRSS satellite, in 1986 [14] and the Japanese dedicated orbiting VLBI satellite HALCA in 1997[15]; both of these satellites transmitted data to the ground for recording, but did not transmit data in real-time or near-real-time for correlation.

More recent developments in e-VLBI have centered around the use of rapidly expanding global high-speed data networks. In 2001, the MIT Haystack Observatory developed the Mark 5 VLBI data system which allows VLBI data to be flexibly directed to/from either magnetic disks or high-speed networks at rates up to 1Gbps. With DARPA support, and with the cooperation of MIT Lincoln Laboratory and NASA Goddard Space Flight Center, the first modern high-speed e-VLBI experiment was carried out in October 2002 using standard IP networks using antennas in Westford, MA and Greenbelt, MD, with a data transfer speed of 760Mbps to a Mark 4 VLBI correlator at MIT Haystack Observatory in Westford, MA.

E-VLBI has expanded to an international scale, with e-VLBI data transfers between U.S., Japan and Europe, but primarily in non-real-time (i.e. disk buffers at both the stations and at the correlator). A number of experimental full real-time experiments (i.e. no disk buffering at either station or correlator) have been carried out using stations in the U.S. and Europe, but these have been mostly limited to two stations and at data rates less than 512 Mbps. The work described in this paper has led to the first intercontinental real-time e-VLBI at rates as high as 512 Mbps.

e-VLBI development work is now taking place in the U.S., Europe, Japan and Australia. One of the most difficult issues is the physical connection of stations to the global network, since many stations were placed in deliberately remote locations where RF interference from surrounding human activity is minimized. In recent years, there has been good progress in connecting these telescope facilities into the net with advanced high speed telecommunications infrastructure, but many sites remain network challenged. As VLBI data rates continue to climb, (4+ Gbps are on the horizon) pressure will increase to complete the fiber infrastructure to all VLBI telescope sites.

3 The Demonstration: Real-time, global e-VLBI

The objectives of the “Dynamic Resource Allocation in support of Real-time global Radio Astronomy” demonstration at iGRID 2005 was to

- Show how dedicated network resources can benefit globally distributed e-science applications;
- Demonstrate dynamic provisioning of LightPath services using Generalized Multi-Protocol Label Switching (GMPLS) control plane,
- Enable advancements in real-time global e-VLBI correlation
- Debut the HOPI testbed and show how it can drive advances in both networking and applications.

It should be noted that while this paper describes the activities and results from the iGRID 2005 demonstration, the infrastructure assembled and application capabilities developed are part of an ongoing project and will remain in place. An important objective of these activities is to drive new capabilities into existence which can be utilized on a continuing basis to be evolved over time into robust, reliable, and ubiquitous network services and e-science resources.

Toward these goals, we worked with key e-VLBI scientists and applications developers to identify and incorporate colleagues at a number of leading radio-astronomy facilities into an effort to perform multi-continent real-time VLBI correlation at as high a sampling rate as possible. This included e-VLBI experts from the MIT Haystack Observatory (US), the Goddard Geophysical and Astronomical Observatory (US), Jodrell Bank Observatory (UK), Onsala Space Observatory (SE), and the Joint Institute for VLBI in Europe (JIVE) (NL), and the Kashima Research Space Center (JP).

The network project team from the Mid-Atlantic Crossroads in Washington DC likewise incorporated their international networking colleagues to define and establish the necessary network resources to enable this demonstration. A high-level overview of the topology is shown in the Figure 2. The following network organizations were involved: UKLight (United Kingdom), NetherLight (Netherlands), NorthernLight (Scandinavian countries), SUnet (Sweden), JGN2 (Japan), HOPI (United States), and DRAGON (United States).

The demonstration plan was formulated four months prior to the conference. It called for dynamically establishing LightPaths from each of four telescopes around the world to a correlator at the Haystack Observatory outside Boston. The data stream sourced from each telescope was 512Mbps average sustained data rate with bursts up to 700 Mbps, so we planned to provision 1Gbps Ethernet-framed connections between telescopes and correlator. Further, we proposed to distribute GMPLS-capable switches throughout the various transit networks to instantiate the VLBI topology dynamically. We selected six telescope facilities based upon a) known high speed network connectivity, b) availability of the instruments for testing and during the live iGRID demonstration, and c) availability of resources in the transit networks back to Haystack. As is normal, The Plan required modification along the way. Details of the network engineering and final configuration, and the technologies employed are provided in the following section.

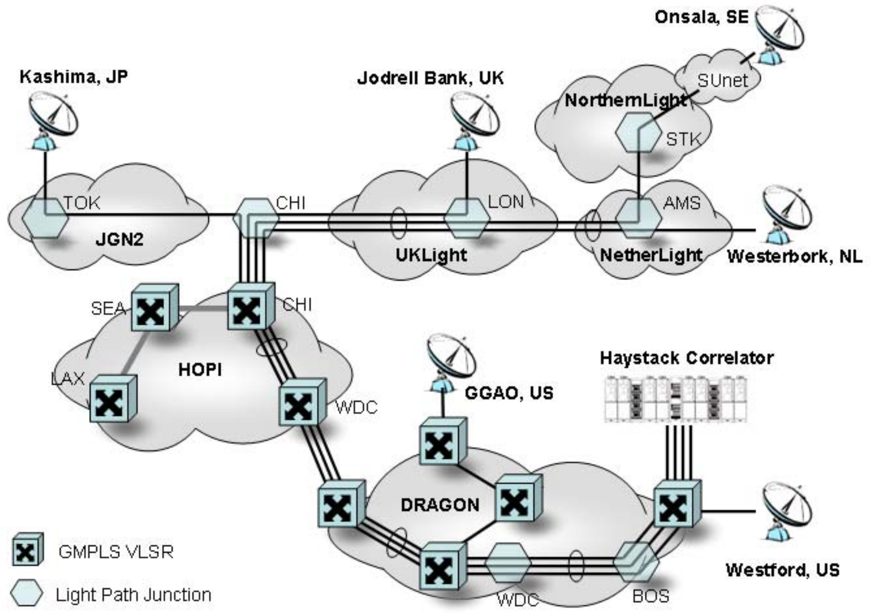

Figure 2 iGRID 2005 real-time e-VLBI global network topology

3.1 Dynamic LightPath Services: The Generalized Multi-Protocol Label Switching Control Plane

The term “LightPath” was coined to refer to the path a specific optical-wavelength channel takes thru a photonic network. These wavelength channels exhibited some very desirable characteristics: First, they are dedicated channels accessible only to the devices attached to the end points, i.e not shared with other users. Second, they offer very deterministic performance, i.e. one can get consistent and repeatable performance across the wavelength channel every time it is used. The term “LightPath” has evolved to include services across other network technologies that exhibit similar characteristics. For instance, TDM circuits provisioned across Sonet/SDH infrastructure also provide these characteristics, as do certain types of Ethernet connections.

Creating the base telecommunications infrastructure needed for global LightPath services is an ambitious, expensive, and international effort. These network resources are themselves relatively scarce and must be shared among similar applications worldwide. Provisioning these resources to suit one application and then another is a painstaking engineering and operational task requiring weeks or months to assemble all of the constituent components across multiple administrative and operational service domains into a transparent high-performance networking environment adapted to a specific application.

In order for these LightPaths to become a broadly available service option, an automated means of establishing these LightPaths is necessary. One proposed means to accomplish this is to employ the Generalized Multi-Protocol Label Swapping (GMPLS) [1] protocols to perform the routing and signaling functions required. Specifically, OSPF-TE and RSVP-TE, protocols used for MPLS based traffic engineering in the IP layer, have been extended and standardized to support a “generalized” hierarchy of switching capabilities including packets (IP layer), Layer 2 (e.g. Ethernet), TDM (Sonet/SDH), wavelength, and fiber.

The NSF-funded DRAGON Project [8, 11], a photonic test bed situated in the Washington DC region, has developed open source and standards-based implementations of the GMPLS protocols. The DRAGON protocol suite runs on a Unix “control PC” and manages a network switching element – say, an Ethernet switch. This control PC participates in the routing and signaling protocols of the dynamic LightPath network, translating protocol events into [re-]configuration commands for the covered switching device. (See Figure 3). This coupled relationship between the Ethernet switch and a GMPLS control PC has acquired the moniker of a “Virtual Label Switching Router” or VLSR.

In addition to the OSPF and RSVP protocols, DRAGON has developed an inter-domain service routing agent called the “Network Aware Resource Broker” (NARB). The NARB listens to OSPF for internal link state updates and summarize this topology information for dissemination to other “peering” LightPath networks. The NARB is responsible for inter-domain path computation and provides hooks to interface with authorization and advanced scheduling tools (future work). These three components, OSPF, RSVP, and NARB, make up the DRAGON control plane.

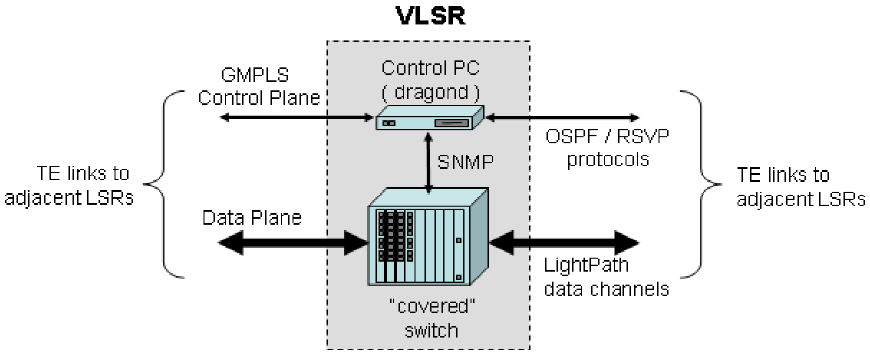

Figure 3 GMPLS "Virtual LSR" and the NARB

This DRAGON control plane is deployed within the DRAGON network – which exists mostly in the Washington DC region but extends all the way to MIT Haystack Observatory facility, and it has also been deployed within the Internet2 HOPI testbed extending across the continental United States. Two VLSRs were deployed internationally, one to the University of Manchester, and the other to the KTHNOC (SUnet) in Stockholm. Unfortunately, these two VLSR were not brought into service in time for iGRID. The iGRID demonstration used the VLSRs within the HOPI and DRAGON networks to setup the VLBI LightPaths between the StarLight facility in Chicago and the Haystack facility outside Boston.

3.2 The Network Engineering: Marshalling network resources across the globe

Besides the control plane, the demonstration required substantial data-plane engineering. The following networks provided substantial resources for the cause:

In addition to the dynamic control plane technology, DRAGON provided key regional network connectivity to support the e-VLBI science and this iGRID demonstration. The DRAGON network is a photonic regional testbed designed to switch ITU wavelengths among approximately ten sites in the Washington DC metro area. DRAGON works closely with the BOSSNet testbed – a DARPA fiber test bed extending between Washington DC and Boston along the US east coast. The MIT Haystack facility in Westford MA, is part of the DRAGON Project and is accessed via the DRAGON network and a wavelength channel across BOSSNet.

The Haystack Observatory hosts the correlator used in this demonstration as well as the nearby Westford Telescope. The Haystack facility connects into DRAGON via an 815 km dedicated wavelength channel currently configured as OC48 Packet over Sonet link supporting all the high speed network access for the e-VLBI work. This IP link is subdivided into three QoS enabled MPLS LSPs of 700 Mbs each. Each LSP is circuit cross-connected at each end through a Juniper Networks router to a dedicated 1Gbps Ethernet interface. While the LSPs were manually configured and static, the GMPLS control plane was used to dynamically map the LSPs to appropriate LightPaths as needed.

The VLBI telescope at the GGAO facility at NASA Goddard Space Flight Center in Greenbelt, MD is also directly connected into the DRAGON infrastructure via 10Gbps Ethernet. GMPLS VLSRs were used to establish the Ethernet channel between NASA Goddard Space Flight Center and the Haystack correlator.

Another key network component of this iGRID demonstration is the Internet2 “Hybrid Optical / Packet Infrastructure” (HOPI) test bed [10]. The HOPI testbed consists of four switching nodes: Washington DC, Chicago, Seattle, and Los Angeles, connected in a linear topology via 10Gbs Ethernet. At each of these POPs, HOPI has deployed a Force10 E600 high performance ethernet switch, a “control PC”, a “performance PC”, and other devices to support operations and future experiments. The HOPI network has deployed a GMPLS control plane using the DRAGON protocol suite. The GMPLS protocols run on the control PC at each of the HOPI nodes and manage the local E600 switch. This capability allows HOPI to dynamically establish Ethernet LightPaths across the US at capacity up to 10Gbs. The performance PCs at each HOPI node have 10Gbps network interfaces cards and are used for performance testing and to assist in fault isolation on the lower-layer LightPaths.

HOPI “peers” with DRAGON in Washington DC. This means the two networks have established physical connectivity, and their respective NARB agents exchange topology and service routing information. RSVP is used to signal across the inter-carrier interface.

HOPI has established numerous cross-connects between the HOPI Chicago node and other networks at the StarLight facility in Chicago. The StarLight facility is a nexus for many international networks; of particular note are UKLight, NetherLight, and JGN2 networks. For iGRID, three 1GE links were established across the StarLight switch between UKLight and HOPI. Another single 1GE link was installed directly between HOPI and JGN2. The international e-VLBI links were provisioned manually and appear to the dynamic HOPI control plane as unusually long static client links (deploying VLSRs internationally took longer than expected due to shipping and customs requirements, so we elected to set up these international links manually for this demonstration.)

The HOPI testbed is a new facility designed to explore next-generation internet architectures – most notably the integration of conventional IP packet networks with emerging technologies such as LightPaths. To this end, HOPI has direct connections to Abilene to both extend the LightPaths into the packet infrastructure as conventional MPLS LSPs, and/or to allow the packet infrastructure to leverage HOPI’s lower-layer circuit capability to develop effective integrated service models. In this iGRID 2005 demonstration, the e-VLBI traffic utilized TCP and UDP transport protocols to transport the data even with the dedicated links. Further, the real-time graphical results of the correlation - a “fringe plot” - were sent via best effort IP to an X-Windows display at the iGRID conference venue in San Diego, CA. The iGRID 2005 conference served as the debut of the HOPI testbed facility.

The dynamics of the HOPI control plane were utilized extensively leading up to the demonstration. The ability to rapidly reconfigure LightPaths between performance verification hosts along the paths and the Mark 5 systems at the end points proved to be a powerful asset.

The UKLight network provided substantial engineering support and network resources to extend LightPaths from Chicago thru London UKLight operates an OC192 SDH circuit from Chicago to London, and another from London to Amsterdam. UKLight also has similar capacity extending from London to the University of Manchester. Three VC-3-13c (~700 Mbps) SDH circuits were manually provisioned: one from a VLSR installed at Manchester University thru London to Chicago that serves the Jodrell Bank Observatory, and two from Amsterdam thru London to Chicago that serve the Westerbork and Onsala facilities(See Figure 2).

The NetherLight network and the Amsterdam open exchange (operated by SURFnet) also played an important role in this iGRID demonstration. NetherLight operates an OC192 SDH circuit between Amsterdam and Chicago, and made this available for the iGRID demonstration as well. The Amsterdam exchange point provided cross-connects between the Westerbork 1GE fiber path and the UKLight circuit to Chicago. Amsterdam also enabled the cross-connect between NorthernLight (serving Onsala) and UKLight circuit to Chicago. The NetherLight capacity connecting directly to Chicago was in use for other demonstrations and so was considered a backup path for e-VLBI purposes.

The Westerbork Observatory is northeastern NL, ~20km from the offices of the Joint Institute for VLBI in Europe (JIVE). Dark fiber links carrying 1Gbps Ethernet normally link Westerbork to the correlator at JIVE. For the iGRID demonstration, a 1GE link was established between Westerbork (via JIVE) to the Amsterdam exchange. This LightPath was terminated on a production VLAN at JIVE which created at least one surprise for the demonstration team (details below).

In order to connect the Onsala Space Observatory in Sweden into the global real-time application, the NorthernLight network and the Swedish Universities networks (SUnet) were utilized. NorthernLight is operated by NorduNet, a consortium of R&E networks within the Scandinavian countries of Europe. A 1Gbps Ethernet-framed TDM channel was established over NorthernLight between Stockholm and Amsterdam. The Onsala facility does not have direct LightPath capabilities to Stockholm (yet), but does have 1GE access to the SUnet routed IP network, which proved to have ample capacity and performance for our purposes. So the global LightPath serving Onsala was terminated in Stockholm on a SUnet router. Like the other LightPaths established in support of the iGRID demo, the NorthernLight channel carried GFP encapsulated Ethernet frames from Stockholm to Amsterdam. In Amsterdam, the channel was switched at the Ethernet layer thru the Amsterdam exchange to be re-encapsulated onto a UKLight channel to Chicago.

The Kashima Observatory was connected into the JGN2 network at 1 GE. A LightPath was provisioned at the Ethernet layer thru JGN2 and emerged in Chicago. The trans-Pacific OC192 this link used was shared with several other more voracious applications for iGRID. Due to some network equipment limitations close to the radio telescope, an MTU size larger than 1500 bytes could not be used. Some minor packet loss was observed during testing that we suspect had to do with the shared traffic. The Kashima observatory also uses a different data-acquisition system which requires on-the-fly data-format translation to the Mark 5 data format. TCP performance across this longest link was also an issue. Even minor packet drops degraded end-to-end throughput well below the 512Mbps desired. All of these issues took longer to resolve than expected and we were unable to get Kashima station into the correlation in time for the iGRID demonstration. (Note: all of these issues have since been resolved.)

4. Challenges and Futures

Creating a global real-time application such as e-VLBI poses many challenges.

The first challenge is the simple act of scheduling the facilities for testing. The Mark 5 systems are an integral part of the VLBI program and are often unavailable for testing. In many instances the engineering team was able to use iperf servers along the path(s) to verify network performance, but this in itself is not sufficient to verify end-to-end functionality of the application. The individual Mark 5 systems are often several years old and so do not incorporate the latest processor technologies. Hardware and software compatibilities as well as limited development support conspire to limit how quickly new releases of OS (and network protocols and drivers) can be updated. Thus, tuning the application requires insuring that the actual hosts that will be driving the streams are part of the testing procedure. In addition, even when a Mark 5 host is available for testing, the telescope itself may be applied to other observations. Valid prerecorded and synchronized data streams must be used in order to verify correct end-to-end operation, particularly when translating other recording formats on the fly.

Second, the e-VLBI application was developed to utilize conventional transport protocols over IP networks. This is a reasonable approach, since most transfers occur over shared commodity networks using bulk file transfers, but does not work so well for real-time e-VLBI correlations. Standard TCP has well documented issue over high-latency, high-capacity links [3]. Even very minor packet loss significantly impacts the end-to-end TCP performance. Such packet loss can occur due to occasional link errors corrupting a packet or where dedicated network resources are not suitably allocated and provisioned for each LightPath. Reliable protocols [2,4,5,6,7,9] designed to maintain throughput in the face of minor loss are being explored for integration into the real-time e-VLBI process.

Another application issue that is present in real time e-VLBI is proper stream synchronization. In conventional VLBI where the data streams are originating from recorded media local to the correlator, synchronizing the streams is a matter of local management. However, in real-time e-VLBI processing, buffers must be managed to account for the differing latencies between the data streams arriving from different stations. Further, packet-loss recovery schemes that retransmit lost data require substantially larger buffering still. Ironically, since the data are essentially noise, some lost data can be tolerated during correlation However, lost data must still be detected in order to maintain the pipeline synchronization between streams within the correlator.

Since the LightPath paradigm is in its infancy, many of the transport services incorporated in the end-to-end connections have minor but important differences in service characteristics. For instance, some links along the Ethernet paths provisioned for the iGRID demonstration did not support large MTUs. This particular issue can make substantial difference in overall performance of the Mark 5s. MTU size varied for the demo from 1500 bytes to 9000 bytes on different links. In other cases, the provisioned path, while dedicated, did not take into account typical burstiness of the data stream. Overall throughput was 512 Mbps, but the TCP streams could burst to as much as 700 Mbps. The burstiness of the TCP stream was a combination of application performance and links in the path that transited shared best-effort network links.
.
Another issue arose from interactions with production services. Collaborating research activities at Jodrell Bank and Westerbork had previously established a common VLAN that served both sites. The iGRID LightPaths from each of these two sites to the Haystack correlator had inadvertently been joined at the HOPI border in Chicago during testing. This caused a broadcast storm of ARP requests which had a “negative impact” on production traffic at gateway router in Amsterdam. The spanning tree had been disabled on the iGRID LightPaths in order to speed the provisioning process.

Performance verification of the network links proved to be a challenge. Traditional methods and tools used within the IP network are not well suited (or do not work at all) when the data plane path does not cross IP routers. Traceroute does not expose the lower layer path and so end-to-end connectivity must be localized using knowledge of the intended path and a binary-search technique. This involves breaking the end-to-end path to insert iperf servers at test points. Since the Ethernet LightPaths provisioned for this demonstration utilized VLAN port mapping at each switch to establish the path, inserting test points was possible, albeit a manual process. When testing several paths, a complex set of connections were created which sometimes resulted in “negative impacts”. The techniques used for an Ethernet VLAN environment may not be suitable for TDM or wavelength services. This performance verification issue is a major problem and begs for a more formalized approach to define and implement protocols and features that can simplify and automate the detection and localization of service faults.

An area for future collaboration will be extending the GMPLS control plane into other networks. We have already begun discussions and plans to establish a persistent TDM circuit topology that can be used for further testing of dynamic provisioning. We plan to deploy GMPLS VLSRs at key points in the iGRID e-VLBI topology that will enable experimenting with flexible and dynamic setup of paths between various telescopes and correlators.

At the end of the day, the iGRID e-VLBI team was successful in doing real-time correlation across two continents and three telescopes at 512 Mbps/station. We successfully did real-time correlations between Westford, Goddard, and Onsala. The Jodrell Bank antenna was configured to listen to a different radio-frequency band and was unable to participate in the iGRID correlation. We had hoped to bring in Westerbork and Kashima, but we were unable to complete network testing and debugging with enough time left for application tuning. The Kashima integration proved most challenging in dealing with format translation and high latency. But even these issues will be resolved very soon.

This demo was an international effort in which all the participating sites contributed significantly. The cooperation and collaboration exhibited by all of the personnel portends well for follow on work to complete the types of application engineering and network service provisioning that is necessary to make real-time e-VLBI a reliable and robust capability over which we can extend the science capabilities of VLBI.

The authors would like to recognize the following individuals and organizations who share in the success of this exciting and continuing effort: Chris Tracy (MAX/HOPI), Xi Yang (USC/ISI East), Ralph Spencer (Jodrell Bank Observatory), Richard Hughes-Jones (University of Manchester), Arpad Szomoru (JIVE), Lars Fisher (NorduNet), John Graham (UKLight), Rudiger Haas (Onsala Space Observatory), Bengt Gorden (SUnet), Tony Foley (Westerbork Observatory), Steve Thorpe (NCREN/HOPI), Kevin Dudevoir , Mike Titus and Jason Soohoo (MIT Haystack).

References

[1] L. Berger (editor), “Generalized Multiprotocol Label Switching”, IEEE RFC 3471, Jan 2003

[2] Y. Gu and R. Grossman, "UDT (UDP based Data Transfer Protocol): An Application Level Transport Protocol for Grid Computing," in Second International Workshop on Protocols for Fast Long-Distance Networks, Argonne, IL, 2004. Available: http://dsd.lbl.gov/DIDC/PFLDnet2004/papers/Grossman.pdf

[3] M. Mathis, J. Semske, J. Mahdavi, and T. Ott, "The macroscopic behavior of the TCP congestion avoidance algorithm," Computer Communication Review, volume 27, number 3, July 1997. Available: http://www.psc.edu/networking/papers/model_ccr97.ps

[4] C. Jin, D. Wei, S. H. Low, G. Buhrmaster, J. Bunn, D. H. Choe, R. L. A. Cottrell, J. C. Doyle, H. Newman, F. Paganini, S. Ravot, and S. Singh, "FAST Kernel: Background theory and experimental results," presented at the First International Workshop on Protocols for Fast Long-Distance Networks (PFLDnet 2003), February 3-4, 2003, CERN, Geneva, Switzerland. Available: http://netlab.caltech.edu/pub/papers/pfldnet.pdf

[5] C. Jin, D. Wei, S. H. Low, G. Buhrmaster, J. Bunn, D. H. Choe, R. L. A. Cottrell, J. C. Doyle, W. Feng, O. Martin, H. Newman, F. Paganini, S. Ravot, and S. Singh, "FAST TCP: From Theory to Experiments," IEEE Network Magazine, Volume 19, Issue 1, January/February 2005. Available: http://ieeexplore.ieee.org/iel5/65/30145/01383434.pdf

[6] J. Wang, D. X. Wei, and S. H. Low, "Modeling and Stability of FAST TCP," in Proceedings of IEEE INFOCOM, March 2005. Available: http://www.cds.caltech.edu/~jiantao/doc/infocom05.pdf

[7] L. Xu, K. Harfoush, and I. Rhee, "Binary Increase Congestion Control for Fast Long Distance Networks," in Proceedings of IEEE INFOCOM, March 2004. Available: http://www.ieee-infocom.org/2004/Papers/52_4.PDF

[8] X. Yang, T. Lehman, C. Tracy, J. Sobieski, P. Torab, S. Gong, B. Jabbari, "Policy-Based Resource Management and Service Provisioning in GMPLS Networks", submitted for publication, July 2005. Available: http://dragon.maxgigapop.net/papers

[9] Y. Zhang and T. Henderson, "An Implementation and Experimental Study of the eXplicit Control Protocol (XCP)," in Proceedings of IEEE INFOCOM, March 2005. Available: http://bat710.univ-lyon1.fr/~cpham/TCP/extended.pdf

[10] "The Hybrid Optical and Packet Infrastructure (HOPI) Testbed", http://networks.internet2.edu/hopi

[11] T. Lehman, J. Sobieski, B. Jabbari, "DRAGON: A Framework for Service Provisioning in Heterogeneous Grid networks", in IEEE Communications Magazine, Vol 44 No. 3, March 2006, Available thru http://www.comsoc.org/livepubs/ci1/index.html

[12] Yen, J. L.; Kellermann, K. I.; Rayhrer, B.; Broten, N. W.; Fort, D. N.; Knowles, S. H.; Waltman, W. B.; Swenson, G. W., Jr., "Real-time, very-long-baseline interferometry based on the use of a communications satellite", Science, vol. 198, Oct. 21, 1977, p. 289-291.

[13] Taizoh Yoshino, Jun Amagai, Kouichi Sebata, Masato Furuya, Tetsuro Kondo1, Noriyuki Kurihara, Yasuhiro Koyama, Mamoru Sekido, and Ryuichi Ichikawa,Yukio Takahashi, Hitoshi Kiuchi, and Akihiro Kaneko "KSP VLBI System and its Measurement Acuracy" Communications Research Laboratory

[14] Levy, G. S.; Ulvestad, J. S.; Edwards, C. D.; Dinardo, S. J.; Stavert, L. R.; Ottenhoff, C. H.; Linfield, R. P.; Whitney, A. R.; Cappallo, R. J.; Rogers, A. E. E.; Hirabayashi, H.; Morimoto, M.; Inoue, M.; Jauncey, D. L.; Nishimura, T. "VLBI using a telescope in Earth orbit. II - Brightness temperatures exceeding the inverse Compton limit" in Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 336, Jan. 15, 1989, p. 1105-1112.

[15] H. Hirabayashi, et al, "Overview and Initial Results of the Very Long Baseline Interferometry Space Observatory Programme" in Science 18 September 1998: Vol. 281. no. 5384, pp. 1825 – 1829 DOI: 10.1126/science.281.5384.1825

[16] Levine, J. I.; Whitney, A. R. "Mark-Iii Real-Time Fringe-Detection System" RADIO INTERFEROMETRY TECH. FOR GEODESY JUN1980 P.333, 1980