WIDE Technical-Report in 2006

Tokyo Lambda Exchange - T-LEX - Status Report

wide-tr-kato-tlex-status-report-00.pdf

WIDE Project : http://www.wide.ad.jp/

If you have any comments on this document, please contact to ad@wide.ad.jp

Title: Tokyo Lambda Exchange — T-LEX — Status Report
Author: Akira Kato
Date: 2006/01/31

1 Brief History

IEEAF (Internet Educational Equal Foundation)[1] is a non-profit organization to accept donations to provide and stimulate real benefits for educational institutions and their learning communities.

The first international link donated by Tyco Telecom to IEEAF has been operational since September 2003 between New York and Amsterdam in order to support demonstration for the iGRID2003 conference held in Amsterdam. The link is actually a combination of an unprotected OC-192c and a protected OC-12c. In order to operate the circuits, UCAID worked as a steward in the U.S. side and SURFnet served in Netherlands side. These stewards offer appropriate gear and operational collaboration to make the link usable.

On the other hand, the IEEAF Pacific link was established in January 2003 with a set of circuits, an unprotected OC-192c and a protected OC-12c, donated by Tyco Telecom. It was expected to support demonstrations in APAN meeting which was held in Fukuoka, Japan. Pacific Northwest Gigapop[2] operated by the University of Washington[3] served as an U.S. side steward and WIDE Project[4] served as a Tokyo side steward. While NTT Communications donated two pairs of dark fiber between Tyco Tokyo TelExchange and NTT Otemachi building, it failed to provide bandwidth to the APAN Fukuoka through the IEEAF circuit due to lack of appropriate gear on both ends.

In the next APAN meeting in Busan, Korea held in August 2003, transmission of a compressed HDTV stream was sent from Portland to Busan by using the OC-12c circuit. From Tokyo, the stream was sent to Busan through JGN and Genkai circuits.

After the APAN Fukuoka meeting, the stewards on the IEEAF Pacific link worked hard to provide appropriate gear by March 2004 in each side. T-LEX[5] has been established to support the IEEAF Pacific link then. After transition of the main circuit, OC-192c, to be a clear channel, the IEEAF Pacific link has been operational since July 2004.

There are four OC-192c class circuits between Japan and the United States. A SINET circuit between Tokyo and New York and a TransPAC2 circuit between Tokyo and Los Angeles are operated at layer-3. An JGN2 circuit between Tokyo and Chicago is operated at layer-2 and layer-3. The IEEAF Pacific circuit is the only lambda capable one to the U.S.

The IEEAF links are part of GLIF[6] — Global Lambda Integraged Facility —. T-LEX is able to serve as layer-1, layer-2, and layer-3 exchange between the IEEAF Pacific link and Japanese R&D networks. It participates GOLE — GLIF Open Lambda Exchange – as well.

2 Current Configuration

We have a Cisco ONS-15454, a optical multiservice provisioning platform, at the center of T-LEX as shown in Figure 1. It terminates the circuits from Tyco Tokyo (or VSNL Tokyo after acquisition of Tyco assets by VSNL) with an OC-192 port at 1550nm and an OC-12 port at 1310nm.

While it is possible to exchange lambdas or sublambdas with Japanese R&D networks, none of them

[1] http://www.ieeaf.org/
[2] http://www.pnwgigapop.net/
[3] http://www.washington.edu/
[4] http://www.wide.ad.jp/
[5] http://www.t-lex.net/
[6] http://www.glif.is/

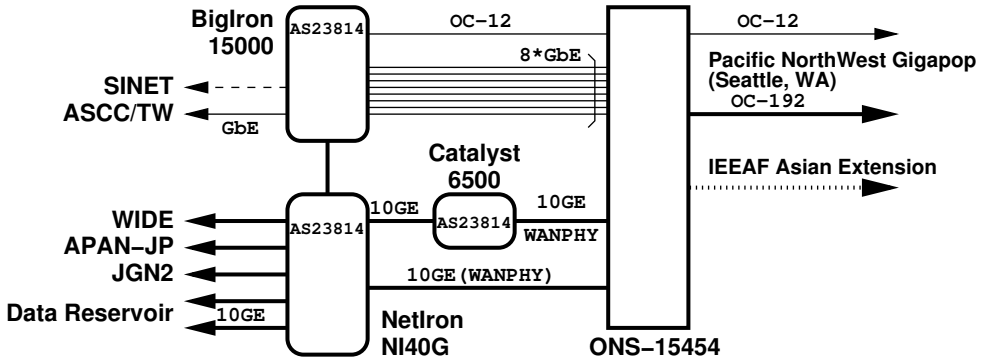

Figure 1: Current Configuration

are offering layer-1 service to the users at this moment. So it is necessary to terminate the lambdas at layer-2 or layer-3 to interconnect with Japanese R&D networks. For this reason, we have a Foundry Bigiron4000 donated by Mitsui and Co. in 2004. We upgraded the chassis to 15slot version so that it can accommodate more 10GE ports. At the bootstrap of T-LEX, Cisco Catalyst6500 with sup730 has been introduced in order to support a 2.4Gbps POS link, which was expected to be required for an uncompressed HDTV transmission.

While Bigiron series switches are able to accommodate 10GEs, they are not able to provide wire-rate switching and true jumbo frame support. During the experiment of long-haul 10GE WANPHY described below, we found that the link didn't work well. NICT has allowed us to use a Netiron 40G switch upon our request. By introduction of additional modules supplied by WIDE Project, the NI40G is now the main switch in T-LEX and offers layer-2 and dual-stack layer-3 service. The NI40G supports 10GE WAN-PHY/LANPHY only.

Currently no sophisticated control plane technology such as UCLP or GMPLS is introduced to T-LEX yet. Its configuration is changed by the operators. We have only a single lambda capable circuit, which is the OC-192c to Seattle. No single-domain control plane technology is useful. With the current users' demands of the configuration change, it is possible to support them by the operators. In order to introduce a control plane technology, we may need to acquire an optical cross connect based on MEMS technology.

In order to support control access between Tokyo and Seattle, and to support direct peering with R&D networks in Seattle, a pair of Foundry Bigirons are used to terminate the OC-12c circuit. The Bigiron 4000 installed in Seattle side has access to the Pacific Wave[7] switch in GbE.

3 Supported Activities

After bootstrap of T-LEX, it supported number of experiments and demonstrations. Brief descriptions are given below. Besides these non-sustainable events, T-LEX is accomodating layer-2 interconnect at GbE between ASCC from Taiwan and KEK in Tsukuba via JGN2.

3.1 SC2003

The first activity which used the OC-192c circuit was Data Reservoir Project[8] (DR) which has been directed by Prof. Kei Hiraki from the University of Tokyo. DR participated SC2003 held in Phoenix in November 2003. As we had no gear to support OC-192c in both ends of the OC-192c circuit, we con-

[7] http://www.pacificwave.net/

[8] http://data-reservoir.adm.s.u-tokyo.ac.jp/

nected it to a Juniper T-320 operated by DR directly in Tokyo. In Portland side, we asked Tyco Telecom to establish a loopback there. By using policy based routing, DR established aggregated performance of 7.56Gbps over about 24,000km links with multiple TCP streams.

3.2 First OC-192c between Tokyo and CERN

The first global scale activity we have participated was a long-haul 10GE WANPHY experiment between Tokyo and CERN in Switzerland with other GLIF members of Pacific Northwest Gigapop, CANARIE, MANLAN, and SURFnet. A 10GE WANPHY link was established over a 18,600km OC-192c circuit between the layer-2 switches in T-LEX and in CERN. We invited Data Reservoir Project to fill the pipe with data. While it has not been approved by LSR, DR recorded 7.56Gbps with the standard length of Ethernet frames between Tokyo and CERN.

3.3 SC2004

In SC2004 held in Pittsburgh in November 2004, DR performed data transfer from a set of DR machines in CERN to another set of DR machines in the venue. The Ethernet frames were sent along a path of SURFnet, CANARIE, IEEAF Pacific to T-LEX, then sent to Chicago via an APAN router in order to reach Pittsburgh. The communication path was approximately 31,200km and DR marked 7.21Gbps with a single TCP stream.

3.4 Christmas of 2004

While DR group recorded an outstanding result in SC2004, it has was not approved as Internet2 Land Speed Record[9] due to its LSR rules. In the Christmas evening of 2004, DR tried to perform a data transmission between a pair of units installed at T-LEX with a world-wide high performance network links — IEEAF, CANARIE, SURFnet, Abilene, and JGN2. It marked 7.21Gbps by a single TCP stream by using standard 1500byte Ethernet frames. The corresponding BW-distance product of 216.3 Pbm/s was approved as LSR in IPv4 single stream class well as multiple stream class.

3.5 JGN2 Symposium

JGN2[10] holds national symposium twice a year and it held one in Osaka in January 2005. While Prof. Larry Smarr from UCSD was expected to give a keynote speech, he wasn't make a trip to Osaka. Instead, he gave a keynote remotely from Seattle by using an uncompressed HDTV stream. As this was a JGN2 symposium, it was expected to use JGN2 OC-192c international circuit between Tokyo and Chicago. But in a few days before the symposium U.S. domestic portion of the fiber cut happened due to a flood in Nevada area. The IEEAF Pacific link was used instead of the JGN2 link. In this event, two GbEs were used to carry an uncompressed HDTV stream between a pair of systems developed by Research Channel[11] and the participants of JGN2 symposium in Osaka enjoyed the very clear video and audio of Prof. Smarr's keynote.

3.6 Aichi World Exposition

Aichi World Exposition was held in Aichi, Japan during March to September, 2005. As one of the closing events, a group of Prof. Osamu Nakamura from Keio University demonstrated a distance musical collaboration over the lambda networks between Amsterdam, Netherlands and the venue. It used an OC-48 lightpath between Amsterdam and T-LEX where two GbEs were carried, and i-Visto was used to transmit the uncompressed HDTV stream. With Internet Metronome technology, the jazz jam session was successfully given in spite of 120ms one-way delay.

3.7 iGRID2005

In iGRID2005 held in September 2005 in San Diego, there were two activities which depended on the

[9] http://lsr.internet2.edu/

[10] http://www.jgn.nict.go.jp/e/

[11] http://www.researchchannel.org/

IEEAF Pacific circuit and T-LEX. One of them were Data Reservoir and in this case they failed to mark a significant record but they got valuable data toward the future improvements. Another activity was n-way uncompressed HDTV demonstration chaired by Mr. Mike Wellings from Research Channel. WIDE Project participated from Tokyo to this innovative demonstration where a number of uncompressed HDTV streams were sent to the venue from various points in the world. On the other hand, IPv4 multicasting was used to feed the uncompressed HDTV stream from the venue. Due to the time difference between Tokyo and San Diego, no interactive session was given but the audience saw head lights of cars on the express way in Tokyo or a traffic jam in the morning commuting hours.

3.8 Around SC|05

Based on the experiences from iGRID2005, Data Reservoir team marked significant record before SC|05 began using a similar circuit to one in the Christmas 2004. They performed 5.58Gbps by using a single stream of TCP over IPv6. In this case no TCP offloading was used because no NIC supports IPv6 TCP offloading at this moment. 167.4Pbm/s was approved as a single stream/multiple streams classes of IPv6 LSR.

Although DR was one of the finalist in the Bandwidth Challenge, it failed to win an award in SC|05. WIDE Project also participated to the n-way uncompressed HDTV streams demonstration by Research Channel.

4 Future Plans

It experienced that the configuration of a lightpath can be done in a day or so provided if enough level of negotiation and coordination had been done in advance. But it also experienced that the debugging of the lightpath was not that easy because there were no sophisticated debugging tools available other than making a loopback in the middle. This fact suggested us that we needed more advanced monitoring and management frameworks to ensure the quality of the lightpaths in addition to the sophisticated control plane technology. We are going to develop T-LEX as one of the mornitorable exchange so that each user may able to see how their packets are sent over T-LEX remotely with appropriate authentication.

On the other hand, we are expecting that some of the Japanese R&D networks will support lambda based services in their limited topologies. In this case, we may need to support as many lightpaths as possible including IEEAF Asian extension link which is expected to be available in near future. As the current optical multiservice provisioning platform, Cisco ONS-15454, supports up to only four OC-192c interfaces, we may need to introduce advanced gear to support more than four OC-192c interfaces or by introduction of optical cross connect so that up to four OC-192c can be used simultaneously.