Canadian Networks for Particle Physics Research

2009 Report to the Standing Committee on Interregional Connectivity, ICFA Panel
January 2009

This report describes the status and plans of the Canadian network used for particle physics research in 2009[1]. We describe the status of the CANARIE network infrastructure. The ATLAS Tier 1 Centre at TRIUMF and ATLAS real-time remote processing system are highlighted as key projects using high-speed networks.

CANARIE Network

The CANARIE network is Canada's national research and education network infrastructure, designed and operated by CANARIE through funding provided by the Government of Canada. The CANARIE network is a flexible and scaleable network infrastructure upon which a wide variety of users can custom-build independent networks in accordance with their application needs and organizational constraints.

To support the widest range of research and application innovation over networks and to support network research, the CANARIE network infrastructure is made up of "low-layer building blocks" (specifically, layer 1 as defined in the ISO Open Systems Interconnection reference model). The "low-layer building blocks" allow resource sharing at the lowest layers, thus permitting the sustained operation of many independent higher-layer networks. Each of the layers use switching methods (circuit or packet) and associated protocols in a topological configuration to best meet users' and future users' varying requirements. This is one of the key architectural principles on which the network has been designed.

Figure 1 illustrates CANARIE's current network infrastructure. The CANARIE network continuously evolves to meet users' demands. In 2006 and 2007, CANARIE acquired dark fibre to build two Dense Wavelength Division multiplexing (DWDM) networks. The Western DWDM system links Seattle, Victoria, Vancouver and, Calgary. The Eastern DWDM links Chicago, Detroit, Toronto, Ottawa, Montreal, and New York City. CANARIE operates a SONET network infrastructure over the two DWDM systems.

Between the geographic expanse of the two DWDMs and eastward beyond Quebec, the connectivity is assured through the lease of 10 Gbps wavelengths from carriers. These 10 Gbps wavelengths are terminated on CANARIE's optical switches, making the carrier infrastructure transparent to the users. Additional 10 Gbps wavelengths can be added to an existing or new route as demand increases.

[1] Additional information can be obtained by contacting Dr. R. Sobie, Director of HEPNET/Canada (rsobie@uvic.ca)

CANARIE-lit wavelengths

CANARIE-lit wavelengths are possible because in 2006 and 2007 CANARIE undertook in-house projects under which dark fibre was purchased and lit with CANARIE-owned and operated DWDM system. The two DWDM systems, depicted in blue in Figure 1, were built based on ROADM (Reconfigurable Optical Add-Drop Multiplexer) equipment, the latest in DWDM technology. ROADM enables remote software reconfiguration of wavelength routing and permitting low-cost operations as compared to earlier DWDM technologies. The ROADM hardware supports up to 72 x 10 Gbps or 40 Gbps wavelengths with the client interface framed in either 10 GbE or OC-192 SONET, and will support 100 Gbps wavelength when hardware is available.

The first ROADM project, called the "Eastern Canadian ROADM project", established a 2700 km, multi-degree optical mesh network along the busiest corridor of the CANARIE Networks: Chicago – Toronto – Ottawa – Montréal – NYC with the two new points of presence in Windsor and in Boston. The infrastructure between Windsor – Toronto – Ottawa is being shared with the Ontario-based ORAN (i.e. Optical Regional Advanced Network of Ontario (ORANO)).

The second ROADM project, called the "Western Canadian ROADM project", established a 1500 km, multi-degree optical mesh network from Seattle – Victoria – Vancouver – Kamloops – Calgary, with a spur from Kamloops to Kelowna. The infrastructure is being shared with the British Columbia and Alberta ORANs (i.e. BCNet and Cybera).

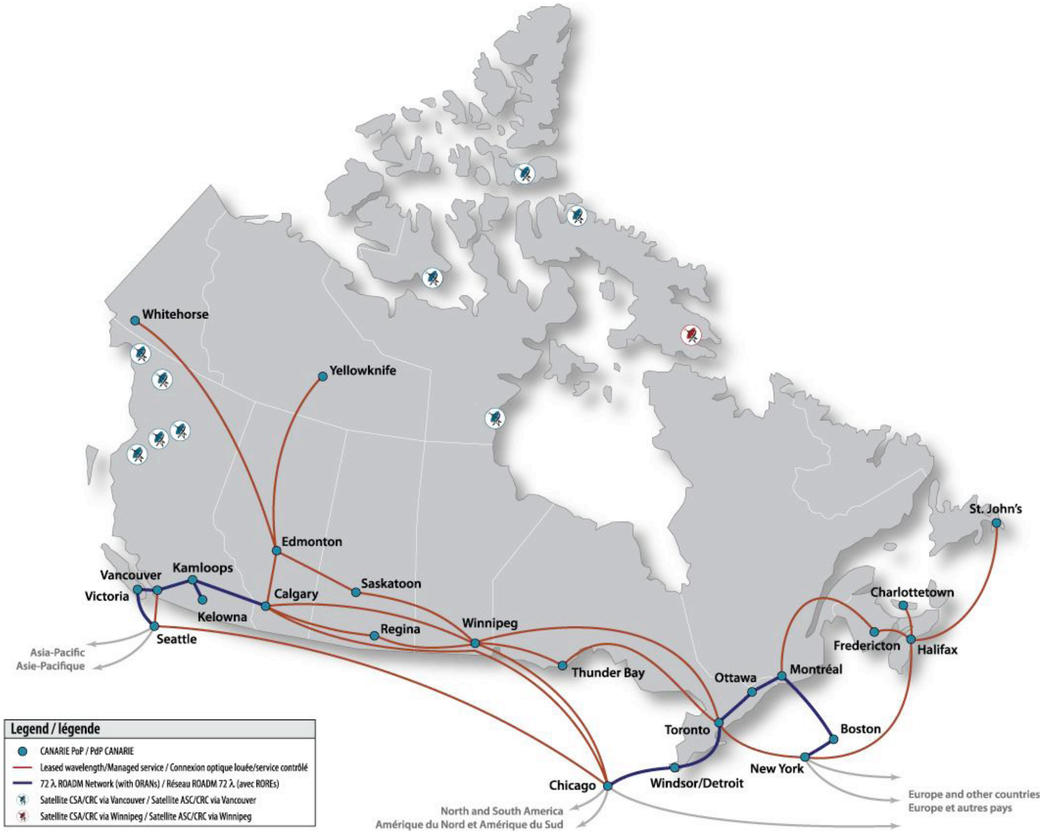
Whitehorse
Yellowknife
Edmonton
Saskatoon
Vancouver
Kamloops
Victoria
Kelowna
Calgary
Regina
Winnipeg
Seattle
Asia-Pacific
Asie-Pacifique
Thunder Bay
St. John's
Charlottetown
Fredericton
Halifax
Ottawa
Montréal
Toronto
Boston
New York
Windsor/Detroit
Chicago
North and South America
Amérique du Nord et Amérique du Sud
Europe and other countries
Europe et autres pays
Legend / légende
CANARIE PoP / PdP CANARIE
Leased wavelength/Managed service / Connexion optique louée/service contrôlé
72 λ ROADM Network (with ORANs) / Réseau ROADM 72 λ (avec ROREs)
Satellite CSA/CRC via Vancouver / Satellite ASC/CRC via Vancouver
Satellite CSA/CRC via Winnipeg / Satellite ASC/CRC via Winnipeg

Figure 1: The CANARIE network

Wavelength swapping

Beside CANARIE-lit and leased wavelengths, CANARIE has wavelength swapping agreements with like-minded organizations in exchange for excess capacity on the CANARIE Network.

At present, CANARIE has wavelength swapping agreements with National Lambda Rail (NLR) and SURFNet. CANARIE is currently working on a swapping agreement with NORDUnet to exchange sub 10Gbps network capacity. Under that new agreement, NORDUnet would provide 2 Gbps capacity from St. John‘s, Newfoundland to Amsterdam, through Reykjavik and Copenhagen, in exchange for 2 Gbps network capacity between St. John‘s, Newfoundland and New York City. This 2 Gbps capacity through Iceland will provide a physical diverse path over the Atlantic to Europe as the current cross-Atlantic wavelengths are all through New York.

CANARIE lightpath services

CANARIE lightpath services (a.k.a. private-line in industry) are delivered on GbE and 10 GbE client interfaces over SONET or a dedicated wavelength. 10Gps SONET wavelengths are usually partitioned into smaller capacity channels, from 155 Mbps to GbE, and up to a full 10 GbE. With the new ROADM networks, lightpaths can also be 10GbE wavelengths that can be dropped directly into researchers' equipment thus bypassing CANARIE optical switches. This reduces the total cost (CANARIE and End-User) of bringing up a 10GbE lightpath. Canadian researchers will have increased capability to develop and utilize high bandwidth applications in leading edge national and international collaborations.

CANARIE IP Network services

A number of lightpaths on CANARIE‘s network infrastructure are used to provide traditional IP services. The last and third layer of the CANARIE network is used to provide network service, with full and equal support for IPv4 and IPv6 unicast and multicast routing. Internally the IP network is comprised of five major routing nodes, which are located in Calgary, Winnipeg, Toronto, Montreal, and Halifax. A sixth smaller node was in Edmonton to aggregate IP network traffic for the Yukon Territory and Northwest Territory GigaPoPs.

In the early of 2009, CANARIE issued a Request for Proposals for the Provision of IP Routing Equipment to upgrade the core routing equipment for the CANARIE Network at the major five routing nodes. CANARIE selected the Juniper MX-480 switch routing platform because it offered a superior technical solution in compliance with the criteria of technical requirement and support to research. Key features of the Juniper MX-480 are: 10Gbps interfaces, not only L3 (IP networking) but L2 (Ethernet switching), 240G full duplex backplane, Logical/Virtual Routers. This choice provides a strong solution to maintain support of the CANARIE Network IP routing needs beyond 5 years. The new routing layer will provide CANARIE with the means to offer the network services required by advanced users.

There are seven major network segments and one minor interconnecting the routing nodes. During the deployment of the new routing platform, the capacity of the seven major segments has been increased to a minimum of 2 Gbps except the Toronto-Montreal segment, which operates at 10Gbps. To keep the network segments uncongested; capacity can be increased in increments of 155Mbps up to 10Gbps.

In addition to the eight internal network segments, there are five external network segments which extend to international R&E layer 2 exchanges: the Pacific Wave in Seattle, StarLight in Chicago, and Manhattan Landing (MANLAN) in New York. The capacity of these external segments has also been increased.

Figure 2 depicts the CANARIE IP network physical connectivity. Through international peerings at the international exchanges and through transit of R&E routes through some peers, IP connectivity truly is global in reach.

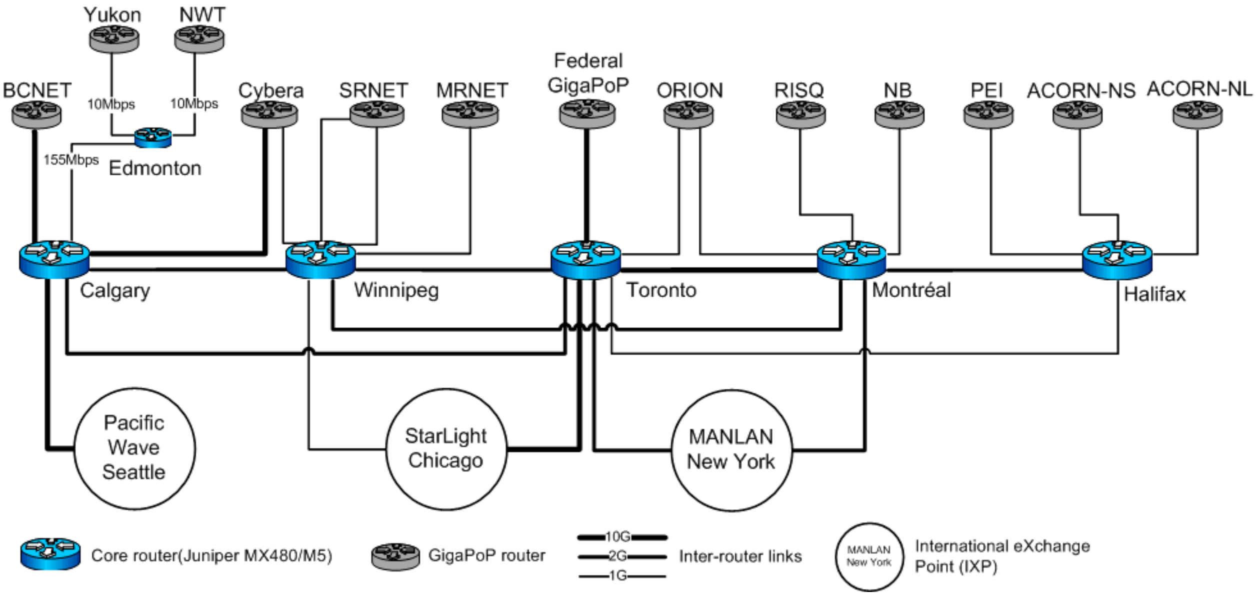

Figure 2: The CANARIE IP network

International collaboration

The CANARIE Network is connected to GLIF (the Global Lambda Integrated Facility) through three International eXchange Points (IXPs): Pacific Wave, StarLight, and MANLAN (Manhattan Landing Exchange Point). This connectivity provides CANARIE with the capacity to create dedicated lightpaths throughout most of the world, enabling researchers to perform collaborative research.

GLIF is a technical forum where lightpath-enabled network resource providers work together to develop inter-operable international infrastructure. CANARIE is a funding member of GLIF. By identifying equipment, connection requirements, necessary engineering functions and services, and effective end-to-end processes, GLIF presents an integrated perspective of the networks to the end user.

In the last year, the GLIF Technical Issues Working Group has met twice and constituted four task forces to work on specific issues. These are:

- End-to-end lightpath Global ID
- perfSONAR end-to-end monitoring infrastructure
- Service Level Specification (SLS)
- Dynamic lightpath at GLIF open-lightpath-exchanges (GOLEs)
- GLIF network interface (GNI) API

CANARIE has been actively involved in end-to-end lightpath Global ID and perfSONAR end-to-end monitoring infrastructure task forces for the past year. CANARIE will continue to be involved in both task forces. In the last GLIF meeting, CANARIE has presented the challenges of maintaining and facing the end of support of the current GOLE equipment. CANARIE has volunteered to work on a draft proposal to better define the needs of the next generation GOLE architecture and equipment.

ATLAS Tier 1 Computing Centre at the TRIUMF Laboratory

TRIUMF, Canada's National Laboratory for Nuclear and Particle Research, has built a Tier-1 (T1) Computing Centre for the ATLAS experiment in Canada. The TRIUMF Centre is linked to the LHC Worldwide Computing Grid (WLCG) and provides an interface to a grid of computing resources at universities across Canada.

In July 2005 CANARIE signed a Memorandum of Understanding (MOU) with HEPnet/Canada, ATLAS Canada and TRIUMF to provide the high energy physics community with a dedicated 10 Gbps APN across Canada and initial 5 Gbps lightpath to the CERN Tier-0 (T0) Centre. This lightpath became active in December 2006.

The TRIUMF T1 to CERN T0 circuit, depicted in Figure 3, runs over the CANARIE infrastructure until it disembarks North America in New York City. Each T1 site must use a small or series of small publicly routable Classless Inter-Domain Routing (CIDR) blocks as only traffic from the Large Hadron Collider Private Optical Network (LHCOPN) address space is allowed to flow over the network. Exchange of routing information is performed using Border Gateway Protocol (BGP) at the T1 and T0 institutions. The 5 Gbps lightpath, terminated on 10GbE LANPHY interfaces in both ends, transits Canada west to east. The lightpath travels over the BCNet network from TRIUMF to the CANARIE PoP at UBC, and continues along CANARIEs network, then debarks North America at the MANLAN transit exchange in New York City. The lightpath enters Europe on SURFnet in Amsterdam and then transits Geant2 network to CERN.

The TRIUMF T1 will hold 4.3% of the ATLAS data, and it is anticipated that a 5 Gbps link will be sufficient for the first few years of LHC operation. However should the demand increase beyond 5Gbps, the lightpath capacity can be increased in 155Mbps increments up to the full 10 Gbps.

The backup link for the primary 5 Gbps link passes from the Vancouver CANARIE OME to the Pacific Northwest Gigapop in Seattle, then via Chicago to Amsterdam. This link provides an alternate fibre path across North America and the Atlantic. Even at a lower capacity of 1 Gbps it is expected to be able to temporarily handle loads while the 5 Gbps link is restored. The tertiary backup link added in 2008 travels via Victoria to Pacific Northwest Gigapop and then to the Brookhaven National Laboratory (BNL). In the event of the failure of the primary and secondary links traffic will be carried via the US LHC Network to CERN. The tertiary backup also acts as T1 to T1 link which is particularly advantageous because BNL will host 25% of the ATLAS Data. In 2008 an additional T1 to T1 link was established with SARA (TRIUMF's Tier-1 partner) following the same path as the primary 5 Gbps link. The 2009 delays in LHC startup caused network configurations to remain relatively static throughout the year.

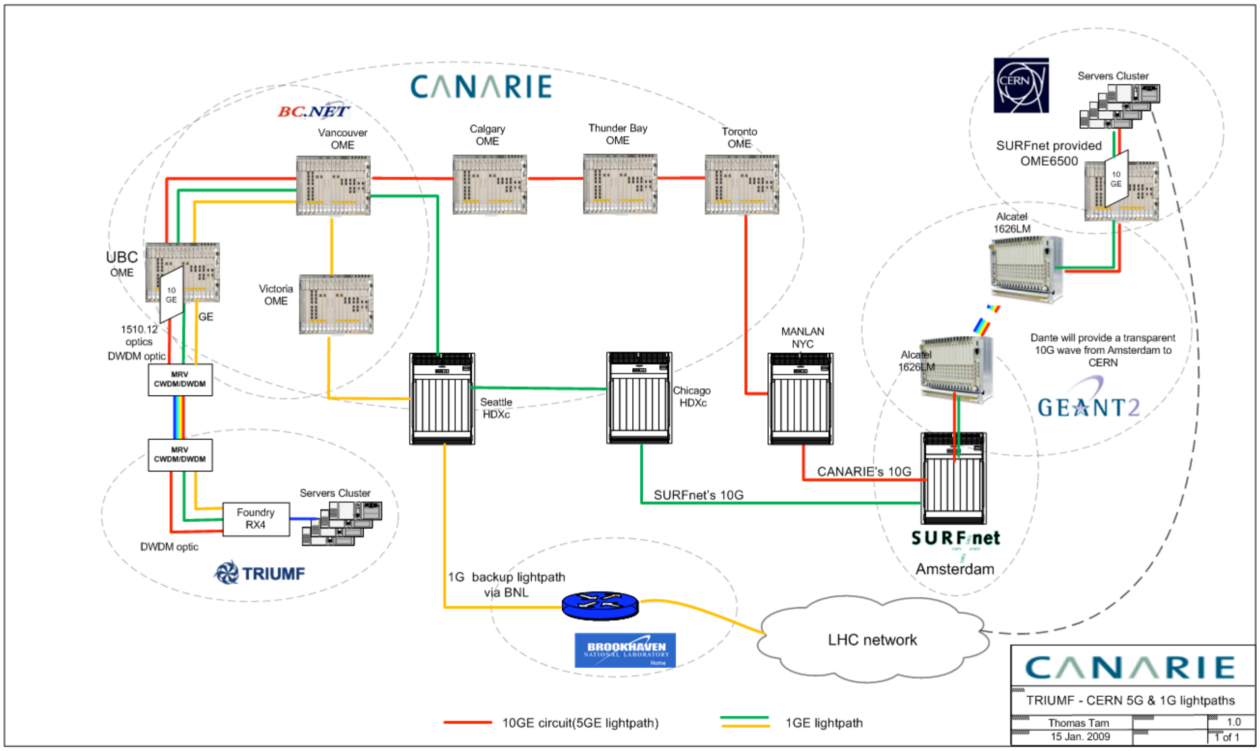

Figure 3: TRIUMF – CERN 5 Gbps and 1Gbps lightpaths

Canada hosts Tier 2 centres at the University of Victoria, University of Alberta, University of Toronto, Simon Fraser University and McGill University. The T1 to T2 connections in Canada follow the successful LHCOPN model with TRIUMF at the centre of a star pattern. Four of the T2 institutions have a 1 Gbps lightpath to TRIUMF while Simon Fraser University shares a 1 Gbps circuit with the WestGrid computing consortium. Each path, excepting Simon Fraser University, is carried on CANARIE ROADM network.

In Canada a T2 to T1 lightpaths transit the CANARIE ROADM from the POP nearest to the Tier-2 to the University of British Columbia(UBC) Life Sciences ROADM POP near TRIUMF. Three Tier-2s: the University of Victoria, McGill University and the University of Toronto use CWDM multiplexing equipment to arrive at a CANARIE POP. A typical example is the University of Victoria (Victoria, British Columbia) located near the southern tip of Vancouver Island. The University of Victoria connects to the Victoria CANARIE ROADM POP in the Victoria Transit Exchange by traversing an MRV CWDM Mux (a distance of approximately 8 km) and is then carried via the ROADM network before disembarking at the University of British Columbia Life Sciences CANARIE ROADM POP near TRIUMF. The lightpath again transits a MRV CWDM Mux to arrive at TRIUMF a few kilometers away.

Simon Fraser University is the exceptional case due to its proximity to TRIUMF. SFU is located in the Vancouver area and is able to connect to TRIUMF exclusively via the BCNet network at the Vancouver Transit Exchange in Habour Centre in downtown Vancouver.

Real-time Remote Processing Systems and Grids for the Atlas High Level Trigger

There are continuing efforts in Canada to evaluate the use of remote processing farms as part of the Event Filter system of the ATLAS High Level Trigger. The benefit of employing such resources is clear given that ATLAS may be low in on-site computing resources especially at the beginning of data taking, signaling a clear requirement for direct off-site traffic for calibration and monitoring tasks. Institutes contributing to this effort include the University of Alberta along with groups from CERN, Manchester, Krakow, Copenhagen and NIKHEF.

The computing farms being developed will interface to the online dataflow and receive data using an environment identical to a standard processing task (PT) used by the Event Filter to make trigger decisions. The key difference is that a monitoring task (MT) does not have to process every event. Hence the MTs are only fed as many events as they can handle and do not interfere with the normal dataflow. This makes them ideal for an offsite situation since, in the event of a serious network problem, they would not prevent ATLAS from taking data. Locating these monitoring resources offsite means that the limited power and cooling for CERN-based computing at point 1 can be dedicating to triggering. In addition, having substantial computing resources with access to online events provides an ideal testbed for new trigger algorithms which can be tested with millions of real events before being deployed in the EF farm at CERN and also provides additional CPU for monitoring tasks that otherwise might take too long to gain useful statistics such as cell level monitoring of the ATLAS LAr calorimeter.

The full Event Filter farm requires roughly 3,000 2.3 GHz quad-core CPUs. Making the modest assumption that 1% of events should be monitored then 120 2.3 GHz cores will be required. Currently the University of Alberta has a 200 2GHz core cluster, which is used for testing and development of the remote farms. This will be sufficient to run one monitoring algorithm on 1% of the data with spare capacity for testing new algorithms as well as for additional overheads such as the data quality-monitoring framework (DQMF).

The development strategy for the remote farms is split into three phases:

1. Run a saved bytestream data file through a remote system running online monitoring code.
2. Use automatic scripts to fetch recent bytestream data directly from the CASTOR disk store, before it is written to tape, and process them remotely.
3. Fully integrate with the online dataflow at point 1 and receive data directly from the SFIs in the same manner as the local monitoring machines do currently.

The first phase was completed in 2008. In 2009 work continued on the second phase of automatically retrieving data from CASTOR to be processed by the remote site running the online monitoring code. The group also continued the investigation of integrating the ATLAS High Level Trigger Event Filter system with Grid technologies based upon a pilot-job system, an endeavor led by the Krakow group. Concurrently NIKHEF has continued to focus on the necessary modifications to the T/DAQ software that would enable tagging of events at Level-2 component of the trigger decision and subsequent routing to external sites for processing.