TRIUMF Site Report

Steve McDonald

HEPiX 2009, Sweden Umea

Organizational Changes

- 2 Retirements
  - Renee Poutissou - Dec (DAQ group leader)
  - Joe Chuma - May

- 1 Resign

  - Rod Walker - Oct ,LMU Munchen Germany (ATLAS Grid Computing)

  - Position filled by Di Qing (CERN/ASGC T1)

- TRIUMF Computing is 26 people in total down by 2

  - 9 of which are dedicated entirely to ATLAS Tier1 Center operations

  - No replacements yet or even approved

Bacula Tape Backup

- After using Amanda for the past 6 years for our site backup, we have started to replace it with Bacula.

- Improved archiving and scheduling capabilities

- GUI to access and restore files (can be user initiated with appropriate privileges)

- Certificate based auth between server/clients

  - Client initiated backups "feature request"

- Data stored in SQL postgres/mysql

- More efficient use of network bandwidth

- Clients for Linux / Windows / MacOSX

- Being actively developed/improved lots of documentation

Laptop Backup

- Starting Summer of last year can now offer a laptop backup solution to users, didn't really have anything convenient before.

- Product is from Atempo http://www.atempo.com/

- CDP (Continuous Data Protection)

- Support for Windows, MacOSX, Linux

- Entire Filesystem or files/directories based on set of pre-defined policies.

- User restorations via web interface

- File changes on the laptops are cached and propagated to the server whenever the laptop gets an internet connection.

TRUMF Core Network Upgrade

- Until ~5yrs ago, TRIUMF's network was one BIG flat Layer 2 /16 broadcast domain, something I inherited and wanted to change back in 2000.

- Difficult because of the way they had assigned the initial IP addresses required renumbering ~300 IP, 200 of which were on the Controls network. Also every machine on site had to have a netmask change

- TRIUMF accelerator operates 9 month of the year,

- Networking group is smaller than small just me and 1 other

- But we now have a fully routed and meshed network core

- OSPF internally

- BGP on the Edge for the ATLAS T/T1/T2 lightpaths

- By the end of this summer will also have the edge switches with fully redundant RSMLT paths to the core
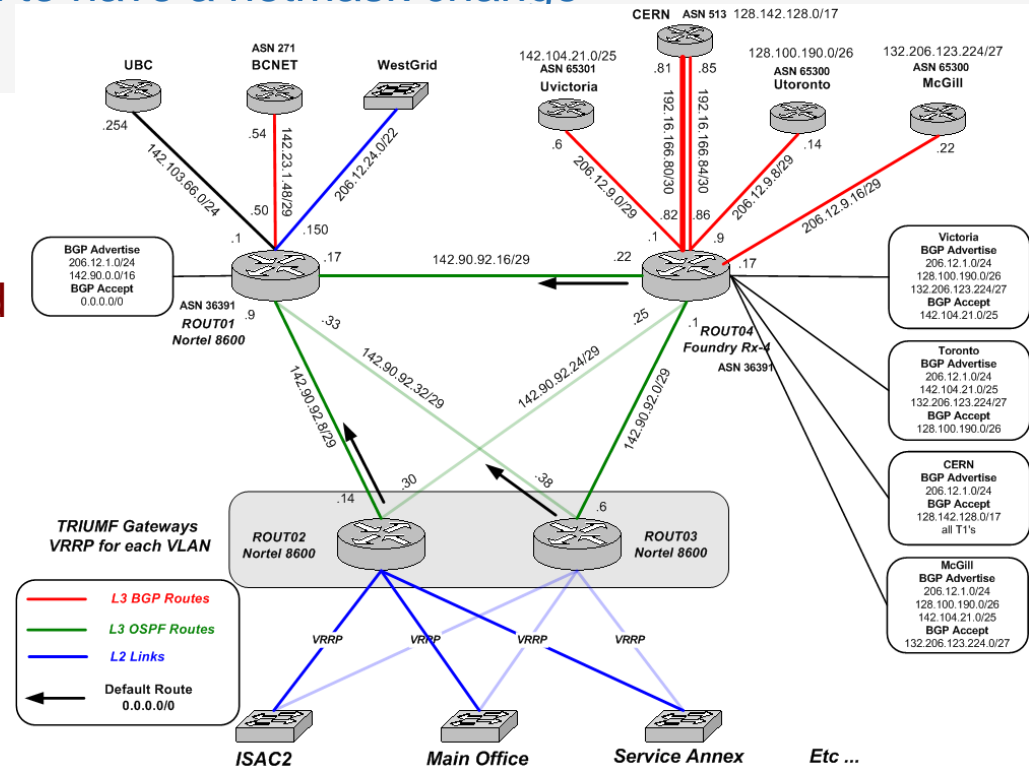
"Self Bricking" Seagate Disks

- In Dec of 2008 Seagate announced a problem affecting various Seagate drives

  - *"A firmware issue has been identified that affects a small number of Seagate Barracuda 7200.11 hard drive models which may result in data becoming inaccessible after a power-off/on operation. The affected products are Barracuda 7200.11, Barracuda ES.2 SATA, and DiamondMax 22"*

- An inventory of Seagate disks purchased by TRIUMF identified ~40 of 400 disks, 10% were affected.

- 2 of these 40 "Self-Bricked" during the shutdown procedure to perform the necessary firmware patch.

Unified Storage

- Sun 7410 Unified Storage

- Based on ZFS, RAID-Z DP

- Provides iSCSI, NFS, in kernel CIF

- Very nice intuitive management GUI

- Hybrid storage pools, DRAM, SSD, SATA

- File & block level snapshots, D-trace Analytics

- Shared storage for Xen, openVZ VM fail-over, mail and homes dirs

- 10GbE, Snapshots, Clustering (is available), based on OSS.

- Scales to 500TB (purchased 10TB), Thin provisioning

- Very similar to low end Net App 2020/2050 but with better scalability less filesystem overhead, and more built-in features.
  - Net App 2020 only scales to 57TB Raw/27 TB useable
 - Net App 2050 ~120TB Raw/65 TB useable
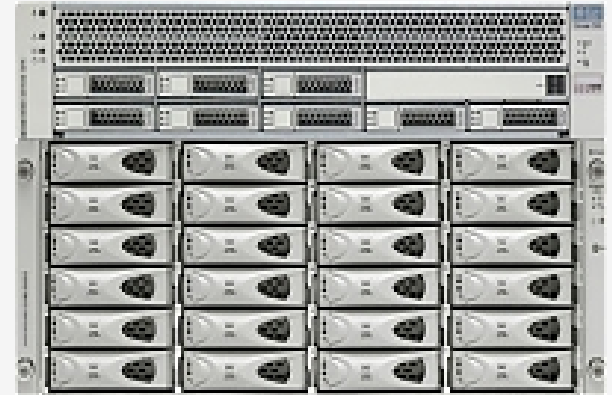
Virtualization @ TRIUMF

- Virtualization - Xen and openVZ
  - Starting to use openVZ where appropriate, single kernel for all Linux OS's, Container based, much easier mgmt, better performance, less overhead than traditional VM.
  - Good if you only want to host Linux VPS (Virtual Private Servers)
  - Still need Xen for some Windows VM's. Also use HyperV for windows VM's
  - openVZ allows more control over resources, very little overhead, 1-2% at the virtualization layer, Networking is also much simpler
- Saw VERY slow performance on one of our AMD Virtual Host servers. Turned out to be a RHEL problem with XEN Guest kenels If dom0 started cpuspeed daemon the Xen guest's CPU does not increase under load, It would stay fixed at 1GHz it also affected on latency.

Web Presence

- February Thu 12th launched a new public web face for TRIUMF

- Based on Drupal 6

- Contracted an external company Xeno-media based in Chicago. Xeno has a history of working with Scientific laboratories, Fermilab, SLAC, CERN, USLHC, Symmetry Magazine, Interactions.org

- Approximately 6 month to roll-out first stage, will continue to migrate old content over time.
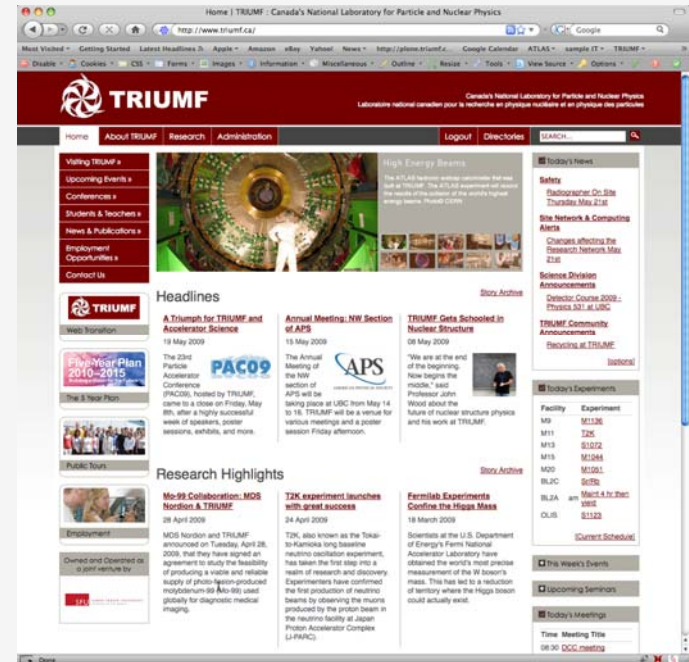
Mac OSX support

- In the past TRIUMF Computing Services has always said "we do NOT support MacOSX", what support we do provide is a "best effort".

- Despite the fact that several of us even in computing have begun using Mac OSX on laptops

- While researching our calendaring options we surveyed what OS/clients our users were using to connect to our mailserver. The results were somewhat surprising.

  – Windows 60%, Linux 20%, MacOSX 20%

ATLAS Tier 1

- Operating very well with minimal 'but dedicated' staff
  - 5 System Support, 3 User Support, 1 Tech
- No new additions yet, 2009 RFP was delayed
- Have experienced a couple of issues that really tested the automated worker nodes shutdown scripts,
  - Cooling leak (very slow no alarms until too late.)
  - Multiple site power failures of backup diesel generator

- Despite these occurrences continue to meet the MOU requirement and are providing >98% reliability.
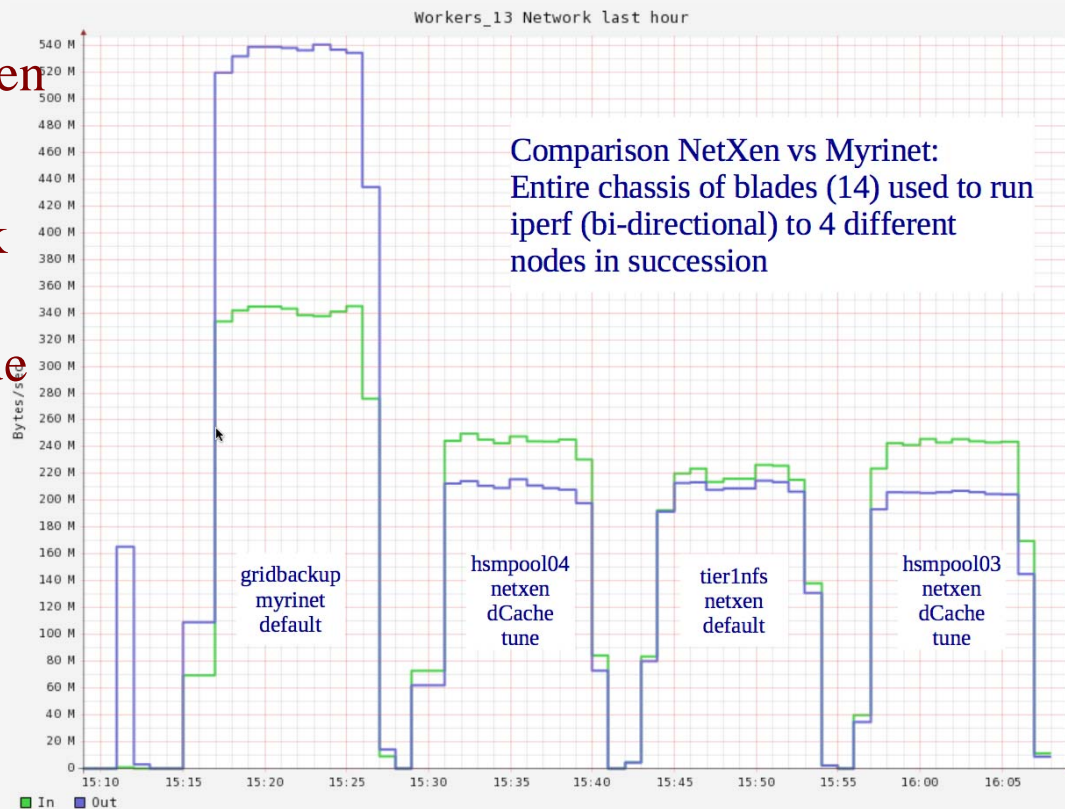

TRIUMF availability from May-08 to May-09
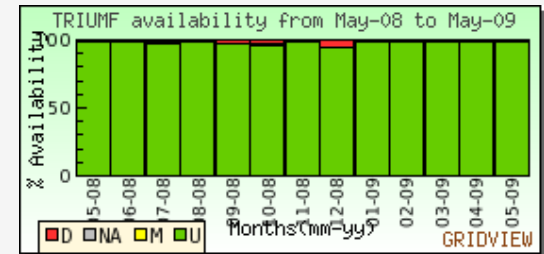
ATLAS Tier 1 Upgrades

	Existing	**2009**	**2010**
CPU (HEP-SPEC-06)	6640	**4000**	**3000**
Disk (TB -usable)	720	**700**	**800**
Tape (TB)	554	**-**	**200**

- RPF originally scheduled for last Fall has just gone out, due to delayed LHC startup, closes June 12th

- Option to purchase 2010 resources under same RFP, even from a different vendor, will almost certainly be Dual Quad Core, 2TB disks.

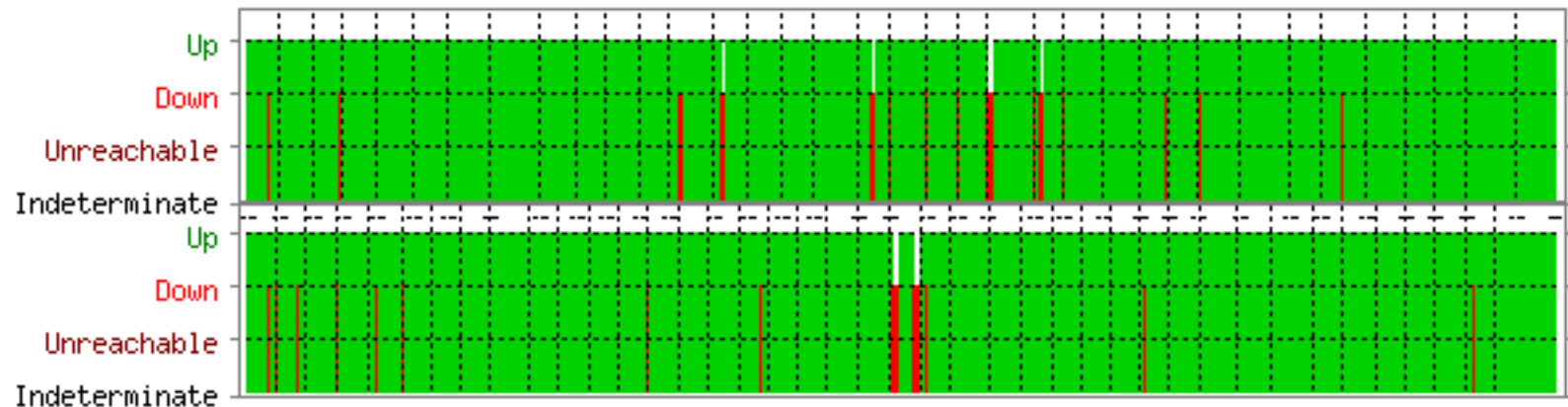- A second ORACLE RAC will also be purchased in the Fall to support the TAGs database

Myrinet / NetXen 10G cards

• IBM cluster shipped with netxen 10G cards

• Initial driver problems network would lockup under heavy load, new firmware corrected this issue

• 4 or 12 netxen cards replaced

• 2 corrupting data

• 1 did not come back after a planned power outage

• 1 node would not boot reliably

• After a report at HEPiX - CERN are no looking at the Myrinet Cards

Iperf tests from one blade chassis to various 10G servers

Workers_13 Network last hour

Comparison NetXen vs Myrinet:
Entire chassis of blades (14) used to run iperf (bi-directional) to 4 different nodes in succession

gridbackup myrinet default

hsmpool04 netxen dCache tune

tier1nfs netxen default

hsmpool03 netxen dCache tune

Frequency of Primary & Secondary network path failures

Graph of primary & secondary path availability 2008

- Almost never happen at the same time! ***Or do they?***
- Last time this graph was presented both of TRIUMF's primary and secondary paths to CERN we interrupted within one hour.
- Failover through BNL, BNL also provides a redundant path in Europe via France, rather than the heavily populated path through Basel.
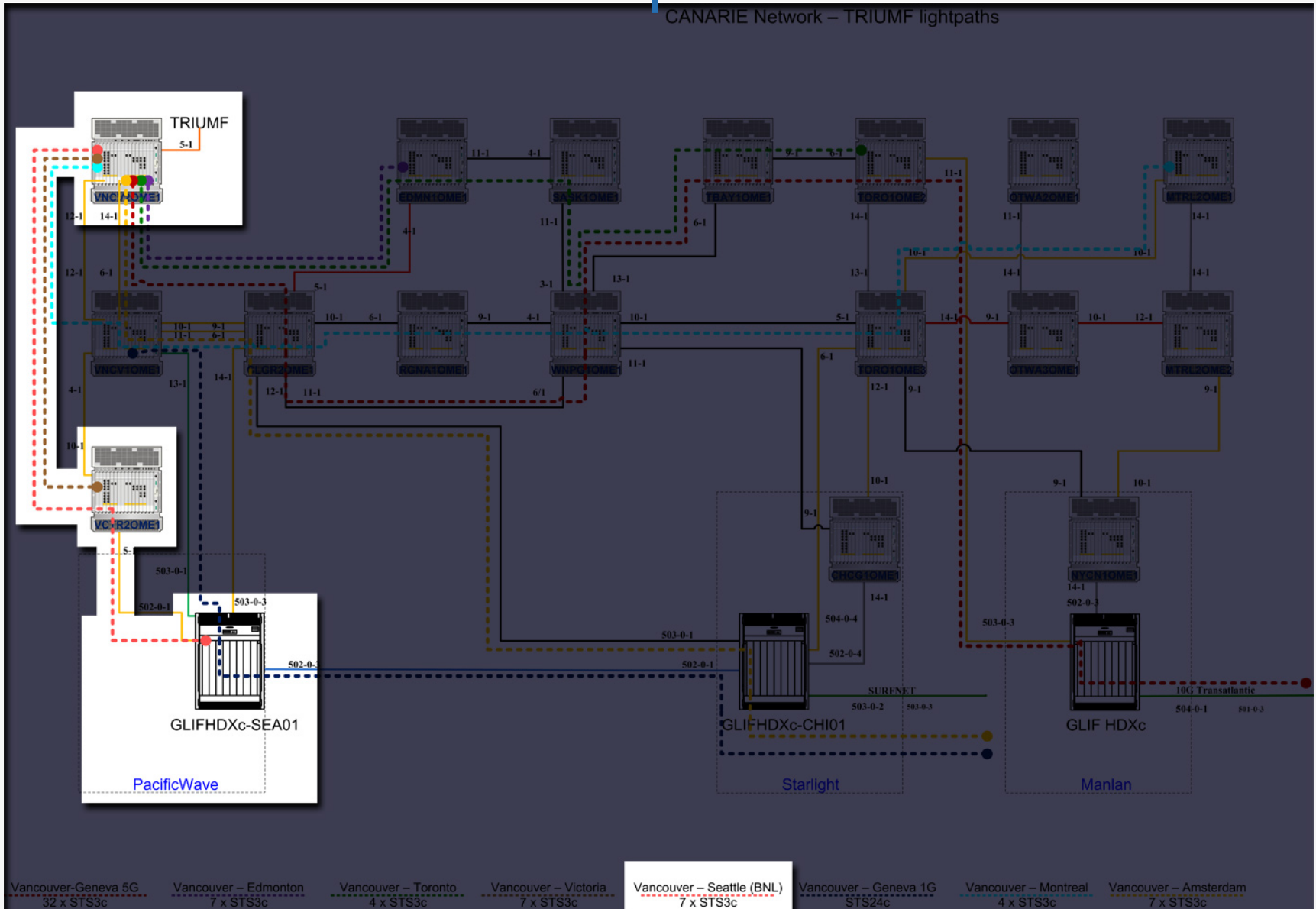- Lot of work gone into making paths physically diverse and redundant

LHCOPN
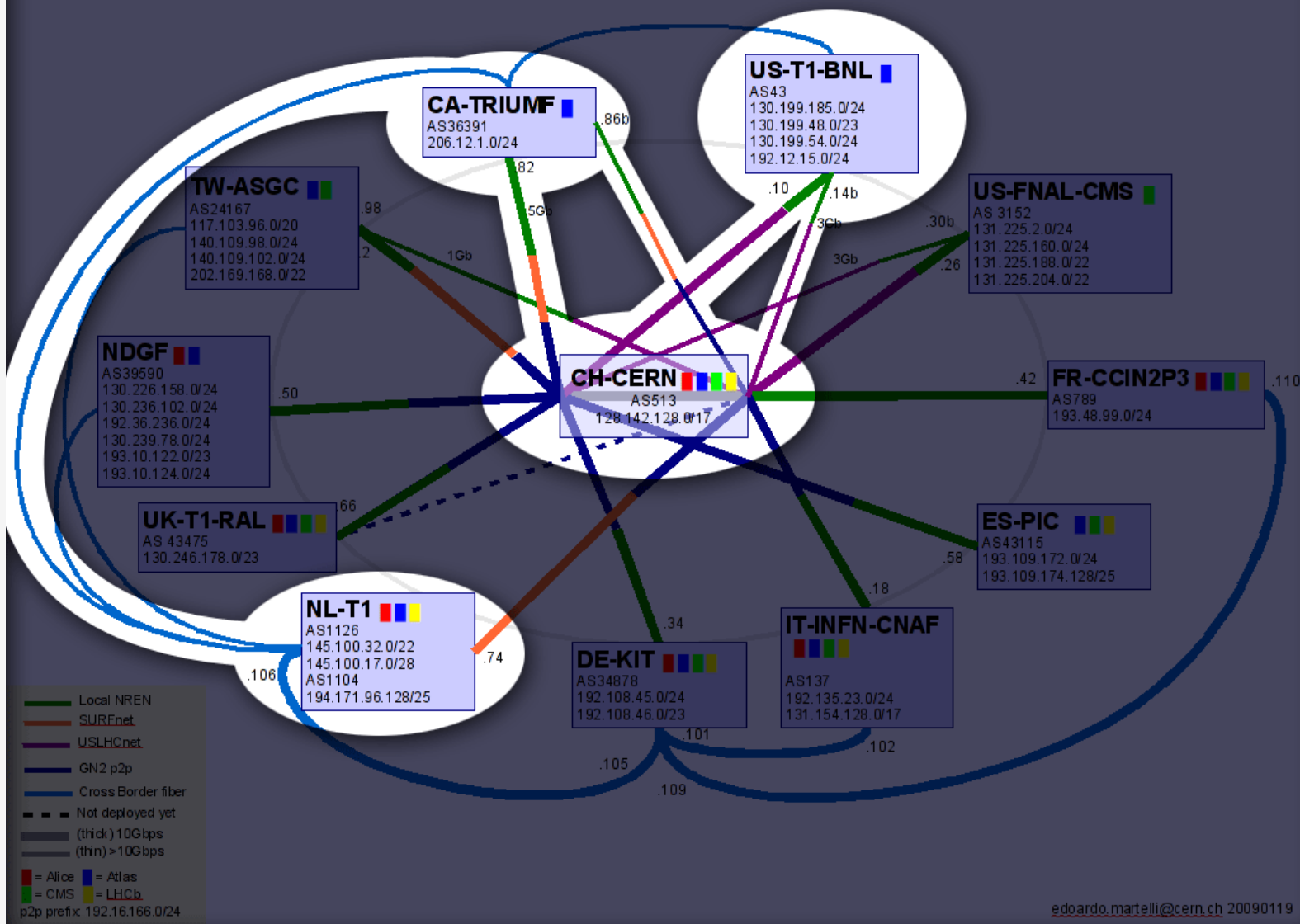
Canadian Optical Network

European Optical Network
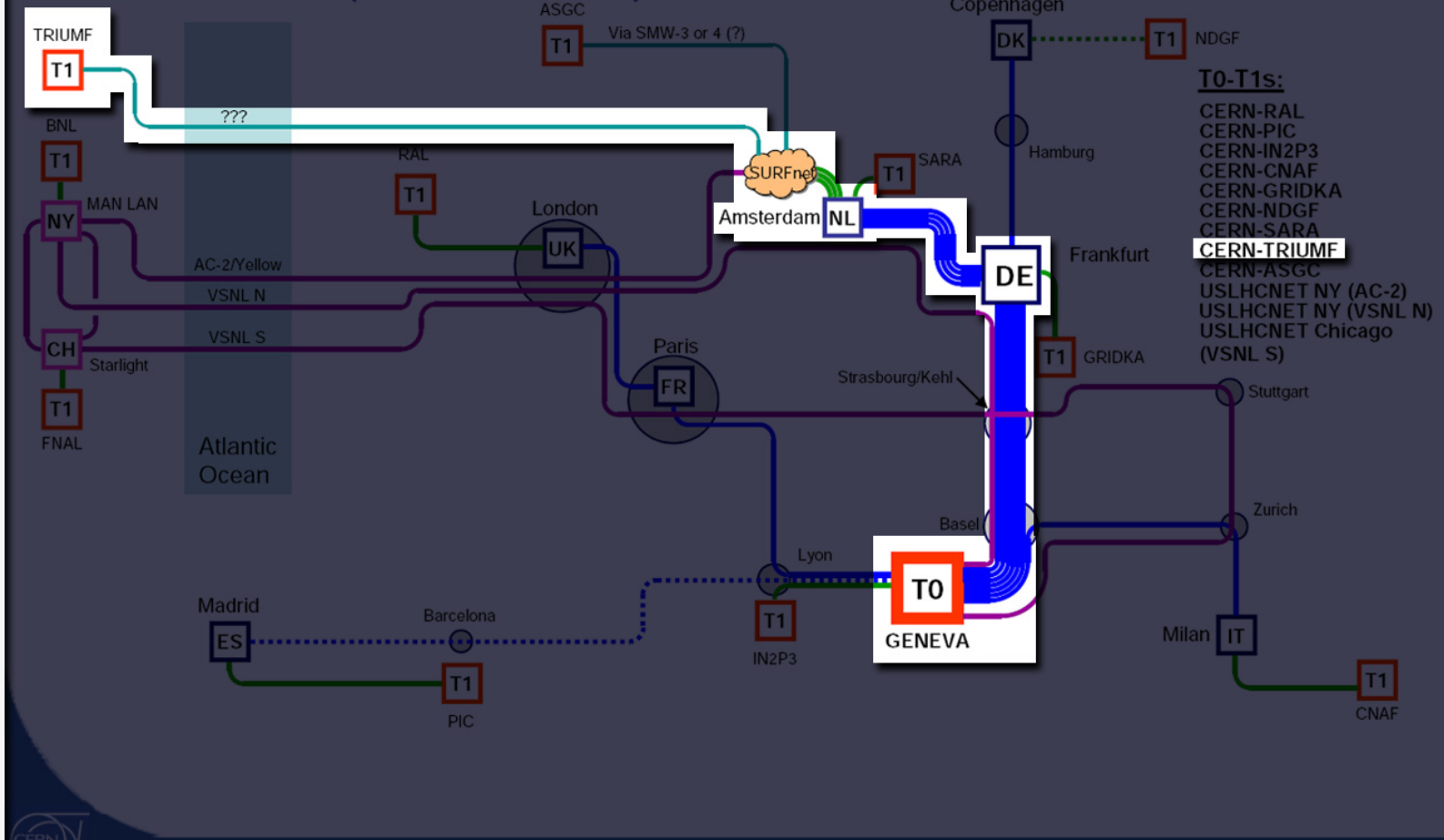
However …
best laid plans can not thwart University Development

All of TRIUMF external network connection and ATLAS lightpaths pass through this shack. Destined to be the entrance to a large supermarket.
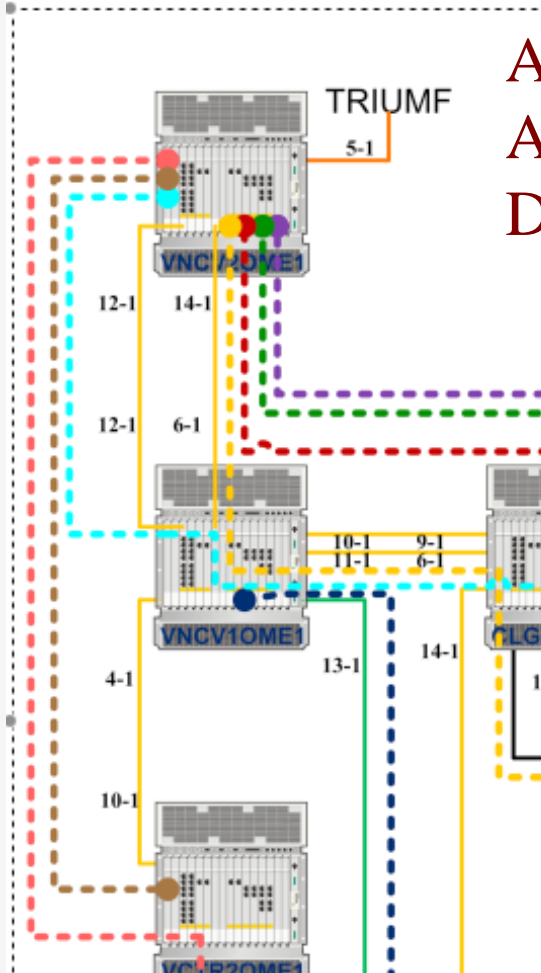
Thank you

TRIUMF

**4004 Wesbrook Mall
Vancouver, B.C. Canada V6T 2A3
Tel: 604 222-1047   Fax: 604 222-1074
www.triumf.ca**

Merci