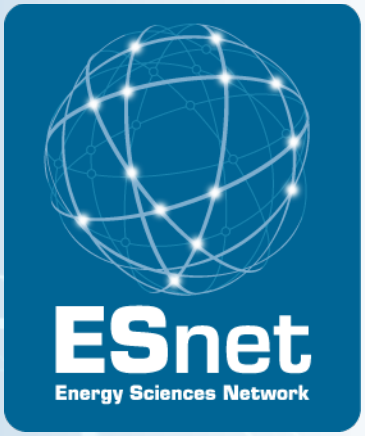


LHCONE Overlay Networks

Mike O'Connor

ESnet Engineering Group

LHCOPN and LHCONE Joint Meeting

Washington DC

Jan. 31, 2013

The LHCONE Overlay

The LHC global community has implemented the LHCONE network as a Virtual Routing and Forwarding (VRF) overlay network.

In the general case, an overlay network exploits available capacity in a common network substrate, sharing resources with other networks and services of various types at multiple network layers.

At the provider edge, customers and peer networks have no means of differentiating a network overlay from a typical layer three routed network implementation. This particular overlay network attribute is an advantage since common NOC procedures require little if any adaptation to achieve a well supported production service.

Why Choose an Overlay?

The privacy and service characteristics of a “walled garden” multipoint internet correlate favorably with the requirements necessary to serve the well defined global LHC collaboration. Employing large scale data movement as an integral design element of the LHC compute model, a well managed high performance network of networks becomes an obvious choice when building a cohesive globally dispersed network footprint.

To share or not to share?

Obviously the cost and effort required to design, procure and deploy a completely new dedicated infrastructure across the potentially hundreds of of multi-national stake holders was not realistic. The obvious choice was to take advantage of existing capacity in general purpose R&E networks in a targeted and well defined fashion that economically addressed the requirements of the LHC computing model.

Why Bother With an Overlay?

Shared resources are economical but why not just connect through the general purpose R&E Internet?

LHCONE flows by definition can be easily and accurately characterized as “science” traffic, specifically between the well known set of LHC collaborators and defined by any instance of the complete VRF routing table.

Once defined and distinct, the flows through the LHCONE are available to be traffic engineered across preferred paths, under oceans or through regionals. High performance customer perimeters can be traversed using a Science DMZ, avoiding bottlenecks in route to specific purpose built data transfer nodes. The inherent agility of the L3 overlay as a distinct and manageable object is an important advantage.

The L3 VPN technique is well supported by open standards and available across many routing platforms, providing a high performing, reliable and extensible implementation.

LHCONE Today

Based on the responses to my survey, it's clear that the production LHCONE network is implemented as a virtual layer 3 overlay network.

The LHCONE is:
A set of physical and virtual router interfaces.
A routing table containing common reachability information.
A relatively small number of dedicated lambdas, typically 10G.
Built on open standards over a broad range of routing platforms
Alcatel-Lucent, Brocade, Cisco, Juniper, various models too numerous to mention.

When considered in proportion to the infrastructure of any given NSP, less than 5% network infrastructure is 100% dedicated to the LHCONE service.

Challenges

In the ESnet implementation which is predominantly Juniper today, all VRF routes are common to the L3VPN table and by default are all routed using the same metrics so routes to the same region tend to have the same defined next hop even across different VRFs. This makes it challenging to implement a dynamic **traffic separation** scheme on a per VRF basis. A network with many VRF implementations may find this essential.

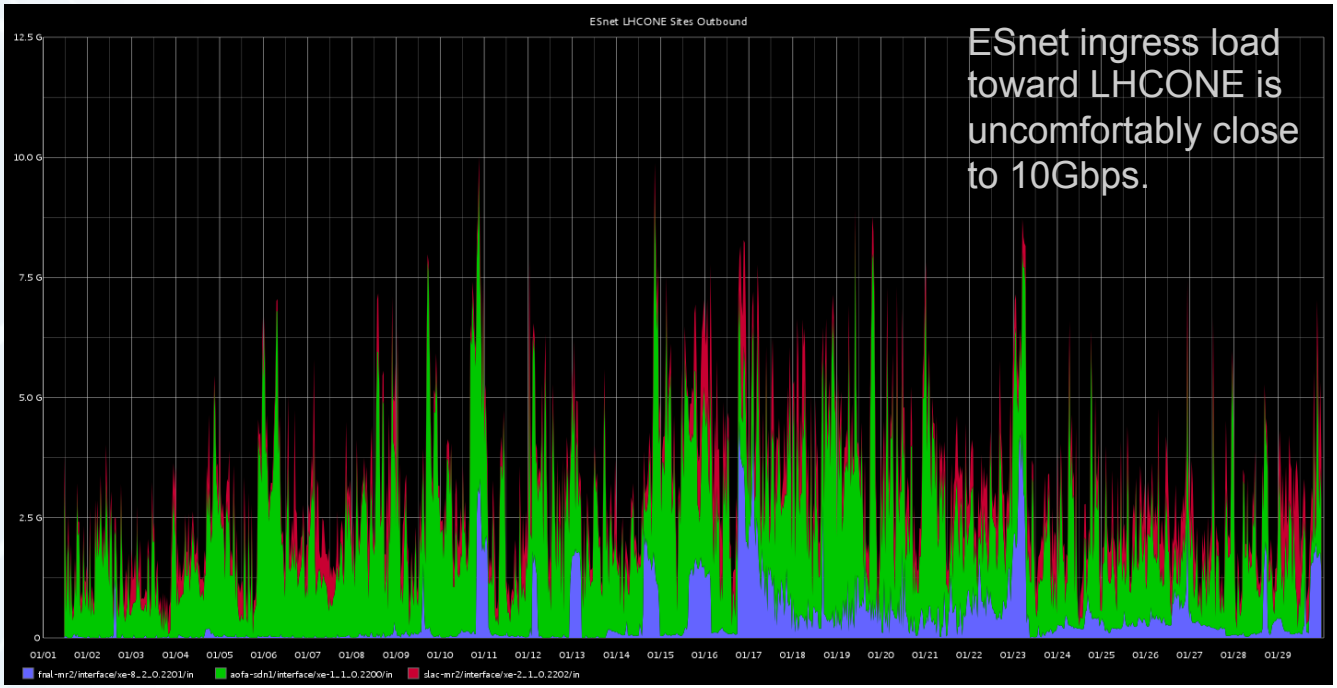
Since flows within a particular NSP network are tagged LSPs, routed dynamically using IGP metrics, obtaining **traffic statistics** within an autonomous system can also be a challenge. Currently ESnet LHCONE specific statistics are collected at the edges of our network.

Since VRF reachability information needs to be exchanged and stored throughout the hosting network's routing infrastructure, certain configuration changes are **service impacting** due to the addition of routing tables such as family inet-vpn.

ESnet traffic into LHCONE Jan. 2013

In the past six months, the LHCONE routing table has doubled in size to include 33 autonomous systems serving 115 uniquely routed prefixes.

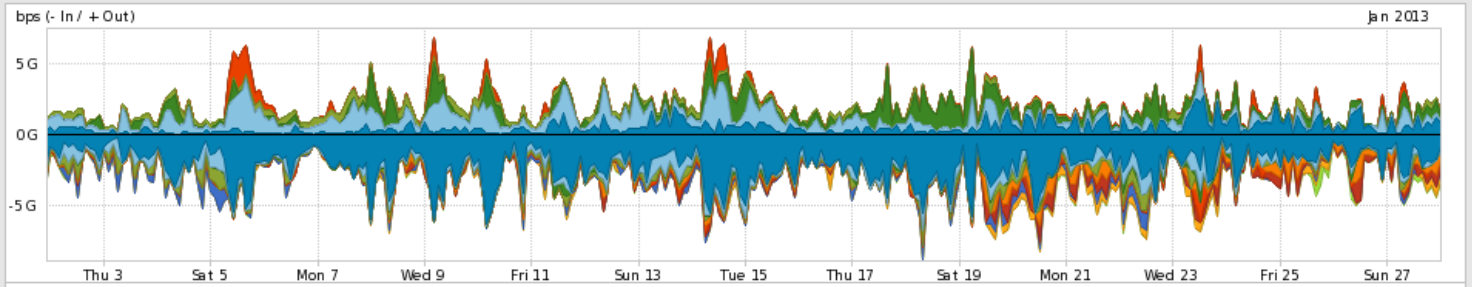
12.5Gbps Scale



ESnet ingress load toward LHCONE is uncomfortably close to 10Gbps.

BNL, FNAL, SLAC

ESnet LHCONE MANLAN Flow Report Jan. 2013

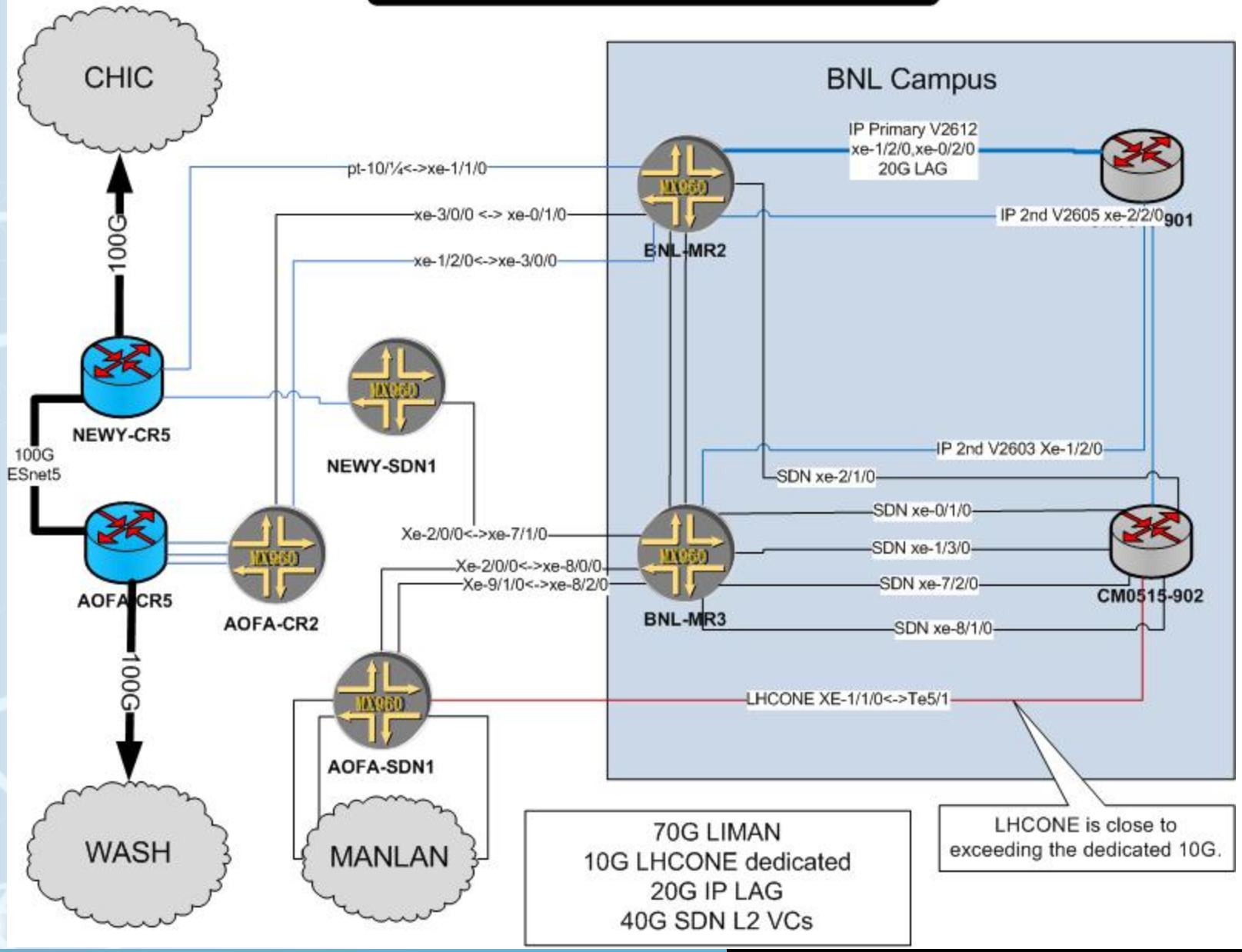


Select All | Clear All | Update

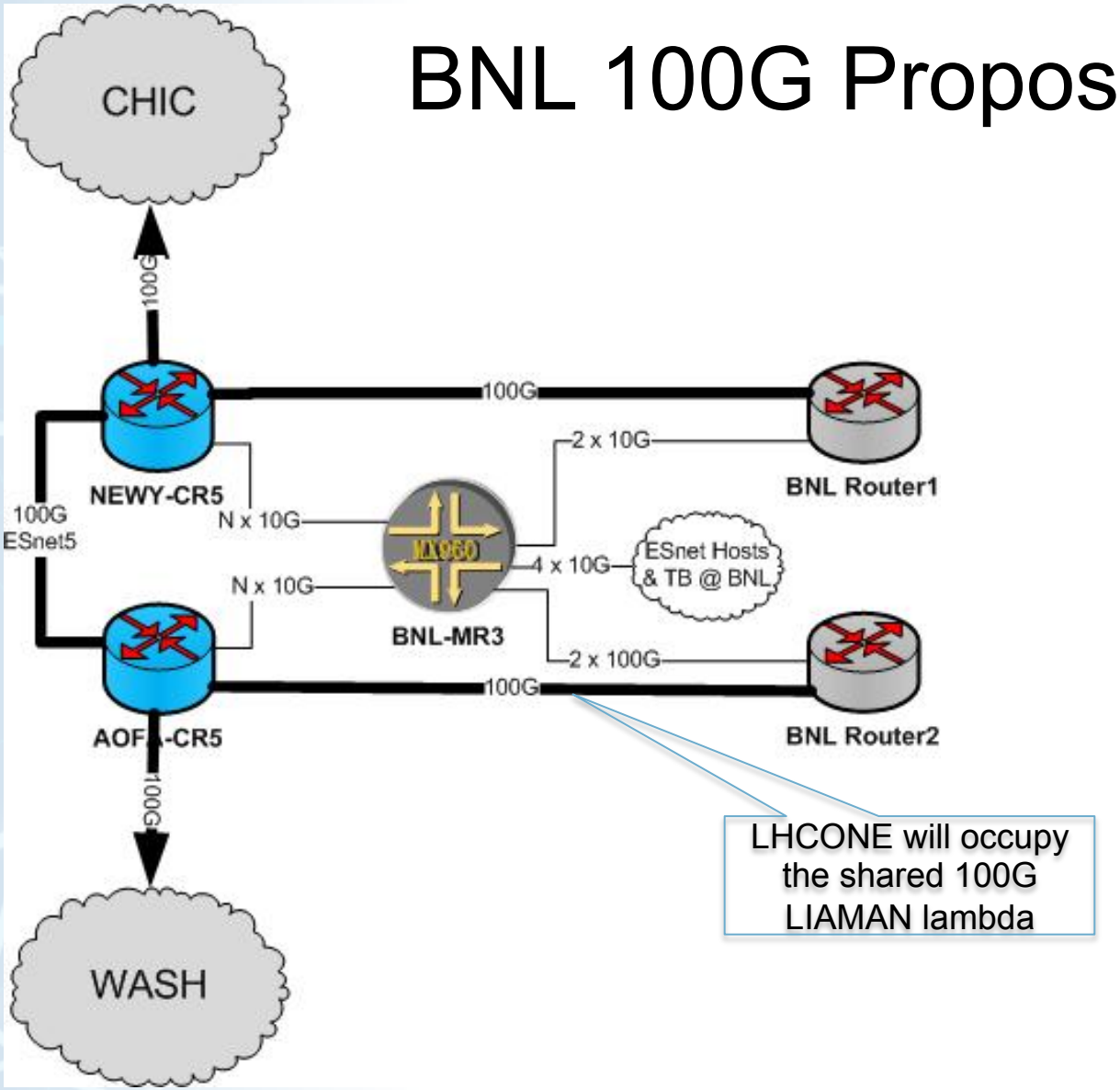
Average | Max | PCT95

Interface	As	ASN	In	Out	Total (In + Out)
aofa-sdn1.es.net xe-1/1/0.2200-BNL *aofa-sdn1->bnl-site-lhcone:10ge::show:intercloud*	BNL	43	2.15 Gbps	516.20 Mbps	2.67 Gbps
aofa-sdn1.es.net xe-7/1/0.2001-CANARIE-NTN *aofa-sdn1->internet2-lhcone:10ge::show:intercloud*	CANARIE-NTN	6509	349.43 Mbps	711.86 Mbps	1.06 Gbps
aofa-sdn1.es.net xe-7/1/0.2004-ASGARR *aofa-sdn1->geant-lhcone:10ge::show:intercloud*	ASGARR	137	64.39 Mbps	513.02 Mbps	577.41 Mbps
aofa-sdn1.es.net xe-7/1/0.2001-BCNET *aofa-sdn1->internet2-lhcone:10ge::show:intercloud*	BCNET	271	211.89 Mbps	247.97 Mbps	459.86 Mbps
aofa-sdn1.es.net xe-7/1/0.2004-DFN-IP *aofa-sdn1->geant-lhcone:10ge::show:intercloud*	DFN-IP	680	182.12 Mbps	0.00 bps	182.12 Mbps
aofa-sdn1.es.net xe-7/1/0.2001-UTORONTO *aofa-sdn1->internet2-lhcone:10ge::show:intercloud*	UTORONTO	239	40.41 Mbps	134.95 Mbps	175.36 Mbps
aofa-sdn1.es.net xe-7/1/0.2004-DESY-HAMBURG *aofa-sdn1->geant-lhcone:10ge::show:intercloud*	DESY-HAMBURG	1754	171.22 Mbps	0.00 bps	171.22 Mbps
aofa-sdn1.es.net xe-7/1/0.2001-TRIUMF *aofa-sdn1->internet2-lhcone:10ge::show:intercloud*	TRIUMF	36391	168.12 Mbps	0.00 bps	168.12 Mbps
aofa-sdn1.es.net xe-7/1/0.2011-CERN *aofa-sdn1->cern-lhcone:10ge::show:intercloud*	CERN	513	124.99 Mbps	0.00 bps	124.99 Mbps
aofa-sdn1.es.net xe-7/1/0.2004-IN2P3 *aofa-sdn1->geant-lhcone:10ge::show:intercloud*	IN2P3	789	31.97 Mbps	5.06 Kbps	31.98 Mbps
aofa-sdn1.es.net xe-7/1/0.2001-U-ALBERTA	U-ALBERTA	3359	383.12 Kbps	14.58 Mbps	14.96 Mbps

BNL Connections to NY POPs
Jan. 11, 2013



BNL 100G Proposal



The 100G Advantage

Overlay and other forms of virtual networking are only viable within the context of abundant bandwidth in the network substrate.

Deploying dedicated 10G lambdas has been a cost effective approach in certain cases, an example of which is BNL to MANLAN dedicated 10G lambda over the dark fiber LIMAN ring in the Esnet.

However, as the LHCONE network utilization continues to increase, these dedicated portions will need to be upgraded or risk becoming bottlenecks. As one of the first steps in deploying their 100G service, BNL intends to migrate their LHCONE connectivity off of the dedicated 10G and on to their common 100G service, effectively increasing the optimal LHCONE transport to MANLAN by approximately one order of magnitude.

Long Term Virtual Viability

How long will the LHCONE overlay fit in a 10G network substrate? The two US LHC tier 1 sites are already driving their LHCONE 10G connections over 50% on a regular basis.

When the Esnet was upgraded to a 100G core the virtualized portions of the ESnet LHCONE L3 overlay gained room to grow beyond 10G.

One of the first steps in the BNL 100G deployment will be to migrate off of the dedicated 10G circuit to MANLAN and on to a common shared 100G infrastructure.

A 100G network substrate will be essential for deploying various kinds of virtualized networks to address the needs of the growing number of distributed scientific collaborations world-wide.

Summary

The LHCONE today deploys dedicated resources where necessary and shared resources everywhere else.

Typically less that 5% dedicated infrastructure overall.

The LHCONE is growing rapidly, routing table has doubled in size over the last six months.

The common general R&E Internet operational model easily translates quickly and to supporting virtualized overlay networks.

100 Gbps is a tremendous advantage, eliminates the headroom constraints, paves the way for advanced virtual networks of the future, and increases the lifespan of the LHCONE as it continues to evolve.

Questions?

Michael O'Connor

ESnet Network Engineer

moc@es.net

631 344-7410