



# LHC Open Network Environment *LHCONE*

**Artur Barczyk**

**Caltech**

**David Foster**

**CERN**

**April, 2011**





# INTRODUCTION





# Overview

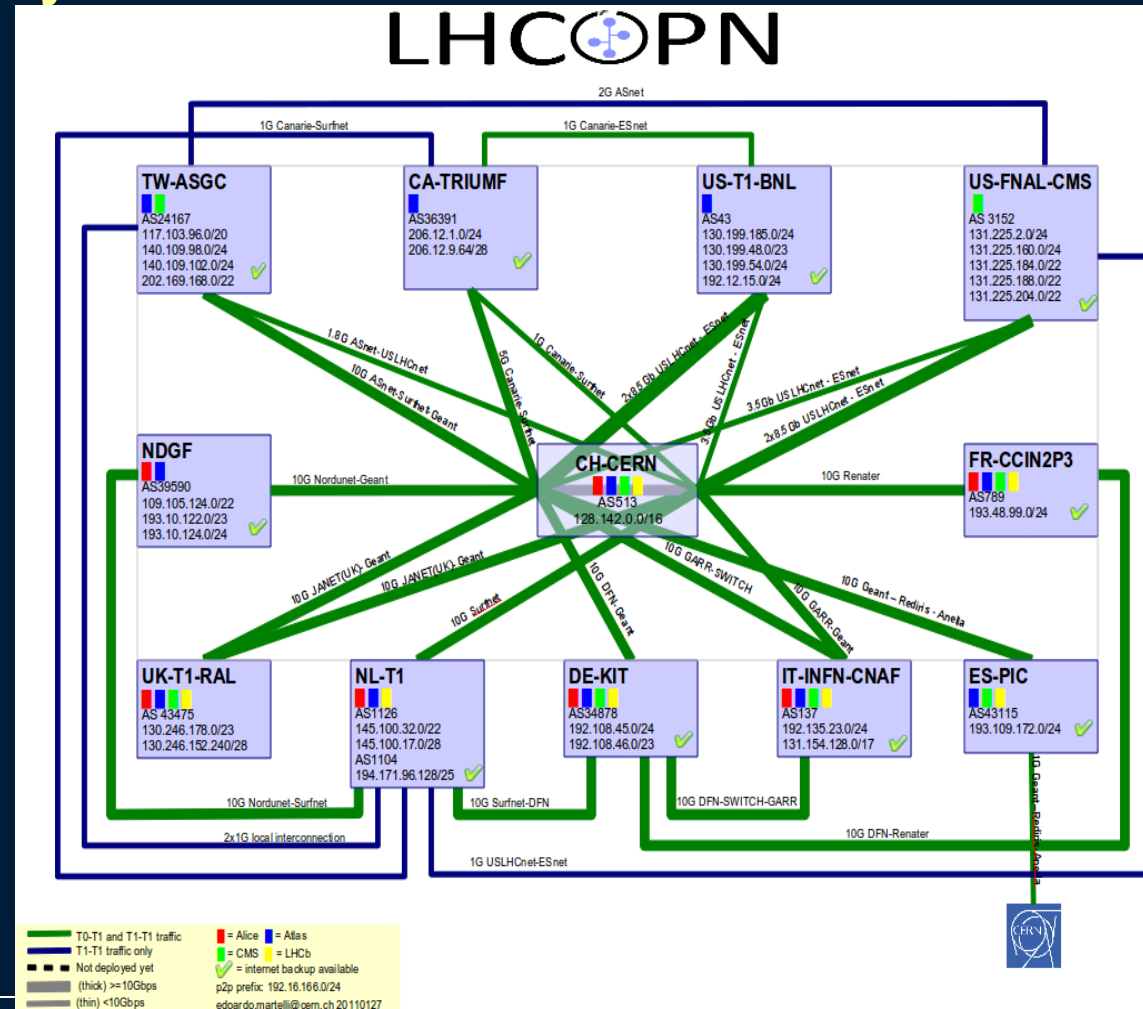
- **Started with Workshop on Transatlantic Connectivity for LHC experiments**
  - June 2010 @ CERN
- **Same time as changes in the computing models were being discussed in the LHC experiments**
- **Experiments provided a requirements document (Oct 2010)**
  - Tasked LHCOPN with providing a proposal
- **LHCT2S group was formed from within the LHCOPN**
- **LHCT2S Meeting in Geneva in January 2011**
  - Discussion of 4 proposals, led to formation of a small working group drafting an architectural proposal based on these 4 documents
- **LHCOPN Meeting in Lyon in February 2011**
  - Draft architecture approved, finalised as “v2.2”





# The LHCOPN

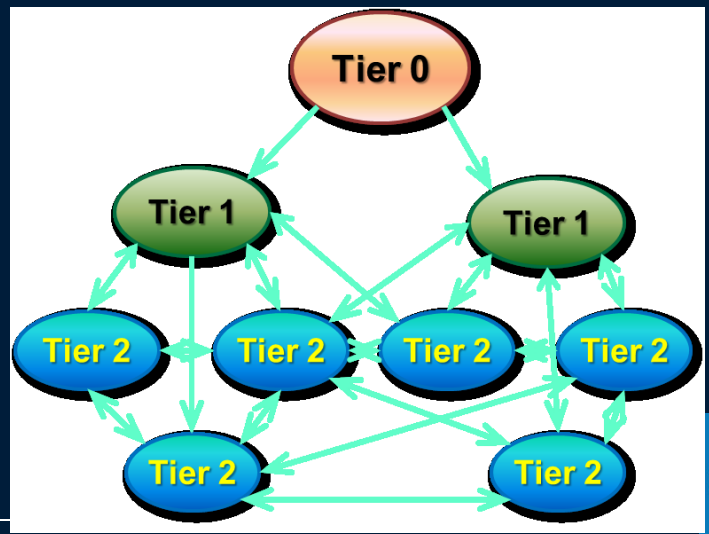
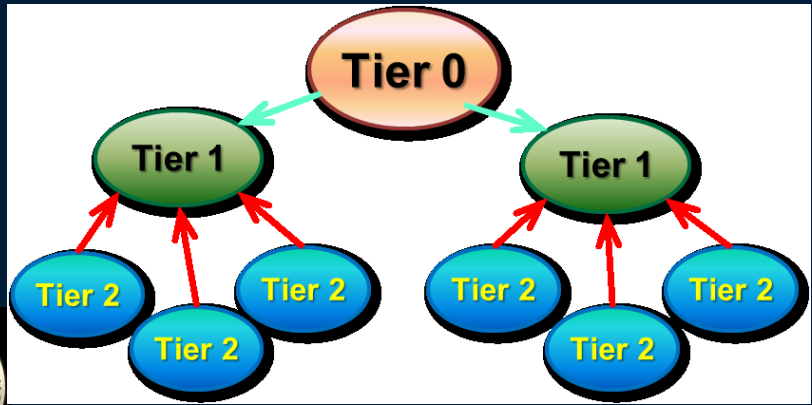
- **Dedicated network resources for Tier0 and Tier1 data movement**
- **130 Gbps total Tier0-Tier1 capacity**
- **Simple architecture**
  - Point-to-point Layer 2 circuits
  - Flexible and scalable topology
- **Grew organically**
  - From star to partial mesh
  - Open to technology choices
    - have to satisfy requirements
- **Federated governance model**
  - Coordination between stakeholders
  - No single administrative body required





# LHCONE architecture to match the computing models

- 3 recurring themes:
  - **Flat(ter) hierarchy**: Any site can use any other site as source of data
  - **Dynamic data caching**: Analysis sites will pull datasets from other sites “on demand”, including from Tier2s in other regions
    - Possibly in combination with strategic pre-placement of data sets
  - **Remote data access**: jobs executing locally, using data cached at a remote site in quasi-real time
    - Possibly in combination with local caching
- **Expect variations by experiment**





***LHCONE***

**HTTP://LHCONE.NET**

**The requirements, architecture, services**





# Requirements summary (from the LHC experiments)

---

- **Bandwidth:**

- Ranging from 1 Gbps (Minimal site) to 5-10Gbps (Nominal) to N x 10 Gbps (Leadership)
- No need for full-mesh @ full-rate, but several full-rate connections between Leadership sites
- Scalability is important,
  - sites are expected to migrate **Minimal → Nominal → Leadership**
  - Bandwidth growth: Minimal = 2x/yr, Nominal&Leadership = 2x/2yr

- **Connectivity:**

- Facilitate good connectivity to so far (network-wise) under-served sites

- **Flexibility:**

- Should be able to include or remove sites at any time

- **Budget Considerations:**

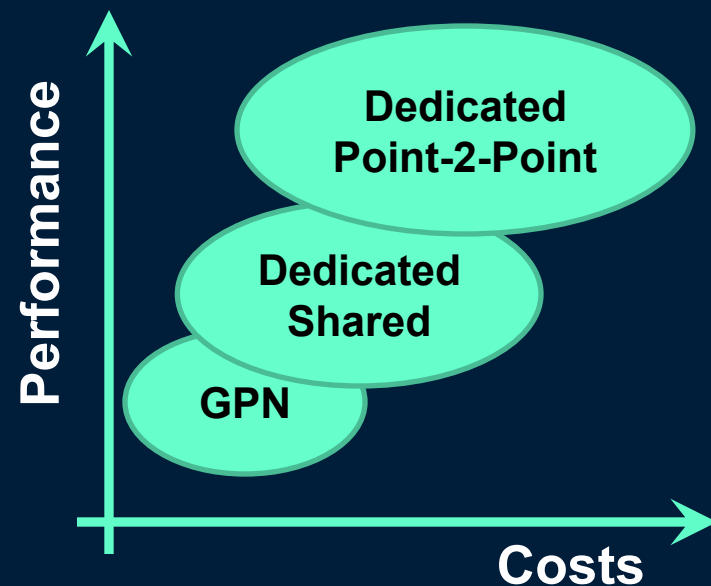
Costs have to be understood, solution needs to be affordable





# Design Considerations

- So far, T1-T2, T2-T2, and T3 data movements have been using **General Purpose Network infrastructure**
  - Shared resources (with other science fields)
  - Mostly best effort service
- **Increased reliance on network performance** → need more than best effort service
  - Dedicated resources
- **Collaboration on global scale, diverse environment, many parties**
  - Solution to be **open, neutral** and **diverse**
  - Scalable in bandwidth, extent and scope
  - ⇒ Open Exchange Points
- **Choose the most cost effective solution**



**Organic activity, growing over time according to needs**







# LHCONE Architecture

- **Builds on the Hybrid network infrastructures and Open Exchanges**
  - As provided today by the major R&E networks on all continents
  - To build a global unified service platform for the LHC community
- **Make best use of the technologies and best current practices and facilities**
  - As provided today in national, regional and international R&E networks
- **LHCONE's architecture incorporates the following building blocks**
  - Single node **Open Exchange Points**
  - Continental / regional **Distributed Open Exchange Points**
  - **Interconnect Circuits** between exchange points
- **Continental and Regional Exchange Points are likely to be built as distributed infrastructures with access points located around the region, in ways that facilitate access by the LHC community**
  - Likely to be connected by allocated bandwidth on various (possibly shared) links to form LHCONE





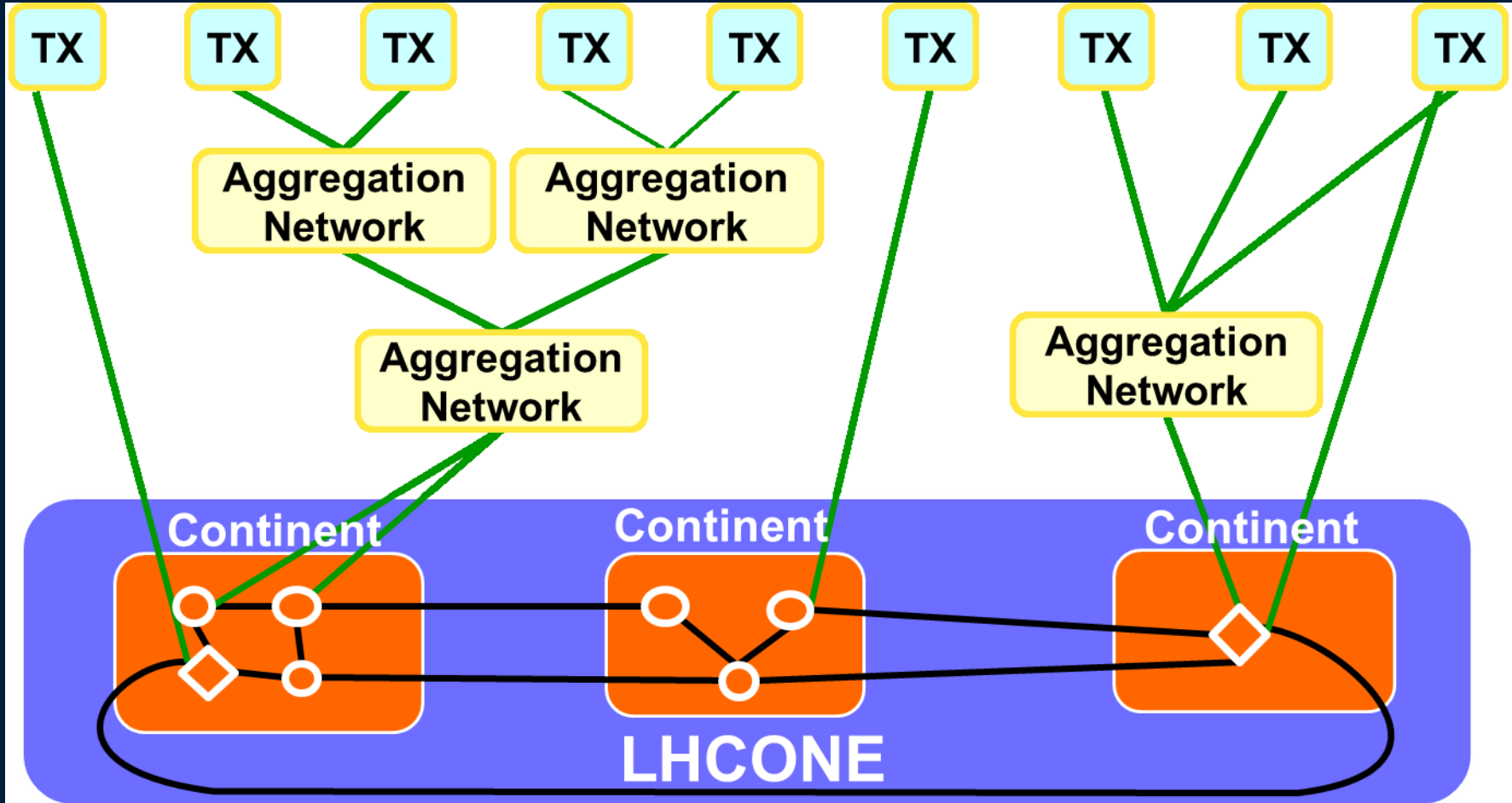
# LHCONE Access Methods

- **Choosing the access method to LHCONE, among the viable alternatives, is up to the end-site (a Tier1, 2 or 3), in cooperation with site and/or regional network**
- **Alternatives may include**
  - Static or dynamic circuits
  - Dynamic circuits with guaranteed bandwidth
  - Fixed lightpath(s)
  - Connectivity at Layer 3, where appropriate and compatible with the general purpose traffic
- **Tier-1/2/3s may connect to LHCONE through aggregation networks**





# High-level Architecture



Single node Exchange Point

Distributed Exchange Point





# LHCONE Network Services

Offered to Tier1s, Tier2s and Tier3s

- **Shared Layer 2 domains: separation from non-LHC traffic**
  - IPv4 and IPv6 addresses on shared layer 2 domain including all connectors
  - Private shared layer 2 domains for groups of connectors
  - Layer 3 routing is up to the connectors
    - A Route Server per continent is planned to be available
- **Point-to-point layer 2 connections: per-channel traffic separation**
  - VLANs without bandwidth guarantees between pairs of connectors
- **Lightpath / dynamic circuits with bandwidth guarantees**
  - Lightpaths can be set up between pairs of connectors
  - Circuit management: DICE IDC & GLIF Fenius now, OGF NSI when ready
- **Monitoring: perfSONAR archive now, OGF NMC based when ready**
  - Presented statistics: current and historical bandwidth utilization, and link availability statistics for any past period of time
- **This list of services is a starting point and not necessarily exclusive**

**LHCONE** does not preclude continued use of the general R&E network infrastructure by the Tier1s, Tier2s and Tier3s - where appropriate





# LHCONE Policy Summary

- It is expected that LHCONE policy will be defined and may evolve over time in accordance with the governance model
- **Policy Recommended for LHCONE governance**
  - Any Tier1/2/3 can connect to **LHCONE**
  - Within **LHCONE**, transit is provided to anyone in the Tier1/2/3 community that is part of the **LHCONE** environment
  - Exchange points must carry all LHC traffic offered to them (and only LHC traffic), and be built in carrier-neutral facilities so that any connector can connect with its own fiber or using circuits provided by any telecom provider
  - Distributed exchange points: same as above + the interconnecting circuits must carry all the LHC traffic offered to them
  - **No additional restrictions can be imposed on LHCONE by the LHCONE component contributors**
- **The Policy applies to LHCONE components, which might be switches installed at the Open Exchange Points, or virtual switch instances, and/or (virtual) circuits interconnecting them**

Details at <http://lhcone.net>





# LHCONE Governance Summary

- **Governance is proposed to be similar to the LHCOPN, since like the LHCOPN, LHCONE is a community effort**
  - Where all the stakeholders meet regularly to review the operational status, propose new services and support models, tackle issues, and design, agree on, and implement improvements
- **Includes connectors, exchange point operators, CERN, and the experiments, in a form to be determined.**
- **Defines the policies of LHCONE and requirements for participation**
  - It does not govern the individual participants
- **Is responsible for defining how costs are shared**
- **Is responsible for defining how resources of LHCONE are allocated**

Details at <http://lhcone.net>





# THE LHCONE IMPLEMENTATION





# Starting Point

- **Based on CMS and Atlas use case document**
  - Tier1/2/3 sites, spanning 3 continents
  - **Measurable success criteria**
    - Improved analysis (to be quantified by the experiments)
    - Sonar/Hammercloud tests between LHCONE sites
- **Use currently existing infrastructure as much as possible**
  - Open Exchange Points
  - Deployed bandwidth, allocated for LHCONE where possible
- **Build out according to needs, experience and changes in requirements**
- **Need close collaboration between the experiments and the LHCOPN community!**

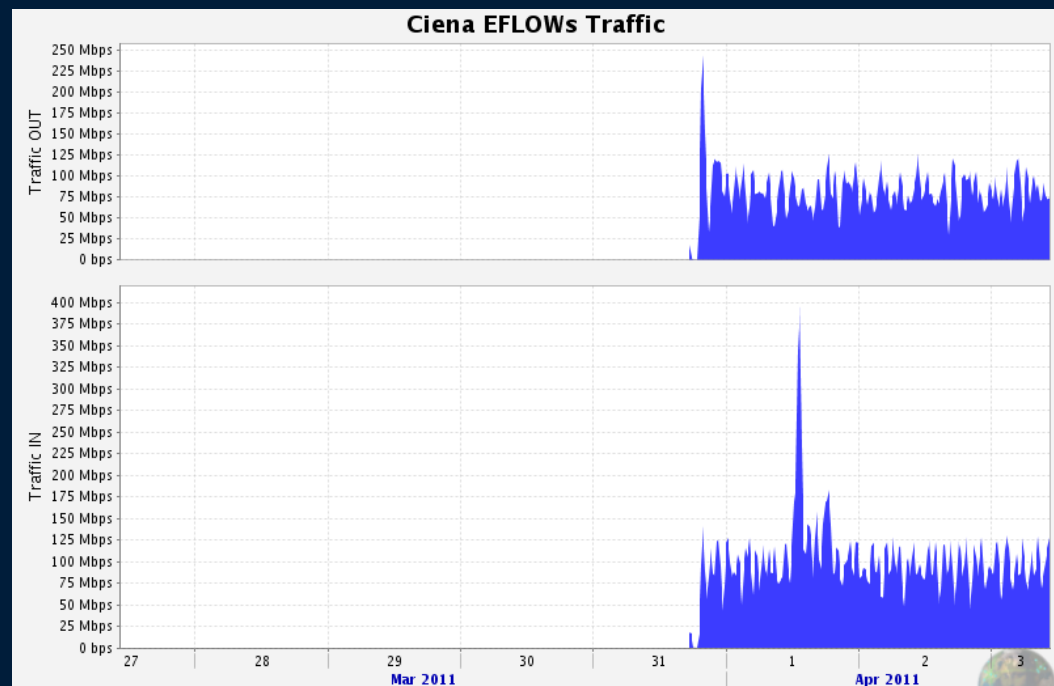






# Status Today

- **LHCONE launched on March 31<sup>st</sup>**
  - Started shared VLAN service including two initial sites: CERN, Caltech
  - Verified connectivity and routing
  - First data exchanged on March 31<sup>st</sup>
- **Active Open Exchange Points:**
  - CERNLight, NetherLight, StarLight (and MANLAN)
- **Route server for IP-layer connectivity installed and operational at CERN**
  - End-site's border router peers with the route server, but data path is direct between the sites – no data flow through route server!



**Monitoring of Transatlantic segment (US LHCNet)**





# What's next?

- **Organic growth** - LHCONE is open to participation
- **We will be working with Atlas and CMS operations teams and LHC sites on including next sites**
  - In collaboration with regional and national networks
- **Next step is to arrive at significant build-out by Summer 2011**
  - Connect the right mix of T1/T2/T3 sites (per experiment)
  - Include several geographical regions
  - Achieve measurable effect on operations and analysis potential
  - (Continuous) feedback between experiments and networks





# How to connect your site

- Procedure and details vary by site
- Generic steps:
  - Get Layer 2 connection to an open exchange point
    - Shared or dedicated bandwidth
    - Exchange point already in LHCONE?
      - YES: just ask to have your connection included in vlan 3000
      - NO: need to work out core connectivity
  - Set up routing at your site
    - Get your IP (v4 and/or v6) from Edoardo Martelli, IT/CS
    - Peer with the route server
      - Currently one installed at CERN, more to come
    - Alternative: set up peerings only with the remote sites you choose
    - Announce only the relevant subnet (e.g. the SEs)
    - More details and contacts on the LHCONE twiki (see <http://lhcone.net>)





# Summary

- **LHCONE** is a robust and scalable solution for a global system serving LHC's Tier1, Tier2 and Tier3 sites' needs
  - Fits the new computing models
  - Based on a **switched core with routed edge** architecture
  - IP routing is implemented at the end-sites
- Core consists of sufficient number of strategically placed **Open Exchange Points** interconnected by properly sized trunks
  - Scaling rapidly with time as in requirements document
- Initial deployment to use predominantly static configuration (shared VLAN & Lightpaths),
  - later predominantly using dynamic resource allocation
- Seed implementation interconnecting an initial set of sites has started

**Organic growth; we're ready to connect sites!**





**THANK YOU!**

**<http://lhcone.net>**

**[Artur.Barczyk@cern.ch](mailto:Artur.Barczyk@cern.ch)**

**[David.Foster@cern.ch](mailto:David.Foster@cern.ch)**

