LHCONE status and future

Alice workshop

Tsukuba, 7th March 2014

Edoardo.Martelli@cern.ch

Summary

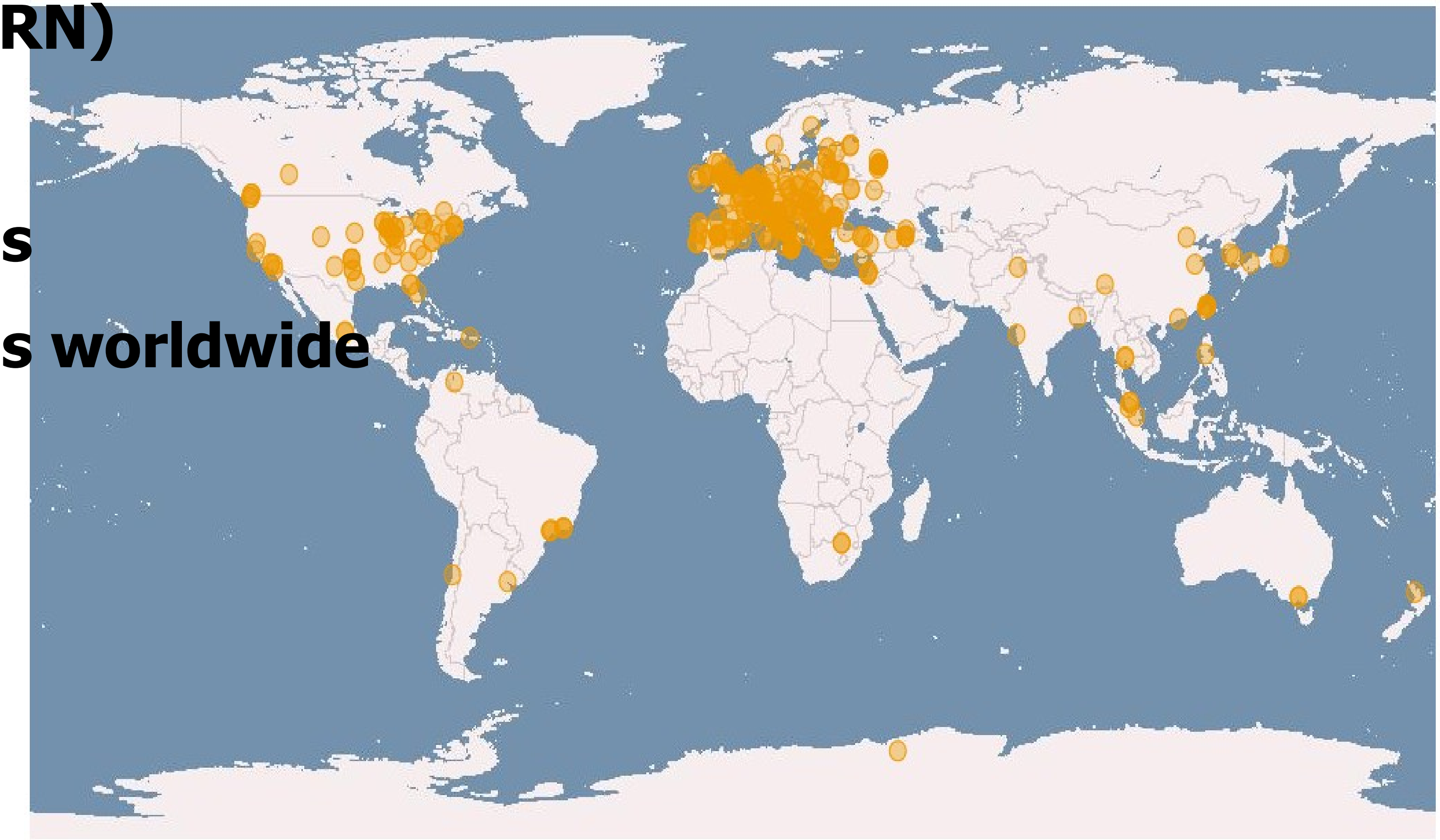
- Networking for WLCG
- LHCOPN
- LHCONE
  - services
  - how to join
- LHCONE in Asia

Networking for WLCG

Worldwide LHC Computing Grid

WLCG sites:

- 1 Tier0 (CERN)
- 13 Tier1s
- ~170 Tier2s
- >300 Tier3s worldwide



WLCG
Worldwide LHC Computing Grid

Planning for Run2

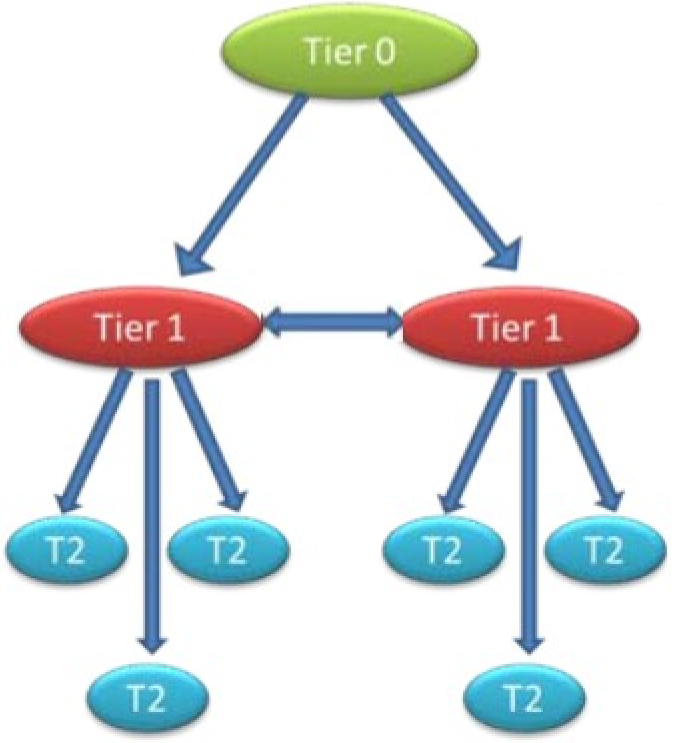
“The Network infrastructure is the most reliable service we have”

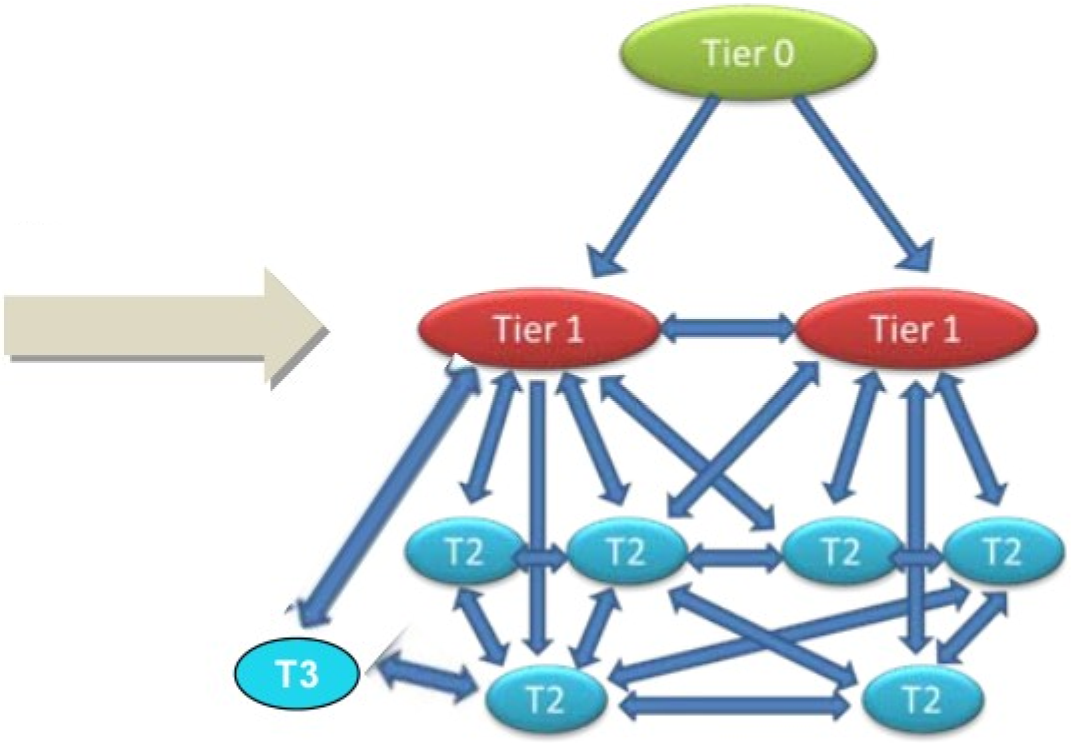
“Network Bandwidth (rather than disk) will need to scale more with users and data volume”

“Data placement will be driven by demand for analysis and not pre-placement”

Ian Bird, WLCG project leader

Computing model evolution





Original MONARCH model

Model evolution

Technology Trends

- Commodity Servers with 10G NICs
- High-end Servers with 40G NICs
- 40G and 100G interfaces on switches and routers

Needs for 100Gbps backbones to host large data flows >10Gbps and soon >40Gbps

Role of Networks in WLCG

Computer Networks even more essential component of WLCG

Data analysis in Run 2 will need more network bandwidth between any pair of sites

LHCOPN
LHC Optical Private Network

What LHCOPN is:

Private network connecting **Tier0 and Tier1s**

Reserved to LHC data transfers and analysis

Dedicated large bandwidth links

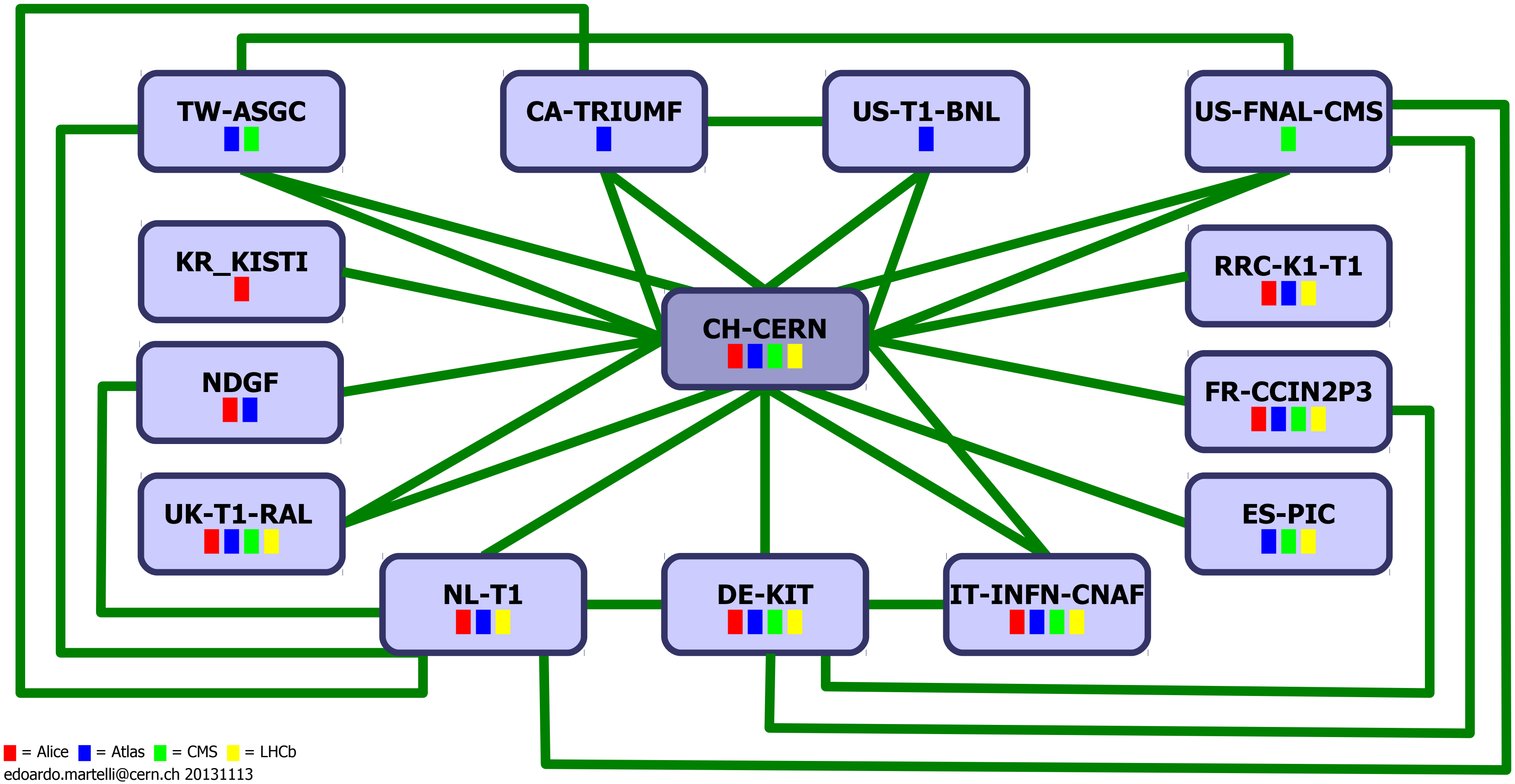
Highly resilient architecture

A collaborative effort

Layer3: Designed, built and operated by the Tier0-Tier1s community

Layer1-2: Links provided by Research and Education network providers: Asnet, ASGCnet, Canarie, DFN, Esnet, GARR, Geant, JANET, Kreonet, Nordunet, Rediris, Renater, Surfnet, SWITCH, TWAREN, USLHCnet

Topology



🟥 = Alice 🟦 = Atlas 🟩 = CMS 🟨 = LHCb

edoardo.martelli@cern.ch 20131113

Technology

- Single and bundled long distance 10G Ethernet links
- Multiple redundant paths. Star and Partial-Mesh topology
- BGP routing: communities for traffic engineering, load balancing.
- Security: only declared IP prefixes can exchange traffic.

LHCOPN future

- The LHCOPN will be kept as the main network to exchange data among the Tier0 and Tier1s

- Links to the Tier0 may be soon upgraded to multiple 10Gbps (waiting for Run2 to see the real needs)

LHCONE
LHC Open Network Environment

New computing model impact

- Better and more dynamic use of storage
- Reduced load on the Tier1s for data serving
- Increased speed to populate analysis facilities

Needs for a faster, predictable, pervasive network connecting Tier1s and Tier2s

Requirements from the Experiments

- Connecting any pair of sites, regardless of the continent they reside

- Site's bandwidth ranging from 1Gbps (Minimal), 10Gbps (Nominal), to 100G (Leadership)

- Scalability: sites are expected to grow

- Flexibility: sites may join and leave at any time

- Predictable cost: well defined cost, and not too high

LHCONE concepts

- Serving any LHC sites according to their needs and allowing them to grow

- Sharing the cost and use of expensive resources

- A collaborative effort among Research & Education Network Providers

- Traffic separation: no clash with other data transfer, resource allocated for and funded by the HEP community

Governance

LHCONE is a community effort.

All stakeholders involved: TierXs, Network Operators, LHC Experiments, CERN.

LHCONE services

L3VPN (VRF): routed Virtual Private Network - *operational*

P2P: dedicated, bandwidth guaranteed, point-to-point links - *development*

perfSONAR: monitoring infrastructure

LHCONE L3VPN

What LHCONE L3VPN is:

Layer3 (routed) Virtual Private Network

Dedicated worldwide backbone connecting **Tier1s, T2s and T3s** at high bandwidth

Reserved to LHC data transfers and analysis

Advantages

Bandwidth dedicated to LHC data analysis, no contention with other research projects

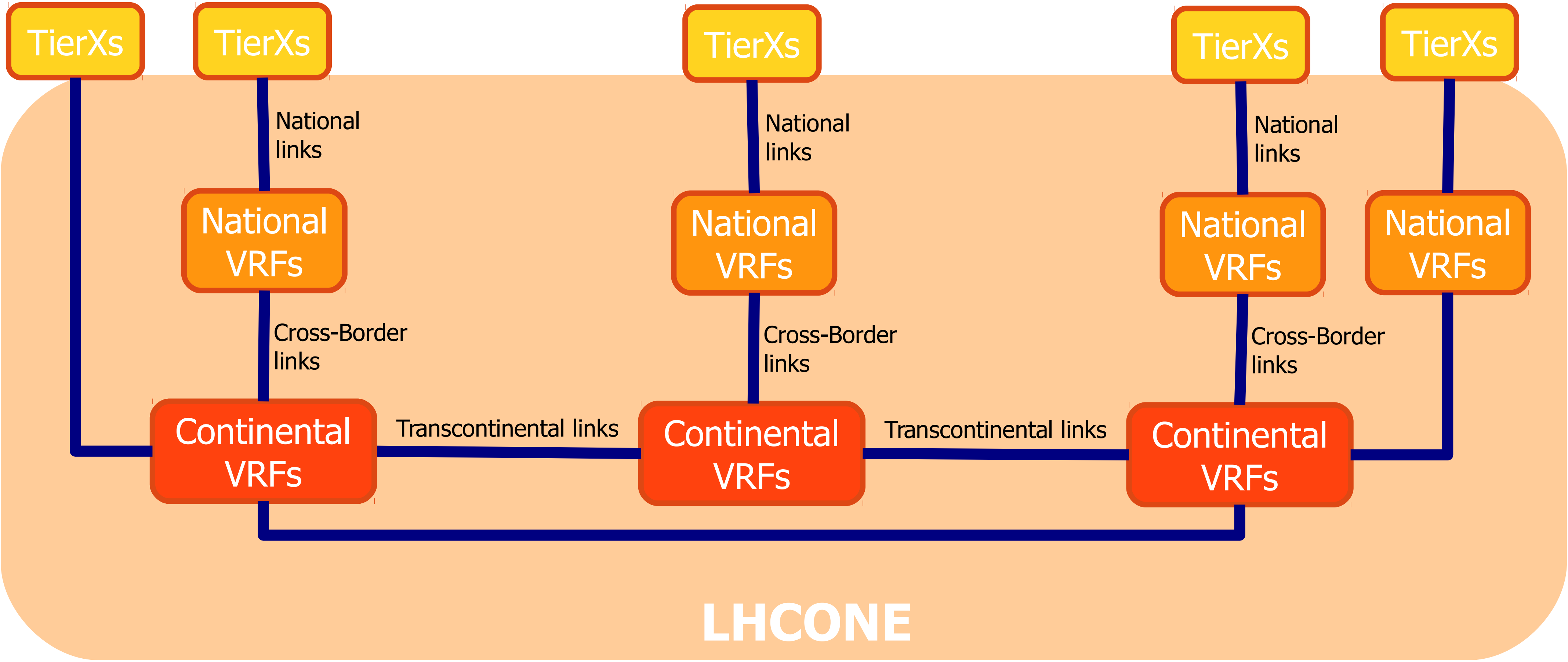
Well defined cost tag for WLCG networking

Trusted traffic that can bypass firewalls

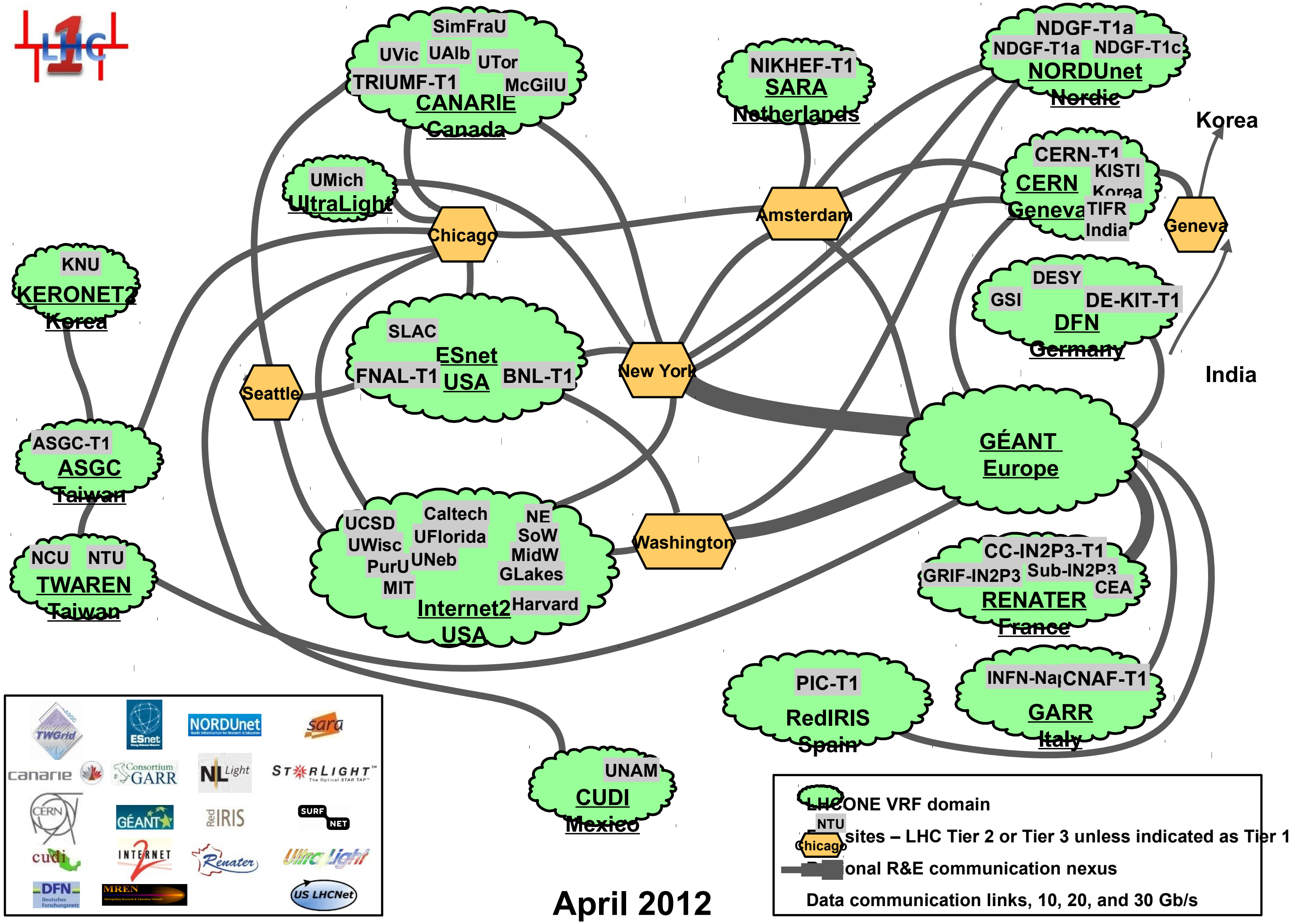
LHCONE L3VPN architecture

- TierX sites connected to National-VRFs or Continental-VRFs
- National-VRFs interconnected via Continental-VRFs
- Continental-VRFs interconnected by trans-continental/trans-oceanic links

Acronyms: **VRF** = Virtual Routing Forwarding (virtual routing instance)



Current L3VPN topology



LHCONE VRF domain

sites – LHC Tier 2 or Tier 3 unless indicated as Tier 1

onal R&E communication nexus

Data communication links, 10, 20, and 30 Gb/s

See http://lhcone.net for details.

April 2012

credits: Joe Metzger, ESnet

Status

Over 15 national and international Research Networks

Several Open Exchange Points including NetherLight, StarLight, MANLAN, CERNlight and others

Trans-Atlantic connectivity provided by ACE, GEANT, NORDUNET and USLHCNET

~50 end sites connected to LHCONE:
- 8 Tier1s
- 40 Tier2s

Credits: Mian Usman, Dante
More Information: https://indico.cern.ch/event/269840/contribution/4/material/slides/0.ppt

Operations

Usual Service Provider operational model: a TierX must refer to the VRF providing the local connectivity

Bi-weekly call among all the VRF operators and concerned TierXs

How to join the L3VPN

Pre-requisites

The TierX site needs to have:

- Public IP addresses

- A public Autonomous System (AS) number

- A BGP capable router

How to connect

The TierX has to:

- Contact the Network Provider that runs the closest LHCONE VRF

- Agree on the cost of the access

- Lease a link from the TierX site to the closest LHCONE VRF PoP (Point of Presence)

- Configure the BGP peering with the Network Provider

TierX routing setup

- The TierX announce only the IP subnets used for WLCG servers

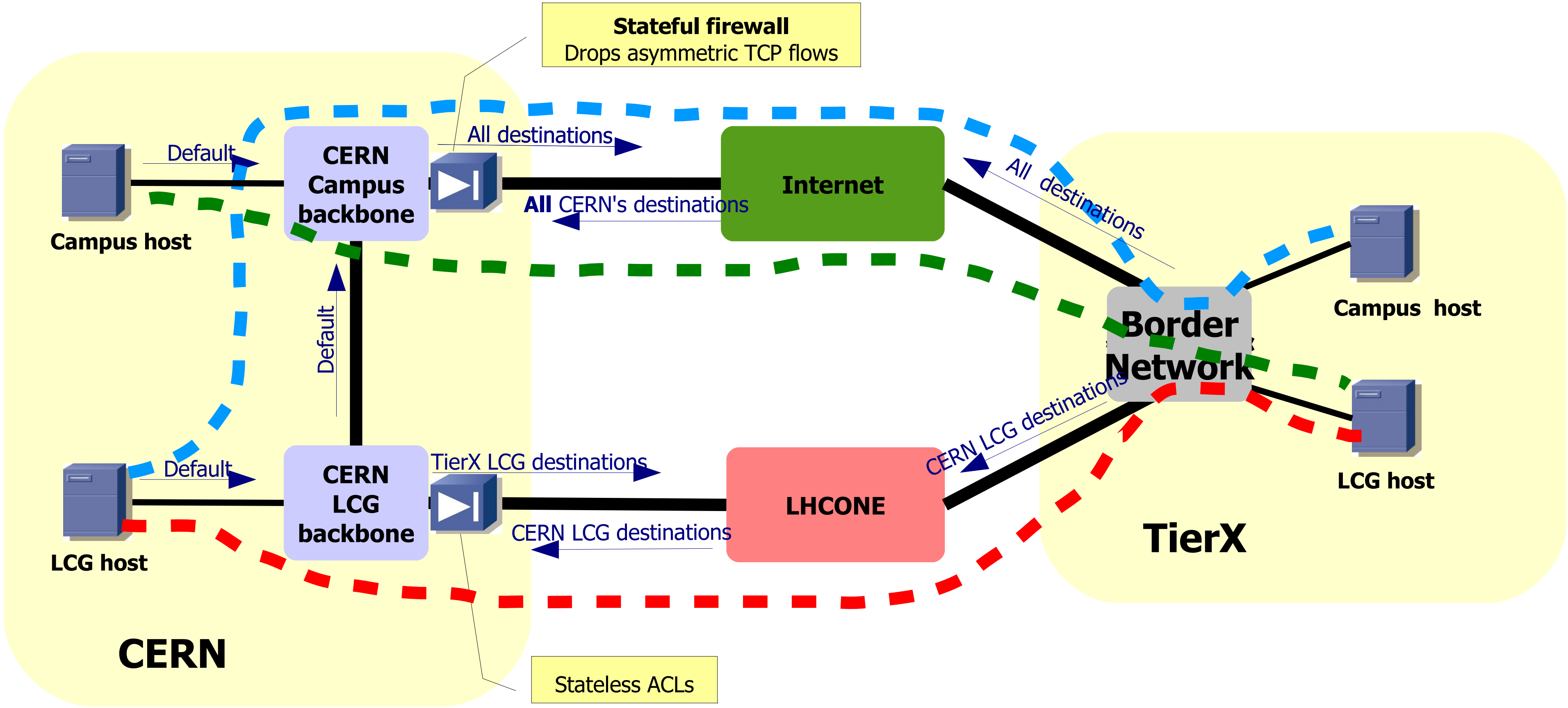
- The TierX accepts all the prefixes announced by the LHCONE VRF

- The TierX **must** assure traffic symmetry: injects only packets sourced by the announced subnets

- LHCONE traffic may be allowed to bypass the central firewall (up to the TierX to decide)

Symmetric traffic is <u>essential</u>

Beware: statefull firewalls discard unidirectional TCP connections



🟥 LHCONE host to LHCONE host

🟦 CERN's LHCONE host to TierX not LHCONE host

🟩 CERN's not LHCONE host to TierX's LHCONE host

Symmetry setup

To achieve symmetry, a TierX can use one of the following techniques:

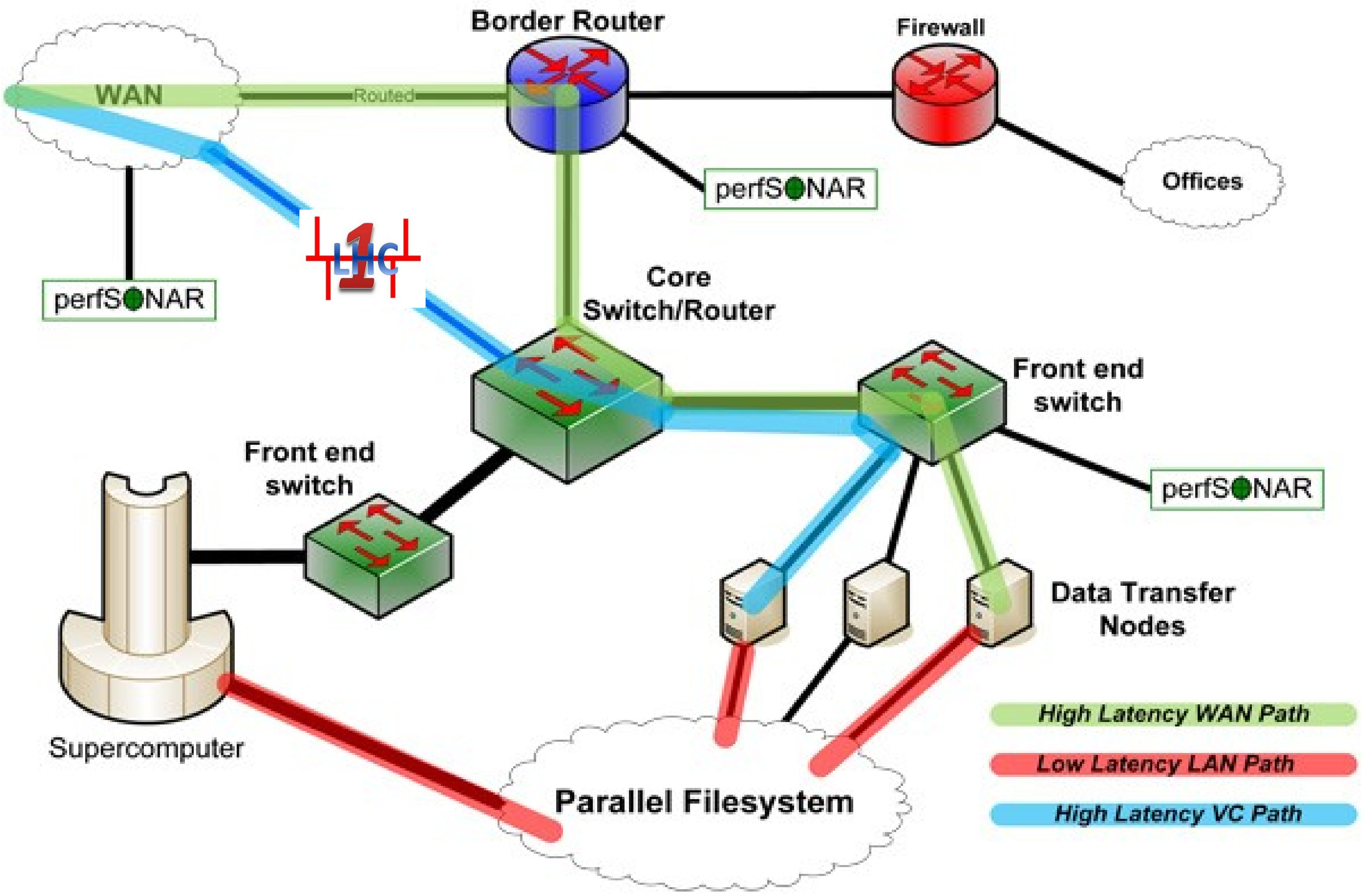
- Policy Base Routing (source-destination routing)

- Physically Separated networks

- Virtually separated networks (VRF)

- Scienze DMZ

Scienze DMZ



http://fasterdata.es.net/science-dmz/science-dmz-architecture/

LHCONE P2P
Guaranteed bandwidth point-to-point links

What LHCONE P2P is (will be):

On demand point-to-point (P2P) link system over a multi-domain network

Provides P2P links between any pair of TierX

Provides dedicated P2P links with guaranteed bandwidth (protected from any other traffic)

Accessible and configurable via software API

Status

Work in progress: still in design phase

Challenges:

- multi-domain provisioning system

- intra-TierX connectivity

- TierX-TierY routing

- interfaces with WLCG software

LHCONE perfSONAR

What LHCONE perfSONAR is

LHCONE Network monitoring infrastructure

Probe installed at the VRFs interconnecting points and at the TierXs

Accessible to any TierX for network healthiness checks

perfSONAR

- framework for active and passive network probing

- developed by Internet2, Esnet, Geant and others

- two interoperable flavors: perfSONAR-PS and perfSONAR-MDM

- WLCG recommended version: perfSONAR-PS

Status

Endorsed by WLCG to be a standard WLCG service

Probes already deployed in many TierXs.

Being deployed in the VRF networks

Full information:
https://twiki.cern.ch/twiki/bin/view/LCG/PerfsonarDeployment

LHCONE-L3VPN in Asia

Connectivity status

Only few sites connected to LHCONE-L3VPN in ASIA via ASGC or with direct link to the US or Europe

Connectivity between ASIA and North America not scarce, but transit to Europe may not be adequate

Un-coordinated effort

Working together

ASCG is willing to share the use of their links to North-America and Europe with other Asian TierXs

Anyone interested to connect to the Asian LHCONE or share their trans-continental links, please get in touch with us

Anyway: You have to tune!

TCP Throughput <= TCPWinSize / RTT

Tokyo-CERN RTT (Round Trip Time): 280 ms
Default Max TCPWinSize for Linux = 256KBytes (= 2.048Mbit)

Tokyo-CERN throughput <= 2.048Mb / 0.280sec = 7.31Mbps :-(

Remote TierXs must tune server and client TCP Kernel parameters to get decent throughput!

LHCONE evolution

LHCONE evolution

- VRFs have started upgrading internal links and links to TierXs to 100Gbps

- VRFs interconnecting links will be upgraded to 100Gbps. 100Gbps Transatlantic link being tested.

- Operations need to be improved, especially how to support a TierX in case of performance issue

- perfSONAR deployment will be boosted

LHCONE evolution

- LHCONE-P2P take off still uncertain

- LHCONE-L3VPN must be better developed in ASIA

Conclusions

Conclusions

- New Computing Models will relay even more on good and abundant network connectivity

- TierXs need to improve their network connectivity

- LHCONE-L3VPN is a viable solution already adopted by many Tier1/2s

More information

Last LHCONE workshop:
https://indico.cern.ch/event/289679/

LHCONE websites:
http://lhcone.net
https://twiki.cern.ch/twiki/bin/view/LHCONE/WebHome

Weekly audio conference:
Monday 14:30 GMT, alternating every second week
architecture and operations

Mailing lists:
lhcone-operations@cern.ch
lhcone-architecture@cern.ch

Questions?