Progress on routing

Utrecht - 21st September 2006

edoardo.martelli@cern.ch

Recent achievements

ASGC: backup link via Starlight-USLHCnet active and used.

BNL: primary and backup links active and used

Gridka: Karlsruhe-Geneva lightpath active and used for part of the traffic. BGP configuration just started.

Gridka-CNAF: Karlsruhe-Milano lightpath deployed but not configured for backup connectivity yet. Waiting for configuring BGP.

Recent achievements - continue

RAL: Didcot-Geneva lightpath provisioned but not configured yet.

SARA: First GN2 Amsterdam-Geneva lightpath provisioned but not used yet. Procedure to install a SURFnet's Nortel OME6500 (DWDM box) at CERN has started.

Jumbo frames: implemented in most of the LHCOPN's links and in the LCG network backbone at CERN. CERN's machines still using 1500B MTU; compatibility tests are ongoing.

Bandwidth

http://network-statistics.web.cern.ch/network-statistics//ext/?p=sc

IT - Service Challenges / LHCopn...

IT Department

CERN Home IT Home Phonebook Need Help?

IT Home > Network Monitoring > Network Statistics > Service Challenges / LHCopn Traffic Monitoring

External Network | Service Challenges | Campus Network

Find

All CERN

IT Department

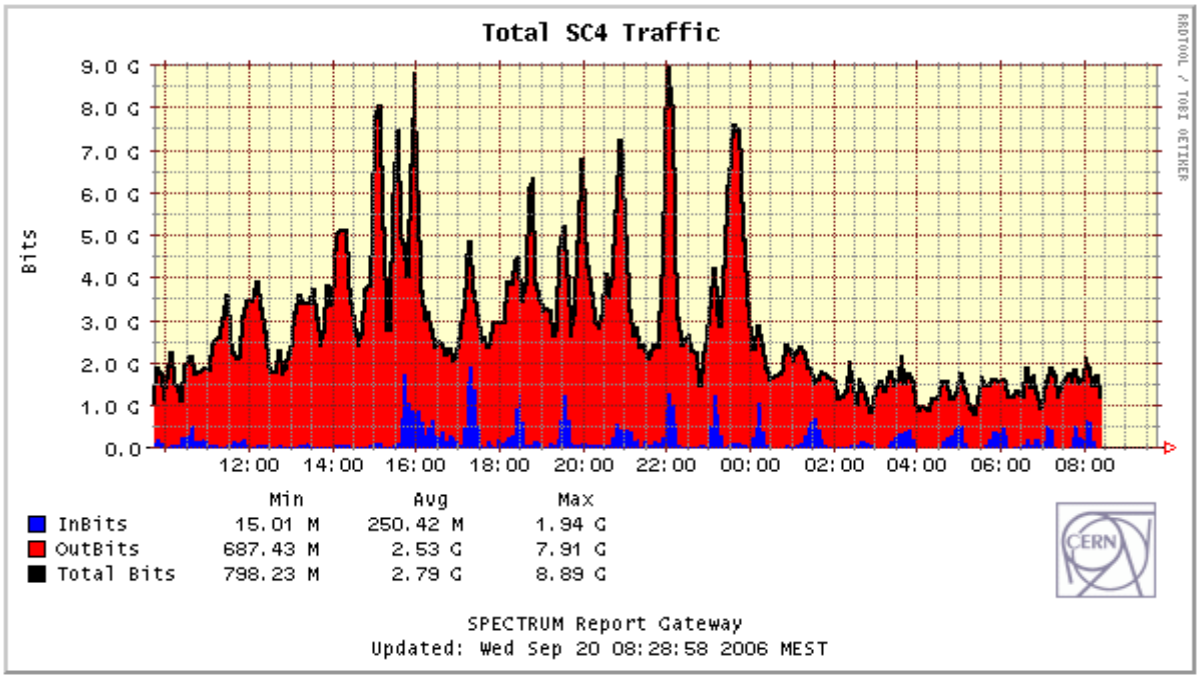
Service Challenges Tier-1 Sites

- SC4 Total Traffic
- ASGC-Taiwan 1
- ASGC-Taiwan 2
- ASGC-Taiwan Backup
- CNAF
- FNAL Primary
- FNAL Backup
- BNL Primary
- BNL Backup
- GRIDKA
- RAL
- SARA
- TRIUMF 1
- TRIUMF 2
- IN2P3
- NDGF-PIC-DESY 1 (Geant2)
- NDGF-PIC-DESY 2 (Geant2)

Service Challenges / LHCopn Traffic Monitoring

Service Challenge 4 has started with the goal of demonstrating that all of the offline data processing requirements expressed in the experiments Computing Models, from raw data taking through to analysis, can be handled by the Grid at the full nominal data rate of the LHC.

The current throughput for each Tier0-Tier1 link is available on the left side menu.



Service Challenges / LHCopn Traffic Monitoring

LHCopn and LCG

- LHCopn Twiki
- Network Map
- Prefixes
- LCG Networking Group
- LCG Service Challenge 4 Twiki

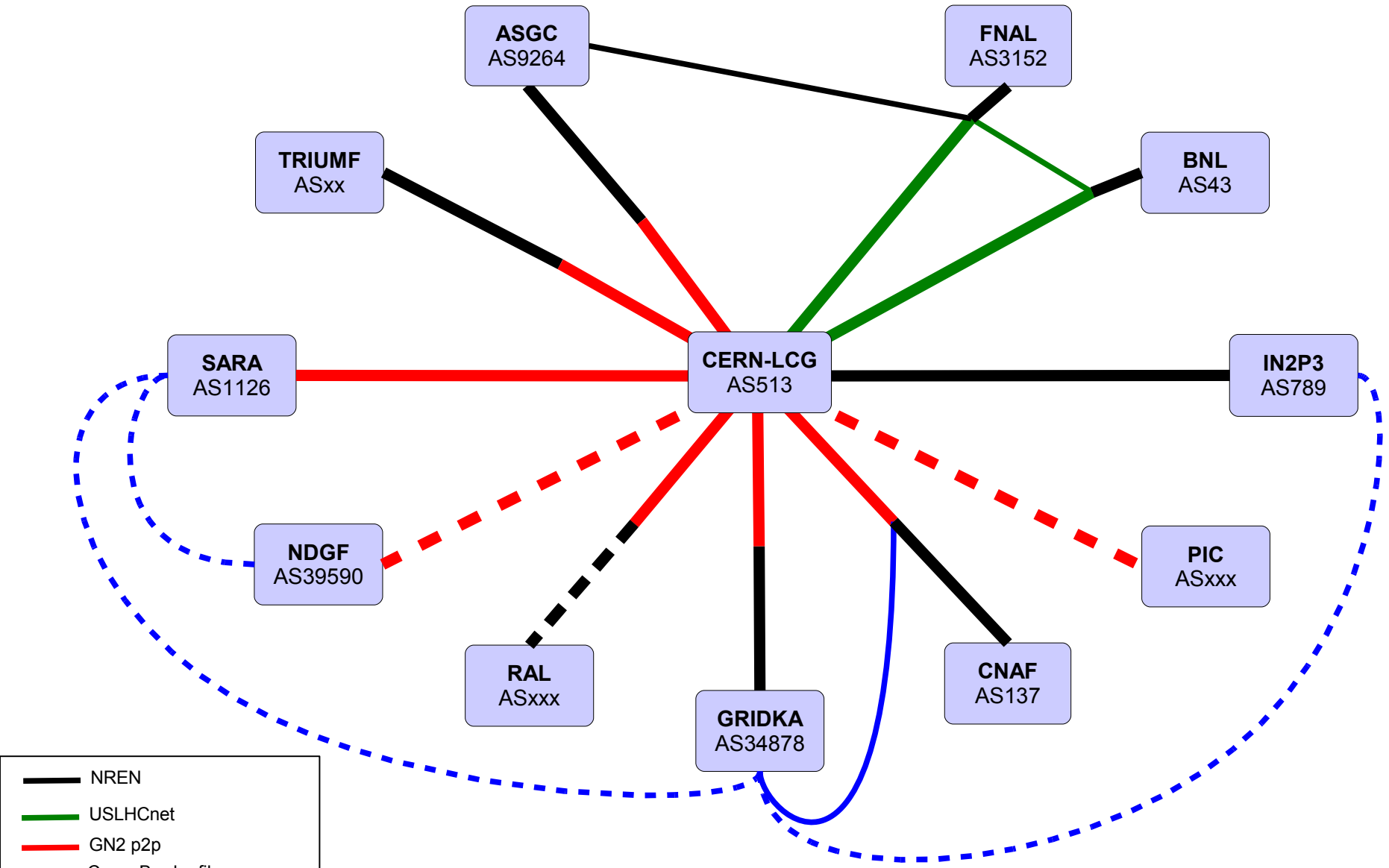
Monitoring Links

- MonALISA for LHCopn
- GridView
- SPECTRUM Service Dashboard (CERN only)

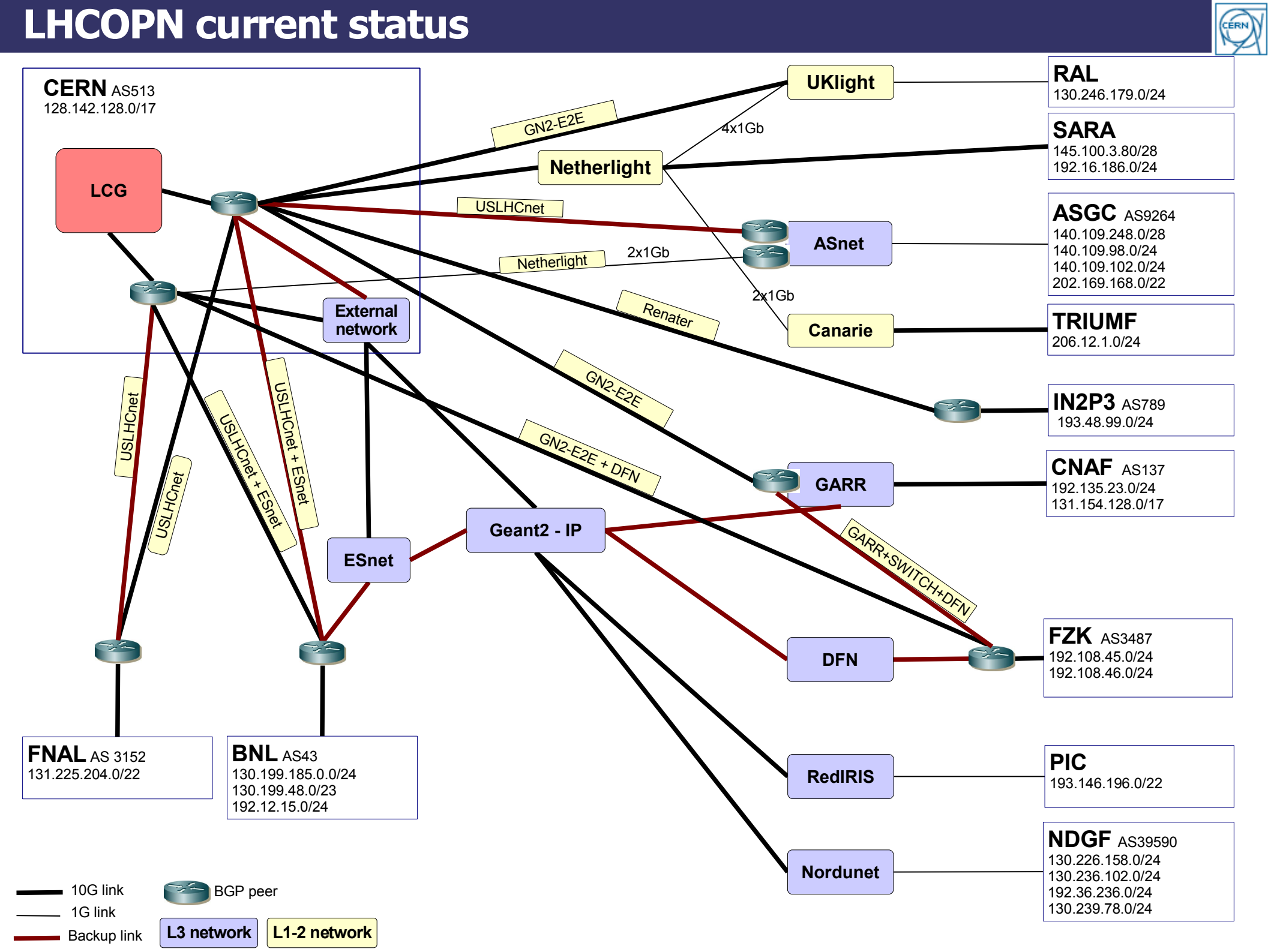
Feedback: Nikos Trikoupis and Juraj Sucik
Last update: Monday, 10-Jul-2006 17:37

About this Site
Copyright CERN

LHCOPN links (active and foreseen)



LHCOPN current status



Routing document: addressing, routing, backup

LHCopnRoutingDoc < LHCOPN < ...

Search

Webs

ABATBEA
ACPP
ADCgroup
ALICE
ALPHA
AliceSPD
ArdaGrid
AthenaFCalTBAna
Atlas
CERNSearch
CMS
Controls
DESgroup
DefaultWeb
EGEE
ELFms
ETICS
EgeePtf
FIOgroup
HCC
ItCommTeam
Know
LAr
LCG
LCGAAWorkbook
LHCAtHome
LHCOPN
LHCb
LHCgas
LinuxSupport
Main
PSSGroup
Plugins
RCTF
SPI
SRMDev
Sandbox
SocialActivities
SupComp05
TWiki

LHCOPN - IP addressing and routing

Scope

This document gives some guidelines about the configuration of the LHCOPN network. It focuses on IP addressing and routing aspects; it uses and complements the instructions already provided by the LCG Network Architecture document. It also reports all the decisions taken by the LHCOPN routing working group.

IP addressing

- **Use of public IP addresses**: Every Tier must assign publicly routable IP addresses to the machines that need to be reached over the T0-T1 links; these IP networks are referred as " *LHC prefixes* ". A LHC prefix cannot be a RFC1918 network.
- **Aggregation in few prefixes**: every Tier must aggregate the address space dedicated to the LHCOPN traffic into one or few CIDR blocks.
- **Addresses for the T0-T1 links**: the network 192.16.166.0/24 has been allocated for the addressing of the point-to-point links between the T0 and the Tier1s. Please refer to LHCopnTables#AnchorASnumbersLHCprefixes for the allocation of the /30.
- **Security**: for security reason, only packets with source and destination IP address that belong to one of the LHCOPN prefixes can transit in the LHCOPN. Please refer to the Security policies document for more information about security in the LHCOPN.
- **LHCOPN prefixes repository**: the list of the LHCOPN prefixes is saved in the RIPE route-set object RS-LHCOPN. The object is mantained by CERN; all the request for modification must be sent to extip@cernNOPAMSMAND.ch.

Routing

- **Topology**: the LHCOPN is essentially a star with the Tier0 at the centre. Connections between pairs of T1s are also admitted.
- **Routing**: routing is ensured by BGP (see below); usage of static routes and/or default route is discouraged.
- **Global connectivity**: machines connected to the LHCOPN might need global connectivity. Sites that don't use dedicated routers for the LHCOPN can take advantages of Policy Based Routing to correctly steer only the LHC traffic. Every Tier is responsible to correctly route its traffic towards the correct upstream.
- **T1 to T1 transit via T0**: data transfers between T1 centres transiting to the T0 is technically feasible. However, at the time of writing there is not such request and it is not implemented.
- **T0-T2 traffic**: at the time of writing it is not allowed inside the LHCOPN.

Backup connectivity

There are severeal options for backup connectivity:

- **Direct lightpath**: a second direct lightpath between the T1 and the T0.
- **Mutual backup**: pairs of T1s directly connected with a lightpath can use it also for mutual backup.
- **Backup of last resort**: backup via Layer 3 paths across NRENs and Research Backbones. This option is discouraged as it might heavily interfere with the traffic normally flowing through those backbones. However it is considered acceptable, especially in the warm-up phase of the LHCOPN.
  Traffic sent to Geant2-IP should be marked for the LBE service (DSCP=8, TOS=0x20, see here for details).

Routing document: BGP

LHCopnRoutingDoc < LHCOPN < ...

BGP setup

BGP is the routing protocol that manage the routing in the LHCOPN. External BGP peerings are established among the T0 and the T1s.

- **T0's BGP speakers**: the T0's BGP speakers are two routers connected to the CERN's LCG backbone and terminating all the T0-T1 links.
- **T1's BGP speaker**: the T1's BGP speaker is the router that terminates the T0-T1 link in the T1 side.

Guidelines for primary connectivity between the T0 and every T1

- **AS number**: An Autonumous System number is necessary to establish eBGP peers. Tiers need to use a valid public AS number; if they don't have one, they should contact their upstream NREN or their LIR to obtain one. The list of AS numbers used is in LHCopnTables#AnchorASnumbersLHCprefixes.
- **T1's announces to the T0** : every T1 announces its own LHC prefixes.
- **T0 announces to every T1**: the T0 announces its own LHC prefixes.
- **Prefixes accepted by the T1**: a T1 must accepts the T0's prefixes.
- **Prefixes accepted by the T0**: the T0 accepts only the LHC prefixes related to the peering T1.

Guidelines for backup connectivity

In case of a T1 with two direct lightpath to the T0:

- **Metrics:** the T0 and the T1 will set appropriate BGP MED values in order to prefer the main connection rather than the backup one.
- **Annunces:** except for the metrics, the announces are the same on the primary and the backup link.

In case of pair of T1s providing mutual backup via a lightpath between the two (triangle):

- **Peering:** The two directly connected T1s must establish an eBGP peering over the direct link
- **T1's announces to the peering T1:** each T1 announces the T0's prefixes to the peering T1 (together with its own prefixes)
- **T1's announces to the T0:** each T1 announces the peering T1's prefixes to the T0 (together with its own prefixes)
- **Prefixes accepted by the T1 from the T1**: each T1 must accept the peering T1's prefixes and the T0's prefixes
- **Prefixes accepted by the T0**: the T0 must accept the prefixes of both the T1s

In case of backup via generic Internet *[not recommended]*:

- **Announces to the generic Internet**: T0 and T1s have to announce their LHC prefixes to their upstream networks (GÉANT2, Abilene, ESnet, for instance). Most probably the LHC prefixes are part of networks already announced; in this case there is no need to announce the more specific LHCOPN prefix.
- Special care must be taken by each Tier to not leak out BGP prefixes that belong to other Tiers.

Guidelines T1-T1 transit via the T0 *[not implemented yet]*

- **T0 announces to every T1**: The T0 announces all the T1s' prefixes received (together with its own prefixes)
- **Prefixes accepted by the T1**: each T1 accepts the other T1s prefixes that want to reach via the T0. Unnecessary prefixes had better to be filtered out

Private AS numbers

RAL has got from Janet a private ASN to use: is it acceptable or should they ask for a public one?

SARA and RAL share a T2 without a public ASN. In case it wants to use the LHCOPN, is a private ASN acceptable? Btw, are T2s allowed to use the LHCOPN?

Security

It seems an incident has already happened inside the LHCOPN. Is anyone strengthening the security on the LHCOPN borders?

Anyone is already using access lists that filters at TCP and UDP level?

What about an LHCOPN operation officer? It could try to check the implementation of the Security Working group recommendations, or/and liaises the security officers of the LHCOPN members.

Tier2s allowed in the LHCOPN?

T0-T1 links bandwidth utilization can run out of control.

Management of the security ACLs can become complicated.

Security will decrease. Probably the model should be changed.

If private ASNs become popular, a sort of registration authority will be necessary.