LHCOPN routing w.g.

Amsterdam - 31st January 2006

edoardo.martelli@cern.ch

Latest achievement

- **CNAF**: connected with 10Gbps lightpath provided by GN2; BGP configured, backup via GN2-IP
- **FNAL**: connected with primary and backup 10Gbps lightpaths provided by USLHCnet; BGP configured.
- **IN2P3**: connected with 10Gbps lightpath provided by Renater; BGP configured, backup via GN2-IP
- **SARA**: 10Gbps lightpath provided by Surfnet but still no BGP configured. No automatic backup.

Latest achievement

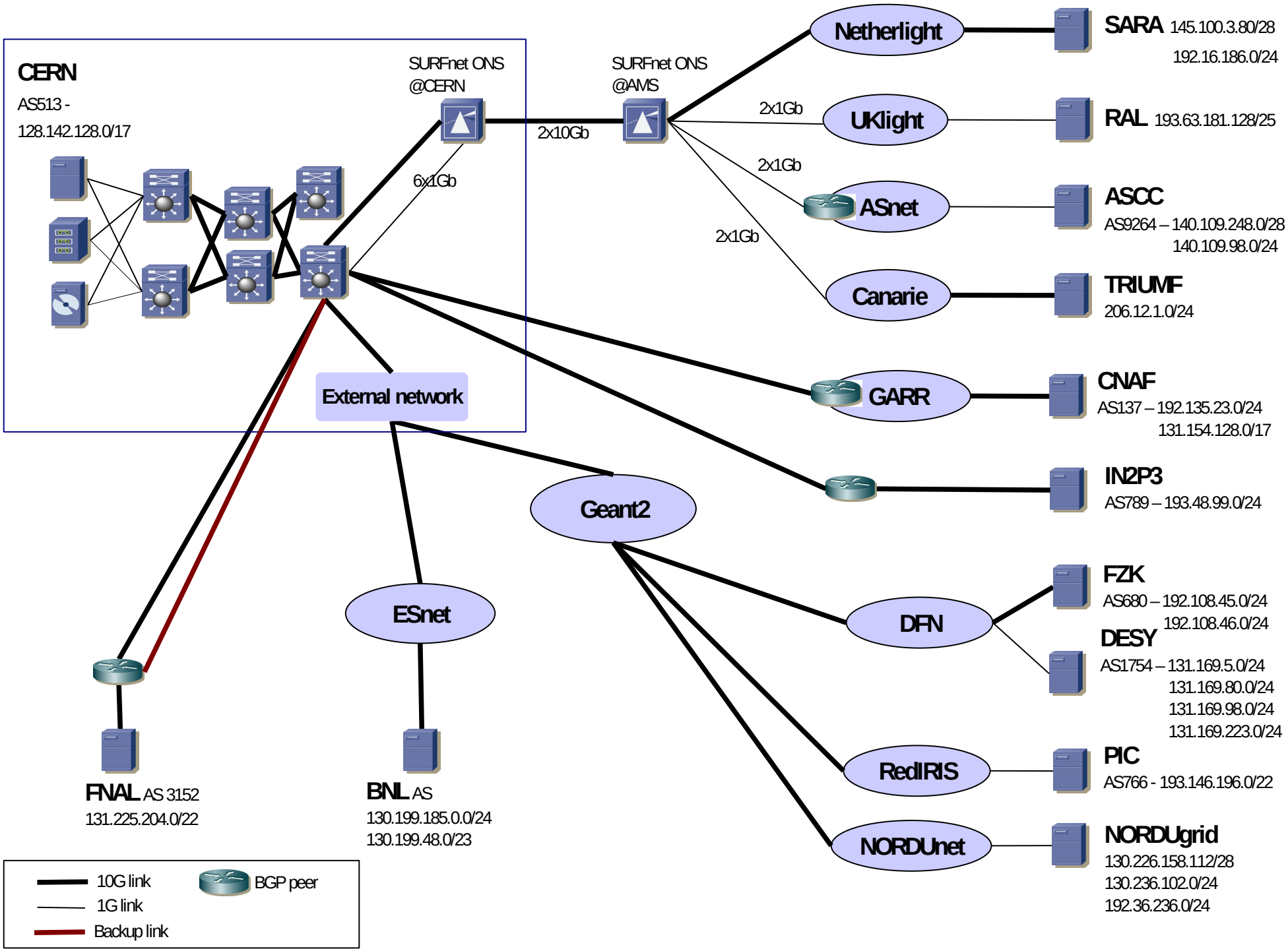
- **CERN** started the deployment of the planned Tier0 architecture:

- six Force10 of the LCG backbone core are already operational;

- CERN Service Challenge 3's machines reconnected to the LCG backbone

- every new machine deployed is immediately connected to the LCG backbone.

LHCOPN status



Tier0

The LCG backbone at CERN

Tier0 construction

Implementation status

Some sites have already implemented the suggested architecture and use BGP to route the traffic.

/30 from the 192.16.166.0/24 are used for the addressing of the direct lightpaths. Removal of the legacy /30 is on going.

Only T0-T1 traffic is allowed at the moment; no transit.

Security ACLs in place on the T0's interfaces facing the Tier1s. Status at the T1s is unknown.

ACLs are at IP level for the time being (src-dst ip addresses).

LHCOPN prefixes

The list of IP prefixes allowed in the LHCOPN is stored in the RIPE database in the route-set object **RS-LHCOPN.**

To retrieve it:

http://www.ripe.net/perl/whois?&searchtext=RS-LHCOPN

CERN mantains the object. Please contact extip@cern.ch for any request.

Next steps

Implement BGP routing for all the sites: static routing and multi-site LAN don't allow automatic backups, are difficult to debug, require lot of reconfiguration effort in case of prefix change.

Implement T1-T1 transit via T0 (is it needed?)

Deployment of the second router at CERN that will terminate the T1s' lightpaths.

Keep deploying the direct lightpaths as soon as they are available.

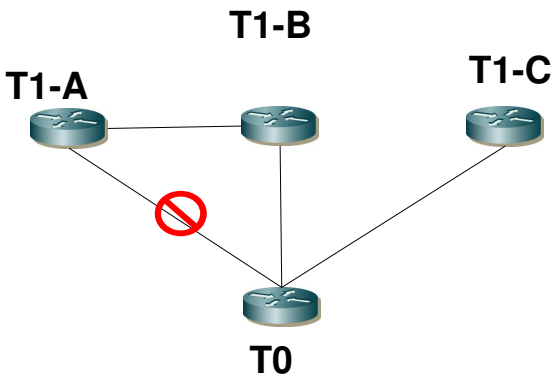
Open questions

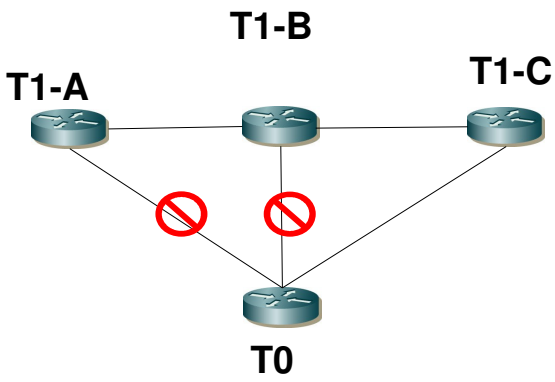
- T1-T1 transit via T0 is needed?

- Backup over Layer3 connections: should be implemented (kept, actually) or dropped?

- Tier2 access?

- Direct T1-T1 links should be made available to all the LHCOPN members?





Tier0 - CERN

Technical contact: Edoardo Martelli and David Gutierrez,

extip@cern.ch

Operations contact: noc@cern.ch +41 22 767 4011 (24x7)

AS number: AS513

LHC prefixes: 128.142.128.0/17

Notes:

- Force10 E1200 routers. 1Gb Ethernet (T, SX, LX), LANPHY (SX, LX, EX) and WANPHY (LX, EX) interfaces available.

Tier1 - ASCC

Technical contact: Yu Lin Chang ylchang@gate.sinica.edu

Operations contact:

Primary link: Concatenad lightpath: Taipeh-Amsterdam (10G?, Provider?) - Amsterdam-Geneva (10G, GN2 ?).- WANPHY?

Backup link:

AS number: AS9264

LHC prefixes: 140.109.248.0/28 - 140.109.98.0/24

Notes:

Tier1 - BNL

Technical contact: John Bigrow big@bnl.gov

Operations contact: itdhelp@bnl.gov

Primary link: Concatenated lightpath: Brookhaven-NYManlan (10G? Esnet), NYManlan-Geneva (10G, USLHCnet). LANPHY?

Backup link: Concatenated lightpath: Brookhaven-NYManlan (10G? Esnet), NYManlan-Chicago-Geneva (10G, USLHCnet). LANPHY?

AS number: AS43

LHC prefixes: 130.199.185.0/24 – 130.199.48.0/23 192.12.15.0/24

Notes:

Tier1 - CNAF

Technical contact: Marco.Marletta@garr.it

Luca.DellAgnello@cnaf.infn.it

Operations contact:

Primary link: Concatenated Lightpath: Bologna-Milano (10G, GARR), Milano-Geneva (10G, GN2). 10G-baseLX

Backup link: Concatenated lightpath: Bologna-Milano (?), Milano-Zurich-Geneva (10G, SWITCH). LANPHY?

AS number: AS137

LHC prefixes: 131.154.128.0/17 - 192.135.23.0/24

Notes:

- Direct link with Gridka
- Peering with the GARR router in Milano
- Policy Based Routing used inside GARR network

Tier1 - FNAL

Technical contact: Vyto Grigaliunas vyto@fnal.gov, Andrey Bobyshew bobyshew@fnal.gov

Operations contact: helpdesk@fnal.gov +1-630-840-2345 (8x5)

Primary link: Concatenated lightpath: Batavia-Chicago (10G, FNAL), Chicago-Geneva (10G, USLHCnet). 10GbaseLX.

Backup link: Concatenad lightpath Batavia-Chicago (10G, FNAL), Chicago-NYC-Geneva (10G, USLHCnet).

AS number: AS3152

LHC prefixes: 131.225.204.0/22

Notes:

- Policy Based Routing used inside FNAL network

Tier1 - Gridka

Technical contact: Bruno.Hoeft@iwr.fzk.de

Operations contact:

Primary link: Concatenated lightpath: Karlsruhe-Frankfurt(10G, DFN), Frankfurt-Geneva (10G, GN2). 10GbaseLX.

Backup link: Concatenated lighpath provided bt DFN and SWITCH. Interface?

AS number: AS34878

LHC prefixes: 192.108.45.0/24 192.108.46.0/24

Notes:

- Direct link with GARR and SARA
- Policy Based Routing used inside FZK network

Tier1 - IN2P3

Technical contact: Jerome.Bernier@in2p3.fr - Simon Frank simon@renater.fr

Operations contact: noc@in2p3.fr +41 33 78 93 08 80

Primary link: Direct lightpath: Lyon-Geneva (10G, Renater). 10GbaseLX

Backup link: L3 path viar Renater and Geant

AS number: AS789

LHC prefixes: 193.48.99.0/24

Notes:

Tier1 - NORDUgrid

Technical contact: Per Nihlen

Operations contact:

Primary link: Concatenated lightpath

Backup link:

AS number:

LHC prefixes: 130.226.158.64/26 130.236.102.0/24 192.36.236.0/24

Notes:

Tier1 - PIC

Technical contact: Esther.Robles@rediris.es, delfino@pic.es

Operations contact: network@pic.es

Primary link: Concatenated lightpath: Barcelona-Madrid (10G, Rediris), Madrid-Geneva (10G,GN2). LANPHY

Backup link: MPLS tunnel over Rediris + GN2

AS number: AS766

LHC prefixes: 193.146.196.0/22

Notes:

Tier1 - RAL

Technical contact:

Operations contact:

Primary link: Concatenated lightpath? Interface?

Backup link:

AS number: 786

LHC prefixes: 193.63.181.128/25 130.246.179.0/24

Notes:

Tier1 - SARA

Technical contact: Pieter de Boer pieter@sara.nl - Hanno.Pet@sara.nl

Operations contact: nic@sara.nl, nrg@sara.nl +312-5028000

Primary link: Direct lightpath: Amsterdam-Geneva (10G, Netherlight). WANPHY

Backup link:

AS number: AS1126

LHC prefixes: 145.100.3.80/28

Notes:

Tier1 - TRIUMF

Technical contact: Steven.McDonald@triumf.ca

Operations contact:

Primary link: Concatenated lightpath: Vancouver-Amsterdam (10G, ?), Amsterdam-Geneva (10G, GN2). Interface?

Backup link:

AS number:

LHC prefixes: 206.12.1.0/24

Notes:

The End

Additional slides

LHCOPN architecture – IP addressing

- Every T1 and the T0 must allocate publicly routable IP address space to the machines that need to be reached over the T0-T1 links (the "LHC prefixes").
- LHC prefixes should be aggregated into a single CIDR block for every T1; if this is not possible, only a very small number of CIDR blocks per T1 would be acceptable.
- LHC prefixes should be dedicated to the LHC network traffic.
- If a T1 can't dedicate a prefix but need to use its whole address space, then it will be its own responsibility to inject in the LHCOPN only the traffic directed to the LHC-OPN resources.
- LHC prefixes can be carved as a CIDR block from a T1's existing allocations or obtained as new allocation from the appropriate RIR through already established channels.
- LHC prefixes cannot be from RFC1918 and related (like RFC3330) addresses.
- T0 will allocate /30 prefixes for the addressing of the T0-T1 links, i.e. the links that connect to CERN up to the first BGP speaker in the path.
- Every T1 and T2 interested in exchanging traffic directly with the T0 is required to provide the T0 with the list of its LHC prefixes.

LHCOPN architecture – BGP

External BGP peerings will be established between T0 and each T1. More precisely, the T1 peer is the BGP speaker directly connected to the T0 on behalf of a specific T1

Guidelines for the BGP configuration:

· The Tiers will use the AS number of the entity that provides the LHC prefixes to them or the AS number of their standard upstream NREN.

· Every T1 will announce its own LHC prefixes to T0.

· T0 will announce its LHC prefixes to every peering T1.

· T0 will accept only the LHC prefixes related to a specific T1, i.e. the T1's own LHC prefixes, plus LHC prefixes of any T1 or T2 for which that T1 is willing to provide transit for.

· T0 will re-announce to all the T1s all the LHC prefixes received in BGP.

· T1 will accept T0's prefixes, plus, if desired, all or some selected T1's prefixes .

· T0 and T1s should announce their LHC prefixes to their upstream continental research networks (GÉANT2, Abilene, ESnet) in order to allow connectivity towards the T2s.

- Special care must be taken by each Tier to not leak out BGP prefixes that belong to other Tiers.

LHCOPN architecture – More on routing

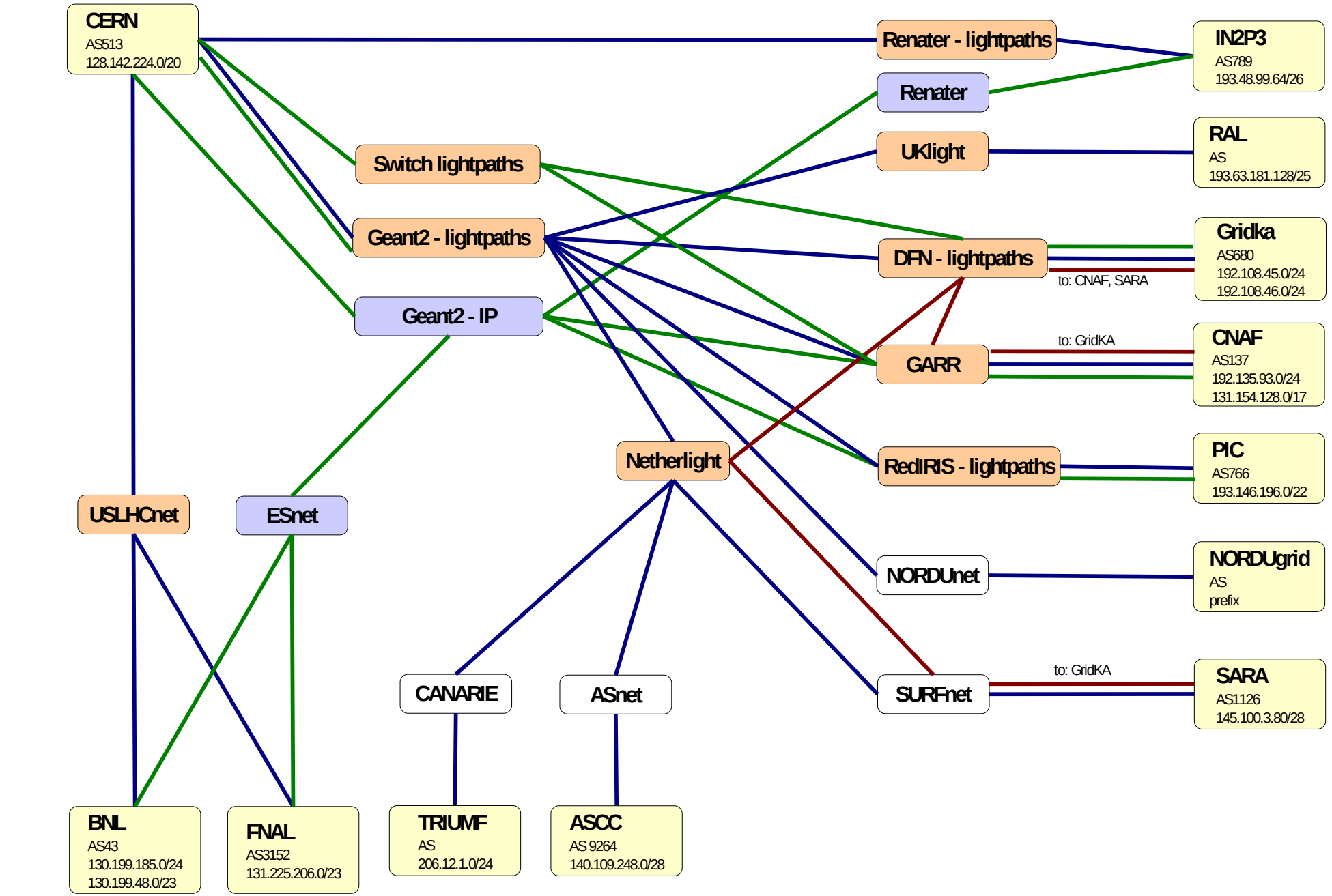
Some recommendations:

· Usage of static routes is not advisable.

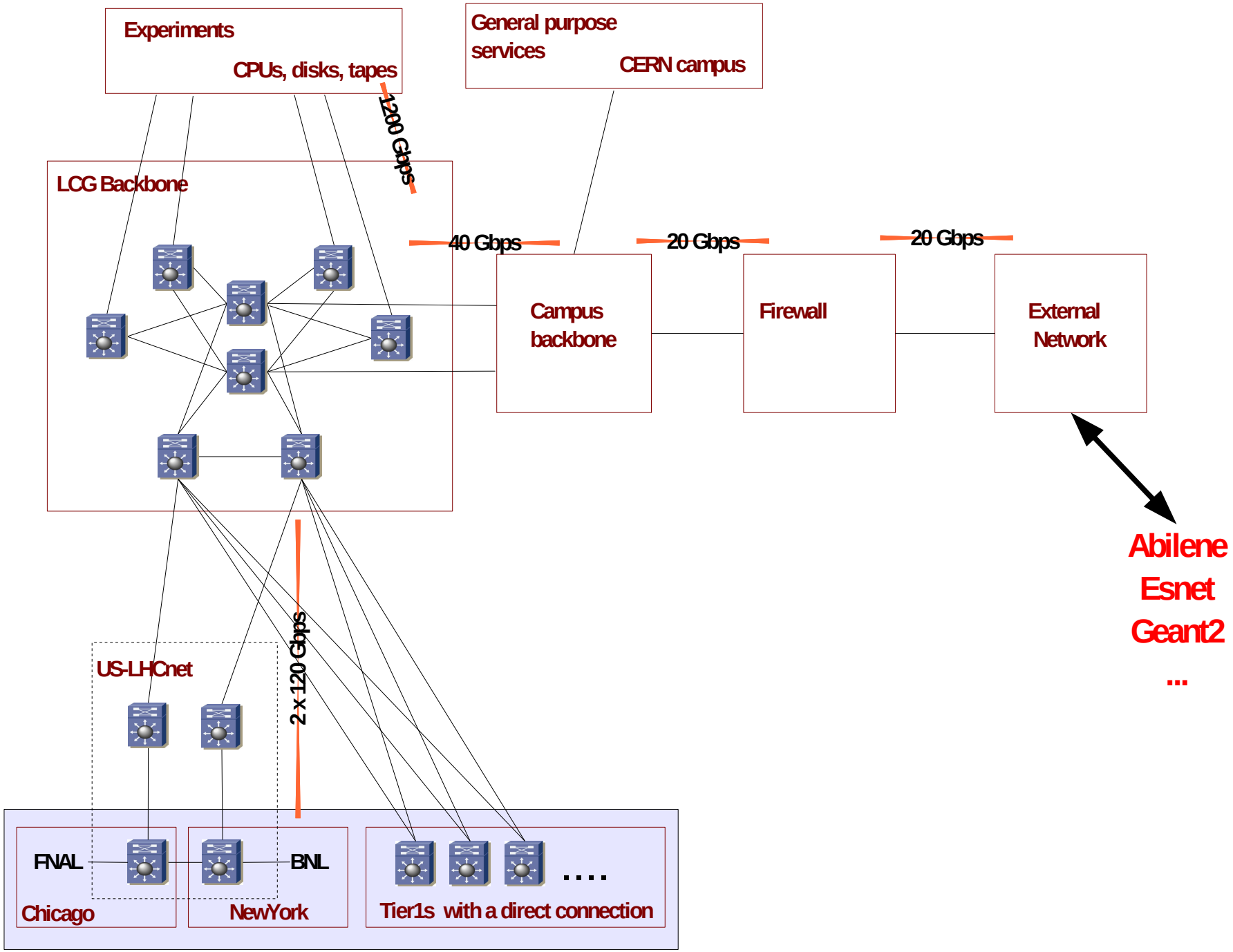
· No default route must be used in T1-T0 routing.

· It is the responsibility of every Tier to make sure that any of its own machines within the LHC prefix ranges can reach any essential service (for instance the DNS system).

LHCOPN Overview



Tier0 - CERN


Experiments
CPUs, disks, tapes
General purpose services
CERN campus
1200 Gbps
LCG Backbone
40 Gbps
20 Gbps
20 Gbps
Campus backbone
Firewall
External Network
Abilene
Esnet
Geant2
...
US-LHCnet
2 x 120 Gbps
FNAL
BNL
Chicago
NewYork
Tier1s with a direct connection
....