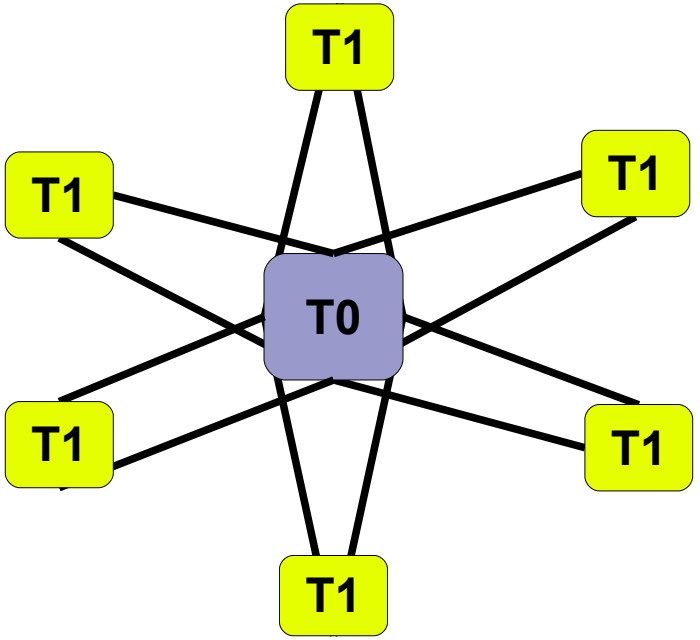
Effective Layer3 resilience

Cambridge - 16th June 2006

edoardo.martelli@cern.ch

Every Tier1 has at least two direct 10G links to the Tier0

A BGP peering is established over each direct lightpaths.



Reliable topologies

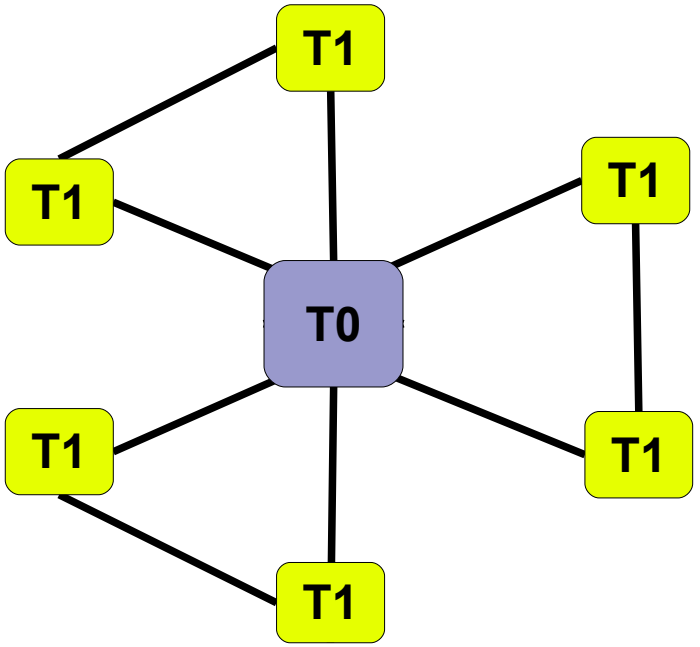
Every Tier1 has at least two direct 10G links to the Tier0

Pros:
- high reliability
- easy to monitor and keep operational
- possibility to exploit double the bandwidth in case of traffic burst
- no dependencies among Tier1s

Cons:
- expensive (more links and interfaces)

Pairs of Tier1s directly connected with a 10G link can share it for mutual backup

- A BGP peering is also established between directly connected Tier1s.
- Each Tier1 grants transit to the peering partner.



Reliable topologies

Pairs of Tier1s directly connected with a 10G link can share it for mutual backup

Pros:
- good reliability
- no unused links to be purchased
- less interfaces at the Tier0

Cons:
- configuration dependencies among Tier1s
- possible congestion in a T1-T0 link when the other one is faulty

How to build a reliable and effective LHCOPN?

- avoid unnecessary hops
- keep it separated by other domains
- reduce dependencies among Tier1s
- carefully consider all the bandwidth contenders

L2 connectivity

A wide offer of L2 connectivity:

- lambda from GN2 P2P service

- lambda from NREN owning dark fibres

- lambda over cross border fibres

- VLANs (even over transatlantic links)

- MPLS tunnels over transit network

- ....

But only in few cases a single occurrence can connect the Tier1's router to the Tier0's one. Most likely a single lithpath will require to stitch several flavours of links.

Link stitching difficulties:

Stitching together two links without changing OSI layer can be problematic sometimes:

- different nature: SDH and Ethernet, 1310nm and 1550nm, VLAN and MPLS, GRE and CCC....

- termination points in two disconnected locations

- incompatible devices (not completely standard)

- frame translator not available (can be expensive)

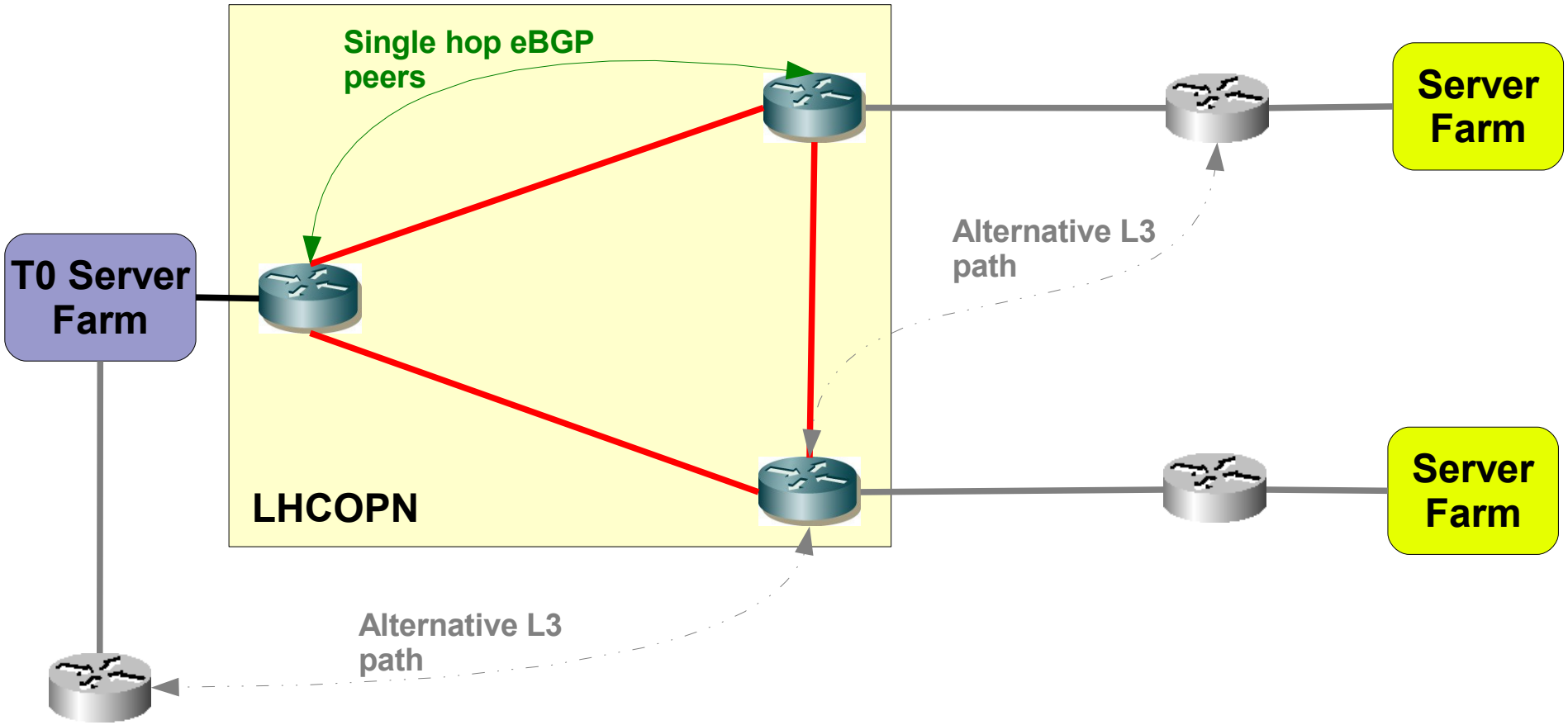
- ...

Link under utilized

It can happen that some lightpaths are under utilized, especially when they are dedicated to backup connectivity.

Splitting them in several L3 hops to better exploit the available bandwidth can be a tempting option.

Shared routers, not direct lightpaths



PBR is necessary to steer only the LHC traffic into the LHCOPN.

PBR: a real friend?

- can it detect if the next-hop is not reachable?

- if yes, what decision can make?

- that special PBR feature is supported in the router's OS?

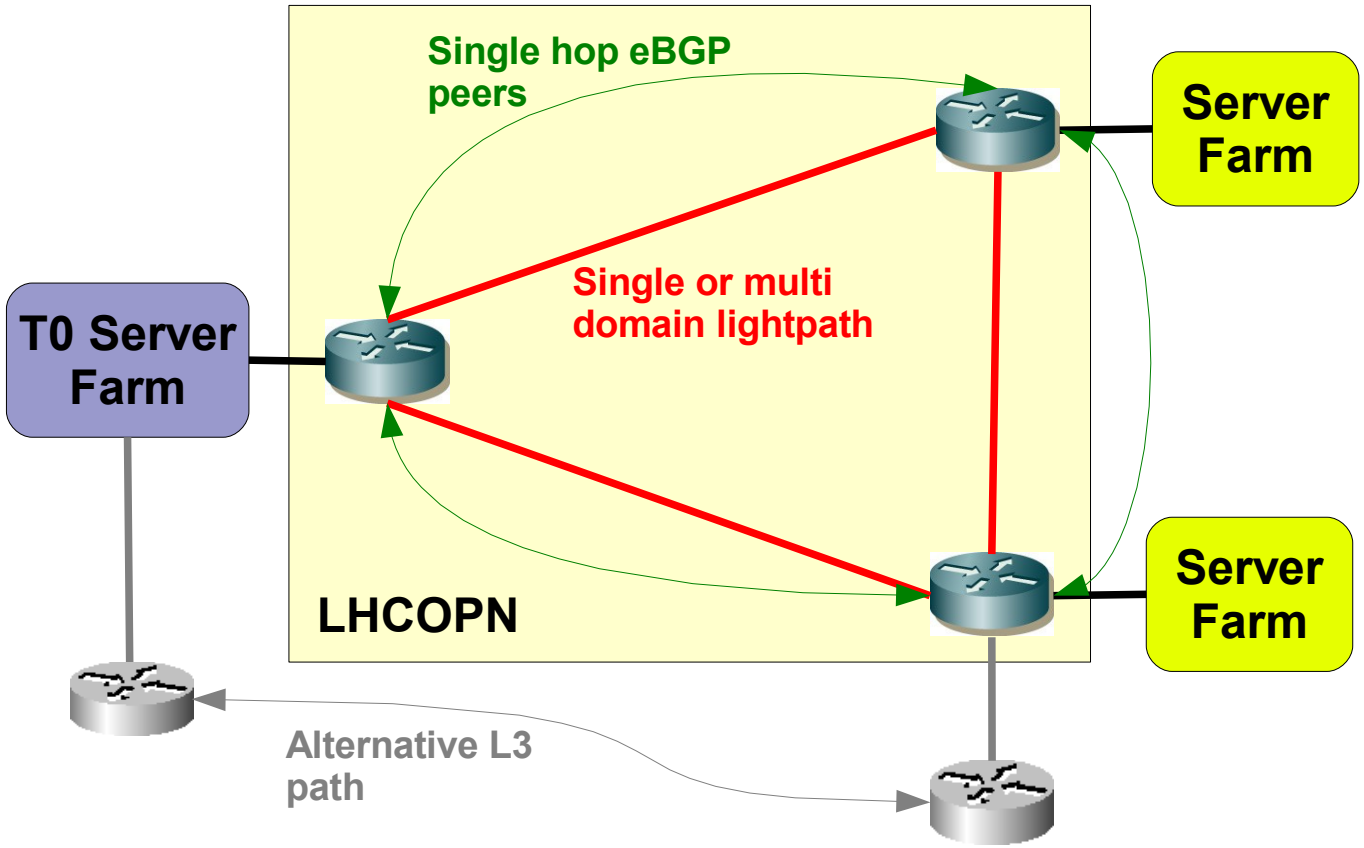
- is the packet forwarding made by the hardware?

- is it simple to add/remove prefixes to the policy rules?

- have all the possible network condition been evaluated?

PBR is almost static routing: not the best tool to deal with link failures.

Direct lightpaths, dedicated routers, BGP speakers close to the server farms



Straightforward to keep control of the traffic paths with dynamic routing

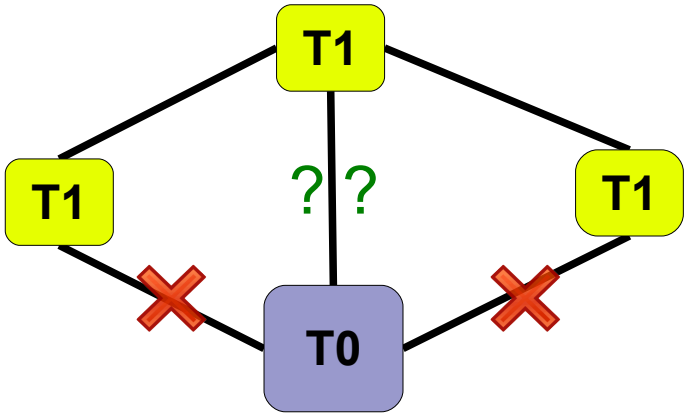
Know your backup mates:

- who are the bandwidth contenders?

- have they correctly updated their configurations?

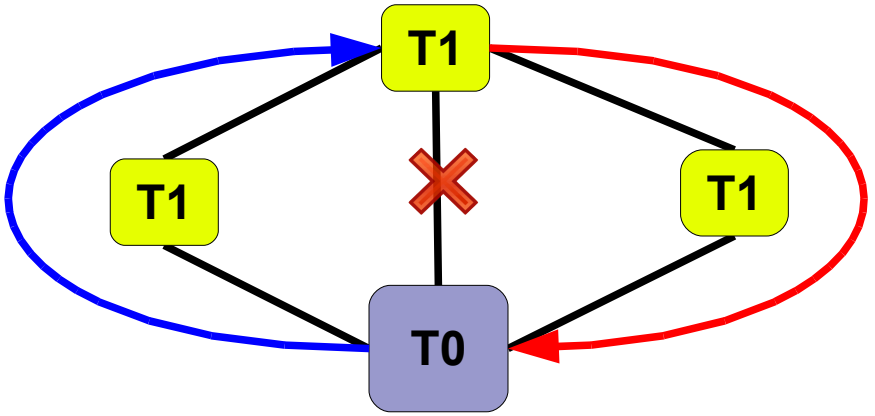
Tier1 offering backup to two others Tier1s

In case two links fail at the same time, will the only link left able to sustain all the traffic?



Tier1 offering backup to two others Tier1

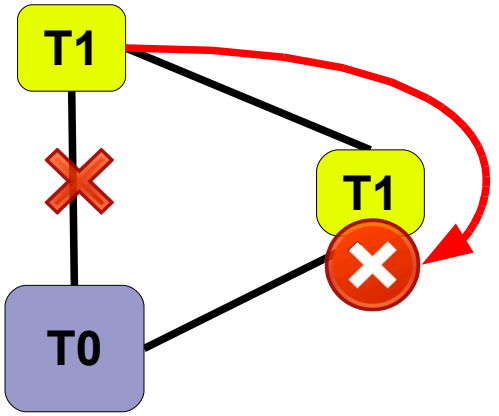
There can be two equal cost paths, and thus asymmetric routing: has this option be considered or could it be a problem?



Consistent filters?

Tier1s providing mutual backup relies on each other filters configuration

mutual backups require more care in keeping Access Lists and BGP filters up-to-date



Best practice

The LHCOPN's routers should be dedicated to the LHCOPN

- it can be difficult to keep traffic of different L3 domains separated when they share the same routing table.

- with a single shared hardware resource, an issue in a domain can impact all.

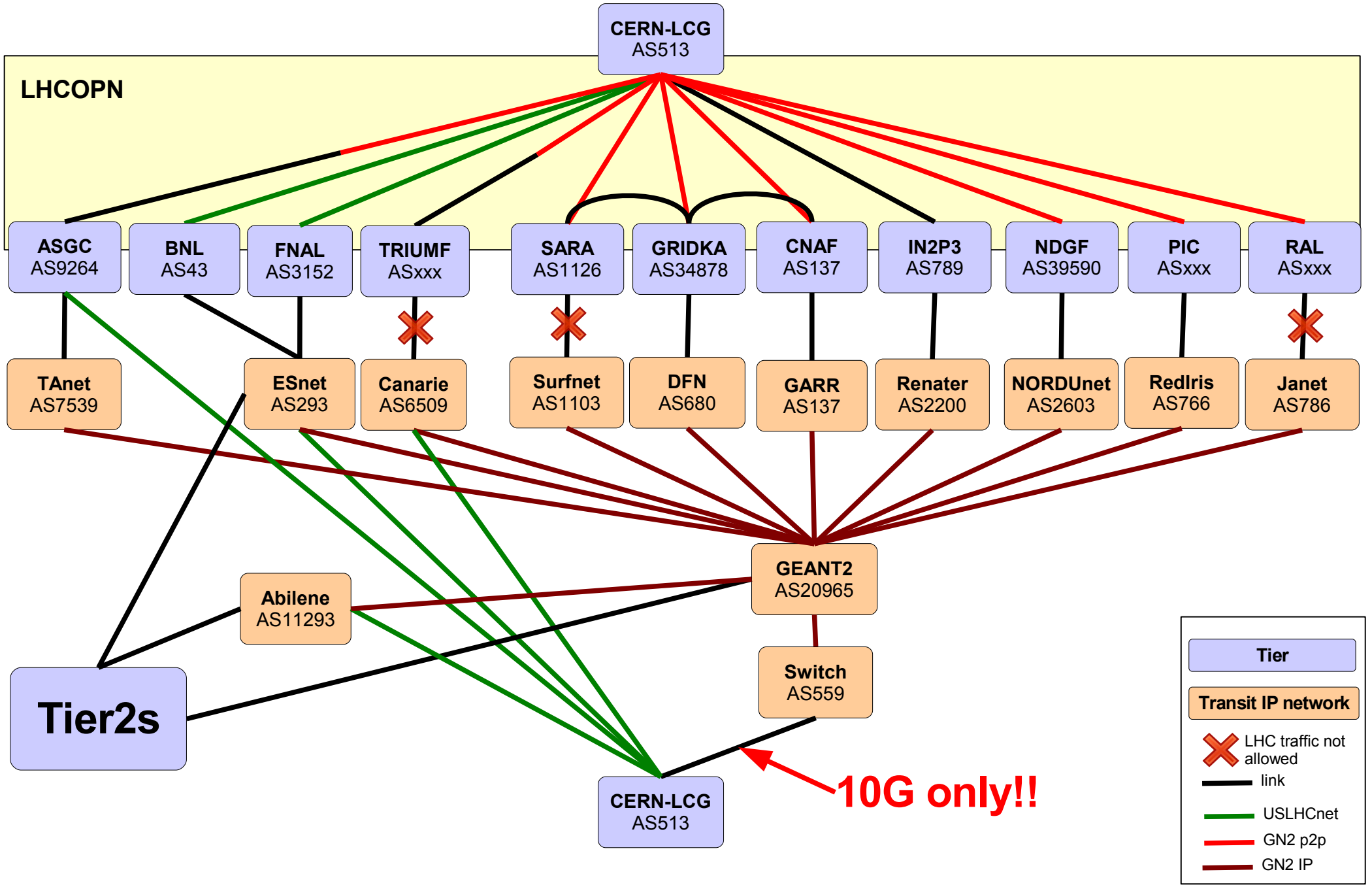
- avoid PBR. Virtual routers can be a better option

The LHCOPN's links should be dedicated to the LHCOPN

- bandwidth is a limited resources: what is spare today can become suddenly necessary.

Backup of last resort via L3 backbones: a risky option?

LHCOPN + Layer3 connectivity



Backup of last resort

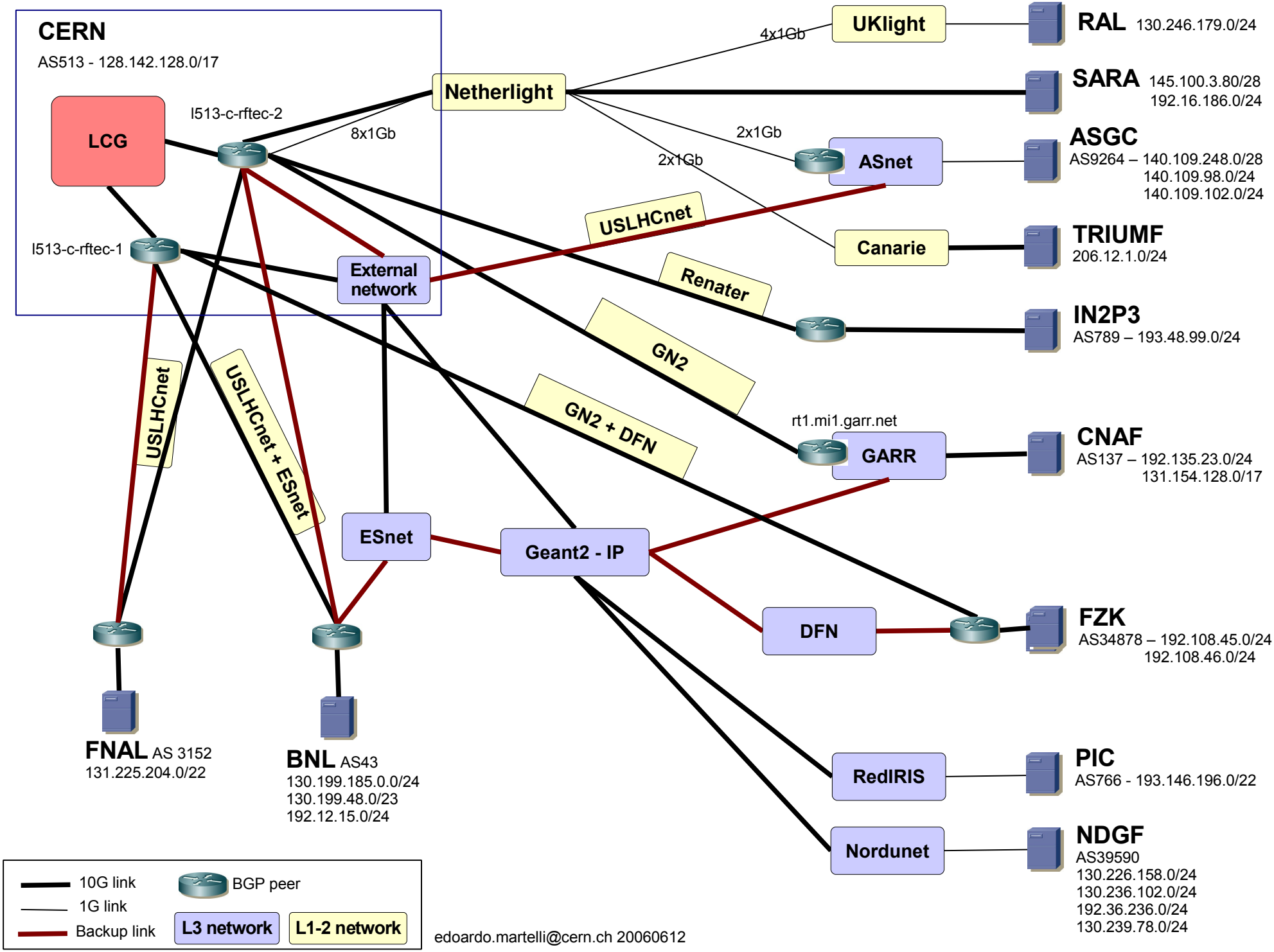
Backup of last resort cab be achieved on existing L3 path (via Gean2-IP, ESnet, ...)

Pros:
- most of it already in place (for several Tier1 it has been the main connectivity in the Service Challenges)
- cheaper than a dedicated backup: no additional connectivity to be purchased

Cons:
- limited available bandwidth: most likely to have contentions
- expensive bandwidth
- firewall constrains for some sites

Opinions?

LHCOPN status



edoardo.martelli@cern.ch 20060612