

End site challenges: Router virtualization and lightpaths

Peter Tavenier

SARA Computing and Networking Services

peter.tavenier@sara.nl



Agenda

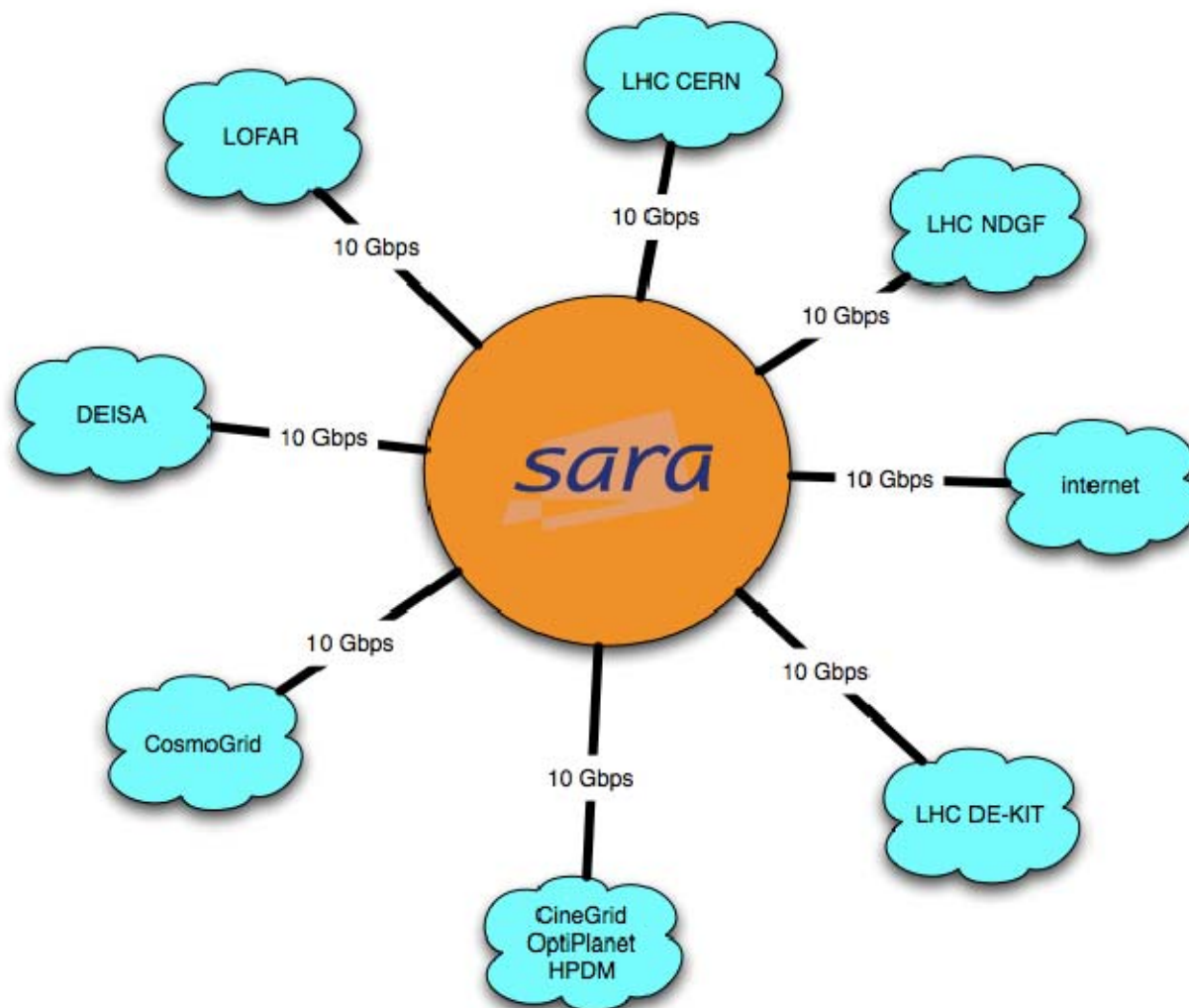
- ▀ About SARA and projects
- ▀ Lightpath challenges
- ▀ Requirements for SARA's network
- ▀ Upgrade path
- ▀ Virtual routing and router configuration example
- ▀ How we did it
- ▀ conclusions



About SARA

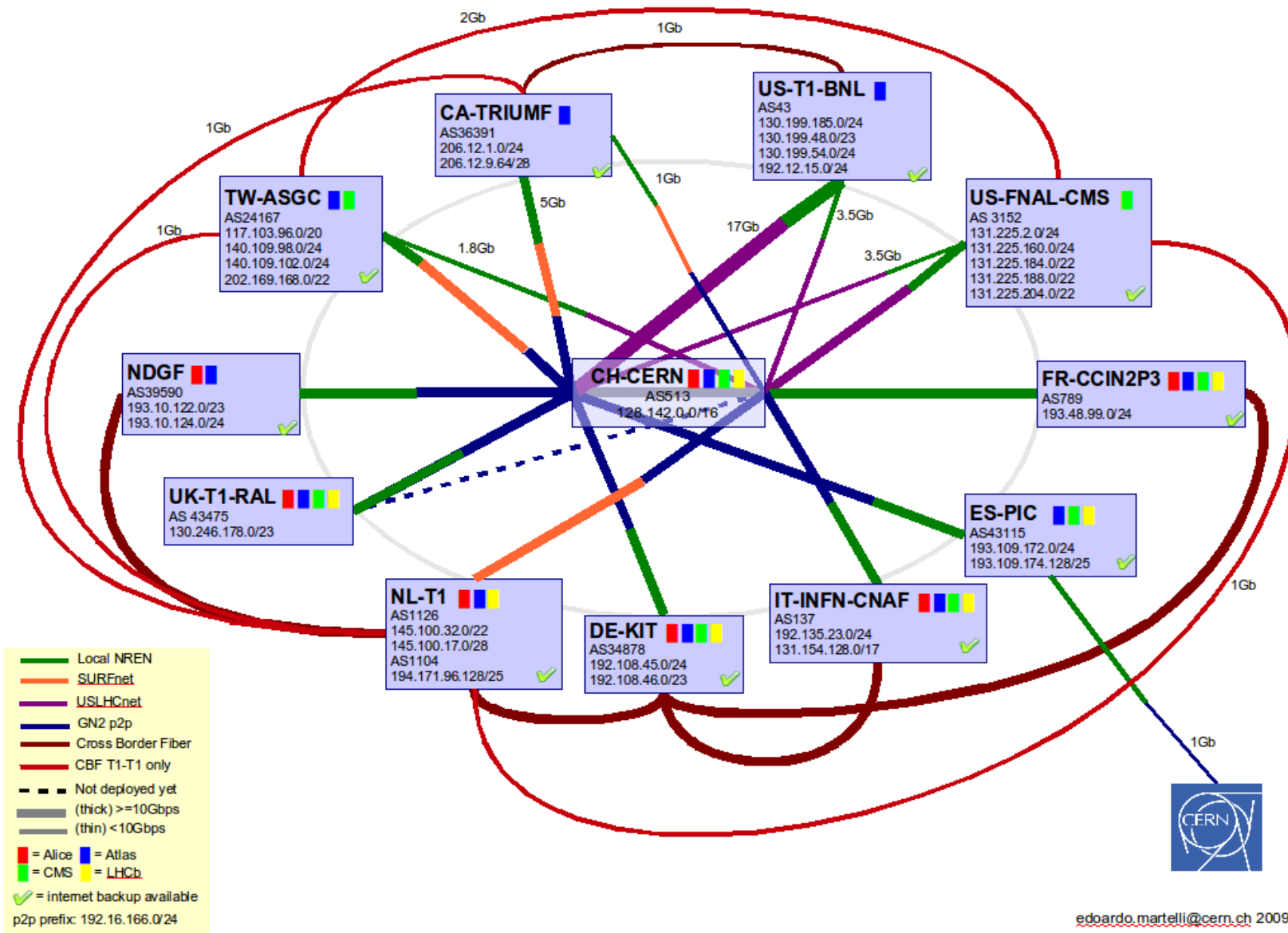
- ▶ **Computing and Networking services**
- ▶ **Houses and operates national supercomputer Huygens**
- ▶ **Houses and operates national cluster Lisa**
- ▶ **LightHouse (joint lab of SARA, UvA and SURFnet for optical networking experiments and demos)**
- ▶ **SURFnet's subcontractor for SURFnet6 NOC**
- ▶ **SURFnet's subcontractor for Netherlight NOC**
- ▶ **One of the co-location sites of the AMS-IX**
- ▶ **CERN LHC Tier-1 site**
- ▶ **LOFAR Tier-1 site**
- ▶ **Life Science Grid clusters**

SARA's lightpath connectivity



LHC OPN Tier-1 site

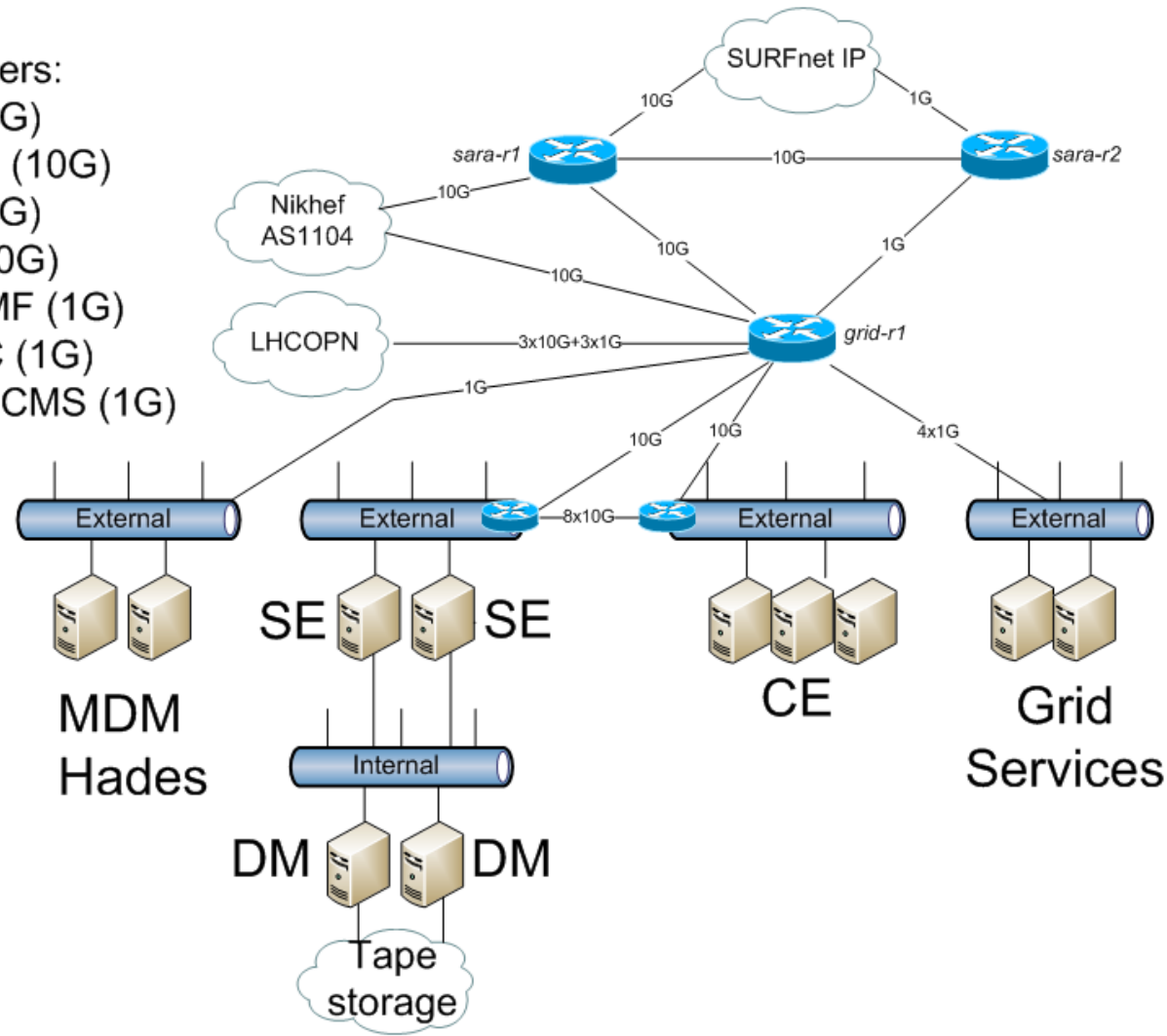
LHCOPN – current status



edoardo.martelli@cern.ch 20091103

LHCOPN peers:

- Nikhef (10G)
- CH-CERN (10G)
- NDGF (10G)
- DE-KIT (10G)
- CA-TRIUMF (1G)
- TW-ASGC (1G)
- US-FNAL-CMS (1G)



LOFAR Tier-1 Site

- ▶ **LOW Frequency ARray**
- ▶ **Radiotelescope**
- ▶ **Consists of Sensor Fields**
- ▶ **Data Storage at SARA**



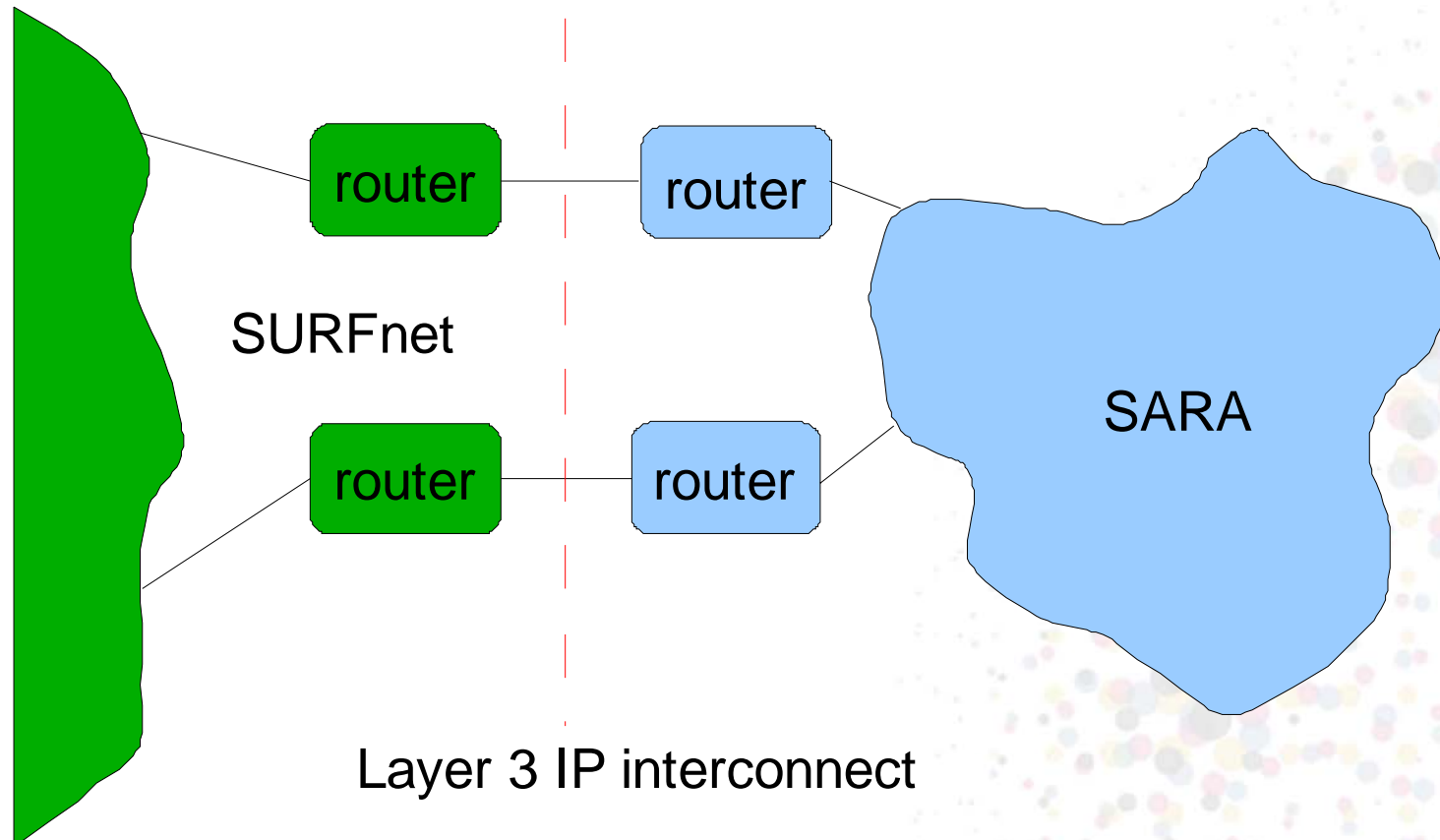
IMAU climate model

- ▶ Rendering at SARA
- ▶ Visualization at IMAU
- ▶ Connected with a SURFnet6 1G lightpath



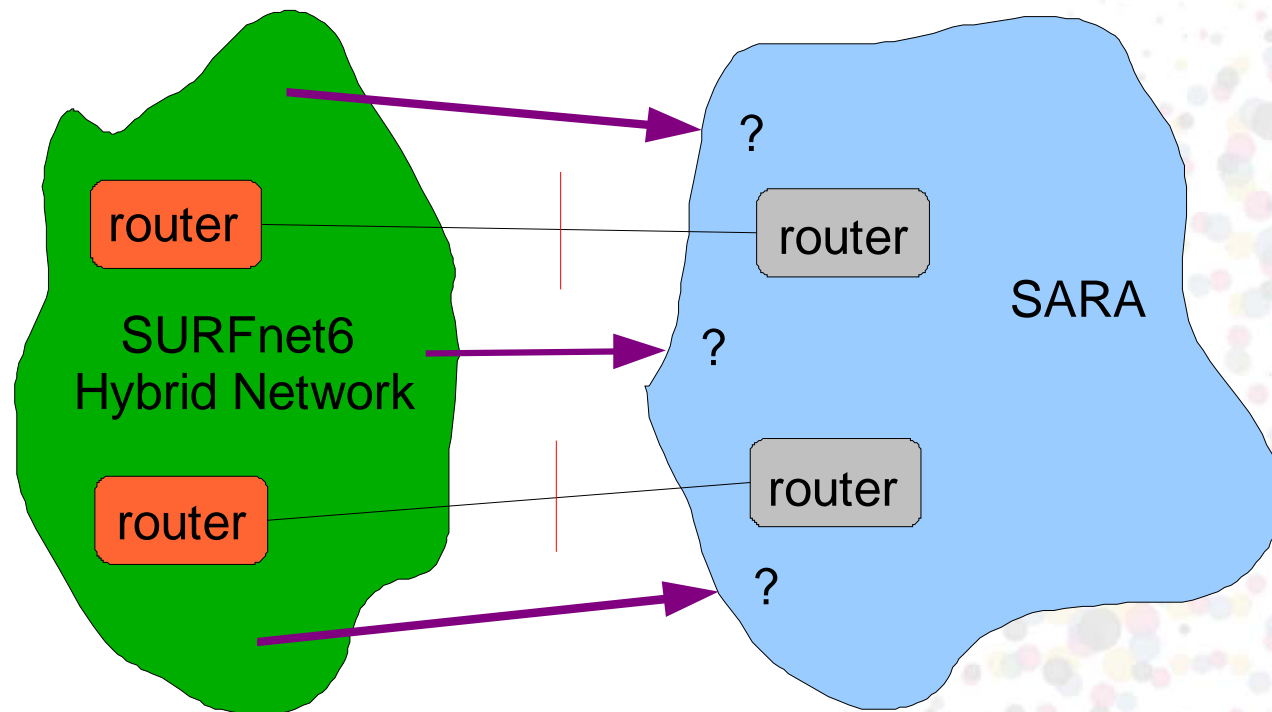
Traditional ISP connection

- Router to router connection
- A layer 3 IP interconnect



Introduction of Lightpaths

- ▀ Router to router connections
 - ▀ A layer 3 IP interconnect
- ▀ Lightpaths connection
 - ▀ Layer 1/2 connection to ?



Lightpath Challenges

- ▶ **Interconnect sites at L2 or at L3?**
- ▶ **How to handle security?**
- ▶ **How to handle addressing?**
- ▶ **How to protect against configuration errors and accidents at other site?**

Layer2 versus Layer3

- **L2 pros**
 - Cheap Ethernet switches
- **L2 cons**
 - No IP ACLs
 - Mixing of administrative domains
 - ▶ One broadcast domain, one IP subnet
 - Broadcast storms
- **L3 pros**
 - Well-known (we know how to do this between sites)
 - Supports ACLs and firewall
 - Easier fault resolution
 - ▶ Ping, traceroute, router reachability
 - Policy based routing
- **L3 cons**
 - Routers (and L3 switches) usually more expensive

SARA's requirements

- ▶ **Keep services separated**
 - ▶ Access to one service does not mean access to another service, unless explicitly allowed
- ▶ **No (accidental) connectivity between lightpaths via SARA**
- ▶ **No (accidental) Internet connectivity via SARA**
- ▶ **Solution must scale to multiple services and multiple lightpath peer sites**
- ▶ **Solution must support multiple 10G connections**
- ▶ **No big routing tables on the servers**
 - ▶ Only a default gateway
- ▶ **Segmenting the routing tables**
 - ▶ E.g. No LHCOPN prefixes in global routing table

Problems encountered in LHCOPN

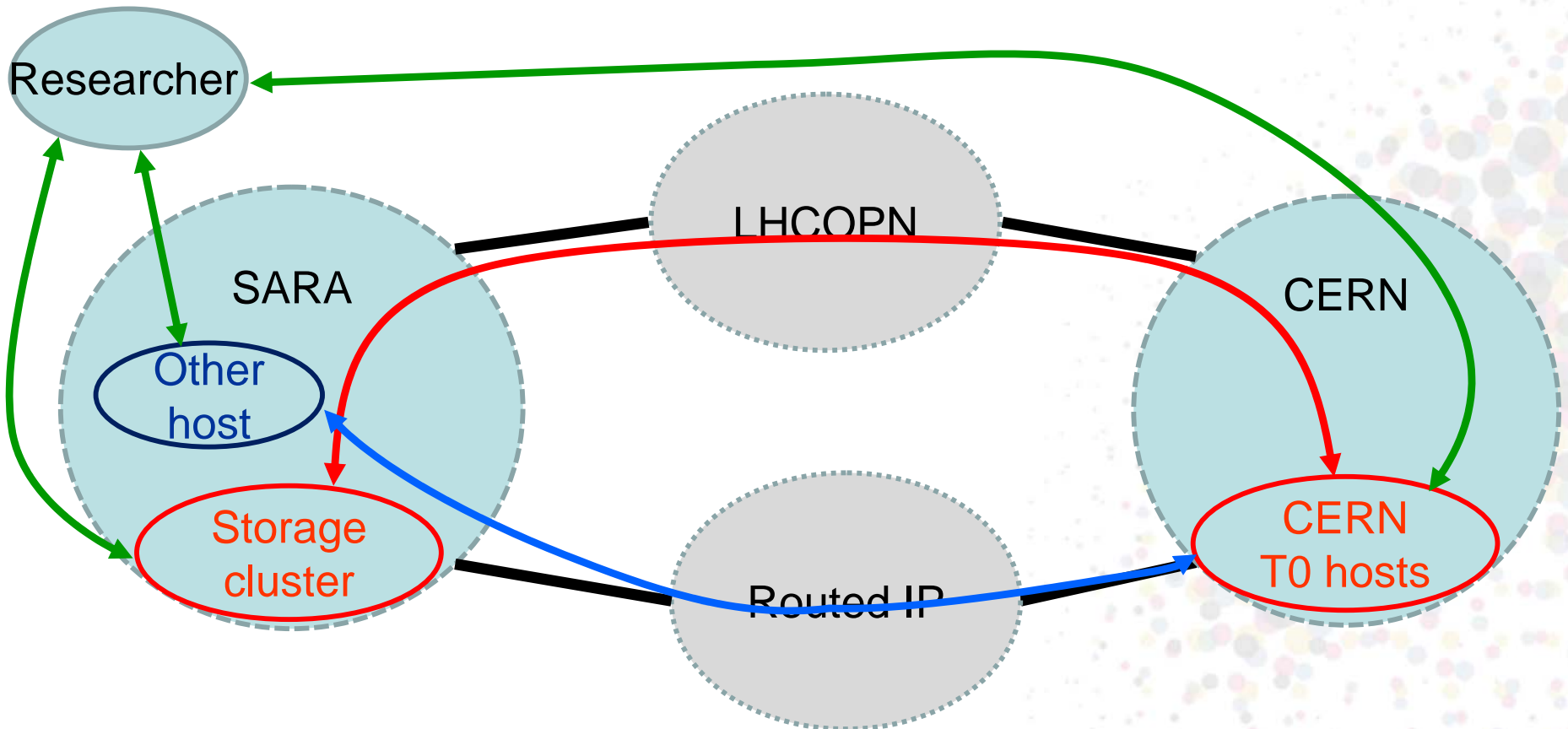
- ▶ Only storage servers traffic allowed on the LHCOPN
- ▶ Other hosts and servers must reach CERN via Internet
- ▶ Traditional destination based routing does not work
- ▶ We needed to find a good, scalable solution

SARA's choices

- ▶ **Interconnect at L3**
 - ▶ L2 only for a few very simple cases
- ▶ **BGP routing**
 - ▶ BGP detects when peer is unreachable
 - ▶ BGP needed when there are multiple paths
- ▶ **Routing segmentation**
 - ▶ Put each lightpath project in its own virtual router
 - ▶ Good way to keep projects and services separated
- ▶ **Powerful routing policy language**

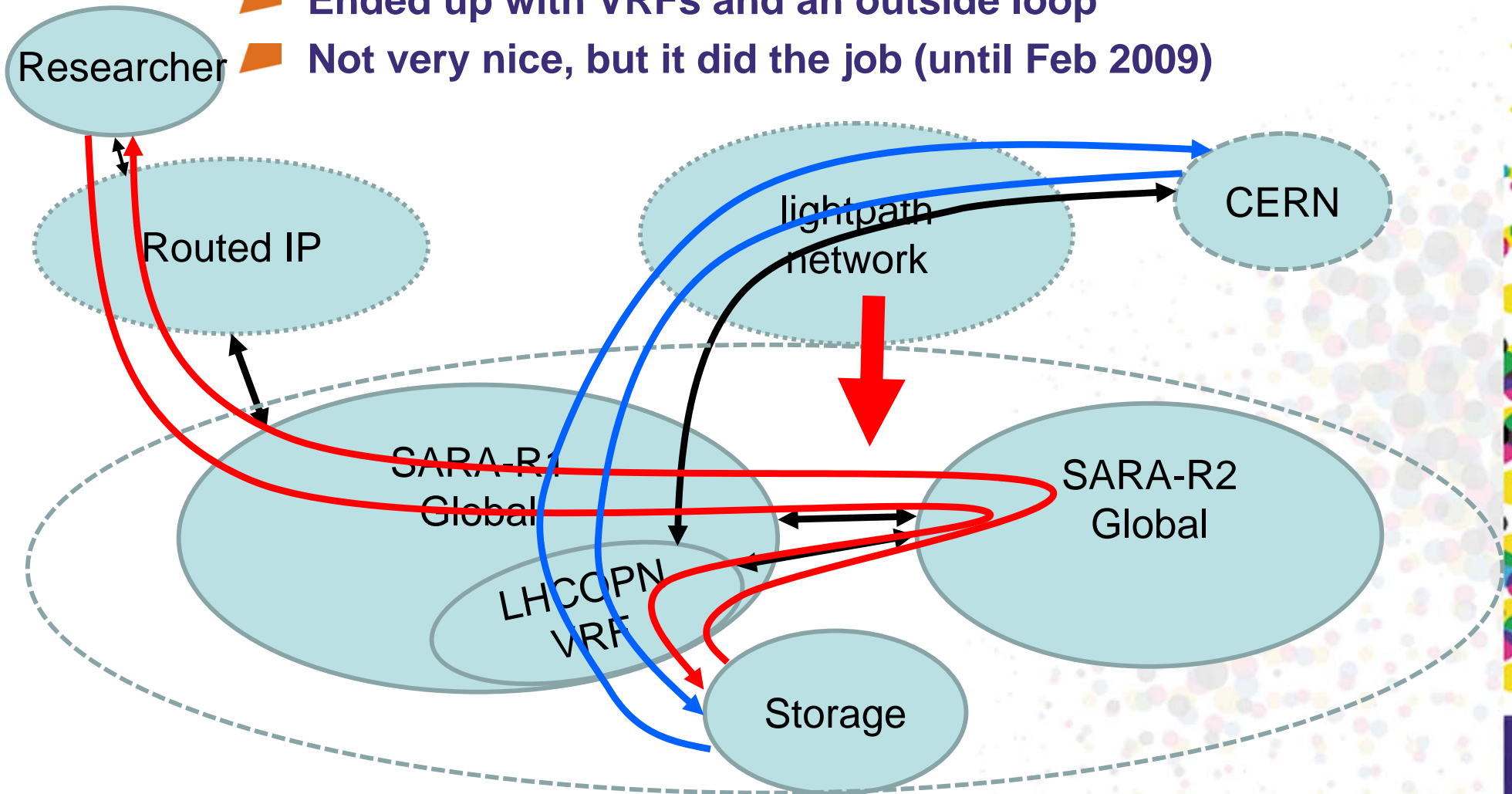
How it all started

- Back in April 2007 and January 2008 description of our problem/setup
- Struggling to keep traffic flows separated



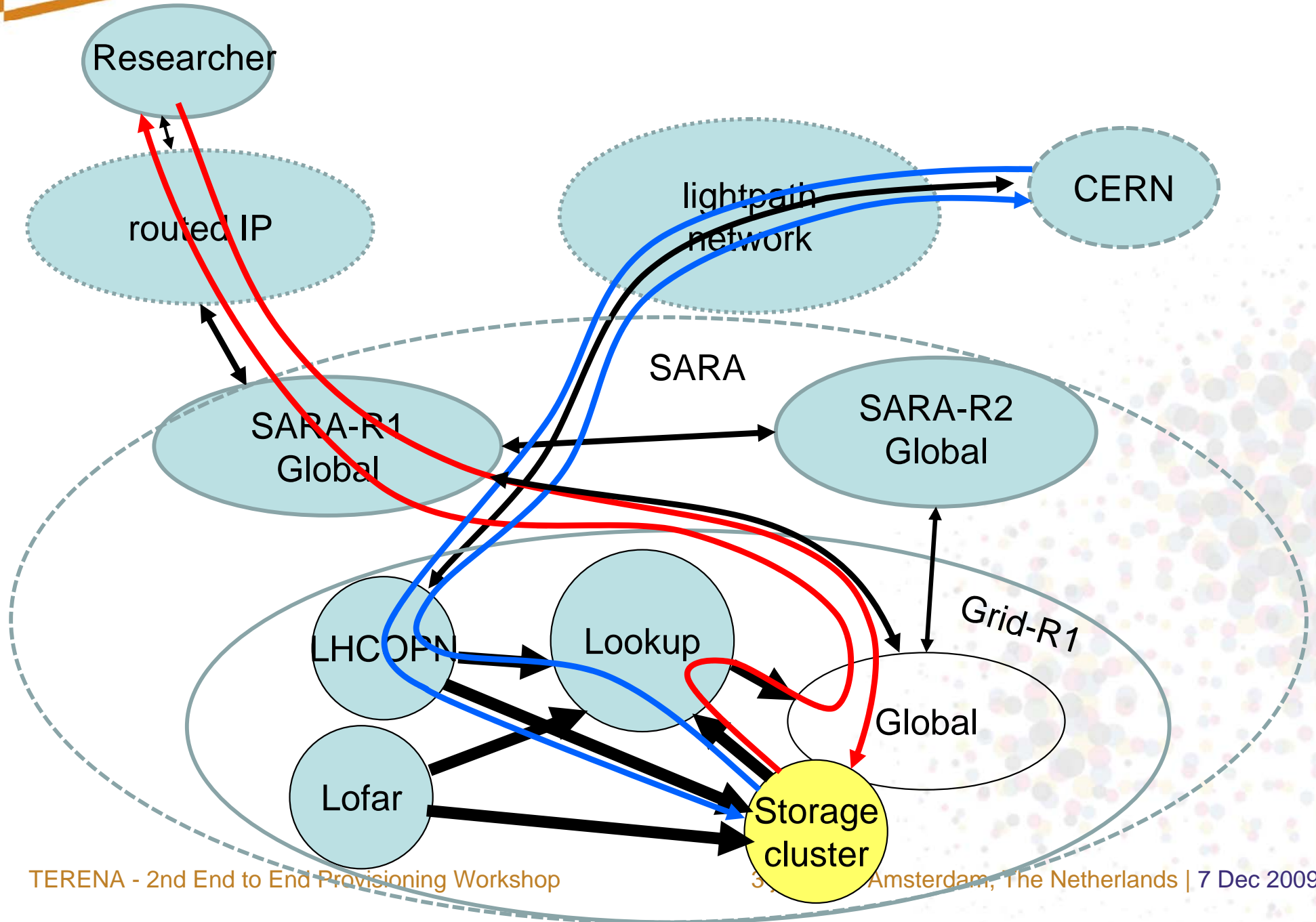
Network status until Feb 2009

- We thought of Policy-Based Routing
- Ended up with VRFs and an outside loop
- Not very nice, but it did the job (until Feb 2009)





Current network status





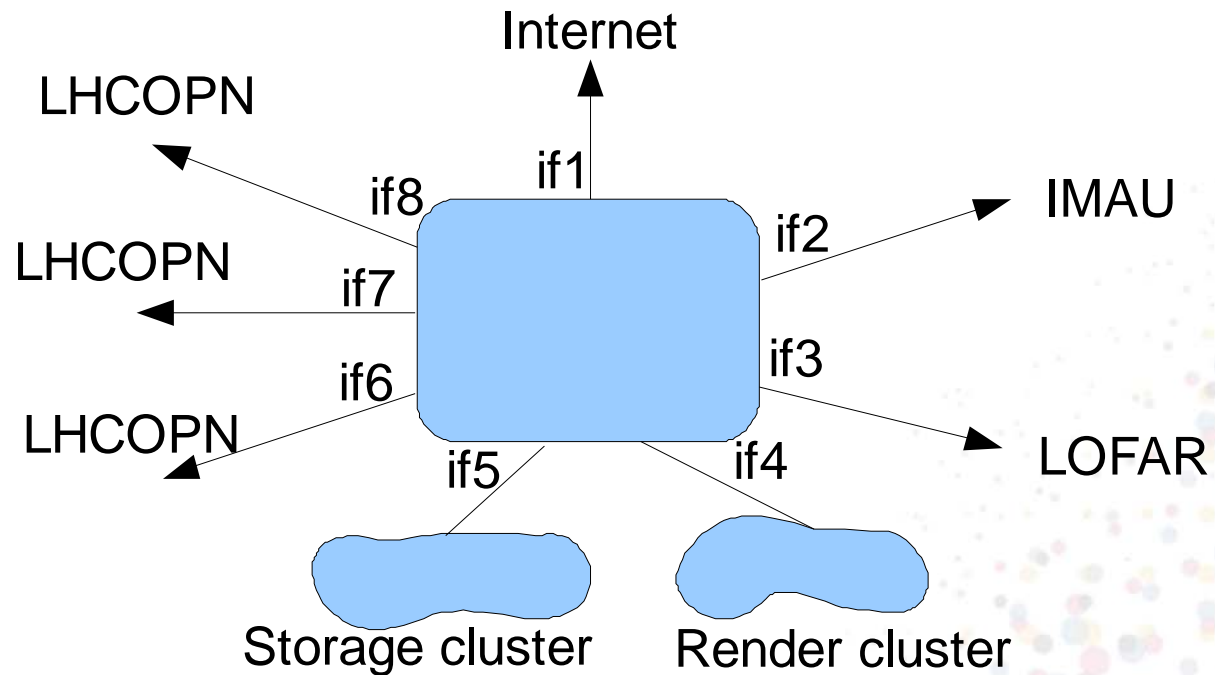
Juniper VRF config

```
routing-instances LHCOPN
  instance-type vrf;
  vrf-export lhcopn-export;
  vrf-target target:1126:2;
  routing-options rib LHCOPN.inet.0 static route 145.100.32.0/22 next-
  table inet.0
  routing-options auto-export

routing-instances storagecluster-shared
  instance-type vrf
  route-distinguisher 1126:999
  vrf-import storagecluster-import
  vrf-target target:1126:999
  routing-options static route 0.0.0.0/0 next-table inet.0
  routing-options auto-export

policy-statement lhcopn-export then community add lhcopn accept
policy-statement storagecluster-import from community lhcopn then accept
firewall family inet filter grid-lan-in term storagecluster-shared then
  routing-instance storagecluster-shared
```

Virtual Routing



Global Table: if1, if4, if5
 VR1 (LHCOPN): if6, if7, if8
 VR2 (IMAU): if2
 VR3 (LOFAR): if3

VRFs on a Juniper

- ▶ Every OPN in its own VRF with static to the storage cluster.
- ▶ Routes from a VRF exported with a tag.
- ▶ Lookup VRF imports routes from the other VRFs and has a default to the global.
- ▶ Storage cluster does its route lookup in the lookup VRF
- ▶ Route lookup takes the most specific route out of a VRF and otherwise uses the default to the global, in the global a sequential route lookup is performed, for the best route there.
- ▶ Since no routes between OPN VRFs are exchanged there is absolutely no risk of traffic leaking between OPNs
- ▶ In addition the global doesn't know anything of the OPNs
- ▶ And it's scalable!

How we got here

- ▶ **Wrote an extensive document with requirements and send it to several vendors and asked them to come up with a proposal.**
- ▶ **Juniper MX had a stronger and more scalable routing policy solution then the Cisco 6500/7600.**
- ▶ **We also asked the vendors for a POC session, the one we had with Juniper was very useful.**
- ▶ **In the end the Juniper MX960 was left**

Conclusions

- ▶ Supporting multiple lightpaths and multiple services is *not a trivial task*
- ▶ Virtual routing is a *scalable* way to handle the routing and keep services and lightpath peers *separated*
- ▶ Routing requirements often result in the choice for BGP



Questions?

Thank you, any questions...



Or send an email to peter.tavenier@sara.nl