

Deisa-GÉANT2 Workshop

# An Overlay Network for DEISA

**4<sup>th</sup> July 2005**

**München, Germany**

**Maarten Büchli**

**Network Engineering & Planning, DANTE**

## Outline

- **GÉANT2 network**
- **Deisa network: now and in the future**
- **Network services**
- **Ethernet switches**
- **Monitoring**

# GÉANT2 Network

# GÉANT2 Network

- Connectivity
  - Dark fibre
  - Managed wavelengths and circuits
- Equipment
  - DWDM transmission equipment
  - Optical crossconnects
  - IP routers
- Most of the network will be ready by end 2005

# Optical cross connect

## Features

Fundamentally SONET/SDH cross-connects (up to 640G)

All have GE client interfaces

10GE being developed (most appear during 2005)

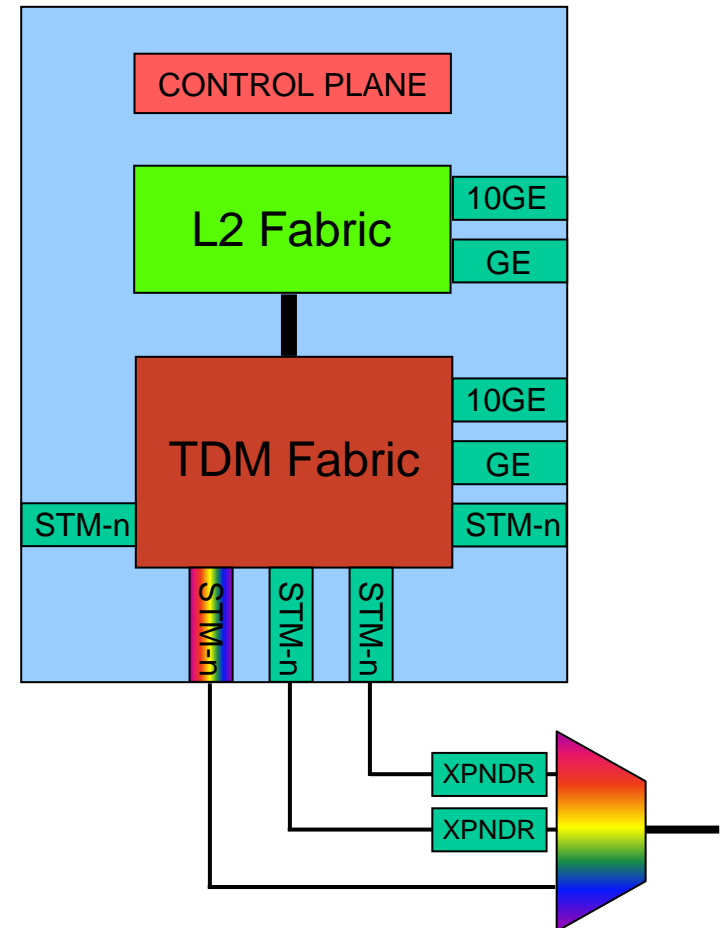
Support “recent” enhancements:

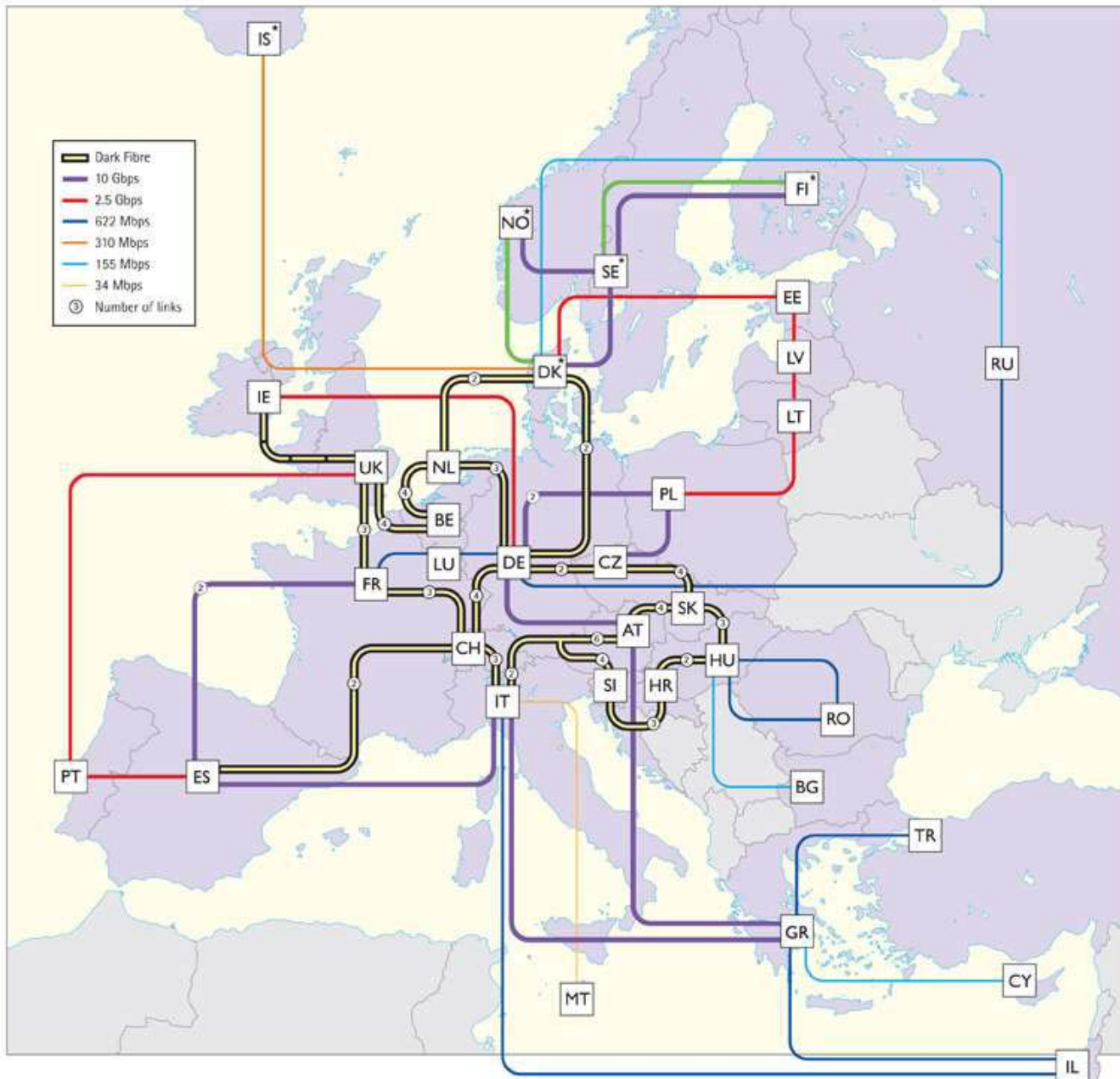
- GFP-F / VCAT / LCAS

L2 switching capability being added (typically during 2005)

Coloured interfaces on some

G.ASON/GMPLS control planes





# GÉANT2 Services

- IP/MPLS services
  - IPv4, IPv6 and MPLS
  - Unicast and multicast
  - Quality of Service
- Point-to-point services
  - Gigabit ethernet
  - 10 gigabit ethernet LAN PHY
  - SDH/SONET

# Protection point-to-point services

- Protection improves service availability
  - node/link failure
  - planned maintenance
- Protection implemented as two diversely routed lightpaths
  - provision the double amount of capacity
  - double the amount of interfaces
  - endpoints are responsible for failure detection and switchover

# Deisa network: where are we now?

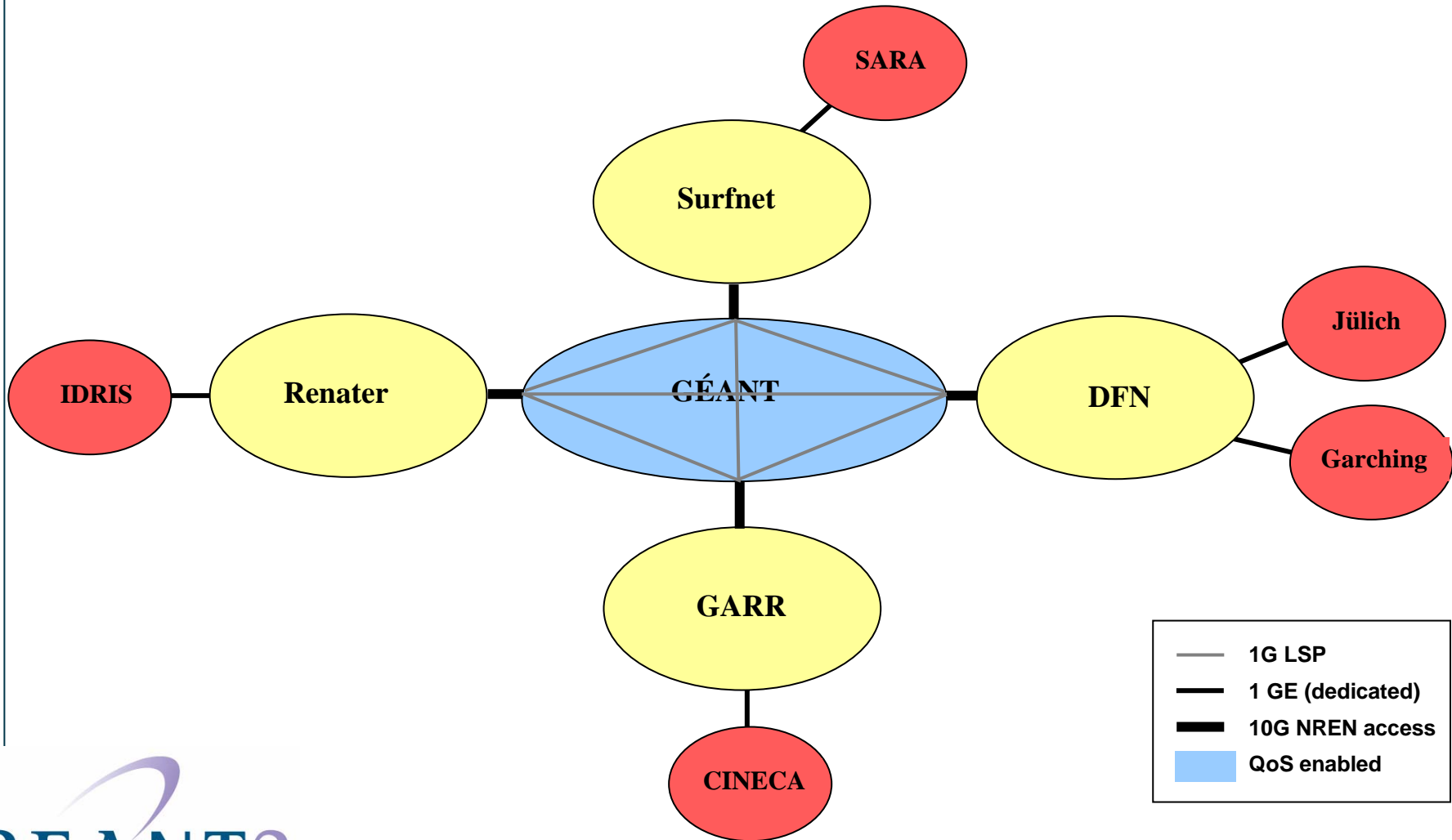
## Connected sites

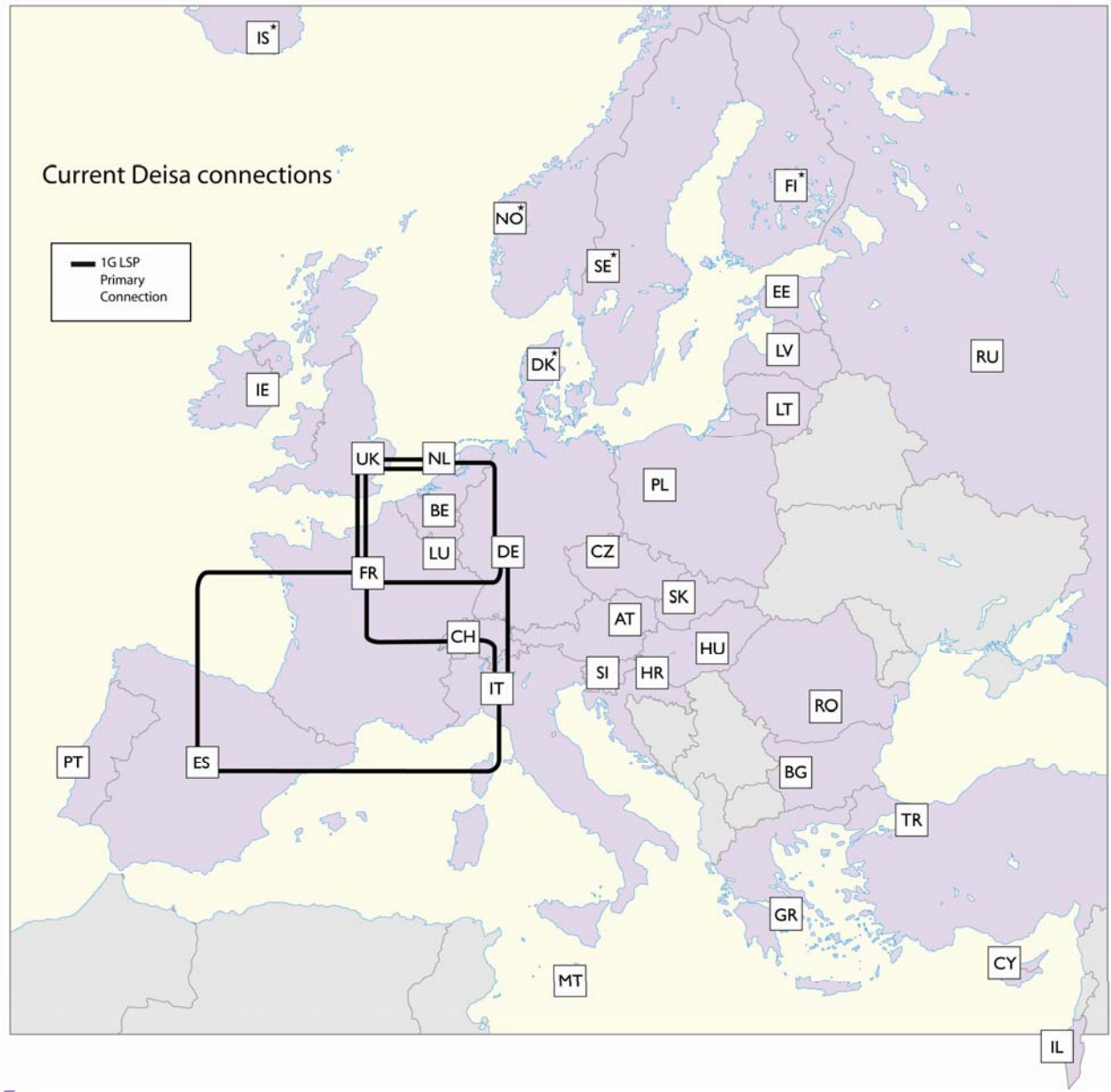
- Consorzio Interuniversitario (CINECA), Bologna, Italy
- Institut du Développement et des Ressources en Informatique Scientifique (IDRIS-CNRS), Orsay, France
- Forschungszentrum Jülich (FZJ), Jülich, Germany
- Rechenzentrum Garching of the Max Planck Society (RZG) , Garching, Germany
- SARA Computing and Networking Services , Amsterdam, The Netherlands

# DEISA Network

- Each end site has a 1GE connection to their NREN that is dedicated to DEISA
- NRENs use best-effort IP
- GÉANT a mesh of MPLS LSPs has been implemented
  - fully specified paths for traffic engineering
  - quality of service (premium)
  - backup LSPs

# DEISA Network





# Premium IP

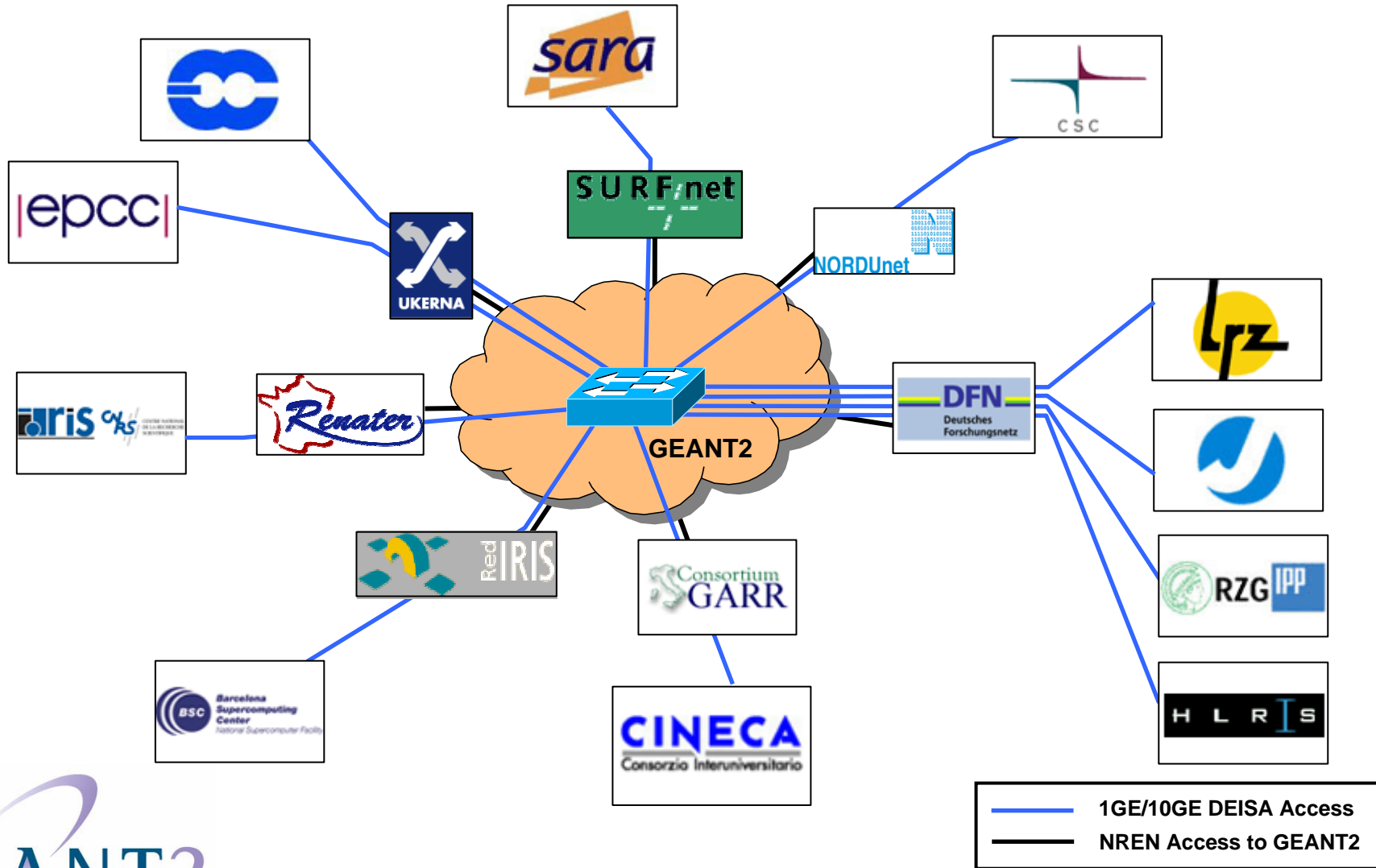
- Currently 1G for each LSP
- Maximum amount of Premium IP is 10% of the link
  - e.g. max 1G premium traffic on a 10G link
- Create some capacity for other reservations
  - can the capacity be decreased?
  - switch to best effort forwarding?

# Deisa network: Proposal for Phase 2

## Solution for Deisa

- 11 end sites
  - DE (4x), UK (2x), NL, FR, ES, IT, FI
- Star topology
  - scalable
  - does not optimise delay between Deisa sites
- 1GE or 10GE dedicated connections
- Ethernet switch in Frankfurt GÉANT2 POP
- Design recommendation at architecture workshop was to use unprotected connectivity
- Jumbo frames?

# Deisa topology proposal

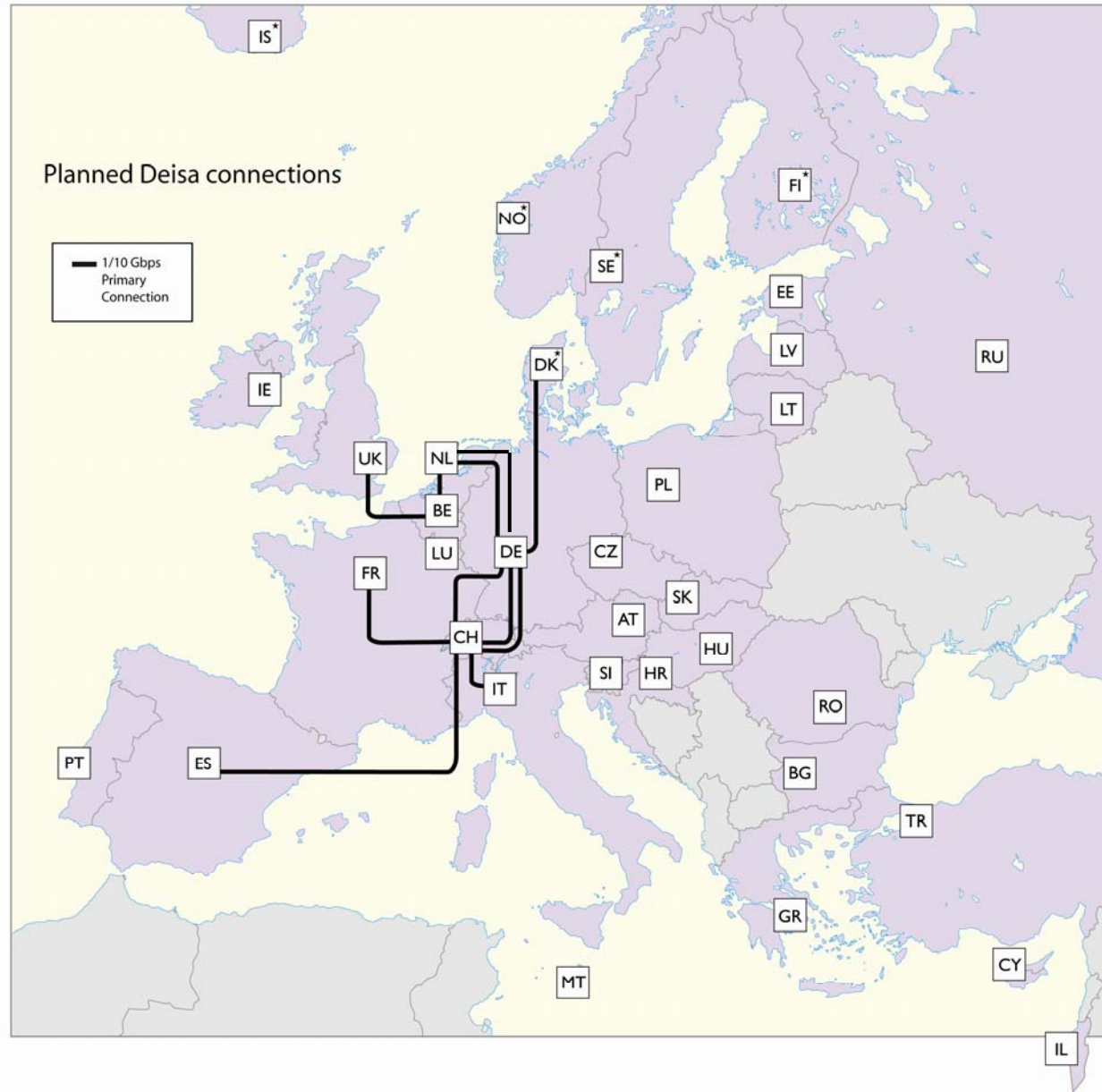


# NREN Network Services

- 1GE connections available now
  - DFN, via transmission system
  - GARR,
  - NORDUnet, SONET/SDH
  - Rediris, MPLS L2VPN (best effort)
  - Renater,
  - Surfnet, SONET/SDH
  - UKERNA, SONET/SDH
- 10GE connections become available between August 2005 and mid 2006

# GÉANT2 Network Services

- Gigabit ethernet
  - Based on switched infrastructure (GFP encapsulation)
  - NREN connects to Optical Crossconnect with a GE or SDH/SONET interface
  - Jumbo frame support
- 10 GE LAN PHY
  - Based on (unprotected) wavelengths
  - NREN connects to 10GE DWDM transponder
  - 10GE via switch at end 2005 (small rate mismatch between STM-64 and 10GE LAN PHY)



# Ethernet switches

# Requirements

- Number of GE and 10GE ports, etc.
  - 11 Deisa sites connected with 1/10G
  - will all sites connect with 10GE in the future?
  - more Deisa partners to join in the future?
- Timelines to upgrade from 1GE to 10 GE
- Performance (line rate per port, backplane capacity, no reordering of frames)
- Jumbo frame support
- Resiliency, dual power supply, etc.
- Other requirements, e.g. VLANs?

# Procurement

- Switch procured and managed by DANTE on behalf of Deisa
  - public tender procedure
- First research on switches shows:
  - Force10 E300/600
  - Foundry FastIron
  - Extreme Black Diamond
  - Alcatel 7450 ESS
  - Cisco Catalyst 6509
- Approximate cost for switch chassis, power supply, etc., 4x10GE and GE ports and optics: 100 k€
- Additional cost per 10GE port ~10 k€

# Monitoring

# Monitoring? What for?

- There are three main reasons for which we are interested in monitoring:
  - **Lightpath set-up**  
Verify that the lightpath is working between two sites and accept the service.
  - **Day-to-day monitoring**  
Have information about the availability and the behavior of the lightpath  
Be notified when changes occurs.  
Get historical information (used as reference).
  - **Diagnostic**  
Where is the problem? (end-host, lightpath, local LAN, server, OS, transport protocol, application, or a combination)  
If this is along the lightpath, where is it located?

# What to Monitor?

- Each network layer provide different but complementary information.
  - Even though, you won't always be able to monitor everything you wish.
  - Different stack of technology will provide different type of data.
- Note: Having a site connected to both a lightpath and to the Internet represent a major routing risk if the routing is not properly addressed
  - Lightpath traffic over Internet
  - Internet traffic over the lightpath

## What can you monitor per layers?

- IP layer
  - OWD/Jitter/packet loss/traceroute between end-sites
    - Usefulness mostly for providing the daily “heartbeat” of the lightpath
    - historical purpose (lightpath e2e uptime)
    - verify current behavior against past one
    - if the lightpath path changes
    - When a major problem occurs
    - Most of the time there should not be any changes
    - except if a major event is happening
    - or if the lightpath is carried over a packet switched network.
    - Traceroute ideal to check the IP path (over the lightpath or over the IP network)

# What can you monitor per layer?

- IP layer

- TCP/UDP throughput tests between end-sites

- Useful to verify that the lightpath is working fine

- Verify the throughput that a given threshold is reachable when bringing into service a lightpath, but also during it's operation. (easy for 1Gbps, much less easy for 10Gbps)

- Find out queuing burst problems which can impact large TCP transfer

- Gives indication when there are low rate errors

- Pretty handy to diagnose a problem (along the network or end-to-end) then, if other of such boxes deployed along the lightpath are available, along which to sub-path the problem is located.

- Not easy to deploy such boxes along the path

- What about having one connected to the Deisa ethernet switch? This would allow tests between and end-site and the central switch.

- Main drawback, it generates a large amount of traffic which can impact the production traffic. (to be run at well chosen scheduled time)

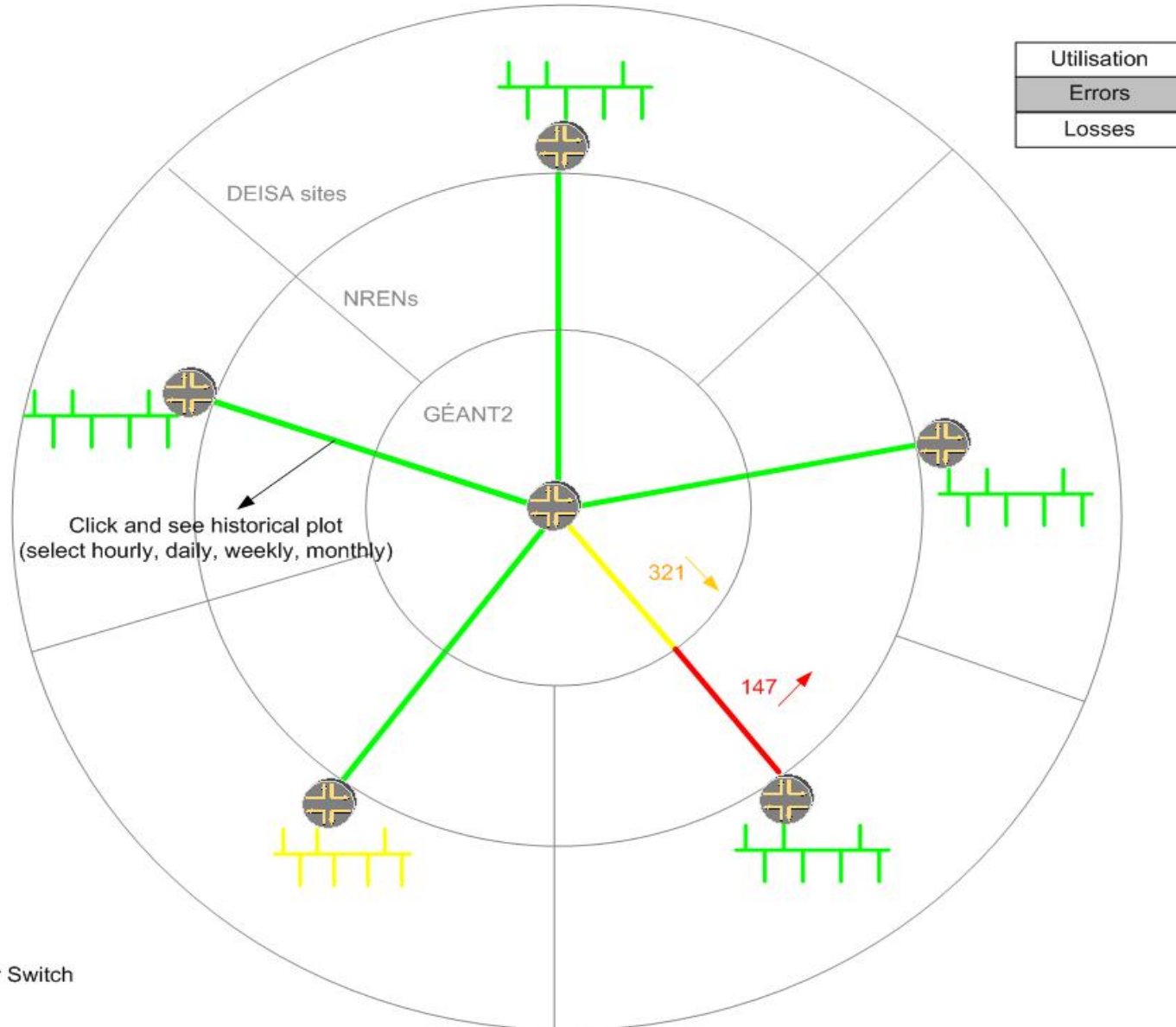
## What can you monitor per layer?

- Ethernet
  - Provides information about the load of a circuit/VLAN or the amount of loss/errors (drops, runts, CRC) between two switches.
  - Drawback: on a WAN, if the problem is coming from the circuit itself, it doesn't indicate where a problem exactly is (difficult to troubleshoot).
  - No capabilities of verifying the traffic from one end-site to another one (except if full mesh of VLANs or if switch with IP capabilities)
- SDH
  - Ease the fault localisation along a circuit, but don't provide any information about the amount of traffic seen.
- WDM
  - Less information to find out where a problem is coming from along a circuit than SDH.

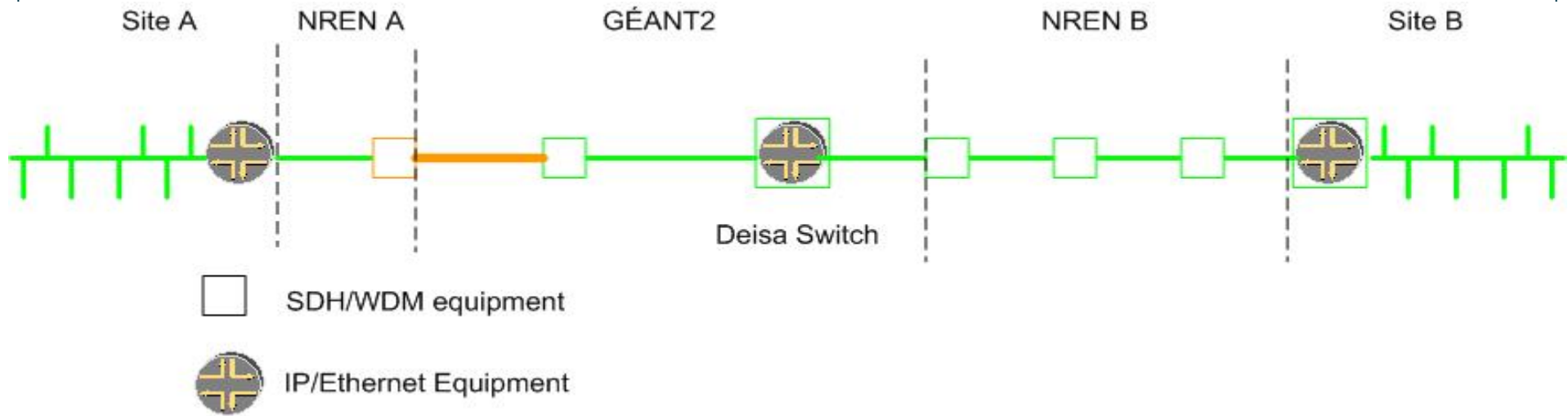
## What can you monitor per layer?

- The University of Erlangen is working within the GN2-JRA1 to set up a server which can do both the OWD and the TCP throughput tests.
- GN2-JRA1 is also building a framework to exchange monitoring information using the GGF NM-WG request and reply schema. So there is no need to adapt a visualisation tool to different network APIs a single one is used.
- The framework will also export sub-IP information.

# Lightpath Health



# Path overview



# Matrix

	A	B	C	D
Site A		3	29	-3
Site B	2		6	10
Site C	40	-6		3
Site D	-78	10	3	

Packet loss
OWD
Jitter
TCP Throughput
IP traffic

Click and see historical plot  
(select hourly, daily, weekly, monthly)

- The matrix can also be visually displayed (as it was done for the ATM PVCs the former purgatorio tool)  
Ideally, the tool maintainer should have an interface to add a new sites (location, name, IP addresses) and have the different visualisation automatically updated

**Questions...**