LHC high-level network architecture

Erik-Jan Bos

Director of Network Services

SURFnet, The Netherlands

T0/T1 network meeting

NIKHEF/SARA, Amsterdam, The Netherlands; April 8, 2005

GigaPort

SURF net

GigaPort

Contents

- **History and mission**

- **A proposed high-level architecture**

- **Further steps**

SURF/net

- **January 20 & 21, 2005 meeting in Amsterdam chaired by David Foster:**
  - Presentations by the experiments
  - Presentations by some network orgs
  - Conclusion: Move from bottom up to top down
  - Consensus on small task force for proposing LHC high-level network architecture

- **Initial proposed people: Don Petravick, Kors Bos, David Foster, Paolo Moroni, Edoardo Martelli, Roberto Sabatino, Erik-Jan Bos (volunteered to be chair)**

GigaPort

First steps to the architecture

- **Assumptions:**
  - **High-volume data streams**
  - **Continuous data streams**
  - **Keep It Simple**

- **Stay as low in the stack as you can (see January presentations)**
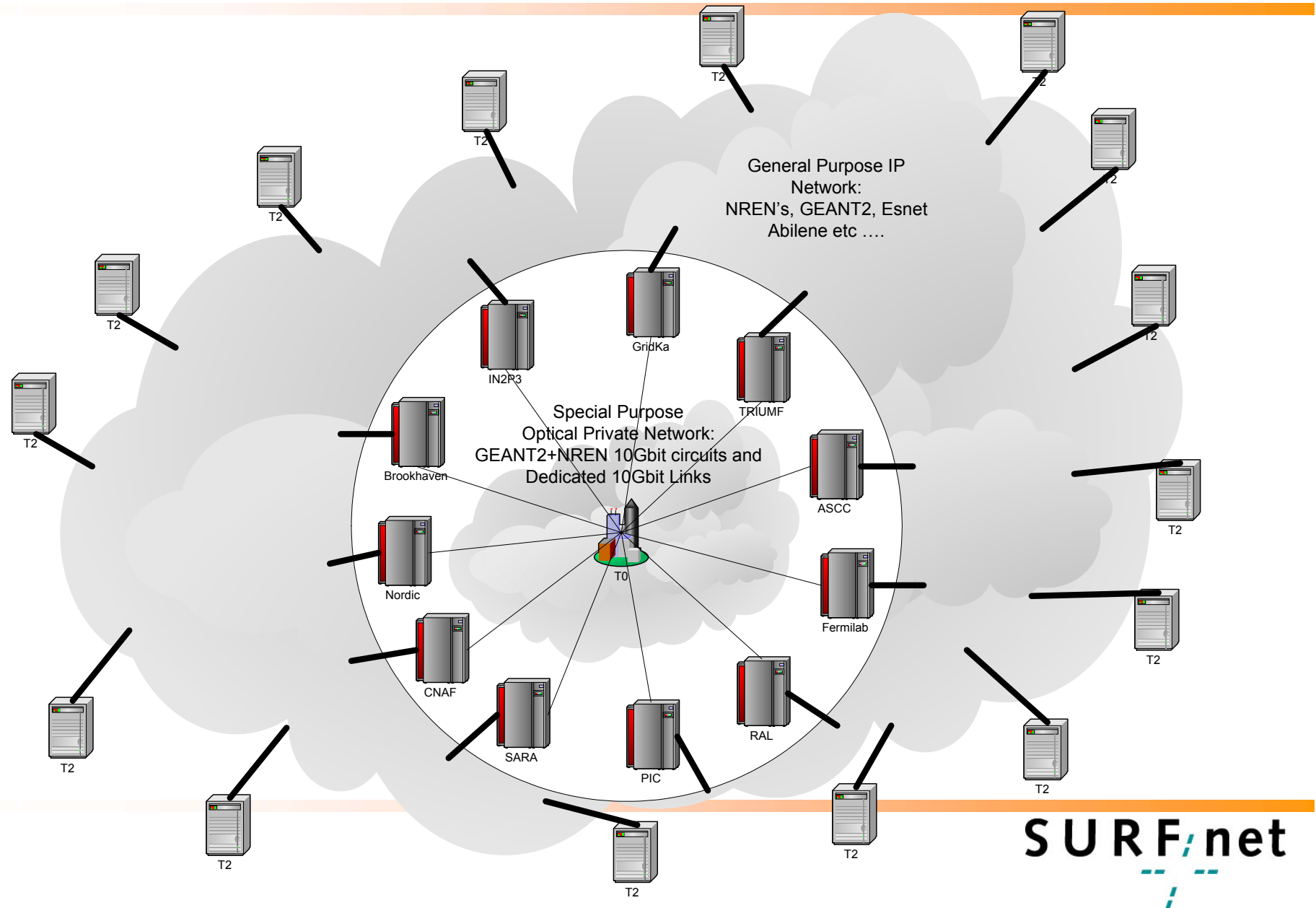
SURF\net

Security considerations

- **Important to address security concerns already in the design phase**

- **Architecture will be kept as protected as possible from external access**

- **At least in the beginning, access from trusted sources (i.e. LHC prefixes) will not be restricted**
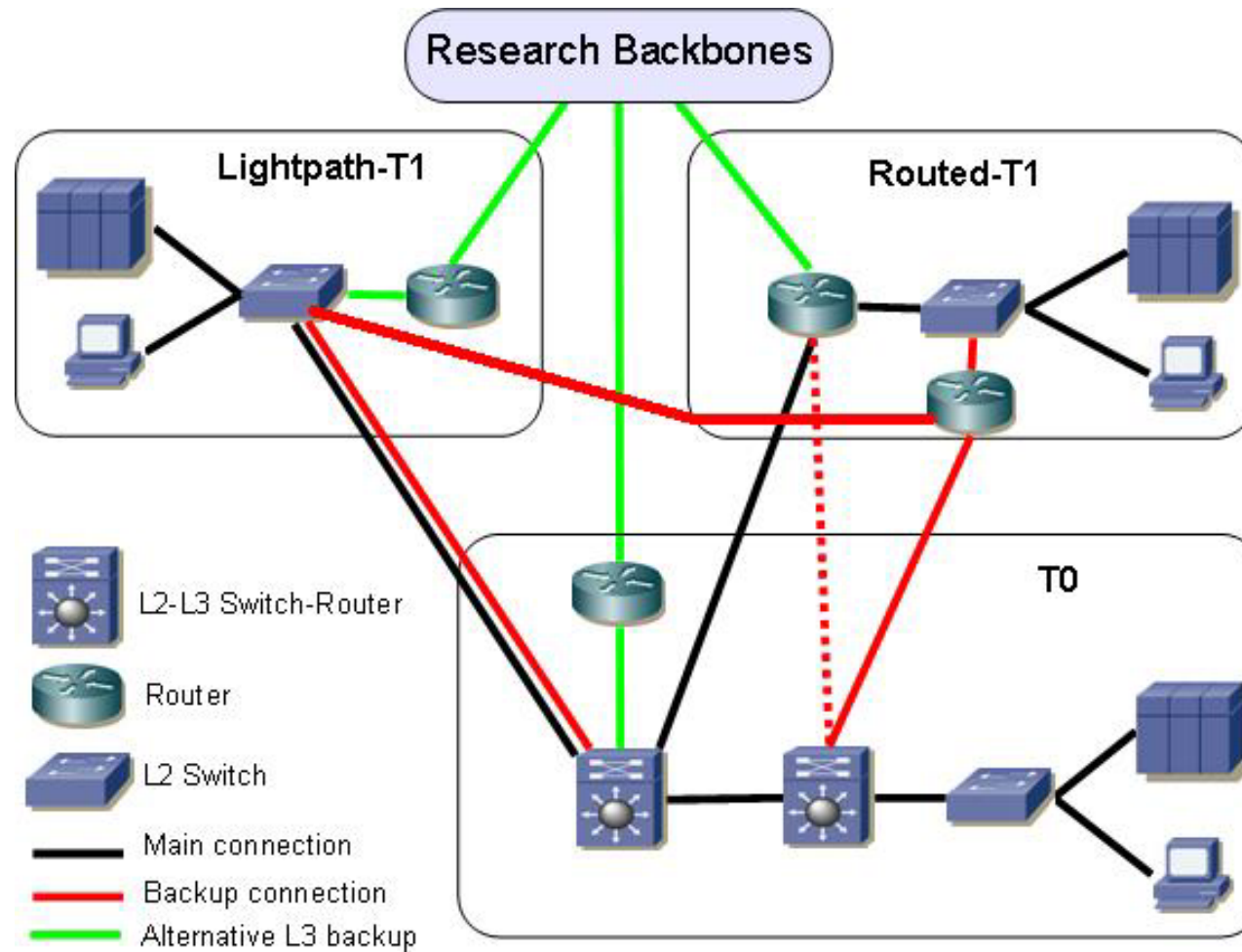
A proposed high-level architecture (1)

- **Optical Private Network, consisting of dedicated 10G paths between T0 and each T1, two flavors:**
    - **"Light path T1"**
    - **"Routed T1"**

- **Special measures for back-up for T0-T1, to be filled-in later**

- **T0 preferred interface is 10Gbps Ethernet LAN-PHY**

A proposed high-level architecture (2)

General Purpose IP
Network:
NREN's, GEANT2, Esnet
Abilene etc ….

Special Purpose
Optical Private Network:
GEANT2+NREN 10Gbit circuits and
Dedicated 10Gbit Links

GridKa

IN2P3

TRIUMF

Brookhaven

ASCC

Nordic

Fermilab

CNAF

T0

SARA

PIC

RAL

T2 (multiple)

A proposed high-level architecture (3)

Light Path definition

- **Definition:**
  "(i) a point to point circuit based on WDM technology or (ii) a circuit-switched channel between two end points *with deterministic behaviour* based on TDM technology or (iii) concatenations of (i) and (ii)"

- **So: A layer 1 connection with Ethernet framing**

- **Document contains examples**

Light Path T1

- **Uses a dedicated light path, at 10G, to the interface at T0**

- **Possible implementation for a European T1:**
  - **10GE LAN PHY at T0 awaiting the T1**
  - **10GE LAN PHY at T1 for the connection to T0**
  - **T1 connects to NRN at 10GE LAN PHY**
  - **NRN connects to GÉANT2 at 10GE LAN PHY or 10G SONET (with GFP-F mapping)**
  - **GÉANT2 connects to T0 at 10GE LAN PHY**

- **CIDR address block of T1 on this interface**

Routed T1

- **BGP peering established between the T0's router and the T1's router site using external BGP (eBGP)**

- **Possible implementation for a non-European T1:**
  - **10GE LAN PHY at T0 awaiting the T1 (10GE WAN PHY to be discussed with CERN, to avoid extra box in Geneva)**
  - **Connection to an intercontinental wave from a commercial carrier**
  - **Connected to a router of the NRN on 10GE WAN PHY**
  - **T1 connected to NRN at 10G**

What does this mean for you? (1)

- **T1 will be responsible for organising the physical connectivity from the T1's premises to the T0's computer centre**

- **Party to contact and to get involved: Your local NRN (European NREN, ESnet, CANARIE, or ASnet)**

- **European NRENs:**
  - **Will sync with DANTE**
  - **DANTE to connect to T0**
  - **One primary 10G light path per Tier1 and a back-up path**

What does this mean for you? (2)

- ## Non-European Tier1s, e.g.:
  - Have dedicated bandwidth into CERN, or
  - Connect to an open optical exchange in Europe, like NetherLight, CzechLight, NorthernLight or UKLight and ask DANTE for a 10G light path between the *Light and CERN

SURF/net

GigaPort

Envisioned T0-T1 provisioning

Name of T1	LP/Routed	T0 Interface and intervening networks
ASCC	Routed	10GE LAN, ASNet, NetherLight|GÉANT2
BNL	Routed	10G SONET, LHCnet*, ESnet
CNAF	Light Path	10G LAN, GÉANT2, GARR
FNAL	Routed	10G SONET, LHCnet*, ESnet
IN2P3	Light Path	10G LAN, RENATER3
GridKa	Ligth Path	10G LAN, GÉANT2, X-WiN
SARA	Light Path	10G LAN PHY, GÉANT2, SURFnet6
NorduGrid	Light Path	GÉANT2, NORDUnet, Nordic NRNs ?
PIC	Light Path	10G LAN, GÉANT2, RedIRIS, Catalan Net
RAL	Light Path	10G LAN, GÉANT2, SuperJANET5
TRIUMF	Light Path	CA*net 4, ?

* = CALTECH-CERN transatlantic links

SURFnet

Planning

- **Start date for physics traffic is June 2007**

- **T1s are encouraged to proceed with provisioning well before that date, ideally already within 2005**

- **Nevertheless, T1s must be ready at full bandwidth not later than Q1 2006, to be in place for the mid-2006 SC.**

"LHC Network Operations", discussion

- **Distributed Operations:**
  - Every Tier is responsible to monitor and assure the functionality of its own equipment and line(s)
  - Parties involved: Tiers, DANTE, NRNs, *Light operators
  - Communication infrastructure in place

- **Centralised Operations:**
  - LHC Helpdesk and/or NOC
  - Ultimately, the LHC NOC does all configuration, trouble shooting, and fixing

- **Hybrid Operations:**
  - Central LHC health & volume monitoring capability
  - Each Tier or network organization has responsibility

A word on future growth

- **Some light path math (theoretical):**
  - 10 Gbit/s   ~ 10^{14} byte/day or 100 Tbyte/day
  - Eleven 10G light paths -> more than 1 petabyte/day or roughly half an exabyte/year

- **In case a 10G is not sufficient:**
  - Order a second 10G between T0 and T1
  - Preferably on a separate physical path
  - Architecture fully allows for this

SURFnet

Items for further discussion

- **Agree with T0 about the physical interface for the T0-T1 link**

- **Verify that the proposed addressing set-up is compatible with the grid software (e.g. can the servers be grouped in the same CIDR block?)**

- **Inform T0 about the AS number used**

- **Check if it is possible to establish an environment without default route**

- **Verify if the proposed security model is compatible with the Grid applications**

- **Decide a backup strategy in case an alternate path at full speed is not available: tolerate a few hours stop or prefer low performance on general purpose research backbones.**

Next Steps

- **Get comments in on version 1.0 of the document**

- **Together with results of the discussion write the final version 2.0**

- **T1s must start to work with their NRNs**

- **NRNs must work on dedicated bandwidth with DANTE for GÉANT2 light paths and/or commercial carriers and/or open optical exchange operators**

Acknowledgements

- **Thanks to (alphabetically):**
  - Kors Bos (NIKHEF)
  - Hans Döbbeling (DANTE)
  - David Foster (CERN)
  - Bill Johnston (ESnet)
  - Donna Lamore (FNAL)
  - Edoardo Martelli (CERN)
  - Paolo Moroni (CERN)
  - Don Petravick (FNAL)
  - Roberto Sabatino (DANTE)
  - Karin Schauerhammer (DFN)
  - Klaus Ullmann (DFN)

Questions?

SURFnet