GigaPort-RON SAC 2008

From Routed to Hybrid Networking

Cees de Laat

University of Amsterdam

System &Network Engineering @ UvA

update 2008

- group has 4 sections
  - Advanced Networking (GP, EU, TNO)
    - Paola Grosso
  - Security (GP, EU, VL-e, SurfWorks)
    - Guido van ‘t Noordende
  - Sensor Grids - Intelligent networks (TNO)
    - Rob Meijer
  - Master SNE education (GP)
    - Karst Koymans
- 25 people - 19 fte
  - people leaving (LG, FD, DM, MC)
- Home @ Science Park Amsterdam, co-located with:
  - NIKHEF (together with SARA LHC Tier-1 center, BigGrid)
  - SARA (SN6-NOC, NetherLight, SN6-core location, LightHouse)
  - AMS-IX
  - UvA Science faculty (Dutch e-Science program VL-e)

GP - Plans 2004-2008

1. Hybrid networking structure
   - Network Architecture
   - Optical Internet Exchange Architecture
   - Network Modeling <NDL, Pathfinding>
   - Fault Isolation
2. Network transport protocols
   - UDP - TCP
   - Protocol testbed
   - LinkLocal Addressing
3. Optical networking applications
   - StarPlane
   - eVLBI
   - Smallest University for proof of concepts
   - CineGrid
   - CosmoGrid
4. Authorization, Authentication and Accounting in Networking and Grids
   - AAA & schedule server
   - WS security
   - Multi domain token based implementations
   - Cross domain LightPath setup
5. Testbed LightHouse, SC0X, iGrid, GLIF, OGF, Terena, ...

A. Lightweight users, browsing, mailing, home use

Need full Internet routing, one to all

B. Business/grid applications, multicast, streaming, VO's, mostly LAN

Need VPN services and full Internet routing, several to several + uplink to all

C. E-Science applications, distributed data processing, all sorts of grids

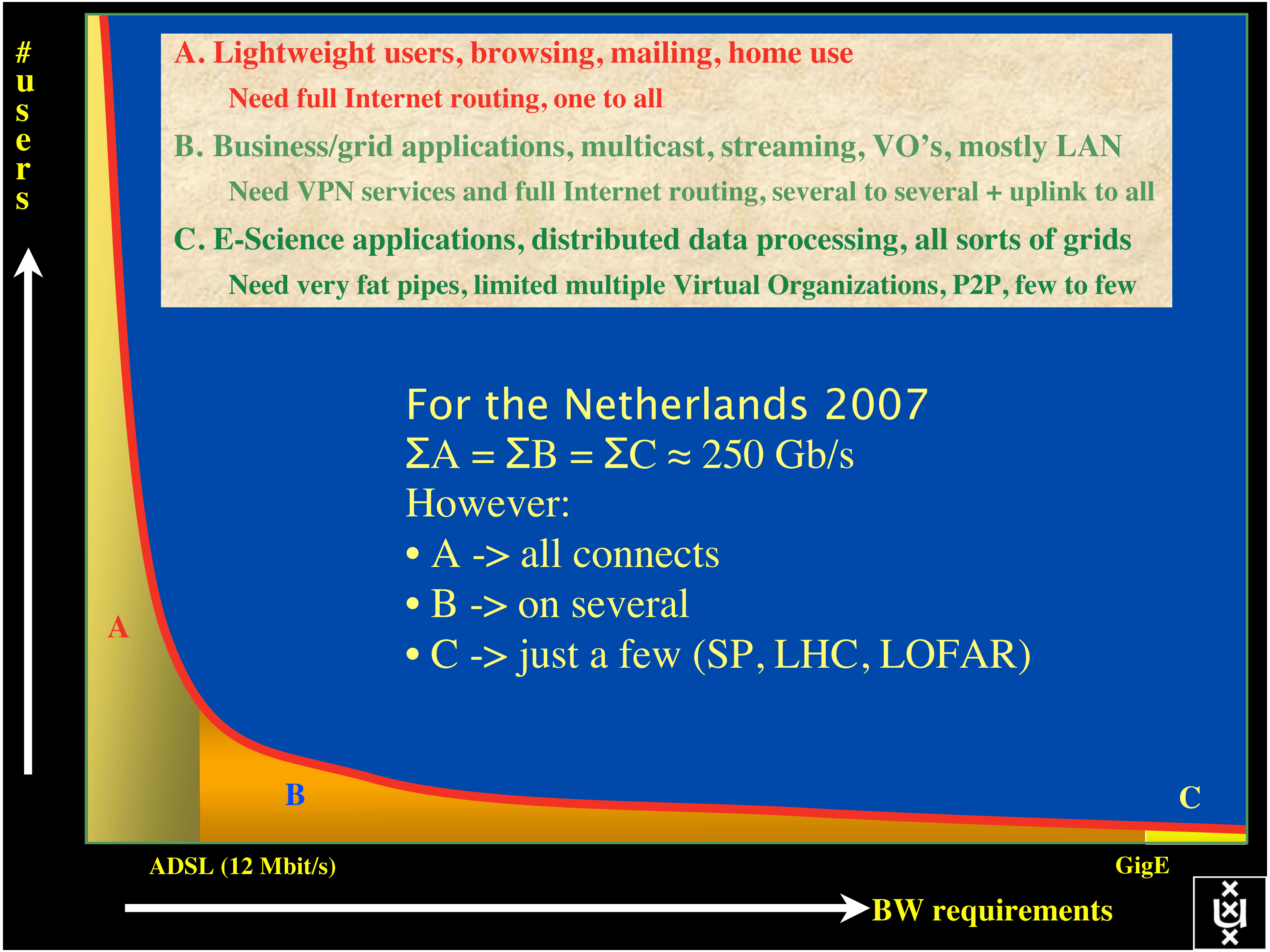
Need very fat pipes, limited multiple Virtual Organizations, P2P, few to few

For the Netherlands 2007

ΣA = ΣB = ΣC ≈ 250 Gb/s

However:

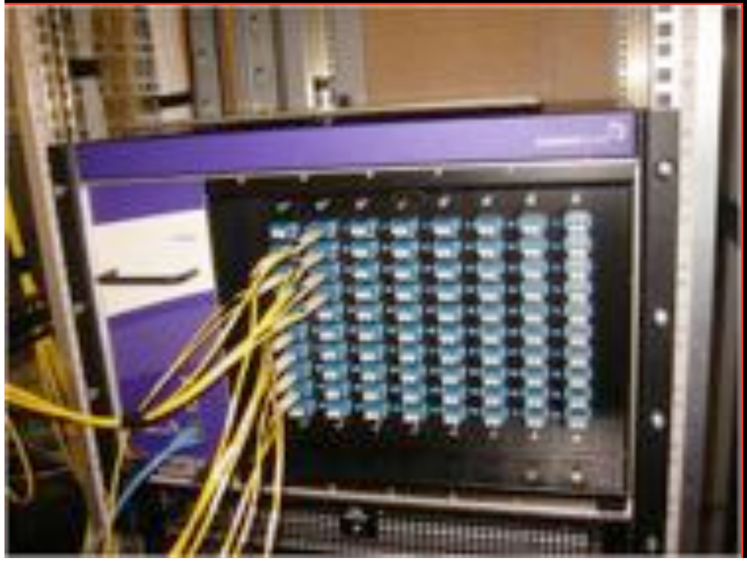
- A -> all connects
- B -> on several
- C -> just a few (SP, LHC, LOFAR)



Towards Hybrid Networking!

- Costs of photonic equipment 10% of switching 10 % of full routing
  - for same throughput!
  - Photonic vs Optical (optical used for SONET, etc, 10-50 k$/port)
  - DWDM lasers for long reach expensive, 10-50 k$
- Bottom line: look for a hybrid architecture which serves all classes in a cost effective way
  - map A -> L3 , B -> L2 , C -> L1 and L2
- Give each packet in the network the service it needs, but no more !

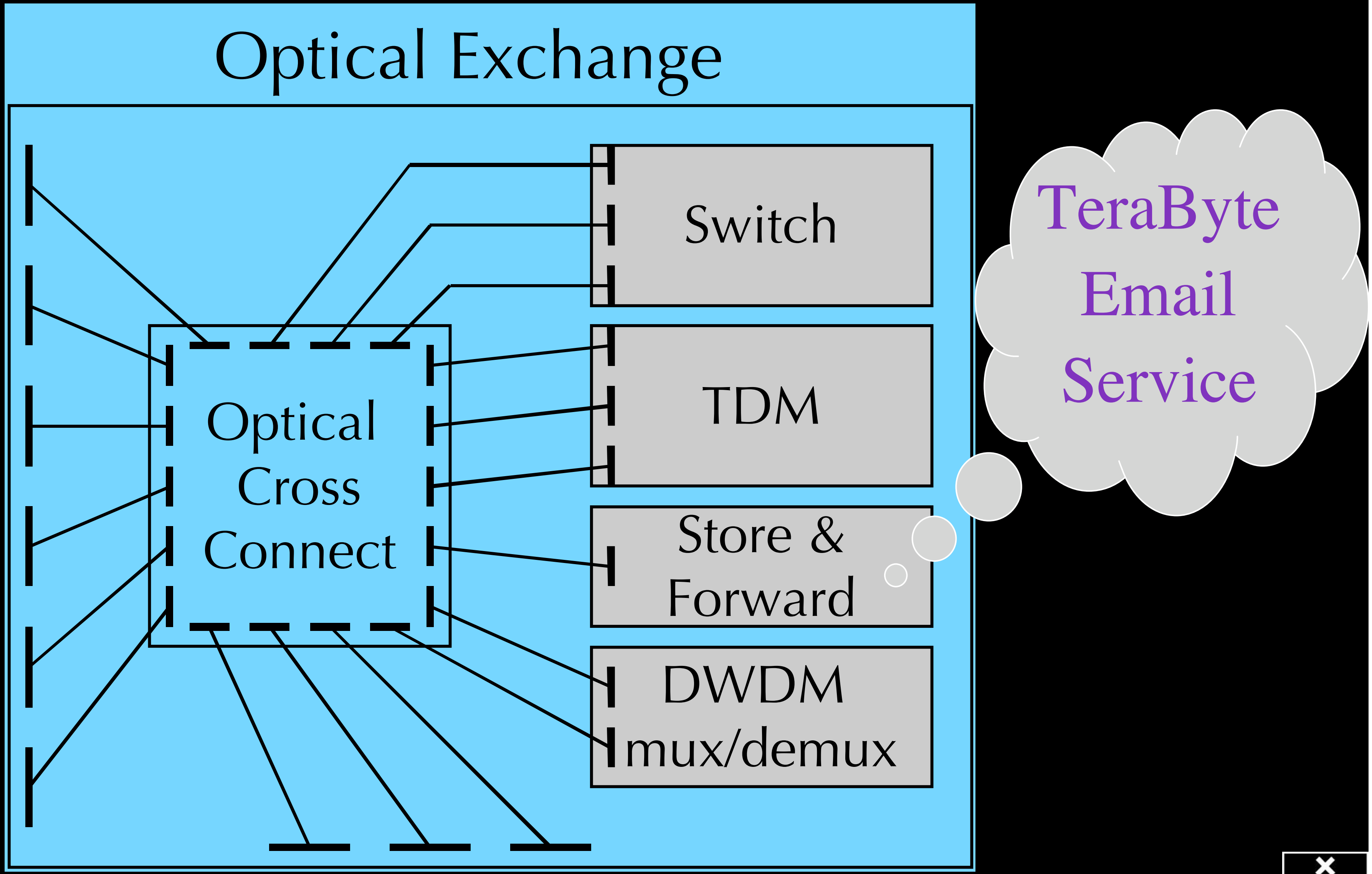
L1 ≈ 2-3 k$/port



L2 ≈ 5-8 k$/port

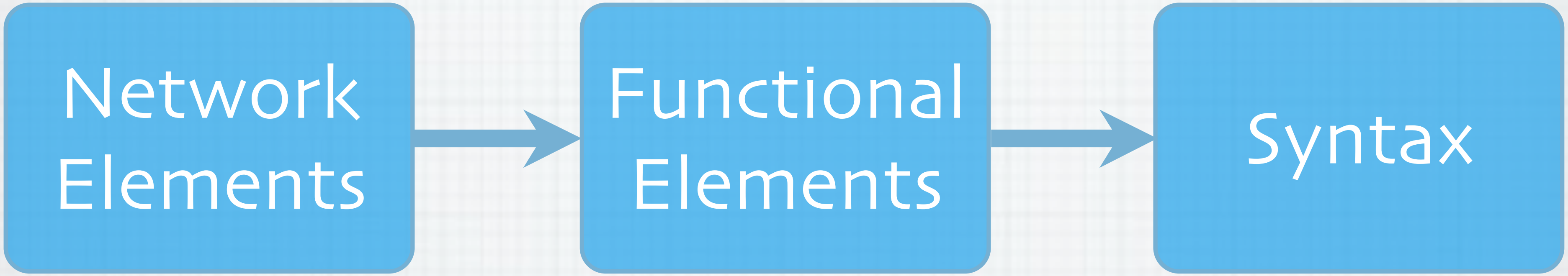
L3 ≈ 75+ k$/port

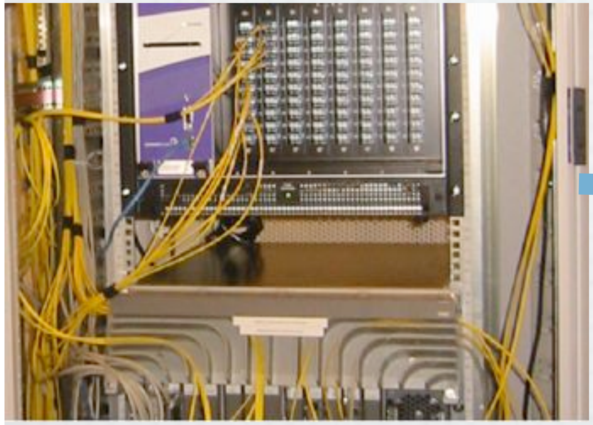
Optical Exchange as Black Box

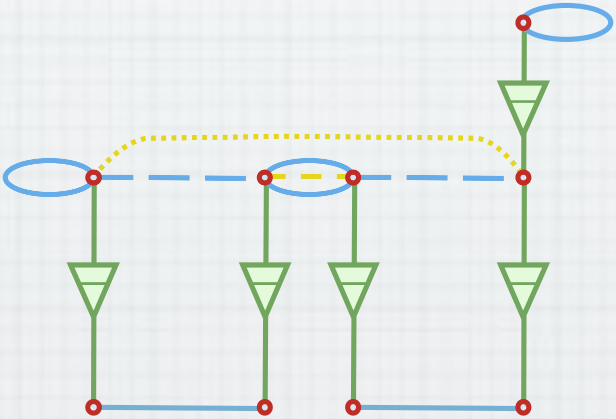


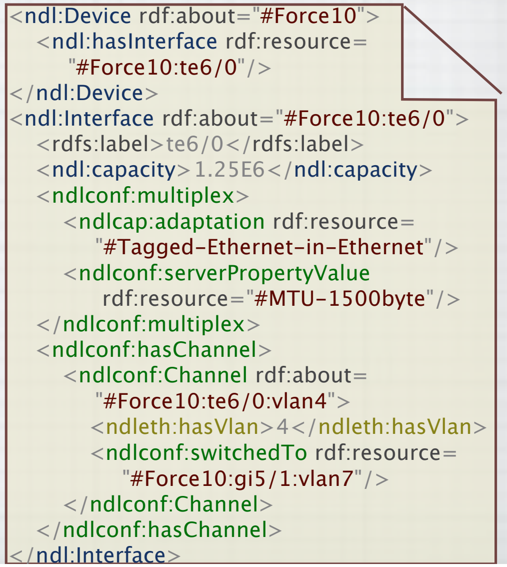
Ref: gridnets paper by Freek Dijkstra, Cees de Laat

The Modelling Process





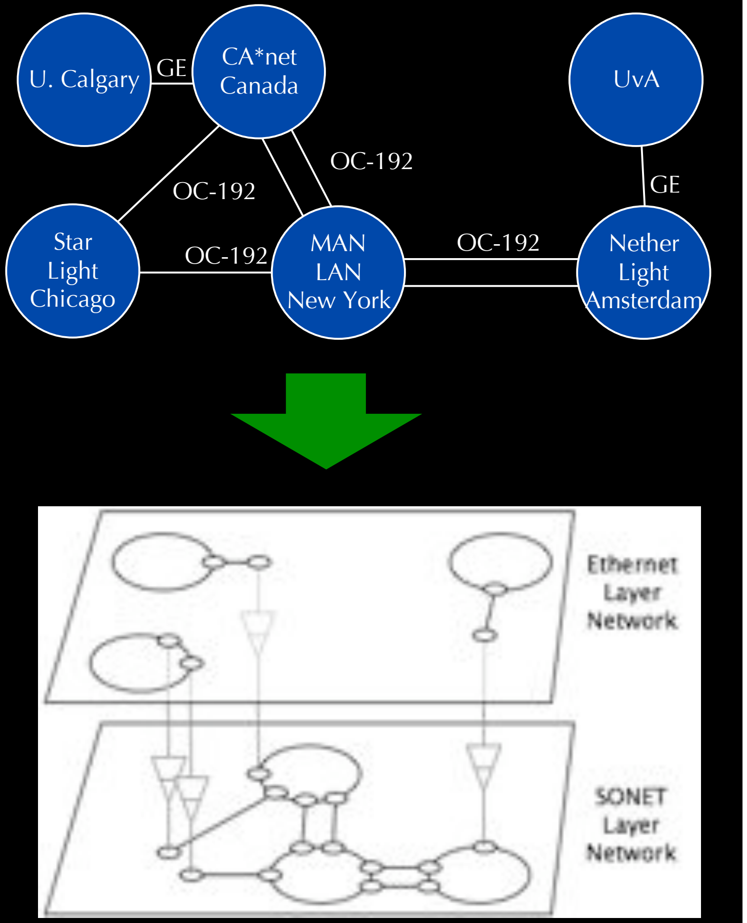




```
<ndl:Device rdf:about="#Force10">
  <ndl:hasInterface rdf:resource=
    "#Force10:te6/0"/>
</ndl:Device>
<ndl:Interface rdf:about="#Force10:te6/0">
  <rdfs:label>te6/0</rdfs:label>
  <ndl:capacity>1.25E6</ndl:capacity>
  <ndlconf:multiplex>
    <ndlcap:adaptation rdf:resource=
      "#Tagged-Ethernet-in-Ethernet"/>
    <ndlconf:serverPropertyValue
      rdf:resource="#MTU-1500byte"/>
  </ndlconf:multiplex>
  <ndlconf:hasChannel>
    <ndlconf:Channel rdf:about=
      "#Force10:te6/0:vlan4">
      <ndleth:hasVlan>4</ndleth:hasVlan>
      <ndlconf:switchedTo rdf:resource=
        "#Force10:gi5/1:vlan7"/>
    </ndlconf:Channel>
  </ndlconf:hasChannel>
</ndl:Interface>
```

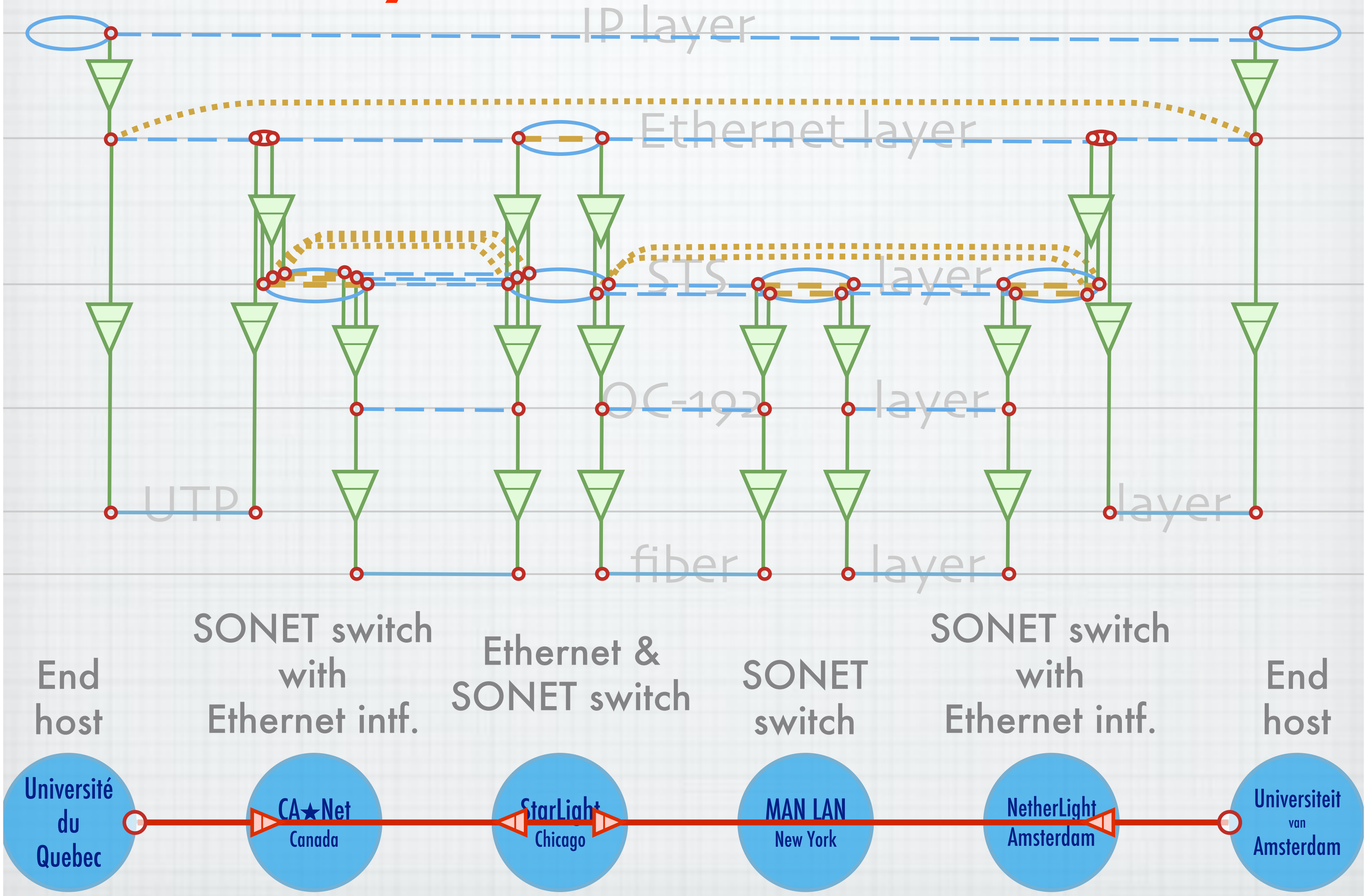
NDL Multilayer Extension

- ITU-T G.805 describes functional elements (e.g. adaptation, termination functions, link connections, etc.) to describe **network connections**.
- We extended these function elements (e.g. with potential adaptation functions) to describes **networks**.
- We created a model to map actual network elements to functional elements.
- Defined a simple algebra to define correctness of a connection
- We created a NDL extension to describe these functional elements.



Simplified model to map network elements to functional elements

Multi-layer extensions to NDL



OGF NML-WG

Open Grid Forum - Network Markup Language workgroup

Chairs:

- Paola Grosso – Universiteit van Amsterdam
- Martin Swany – University of Delaware

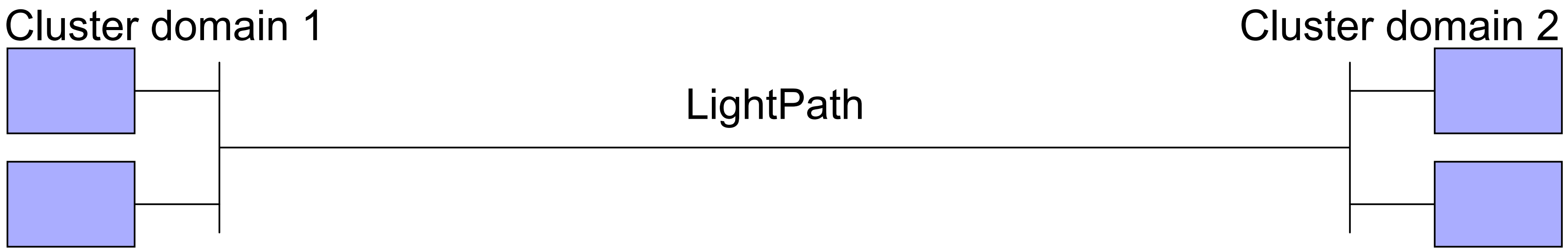
Purpose:

To describe network topologies, so that the outcome is a standardized network description ontology and schema, facilitating interoperability between different projects.

https://forge.gridforum.org/sf/projects/nml-wg

IP configuration in Optical Networks

- Problem: After a LightPath has been created, time is spent to manually configure IP addresses. Can this be done automatically?
- DHCP will not work out-of-the-box, since it is not clear which domain should run it.
- Possible solution: self-assigned IP addresses (RFC3927 for IPv4 or RFC1971 for IPv6)
- How to discover the target IP address or service?

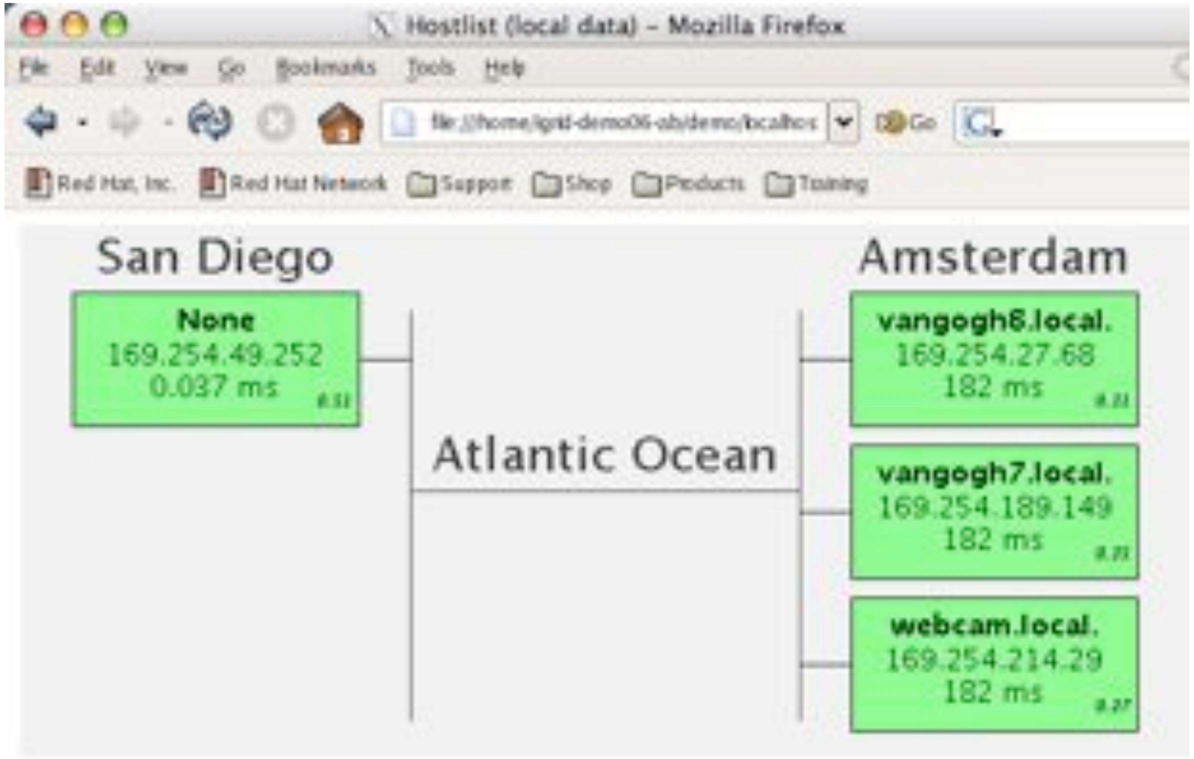


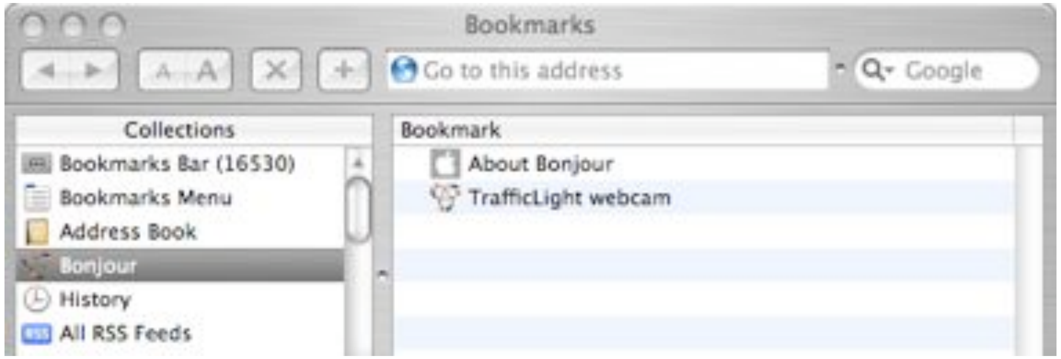
Technologies and Implementations

- Use Zero Configuration protocols
  - **Automatic configuration of IP addresses**
    - RFC3927 for IPv4 or RFC1971 for IPv6
  - **Name lookup of hosts**
    - Multicast DNS (mDNS) or Link-Local Multicast Name Resolution (LLMNR)
  - **Discovery of services**
    - DNS Service Discovery (DNS-SD), or Simple Service Discovery Protocol (SSDP, in UPnP), or Service Location Protocol (SLP) (or even UDDI, SDP, Salutation, or Jini)
- Three software suites, used multiple implementations:
  - RFC3927: ZCIP and autoip for Linux, native in OS X and Windows
  - mDNS: mDNSResponder, tmdns, and Porchdog mDNS
  - hooking gethostby*() to use mDNS: tmdns and libnss_mdns

Demonstration

- Used broadcast ping to discover hosts
- Used multicast DNS and gethostbyaddr() hook to discover hostnames
- Tested IP collisions
- Also demonstrated service discovery through DNS

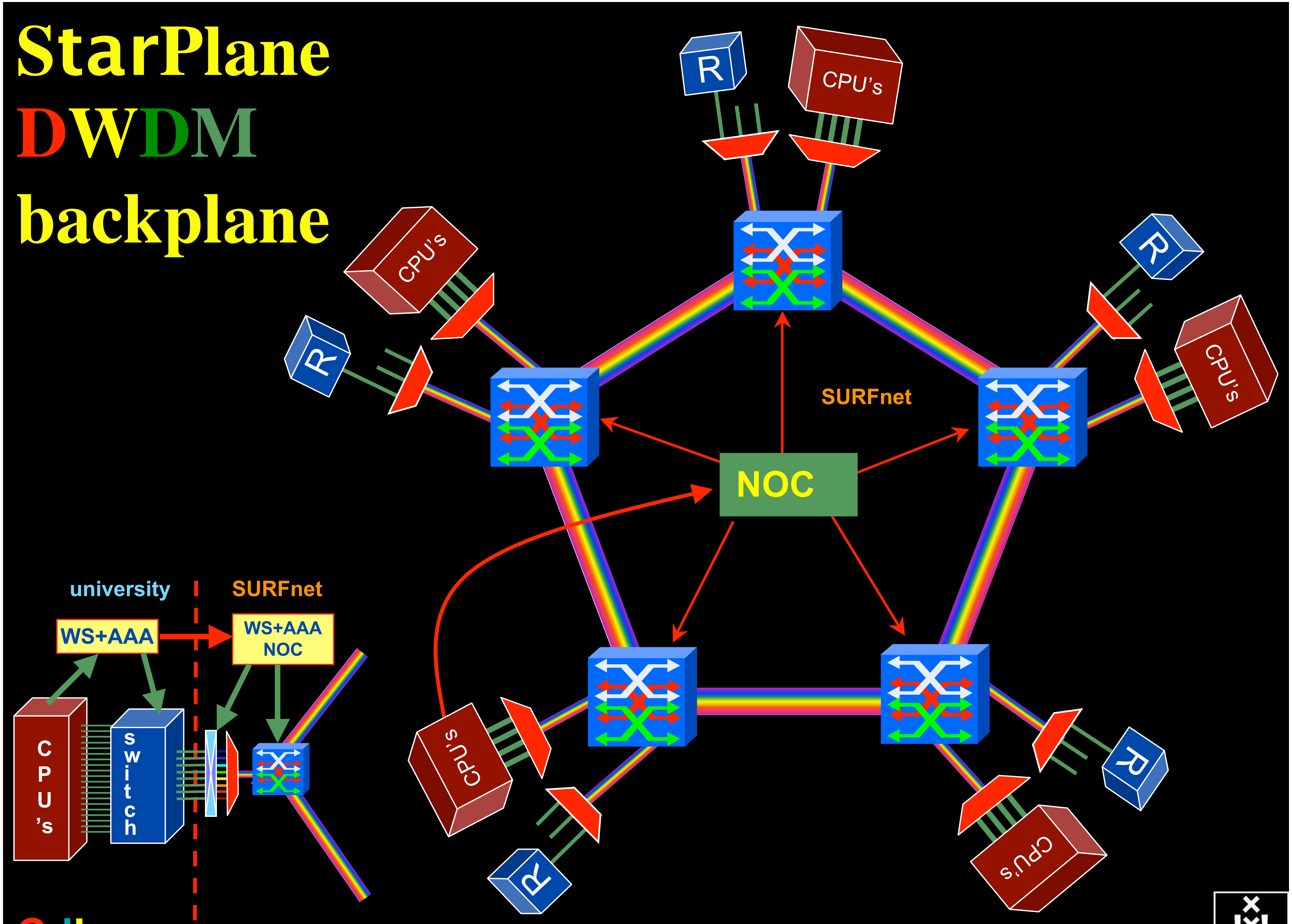




Bookmarks Bar (16530)
Bookmarks Menu
Address Book
Bonjour
History
All RSS Feeds

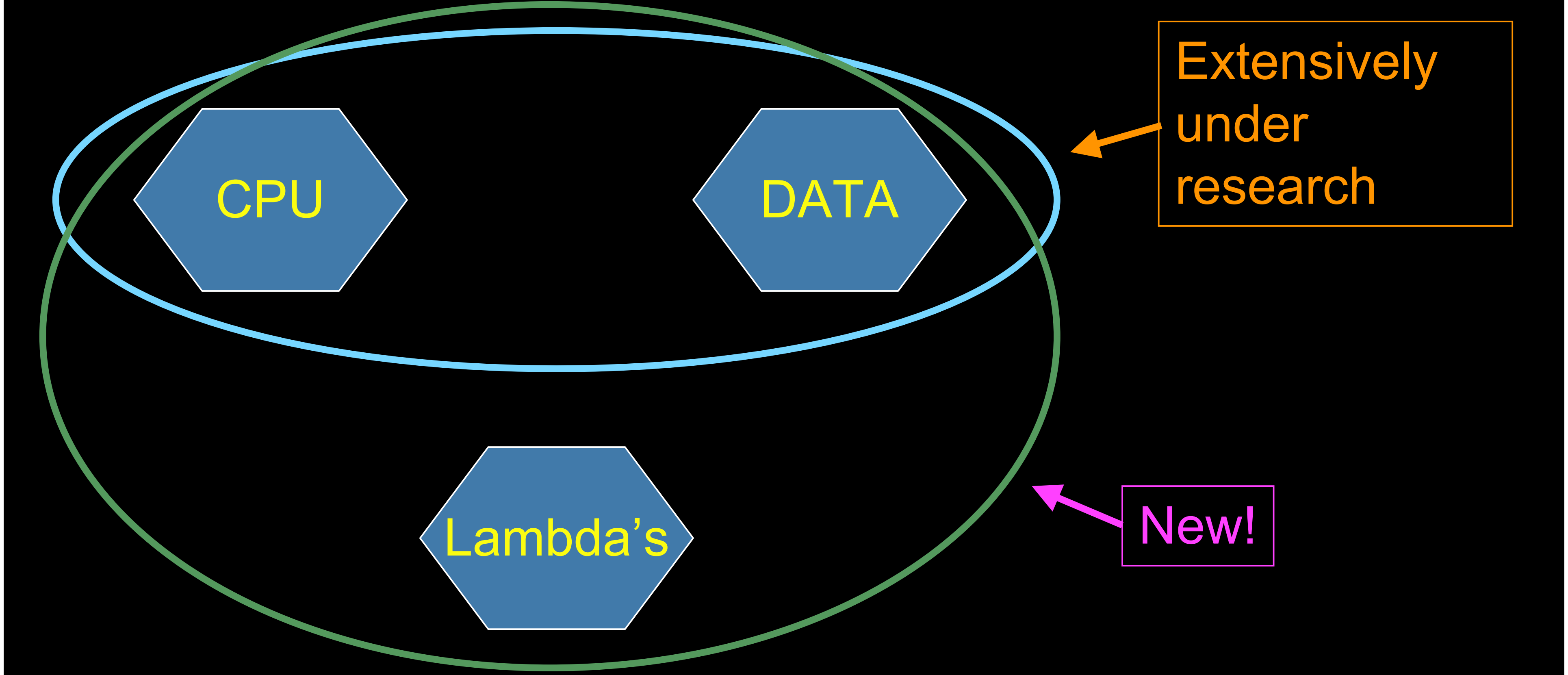
About Bonjour
TrafficLight webcam

StarPlane DWDM backplane



CdL

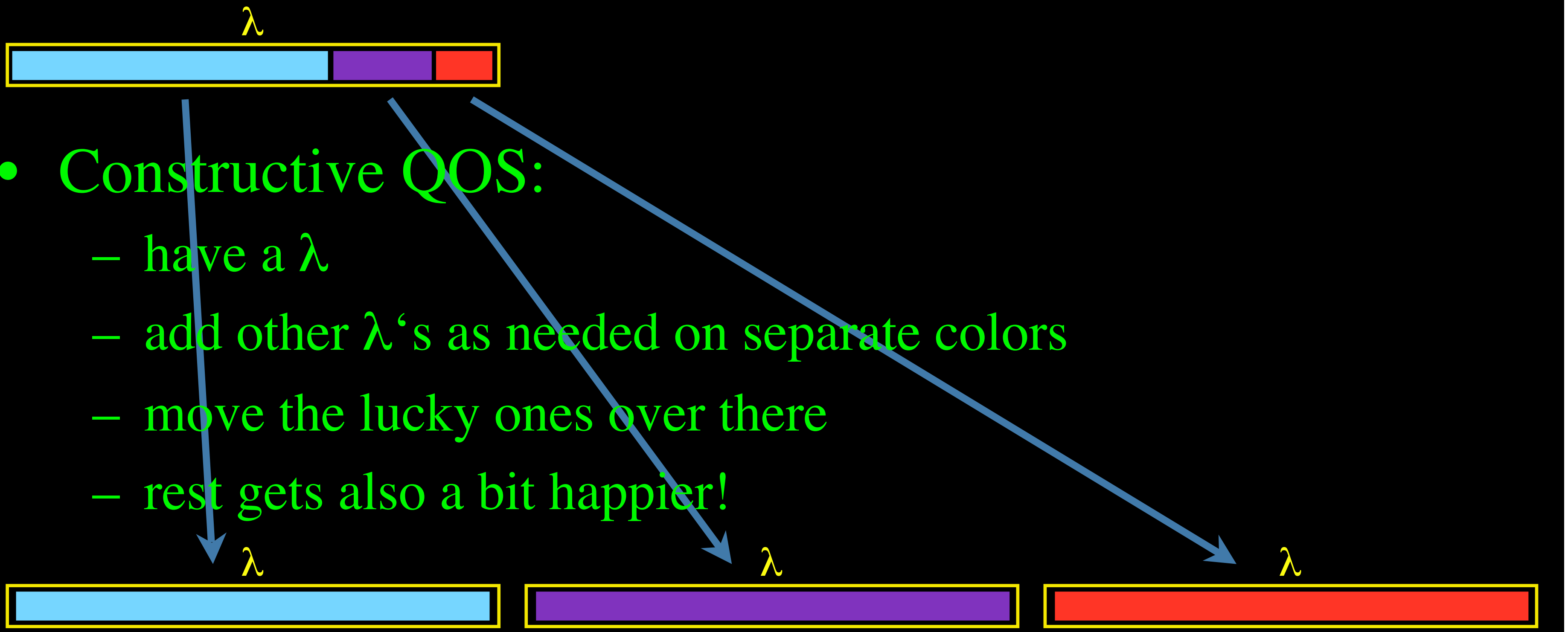
GRID Co-scheduling problem space



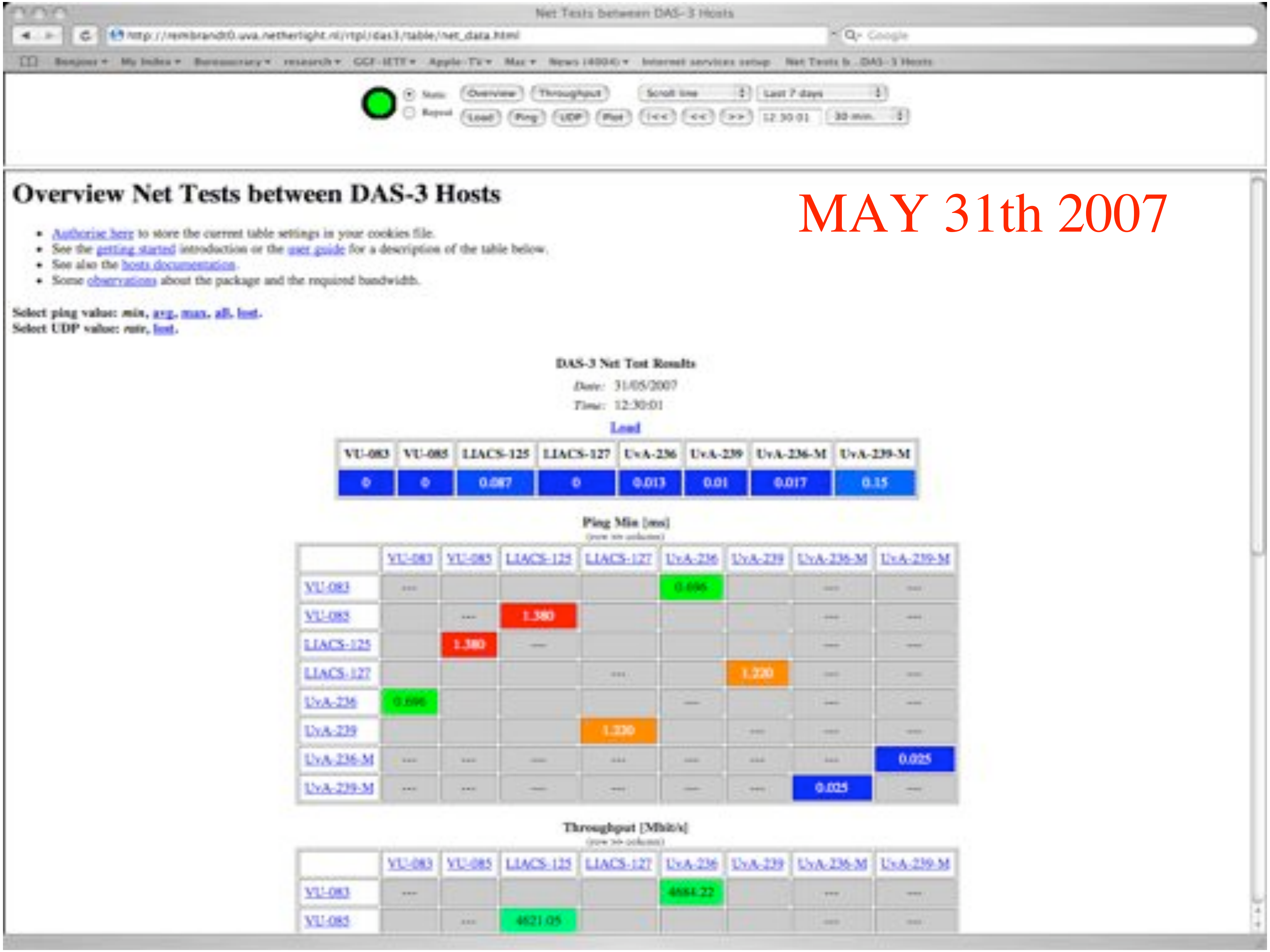
The StarPlane vision is to give flexibility directly to the applications by allowing them to choose the logical topology in real time, ultimately with sub-second lambda switching times on part of the SURFnet6 infrastructure.

QOS in a non destructive way!

- Destructive QOS:
  - have a link or λ
  - set part of it aside for a lucky few under higher priority
  - rest gets less service



- Constructive QOS:
  - have a λ
  - add other λ‘s as needed on separate colors
  - move the lucky ones over there
  - rest gets also a bit happier!



● Static ○ Repeat

Scroll line · Last 7 days · 12:30:01 · 30 min.

MAY 31th 2007

Overview Net Tests between DAS-3 Hosts

- Authorise here to store the current table settings in your cookies file.
- See the getting started introduction or the user guide for a description of the table below.
- See also the hosts documentation.
- Some observations about the package and the required bandwidth.

Select ping value: ***min***, **avg**, **max**, **all**, **lost**.
Select UDP value: ***rate***, **lost**.

DAS-3 Net Test Results

Date: 31/05/2007

Time: 12:30:01

Load

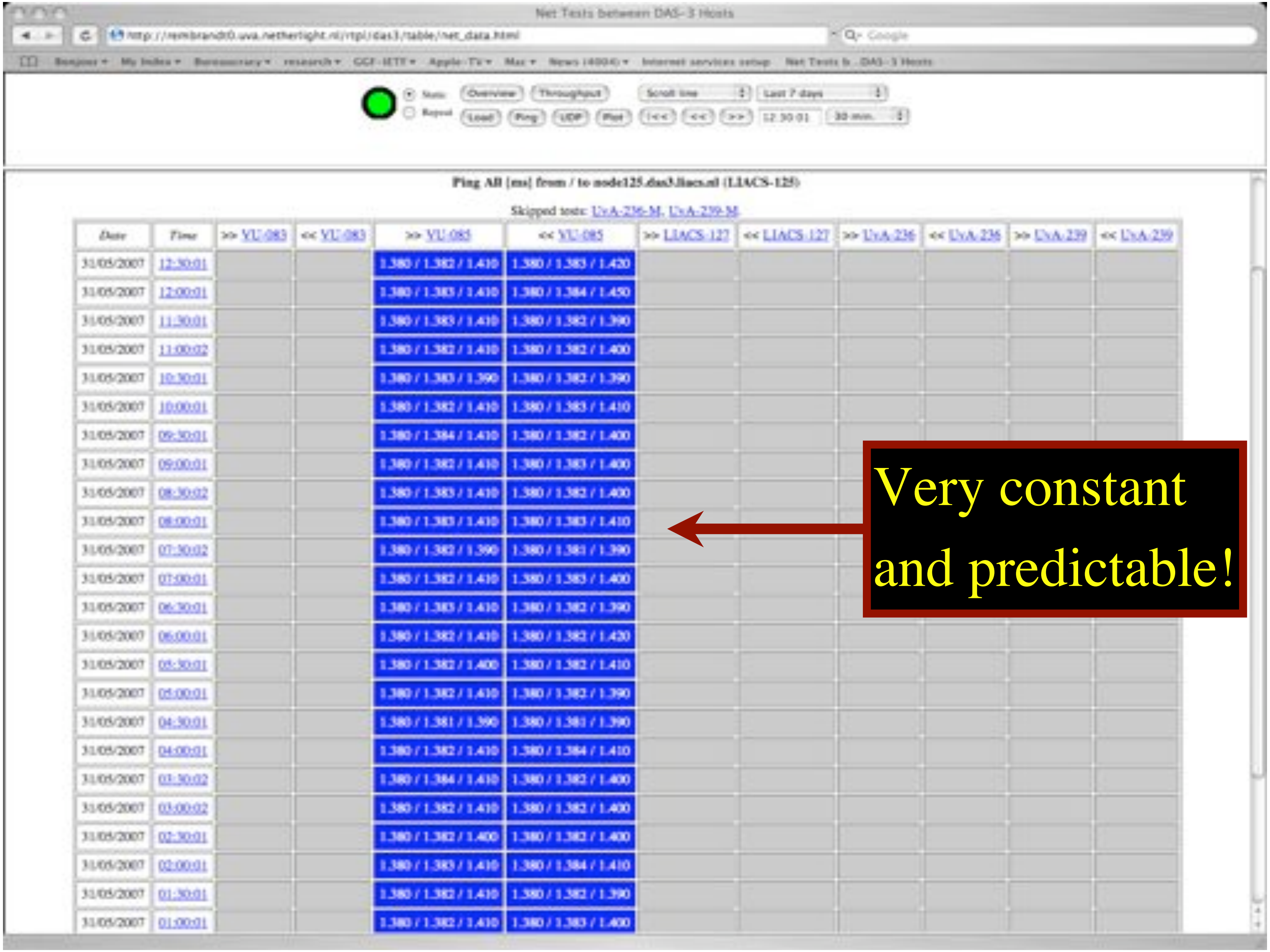
VU-083	VU-085	LIACS-125	LIACS-127	UvA-236	UvA-239	UvA-236-M	UvA-239-M
0	0	0.087	0	0.013	0.01	0.017	0.15

Ping Min [ms]
(row >> column)

	VU-083	VU-085	LIACS-125	LIACS-127	UvA-236	UvA-239	UvA-236-M	UvA-239-M
VU-083	---				0.696		---	---
VU-085		---	1.380				---	---
LIACS-125		1.380	---				---	---
LIACS-127				---		1.220	---	---
UvA-236	0.696				---		---	---
UvA-239				1.220		---	---	---
UvA-236-M	---	---	---	---	---	---	---	0.025
UvA-239-M	---	---	---	---	---	---	0.025	---

Throughput [Mbit/s]
(row >> column)

	VU-083	VU-085	LIACS-125	LIACS-127	UvA-236	UvA-239	UvA-236-M	UvA-239-M
VU-083	---				4684.22		---	---
VU-085		---	4621.05				---	---


Net Tests between DAS-3 Hosts
http://rembrandt0.uva.netherlight.nl/rtpi/das3/table/net_data.html
Google
Overview
Throughput
Scroll time
Last 7 days
Load
Ping
UDP
Plot
|<<
<<
>>
12:30:01
30 min.

Ping All [ms] from / to node125.das3.liacs.nl (LIACS-125)

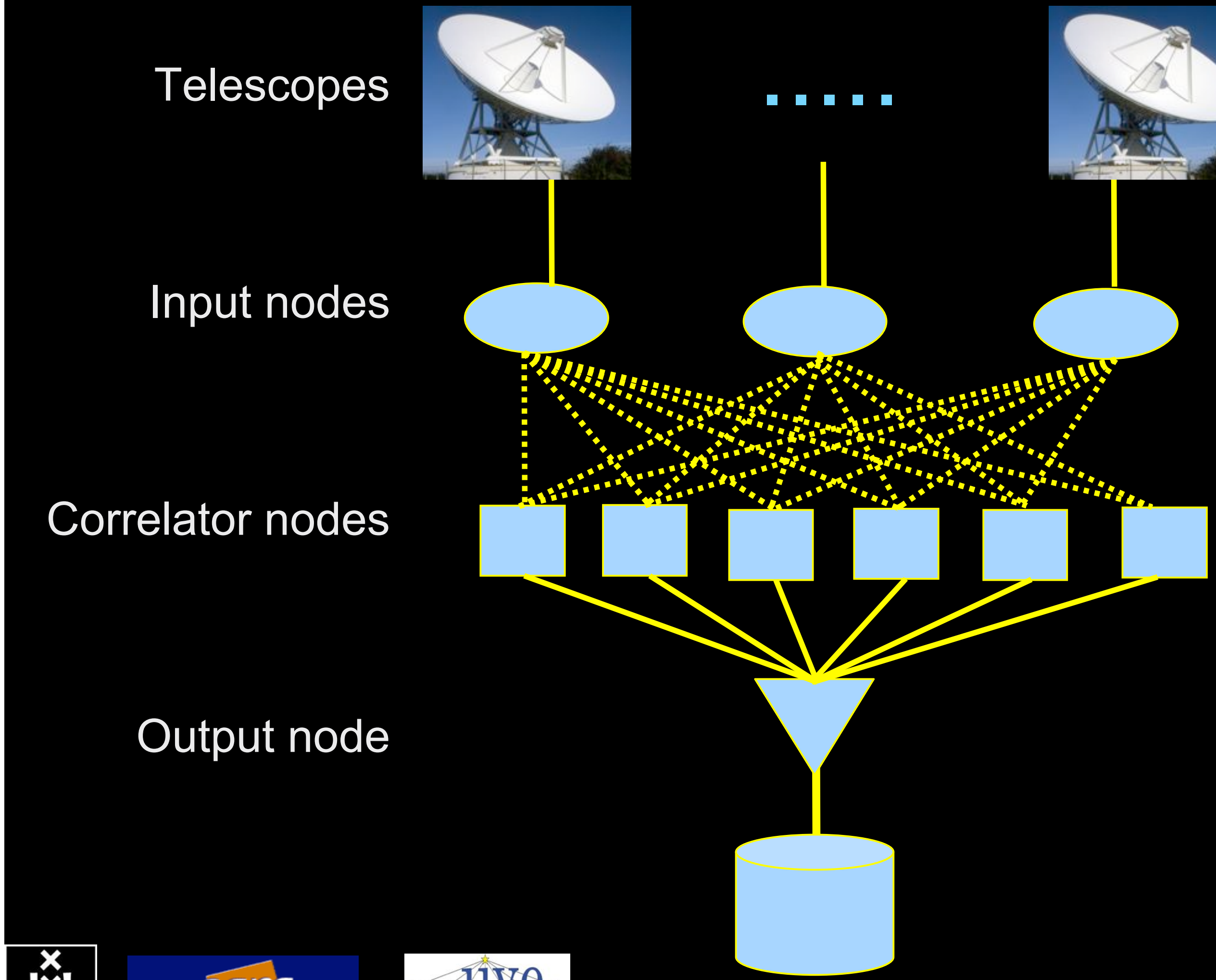
Skipped tests: UvA-236-M, UvA-239-M

Date	*Time*	>> VU-083	<< VU-083	>> VU-085	<< VU-085	>> LIACS-127	<< LIACS-127	>> UvA-236	<< UvA-236	>> UvA-239	<< UvA-239
31/05/2007	12:30:01			1.380 / 1.382 / 1.410	1.380 / 1.383 / 1.420						
31/05/2007	12:00:01			1.380 / 1.383 / 1.410	1.380 / 1.384 / 1.450						
31/05/2007	11:30:01			1.380 / 1.383 / 1.410	1.380 / 1.382 / 1.390						
31/05/2007	11:00:02			1.380 / 1.382 / 1.410	1.380 / 1.382 / 1.400						
31/05/2007	10:30:01			1.380 / 1.383 / 1.390	1.380 / 1.382 / 1.390						
31/05/2007	10:00:01			1.380 / 1.382 / 1.410	1.380 / 1.383 / 1.410						
31/05/2007	09:30:01			1.380 / 1.384 / 1.410	1.380 / 1.382 / 1.400						
31/05/2007	09:00:01			1.380 / 1.382 / 1.410	1.380 / 1.383 / 1.400						
31/05/2007	08:30:02			1.380 / 1.383 / 1.410	1.380 / 1.382 / 1.400						
31/05/2007	08:00:01			1.380 / 1.383 / 1.410	1.380 / 1.383 / 1.410						
31/05/2007	07:30:02			1.380 / 1.382 / 1.390	1.380 / 1.381 / 1.390						
31/05/2007	07:00:01			1.380 / 1.382 / 1.410	1.380 / 1.383 / 1.400						
31/05/2007	06:30:01			1.380 / 1.383 / 1.410	1.380 / 1.382 / 1.390						
31/05/2007	06:00:01			1.380 / 1.382 / 1.410	1.380 / 1.382 / 1.420						
31/05/2007	05:30:01			1.380 / 1.382 / 1.400	1.380 / 1.382 / 1.410						
31/05/2007	05:00:01			1.380 / 1.382 / 1.410	1.380 / 1.382 / 1.390						
31/05/2007	04:30:01			1.380 / 1.381 / 1.390	1.380 / 1.381 / 1.390						
31/05/2007	04:00:01			1.380 / 1.382 / 1.410	1.380 / 1.384 / 1.410						
31/05/2007	03:30:02			1.380 / 1.384 / 1.410	1.380 / 1.382 / 1.400						
31/05/2007	03:00:02			1.380 / 1.382 / 1.410	1.380 / 1.382 / 1.400						
31/05/2007	02:30:01			1.380 / 1.382 / 1.400	1.380 / 1.382 / 1.400						
31/05/2007	02:00:01			1.380 / 1.383 / 1.410	1.380 / 1.384 / 1.410						
31/05/2007	01:30:01			1.380 / 1.382 / 1.410	1.380 / 1.382 / 1.390						
31/05/2007	01:00:01			1.380 / 1.382 / 1.410	1.380 / 1.383 / 1.400						

Very constant and predictable!

The SCARIe project

SCARIe: a research project to create a Software Correlator for e-VLBI.
VLBI Correlation: signal processing technique to get high precision image from spatially distributed radio-telescope.



To equal the hardware correlator we need:

16 streams of 1Gbps

16 * 1Gbps of data

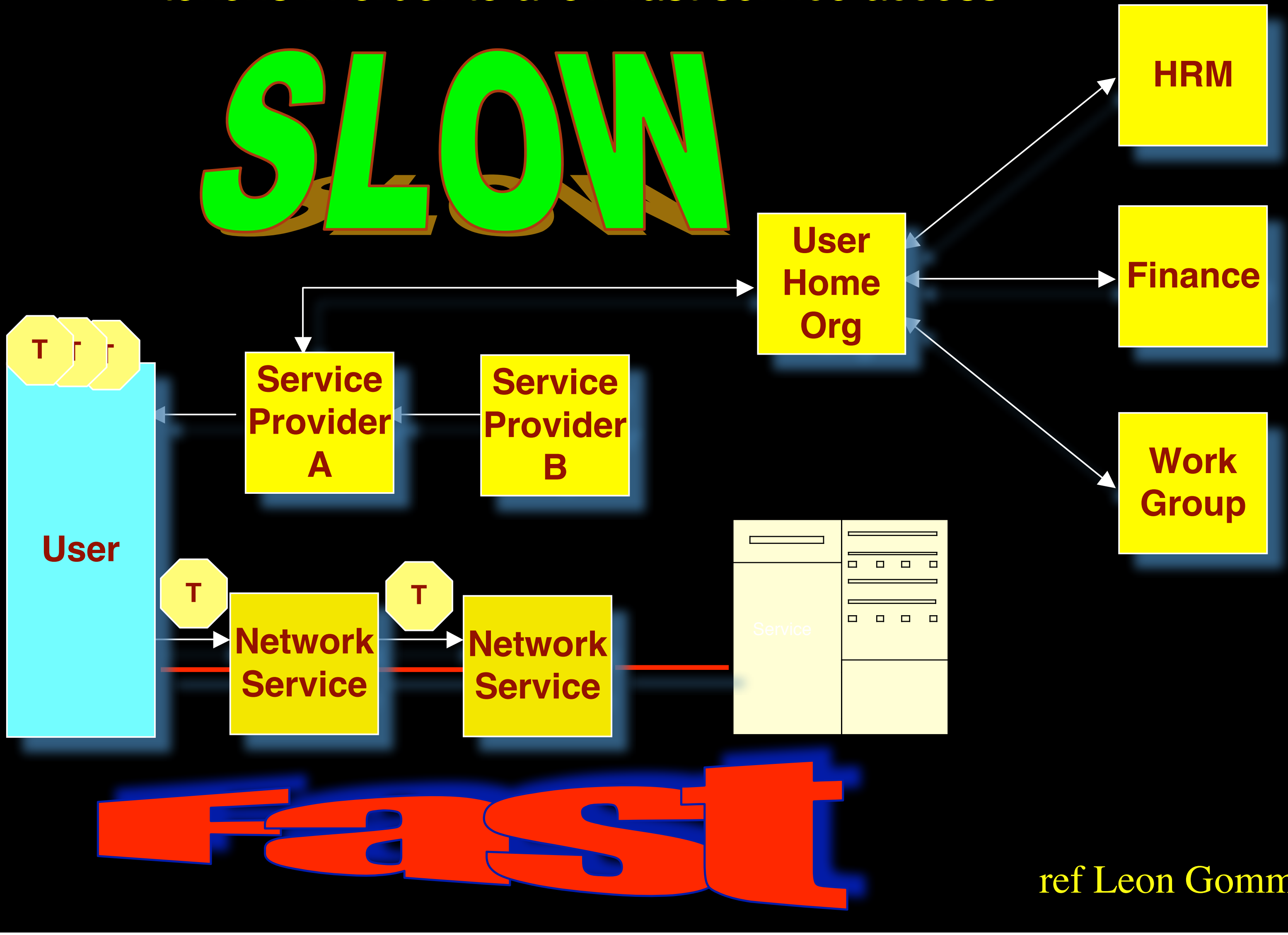
2 Tflops CPU power

2 TFlop / 16 Gbps =

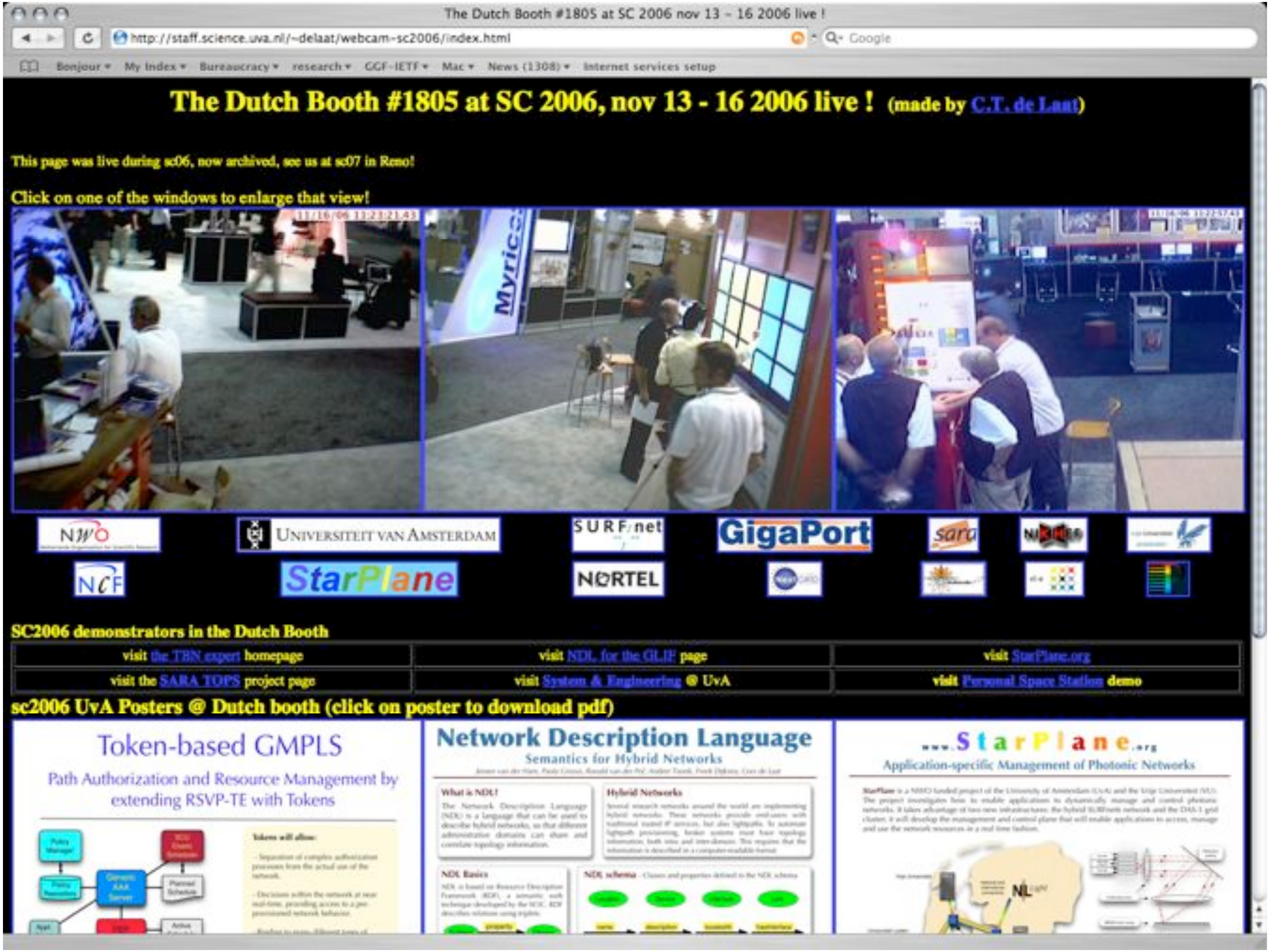
1000 flops/byte

THIS IS A DATA FLOW PROBLEM !!!

Use AAA concept to split (time consuming) service authorization process from service access using secure tokens in order to allow fast service access.



ref Leon Gommans



The Dutch Booth #1805 at SC 2006 nov 13 - 16 2006 live !
http://staff.science.uva.nl/~delaat/webcam-sc2006/index.html
Google
Bonjour ▾
My Index ▾
Bureaucracy ▾
research ▾
GGF-IETF ▾
Mac ▾
News (1308) ▾
Internet services setup

The Dutch Booth #1805 at SC 2006, nov 13 - 16 2006 live ! (made by C.T. de Laat)

This page was live during sc06, now archived, see us at sc07 in Reno!

Click on one of the windows to enlarge that view!

SC2006 demonstrators in the Dutch Booth

visit the TBN expert homepage	visit NDL for the GLIF page	visit StarPlane.org
visit the SARA TOPS project page	visit System & Engineering @ UvA	visit Personal Space Station demo

sc2006 UvA Posters @ Dutch booth (click on poster to download pdf)

Token-based GMPLS

Path Authorization and Resource Management by extending RSVP-TE with Tokens

Tokens will allow:

- Separation of complex authorization processes from the actual use of the network.

- Decisions within the network at near real-time, providing access to a pre-provisioned network behavior.

- Binding to many different types of

Network Description Language

Semantics for Hybrid Networks

Jeroen van der Ham, Paola Grosso, Ronald van der Pol, Andree Toonk, Freek Dijkstra, Cees de Laat

What is NDL?

The Network Description Language (NDL) is a language that can be used to describe hybrid networks, so that different administrative domains can share and correlate topology information.

Hybrid Networks

Several research networks around the world are implementing hybrid networks. These networks provide end-users with traditional routed IP services, but also lightpaths. To automate lightpath provisioning, broker systems must have topology information, both intra and inter-domain. This requires that the information is described in a computer-readable format.

NDL Basics

NDL is based on Resource Description Framework (RDF), a semantic web technique developed by the W3C. RDF describes relations using triples.

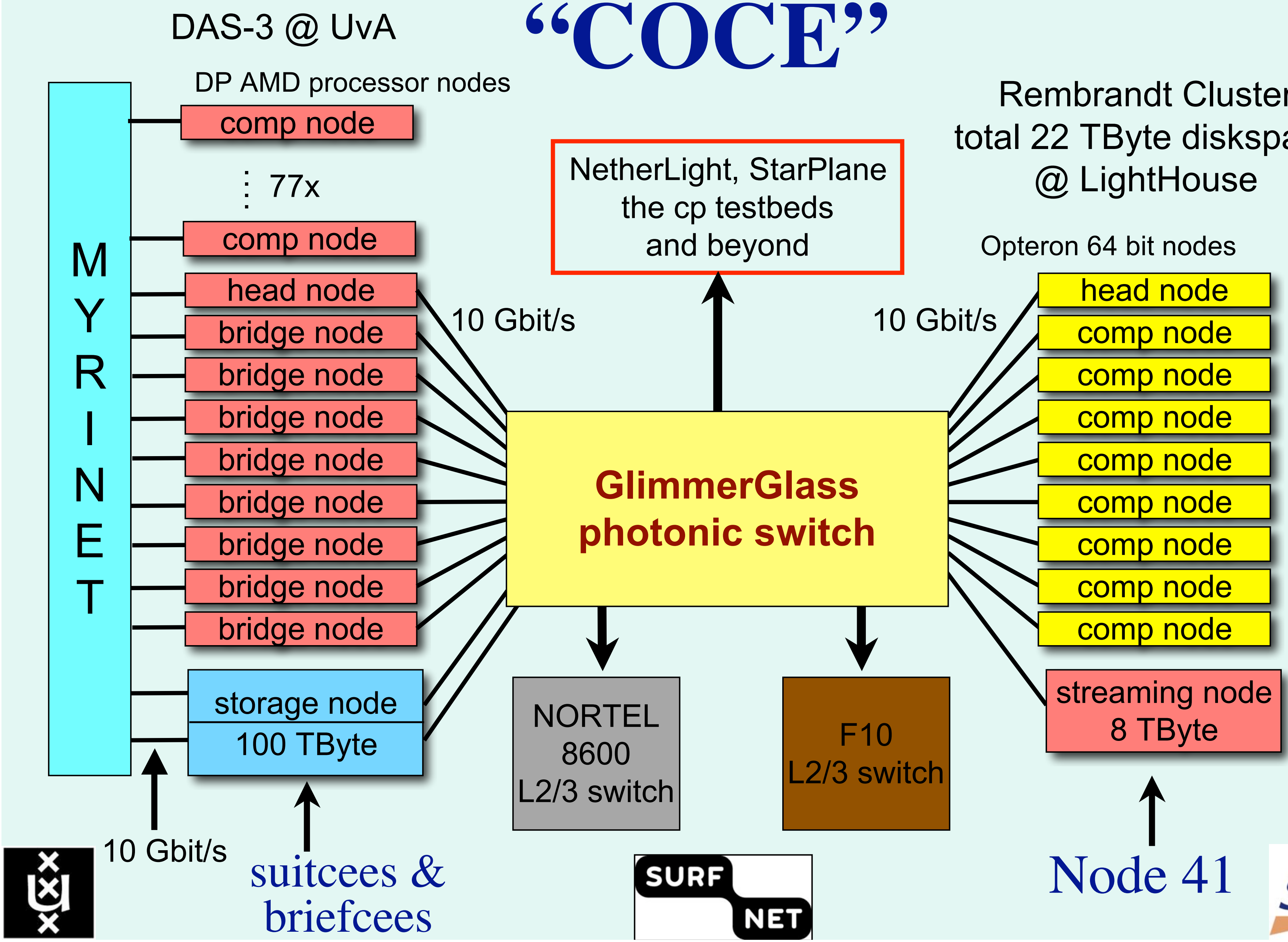
NDL schema - Classes and properties defined in the NDL schema

www.StarPlane.org

Application-specific Management of Photonic Networks

StarPlane is a NWO funded project of the University of Amsterdam (UvA) and the Vrije Universiteit (VU). The project investigates how to enable applications to dynamically manage and control photonic networks. It takes advantage of two new infrastructures: the hybrid SURFnet6 network and the DAS-3 grid cluster. it will develop the management and control plane that will enable applications to access, manage and use the network resources in a real time fashion.

Amsterdam CineGrid S/F node "COCE"



DAS-3 @ UvA

DP AMD processor nodes

MYRINET

comp node

77x

comp node

head node

bridge node

bridge node

bridge node

bridge node

bridge node

bridge node

bridge node

bridge node

storage node

100 TByte

10 Gbit/s

suitcees & briefcees

NetherLight, StarPlane
the cp testbeds
and beyond

10 Gbit/s

GlimmerGlass
photonic switch

NORTEL
8600
L2/3 switch

F10
L2/3 switch

Rembrandt Cluster
total 22 TByte diskspace
@ LightHouse

Opteron 64 bit nodes

10 Gbit/s

head node

comp node

comp node

comp node

comp node

comp node

comp node

comp node

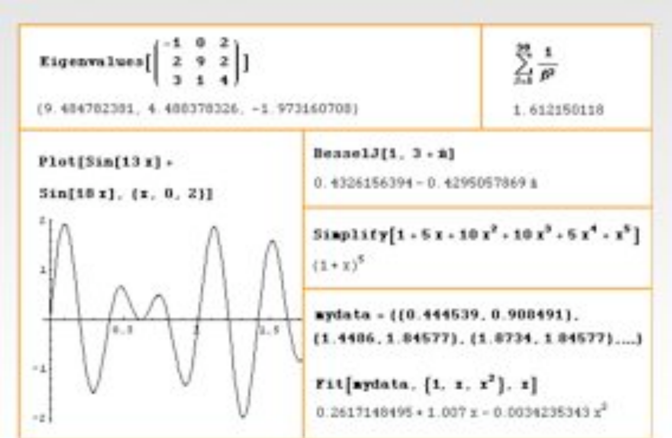
comp node

streaming node
8 TByte

Node 41

User Programmable Virtualized Networks allows the results of decades of computer science to handle the complexities of application specific networking.

- The network is virtualized as a collection of resources
- UPVNs enable network resources to be programmed as part of the application
- Mathematica, a powerful mathematical software system, can interact with real networks using UPVNs



```
Eigenvalues[( -1 0 2
              2 9 2
              3 1 4 )]
{9.484782381, 4.488378326, -1.973160708}
```

$\sum_{j=1}^{30} \frac{1}{j^2}$

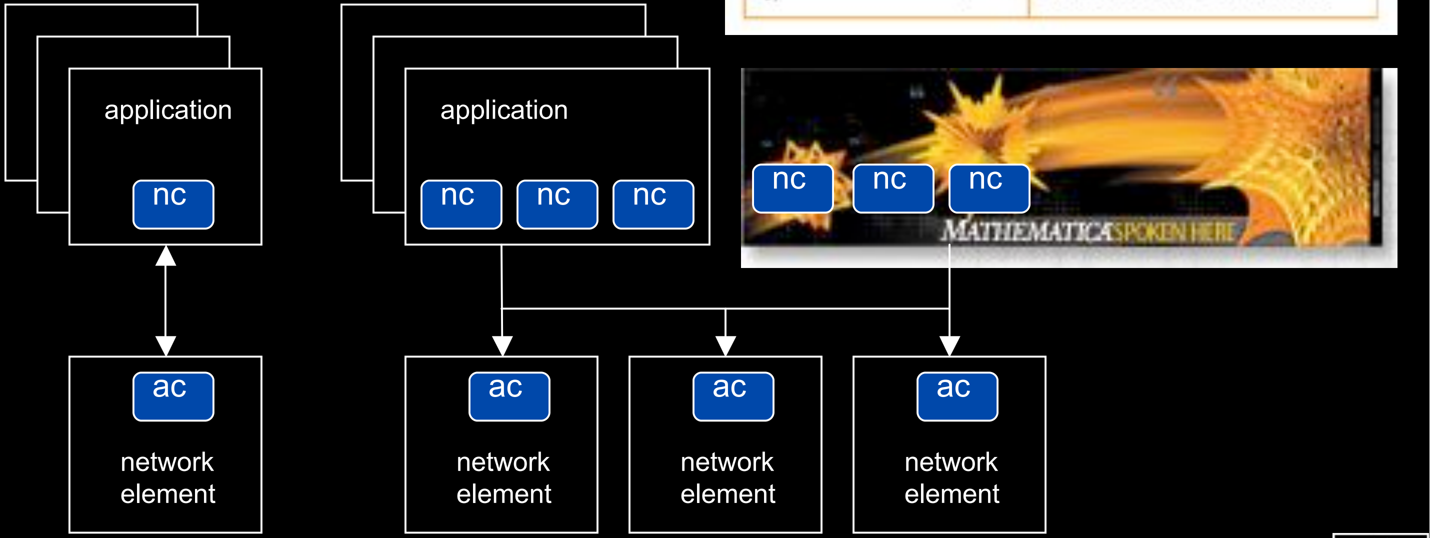
```
1.612150118
```

```
Plot[Sin[13 x] + Sin[18 x], {x, 0, 2}]
```

```
BesselJ[1, 3 + i]
0.4326156394 - 0.4295057869 i
```

```
Simplify[1 + 5 x + 10 x^2 + 10 x^3 + 5 x^4 + x^5]
(1 + x)^5
```

```
mydata = {{0.444539, 0.908491},
{1.4486, 1.84577}, {1.8734, 1.84577},...}
Fit[mydata, {1, x, x^2}, x]
0.2617148495 + 1.007 x - 0.0034235343 x^2
```

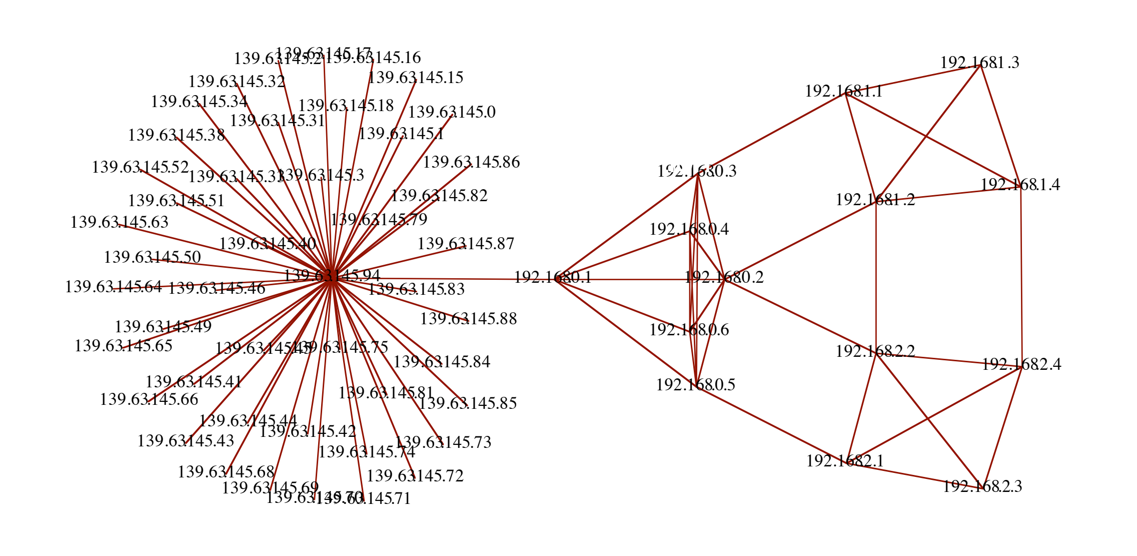


Mathematica enables advanced graph queries, visualizations and real-time network manipulations on UPVNs

Topology matters can be dealt with algorithmically
Results can be persisted using a transaction service built in UPVN

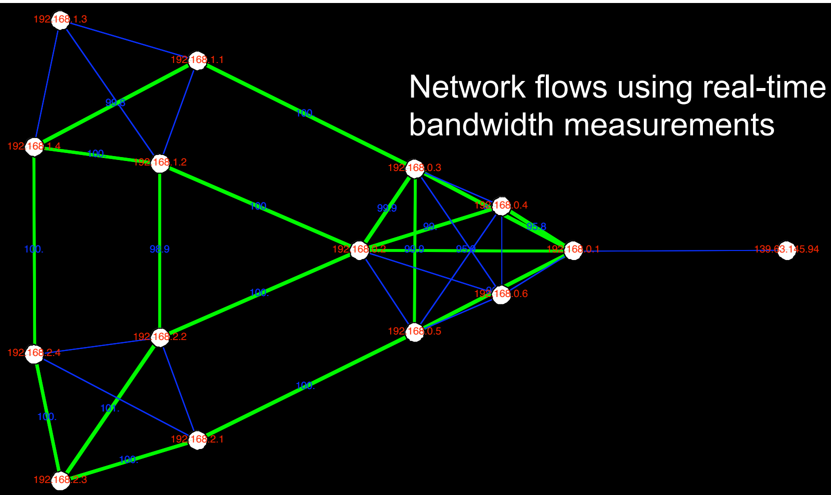
Initialization and BFS discovery of NEs

```
Needs["WebServices`"]
<<DiscreteMath`Combinatorica`
<<DiscreteMath`GraphPlot`
InitNetworkTopologyService["edge.ict.tno.nl"]

Available methods:
{DiscoverNetworkElements,GetLinkBandwidth,GetAllIpLinks,Remote,
NetworkTokenTransaction}

Global`upvnverbose = True;
AbsoluteTiming[nes = BFSDiscover["139.63.145.94"];][[1]]
AbsoluteTiming[result = BFSDiscoverLinks["139.63.145.94", nes];][[1]]

Getting neigbours of: 139.63.145.94
Internal links: {192.168.0.1, 139.63.145.94}
(...)
Getting neigbours of:192.168.2.3
Internal links: {192.168.2.3}
```



Transaction on shortest path with tokens

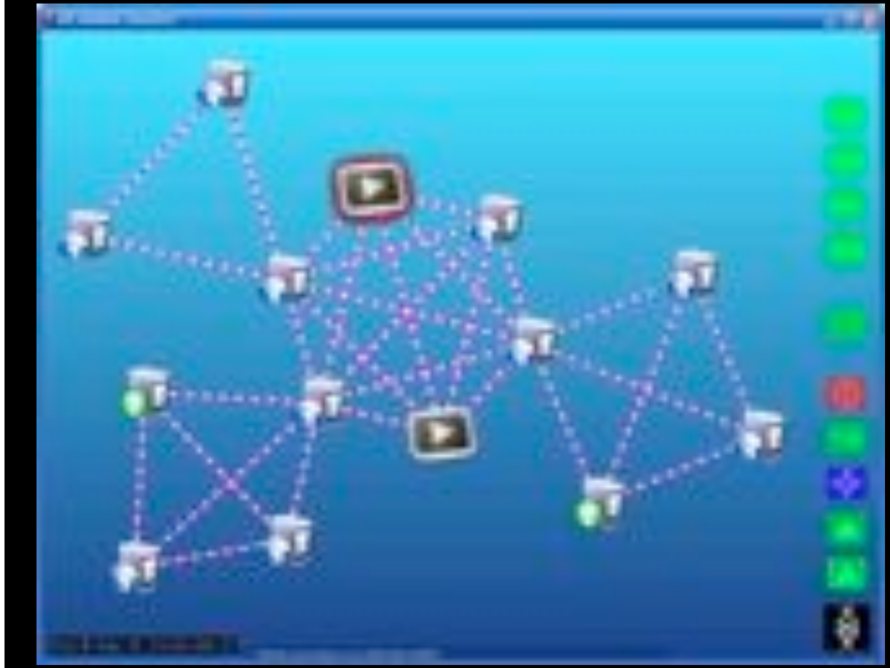
```
nodePath = ConvertIndicesToNodes[
                    ShortestPath[  g,
                                        Node2Index[nids,"192.168.3.4"],
                                        Node2Index[nids,"139.63.77.49"]],
                                        nids];
Print["Path: ", nodePath];
If[NetworkTokenTransaction[nodePath, "green"]==True,
          Print["Committed"], Print["Transaction failed"]];

Path:
{192.168.3.4,192.168.3.1,139.63.77.30,139.63.77.49}

Committed
```

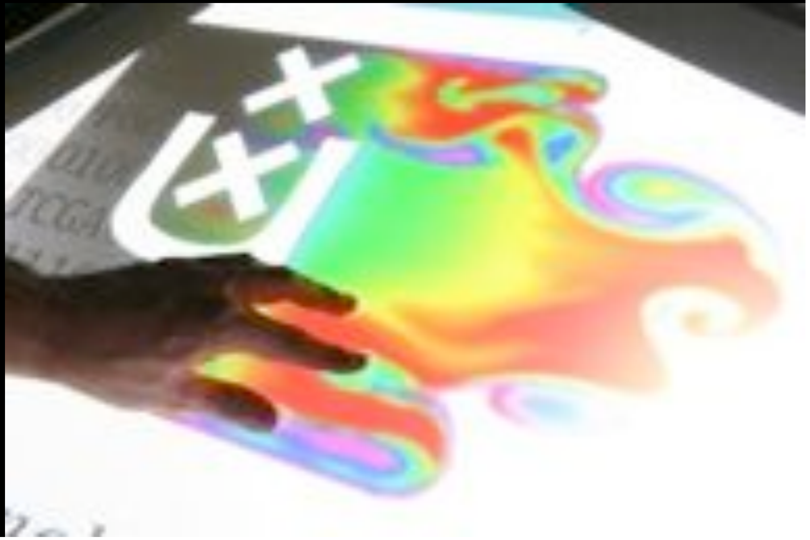


Network flows using real-time bandwidth measurements

ref: Robert J. Meijer, Rudolf J. Strijkers, Leon Gommans, Cees de Laat, User Programmable Virtualiized Networks, accepted for publication to the IEEE e-Science 2006 conference Amsterdam.

TouchTable Demonstration @ SC08





Scientific Publications

- Some publications this year:
  - Larry Smarr, Maxine Brown, Cees de Laat, Editorial: "Special Section: OptIPlanet - The OptIPuter Global Collaboratory", FGCS, Vol 25, issue 2, feb 2009, pages 109-113
  - Leon Gommans, Li Xu, Fred Wan, Yuri Demchenko, Mihai Cristea, Robert Meijer, Cees de Laat , Multi-Domain Lightpath Authorization using Tokens, FGCS, Vol 25, issue 2, feb 2009, pages 153-160
  - Freek Dijkstra, Jeroen J van der Ham, Paola Grosso, Cees de Laat, "A Path Finding Implementation for Multi-Layer Networks", FGCS, Vol 25, issue 2, feb 2009, pages 142-146
  - Paola Grosso, Damien Marchal, Jason Maassen, Eric Bernier, Li Xu, Cees de Laat, "Dynamic Photonic Lightpaths in the StarPlane Network", FGCS, Vol 25, issue 2, feb 2009, pages 132-136
  - Yuri Demchenko, Fred Wan, Mihai Cristea, Cees de Laat, "Authorisation Infrastructure for On-Demand Network Resource Provisioning", Grid2008 Conference - September 29 - October 1, 2008, Tsukuba, Japan, accepted for publication.
  - Jeroen van der Ham, Freek Dijkstra, Paola Grosso, Ronald van der Pol, Andree Toonk, Cees de Laat, "A distributed topology information system for optical networks based on the semantic web", Elsevier Journal on Optical Switching and Networking,Volume 5, Issues 2-3, June 2008, Pages 85-93
  - Freek Dijkstra, Bert Andree, Karst Koymans, Jeroen van der Ham, Paola Grosso, Cees de Laat, "Multi-Layer Network Model Based on ITU-T G.805", Elsevier Computer Networks, Mar 2008
  - Y. Demchenko and L. Gommans and C.T.A.M. de Laat, "Extending role based access control model for distributed multidomain applications", IFIP International Federation for Information Processing, Volume 232, pages 301-312.
- About 8 - 10 publications/year in journals and conf records.
  - see http://www.science.uva.nl/~delaat/pubs.html
- About 15 talks/year, many invited.
  - see http://www.science.uva.nl/~delaat/talks.html

International presence

- GLIF
- OGF
  - GFSG
  - GHPN-WG
  - NSI-WG
  - NML-WG
- IETF
- ONT workshop organization
- IRNC workshop
- FIRE Expert team

The HighLights

- StarPlane first DRAC WSS flip nov 2008
- NDL Multilayer pathfinding is being adopted
- Multi domain simulation NDL
- NDL & PROLOG
- Token based networking for inter domain GMPLS
- TBN solves problems for PhosPhorus-I2 interworking
- DRAC - IDC - Harmony LightPath setup
- SCARIe AuthoBAHN StarPlane demo
- HPDMnet High Quality video switching
- CineGrid Streaming, Storage and Forwarding
- Dark fiber SARA and SNE master extended to Oslo
- Programmable network demonstration with touch-table
- CineGrid portal streaming with PBT for QoS

Questions ?

SURF NET

StarPlane