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)
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
\[ \Sigma A = \Sigma B = \Sigma C \approx 250 \text{ Gb/s} \]

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

Optical Exchange

- Switch
- TDM
- Store & Forward
- DWDM mux/demux

Ref: gridnets paper by Freek Dijkstra, Cees de Laat
The Modelling Process

Network Elements -> Functional Elements -> Syntax

Example code:
```xml
<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">
      <ndl:hasVlan>4</ndl:hasVlan>
      <ndlconf:switchedTo rdf:resource="#Force10:gi5/1:vlan7"/>
    </ndlconf:Channel>
  </ndlconf:hasChannel>
</ndl:Interface>
```
• 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
StarPlane
DWDM
backplane
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 $\lambda$
  – set part of it aside for a lucky few under higher priority
  – rest gets less service

• Constructive QOS:
  – have a $\lambda$
  – add other $\lambda$’s as needed on separate colors
  – move the lucky ones over there
  – rest gets also a bit happier!
Very constant and predictable!
**The SCARLe project**

**SCARLe:** 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.
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
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 neighbours of: 139.63.145.94
Internal links: {192.168.0.1, 139.63.145.94}
(...)
Getting neighbours 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];

Path: {192.168.3.4,192.168.3.1,139.63.77.30,139.63.77.49}
Committed
```

ref: Robert J. Meijer, Rudolf J. Strijkers, Leon Gommans, Cees de Laat, User Programmable Virtualized Networks, accepted for publication to the IEEE e-Science 2006 conference Amsterdam.
TouchTable Demonstration @ SC08
Scientific Publications

• Some publications this year:

• 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 ?