Beyond Hybrid Networking

Cees de Laat

University of Amsterdam
Trends

• We have made baby-steps on the path to optical networking
  – Still many mails and phone calls

• See several trends:
  – lambda’s get fatter and cheaper
  – photonic technology cheap per bandwidth
  – 100 Gb trials ongoing
  – embedded computation capacity increasing
  – latency and high bandwidth congestion avoidance conflict
  – ethernet is enhanced with circuit properties (PBT)
  – applications need more and more predictable behaviour
Grid in operation
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
- \( \frac{2 \text{ TFlop}}{16 \text{ Gbps}} = 1000 \text{ flops/byte} \)

**THIS IS A DATA FLOW PROBLEM !!!**
CineGrid
# CineGrid Format - Numbers - Bits

<table>
<thead>
<tr>
<th>Format</th>
<th>X</th>
<th>Y</th>
<th>Rate</th>
<th>Color bits/pix</th>
<th>Frame pix</th>
<th>Frame MByte</th>
<th>Flow MByte/s</th>
<th>Stream Gbit/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>720p HD</td>
<td>1280</td>
<td>720</td>
<td>60</td>
<td>24</td>
<td>921600</td>
<td>2.8</td>
<td>170</td>
<td>1.3</td>
</tr>
<tr>
<td>1080p HD</td>
<td>1920</td>
<td>1080</td>
<td>30</td>
<td>24</td>
<td>2073600</td>
<td>6.2</td>
<td>190</td>
<td>1.5</td>
</tr>
<tr>
<td>2k</td>
<td>2048</td>
<td>1080</td>
<td>24</td>
<td>36</td>
<td>2211840</td>
<td>10</td>
<td>240</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>48</td>
<td></td>
<td></td>
<td></td>
<td>480</td>
<td>2.4</td>
</tr>
<tr>
<td>SHD</td>
<td>3840</td>
<td>2160</td>
<td>30</td>
<td>24</td>
<td>8294400</td>
<td>25</td>
<td>750</td>
<td>6.0</td>
</tr>
<tr>
<td>4k</td>
<td>4096</td>
<td>2160</td>
<td>24</td>
<td>36</td>
<td>8847360</td>
<td>40</td>
<td>960</td>
<td>7.6</td>
</tr>
</tbody>
</table>

Note: this is excluding sound!  
Note: these are raw uncompressed data rates!
The playfield ⇨ GLIF
Multi-layer Network Description Language
What constitutes a Tb/s network?

CALIT2 has 8000 Gigabit drops ->? Terabit Lan?

Look at 80 core Intel processor
- cut it in two, left and right communicate 8 TB/s

Think back to teraflop computing!
- MPI makes it a teraflop machine

Massive parallel channels in hosts, NIC’s

TeraApps programming model supported by
- TFlops -> MPI / Globus
- TBytes -> OGSA/DAIS
- TPixels -> SAGE
- TSensors -> LOFAR, LHC, LOOKING, CineGrid, ...
- Tbit/s -> ?
Need for discrete parallelism

- it takes a core to receive 1 or 10 Gbit/s in a computer
- it takes one or two cores to deal with 10 Gbit/s storage
- same for Gigapixels
- same for 100’s of Gflops
- Capacity of every part in a system seems of same scale
- look at 80 core Intel processor
  - cut it in two, left and right communicate 8 TB/s
- massive parallel channels in hosts, NIC’s
- Therefore we need to go massively parallel allocating complete parts for the problem at hand!
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 neighbors of: 139.63.145.94
Internal links: {192.168.0.1, 139.63.145.94}
(...)
Getting neighbors of: 192.168.2.3
Internal links: {192.168.2.3}

Transaction on shortest path with tokens

```
nodePath = ConvertIndicesToNodes[
    ShortestPath[
        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
```

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.
Programmable Deterministic Service
**Recommendations**

1) The current draft fire report lacks vision, goals, does not define stakeholders, defines testbed but for what? Why?

2) Let research determine the testbed, not other way around! Allow only testbed if R&D, goals, vision is defined.

3) Look at the 10 year timescale in Internet development, underlying methods (paradigms) and technology.

4) 100 Gbit/s - Tb/s nets, look at grooming!

5) Populate the infrastructure with programmable L1, L2 and L3 objects and the virtualization towards applications

6) take into account the new internet exchanges (GOLE’s)

7) Security infrastructure <=> infrastructure security

8) Do not demand too many different details in a proposal. Writing a proposal is now already a huge effort.
Questions?

I did not talk about StarPlane
AAA Security
...