ClearStream

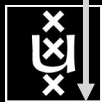
Cosmin Dumitru

Ralph Koning

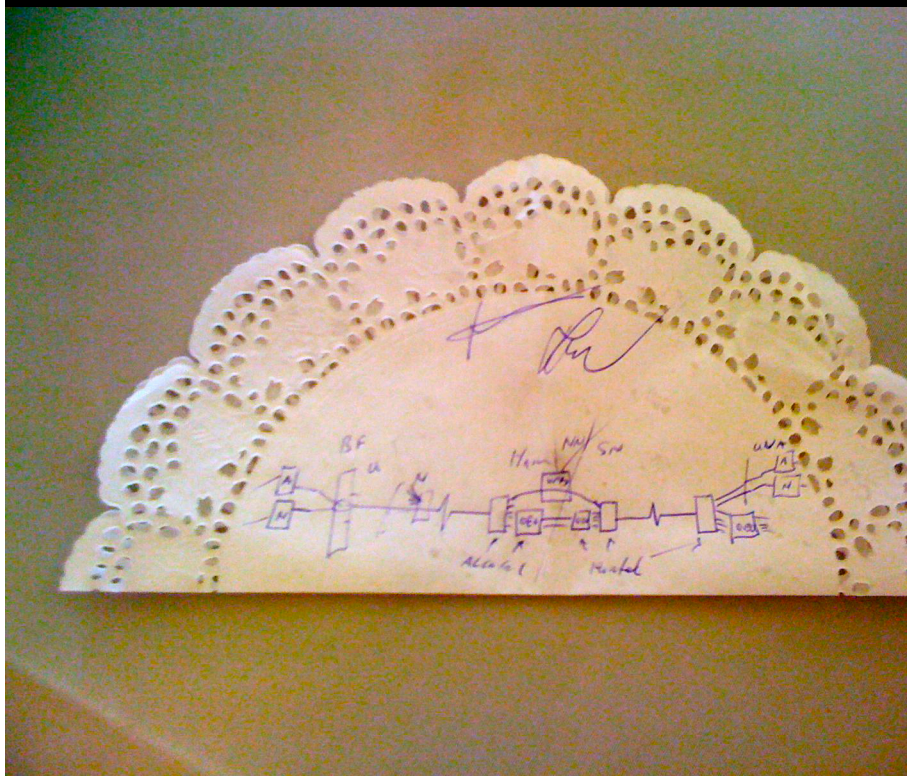
Cees de Laat

and many others (see poster)

University of Amsterdam



Alien light From idea to realisation!



40Gb/s alien wavelength transmission via a multi-vendor 10Gb/s DWDM infrastructure

NCF

Alien wavelength advantages

- Direct connection of customer equipment[1] → cost savings
- Avoid OEO regeneration → power savings
- Faster time to service[2] → time savings
- Support of different modulation formats[3] → extend network lifetime

Alien wavelength challenges

- Complex end-to-end optical path engineering in terms of linear (i.e. OSNR, dispersion) and non-linear (FWM, SPM, XPM, Raman) transmission effects for different modulation formats.
- Complex interoperability testing.
- End-to-end monitoring, fault isolation and resolution.
- End-to-end service activation.

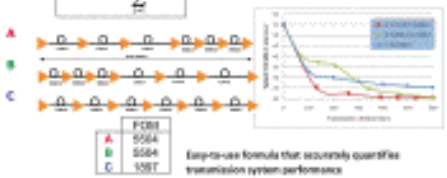
In this demonstration we will investigate the performance of a 40Gb/s PM-QPSK alien wavelength installed on a 10Gb/s DWDM infrastructure.

New method to present fiber link quality, FoM (Figure of Merit)

In order to quantify optical link grade, we propose a new method of representing system quality: the FOM (Figure of Merit) for concatenated fiber spans.

$$FOM = \sum_{n} 10^{\left[\frac{L_n}{10}\right]}$$

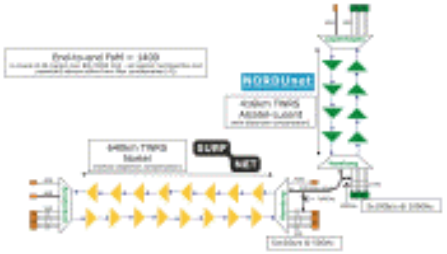
L_n: span losses in dB
N: number of spans



Easy-to-use formula that accurately quantifies transmission system performance

Transmission system setup

JOINT SURFnet/NORDUnet 40Gb/s PM-QPSK alien wavelength DEMONSTRATION.



Test results

Error-free transmission for 23 hours, 17 minutes → BER < 3.0 10^{-15}

Conclusions

- We have investigated experimentally the all-optical transmission of a 40Gb/s PM-QPSK alien wavelength via a concatenated native and third party DWDM system that both were carrying live 10Gb/s wavelengths.
- The end-to-end transmission system consisted of 1056 km of TWRS (TrueWave Reduced Slope) transmission fiber.
- We demonstrated error-free transmission (i.e. BER below 10-15) during a 23 hour period.
- More detailed system performance analysis will be presented in an upcoming paper.

NORTEL NORDUnet SURF NET

REFERENCES

[1] "OPERATIONAL SOLUTIONS FOR AN OPEN DWDM LAYER", G. GRIFFITH ET AL., OFC'2009 | [2] "NEXT OPTICAL TRANSPORT SERVICES", BARBARA S. SMITH, OFC'09 | [3] "OPEX SAVINGS OF ALL-OPTICAL CORE NETWORKS", ANDREW LORD AND CARL ENGLISH, ECOC'2009 | [4] SURFnet/GUARANTEE INTERNAL COMMUNICATION

ACKNOWLEDGEMENTS

WE ARE GRATEFUL TO NORDUNET FOR PROVIDING US WITH BANDWIDTH ON THEIR CONNECTION FOR THIS EXPERIMENT, AND ALSO FOR THEIR SUPPORT AND ASSISTANCE DURING THE EXPERIMENTS. WE ALSO ACKNOWLEDGE TERAPLA AND HUAWEI FOR THEIR INTEROPERABILITY AND SIMULATION SUPPORT.

GLIF 2010

40 Gbps Lambda Based on Ethernet From UVA cluster To CERN cluster

ClearStream

End-to-End Ultra Fast Transmission Over a Wide Area 40 Gbit/s Lambda

University of Amsterdam
Cosmin Dumitru
Cees de Laat
Ralph Koning
SURFnet
Erik-Jan Bos
Gerben van Malenstein
Ciena
David Yeung
Jan-Willem Ellion
Harry Peng
Kevin McKernan
Martin Bluethner
VU University Amsterdam
Kees Verstoep
Henri Bal
Mellanox
Erez Cohen
Bill Lee

Utilizing shared expertise in advanced photonic, leading edge hardware and high-performance computing, the team created a network application testbed using the 1650 km Cross Border Fiber between NetherLight and CERNLight, lit by SURFnet, connecting servers equipped with 40 Gigabit Ethernet network interface at the University of Amsterdam to remote servers with corresponding interfaces at GLIF 2010 in Geneva.

Network Setup

The Mellanox ConnectX-2 EN 40GbE is the first network interface that allows single stream ethernet transport far exceeding the common 10Gbps boundary limit. The achieved throughtput is 26Gbps from CPU to CPU which is the practical limit of the PCI-E interface.

The network infrastructure is based on Ciena's Optical Multiservice Edge (OME) 6500 equipped with 40 GbE interfaces, which enables data speeds to be seamlessly upgraded from 10 Gbps to 40 Gbps.

Application Setup

The DiVinE application is MPI based and in this setup uses TCP/IP as its network backend. DiVinE's runtime system is optimized to achieve good performance despite the very intensive traffic rate and high WAN latency over long distance.

We also use a server with basic UDP and TCP test tools to tune and measure capacities. Going beyond 10 Gbps leads to new challenges in applications, operating system tuning and system architecture design as new bottlenecks appear.

Special attention needs to be given to the setup of multi-core machines in order to have the best I/O performance and maximize the network throughput. During the demo the PCI-E x8 2.0 interface of the network card is saturated when using UDP or TCP traffic.

DiVinE

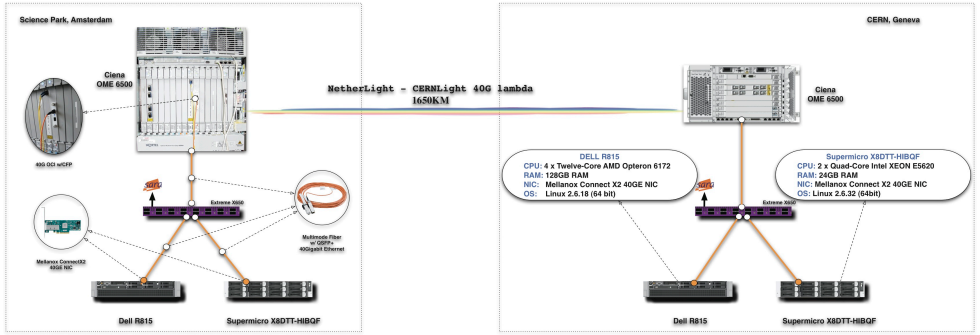
DiVinE is a tool for LTL model checking and reachability analysis of discrete distributed systems. The tool is able to efficiently exploit the aggregate computing power of multiple network-interconnected multi-cored workstations in order to deal with extremely large verification tasks.

Cluster-in-a-box

The Dell R815 is a 2U server powered by 48 AMD Opteron 6100 cores which make it as one of the densest x64 servers available on the market and is used to run the DiVinE application.

High Performance Node

Using a flexible I/O architecture, the Supermicro X8DTT with two quad-core Intel E5620 CPUs, allows extreme speeds of over 25 Gbps to be reached.



System and Network Engineering Research Group, Universiteit van Amsterdam
http://science.uva.nl/research/sne

Universiteit van Amsterdam · SURFnet · ciena · Mellanox Technologies · vrije Universiteit amsterdam · CERN · sara · extreme networks

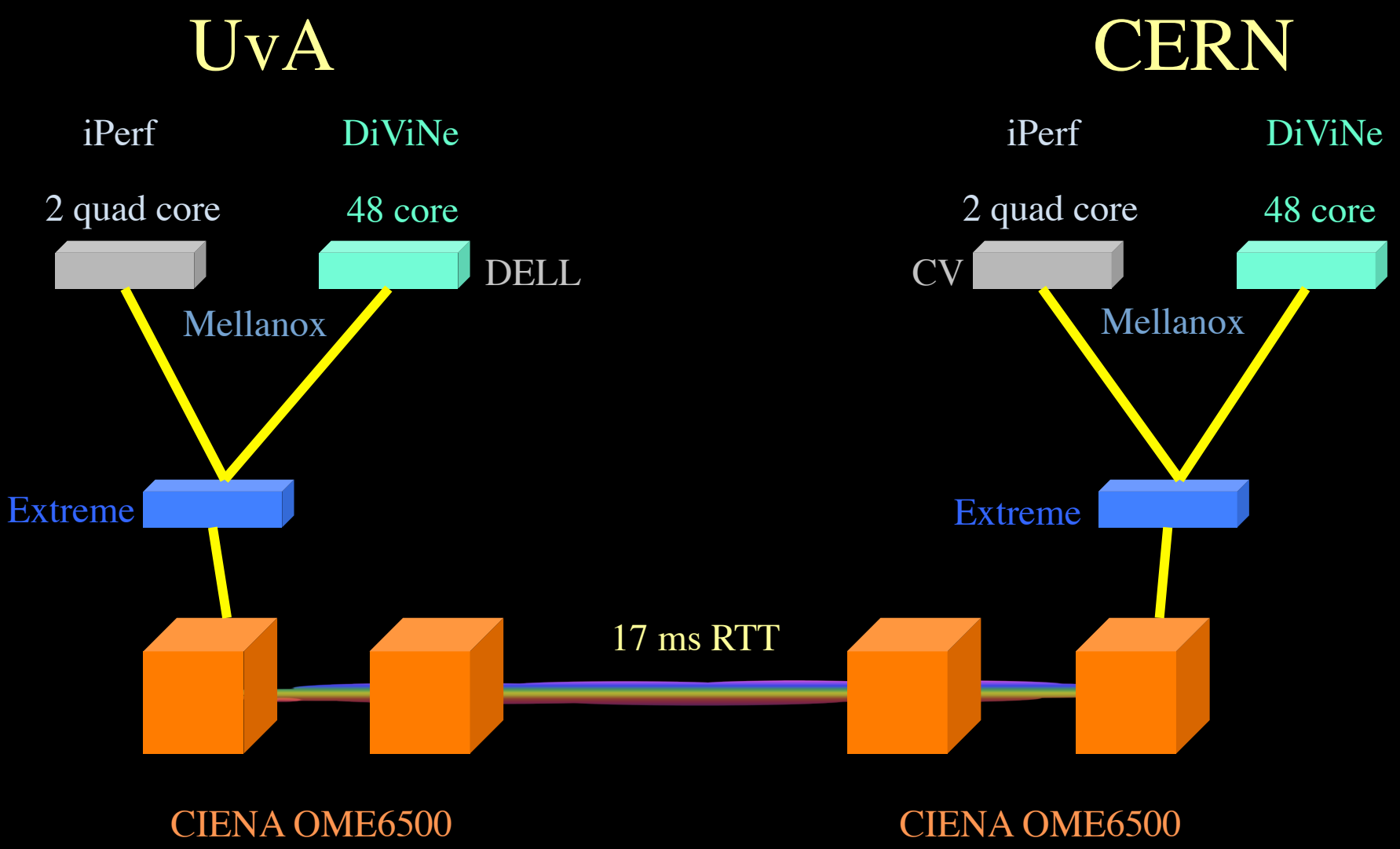
Goals

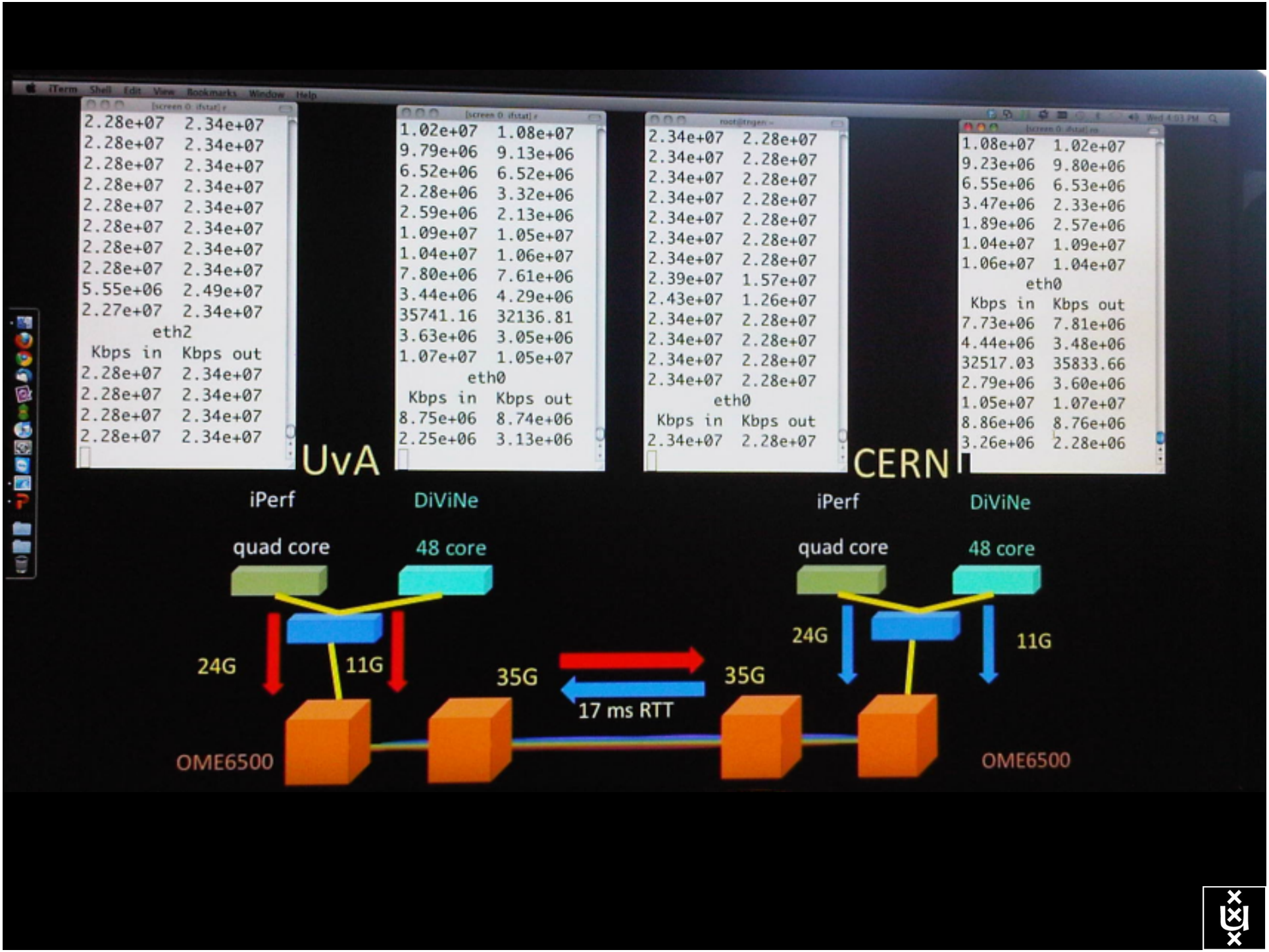
- To demonstrate single stream single wave performance end to end operating without any obstruction.
- To break the 10 Gbps barrier.
- To test a real application on this infrastructure.

What we show

- Single server to single server performance memory to memory
  - Single stream single Lambda TCP
  - Multiple stream single Lambda TCP
  - UDP streaming
- 48 core system to 48 core system
  - Running DiViNe model checker
  - Many small messages
  - Cluster in a box!

Setup


UvA
CERN
iPerf
DiViNe
iPerf
DiViNe
2 quad core
48 core
2 quad core
48 core
DELL
CV
Mellanox
Mellanox
Extreme
Extreme
17 ms RTT
CIENA OME6500
CIENA OME6500


2.28e+07 2.34e+07
2.28e+07 2.34e+07
2.28e+07 2.34e+07
2.28e+07 2.34e+07
2.28e+07 2.34e+07
2.28e+07 2.34e+07
2.28e+07 2.34e+07
2.28e+07 2.34e+07
5.55e+06 2.49e+07
2.27e+07 2.34e+07
eth2
Kbps in Kbps out
2.28e+07 2.34e+07
2.28e+07 2.34e+07
2.28e+07 2.34e+07
2.28e+07 2.34e+07
1.02e+07 1.08e+07
9.79e+06 9.13e+06
6.52e+06 6.52e+06
2.28e+06 3.32e+06
2.59e+06 2.13e+06
1.09e+07 1.05e+07
1.04e+07 1.06e+07
7.80e+06 7.61e+06
3.44e+06 4.29e+06
35741.16 32136.81
3.63e+06 3.05e+06
1.07e+07 1.05e+07
eth0
Kbps in Kbps out
8.75e+06 8.74e+06
2.25e+06 3.13e+06
2.34e+07 2.28e+07
2.34e+07 2.28e+07
2.34e+07 2.28e+07
2.34e+07 2.28e+07
2.34e+07 2.28e+07
2.34e+07 2.28e+07
2.34e+07 2.28e+07
2.39e+07 1.57e+07
2.43e+07 1.26e+07
2.34e+07 2.28e+07
2.34e+07 2.28e+07
2.34e+07 2.28e+07
2.34e+07 2.28e+07
eth0
Kbps in Kbps out
2.34e+07 2.28e+07
1.08e+07 1.02e+07
9.23e+06 9.80e+06
6.55e+06 6.53e+06
3.47e+06 2.33e+06
1.89e+06 2.57e+06
1.04e+07 1.09e+07
1.06e+07 1.04e+07
eth0
Kbps in Kbps out
7.73e+06 7.81e+06
4.44e+06 3.48e+06
32517.03 35833.66
2.79e+06 3.60e+06
1.05e+07 1.07e+07
8.86e+06 8.76e+06
3.26e+06 2.28e+06
UvA
CERN
iPerf
DiViNe
iPerf
DiViNe
quad core
48 core
quad core
48 core
24G
11G
35G
35G
17 ms RTT
24G
11G
OME6500
OME6500

Preliminary results

- Single flow iPerf 1 core -> 21 Gbps
- Single flow iPerf 1 core <> -> 15+15 Gbps
- Multi flow iPerf 2 cores -> 25 Gbps
- Multi flow iPerf 2 cores <> -> 23+23 Gbps
- DiViNe <> -> 11 Gbps
- Multi flow iPerf + DiVine -> 35 Gbps
- Multi flow iPerf + DiVine <> -> 35 + 35 Gbps

Hybrid computing

Routers ⇐ ⇒ Supercomputers

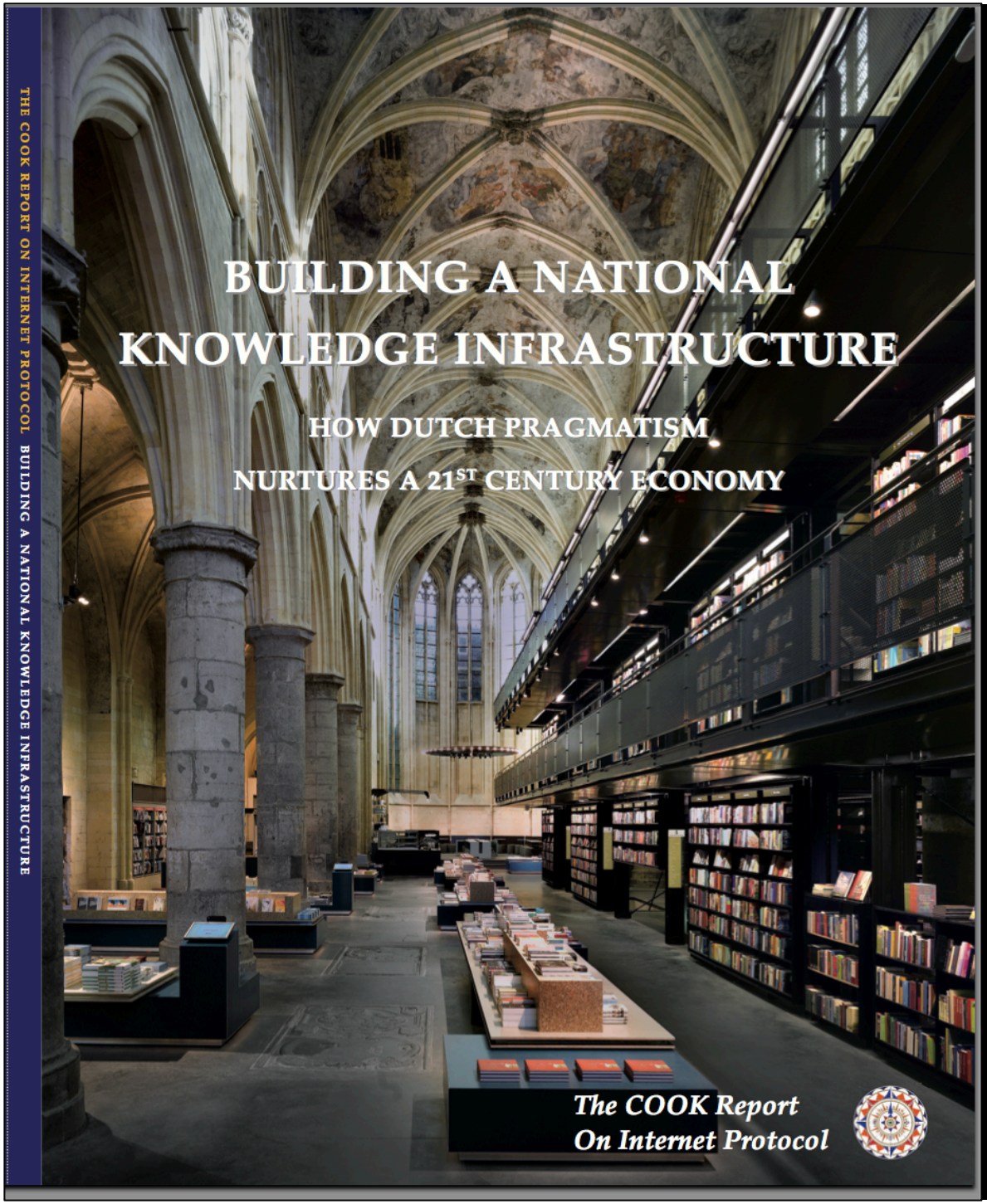
Ethernet switches ⇐ ⇒ Grid & Cloud

Photonic transport ⇐ ⇒ GPU's

What matters:

Energy consumption/multiplication

Energy consumption/bit transported



Questions ?

CookReport
feb 2009 and feb-mar 2010

november '08
interview with
Kees Neggers (SURFnet),
Cees de Laat (UvA)

and furthermore
on november '09

Wim Liebrandt (SURF),
Bob Hertzberger (UvA) and
Hans Dijkman (UvA)

BSIK projects
GigaPort &
VL-e / e-Science

SURF NET

ext.delaat.net