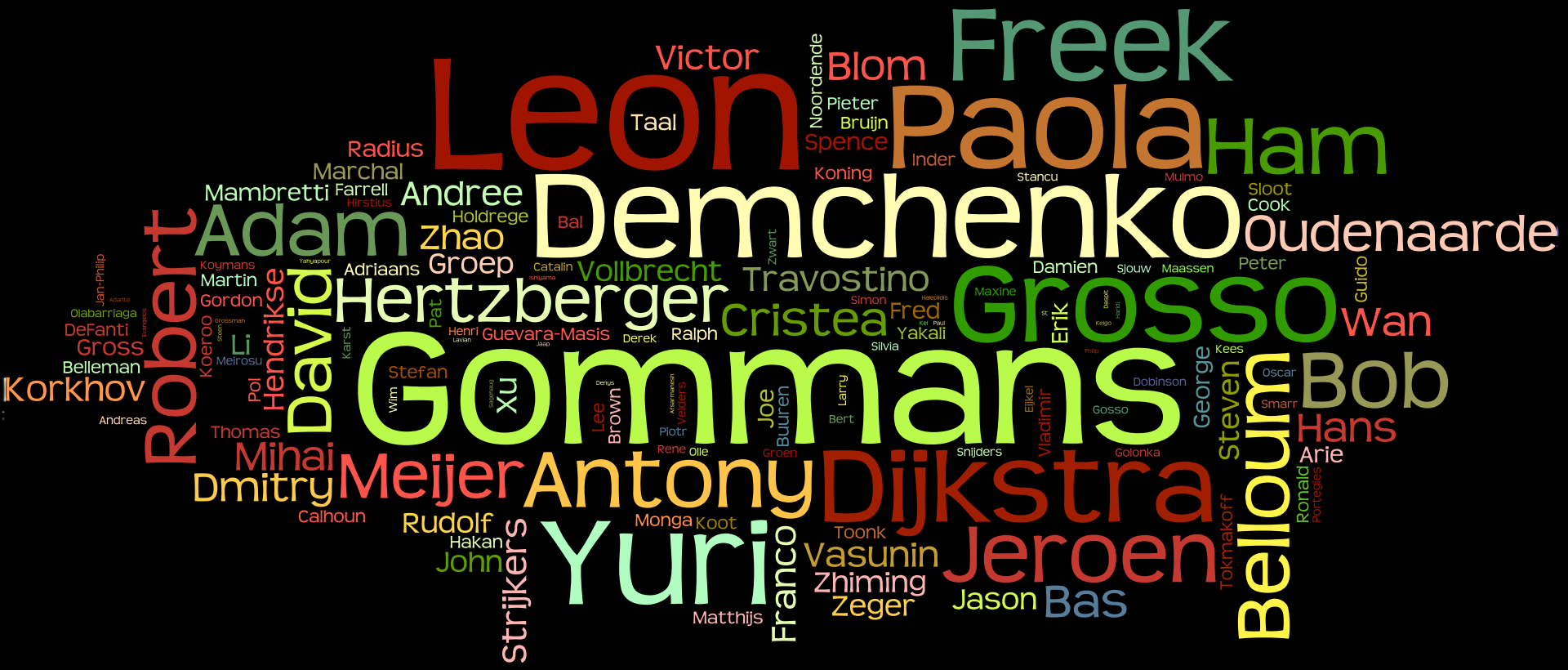
Smart Cyber Infrastructure for Big Data Processing

Paola Grosso & Cees de Laat



Data Science Research Center Amsterdam

COMMIT/

ENVRI

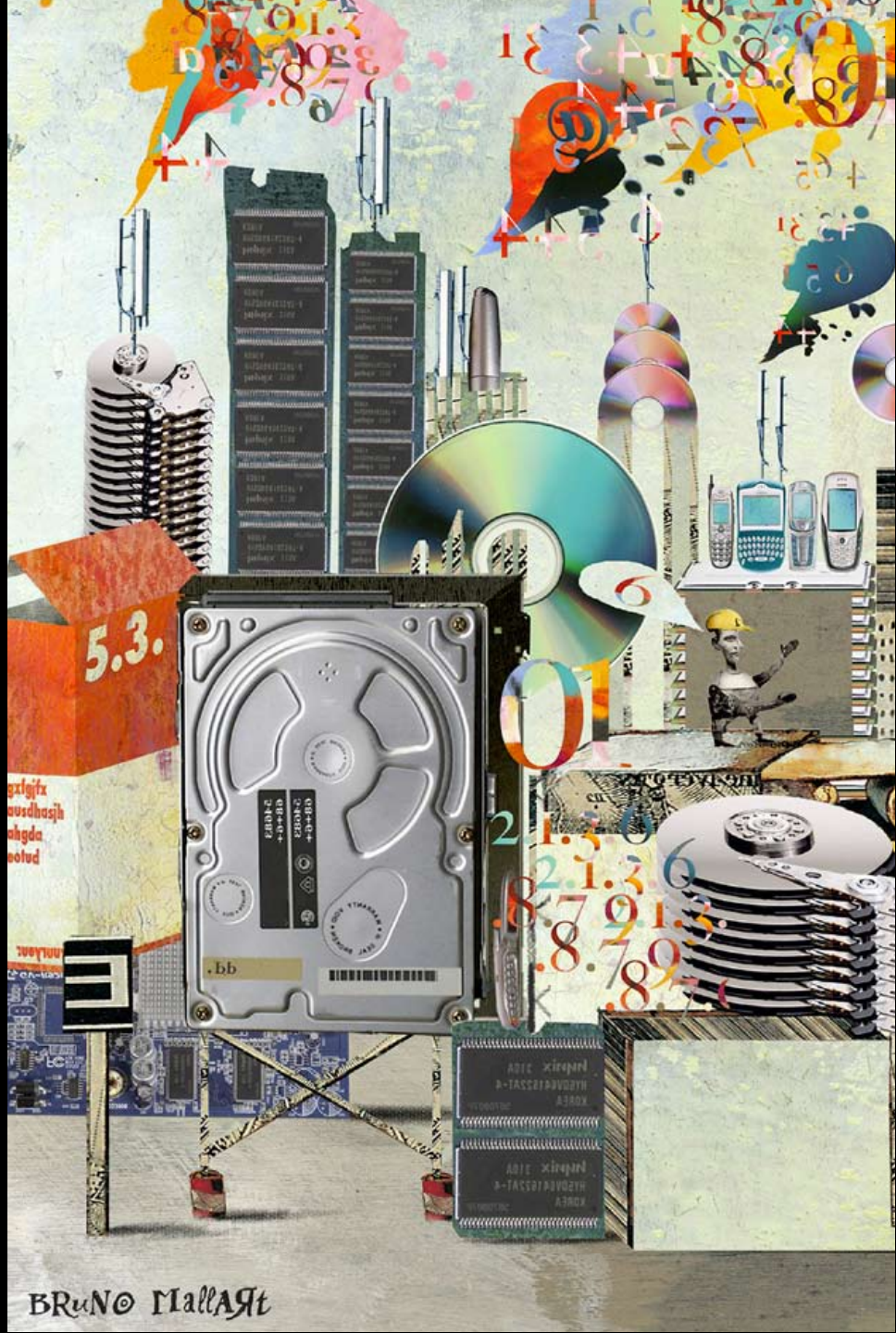
From King's Dutch Academy of Sciences

The Dutch Research Agenda

"Information technology (IT) now permeates all aspects of public, commercial, social, and personal life. bank cards, satnav, and weather radar... IT has become completely indispensable."

"But to guarantee the reliability and quality of constantly bigger and more complicated IT, we will need to find answers to some fundamental questions!"

https://www.knaw.nl/nl/actueel/publicaties/the-dutch-research-agenda/@@download/pdf_file/20111029.pdf



Reduction of Complexity by Integration

By combining services such as telephony, television, data, and computing capacity within a single network, we can cut down on complexity, energy consumption and maintenance.

- How can we describe and analyze complex information systems effectively?
- How can we specify and measure the quality and reliability of a system?
- How can we combine various different systems?
- How can we design systems in which separate processors can co-operate efficiently via mutual network connections within a much larger whole?
- Can we design information systems that can diagnose their own malfunctions and perhaps even repair them?
- How can we specify, predict, and measure system performance as effectively as possible?

SNE addresses a.o. the **highlighted** questions!

http://www.knaw.nl/Content/Internet_KNAW/publicaties/pdf/20111029.pdf

Mission

Can we create smart and safe data processing infrastructures that can be tailored to diverse application needs?

- *Capacity*
- *Capability*
- *Security*
- *Sustainability*
- *Resilience*

Mission

Can we create smart and safe data processing infrastructures that can be tailored to diverse application needs?

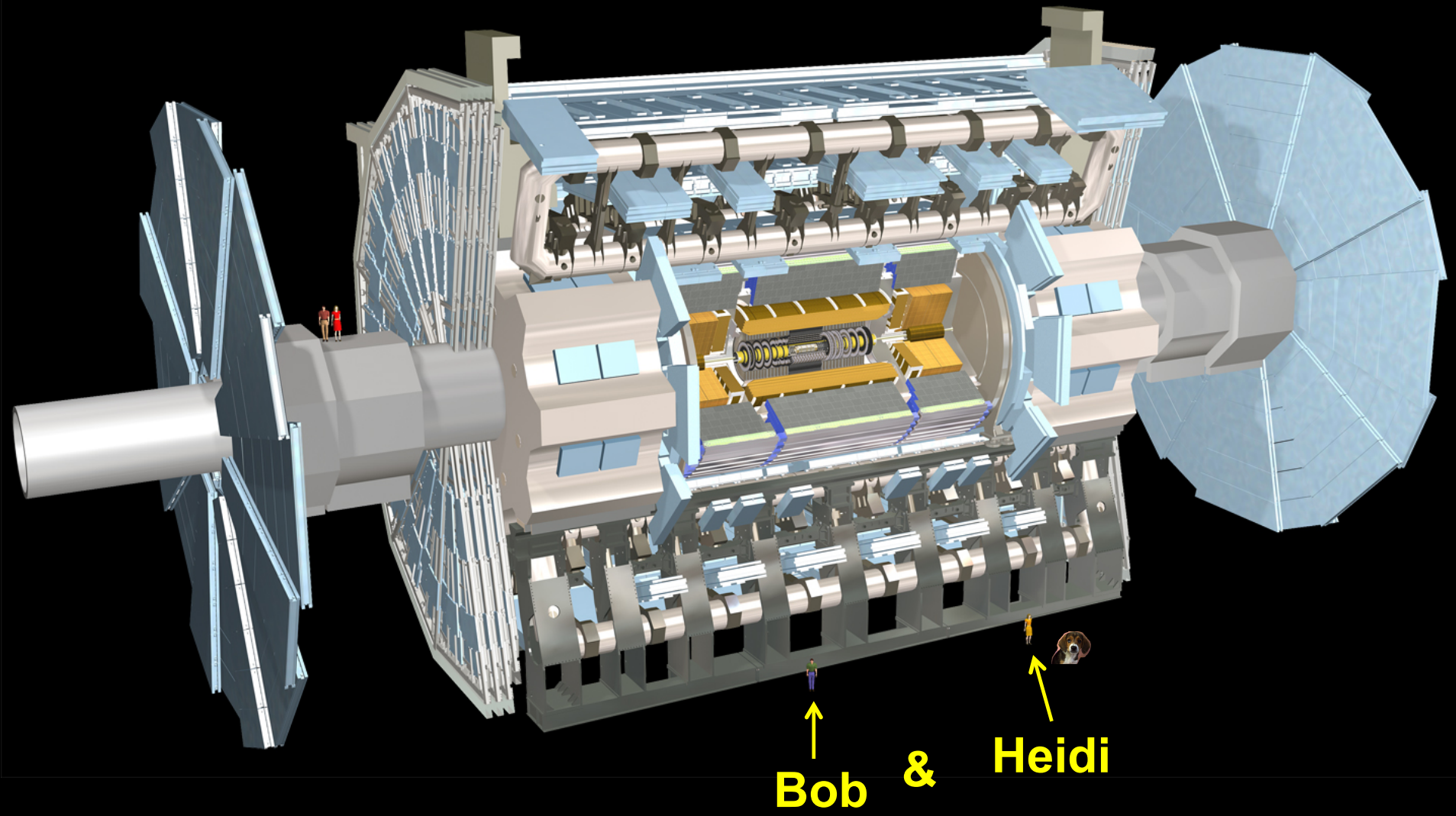
- *Capacity*
  - *Bandwidth on demand, QoS, architectures, photonics, performance*
- *Capability*
  - *Programmability, virtualization, complexity, semantics, workflows*
- *Security*
  - *Anonymity, integrity of data in distributed data processing*
- *Sustainability*
  - *Greening infrastructure, awareness*
- *Resilience*
  - *Systems under attack, failures, disasters*

Mission

Can we create smart and safe data processing infrastructures that can be tailored to diverse application needs?

- ***Capacity***
  - ***Bandwidth on demand, QoS, architectures, photonics, performance***
- *Capability*
  - *Programmability, virtualization, complexity, semantics, workflows*
- *Security*
  - *Anonymity, integrity of data in distributed data processing*
- *Sustainability*
  - *Greening infrastructure, awareness*
- *Resilience*
  - *Systems under attack, failures, disasters*

ATLAS detector @ CERN Geneve



What Happens in an **Internet Minute?**

And **Future Growth** is **Staggering**

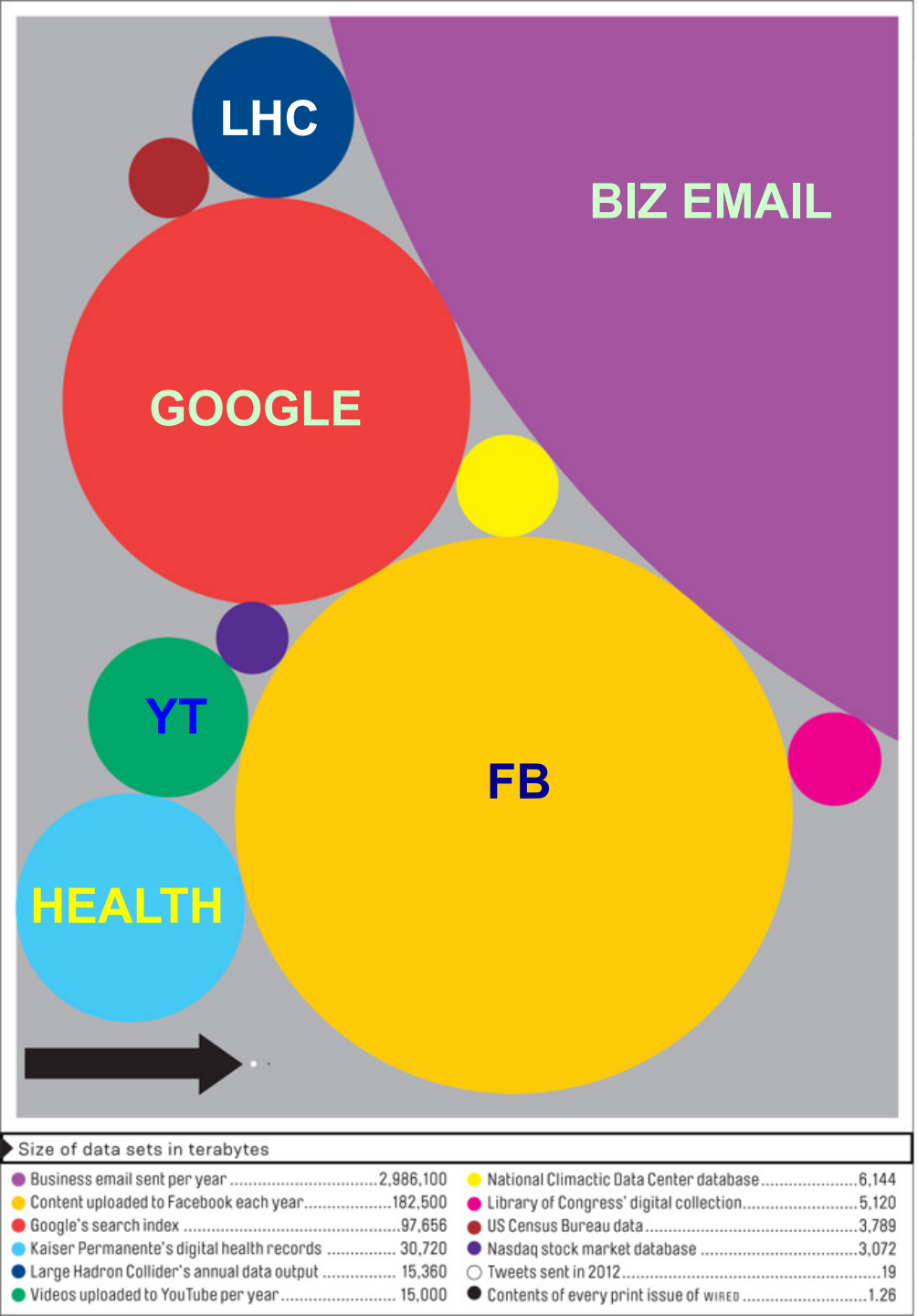
Today, the number of networked devices = the global population

By **2015**, the number of networked devices = 2x the global population

In **2015**, it would take you **5 years**

to view all video crossing IP networks each **second**

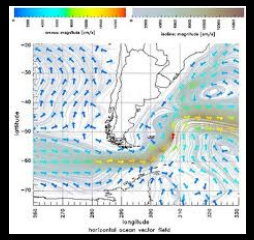
intel

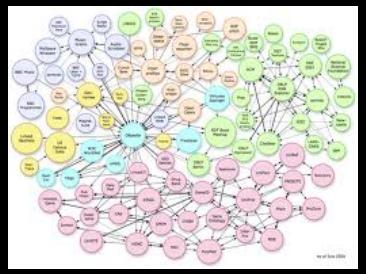


There
is
always
a
bigger
fish

Internet developments

... more data!





... more realtime!

... more users!

Internet developments

... more data!

Speed

Volume

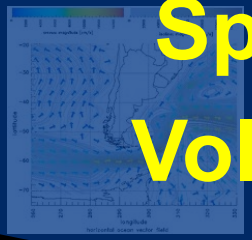
Deterministic

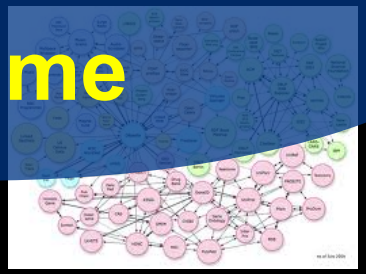
Real-time

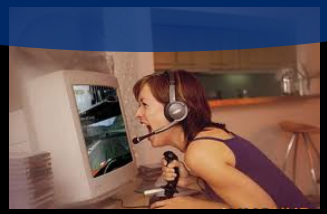
Scalable

Secure

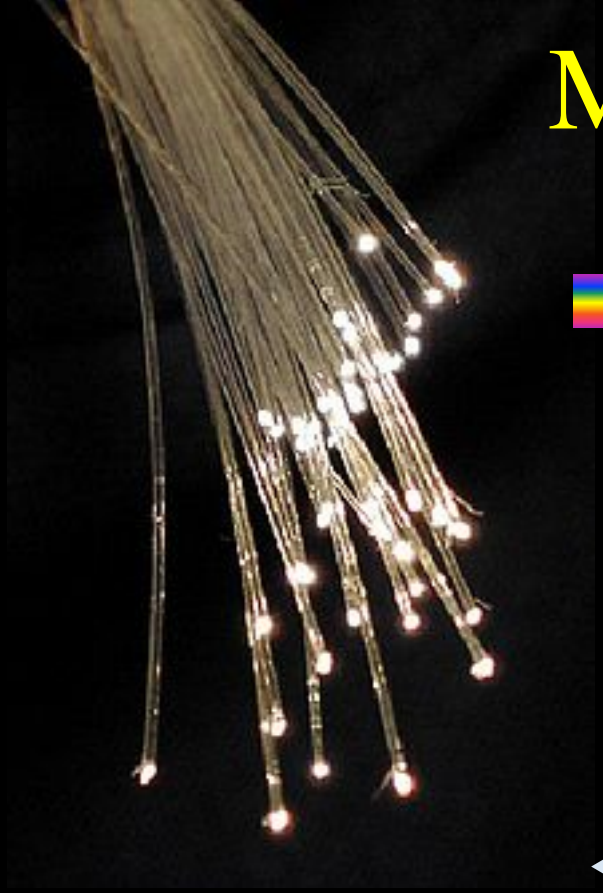
... more users!

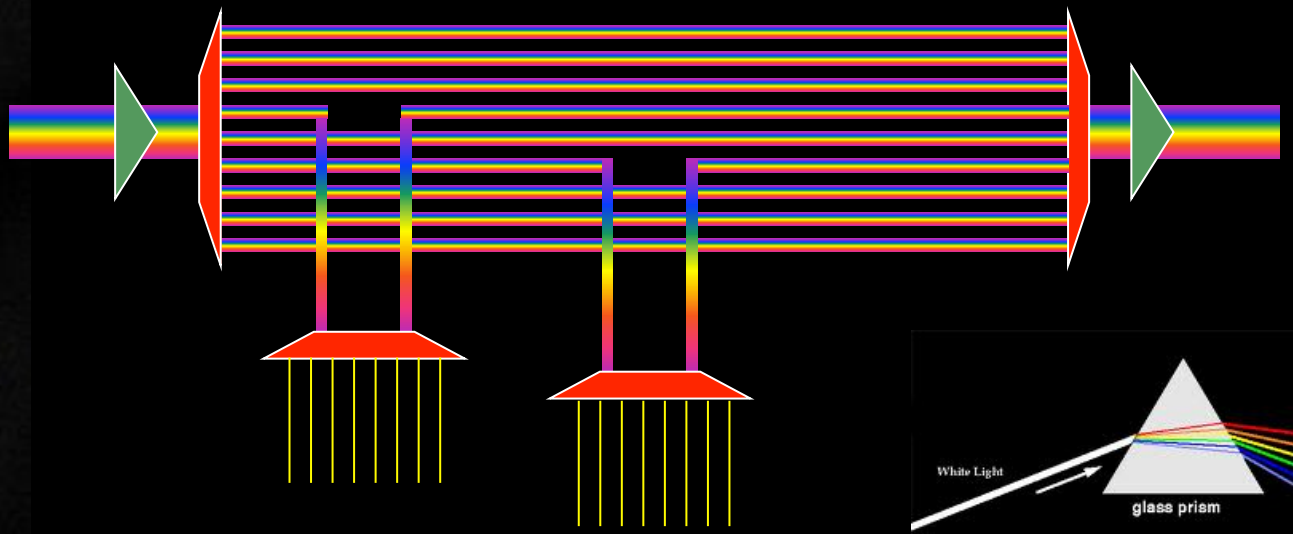


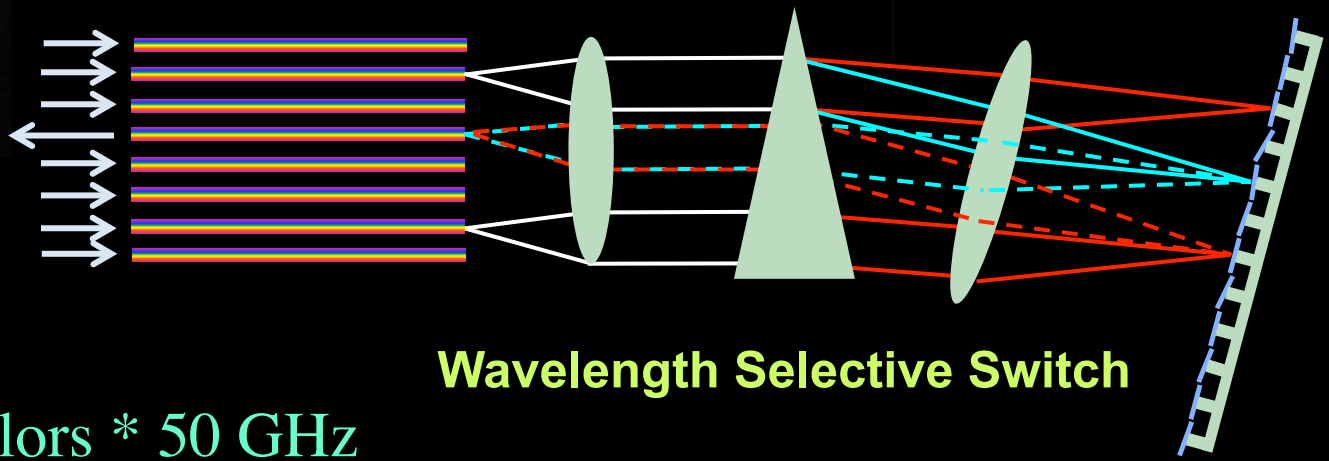




Multiple colors / Fiber







Wavelength Selective Switch

Per fiber: ~ 80-100 colors * 50 GHz

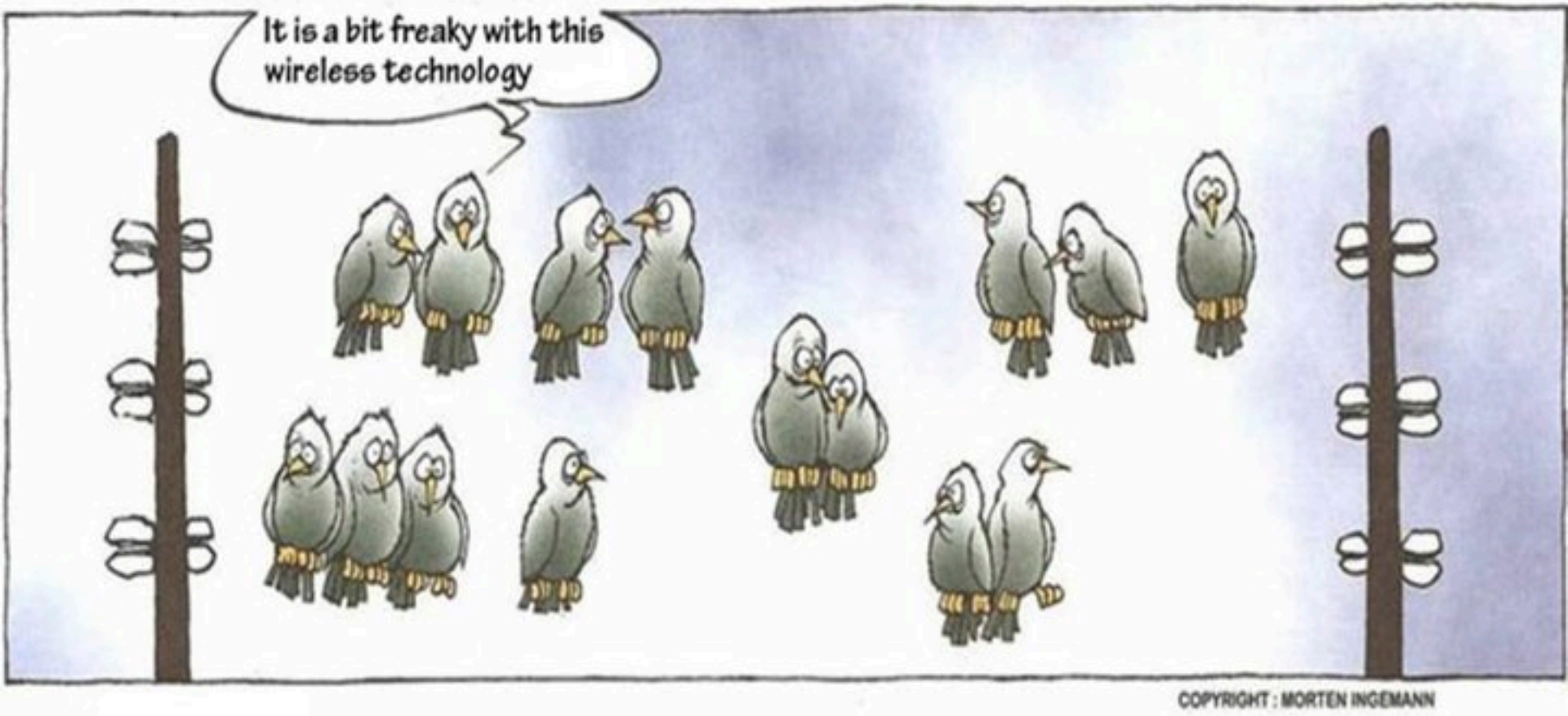
Per color: 10 – 40 – 100 Gbit/s

BW * Distance ~ $2*10^{17}$ bm/s

New: Hollow Fiber!

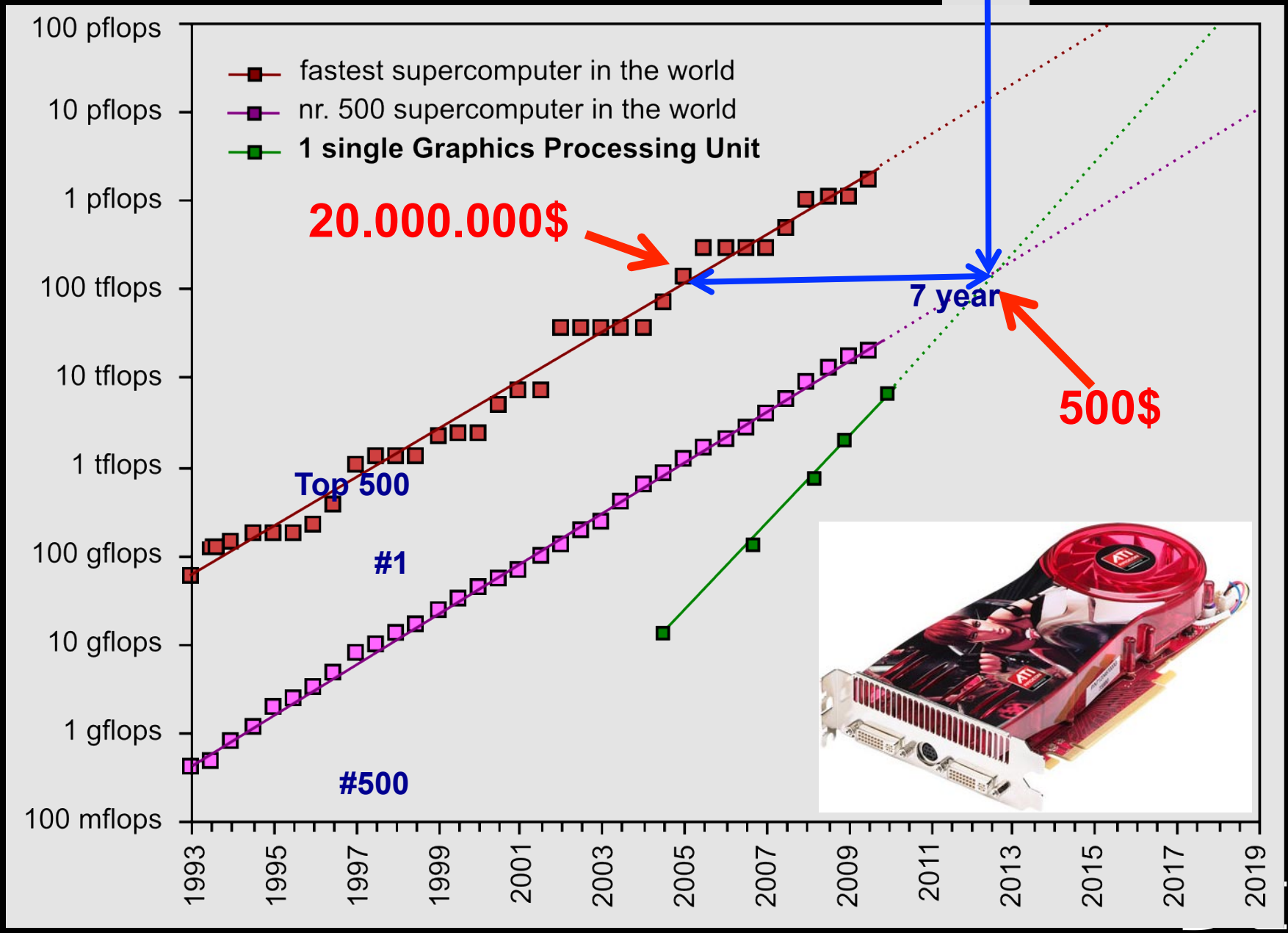
➔ less RTT!

Wireless Networks



protocol LAN due to the easy comparison and convenience in the **digital home**. While consumer PC products has just started to migrate to a much higher bandwidth of 802.11n wireless LAN now working on next-generation standard definition is already in progress.

GPU cards are distruptive!

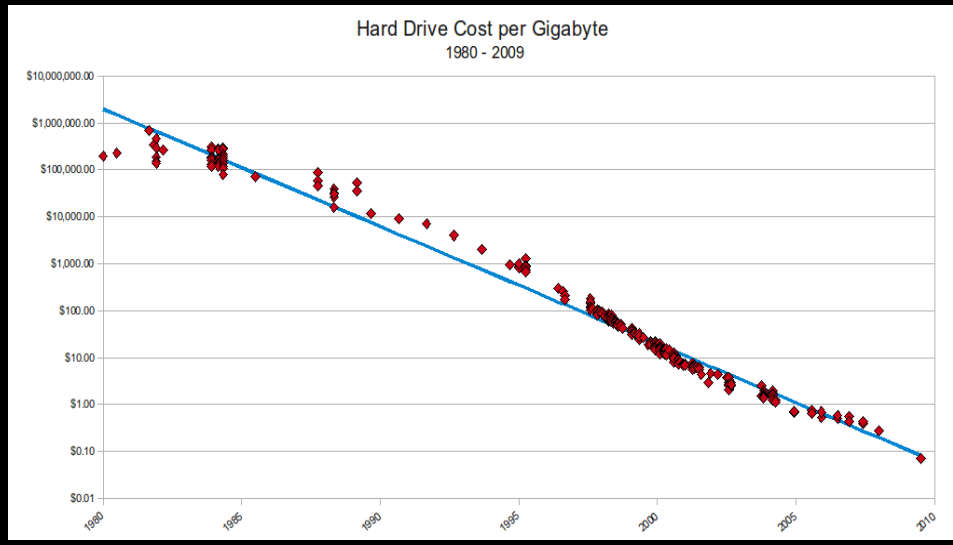


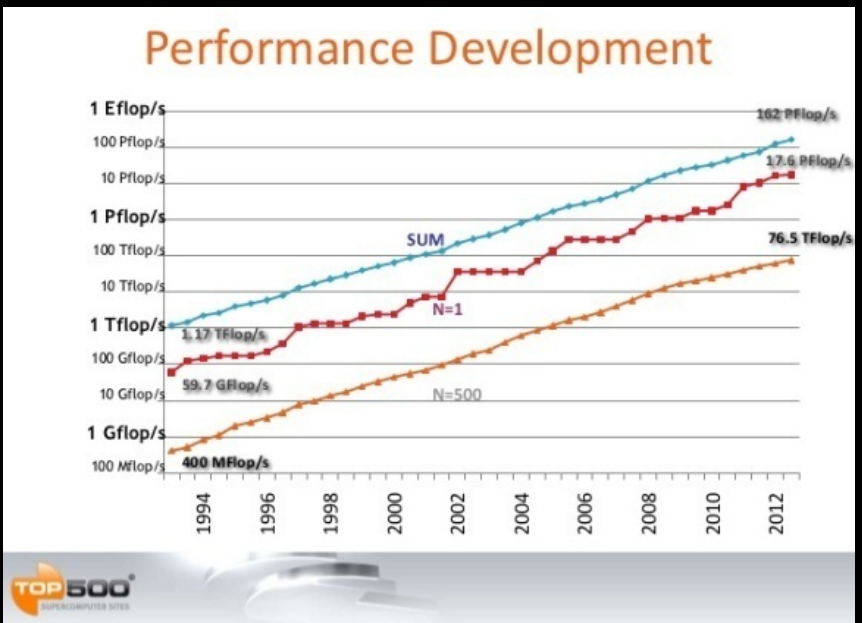
Reliable and Safe!

This omnipresence of IT makes us not only strong but also vulnerable.

- A virus, a hacker, or a system failure can instantly send digital shockwaves around the world.

The hardware and software that allow all our systems to operate is becoming bigger and more complex all the time, and the capacity of networks and data storage is increasing by leaps and bounds.



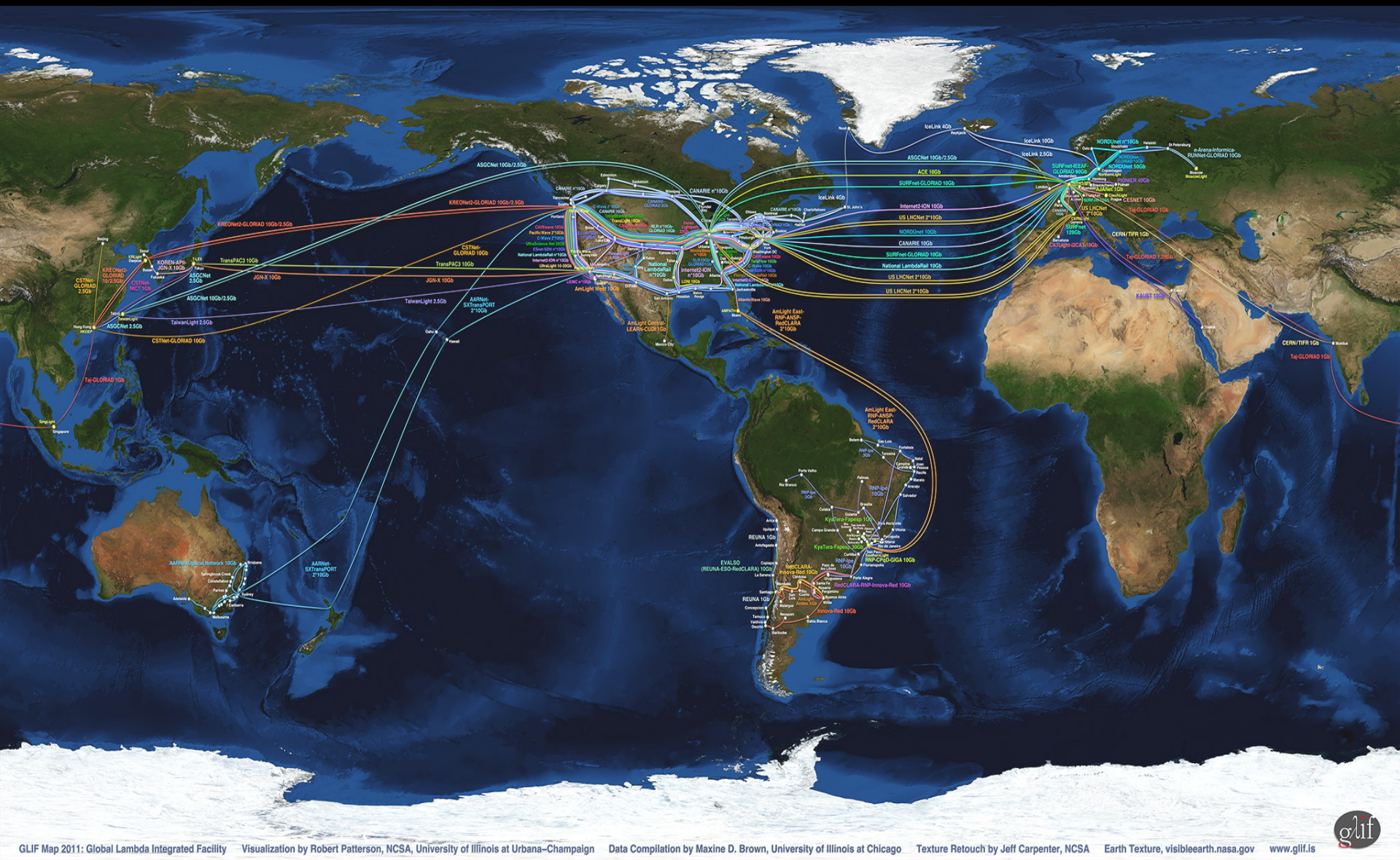


We will soon reach the limits of what is currently feasible and controllable.

https://www.knaw.nl/shared/resources/actueel/publicaties/pdf/20111029.pdf

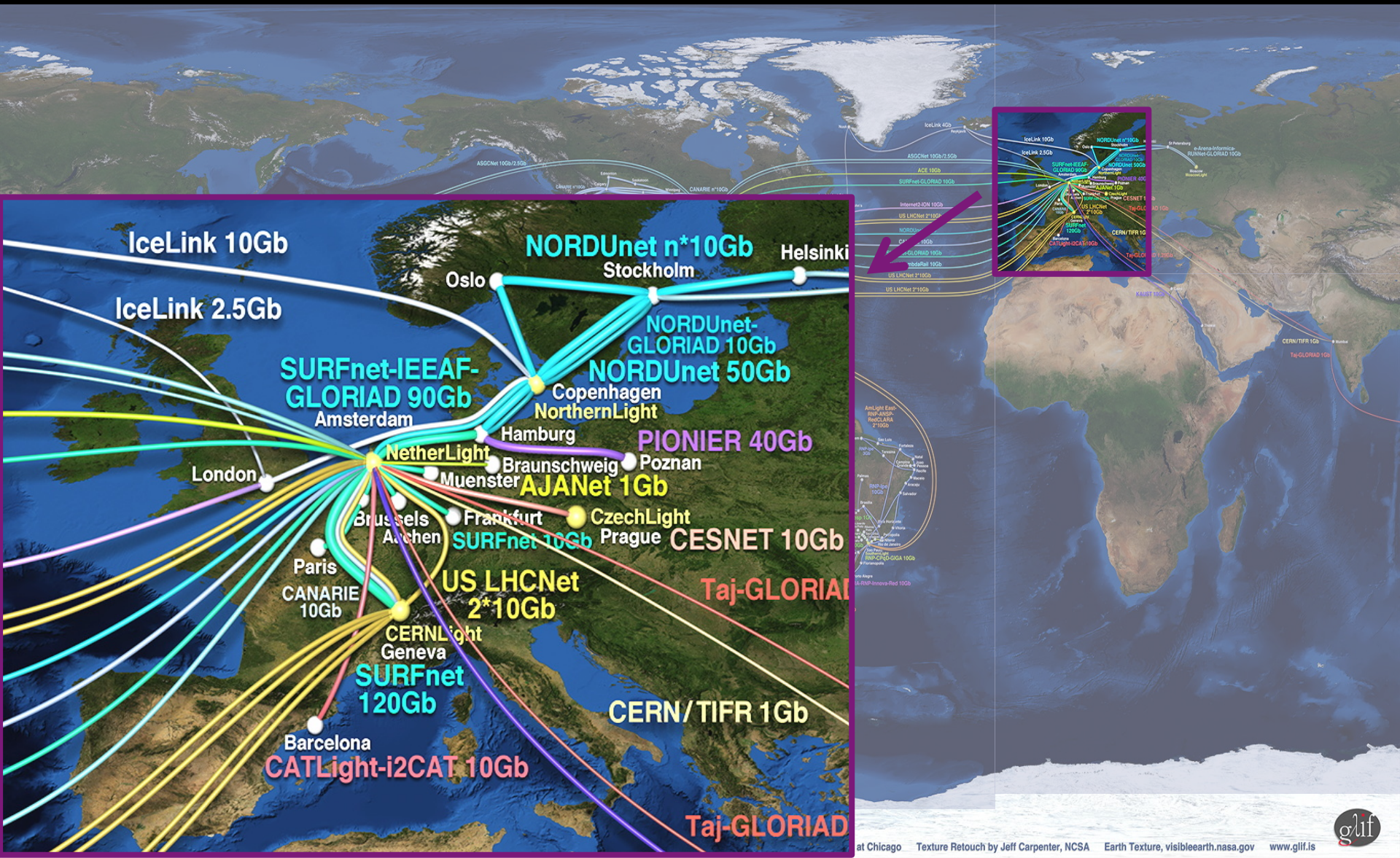
The GLIF – LightPaths around the World

F Dijkstra, J van der Ham, P Grosso, C de Laat, “A path finding implementation for multi-layer networks”, Future Generation Computer Systems 25 (2), 142-146.



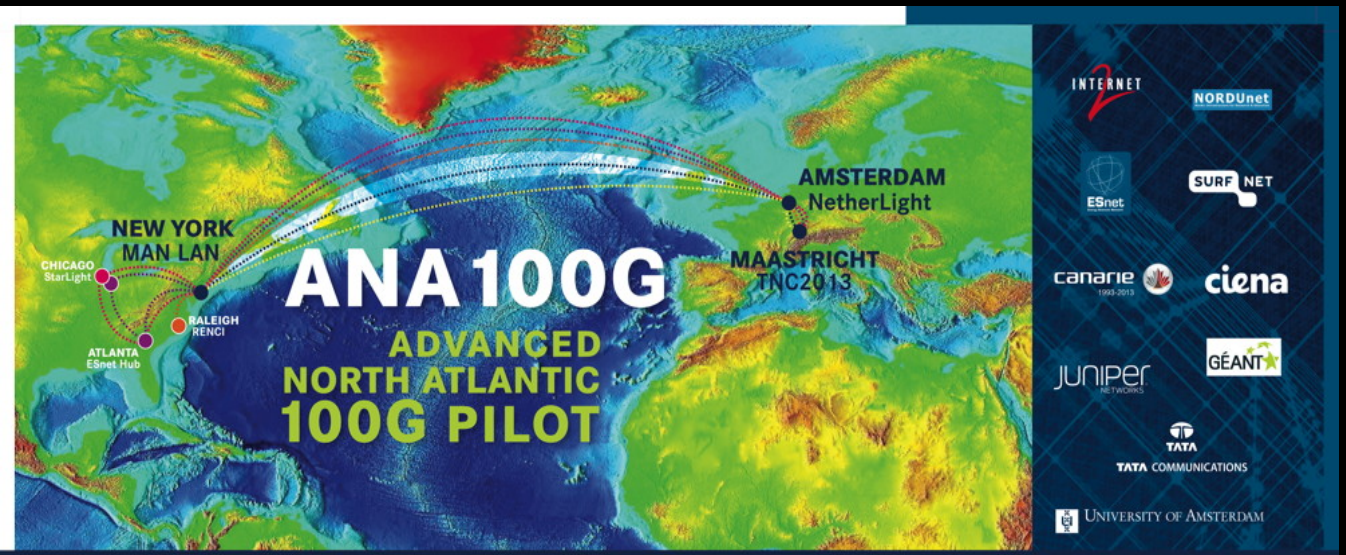
Amsterdam is a major hub in The GLIF

F Dijkstra, J van der Ham, P Grosso, C de Laat, “A path finding implementation for multi-layer networks”, Future Generation Computer Systems 25 (2), 142-146.



ExoGeni @ OpenLab - UvA

Installed and up June 3th 2013



TNC2013 DEMOS JUNE, 2013

DEMO	TITLE	OWNER	AFFILIATION	E-MAIL	A-SIDE	Z-SIDE	PORTS(S) MAN LAN	PORTS(S) TNC2013	DETAILS
1	Big data transfers with multipathing, OpenFlow and MPTCP	Ronald van der Pol	SURFnet	ronald.vanderpol@surfnet.nl	TNC/MECC, Maastricht NL	Chicago, IL	Existing 100G link between internet2 and ESnet	2x40GE (Juniper)+ 2x 10GE (OME6500)	In this demonstration we show how multipathing, OpenFlow and Multipath TCP (MPTCP) can help in large file transfers between data centres (Maastricht and Chicago). An OpenFlow application provisions multiple paths between the servers and MPTCP will be used on the servers to simultaneously send traffic across all those paths. This demo uses 2x40G on the transatlantic 100G link. ESnet provides 2x40G between MAN LAN and StarLight, ACE and USLHCnet provide additional 10GEs.
2	Visualize 100G traffic	Inder Monga	ESnet	imonga@es.net					Using an SNMP feed from the Juniper switch at TNC2013,and/or Brocade ALZS node in MANLAN, this demo would visualize the total traffic on the link, of all demos aggregated. The network diagram will show the transatlantic topology and some of the demo topologies.
3	How many modern servers can fill a 100Gbps Transatlantic Circuit?	Inder Monga	ESnet	imonga@es.net	Chicago, Ill	TNC showfloor	1x 100GE	8x 10GE	In this demonstration, we show that with the proper tuning and tool, only 2 hosts on each continent can generate almost 80Gbps of traffic. Each server has 4 100 NICS connected to a 40G virtual circuit, and has iperf3 running to generate traffic. ESnet's new "iperf3" throughput measurement tool, still in 'beta', combines the best features from other tools such as iperf, nuttcp, and netperf. See: https://my.es.net/demos/tnc2013/
4	First European ExoGENI at Work	Jeroen van der Ham	UvA	vdham@uva.nl	RENCI, NC	UvA, Amsterdam, NL	1x 10GE	1x 10GE	The ExoGENI racks at RENCI and UvA will be interconnected over a 10G pipe and be on continuously, showing GENI connectivity between Amsterdam and the rest of the GENI nodes in the USA.
5	Up and down North Atlantic @ 100G	Michael Enrico	DANTE	michael.enrico@dante.net	TNC showfloor	TNC showfloor	1x 100GE	1x 100GE	The DANTE 100GE test set will be placed at the TNC2013 showfloor and connected to the Juniper at 100G. When this demo is running a loop @ MAN LAN's Brocade switch will ensure that the traffic sent to MAN LAN returns to the showfloor. On display is the throughput and RTT (to show the traffic traveled the Atlantic twice)

Connected via the new 100 Gb/s transatlantic To US-GENI

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.

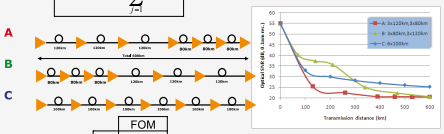
In this demonstration we will investigate the performance of a 40Gb/s PM-QPSK alien wavelength installed on a 10Gb/s DWDM infrastructure.

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.

$$FOM = \sum_{j=1}^{N} 10^{\left(\frac{L_j}{10}\right)}$$

Lj, span losses in dB
N, number of spans

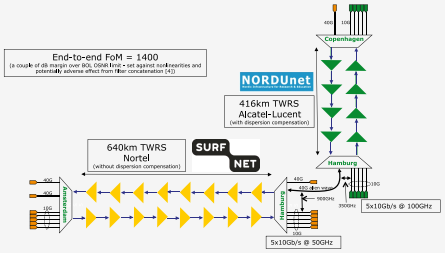


	FOM
A	5504
B	5504
C	1897

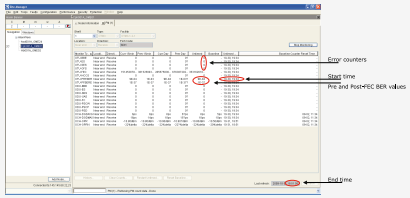
Easy-to-use formula that accurately quantifies transmission system performance

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^{-16}

Conclusions

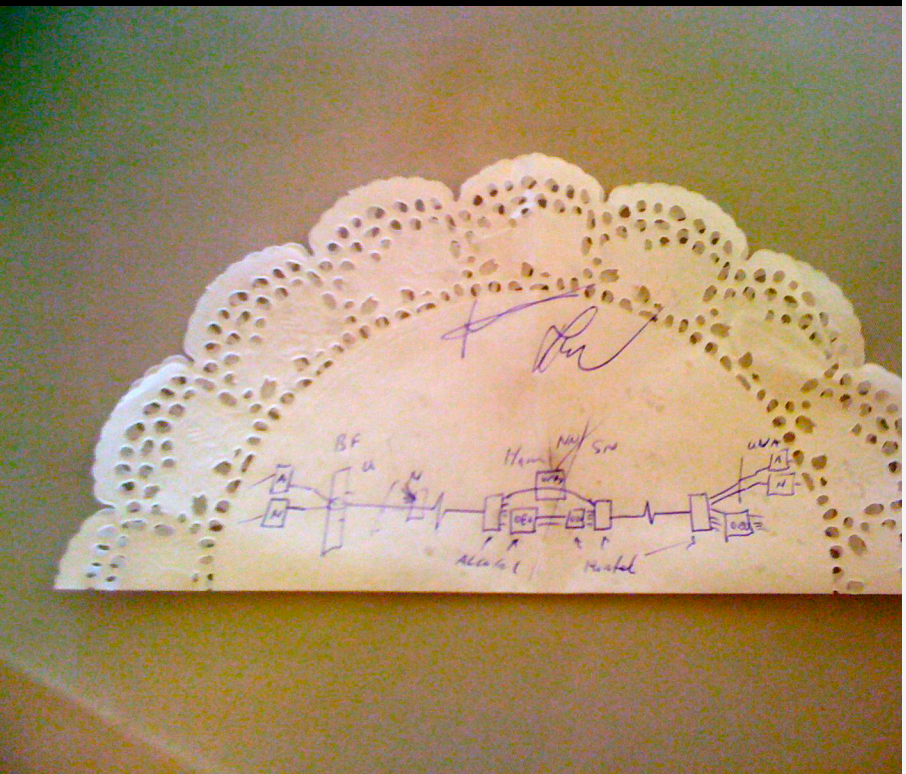
- We have investigated experimentally the all-optical transmission of a 40Gb/s PM-QPSK alien wavelength via a concatenated native and third party DWDM system that both were carrying live 10Gb/s wavelengths.
- The end-to-end transmission system consisted of 1056 km of TWRS (TrueWave Reduced Slope) transmission 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 telindus together with ISTY SURF NET

REFERENCES [1] "OPERATIONAL SOLUTIONS FOR AN OPEN DWDM LAYER", O. GERSTEL ET AL, OFC'2009 | [2] "AT&T OPTICAL TRANSPORT SERVICES", BARBARA E. SMITH, OFC'09
[3] "OPEX SAVINGS OF ALL-OPTICAL CORE NETWORKS", ANDREW LORD AND CARL ENGINEER, ECOC2009 | [4] NORTEL/SURFNET INTERNAL COMMUNICATION

ACKNOWLEDGEMENTS WE ARE GRATEFUL TO NORDUNET FOR PROVIDING US WITH BANDWIDTH ON THEIR DWDM LINK FOR THIS EXPERIMENT AND ALSO FOR THEIR SUPPORT AND ASSISTANCE DURING THE EXPERIMENTS. WE ALSO ACKNOWLEDGE TELINDUS AND NORTEL FOR THEIR INTEGRATION WORK AND SIMULATION SUPPORT

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.

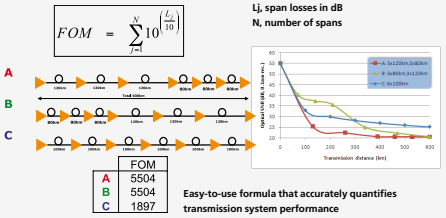
In this demonstration we will investigate the performance of a 40Gb/s PM-QPSK alien wavelength installed on a 10Gb/s DWDM infrastructure.

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.

$$FOM = \sum_{j=1}^{N} 10^{\left(\frac{L_j}{10}\right)}$$

Lj, span losses in dB
N, number of spans

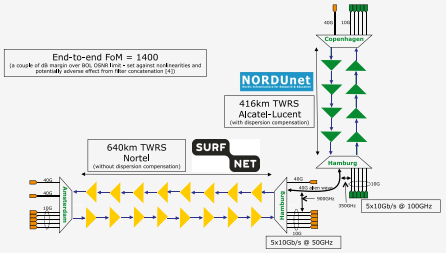


	FOM
A	5504
B	5504
C	1897

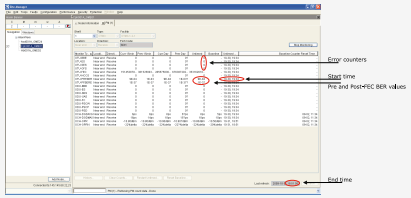
Easy-to-use formula that accurately quantifies transmission system performance

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^{-16}

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 telindus together with ISIT SURF NET

REFERENCES [1] "OPERATIONAL SOLUTIONS FOR AN OPEN DWDM LAYER", O. GERSTEL ET AL, OFC'2009 | [2] "AT&T OPTICAL TRANSPORT SERVICES", BARBARA E. SMITH, OFC'09
[3] "OPEX SAVINGS OF ALL-OPTICAL CORE NETWORKS", ANDREW LORD AND CARL ENGINEER, ECOC2009 | [4] NORTEL/SURFNET INTERNAL COMMUNICATION

ACKNOWLEDGEMENTS WE ARE GRATEFUL TO NORDUNET FOR PROVIDING US WITH BANDWIDTH ON THEIR DWDM LINK FOR THIS EXPERIMENT AND ALSO FOR THEIR SUPPORT AND ASSISTANCE DURING THE EXPERIMENTS. WE ALSO ACKNOWLEDGE TELINDUS AND NORTEL FOR THEIR INTEGRATION WORK AND SIMULATION SUPPORT

Mission

Can we create smart and safe data processing infrastructures that can be tailored to diverse application needs?

- *Capacity*
  - *Bandwidth on demand, QoS, architectures, photonics, performance*
- ***Capability***
  - ***Programmability, virtualization, complexity, semantics, workflows***
- *Security*
  - *Anonymity, integrity of data in distributed data processing*
- *Sustainability*
  - *Greening infrastructure, awareness*
- *Resilience*
  - *Systems under attack, failures, disasters*

INDL

An effort started in 2010 (in parallel with our involvement in the FP7 projects Geysers and NOVI).

The goal was to capture the concept of virtualization in computing infrastructures and to describe the storage and computing capabilities of the resources.

A key feature is the decoupling of virtualization, connectivity and functionalities.

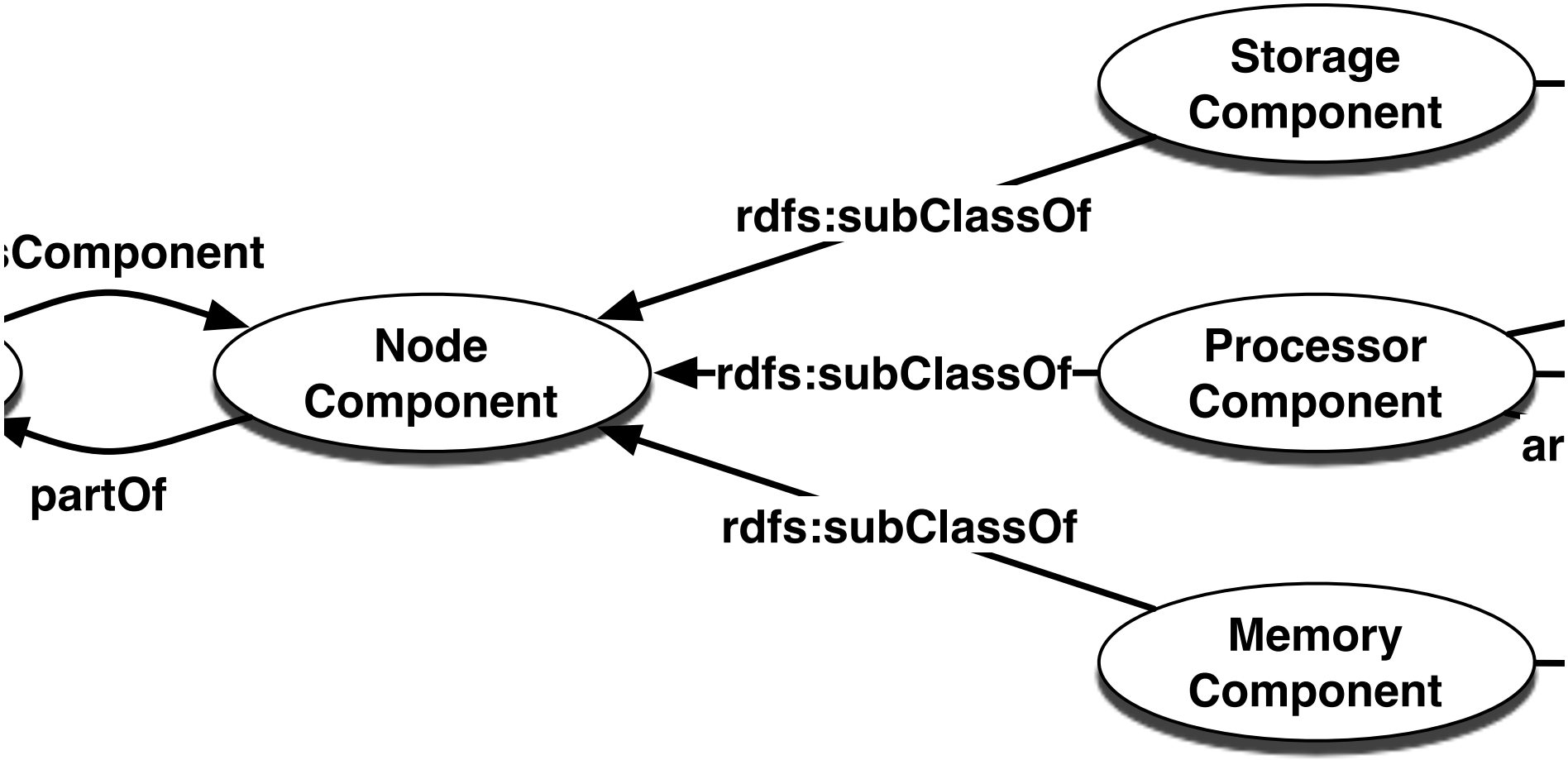
It is built upon the NML ontology.

It uses the nml:node concept as basic entity to describe resources in computing infrastructures.

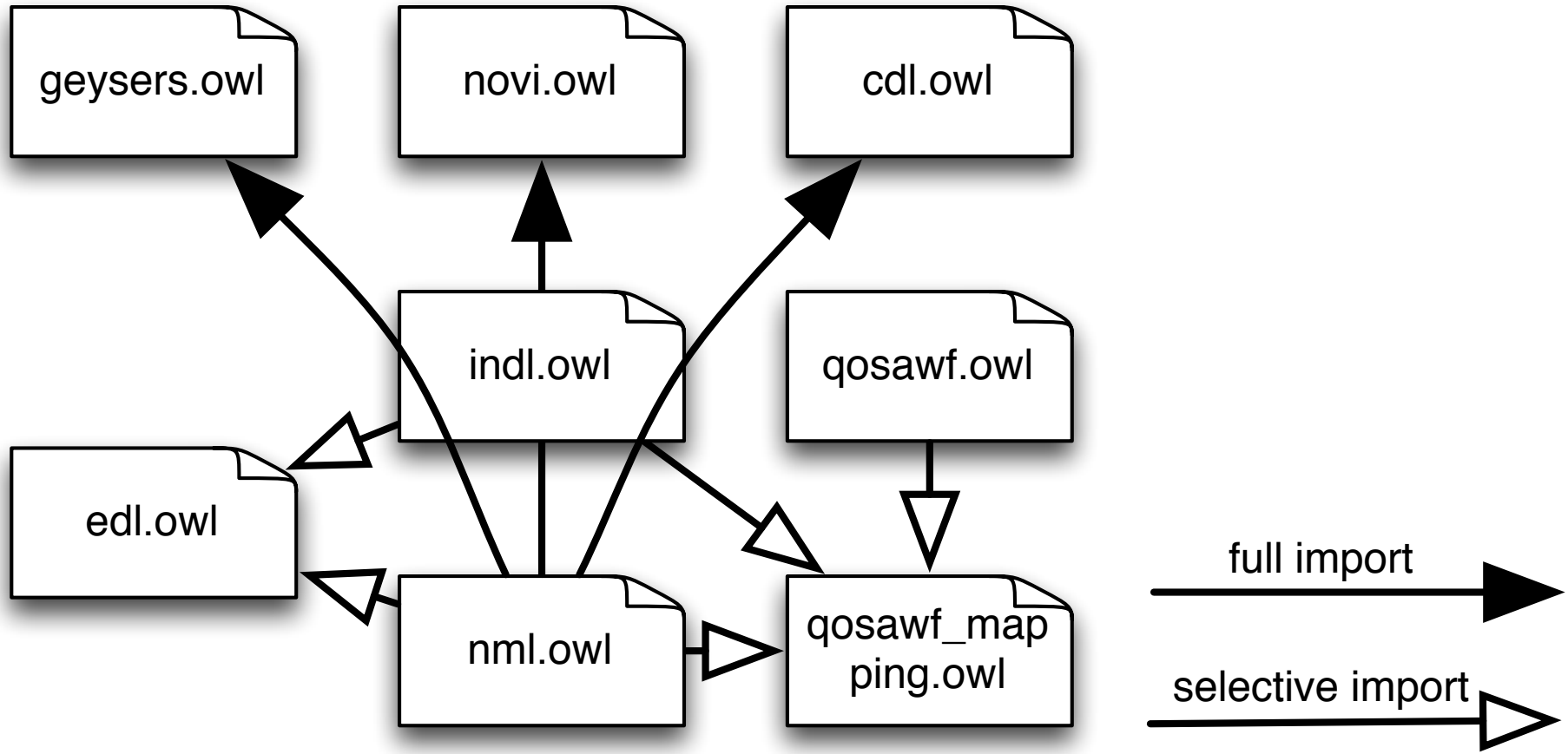
It can be used as:

- a stand-alone model (i.e. without any network descriptions),
- in combination with NML by importing the NML ontology into the INDL definition.

Node components



Our connecting models



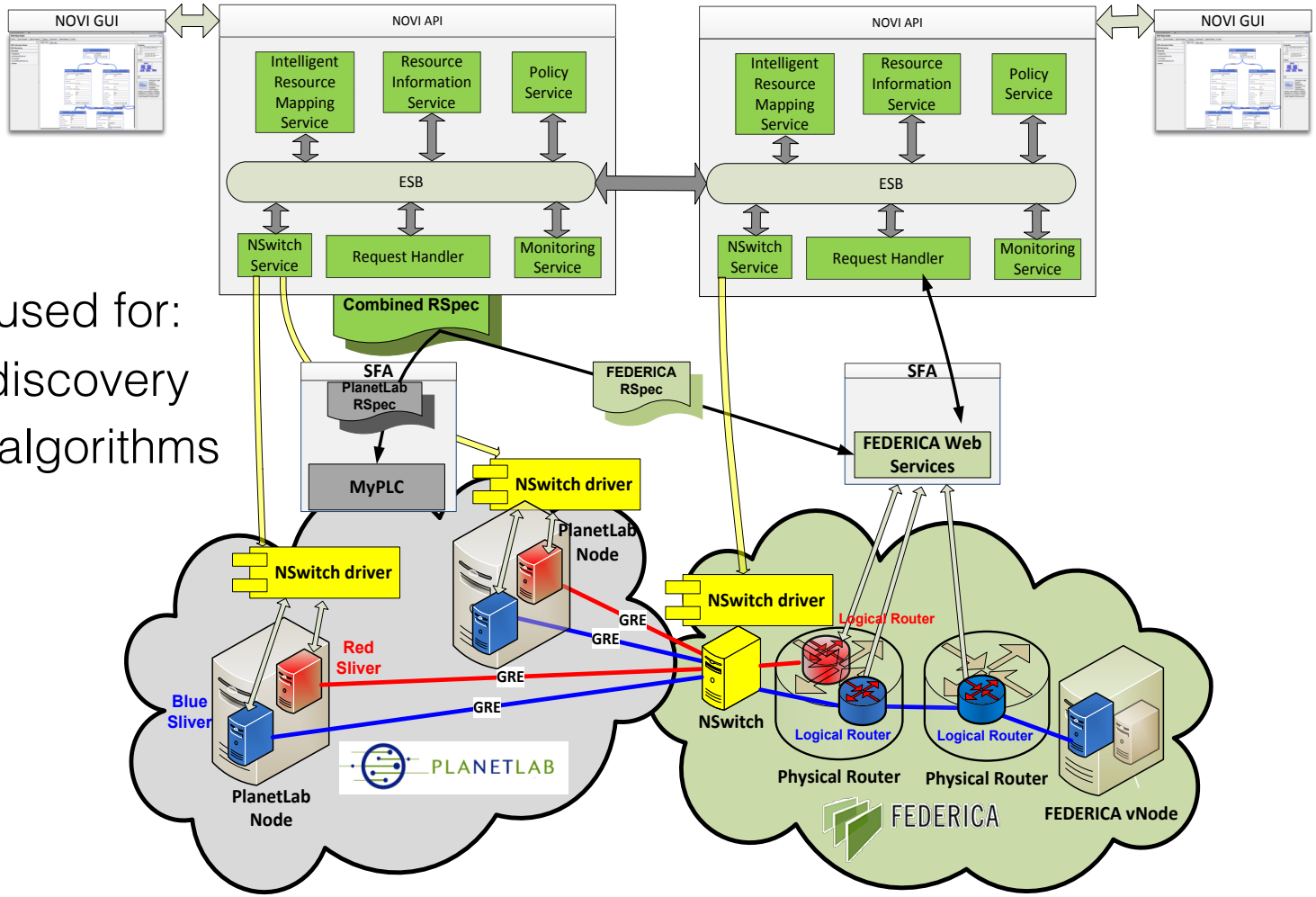
COMMIT/

ENVRI

NOVI Federation

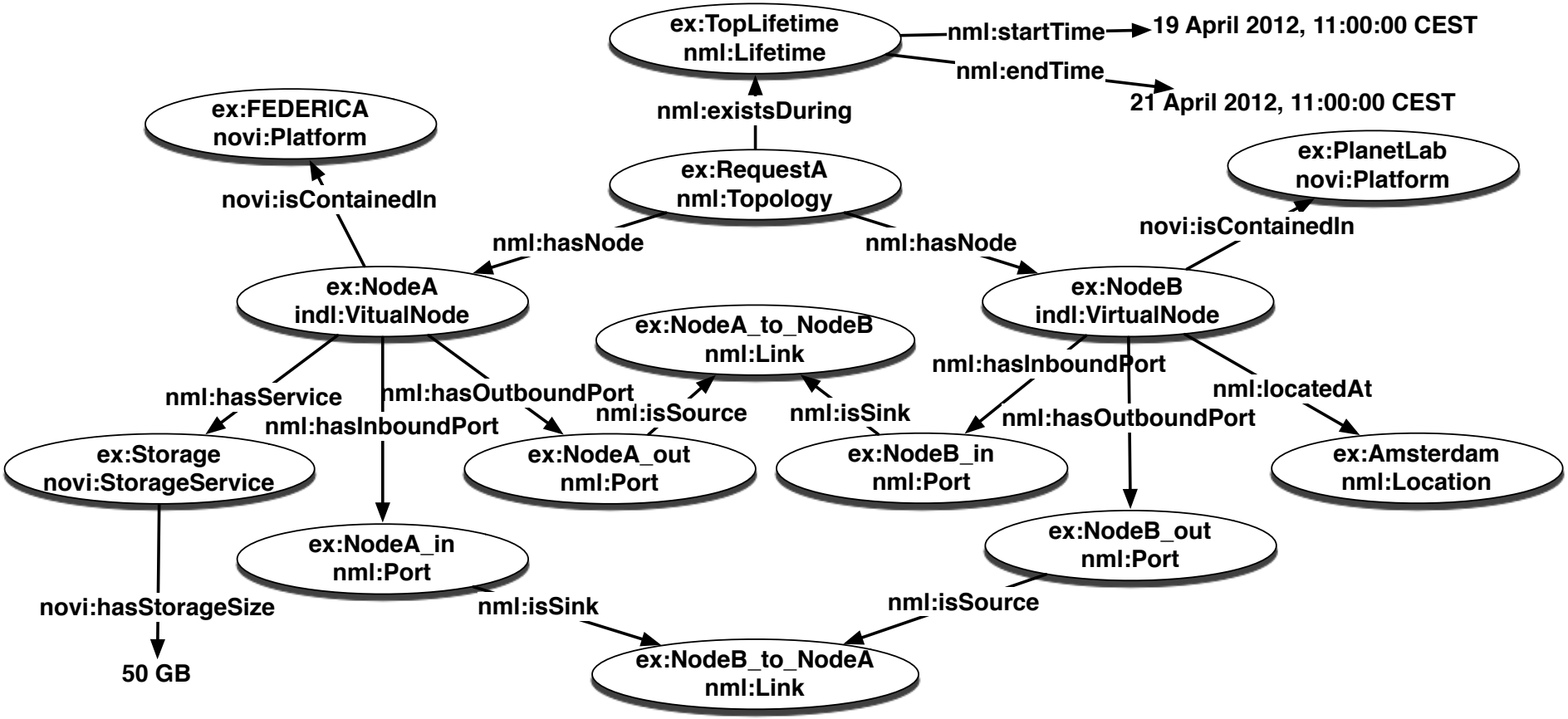
Our model is used for:

- Resource discovery
- Embedding algorithms



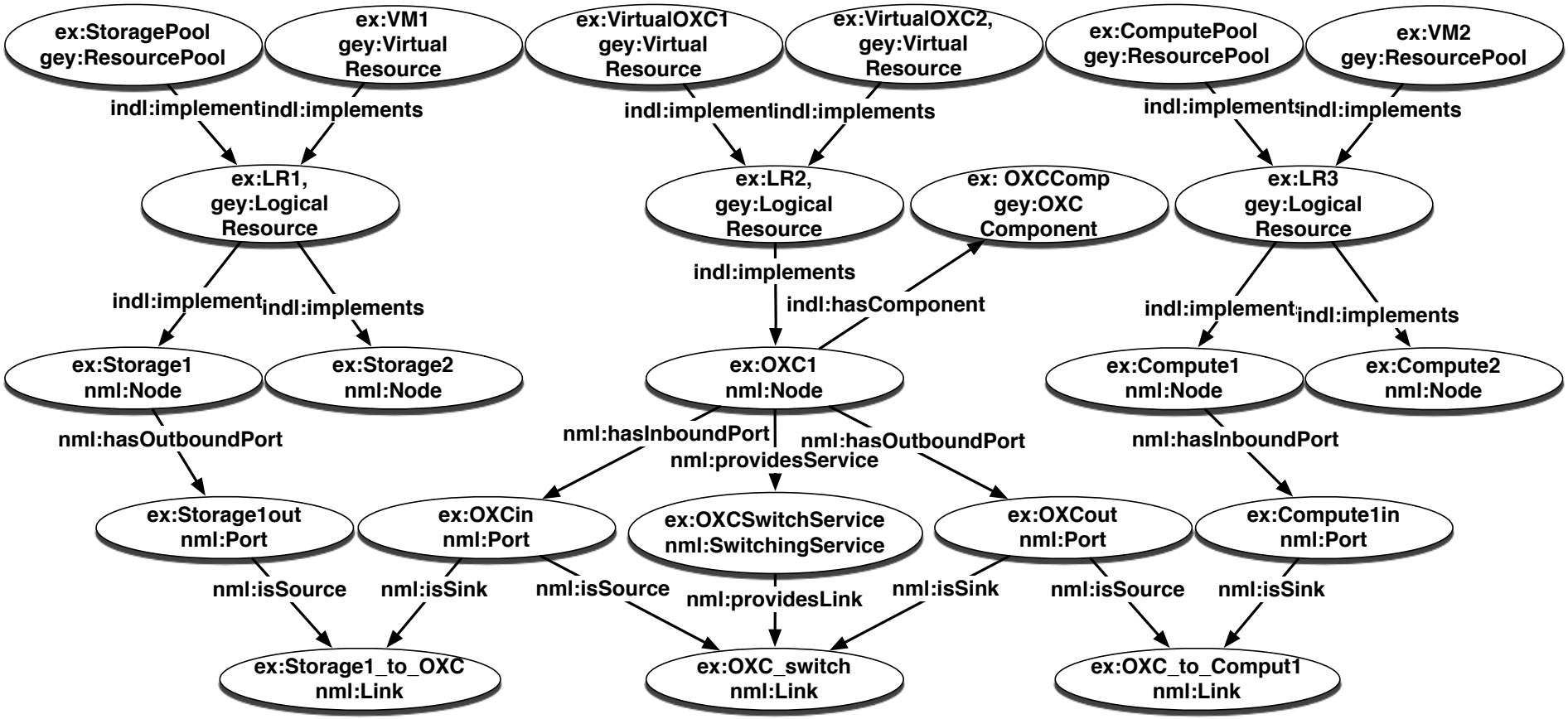
INDL use in NOVI

- Two nodes in the NOVI federation:



INDL in Geysers

- The virtualization model:



NML and NSI

NML - Network Markup Language and NSI – Network Service Interface

within the OGF.

- See: "Network Markup Language Base Schema version 1"

The Network Markup Language purpose is to create a functional description of multi-layer and multi-domain networks.

It can be used for aggregated or abstracted topologies.

Under development: the Network Service Interface Topology Extensions (Draft OGF Standard)


Hey, sit still.
We're almost done. Shhh...

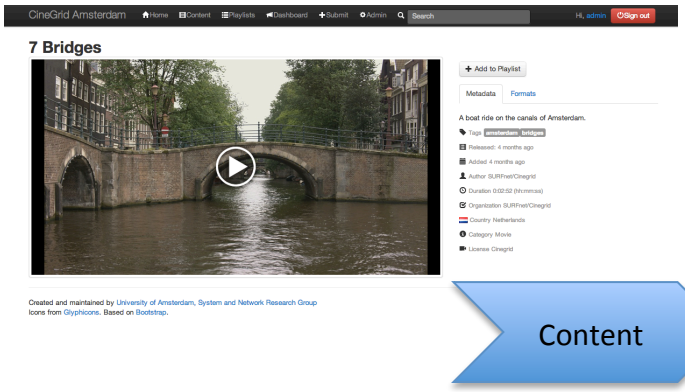
Content Metadata

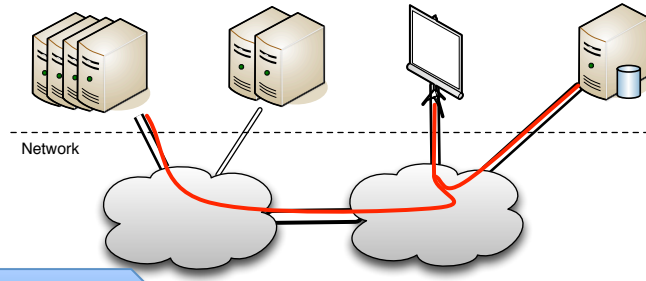
Resource Descriptions

Programmable Infrastructure

CineGrid Portal

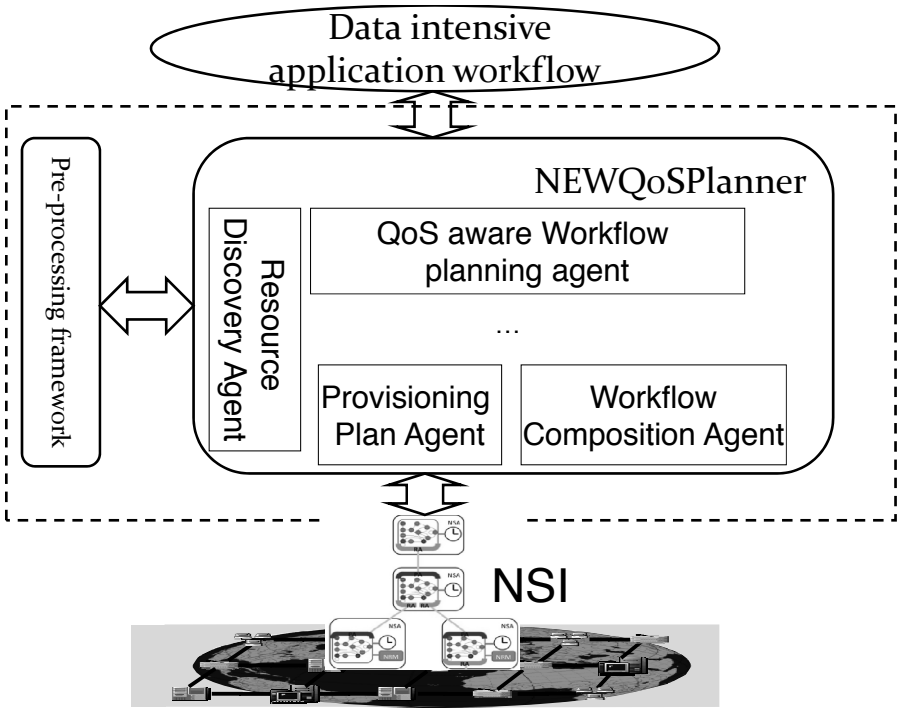
Unified orchestration of distributed CineGrid resources

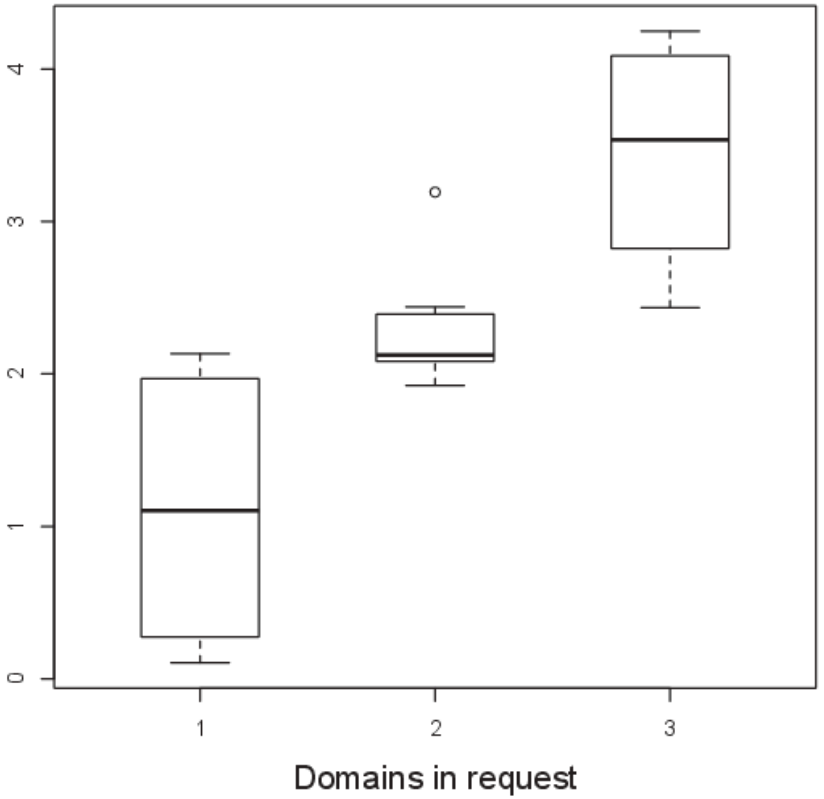




NewQOSPlanner

The NSI – Network Service Interface – creates on the fly connections between domains.

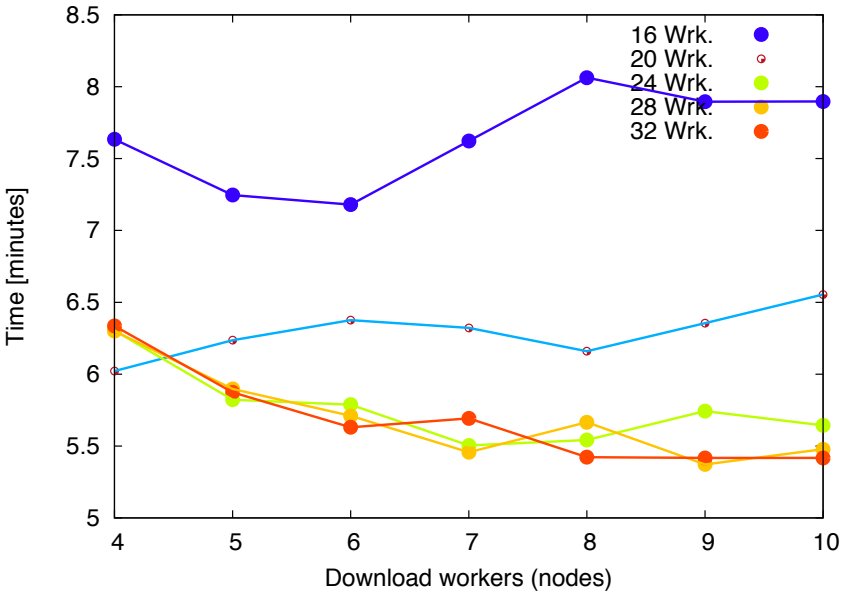


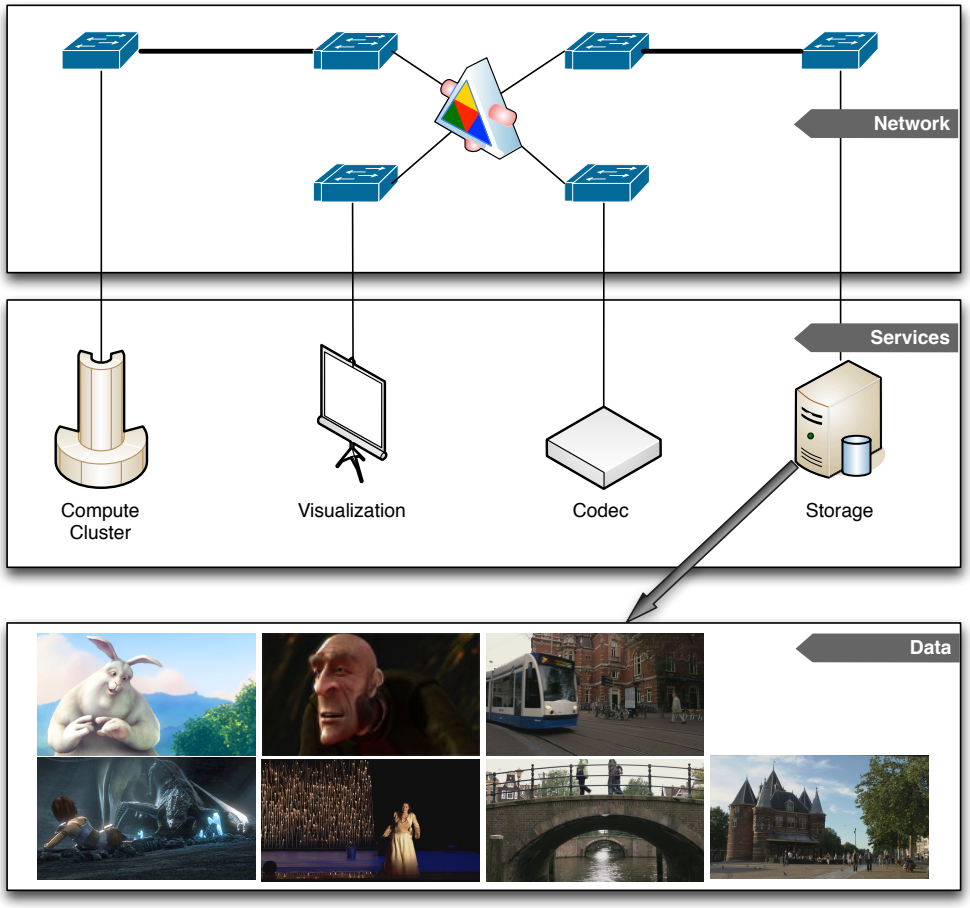


Z. Zhao, J. v/d Ham, A. Taal, R. Koning, P. Grosso and C. de Laat
Planning data intensive workflows on inter-domain resources using the Network Service Interface (NSI) In: WORKS 2012

HyperFlow

Encoding times improve as the end nodes are connected via dynamic lightpaths





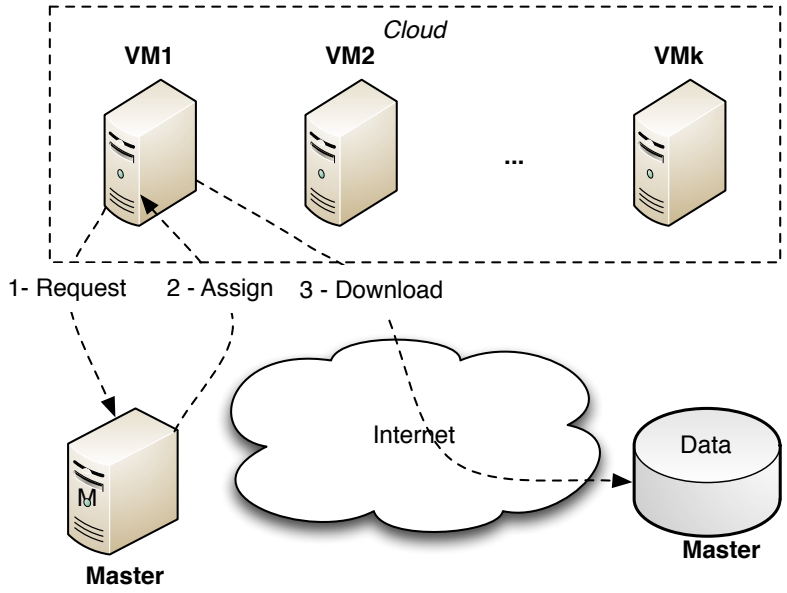
C. Dumitru, Z. Zhao, P. Grosso and C. de Laat
HybridFlow: Towards Intelligent Video Delivery and Processing Over Hybrid Infrastructures
(In CTS 2013))

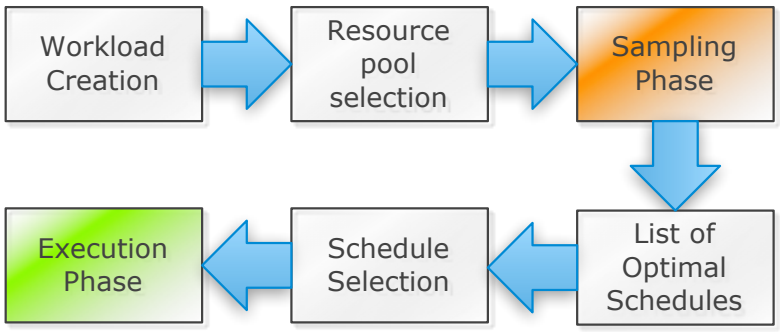
Processing CineGrid with Clouds
A queuing model approach

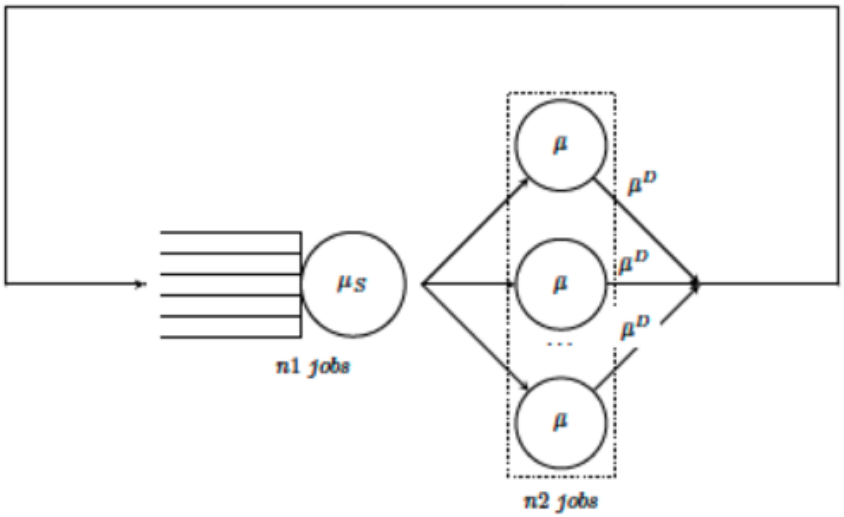
Process large amount of independent data
- Bags-of-Tasks + Data = Bags-of-Data
- Example : Image processing
- Independent files
- Large sizes (10-100s of MBs)

Idea: rent resources
- scaling up (more resources)
- scaling out (more powerful resources)
- Which option ?
- How many ?

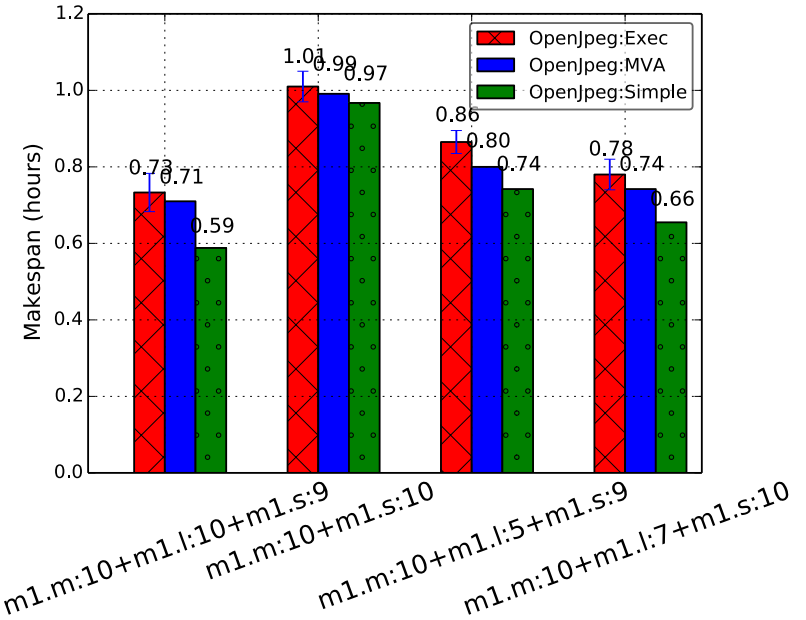
Requirements:
- Within time
- Within budget
- Simple, if possible

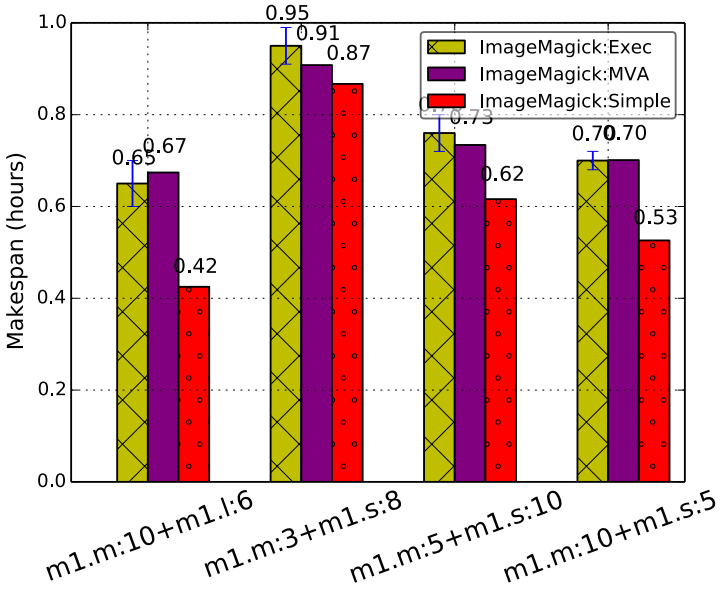


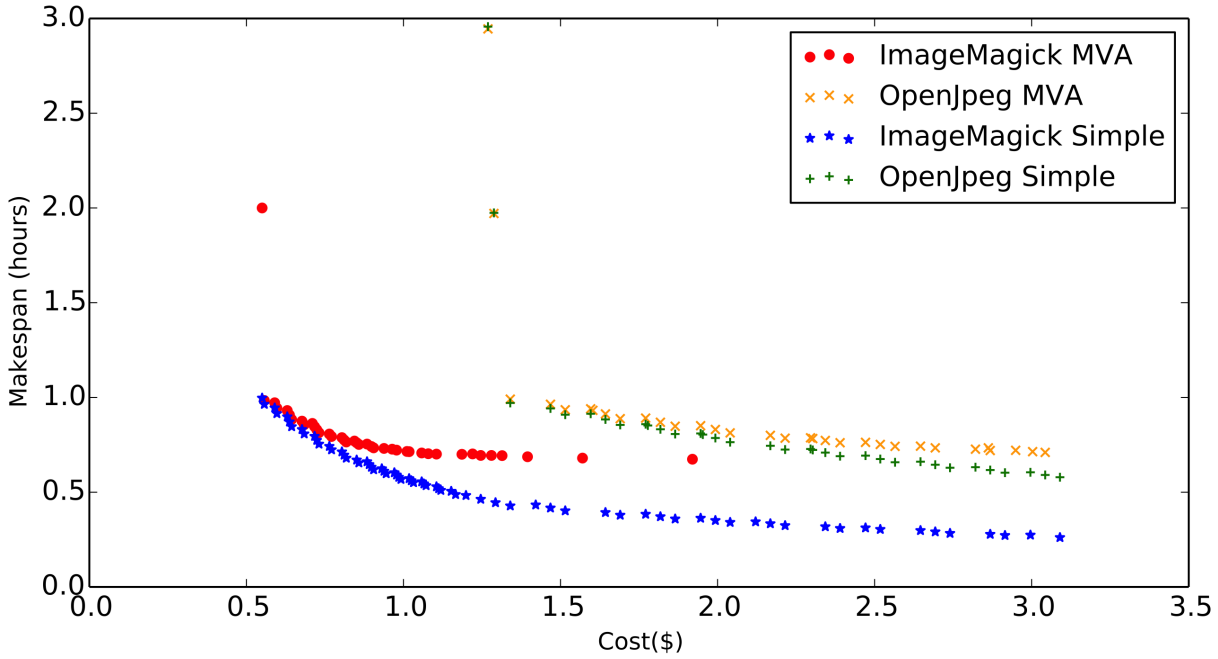




Processing in the Cloud: Mean Value Analysis, Pareto fronts

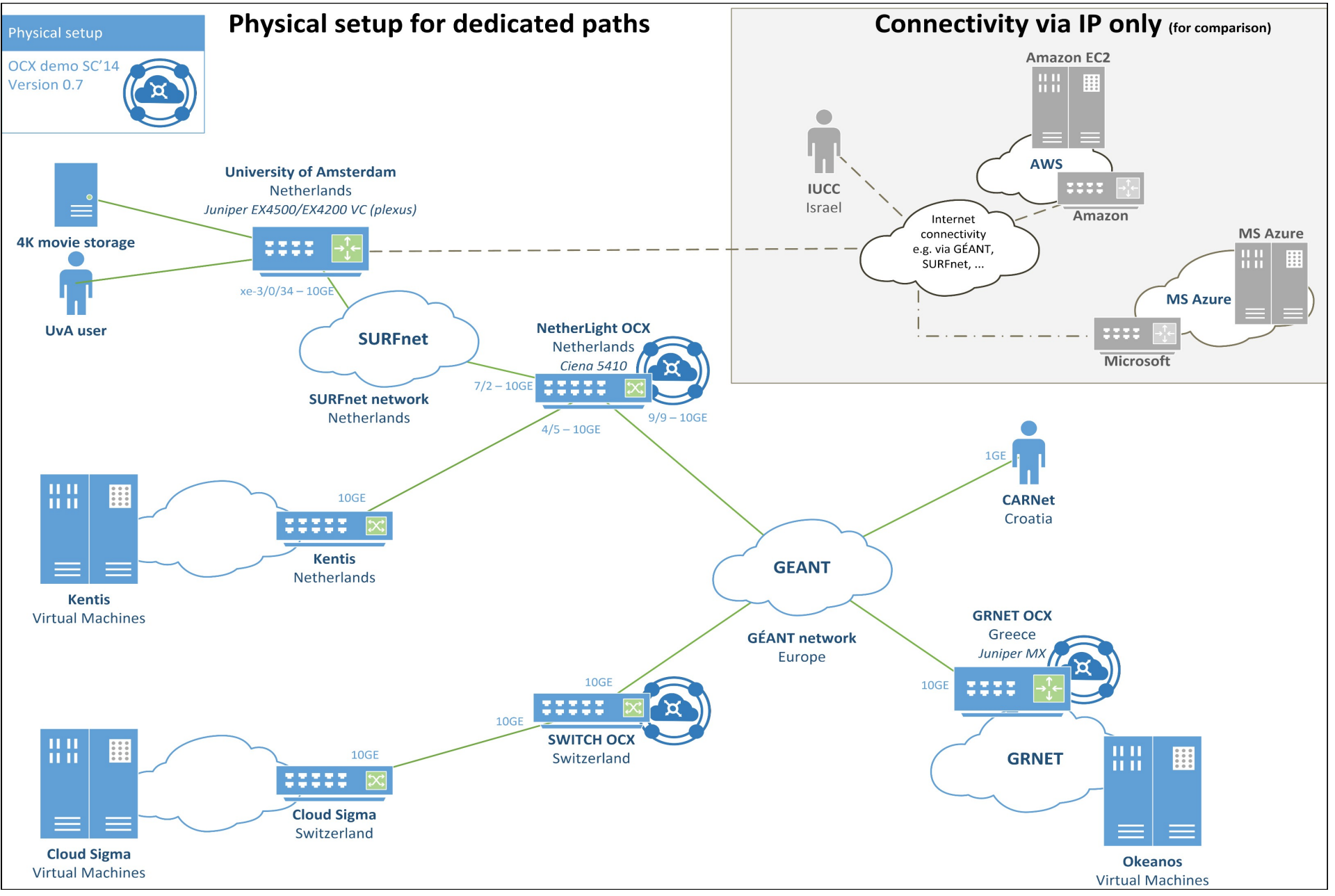






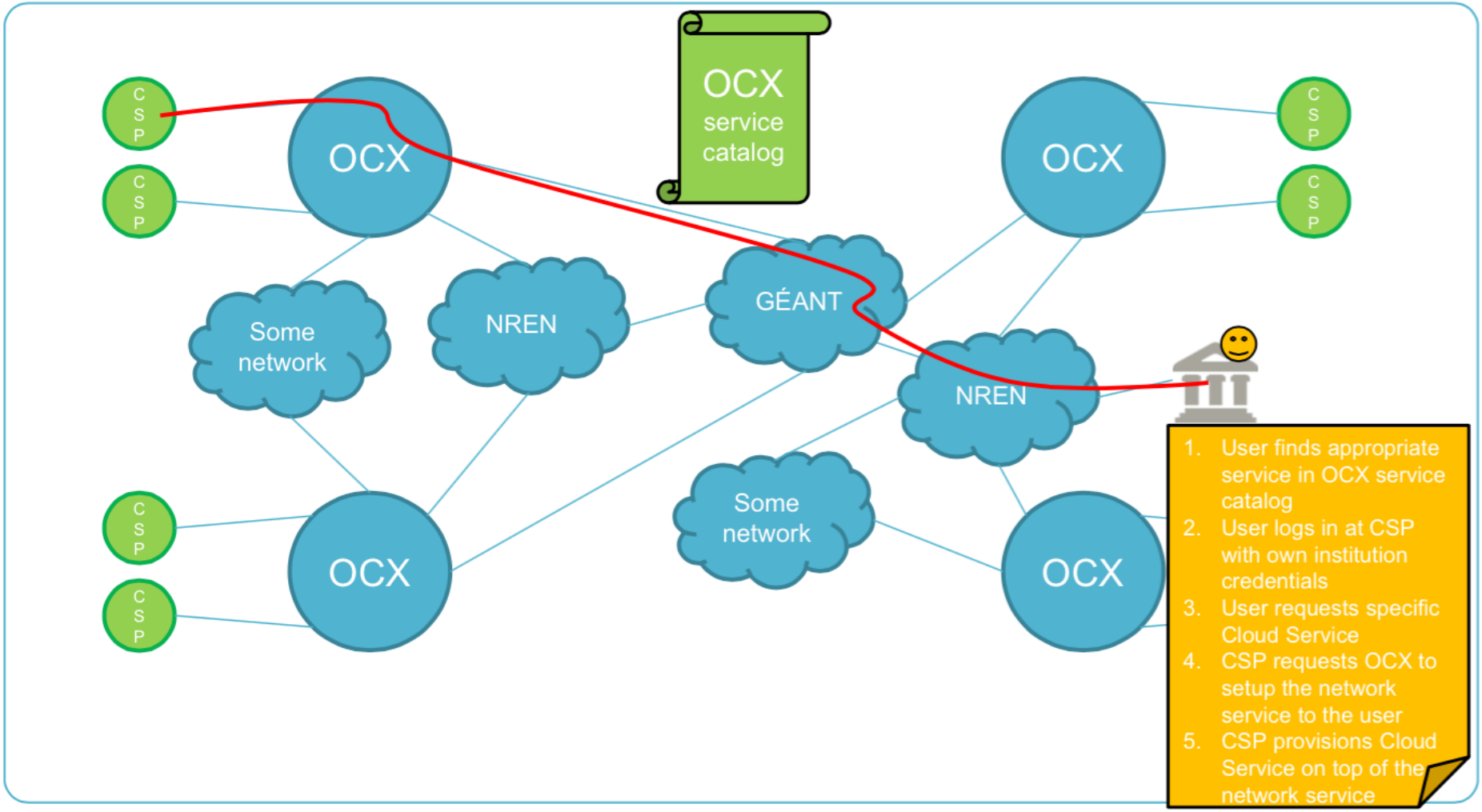
Cosmin Dumitru, Ana-Maria Oprescu, Miroslav Zivkovic, Rob van der Mei, Paola Grosso, Cees de Laat, "A Queueing Theory Approach to Pareto Optimal Bags-of-Tasks Scheduling on Clouds", August 27 2014 , Euro-Par 2014, Porto, Portugal

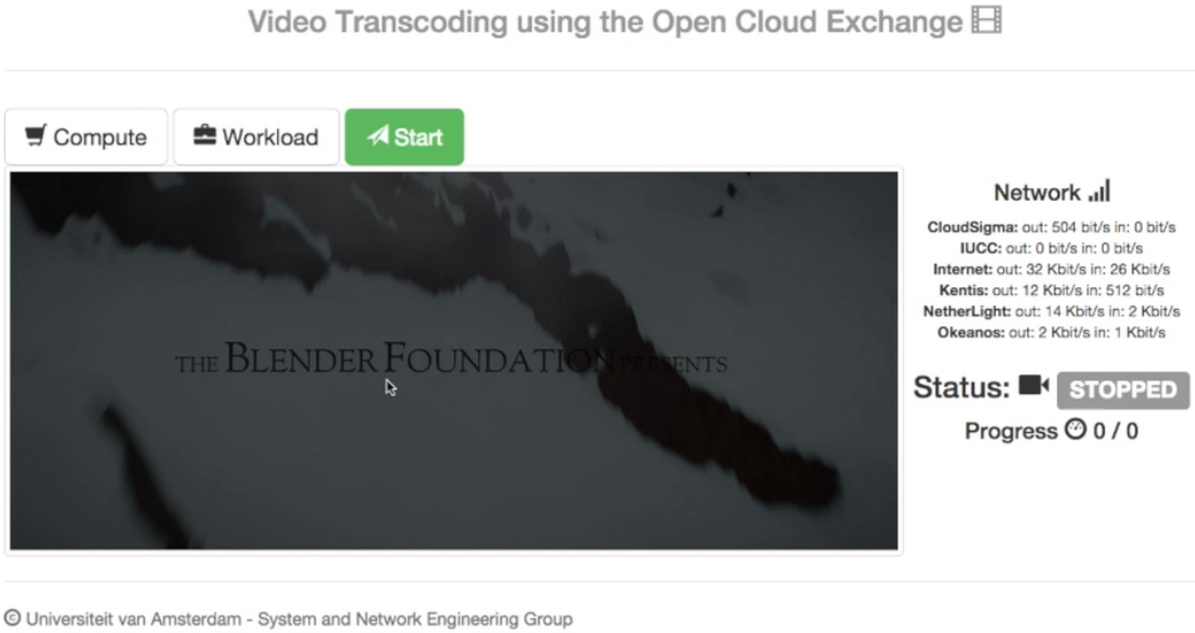
Demo @ SC14


Physical setup
OCX demo SC'14
Version 0.7
Physical setup for dedicated paths
Connectivity via IP only (for comparison)
Amazon EC2
AWS
Amazon
IUCC
Israel
Internet connectivity e.g. via GÉANT, SURFnet, ...
MS Azure
MS Azure
Microsoft
University of Amsterdam
Netherlands
Juniper EX4500/EX4200 VC (plexus)
4K movie storage
UvA user
xe-3/0/34 – 10GE
SURFnet
SURFnet network
Netherlands
NetherLight OCX
Netherlands
Ciena 5410
7/2 – 10GE
9/9 – 10GE
4/5 – 10GE
1GE
CARNet
Croatia
10GE
Kentis
Netherlands
Kentis
Virtual Machines
GEANT
GÉANT network
Europe
GRNET OCX
Greece
Juniper MX
10GE
10GE
10GE
SWITCH OCX
Switzerland
GRNET
10GE
Cloud Sigma
Switzerland
Cloud Sigma
Virtual Machines
Okeanos
Virtual Machines

Open Cloud eXange

OCX



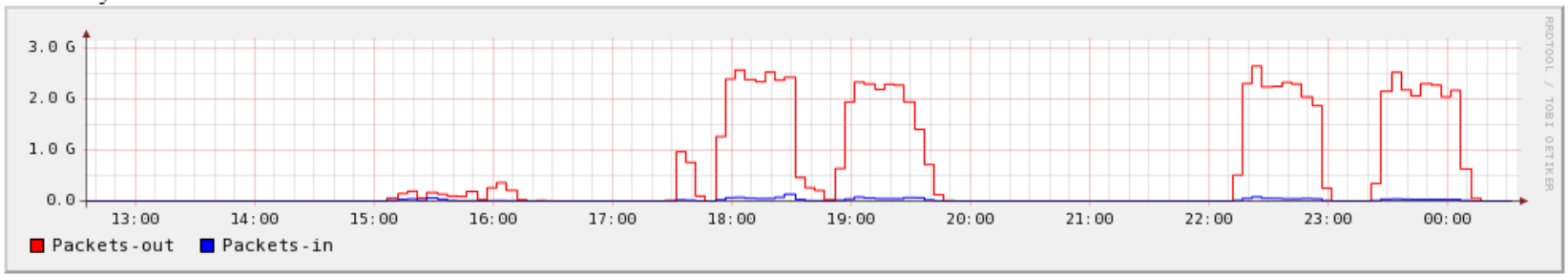


OCX @ SC14

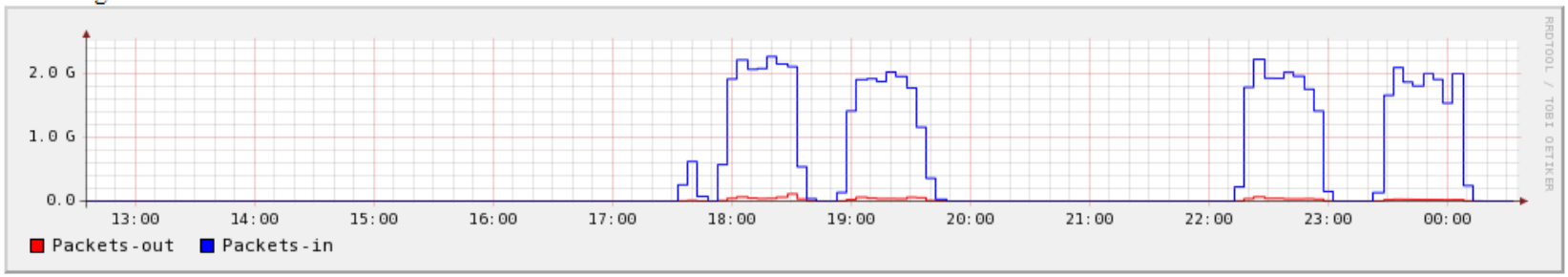
Also: http://sc.delaat.net/sc14/demo-ocx.html

SC14

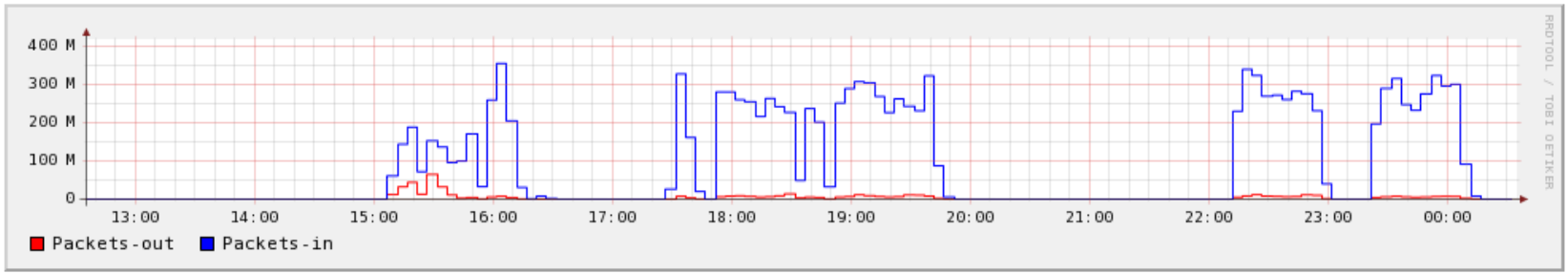
University of Amsterdam



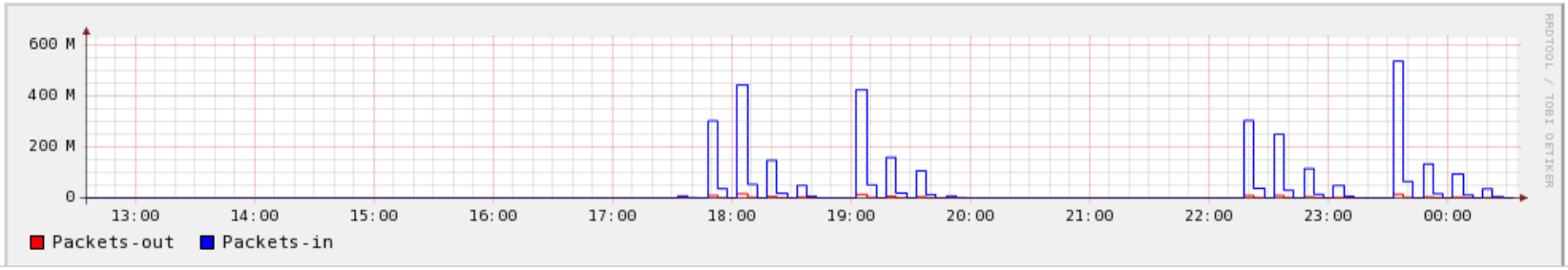
Cloud Sigma



Kentis



Okeanos



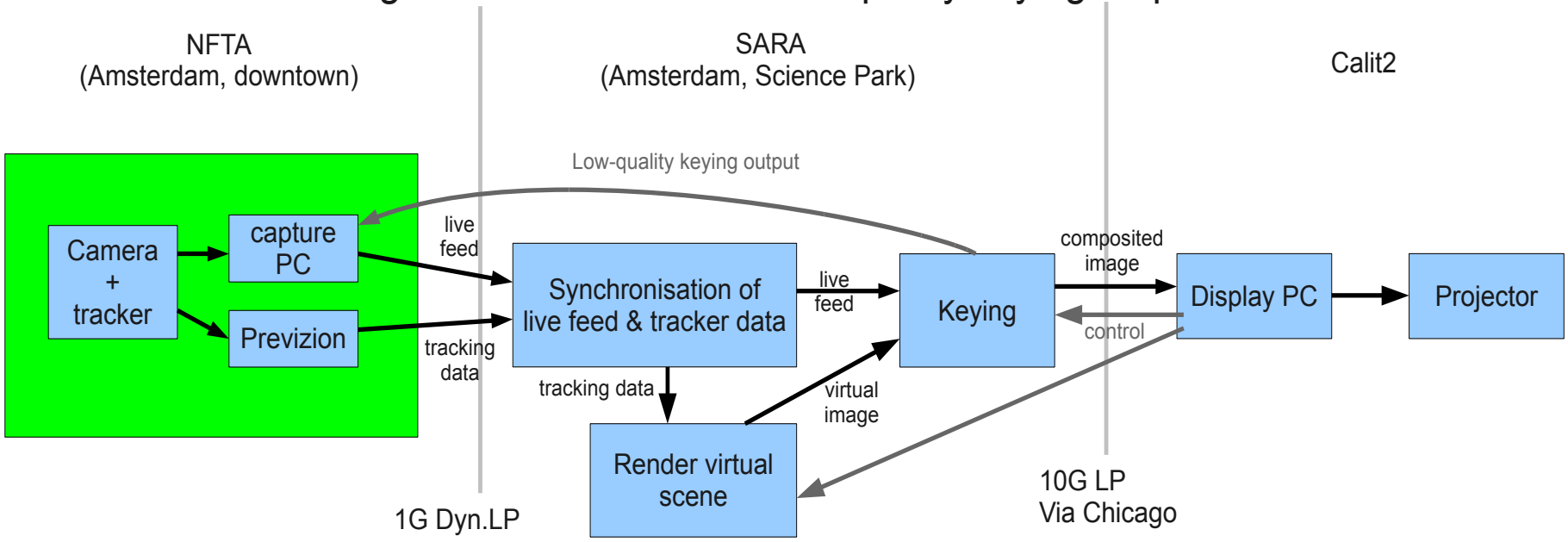
Directing Remote Live Shoot of Virtual Set Acting with Live Compositing in the Cloud

Live action camera, actors, green screen at NFTA (Amsterdam #1)
Virtual set compositing at SARA (Amsterdam #2)
Remote viewing and direction at UCSD/Calit2 Vroom (San Diego)

Real Time Rendering Workflow

Demo setup

- Three locations
  1) NFTA: greenscreen studio, Previzion, camera(+man), actress (+ dress)
  2) SARA: render node for keying, virtual scene rendering
  3) Calit2: keying controls, projection of final output, director
- Two lightpaths in between
- Video-conferencing for communication + low quality keying output back to NFTA



Movie Making on the GLIF

Direction

- Distributed Comp -> Grid -> Cloud -> Big Data
- Lego Block approach
- Application as a Service
- Elastic Cloud
- Determinism & Real Time?
- CineGrid ToolBox
- Storage
- Deep Storage
- Very Deep Storage

Scientific Publications: FGCS Special Issue on CineGrid!

Volume 27, Issue 7,june 2011

Guest Editors: Naohisa Ohta & Paul Hearty & Cees de Laat

Editorial: CineGrid: Super high definition media over optical networks.

1. Real-time long-distance transfer of uncompressed 4K video for remote collaboration.
2. Media Network (HPDMnet): An advanced international research initiative and global experimental testbed.
3. Producing and streaming high resolution digital movies of microscopic subjects.
4. Enabling multi-user interaction in large high-resolution distributed environments.
5. Tri-continental premiere of 4K feature movie via network streaming at FILE 2009.
6. A collaborative computing model for audio post-production.
7. Design and implementation of live image file feeding to dome theaters.
8. Beyond 4K: 8K 60p live video streaming to multiple sites.
9. Using ontologies for resource description in the CineGrid Exchange.
10. CineGrid Exchange: A workflow-based peta-scale distributed storage platform on a high-speed network.
11. CSTP: A parallel data transfer protocol using cross-stream coding.
12. Multi-point 4K/2K layered video streaming for remote collaboration.

The constant factor in our field is Change!

The 50 years it took Physicists to find one particle, the Higgs, we came from:

Assembler, Fortran, COBOL, VM, RSX11, Unix, c, Pascal, SmallTalk, DECnet, VMS, TCP/IP, c++, Internet, WWW, ATM, Semantic Web, Photonic networks, Google, Grid, Phyton, FaceBook, Twitter, Cloud, SDN, Data^3, App's

to:

DDOS attacks destroying Banks and BitCoins!

Conclusion:

Need for Safe, Smart, Resilient Sustainable Infrastructure.

Questions?

http://delaat.net

http://sne.science.uva.nl

http://www.os3.nl/

http://i4dw.nl/

http://dsrc.nl/

http://sne.science.uva.nl/openlab/

http://pire.opensciencedatacloud.org

http://staff.science.uva.nl/~delaat/pire/

https://rd-alliance.org

http://envri.eu

Data Science Research Center Amsterdam

Supported by: