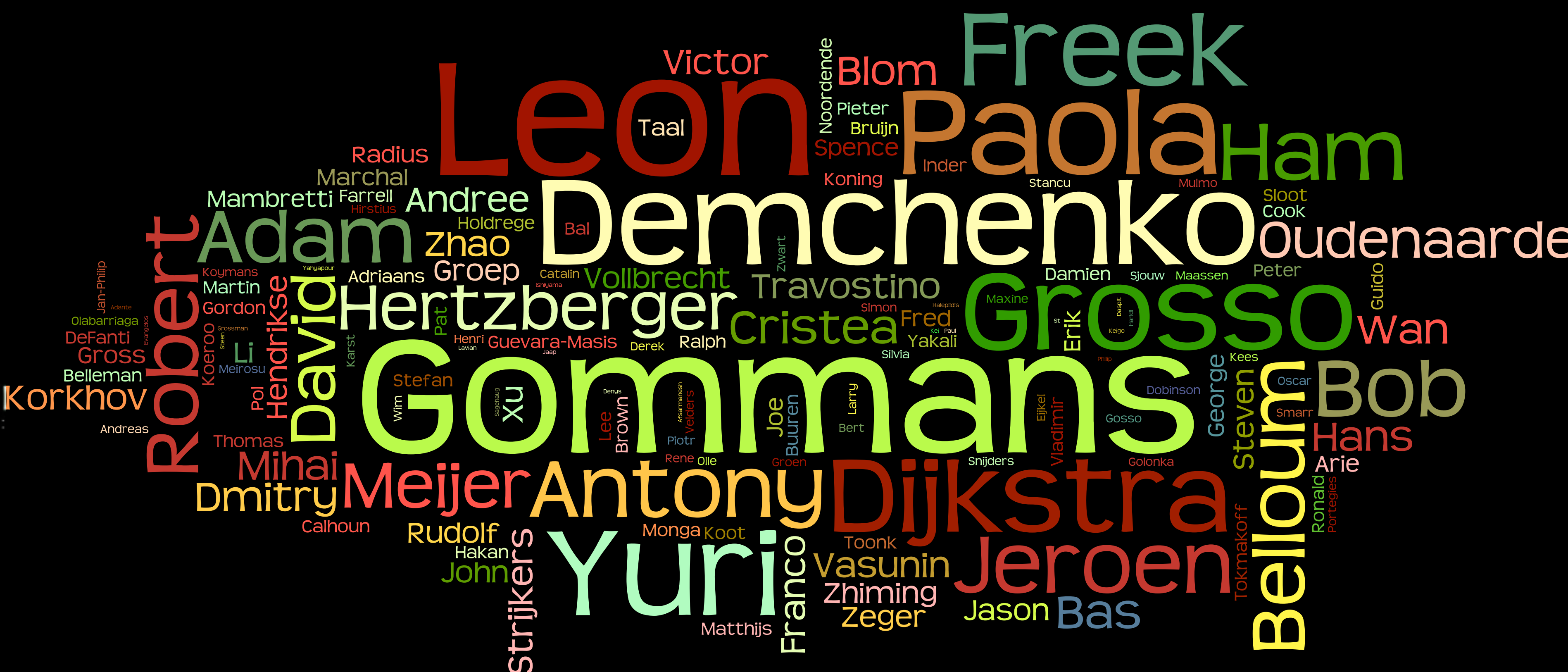
System and Network Engineering Research for Big Data Sciences

Cees de Laat



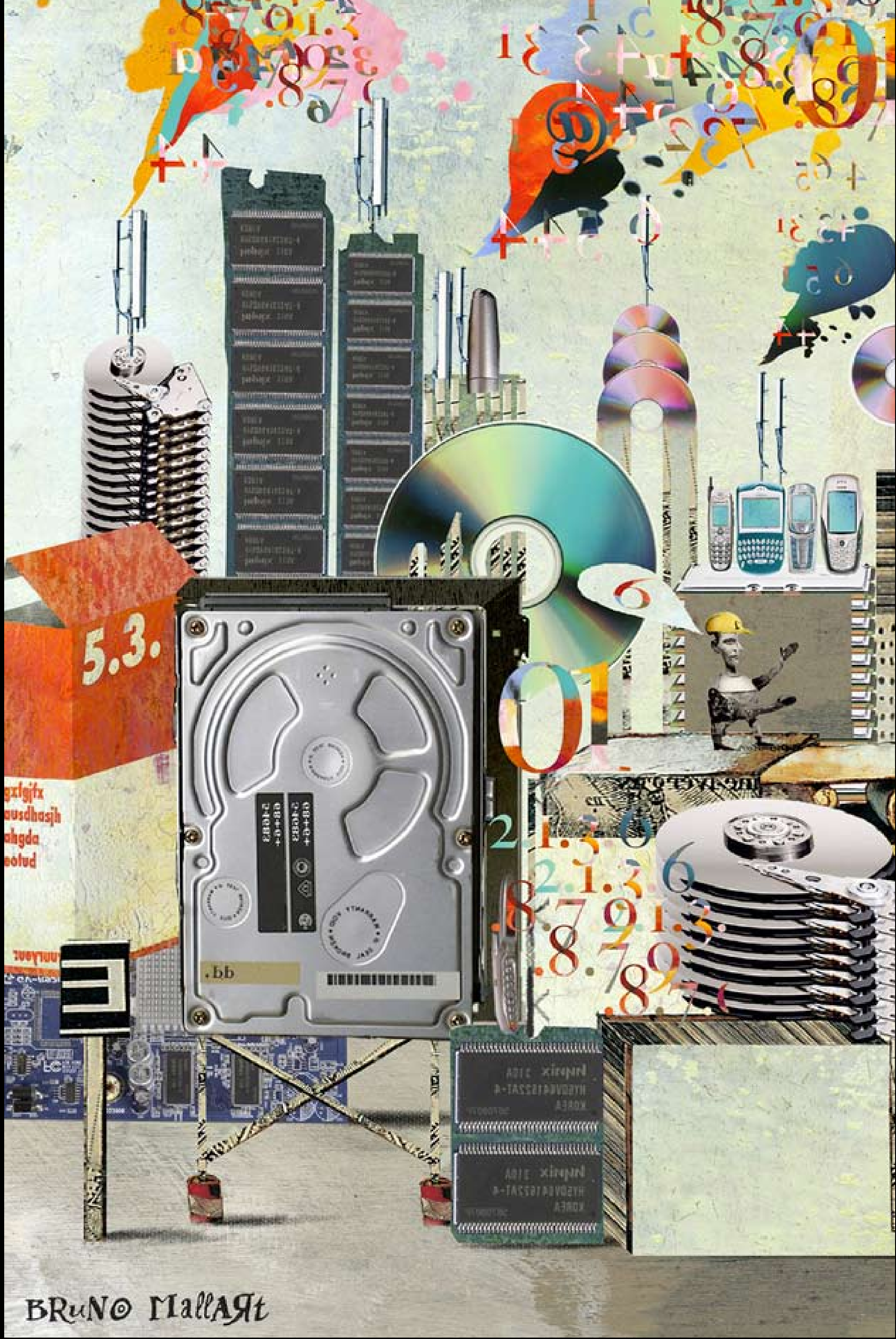
From King's Dutch Academy of Sciences

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?

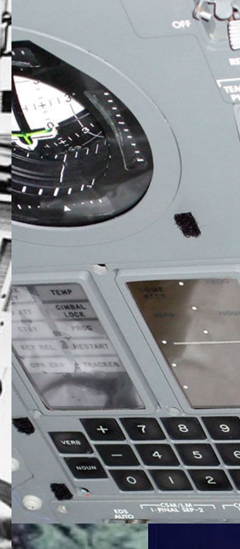
SNE addresses a.o. the **highlighted** questions!

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

Mission SNE

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

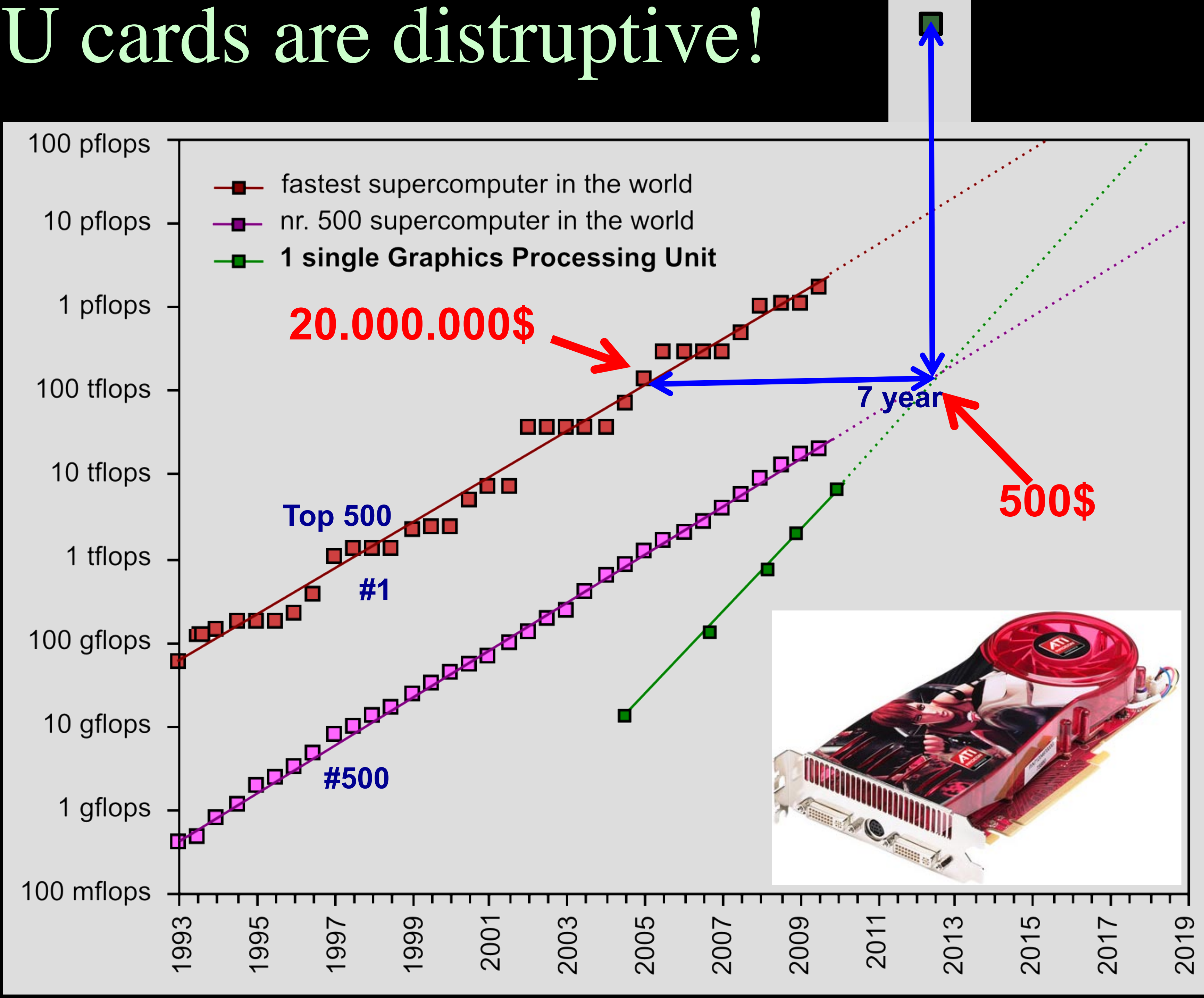
- *Capacity*
  - *Bandwidth on demand, QoS, architectures, photonics, performance, GPU's*
- *Capability*
  - *Programmability, virtualization, complexity, semantics, workflows*
- *Security*
  - *Authorization, Anonymity, integrity of data in distributed data processing*
- *Sustainability*
  - *Greening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


TEMP
GIMBAL LOCK
PROG
RESTART
TRACKER
VERB
NOUN

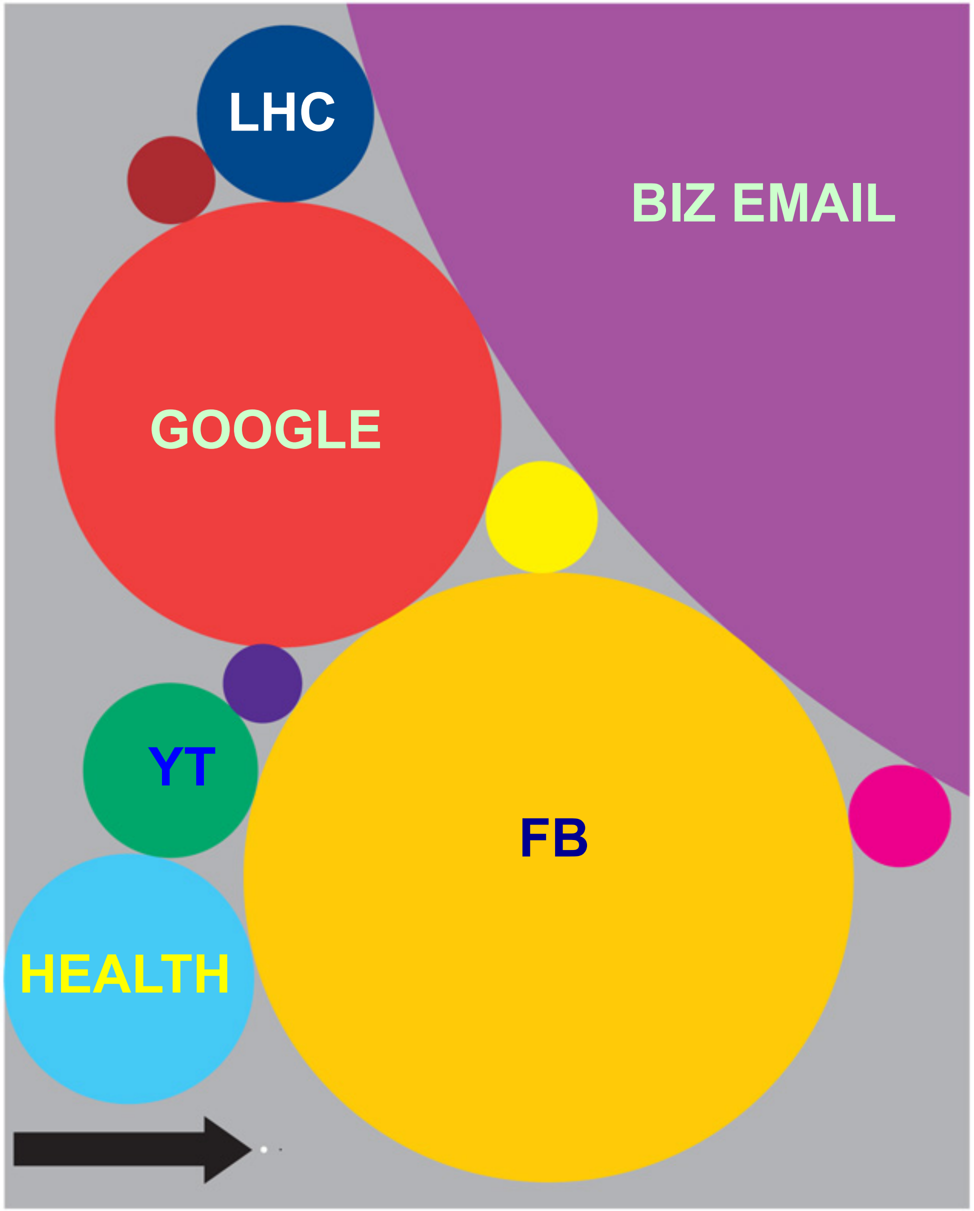

UPLINK ACTY
TEMP
AUTO
GIMBAL LOCK
PROG
RESTART
NO ATT
TRACKER
STBY
KEY REL
OPR ERR
VERB
NOUN
CLR
ENTR
RSET


AT&T
9:42 AM
SMS
Text
Tuesday
9
Calendar
Photos
Camera
YouTube
Stocks
280
Maps
73°
Weather
Clock
Calculator
Notes
Settings
iTunes
App Store
Phone
Mail
Safari
iPod
UPLINK ACTY
TEMP
AUTO
GIMBAL LOCK
PROG
RESTART
NO ATT
TRACKER
STBY
KEY REL
OPR ERR
VERB
NOUN
CLR
ENTR
RSET

GPU cards are distruptive!



What Happens in an **Internet Minute?**

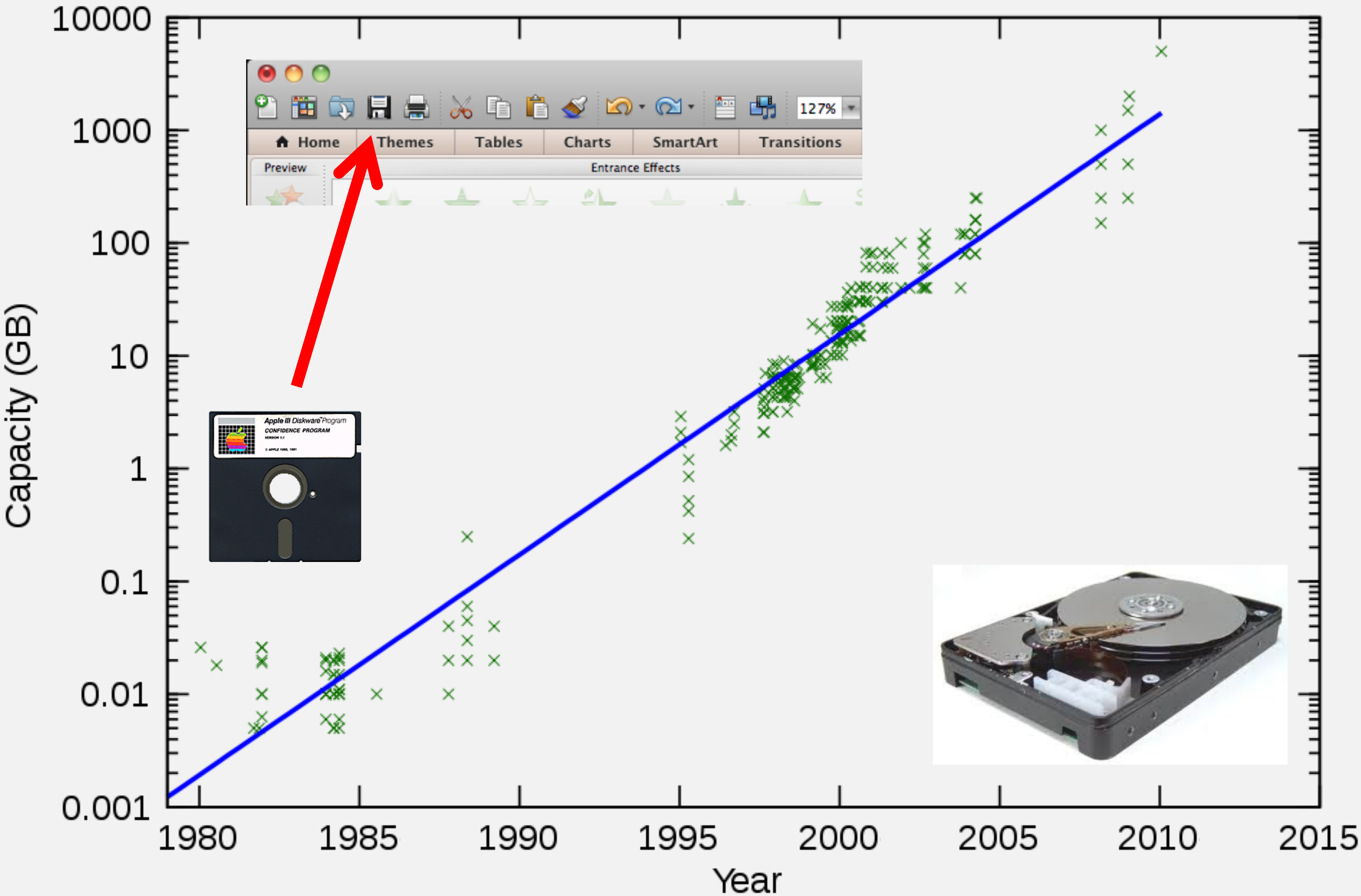


Size of data sets in terabytes

Data set	Terabytes
Business email sent per year	2,986,100
Content uploaded to Facebook each year	182,500
Google's search index	97,656
Kaiser Permanente's digital health records	30,720
Large Hadron Collider's annual data output	15,360
Videos uploaded to YouTube per year	15,000
National Climactic Data Center database	6,144
Library of Congress' digital collection	5,120
US Census Bureau data	3,789
Nasdaq stock market database	3,072
Tweets sent in 2012	19
Contents of every print issue of WIRED	1.26

There is always a bigger fish

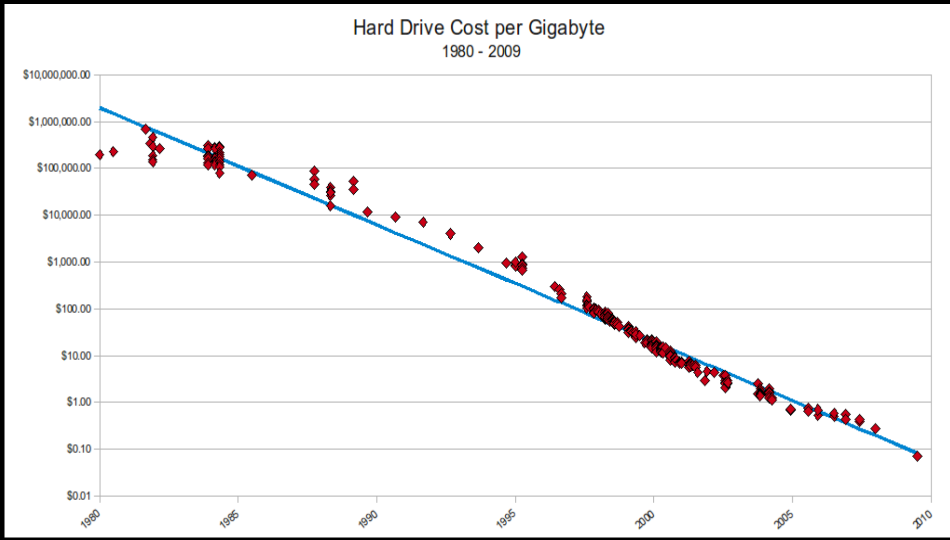
Data storage: doubling every 1.5 year!



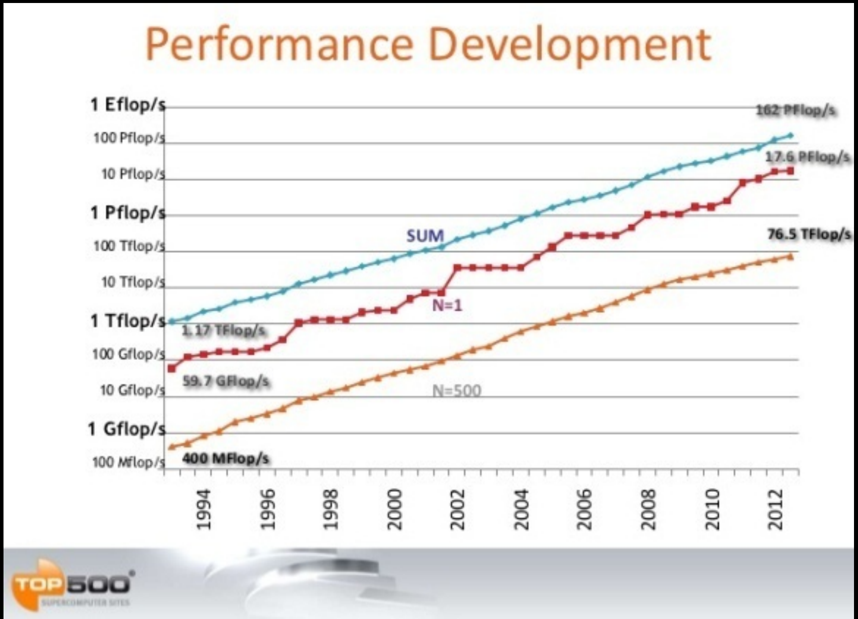
Reliable and Safe!

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.



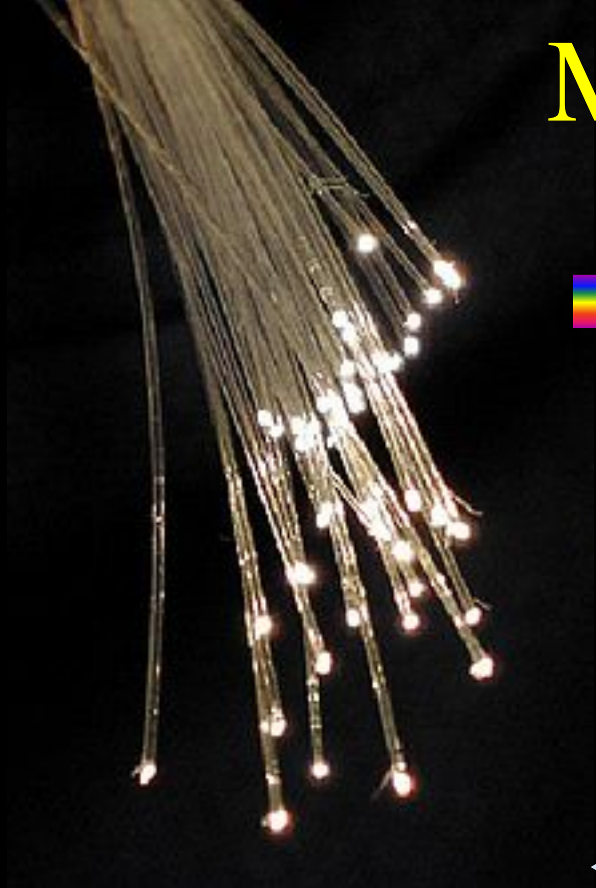
The hardware and software that allow all our systems to operate is becoming bigger and more complex all the time, and the capacity of networks and data storage is increasing by leaps and bounds.

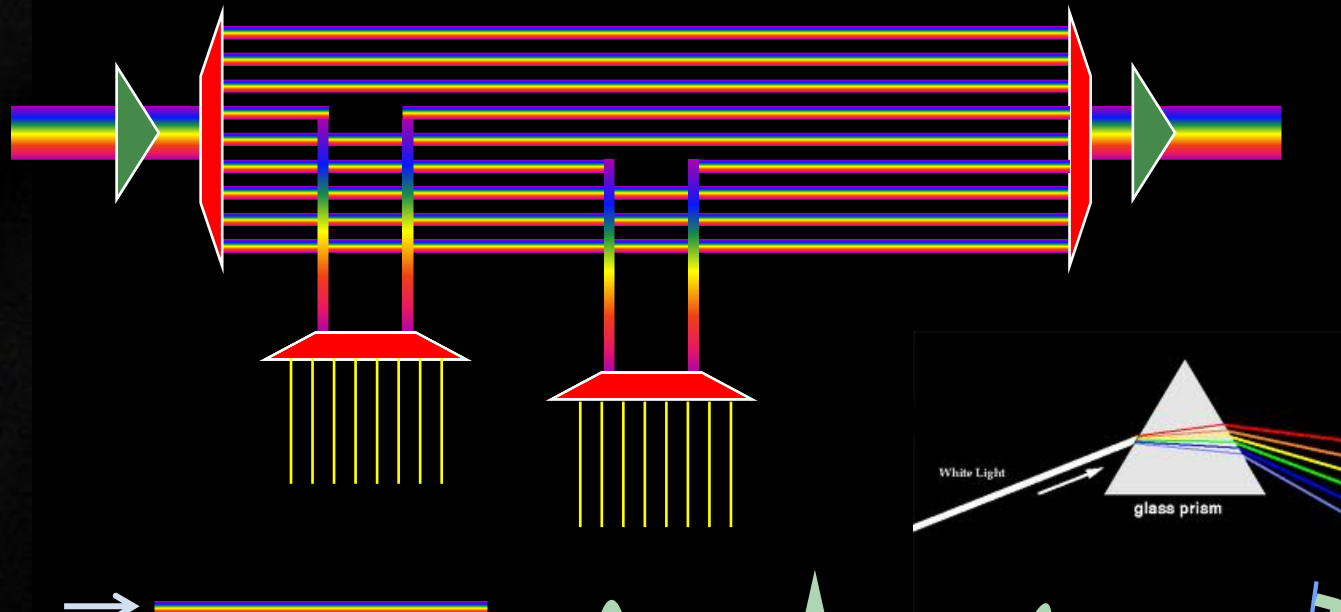


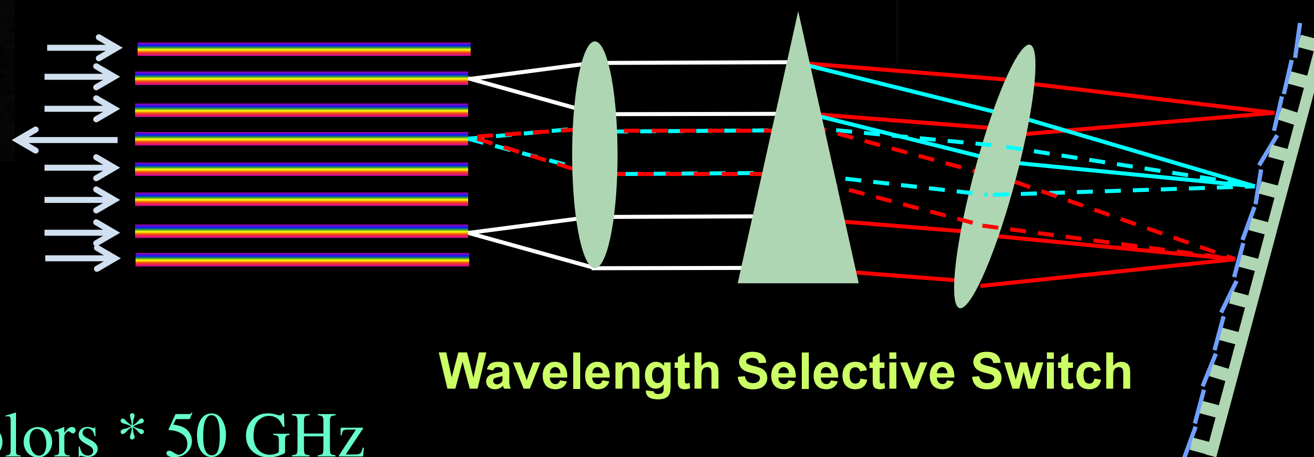
We will soon reach the limits of what is currently feasible and controllable.

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

Multiple colors / Fiber







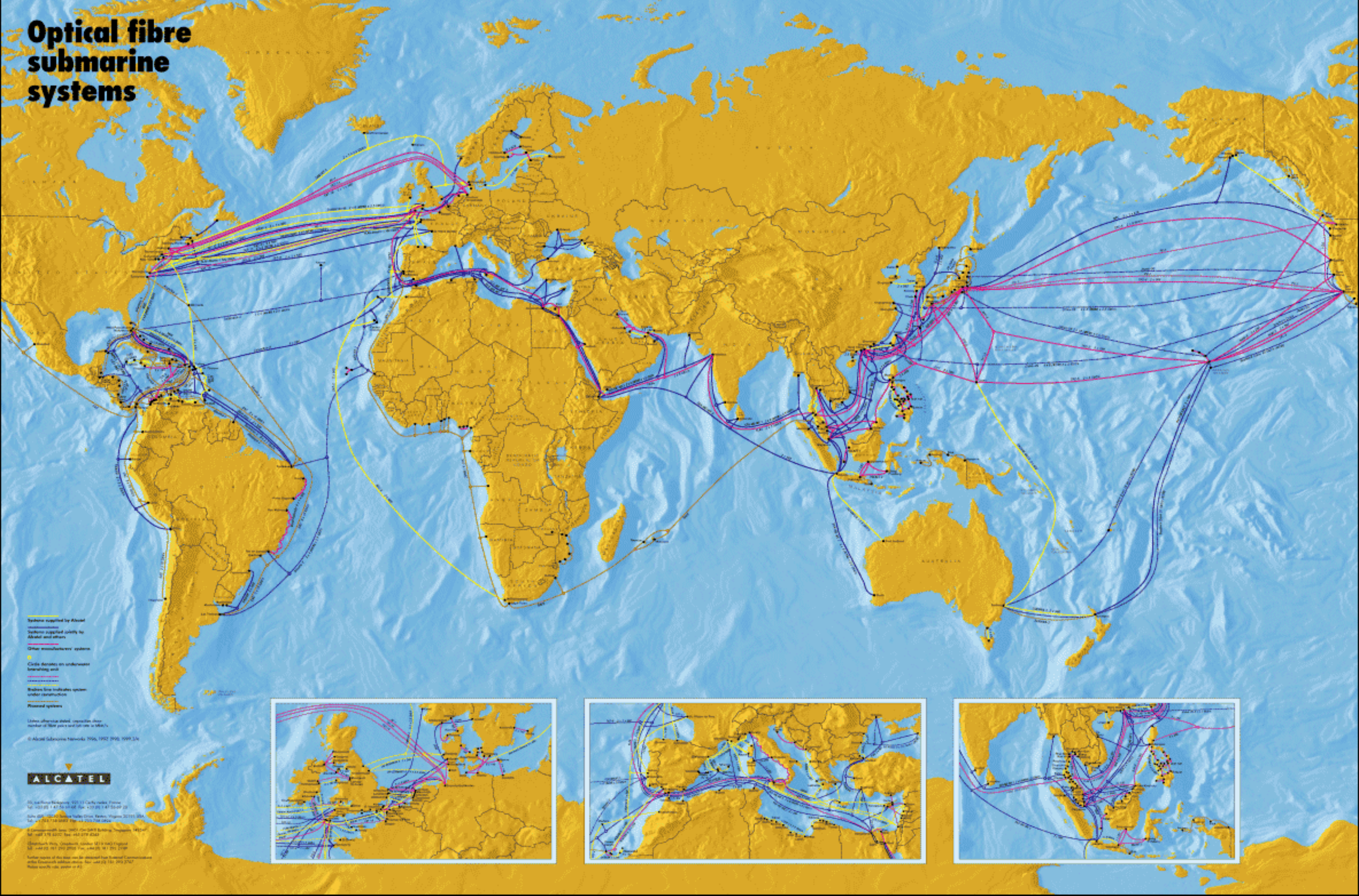
Per fiber: ~ 80-100 colors * 50 GHz

Per color: 10 – 40 – 100 – 200 ... Gbit/s

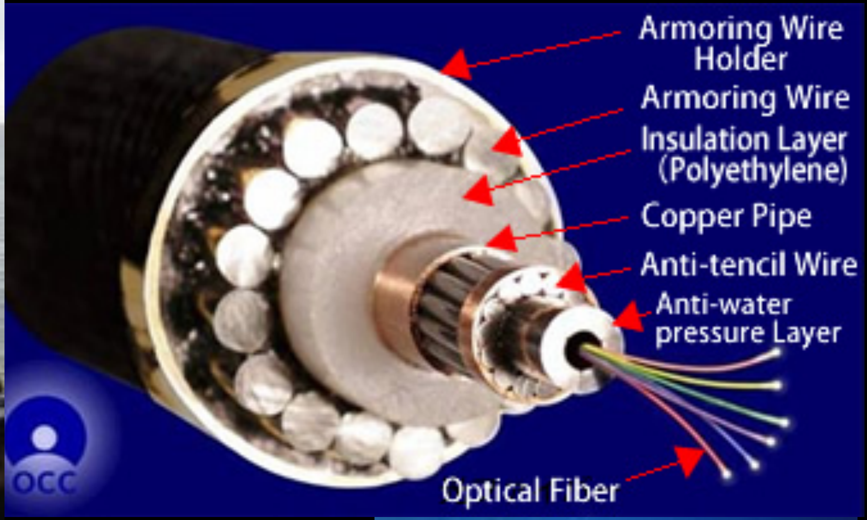
BW * Distance ~ $2*10^{17}$ bm/s

New: Hollow Fiber!

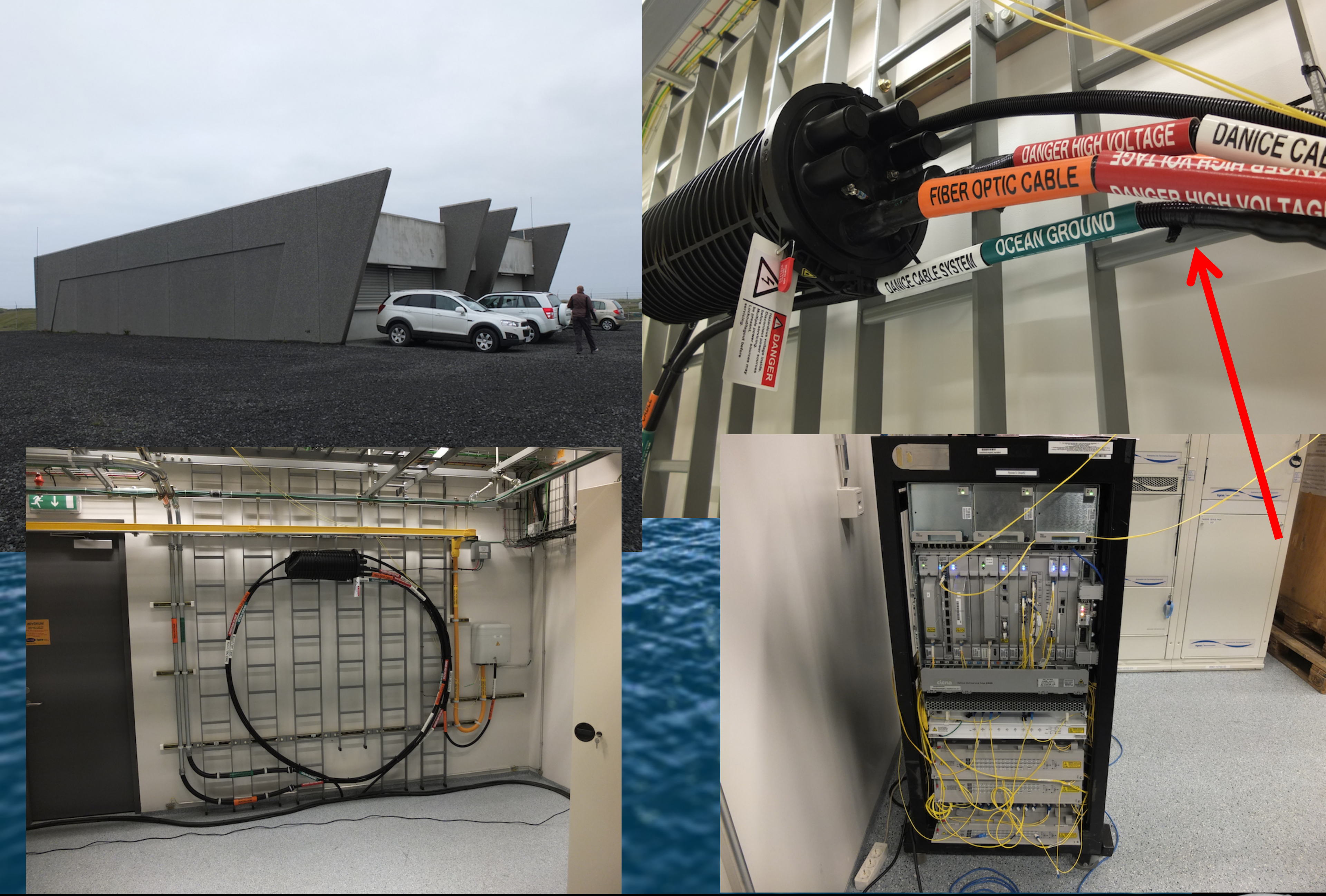
➔ less RTT!



Undersea Cable System

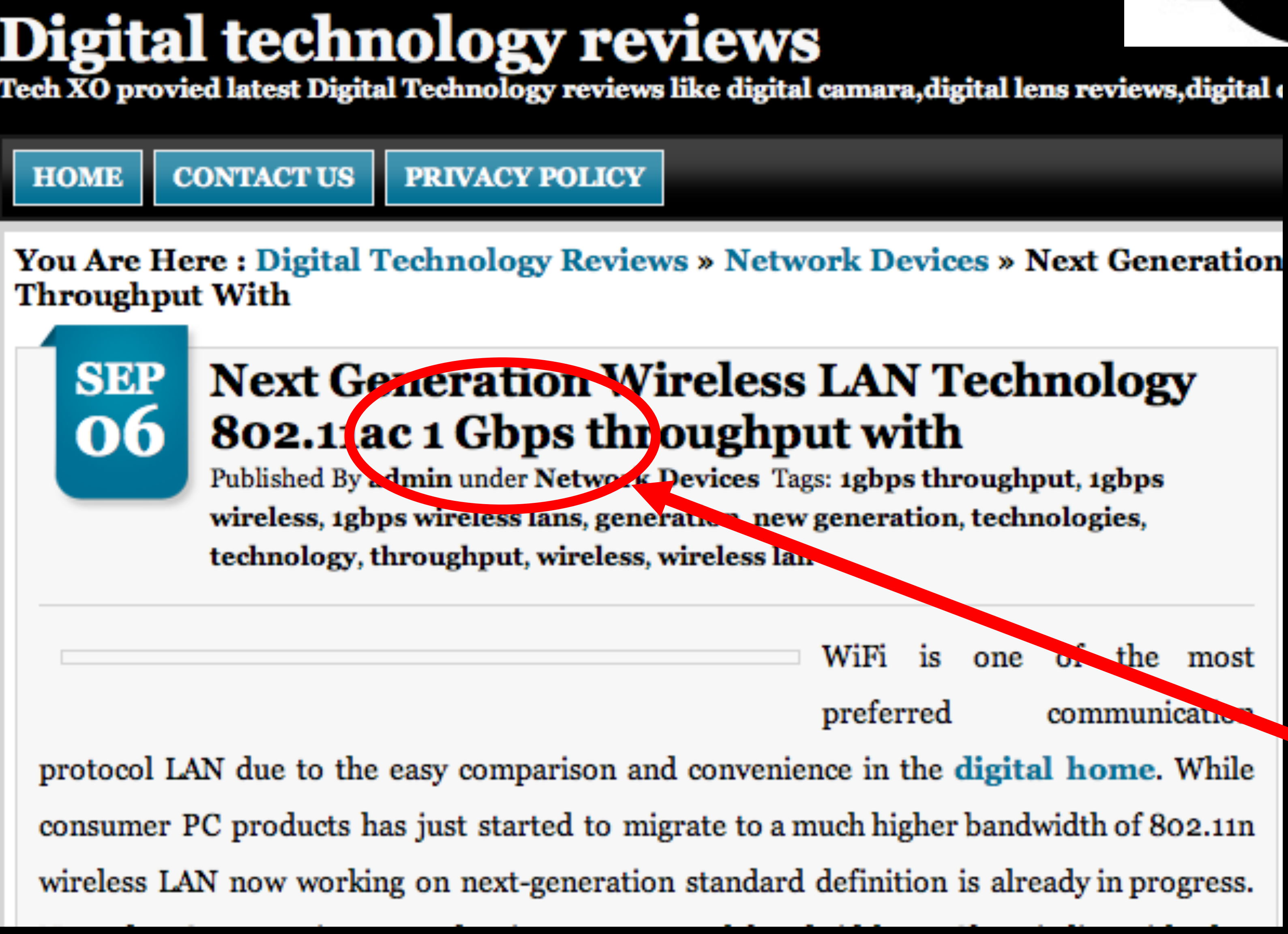


A **cable landing station** may or may not be required, depending on whether, for example, the submarine cable requires power to power submarine repeaters or amplifiers. The voltages applied to the cables can be high **3,000 to 4,000 volts** for a typical trans-Atlantic telecommunications cable system, and 1,000 volts for a cross-channel telecommunications cable system. Submarine power cables can operate at many kilovolts: for example, the[Fenno-Skan power cable operates at 400 kV DC.]



Undersea Cable HV

Wireless Networks



Digital technology reviews

Tech XO provied latest Digital Technology reviews like digital camara,digital lens reviews,digital

HOME | CONTACT US | PRIVACY POLICY

You Are Here : Digital Technology Reviews » Network Devices » Next Generation Throughput With

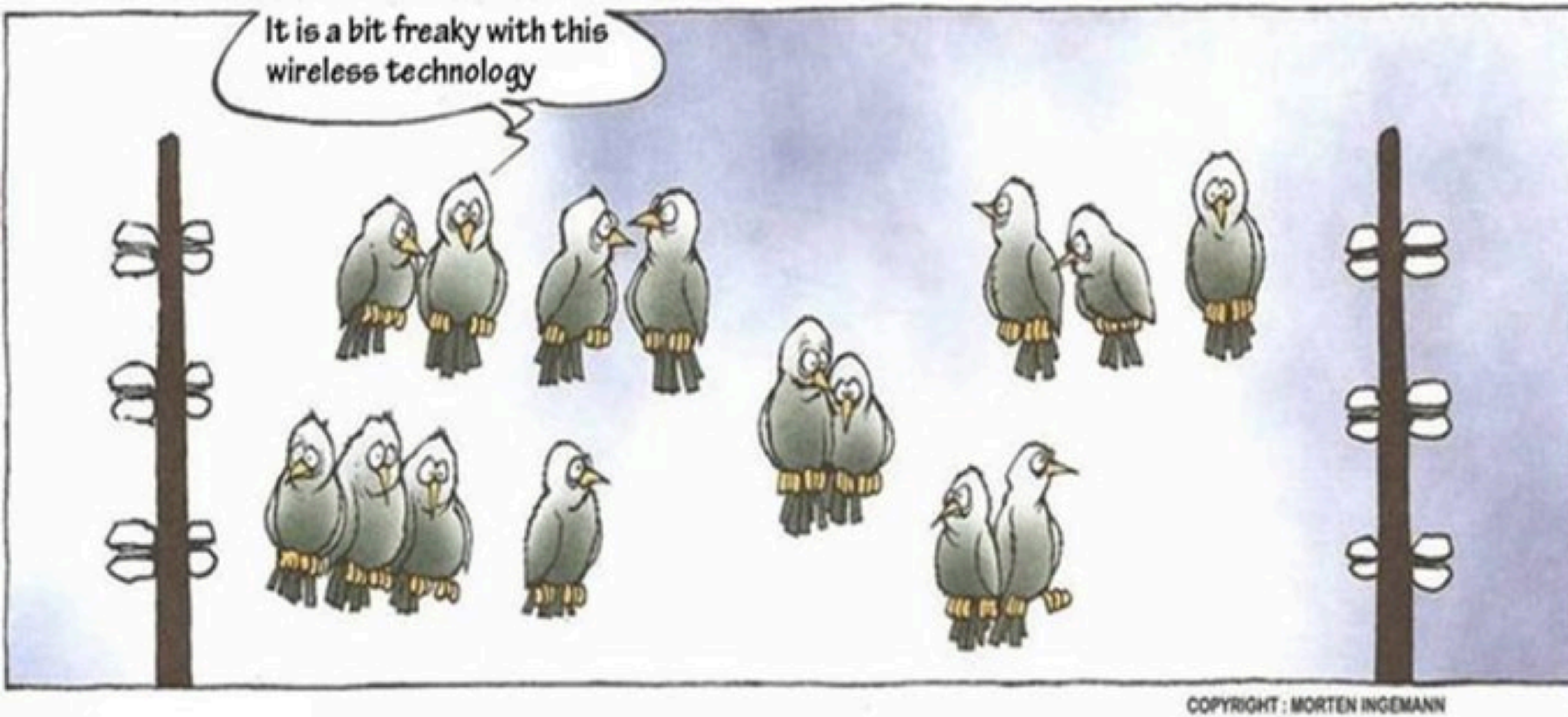
SEP 06

Next Generation Wireless LAN Technology 802.11ac 1 Gbps throughput with

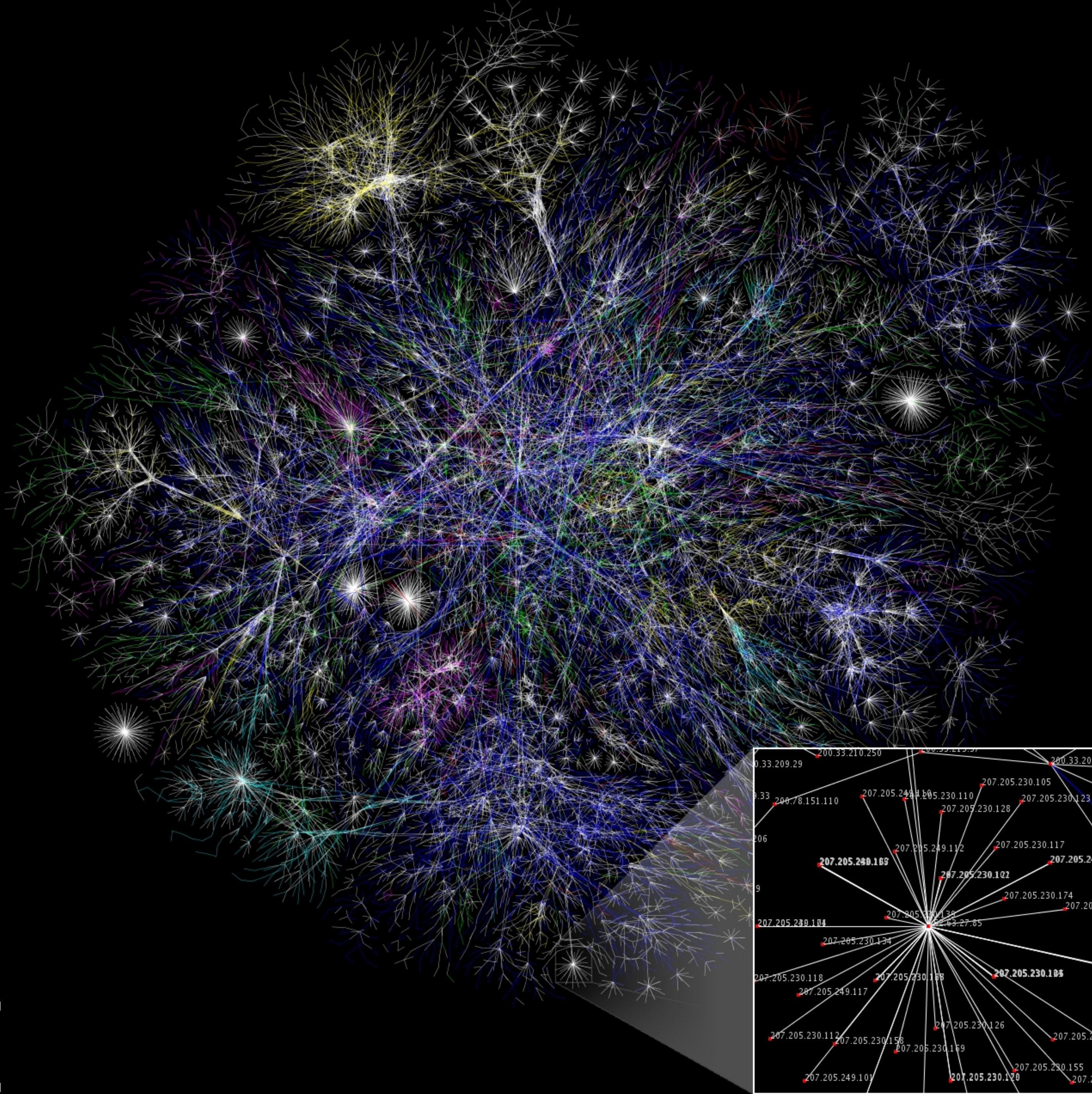
Published By admin under Network Devices Tags: 1gbps throughput, 1gbps wireless, 1gbps wireless lans, generation, new generation, technologies, technology, throughput, wireless, wireless lan

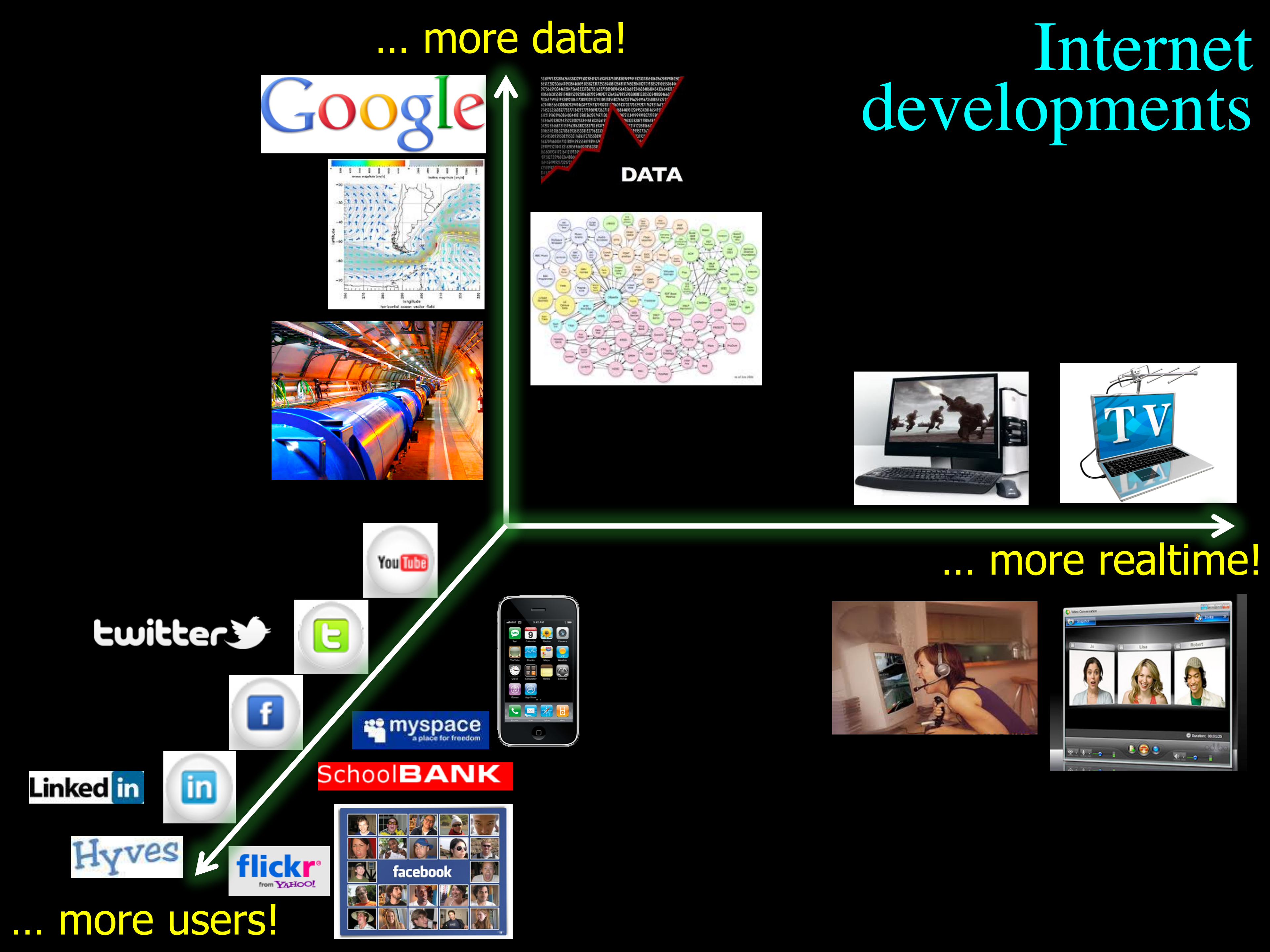
WiFi is one of the most preferred communication protocol LAN due to the easy comparison and convenience in the digital home. While consumer PC products has just started to migrate to a much higher bandwidth of 802.11n wireless LAN now working on next-generation standard definition is already in progress.

Wireless Networks

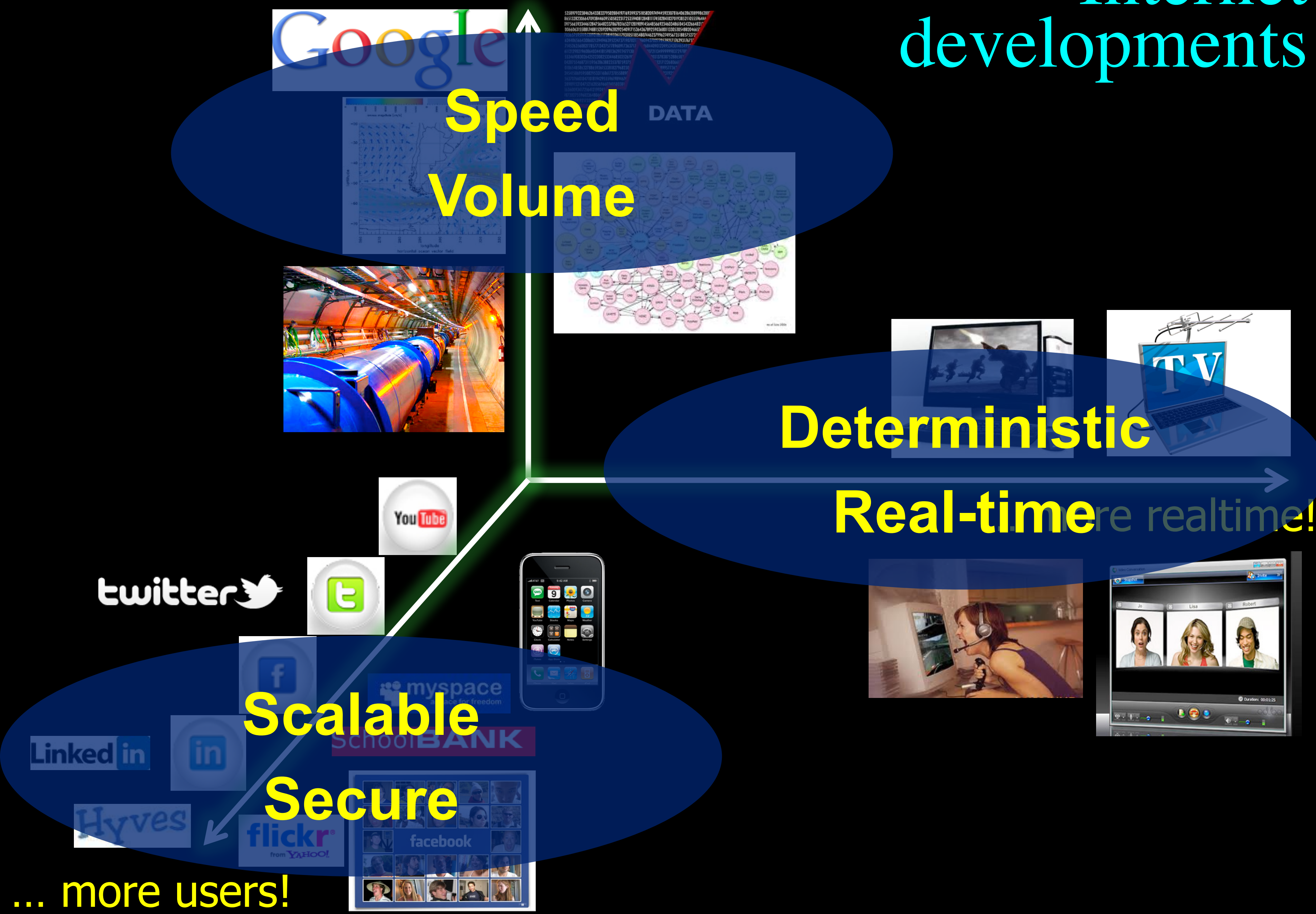


protocol LAN due to the easy comparison and convenience in the digital home. While consumer PC products has just started to migrate to a much higher bandwidth of 802.11n wireless LAN now working on next-generation standard definition is already in progress.

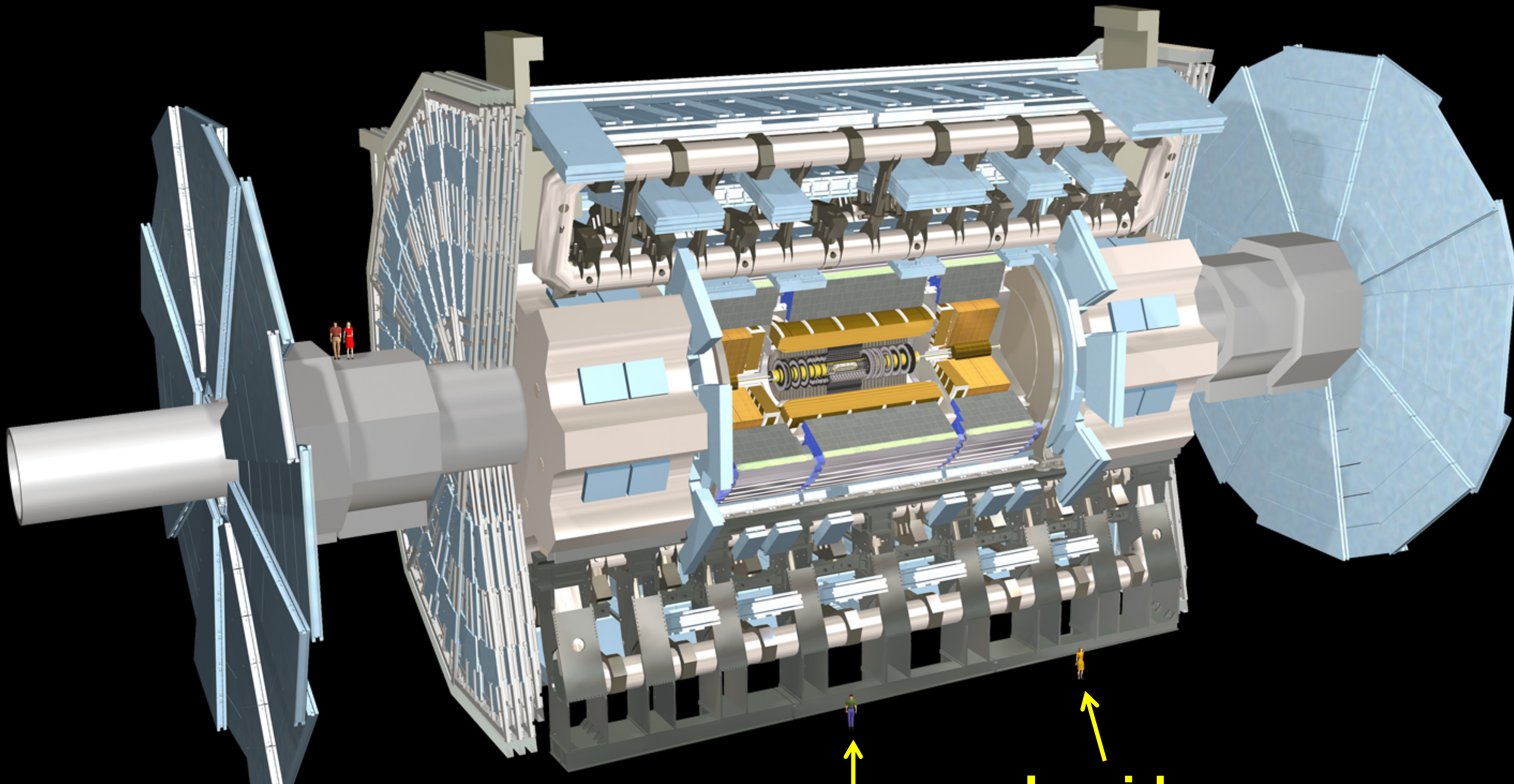

200.33.210.250
200.33.208
200.78.151.110
207.205.230.105
207.205.230.123
207.205.230.128
207.205.249.112
207.205.230.117
207.205.230.174
207.205.230.134
207.205.230.118
207.205.249.117
207.205.230.112
207.205.230.158
207.205.230.169
207.205.230.126
207.205.230.155
207.205.249.101


Internet developments
... more data!
Google
DATA
TV
... more realtime!
YouTube
twitter
myspace
a place for freedom
SchoolBANK
LinkedIn
Hyves
flickr
from YAHOO!
facebook
... more users!

Internet developments



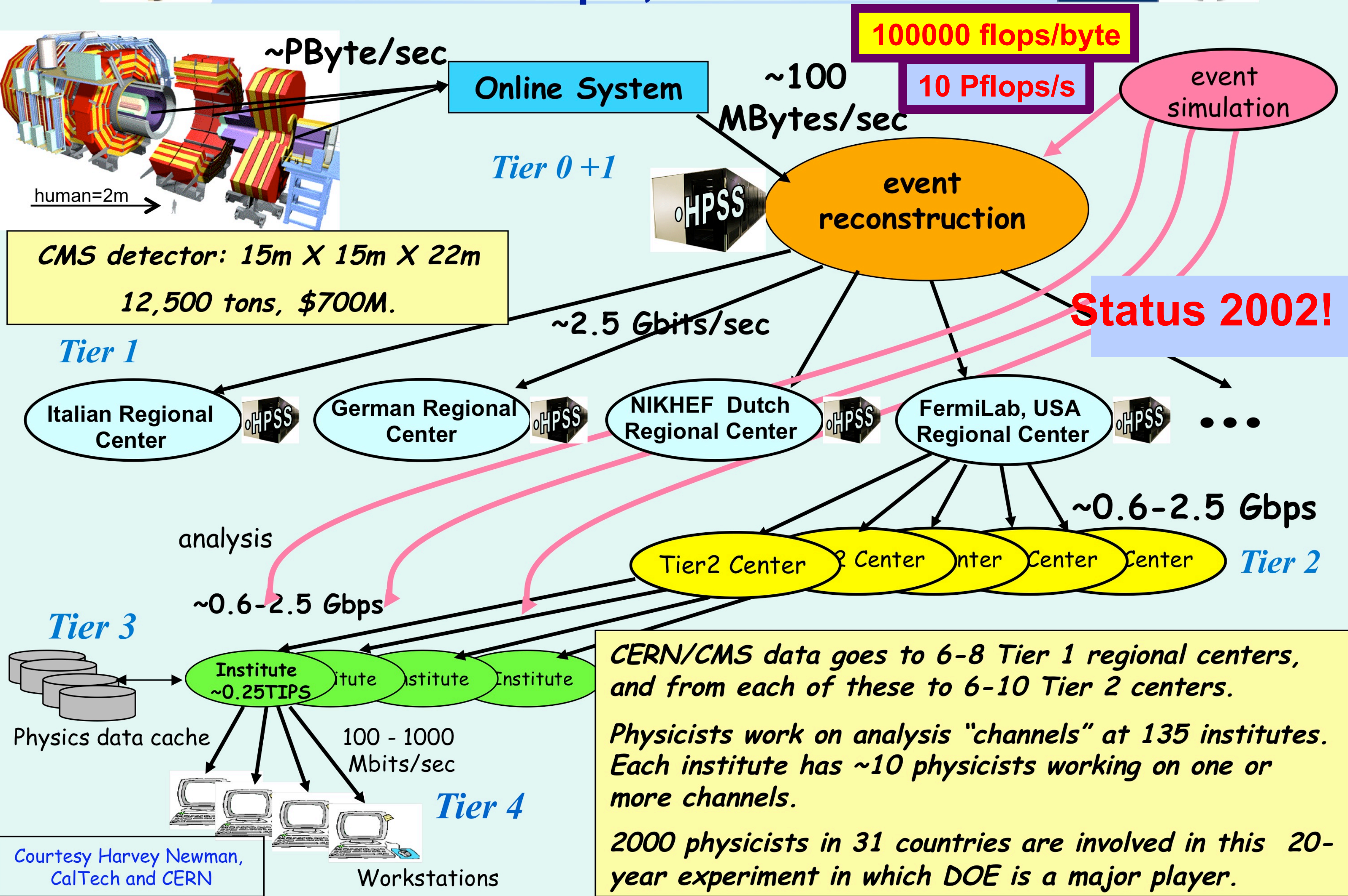
ATLAS detector @ CERN Geneve



ATLAS detector @ CERN Geneve

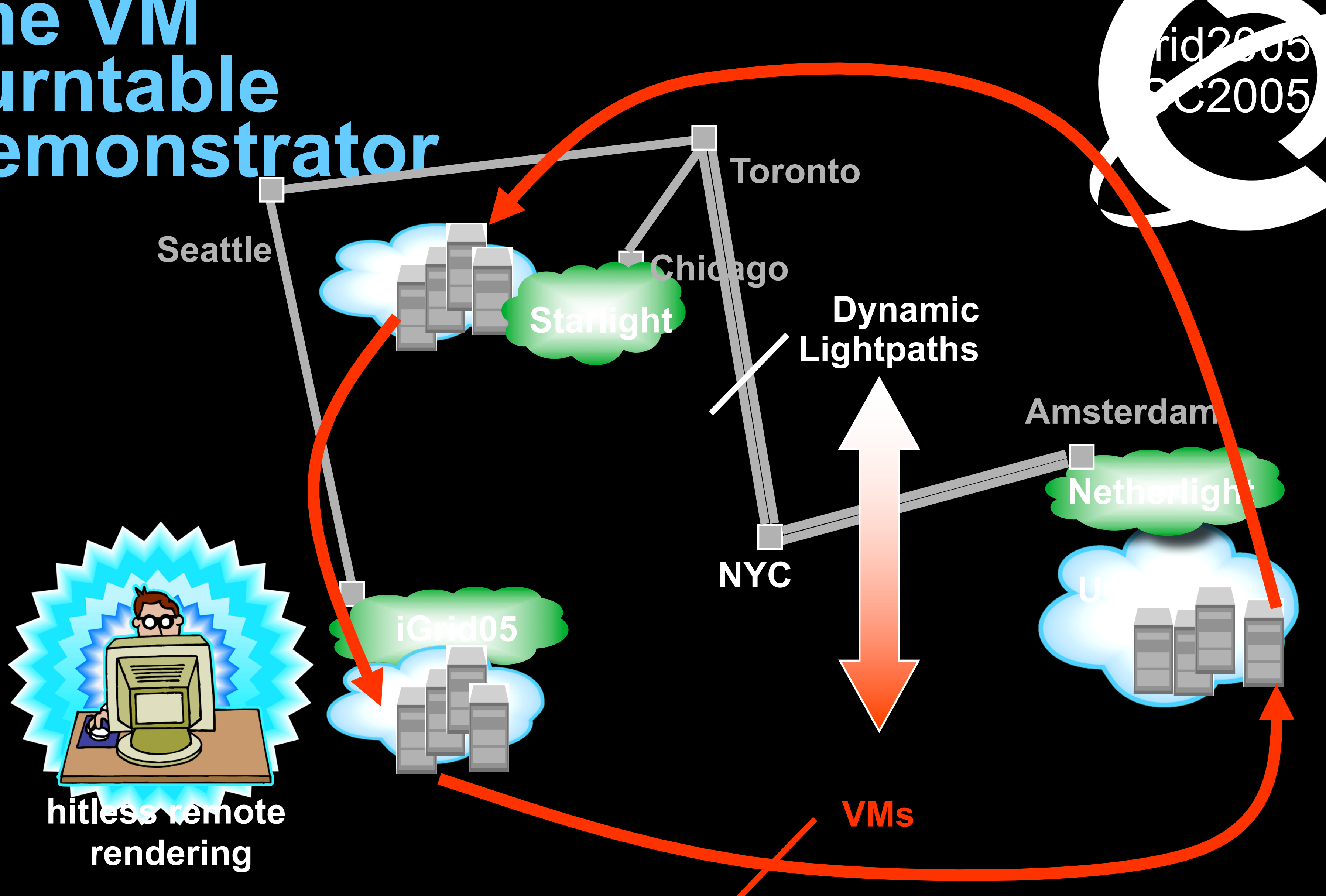
LHC Data Grid Hierarchy

CMS as example, Atlas is similar



Big and small flows don't go well together on the same wire! ☹

The VM Turntable Demonstrator



The VMs that are live-migrated run an iterative search-refine-search workflow against data stored in different databases at the various locations. A user in San Diego gets hitless rendering of search progress as VMs spin around

Experiment outcomes

- We have demonstrated seamless, live migration of VMs over MAN/WAN
- For this, we have realized a network service that
  - Exhibits predictable behavior; tracks endpoints
  - Flex bandwidth upon request by credited applications
  - Doesn't require peak provisioning of network resources
- Pipelining bounds the downtime in spite of high RTTs
  - San Diego – Amsterdam, 1GE, RTT = 200 msec, downtime <= 1 sec
  - Back to back, 1GE, RTT = 0.2-0.5 msec, downtime = ~0.2 sec*
  - ***Clark et al. NSDI 05 paper. Different workloads**
- VM + Lightpaths across MAN/WAN are deemed a powerful and general alternative to RPC, GRAM approaches
- We believe it's a representative instance of active

Alien light
From idea to realisation!

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

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.

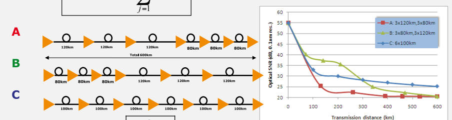
In this demonstration we will investigate the performance of a 40Gb/s PM-QPSK alien wavelength installed on a 10Gb/s DWDM infrastructure.

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

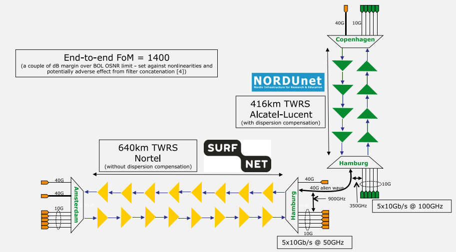


	FOM
A	5504
B	5504
C	1897

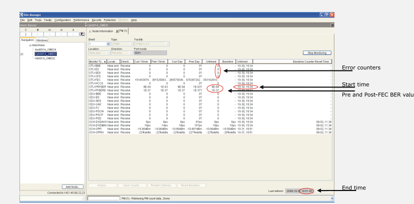
Easy-to-use formula that accurately quantifies transmission system performance

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.

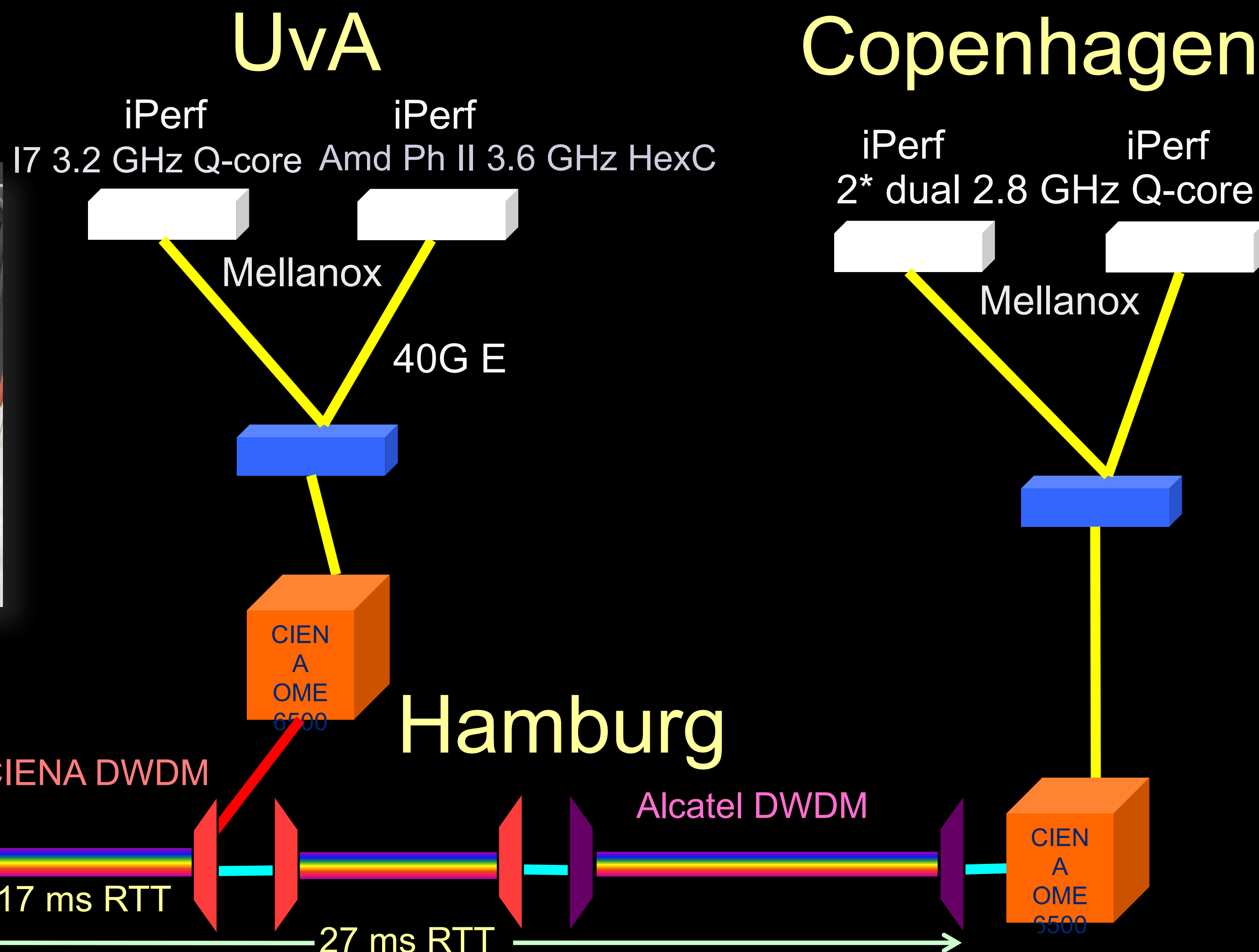
NORTEL NORDUnet telindus together with ISIT SURF NET

REFERENCES [1] "OPERATIONAL SOLUTIONS FOR AN OPEN DWDM LAYER", O. GERSTEL ET AL, OFC'2009 I [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 I [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

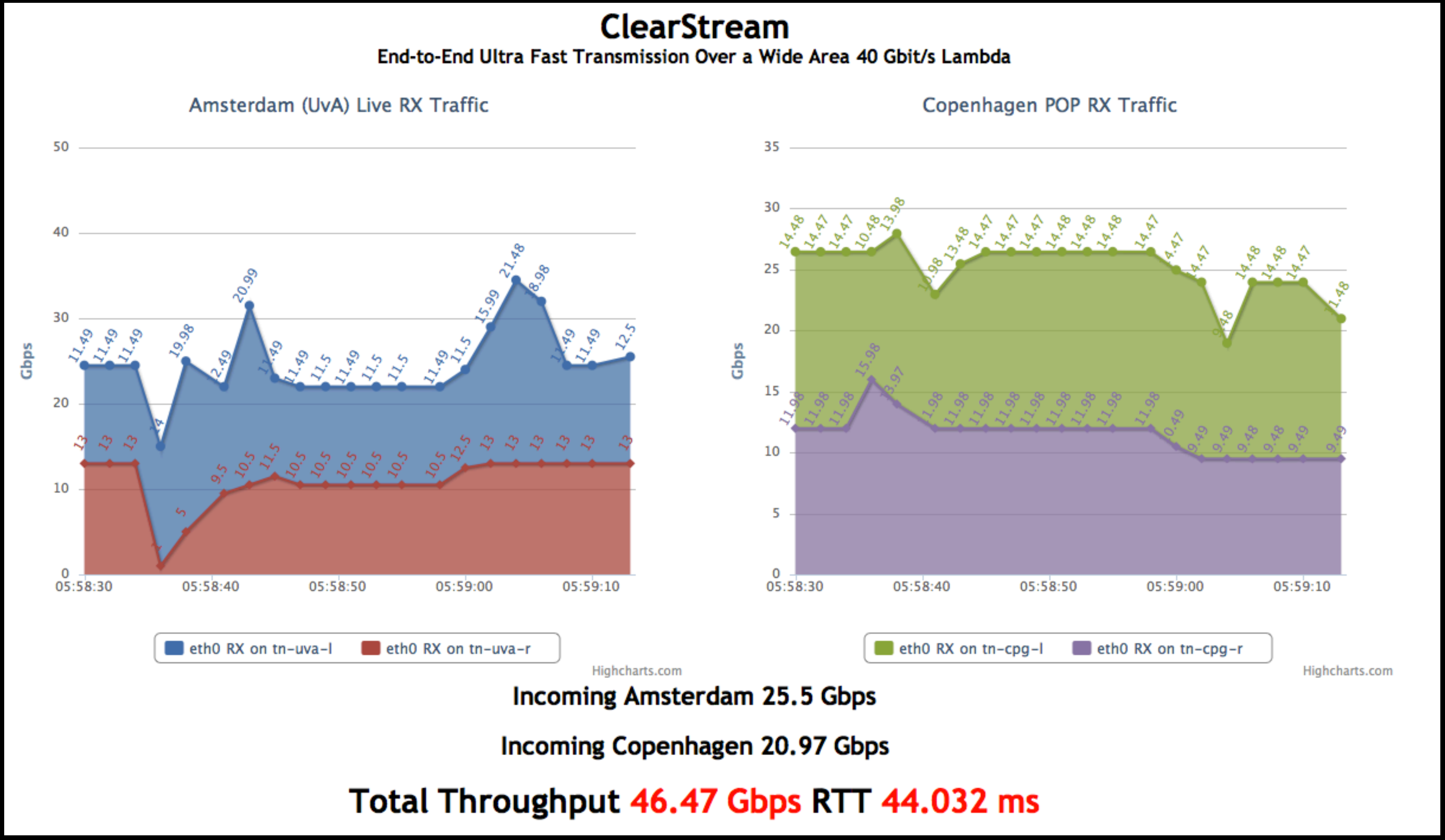
ClearStream @ TNC2011

Setup codename: FlightCees

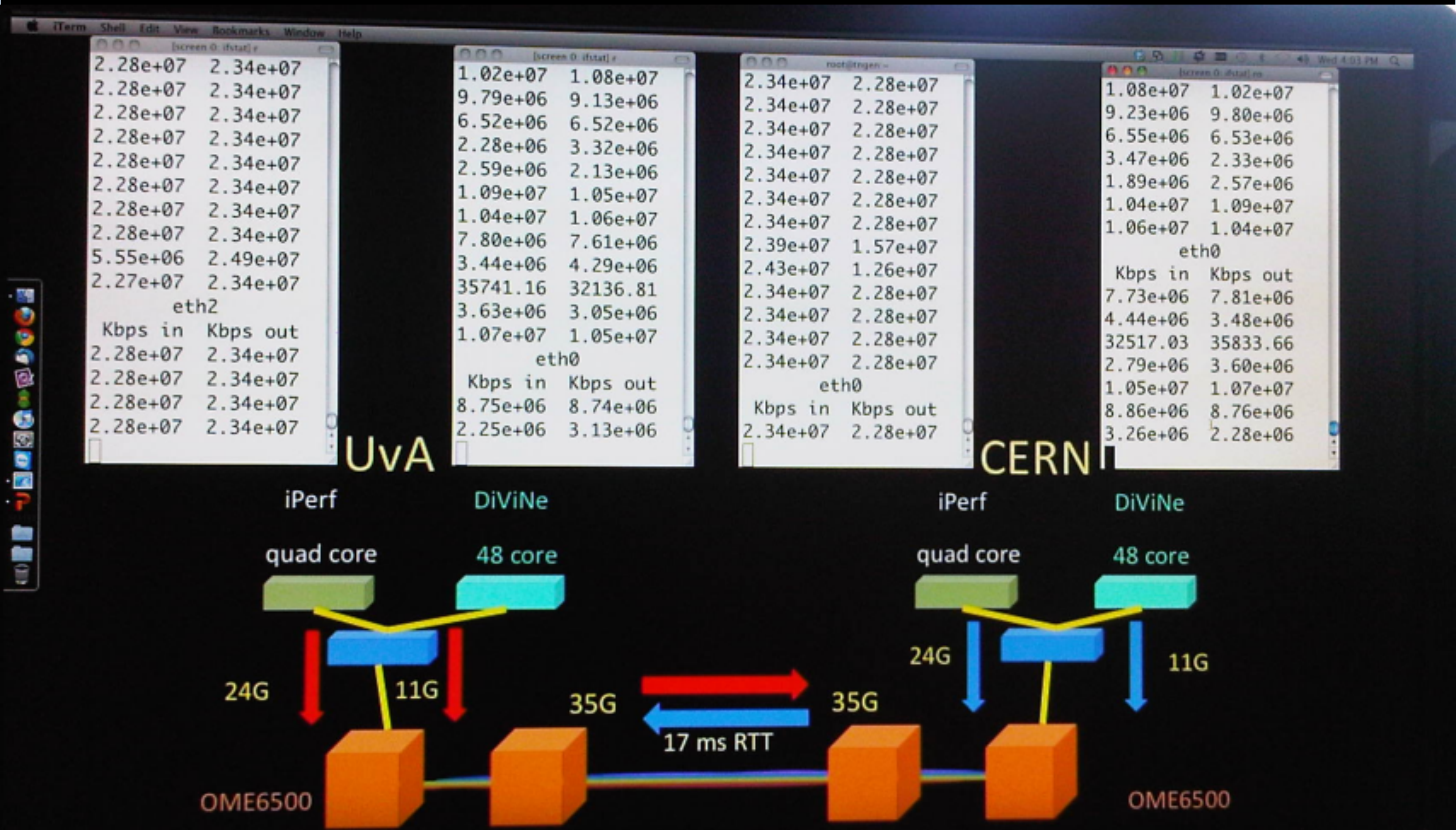


Amsterdam – Geneva (CERN) – Copenhagen – 4400 km (2700 km alien light)

Visit CIENA Booth
surf to http://tnc11.delaat.net



From GLIF October 2010 @ CERN



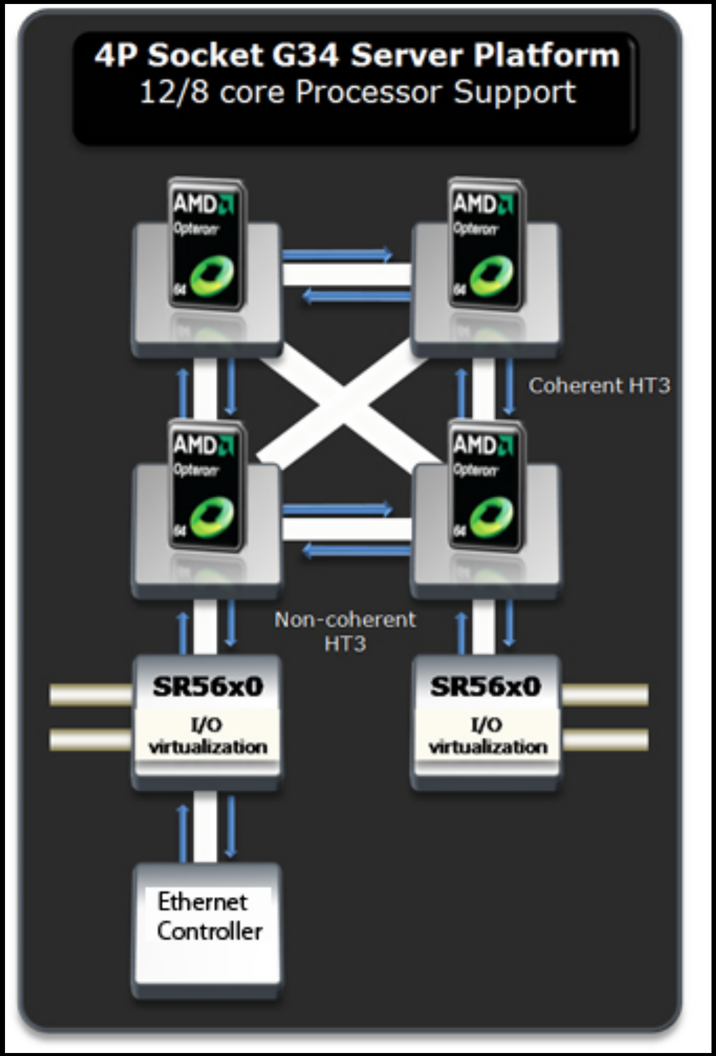
Results (rtt = 17 ms)

- 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

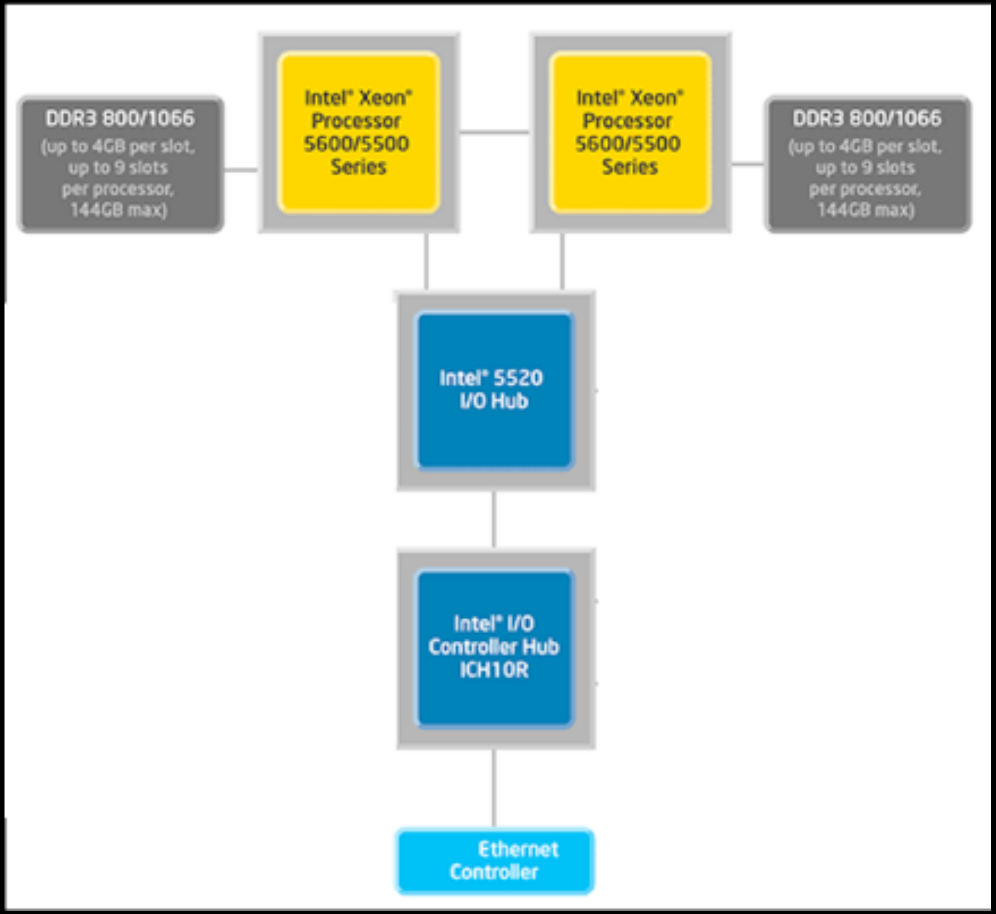
Performance Explained

- Mellanox 40GE card is PCI-E 2.0 8x (5GT/s)
- 40Gbit/s raw throughput but ….
- PCI-E is a network-like protocol
  - 8/10 bit encoding -> 25% overhead -> 32Gbit/s maximum data throughput
  - Routing information
- Extra overhead from IP/Ethernet framing
- Server architecture matters!
  - 4P system performed worse in multithreaded iperf

Server Architecture

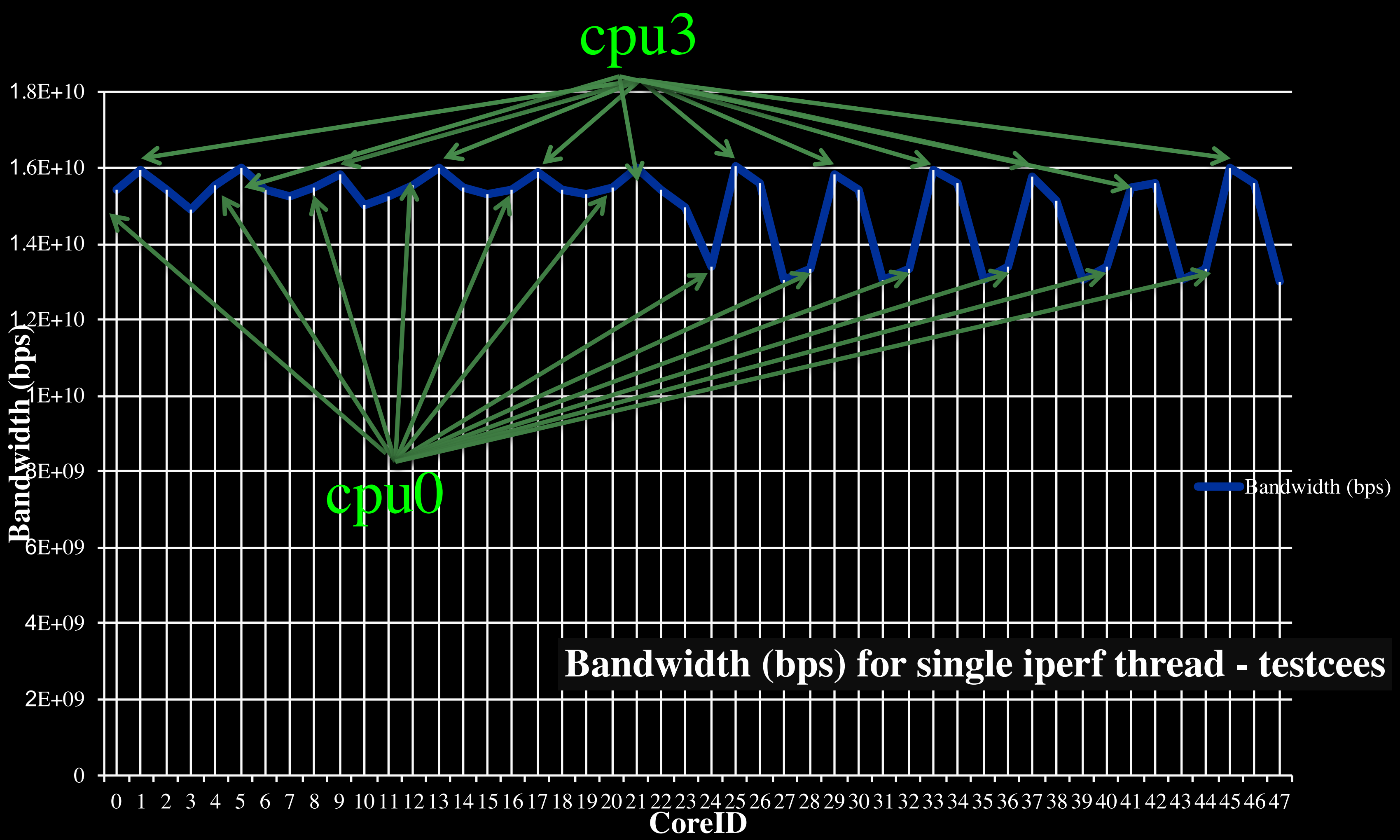


DELL R815
4 x AMD Opteron 6100



Supermicro X8DTT-HIBQF
2 x Intel Xeon

CPU Topology benchmark



We used numactl to bind iperf to cores

SARNET: Security Autonomous Response with programmable NETworks

Cees de Laat
Leon Gommans, Rodney Wilson, Rob Meijer
Tom van Engers, Marc Lyonais, Paola Grosso, Frans Franken,
Ameneh Deljoo, Ralph Koning, Ben de Graaff, Stojan Trajanovski

SNE

UNIVERSITY OF AMSTERDAM

TNO

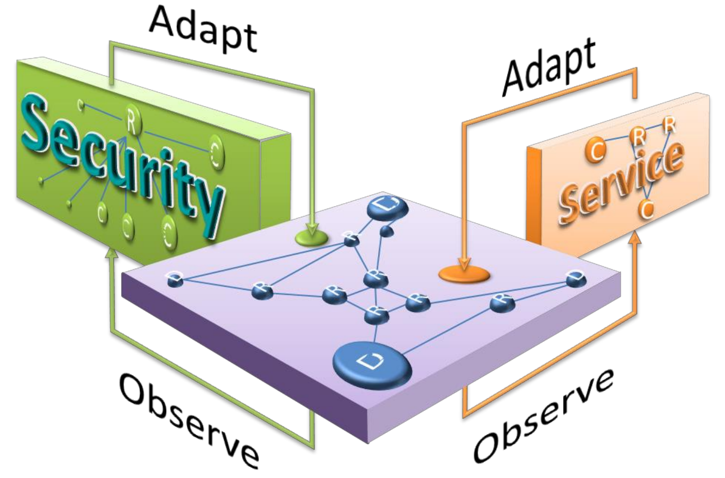
ciena

AIRFRANCE KLM

NWO
Netherlands Organisation for Scientific Research

COMMIT/

Cyber security program



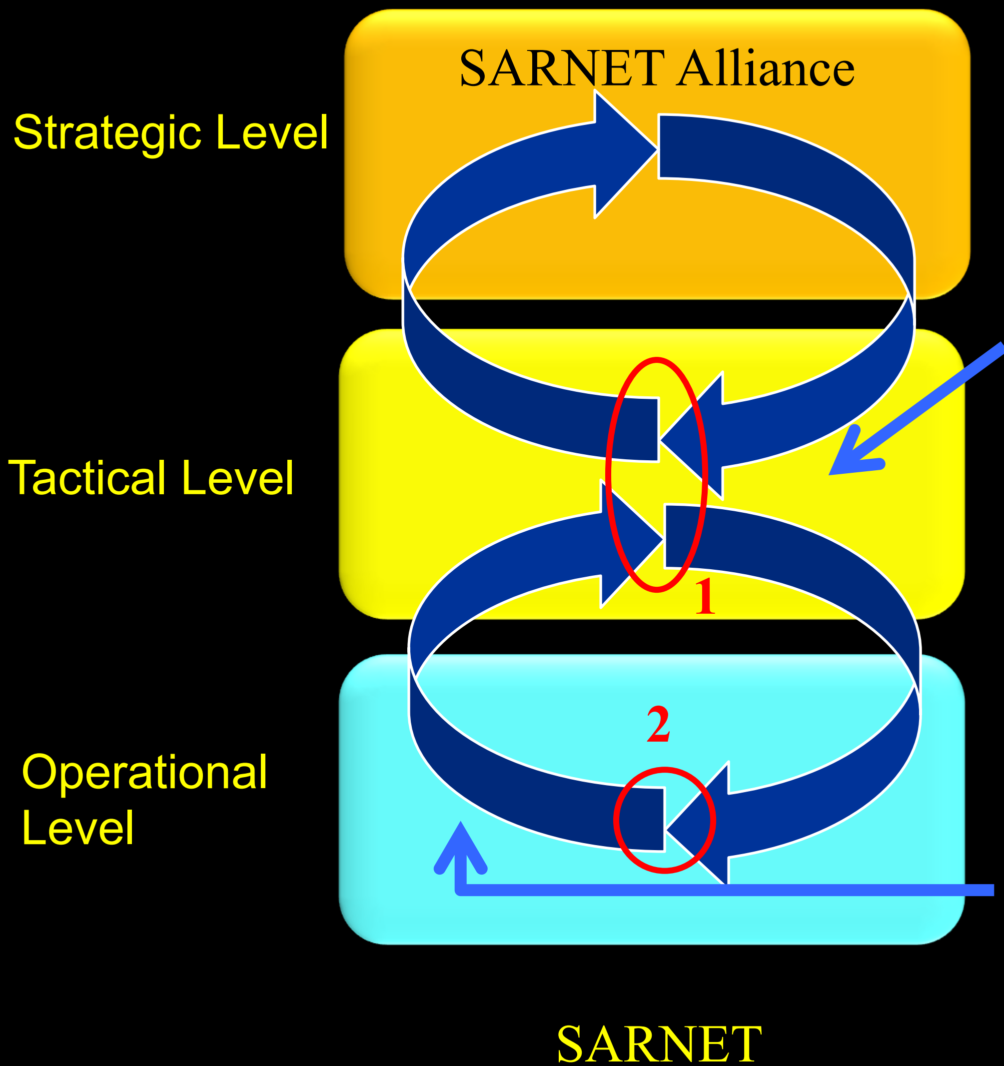
Research goal is to obtain the knowledge to create ICT systems that:

- model their state (situation)
- discover by observations and reasoning if and how an attack is developing and calculate the associated risks
- have the knowledge to calculate the effect of counter measures on states and their risks
- choose and execute one.

In short, we research the concept of networked computer infrastructures exhibiting SAR: Security Autonomous Response.

Context & Goal

Security Autonomous Response NETwork Research

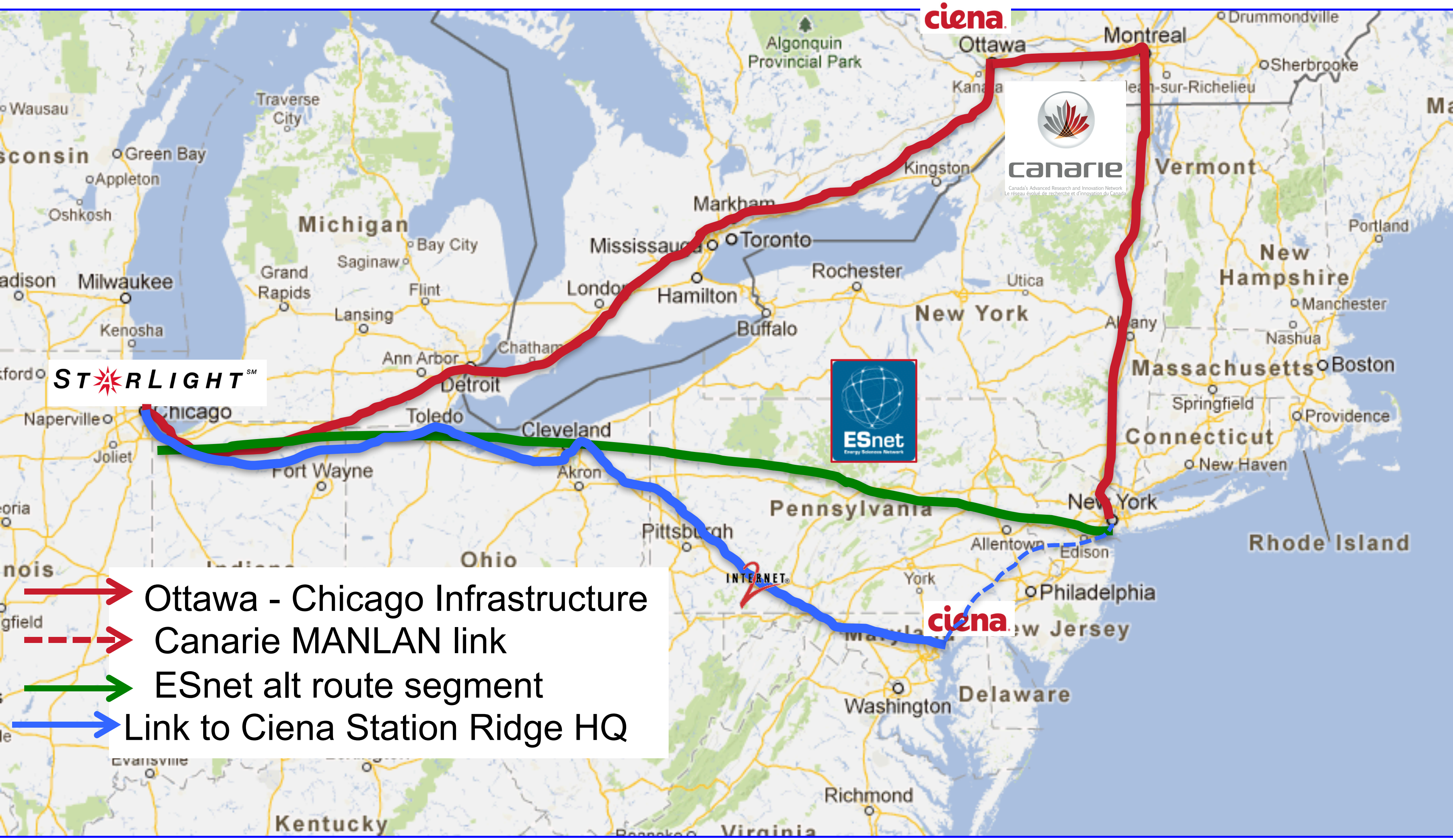


Ameneh Deljoo (PhD):
Why create SARNET Alliances? Model autonomous SARNET behaviors to identify risk and benefits for SARNET stakeholders

Stojan Trajanovski (PD):
Determine best defense scenario against cyberattacks deploying SARNET functions (1) based on security state and KPI information (2).

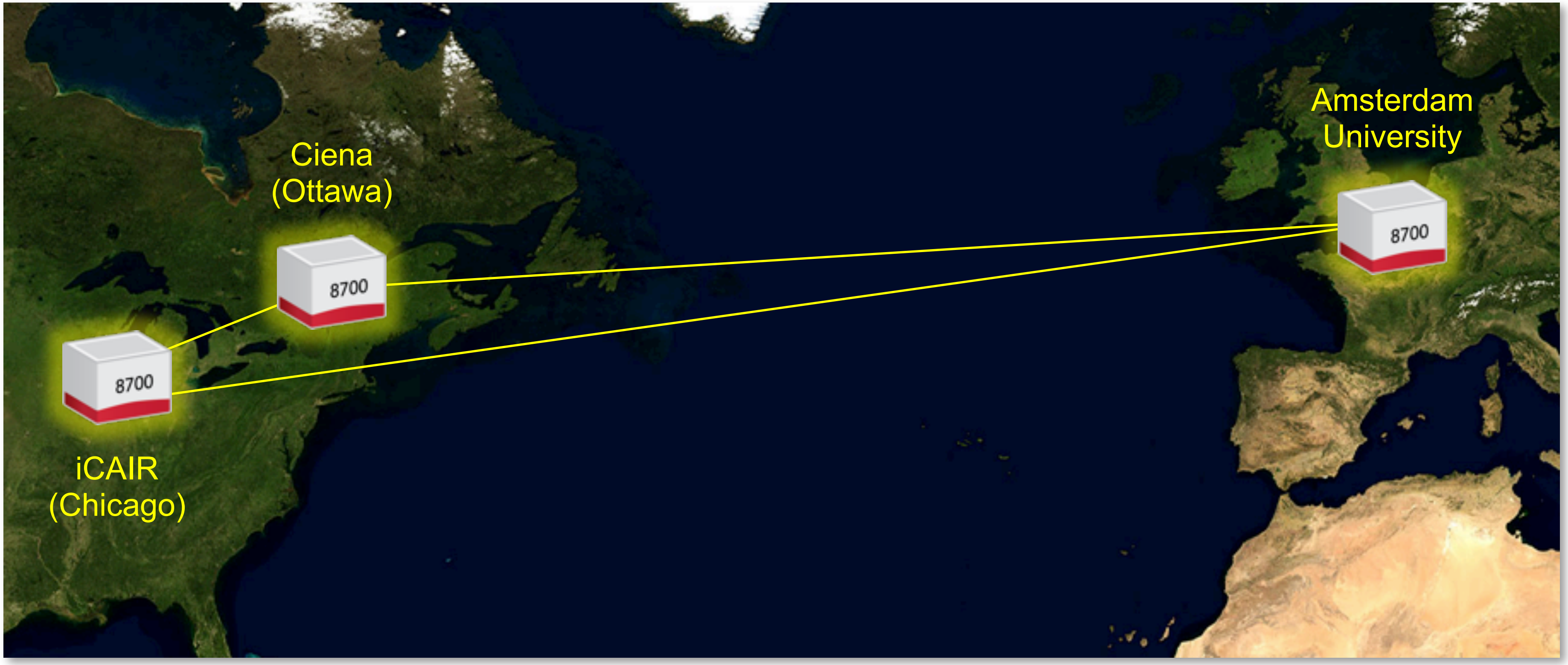
Ralph Koning (PhD)
Ben de Graaff (SP):
1. Design functionalities needed to operate a SARNET using SDN/NFV
2: deliver security state and KPI information (e.g cost)

Ciena's CENI topology



CENI, International extension to University of Amsterdam

Research Triangle Project.   Operation Spring of 2015



National Science Foundations ExoGENI racks, installed at UvA (Amsterdam), Northwestern University (Chicago) and Ciena's labs (Ottawa), are connected via a high performance 100G research network and trans-Atlantic network facilities using the Ciena 8700 Packetwave platform. This equipment configuration is used to create a computational and storage test bed used in collaborative demonstrations.

Position of demo @ SC15

Objective

- To get a better understanding for cyber attack complexity by visually defend a network suffering from basic volumetric attacks.
- To find a way to visualize future research in automated response.

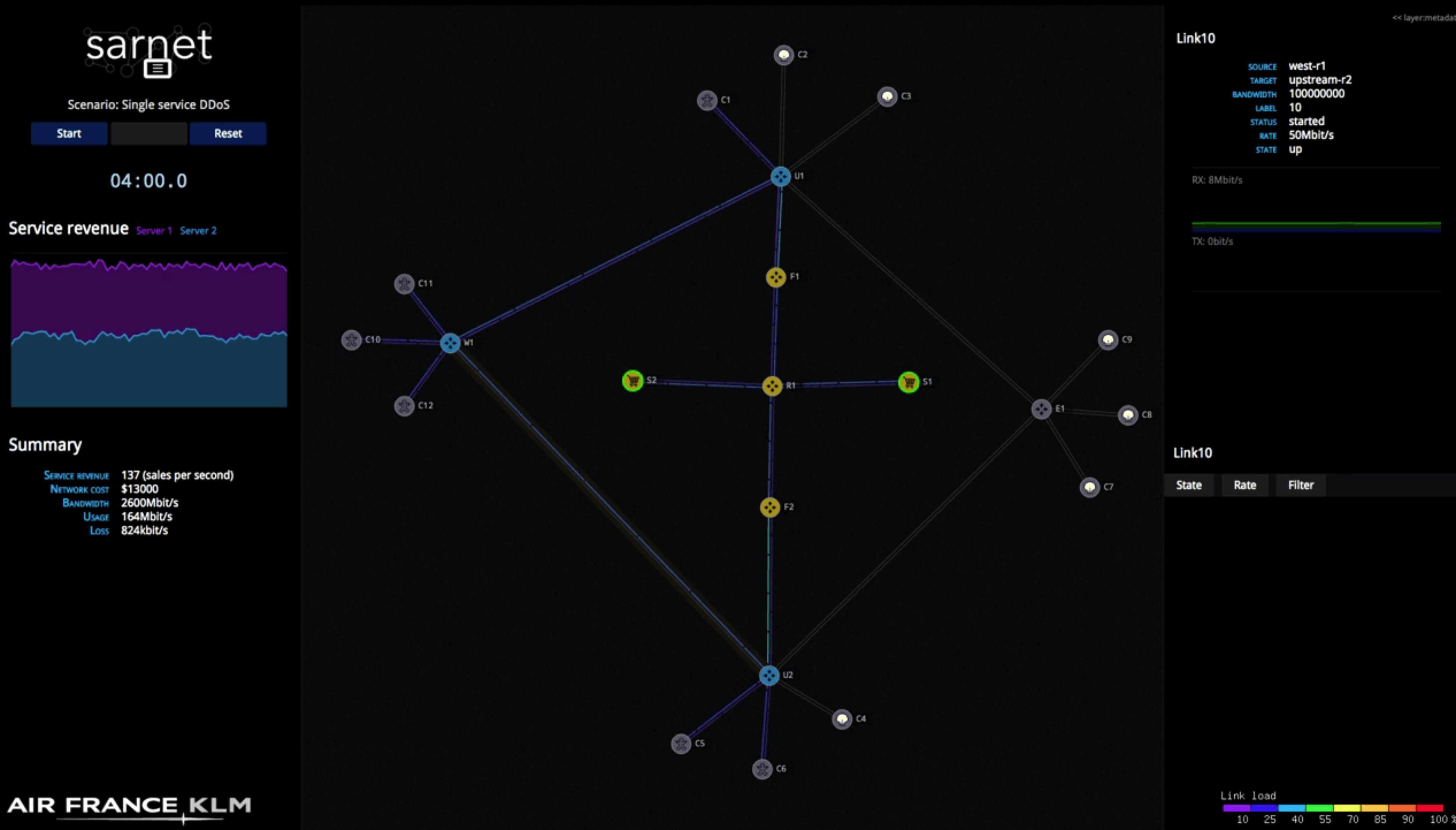
Demo highlights

- Pre-programmed attack scenarios that are able to show defense functions.
- Virtual sales + income from web services
- Defense cost

DDoS Defence functions.

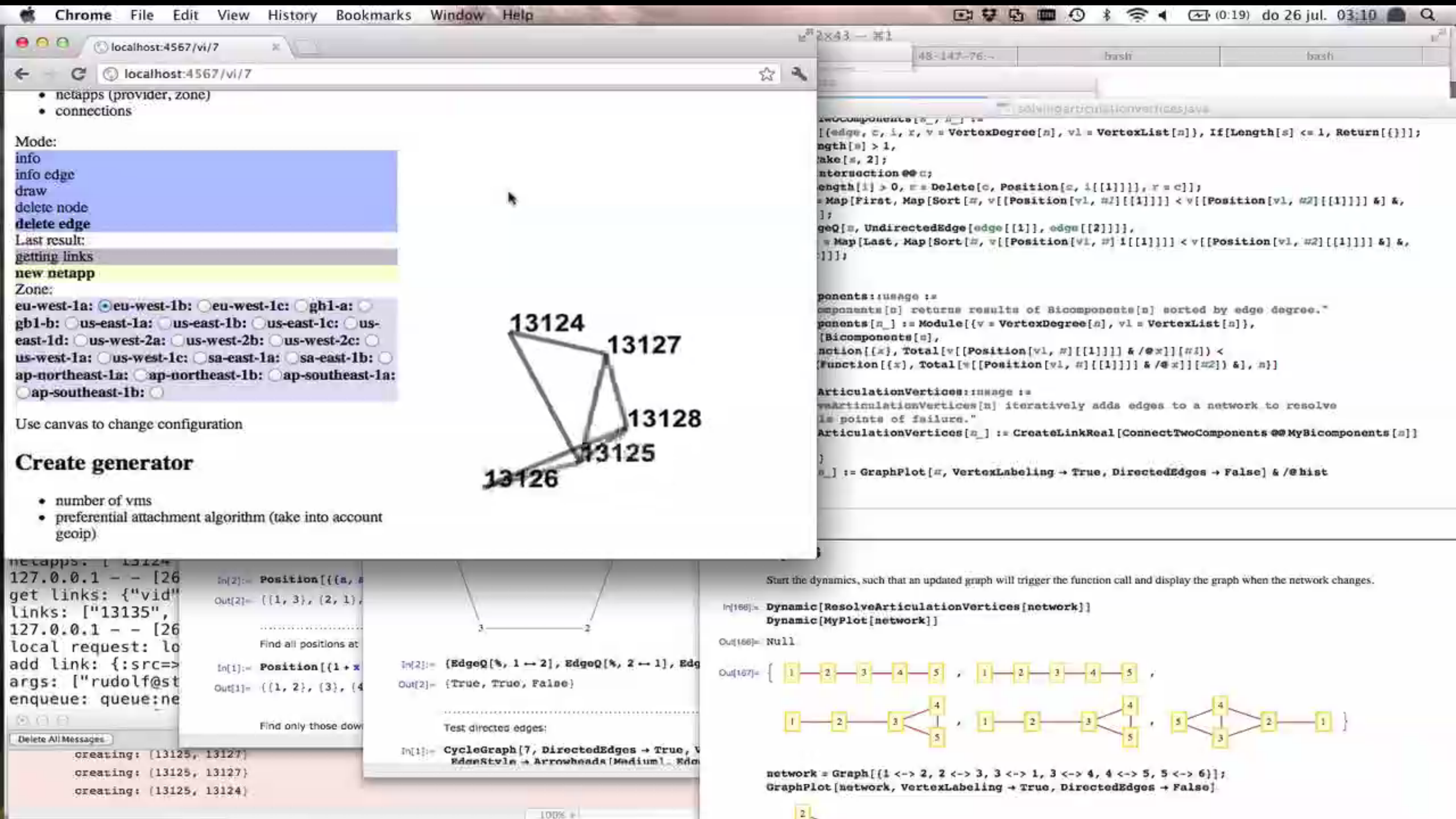
- Filtering
- Blocking
- Resource Scaling

Demo



University of Amsterdam

Basic operating system loop



Service Provider Group framework

A Service Provider Group (SPG) is an organisation structure providing a defined service only available if its members collaborate.

Examples:

Internet2 NET+

SKYTEAM

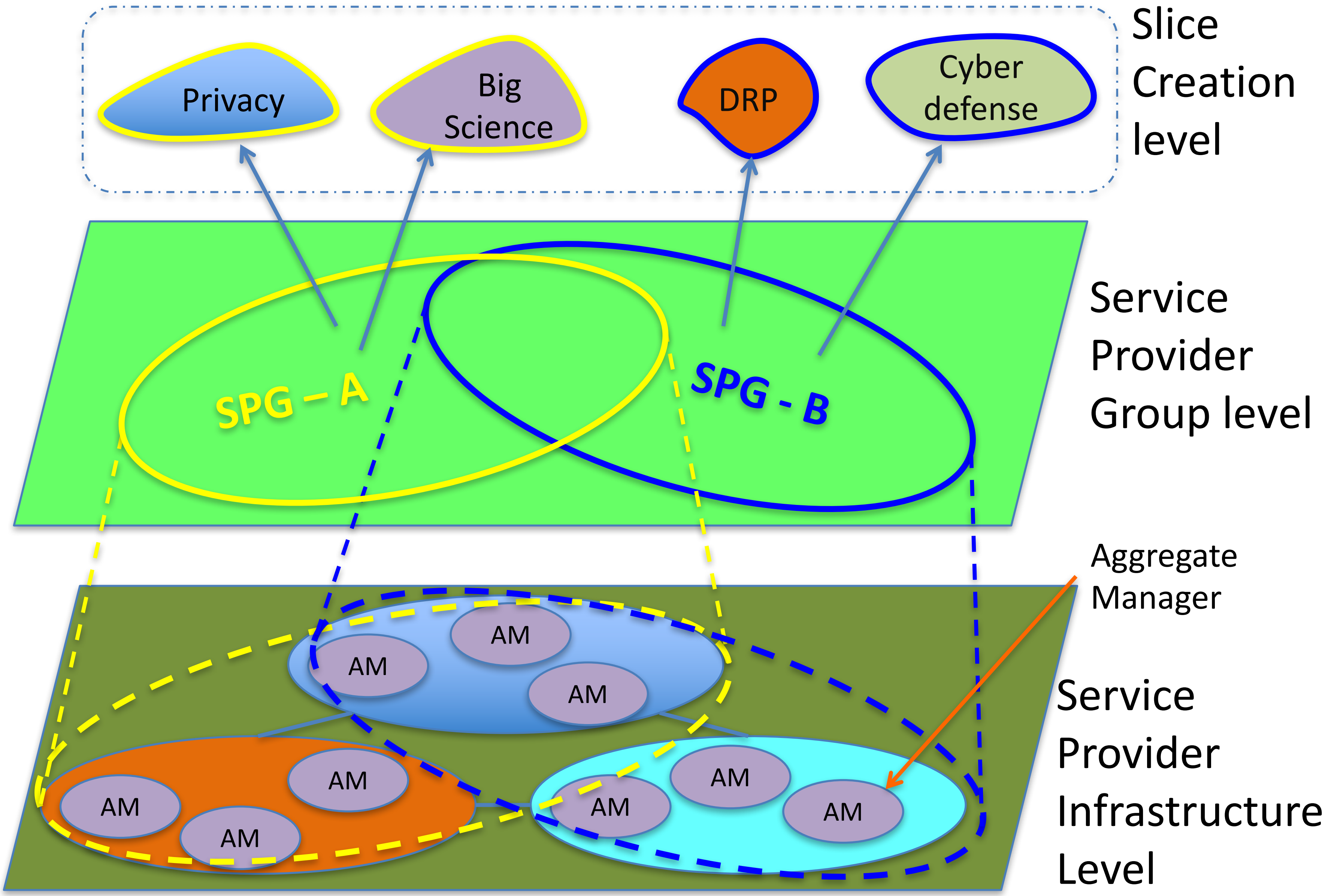
geni
Exploring Networks of the Future

glif

eduroam

MasterCard

Envisioned role of the SPG: define slice archetypes?



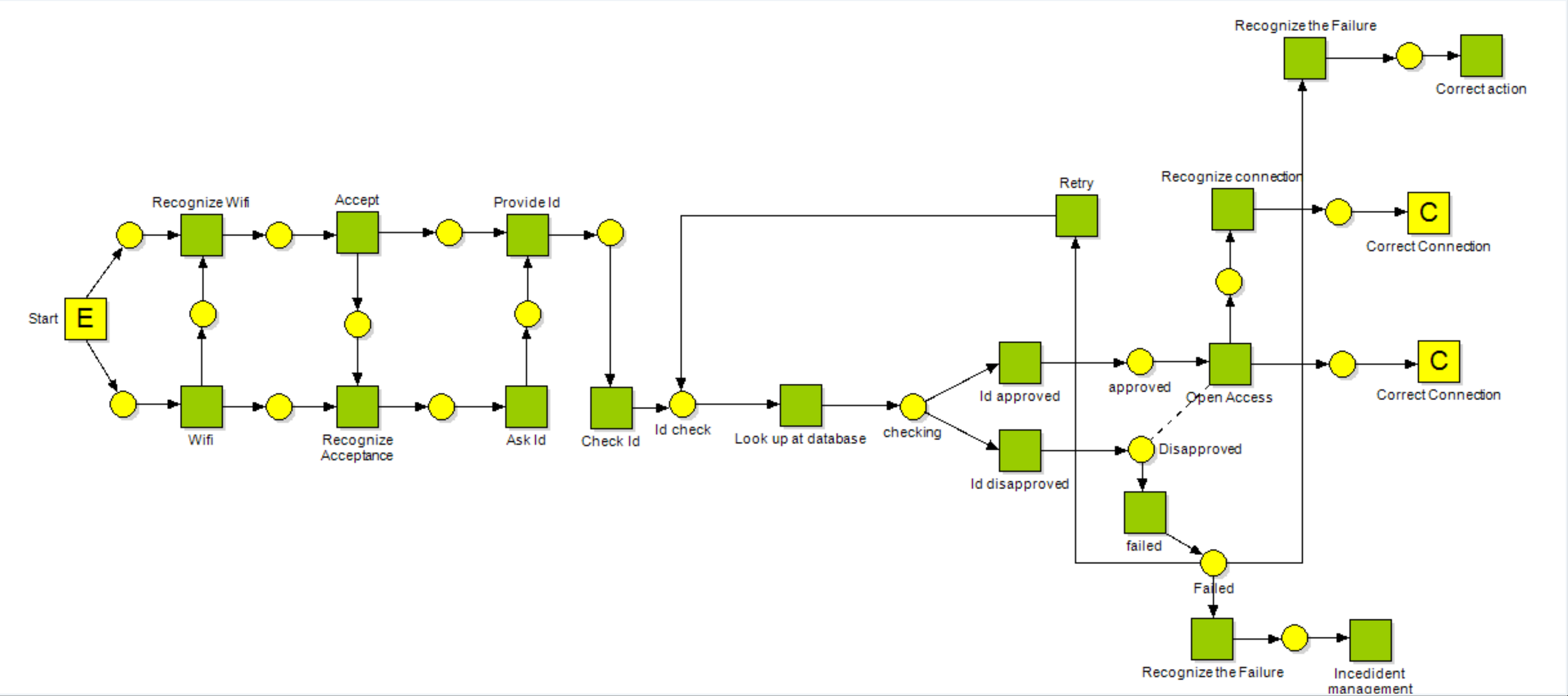
Agent Based Modelling Framework

	Main component
Signal layer	Message / Act
Action layer	Action / Activity
Intentional layer	Intention
Motivational layer	Motive

In our model, we refer to four layers of components:

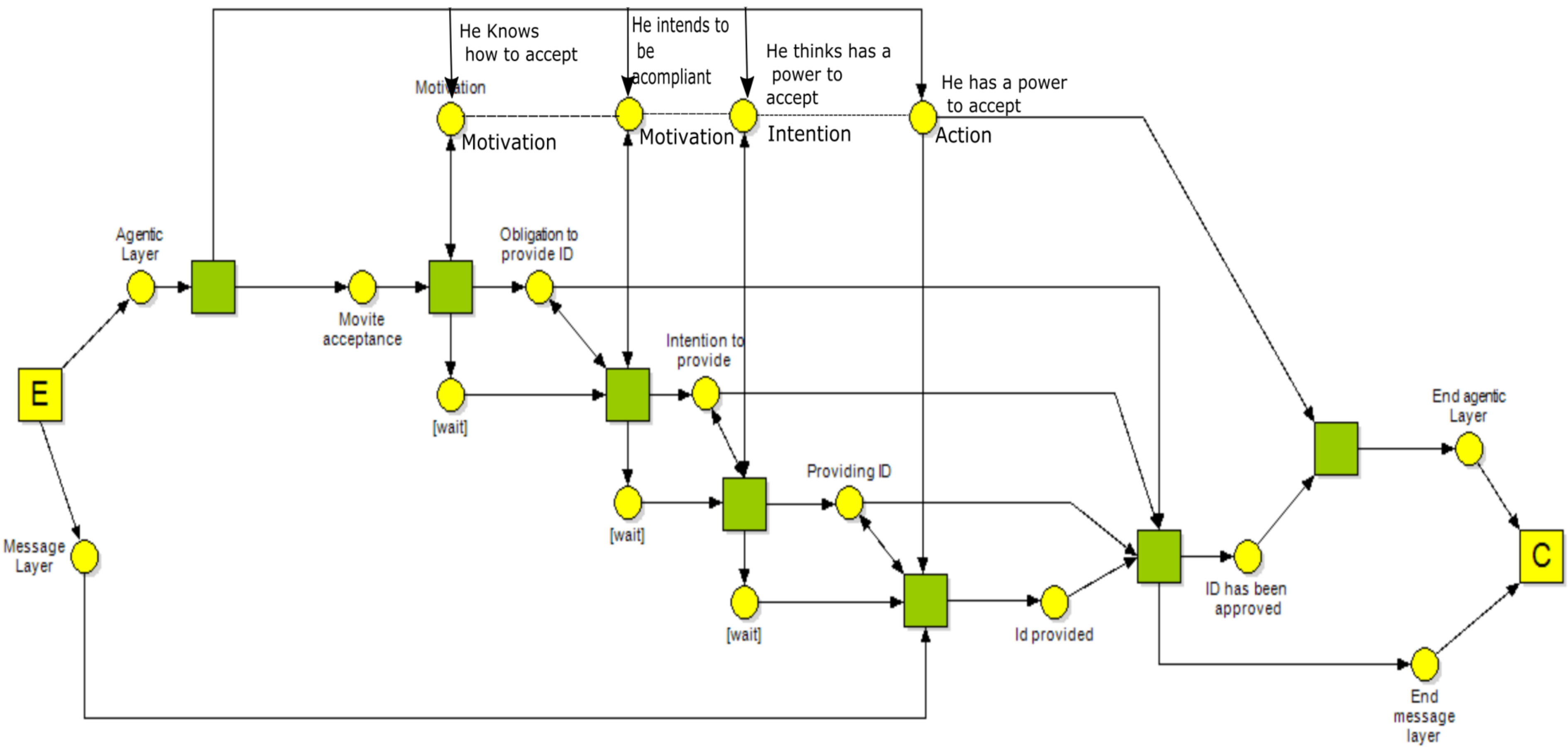
- the signal layer— describes acts, side-effects and failures showing outcomes of actions in a topology.
- the action layer—actions: performances that bring a certain result,
- the intentional layer—intentions: commitments to actions, or to build up intentions,
- the motivational layer—motives: events triggering the creation of intentions.

Simplified Eduroam case at signalling layer



Petri net of EduRoam Case
(first step)

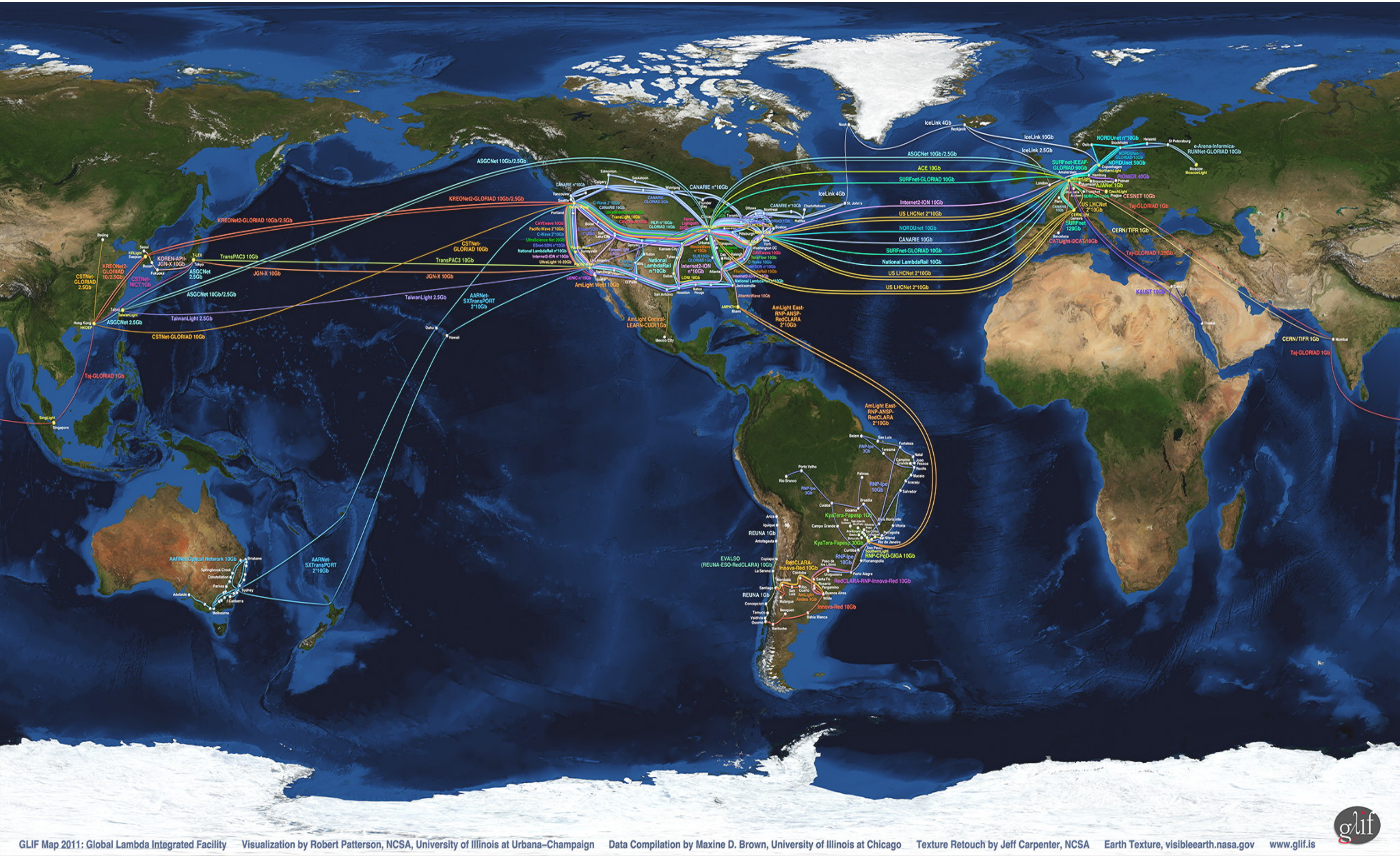
Describing Intentions, Motivations and Actions



Petri net of EduRoam Case

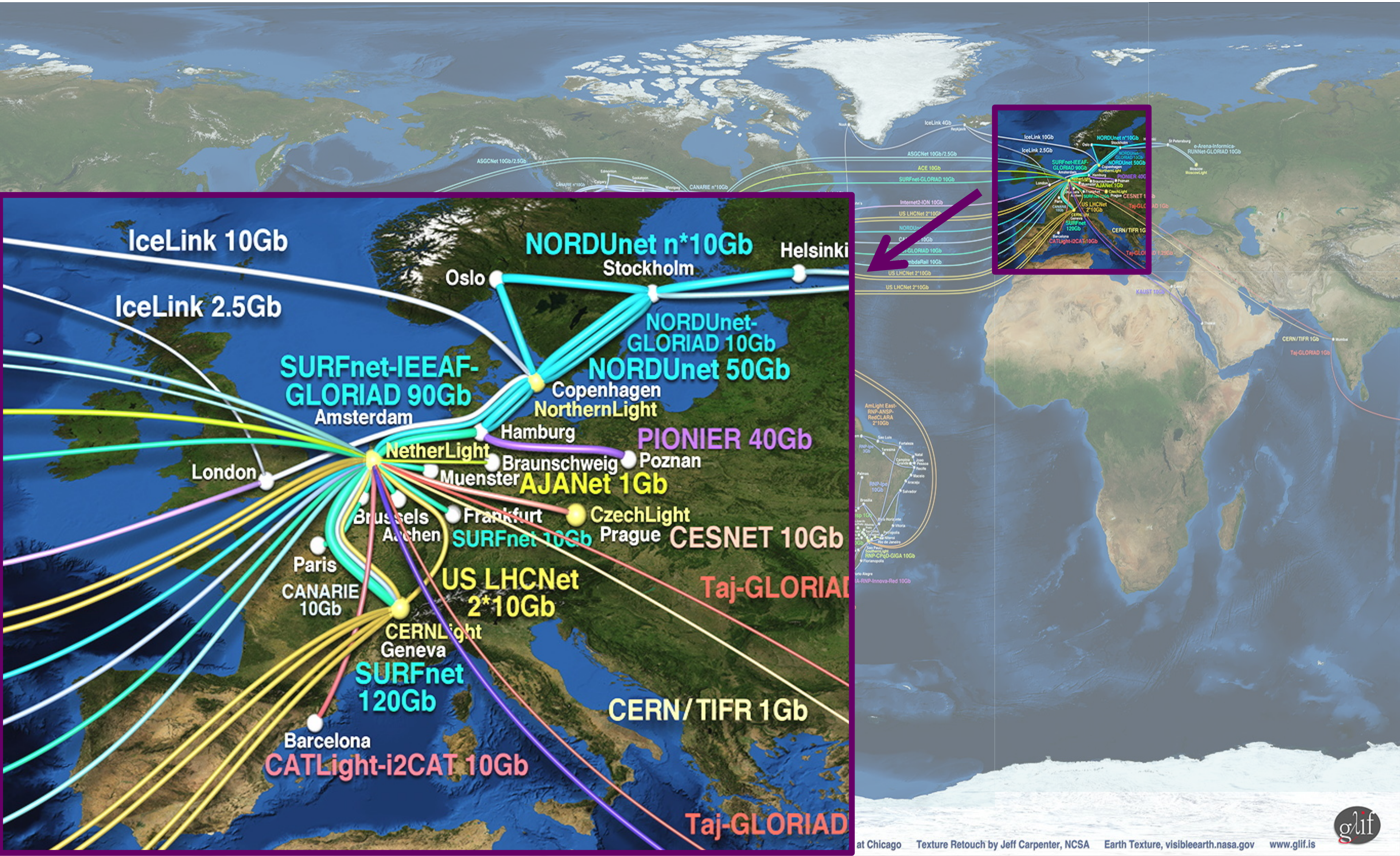
The GLIF – LightPaths around the World

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.

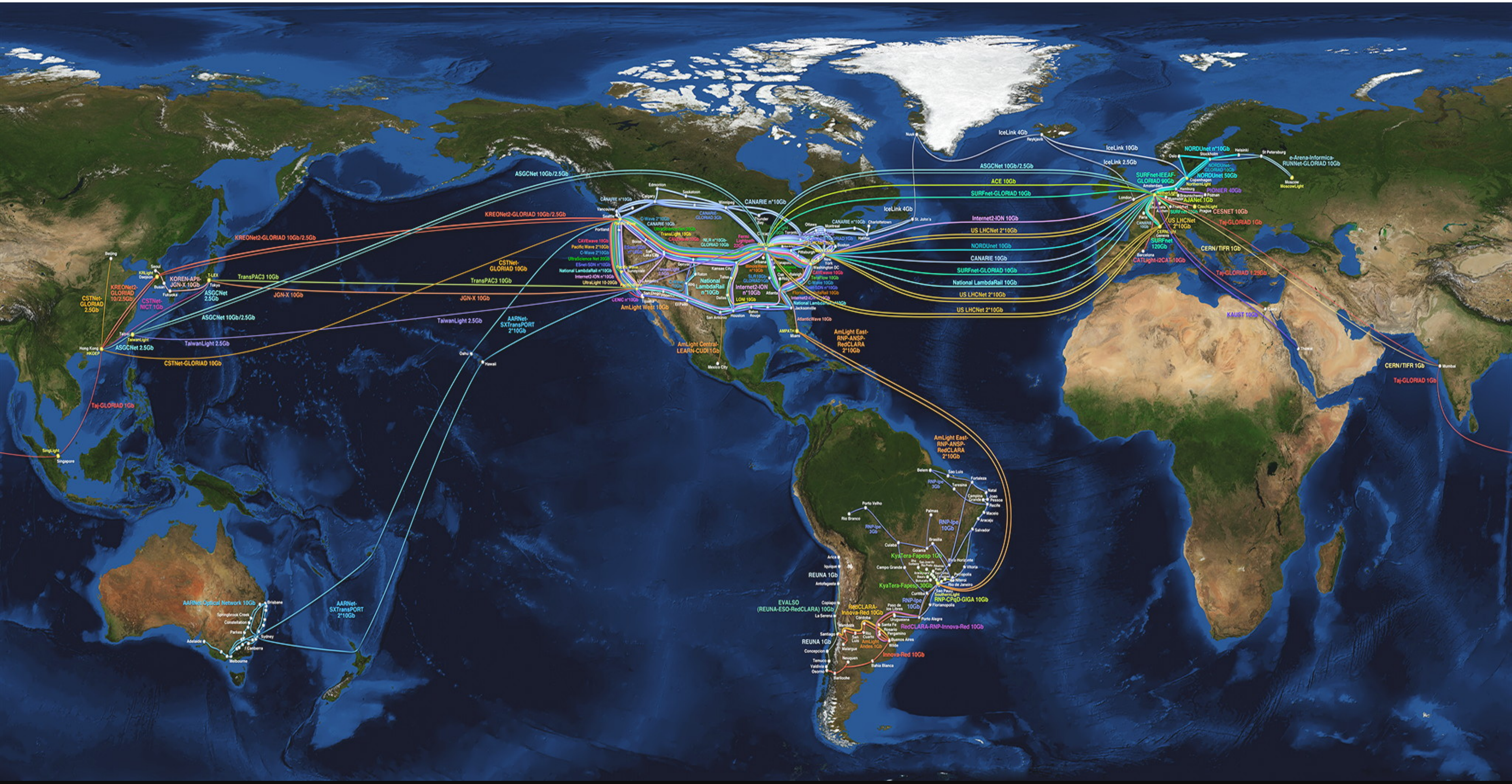


The GLIF – LightPaths around the World

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.



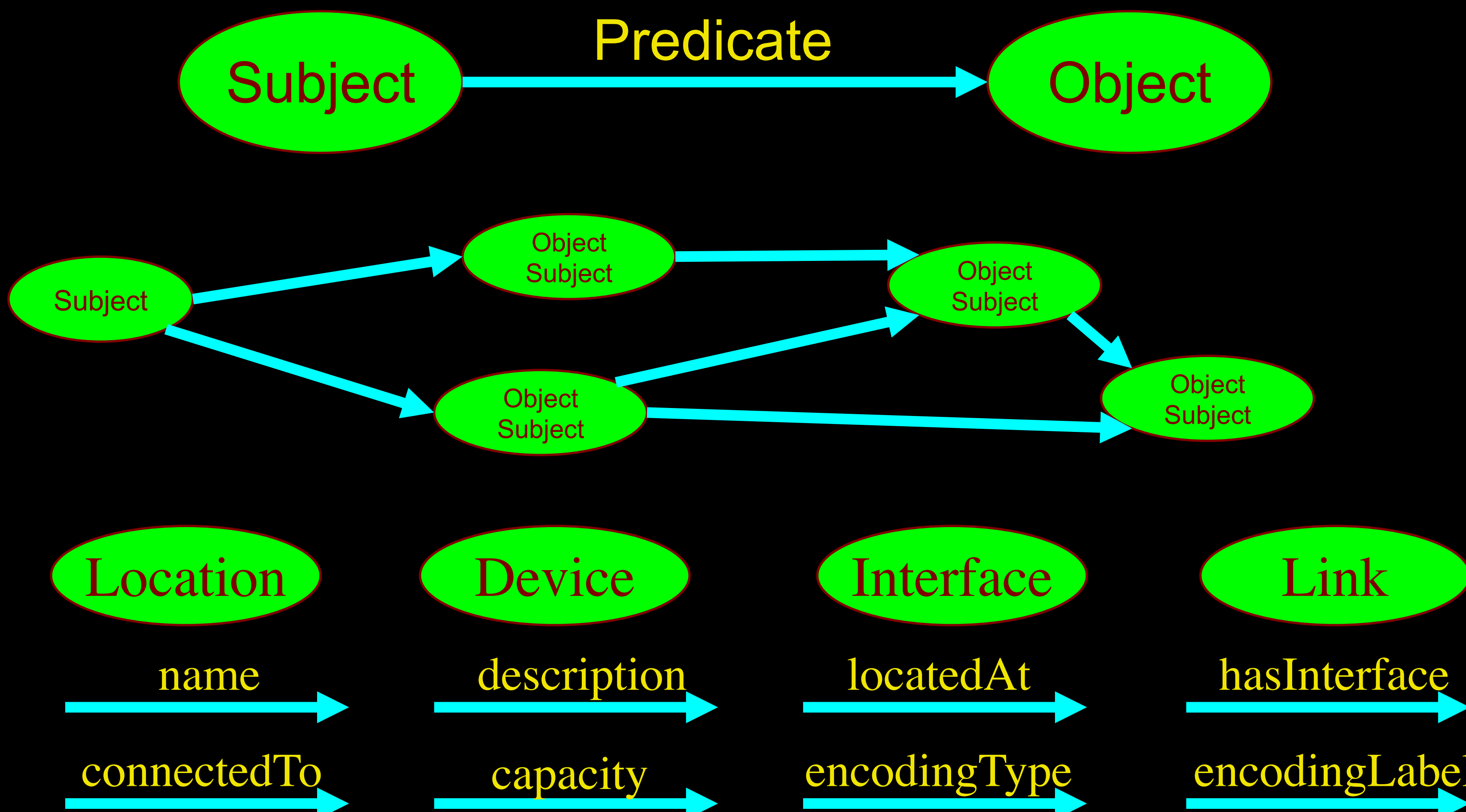
The GLIF – LightPaths around the World



We investigate: TOMTOM for complex networks!

LinkedIN for Infrastructure

- From semantic Web / Resource Description Framework.
- The RDF uses XML as an interchange syntax.
- Data is described by triplets (Friend of a Friend):



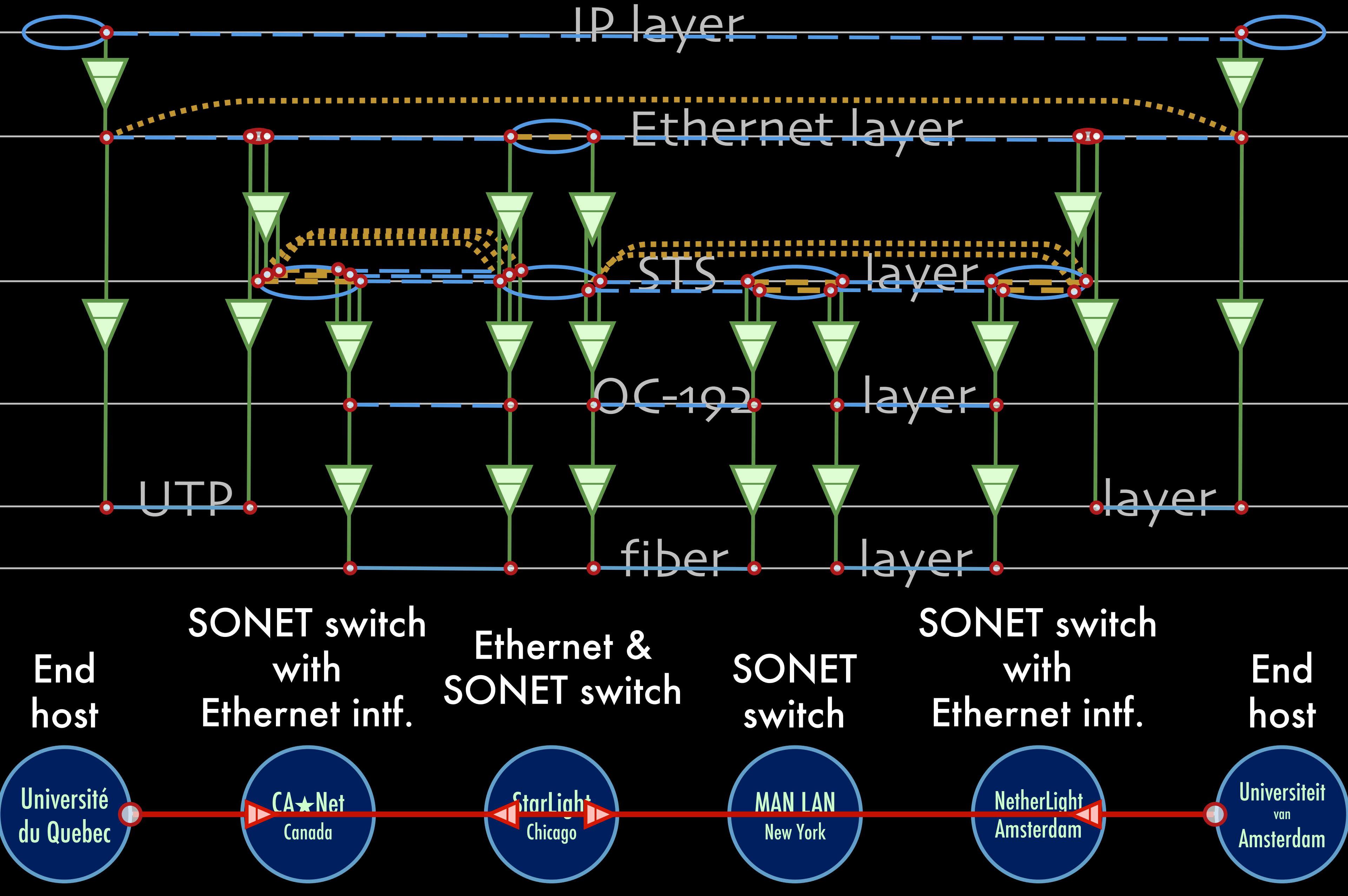
NetherLight in RDF

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
      xmlns:ndl="http://www.science.uva.nl/research/air/ndl#">
<!-- Description of Netherlight -->
<ndl:Location rdf:about="#Netherlight">
      <ndl:name>Netherlight Optical Exchange</ndl:name>
</ndl:Location>
<!-- TDM3.amsterdam1.netherlight.net -->
<ndl:Device rdf:about="#tdm3.amsterdam1.netherlight.net">
      <ndl:name>tdm3.amsterdam1.netherlight.net</ndl:name>
      <ndl:locatedAt rdf:resource="#amsterdam1.netherlight.net"/>
      <ndl:hasInterface rdf:resource="#tdm3.amsterdam1.netherlight.net:501/1"/>
      <ndl:hasInterface rdf:resource="#tdm3.amsterdam1.netherlight.net:501/3"/>
      <ndl:hasInterface rdf:resource="#tdm3.amsterdam1.netherlight.net:501/4"/>
      <ndl:hasInterface rdf:resource="#tdm3.amsterdam1.netherlight.net:503/1"/>
      <ndl:hasInterface rdf:resourc
      <ndl:hasInterface rdf:resourc
      <ndl:hasInterface rdf:resourc
      <ndl:hasInterface rdf:resourc
      <ndl:hasInterface rdf:resourc
      <ndl:hasInterface rdf:resourc
      <ndl:hasInterface rdf:resourc
      <ndl:hasInterface rdf:resourc
```

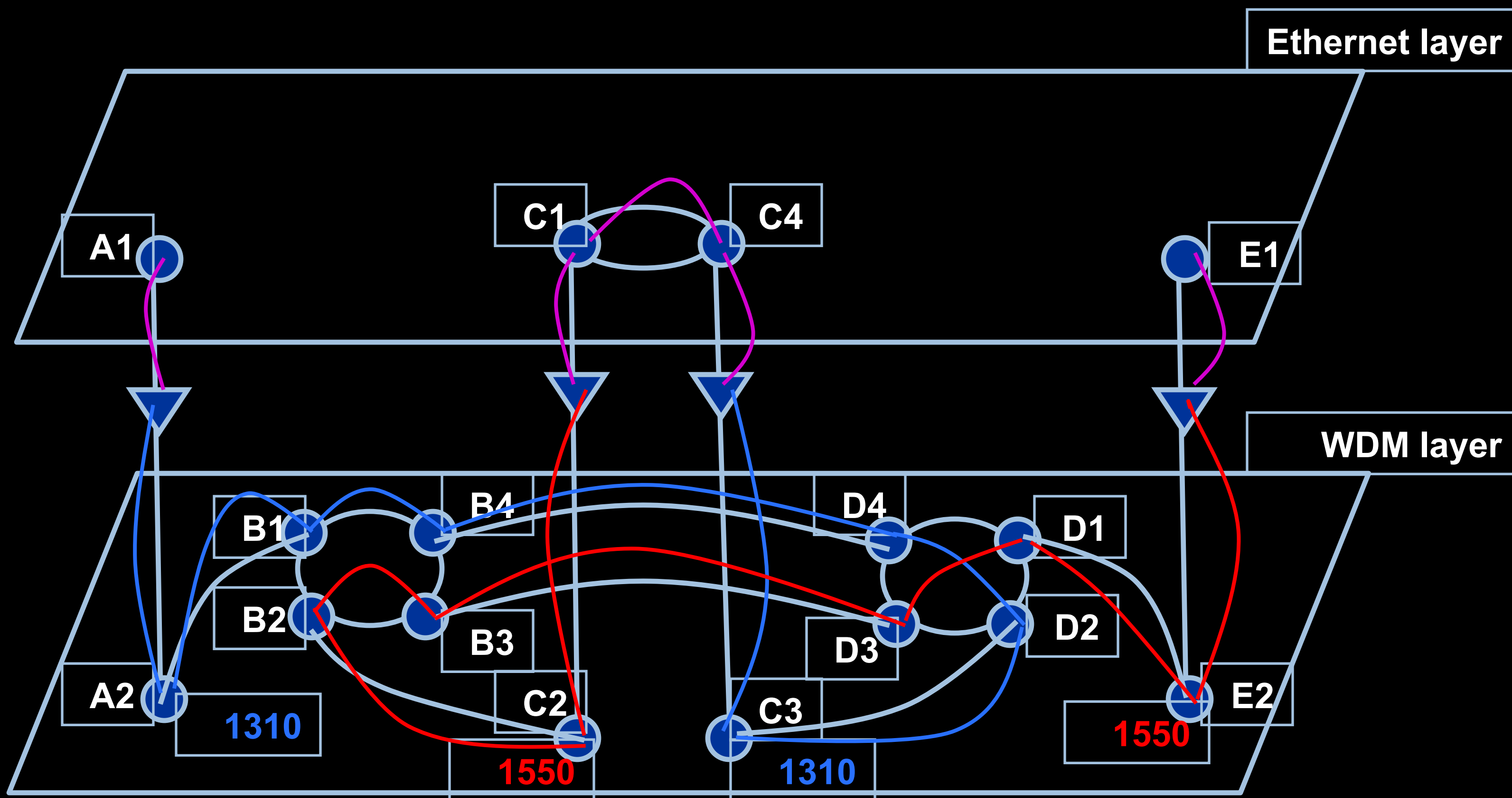
```
<!-- all the interfaces of TDM3.amsterdam1.netherlight.net -->

<ndl:Interface rdf:about="#tdm3.amsterdam1.netherlight.net:501/1">
<ndl:name>tdm3.amsterdam1.netherlight.net:POS501/1</ndl:name>
<ndl:connectedTo rdf:resource="#tdm4.amsterdam1.netherlight.net:5/1"/>
</ndl:Interface>
<ndl:Interface rdf:about="#tdm3.amsterdam1.netherlight.net:501/2">
<ndl:name>tdm3.amsterdam1.netherlight.net:POS501/2</ndl:name>
<ndl:connectedTo rdf:resource="#tdm1.amsterdam1.netherlight.net:12/1"/>
</ndl:Interface>
```

Multi-layer descriptions in NDL



Multi-layer Network PathFinding



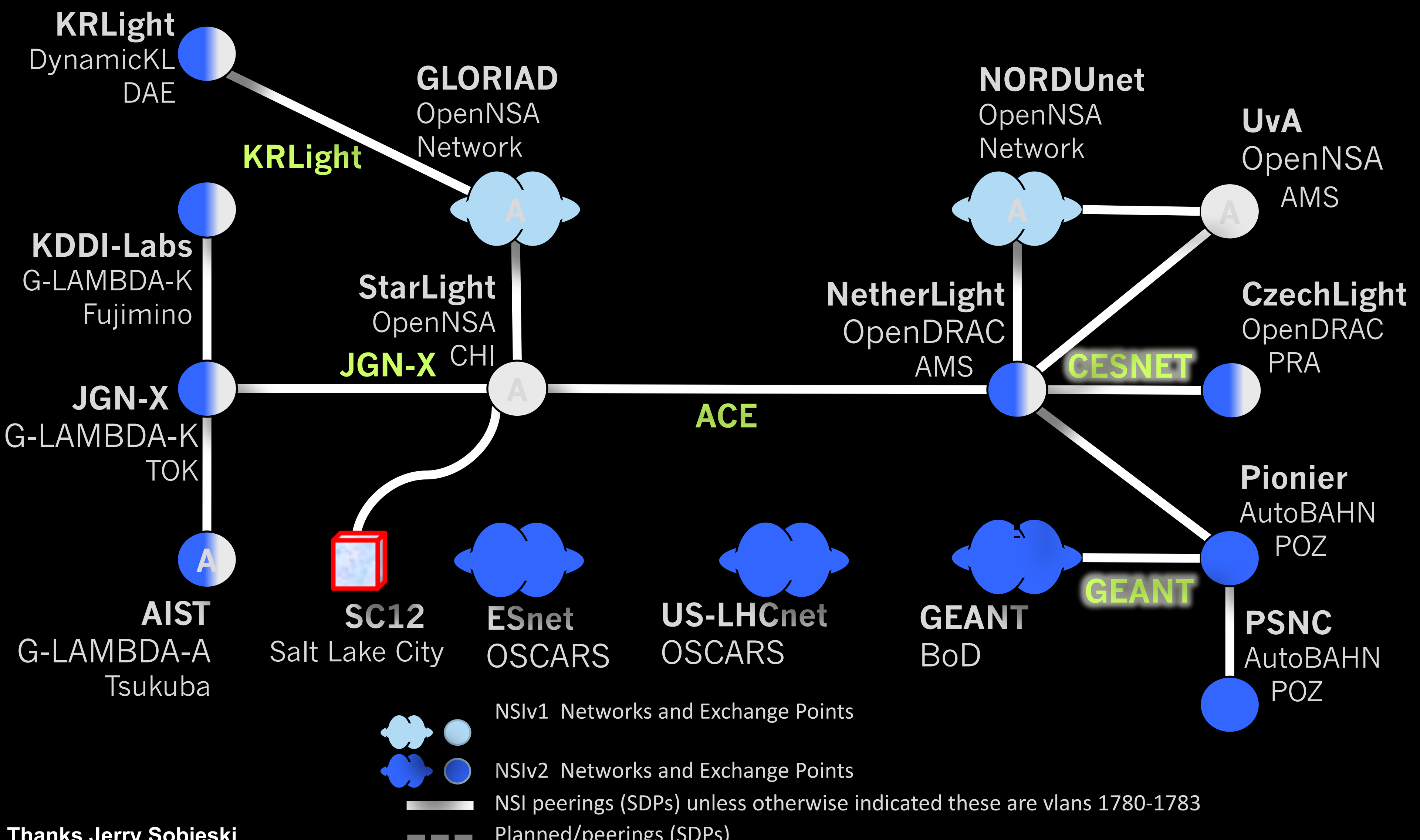
Path between interfaces A1 and E1:
A1-A2-B1-B4-D4-D2-C3-C4-C1-C2-B2-B3-D3-D1-E2-E1

Scaling: Combinatorial problem

Automated GOLE + NSI

Joint NSI v1+v2 Beta Test Fabric   Nov 2012

Ethernet Transport Service



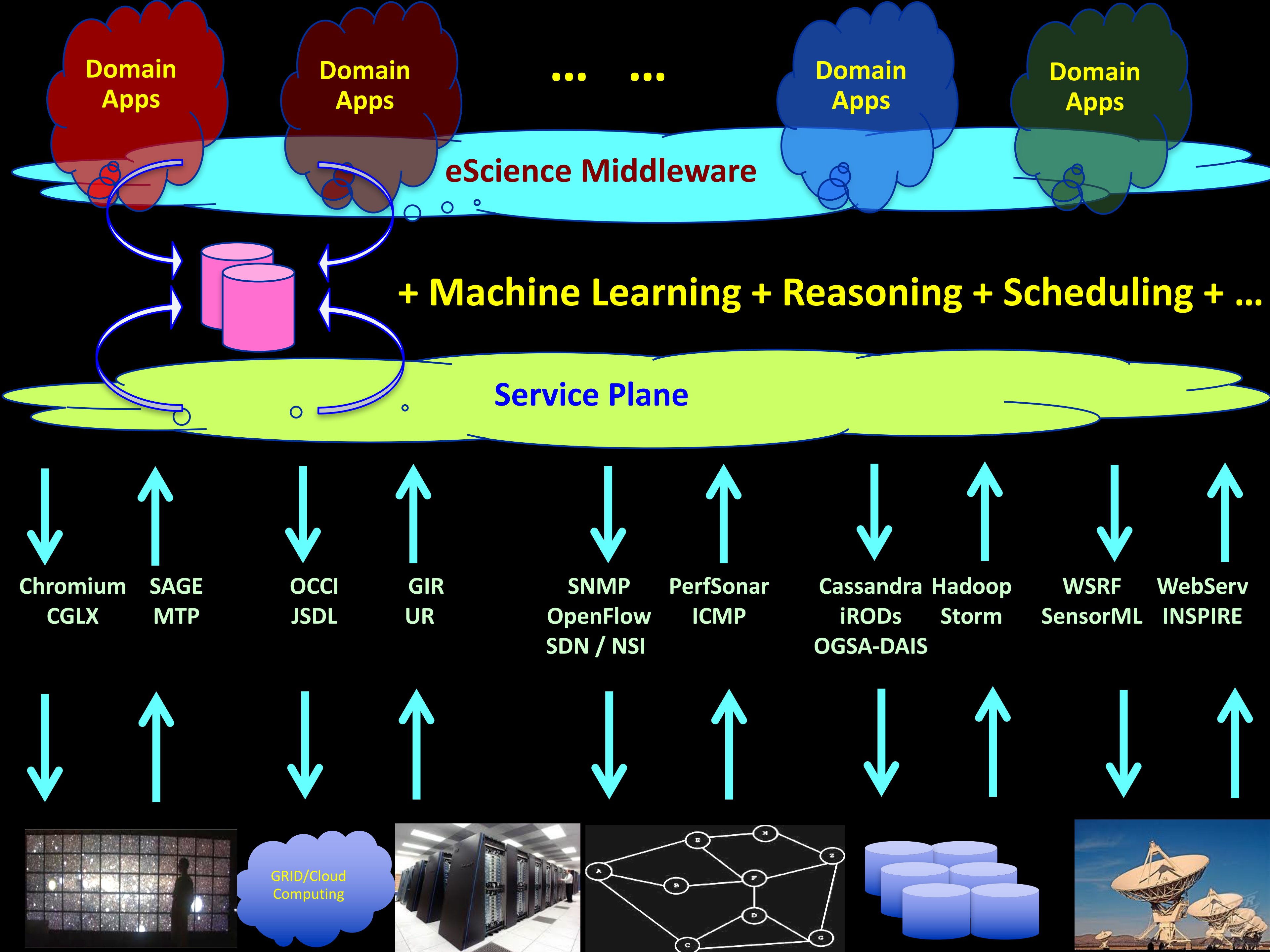
Thanks Jerry Sobieski

I want to →

"Show Big Bug Bunny in 4K on my Tiled Display using green Infrastructure"

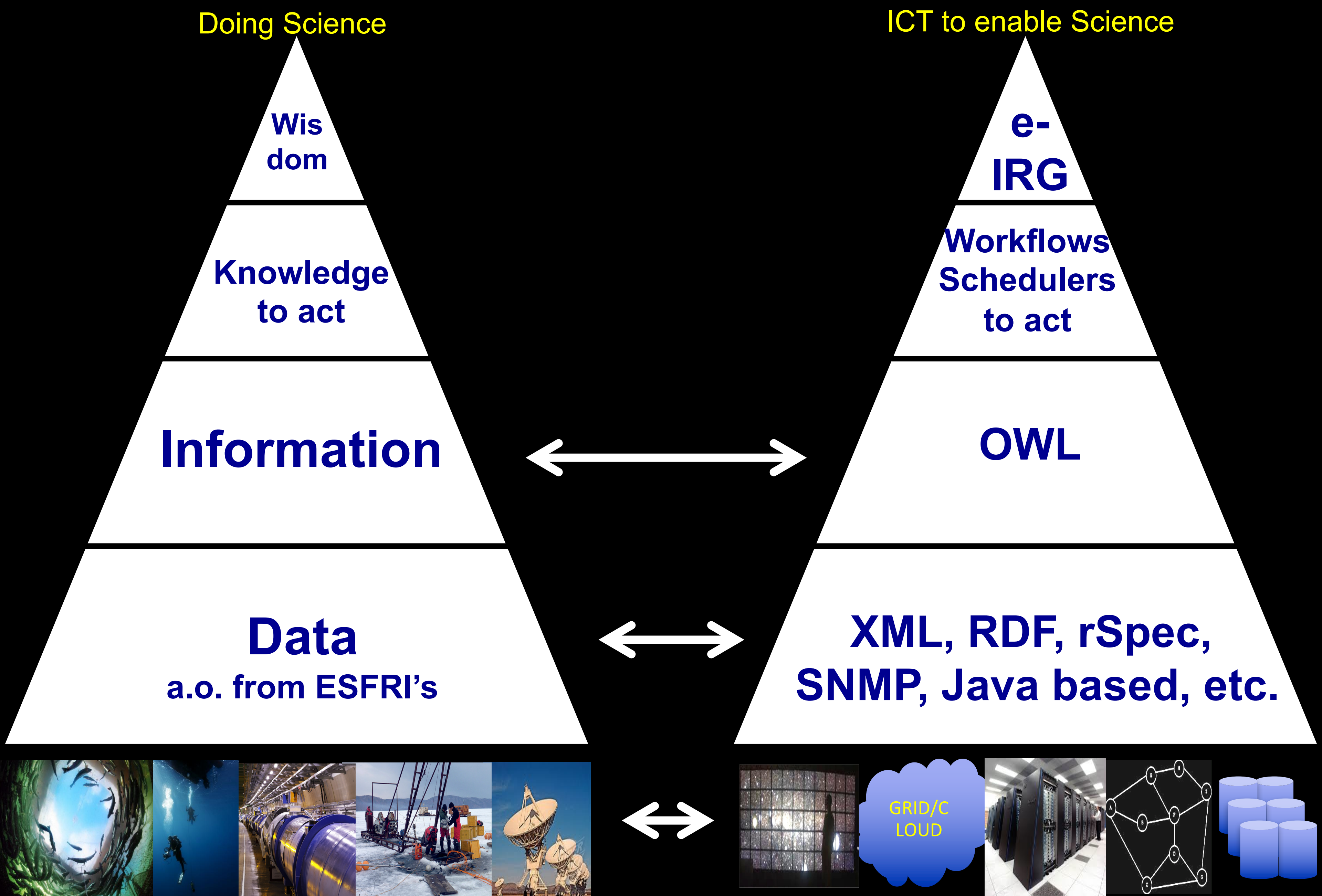
- Big Bugs Bunny can be on multiple servers on the Internet.
- Movie may need processing / recoding to get to 4K for Tiled Display.
- Needs deterministic Green infrastructure for Quality of Experience.
- Consumer / Scientist does not want to know the underlying details.

➔ His refrigerator also just works!

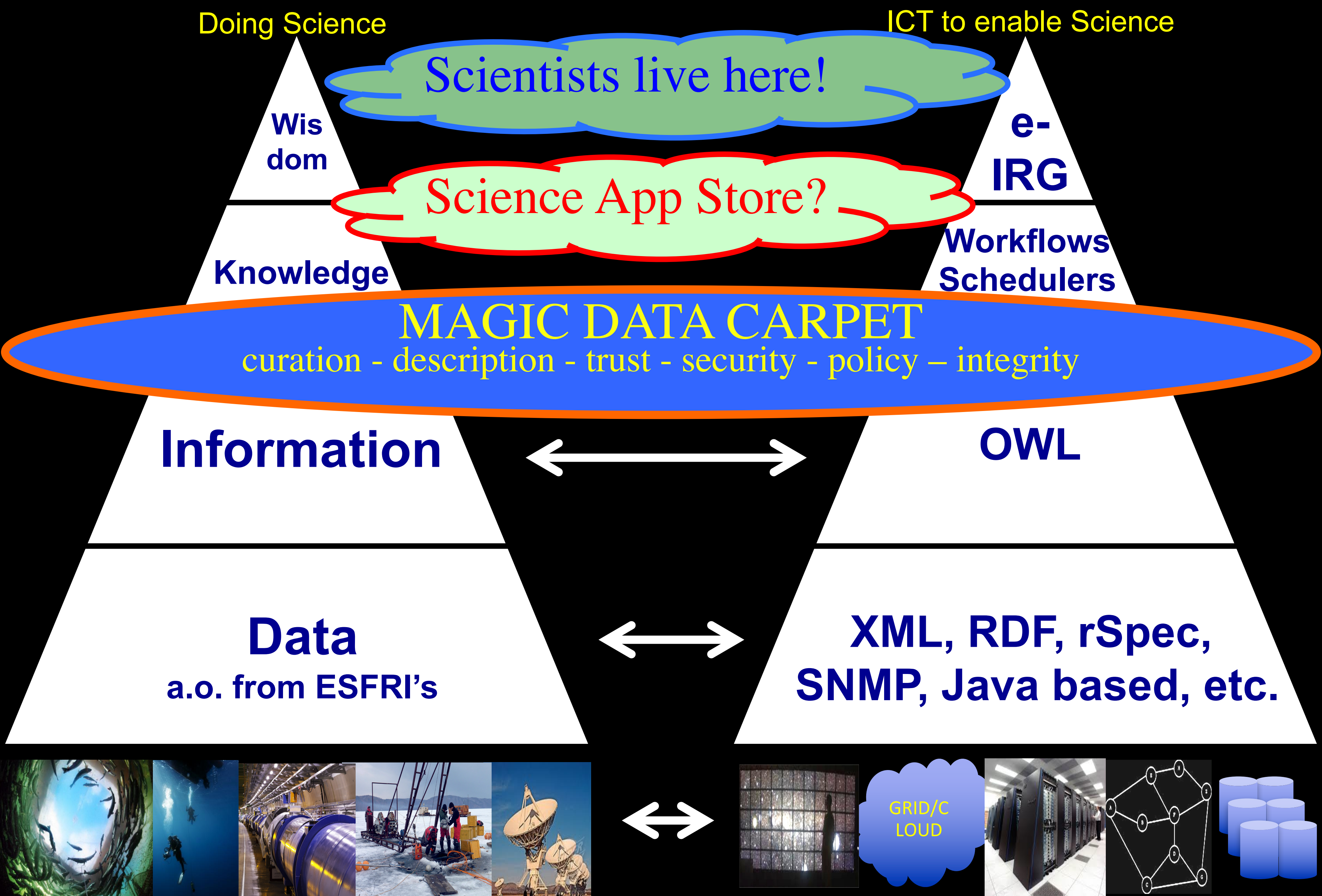


Domain Apps
Domain Apps
... ...
Domain Apps
Domain Apps
eScience Middleware
+ Machine Learning + Reasoning + Scheduling + ...
Service Plane
Chromium CGLX
SAGE MTP
OCCI JSDL
GIR UR
SNMP OpenFlow SDN / NSI
PerfSonar ICMP
Cassandra iRODs OGSA-DAIS
Hadoop Storm
WSRF SensorML
WebServ INSPIRE
GRID/Cloud Computing

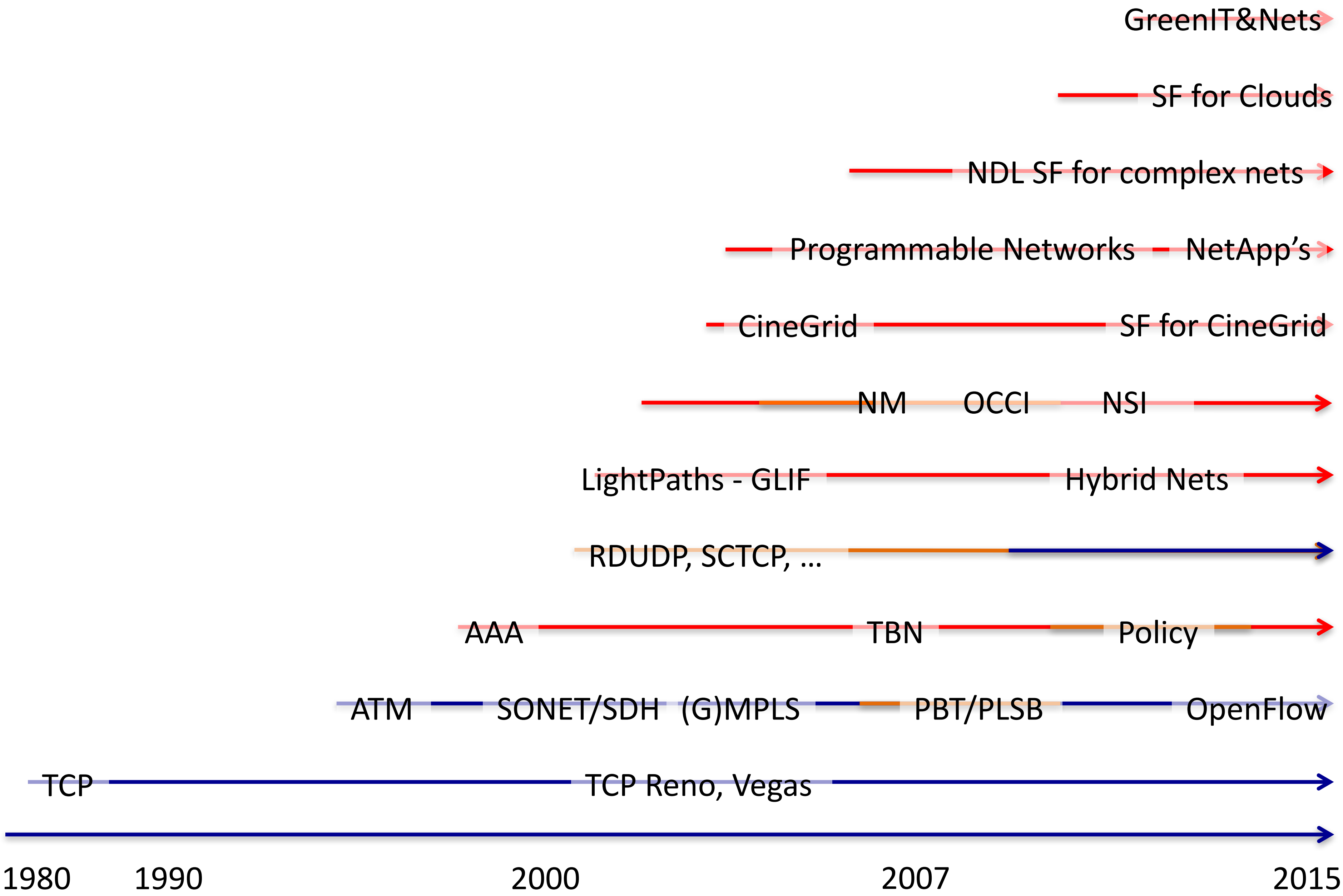
The Big Data Challenge



The Big Data Challenge

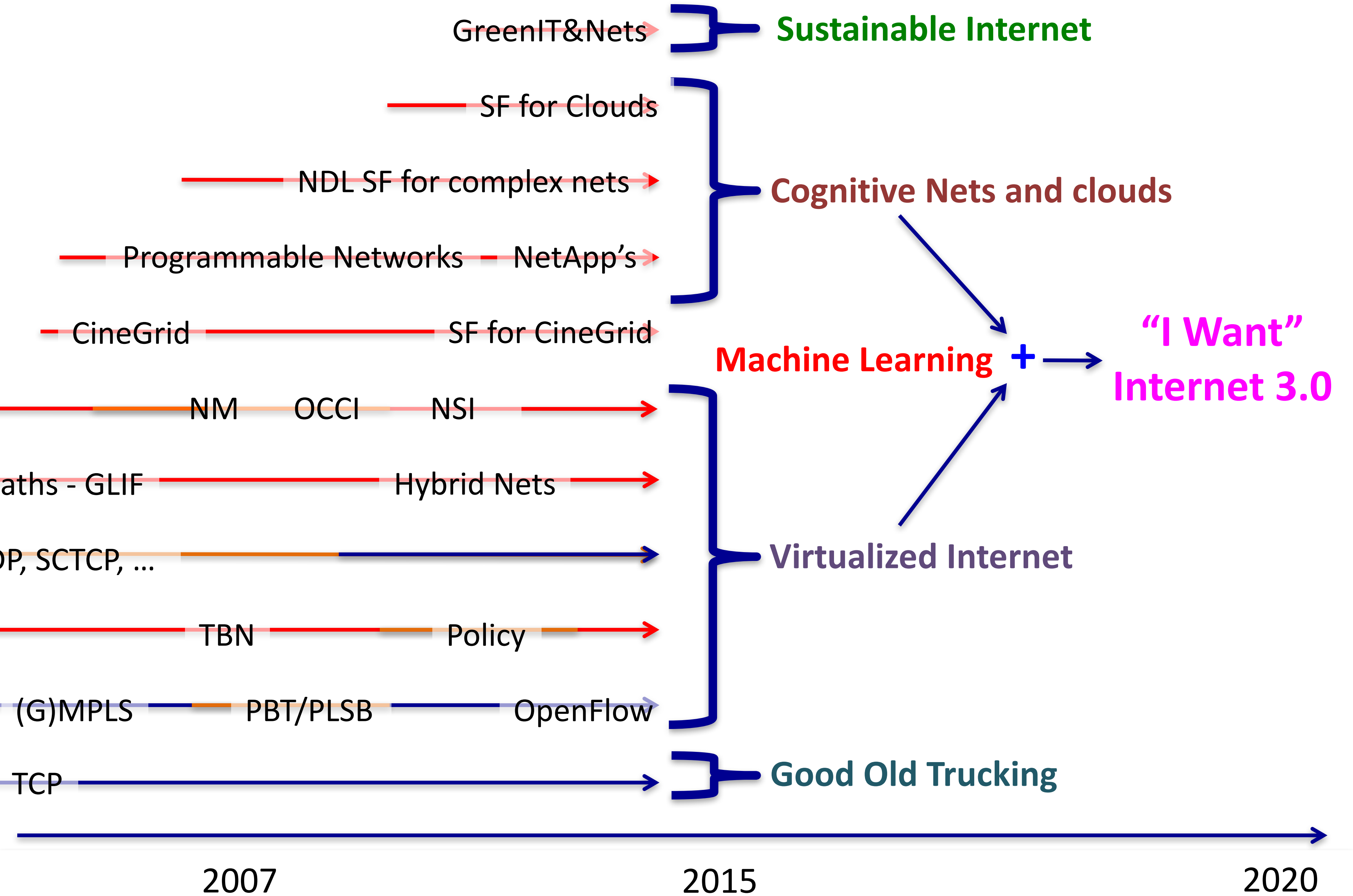


TimeLine



GreenIT&Nets
SF for Clouds
NDL SF for complex nets
Programmable Networks
NetApp's
CineGrid
SF for CineGrid
NM
OCCI
NSI
LightPaths - GLIF
Hybrid Nets
RDUDP, SCTCP, ...
AAA
TBN
Policy
ATM
SONET/SDH
(G)MPLS
PBT/PLSB
OpenFlow
TCP
TCP Reno, Vegas
1980
1990
2000
2007
2015

TimeLine


GreenIT&Nets
Sustainable Internet
SF for Clouds
NDL SF for complex nets
Programmable Networks
NetApp's
Cognitive Nets and clouds
CineGrid
SF for CineGrid
Machine Learning
+
"I Want"
Internet 3.0
NM
OCCI
NSI
aths - GLIF
Hybrid Nets
P, SCTCP, …
Virtualized Internet
TBN
Policy
(G)MPLS
PBT/PLSB
OpenFlow
TCP
Good Old Trucking
2007
2015
2020

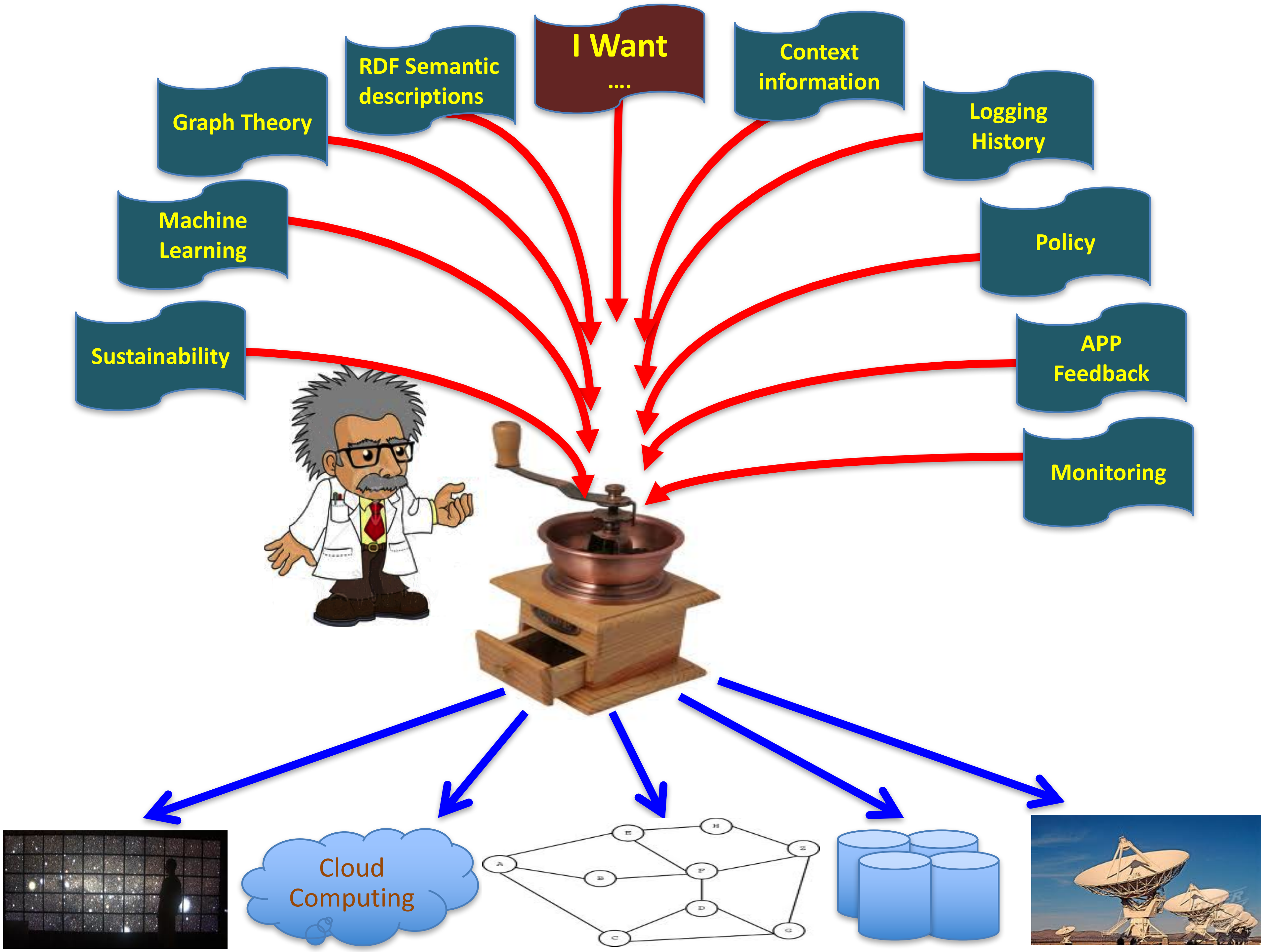
TimeLine

- Sustainable Internet
- Cognitive Nets and clouds
- lachine Learning + → "I Want" Internet 3.0
- Virtualized Internet
- Good Old Trucking

I retire

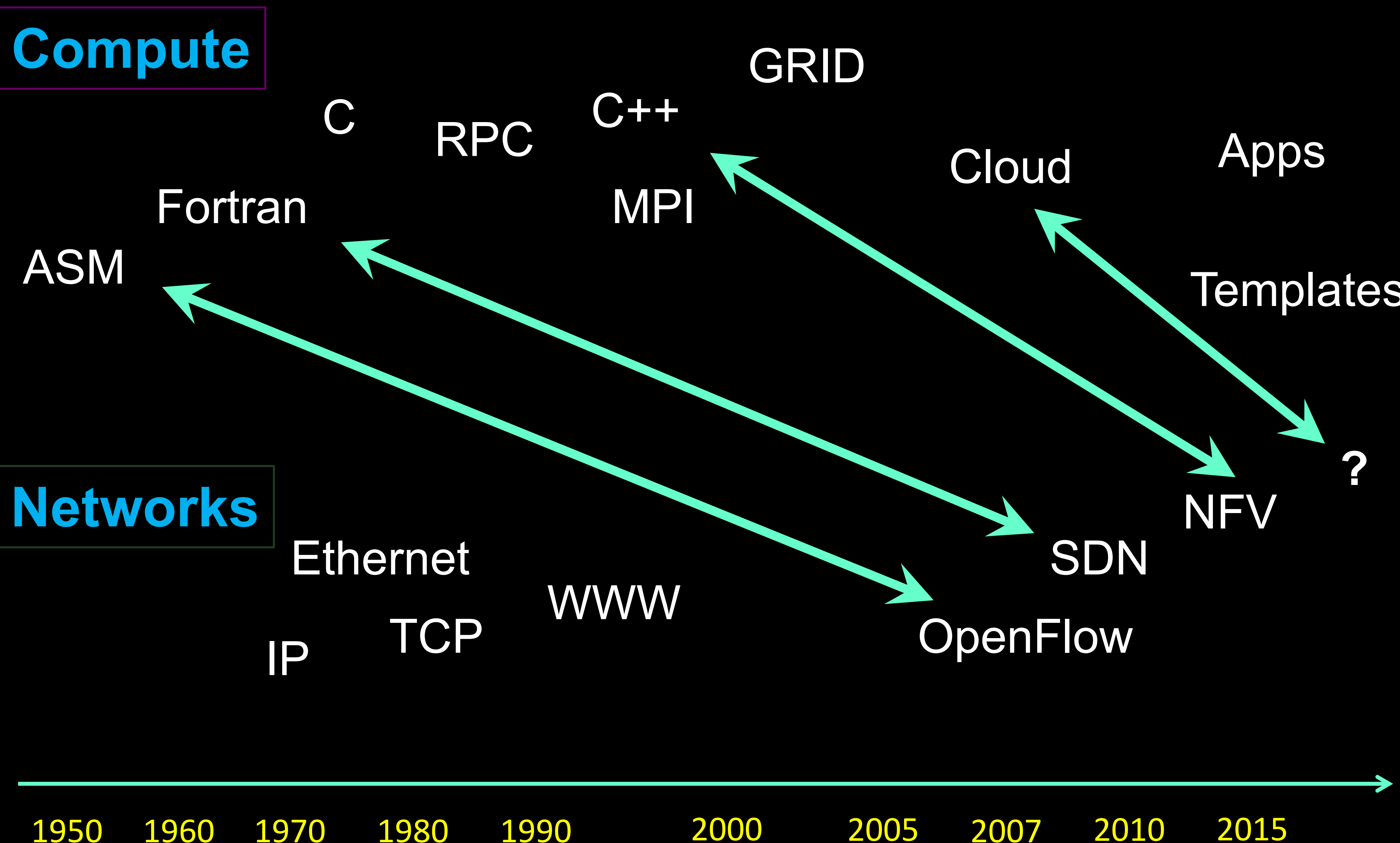
2020

2040



I Want ….
RDF Semantic descriptions
Context information
Graph Theory
Logging History
Machine Learning
Policy
Sustainability
APP Feedback
Monitoring
Cloud Computing

TimeLine



Compute
GRID
C
C++
RPC
Apps
Cloud
Fortran
MPI
ASM
Templates
?
NFV
Networks
SDN
Ethernet
WWW
OpenFlow
TCP
IP
1950
1960
1970
1980
1990
2000
2005
2007
2010
2015

Why?

Because we can!

Paper #1 + Q's

TRANSLIGHT

A GLOBAL-SCALE LAMBDAGRID FOR E-SCIENCE

This global experiment wants to see if high-end applications needing transport capacities of multiple Gbps for up to hours at a time can be handled through an optical bypass network.

Tom DeFanti, Cees de Laat, Joe Mambretti, Kees Neggers, Bill St. Arnaud.

Communications of the ACM, Volume 46, Issue 11 (November 2003), Pages: 34 – 41.

http://delaat.net/pubs/2003-j-6.pdf

Paper #1 + Q's

- Q1: This article is now 10 years old. Back then Twitter did not exist. What do you think will be the drivers for network capacity demand in Science and Society 10 years from now?

- Q2: List arguments why one would use photonic networks directly in science applications and arguments why not tu use photonics directly but use current Internet.

- Q3: This question is not directly from this paper but fun to figure out via search on the web: Fiber cable systems under the ocean are very expensive and cost 100's of millions to put in place. How many fibers do they put in one cable and why that amount?

Paper #2 + Q's

Seamless Live Migration of Virtual Machines over the MAN/WAN

F. Travostino, P. Daspit, L. Gommans, C. Jog, C.T.A.M. de Laat, J. Mambretti, I. Monga, B. van Oudenaarde, S. Raghunath and P.Y. Wang

Future Generation Computer Systems, Volume 22, Issue 8, October 2006, Pages 901-907.

http://delaat.net/pubs/2006-j-5.pdf

Paper #2 + Q’s

- Q1: When migrating VM’s as described in the paper, what are the related network connectivity challenges for the running VM's?

- Q2: Nowadays VM migration and load balancing are more or less standard in cloud environments. List a number of modern similar functionalities in current cloud providers and compare features among those and the method described in the paper.

- Q3: List a number of applications that could take advantage of the described migration methods. Do not only list what the paper described more than ten years ago before the word cloud was applied in the computer science, but take into consideration current technologies and trends, such as IOT, Smart City, autonomous driving cars,

The constant factor in our field is Change!

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

"Fortran goto", Unix, c, SmallTalk, DECnet, TCP/IP, c++, Internet, WWW, Semantic Web, Photonic networks, Google, grid, cloud, Data^3, App

to:

DDOS attacks destroying Banks and Bitcoins.

Conclusion:

Need for Safe, Smart, Resilient Sustainable Infrastructure.

CHANGE!