CERN Advanced Networking

CERN, 18th July 2014

edoardo.martelli@cern.ch

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Summary

- IT-CS
- CERN networks
- LHC Data Challenge
- WLCG
- LHCOPN and LHCONE
- Openlab
- Conclusions

IT-CS
Communication systems

IT-CS

The IT-CS group is responsible for all communication services in use at CERN for data, voice and video

http://information-technology.web.cern.ch/about/organisation/communication-systems

IT-CS organization

Communication Systems		
Group Leader	**Tony Cass**	
Deputy Group Leader	**Frederic Chapron**	

Communications Deployment CD
Section Leader
John Shade
Leszek Borakiewicz
Stephane Casenove
Marc Collignon
Maryse da Costa
Daniel Francart
Mathieu Pierre Largeron
Federico Lerda
Roland Schuster
Mohssen Souayah
Loic Vittoz
Baptiste Wentzler

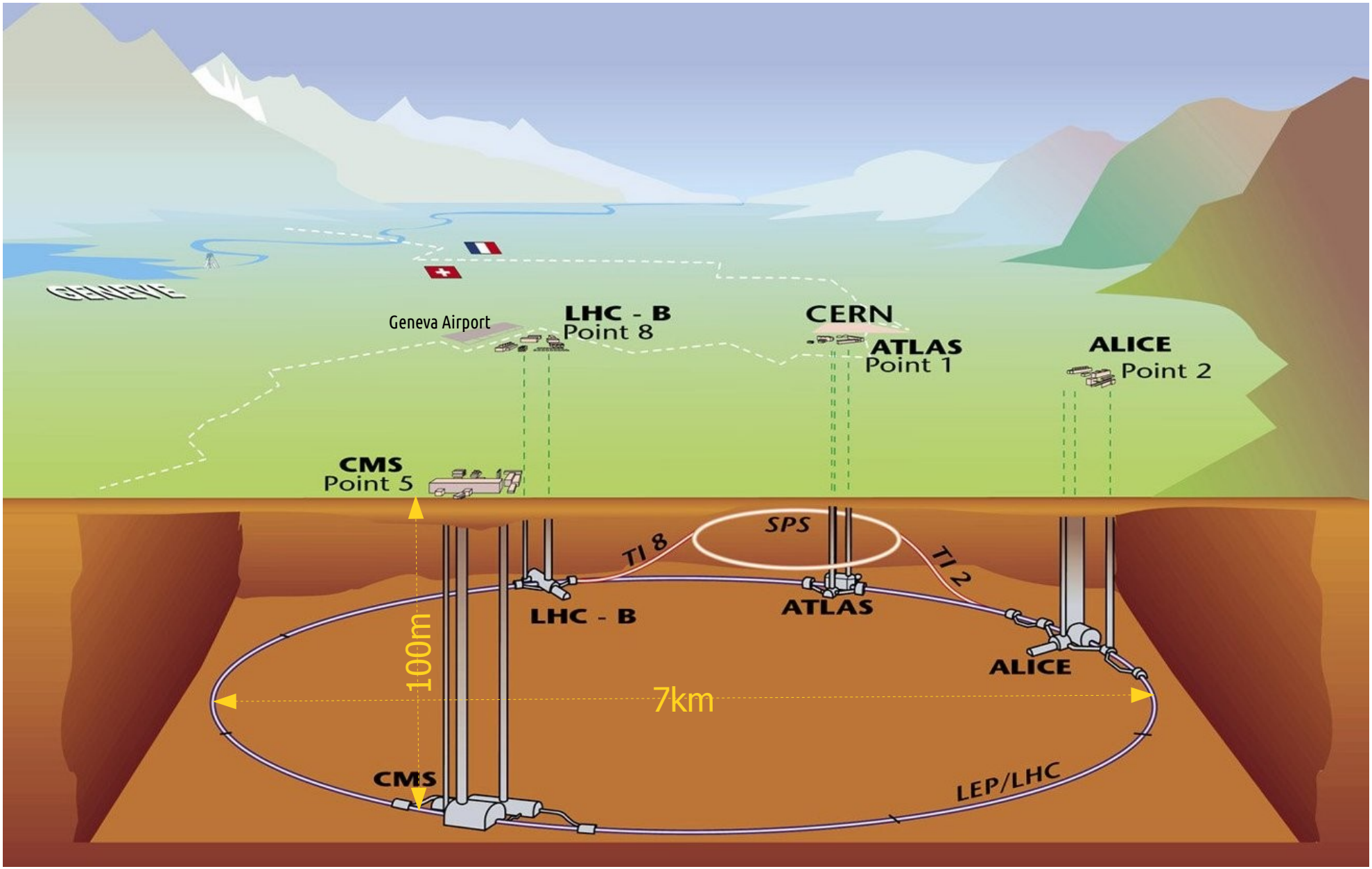
Communications Engineering CE
Section Leader
Edoardo Martelli
Juan Cristobal Boullosa
Sebastien Ceuterickx
Anastasios Fousekas
David Gutierrez Rueda
Carles Kishimoto Bisbe
Mortadha Jouini
Veronique Lefebure
Virginie Longo
Eric Sallaz
Dan Savu
Adam Wojciech Sosnowski
Stefan Nicolae Stancu
Caltech team (Atrur Barczyk)

Communications Support CS
Section Leader
Frederic Chapron
Stefano Agosta
Marta Antosik
Gyorgy Balazs (op manager)
Alberto Garcia Molero
Aurelie Pascal
Jean-Pierre Puget
Anna Raczynska-Tartivel
Rodrigo Sierra Moral
Francisco Valentin Vinagrero

Communications Tools CT
Section Leader
Jose Carlos Luna Duran
Raul Garcia Martinez
Alexandru Doru Grosu
Milosz Marian Hulboj
Marwan Khelif
Marek Pomocka
Miguel Filipe Santos Pinto

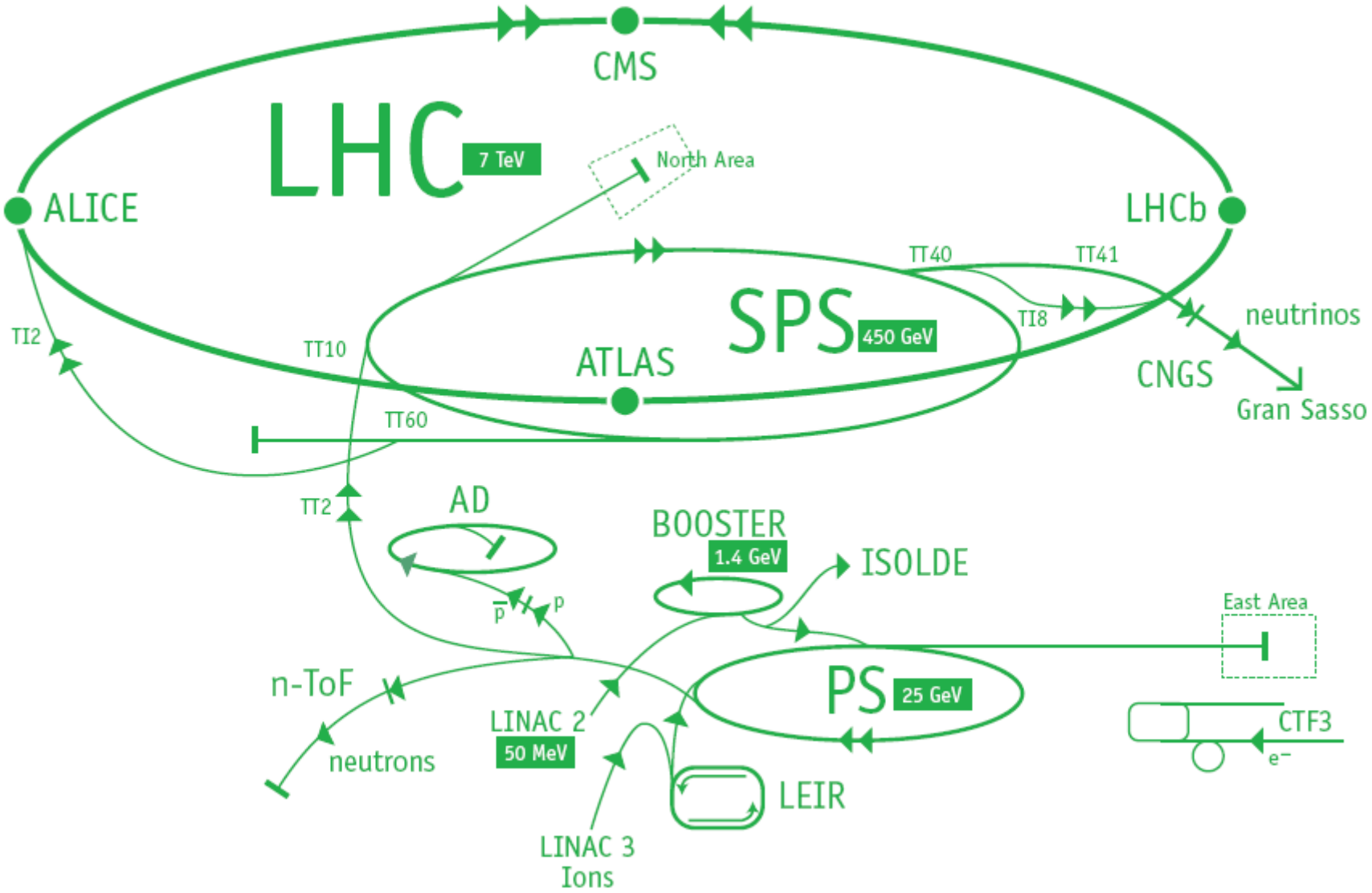
CERN

CERN Location Points

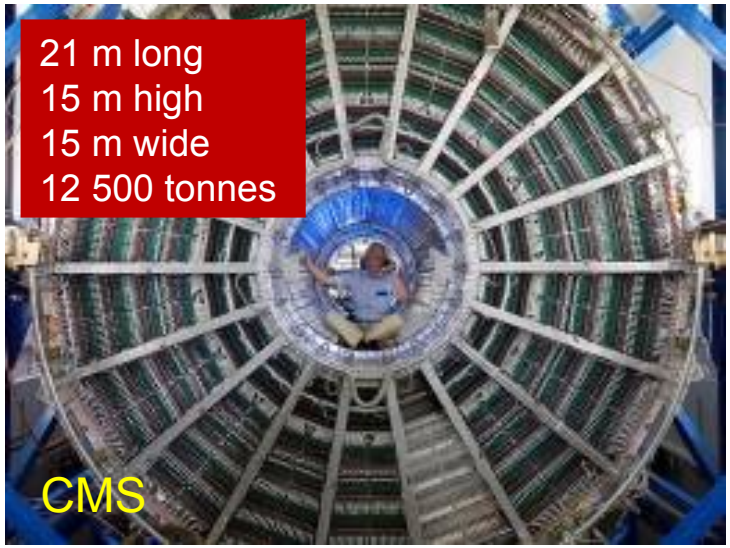


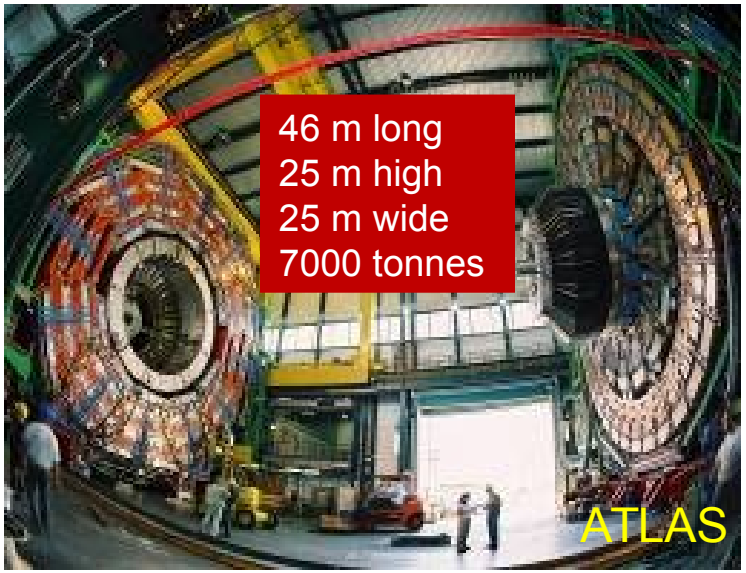
CERN bird's eye view

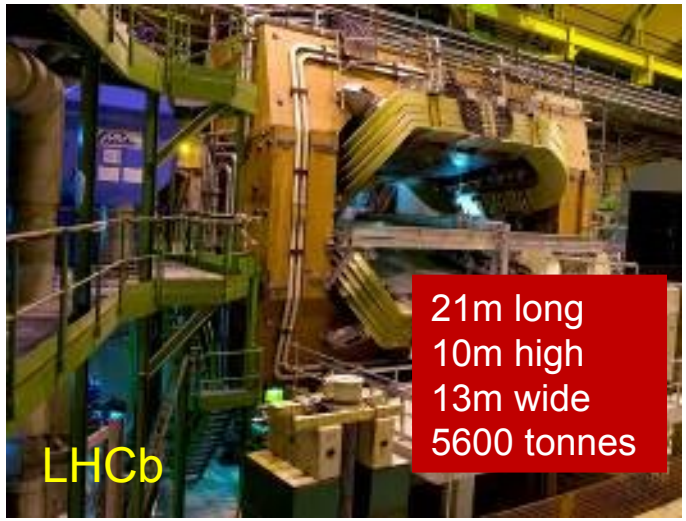
CERN accelerator complex

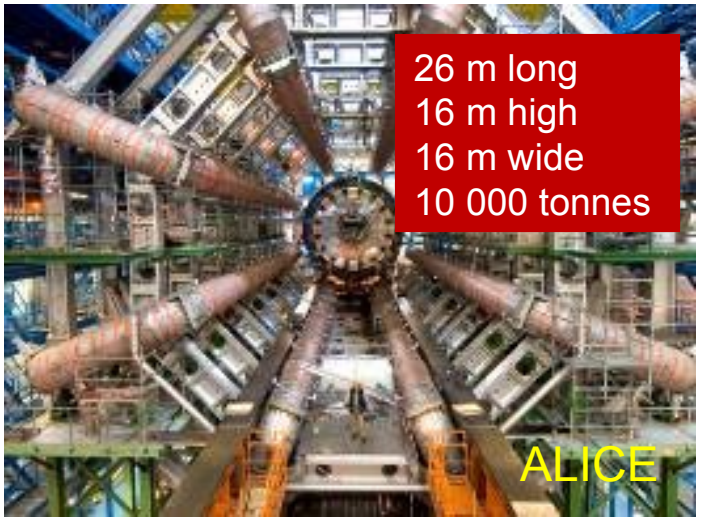


LHC Experiments









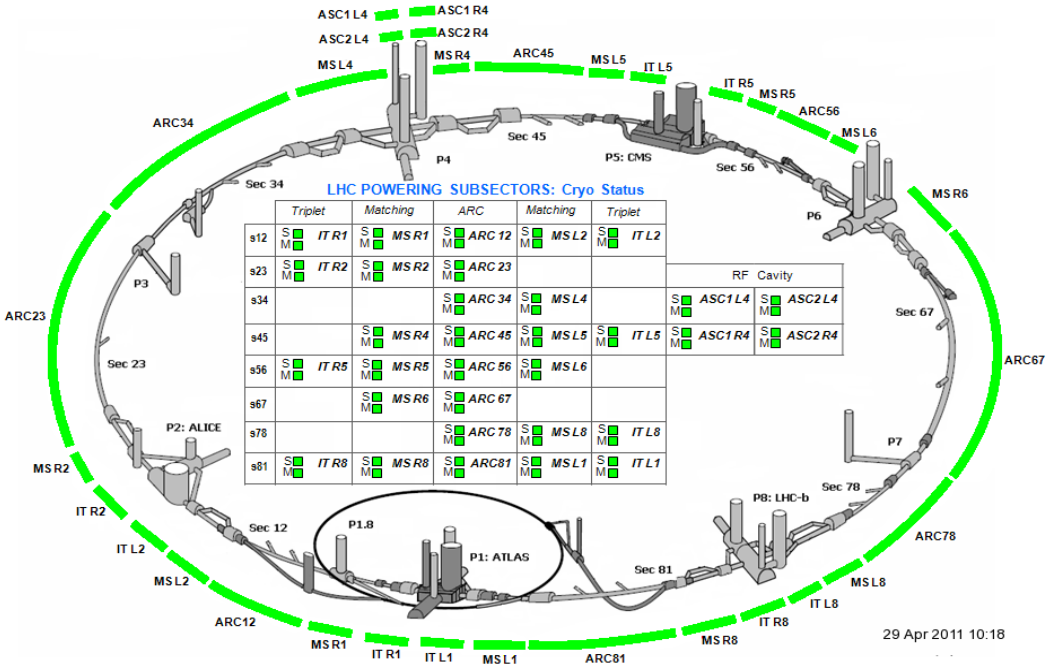
Networks at CERN

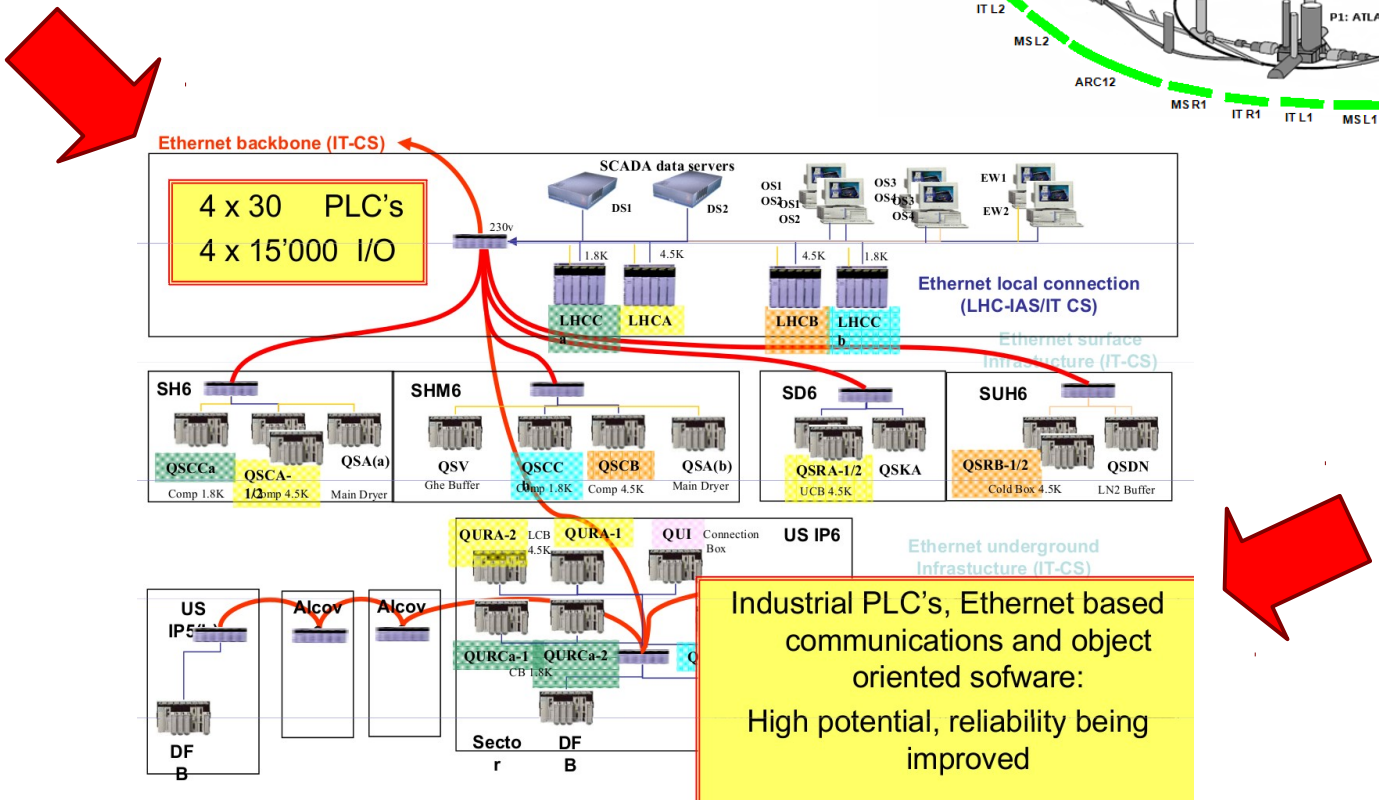
High Energy Physics over IP

Most of the CERN infrastructure is controlled and managed over a pervasive **IP network**

Cryogenics

27Km of pipes at -271.11° C by means of 700.000 litres of Helium: **controlled over IP**

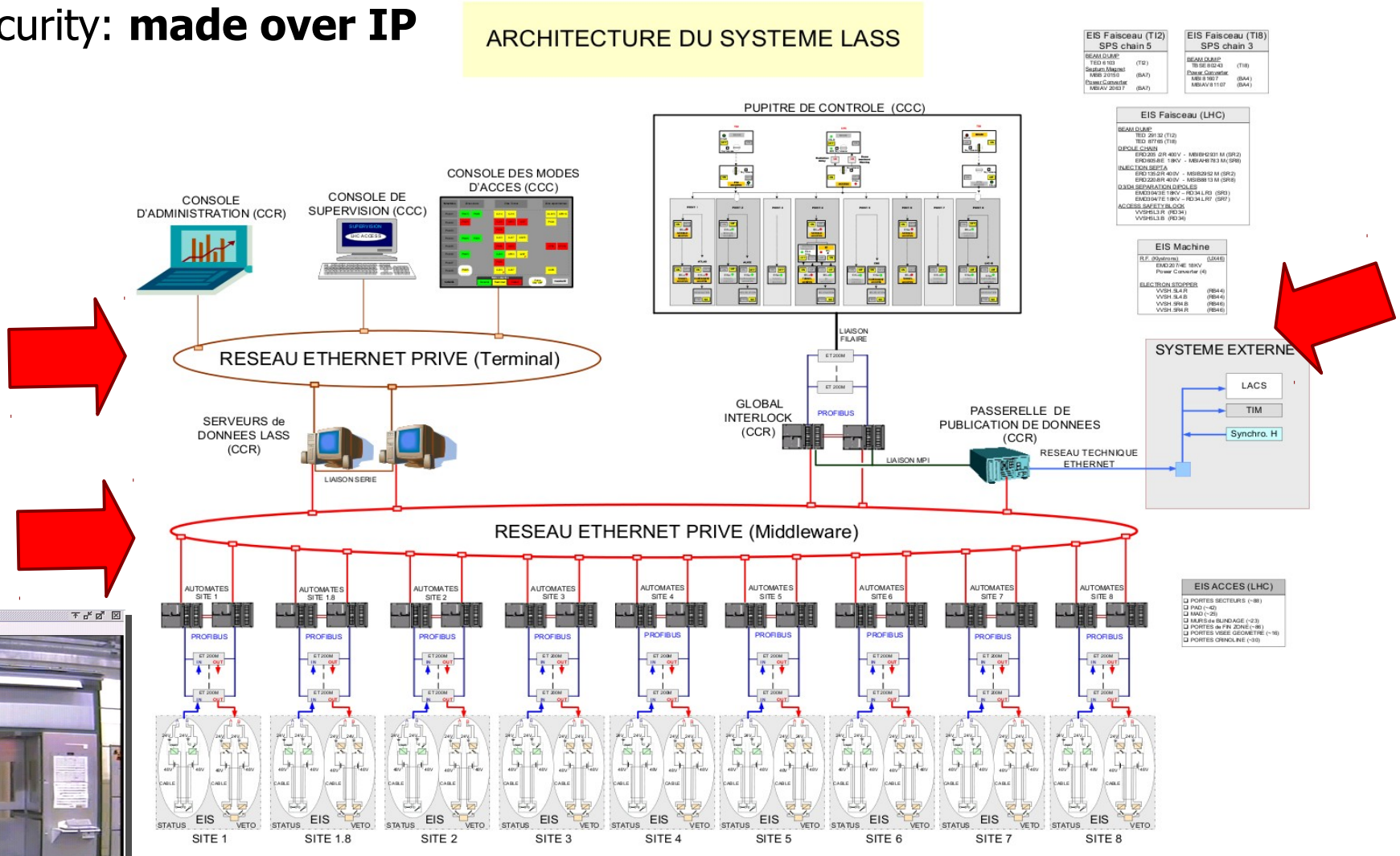


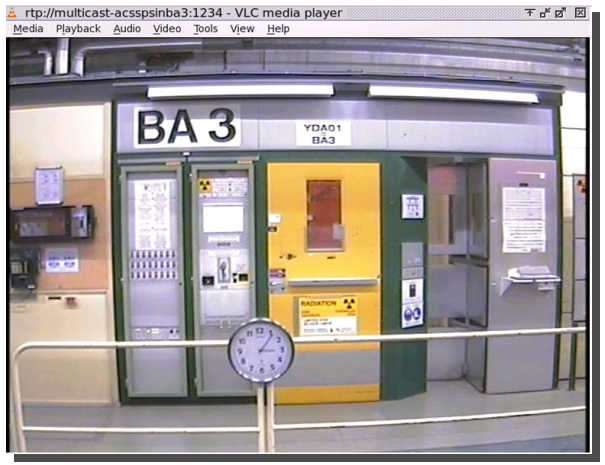


Source: http://te-dep-crg-oa.web.cern.ch/te-dep-crg-oa/te-crg-oa_fichiers/cryolhc/LHC%20Cryo_BEOP_lectures2009.pdf

Access control

Safety and Security: **made over IP**





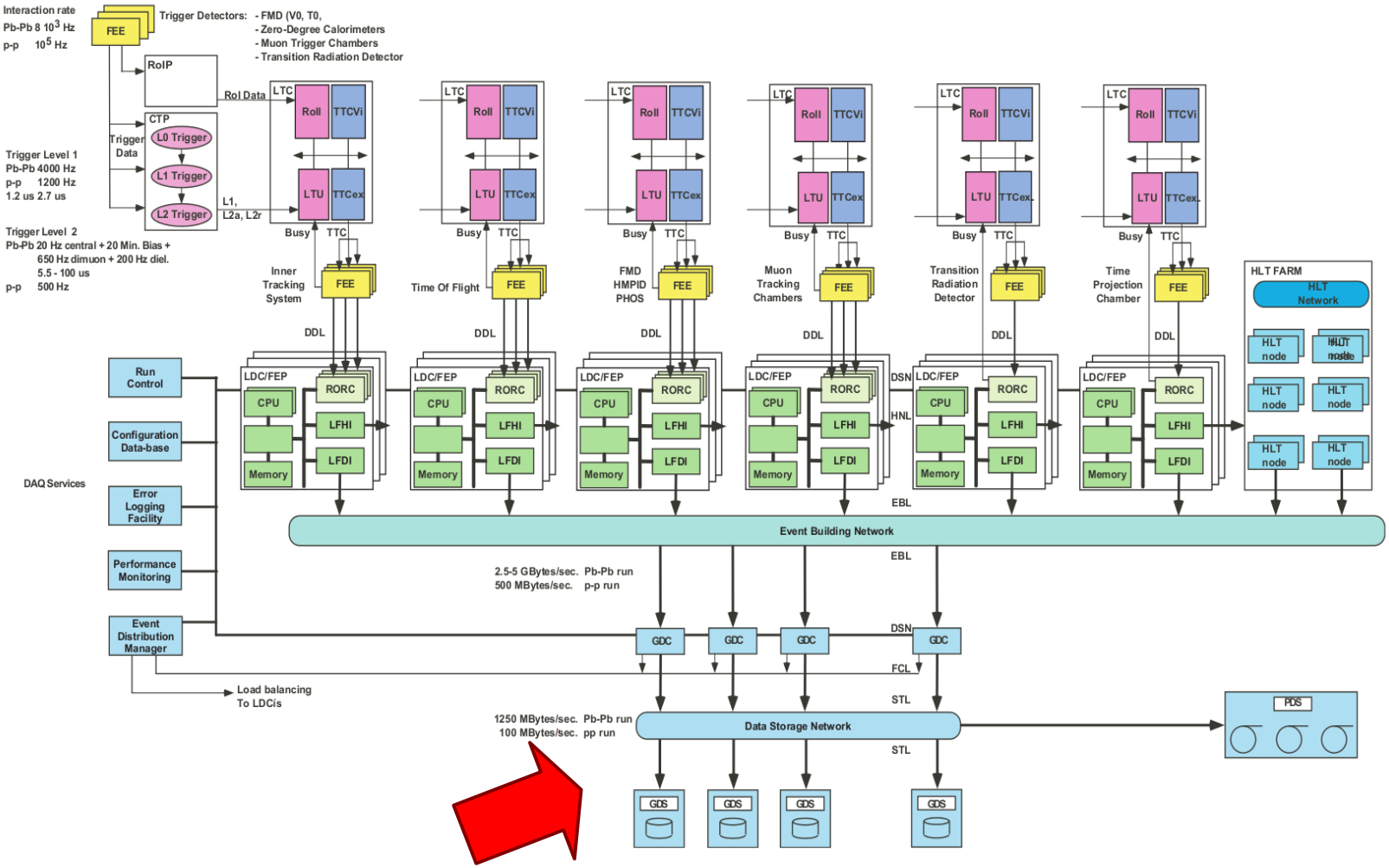
Source:https://edms.cern.ch/file/931641/1/LASS-LACS_IHM.pdf

Remote inspections

Remote inspection of dangerous areas: robots controlled and giving feedback over **WiFi and GSM IP networks**

DAQ: Data Acquisition

A constant stream of data from the four Detectors to disk storage

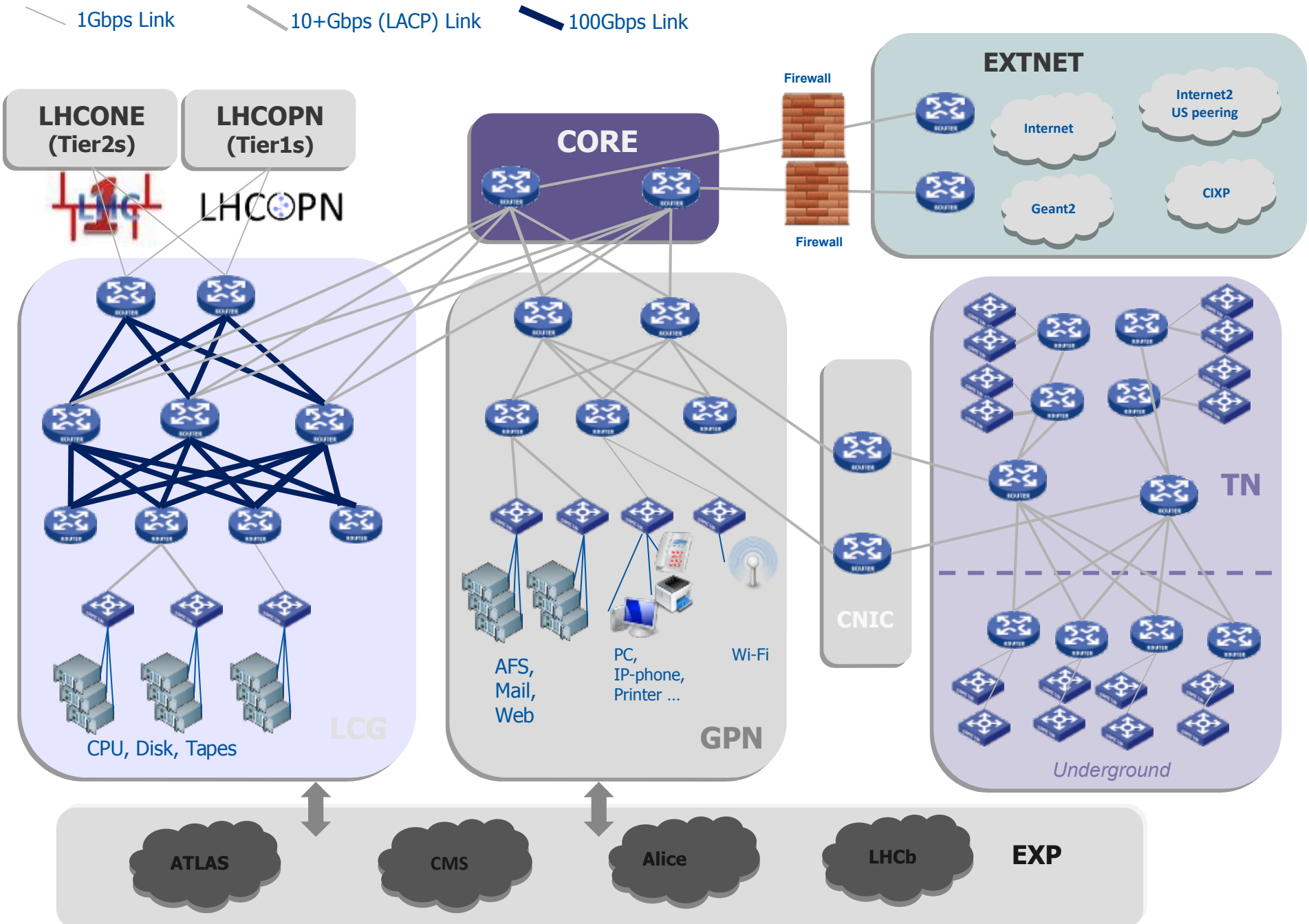


Source: http://aliceinfo.cern.ch/Public/Objects/Chapter2/DetectorComponents/daq_architecture.pdf

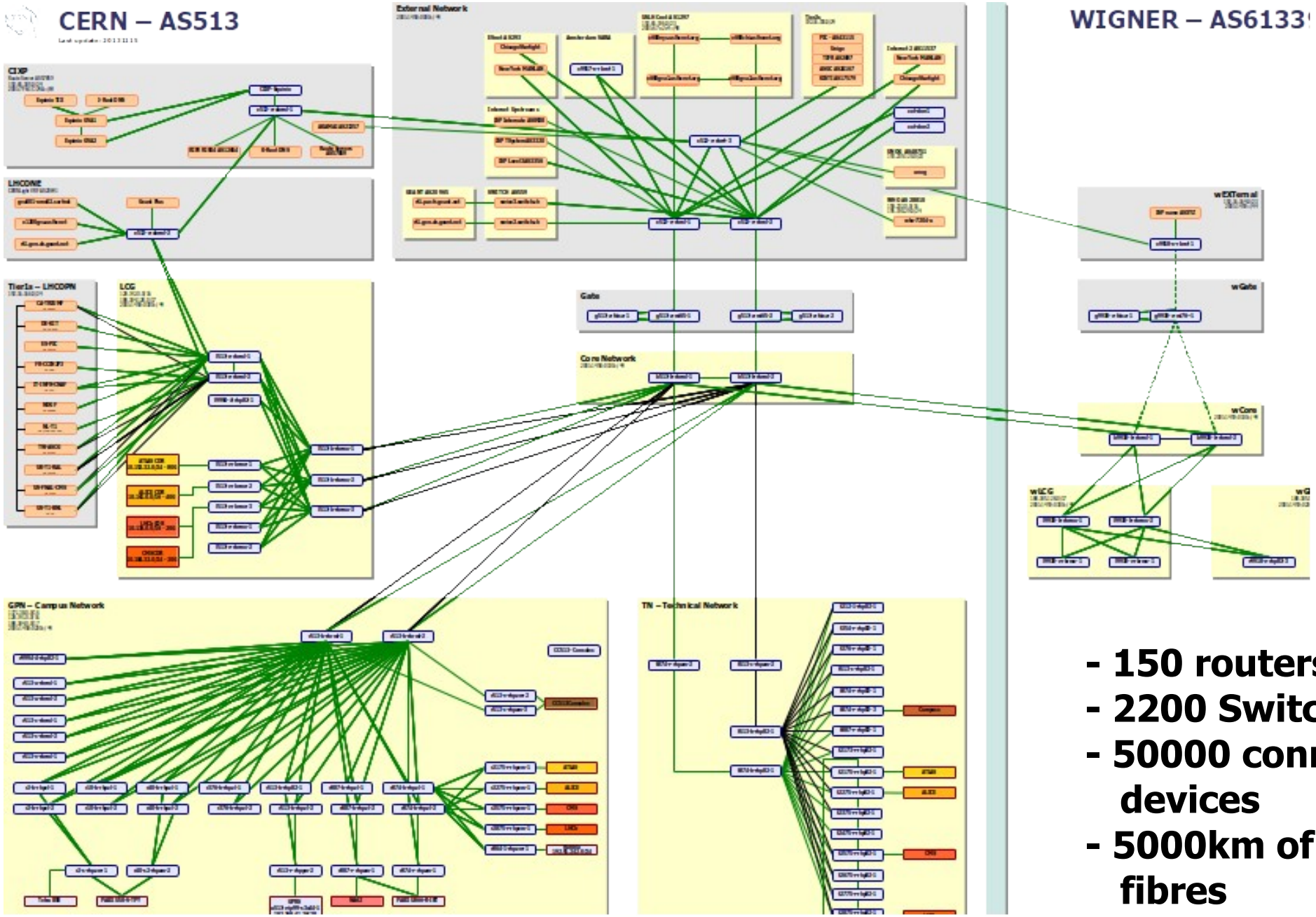
CCC: CERN Control Centre

The nerve centre of the particle accelerator: over IP

CERN data network: Domains



CERN data network: details



- 150 routers
- 2200 Switches
- 50000 connected devices
- 5000km of optical fibres

WIFI surface network

802.11b/g/a/n 2.4 and 5 GHz

Networks for general staff and Eduroam

>1250 WIFI Access Points deployed (June 2014)

300 new Access Points deployed every year

4500/7000 (avg/max) devices connected every day

>300 buildings and >250000 m2 covered

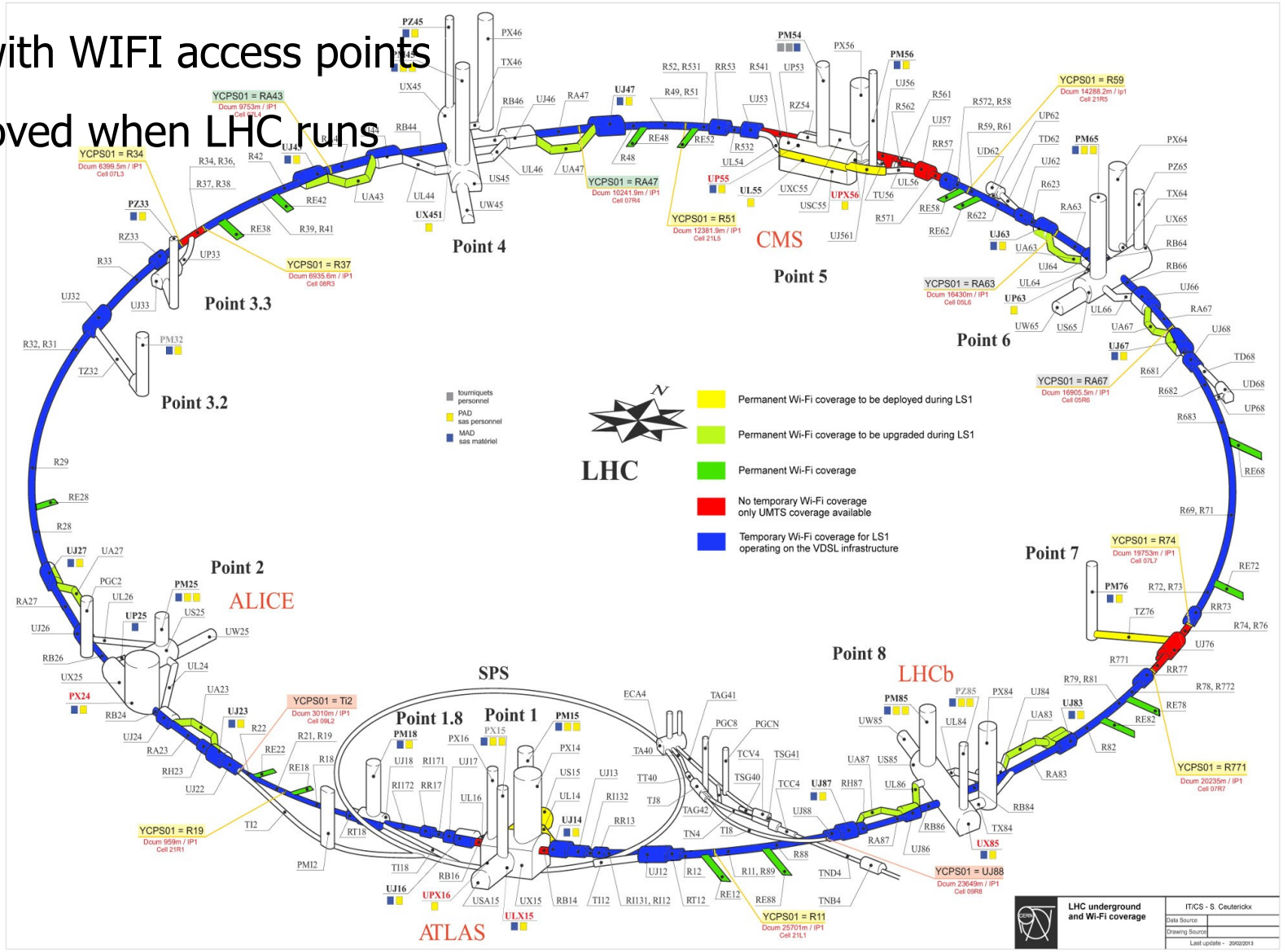
VDSL and WIFI in the LHC tunnel

VDSL network to connect WIFI Access point in the LHC tunnel

VDSL plugs every 100m in LHC tunnel

~280 VDSL modems with WIFI access points

Modems and APs removed when LHC runs

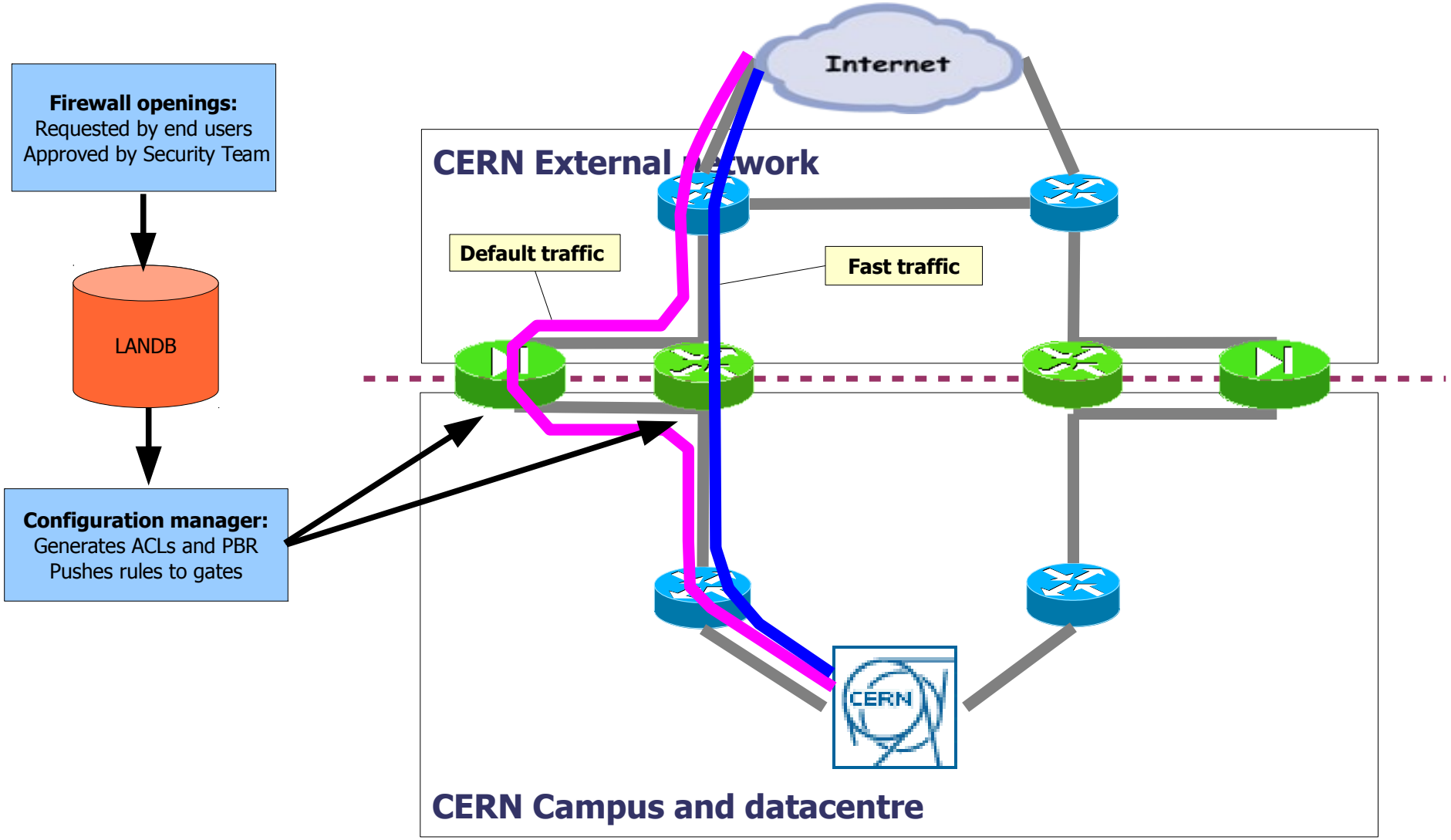


SDN Firewall

Firewall rules stored in network database

Routers and Firewall configuration updated very 15 minutes

Statefull firewall bypass for large, well known flows



IPv6

IPv6 dual stack network deployment completed in March 2014!

More information:

http://cern.ch/ipv6

Fixed Telephony

4 PBX and 15 IP-PBXs

12000 fixed users (10000 traditional + 2000 IP)

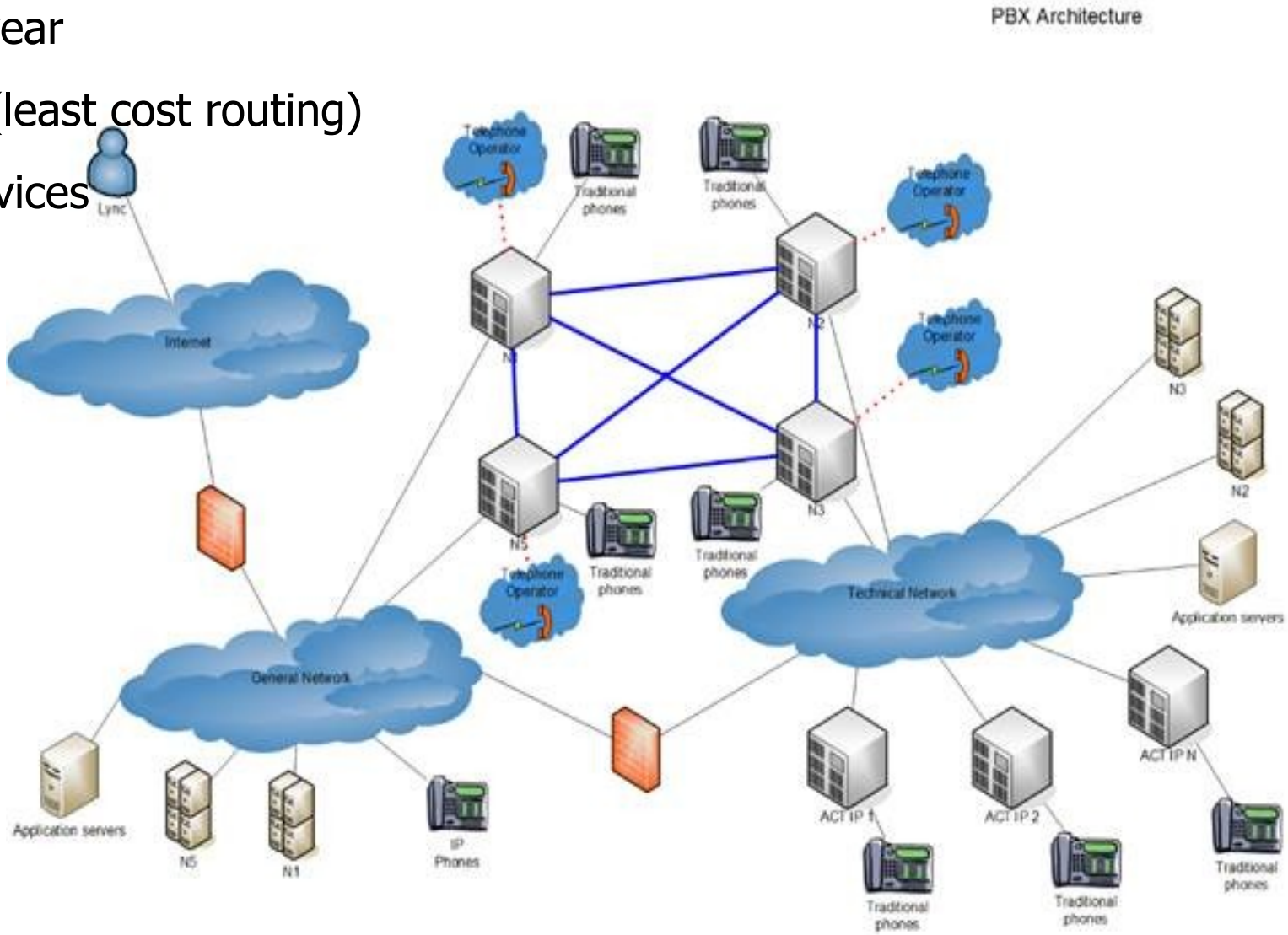
2.5M outgoing calls/year

5 telecom operators (least cost routing)

Audio conference services

IP telephony services

Emergency phones



Mobile Telephony

Single contract with our current Swiss operator (Sunrise)

2G & 3G coverage (LTE being tested)

Closed VPN comprising around 6000 mobiles

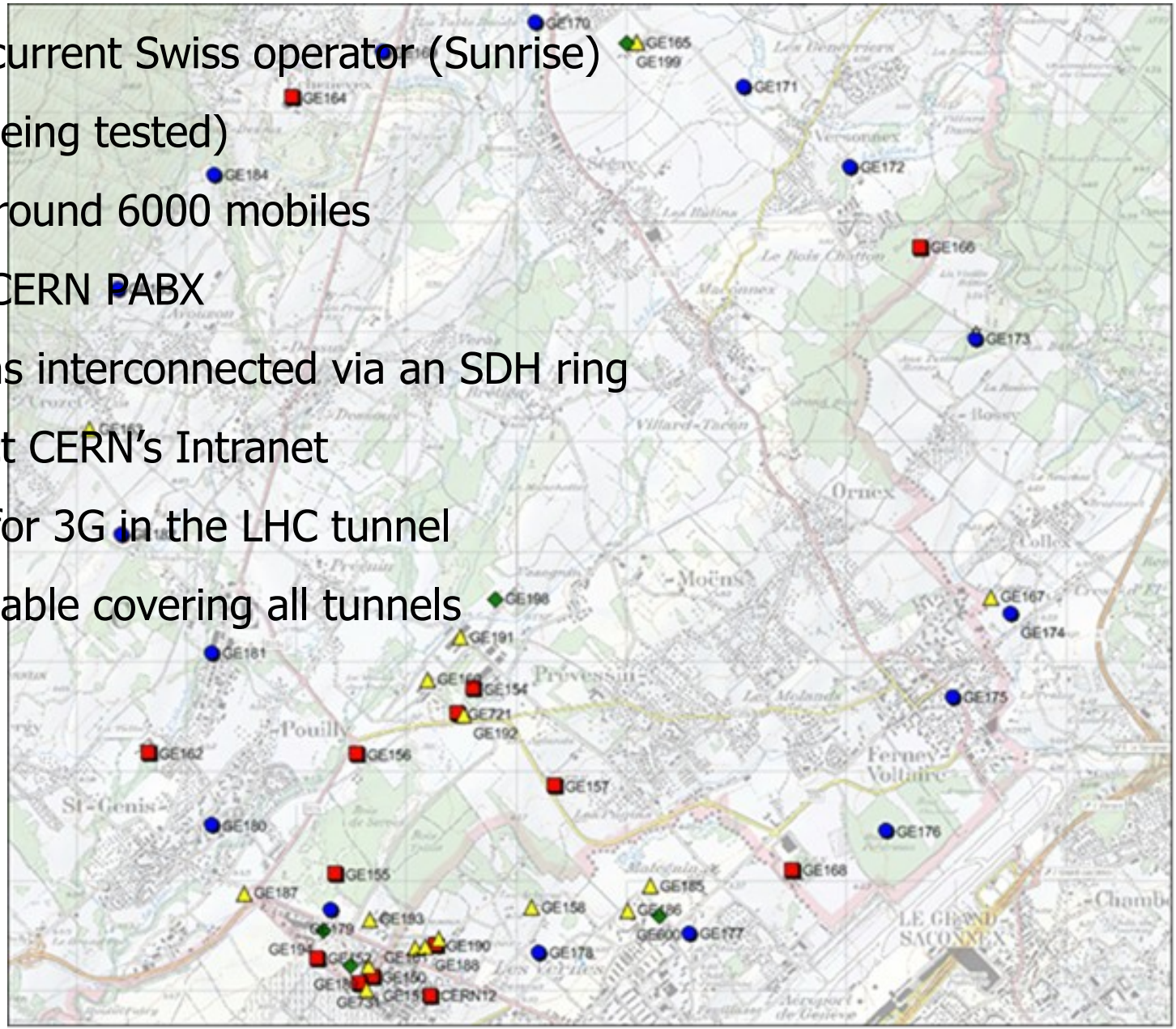
GSM calls routed by the CERN PABX

>40 Transmission Stations interconnected via an SDH ring

Dedicated APN to connect CERN's Intranet

CERN's repeater system for 3G in the LHC tunnel

>50 km of leaky feeder cable covering all tunnels



Tetra safety radio network

5 TETRA frequencies

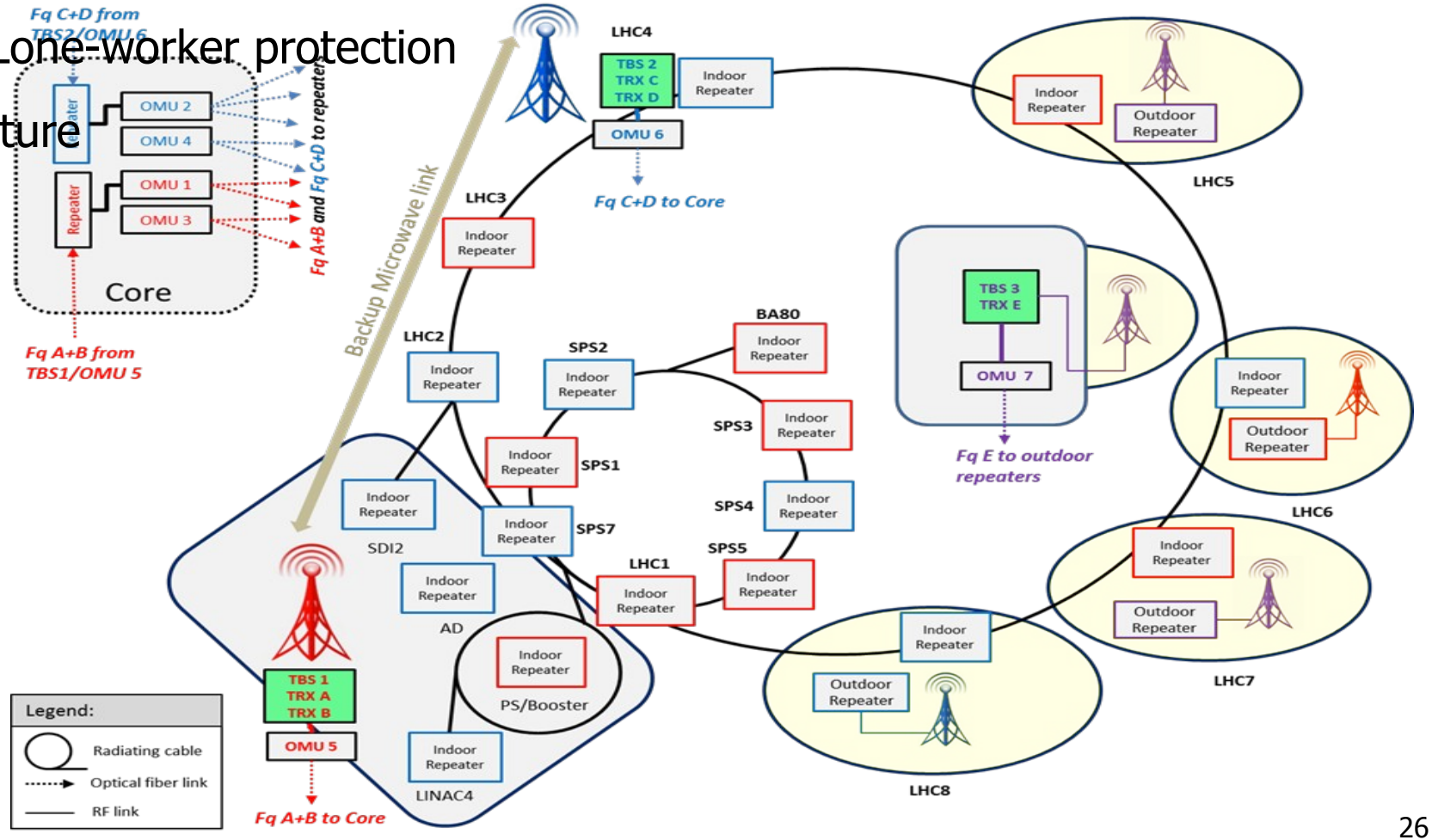
400 km2 and underground facilities covered

Interconnection with CERN PBX

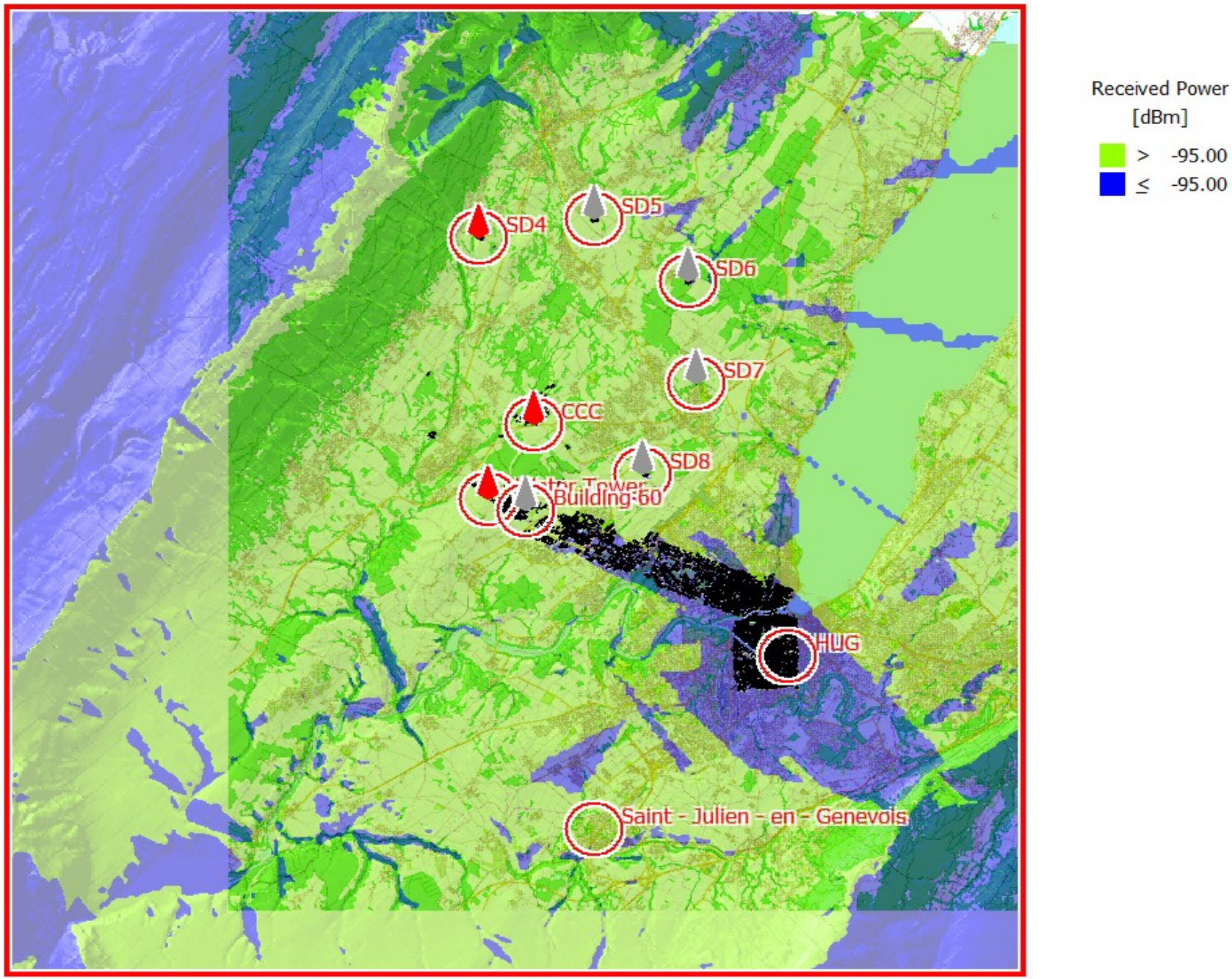
Gateways with VHF and Swiss and French TETRAPOL

Geo-localization & Lone-worker protection

Redundant architecture

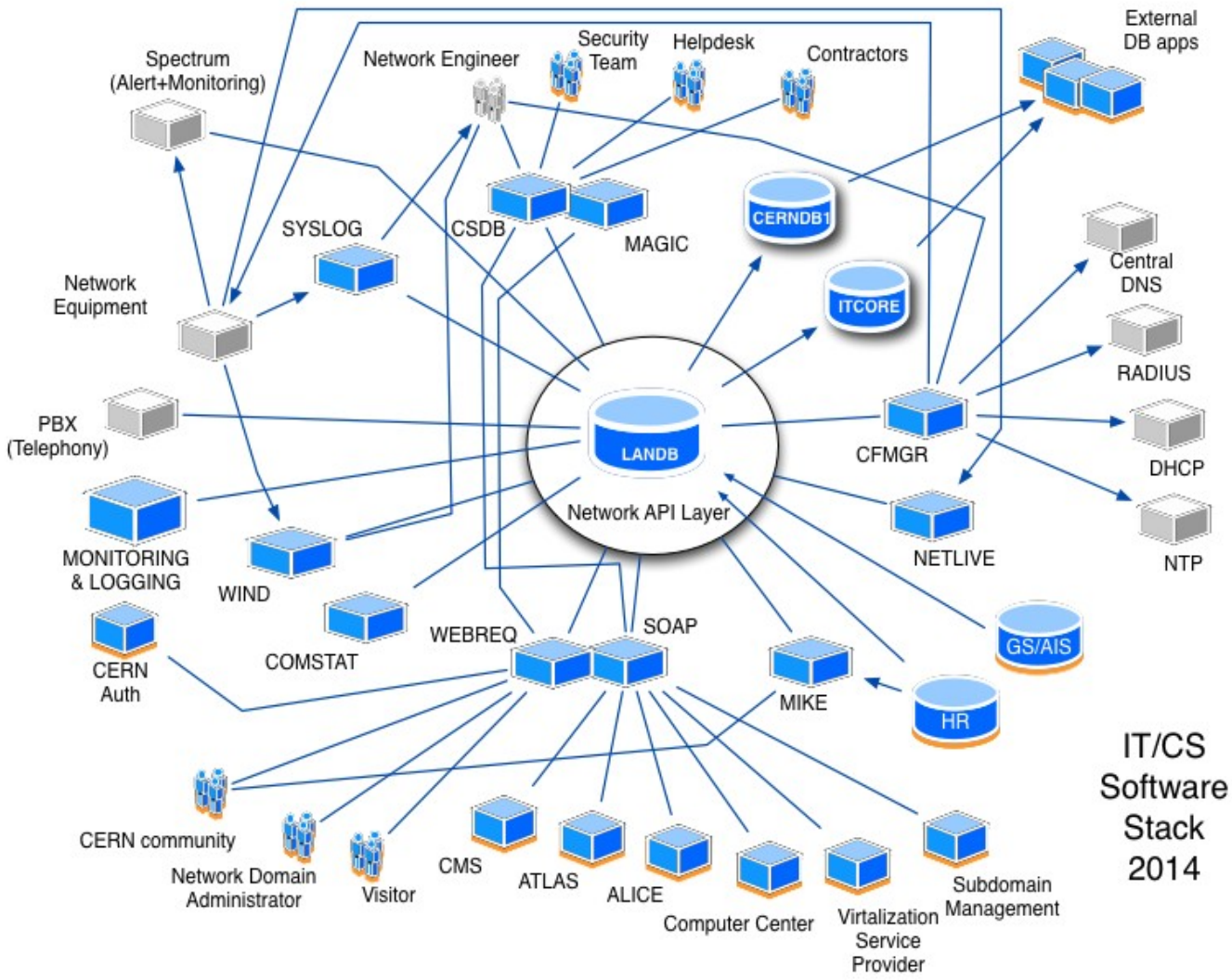


Tetra coverage



Network Provisioning and Management System

- >250 Database tables
- ~200,000 Registered devices
- >1,000,000 lines of codes
- >15 years of development

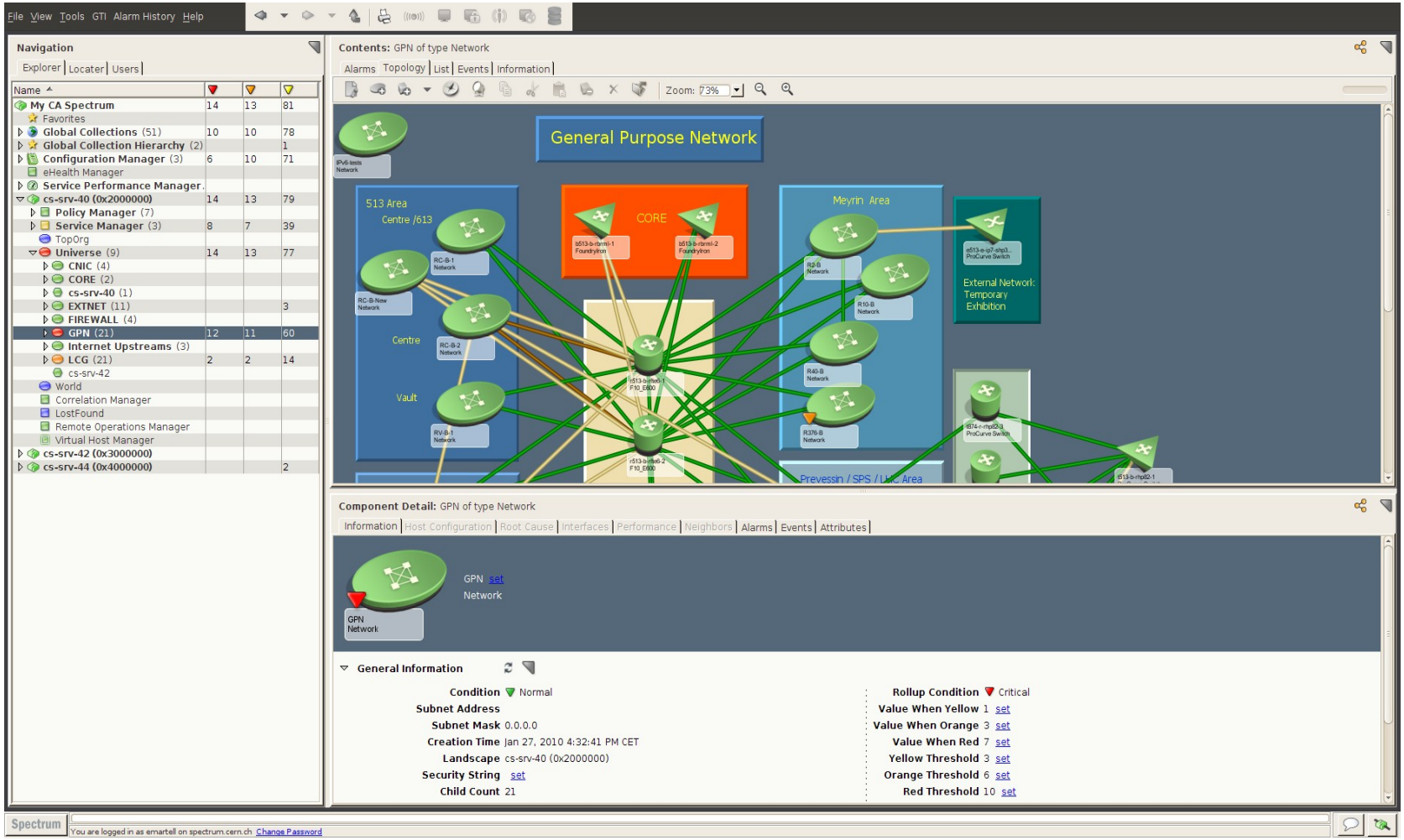


Monitoring and Operations

The whole network is monitored and operated by the CERN NOC:

- 7500 devices

- 150 alarms treated per day



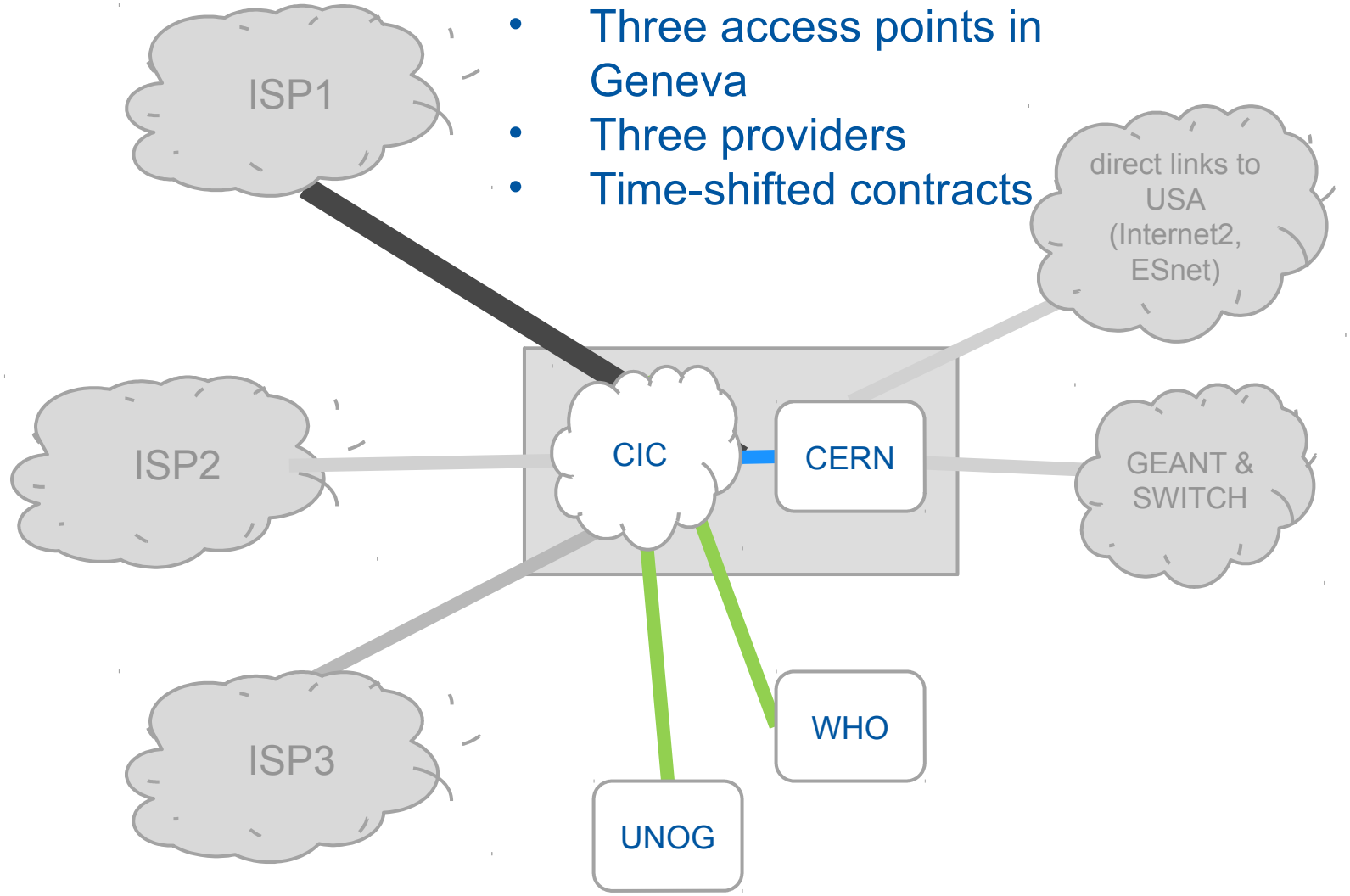
CIXP Internet eXchange Point

- one of the first IXP in Europe
- >60 members
- three locations in Geneva
- extension to Zurich

CIC - CERN Internet Consortium

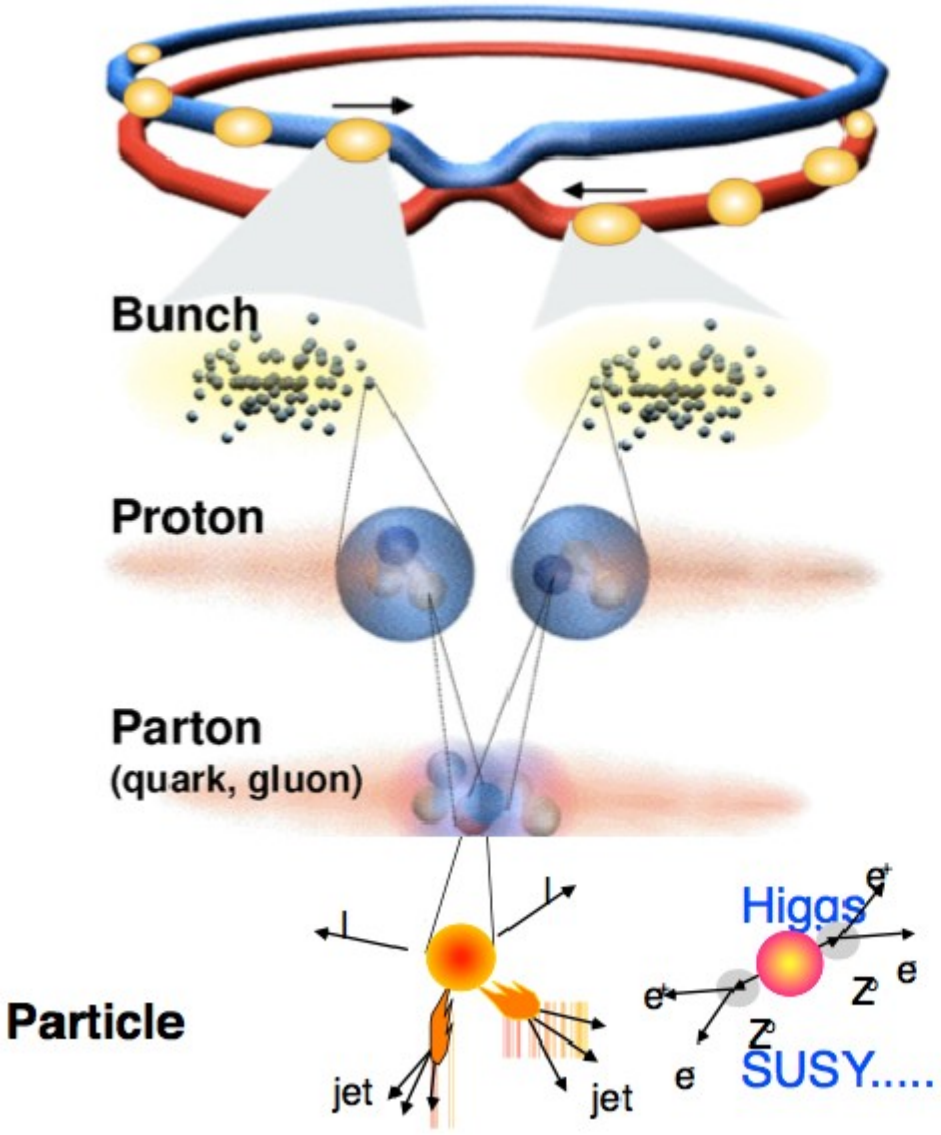
- **Internet Transit Service for International Organizations in Geneva**

- Redundancy
- Three access points in Geneva
- Three providers
- Time-shifted contracts



LHC Data Challenge

Collisions in the LHC

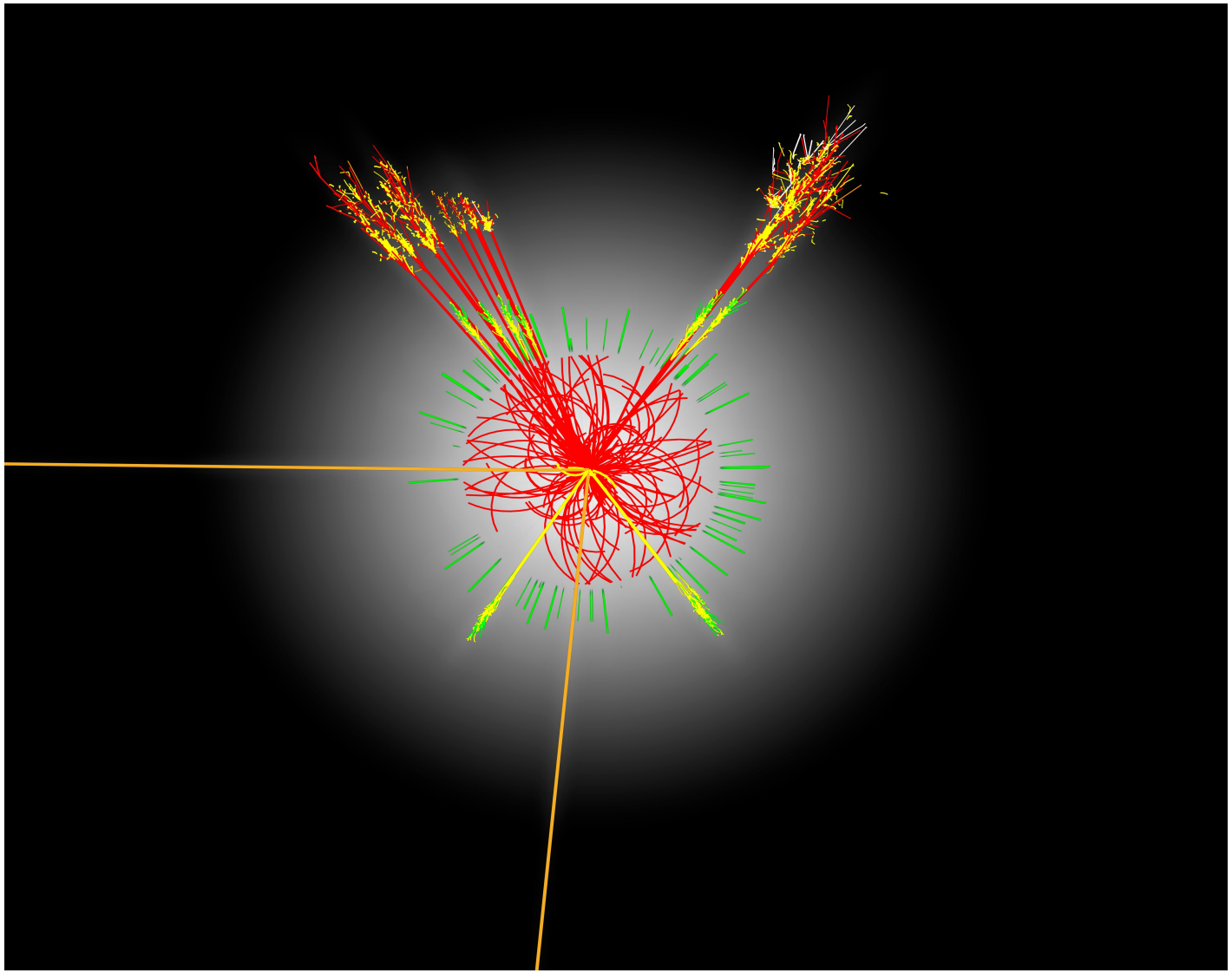


Proton - Proton	2808 bunch/beam
Protons/bunch	10^{11}
Beam energy	7 TeV (7×10^{12} eV)
Luminosity	$10^{34} cm^{-2} s^{-1}$
Crossing rate	40 MHz
Collision rate ≈	10^{7}-10^{9}

New physics rate ≈ .00001 Hz

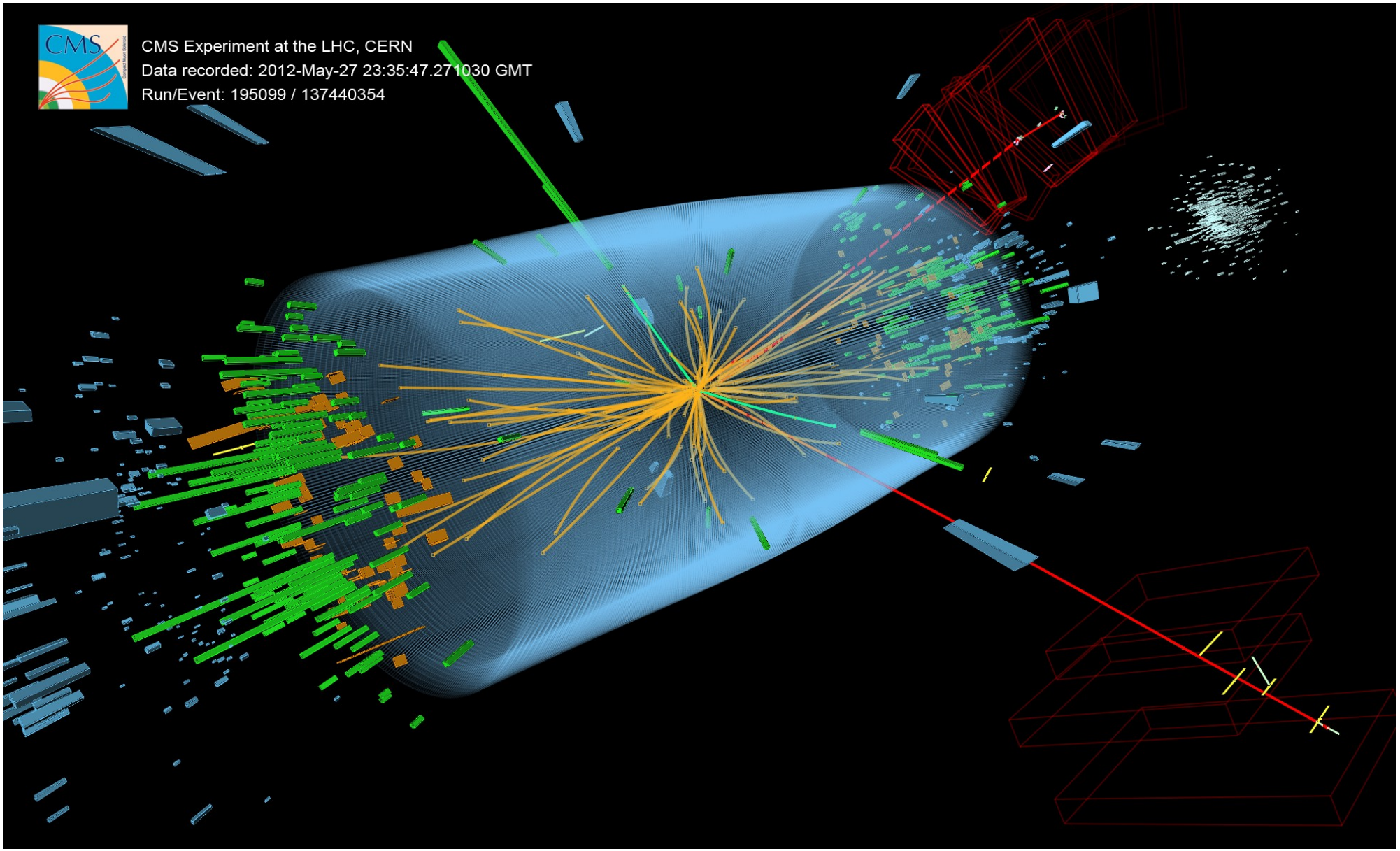
Event selection:
1 in 10,000,000,000,000

Comparing theory...



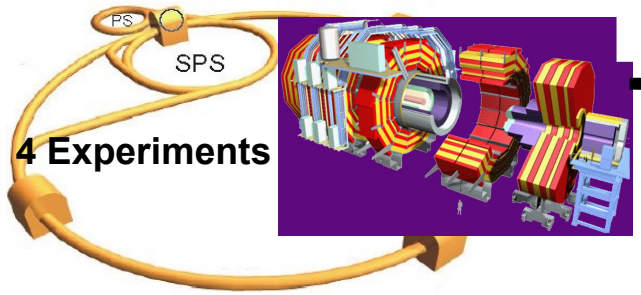
Simulated production of a Higgs event in ATLAS

.. to real events

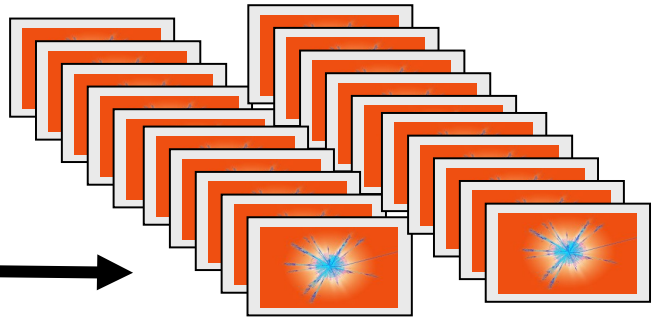


Higgs event in CMS

Data flow

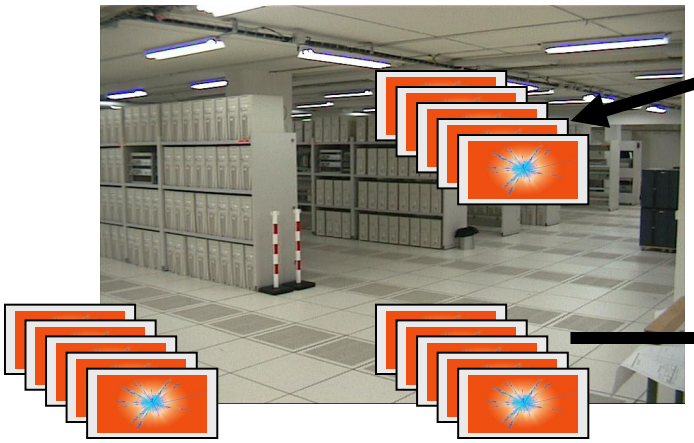


24 Pbit/s

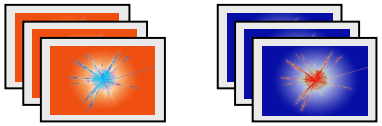


Filter and first selection

40 Gbit/s
to the CERN computer center



Create sub-samples



10 Tbit/s
Distributed + local

World-Wide Analysis

Physics
Explanation of nature

100 Gbit/s

40 Gbit/s

Store on disk and tape

Export copies

$$\sigma_{f\bar{f}} \approx \sigma^0_{f\bar{f}} \times \frac{s\Gamma_Z^2}{(s-m_Z^2)^2 + s^2\Gamma_Z^2/m_Z^2} \quad \text{with}$$

$$\sigma^0_{f\bar{f}} = \frac{12\pi}{m_Z^2}\frac{\Gamma_{ee}\Gamma_{ff}}{\Gamma_Z^2} \quad \text{and} \quad \Gamma_{ff} = \frac{G_F m_Z^3}{6\pi\sqrt{2}} \times (v_f^2 + a_f^2)$$

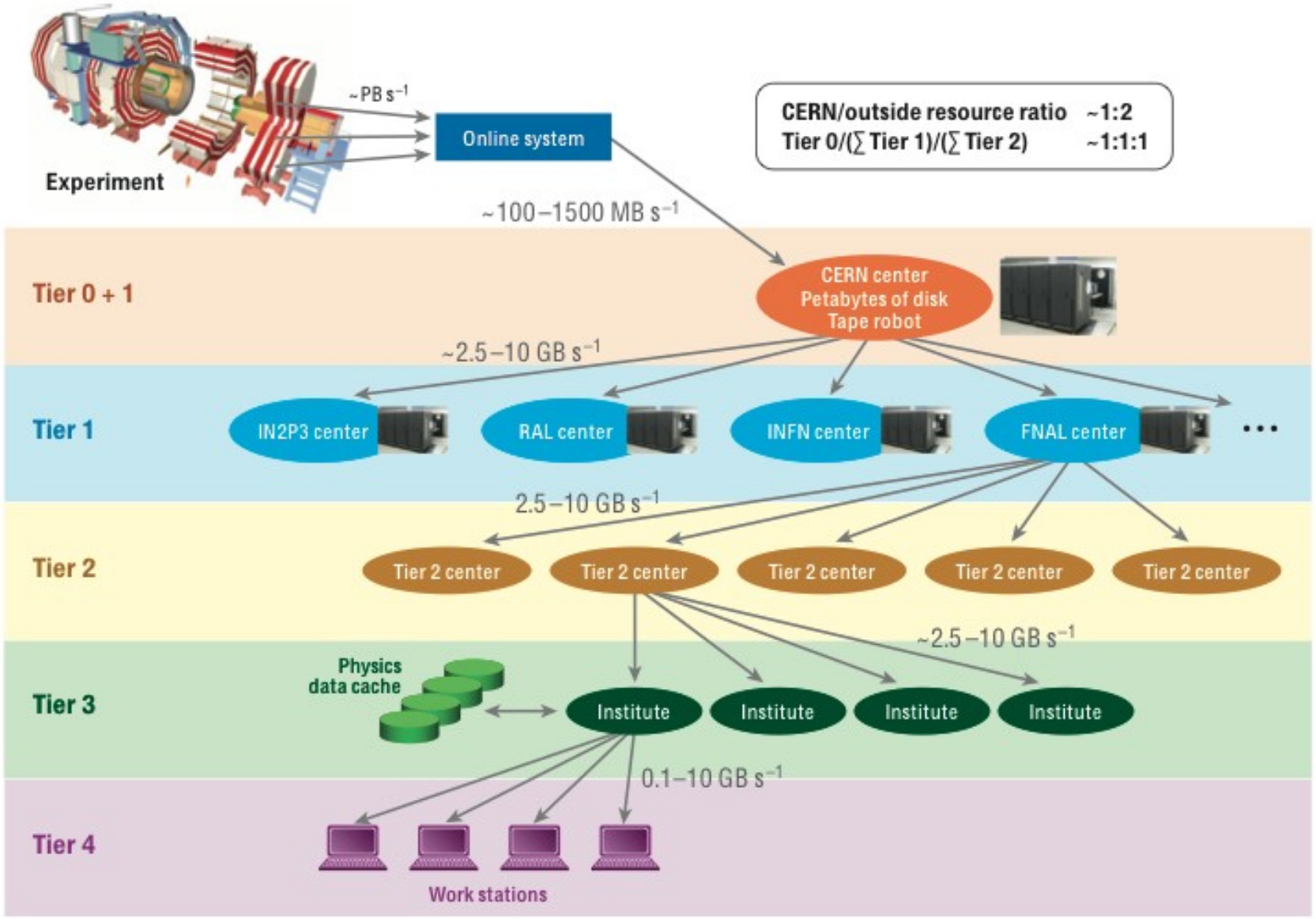
Data Challenge

- 40 million collisions per second

- After filtering, 100 collisions of interest per second

- 10^{10} collisions recorded each year
  = 15 Petabytes/year of data

Computing model



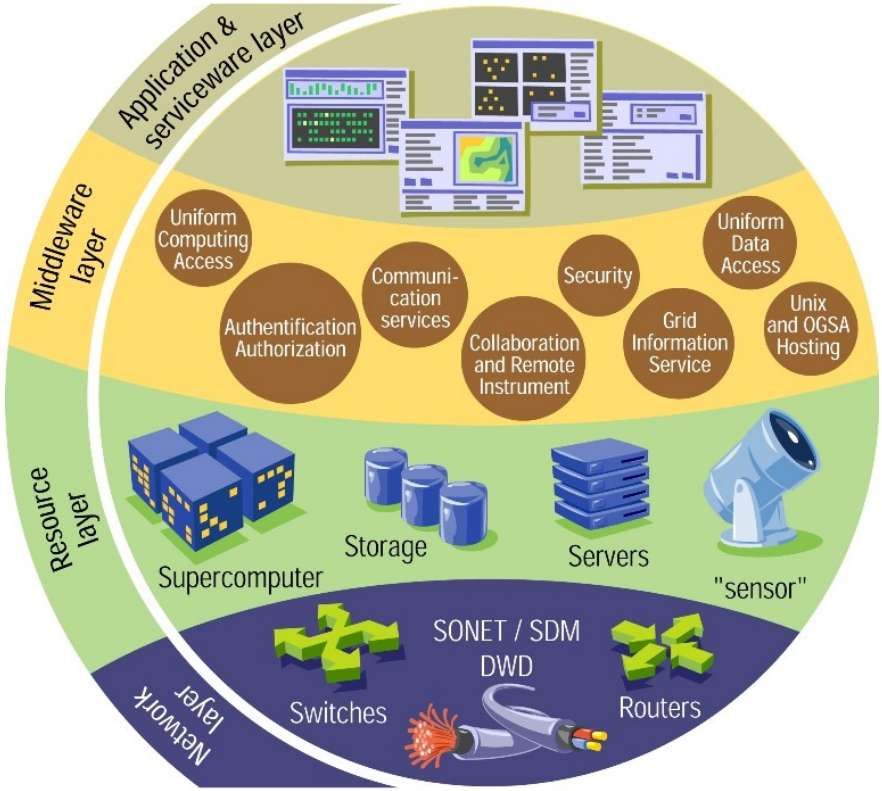
WLCG

Worldwide LHC Computing Grid

WLCG
Worldwide LHC Computing Grid

Distributed Computing Infrastructure for LHC experiments

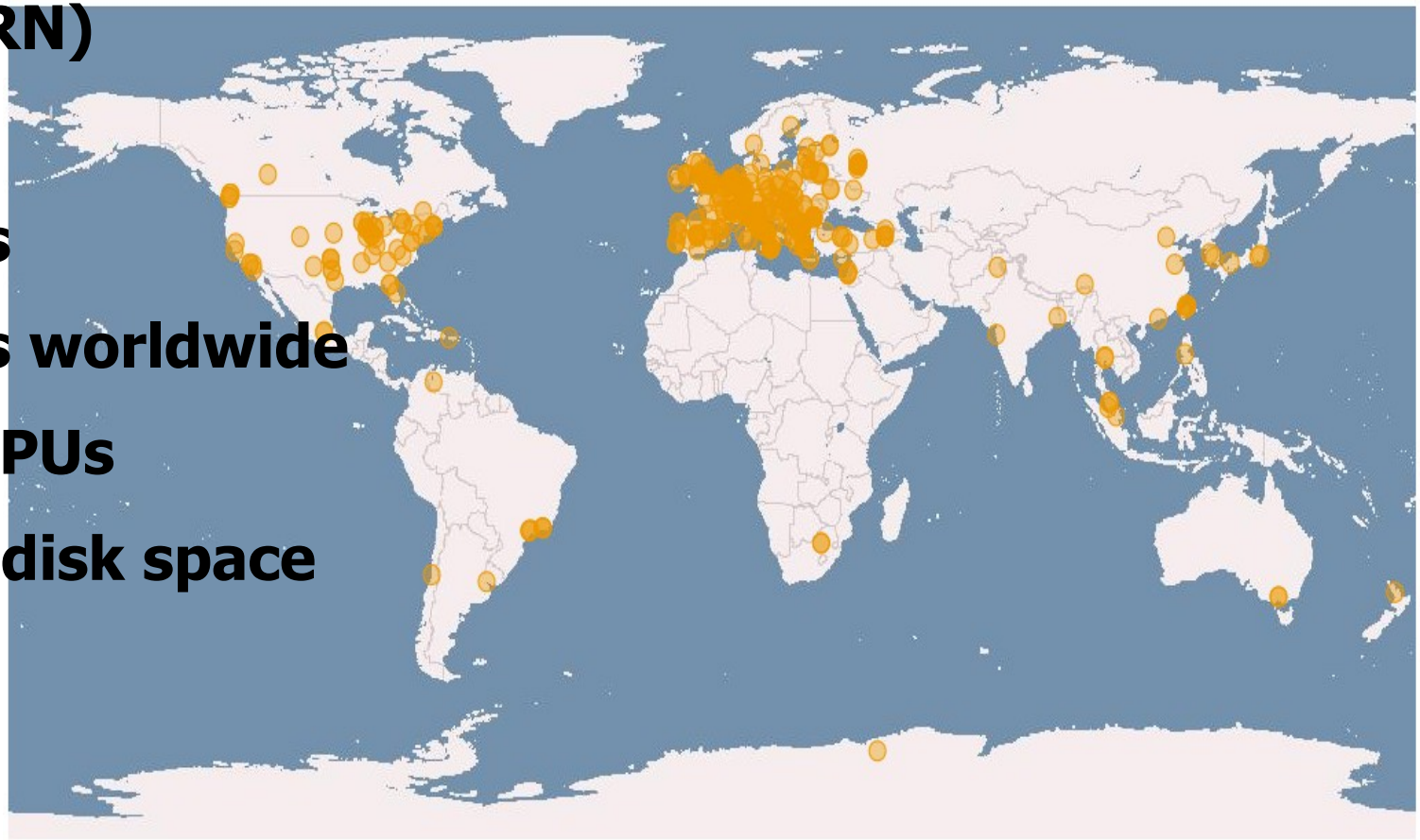
Collaborative effort of the HEP community



WLCG resources

WLCG sites:

- 1 Tier0 (CERN)
- 13 Tier1s
- ~140 Tier2s
- >300 Tier3s worldwide
- ~250,000 CPUs
- ~ 150PB of disk space



CERN Tier0 resources

Servers	**11000**
Processors	**18000**
Cores	**90000**

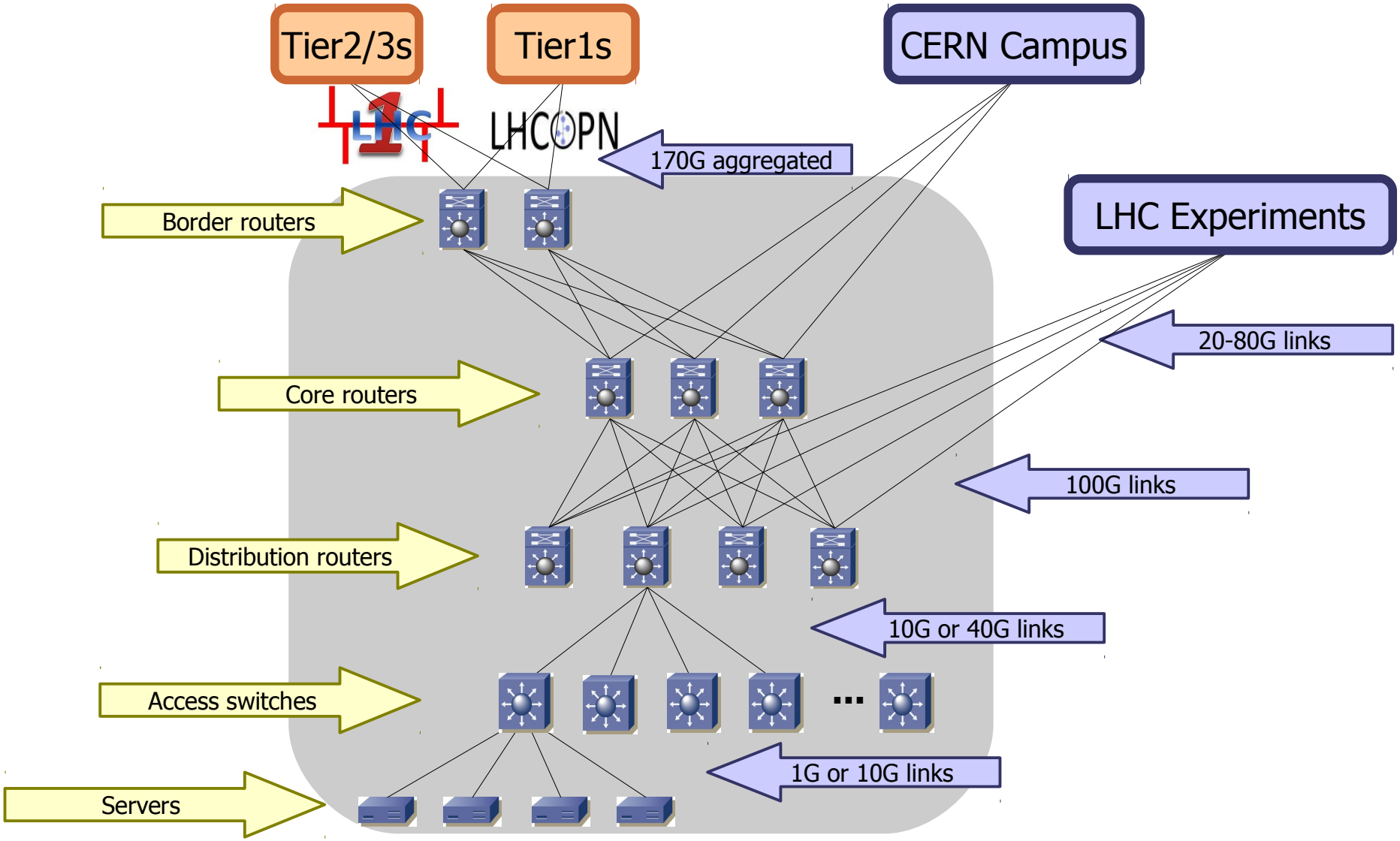
Disks	**75000**
Raw Disk Capacity (TB)	**110000**
Memory Modules	**64000**

Tape drives	**160**
Tape cartridges	**51000**
Tape slots	**65000**
Tape capacity(TB)	**92000**

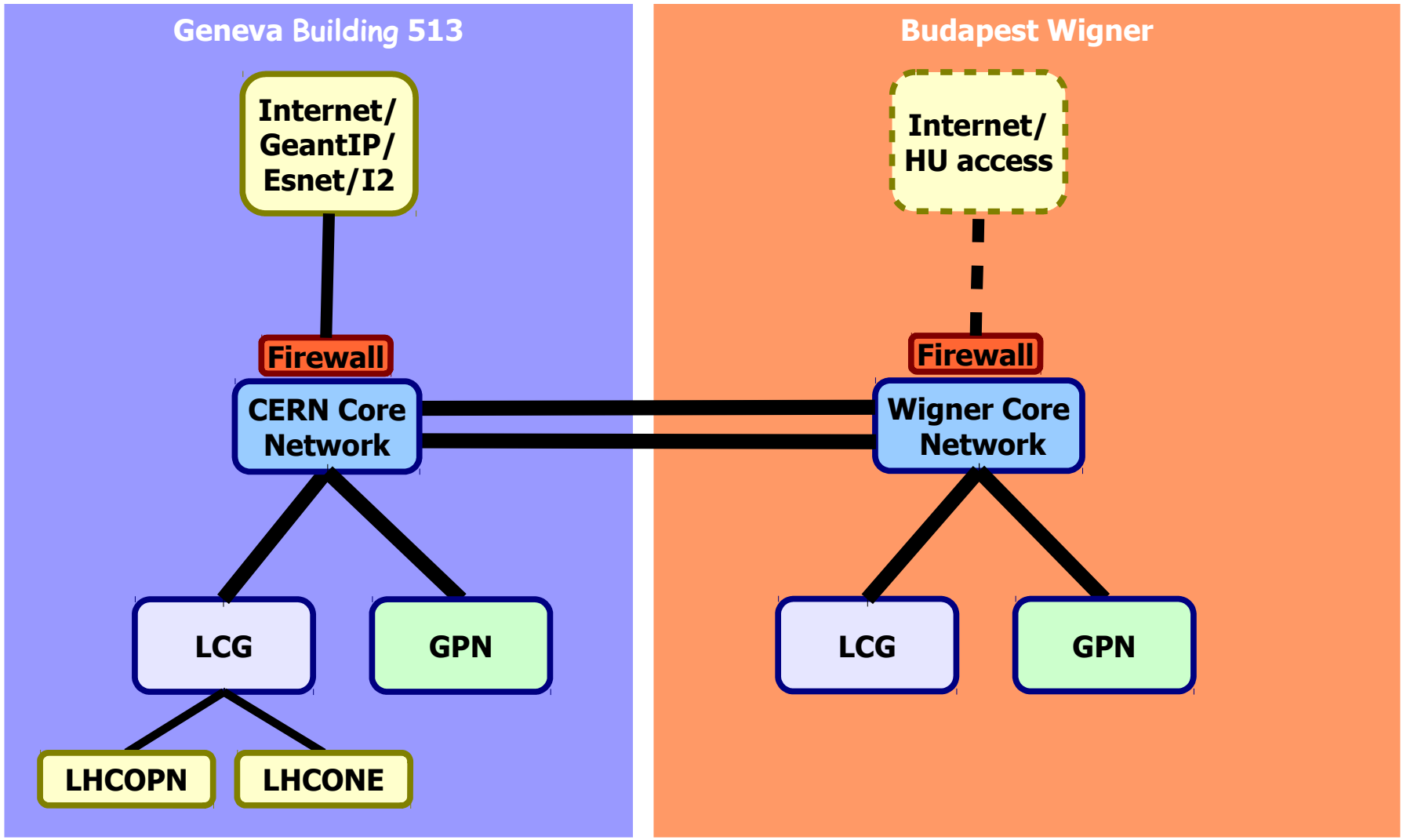
High Speed routers	**20**
Ethernet switches	**500**
10Gbps ports	**3000**
100Gbps ports	**50**

January 2014

CERN Tier0 LCG network



Wigner datacentre



Wigner resources

Wigner Data Center	2013	2014
Power	~900KW	~1200KW
Racks	90	120
Routers	6	10+Firewall
100Gbps ports	18	18
Switches	140	210
Servers	~1200	~1800
L2 Switch	**2013**	**2014**
1Gbps ports	3072	4608
10Gbps ports	528	792

2x100Gbps circuits CERN-Wigner

Trends

Virtualization mobility (Software Defined Networks)

Commodity Servers with 10G NICs

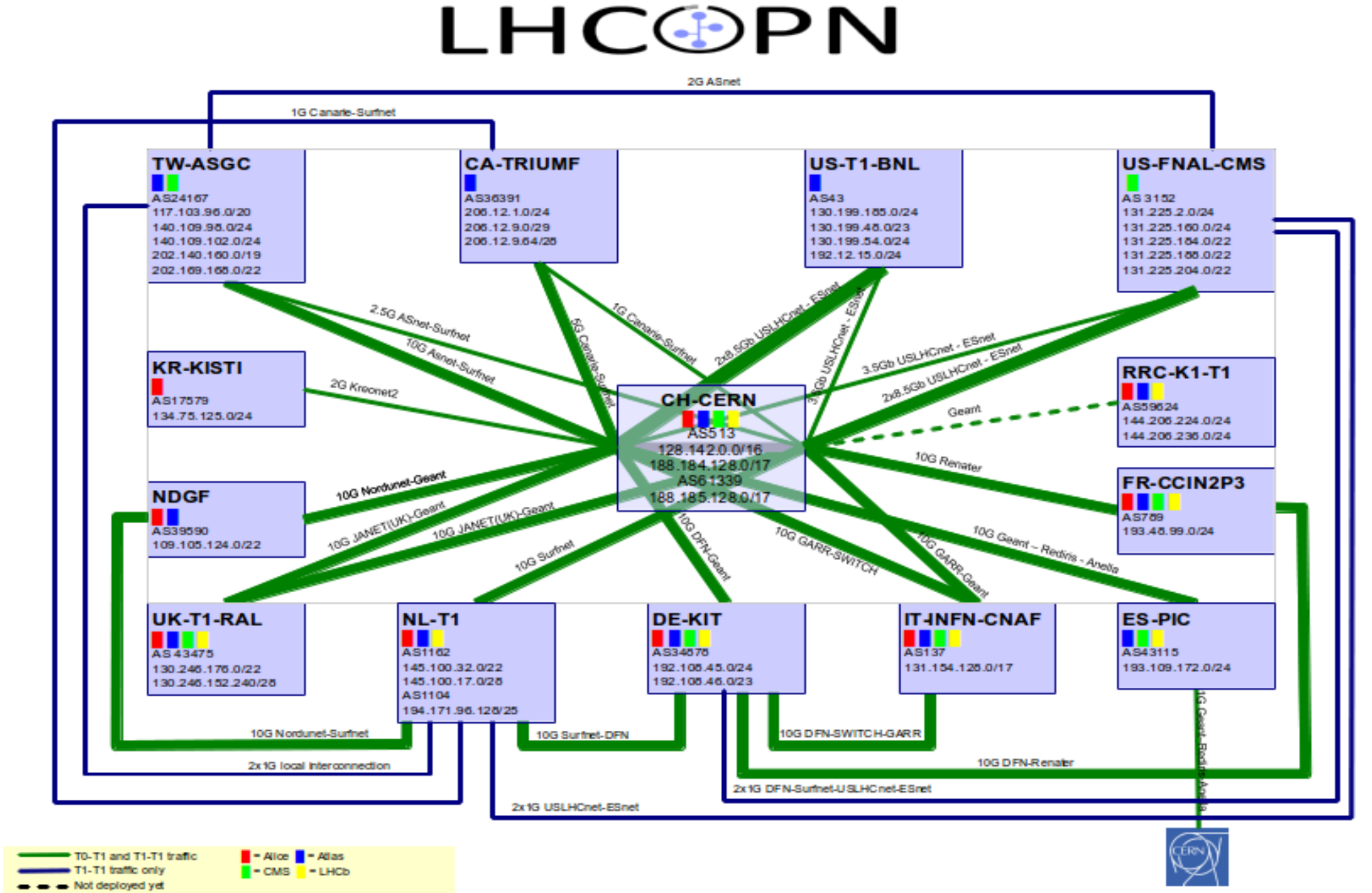
High-end Servers with 40G NICs

40G and 100G interfaces on switches and routers

LHCOPN

LHC Optical Private Network

Tier0-Tier1s network



A collaborative effort

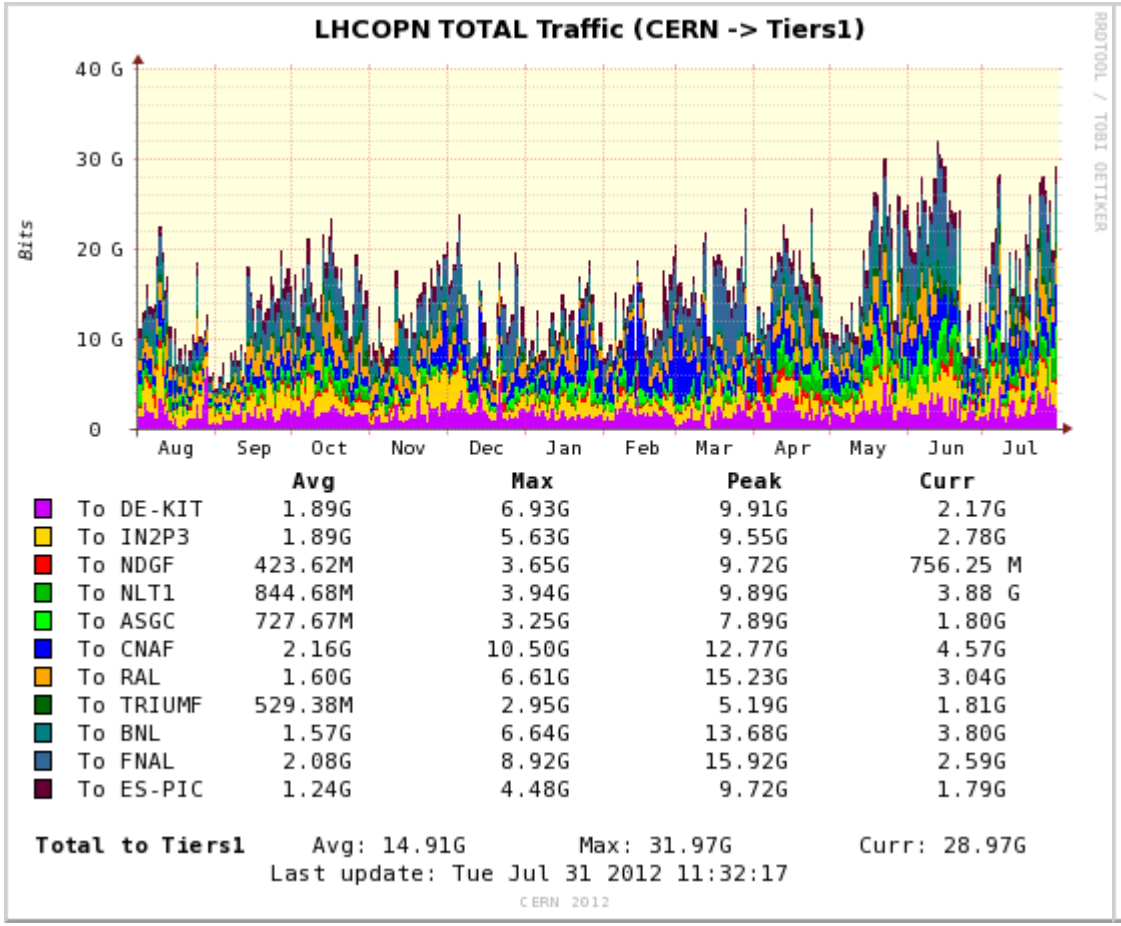
Designed, built and operated by the Tier0-Tier1s community

Links provided by the Research and Education network providers: Geant, USLHCnet, Esnet, Canarie, ASnet, Nordunet, Surfnet, GARR, Renater, JANET.UK, Rediris, DFN, SWITCH

Technology

- Single and bundled long distance 10G ethernet links
- Multiple redundant paths. Star+PartialMesh topology
- BGP routing: communities for traffic engineering, load balancing.
- Security: only declared IP prefixes can exchange traffic.

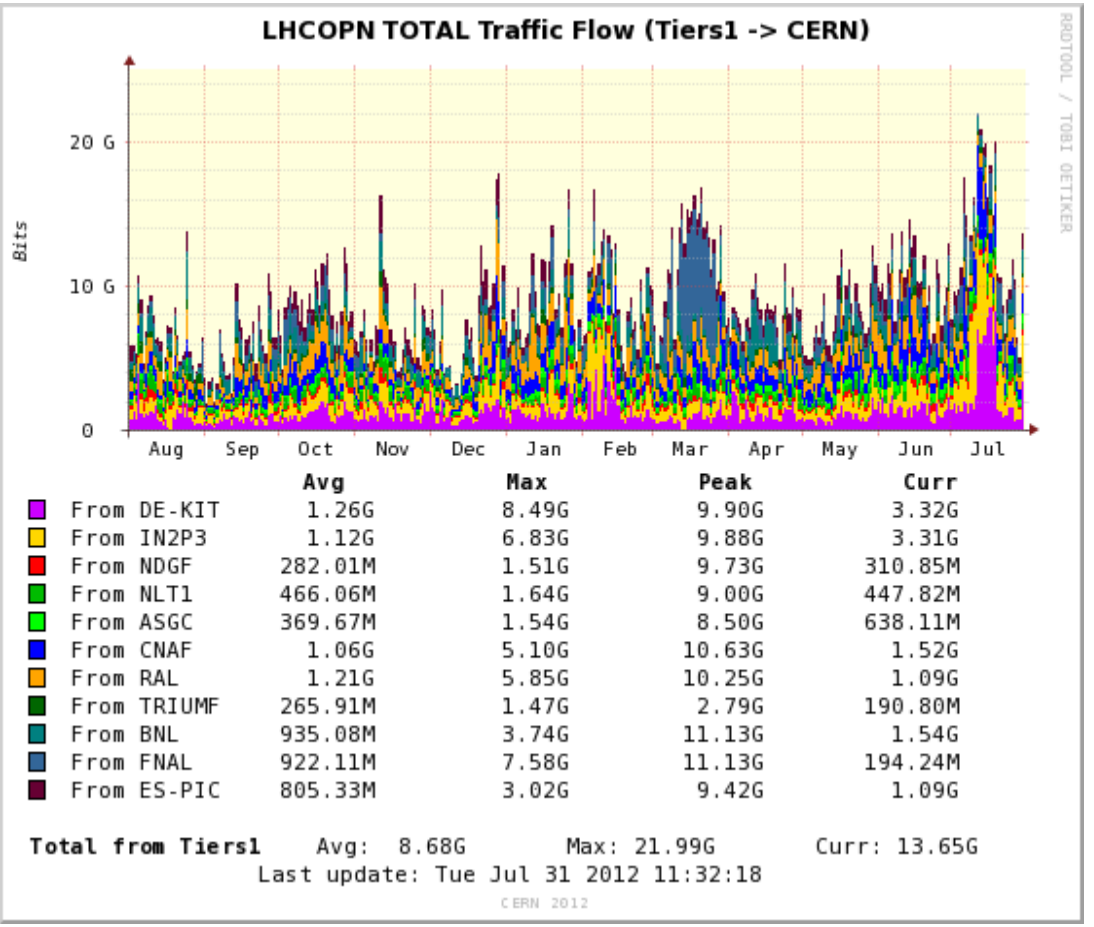
Traffic to the Tier1s



	Avg	Max	Peak	Curr
To DE-KIT	1.89G	6.93G	9.91G	2.17G
To IN2P3	1.89G	5.63G	9.55G	2.78G
To NDGF	423.62M	3.65G	9.72G	756.25 M
To NLT1	844.68M	3.94G	9.89G	3.88 G
To ASGC	727.67M	3.25G	7.89G	1.80G
To CNAF	2.16G	10.50G	12.77G	4.57G
To RAL	1.60G	6.61G	15.23G	3.04G
To TRIUMF	529.38M	2.95G	5.19G	1.81G
To BNL	1.57G	6.64G	13.68G	3.80G
To FNAL	2.08G	8.92G	15.92G	2.59G
To ES-PIC	1.24G	4.48G	9.72G	1.79G

Total to Tiers1 Avg: 14.91G Max: 31.97G Curr: 28.97G

Last update: Tue Jul 31 2012 11:32:17

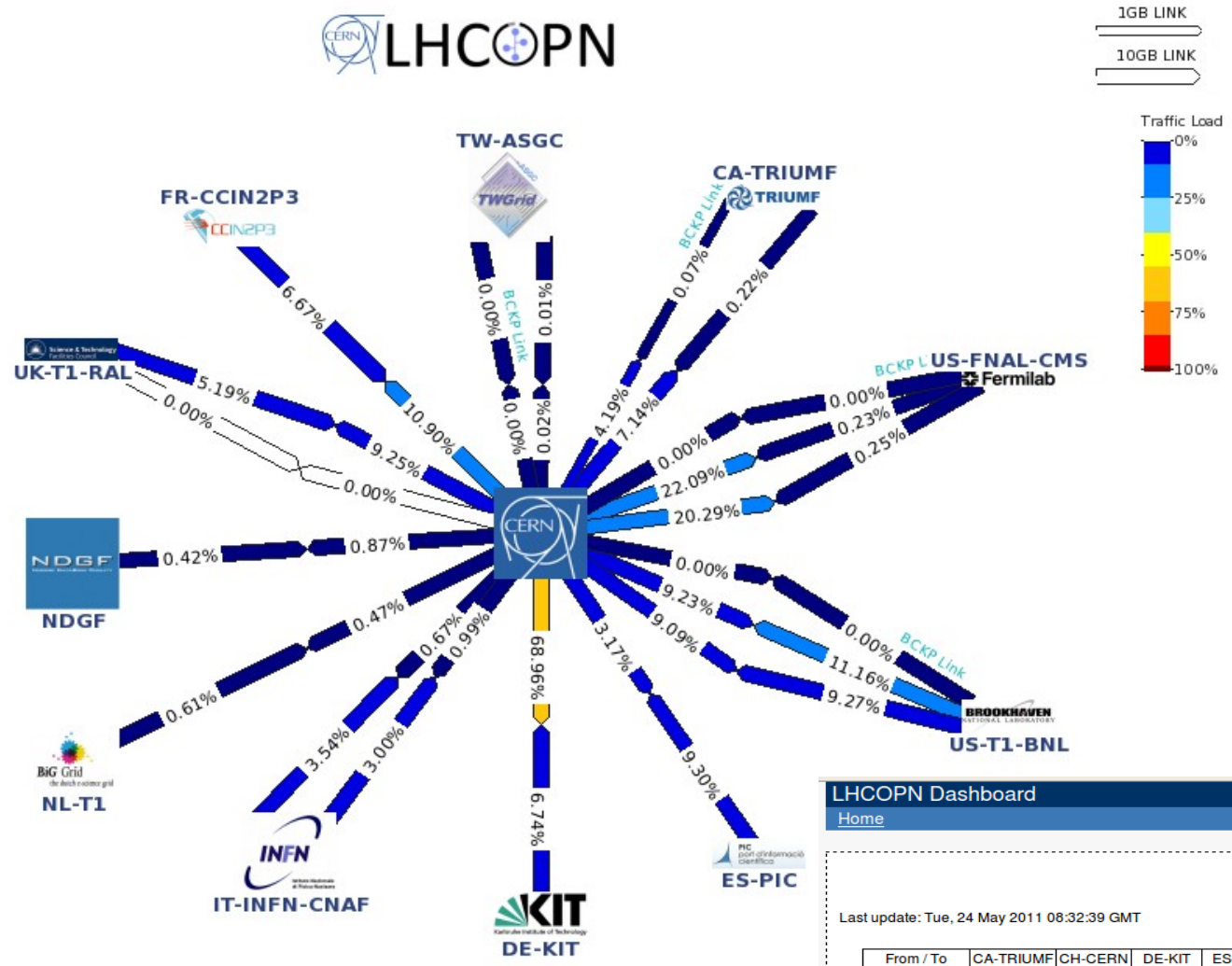


	Avg	Max	Peak	Curr
From DE-KIT	1.26G	8.49G	9.90G	3.32G
From IN2P3	1.12G	6.83G	9.88G	3.31G
From NDGF	282.01M	1.51G	9.73G	310.85M
From NLT1	466.06M	1.64G	9.00G	447.82M
From ASGC	369.67M	1.54G	8.50G	638.11M
From CNAF	1.06G	5.10G	10.63G	1.52G
From RAL	1.21G	5.85G	10.25G	1.09G
From TRIUMF	265.91M	1.47G	2.79G	190.80M
From BNL	935.08M	3.74G	11.13G	1.54G
From FNAL	922.11M	7.58G	11.13G	194.24M
From ES-PIC	805.33M	3.02G	9.42G	1.09G

Total from Tiers1 Avg: 8.68G Max: 21.99G Curr: 13.65G

Last update: Tue Jul 31 2012 11:32:18

Monitoring



LHCOPN Dashboard

Home

Current status

Last update: Tue, 24 May 2011 08:32:39 GMT

From / To	CA-TRIUMF	CH-CERN	DE-KIT	ES-PIC	FR-CCIN2P3	IT-INFN-CNAF	NDGF	NL-T1	TW-ASGC	UK-T1-RAL	US-FNAL-CMS	US-T1-BNL
CA-TRIUMF												
CH-CERN												
DE-KIT												
ES-PIC												
FR-CCIN2P3												
IT-INFN-CNAF												
NDGF												
NL-T1												
TW-ASGC												
UK-T1-RAL												
US-FNAL-CMS												
US-T1-BNL												

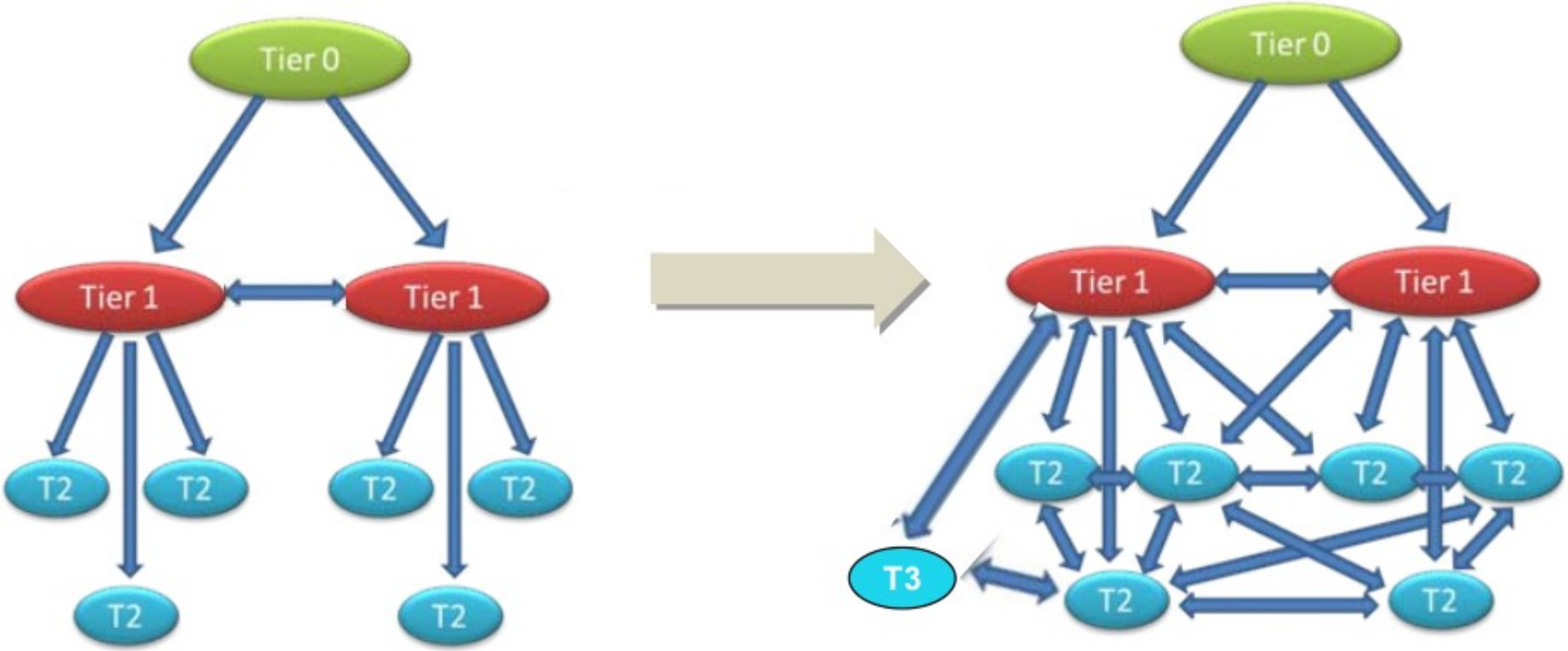
JSON feed

Legend: OK | Deviation from Baseline | Critical | Unknown

LHCONE

LHC Open Network Environment

Computing model evolution



Original MONARCH model

Model evolution

LHCONE concepts

- Serving any LHC sites according to their needs and allowing them to grow

- Sharing the cost and use of expensive resources

- A collaborative effort among Research & Education Network Providers

- Traffic separation: no clash with other data transfer, resource allocated for and funded by the HEP community

LHCONE services

L3VPN (VRF): routed Virtual Private Network - *operational*

perfSONAR: monitoring infrastructure

What LHCONE L3VPN is:

Layer3 (routed) Virtual Private Network

Dedicated worldwide backbone connecting **Tier1s, T2s and T3s** at high bandwidth

Reserved to LHC data transfers and analysis

Advantages

Bandwidth dedicated to LHC data analysis, no contention with other research projects

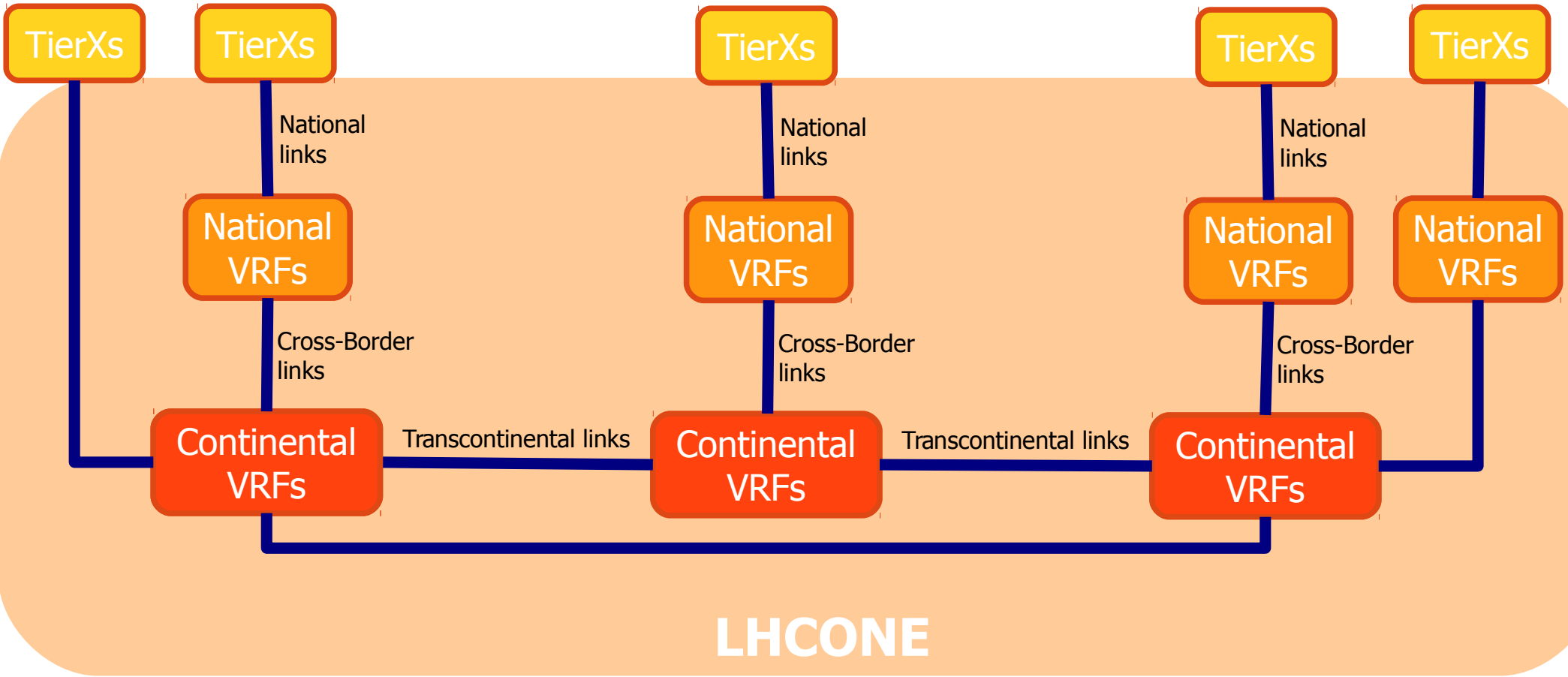
Well defined cost tag for WLCG networking

Trusted traffic that can bypass firewalls

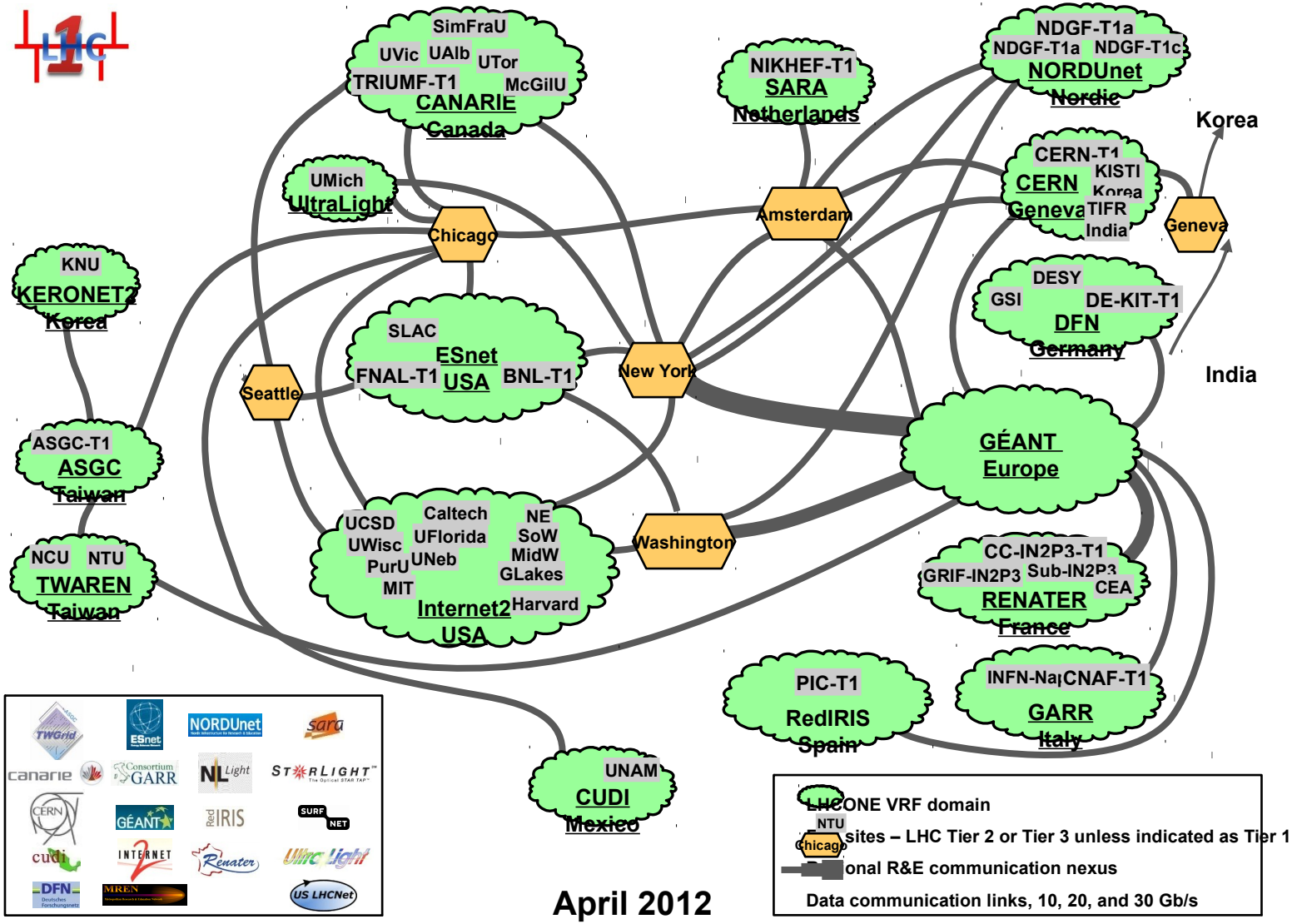
LHCONE L3VPN architecture

- TierX sites connected to National-VRFs or Continental-VRFs
- National-VRFs interconnected via Continental-VRFs
- Continental-VRFs interconnected by trans-continental/trans-oceanic links

Acronyms: **VRF** = Virtual Routing Forwarding (virtual routing instance)

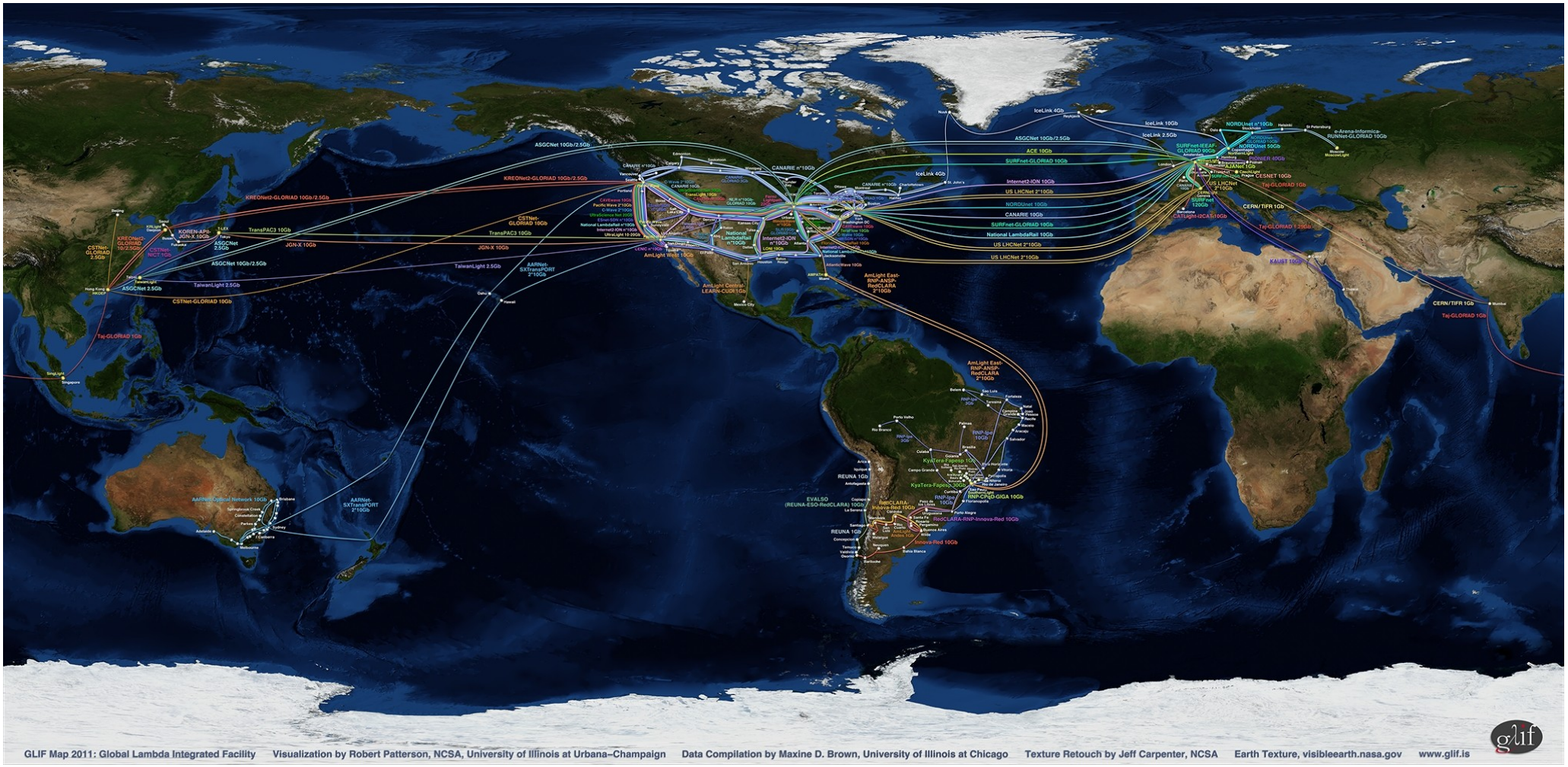


Current L3VPN topology



credits: Joe Metzger, ESnet

The underlying infrastructure



Status

Over 15 national and international Research Networks

Several Open Exchange Points including NetherLight, StarLight, MANLAN, CERNlight and others

Trans-Atlantic connectivity provided by ACE, GEANT, NORDUNET and USLHCNET

~50 end sites connected to LHCONE:
- 8 Tier1s
- 40 Tier2s

Credits: Mian Usman, Dante
More Information: https://indico.cern.ch/event/269840/contribution/4/material/slides/0.ppt

What LHCONE perfSONAR is

LHCONE Network monitoring infrastructure

Probe installed at the VRFs interconnecting points and at the TierXs

Accessible to any TierX for network healthiness checks

perfSONAR

- framework for active and passive network probing

- developed by Internet2, Esnet, Geant and others

- two interoperable flavors: perfSONAR-PS and perfSONAR-MDM

- WLCG recommended version: perfSONAR-PS

Status

Endorsed by WLCG to be a standard WLCG service

Probes already deployed in many TierXs.

Being deployed in the VRF networks

Full information:
https://twiki.cern.ch/twiki/bin/view/LCG/PerfsonarDeployment

Openlab and IT-CS

Openlab project:
CINBAD

CINBAD

CERN **I**nvestigation of **N**etwork **B**ehaviour and **A**nomaly **D**etection

Project Goals:

Understand the behaviour of large computer networks (10'000+ nodes) in High Performance Computing or large Campus installations to be able to:

- detect traffic anomalies in the system
- perform trend analysis
- automatically take counter measures
- provide post-mortem analysis facilities

Resources:

- In collaboration with HP Networking
- Two Engineers in IT-CS

Started in 2007, completed in 2010

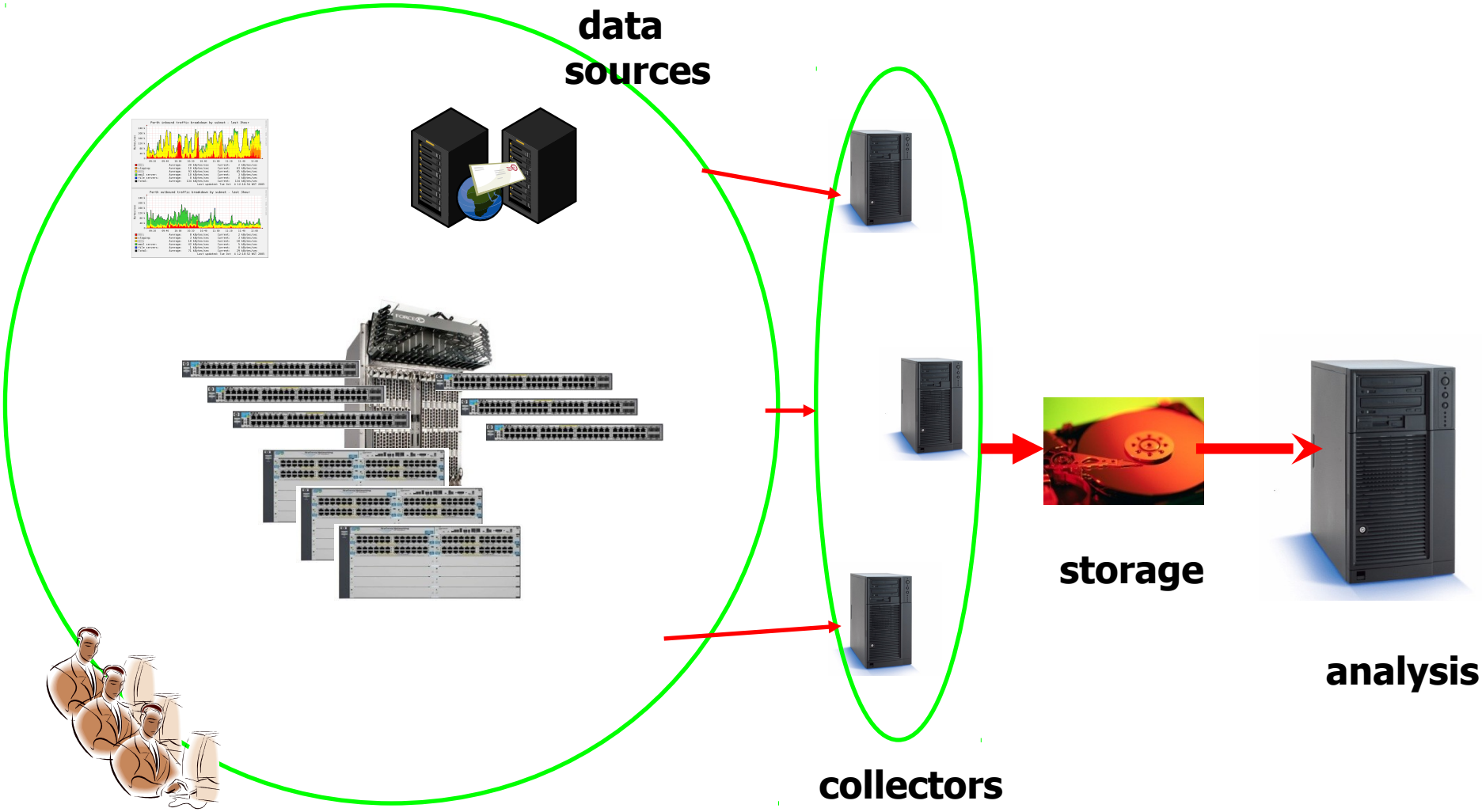
Results

Project completed in 2010

For CERN:
Designed and deployed a complete framework (hardware and software) to detect anomalies in the Campus Network (GPN)

For HP:
Intellectual properties of new technologies used in commercial products

CINBAD Architecture



Openlab project:
WIND

WIND

Wireless **I**nfrastructure **N**etwork **D**eployment

Project Goals

- Analyze the problems of large scale wireless deployments and understand the constraint
- Simulate behaviour of WLAN
- Develop new optimisation algorithms

Resources:

- In collaboration with HP Networking
- Two Engineers in IT-CS

Started in 2010, completed in 2013

Needs

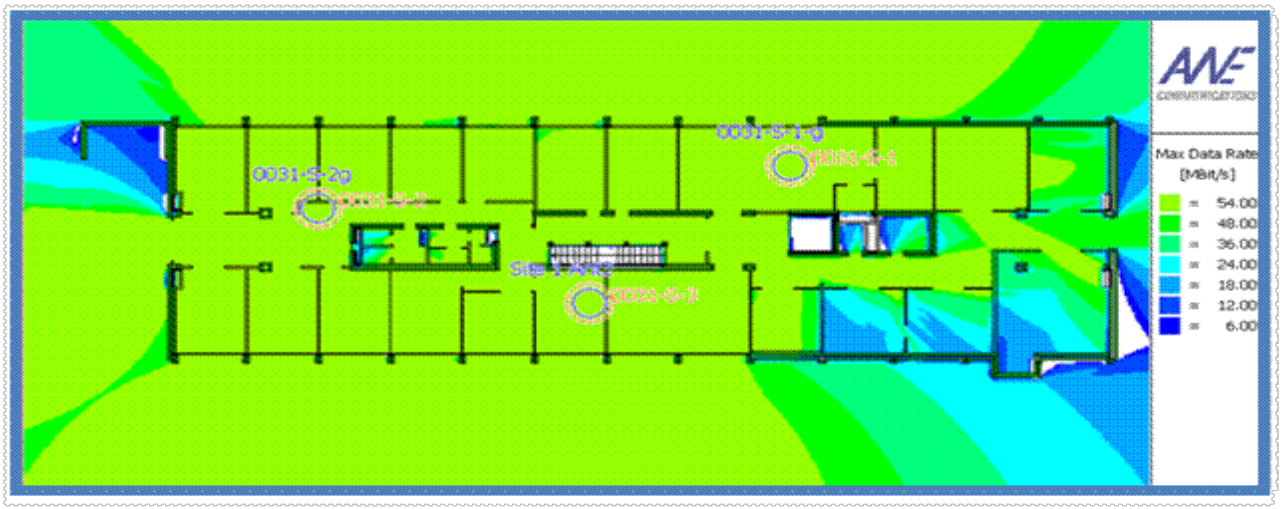
Wireless LAN (WLAN) deployments are problematic:

- Radio propagation is very difficult to predict
- Interference is an ever present danger
- WLANs are difficult to properly deploy
- Monitoring was not an issue when the first standards were developed
- When administrators are struggling just to operate the WLAN, performance optimisation is often forgotten

Example: Radio interferences

Max data rate in 0031-S: The APs work on the same channel

Max data rate in 0031-S: The APs work on 3 independent channels



Results

Project completed in 2013

For CERN:
Extend monitoring and analysis tool which allows to:
- streamline WLAN design
- run accurate troubleshoot
- isolate misbehaving clients

For HP:
Intellectual properties of new technologies used in commercial products

Openlab project:
ViSION

ViSION

Project Goals:

- Develop a SDN traffic orchestrator using OpenFlow

Resources:

- In collaboration with HP Networking
- Two Engineers in IT-CS

Started in 2012, ended in 2013

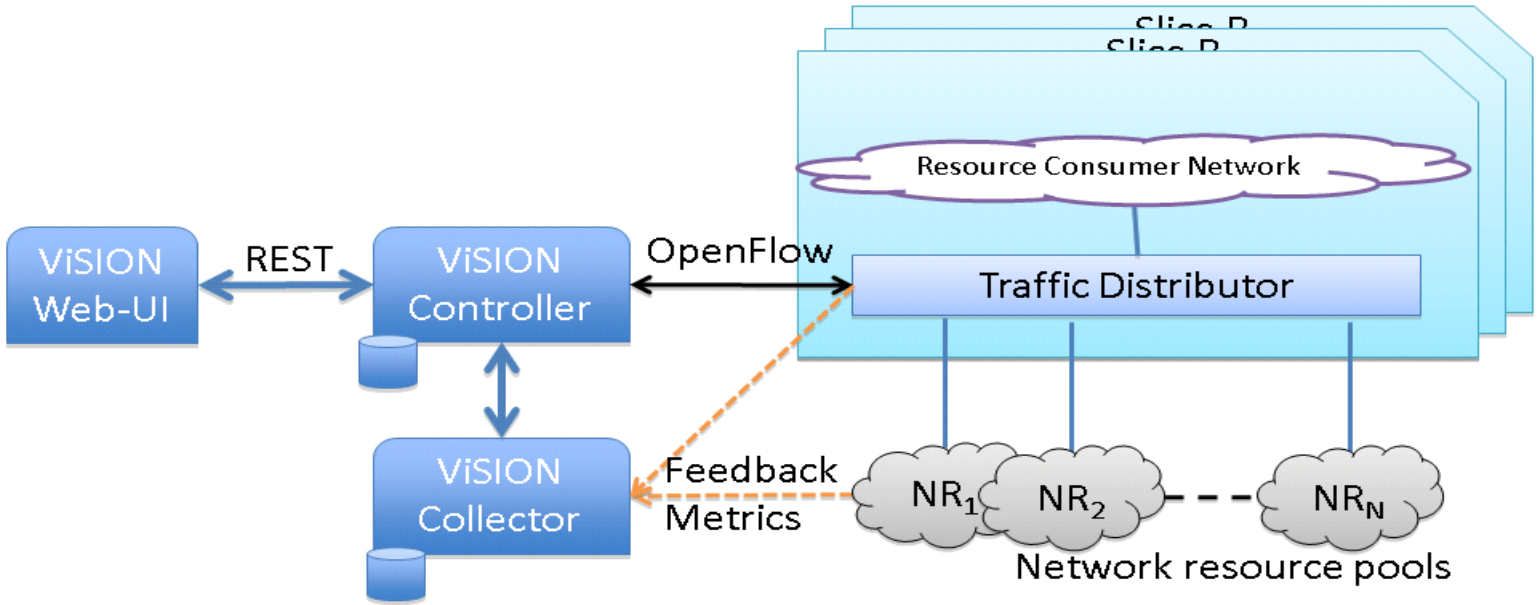
Goals

SDN traffic orchestrator using OpenFlow:

- distribute traffic over a set of network resources
- perform classification (different types of applications and resources)
- perform load sharing (similar resources).

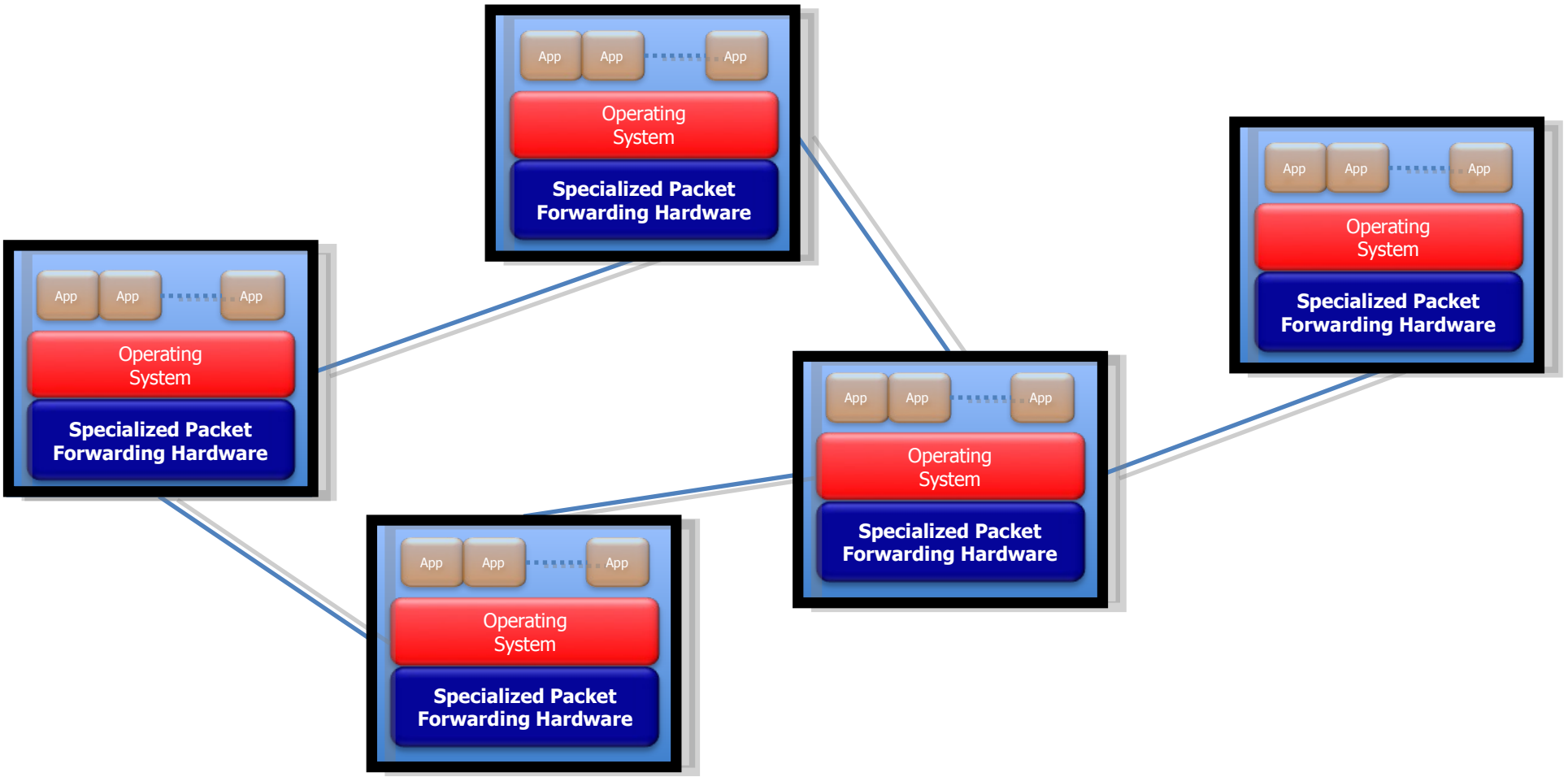
Benefits:

- improved scalability and control than traditional networking technologies

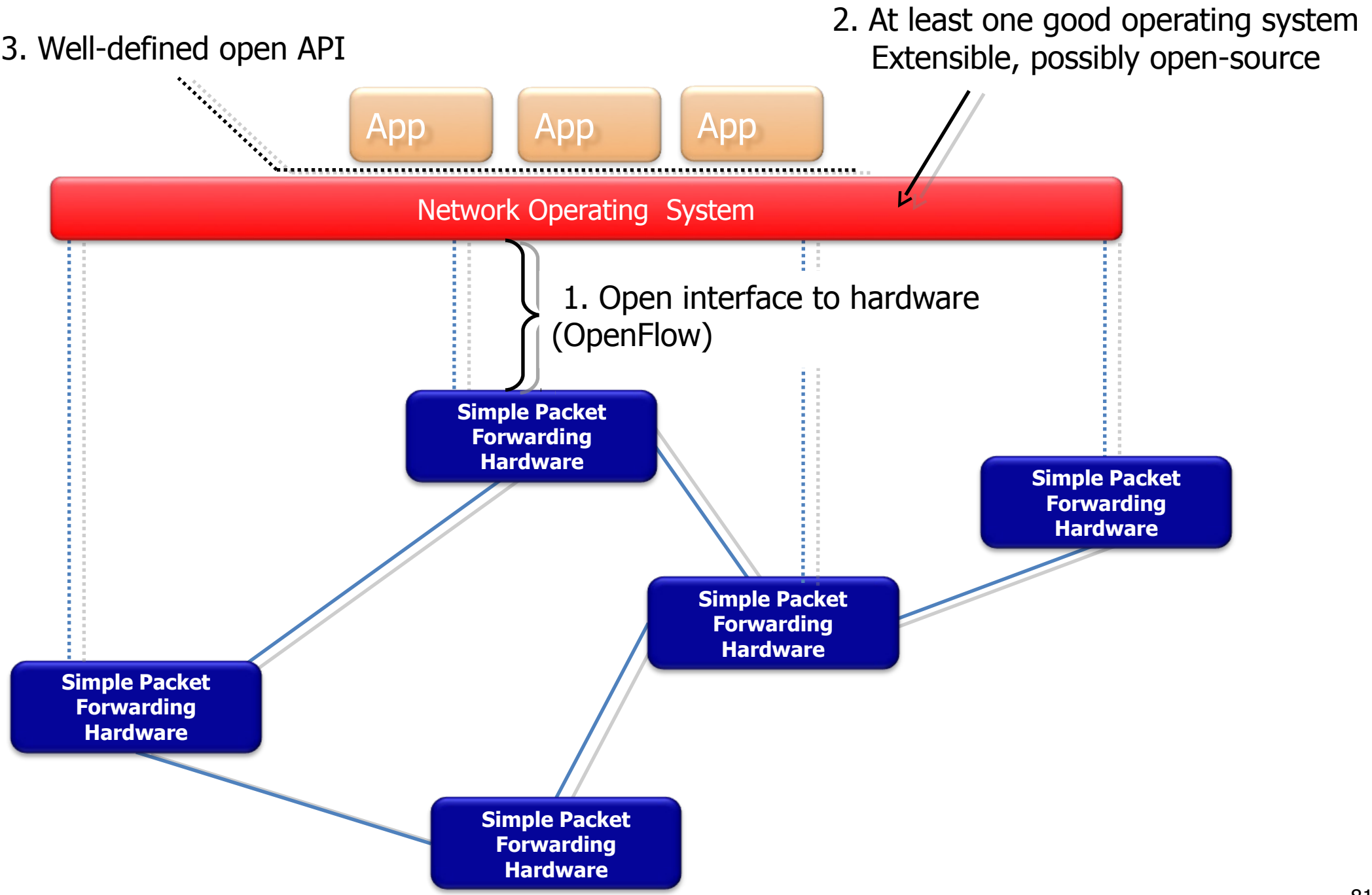


From traditional networks...

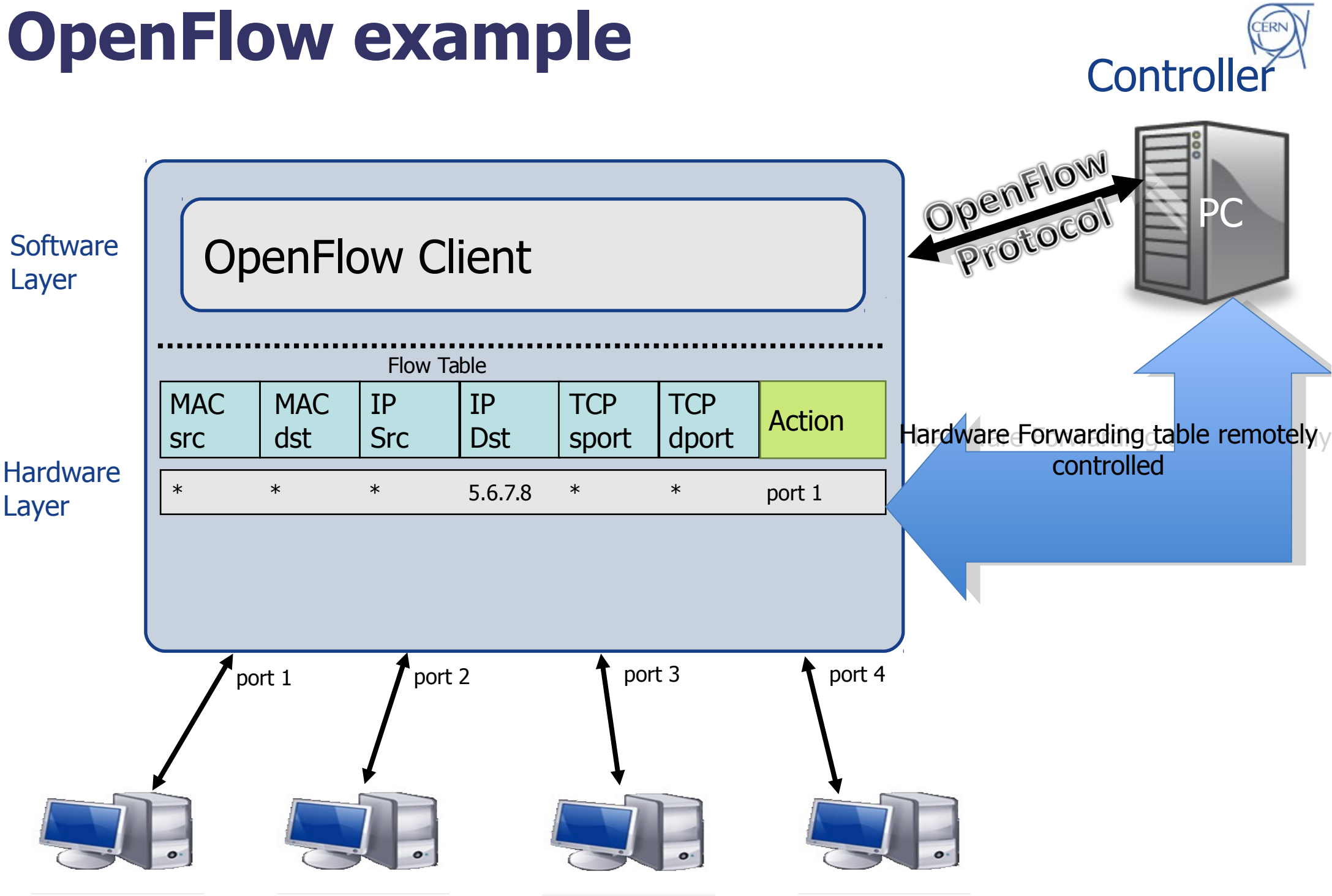
Closed boxes, fully distributed protocols



.. to Software Defined Networks (SDN)



OpenFlow example



Conclusions

Conclusions

- Telecommunication Networks are essential components of the LHC instrument

- High Reliability and Large Capacity are they key factors

- Challenging to keep up with continuously evolving technologies

Credits

Artur Barczyk (LHCONE)

Aurelie Pascal (Tetra)

Dan Savu (VISION)

Carles Kishimoto (CC 513)

David Gutierrez (Wigner)

John Shade (CIC and CIXP)

Jose Luna Durand (CSDB)

Milosz Hulboj (WIND, CINBAD)

Rodrigo Sierra Moral (Telephony, Tetra)

Ryszrard Jurga (CINBAD)

Sebastien Ceuterickx (WIND, WIFI)

Stefan Stancu (VISION)

Vlad Lapadatescu (WIND)