The HEPiX IPv6 working group

David Kelsey (STFC UKRI)

HEPiX IPv6 WG meeting, CERN, 16-17 Jan 2020

On behalf of all co-authors in the HEPiX IPv6 working group

Active in HEPiX IPv6 Working Group – last 12 months

- M Babik (CERN), M Bly (RAL), T Chown (Jisc), D Christidis (U Texas/ATLAS), J Chudoba (Prague), C Condurache (RAL/EGI.eu), T Finnern (DESY), C Grigoras (CERN/ALICE), B Hoeft (KIT), D P Kelsey (RAL), R Lopes (Brunel), F López Muñoz (PIC), E Martelli (CERN), A Manzi (CERN), R Nandakumar (RAL/LHCb), K Ohrenberg (DESY), F Prelz (INFN), D Rand (Imperial), A Sciabà (CERN/CMS)
- Many more in the past, and others join from time to time
- *and thanks also to WLCG operations, WLCG sites, LHC experiments, networking teams, monitoring groups, storage developers…*

Outline

- History
  - Phase 1 2011-2016
  - Phase 2 2016-2020
- Current status (see later talks for all the details)
- The future - Phase 3: IPv6-only networking
- Summary

Science & Technology Facilities Council | UK Research and Innovation

HEPiX

The deployment of IPv6 data storage on WLCG and UK GridPP

David Kelsey

(Head of Particle Physics Computing Group)

STFC Rutherford Appleton Laboratory

- UK Research and Innovation

Talk at UKNOF42, London, 15 Jan 2019

GridPP UK Computing for Particle Physics

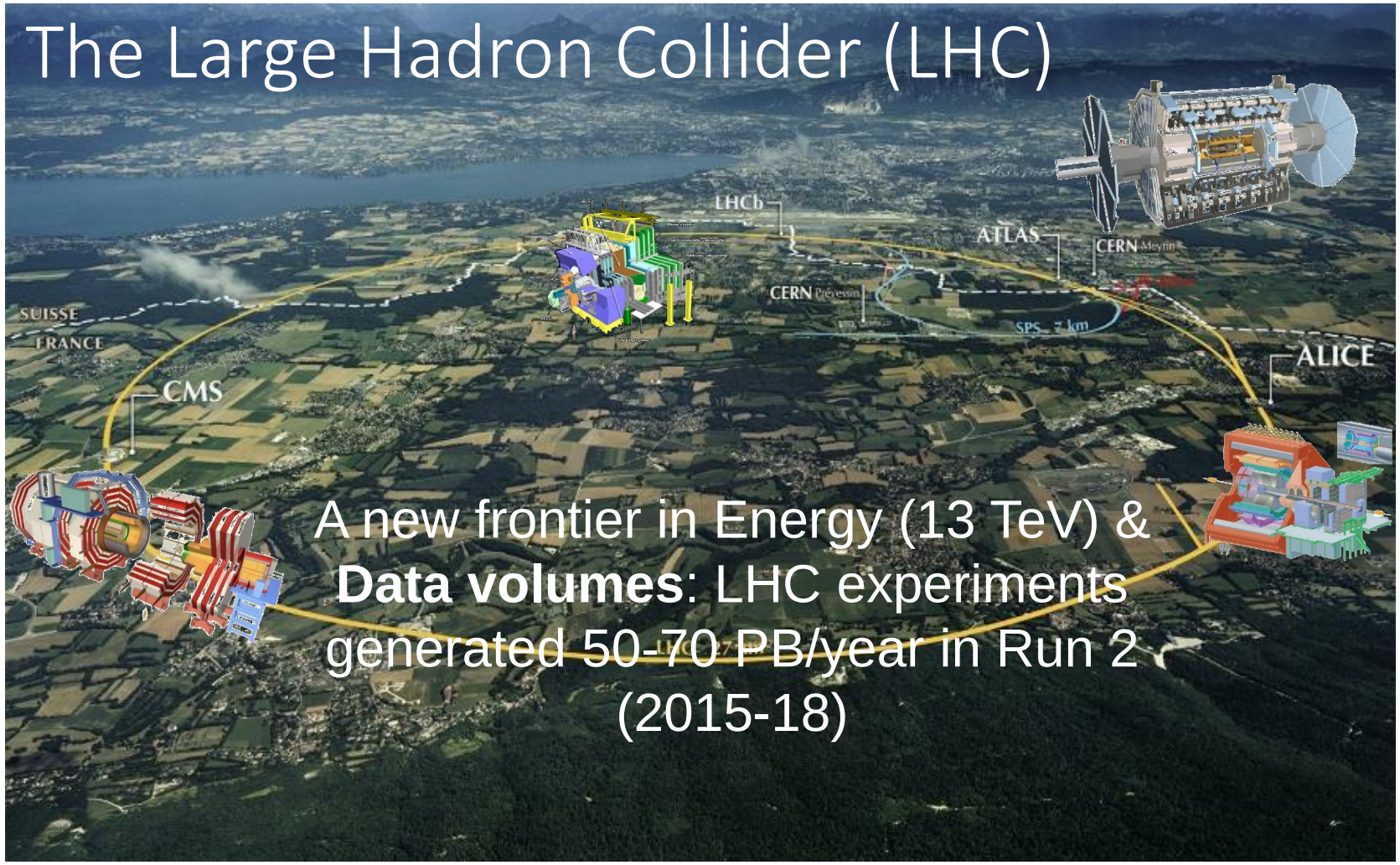
WLCG Worldwide LHC Computing Grid

David Kelsey

- Experimental particle physicist – moved to IT
- Lead computing group in Particle Physics Dept, STFC-RAL
- Trust, security & identity coordination roles in WLCG, GridPP, EGI, EOSC-hub & AARC
  - Including coordination bodies WISE, IGTF, REFEDS, …
- Chair of the HEPiX IPv6 Working Group
  - HEPiX is a worldwide body of HEP IT specialists
- *Note – during 1990s – worked on transition of worldwide HEP/ESA/NASA DECNET from Phase IV to OSI (Phase V)*

Large Hadron Collider (LHC) at CERN, WLCG & UK GridPP

The Large Hadron Collider (LHC)



A new frontier in Energy (13 TeV) &
Data volumes: LHC experiments generated 50-70 PB/year in Run 2 (2015-18)

Physics results (Run1) including…

In July 2012 >

Higgs boson-like particle discovery claimed at LHC

COMMENTS (1665)

By Paul Rincon

Science editor, BBC News website, Geneva

The moment when Cern director Rolf Heuer confirmed the Higgs results

Cern scientists reporting from the Large Hadron Collider (LHC) have claimed the discovery of a new particle consistent with the Higgs boson.

Relat

Nobel Prize in Physics 2013: F. Englert & P. Higgs

Worldwide LHC Computing Grid (WLCG)

- The WLCG is a global collaboration
- more than 170 computing centres in 42 countries
- Its mission is to store, distribute and analyse the data generated by the LHC experiments
- Sites hierarchically arranged with three tiers:
  - Tier-0 at CERN (and Wigner in Hungary)
  - 13 Tier-1s (mainly national laboratories, incl RAL in UK)
  - >150 Tier-2s (generally university physics laboratories)

WLCG Tiers Hierarchy

- **Tier-0 (CERN and Hungary): data recording, reconstruction and distribution**
- **Tier-1: permanent storage, re-processing, analysis**
- **Tier-2: Simulation,**
- **end-user analysis**

- **~750k CPU cores**
- **~ 1 EB storage**
- **> 2 million jobs/day**
- **10-100 Gbps links**

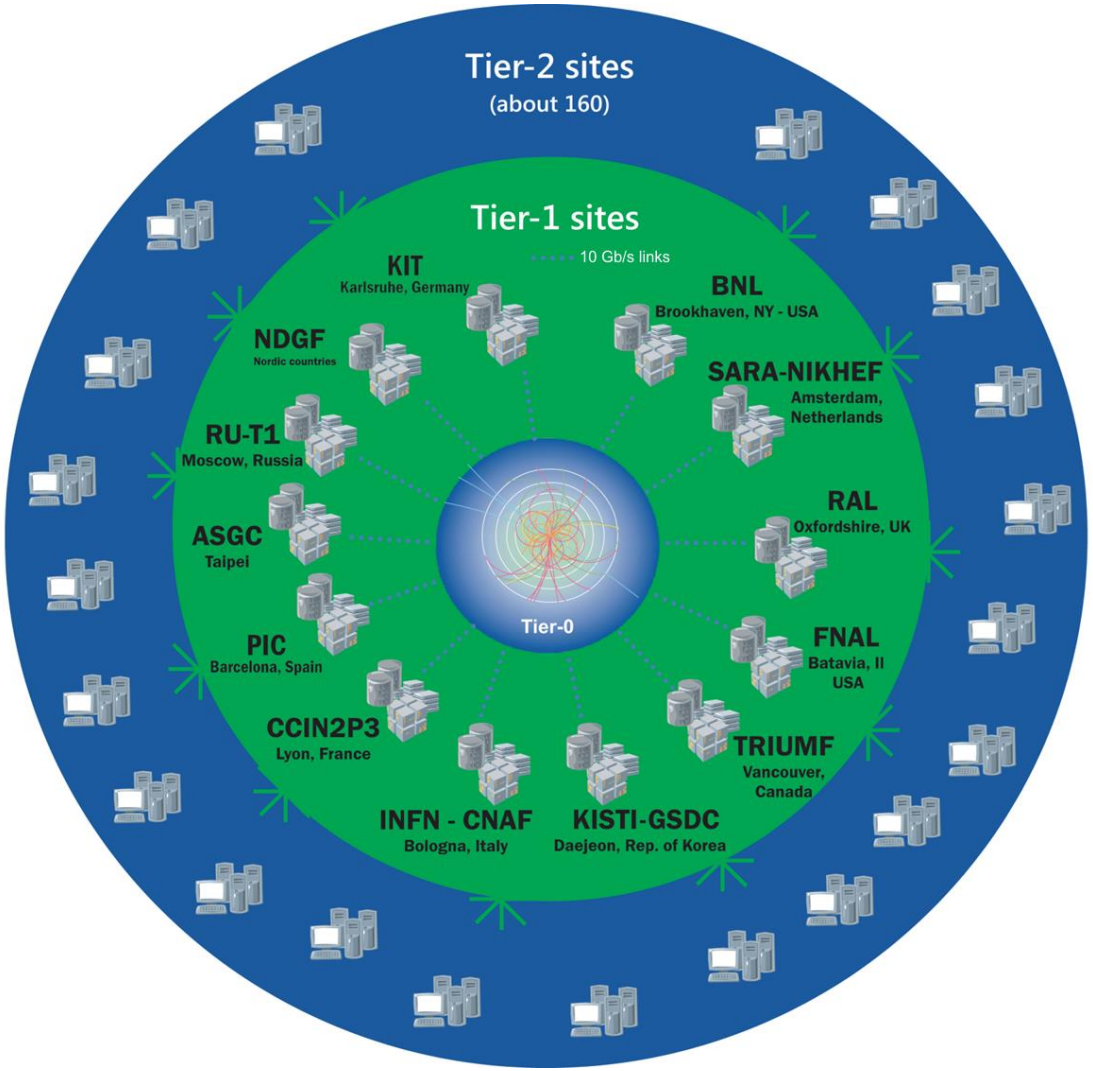


Image from 2014

WLCG sites

WLCG Data Transfers

Data transfers in WLCG

- From Tier-0 to Tier-1s
- From Tier-1s to Tier-2s
- Requirements – Fast and reliable!
- Multiple protocols and implementations, but the standard approach is:

FTS3 and GridFTP

Bulk data transferred between storage clusters with the File Transfer Service (FTS3) using GridFTP

- Also data transfer from federated data storage using a HEP-specific protocol called XrootD
- direct access to data by an analysis job at one site from storage at another

Why should WLCG use IPv6?

Why IPv6?

- Survey of 18 major HEP sites (Sep 2010) – IPv6 readiness
  - National NRENs ready, Universities and Labs not ready
  - Some reported lack of IPv4 address space, including CERN
- HEPiX meeting – Cornell, Ithaca NY – Nov 2010
  - Projected IANA IPv4 address exhaustion
  - Sep 2010 – memo from US Federal CIO to all Exec depts (incl DOE)
- Offers of opportunistic CPU resources which could be IPv6-only
  - Experiments want to be able to use them
- Recognition that much of our middleware, software and technology was not yet IPv6 capable
- HEPiX decided to create a working group (started April 2011)
  - No specific funding – but motivated, competent volunteers!

Preparatory work during 2011-2016

HEPiX IPv6 Working Group

- Phase 1 – full analysis of work to be done
  - Applications, system and network tools, operational security
  - Create and operate a distributed test-bed
    - No interference with WLCG production data analysis!
  - Propose timetable and plan for transition

2012

- CERN announces shortages of routable IPv4 addresses
  - explosion of virtualisation
- Active HEPiX IPv6 test-bed with ~ 12 sites
  - engagement of all 4 LHC experiments
- Testing regular data transfers across the testbed
- Testing dual-stack services (production) at Imperial College London
- Concluded not able to support IPv6-only clients until at least 2014

At CHEP2013 conference

- > 2 PB data transferred over IPv6 in last 6 months
- Success rate > 87%
- Very High!

GridFTP IPv6 data transfer mesh

2013-14 Data Management

- Testing the important data transfer protocols, technology and data storage/file systems
  - For IPv6-readiness
- GridFTP, DPM, dCache, xRootD, OpenAFS, FTS, CASTOR
  - Found many problems needing work
    - Worked closely with developer community
- Concluded IPv6 support will be much later than 2014!

2015

- At CHEP conference in April 2015
  - 75% of Tier-1 sites are IPv6-ready (but only 20% of Tier2)
  - 10% of sites now reporting lack of IPv4 addresses
- Most important IPv6-only use case
  - Sites, Clouds providing CPU (virtual machines)
  - Opportunistic resources may be IPv6-only
  - Need dual-stack federated storage services
  - And dual-stack central WLCG and Experiment services

The transition 2016-2020

2016

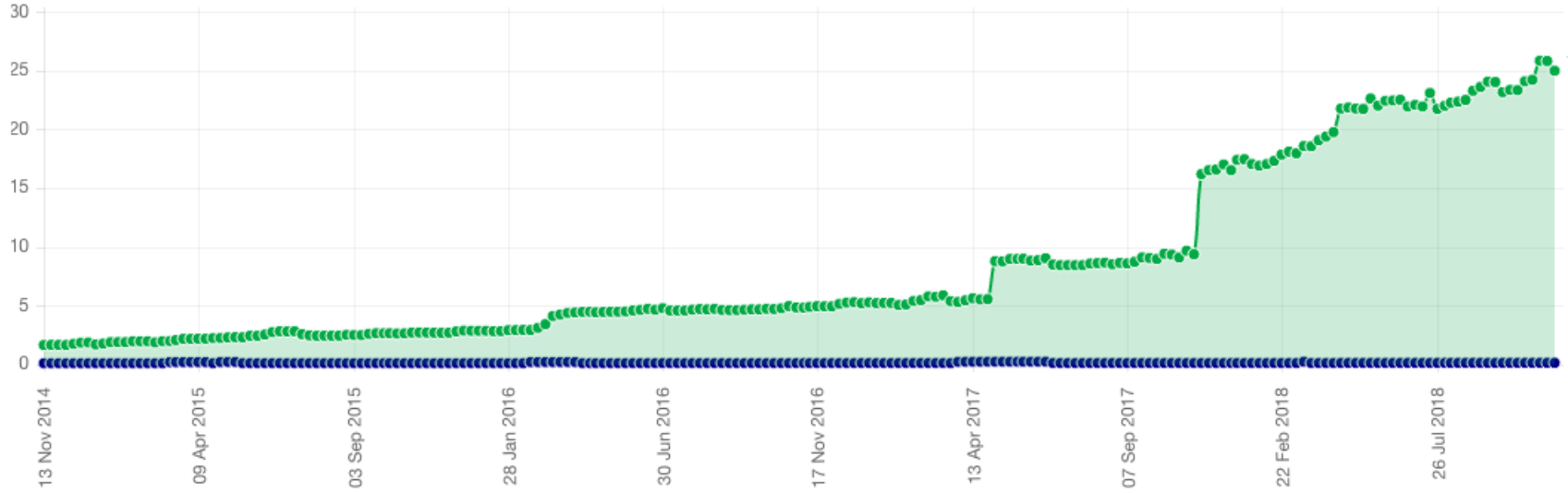
- Continue to push for
  - deployment of production dual-stack data services
  - LHCOPN (Tier0-Tier1 private network)
    - IPv6 peering everywhere
- perfSONAR – end to end network monitoring – dual-stack
- Move central services and central monitoring to IPv6
- Wrote guidance on IPv6 security for WLCG sites
- Deployment timetable approved by WLCG Management Board (Sep 2016)

WLCG – IPv6 deployment

Plan approved by WLCG Management Board

- April 2017 – support for IPv6-only CPU starts
  - Tier-1s to provide dual-stack storage (in testbed)
- April 2018
  - Tier-1 dual-stack storage in production mode
- By end of LHC Run 2 (end 2018)
  - A large number of Tier-2s to migrate storage to IPv6
  - All requested to do this

Growth of dual-stack hosts in the WLCG (Jan 2019)



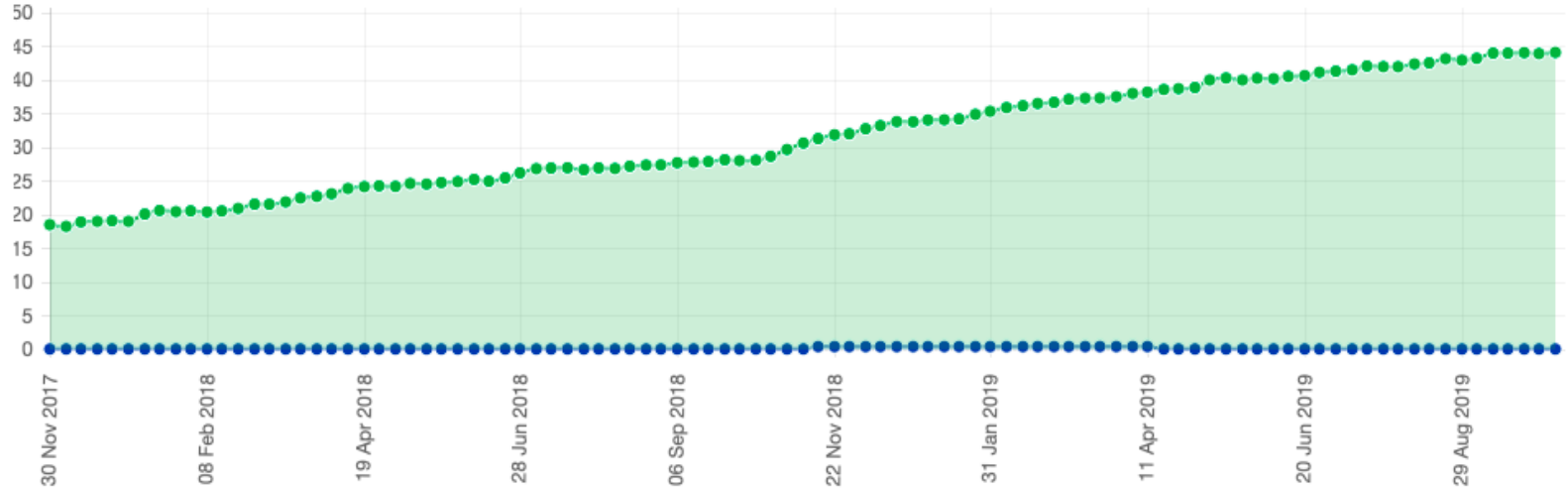
All services, not just storage

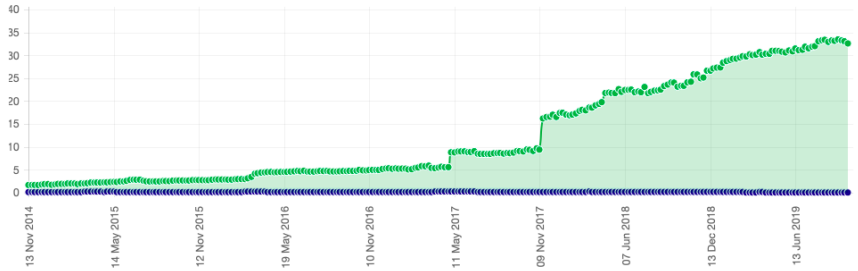
Fraction of endpoints listed in the CERN central BDII (lcg-bdii.cern.ch) where the DNS returns a dual-stack IPv6-IPv4 (A+AAAA) resolution (green line) or an IPv6-only resolution (blue line).
(http://orsone.mi.infn.it/~prelz/ipv6_bdii/).

WLCG services status (dual-stack) OCT 2019

all services, not just storage

results from the LHC Experiment VO feeds (https://orsone.mi.infn.it/~prelz/ipv6_vofeed/) (~44%)

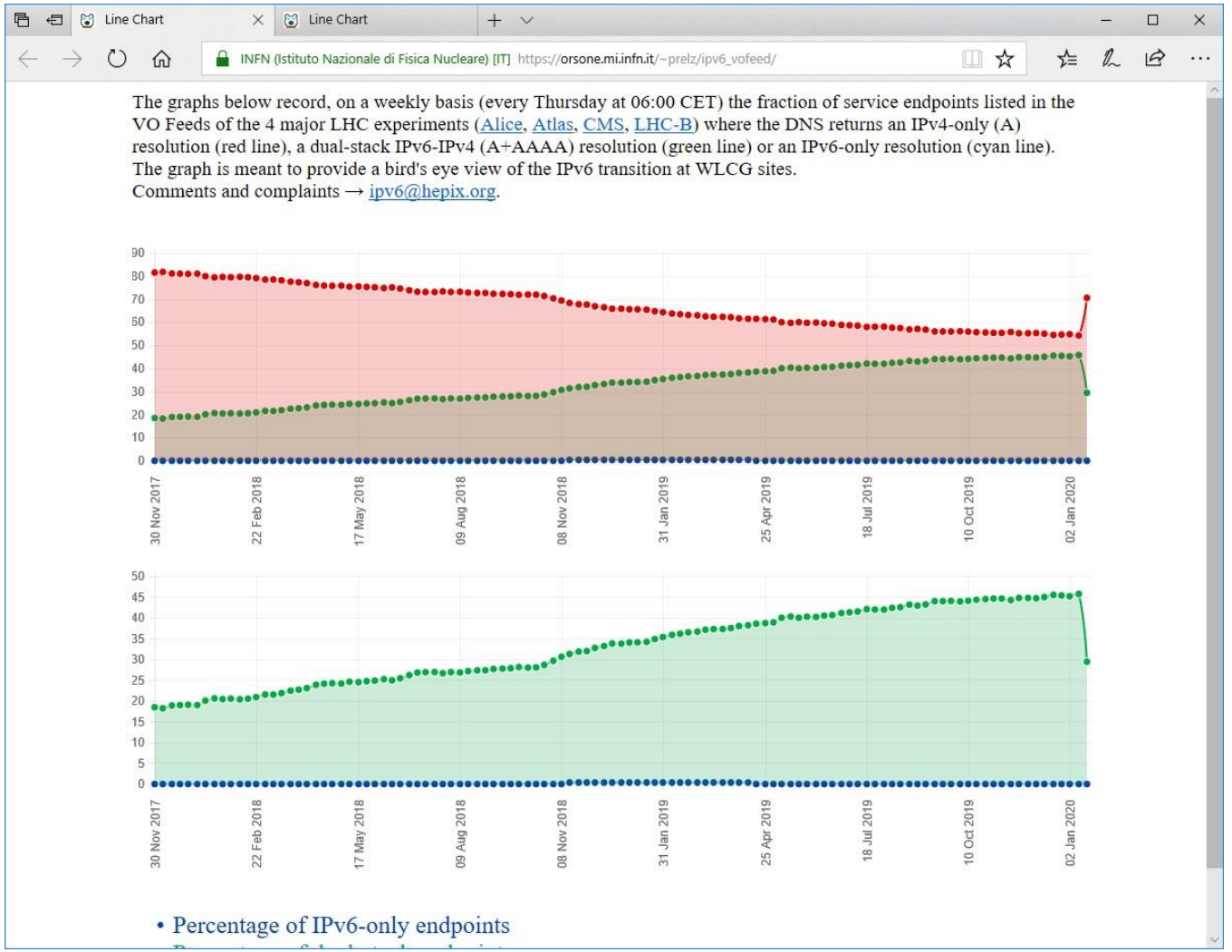




Francesco Prelz

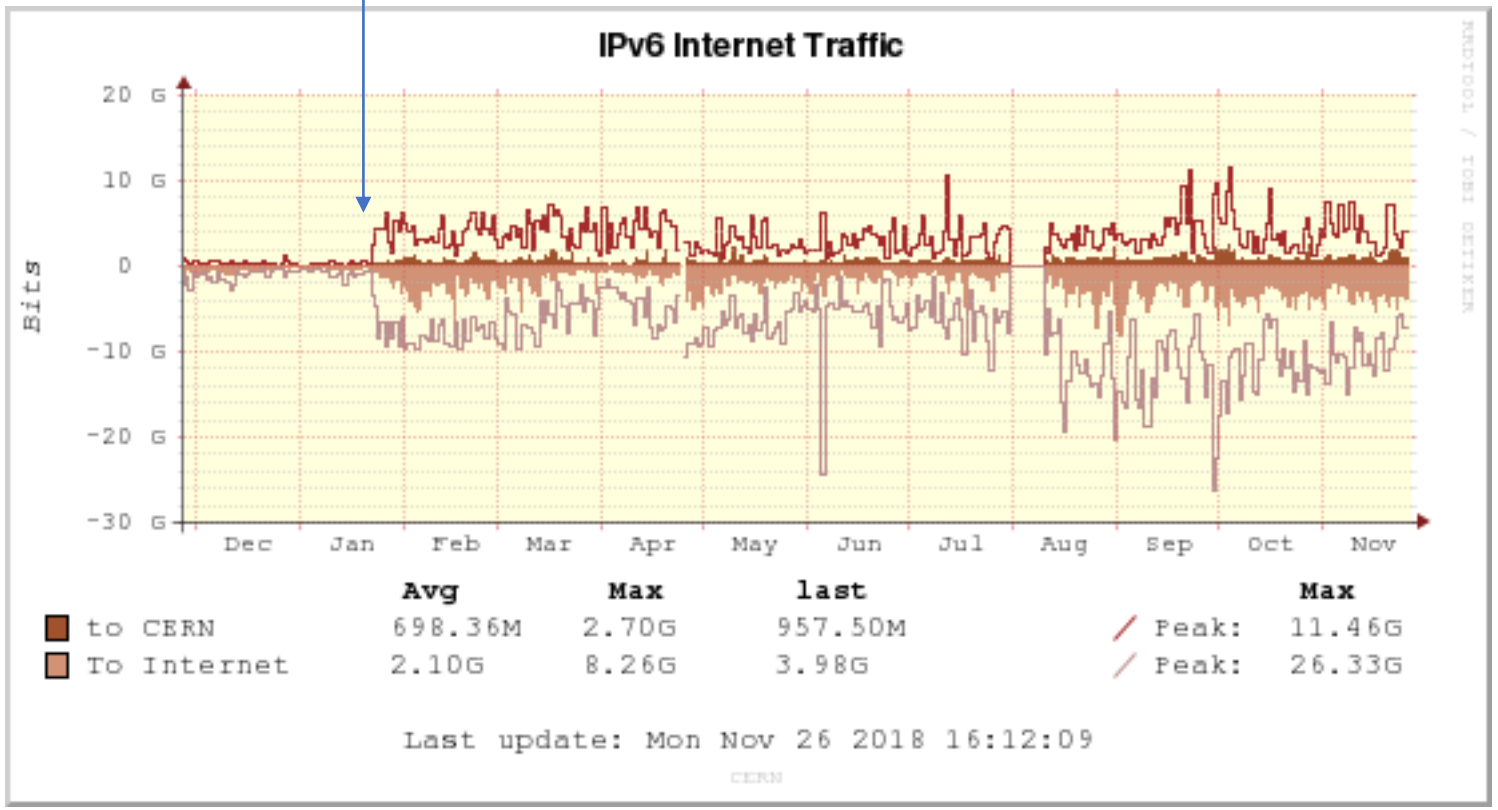
Fraction of endpoints listed in the CERN central BDII (lcg-bdii.cern.ch) where the DNS returns a dual-stack IPv6-IPv4 (A+AAAA) resolution (green line) or an IPv6-only resolution (blue line). (https://orsone.mi.infn.it/~prelz/ipv6_bdii/).

And now today!



Turning on IPv6 on CERN Tier-0 disk storage (EOS) in Jan 2018

Non-LHCOPN/non-LHCONE traffic



Not everything went smoothly!

Problems & lessons learned

- Many blocking issues outside of our own control
  - Both software and site networking teams
- Developers claim that software is fully IPv6-compliant!
- Software/protocols fixed-size storage for IP addresses
- Software/protocols assume single address (as in IPv4)
- Performance differences between IPv4 & IPv6
  - IPv6 must perform at least as well
- Have to understand cases where fraction of IPv6 is smaller than expected
  - Preference for IPv6 over IPv4 must be established
- Can be lots of development effort and testing is not easy when no other positive change re functionality
- Sys admins, operations staff, security team, developers
  - All need TRAINING and experience

Tier-0/Tier-1/LHCOPN/LHCONE status (see talk by Bruno Hoeft)

Network and pS at Tier-1's (Oct 2019)

Bruno Hoeft

- All sites connected to LHCONE
- TW-ASGC Perfsonar server currently down
- RRC-KI-T1 is connected to with IPv6 to LHC[OPN/ONE], the perfsonar server is running from beginning of this week (Oct. 14)

improved

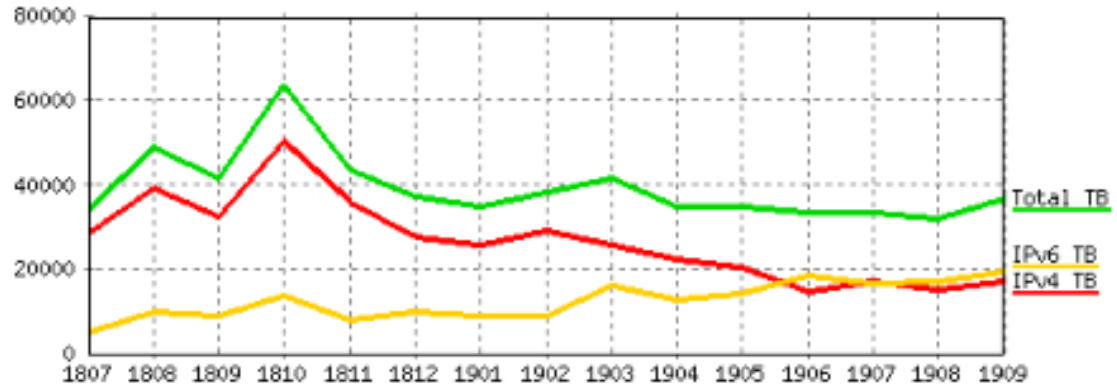
Tier-1	LHCOPN	LHCONE	IPv6 Perfsonar
CA-TRIUMF	OK	OK	LHC[OPN/ONE]
CH-CERN (Tier-0)	OK	OK	LHC[OPN/ONE]
DE-KIT	OK	OK	LHC[OPN/ONE]
ES-PIC	OK	OK	LHC[OPN/ONE]
FR-CCIN2P3	OK	OK	LHC[OPN/ONE]
IT-INFN-CNAF	OK	OK	LHC[OPN/ONE]
KR-KISTI	OK	OK	LHC[OPN/ONE]
NGDF	OK	OK	LHC[OPN/ONE]
NL-T1 - NIKHEF	OK	OK	--
NL-T1 - Sara-Matrix	OK	OK	LHC[OPN/ONE]
RRC-KI-T1	OK	OK	LHC[OPN]
RRCC-JINR-T1	OK	OK	LHC[OPN/ONE]
TW-ASGC	OK	OK	--
UK-T1-RAL	OK	OK	LHC[OPN]
US-T1-BNL	OK	OK	LHC[OPN/ONE]
US-T1-FNAL	OK	OK	LHC[OPN/ONE]

IPv6 traffic on LHCOPN & LHCONE at CERN

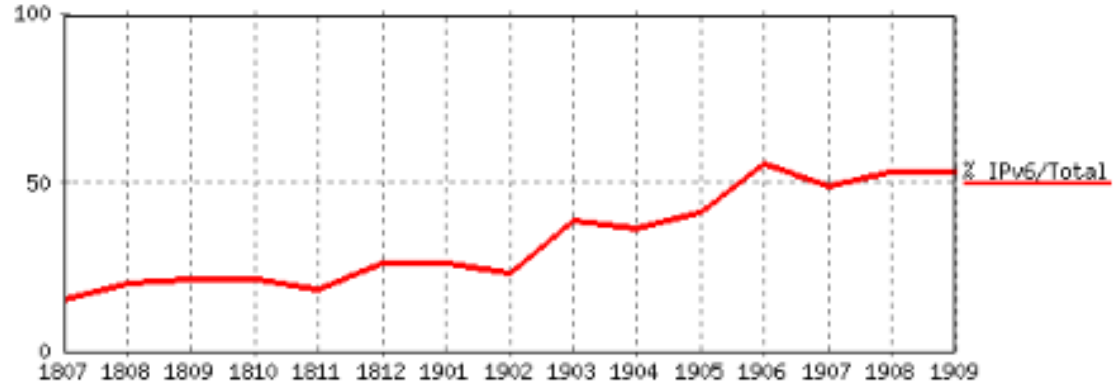
LHCOPN and LHCONE IPv4 and IPv6 traffic volumes seen at CERN Tier0

Edoardo Martelli

IPv4 and IPv6 traffic volumes month by month



Percentage of IPv6 traffic over the total



IPv6 traffic on LHCOPN & LHCONE as seen at CERN

- > 50% of all traffic is IPv6
- From June 2019 onwards

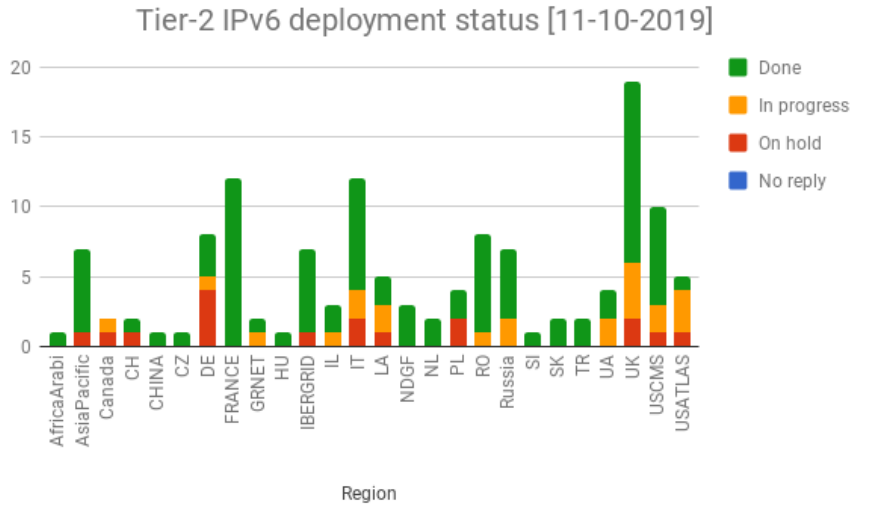
LINK to these plots

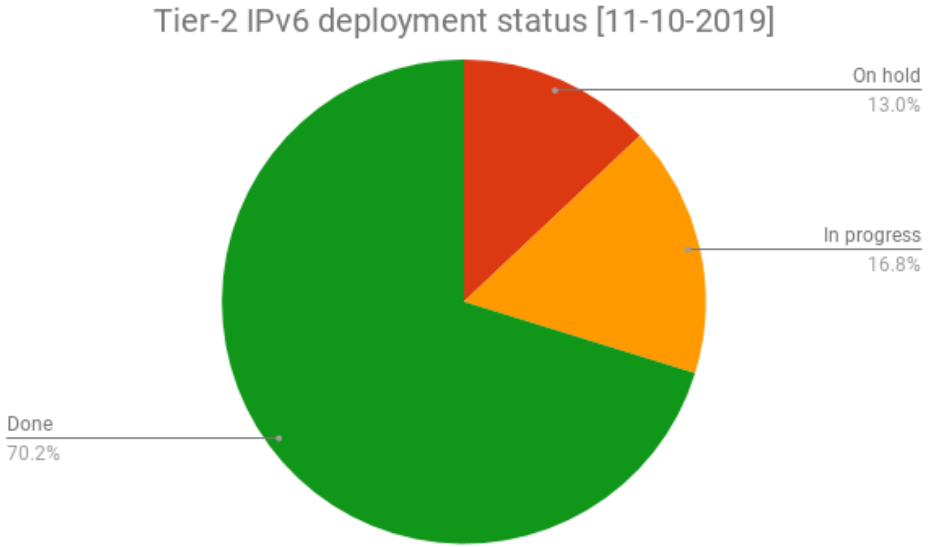
Tier2 status
(see talk by Andrea Sciaba)

Tier-2s: GGUS tickets to all Tier-2 sites

Andrea Sciaba

- WLCG set a target for end 2018 for deploying IPv6 on storage systems (and perfSONAR)
- The deployment campaign was launched in November 2017
  - GGUS tickets sent to all non-US sites
  - US sites are tracked via the experiments
  - Sites made aware of the WLCG plans and asked to report plans and give updates
- Steady progress (status)
  - About 70% of T2 sites have storage on dual stack



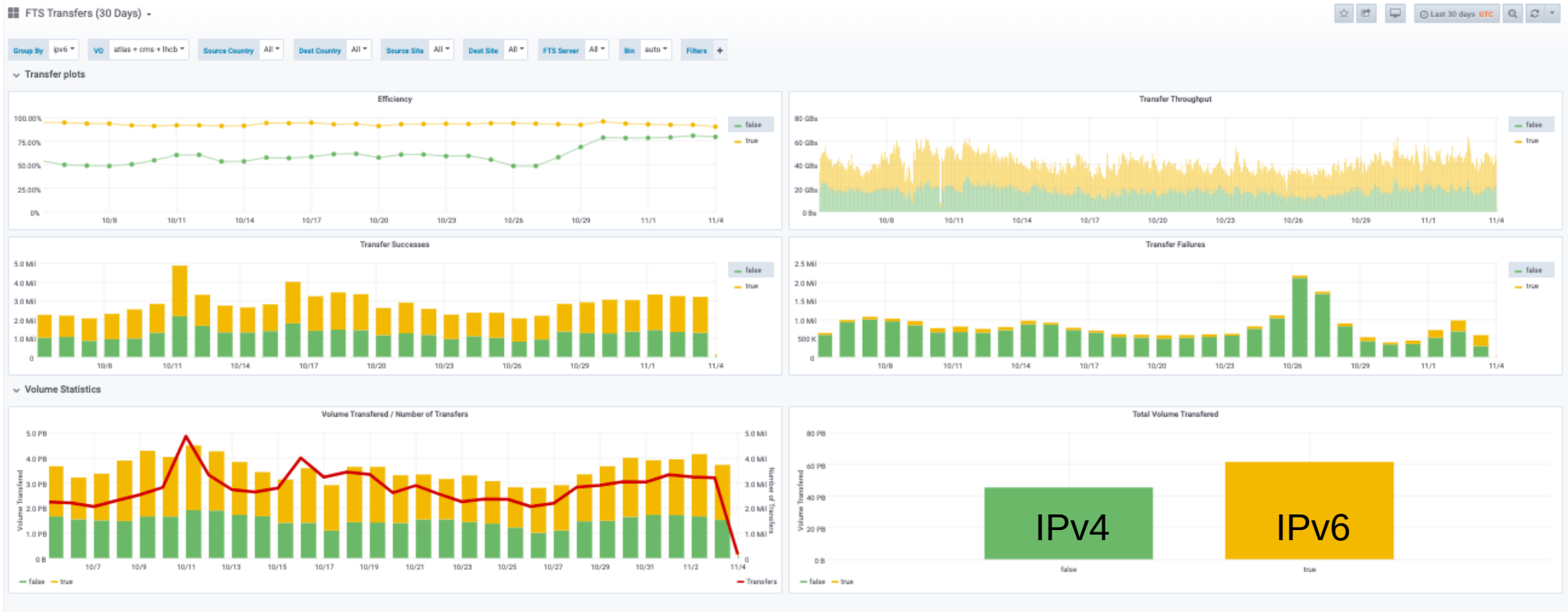


Data transfers over IPv6 (FTS)

(Note: xRootD and HTTP transfers not yet instrumented to track IPv6 vs IPv4, but then only a small fraction of FTS transfers use xRootD or HTTP)

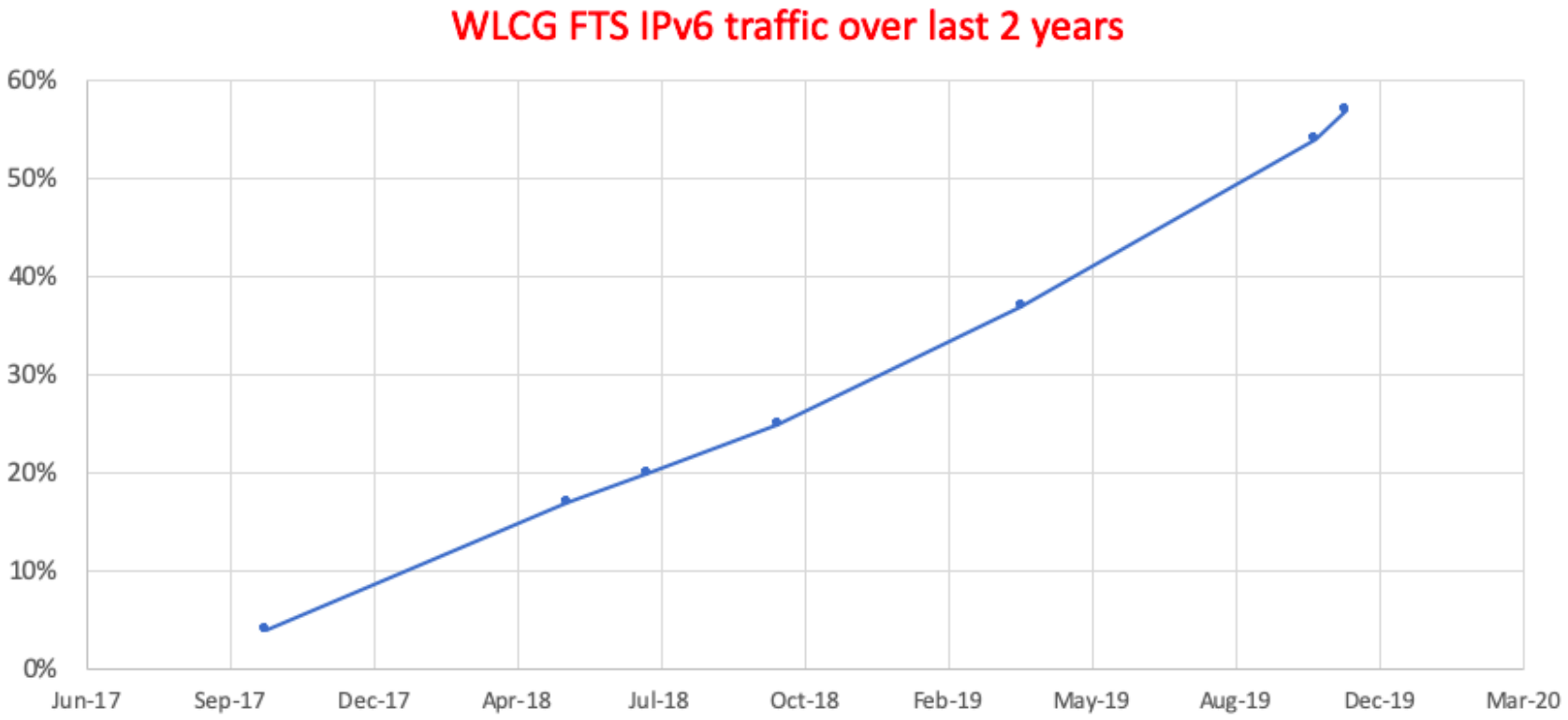
FTS transfer monitoring - last 30 days (7 Nov 2019)

Approximately 57% of data transferred via FTS in the last 30 days went over IPv6



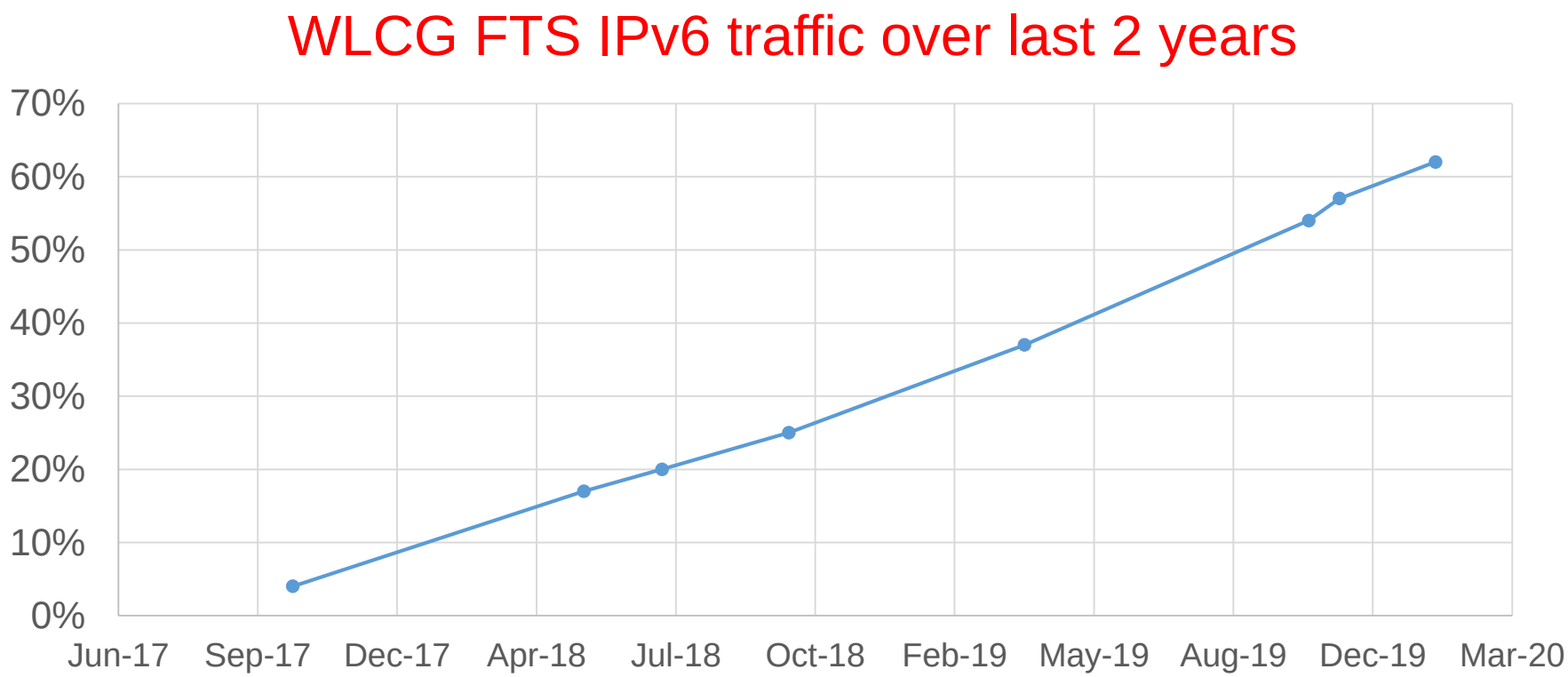
https://monit-grafana.cern.ch/

% of FTS traffic over IPv6 - last 2 years



Data points:
Reports at HEPiX meetings and CHEP18/CHEP19

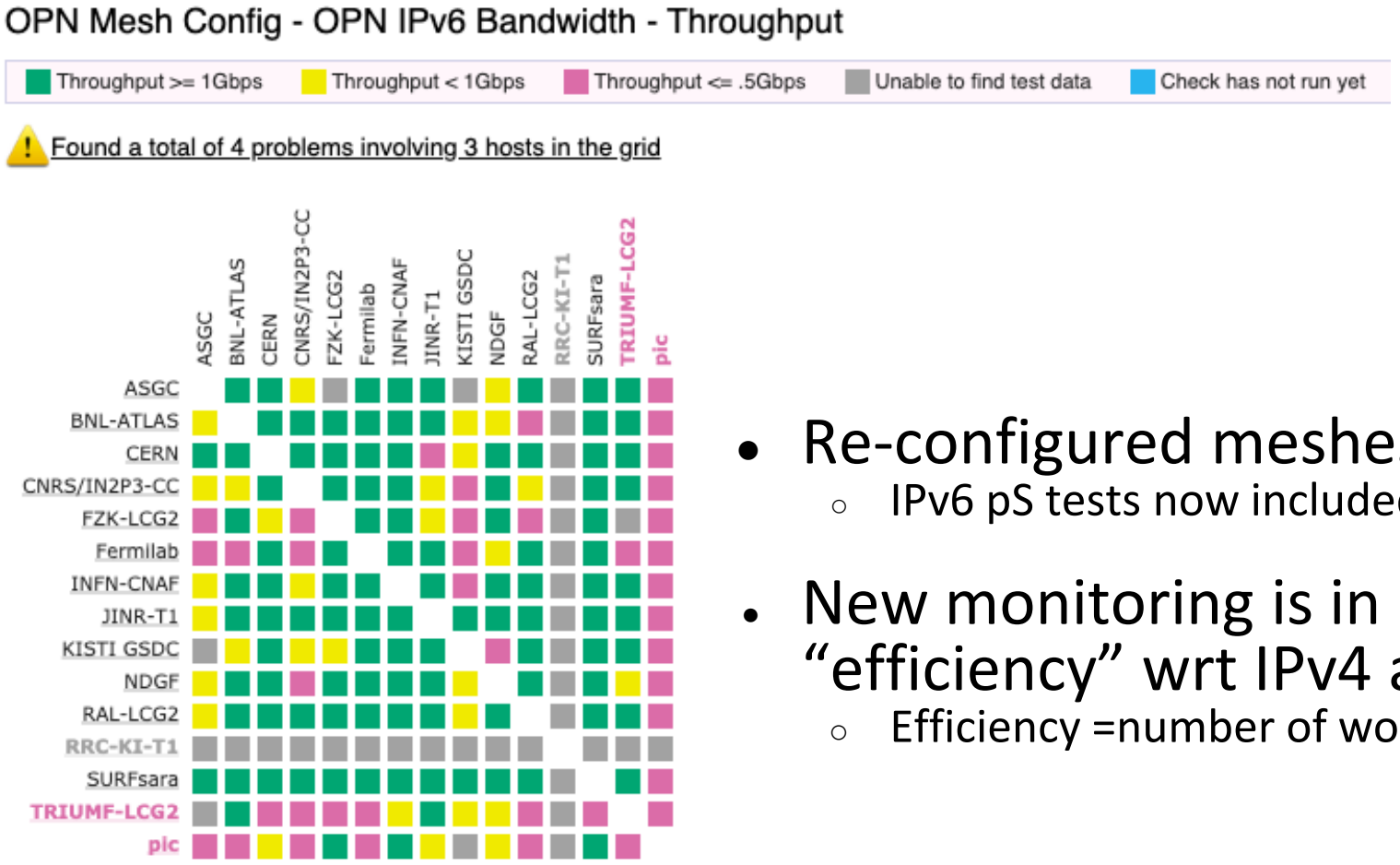
%FTS – yesterday (62%)



Monitoring (ETF, perfSONAR)
(see talk later – Marian Babik/Duncan Rand)

perfSONAR IPv6 Mesh

Marian Babik & Duncan Rand



[LINK](#) to dashboard

- Re-configured meshes
  - IPv6 pS tests now included in all meshes
- New monitoring is in place which reports test “efficiency” wrt IPv4 and IPv6
  - Efficiency =number of working destinations / total destinations

Future plans

IPv6-only networking

- Our main use case - ready for use of (opportunistic) IPv6-only CPU

- BUT there are other drivers for IPv6-only:
  a) new CERN machine room and lack of public IPv4 addresses
     - Use only IPv6 addresses for external public networking?

  b) multiONE (different communities using LHCONE)
     - multiple overlay networks
       - different addresses for each community
     - sites likely have lack of sufficient IPv4 address space

IPv6-only networking (2)

- Running a dual-stack IPv4/IPv6 infrastructure is complex
- Large companies (e.g. Facebook, EE/BT) use IPv6-only internally
  - Then use tools like NAT64/DNS64/464XLAT for legacy world
- CERN EOS infrastructure also uses IPv6-only internally

When/how do we simplify and move to IPv6-only in WLCG?

IPv6 working group Phase 3

- The fraction of data transfers on IPv6 is getting much larger (>50%)
- When the amount of IPv4 traffic on LHCOPN is close to zero
  - Turn off IPv4 entirely on LHCOPN?
  - simplify routing tables and tracking problems is easier
- MultiONE/LHCONE may also be using IPv6-only

IPv6-only networking on WLCG?

- Fixing dual-stack endpoints that prefer to use IPv4 and not IPv6
- Turning off IPv4 at a Site is clearly their own decision
  - they may have many other needs (not WLCG) for IPv4
- We need to include experience of sites already doing IPv6-only CPU
  - UKI-LT2-Brunel - successful IPv6-only cluster for LHCb, ATLAS, CMS & LSST
  - Also SiGNET, T2_US_Nebraska, UKI-T2-QMUL, …
  - More testing is essential
- Transition tools such as NAT64 can be used once core is IPv6-only

- WLCG may need a date for “end of support” for IPv4-only clients
  - e.g. start of LHC Run4?

Summary

- WLCG is ready to support use of IPv6-only CPU resources
  - Good steady progress towards this goal!
- Tier-1s should all have production storage accessible over IPv6
  - 96% of Tier-1 storage is available via IPv6
- Tier-2s 70% sites done
  - 73% of Tier-2 storage is dual-stack
- ~62% of FTS transfers today over IPv6
- ~55% LHCOPN+LHCONE traffic observed at CERN is IPv6
- WG Phase 3 - we are planning for move to IPv6-only services
- ***message to new research communities - build on IPv6 from start***

Questions?