

IPv6 in the WLCG

Duncan Rand
Imperial College London

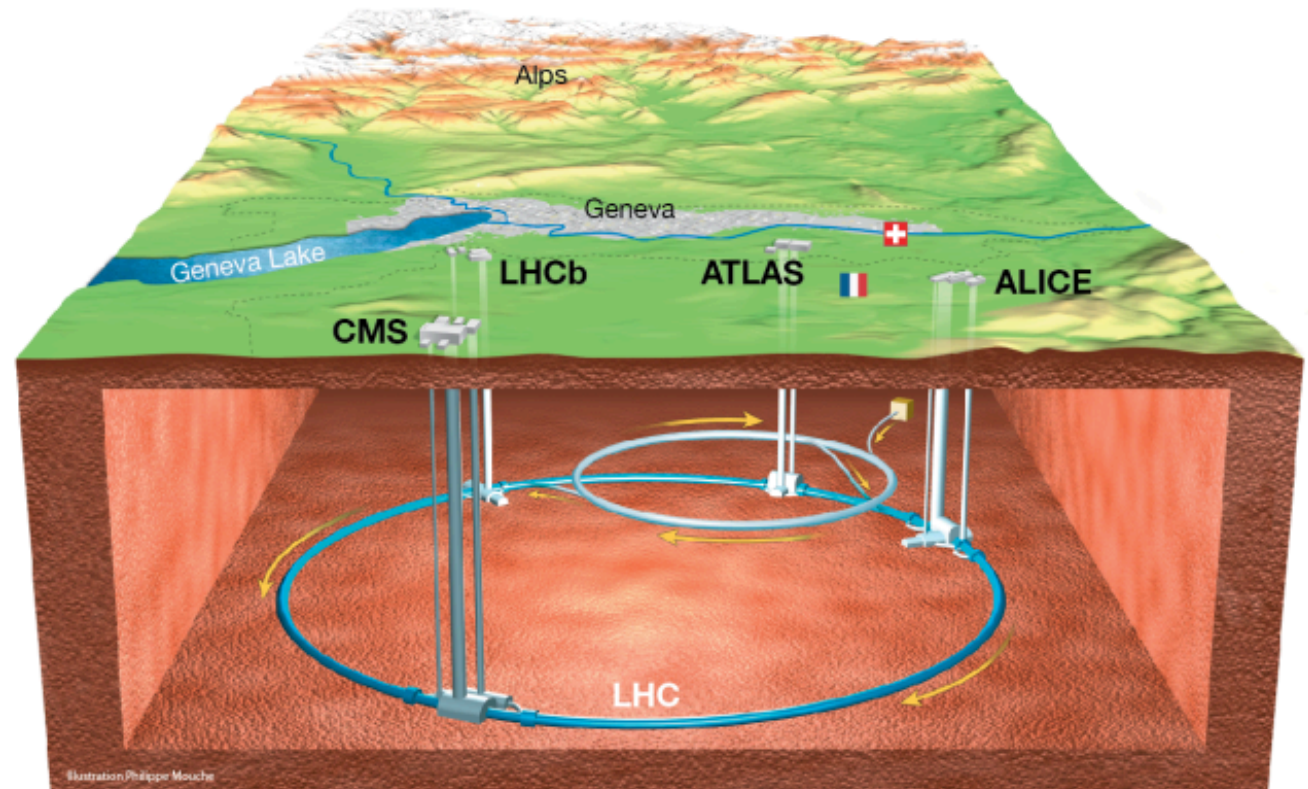
UK IPv6 Council Meeting, London, December 2018

Contents

- The Large Hadron Collider (LHC)
- The Worldwide LHC Computing Grid (WLCG)
- GridPP
- Why IPv6?
- IPv6 deployment plan
- Status of the WLCG sites
- Data transfer with the File Transfer Service (FTS)
- Network monitoring with perfSONAR
- Summary

The Large Hadron Collider (LHC)

- The LHC is located at CERN on the Franco-Swiss border
- Proton proton and heavy ion collider with four main experiments
- Two general purpose: ATLAS and CMS
- Two specialist: LHCb and ALICE (heavy ions)
- During Run 1 at 8 TeV: found the Higgs particle in 2012
- Started Run 2 in 2015 at 13 TeV, just finished it on Monday 3rd December
- Computing for LHC experiments carried out by the Worldwide LHC Computing Grid (WLCG or 'the Grid')

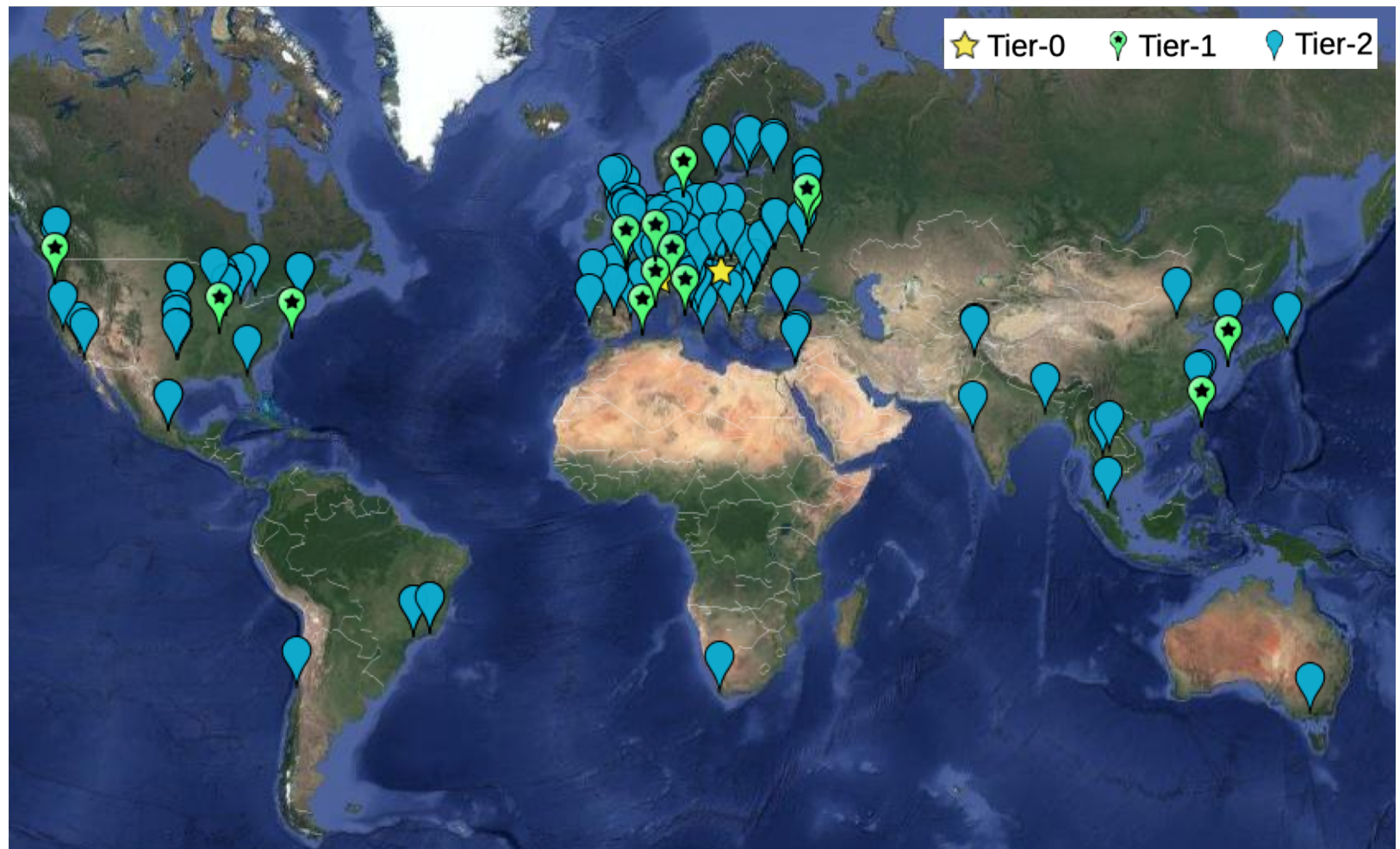


© 2014-2018 CERN

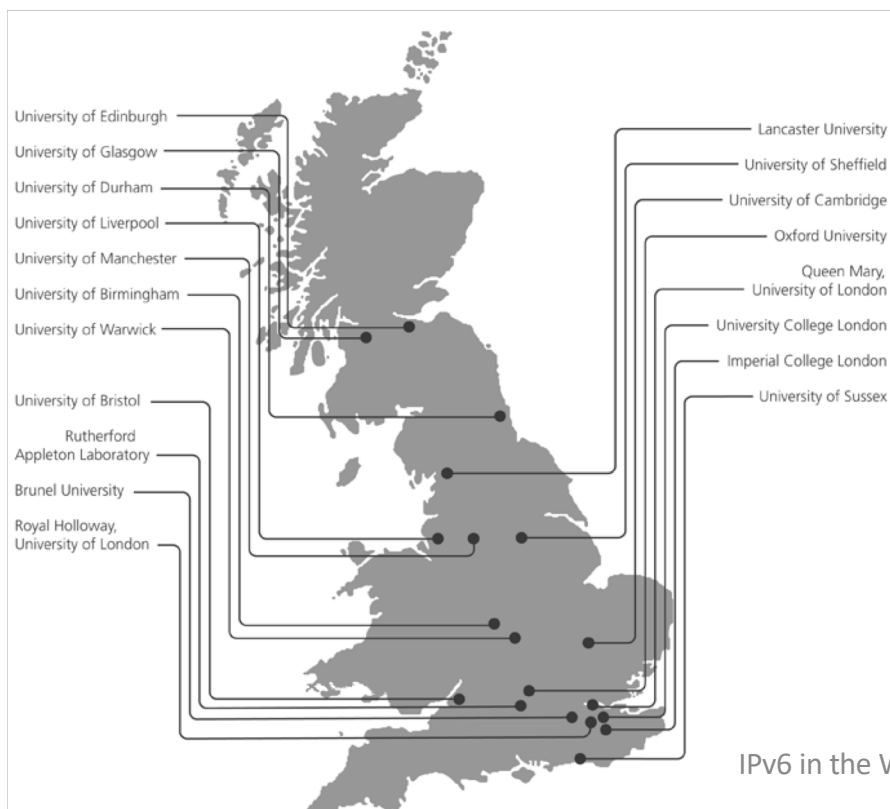
Worldwide LHC Computing Grid (WLCG)

- The WLCG is a global collaboration of more than 170 computing centres in 42 countries.
- Its mission is to provide global computing resources to store, distribute and analyse the ~50-70 petabytes of data generated per year by the LHC experiments

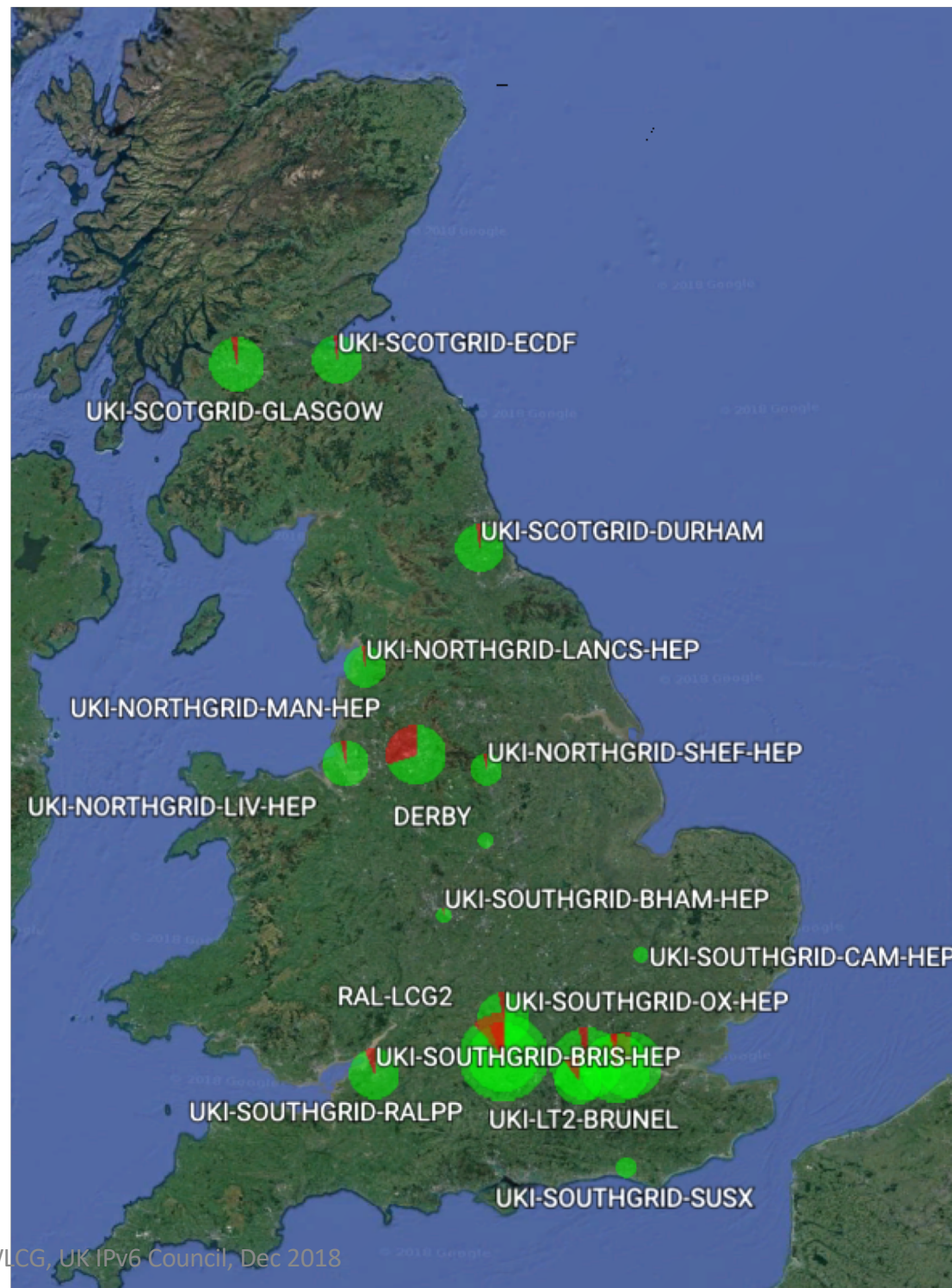
- Sites hierarchically arranged
- Tier-0 at CERN (and Wigner in Hungary)
- 14 Tier-1s (mainly national laboratories)
- 149 Tier-2s (generally university physics laboratories)



GridPP is a collaboration of nineteen institutes providing data-intensive distributed computing resources for the UK High Energy Physics community and the UK contribution to the WLCG



IPv6 in the WLCG, UK IPv6 Council, Dec 2018



WLCG Tiers Hierarchy

- Initial modelling of LHC computing requirements suggested a hierarchical tier-based data management and transfer model
- Data exported from Tier-0 at CERN to each Tier-1 and then on to Tier-2s
- However better than expected network bandwidth means that the LHC experiments have been able to relax this hierarchy
- Now data is transferred in an all-to-all mesh configuration
- Data often transferred across multiple domains
- e.g. a CMS transfer to Imperial College London might come from Fermilab near Chicago

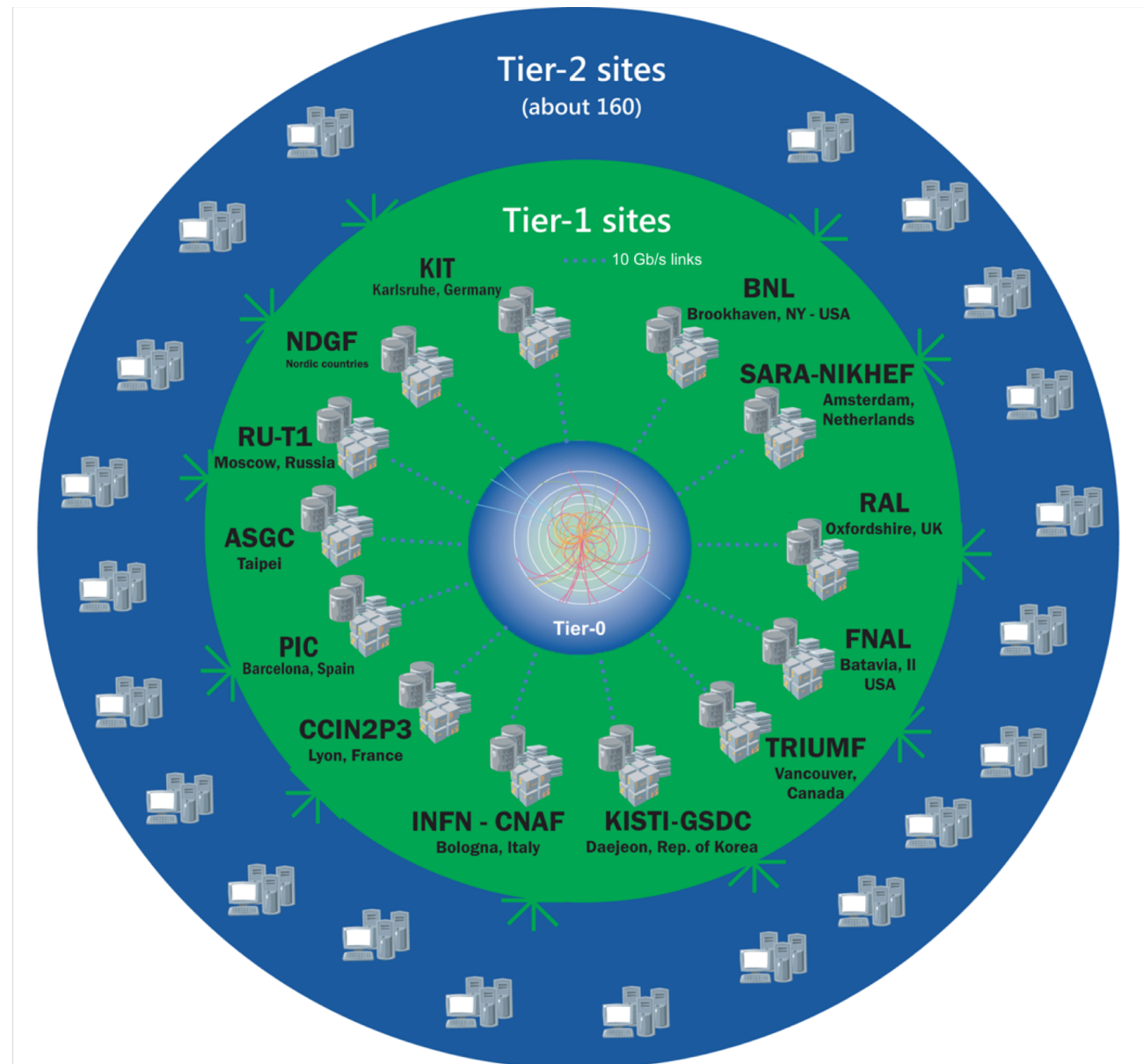


Image from 2014

WLCG Site Operation

- WLCG resources at a site generally consist of
 - a large compute cluster (typically several thousand cores)
 - a disk storage cluster (typically a few petabytes)
- Bulk data is transferred between storage clusters with the File Transfer Service (FTS3) using GridFTP
- Computing jobs arrive at the site and produce simulated data or process some of the data stored locally
- Also possible for a job at one site to access data directly from storage at another
- For example QMUL reads CMS data from storage at Imperial College
- It is envisaged that the use of such remote reading of data is likely to increase in the future

Why IPv6?

- The WLCG is generally open to new compute resources
- We might get an offer of opportunistic CPU resources which are IPv6-only – want to be able to use them
- So the main goal is to make the data at the sites accessible by clients running on IPv6-only machines
- Also for pledged resources, sites running out of IPv4 addresses and to avoid use of NAT
- Initial deployment plan
 - Make experiment central services dual-stack
 - Make some test worker nodes IPv6-only
 - Deploy perfSONAR network monitoring at sites
 - Make site storage accessible over IPv6

WLCG deployment plan: timeline

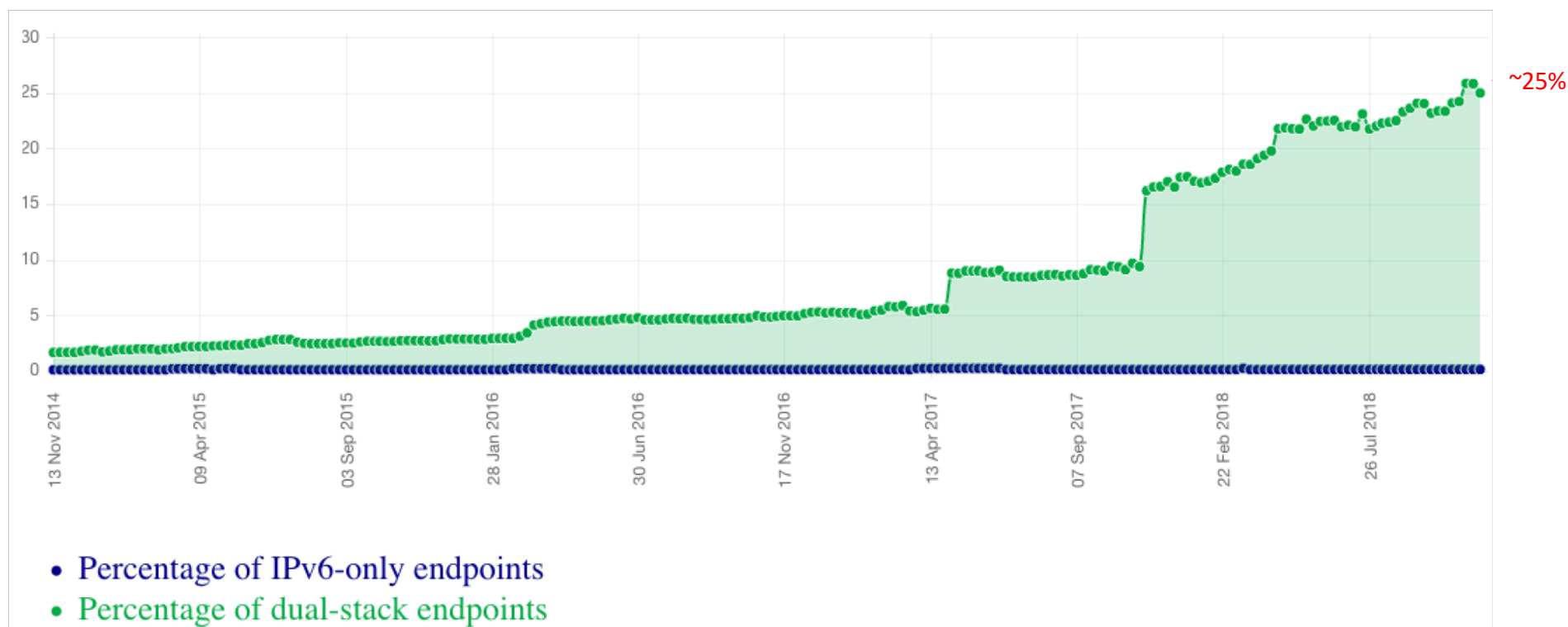
- By April 1st 2017
 - Sites can provide IPv6-only CPUs if necessary
 - Tier-1's must provide dual-stack storage access with sufficient performance and reliability
 - At least in a testbed setup
 - Stratum-1 service at CERN must be dual-stack
- By April 1st 2018
 - A dedicated ETF infrastructure to test IPv6 services must be available
 - ATLAS and CMS must deploy all services interacting with WNs in dual-stack
 - All the above, without disrupting normal WLCG operations
- By April 1st 2018
 - Tier-1's must provide dual-stack storage access in production with increased performance and reliability
 - Tier-1's must upgrade their Stratum-1 and FTS to dual-stack
 - The official ETF infrastructure must be migrated to dual-stack
 - GOCDB, OIM, GGUS, BDII should be dual-stack
- By end of Run2
 - A large number of sites will have migrated their storage to IPv6
 - The recommendation to keep IPv4 as a backup will be dropped

13/09/2017

HEPiX IPv6 WG

5

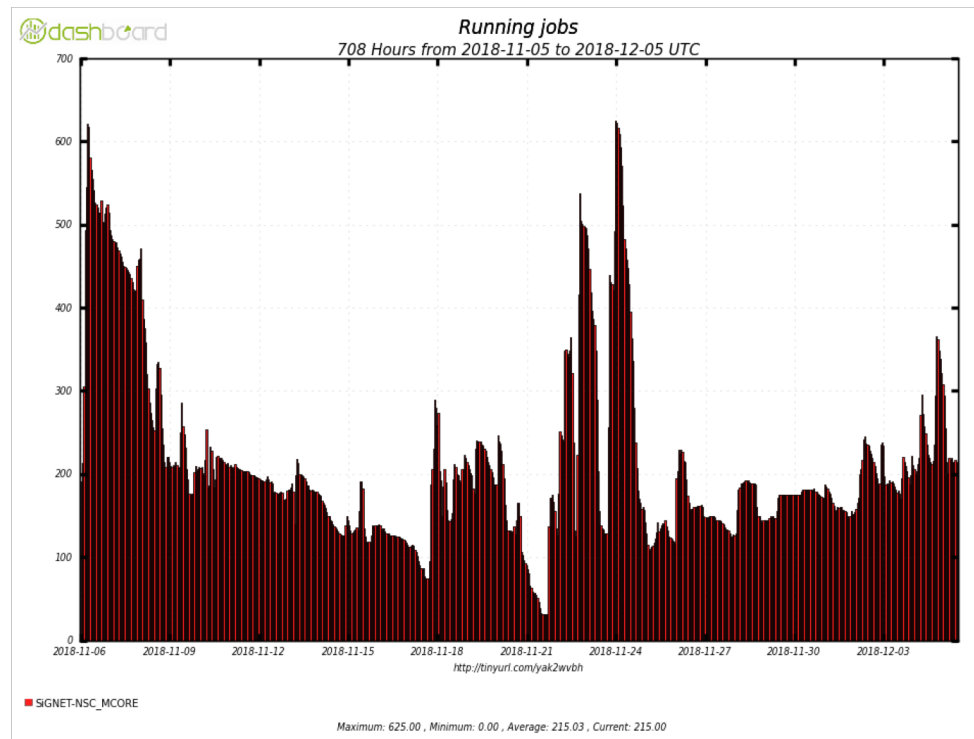
Growth of dual-stack hosts in the WLCG



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/).

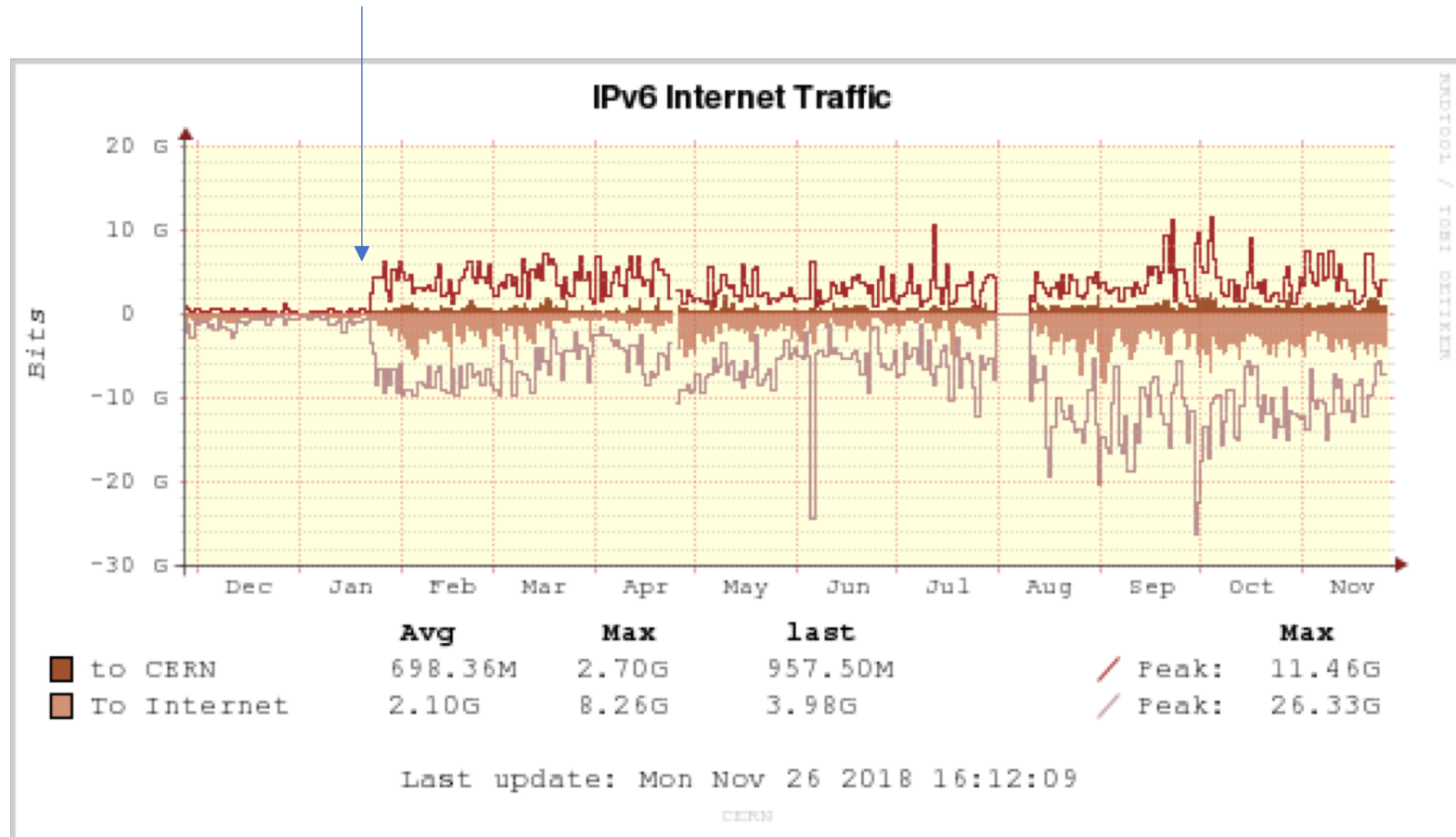
IPv6-only compute

- Need to be ready for a possible offer of IPv6-only compute resources
- Testing IPv6-only worker nodes for
 - CMS at Brunel University London
 - ATLAS at the Jozef Stefan Institute, Slovenia (running production jobs)



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

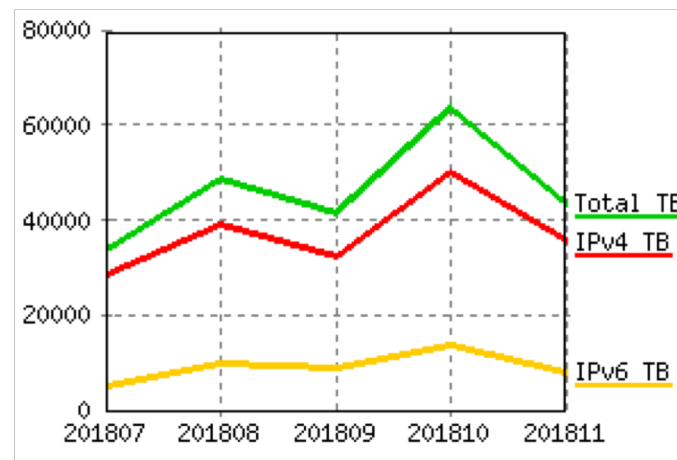
Non-LHCOPN/non-LHCONE traffic



Network traffic flowing through CERN

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

Period	IPv4 TB	IPv6 TB	Total TB	% IPv6/Total	Note
201807	28464	5202	33666	15.45	
201808	39139	9666	48805	19.81	estimation, some data missing
201809	32562	8714	41276	21.11	estimation, some data missing
201810	50016	13686	63702	21.48	estimation, some data missing
201811	35470	7940	43410	18.29	



Tier-1 and Tier-2 dual-stack roll-out

- Nine Tier-1s have dual-stack storage and five IPv4
- Tier-2 sites were requested to deploy dual-stack perfSONAR and storage by end of Run 2 (end of 2018)
- Submitted a ticket to each site in autumn 2017 requesting timescale for deployment of IPv6 and details of steps
- Following up with assistance, checking deployment etc
- Several sites are waiting for their campus network infrastructure to become IPv6-ready
- Only a few sites where the problem is at the Grid service level

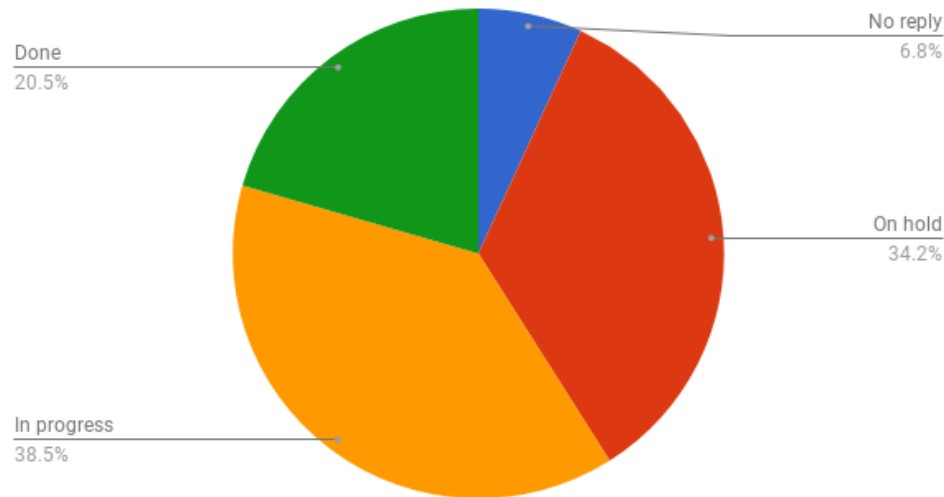
Tier-2s: GGUS tickets submitted to 115 Tier-2 sites

Feb 2018

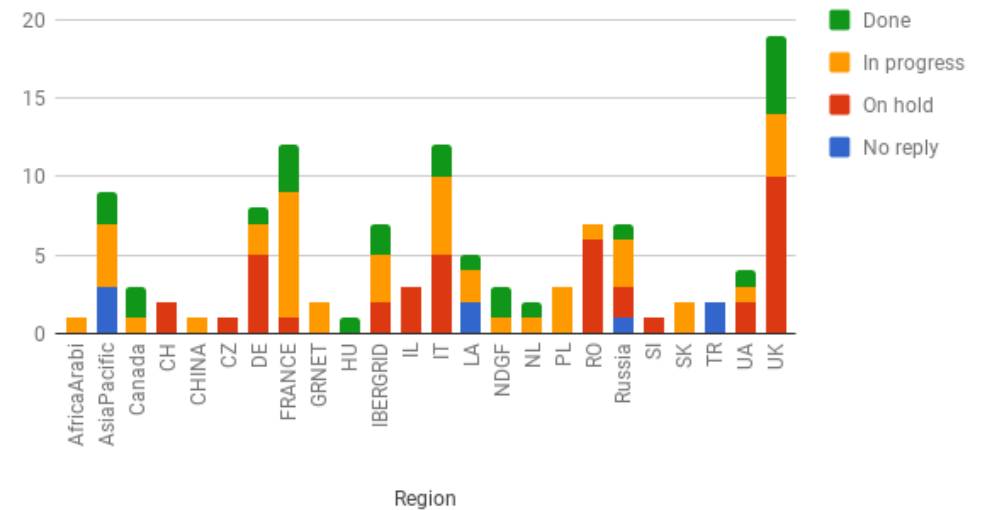
20% Tier-2s with dual-stack perfSONAR and storage

https://twiki.cern.ch/twiki/bin/view/LCG/WlcgIpv6#WLCG_Tier_2_IPv6_deployment_stat

Tier-2 IPv6 deployment status [26-02-2018]



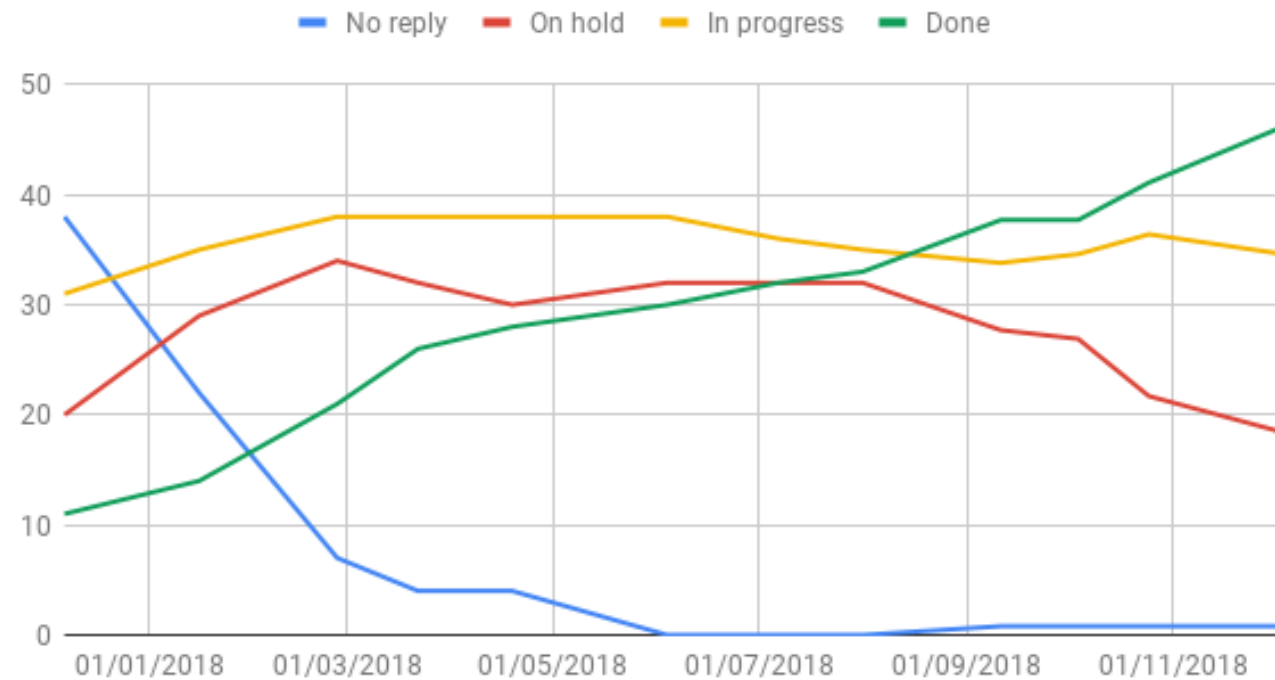
Tier-2 IPv6 deployment status [26-02-2018]



Tier-2 evolution

Tier-2 IPv6
deployment ticket
states (%)

Status vs. time

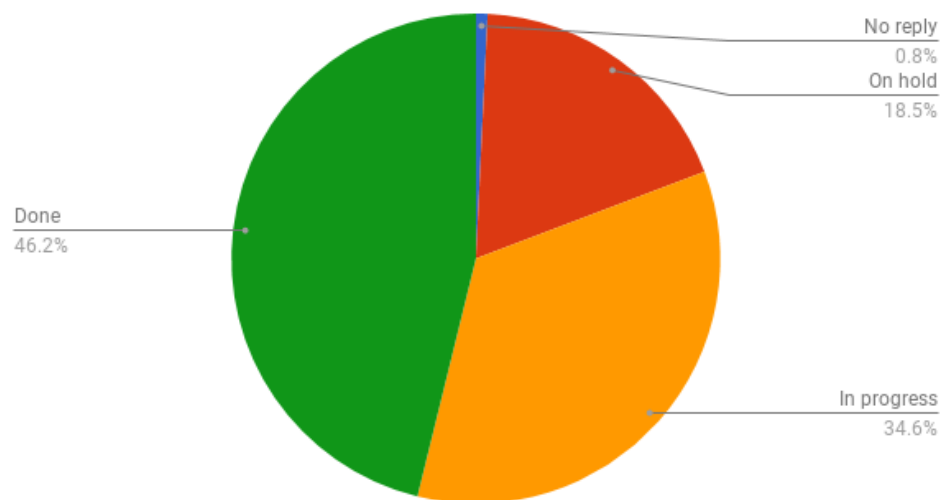


Tier-2s: GGUS tickets submitted to 115 Tier-2 sites

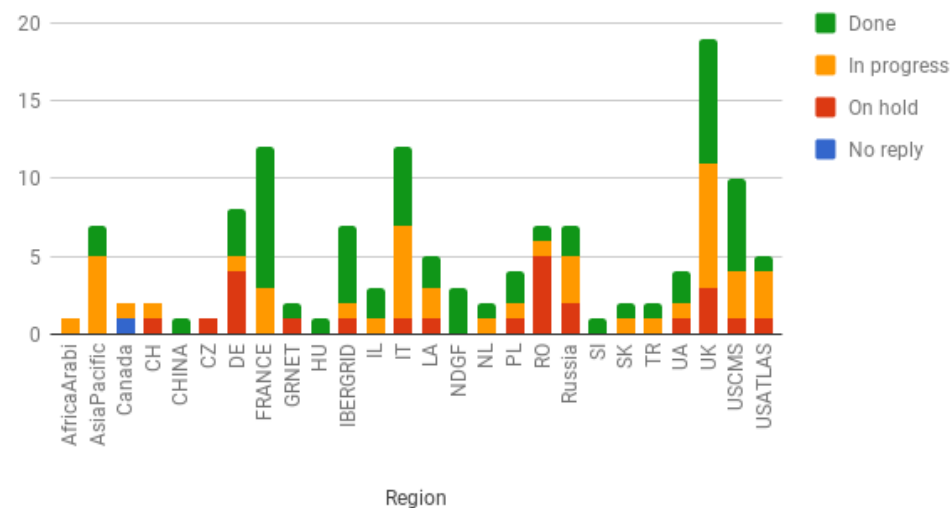
46% of Tier-2s with dual-stack perfSONAR and storage

https://twiki.cern.ch/twiki/bin/view/LCG/WlwgIpv6#WLCG_Tier_2_IPv6_deployment_stat

Tier-2 IPv6 deployment status [04-12-2018]



Tier-2 IPv6 deployment status [04-12-2018]



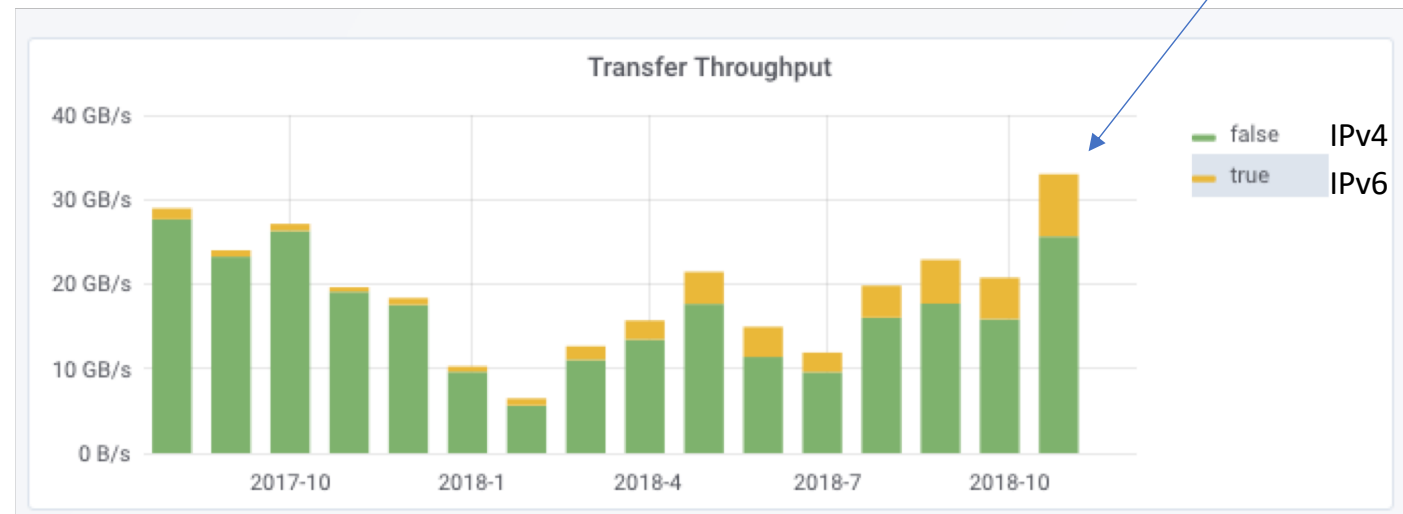
Proportion of Storage accessible over IPv6 (Dec 2018)

Experiment	Fraction of Tier-2 storage accessible via IPv6
ALICE	49%
ATLAS	37%
CMS	65%
LHCb	33%
WLCG	48%

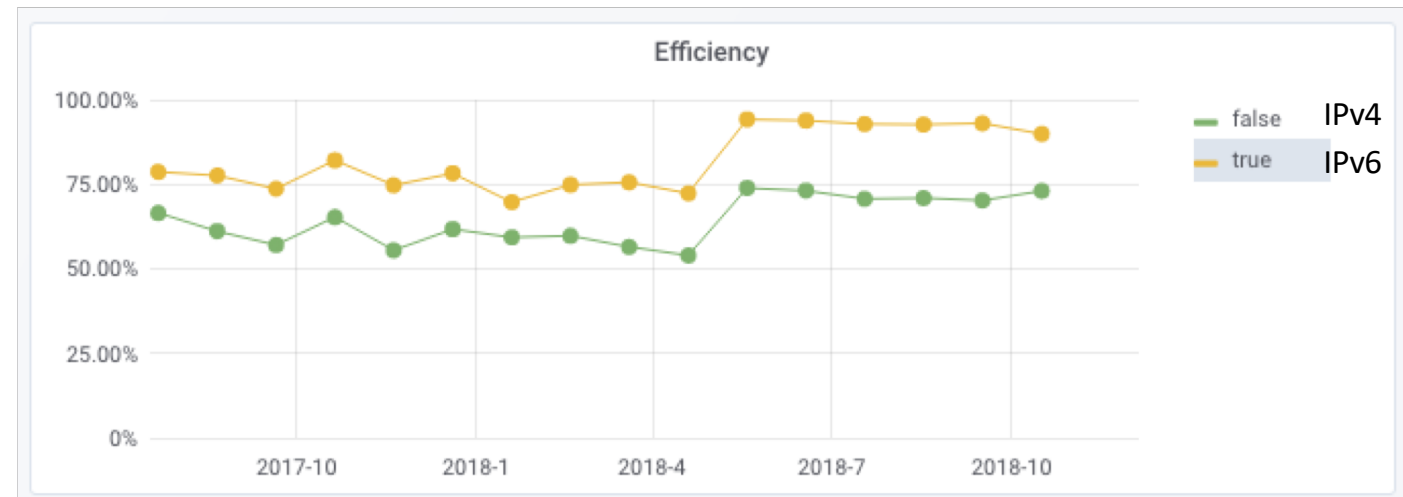
Country	Fraction of Tier-2 storage accessible via IPv6
UK	53%

Bulk data transfer using the File Transfer Service (FTS)

Aggregate WLCG
transfer rates
(Gigabytes/s)



Transfers over
IPv6 appear to be
more reliable

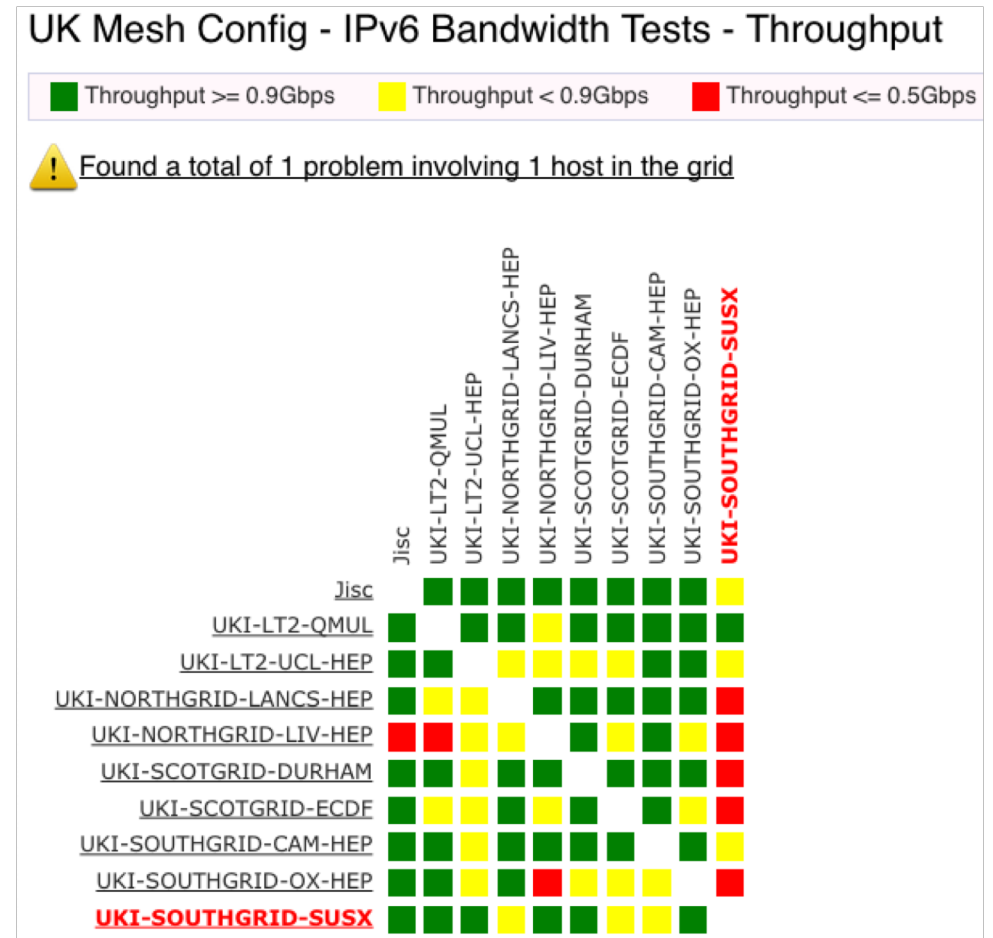
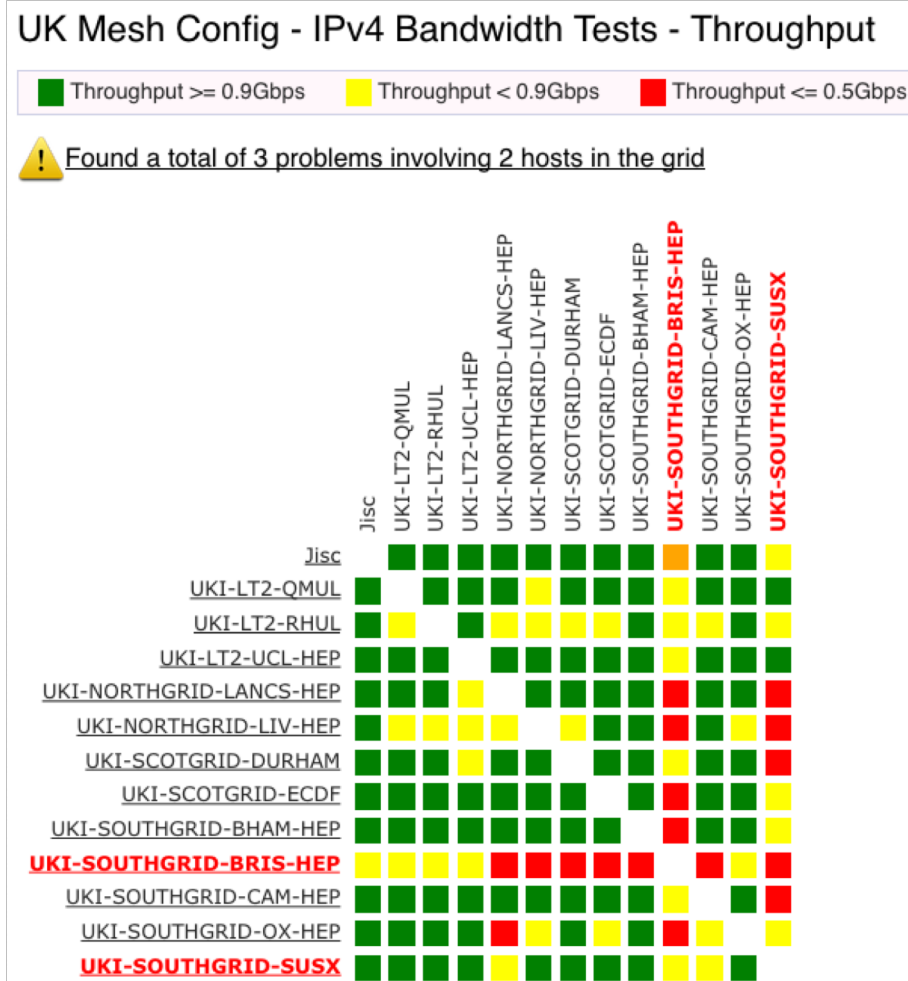


perfSONAR

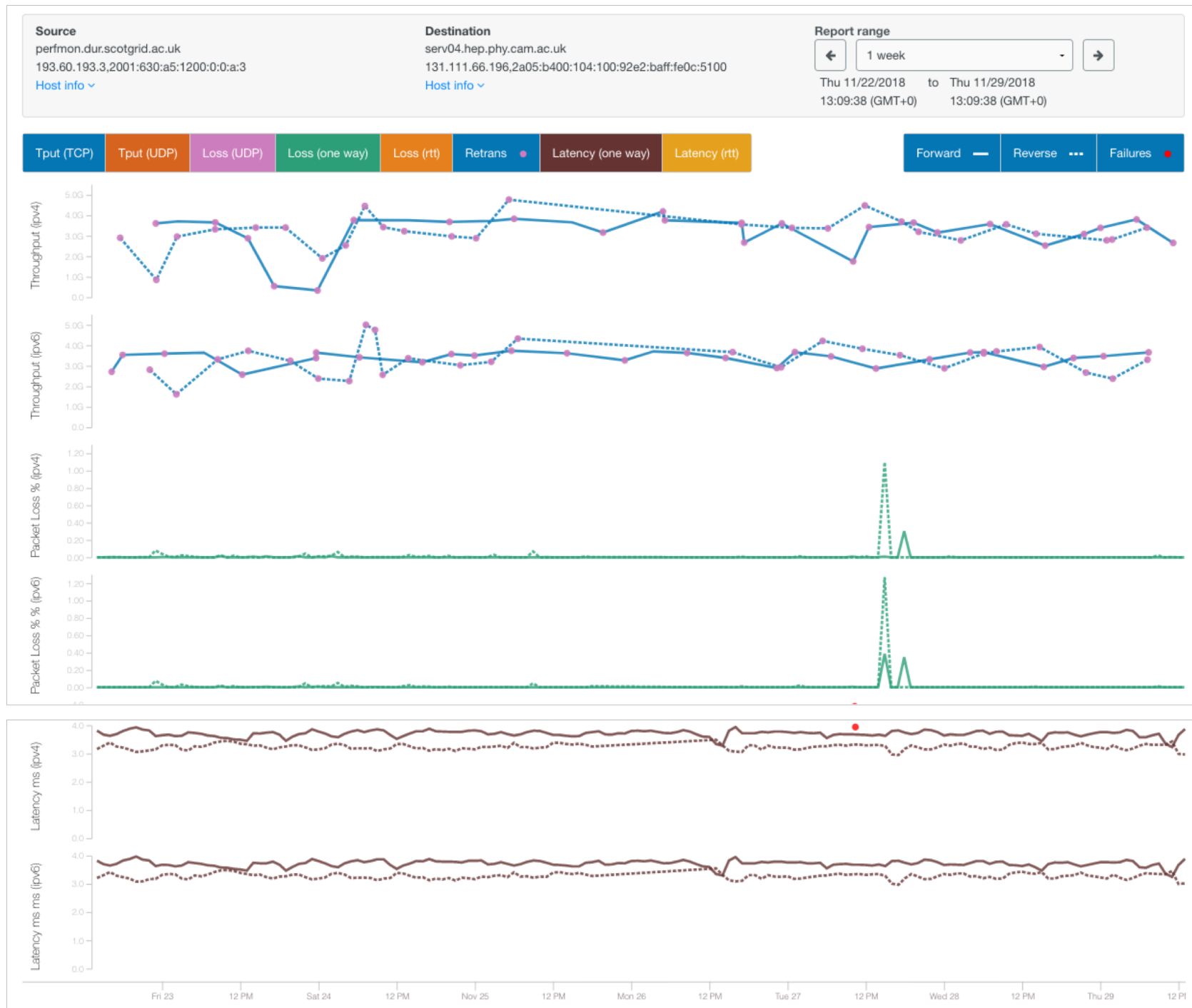
- Network monitoring tool developed by Esnet, GEANT, Indiana University and Internet2
- 'perfSONAR is a widely-deployed test and measurement infrastructure that is used by science networks and facilities around the world to monitor and ensure network performance.'
- <http://www.perfsonar.net/about/what-is-perfsonar/>
- WLCG goals with perfSONAR
 - Find and isolate “network” problems; alerting in time
 - Characterize network use such as finding base-line performance
 - In the future: provide a source of network metrics for higher level services
- perfSONAR is IPv6 compatible

perfSONAR dashboards

- Each WLCG site requested to deploy perfSONAR
- WLCG has meshes for a variety of groupings e.g. the LHCOPN, CMS and ATLAS
- UK also runs dual-stack one: throughput, latency, loss, traceroute
- Gives insight into network performance over IPv4 and IPv6 within UK



Example perfSONAR results: Durham to Cambridge



GridPP Network Tests

- Jobs are sent to a WN at each site to read 1GB, 2GB and 3GB files from each site's Storage Element (SE) using various protocols. The files have been previously replicated to all SE. The table shows average bandwidth (in MB/s) into the worker nodes computed from the times taken for each combination (including the local SE).
- Test over IPv6 also
- Transfers are made with lcg-cp, gfal-copy, curl and xrdcp over IPv4 and IPv6 (where relevant)
- Also recording the percentage of UK CPU and storage available over IPv6
- UK currently has 53% of disk storage available over IPv6

Site	Capacities				Network						
	CPU	Core	HS06	Disk	lcg	gfal4	gfal6	http4	http6	xroot4	xroot6
Brunel	359	5770	68951	1377		40.1	44.5	31.6	37.2	44.4	89.3
Imperial	716	5718	56664	4965	25.0	12.8	25.9	12.6	20.6	14.8	33.9
QMUL	361	4000	62351	5031	13.1	14.6		15.4		13.1	
RHUL	442	4624	48121	1460	63.2	39.0		39.3		32.9	
								65.7		61.7	
UCL	0	0	0	0							
Lancaster	420	3360	48384	3074							
Liverpool	173	1816	18466	1425	6.6	6.1		8.6		6.2	
					136.7			28.5		52.5	
Manchester	219	4297	46010	4544	105.6	52.3	60.8	45.4	45.8	73.9	76.7
								52.4		75.2	
Sheffield	100	800	10560	531	74.1	72.0		72.4		81.0	
Durham	423	3400	45560	423	24.8	22.0	18.6	29.5	31.5	21.0	20.3
Edinburgh	66	528	6811	2208		76.1		103.0		126.8	
Glasgow	629	5032	43980	3816	12.4	16.4		18.1		20.2	
								31.4		37.0	
Birmingham	152	1584	16996	260				86.9		123.7	
Bristol	82	1320	14744	713	28.8	30.7		24.1		54.3	
Cambridge	78	528	6146	264	54.9	46.1		38.5		54.6	
								44.2		52.0	
Oxford	407	3256	33586	939	31.0	31.1		27.9		47.0	
RAL PPD	509	4668	46680	3425	16.0	16.2		14.5		18.9	
Sussex	71	568	5583	84	8.7	6.3		9.3		7.9	
CLOUD											
RAL Tier-1	2285	27424	274240	12819	18.1	13.9		10.4		16.3	
Tier-2 Totals:	5207	51269	579593	34539							
IPv6 Totals:	1540	15735	180255	18220							
IPv6 Percent:	30%	31%	31%	53%							

<https://pprc.qmul.ac.uk/~lloyd/gridpp/ukgrid.html>

Summary

- The WLCG needs to be ready for an offer of opportunistic IPv6-only CPU resources
- We are slowly but surely making our computing service IPv6 ready
- IPv6-only worker nodes at one WLCG site are already running production jobs
- Tier-1s and Tier-2s should be providing production storage accessible over IPv6 (64% and 46% are, respectively)
- This means ~48% of LHC data is now accessible over IPv6
- The volume of data transferred over IPv6 has increased over the last year, ~20-25% of bulk data transfers now go over IPv6
- ~50% of WLCG perfSONAR hosts are now reporting 'IPv6-enabled'
- Finally, one hopefully positive, side-effect is that this is encouraging IPv6 adoption in a large number (~170) of research institutes worldwide

Acknowledgements

- Reported work done by the HEPiX IPv6 Working Group and WLCG IPv6 Task Force and many others in the WLCG

