Imperial College London

GridPP
UK Computing for Particle Physics

IPv6 update

Duncan Rand

Imperial College London

LHCOPN/LHCONE meeting Abingdon, March 2018

HEPiX

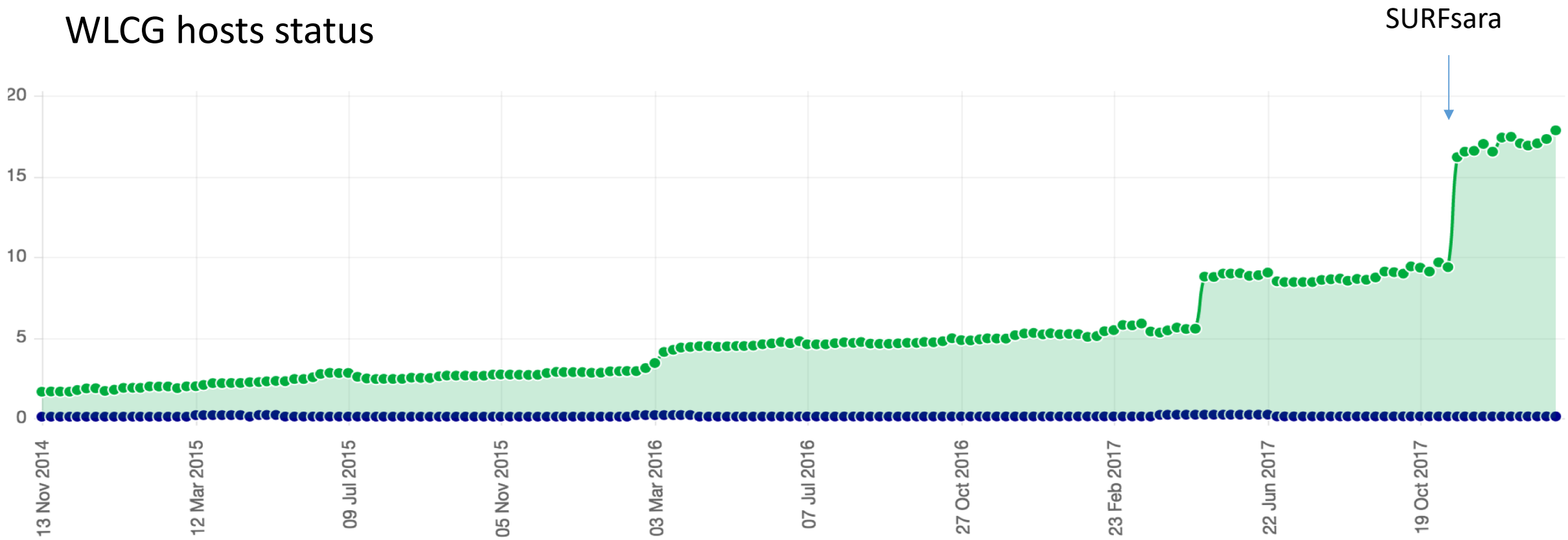
- Reporting work done by HEPiX IPv6 Working Group and WLCG IPv6 Task Force and many others such as site admins, CERN and LHC VO central support teams
- Many slides taken from recent HEPiX IPv6 Working Group face to face meeting
  - https://indico.cern.ch/event/676532/

- Why IPv6?
  - The main motivation is to make the data at the site accessible by clients running on IPv6-only machines
  - WLCG might get an offer of opportunistic CPU resources which are IPv6-only
  - Also for pledged resources, sites running out of IPv4 addresses and to avoid use of NAT

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
  - 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

WLCG hosts status



- Percentage of IPv6-only endpoints
- Percentage of dual-stack endpoints

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/). See also corresponding results from the LHC Expts VO feeds (http://orsone.mi.infn.it/~prelz/ipv6_vofeed/)

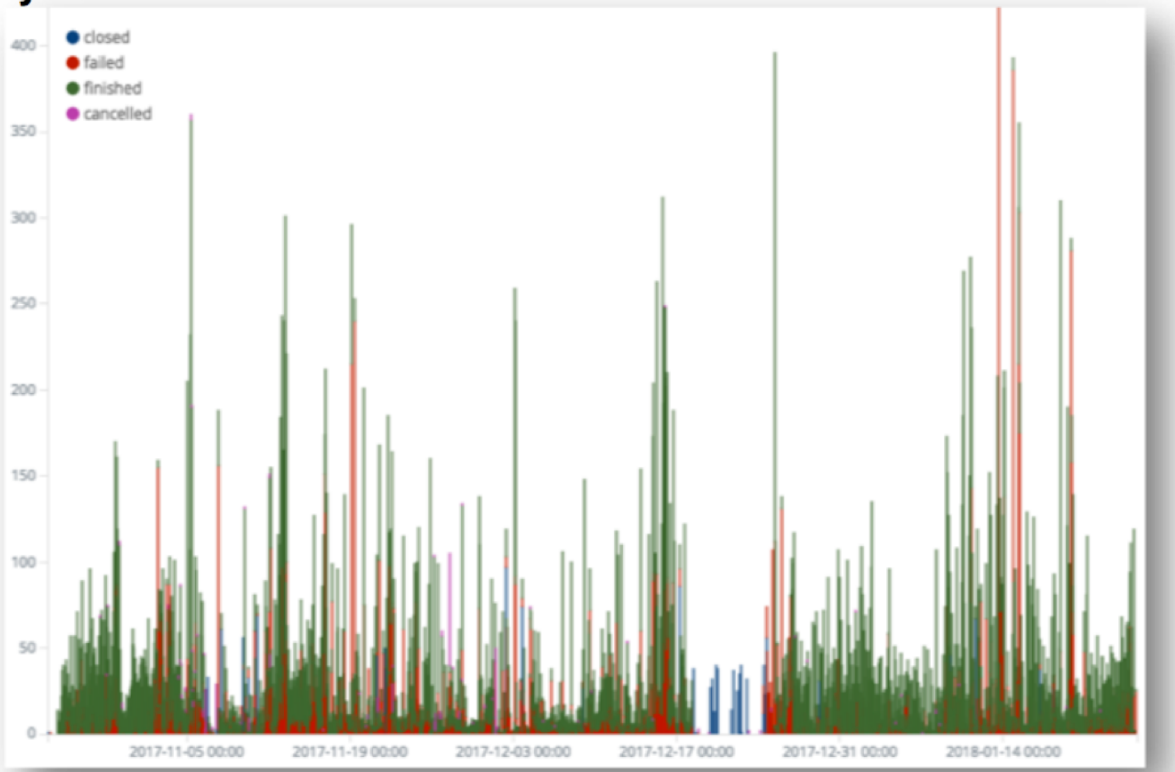
IPv6 only CPU

- ATLAS jobs are working on IPv6 only CPU at SiGNET
- All necessary services were made dual-stack

- SiGNET NSC cluster has IPv6 only CPU and is running production jobs.

Jobs run over the last month

400 →

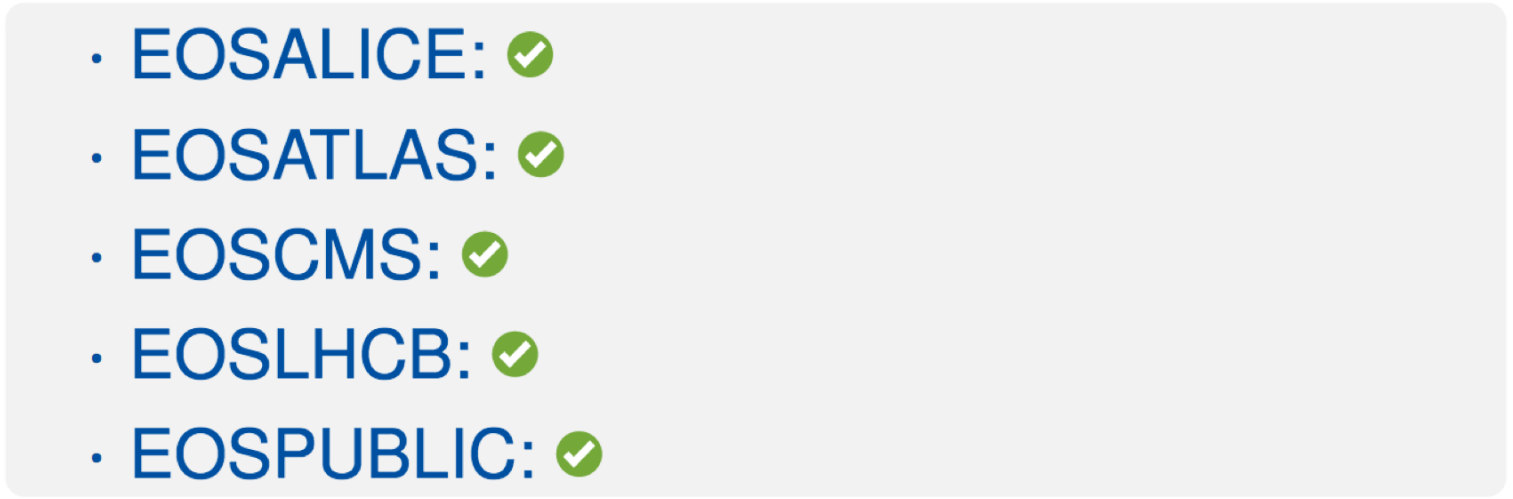


Alastair Dewhurst, 25th January 2018

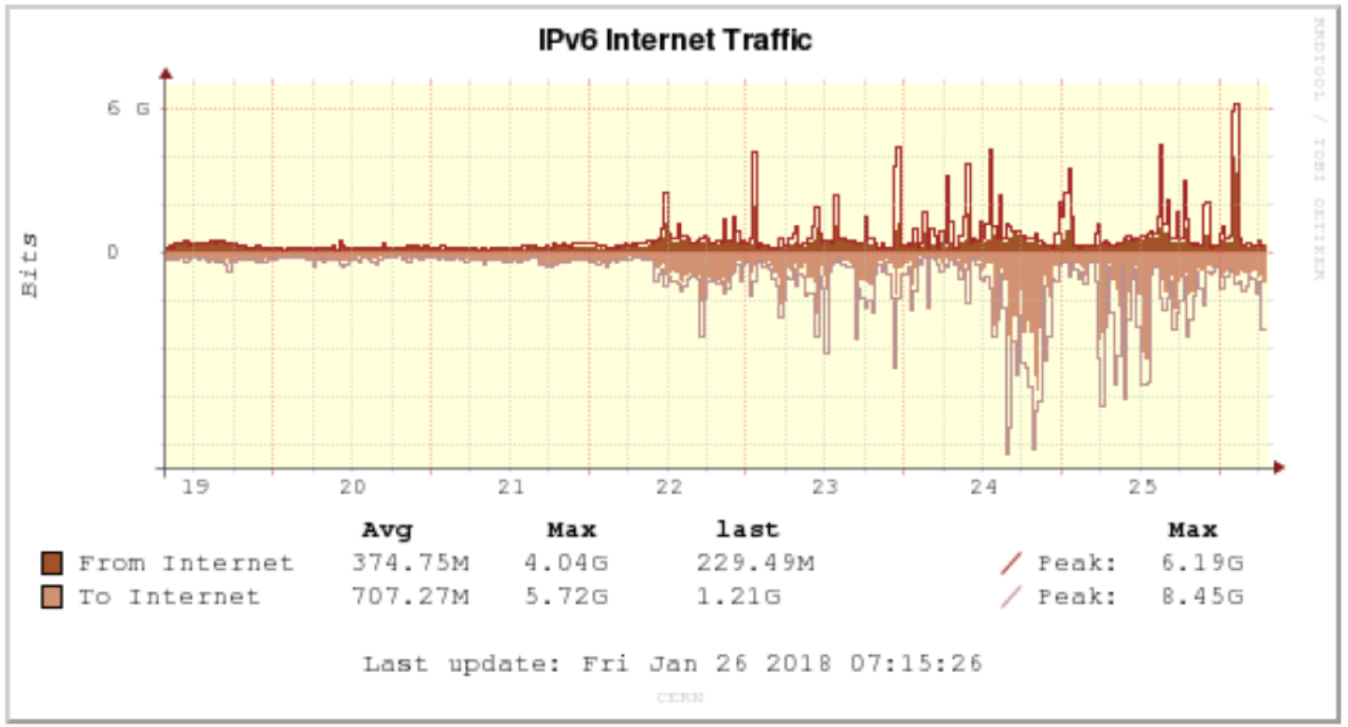
CERN Storage update
(Hervé Rousseau)

EOS Instances IPv6 status

- CERN EOS instances are now all dual-stack
- EOSALICE and EOSLHCB made dual-stack at end of 2017
- EOSATLAS and EOSCMS done at beginning of 2018

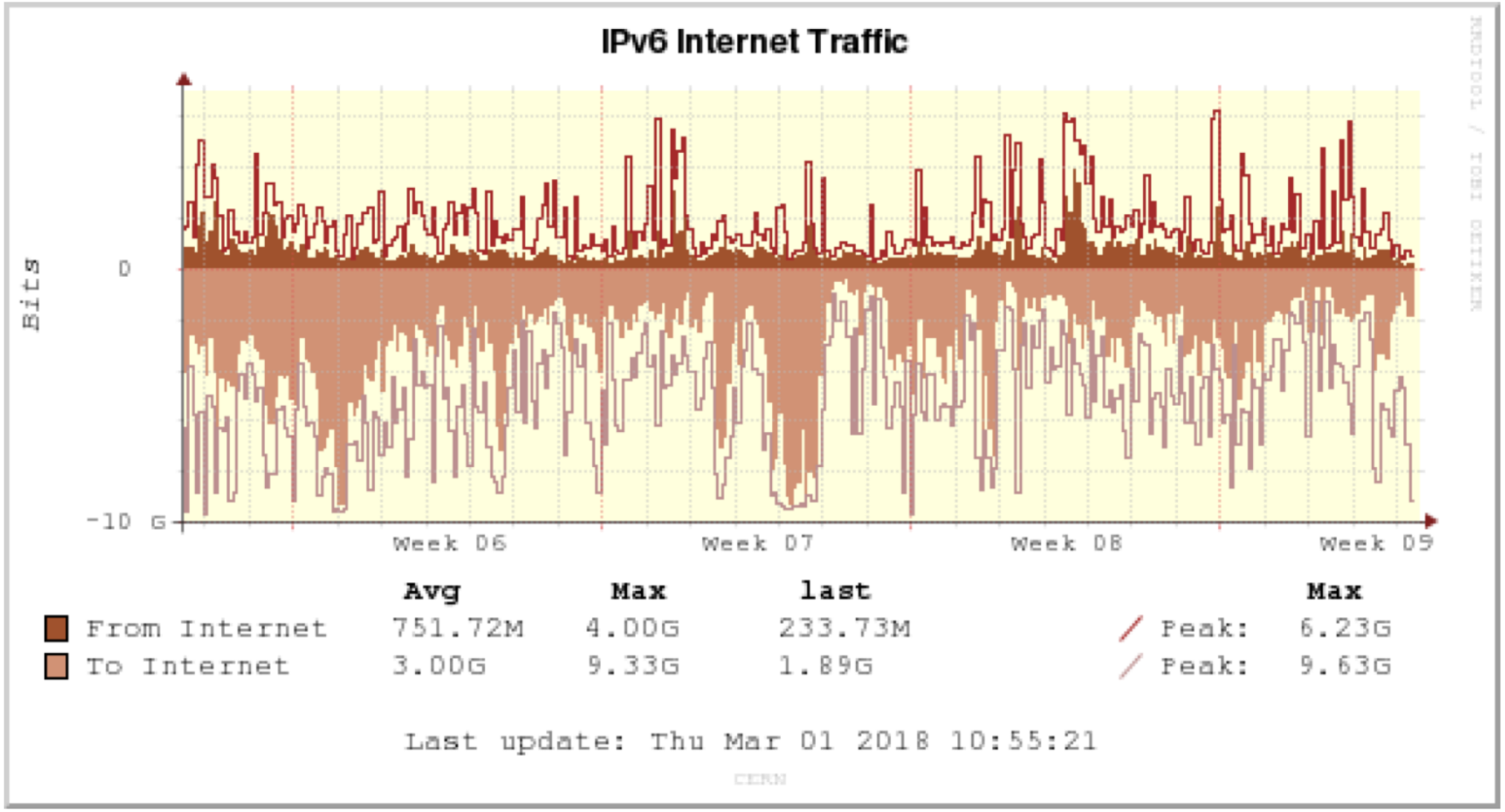


EOS Instances IPv6 status



Something changed this week

Non-LHCOPN/non-LHCONE traffic February 2018



https://tinyurl.com/ok773t2

Tier-1 Status

- All Tier 1, except **KR-KISTI-GSDC,** are now peering with LHCOPN over IPv6
- Storage gradually being made dual-stack e.g. RAL Echo storage done last week
- Currently transferring over IPv6: IN2P3, INFN, JINR, NDGF, RAL, SARA, ASGC, pic
- Transferring over IPv4: BNL, KIT, RRC-KI, FNAL, NIKEF, TRIUMF

- FTS servers IPv4: FNAL, BNL

Sites IPv6 connectivity

Name	Type	LHCOPN IPv6 peering	LHCONE IPv6 peering	LHCONE IPv6 peers	NREN IPv6 peers	IPv6 LAN	dualstack perfSONAR	dualstack storage percentage by 1st April 2017	dualstack storage percentage by 31st May 2017	dualstack storage percentage by 31st July 2017	Network Statistics	Notes
UKI-LT2-IC-HEP	Tier2		Yes	Janet	Janet	Yes		100				
TRIUMF	Tier1	Yes	Yes	Canarie	BCNET	Yes	Yes	0				TRIUMF MW readiness storage is 100% running on dual stack since April 26. And data transfer is primarily using ipv6 now.
RAL	Tier1	Yes			JANET	Yes	Yes	0	20	100		IPv6 storage connectivity referes to ECHO/Ceph, Castor will not (ever) be dual stack.
praguelcg2 (FZU)	Tier2	No	Yes	CESNET	CESNET	Yes	Yes	100			http://netreport.cesnet.cz/netreport/hep-cesnet-experimental-facility2/	
NL-T1-Nikhef	Tier1	Yes	No	GEANT	SURFnet	Yes	No	0	0	100	http://beer.nikhef.nl/cgi-bin/grapher.cgi?target=%2Fparkwachter.ipmi.nik...	Full dual stack coming soon; direct peering with CERN is up.

https://hepix-ipv6.web.cern.ch/sites-connectivity

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

- Request to deploy dual-stack perfSONAR and storage by end of Run 2 (end of 2018)
- Engaging with sites - requesting timescale for deployment of IPv6 and details of steps
- Following up with assistance, checking deployment etc
- GGUS ticket details: https://tinyurl.com/y9tfd5oo

Description: IPv6 deployment at WLCG Tier-2 sites

Detailed Description:

Dear site support,

The WLCG management and the LHC experiments approved a deployment plan for IPv6 that requires that Tier-2 sites deploy dual-stack connectivity (IPv4+IPv6) at least on their perfSONAR and storage systems by the end of Run2 (end of 2018).
The main motivation is to make the data at the site accessible by clients running on IPv6-only machines, which might well soon be the case soon for opportunistic (and not only) resources.
All experiments therefore require that the vast majority of sites (if not all) offer IPv6 connectivity to their storage.

The purpose of this ticket is to track the status of the IPv6 deployment process at your site and will be closed only when your storage has been successfully tested to work via GridFTP, xrootd and WebDAV (depending on the VOs supported) from an IPv6-only client, ensuring that the transfer performance is as good as via IPv4, as monitored by perfSONAR.

Before that, we would ask you to answer with this information:

- your estimate of the timescale for the deployment;
- a few details about the required steps;

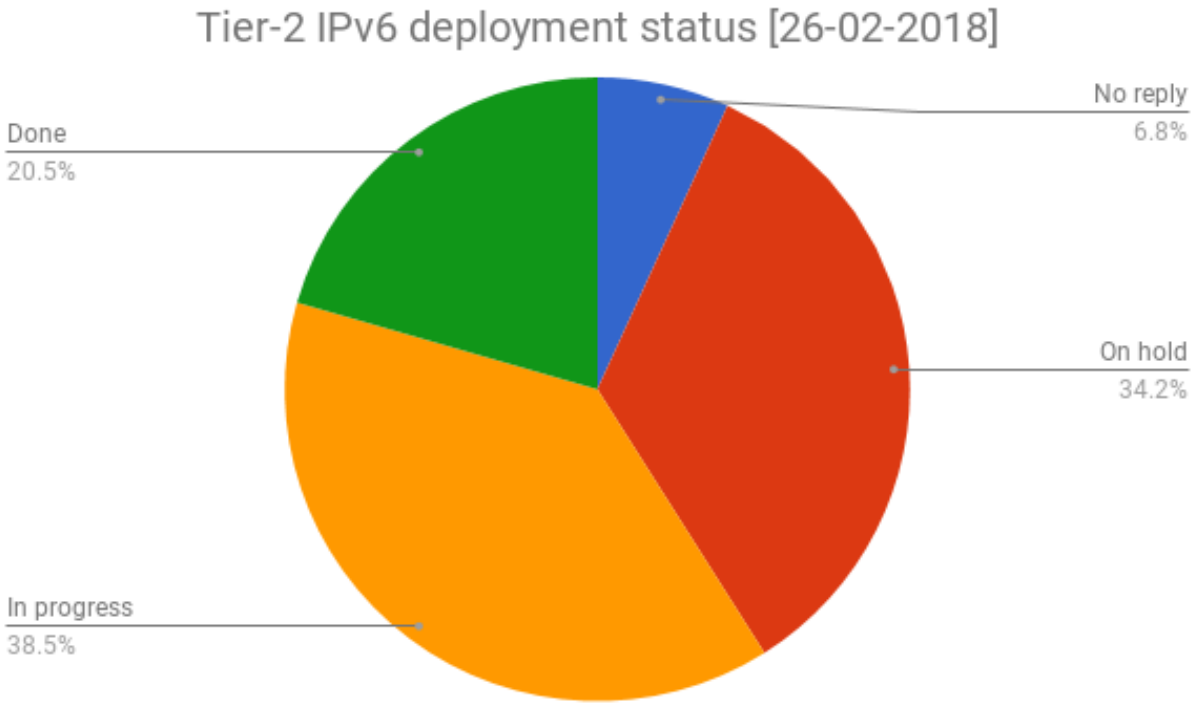
and to add comments to this ticket whenever progress has been made.
In the unfortunate case it becomes evident that the deadline cannot be met, we would appreciate if you could explain what are the obstacles and still give an estimate for the time of completion.

The final step of the deployment will be a test conducted by somebody in the LHC VO(s) supported by your sites. The relevant contact people will be monitoring the ticket and will make sure that the tests are performed.

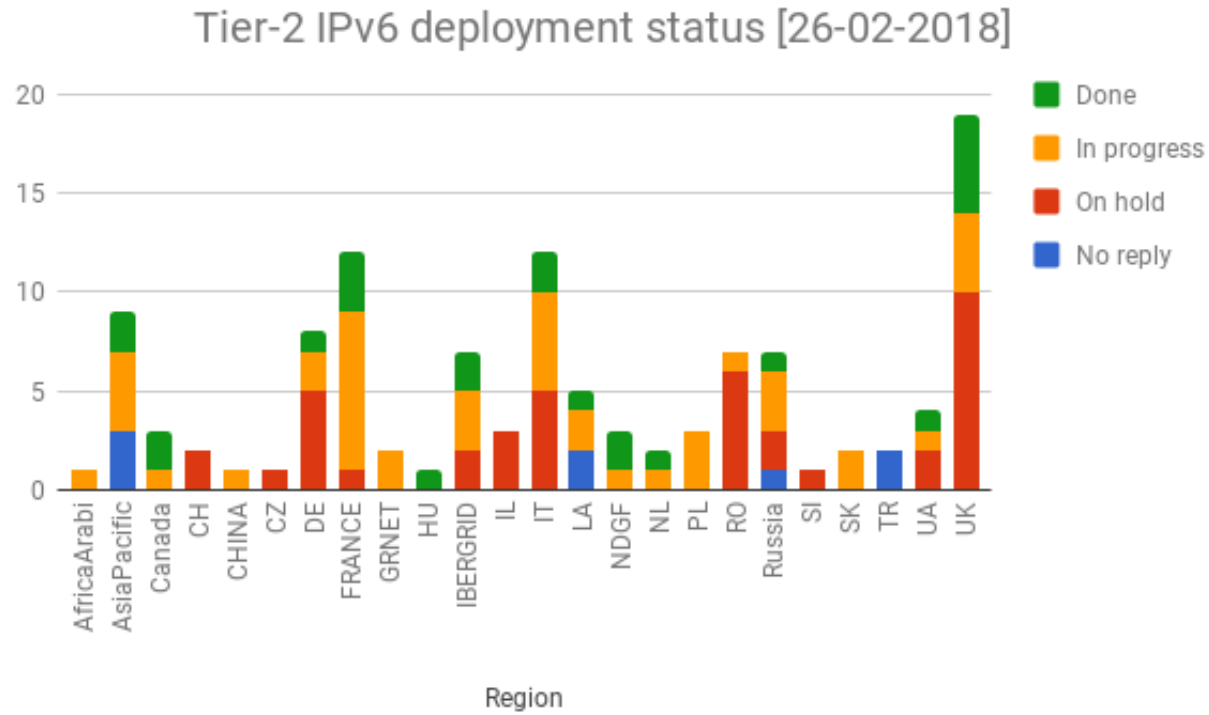
For questions and request for help, you can contact the 'WLCG IPv6' support unit in GGUS.

The WLCG IPv6 deployment task force

A fifth of Tier-2s now with dual-stack perfSONAR and storage

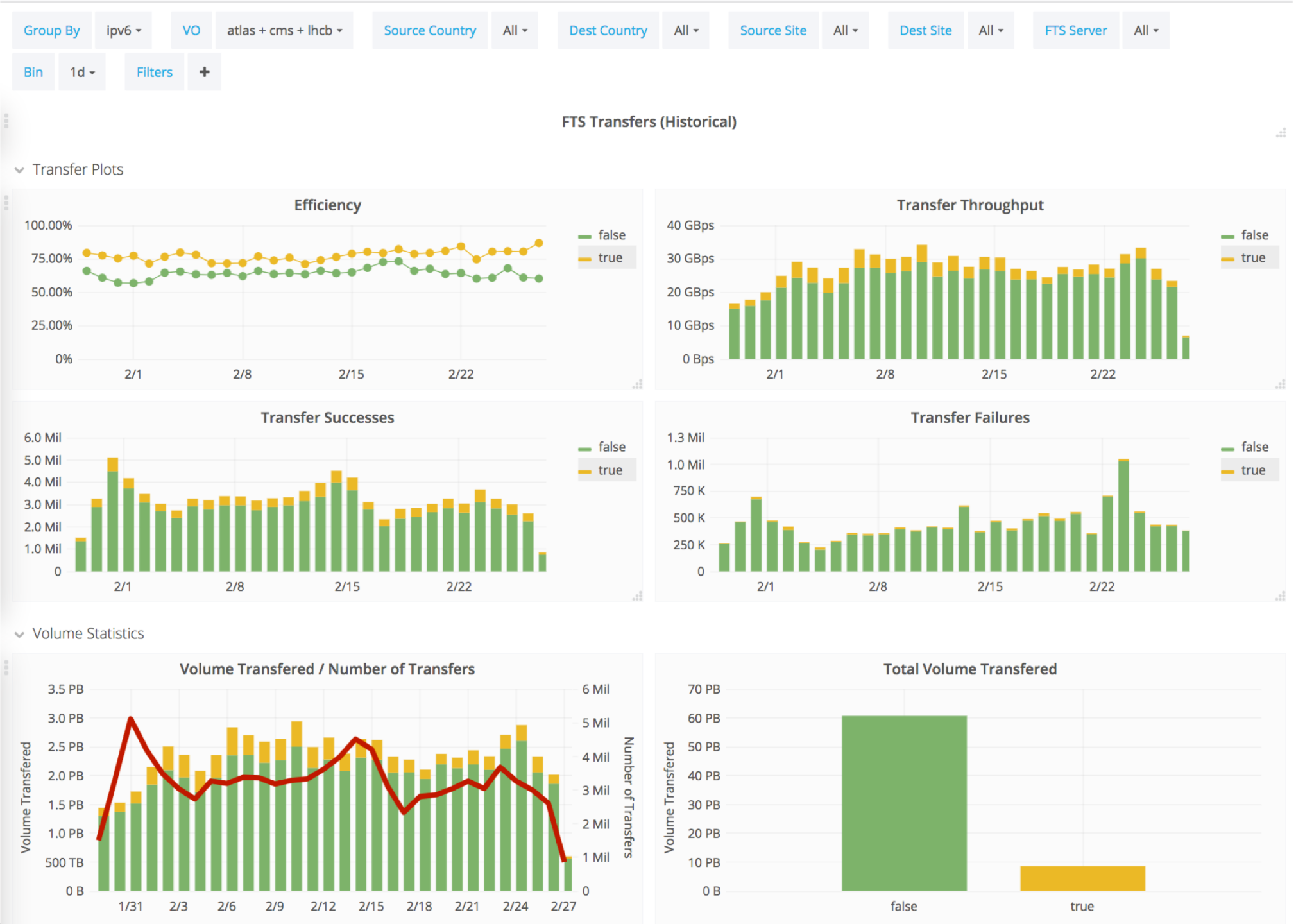


https://twiki.cern.ch/twiki/bin/view/LCG/WlcgIpv6#WLCG_Tier_2_IPv6_deployment_stat



FTS transfer monitoring

- FTS transfers can now be displayed indicating whether IPv6 (true) or IPv4 (false) was used
- Approximately 12% of data transferred in the last 30 days went over IPv6
- Transfers over IPv6 appear to be more efficient – investigating why
- Working on getting similar data for xrootd transfers



Marian Babik (CERN)

perfSONAR News

- 4.0.2 was released in November with many bug-fixes and performance improvements
  - Most of our infrastructure updated, we already have around 60 instances on CC7
  - However we're still seeing issues with performance and also stability
  - perfSONAR F2F dev meeting will take place in Amsterdam 7-8th March
- Currently ongoing campaign to update all perfSONARs to CC7
  - 4.1 release will not have SL6 packages and SL6 support will be dropped 6 months after 4.1 (Q3 2018)
- New Grafana dashboards introduced - single place to visualize all results
  - Currently IPv4-only, but working towards enabling IPv6 as well with side-by-side comparison
- IPv6 configured via dual-stack mesh maintained by Duncan
  - Out of 270 hosts monitored, 103 are dual-stack now
  - IPv6 performance analysis - credits to Brian Davies
  - Few cases with IPv6 issues were also reported to the WLCG Network Throughput support
    - Most notable was IC to SARA, which was due to a firmware issue

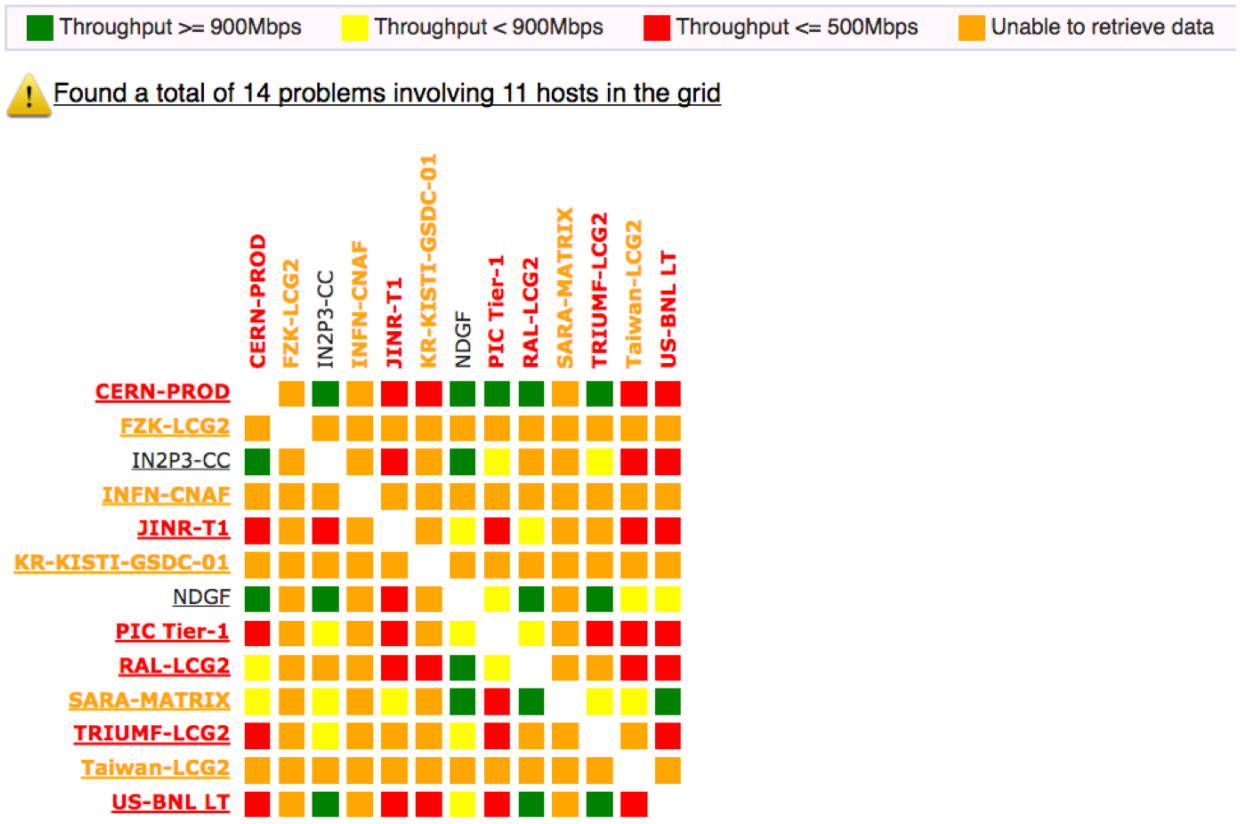
https://tinyurl.com/y8ftd9sh

perfSONAR Dual-stack Mesh

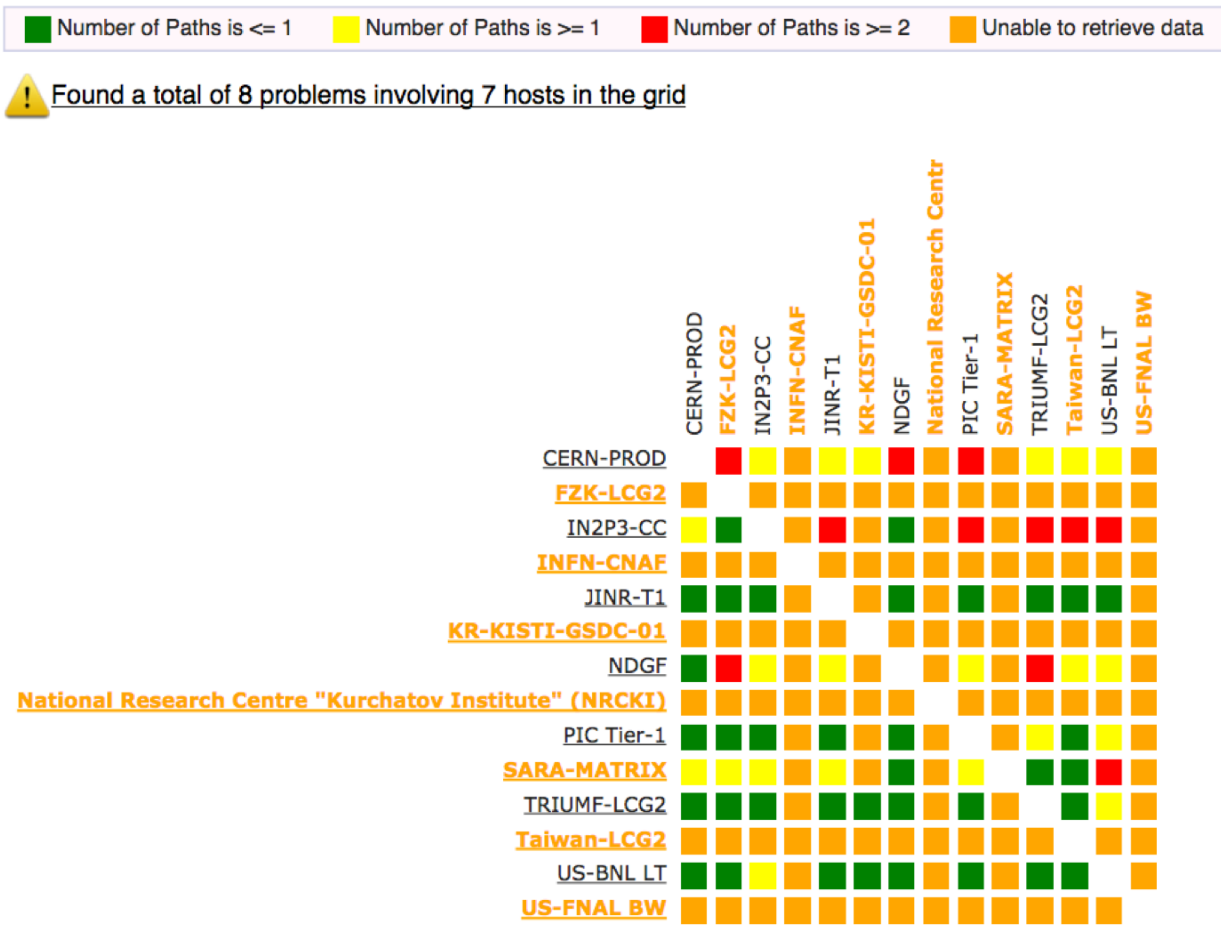
- Dual-stack mesh was created to follow up IPv4 vs IPv6 performance
  - Ideally we would like to run both IPv4 and IPv6 tests for all dual stack perfSONARs
  - This would however increase the amount of testing by factor 2 on already stretched resources
- Proposal to replace dual-stack mesh is needed
  - We need to have a way to monitor IPv6 performance while at the same time avoid collapse of the infrastructure due to overload
  - The most resource sensitive monitoring is latency as it runs continuously
  - Throughput and packet trace could in principle scale by factor 2
  - Ideas:
    - Create dual-stack LHCOPN with both IPv4 and IPv6 for all tests
    - Change all current experiments meshes to contain IPv6 throughput and tracepath
    - Create a dedicated IPv6/IPv4 latency mesh that would only be used for debugging specific cases
    - Alternatively, change experiments meshes to measure IPv4 on IPv4-only paths and IPv6 on IPv6-only paths, however this way we would be missing IPv4 results on IPv6-enabled paths

- LHCOPN mesh is now additionally testing throughput and traceroute (but not latency/loss) over IPv6

OPN Mesh Config - IPv6 TCP BWCTL Test Between OPN Bandwidth Hosts

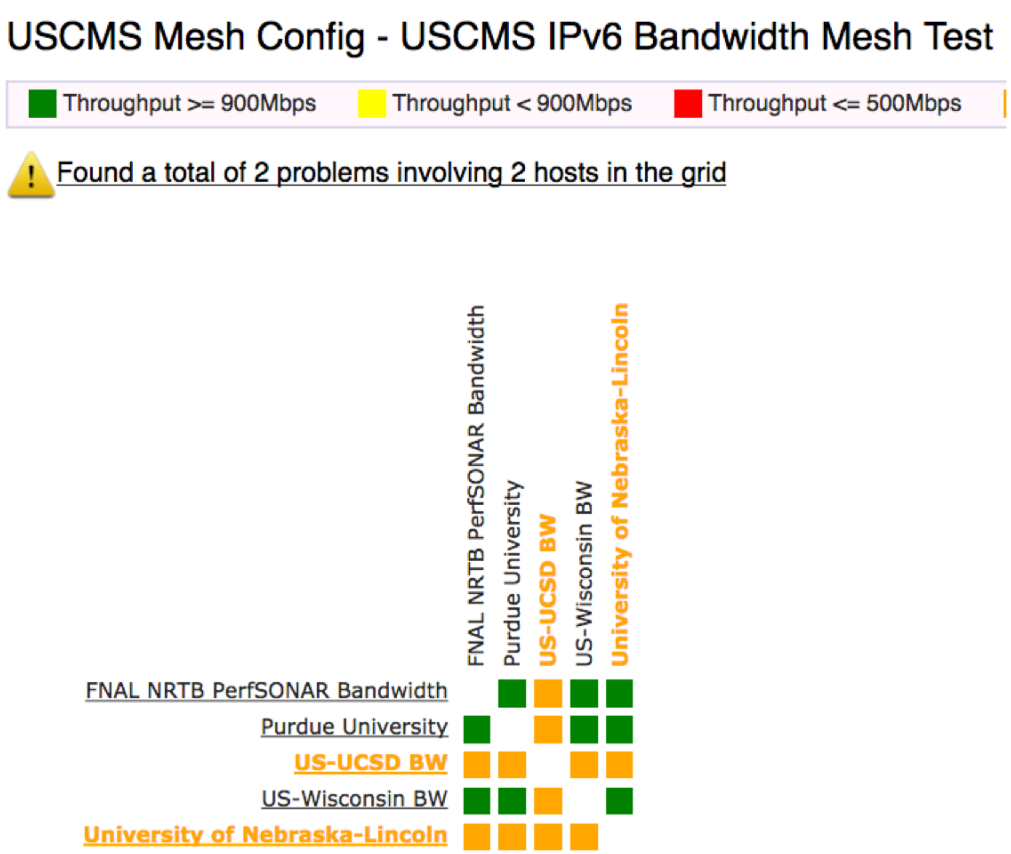


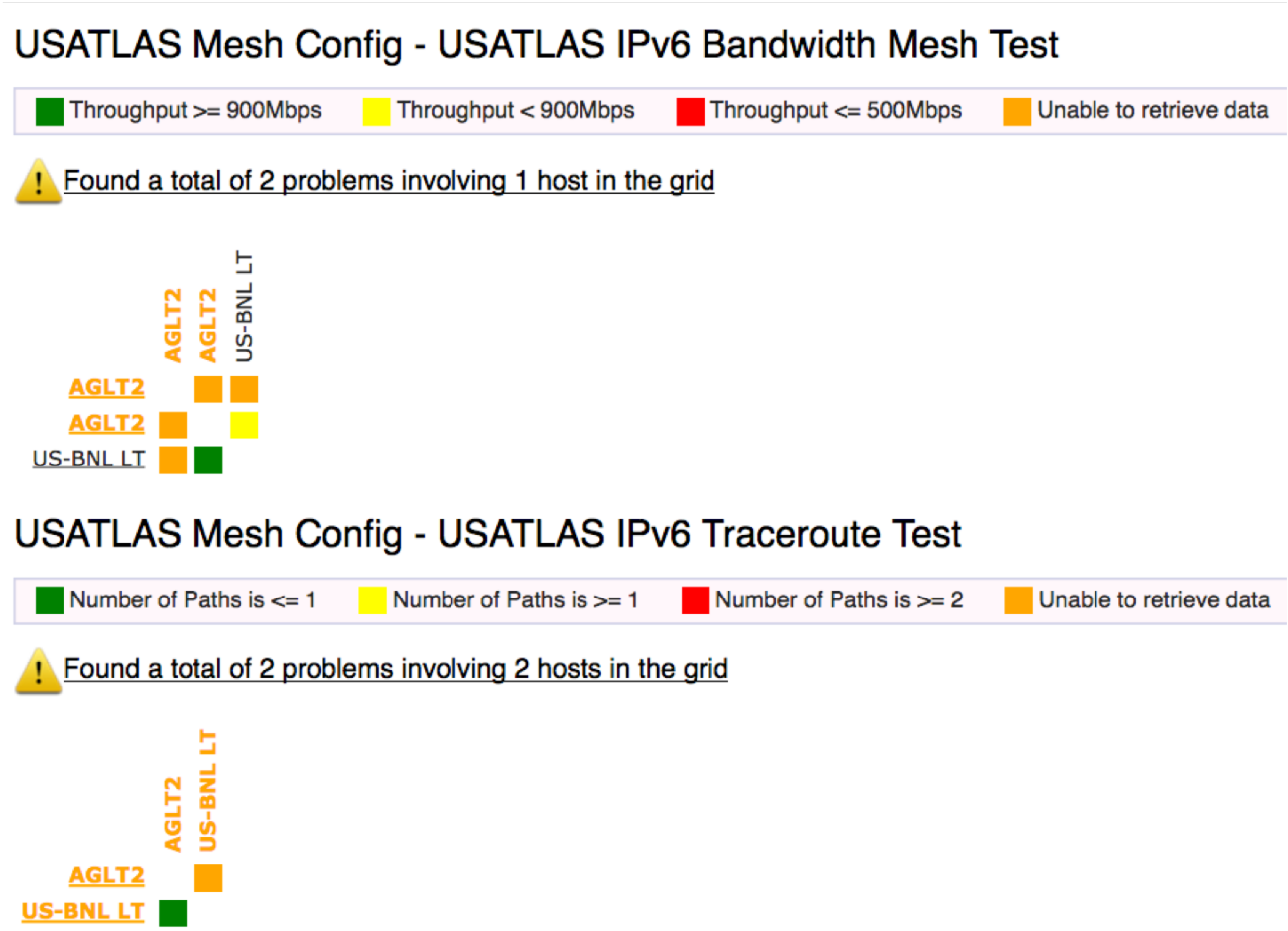
OPN Mesh Config - IPv6 Traceroute Test Between OPN Bandwidth Hosts



http://psmad.grid.iu.edu/maddash-webui/index.cgi?dashboard=OPN%20Mesh%20Config

- USCMS, USATLAS now also testing throughput and traceroute (but not latency/loss) over IPv6 where enabled
- Main LHC VO meshes still testing over IPv4 – will be addressed soon





ETF News

- ETF IPv6 instance was established some time ago to take measurements for IPv6-enabled services - running exact same tests as we do in production
- Proposal was sent to the experiments in August to create dedicated instances for IPv6-only testing (replacing ETF IPv6 instance)
  - CMS and LHCb IPv6-only instances were created
    - https://etf-cms-ipv6.cern.ch/etf/check_mk/
    - https://etf-lhcb-ipv6.cern.ch/etf/check_mk/
  - Both are currently dual-stack, but will be switched to IPv6-only as soon as this is supported
  - Missing input from ATLAS, some input received from ALICE, but not request was made yet
- For both LHCb and CMS we can start publishing results to SAM3
  - Note: current worfklow ETF -> message-bus -> SAM3 -> WLCG reports
  - Publishing to SAM3 needs to be tested as the aggregation algorithm needs to be tuned to support getting results from IPv4-only/IPv6-only instances or both at the same time
  - We have a technical proposal how to enable this and will start testing it in QA soon

UK GridPP Dirac Network Test Results

- Jobs are sent to a WN at each site to read 1GB, 2GB and 3GB files from each site's SE using various protocols. The files have been previously replicated to all SEs. The table shows average bandwidth computed from the times taken for each combination (including the local SE).
- Now upgraded to 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

	Capacities				Network						
Site	**CPU**	**Core**	**HS06**	**Disk**	**lcg**	**gfal4**	**gfal6**	**http4**	**http6**	**xroot4**	**xroot6**
Brunel	266	4271	51039	1438	30.9	35.2	46.0	30.3	37.9	22.9	36.6
Imperial	716	5718	56664	4988	12.1	13.1	28.6	12.8	33.0	10.0	40.9
QMUL	424	4248	46797	5029	6.3	6.8		8.1		7.1	
RHUL	422	3824	38312	1460	13.5	12.8		19.9		20.4	
					50.8			38.9		41.9	
UCL	0	0	0	0	14.9			19.2		14.6	
Lancaster	400	3200	46080	3074				79.1			
Liverpool	289	2498	26634	1425	15.0	6.7		7.7		7.2	
					25.8			24.8		34.5	
Manchester	238	3416	34714	3447	72.7	52.8	71.1	47.9	69.0	88.6	71.1
					69.2			52.3		73.2	
Sheffield	100	800	10560	531	70.3	68.4		59.0		65.5	
Durham	423	3400	45560	346	8.0	9.2	9.6	9.6	7.6	10.2	6.2
Edinburgh	66	528	6811	2148		119.3		88.8		147.8	
Glasgow	629	5032	43980	3900	7.5	9.5		8.0		8.5	
					11.6			11.9		11.4	
Birmingham	152	1584	16996	903							
					89.5			73.7		99.9	
Bristol	82	1320	14744	606	41.1	39.7	12.2	33.7		46.7	
Cambridge	78	528	6146	264	43.6	40.0		33.7		41.9	
					43.1			41.5		39.6	
Oxford	409	3256	33067	939	37.2	54.3		29.7		24.9	
RAL PPD	465	4260	42600	3020	4.6	5.0		6.8		5.0	
Sussex	71	568	5583	84	4.4	6.1		6.4		4.3	
CLOUD											
RAL Tier-1	1670	20041	200410	14087	3.6	4.0		6.2		3.9	
Tier-2 Totals:	5230	48451	526287	33602							
IPv6 Totals:	1555	15570	175450	14529							
IPv6 Percent:	30%	32%	33%	43%							

http://pprc.qmul.ac.uk/~lloyd/gridpp/ukgrid.html

Q: is IPv6 faster or slower than IPv4?

Dest LCG.UKI-LT2-IC-HEP.uk (MB/s)

	lcg	gfal_ipv4	gfal_ipv6	http_ipv4	http_ipv6	xrootd_ipv4	xrootd_ipv6
UKI-LT2-Brunel-disk	40.1	12.5	39.8	13.7	37.9	13.2	50.1
UKI-LT2-IC-HEP-disk	104.5	69.9	92.7	61.9	90.7	_	_
UKI-LT2-QMUL2-disk	7.0	20.3	35.9	27.8	43.0	17.7	51.7
UKI-LT2-RHUL-disk	14.1	12.7		13.5	_	_	_
UKI-NORTHGRID-LANCS-HEP-disk	10.8	5.8	15.3	7.6	13.3	7.4	24.6

Dest LCG.UKI-LT2-Brunel.uk (MB/s)

	lcg	gfal_ipv4	gfal_ipv6	http_ipv4	http_ipv6	xrootd_ipv4	xrootd_ipv6
UKI-LT2-Brunel-disk	69.9	87.7	75.6	98.2	58.2	92.9	96.8
UKI-LT2-IC-HEP-disk	51.3	67.0	66.8	48.6	72.1	_	_
UKI-LT2-QMUL2-disk	26.1	30.2	25.7	35.6	28.8	56.3	43.9
UKI-LT2-RHUL-disk	38.7	20.2		40.2	_	_	_
UKI-NORTHGRID-LANCS-HEP-disk	43.3	5.1	5.5	12.2	8.3	6.7	15.0

Dest LCG.UKI-NORTHGRID-MAN-HEP.uk (MB/s)

	lcg	gfal_ipv4	gfal_ipv6	http_ipv4	http_ipv6	xrootd_ipv4	xrootd_ipv6
UKI-LT2-Brunel-disk	34.4	36.2	56.4	68.4	55.8	62.5	64.5
UKI-LT2-IC-HEP-disk	115.2	55.4	50.0	68.5	122.8	_	_
UKI-LT2-QMUL2-disk	34.0	38.5	50.8	40.9	33.6	36.9	65.4
UKI-LT2-RHUL-disk	41.2	58.3		43.1	_	_	_
UKI-NORTHGRID-LANCS-HEP-disk	83.9	12.5	105.9	19.9	7.2	52.5	29.7

http://pprc.qmul.ac.uk/~lloyd/gridpp/nettest_lcg.html

Summary

- WLCG needs to be ready for any opportunistic IPv6-only CPU resources
- Tier-1s should be providing production storage accessible over IPv6 by April 2018 (8/14 are currently able)
- Tier-2s by the end of 2018 (~20% done)
- ATLAS now running production jobs on IPv6-only worker nodes
- CERN EOS instances are now all dual-stack
- Approximately 12% of FTS transfers go over IPv6 (which appear more efficient/reliable)
- 111/271 perfSONAR hosts now reporting IPv6-enabled
- OPN, USCMS, USATLAS, UK and Belle II perfSONAR meshes now dual-stack
- Some interest in examining performance of IPv6 compared to IPv4