

IPv6 deployment on WLCG

David Kelsey

RAL, STFC, UK Research and Innovation

LHCOPN-LHCONE meeting, Jodrell Bank, 19 March 2025

On behalf of all members of the HEPiX IPv6 working group - (many thanks all!)

G Attebury (UNL), M Babik (CERN), M Bly (RAL), N Buraglio (ESnet),
T Chown (Jisc), D Christidis (CERN/ATLAS), J Chudoba (FZU Prague),
P Demar (FNAL), J Flix Molina (PIC), C Grigoras (CERN/ALICE), B Hoeft (KIT),
H Ito (BNL), D P Kelsey (RAL), E Martelli (CERN), S McKee (U Michigan),
C Misa Moreira (CERN), R Nandakumar (RAL/LHCb), K Ohrenberg (DESY),
F Prelz (INFN), D Rand (Imperial), A Sciabà (CERN/CMS), T Skirvin (FNAL),
C J Walker (Jisc)

- Underlined - attending this meeting
- Many more in the past, and members join/leave from time to time
- **many thanks** also to *WLCG operations, WLCG sites, LHC experiments, networking teams, monitoring groups, storage developers...*

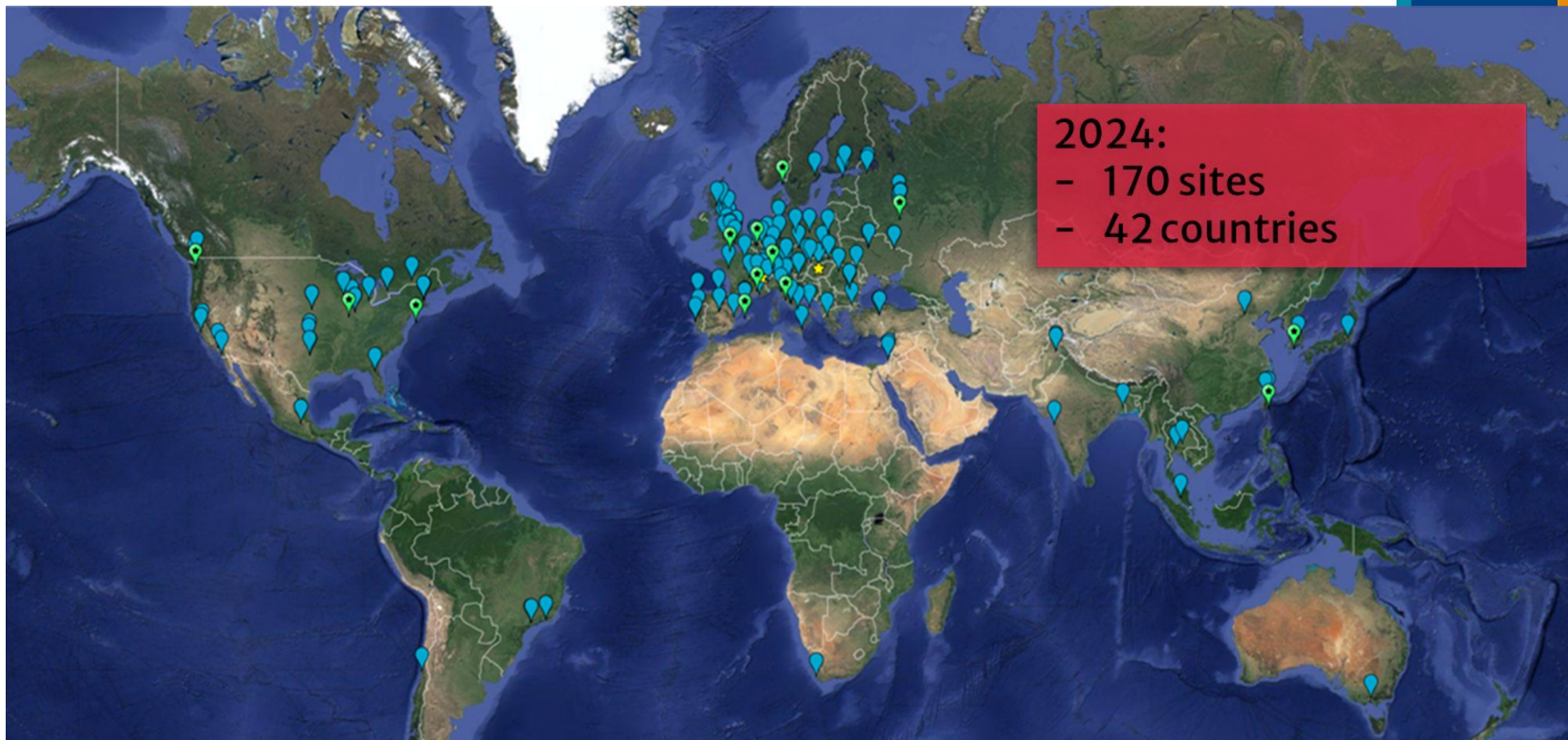
Outline

- The HEPiX IPv6 working group and WLCG - a reminder
- IPv6 on data storage
- IPv6 on CPU & worker nodes
- Identifying and removing use of IPv4
- Plans for IPv6-only WLCG (& IPv6-Mostly)
- Summary & lessons learned

HEPiX IPv6 working group - History and drivers for use of IPv6

- [Phase 1 - 2011-2016](#) - analysis, investigations, testbed, fix storage
- [Phase 2 - 2017-2023](#) - deploy IPv6 on WLCG data storage
- [Phase 3 - 2019-onwards](#) - move towards IPv6-only
 - [2023 onwards](#) - deploy IPv6 on CPU services and WLCG clients
- Sites running out of routable IPv4 addresses (avoid NAT)
 - Use IPv6 addresses for external public networking
- To be ready to support use of IPv6-only CPU clients
- There are **other drivers** for IPv6:
 - scitags.org – packet marking (in header of IPv6 packets)
 - USA Federal Government – [directive](#) on “IPv6-only” (Nov 2020)

Worldwide LHC computing Grid (WLCG)



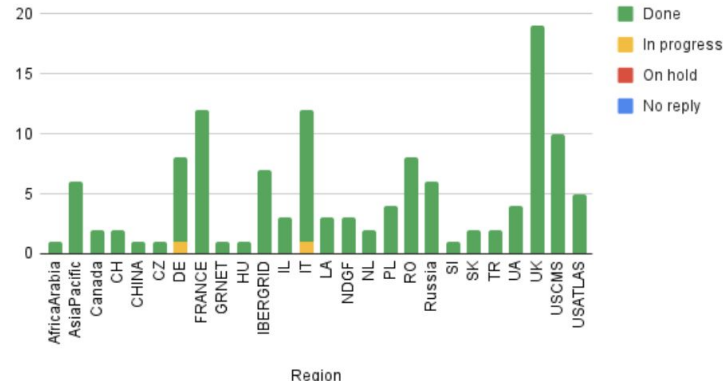
IPv6 on WLCG Storage (Tier2s)

- “IPv6 on storage services” started in 2017
- Goal to support IPv6-only WNs
- Main reason for delay - the institute networking
- **Today, essentially complete**

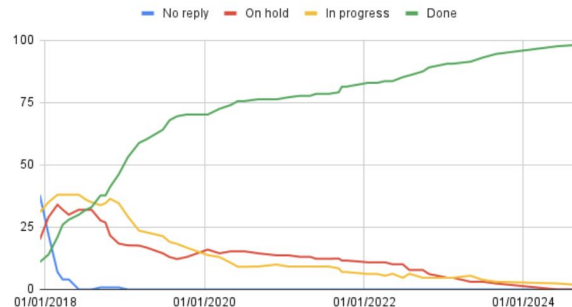
VO	T2 storage on IPv6 (%)
ALICE	94
ATLAS	98
CMS	100
LHCb	100
WLCG	98

(checked on 15-10-2024)

Tier-2 IPv6 deployment status [15-10-2024]



Status vs. time



Status always visible from a [twiki page](#)

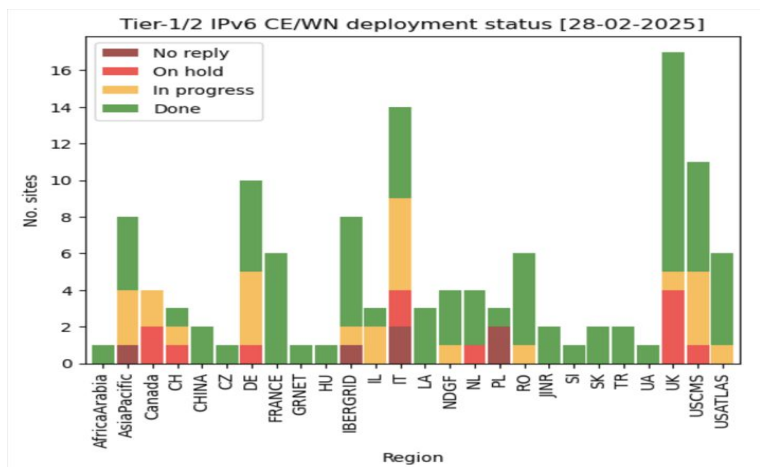
IPv6 on CPU services and Worker Nodes

<https://twiki.cern.ch/twiki/bin/view/LCG/WlcgIpv6>

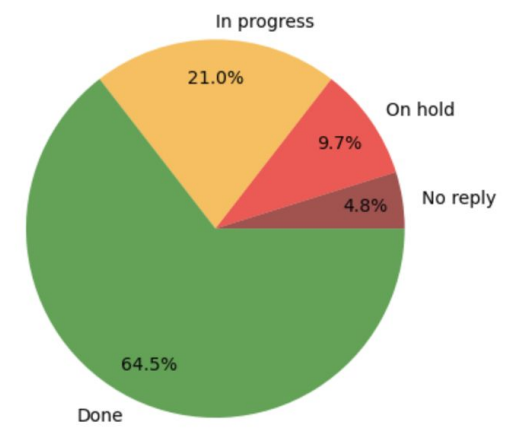
CPU - GGUS support ticket campaign

- Eliminate IPv4 data transfers - between WNs and remote storage
 - Approved by WLCG MB in October 2023
- Launched end of November 2023 to all WLCG sites
- **“Please deploy dual-stack connectivity (IPv4+IPv6) on your computing services (computing elements and worker nodes) as soon as possible and by 30 June 2024 at the latest”**
- Provide estimates for timescale and details on the necessary steps
 - If cannot meet the deadline, then explain why

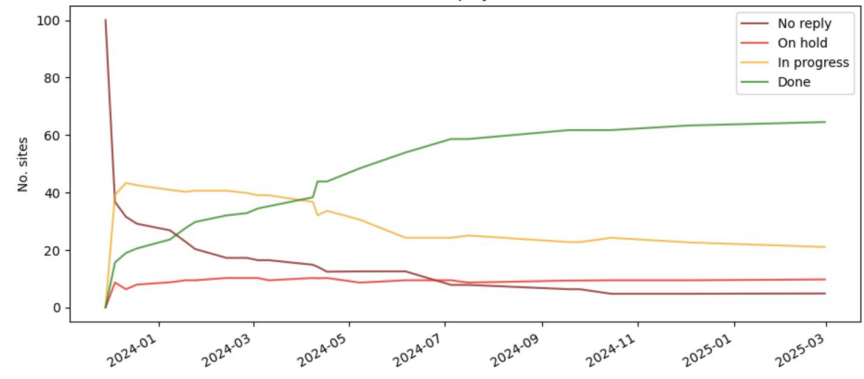
CPU/WN IPv6 status



Tier-1/2 IPv6 CE/WN deployment status [28-02-2025]



Tier-1/2 IPv6 CE/WN deployment status vs. time



65% sites are done
 Status always visible from a [twiki page](#)

Issues reported with enabling IPv6 for WNs/CEs

Common examples:

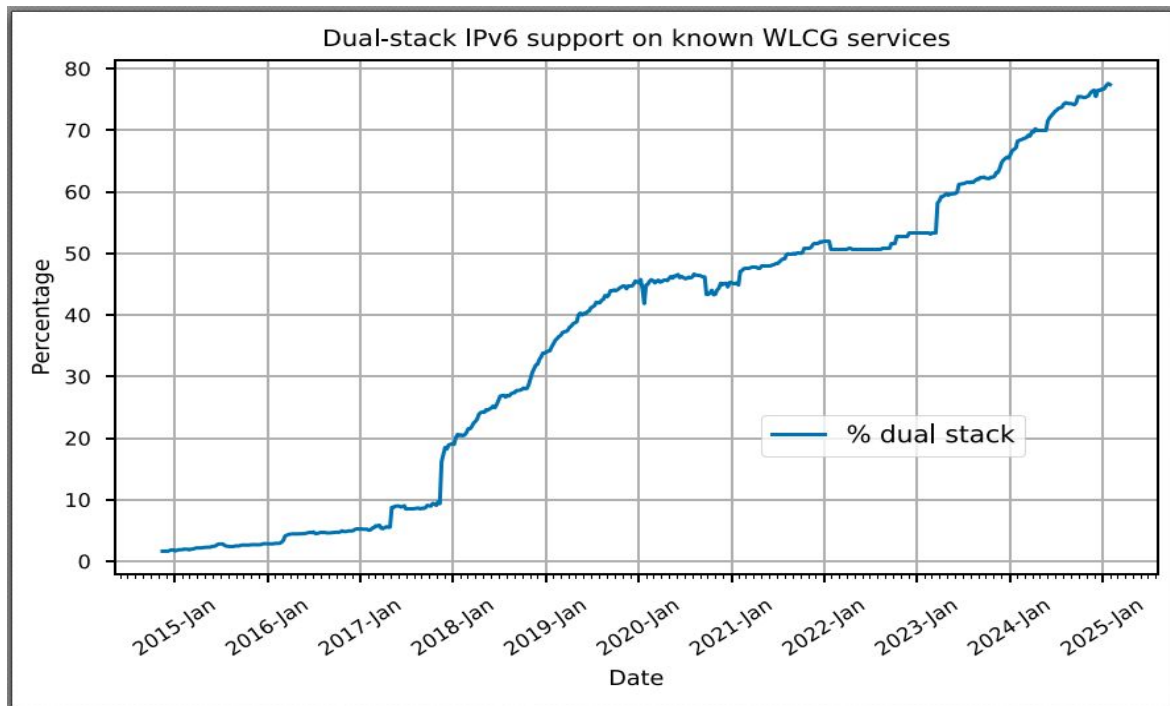
- Enabling IPv6 needs to be coupled to other changes
 - new Data Centres, OS updates, hardware, internal routing changes
- Some sites have more complex/special configurations to consider with respect to NAT for IPv4 and global IPv6
 - e.g., needing to replace IPv4 NAT(s) with dual-stack router(s)
- Other priorities: WLCG authZ tokens or handling the CentOS 7 end-of-life
- Concern that WNs currently behind IPv4 NAT will become more “exposed”

Also worth noting that some sites keen to go IPv6-only now (though not yet recommended), e.g., Brunel University is currently piloting an IPv6-only cluster

And US ATLAS (see Shawn’s talk)

All WLCG services - “VOfeeds”

https://orson.ei.infn.it/~prelz/ipv6_vofeed/



6 Mar 2025:
~78% IPv6/IPv4
~22% IPv4

ALICE experiment monitoring

`curl` connectivity to CERN servers



% IPv6 WNs
since May 2022

Identifying use of IPv4

During WLCG DC24 - and ongoing too

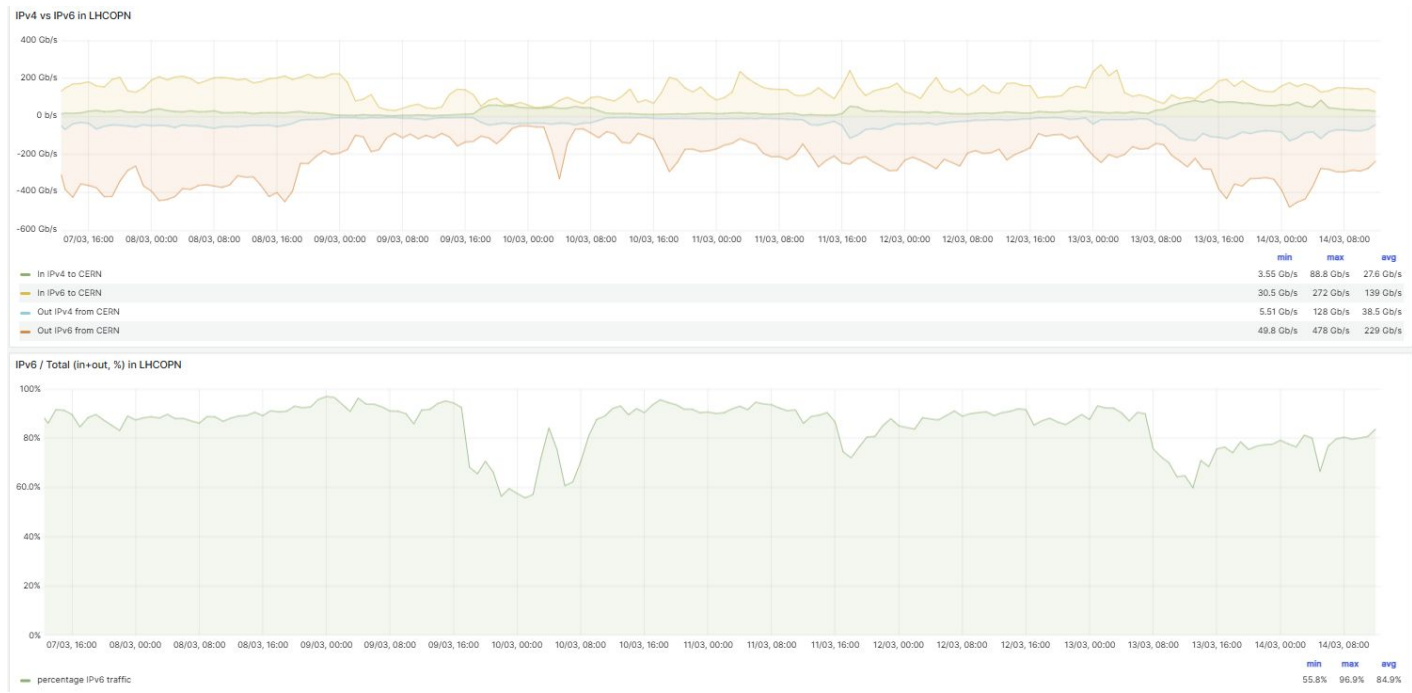
- Work to study the LHCOPN link between CERN and KIT
- Understand when and why IPv4 is being used
- Early on - large IPv4 transfer seen to ALICE at CERN
 - Failed transfers on IPv6 failing over to use of IPv4
- Later - some transfers from KIT to NL-T1
 - All end-points were dual-stack but NL-T1 preferred IPv4 to avoid some observed problems with many concurrent IPv6 streams
- Then XRootD file transfers from CERN
 - Squid at KIT - all would work if IPv6-only but often fails back to IPv4
- Lots of detailed investigations - and STILL ongoing

UK Mini Data Challenge 5-8 March 2025



- Data Challenge CERN to RAL Tier 1 (LHCOPN) - this is **one** of the two 100 Gbps circuits
 - Normal production traffic plus injected traffic CERN to RAL (to saturate 2 *100 Gbps)
 - Total IPv4 traffic CERN to RAL 153 Gb; total IPv6 traffic CERN to RAL 6.0 Tb
- ~98% IPv6**

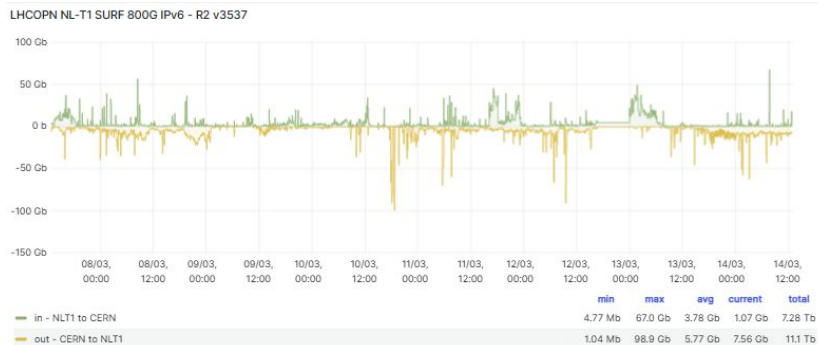
%IPv6 on LHCOPN (CERN) 7-14 Mar 2025



LHCOPN NL T1 March 2025

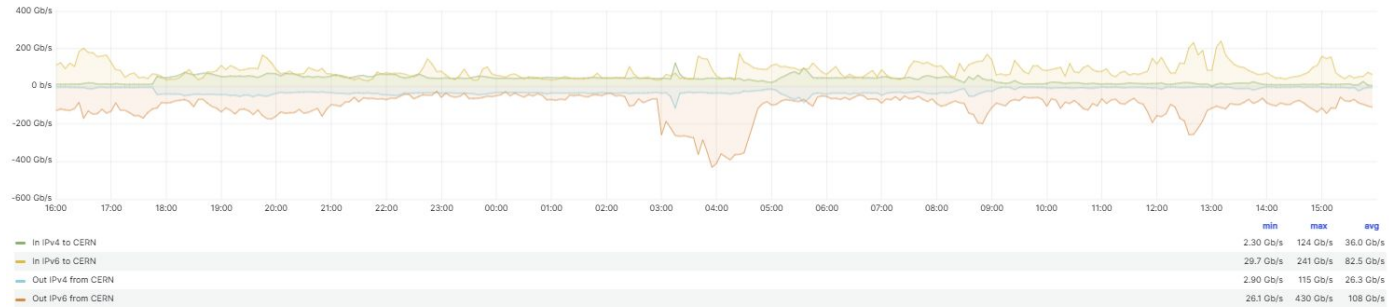


Large IPv4 traffic - CERN to NL
13 and 14 March 2025



LHCOPN all - 9 to 10 Mar 2025 (16:00 to 16:00)

IPv4 vs IPv6 in LHCOPN



IPv6 / Total (in+out, %) in LHCOPN



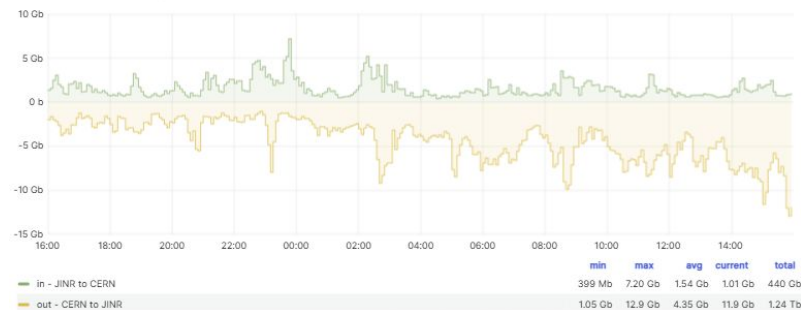
LHCOPN 9Mar to 10 Mar - RU-JINR-T1

LHCOPN RU-JINR-T1 primary IPv4 - R1 v511

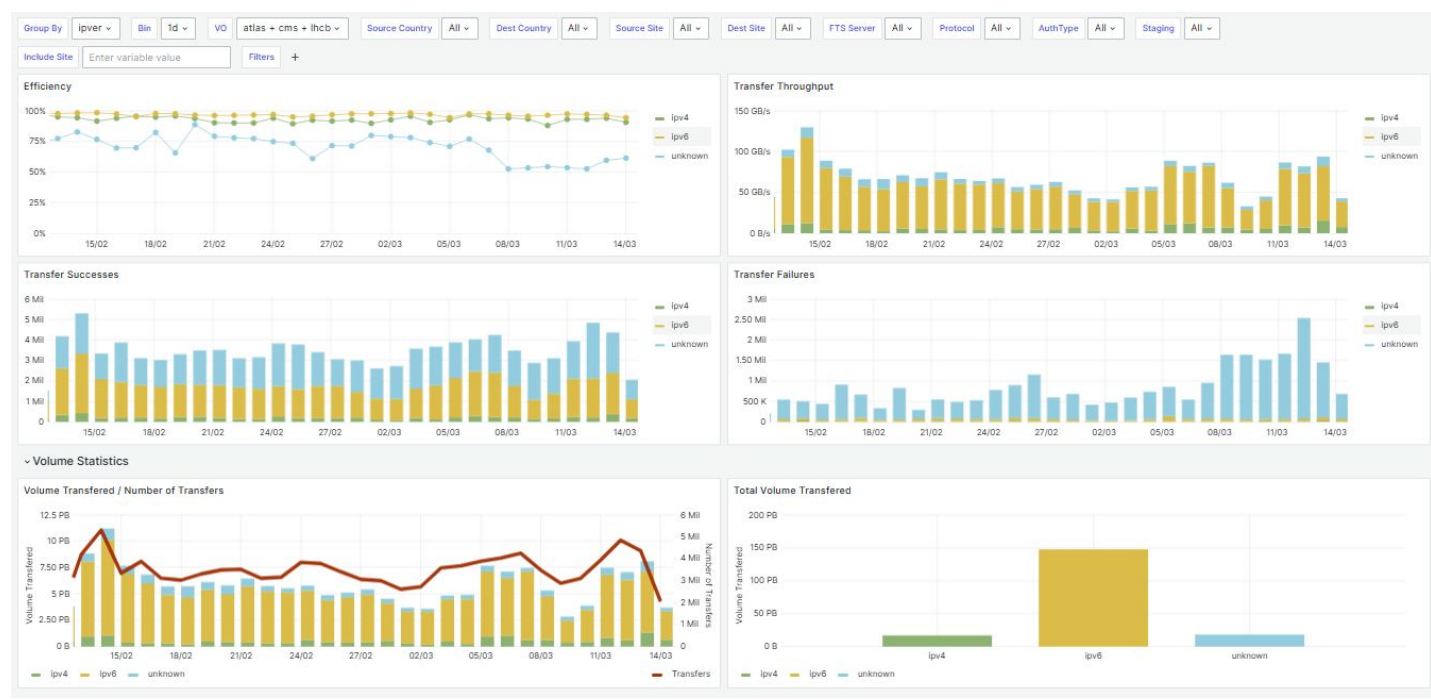


30 Gbps IPv4 CERN to JINR
From 18:00 until 09:00

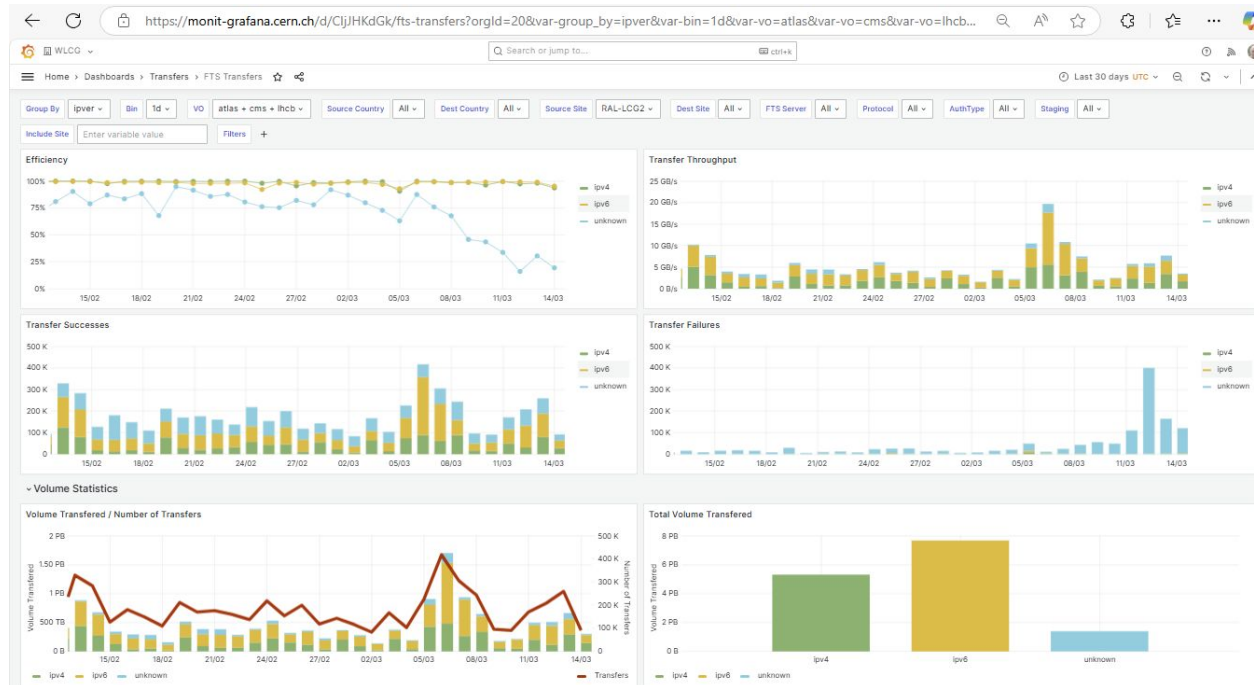
LHCOPN RU-JINR-T1 primary IPv6 - R1 v514



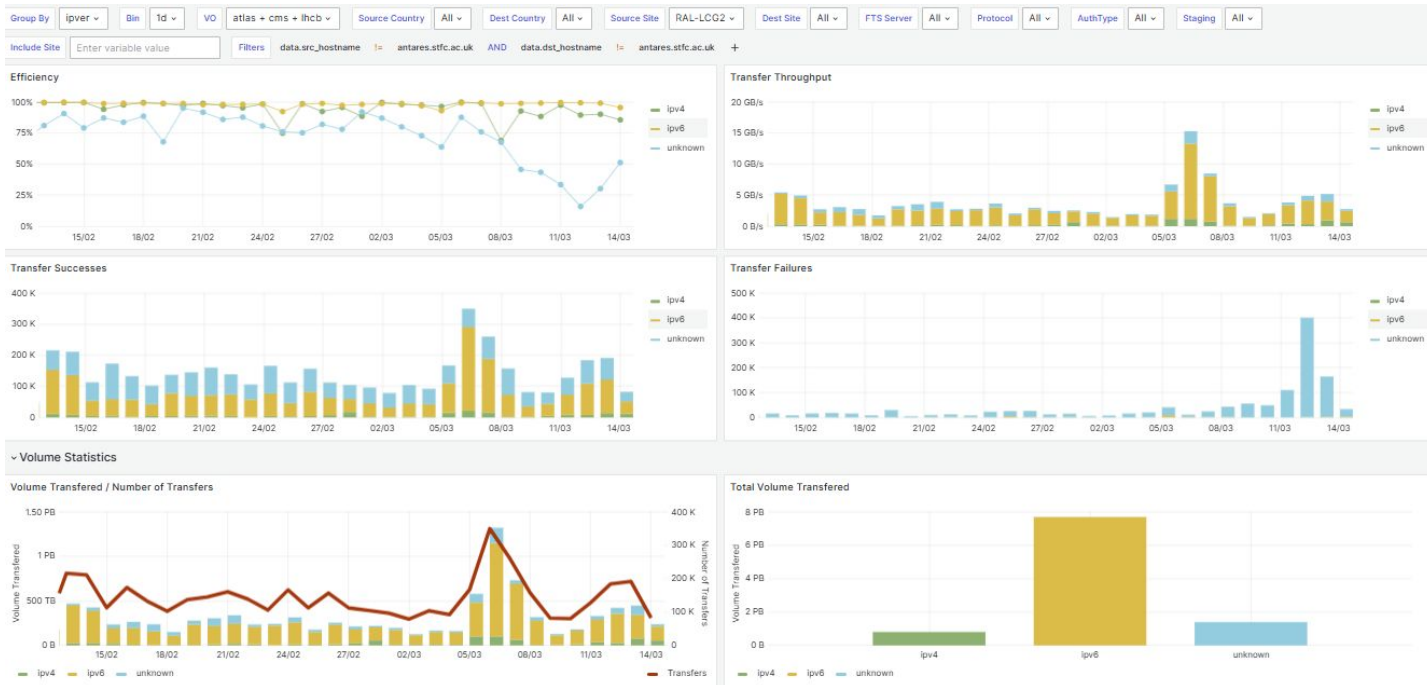
FTS Last 30 days All LHC VOs - All Sites



Same as last slide but Source RAL-LCG2 (but this includes RAL Tape which is IPv4-only)

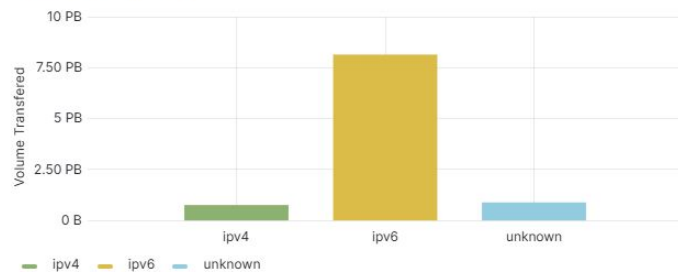


Same as last slide but Source RAL-LOG2 (Here exclude the RAL Tape which is IPv4-only)



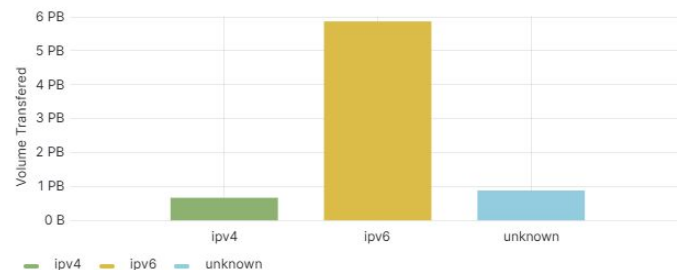
Same as before - various source Tier 1s

Total Volume Transferred

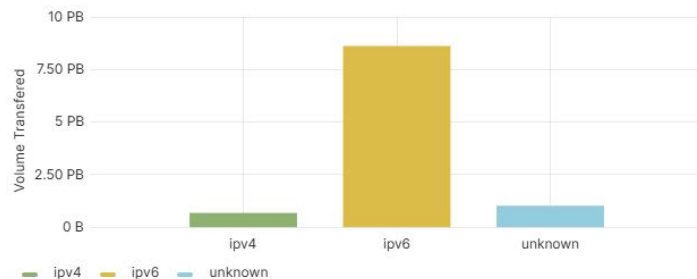


<- DE
IT->

Total Volume Transferred

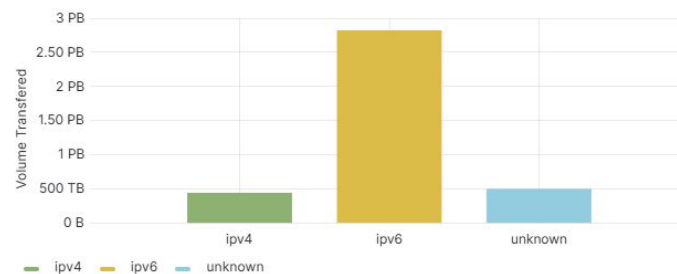


Total Volume Transferred



<- FR
NL ->

Total Volume Transferred



Plans for IPv6-only WLCG

IPv6-only on WLCG (CHEP2019)

<https://doi.org/10.1051/epjconf/202024507045>

- The end point of the transition from IPv4 is an **IPv6-only** WLCG core network - already agreed by WLCG MB - not a dual-stack network!
- To **simplify** operations
 - Dual-stack infrastructure is the most complex
 - Reduced complexity reduces chance of making security errors
 - And dual-stack does not reduce pressure on IPv4 address space
- Large infrastructures (e.g. Facebook, Microsoft,...) use IPv6-only internally
- This is the goal we are still working towards:
 - i.e. *all Wide-area data transfers over IPv6*
- **Timetable** still to be defined - but aiming for “well before LHC Run 4”

Plans for “IPv6-only” WLCG

The plan (LHCOPN):

- By end Run 3 encourage the deployment of IPv6 on **all** WLCG services (today ~80%)
- We need monitoring in place to continue to identify/remove IPv4 on LHCOPN
- Propose IPv4 peering removed from LHCOPN as soon as possible/sensible

For transfers over LHCONE:

- Identify/remove ongoing use of IPv4 on LHCONE network (WLCG traffic)
- We need appropriate monitoring to achieve this
- Prepare the WLCG LHCONE overlay to be IPv6-only (if can be shown to be sensible)

End point of the IPv6 work:

- Complete all above during the first half of the LHC Long Shutdown 3
- Well before start HL-LHC Run 4 - exact date still to be agreed with the WLCG MB

IPv6-Mostly

see <https://datatracker.ietf.org/doc/draft-ietf-v6ops-6mops/>

The current approach in IETF to remove legacy IPv4 devices

- Hosts with CLAT support to turn off IPv4 when NAT64 is available
- The host uses DHCPv4 Option 108 to confirm that the network supports NAT64
- The resulting IPv6-only traffic is then either sent natively to IPv6-enabled destinations or via the CLAT for IPv4-only destinations
- See IETF RFCs: 6146, 6877, 8925

Has been deployed in production on the eduroam/WiFi at Imperial College London (UK)

- A typical week - devices seen on IPv6-only (71K of 87K devices)
- used DHCPv6 Option 108 and CLAT support included in iOS, Android and macOS
- as support grows (Windows CLAT expected soon) the percentage of IPv6 will grow
- legacy devices will be either dual-stack or in some cases IPv4-only

This approach is aimed at user-centric networks but could be useful in WLCG server networks

- given that Linux CLAT implementations are available

Working group observations/questions:

- When should perfSONAR stop performing IPv4 tests?
 - Just for WLCG dashboards?
 - To remove more IPv4 traffic from LHCOPN/LHCONE
- Can we add “IPv4 versus IPv6” traffic split in the WLCG Site egress monitoring network I/O (for DC24) (every minute)?
 - Rather than just total traffic
- Is there evidence that IPv6 networking/routing is more energy efficient than IPv4? (carbon footprint)
 - No NAT boxes
 - Linux kernel IPv6 has been optimised
 - Packet forwarding more efficient - fixed length headers etc.

Summary - and lessons learned

- WLCG already supports use of IPv6-only clients
- Majority of WLCG data transfers already use IPv6
- Campaign for IPv6 on CPU and WN's - well underway
- We still observe ongoing use of legacy IPv4 on LHCOPN/ONE
 - We continue to chase “preference of IPv6” in service configuration
- Complete the move to IPv6-only well before start HL-LHC Run 4
- ***Message to WLCG & LHC experiments:***
 - ***Deploy IPv6 on all services & clients and prefer its use***
- Main lessons learned (for other Research Communities)
 - Everything takes longer than hoped for - start early (or **start with IPv6**)
 - For all Research Communities using MultiONE - **encourage “IPv6-only”**

Questions, Discussion?

Backup slides

“Obstacles” to IPv6

There are many reasons stopping the full use of IPv6/IPv4

- Dual stack is an essential step on the journey to IPv6-only

The Obstacles that we have been addressing:

1. **WLCG Sites not yet deployed IPv6 networking** ~done
 2. **Sites have IPv6 but Tier-2 has no dual-stack storage** ~done
 3. **IPv6 monitoring not available or broken**
 4. **Service is dual-stack but IPv4 still being used**
- Monitoring is essential
 - We continue to chase these problems

Obstacles to IPv6 - being addressed

5. **Non-storage services not yet dual-stack**
 - a. ~75% of all WLCG services are dual-stack today, we need 100%
6. **WLCG client CPU (worker nodes, VMs, containers) some IPv4-only**
 - a. GGUS ticket campaign well underway
7. **Services/clients outside of WLCG Tier-1/Tier-2 not yet addressed**
 - a. Tier-3, Public/Commercial Clouds, Analysis facilities, Experiment portals...
8. **Use of new or evolving technologies not yet tested or tracked**
 - a. New CPU architectures (GPU, non-x86, ...), container orchestration, ...
9. **Staffing issues can be an obstacle**
 - a. Lack of effort, lack of IPv6 training/knowledge, pressure of other work

Some plots: IPv6 and IPv4 traffic
on LHCOPN (5 to 9 Oct 2024)
(and compare with CHEP2023)

Will skip these if no time to show

LHCOPN - %IPv6 traffic - shown at CHEP2023

7 April to 7 May 2023 - shows drops in %IPv6



100% ←

Max
99.5%

Avg
95.3%

Min
70.9%

XRootD file transfer from CERN



```
2024-02-20 06:50:17.012 22.500 TCP 128.142.56.61 59332 192.108.47.90 1094 2.7 M 4.1 G 1.5 G 1499 1
2024-02-20 06:02:38.012 16.000 TCP 128.142.57.111 40594 192.108.47.89 1094 2.7 M 4.1 G 2.1 G 1499 1
2024-01-31 09:33:31.833 11.653 TCP 128.142.63.105 43670 192.108.46.89 1094 2.8 M 4.2 G 2.9 G 1498 1
```

Summary: total flows: 597053, total bytes: 33.0 TeraByte

1625 Server at CERN

```

      • cvmfs-sq4.gridka.de. dual-stack
      • cvmfs-sq1.gridka.de. dual-stack
      • cvmfs-sq3.gridka.de. dual-stack
      • cvmfs-sq5.gridka.de. dual-stack
      • cvmfs-sq6.gridka.de. dual-stack
      • cvmfs-sq2.gridka.de. dual-stack
      • frontier-sq1.gridka.de. dual-stack
      • fw-nat-inside-outside.gridka.de.

```

8 Storage Server at DE-KIT (XRootD Port – 1094):

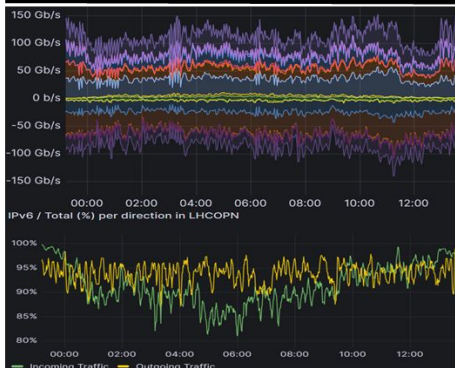
```

      • f01-032-114-e.gridka.de.
      • f01-124-110-e.gridka.de. dual-stack
      • f01-124-159-e.gridka.de. dual-stack
      • f01-124-160-e.gridka.de. dual-stack
      • f01-124-161-e.gridka.de. dual-stack
      • f01-125-159-e.gridka.de. dual-stack
      • f01-125-160-e.gridka.de. dual-stack
      • f01-125-161-e.gridka.de. dual-stack

```

Only 16 Server at KIT

Green line - CERN to KIT %IPv6 70 to 80%



```
2024-02-20 23:26:09.012 0.500 TCP 128.142.249.74 38908 192.108.68.144 1094 12 2266 36256 188 1
2024-02-21 02:39:33.262 0.250 TCP 128.142.240.76 55700 192.108.46.89 1094 10 706 22592 70 1
```

Summary: total flows: 1460049, total bytes: 43.1 TeraByte

2426 Server at CERN

Squid service

Port 3401

```

cvmfs-sq4.gridka.de.
cvmfs-sq1.gridka.de.
cvmfs-sq3.gridka.de.
cvmfs-sq5.gridka.de.
cvmfs-sq6.gridka.de.
cvmfs-sq2.gridka.de.
frontier-sq1.gridka.de.
fw-nat-inside-outside.gridka.de.

```

XRootD Port 1094

```

f01-124-109-e.gridka.de.
f01-124-110-e.gridka.de.
f01-124-112-e.gridka.de.
f01-124-155-e.gridka.de.
f01-124-159-e.gridka.de.
f01-124-160-e.gridka.de.
f01-124-161-e.gridka.de.
f01-125-109-e.gridka.de.
f01-125-110-e.gridka.de.
f01-125-155-e.gridka.de.
f01-125-159-e.gridka.de.
f01-125-160-e.gridka.de.
f01-125-161-e.gridka.de.
f01-125-161-e.gridka.de.
f01-117-137-e.gridka.de.
f01-152-140-e.gridka.de.
f01-152-191-e.gridka.de.
f01-152-192-e.gridka.de.

```

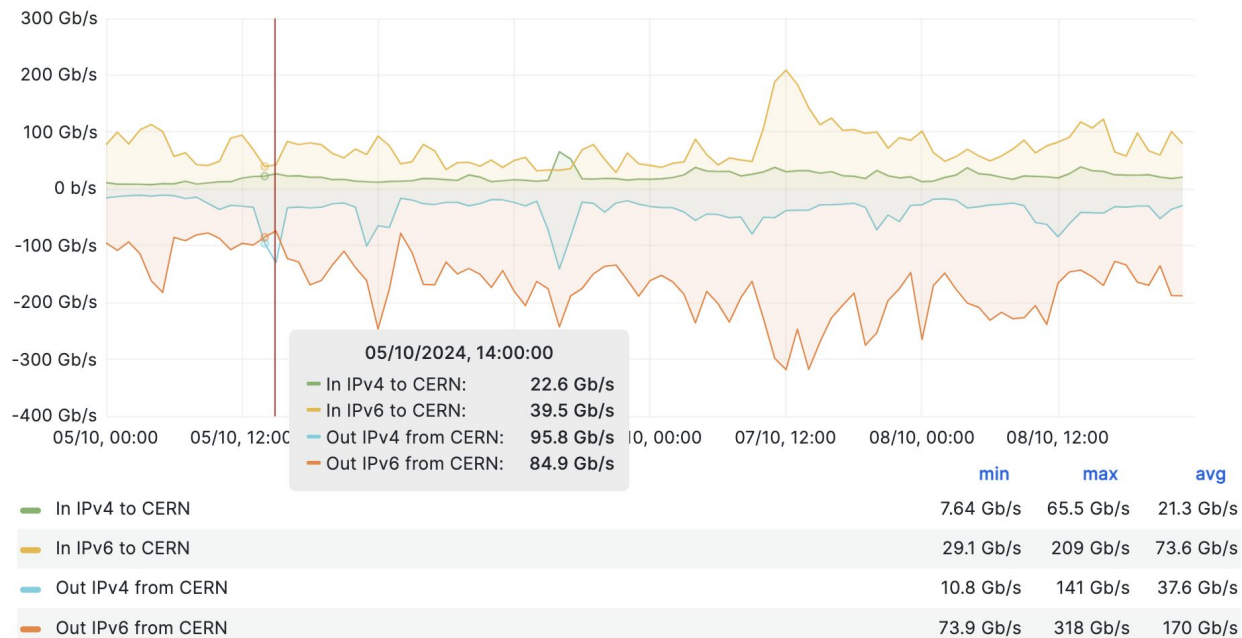
Only 25 Server at KIT

Green line - and again

LHCOPN total traffic, split IPv4 & IPv6 (as seen at CERN)

<https://monit-grafana-open.cern.ch/d/cumEJJb4z/lhcopn-one-ipv6-vs-ipv4?orgId=16&from=1728079200000&to=1728424799000>

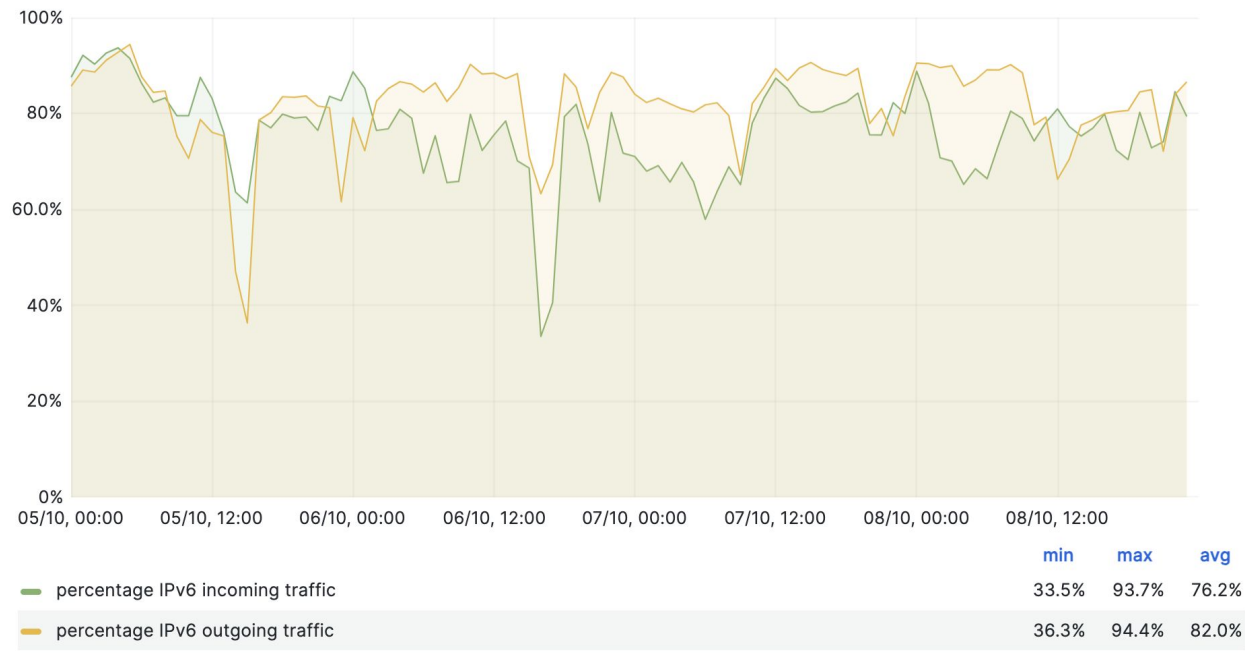
IPv4 vs IPv6 in LHCOPN



- 5 to 9 Oct 2024
- IPv6 Out of CERN
 - Avg 170 Gbps
- IPv4 Out of CERN
 - Avg 37.6 Gbps
- BUT
 - Large IPv4 peaks, e.g.
 - 5/10 @ 14:00
 - Out 95.8 Gbps

%IPv6 traffic - generally high - but large drops down to ~40%

IPv6 / Total (%) per direction in LHCOPN

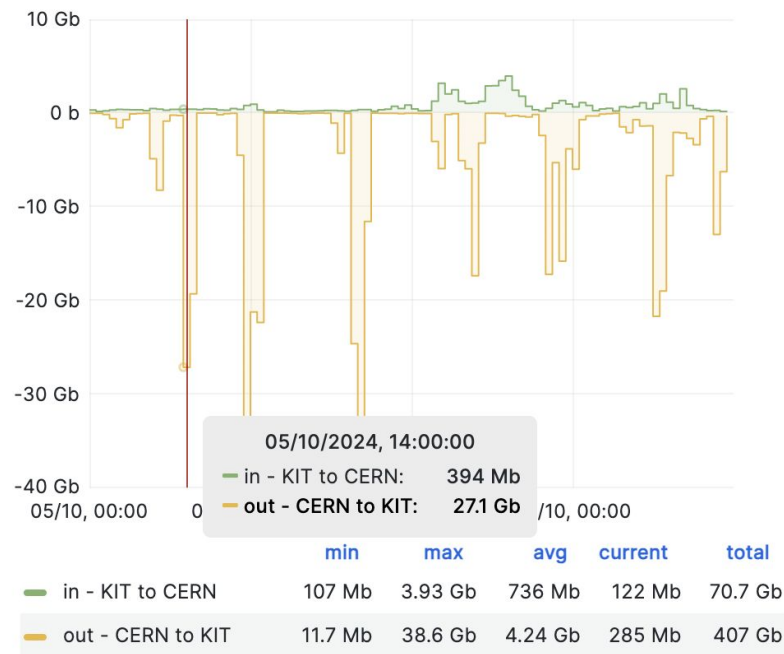


- %IPv6
- In - avg 76.2%
 - Min 33.5%
- Out - avg 82.0%
 - Min 36.3%

LHCOPN traffic (CERN- KIT) German Tier1 - large IPv4 peaks

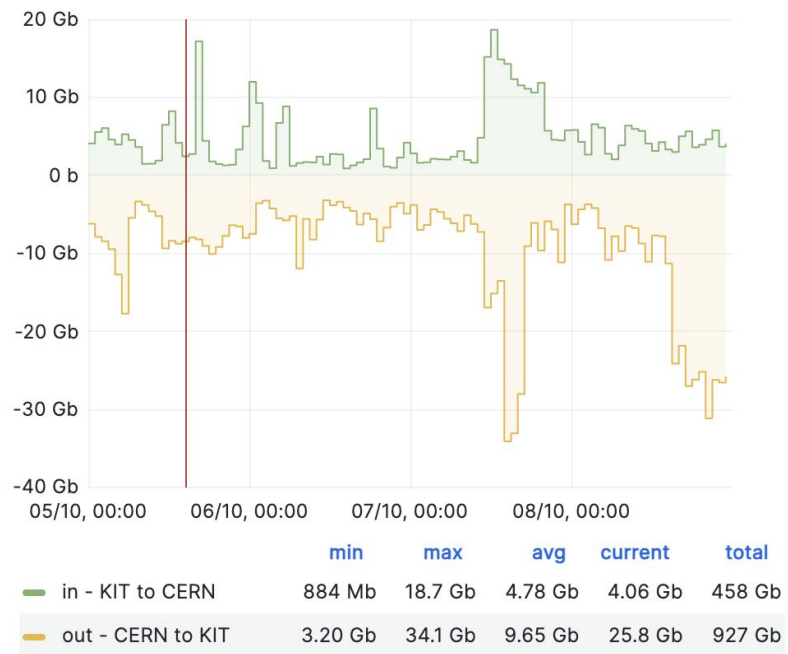
IPv4 plot

LHCOPN DE-KIT 100G IPv4 - R1 v3512



IPv6 plot

LHCOPN DE-KIT 100G IPv6 - R1 v3530



This traffic shown as an example - back in Oct 2024

What are these large peaks of IPv4?

- Not easy
- Need access to Netflow data
- Study IP addresses and Port numbers
 - Aim to identify LHC Experiment
 - Source and Destination address
 - Type of data transfer
- Work in progress
 - But some evidence of Frontier/CVMFS/Squid, etc....