Overcoming obstacles to IPv6 on WLCG

Jiri Chudoba¹, Bruno Hoeft², David Kelsey³, Edoardo Martelli⁴
ISGC2023, Taipei, 24 March 2023

1 - Institute of Physics of the Czech Academy of Sciences (FZU), Czech Republic
2 - Karlsruhe Institute of Technology (KIT), Germany
3 - UKRI Science and Technology Facilities Council (STFC UK), United Kingdom
4 - Conseil Européen pour la Recherche Nucléaire (CERN), Switzerland
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), N Buraglio (ESnet), T Chown (Jisc), D Christidis (U Texas/ATLAS), J Chudoba (FZU Prague), C Condurache (EGI.eu), P Demar (FNAL), J Flix (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), E Simmonds (FNAL), T Skirvin (FNAL)

• 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

• Introduction - WLCG & global IPv6 traffic
• The HEPiX IPv6 working group
• Deployment of IPv6/IPv4 dual-stack storage
• IPv6 monitoring
• Plans for IPv6-only WLCG
• Obstacles to IPv6 on WLCG
• Overcoming those obstacles
• Summary
Introduction - WLCG & global IPv6 traffic
Worldwide LHC Computing Grid (WLCG)
Computing for the CERN Large Hadron Collider

- The WLCG is a global collaboration
  - > 170 computing centres, > 40 countries
- Its mission is to **store, distribute** and **analyse** the data generated by the LHC experiments
- Sites hierarchically arranged in three tiers:
  - Tier-0 at CERN
  - 14 Tier-1s (national laboratories)
  - ~ 160 Tier-2s (universities)
- > 1M CPU cores (> 2M jobs per day)
- > 1EB of data storage
Worldwide LHC computing Grid (WLCG)

2022:
- 170 sites
- 42 countries
LHCOPN - Optical Private Network

https://twiki.cern.ch/twiki/bin/view/LHCOPN/OverallNetworkMaps
LHCONE map
HEP Networking - uses GÉANT

And also ESnet & Internet2 (USA), NORDUnet and many other national networks
General IPv6 traffic continues to grow (Google and Facebook)

Google

Facebook
(and also) IPv6 users -6lab.cisco.com
The HEPiX IPv6 working group
The HEPiX IPv6 Working Group

- Started in April 2011
  - some HEPiX sites running out of IPv4 addresses
  - IANA projecting imminent IPv4 address exhaustion
  - Moving to support IPv6 would not be fast - better start now!
- **Phase 1** - 2011-2016 - full analysis, investigations, ran a testbed
  - lots of work by storage developers to be IPv6-capable
- **Phase 2** - 2017-2020 - deploy dual-stack storage on WLCG
  - in production
- **Phase 3** - 2021-onwards - plan for IPv6-only
  - investigate reasons for data transfers over IPv4

[https://www.hepix.org/e10227/e10327/e10326/](https://www.hepix.org/e10227/e10327/e10326/)
[https://indico.cern.ch/category/3538/](https://indico.cern.ch/category/3538/) (meetings)
Drivers for use of IPv6

• 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

• BUT there are other drivers for IPv6 too
  • scitags.org – packet marking (in header of IPv6 packets)
    • Research Networking Technical Working Group (RNTWG)
  • USA Federal Government – directive on “IPv6-only” (Nov 2020)
  • multiONE (several LHCONE for different communities)
    • uses the scitags marks in header flow label for policy based routing
US Government IPv6 Mandate
(from Phil Demar and Nick Buraglio)

FY23
• All new federal systems to be IPv6 enabled at deployment.
• 20% of all networked federal systems IPv6-only

FY24
• 50% of all networked federal systems IPv6-only

FY25
• 80% of all networked federal systems IPv6-only
• Identify, plan, schedule retirement/replacement of remaining networked systems that cannot be converted to IPv6-only

• US National Labs (T1s) are included; but university-run T2s are not subject to the mandate
Deployment of IPv6/IPv4 dual-stack storage
IPv6/IPv4 deployment at WLCG Tier-2 sites

- The deployment campaign was launched in November 2017
- Steady progress (status)
  - \(~91\%\) of Tier-2s have dual stack storage
  - 91% of storage

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Fraction of T2 storage accessible via IPv6</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALICE</td>
<td>90%</td>
</tr>
<tr>
<td>ATLAS</td>
<td>89%</td>
</tr>
<tr>
<td>CMS</td>
<td>96%</td>
</tr>
<tr>
<td>LHCb</td>
<td>79%</td>
</tr>
<tr>
<td>Overall</td>
<td>91%</td>
</tr>
</tbody>
</table>
Tier-2 evolution of dual-stack

Status vs. time

Reason of delay [13-02-2023]

No dual stack
9.1%

Network
90.9%
IPv6 monitoring
Importance of monitoring

- Must monitor
  - deployment of IPv6-capable services
  - fraction of data transfers taking place over IPv6
- Monitoring implementations used for IPv6
  - perfSONAR
  - ETF - experiment test framework
  - FTS (File Transfer Service)
  - Network utilisation and traffic plots
    - e.g. IPv6 versus IPv4 on LHCOPN/LHCON as seen at the CERN routers
- But in recent years some existing monitoring has ceased to work
% of WLCG FTS data traffic over IPv6

- Some FTS protocols, e.g. DAVS still not able to monitor IPv6 traffic
- DAVS excluded from this plot
  - and explains end date here
- GSIFTP & SRM are OK

![Graph showing WLCG FTS data transfers - percentage of IPv6 traffic.](image)
Broken FTS IPv6 monitoring

- Experiments **no longer** using GSIFTP & SRM
  - Moving to HTTP/WebDAV
- Reason for no plots in 2022
- IPv6 HTTP/WebDAV is not fully “visible” in our FTS monitoring!

**In the next slide**
- IPv4 numbers are wrong!
- WebDAV monitor is being made capable of splitting IPv6 from IPv4
- In past, wrongly assumed “IPv4” when the IP version is “unknown”
- Conclusion – the **amount of IPv6 traffic in 2022 is UNKNOWN**
- This is now being fixed - but will take time to deploy
II. FTS & IPv6 Monitoring – the bad

“IPv4” includes “unknown”

Charts plotted using the FTS Aggregated data
IPv6 traffic on LHCOPN/LHCONE at CERN

- Problems with data from April 2022 onwards
- May not be accurate from April 22
- Ratio IPv6/IPv4 may be correct
- Required fix from vendor
- Next slides fixed (after 3 Mar 23)

LINK to these plots
“Fixed” plots from CERN (>3 Mar 23)
“Fixed” plots (2) from CERN (>3 Mar 23)

Average rates more than 85% IPv6 in each direction
Examples from a WLCG site (FZU)

FZU (praguelcg2) network monitoring
Examples from a WLCG site (FZU)
Examples from a WLCG site (FZU)
Plans for IPv6-only WLCG
WLCG - from dual-stack to IPv6-only (CHEP2019) https://doi.org/10.1051/epjconf/202024507045

• The end point of the transition from IPv4 is an IPv6-only WLCG core network
• To simplify operations
  • Dual-stack infrastructure is the most complex
  • Dual-stack has more security threat vectors
• Large infrastructures (e.g. Facebook) use IPv6-only internally
• The plan - the goal we are working towards
  • IPv6-only for the majority of WLCG services and clients
  • With ongoing support for IPv4-only clients where needed
• Timetable to be defined
Obstacles to IPv6 on WLCG
Why: IPv4 transfers on LHCOPN?

- Tier-1s are dual-stack, but IPv4 often still used for transfers
  - Site/experiment issues
    - Old software stacks (legacy deployments)
  - both ends dual-stack but configuration prefers IPv4
  - transfers are to/from WN's - and the WN's are IPv4-only
- IPv6 WG has been analyzing Tier-1 top-talkers over IPv4
  - understand reasons for IPv4 and request fixes to problems
- encourage all sites to deploy CPU as dual-stack or IPv6-only
- encourage all sites and all experiments to "prefer" IPv6
## The obstacles to IPv6

<table>
<thead>
<tr>
<th>Obstacles to IPv6 on WLCG</th>
<th>Within IPv6 WG or WLCG MB control?</th>
<th>Status</th>
<th>Possible actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Institute/Site not yet providing IPv6 network support</td>
<td>No</td>
<td>&gt;90% WLCG sites do support IPv6 networking</td>
<td>WG, MB and experiments can continue to “encourage”</td>
</tr>
<tr>
<td>Site provides IPv6 network support but still no dual-stack storage</td>
<td>Yes - MB mandate</td>
<td>&gt;90% sites offering at least one dual-stack storage service</td>
<td>Experiments can request deployment of dual-stack storage. MB could “demand”</td>
</tr>
<tr>
<td>Non-storage services not IPv6 capable</td>
<td>Partial - no MB mandate</td>
<td>Not yet tracked. We have evidence suggesting ~60% of these services are dual-stack</td>
<td>Sites and experiments continue controlling and negotiating constraints that prevent IPv6. WG can continue to encourage</td>
</tr>
</tbody>
</table>

WG = HEPiX IPv6 Working Group       MB = WLCG Management Board
# The obstacles to IPv6 (2)

<table>
<thead>
<tr>
<th>Obstacles to IPv6 on WLCG</th>
<th>Within IPv6 WG or WLCG MB control?</th>
<th>Status</th>
<th>Possible actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>WLCG clients (CPU) not IPv6 capable</td>
<td>Partial - no MB mandate</td>
<td>Not yet tracked. WG encourages sites. Some sites have fully deployed dual-stack worker nodes</td>
<td>Sites and experiments continue controlling and negotiating constraints that prevent IPv6. WG can continue to encourage</td>
</tr>
<tr>
<td>Monitoring is not available to track use of IPv6 by data transfers</td>
<td>No - but MB supports the need for this</td>
<td>Partially available. FTS now handles “IPvversion”; Storage must provide “IPvversion” on file close</td>
<td>WG to continue to encourage. Storage developers to provide IP version in logs and new implementations to be deployed by sites</td>
</tr>
<tr>
<td>Service is IPv6 capable but chooses not to use IPv6 (configuration error or deliberate choice).</td>
<td>Partial</td>
<td>A WG priority during 2023. We are actively seeking, tracking and chasing.</td>
<td>Experiments either control the configuration or can request configuration changes. WG to request.</td>
</tr>
</tbody>
</table>

Should we ask sites to move straight to IPv6-only CPU?
Overcoming those obstacles

Also - see talk at this conference

- IPv4 to IPv6 Worker Node migration in WLCG by Bruno Hoeft (KIT)
Why is IPv4 used between dual-stack endpoints?

**HTCondor job logs** During summer of 2022
- study in June (17.8K jobs) for one LHC experiment
  - many HTCondor Scheduler Daemons not dual-stack (close to zero for “analysis” jobs)
- experiment was informed
  - they configure IPv6 preference on Schedds
- second study in September (16K jobs) showed a significant improvement
  - “analysis” now has 19% of the dual-stacked Schedds

Study of the **IPv4 and IPv6 Top-talkers** (each month) on LHCOPN/LHCONE (at CERN)
- Found many different sites and endpoints using IPv4
- Not easy to find protocols/services
- 2 improvements will help
  - weekly basis rather than monthly
  - log port numbers and not just IP addresses

Work continues on these studies
Some other obstacles fixed

Data transfers into USA/ATLAS Great Lakes Tier 2 (AGTL2)
Found to use IPv4 even when both ends dual-stack (dCache/WebDAV)
java.net.preferIPv6Addresses (default: false) - Now set to “true”
Fixed at 17:00 on 14 Feb 2022 (confirmed in the plot!)
This fix is essential for all dCache instances - fixed in v7.2.11

IPv6 is yellow

Some FTS monitoring now able to distinguish IPv6 from IPv4

ATLAS & CMS HTTP transfers into CERN (last year) – IPv6
showing from August 2022 onwards
Summary
Summary

• WLCG is ready to support use of IPv6-only clients
• Tier-1s all have production storage accessible over IPv6
• Tier-2s >90% sites are IPv6 capable
• Monitoring data transfers is essential - broken and being fixed
• We have investigated obstacles to IPv6 in WLCG (seeking fixes)
  • e.g. Why do two dual-stack endpoints use IPv4 between them?
• Phase 3 – we are planning for move to IPv6-only services
  • Dual-stack is NOT the desired end-point!

• message to new research communities - build on IPv6 from start
Questions, Discussion?
Backup slides
The CERN Large Hadron Collider (LHC)

How do you get from this to this?

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

4 July 2012

Higgs boson-like particle discovery claimed at LHC

By Paul Rincon
Science editor, BBC News website, Geneva

Cern scientists reporting from the Large Hadron Collider (LHC) have claimed the discovery of a new particle consistent with the Higgs boson.
Imperial London - LHCONE - 100 Gbps on IPv6

https://shapingthefutureofjanet.jiscinvolve.org/wp/uncategorized/100gbps-of-cern-data-over-ipv6-on-the-janet-network/

Figure 1 — Imperial monitoring shows the two-hour period where the 100G link was filled and where 100% of the LHCONE traffic was IPv6.

Figure 2 — The traffic levels seen in the network view correspond to those seen by the WLCG File Transfer Service (FTS) visualization tools.

Figure 3 — It was also interesting to see this traffic reflected in the monitoring platform for the GÉANT pan-European research and education backbone network.
Messages to WLCG sites

WLCG MB statements (July 2021)

• Deployment of dual stack storage remains the priority
  • this is a prerequisite to fully supporting IPv6-only WNs
• All sites and regions should plan accordingly and as soon as possible
• The final goal is IPv6-only (timetable to be agreed later)

Encouragement: IPv6 WG to all sites and experiments
• deploy all WN, VM, containers, local services as dual-stack
• Configure to enable and "prefer" IPv6 transfers
New “fixed” plots from CERN

IPv4 vs IPv6 in LHCOPN

- In IPv4 to CERN: min 607 Mb/s, max 55.0 Gb/s, avg 18.7 Gb/s
- In IPv6 to CERN: min 9.24 Gb/s, max 73.6 Gb/s, avg 32.7 Gb/s
- Out IPv4 from CERN: min 2.12 Gb/s, max 59.6 Gb/s, avg 18.1 Gb/s
- Out IPv6 from CERN: min 17.9 Gb/s, max 150 Gb/s, avg 77.6 Gb/s

URL: [https://monit-grafana-open.cern.ch/d/cumEJJb4z/lhcopn-one-ipv6-vs-ipv4?orgId=16](https://monit-grafana-open.cern.ch/d/cumEJJb4z/lhcopn-one-ipv6-vs-ipv4?orgId=16)
New “fixed” plots from CERN (2)