EPJ Web of Conferences **245**, 07045 (2020) https://doi.org/10.1051/epjconf/202024507045
CHEP 2019

IPv6-only networking on WLCG

Marian Babik[1], *Martin* Bly[2], *Tim* Chown[3], *Dimitrios* Christidis[4], *Jiři* Chudoba[5], *Catalin* Condurache[6], *Thomas* Finnern[7], *Terry* Froy[8], *Costin* Grigoras[1], *Kashif* Hafeez[2], *Bruno* Hoeft[9], *David* Kelsey[2*], *Raul* Lopes[10], *Fernando* López Muñoz[11,12], *Edoardo* Martelli[1], *Raja* Nandakumar[2], *Kars* Ohrenberg[7], *Francesco* Prelz[13], *Duncan* Rand[14], and *Andrea* Sciabà[1]

[1]European Organization for Nuclear Research (CERN), CH-1211 Geneva 23, Switzerland
[2]UKRI STFC Rutherford Appleton Laboratory, Harwell Campus, Didcot OX11 0QX, United Kingdom
[3]JISC, Lumen House, Library Avenue, Harwell Campus, Didcot OX11 0SG, United Kingdom
[4]University of Texas at Arlington, Arlington TX, United States of America
[5]Institute of Physics, Academy of Sciences of the Czech Republic, Na Slovance 2 182 21 Prague 8, Czech Republic
[6]EGI Foundation, Science Park 140, 1098 XG Amsterdam, The Netherlands
[7]Deutsches Elektronen-Synchrotron DESY, Notkestraße 85, D-22607 Hamburg, Germany
[8]Queen Mary University of London, Mile End Road, London E1 4NS, United Kingdom
[9]Karlsruhe Institute of Technology, Hermann-von-Helmholtz-Platz 1, D-76344 Eggenstein-Leopoldshafen, Germany
[10]College of Engineering, Design and Physical Sciences, Brunel University London, Uxbridge UB8 3PH, United Kingdom
[11]Port d'Informació Científica, Campus UAB, Edifici D, E-08193 Bellaterra, Spain
[12]Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain
[13]INFN, Sezione di Milano, via G. Celoria 16, I-20133 Milano, Italy
[14]Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom

Abstract. The use of IPv6 on the general Internet continues to grow. The transition of the Worldwide Large Hadron Collider Computing Grid (WLCG) central and storage services to dual-stack IPv6/IPv4 is progressing well, thus enabling the use of IPv6-only CPU resources as agreed by the WLCG Management Board and presented by us at earlier CHEP conferences. During the last year, the HEPiX IPv6 Working Group has continued to chase and support the transition to dual-stack services. We present the status of the transition and some tests that have been made of IPv6-only CPU showing the successful use of IPv6 protocols in accessing WLCG services. The dual-stack deployment does however result in a networking environment which is more complex than when using just IPv6. The group is investigating the removal of the IPv4 protocol in places. We present the areas where this could be useful together with our future plans.

1 Introduction

The HEPiX IPv6 Working Group [1] has been investigating the many issues related to the move of the Worldwide LHC Computing Grid (WLCG) services to dual-stack IPv6/IPv4

*e-mail: david.kelsey@stfc.ac.uk

networking, thus enabling the use of IPv6-only CPU resources as agreed by the WLCG Management Board and presented by us at CHEP2018 [2].

The dual-stack deployment does however result in a networking environment which is more complex than when using just IPv6. Some WLCG services, e.g. the EOS storage system at CERN [3], are already using IPv6-only for internal communication, where possible. Several Broadband/Mobile-phone companies, such as T-Mobile in the USA and BT/EE in the UK, now use IPv6-only networking with connectivity to the IPv4 legacy world enabled by the use of NAT64 (RFC 6146 [4]), DNS64 (RFC 6147 [4]) and 464XLAT (RFC 6877 [4]). Large companies, such as Facebook, use IPv6-only networking within their internal networks, there being good management and performance reasons for this. Based on these examples of IPv6-only networking, we have therefore been motivated to investigate the future removal of the IPv4 protocol in places within the WLCG infrastructure.

This paper presents the status of the WLCG transition to dual-stack services, together with our work and plans for moving to an IPv6-only networking environment for WLCG.

2 Status of the transition to dual-stack storage

The long process of enabling the protocol IPv6 at LHC started already 10 years ago in 2010. Today, after extensive testing by the HEPiX IPv6 Working Group [5] and the strong support of the storage developer community, the current WLCG storage and grid-middleware applications fully support the use of IPv4 and IPv6 protocols simultaneously; they are dual-stack ready or even protocol agnostic.

2.1 Deployment at Tier-0 and Tier-1s

After the aforementioned ten years the storage environment is almost completely dual-stack ready. At the CERN WLCG Tier-0 and at the 14 Tier-1s, dual-stack IPv6/IPv4 is nearly fully enabled. Only the Tier-1 site at the Kurchatov Institute in Moscow, part of the Russian Federation, is still currently running on IPv4-only. This enables a total of 96% of the Tier-1 storage of WLCG to be accessible via IPv6 as shown in table 1. The set of Tier-1 and Tier-2 sites used by each experiment is different and therefore the fraction of storage available over IPv6 per experiment also differs.

Table 1. Fraction of Tier-1 and Tier-2 storage available over IPv6

	ALICE	ATLAS	CMS	LHCb	Global
Tier-1 storage	78%	96%	100%	94%	96%
Tier-2 storage	86%	59%	89%	75%	74%

The File Transfer Service (FTS [6]) is responsible for distributing the majority of the LHC data across the WLCG infrastructure. The FTS server at FNAL is still currently running in IPv4-preferred mode. There was a long-standing transfer malfunction issue to IPv4-only Tier-2 sites in the USA which is now solved. This last server will be deployed in dual-stack in the near future.

2.2 Deployment at Tier-2 sites

The deployment of IPv6 at Tier-2 sites is still proceeding even after the deadline expired at the end of 2018. It was decided not to give the deadline a formal extension, but just to encourage all remaining sites to complete the IPv6 deployment “as soon as possible”: the

main motivations were that *a)* sites behind schedule were encountering objective difficulties and *b)* the most effective deadline would be imposed by the experiments themselves, if they wished, for example, to require IPv6 for production. This choice was confirmed by the steady progress observed during 2019, as it can be seen in figure 1.

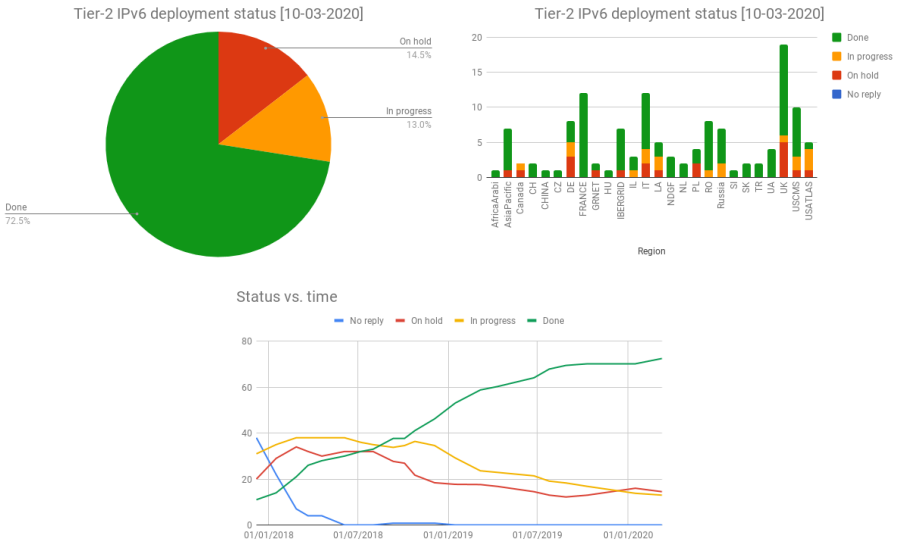

Figure 1. (left) Tier-2 deployment status by site globally, (right) by region, and (bottom) time evolution

The time evolution of the site status shows a steady increase of the number of sites that have deployed IPv6, until a more recent slowdown. This is consistent with the hypothesis that the remaining sites are those facing the biggest difficulties. The progress of each Tier-2 site is recorded in a support tracking system with a separate ticket assigned to each site. A detailed analysis of these tickets shows that, in many cases, sites need to wait for the IPv6 deployment on site, which often depends on people different from the WLCG site staff. The fraction of the Tier-2 storage that is accessible via IPv6 is shown in table 1 for each experiment, and significant differences are apparent. Two experiments (ALICE and CMS) are very close to having all their Tier-2 storage on IPv6, LHCb has little Tier-2 storage to begin with due to their particular computing model and ATLAS is getting better, but still far from the goal.

2.3 LHCOPN and LHCONE

The Large Hadron Collider Optical Private Network (LHCOPN [7]) and the LHC Open Network Environment (LHCONE [7]) are both virtual private networks serving the LHC Experiments. Since the end of 2016 both networks are dual-stack ready. LHCOPN is a CERN (Tier-0) centric star network mainly deployed for the distribution of the raw detector data to the Tier-1s. Even though the majority of Tier-1s are dual-stack ready and the IPv6 protocol is preferred, we still observe transfers over IPv4, in part because the FTS server at FNAL is still running in IPv4-mode. The LHCONE network consists of approximately 140 sites connected through Virtual Routing and Forwarding implementations at 26 different network service providers. The network itself has been IPv6-ready for several years. The connected end sites are gradually becoming IPv6-ready.

2.4 WLCG data transfers

For more than the last two years, since we started encouraging the Tier-2 transition, we have been regularly tracking the fraction of WLCG data transfers that take place over IPv6. We have been able to use the IPv6 protocol filter in the monitoring of the total WLCG FTS [6] data transfers. The fraction of WLCG FTS data transfers over IPv6 as a function of date is shown in figure 2.

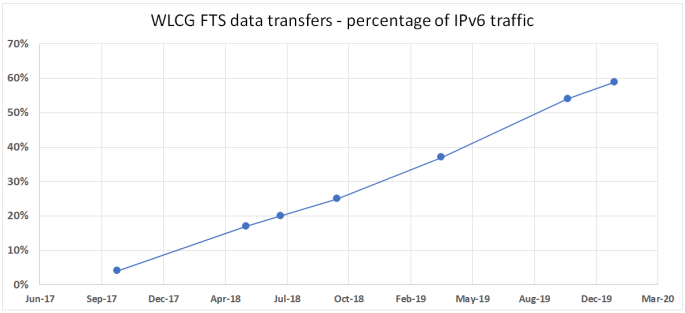

Figure 2. Percentage of FTS data transfers over IPv6. Each data point shows the average percentage over the previous 30 days

We have been aware of the fact that some data transfers between systems have been taking place over IPv4 even when both ends are dual-stack enabled. In the majority of cases this has turned out to be due to configuration settings which have either deliberately or accidentally been set to prefer IPv4. Preferring IPv6 reduces the use of IPv4 as a step to the end game of removing IPv4. We note that when transfers do take place over IPv6, they do so in such a way that the LHC experiments do not notice any difference in behaviour compared with transfers over IPv4.

3 IPv6-only networking

A few years ago, RFC 6586 [4] reported on a survey on IPv6-only networking for mainstream applications (gaming, telephony, multimedia, etc.) and observed that *"it is possible to employ IPv6-only networking"* and that *"for large classes of applications there are no downsides or the downsides are negligible"*. This, along with the good working relations we established over the past years with the HEP software stack developers, encouraged us to test scenarios where the transient complication of running and managing two independent network stacks is eventually over and we are 'back' to running just IPv6.

3.1 Aims of moving to IPv6-only and issues to be tackled

A dual-stack IPv6/IPv4 setup includes many components and services that need to be deployed twice *and kept in sync*: firewall rules and access lists, address assignment services, routing rules, network monitoring, diagnostics and intrusion detection infrastructures; to name just a few. Removing this duplication is highly desirable both for better maintainability and cost-saving. However, this *requires* a technical solution to access any site and service that may remain accessible via IPv4 *only*. This trailing remainder of sites will be hopefully shrinking but will likely exist for a very long time (see e.g. [8] and references therein). It is actually expected that after large blocks of public IPv4 addresses have started to be returned

to the market, their market price will decrease and offset any economic drive connected to the IPv4 address shortage, *relieving* the pressure for migration.

A standard solution for accessing IPv4-only services from an IPv6-only network is the deployment of DNS64 (RFC 6147 [4]) and stateful NAT64 (RFC 6146 [4]) services. DNS64 maps names that are resolved to an IPv4 address only (‘`A`’ records) to a synthesized IPv6 address composed of a default prefix[1] plus the four bytes of the IPv4 address. Traffic towards the DNS64 prefix is then routed to a NAT64 service attached to the public IPv4 network. This will in turn map the source ports, translate the IP packets to IPv4 and convert any return traffic back to IPv6. While NAT64 and DNS64 services have started to be incorporated by several technology developers (especially in the ‘carrier-grade’ NAT appliance market), just a few open-source reference implementations exist for UNIX, with JOOL [9] apparently being the only one under active development.

While IPv6-only environments can present a few operational challenges that can be worked around[2], one-step (6→4) address translation techniques do and will fail whenever IPv4 literal addresses are explicitly handled, stored or signaled by network applications or protocols. We feel that the time is ripe to start identifying this class of applications and protocols and direct an early effort at cleaning them of *any* usage or reference to IPv4 literals. While two-step (4→6→4) address translation techniques such as 464XLAT[3] are currently being added to network stacks especially at the request of telephony carriers that operate IPv6-only networks, we see this extra indirection as an (inefficient) workaround that just hides issues that should be fixed at the application level. Locating these issues as early as possible motivates the experimental operation of typical WLCG sites with IPv6-only networking, as described later (§3.3).

3.2 The case for an IPv6-only LHCOPN

The LHCOPN network implemented IPv6 quite early during its development. Since the start of IPv6 support in the EOS storage service, a large fraction of the data transfers carried by this network have changed Internet protocol, moving from IPv4 to IPv6. Since June 2019, LHCOPN carries more IPv6 packets than IPv4, as shown in figure 3.

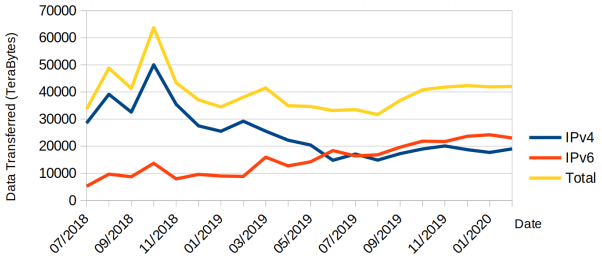

Figure 3. LHCOPN and LHCONE IPv4-IPv6 traffic distribution as seen on the CERN routers[10]

It could be envisaged that in the near future, once all the Tier-1s will have implemented dual-stack storage services, the LHCOPN could be turned into an IPv6-only network. There are some advantages that an IPv6-only LHCOPN could bring:

[1] Usually `64:ff9b::`.

[2] Some OS-specific network management tools, firewall appliances and network monitoring and diagnostic tools were found to be defective or immature, see RFC 6586 [4] for details.

[3] See RFC 6877 [4]. 464XLAT keeps a private IPv4 address assigned to devices connected to IPv6-only networks and performs an additional address translation at the device level.

- Increased security: LHCOPN links connect directly into Tier-1 data-centres, often bypassing border firewalls. Removing one protocol would decrease the attack surface;
- Simpler operations: maintaining one transmission protocol would simplify the operation of the networks and the resolution of problems;
- More addresses: IPv6 provides a larger number of addresses which can be used to avoid NAT in all situations.

The HEPiX IPv6 Working Group will encourage the LHCOPN community to move to IPv6-only as soon as possible.

3.3 Testing of IPv6-only

An IPv6-only WLCG production cluster, composed of an ARC-CE head node, two worker nodes and three (SQUID-based) web cache nodes has been in operation at Brunel University since March 2018. Given the value we place on early detection of IPv4-only code sections (especially non-address-translatable constructs such as the use of IPv4 literals in data structures and signaling; see above, §3.1), no transition techniques (e.g. NAT64/DNS64) were used for this infrastructure.

WLCG production jobs for three (out of four) major LHC experiments were routed to this IPv6-only cluster, with LHCb jobs running successfully since 2018, CMS jobs (submitted via a dedicated queue) running successfully in 2019, and ATLAS jobs, also handled by a special IPv6-only queue, requiring an in-depth, and still partly on-going, investigation of issues mainly within the Frontier [11] distributed database service.

This reality check does confirm that IPv4 is still *required* in part of the WLCG software base, with services failing in case IPv4 connectivity cannot be established. While the development time that has been spent in early troubleshooting and linting of these cases will definitely be rewarded as the transition progresses, we plan to complement this study with an assessment on how many of the residual issues aren't or cannot be covered by available address-translation techniques.

4 Conclusions and future plans

We have presented the status of the WLCG transition to the use of dual-stack IPv6/IPv4. The Tier-1 transition is nearly complete and more than 70% of the Tier-2 storage is available over IPv6. The transition will only be completed once we remove the complexity of dual-stack networking and the WLCG core infrastructure is IPv6-only.

Insufficiently tested or immature code and the requirement that IPv6-based tools and infrastructures perform at least equally well as their IPv4 counterparts have been the opposite, conflicting poles of every IPv6 deployment effort so far. This continues to be true in the process of completing the WLCG transition. We conclude that testing activities, and the consequent early detection of further application development needs, will keep the working group busy. We plan to increase the number of sites and stakeholders involved in testing IPv6-only scenarios. The aim is to stress-test existing networking software components that implement any needed transition protocol (especially NAT64 and DNS64, as their implementations under current maintenance are rare) and detect residual uses of IPv4 literals or IPv4-specific APIs in both applications and network protocols as early as possible.

Any use of IPv4 that cannot respond properly to a NAT64-mediated transaction[4] should be seen as an issue to be reported, tracked and addressed by developers: we plan to deal with

[4]More complex and inefficient address translation solutions such as the deployment of 'customer'-side address translation for RFC 6877 [4] 464XLAT or RFC 7597/9 [4] MAP-E/T should be seen as options of last resort, see §3.1 above.

these just as we did with the lack of IPv6 support or the incorrect address selection strategies we were able to identify so far.

Once we are confident that IPv6-only scenarios work well and that all issues found with the use of transition protocols have been fixed, we will propose a timetable for the deployment of an IPv6-only networking environment for WLCG.

References

[1] S. Campana et al, J. Phys. Conf. Ser. **513**, 062026 (2014)

[2] M. Babik et al, J. Phys. Conf. Ser. **214**, 08010 (2019)

[3] A. J. Peters et al, J. Phys. Conf. Ser. **664**, 042042 (2015)

[4] All Internet Engineering Task Force Requests For Comments (RFC) documents are available from URLs such as https://www.ietf.org/rfc/rfcNNNN.txt where NNNN is the RFC number, for example `https://www.ietf.org/rfc/rfc2460.txt`

[5] J. Bernier et al, J. Phys. Conf. Ser. **664**, 052018 (2015)

[6] A. A. Ayllon et al, J. Phys. Conf. Ser. **513**, 032081 (2014)

[7] E. Martelli et al, J. Phys. Conf. Ser. **664**, 052025 (2015)

[8] M. Nikkhah and R. Guérin, IEEE/ACM Transactions on Networking, **24(4)**, 2291 (2016)

[9] NIC Mexico, (2019) "Jool: SIIT and NAT64", https://www.jool.mx/en/about.html

[10] LHCOPN and LHCONE traffic flows on the CERN border routers, https://twiki.cern.ch/twiki/bin/view/LHCOPN/LHCOPNEv4v6Traffic

[11] Barry Blumenfeld et al, J. Phys. Conf. Ser. **396**, 052014 (2012)