Anycast
DNS64 + NAT64

Terry Froy
`<t.froy@qmul.ac.uk>`

School of Physics and Astronomy
Queen Mary University of London

IPv6 Council, London, December 2018

GridPP
UK Computing for Particle Physics

Queen Mary
University of London

Queen Mary University of London

- Research-focused higher education institution.

- Four main campuses in London.

- 21 academic departments.

- 25,000+ post- and undergraduate students.

- Approximately 4,500 staff.

  `<http://www.qmul.ac.uk>`

Terry Froy
(t.froy@qmul.ac.uk)
7th December 2018

Anycast DNS64 + NAT64
IPv6 Council, London, Dec. 2018

2 / 23

I suffer from Multiple Hat Syndrome...

- By weekday, I wear my 'GridPP hat' as the Cluster Systems Manager at QMUL.
  - The GridPP cluster at QMUL uses, by design, a non-anycast and non-resilient version of the implementation described here.

- By evening (and most weekends), I wear my 'commercial ISP hat' as Technical Director at Spilsby Internet Solutions.
  - The implementation used on my commercial network is anycast and resilient, therefore of more likely use to this audience.
  - It is this implementation which is described in this talk.

Terry Froy
(t.froy@qmul.ac.uk)
7th December 2018

Anycast DNS64 + NAT64
IPv6 Council, London, Dec. 2018

3 / 23

IPv6 @ Spilsby Internet Solutions

- Wireless ISP in deepest darkest Lincolnshire.

- Fixed-line ADSL/FTTC services across the UK.

- Started in 2003 with a 512Kbit/s circuit from Pipex and Pipex Legacy IP PA space.

- IPv6 from the very beginning (via 6in4 tunnels).

- Became RIPE LIR (uk.spilsby) in 2007:
  - Obtained ASN, /21 IPv4 and /29 IPv6

- Multi-homed shortly thereafter.

- Native IPv6 (no tunnels) since 2008.

Terry Froy
(t.froy@qmul.ac.uk)
7th December 2018

Anycast DNS64 + NAT64
IPv6 Council, London, Dec. 2018

4 / 23

Motivation for DNS64/NAT64

- Our "core" is not dual stack – per PoP:
  - Two IPv6 core routers
  - Two Legacy IP core routers
- Dual stack adds unnecessary complexity.
- No expectation that users need to deploy Legacy IP.
- There was a research paper in it ☺

Terry Froy
(t.froy@qmul.ac.uk)
7th December 2018

Anycast DNS64 + NAT64
IPv6 Council, London, Dec. 2018
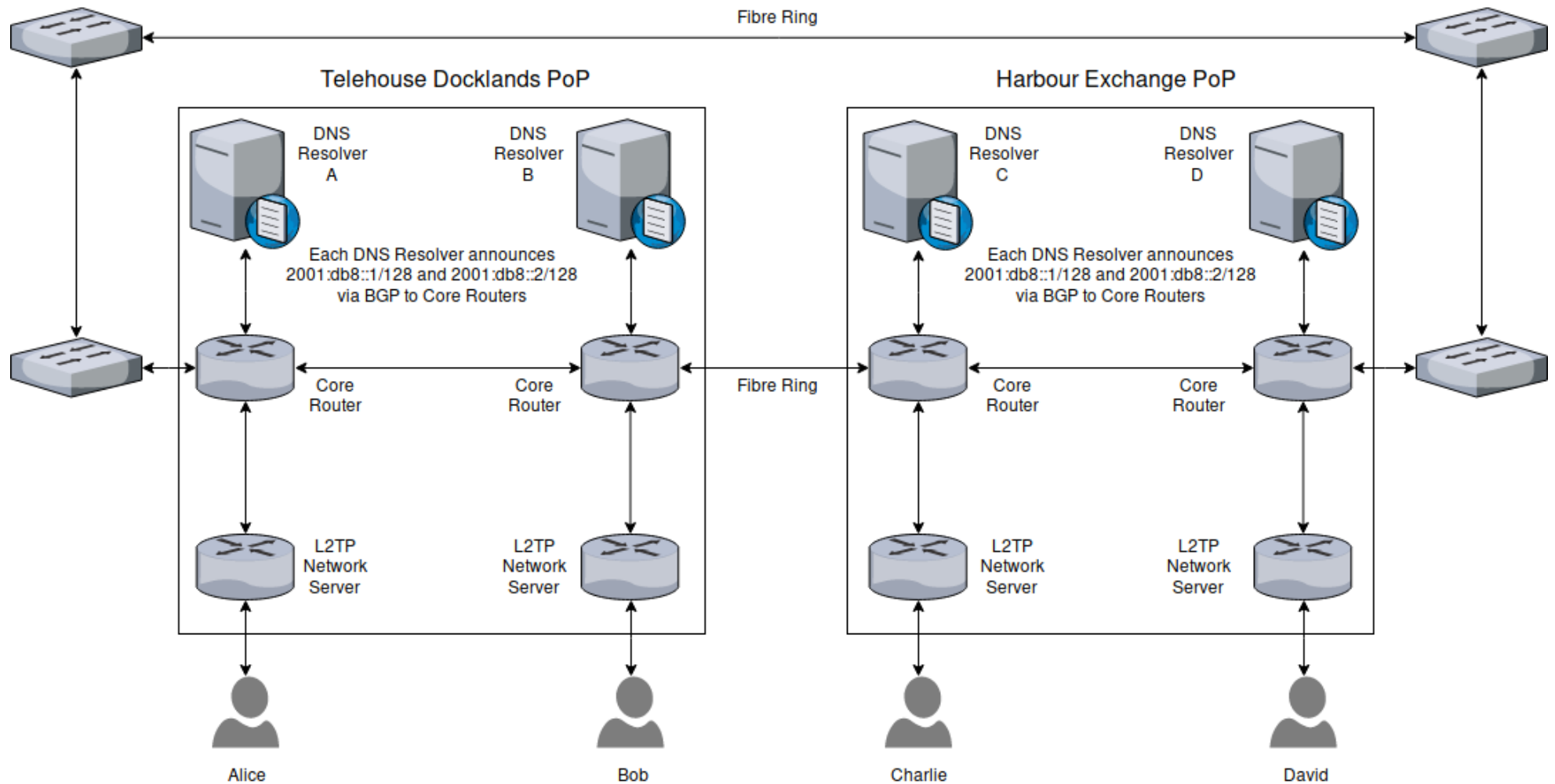
5 / 23

What is Anycast ?

- RFC4786 defines it thusly:

  "Anycast: the practice of making a particular Service Address available in multiple, discrete, autonomous locations, such that datagrams sent are routed to one of several available locations."

- Multiple instances.

- End users see one instance at a time.

- Failover is automatic for end users and handled at the routing protocol layer.

Terry Froy
(t.froy@qmul.ac.uk)
7th December 2018

Anycast DNS64 + NAT64
IPv6 Council, London, Dec. 2018

6 / 23

How does Anycast work ?

Terry Froy
(t.froy@qmul.ac.uk)
7th December 2018

Anycast DNS64 + NAT64
IPv6 Council, London, Dec. 2018

What is DNS64 ?

- RFC6147 defines it thusly:

  "DNS64 is a mechanism for synthesizing AAAA records from A records. DNS64 is used with an IPv6/IPv4 translator to enable client-server communication between an IPv6-only client and an IPv4-only server, without requiring any changes to either the IPv6 or the IPv4 node, for the class of applications that work through NATs."

- TLDR; A DNS resolver service that tells "lies".

Terry Froy
(t.froy@qmul.ac.uk)
7th December 2018

Anycast DNS64 + NAT64
IPv6 Council, London, Dec. 2018

8 / 23

When does DNS64 work ?

- When queried for an A record and the authoritative DNS servers supply one, it does not "lie".

- When queried for a AAAA record and the authoritative DNS servers supply one, it does not "lie".

- When queried for a AAAA record and the authoritative DNS servers do **<u>not</u>** supply one, this is when the DNS64 logic in a suitably-configured DNS resolver will kick in.

Terry Froy
(t.froy@qmul.ac.uk)
7th December 2018

Anycast DNS64 + NAT64
IPv6 Council, London, Dec. 2018

9 / 23

DNS64 [RFC6147] – How It Works(tm)

- The nameservers for qmul.ac.uk will only return an A record for www.qmul.ac.uk - they don't serve a AAAA record.

- Our IPv6-only hosts query DNS64-enabled resolvers (powered by the awesome open-source PowerDNS Recursor 4.x).

- A regular DNS resolver merely relays questions to servers and answers to clients (optionally caching the answers in the process).

- A DNS64-enabled resolver behaves slightly differently:

  - Client asks DNS64-enabled resolver for AAAA record for www.qmul.ac.uk.

  - DNS64-enabled resolver asks qmul.ac.uk nameserver which says 'No AAAA record for www.qmul.ac.uk'.

  - DNS64-enabled resolver asks for A record instead.

  - qmul.ac.uk nameserver responds with 161.23.16.98.

Terry Froy
(t.froy@qmul.ac.uk)
7th December 2018

Anycast DNS64 + NAT64
IPv6 Council, London, Dec. 2018

10 / 23

DNS64 [RFC6147] – Base 10 vs. Base 16

- Client asked for an AAAA record though.

- Our DNS64-enabled resolver 'fixes' this by converting the Legacy IP address 161.23.16.98 contained in the A record into hexadecimal:

DEC	161	23	16	98
HEX	A1	17	10	62

- The DNS64-enabled resolver 'synthesizes' a AAAA record by appending this 32-bit hexadecimal representation of the Legacy IP address to a /96 IPv6 prefix:

  - 64:ff9b::[/96] Well-Known NAT64 Prefix [RFC6052]

  - ::a117:1062 www.qmul.ac.uk

- The DNS64-enabled resolver returns an AAAA record of 64:ff9b::a117:1062 to the client.

Terry Froy
(t.froy@qmul.ac.uk)
7th December 2018

Anycast DNS64 + NAT64
IPv6 Council, London, Dec. 2018

11 / 23

GridPP
UK Computing for Particle Physics

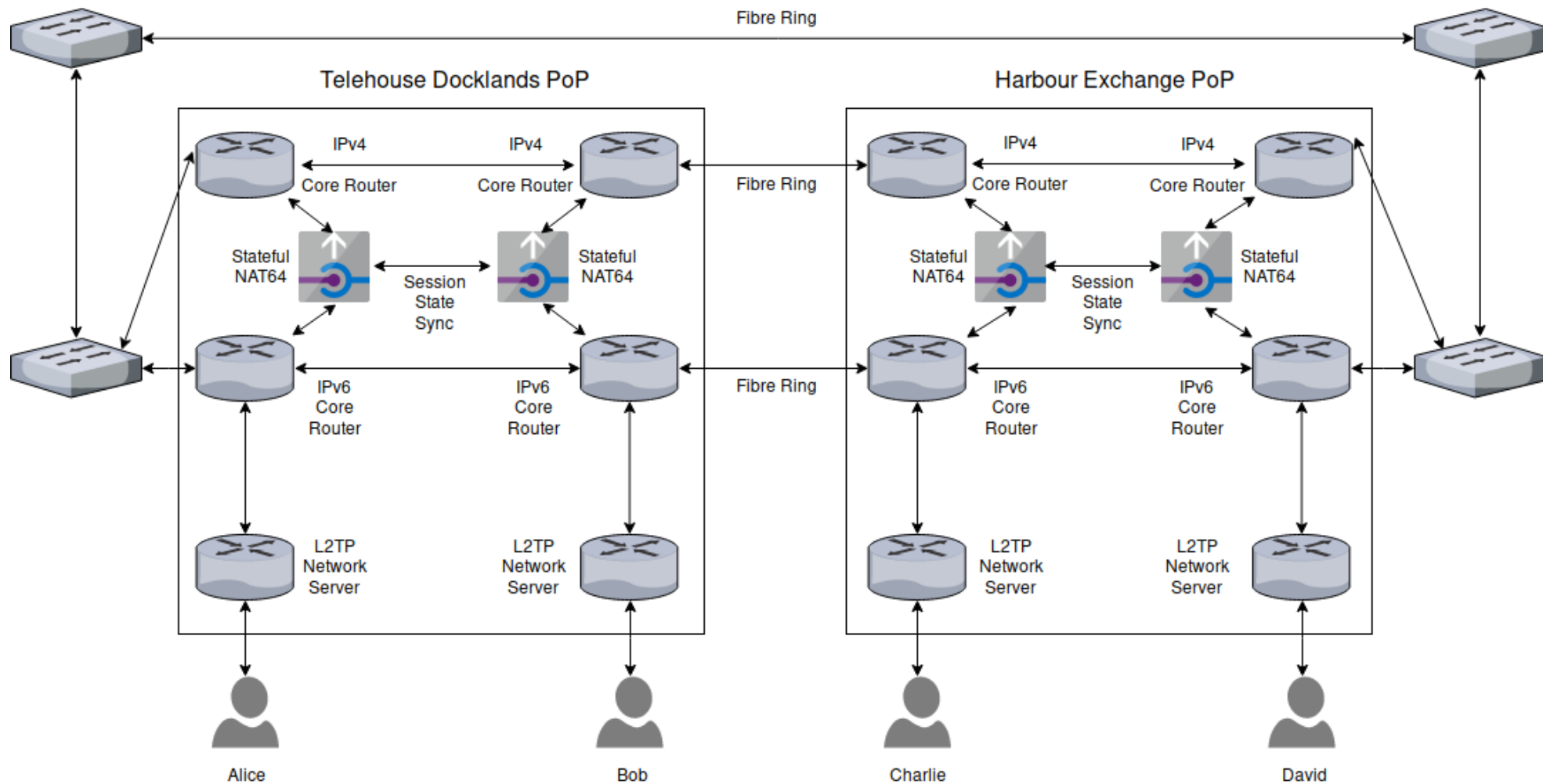Queen Mary
University of London

What is NAT64 ?

- RFC6146 defines it thusly:

- "Stateful NAT64 is a mechanism for translating IPv6 packets to IPv4 packets and vice versa. The translation is done by translating the packet headers according to the IP/ICMP Translation Algorithm defined in [RFC6145]."

- NAT64 can also be "stateless" as per RFC7915.

- "stateful" NAT64 can be considered somewhat similar to a masquerade NAT (or PAT).

- "stateless" NAT64 can be considered somewhat similar to a static NAT (one-to-one mapping).

Terry Froy
(t.froy@qmul.ac.uk)
7th December 2018

Anycast DNS64 + NAT64
IPv6 Council, London, Dec. 2018

12 / 23

NAT64 [RFC6146] – How It Works(tm)

- We run two "stateful" NAT64 translators per PoP.

- These run stock CentOS 7.6 and JooL 3.5.8.

- The Well-Known NAT64 Prefix 64:ff9b::/96 is advertised from each translator via BGP to its' IPv6 Core Router using ExaBGP.

- A PoP-specific Legacy IP prefix is advertised from each translator via BGP to its' Legacy IP Core Router using ExaBGP.

- A dedicated IPv6 link-local only VLAN between the translators distributing session state via multicast.

- What does this look like ?

Terry Froy
(t.froy@qmul.ac.uk)
7th December 2018

Anycast DNS64 + NAT64
IPv6 Council, London, Dec. 2018

13 / 23

NAT64 [RFC6146] – How It Looks(tm)

Terry Froy
(t.froy@qmul.ac.uk)
7th December 2018

Anycast DNS64 + NAT64
IPv6 Council, London, Dec. 2018

14 / 23

Does It Work ?

```
[tez@cybernoid ~]$ ip route ls
[tez@cybernoid ~]$ ping6 -c 4 www.qmul.ac.uk
PING www.qmul.ac.uk(host-161-23-16-98.qmul.ac.uk (64:ff9b::a117:1062)) 56 data bytes
64 bytes from host-161-23-16-98.qmul.ac.uk (64:ff9b::a117:1062): icmp_seq=1 ttl=234 time=11.1 ms
64 bytes from host-161-23-16-98.qmul.ac.uk (64:ff9b::a117:1062): icmp_seq=2 ttl=234 time=10.6 ms
64 bytes from host-161-23-16-98.qmul.ac.uk (64:ff9b::a117:1062): icmp_seq=3 ttl=234 time=9.93 ms
64 bytes from host-161-23-16-98.qmul.ac.uk (64:ff9b::a117:1062): icmp_seq=4 ttl=234 time=11.1 ms

--- www.qmul.ac.uk ping statistics ---
4 packets transmitted, 4 received, 0% packet loss, time 3003ms
rtt min/avg/max/mdev = 9.931/10.712/11.130/0.503 ms

[tez@cybernoid ~]$ wget http://www.qmul.ac.uk/
--2018-12-05 17:06:12--  http://www.qmul.ac.uk/
Resolving www.qmul.ac.uk (www.qmul.ac.uk)... 64:ff9b::a117:1062, 161.23.16.98
Connecting to www.qmul.ac.uk (www.qmul.ac.uk)|64:ff9b::a117:1062|:80... connected.
HTTP request sent, awaiting response... 302 Found
Location: https://www.qmul.ac.uk/ [following]
--2018-12-05 17:06:12--  https://www.qmul.ac.uk/
Connecting to www.qmul.ac.uk (www.qmul.ac.uk)|64:ff9b::a117:1062|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 46459 (45K) [text/html]
Saving to: 'index.html'

index.html              100%[=======================================>]  45.37K  --.-KB/s    in 0.02s

2018-12-05 17:06:12 (2.72 MB/s) - 'index.html' saved [46459/46459]
```

Does It Work ?

```
[tez@cybernoid ~]$ traceroute6 www.esc.qmul.ac.uk
traceroute to www.esc.qmul.ac.uk (64:ff9b::8a25:3324), 30 hops max, 80 byte packets
 1  router (2a01:568:302d::1)  1.596 ms  1.570 ms  1.539 ms
 2  vlan70.gradius.spilsby.net.uk (2a01:568:6000:70::2)  7.830 ms  7.819 ms  8.582 ms
 3  vlan70.thn-gw1.lon.spilsby.net.uk (2a01:568:6000:70::1)  8.826 ms  8.772 ms  8.745 ms
 4  thn-nat64-gw1.lon.spilsby.net.uk (2a01:568:6000:525::2)  9.235 ms  9.185 ms  9.148 ms
 5  vlan526.thn-gw1.lon.spilsby.net.uk (64:ff9b::4f62:20f5)  9.380 ms  9.361 ms  9.295 ms
 6  lon2.linx.thn-gw1.lon.acme.net.uk (64:ff9b::c342:ed3b)  10.663 ms  9.013 ms  8.915 ms
 7  linx-gw2.ja.net (64:ff9b::c342:ec0f)  8.908 ms  8.559 ms  8.477 ms
 8  ae23.londhx-sbr1.ja.net (64:ff9b::9261:23a5)  8.395 ms  8.548 ms  8.523 ms
 9  ae29.londpg-sbr2.ja.net (64:ff9b::9261:2102)  9.233 ms  9.611 ms  9.491 ms
10  ae26.londpg-ban1.ja.net (64:ff9b::9261:23ea)  9.468 ms  9.460 ms  9.443 ms
11  64:ff9b::9261:8fdb (64:ff9b::9261:8fdb)  10.524 ms  10.471 ms  10.923 ms
12  * * *
13  * * *
14  * * *
15  * * *
16  * * *
17  * * *
18  * * *
19  * * *
20  * * *
21  * * *
22  dewberry-ebr3.core-net.qmul.ac.uk (64:ff9b::8a25:3cc)  10.524 ms  10.526 ms  10.213 ms
23  hepweb2.ph.qmul.ac.uk (64:ff9b::8a25:3324)  10.452 ms  54.834 ms  10.360 ms
```

![GridPP — UK Computing for Particle Physics]

![Queen Mary University of London]

Terry Froy
(t.froy@qmul.ac.uk)
7th December 2018

Anycast DNS64 + NAT64
IPv6 Council, London, Dec. 2018

16 / 23

NAT64 [RFC6146] – Summary

- What does this provide us ?
  - Elimination of dual-stack.
  - Permits IPv6-only users to talk to Legacy IP resources.
  - Efficient use of our scarce Legacy IP resources.
  - Complexities of dual-stack are constrained to NAT64 translators.
- What doesn't work ?
  - Legacy IP literals [i.e. http://192.0.2.1/]
  - Plain ol' FTP [JooL lacks support for RFC6384]

Terry Froy
(t.froy@qmul.ac.uk)
7th December 2018

Anycast DNS64 + NAT64
IPv6 Council, London, Dec. 2018

17 / 23

Advice for those attempting this...

- Build each component in turn:
  - DNS64-capable resolvers (PowerDNS Recursor, Knot, Unbound, BIND)
  - First NAT64 instance
  - Second and subsequent NAT64 instance(s)

- Build **<u>one</u>** non-resilient NAT64 instance first and bring up ExaBGP.

- Once tested and working, build second instance then implement session sync.

- Bring up ExaBGP on second instance and test.

- Rinse and repeat as appropriate.

Terry Froy
(t.froy@qmul.ac.uk)
7th December 2018

Anycast DNS64 + NAT64
IPv6 Council, London, Dec. 2018

18 / 23

What Next ?

- Resilient 464XLAT [RFC6877] was put into production last month.

  – Dual-stack LAN and single-stack IPv6 on WAN.

- Currently in testing with staff and 'techie' customers.

- So far… It Just Works™

- Some development work required to automate static address/port mappings.

- Performance testing @ 100Gbit/s for the LOLz ☺

Terry Froy
(t.froy@qmul.ac.uk)
7th December 2018

Anycast DNS64 + NAT64
IPv6 Council, London, Dec. 2018

19 / 23

Relevant RFCs

- NAT64 [RFC6146]

- DNS64 [RFC6147]

- Discovery of the IPv6 Prefix Used for IPv6 Address Synthesis [RFC7050]

- Scenarios and Analysis for Introducing IPv6 into ISP Networks [RFC4029]

- An FTP Application Layer Gateway (ALG) for IPv6-to-IPv4 Translation [RFC6384]

- 464XLAT [RFC6877]

Terry Froy
(t.froy@qmul.ac.uk)
7th December 2018

Anycast DNS64 + NAT64
IPv6 Council, London, Dec. 2018

20 / 23

Special Acknowledgements

- PowerDNS.COM BV [http://www.powerdns.com/] for their awesome PowerDNS Recursor software [http://www.powerdns.com/recursor.html]

- ITESM [http://www.itesm.mx/] and NIC Mexico [http://www.nicmexico.mx/] for their work on JooL [http://www.jool.mx/]

Terry Froy
(t.froy@qmul.ac.uk)
7th December 2018

Anycast DNS64 + NAT64
IPv6 Council, London, Dec. 2018

21 / 23

Special Acknowledgements

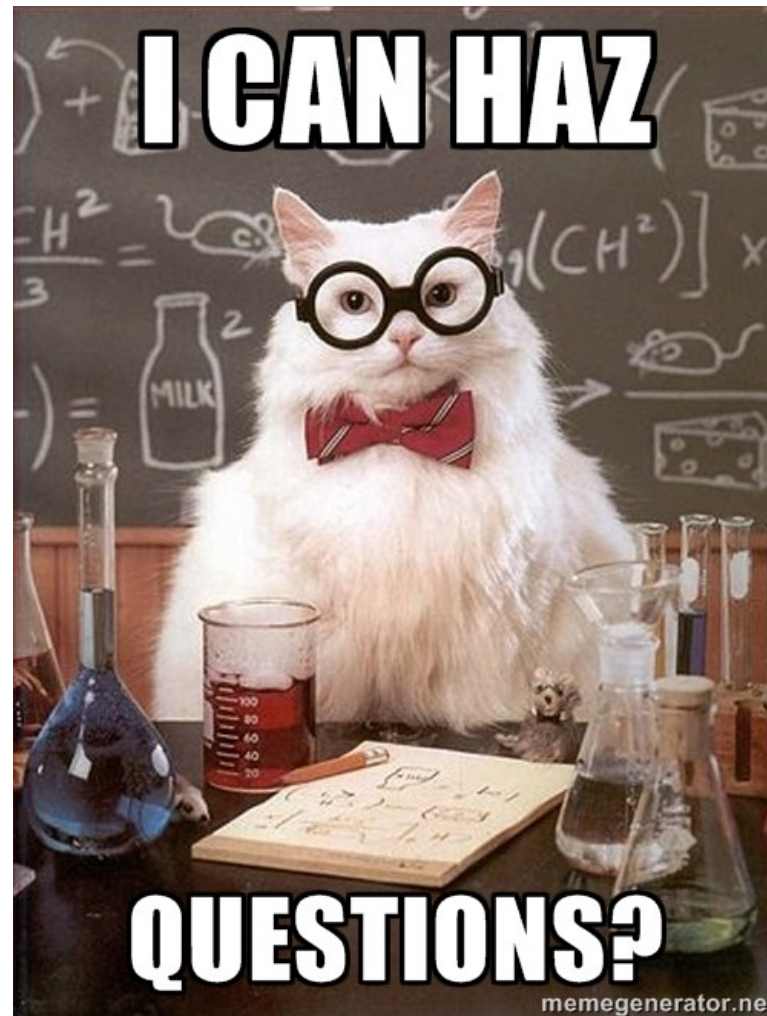- ExaBGP [https://github.com/Exa-Networks/exabgp/] for their absolutely wonderful 'BGP swiss-army knife'

Terry Froy
(t.froy@qmul.ac.uk)
7th December 2018

Anycast DNS64 + NAT64
IPv6 Council, London, Dec. 2018

22 / 23

Questions ?

Terry Froy
(t.froy@qmul.ac.uk)
7th December 2018

Anycast DNS64 + NAT64
IPv6 Council, London, Dec. 2018

23 / 23