IPv6 deployment at Imperial

...

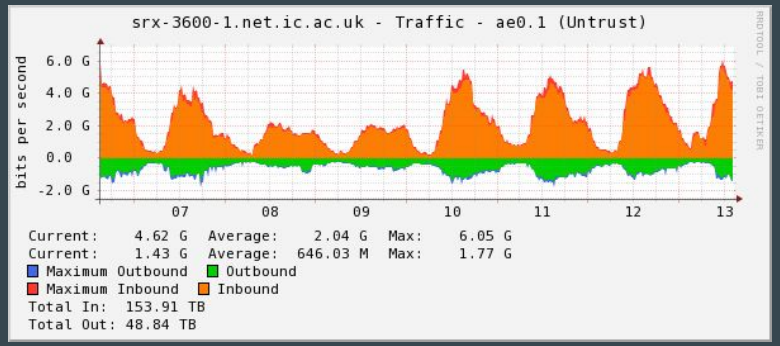
Phil Mayers <p.mayers@imperial.ac.uk>

About Imperial

- 14,700 students, 8,000 staff
- Focused on science, engineering, medicine and business
- 6 major campuses in London, also Silwood Park, and medical sites
- Various downstream customers (Museums, NHS trusts, Learned societies)
- Substantial e-Science work - IT infrastructure is important

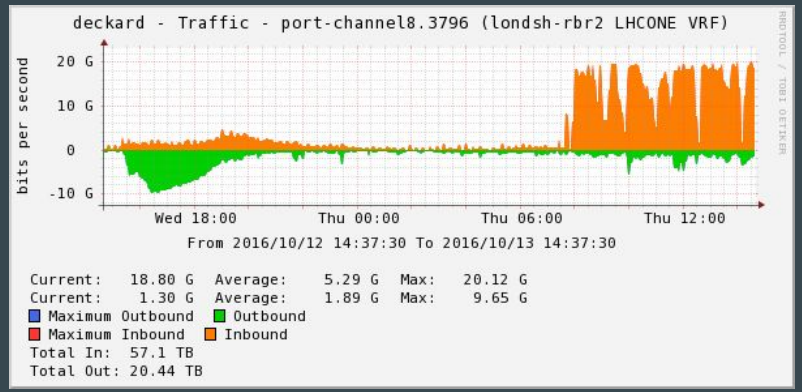
Campus network

- 60k devices on-net including PCs, WiFi/BYOD, SCADA, VoIP, etc.
- 18k simultaneous wifi users at peak
- Internet to campus throughput ~2Gbit/s average, ~6Gbit/s peak (Oct 2016)
  - This is just the “normal” traffic - web, email, etc. - excludes high-throughput users



Research traffic

- e-Science big hitters - High Energy Physics group
  - Increasing focus on IPv6 for this area
  - Rates of up to 40Gbit/s - could easily go higher



Story of our IPv6

- Long process - started experimenting in 2003
  - Initially using IP6-in-IP tunnel
  - Upstream provider was outsourced at the time - little appetite
- Mid 2008 - Deployments to select servers & test subnets
- April 2010 - Upstream native IPv6
- June 2010 - Mass deployment to clients started
- Early 2011 - Big push for World IPv6 day
  - Enabled the college website, email, DNS
- 2011-2012 - Servers & service deployment ongoing
- Sep 2013 - WiFi platform IPv6-enabled
- Spanned several generations of equipment & procurement

IPv6 day - 8 June 2011

- First big test - coordinated, worldwide enabling of IPv6
  - Google, Facebook, etc.
- Pushed hard - along with others in UK HE community - to participate
- College website was v6-enabled via v6-to-v4 NAT
- Deployment to client networks already ~75% complete
- On the day:
  - ~15% of traffic IPv6, ~5,500 machines doing IPv6 to the internet
  - A couple of minor issues relating to path MTU on the website hack
  - No issues raised by customers or externals
- Comprehensive success, in our view
  - Following years IPv6 launch was even easier - work already done

Current status

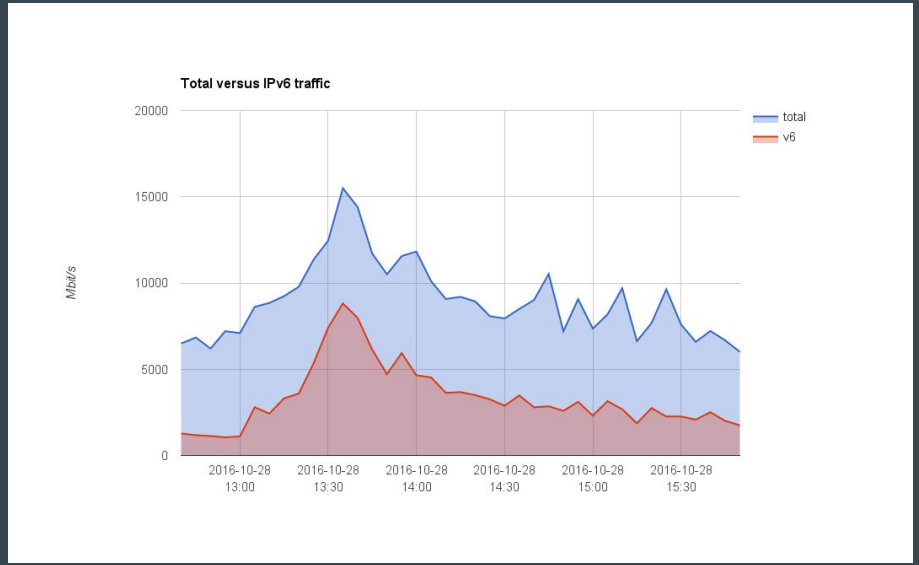
- Deployed - production ready, full SLA coverage equivalent to IPv4
- All new services in datacentre have dual-stack SLB by default
  - No reported issues
- Older COTS / ERP stack software not retrofitted
  - No “certification” from vendors - Grr!
- Various cloud services accessed predominantly over IPv6
  - Office 365
- IPv6 parity mandatory in all equipment procurement
  - Not yet filtered through to software & services procurement - ongoing

Results

- Average 20-40% of traffic by volume over IPv6
  - Depends on time, and counting methodology
  - 26 Oct 13:00 - 27 Oct 13:00, a typical 24h, no big eScience runs
  - IPv4 - 25TB, 27Gpkt, 208M flows
  - IPv6 - 6.5TB, 7.5Gpkt, 60M flows
- Large content providers like Google/Youtube, Facebook significant
  - Same period, major IPv6 sources - Google 3.7TB, Facebook 0.6TB
- As noted, Office 365 infrastructure primarily accessed over IPv6
  - Exchange, Sharepoint, Project, Lync

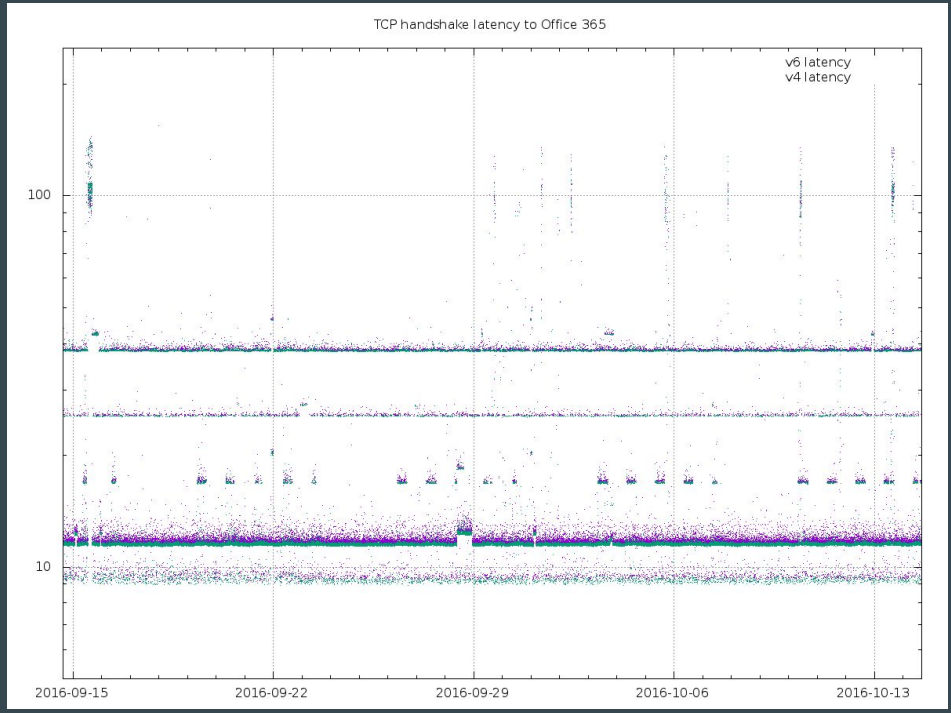
e-Science

- High Energy Physics dataflows - growing rapidly
- Fri Oct 28th @ 13:30 - 7Gbit/s IPv6
  - Not the biggest we have seen
  - Just the one I have a graph for...
- CERN & Brunel
- WLCG planning for IPv6-only capability in the near future
- Sheer quantity of compute and storage has exceeded IPv4 capacity



Cloud services & Latency

- Latency to Office 365
- Banding - different MS DCs
- Very little v4/v6 difference
- Similar results to other cloud providers



Why do this?

- Imperial is fortunate enough to have adequate IPv4 space - so why bother?
- We will run out of address space
  - Fortunate as we are, device count growth rate is astonishing
  - Projected WiFi BYOD growth could consume all remaining v4 IPs in ~5 years
  - IoT concerns - assuming the Internet hasn't been destroyed by hacked toasters of course
- Avoidance of surprise
  - Research tends to generate new requirements on short notice
  - Avoid being the blocker for your customers
  - Example: HEP community moving to IPv6 - IPv4 has run out for them!
- Don't believe in being last
- Done right, the cost is not high
  - Conversely, cost of having to do a rapid deployment could be significant

Choices we made

- Address configuration model - SLAAC not DHCP
  - DHCPv6 not widely available on clients at the time of deployment
  - SLAAC not overly problematic, no real impetus to move now
  - Manual well-known suffixes for servers
- Dual-stack - no current use-case for IPv6-only / NAT64 / 464xlat
  - But watching intently
- SLB - dual-stack VIP, single family backend
- Parity - same network equipment, paths, upstream
  - Final config - the rollout had various interim elements
- Whole network, not just select parts - student residences as well
  - Xbox One IPv6 gaming support - exemplar in the field, better UX for customers

SLAAC vs. DHCPv6

- No interest in the protocol drama any more
  - Please don’t ask me about it...
- SLAAC was available and worked for us at the time
  - DHCPv6 issues with L3VPN relay - software bug, not inherent
- Clients self-generate addresses, plural
  - These days, very likely >1 - privacy addressing
- Issues to consider
  - Need to track address usage for abuse, legal reasons
  - Addresses not “pretty” or “memorable”
  - Reverse DNS - not important for clients IMO
  - DNS-over-IPv6 - RFC 6106 sparsely supported
  - Address count growth

Address tracking

- Need to track (time, IP) -> machine mapping for abuse & legal reasons
- Lots of solutions
- Router neighbour tables
  - See for example https://nav.uninett.no/wiki/start
- DHCPv6 server logs
  - If using DHCPv6 of course
- Layer-2 switch FHS / radius accounting logs
  - Vendor-dependent
- Directly observe L2 ND/NS via span/mirror, or sampling e.g. sFlow
  - See for example https://github.com/jimdigriz/slaacer

Address tracking at Imperial

- Router neighbour tables
  - Bespoke system, pre-dates IPv6 rollout, ARP for IPv4
- Postgresql DB with inet datatype - transparently supported IPv6
- Consider tracking the IPv6 link-local addresses
  - Clients may talk to each other over link-local
  - You might find you have to trace abuse via LL
- Watch for address count growth
  - Temp/privacy addresses - many more than you'd expect in IPv4
  - Certain vendors cycle these addresses quite rapidly... not clear why
  - We see rare cases of extreme address counts - not operationally problematic, but odd...
  - Cheap & fast storage solve the database size issue for us

Address tracking - unusual clients

- Real client
- Lots of addresses
- >5000 in 24 hours
- No idea why...

¯_(ツ)_/¯

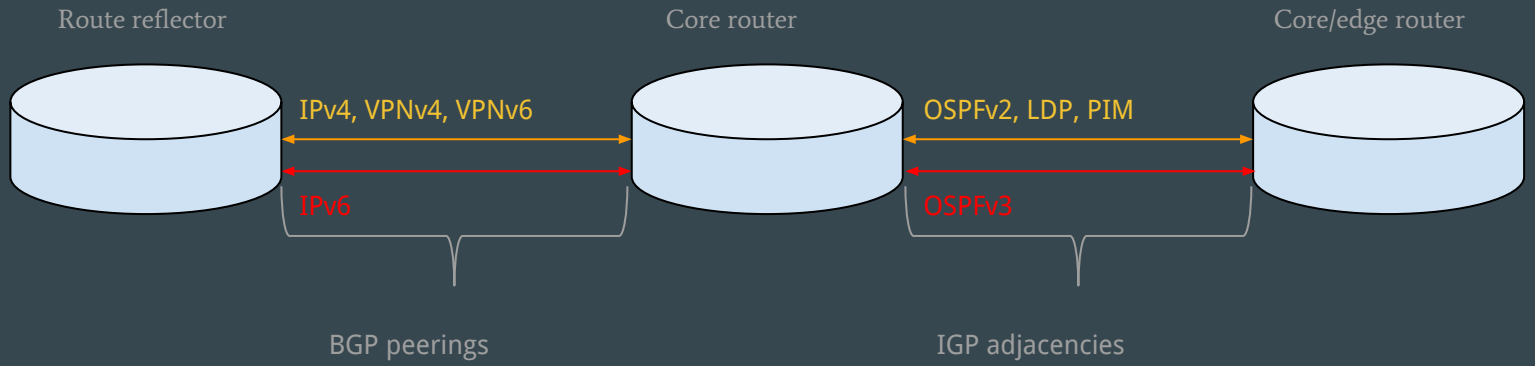
ip	mac	first	last
2001:630:12:107b:296f:16a9:48e2:7ebe	00:24:9b	2016-10-27T15:20:28+01:00	2016-10-27T15:20:28+01:00
2001:630:12:107b:6580:4eff:fa1d:ecb5	00:24:9b	2016-10-27T15:05:26+01:00	2016-10-27T15:20:28+01:00
2001:630:12:107b:3c20:bec5:60e1:5da1	00:24:9b	2016-10-27T15:10:27+01:00	2016-10-27T15:20:28+01:00
2001:630:12:107b:4c24:6c53:578f:425b	00:24:9b	2016-10-27T15:20:28+01:00	2016-10-27T15:20:28+01:00
2001:630:12:107b:9ceb:5ae:51e3:4b0a	00:24:9b	2016-10-27T15:10:27+01:00	2016-10-27T15:20:28+01:00
2001:630:12:107b:1963:8d8c:a1db:186c	00:24:9b	2016-10-27T15:15:27+01:00	2016-10-27T15:20:28+01:00
2001:630:12:107b:d40d:9ef8:810c:efa0	00:24:9b	2016-10-27T15:15:27+01:00	2016-10-27T15:20:28+01:00
2001:630:12:107b:89d2:17d5:ced5:b18f	00:24:9b	2016-10-27T15:10:27+01:00	2016-10-27T15:20:28+01:00
2001:630:12:107b:5162:1b47:3446:a331	00:24:9b	2016-10-27T15:20:28+01:00	2016-10-27T15:20:28+01:00
2001:630:12:107b:1965:d324:7980:7992	00:24:9b	2016-10-27T15:15:27+01:00	2016-10-27T15:20:28+01:00
2001:630:12:107b:586e:c2fe:e1da:b419	00:24:9b	2016-10-27T15:10:27+01:00	2016-10-27T15:20:28+01:00
2001:630:12:107b:4854:a088:be12:9d98	00:24:9b	2016-10-27T15:05:26+01:00	2016-10-27T15:20:28+01:00
2001:630:12:107b:34b1:ad1d:2503:d3e1	00:24:9b	2016-10-27T15:15:27+01:00	2016-10-27T15:20:28+01:00
2001:630:12:107b:a564:3b68:a5c4:eb89	00:24:9b	2016-10-27T15:15:27+01:00	2016-10-27T15:20:28+01:00
2001:630:12:107b:4973:2f5b:fb47:f62e	00:24:9b	2016-10-27T15:10:27+01:00	2016-10-27T15:20:28+01:00
2001:630:12:107b:96a:3686:88e4:3e4d	00:24:9b	2016-10-27T15:15:27+01:00	2016-10-27T15:20:28+01:00
2001:630:12:107b:79f3:8e08:cc36:7d8a	00:24:9b	2016-10-27T15:15:27+01:00	2016-10-27T15:20:28+01:00
2001:630:12:107b:8010:8eb0:b51f:15f8	00:24:9b	2016-10-27T15:05:26+01:00	2016-10-27T15:20:28+01:00
2001:630:12:107b:3007:8a:beef:ce19	00:24:9b	2016-10-27T15:15:27+01:00	2016-10-27T15:20:28+01:00
2001:630:12:107b:81b5:686e:7b0b:f1a0	00:24:9b	2016-10-27T15:20:28+01:00	2016-10-27T15:20:28+01:00
2001:630:12:107b:493e:83c0:848c:c803	00:24:9b	2016-10-27T15:10:27+01:00	2016-10-27T15:20:28+01:00

Dual-stack rollout

- Currently no driver for IPv6-only subnets
  - On 3-year timescale we expect dual-stack to be pervasive
- Could alleviate pressure in some upcoming areas
  - Container-based services, container-based desktops (app virtualisation)
  - SCADA? - problematic, barely does IPv4 properly
  - IoT - potentially, but so far appalling software quality, dreading poor IPv6 support
- IPv6-only WiFi would be a big help, as the client count is very high
  - Needs to be very reliable though - perception is it's not quite there yet on BYOD
  - Perhaps that's untrue? Comments welcome!
- Core routing - next slide

Core routing

- Separate OSPFv2 & v3, BGP, LDP for MPLS L3VPN
  - /112 for router p2p if you really want to know; aesthetically pleasing!
- Only notable element - MPLS L3VPN used for segmentation
  - Bulk of edge networks are therefore 6vPE w/ IPv4 provider control-plane
  - Switch to native IPv6 via “normal” IGP/BGP on leaving firewalls



SLB

- Previous SLB vendor supported IPv6 - used for v6 launch
  - No real issues, product range now EoL
- Current vendor supports IPv6 very well
- SLB services are dual-stack at client-facing VIP
  - Random sample ~9k IPv6 connections, ~14k IPv4 connections
- Backends are mostly v4-only
  - SLB does v6-v4 translation, adds X-Forwarded-For: HTTP header
  - Choice of v4 backend based on lowest upheaval during transition
- Going forward, some backends are v6-only backend
  - Mail relays, IPAM - controlled by my team, easy to do
  - See also https://fud.no/talks/ - v4 as a “only on SLB VIP” model

Whole Network - Wireless & Residences

- Wireless - one of the last client systems to deploy IPv6
  - Start of academic year 2013
- Wireless vendor had no RA guard
  - Proved especially problematic on WiFi - Internet Connection Sharing to wired
  - Clients would be trying broken IPv6
  - Will discuss later
- Had to use DNS AAAA-blacklist
- Solved in later release - works now without issues
- Residences - no real issues despite prevalence of unmanaged devices
  - RA guard and DHCP snooping a MUST however!
  - Did allow Teredo back in for Xbox One p2p fallback networking
  - Closely watching IETF stuff for appropriate IPv6 residence security posture

Issues faced

- Important to note: these were not huge problems for us
  - Context only, do not be discouraged - be aware
- Layer-2 first-hop security
  - Rogue router advertisements and DHCP servers
  - Usually accidental via Internet Connection Sharing
- Broken external websites
  - IPv6 in DNS but not responding - browser-based happy eyeballs solved this
  - Answering over IPv6 but with bad content - sadly, still seeing this
- IPv6 do-not-serve blacklists at content providers
- Address counts & table sizes
- Bespoke systems

Layer-2 first-hop security

- Internet Connection Sharing
  - a.k.a Infernal Connection Shenanigans - least helpful “feature” ever?
  - Rogue RA/DHCPv6 from connectivity via tunnels, or other wired/wireless interfaces
- Native IPv6 ameliorated this
  - All hail RFC 3484/6724 address selection rules
  - Also, set native router-preference to “high” just in case
- If you lack native IPv6 and RA/DHCPv6 guard, this can be a problem
- Various platform limits, but finally got “stateless” DHCPv6 & RA guard
  - ACL dropping ICMPv6 type 134 - hardware ACL TCAM hassles, overcame these
  - DHCPv6 dropped by UDPv6 port match - simpler
- Mandate relevant sections of RIPE-554 in procurement

Broken websites

- One of the few areas which generate ongoing support load
  - Very infrequent, but non-zero
- External websites which are reachable over IPv6 but serving invalid content
- .eu I am looking at you
- .gov you can stop smiling as well
- Customers perceive your network is broken
  - “Works from my phone / home ADSL / other places”
- Increased number of IPv6 access networks will hopefully stop this
- Very, very rarely, we “fix” this using DNS RPZ to strip the AAAA
  - Dislike doing this intensely, hides the problem, misaligned incentives

IPv6 blacklists

- Various providers - notably Google and Facebook
- Detect clients with broken IPv6 with various black magic tools
  - Backtrack to the client DNS server
  - Stop serving AAAA to that DNS server
- End sites just see a drop in IPv6 usage
- No real feedback for end sites on triggering events
  - Understandable - content providers would incur a lot of work and have to expose potentially sensitive logging information
- Not sure these are still in use? Issue largely historic for Imperial

Blacklist triggers

- DNS server handles disparate clients
  - In our case, main Imperial network as well as downstream unmanaged customers
  - Solution: split them onto separate resolver query sources
- Clients in sections of the network with spotty IPv6
  - Such as the aforementioned wireless issues
  - Solution: deploy the IPv6!
  - Alternatively, AAAA-blacklist - very short term, hides not solves the problem
- Lack of parity
  - Example: excessive loss, latency on IPv6 compared to IPv4
  - Solution: aim for parity
- Combination of first two solved our issues

Table sizes

- IPv6 addresses are 4x the size of an IPv4 address
  - Devices may have comparatively limited IPv6 FIB
  - And/or - FIB may be statically partitioned with low IPv6 capacity
  - Overrun can require a reboot to fix
- Consider the number of adjacent hosts
- Check with your vendor for scaling and dynamic/static limits
  - Be very careful of misleading claims about concurrent v4/v6 routing and adjacency sizes
  - Does a host consume a route? Does a v6 host/route consume 2 or 4 v4 host/route slots?
- Cause of one outage at our site - FIB exception on older platform
  - Triggered by wireless network - very busy, lots of connected addresses
- Suggest budgeting for at least 3x number of connected clients as IPv6 addresses

Neighbour churn

- It will be busier than IPv4
- Watch control-plane load
  - Default ND refresh timers may be inappropriate
- ~18k associated WiFi clients leads to:

```
wlan-rt1#sh ipv6 neighbors vrf [redacted] statistics
IPv6 ND Statistics
 Entries 27015, High-water 30369, Gleaned 30329332, Scavenged 44275425, Static 0
 Entry States
   INCMP 144  REACH 11668  STALE 13266  GLEAN 1681  DELAY 171  PROBE 85
 Resolutions
   Requested 69387649, timeouts 140350557, resolved 28376295, failed 40390005
   In-progress 144, High-water 274, Throttled 0, Data discards 26297438
 NUD
   Requested 194464215, timeouts 47865620, resolved 179439572, failed 15024536
   in-progress 256, high-water 256, throttled 32198849, current queue 1091, queue high-water 5736
wlan-rt1#
```

- ...and the multiplier will likely go up over time

Bespoke systems

```
create table log (ip varchar(15) ...);

drop will_to_live;
```

- Try hard not to have these problems ;o)
- Fortunate at Imperial - most systems using postgres/inet, transparent to IPv6
- Occasional tweaks to client-side validation e.g. webapp javascript
- One example: bespoke IPAM system, feeds DNS, DHCP, firewall
  - ~300 lines of code, ~1 hour to IPv6-enable
  - Almost entirely form validation

Support costs

- Very low - modern IPv6 stacks and browsers with happy eyeballs are well behaved
- Very rare to investigate an IPv6 issue - about the same as IPv4 once mature
  - 26 incidents to our Service Desk since 1 Jan 2016 mentioning IPv6
  - Vast majority unrelated on postmortem analysis
  - Couple of incidents of IPv4 being broken and only connectivity over IPv6!
- No substantial engineering cost to maintaining IPv6 in our experience
  - Marginal cost ensuring parity in procurement, but that's infrequent activity
- Educate front-line staff that "disabling IPv6" is not a solution
  - Rare problems should be known and solved, not hidden

Procurement

- Hopefully you have been mandating and testing for IPv6 parity for some time
- If not, start now
- RIPE-554 an excellent start, but not a panacea
  - You will have to test, and to test you'll need knowledge
- Signal - firmly - to vendors that you won't accept 2nd class IPv6
  - Without those signals, the market may backslide
- Ensure you have a working rollout or testbed, to compare against

What now?

- If you have already deployed IPv6 - such as Imperial:
  - Identify areas where coverage isn't great - old software, equipment
  - Correlate with refresh cycles
  - Identify route forward - deprecate, replace, upgrade/fix, ignore
  - Continue to grow coverage
- If you have not deployed yet:
  - Establish a testbed ASAP to gain experience
  - Identify critical path items - upstream, core, firewall
  - Deploy incrementally, possibly in concert with hardware/software refresh cycles
  - Set achievable goals - don't get bogged doing too much
- If you're not intending to deploy:
  - I'm out of advice for you... IPv6 is not going away. Please reconsider!

Thanks!

Feel free to contact me with any questions