Imperial College
London

IPv6 Only at Imperial

David Stockdale
ICT Networks Group
david@imperial.ac.uk

Imperial College
London

Facts and figures

- Over 65,000 unique hosts on wired network
- Over 60,000 unique hosts on wireless network
- Over 20,000 concurrent wireless clients at peak time
- 2x100G to Janet
- Most hosts within VRFs (MPLS L3VPNs)
- Firewalls between VRFs
- No NAT(44)

Our current position

- ~35% of our Internet traffic IPv6 (nearly 50% on BYOD)
- Dual stack on production, guest & BYOD (including wireless)
- AAAAs on most load-balanced services
- Other services enabled:
  - Home directories (>95% IPv6!)
  - 10PB research data storage (~~IPv6 only~~)
  - Mail, DNS, HEP systems
- SLAAC rather than DHCPv6
- Feature parity mandated in tenders

HPC refresh

- Multi-year programme to replace HPC estate
- Year 1: 7 racks, 30 servers in each
- By the end: 40 racks, 1PFLOPS
- Existing servers 1/10G Ethernet plus Infiniband
- New servers 100G Ethernet with RoCE
- Speaks to IPv6 enabled research data storage

HPC refresh

- An opportunity to go IPv6 only!
- How hard can it be!?
- Pesky IPv4 only PBS for starters
- 2x spine switches and leaf per rack
- EBGP, ASN per switch
- /64 IPv6 and /24 IPv4 per leaf
- MP-BGP between switches, IPv6 sessions only
- IPv6 only… to rest of College network
- /32 IPv4 route on servers via local gateway for PBS

HPC refresh

- So far, so good!
- How do we boot them?
- ...DHCPv6 ...and UEFI
- SLAAC, RDNSS
- Plan A:
  - Stateless DHCPv6 server on switch returning PXE options
  - Server: “I only support DUID-UUID”
  - Switch: “Unsupported DUID type 4”
  - Us: :-(

HPC refresh

- Plan B:
  - Stateless DHCPv6 relay to Kea returning PXE options
  - Server: “Send me a Bootfile URL in option 59”
  - Kea: “Sure thing”
  - Server: “And reflect Vendor Class option 16 back at me”
  - Kea: “What?”

HPC refresh

- Plan C:
  - Stateless DHCPv6 relay to ISC returning PXE options
  - Server: “I see your RA. SLAAC address, done”
  - Server: “Other Configuration Flag set? Will do”
  - Server: “Managed Address Configuration unset…”
  - Server: “...Request! Request! Request!”

HPC refresh

- Plan D:
  - Stateful* DHCPv6 relay to ISC
  - Success!

- Sort of… hello iPXE
- iPXE: “Information-Request!”
- Switch: *drops packet*
- iPXE: “RA said I needed that reply. I’ll wait… forever...”

HPC refresh

- iPXE recompiled to not send Information-Requests
- iPXE: “I don’t care, the NIC wasn’t initialised anyway”

- iPXE recompiled with bodgetastic initial sleep
- Us, 2 days in: “OMG it’s actually booting!!!”

HPC refresh

- Machines booted!
- How to talk to rest of the world?
- ...NAT64/DNS64
- Presenting software exhibit A
- Exhibit A: “Where’s my licence server?”
- DNS: “Here A or here AAAA”
- Exhibit A: “I’ll go with the A then”
- One AAAA only later… Fixed!

Imperial College London

HPC refresh

- Presenting software exhibit B
- Exhibit B: “Where’s my licence server?”
- DNS: “Here A or here AAAA”
- Exhibit B: “What’s a AAAA? What’s IPv6?”
- yum install clatd
- Exhibit B: *springs into life*
- Everyone: “Let’s party like it’s 1980 again!”

HPC refresh

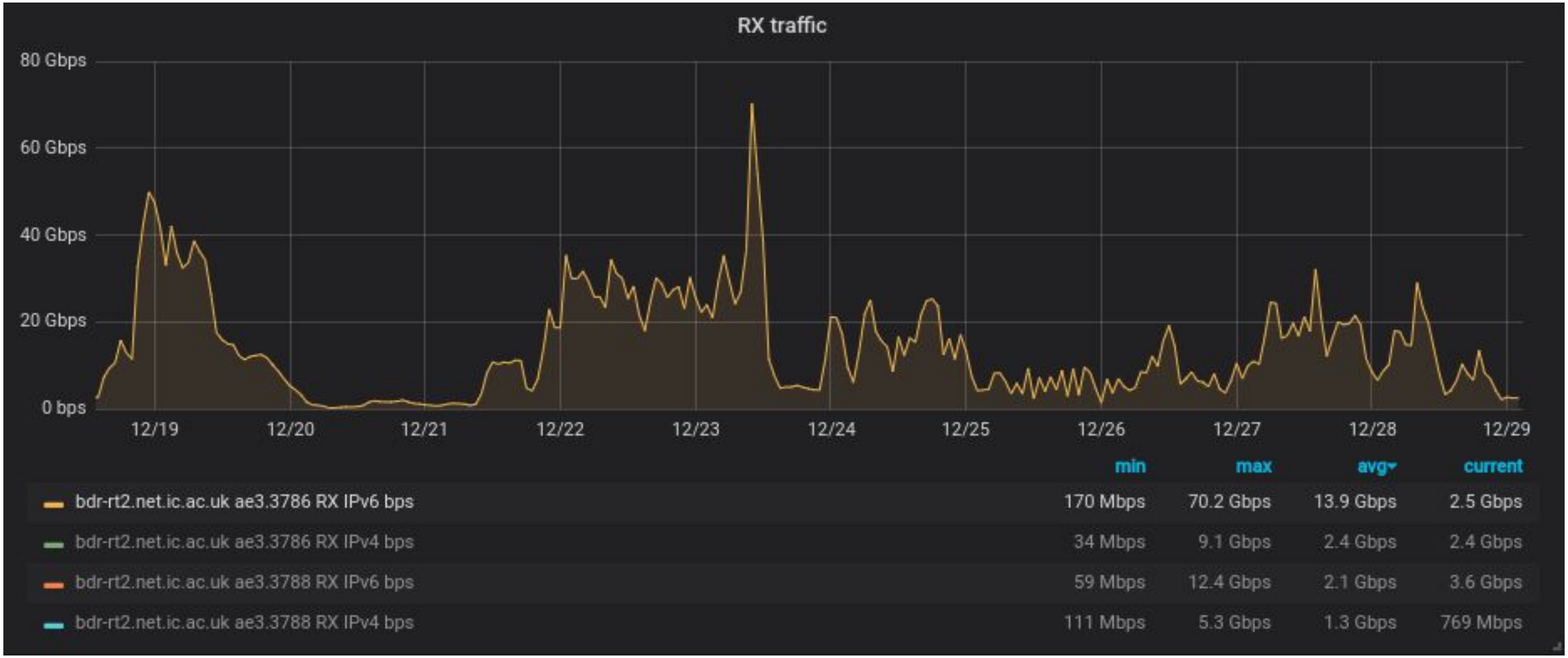
- We have a fully functional system
- One last thing… that RoCE thing

- All 7 racks now in full production use
- Looking at options for PBS

What next?

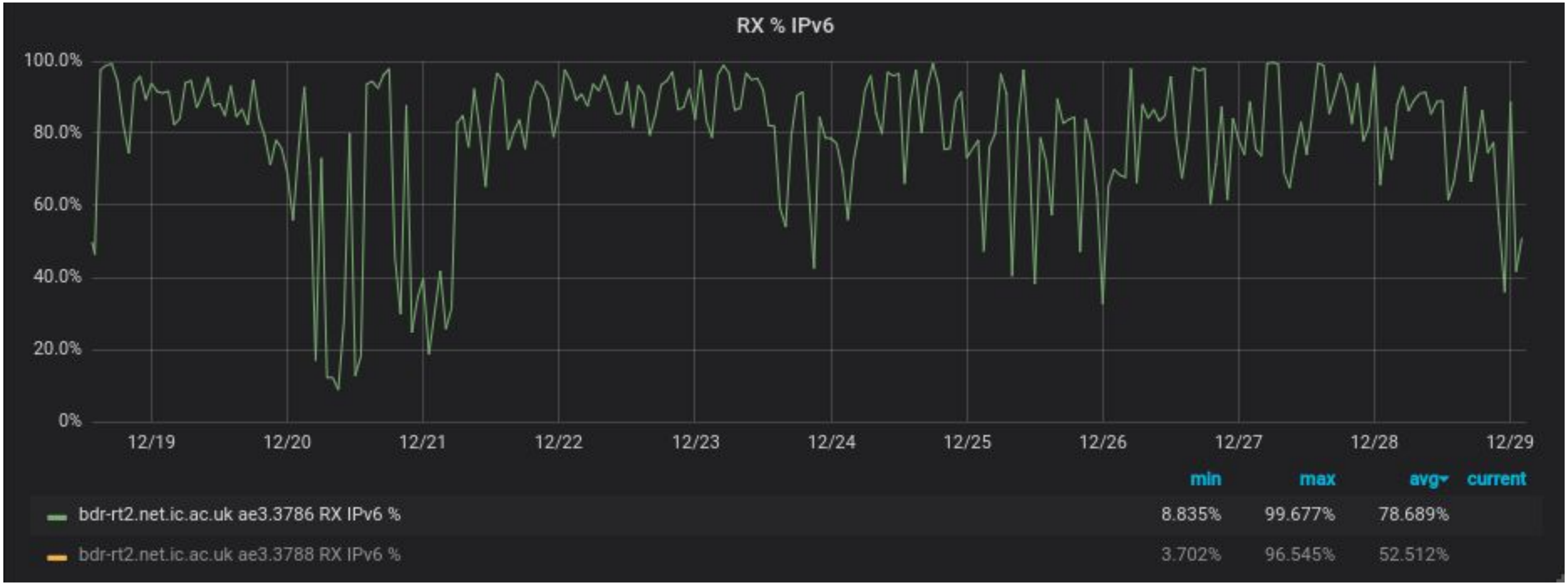
- IPv6 enable remaining services
- NAT64/DNS64 trials on wireless and wired
- IPv6 only internal services
- External services fronted by load-balancers
- DHCPv6 in DC, PXE
- Retire IPv4
- Free up IPv4 address space - $$$!

HEP Internet traffic



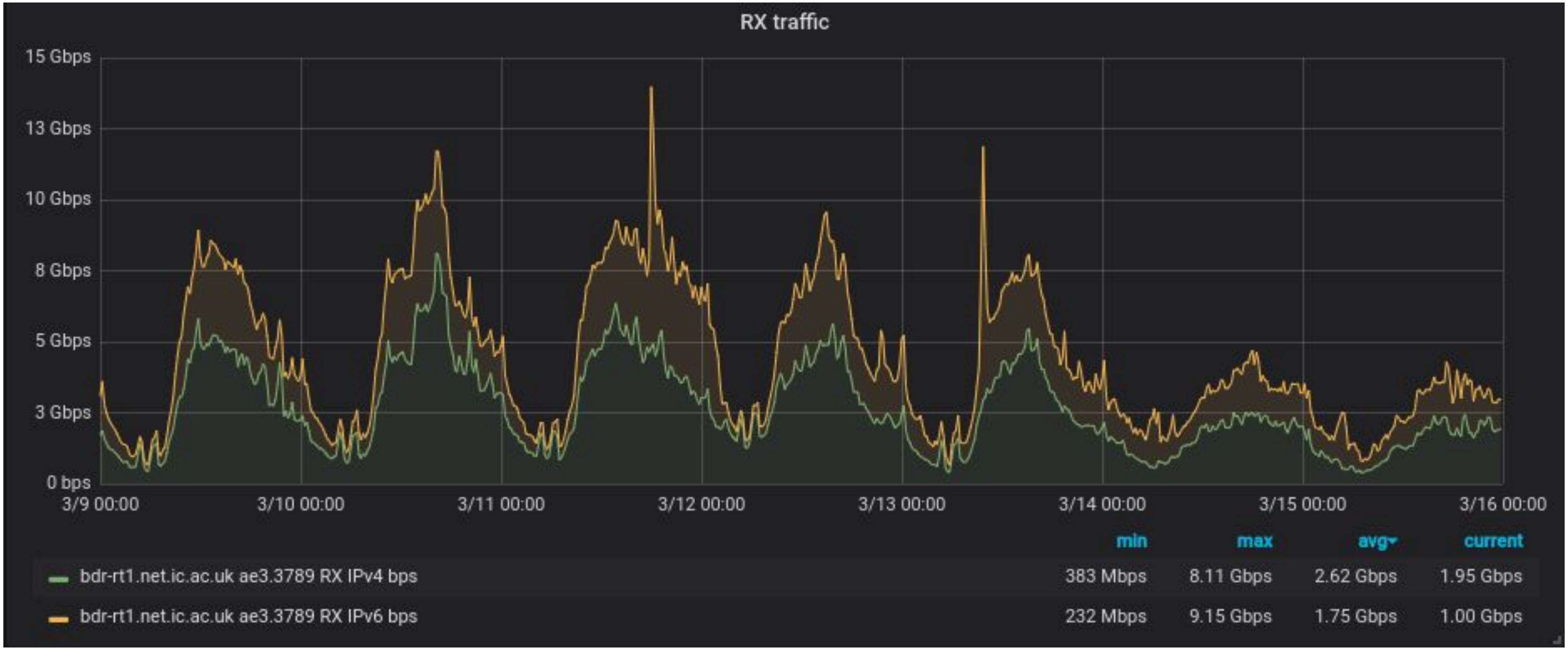
	min	max	avg▾	current
🟨 bdr-rt2.net.ic.ac.uk ae3.3786 RX IPv6 bps	170 Mbps	70.2 Gbps	13.9 Gbps	2.5 Gbps
🟩 bdr-rt2.net.ic.ac.uk ae3.3786 RX IPv4 bps	34 Mbps	9.1 Gbps	2.4 Gbps	2.4 Gbps
🟧 bdr-rt2.net.ic.ac.uk ae3.3788 RX IPv6 bps	59 Mbps	12.4 Gbps	2.1 Gbps	3.6 Gbps
🟦 bdr-rt2.net.ic.ac.uk ae3.3788 RX IPv4 bps	111 Mbps	5.3 Gbps	1.3 Gbps	769 Mbps

HEP Internet traffic



	min	max	avg▾	current
🟩 bdr-rt2.net.ic.ac.uk ae3.3786 RX IPv6 %	8.835%	99.677%	78.689%	
🟨 bdr-rt2.net.ic.ac.uk ae3.3788 RX IPv6 %	3.702%	96.545%	52.512%	

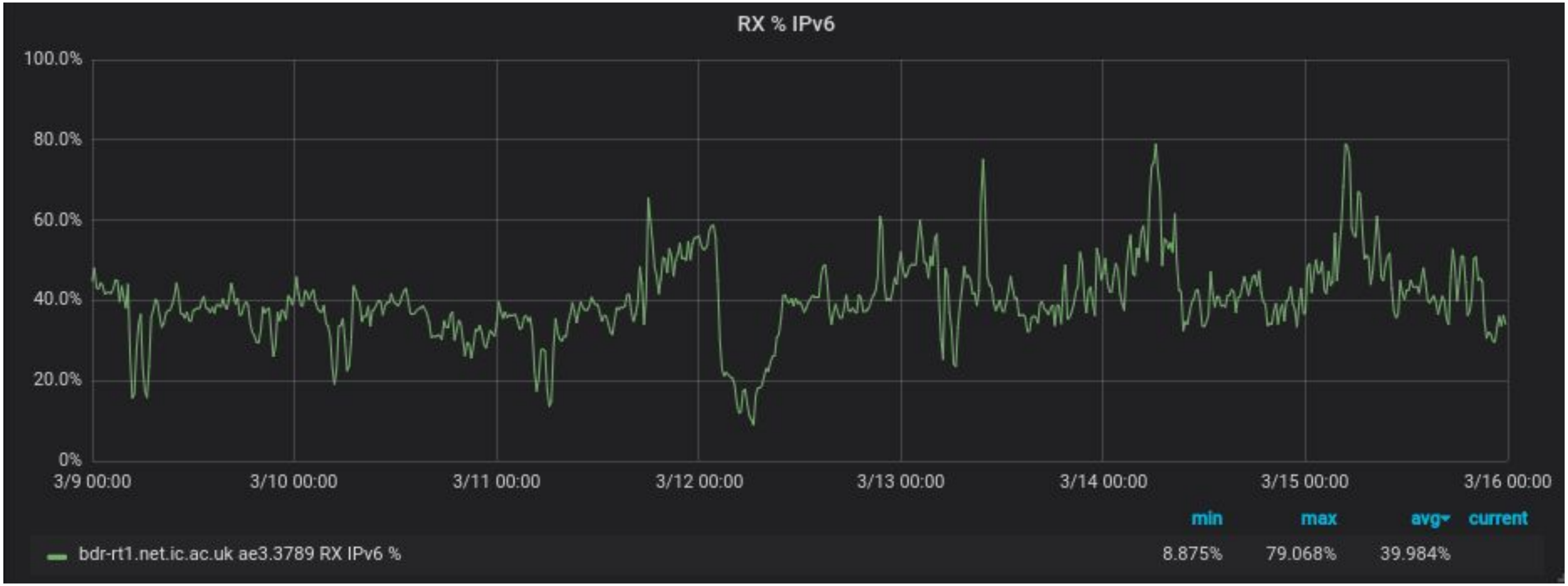
College Internet traffic



	min	max	avg▾	current
🟩 bdr-rt1.net.ic.ac.uk ae3.3789 RX IPv4 bps	383 Mbps	8.11 Gbps	2.62 Gbps	1.95 Gbps
🟨 bdr-rt1.net.ic.ac.uk ae3.3789 RX IPv6 bps	232 Mbps	9.15 Gbps	1.75 Gbps	1.00 Gbps

Imperial College
London

College Internet traffic


RX % IPv6
3/9 00:00
3/10 00:00
3/11 00:00
3/12 00:00
3/13 00:00
3/14 00:00
3/15 00:00
3/16 00:00
min
max
avg
current
bdr-rt1.net.ic.ac.uk ae3.3789 RX IPv6 %