Dropping in 80Gbits (sort of) of Stateful Firewalling with OpenBSD

(PF, OpenOSPF)

UKNOF 37, Manchester

Caveats

I am not pushing 80Gbits yet *(sorry if you were expecting Netflix levels of awesome)*

See: Sort of

Who am I?

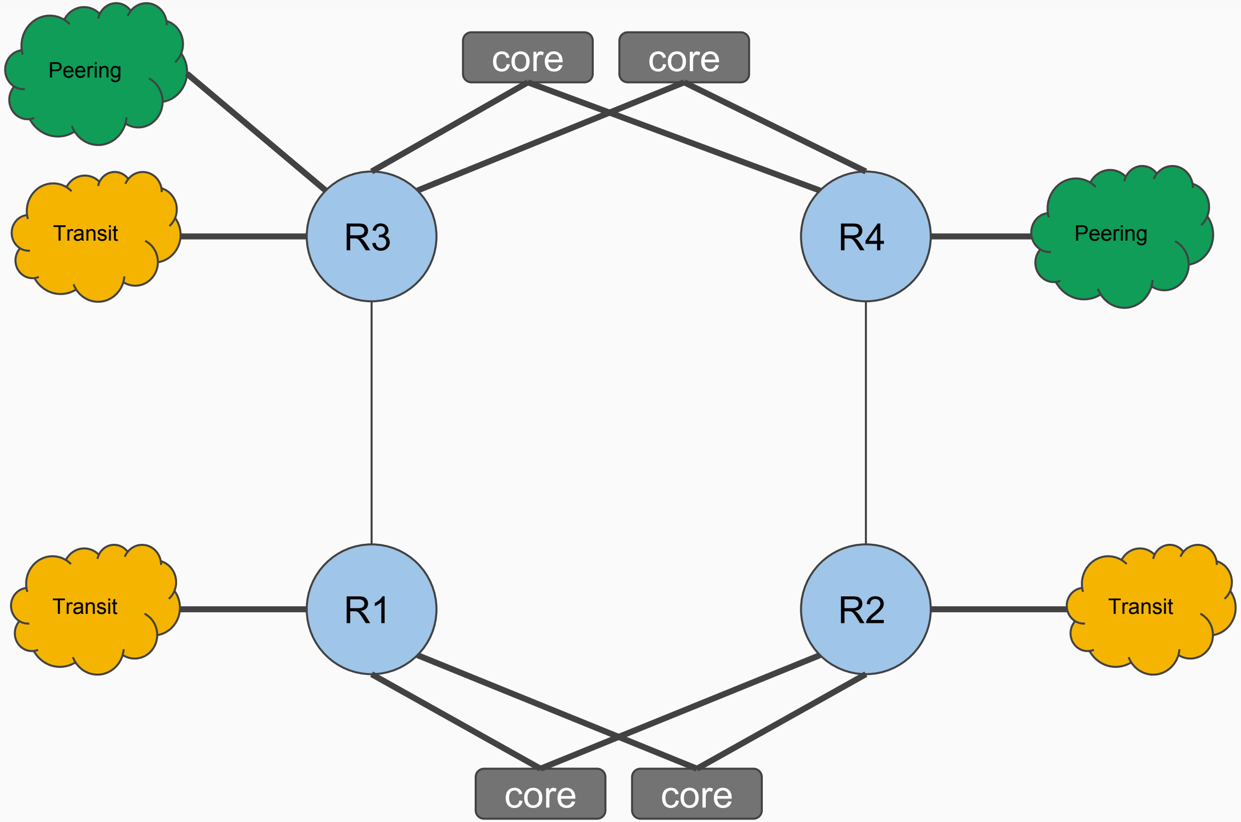
Gareth Llewellyn

@NetworkString | gareth@networksaremadeofstring.co.uk

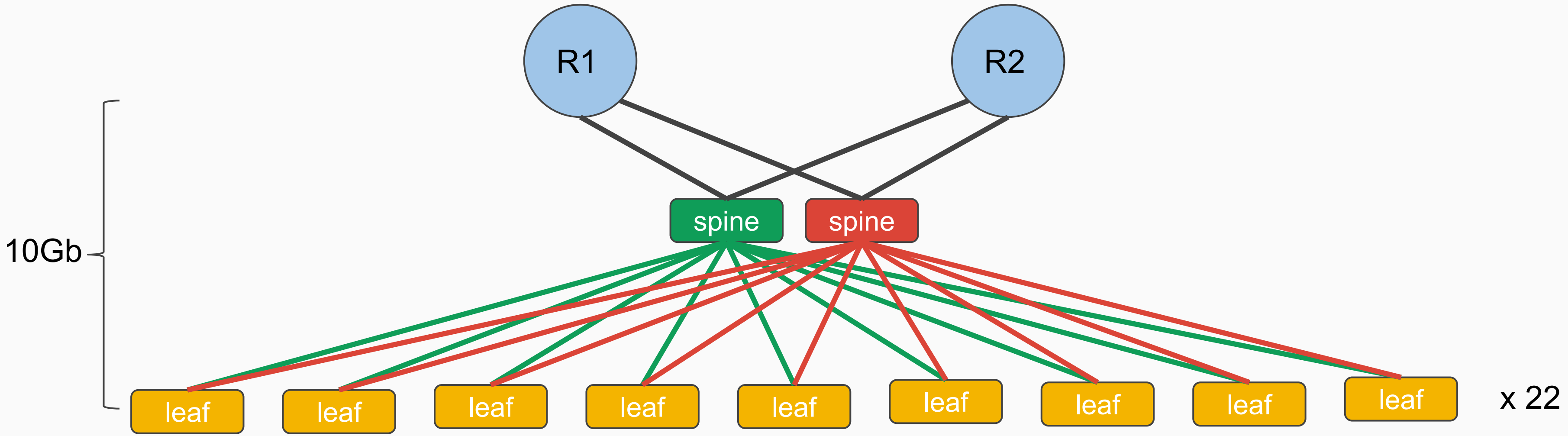
Currently operates AS28715 | **Presentation is about** AS202119

AS28715	Non-profit for operating Tor Exits / Relays
AS202119	$DayJob - 1

Stateless


Peering
core
core
Transit
R3
R4
Peering
R1
R2
Transit
Transit
core
core

Stateless


R1
R2
spine
spine
10Gb
leaf
leaf
leaf
leaf
leaf
leaf
leaf
leaf
leaf
x 22

Stateless

R1	Cisco ASR 1002-x		
R2	Cisco ASR 1002-x		
R3	Cisco ASR 1004		
R4	Cisco ASR 1004		
Core 1	Arista 7050S-52	(52x 10Gb)	
Core 2	Arista 7050-128x	(96x 10Gb	8x 40Gb)
Leaf	Arista 7048T	(48x 1Gb	4x 10Gb)

And then there was SOC II

SOC II

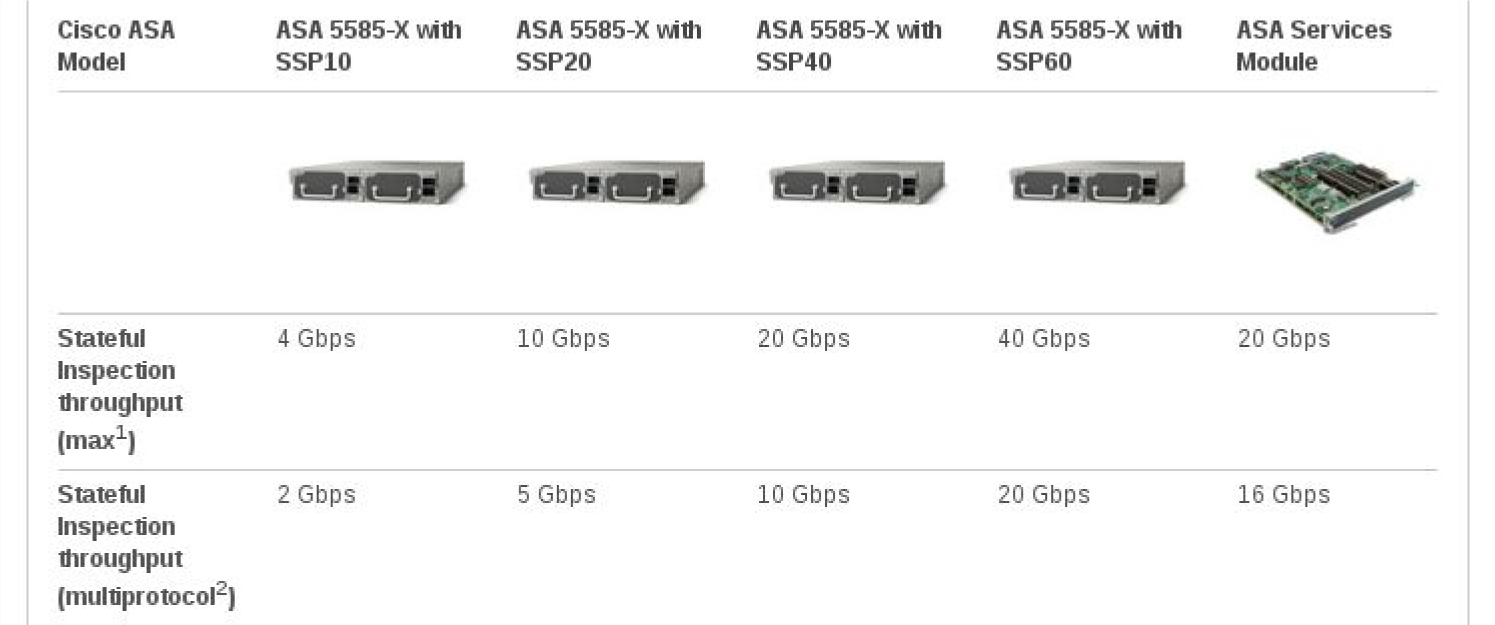
- A stateful inspection firewall shall exist between the Internet and all assets.

- Firewalls shall be configured to allow explicitly approved services and protocols into and out of the environment, with default deny-all.

Requirements

- 1:1 contention within a DC (leaf / spine)
- Didn’t want to have 20Gbits+ of routing capacity constrained by firewalls
- Not cost the earth

Gathering Quotes



Cisco ASA Model	ASA 5585-X with SSP10	ASA 5585-X with SSP20	ASA 5585-X with SSP40	ASA 5585-X with SSP60	ASA Services Module
Stateful Inspection throughput (max[1])	4 Gbps	10 Gbps	20 Gbps	40 Gbps	20 Gbps
Stateful Inspection throughput (multiprotocol[2])	2 Gbps	5 Gbps	10 Gbps	20 Gbps	16 Gbps

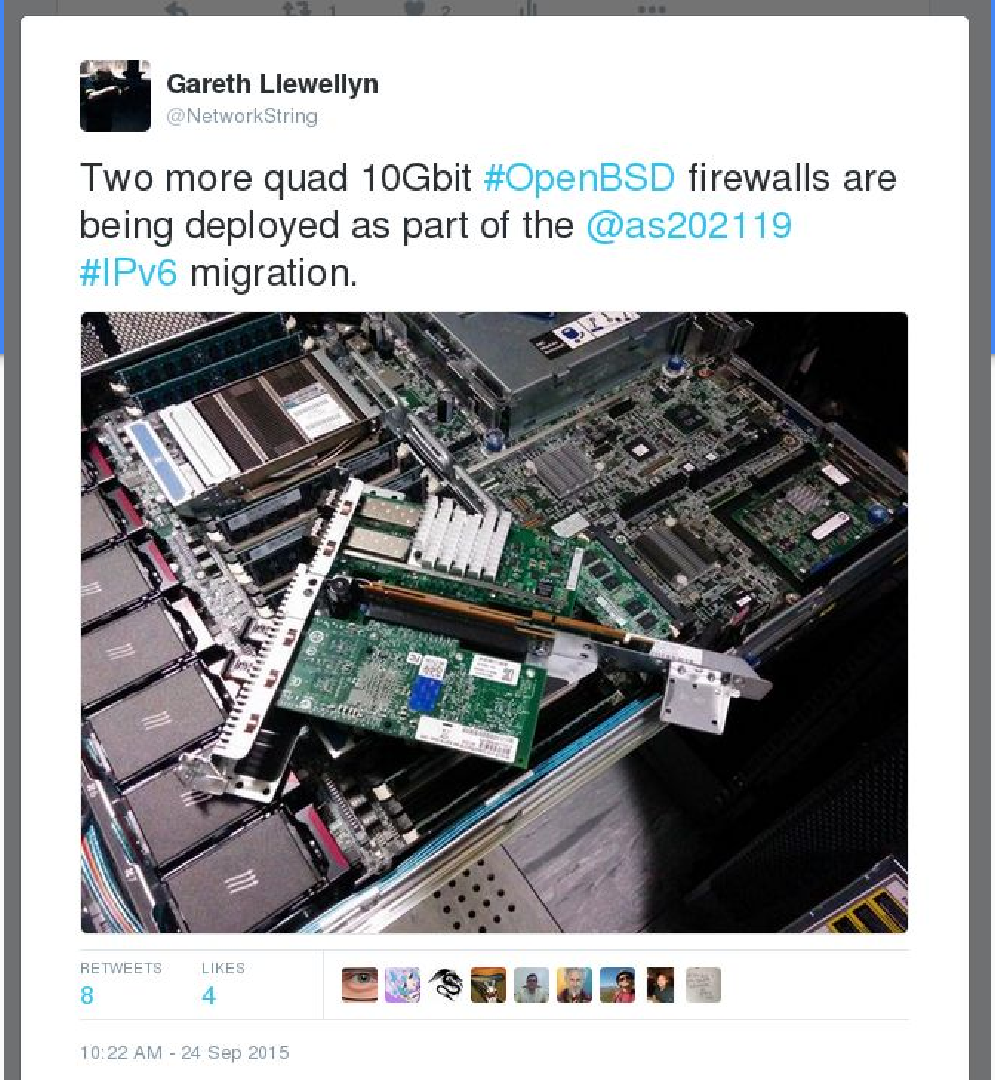
Nope nope nope nope nope

Enter Stage Left: Puffy

Platform

- Stock server was a DL360p Gen8
  - 2x PCI-E slots (x16 + x8)
  - Dual Xeon(R) CPU E5-2630 CPUs
  - 32Gb of RAM amd64
  - 4x 1Gb NICs bge(4)
- Added 2x Intel x520 NICs (2x 10Gb SX) ix(4)
- Hundreds of servers in the DC *(plenty of warm spares if waiting for RMA)*
  - HP DL360p “Core” platform
  - Dell C8000 SW sled “Core” platform
  - Dell C8000 DW sled DB servers
  - Dell R720 Hadoop

Platform



Gareth Llewellyn
@NetworkString

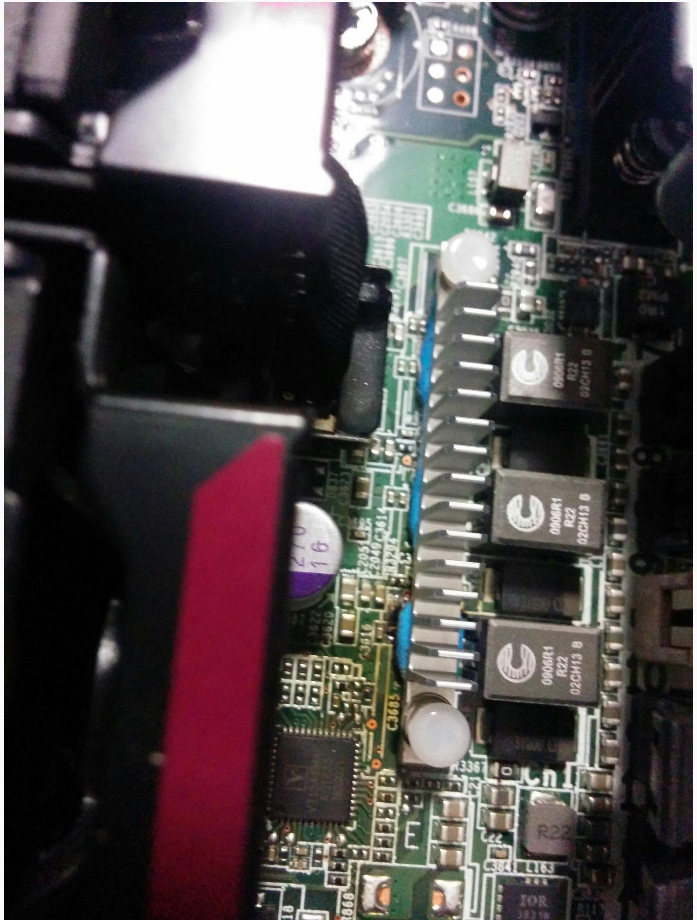
Two more quad 10Gbit #OpenBSD firewalls are being deployed as part of the @as202119 #IPv6 migration.

RETWEETS 8

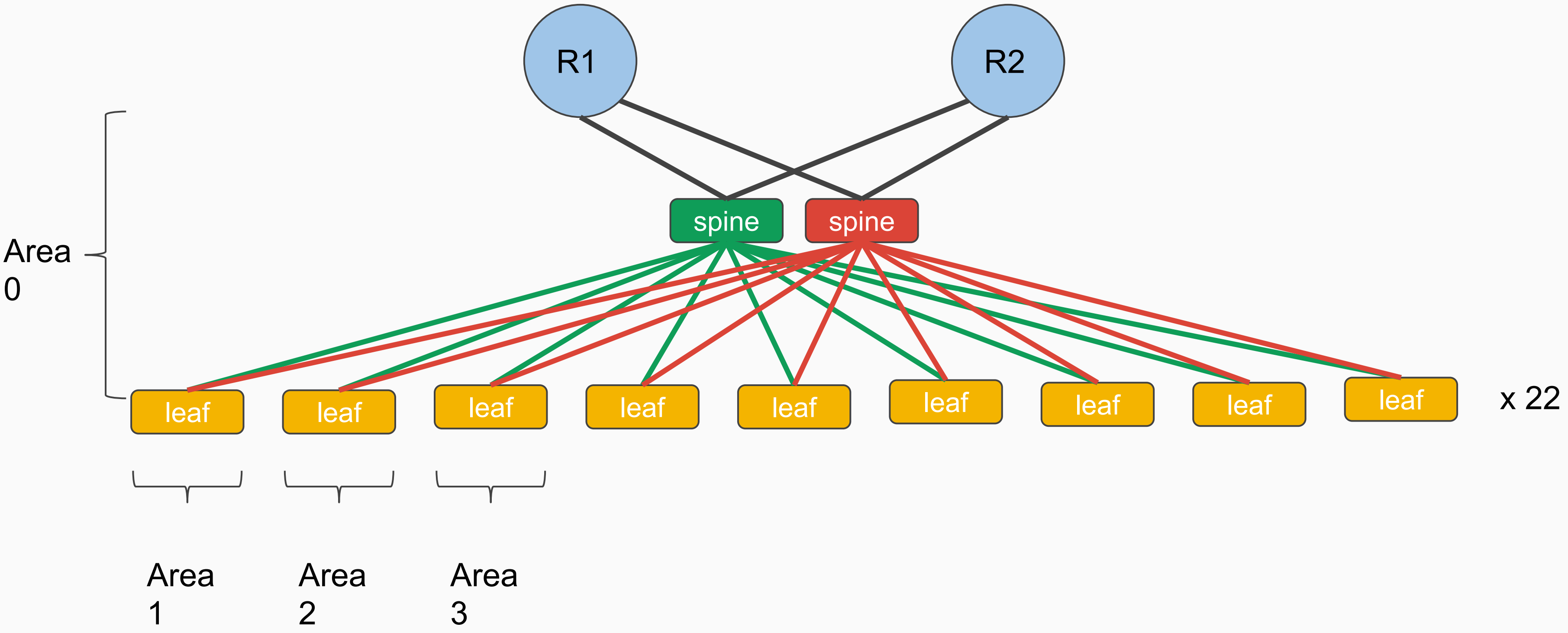
LIKES 4

10:22 AM - 24 Sep 2015

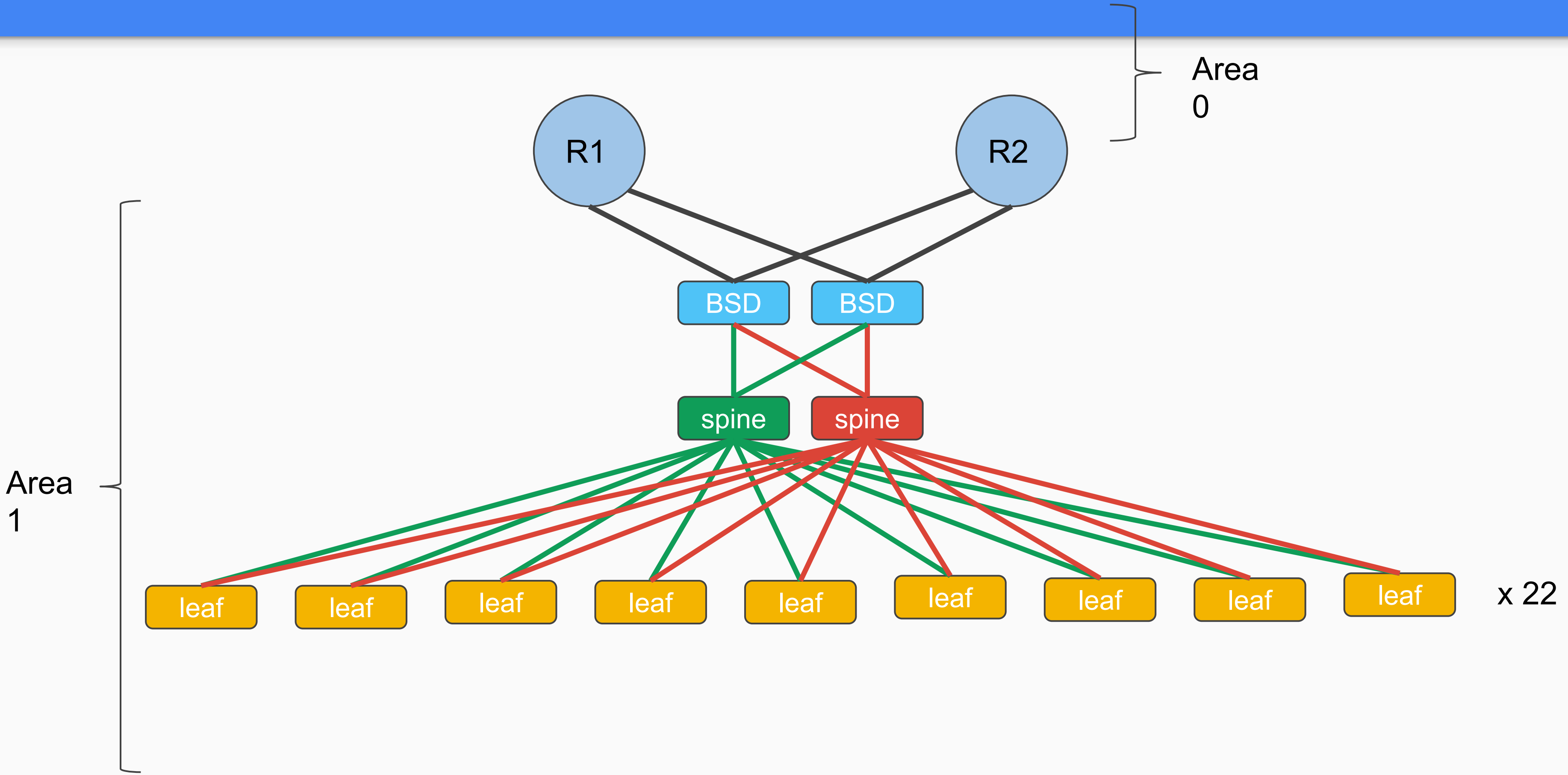
SOAK Testing - Good job we have those spares...



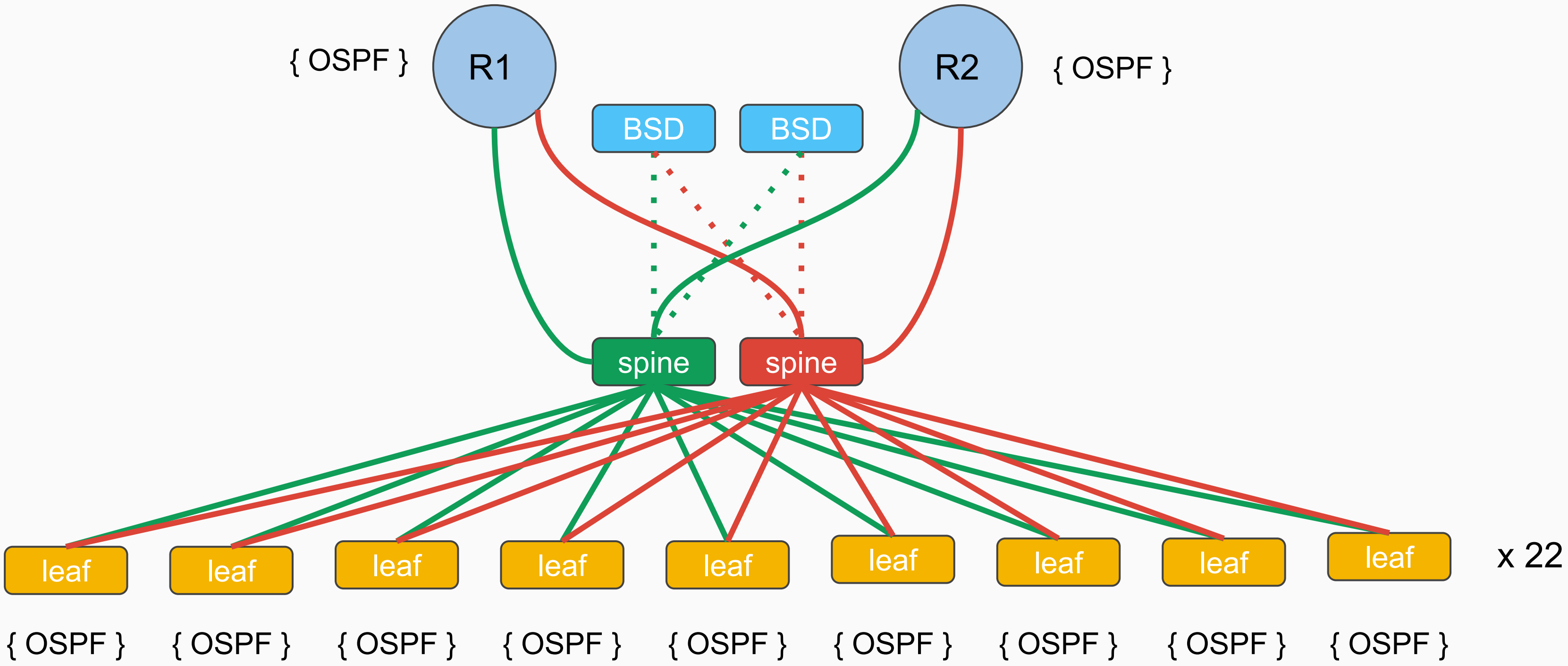
Transition - Starting Point



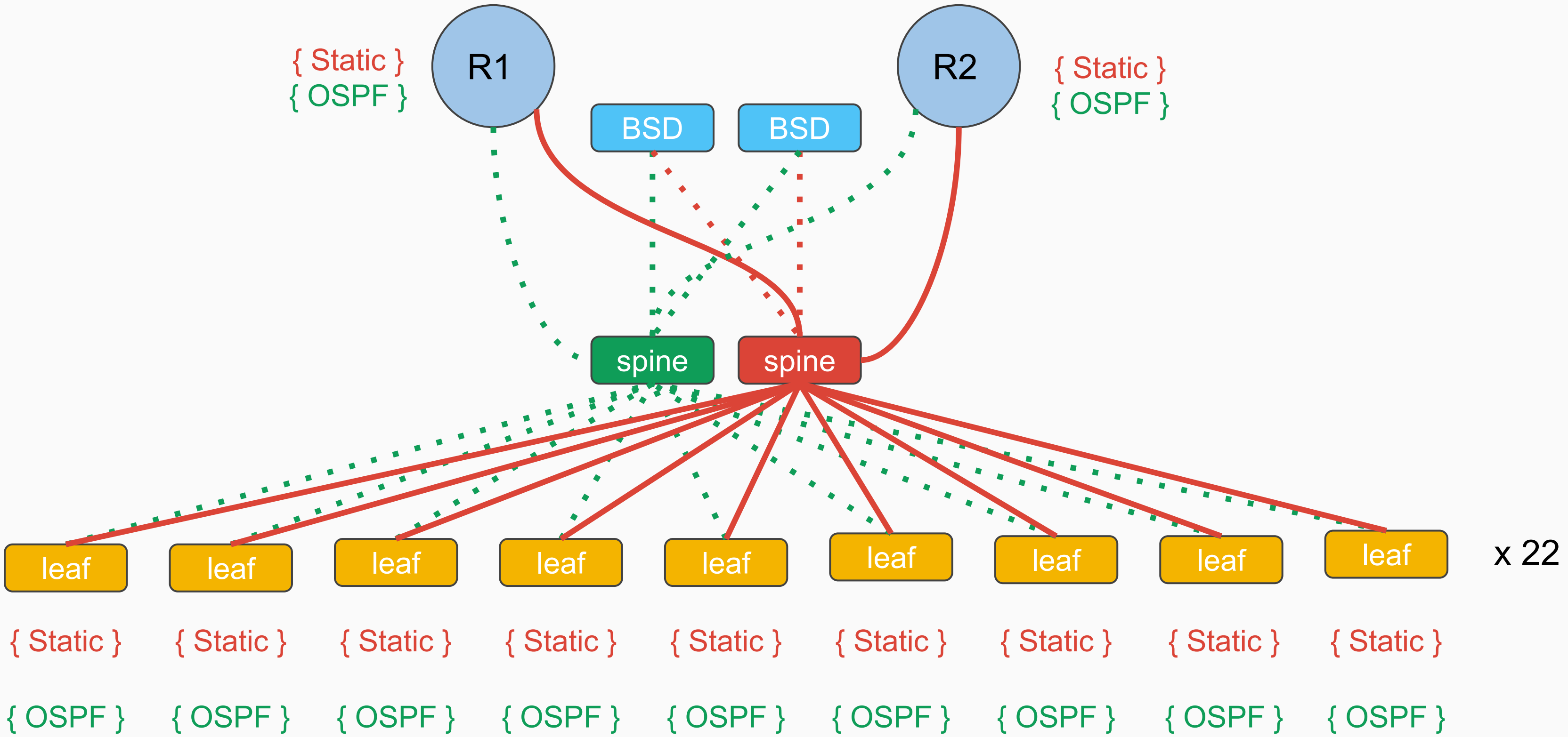
Transition - Finish Point



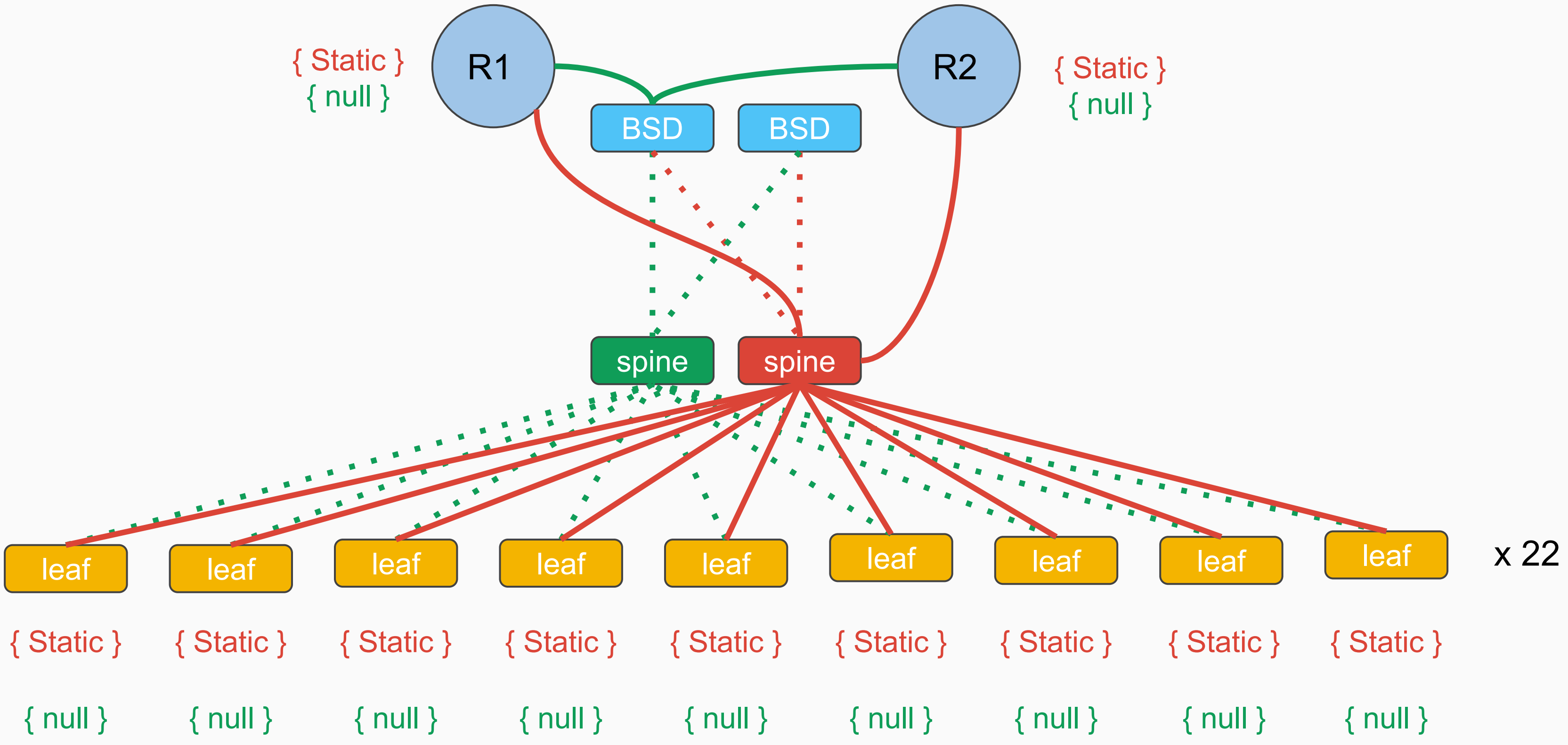
Transition - Drop in the BSDs



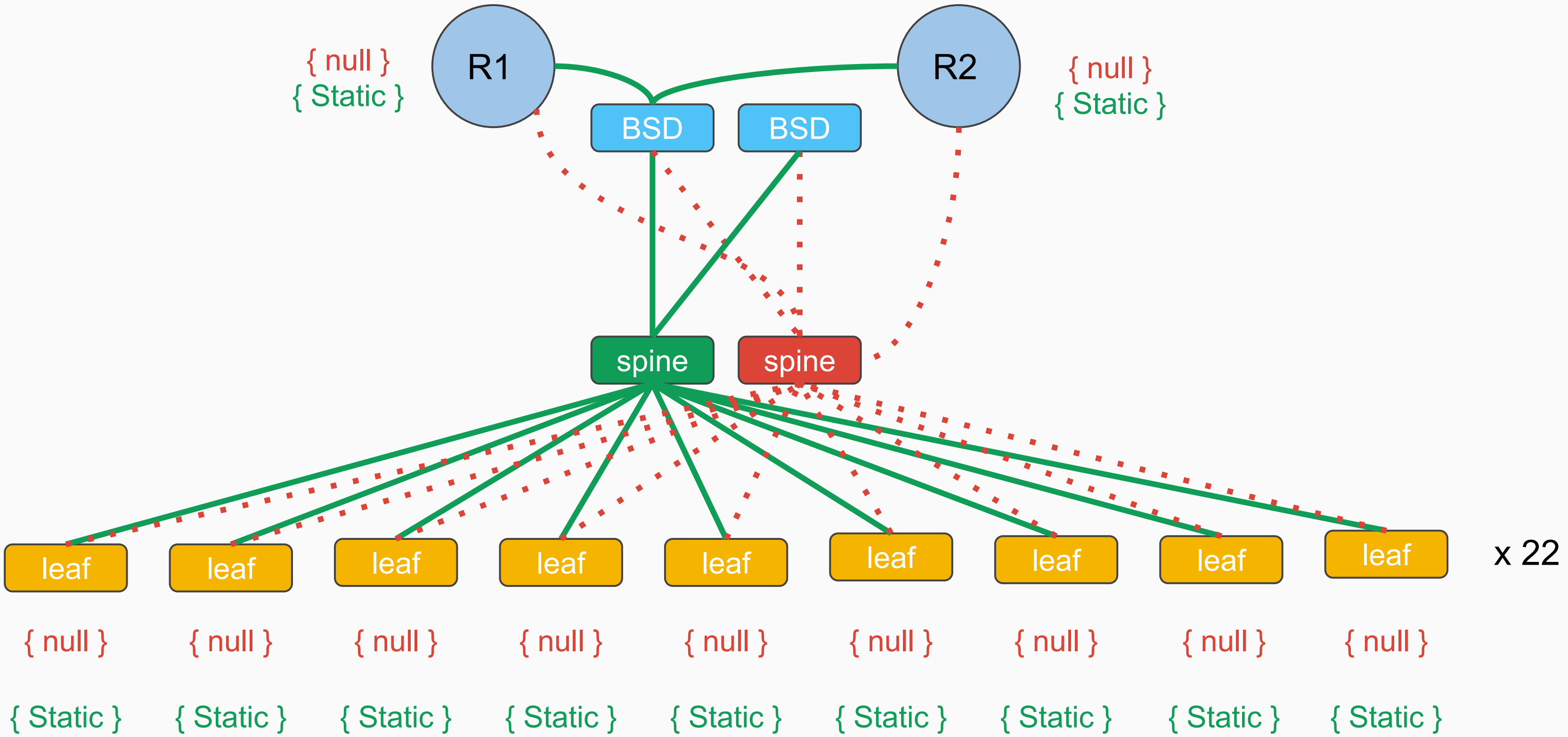
Transition - Statics



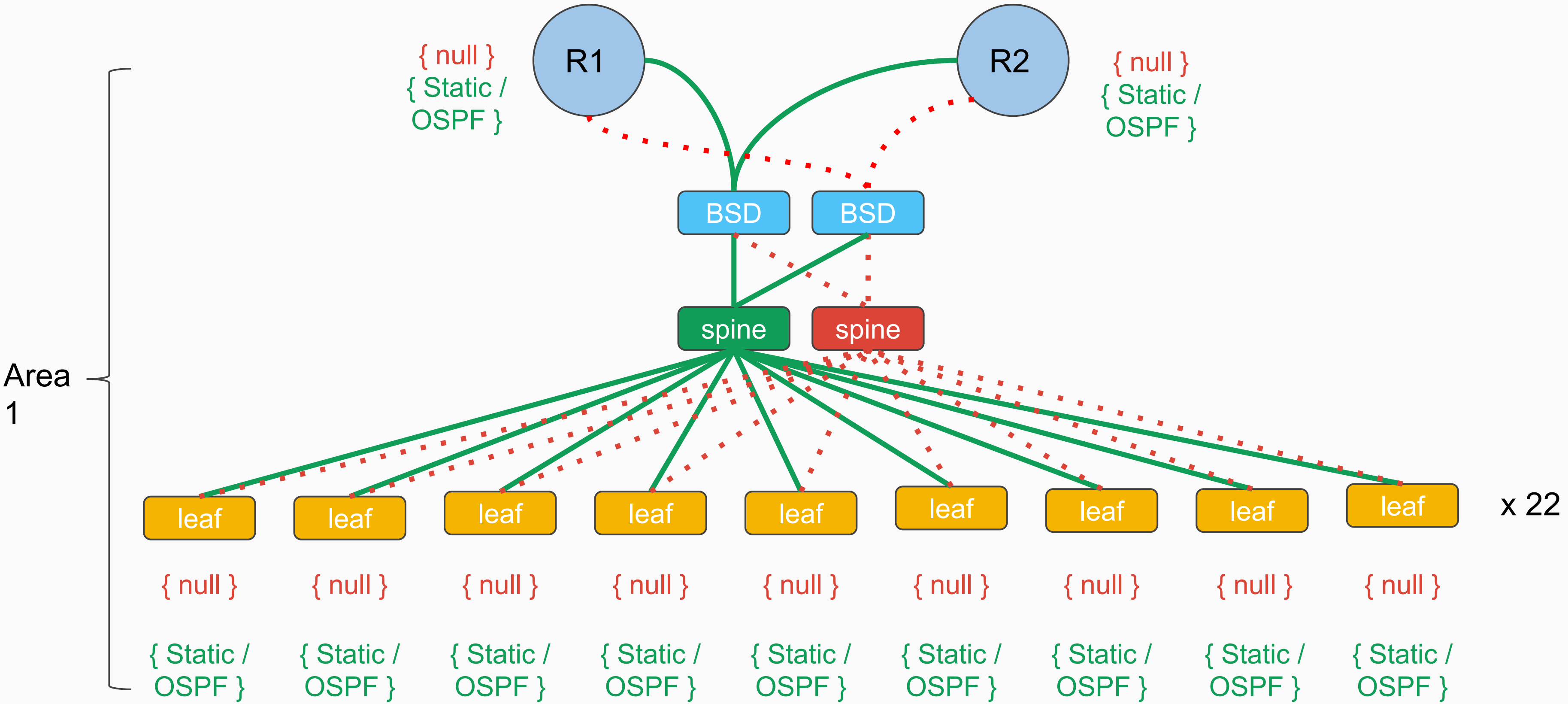
Transition - Statics



Transition - Statics



Transition - Statics



Explosion

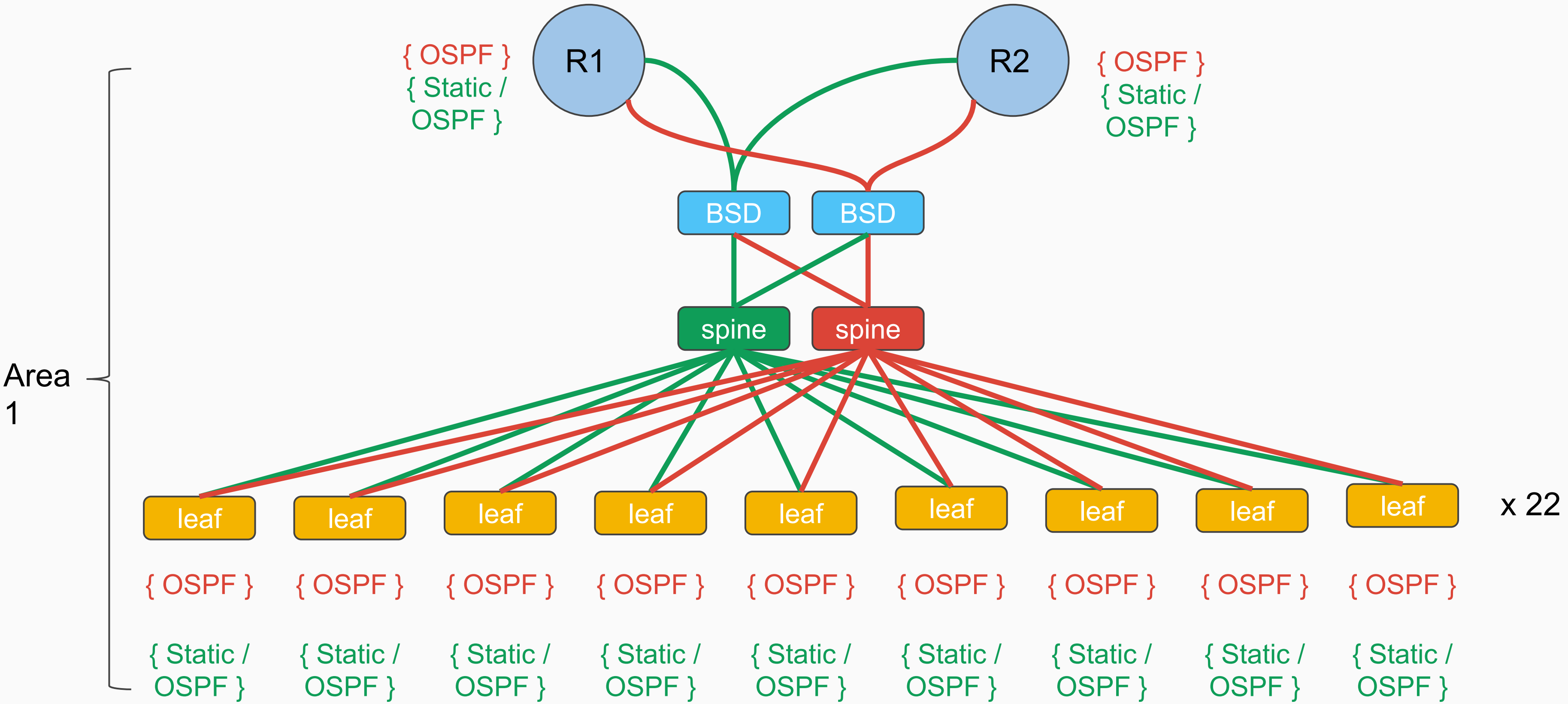
Arista switches started to arbitrarily null route OSPF learnt networks and/or dumping their routing tables.

Explained as: A difference between the way GateD based routers and other devices behave when they receive LSU with the same SEQ number.

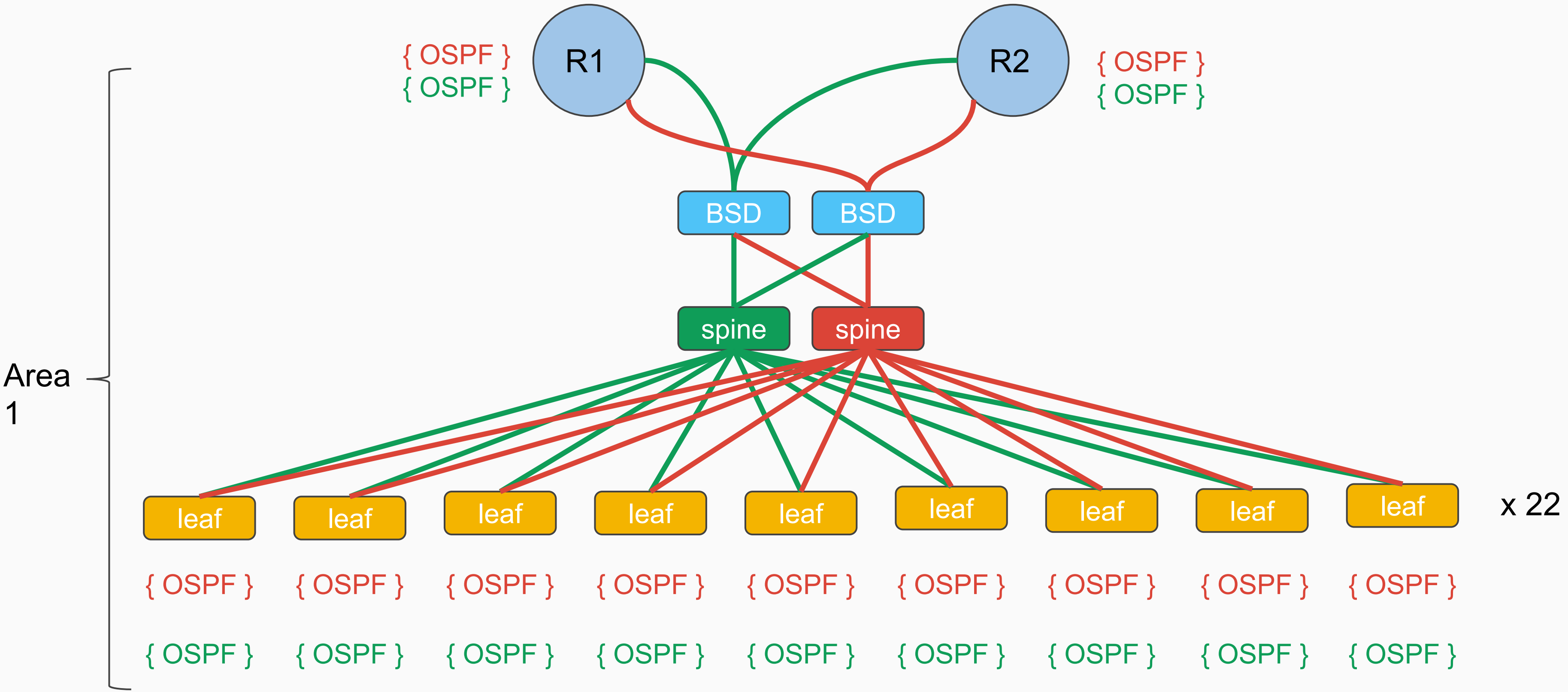
Effectively a difference between Cisco / OpenOSPFd / Arista in regards to checksumming LS updates.

Arista bug 119845 was created

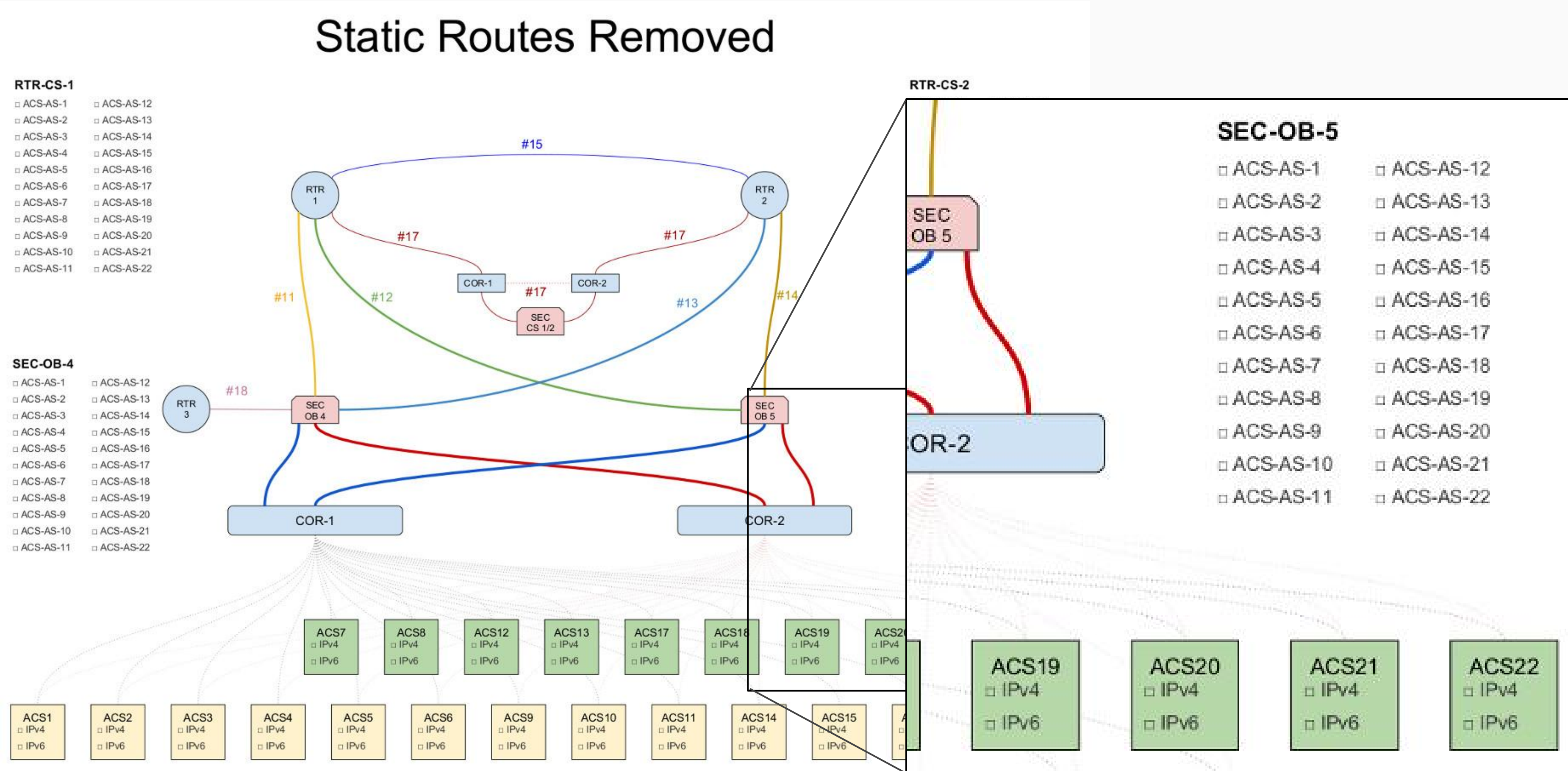
Transition - Statics



Transition - OSPF



Literal Checklists



Static Routes Removed

RTR-CS-1

- □ ACS-AS-1
- □ ACS-AS-2
- □ ACS-AS-3
- □ ACS-AS-4
- □ ACS-AS-5
- □ ACS-AS-6
- □ ACS-AS-7
- □ ACS-AS-8
- □ ACS-AS-9
- □ ACS-AS-10
- □ ACS-AS-11
- □ ACS-AS-12
- □ ACS-AS-13
- □ ACS-AS-14
- □ ACS-AS-15
- □ ACS-AS-16
- □ ACS-AS-17
- □ ACS-AS-18
- □ ACS-AS-19
- □ ACS-AS-20
- □ ACS-AS-21
- □ ACS-AS-22

SEC-OB-4

- □ ACS-AS-1
- □ ACS-AS-2
- □ ACS-AS-3
- □ ACS-AS-4
- □ ACS-AS-5
- □ ACS-AS-6
- □ ACS-AS-7
- □ ACS-AS-8
- □ ACS-AS-9
- □ ACS-AS-10
- □ ACS-AS-11
- □ ACS-AS-12
- □ ACS-AS-13
- □ ACS-AS-14
- □ ACS-AS-15
- □ ACS-AS-16
- □ ACS-AS-17
- □ ACS-AS-18
- □ ACS-AS-19
- □ ACS-AS-20
- □ ACS-AS-21
- □ ACS-AS-22

RTR-CS-2

SEC-OB-5

- □ ACS-AS-1
- □ ACS-AS-2
- □ ACS-AS-3
- □ ACS-AS-4
- □ ACS-AS-5
- □ ACS-AS-6
- □ ACS-AS-7
- □ ACS-AS-8
- □ ACS-AS-9
- □ ACS-AS-10
- □ ACS-AS-11
- □ ACS-AS-12
- □ ACS-AS-13
- □ ACS-AS-14
- □ ACS-AS-15
- □ ACS-AS-16
- □ ACS-AS-17
- □ ACS-AS-18
- □ ACS-AS-19
- □ ACS-AS-20
- □ ACS-AS-21
- □ ACS-AS-22

Pain Points

PFSYNC

DDOS

Syncing Rules

pfsync(4)

- Asynchronous Routing
  - Dropped packets
- 4(8)x 10Gbit interfaces vs 1x 1Gb syncdev
  - Can’t increase *maxupd* too much
- Dirty hack
  - OSPF weights
  - Let TCP / applications retry in the event of a failure

Pain Points

PFSYNC

DDOS

Syncing Rules

DDOS

- ~11Gbit/s of additional traffic
  - Weekly
  - 99% DNS Reflection
  - Lasts an hour or two
- PF did not like this
- Had to hand back off to the ASRs

Pain Points

PFSYNC

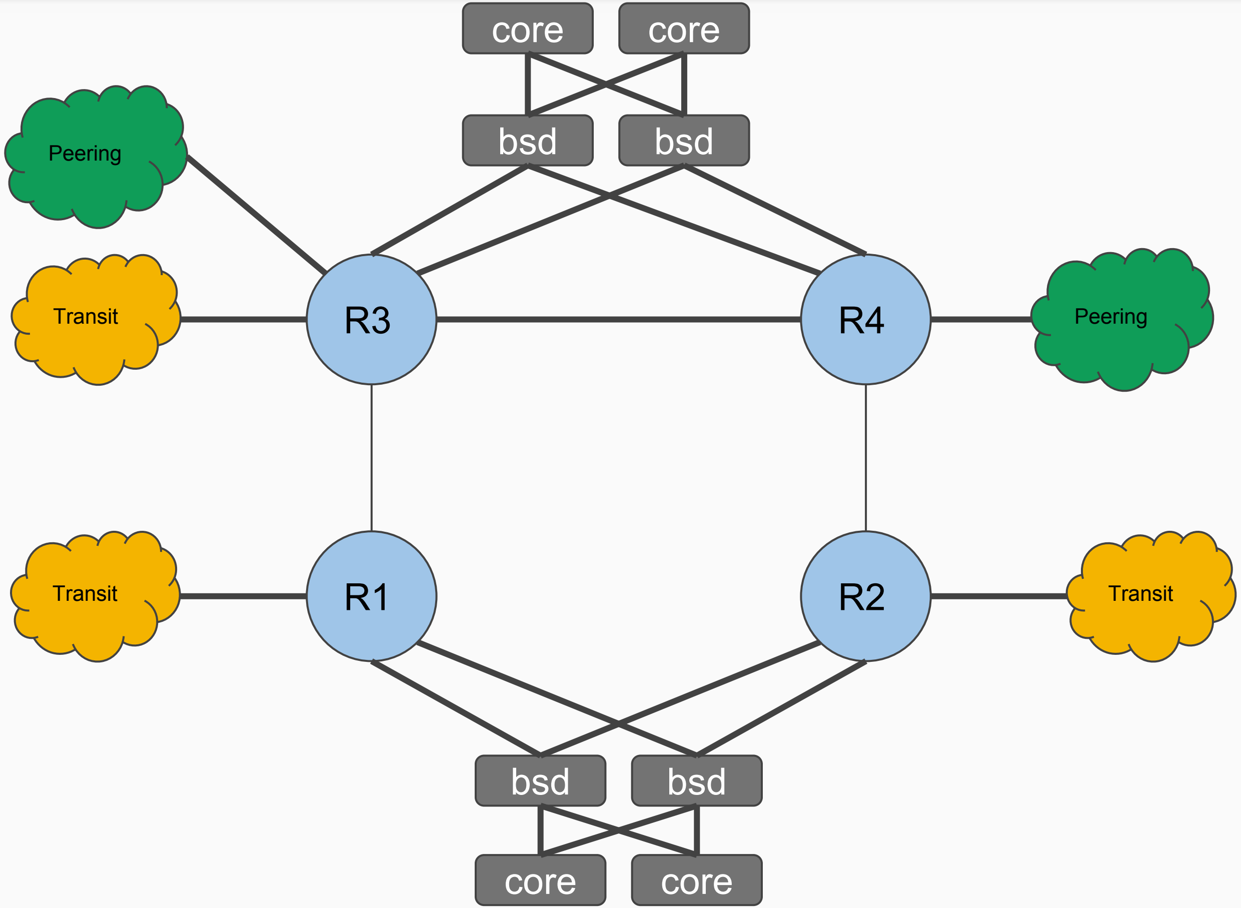
DDOS

Syncing Rules

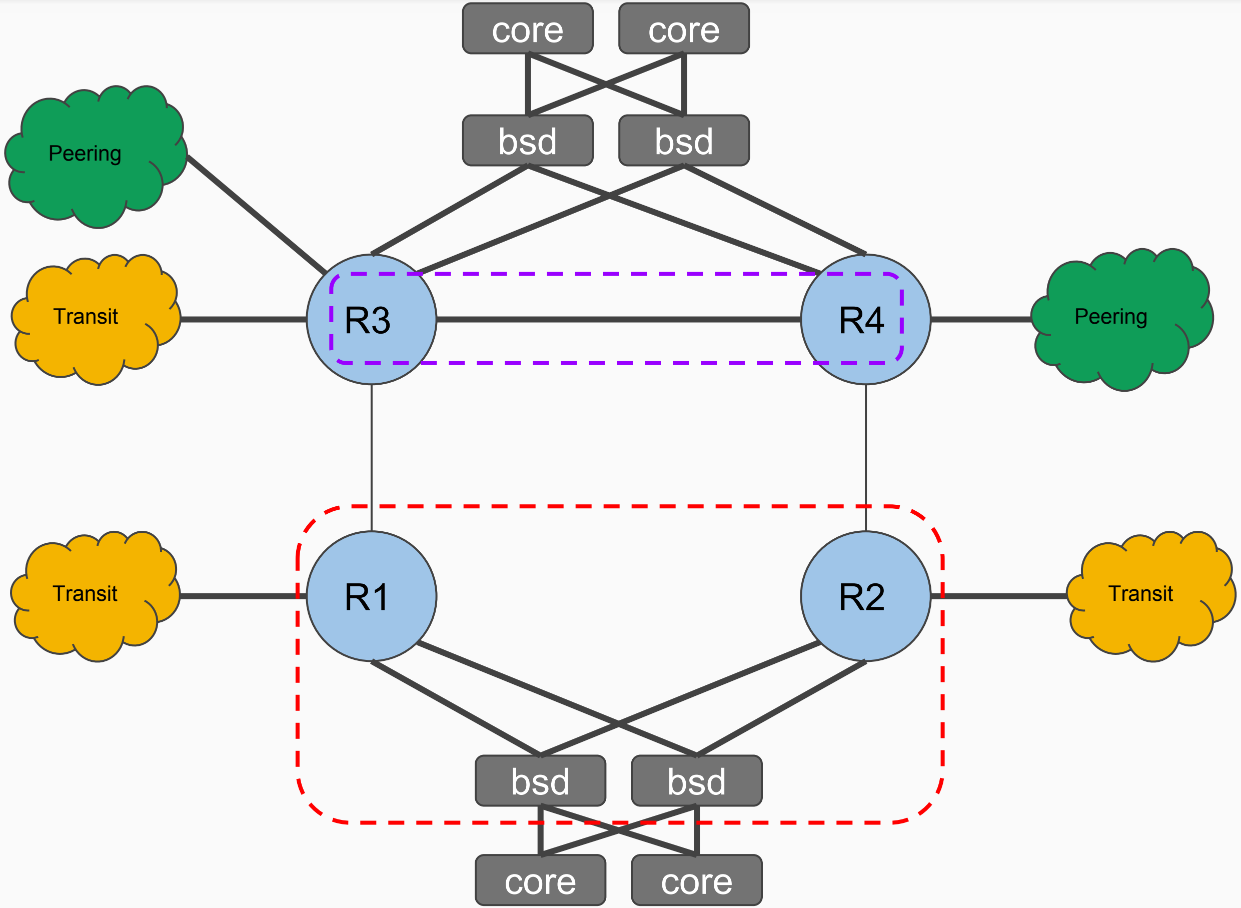
Syncing Rules

- We use Chef on all other servers
- Currently
  - Make a change on the ‘primary’ *(remember OSPF hack)*
  - Then on the secondary
- Need a better way
  - Chef
  - pf tables + magic

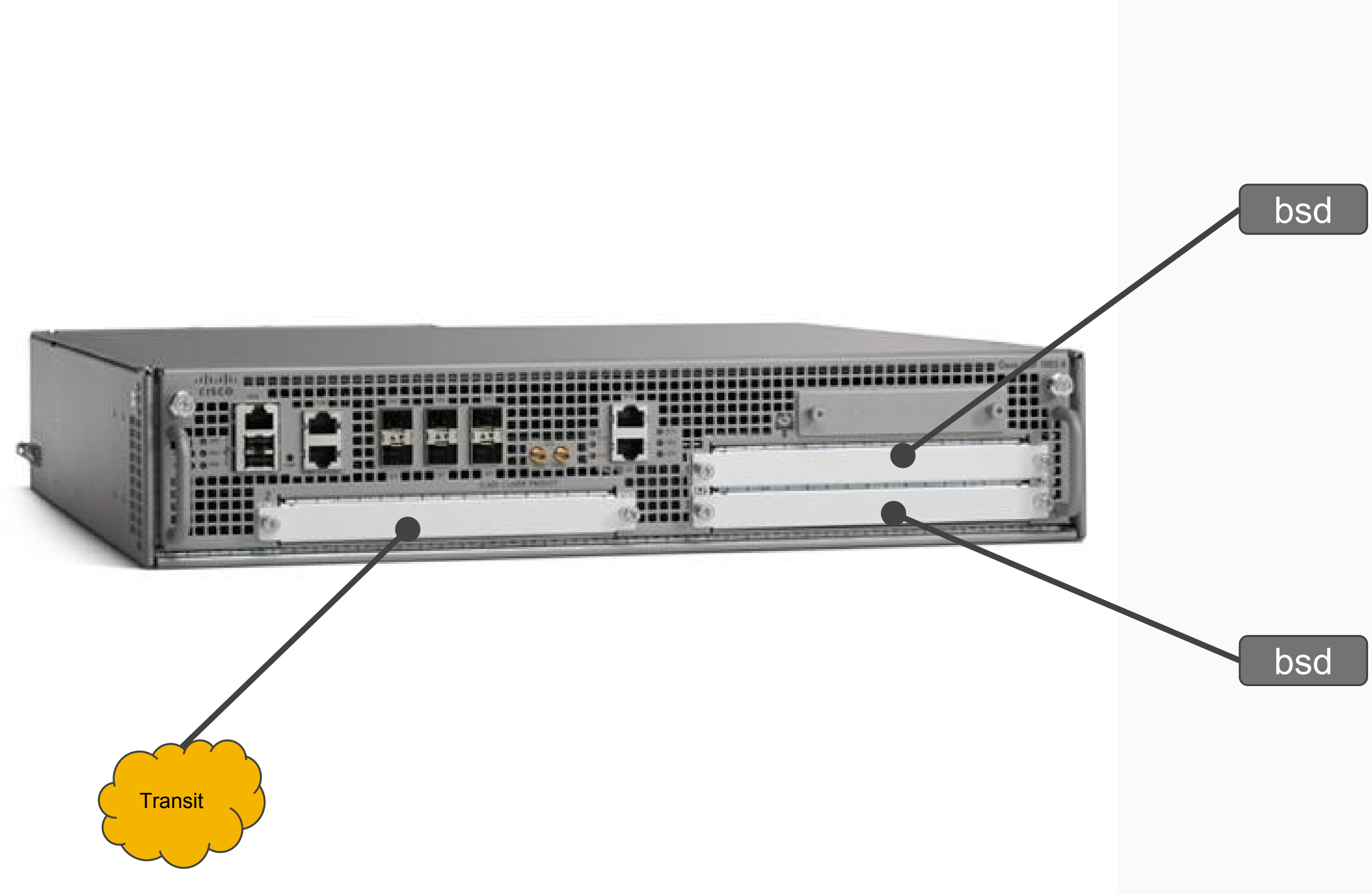
Wahoo


core
core
bsd
bsd
Peering
Transit
R3
R4
Peering
Transit
R1
R2
Transit
bsd
bsd
core
core

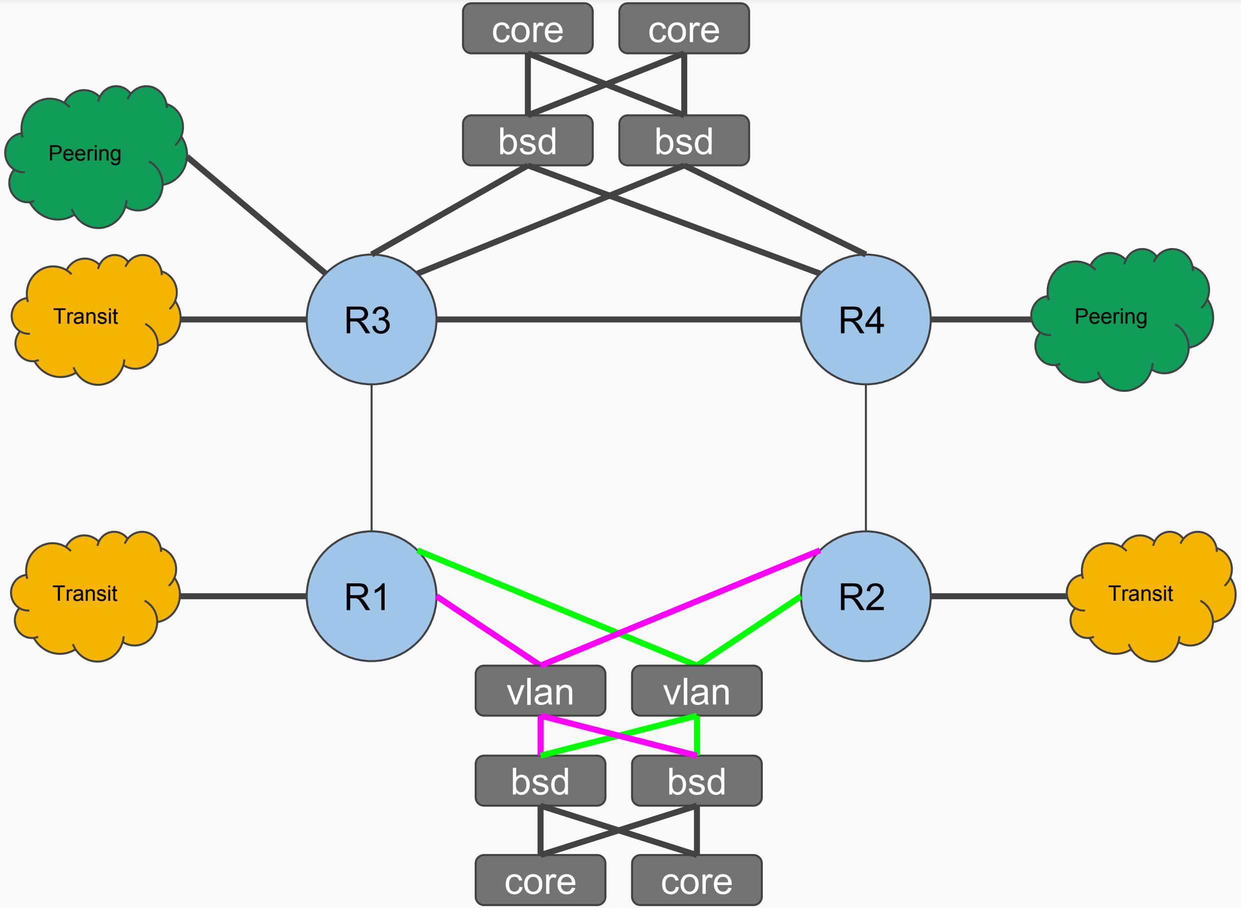
Wahoo - Not So Much



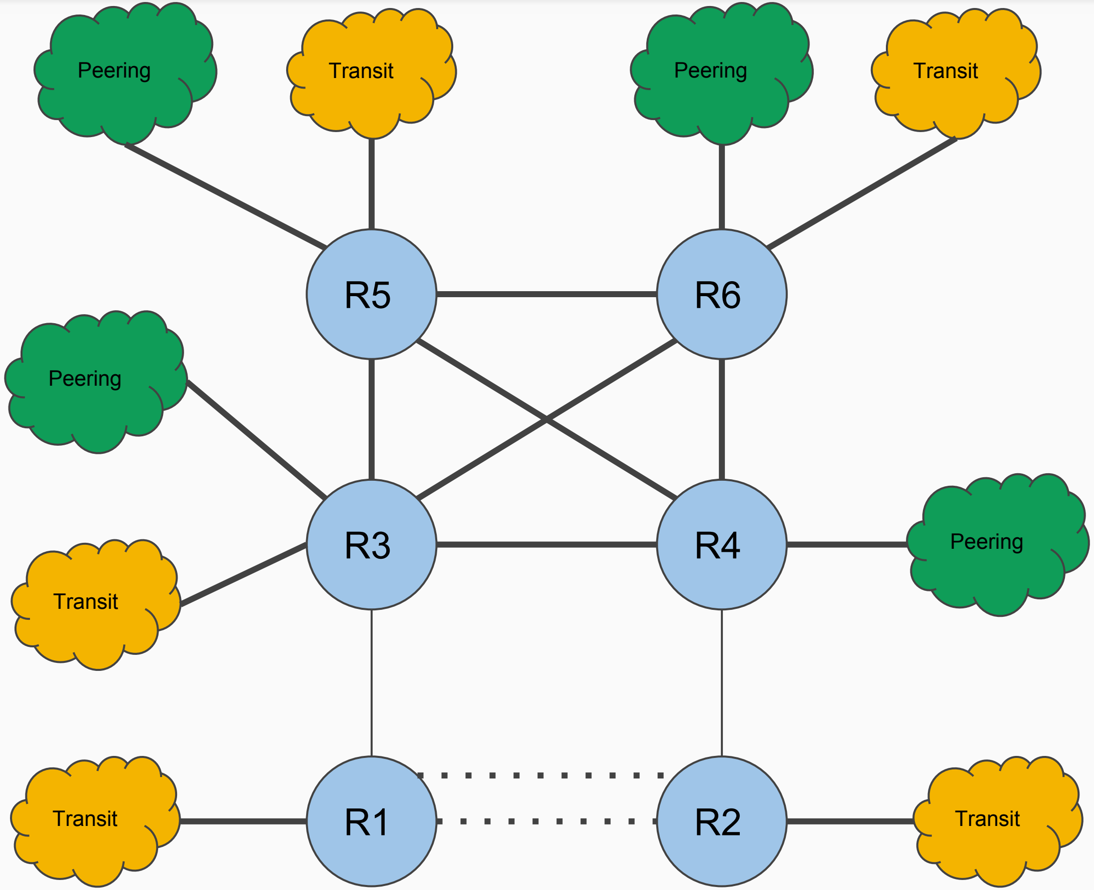
ASR 1002-X



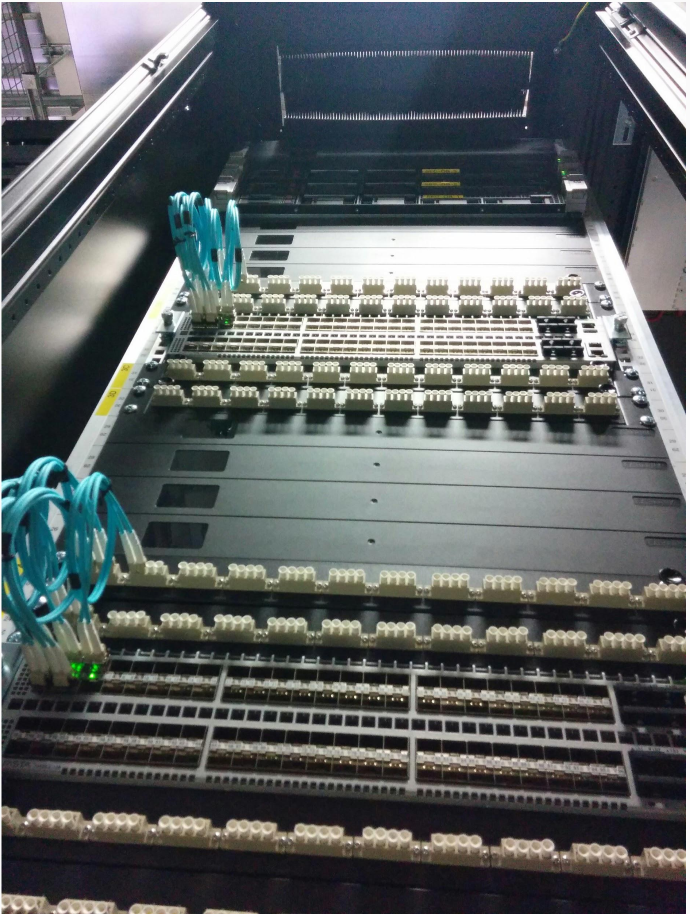
Wahoo - Well, it works

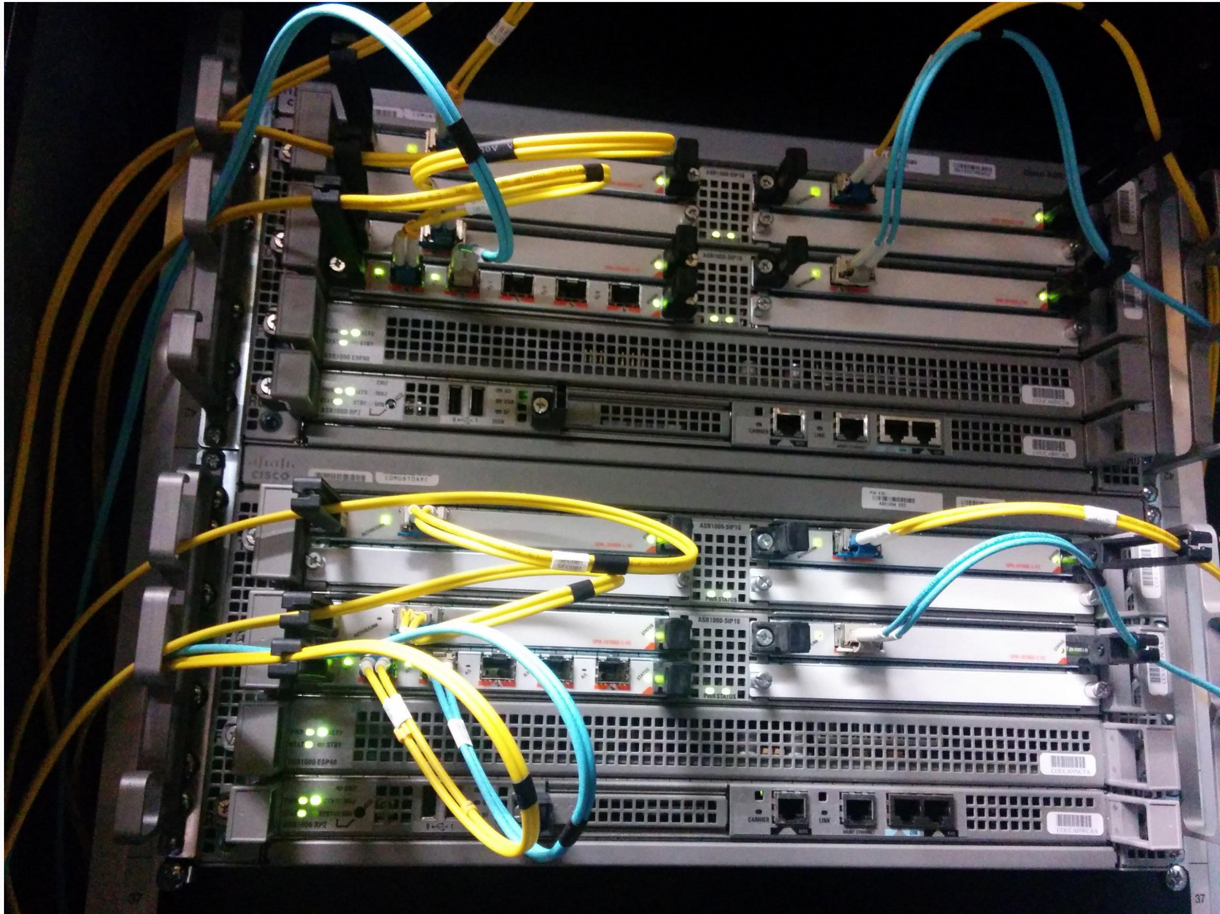


Next Steps

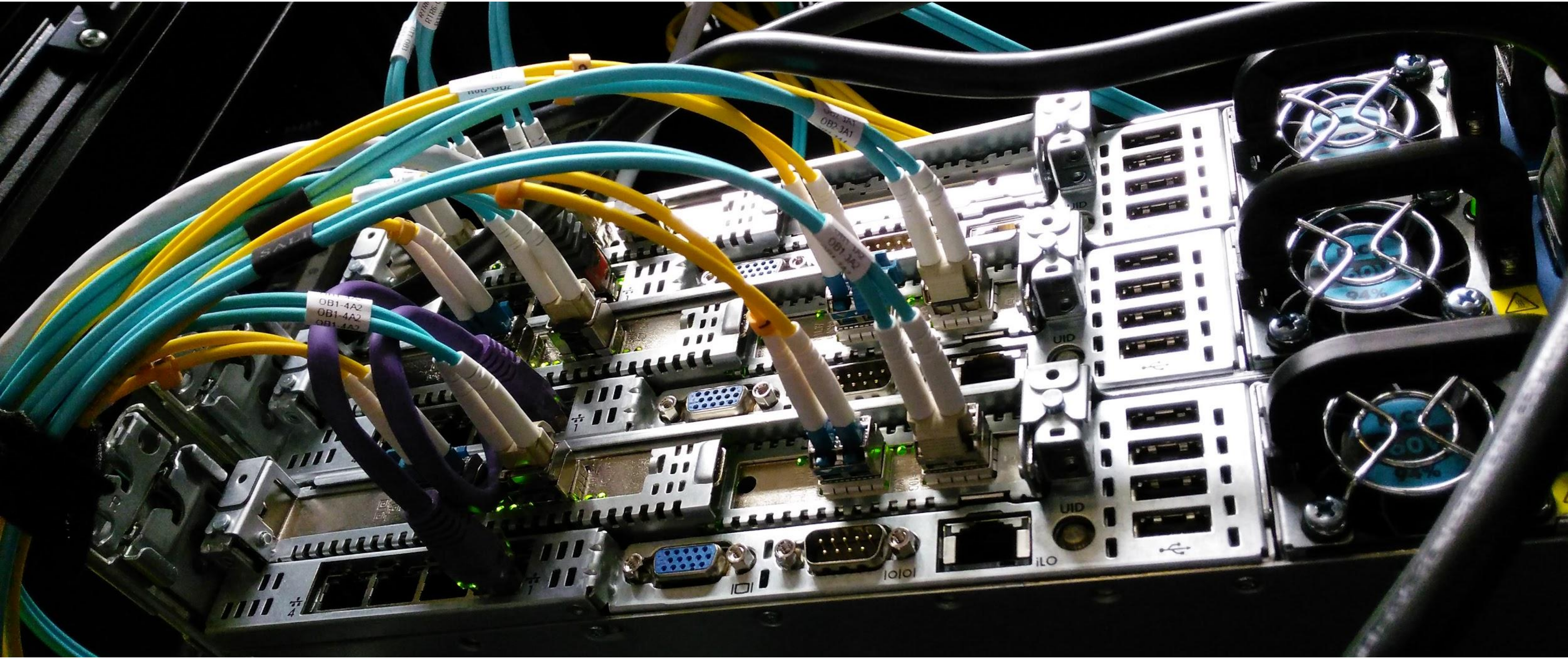

Peering
Transit
Peering
Transit
R5
R6
Peering
R3
R4
Peering
Transit
Transit
R1
R2
Transit

Pictures - Because





Photos



The first time buying an operating system…

Was FOSS

OpenBSD 5.6
Multi-Platform Ultra Secure Operating System
Free, functional and secure
Multi-Platform Ultra Secure Operating System
OPENBSD 5.9
FREE FUNCTIONAL AND SECURE
MULTI-PLATFORM ULTRA SECURE OPERATING SYSTEM
OPENBSD 6.0
CD1: i386, alpha, hppa
CD2: amd64, macppc, and songs
CD3: sparc64, source code
CD can boot on i386, amd64, macppc
Other architectures available on the Internet

Questions?