



IPv6 TRANSITION FOR THE ENTERPRISE

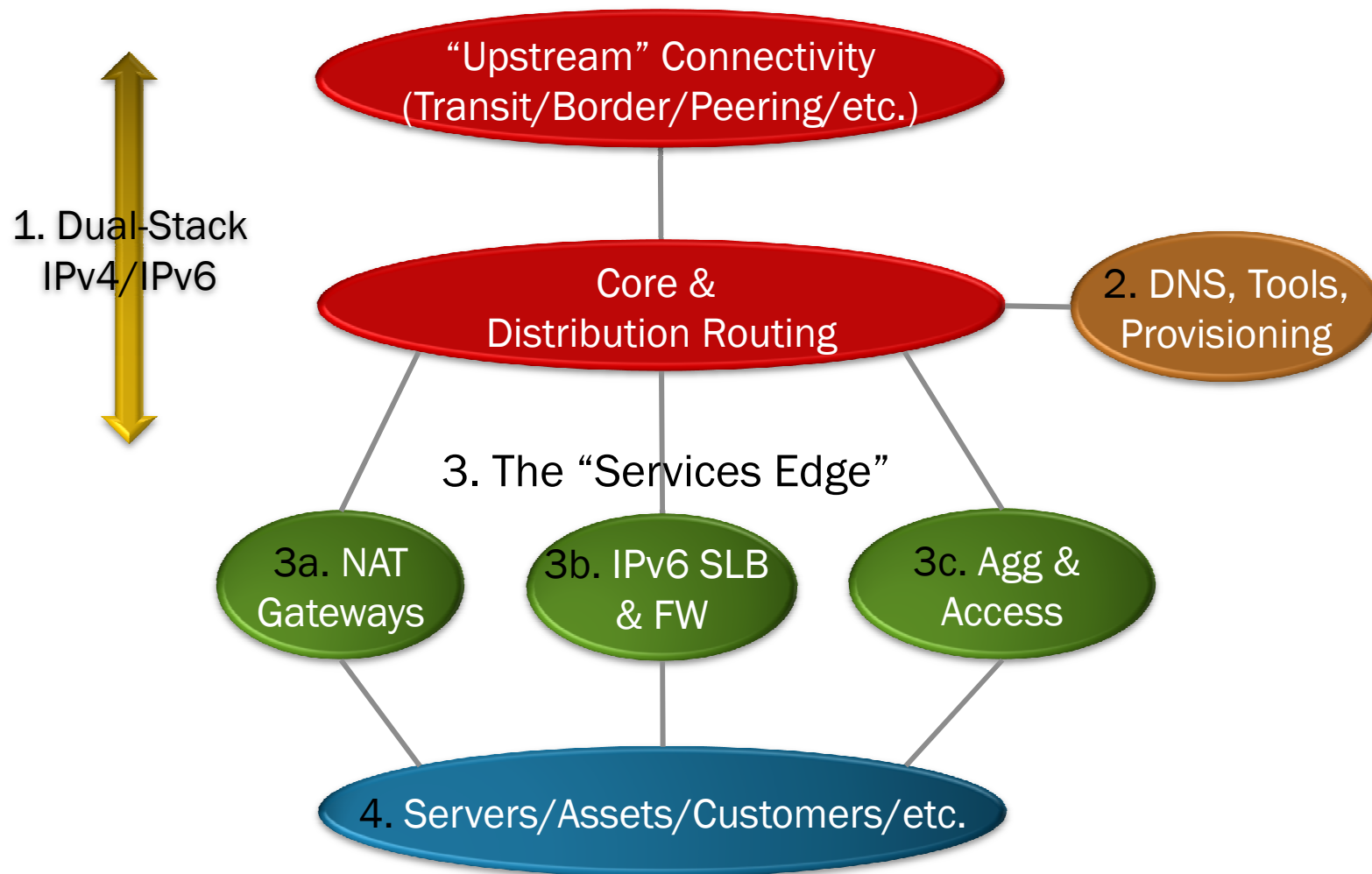
Brief selected topics

Jeff Hartley, SP ADP SE



Observations on IPv6 Deployment Trends

“Where do successful sites commonly deploy first?”



Observations on IPv6 Deployment Trends

“Where do sites commonly deploy first?”

- Numbering and routing v6 (at least) down through the Distribution layer is always a good place to start
- While DNS, tools, and provisioning services (i.e., Ops) are often thought of LAST, the sites that deploy v6 there early are the most successful.
- Numerous options exist for migrating everything else gradually – matching existing refresh cycles makes sense.
- High Availability options need to be reconsidered as each class of use-case (application/service/customer) is considered for IPv6 deployment.

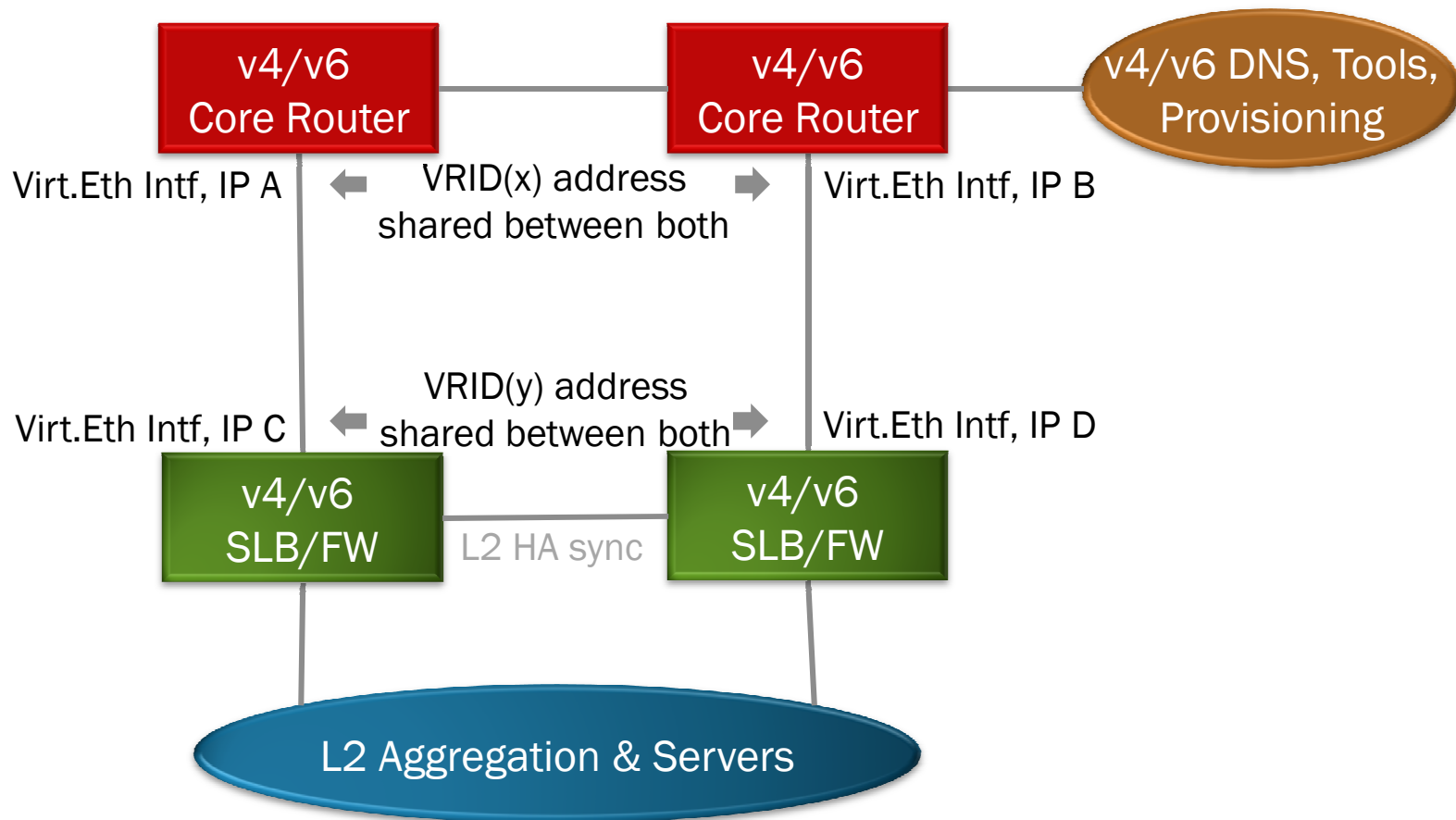


Services edge interconnections in dual-stack environments



Services edge: adding v6 services

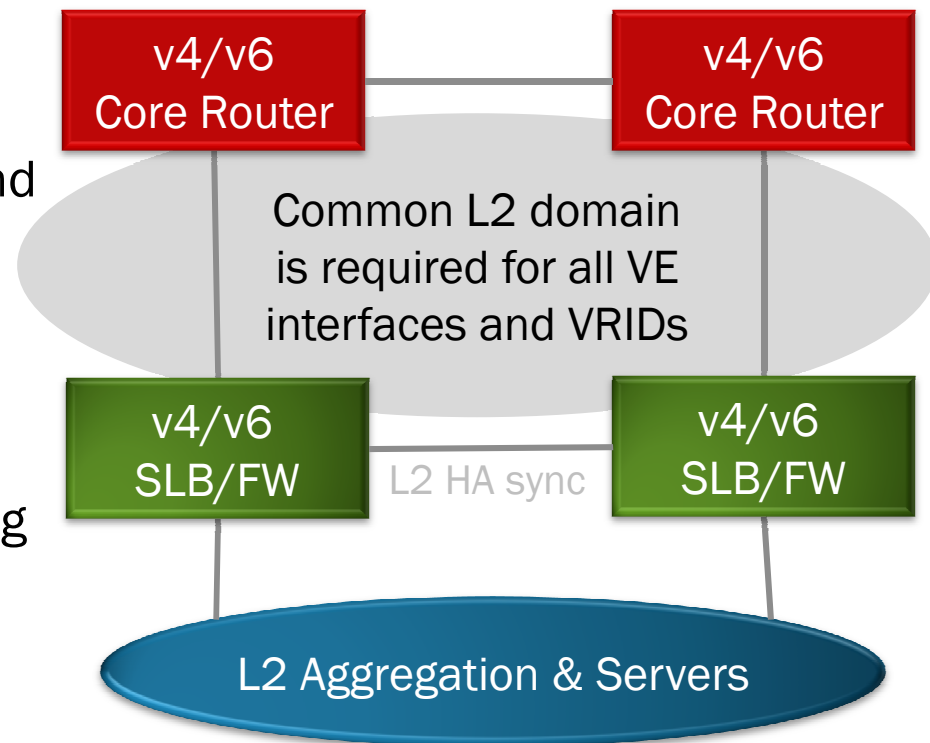
Solution: Use FHRP (VRRP-e in this case) for both IPv4 and IPv6



Services edge: adding v6 services

Solution: Use FHRP (VRRP-e) for both IPv4 and IPv6

- v4 & v6 VRIDs are added to VEs
- Static routes for all v4 & v6 subnets are set on the Core routers for all subnets below, and redistributed into OSPFv2 & OSPF v3
- Multi-Chassis types of devices used for Core can appear as a single large router, thus reducing the scope of the inter-device L2 domain
- As many pairs of FWs/SLBs/etc. can be homed in this dual-stacked VRRP domain as space permits.



“Hey, why can’t I use a Global (formerly ‘public’) IP here?”

An extremely common question during deployments...

- The VRID is a Link-Local address; as with VRRPv3 draft RFC
- Remember that “routing is based on destination!”
- All IPv6 hosts have Link-Local addresses by default, so they’re “all in the same subnet”
- Only the owner announces RAs for the VRID
- It’s handy to use an easily-identifiable address scheme, instead of EUI-64 of the standard virtual MAC (00-00-5E-00-02-XX), which should not be used.
- RFC 5798 – IPv6 routers running VRRP MUST create their Interface Identifiers in the normal manner (e.g., "Transmission of IPv6 Packets over Ethernet Networks" [[RFC2464](#)]). They MUST NOT use the virtual router MAC address to create the Modified Extended Unique Identifier (EUI)-64 identifiers. This VRRP specification describes how to advertise and resolve the VRRP router's IPv6 link-local address and other associated IPv6 addresses into the virtual router MAC address. – IANA has assigned an IPv6 link-local scope multicast address for VRRP for IPv6. The IPv6 multicast address is as follows:
FF02:0:0:0:0:0:12 The values assigned address should be entered into [Section 5.1.2.2](#). The IANA has reserved a block of IANA Ethernet unicast addresses for VRRP for IPv6 in the range 00-00-5E-00-02-00 to 00-00-5E-00-02-FF (in hex).



Services edge: adding v6 services

Summarizing dual-stack VRRP (or VRRP-e, HSRP, NSRP, etc.)

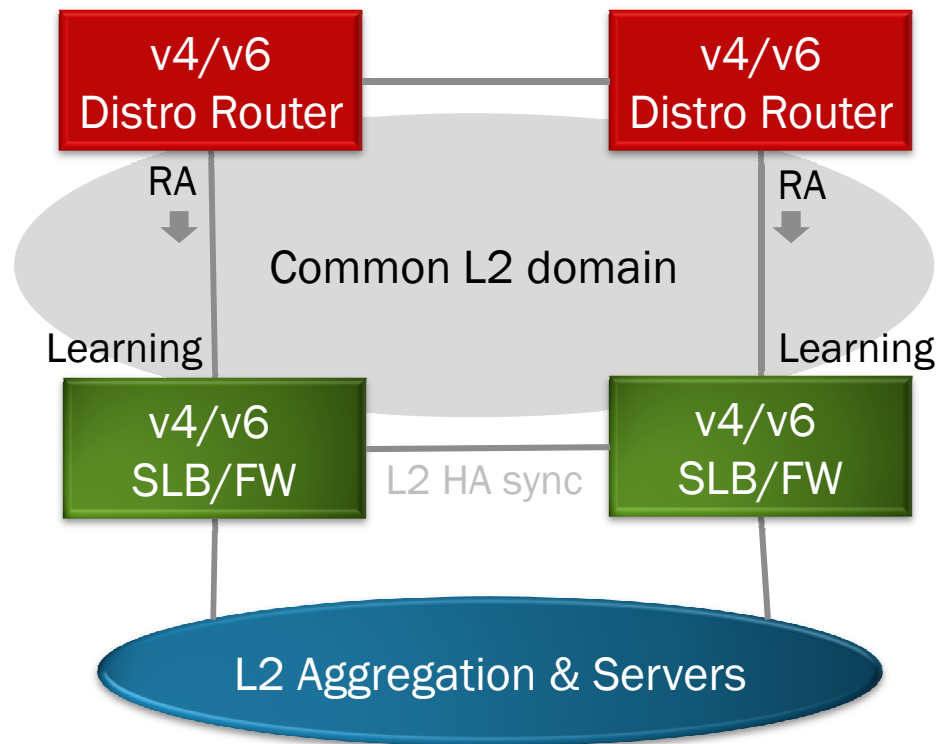
- Conclusion: It works, it's very similar to existing best common practices, and people understand it with only having to learn a slight modification to existing practices.
- The vendor-specific variants of VRRP are starting to add custom options, such as permitting global IPs as the VRID or multiple IPs per VRID.
- ...Is there another, even simpler alternative provided via IPv6 standards?



RA-based default route learning

Solution: It's not just for servers & desktops!

- While this doesn't change v4 requirements, there is no need for the VRID's shared/virtual IP.
- Both def.routes exist in table simultaneously; it's just a matter of tuning configs to send, receive, & learn at appropriate priorities and intervals.



RA-based route learning: Vendor extended

From the RFC 4861 text:

- MaxRtrAdvInterval

The maximum time allowed between sending unsolicited multicast Router Advertisements from the interface, in seconds. MUST be no less than 4 seconds and no greater than 1800 seconds.

Default: 600 seconds

- MinRtrAdvInterval

The minimum time allowed between sending unsolicited multicast Router Advertisements from the interface, in seconds. MUST be no less than 3 seconds and no greater than $.75 * \text{MaxRtrAdvInterval}$.

Default: $0.33 * \text{MaxRtrAdvInterval}$, if $\text{MaxRtrAdvInterval} \geq 9$ seconds; otherwise, the Default is MaxRtrAdvInterval .



RA-based route learning: Vendor extended

Vendor extensions

- Brocade example: A multiplier between 0.5 and 1.5 is applied to the configured RA-interval when an interface comes up or the interval value is modified. Thus a system doesn't have to simultaneously process all RAs at the exact same time every interval.
- Cisco example: Some OS versions allow for manually specifying sub-second intervals, although receiving stations need to be expecting the load.
- The current *Actual* RA-interval value for any given interface can be seen via the usual "show..." (ex: "show ipv6 int") commands
- Always disable/filter RAs where you don't need/want them!



RA-based route learning

Example

- An upstream box **sending** out a high-preference RA every 1 sec, valid for 5 sec:
`(config)# interface ethernet 3/1`
`(config-if-e1000-3/1)# ipv6 nd ra-interval 1`
`(config-if-e1000-3/1)# ipv6 nd ra-lifetime 7`
`(config-if-e1000-3/1)# ipv6 nd preference high`
- RA lifetime needs to be greater than or equal to the interval -- you don't want the validity expiring before another one is sent!
- Here's an example of a box **listening** for RAs to use as default routers:
`(config)# interface ethernet 3/1`
`(config-if-e1000-3/1)# ipv6 nd ra-route`
`(config-if-e1000-3/1)# ipv6 nd router-reachable-time 4`
- router-reachable-time is the # sec to wait before using the next valid RA.
- Tie-breaker for multiple RAs of the same priority is "first one in wins".
- "show ipv6 ra-route" (as well as simple "show ipv6") will display useful information.



Example: RA-based route learning (Router1)

Intel(R) 82577LM Gigabit Network Connection (Microsoft's Packet Scheduler) [Wireshark 1.6.1 (SVN Rev 38096 from /trunk-1.6)]

Filter: icmpv6 Expression: Clear Apply

No.	Time	Source	Destination	Protocol	Length	Info
1	0.000000	fe80::21b:edff:fe05:9347	ff02::1	ICMPv6	118	Router Advertisement from 00:1b:ed:05:93:47
2	0.379562	fe80::21b:edff:fe05:9607	ff02::1	ICMPv6	118	Router Advertisement from 00:1b:ed:05:96:07
3	1.000269	fe80::21b:edff:fe05:9347	ff02::1	ICMPv6	118	Router Advertisement from 00:1b:ed:05:93:47
4	1.379558	fe80::21b:edff:fe05:9607	ff02::1	ICMPv6	118	Router Advertisement from 00:1b:ed:05:96:07
5	2.001879	fe80::21b:edff:fe05:9347	ff02::1	ICMPv6	118	Router Advertisement from 00:1b:ed:05:93:47
6	2.379068	fe80::21b:edff:fe05:9607	ff02::1	ICMPv6	118	Router Advertisement from 00:1b:ed:05:96:07
7	2.999832	fe80::21b:edff:fe05:9347	ff02::1	ICMPv6	118	Router Advertisement from 00:1b:ed:05:93:47
8	3.379299	fe80::21b:edff:fe05:9607	ff02::1	ICMPv6	118	Router Advertisement from 00:1b:ed:05:96:07
9	3.999786	fe80::21b:edff:fe05:9347	ff02::1	ICMPv6	118	Router Advertisement from 00:1b:ed:05:93:47
10	4.378961	fe80::21b:edff:fe05:9607	ff02::1	ICMPv6	118	Router Advertisement from 00:1b:ed:05:96:07
11	4.999970	fe80::21b:edff:fe05:9347	ff02::1	ICMPv6	118	Router Advertisement from 00:1b:ed:05:93:47
12	5.379529	fe80::21b:edff:fe05:9607	ff02::1	ICMPv6	118	Router Advertisement from 00:1b:ed:05:96:07
13	5.999826	fe80::21b:edff:fe05:9347	ff02::1	ICMPv6	118	Router Advertisement from 00:1b:ed:05:93:47
14	6.378964	fe80::21b:edff:fe05:9607	ff02::1	ICMPv6	118	Router Advertisement from 00:1b:ed:05:96:07
15	6.999665	fe80::21b:edff:fe05:9347	ff02::1	ICMPv6	118	Router Advertisement from 00:1b:ed:05:93:47

Frame 2: 118 bytes on wire (944 bits), 118 bytes captured (944 bits)

Ethernet II, Src: 00:1b:ed:05:96:07 (00:1b:ed:05:96:07), Dst: 33:33:00:00:00:01 (33:33:00:00:00:01)

Internet Protocol Version 6, Src: fe80::21b:edff:fe05:9607 (fe80::21b:edff:fe05:9607), Dst: ff02::1 (ff02::1)

0110 = Version: 6

.... 0000 0000 = Traffic class: 0x00000000

.... 0000 0000 0000 0000 0000 0000 = Flowlabel: 0x00000000

Payload length: 64

Next header: ICMPv6 (0x3a)

Hop limit: 255

Source: fe80::21b:edff:fe05:9607 (fe80::21b:edff:fe05:9607)

[Source SA MAC: 00:1b:ed:05:96:07 (00:1b:ed:05:96:07)]

Destination: ff02::1 (ff02::1)

Internet Control Message Protocol v6

Type: Router Advertisement (134)

Code: 0

Checksum: 0xe887 [correct]

Cur hop limit: 64

Flags: 0x08

0... .. = Managed address configuration: Not set

.0.. .. = Other configuration: Not set

..0. = Home Agent: Not set

...0 1... = Prf (Default Router Preference): High (1) ← Preference set to HIGH

.... .0.. = Proxy: Not set

.... ..0. = Reserved: 0

Router lifetime (s): 1800 ← Lifetime not tuned down yet

Reachable time (ms): 0

Retrans timer (ms): 0

ICMPv6 option (Source link-layer address : 00:1b:ed:05:96:07)

ICMPv6 option (MTU : 1500)

ICMPv6 option (Prefix information : 2001:db8:0:b::/64)

Ready to load or capture Packets: 206 Displayed: 204 Marked: 0 Dropped: 0 Profile: Default



Example: RA-based route learning (Router2)

The image shows a Wireshark packet capture of ICMPv6 Router Advertisement (RA) messages. The packet list at the top shows 15 packets, all of which are Router Advertisements from source fe80::21b:edff:fe05:9347 to destination ff02::1. The packet details pane for the selected packet (No. 3) shows the following structure:

- Ethernet II, Src: 00:1b:ed:05:93:47 (00:1b:ed:05:93:47), Dst: 33:33:00:00:00:01 (33:33:00:00:00:01)
- Internet Protocol version 6, Src: fe80::21b:edff:fe05:9347 (fe80::21b:edff:fe05:9347), Dst: ff02::1 (ff02::1)
- 0110 = Version: 6
- 0000 0000 = Traffic class: 0x00000000
- 0000 0000 0000 0000 = Flowlabel: 0x00000000
- Payload length: 64
- Next header: ICMPv6 (0x3a)
- Hop limit: 255
- Source: fe80::21b:edff:fe05:9347 (fe80::21b:edff:fe05:9347)
- [Source SA MAC: 00:1b:ed:05:93:47 (00:1b:ed:05:93:47)]
- Destination: ff02::1 (ff02::1)
- Internet Control Message Protocol v6
- Type: Router Advertisement (134)
- Code: 0
- Checksum: 0xee0f [correct]
- Cur hop limit: 64
- Flags: 0x00
- 0... .. = Managed address configuration: Not set
- .0.. .. = Other configuration: Not set
- ..0. = Home Agent: Not set
- ...0 = Prf (Default Router Preference): Medium (0) ← Preference set to MEDIUM
-0.. = Proxy: Not set
-0. = Reserved: 0
- Router lifetime (s): 1800
- Reachable time (ms): 0
- Retrans timer (ms): 0
- ICMPv6 option (Source link-layer address : 00:1b:ed:05:93:47)
- ICMPv6 option (MTU : 1500)
- ICMPv6 option (Prefix information : 2001:db8:0:b::/64)



Example: RA-based route learning

Showing two RAs received from upstream routers, then one is disabled

```
telnet@lab-ADX3#sho ipv6 int
Interface      Status      Routing   Global Unicast Address
Eth 8          up/up      fe80::21b:edff:fe3c:50c7
                2001:db8:0:b::c/64

telnet@lab-ADX3#sho ipv6 route
Type  IPv6 Prefix      Next Hop Router      Interface  Dis/Metric
D      ::/0             fe80::21b:edff:fe05:9607  e 8        254/1
C      2001:db8:0:b::/64      ::                      e 8        0/0

telnet@lab-ADX3#sho ipv6 ra-route
Best Default Router
Link Local Address      : fe80::21b:edff:fe05:9607
Router lifetime         : 4          ← configured ra-lifetime of 5 seconds (last was 1sec ago)
Interface               : 8
Preference               : HIGH

Backup Routers 1
Link Local Address      : fe80::21b:edff:fe05:9347
Router lifetime         : 4
Interface               : 8
Preference               : MEDIUM

<<<Disabled intf on box w/HIGH priority here>>>

telnet@lab-ADX3#sho ipv6 ra-route
Best Default Router
Link Local Address      : fe80::21b:edff:fe05:9347
Router lifetime         : 4
Interface               : 8
Preference               : MEDIUM

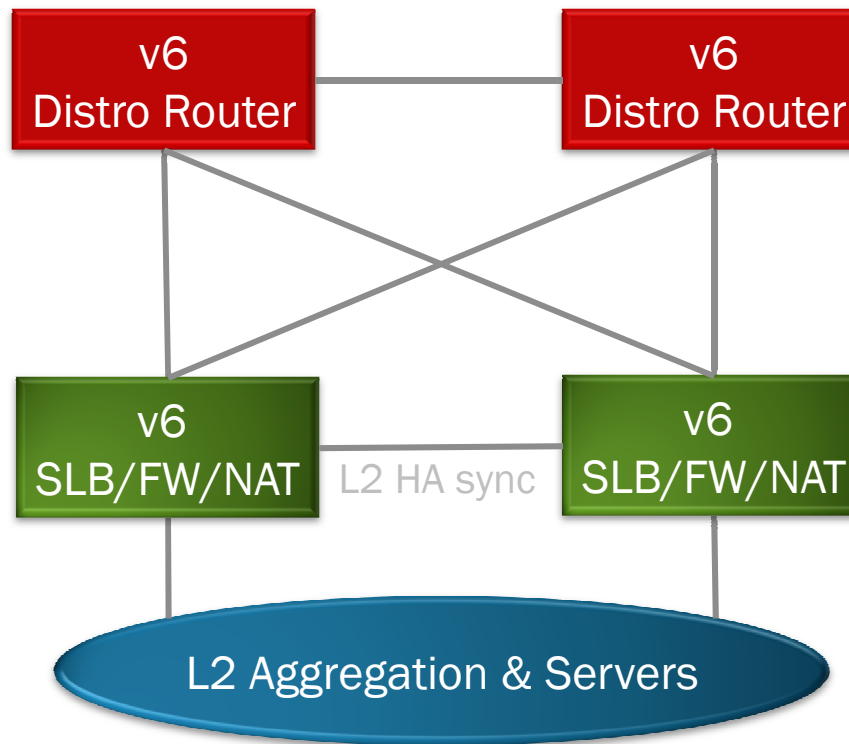
telnet@lab-ADX3#sho ipv6 route
Type  IPv6 Prefix      Next Hop Router      Interface  Dis/Metric
D      ::/0             fe80::21b:edff:fe05:9347  e 8        254/1
C      2001:db8:0:b::/64      ::                      e 8        0/0
```



ECMP – Scalable HA design

Solution: Equal Cost Multi-Path routing

- Each stateful device is connected to every router, such that any router has equal-cost paths through either device. Devices must operate in a synchronous, active-active mode.



ECMP overview

- “Standard” Routing decision: When the IPv6 route table contains more than one path to a given destination, the router selects the path with the lowest cost for insertion into the active routing table.
- ECMP decision: If more than one path with the lowest cost exists, all of these paths are inserted into the active routing table, up to the configured maximum number of paths.
- If the path selected by the device becomes unavailable, the IPv6 neighbor should change state and trigger the update of the destination in hardware.



ECMP overview

(continued)

- Which of the available paths to send any given packet is based on a simple hash of various L2/L3 header field. For example: XOR (S.MAC, S.IP, D.MAC, Flow label)
- Any routing protocol works as long as metrics/costs are equal for multiple paths
- ECMP is intrinsically stateless, so if passing through HA clusters of stateful devices (SLB, FW, NAT GW, etc.) then those devices must synchronize state tables in case the hashing algorithm splits related flows.
- Enabled by default on some platforms (including Brocade); can be disabled with "no ipv6 load-sharing".



NAT Technologies for the Enterprise



The “Chicken and Egg” deployment conundrum

“We can’t migrate all this content until there are IPv6 users to make use of it!”



CONTENT:

Datacenter, Content
Delivery Networks,
Managed Hosting, 'Net-
facing Enterprise Apps,
Media

“We can’t migrate all these users until there is IPv6 content to consume!”



ACCESS:

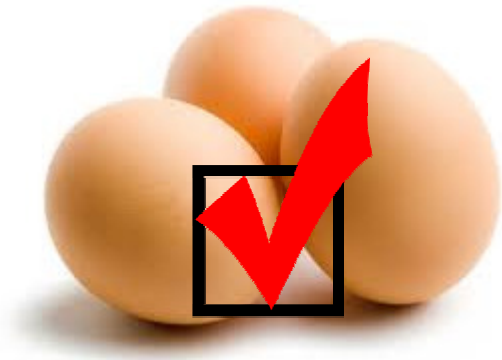
Broadband, Desktops,
Wholesale IP, Mobility,
Metro/access nets, “The
Eyeballs”

Egg Beats Chicken

Content is trivial to IPv6-enable – but IPv6 access lags for good reasons!

CONTENT:

Datacenter, Content Delivery Networks, Managed Hosting,
'Net-facing Enterprise Apps,
Media



Applicable tech:

Stateful NAT64
Stateless NAT64
SLB “6-6-4” and “4-4-6”

ACCESS:

Broadband, **Desktops**, Wholesale IP, Mobility, Metro/access networks, “The Eyeballs/\$”



Applicable tech:

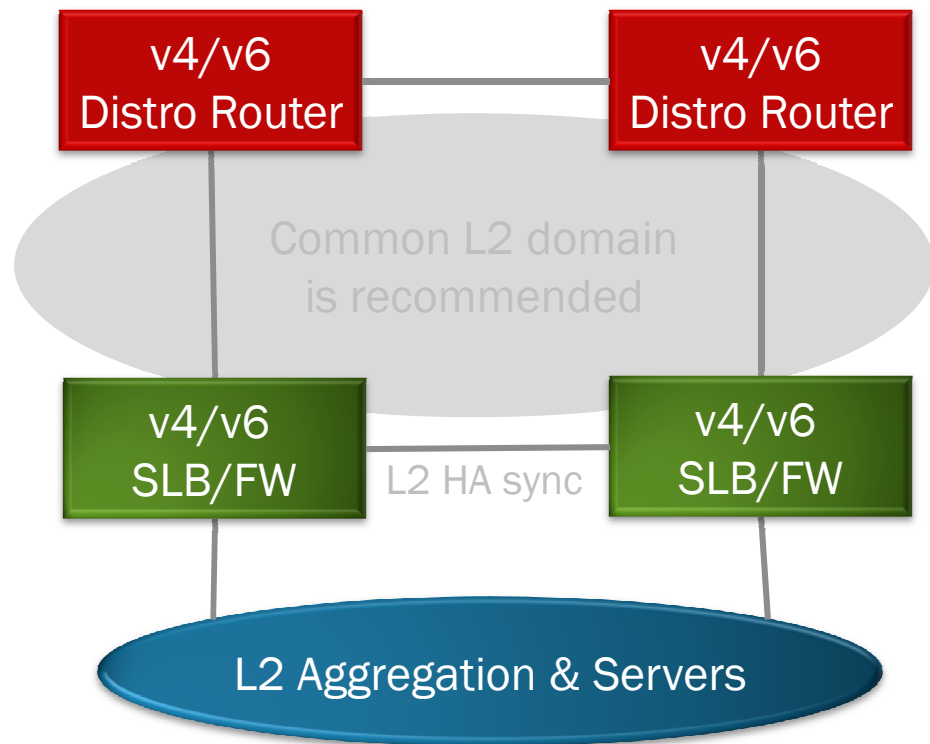
Stateful NAT64+DNS64
Stateless “NAT46”
...but dual-stack is the best path
for Enterprise Desktops



SLB & FW-based IPv6 deployments

“simple NAT” use-cases, combined with base function of device

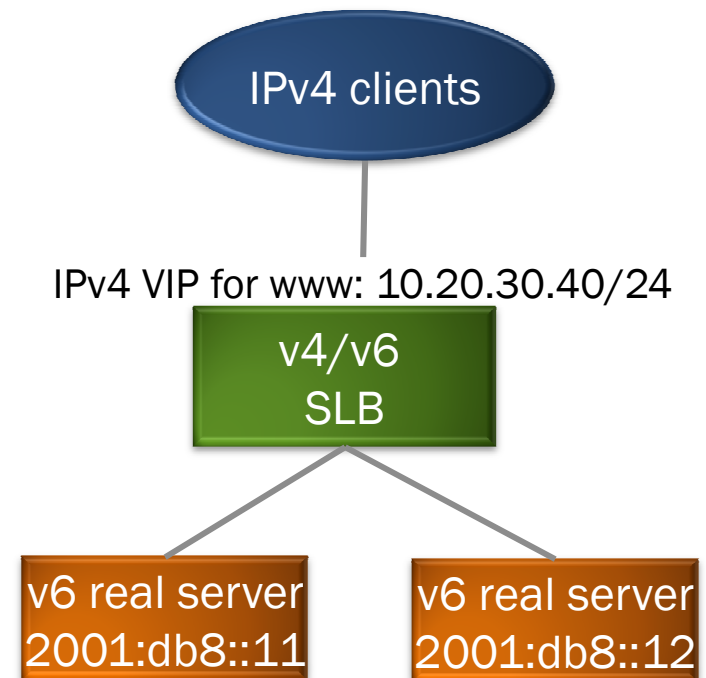
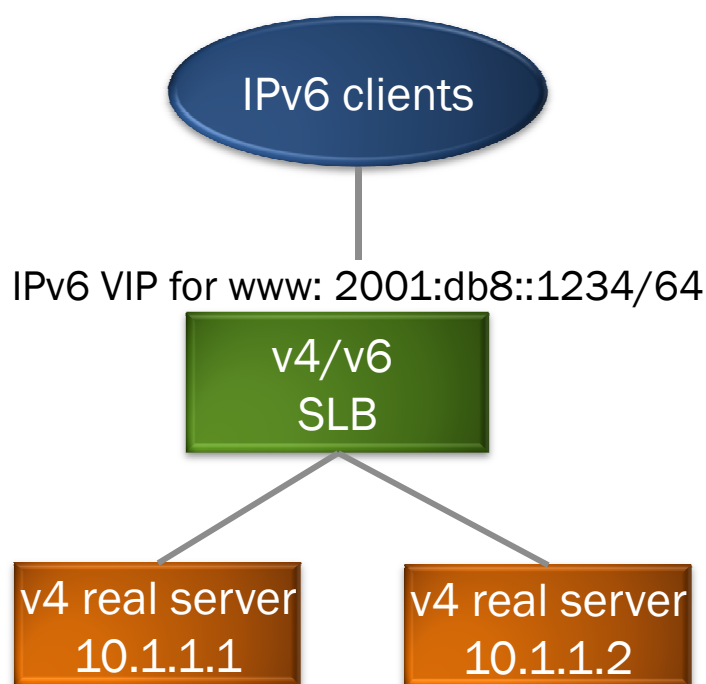
- These are typically trivial, since nothing changes about what functions these devices serve.
- The only common challenge is bringing IPv6 connectivity that deep into the network – SLBs and FWs usually live adjacent to the applications they support.
- Lesson Learned: Bring the firewall testing in AFTER the application testing is complete & v6-validated!



SLB-based IPv6 Transition Deployments

Observation from the field

- Proven to be a truly cost-effective transition strategy, since Operational processes barely change & little training is required.
- IPv6 performance/scalability requires testing if using same HW.



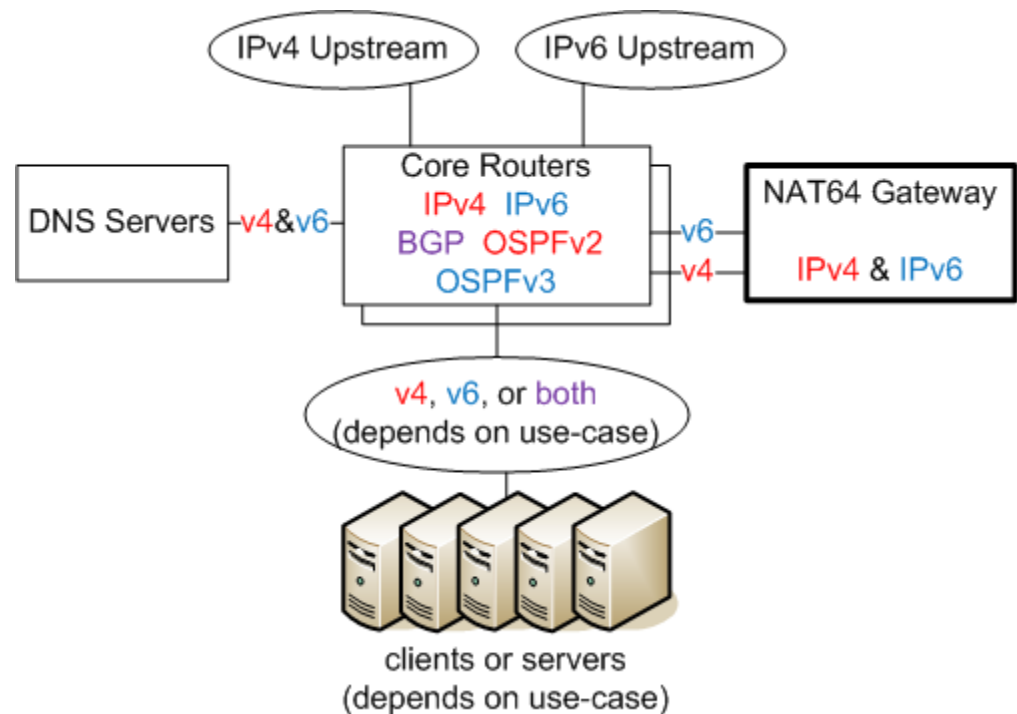
NAT Gateway Topology for Enterprise

In-line deployment is far too disruptive, and provides NO benefit

- **Use a Routed topology**

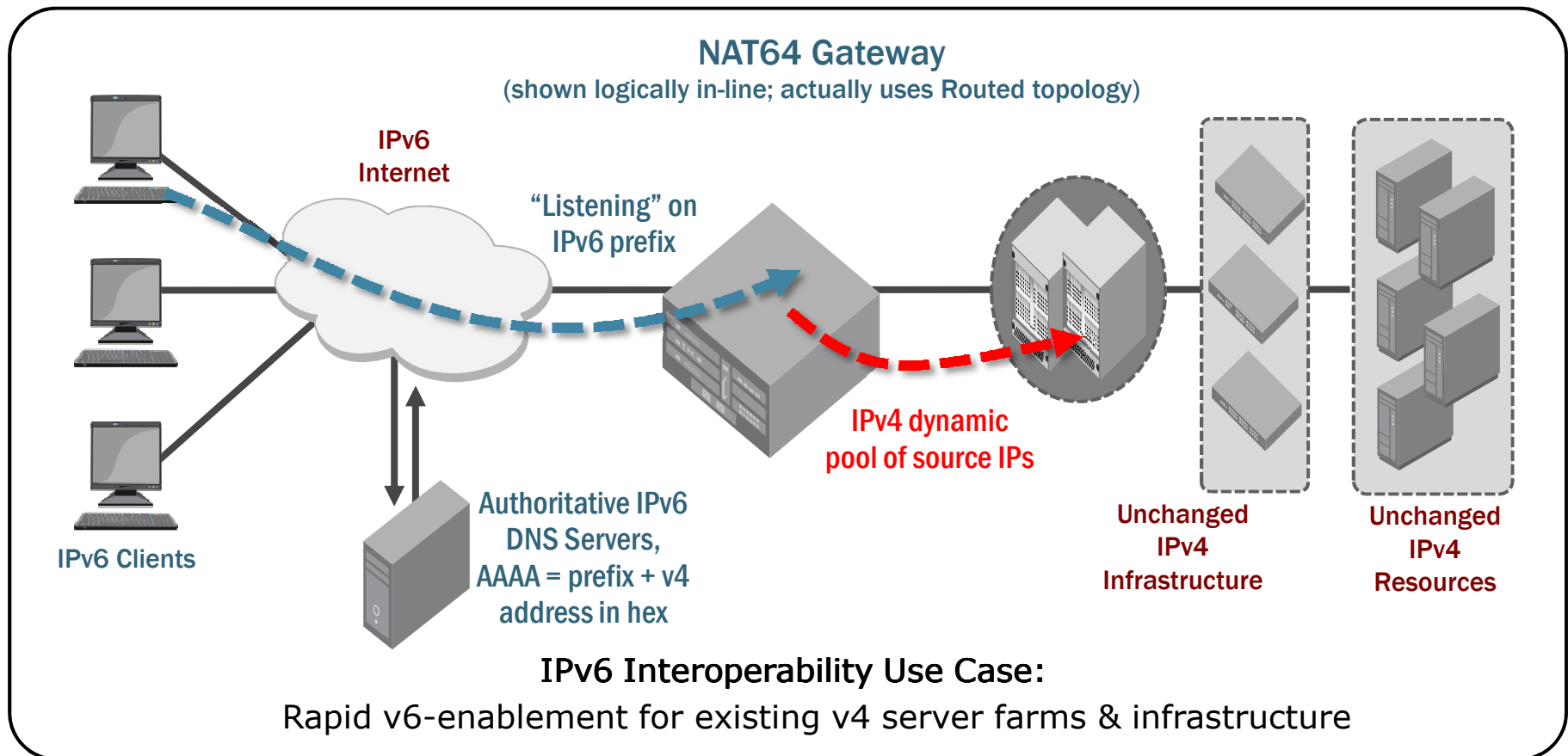
(simplified example diagram):

- Trivial to add to any network;
“Let the routing do the work for you” – instant prod-trial!
- Trivial to maintain & replace
- Only NAT traffic needs to go through it, and NAT traffic can be easily re-routed to another NAT GW for maint.
- Trivial to “scale horizontally”;
service & project owners can expand their own specific gateways



Stateful NAT64: Enterprise use-cases

This uses normal IPv6 DNS AAAA records



NAT64 Gateway config tips

- OSPFv2 & v3 is the most popular choice for dynamically routing the prefixes. But it's certainly okay to simply point static routes to the NAT64 Gateway from the upstream router, and use either FHRPs or RA learning on the GW itself.
- Tip for IPv4 NAT pool scaling: Reserve a /24 per site/pair, but actually configure and deploy less (such as $\frac{1}{2}$ or $\frac{1}{4}$ of the space).
- Same for IPv6 prefixes: Reserve a /64 any time you use less!
- As always, dual-stacking DNS servers to deliver normal AAAA records is an important (yet surprisingly simple) process. (Optionally upgrade your DNS infrastructure if appropriate.)





Thank You

