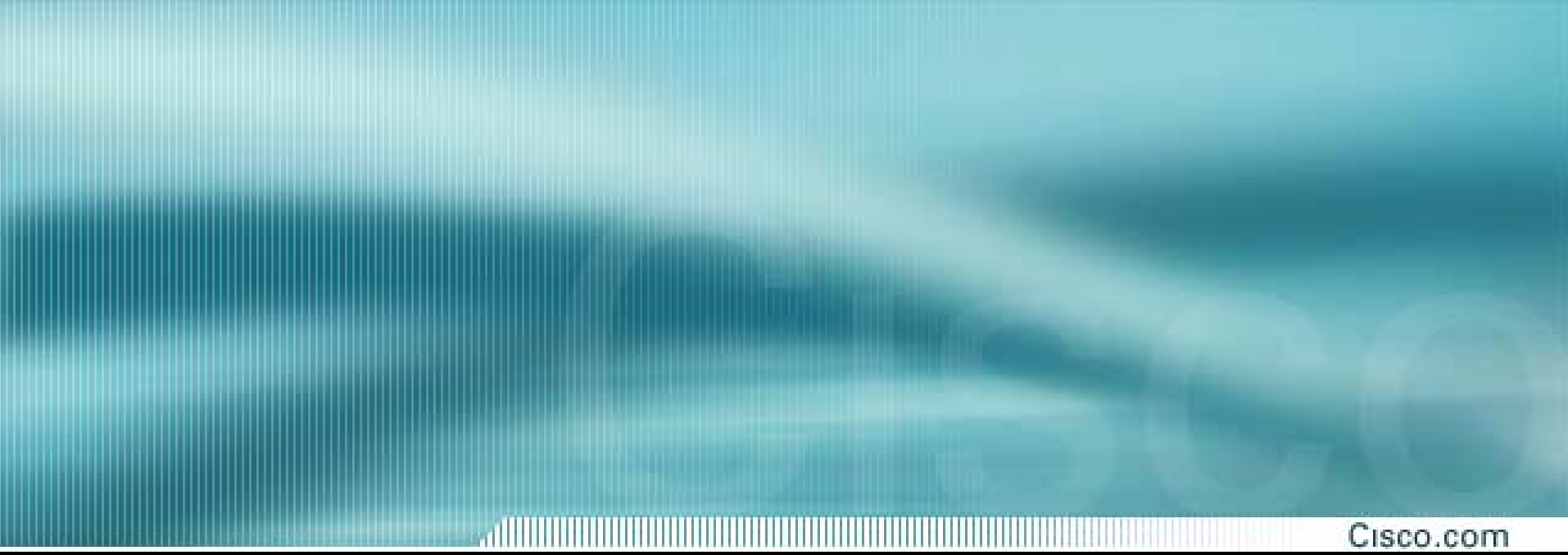


2547 L3VPN Control Plane Scaling

Apricot Kyoto Japan 2005

Robert Raszuk (raszuk@cisco.com)

Presented by

Mukhtiar Shaikh (mshaikh@cisco.com)

L3VPN Control plane scaling options

Cisco.com

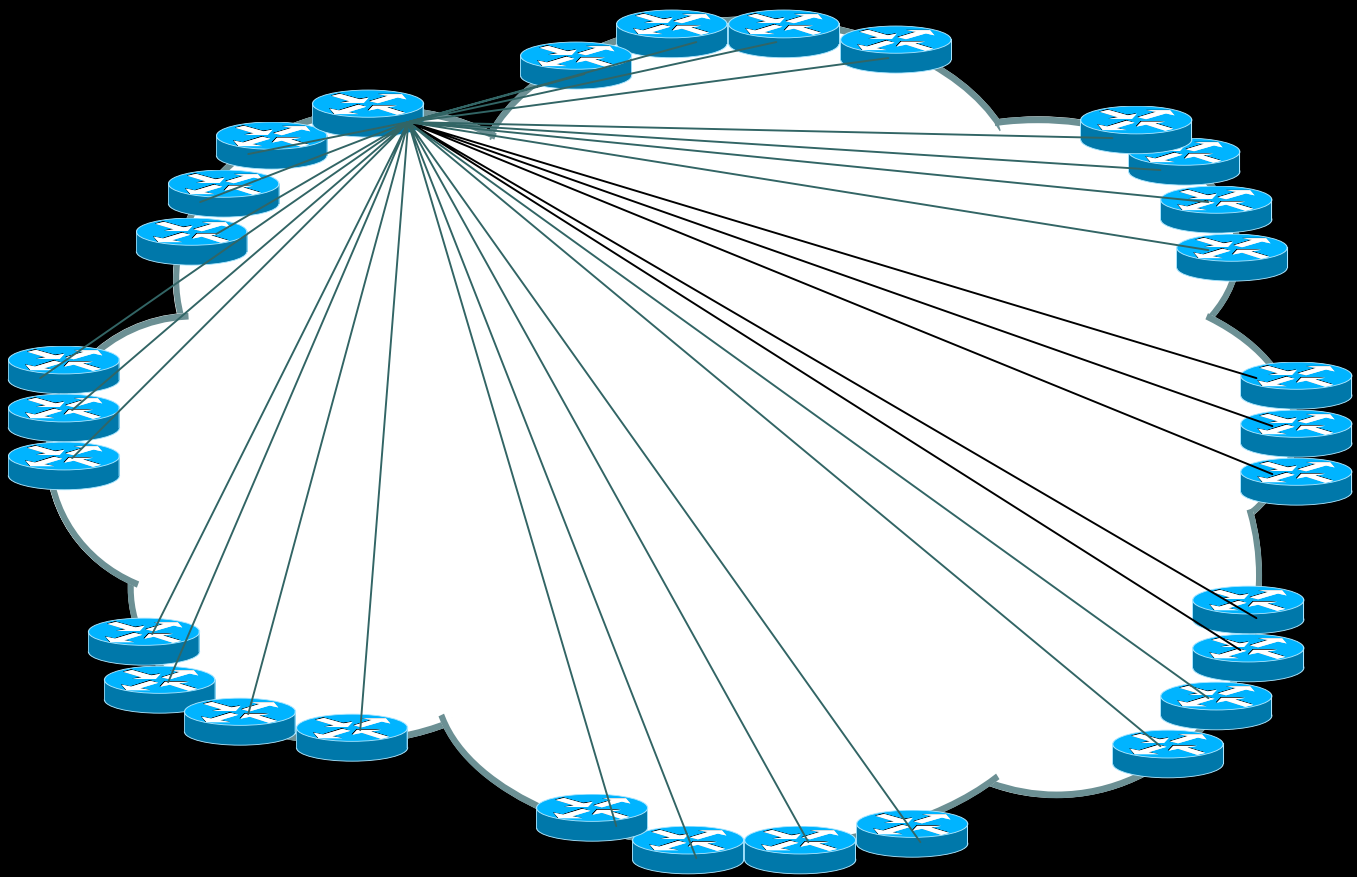
- Current L3VPNs intra-as distribution models
- Option #1 – Full mesh of PEs – Ultimate scaling
- Option #2 – CSC+ - Architectural scaling
- Option #3 – High capacity RRs – Forced
- Option #4 – VPNv4 filtering - semi-distributed
- Option #5 – VPNv4 filtering – geographic provisioning

Current L3VPNs intra-as distribution models

- With basic 2547 model SPs take all VPN customer routes and convert them into VPNv4 address space (by 8 byte of RD prepend)
- Dealing with this monolithic VPNv4 address bundle is becoming an issue during their distribution intra and inter-as and may only get worse with the L3VPN growth of demand
- The biggest L3VPN networks do reach up to 500K VPN routes today and still growing. Long term requirements reach today 5-10M VPN routes for largest providers.

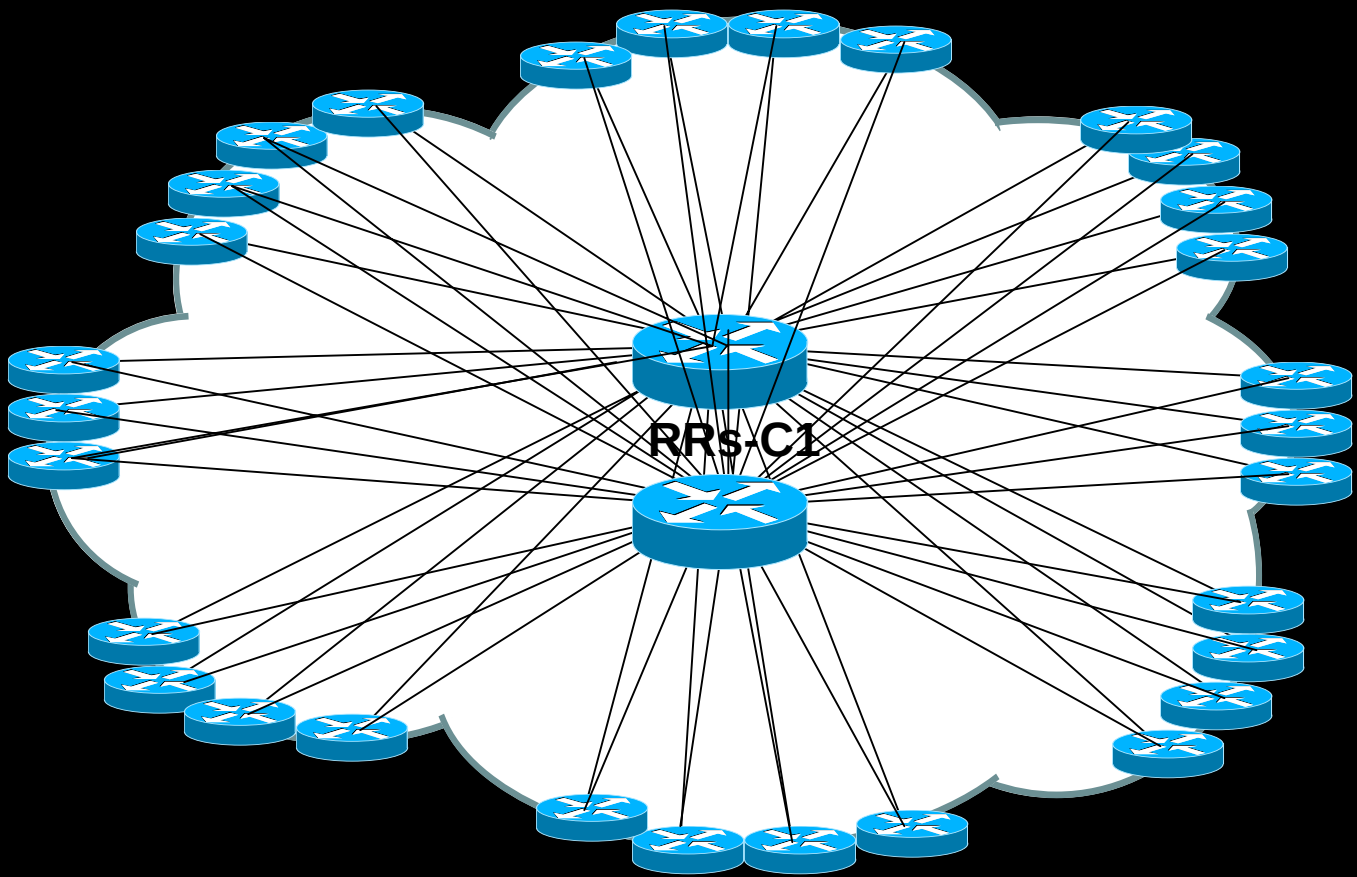
Current L3VPNs intra-as distribution models

- Today's topology models – Full Mesh
- Sessions for single PE shown



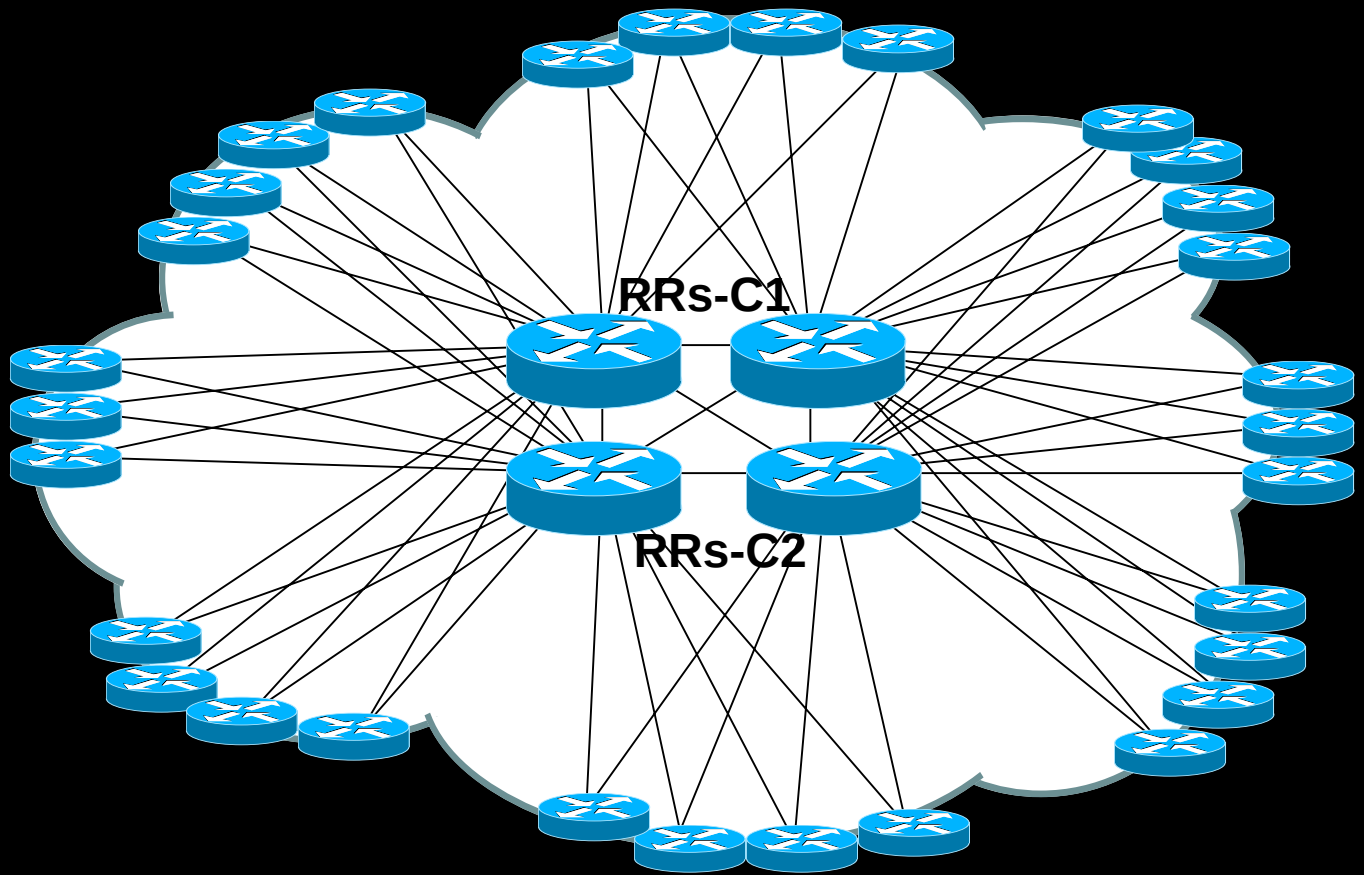
Current L3VPNs intra-as distribution models

- Today's topology models – single RR cluster



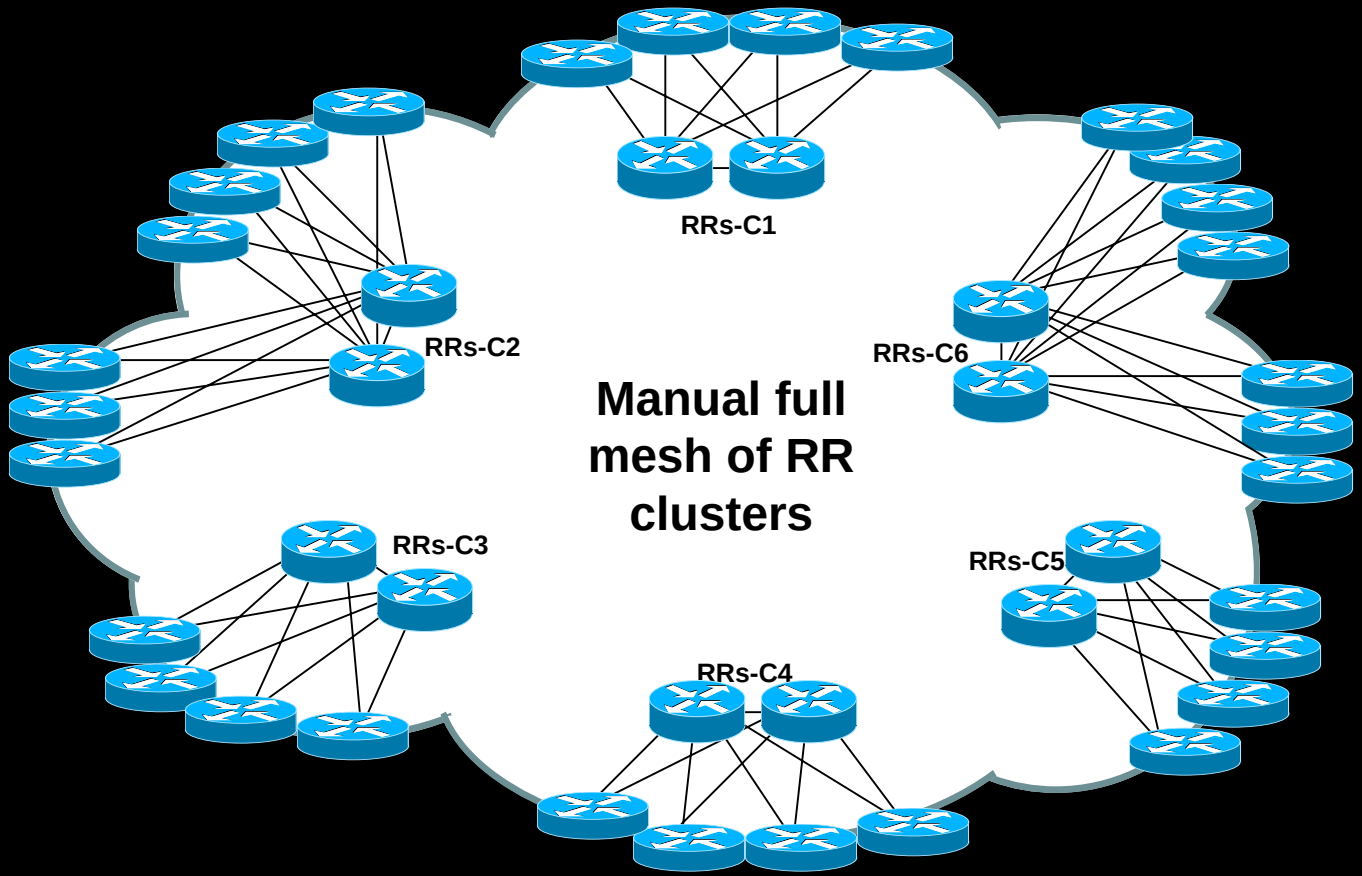
Current L3VPNs intra-as distribution models

- Today's topology models – dual RR cluster



Current L3VPNs intra-as distribution models

- Today's topology models – multi RR cluster



L3VPN Control plane scaling options

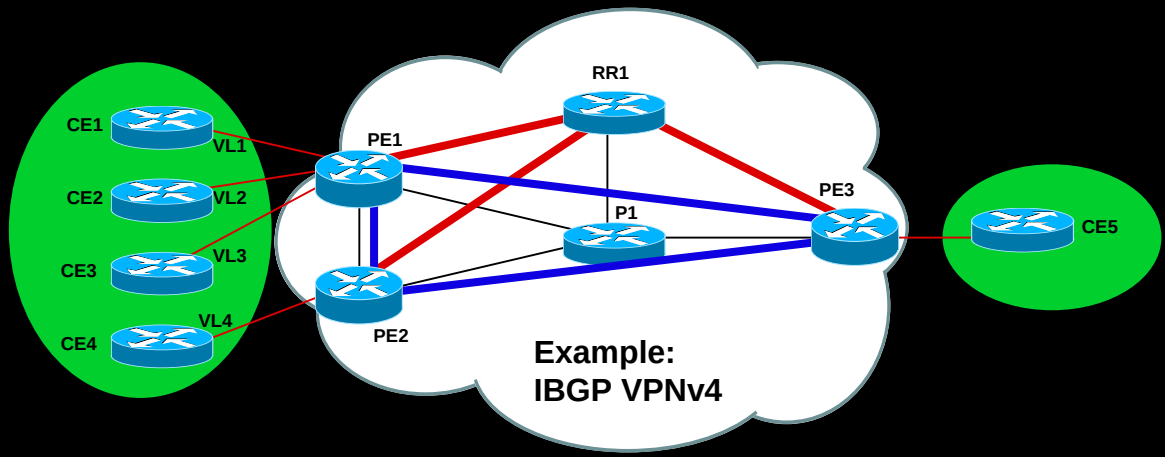
- Current L3VPNs intra-as distribution models
- Option #1 – Full mesh of PEs – Ultimate scaling
- Option #2 – CSC+ - Architectural scaling
- Option #3 – High capacity RRs – Forced
- Option #4 – VPNv4 filtering - semi-distributed
- Option #5 – VPNv4 filtering – geographic provisioning

Option #1 – Full mesh of PEs – Ultimate scaling

- Automation of BGP PE-PE peering via BGP Build in Peer Discovery removes the need to manually configure full mesh of PEs.
- Elimination of RRs removes the network devices bottle necks in control plane route distribution
- Amount of information send & recieved by PEs can be balanced by RT based PE-PE filter list propagation & dynamic update groups.

Option #1 – Full mesh of PEs – Ultimate scaling

- With recommended different RD per vrf model no information reduction due to best path run on RRs
- Single RR cluster can carry and handle all peer discovery information
- Full PE mesh made posible with Automated BGP Build in Peer Discovery:



Option #1 – Full mesh of PEs – Ultimate scaling

- Transition to full mesh does not require RD renumbering when (with the same RD for all VPN sites) IBGP multipath is required.
- Reduced “native” convergence as well as possibility to use created IBGP mesh for Virtual Links ID propagation from Accelerated BGP Convergence (goal sub-second BGP convergence).

L3VPN Control plane scaling options

Cisco.com

- Current L3VPNs intra-as distribution models
- Option #1 – Full mesh of PEs – Ultimate scaling
- Option #2 – CSC+ - Architectural scaling
- Option #3 – High capacity RRs – Forced
- Option #4 – VPNv4 filtering - semi-distributed
- Option #5 – VPNv4 filtering – geographic provisioning

Option #2 – CSC+ - Architectural scaling

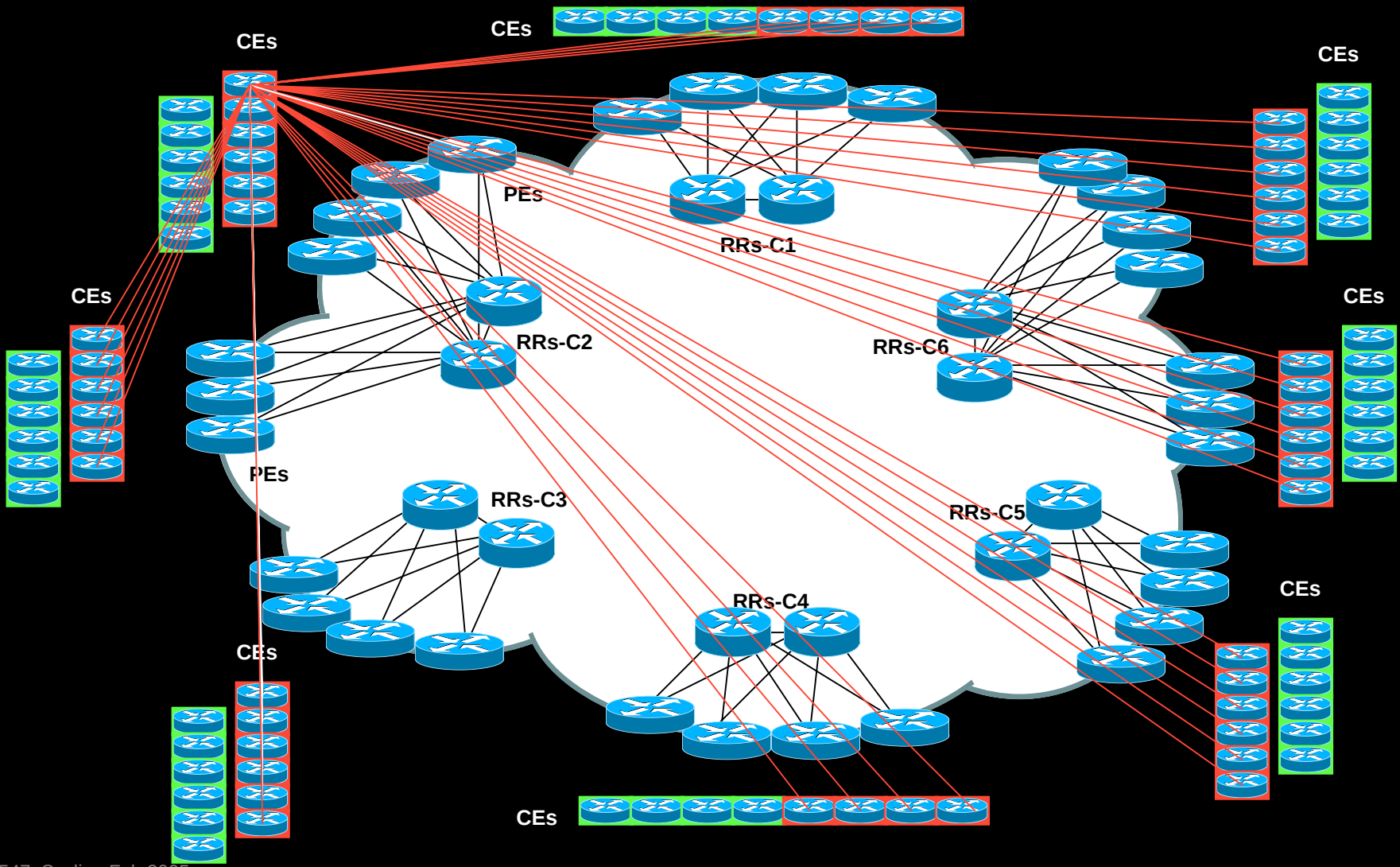
- **Basic 2547 model of operaton:**

- Customer routes are accepted into PEs
- PEs apply RT based export policies & RDs which turn per VPN IPv4 customer separated routes into one big block of “provider owned” vpnv4 routes
- Provider takes responsibility of propagating “combined” customer routes between his PEs
- In order to scale the propagation an significant effort is made to again separate VPN routes into groups
- PEs do the inbound filtering accepting only routes which match import RTs

Option #2 – CSC+ - Architectural scaling

- The alternative is classic Carrier's Carrier model:
- No IP lookup in the VRFs – only label switching at PEs with CE running MPLS encapsulation to PE
- No L3 Routing information from customers at PEs both control & forwarding plane (except next hop information)
- Automated basic 2547 full mesh of customer site interconnectivity without per site state (the case with L2 tunnels)
- „Out of band" L3 routing information exchange between customer sites

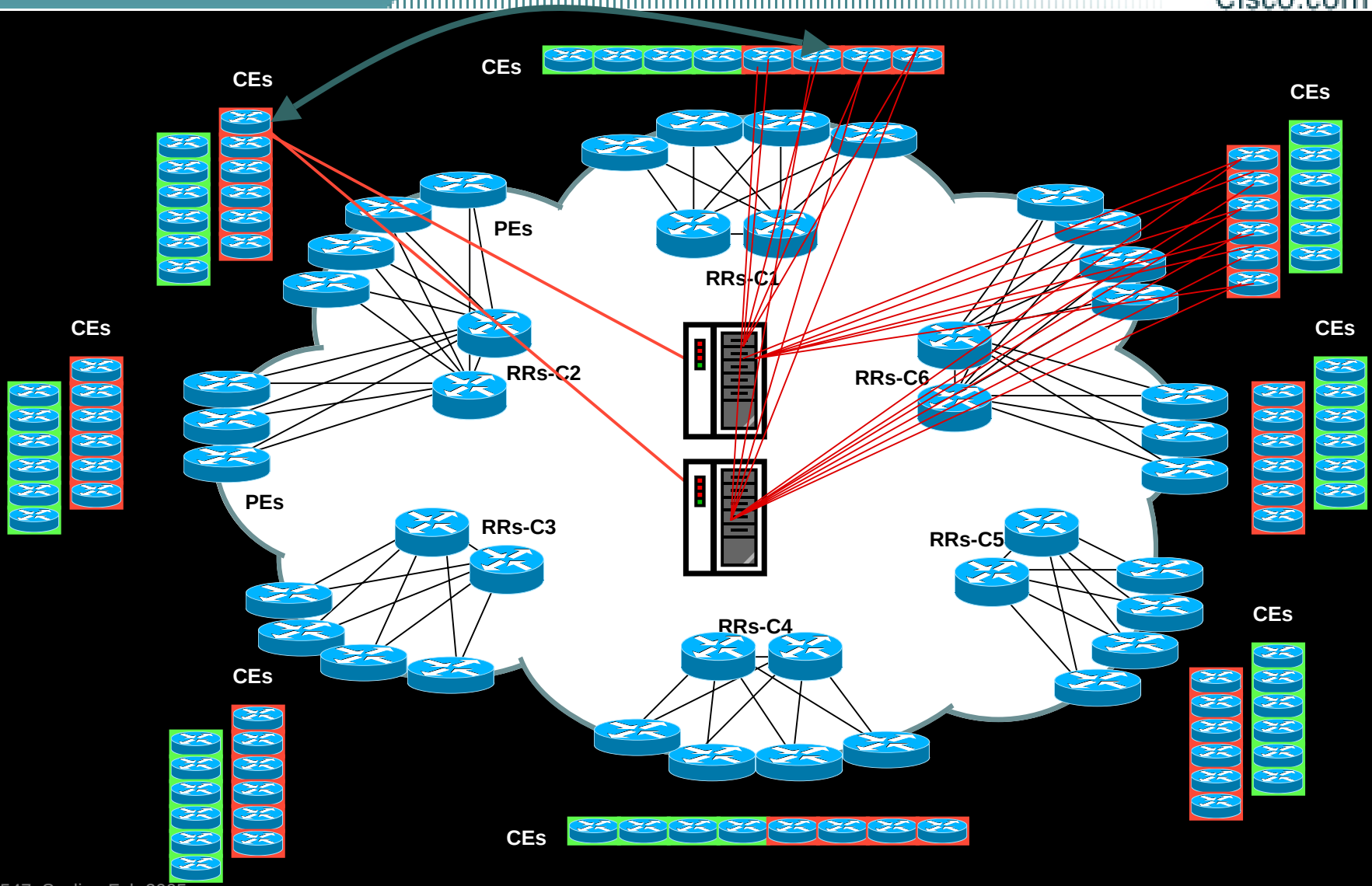
Option #2 – Classic CSC model:



Option #2 – CSC+ - Architectural scaling

- The main issue with classic CSC is the requirement for customer to exchange their routing themselves. Some customer find their own route controling a benefit.
- For those who do not and for L3VPN CE managed services CSC+ can offer a route server model for propagating customer routes via logical independent partitions.
- No need to partition VPNv4 routes as the routes are never converted to VPNv4 !!! They remain IPv4 only.

Option #2 – CSC+ - Architectural scaling



Option #2 – CSC+ - Architectural scaling

- Customer next hop information are still propagated via existing vpnv4 RRs
- Extranets can be supported by selective import/export of IPv4 routes between Route Server partitions
- For the simplest Route Server with logical partitions one can use any router + VRF-Lite
- Possible further work to eliminate requirement for two BGP sessions from CEs

Option #2 – CSC+ - Architectural scaling

- The logical route server partitions are reachable within customer VPN space
- They could be managed by customers or by providers
- Customers can utilize those for new additional value add services DNS, DHCP, route monitoring etc ...
- As routes are partitioned requirement to go to 64bit based memory access model in the route server code is no longer a scaling necessity

L3VPN Control plane scaling options

- Current L3VPNs intra-as distribution models
- Option #1 – Full mesh of PEs – Ultimate scaling
- Option #2 – CSC+ - Architectural scaling
- Option #3 – High capacity RRs – Forced
- Option #4 – VPNv4 filtering - semi-distributed
- Option #5 – VPNv4 filtering – geographic provisioning

Option #3 – High capacity RRs – Forced

- The approach to just increase capacity of route reflectors
- Long term requires 64 bit based memory access (> 2GB address space requirement)
- Not clear if this is sufficient approach for long term L3VPN customer route distribution
- Lack of architectural approach may cause problem shift model from RRs to PEs (Option #2 avoids this)

L3VPN Control plane scaling options

Cisco.com

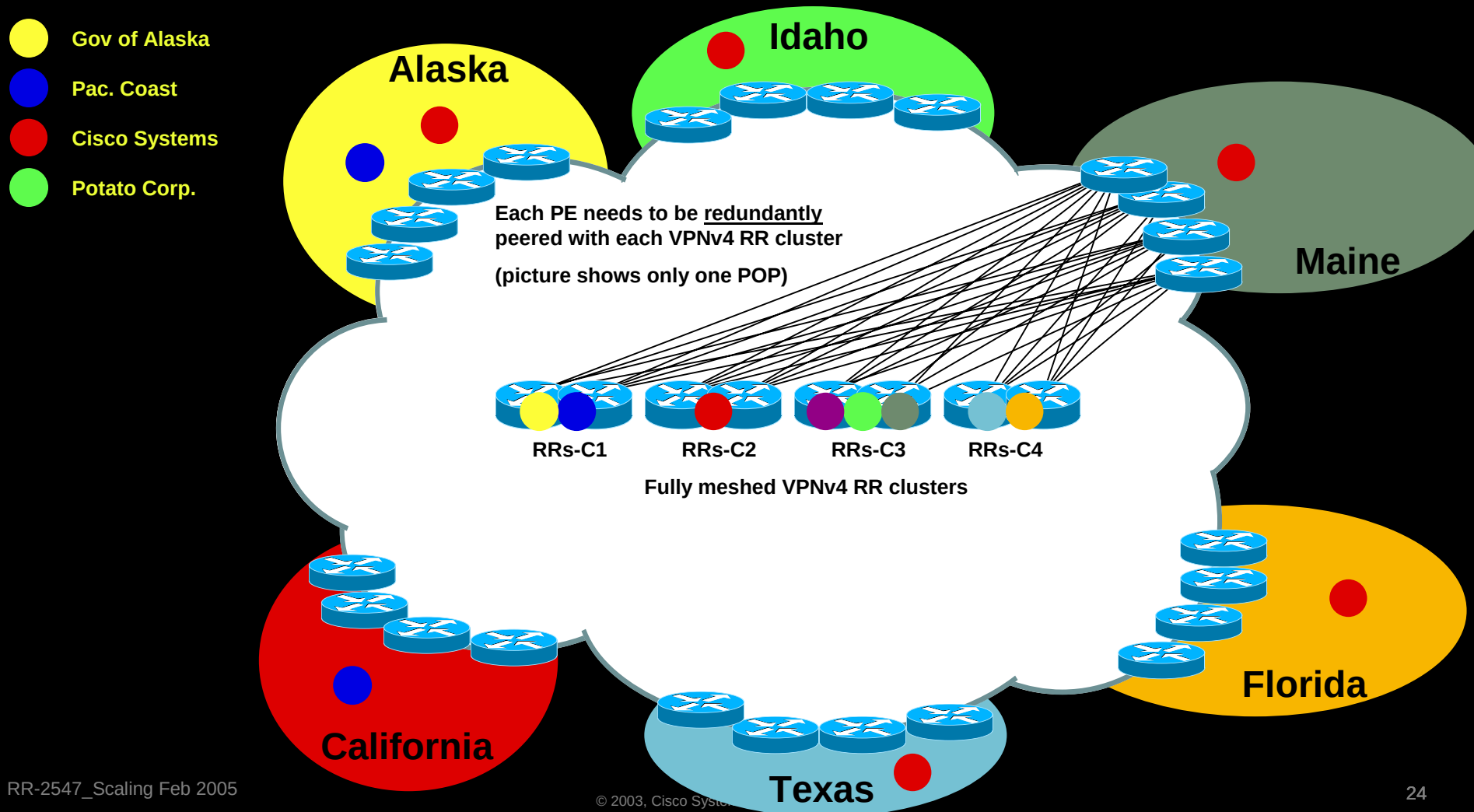
- Current L3VPNs intra-as distribution models
- Option #1 – Full mesh of PEs – Ultimate scaling
- Option #2 – CSC+ - Architectural scaling
- Option #3 – High capacity RRs – Forced
- Option #4 – VPNv4 filtering - semi-distributed
- Option #5 – VPNv4 filtering – geographic provisioning

Option #4 – VPNv4 filtering - semi-distributed

- Relies on partitioning VPNv4 routes into multiple distribution groups selected by RT extended community and reflecting those by multiple VPNv4 RR clusters (rr-group command)
- Requires connecting each PE to all VPNv4 RR clusters !
- Uses route target filtering propagation draft-ietf-l3vpn-rt-constrain or ext community ORF for RT ext community propagation

Option #4 – VPNv4 filtering - semi-distributed

Only VPN routes required are preset in corresponding RR clusters :



Option #4 – VPNv4 filtering - semi-distributed

- Can become difficult to maintain due to number of RR clusters required to get any signifcant benefit and manual PE – RR clusters meshing
- Due to the VPNv4 routes carrying more then one RT ext community it may be still very difficult to enforce the fact that given vpnv4 route is stored only on one route reflector cluster
- Good corelation of rr-groups (RT ranges) allocation with the provisioning system is a must for good effects

L3VPN Control plane scaling options

Cisco.com

- Current L3VPNs intra-as distribution models
- Option #1 – Full mesh of PEs – Ultimate scaling
- Option #2 – CSC+ - Architectural scaling
- Option #3 – High capacity RRs – Forced
- Option #4 – VPNv4 filtering - semi-distributed
- Option #5 – VPNv4 filtering – geographic provisioning

Option #5 – VPNv4 filtering – geographic provisioning

- Similar model to Option #4
- Does not require connecting PEs to all RR VPNv4 clusters
- Does not require configuration of RT groups on RRs
- Relies on correct geographic match between RR clusters and VPN site membership
- Uses rt filtering (rt-constrain) propagation between vpnv4 RR clusters

Option #5 – VPNv4 filtering – geographic provisioning

Only VPN routes required are preset in localized RRs clusters: