Optimal Route Reflection

Technology and Deployment

Phil Bedard – TME @Cisco

NANOG70

6/6/2017

CISCO

Quick BGP/RR Refresher

- RR is a widely deployed IBGP network scaling solution, RFC 4456
- Prefixes received from client reflected to client and non-client peers
- RRs use the same best-path selection process as normal BGP speakers
- IBGP best path selection (generalized, varies by vendor)
  1. LOCAL_PREF
  2. AS_PATH length
  3. ORIGIN
  4. MED
  5. **Lowest IGP metric to reach resolved next-hop address**

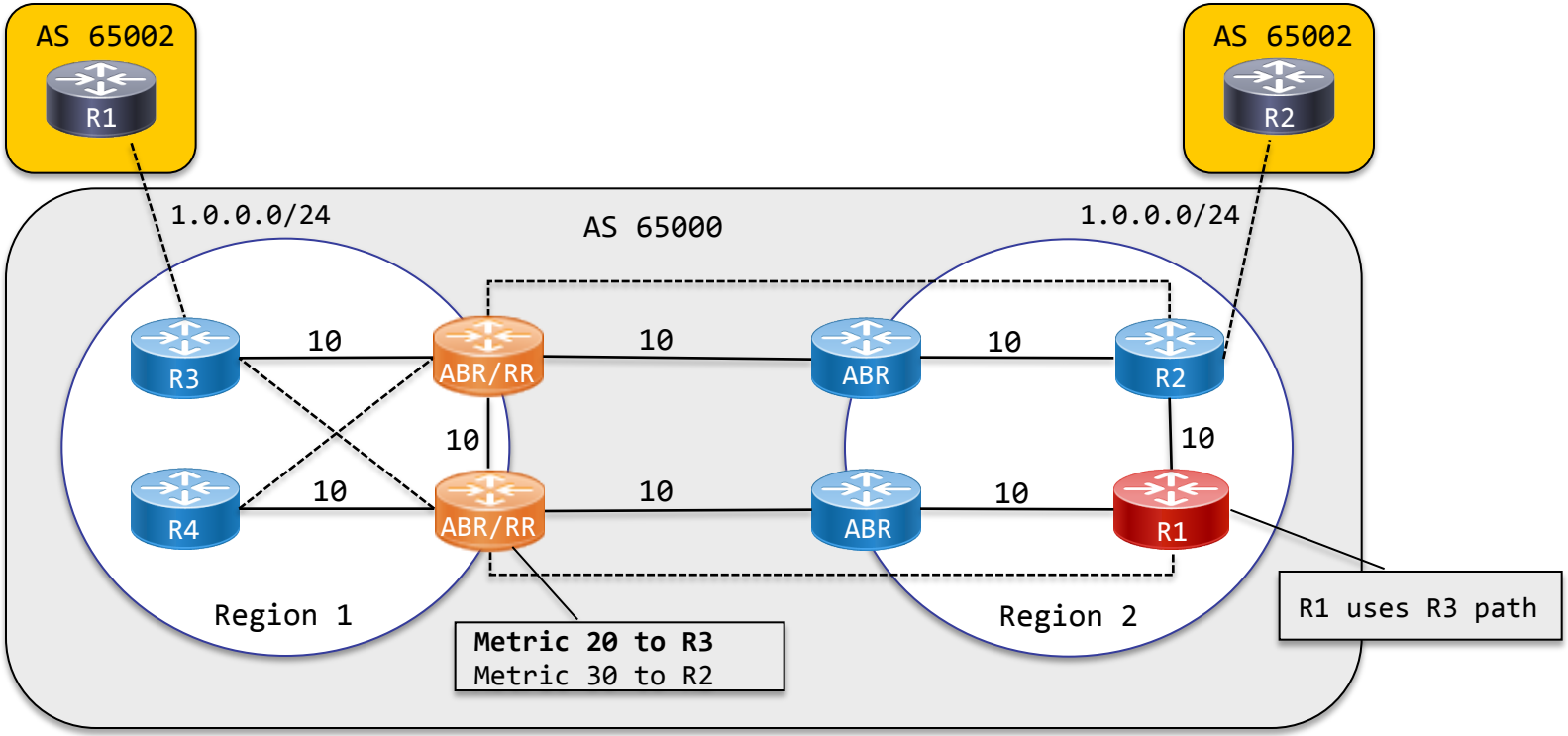
  …

  n. Router ID

- 5 is very typical in IBGP, ensures exit point is the lowest metric cost
  - Commonly referred to as hot potato routing

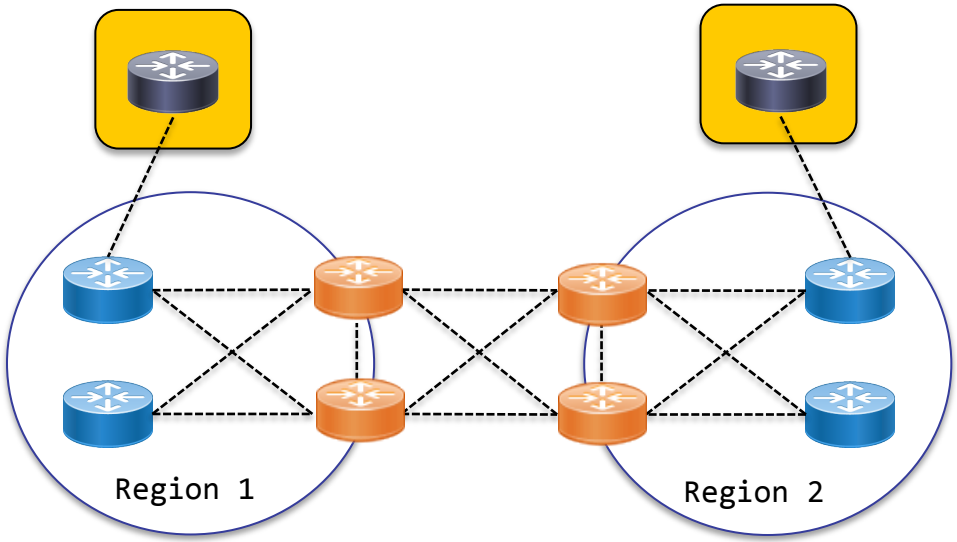
Problem: Suboptimal RR Path Reflection

Caused when RR is not topologically close to clients



Suboptimal RR Path Reflection – Non-ORR Solutions

- Hierarchical RRs
  - RRs in close proximity to clients
  - Widely deployed solution
  - Limits RR deployment location
- Advertise multiple paths via RR
  - Diverse-path / Shadow RR
  - Add-path
- Internet in a VRF w/unique RDs
- Solution for network proximity issue:
  - Tunnels (GRE/IPIP)



Additional Motivating Factors

Centralized RR

- Decouple RIB/FIB, FIB programming on in-path RR can cause issues
- Simplify network forwarding path routers by removing RR function
- Convergence improvements by eliminating chain of RRs
- Reduce number of policy control points
- Centralized RR has been widely deployed for VPN NLRI where RD per PE solves the single route selection / optimality issue

vRR on x86

- Route reflection is a software control-plane function
- Stable, starting to see more widespread deployment
- Scale, speed, automation. XRv-9000 vRR has a RIB capacity of 70M prefixes
- Flexible deployment options, IE: RR per service
- Feature velocity

What is BGP Optimal Route Reflection?

draft-ietf-idr-bgp-optimal-route-reflection, introduced in 2011

Current IETF draft covers two types of ORR

1. Optimal BGP path selection based on client IGP perspective
    - Purely a software solution, no BGP protocol changes
2. Optimal BGP path selection based on policy, such as choosing a specific network exit point for a client based on traffic engineering data

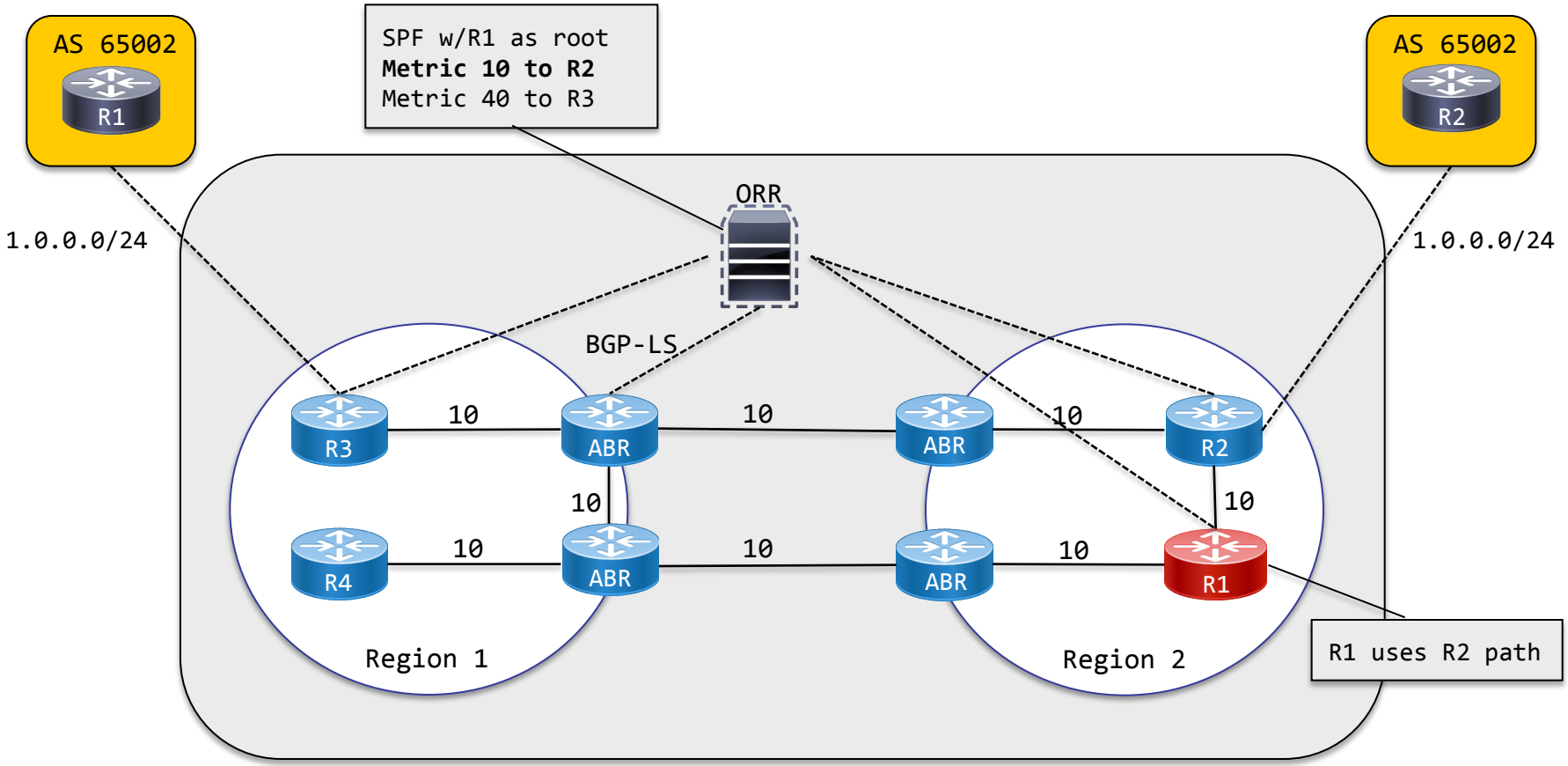
This talk is mostly about option 1. Option 2 can be combined with 1 and is very loosely defined and leaves open the option arbitrary policy.

No requirement client use the path, local BGP paths or policy could override advertisement.

How does BGP-ORR work?

- Disclaimer: Cisco implementation
- Topology data acquired via IS-IS, OSPF, or BGP-LS
- ORRSPF process
  - Extracts Node ID information from IGP topology
  - Runs SPF to each node with client as root
  - Store IGP costs in a ORR group RIB for use during BGP best path selection
- BGP best path selection references group-specific NH table during IGP step
  - Route with lowest NH metric is chosen and advertised to client
- ORRSPF is independent of traditional BGP code, only runs if ORR is enabled
- ORR is fully compatible with other enhancements like add-path

How does BGP ORR work?



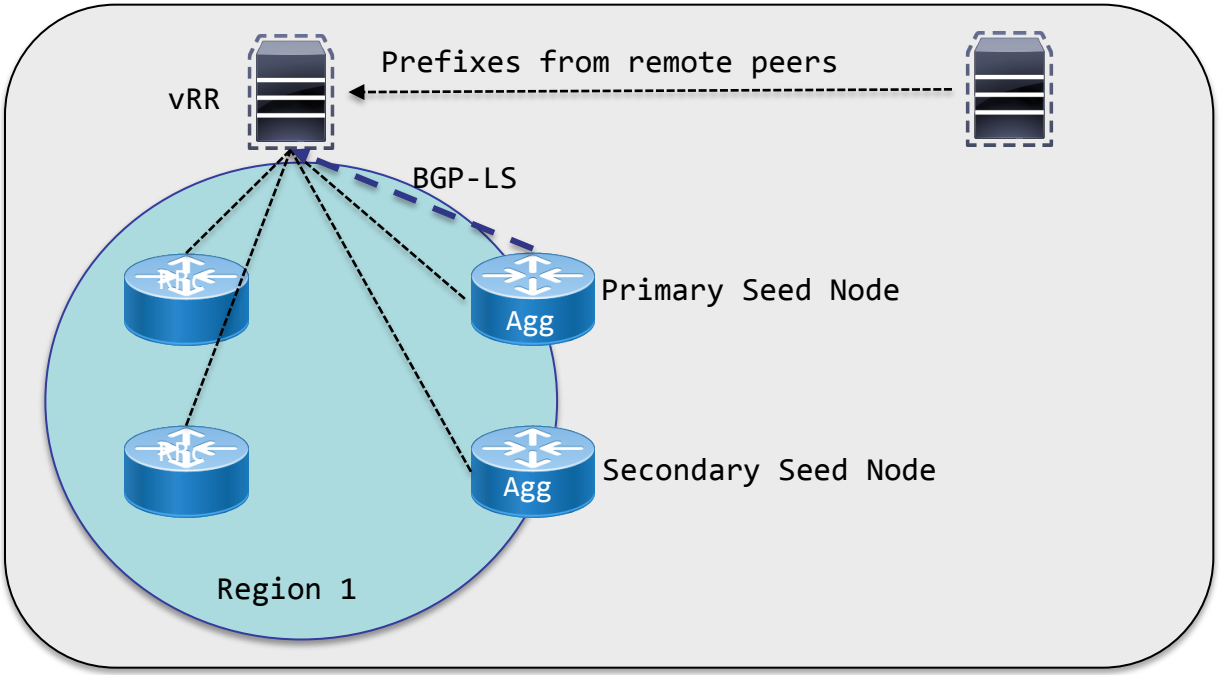
How does BGP-ORR work?

I wrote a simple ORR implementation built on ExaBGP to illustrate flow

- Two ExaBGP instances, one for ingesting external routes/topology and one to announce optimized routes to clients
- Topology data is ingested using BGP-LS, decoded into JSON via ExaBGP, and fed to external application
- Routes from peers are also fed into application, associated with NH node IPs
- Info is stored ephemerally and available via REST interface
- Client ExaBGP application listens for new clients, queries first application to acquire optimized route table, and advertises routes to specific client via ExaBGP

ORR Group Optimization

- Nodes are grouped based on seed IP addresses
- Seed IP is used as root node for SPF calculations for all nodes in group, does not need to be BGP peer (IE: if ABR does not speak BGP)
- IGP metrics to nodes behind agg routers are not important for reaching remote prefixes



Current known vendor implementations

- Cisco
  - Introduced IOS-XR 5.3.1
  - IS-IS, OSPF, BGP-LS
- Juniper
  - Introduced Junos 15.1
  - IS-IS, OSPF
- Nokia
  - SR OS 15.0.R1
  - OSPF, IS-IS, BGP-LS

Vendor configuration

Junos OS

```
[edit protocols bgp]
group peer-group {
           optimal-route-reflection
           igp-primary address
           igp-backup address
      }
}

show igp bgp-orr
```

SR OS

```
config router bgp
  optimal-route-reflection
    location location-id
      primary-ip-address address
      secondary-ip-address address
      tertiary-ip-address address

  group peer-group
    cluster cluster-id orr-location location-id
    allow-local-fallback if no reachable paths

show router bgp optimal-route-reflection
```

IOS-XR

```
router bgp asn
   address-family ipv4 unicast
      optimal-route-reflection orr-group pri sec ter

router bgp asn
   neighbor neighbor-ip
      address-family ipv4 unicast
         optimal-route-reflection orr-group

show orrspf database orr-group
```

Current ORR Limitations (not exhaustive)

- No dynamic peer grouping
  - All vendor solutions require static seed nodes
  - If you want specific ORR for every client, results in many groups
- Potential scale limitations
  - Implementation requires creating a new IGP RIB for each ORR client group
  - 10k node flat IGP with a number of unique groups is likely problematic
  - BGP NLRI and next-hop changes trigger ORR SPF calculations
- Synchronization Issues
  - IGP/BGP-LS convergence needs to take place before ORR can advertise routes
- Hierarchical IGP
  - SPF across area boundaries is not well supported
  - Not a typical issue in many networks, the upstream ABRs would be the root for the SPF calculations, NH is rewritten to ABR
  - More of an issue when we look at solutions like EPE

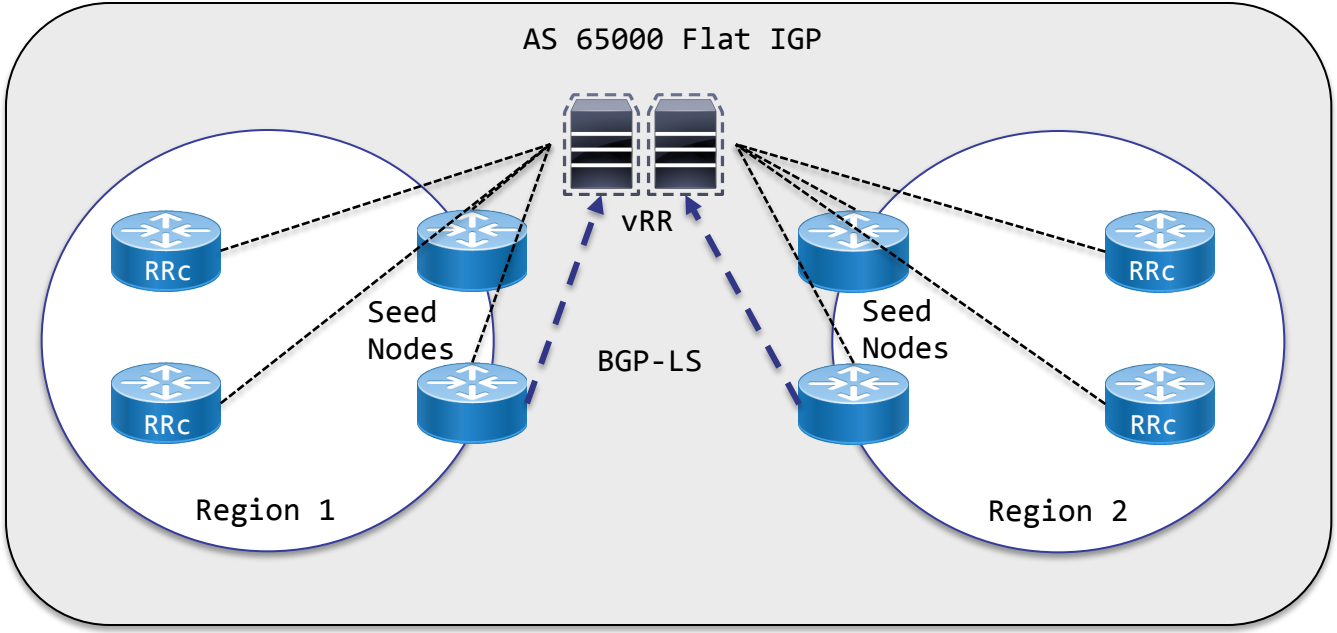
How do I deploy ORR in my network?

- Recommend Virtual Route Reflectors
  - Can be provisioned anywhere x86 compute exists
  - ORR enables geo-redundancy
  - Similar redundancy guidelines to current RRs such as dual-homing clients
  - vRR deployment does not require SW upgrades on client routers

- Topology Acquisition
  - Recommend BGP-LS if available, may require SW upgrade
  - Can run a virtual router running BGP-LS with adj to network, advertise to vRR
  - IS-IS or OSPF adjacencies can be used via tunnels or direct adjacencies to vRR if necessary

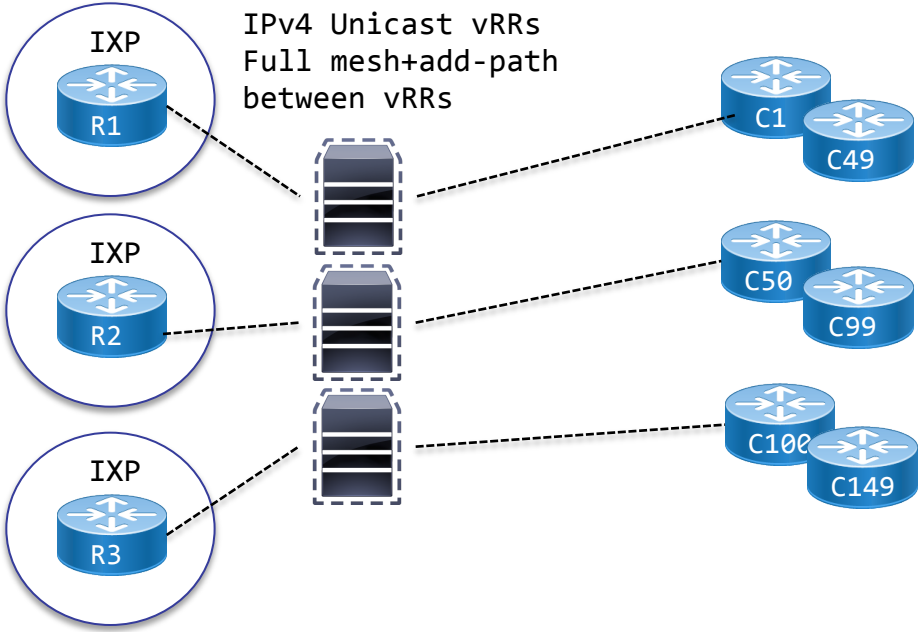
- Client Migration
  - Same process as migration to RRs, well documented

ORR Deployment – Simple Example



- BGP-LS is used to acquire topology data, vRR out of forwarding path
- IGP adjacency to network could also be used
- Use add-path to add resiliency and faster re-convergence

ORR Deployment – Cluster Scaling



- IXP routers are clients. IXP router peers with all vRRs, use full mesh+add-path between vRRs, or use an intermediate route-server to feed vRRs
- On x86 vRRs can scale to hundreds of ORR clients*
- Easy to spin up additional vRRs

ORR Future

- Enhancements to improve performance and scale
- Dynamic grouping of common ORR peers via topology inspection
- Extend “Optimal” to include additional criteria
  - As networks continue to use BGP to optimize networks, ORR will likely encompass more than just simple virtual SPF calculations
  - User-defined policies drive best path selection

Q & A

Appendix – Sample XR config and logs

```
hostname orr
!
interface Loopback0
 ipv4 address 192.168.0.50 255.255.255.255
!
interface GigabitEthernet0/0/0/0
 ipv4 address 192.168.3.2 255.255.255.252
!
route-policy no_install
 drop
end-policy
!
route-policy reset_nh_local
 set next-hop 192.168.0.2
 pass
end-policy
```

```
router bgp 65501
 ibgp policy out enforce-modifications
 address-family ipv4 unicast
  optimal-route-reflection orr1 192.168.0.2
  optimal-route-reflection orr2 192.168.0.3
  table-policy no_install
 !
 address-family link-state link-state
!
neighbor 192.168.0.1
 remote-as 65501
 local address 192.168.3.2
 address-family ipv4 unicast
 !
!
neighbor 192.168.0.2
 remote-as 65501
 local address 192.168.3.2
 address-family link-state link-state
 !
!
neighbor 192.168.0.3
 remote-as 65501
 local address 192.168.3.2
 address-family ipv4 unicast
  optimal-route-reflection orr2
  route-policy reset_nh_local out
  route-reflector-client
```

Appendix – Sample XR config and show cmds

```
RP/0/RP0/CPU0:orr#show orrspf database
Thu May 25 07:59:10.979 UTC

ORR policy: orr1, IPv4, RIB tableid: 0xe0000004
Configured root: primary: 192.168.0.2, secondary: NULL, tertiary: NULL
Actual Root: 192.168.0.2, Root node: 1921.6800.0002.0000

ORR policy: orr2, IPv4, RIB tableid: 0xe0000005
Configured root: primary: 192.168.0.3, secondary: NULL, tertiary: NULL
Actual Root: 192.168.0.3, Root node: 1921.6800.0003.0000RP/0/RP0/CPU0

orr#show orrspf database detail
Thu May 25 07:59:42.157 UTC

ORR policy: orr1, IPv4, RIB tableid: 0xe0000004
Configured root: primary: 192.168.0.2, secondary: NULL, tertiary: NULL
Actual Root: 192.168.0.2, Root node: 1921.6800.0002.0000

Prefix                                        Cost
192.168.0.1                                   10
192.168.0.2                                   0
192.168.0.3                                   10

ORR policy: orr2, IPv4, RIB tableid: 0xe0000005
Configured root: primary: 192.168.0.3, secondary: NULL, tertiary: NULL
Actual Root: 192.168.0.3, Root node: 1921.6800.0003.0000

Prefix                                        Cost
192.168.0.1                                   20
192.168.0.2                                   10
192.168.0.3                                   0
```