

BGP Diverse Paths

draft-ietf-grow-diverse-bgp-paths-dist-02

Keyur Patel

RIPE 61, November 2010, Rome, Italy

Motivation

- BGP protocol defined in RFC4271 provides a mechanism to announce only a single path aka a BGP bestpath
- Announcement of multiple paths within BGP is gradually becoming a requirement
 - Used for BGP Fast Convergence
- Different solutions available to announce multiple paths with BGP
 - BGP Full mesh + BGP Best External
 - BGP Additional Paths
 - Different RD approach in BGP VPNs
 - BGP Diverse paths

BGP Full Mesh + BGP Best External

- BGP full mesh explained at length in BGP RFC 4271
 - Assures path diversity in hot potato routing
 - Needs BGP Best External support for cold potato routing
- BGP Best External explained at length in draft-ietf-idr-best-external-02.txt
- All BGP speaking routers establish peering with each other
 - Results in TCP state explosion
 - BGP can store as many paths as sessions

BGP Add Paths

- BGP Add paths explained at length in draft-walton-bgp-addpaths-06.txt
- Mechanism to announce multiple paths over a single neighbor session
- New BGP protocol encoding and a capability
 - Requires a network-wide upgrade
- Operators need to figure out how many paths and what paths to announce
 - Can the edge router handle a load of more than 2 paths?

VPNs & Different RD Approach

- L3VPNs can announce multiple paths using different RD approach
 - Implementation tweak: - SPs can configure multiple RDs for their multi-homed customers
 - Each unique RD creates separate set of customer prefixes and thereby ensures multiple paths
- Works well and deployed in most of the L3VPN environment
 - No reason to look into BGP Add Paths or BGP Diverse paths

BGP Diverse Paths

- RR based solution
 - Does NOT require an upgrade of an entire network
- Does not require any protocol changes
- Applicable to all BGP AFI/SAFIs
 - Targeted mostly for IP hop-by-hop as well as tunneled networks
- Diverse paths are announced either using
 - Shadow RRs that calculate only diverse paths
 - Shadow sessions that announce only diverse paths

BGP Diverse Paths (RR Functionality)

- RRs need to pre-compute diverse paths as part of its bestpath processing
- RR can operate as if it was a diverse/backup RR and only announce diverse paths to its configured neighbors
 - IGP metric check needs to be disabled in order to deterministically compute diverse paths
- Alternatively RRs can track session types and announce appropriate diverse path accordingly
 - Requires configuration of new addresses for diverse path sessions (one address per nth diverse path announced to all BGP neighbors)
- RRs in forwarding plane needs to pre-compute and install diverse paths in its RIB and FIB
 - Covers hot potato routing as well as cold potato routing

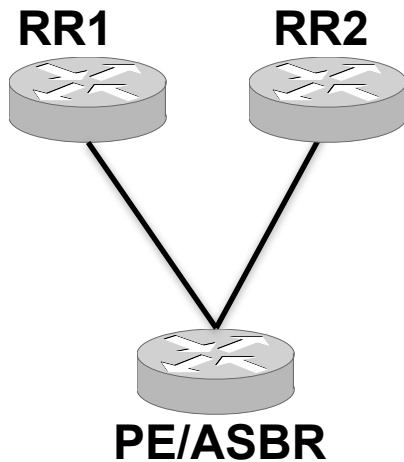
Diverse Paths PE/ASBR Functionality

- PE/ASBR may need configure additional sessions with RRs
 - Does NOT require any software upgrade
 - Can reuse its existing peering address for diverse path sessions
- PE/ASBR treats diverse paths as just another BGP path received
- PE/ASBR can use diverse paths to pre-compute and install backup paths in RIB and FIB
- PE/ASBR needs to account for
 - Additional sessions to enable diverse paths
 - Additional memory to store diverse paths

RR Deployment Models

Today's case:

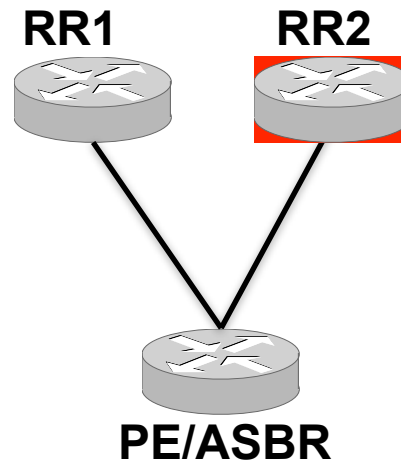
10/8 - P1*, P2



10/8 - P1 via RR1
- P1 via RR2

Diverse RR:

10/8 - P1*, P2

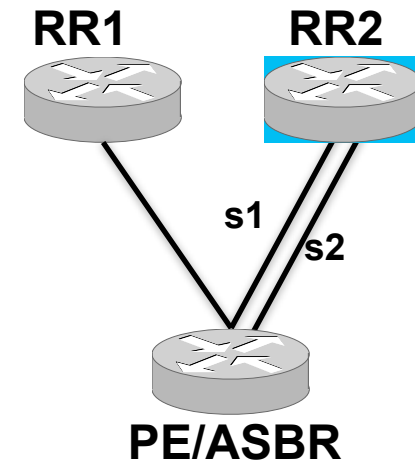


10/8 - P1 via RR1
- P2 via RR2

RR1 & RR2 same
IGP location or IGP
metric ignore in best
path

Diverse sessions:

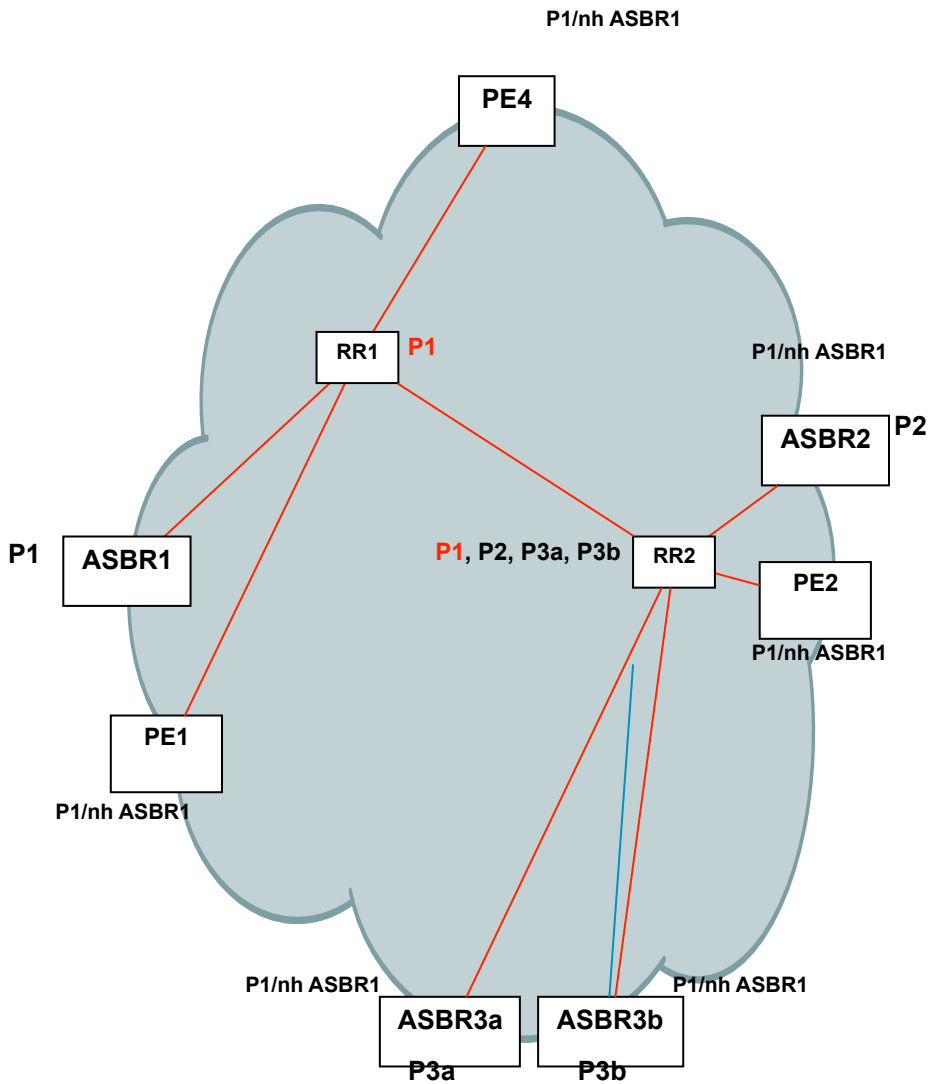
10/8 - P1*, P2



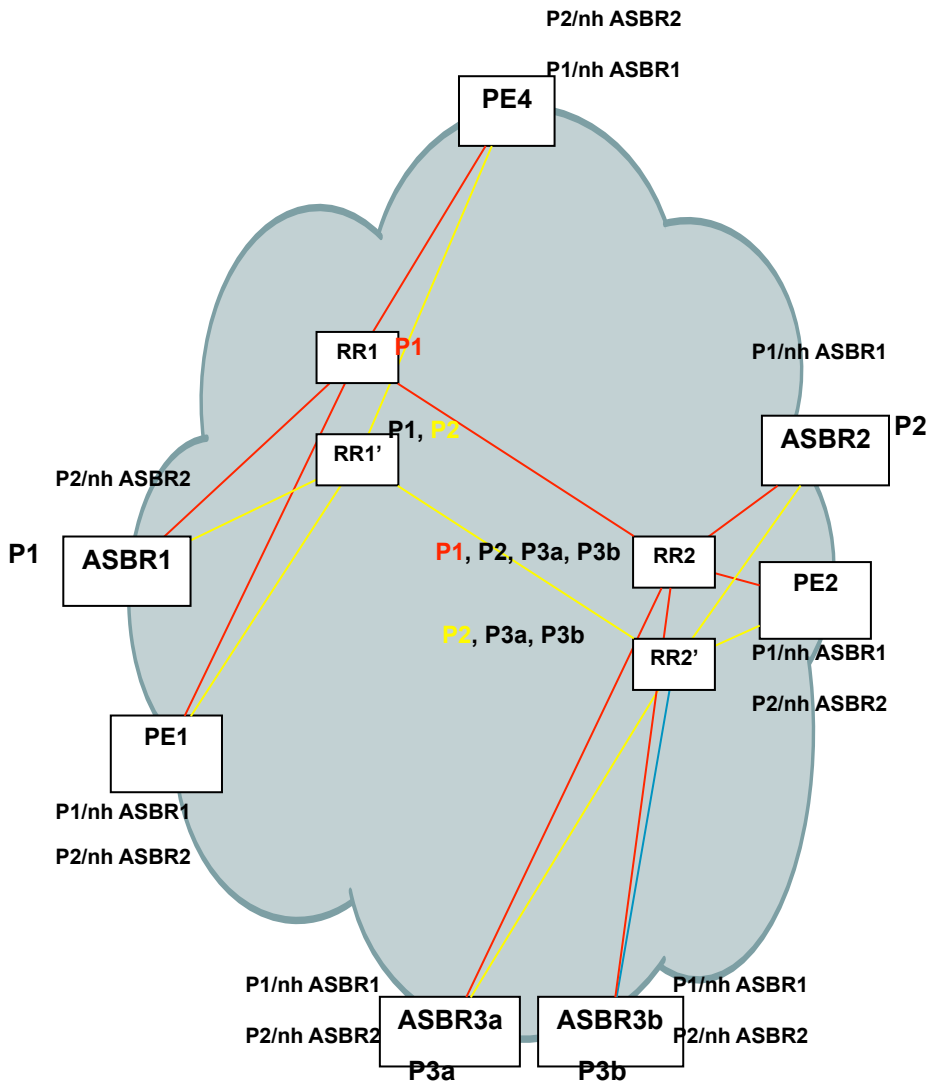
10/8 - P1 via RR1
- P1 via RR2 s1
- P2 via RR2 s2

Possible to peer both
sessions s1 and s2 to
the same loopback on
PE/ASBR.

IBGP Network With Flat Igp & Multiple RR



IBGP Network With Flat Igp, Multiple RR &

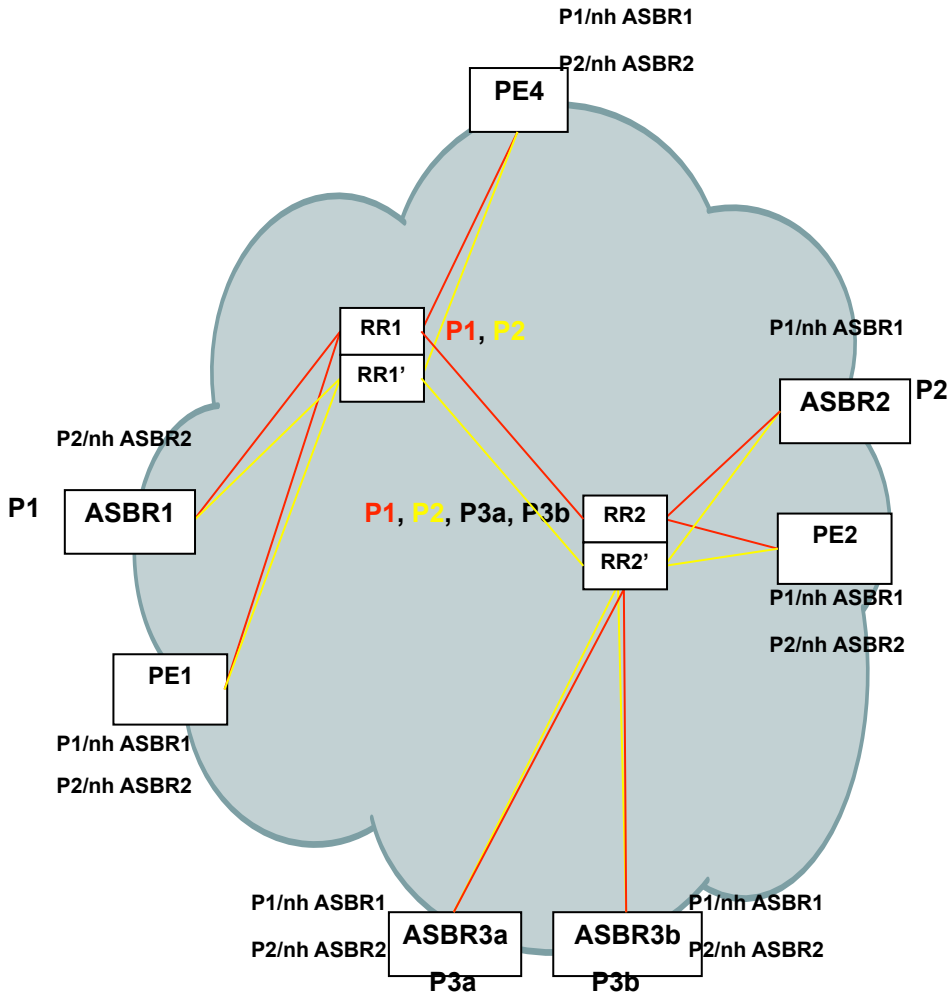


- RR1' and RR2' are **shadow RRs**
- RR1, RR1' & RR2, RR2' are in different clusters
- They are configured to calculate and advertise diverse path to it's clients
- They can do it on a per AFI/SAFI basis
- Same IGP metric as best RRs or IGP metric disabled on both
- P1 overall best and P2 second best/diverse

Deployment Considerations:

- No upgrade of any existing PE/ASBR is required
- Additional IBGP session per each diverse path
- Additional shadow RR per cluster
- Works within flat domain or within each area of hierarchical network

IBGP Network With Flat Igp And Multiple RR Clusters

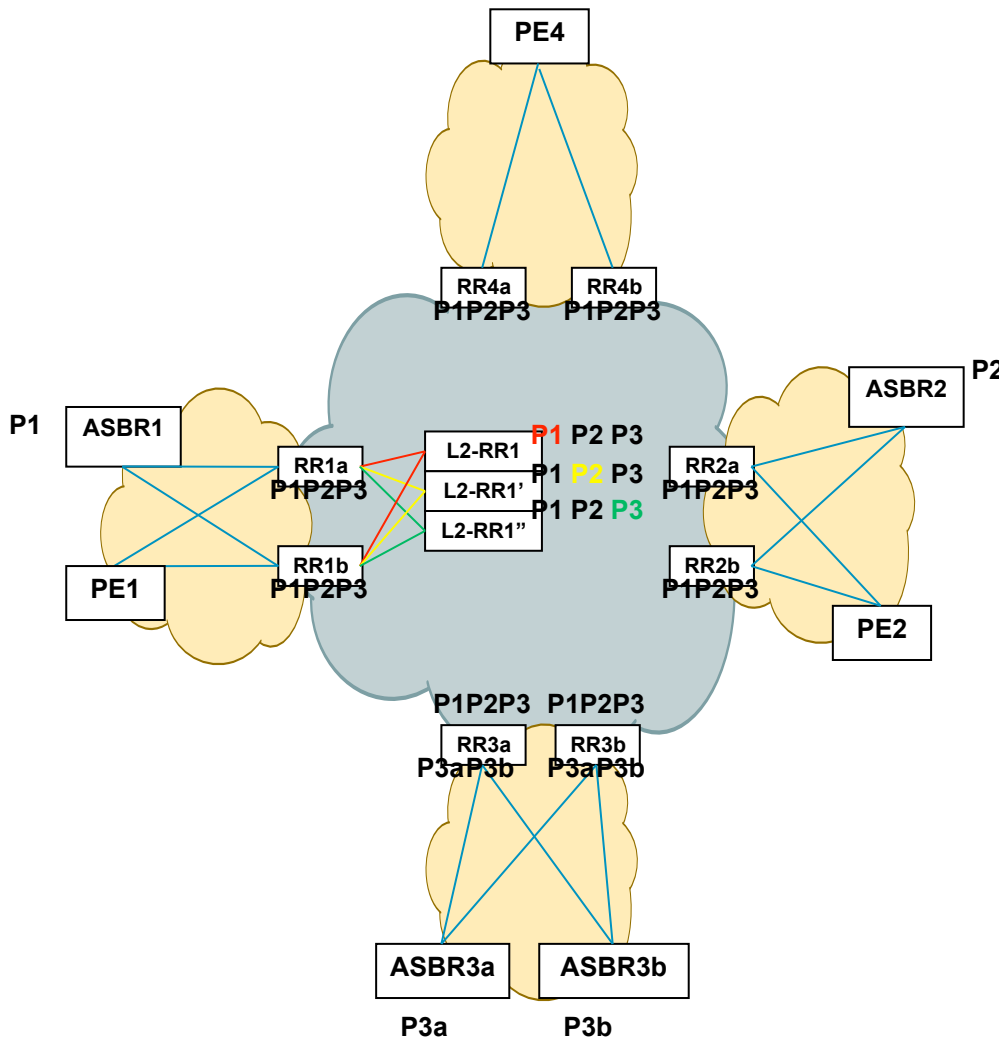


- RRs and RRs' are **same RRs**
- RR1, RR1' & RR2, RR2' are in different clusters
- They are configured to calculate and advertise diverse path to it's clients on a per neighbor basis
- They can do it on a per AFI/SAFI basis
- Same IGP metric as best RRs
- P1 overall best, P2 second best/diverse

Deployment Considerations:

- No upgrade of any existing PE/ASBR is required
- Additional IBGP sessions per each diverse path
- No additional Shadow RRs required
- Works within flat domain or within each area of hierarchical network

Hierarchical RRs



- Each shadow RR is configured to calculate and advertise diverse path to it's clients (POP RRs) on a per neighbor basis
- They can do it on a per AFI/SAFI basis
- Any encapsulation can be used within each area IP or MPLS (option).
- Pop RRs replicate the design in the same manner

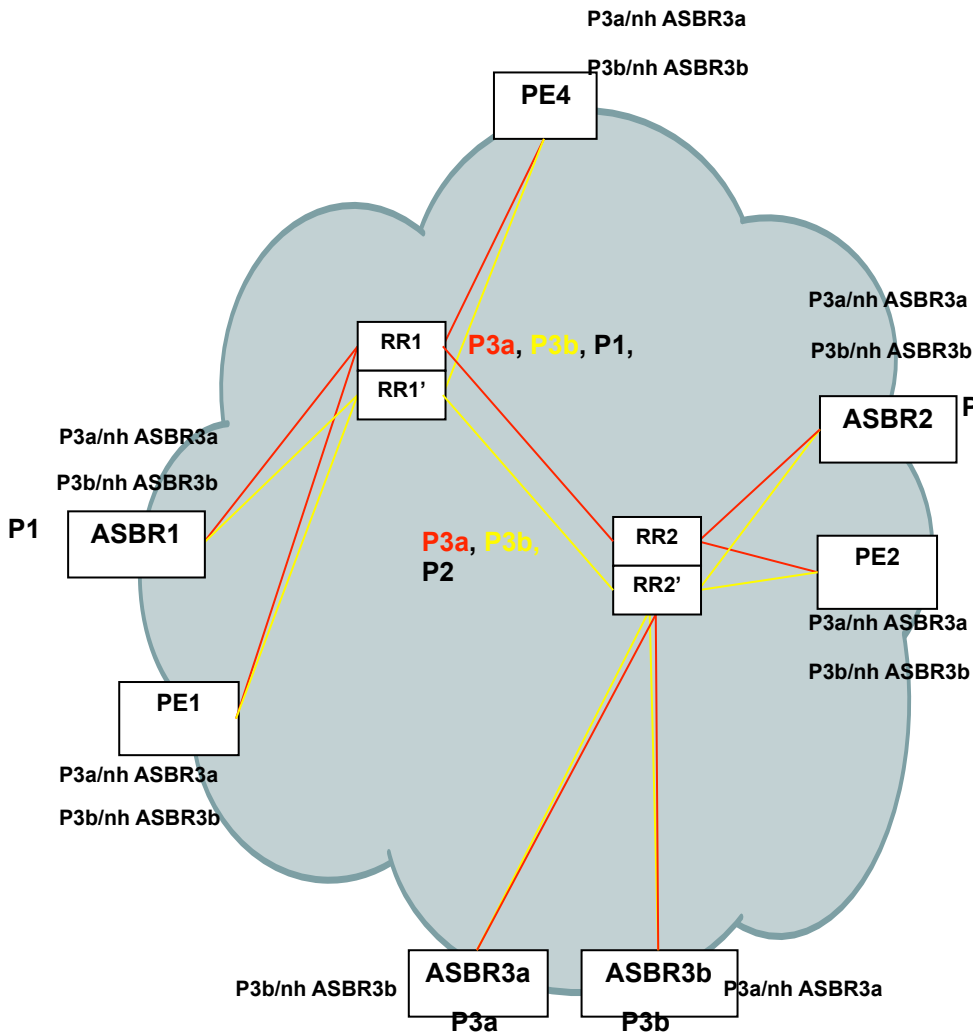
Deployment Consideration:

- No upgrade of any existing PE/ASBR is required
- Additional IBGP sessions per each diverse path

Diverse Paths & BGP Best External

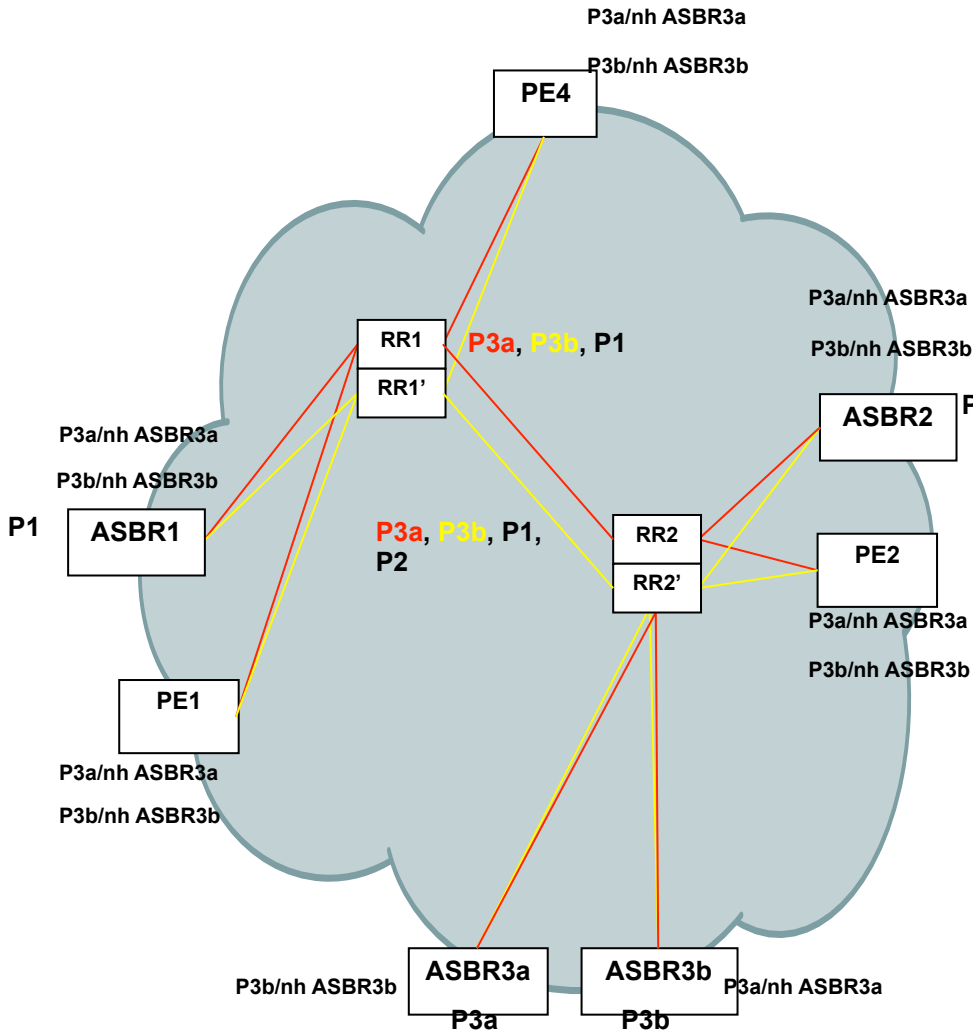
- Intra-cluster routes may not always get announced with diverse paths
 - Diverse paths does NOT change any BGP route propagation rules
 - IBGP network still has multiple diverse paths for a given prefix
 - Not an issue within a Cluster
- Usage of BGP Best External between RRs can assure announcements of intra cluster routes
 - Reduce number of sessions between full meshed RRs
 - Achieve higher path diversity within the network

Diverse Paths With Missing Intra-Cluster routes



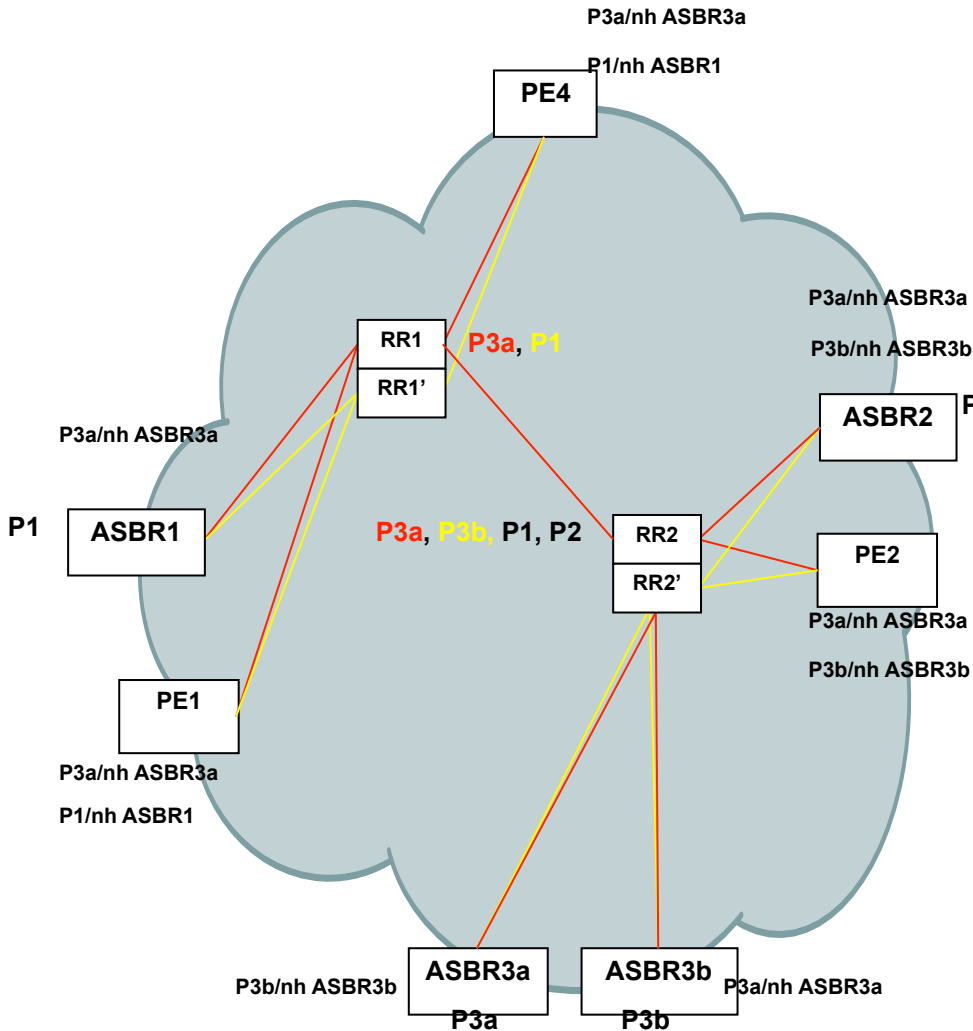
- RR1 and RR1' are **same RRs**
- RR1, RR1' & RR2, RR2' are in different clusters
- They are configured to calculate and advertise diverse path to it's clients on a per neighbor basis
- They can do it on a per AFI/SAFI basis
- Same IGP metric as best RRs
- Cold Potato Routing (Local Prefs win)
- P3a overall best, P3b second best/diverse
- RR1 and RR1' got 3 paths
- RR2 and RR2' got 3 paths
- Network still has path diversity

BGP Diverse Paths + BGP Best External



- RRs and RRs' are **same RRs**
- RR1, RR1' & RR2, RR2' are in different clusters
- They are configured to calculate and advertise diverse path to it's clients on a per neighbor basis
- They can do it on a per AFI/SAFI basis
- Same IGP metric as best RRs
- Cold Potato Routing (Local Prefs win)
- P3a overall best, P3b second best/diverse
- RR1 and RR1' got 3 paths
- RR2 and RR2' got 4 paths
- Network still has path diversity

BGP Diverse Paths + BGP Best External



- RRs and RRs' are **same RRs**
- RR1, RR1' & RR2, RR2' are in different clusters
- They are configured to calculate and advertise diverse path to it's clients on a per neighbor basis
- No diverse session between RRs
- Best External enabled between RRs
- They can do it on a per AFI/SAFI basis
- Same IGP metric as best RRs
- P3a overall best, P1 & P3b second best/diverse
- RR1 and RR1' got 2 paths
- RR2 and RR2' got 4 paths
- Network still has path diversity
- Single session between RRs

Acknowledgements

Special Thanks to:

Randy Bush

Robert Raszuk

Chris Cassar

Satish Mynam

Selma Yilmaz

Questions?