Diverse BGP Path Distribution version 00

draft-raszuk-diverse-bgp-path-dist-00

Robert Raszuk, Keyur Patel, Isidor Kouvelas

Rex Fernando, Danny McPherson

Agenda:

- **Background**
- **Idea presentation**
- **Discussion**

Background ....

- There are many applications today that require presence of other then BGP best path in BGP speakers.

- There are also various ways to acomplish this:

1. **Full mesh** (Can be accomplished manually or with automated IBGP auto discovery)
2. **Add-path** (Universal solution, but has few issues: requires to keep per path vs per prefix advertised state (memory price), requires quite heavy implementation changes to bgp and worst requires full network upgrade: PEs, ASBRs, RRs → May take years to be fully deployed)
3. **Nth best path propagation by route reflectors to regular non upgraded/non upgradable clients**

Let's discuss N-th best path solution. It can allow gradual add-path deployment while meeting most critical customer's requirements in a very fast way.

Today's RRs & N-th best path deployment

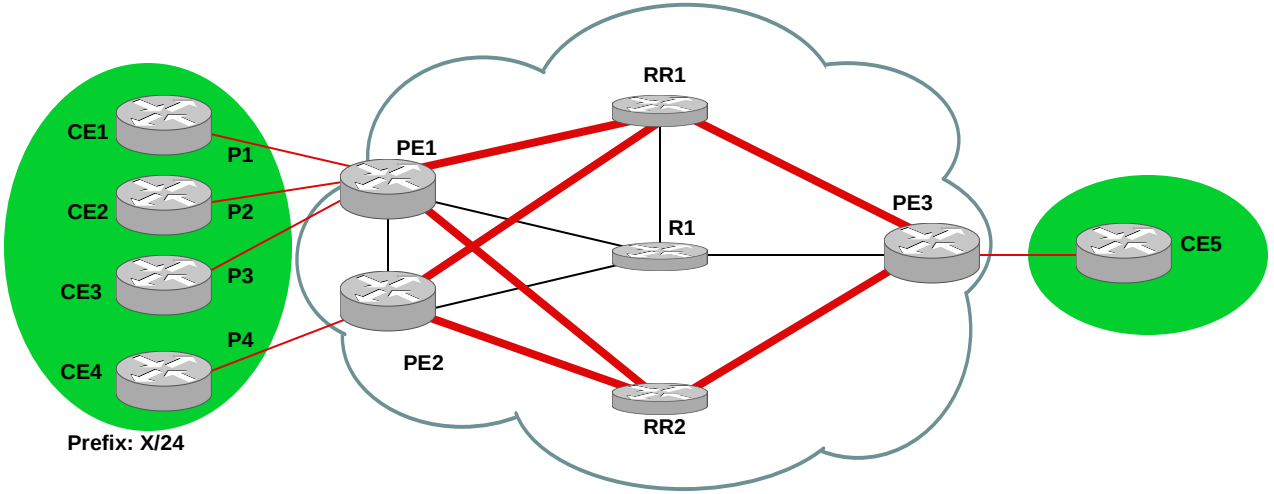
- **Traditional IPv4/IPv6 RR design**

  → RRs are in the data plane path most othen on the POP to Core boundary .. Additonal RR's would be added on the same IGP locations or current RRs would be upgraded to serve new application without need to deploy new platforms

  - **Control plane only RRs usually in the core of the network**

  → RRs are not in the data path, are control plane devices, IGP metric to next hop step in best path to be disabled ... New RR(s) added as a control plane devices or current RRs upgraded to server new application.

Typical IBGP Network design with control plane RRs:



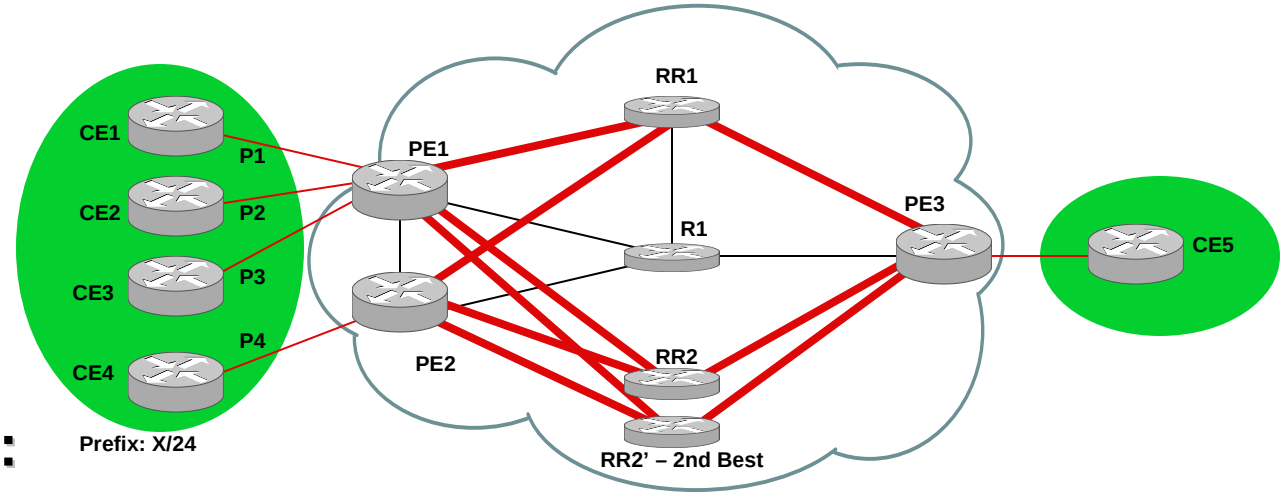
Today:

- **RRs would have P4 & P[1|2|3] unless MED/Local Pref forces P4 on PE1 to be selected as best !**
- **Best external on PE1 will allow P[1|2|3] to be propagated to RRs regardless if P4 carries better local pref/med.**
- **Now RRs select the best and advertise it to PE3**
- **PE3 ends up holding two paths (via RR1 & RR2) with in most cases the same NH – PE1 or PE2 assuming PEs do next hop self.**

Agenda:

- Background
- Idea presentation
- Discussion

IBGP Network design with RRs & 2nd best path:



Proposal:

- **Let's make the RR2 to select the best and let's introduce RR2' to select 2nd best and advertise it to it's clients**
- **Clients do not need to be upgraded ! Just one regular extra IBGP session is needed for each additional path.**
- **RR2' would select 2nd best path ... Different from primary best path on RR2 by next hop & router_id/originator_id.**
- **Best external on PE1 makes sure that RR2 has other path to select from available**
- **Now PE3 has still two paths but both from different exit points**

BGP Nth best path - discussion:

- **Please note that RR2 & RR2' may be the same physical route reflector distributing to both add-path capable peers and non add-path capable peers multiple BGP paths.**

- **In the latter case the distribution of additional paths happens over dedicated IBGP sessions .. One session for each additional bgp path bound to different RR2 loopback address. (No new loopbacks on the PEs needed !)**

- **We need the same functionality even if we will support add-path as the same code will be required on each BGP speaker to select 2nd ... Nth best path ... To make the local decision which other paths are to be installed in RIB/FIB**

IBGP Network design with RRs & Nth best path :

- **This solution does not require any new protocol changes.**
- **This solution is applicable to any AFI/SAFI**
- **This solution also addresses the 2547 Inter-AS option B ASBR redunddancy issue without requiring RD rewrite**
- **RR2' can also be just a single logical/virtual router or another instance of BGP process running on a primary RR**
- **Nth best RR should be co-located with primary best RRs to address the case of IGP distance to BGP nh influence. Alternatively especially in the control plane only RRs IGP distance to BGP nh step should be relaxed/skipped from best path.**

Agenda:

- Background
- Idea presentation
- Discussion

BGP Nth best path – discussion ....

- **This is a tool for IBGP multipath, reduction of BGP churn as well as key component for BGP fast connectivity restoration for any arbitrary peering topology.**

Acknowledgments

- **The authors would like to thank Bruno Decraene and Eric Rosen for their valuable input.**

Thank you !