

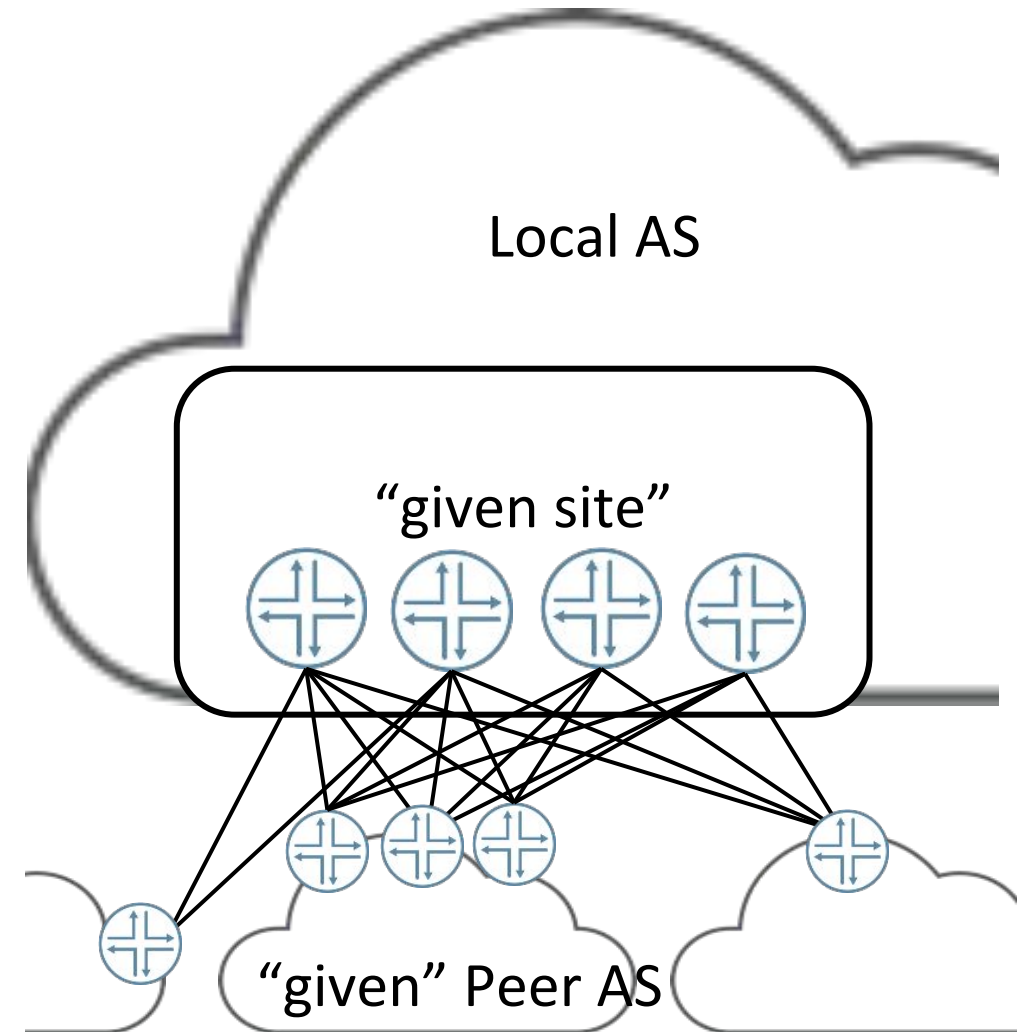
Use of Abstract NH in Scale-Out peering architecture

[draft-szarecki-grow-abstract-nh-scaleout-peering-00](#)

R. Szarecki, Ed.
K. Vairavakkalai
N. Venkataraman

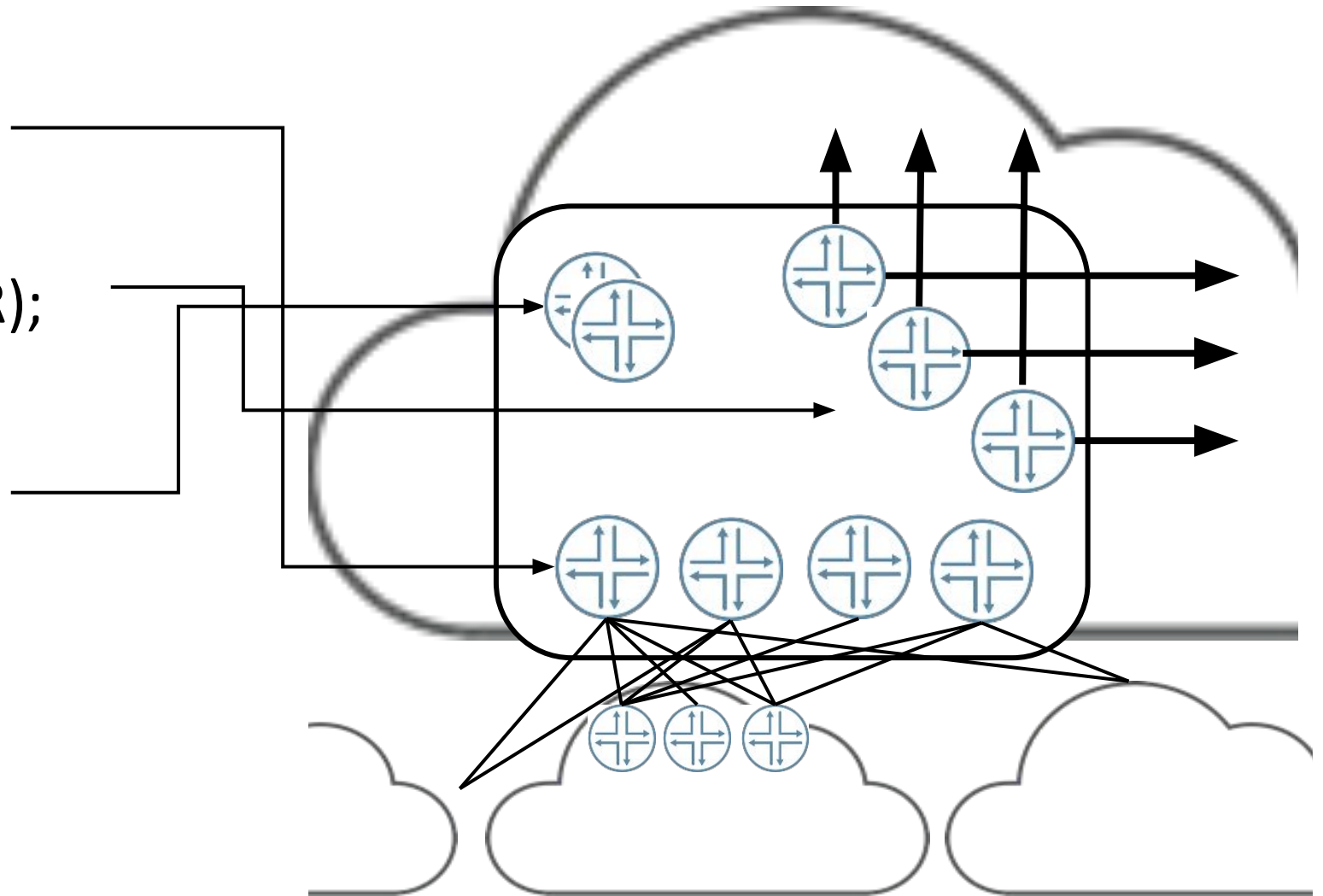
Scale-out peering

- Given Peer AS is connected to multiple ASBRs at given site (of local AS)
 - All ASBRs are set to construct ECMP toward given Peer AS (egress ECMP)
 - All ASBRs sends equal routing information to given Peer AS (desired ingress ECMP)
- Principles
 - Provides required B/W with Load Balancing (ECMP, BGP multipath)
 - Provides site-local N+1 redundancy
 - Scale-out ASBRs w/o impacting other sites
 - Restore optimal traffic in seconds, not 10's minutes.

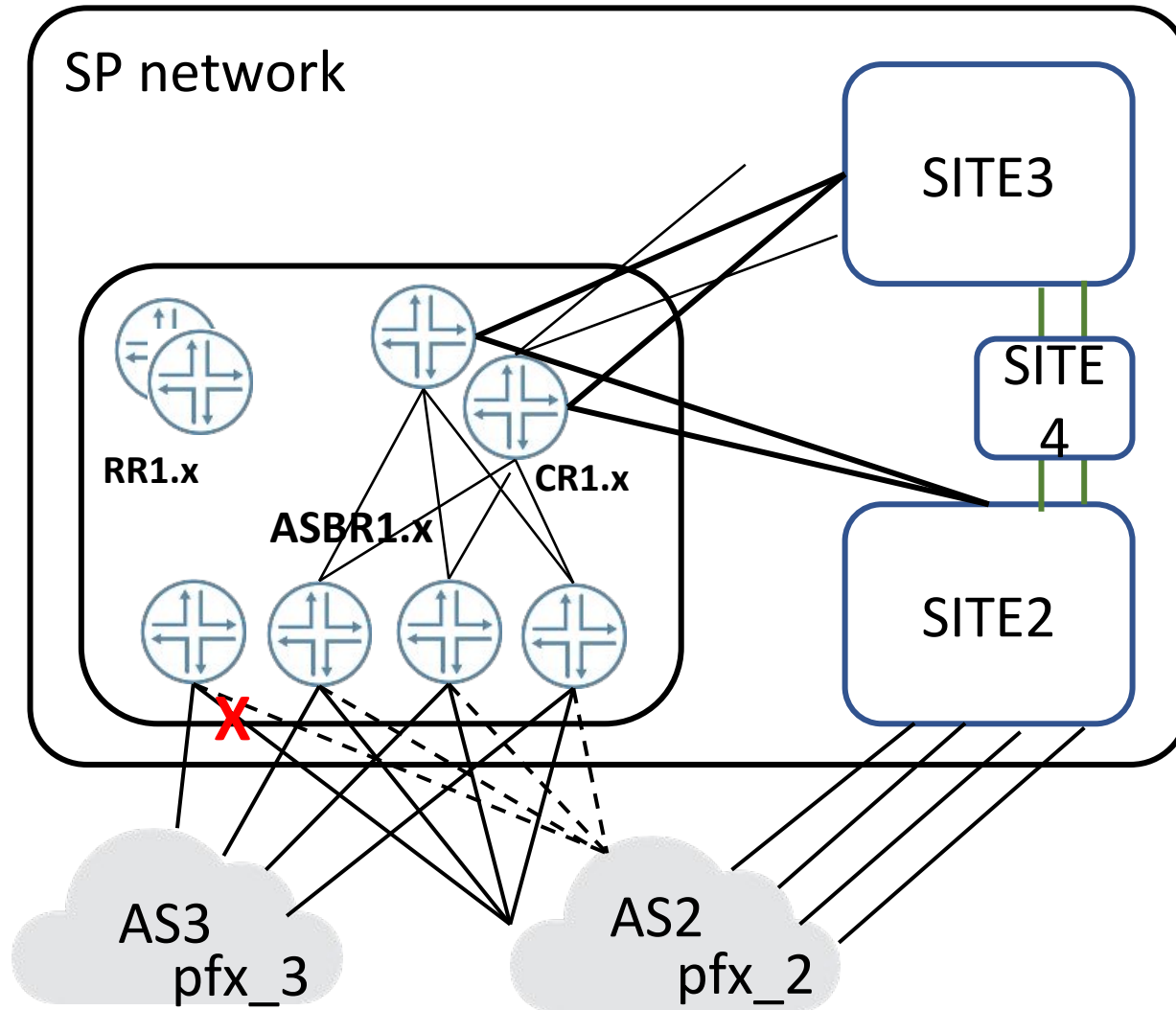


Scale-Out Peering Site

- Multiple (N) ASBRs;
- 1-2-M Core Routers(CR);
- Route Reflectors(RR)



Example network



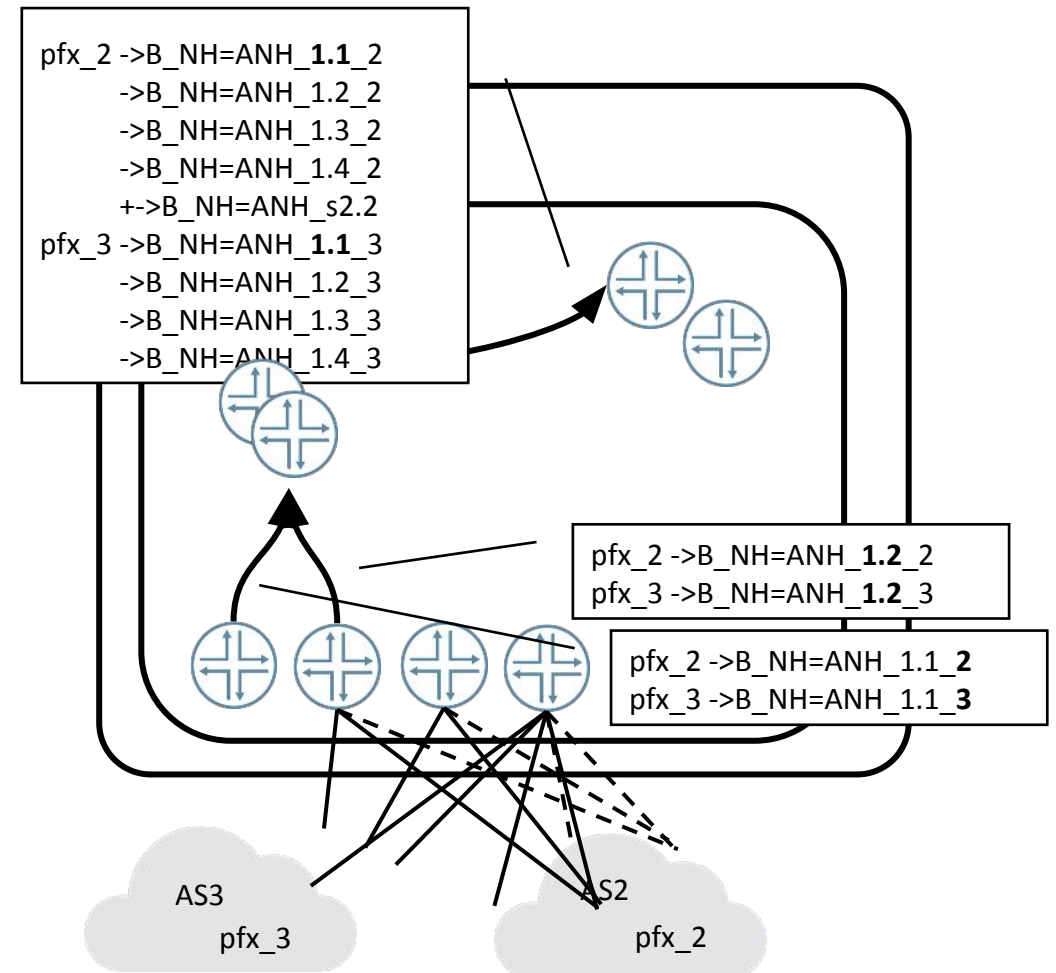
- Native IP network.
- BGP:
 - one active path advertised among RR.
 - multipath, ADD-PATH w/ ASBR and CR
 - ASBR – only active PATH to RR
- Classic NH shortages
 - NHS requires path withdraw when ASBR loos all sessions.
 - NH unchanged:
 - Session DOWN, IF up cease
 - Temporal move to SITE 2 @ single session DOWN

Abstract Next-Hop (ANH)

- No protocol changes. Same old good BGP!
- Arbitrary IP/32 address:
 - set as B_NH when path form (member of) **sub-set** of eBGP advertised to iBGP.
 - CONDITIONALLY inserted into IGP
 - When at least one eBGP session form **sub-set** is ESTABLISHED/Converged
- Sub-set of eBGP sessions – configuration, up to operator's decision: E.g.
 - all sessions on given ASBR with same peer AS
 - all sessions on given ASBR with same Transit providers
 - all sessions on given SITE with same peer AS
- Generic concept, not only for scale-out peering.

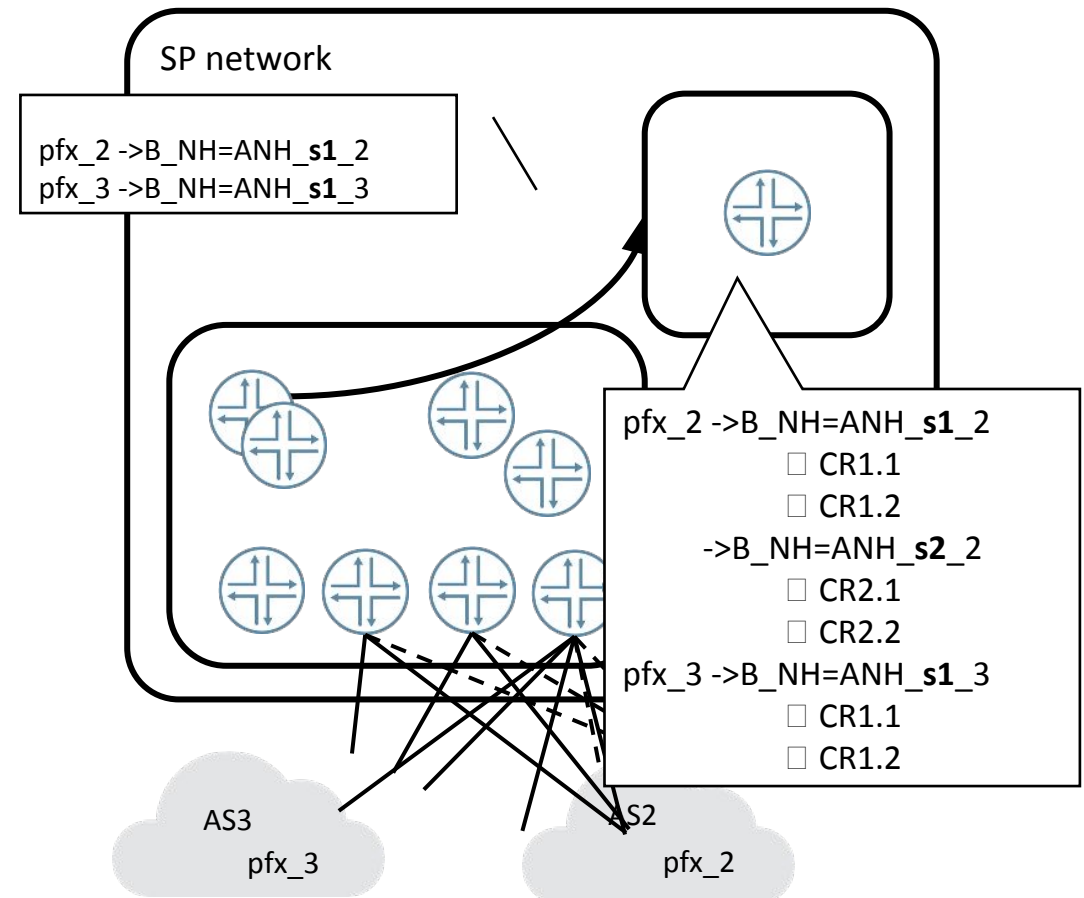
ASBR operation w/ ANH – intra-site

- eBGP sessions sub-set - all session w/ AS2 from ASBR.
- ASBR-PeerAS-ANH (**AP-ANH**) - unique per ASBR, per PeerAS – (**ANH_1.1_2**)
- RR1.x
 - gets **1** path from each ASBR w/ B_NH==AP-ANH (regardless of # eBGP session w/ AS2)
 - advertise ADD-PATH (5) to CR1.x and ASBR
- CR1.x load-balance among 4 B_NHs



ASBR operation w/ ANH – AS-wide

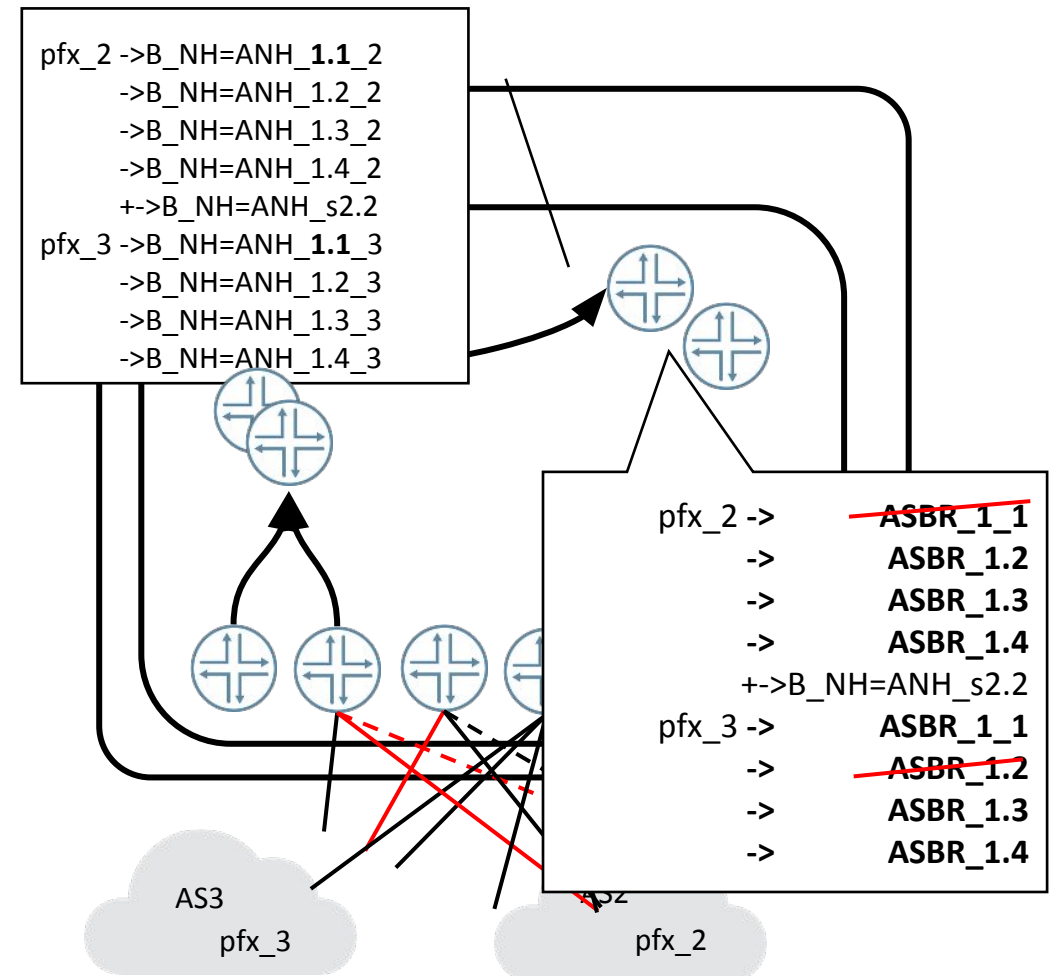
- eBGP sessions set - all session w/ AS2 from all ASBR at site.
- Site-PeerAS-ANH (**SP-ANH**) - unique per SITE, per PeerAS – (**ANH_s1_2**)
- RR1.x - advertise one path to other sites (RRy.x) w/ B_NH:=SP-ANH
- ASBR1.x – insert SP-ANH into IGP if its AP-ANH is active.
- CR/CS at other sites – resolves SP-ANH via IGP- load balance among all CR1.x/CRy.x



Failures

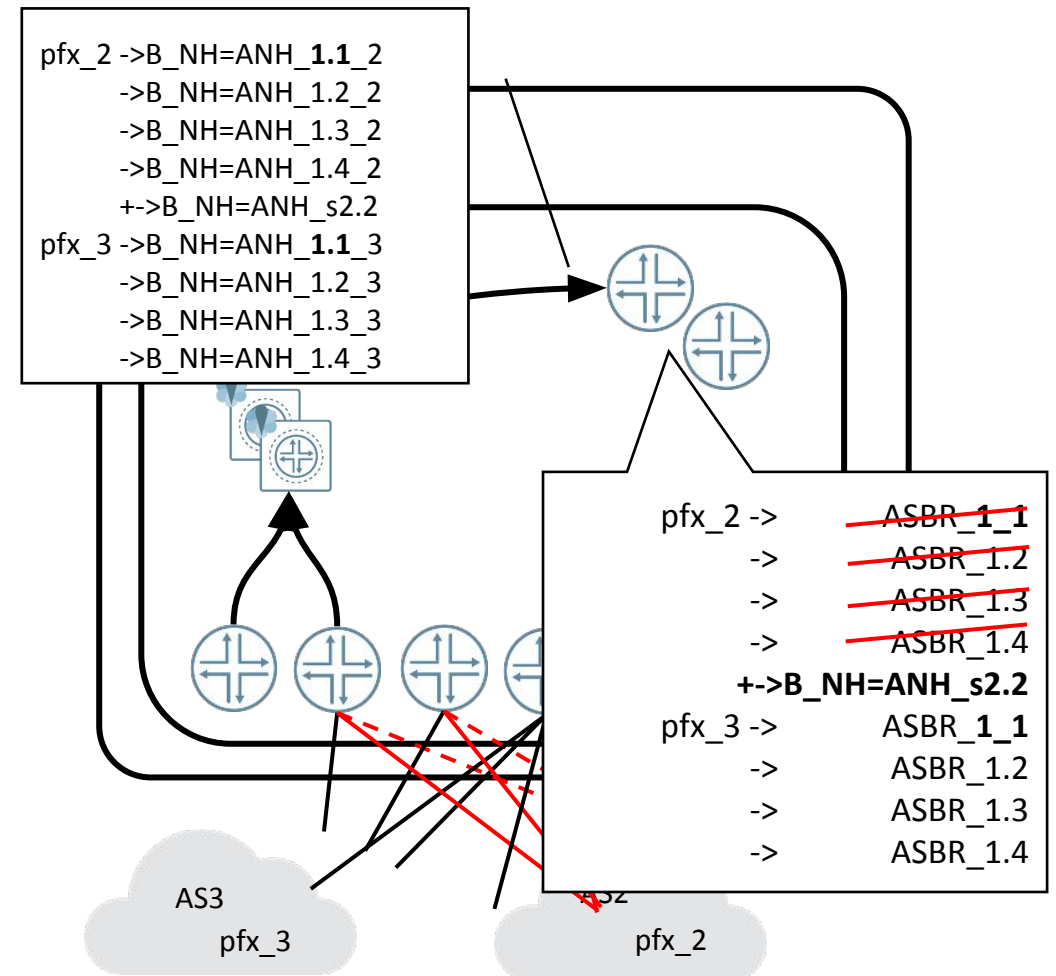
ASBR operation w/ ANH – intra-site

- One session with AS2
 - AP-ANH on iBGP not changed. Other attributes unchanged.
 - No Update send to RR.
 - like NHS
- One session with AS3 (@ ASBR1.2)
 - AP-ANH unreachable in IGP. Path invalid. CR remove path form ECMP group. CR sent to un-affected ASBR only.
 - ASBR1.1 (slowly) withdraws paths
 - like NH unchanged (peer IP)
- All session with AS2 (@ ASBR1.1)
 - AP-NH unreachable in IGP. Path invalid. CR remove path form ECMP group. CR sent to un-affected ASBR only.
 - ASBR1.1 (slowly) withdraws paths



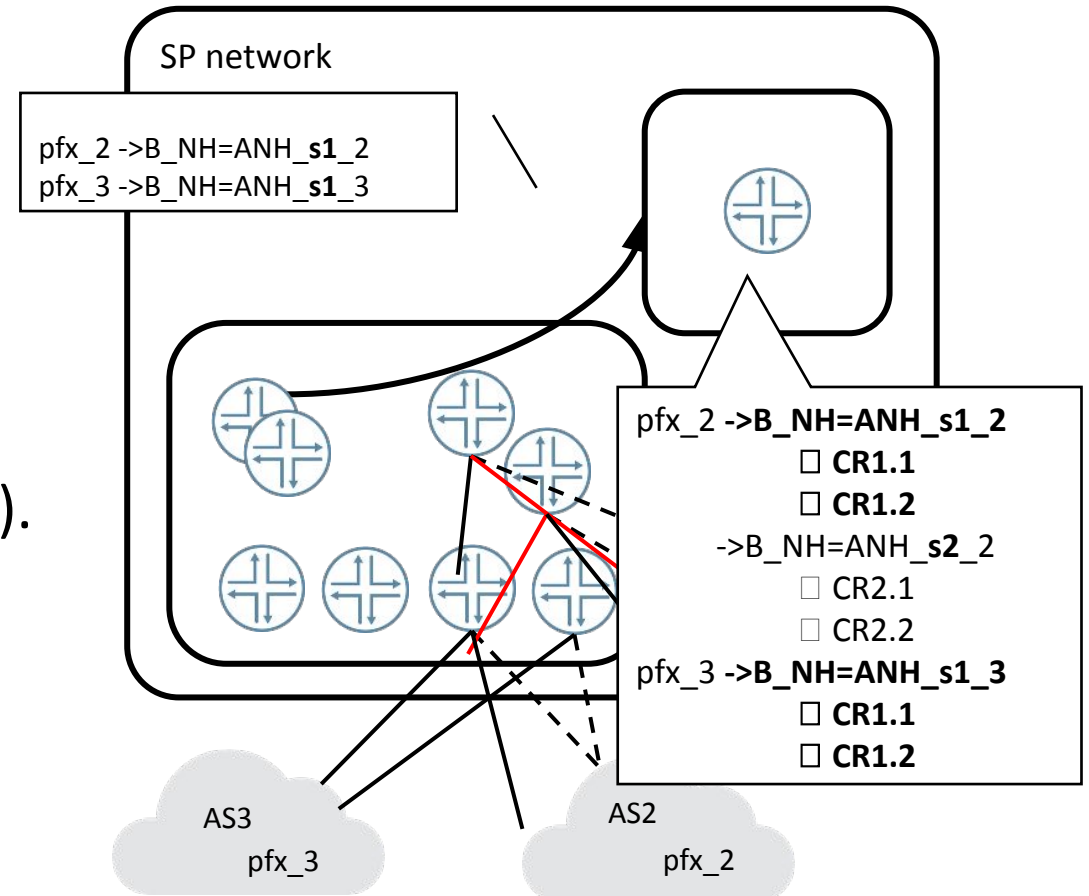
ASBR operation w/ ANH – intra-site

- All session with AS2 on all ASBR_1.x
 - All AP-ANHs unreachable in IGP.
 - Path invalid. CR remove 4 path form ECMP group.
 - CR sent traffic to other site.
 - ASBR1.x (slowly) withdraws paths



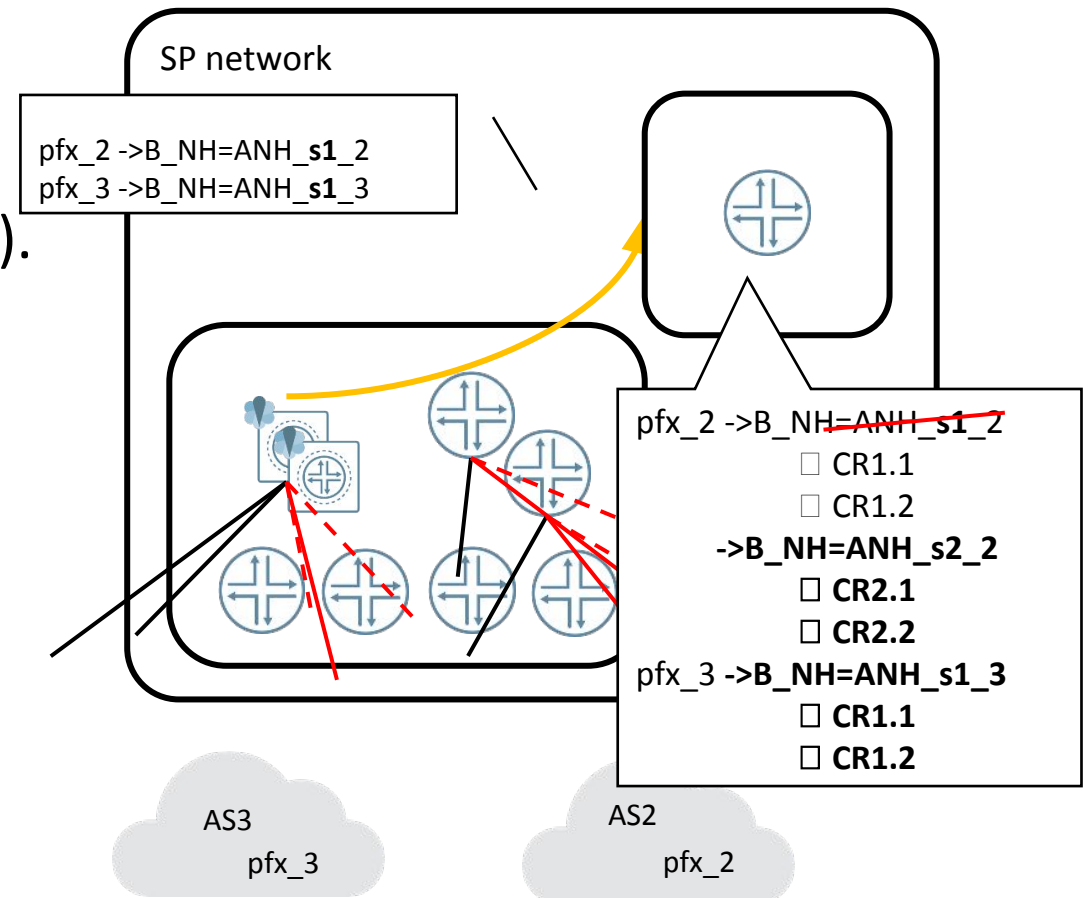
ASBR operation w/ ANH – AS-WIDE

- One session with AS2
 - No Update/Withdraw send to among RR.
 - like NHS
- One session with AS3 (@ ASBR1.2)
 - SP-ANH reachable in IGP.
 - No FIB changes on other sites (same B_NH).
 - ORIGINATOR ID changed - Update send to among RR.



ASBR operation w/ ANH – AS-WIDE

- All session with AS2 (@ ASBR1.1)
 - SP-ANH reachable in IGP.
 - No FIB changes on other sites (same B_NH).
- All session with AS2 on all ASBR_1.x
 - AP-ANH and SP-ANH unreachable in IGP.
 - Path invalid. CR/CS remove 4 path form ECMP group.
 - At other site CR/CS sent traffic to other site.
 - ASBR1.x (slowly) withdraws paths



SUMMARY

- Scale-out solves challenges with Bandwidth, redundancy, RTT, etc.
- At cost of scale-up Control Plane states.
- New construct and practices in protocol configuration/network design needed.
 - Abstract NH is useful construct.
 - Use it to control BGP scale in scale-out peering is example of practices.

Next steps

- Seek for more feedback (thx Robert and Ron)
- Heading toward INFORMATIONAL