Avoid Blackholing: Selective BGP Next-Hop resolution

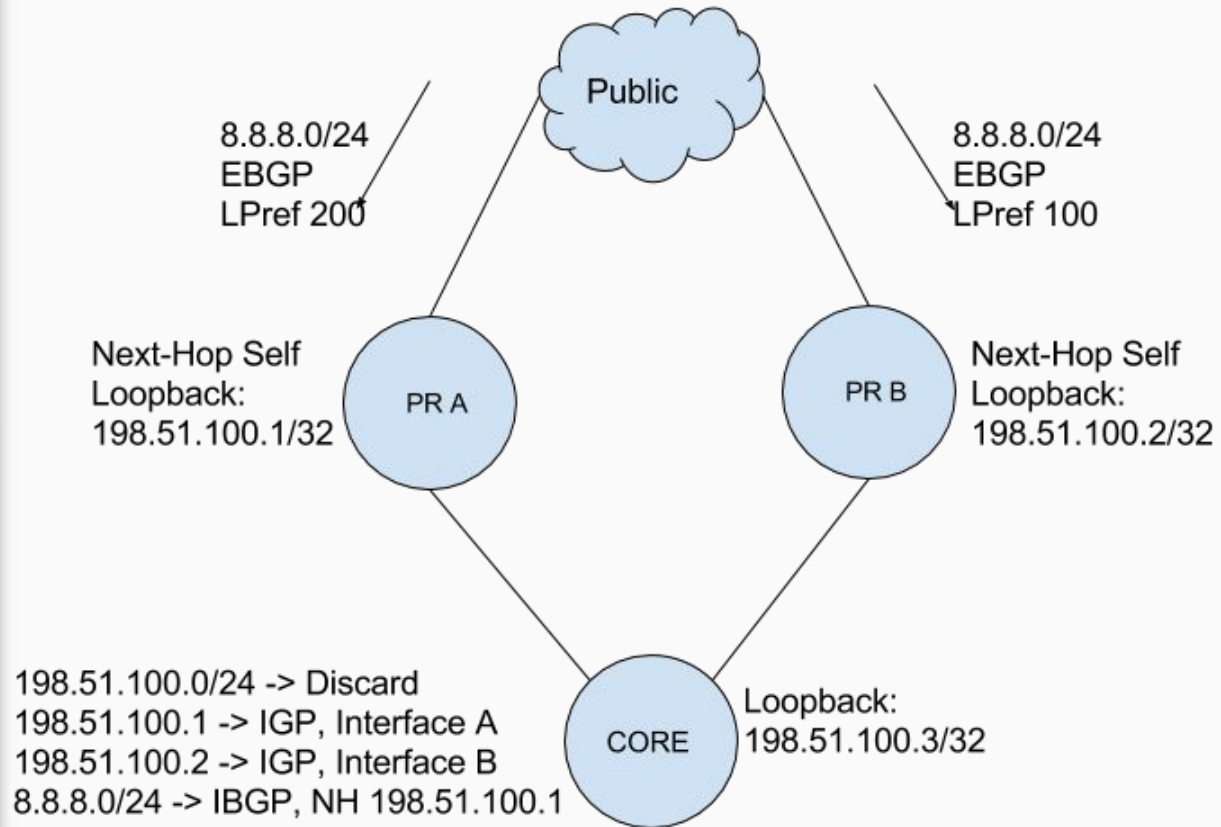Oliver Herms @ DENOG 9, <oliver.herms@exaring.de>

Example

PR-A and PR-B receive the same prefix 8.8.8.0/24 on EBGP sessions

Both apply Next-Hop Self before advertising prefix to CORE

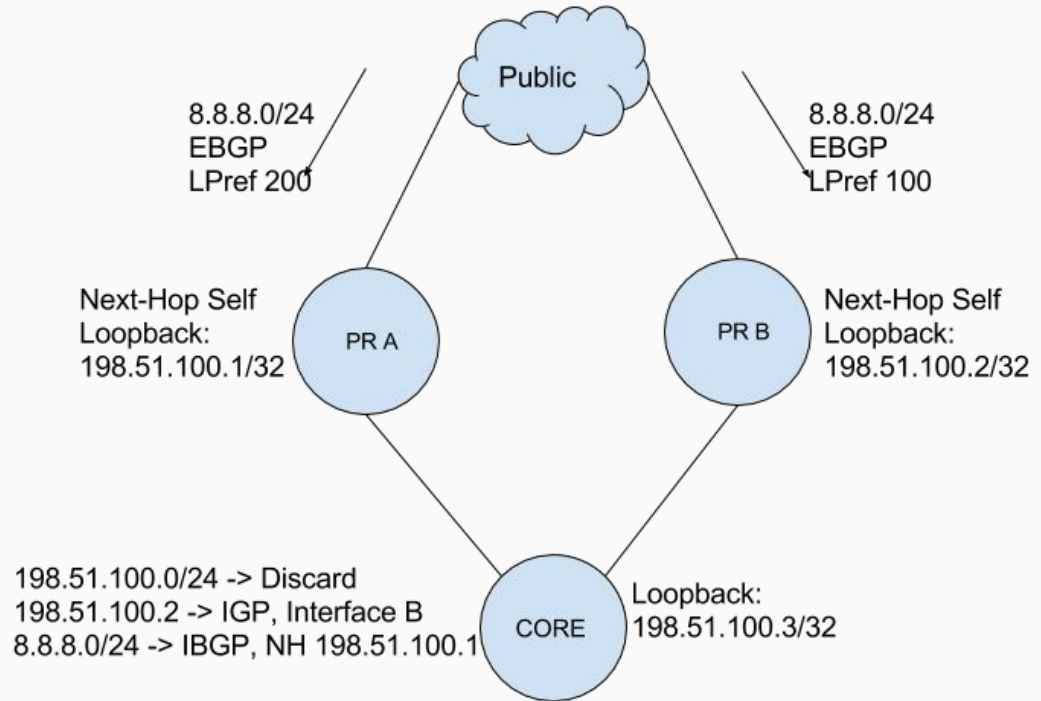CORE uses loopback addresses of PR-A and PR-B as Next-Hops resolved via IGP

Note: Loopback addresses come from an address range that is configured as a discard route!

Public

8.8.8.0/24
EBGP
LPref 200

8.8.8.0/24
EBGP
LPref 100

Next-Hop Self
Loopback:
198.51.100.1/32

PR A

Next-Hop Self
Loopback:
198.51.100.2/32

PR B

198.51.100.0/24 -> Discard
198.51.100.1 -> IGP, Interface A
198.51.100.2 -> IGP, Interface B
8.8.8.0/24 -> IBGP, NH 198.51.100.1

CORE

Loopback:
198.51.100.3/32

Problem

When PR-A fails:

➔ PR-CORE link goes down
➔ IGP converges quickly
➔ PR-A's loopback disappears from CORE's routing table
➔ IBGP sessions stays up until keepalive expires
➔ 8.8.8.0/24's NH becomes 198.51.100.0/24 -> Discard
➔ OOPS
➔ Converges after a keepalive interval is over

8.8.8.0/24
EBGP
LPref 200

Public

8.8.8.0/24
EBGP
LPref 100

Next-Hop Self
Loopback:
198.51.100.1/32

PR A

PR B

Next-Hop Self
Loopback:
198.51.100.2/32

198.51.100.0/24 -> Discard
198.51.100.2 -> IGP, Interface B
8.8.8.0/24 -> IBGP, NH 198.51.100.1

CORE

Loopback:
198.51.100.3/32

Solution: Selective Next-Hop resolution

➔ Routing policy applied by BGP when resolving Next-Hops
   ◆ Remember: Only routes with **valid Next-Hops** are being processed!
➔ **Rule: Do not use discard routes as Next-Hops in BGP NH resolution**
➔ Problem: Matching "from discard" not available in JunOS
   ◆ Solution: Indirect Next-Hops
   ◆ 192.0.2.1 discard
      198.51.100.0/24 -> 192.0.2.1
   ◆ Matching rule: "from next-hop 192.0.2.1 then reject"

Config Example

```
oherms@CORE> show configuration routing-options resolution
rib inet.0 {
    import NH-RESOLUTION; }

oherms@CORE> show configuration policy-options policy-statement NH-RESOLUTION
term DENY-BGP {
    from protocol bgp;
    then reject; }
term DENY-DISCARDS {
    from next-hop 192.0.2.1;
    then reject; }
term DEFAULT {
    then accept; }

oherms@CORE> show configuration routing-options static
route 198.51.100/24 {
    next-hop 192.0.2.1;
    resolve; }
route 192.0.2.1/32 discard;
```

Thanks for your attention

Expected Questions:

Does this only work on JunOS?
Afaik, yes!

Does this work with other vendors?
We don't know. Tell us if you find out more!

What about BFD for IBGP sessions?
Consumes more resources. Depends on timers.
Probably slower.

Follow us:

@waipuTech on Twitter

We are hiring kick ass Network Engineers with coding skills. Talk to @TimKleefass or me @taktv6.

EXARING AG