2006-2: Micro-allocations for Internal Infrastructure

Jason Schiller
schiller@uu.net
Chris Morrow
chris@uu.net

Overview

- 2006-2 to allow for a small additional non-contiguous IPv6 allocation for internal infrastructure in addition to pre-existing IPv6 Aggregate
- Remove BGP convergence issue
  - –3 min black-holing
- Address security considerations

BGP Re-convergence Problem

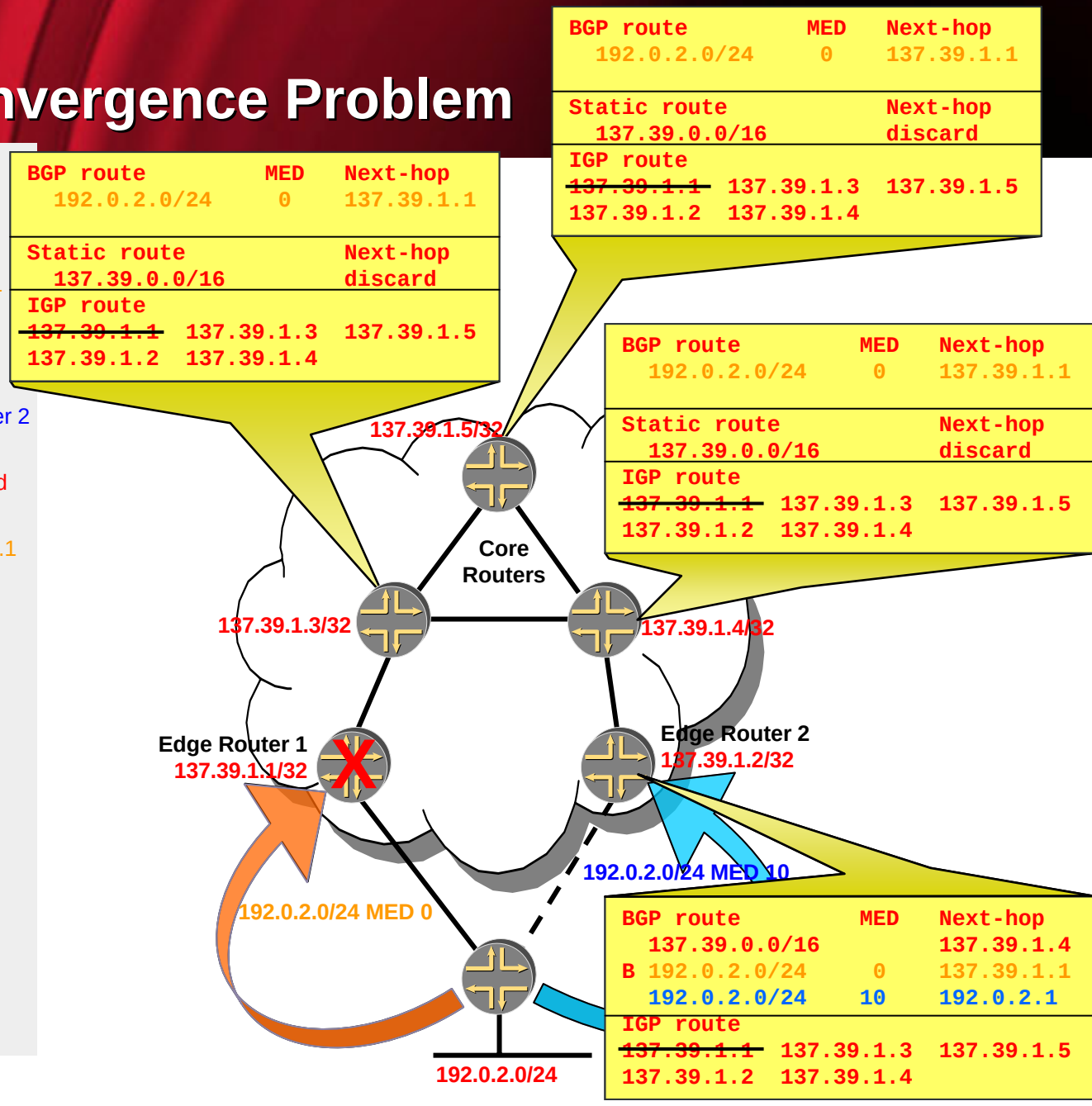
- If a route to a destination has a protocol next-hop that is reachable through a pull-up or less specific route, then the route to that destination will never be invalidated due to next-hop unreachability
- Must wait for the iBGP sessions with the failed edge device to time out (up to 3 min hold timer)
- If your routing table has a less specific route for your BGP protocol Next-hops then you have this problem

BGP Re-convergence Problem

- Take a multi-homed customer with prefix 192.0.2.0/24 connected to two different ISP edge routers (edge router 1 and edge router 2)
- Assume the connection to edge router 1 is a primary link with an eBGP announcement of 192.0.2.0/24 with a MED of 0
- Assume the connection to edge router 2 is a secondary link with an eBGP announcement of 192.0.2.0/24 with a MED of 10
- Assume both edge routers set next-hop self
- Assume that there is a “pull-up” or aggregate route that is less specific than the edge routers’ loopback IP address

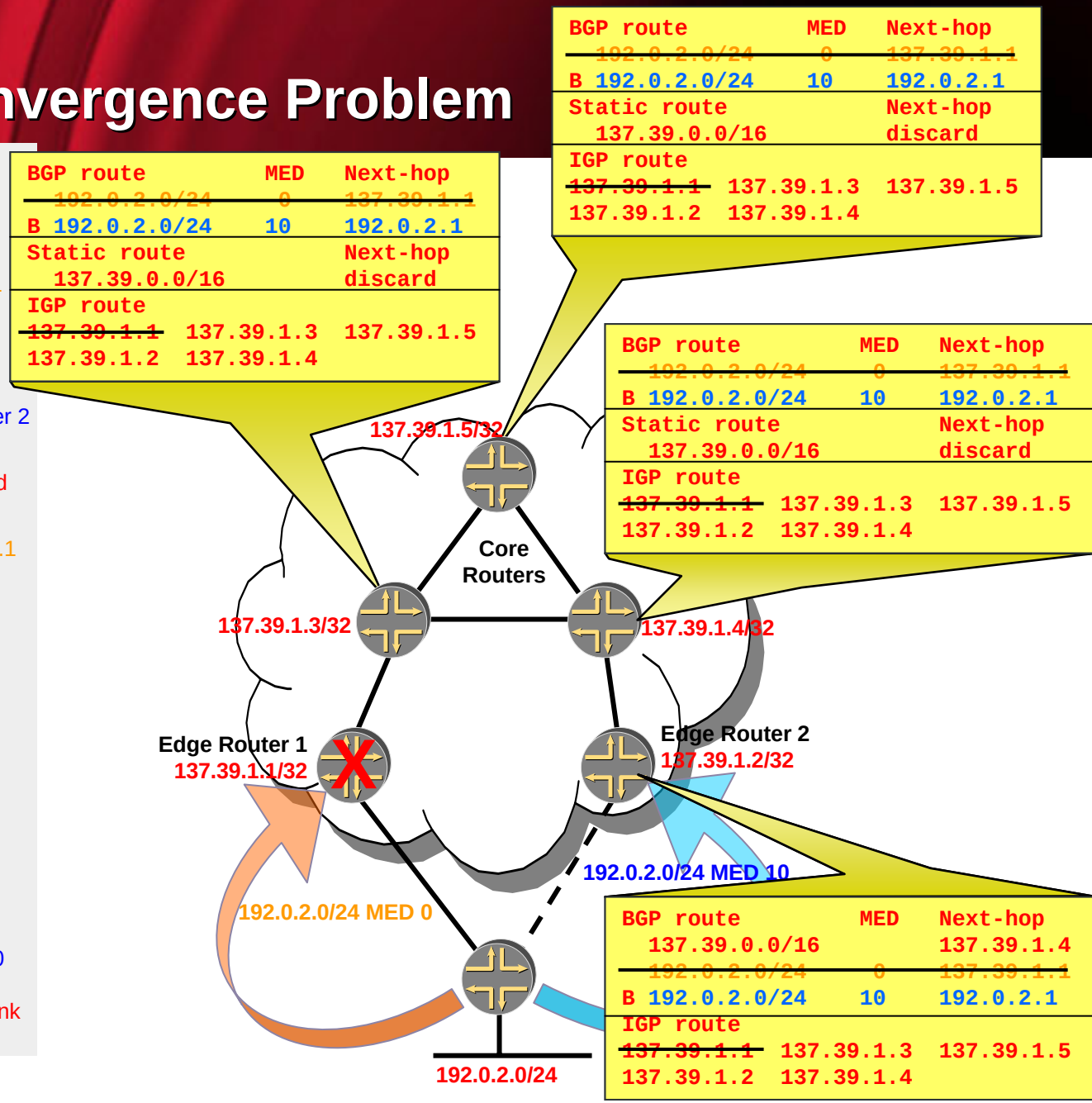
BGP Re-convergence Problem

- Router /32 loopbacks are in the IGP
- 137.39.0.0/16 is a “pull-up” route in the core
- “Pull-up” routes are re-distributed into BGP
- Customer advertises 192.0.2.0/24 MED of 0 via eBGP across primary link to Edge Router 1
- Edge Router 1 advertises 192.0.2.0/24 MED 0 via iBGP and sets next-hop self (137.39.1.1)
- Customer advertises 192.0.2.0/24 MED of 10 via eBGP across secondary link to Edge Router 2
- Edge Router 2 learns 192.0.2.0/24 MED 10 but the best path is 192.0.2.0/24 MED 0 learned via iBGP, so no announcement is made
- All routers select 192.0.2.0/24 MED 0 137.39.1.1 as the best BGP route to the customer
- Edge Router 1 fails
- 137.39.1.1/32 is removed from IGP
- The network still has best BGP route 192.0.2.0/24 MED 0 NH 137.39.1.1
- 137.39.1.1 is reachable through the route to 137.39.0.0/16
- Traffic is drawn to core routers and discarded



BGP Re-convergence Problem

- Router /32 loopbacks are in the IGP
- 137.39.0.0/16 is a "pull-up" route in the core
- "Pull-up" routes are re-distributed into BGP
- Customer advertises 192.0.2.0/24 MED of 0 via eBGP across primary link to Edge Router 1
- Edge Router 1 advertises 192.0.2.0/24 MED 0 via iBGP and sets next-hop self (137.39.1.1)
- Customer advertises 192.0.2.0/24 MED of 10 via eBGP across secondary link to Edge Router 2
- Edge Router 2 learns 192.0.2.0/24 MED 10 but the best path is 192.0.2.0/24 MED 0 learned via iBGP, so no announcement is made
- All routers select 192.0.2.0/24 MED 0 137.39.1.1 as the best BGP route to the customer
- Edge Router 1 fails
- 137.39.1.1/32 is removed from IGP
- The network still has best BGP route 192.0.2.0/24 MED 0 NH 137.39.1.1
- 137.39.1.1 is reachable through the route to 137.39.0.0/16
- Traffic is drawn to core routers and discarded
- After 3 mins the iBGP sessions with Edge Router 1 time out
- Route for 192.0.2.0/24 MED 0 is retracted
- Edge Router 2 route for 192.0.2.0/24 MED 10 is now best. It advertises 192.0.2.0/24 MED 10 via iBGP and sets next-hop self (137.39.1.2)
- Traffic is forwarded to CPE across secondary link



BGP Re-convergence Solution

- BGP next-hops are not aggregated
- The aggregate of the BGP next-hops are not announced to the Internet

- Router /32 loopbacks are in the IGP
- 137.39.0.0/16 is a “pull-up” route in the core
- “Pull-up” routes are re-distributed into BGP

- Customer advertises 192.0.2.0/24 MED of 0 via eBGP across primary link to Edge Router 1

- Edge Router 1 advertises 192.0.2.0/24 MED 0 via iBGP and sets next-hop self (157.130.1.1)

- Customer advertises 192.0.2.0/24 MED of 10 via eBGP across secondary link to Edge Router 2

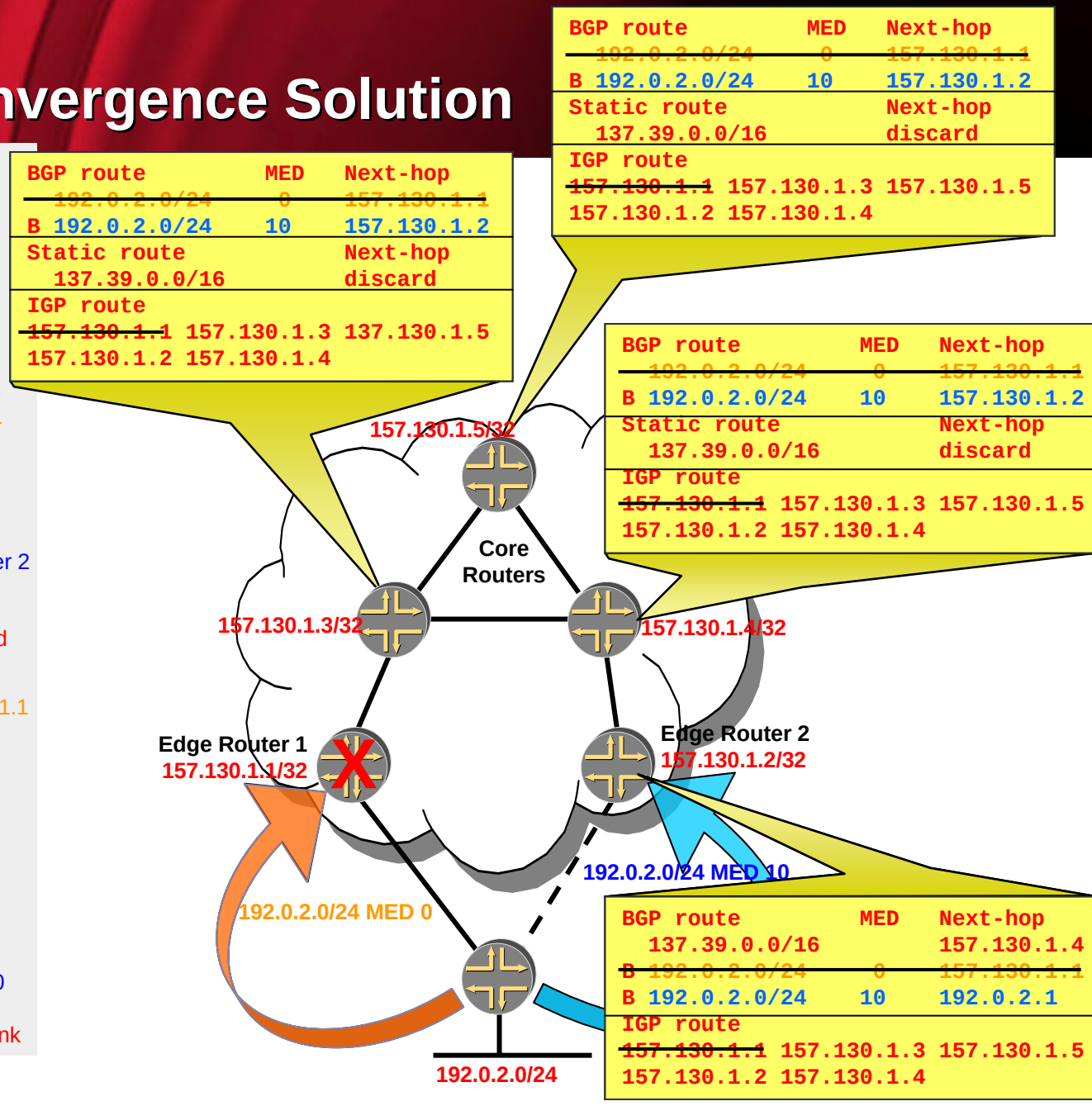
- Edge Router 2 learns 192.0.2.0/24 MED 10 but the best path is 192.0.2.0/24 MED 0 learned via iBGP, so no announcement is made

- All routers select 192.0.2.0/24 MED 0 157.130.1.1 as the best BGP route to the customer

- Edge Router 1 fails

- 157.130.1.1/32 is removed from IGP
- The best BGP route 192.0.2.0/24 MED 0 has an unreachable next-hop (157.130.1.1) and is invalidated

- Edge Router 2 route for 192.0.2.0/24 MED 10 is now best. It advertises 192.0.2.0/24 MED 10 via iBGP and sets next-hop self (157.130.1.2)
- Traffic is forwarded to CPE across secondary link



Solution Considerations and IPv6

- BGP re-convergence solution required non-aggregated prefixes for BGP next-hops
- ARIN policy provides for only a single IPv6 block
- Aggregating a portion of that block either (ex. /28)
  - Creates many prefixes (/29, /30, /31 and a /32 routed on the Internet)
  - Wastes half of the space (/29 aggregated and routed on the Internet)
- Both approaches add to the global routing table, and do not uphold the principle that IPv6 address aggregation is important for IPv6 stewardship

Security Considerations

- Non routed internal only addresses can be used for internal only services
  - iBGP
  - SNMP
  - Radius / TACACS
  - OOB management
- Two tiered approach to network security
  - Can reduce many attacks to internal infrastructure in control plane by not routing the internal address
  - Additional forwarding filters can be easily constructed by the uniqueness of internal only address block

Private Address Considerations

- Private addresses in traceroutes across the public Internet may create confusion
- If routers source ICMP messages with private addresses, and there is wide spread packet filtering of private addresses, then additional problems and confusion may result
- For reverse DNS to work for private addresses requires split plane DNS and hijacking of IANA's authority of the reverse zones

Policy Language Considerations from PPML

- Remove IPv6 references to NRPM 4.4
- Remove references that internal infrastructure MUST NOT be routed on the Internet as ARIN does not set routing policy
- Strengthen references that internal infrastructure MUST NOT be routed on the Internet and the space will be revoked if it is
- Change section 6.10.2 to reflect only root DNS servers need golden space, and possibly anycast
- Add section 6.10.4 discussing RIR and IANA micro-allocations
- Require each type of micro-allocation to have its own unique block