Multihoming with IPv6-to-IPv6 Network Prefix Translation (NPTv6)

draft-bonica-v6-multihome-02

IETF 83

March 29, 2012

Motivation

- Demonstrate a multihoming architecture for sites that deploy NPTv6
- Architectural Goals
  - Redundancy
  - Transport-layer survivability
  - Load balancing
  - Address independence
  - Prevent excessive growth of global routing tables

Motivation (continued)

- Updates Section 2.4 of RFC 6296
  - Demonstrates an alternative multihoming architecture
  - Does not provide transport-layer survivability

NPTv6

- One-to-one translation between inside and outside addresses
  - No attempt to conserve outside address space
- Algorithmic translation
  - Overwrite high order bits
- Stateless translation
- Checksum neutral
- No requirement for routing symmetry
- By default, supports inbound connection requests

Architectural Aspects

- Topology
- Addressing
- Translation
- Routing
- DNS
- Recovery
- Load Balancing

Topology

```
          Upstream                Upstream
       Provider #1             Provider #2
         /     \                 /       \
        /       \               /         \
       /     +------+       +------+       \
  +------+   |Backup|       |Backup|   +------+
  |  PE  |   |  PE  |       |  PE  |   |  PE  |
  |  #1  |   |  #1  |       |  #2  |   |  #2  |
  +------+   +------+       +------+   +------+
      |                                    |
      |                                    |
   +------+                             +------+
   |NPTv6 |                             |NPTv6 |
   |  #1  |                             |  #2  |
   +------+                             +------+
      |                                    |
      |                                    |
------------------------------------------------------
                   Internal Network
```

Provider Addressing

- A Regional Internet Registry (RIR) assigns Provider Address Block (PAB) #1 to Upstream Provider #1
- Likewise, an RIR assigns PAB #2 to Upstream Provider #2
- Both upstream providers assign address space to their customers from their PABs
  - Assignments are called Customer Network Blocks (CNB)

Provider Addressing (continued)

- Upstream Provider #1 assigns CNB #1 to the site
  - CNB #1 is a /59
- Upstream Provider #2 assigns CNB #2 to the site
  - CNB #2 is a /60
- The site does not number any of its resources from CNB #1 or CNB #2
- Site resources are accessed using CNB #1 and CNB #2 addresses

Site Addressing

- Site obtains a Site Address Block (SAB)
  - From ULA or other source
  - SAB is a /58
- SAB Partitioning
  - Lower half: Same size as CNB #1 (/59)
  - Third quarter: Same size as CNB #2 (/60)
  - Fourth quarter

Site Numbering

- Hosts numbered from the lower half of the SAB normally receive inbound traffic from Upstream Provider #1
- Hosts numbered from the third quarter of the SAB normally receive inbound traffic from Upstream Provider #2
- Hosts numbered from the fourth quarter of the SAB are not accessible from outside of the site

Site Numbering (continued)

- Selected hosts can receive inbound traffic from both Upstream Provider #1 and Upstream Provider #2
  - These hosts have multiple SAB addresses
  - At least one address is drawn from the lower half of the SAB
  - At lease one address is drawn from the third quarter of the SAB

Translation: Inbound Traffic

- If the 59 high-order bits of the destination address match the 59 high-order bits of CNB #1, overwrite those bits with 59 bits that identify the lower half of the SAB
- If 60 high-order bits of the destination address match the 60 high-order bits of CNB #2, overwrite those bits with 60 bits that identify the third quarter of the SAB
- Else silently discard

Translation: Outbound Traffic

- If the 59 high-order bits of the source address match the 59 bits that identify the lower half of the SAB, overwrite those bits with the 59 high order bits of CNB #1
- If the 60 high-order bits of the source address match the 60 bits that identify the third quarter of the SAB, overwrite those bits with the 60 high order bits of CNB #2
- Else silently discard

Routing

- PE #1 learns a route to CNB # 1 with next-hop equal to outside interface of NPTv6 #1
  - Overwrites next-hop with its own loopback address
  - Set LOCAL PREF high
  - Distributes throughout Upstream Provider #1 using iBGP
- Backup PE #1 learns a route to CNB # 1 with next-hop equal to outside interface of NPTv6 #2
  - Overwrites next-hop with its own loopback address
  - Set LOCAL PREF high
  - Distributes throughout Upstream Provider #1 using iBGP

Routing Continued

- Upstream Provider #1 advertises PAB #1 into global Internet
  - Nothing else (not the SAB, CNB #1 or CNB #2)
- Upstream Provider #2 deploys similar routing arrangement
- Two default routes circulate within the site
  - Next-hop is inside interface of NPTv6 #1
  - Next-hop is inside interface of NPTv6 #2

DNS

- Site publishes AAAA records associating each resource with all of its CNB addresses
- Sufficient but suboptimal
  - Hairpinning
- Workarounds available
  - Currently deployed in RFC 1918 networks

Recovery

- When PE #1 loses its direct link to NPTv6 #1, it withdraws its advertisement for CNB #1
- Backup PE #1 attracts traffic destined for CNB #1 to itself
- Backup PE #1 forwards traffic to NPTv6 #2 through tunnel
- NPTv6 #2 translates and forwards into site
- Return traffic traverse NPTv6 #2 and Upstream Provider #2
- Convergence is achieved as soon as PE #1’s withdrawal propagates throughout Upstream Provider #1 network

Load Balancing

- Outbound
  - Controlled by site
  - Traffic can exit through either NPTv6 gateway
- Inbound: connections originating within site
  - Originating host selects one of its source addresses
  - Selected address determines path or return traffic
- Inbound: connections originating outside of the site
  - Originating host selects one of the addresses advertised in DNS
  - Selected address determines traffic path
  - Site influences selection by controlling the number of addresses that it advertises from each part of the SAB

Conclusion

- Please provide review
- Please adopt draft
- Intended publication status is EXPERIMENTAL