Controlled IPv6 deaggregation by large organizations

RIPE 69
3 november 2014, London

Iljitsch van Beijnum

The IPv6 routing table today

- Size of the routing table:
  - Currently ~ 19000 prefixes
  - Growing at about 4000 prefixes/year
- However, more specifics are growing at 57% per year:
  - Jan 2013: 3049 of 11500: 27%
  - Jan 2014: 4799 of 16100: 29%

Source: http://www.potaroo.net/presentations/2014-02-09-bgp2013.pdf

An example... (1)

```
*    2001:2B8::/32       0 6939 9957 17832 i
*    2001:2B8:2::/48     0 6939 9957 17832 i
*    2001:2B8:11::/48    0 6939 9957 17832 i
*    2001:2B8:16::/48    0 6939 9957 17832 i
*    2001:2B8:17::/48    0 6939 9957 17832 i
*    2001:2B8:19::/48    0 6939 9957 17832 i
*    2001:2B8:20::/48    0 6939 9957 17832 i
*    2001:2B8:21::/48    0 6939 9957 17832 i
*    2001:2B8:22::/48    0 6939 9957 17832 i
*    2001:2B8:26::/48    0 6939 9957 17832 i
*    2001:2B8:28::/48    0 6939 9957 17832 i
*    2001:2B8:30::/48    0 6939 9957 17832 i
*    2001:2B8:31::/48    0 6939 9957 17832 i
*    2001:2B8:32::/48    0 6939 9957 17832 i
*    2001:2B8:35::/48    0 6939 9957 17832 i
*    2001:2B8:36::/48    0 6939 9957 17832 i
```

An example... (2)

```
*    2001:2B8:37::/48    0 6939 9957 17832 i
*    2001:2B8:39::/48    0 6939 9957 17832 i
*    2001:2B8:40::/48    0 6939 9957 17832 i
*    2001:2B8:43::/48    0 6939 9957 17832 i
*    2001:2B8:45::/48    0 6939 9957 17832 i
*    2001:2B8:48::/48    0 6939 9957 17832 i
*    2001:2B8:49::/48    0 6939 9957 17832 i
*    2001:2B8:50::/48    0 6939 9957 17832 i
*>   2001:2B8:51::/48    0 6939 9957 17832 i
*>   2001:2B8:52::/48    0 6939 9957 17832 i
*>   2001:2B8:53::/48    0 6939 9957 17832 i
*    2001:2B8:90::/48    0 6939 9957 17832 1237 i
*    2001:2B8:94::/48    0 6939 9957 17832 1237 i
*    2001:2B8:9A::/48    0 6939 9957 17832 1237 i
*    2001:2B8:9C::/48    0 6939 9957 17832 1237 i
*    2001:2B8:9D::/48    0 6939 9957 17832 1237 i
```

An example... (3)

```
*    2001:2B8:A0::/48    0 6939 9957 17832 1237 i
*    2001:2B8:A4::/48    0 6939 9957 17832 1237 i
*    2001:2B8:B0::/48    0 6939 9957 17832 1237 i
*    2001:2B8:B2::/48    0 6939 9957 17832 1237 i
*    2001:2B8:B4::/48    0 6939 9957 17832 1237 i
*    2001:2B8:B6::/48    0 6939 9957 17832 1237 i
*    2001:2B8:B8::/48    0 6939 9957 17832 1237 i
*    2001:2B8:BA::/48    0 6939 9957 17832 1237 i
*    2001:2B8:BC::/48    0 6939 9957 17832 1237 i
*    2001:2B8:BE::/48    0 6939 9957 17832 1237 i
*    2001:2B8:C0::/48    0 6939 9957 17832 1237 i
*    2001:2B8:C2::/48    0 6939 9957 17832 1237 i
*    2001:2B8:C4::/48    0 6939 9957 17832 1237 i
*    2001:2B8:C6::/48    0 6939 9957 17832 1237 i
*    2001:2B8:C8::/48    0 6939 9957 17832 1237 i
*    2001:2B8:CA::/48    0 6939 9957 17832 1237 i
```

An example... (4)

```
*    2001:2B8:CC::/48    0 6939 9957 17832 1237 i
*    2001:2B8:CE::/48    0 6939 9957 17832 1237 i
*    2001:2B8:D0::/48    0 6939 9957 17832 1237 i
*    2001:2B8:D2::/48    0 6939 9957 17832 1237 i
*    2001:2B8:D4::/48    0 6939 9957 17832 1237 i
*    2001:2B8:D6::/48    0 6939 9957 17832 1237 i
*    2001:2B8:DC::/48    0 6939 9957 17832 1237 i
*    2001:2B8:E6::/48    0 6939 9957 17832 1237 i
*    2001:2B8:ED::/48    0 6939 9957 17832 1237 i
*    2001:2B8:EF::/48    0 6939 9957 17832 1237 i
*    2001:2B8:F2::/48    0 6939 9957 17832 i
*    2001:2B8:200::/48   0 6939 9957 17832 i
*    2001:2B8:380::/48   0 6939 9957 17832 1237 i
```

An example... (5)

```
inet6num:       2001:02B8::/32
netname:        NGINET-KRNIC-KR-20010115
descr:          NGInet(Next Generation Internet Network) is
                the national-wide
descr:          Internet service provider for public
                oganizations
country:        KR
```

What is this?

- Traditionally, types of addresses:
  - Provider Aggregatable (PA): used by ISPs
  - Provider Independent (PI): used by end users
- However, large organizations find it useful to have one big PA-like prefix
- But: their offices connect to different ISPs!
  - because they operate in many countries
  - or they have largely independent subunits

So: deaggregation

- So organizations such as:
  - Big multinationals
  - Governments
- Become "enterprise LIRs" and obtain a PA prefix
- Then subunits advertise deaggregates / more specifics of that PA block
  - Towards different ISPs
  - In different locations

Is this a problem for the internet community?

- Not today!
  - IPv6 table is still small
- But people get large blocks so possible to source many deaggregates
  - No obvious way to filter on prefix length
- IPv6 is going to be around for a long time
- IPv4 has shown that mistakes early on are hard to clean up later

Does this work well for those organizations?

- Mostly
- However, deaggregates may be filtered
  - Filtering is inconsistent because there is no agreed "safe" prefix length for IPv6
  - (like /24 in IPv4)

The alternative: PI

- Having many deaggregates in the IPv6 routing table is not great
- But what if they request a PI prefix for each office/organizational subunit?
  - Same effect on the routing table
    - but no opportunities to filter/aggregate
  - Worse for the organization: no easy way to identify the organization's address space

What do we do?

- Nothing?
  - Suboptimal for routing table size
  - Suboptimal for the organizations involved
  - May even hinder IPv6 deployment?
- Start a conversation between enterprise LIRs and network operators?
  - Give enterprise LIRs guidance on what will work
  - Give network operators tools to control table size

The document

- You need a "draft" to get discussion in the IETF
  - So I wrote draft-van-beijnum-grow-controlled-deagg-00
  - started a discussion on the v6ops (IPv6 operations) and grow (global routing operations) working group mailing lists
- This does not mean the IETF is the best or only place to have this discussion
  - (But IETF is global, RIRs are regional...)

The idea

- Allow enterprise LIRs to set up an "aggregate of last resort" (AoLR)
  - So traffic has a place to go if deaggregates are filtered
- Tag deaggregates with BGP communities
  - Indicate that it's safe to filter if needed
  - Indicate where the deaggregate comes from
    - May want to allow "close" deaggregates but filter ones from far away

Aggregate of last resort

- ISP A injects the entire prefix in BGP
- ISPs B, C, D, ... (and maybe A) provide connectivity towards subunits of the organization
- B - D interconnect with A
- A accepts the deaggregates from B - D
- So the rest of the internet delivers packets to A
- A hands over the packets to B - D
  - so A only carries the packets a relatively short distance

Aggregate of last resort (2)

- This works well if A is a large world-wide network
- However, B - G can be smaller regional or national networks
- A would have to be paid to provide this service
  - But can now be held accountable!
- (Multiple ISPs can provide the AoLR service if desired)
- Rest of the internet can safely filter the deaggregates

Location in BGP community

- A BGP "community" is simply a label attached to a prefix
  - 702:120 or NO_EXPORT
- In Europe we probably don't care about Korean deaggregates
- We Europeans just send the traffic in the general direction of Korea and once the packets get closer, the deaggregates will be there

Location in BGP community (2)

- GPS coordinates in BGP communities
  - Precision is 1 degree, ~ 100 km
- Not subject to change or political controversy!
- *Somewhat* human readable/understandable:
  - Example: xxxx1:53013

- Maybe express filters as geographic areas in the future if router vendors add this to their routers
- But can work <u>today</u>!

Selective filtering

- Everyone decides which deaggregates to carry
  - Big routers? Maybe carry them all
  - Small routers? Maybe carry none of them
  - Regional network? Maybe only carry deaggregates announced in the region
  - World-wide network? Maybe each router only carries deaggregates announced in the same region
    - so the network as a whole carries all deaggregates
    - but individual routers don't

Questions?