

IP Address Aggregation Practices

James Rice

<james_r-uknof@jump.org.uk>

UKNOF 1

London, Wednesday 25th May 2005

Overview

- This is not about when we are going to run out of IPv4 space, see Geoff Huston presentations about that.
- This is about best practice for aggregation, and starting off investigating ways to encourage aggregation / discourage the current amount of de-aggregation
- Not just a one way presentation. Intended to spark a discussion forum regarding aggregation and filtering.

What we should do

- Aggregate our PA space as far as is possible, at least to RIR allocation boundaries
- If more specifics are needed for load balancing across multiple links to an upstream, then still announce the aggregate, and tag the more specifics with no-export
- If you're going to deploy a new site that isn't connected to your network, get some PI space, don't de-aggregate your PA space.

Current table state

- 179,121 prefixes visible at route-views
- 162,634 prefixes from APNIC analysis
- 158,685 prefixes from CIDR report
- 143,675 prefixes accepted by Jump
- 94,526 prefixes after maximum aggregation

Prefix-length filtering

- RIRs publish a list of minimum allocation sizes for each given /8 they allocate from.
- <http://www.apnic.net/db/min-alloc.html>
- <https://www.ripe.net/ripe/docs/smallest-alloc-sizes.html>
- <http://www.arin.net/statistics/index.html#cidr>

Filtering based on RIR allocations

- Verio did this (Randy Bush era), Verio carried about 65,000 prefixes with few problems, when the global table size was in excess of 100K prefixes.
- Requires care, to minimise collateral damage
- Little business justification for doing this, as benefits aren't seen right away, but problems might be.
- Engineering justification isn't business justification.

Why bother aggregating?

- Announcing aggregates of at least the RIR minimum allocation size for that /8 allows networks to apply prefix-length filters, which reduces the chances of accidental or malicious more specific announcing having a large effect.
- Decreases convergence times
- Increases the lifespan of hardware –
Memory, Forwarding Database limit.

Malicious / Accidental Hijacking

- Case study: BBC
- Announcing 212.58.192.0/19
- One morning, 212.58.240.0/24 starts being announced somewhere in Asia. This contains the New York www.bbc and news.bbc server farms.
- All heck breaks loose.
- Widespread filtering on RIR boundaries would have seriously limited this damage.

Convergence times

- Case study: Cisco Sup2A/MSFC2
- Hardware forwarding performance may be great, but:
- Takes ~11 seconds to populate the hardware with 100K forwarding entries.

Hardware life

- Example: Cisco SUP720
- Recent, modern equipment
- Maximum 256K HW IPv4 Unicast routes

We can at least

- Educate our transit customers
- Ensure any prefixes we transit are aggregated as much as possible
- Attempt to engage clue with any peers announcing routes which aren't aggregated as much as possible.

Co-ordinated filtering

- Is anyone interested in getting together for producing a co-ordinated prefix-length based filter list, which could be applied by interested networks, and has a maximal gain in excess route reduction with a minimal of collateral damage?

Discussion

- Thoughts?
- Comments?
- Please discuss