Open issues with ipv6 routing/multihoming

Vince Fuller, Dave Meyer

Session Objectives

- A brief look at how we got where we are today
- Define “locator”, “endpoint-id”, and their functions
- Explain why these concepts matter and why this separation is a good thing
- Understand that IPv4 and ipv6 co-mingle these functions and why that doesn’t scale
- Determine if this community is interested in looking at a solution to the scaling problem

A brief history of Internet time

- Recognition of exponential growth – late 1980s
- CLNS as IP replacement – December, 1990 IETF
- ROAD group and the “three trucks” – 1991-1992
  - Running out of “class-B” network numbers
  - Explosive growth of the “default-free” routing table
  - Eventual exhaustion of 32-bit address space
  - Two efforts – short-term vs. long-term
  - More at “The Long and Winding ROAD” http://rms46.vlsm.org/1/42.html
- Supernetting and CIDR – 1992-1993

A brief history of Internet time (cont'd)

- IETF "ipng" solicitation – RFC1550, Dec 1993
- Direction and technical criteria for ipng choice – RFC1719 and RFC1726, Dec 1994
- Proliferation of proposals:
  - TUBA – RFC1347, June 1992
  - PIP – RFC1621, RFC1622, May 1994
  - CATNIP – RFC1707, October 1994
  - SIP – RFC1710, October 1994
  - NIMROD – RFC1753, December 1994
  - ENCAPS – RFC1955, June 1996
- Choice came down to politics, not technical merit
  - Hard issues deferred in favor of packet header design

Identity - “what’s in a name”?

- Think of an “endpoint-id” as the “name” of a device or process that is communicating over a network
- In the real world, this is something like “Dave Meyer” - “who” you are
- A “domain name” can be used as a human-readable way of referring to an endpoint-id

Desirable properties of endpoint-IDs

- **Persistence: long-term binding to the thing that they name**
  - **These do not change during long-lived network sessions**
- **Ease of administrative assignment**
  - **Assigned to and by organizations**
  - **Hierarchy is along these lines (like DNS)**
- **Portability**
  - **IDs remain the same when an organization changes provider or otherwise moves to a different point in the network topology**

Locators – “where” you are in the network

- Think of the source and destination “addresses” used in routing and forwarding
- Real-world analogy is street address (i.e. 3700 Cisco Way, San Jose, CA, US) or phone number (408-526-7128)
- Typically there is some hierarchical structure (analogous to number, street, city, state, country or NPA/NXX)

Desirable properties of locators

- Hierarchical assignment according to network topology (“isomorphic”)
- Dynamic, transparent renumbering without disrupting network sessions
- May be abstracted to reduce unwanted state
  - Real-world analogy: don’t need to know exact street address in Australia to travel toward it from San Jose
  - Variable-length addresses or prefixes, etc.
- Possibly applied to traffic without end-system knowledge (effectively, like NAT but without breaking the sacred End-to-End principle)

Why should I care about this?

- In IPv4 and ipv6, there are only “addresses” which serve as both endpoint-ids and locators
- This means they don’t have the desirable properties of either:
  - Assignment to organizations is painful because use as locator constrains it to be topological (“provider-based”)
  - Exceptions to topology create additional, global routing state - multihoming is painful and expensive
  - Renumbering is hard – DHCP isn’t enough, changing address disrupts sessions, weak authentication used, source-based filtering, etc.
- Doesn’t scale for large numbers of “provider-independent” or multi-homed sites

Why should I care (continued)?

- **The really scary thing is that the scaling problem won't become obvious until (and if) ipv6 becomes widely-deployed**
  - **Larger ipv6 address space could result in orders of magnitude more prefixes (depending on allocation policy, provider behavior, etc.)**
  - **NAT is effectively implementing id/locator split – what happens if the ipv6 proponents' dream of a "NAT-free" Internet is realized?**
  - **Scale of IP network is still relatively small**
  - **Re-creating the "routing swamp" with ipv6 would be… ugly/bad/disastrous; it isn't clear what anyone could do to save the Internet if that happens**
- **Sadly, this has been mostly ignored for 10 years**

Can ipv6 be fixed? (and what is GSE, anyway?)

- Can we keep ipv6 packet formats but implement the identifier/locator split?
- Mike O'Dell proposed this in 1997 with 8+8/GSE

  http://ietfreport.isoc.org/idref/draft-ietf-ipngwg-gseaddr
- Basic idea: separate 16-byte address into 8-byte EID and 8-byte "routing goop" (locator)
  - Change TCP/UDP to only care about EID (requires incompatible change to tcp6/udp6)
  - Allow routing system to modify RG as needed, including on packets "in flight", to keep locators isomorphic to network topology

GSE benefits

- Achieves goal of EID/locator split while keeping most of ipv6 and without requiring a new database for EID-to-locator mapping
- Allows for scalable multi-homing by allowing separate RG for each path to an end-system; unlike shim6, does not require transport-layer complexity to deal with multiple addresses
- Renumbering can be fast and transparent to hosts (including for long-lived sessions) with no need to detect failure of usable addresses

GSE issues

- Incompatible change needed to tcp6/udp6
  - in 1997, no installed base and plenty of time for transition
  - may be more difficult today
- Purists argue violation of end-to-end principle
- Perceived security weakness of trusting “naked” EID (Steve Bellovin says this is a non-issue)
- Mapping of EID to EID+RG may add complexity to DNS, depending on how it is implemented
- Scalable TE not in original design; will differ from IPv4 TE, may involve “NAT-like” RG re-write
- Currently not being pursued (expired draft)

What about shim6/multi6?

- Approx 3-year-old IETF effort to retro-fit an endpoint-id/locator split into the existing ipv6 spec
- Summary: end-systems are assigned an address (locator) for each connection they have to the network topology (each provider); one address is used as the id and isn't expected to change during session lifetimes
- A "shim" layer hides locator/id split from transport (somewhat problematic as ipv6 embeds addresses in the transport headers)
- Lots of complexity around locator pair selection, addition, removal, testing of liveness, etc.

What about shim6/multi6? (continued)

- Some perceive as an optional, “bag on the side” rather than a part of the core architecture… but maybe that is just us
- What do you, the SP community, think of shim6? Will it solve your problems and help make ipv6 both scalable and deployable in your network?
- Feedback thus far (especially on NANOG mailing list): nobody seems to like it
  - SP objection: doesn’t allow site-level traffic-engineering in manner of IPv4; TE may be doable but will be very different and will add greater dependency on host implementations and administration
  - Hosting provider objection: requires too many addresses and too much state in web servers
  - End-users: still don’t get “provider-independent addresses” so still face renumbering pain

What if nothing is changed?

- How about a “thought experiment”?
- Make assumptions about ipv6 and Internet growth
- Take a guess at growth trends
- Pose some questions about what might happen
- What is the “worst-case” scenario that providers, vendors, and users might face?

My cloudy crystal ball: a few assumptions

- ipv6 will be deployed in parallel to IPv4 and will be widely adopted
- IPv4 will be predominant protocol for near-to-mid term and will continue to be used indefinitely
- IPv4 routing state growth, in particular that for multi-homed sites, will continue to grow at a greater than linear rate up to or beyond address space exhaustion; ipv6 routing state growth curve will be similar - driven by multihoming
- As consequence of above, routers in the “DFZ” will need to maintain full routing/forwarding tables for both IPv4 and ipv6; tables will continue to grow and will need to respond rapidly in the face of significant churn

A few more assumptions

- prefix assignments will be large enough to allow virtually all organizations to aggregate addresses into a single prefix; in only relatively few cases (consider acquisitions, mergers, etc.) will multiple prefixes need to be advertised for an organization into the “DFZ”
- shim6 will not see significant adoption beyond possible edge use for multi-homing of residences and very small organizations
- IPv4-style multi-homing will be the norm for ipv6, implying that all multi-homed sites and all sites which change providers without renumbering will need to be explicitly advertised into the “DFZ”

A few more assumptions

- as the Internet becomes more mission-critical a greater fraction of organizations will choose to multi-home
- IPv4-style traffic engineering, using more-specific prefix advertisements, will be performed with ipv6; this practice will likely increase as the Internet grows
- Efforts to reduce the scope of prefix advertisements, such as AS_HOPCOUNT, will not be adopted on a large enough scale to reduce the impact of more-specifics in the "DFZ"

Questions to ask or worry about

- How much routing state growth is due to organizations needing multiple IPv4 prefixes? Some/most of these may be avoided with ipv6.
- As a result of available larger prefixes, will the number of prefixes per ASN decrease toward one? What is the likelihood that ASN usage growth will remain linear? (probably low)
  - Today, approximately 30,000 ASNs in use
- How much growth is due to unintentional more-specifics? These may be avoided with ipv6.
- How much growth is due to TE or other intentional use of more-specifics? These will happen with ipv6 unless draconian address allocation rules are kept (which is unlikely)
  - This appears to be an increasing fraction of the more-specifics
- What's the routing state "churn rate" and is it growing, shrinking, or remaining steady? (growing dramatically)

More to worry about

- How bad are the growth trends?
  - Prefixes: 100K to 170K in 2005
    - ➢increase to ~370K within 5 years
    - ➢global routes only – each SP has additional internal routes
  - Churn: 0.7M/0.4M updates/withdrawals per day
    - ➢increase to 2.8M/1.6M within 5 years
  - CPU use: 30% at 1.5Ghz (average) today
    - ➢increase to 120% within 5 years
- These are guesses based on a limited view of the routing system and on low-confidence projections (cloudy crystal ball); the truth could be worse, especially for peak demands
- See Geoff's and Jason's presentations for more numbers
- Trend lines look exponential or quadratic; this is bad…

What's next?

- **Is there a real problem here or are we worrying about nothing?**
- **Is it worth doing an IAB-sponsored experiment, workshop, or other IETF-sanctioned activity along these lines to re-examine GSE or explore other solutions?**
- **Is the Internet operations community interested in looking at this problem and working on a solution?**
- **Should we socialize this anywhere else?**
- **Any other suggestions?**
- **Read and join the discussion at**

  architecture-discuss@ietf.org

Recommended Reading

"Endpoints and Endpoint names: A Proposed Enhancement to the Internet Architecture", J. Noel Chiappa, 1999, http://users.exis.net/~jnc/tech/endpoints.txt

"On the Naming and Binding of Network Destinations", J. Saltzer, August, 1993, published as RFC1498, http://www.ietf.org/rfc/rfc1498.txt?number=1498

"The NIMROD Routing Architecture", I. Castineyra, N. Chiappa, M. Steenstrup. February 2006, published as RFC1992, http://www.ietf.org/rfc/rfc1992.txt?number=1992

"2005 – A BGP Year in Review", G. Huston, APRICOT 2006, http://www.apnic.net/meetings/21/docs/sigs/routing/routing-pres-huston-routing-update.pdf