

# Routing Scalability: Separation or Elimination?

---

Dan Jen, Daniel Massey, Michael Meisel, Lan Wang, Beichuan Zhang, and Lixia Zhang

Talk by Lixia Zhang  
Routing Research Group  
IETF72, August 1, 2008

# The Goal of This Talk

Contributing to RRG's goal of design convergence  
(with acknowledgment to many people for input)

- Understanding the design space by carefully studying all proposals
- Identify commonality/differences at the highest branching points
- This talk does not describe new design, but to
  - Solicit consensus on direction to march forward
  - Articulate an overall task list (for later discussion)

# Scalable Routing

- Being able to control the scale of the routing system
  - The ability to control, rather than any specific numbers
- Allowing the global transit core to route on aggregatable prefixes only
  - Provider-assigned (PA) addressing
- Two ways to get there
  - Separation
  - Elimination

# Separation

- Separating edge prefixes from the transit core
- Edge network prefixes removed from global routing system
  - APT, IVIP, LISP, TRRP: Map & Encap
  - Six/One Router: map & translation
- Requires a mapping system to glue the edges to the middle

# Elimination

- Pushing multiple PA addresses all the way into the hosts of multihomed sites
- SHIM6
  - Multiple PA addresses stop at shim layer in a host
  - Lots of hard work has been done here
- Multipath transport
  - As in Mark Handley's proposal: transport layer can make use of multiple PA addresses
- ILNP

# Separation, or elimination?

If separation:

- Need to work out a mapping system design
  - Map an edge destination address to the edge network's attachment point to the transit core
  - Mapping info must be distributed to all entry points to the core
- Need to decide between encapsulation and translation

# Separation, or elimination?

- Need to develop effective detection and recovery mechanism for failures occurring between the core and edge networks, because they are
  - longer reflected in global routing
  - proposed not to be reflected in the mapping system
  - Need solutions that can detect failures and switch to alternate path promptly (whenever available)
- Need an incremental deployment plan

# Separation, or elimination?

If elimination:

- No new work need to be done at network layer
- however there is a conservation of hard work
  - Effective handling of multiple addresses by host/transport
  - host changes
  - site renumbering when changing providers

# Which way to go?

- Some people believe renumbering is nonstarter
- Some people believe all hosts can be changes within reasonable time frame
- The real answer: The future is uncertain

# If we choose elimination

- And indeed all edge networks take in PA addresses in next 5 years
- We would succeed without working hard!
  - Of course Handley and friends will work hard to roll out multipath transport, and
  - Sites will have to adopt multiple-addressing and renumbering
- But what if we guessed wrong?
  - In next 5 years
    - IPv4 routing table will continue to grow
    - IPv6 deployment would progressing
    - Interent could be facing routing scalability crisis...

# If we choose separation

- We will have to work really hard to solve the multiple major challenges
- If we choose wrong: all the hard work would be wasted!
  - But we don't do any worse than that
- If we choose right: the hard work will be worthwhile
  - Resolving a decades long problem

See <ftp://ftp.ietf.org/ietf-online-proceedings/95jul/presentations/allocation/pre.allocation.txt>

# IETF33 Plenary: IP Address Allocation

## (July 1995)

- up to now, the IP address has served as an invariant, unique identification for the end host. TCP design makes use of this assumption, so do many other protocols and applications.
- As a result, nobody today has a complete list of all the possible places in the protocol architecture that have the IP address hard wired or embedded in it.
- Therefore, contradicting Peter(Ford)'s assumption that most customers do not care about permanent IP addresses, dynamically changing addresses, as required by provider-based assignment, changes the architecture we used to know and causes serious problems at the user ends.

# Additional Benefits from Separation

- Raise the barrier against attacks
  - Unwanted traffic could be filtered at the border between edge networks and the transit core
  - Prevent end hosts from sending directly to transit routers
- Provide scaling benefits while multipath transport research getting ready over time
- Scale routing *without dependency* on the assumption that all/majority sites would adopt PA addresses in any given time frame (if ever)

# Costs of Separation

- As mentioned earlier
  - significant amount of work to be done!
  - New complexity into the system
    - Potentially also help reduce some of existing complexity
- We are mindful of this; research challenge for coming months and years
- Good engineering design can hopefully enable a gradual roll out with visible benefit along the way

# We also want the benefits of Multipath Transport

- End hosts can use multiple paths simultaneously
- End hosts can choose their favorite path(s)
- End hosts see the end-to-end picture in load balancing
- End hosts can adapt to changes quickly
- End-to-end resilience against individual path failures

# Separation + Multipath Transport

- It is important to support multipath transport
- Separation works well with Multipath transport
  - Edge multihomed site can split its prefix into multiple subprefixes
  - Each subprefix corresponds to one of the site's providers
- Use separation for scalability
  - Map each PI subprefix to the desired provider
  - PI prefixes still stay out of the global routing table

# Benefits of Combining Separation with Multipath Transport

- Get all of the benefits of both schemes
- For solving routing scalability, cost is aligned with benefit
- Separation enables edge networks to engineer traffic under multipath transport
  - Edge networks decide what path choices their end users have
  - Independent of what providers they connect to

# “Mobility as next BIG thing”

“Support 10 billion flying toasters”

- If we choose separation,
- We will design a mapping system
- can it be used to support mobility?
  - Different opinions exist
  - Further investigation needed

# Putting all pieces together: Tasks

(first identify tasks, then figure out who owns what)

- Develop a separation solution
  - work out a mapping system design
  - figure out Encapsulation, or Translation
  - develop effective failure detection and recovery mechanisms
- Multipath transport progressing in parallel
  - Maximizing the benefit of resource pooling
- Clarify name space: separating the two
  - Node identifiers ← we do not have one deployed today (if DNS name does not fit the bill); do we need one tomorrow? How to get it? (and what properties should they have?)
  - IP address (further separation between edge/transit addr's?)
- Understand the interplay among the above
- Reach consensus, start drafting working plan

# Thank You!

- Questions? Comments?