APT Incremental Deployment

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

http://www.cs.ucla.edu/~meisel/draft-apt-incremental-00.txt
Why This Talk

• Incremental deployability is one of major factors in APT design

• Something useful learned from the exercise
  – still in exploration/comparison stage
  – So this talk differs from the earlier draft...

• Come here to share and discuss

• Feedback most welcome
Basic Ideas for Incremental Deployment

• Align benefit with deployment cost: ISPs benefit, they should deploy

• Day-0
  – Must be a unilateral decision to turn on APT
    • Map-n-encap: need both tunnel points under one party’s control
  – Must provide incentives for the first mover
    • Being able to reduce BGP table size: remove internal customers’ prefixes from routing to mapping
Day-1: Expect a few APT regions in a BGP world

- Benefit first mover: remove an APT island’s internal customers’ prefixes from routing to mapping
- No harm/no change to the rest: inject those prefixes into BGP table outside APT island
Terminology

• APT AS: A transit AS that has deployed APT
• APT Island: A topologically connected set of APT ASes
  – The smallest possible island: a single AS
• Island Mapping Table: all the mapping entries for the customer sites of a given APT island
  – Each entry maps an edge network prefix to their APT provider ETRs
  – Every APT AS in the island stores the full island mapping table
Benefits for a Single Island

- Prefixes for Sites 1, 2, 3, and 4 removed from ISP1 and ISP2’s BGP tables
  - Potentially large reduction in BGP table size
    - The reduced entries moved to the mapping table
- Offer benefits to such customers (next slide)
Benefits to Edge Networks

• For edge networks with only APT providers
  – Provider-independent addressing
  – Can explicitly express traffic engineering preferences
    (accommodates edge multihoming with both APT and non-APT providers)

• No changes required in edge networks
  – APT is deployed entirely in transit networks

• Some cost to transit ASes
  – Management of APT Default Mappers
  – Additional complexity of APT/BGP interactions
APT and Non-APT Interaction: Non-APT to APT (1/2)

- How can Site5 reach Site3?
  - Site3’s prefix is in APT Island 1’s island mapping table
  - But Site5 and ISP3 don’t understand APT
  - ISP2 must announce Site3’s prefix into BGP
• How Can Site5 reach Site3?
  - ISP2’s Default Mapper (DM) gets Site3’s mapping
  - ISP2’s DM announces all APT edges’ prefixes into BGP
  - ISP3 receives and propagates the routes via BGP
APT and Non-APT Interaction: APT to Non-APT (1/2)

How can Site1 reach Site5?
- Site1 routes packets through ISP1
- But Site5 is not in the APT mapping table...
• How can Site1 reach Site5?
  • PE routers and DMs in APT Island 1 still have BGP tables that store non-APT prefixes
  • ISP1 forwards packets to the BGP next hop
Communication between APT Islands (1/2)

• How can Site1 and Site7 communicate?
  – Both Site1 and Site7 are connected to APT islands
  – Isolated APT islands don’t share mappings
  – ISP1 does’t have an APT mapping table entry for Site7
Communication between APT Islands (2/2)

• How can Site1 and Site7 communicate?
  - ISP3 has a BGP route to Site7
  - ISP1 learns a BGP route to Site7 from ISP3
  - ISP1 can route to Site7 using the BGP route
    (ISP1 does not know or care that Site7 is in an APT island)
APT Islands merge ➔ even smaller BGP tables

- APT ASes in the same island have the same island mapping table
  - APT ASes in different islands do not (for now)
- Topologically connected islands can merge
  - Their mapping tables merge
  - BGP tables at all the routers in the island shrink
- Future work: allowing topologically unconnected islands to merge
  - eliminate separate islands
Inside an APT Island

• Nodes labeled “X” are border routers
• Nodes labeled “DM” are default mappers
• How does the ITR decide where to forward a packet?
  – We are currently examining a few alternatives
  – The following is our favorite scheme as of now
    • Note that this is different from our draft
Sending to Non-APT Networks

- **Site1 to Site5**
  - Site5 is not attached to an APT network
  - The ITR has a BGP route to Site5
  - Packets are simply routed via BGP (not tunneled)
Sending to Non-APT Networks

• Site1 to Site5
  - Site5 is not attached to an APT network
  - The ITR has a BGP route to Site5
  - Packets are simply routed via BGP (not tunneled)
• Site1 to Site5
  - Site5 is not attached to an APT network
  - The ITR has a BGP route to Site5
  - Packets are simply routed via BGP (not tunneled)
• Site1 to Site5
  - Site5 is not attached to an APT network
  - The ITR has a BGP route to Site5
  - Packets are simply routed via BGP (not tunneled)
Sending to Non-APT Networks

- **Site1 to Site5**
  - Site5 is not attached to an APT network
  - The ITR has a BGP route to Site5
  - Packets are simply routed via BGP (not tunneled)
Sending to Non-APT Networks

• Site1 to Site5
  - Site5 is not attached to an APT network
  - The ITR has a BGP route to Site5
  - Packets are simply routed via BGP (not tunneled)
• ITR receives a packet for Site4
  - First look in the BGP table
  - Site4 is only connected to the APT island; ITRs in the island don’t keep Site4’s prefixes in their BGP tables
ITR Lookups for APT-Only Sites

- ITR receives a packet for Site4
  - Next check mapping cache
  - On the event of a cache miss, use the DM
  - This is normal APT behavior
• Assume there is a cache hit
  – Tunnel to the ETR
ITR Lookups for APT-Only Sites

- Assume there is a cache hit
  - Tunnel to the ETR
ITR Lookups for APT-Only Sites

- Assume there is a cache hit
  - Tunnel to the ETR
ITR Lookups for Sites Multihomed with APT and Non-APT Networks

- What if Site4 multihomes with ISP3?
ITR Lookups for Sites Multihomed with APT and Non-APT Networks

- ITR receives a packet for Site4
  - Now ITRs in the island have a BGP route to Site4
  - But we don’t want ITRs to only use the route through ISP3…
Recall: DM generates BGP announcements for all prefixes in the mapping table
- Special tag for sites multihomed with APT and non-APT nets
- ITRs store these in their RIB-In

- But still drop BGP routes for APT-only sites, which use a different tag
ITR Lookups for Sites Multihomed with APT and Non-APT Networks

- If the ISP3 BGP route is preferred
  - Forward using BGP table
- If the APT BGP route is preferred
  - Normal APT behavior
ITR Lookups for Partial APT Networks

- What if ISP1 starts with partial APT deployment?
  - Now Site1 and Site2 are APT-only site
  - Site3 is non-APT site
• DM injects Site1 & 2’s prefixes into BGP
  – So all other sites can reach them
• Data between Site1-Site2 is tunneled
• Data to all other destinations use BGP table to send
Thank You!

• Questions?