On-Demand Dynamic Route Optimization Between Tunnel Endpoints

IETF RTGWG – March 28, 2011

Fred L. Templin
Boeing Research & Technology
fred.l.templin@boeing.com

Document Status

- **New Section 5.14 of ‘draft-templin-intarea-vet’**
- **Submitted March 14, 2011**
- **Originally specific to ISATAP tunnels**
- **Now being generalized to route optimization on any large links**

Route Optimization for Large Links - Problem Statement

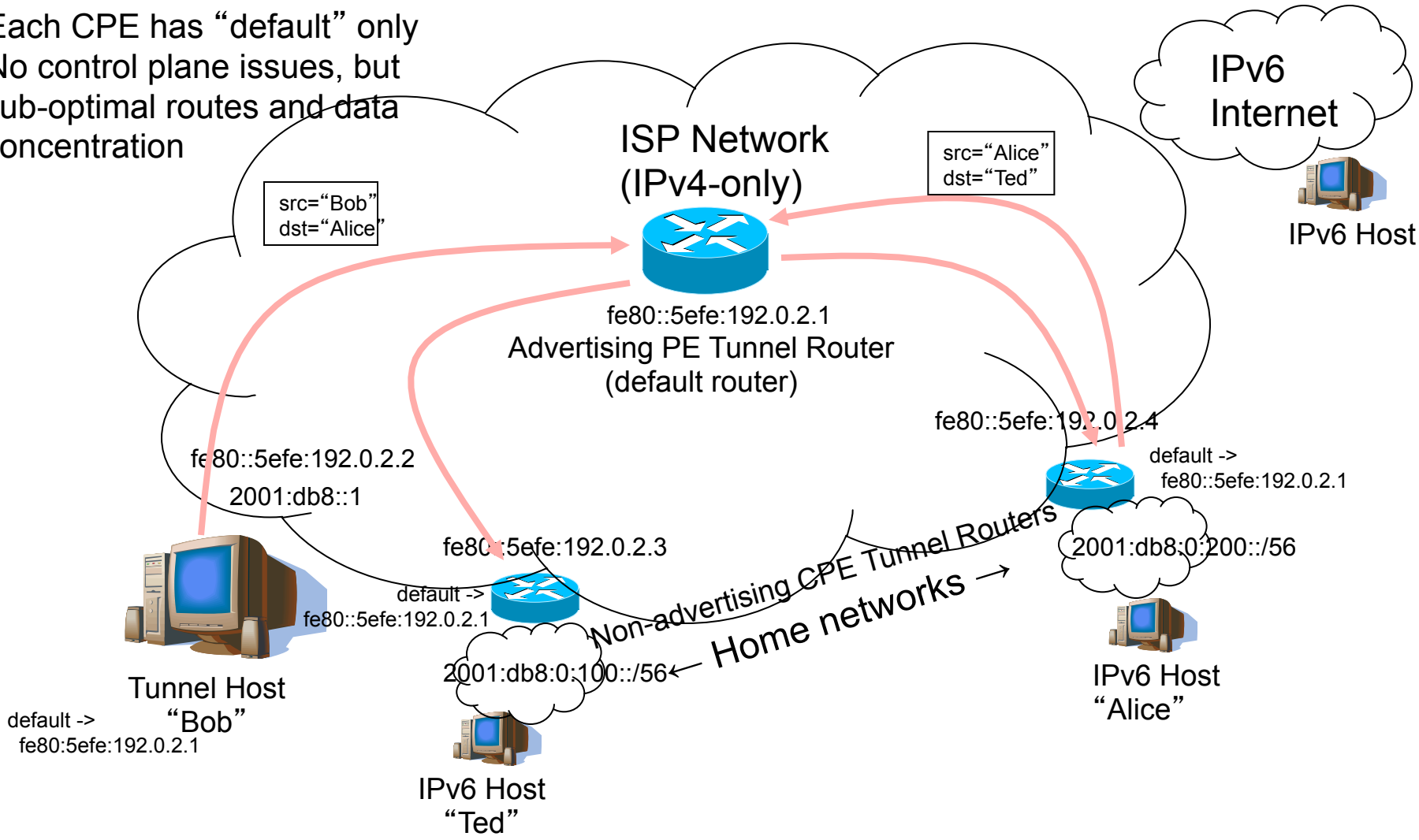
- **Some links connect many hosts and routers**
  - **large campus LANs**
  - **bridged enterprise networks**
  - **cellular provider networks**
  - **aviation networks**
  - **disaster relief / defense networks**
  - **Non-Broadcast, Multiple Access (NBMA) tunnels (e.g., ISATAP, 6over4, 6rd, vet, etc.)**
- **Traditional IGPs (e.g., RIPng, OSPFv3, etc.) don't scale well when the number of routes/routers is very large, or when many nodes are mobile**
- **Default routing via "hub" routers works, but:**
  - **Triangular routing**
  - **Hubs forced to bear load in the data plane**

Specific Example – Tunnels over ISP Network

- Many ISP networks still IPv4-only
- Growing customer requirement for IPv6
- IPv6-in-IPv4 tunnels seen as a near-term solution:
  - Tunnel as virtual NBMA link connecting many thousands of CPE routers
- PE routers in hub-and-spokes, mesh, or partial mesh
- Default routing via PE router works, but:
  - Traffic between CPEs takes longer paths than necessary
  - PE routers have to bear considerable load
- If each CPE is delegated IPv6 address/prefix, need a way to discover inter-CPE routes

NBMA Tunneling over ISP Network (ISATAP example)

- Each CPE has "default" only
- No control plane issues, but sub-optimal routes and data concentration

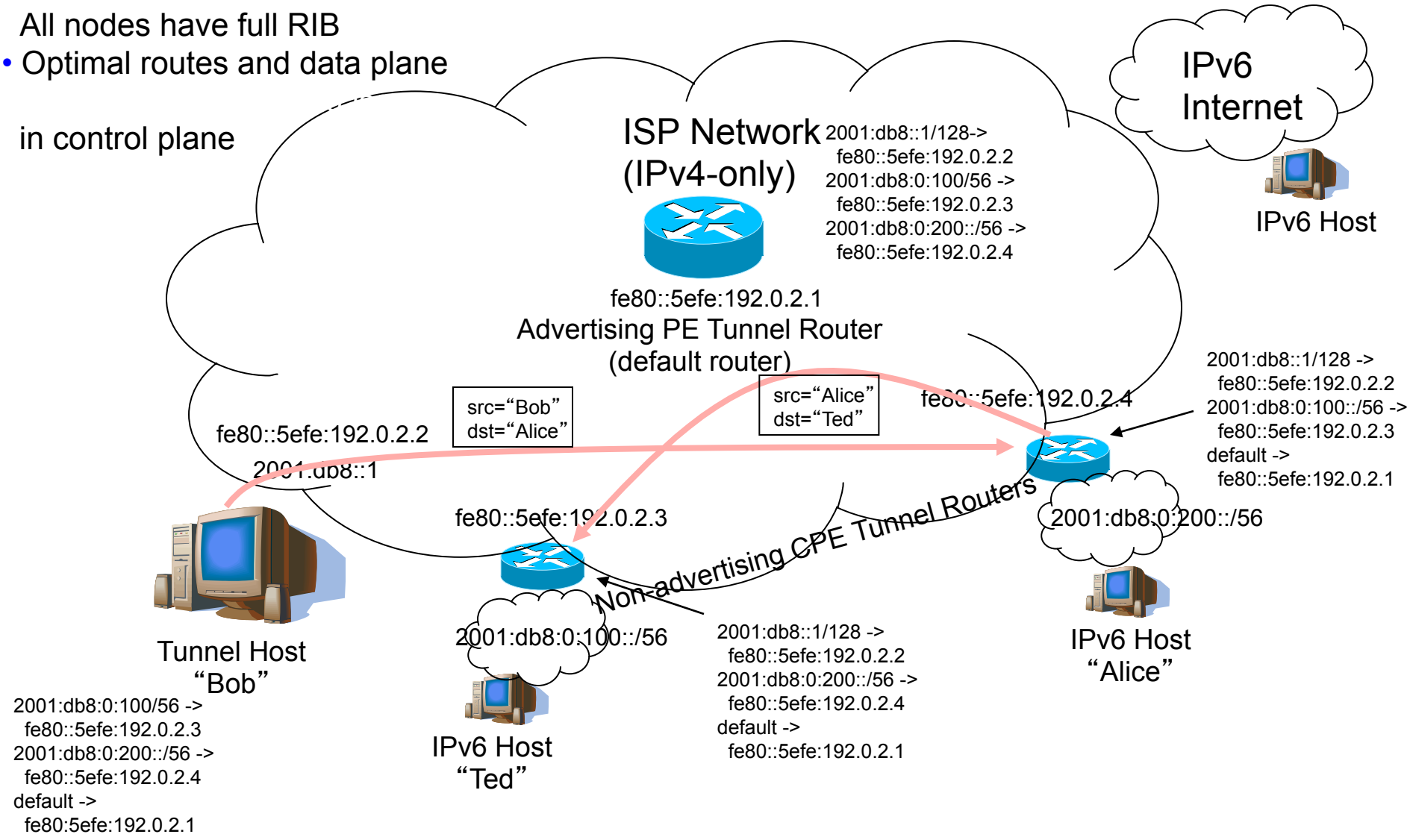


Dynamic IPv6 IGP Between Tunnel Routers

All nodes have full RIB

- Optimal routes and data plane

in control plane



Requirements for Dynamic Routing on Large Links

- **R1: Zero configuration on link nodes**
- **R2: Security based on chain-of-trust**
- **R3: Scale to support lots of nodes**
- **R4: Off-load performance-critical hub routers**
- **R5: Direct node-to-node route optimization**
- **R6: Route optimization for both routers and hosts**
- **R7: Support multiple levels of hierarchy**
- **R8: Do not circumvent ingress filtering**
- **R9: Do not expose packets to loss due to black holes**
- **R10: Support mobility**

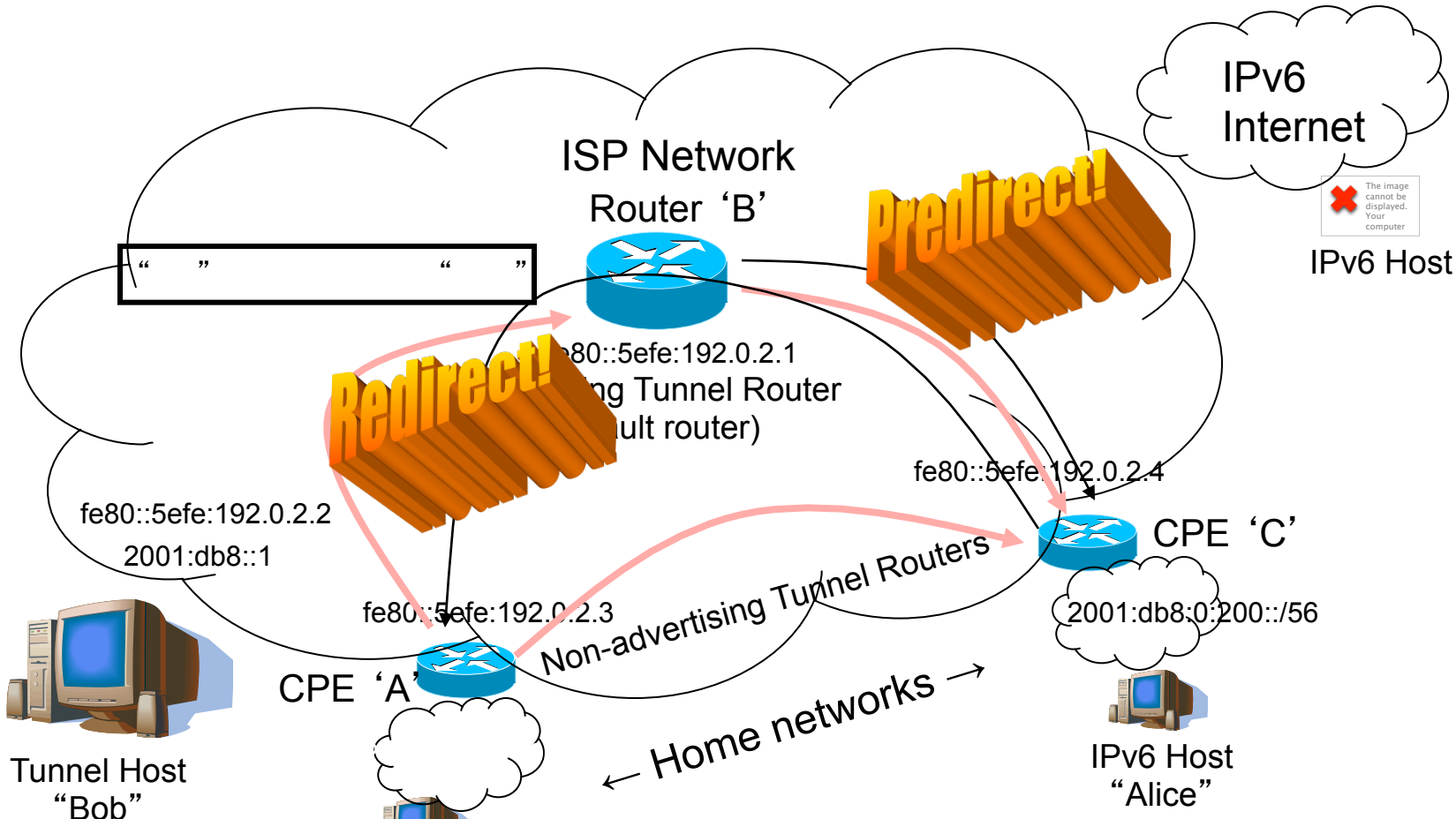
Idea – Use ICMP Redirects

- **ICMP Redirect mechanism works between neighbors on multiple access links (i.e., physical or virtual)**
- **However, classical 1-way redirection does not support coordination between neighbors:**
  - **target has no way of knowing that the source is authorized to produce packets using a given source address (source could be spoofing)**
  - **source has no way of knowing that the target is prepared to accept its packets directly**
- **Also, ICMP Redirection doesn't support router-to-router redirects**

Solution – "Augmented" Redirection

- Initial packets from source go through "anchor" default router that is trusted by both nodes
- Anchor sends PREDIRECT message forward to target:
  - target accepts the PREDIRECT since it trusts the anchor
  - tartget sends back a REDIRECT via the anchor
- Anchor proxies the REDIRECT back to the source:
  - source accepts the REDIRECT since it trusts the anchor
  - source sends future packets directly to target
- Redirects / Predirects include Route Information Options (RIOs) that include prefix/length instead of just destination
- Redirects / Predirects can be router-to-router (i.e. and not just router-to-host)

Augmented Redirection in ISP Network



1) ‘A’ sends packet to default router ‘B’
2) ‘B’ forwards packet; sends PREDIRECT to ‘C’
3) ‘C’ accepts PREDIRECT; returns a REDIRECT
4) ‘B’ PROXIES REDIRECT back to ‘A’
5) ‘A’ forwards future packets directly to ‘C’

Protocol Details

- On receipt of PREDIRECT, Target creates FIB entry in "FILTERING" state and sets a 40sec timer
  - Target accepts packets only while in FILTERING state
- On receipt of REDIRECT, Source creates FIB entry in "FORWARDING" state and sets a 30sec timer
  - Source forwards packets only while in FORWARDING state
- State cleared when timers expire
  - Future packets from source re-start redirection process
- Source can send periodic Predirects to elicit Redirects from Target
  - Keeps route optimized while packets are flowing

Additional Considerations

- **Route Optimization is asymmetric in the forward direction from source to target**
  - **If target sends packets in the reverse direction, a separate route optimization process is used (both directions managed independently)**
- **Mobility supported:**
  - **if final destination moves to a new network point of attachment, Target delivers the packet and returns a NACK**
- **Backward compatibility supported:**
  - **Based on standard ICMP Redirect messages (two new bits taken Reserved field)**
  - **Legacy nodes harmlessly ignore the messages and continue to use the default router**

Next Steps

- **Publish standalone document generalized to route optimization over any multiple access link that supports redirection**