IPv6 DNS, Routing, and Multihoming

Jeff Doyle

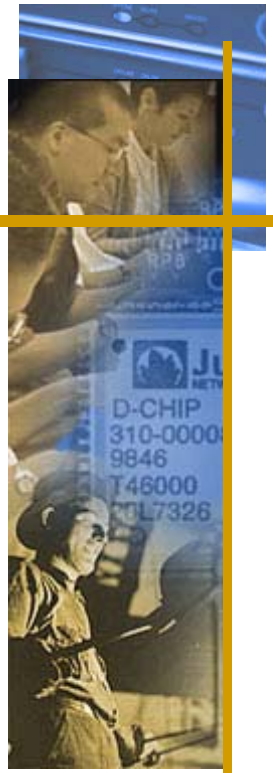
IPv6 Solutions Manager

jeff@juniper.net

Agenda

- DNS
- Routing IPv6
- IPv6 and Multihoming

Juniper NETWORKS



Agenda

- DNS
- Routing IPv6
- IPv6 and Multihoming

Transition Issues:
DNS

- Namespace fragmentation
  - Some names on IPv4 DNS, others on IPv6 DNS
  - How does an IPv4-only host resolve a name in the IPv6 namespace, and vice versa?
  - How does a dual-stack host know which server to query?
  - How do root servers share records?
- MX records
  - How does an IPv4 user send mail to an IPv6 user and vice versa?
- Solutions:
  - Dual stacked resolvers
  - Every zone must be served by at least one IPv4 DNS server
  - Use translators
    - NAT-PT does not work for this
    - totd: proxy DNS translator
- Some DNS transition issues discussed in RFC 1933, Section 3.2

DNS AAAA Records

- RFC 1886
- BIND 4.9.4 and up; BIND 8 is recommended
- Simple extension of A records
  - Resource Record type = 28
  - Query types performing additional section processing (NS, MX, MB) redefined to perform both A and AAAA additional section processing
- ip6.int, ip6.arpa analogous to in-addr.arpa for reverse mapping
  - IPv6 address represented in reverse, dotted hex nibbles

AAAA record:

```
homer        IN        AAAA        2001:4210:3:ce7:8:0:abcd:1234
```

PTR record:

```
4.3.2.1.d.c.b.a.0.0.0.0.8.0.0.0.7.e.c.0.3.0.0.0.0.1.2.4.1.0.0.2.ip6.int.   IN   PTR   homer.simpson.net
```

- RFC 3152 deprecates ip6.int in favor of ip6.arpa

DNS A6 Records

- Proposed alternative to AAAA records
  - RFC 2874
  - Resource Record type = 38
- A6 RR can contain:
  - Complete IPv6 address, or
  - Portion of address and information leading to one or more prefixes
- Supported in BIND 9
- More complicated records , but easier renumbering
  - Segments of IPv6 address specified in chain of records
  - Only relevant records must be changed when renumbering
  - Separate records can reflect addressing topology

A6 Record Chain

Queried Name: homer.simpson.net

```
$ORIGIN  simpson.net
homer    IN      A6      64      ::8:0:abcd.1234   sla5.subnets.simpson.net.

$ORIGIN  subnets.simpson.net
sla5     IN      A6      48      0:0:0:ce7::       site3.sites.net.

$ORIGIN  sites.net
site3    IN      A6      32      0:0:3::           area10.areas.net.

$ORIGIN  areas.net
area10   IN      A6      24      0:10::            tla1.tlas.net.

$ORIGIN  tlas.net
tla1     IN      A6      0       2001:4200::
```

Returned Address: 2001:4210:3:ce7:8:0:abcd:1234

Bitstring Labels

- New scheme for reverse lookups
- Bitstring Labels: RFC 2874
- Bitstring Labels for IPv6: RFC 2673

Examples:

Address: 2001:4210:3:ce7:8:0:abcd:1234

Bitstring labels:

\[x2001421000030ce700080000abcd1234/128].ip6.arpa.

\[x00080000abcd1234/64].\[x0ce7/16].\[x20014210/48].ip6.arpa.

- Pro:
  - More compact than textual (ip6.int) representation
- Con:
  - All resolvers and authoritative servers must be upgraded before new label type can be used
- RFC 3152 deprecates ip6.int in favor of ip6.arpa

DNAME

- DNAME: RFC 2672
- DNAME for IPv6: RFC 2874
- Provides alternate naming to an entire subtree of domain name space
  - Rather than to a single node
- Chaining complementary to A6 records
- DNAME not much more complex than CNAME
- DNAME changed from Proposed Standard to Experimental status in RFC 3363

DNAME Reverse Lookup

Queried Address: 2001:4210:3:ce7:8:0:abcd:1234

$ORIGIN ip6.arpa. \[x200142/24]	IN	DNAME	ip6.tla.net
$ORIGIN ip6.tla.net \[x10/8]	IN	DNAME	ip6.isp1.net
$ORIGIN ip6.isp1.net \[x0003/16]	IN	DNAME	ip6.isp2.net
$ORIGIN ip6.isp2.net \[x0ce7/16]	IN	DNAME	ip6.simpson.net
$ORIGIN ip6.simpson.net \[x00080000abcd1234/64]	IN	PTR	homer.simpson.net

Returned Name: homer.simpson.net

AAAA or A6?

- Good discussion of tradeoffs in RFC 3364
- AAAA Pros:
  - Essentially identical to A RRs, which are backed by extensive experience
  - "Optimized for read"
- AAAA Cons:
  - Difficult to inject new data
- A6 Pros:
  - "Optimized for write"
  - Possibly superior for rapid renumbering, some multihoming approaches (GSE-like routing)
- A6 Cons:
  - Long chains can reduce performance
  - Very little operational experience
- A6 RRs changed from Proposed Standard to Experimental status in RFC 3363
  - AAAA preferred for production deployment

Agenda

- DNS
- **Routing IPv6**
- IPv6 and Multihoming

Juniper NETWORKS

MTU Path Discovery

- IPv6 routers do not fragment packets
- IPv6 MTU must be at least 1280 bytes
  - Recommended MTU: 1500 bytes
- Nodes should implement MTU PD
  - Otherwise they must not exceed 1280 bytes
- MTU path discovery uses ICMP "packet too big" error messages

Configuration Example:
Static Route

```
[edit routing-options]
ps@R1# show
rib inet6.0 {
    static {
        route 3ffe::/16 next-hop 2001:468:1100:1::2;
    }
}
```

RIPng

- RFC 2080 describes RIPngv1, not to be confused with RIPv1
- Based on RIP Version 2 (RIPv2)
- Uses UDP port 521
- Operational procedures, timers and stability functions remain unchanged
- RIPng is not backward compatible to RIPv2
- Message format changed to carry larger IPv6 addresses

Configuration Example:
RIPng

```
[edit protocols]
lab@Juniper5# show
ripng {
    group external_neighbors {
        export default_route;
        neighbor ge-0/0/0.0;
        neighbor ge-0/0/1.0;
        neighbor ge-0/0/2.0;
    }
    group internal_neighbors {
        export external_routes;
        neighbor ge-1/0/0.0;
    }
}
```

IS-IS

- draft-ietf-isis-ipv6-02.txt, Routing IPv6 with IS-IS
- 2 new TLVs are defined:
  - IPv6 Reachability (TLV type 236)
  - IPv6 Interface Address (TLV type 232)
- IPv6 NLPID = 142

Configuration Example:
IS-IS for IPv6 Only

- By default, IS-IS routes both IPv4 and IPv6

```
lab@Juniper5# show
isis {
    no-ipv4-routing;
    interface ge-0/0/1.0;
    interface ge-0/0/2.0;
}
```

OSPFv3

- Unlike IS-IS, entirely new version required
- RFC 2740
- Fundamental OSPF mechanisms and algorithms unchanged
- Packet and LSA formats are different

OSPFv3 Differences from OSPFv2

- Runs per-link rather than per-subnet
  - Multiple instances on a single link
- More flexible handling of unknown LSA types
- Link-local flooding scope added
  - Similar to flooding scope of type 9 Opaque LSAs
  - Area and AS flooding remain unchanged
- Authentication removed
- Neighboring routers always identified by RID
- Removal of addressing semantics
  - IPv6 addresses not present in most OSPF packets
  - RIDs, AIDs, and LSA IDs remain 32 bits

OSPFv3 LSAs

Type	Description
0x2001	Router-LSA
0x2002	Network-LSA
0x2003	Inter-Area-Prefix-LSA
0x2004	Inter-Area-Router-LSA
0x2005	AS-External-LSA
0x2006	Group-Membership-LSA
0x2007	Type-7-LSA (NSSA)
0x2008	Link-LSA
0x2009	Intra-Area-Prefix-LSA

Configuration Example:
OSPFv3

```
[edit protocols]
lab@Juniper5# show
ospf3 {
    area 0.0.0.0 {
        interface ge-1/1/0.0;
    }
    area 192.168.1.2 {
        interface ge-0/0/1.0;
        interface ge-0/0/2.0;
    }
}
```

Multiprocotol BGP-4

- MBGP defined in RFC 2283
- Two BGP attributes defined:
  - Multiprotocol Reachable NLRI advertises arbitrary Network Layer Routing Information
  - Multiprotocol Unreachable NLRI withdraws arbitrary Network Layer Routing Information
  - Address Family Identfier (AFI) specifies what NLRI is being carried (IPv6, IP Multicast, L2VPN, L3VPN, IPX…)
- Use of MBGP extensions for IPv6 defined in RFC 2545
  - IPv6 AFI = 2
- BGP TCP session can be over IPv4 or IPv6
- Advertised Next-Hop address must be global or site-local IPv6 address
  - And can be followed by a link-local IPv6 address
  - Resolves conflicts between IPv6 rules and BGP rules

Example Configuration:
BGP

```
[edit protocols]
lab@Juniper5# show
bgp {
    group IPv6_external {
        type external;
        import v6_externals;
        family inet6 {
            unicast;
        }
        export v6_routes;
        peer-as 65502;
        neighbor 3ffe:1100:1::b5;
    }
    group IPv6_internal {
        type internal;
        local-interface lo0.0;
        family inet6 {
            unicast;
        }
        neighbor 2001:88:ac3::51;
        neighbor 2001:88:ac3::75;
    }
}
```

Agenda

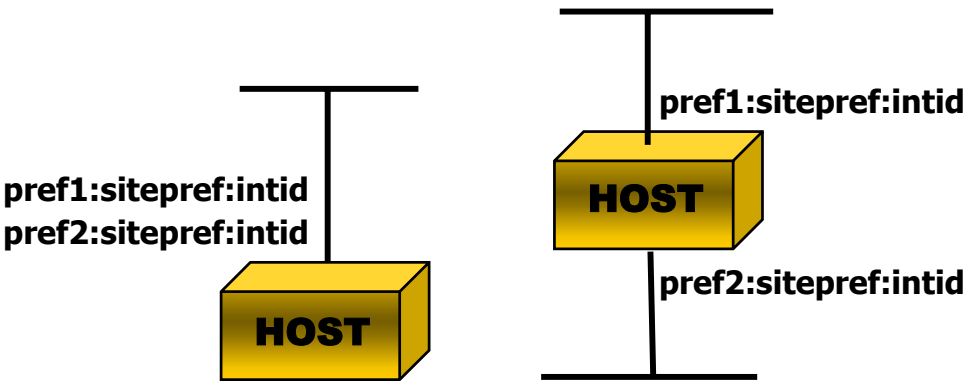
- DNS
- Routing IPv6
- **IPv6 and Multihoming**

Juniper NETWORKS

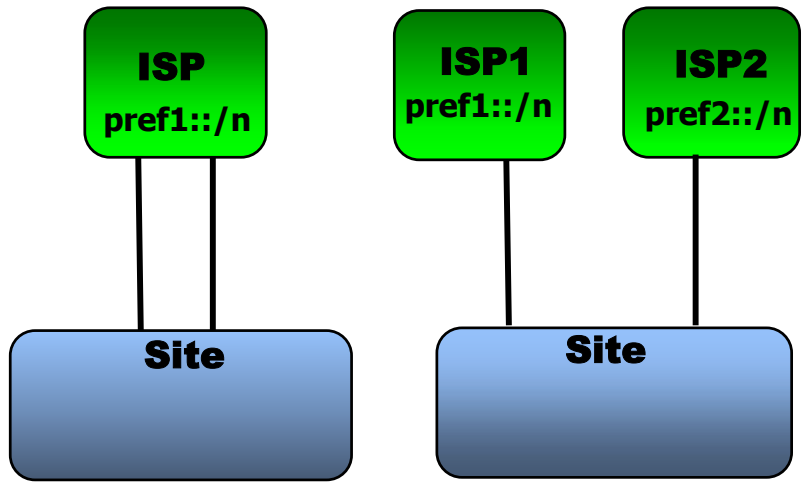
What is Multihoming?

- **Host multihoming**
  - **More than one unicast address on an interface**
  - **Interfaces to more than one network**
- **Site multihoming**
  - **Multiple connections to the same ISP**
  - **Connections to multiple ISPs**

Host Multihoming



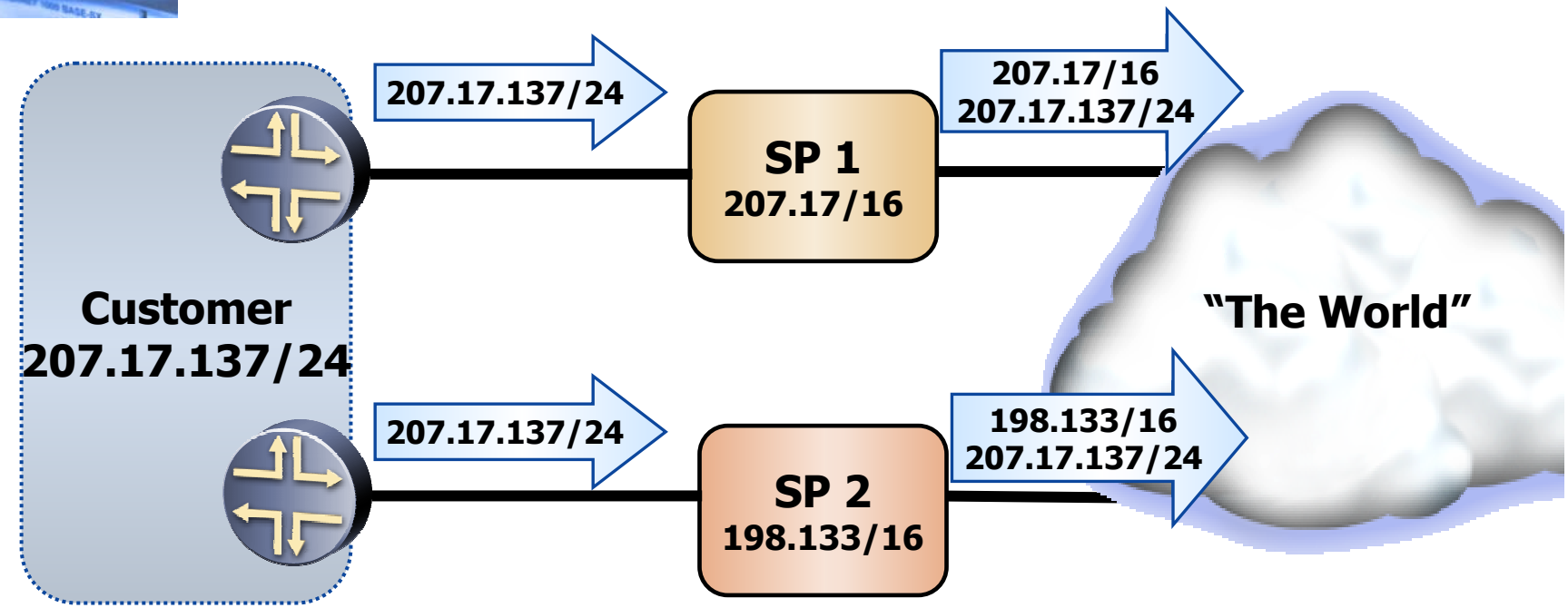
Site Multihoming



Why Multihome?

- Redundancy
  - Against router failure
  - Against link failure
  - Against ISP failure
- Load sharing
- Local connectivity across large geography
- Corporate or external policies
  - Acceptable use policies
  - Economics

The Multihoming Problem



- ISP2 must advertise additional prefix
- ISP1 must "punch a hole" in its CIDR block
- Contributes to routing table explosion
- Contributes to Internet instability
  - Due to visibility of customer route flaps
  - Due to increased convergence time
- Same problem can apply to provider-independent (PI) addresses

IPv6 and The Multihoming Problem

- IPv6 does not have a set solution to the problem
- Currently, 6Bone disallows IPv4-style multihoming (RFC 2772)
  - ISPs cannot advertise prefixes of other ISPs
  - Sites cannot advertise to upstream providers prefixes longer than their assigned prefix
- However, IPv6 offers the possibility of one or more solutions
  - Router-based solutions
  - Host-based solutions
  - Mobile-based solutions
  - Geographic or Exchange-based solutions

Multihoming Requirements

Requirements for IPv6 Site-Multihoming Architectures
(draft-ietf-multi6-multihoming-requirements-03)

- **Must support redundancy**
- **Must support load sharing**
- **Protection from performance difficulties**
- **Support for multihoming for external policy reasons**
- **Must not be more complex than current IPv4 solutions**
- **Re-homing transparency for transport-layer sessions (TCP, UDP, SCTP)**
- **No impact on DNS**
- **Must not preclude packet filtering**
- **Must scale better than IPv4 solutions**
- **Minor impact on routers**
- **No impact on host connectivity**
- **May involve interaction between hosts and routers**
- **Must be manageable**
- **Must not require cooperation between transit providers**

Possible Solution #1: Do Nothing

- **Allow Internet default free zone (DFZ) to continue to grow**
- **Put responsibility on router vendors to keep increasing memory, performance to compensate**

Pros:
- As simple as it gets
- No special designs, policies, or mechanisms needed

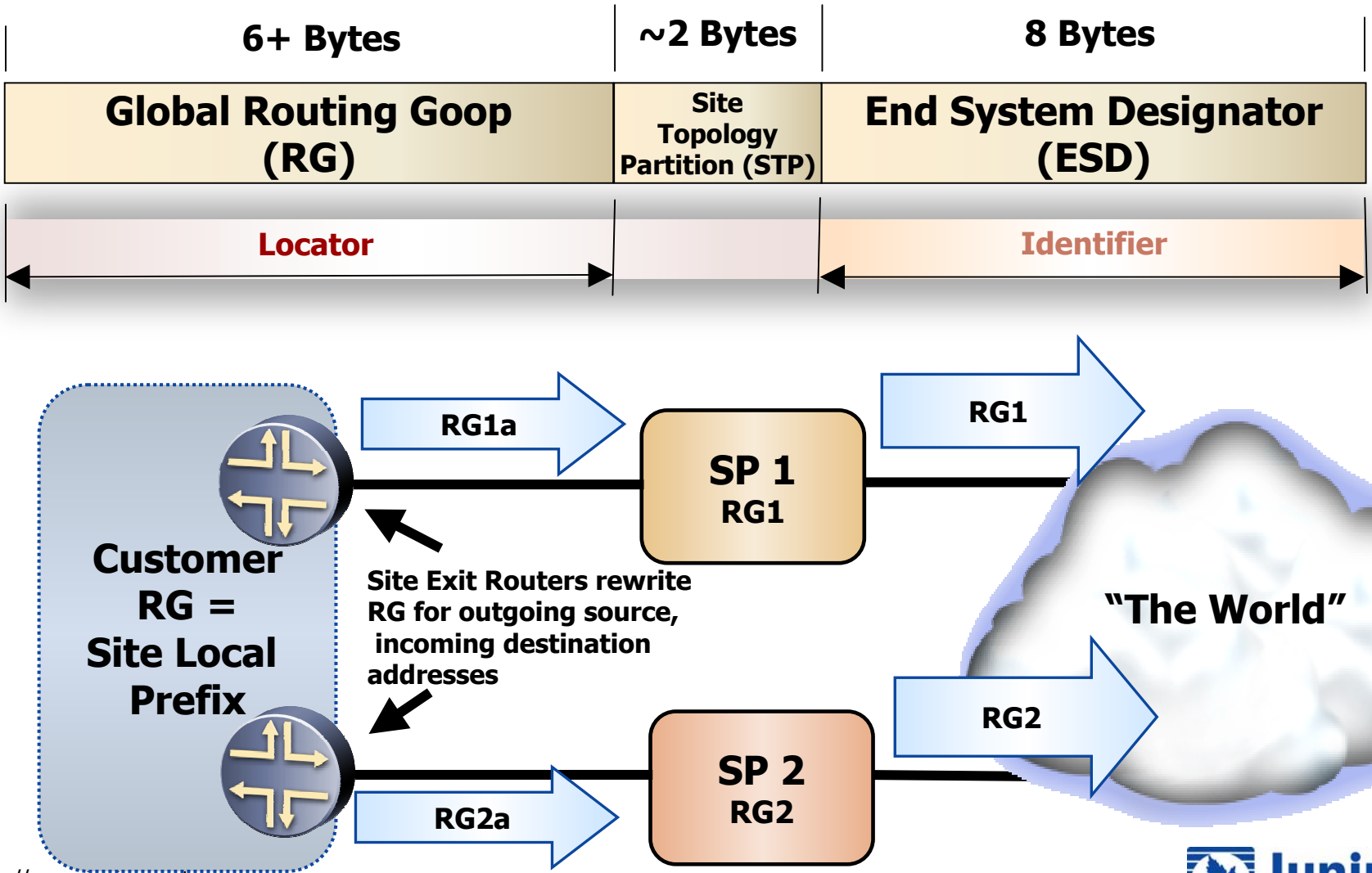
Cons:
- Does nothing to increase Internet stability
- Large routing tables = Large convergence times
- No guarantee vendors can continue to stay ahead of the curve

Possible Solution #2: GSE/8+8

GSE: Global, Site, and End System Address Elements
(draft-ipng-gseaddr-00.txt)
(draft-ietf-ipngwg-esd-analysis-05.txt)

- Router-based solution
- Key concepts:
  - Distinct separation of Locator and Identifier entities in IPv6 addresses
  - Rewriting of locator (Routing Goop) at Site Exit Router
  - Identifier (End System Designator) is globally unique
  - DNS AAA records and RG records

Possible Solution #2: GSE/8+8



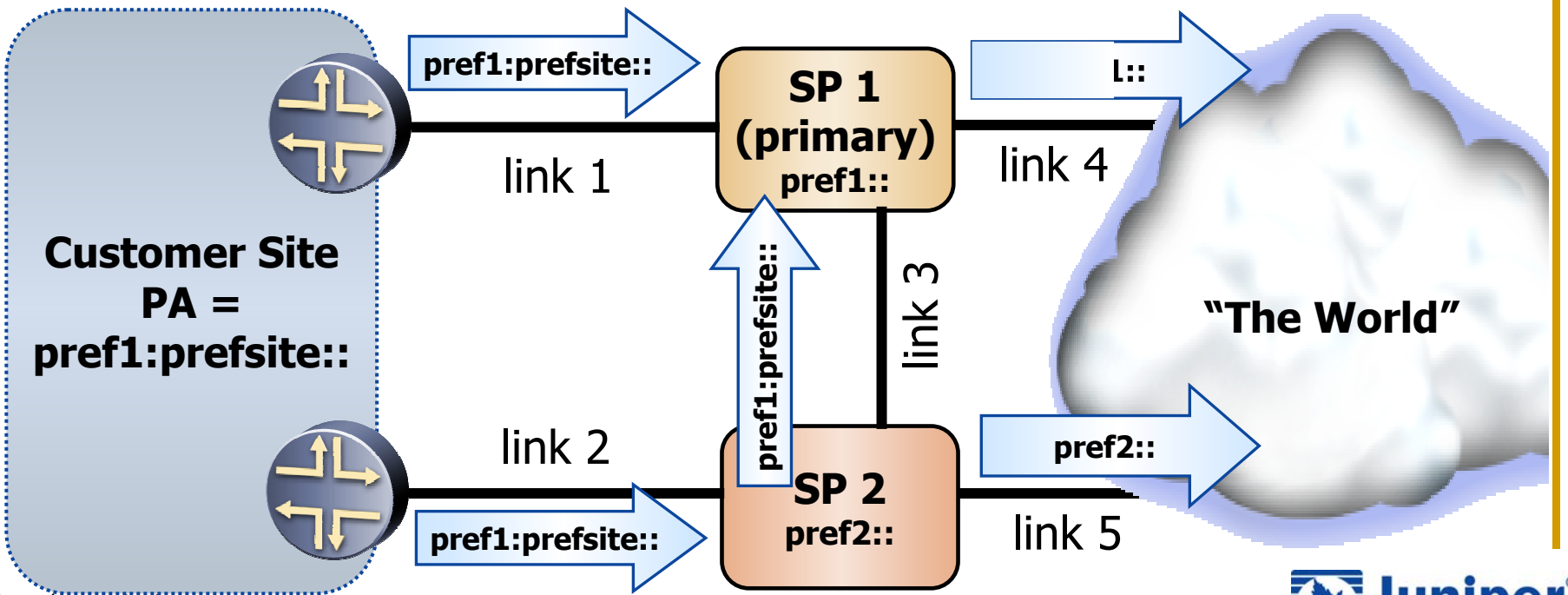
Possible Solution #2: GSE/8+8

- GSE as proposed rejected by IPng WG in 1997
  - Thought to introduce more problems than it solved
    - "Separating Identifiers and Locators in Addresses: An Analysis of the GSE Proposal for IPv6" (draft-ietf0ipngwg-esd-analysis-04.txt)
  - But, concept is still being discussed

Possible Solution #3: Multihoming with Route Aggregation

(draft-ietf-ipngwg-ipv6multihome-with-aggr-01.txt)

- Router-based solution
- Customer site gets PA from primary ISP
- PA advertised to both ISPs, but not upstream
- PA advertised from ISP2 to ISP1



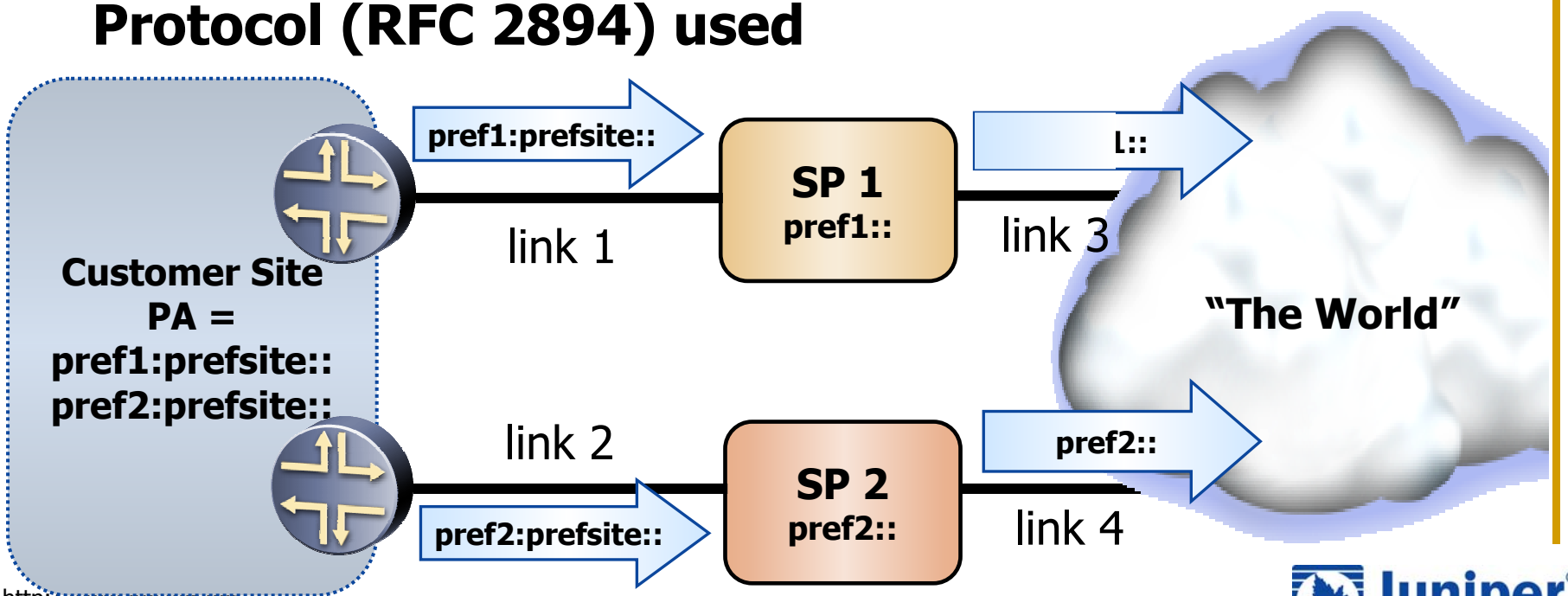
Possible Solution #3: Multihoming with Route Aggregation

- **Pros:**
  - No new protocols or modifications needed
  - Fault tolerance for links 1 and 2
  - Load sharing with ISPs 1 and 2
  - Link failure does not break established TCP sessions
- **Cons:**
  - No fault tolerance if ISP1 or link 4 fails
  - No load sharing if link 3 fails
  - Problematic if link 3 must pass through intermediate ISP
  - Assumes ISP1 and ISP2 are willing to provide link 3 and appropriate route advertisements

Possible Solution #4: Multihoming Using Router Renumbering

(draft-ietf-ipngwg-multi-isp-00.txt)

- Router-based solution
- All customer device interfaces carry addresses from each ISP
- Router Advertisements and Router Renumbering Protocol (RFC 2894) used



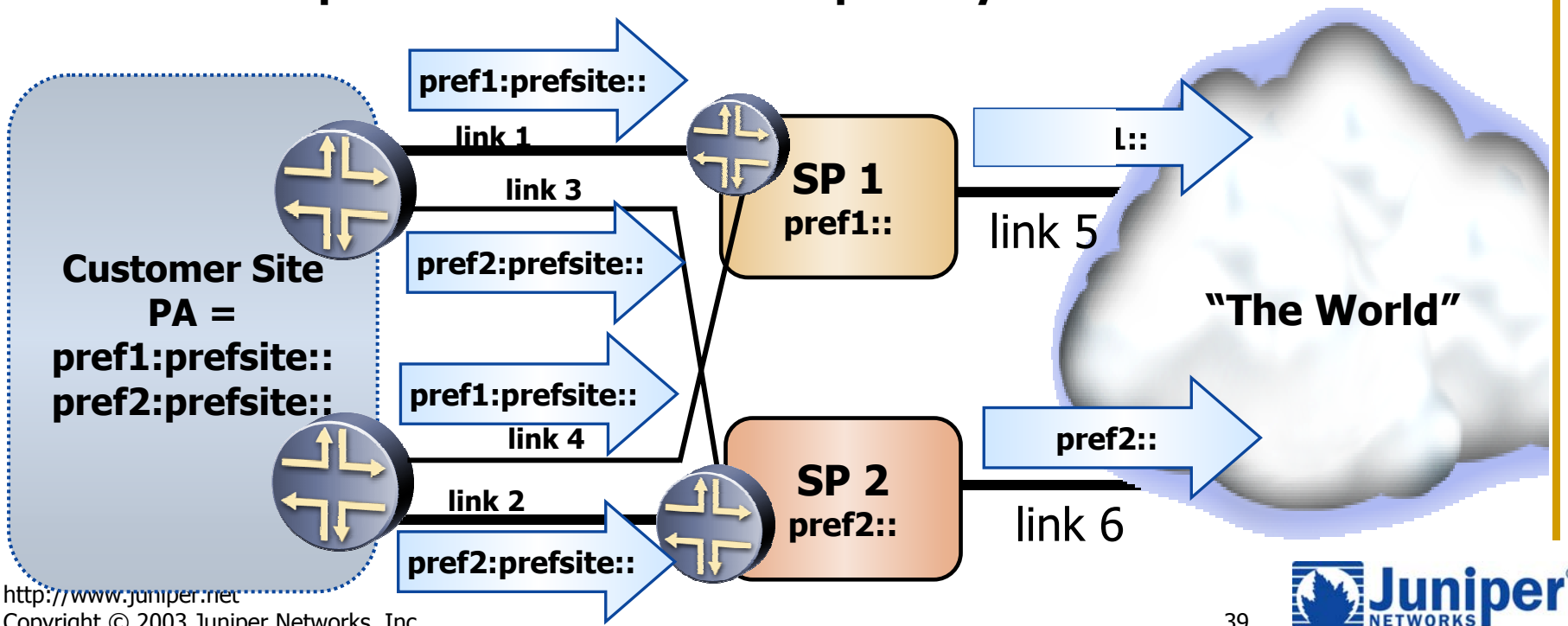
Possible Solution #4: Multihoming Using Router Renumbering

- **If an ISP fails:**
  - Site border router detecting failure sends RAs to deprecate ISP's delegated addresses
  - Router Renumbering Protocol propagates information about deprecation to internal routers
- **Pros:**
  - No new protocols or modifications needed
  - Fault tolerance for both links and ISPs
- **Cons:**
  - No clear criteria for selecting among multiple interface addresses
  - No clear criteria for load sharing among ISPs
  - Link or ISP failure breaks established TCP sessions

Possible Solution #4: Multihoming Support at Site Exit Routers

(RFC 3178)

- Router-based solution
- Links 3 and 4 (IP in IP tunnels) configured as secondary links
- Primary and secondary links on separate physical media for link redundancy
- Prefixes advertised over secondary links have weak preference relative to prefixes advertised over primary links



Possible Solution #4: Multihoming Support at Site Exit Routers

- **Pros:**
  - No new protocols or modifications needed
  - Link fault tolerance
  - Link failure does not break established TCP sessions
- **Cons:**
  - No fault tolerance if ISP fails
  - No clear criteria for selecting among multiple interface addresses
  - No clear criteria for load sharing among ISPs

Possible Solution #5: Host-Centric IPv6 Multihoming

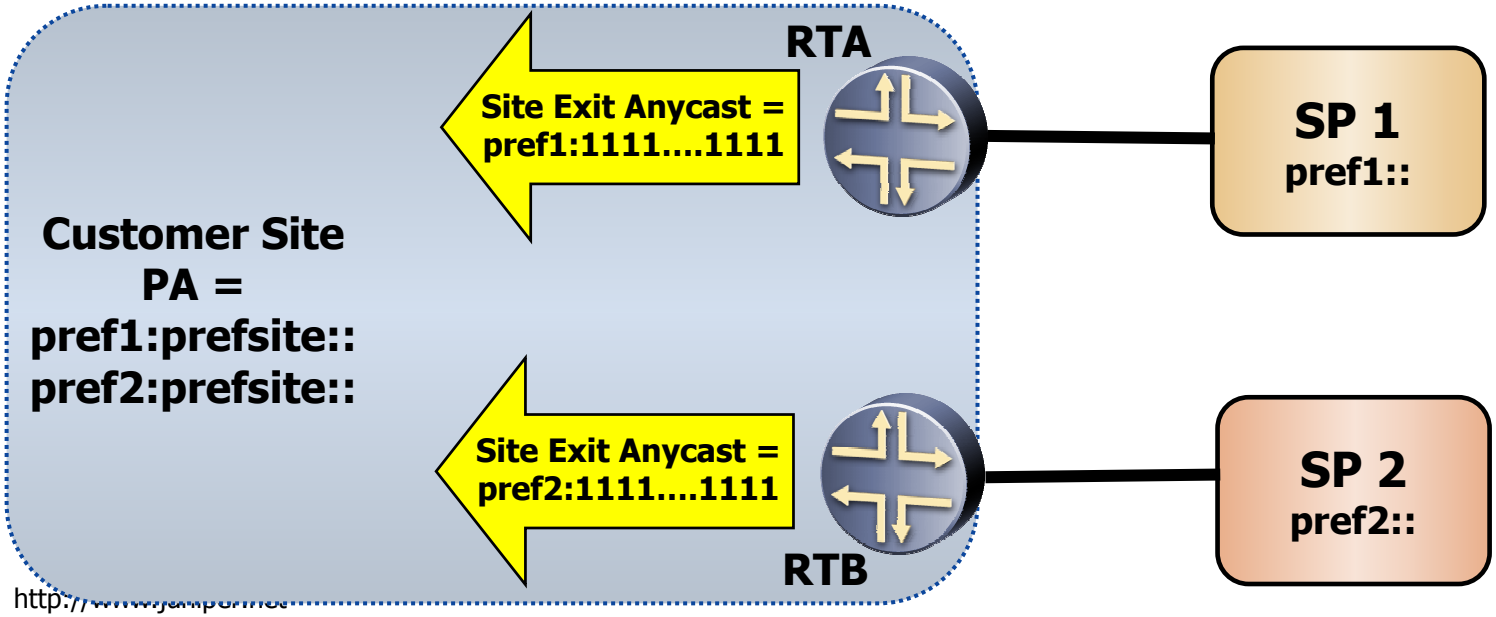
(draft-huitema-multi6-hosts-01.txt)

- Host- *and* router-based solution
- Key Concepts:
  - Multiple addresses per host interface
  - Site exit router discovery
  - Site exit anycast address
  - Site exit redirection
    - New Site Exit Redirection ICMP message defined

Possible Solution #5: Host-Centric IPv6 Multihoming

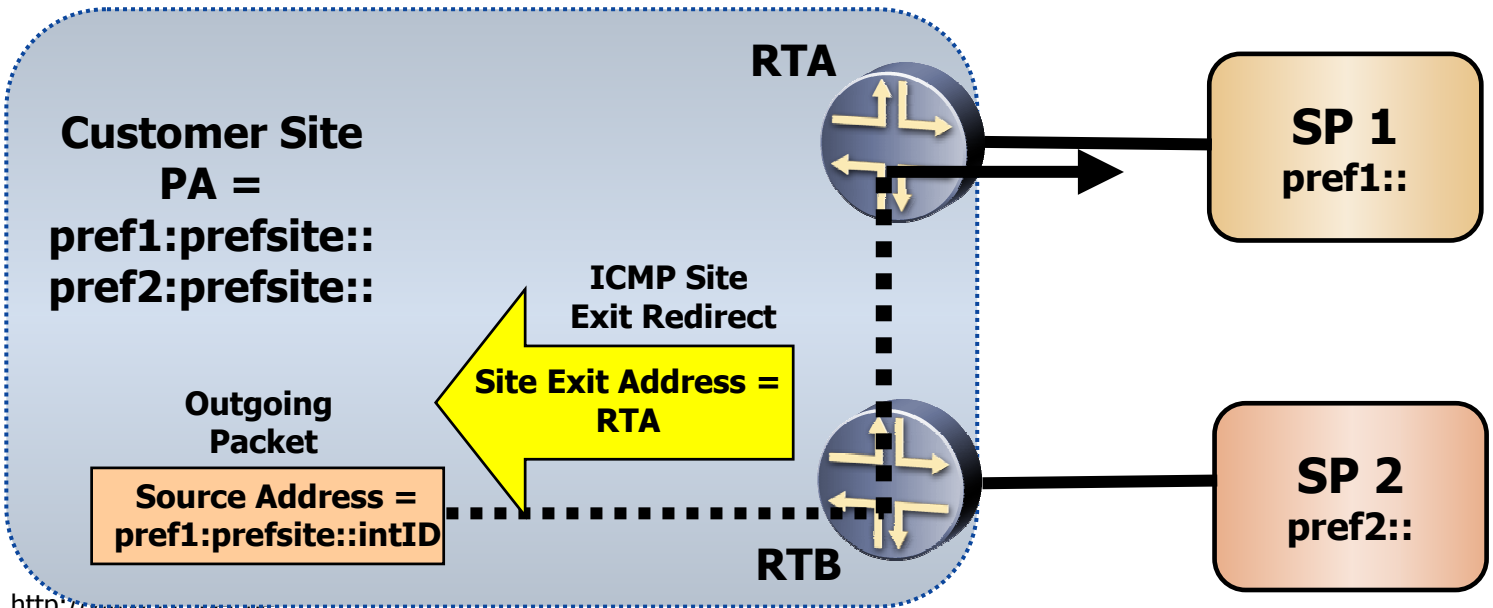
- Site anycast address indicates site exit address
- Site anycast address advertised via IGP
- Hosts tunnel packets to selected site exit router

L bits	128 – L bits
Site Prefix	All Ones (1111.....1111)



Possible Solution #5: Host-Centric IPv6 Multihoming

- **Site redirection:**
  1. Tunnels created between all site exit routers
  2. Source address of outgoing packets examined
  3. Packet tunneled to correct site exit router
  4. Site exit redirect sent to host



Possible Solution #5: Host-Centric IPv6 Multihoming

- **Pros:**
  - Fault tolerant of link, router, and ISP failure
  - Overcomes problem of ingress source address filtering at ISPs
- **Cons:**
  - Requires new ICMP message
  - Requires modification to both routers and hosts
  - Tunneling can become complex
    - Between site exit routers
    - Hosts to all site exit routers

Possible Solution #6: Geographically Aggregatable PI Address Space (GAPI)

(draft-py-multi6-gapi-00.txt)

- Address allocation framework only
  - Does not specify a multihoming mechanism
- Assigns a /16 over geographical entities
- Geographical entities assign /32s
- Example:
  - Top /32 zone: Sub-continents
    - Highest level of aggregation
  - Next level: Countries
    - Assigned range of /32 blocks based on population
  - Lowest level: Metropolitan areas
    - Cities have one or more metro areas
    - Small towns and rural areas away from metro areas allocated from country level

1. China
2. Continental Asia
3. India
4. Northern Africa
5. Asian Islands
6. Western Europe
7. North America
8. South America
9. Eastern Europe
10. Middle East
11. Soutern Africa
12. Central America
13. Oceana

Proposed Geographic Mechanisms

- Multihoming Aliasing Protocol (MHAP)
  - (draft-py-mhap-01a.txt)
- Provider-Internal Aggregation Based on Geography to Support Multihoming in IPv6
  - (draft-van-beijnum-multi6-isp-int-aggr-00.txt)
- An IPv6 Provider-Independent Global Unicast Address Format
  - (draft-hain-ipv6-pi-addr-04.txt)
  - (draft-hain-ipv6-pi-addr-use-04.txt)

Other IPv6 Multihoming Issues

- How does a host choose between multiple source and destination addresses?
  - See draft-ietf-ipv6-default-addr-select-09
- How are DNS issues resolved?
  - See RFC 2874, "DNS Extensions to Support IPv6 Address Aggregation and Renumbering," section 5.1, for DNS proposals for multihoming