Enterprise Multihoming using Provider-Assigned Addresses without Network Prefix Translation: Requirements and Solution

Draft-bowbakova-rtgwg-enterprise-pa-multihoming-00

F. Baker, C. Bowers, J. Linkova

IETF96, Berlin, July 2016

Problems with PA Multihoming

Q: How to send packets to the correct uplink (BCP38)?

Q: How to implement policies?

Q: How to react to links failure/recovery?

WITHOUT NAT!

Solutions with PA Multihoming

Q: How to send packets to the correct uplink (BCP38)?

A: Source Address Dependent Routing (SADR)

NO NAT!

Q: How to implement policies?

Q: How to react to link failure and recovery?

A: Influence source address & next-hop selection on hosts

Requirements/Expectations

Hosts have addresses from 2 or more non-overlapping blocks

Packets are sent to an ISP only if src address belongs to PA space of that ISP

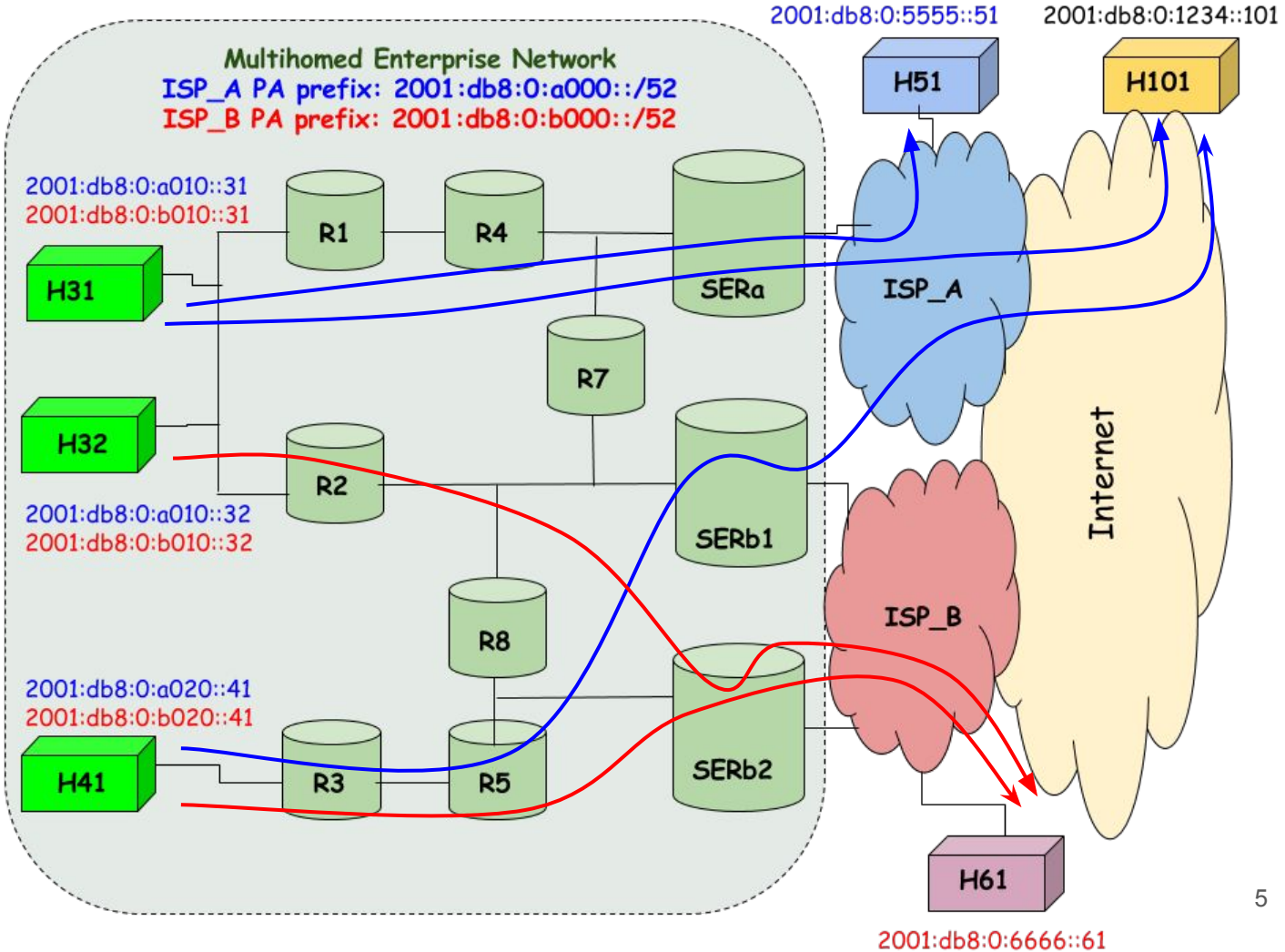
“No uplink for this source” is signalled to hosts

Hosts are expected to properly select a source address

Different DA might require different sources

Intra-site communication is not affected

Example Topology



Part 1: The Network

Source Address Dependent Routing

SADR: Overview

SADR-capable Routers have:

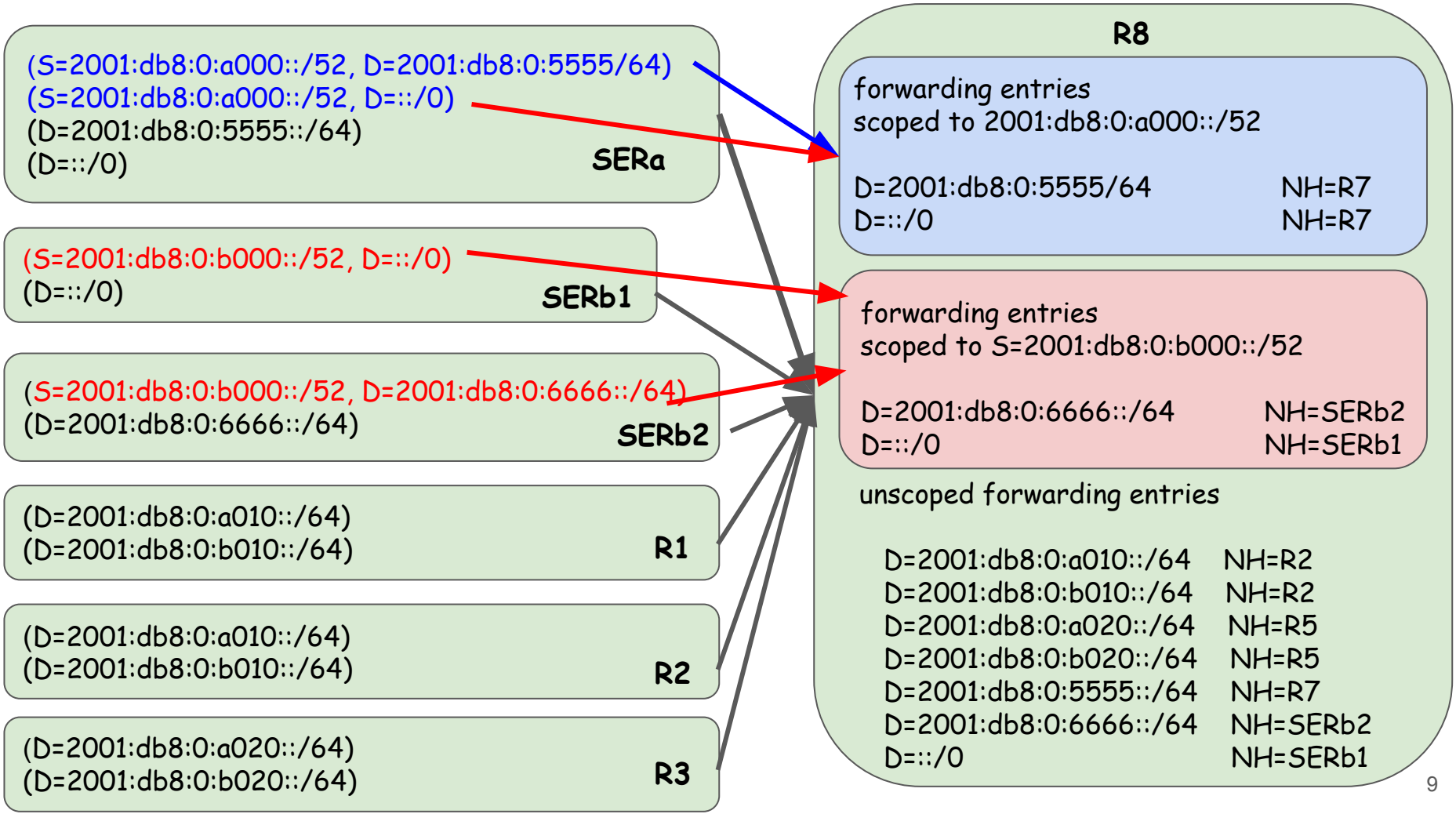
- forwarding tables scoped to given prefixes
- unscoped (scoped to S=::/0) forwarding table
  - might not be required if all routers support SADR

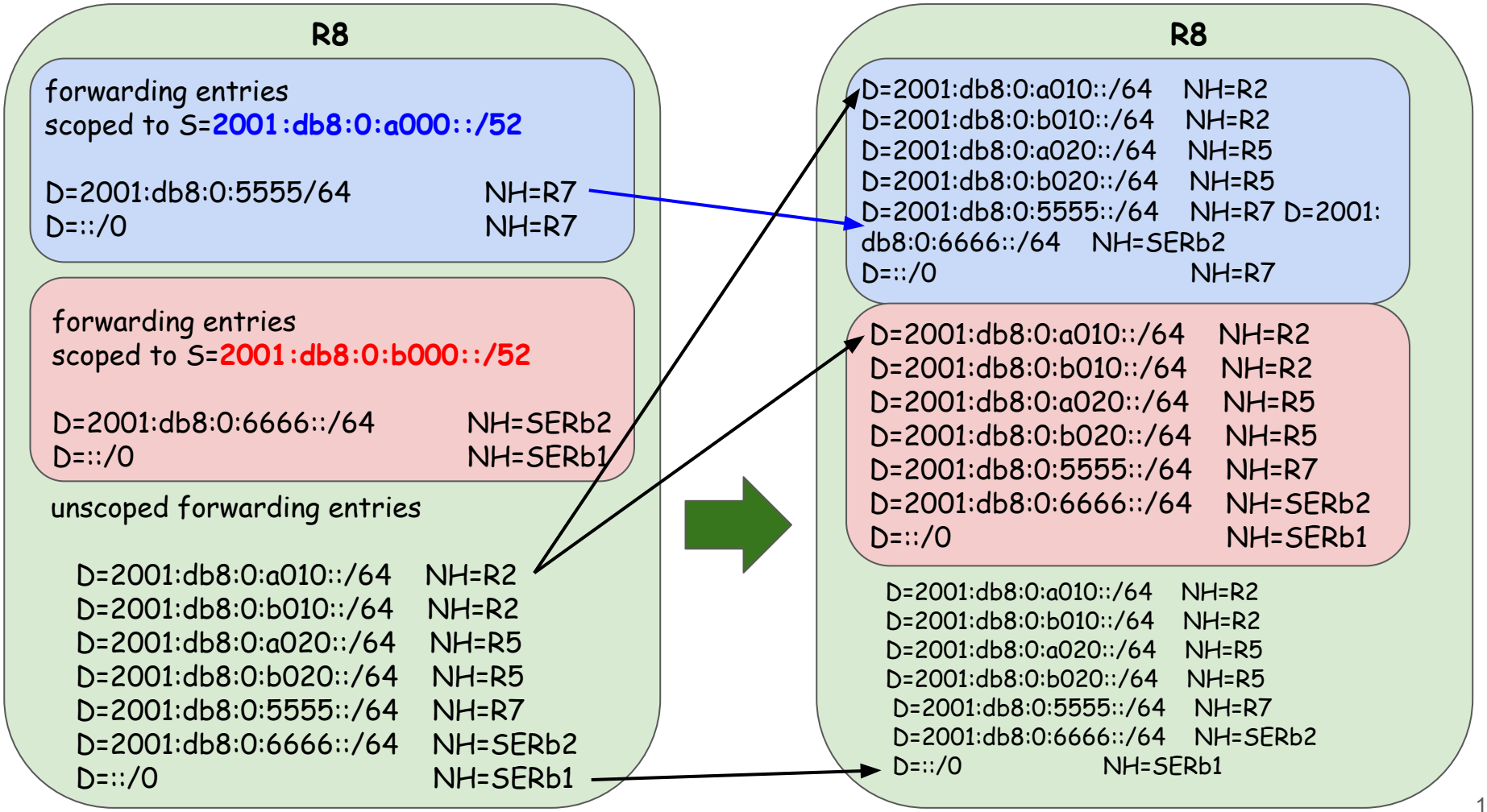
Incremental Rollout:

- At least Site Edge Routers (SERs) support SADR
- Other routers can do destination-based routing
  - Traffic path might be suboptimal and tunnels might be required

Creating Scoped Tables

1. Compute the next-hops for the source-prefix-scoped destination prefixes using only routers in the connected SADR domain (source-prefix-scoped forwarding table)
2. Compute the next-hops for the unscoped destination prefixes using all routers in the IGP (unscoped forwarding table)
3. Augment each source-prefix-scoped forwarding table with unscoped forwarding table entries based on the following rule. If the destination prefix of the unscoped forwarding entry exactly matches the destination prefix of an existing source-prefix-scoped forwarding entry (including destination prefix length), then do not add the unscoped forwarding entry. If the destination prefix does NOT match an existing entry, then add the entry to the source-prefix-scoped forwarding table.

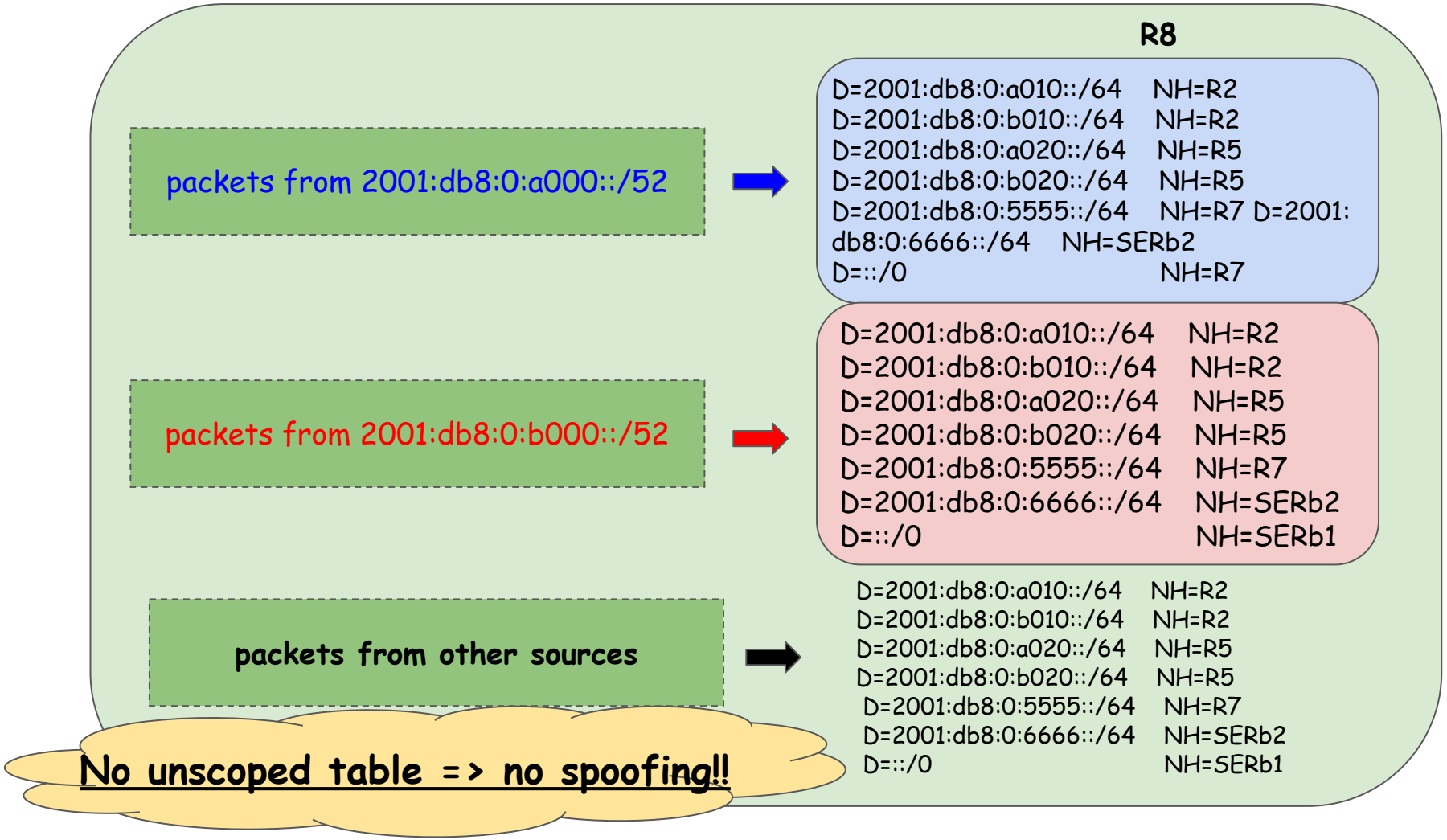

(S=2001:db8:0:a000::/52, D=2001:db8:0:5555/64)
(S=2001:db8:0:a000::/52, D=::/0)
(D=2001:db8:0:5555::/64)
(D=::/0)
SERa
(S=2001:db8:0:b000::/52, D=::/0)
(D=::/0)
SERb1
(S=2001:db8:0:b000::/52, D=2001:db8:0:6666::/64)
(D=2001:db8:0:6666::/64)
SERb2
(D=2001:db8:0:a010::/64)
(D=2001:db8:0:b010::/64)
R1
(D=2001:db8:0:a010::/64)
(D=2001:db8:0:b010::/64)
R2
(D=2001:db8:0:a020::/64)
(D=2001:db8:0:b020::/64)
R3
R8
forwarding entries
scoped to 2001:db8:0:a000::/52
D=2001:db8:0:5555/64 NH=R7
D=::/0 NH=R7
forwarding entries
scoped to S=2001:db8:0:b000::/52
D=2001:db8:0:6666::/64 NH=SERb2
D=::/0 NH=SERb1
unscoped forwarding entries
D=2001:db8:0:a010::/64 NH=R2
D=2001:db8:0:b010::/64 NH=R2
D=2001:db8:0:a020::/64 NH=R5
D=2001:db8:0:b020::/64 NH=R5
D=2001:db8:0:5555::/64 NH=R7
D=2001:db8:0:6666::/64 NH=SERb2
D=::/0 NH=SERb1


R8
forwarding entries
scoped to S=2001:db8:0:a000::/52
D=2001:db8:0:5555/64 NH=R7
D=::/0 NH=R7
forwarding entries
scoped to S=2001:db8:0:b000::/52
D=2001:db8:0:6666::/64 NH=SERb2
D=::/0 NH=SERb1
unscoped forwarding entries
D=2001:db8:0:a010::/64 NH=R2
D=2001:db8:0:b010::/64 NH=R2
D=2001:db8:0:a020::/64 NH=R5
D=2001:db8:0:b020::/64 NH=R5
D=2001:db8:0:5555::/64 NH=R7
D=2001:db8:0:6666::/64 NH=SERb2
D=::/0 NH=SERb1
R8
D=2001:db8:0:a010::/64 NH=R2
D=2001:db8:0:b010::/64 NH=R2
D=2001:db8:0:a020::/64 NH=R5
D=2001:db8:0:b020::/64 NH=R5
D=2001:db8:0:5555::/64 NH=R7 D=2001:
db8:0:6666::/64 NH=SERb2
D=::/0 NH=R7
D=2001:db8:0:a010::/64 NH=R2
D=2001:db8:0:b010::/64 NH=R2
D=2001:db8:0:a020::/64 NH=R5
D=2001:db8:0:b020::/64 NH=R5
D=2001:db8:0:5555::/64 NH=R7
D=2001:db8:0:6666::/64 NH=SERb2
D=::/0 NH=SERb1
D=2001:db8:0:a010::/64 NH=R2
D=2001:db8:0:b010::/64 NH=R2
D=2001:db8:0:a020::/64 NH=R5
D=2001:db8:0:b020::/64 NH=R5
D=2001:db8:0:5555::/64 NH=R7
D=2001:db8:0:6666::/64 NH=SERb2
D=::/0 NH=SERb1

Packet Forwarding

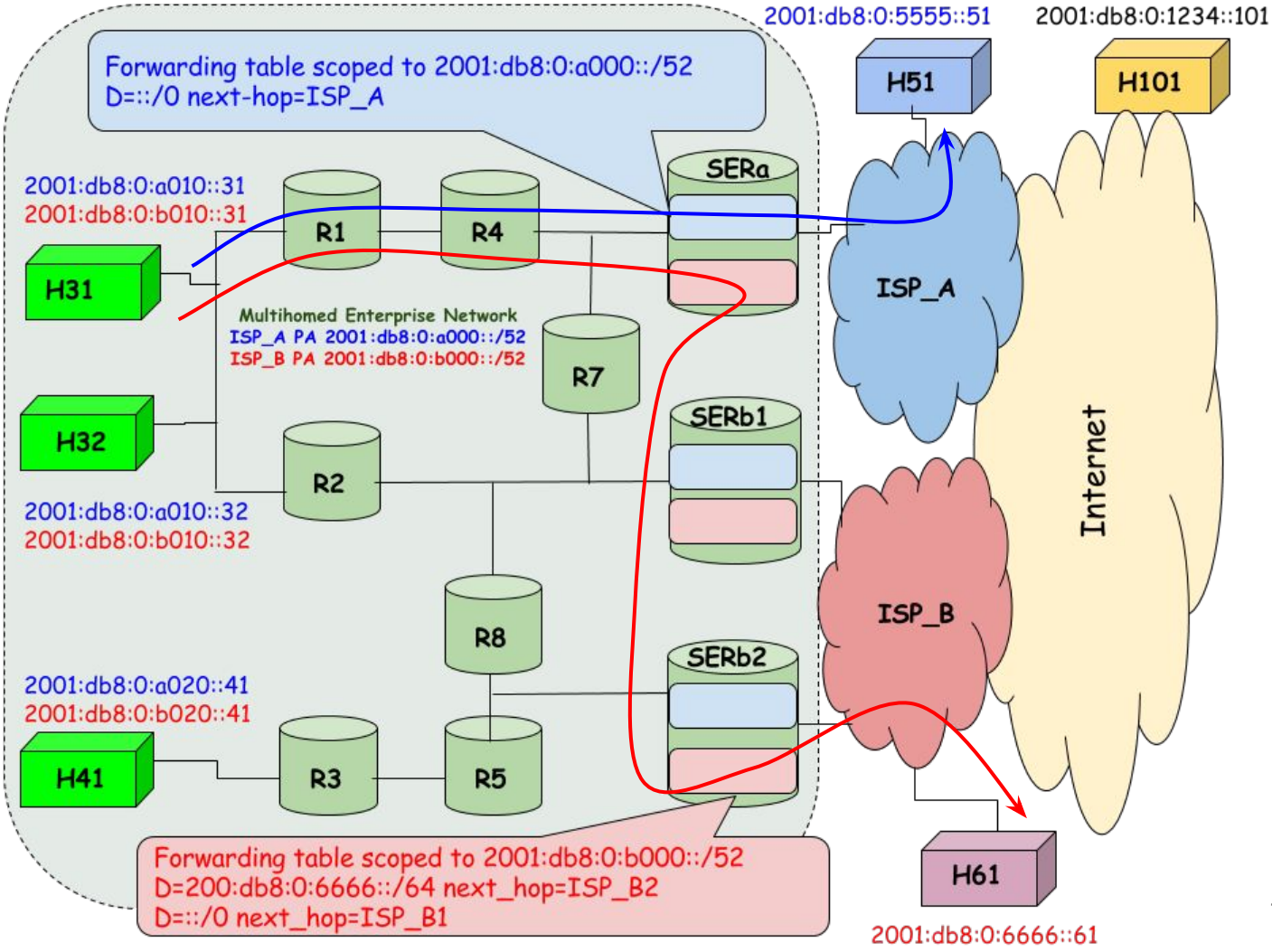
If the source address of the packet matches one of the source prefixes, then look up the destination address of the packet in the corresponding source-prefix-scoped forwarding table to determine the next-hop for the packet.

If the source address of the packet does NOT match one of the source prefixes, then look up the destination address of the packet in unscoped forwarding table to determine the next-hop for the packet.


R8
packets from 2001:db8:0:a000::/52
D=2001:db8:0:a010::/64 NH=R2
D=2001:db8:0:b010::/64 NH=R2
D=2001:db8:0:a020::/64 NH=R5
D=2001:db8:0:b020::/64 NH=R5
D=2001:db8:0:5555::/64 NH=R7 D=2001:
db8:0:6666::/64 NH=SERb2
D=::/0 NH=R7
packets from 2001:db8:0:b000::/52
D=2001:db8:0:a010::/64 NH=R2
D=2001:db8:0:b010::/64 NH=R2
D=2001:db8:0:a020::/64 NH=R5
D=2001:db8:0:b020::/64 NH=R5
D=2001:db8:0:5555::/64 NH=R7
D=2001:db8:0:6666::/64 NH=SERb2
D=::/0 NH=SERb1
packets from other sources
D=2001:db8:0:a010::/64 NH=R2
D=2001:db8:0:b010::/64 NH=R2
D=2001:db8:0:a020::/64 NH=R5
D=2001:db8:0:b020::/64 NH=R5
D=2001:db8:0:5555::/64 NH=R7
D=2001:db8:0:6666::/64 NH=SERb2
D=::/0 NH=SERb1
No unscoped table => no spoofing!!

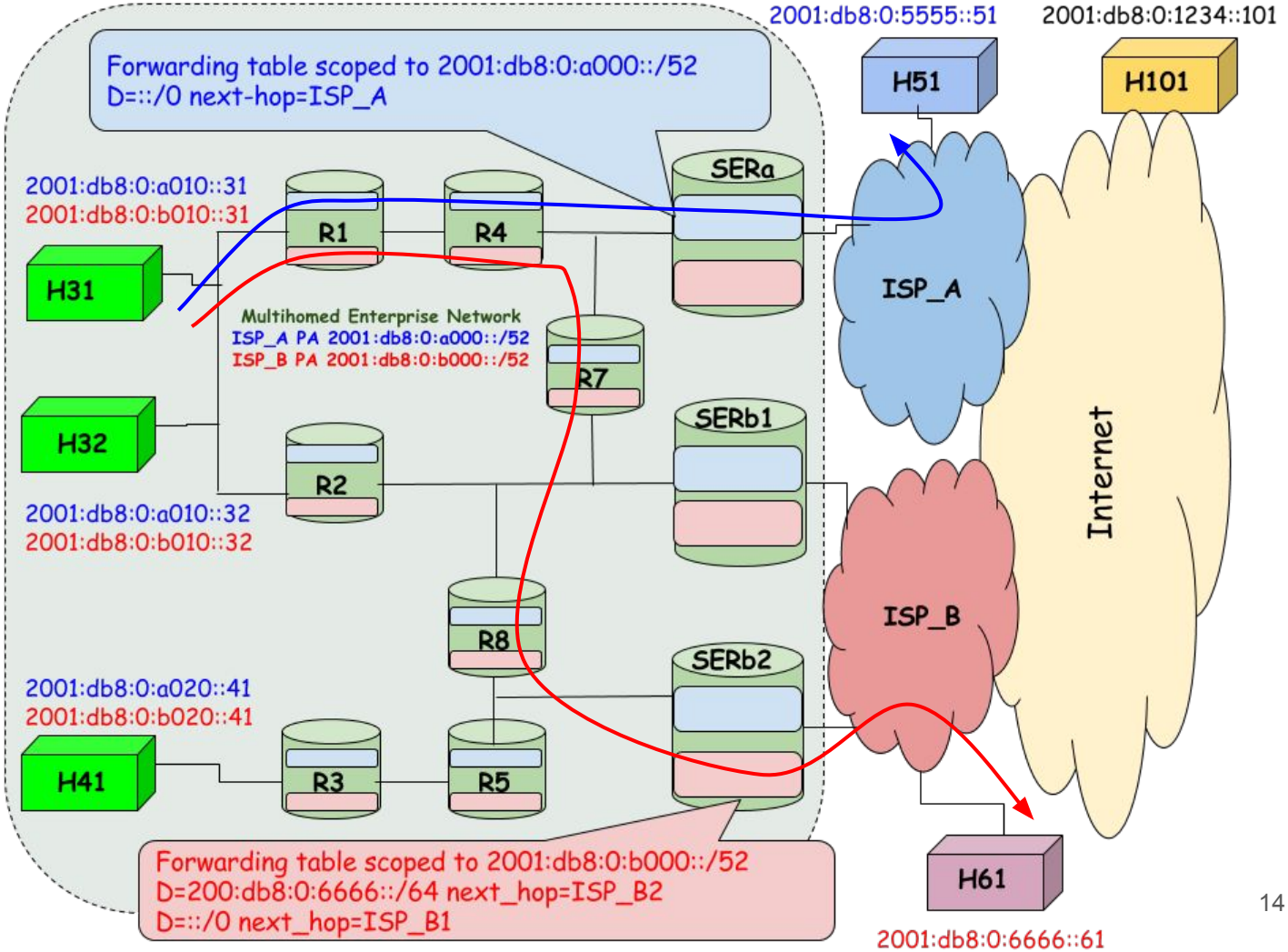
Incremental Deployment

Step 1
SERs Only



Incremental Deployment

Final Step: All Routers



Part 2: The Host

Source Address Selection

Source Address Selection Rules (RFC6724)

Rule 1: Prefer same address

Rule 2: Prefer appropriate scope.

Rule 3: Avoid deprecated addresses.

Rule 4: Prefer home addresses.

Rule 5: Prefer outgoing interface.

Rule 5.5: Prefer addresses in a prefix advertised by the next-hop

Rule 6: Prefer matching label.

Rule 7: Prefer temporary addresses.

Rule 8: Use longest matching prefix.

How to Influence Source Address Selection

DHCPv6: Labels Table Distribution (Rule 6)

SLAAC (RAs):

Next-hop (Rule 5.5)

Preferred vs. deprecated addresses (Rule 3)

ICMPv6: Signalling "incorrect" source address back to hosts

Distributing Address Selection Policy Using DHCPv6

Modifying the labels to influence source address selection (RFC7078)

E.g. "*Use **2001:db8:0:a000::/52** to access Internet and*

***2001:db8:0:b000::/52** to access ISP_B services such as*

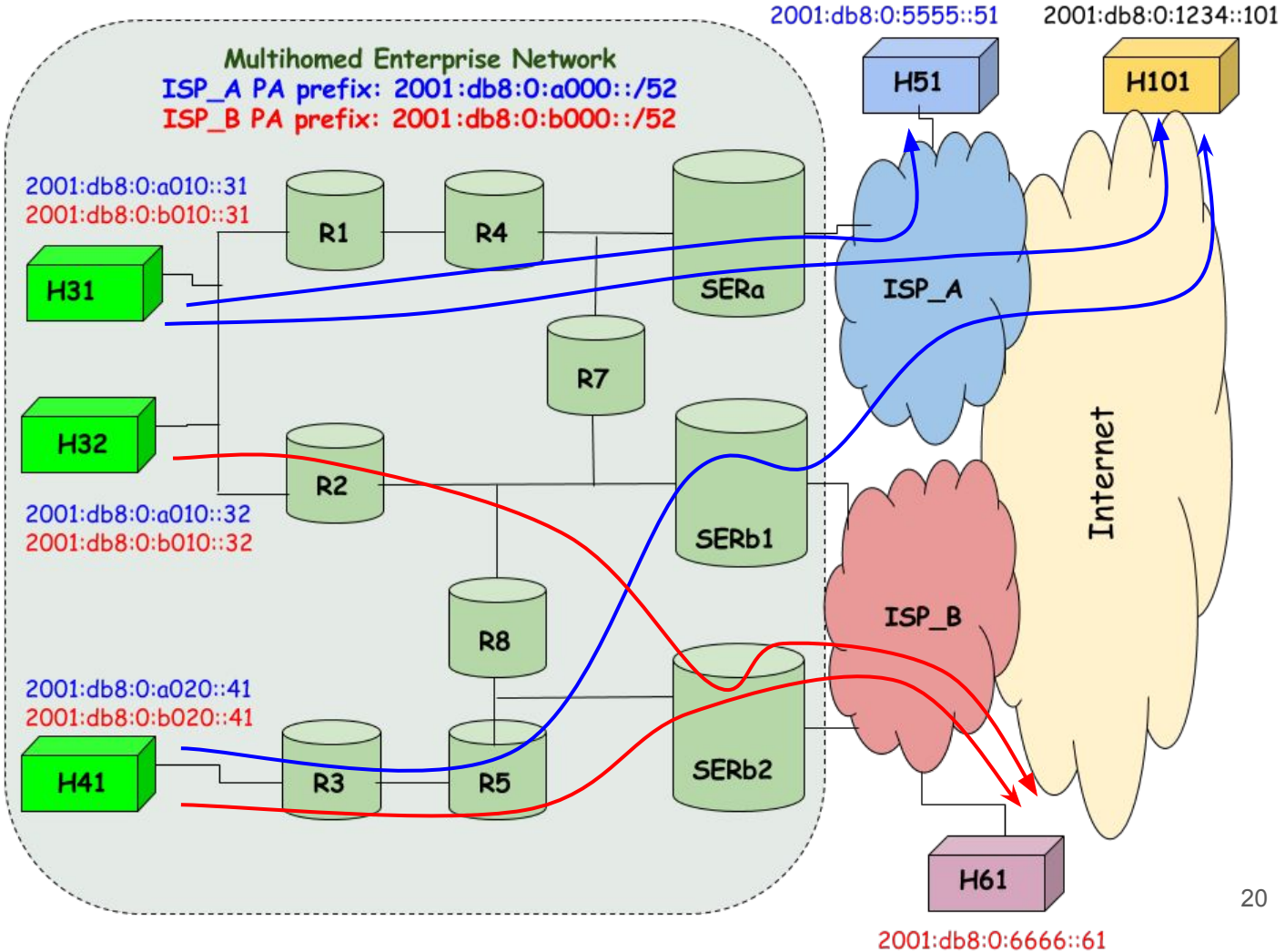
H61 (2001:db8:0:6666::61)"

Prefix	Label
2001:db8:0:6666::/64	33
2001:db8:0:b000::/52	33

DHCPv6 & RFC7078: Drawbacks

- DHCPv6 support is not a hard requirement for hosts
- RFC7078 is not widely implemented
- Policy configuration might be complex
- How to react to topology changes?
- Failover time
- Scalability

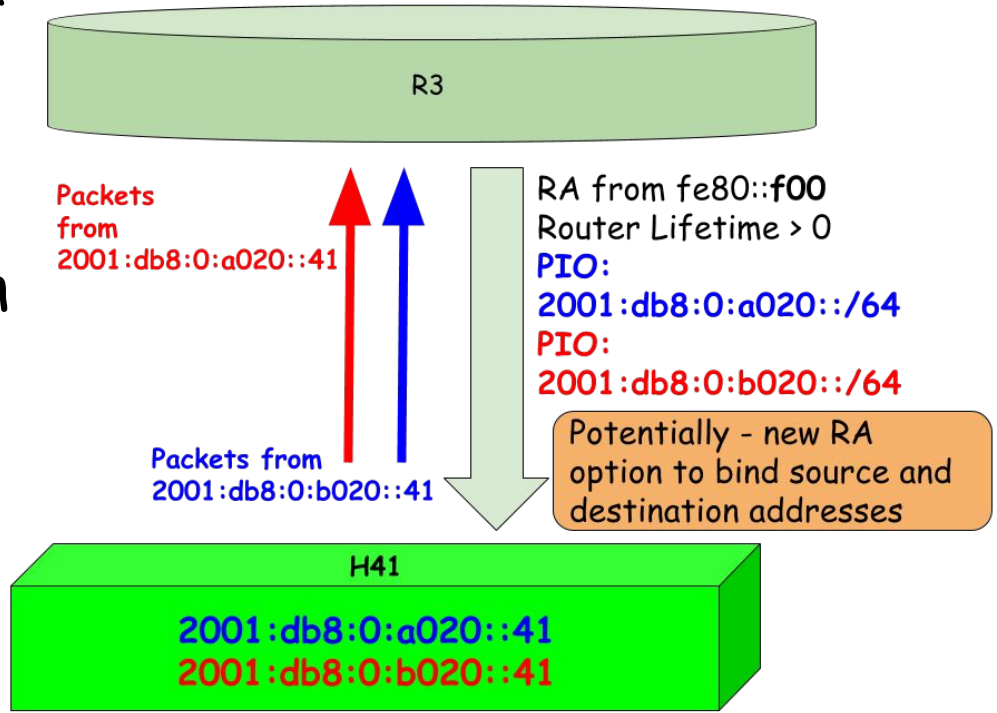
Example Topology



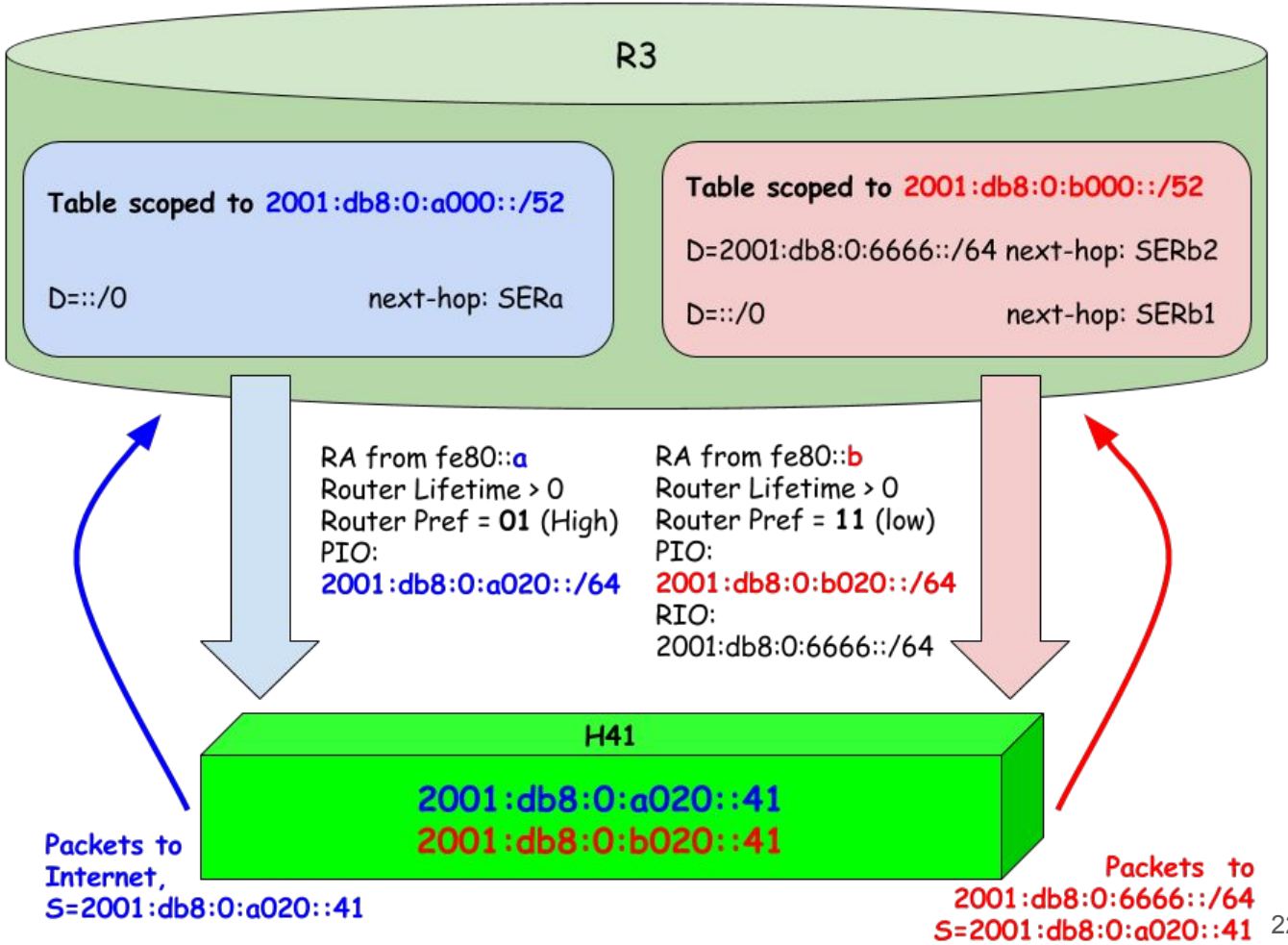
RAs and Source Addresses Selection

Long term solution: invent a new RA option to associate source and destination?

Can we have tactical solution w/o too many changes on hosts?

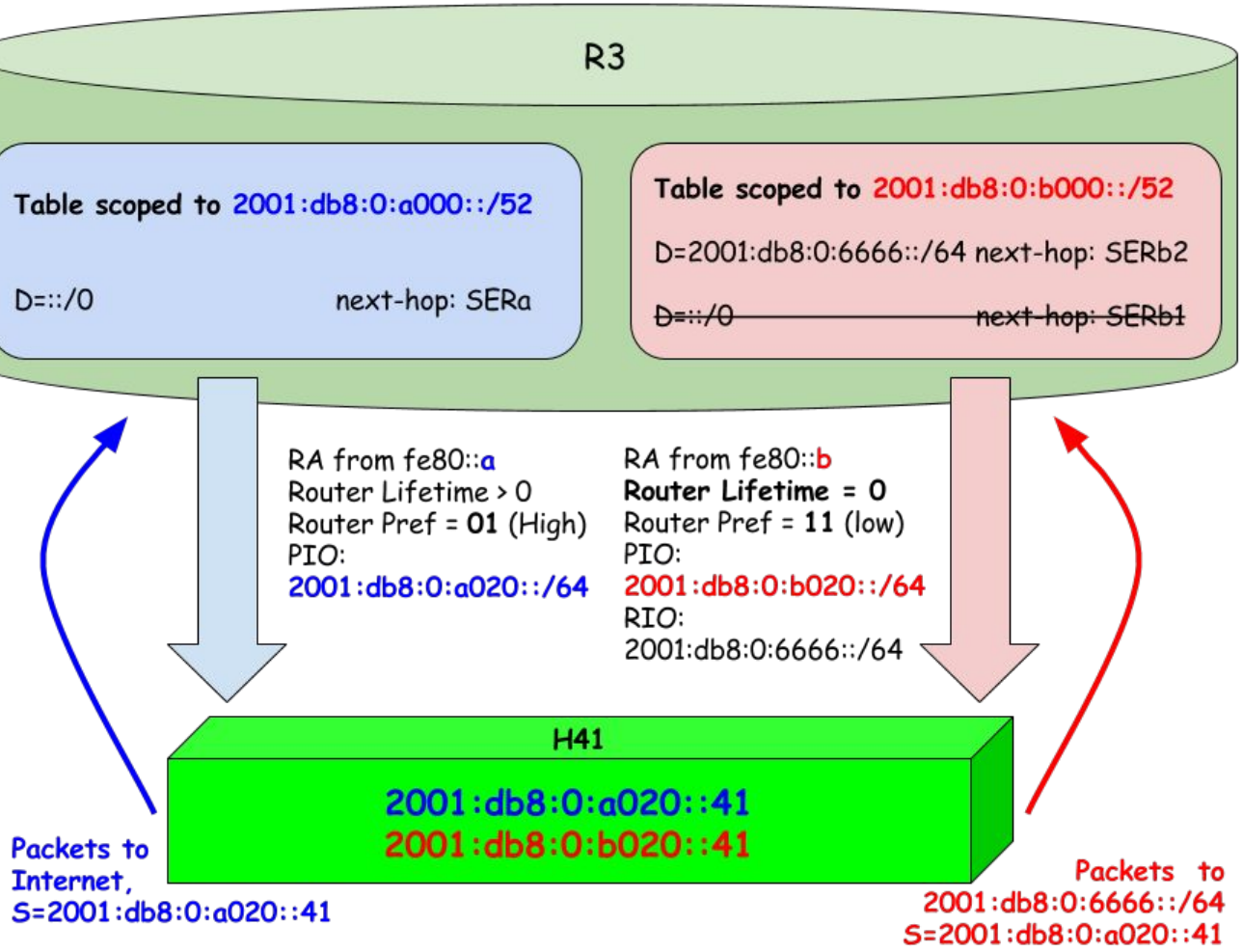


"Scoped" RAs, Router Pref and RIOs *(all uplinks are up)*

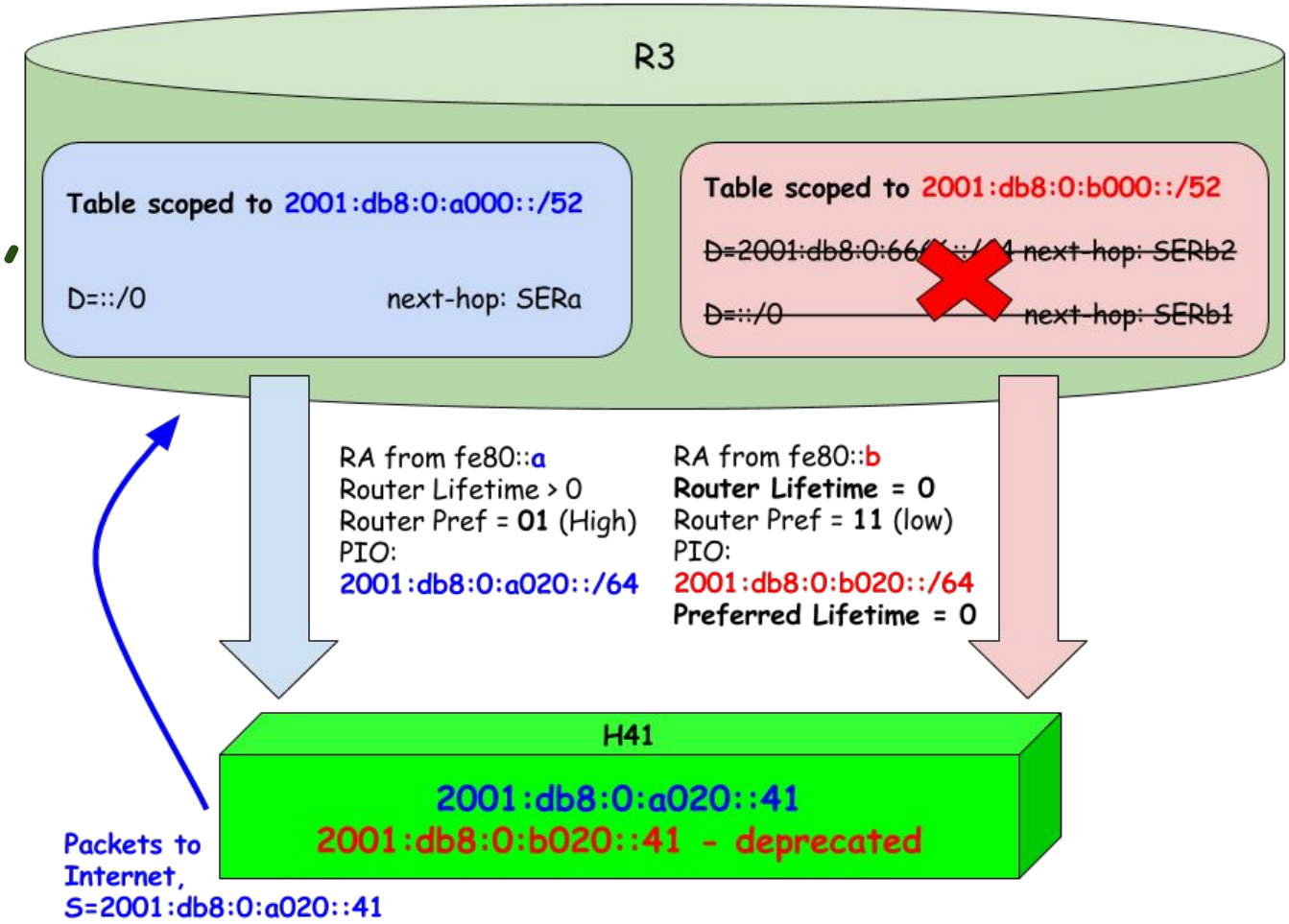


Scoped RAs, Router Pref and RIOs

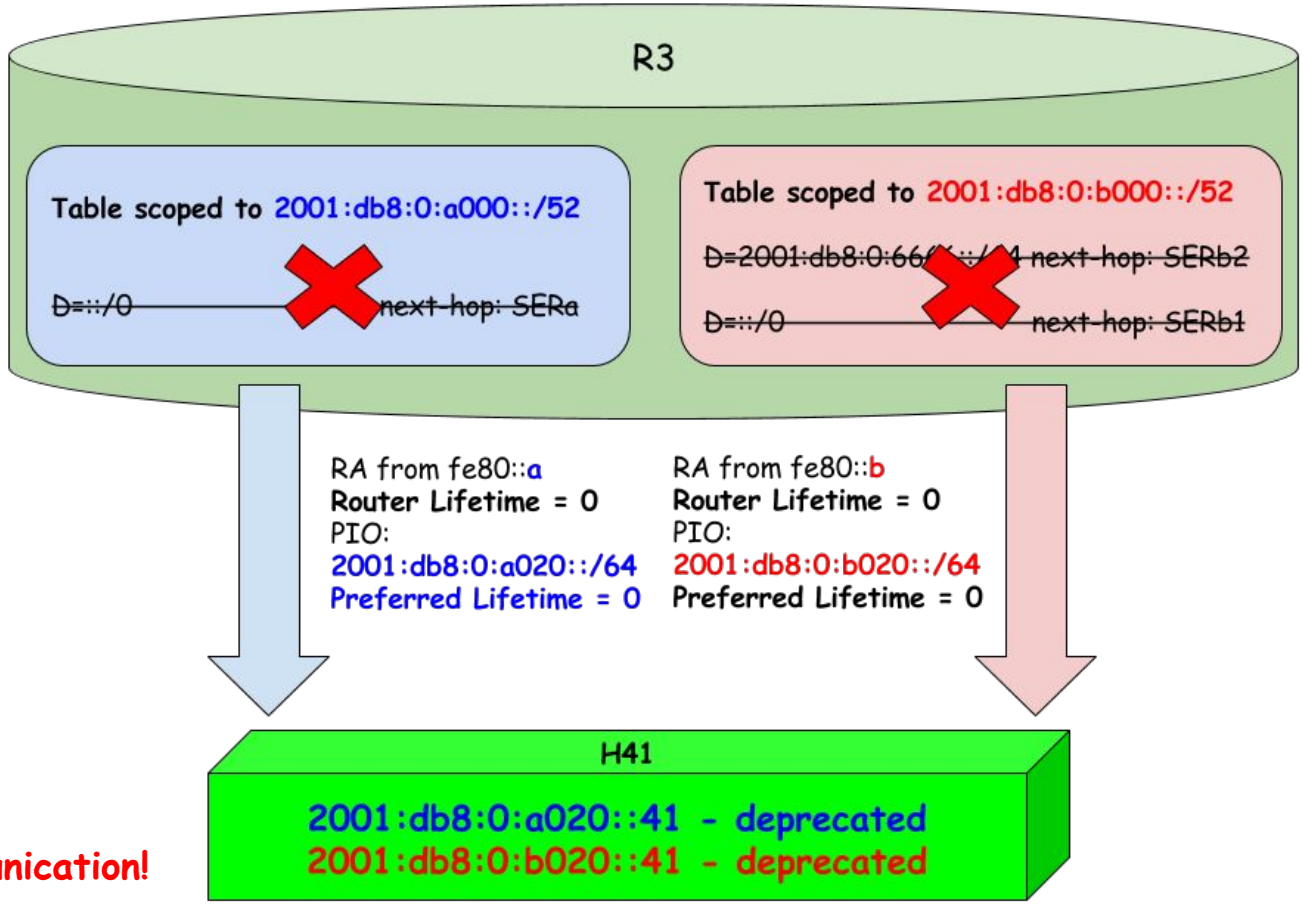
(SERb1 uplink is down)



Scoped RAs, Router Pref and RIOs *(all ISP_B uplink down)*



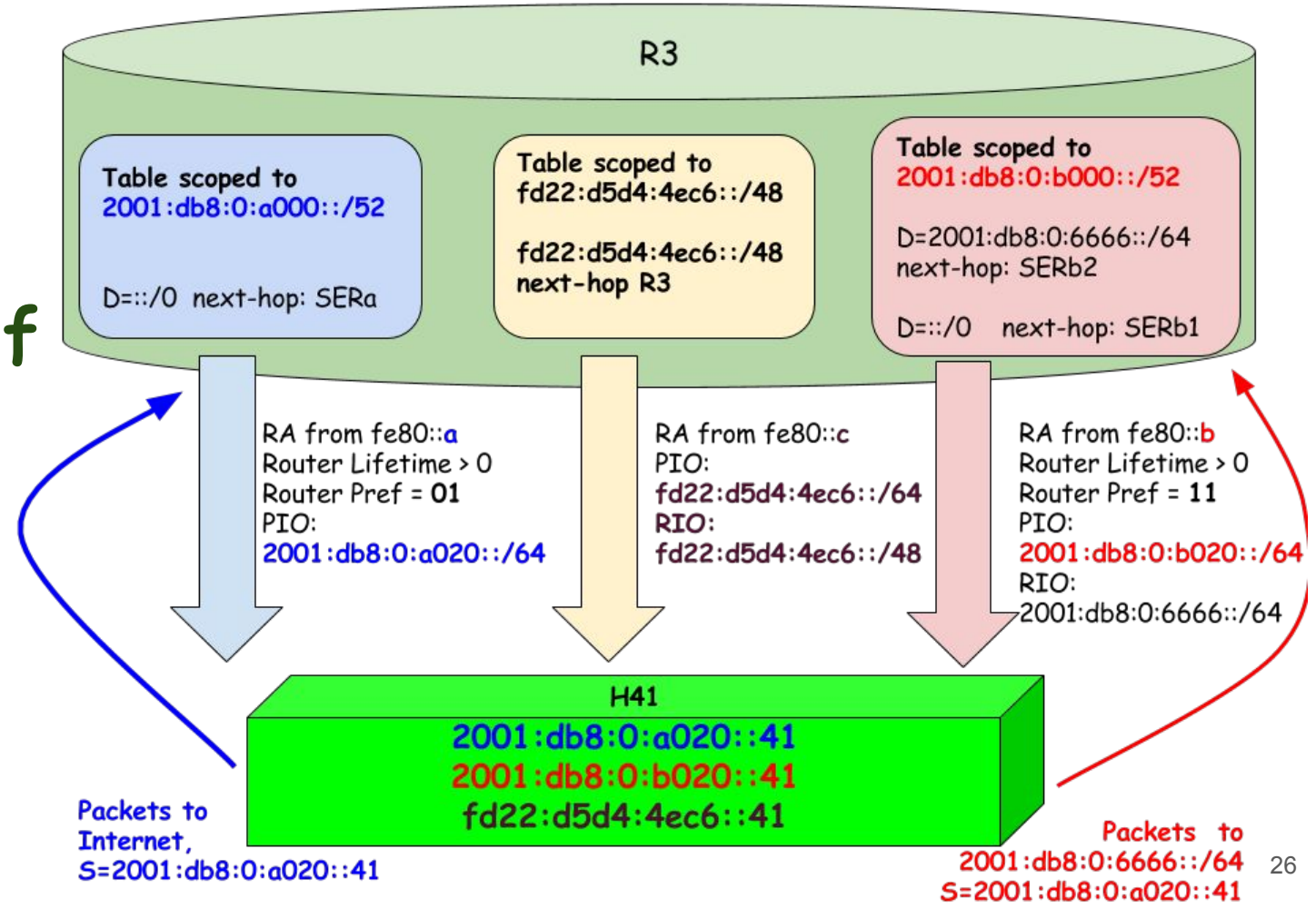
Scoped RAs, Router Pref and RIOs *(all uplinks down)*



Broken intra-site communication!

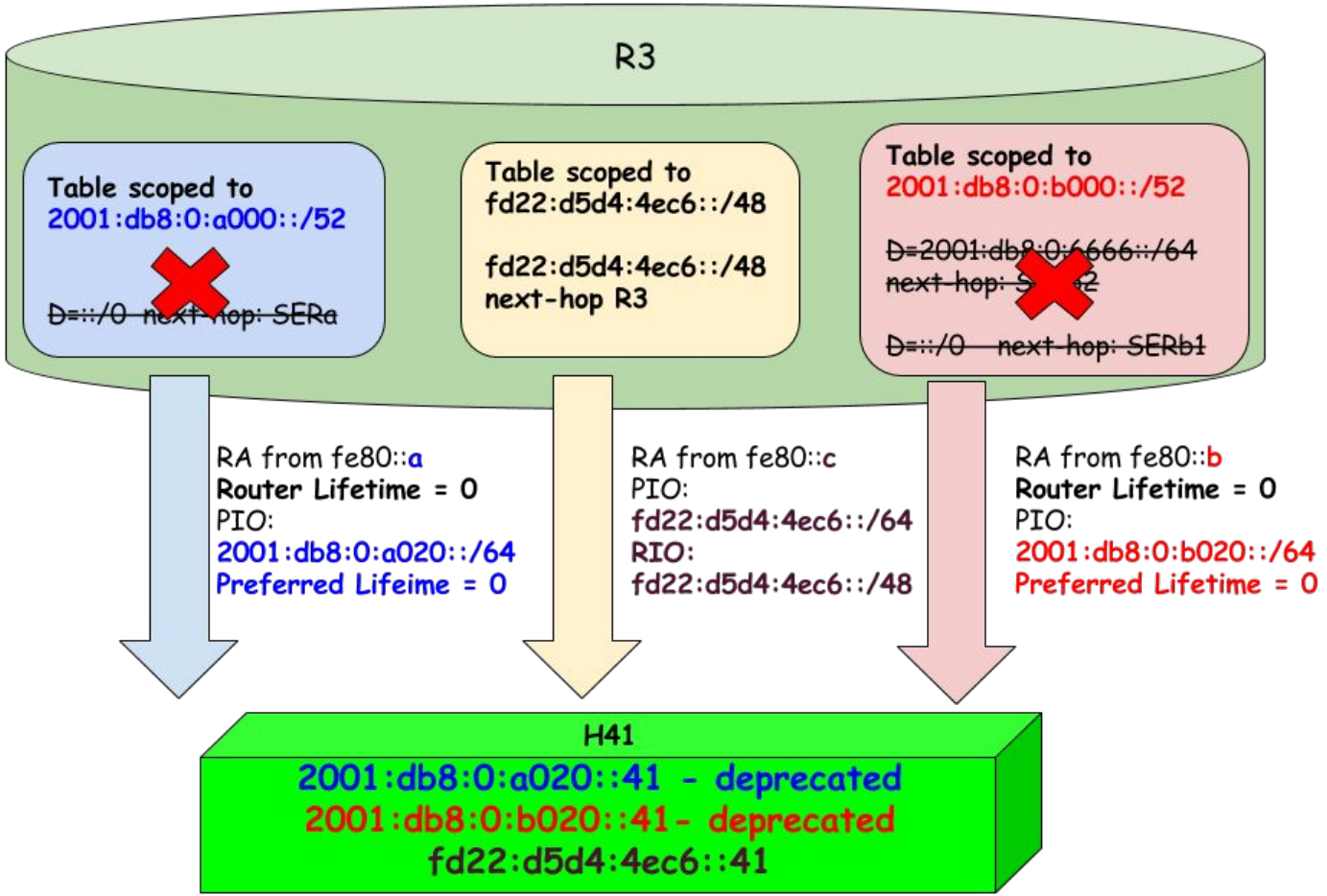
Scoped RAs, Router Pref and RIOs

ULA Usage



Scoped RAs, Router Pref and RIOs

ULA Usage For intra-site communication



ICMPv6

ICMPv6 Destination Unreachable (Type 1, Code 5): "Source address failed ingress/egress policy"

Potential Issues:

- Scalability?
- Delay (trying all available IPs)?
- If the "right" prefix stored in the Destination Cache: for how long?
  - Failed uplink recovery scenario
- Shall the host be informed of the "right" prefix to use?

How hosts behave in real world? - to be investigated

Summary: Network

- SADR allows network to send packets to the "right" egress point
- SADR can be deployed incrementally
- MUST be enabled on the edge
- Enabling on first-hop routers helps to control address selection on hosts

Summary: Source Address Selection on Hosts

- SADR-capable routers sending scoped RAs allow hosts to select the correct source address
- No changes in hosts behaviour are required for hosts supporting *(some testing required)*:
  - RFC4191 (Default Router Preferences and More-Specific Routes)
  - Rule 5.5 of Source Address Selection Algorithm
- If local connectivity is required when all uplinks are down: use ULAs
- ICMPv6 could be used to signal errors

QUESTIONS/COMMENTS?