(my sponsor)

Border Router Discovery Protocol and BRDP Based Routing

Exit routing for multi-homed networks

Teco Boot - 18 November 2008

IPv6 Operations (v6ops)

BRDP Based Routing
marketing ☺ data sheet

- ***Built on well known protocols such as IPv6 Neighbor Discovery and Policy Based Routing***
- ***Works with all IGPs***
- ***Full support for multi-homing with Provider Independent (PI) and Provider Aggregatable (PA) addresses***
- ***Helps scaling the Internet by removing need for PI addresses***
- ***Border Router load balancing***
- ***Automatic ingress filter on first hop routers***
- ***No tunnels, routing headers or any other encapsulation***
- ***Extensible for ad hoc networking***

BRDP Based Routing
marketing ☺ data sheet *page 2*

List of incompatible protocols:

BRDP Based Routing

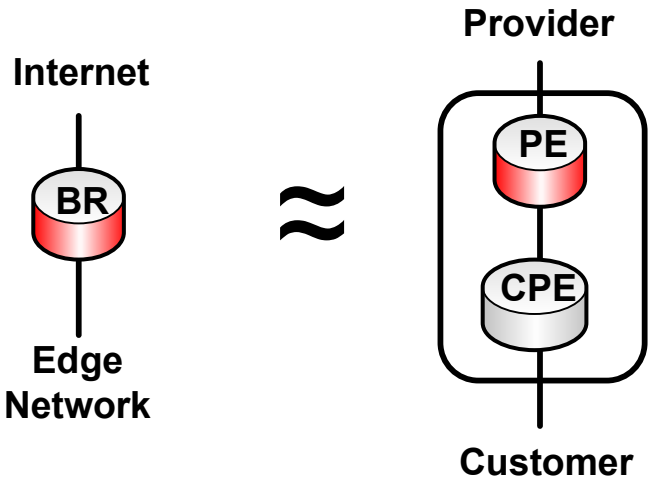
Introduction

Analysis of the problem

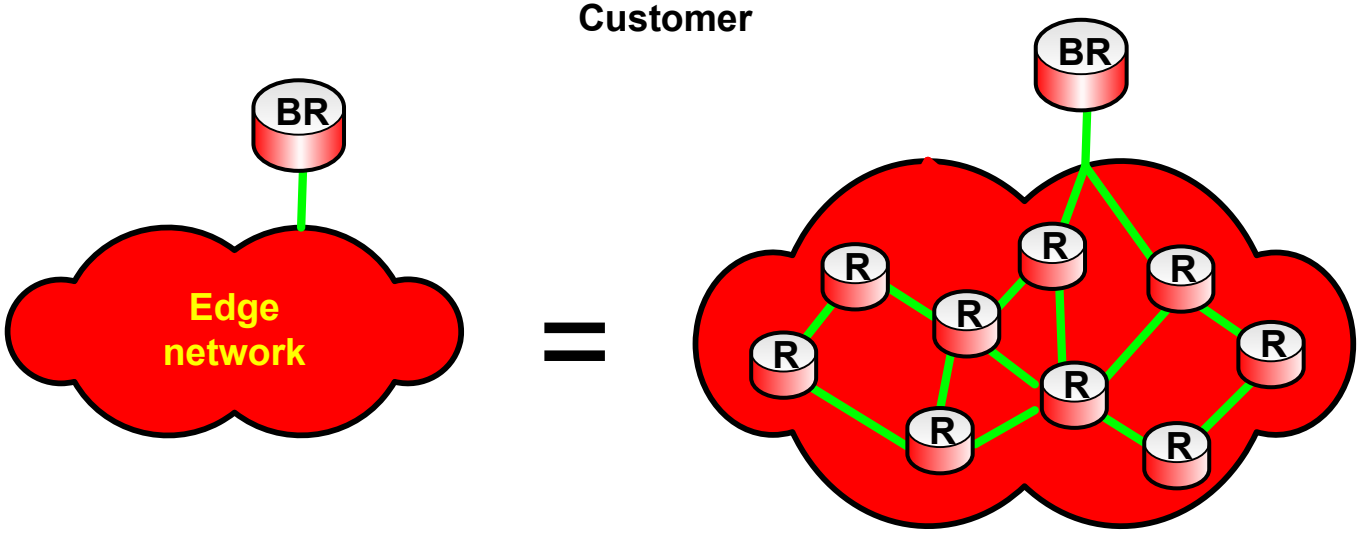
How does it work?

Next steps

My illustrations are simplified

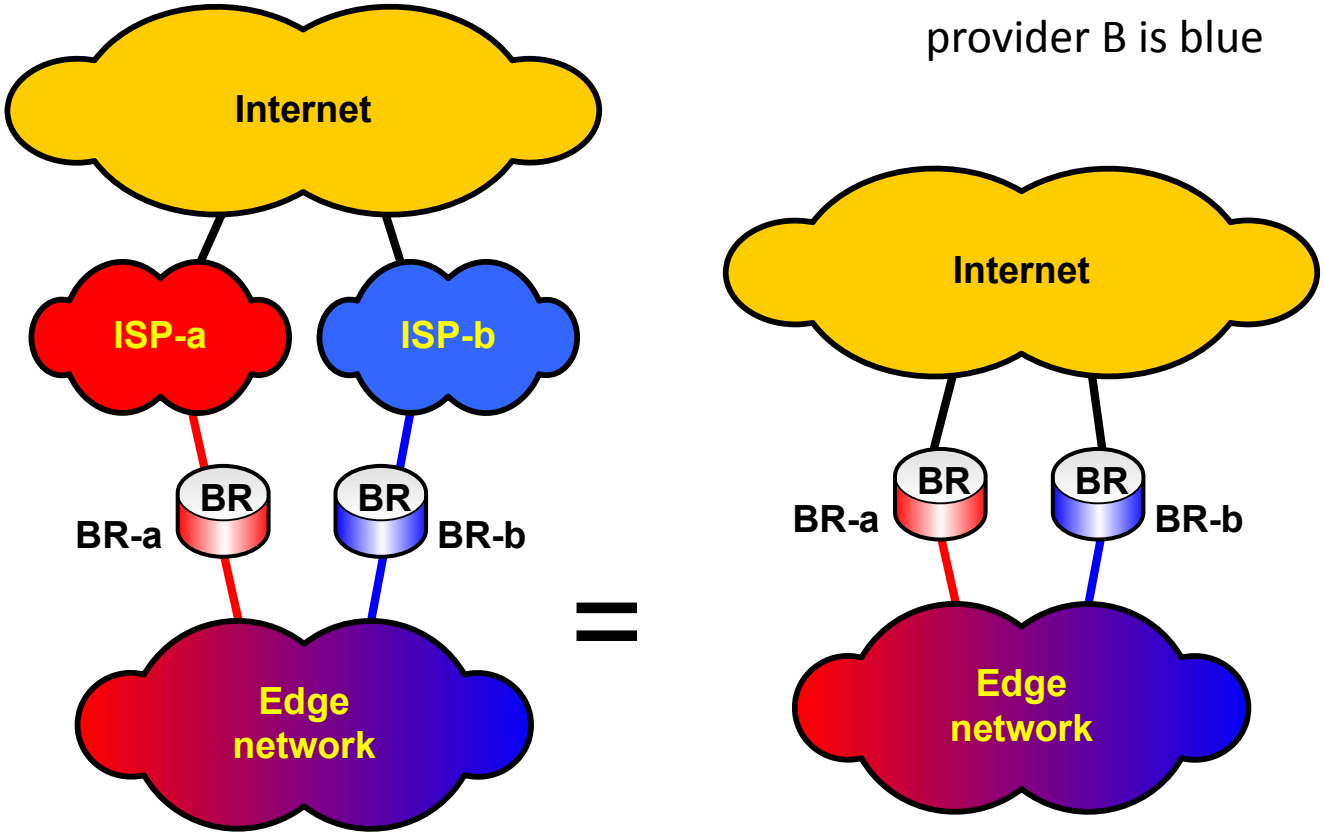


- Border Router (BR) is a router connecting an edge network to an ISP
- In practice, there is a Provider Edge (PE) and Customer Premises Equipment router
- CPE – PE protocols are out of scope for this presentation
- PE may (should!) have ingress filter



My illustrations are simplified

- ISPs have their networks. Details on this are not important for BRDP Based Routing
- Note the colors: Provider A is red, provider B is blue



Problems with multi-homed networks

Problem #1: How do site ***interior routers*** quickly learn which ISP uplink routing prefixes are currently valid ?

Problem #2: If one wants the ***end host*** to make the upstream ISP selection, how does one provide enough information about (bandwidth, cost, current availability, congestion) of the various upstream links to the end hosts so that they can make a good decision ?

Problem #3: If an end host selected an upstream ISP, how can the ***routing system*** direct traffic to this ISP ?

Text problem #1 and #2: Ran Atkinson
See also RFC 5220: *Problem Statement for Default Address Selection in Multi-Prefix Environments: Operational Issues of RFC 3484 Default Rules*
And RFC 5221: *Requirements for Address Selection Mechanisms*

Problems with default gateway routing in a multi-homed network

- Next hop selection is based on destination address
- Multiple DGWs (default gateways):
  - Source has no influence on what DGW is used for sent traffic
  - Three options (at least)
    - Single DGW is used
    - Packet load balancing
    - Flow load balancing
  - Seen from source perspective: used BR (Border Router) is ***guesswork***
- BRs may have ingress filters

- Result:

  In a multi-homed network, outgoing traffic could be blocked on BR when Source Address does not correspond with delegated prefix of used BR

See also “Ingress filtering incompatibility” problem, draft-ietf-shim6-ingress-filtering-00.txt (Oct-2006) and RFC3704 section 4.3. “Send Traffic Using a Provider Prefix Only to That Provider”

BRDP Based Routing

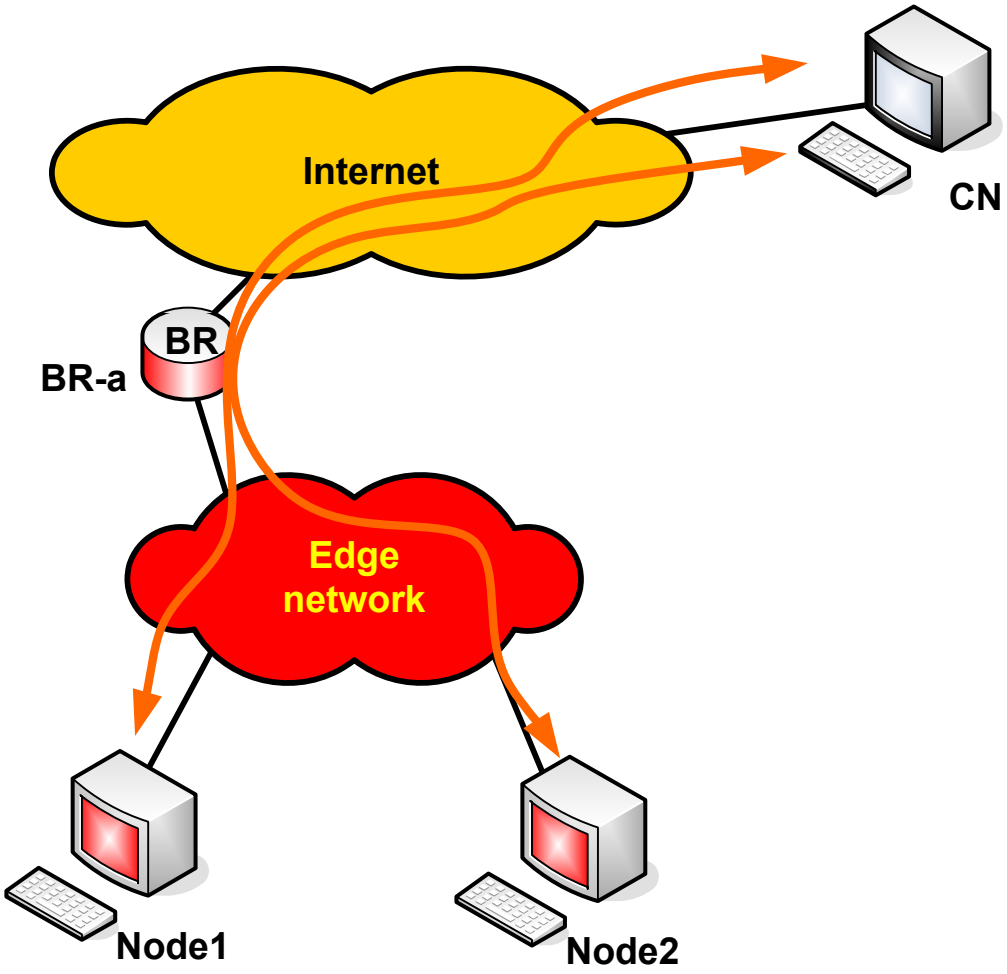
Introduction

Analysis of the problem

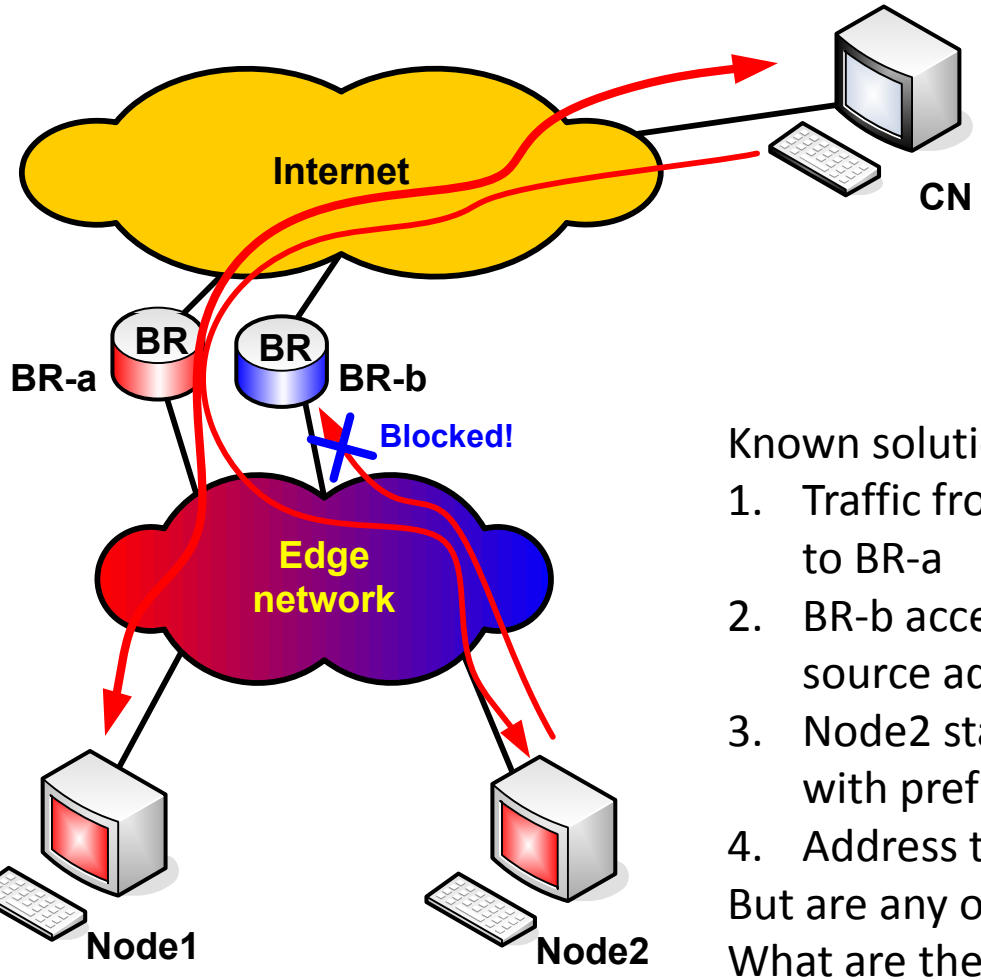
How does it work?

Next steps

Single-homed edge network: no problems ☺



Multi-homed edge network with traffic blocked by ingress filter

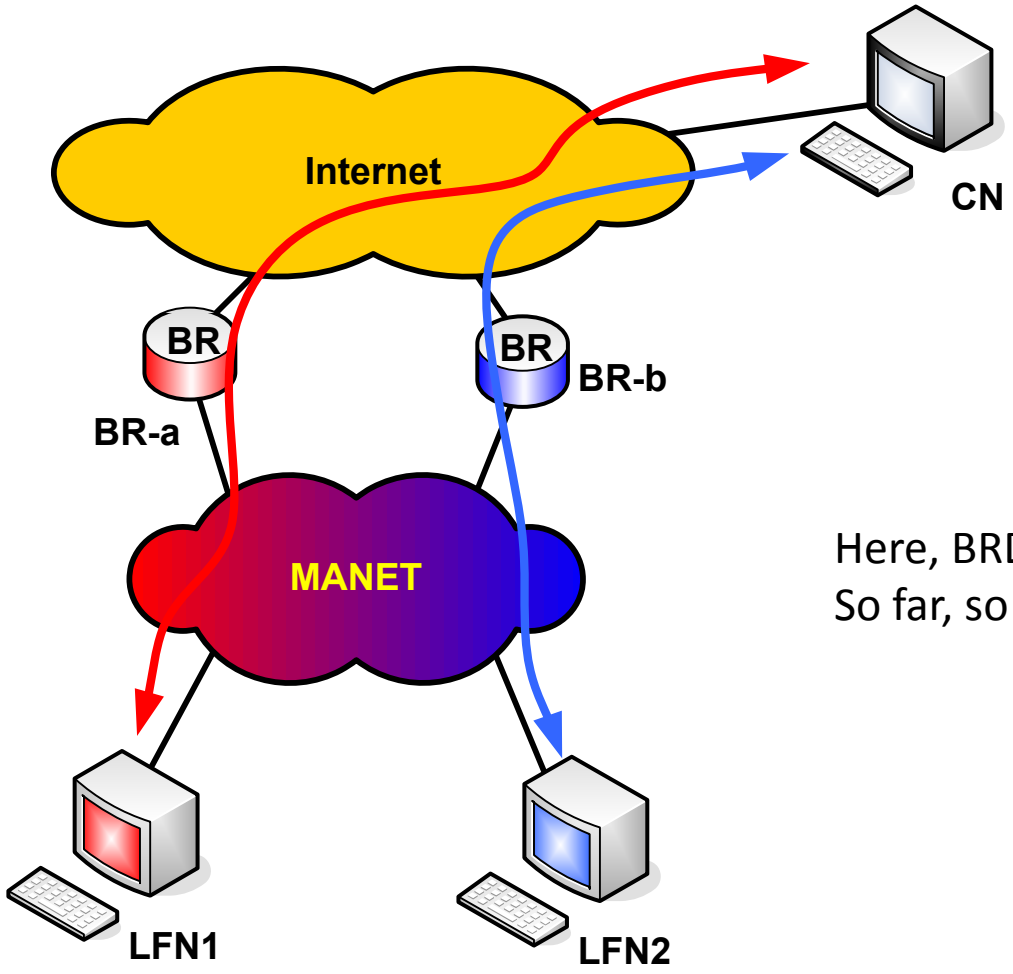


Known solutions:

1. Traffic from Node2 is forwarded to BR-a
2. BR-b accepts BR-a owned source addresses
3. Node2 starts new connection with prefix from BR-b
4. Address translation

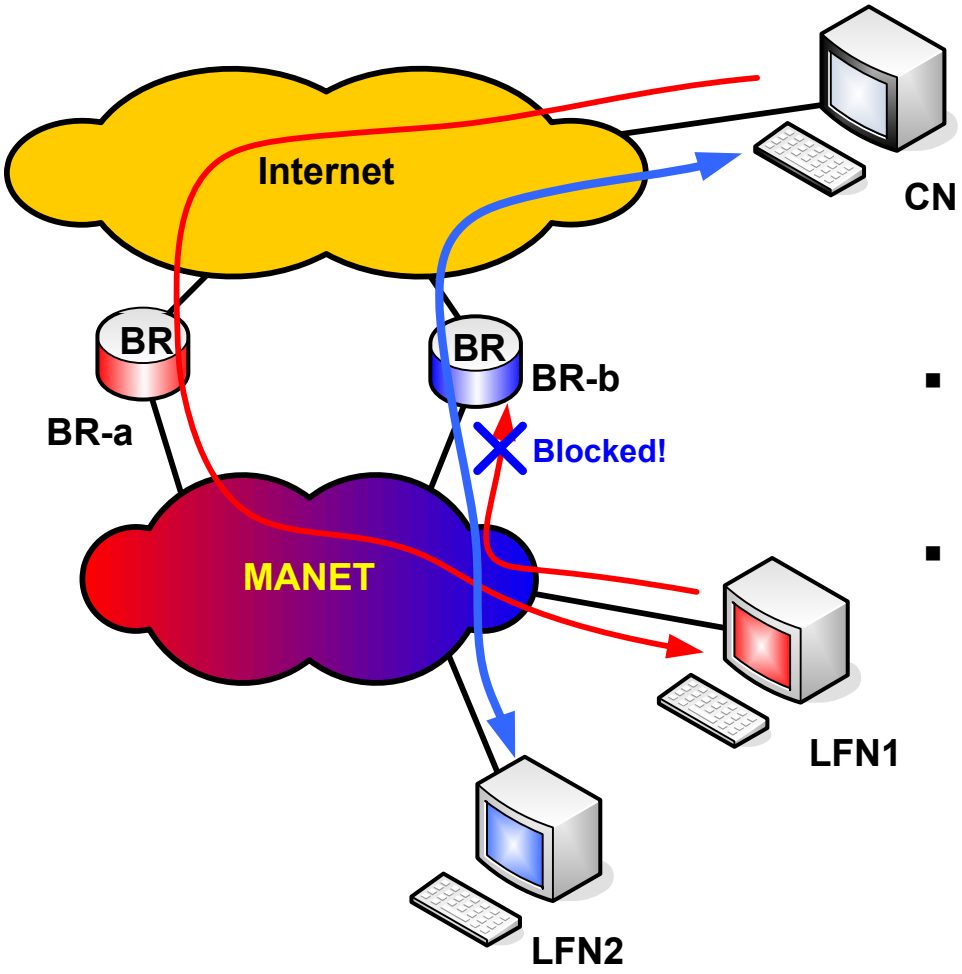
But are any of these implemented?
What are the issues ?

Multi-homed *MANET*



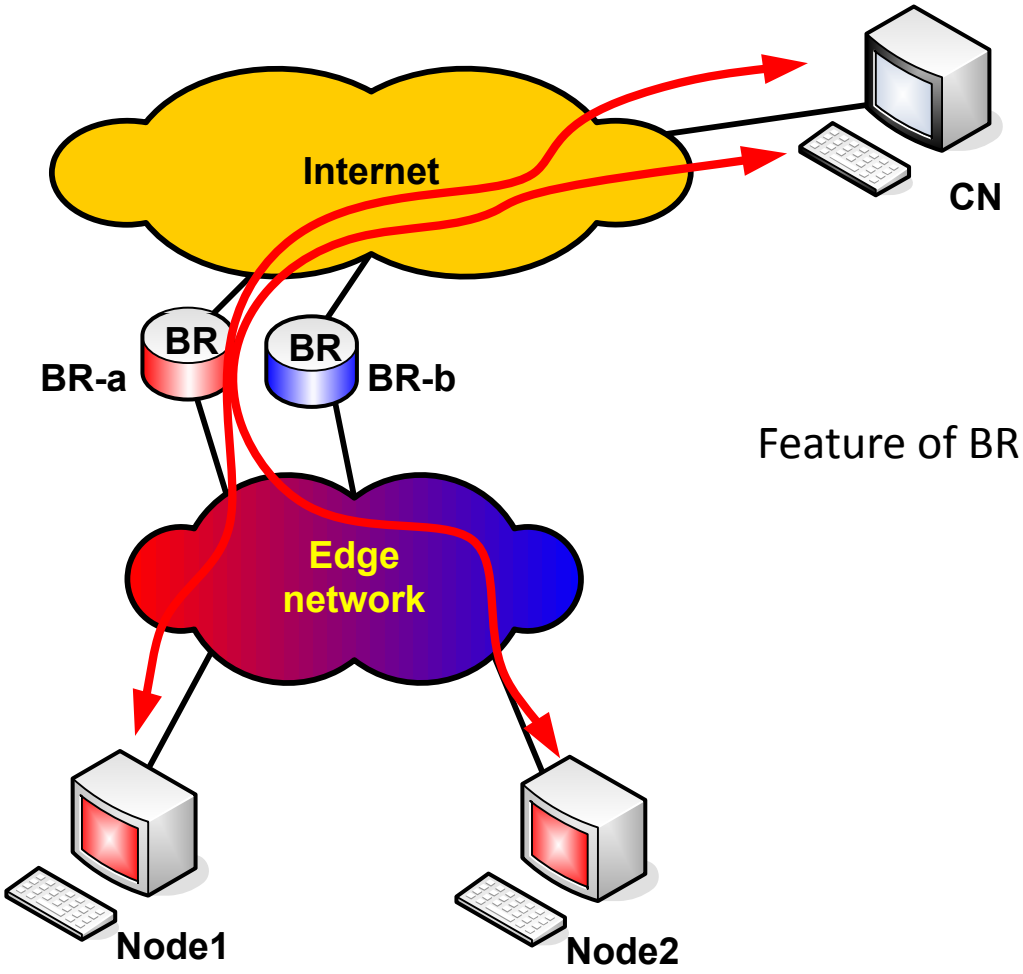
Here, BRDP was born.
So far, so good. ***But:***

In a MANET, things can *move*!



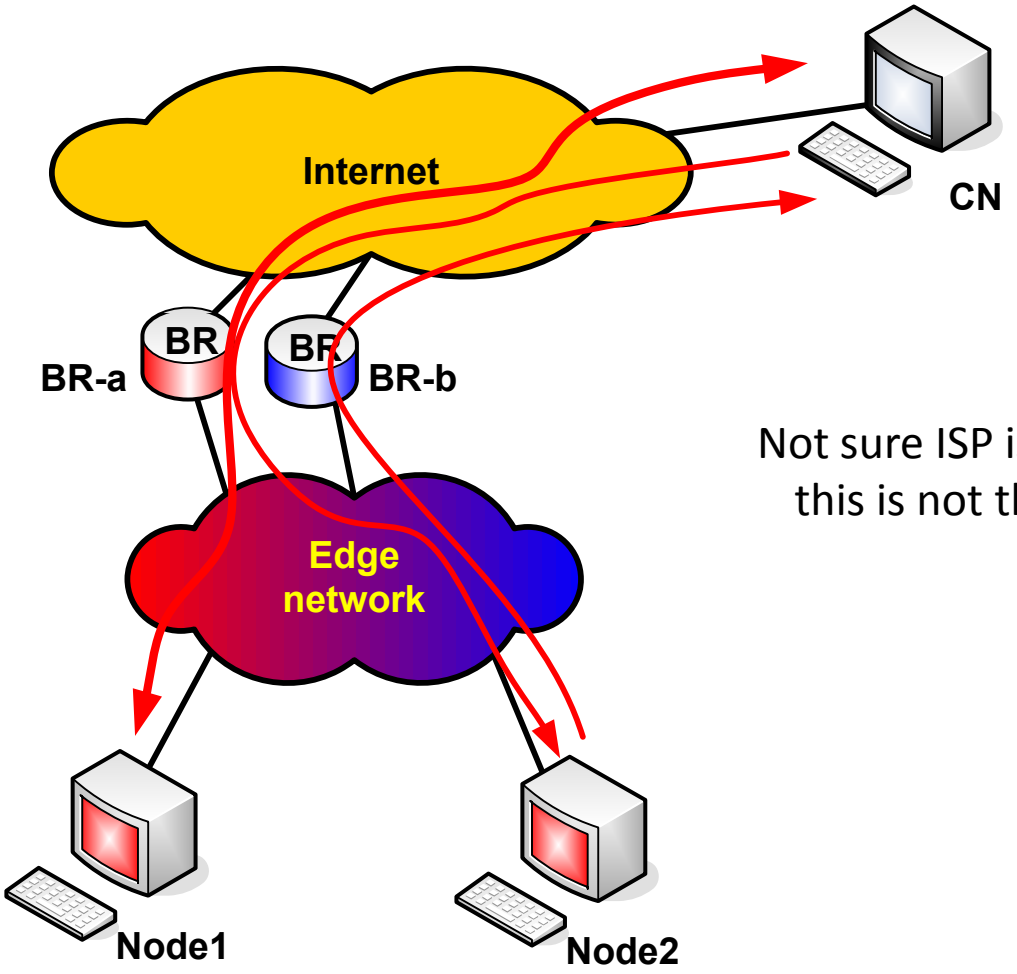
- MANET routing protocol converges to reflect the movement
- LFN is not aware of movement

Solution #1:
Direct traffic to BR that owns SA prefix



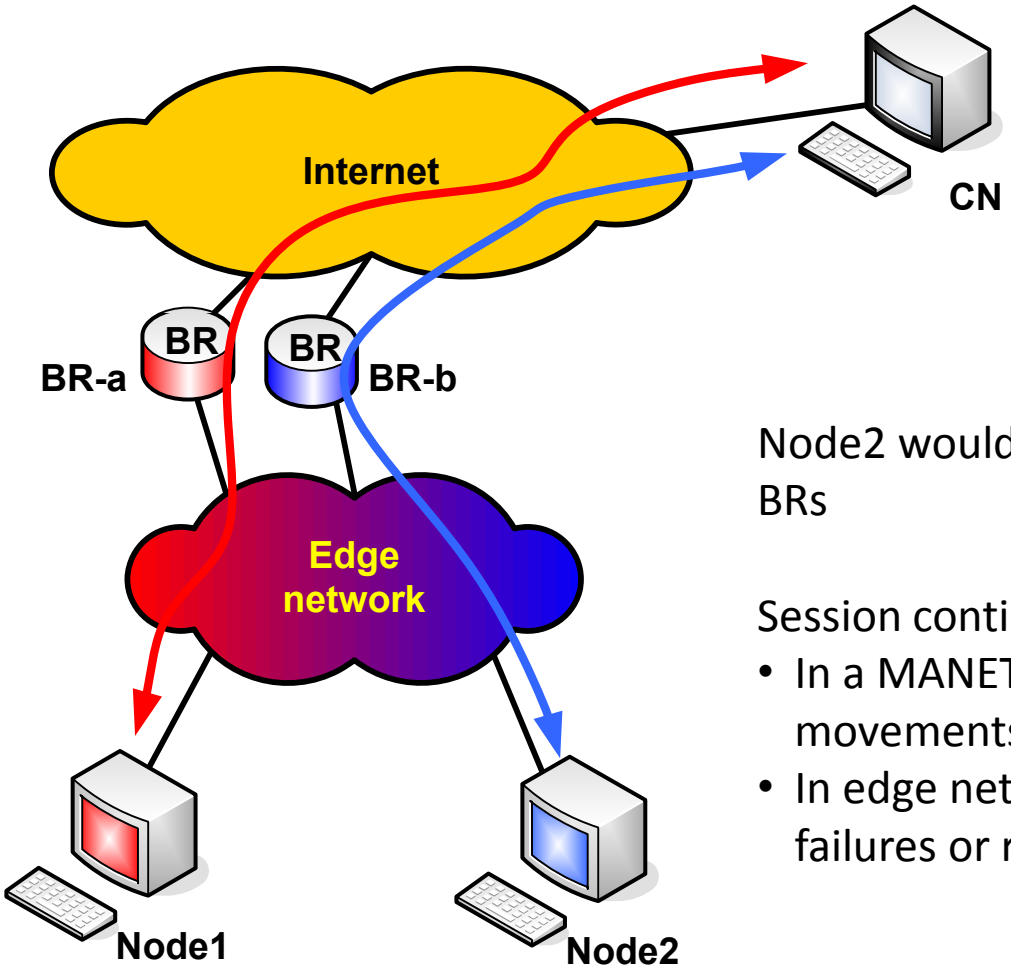
Feature of BRDP Based Routing!

Solution #2:
Adjust BR-b ingress filter for prefix BR-a



Not sure ISP is cooperative, so this is not the goal

Solution #3:
Select a SA that corresponds to BR used

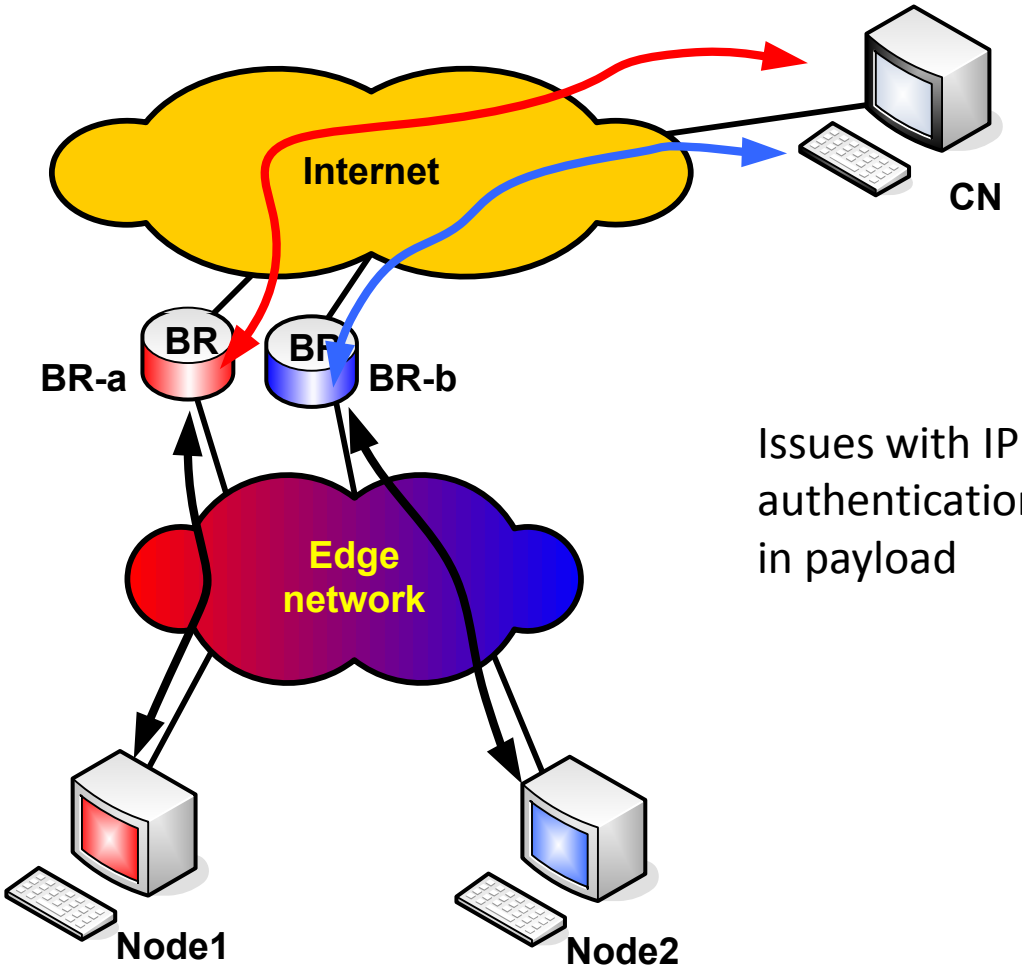


Node2 would use “metrics” of BRs

Session continuity problems:

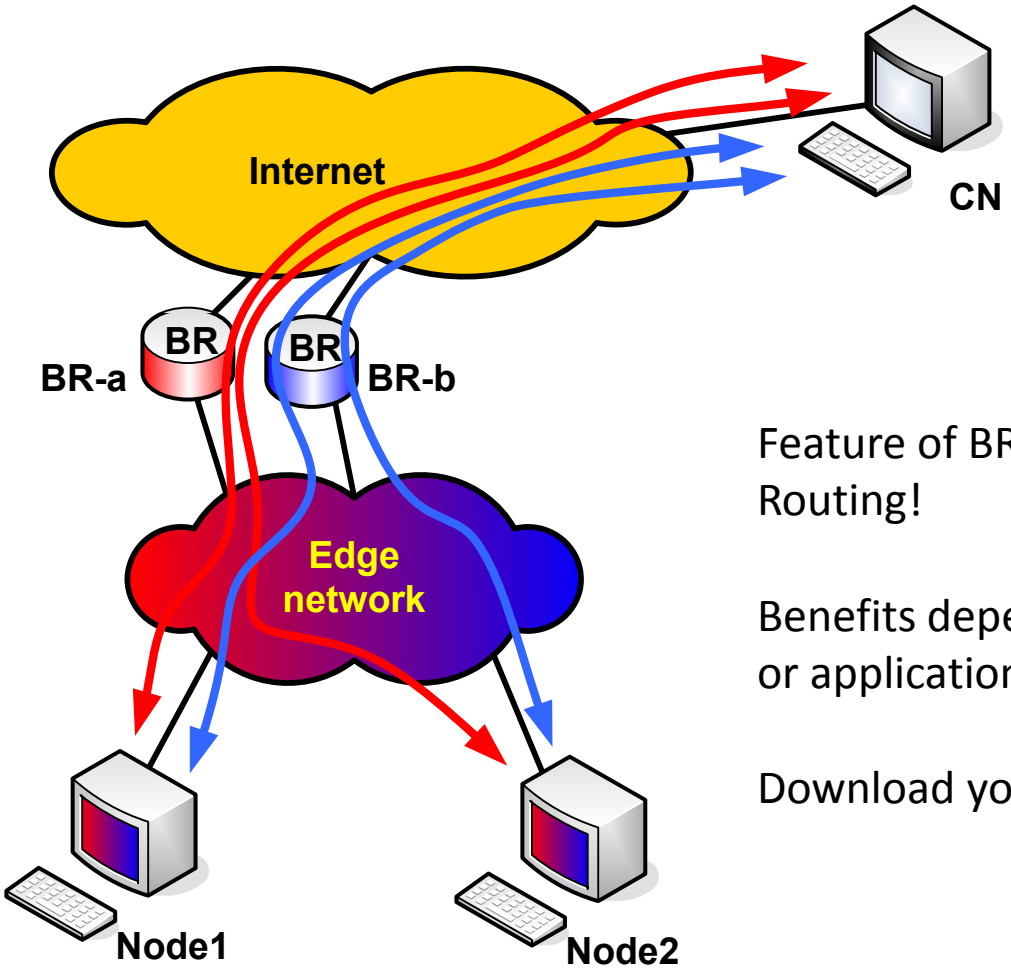
- In a MANET because of movements
- In edge network because of failures or renumbering

Solution #4:
Address translation on BRs



Issues with IP header authentication and addresses in payload

Goal: support for multi-homed edge networks with multi-homed nodes



Feature of BRDP and BRDP Based Routing!

Benefits depend on transport layer or application layer adjustments

Download your files twice as fast!

BRDP Based Routing

Introduction

Analysis of the problem

How does it work?

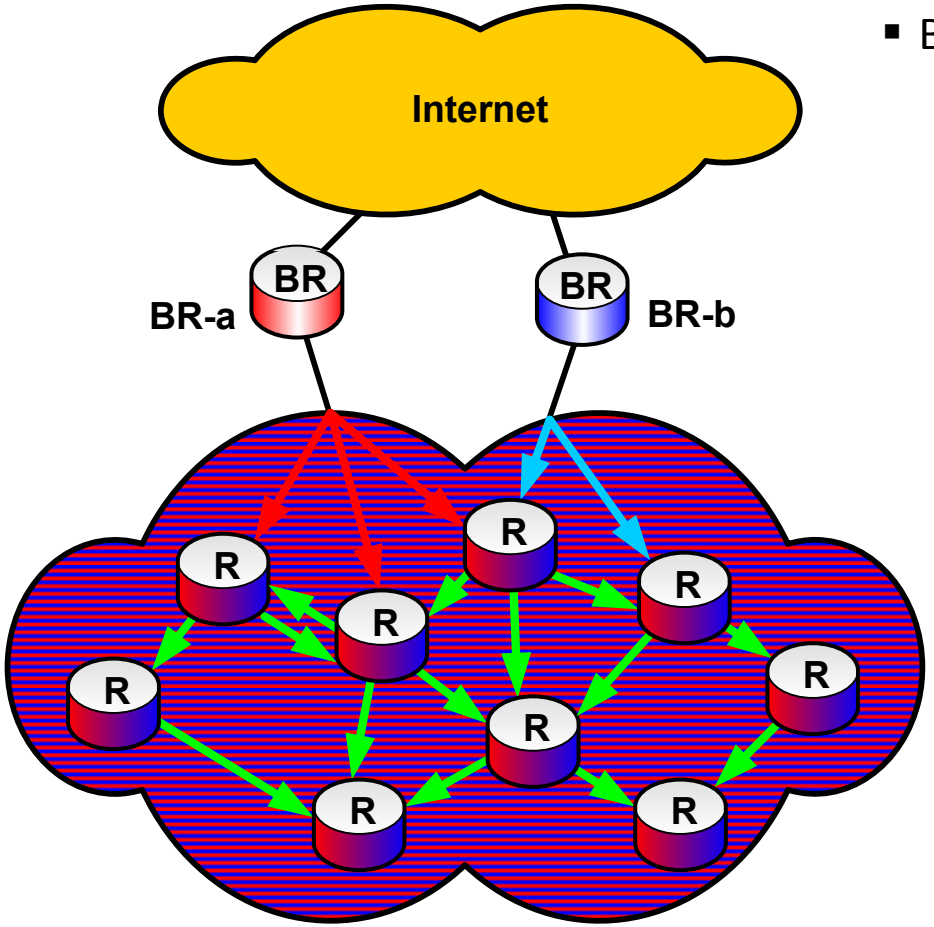
Next steps

Solution for problem #1

How do site ***interior routers*** quickly learn which uplink routing prefixes are currently valid ?

- **Border Router Discovery Protocol (BRDP):**
  - **Provides valid prefix information to interior routers**
  - **Distributes Border Router Information Option (BRIO) via Neighbor Discovery Router Advertisements.**
  - **BRIOs are distributed in the edge Network**
  - **BRIOs provide metrics for paths from and to DFZ (Default-Free Zone)**
  - **BRDP uses link metrics from routing protocol or other source**
    - **Use link metrics for both directions**
  - **BRIOs provide DHCP relay information**
  - **BRDP is a Distance-Vector protocol**

BRIO flooding using Neighbor Discovery Router Advertisements

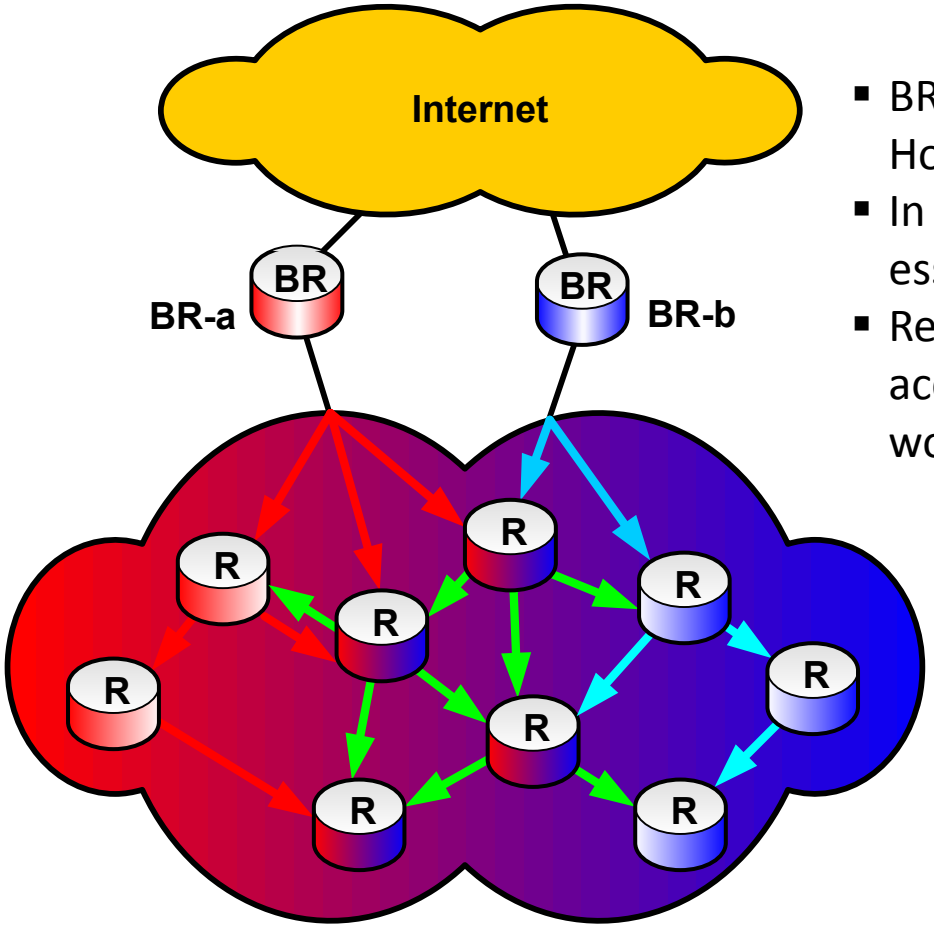


- Routers already send RA
- BRIOs piggy-backed on RA

Results:

1. All routers learn prefixes for BRs, with metric information
2. All routers learn DHCP addresses for relay or prefix delegation

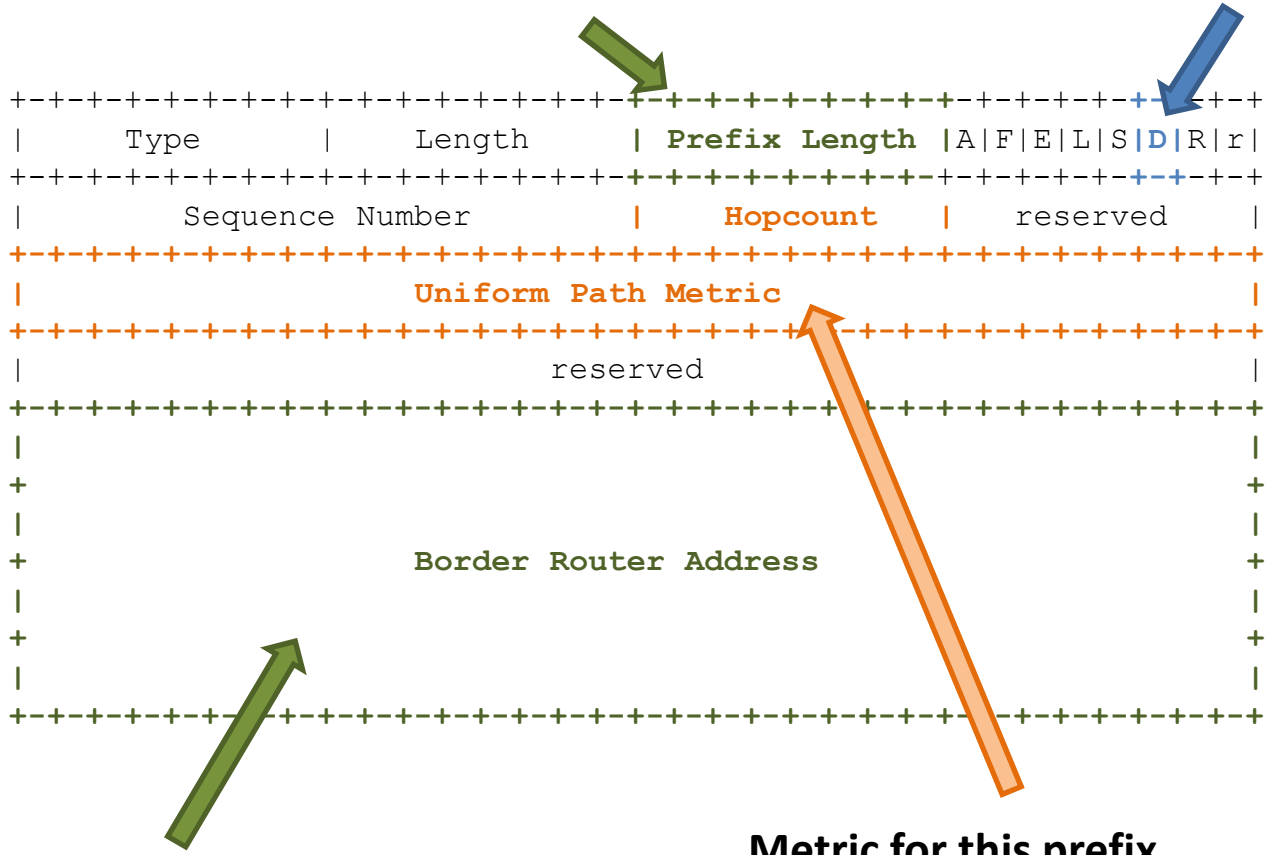
BRIO flooding with scoping (needed in large MANETs)



- BRDP has its roots in Autoconf (Ad-Hoc Network Autoconfiguration).
- In a MANET, overhead reduction is essential
- Reduction of forwarded BRIO set is acceptable, as only the best paths would be used

BRIO format

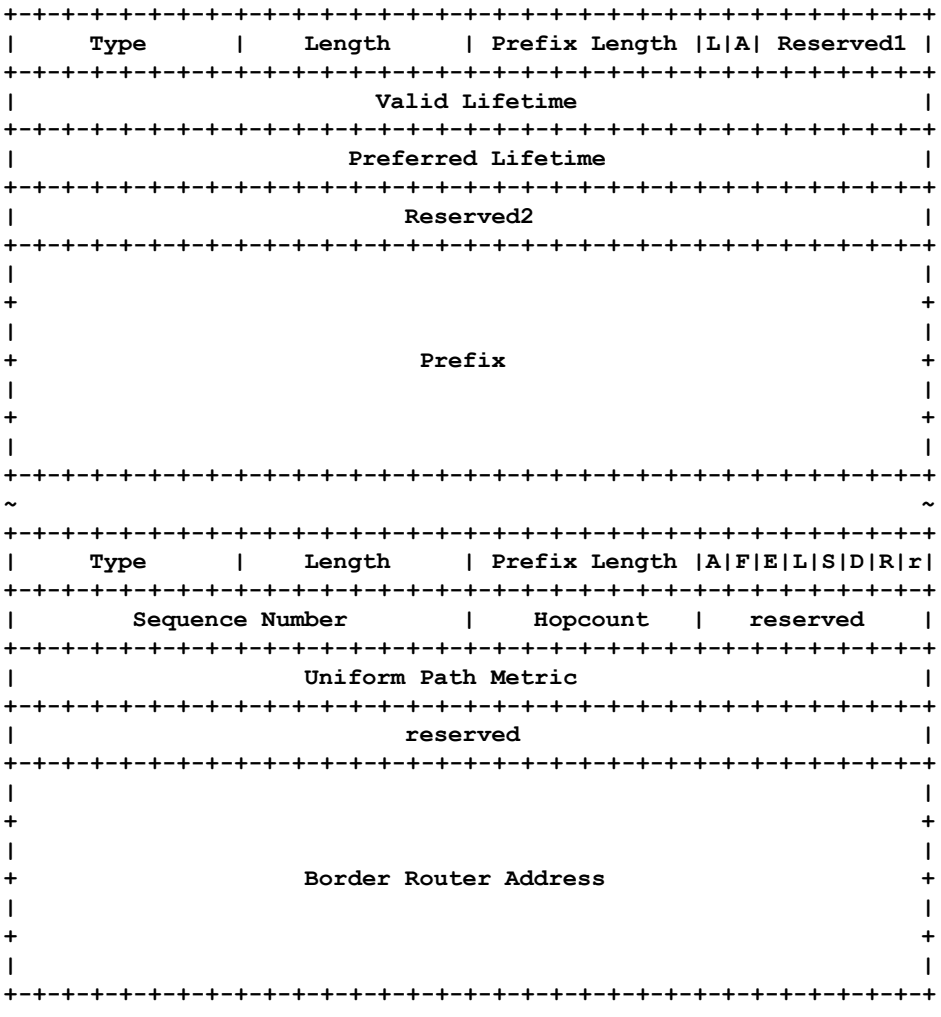
D=1: BR is DHCP server or relay



BR address and valid prefix for this BR

Metric for this prefix
Is a bidirectional metric between
DFZ and this BR

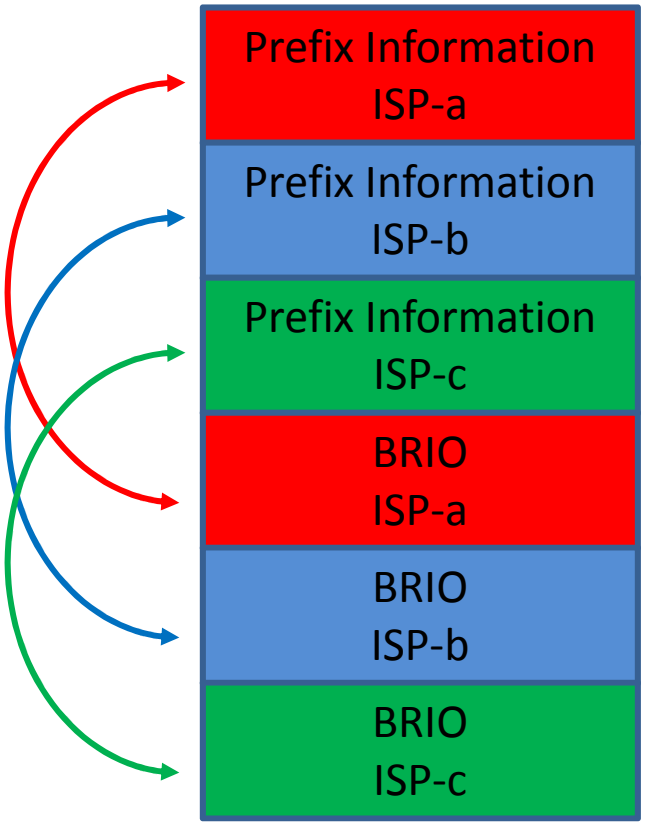
RA with Prefix Information and BRIOs



Prefix Information option
Router initiated

BRIO
BR initiated

RA in edge network with 3 uplinks



OK, the RA packet size increases.

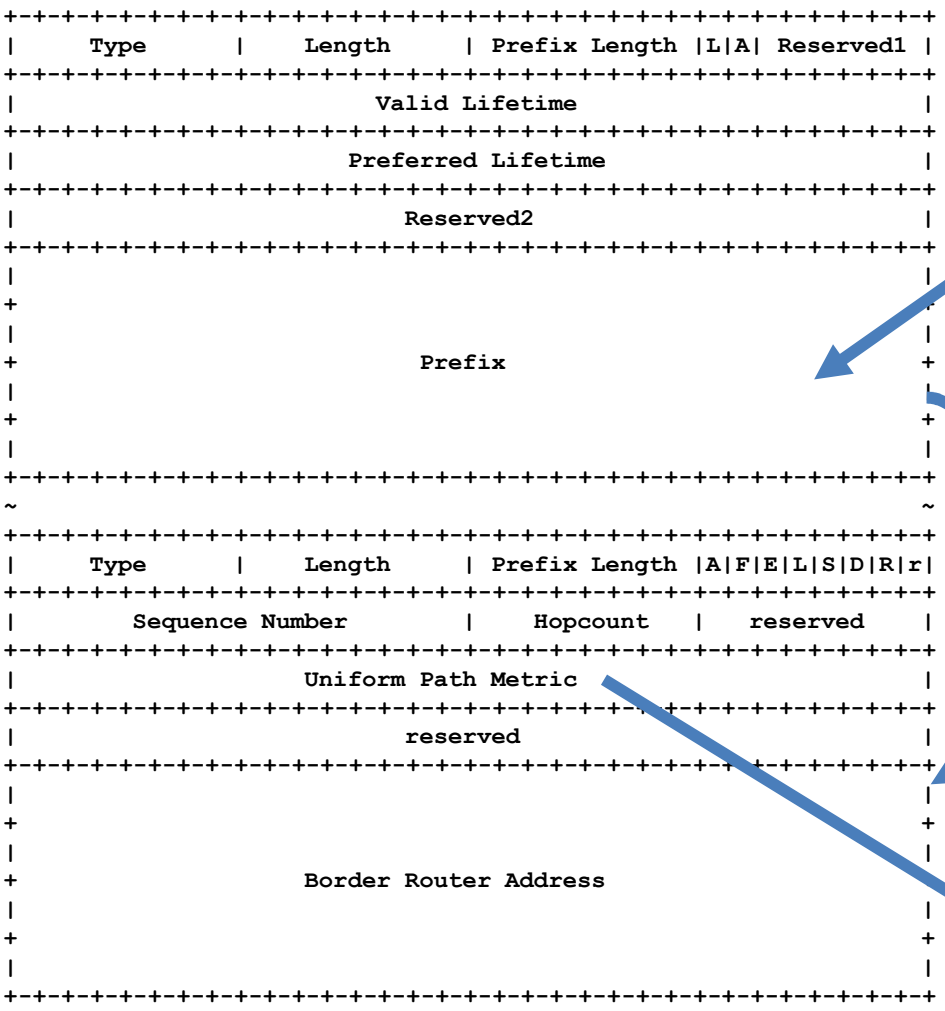
Solution for problem #2

If one wants the ***end host*** to make the upstream ISP selection, how does one provide enough information about the various upstream links (bandwidth, cost, current availability, congestion) to the end hosts so that they can make a good decision ?

- **Prefix Information option in Router Advertisements:**
  - **Advertize only prefixes that have the best BRIO metrics**
  - **Revoked prefixes fade out slowly**

- **Border Router Discovery Protocol (BRDP):**
  - **Hosts may use BRDP information for Source Addresses selection algorithm**
  - **SA selection for new connections and for load balancing over BRs**

Relation between SA, RA Prefix and BRIO



Source Address corresponds to a RA Prefix

RA Prefix corresponds to a BRIO

BRIO metric can be used for SA selection (RFC3484 precedence)

Solution for problem #3

If an end host selected an upstream ISP, how can the ***routing system*** direct traffic to this ISP ?

- **Packets must have an *earmark* for indicating the to-be-used Border Router**
- **Source Address or additional tag (routing header, encapsulation)**
- **BRDP Based Routing:**
  Edge networks, forwarding based on FIB and SA / BRIO cache
- **Border Router Routing Header (work in progress):**
  Ad hoc networks without IGP, forwarding based on
  - **BRDP neighbors, for traffic towards BR**
  - **FIB, maintained by BR Routing Header, for reverse path**

This presentation only discusses BRDP Based Routing

Legacy Internet Routing

RFC1812; Section 5.2.1.2 Unicast:

(5) The forwarder determines the next hop IP address for the packet, **usually** by looking up the **packet's destination** in the router's routing table. This procedure is described in more detail in Section [5.2.4]. This procedure also decides which network interface should be used to send the packet.

RFC1812; 5.2.4.3 Next Hop Address

(5) Default Route: This is a **route to all networks** for which there are **no explicit routes**. It is by definition the route whose prefix length is zero

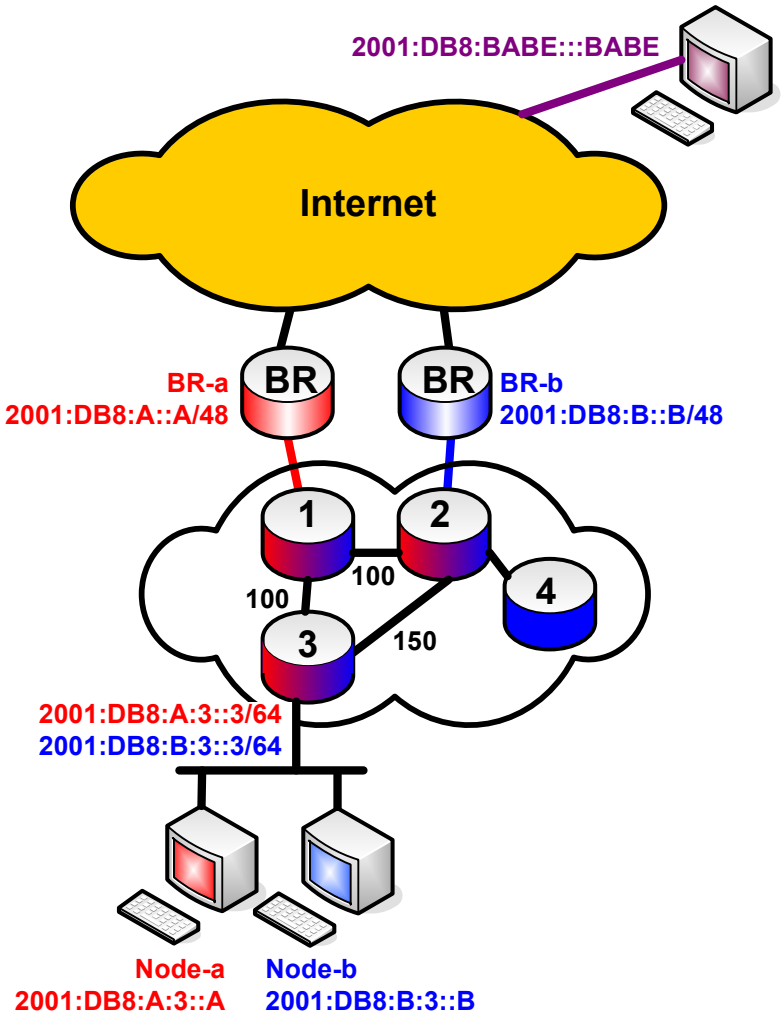
- If a default route exists, legacy routing is used.

BRDP Based Routing

- With BRDP Based Routing, the default route is removed.

- New heuristic for finding a next hop, only used when first FIB lookup didn't find a nexthop:

  (6) BRDP Route: This is a **route to all networks** for which there are **no explicit routes**, and a **default route is not used**. The nexthop IP address is found by means of a Border Router *Information Cache **(BRIO-Cache) lookup** based on the source address and, if a matching BRIO-Cache entry is found, a subsequent **FIB lookup** based on the selected **Border Router address.***

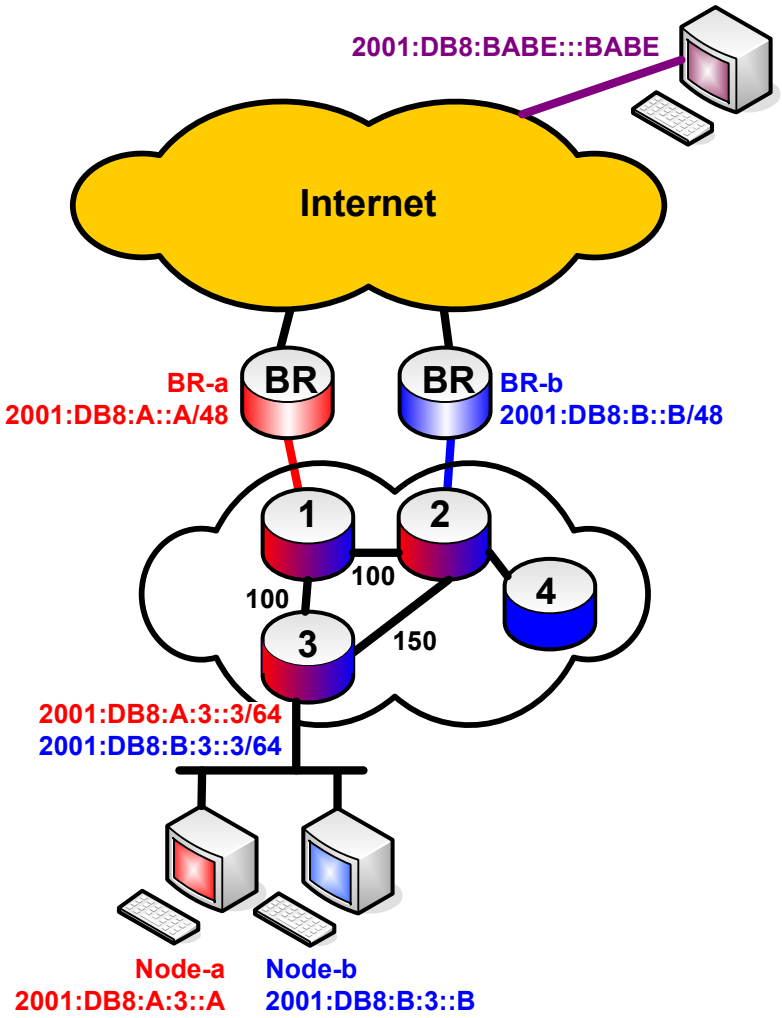


← My example

RFC 5220:

```
                  ==================
                  |    Internet    |
                  ==================
                      |          |
2001:db8:1000::/36    |          | 2001:db8:8000::/36
                  +---+-+      +-+---+
                  | ISP1 |     | ISP2 |
                  +---+-+      +-+---+
                      |          |
2001:db8:1000:::/48   |          | 2001:db8:8000::/48
                 +----+---+ +----+----+
                 | Router1 | | Router2 |
                 +-------+-+ +-+-------+
                         |     |
2001:db8:1000:1::/64     |     | 2001:db8:8000:1::/64
                         |     |
                  -------+-+---+------
                           |
                        +--+---+ 2001:db8:1000:1::100
                        | Host | 2001:db8:8000:1::100
                        +------+
```

Figure 1



Example

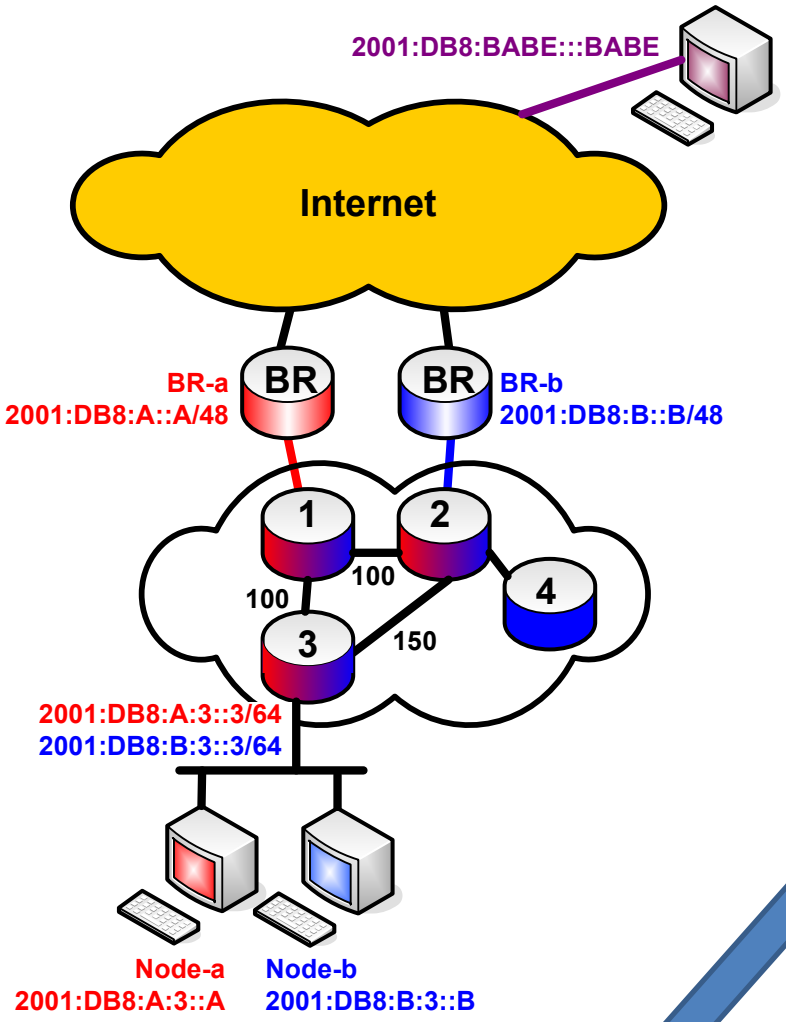
FIB R3:

2001:DB8:A::/64	-> FE80::1	# BR-a
2001:DB8:A:1::/64	-> FE80::1	
2001:DB8:A:2::/64	-> FE80::2	
2001:DB8:A:3::/64	-> local	
2001:DB8:B::/64	-> FE80::2	# BR-b
2001:DB8:B:1::/ 64	-> FE80::1	
2001:DB8:B:2::/ 64	-> FE80::2	
2001:DB8:B:3::/ 64	-> local	
2001:DB8:B:4::/ 64	-> FE80::2	

BRIO Cache R3:

2001:DB8:A::A/48	<= FE80::1 metric 100
2001:DB8:A::A/48	<= FE80::2 metric 250
2001:DB8:B::B/48	<= FE80::2 metric 150
2001:DB8:B::B/48	<= FE80::1 metric 200

Example - FIB



FIB R3:

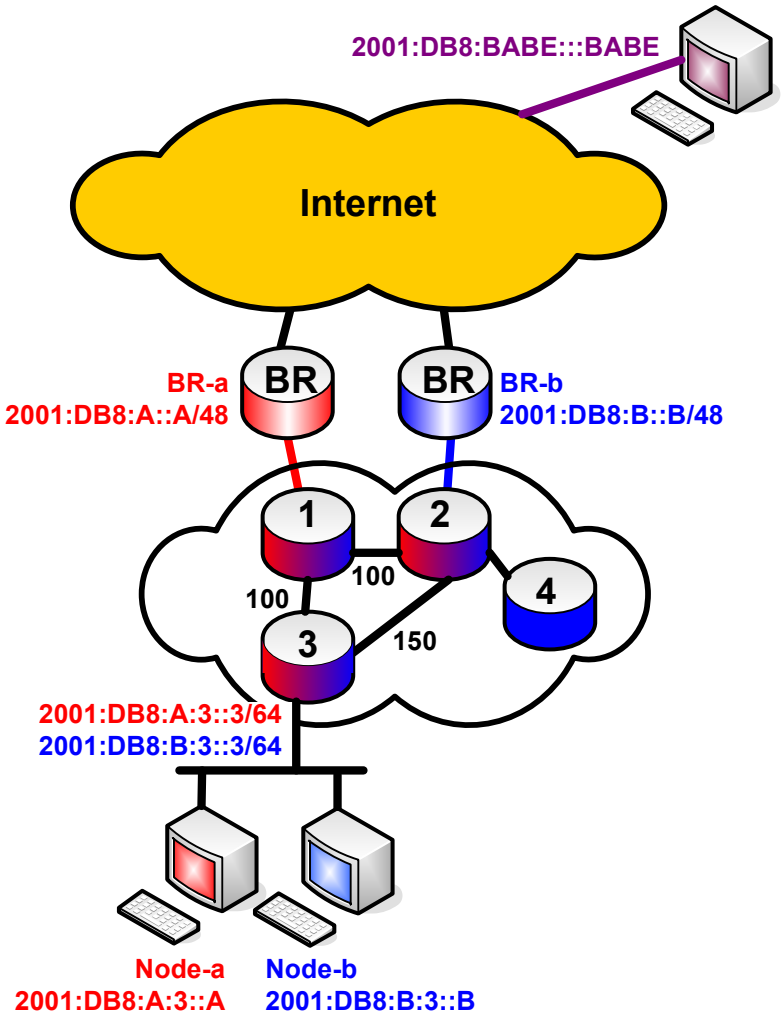
2001:DB8:A::/64	-> FE80::1	# BR-a
2001:DB8:A:1::/64	-> FE80::1	
2001:DB8:A:2::/64	-> FE80::2	
2001:DB8:A:3::/64	-> local	
2001:DB8:B::/64	-> FE80::2	# BR-b
2001:DB8:B:1::/ 64	-> FE80::1	
2001:DB8:B:2::/ 64	-> FE80::2	
2001:DB8:B:3::/ 64	-> local	
2001:DB8:B:4::/ 64	-> FE80::2	

BRIO Cache R3:

2001:DB8:A::A/48	<= FE80::1 metric 100
2001:DB8:A::A/48	<= FE80::2 metric 250
2001:DB8:B::B/48	<= FE80::2 metric 150
2001:DB8:B::B/48	<= FE80::1 metric 200

All internal routes are found in FIB
Forwarding based on DA
No default route !!

Example BRIO-Cache



Border Router information, with sender info and metrics

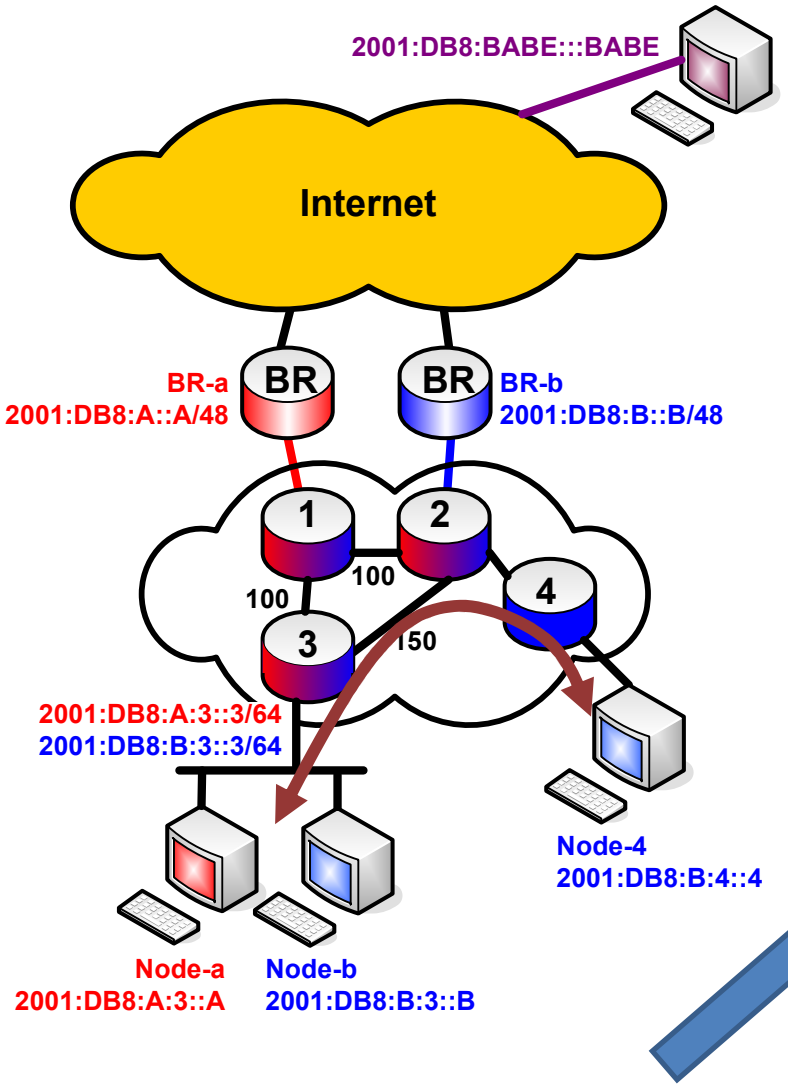
FIB R3:

2001:DB8:A::/64	-> FE80::1	# BR-a
2001:DB8:A:1::/64	-> FE80::1	
2001:DB8:A:2::/64	-> FE80::2	
2001:DB8:A:3::/64	-> local	
2001:DB8:B::/64	-> FE80::2	# BR-b
2001:DB8:B:1::/ 64	-> FE80::1	
2001:DB8:B:2::/ 64	-> FE80::2	
2001:DB8:B:3::/ 64	-> local	
2001:DB8:B:4::/ 64	-> FE80::2	

BRIO Cache R3:

2001:DB8:A::A/48	<= FE80::1 metric 100
2001:DB8:A::A/48	<= FE80::2 metric 250
2001:DB8:B::B/48	<= FE80::2 metric 150
2001:DB8:B::B/48	<= FE80::1 metric 200

Node-a or Node-b sends to Node-4



FIB R3:

2001:DB8:A::/64	-> FE80::1	# BR-a
2001:DB8:A:1::/64	-> FE80::1	
2001:DB8:A:2::/64	-> FE80::2	
2001:DB8:A:3::/64	-> local	
2001:DB8:B::/64	-> FE80::2	# BR-b
2001:DB8:B:1::/ 64	-> FE80::1	
2001:DB8:B:2::/ 64	-> FE80::2	
2001:DB8:B:3::/ 64	-> local	
2001:DB8:B:4::/ 64	-> FE80::2	

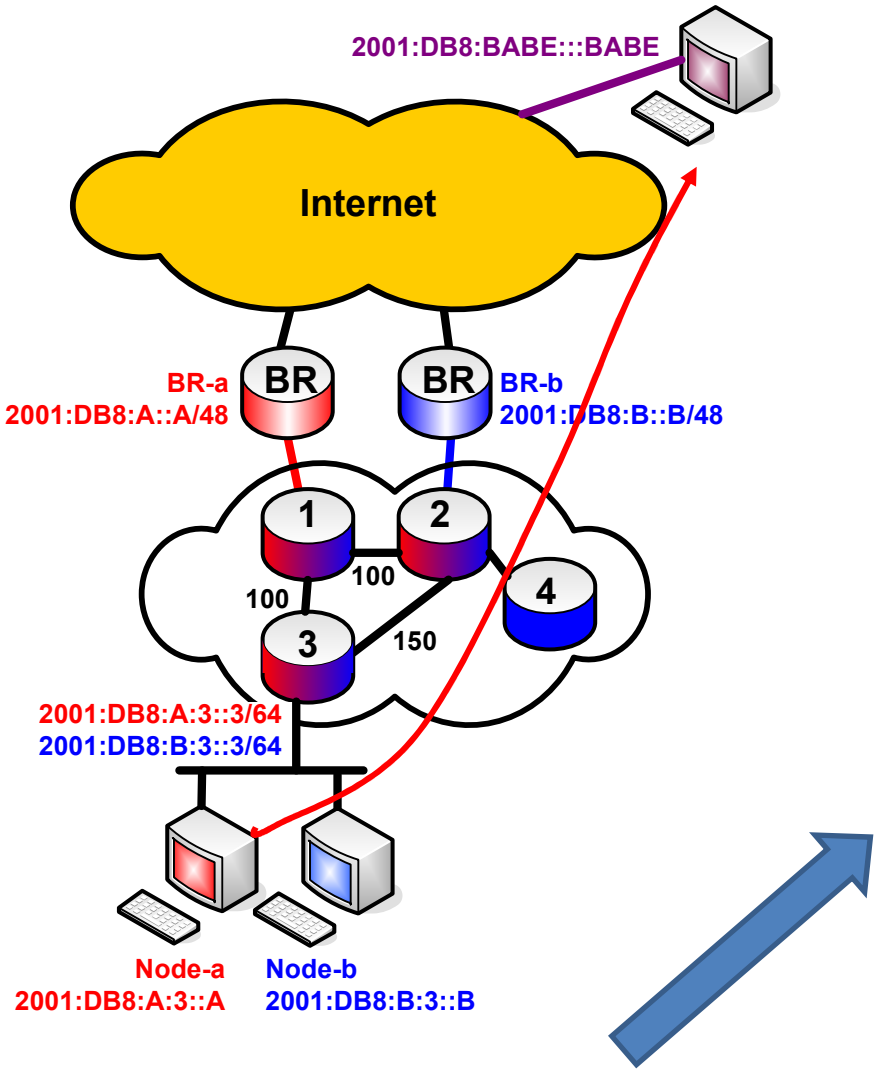
BRIO Cache R3:

2001:DB8:A::A/48	<= FE80::1 metric 100
2001:DB8:A::A/48	<= FE80::2 metric 250
2001:DB8:B::B/48	<= FE80::2 metric 150
2001:DB8:B::B/48	<= FE80::1 metric 200

1) 2001:DB8:B:4::4 in FIB?

Yes

Node-a sends to Babe



FIB R3:

2001:DB8:A::/64	-> FE80::1	# BR-a
2001:DB8:A:1::/64	-> FE80::1	
2001:DB8:A:2::/64	-> FE80::2	
2001:DB8:A:3::/64	-> local	
2001:DB8:B::/64	-> FE80::2	# BR-b
2001:DB8:B:1::/ 64	-> FE80::1	
2001:DB8:B:2::/ 64	-> FE80::2	
2001:DB8:B:3::/ 64	-> local	
2001:DB8:B:4::/ 64	-> FE80::2	

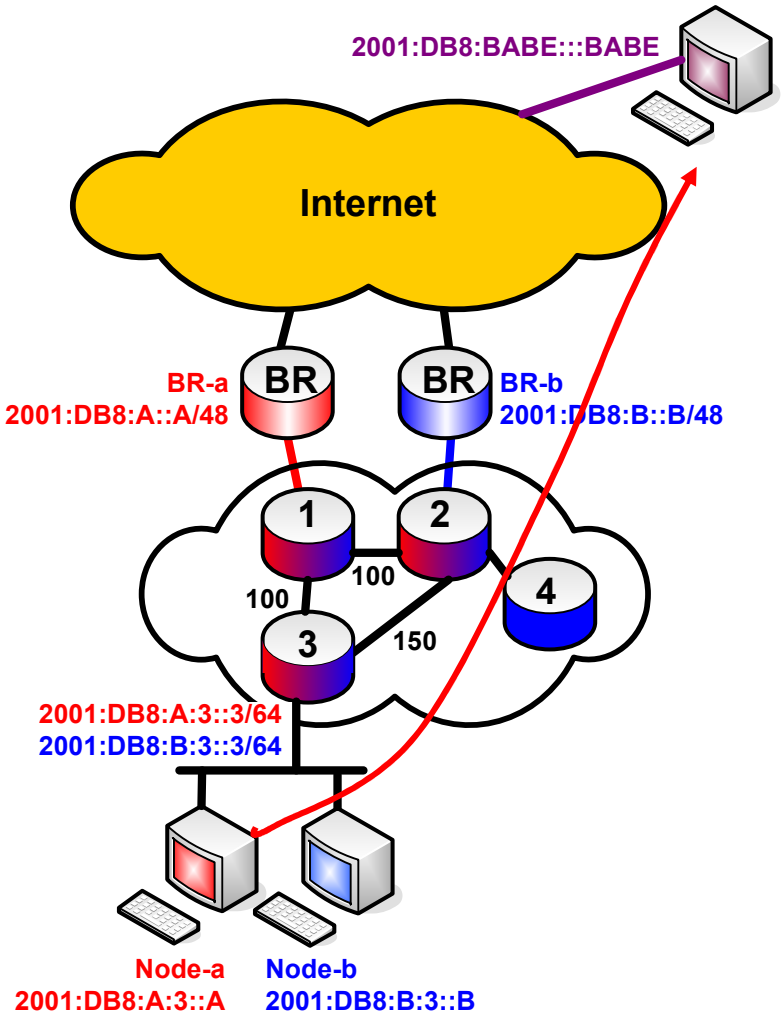
BRIO Cache R3:

2001:DB8:A::A/48	<= FE80::1 metric 100
2001:DB8:A::A/48	<= FE80::2 metric 250
2001:DB8:B::B/48	<= FE80::2 metric 150
2001:DB8:B::B/48	<= FE80::1 metric 200

1) 2001:DB8:BABE::BABE in FIB?

NO

Node-a sends to Babe



FIB R3:

2001:DB8:A::/64	-> FE80::1	# BR-a
2001:DB8:A:1::/64	-> FE80::1	
2001:DB8:A:2::/64	-> FE80::2	
2001:DB8:A:3::/64	-> local	
2001:DB8:B::/64	-> FE80::2	# BR-b
2001:DB8:B:1::/ 64	-> FE80::1	
2001:DB8:B:2::/ 64	-> FE80::2	
2001:DB8:B:3::/ 64	-> local	
2001:DB8:B:4::/ 64	-> FE80::2	

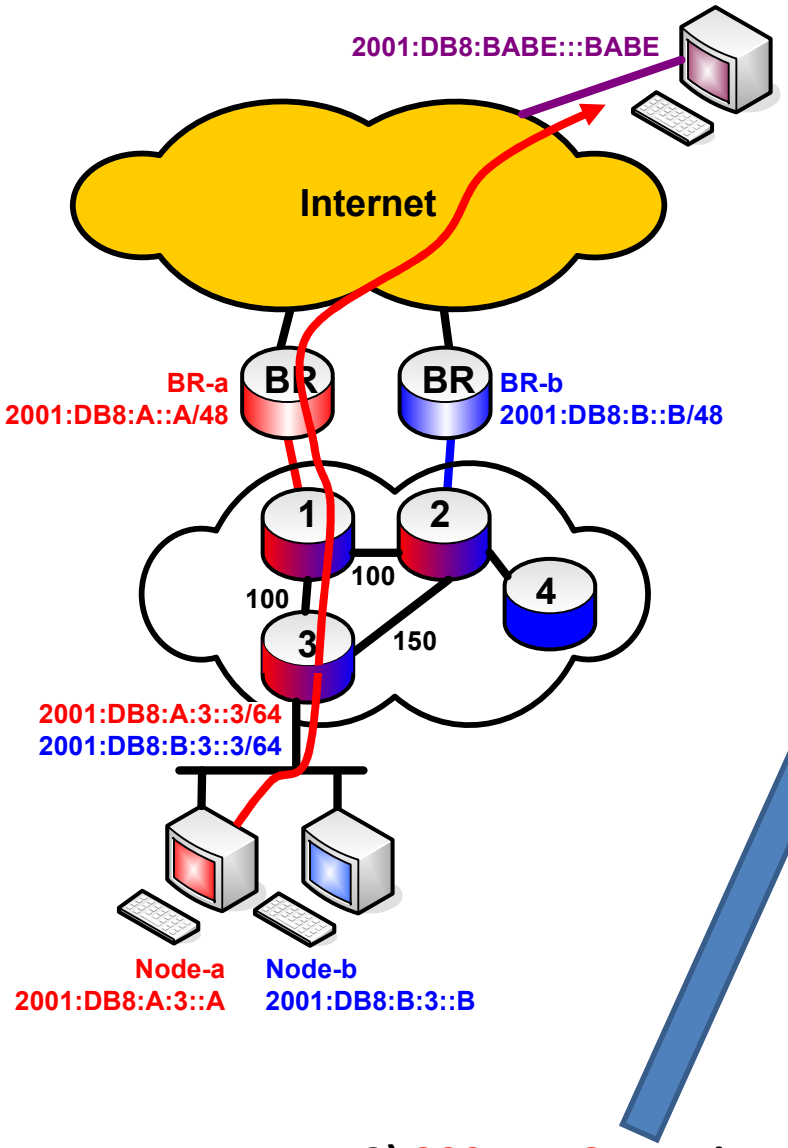
BRIO Cache R3:

2001:DB8:A::A/48	<= FE80::1 metric 100
2001:DB8:A::A/48	<= FE80::2 metric 250
2001:DB8:B::B/48	<= FE80::2 metric 150
2001:DB8:B::B/48	<= FE80::1 metric 200

2) 2001:DB8:A:3::A match in BRIO Cache?

YES: 2001:DB8:A::A

Node-a sends to Babe



FIB R3:

2001:DB8:A::/64	-> FE80::1	# BR-a
2001:DB8:A:1::/64	-> FE80::1	
2001:DB8:A:2::/64	-> FE80::2	
2001:DB8:A:3::/64	-> local	
2001:DB8:B::/64	-> FE80::2	# BR-b
2001:DB8:B:1::/ 64	-> FE80::1	
2001:DB8:B:2::/ 64	-> FE80::2	
2001:DB8:B:3::/ 64	-> local	
2001:DB8:B:4::/ 64	-> FE80::2	

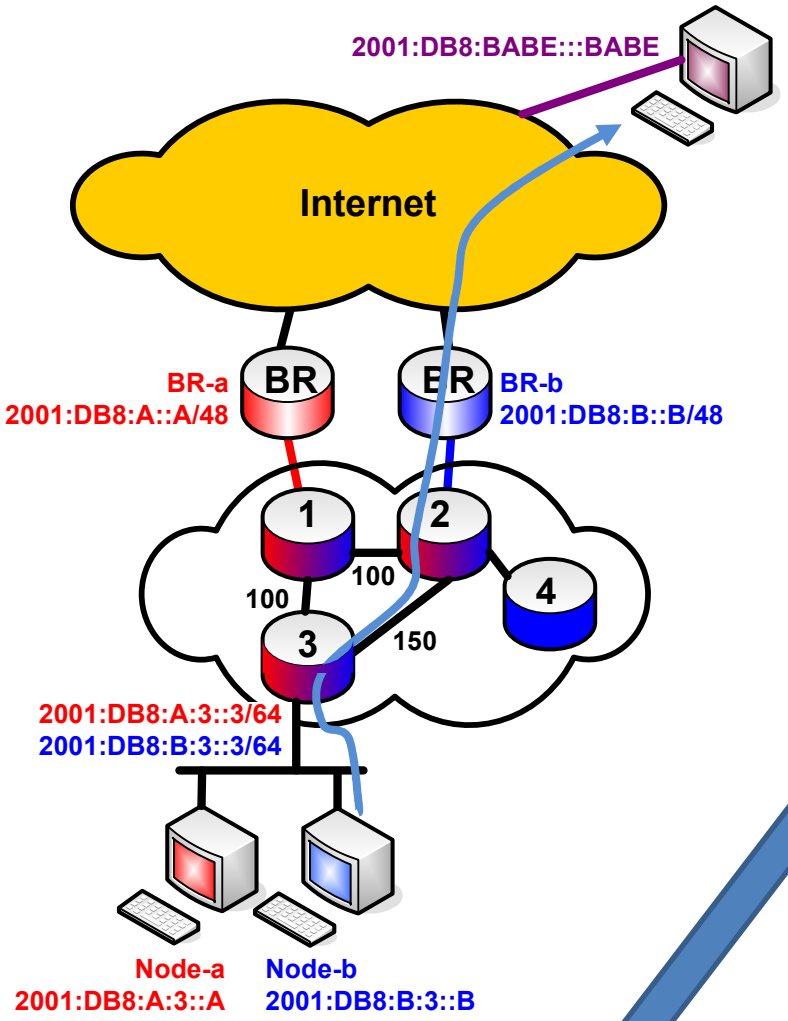
BRIO Cache R3:

2001:DB8:A::A/48	<= FE80::1 metric 100
2001:DB8:A::A/48	<= FE80::2 metric 250
2001:DB8:B::B/48	<= FE80::2 metric 150
2001:DB8:B::B/48	<= FE80::1 metric 200

3) 2001:DB8:A::A in FIB?

YES: 2001:DB8:A::/64 -> FE80::1

Node-b sends to Babe



FIB R3:

2001:DB8:A::/64	-> FE80::1	# BR-a
2001:DB8:A:1::/64	-> FE80::1	
2001:DB8:A:2::/64	-> FE80::2	
2001:DB8:A:3::/64	-> local	
2001:DB8:B::/64	-> FE80::2	# BR-b
2001:DB8:B:1::/ 64	-> FE80::1	
2001:DB8:B:2::/ 64	-> FE80::2	
2001:DB8:B:3::/ 64	-> local	
2001:DB8:B:4::/ 64	-> FE80::2	

BRIO Cache R3:

2001:DB8:A::A/48	<= FE80::1 metric 100
2001:DB8:A::A/48	<= FE80::2 metric 250
2001:DB8:B::B/48	<= FE80::2 metric 150
2001:DB8:B::B/48	<= FE80::1 metric 200

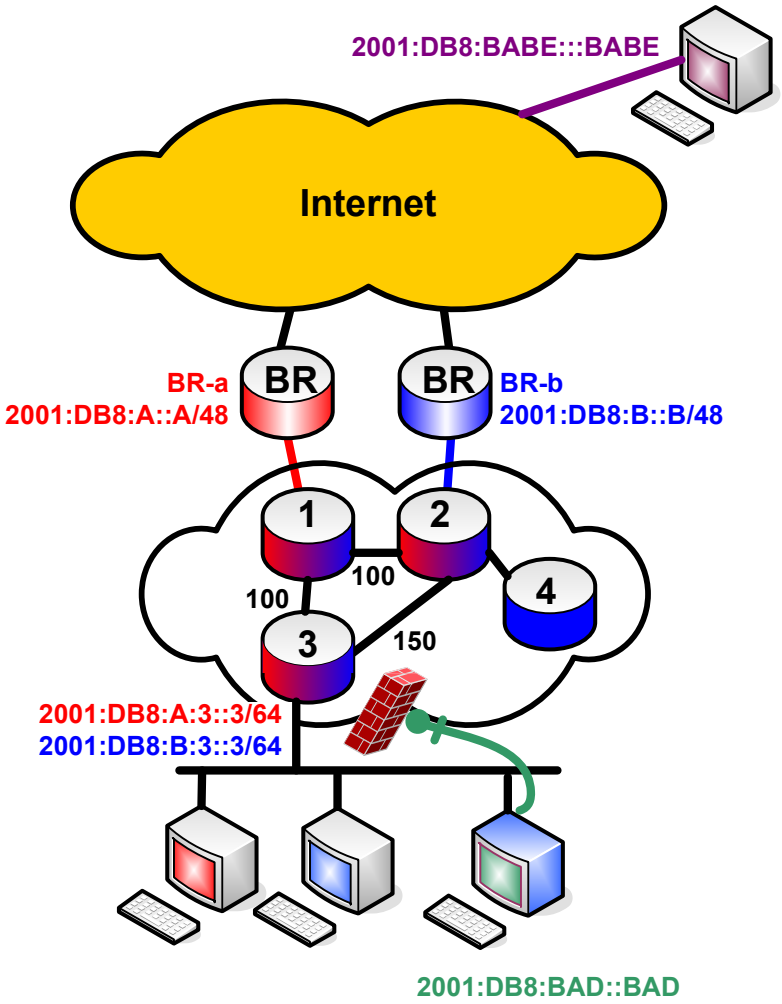
1: 2001:DB8:BABE::BABE in FIB: NO

2: 2001:DB8:B:3::B match in BRIO Cache :

Yes, 2001:DB8:B::B

3: 2001:DB8:B::B in FIB: Yes, FE80::2

Bad sends to Babe



FIB R3:

2001:DB8:A::/64	-> FE80::1	# BR-a
2001:DB8:A:1::/64	-> FE80::1	
2001:DB8:A:2::/64	-> FE80::2	
2001:DB8:A:3::/64	-> local	
2001:DB8:B::/64	-> FE80::2	# BR-b
2001:DB8:B:1::/ 64	-> FE80::1	
2001:DB8:B:2::/ 64	-> FE80::2	
2001:DB8:B:3::/ 64	-> local	
2001:DB8:B:4::/ 64	-> FE80::2	

BRIO Cache R3:

2001:DB8:A::A/48	<= FE80::1 metric 100
2001:DB8:A::A/48	<= FE80::2 metric 250
2001:DB8:B::B/48	<= FE80::2 metric 150
2001:DB8:B::B/48	<= FE80::1 metric 200

1: 2001:DB8:BABE::BABE in FIB: NO

2: 2001:DB8:BAD::BAD match in BRIO Cache : NO

3: DROP

BRDP Based Routing

Introduction

Analysis of the problem

How does it work?

Next steps

Next steps

- Support for IPv4 (not that difficult, but IPv4 is not as flexible as IPv6)
- Finish BRDP implementation (Linux, Opnet)
- Implement BRDP Based Routing (Linux, Opnet)
- Post document for BRDP based Source Address Selection
- Continue research on Border Router Routing Header

- Website:
  http://www.inf-net.nl/brdp.html

- Any help is welcome !
- Any comment is welcome !

Thanks for your attention !