BGP: OPTIMISED ROUTE REFLECTION AND NEXT-HOP SAFI

ILYA VARLASHKIN, EASYNET

ROBERT RASZUK, NTT I3

Current RR behaviour

- Same path(es) for all clients
  - ADD-PATH is not always solution
- Expensive in some scenarios
  - Many small PoP's with many exits need many RR's
- Can't deal with some common failures
- Meaning of Next-Hop Cost is hardcoded
  - IGP cost != business policy

Extension #1:

http://tools.ietf.org/html/draft-ietf-**idr**-bgp-optimal-route-reflection-05

OPTIMAL ROUTE REFLECTION

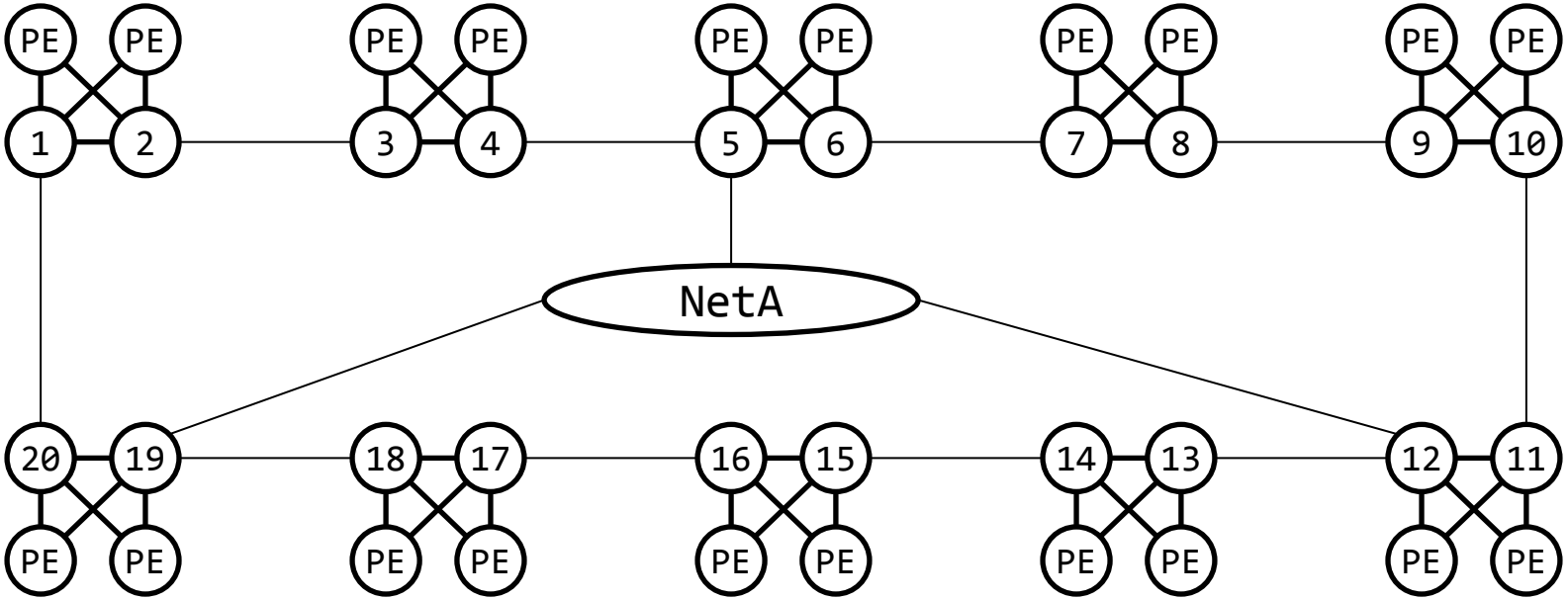
Optimal Route Reflection

- Allows client-specific best path
- Allows arbitrary interpretation of “best” and means to find that out
  - Look in IGP database
  - Fixed administrative preferences
  - *Add your own*
- No protocol modification, only local matter

Use case:

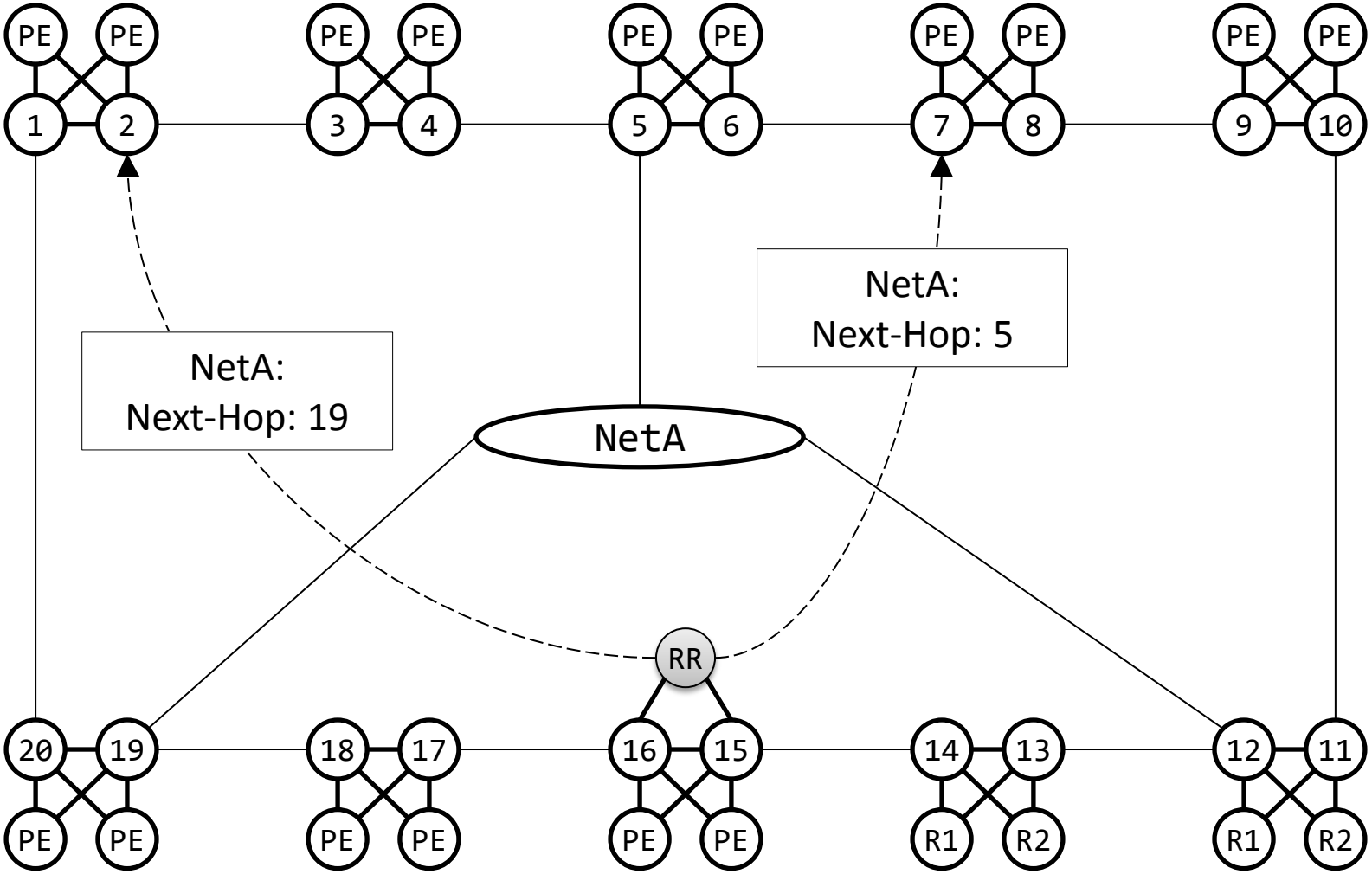
REPLACING FULL-MESH

Where to place RR?



- N small PoP's; N is 10..100
- Distance between PoP's is 10..30ms
- K connections to NetA; K is 3..5
- Either many RR's ($$$) or sub-optimum path

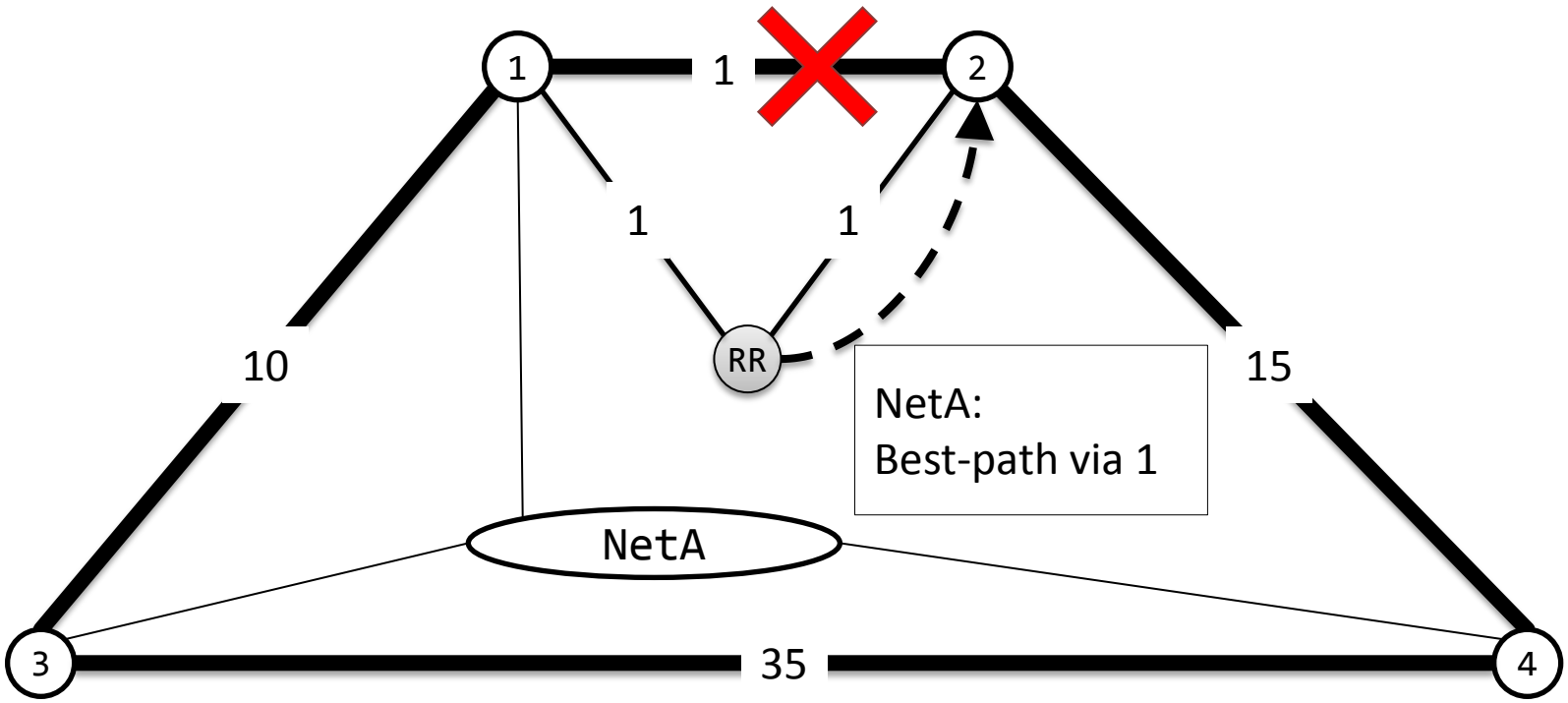
ORR-enabled RR can be anywhere



Use case:

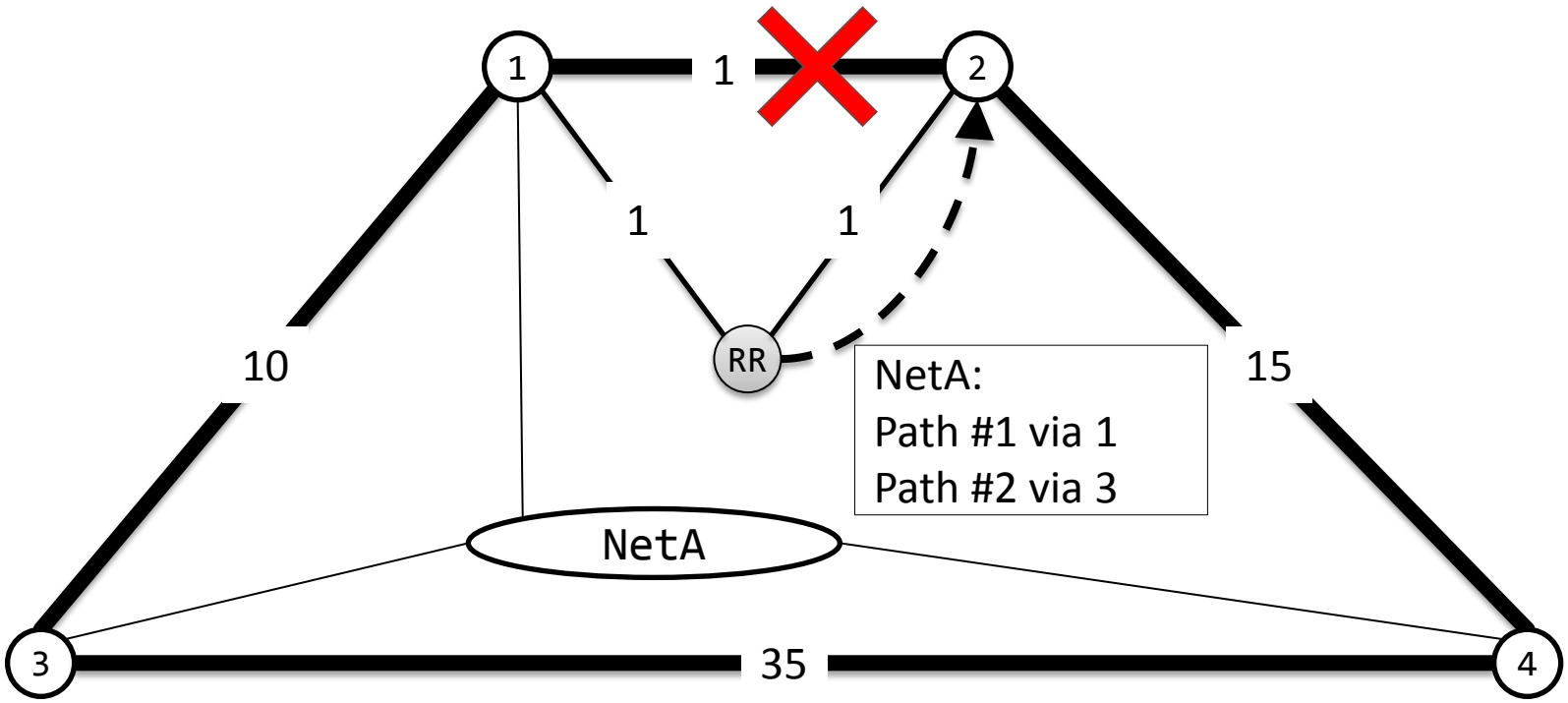
Fixing off-path Route-Reflector

Off-path Route-Reflector



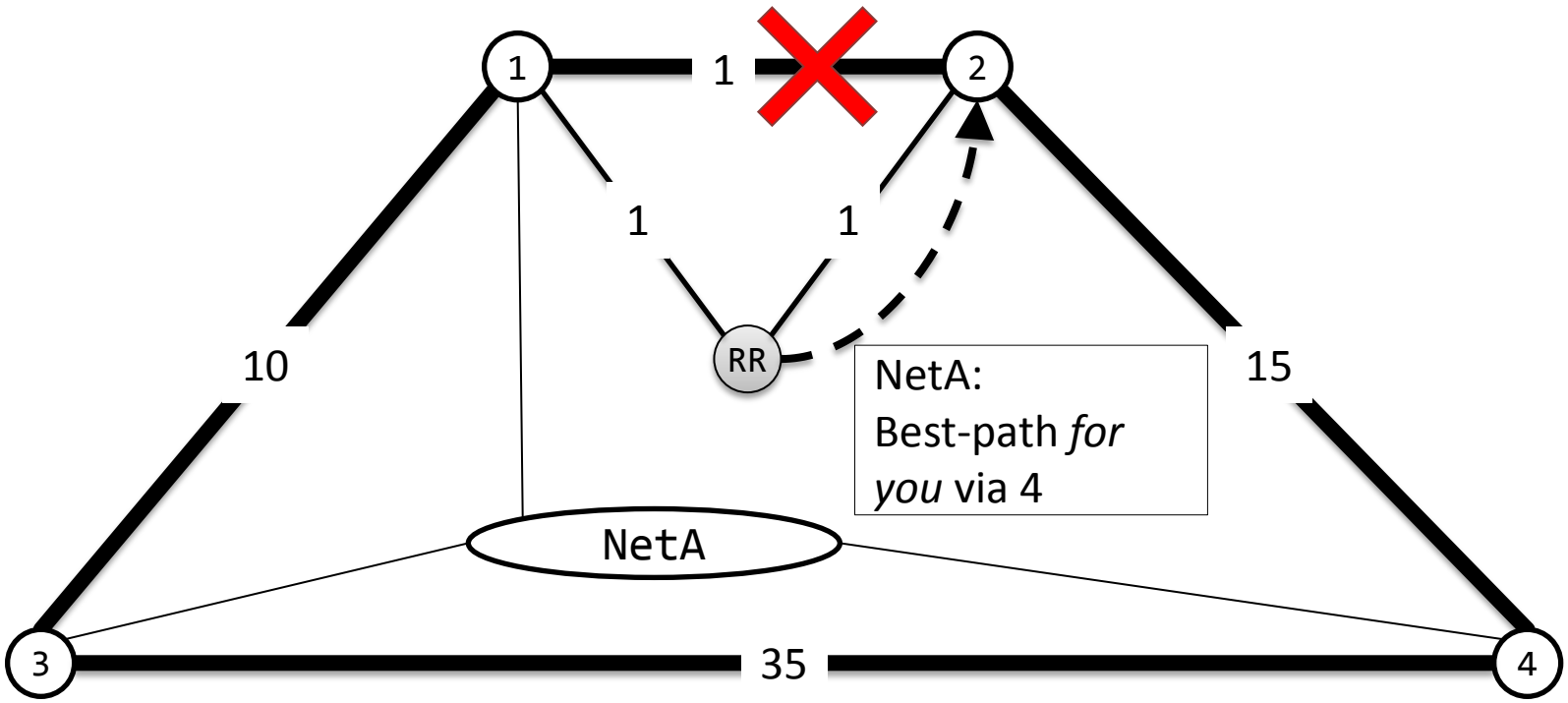
- RR won't care about link [1]-[2] failure
- [2] will go on sight-seeing tour to reach NetA

Off-path Route-Reflector



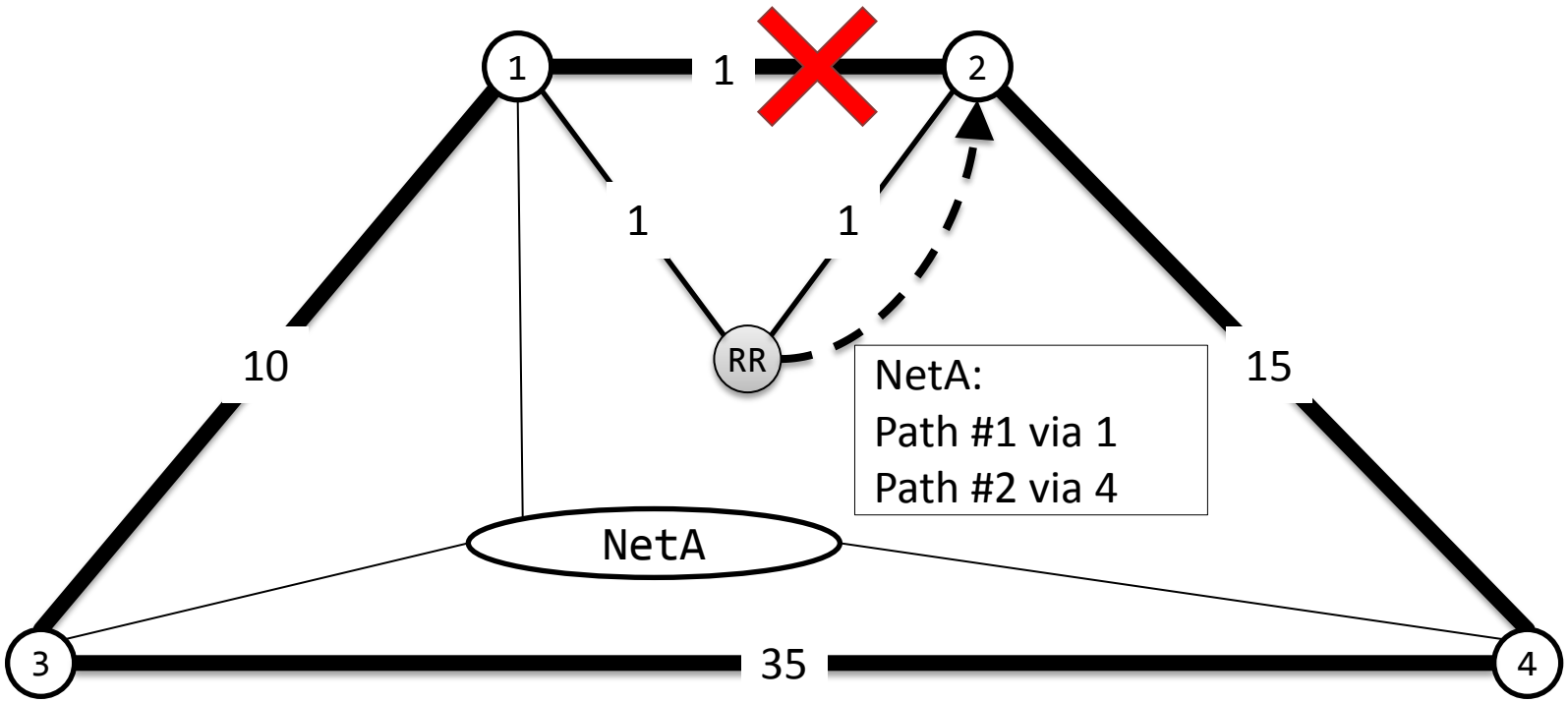
- RR won't care about link [1]-[2] failure
- [2] will go on sight-seeing tour to reach NetA
- ADD-PATH won't be of much help

Off-path Route-Reflector with ORR



- ORR can see link failure and send better path

Off-path Route-Reflector with ORR

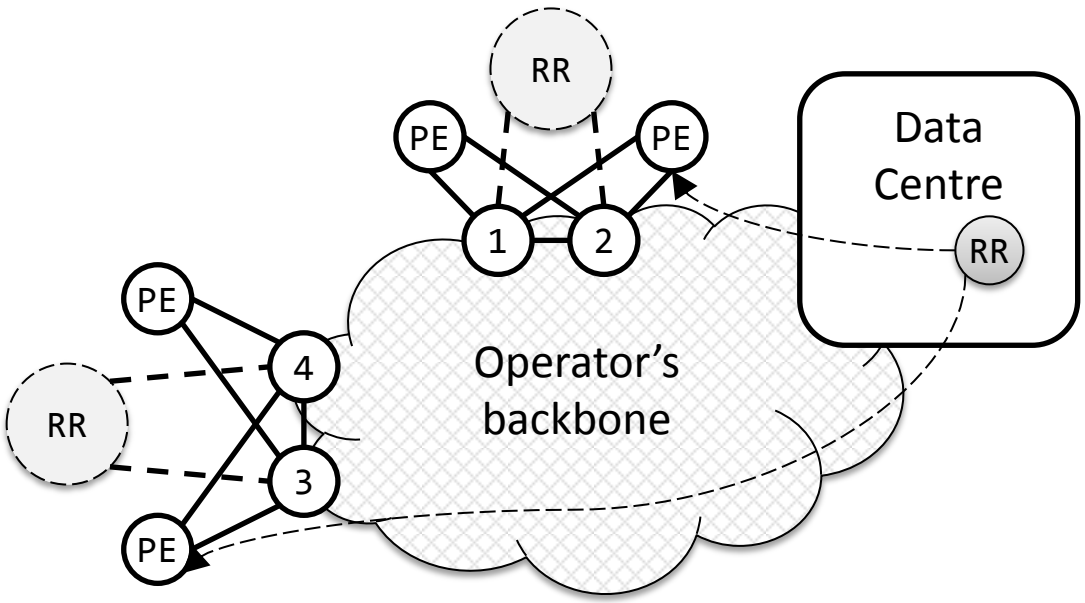


- ORR can see link failure and send better path
- Even better with ADD-PATH

Use case:

Virtual RR placement

Virtual RR placement




```
isis {
  interface list { loopback 1, loopback 2}
}

bgp {
 group iBGP-template {
  remote-as 123
  policy PERMIT_ALL out
  policy PERMIT_ALL in
 }
 group PoP-1 {
  virtual-link R1 cost 1
  virtual-link R2 cost 1
  source-interface loopback 1
  inherit iBGP-template
 }
 group PoP-2 {
  virtual-link R3 cost 1
  virtual-link R4 cost 1
  source-interface loopback 2
  inherit iBGP-template
 }
 neighbor PE1 inherit PoP-1
 neighbor PE2 inherit PoP-2
}
```

- RR physically located on a VM in DC
- Behaves like it's connected at arbitrary PoP

Extension #2:

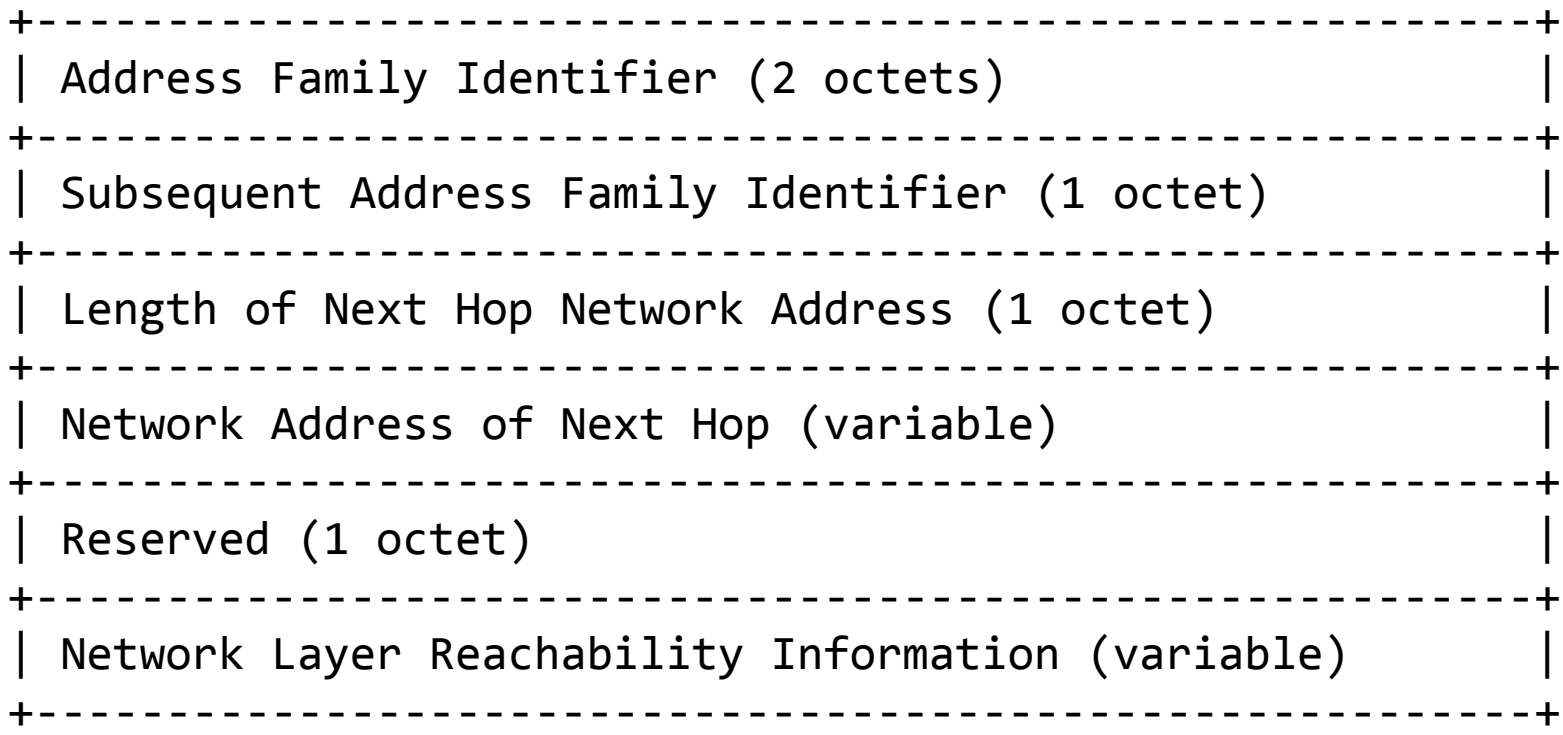
http://tools.ietf.org/html/draft-varlashkin-bgp-nh-cost-02

Next-Hop SAFI

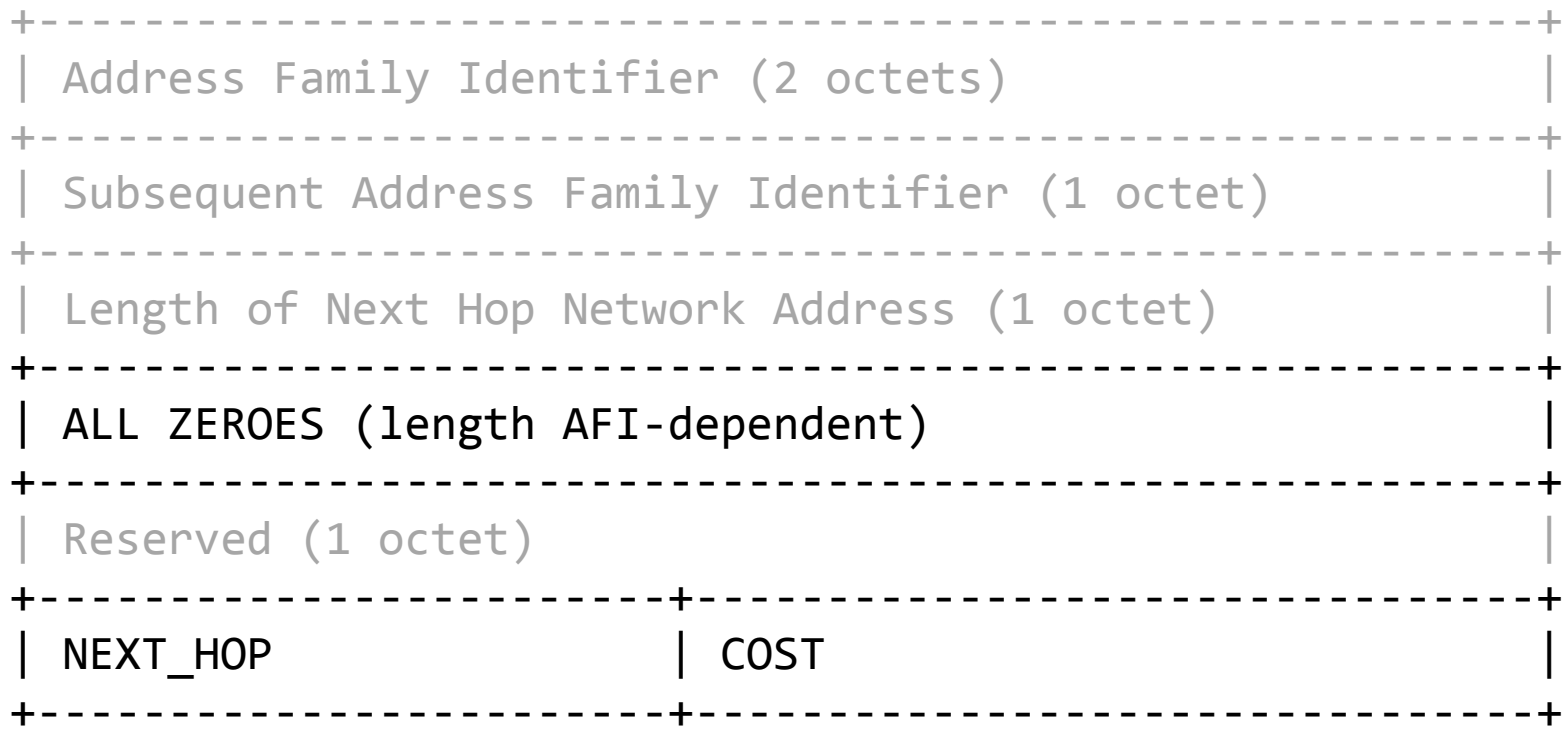
Next-Hop SAFI

- Extends BGP ORR
- Helps where IGP info won't
  - Policy-based exit/Next-Hop
  - Route-servers
- Exchange NH reachability info using BGP
- New SAFI with its own NLRI
  - Relies on ATTRIBUTE 14 (MP_REACH_NLRI)
  - Does not use ATTRIBUTE 15

BGP ATTRIBUTE 14 (RFC4760)



Encoding NH SAFI NLRI

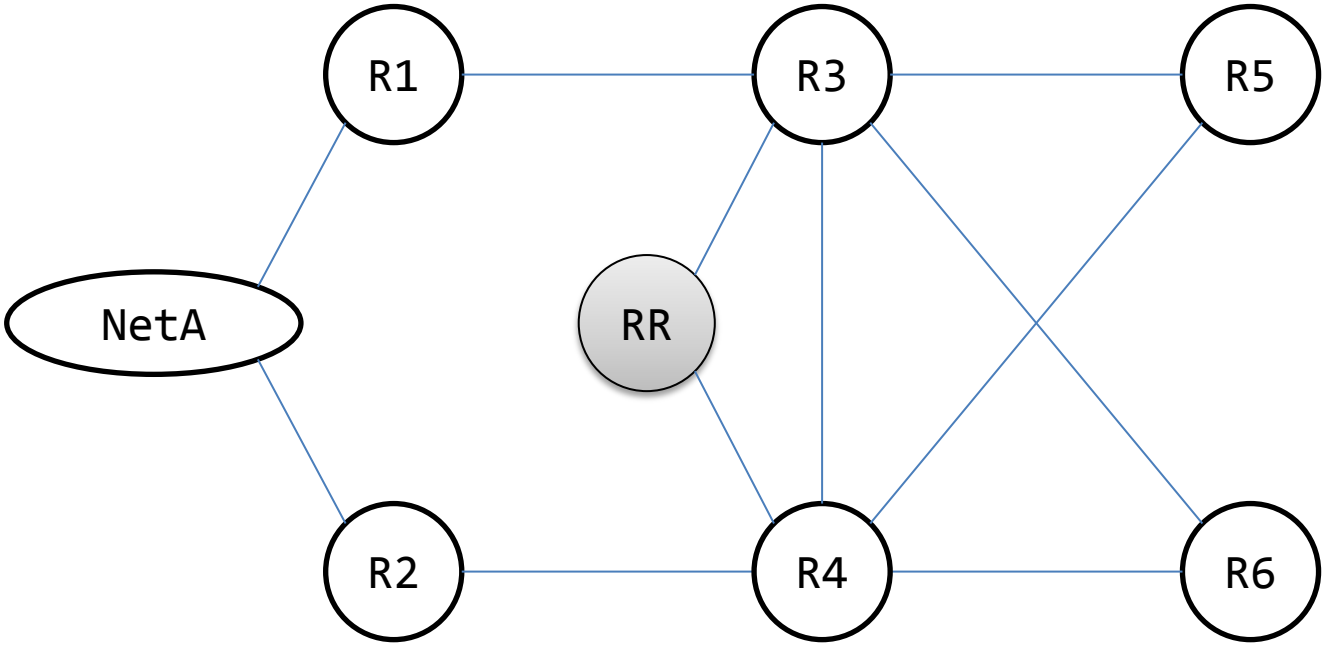


- My cost to NEXT_HOP is COST (uint32_t)
- COST=0xFFFFFFFF – “NEXT_HOP is unreachable”

Use case:

POLICY-BASED EXIT SELECTION

Policy-based exit – the setup



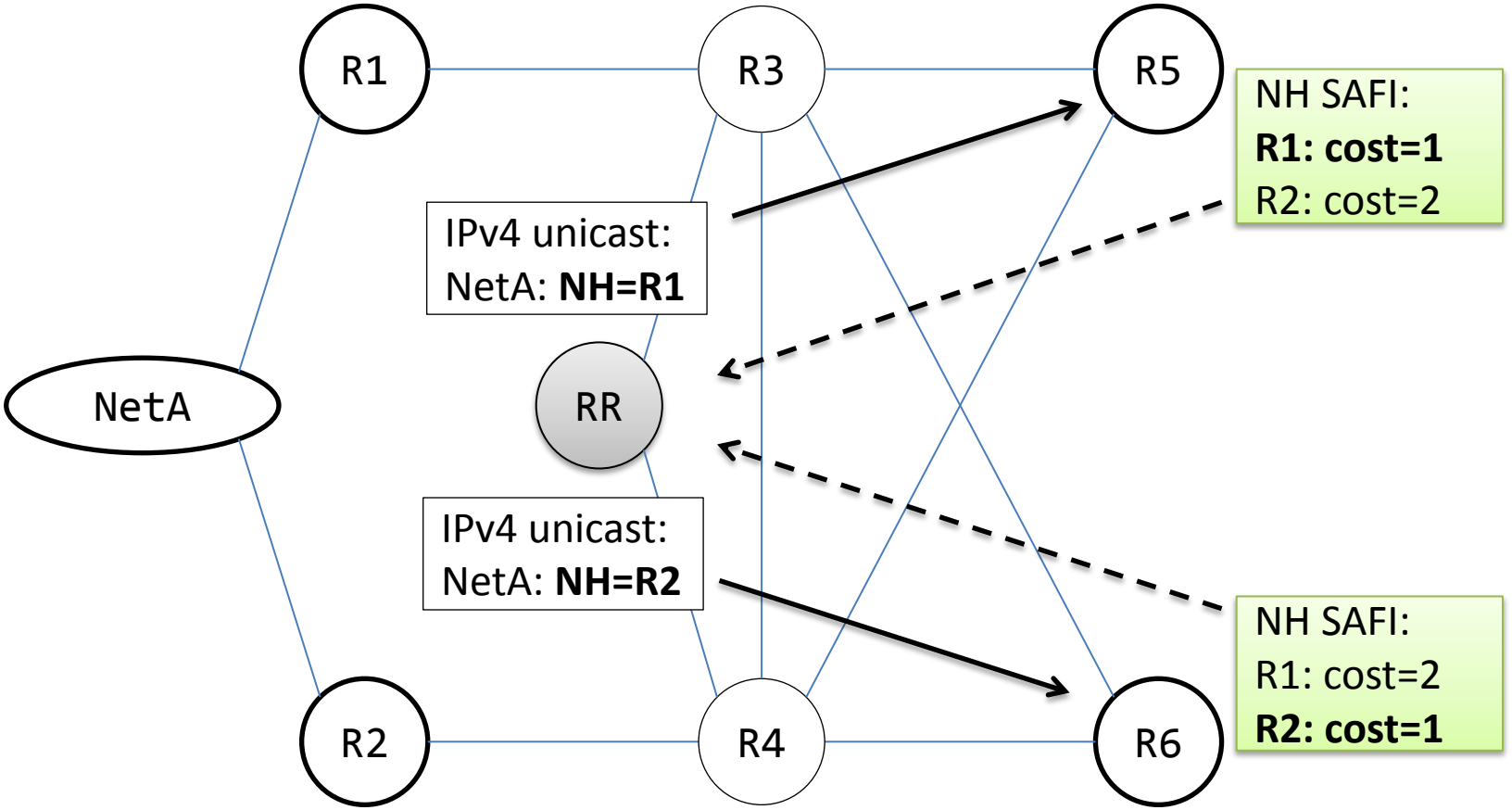
Layer 9 requirements:

- R5 to NetA via R1
- R6 to NetA via R2

Setup:

- RR is a route-reflector
- MPLS end-to-end

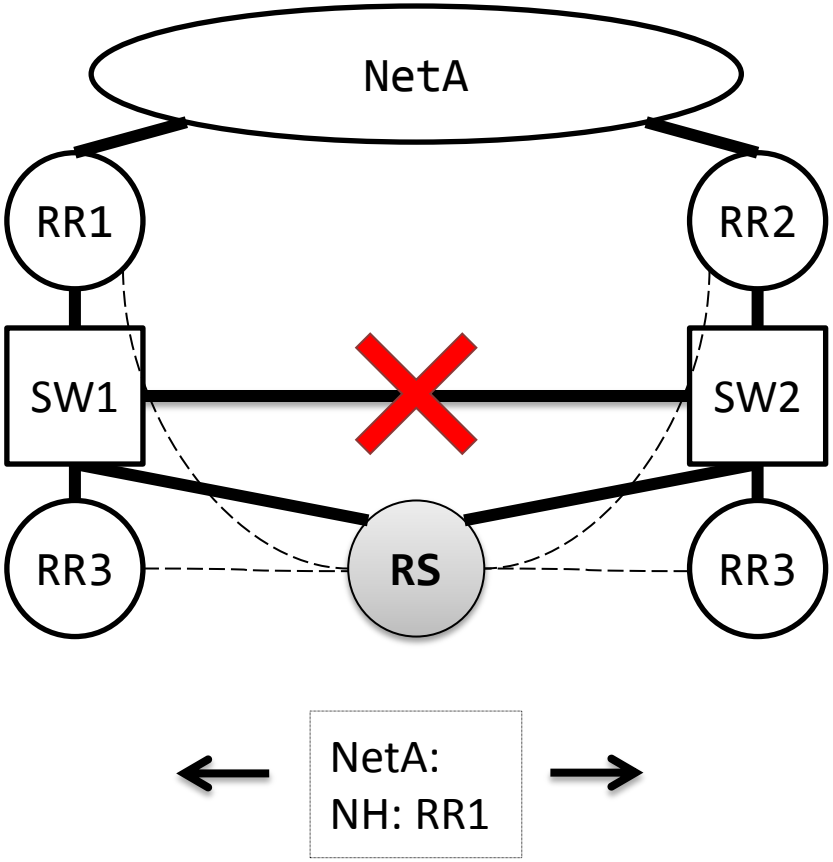
Policy-based exit: ORR + NH SAFI



Use case:

ROUTE-SERVERS AND NH SAFI

Route-servers today



Same best path for everyone

+

No way to convey NH reachability

+

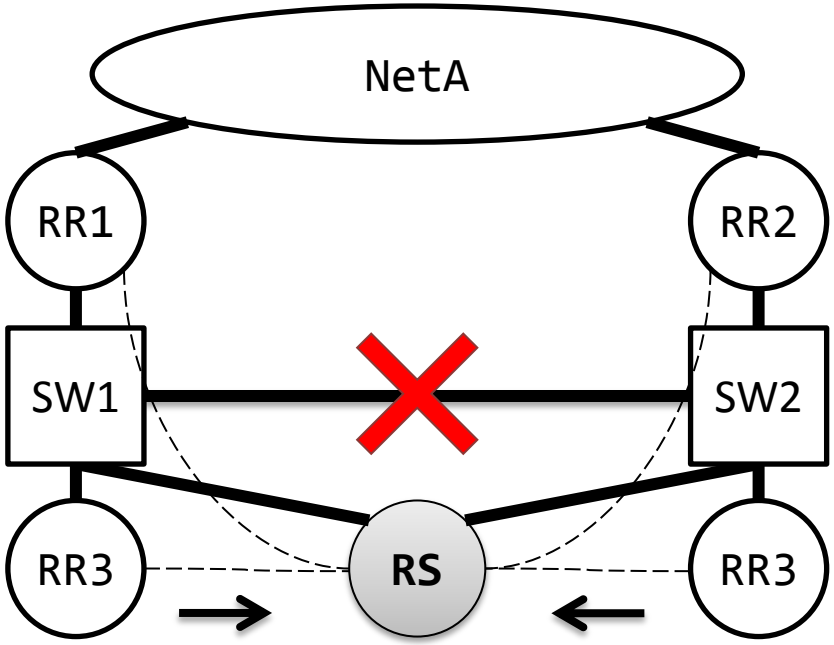
L2 problems

=

Somebody will drop traffic

N.B.: Simplicity of L2 is intentional

Route-servers with NH SAFI



NH SAFI conveys NH reachability

Per-client best route

<u>Unfeasible path avoided</u>

NH SAFI:
RR1=1
RR2=0xFFFFFFFF

NH SAFI:
RR1=0xFFFFFFFF
RR2=1

NetA:
NH: RR1

NetA:
NH: RR2

N.B.: Simplicity of L2 is intentional

THE LAST SLIDE

- Not yet available but there's interest
- If you like it, ask your vendor
- Comments, suggestions, questions?

THANK YOU!!!