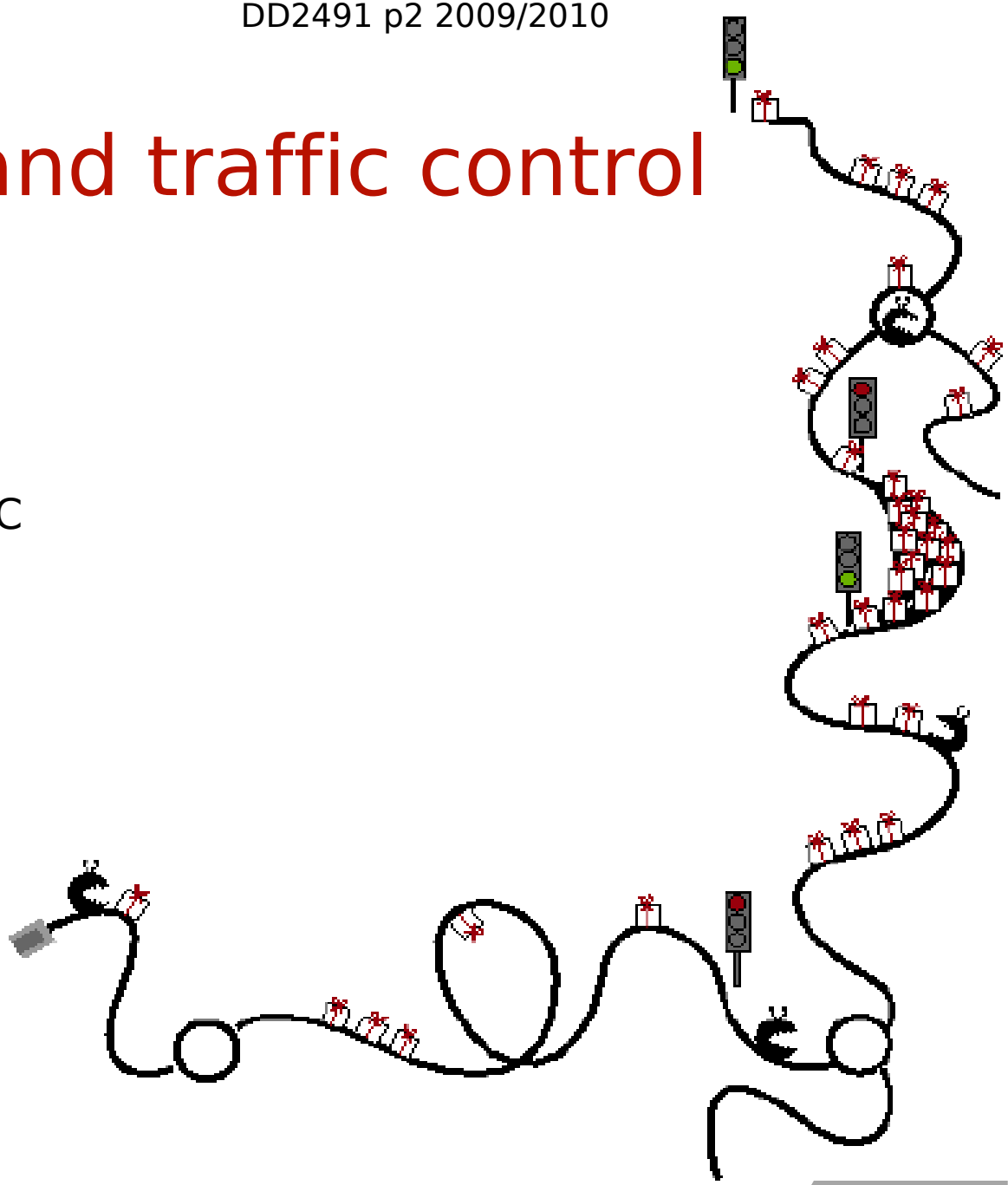
Load balancing and traffic control in BGP

Olof Hagsand KTH /CSC



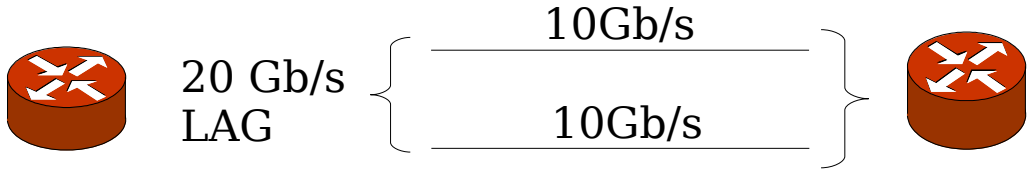
Issues in load balancing

- Load balancing: spread traffic on several paths instead of a single.
- Why?
- Use resources better
  - Can postpone upgrade of infra-structure
- Geography
  - Example: Dont use trans-oceanic links twice
- Cost reasons
  - Balance traffic for optimal price

Redundancy

- Redundancy and load balancing are not always aligned
- Example: if you use load balancing over several links, and one link goes down: can you still forward all traffic on the single link?
- For redundancy it is may sometimes be better to send/announce all traffic on a single link and then have redundant links for backup
- At the edge, it may be easier to get symmetrical routing which is better for filtering

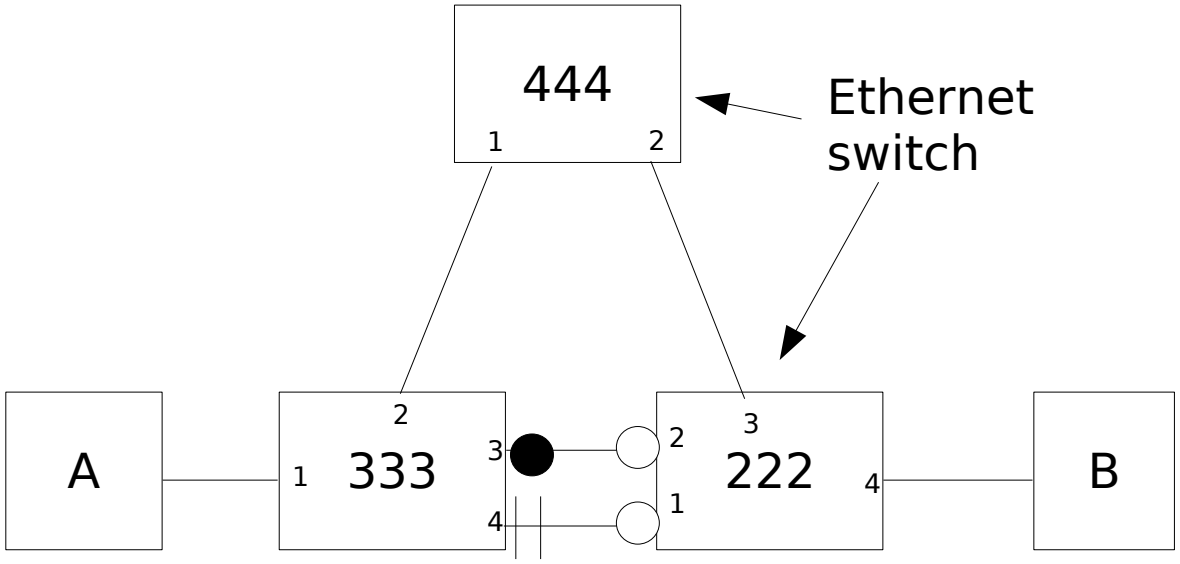
Sub-link-layer load-balancing



- Two physical links is aggregated into a single Link-Aggregate Group.
- Single (20Mb/s) link
- Load balancing normally on flows

Link-layer load-balancing

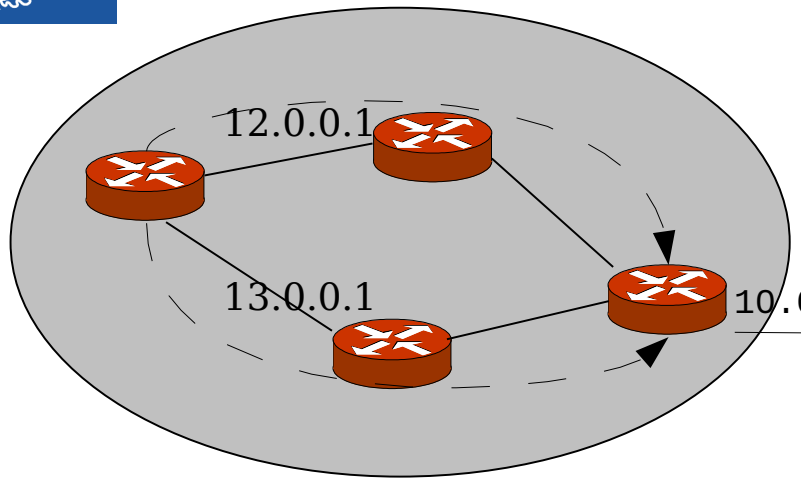
- In (metro) Ethernet, load balancing can not be made: Spanning tree computes a single link
- Example: only one link between 222 and 333 can be used for forwarding.



IGP load balancing in IBGP

- IS-IS / OSPF equal-cost multipath provides load balancing in IBGP

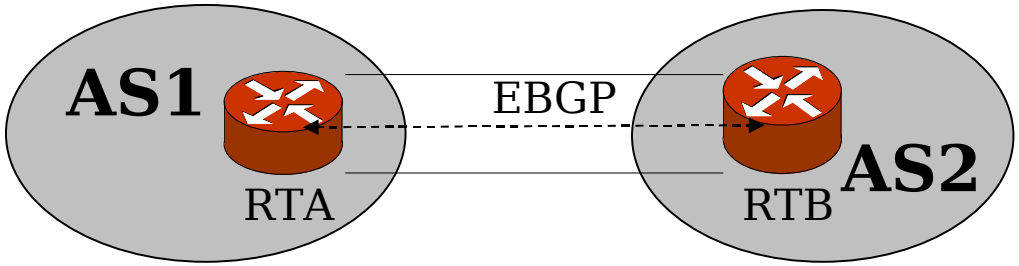
  Since next-hop self uses peering between loop-backs, and the IBGP neighbor may be reachable via more than one nexthop



Route	*Nexthop*	*Metric*	*Protocol*	
130.2.3.0/24	10.0.0.1		IBGP	
10.0.0.1/32	12.0.0.1	10	IGP	Equal-cost multipath
10.0.0.1/32	13.0.0.1	10	IGP	

Loopback peering in EBGP

- Loopback peering can also be used in EBGP, but routing of loopback is then set-up using static routes
  - Uncommon to use IGP between AS

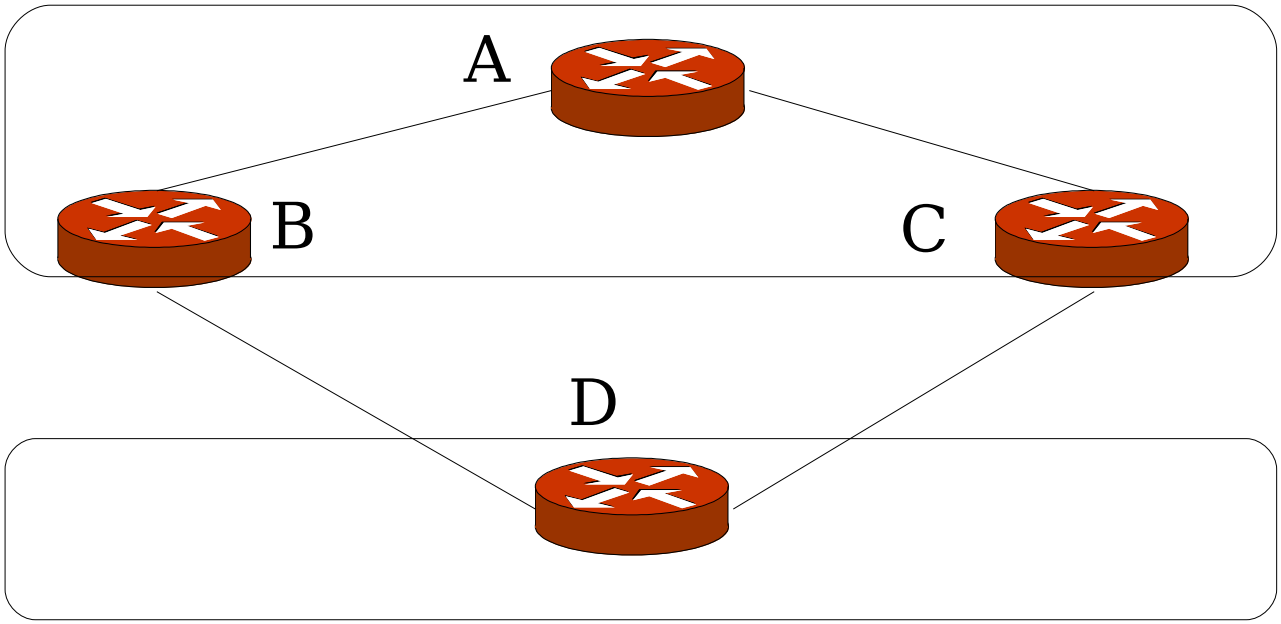


BGP Multipath

- By default, the best selection algorithm in BGP selects *one* route
  - no load balancing from a *single* router to a *single* prefix possible
  - unless “outside” BGP using loopback peering for example
- *BGP multipath* enables load balancing between “equal” paths (to the level of comparing routerids)
- Limited JunOS functionality
  - `set protocol bgp group extern multipath  [multiple-as]`
- By default equal cost multipath
  - Book also describes unequal cost multipath in CISCO

Practical BGP: pages 261-269

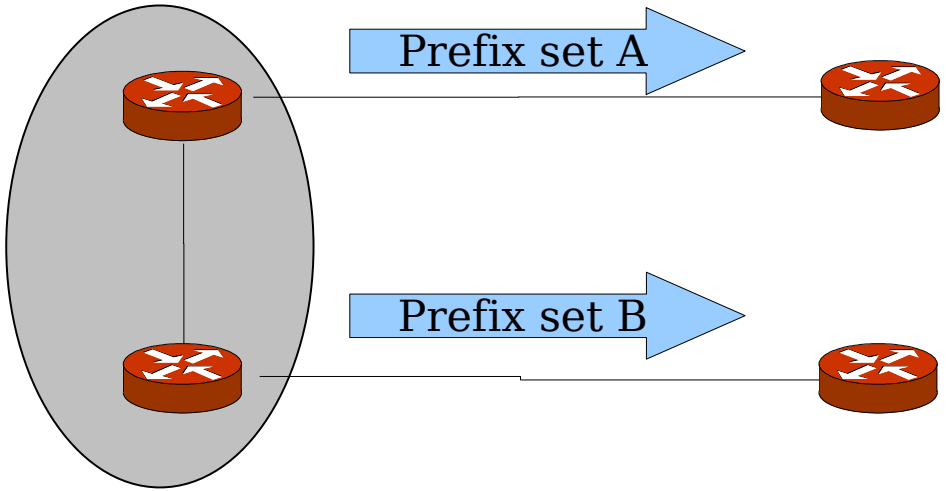
IBGP Multipath example



- Internal (eg from A) vs external (from D) load balancing
- Equal cost vs unequal cost multipath (links between B-D and C-D have different bandwidth).

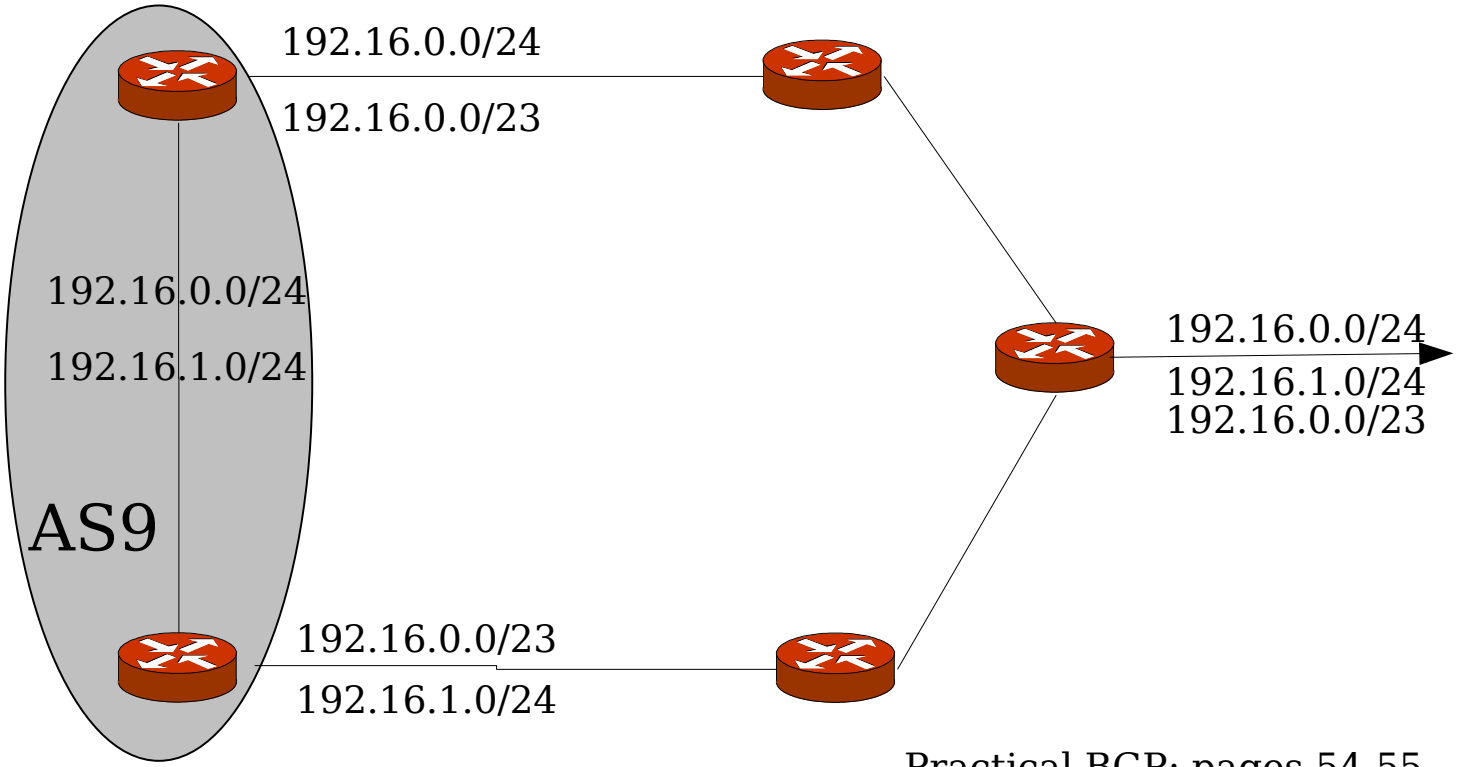
Balance prefix announcements

- For incoming traffic, you can partition your prefixes and anounce different prefix sets on different peerings.
- If you have many transit operators, you can balance the traffic to get optimal price.
- If you want higher granularity of the balancing, it is common to de-aggregate (or refrain from aggregating)
  - Bad for global BGP table sizes



Load balancing + robustness

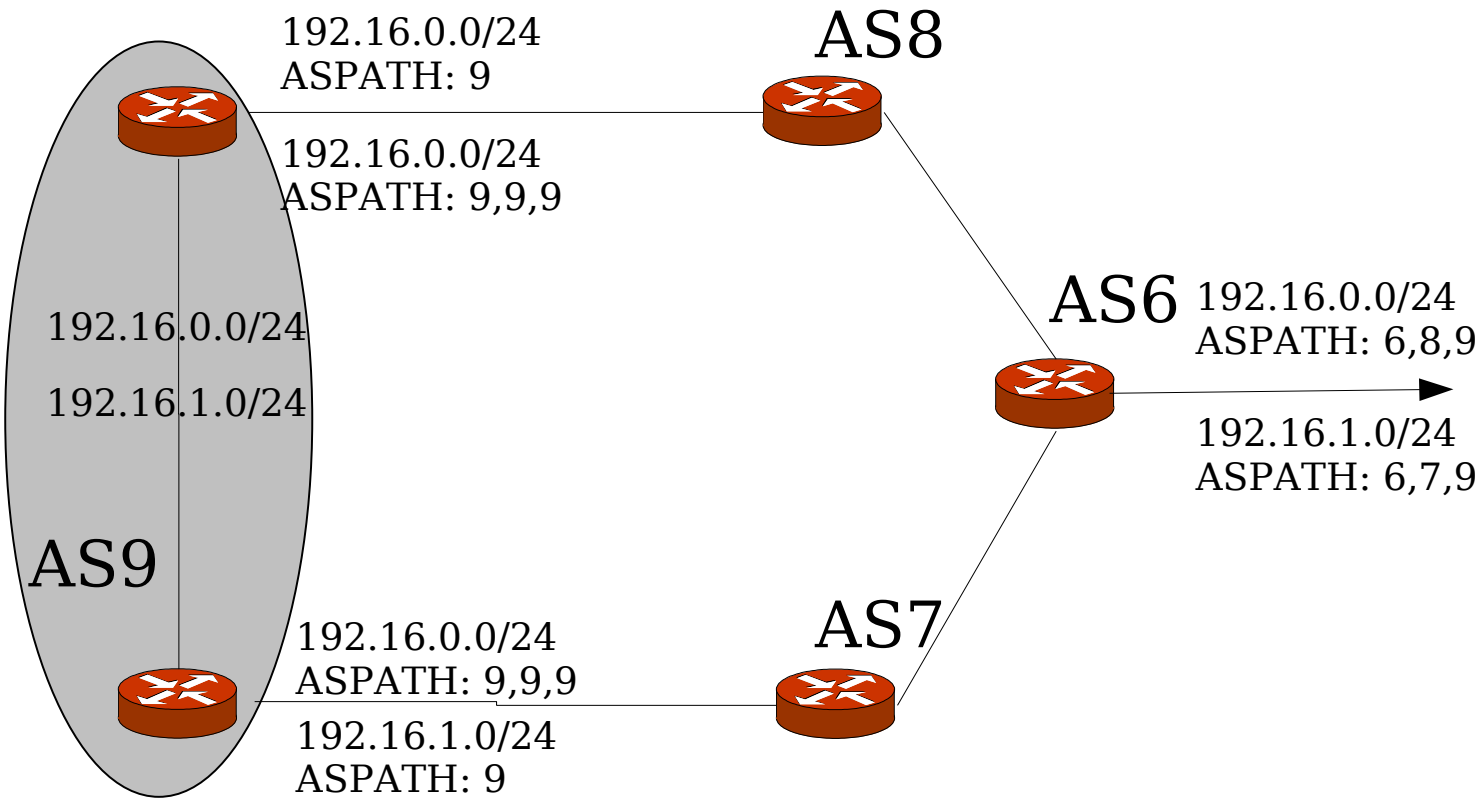
- For robustness, you may wisth to announce all prefixes on all peerings, but use better metric (shorter AS-PATH / longer prefixes) where you prefer traffic
- Example: prefixes



Practical BGP: pages 54-55

Load balancing + robustness

- Example: AS-PATH

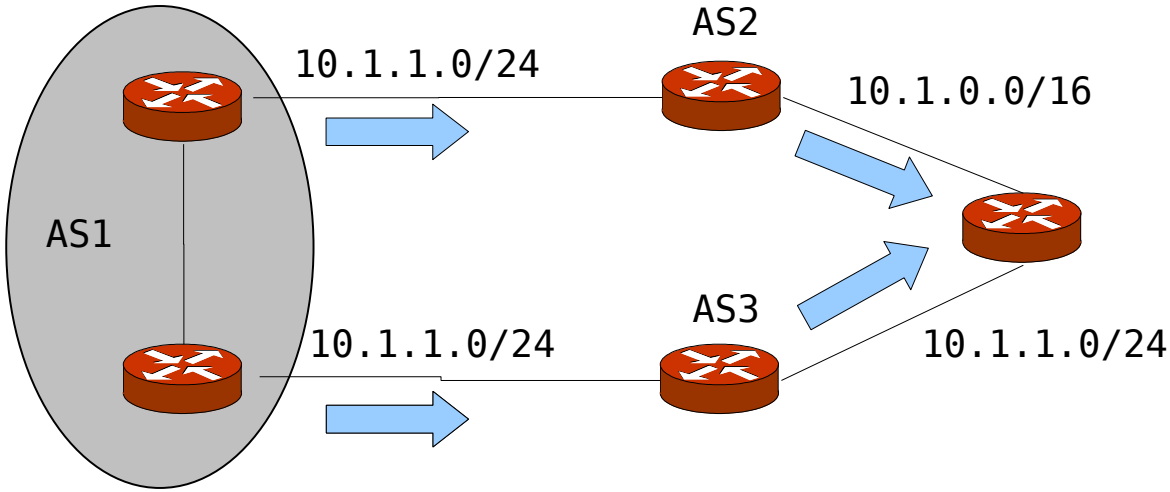


Longer prefix announcements

- Using longer prefixes introduces issues with address allocation and PI/PA addresses
- In the sitiation below where PA addresses are used (AS2:s block), all inbound traffic will flow via AS3
- AS2 must either

  announce the more specific prefix = 'punching a hole' in `10.1.0.0/16` by 'leaking' `10.1.1.0/24`

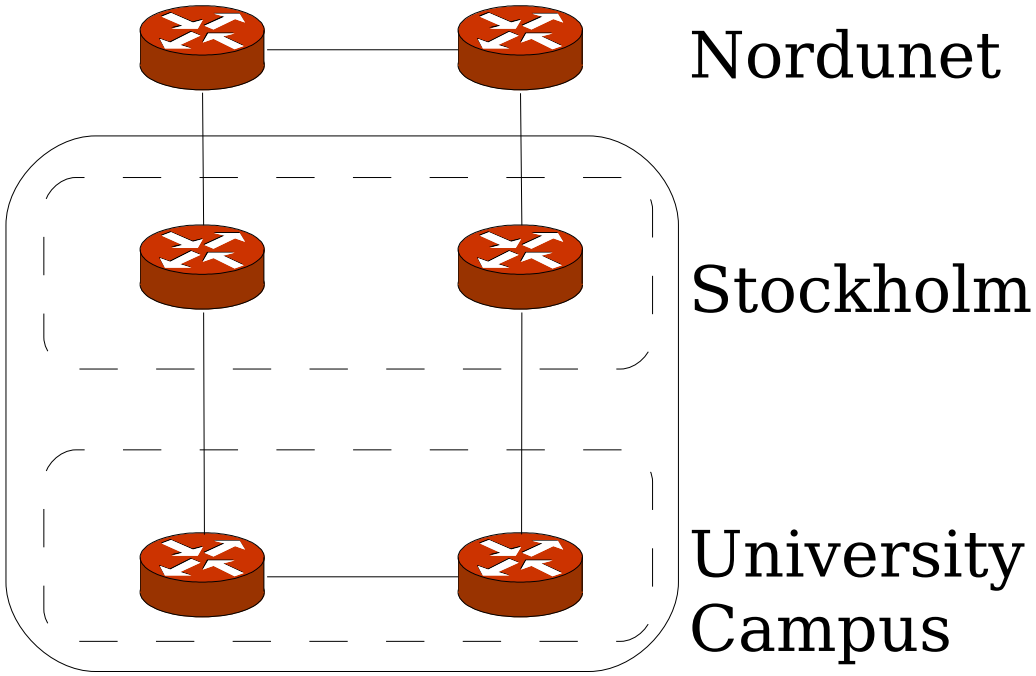
  AS1 must get PI addresses



Practical BGP: pages 56-59

Conditional advertisement

- Advertize a prefix when some (network) condition is met
  - Specific route must be present in routing table
- Example SUNET
  - If IGP reachability fails to campus, do not announce network.



Practical BGP: pages 63-65

Intelligent routing

- A generalization of Conditional advertisement is 'Intelligent routing'
  - External event triggers announcements
- Example 1: Announce an anycast route to a DNS server if you can access an A-record from it
- Example 2: Only announce routes to a site if performance measurement ensures a minimal bandwidth of 10Mb/s

Practical BGP: pages 81-83

Controlling outbound traffic

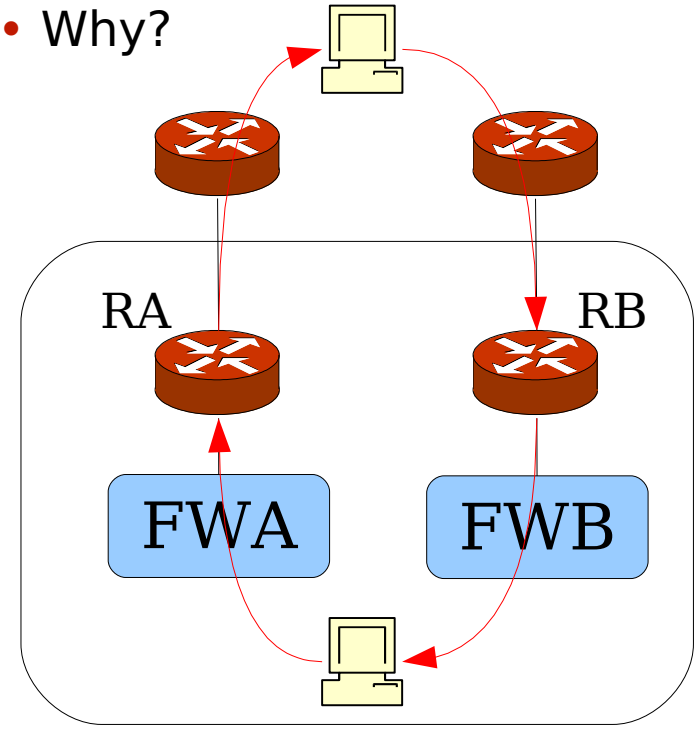
- Outbound control is in general simpler than inbound control since, using IP, you need only control the next-hop.
- No BGP: Announce (default) routes using IGP metrics, load balance using equal-cost multipath
- Control outgoing traffic by filtering input routes
  - accepting partial routes
- Internal multi-path BGP
- Tag with communities for internal use
- or even LOCALPREF (seldom used)

Practical BGP: pages 65-76

Forcing symmetric entry/exit points

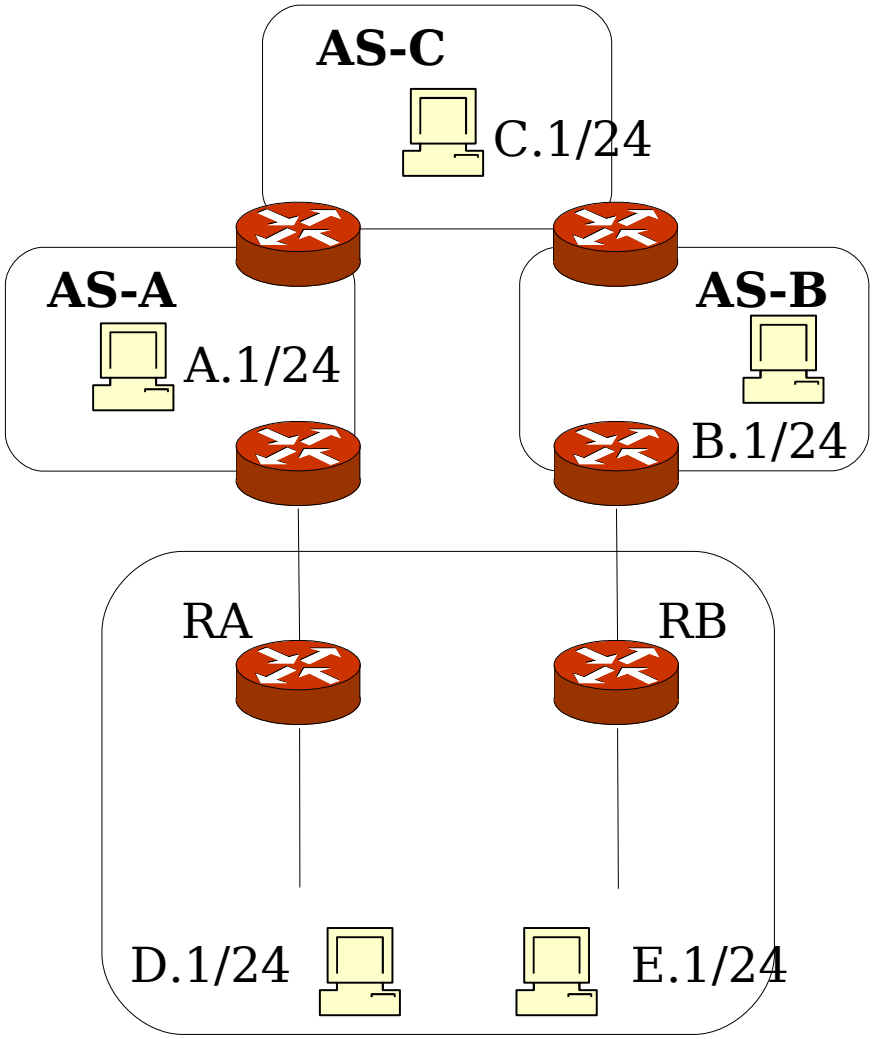
- Asymmetric routing is normal behavior for multi-homed networks
- But enterprise may want to enforce symmetric routing.

  (Providers usually do not have this problem)
- Why?



Practical BGP: pages 77-81

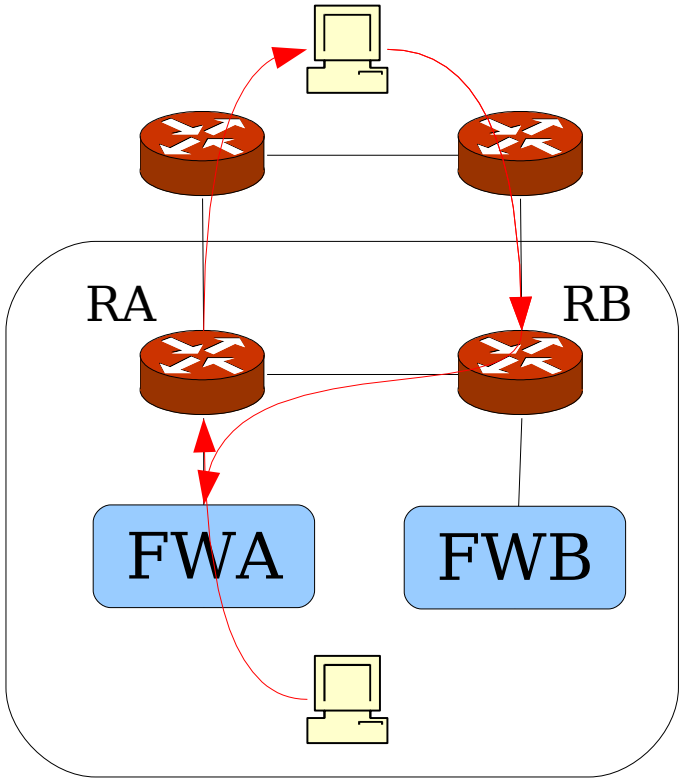
Exercise: Symmetry



- D/24 and E/24 are yout networks that you anounce to AS-A and AS-B
-
- Try to device a *symmetric* solution for D.1 and E.1 communicating with A.1, B.1 and C.1 respectively
- You should have some level of load-balancing.

Forcing symmetric entry/exit points

- Using the topology shown you do not actually require symmetric routing, it is enough to be traffic be symmetric w.r.t the firewalls: the routing could be asymmetric over RA and RB.
- This is usually done by splitting the address space internally and letting the IGP handle symmetric FW access



Load-balancing lab

- EBGP loopback peering
- EBGP over multiple links
- Balance prefixes on multiple EBGP peerings
- Redundancy