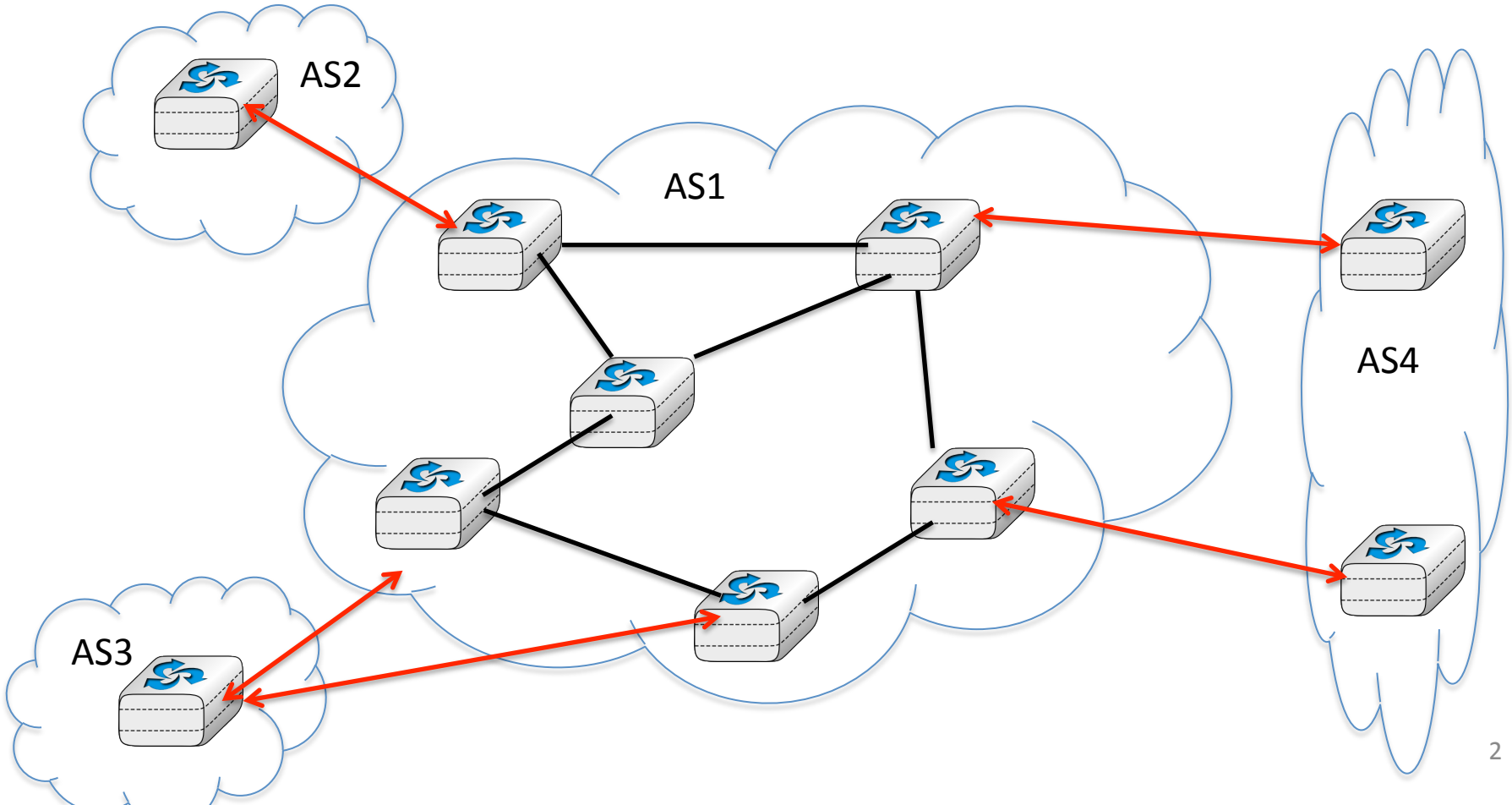
INGI2142

Week 1 : iBGP

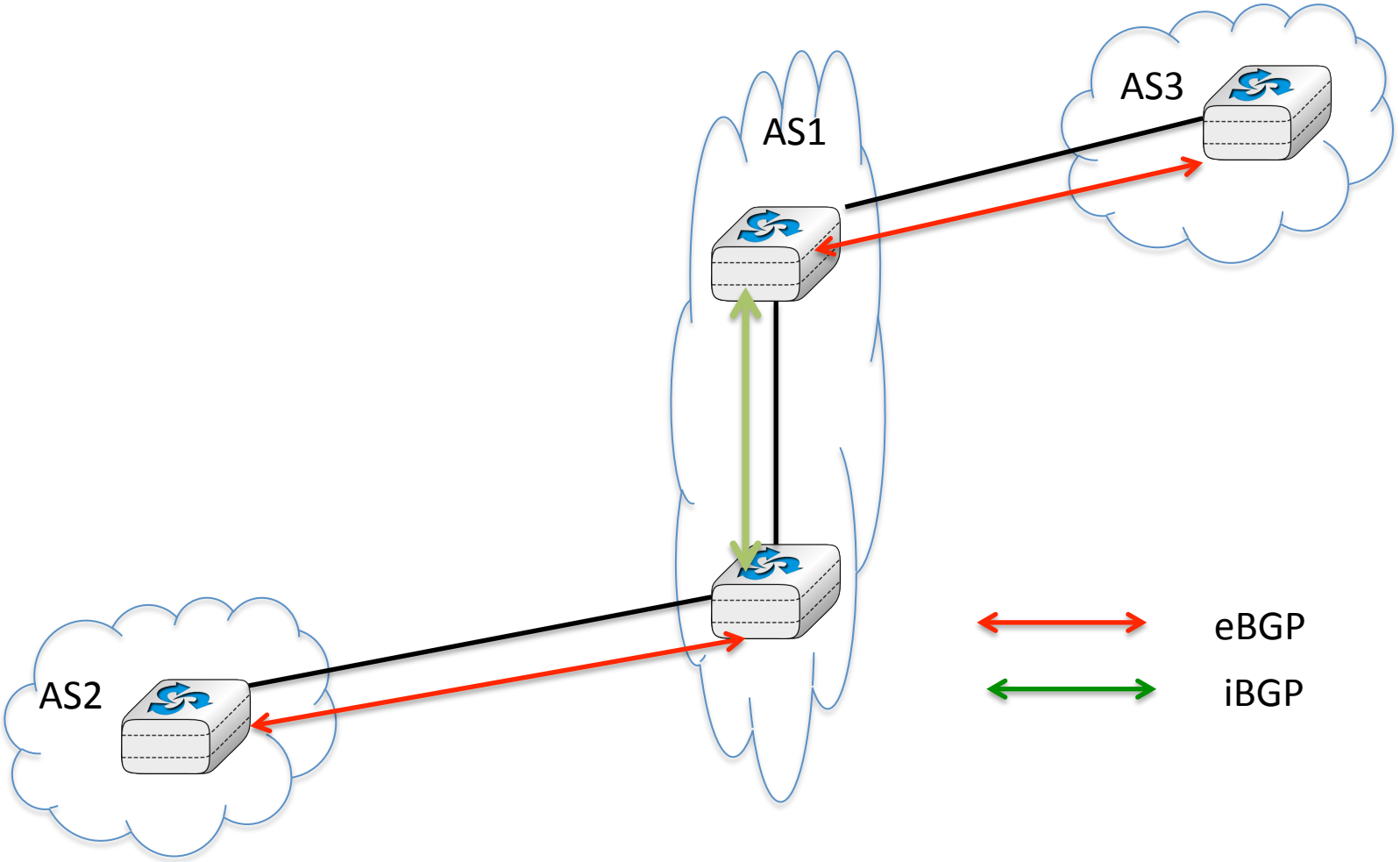
O. Bonaventure, 2016

BGP in large networks

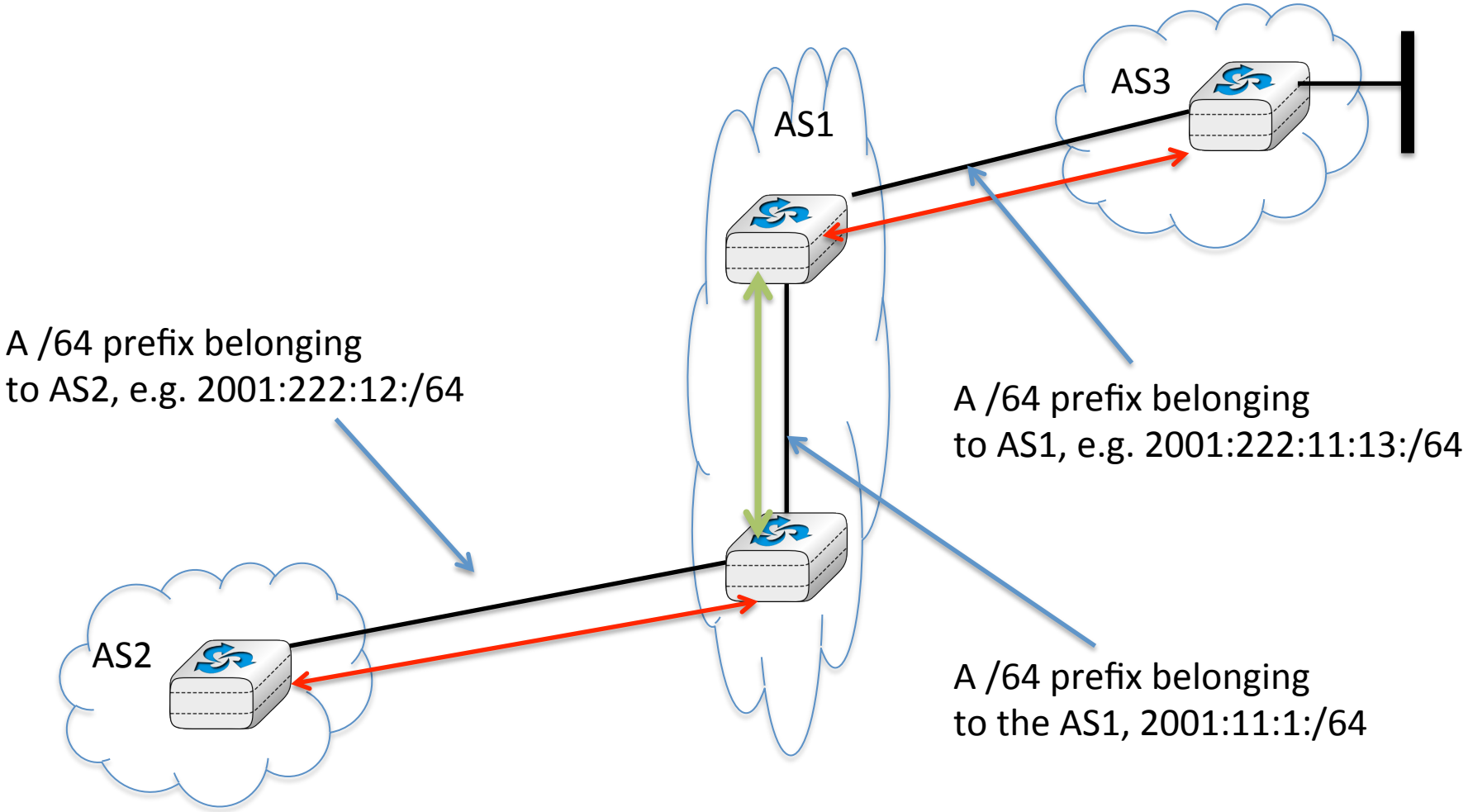
- What happens when a network contains tens, hundreds or thousands of routers ?



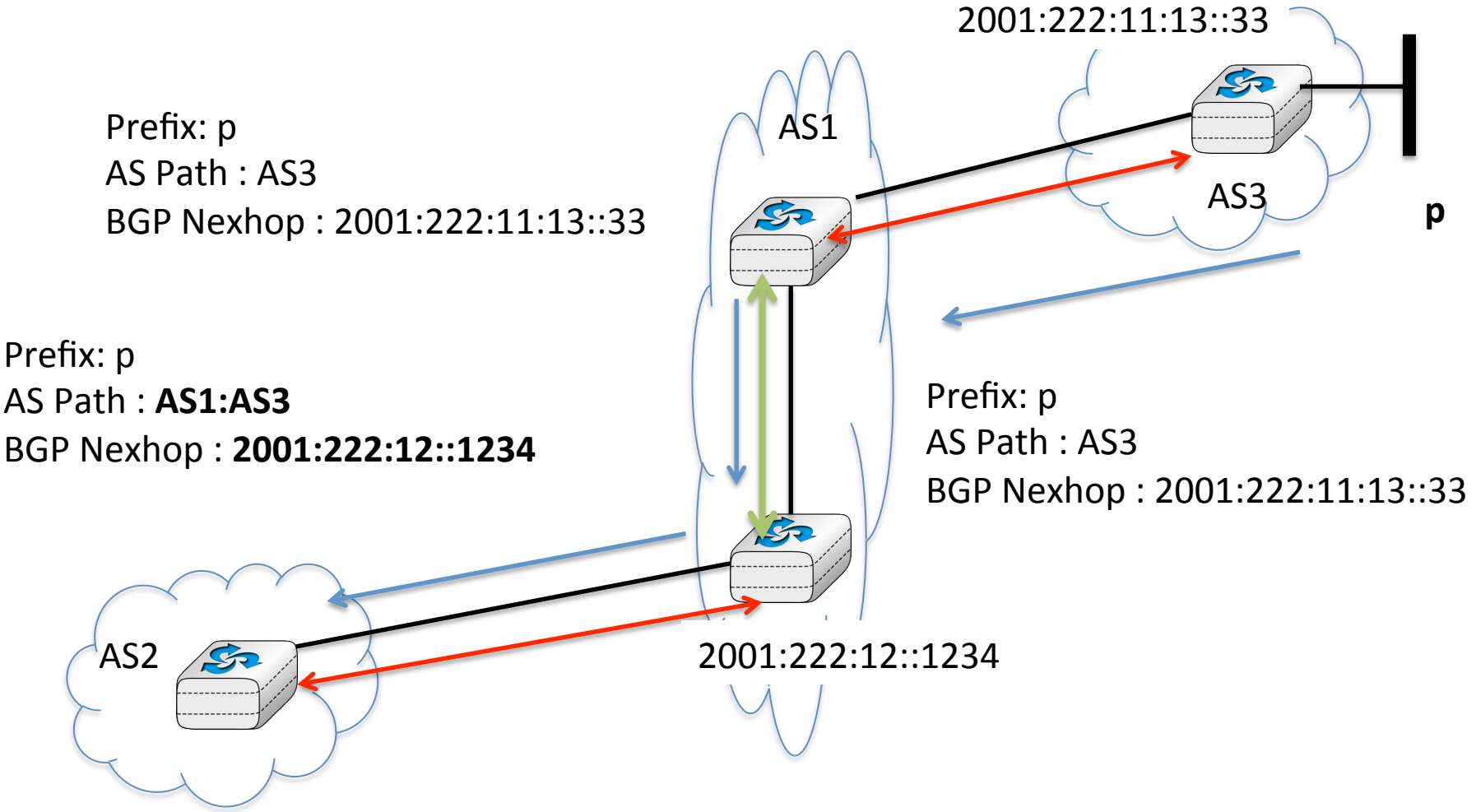
How to distribute BGP routes in a large network ?



IP addresses used on routers

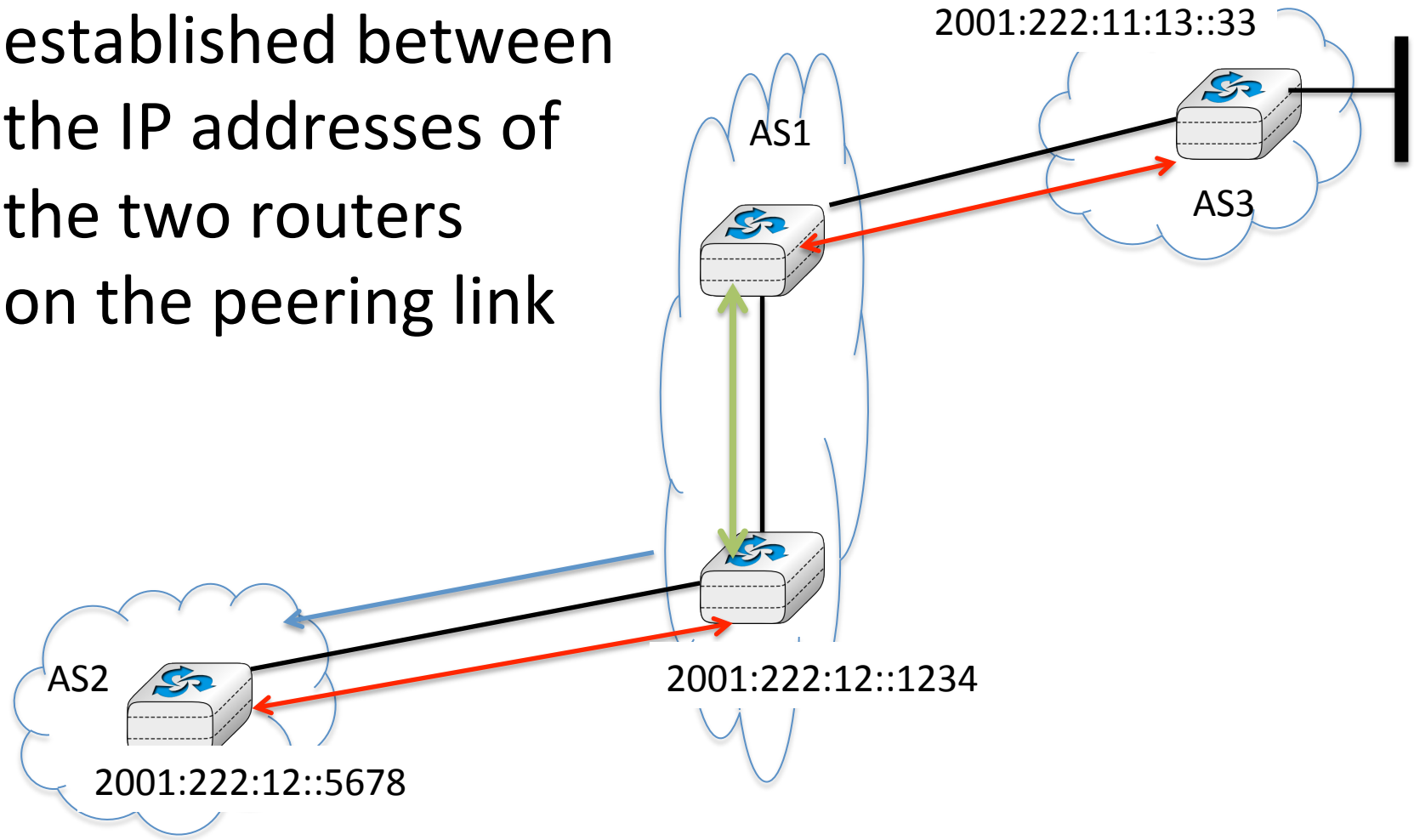


A closer look at the BGP messages



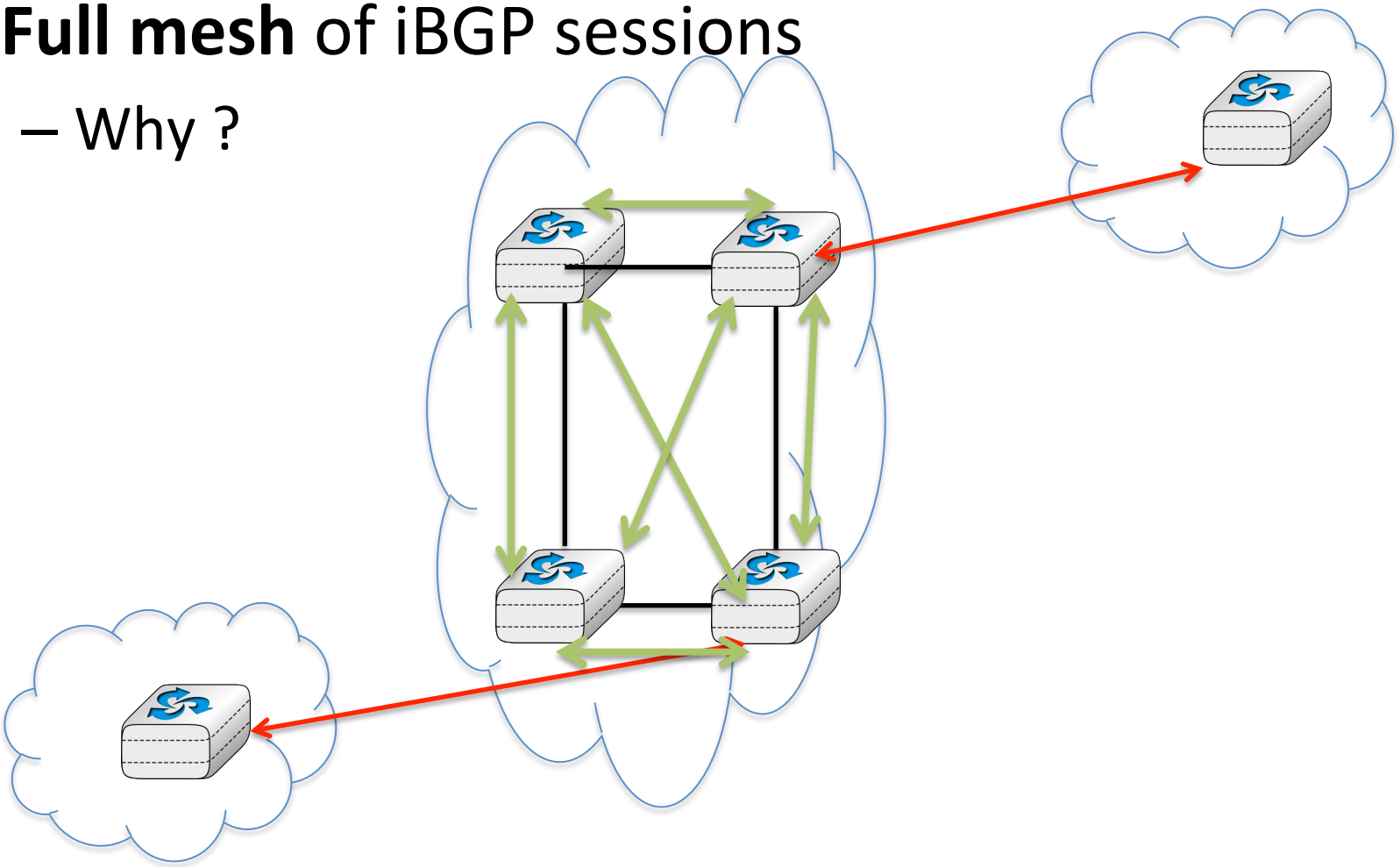
Endpoints of the eBGP sessions

- eBGP sessions are TCP connections established between the IP addresses of the two routers on the peering link

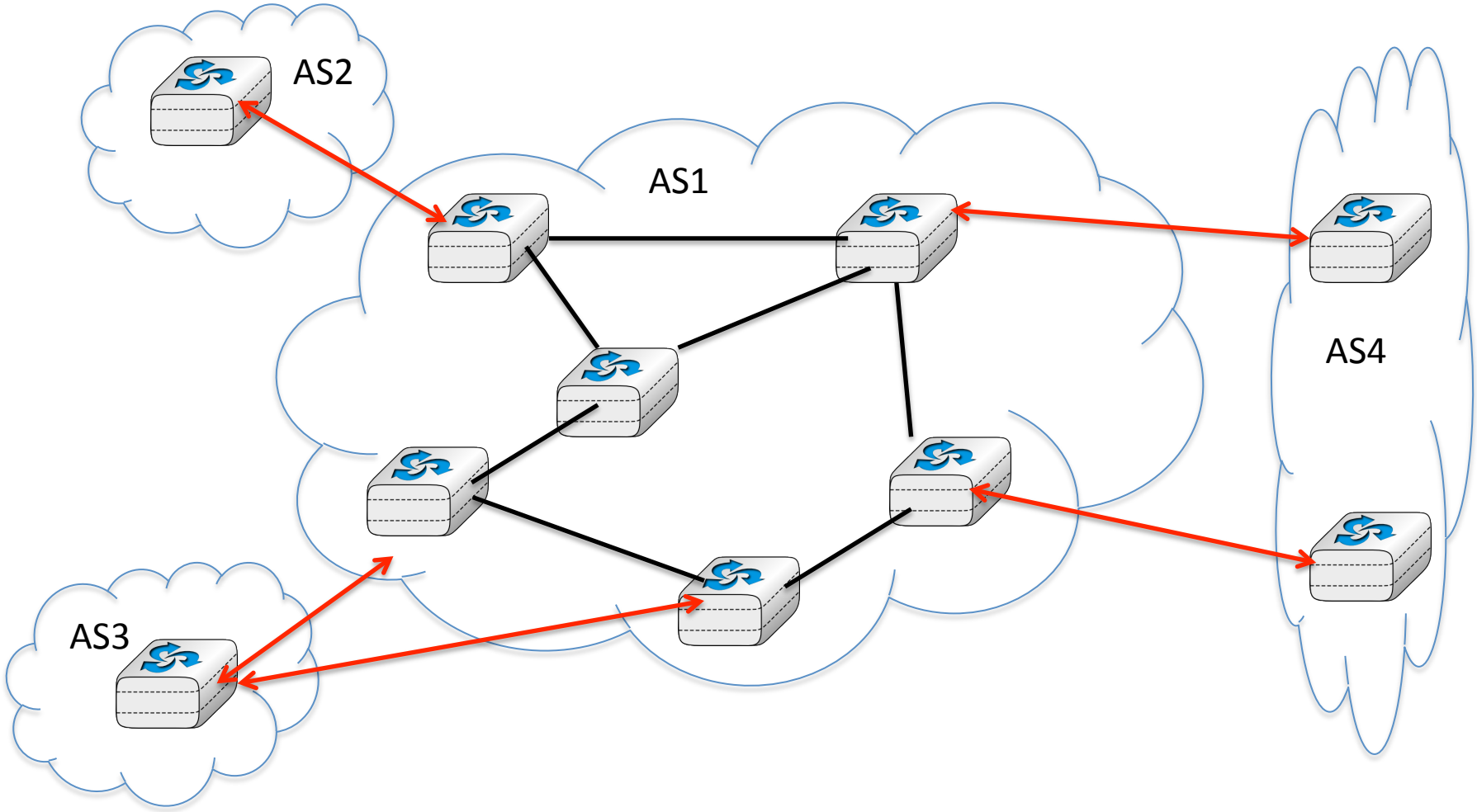


How to distribute BGP routes in a large network ?

- **Full mesh** of iBGP sessions
  - Why ?



Draw iBGP sessions in AS1

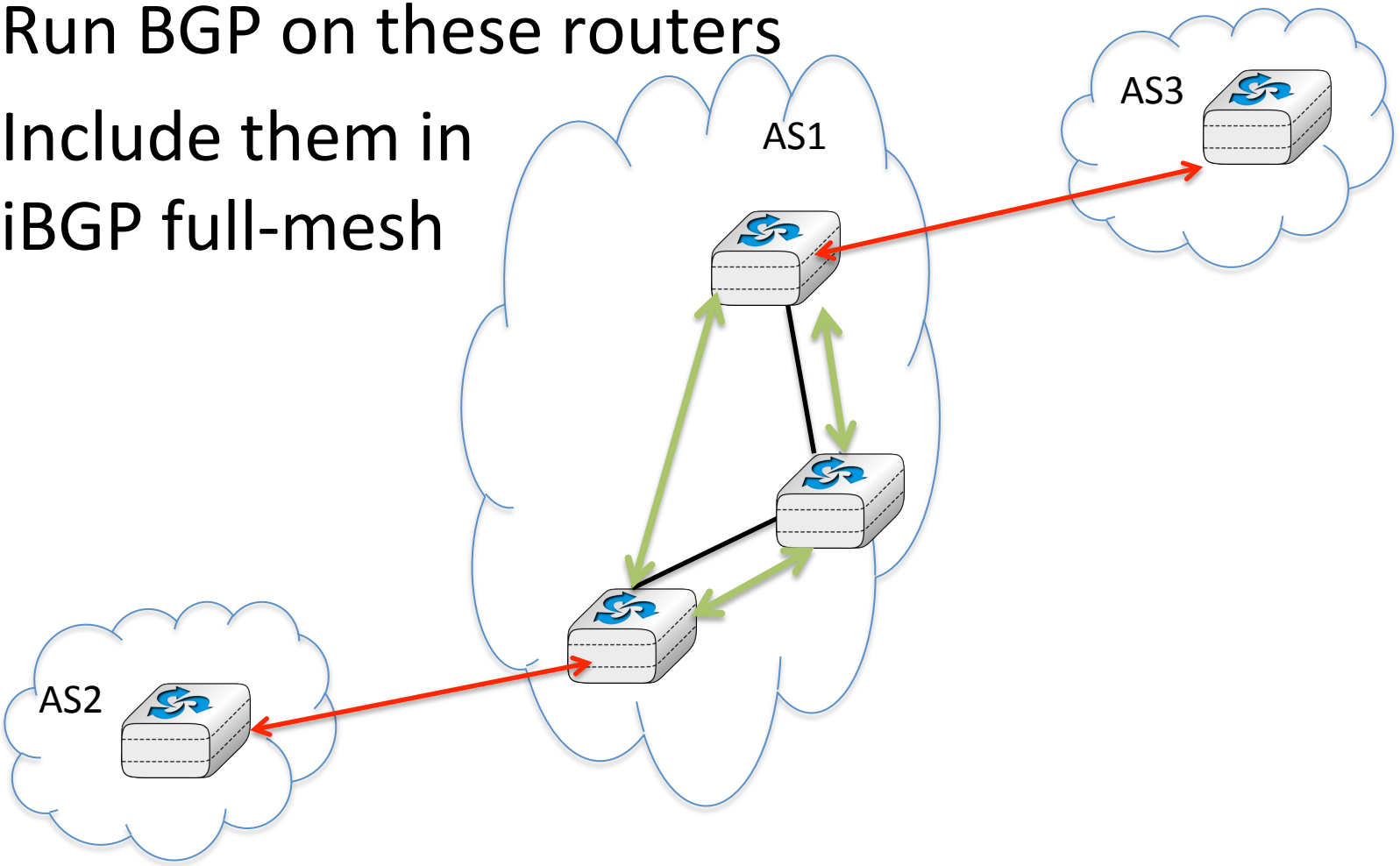


Endpoints of the iBGP sessions

- A router has several IP addresses, on which address should an iBGP session terminate ?
  - Any IP address belonging to the router
    - Address is associated to an interface, if the interface stops, iBGP session stops as well even if the router is still reachable over other interfaces
  - A loopback address
    - A software only interface that is always up and announced through the intradomain routing protocol so that it remains reachable as long as the router has one interface up
    - **Best Current Practice**

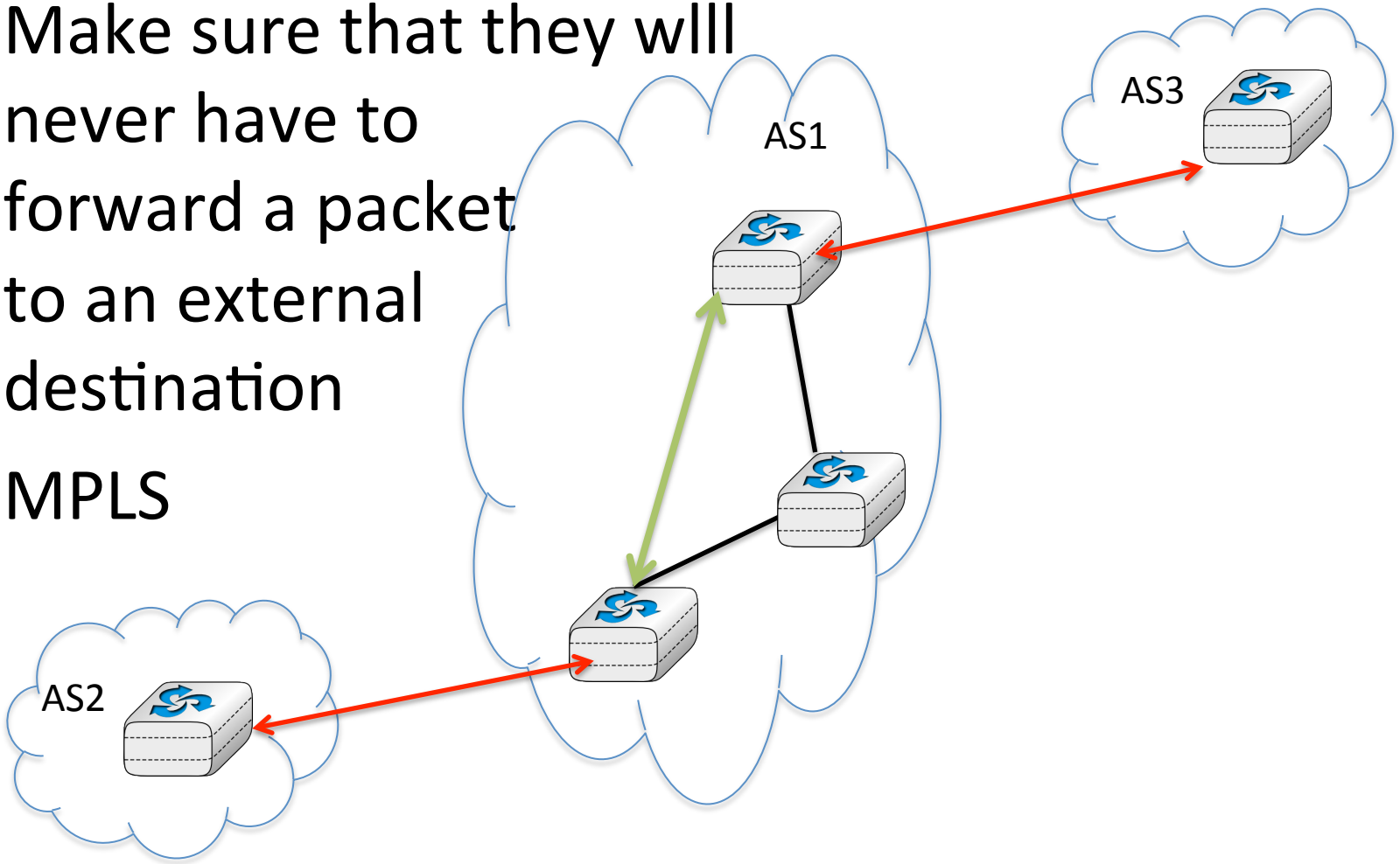
How to deal with routers that are not connected to other ASes ?

- Run BGP on these routers
- Include them in iBGP full-mesh



How to deal with routers that are not connected to other ASes ?

- Make sure that they wlll never have to forward a packet to an external destination
- MPLS

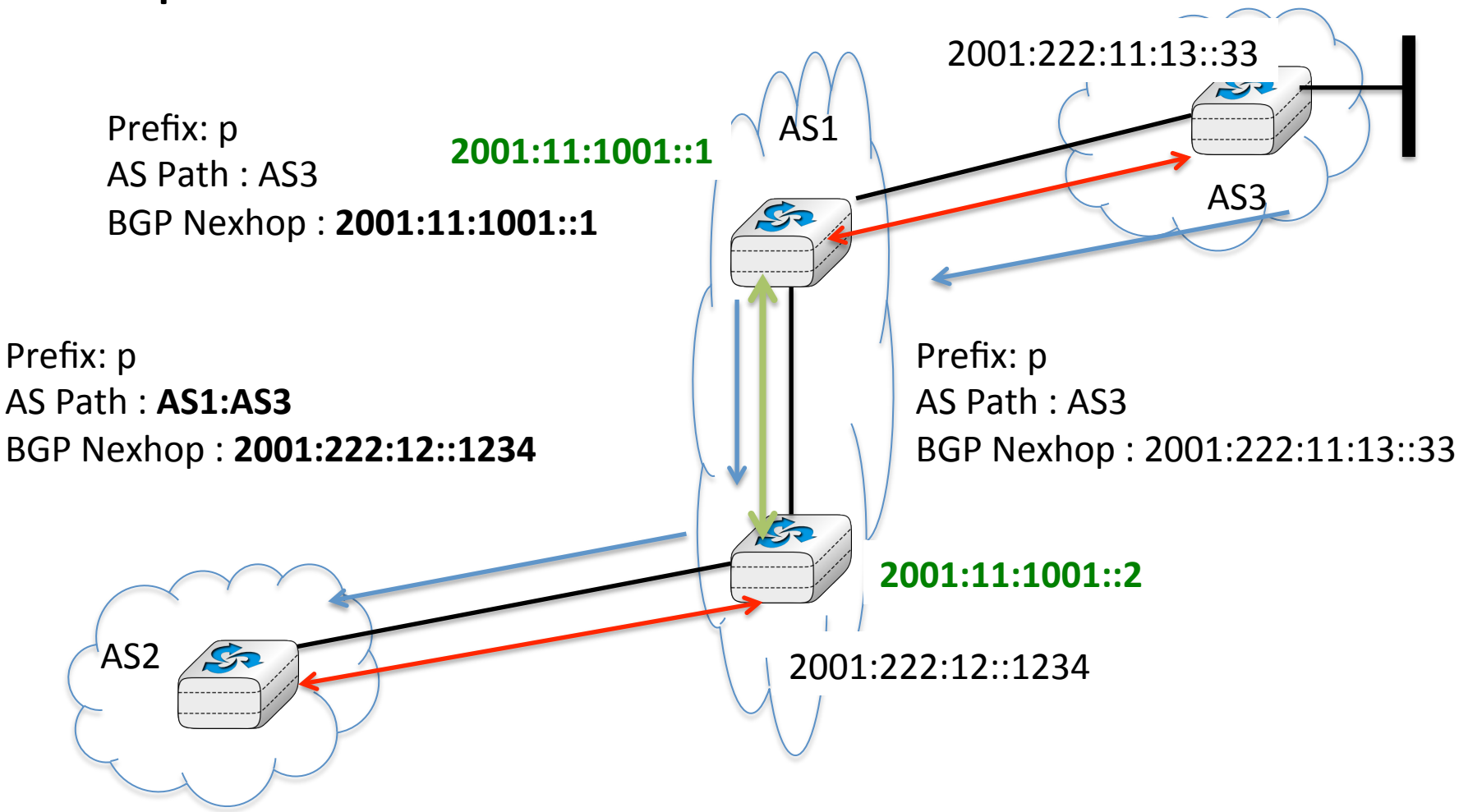


What are the roles of the IGP ?

- The intradomain routing protocol distributes information about the reachability of :
  - IP prefixes associated to internal links between routers of the AS
  - IP addresses associated to loopback interfaces or routers of the AS
  - IP prefixes associated to peering links between a router of this AS and another AS

BGP Nexthop self

- Helps to reduce # routes in the IGP



The BGP decision process

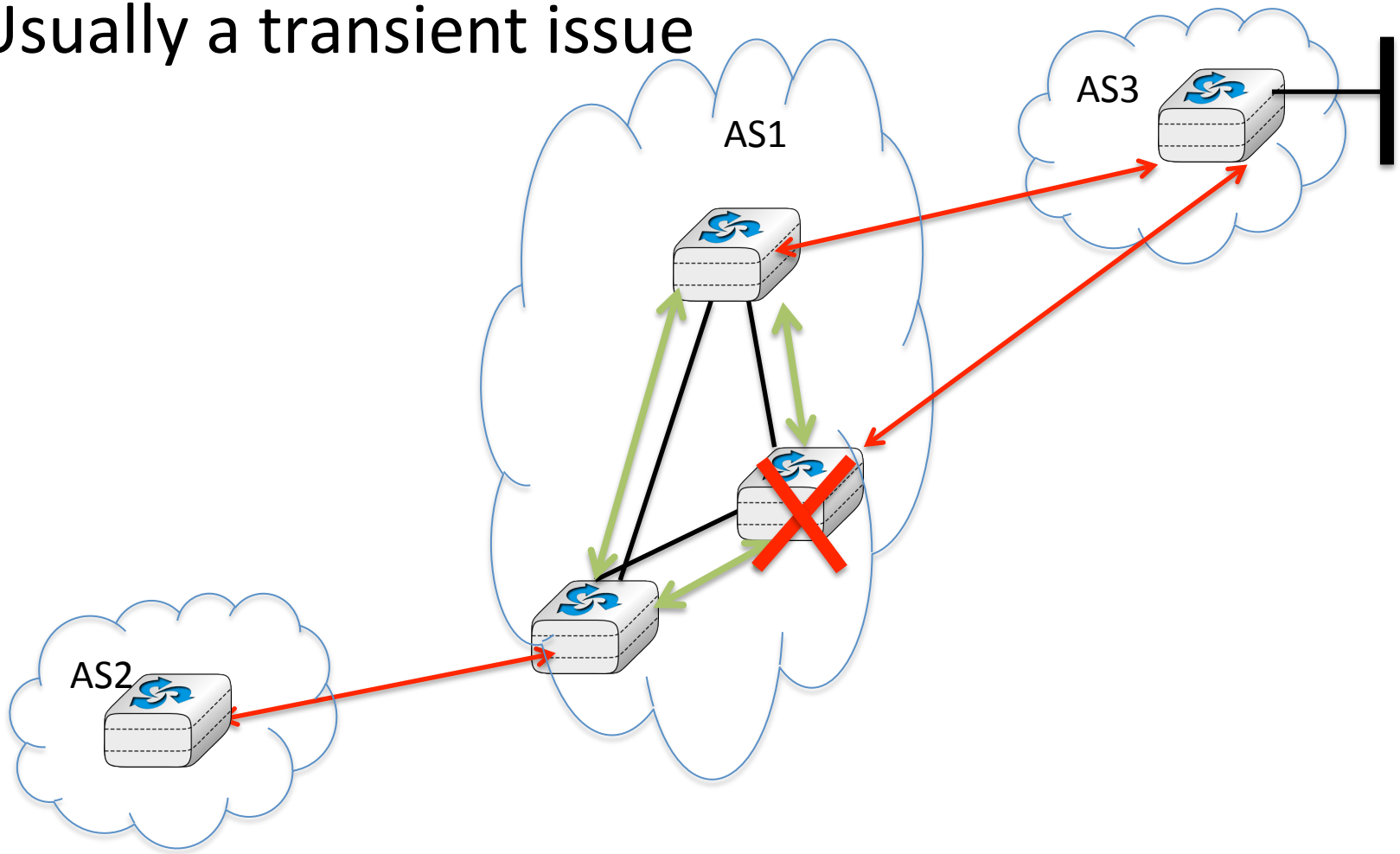
1. Ignore routes having an unreachable BGP nexthop
2. Prefer routes having the highest local-pref
3. Prefer routes having the shortest AS-Path
4. Prefer routes having the smallest MED
5. Prefer routes learned via eBGP sessions over routes learned via iBGP sessions
6. Prefer routes having the closest next-hop
7. Tie breaking rules : prefer route learned from the router with lowest router id

1st step of BGP decision process

- 1. Ignore routes having an unreachable BGP nexthop
  - Why would a BGP route contain an unreachable nexthop ?

Unreachable nexthop

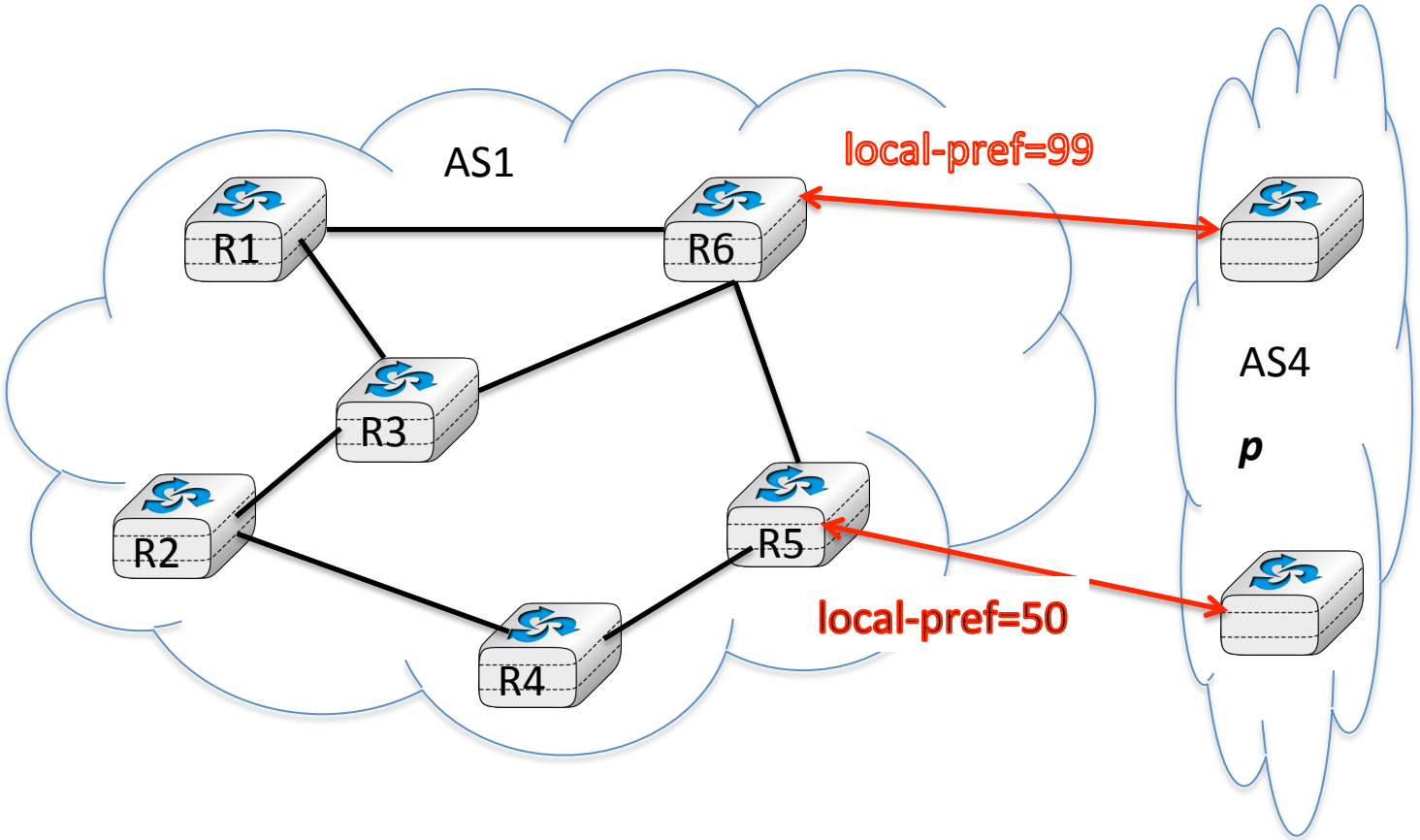
- Usually a transient issue



2nd step of BGP decision process

- 2. Prefer routes having the highest local-pref
  - Implement routing policies
    - Prefer customer routes over shared-cost and provide routers
  - Support backup routes
- Local-pref attribute to added by the import filter on eBGP session and distributed to all routers over iBGP sessions

BGP routes towards prefix *p* on all routers inside AS1



3rd step of BGP decision process

- 3. Prefer routes having the shortest AS-Path
  - Some operators believe that a BGP route with a long AS Path has a lower performance than a route with a longer AS Path
    - This is not always true…

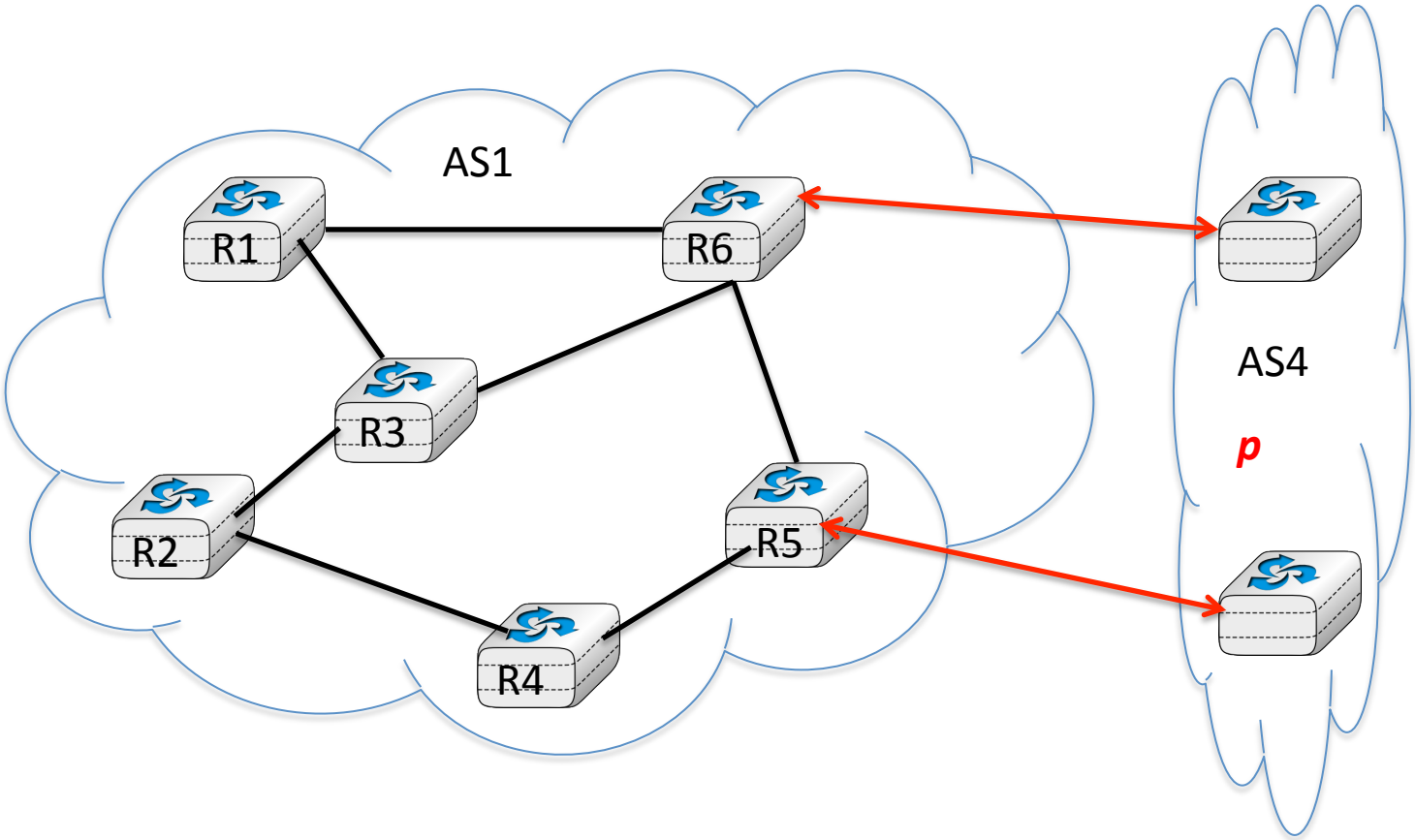
5th step of BGP decision process

- 5. Prefer routes learned via eBGP sessions over routes learned via iBGP sessions
  - Motivation : Hot potato routing
    - Routers should try to forward packets towards external destinations to another AS a quickly as possible

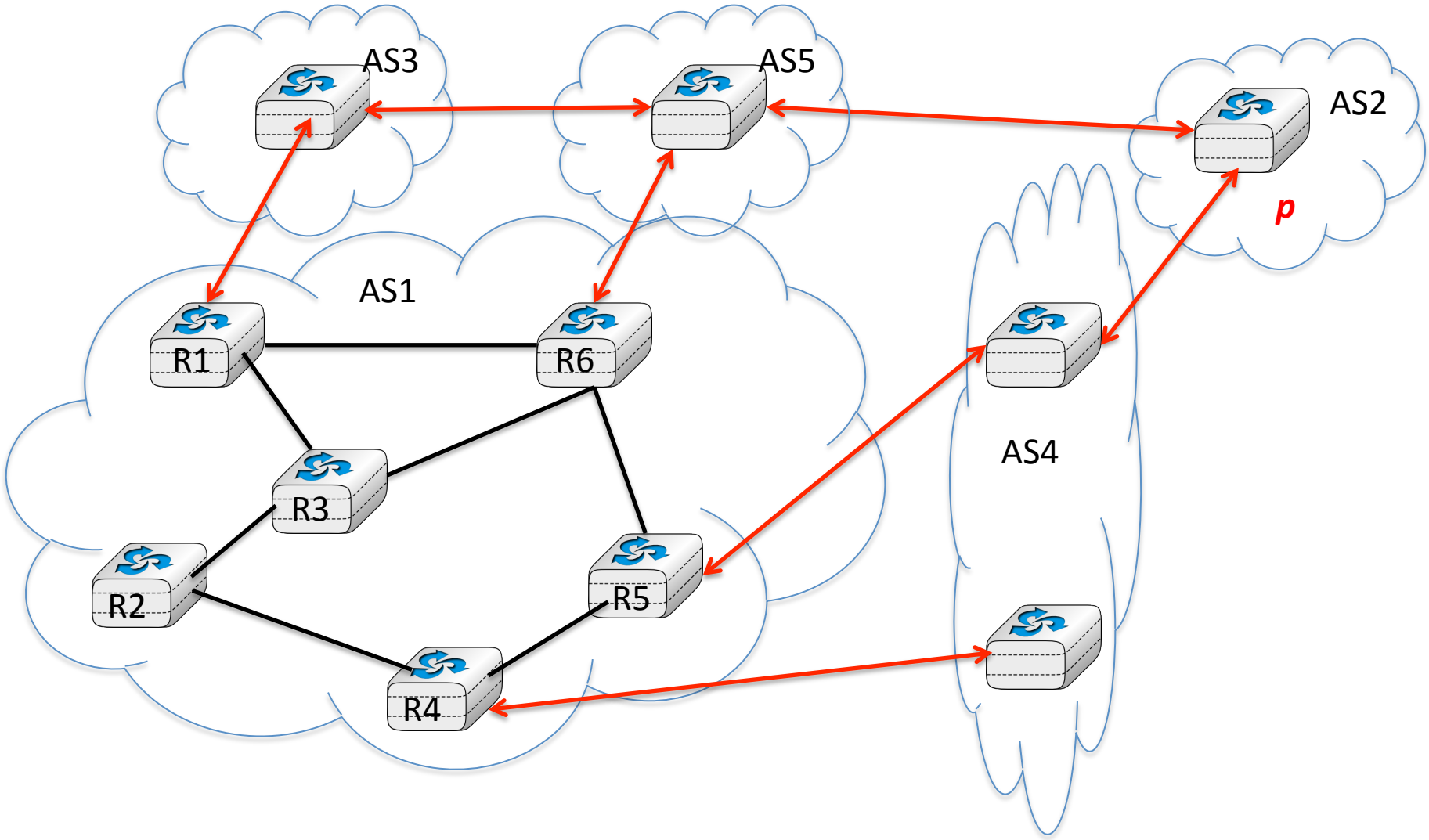
6th step of BGP decision process

- 6. Prefer routes having the closest next-hop
  - Motivation : Hot potato routing
    - Routers should try to forward packets towards external destinations to another AS a quickly as possible

BGP routes towards prefix *p* on all routers inside AS1



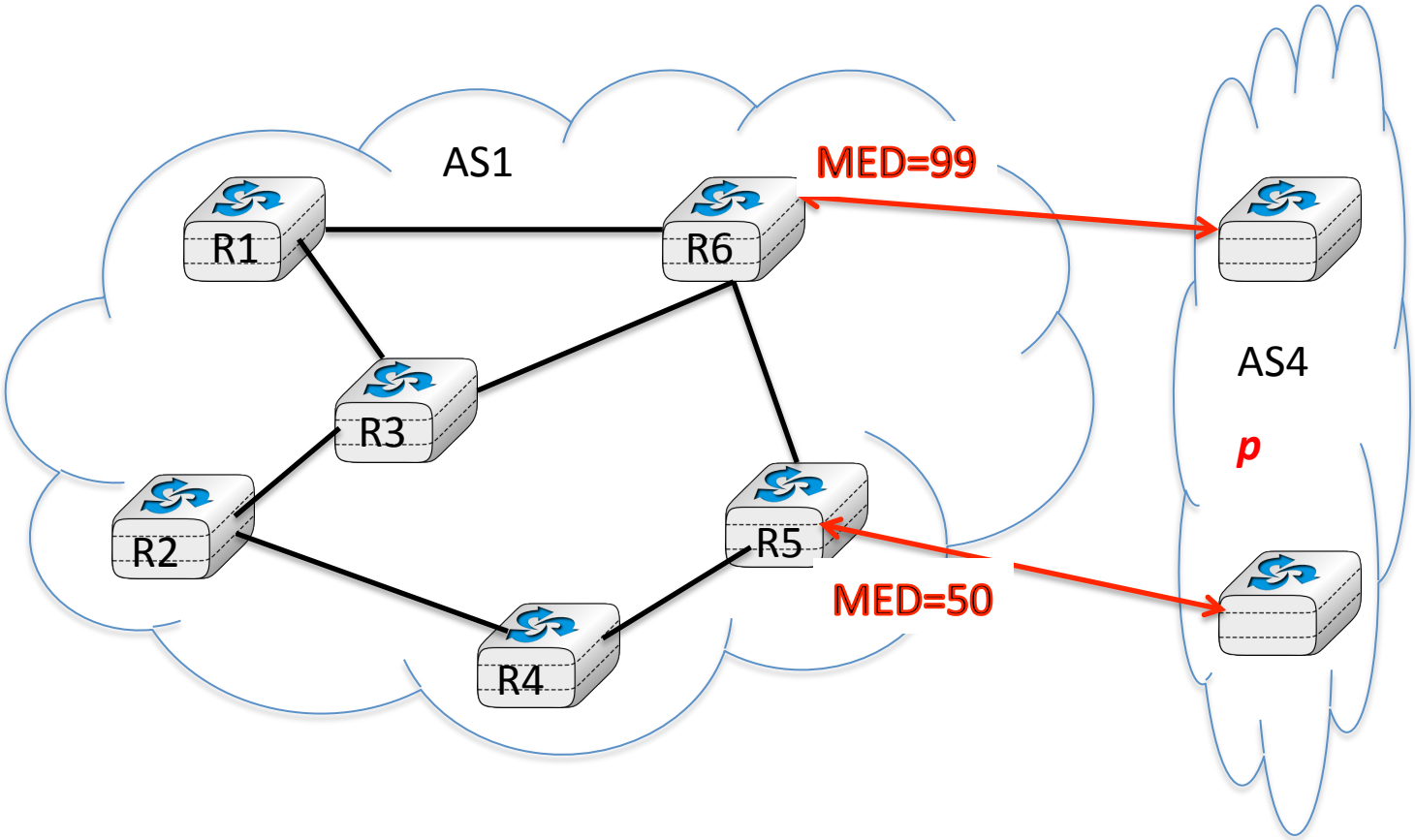
BGP routes towards prefix *p* on all routers inside AS1



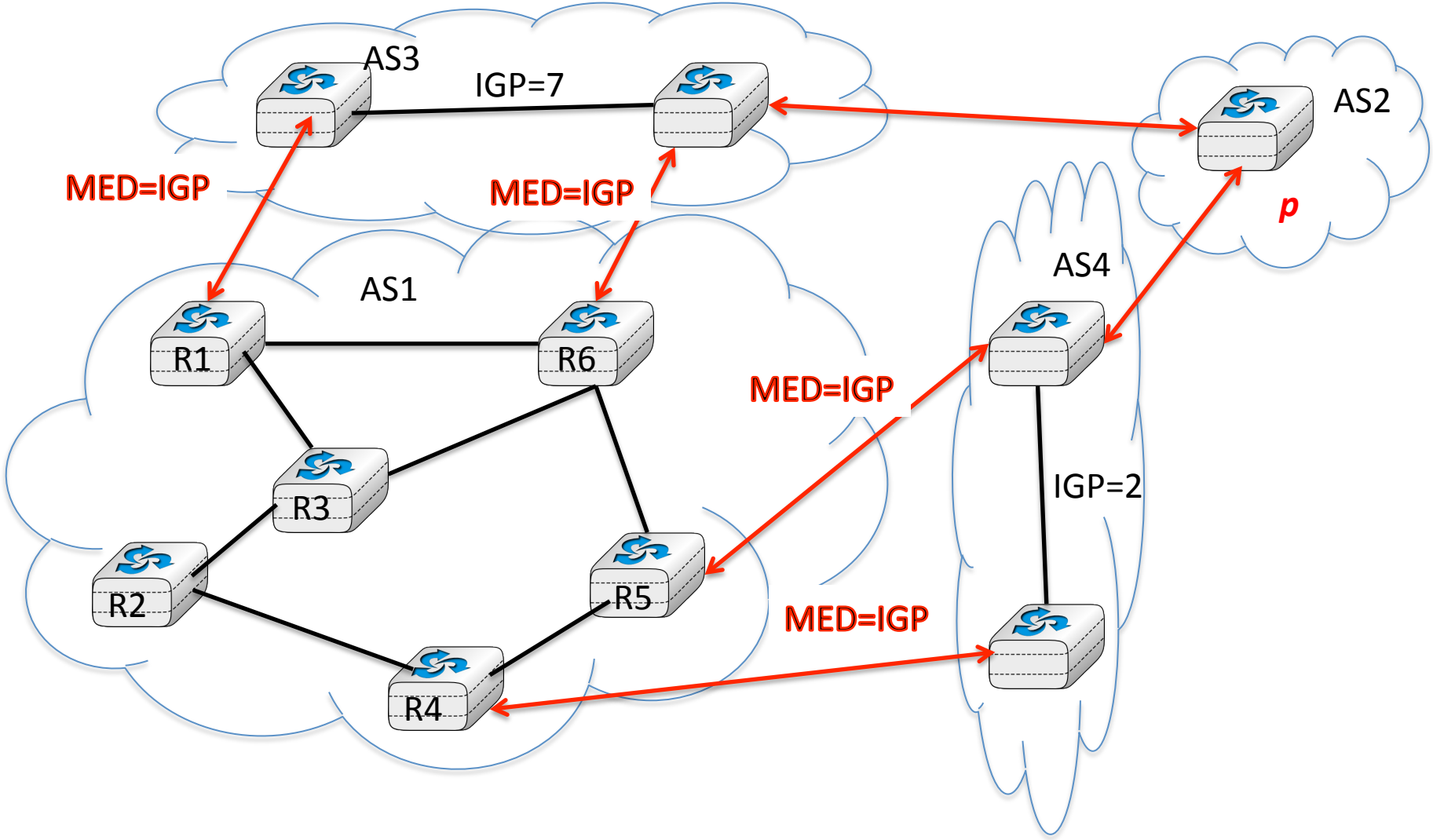
4th step of BGP decision process

- Prefer routes having the smallest MED
- MED means Multi-Exit Discriminator
  - Motivation : Cold potato routing
  - Enable neighbour AS (usually customer) to indicate the best peering link to reach a given prefix

BGP routes towards prefix *p* on all routers inside AS1



BGP routes towards prefix *p* on all routers inside AS1



7^{th} step of the BGP decision process

- 7. Tie breaking rules : prefer route learned from the router with lowest router id
  - Motivation : select a single best route towards each destination prefix
    - This best route is the route that can be advertised over eBGP session
  - Note that recent routers sometimes load balance (BGP Multipath) the traffic towards a given prefix over several routes (having the same BGP attributes)

Differences between iBGP and eBGP

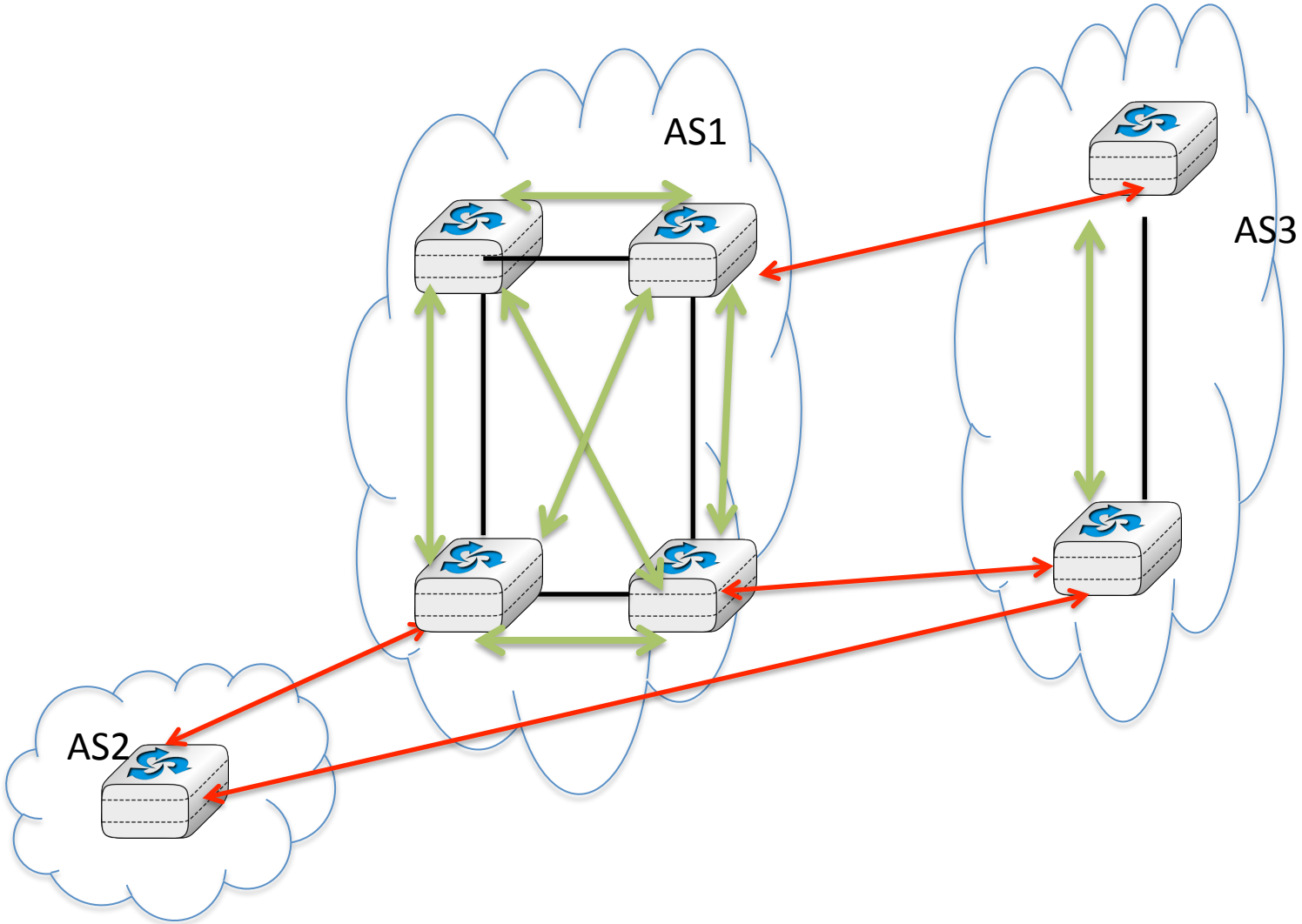
- Which routes are advertised by a router over iBGP and eBGP sessions ?
  - Over an eBGP session, a router advertises its best route towards each destination prefix
    - Provided this advertisement is allowed by the export filter
    - At most one route per destination prefix
  - Over an iBGP session, a router advertises its best route towards each destination prefix provided that this best route was learned over an eBGP session
    - Since iBGP sessions are in full-mesh, there is no need to readvertise a route learned over another iBGP session

Differences between iBGP and eBGP

– Which filters are used over iBGP and eBGP sessions ?

- Import and export filters are used over all eBGP sessions
  - Usually, there is a series of import filters attached to each peering link
  - Filters are usually implemented as modules that are combined together and associated to similar peering links (e.g. customer filter, provider filter, …)
- No filter is applied on iBGP sessions

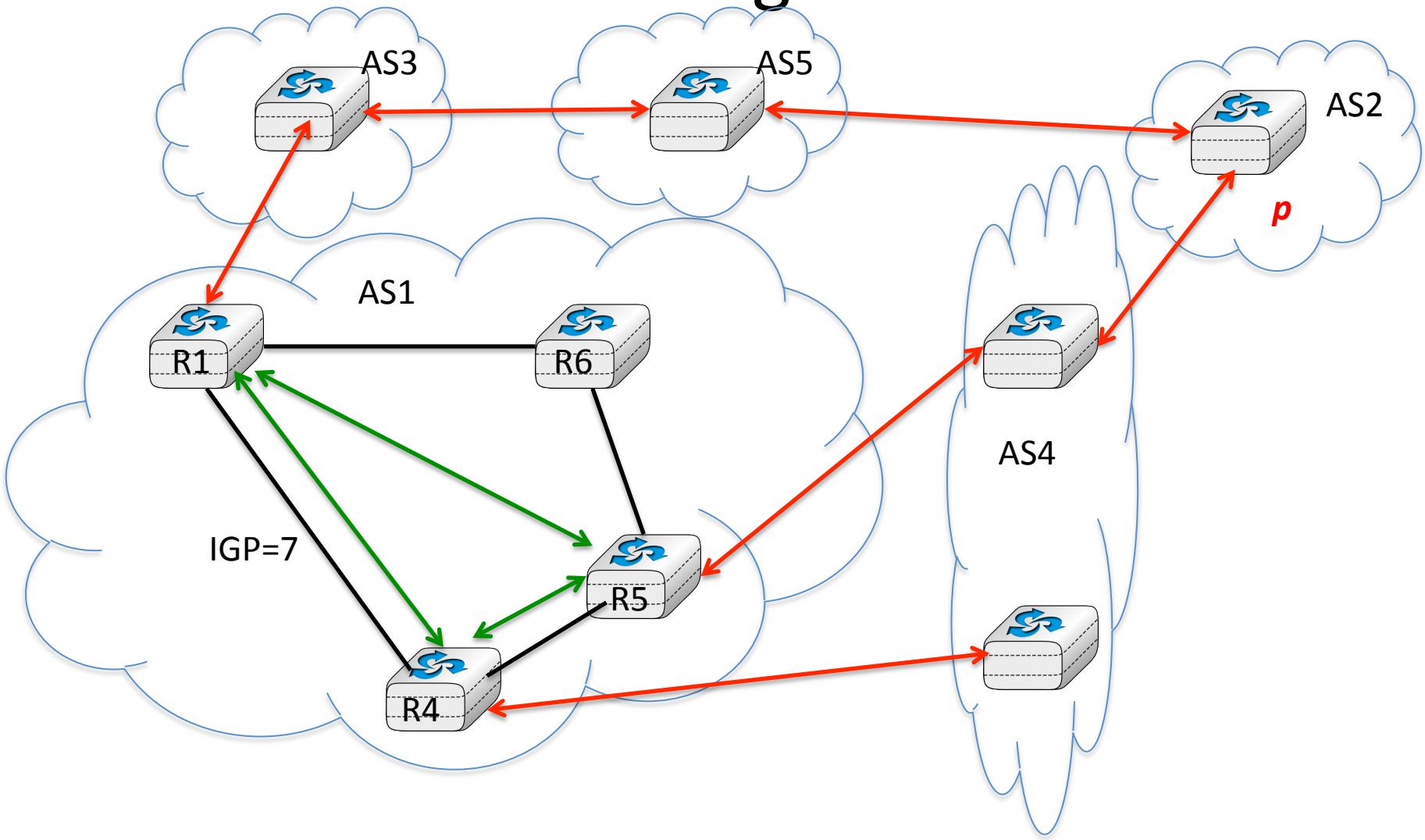
Which BGP routes are known ?



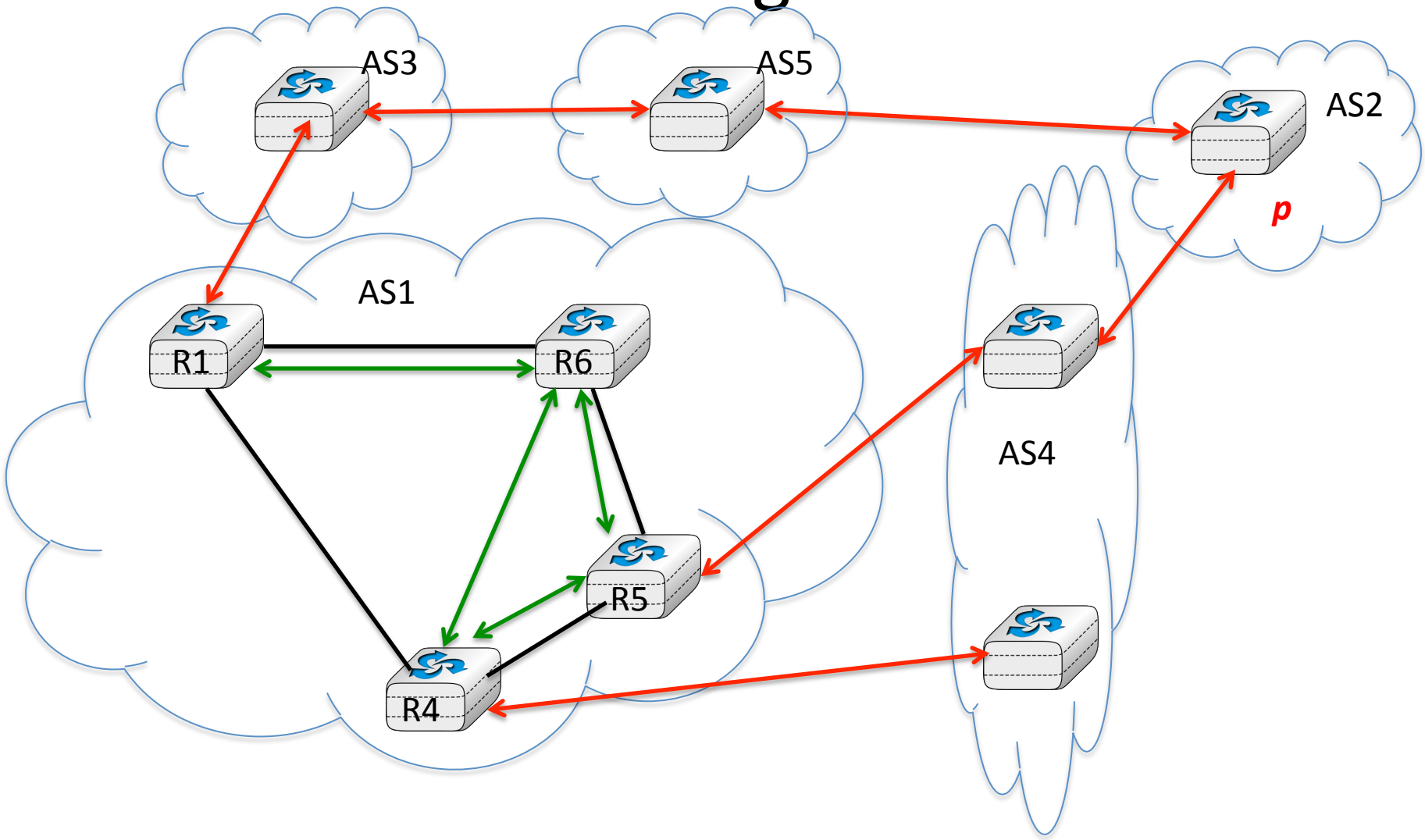
Differences between iBGP and eBGP

- What are the attributes carried by iBGP and eBGP
  - Prefix : iBGP and eBGP
  - AS Path : iBGP and eBGP
    - AS Path is updated when a BGP message is sent over an eBGP session
  - Local-pref : iBGP
    - Local-pref cannot be used on eBGP sessions
  - MED : iBGP and eBGP
  - Nexthop : iBGP and eBGP
    - Nexthop is updated when a BGP message is sent over an eBGP session

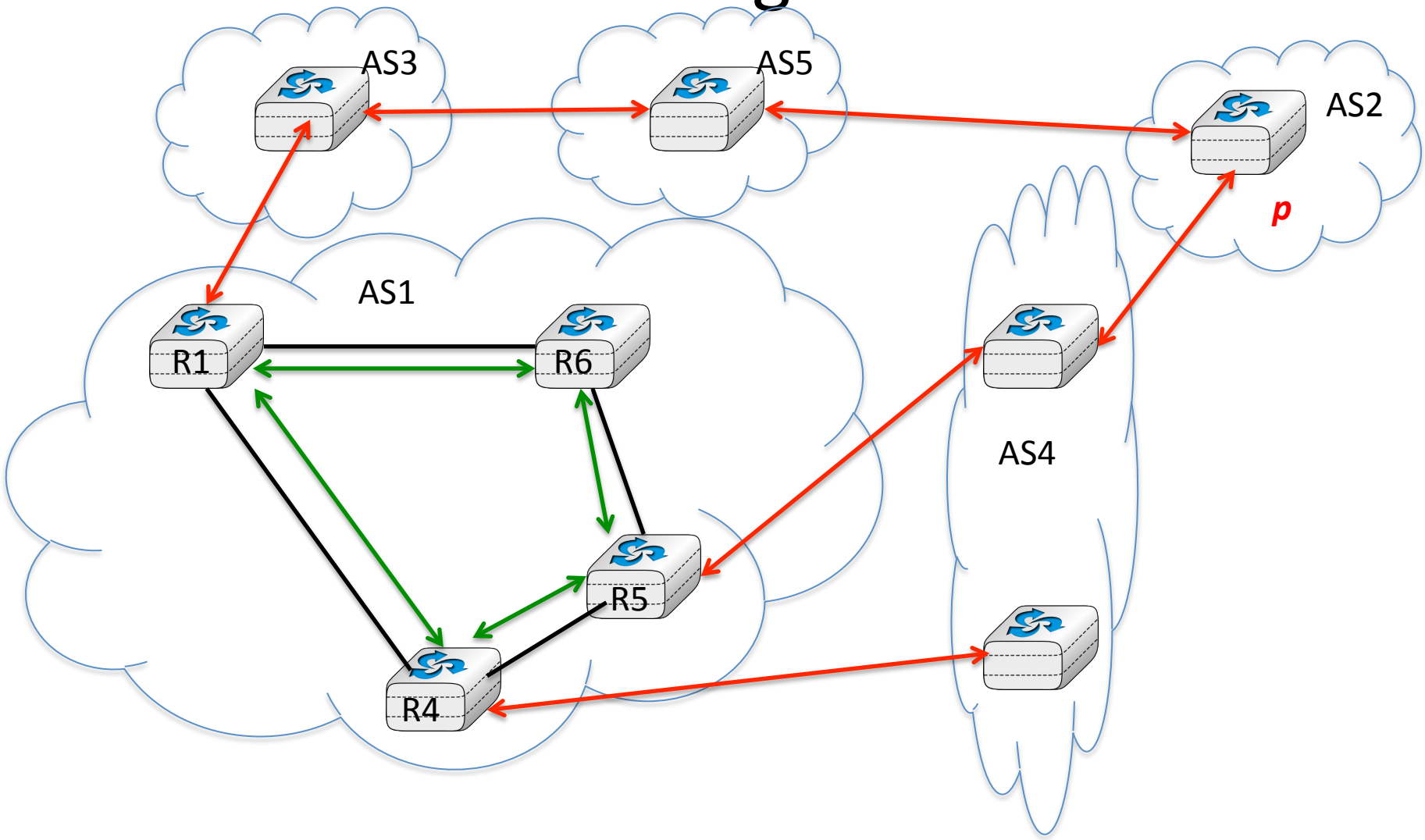
What happens if iBGP sessions are missing ?



What happens if iBGP sessions are missing ?



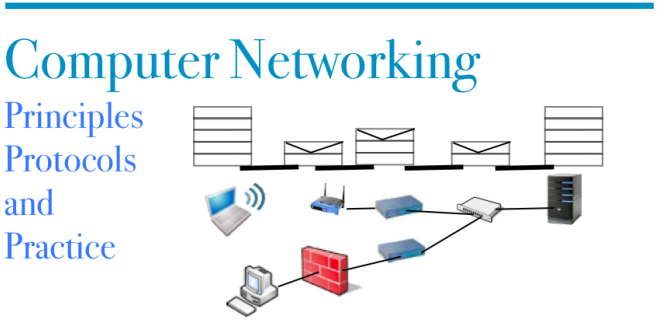
What happens if iBGP sessions are missing ?



Conclusion

- BGP depends on the underlying intradomain routing protocol
  - Establishment of the iBGP sessions
  - Resolution of the BGP nexthop
- iBGP and eBGP play different roles
  - eBGP over sessions with routers in other ASes
  - iBGP sessions (in full mesh) inside an AS
- The BGP decision process ranks routes
  - Hot potato versus cold potato routing

Reading list



1st edition, BGP section

http://cnp3book.info.ucl.ac.be/network/network/#the-border-gateway-protocol

J. Park et al, "*BGP Route Reflection Revisited*", IEEE Communications Magazine, June 2012, http://irl.cs.ucla.edu/~j13park/rr-commag.pdf

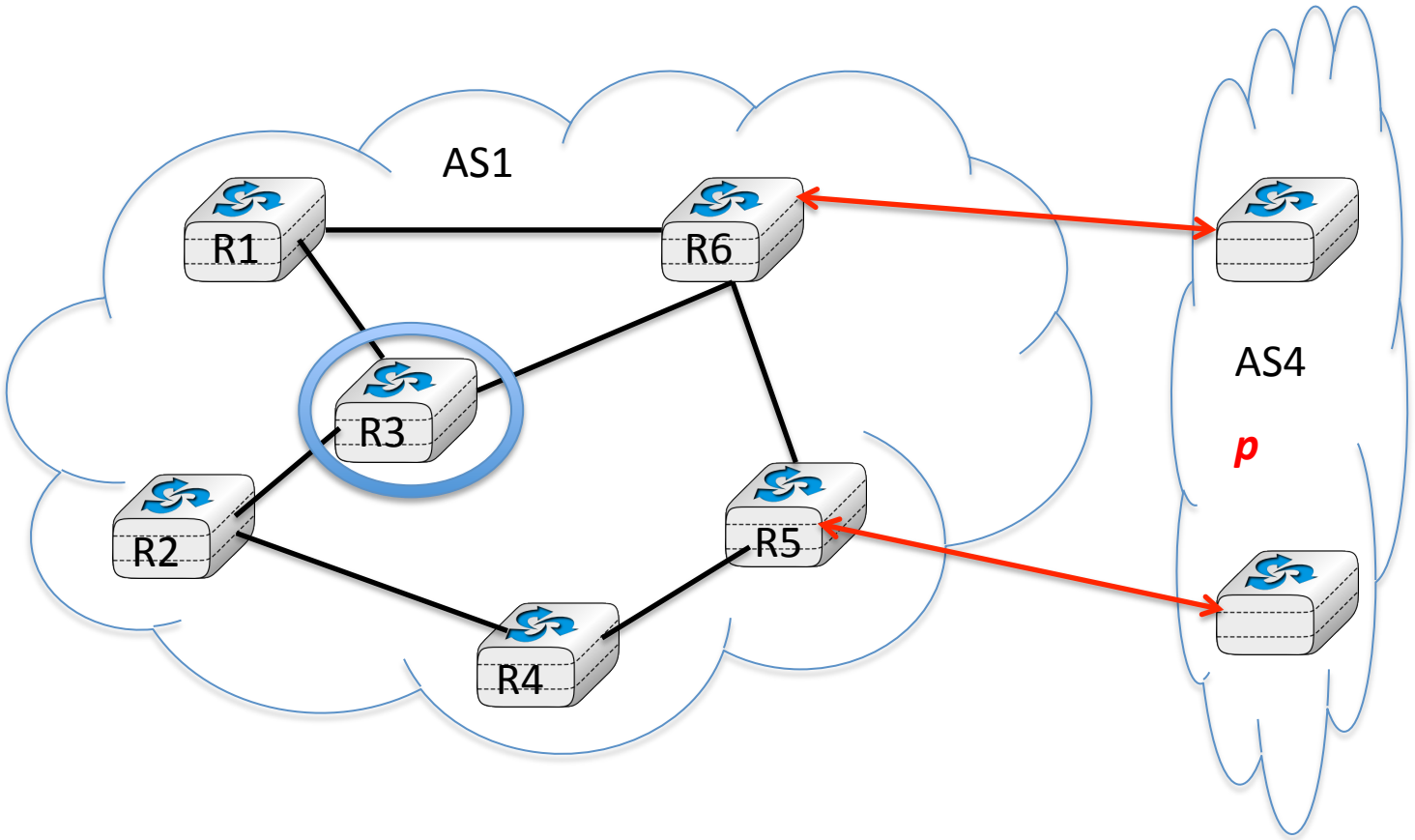
Route Reflectors

- What is a Route Reflector ?
  - Which iBGP attributes differ with RR ?
  - What is their role ?

- What are the advantages of using Route Reflectors compared to iBGP full mesh ?

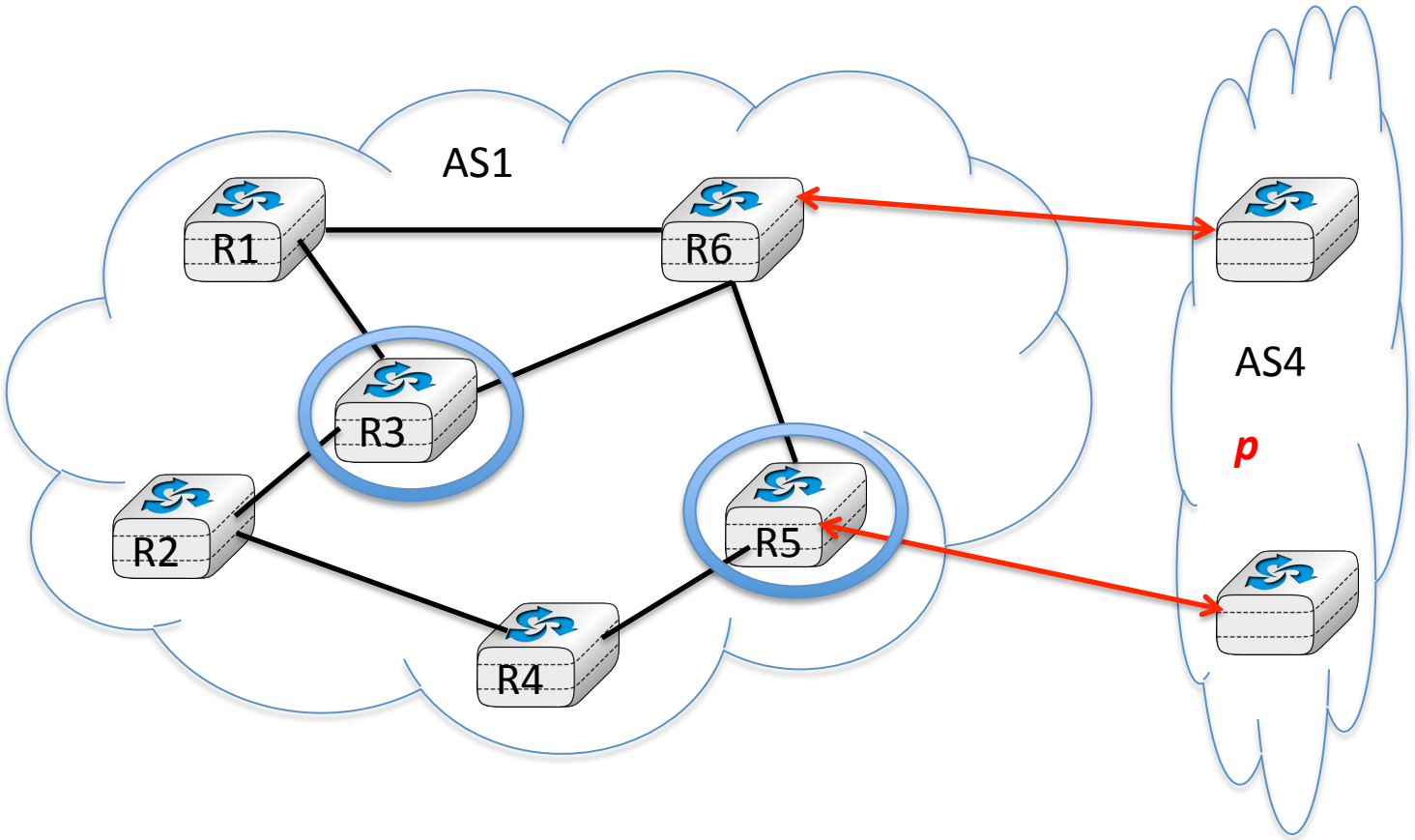
Route Reflectors

- What are the routes learned if R3 acts as RR for the entire AS1 ?

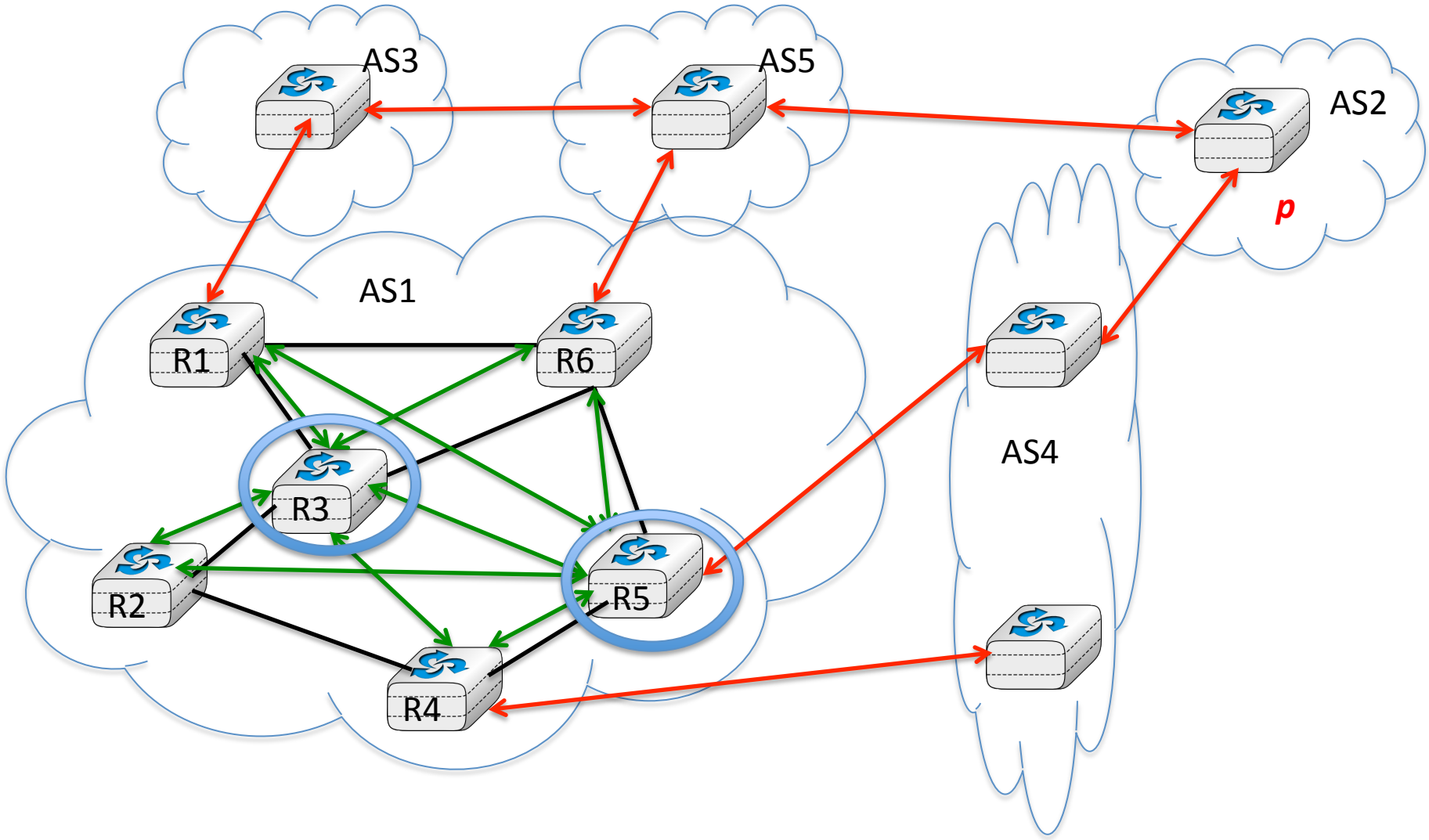


Route Reflectors

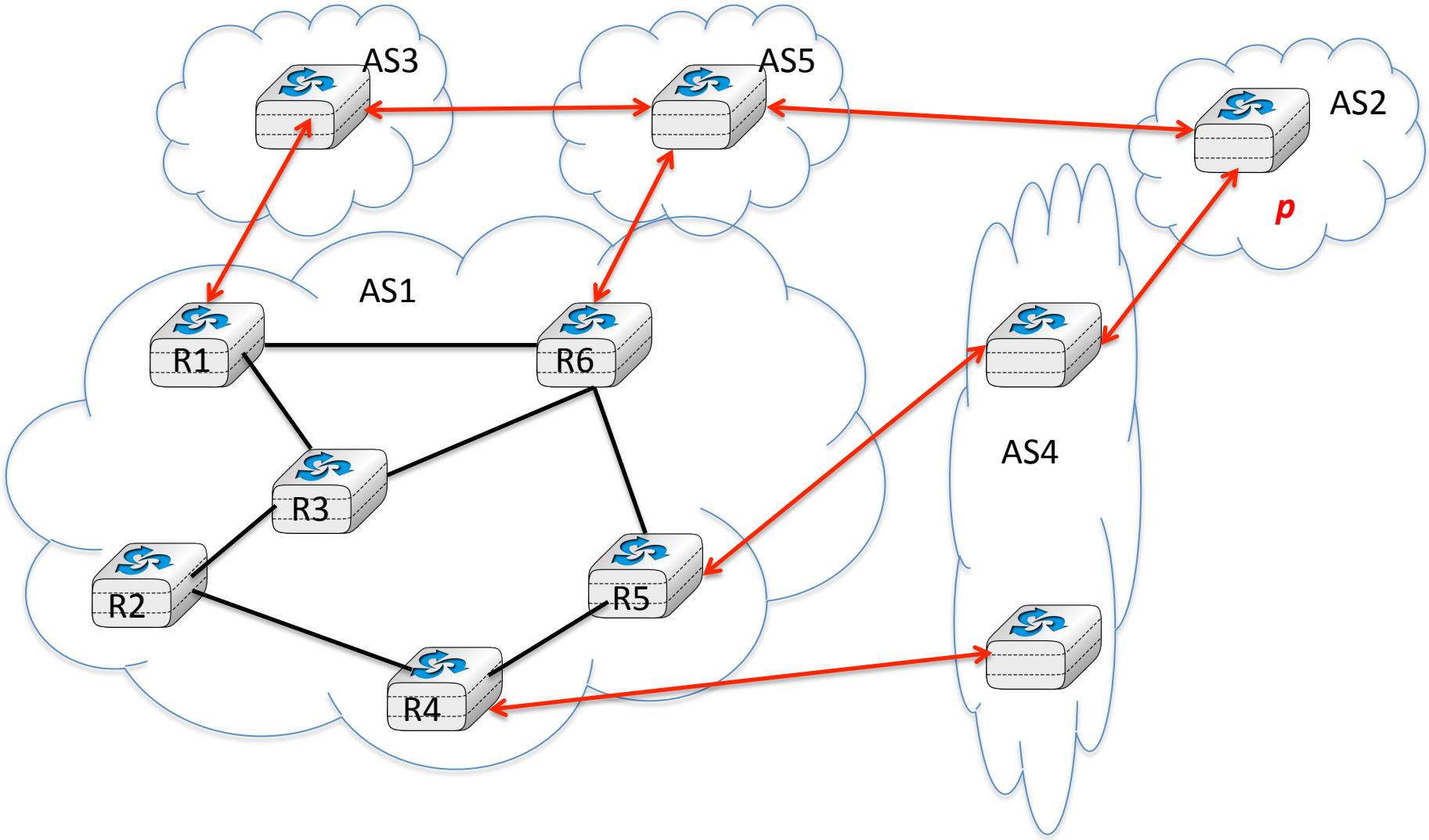
- What are the routes learned if both R3 and R5 act as RR for the entire AS1 ?



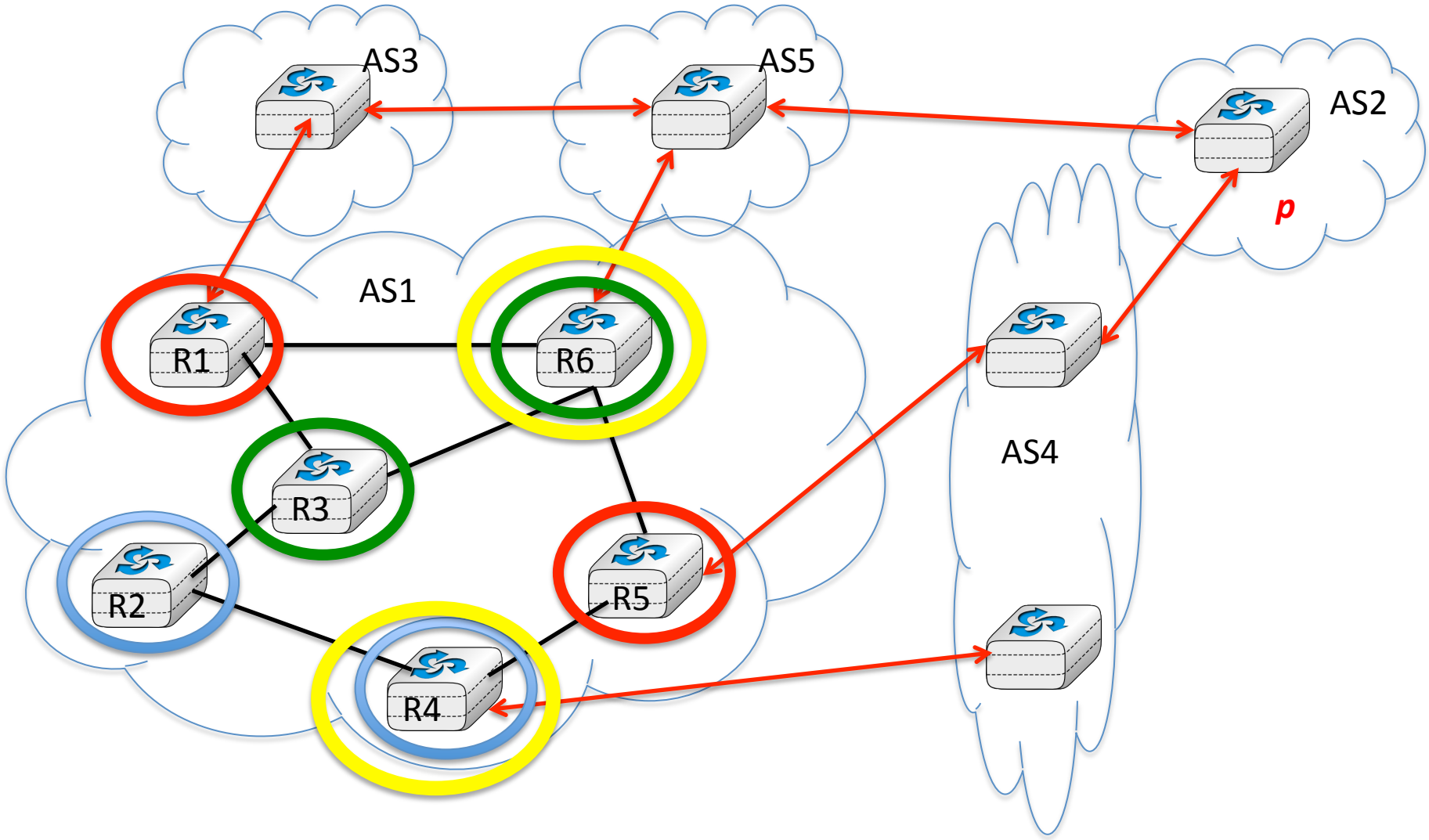
Which routes are selected ?



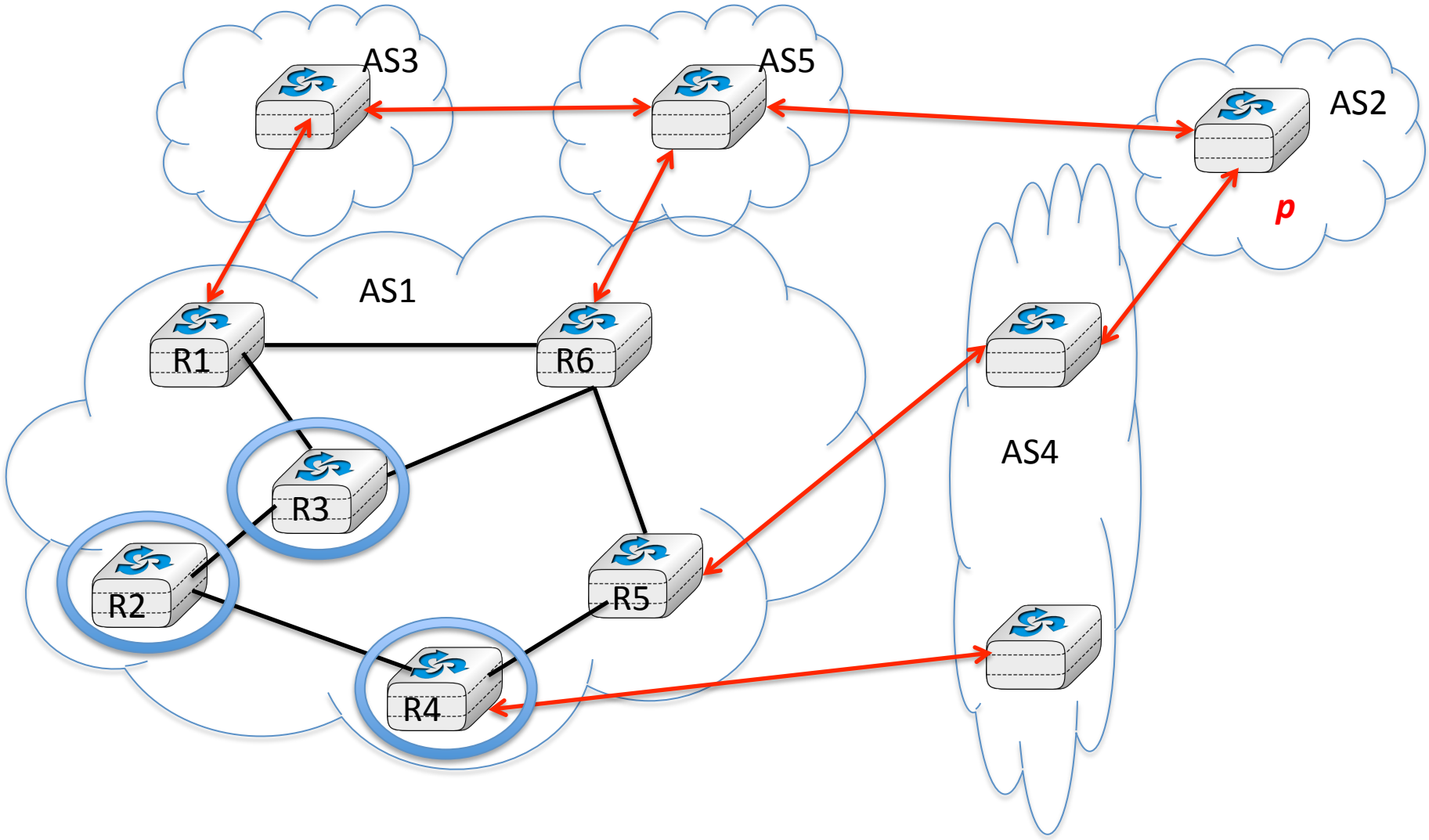
What is the best place for a single RR ?



What is the best location for two RR ?



Three Route Reflectors



Hierarchy of Route Reflectors