

# INGI2142

## Week 2 : BGP Route Reflectors and Traffic Engineering

O. Bonaventure, 2016

# Agenda

- How to scale iBGP to large networks ?
- BGP traffic engineering

# Interactions between IGP and iBGP

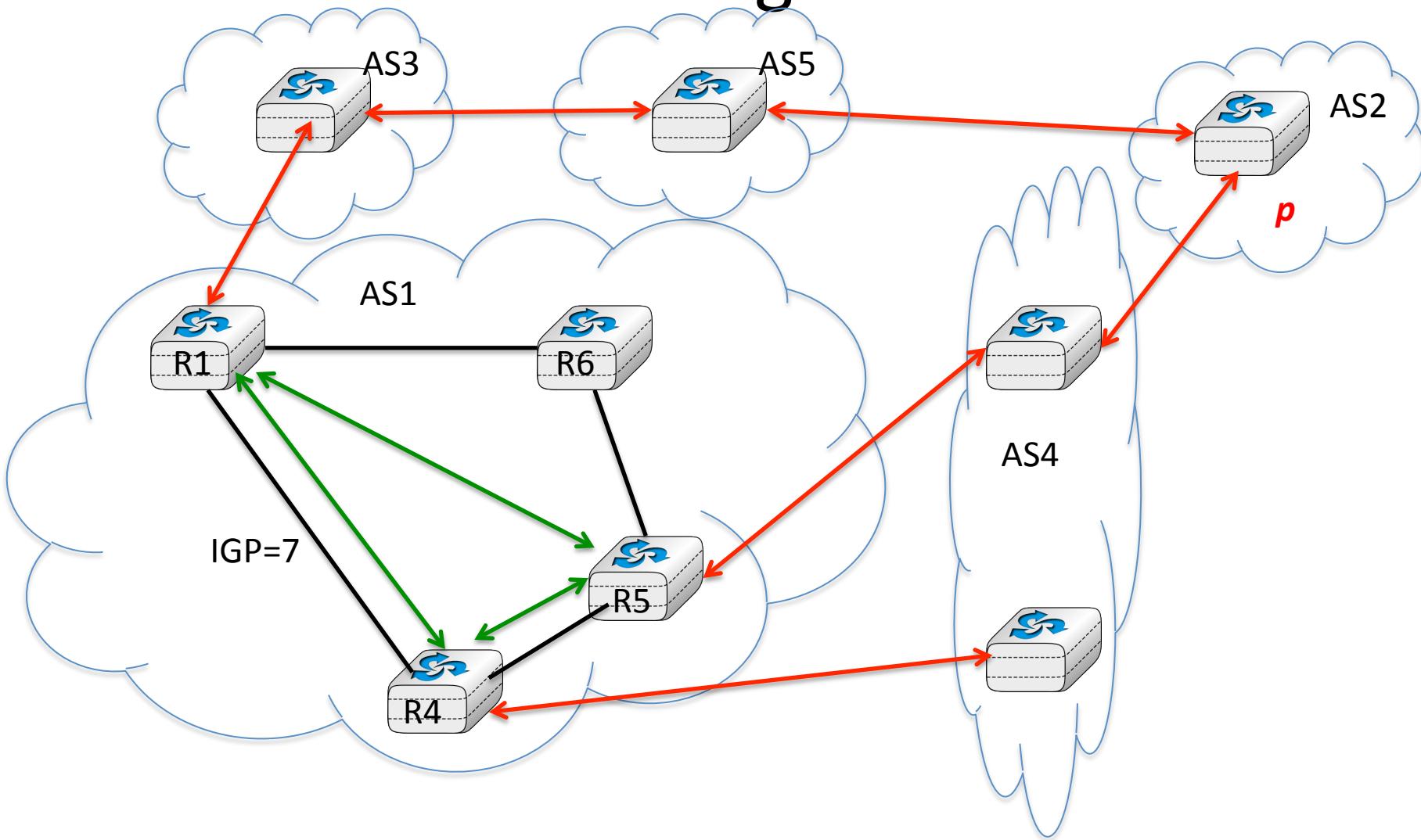
- What are the interactions between iBGP and the intradomain routing protocol ?
  - iBGP sessions are TCP connections whose endpoints are reachable thanks to IGP
    - Endpoints of iBGP sessions are usually loopback interfaces advertised in IGP
  - BGP Nexthops are reachable thanks to IGP
  - BGP decision process uses reachability and IGP cost towards BGP nexthop to rank routes

# Creation of iBGP sessions

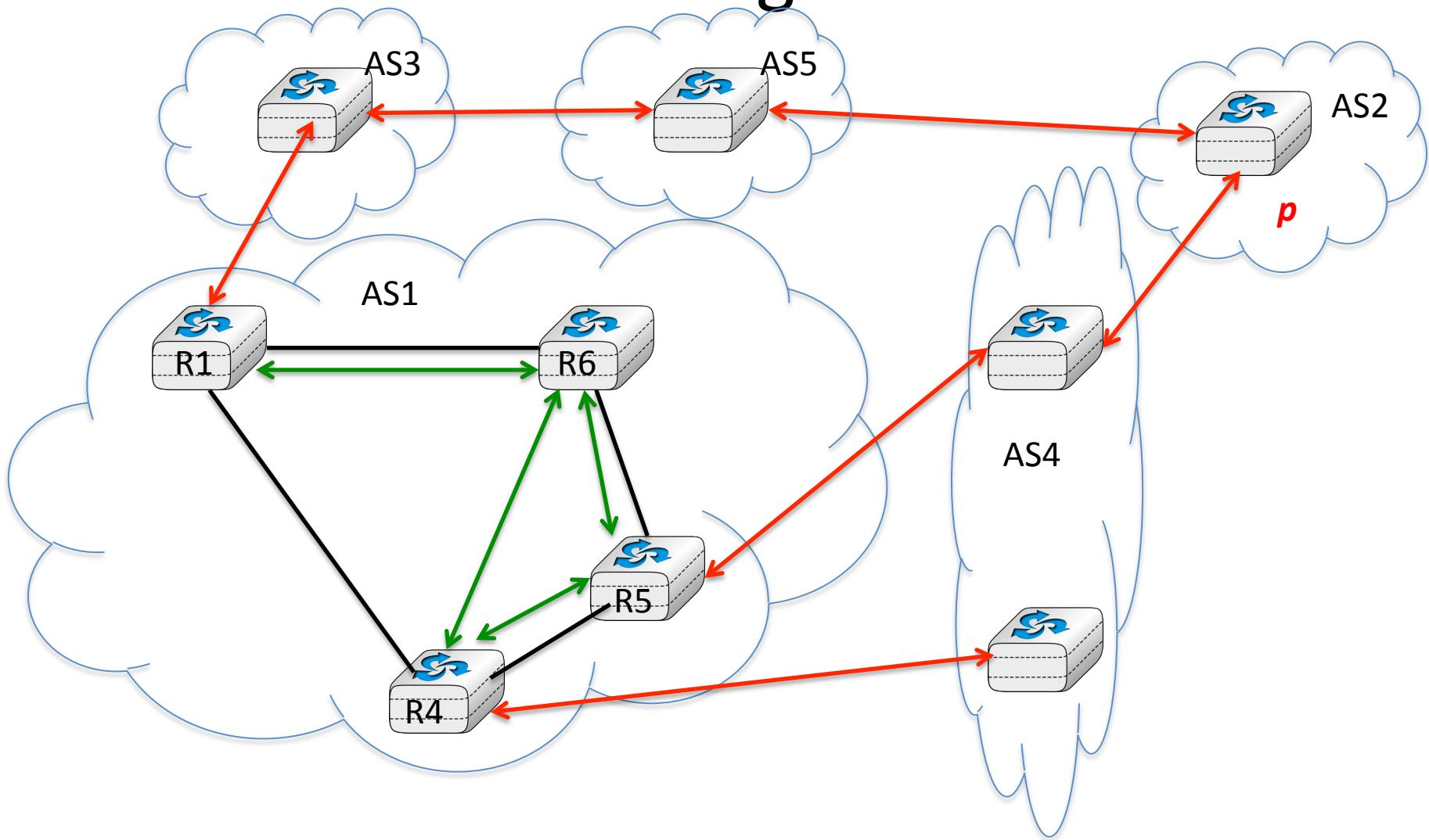
- How are iBGP sessions created on routers ?
  - Usually by manual configuration on each router

```
group INTERNET2-IPv6 {  
    type internal;  
    local-address 2001:468:a::1;  
    family inet6 {  
        any;  
    }  
    export NEXT-HOP-SELF;  
    peer-as 11537;  
    neighbor 2001:468:1::1 {  
        description ATLA;  
    }  
    ...
```

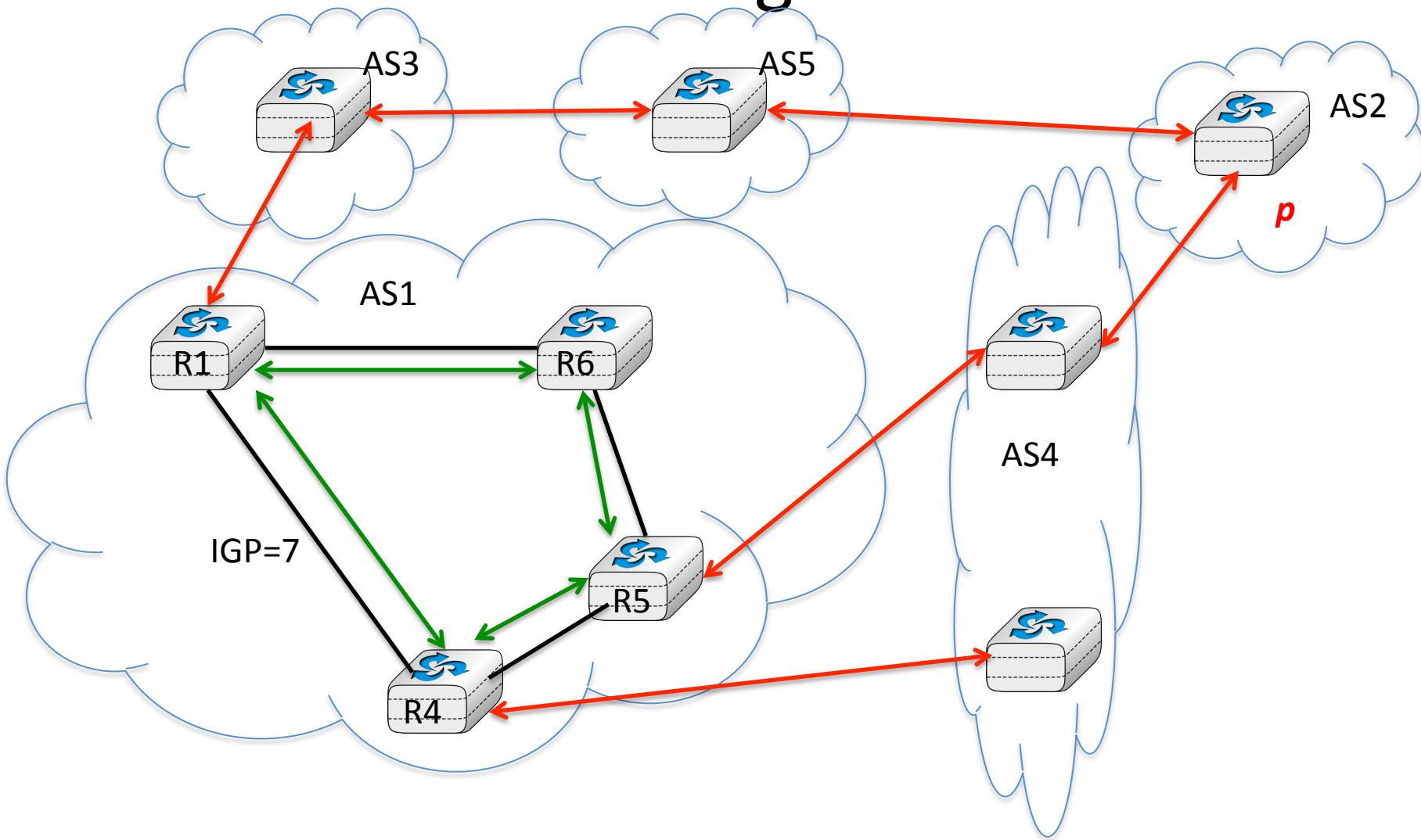
# What happens if iBGP sessions are missing ?



# What happens if iBGP sessions are missing ?



# What happens if iBGP sessions are missing ?



# Scaling issues with iBGP

- In a network containing N routers
  - $N*(N-1)/2$  iBGP sessions need to be manually configured and maintained
- Scalability issues
  - CPU usage to process and send iBGP messages
    - About 600k routes on IPv4 Internet today
  - Number of iBGP sessions on each router
    - TCP state, BGP Keepalives, ...
  - Memory consumption
    - ADJ-RIB-IN, ADJ-RIB-OUT, ...

# Improving iBGP scaling

- Two approaches
  - Route Reflectors
    - A RR is a special iBGP router that is allowed, under specific conditions, to advertise over iBGP sessions routes learned over other iBGP sessions
  - BGP confederations
    - A large AS is divided in smaller (sub-)ASes containing a few tens of routers in iBGP full mesh. The sub-ASes use eBGP to exchange BGP routes

# Route Reflectors

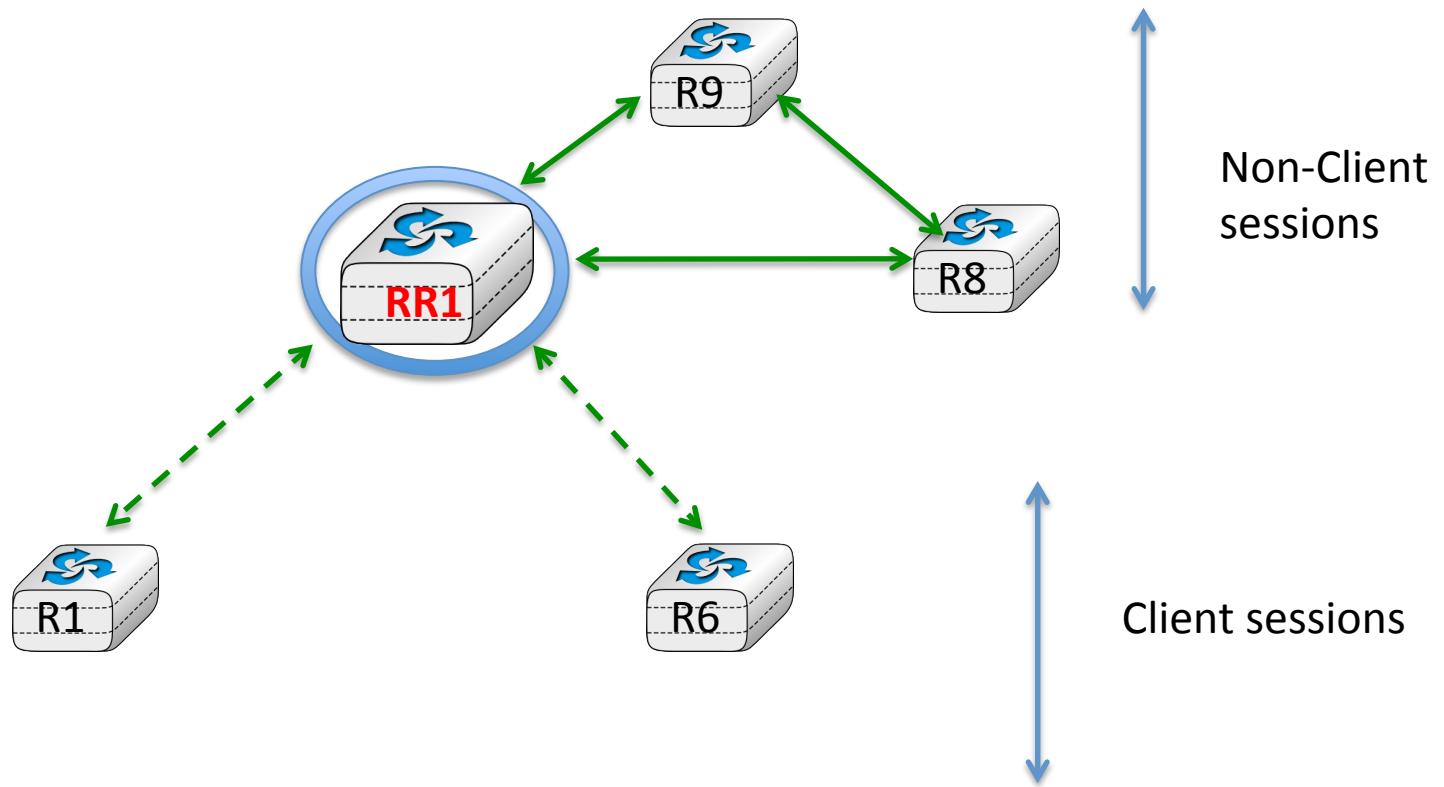
- What is a Route Reflector ?
  - Which iBGP attributes differ with RR ?
  - What is their role ?
- What are the advantages of using Route Reflectors compared to iBGP full mesh ?

# A BGP Route Reflector

- A RR has two types of iBGP neighbours
  - Clients
    - The RR learns routes from its clients, runs the BGP decision process and advertises its best route over all its peers
  - Peers

# How to design an iBGP hierarchy

- A RR has two types of iBGP neighbours



# Operation of Route Reflectors

- Reception of a new route
  - Run BGP decision process on RR
  - If best route has changed
    - If best route was learned from **eBGP** session
      - Advertise the best route to all **iBGP** sessions
    - If best route was learned from an **iBGP** client session
      - Advertise the best route to all **iBGP** sessions (clients and non-clients)
    - If best route was learned from a non-client **iBGP** session
      - Advertise the best route to all client **iBGP** sessions (non-client sessions are assumed to be in full-mesh and will also receive the new route from their own iBGP session)

# Benefits of using RRs

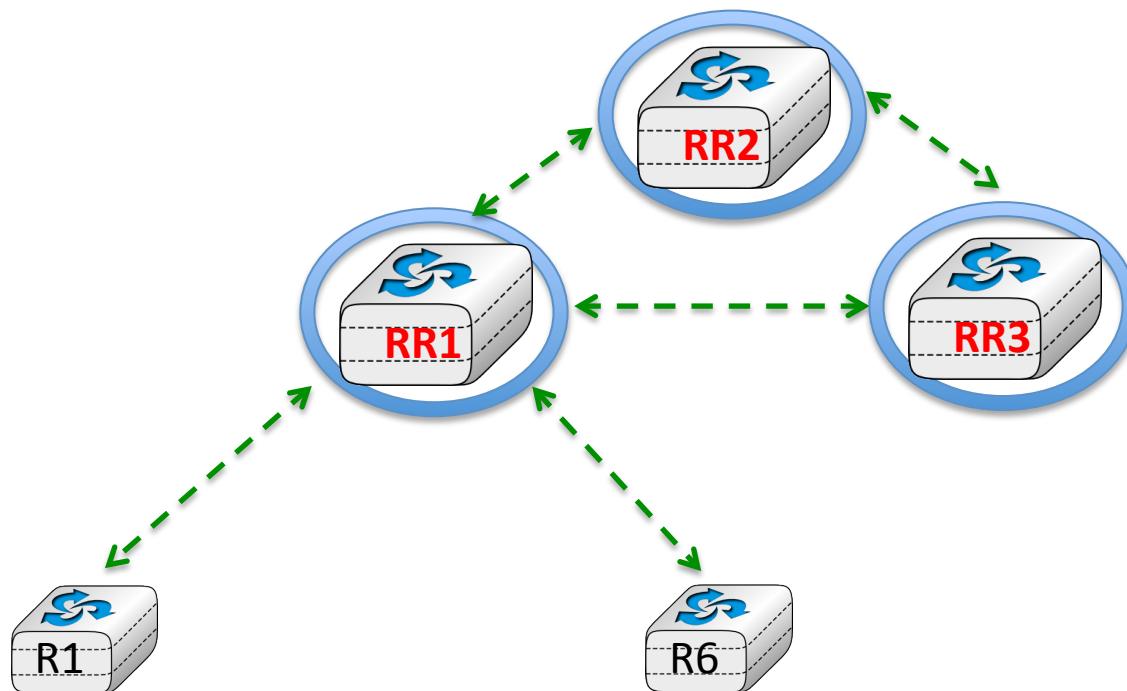
- Simplified configuration
  - A new router can be easily added to the network
- Reduced memory and CPU usage on routers
  - Each router maintains fewer iBGP sessions
  - Route Reflectors do not announce all routes to their BGP clients, reducing their memory usage
  - Route Reflectors are often specialised devices with faster CPU and memory that only run BGP and IGP but do not forward regular packets

# Caveats with Route Reflectors

- Route reflectors hide routes to iBGP clients
  - A RR only advertises its best route towards each prefix over a given iBGP session
  - A RR runs the BGP decision process on the basis of its IGP routing table
    - iBGP clients could select a different best route than their RR
- Route reflectors can increase convergence time after failures

# Caveats with Route Reflectors

- Since RRs advertise iBGP learned routes over iBGP sessions, a badly configured iBGP topology may cause loops



# How to prevent loops

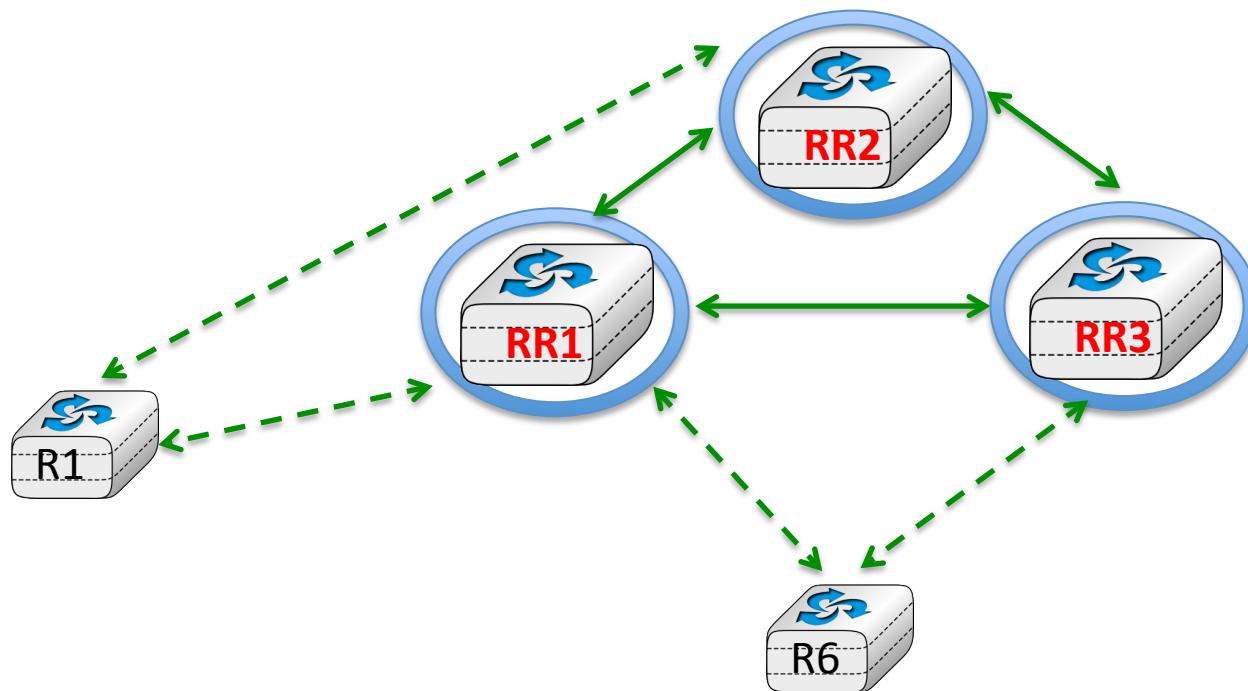
- BGP Route Reflection introduces two new iBGP attributes
  - ORIGINATOR\_ID
    - Set to the router id of the router that injects a route in iBGP
  - CLUSTER\_LIST
    - When a RR receives a route, it checks whether its router id is included in the CLUSTER\_LIST. If yes, the route is rejected.
    - When a RR sends a route over an iBGP session, it adds its router id to the CLUSTER\_LIST

# Fault tolerance

- If a Route Reflector fails, a large number of BGP routers will be affected and lose BGP routes
- How to mitigate this problem ?
  - Use redundant route reflectors that are deployed in pairs
  - Each BGP router is attached to at least two different RRs

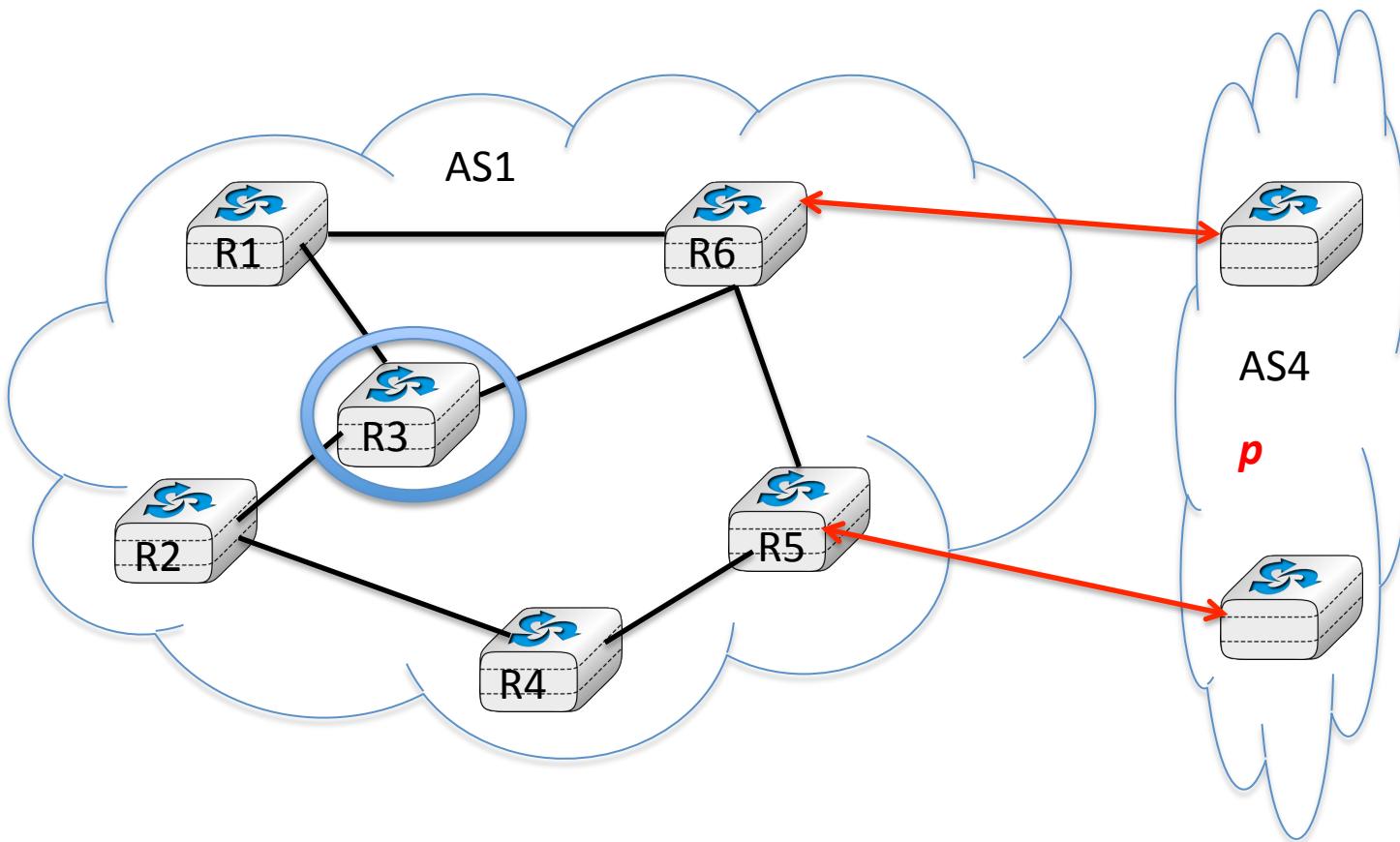
# Fault tolerance

- In practice, each BGP router is usually attached to two RRs that are close to itself in the IGP topology



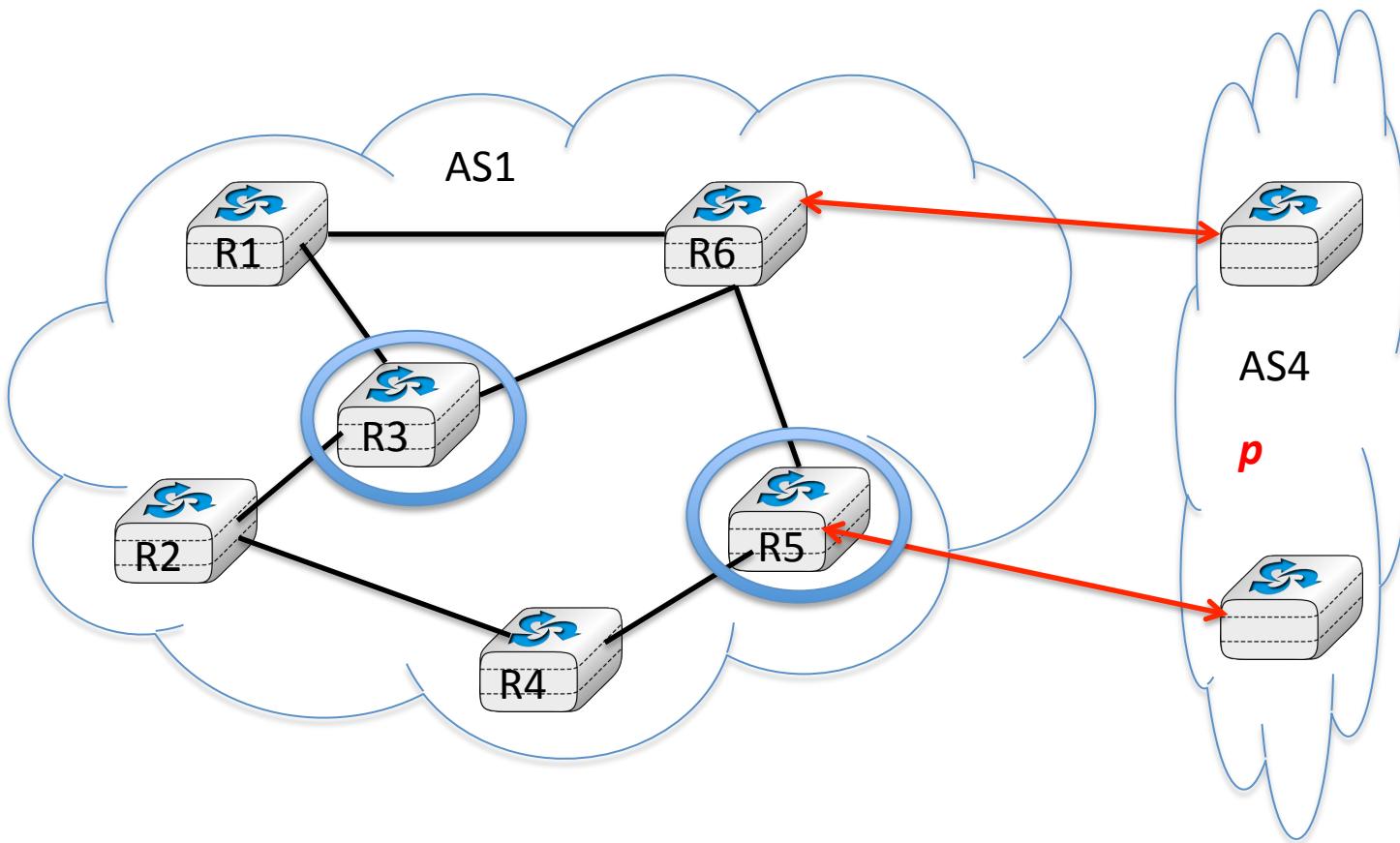
# Route Reflectors

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

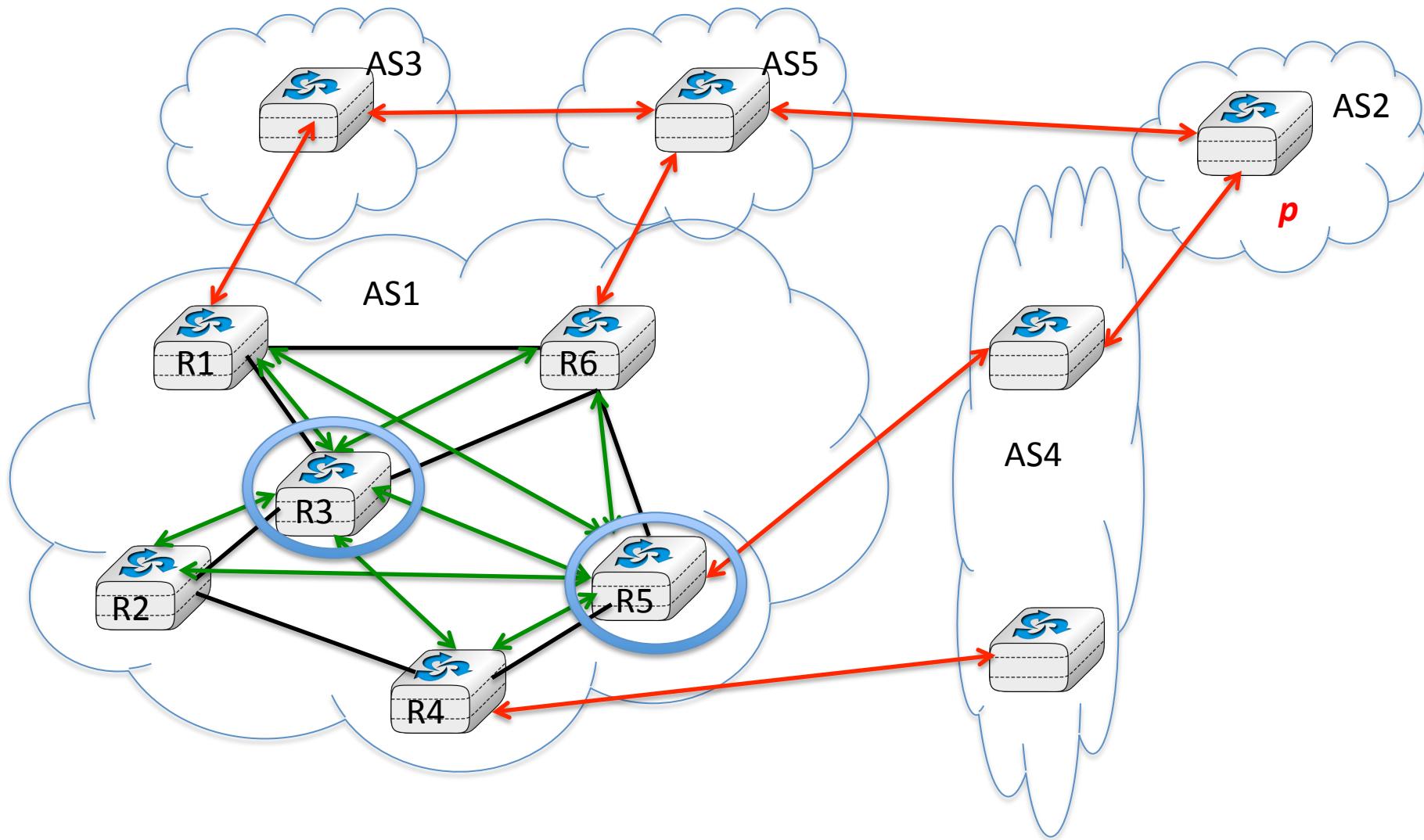


# Route Reflectors

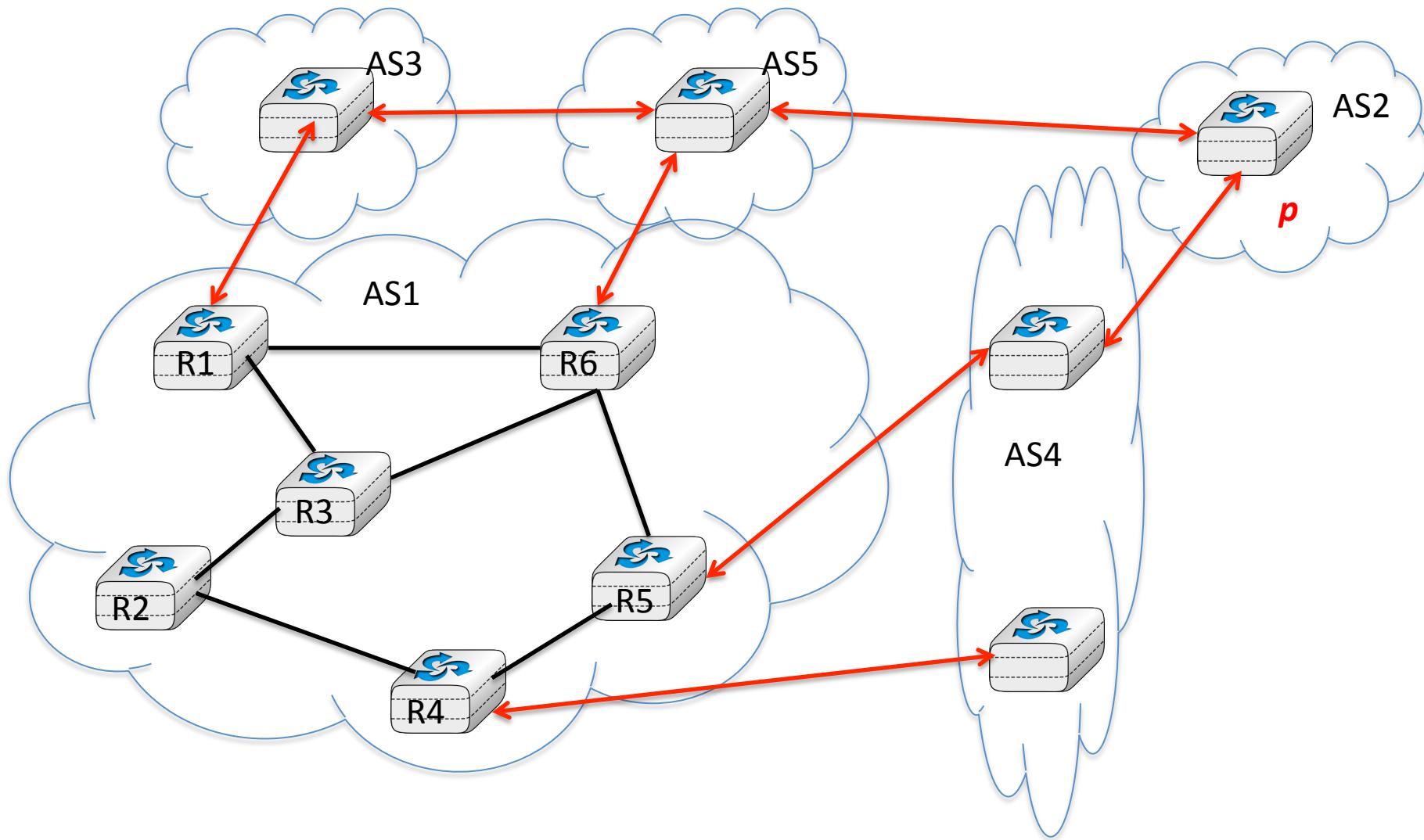
- What are the routes learned if both R3 and R5 act as RR for the entire AS1 ?



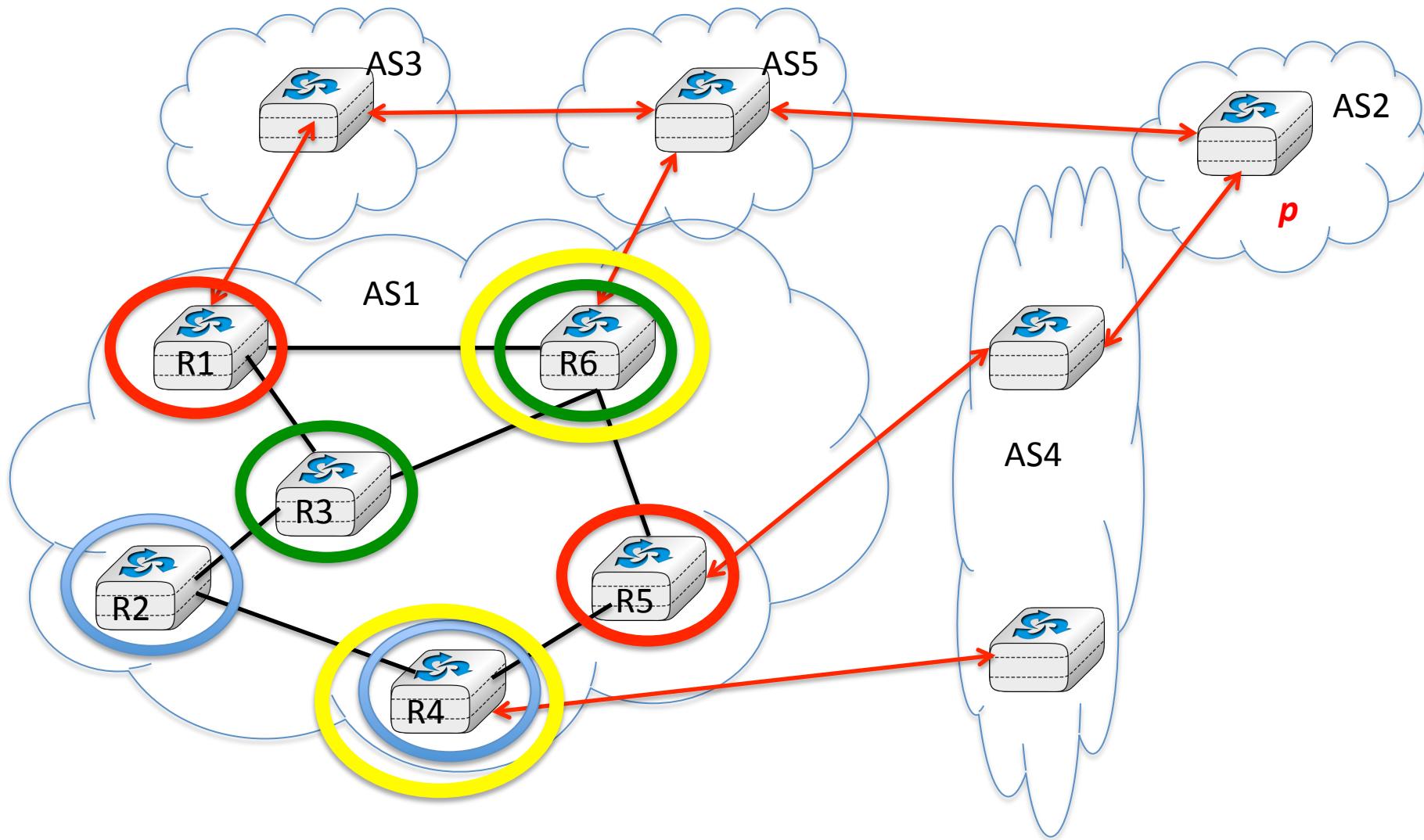
# Which routes are selected ?



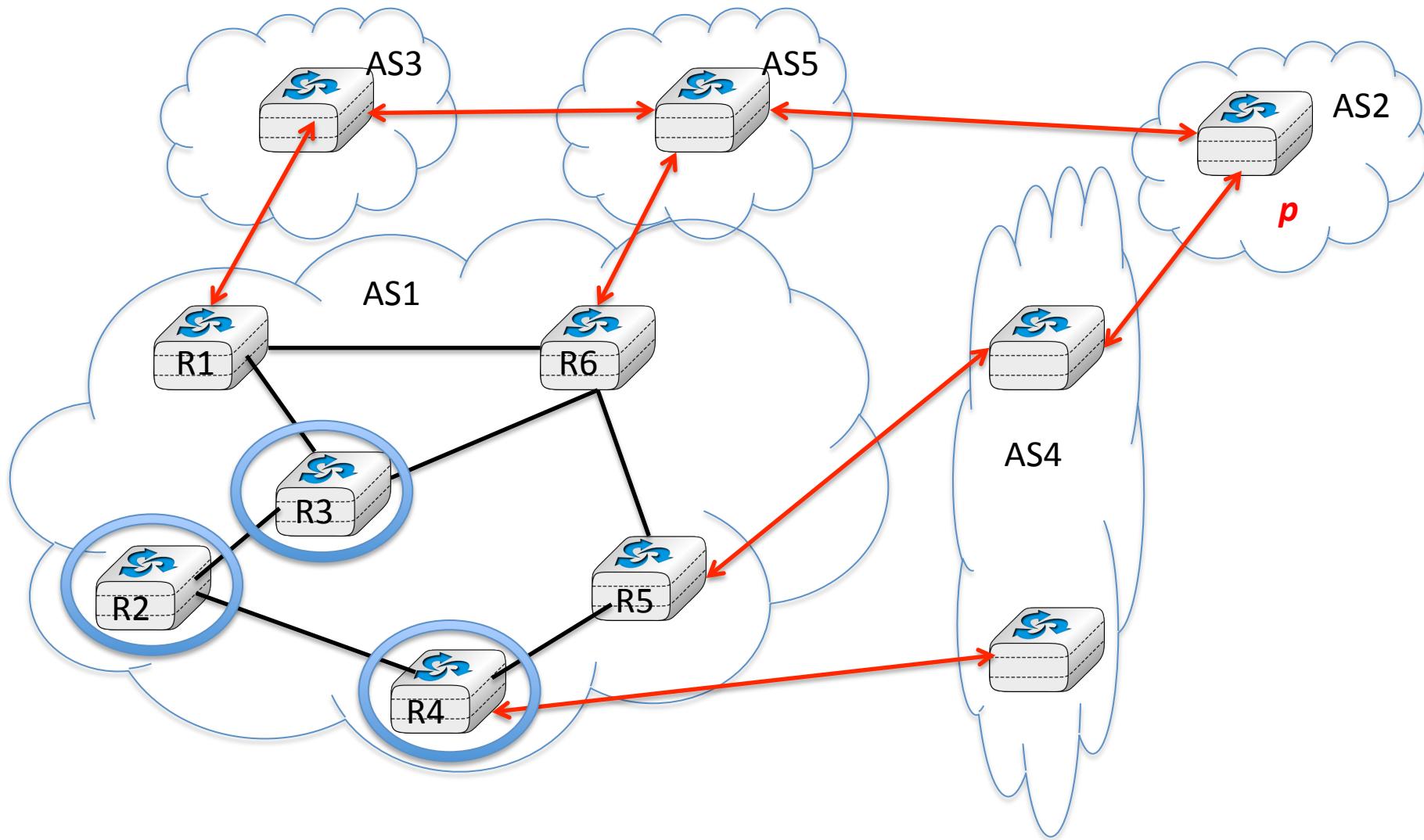
# What is the best place for a single RR ?



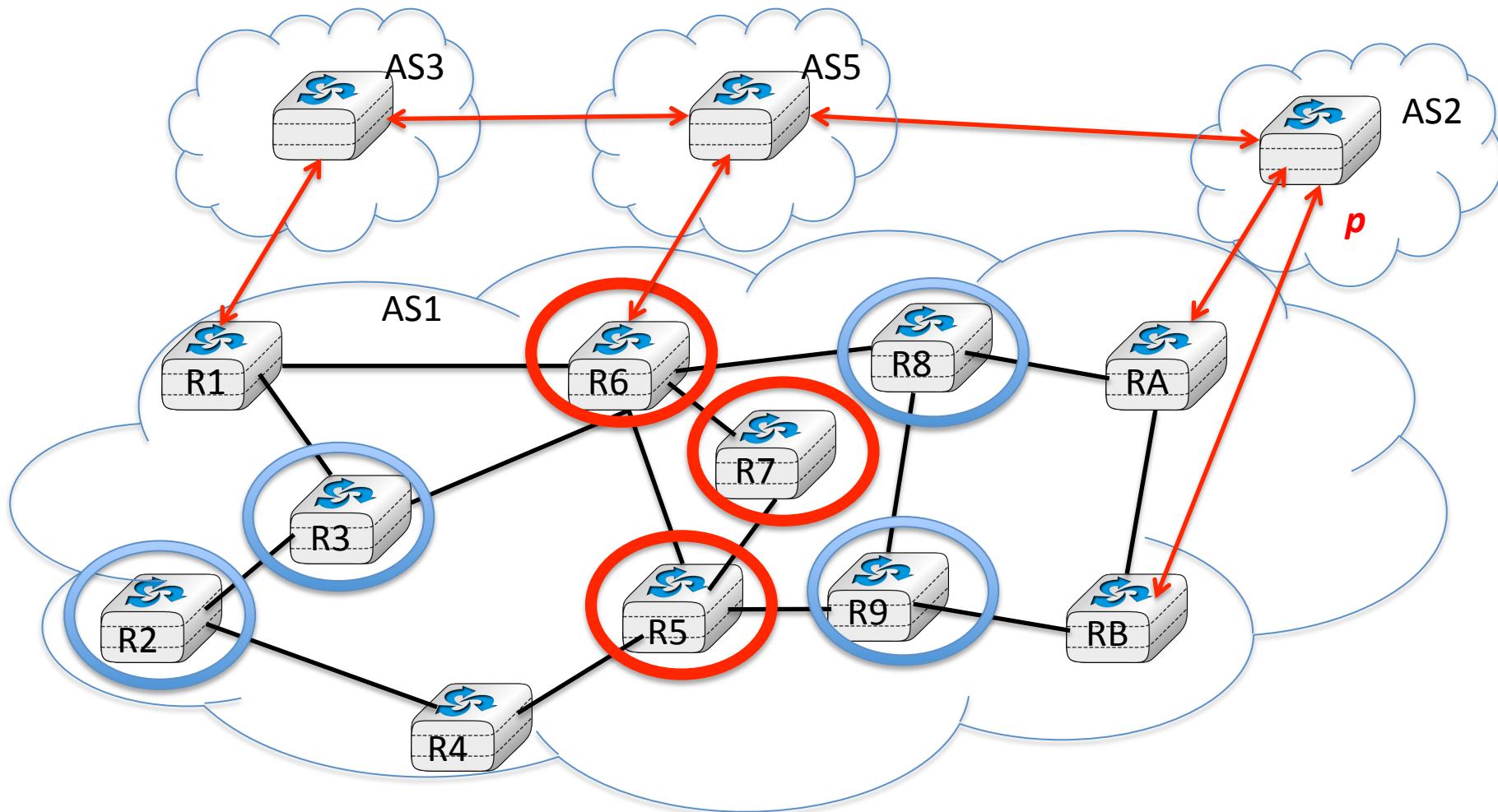
# What is the best location for two RR ?



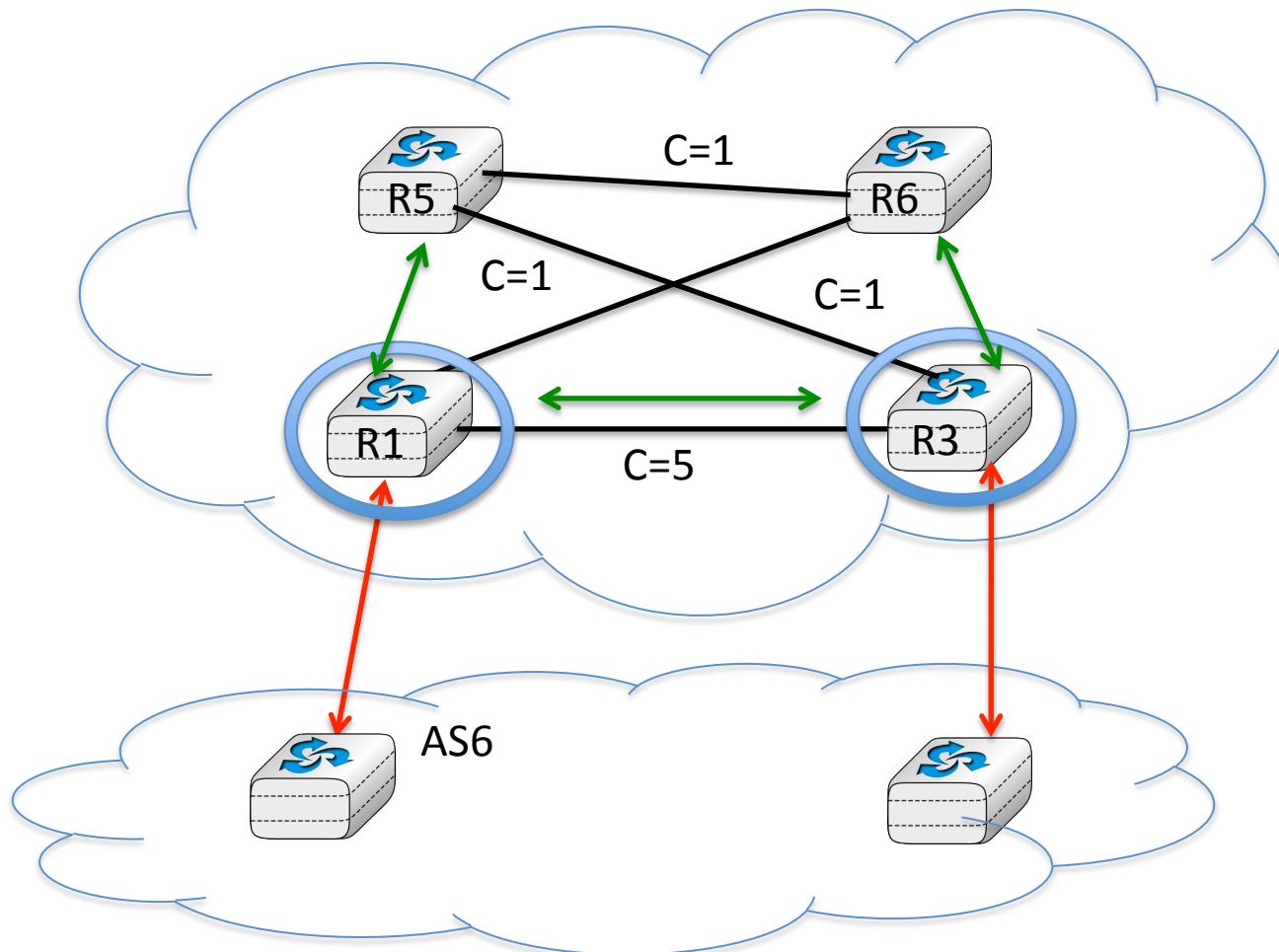
# Three Route Reflectors



# Hierarchy of Route Reflectors



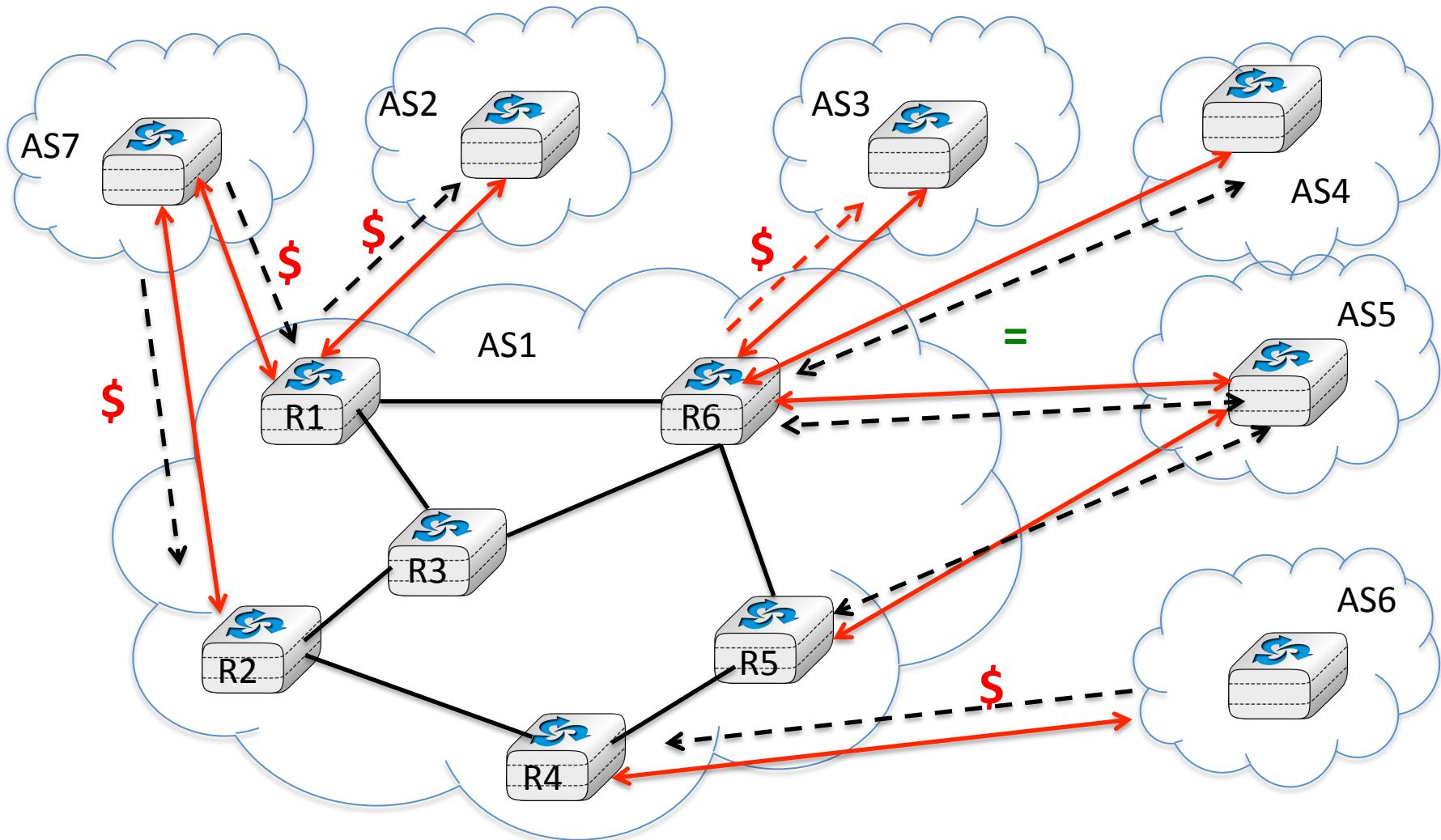
# What are the forwarding paths towards p advertised by AS6 ?



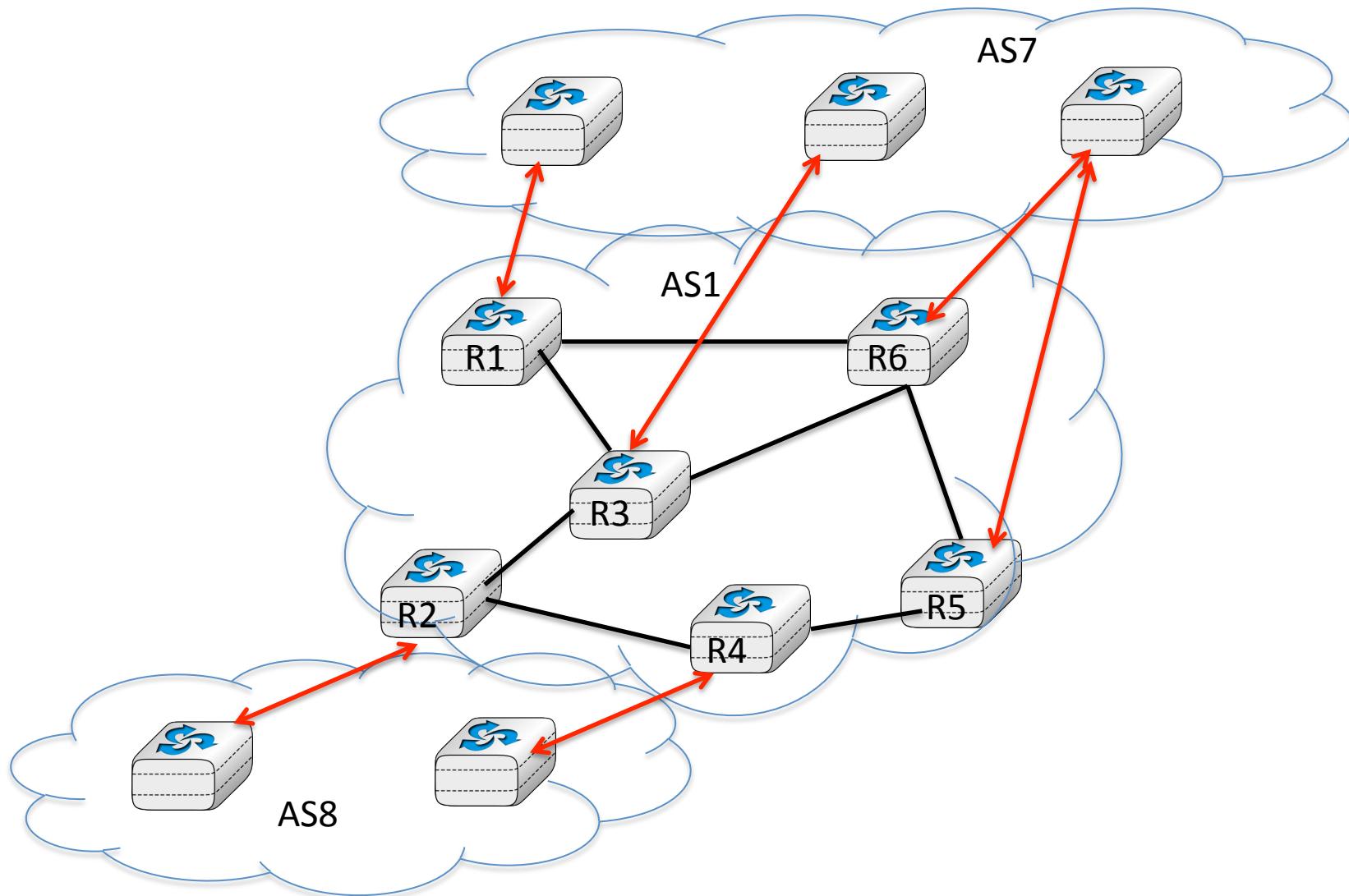
# Agenda

- How to scale iBGP to large networks ?
- BGP traffic engineering

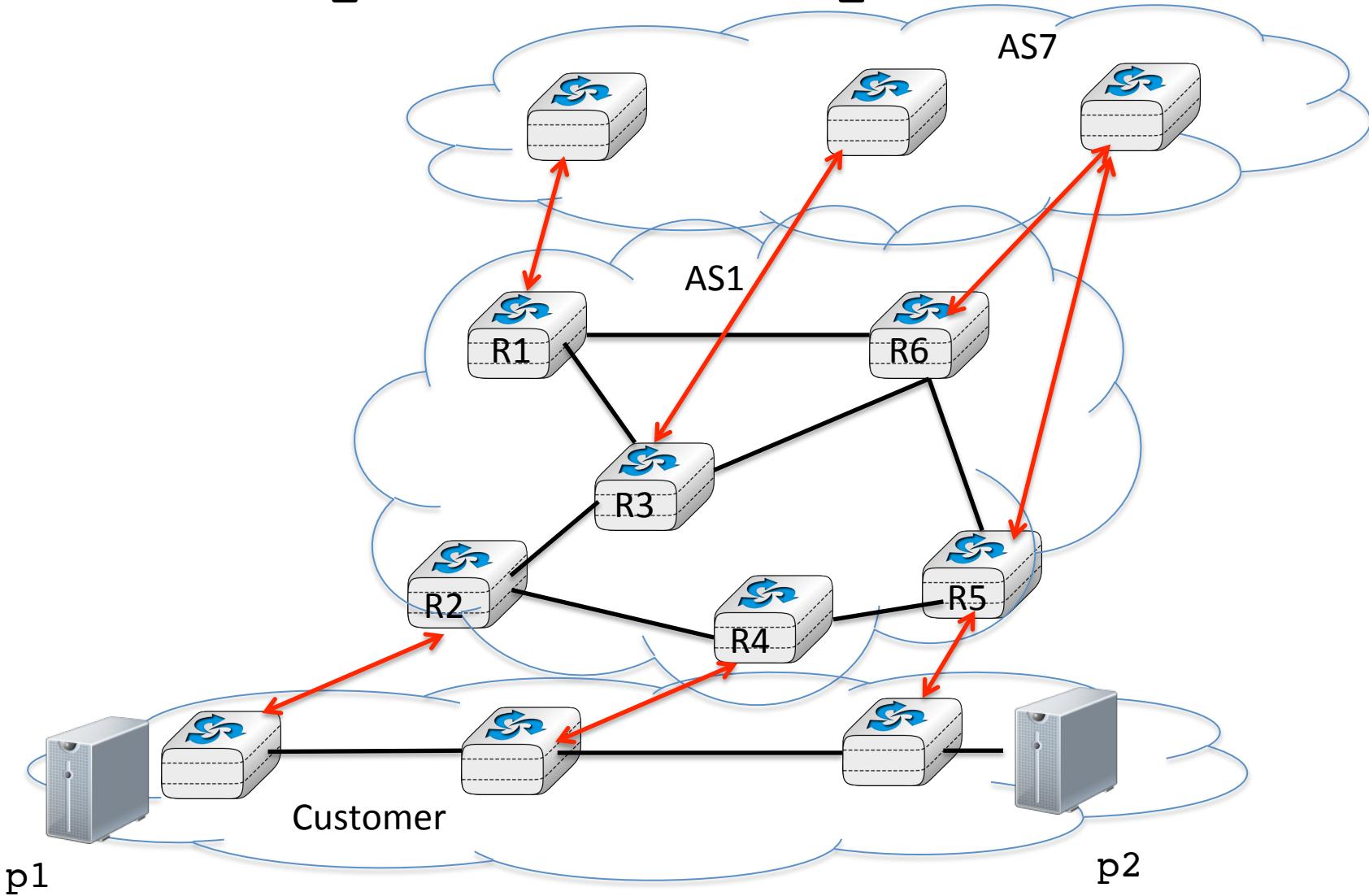
# Prefer customer, ...



# Forwarding paths from AS7 to AS8 and vice-versa

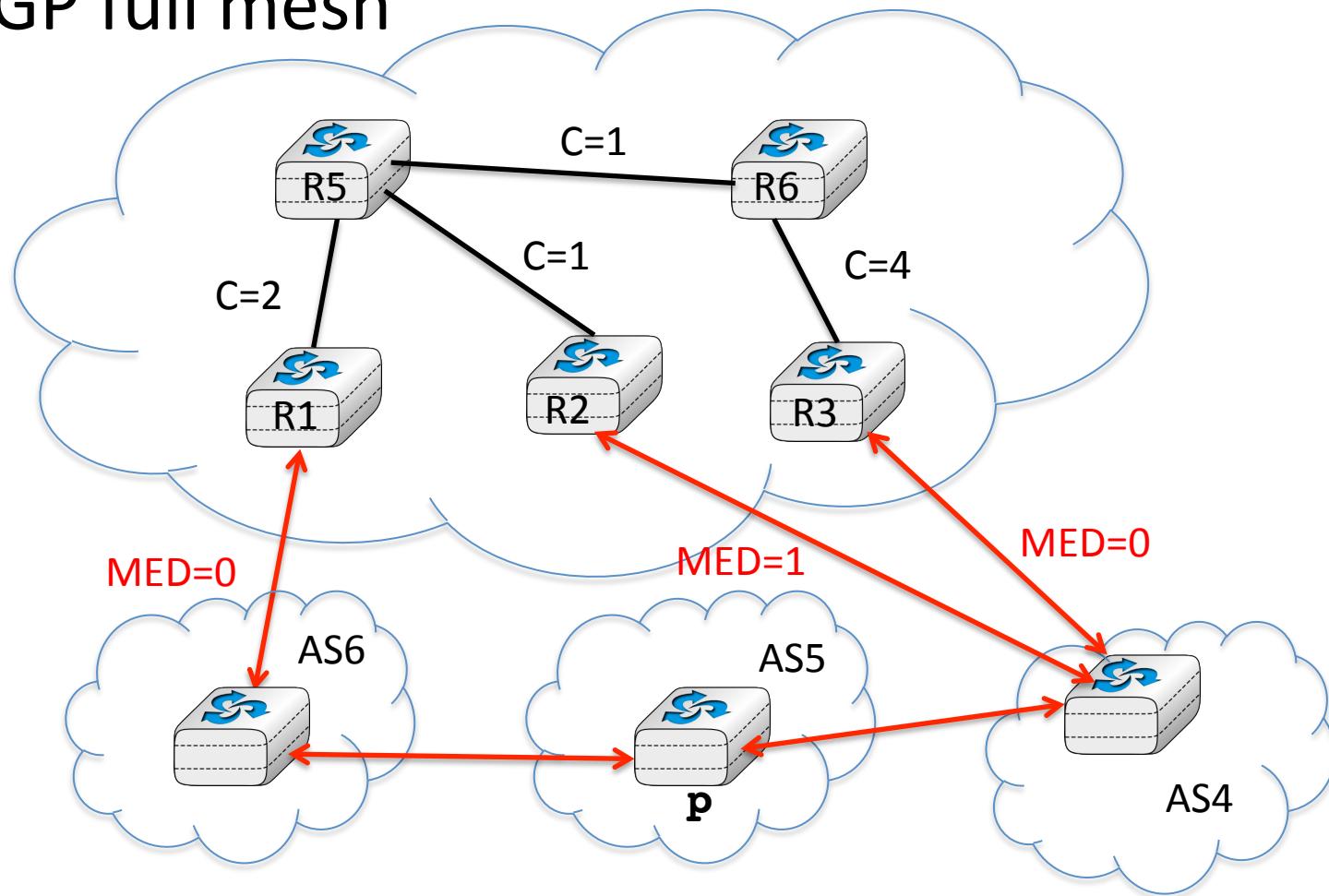


Customer wants packets towards  
p1 via R2 and p2 via R5



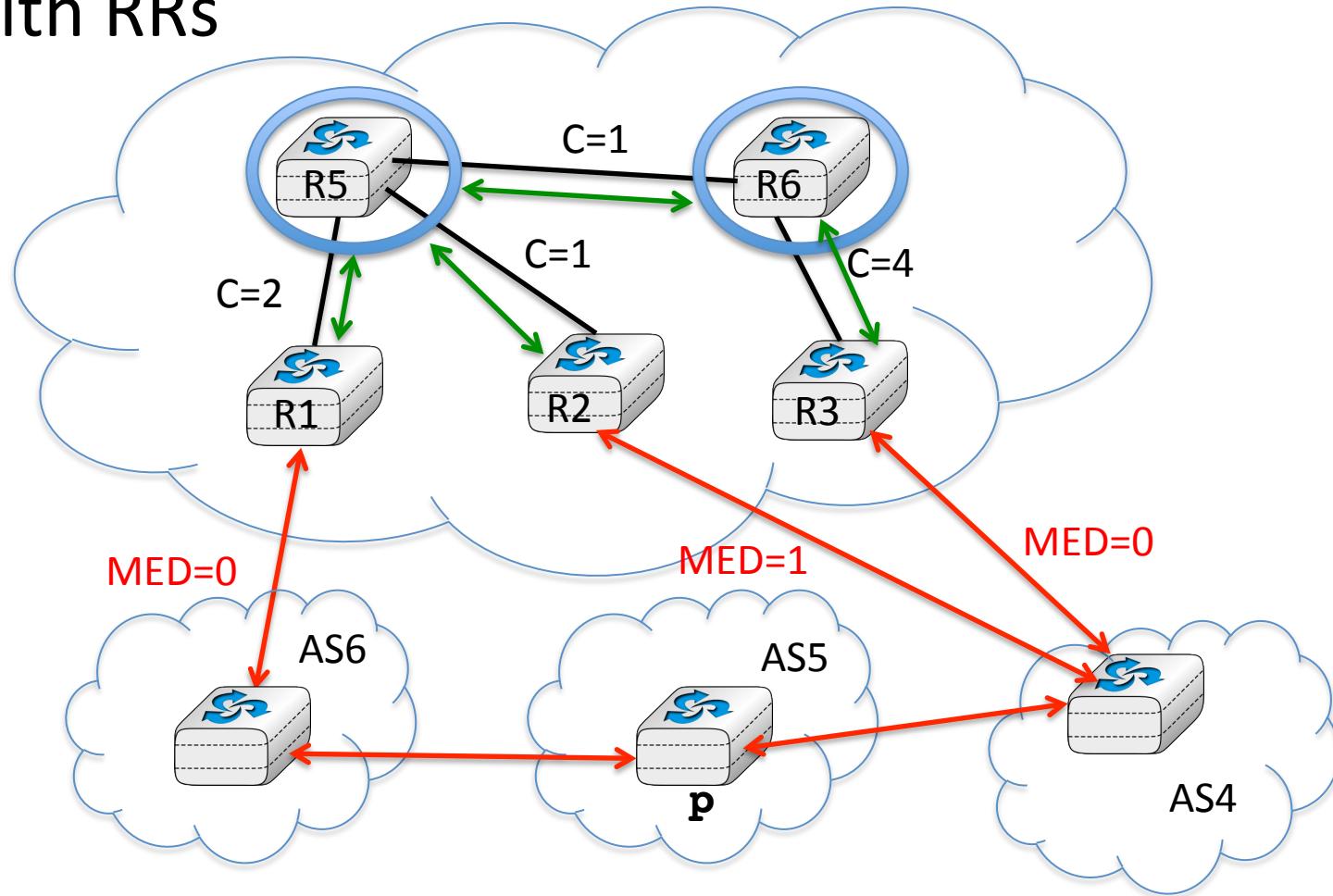
# Fun with MED

- iBGP full mesh



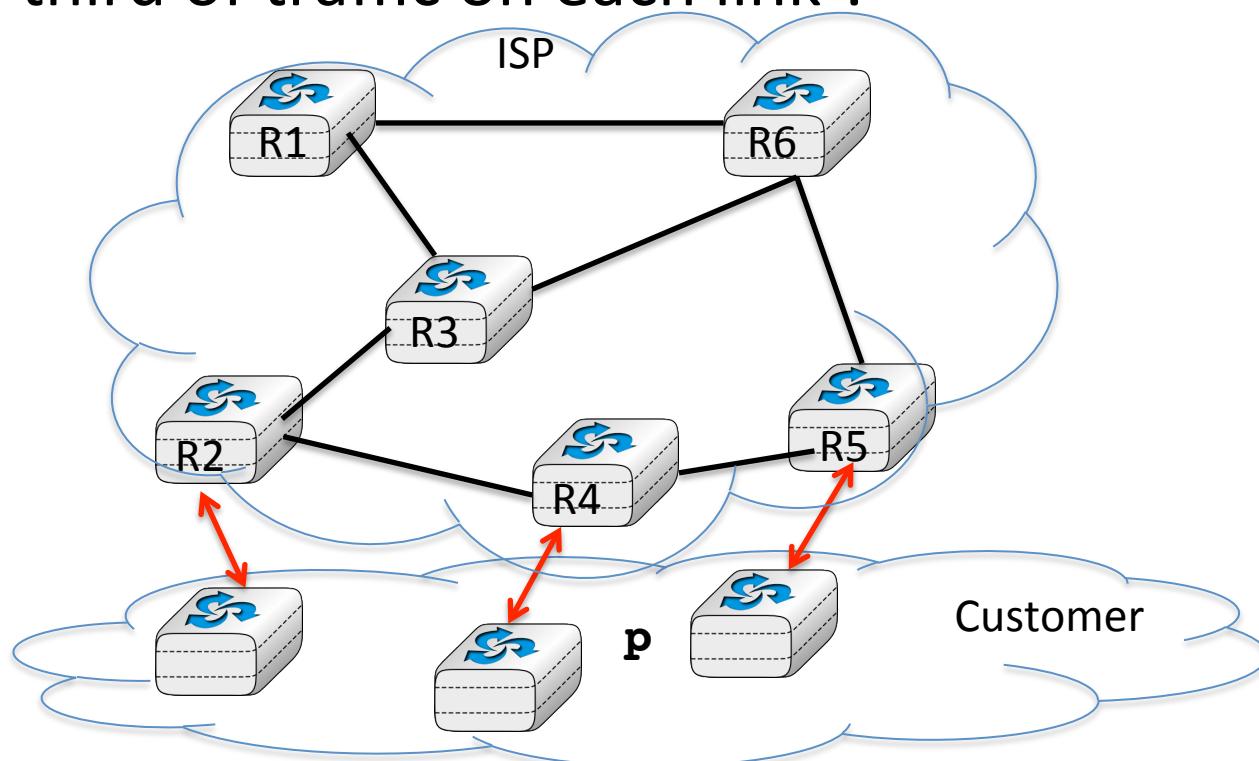
# Fun with MED (2)

- With RRs

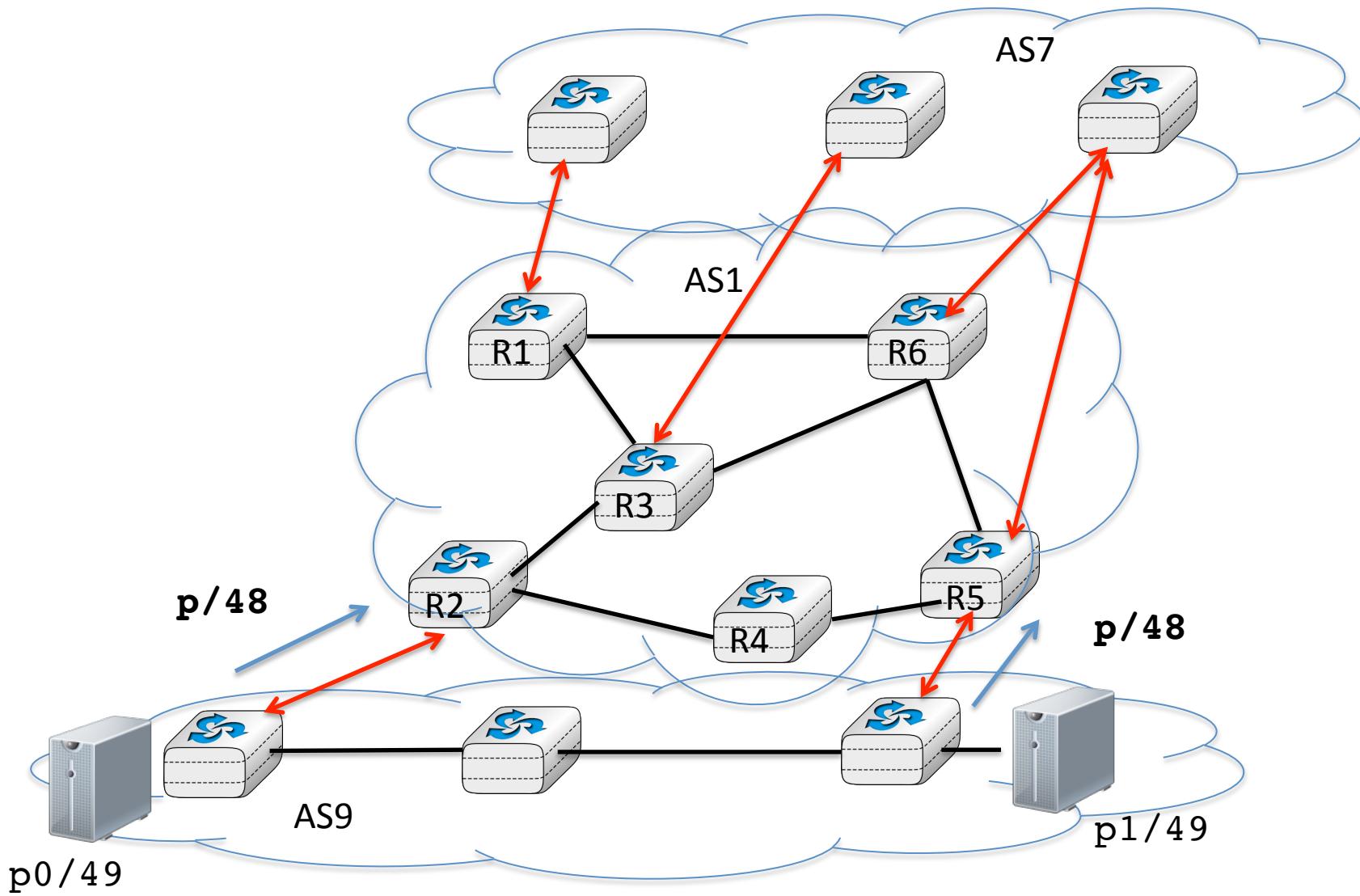


# Incoming traffic engineering

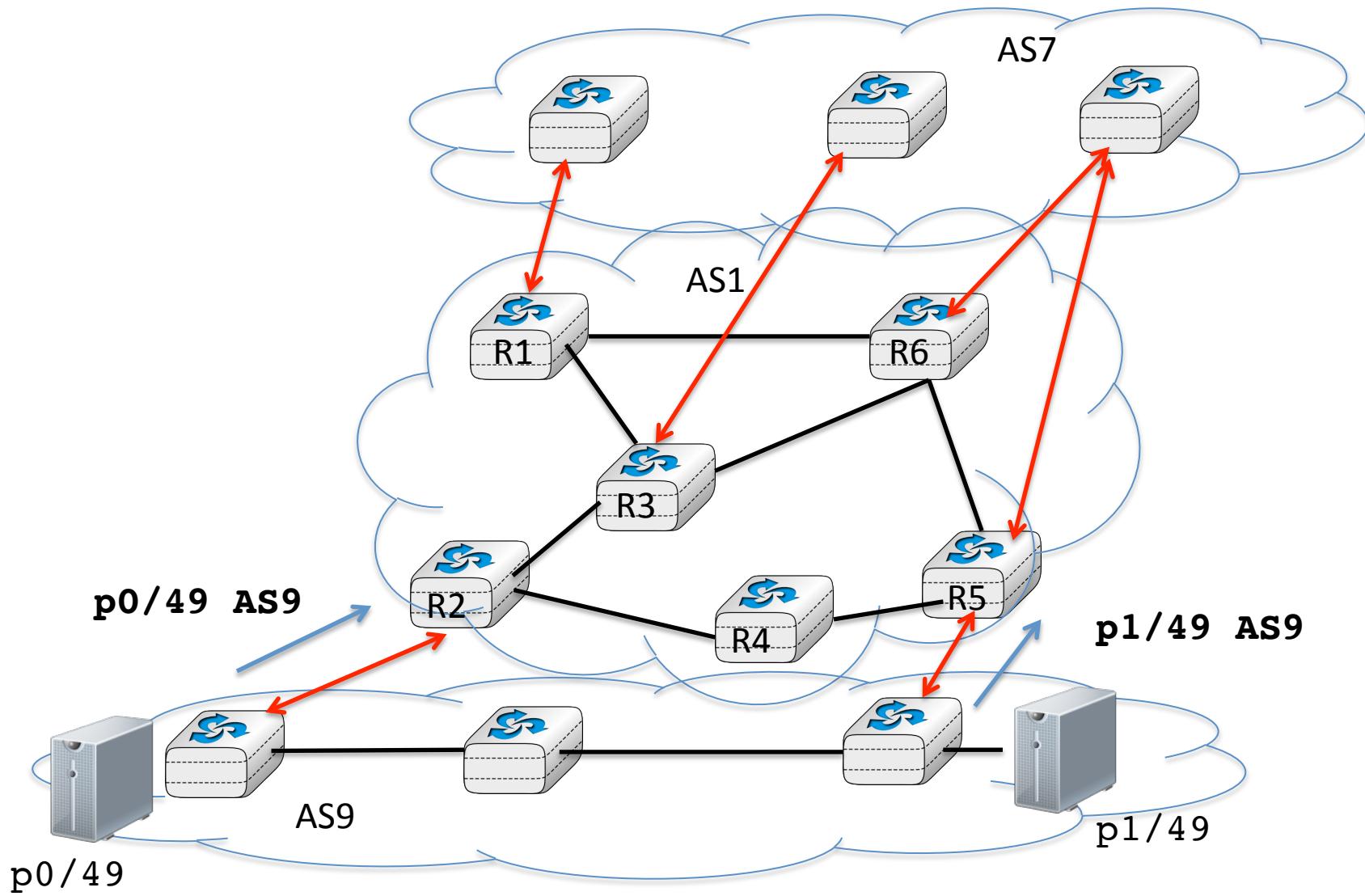
- How can a customer distribute the load on its links ?
  - one third of traffic on each link ?



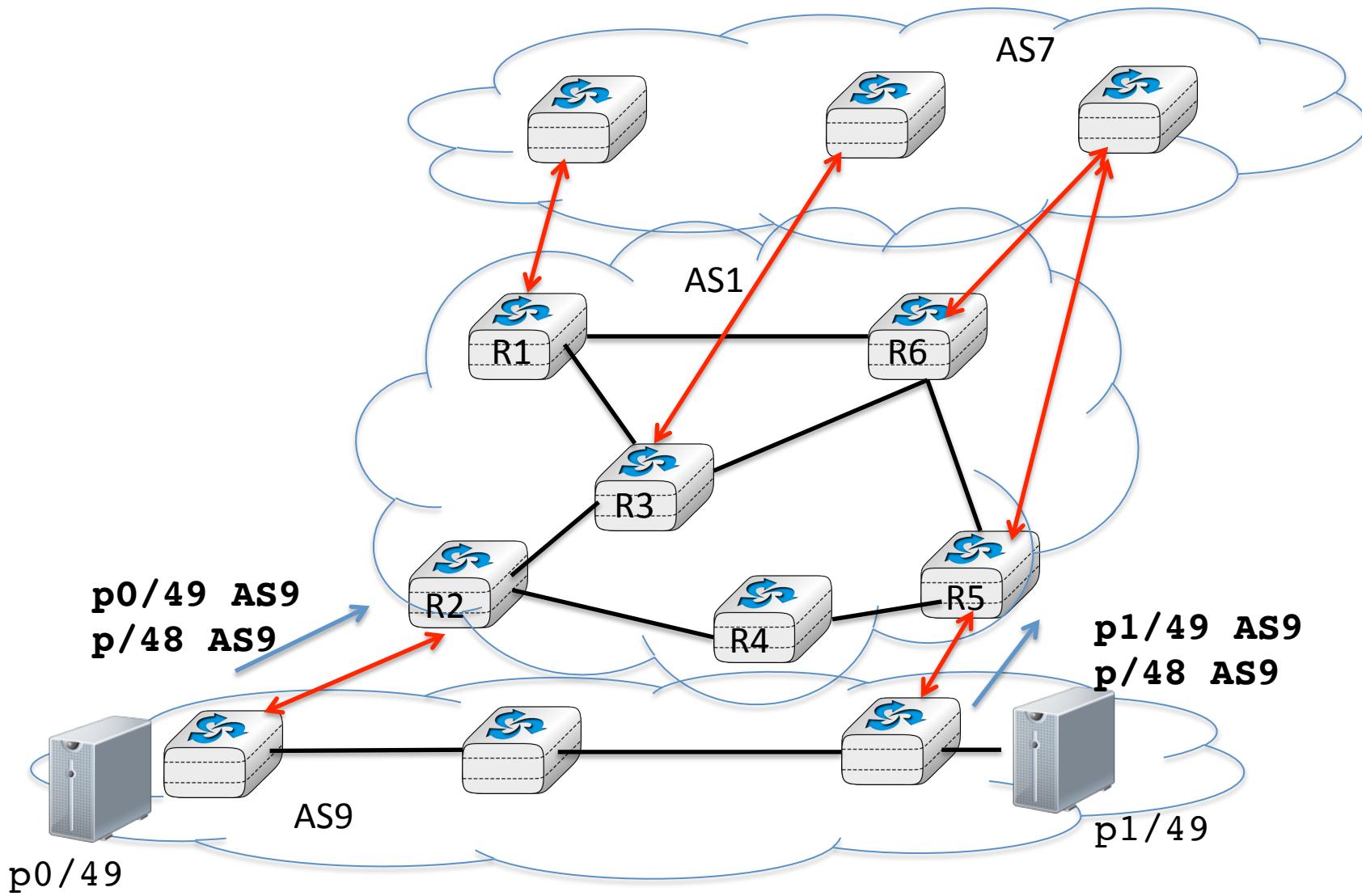
# More specific prefixes



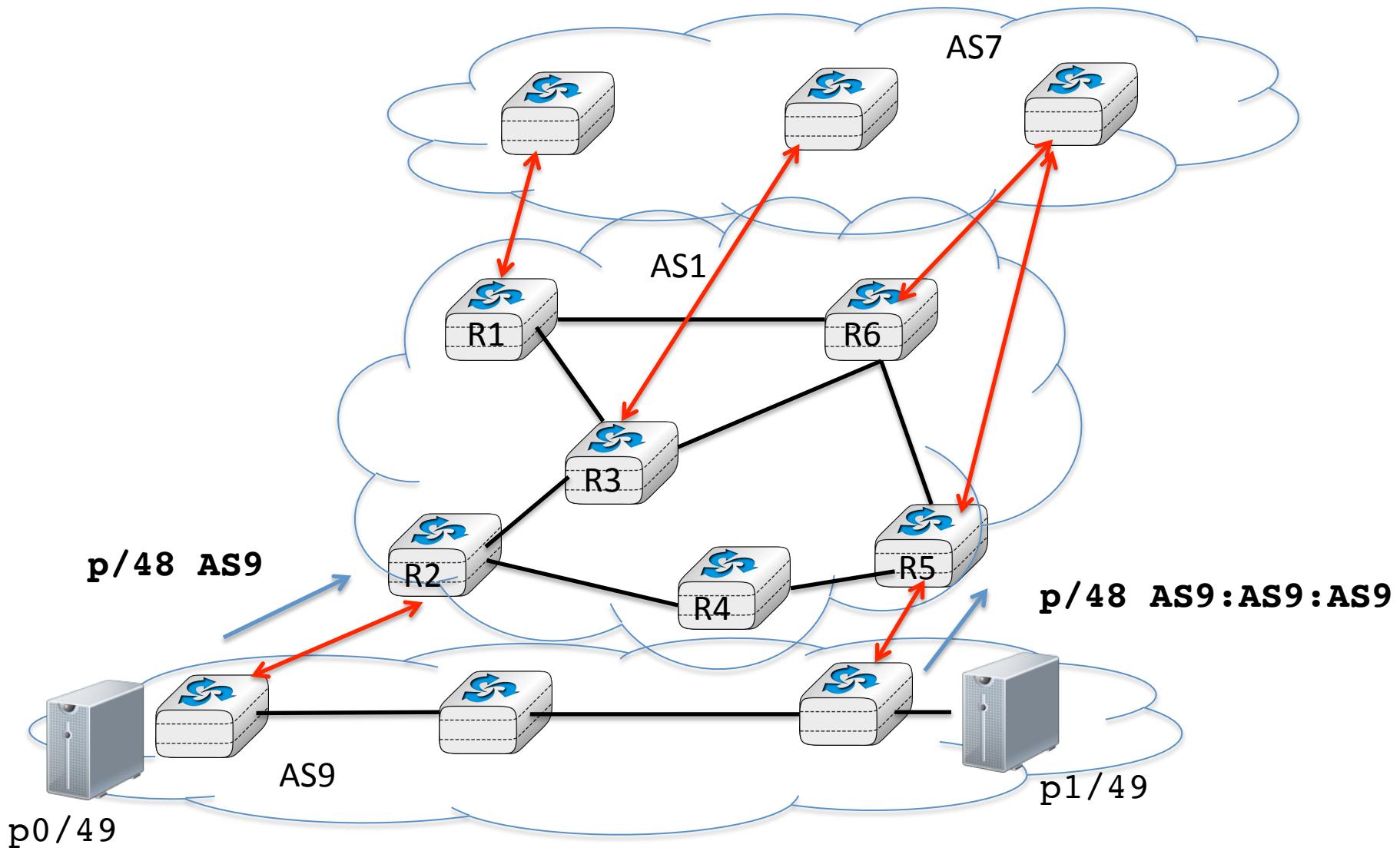
# More specific prefixes



# More specific prefixes



# AS Path prepending



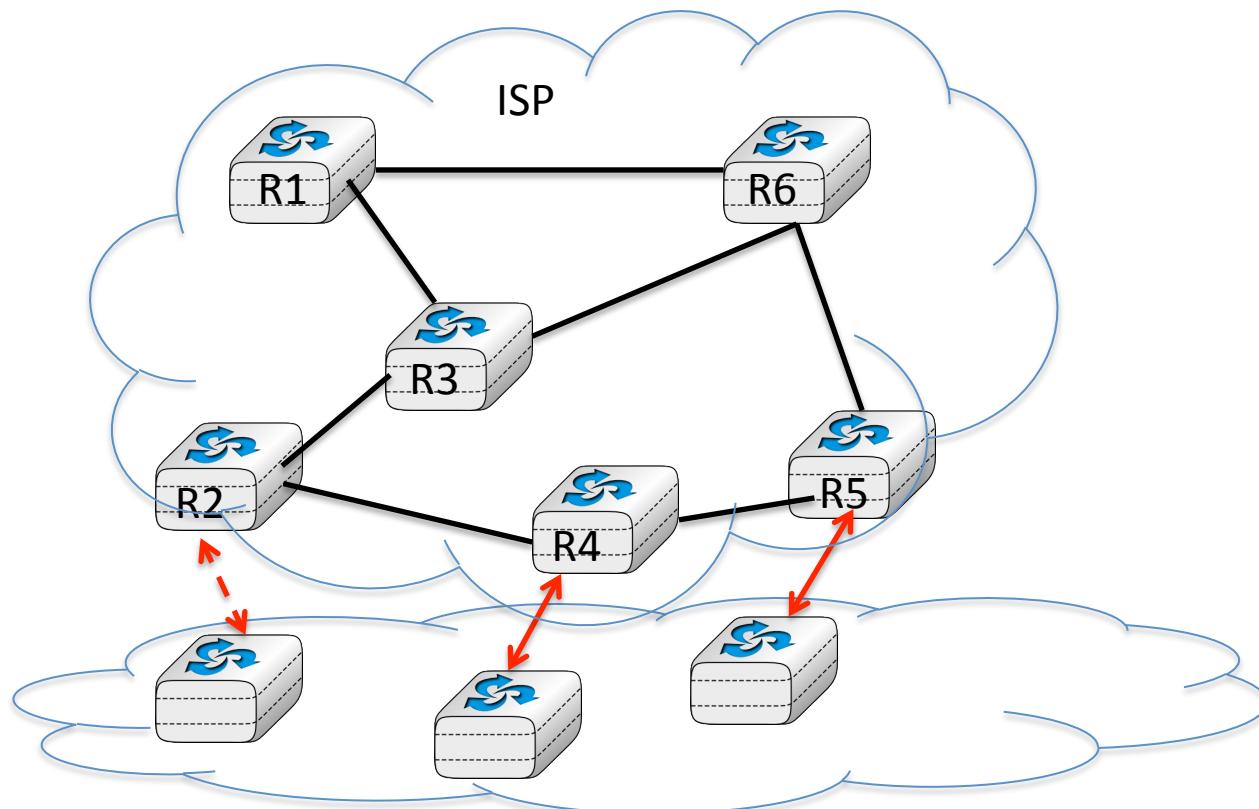
# Traffic engineering

## AS path prepending

- How does this traffic engineering technique works ?
  - Can it control incoming or outgoing traffic ?
  - What are its advantages and drawbacks ?

# Backup links

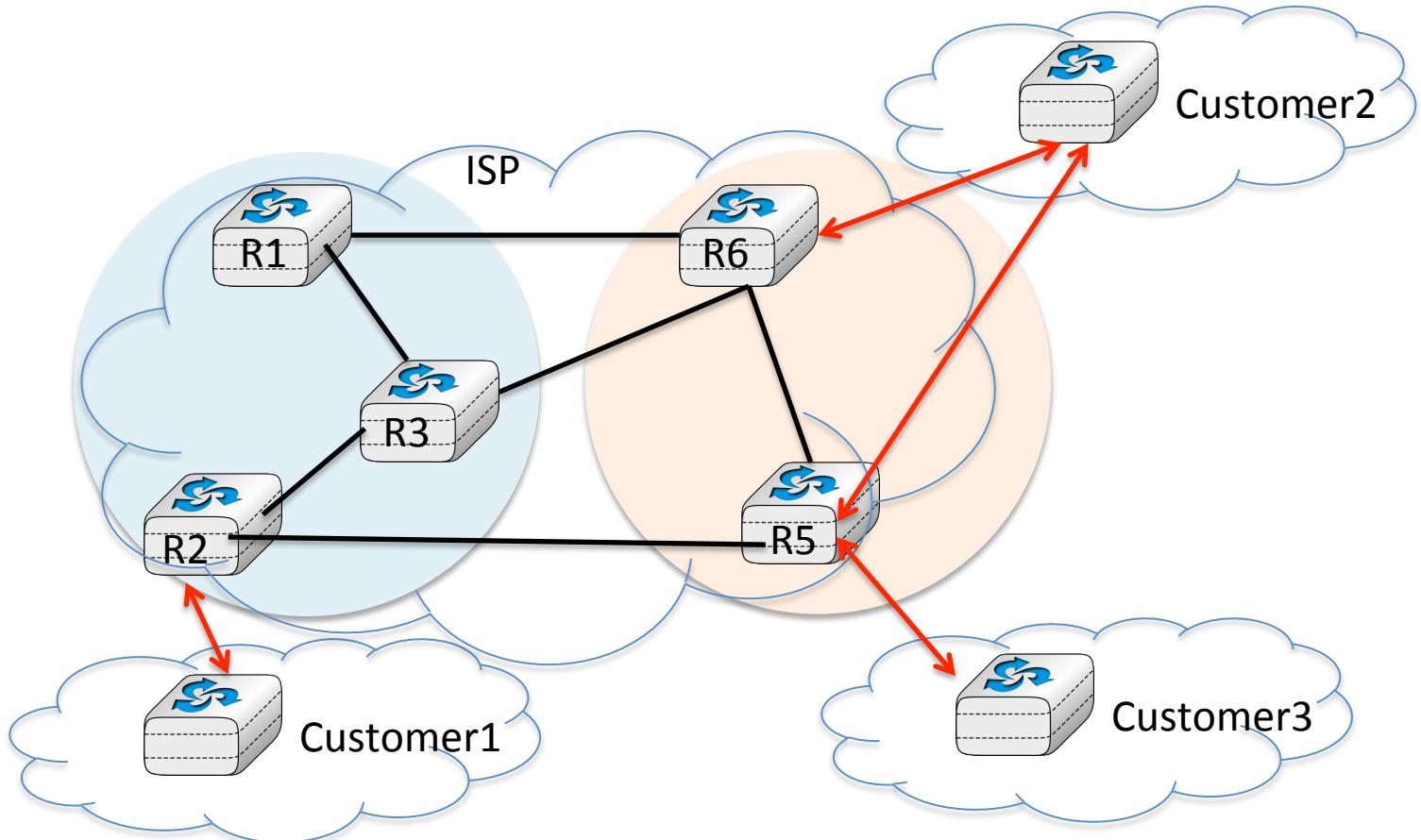
- How can an ISP provide backup services to its customers ?



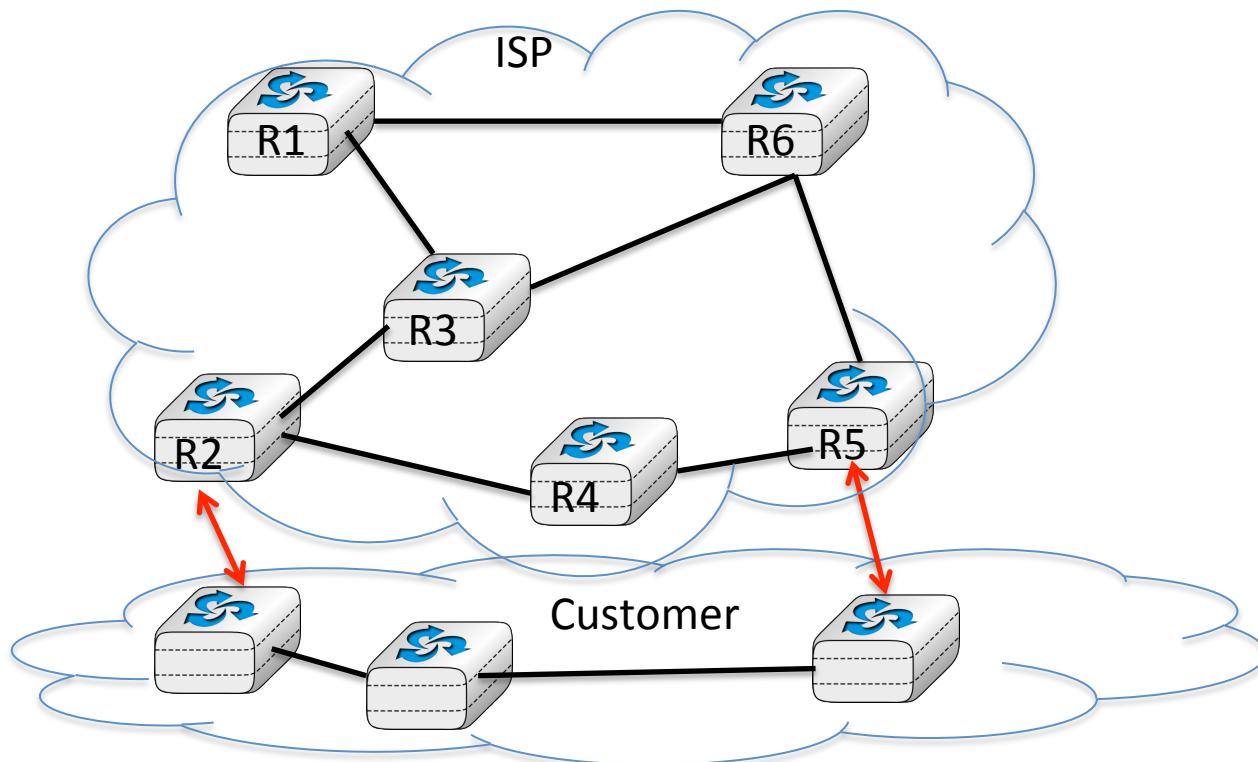
# BGP Communities

- BGP Communities can be attached to BGP routes in import filter
  - To indicate geographical location
  - To indicate type of BGP session
  - ...
- BGP Communities can be attached to BGP routes in export filter
  - To request neighbour AS to treat the route in a specific way

# How to provide restricted transit ?

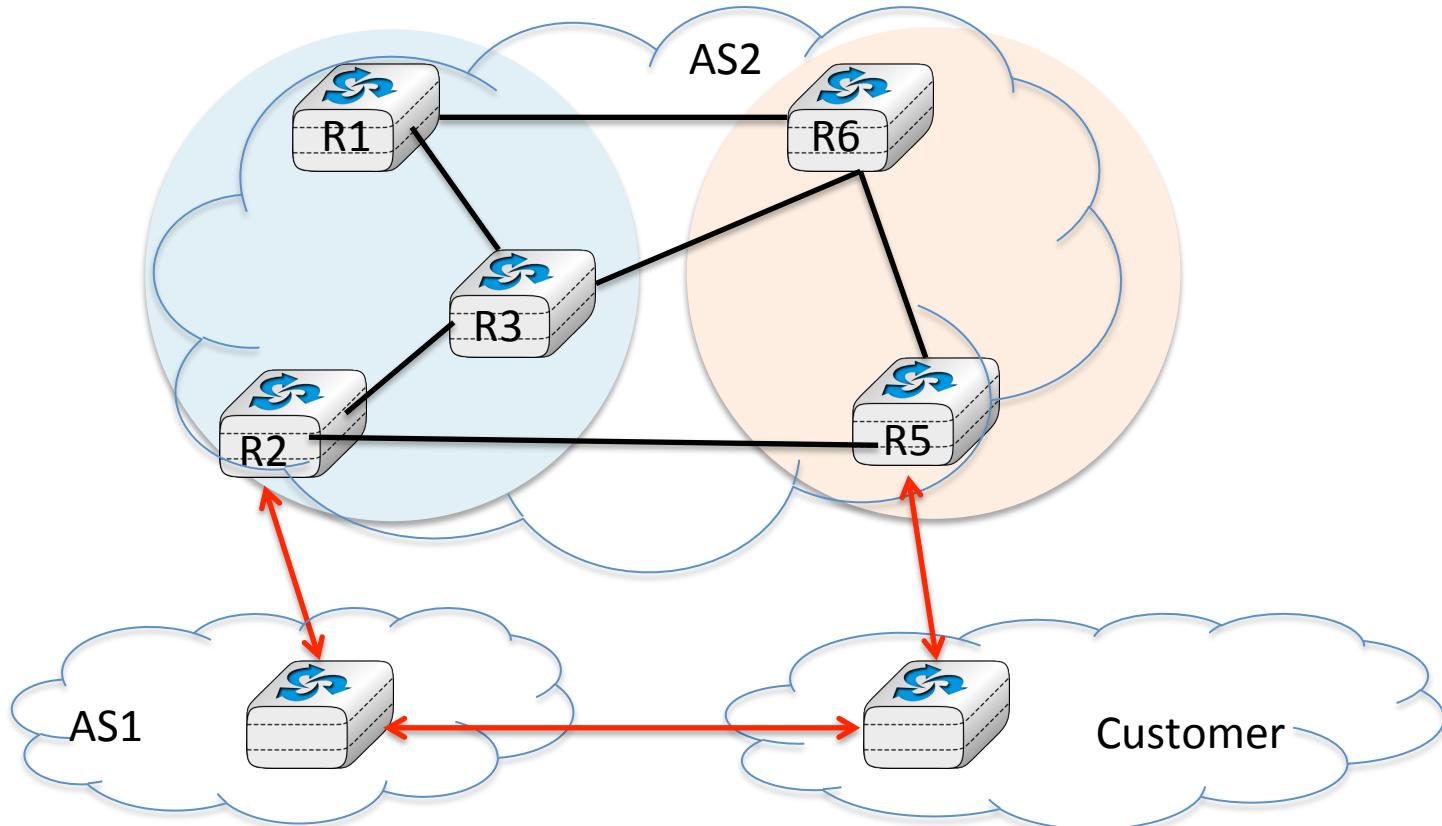


# ISP provides primary connectivity between customer sites



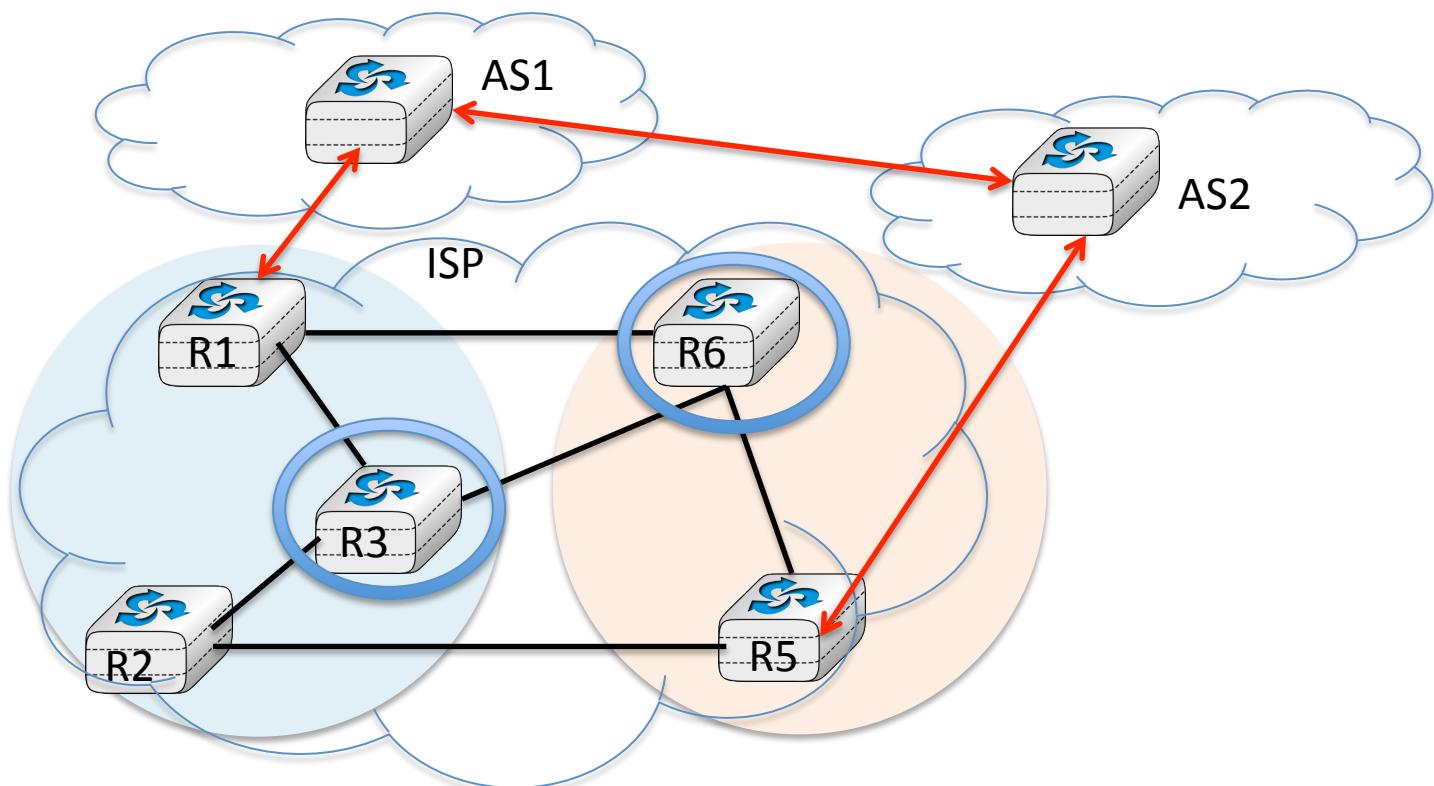
# How to provide richer customer policies ?

- Customer wants to receive packets from US via AS1 and from Europe via AS2



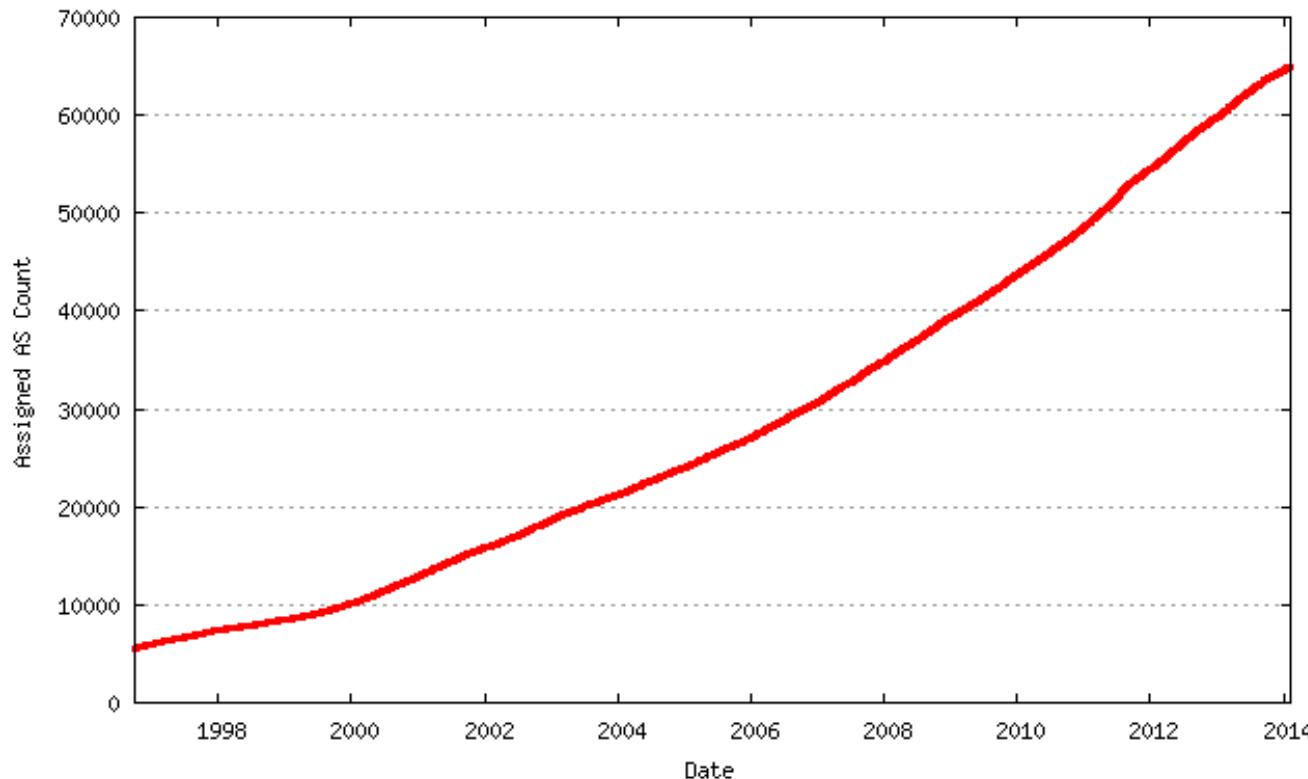
# AS-Path length is not always a synonym of path quality

- How to prefer AS1 in US, AS2 in Europe

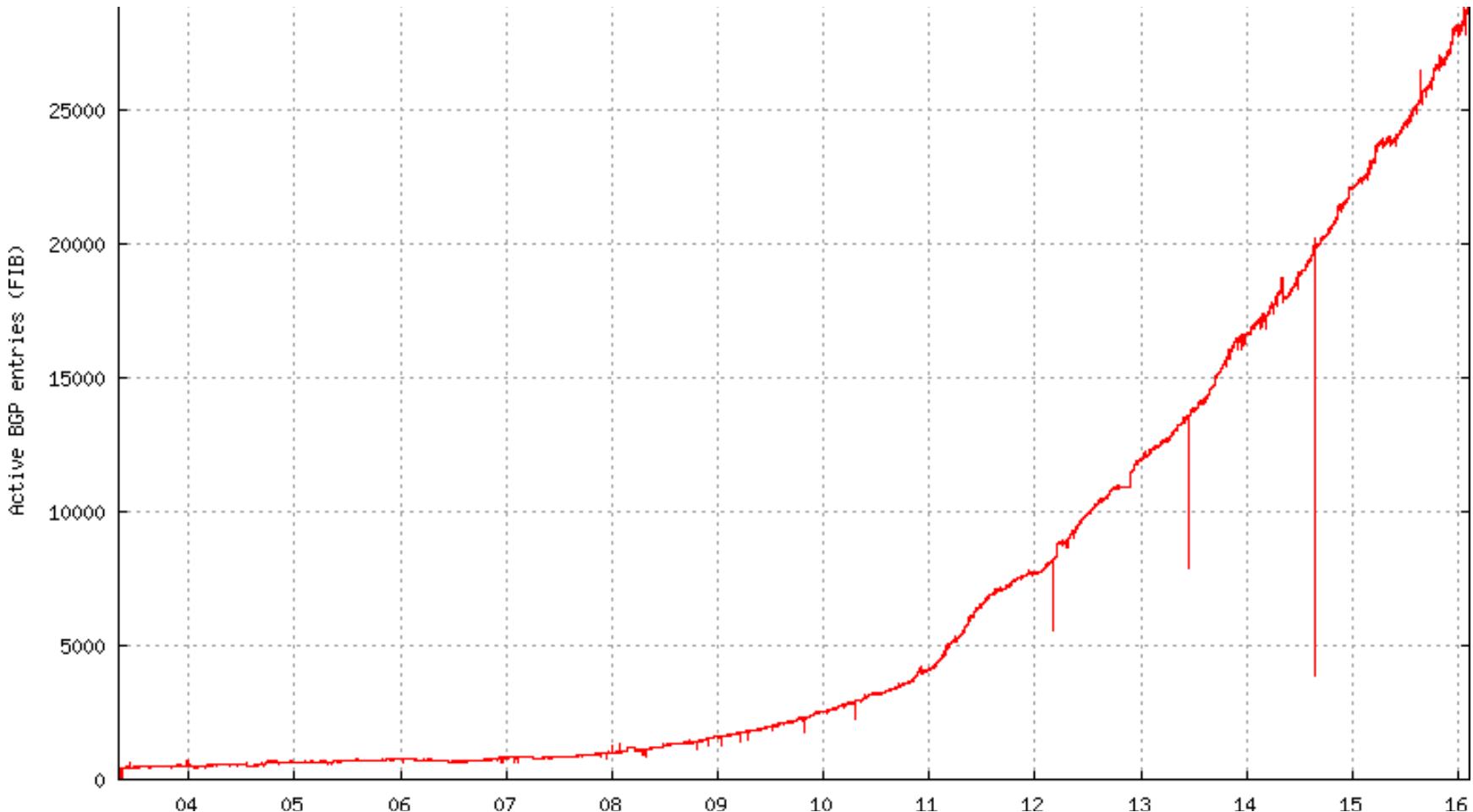


# Things to remember when defining BGP policies

- Any tweaking you do could affects scalability

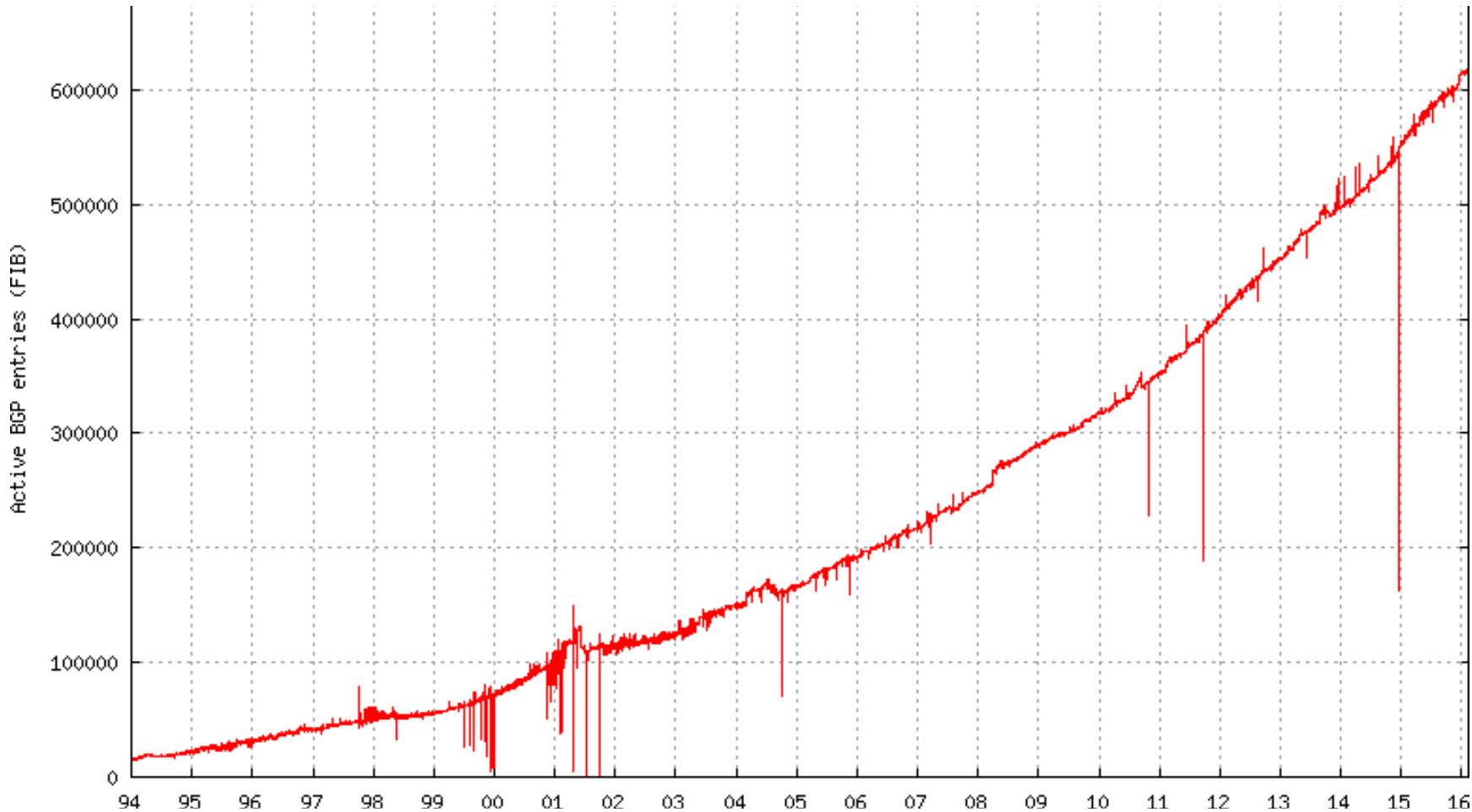


# Size of IPv6 routing tables



Source <http://bgp.potaroo.net/v6/as6447/>

# Size of IPv4 BGP routing tables



Source <http://bgp.potaroo.net/as6447/>

# BGP communities

- Are by default transitive..
- Any BGP community that you add when receiving routes will be advertised all over the Internet
  - you should clean your BGP communities when advertising routes over eBGP, but router configuration languages do not always make this easy

# References

- K. Fster, Application of BGP Communities, The Internet Protocol Journal - Volume 6, Number 2, July 2003
- B. Donnet and O. Bonaventure. On BGP Communities. ACM SIGCOMM Computer Communication Review, 38(2):55-59, April 2008.
  - <http://inl.info.ucl.ac.be/publications/bgp-communities>
- B. Quoitin, S. Uhlig, C. Pelsser, L. Swinnen and O. Bonaventure. Interdomain traffic engineering with BGP. IEEE Communications Magazine Internet Technology Series, 41(5):122-128, May 2003.
  - <http://inl.info.ucl.ac.be/publications/interdomain-traffic-engineering-bgp>