Institut für Kommunikations-Technik

Router Configuration

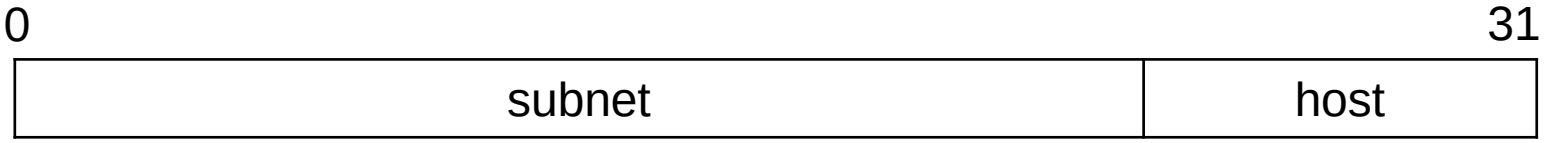
Network Management

Prof. Dr. Panagiotis Papadimitriou

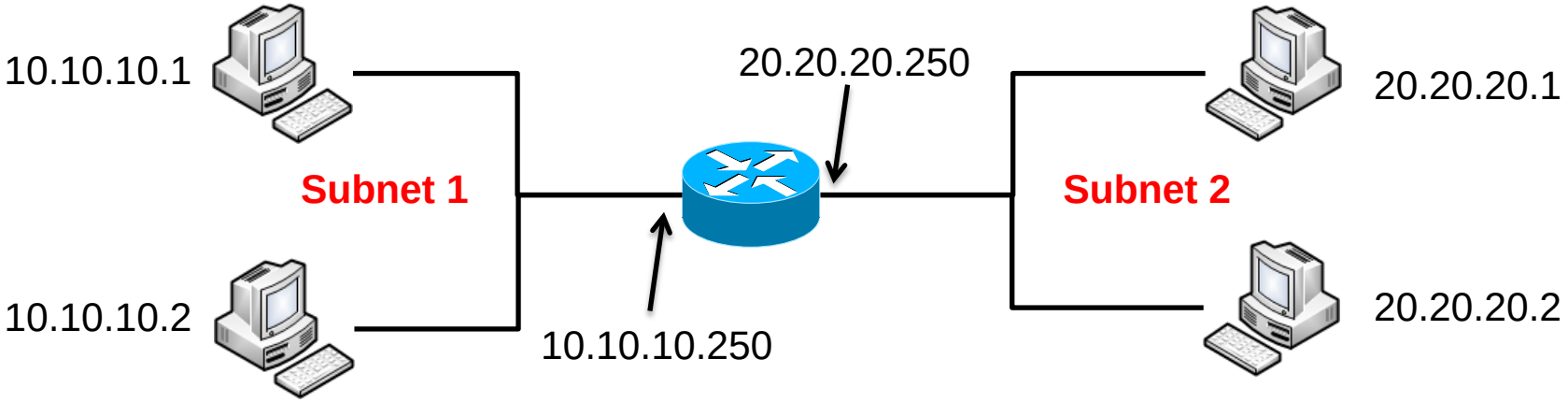
Some slides based on textbook “Computer Networking: Principles, Protocols and Practice”

Subnetting

- An IP address is subdivided into 2 parts:

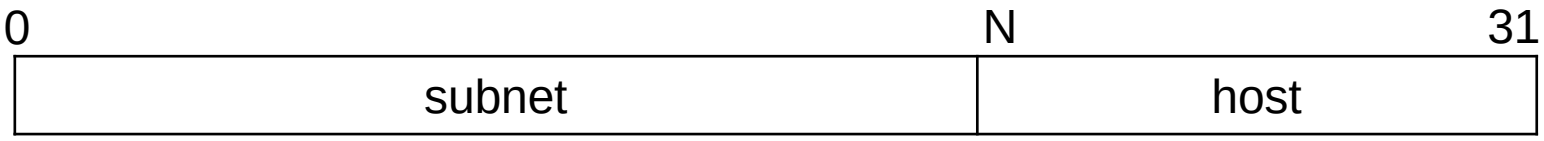


- Subnetting:
  - Device interfaces with the same subnet portion in their IP addresses are grouped into a subnet (isolated network)
  - Interfaces within the same subnet can communicate directly (without a router)
  - Communication across different subnets requires at least one router



Classless Inter-Domain Routing (CIDR)

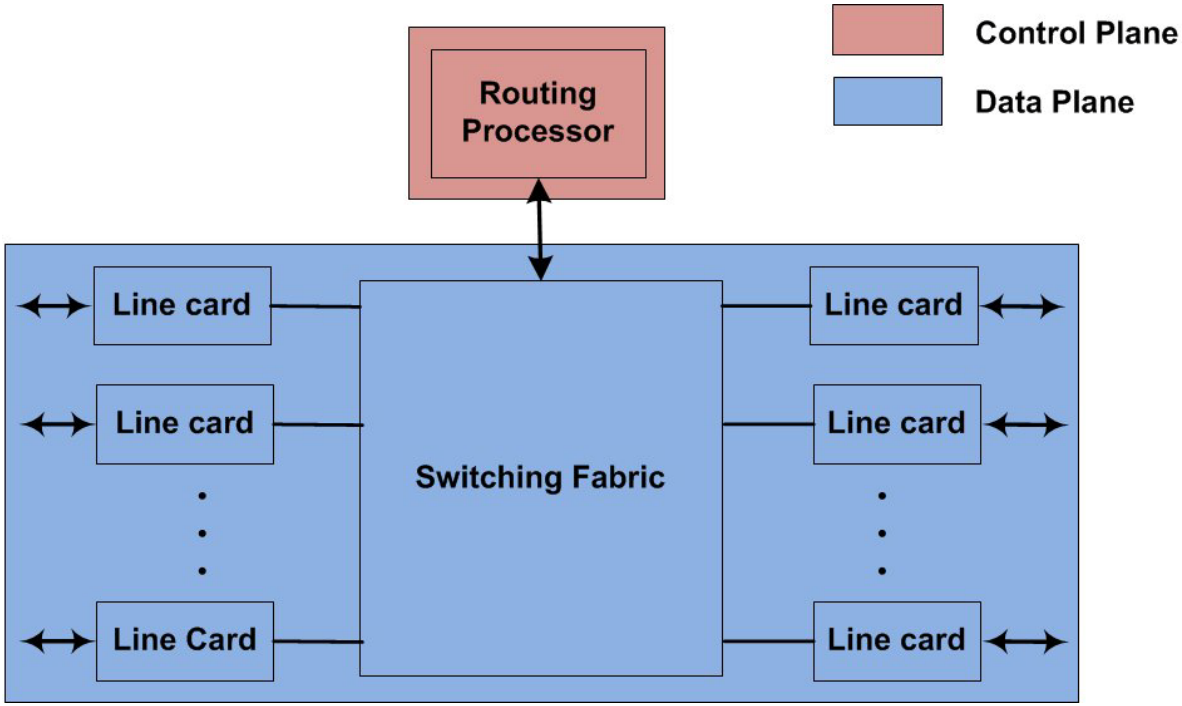
- CIDR replaced Classful Addressing to provide better utilization of IPv4 address space:
  - IP address is subdivided into a subnet portion (of arbitrary length) and a host portion
  - Address format: A.B.C.D/N, where N represents the number of bits for the subnet portion (or prefix)
  - /N is known as the subnet mask (N leftmost bits set to 1 and the rest 32-N bits set to 0)



- Example: 200.10.16.0/23 (subnet mask: 255.255.254.0)

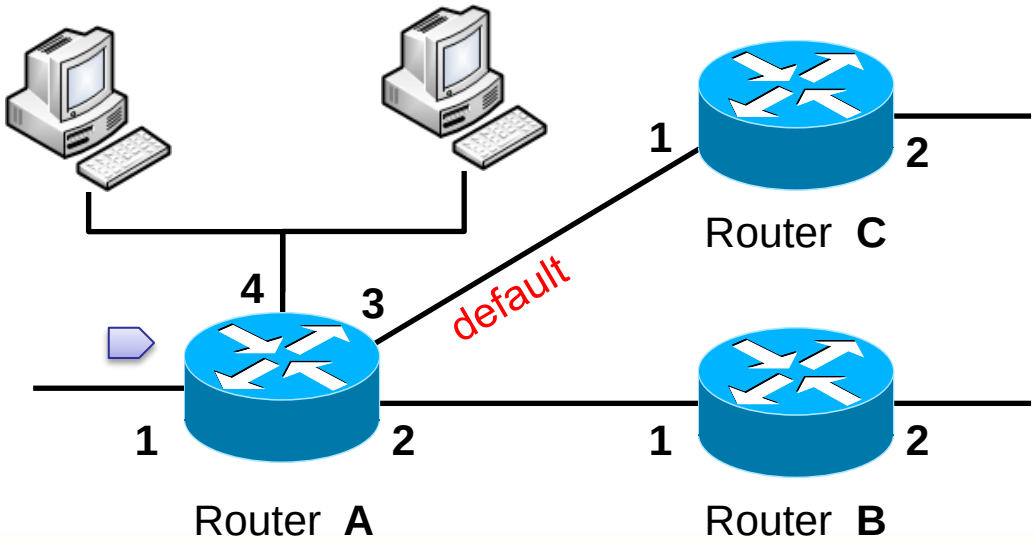
Routers

- Data plane:
  - Forwards packets from input to output
- Control plane:
  - Runs routing protocols to compute the paths that packets will follow



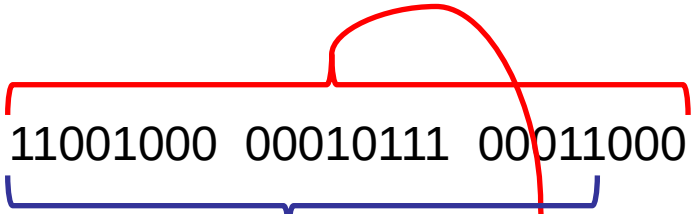
Packet Forwarding Algorithm

if destination subnet matches one of the router interfaces **then**
    forward packet over that interface
else
    if destination subnet is in my forwarding table **then**
        forward packet to Next-Hop router
    else
        forward packet to default router



Longest Prefix Match

- A packet destination IP address may match more than one prefixes in a router forwarding table:

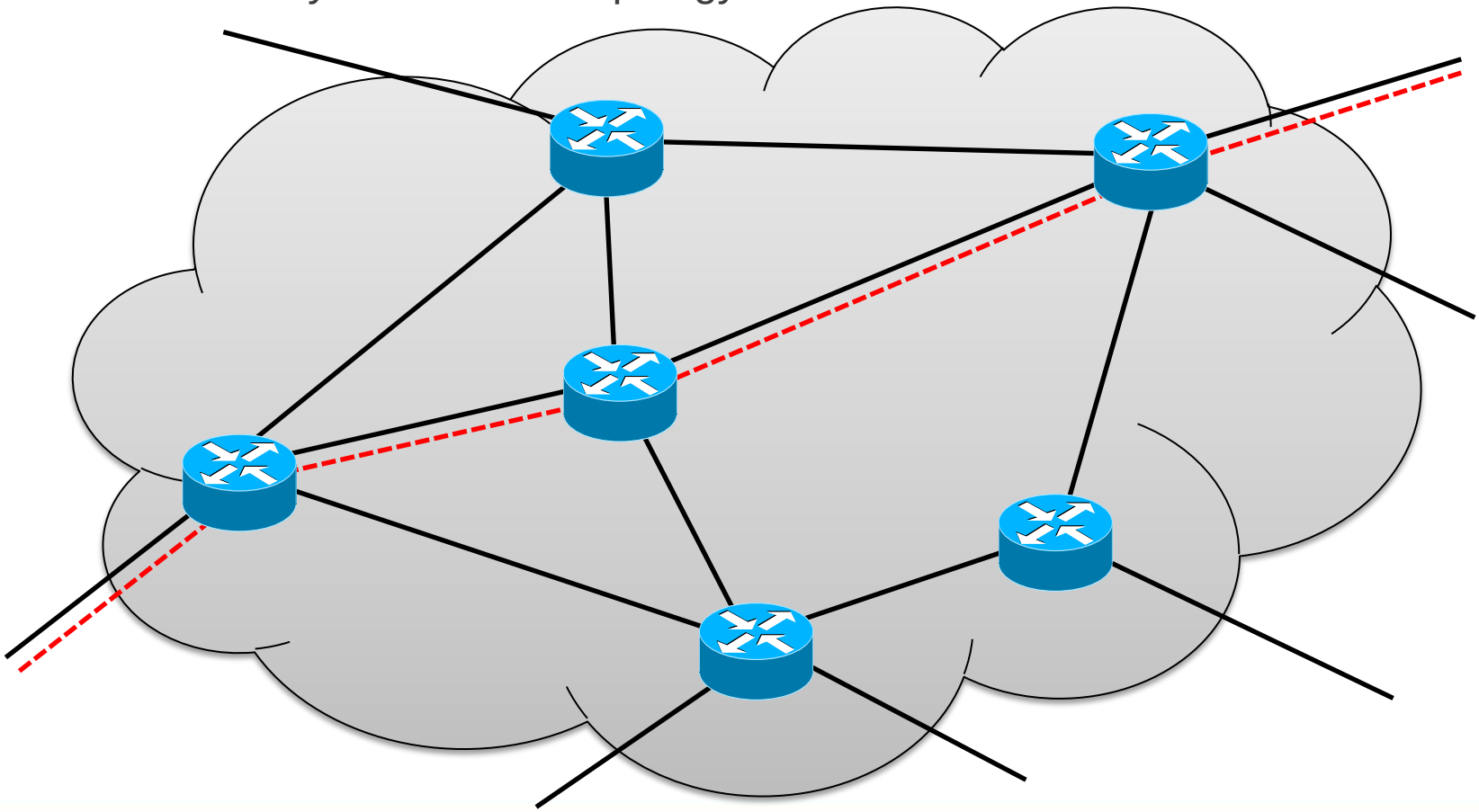


Prefix	Link Interface
11001000 00010111 00010	**1**
11001000 00010111 00011000	**2**
11001000 00010111 00011	**3**
otherwise	**4**

- In the case of multiple matches, the router uses the longest prefix match

Intra-Domain Routing (1)

- Interior Gateway Protocol (IGP):
  - Routing of IP packets inside each domain
  - Knows only the network topology of each domain



Intra-Domain Routing (2)

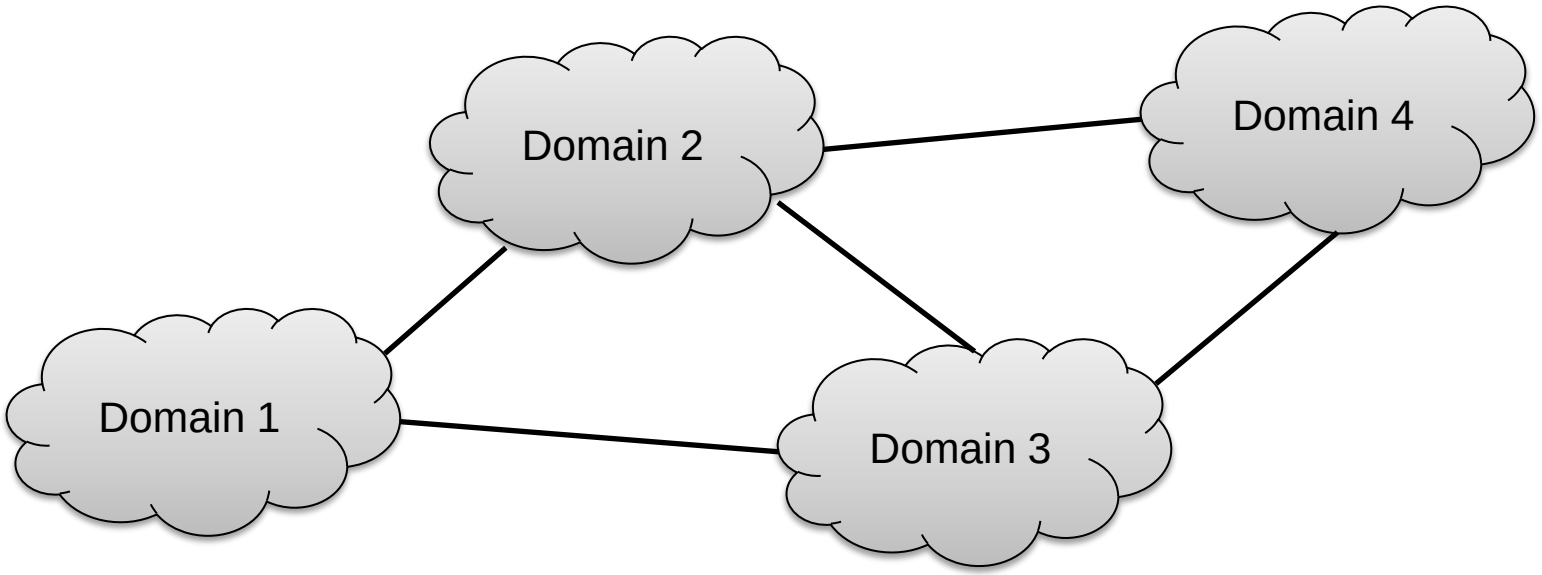
- Goal:
  - Allows routers to forward IP packets along the best path towards their destination
  - Best path corresponds to:
    - shortest path (in terms of delay or number of hops)
    - less loaded path (i.e., more available bandwidth)
- Operation:
  - All routers exchange routing information:
    - Each domain router obtains routing information for the whole domain
    - The network operator or the routing protocol assigns costs to each link

Intra-Domain Routing (3)

- Static routing:
  - Routes are configured and updated statically by the network operator
  - Useful only in very small domains
- Distance vector routing:
  - Routing Information Protocol (RIP)
    - Still widely used in small domains despite its limitations
- Link-state routing:
  - Open Shortest Path First (OSPF)
    - Widely used in enterprise networks
  - Intermediate System – Intermediate System (IS-IS)
    - Widely used by ISPs

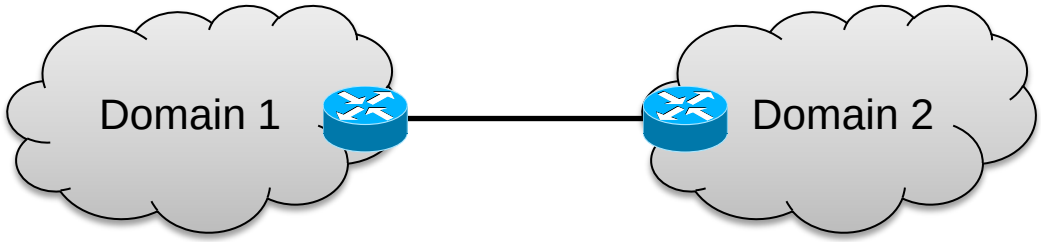
Inter-Domain Routing

- Exterior Gateway Protocol (EGP):
  - Routing of IP packets between domains
  - Each domain is considered as a black-box

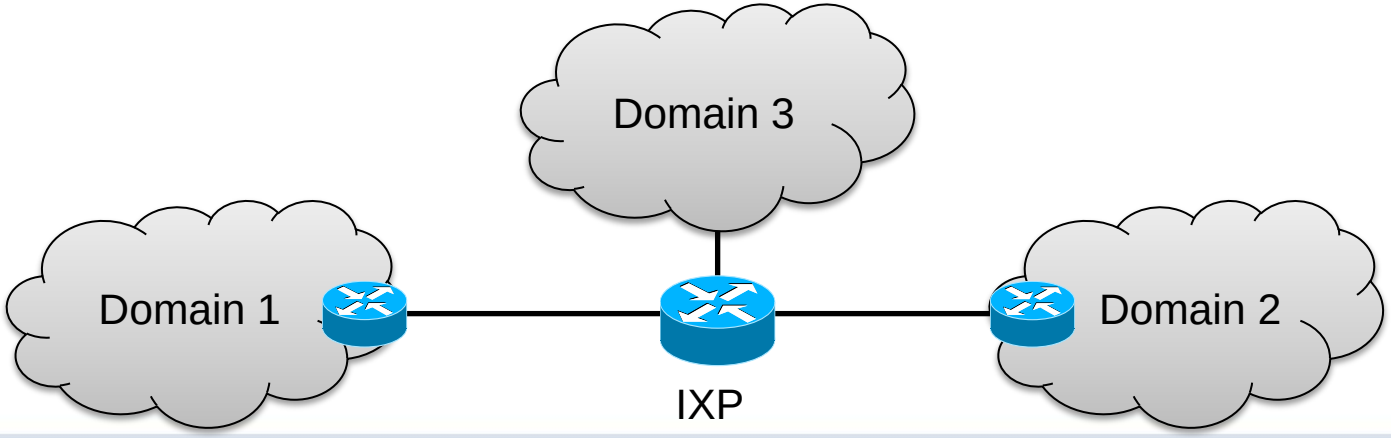


Inter-Domain Links

- Private link:
  - Leased line where the border routers are attached



- Connection via a public interconnection point (a.k.a. IXP):
  - Infrastructure composed of high-speed Ethernet switches that interconnect routers belonging to different domains



IXP Topology

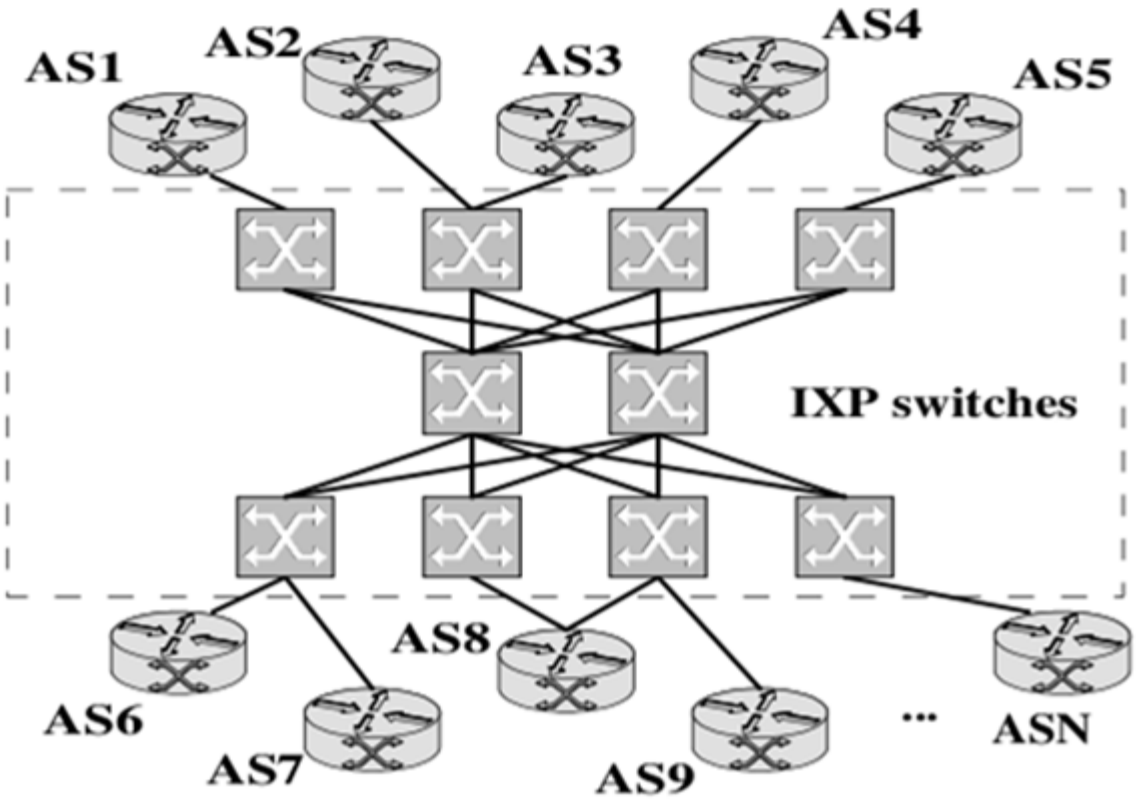
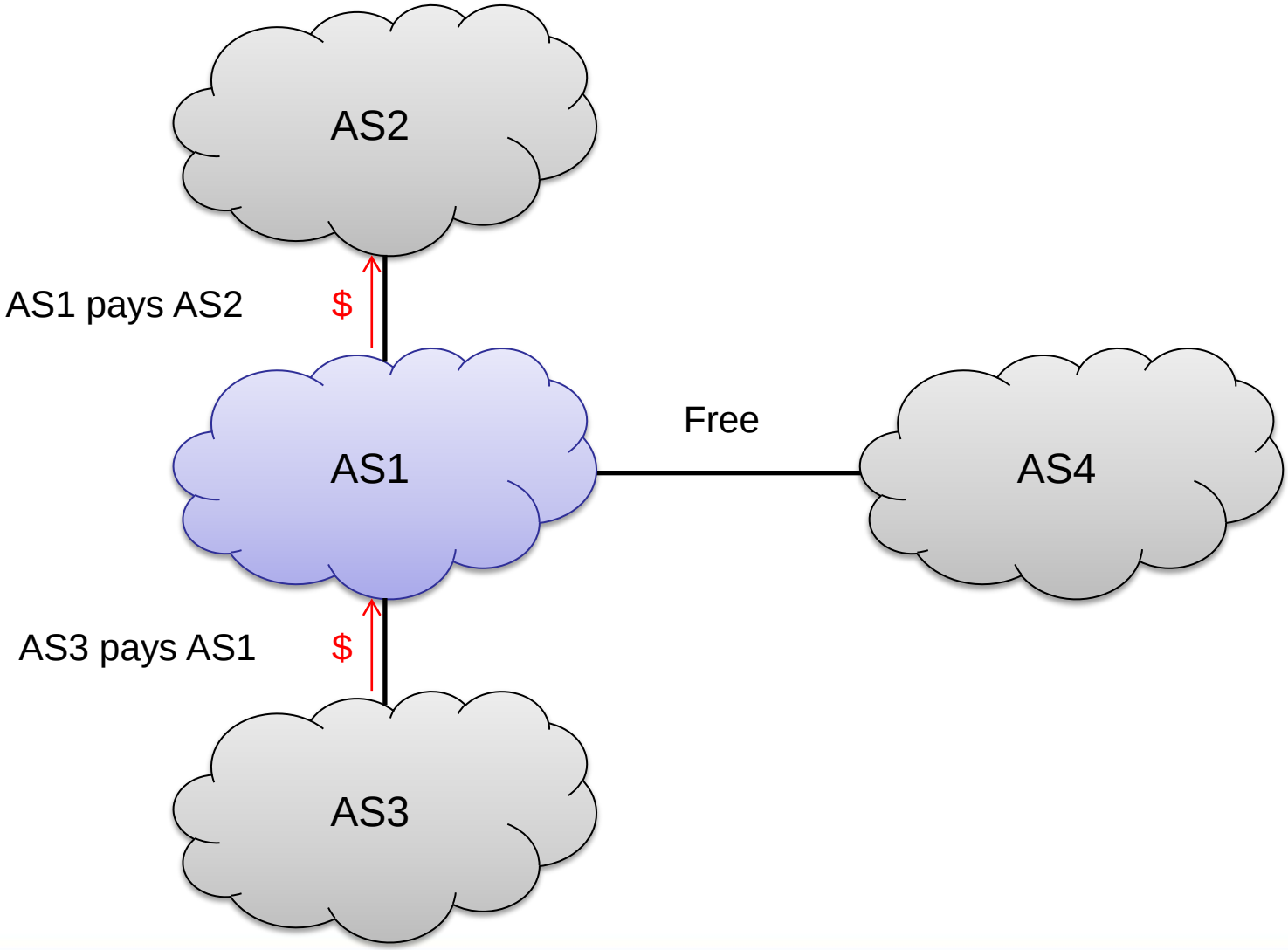


Figure from B. Ager, et al., Anatomy of a Large European IXP

Inter-Domain Routing (1)

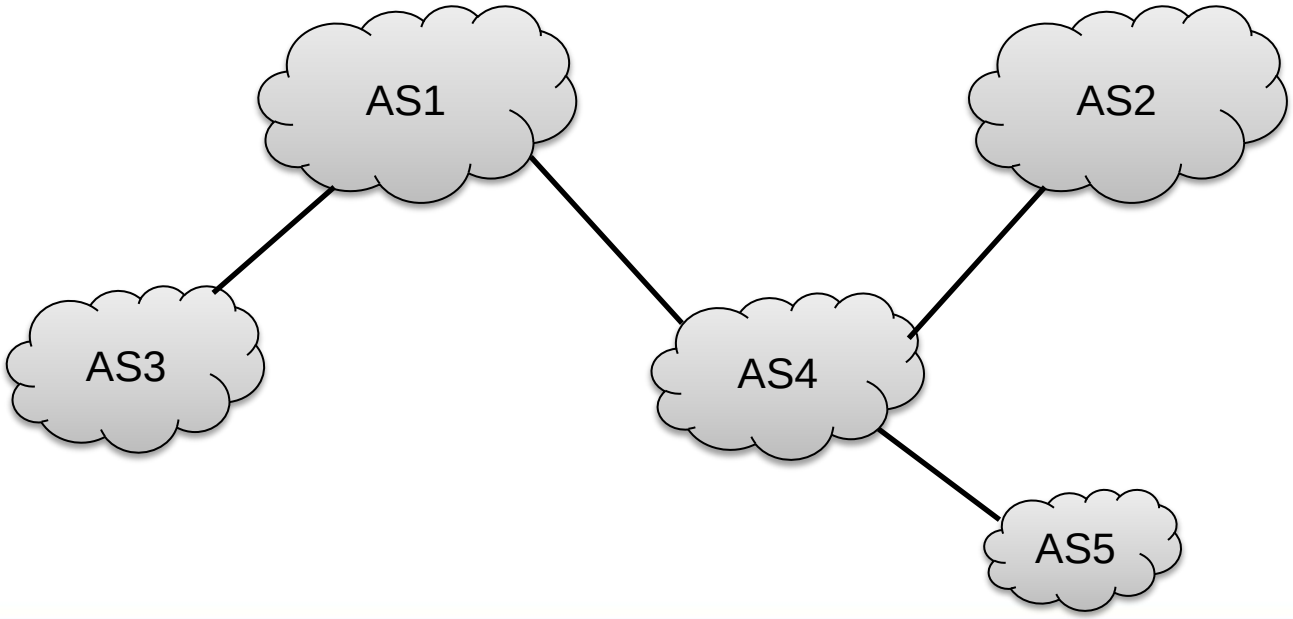
- Goal:
  - Allow routers to forward IP packets along the best path towards their destination through several transit domains while taking into account the routing policies of each domain (with limited knowledge of the topology of these domains)
  - Best path corresponds to:
    - cheapest path (i.e., depends on routing policies)
  - Each domain can specify its routing policy, i.e.:
    - the domains for which it agrees to provide a transit service
    - the method it uses to select the best path to reach each destination

Peering Relationships



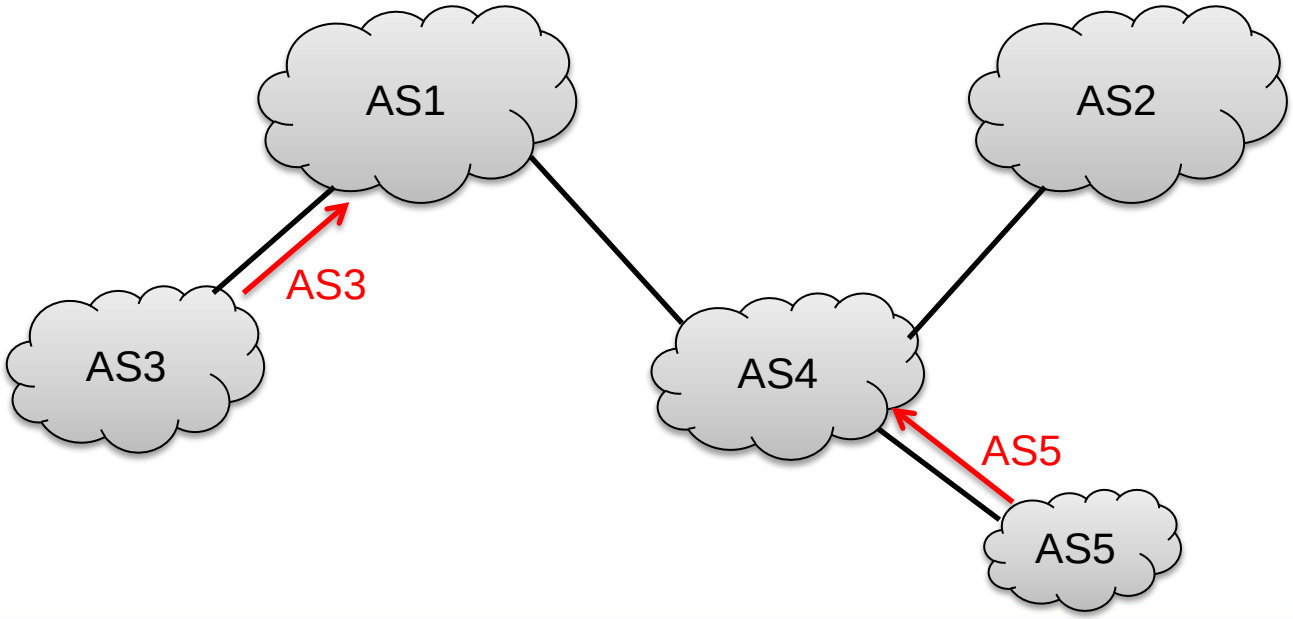
Paid Peering

- Customer advertises to its Provider its internal routes and the routes learned from its own customers
  - The advertised routes will be propagated (via the Provider) to the entire Internet allowing anyone to reach the customer



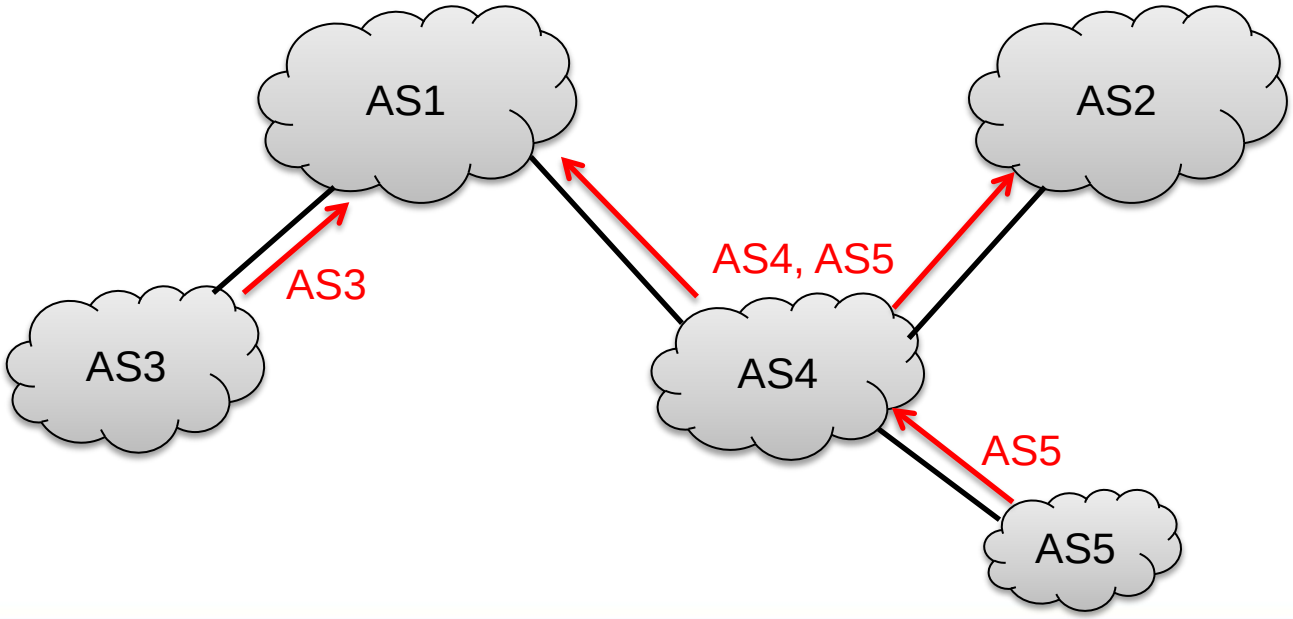
Paid Peering

- Customer advertises to its Provider its internal routes and the routes learned from its own customers
  - The advertised routes will be propagated (via the Provider) to the entire Internet allowing anyone to reach the customer



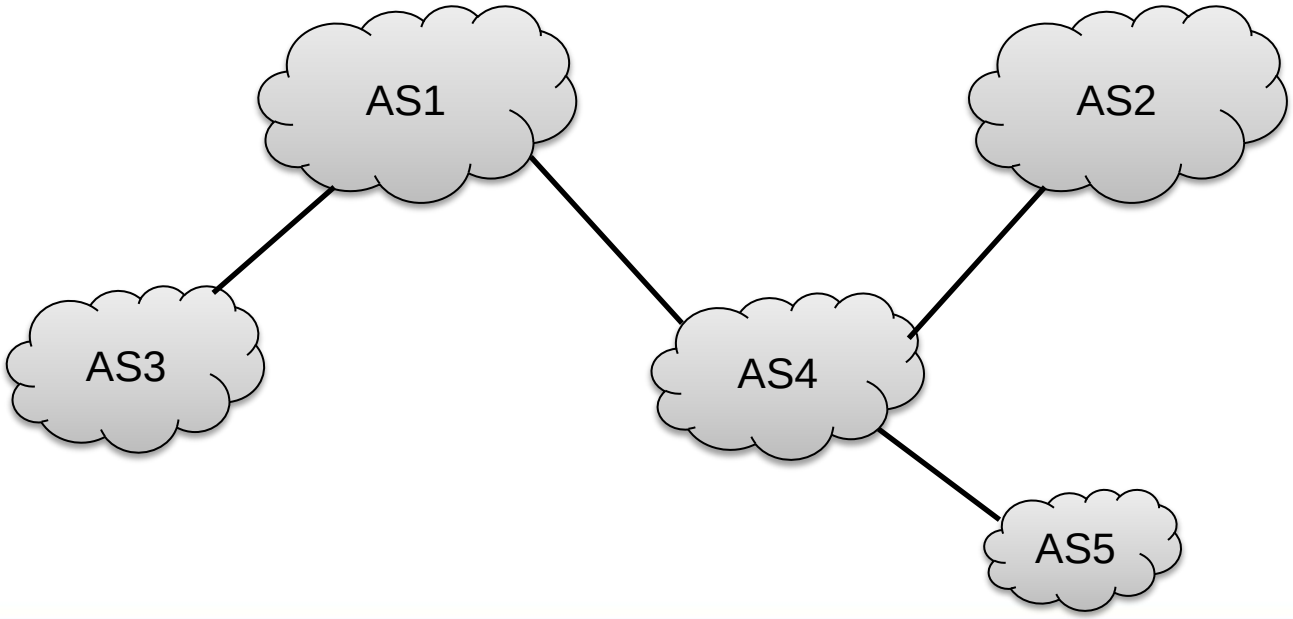
Paid Peering

- Customer advertises to its Provider its internal routes and the routes learned from its own customers
  - The advertised routes will be propagated (via the Provider) to the entire Internet allowing anyone to reach the customer



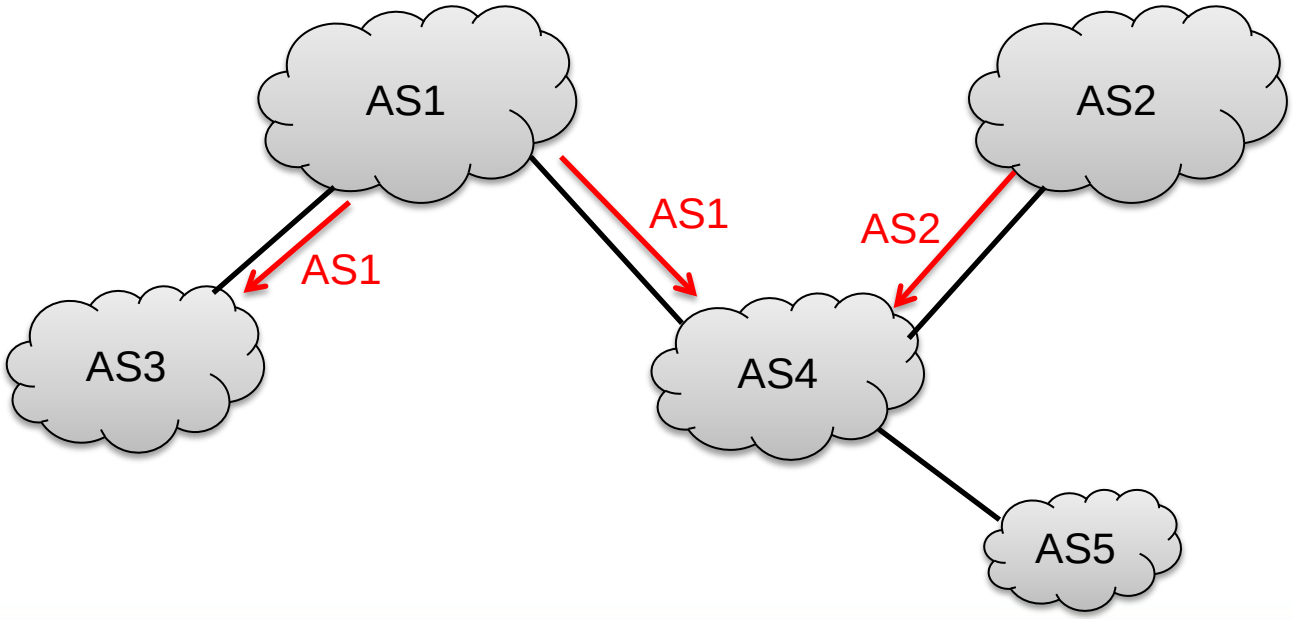
Paid Peering

- Customer advertises to its Provider its internal routes and the routes learned from its own customers
  - The advertised routes will be propagated (via the Provider) to the entire Internet allowing anyone to reach the customer
- Provider advertises all known routes to its customers
  - Each customer will be able to reach anyone in the Internet



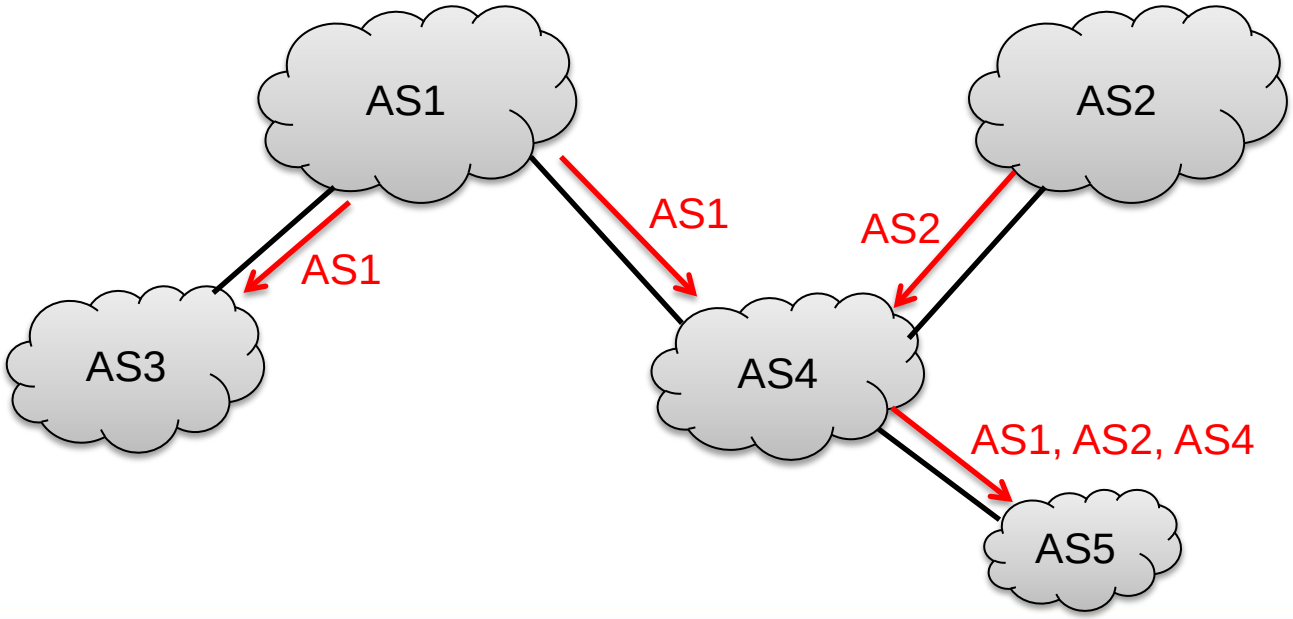
Paid Peering

- Customer advertises to its Provider its internal routes and the routes learned from its own customers
  - The advertised routes will be propagated (via the Provider) to the entire Internet allowing anyone to reach the customer
- Provider advertises all known routes to its customers
  - Each customer will be able to reach anyone in the Internet



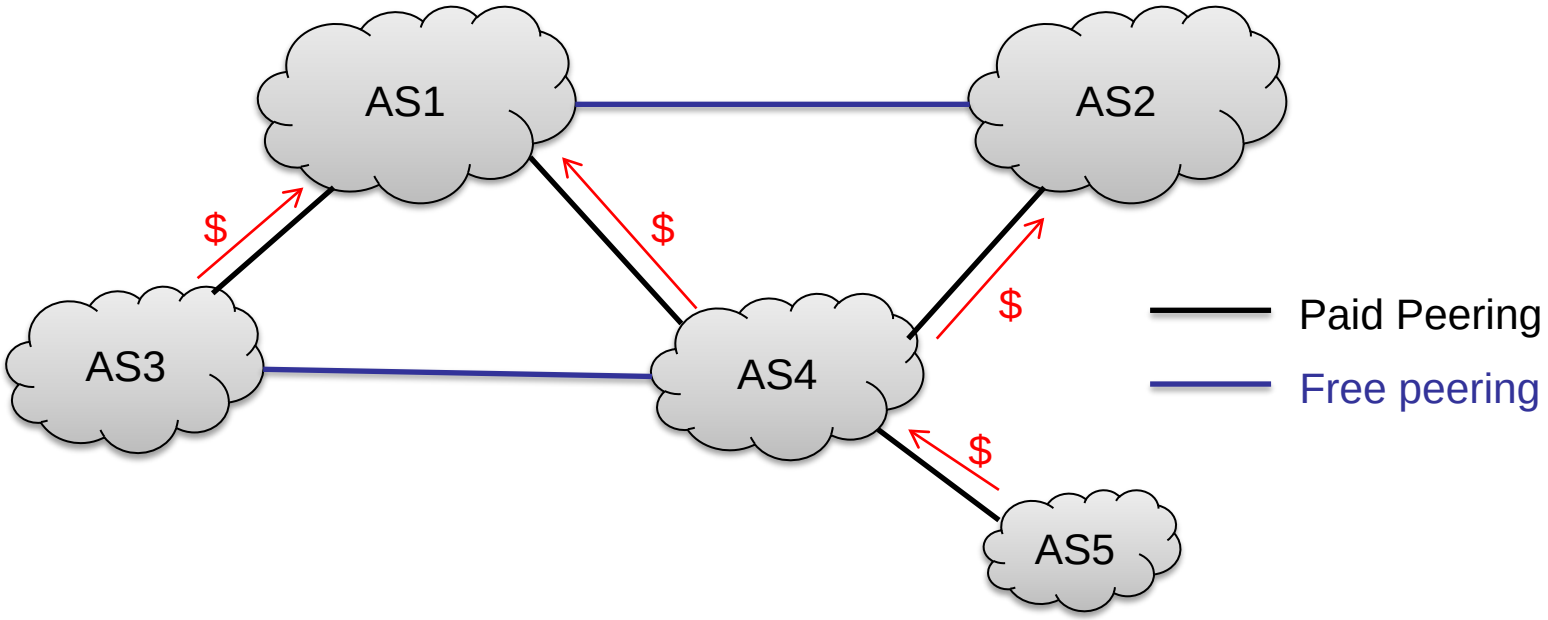
Paid Peering

- Customer advertises to its Provider its internal routes and the routes learned from its own customers
  - The advertised routes will be propagated (via the Provider) to the entire Internet allowing anyone to reach the customer
- Provider advertises all known routes to its customers
  - Each customer will be able to reach anyone in the Internet



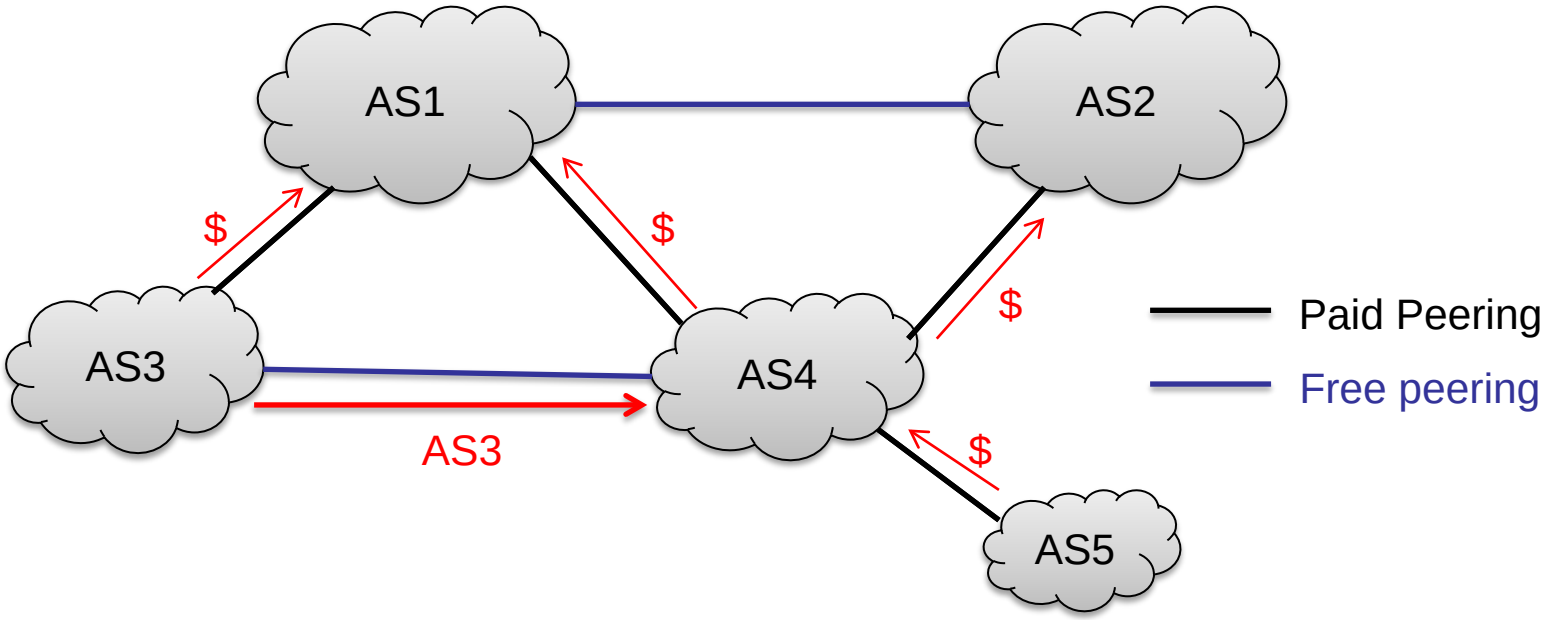
Settlement-Free Peering

- AS4 advertises to AS3 its internal routes and the routes learned from its own customers
  - AS3 will use the shared link to reach AS4 and AS4's customers
  - AS4's providers are nor reachable via the shared link



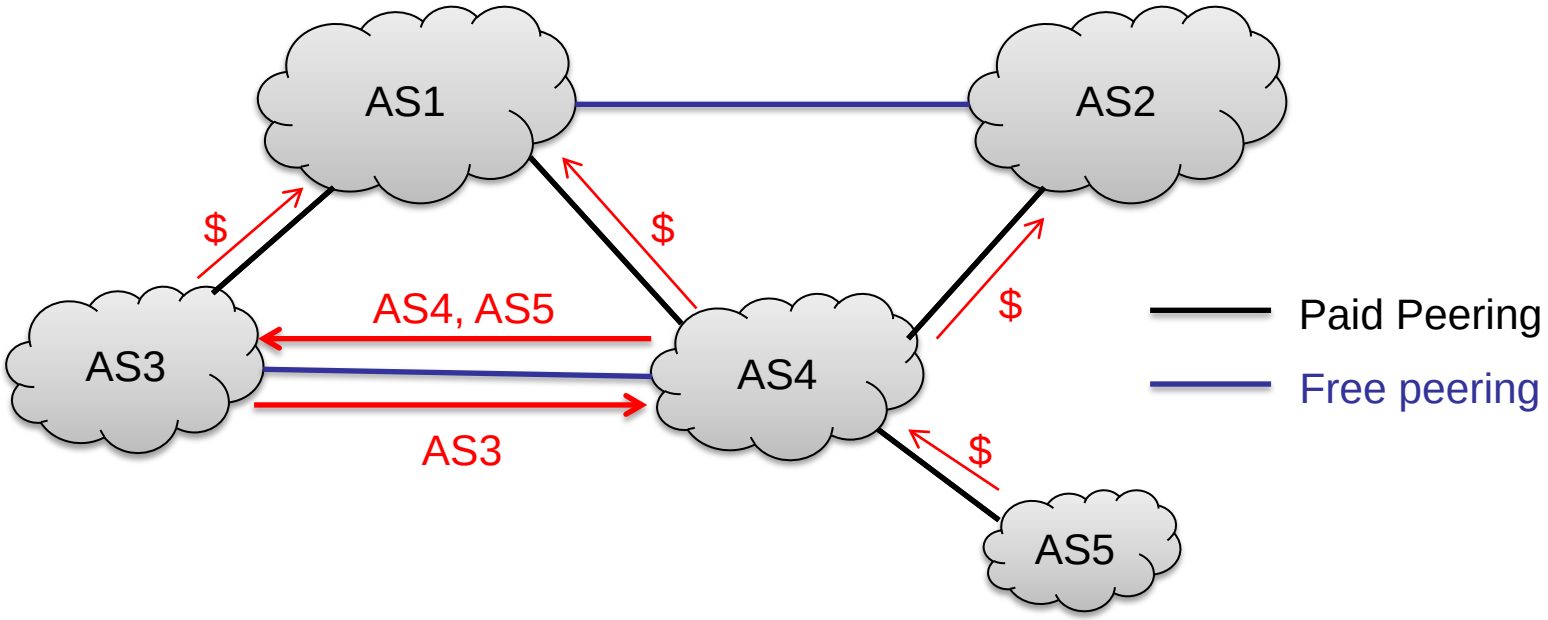
Settlement-Free Peering

- AS4 advertises to AS3 its internal routes and the routes learned from its own customers
  - AS3 will use the shared link to reach AS4 and AS4's customers
  - AS4's providers are nor reachable via the shared link



Settlement-Free Peering

- AS4 advertises to AS3 its internal routes and the routes learned from its own customers
  - AS3 will use the shared link to reach AS4 and AS4's customers
  - AS4's providers are nor reachable via the shared link



Routing Policies

- An AS specifies its routing policy which is composed of:
  - Import filter:
    - Specifies which routes can be accepted by the router among all the received routes from a given peer
  - Export filter:
    - Specifies which routes can be advertised by the router to a given peer
  - Ranking algorithm:
    - Used to select the best route among all routes that the domain has received towards the same destination prefix
- Filters can be defined using the Routing Policy Specification Language (RPSL) [RFC 2622, 2650]

Routing Policy Specification Language (RPSL)

- Import filter:
  - Syntax:
    - `Import: from AS# accept list_of_ASes`
  - Examples:
    - `Import: from AS10 accept AS10 AS25`
    - `Import: from Provider accept ANY`

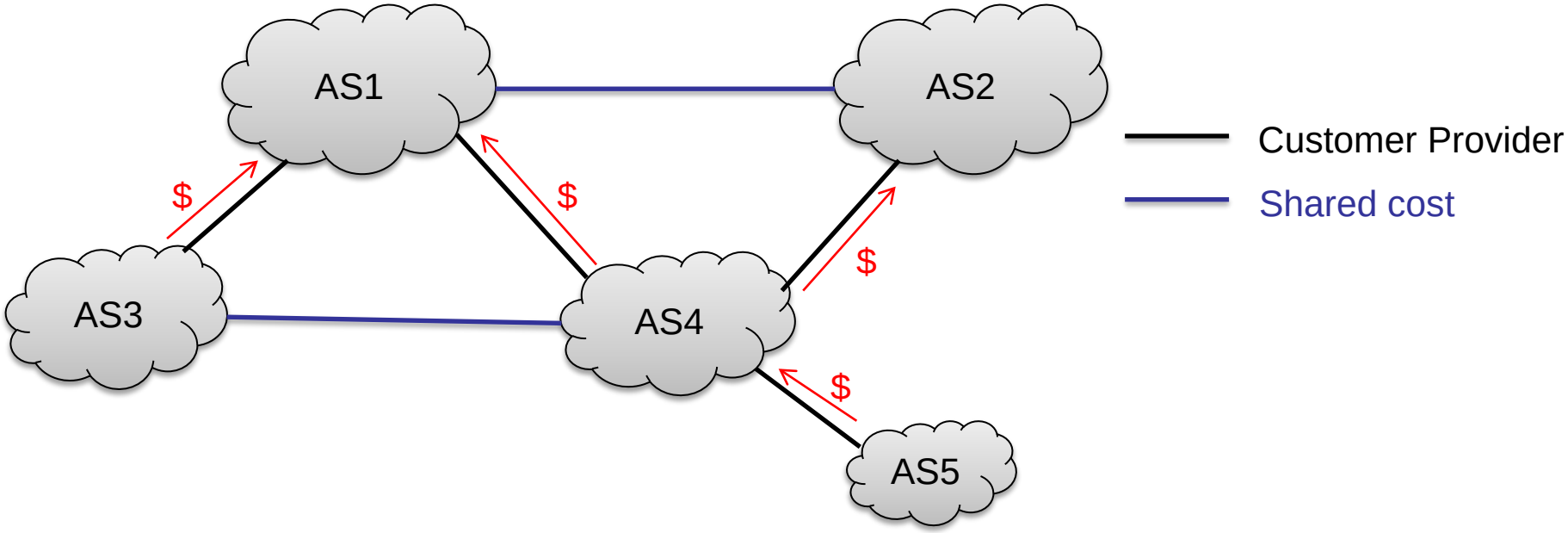
- Export filter:
  - Syntax:
    - `Export: to AS# announce list_of_ASes`
  - Examples:
    - `Export: to Customer announce ANY`
    - `Export: to AS5 announce AS1 AS3`

Bogon Filters

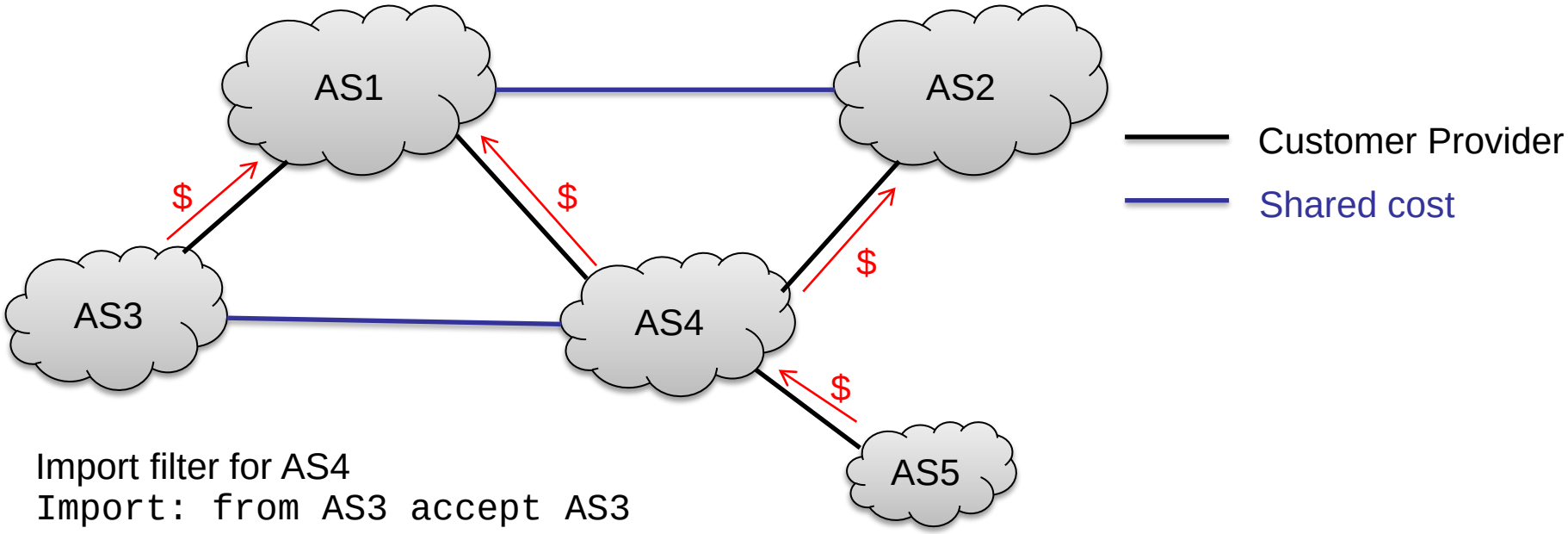
- Filters should ensure that bogon routes are not advertised on the Internet:

- Bogon routes include:
  - private IPv4 prefixes (192.168.0.0/16, 172.16.0.0/12, 10.0.0.0/8)
  - loopback prefix (127.0.0.0/8)
  - IP prefixes that have not yet been allocated by Internet Assigned Numbers Authority (IANA)

RPSL Example



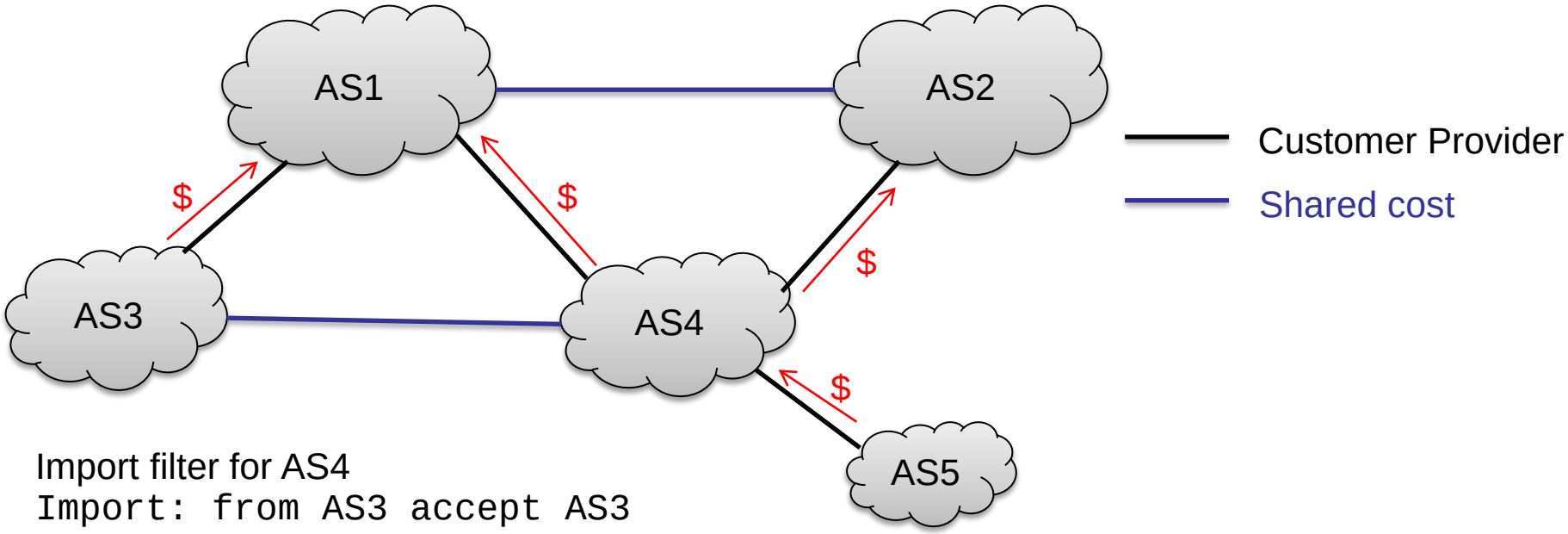
RPSL Example



Import filter for AS4

```
Import: from AS3 accept AS3
Import: from AS5 accept AS5
Import: from AS1 accept ANY
Import: from AS2 accept ANY
```

RPSL Example



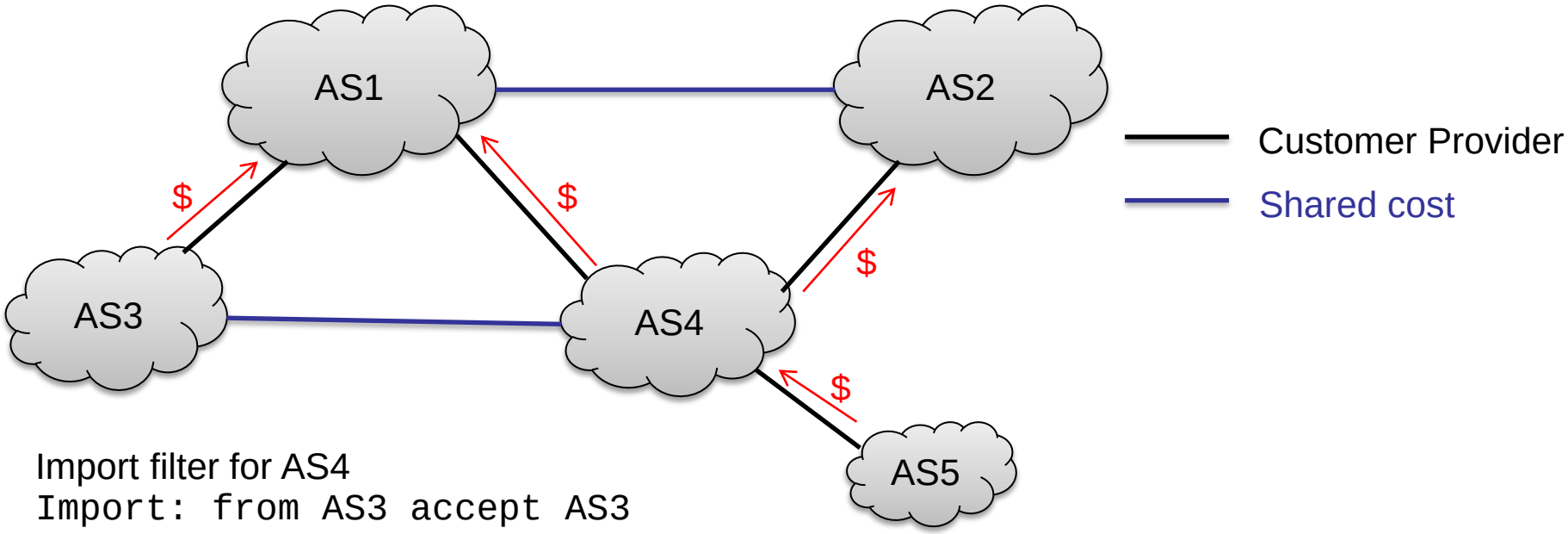
Import filter for AS4

```
Import: from AS3 accept AS3
Import: from AS5 accept AS5
Import: from AS1 accept ANY
Import: from AS2 accept ANY
```

Export filter for AS4

```
Export: to AS3 announce AS4, AS5
Export: to AS5 announce ANY
Export: to AS1 announce AS4, AS5
Export: to AS2 announce AS4, AS5
```

RPSL Example



Import filter for AS4

```
Import: from AS3 accept AS3
Import: from AS5 accept AS5
Import: from AS1 accept ANY
Import: from AS2 accept ANY
```

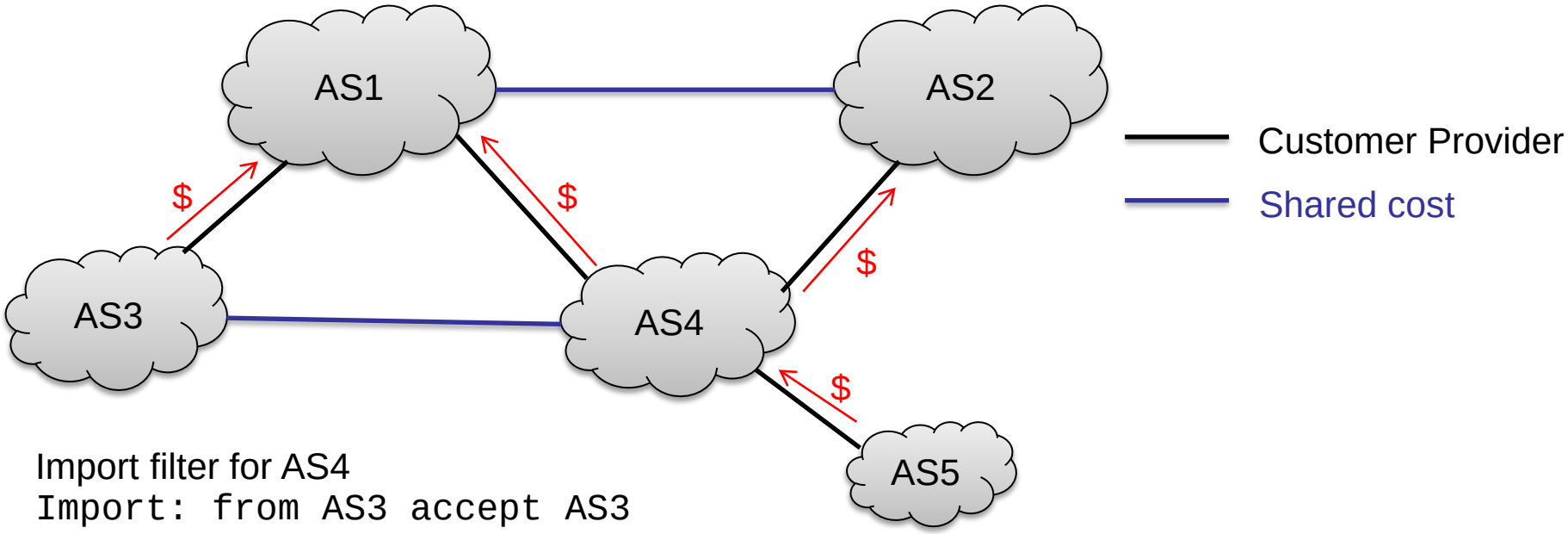
Export filter for AS4

```
Export: to AS3 announce AS4, AS5
Export: to AS5 announce ANY
Export: to AS1 announce AS4, AS5
Export: to AS2 announce AS4, AS5
```

Import filter for AS5

```
Import: from AS4 accept ANY
```

RPSL Example



Import filter for AS4

```
Import: from AS3 accept AS3
Import: from AS5 accept AS5
Import: from AS1 accept ANY
Import: from AS2 accept ANY
```

Export filter for AS4

```
Export: to AS3 announce AS4, AS5
Export: to AS5 announce ANY
Export: to AS1 announce AS4, AS5
Export: to AS2 announce AS4, AS5
```

Import filter for AS5

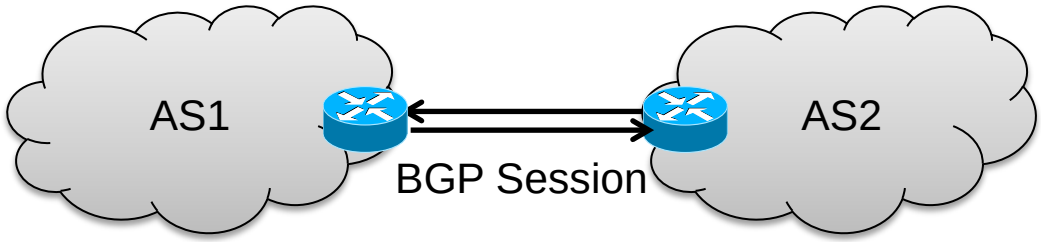
```
Import: from AS4 accept ANY
```

Export filter for AS5

```
Export: to AS4 announce AS5
```

Border Gateway Protocol (BGP)

- BGP is a path-vector protocol that allows ASes to exchange routes:
  - BGP routers exchange routes over BGP sessions
    - A BGP session is established between two routers that belong to two different domains and are directly connected
      - For security reasons, a router never establishes a BGP session that has not been manually configured on the router
    - BGP uses TCP for data delivery

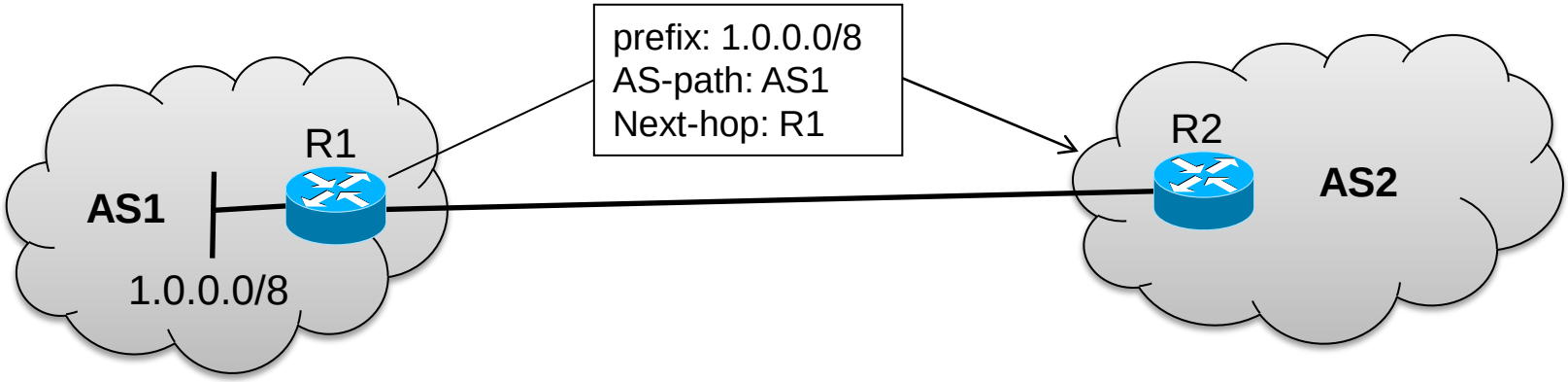


BGP Messages

- OPEN: initialises the BGP session and allows the negotiation of some options
- NOTIFICATION: used to terminate a BGP session, usually because an error has been detected by the BGP peer
  - A router that sends or receives a NOTIFICATION message immediately shutdowns the corresponding BGP session
- UPDATE: used to advertise new or modified routes or to withdraw previously advertised routes
- KEEPALIVE: used to ensure a regular exchange of messages on the BGP session, even when no route changes
  - When a BGP router has not sent an UPDATE message during the last 30 seconds, it will send a KEEPALIVE message to confirm to the other peer that it is still up
  - If a peer does not receive any BGP message during a period of 90 seconds, the BGP session is considered to be down and all the routes learned over this session are withdrawn

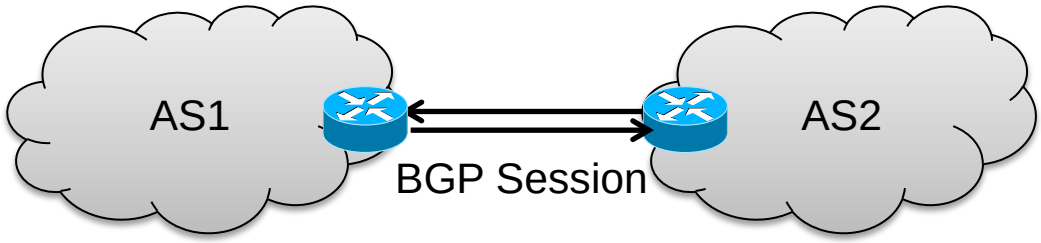
BGP UPDATE

- An UPDATE message is used to advertise or withdraw a route
- An UPDATE message contains:
  - A list of prefixes that are advertised or withdrawn
  - The AS-path (e.g., AS3:AS14:AS7)
  - The IP address of the next-hop router



BGP Operation

- BGP relies on the incremental exchange of path vectors:
  1. BGP session is initially established over TCP connection between peers
  2. Each peer announces all its active routes
  3. BGP routing tables are incrementally updated (routes are added, modified or withdrawn)
  4. If BGP session is open GOTO 3



BGP Session Initialization

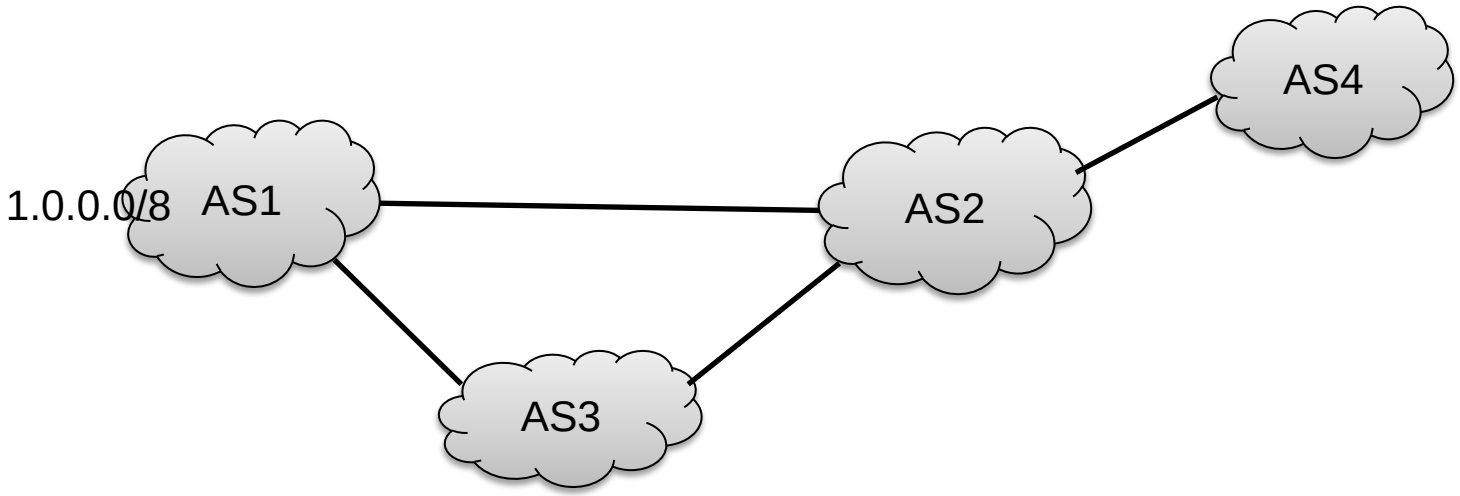

```
Initialize_BGP_Session(RemoteAS, RemoteIP)
{
   /* Initialize and start BGP session */
   /* Send BGP OPEN Message to RemoteIP on port 179 */
   /* Follow BGP state machine */
   /* advertise local routes and routes learned from peers*/
   foreach (destination = d inside BGP Loc-RIB)
   {
      B = build_BGP_UPDATE(d); // B: best path
      S = apply_export_filter(RemoteAS, B);
      if (S <> NULL)
      {  /* send UPDATE message */
         send_UPDATE(S, RemoteAS, RemoteIP)
      }
   }
   /* entire RIB has been sent */
   /* new UPDATE will be sent only to reflect local or distant
      changes in routes */
   ...
}
```

Events during a BGP Session

- Addition of a new route to the routing table:
  - Internal route added manually or announced via IGP from a local router
  - Reception of an UPDATE message from a peer announcing a new or modified inter-domain route
- Removal of a route from the routing table:
  - Internal route removed manually or declared unreachable via IGP
  - Reception of an UPDATE message from a peer declaring an inter-domain route unreachable
- BGP session loss:
  - All routes learnt from this peer are removed from the routing table

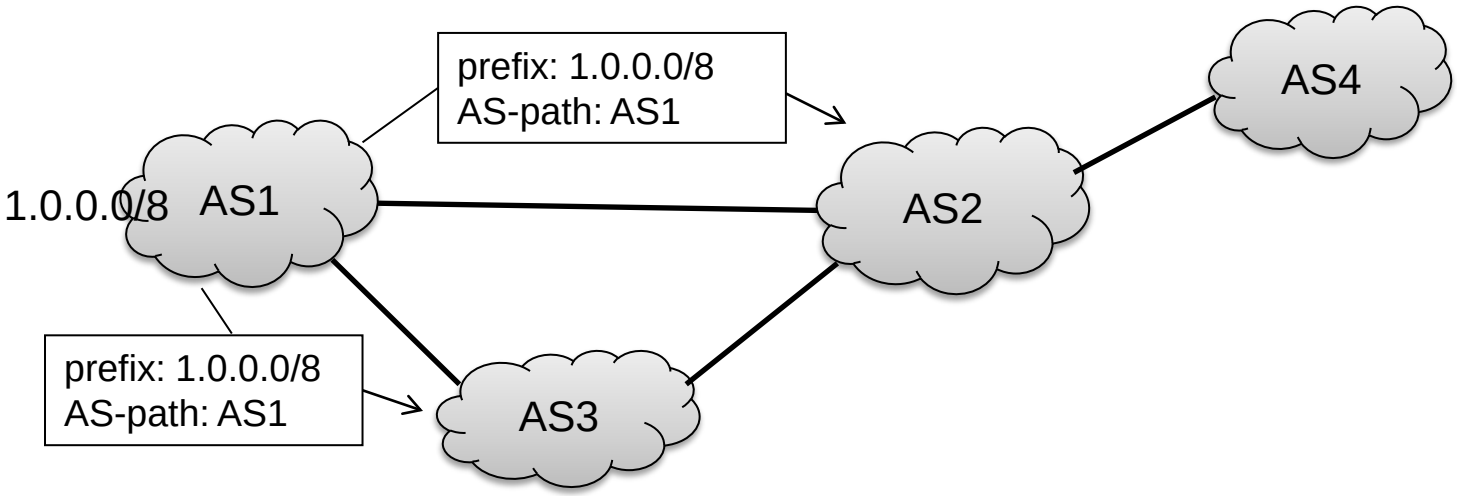
BGP Operation Example

- BGP routers exchange AS-path vectors:



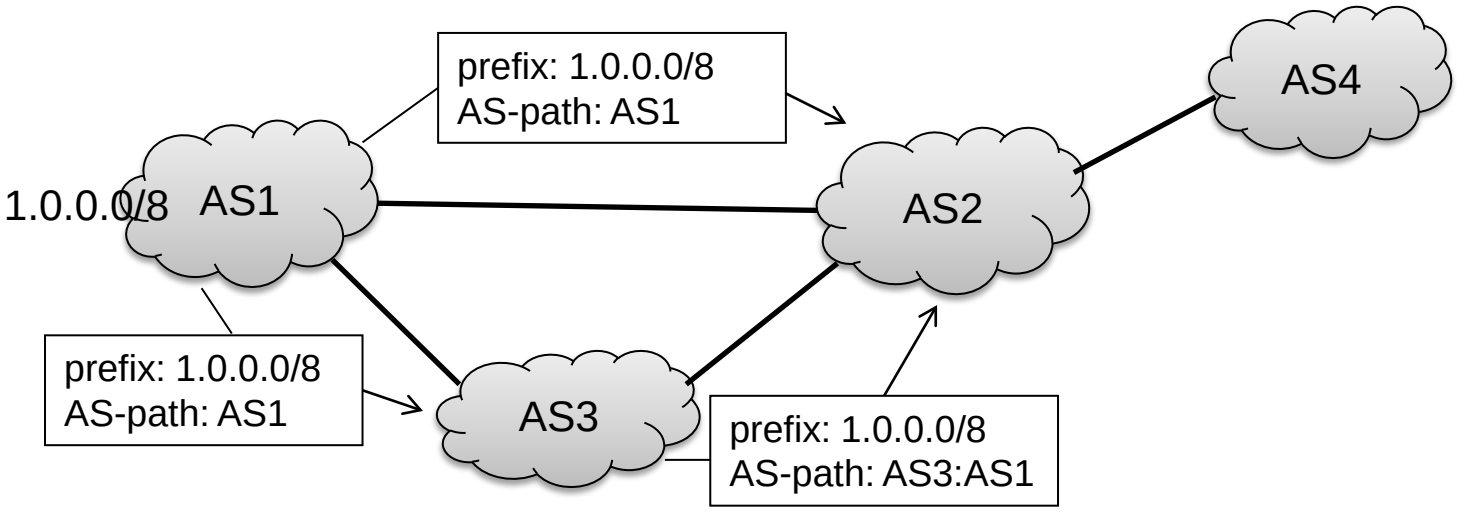
BGP Operation Example

- BGP routers exchange AS-path vectors:
  - AS1 advertises its prefix 1.0.0.0/8



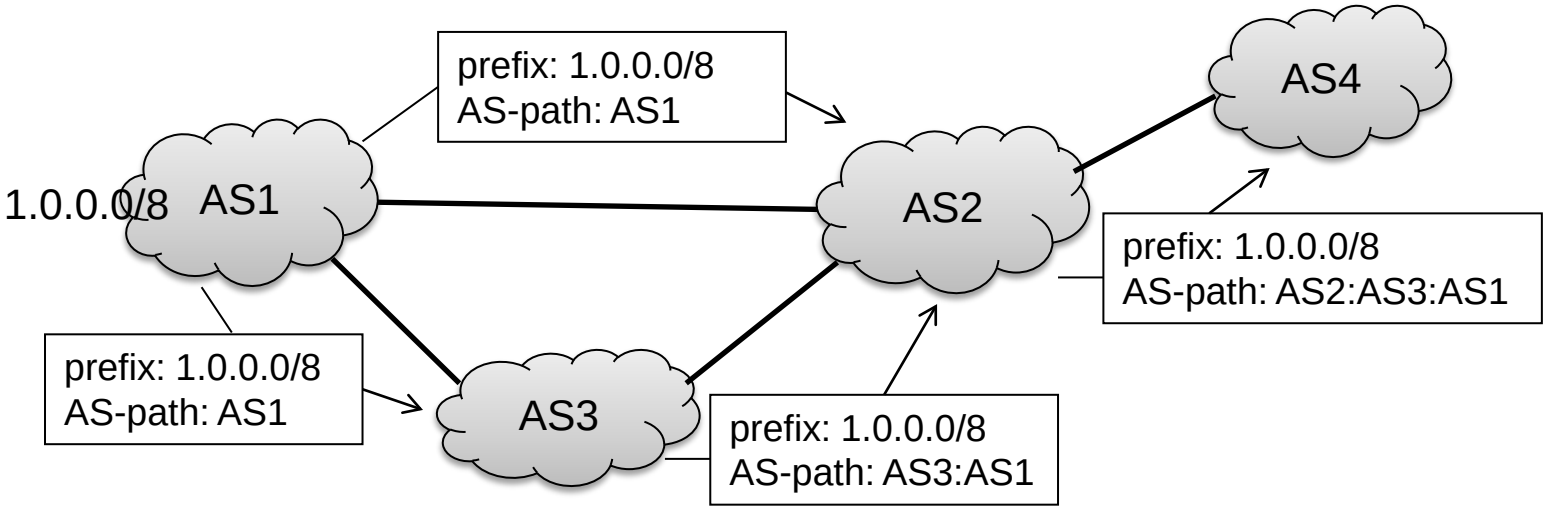
BGP Operation Example

- BGP routers exchange AS-path vectors:
  - AS1 advertises its prefix 1.0.0.0/8
  - AS2 learns 2 routes for this prefix

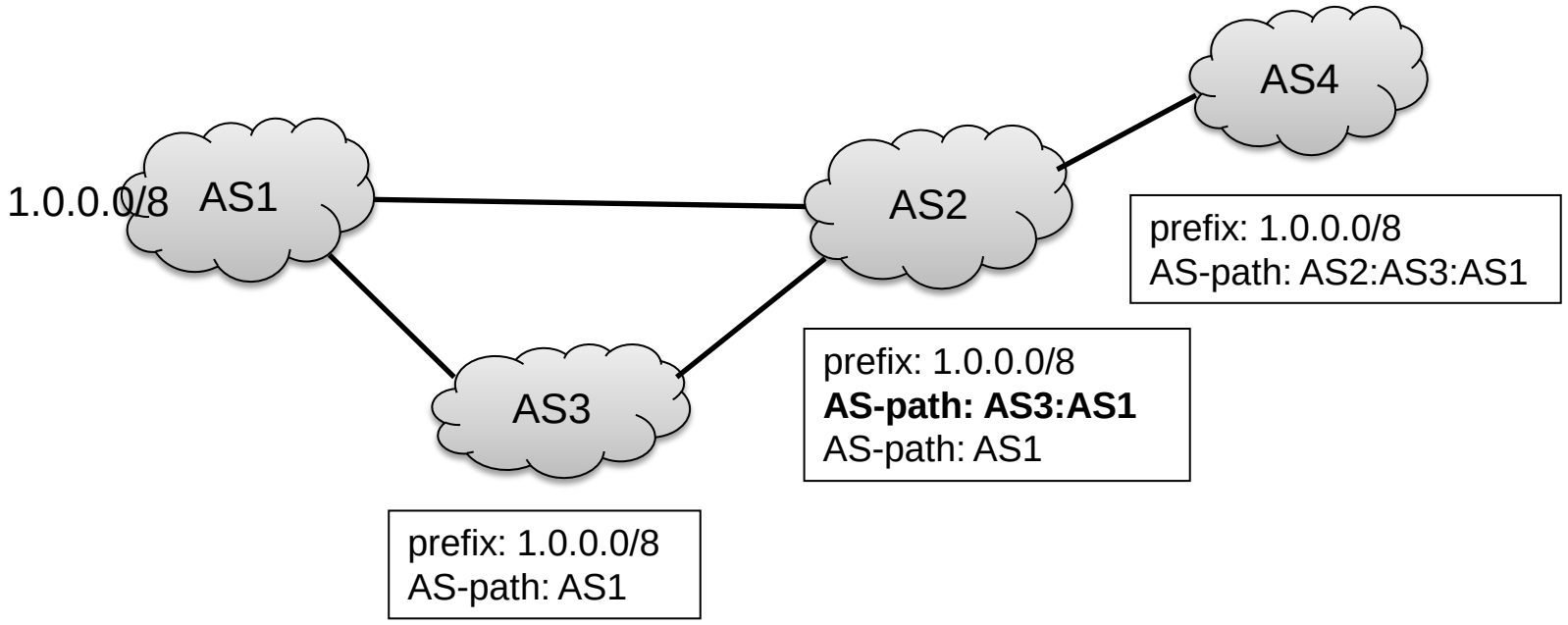


BGP Operation Example

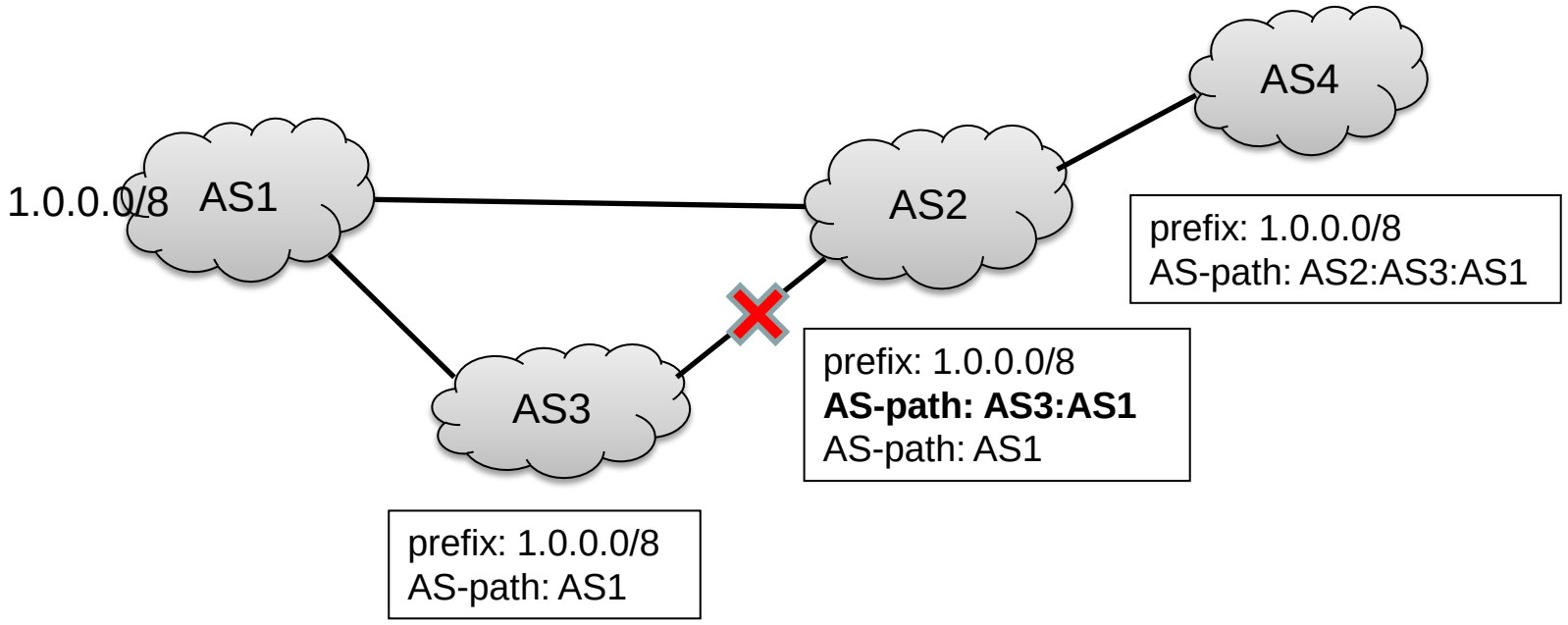
- BGP routers exchange AS-path vectors:
  - AS1 advertises its prefix 1.0.0.0/8
  - AS2 learns 2 routes for this prefix
  - AS2 prefers the route learnt from AS3 and announces this route to AS4



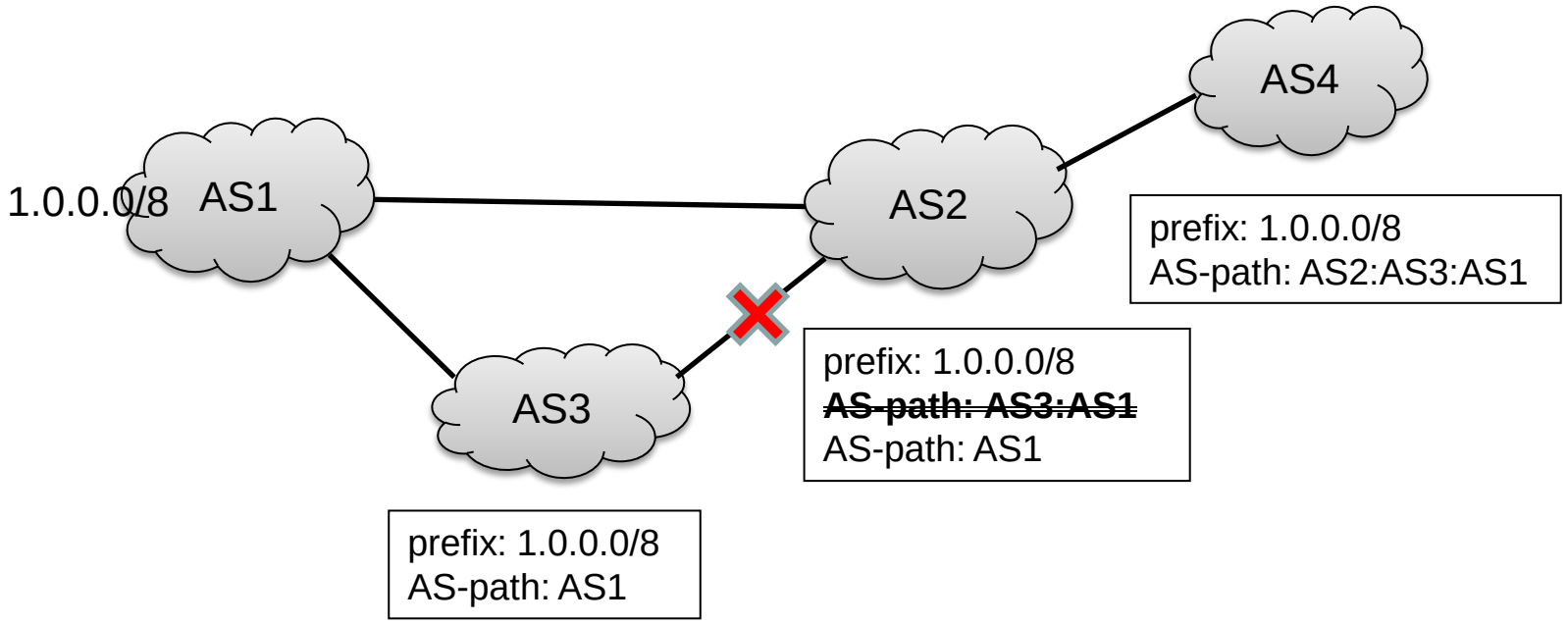
BGP Reaction to Link Failure



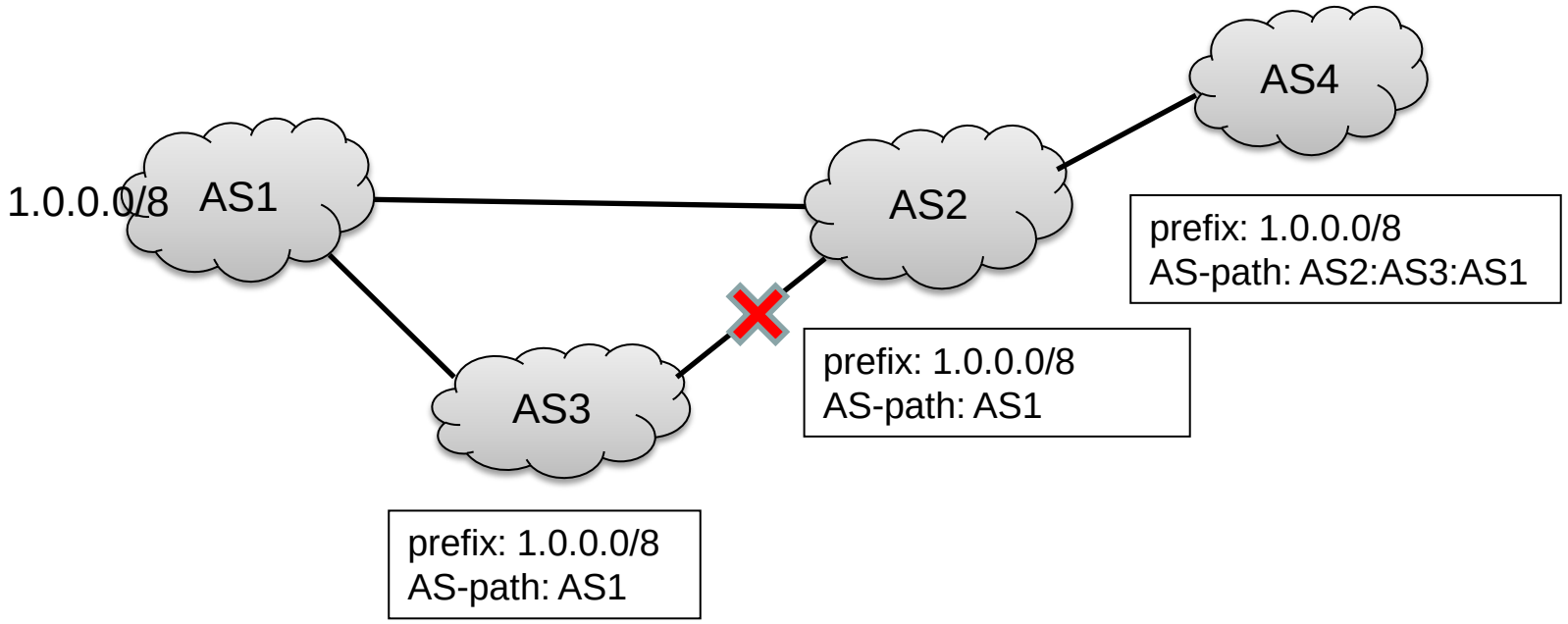
BGP Reaction to Link Failure



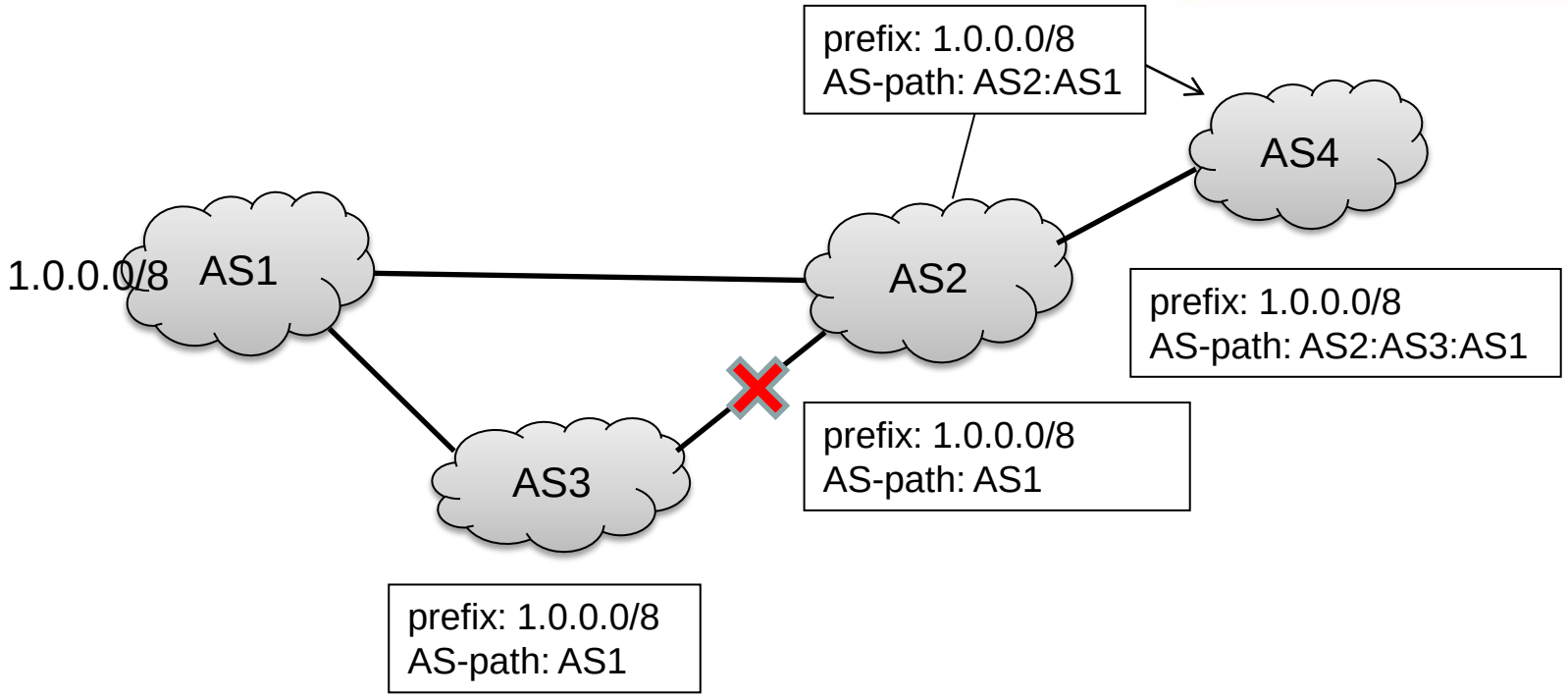
BGP Reaction to Link Failure



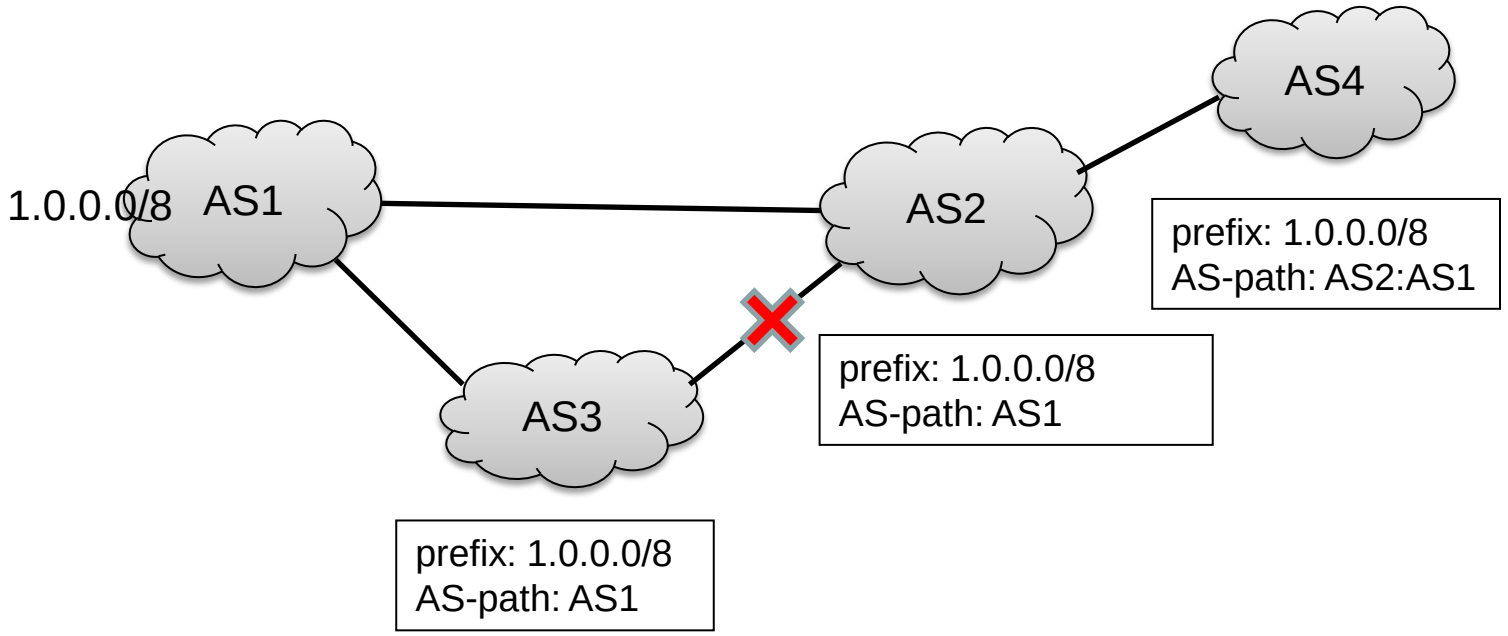
BGP Reaction to Link Failure



BGP Reaction to Link Failure

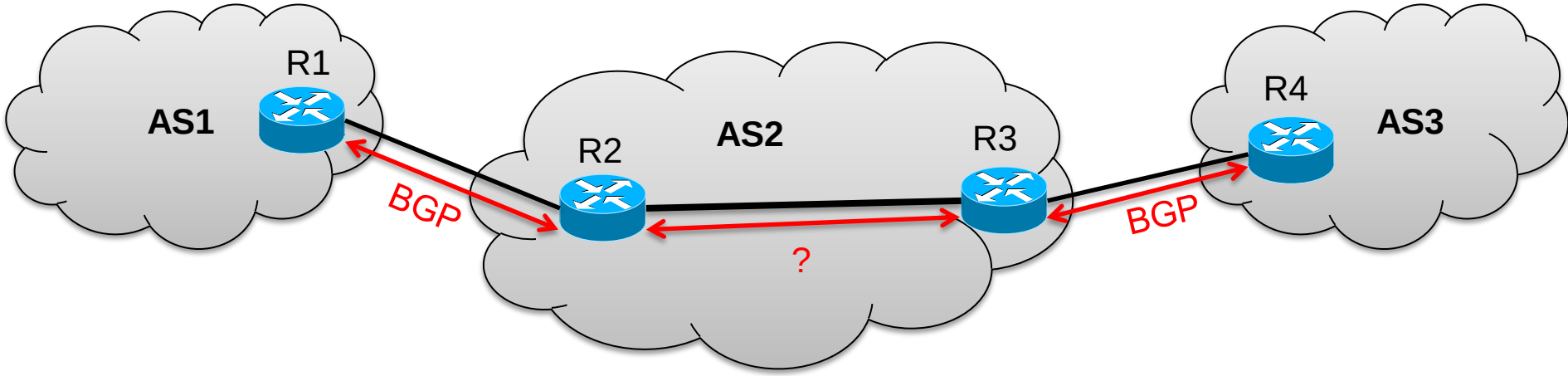


BGP Reaction to Link Failure



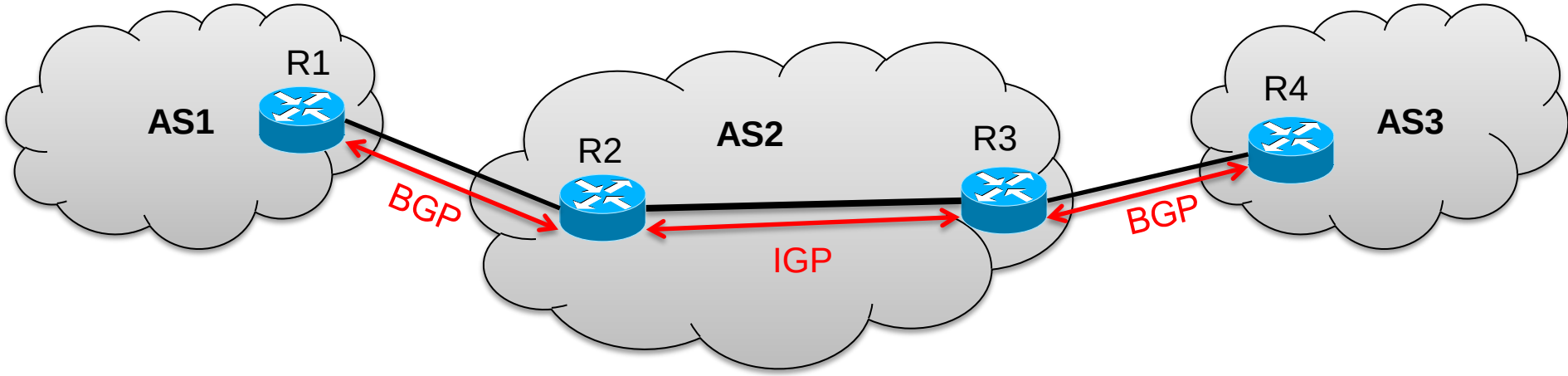
Exchanging Routes within Domains

- Within a domain, multiple routers (e.g., R2, R3) may establish BGP sessions with peers
- How should these routers exchange the routes learnt from their peers via BGP?



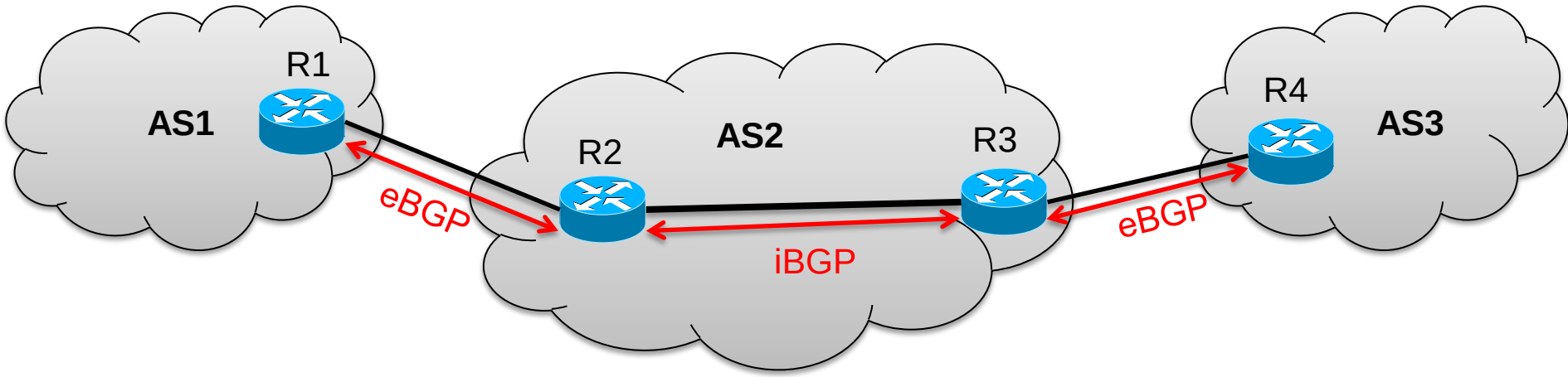
Exchanging Routes within Domains with IGP

- IGP (OSPF, IS-IS or RIP) could be used to advertise within the domain the routes learnt from peers
  - IGP does not support BGP attributes, such as AS-Path
  - IGP is not designed to support hundreds of thousands of routes that a BGP router can receive from peers



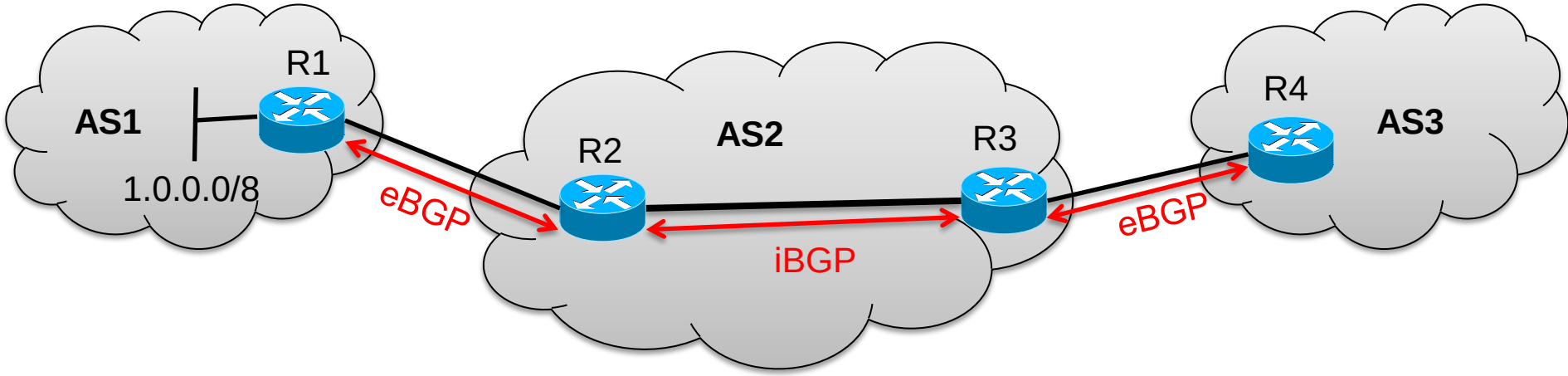
Exchanging Routes within Domains with BGP

- Instead of IGP, BGP can be used to exchange routes between the routers of a domain
- Routers establish two types of BGP sessions:
  - eBGP sessions between routers that belong to different domains
  - iBGP sessions between routers that belong to the same domain



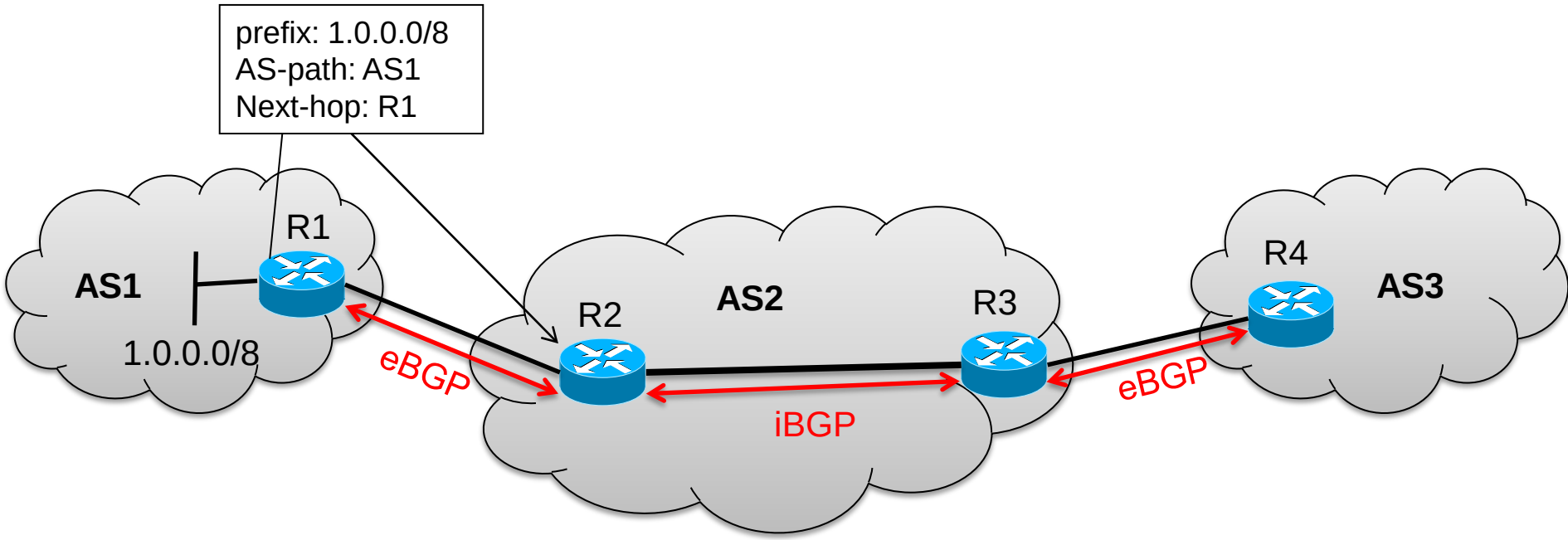
Route Announcement with eBGP and iBGP

- BGP next-hop and AS-path are updated only when a route is advertised over an eBGP session



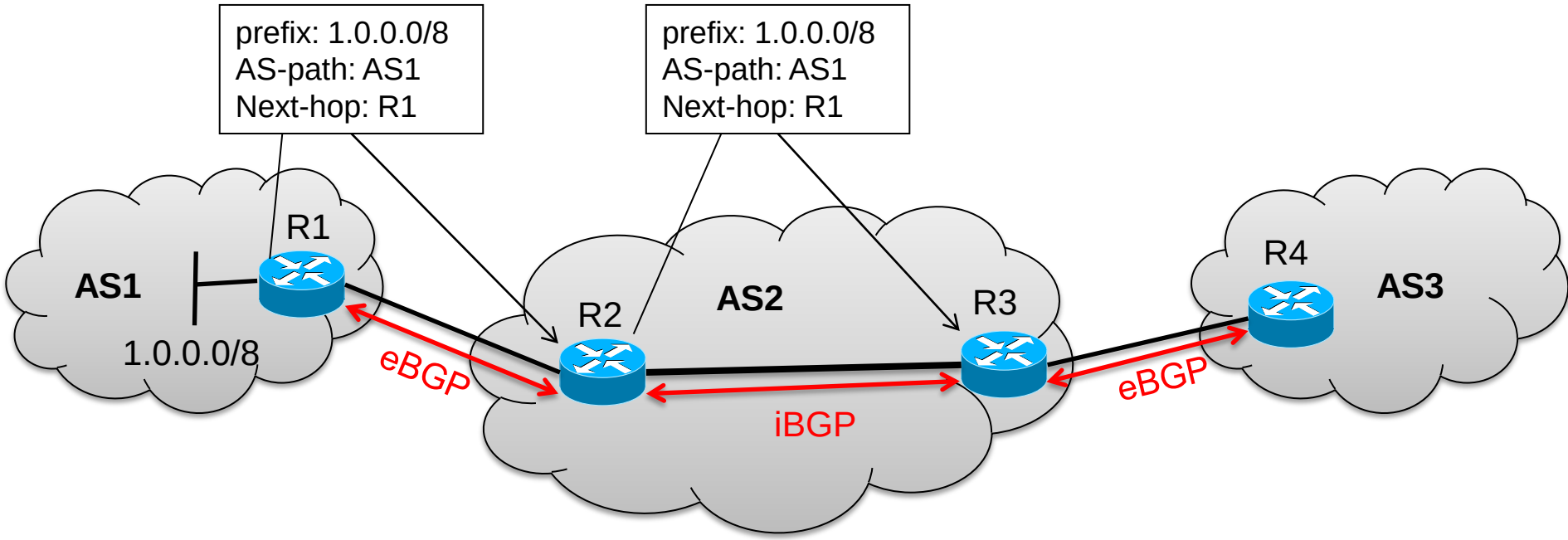
Route Announcement with eBGP and iBGP

- BGP next-hop and AS-path are updated only when a route is advertised over an eBGP session



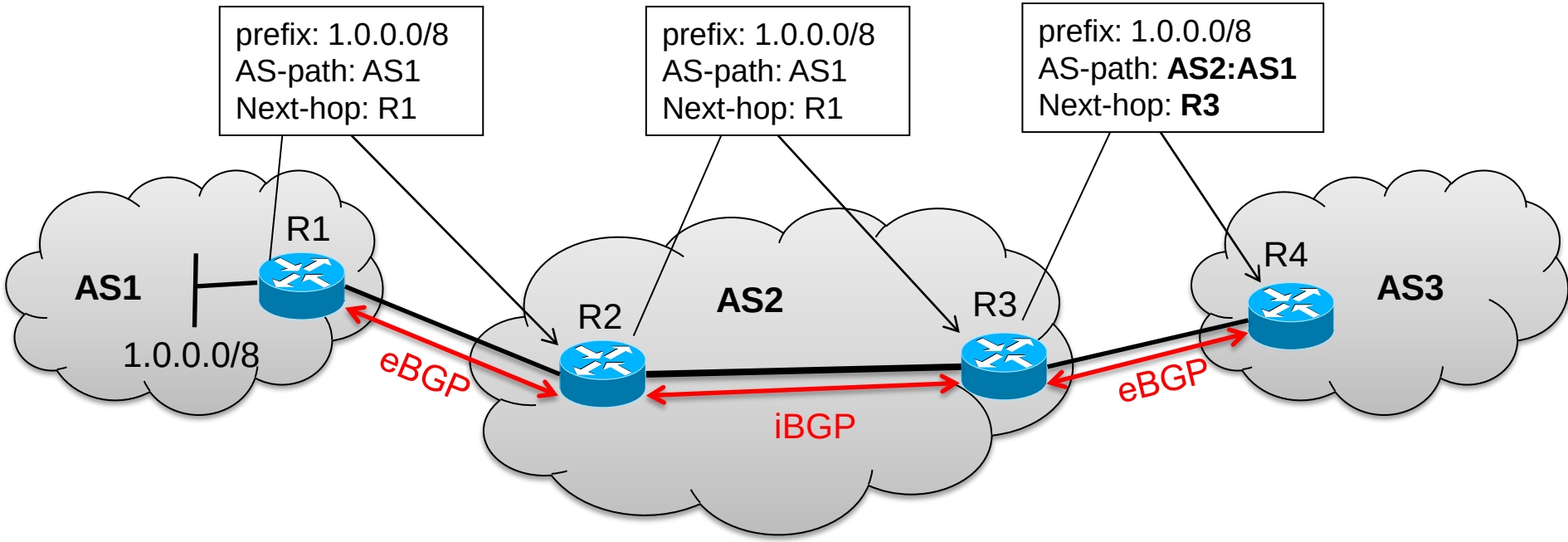
Route Announcement with eBGP and iBGP

- BGP next-hop and AS-path are updated only when a route is advertised over an eBGP session



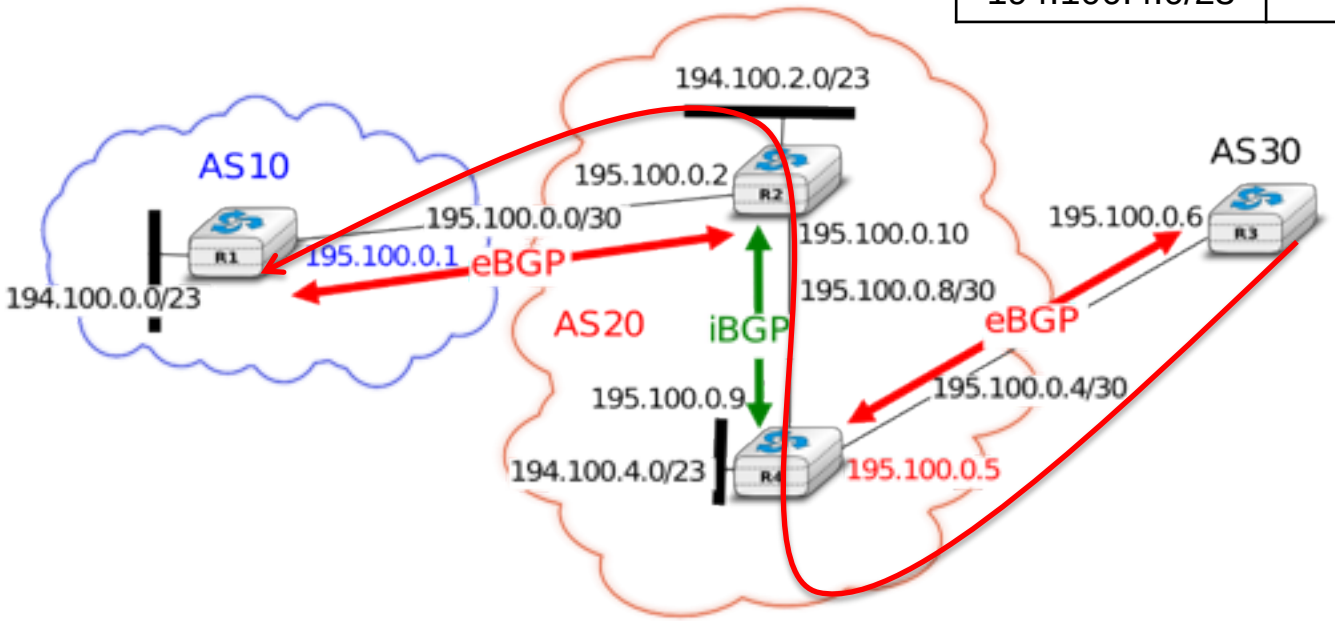
Route Announcement with eBGP and iBGP

- BGP next-hop and AS-path are updated only when a route is advertised over an eBGP session



BGP Routes for R4	
Prefix	**Next-hop**
194.100.0.0/23	195.100.0.1

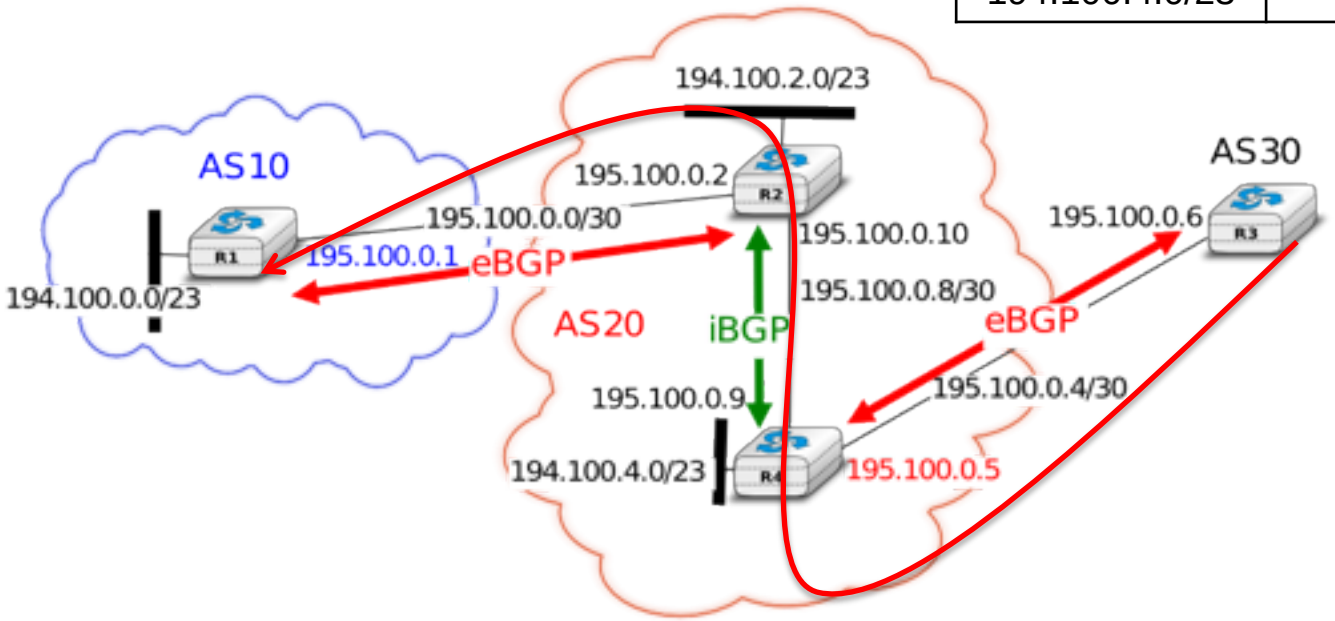
IGP Routes for R4	
Prefix	**Next-hop**
195.100.0.0/30	195.100.0.10
195.100.0.4/30	East
195.100.0.8/30	North
194.100.2.0/23	195.100.0.10
194.100.4.0/23	West



Routing with IGP and BGP

BGP Routes for R4	
Prefix	**Next-hop**
194.100.0.0/23	195.100.0.1

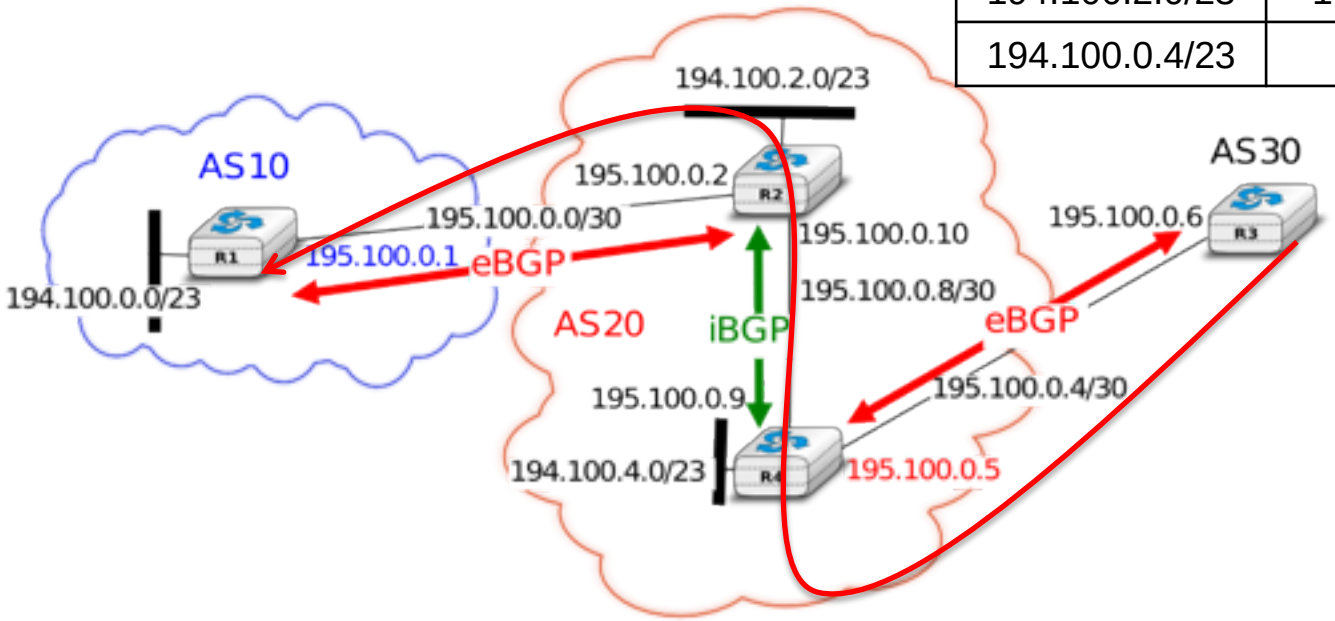
IGP Routes for R4	
Prefix	**Next-hop**
195.100.0.0/30	195.100.0.10
195.100.0.4/30	East
195.100.0.8/30	North
194.100.2.0/23	195.100.0.10
194.100.4.0/23	West



Routing with IGP and BGP

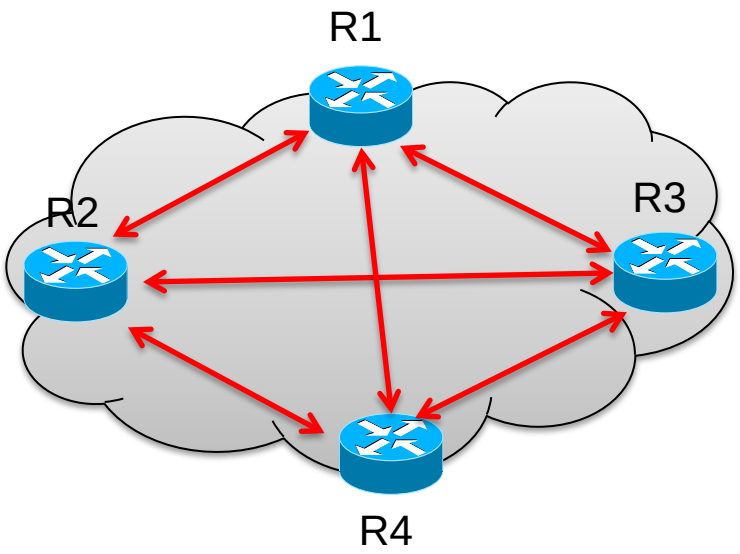
BGP Routes for R4	
Prefix	**Next-hop**
194.100.0.0/23	195.100.0.1

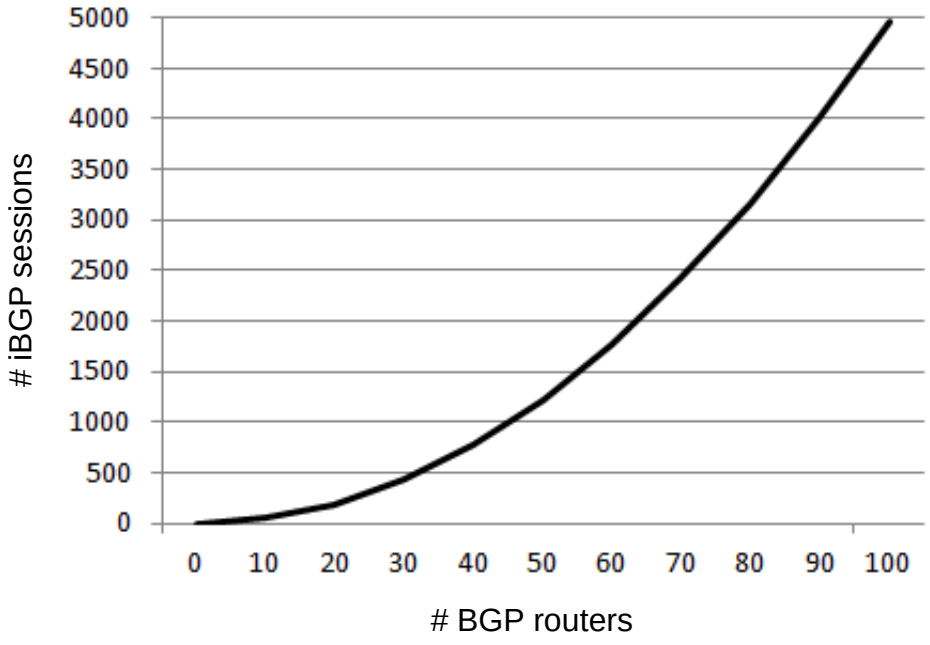
Forwarding Table for R4	
Prefix	**Next-hop**
194.100.0.0/23	195.100.0.10
195.100.0.0/30	195.100.0.10
195.100.0.4/30	East
195.100.0.8/30	North
194.100.2.0/23	195.100.0.10
194.100.0.4/23	West



iBGP Full Mesh

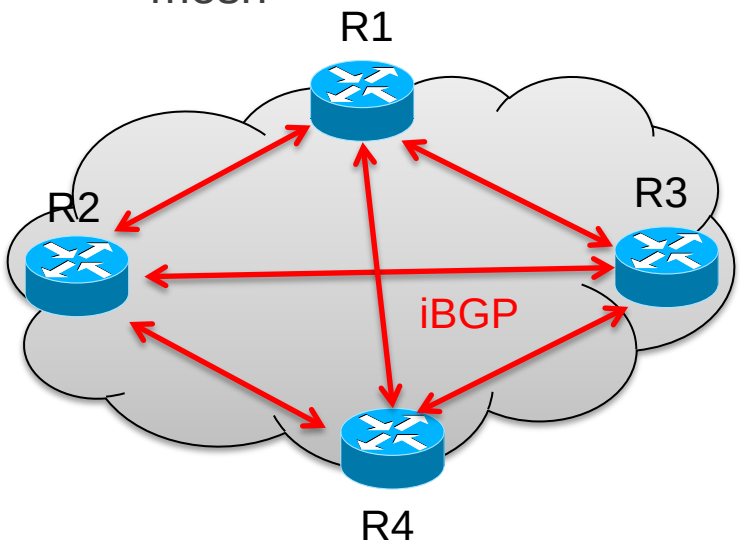
- Each BGP router maintains an iBGP session with each other BGP router within the same domain (full iBGP mesh)
  - N * (N-1) / 2 iBGP sessions for N BGP routers



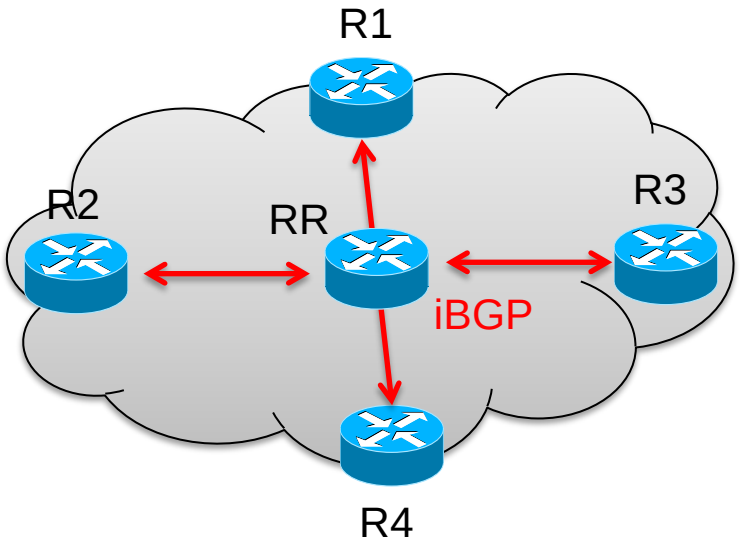


Route Reflectors

- Route reflector (RR) is a router that establishes iBGP sessions with a number of BGP routers in the AS
  - A RR will announce to other iBGP peers all routes learnt from their iBGP clients
- An AS may include multiple route reflectors
  - Multiple route reflectors within an AS peer with each other in a full mesh



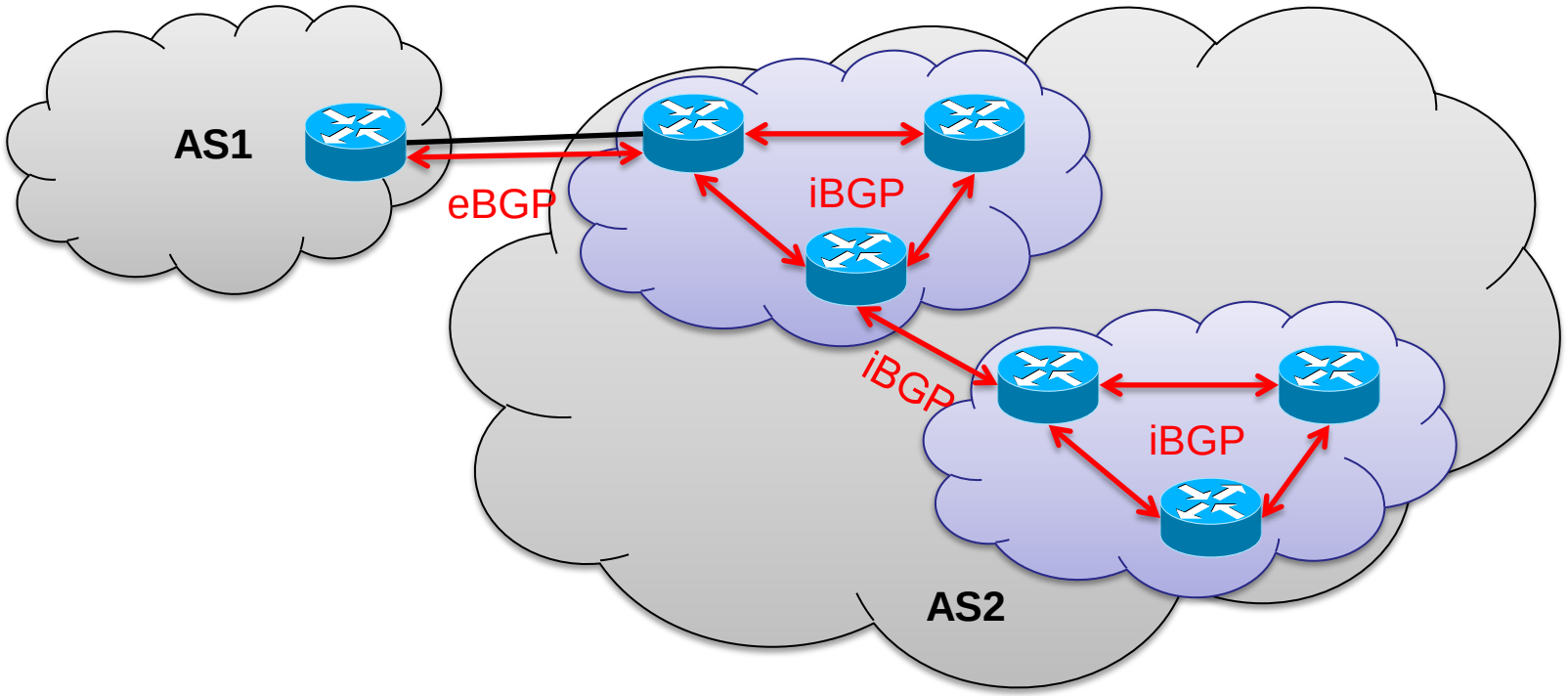
iBGP Full Mesh



Route Reflector

Confederations

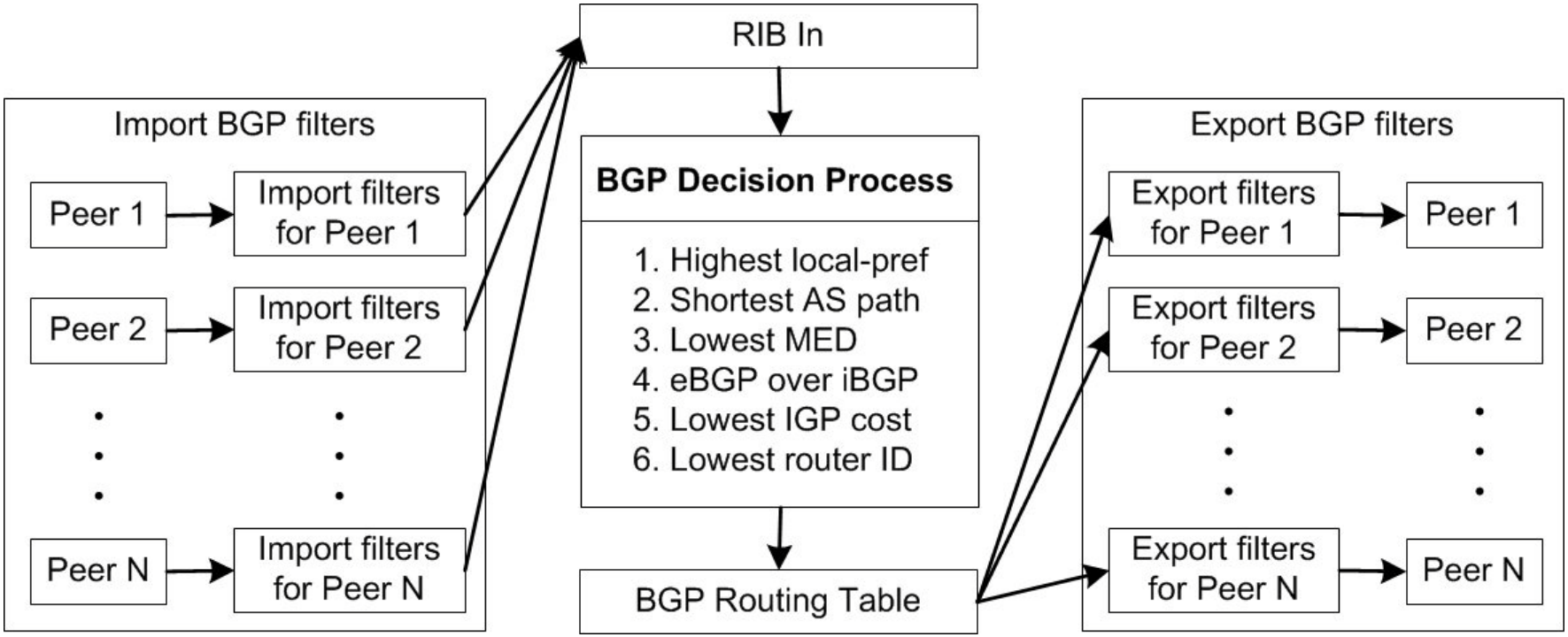
- Confederations reduce iBGP mesh inside an AS by subdividing it into multiple smaller sub-ASes
  - Confederation is not visible to other ASes which still see a single AS
  - BGP next-hop is preserved inside the confederation



Confederation

BGP Router Overview

- The RIB-In contains all internal routes and the acceptable routes learnt from all peers
- The BGP decision process selects the best route towards each destination
- The BGP routing table contains the best route for each destination

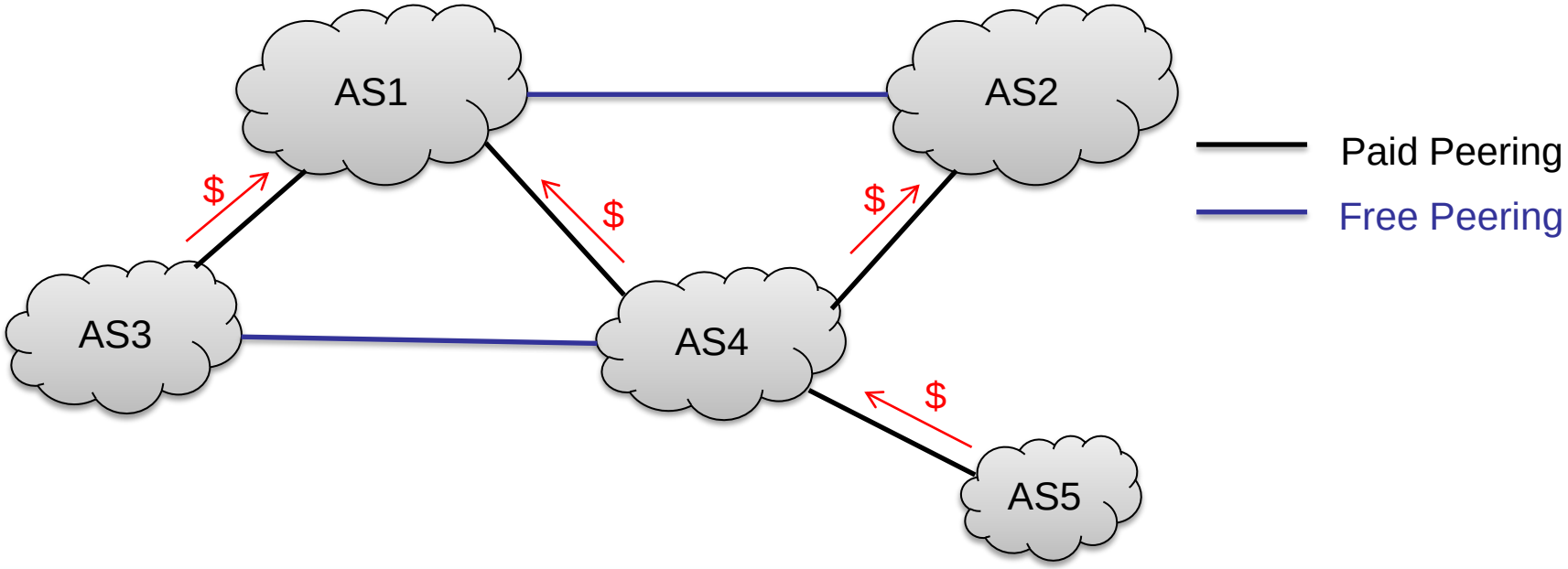


BGP Decision Process

- The BGP decision process may receive more than one route for a given destination
- Routes are removed from the candidate set in order to determine the best route, based on the following steps:
  1. Prefer routes with the highest `local-pref`
  2. Prefer routes with the shortest AS path
  3. Prefer routes with the smallest Multi-Exit Discriminator (MED)
  4. Prefer routes learnt via eBGP sessions over routes learnt via iBGP sessions
  5. Prefer routes with the most proximate next-hop (shortest intra-domain path)
  6. Tie-breaking rule: prefer routes learnt from the router with lowest router ID

Setting local-pref to Routes

- Network operators usually configure the import filters on their BGP routers as follows:
  - high `local-pref` (> 1000) for the routes learnt from a customer
  - medium `local-pref` (500 – 999) for the routes learnt over a shared-cost peering
  - low `local-pref` (< 500) for the routes learnt from a provider



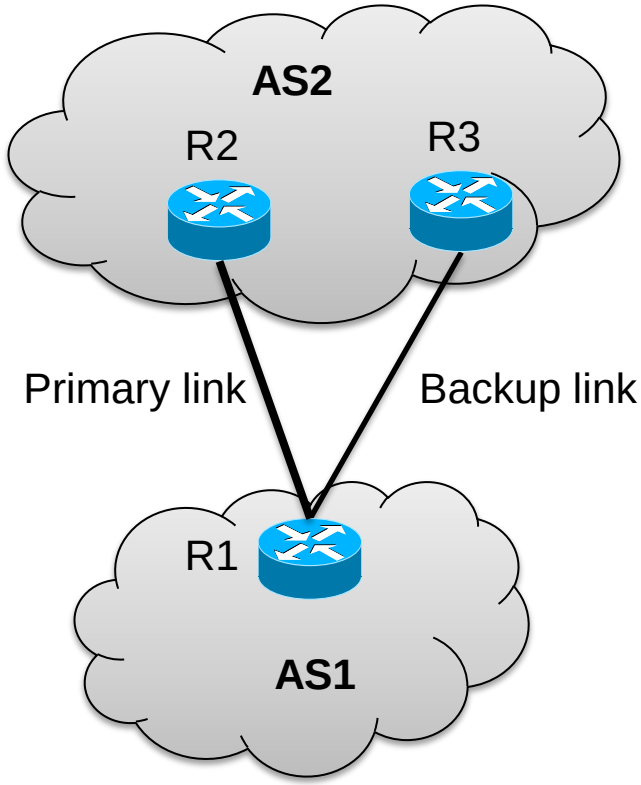
Configuring a Backup Link with BGP

Import filter for AS1

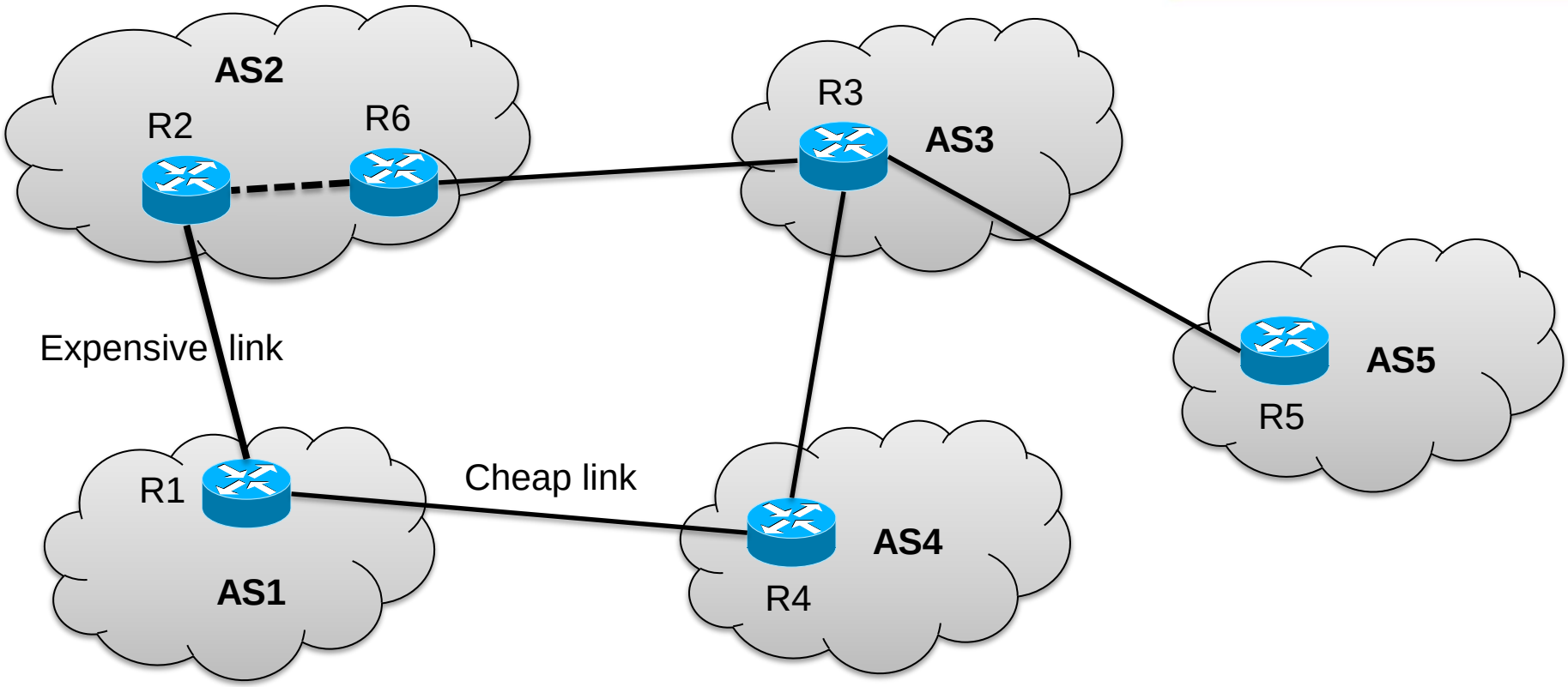
```
Import: from AS2 R2 at R1 set localpref=200;
        from AS2 R3 at R1 set localpref=100;
        accept ANY
```

Import filter for AS2

```
Import: from AS1 R1 at R2 set localpref=1200;
        from AS1 R1 at R3 set localpref=1100;
        accept AS1
```



Setting Link Preferences with BGP

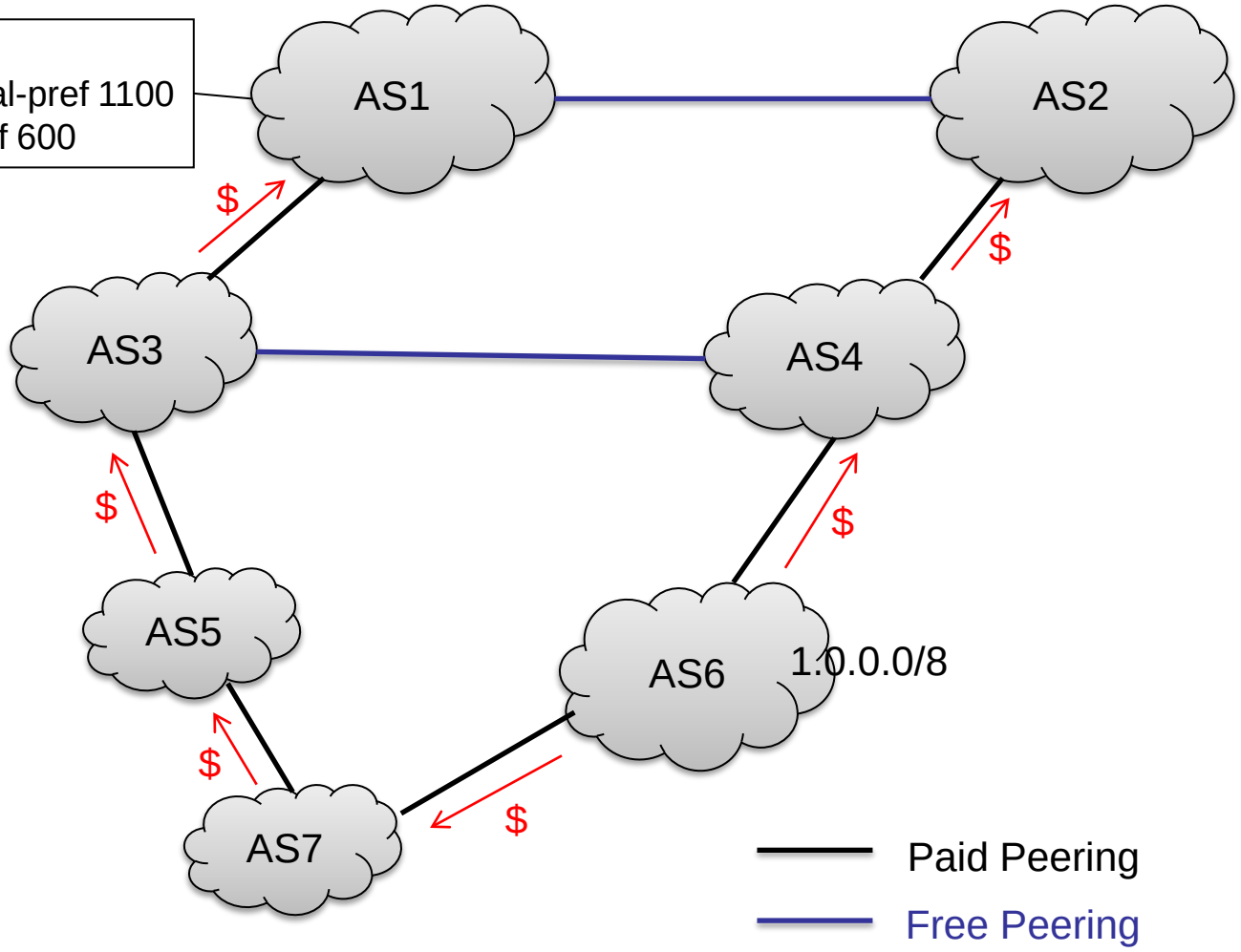


Import filter for AS1

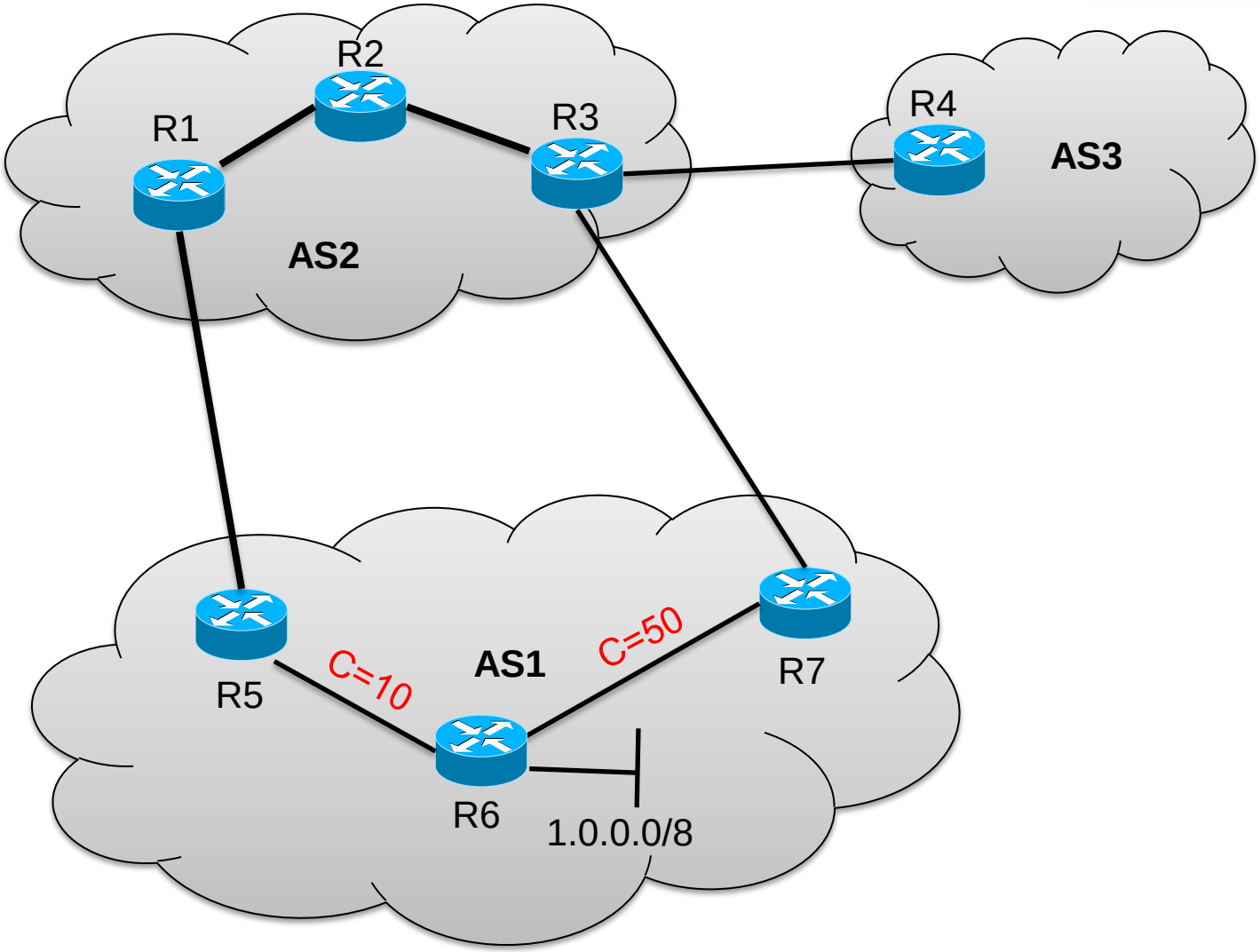
```
Import: from AS2 R2 at R1 set localpref=100;
        from AS4 R4 at R1 set localpref=600;
        accept ANY
```

Asymmetry of Internet Paths

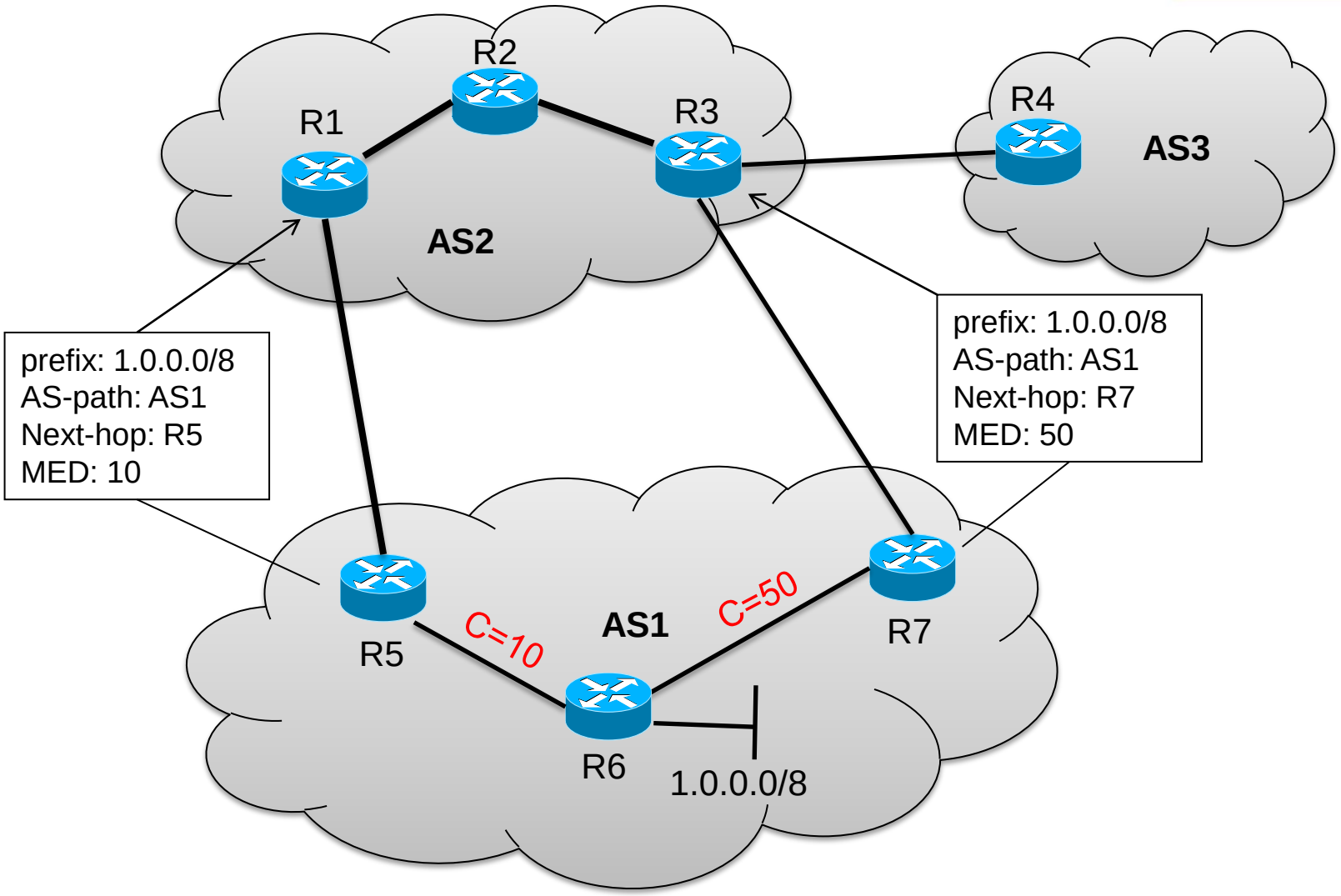
prefix: 1.0.0.0/8
AS-path: **AS3:AS5:AS7:AS6**, local-pref 1100
AS-path: **AS2:AS4:AS6**, local-pref 600



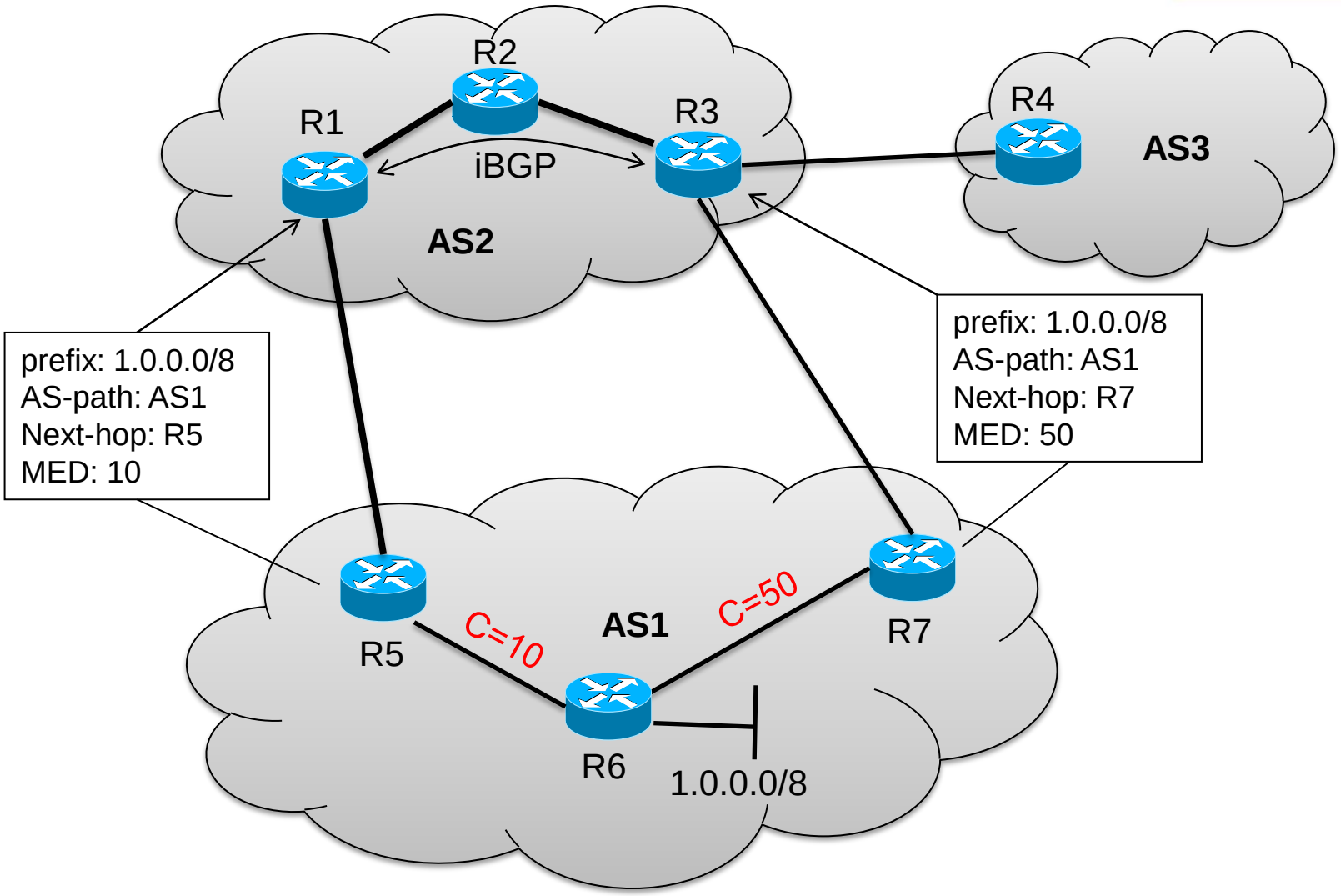
Multi-Exit Discriminator (MED)



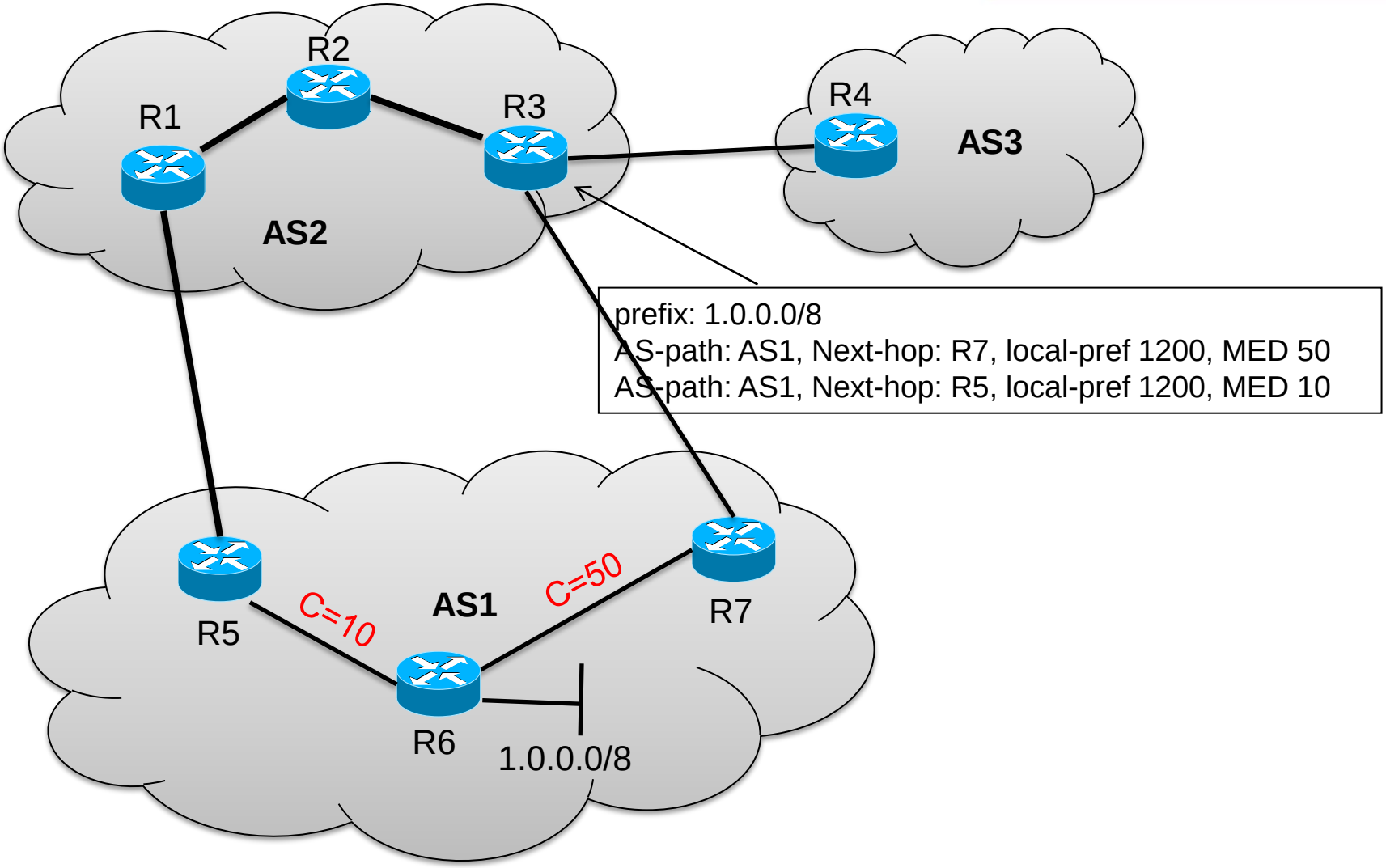
Multi-Exit Discriminator (MED)



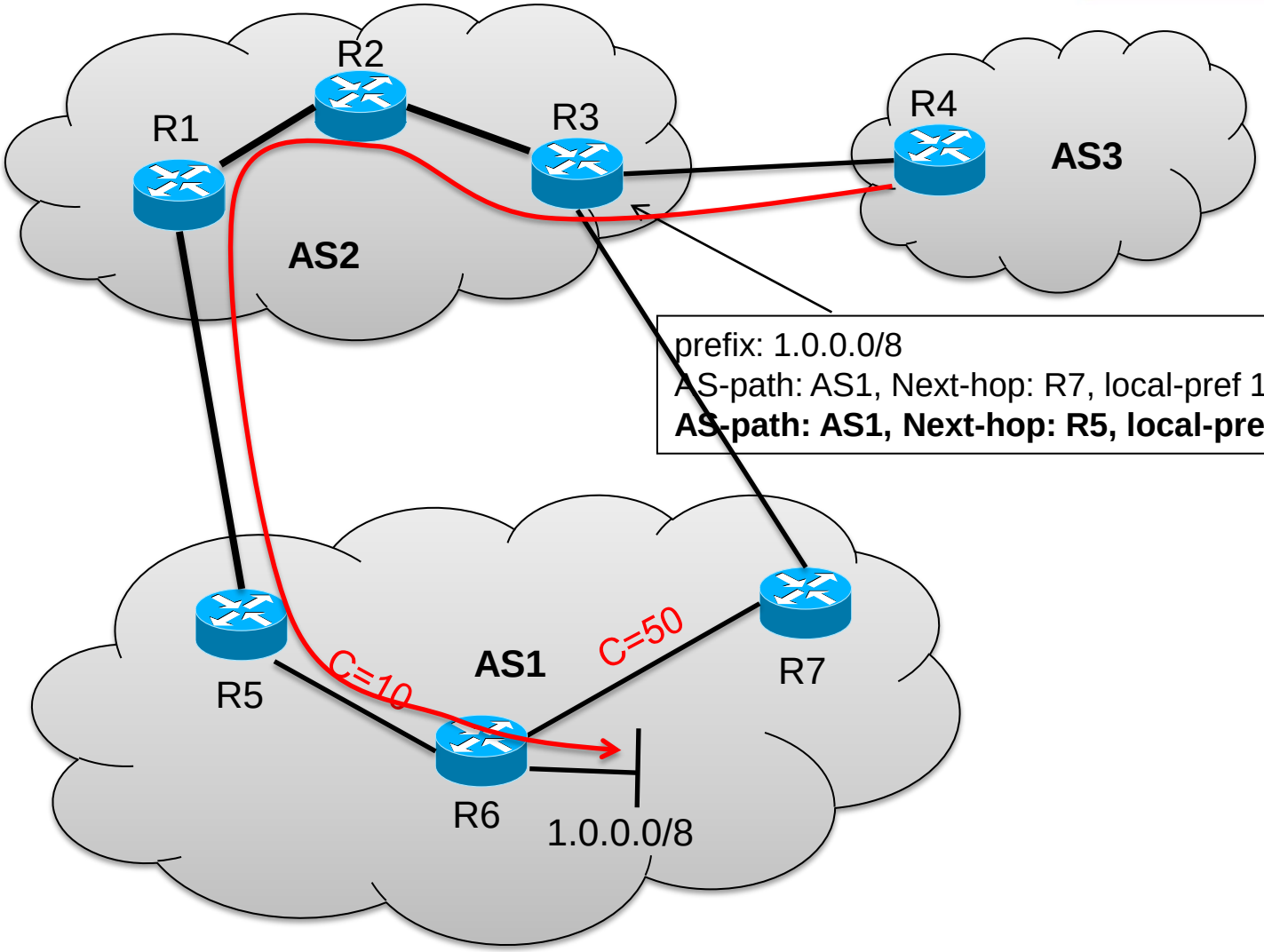
Multi-Exit Discriminator (MED)



Multi-Exit Discriminator (MED)

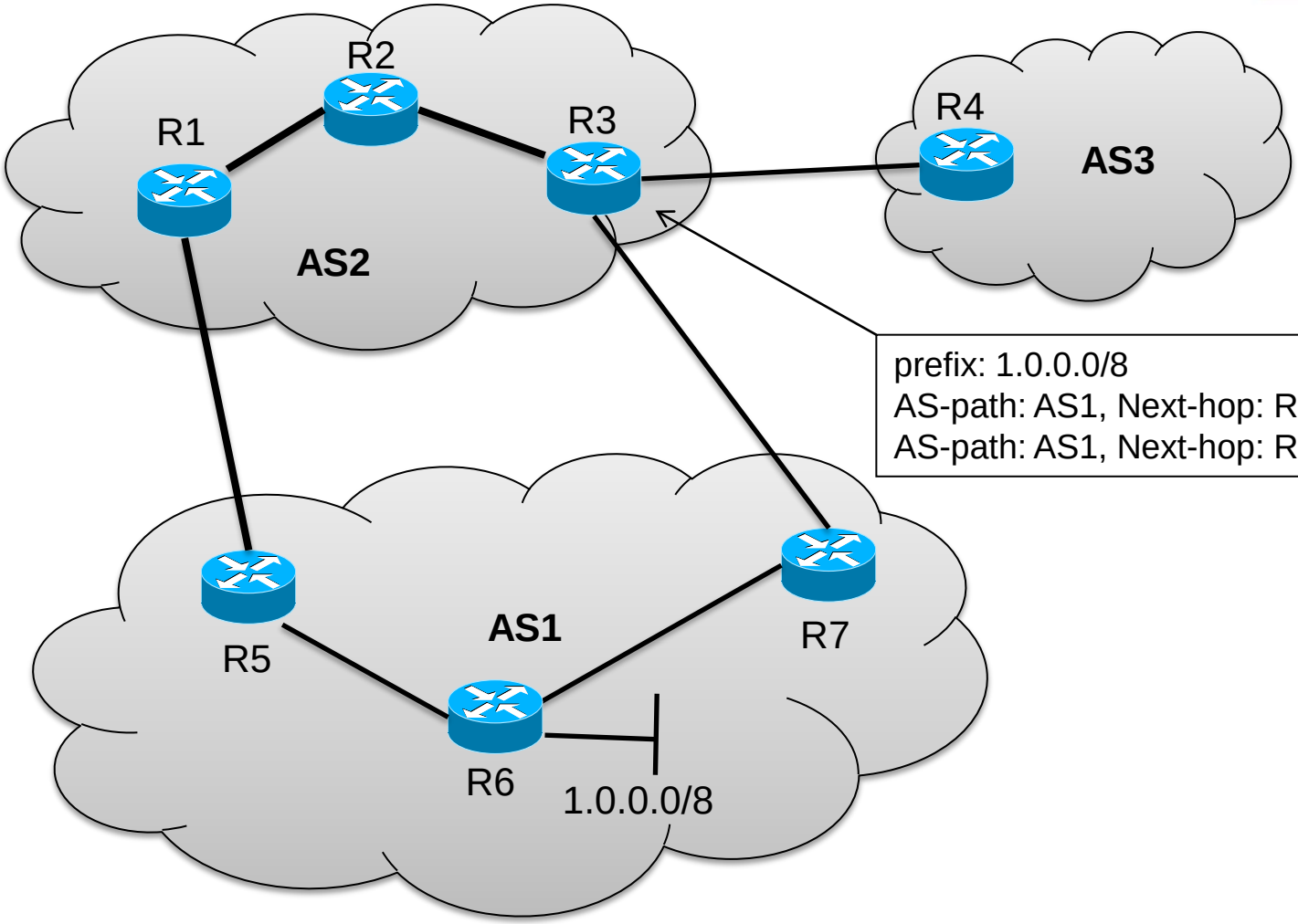


Multi-Exit Discriminator (MED)

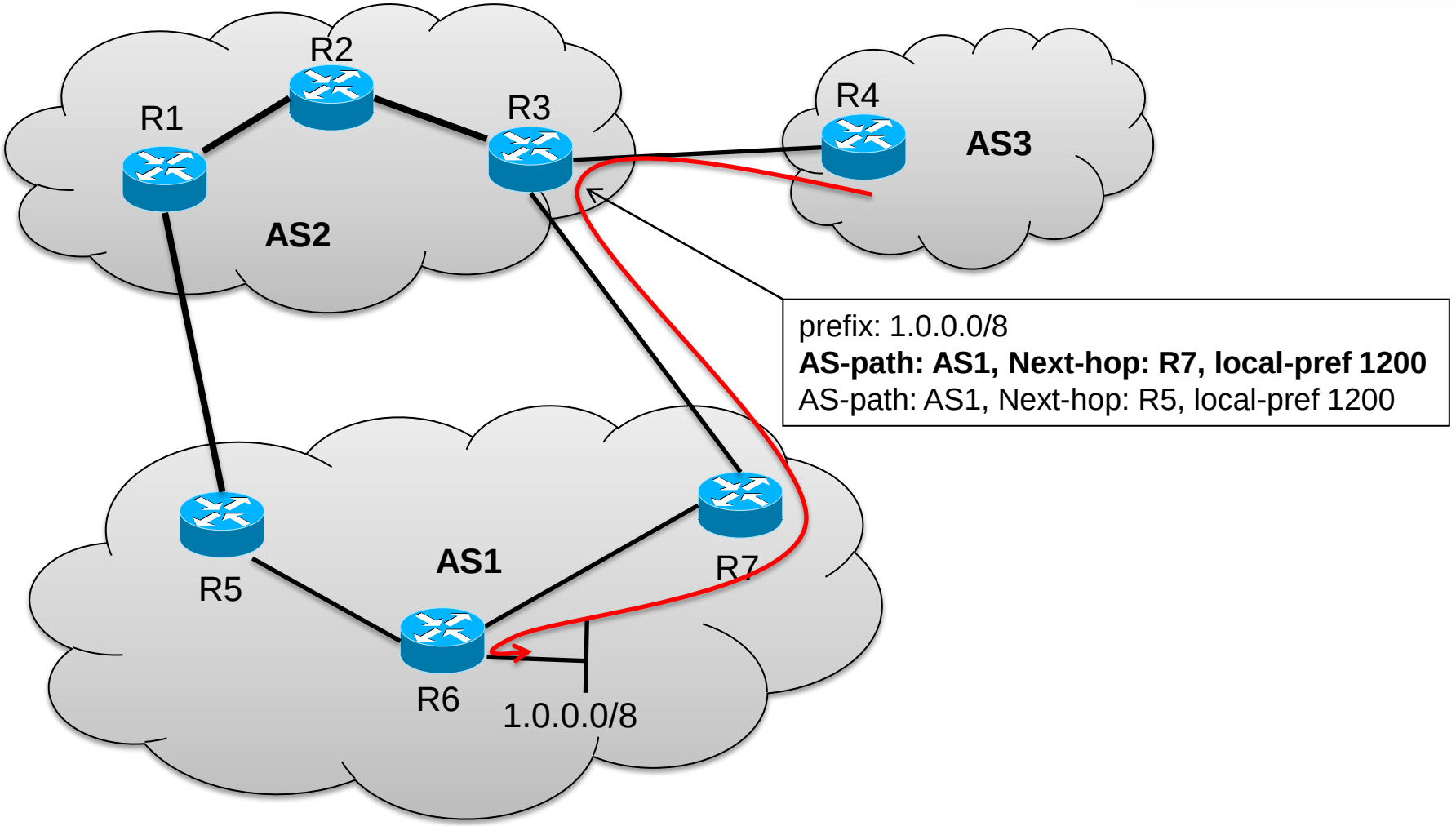


prefix: 1.0.0.0/8
AS-path: AS1, Next-hop: R7, local-pref 1200, MED 50
AS-path: AS1, Next-hop: R5, local-pref 1200, MED 10

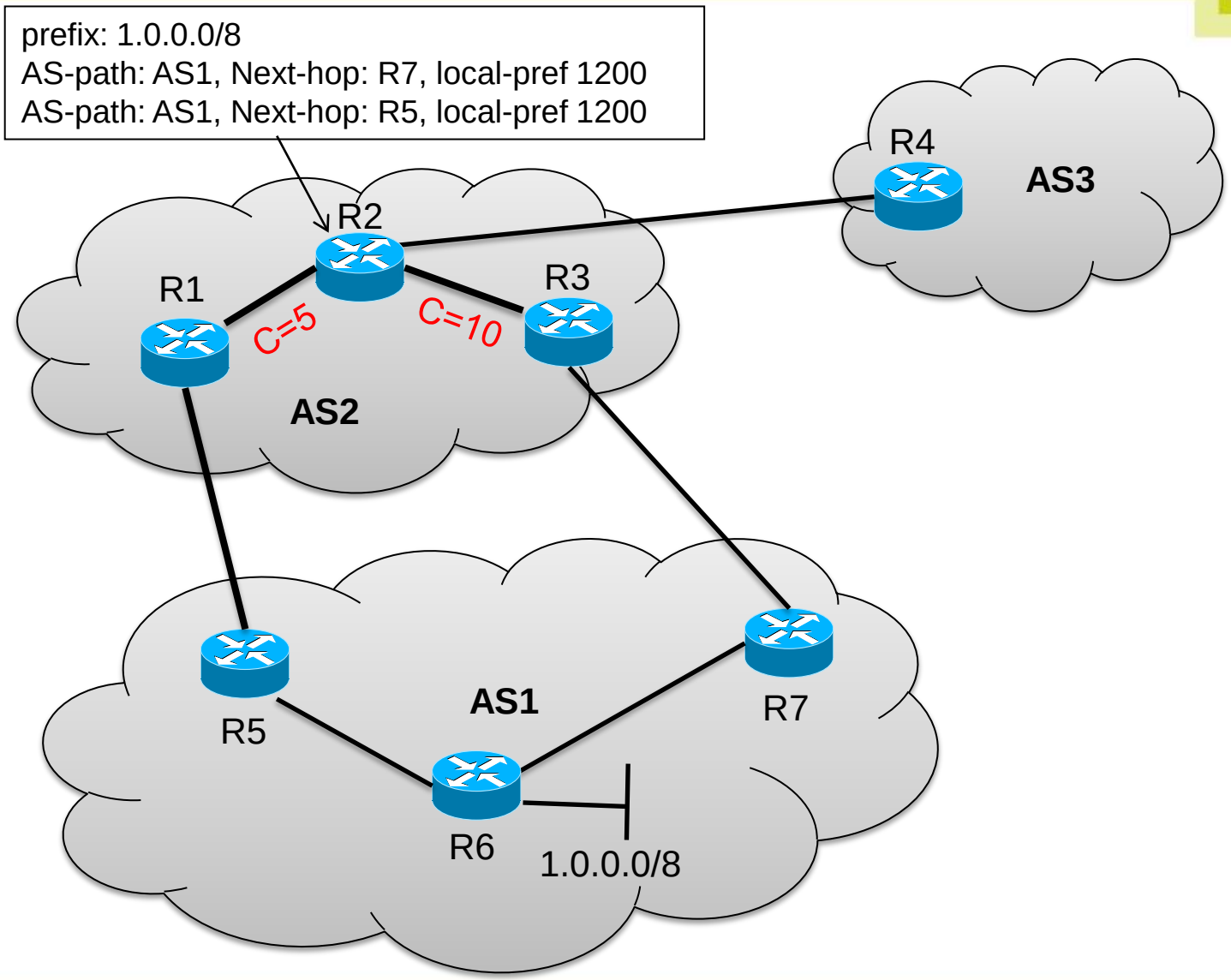
Hot-Potato Routing



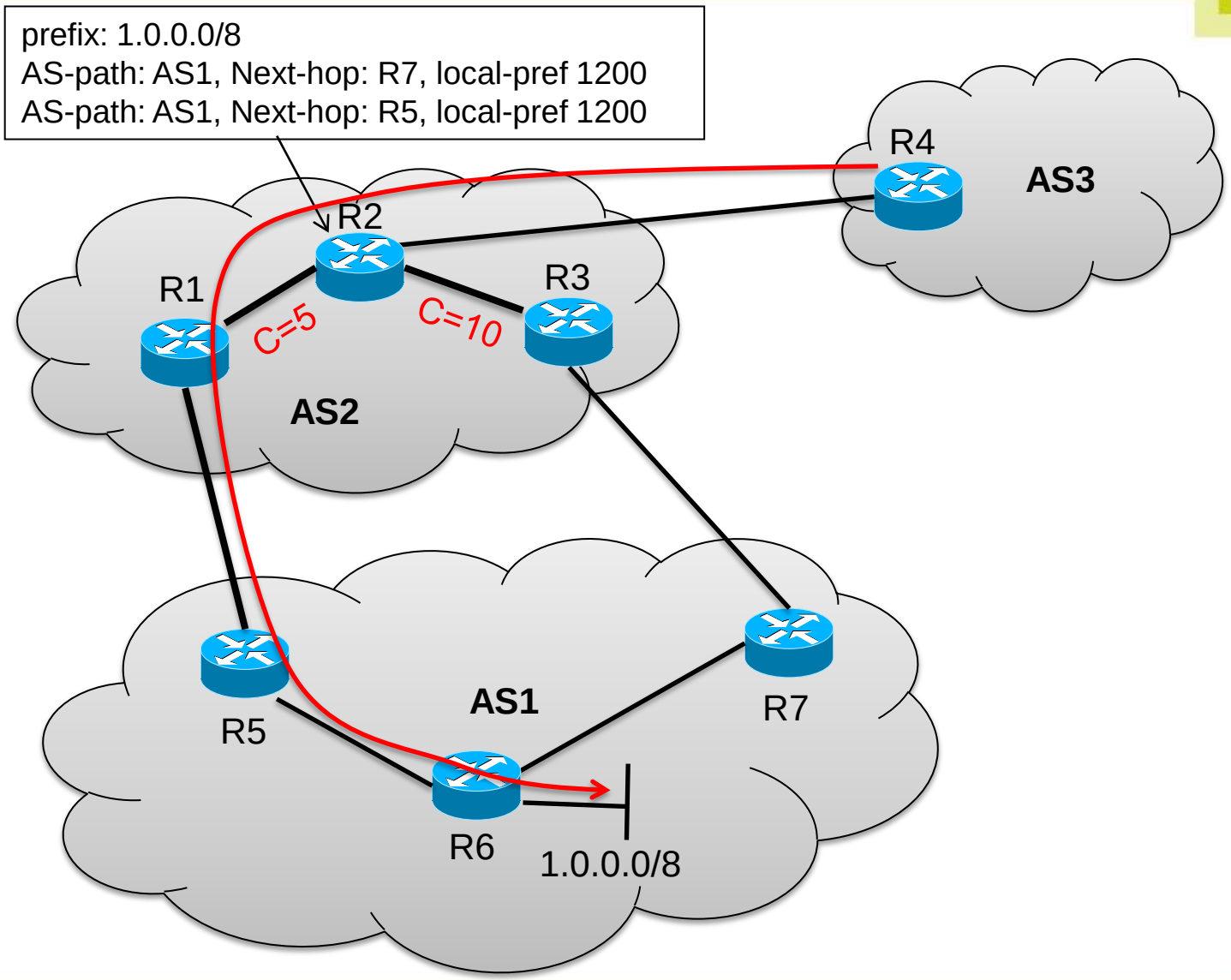
Hot-Potato Routing



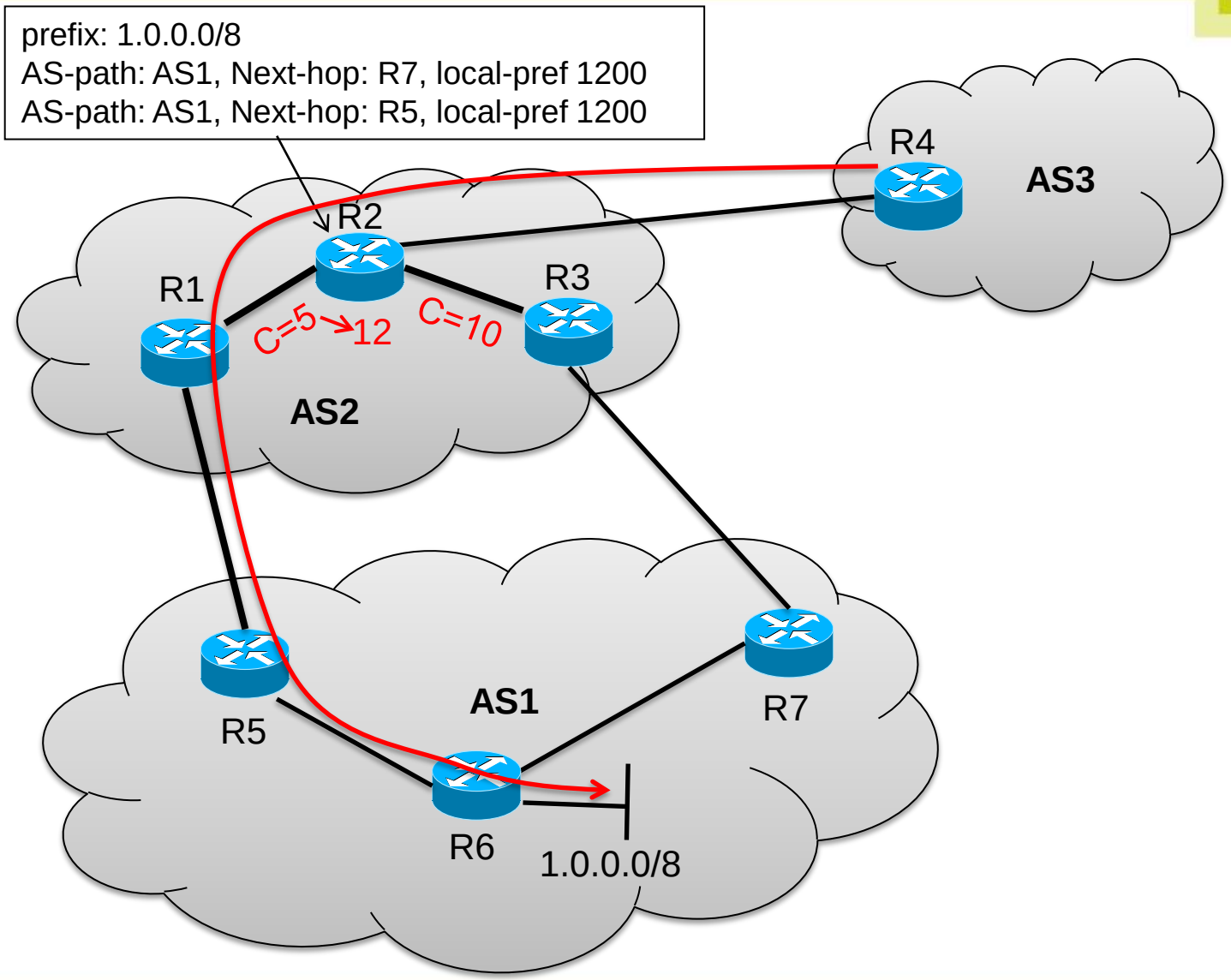
Preferring Shortest Intra-Domain Paths



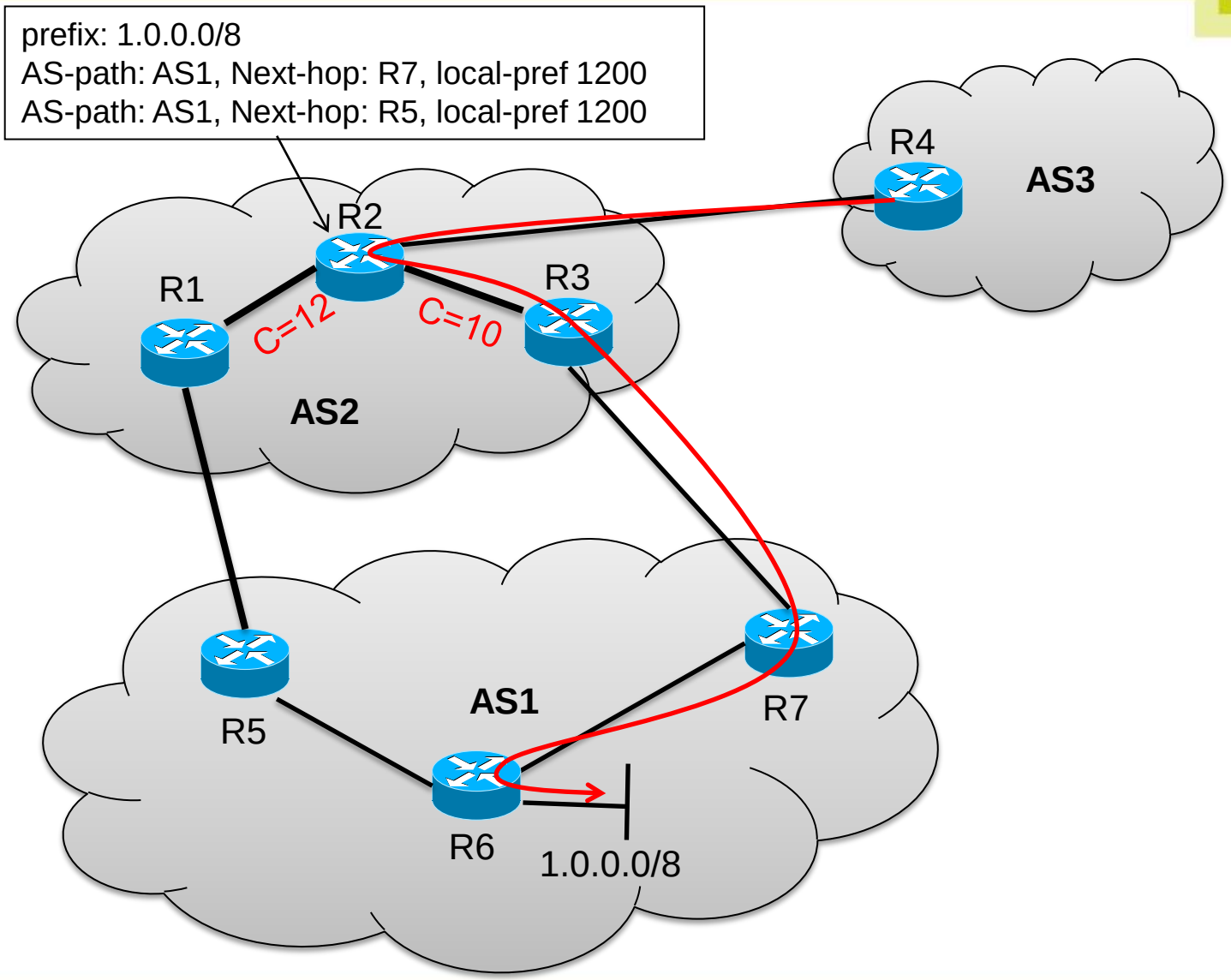
Preferring Shortest Intra-Domain Paths



Interplay Between Intra- and Inter-Domain Routing



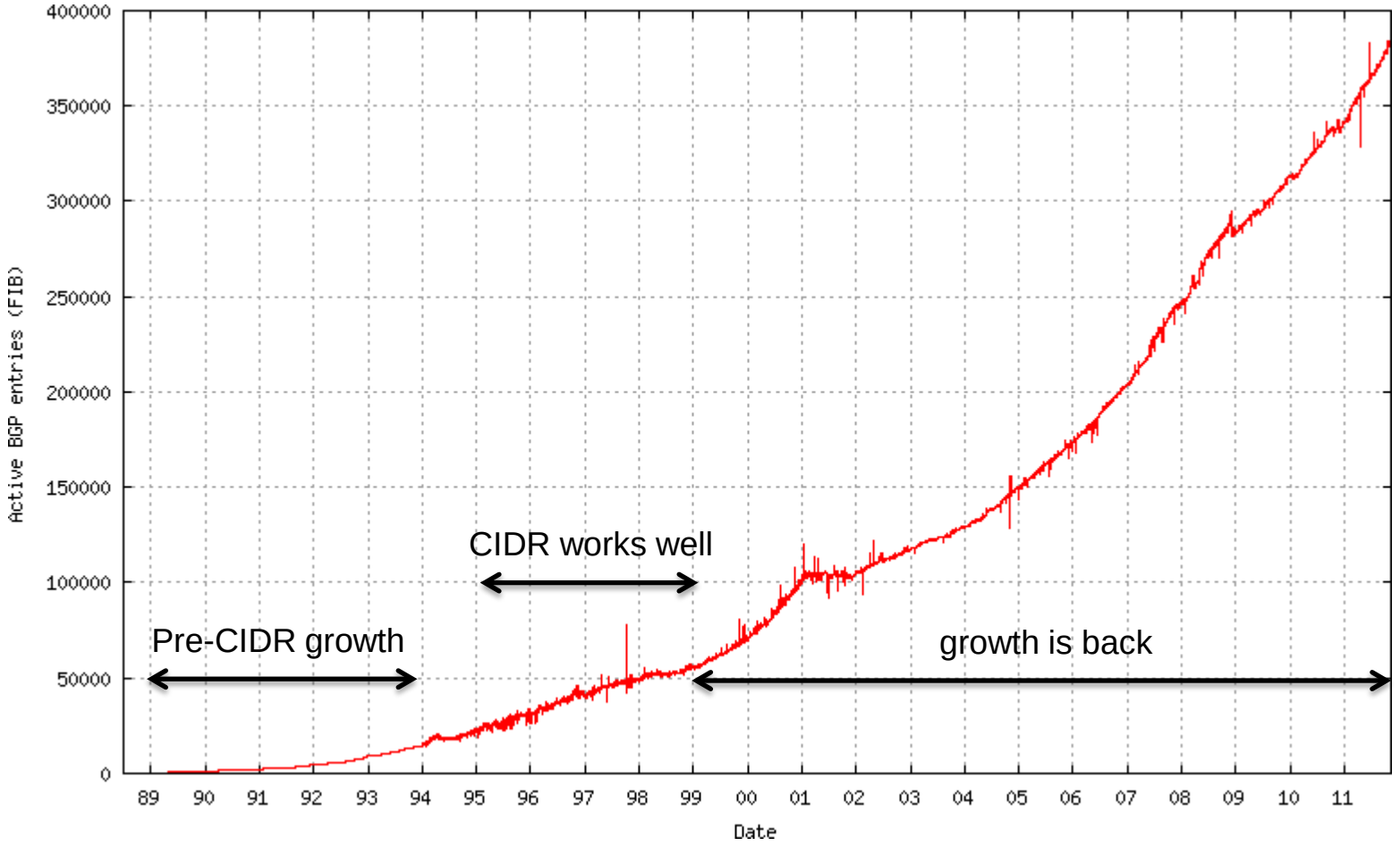
Interplay Between Intra- and Inter-Domain Routing



iBGP vs eBGP

- Differences between iBGP and eBGP:
  - `local-pref` attribute is used only in iBGP sessions
    - Network operators do not want to expose their policies to their peers
  - Over an eBGP session, a router advertises only its best route towards each destination
    - Import and export filters are usually defined and applied for each eBGP session
  - Over an iBGP session, a router advertises only the best routes learnt from eBGP sessions
    - A route learnt from an iBGP session is never advertised over an eBGP session
    - Filters are typically not applied to iBGP sessions

Size of BGP Routing Tables



References

- O. Bonaventure, **Computer Networking: Principles, Protocols and Practice**, http://inl.info.ucl.ac.be/cnp3
- Y. Rekhter, et al., **A Border Gateway Protocol 4 (BGP-4)**, RFC 4271, 2006
- C. Alaettinoglu, et al., **Routing Policy Specification Language (RPSL)**, RFC 2622, 1999
- D. Meyer, et al., **Using RPSL in Practice**, RFC 2650, 1999
- T. Bates, et al., **BGP Route Reflection – An Alternative to Full Mesh iBGP**, RFC 2796, 2000
- R. Teixeira, et al., Dynamics of Hot-Potato Routing in IP Networks, ACM SIGMETRICS 2004
- **BGP Routing Table Analysis Reports**, http://bgp.potaroo.net/