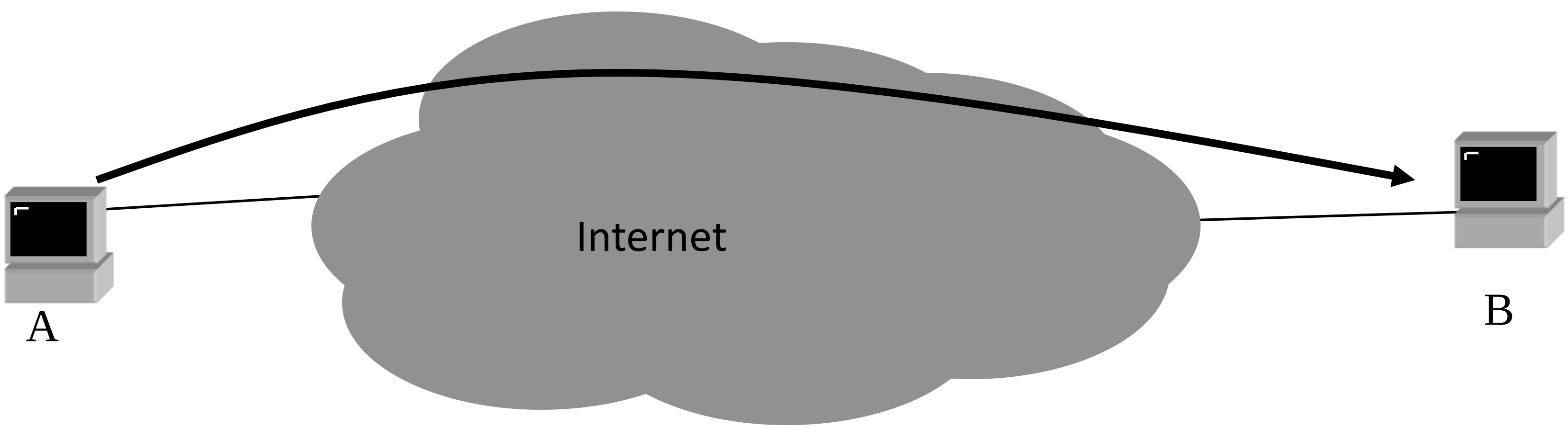
Introduction to IP Routing

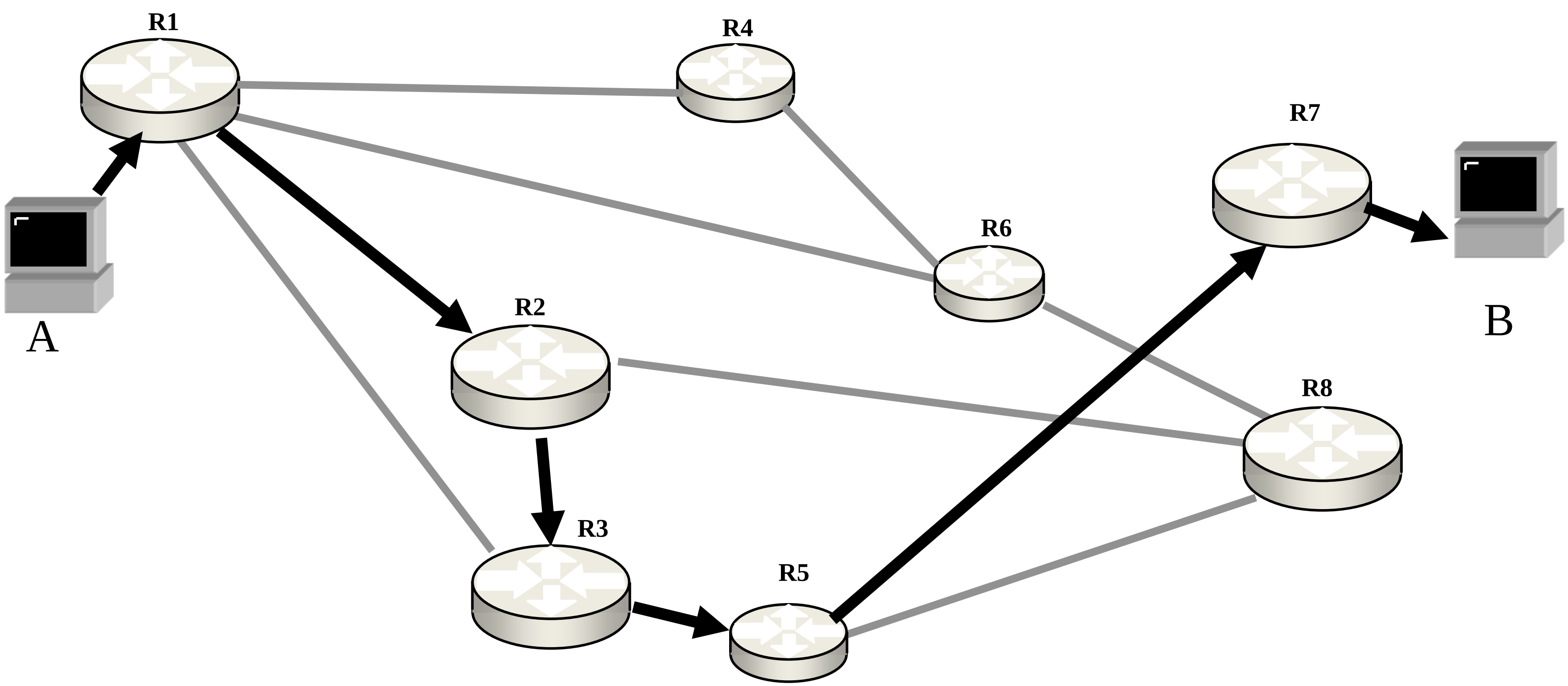
Routing

- How do packets get from A to B in the Internet?



Connectionless Forwarding

- Each router (switch) makes a LOCAL decision to forward the packet towards B



Connectionless Forwarding

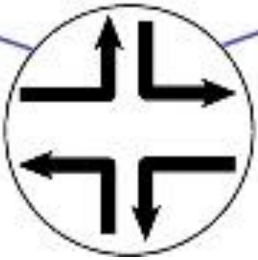
- This is termed *destination-based connectionless forwarding*
- How does each router know the *correct* local forwarding decision for any possible destination address?
  - Through knowledge of the *topology state* of the network
  - This knowledge is maintained by a *routing protocol*

Routing Protocols

- Distribute the knowledge of the current *topology state* of the network to all routers
- This knowledge is used by each router to generate a *forwarding table*, which contains the local switching decision for each known destination address

Forwarding Table

```
nancy@sluggo.lab> show route forwarding-table
Internet:
Destination        Type RtRef Nexthop           Type Index NhRef Netif
10.100.71.0/24     user     0 10.100.67.254     ucst    18 74212 GigE0.0
10.100.71.224/27   user     2 10.100.67.254     ucst    18 74212 GigE0.0
10.250.1.36/30     intf     0 ff.3.0.21         ucst    27     1 so-2/0/0.0
10.250.1.37/32     intf     0 10.250.1.37       locl    26     1
10.250.1.103/32    dest     0 10.250.1.103      bcst    37     1 ge-7/2/0.0
---(more)---
```

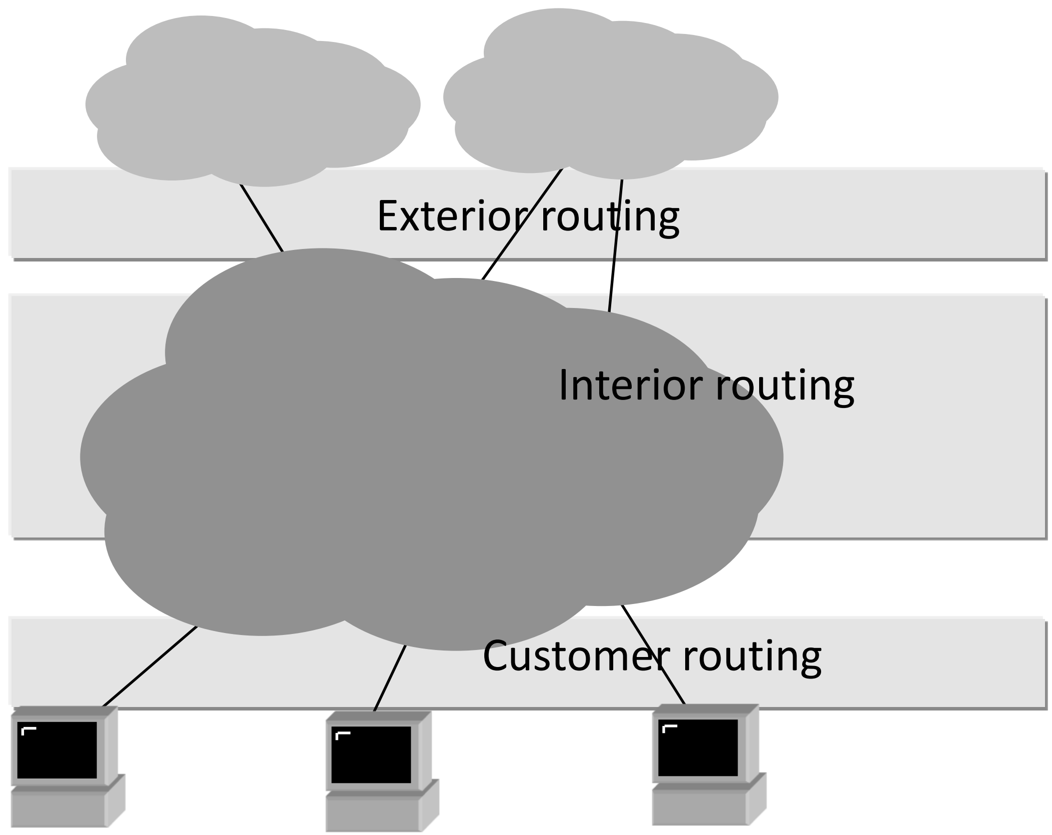


Routing Protocols

- correct operation of the routing state of a network is essential for the management of a *quality* network service
  - accuracy of the routing information
  - dynamic adjustment of the routing information
  - matching aggregate traffic flow to network capacity

ISP Routing Tasks

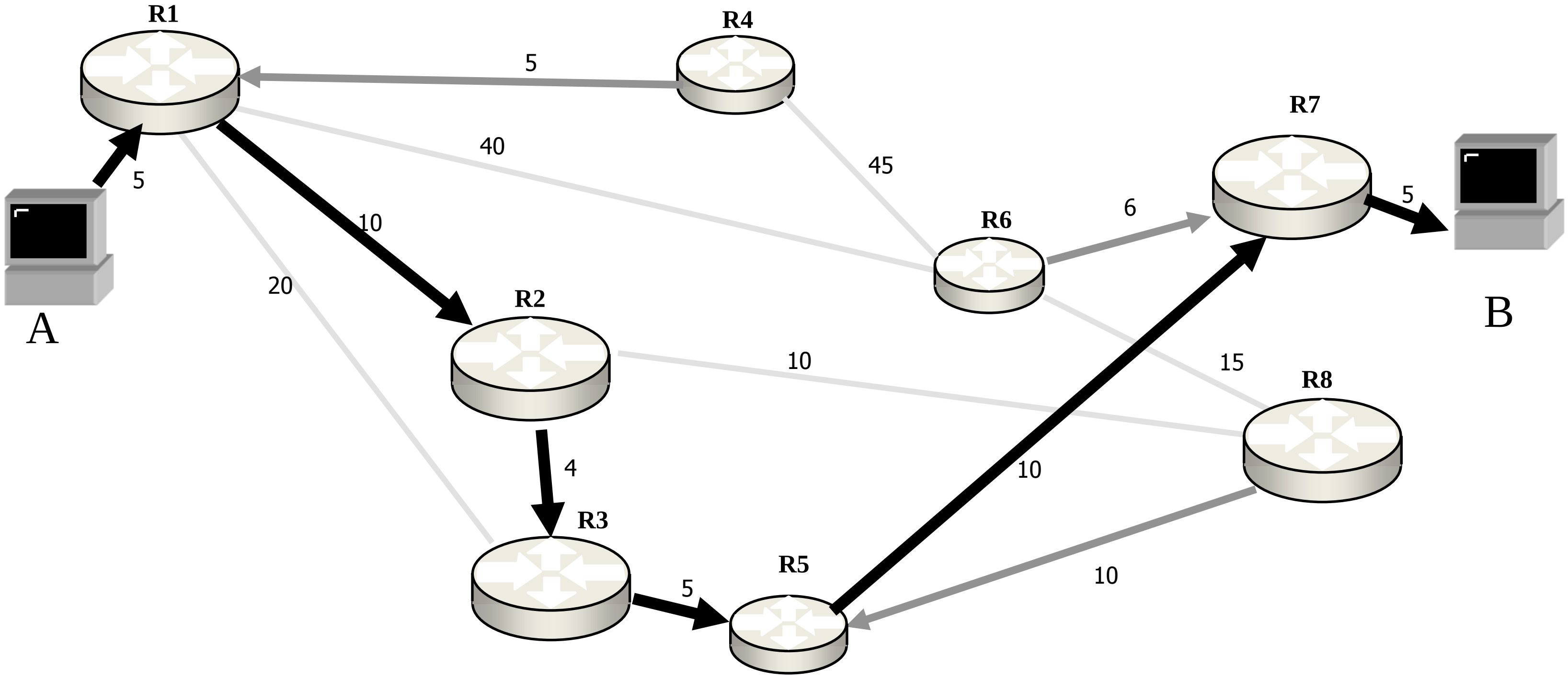
- customers
- internal
- peer / upstream



Interior Routing

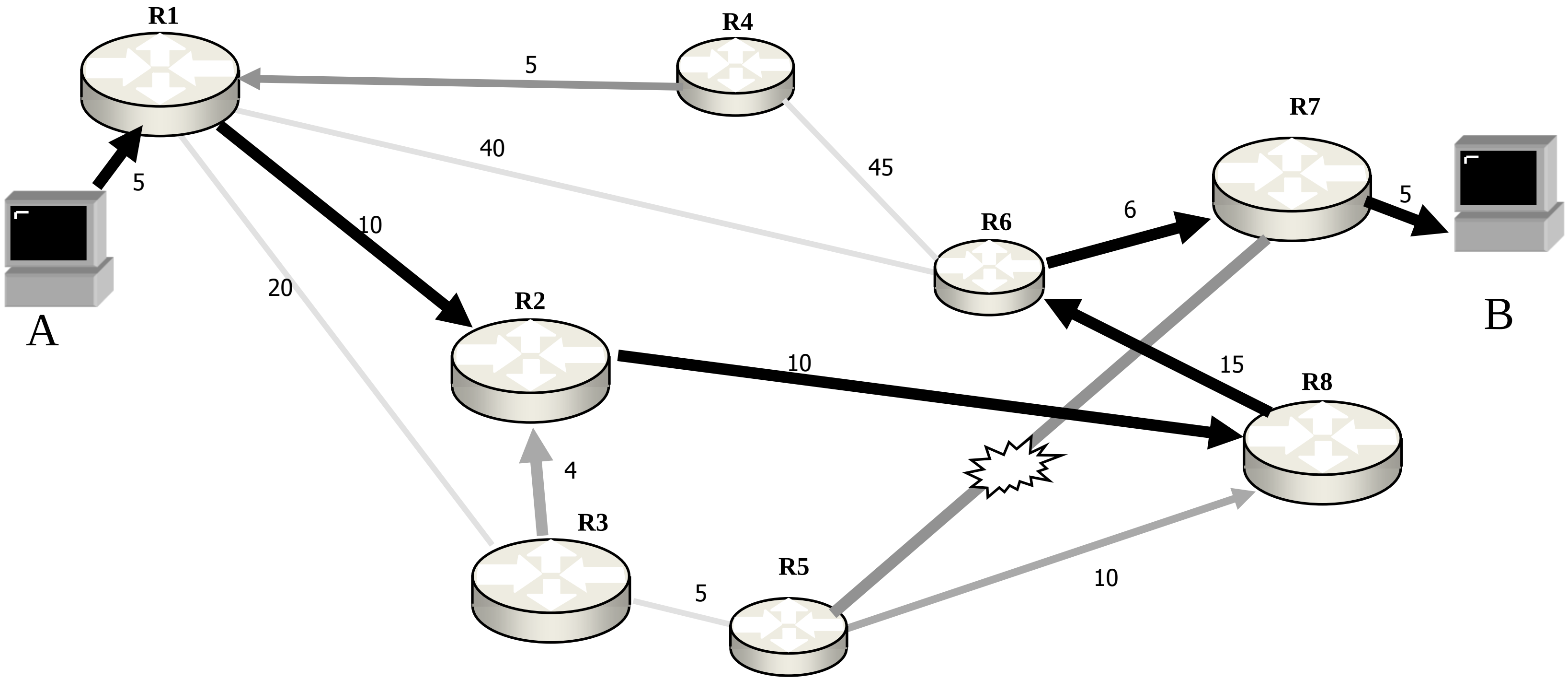
- discovers the *topology* of a network through the operation of a *distributed routing protocol*

Path Selection



Minimum cost from A to B is **39** units

Dynamic Path Adjustment



If R5 – R7 breaks, minimum cost path from A to B is Now **46** units

Interior Routing Protocols

- describe the current network topology
- Routing protocols distribute **how** to reach **address prefix** groups
- Routing protocols function through either
  - *distributed computing model (distance vector)*
  - *parallel computing model (link state)*

Routing Protocols

- Distance Vector Routing Protocols
  - Each node sends its routing table (dest, distance) to all neighbors every 30 seconds
  - Lower distances are updated with the neighbor as next hop
  - cannot scale
  - cannot resolve routing loops quickly
  - RIP is the main offender

Routing Protocols

- Link State Routing Protocols
  - Each link, the connected nodes and the metric is flooded to all routers
  - Each link up/down status change is incrementally flooded
  - Each router re-computes the routing table in parallel using the common link state database.
  - OSPF is the main protocol in use today

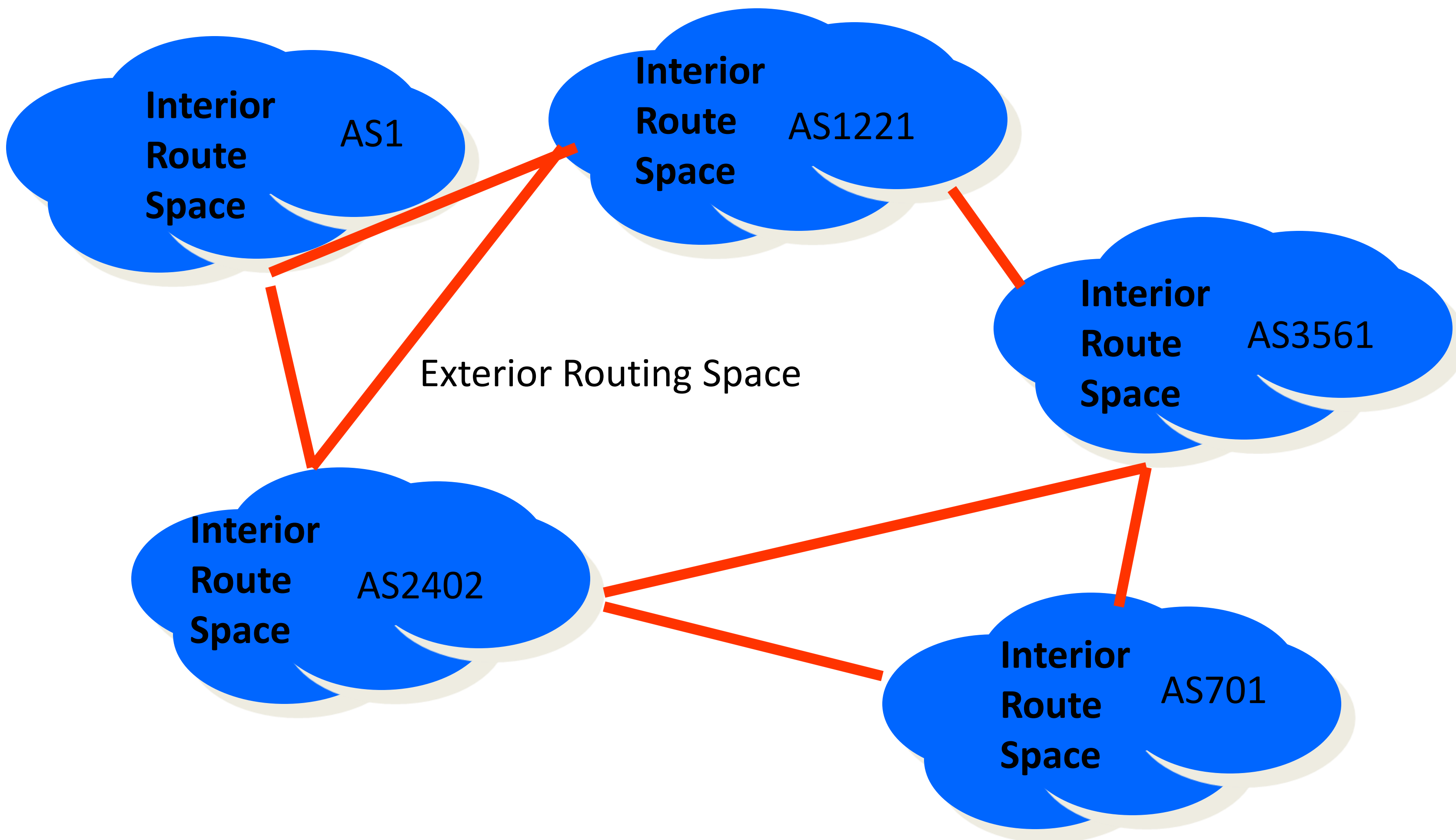
Suggestions

- Just engineering a physical link does not ensure that traffic will flow
  - some system somewhere must provide routing information about how to reach the newly connected network
- Installing backup circuits is easy, making the routing work may not be

Suggestions

- need a clear understanding of how the client networks want their traffic to flow before you can start making routing configuration changes

Interior and Exterior Routing Protocols



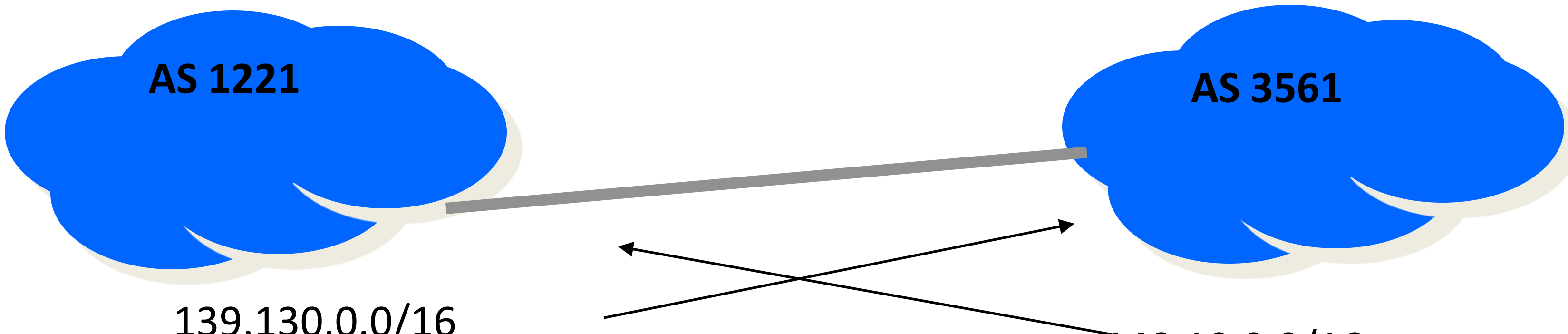
Exterior Routing Protocols

- You tell me all the address prefixes you can reach, but don’t tell me the path you use to get there
  - I’ll tell you the same
- If anything changes, please let me know
- If you tell me an address I’ll send you traffic destined to that address.
  - If I tell you an address I will accept traffic destined to that address

Exterior Routing Protocols

- Border Gateway Protocol version 4 (BGP4)
- Each interior route collection is described by an Autonomous System (AS) number
- Internal topology is hidden
- Routes are announced with associated AS value
  - 139.130.0.0/16 + AS 1221

BGP example



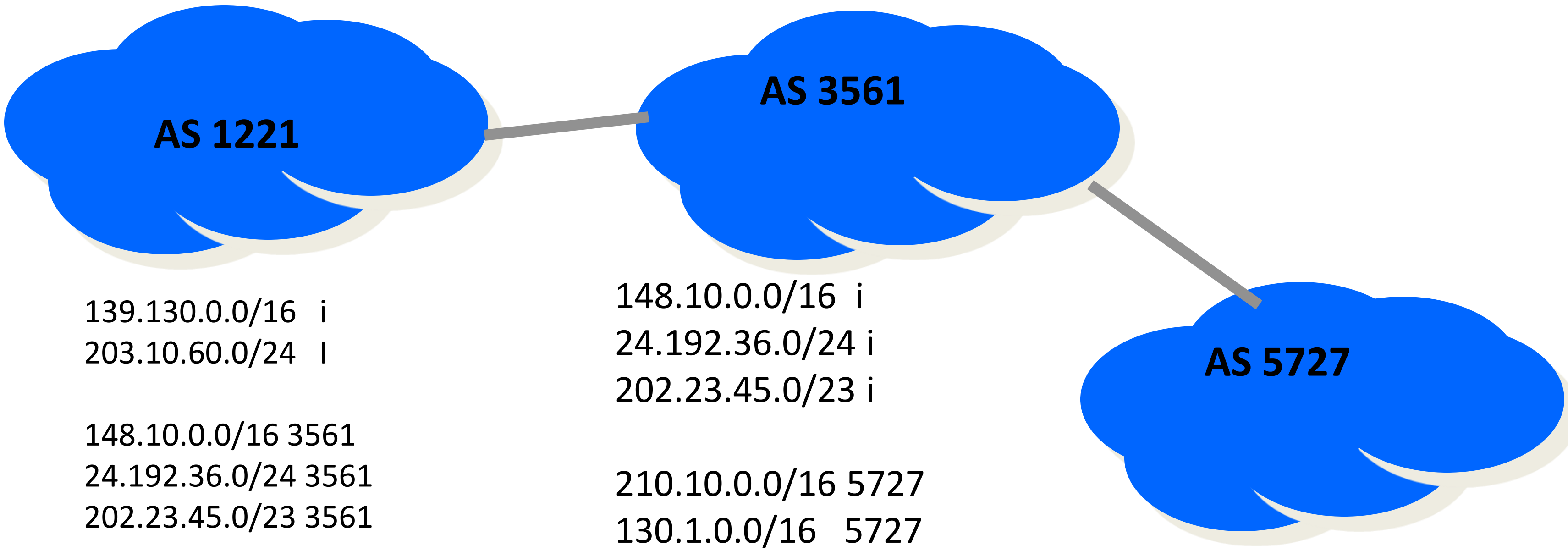
139.130.0.0/16
203.10.60.0/24

148.10.0.0/16 3561
24.192.36.0/24 3561
202.23.45.0/23 3561

148.10.0.0/16
24.192.36.0/24
202.23.45.0/23

139.130.0.0/16 1221
203.10.60.0/24 1221

BGP Example of TRANSIT



AS 1221

139.130.0.0/16 i
203.10.60.0/24 I

148.10.0.0/16 3561
24.192.36.0/24 3561
202.23.45.0/23 3561

210.10.0.0/16 3561,5727
139.1.0.0/16 3561,5727

AS 3561

148.10.0.0/16 i
24.192.36.0/24 i
202.23.45.0/23 i

210.10.0.0/16 5727
130.1.0.0/16 5727

139.130.0.0/16 1221
203.10.60.0/24 1221

AS 5727

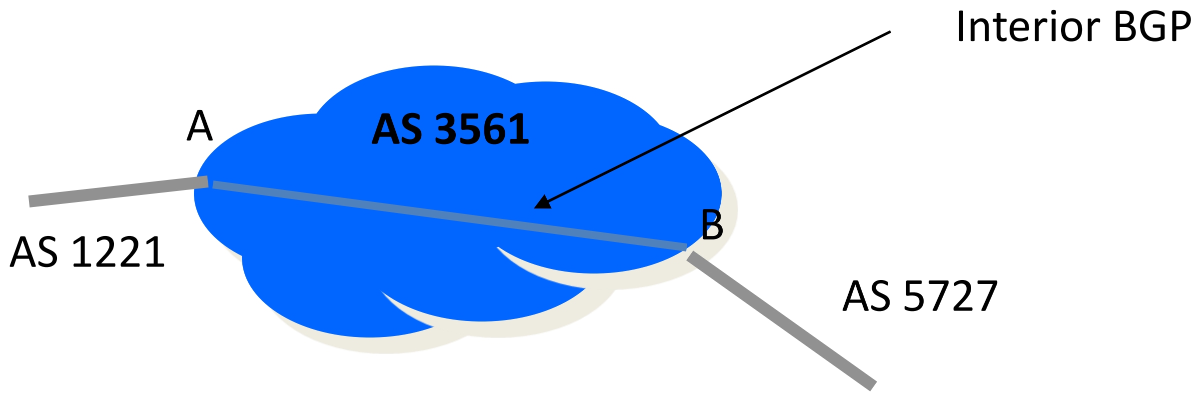
210.10.0.0/16 i
130.1.0.0/16 I

148.10.0.0/16 3561
24.192.36.0/24 3561
202.23.45.0/23 3561

139.130.0.0/16 3561,1221
203.10.60.0/24 3561,1221

Exterior Routing Protocols

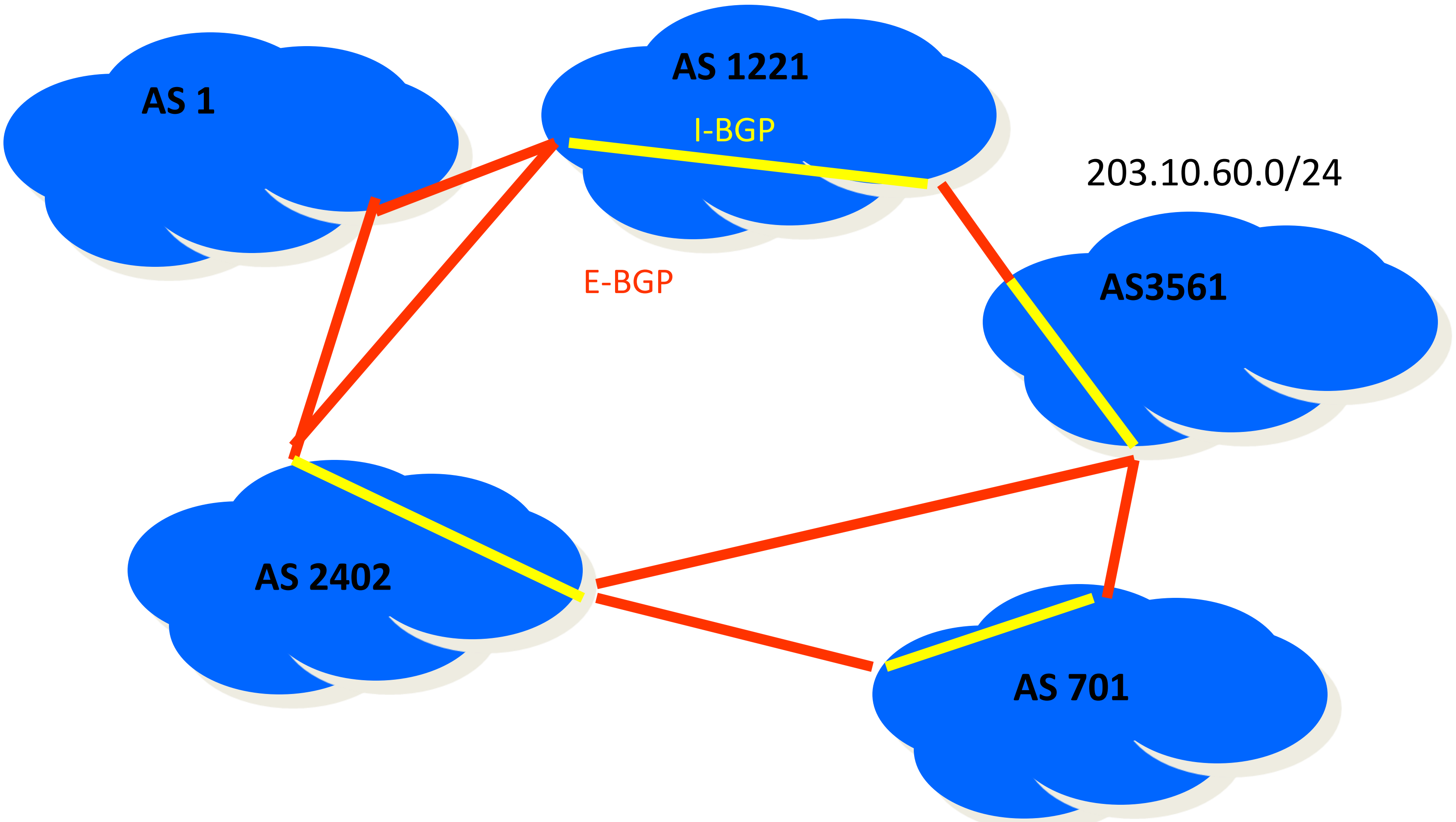
- Internal transit paths use I-BGP



Q: How does router A tell router B about AS1221 addresses?

A: Router A sets an INTERIOR BGP session with router B

Exterior Routing Protocols



Exterior Routing Protocols

- Normally chose minimal AS path length

203.10.60.0/24 701,3561,1221
→ 203.10.60.0/24 5727,1221

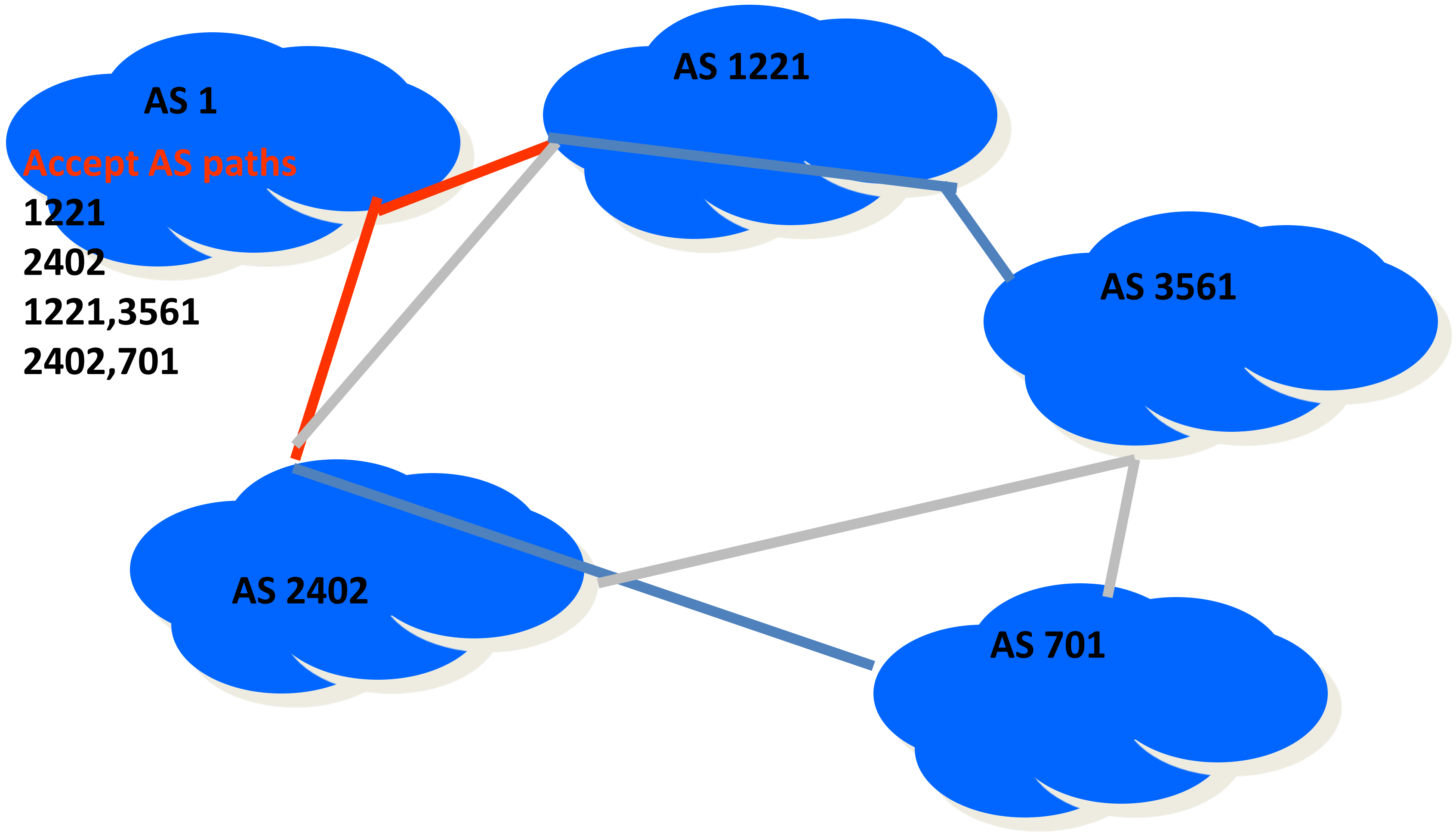
Selected path is via peer session to AS 5727 as this
Is 1 AS shorter that the other path

Exterior POLICY

- How can I share the traffic load between 2 or more exterior providers?
- How can I create a backup link to support my main exterior link?

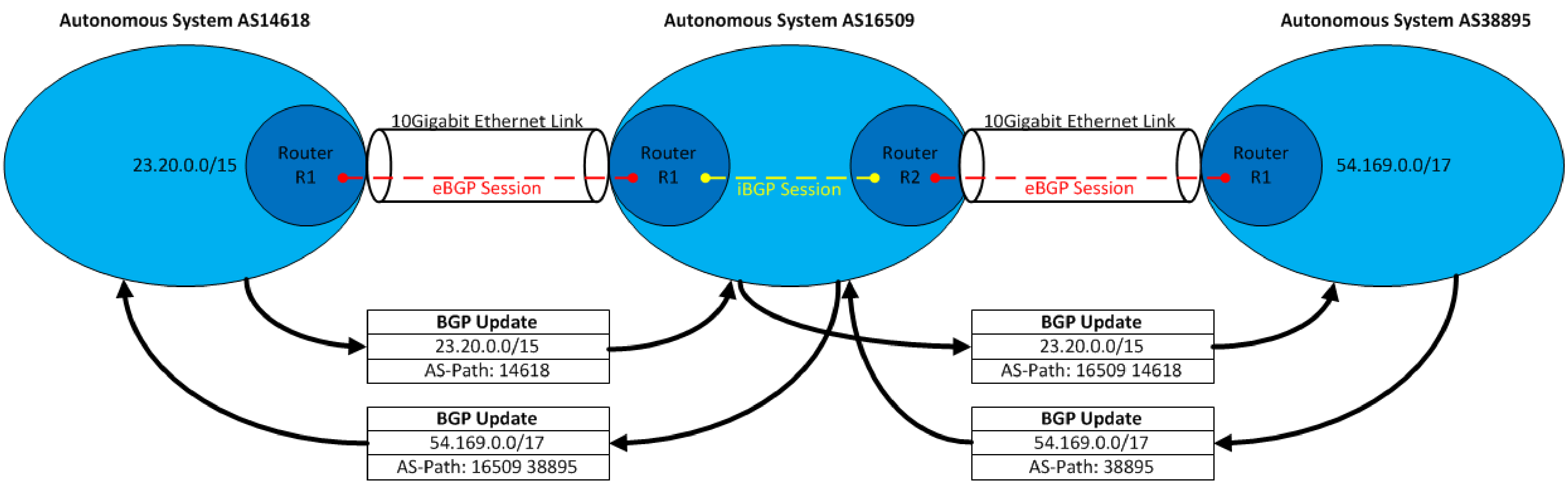
- You can bias minimal path selection by **AS path filter lists** or **community attributes** or **local preferences**

Exterior Routing Protocols plus Policy



Exterior Routing Protocols plus Policy

- policy settings control
  - what you advertise to your immediate peers
  - What you accept from your immediate peers
  - What transits you will accept (send traffic)
- you **cannot** control
  - transit path of received traffic
  - symmetry of transit policy



AS14618 Routing Table
Prefix: 23.20.0.0/15 AS-Path: Locally Originated
Prefix: 54.169.0.0/17 AS-Path: 16509 38895 Next-Hop: R1

AS16509 Routing Table
Prefix: 23.20.0.0/15 AS-Path: 14618 Next-Hop: R1
Prefix: 54.169.0.0/17 AS-Path: 38895 Next-Hop: R2

AS38895 Routing Table
Prefix: 23.20.0.0/15 AS-Path: 16509 14618 Next-Hop: R1
Prefix: 54.169.0.0/17 AS-Path: Locally Originated

Interior Gateway Protocols

Routing Information Protocol (RIP)

- IP only
- distance vector protocol
- slow convergence
- does not carry mask information
- reasonably simple design &configuration
- does not scale (maximum 15 hops)
- poor metrics (hop-count)

Interior Gateway Protocols

The IGRP metric

» always get optimal routing metric vector, not single value

I. bandwidth
II. delay
III. hops
IV. reliability
V. loading

Interior Gateway Protocols

Open Shortest Path First (OSPF)

I. IP only
II. link state protocol
III. fast convergence
IV. design and architecture very complex
V. configuration can be simple

Interior Gateway Protocols

Which to use?

a. Your interior network is actually VERY simple.

b. Your IGP should only carry your routes and your direct customers'

Interior Gateway Protocols

Problems with "classic" protocols

- ❑slow convergence
- ❑count to infinity
- ❑no mask information

Interior Gateway Protocols

Slow convergence

advertisement period

- entire routing table dumped every n seconds

timeout period

- usually 3 times advertisement period

RIP values are normally 30 and 90 seconds!

Count to infinity: hold-down

Link Between A & B is Broken

	A	B	C	D
A	0, -	1, A	2, B	3, C
B	1, B	0, -	2, C	3, D
C	2, B	1, C	0, -	1, C
D	3, B	2, C	1, D	0, -

Count to infinity problem

 Imagine a network with a graph as above.
 As you see in this graph, there is only one link between A and the other parts of the network.
 Now imagine that the link between A and B is cut.
 At this time, B corrects its table.
 After a specific amount of time, routers exchange their tables, and so B receives C's routing table.
 Since C doesn't know what has happened to the link between A and B, it says that it has a link to A with the weight of 2 (1 for C to B, and 1 for B to A -- it doesn't know B has no link to A).
 B receives this table and thinks there is a separate link between C and A, so it corrects its table and changes infinity to 3 (1 for B to C, and 2 for C to A, as C said).
 Once again, routers exchange their tables.
 When C receives B's routing table, it sees that B has changed the weight of its link to A from 1 to 3, so C updates its table and changes the weight of the link to A to 4 (1 for C to B, and 3 for B to A, as B said).
 This process loops until all nodes find out that the weight of link to A is infinity.