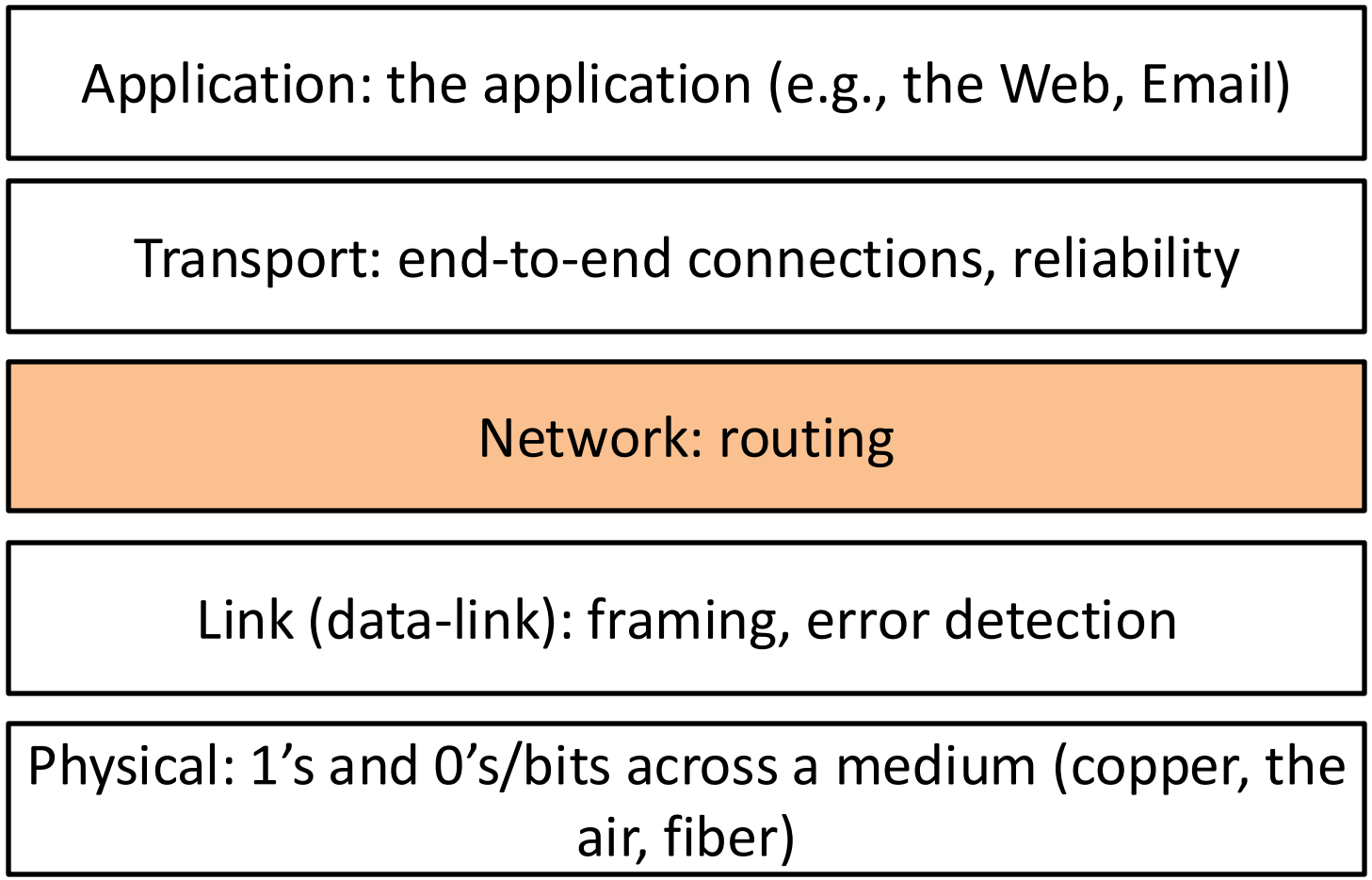
CS 43: Computer Networks

BGP

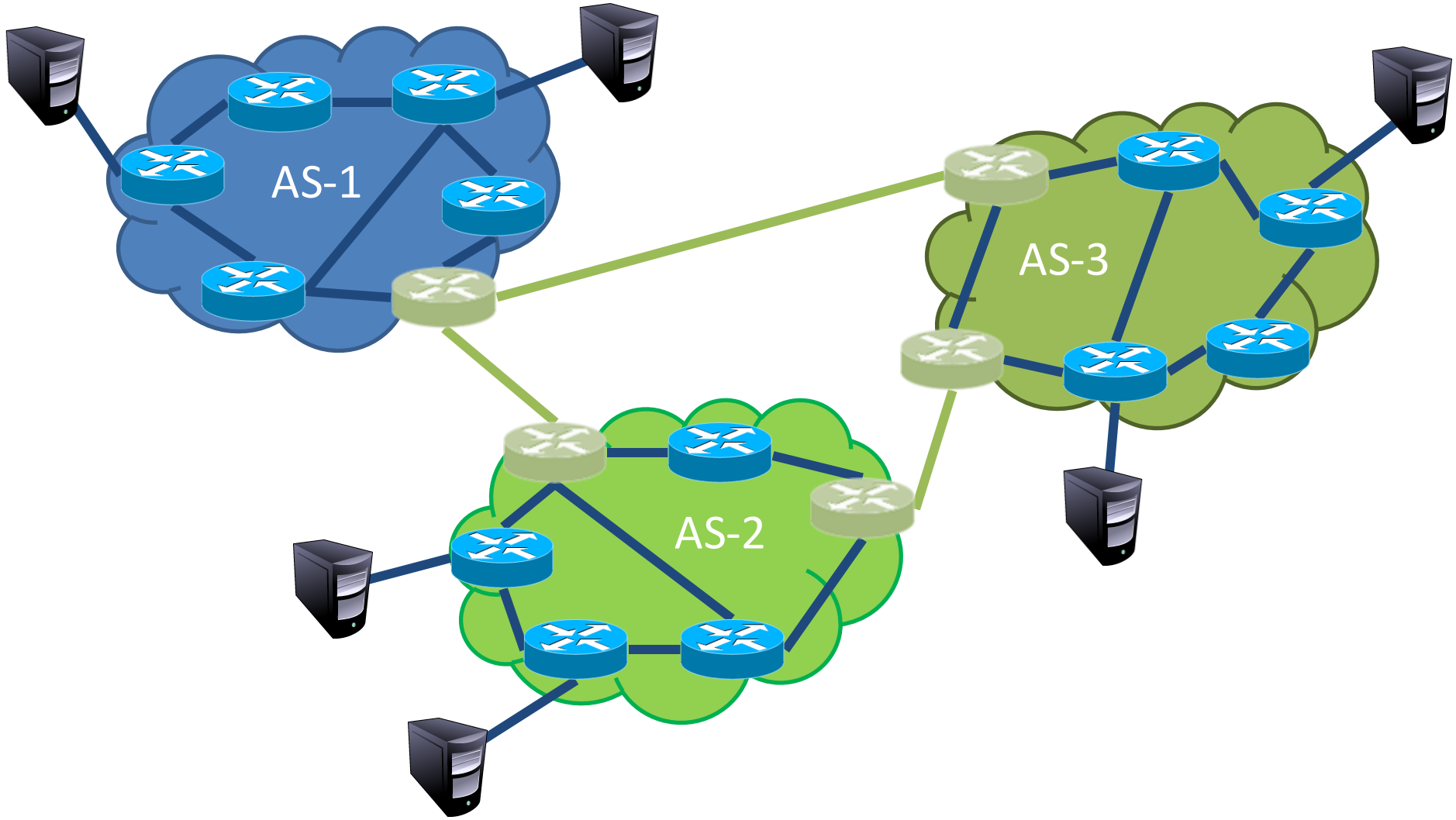
November 26, 2025

Adapted from Slides by: Kurose & Ross, D. Choffnes, K. Webb, J. Rexford

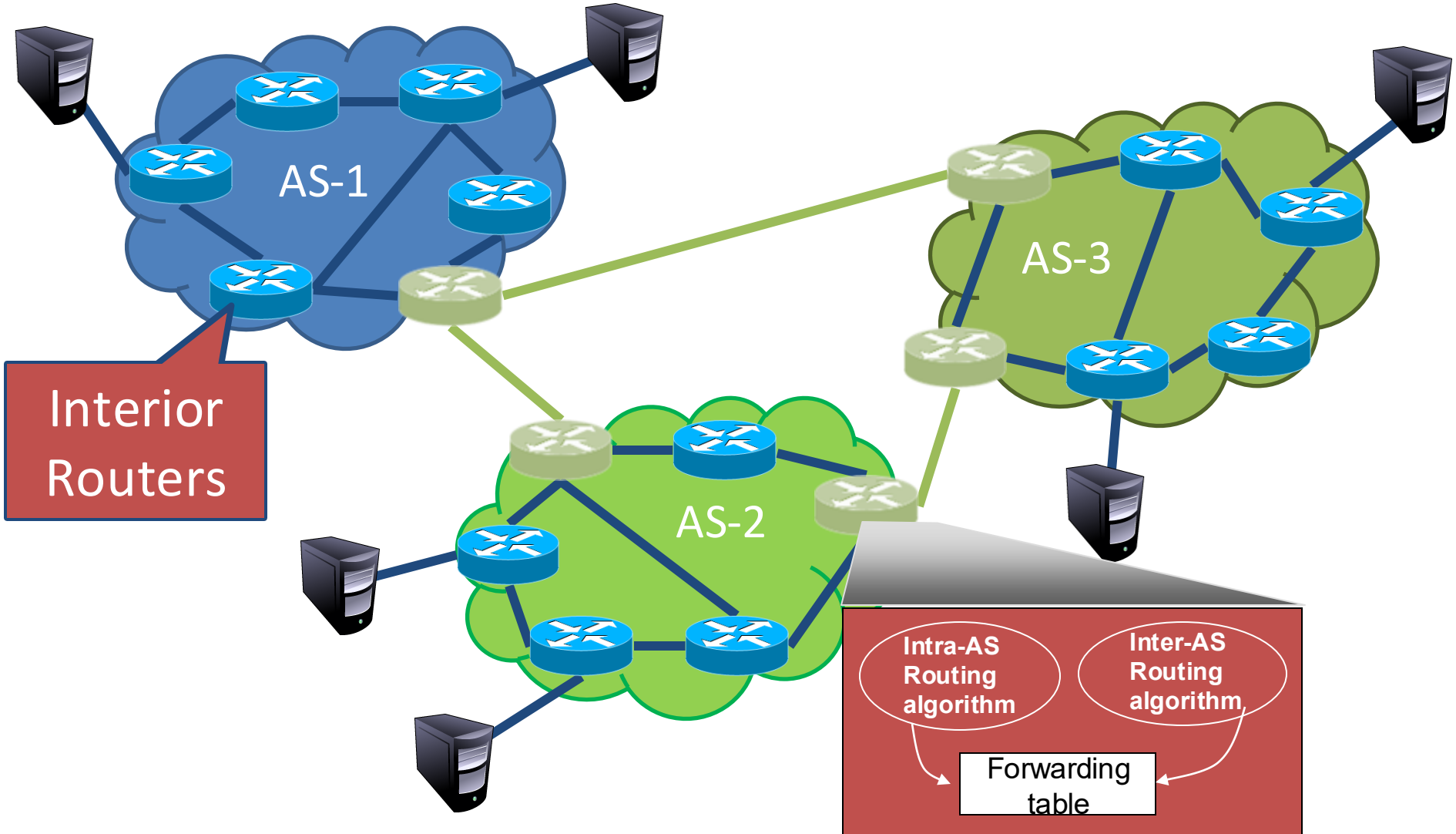
The Network Layer!



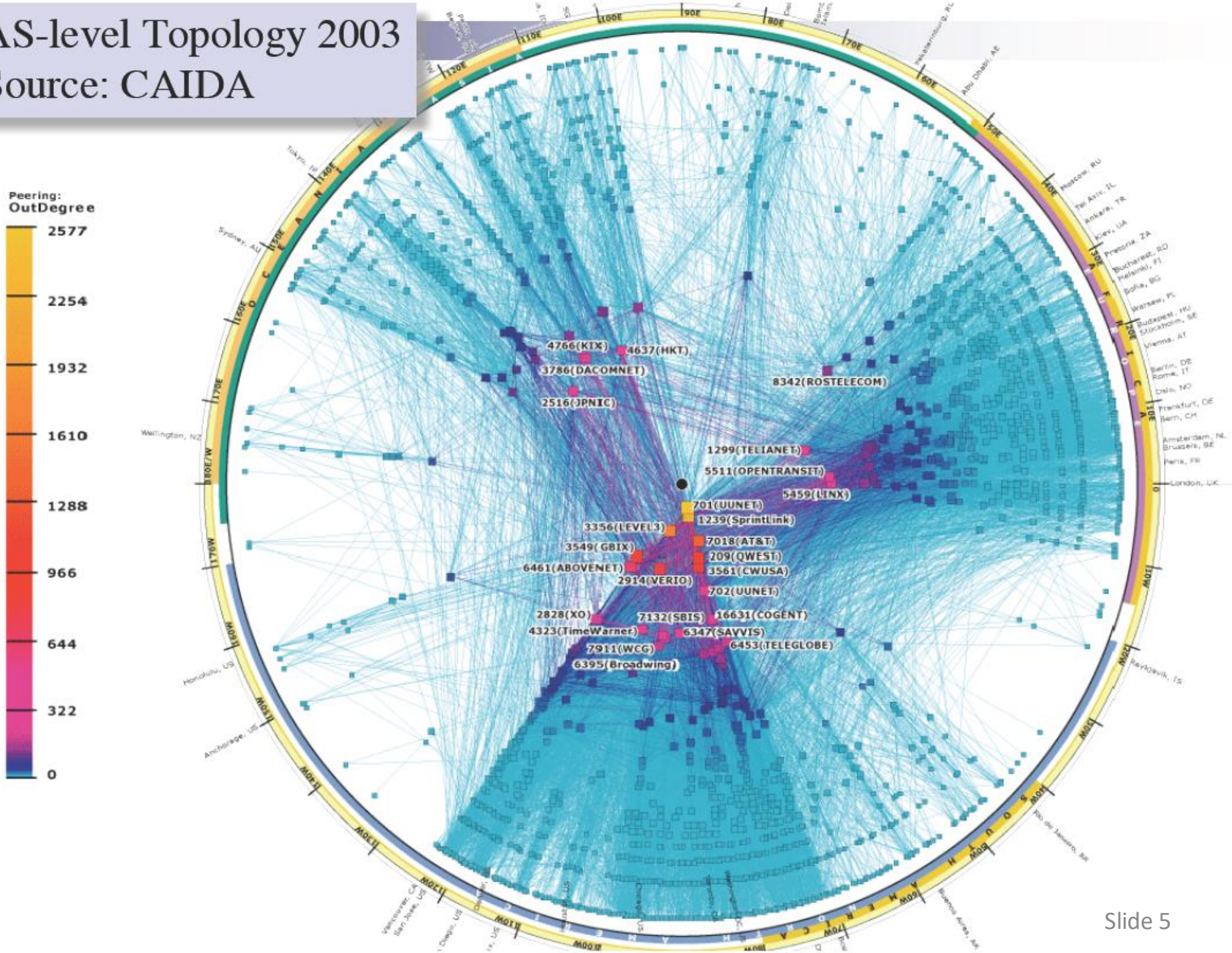
Hierarchical routing: Autonomous Systems



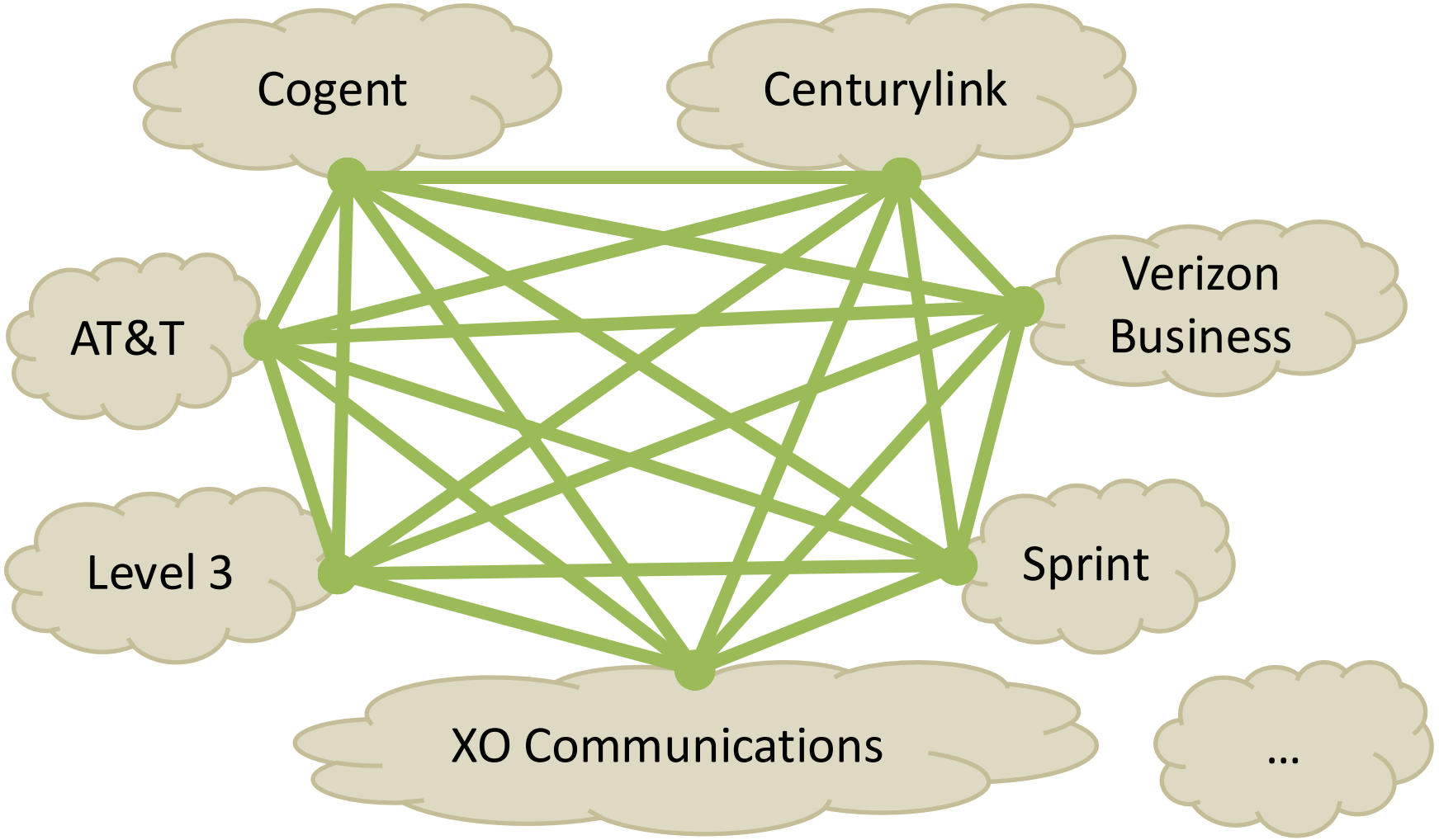
Hierarchical routing: Interconnected ASes



AS-level Topology 2003
Source: CAIDA

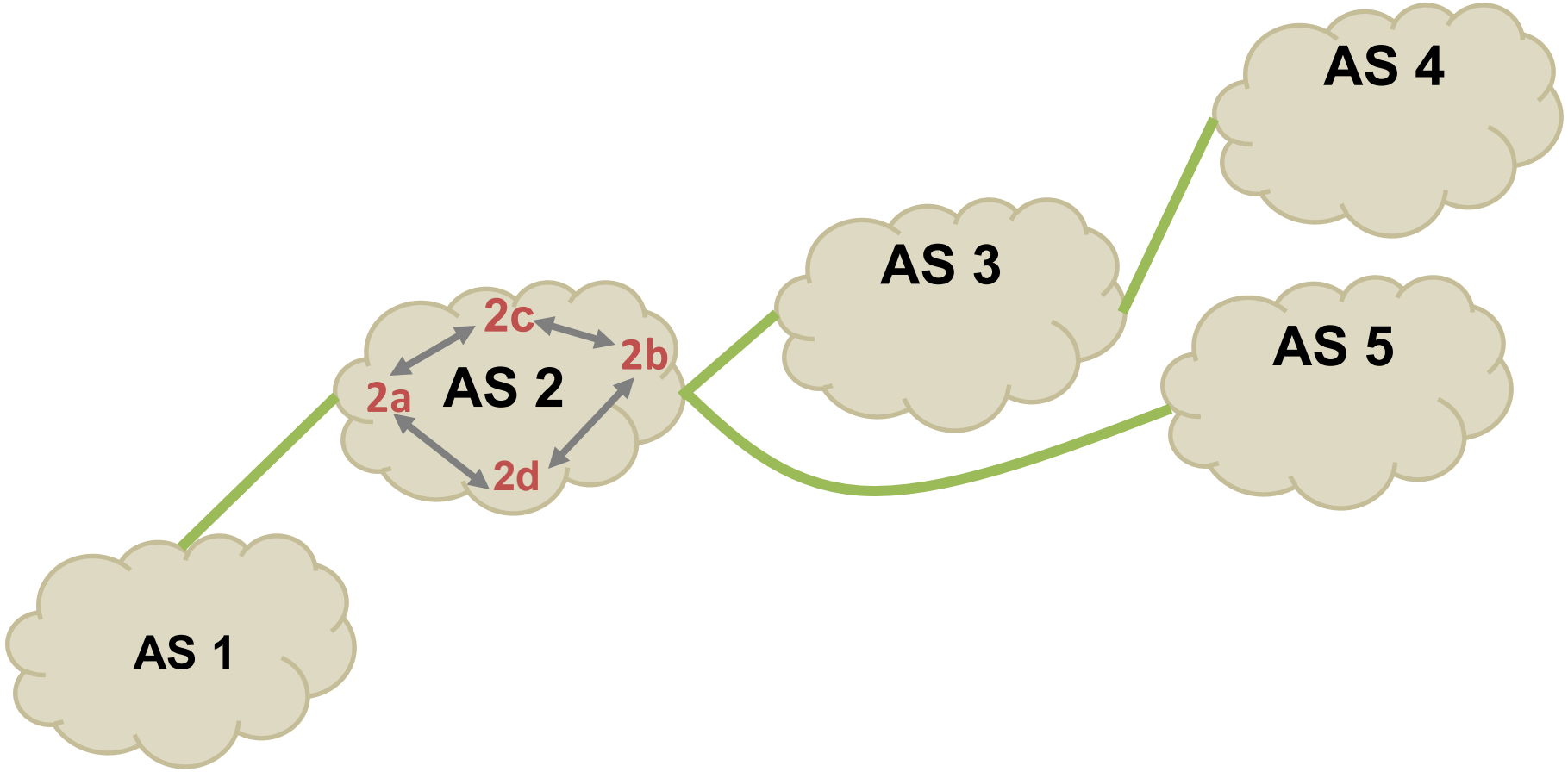


Tier-1 ISP Peering



Path Vector Protocol

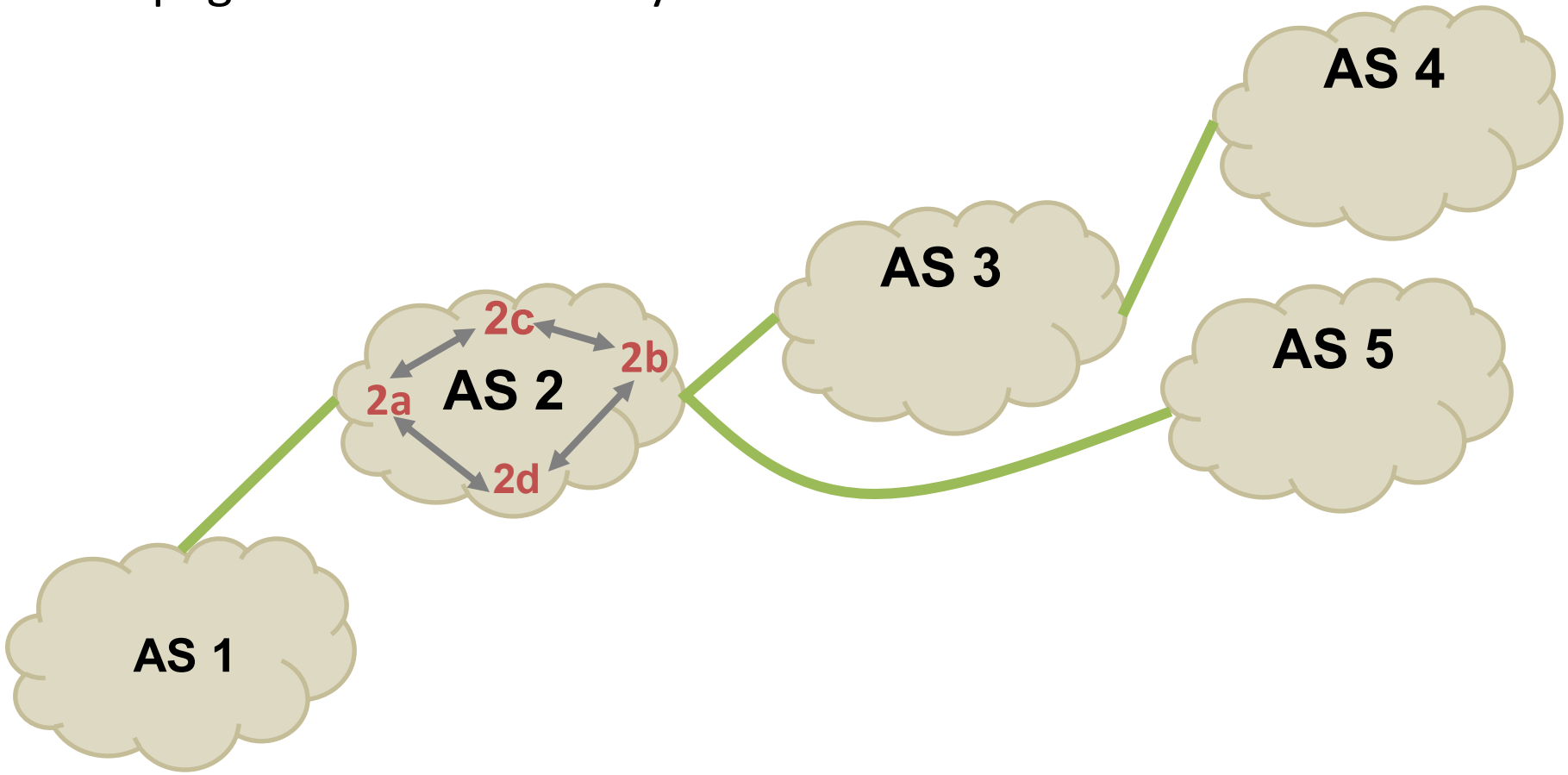
- Key idea: advertise the entire path
  - Distance vector: send *distance metric* per dest d
  - Path vector: send the *entire path* for each dest d



Inter-domain (Inter-ISP) Routing

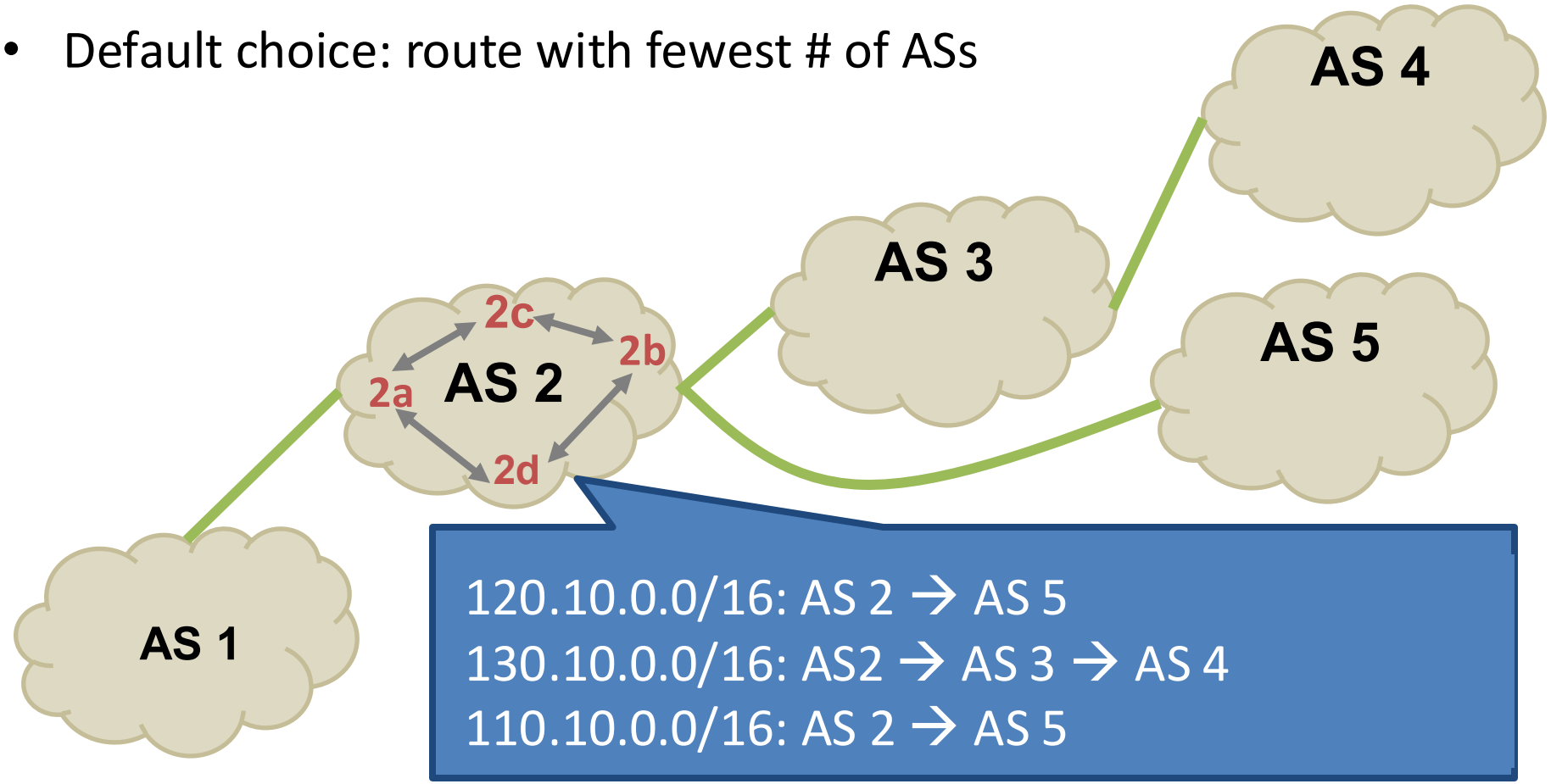
AS2 must:

1. Learn destinations reachable through AS2
2. Propagate this reachability info to all routers in AS2



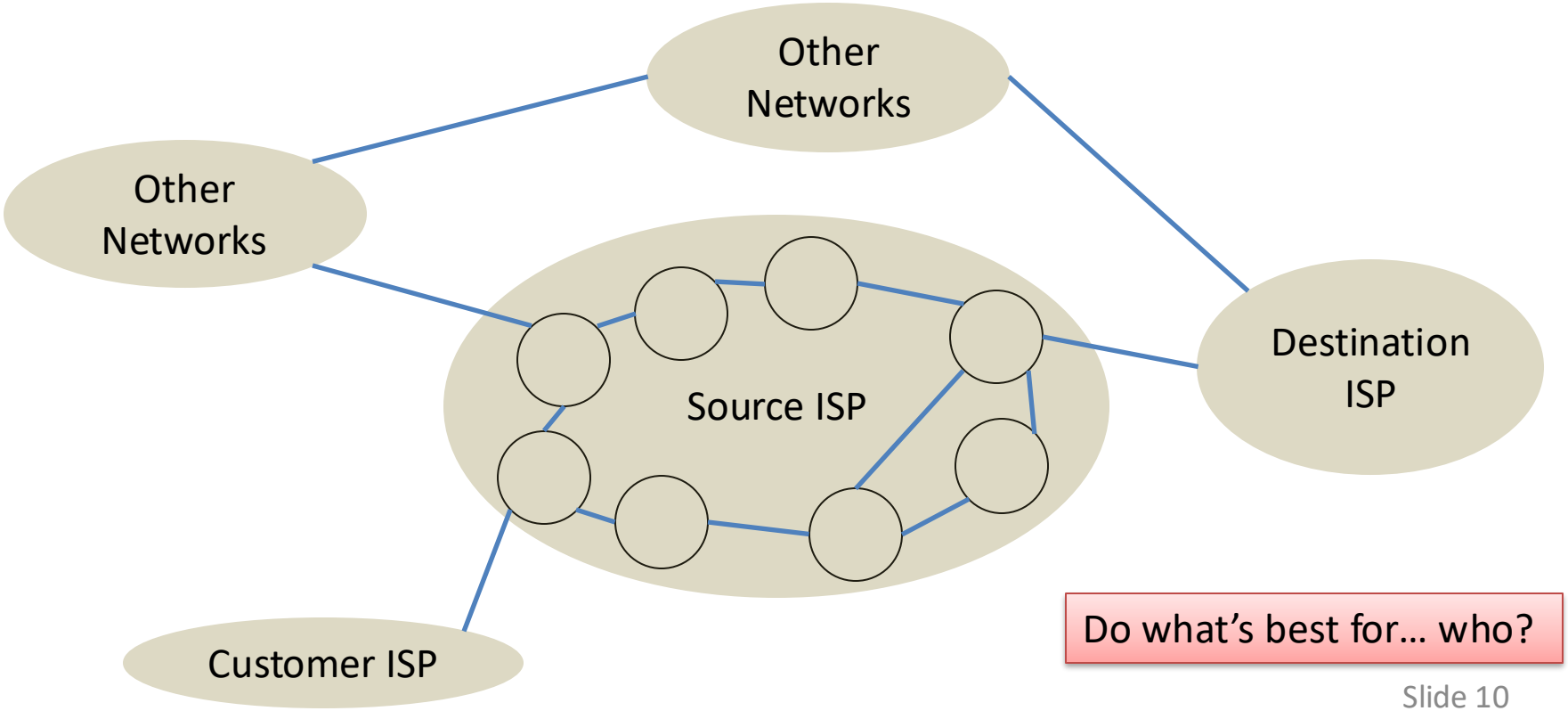
Path Vector Protocol

- AS-path: sequence of ASs a route traverses
  - Like distance vector, plus additional information
- Used for loop detection and to apply policy
- Default choice: route with fewest # of ASs



Routing Policy

- How should the ISP route the customer's traffic to the destination?



Do what's best for… who?

Which routes a BGP router advertises will depend on…

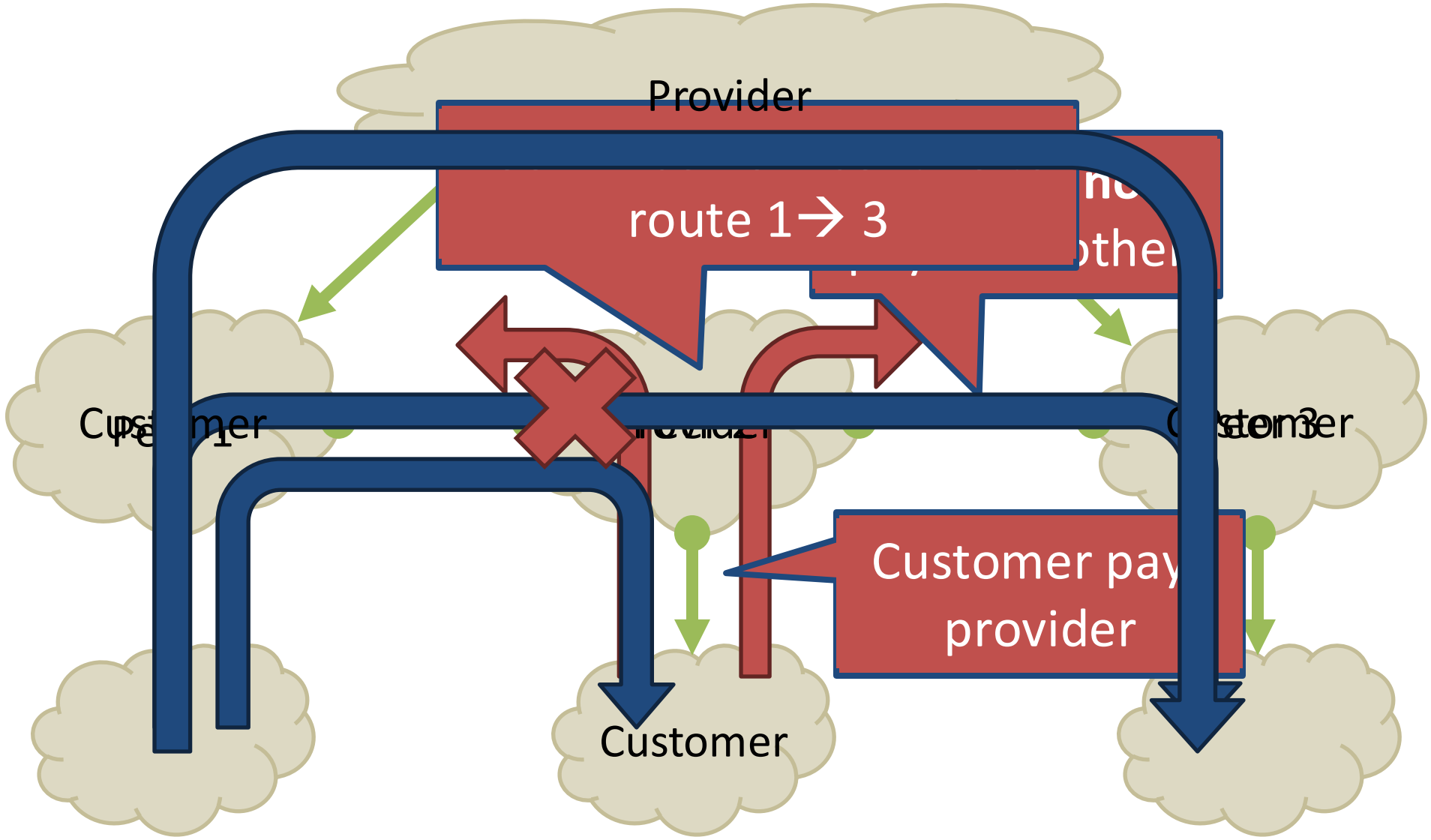
A. which ISPs have contractual agreements.

B. the shortest path to a subnet/prefix.

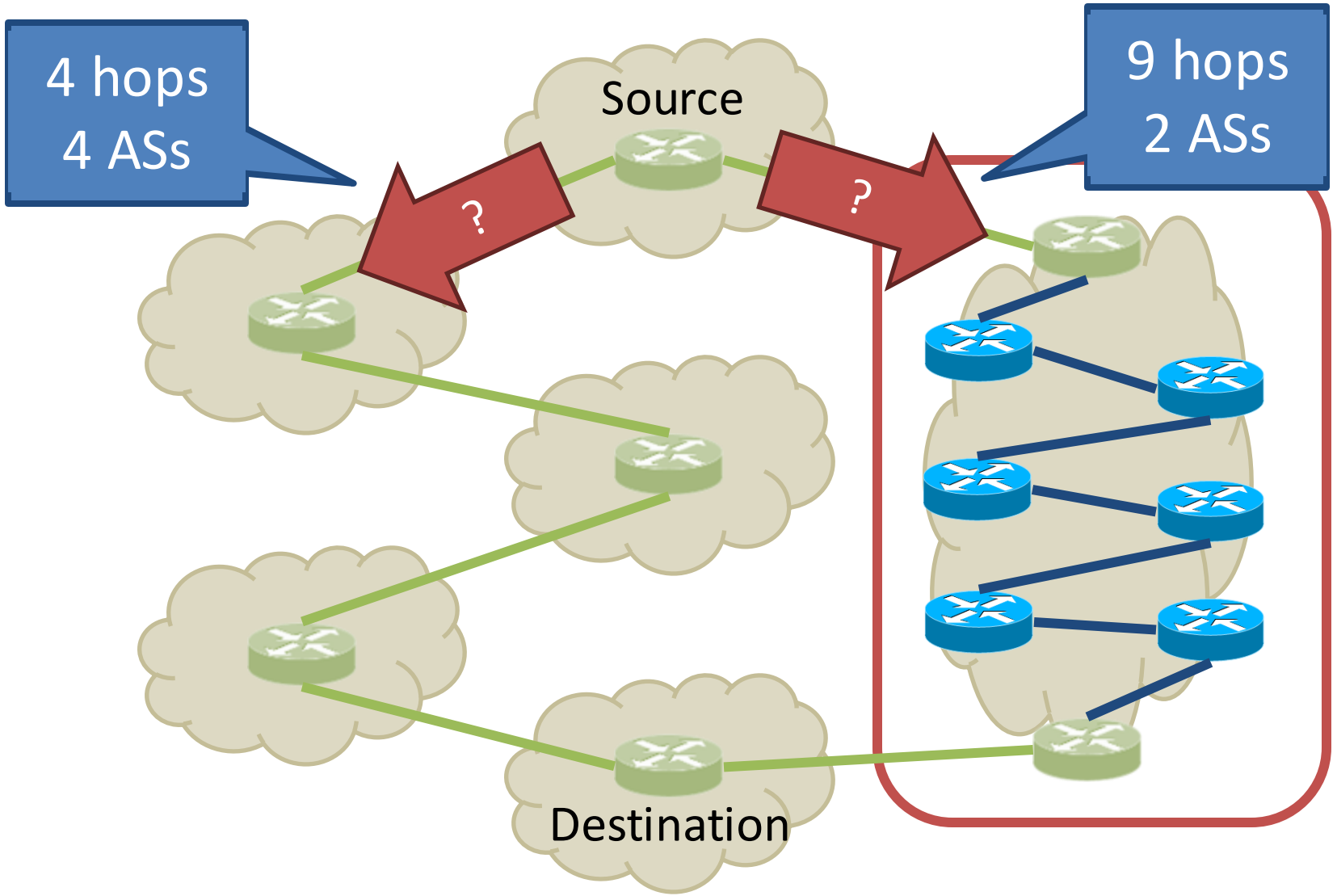
C. which subnets are customers of an ISP.

D. More than one of the above. (which?)

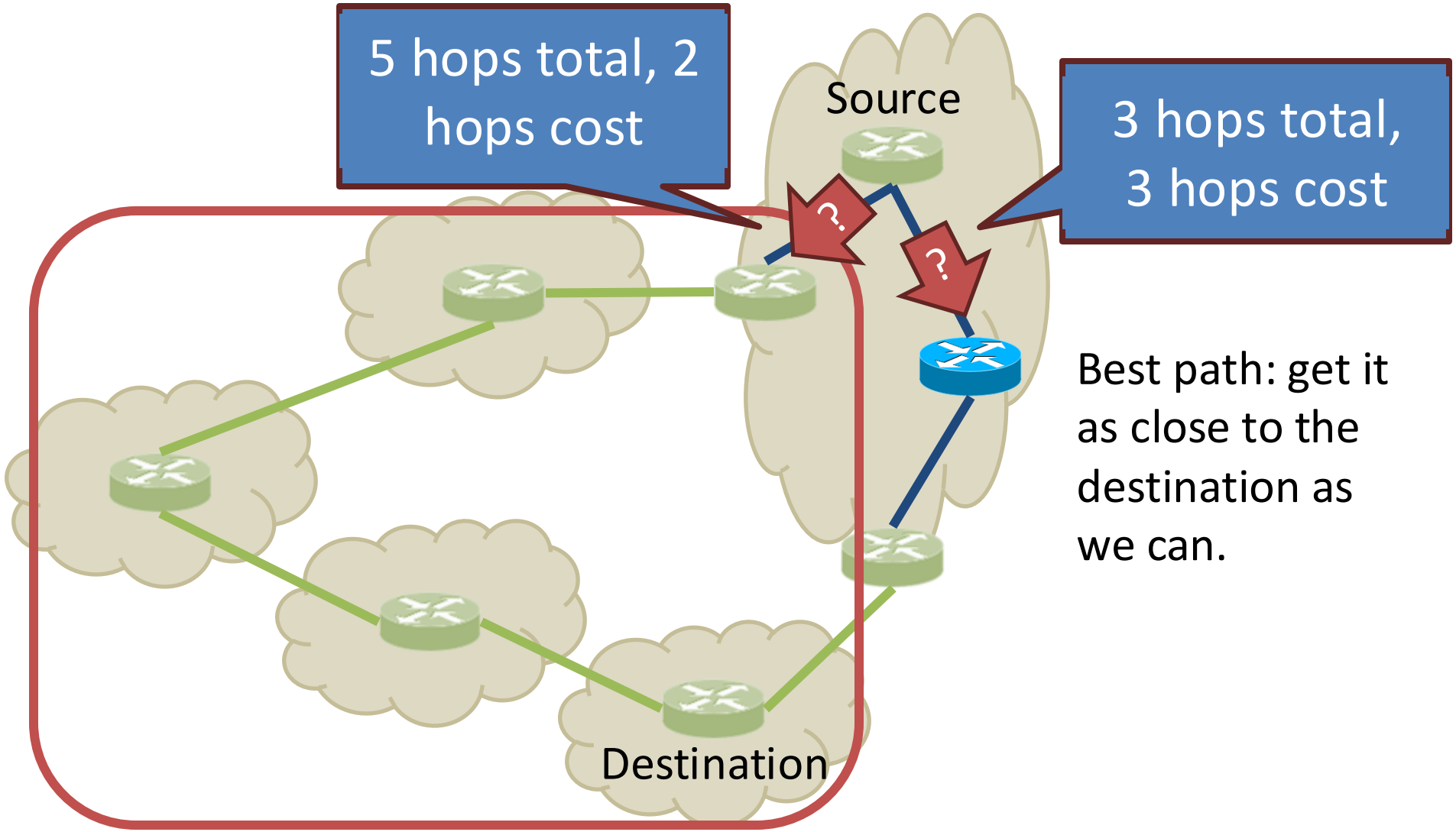
BGP Relationships



Shortest AS Path != Shortest Path

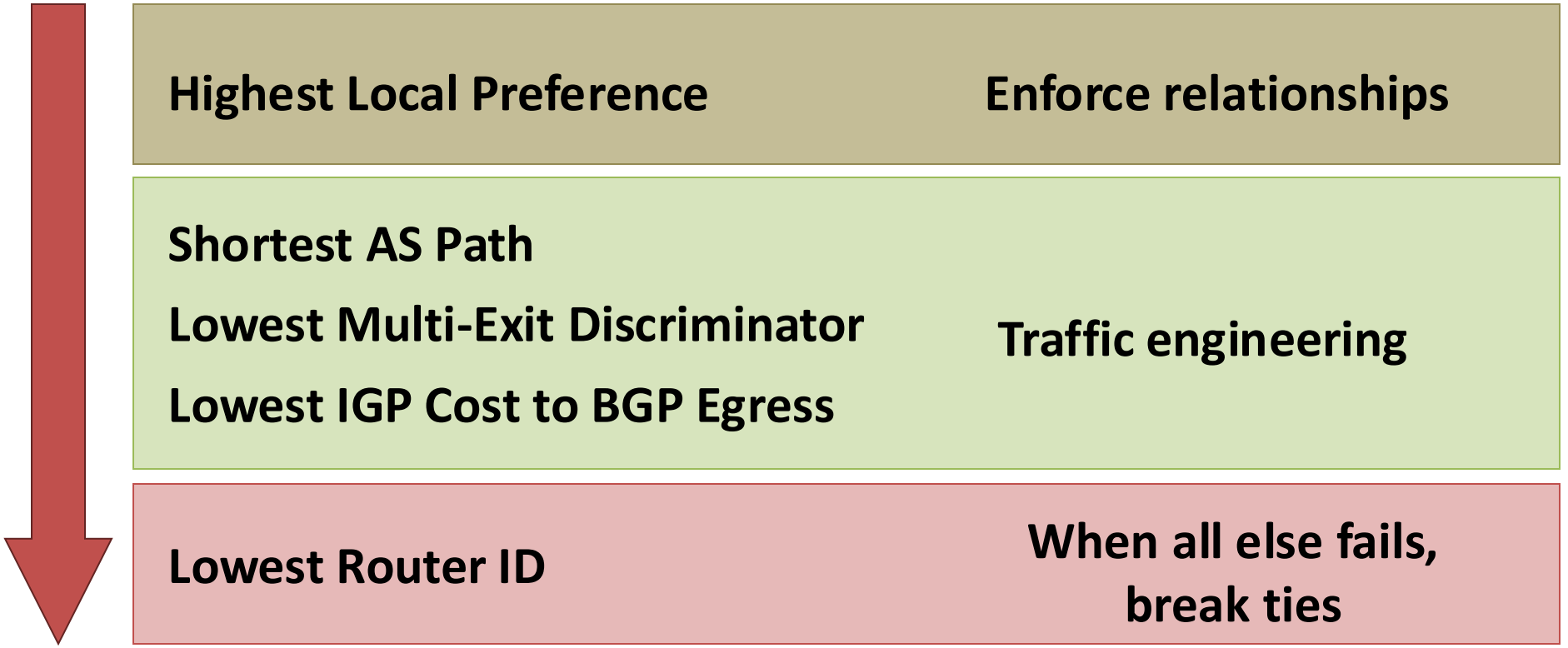


Hot Potato Routing: get rid of packets ASAP!



Best path: get it as close to the destination as we can.

Route Selection Summary



Peering/Interconnection Wars

- Peer

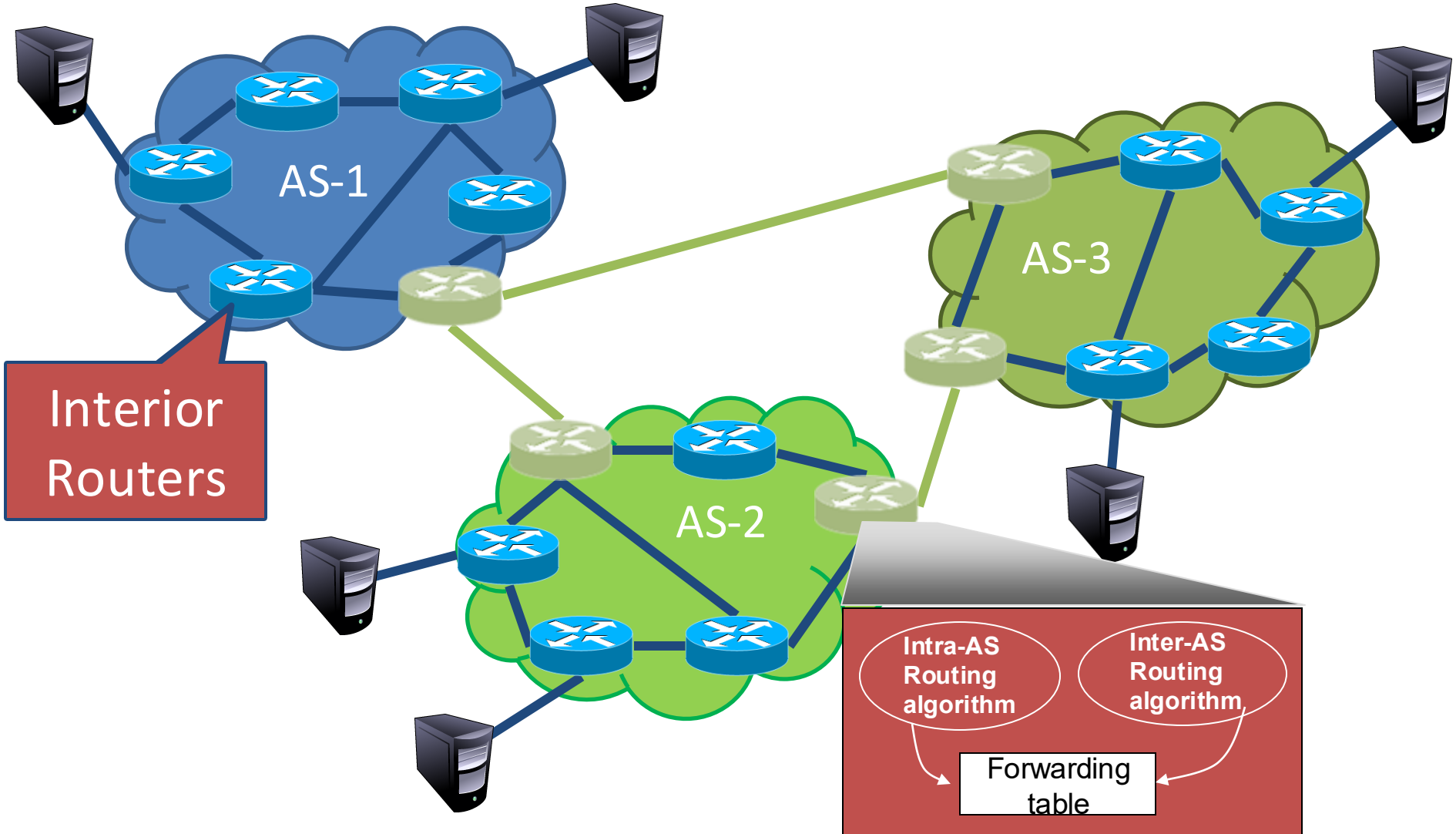
- Reduce upstream costs
- Improve end-to-end performance
- May be the only way to connect to parts of the Internet

- Don't Peer

- You would rather have customers
- Peers are often competitors
- Peering agreements require periodic renegotiation

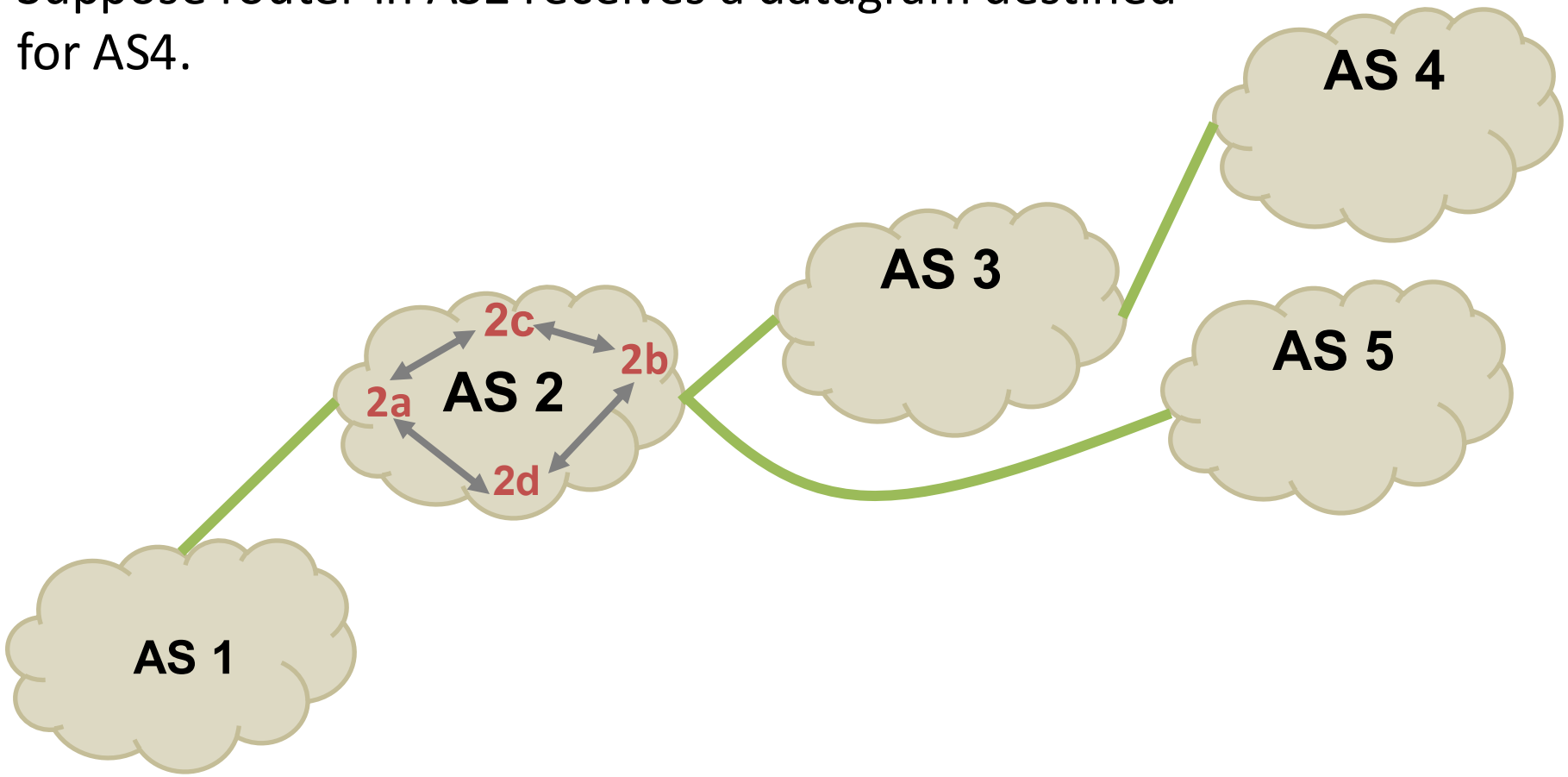
Peering struggles in the ISP world are extremely contentious, agreements are usually confidential

Hierarchical routing: Interconnected ASes



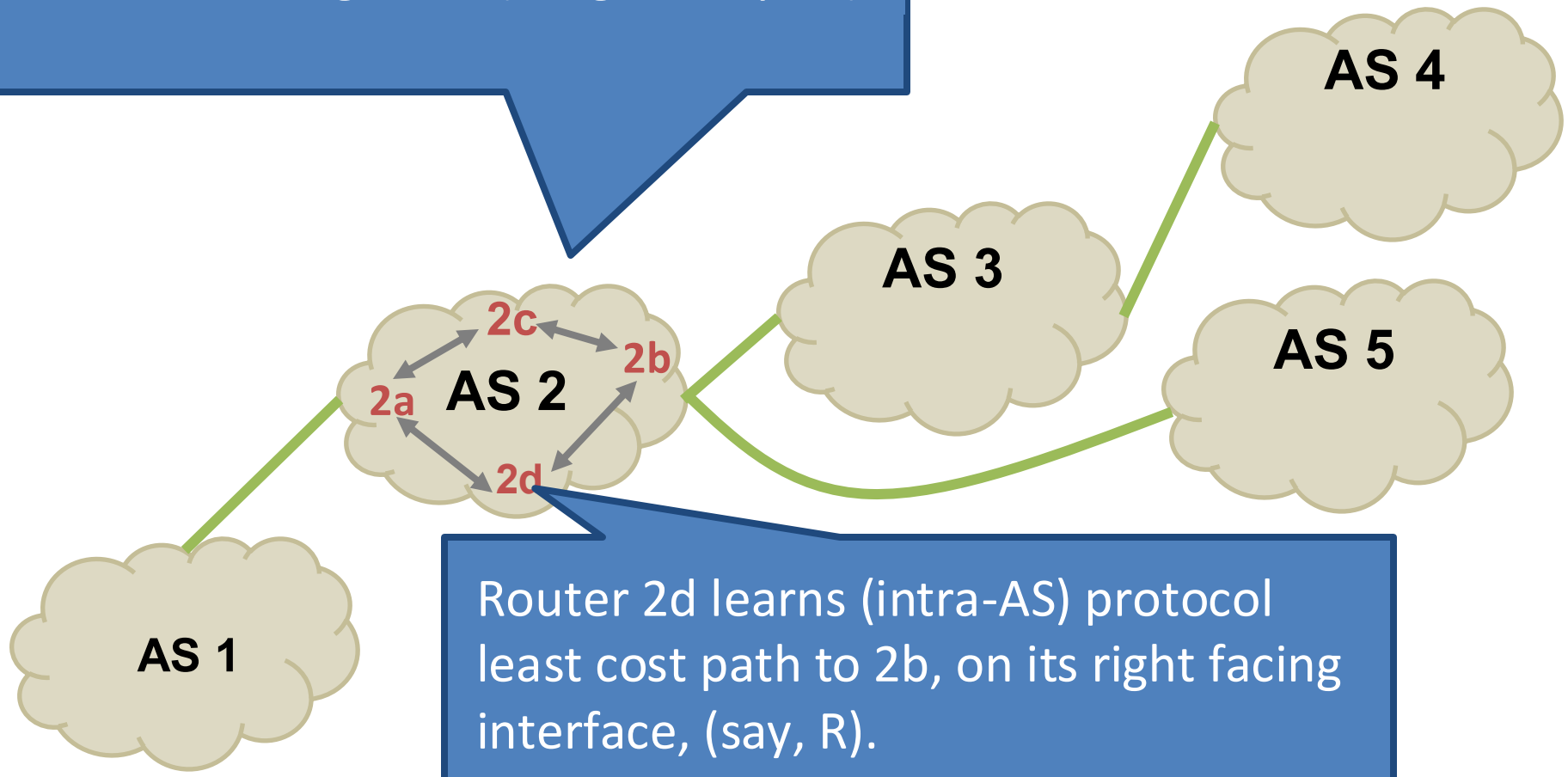
Building the forwarding table in router 2d, for path to AS4

Suppose router in AS2 receives a datagram destined for AS4.



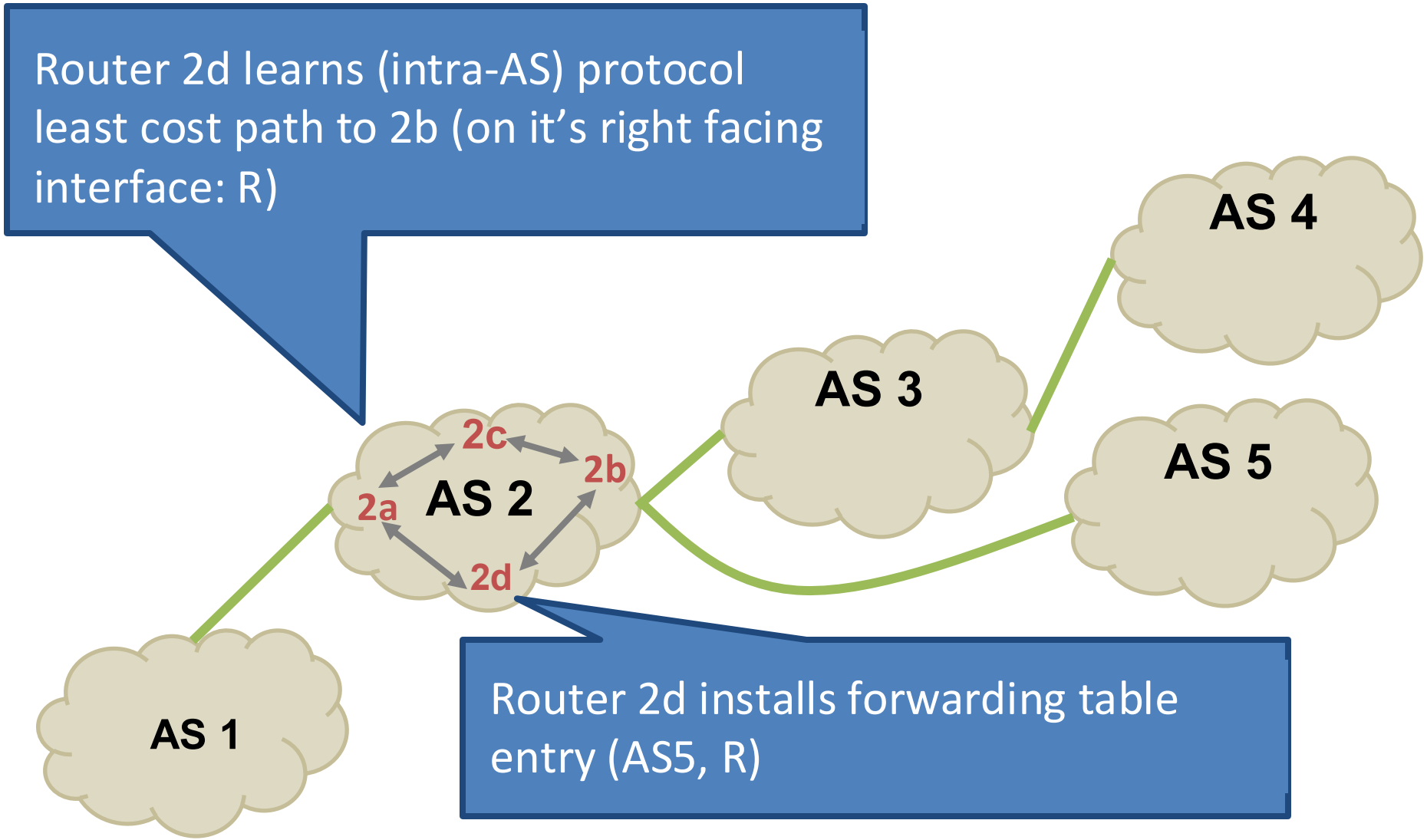
Building the forwarding table in router 2d, for path to AS4

AS2 learns (Inter-AS) protocol that AS4 is reachable through AS3 (via gateway 2b).



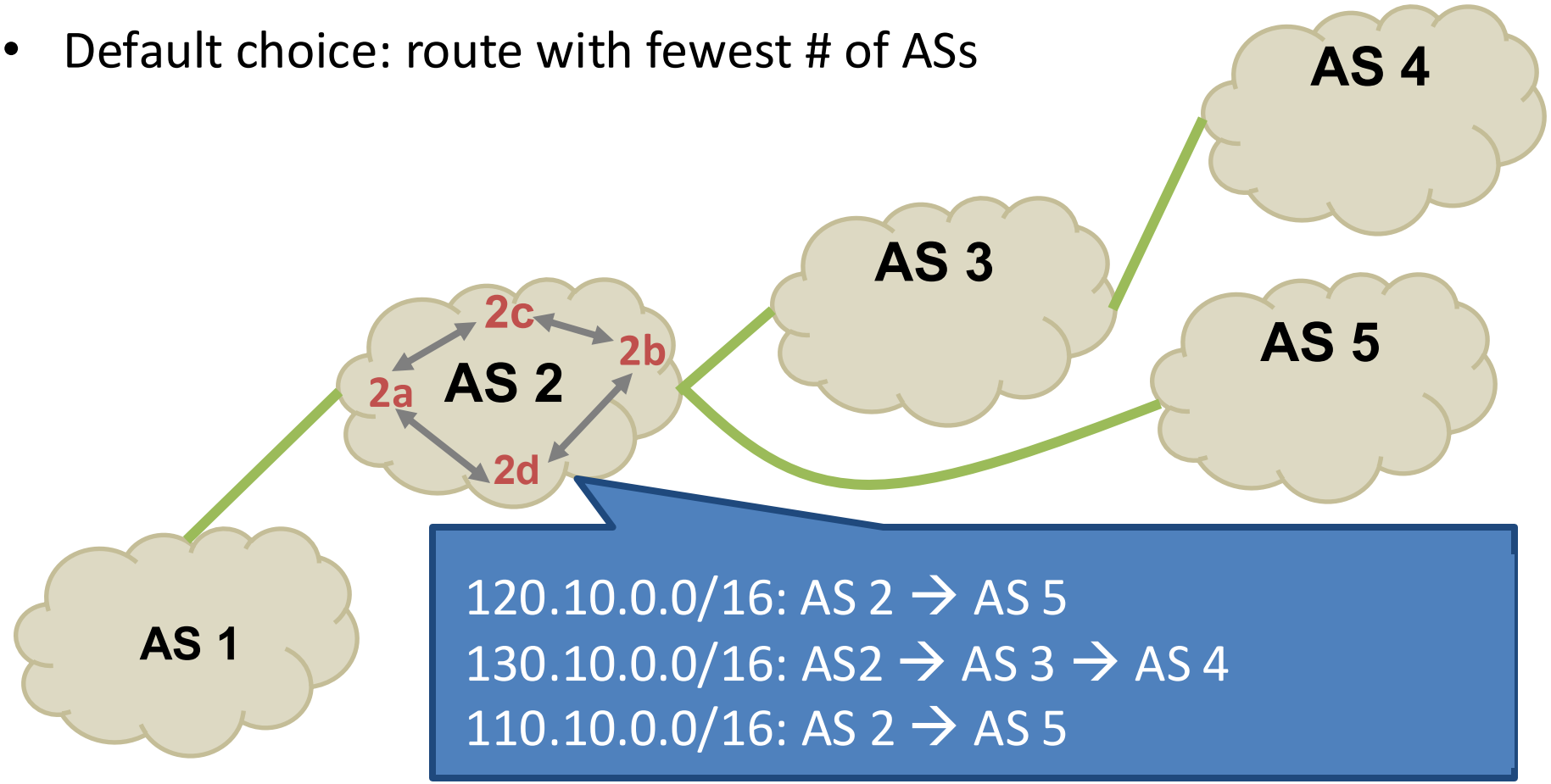
Router 2d learns (intra-AS) protocol least cost path to 2b, on its right facing interface, (say, R).

Building the forwarding table in router 2d, for path to AS 5



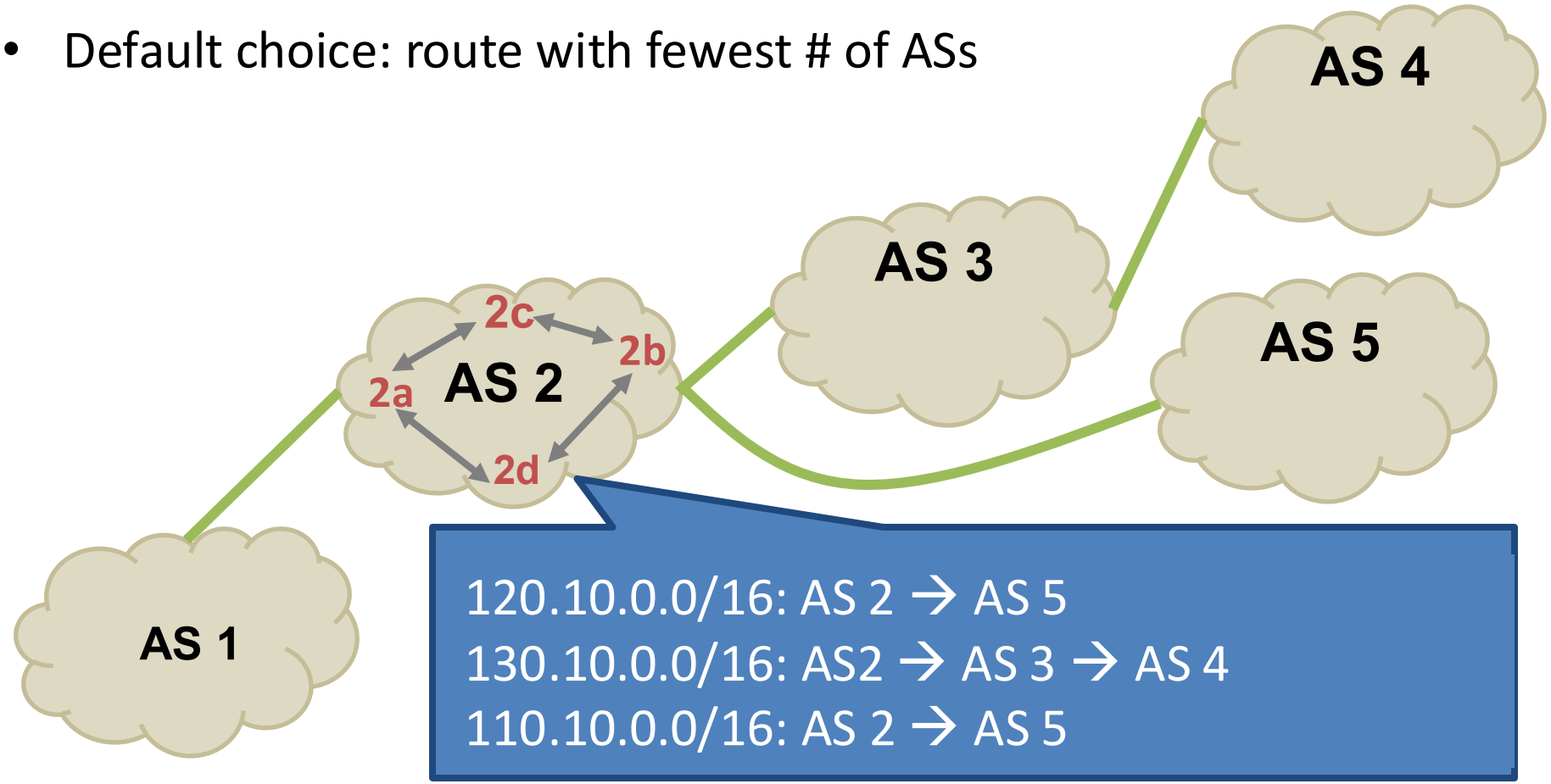
Path Vector Protocol

- AS-path: sequence of ASs a route traverses
  - Like distance vector, plus additional information
- Used for loop detection and to apply policy
- Default choice: route with fewest # of ASs



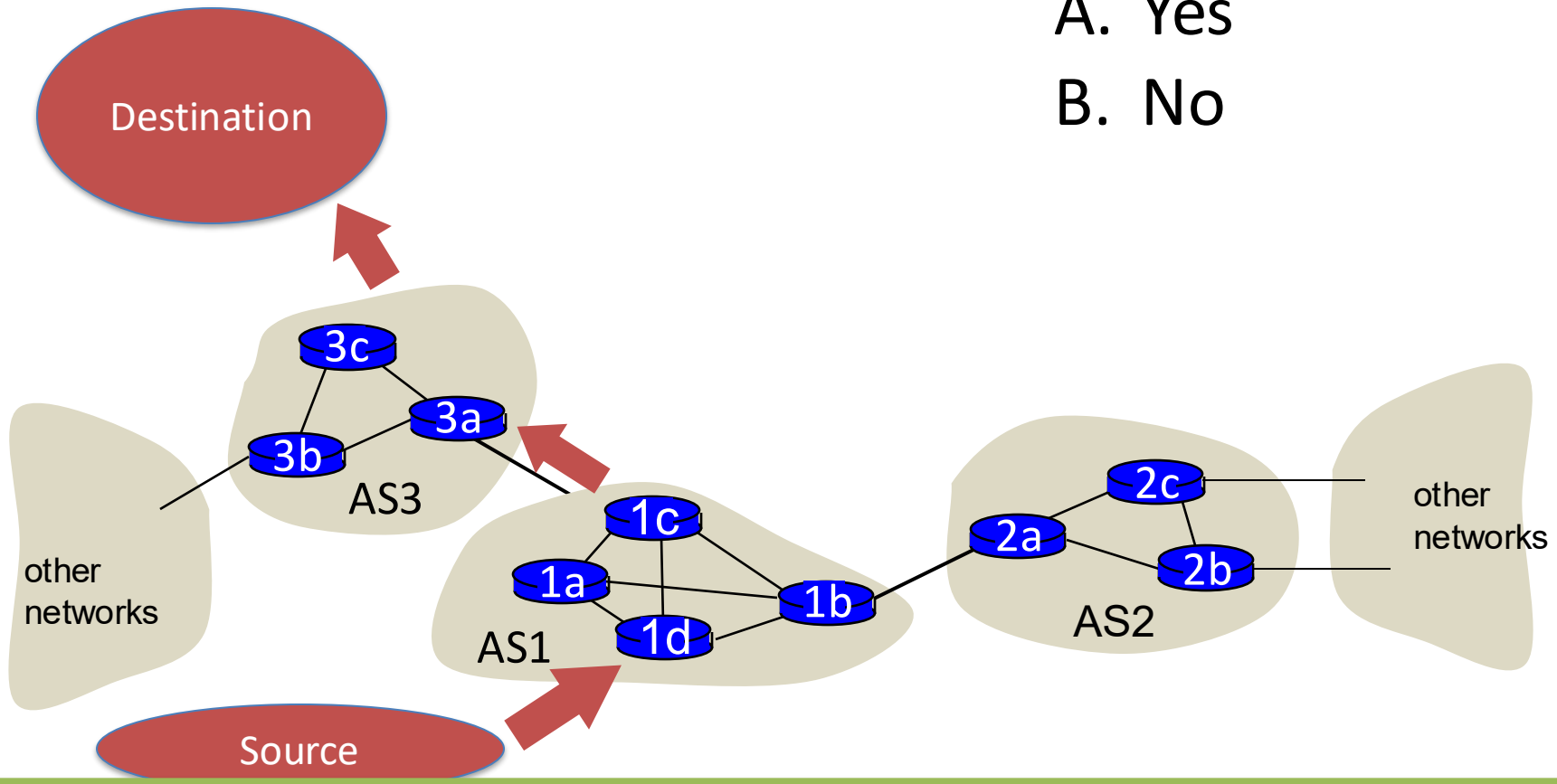
Path Vector Protocol

- AS-path: sequence of ASs a route traverses
  - Like distance vector, plus additional information
- Used for loop detection and to apply policy
- Default choice: route with fewest # of ASs

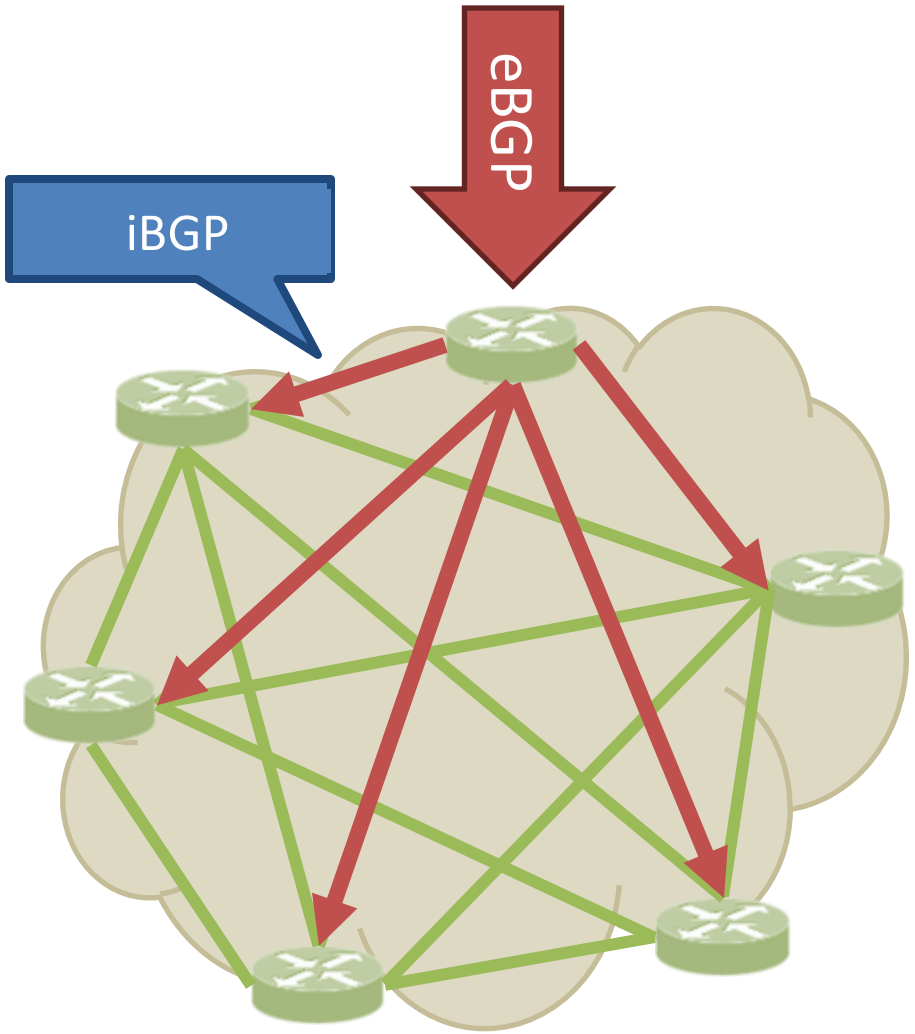


Border routers: exchange AS reachability, Internal routers: exchange intra-AS reachability., Is this sufficient to route from source to destination?

A. Yes
B. No

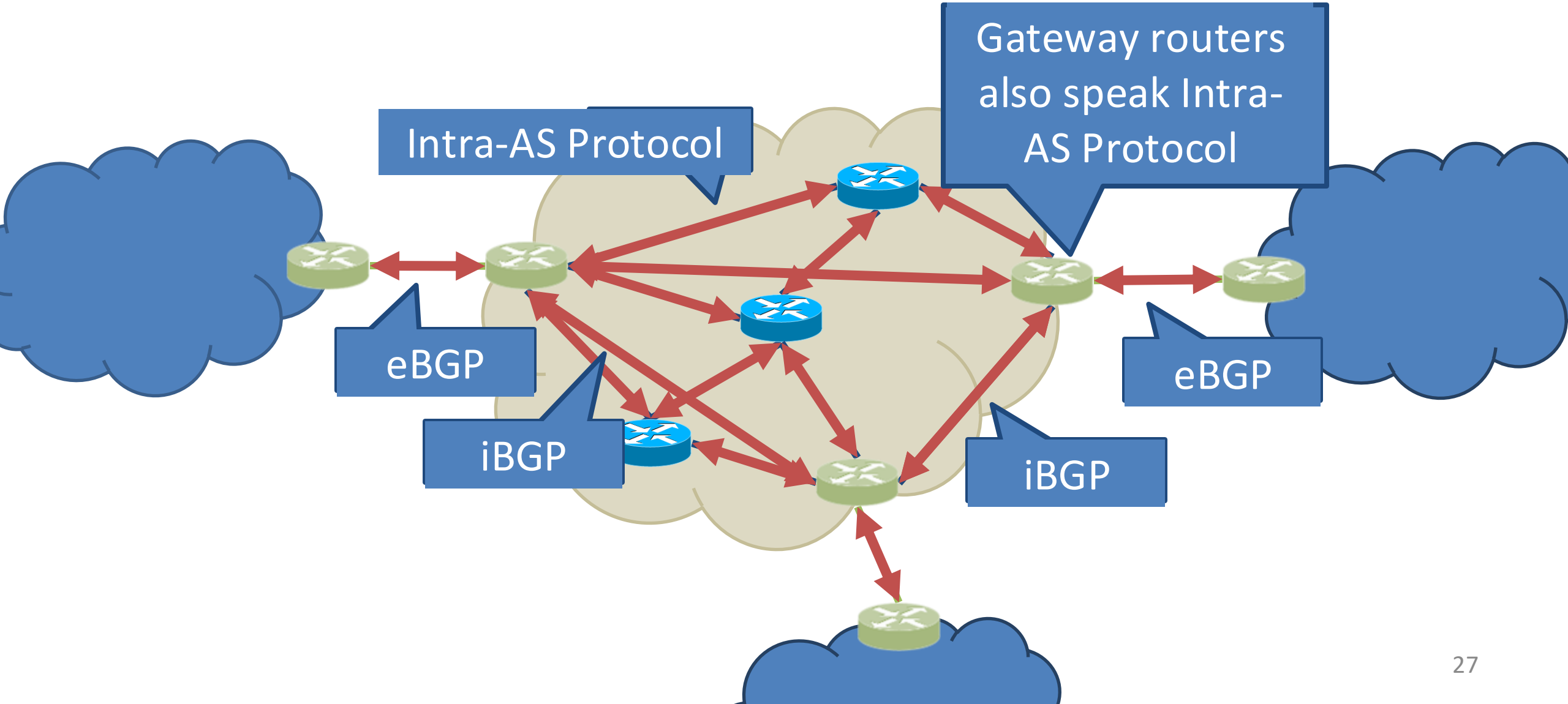


Internet inter-AS routing: BGP



- Question: why do we need iBGP?
  - OSPF does not include BGP policy info
  - Prevents routing loops within the AS
- iBGP updates do not trigger announcements

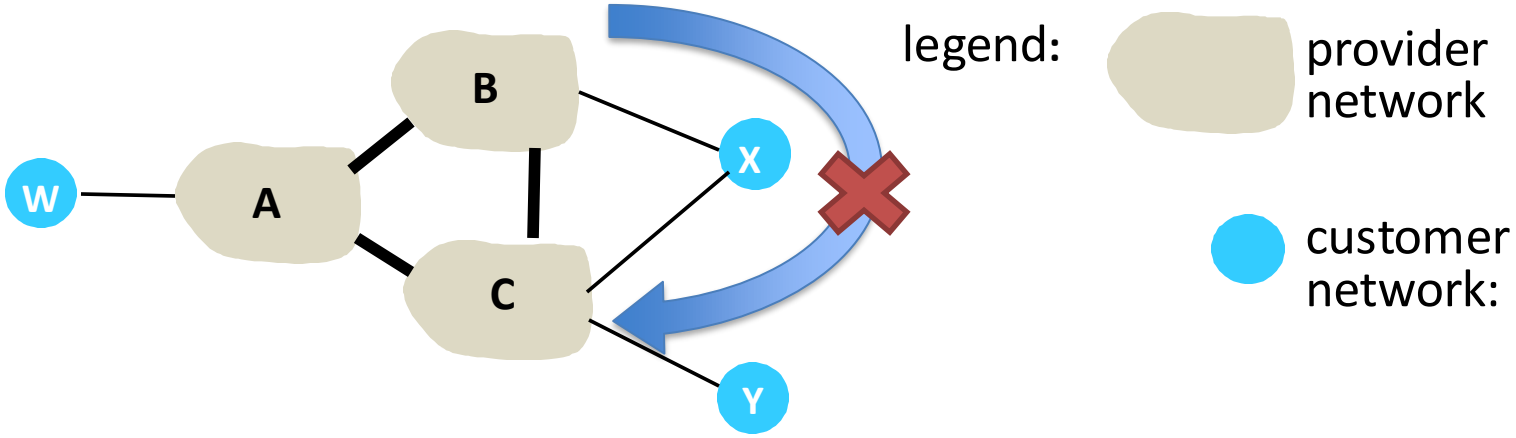
Internet inter-AS routing: BGP



Internet inter-AS routing: BGP

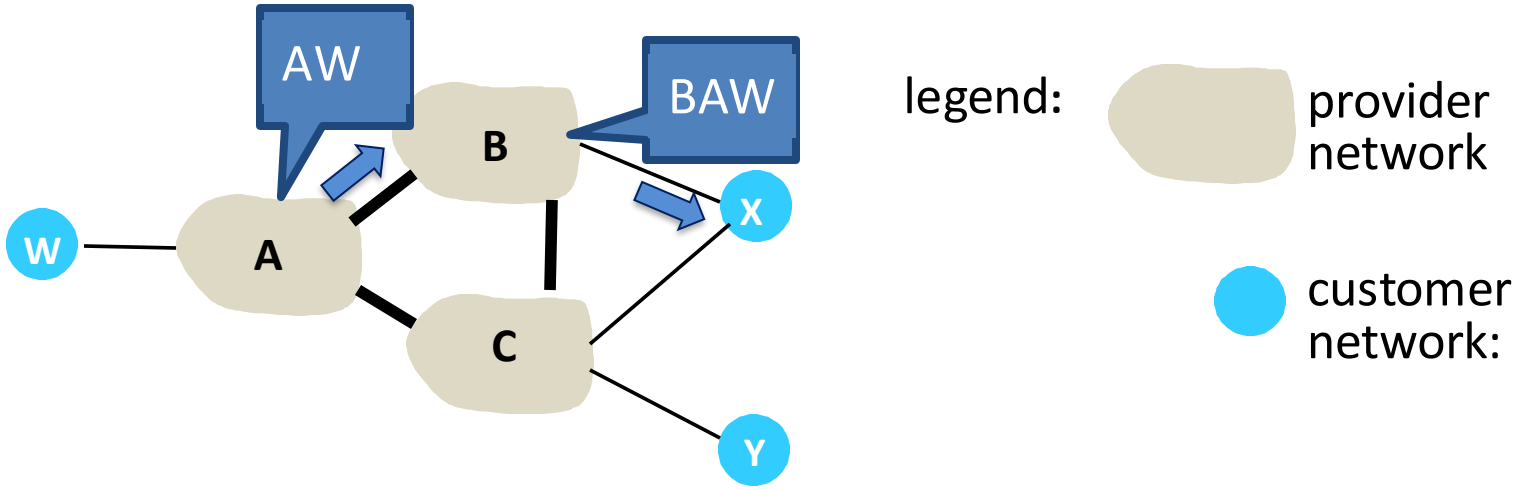
- BGP (Border Gateway Protocol):
  The de facto inter-domain routing protocol
- BGP provides each AS a means to:
  - external BGP: obtain subnet reachability information from neighboring ASs.
  - internal BGP: propagate reachability information to all AS-internal routers.
  - determine "good" routes to other networks based on reachability information and policy.
- Allows a subnet to advertise its prefix to the rest of the Internet

BGP routing policy



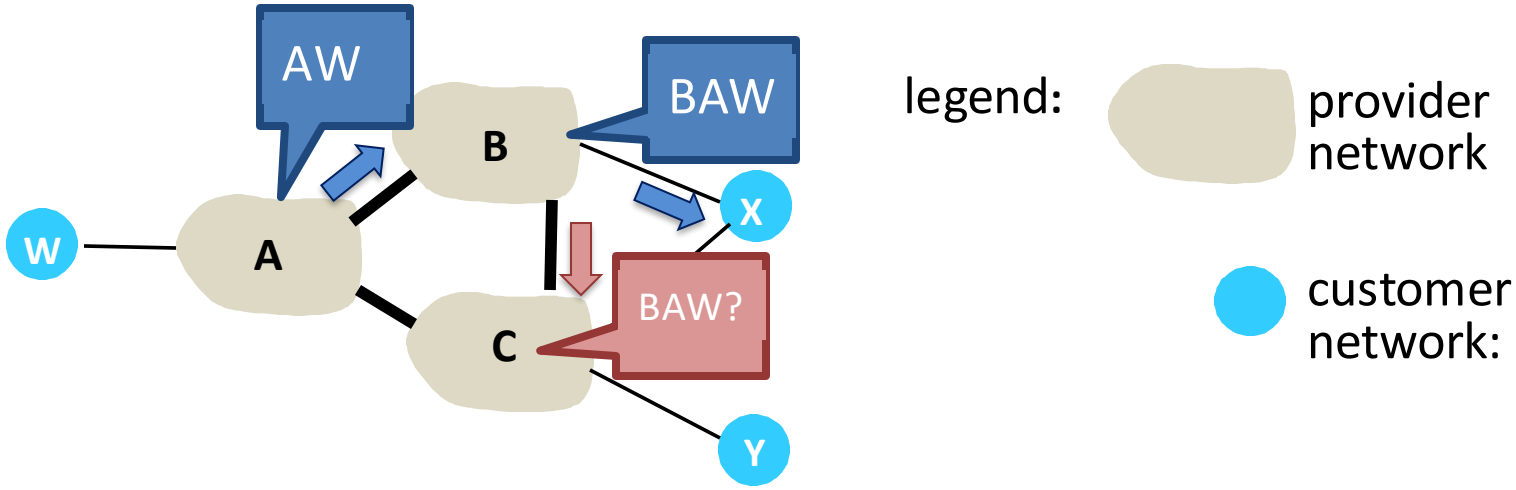
- A,B,C are provider networks
- X,W,Y are customers of the providers
- X is dual-homed: attached to two networks (B and C)
  - X does not want to route from B via X to C
  - .. so X will not advertise to B a route to C

BGP routing policy



- A advertises path AW to B
- B advertises path BAW to X

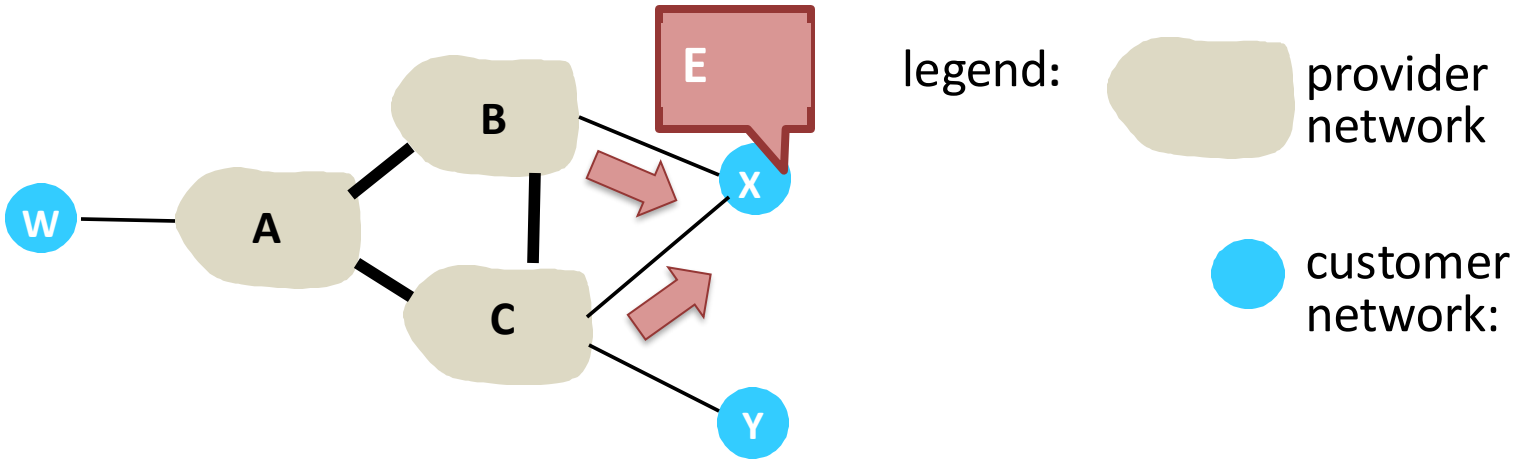
BGP routing policy: Should B advertise path BAW to C?



Should B advertise path BAW to C?

- B gets no "revenue" for routing CBAW since neither W nor C are B's customers
- B wants to force C to route to w via A
- B wants to route *only* to/from its customers!

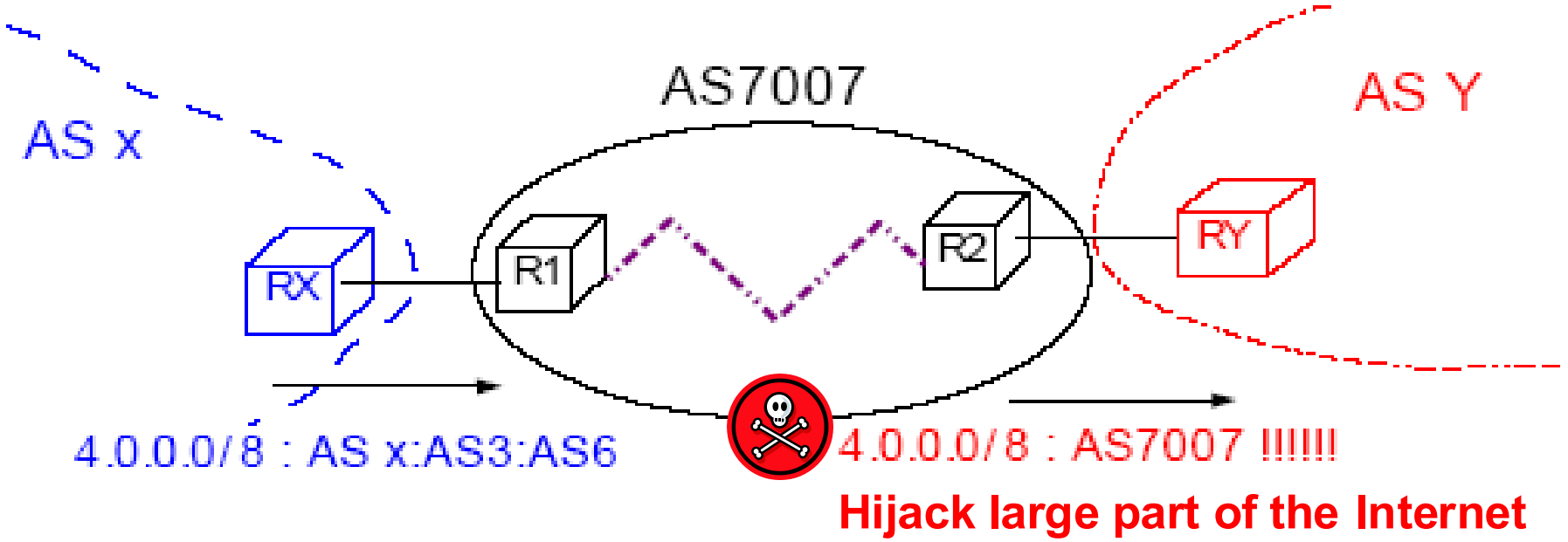
BGP routing policy gone wrong



- x advertises a path to E (that it is not connected to).
- all traffic starts to flow into x from B and C!

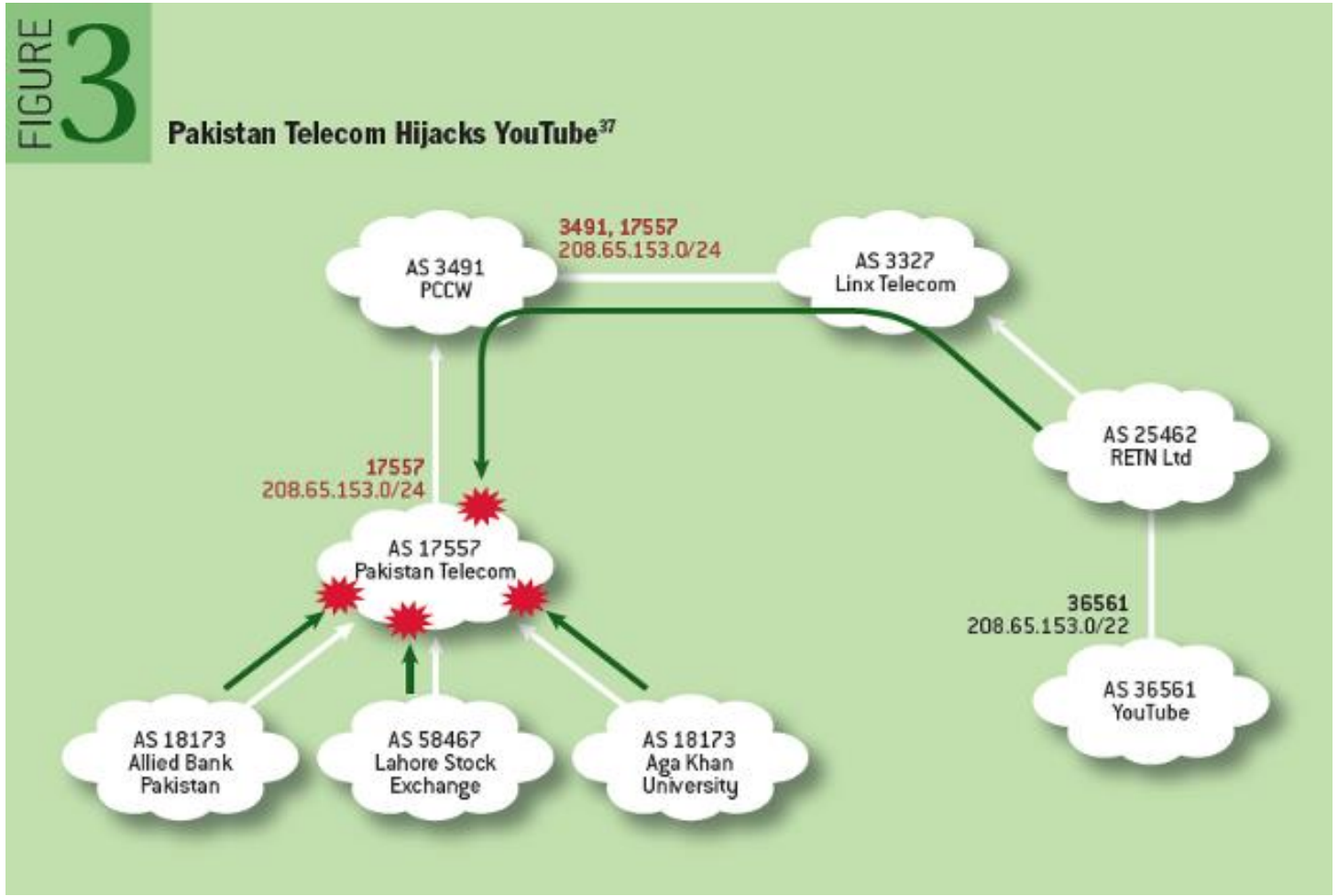
Faulty redistribution can be dangerous!

- AS7007 incident (April, 1997):



https://www.kentik.com/blog/a-brief-history-of-the-internets-biggest-bgp-incidents/

Some "incorrect" route announcements are intentional: BGP Hijacking



- Pakistan Telecom's AS 17557 launched **a subprefix hijack** by originating the subprefix 208.65.153.0/24 of YouTube's prefix 208.65.153.0/22 to its customer ASes in Pakistan
- Traffic destined for YouTube's servers in AS 36561 would instead be **forwarded to the *longer* IP prefix** originated by Pakistan Telecom's AS 17557
- Bogus BGP announcement **leaked out of Pakistan** to large ISP PCCW

Summary

- As we've seen before (DNS), a hierarchy can help manage state storage constraints.
  - intra-AS routing: lots of info about local routes
  - inter-AS routing: less info about far away routes

- BGP: the inter-AS routing protocol for the Internet
  - Decisions often contractual

- BGP advertises AS prefixes, including:
  - entire path of ASes along the way
  - which border router heard the advertisement (Next Hop)

Additional Info: Inter-Domain Routing Challenges

- BGP4 is the only inter-domain routing protocol currently in use world-wide
- Issues?
  - Lack of security
  - Ease of misconfiguration
  - Poorly understood interaction between local policies
  - Poor convergence
  - Lack of appropriate information hiding
  - Non-determinism
  - Poor overload behavior

Additional Info:
Lots of research into how to fix this

- Security
  - BGPSEC, RPKI
- Misconfigurations, inflexible policy
  - SDN
- Policy Interactions
  - PoiRoot (root cause analysis)
- Convergence
  - Consensus Routing
- Inconsistent behavior
  - LIFEGUARD, among others

Additional Info
Why are these still issues?

- Backward compatibility
- Buy-in / incentives for operators
- Stubbornness

Very similar issues to IPv6 deployment

Additional Info:
Why Network Reliability Remains Hard

- Visibility
  - IP provides no built-in monitoring
  - Economic disincentives to share information publicly

- Control
  - Routing protocols optimize for policy, not reliability
  - Outage affecting your traffic may be caused by distant network

- Detecting, isolating and repairing network problems for Internet paths remains largely a slow, manual process