Network Routing:

Link State Routing and BGP

Reading: KR 4.3, 4.5

Link-State Routing

- Net topology, link costs are distributed to all nodes
  - All nodes have same info
  - Thus can compute any types of routes
- Each node computes its shortest paths from itself to all other nodes
  - E.g., use Dijkstra's algorithm
- Link state distribution accomplished via "link state broadcast"

Link State Broadcast

- The hard part is link state broadcast
  - Basic approach: forward a link state (link ID, link status) to all links except the incoming link

- Question: what are the problems the link state broadcast needs to handle?
  - Broadcast loop
  - Ordering of events (link up and down)
  - Network partitioning and then merge

Link State Broadcast

- Each link update is given a sequence number: (initiator, seq#, link, status)
  - The initiator should increase the seq# for each new update
- If the seq# of an update of a link is not higher than the highest seq# a router has seen, drop the update
- Otherwise, forward it to all links except the incoming link

- Each seq# has an age field (why?)
- Updates are sent periodically (why?)

OSPF (Open Shortest Path First)

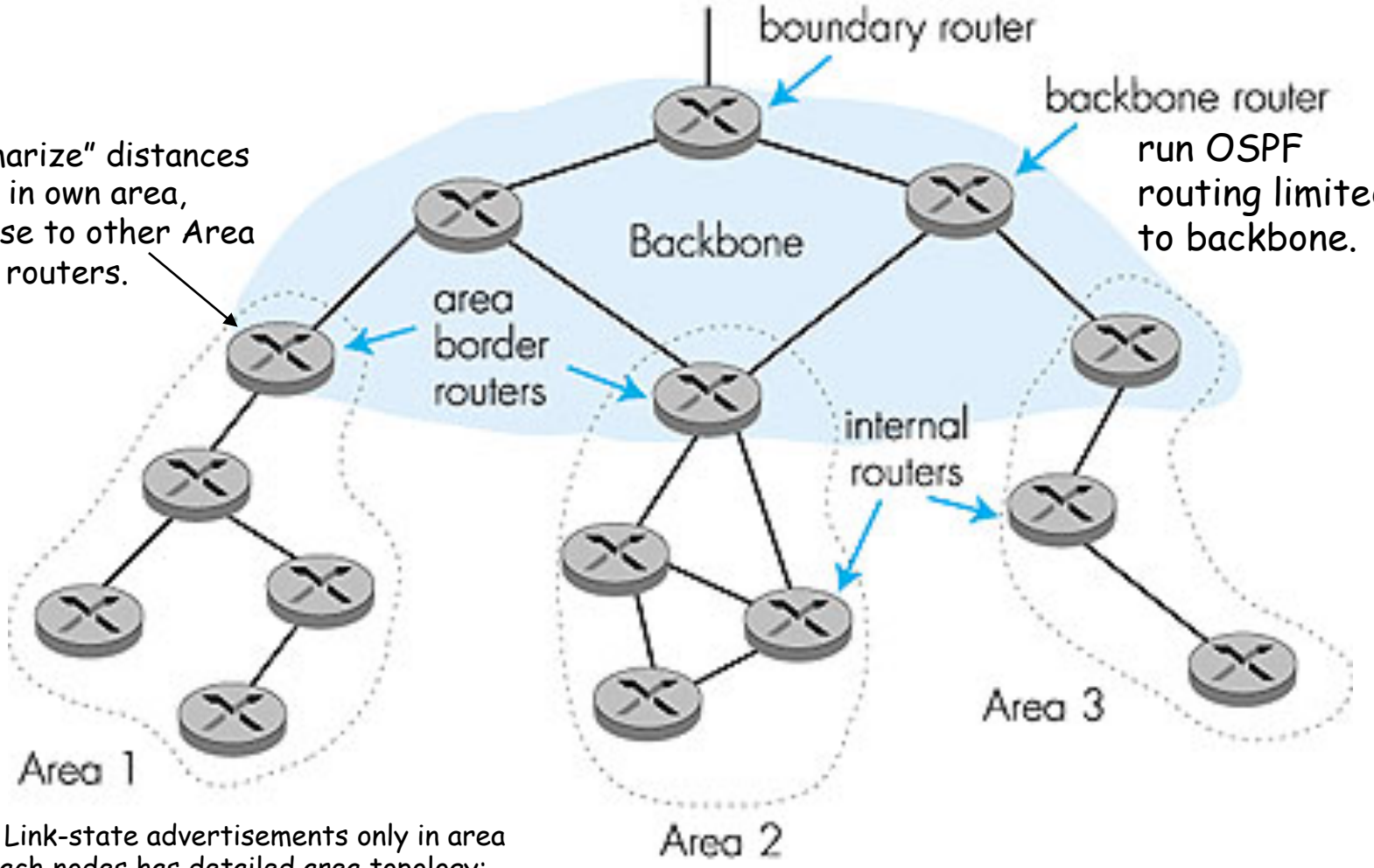
- "Open": publicly available
- Uses Link State algorithm
  - Link state (LS) packet dissemination
  - Topology map at each node
  - Route computation using Dijkstra's algorithm

http://en.wikipedia.org/wiki/Open_Shortest_Path_First

OSPF "Advanced" Features (not in RIP)

- Multiple same-cost paths allowed (only one path in RIP)
- For each link, multiple cost metrics for different Type Of Service (eg, satellite link cost set "low" for best effort; high for real time)
- Security: all OSPF messages authenticated (to prevent malicious intrusion); TCP connections used
- Hierarchical OSPF

Hierarchical OSPF

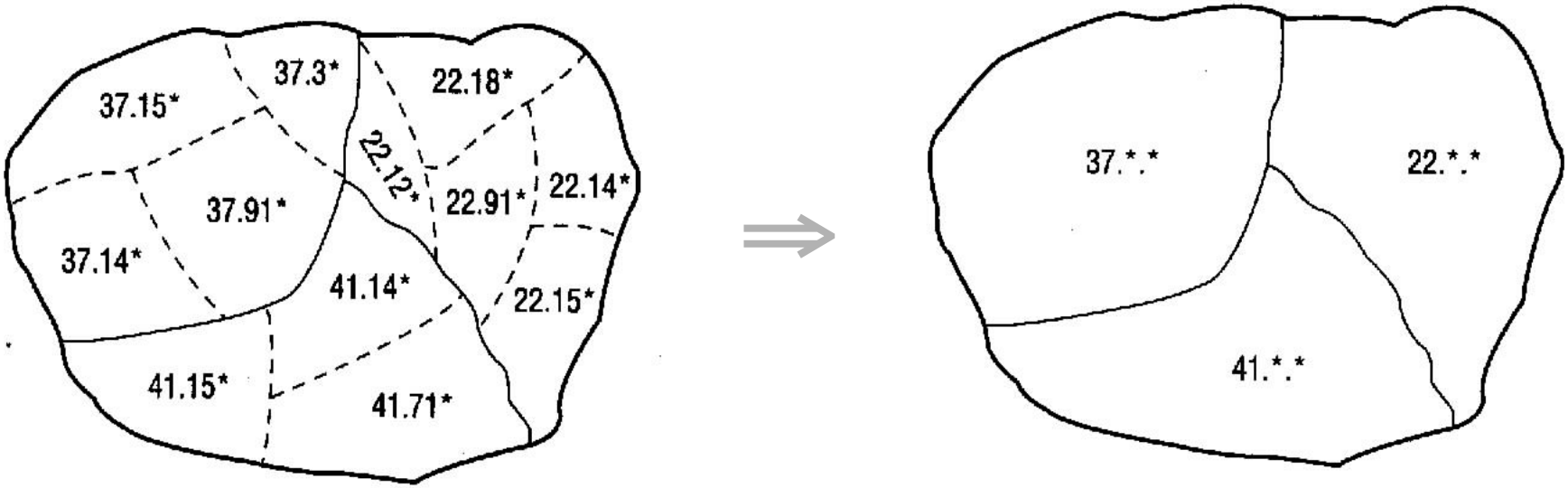


- Link-state advertisements only in area each nodes has detailed area topology;
- only know direction (shortest path) to nets in other areas.

Two-level hierarchy: local area, backbone.

Why Hierarchy?

- Information hiding (filtered) => reduce computation, bandwidth, storage



Discussion: Link State Routing

- What do you like about link state routing?
- What do you not like about link state routing?

Question to think about: which routing protocol (DV or LS) should the Internet use?

Outline

- Recap
- Distance vector protocols
- Link state protocols
- Routing in the Internet
  - overview

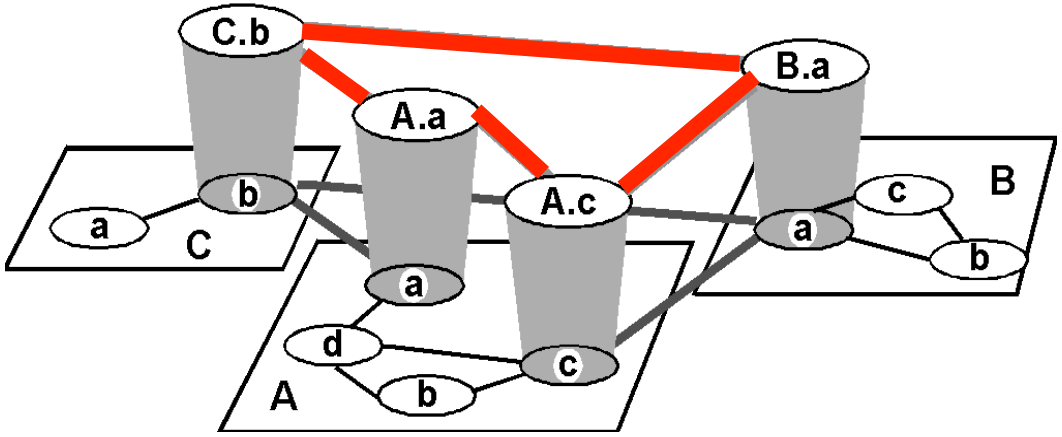
Routing in the Internet

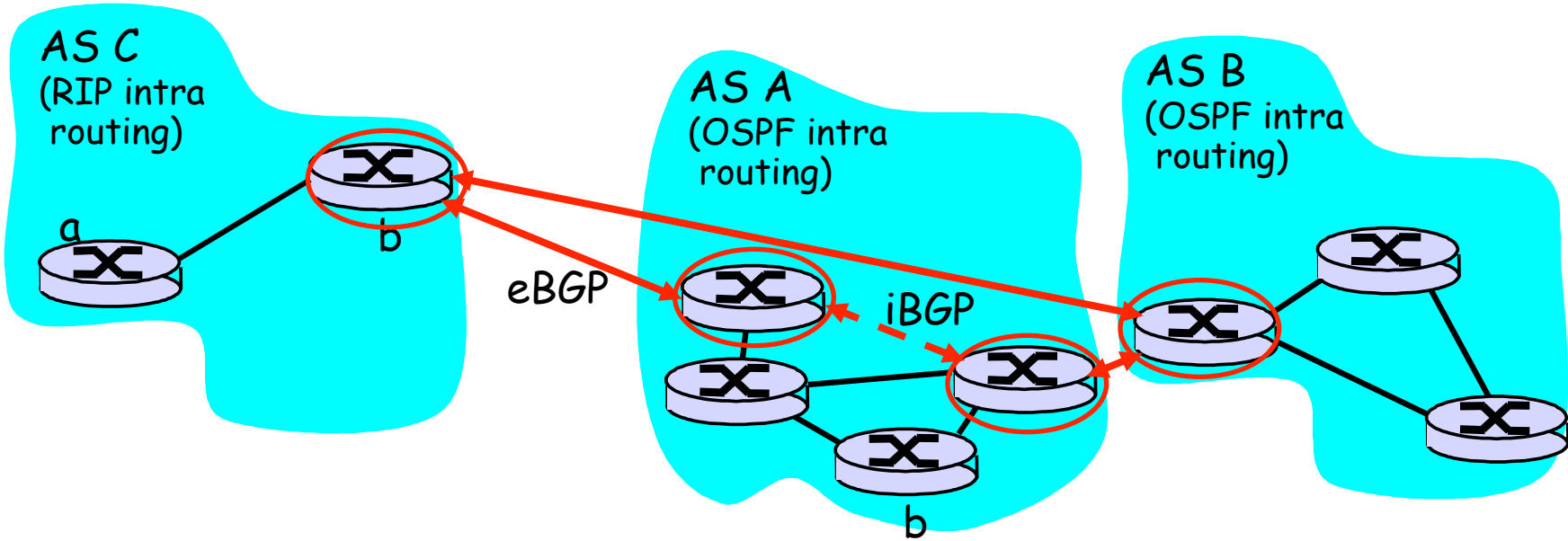
- The Global Internet consists of Autonomous Systems (AS) interconnected with each other
  - An AS is identified by an AS Number (ASN), e.g. Yale ASN is 29

Routing with AS

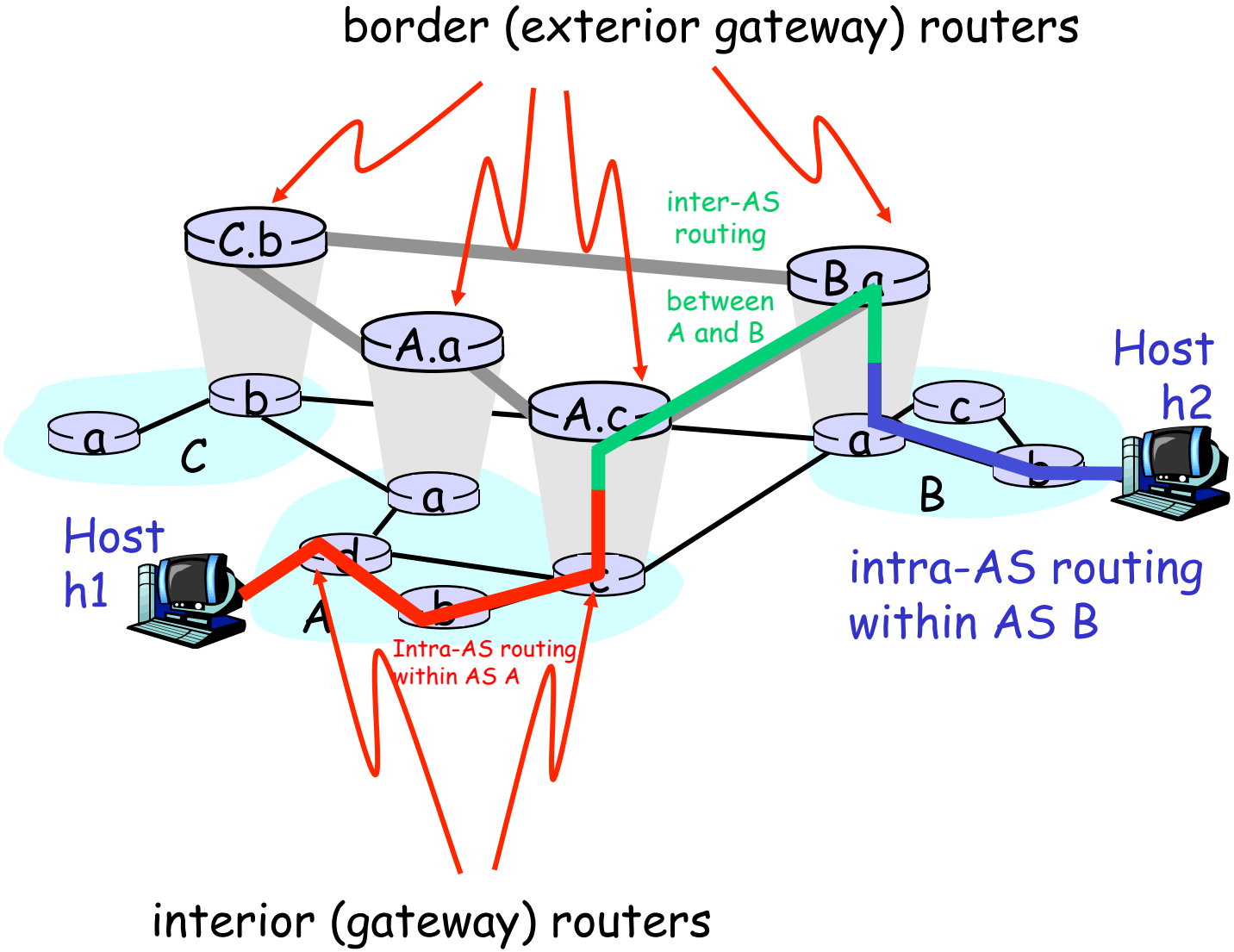
- Intra-AS
  - Different AS's may run different routing protocols
  - A protocol running insides an AS is called an Interior Gateway Protocol (IGP)
    - RIP: Routing Information Protocol
    - OSPF: Open Shortest Path First
    - IS-IS: very similar to OSPF
    - E/IGRP: Interior Gateway Routing Protocol (Cisco)
- Inter-AS
  - A protocol runs among AS's is also called an Exterior Gateway Protocol (EGP)
  - For global connectivity, a single inter-domain routing protocol

Routing in the Internet: Example

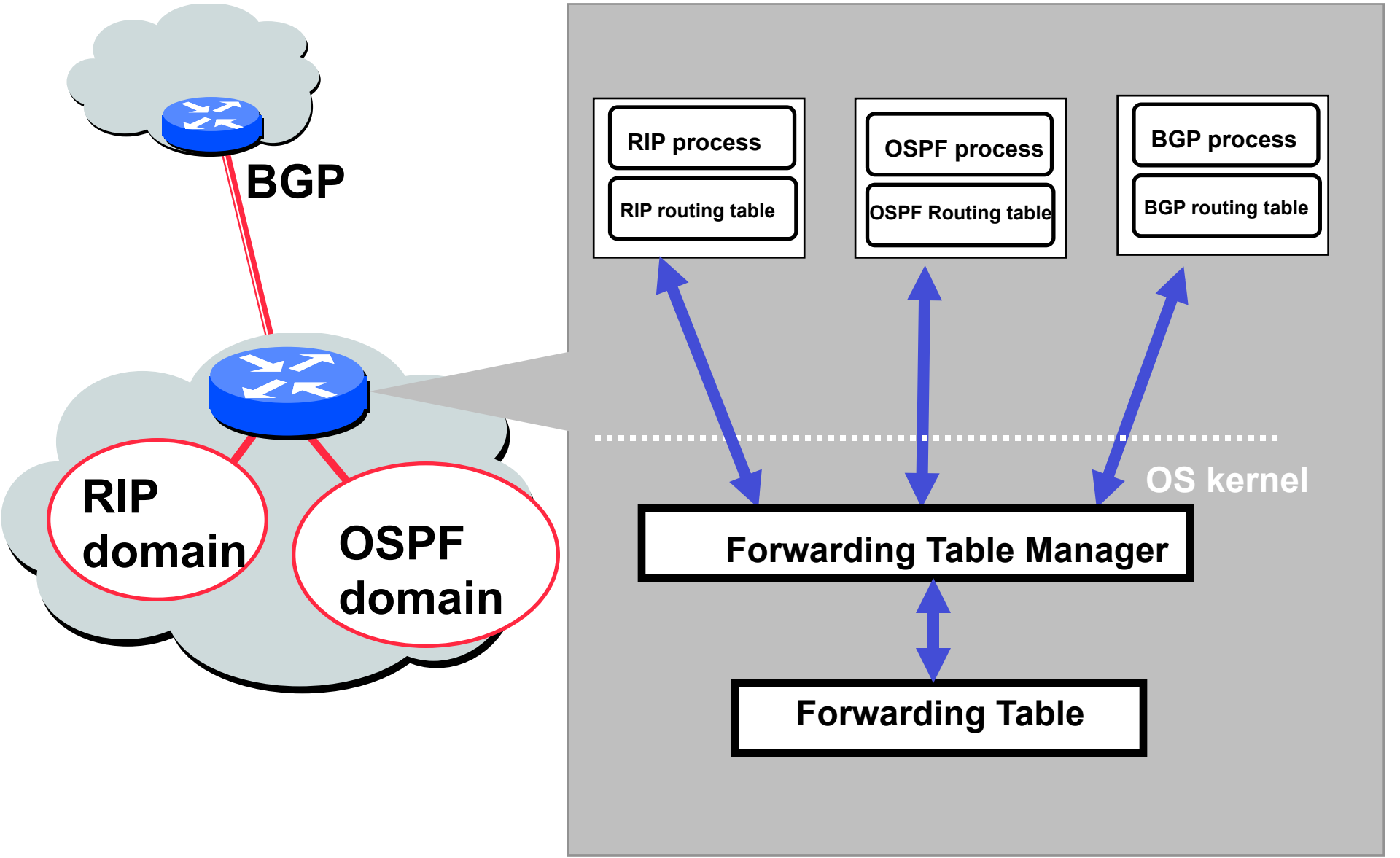




Intra-AS and Inter-AS Routing



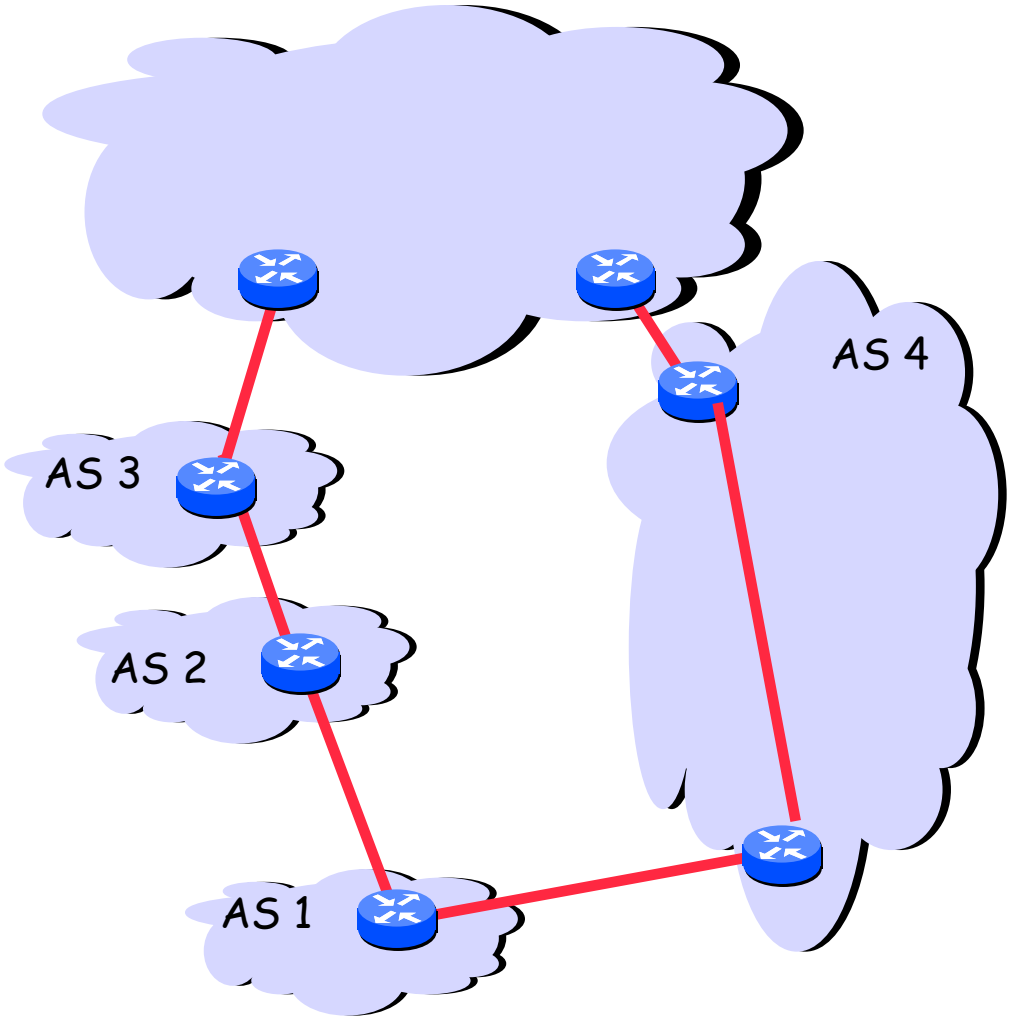
Many Routing Processes on a Single Router



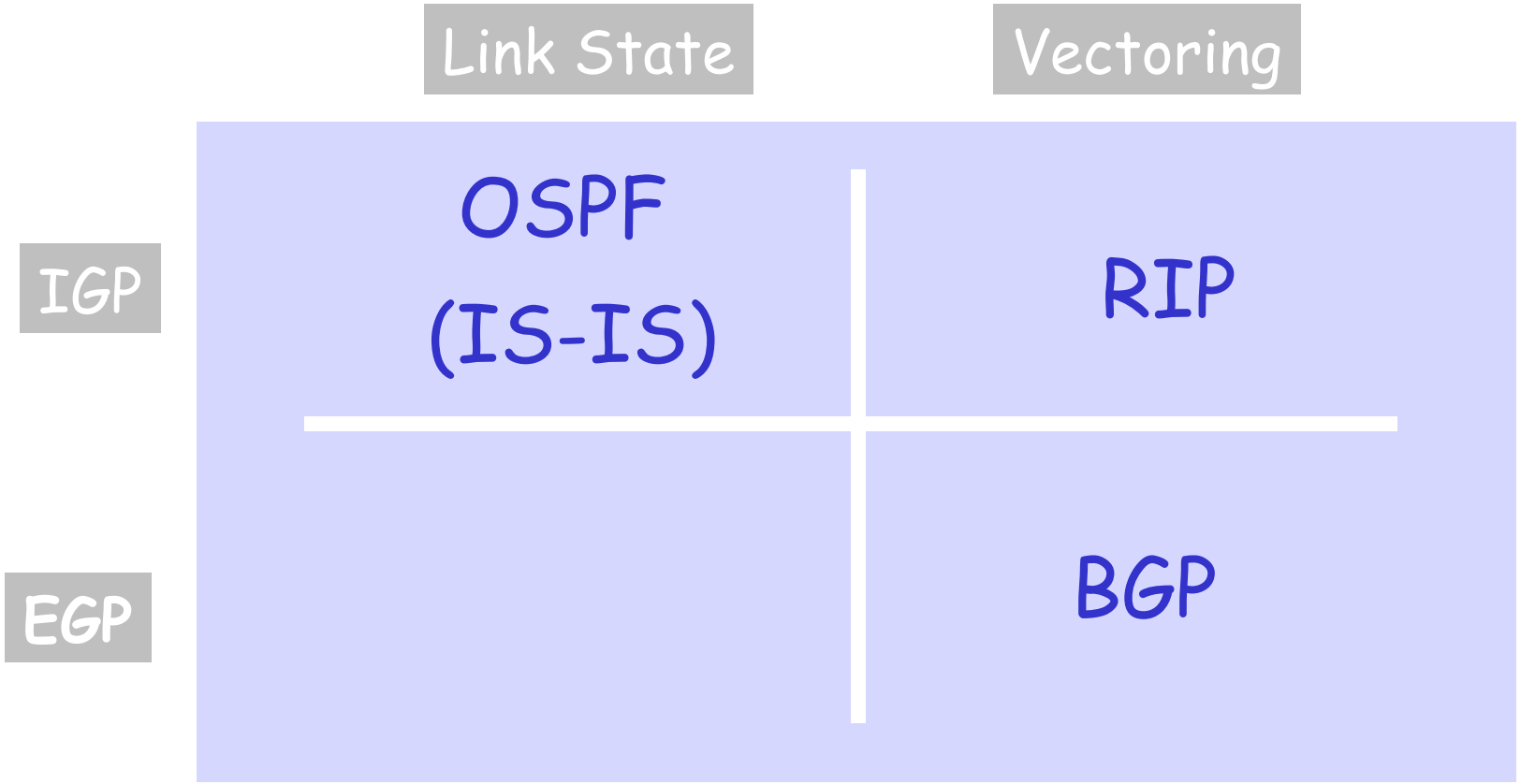
Why Partition into Intra- and Inter-AS Routing?

- This partition allows ASes flexibility to choose their own intra-AS routing protocols
  - Autonomy
- By aggregating many destinations inside an AS into a single destination in inter-domain routing, it improves scalability
  - The partition is a type of *hierarchical* routing
  - Hierarchical routing improves scalability: only a small number of routers are involved with outside

Hierarchical Routing May Pay a Price for Path Quality



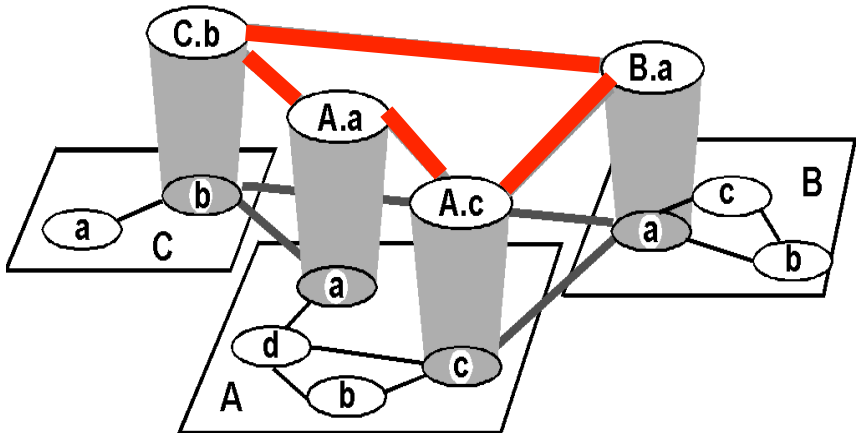
The Gang of Four



Recap: Routing in the Internet

- The Global Internet consists of Autonomous Systems (AS) interconnected with each other
- Routing is divided into intra- and inter-domain routing
  - Intra-AS (intradomain)
    - Different AS's can run different intra-domain routing protocols
  - Inter-AS (interdomain)
    - A single inter-AS protocol: BGP
    - BGP (Border Gateway Protocol) is a Path Vector protocol: a border gateway sends to a neighbor entire path (i.e., a sequence of ASes) to a destination

Routing in the Internet



gateway routers of same AS share learned routes using iBGP.

Gateway routers participate in intradomain to learn internal routes.

AS C (RIP intra routing)

AS A (OSPF intra routing)

AS B (OSPF intra routing)

eBGP

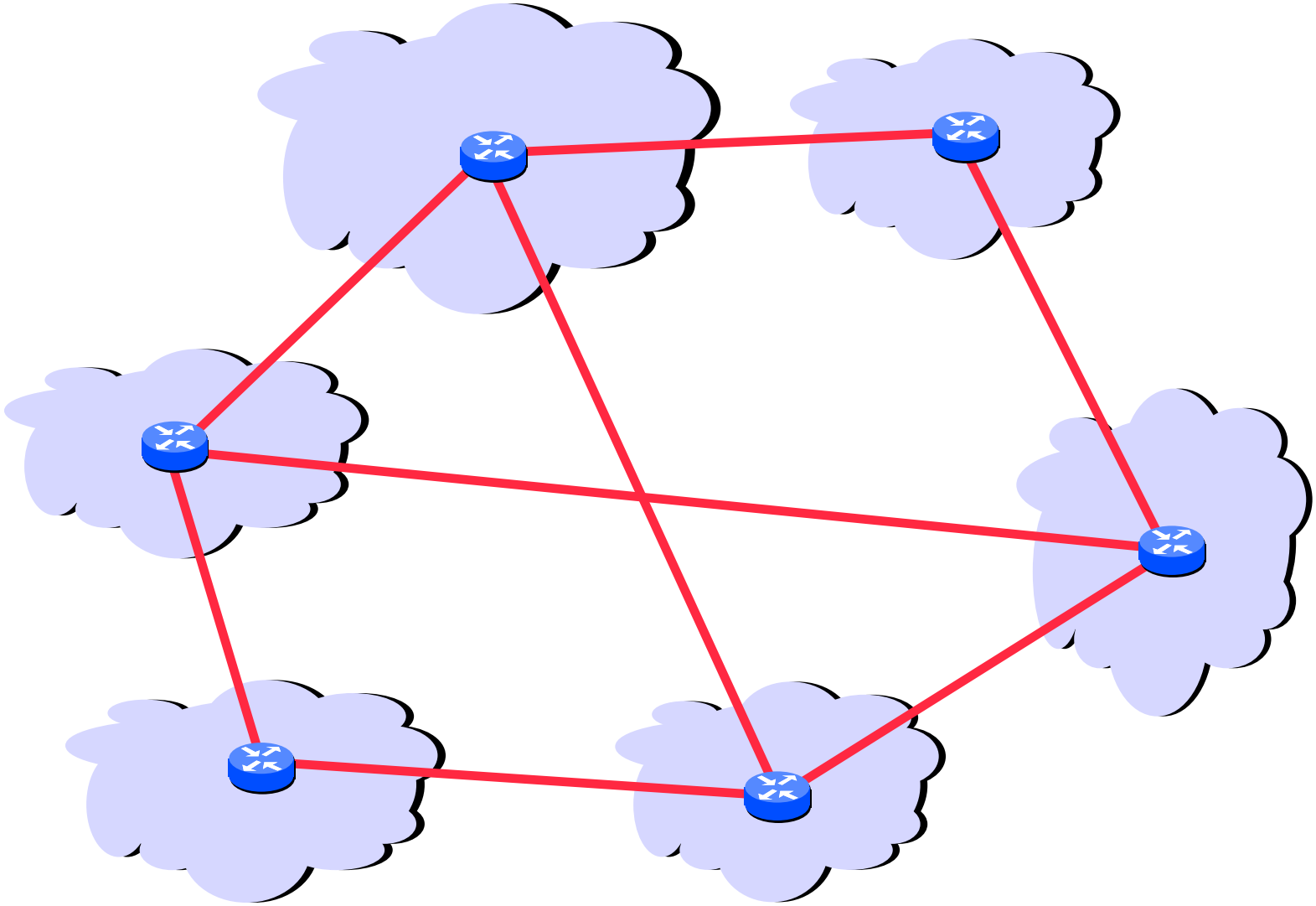
iBGP

Gateway routers of diff. AS's exchange routes using eBGP

Outline

- Recap
- Distance vector protocols
- Link state protocols
- Routing in the Internet
  - BGP

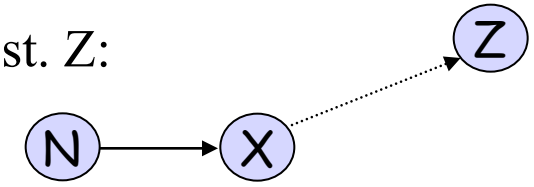
BGP Setup



Internet Interdomain Routing: BGP

- BGP (Border Gateway Protocol): *the* de facto standard
- **Path Vector** protocol:
  - Similar to Distance Vector protocol
  - A border gateway sends to a neighbor *entire path* (i.e., a sequence of ASes) to a destination, e.g.,
    - Gateway X sends to neighbor N its path to dest. Z:

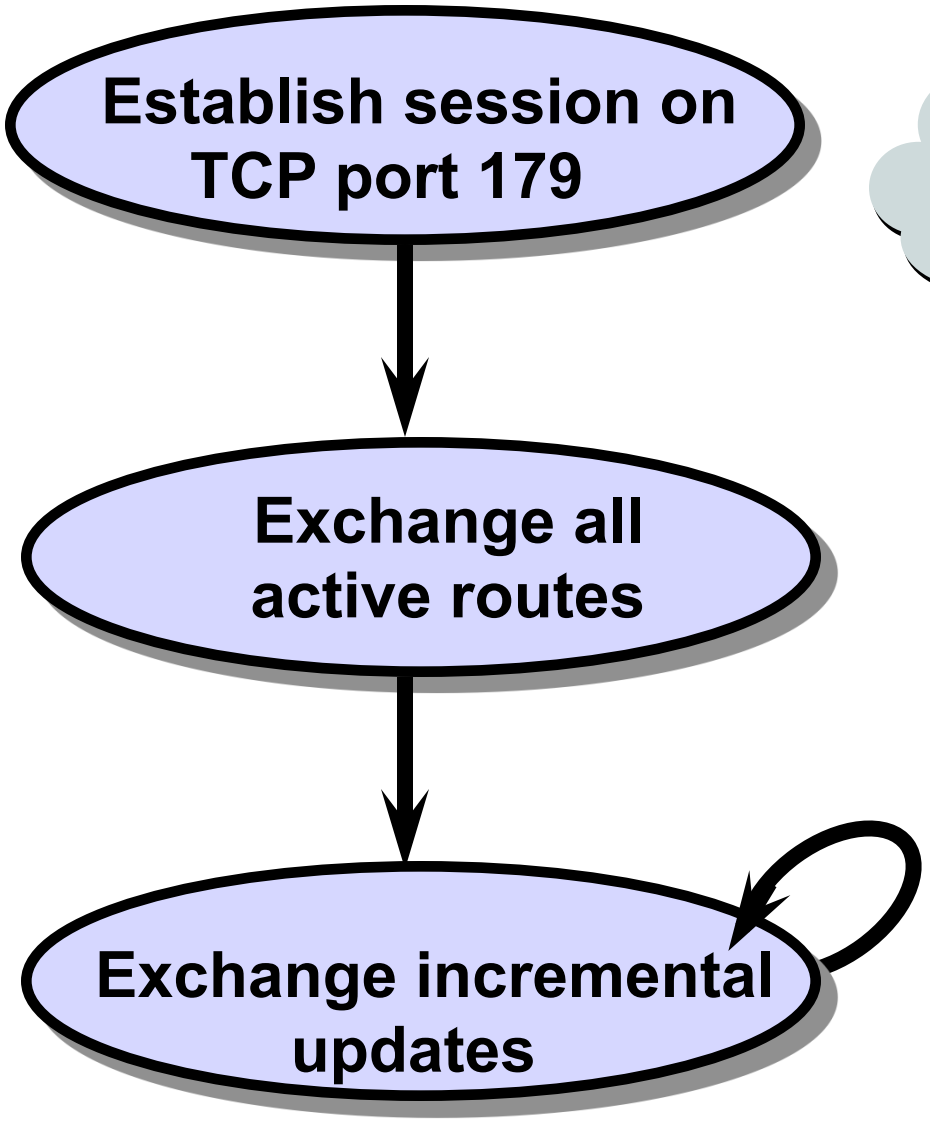
      path (X,Z) = X,Y1,Y2,Y3,…,Z

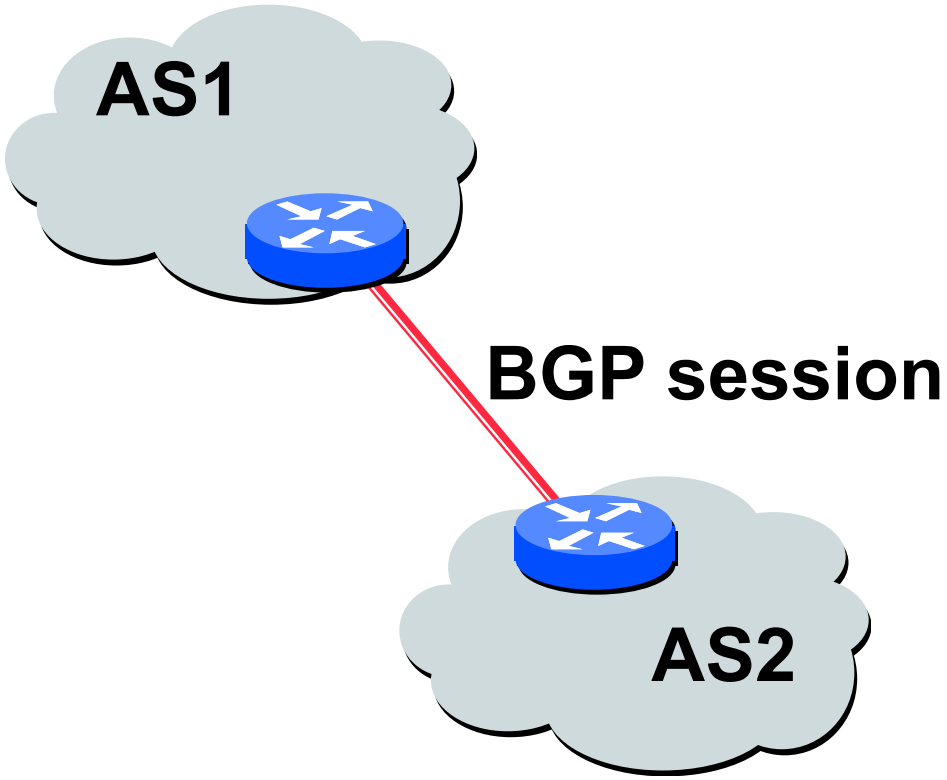


  - If N selects path(X, Z) advertised by X, then:

    path (N,Z) = N, path (X,Z)

BGP Operations (Simplified)





while (connection is ALIVE)
exchange UPDATE message
select best available route
if route changes, export to neighbo

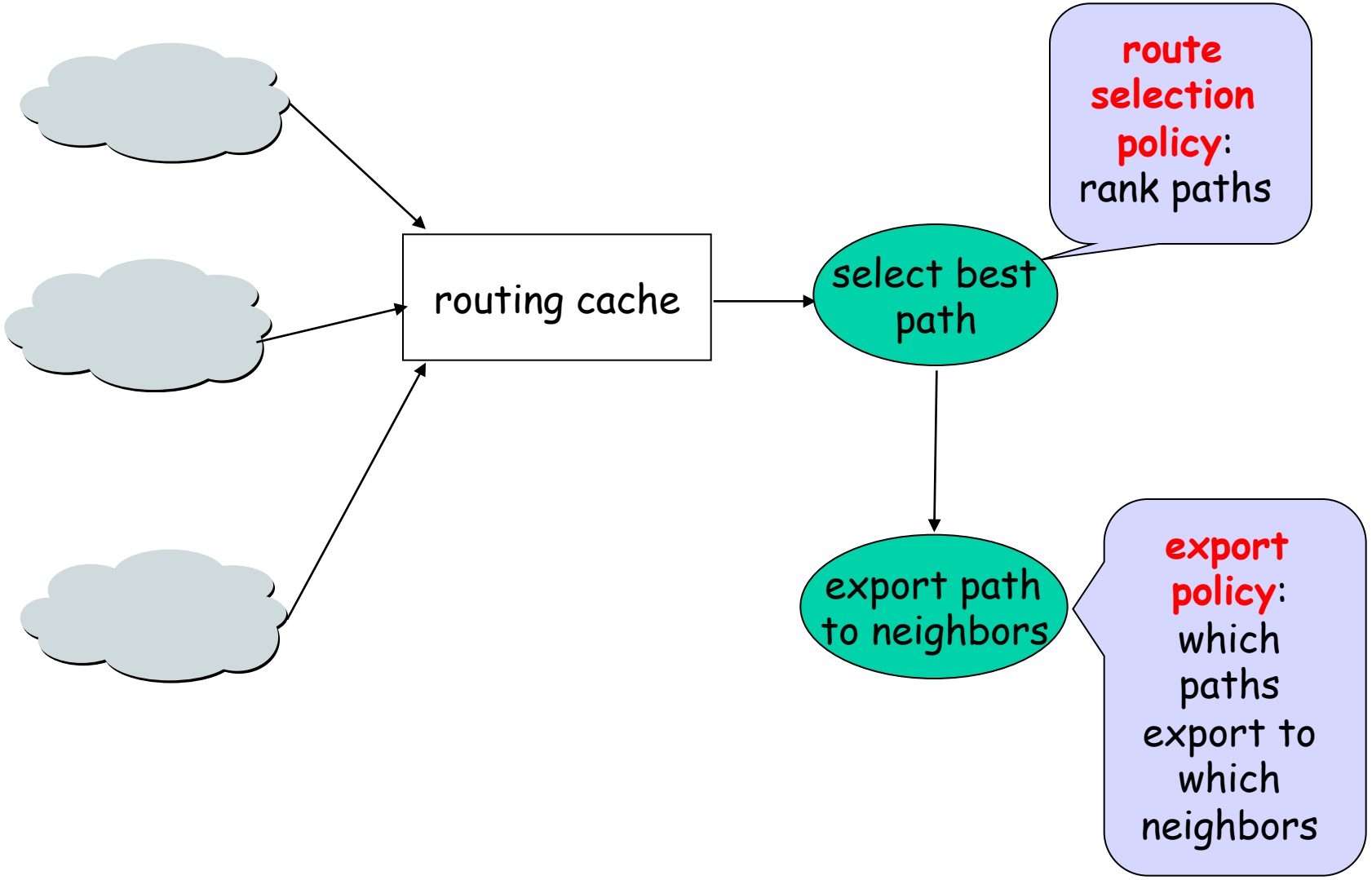
BGP Messages

- Four types of messages
  - OPEN: opens TCP connection to peer and authenticates sender
  - UPDATE: advertises new path (or withdraws old)
  - KEEPALIVE keeps connection alive in absence of UPDATES; also ACKs OPEN request
  - NOTIFICATION: reports errors in previous msg; also used to close connection

Why Path Vector?

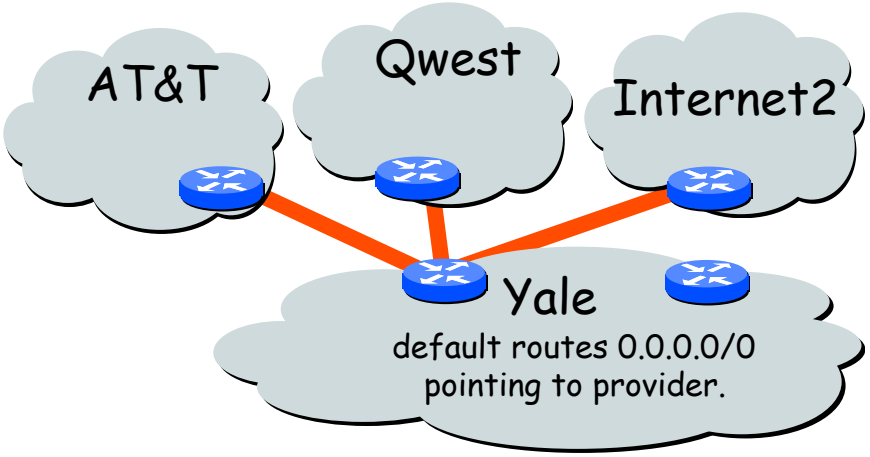
- Path vector prevents counting-to-infinity problem
- Path vector allows an AS to define local policies on the ASes of a given path

BGP Routing Decision Process



BGP Route Selection Policy

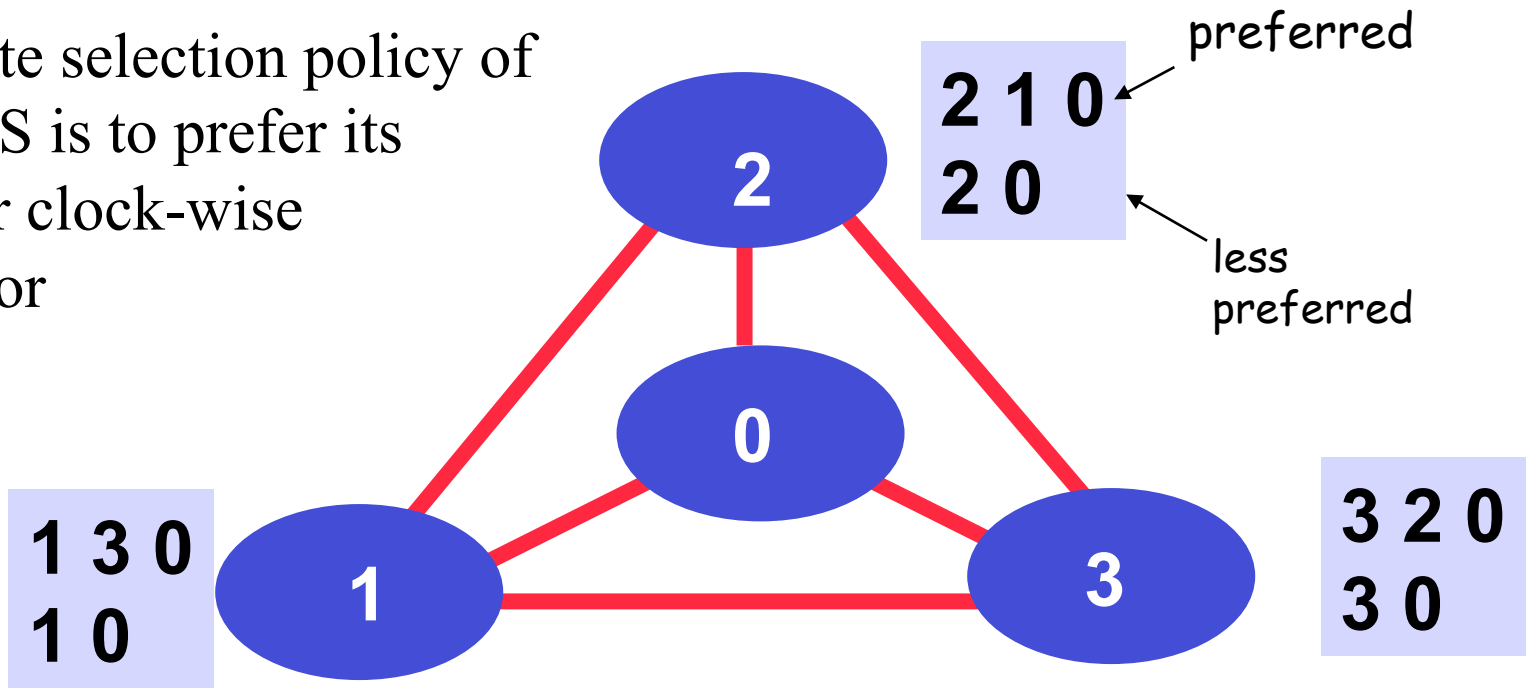
- Typical (Cisco) route selection policy
  - Highest local pref
  - Shortest AS path length
  - Prefer eBGP over iBGP
  - …



Policy Interactions

The BAD GADGET example:

- 0 is the destination
- the route selection policy of each AS is to prefer its counter clock-wise neighbor

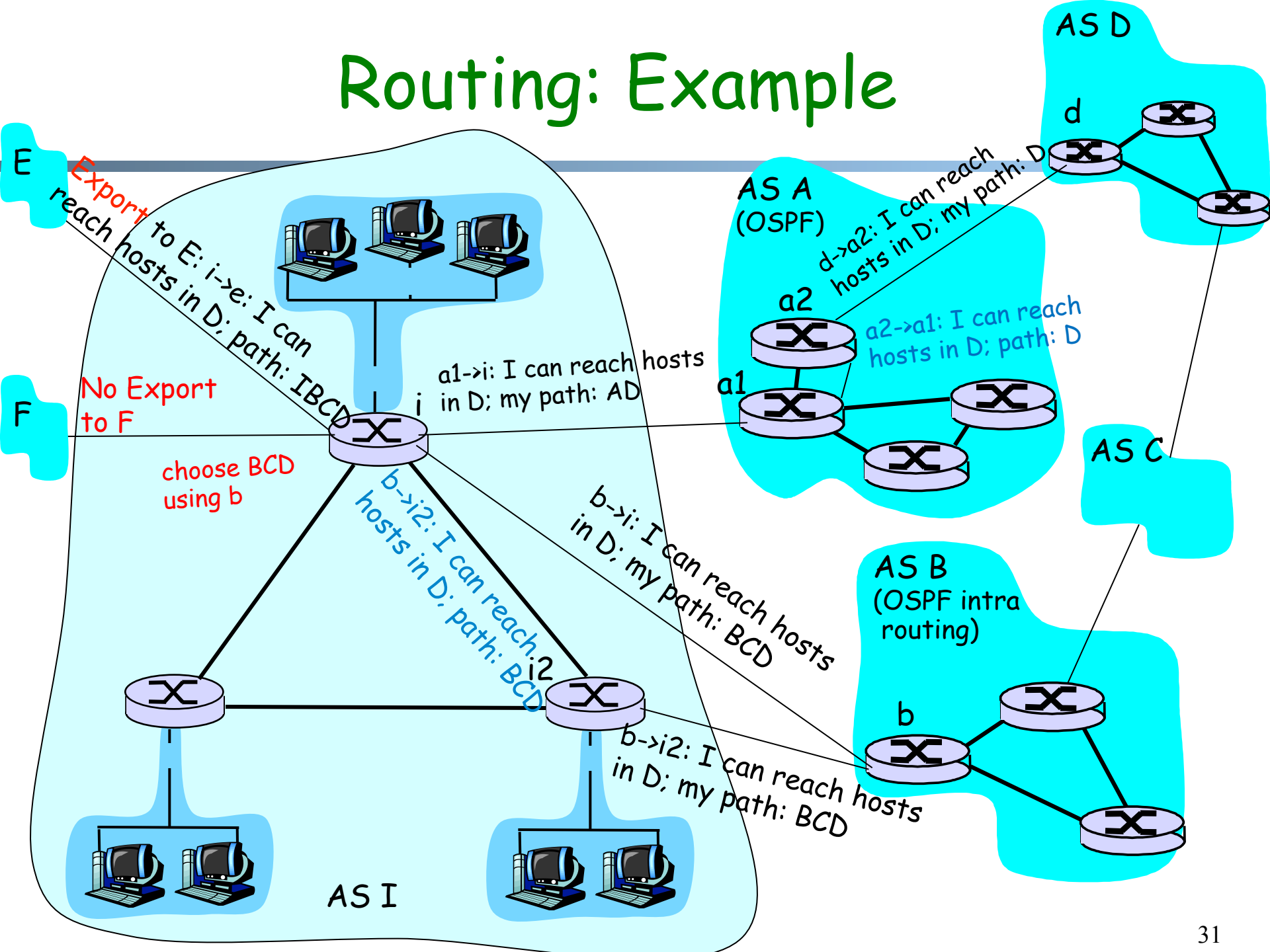


Policy interaction causes routing instability !

BGP Route Export Policies

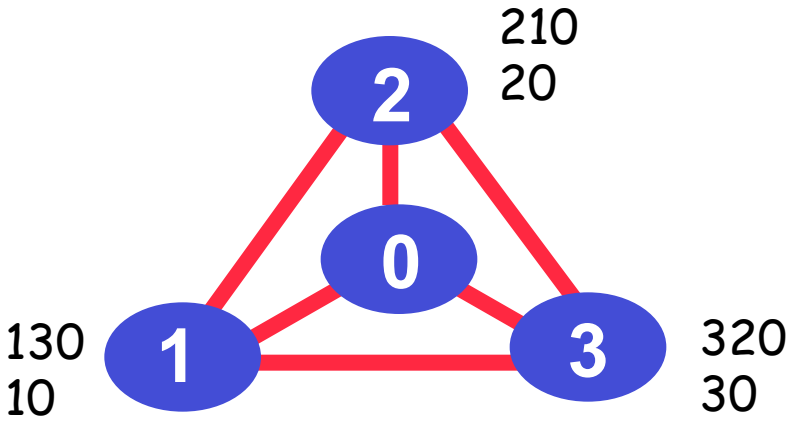
- The export of a path to a neighbor is an indication that the AS is willing to transport traffic for the neighbor
- An AS may not export some routes to some neighbors (more later)

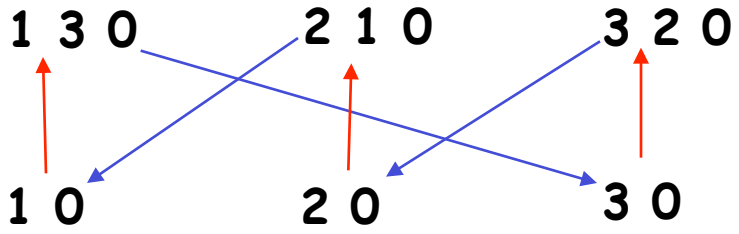
Routing: Example



Understanding Instability: P-Graph

- Nodes in P-graph are feasible paths
- A directed edge from path N_1P_1 to P_1
  - Intuition: to let N_1 choose N_1P_1, P_1 must be chosen and exported to N_1
- A directed edge from a lower ranked path to a higher ranked path
  - Intuition: the higher ranked path should be considered first

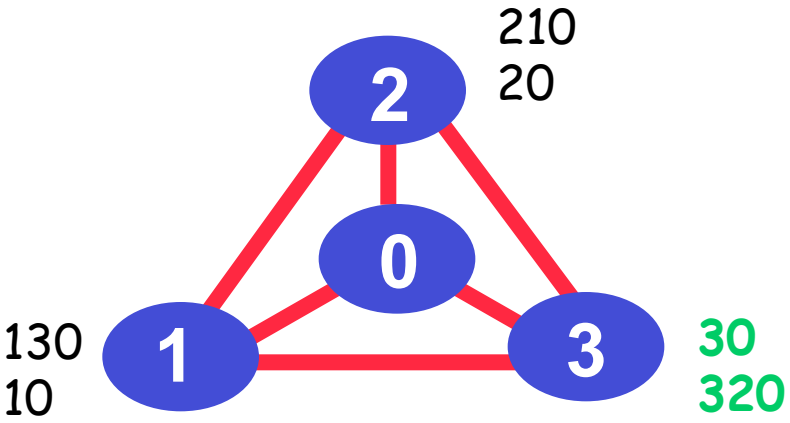


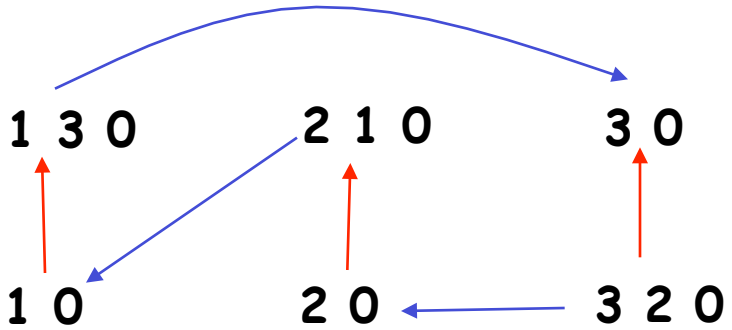


P-graph

Partial Order Graph and Convergence

- If the P-graph has no loop, then BGP policy converges.
  - Intuition: choose the node (i.e., a path) from the partial order graph with no out-going edge, choose the path, remove the path and all other lowered ranked paths of the same node; remove nodes if suffix removed; continue
- Example: suppose we swap the order of 30 and 320



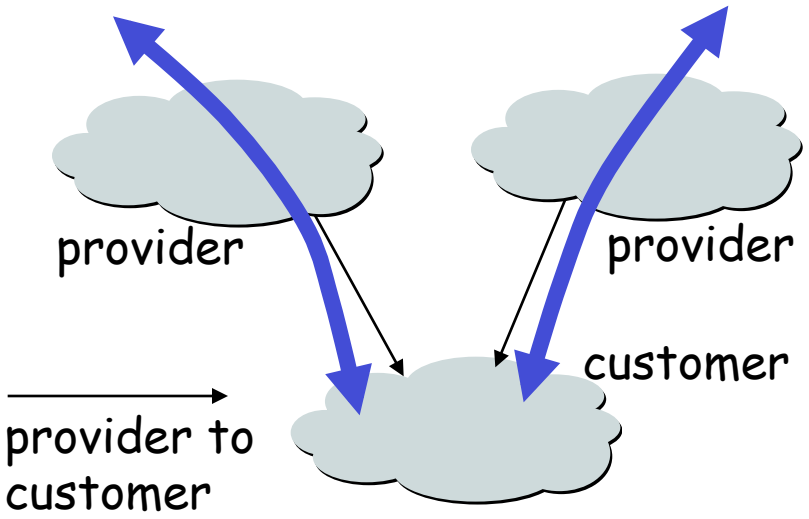


Partial Order Graph and BGP Convergence

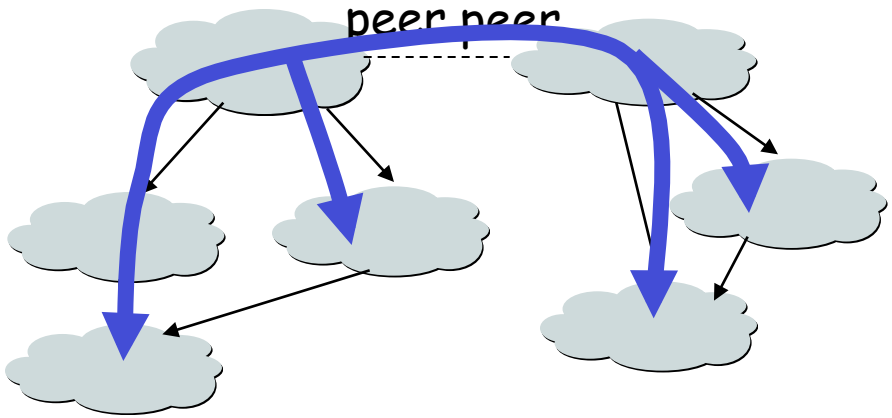
- Preview: A reason we do not often see instability in the Internet is that:
  - The current Internet ISP economy implies no loop in P-graph !

Internet Economy: Two Types of Business Relationship

- *Customer provider relationship*
  - A provider is an AS that connects the customer to the rest of the Internet
  - Customer pays the provider for the transit service
  - E.g., Yale is a customer of AT&T and QWEST



- *Peer-to-peer relationship*
  - Mutually agree to exchange traffic between their respective **customers**
  - There is no payment between peers

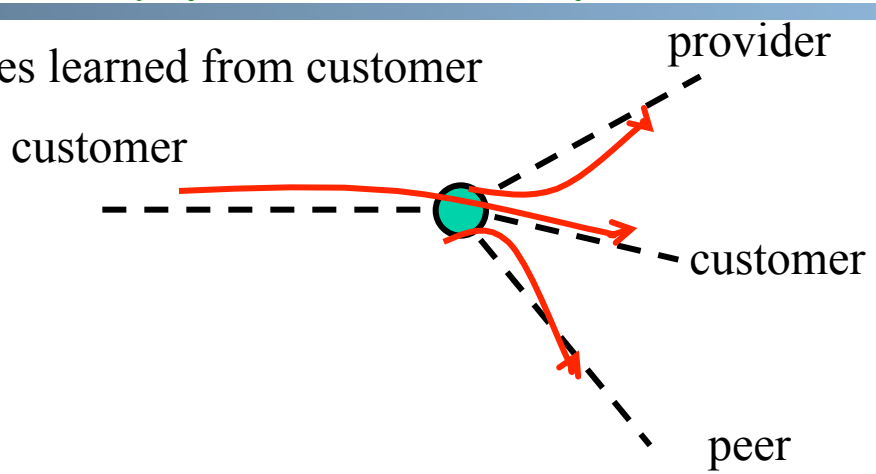


Implication of Business Relationship on Policies

- Route selection (ranking) policy:
  - The typical route selection policy is to prefer customers over peers/providers to reach a destination, i.e., Customer > pEer/Provider

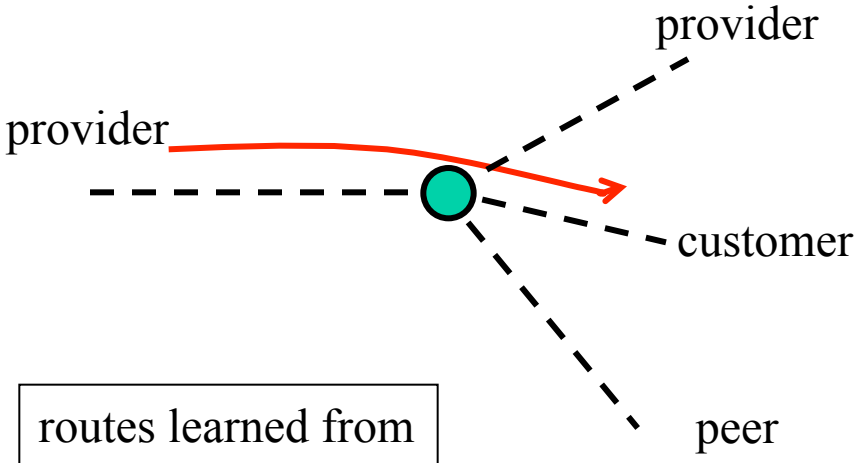
Typical Export Policies

case 1: routes learned from customer



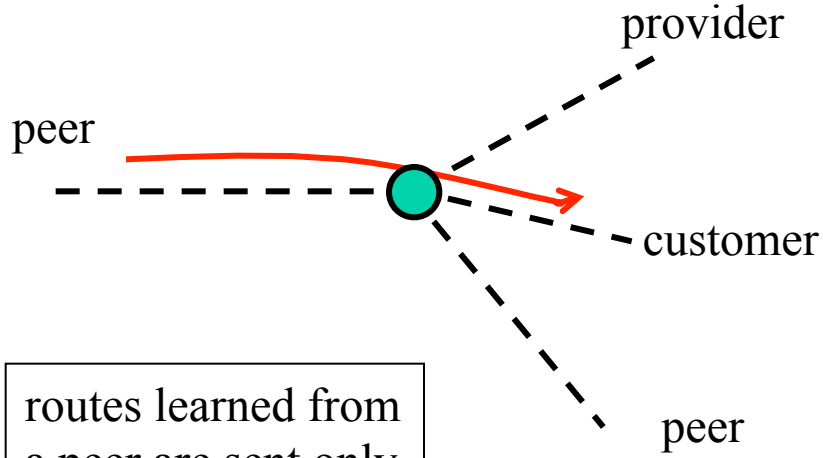
routes learned from a customer are sent to all other neighbors

case 2: routes learned from provider



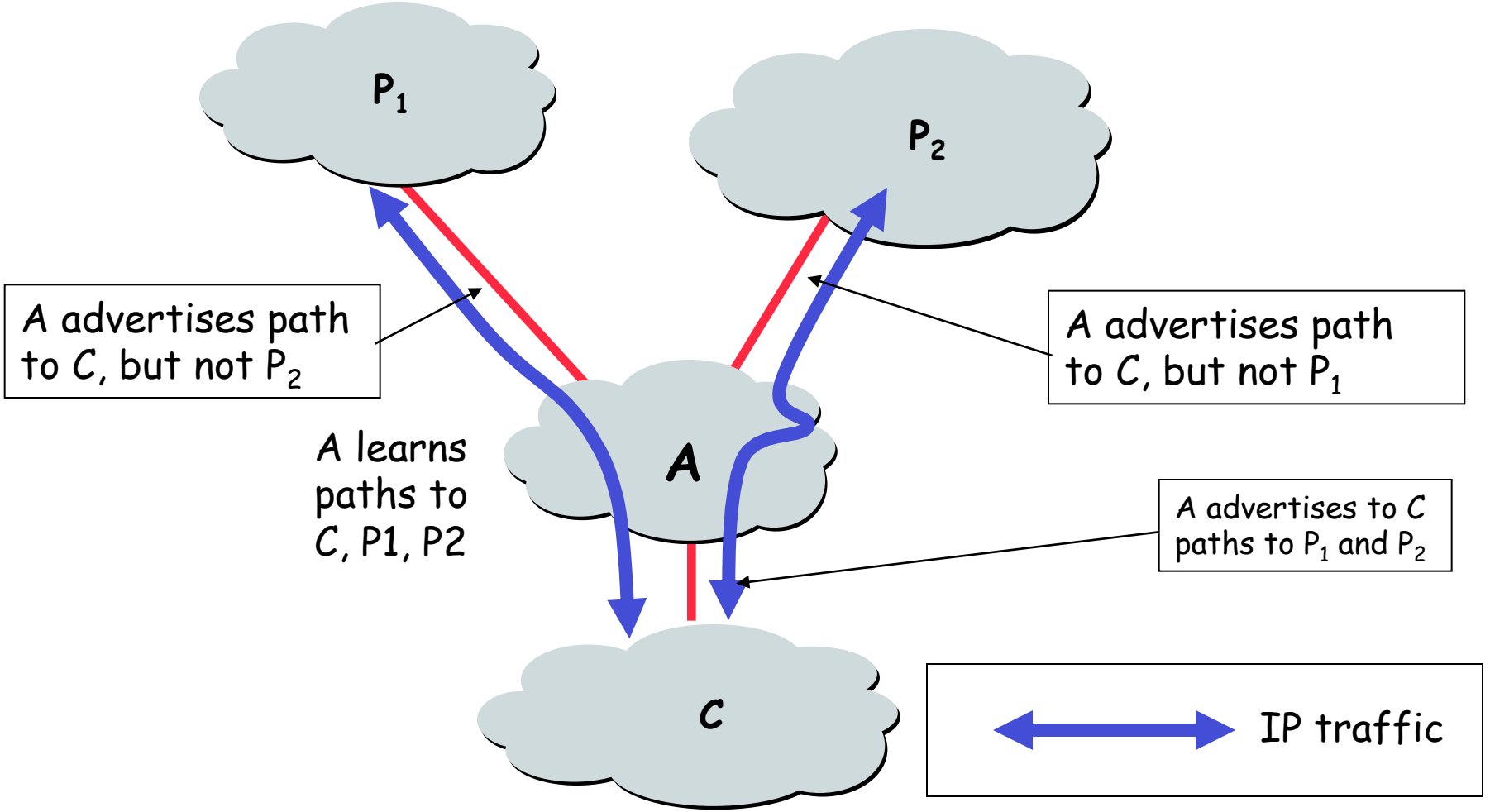
routes learned from a provider are sent only to customers

case 3: routes learned from peer



routes learned from a peer are sent only to customers

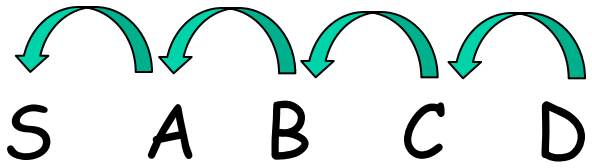
Typical Export -> No-Valley Routing



Suppose P_1 and P_2 are providers of *A*; *A* is a provider of *C*

Typical Export Policies Imply Patterns of Routes

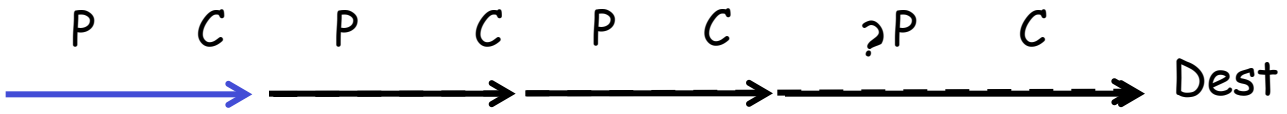
- Assume a BGP path SABCD to destination AS D. Consider the business relationship between each pair:



- Three types of business relationships:
  - PC (provider-customer)
  - CP (customer-provider)
  - PP (peer-peer)

Typical Export Policies Imply Patterns of Routes

- Two invariants of valid BGP routes (with labels representing business relationship)



Reason: only route learned from customer is sent to provider; thus after a PC, it is always PC to the destination

Reason: routes learned from peer or provider are sent to only customers; thus all relationship before is CP

Stability of BGP Routing

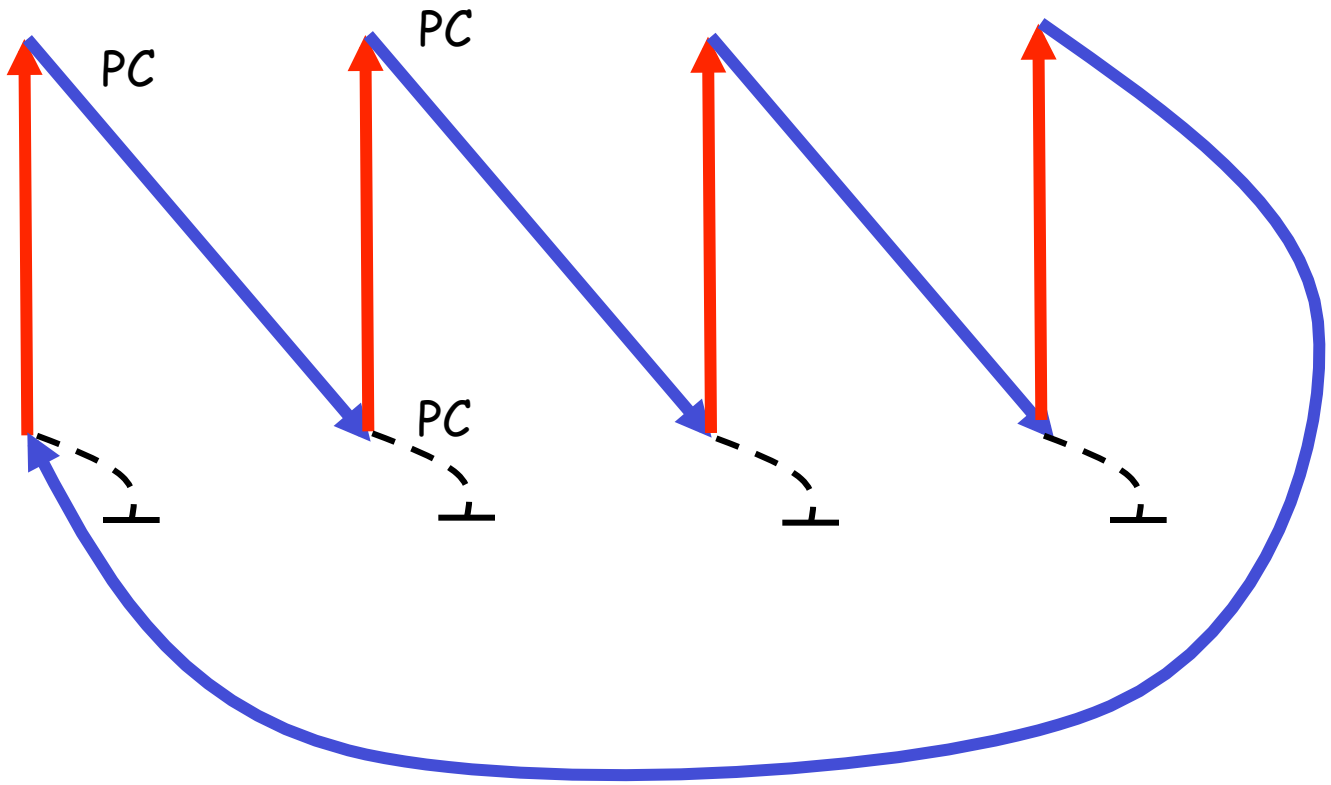
- Suppose
  1. There is no loop formed by provider-customer relationship in the Internet
  2. Each AS uses typical route selection policy:

     C > E/P
  3. Each AS uses the typical export policies

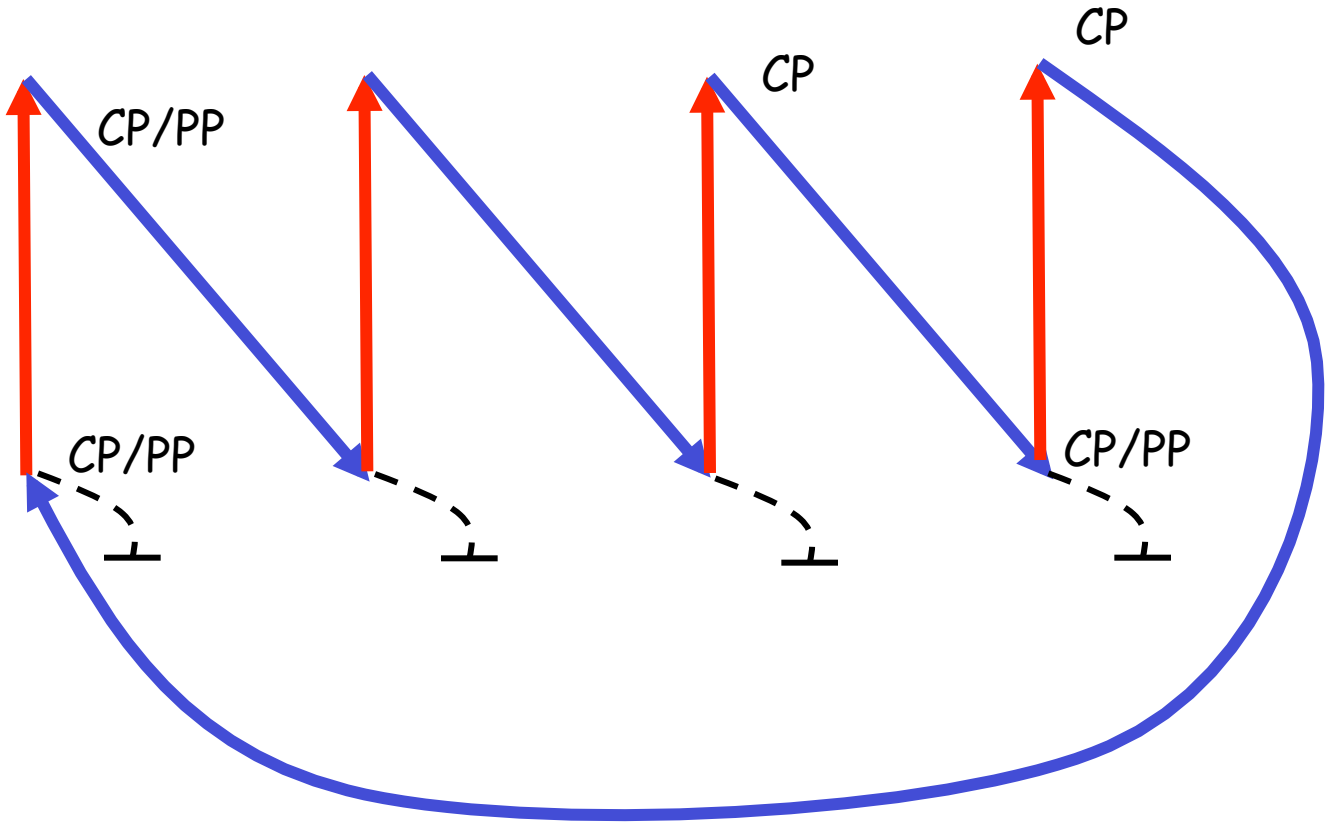
- Then BGP policy routing always converges !

Case 1: A Link is PC

Proof by contradiction. Assume a loop in P-graph. Consider a fixed link. in the loop



Case 2: Link is CP/PP



Routing: Example