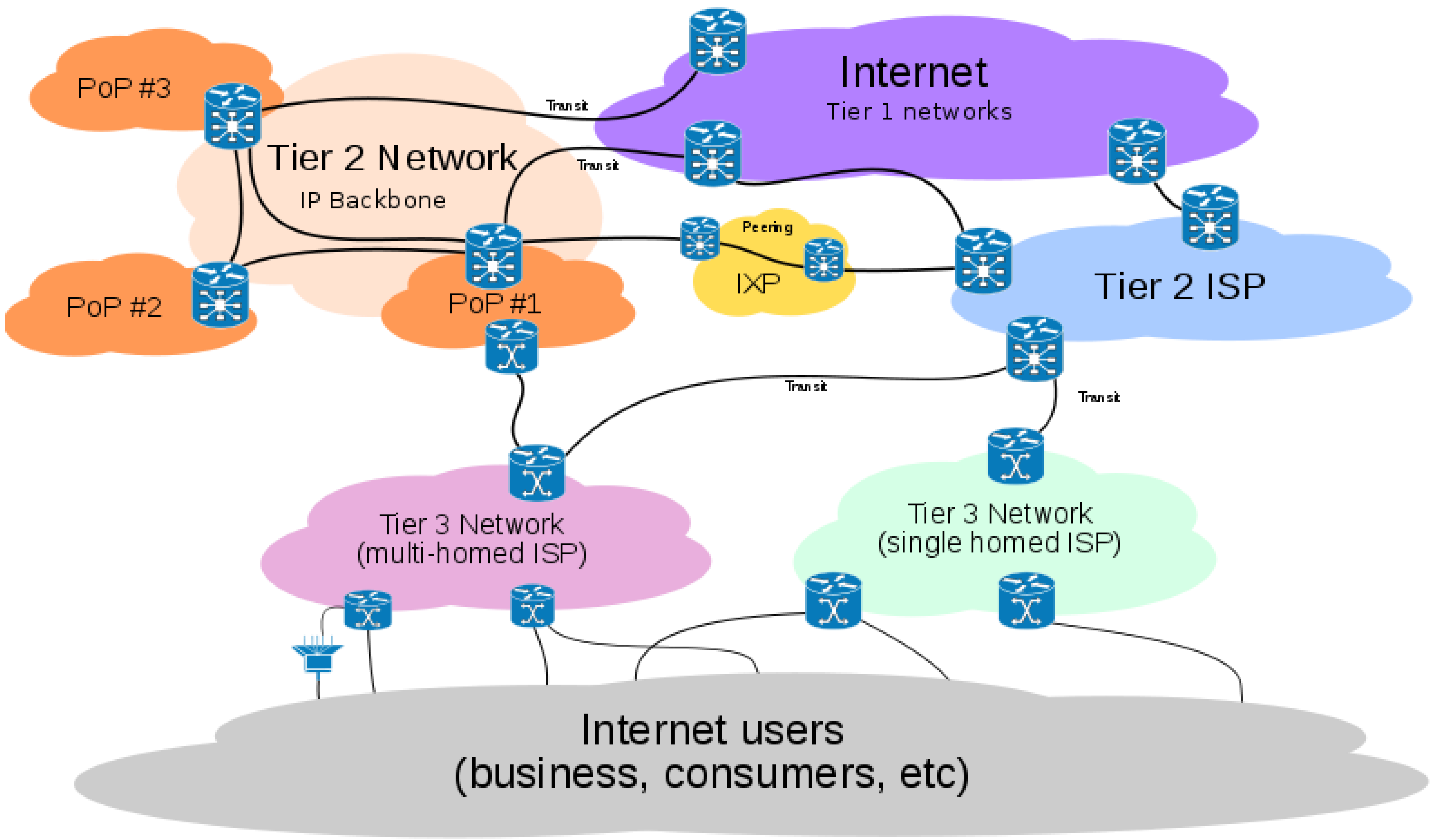
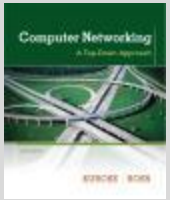
Inter-AS routing



Computer Networking: A Top Down Approach
6th edition
Jim Kurose, Keith Ross
Addison-Wesley



Chapter 4: outline

4.1 Introduction

4.2 Virtual circuit and datagram networks

4.3 What's inside a router

4.4 IP: Internet Protocol

- Datagram format
- IPv4 addressing
- Network Address Translation (NAT)
- DHCP
- ICMP
- IPv6
- IPsec

4.5 Routing algorithms

- Link state
- Distance vector
- Hierarchical routing

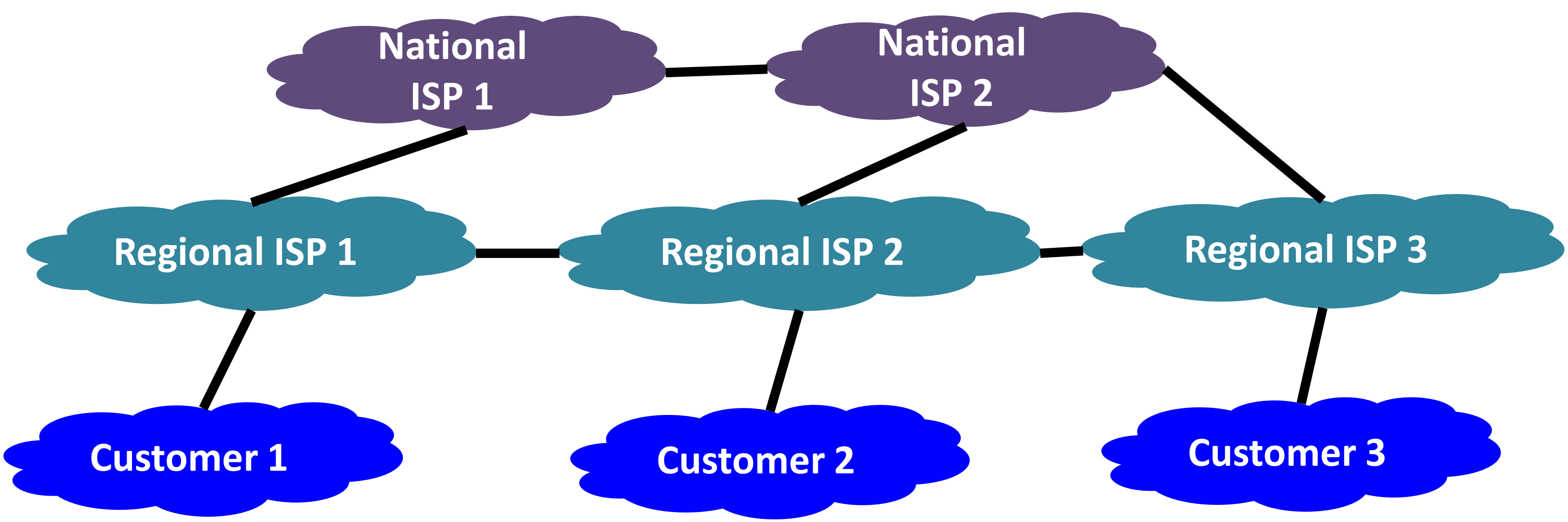
4.6 Routing in the Internet

- RIP
- OSPF
- BGP

4.7 Broadcast and multicast routing

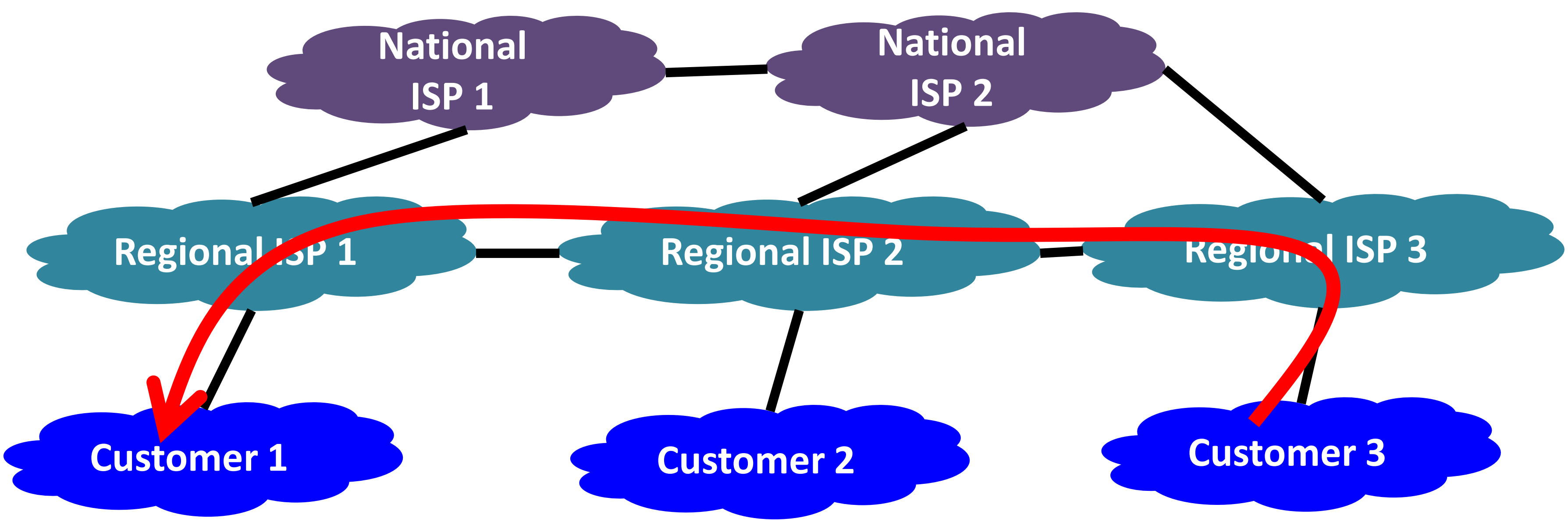
Shortest path routing

- Problems with always taking shortest path:
  - All traffic must travel on shortest path
  - All nodes must do same link cost calculation
  - Not possible to enforce various business rules



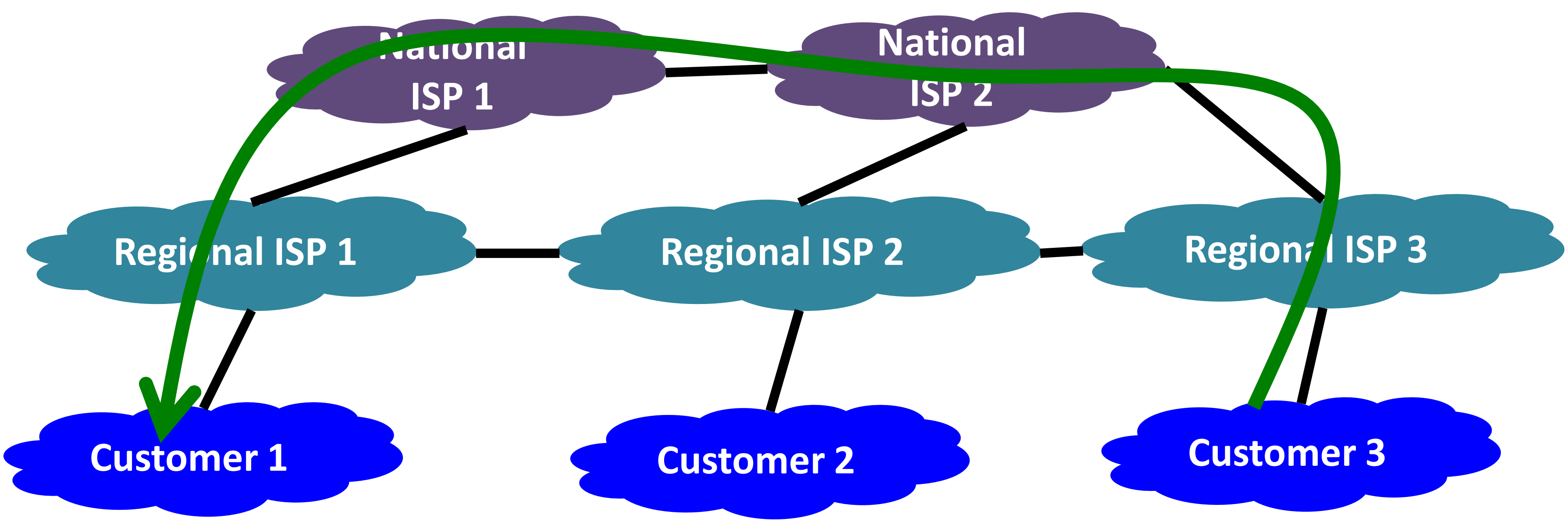
Shortest path routing

- Example: customer 3 talking to customer 1
  - Shortest path transits Regional ISP 2
  - Regional ISP 2 isn't being paid by either customer



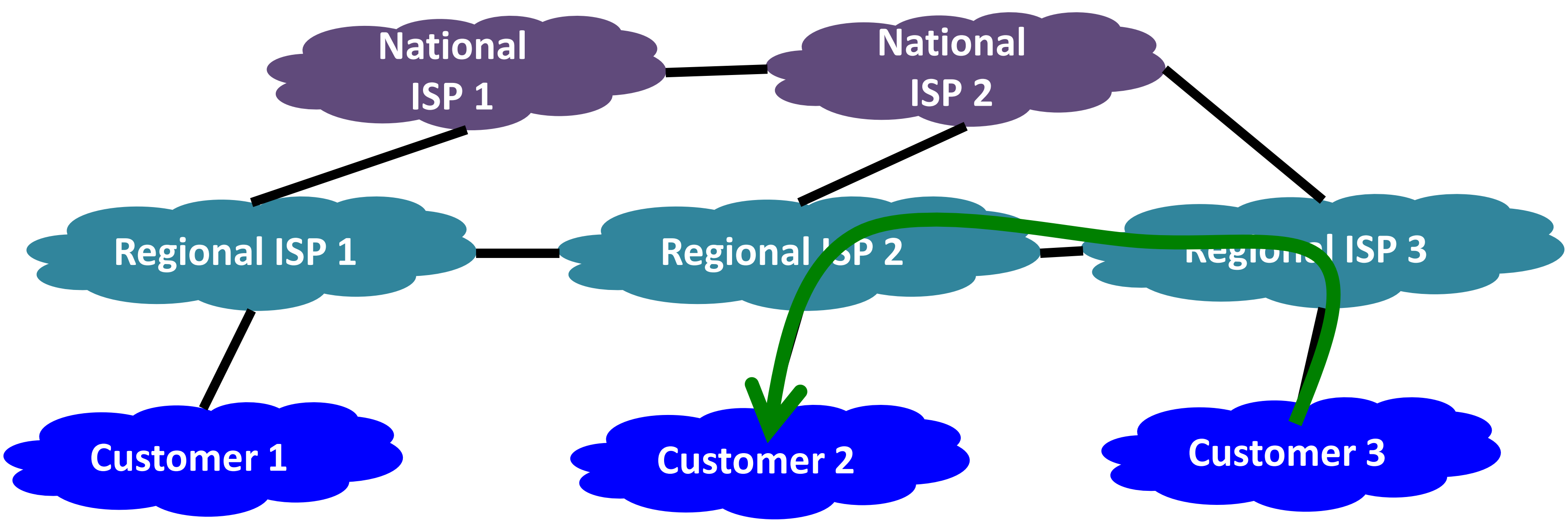
Shortest path routing

- Example: customer 3 talking to customer 1
  - Goes through National ISP 1 & 2
  - Regional 3 is paying National ISP 2
  - Regional 1 is paying National ISP 1



Shortest path routing

- Example: customer 3 talking to customer 2
  - Regional 2 and 3 are peered
  - Avoid going through National ISP 2 since then both regionals would incur expense



Other routing issues

- Policies may be political, security, or economic:
  - Some examples (Tanenbuam):
    - Don't carry commercial traffic on educational network
    - Never send Pentagon traffic through Iraq
    - Use TeliaSonera instead of Verizon because it is cheaper
    - Don't use AT&T in Australia because performance is poor
    - Traffic starting or ending at Apple should not transit Google

Link-state, disadvantages

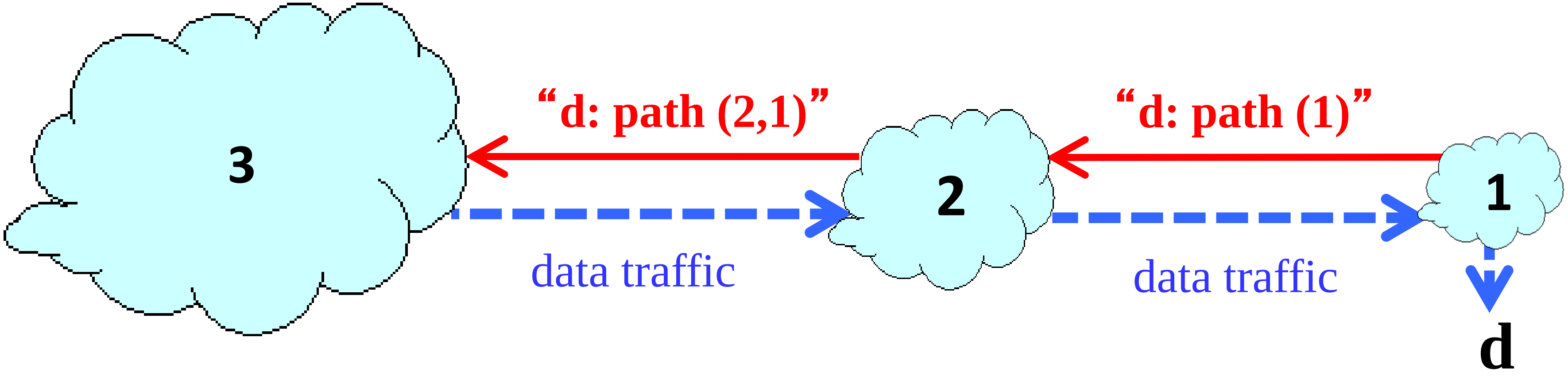
- Floods topology information
  - High bandwidth and storage requirements
  - Nodes divulge potentially sensitive information
- Entire path computed locally
  - High processing overhead for large network
- Distance calculation hides information
  - Everyone has to have a shared notion of link cost
- Typically used within one organization
  - Autonomous System (AS)
    - e.g. university, company, ISP
  - Popular link-state protocols: OSPF, IS-IS

Distance-vector

- Disadvantages:
  - Count to infinity, "bad news travels slow"
  - Slow to converge
  - Hides information that you might need in an inter-AS setting
- Advantages:
  - Summarizes details of network topology
    - Trades optimality for scalability
  - Each node only needs to know about next hop

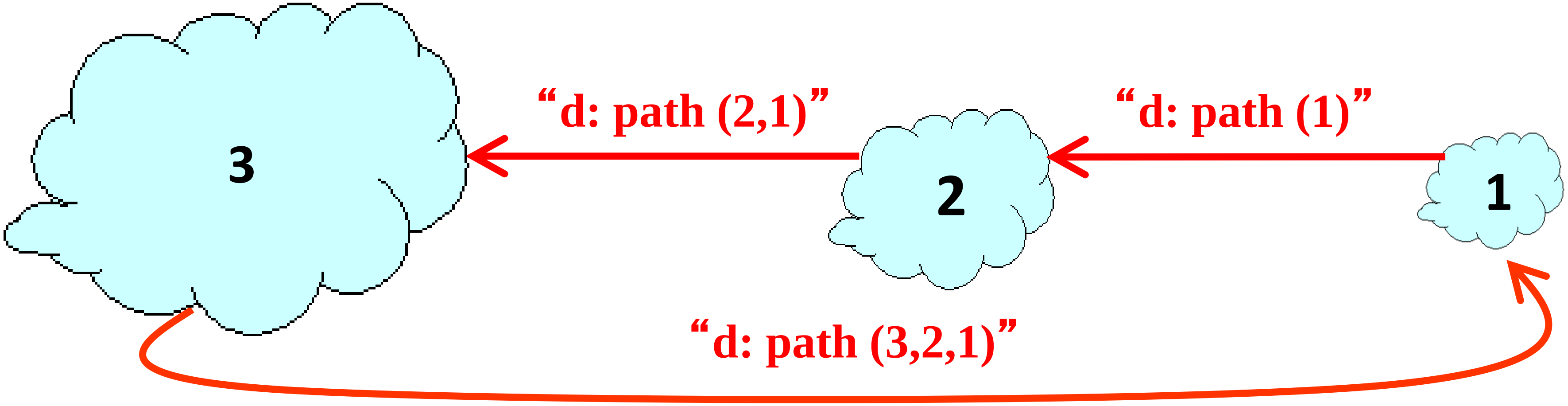
Path-vector routing

- Extension of distance-vector
  - Support flexible routing policies
  - Avoid count-to-infinity problem
- Key idea: advertise the entire path
  - Distance vector: send *distance metric* per destination d
  - Path vector: send the *entire path* per destination d



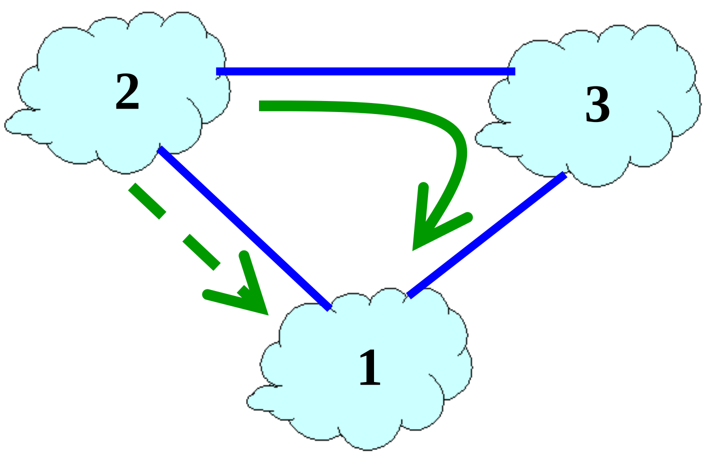
Detecting loops

- Path-vector can easily detect loop
  – Look for your own node ID in the path
  – e.g. node 1 sees itself in path "3, 2, 1"
- Node can discard paths with loops
  – e.g. node 1 drops advertisement

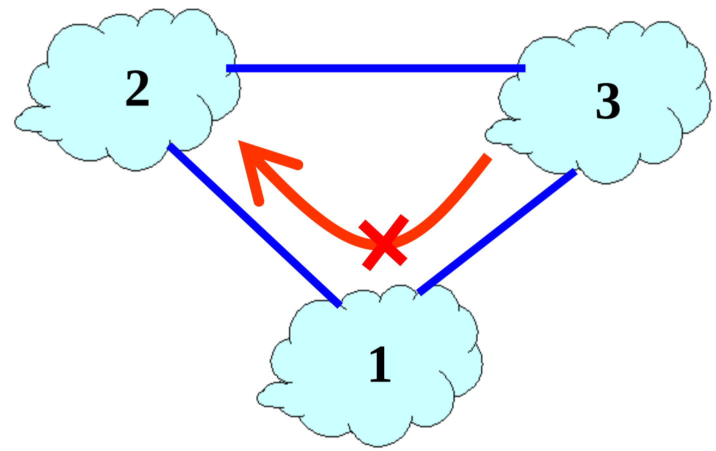


Flexible routing policies

- Each node can apply local policies:
  - Path selection: Which path to use?
  - Path export: Which path to advertise?



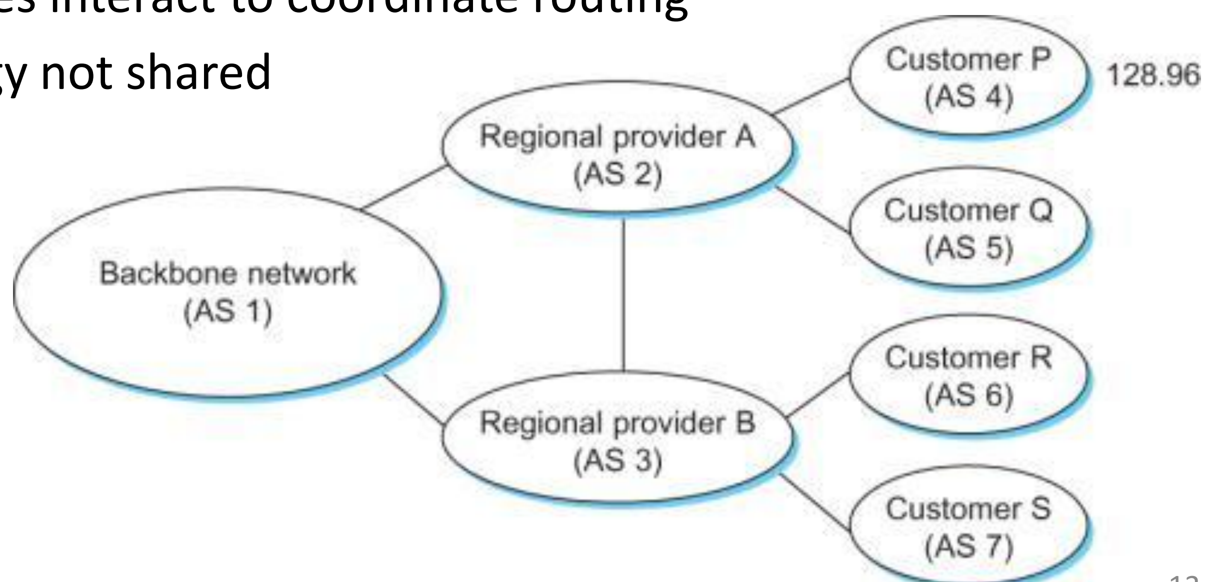
Node 2 may prefer the path "2, 3, 1" over the path "2, 1". Perhaps it is cheaper.



Node 1 may not export the path "1, 2". Perhaps node 1 reserves the 1->2 link for special traffic.

Scaling up and up

- How to scale to the global Internet?
  - Add another level of hierarchy!
  - Routing amongst Autonomous Systems (ASes)
    - Distinct regions of admin control
    - Routers/links managed by a single institution
    - ASes can use policy-based routing
  - Interaction between ASes
    - Neighboring ASes interact to coordinate routing
    - Internal topology not shared

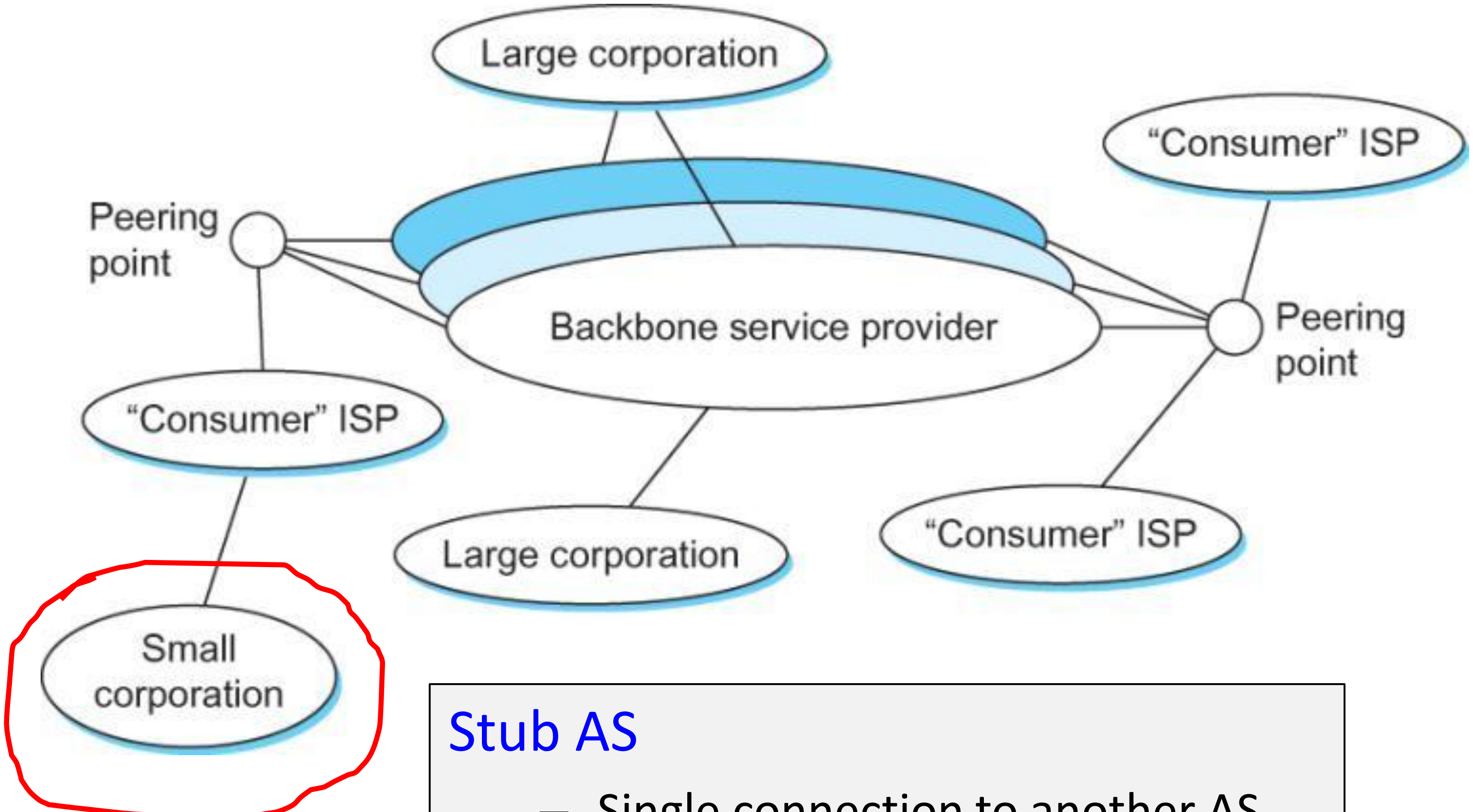


Autonomous System Numbers

- Each AS assigned a unique number
  - Before 2007: AS Numbers 16-bit
  - After 2007: IANA began allocating 32-bit AS numbers
  - Currently over 50,000 allocated

- Level 3: 1
- MIT: 3
- Harvard: 11
- Yale: 29
- Princeton: 88
- AT&T: 7018, 6341, 5074, …
- UUNET: 701, 702, 284, 12199, …
- Sprint: 1239, 1240, 6211, 6242, …
- …

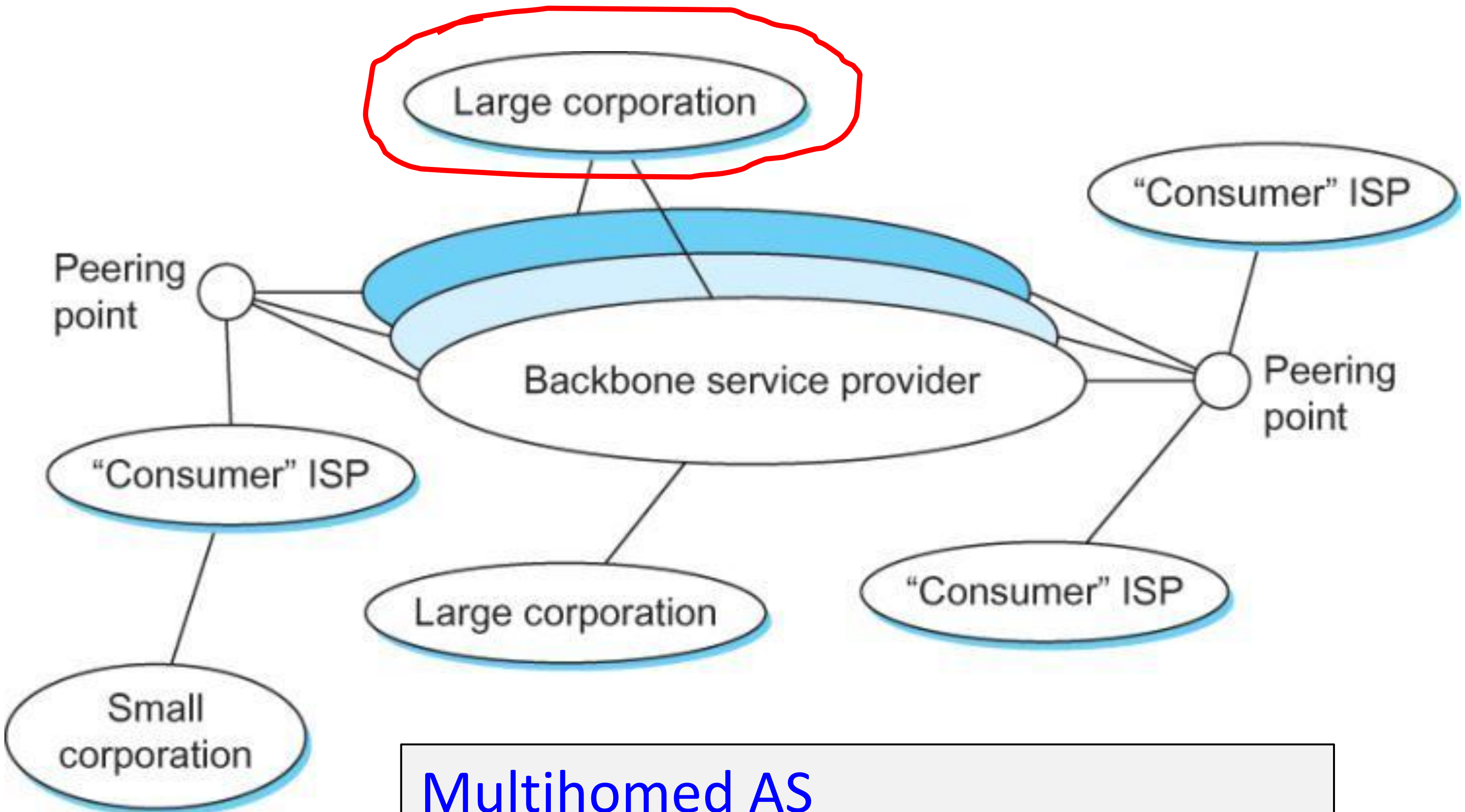
AS stub



Stub AS

- Single connection to another AS
- AS only carries local traffic
- e.g. Small corporation, university

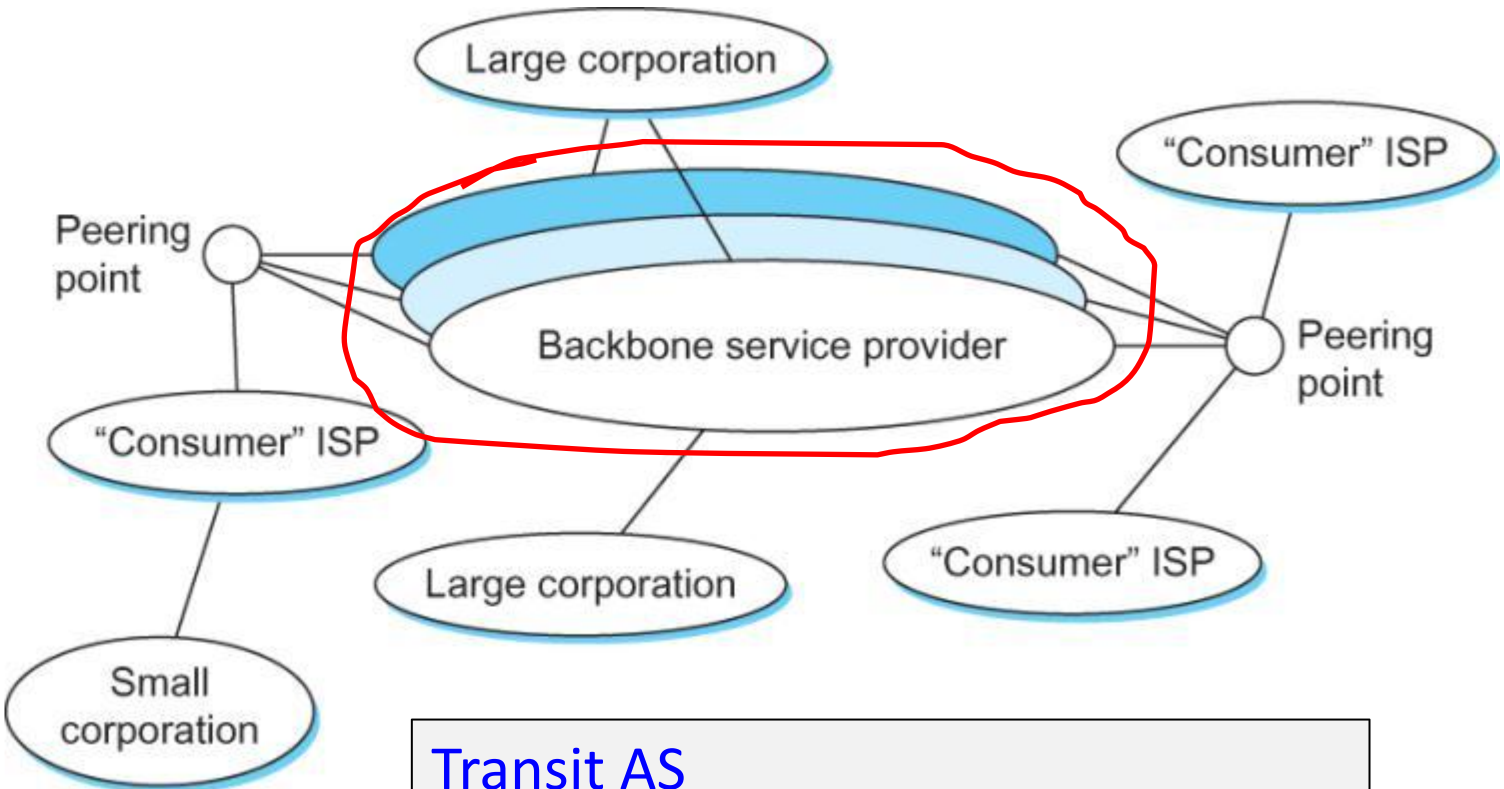
AS multihomed



Multihomed AS

- Connected to multiple ASes
- Refuses to carry transit traffic
- Improves reliability

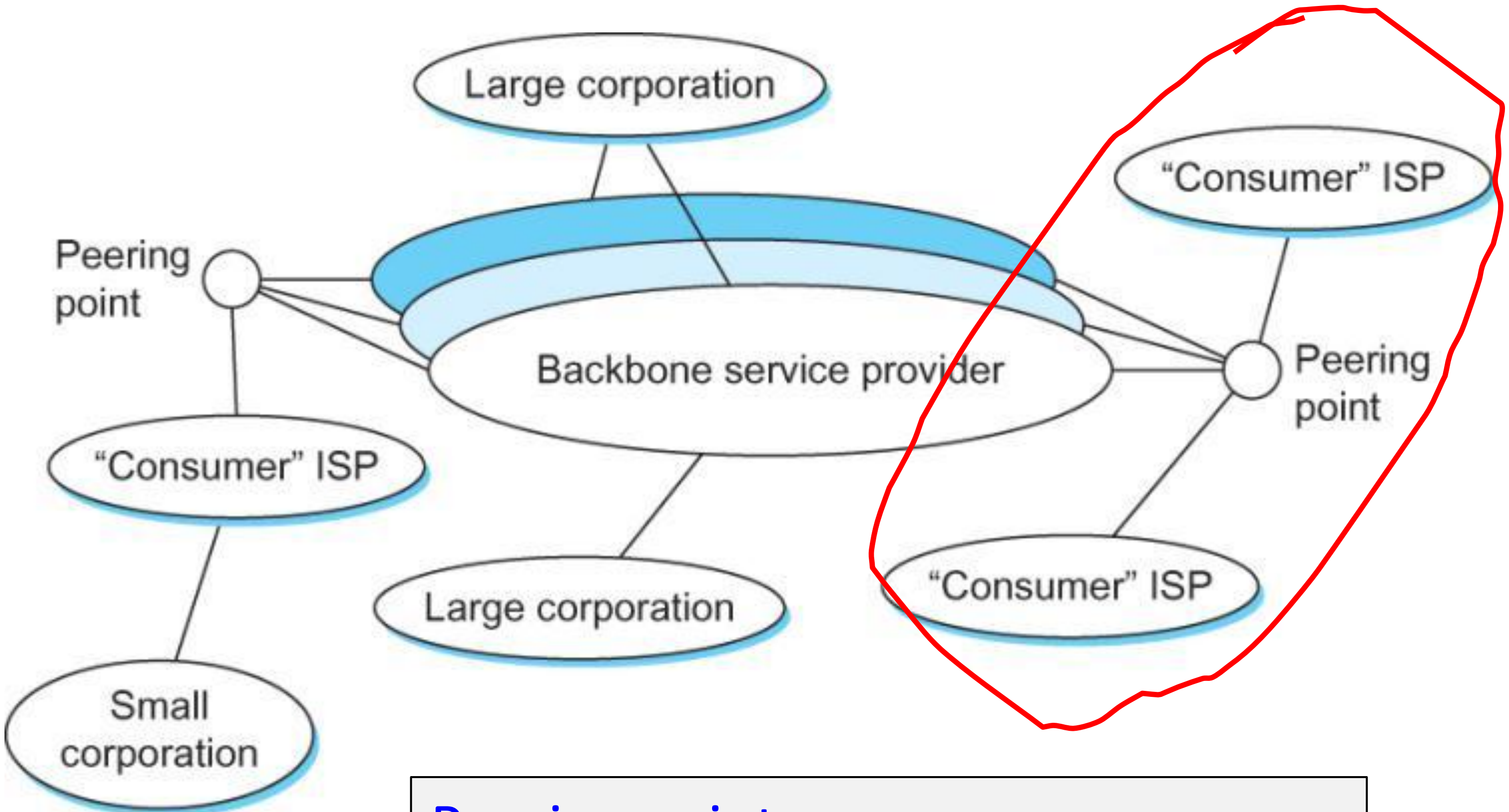
AS transit



Transit AS

- Connected to multiple ASes
- Designed to carry transit and local traffic

Peering point

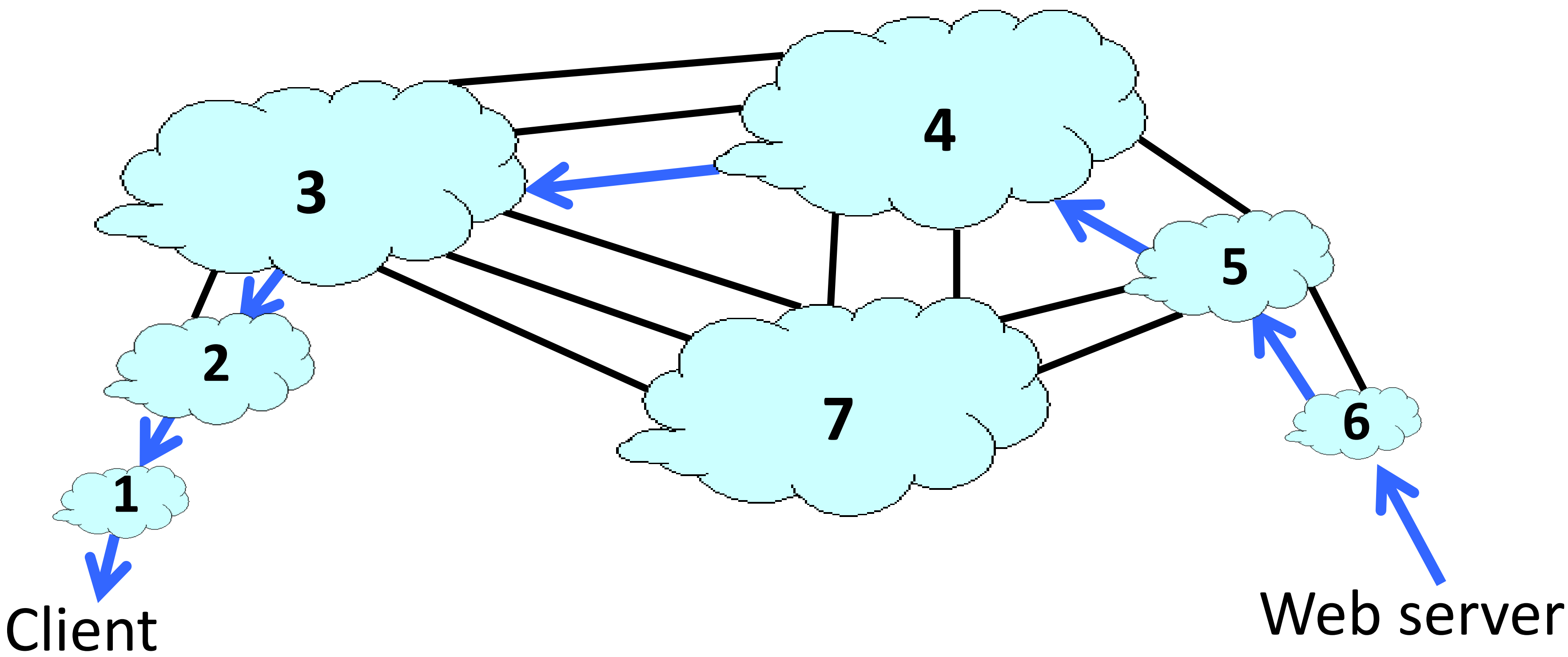


Peering point

- Allows ASes to connect directly, bypassing a transit AS.

Inter-AS routing

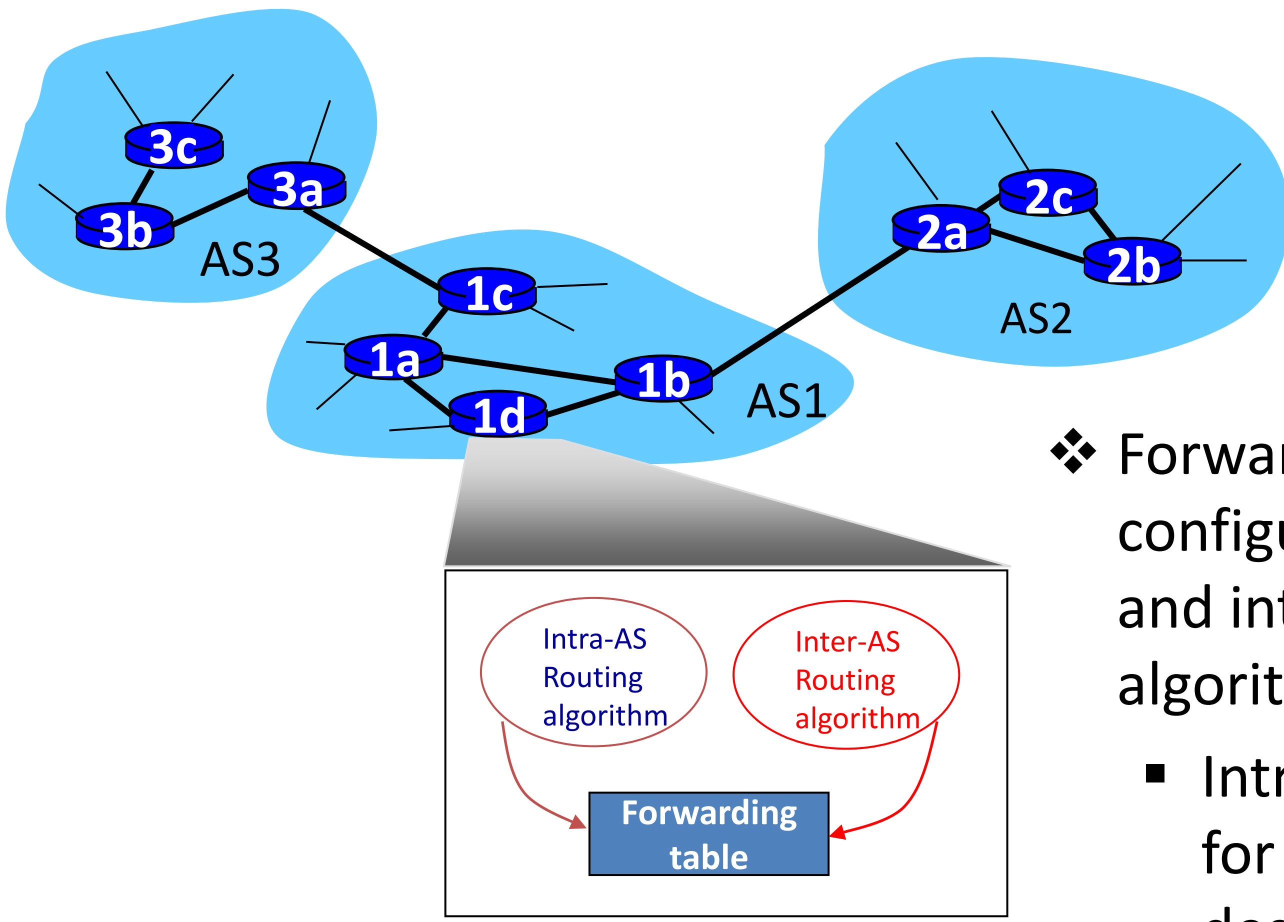
- AS-level topology
  - Destinations are IP prefixes (e.g. 12.0.0.0/8)
  - Nodes are Autonomous Systems (ASes)
  - Edges are links and business relationships



Inter-AS routing challenges

- Scale:
  - IP prefixes: 200,000+
  - ASes: 20K+ visible, 50k+ allocated
  - Routers: millions
- Privacy:
  - ASes don't want other to known topology
  - ASes don't want business relationships exposed
- Policy:
  - No internet-wide notion of link cost metric
  - Need control over where you send traffic, who you send traffic through, etc.

Interconnected ASes



- Forwarding table configured by both intra- and inter-AS routing algorithm
  - Intra-AS sets entries for internal destinations
  - Inter-AS & intra-AS sets entries for external destinations

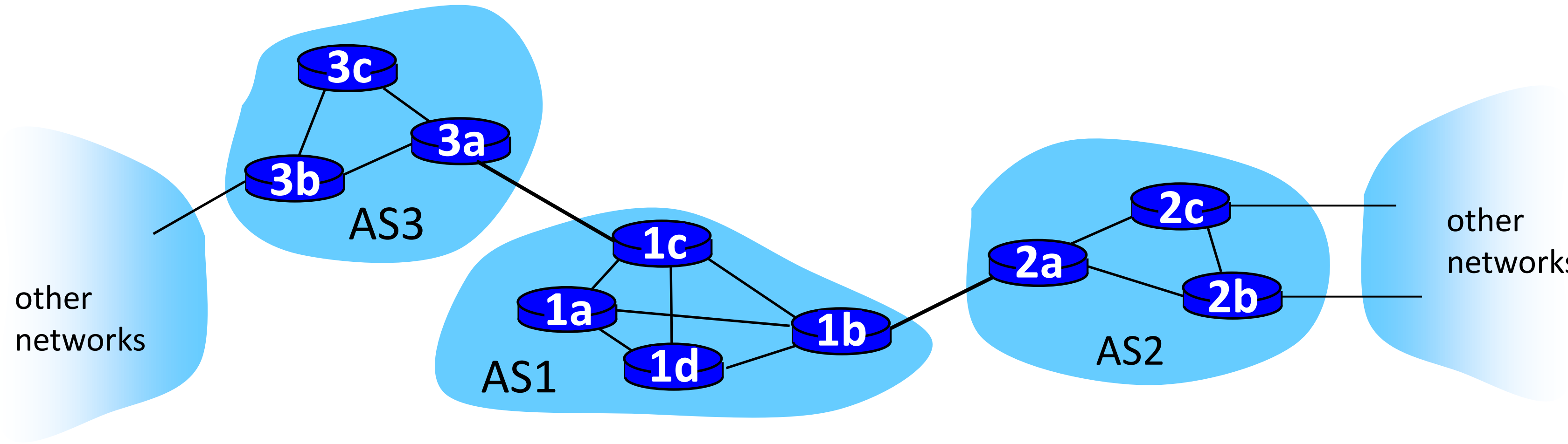
Inter-AS tasks

- Suppose router in AS1 receives datagram destined outside of AS1:
  - Router should forward packet to gateway router, but which one?

AS1 must:

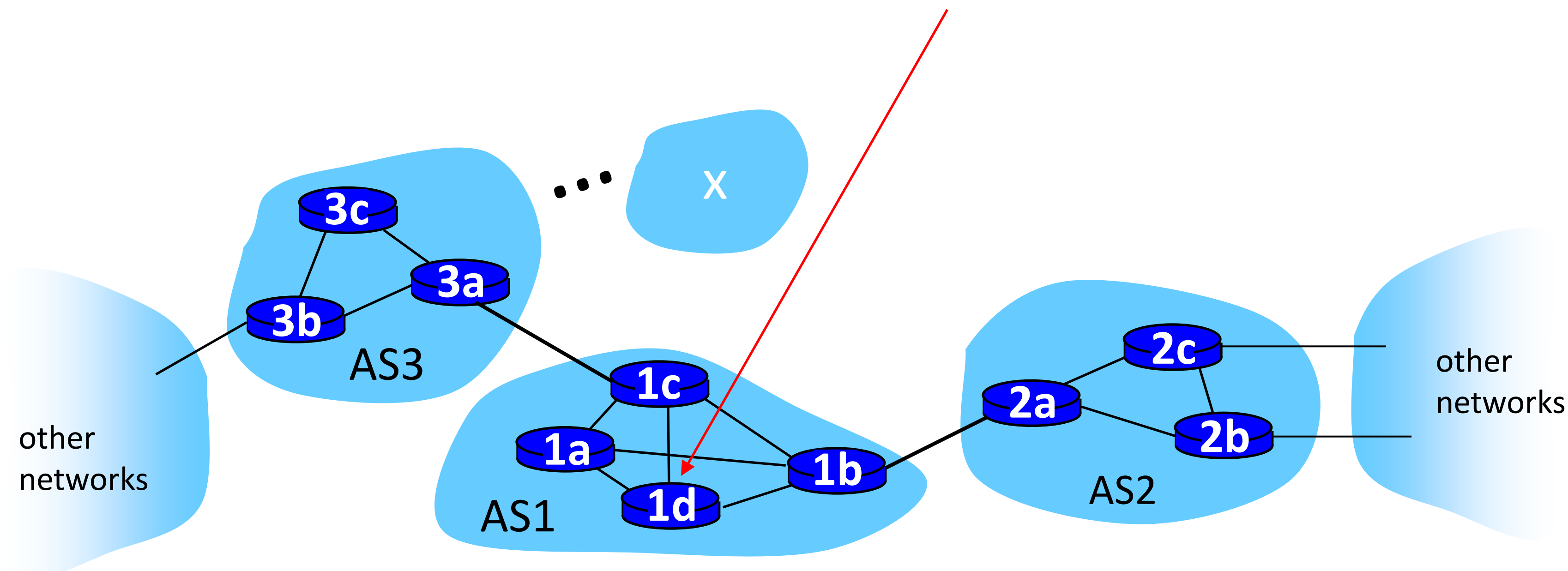
1. Learn which dests are reachable through AS2, which through AS3
2. Propagate this reachability info to all routers in AS1

Job of inter-AS routing!



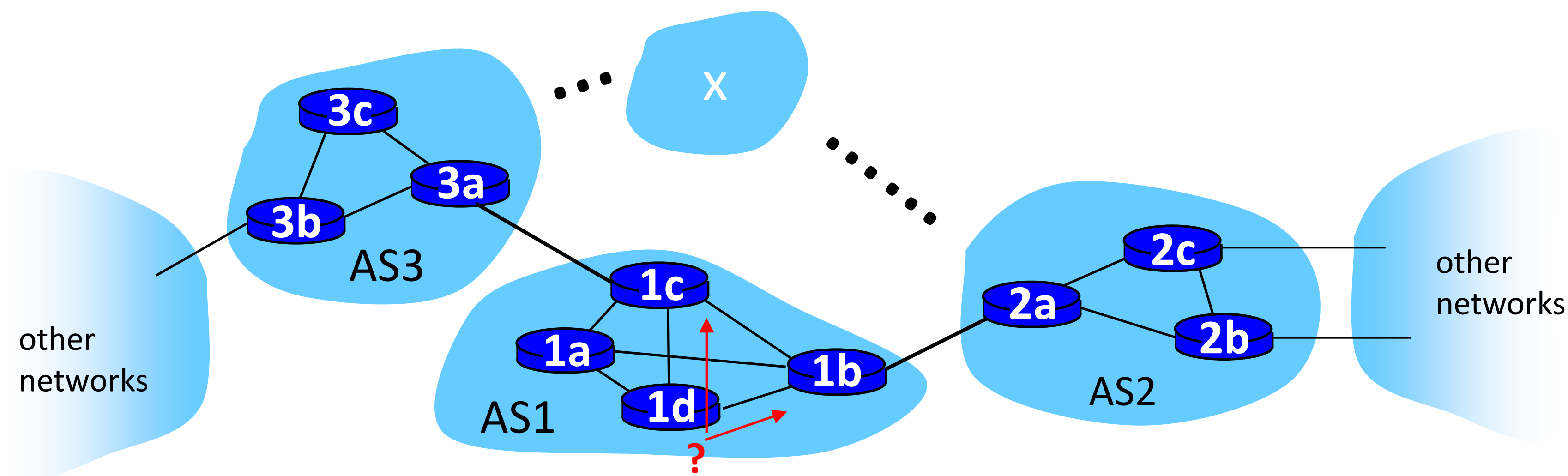
Setting forwarding table in router 1d

- AS1 learns that subnet *x* reachable via AS3 (gateway 1c), but not via AS2
  - Inter-AS protocol propagates info to all internal routers
- Router 1d determines from intra-AS routing info that its interface *I* is on the least cost path to 1c
  - Installs forwarding table entry *(x,I)*



Choosing among multiple ASes

- Now suppose AS1 learns from inter-AS protocol that subnet *x* is reachable from AS3 *and* AS2.
- To configure forwarding table, router 1d must determine towards which gateway it should forward packets for dest x
  - This is also job of inter-AS routing protocol!



Choosing among multiple ASes

- Now suppose AS1 learns from inter-AS protocol that subnet *x* is reachable from AS3 *and* AS2.
- To configure forwarding table, router 1d must determine towards which gateway it should forward packets for dest x
  - This is also job of inter-AS routing protocol!
- *Hot potato routing:* send towards closest router

Learn from inter-AS protocol that subnet *x* is reachable via multiple gateways

→

Use routing info from intra-AS protocol to determine costs of least-cost paths to each of the gateways

→

Hot potato routing: choose the gateway that has the smallest least cost

→

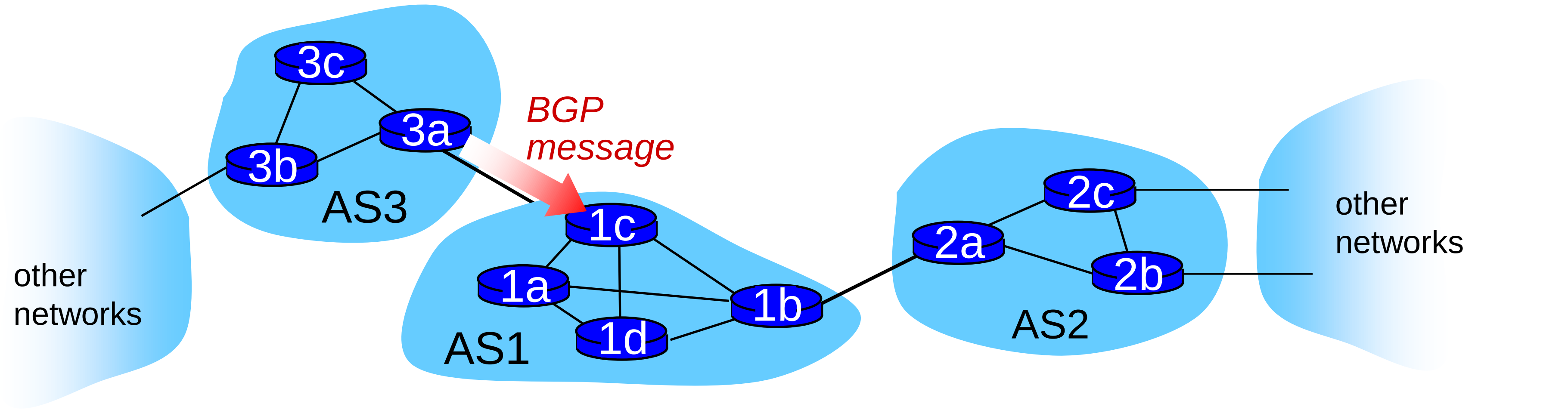
Determine from forwarding table the interface *I* that leads to least-cost gateway. Enter *(x,I)* in forwarding table

Internet inter-AS routing: BGP

- BGP (Border Gateway Protocol):
  - De facto inter-domain routing protocol
  - "glue that holds the Internet together"
- BGP allows each AS to:
  - eBGP: Obtain subnet reachability information from neighboring ASes
  - iBGP: Propagate reachability information to all AS-internal routers
  - Determine good routes to other networks based on reachability information and policy
  - Subnet to advertise its existence : *"I am here"*

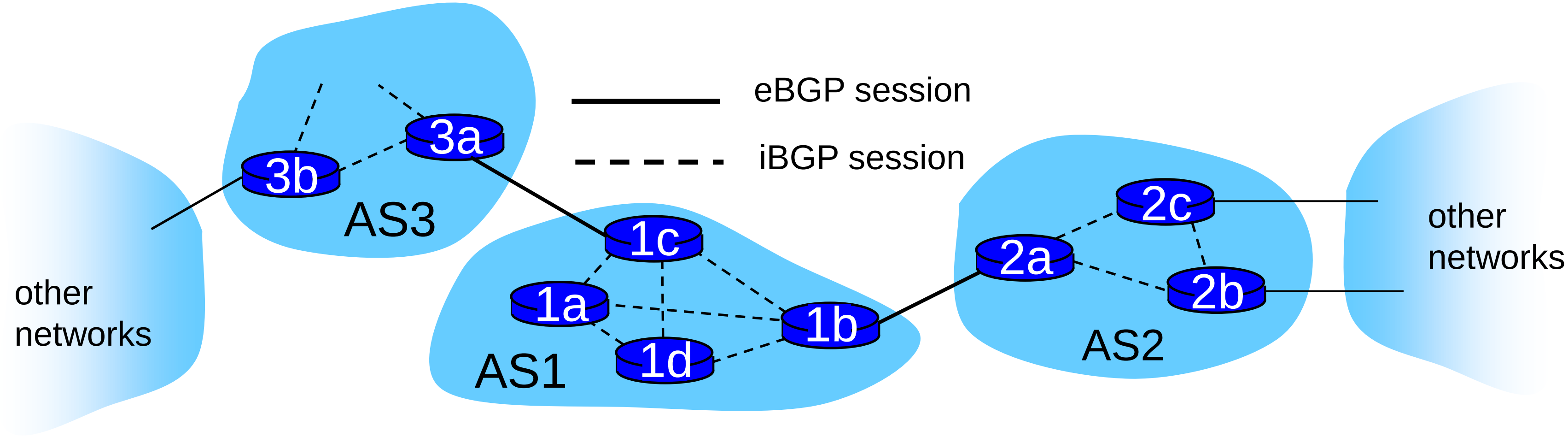
BGP basics

- **BGP session:**
  - Two BGP routers (peers) exchange BGP messages
  - Path vector protocol, advertise paths to different prefixes
  - Exchanged over semi-permanent TCP connections
- When AS3 advertises a prefix to AS1:
  - AS3 *promises* it will forward datagrams towards that prefix
  - AS3 can aggregate prefixes in its advertisement



BGP basics: distributing path info

- Using eBGP session between 3a and 1c, AS3 sends prefix reachability info to AS1.
  - 1c can then use iBGP to distribute new prefix to all routers in AS1
  - 1b can then re-advertise new reachability info to AS2 over 1b-to-2a eBGP session
- When router learns of new prefix, it creates entry for prefix in its forwarding table.



Path attributes and BGP routes

- Advertised prefix includes BGP attributes
  - prefix + attributes = route
  - AS-PATH: contains ASs through which prefix advertisement has passed: e.g., AS 67, AS 17
  - NEXT-HOP: indicates specific internal-AS router to next-hop AS. (may be multiple links from current AS to next-hop-AS)
- Gateway router uses import policy to accept/decline
  - e.g., never route through AS x
  - *policy-based* routing

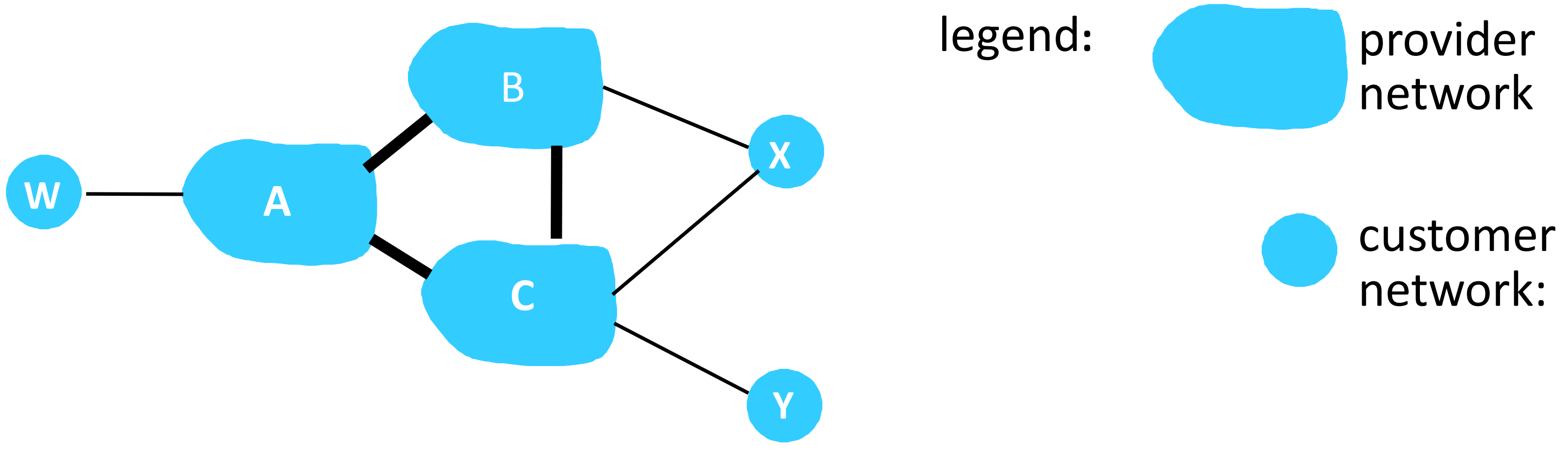
BGP route selection

- Router may learn about more than 1 route to destination AS, selects route based on:
  1. Local preference value attribute: policy decision
  2. Shortest AS-PATH
  3. Closest NEXT-HOP router: hot potato routing
  4. Additional criteria

BGP messages

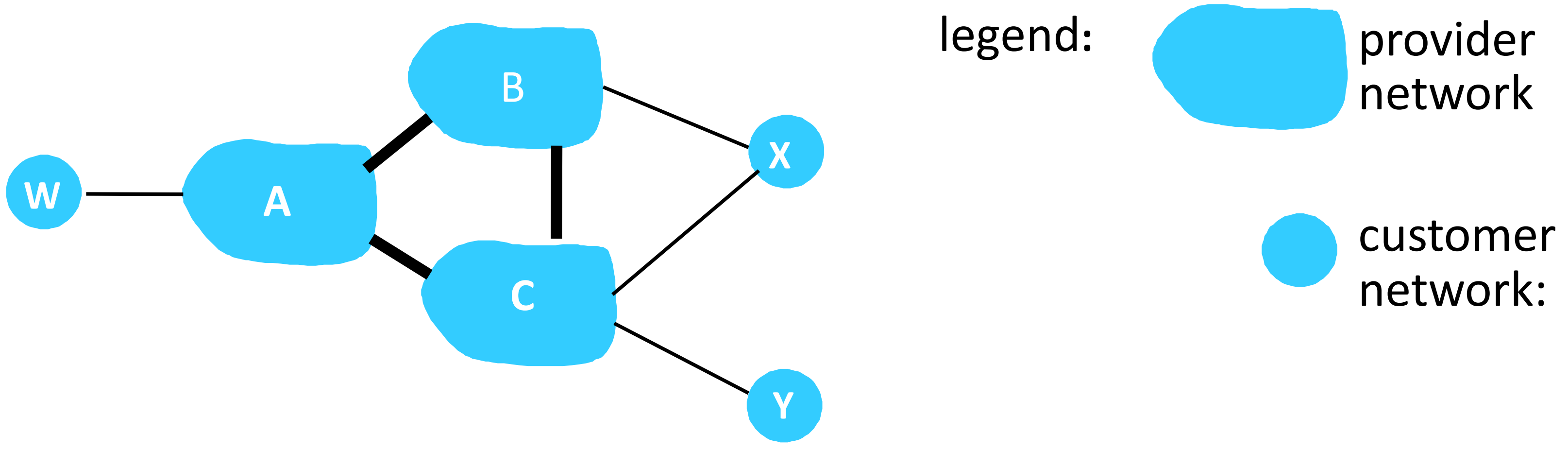
- BGP messages exchanged between peers over TCP connection
- BGP messages:
  - OPEN: opens TCP connection to peer and authenticates sender
  - UPDATE: advertises new path (or withdraws old)
  - KEEPALIVE: keeps connection alive in absence of UPDATES; also ACKs OPEN request
  - NOTIFICATION: reports errors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

BGP routing policy



- A,B,C are *provider networks*
- X,W,Y are customer (of provider networks)
- X is *dual-homed:* attached to two networks
  - 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 (2)



- A advertises path AW to B
- B advertises path BAW to X
- Should B advertise path BAW to C?
  - No way! 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!

Why different Intra-, Inter-AS routing ?

Policy:

- Inter-AS: Admin wants control over how its traffic routed, who routes through its net.
- Intra-AS: Single admin, so no policy decisions needed

Scale:

- Hierarchical routing saves table size, reduced update traffic

Performance:

- Intra-AS: Can focus on performance
- Inter-AS: Policy may dominate over performance

Summary

- Inter-AS routing
  - Scaling routing to Internet scale
  - Routing between independent ASes
  - Allows routing to encode business rules
- Border Gateway Protocol (BGP)
  - A path vector protocol
  - The glue that holds the Internet together