UNIVERSITY OF WATERLOO

CS 456/656
Computer Networks

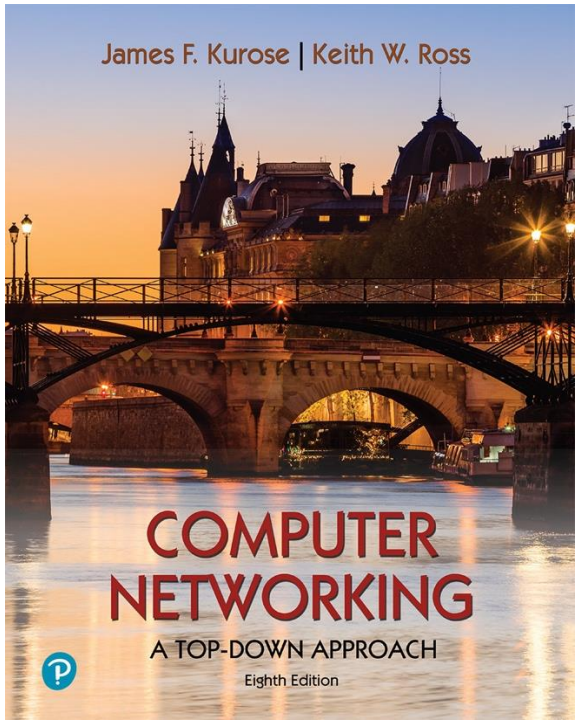
Lecture 12: Network Layer – Part 4

Mina Tahmasbi Arashloo and Uzma Maroof

Fall 2025

A note on the slides

Adapted from the slides that accompany this book.



Computer Networking: A Top-Down Approach
8th edition
Jim Kurose, Keith Ross
Pearson, 2020

Network Layer in the Internet

- The Internet Protocol (IP)
- Internet Routing
  - Hierarchical routing
  - Intra-ISP routing: OSPF
  - Inter-ISP routing: BGP
- Internet Forwarding

Network Layer in the Internet

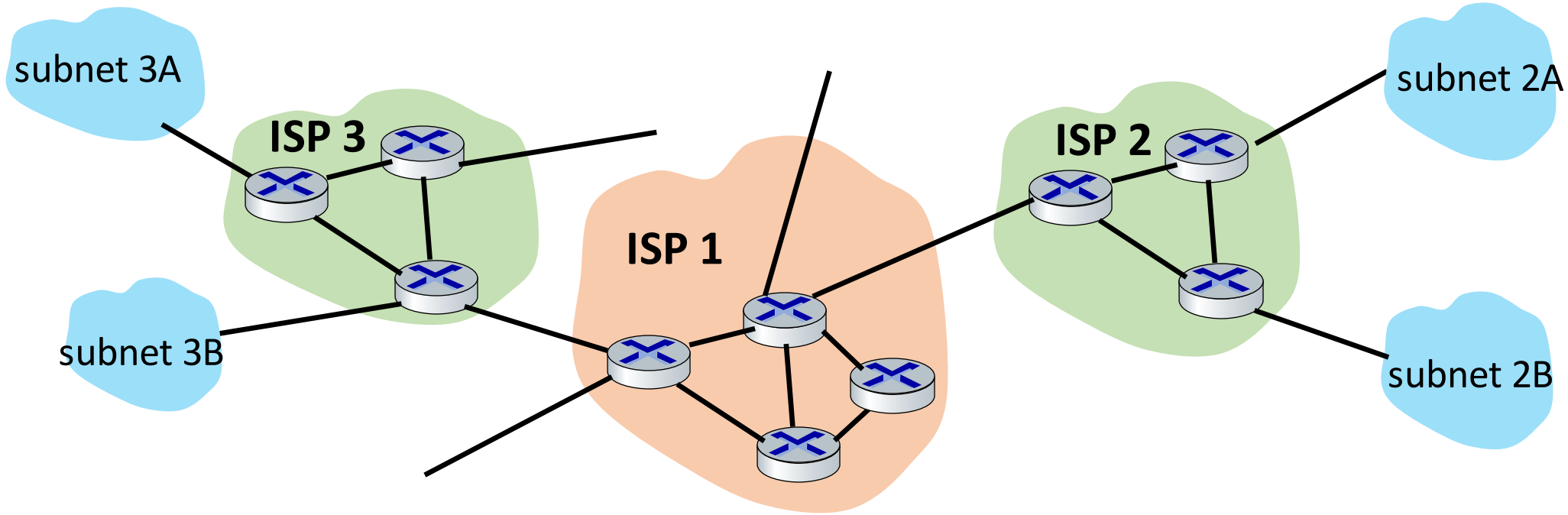
- The Internet Protocol (IP)
- Internet Routing
  - Hierarchical routing
  - Intra-ISP routing: OSPF
  - Inter-ISP routing: BGP
- Internet Forwarding

Challenges in the Internet routing

Our routing study thus far - idealized

- all routers identical
- network “flat”

… not true in practice



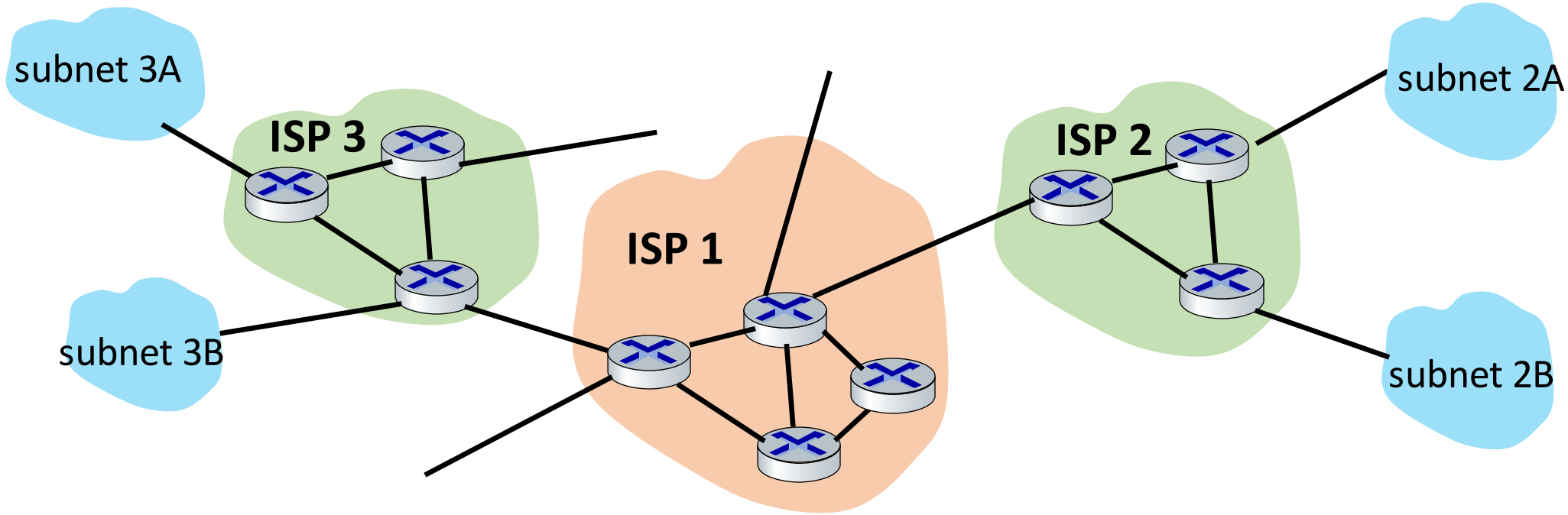
Challenges in the Internet routing

scale: billions of destinations:

- can't store all destinations in routing tables!
- routing table exchange would swamp links!

administrative autonomy:

- Internet: a network of networks
- each network admin may want to control routing in its own network



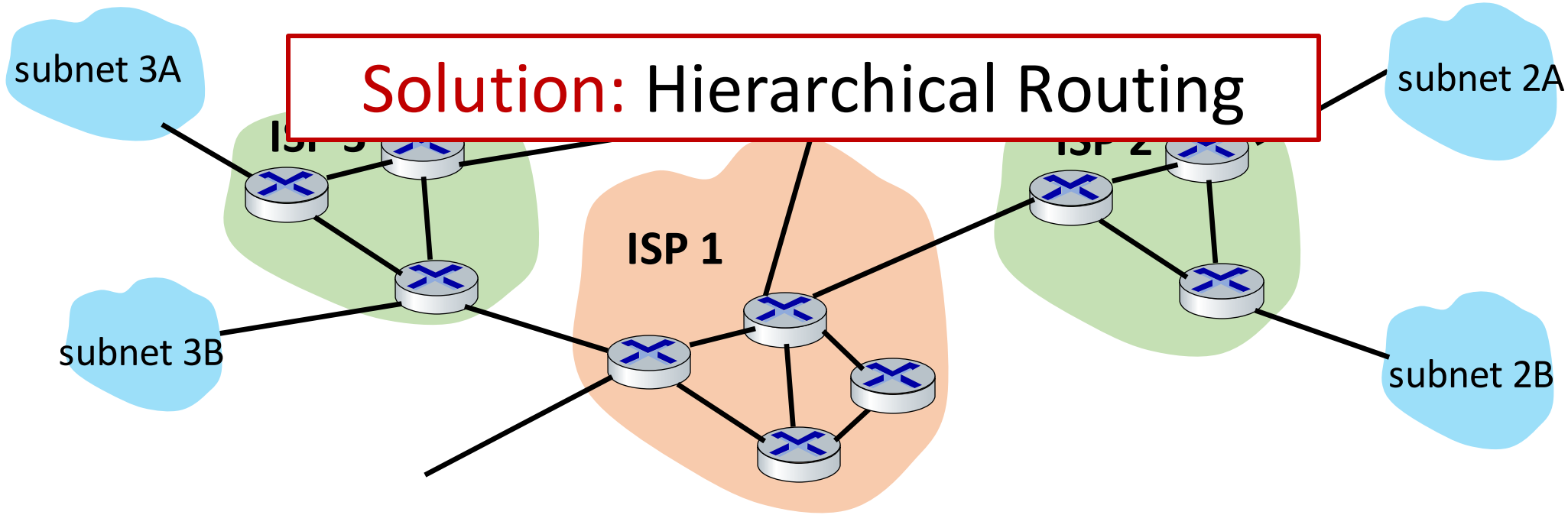
Challenges in the Internet routing

scale: billions of destinations:

- can't store all destinations in routing tables!
- routing table exchange would swamp links!

administrative autonomy:

- Internet: a network of networks
- each network admin may want to control routing in its own network



Hierarchical routing

aggregate routers into regions known as "autonomous systems" (AS) (a.k.a. "domains")

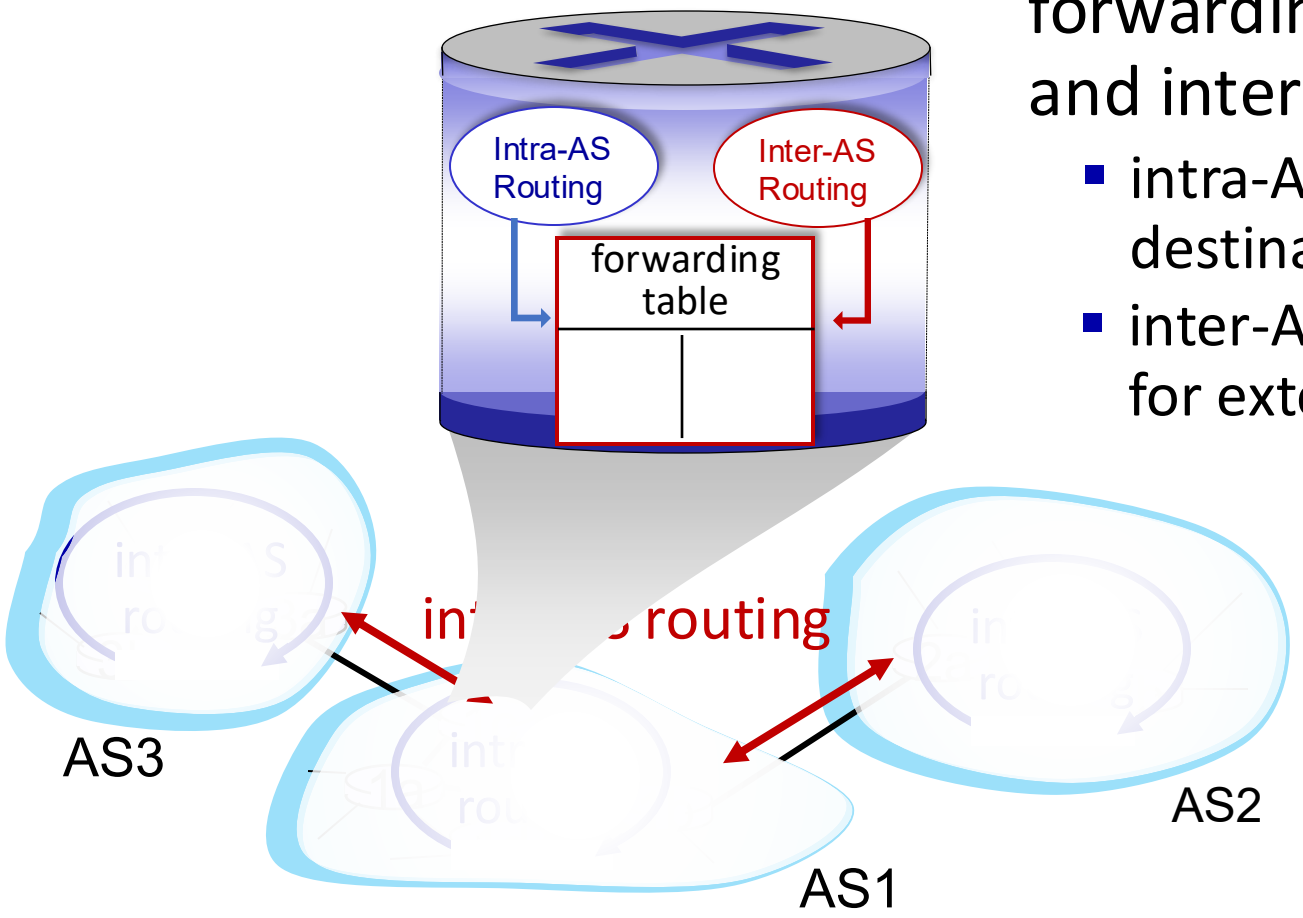
intra-AS (aka "intra-domain"): routing among routers *within same AS ("network")*

- all routers in AS must run same intra-domain protocol
- routers in different AS can run different intra-domain routing protocols
- gateway router: at "edge" of its own AS, has link(s) to router(s) in other AS'es

inter-AS (aka "inter-domain"): routing *among* AS'es

- gateways perform inter-domain routing (as well as intra-domain routing)

Interconnected ASes



forwarding table configured by intra- and inter-AS routing algorithms

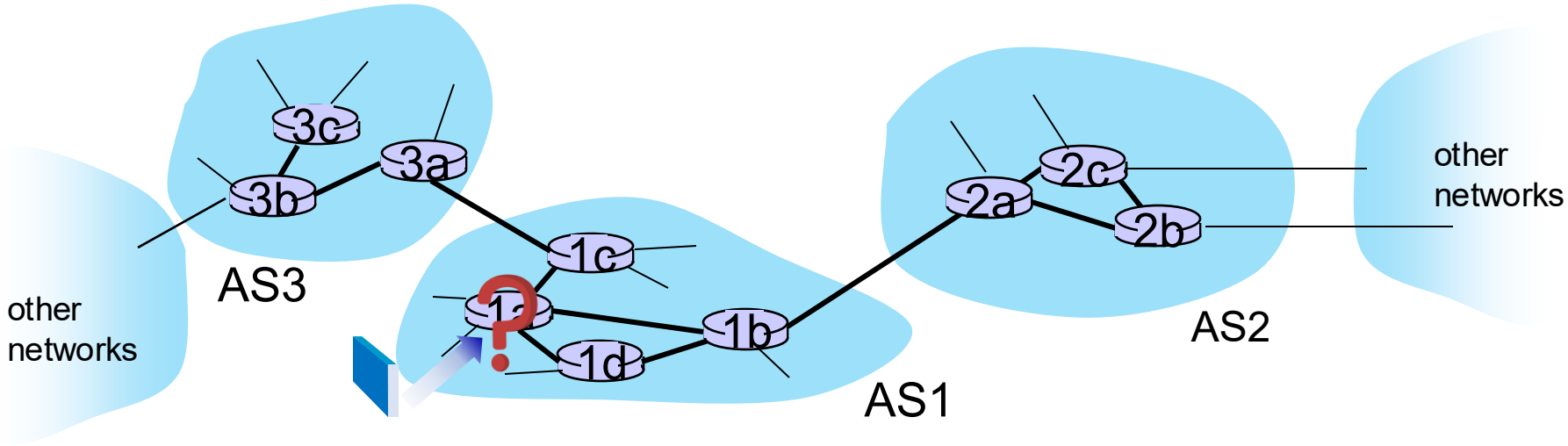
- intra-AS routing determine entries for destinations within AS
- inter-AS & intra-AS determine entries for external destinations

Intra-AS routing's role in an AS's data forwarding

- suppose router in AS1 receives datagram destined inside of AS1:
  - i.e., the destination subnet is attached to one of AS1's routers

AS1 intra-domain routing must:

1. learn which destinations reachable within AS1
2. What are the best routes, within AS1, to get to those destinations

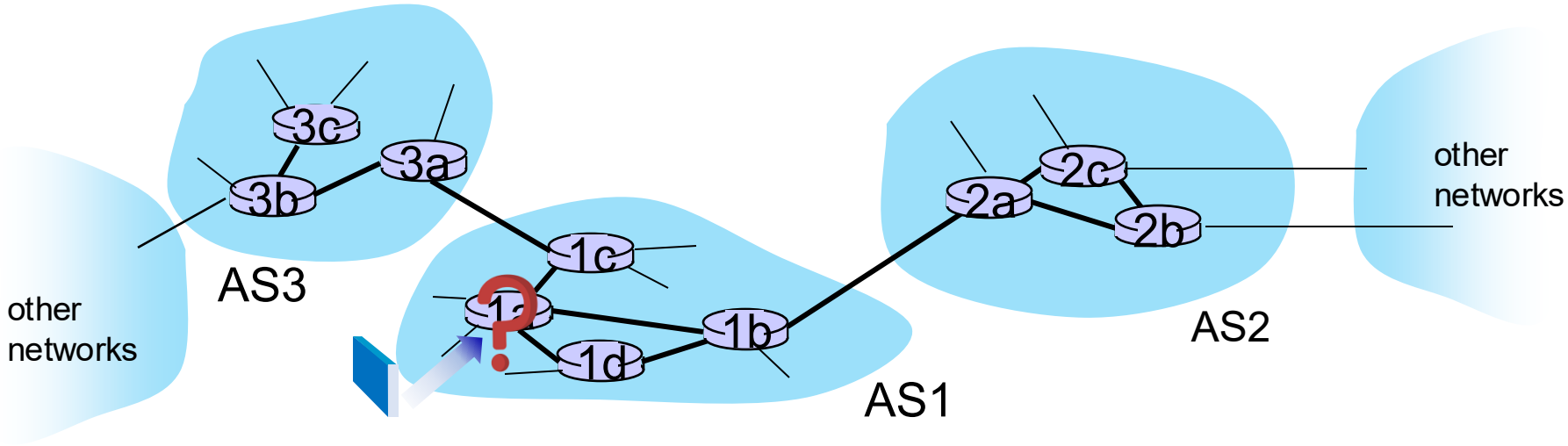


Inter-AS routing's role in an AS's data forwarding

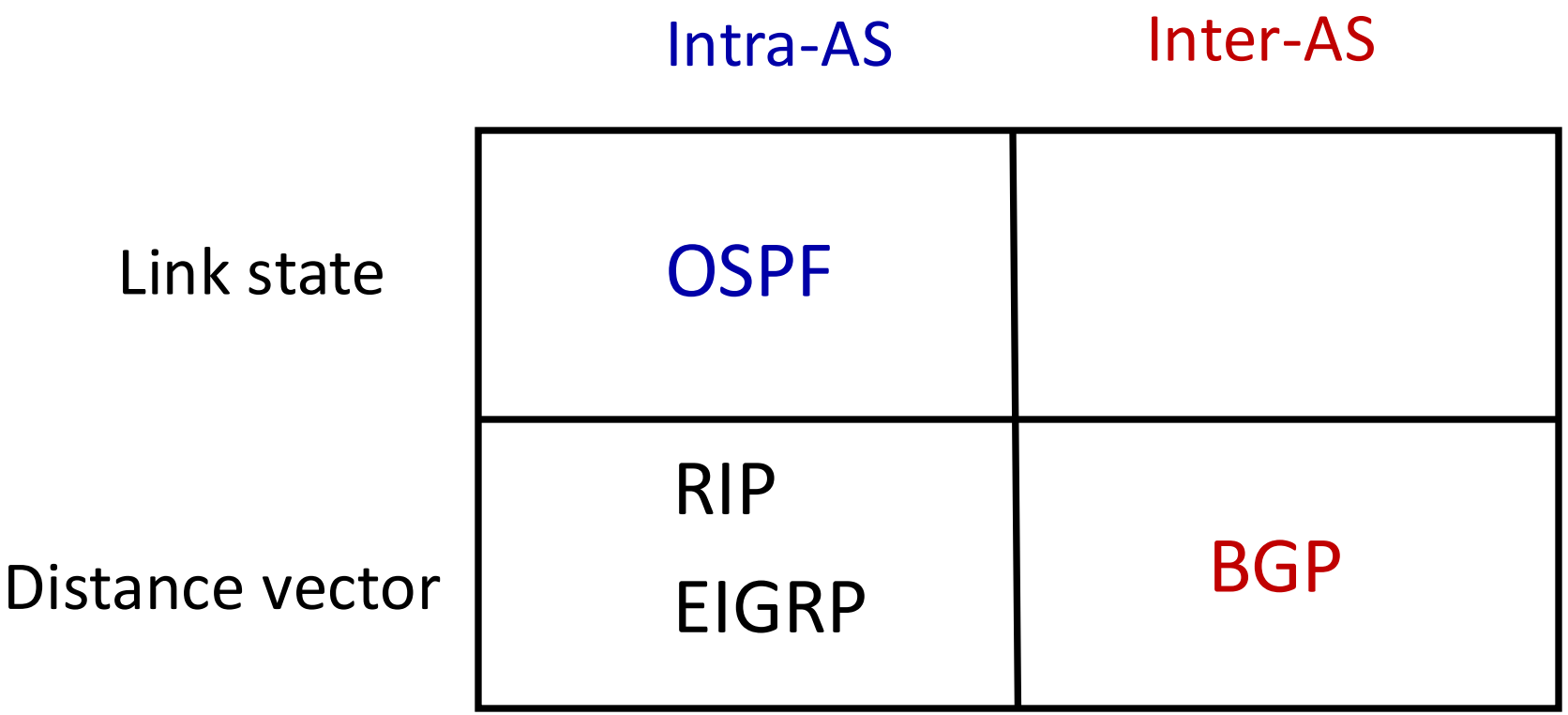
- suppose router in AS1 receives datagram destined outside of AS1:
  - router should forward packet to gateway router in AS1 (e.g., 1c, or 1b), but which one?

AS1 inter-domain routing must:

1. learn which destinations reachable through AS2, which through AS3
2. propagate this reachability info to all routers in AS1



Internet routing protocols



Network Layer in the Internet

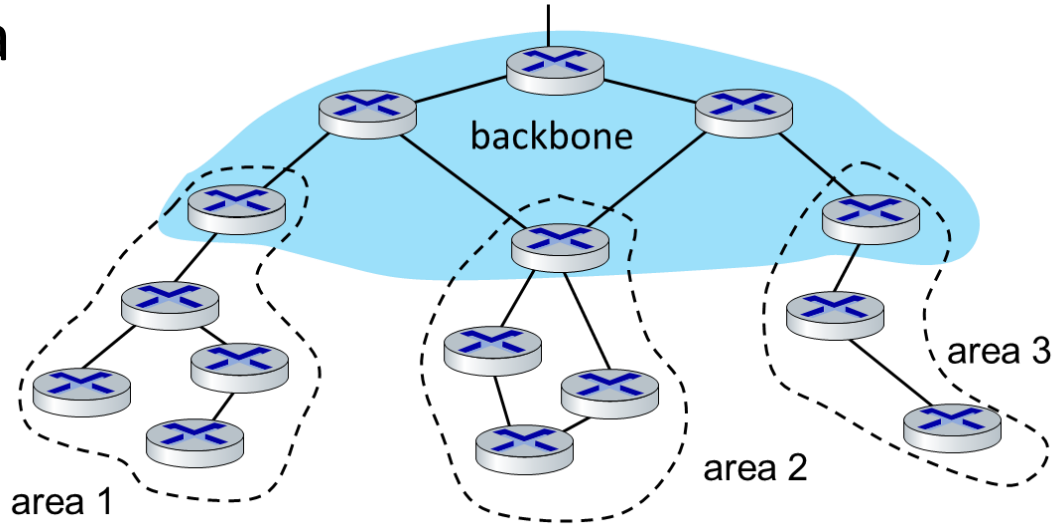
- The Internet Protocol (IP)
- Internet Routing
  - Hierarchical routing
  - Intra-ISP routing: OSPF
  - Inter-ISP routing: BGP
- Internet Forwarding

OSPF (Open Shortest Path First) routing

- "open": publicly available
- classic link-state
  - each router floods OSPF link-state advertisements to all other routers in entire AS
    - directly over IP rather than using TCP/UDP (upper-layer protocol of 89)
    - Responsible for reliable message transfer
  - each router has full topology, uses Dijkstra's algorithm to compute forwarding table

OSPF (Open Shortest Path First) routing

- security: all OSPF messages authenticated (to prevent malicious intrusion)
- link costs are configured by network admin
  - multiple link costs metrics possible: bandwidth, delay
- support for hierarchy within a single AS
  - Two levels: backbone, local area



Other common intra-AS routing protocols

- RIP: Routing Information Protocol [RFC 1723]
  - classic DV: DVs exchanged every 30 secs
  - no longer widely used
- EIGRP: Enhanced Interior Gateway Routing Protocol
  - DV based
  - formerly Cisco-proprietary for decades (became open in 2013 [RFC 7868])

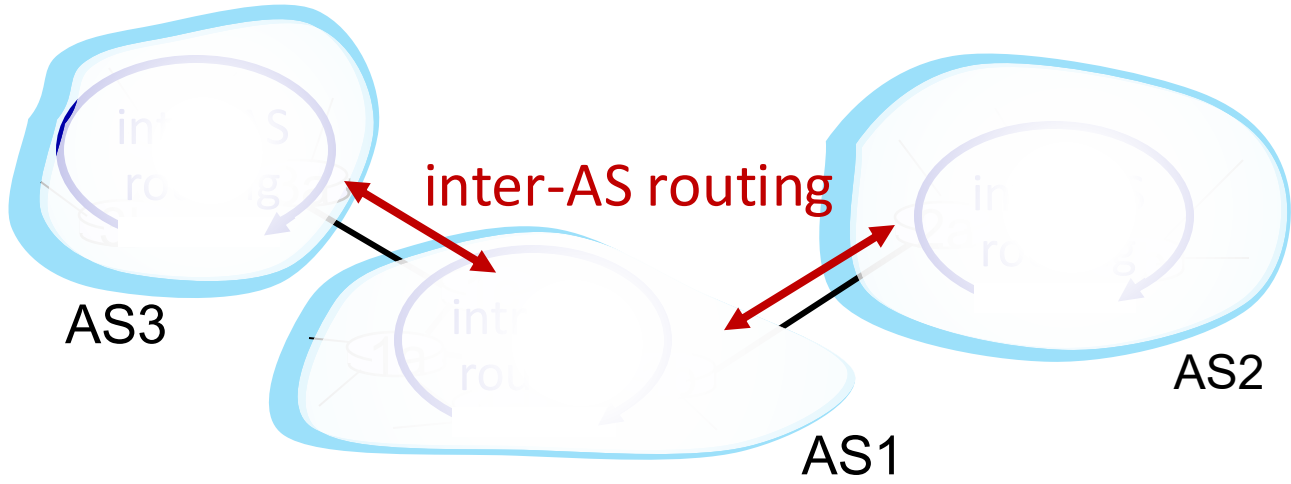
Make sure you know

- What an Autonomous System (AS) is
- What Intra-AS routing means
- What Inter-AS routing means
- How the two work together to find paths from any source to any destination
  - E.g., suppose you have a network with many AS. Suppose you are given a packet whose source is in one AS and its destination in another AS. How would the network of networks know how to forward this packet? What if the destination is in the same AS?
- What are the features of OSPF?

Network Layer in the Internet

- The Internet Protocol (IP)
- Internet Routing
  - Hierarchical routing
  - Intra-ISP routing: OSPF
  - Inter-ISP routing: BGP
- Internet Forwarding

Interconnected ASes



intra-AS (aka “intra-domain”): routing among routers *within same AS (“network”)*

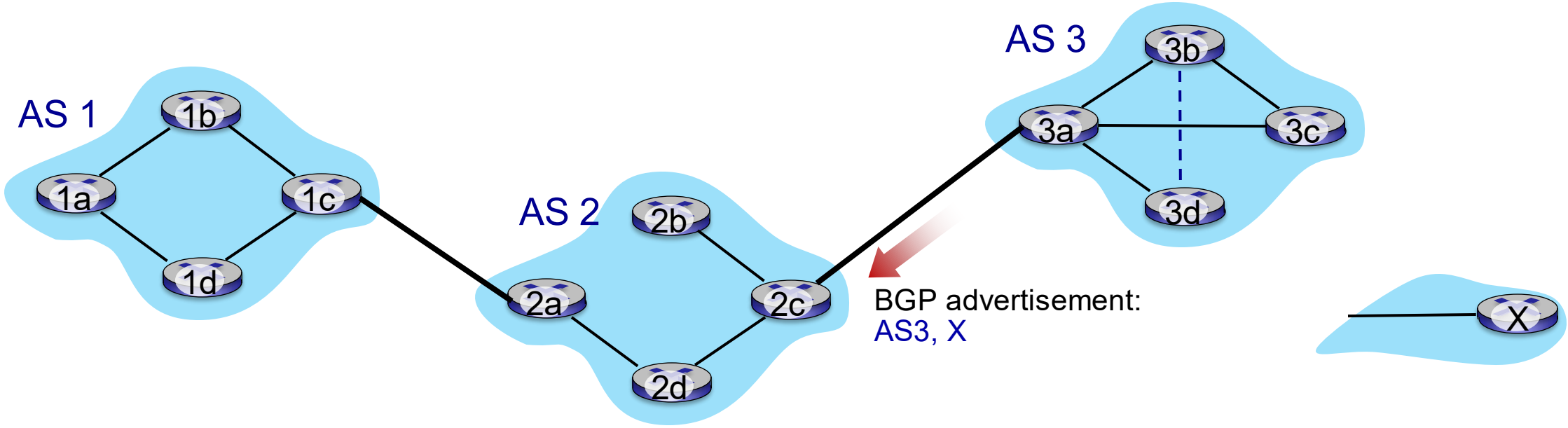
inter-AS (aka “inter-domain”): routing *among* AS’es

Internet inter-AS routing: BGP

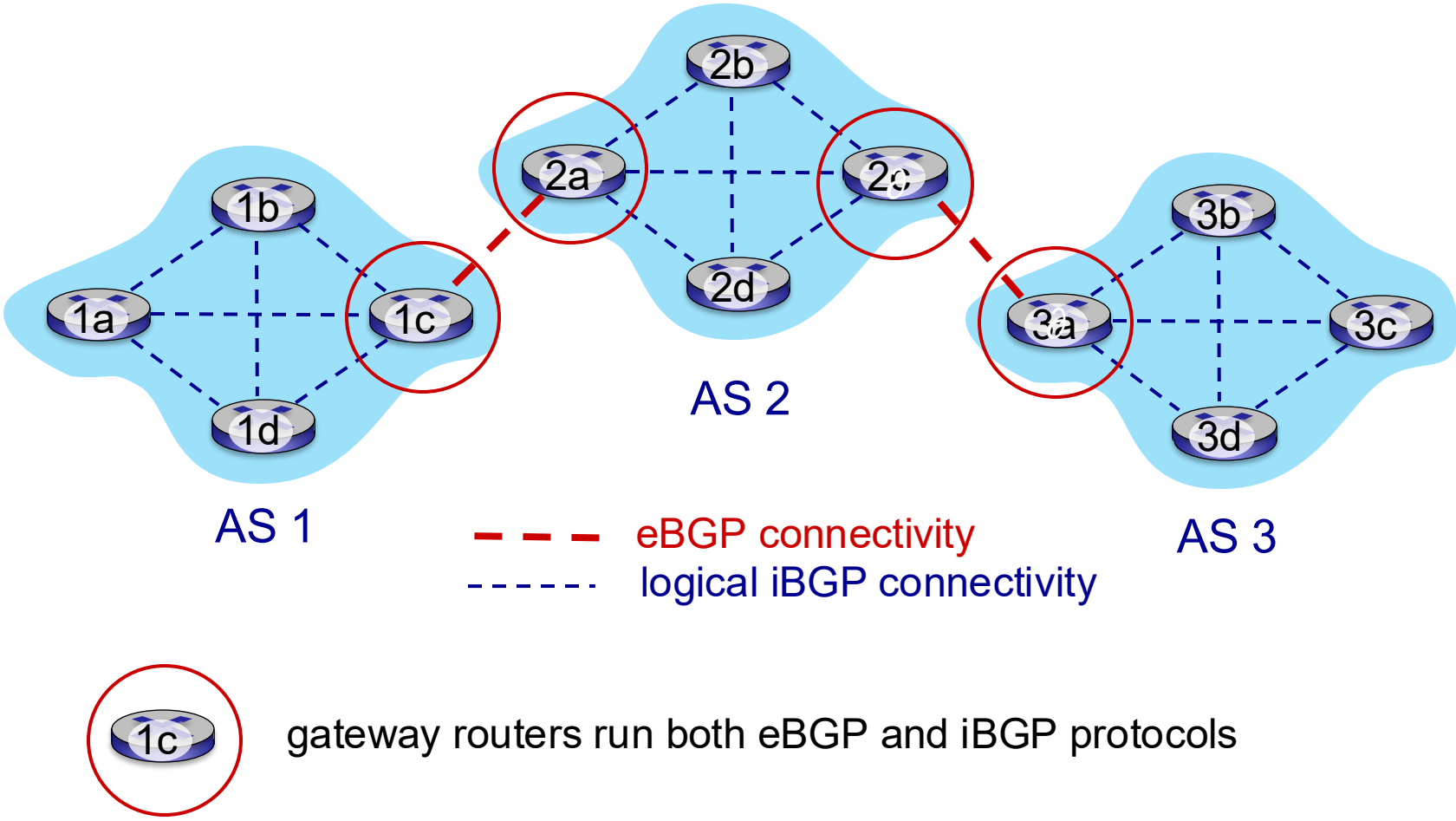
- BGP (Border Gateway Protocol): *the* de facto inter-domain routing protocol
  - "glue that holds the Internet together"
- allows subnet to advertise its existence, and the destinations it can reach, to rest of Internet: *"I am here, here is who I can reach, and how"*
- BGP provides each AS a means to:
  - obtain destination network reachability info from neighboring ASes (eBGP)
  - Decides which routes to use based on reachability information and *policy*
  - propagates reachability information to all AS-internal routers (iBGP)
  - advertise (to neighboring networks) destination reachability info

BGP basics

- BGP session: two BGP routers ("peers") exchange BGP messages over TCP connection:
  - advertising *paths* to different destination network prefixes (BGP is a "path vector" protocol)
- when AS3 gateway 3a advertises path AS3,X to AS2 gateway 2c:
  - AS3 *promises* to AS2 it will forward datagrams towards X

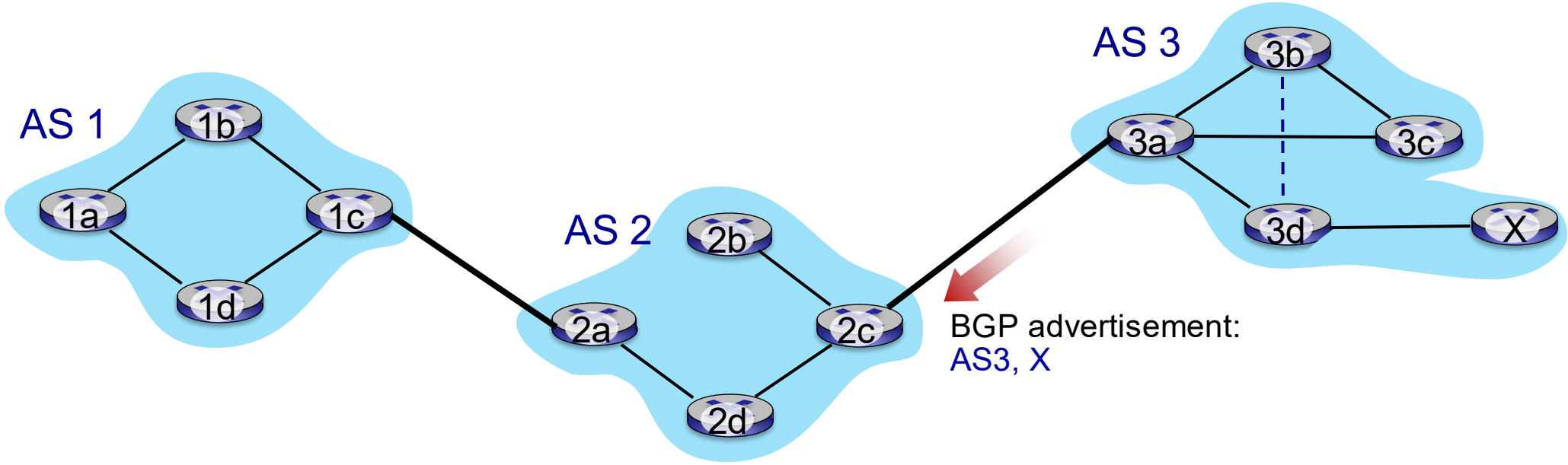


eBGP and iBGP connections



BGP basics

- BGP session: two BGP routers ("peers") exchange BGP messages over TCP connection:
  - advertising *paths* to different destination network prefixes (BGP is a "path vector" protocol)
- when AS3 gateway 3a advertises path AS3,X to AS2 gateway 2c:
  - AS3 *promises* to AS2 it will forward datagrams towards X

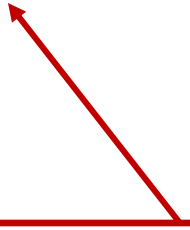


Path attributes and BGP routes

- BGP advertised route: prefix + attributes
  - prefix: destination being advertised
  - two important attributes:
    - AS-PATH: list of ASes through which prefix advertisement has passed
    - NEXT-HOP: indicates specific internal-AS router to next-hop AS

Path attributes and BGP routes

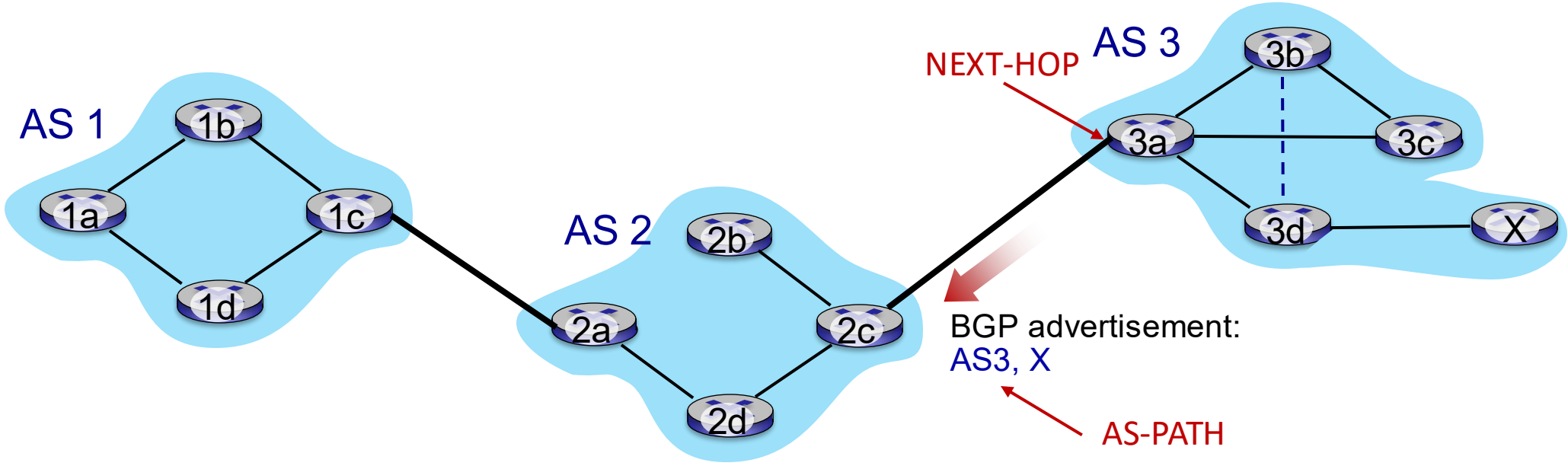
- BGP advertised route: prefix + attributes
  - prefix: destination being advertised
  - two important attributes:
    - AS-PATH: list of ASes through which prefix advertisement has passed
    - NEXT-HOP: indicates specific internal-AS router to next-hop AS



- Path vector routing
- As opposed to distance vector
- Avoids ”count-to-infinity”-like problems (e.g., easy to detect loops in the path)

BGP basics

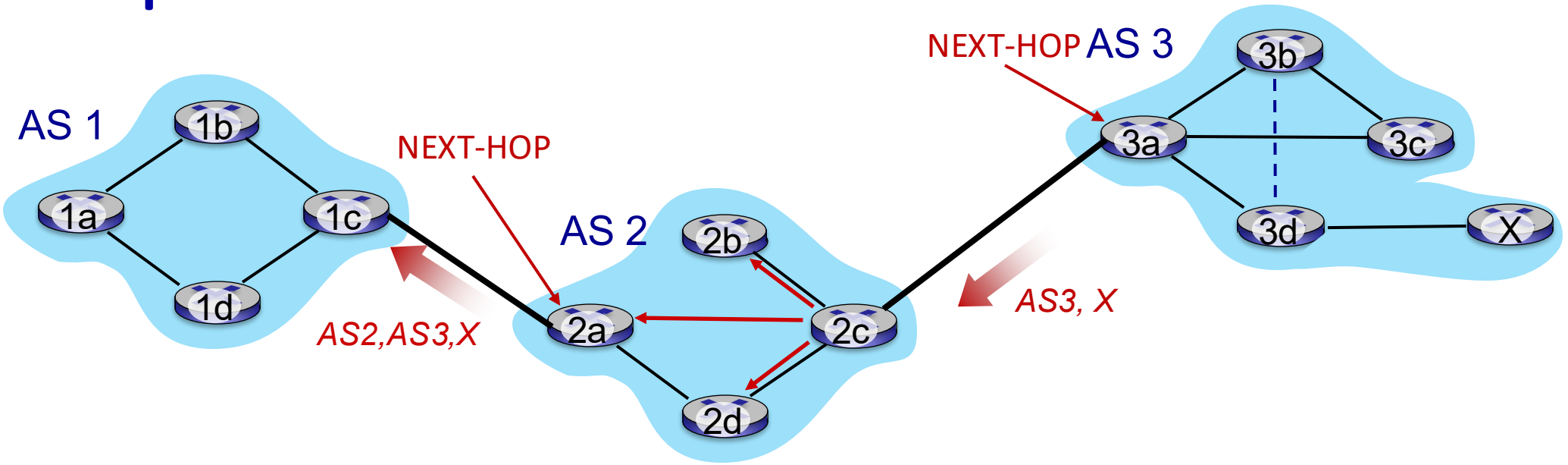
- BGP session: two BGP routers ("peers") exchange BGP messages over TCP connection:
  - advertising *paths* to different destination network prefixes (BGP is a "path vector" protocol)
- when AS3 gateway 3a advertises path AS3,X to AS2 gateway 2c:
  - AS3 *promises* to AS2 it will forward datagrams towards X



Path attributes and BGP routes

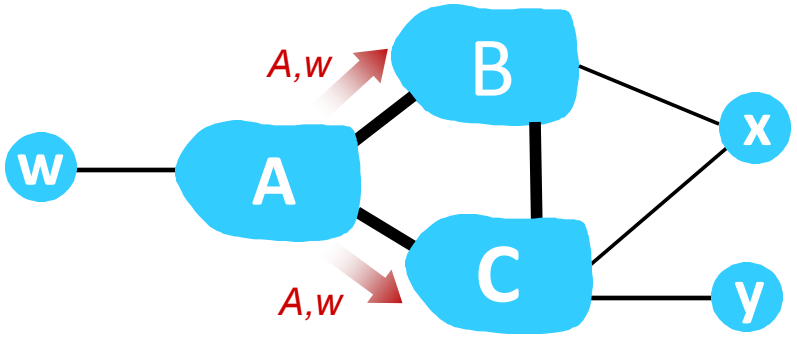
- BGP advertised route: prefix + attributes
  - prefix: destination being advertised
  - two important attributes:
    - AS-PATH: list of ASes through which prefix advertisement has passed
    - NEXT-HOP: indicates specific internal-AS router to next-hop AS
- policy-based routing:
  - gateway receiving route advertisement uses *import policy* to accept/decline path (e.g., never route through AS Y).
  - AS policy also determines whether to *advertise* path to other neighboring ASes

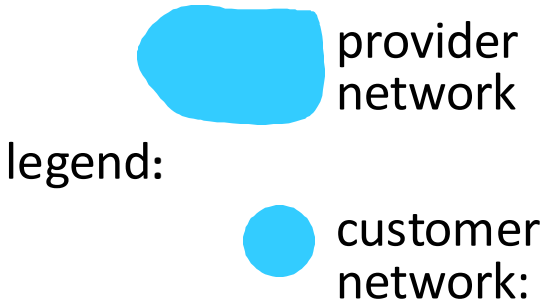
BGP path advertisement



- AS2 router 2c receives path advertisement AS3,X (via eBGP) from AS3 router 3a
- based on AS2 policy, AS2 router 2c accepts path AS3,X, propagates (via iBGP) to all AS2 routers
- based on AS2 policy, AS2 router 2a advertises (via eBGP) path AS2, AS3, X to AS1 router 1c

BGP: achieving policy via advertisements

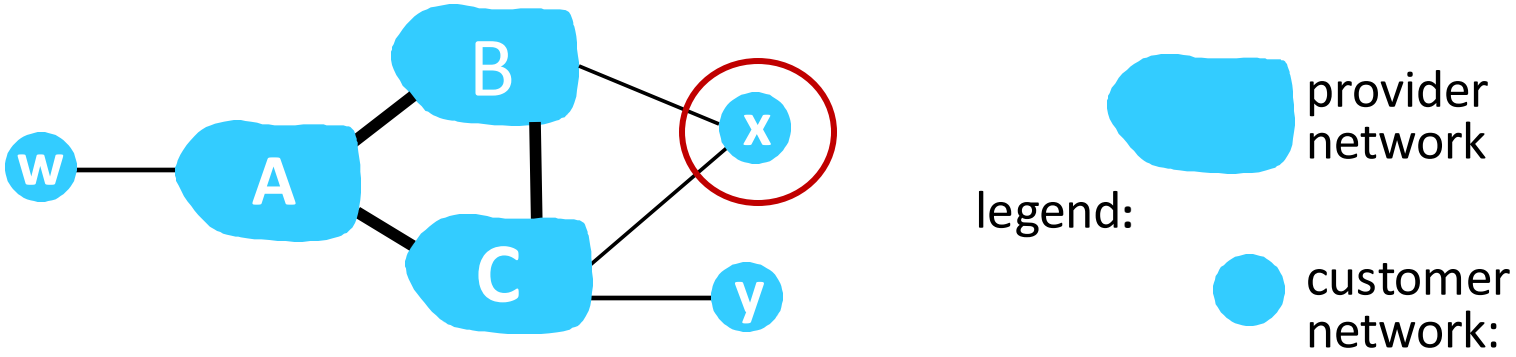




ISP only wants to route traffic to/from its customer networks (does not want to carry transit traffic between other ISPs – a typical "real world" policy)

- A advertises path *A,w* to B and to C
- B *chooses not to advertise B,A,w* to C!
  - B gets no "revenue" for routing *C,B,A,w*, since none of C, A, w are B's customers
  - C does *not* learn about *C,B,A,w* path
- C will route *C,A,w* (not using B) to get to w

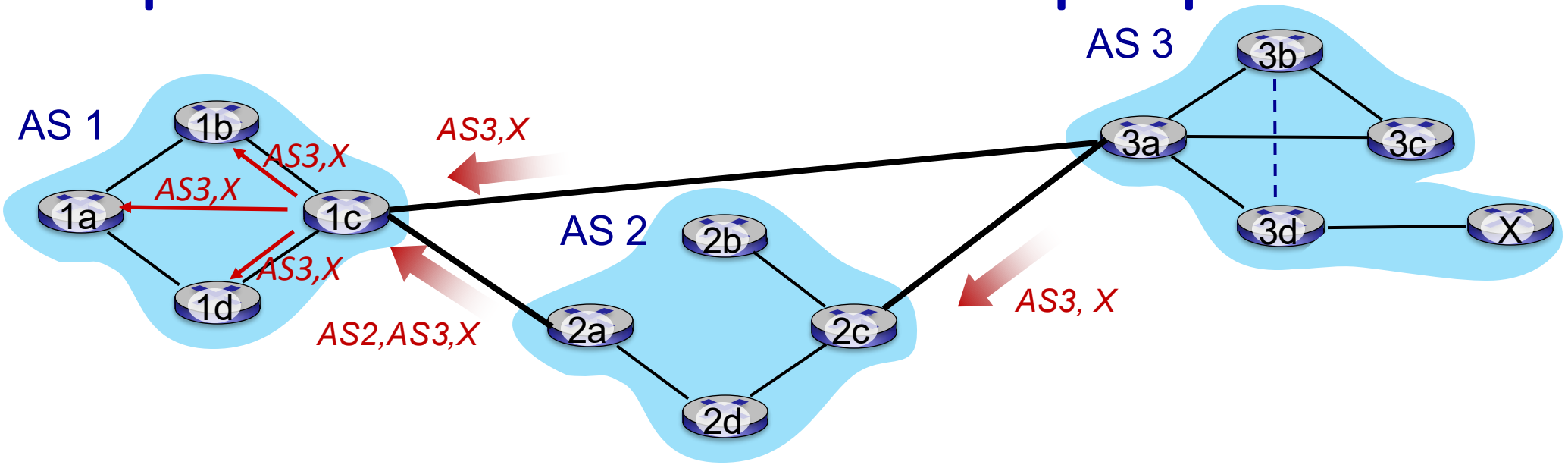
BGP: achieving policy via advertisements (more)



Network only wants to route traffic to/from its customers (does not want to carry transit traffic between other ISPs – a typical "real world" policy)

- A,B,C are provider networks
- x,w,y are customer (of provider networks)
- x is dual-homed: attached to two networks
- *policy to enforce:* x does not want to route from B to C via x
  - .. so x will not advertise to B a route to C

BGP path advertisement: multiple paths



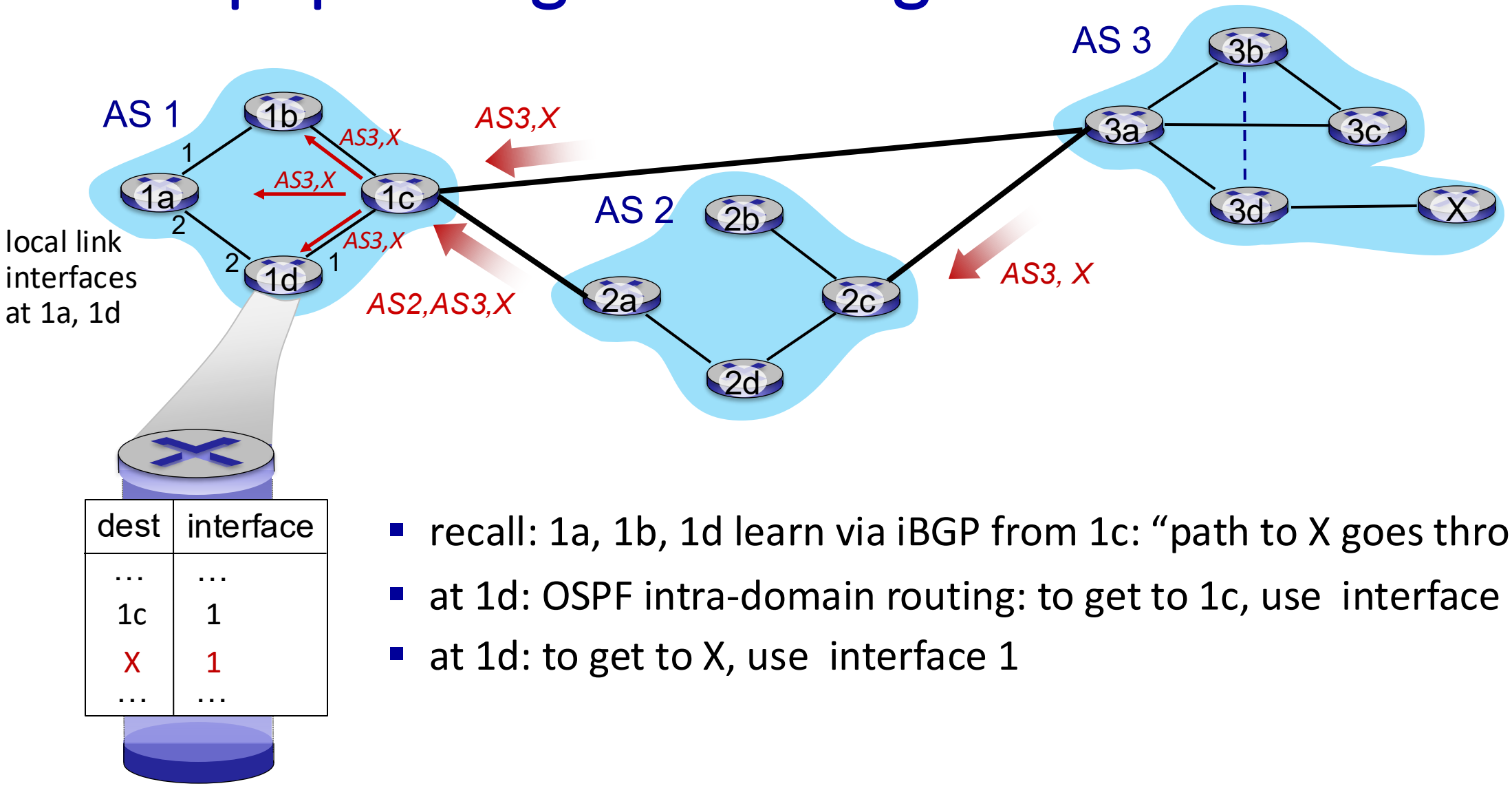
gateway router may learn about multiple paths to destination:

- AS1 gateway router 1c learns path *AS2,AS3,X* from 2a
- AS1 gateway router 1c learns path *AS3,X* from 3a
- based on *policy,* AS1 gateway router 1c chooses path *AS3,X* and advertises path within AS1 via iBGP

BGP route selection

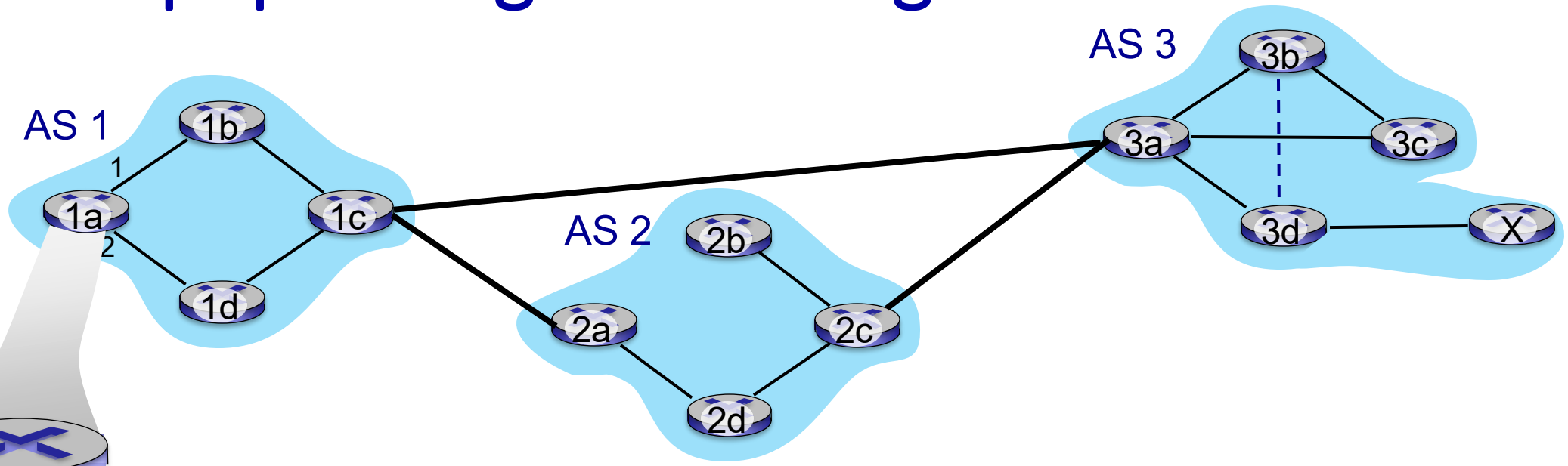
- router may learn about more than one route to destination AS, selects route based on:
  1. local preference value attribute: policy decision
  2. shortest AS-PATH
  3. closest NEXT-HOP router: hot potato routing
  4. additional criteria

BGP: populating forwarding tables



- recall: 1a, 1b, 1d learn via iBGP from 1c: “path to X goes through 1c”
- at 1d: OSPF intra-domain routing: to get to 1c, use interface 1
- at 1d: to get to X, use interface 1

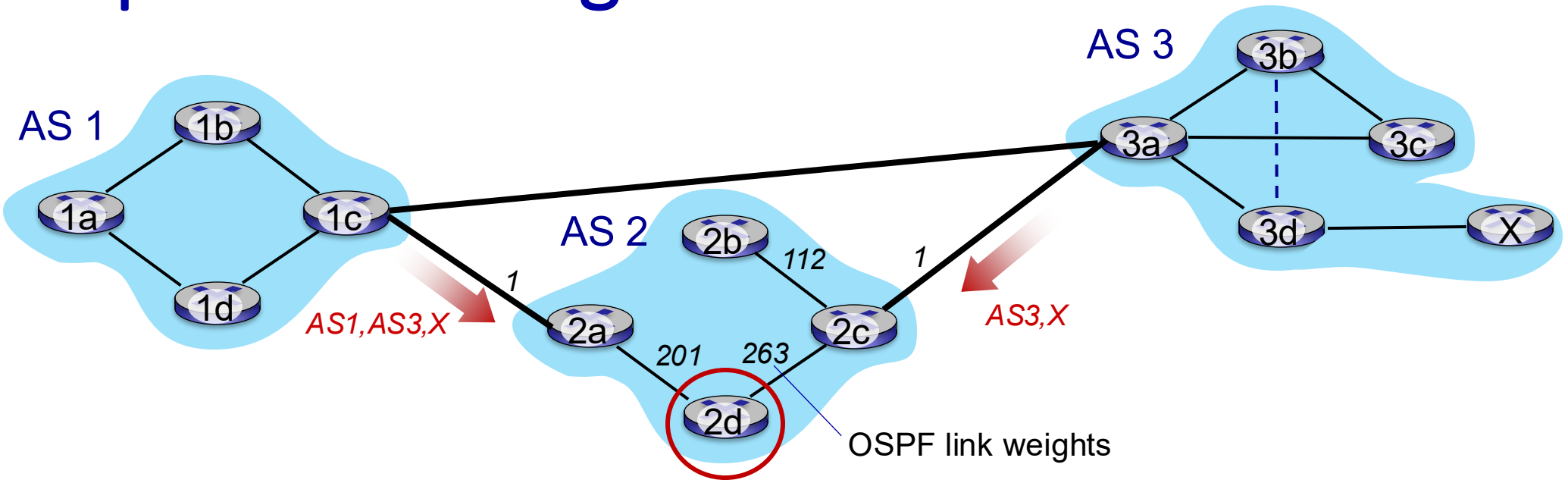
BGP: populating forwarding tables



dest	interface
…	…
1c	2
X	2
…	…

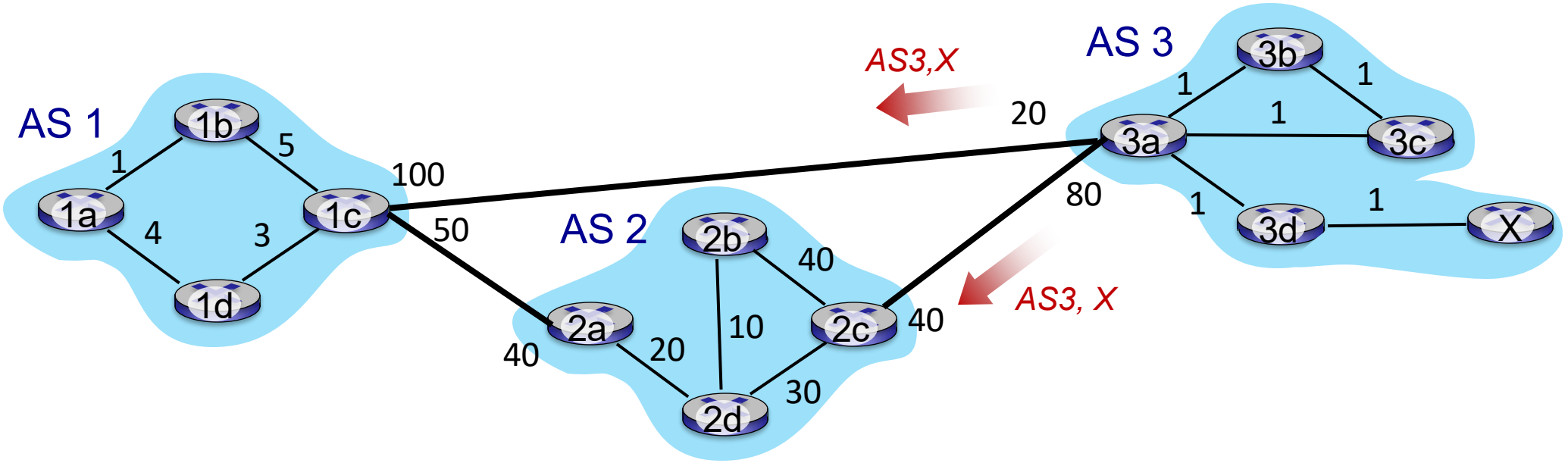
- recall: 1a, 1b, 1d learn via iBGP from 1c: “path to X goes through 1c”
- at 1d: OSPF intra-domain routing: to get to 1c, use interface 1
- at 1d: to get to X, use interface 1
- at 1a: OSPF intra-domain routing: to get to 1c, use interface 2
- at 1a: to get to X, use interface 2

Hot potato routing



- 2d learns (via iBGP) it can route to X via 2a or 2c
- hot potato routing: choose local gateway that has least *intra-domain* cost.
  - If this is the top criteria on the list, 2d will choose 2a, even though more AS hops to *X*): don't worry about how many other AS are left until the destination.

BGP Exercise - 1



- AS2 is AS1's and AS3's customer. It does not want to carry their transit traffic.
- X is connected to subnet 50.60.0.0/16.
- Gateway routers use default BGP route selection criteria list.
- Internal routers prioritize hot potato routing.
- Link weights are shown in the figure
- What are relevant entries in router 3c's forwarding table for forwarding to 50.60.0.0/16? How about for router 2b? How about for router 1a?

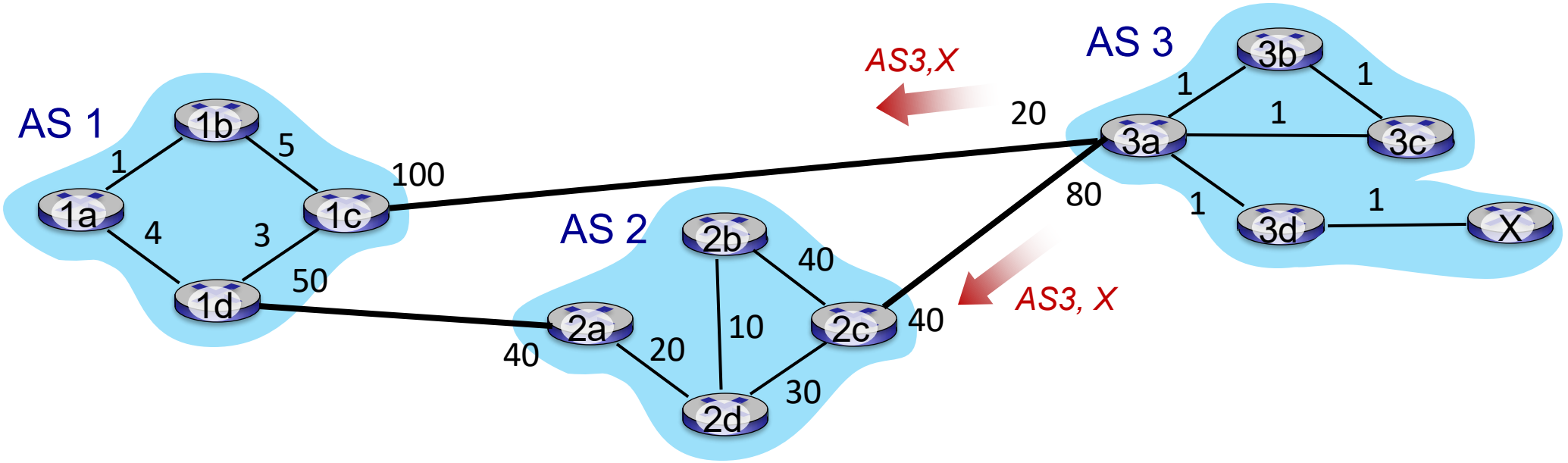
Answer

- For 3c, only intra-domain routing matters (destination subnet is within the same AS).
  - Shortest path to X is through 3a.
- For 2b
  - 2c tells 2b it can reach the subnet with: AS-PATH AS3, NEXT-HOP 3a
  - 2a tells 2b it can reach the subnet with: AS-PATH AS1, AS3, NEXT-HOP 1c
  - 2b is prioritizing hot potato routing
    - Path to 3a costs 80 (knows this through intra-domain routing)
    - Path to 1c costs 70 (knows this through intra-domain routing)
  - 2b will choose the path advertised by 2a. Next hop in 2b's forwarding table will be 2d.

Answer

- For 1a, 1c tells 1a it can reach the subnet with: AS-PATH AS3, NEXT-HOP 3a
- That's the only path to that subnet 1a knows.
- Through intra-domain routing, 1a knows its shortest path is through 1b.

BGP Exercise - 2



- ~~AS2 is AS1's and AS3's customer. It does not want to carry their transit traffic.~~
- X is connected to subnet 50.60.0.0/16.
- Gateway routers use default BGP route selection criteria list.
- Internal routers prioritize hot potato routing.
- Link weights are shown in the figure
- What are relevant entries in router 3c's forwarding table for forwarding to 50.60.0.0/16? How about for router 2b? How about for router 1a?

Why different Intra-, Inter-AS routing ?

policy:

- inter-AS: admin wants control over how its traffic routed, who routes through its network
- intra-AS: single admin, so policy less of an issue

scale:

- hierarchical routing saves table size, reduced update traffic

performance:

- intra-AS: can focus on performance
- inter-AS: policy dominates over performance

What you need to know about BGP

- How BGP routers communicate (i.e., BGP sessions)
  - e.g., over TCP, iBGP vs. eBGP
- What BGP advertisements look like
  - Destination, path, next-hop

What you need to know about BGP

- The BGP pipeline
  - Accepting/importing routes: How does the ”edge” router (gateway) decide which route to accept
  - How the gateway router propagates its information to the other routers in the AS (iBGP)
  - How internal routers decide how to pick paths to the outside destinations
  - How gateway routers decide which routes to advertise to the neighboring AS(s).

What you need to know about BGP

- How do inter- and intra-AS routing interact to populate forwarding tables
  - Specifically, given enough information, you should be able to figure out entries in the forwarding table.
- What hot-potato routing means and the rationale behind it.
- The differences between intra- and inter-AS routing
  - And the rationale behind separating them.

Routing – Wrap-up

- We are done with routing!
- Now you know how routers in the Internet find paths from a source to a destination
  - Intra- and inter-domain routing
  - Link state, distance vector, path vector
  - OSPF, BGP
  - …
- Next:
  - forwarding
  - what does inside a router look like?

Assignment 3 (Nov. 8)

- This assignment involves installing a VM and setting up some software.
- We strongly encourage you to do the setup right away
  - Even if you do not plan to start working on the coding yet.
- The assignment has four parts
  - You can get started on the early parts with what we have covered in lectures so far.
  - We will cover the rest in the next few lectures
    - SDN
    - Middleboxes
    - Link layer concepts: MAC addresses, ARP, etc.