# COMP 445
# Data Communications & Computer networks
# Winter 2022

# Network Layer – Control plane

- ✓ Introduction
- ✓ Routing algorithms
- ✓ Intra-ISP routing: OSPF
- ✓ Routing among ISPs: BGP
- ✓ SDN control plane
- ✓ Internet Control Message Protocol

# Making routing scalable

our routing study thus far:
*idealized*

- all routers identical
- network “flat”

… not true in practice

scale: billions of destinations:

- can’t store all destinations in routing tables!
- exchanging link-state or DV information would swamp links!

administrative autonomy:

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

# Internet approach to scalable routing

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

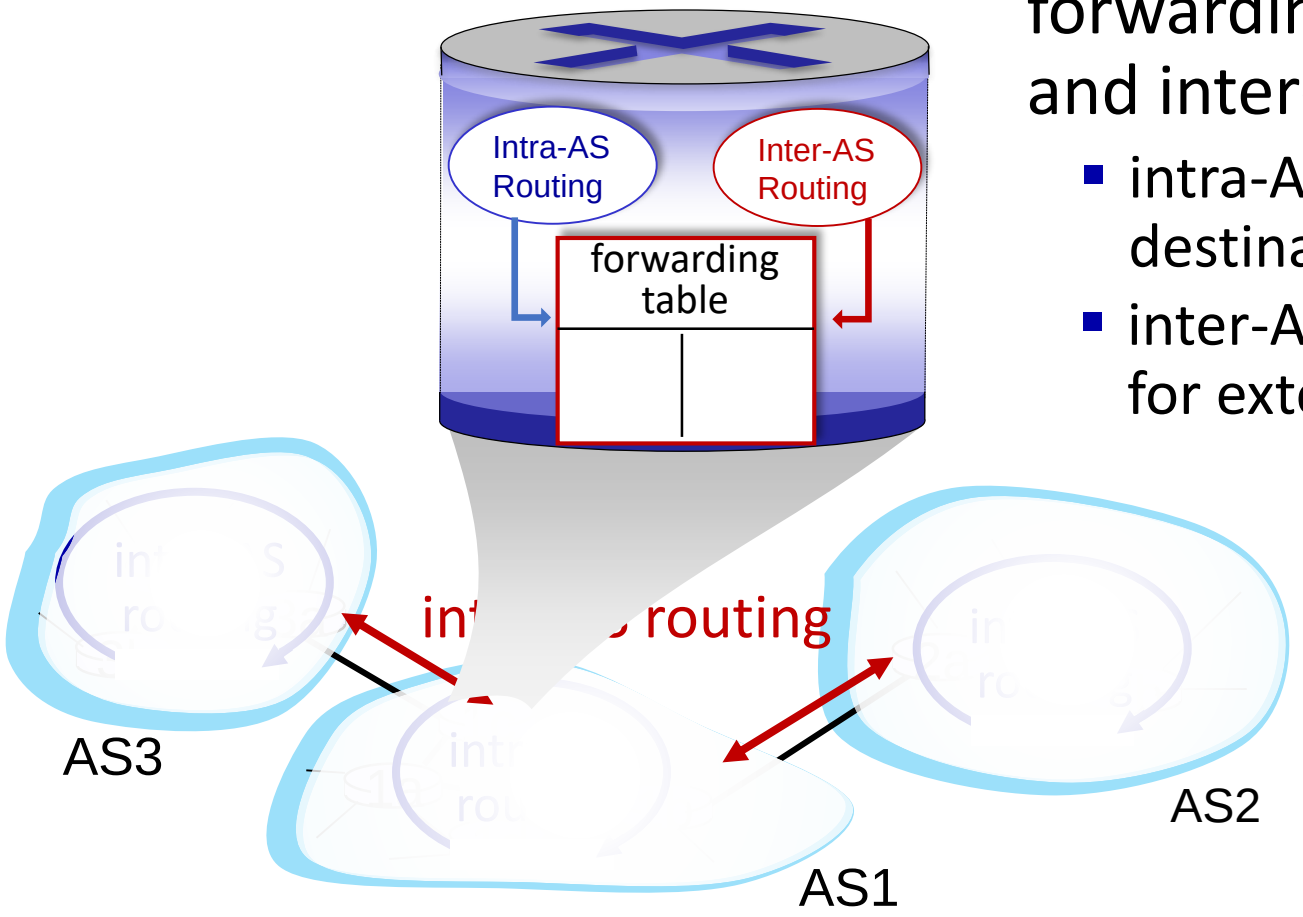
intra-AS (aka "intra-domain"): routing among routers *within same AS ("network")*

- 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

# Inter-AS routing:  routing within an AS

most common intra-AS routing protocols:

- RIP: Routing Information Protocol [RFC 1723]
  - classic DV: DVs exchanged every 30 secs
  - no longer widely used
- OSPF: Open Shortest Path First [RFC 2328]
  - classic link-state routing
  - IS-IS protocol (ISO standard, not RFC standard) essentially same as OSPF
- EIGRP: Enhanced Interior Gateway Routing Protocol
  - DV based
  - formerly Cisco-proprietary for decades
  - became open in 2013 [RFC 7868])

# OSPF (Open Shortest Path First) routing

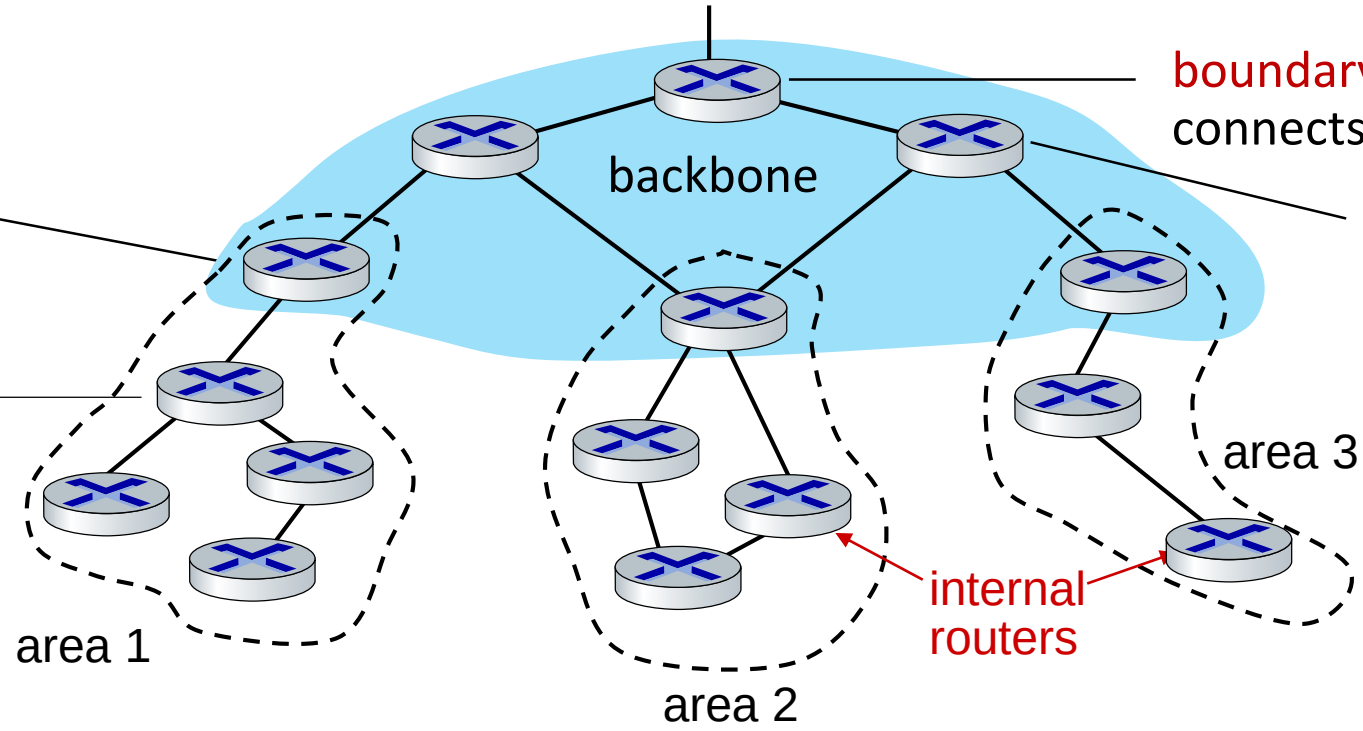
- "open": publicly available
- classic link-state
  - each router floods OSPF link-state advertisements (directly over IP rather than using TCP/UDP) to all other routers in entire AS
  - multiple link costs metrics possible: bandwidth, delay
  - each router has full topology, uses Dijkstra's algorithm to compute forwarding table
- *security:* all OSPF messages authenticated (to prevent malicious intrusion)

# Hierarchical OSPF

- two-level hierarchy: local area, backbone.
  - link-state advertisements flooded only in area, or backbone
  - each node has detailed area topology; only knows direction to reach other destinations

area border routers:
"summarize" distances to destinations in own area, advertise in backbone

local routers:
- flood LS in area only
- compute routing within area
- forward packets to outside via area border router




boundary router:
connects to other ASes

backbone router:
runs OSPF limited to backbone

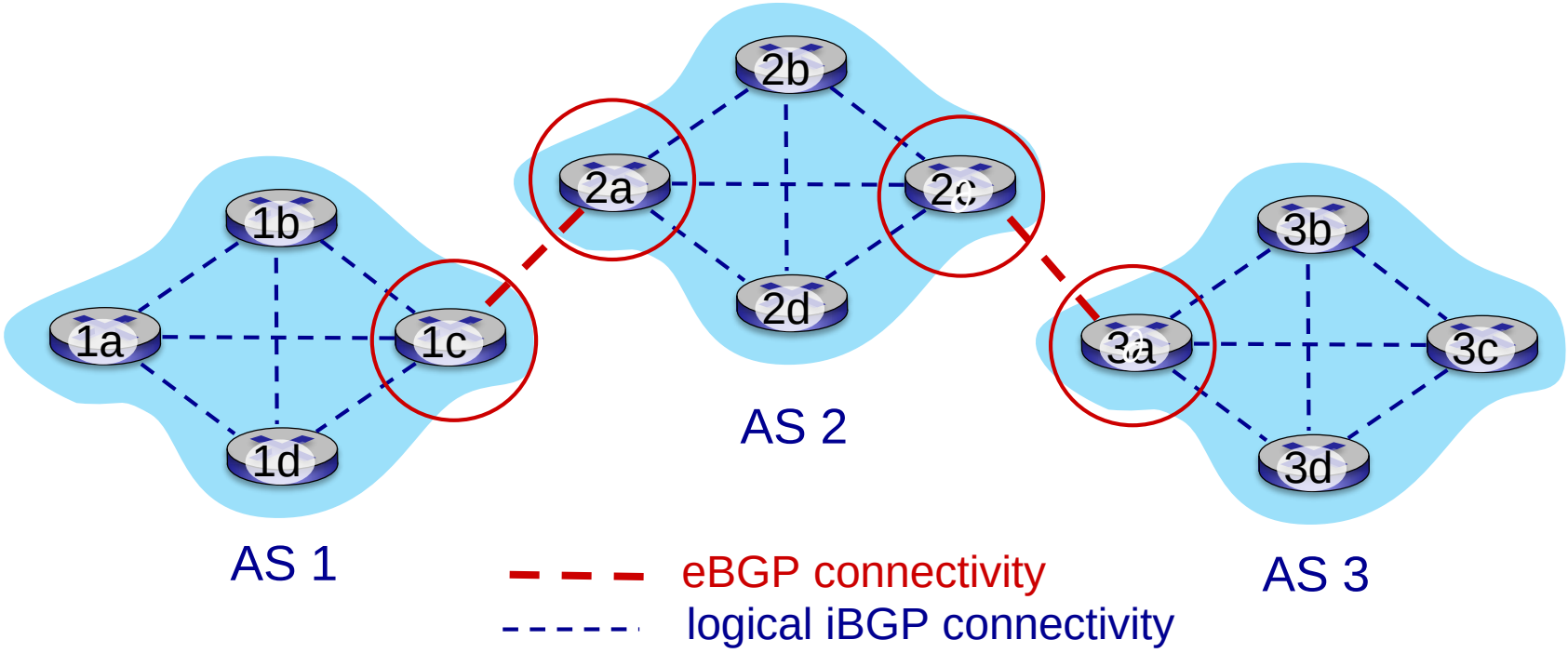
# Network Layer – Control plane

- ✓ Introduction
- ✓ Routing algorithms
- ✓ Intra-ISP routing: OSPF
- ✓ Routing among ISPs: BGP
- ✓ SDN control plane
- ✓ Internet Control Message Protocol

# 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)
  - determine routes to other networks based on reachability information and *policy*
  - propagate reachability information to all AS-internal routers (iBGP)
  - advertise (to neighboring networks) destination reachability info

# eBGP, iBGP connections

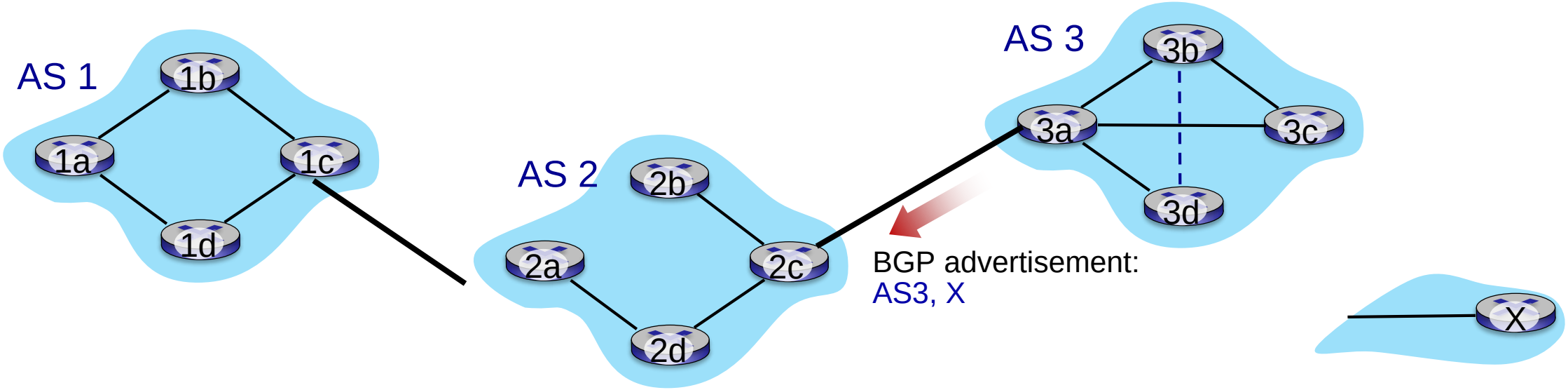




gateway routers run both eBGP and iBGP protocols

# BGP basics

- BGP session: two BGP routers ("peers, speakers") exchange BGP messages over semi-permanent TCP connection:
  - advertising *paths* to different destination network prefixes (e.g., to a destination /16 network)
  - 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



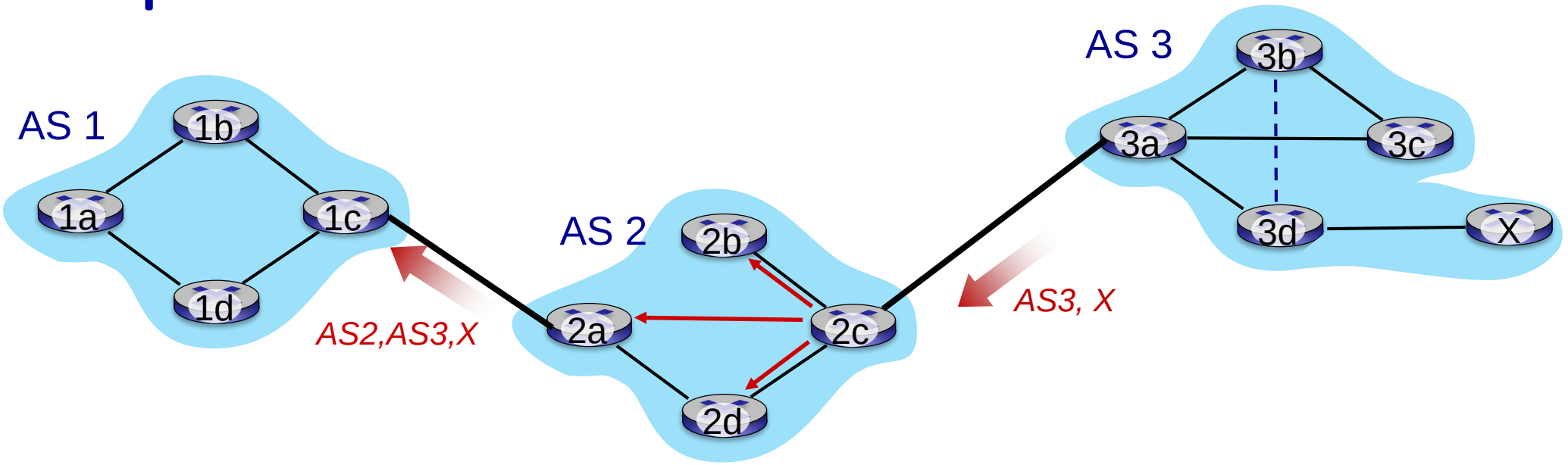
# BGP protocol messages

- BGP messages exchanged between peers over TCP connection
- BGP messages [RFC 4371]:
  - OPEN: opens TCP connection to remote BGP peer and authenticates sending BGP peer
  - **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

# Path attributes and BGP routes

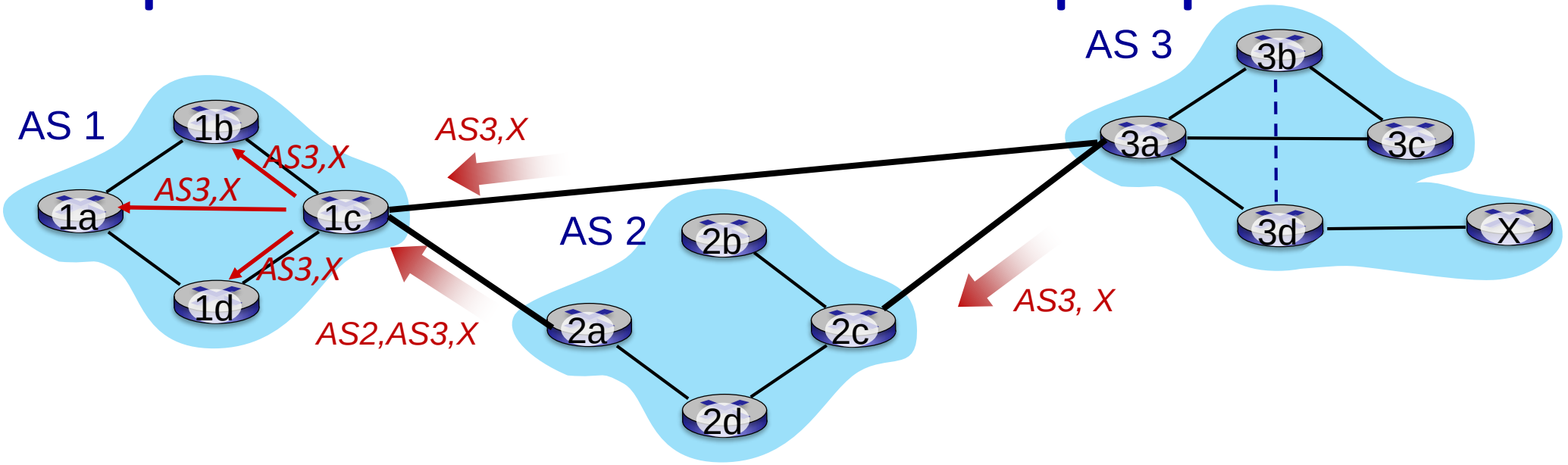
- BGP advertised path:  prefix + attributes
  - path prefix: destination being advertised
  - two important attributes:
    - AS-PATH: list of ASes through which prefix advertisement has passed
    - NEXT-HOP: indicates specific internal-AS router to next-hop AS
- policy-based routing:
  - router receiving route advertisement to destination X uses *policy* to accept/reject a path (e.g., never route through AS W, or country Y).
  - router uses policy to decide whether to *advertise* a path to neighboring AS Z (does router want to route traffic forwarded from Z destined to X?)

# 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

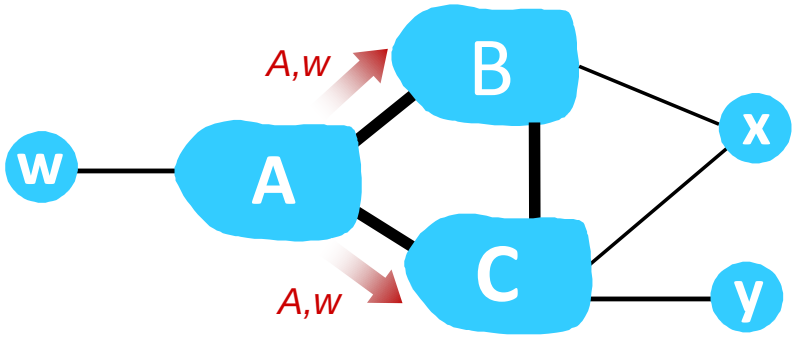
# BGP path advertisement: multiple paths

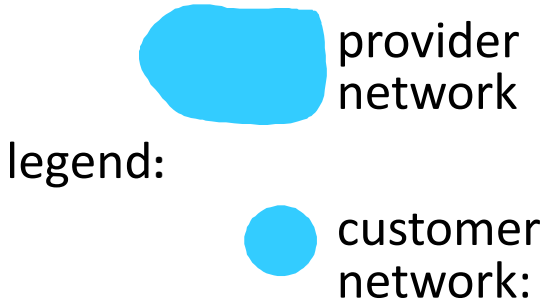



gateway routers may learn about multiple paths to destination:

- 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: achieving policy via advertisements

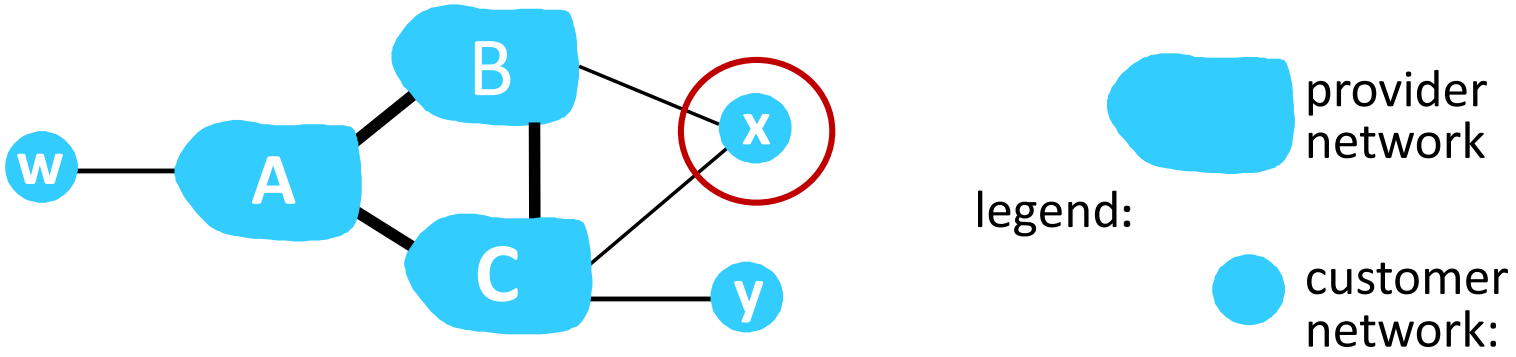



ISP only wants to route traffic to/from its customer networks (does not want to carry transit traffic between other ISPs – a typical "real world" policy)

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

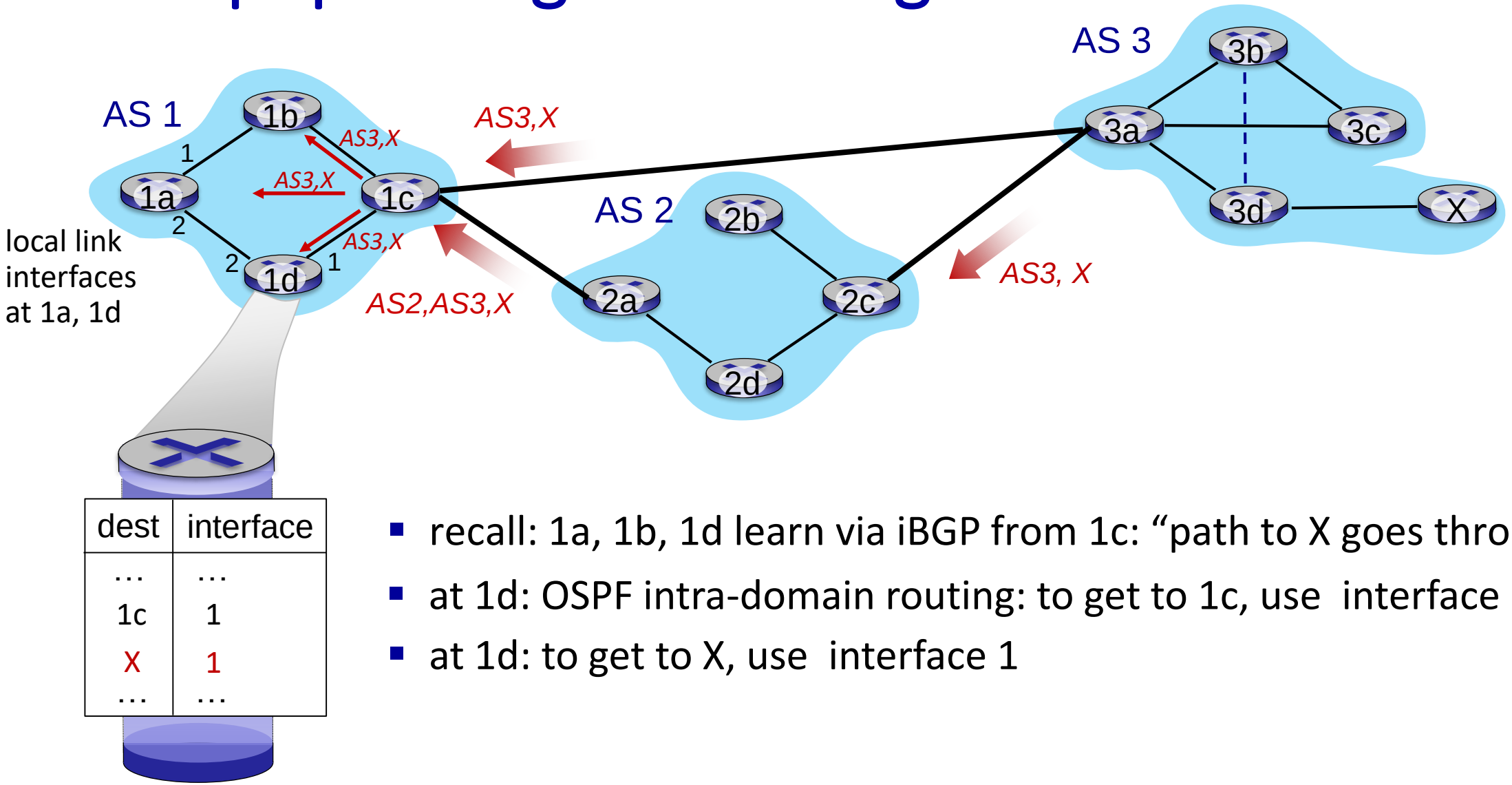
# BGP: achieving policy via advertisements (more)




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)

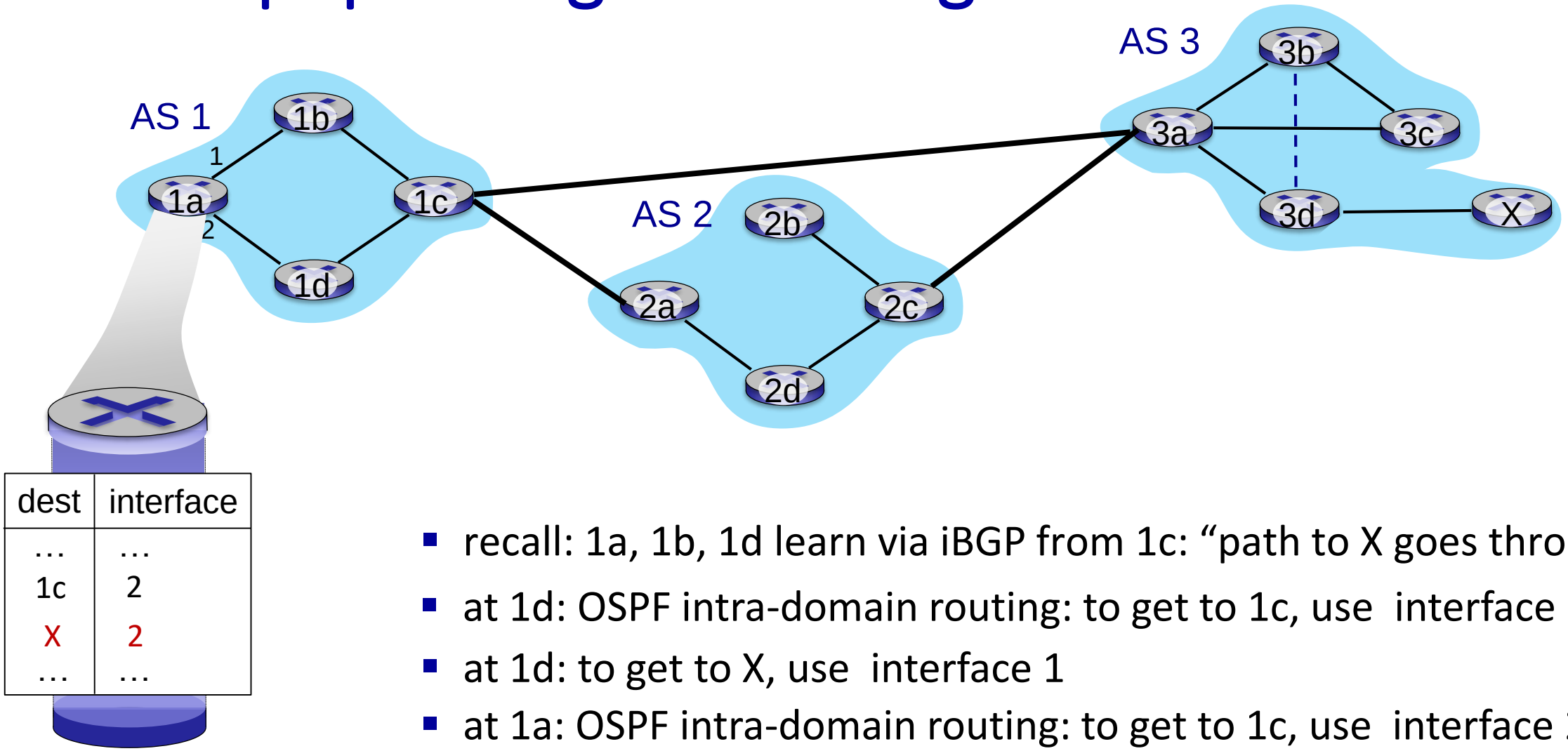
- A,B,C are provider networks
- x,w,y are customer (of provider networks)
- x is dual-homed: attached to two networks
- *policy to enforce:* x does not want to route from B to C via x
  - .. so x will not advertise to B a route to C

# BGP: populating forwarding tables



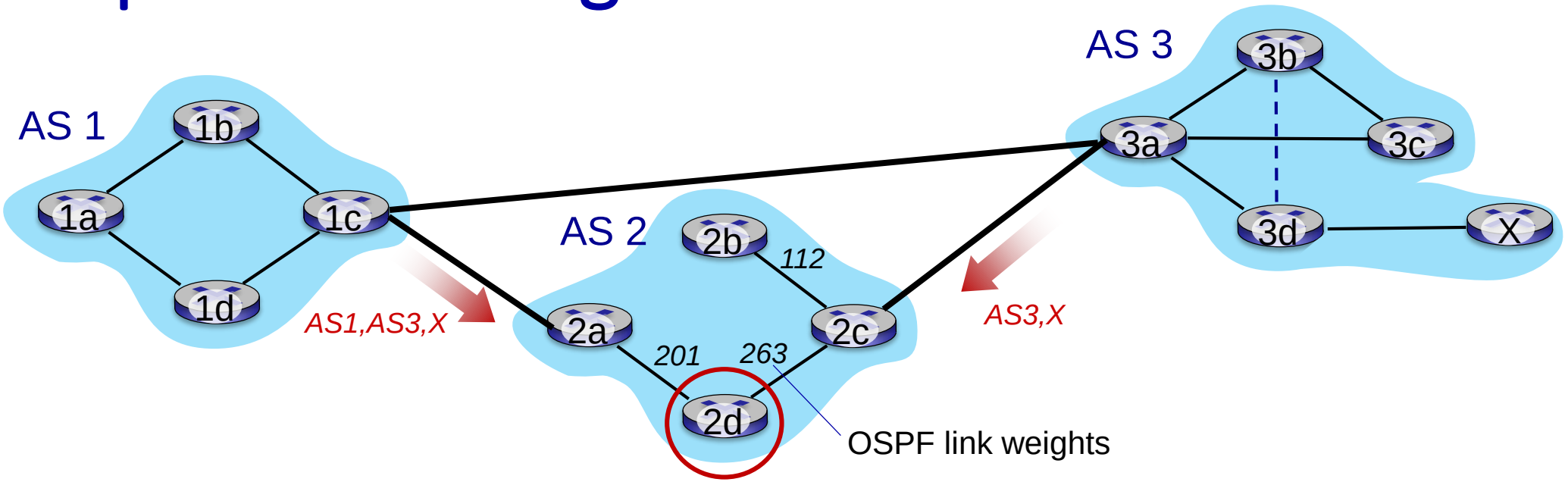

- recall: 1a, 1b, 1d learn via iBGP from 1c: “path to X goes through 1c”
- at 1d: OSPF intra-domain routing: to get to 1c, use interface 1
- at 1d: to get to X, use interface 1

# BGP: populating forwarding tables




- 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 (e.g., 2d chooses 2a, even though more AS hops to *X*): don't worry about inter-domain cost!

# 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: reducing forwarding table size, routing update traffic

- hierarchical routing: limiting the scope of full topological information
- BGP routing to CIDRized destination networks (summarized routes)

performance:

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

# References

Figures and slides are taken/adapted from:

- Jim Kurose, Keith Ross, "Computer Networking: A Top-Down Approach", 7th ed. Addison-Wesley, 2012. All material copyright 1996-2016 J.F Kurose and K.W. Ross, All Rights Reserved

- Jim Kurose, Keith Ross, "Computer Networking: A Top-Down Approach", 8th ed. Pearson, 2020. All material copyright 1996-2020 J.F Kurose and K.W. Ross, All Rights Reserved