Routing in the Internet

To do …

- Intra-AS routing and OSPF
- Inter-AS routing and the Border Gateway Protocol

Slides partially based on J. Kurose and K. Ross, "Computer Networking: A Top Down Approach", 7Ed., 2016

Northwestern

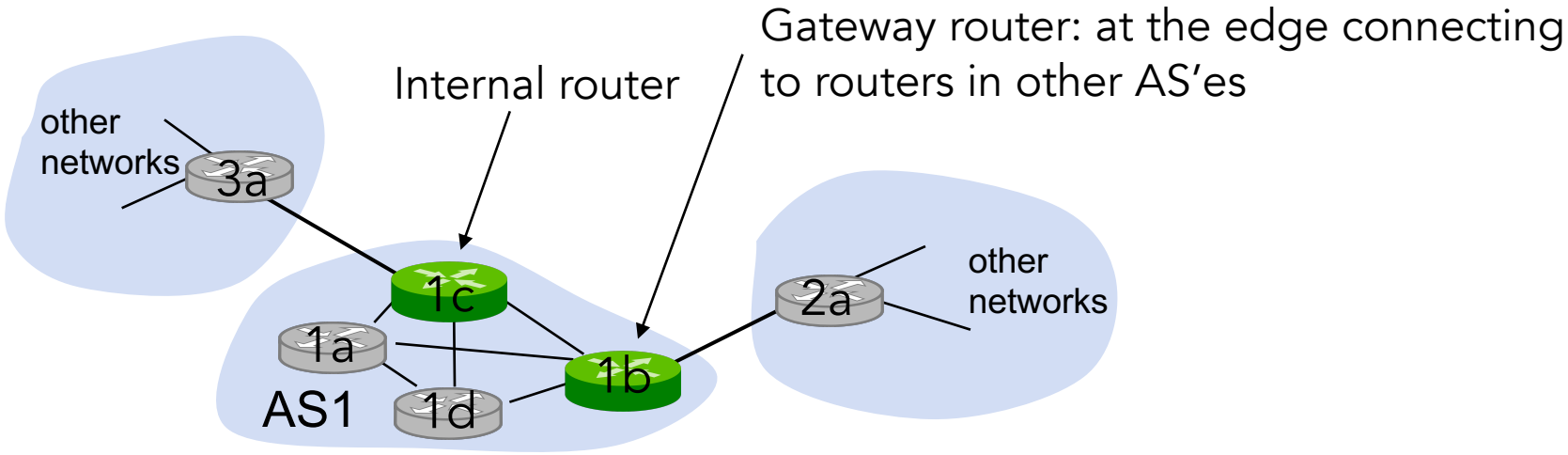
Routing in theory and in practice

- Our routing study thus far – Idealized
  - All routers identical
  - Network "flat"

- … in practice
  - Needs to scale to billions of destinations – Can't store all destinations in routing tables; routing table exchange would swamp links!
  - Need administrative autonomy – The Internet is a network of networks and each admin wants control of its own

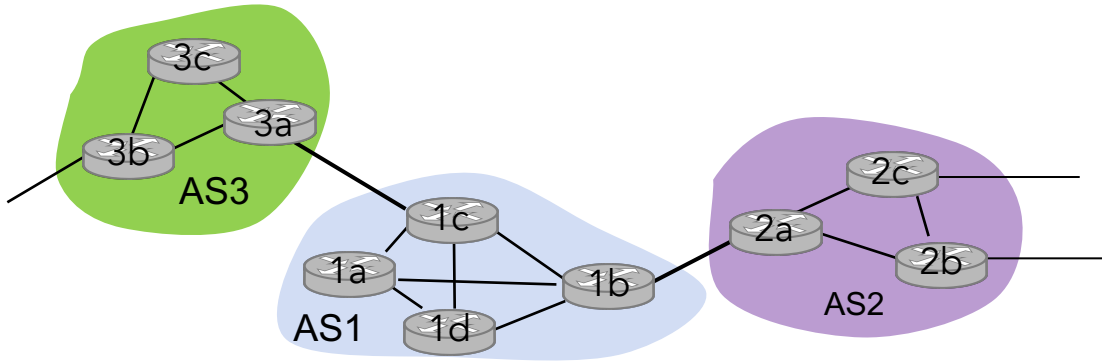
Autonomous systems

- The Internet is divided into autonomous systems
  - Over 92,000 in August 2019 (30% in the US, next Brazil with ~7% …)
  - Each has an AS number, distributed by the ICANN's regional authorities
- Gateway and internal routers
  - Gateway routers at edge of the AS connect to other AS's
  - Routers within an AS run one intra-AS routing protocol



Internet approach to scalable routing

- Intra-AS routing
  - Routing among hosts, routers in same AS ("network")
  - All routers in AS must run same intra-domain protocol
  - Routers in different AS can run different intra-domain routing protocol
  - Gateway router: at "edge" of AS, has link(s) to router(s) in other AS'es
- Inter-AS routing
  - Routing among AS'es
  - Gateways perform inter-domain routing (besides intra-domain routing)



Intra-AS Routing

- Also known as interior gateway protocols (IGP)
- Some common intra-AS routing protocols
  - RIP: Routing Information Protocol
    - The oldest one, started to be implemented in 1969 for ARPANET and CYCLADES; in 1982 was included in Unix BSD which became the basis of many Unix versions
  - OSPF: Open Shortest Path First (IS-IS essentially same as OSPF)
    - *We'll discuss this as example*
  - IGRP: Interior Gateway Routing Protocol
    - Cisco proprietary, until 2016

OSPF (Open Shortest Path First)

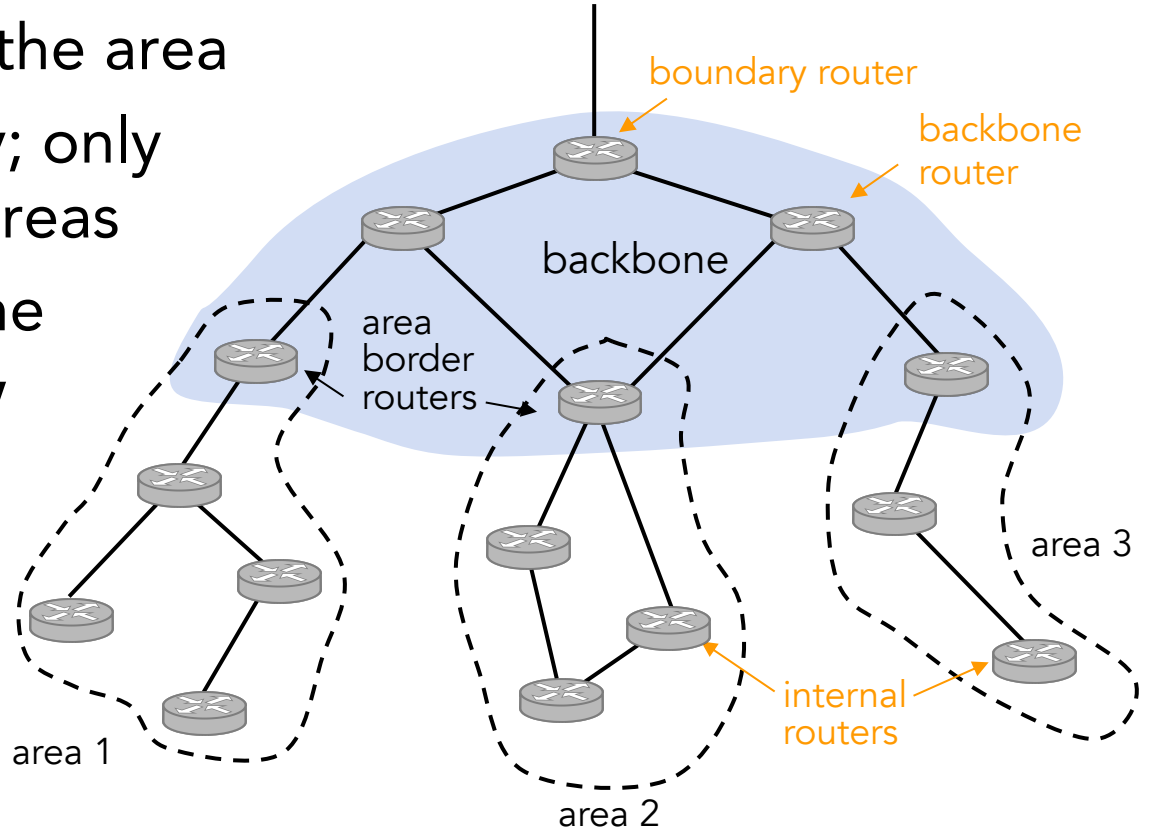
- Open (publicly available) and widely used
- Uses link-state algorithm
  - Link state packet dissemination (over IP)
  - Topology map of the entire AS at each node
  - Route computation using Dijkstra's algorithm
  - Individual link costs set by network administrator (cause and effect)
- Router floods OSPF link-state ads to all other routers in the AS
  - At least every 30 secs even if nothing has changed
  - Carried in OSPF messages directly over IP (rather than TCP or UDP)
  - Link state: for each attached link
- IS-IS routing protocol – nearly identical to OSPF

Some OSPF "advanced" features

- Security - OSPF msgs between routers can be authenticated
  - Simple authentication where routers share a password (not much!)
  - MD5 – Compute hash of a msg content + key and send (content, hash); routers share the key
- Multiple same-cost paths allowed
- For each link, multiple cost metrics for different ToS (e.g., sat. link cost set low for best effort ToS; high for real-time ToS)
- Integrated uni- and multi-cast support
  - Multicast OSPF (MOSPF) uses same topology data base as OSPF
  - Adds a new type of link-state advertisement
- Hierarchical OSPF in large domains …

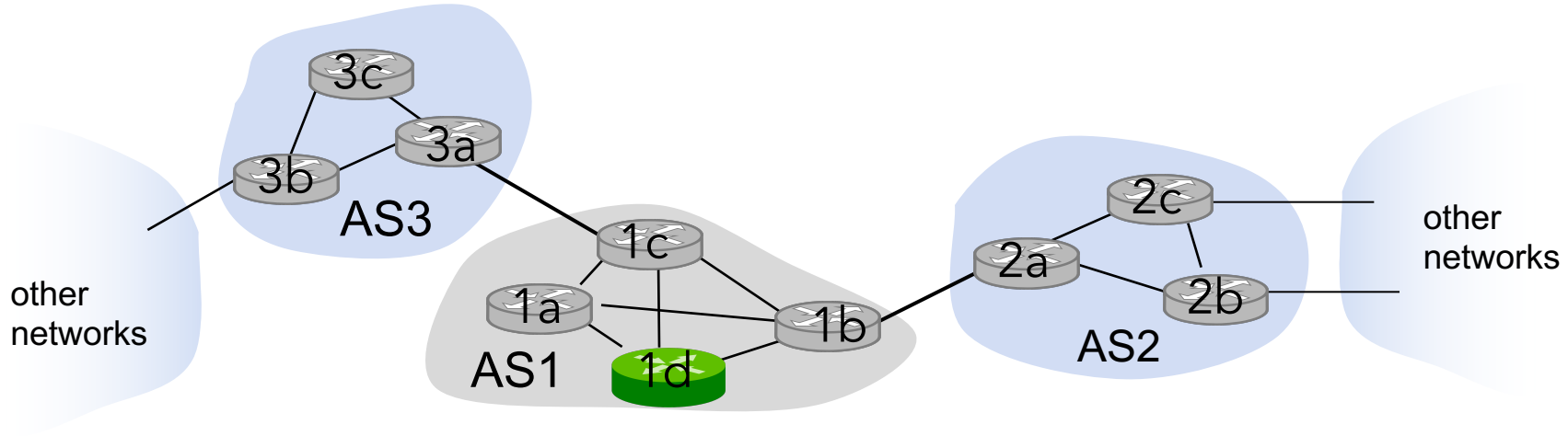
Hierarchical OSPF

- Two-level hierarchy: local area, backbone
  - Link-state advertisements only within the area
  - Each node has detailed area topology; only knows shortest path to nets in other areas
  - Only one OSPF area in AS as backbone
- Area border routers – "summarize" distances to nets in own area, advertise to other ABR
- Backbone routers – run OSPF routing limited to backbone
- Boundary … connect to other AS'es



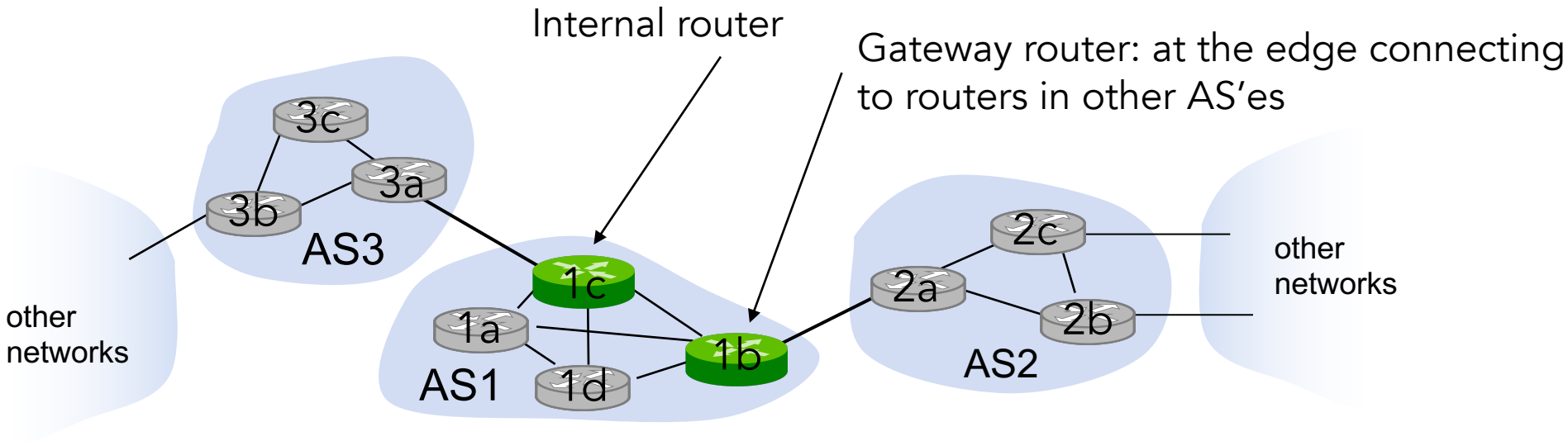
Interconnected ASes

- If router in AS1 receives datagram destined outside of AS1
  - Router should forward packet to gateway router, but which one?
- AS1 must
  - Learn which destinations are reachable through AS2, which through AS3
  - Propagate this reachability info to all routers in AS1



Interconnected ASes

- Every router has a forwarding table
- Table configured by both intra- and inter-AS routing algorithms
  - Intra-AS routing determine entries for destinations within AS
  - Inter-AS & intra-AS determine entries for external destinations



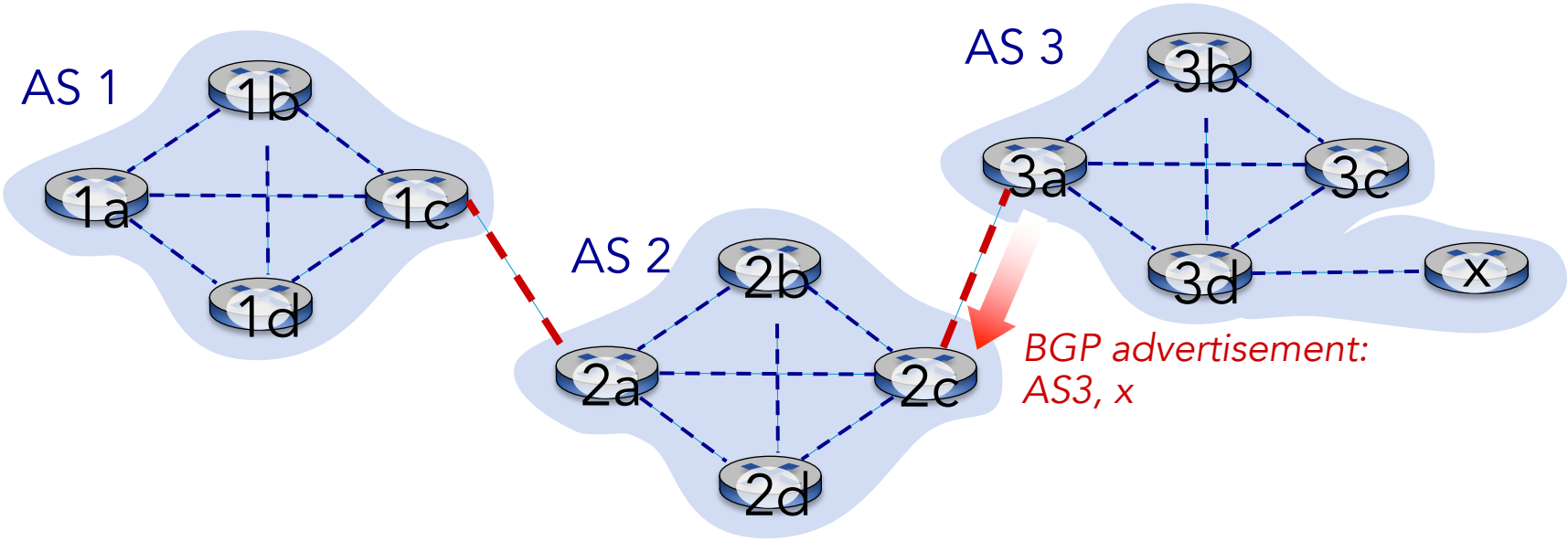
Three AS'es

Internet inter-AS routing: BGP

- BGP (Border Gateway Protocol): the de facto inter-domain routing protocol
  - The glue that holds the Internet together
- BGP provides each AS a means to
  - Obtain subnet reachability information from neighboring AS's
  - Propagate reachability information to all AS-internal routers
  - Determine "good" routes to other networks based on reachability information and policy
- Allows subnet to advertise its existence to rest of the Internet

BGP basics

- BGP session – two BGP routers ("peers") exchange BGP msgs
  - Advertising paths to different destination 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



BGP messages

- BGP messages exchanged between peers over TCP connection
- BGP messages
  - 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

- Advertised prefix includes BGP attributes
  - prefix + attributes = "route"
- Two important attributes
  - AS-PATH: list of ASes through which prefix advertisement has passed
    - Why a list? Loop detection
  - NEXT-HOP: indicates specific internal-AS router to next-hop AS

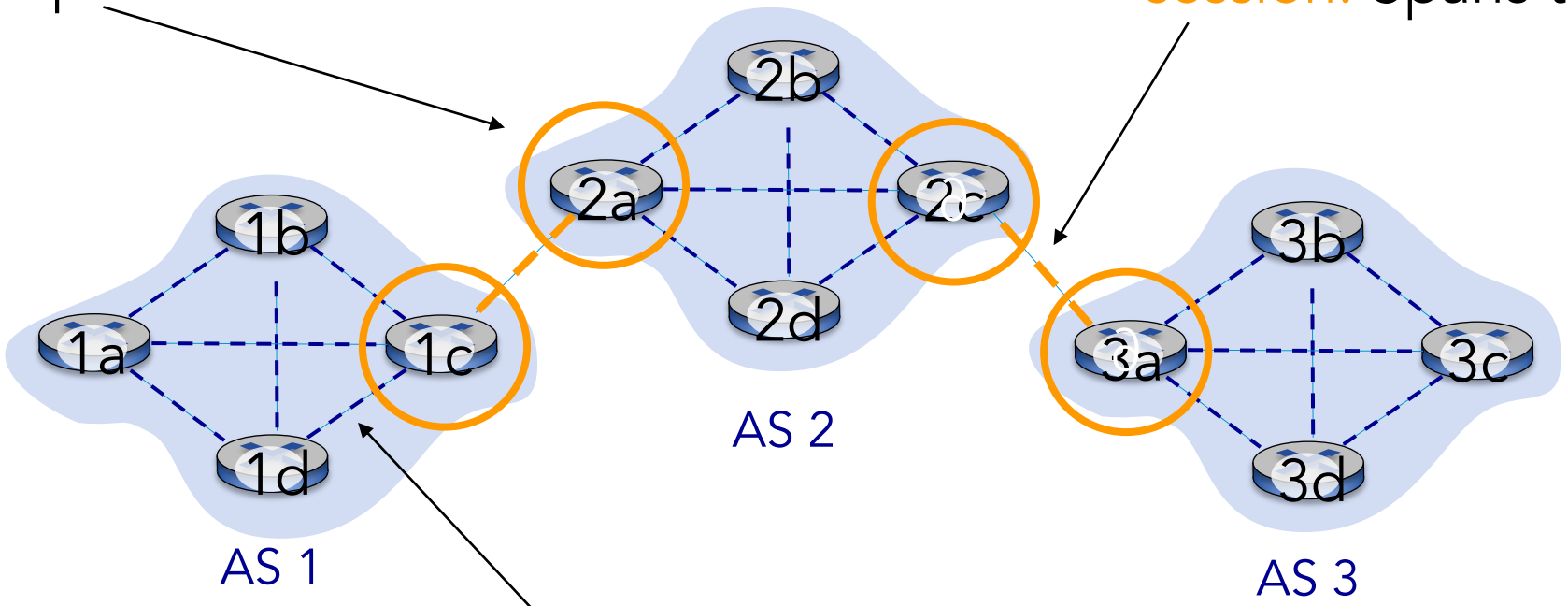
{PREFIX: 43.5.0.0/16, AS-PATH: [AS4, AS65, AS1], NEXT-HOP: 5.6.7.200)}

- Above, a router in AS4 is advertising:
  - *You can send traffic to 43.5.0.0/16 through my router 5.6.7.200, and it will travel through three AS's to get there*

eBGP, iBGP connections

Gateway routers run both eBGP and iBGP protocols

eBGP connection or session: Spans two ASs

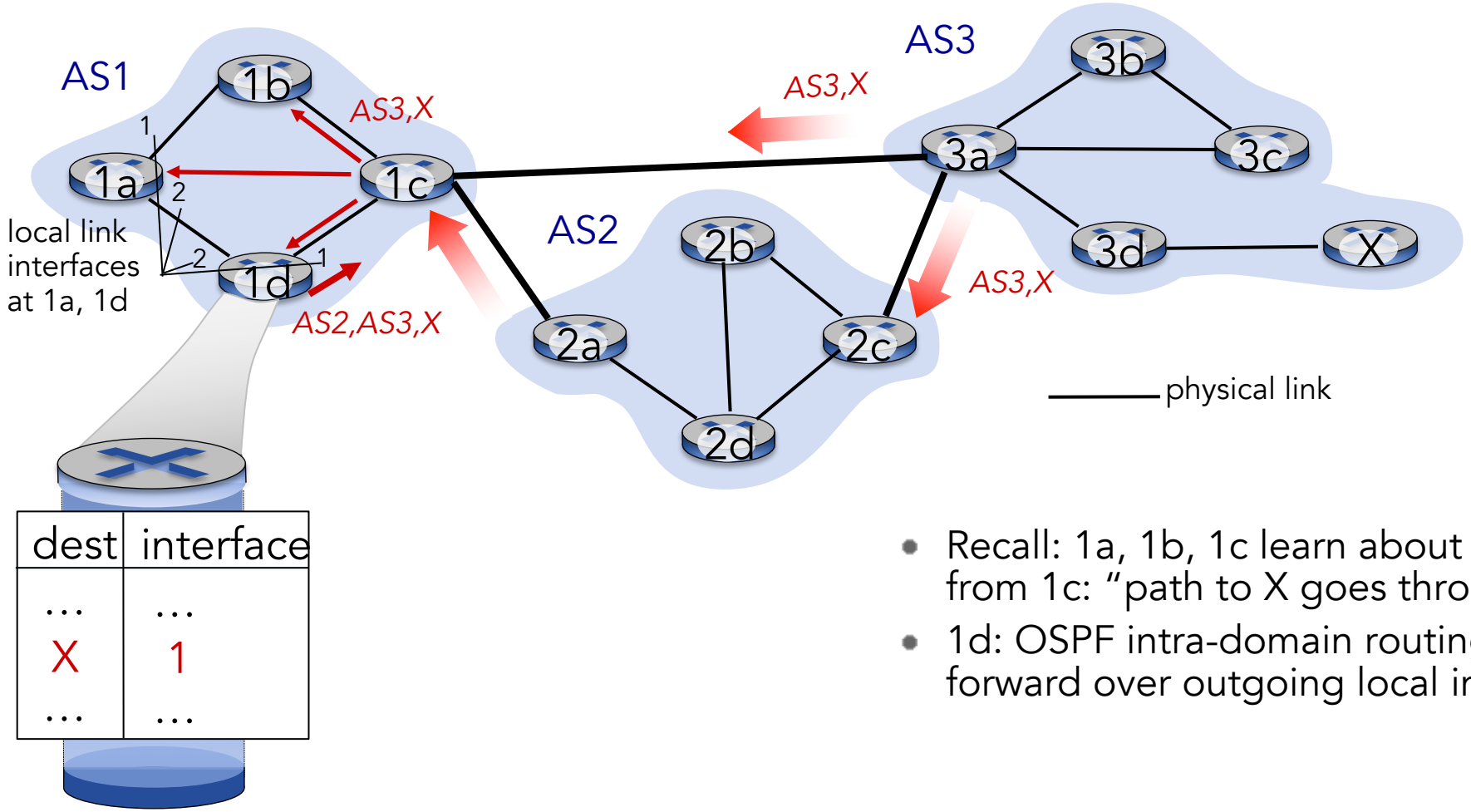


iBGP connection: A BGP session between routers in the same AS

A common configuration

BGP, OSPF, forwarding table entries

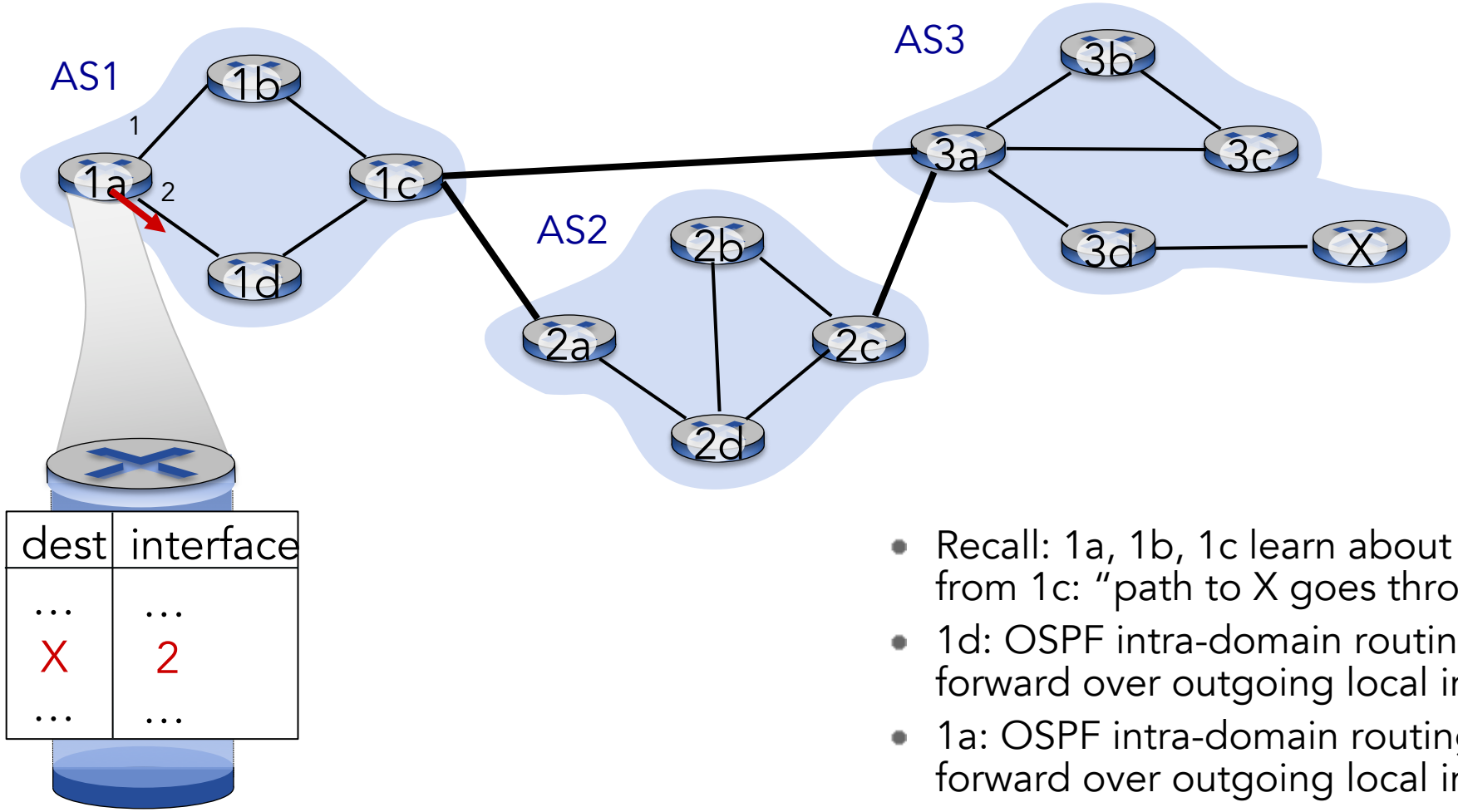
How does router set forwarding table entry to distant prefix?



- Recall: 1a, 1b, 1c learn about dest X via iBGP from 1c: "path to X goes through 1c"
- 1d: OSPF intra-domain routing: to get to 1c, forward over outgoing local interface 1

BGP, OSPF, forwarding table entries

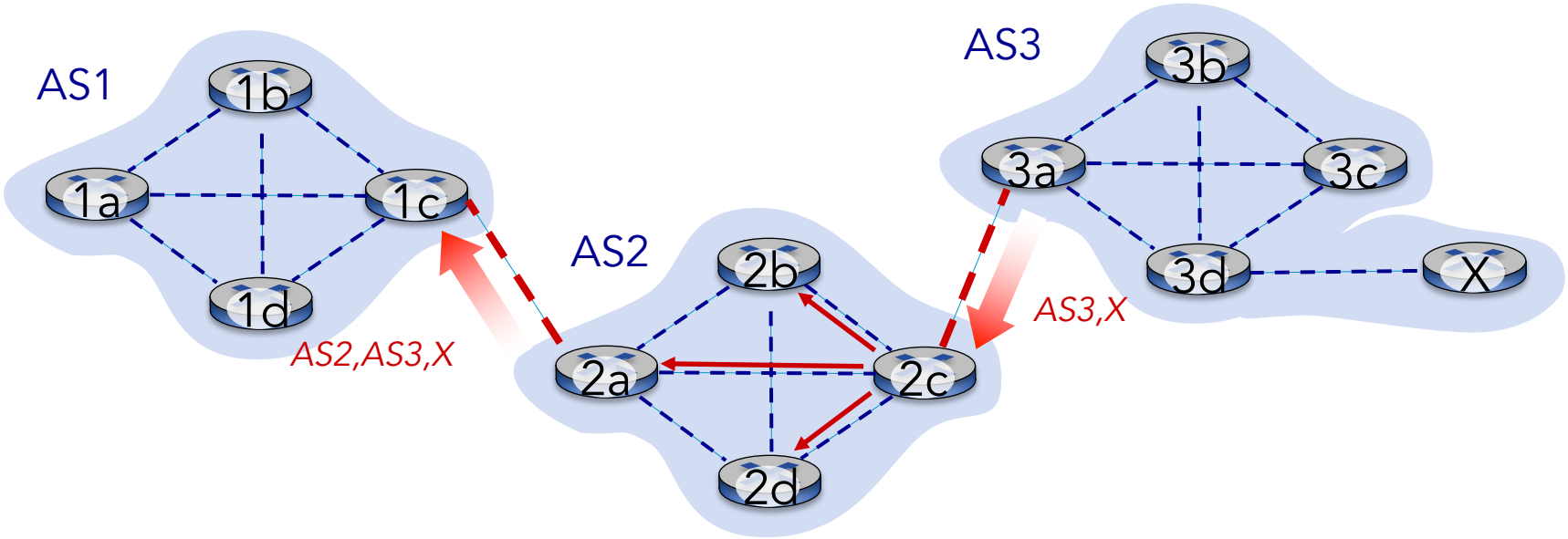
How does router set forwarding table entry to distant prefix?



- Recall: 1a, 1b, 1c learn about dest X via iBGP from 1c: "path to X goes through 1c"
- 1d: OSPF intra-domain routing: to get to 1c, forward over outgoing local interface 1
- 1a: OSPF intra-domain routing: to get to 1c, forward over outgoing local interface 2

BGP path advertisement

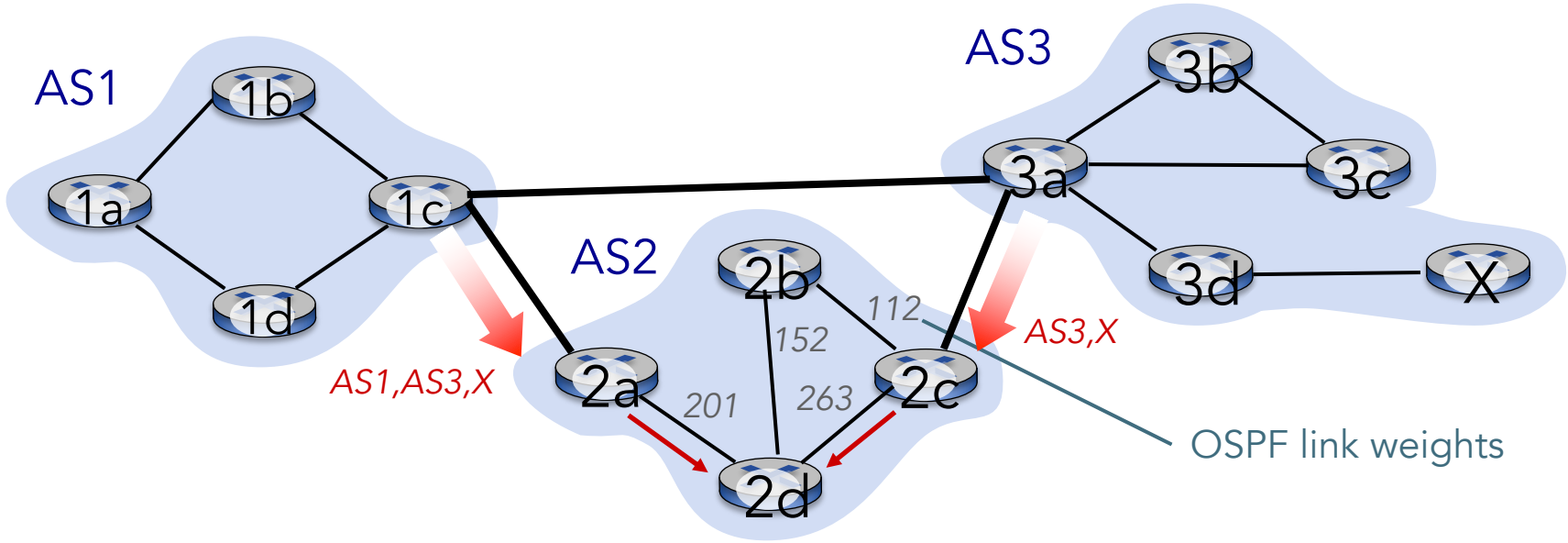
- AS2 2c gets path advertisement AS3,X (via eBGP) from AS3 3a
- Based on AS2 policy, AS2 2c accepts path AS3,X, propagates (via iBGP) to all AS2 routers
- Based on AS2 policy, AS2 2a advertises (via eBGP) path AS2, AS3, X to AS1 1c



Path attributes and BGP routes

- 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 other neighboring Ases
- Route selection
  - AS policy determines local preference for various routes (set a-priori based on financial cost, agreements, etc)
  - Among routes with the highest local preference, choose route with shortest AS-PATH
  - Multiple options remain? *Hot-potato routing*, that is, choose the route whose NEXT-HOP is closest (based on IGP like OSPF).
  - Still options? Use a random tie-breaker (eg., BGP identifier)

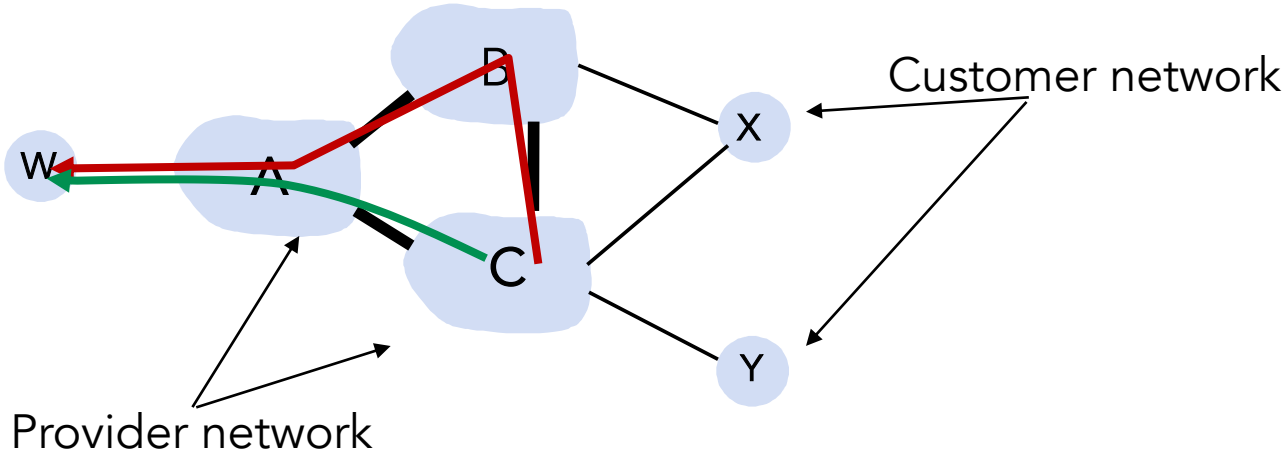
Hot Potato Routing



- 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!
- Potential value of detouring

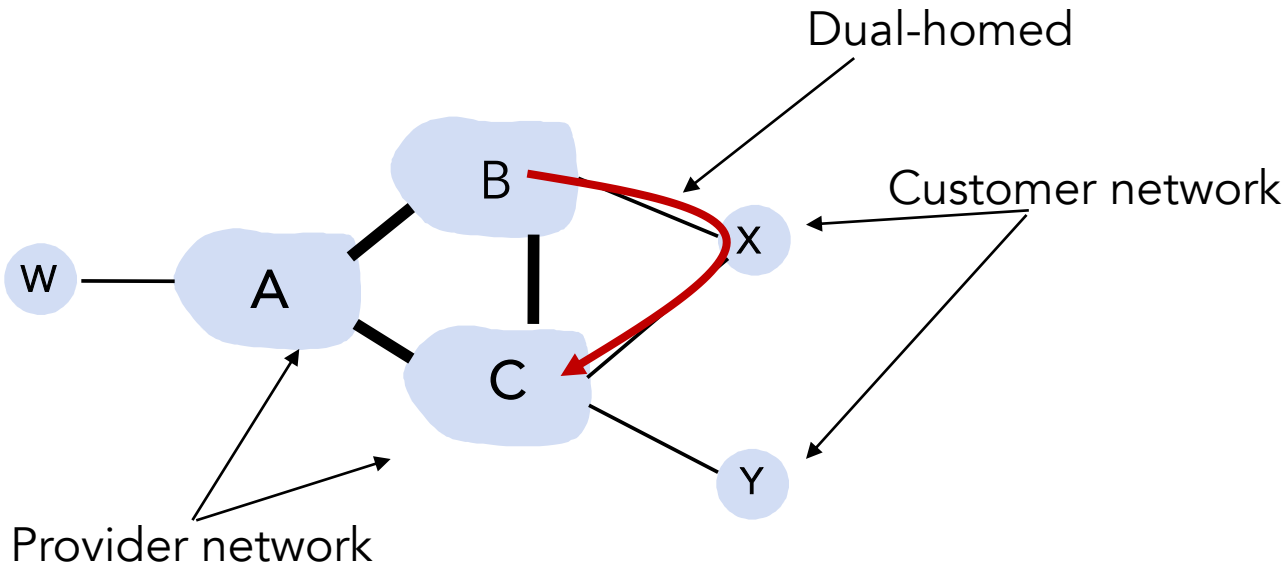
BGP: achieving policy via advertisements

- Suppose an ISP only wants to route traffic to/from its customer networks (does not want to carry transit traffic between other ISPs)
- 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

- X is multihoming (reliability?)
- X does not want to route from B to C via X
  .. so X will not advertise to B a route to C

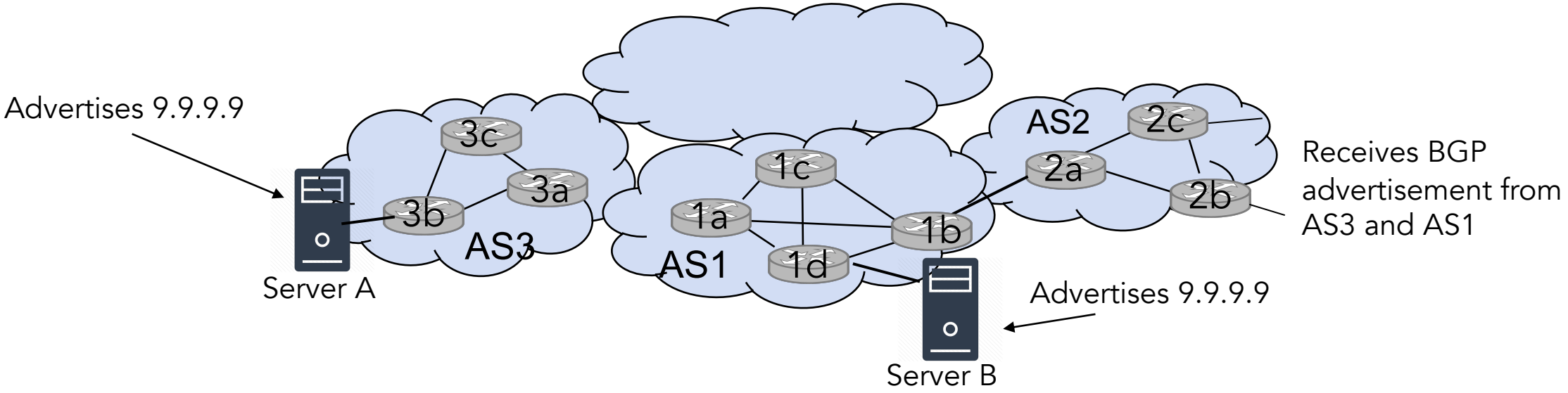


Why different Intra-, Inter-AS routing ?

- Policy
  - Inter-AS: Admin wants control over how their traffic routed, who routes through their network
  - 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

Besides inter-AS routing – IP Anycast

- Commonly used in (root, open) DNS, some CDNs (Edgecast)
- Idea
  - Replicas serve content from different geographic sites with same IP
  - Different routes to the address are announced through BGP
  - Routers consider these to be alternative routes to the same destination, although they are routes to different destinations with the same address
  - As usual, routers select a route by whatever distance metric is in use



IP Anycast

- Advantages
  - Fast (close by, commonly)
  - Resilience (if a server is down, the request gets routed to another one)
  - Attack mitigation
- Disadvantages
  - PoP switch – Mid-connection changing the routing → Use with UDP (connectionless) or with stateless service
  - Inter-domain routing is not guaranteed to be optimal in terms of bandwidth, latency or geographic proximity (at best, connectivity and policy) → Same upstream provider for all?

Recap

- IP addressing allows end-to-end communication on Internet
- Routers have forwarding tables determining which outbound link to forward packets, based on destination IP address
- Routing algorithms define forwarding tables
- Routing is hierarchical
  - IGP (eg., OSPF) determines optimal routes within an AS
    - Can use a centralized (Link State) shortest path algorithm, like Dijkstra's
  - BGP determines routes between AS's
    - Uses a distributed shortest-AS-hop path algorithm (Distance Vector)