Network Layer: Routing Protocols: OSPF & BGP

A/PROF. DUY NGO

Learning Objectives

5.3 intra-AS routing in the Internet: OSPF

5.4 routing among the ISPs: BGP

Making Routing Scalable

our routing study thus far - idealized

- all routers identical
- network “flat”

… **not** true in practice

scale: with billions of destinations:

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

administrative autonomy

- internet = 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 routing

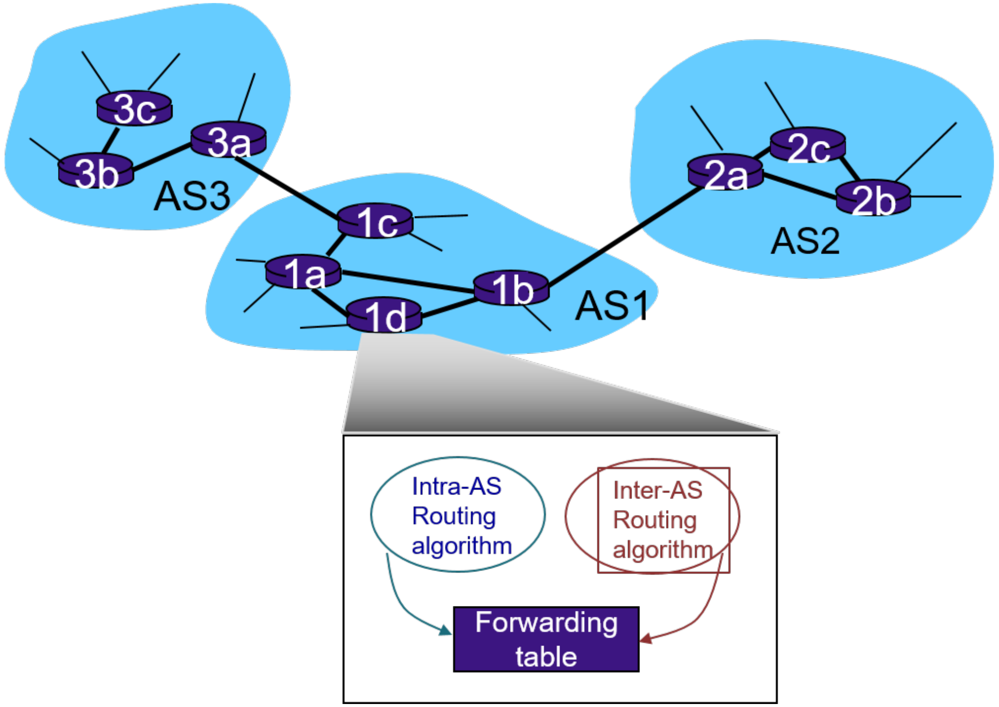
- routing among hosts, routers in same AS ("network")
- all routers in AS must run **same** intra-domain protocol
- routers in **different** AS can run **different** intra-domain routing protocol
- gateway router: at "edge" of its own AS, has link(s) to router(s) in other AS'es

inter-AS routing

- routing among AS'es
- gateways perform inter-domain routing (as well as intra-domain routing)

Interconnected ASes

- forwarding table configured by both intra- and inter-AS routing algorithm
  - intra-AS routing determine entries for destinations within AS
  - inter-AS & intra-AS determine entries for external destinations



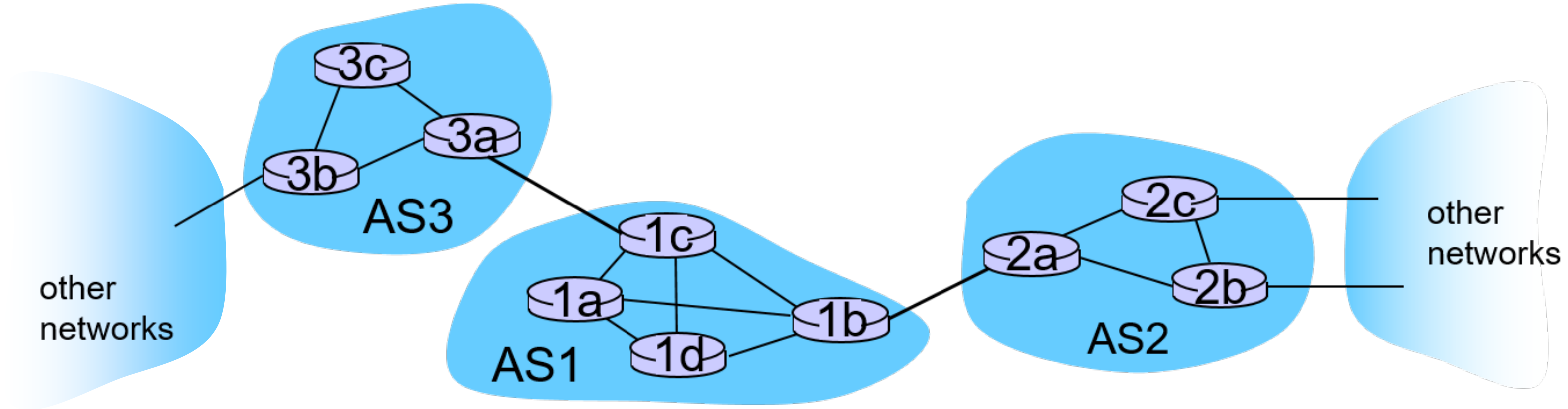
Inter-AS Tasks

- 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!



Intra-AS Routing

- also known as **interior gateway protocols (IGP)**
- most common intra-AS routing protocols:
  - RIP: Routing Information Protocol
  - OSPF: Open Shortest Path First (IS-IS protocol essentially same as OSPF)
  - IGRP: Interior Gateway Routing Protocol (Cisco proprietary for decades, until 2016)

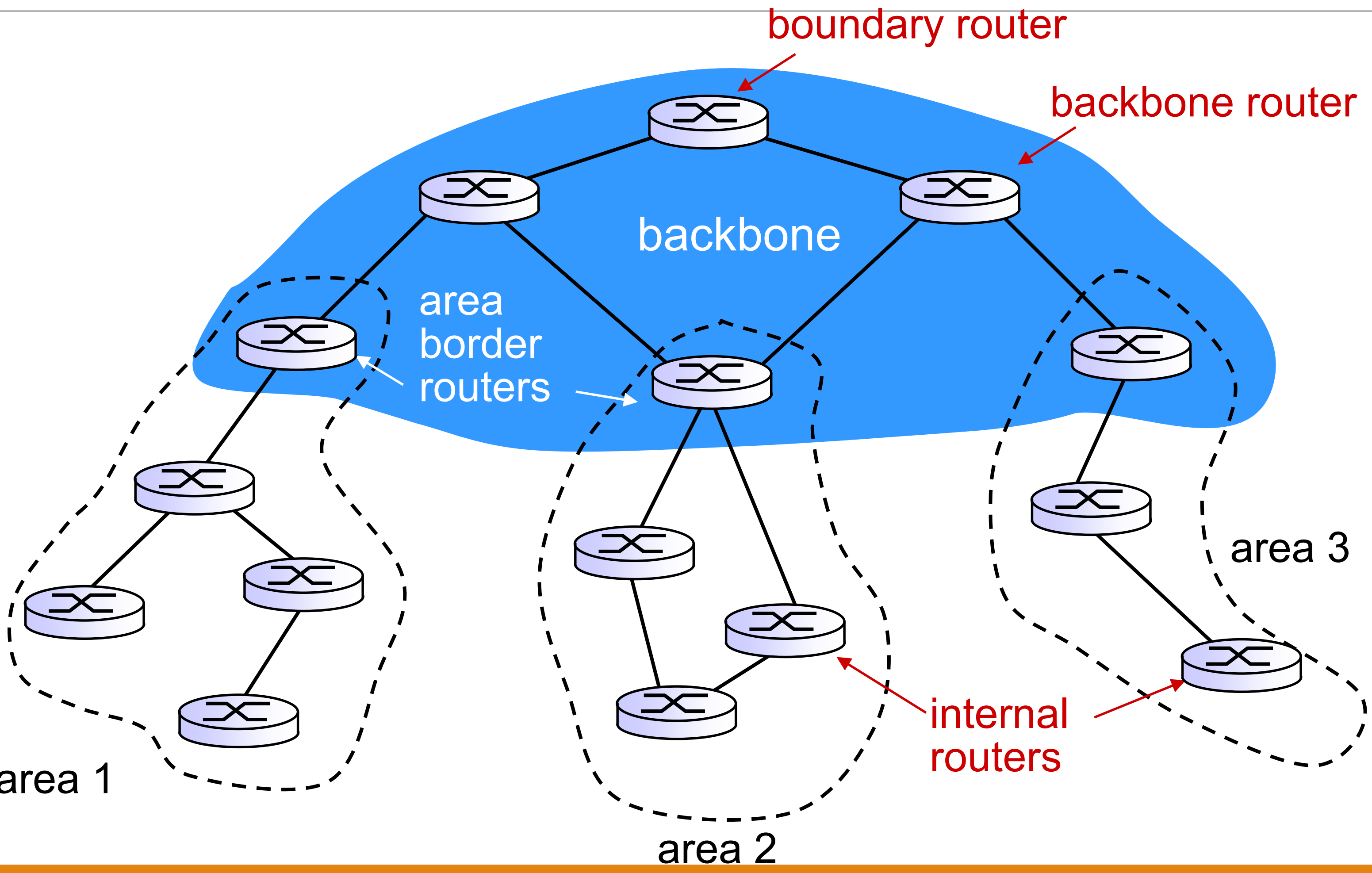
OSPF (Open Shortest Path First)

- “open”: publicly available
- uses link-state algorithm
  - link state packet dissemination
  - topology map at each node
  - route computation using Dijkstra’s algorithm
- router floods OSPF link-state advertisements to all other routers in **entire** AS
  - 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

OSPF "Advanced" Features

- **security***:* all OSPF messages authenticated (to prevent malicious intrusion)
- **multiple** same-cost **paths** allowed (only one path in RIP)
- for each link, multiple cost metrics for different **ToS** (e.g., satellite 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 database as OSPF
- **hierarchical** OSPF in large domains.

Hierarchical OSPF (1 of 2)



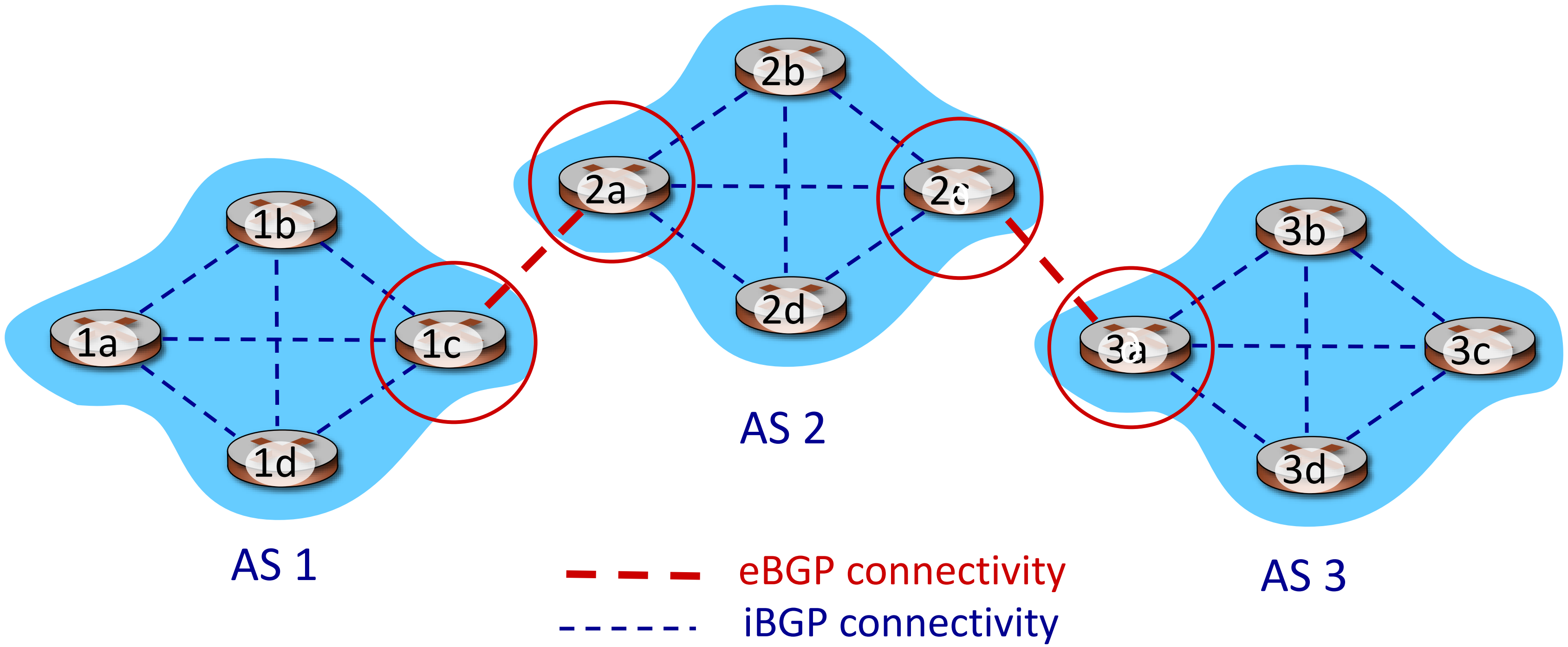
Hierarchical OSPF (2 of 2)

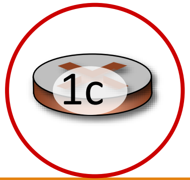
- **two-level hierarchy:** local area, backbone.
  - link-state advertisements only in area
  - each node has detailed area topology; only know direction (shortest path) to nets in other areas.
- **area border routers:** “summarize” distances to nets in own area, advertise to other Area Border routers.
- **backbone routers:** run OSPF routing limited to backbone.
- **boundary routers:** connect to other AS’es.

Internet inter-AS Routing: BGP

- **BGP (Border Gateway Protocol):** the *de facto* inter-domain routing protocol
  - "glue that holds the Internet together"
- BGP provides each router a means 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**
- allows subnet to advertise its existence to the rest of Internet: **"I am here"**

eBGP, iBGP Connections

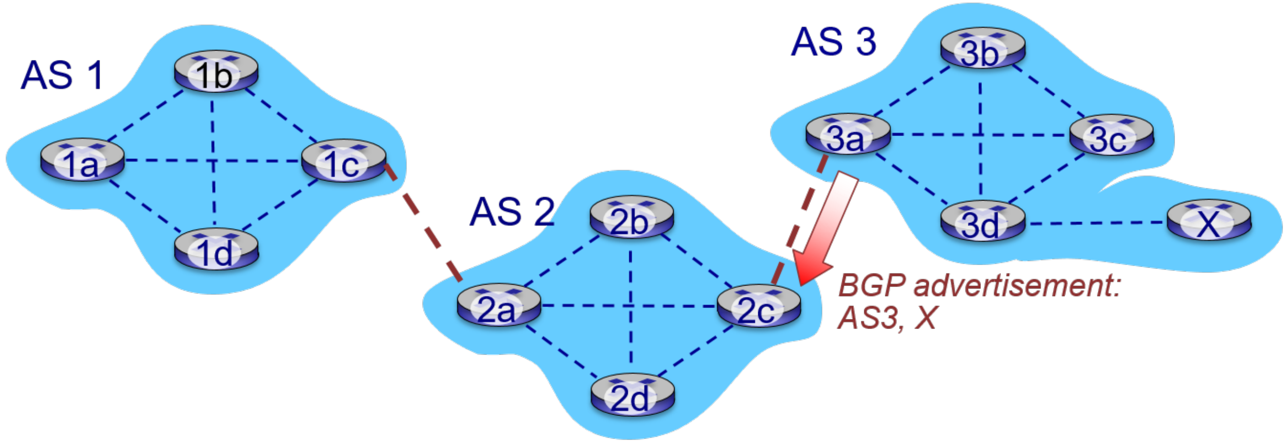




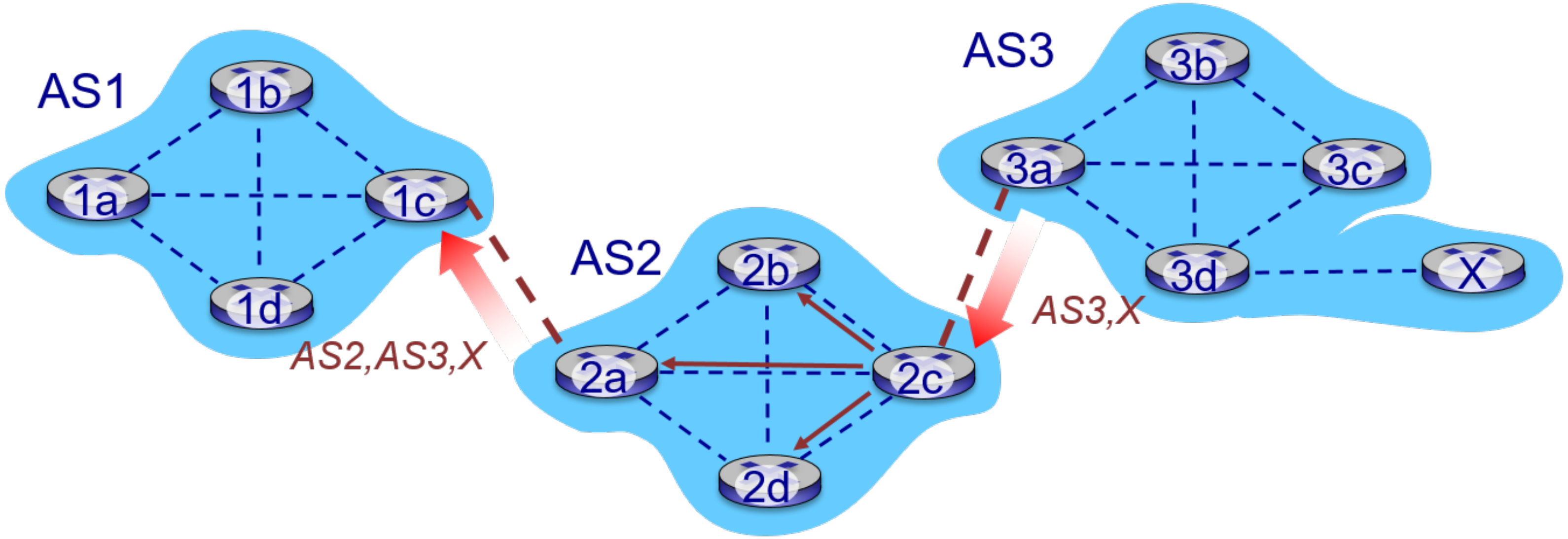
gateway routers run both eBGP and iBGP protocols

BGP Path Advertisement (1 of 3)

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

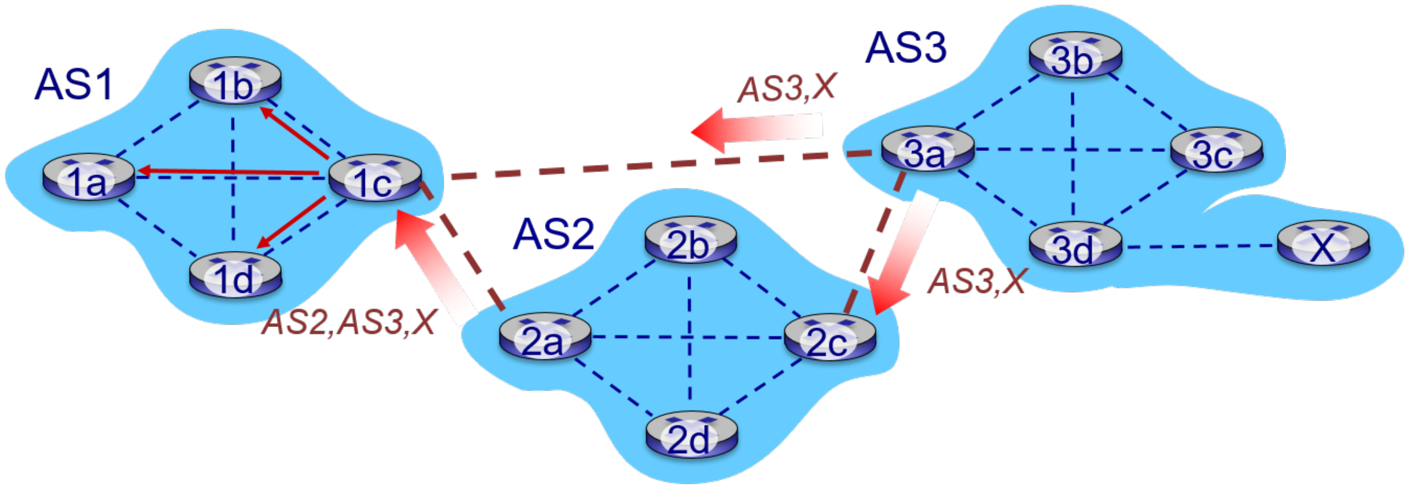


BGP Path Advertisement (2 of 3)



- AS 2 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 (3 of 3)



Gateway router may learn about multiple paths to destination:

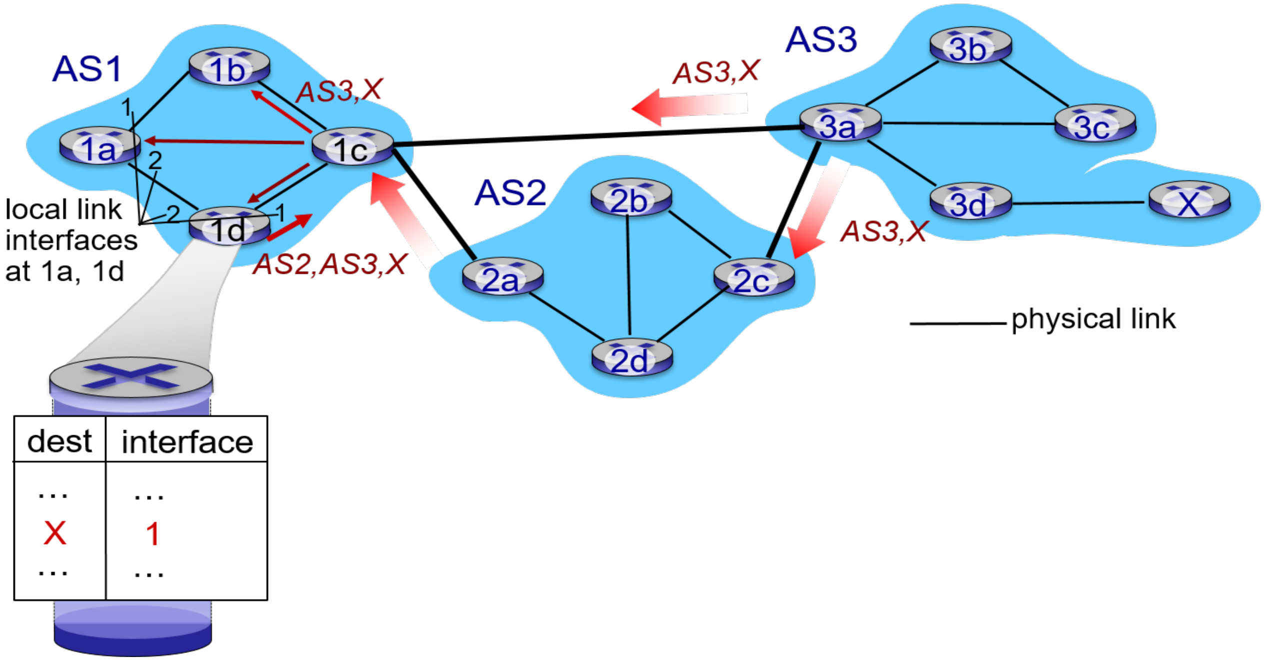
- AS1 gateway router 1c learns path **AS2, AS3, X** from 2a
- AS1 gateway router 1c learns path **AS3, X** from 3a
- Based on policy, AS1 gateway router 1c chooses path **AS3, X, and advertises path within AS1 via iBGP**

Path Attributes and BGP Routes

- Question: Potentially many paths from a given router to a destination subnet. How does a router choose among these paths (and configure it forwarding table accordingly)?
- advertised prefix includes BGP attributes
  - prefix + attributes = “route”
- two important attributes:
  - **AS-PATH:** list of ASes through which prefix advertisement has passed
  - **NEXT-HOP:** IP address of the router interface that begins the AS-PATH
- Each BGP route is written as a list with three components: NEXT-HOP; AS-PATH; destination prefix.
- **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

BGP, OSPF, Forwarding Table Entries (1 of 2)

Q: how does router set forwarding table entry to distant prefix?



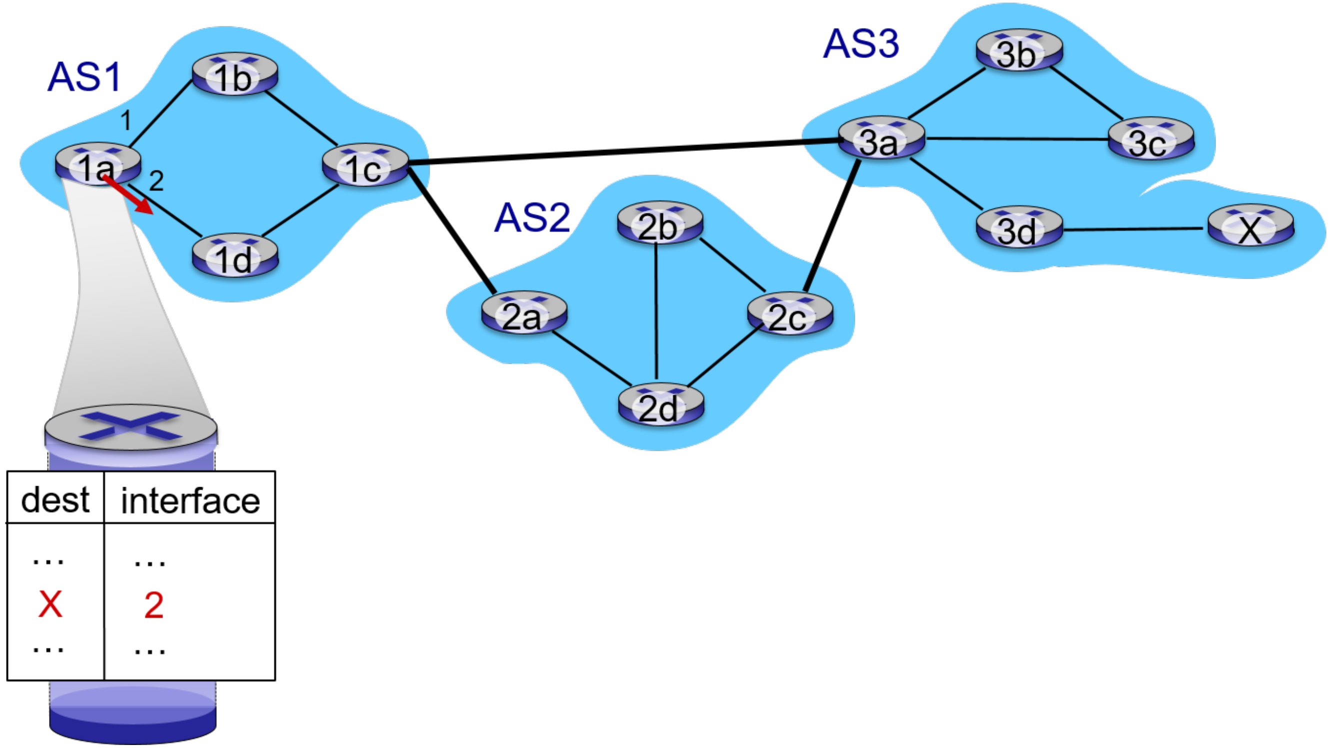
dest	interface
…	…
X	1
…	…

- 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 (2 of 2)

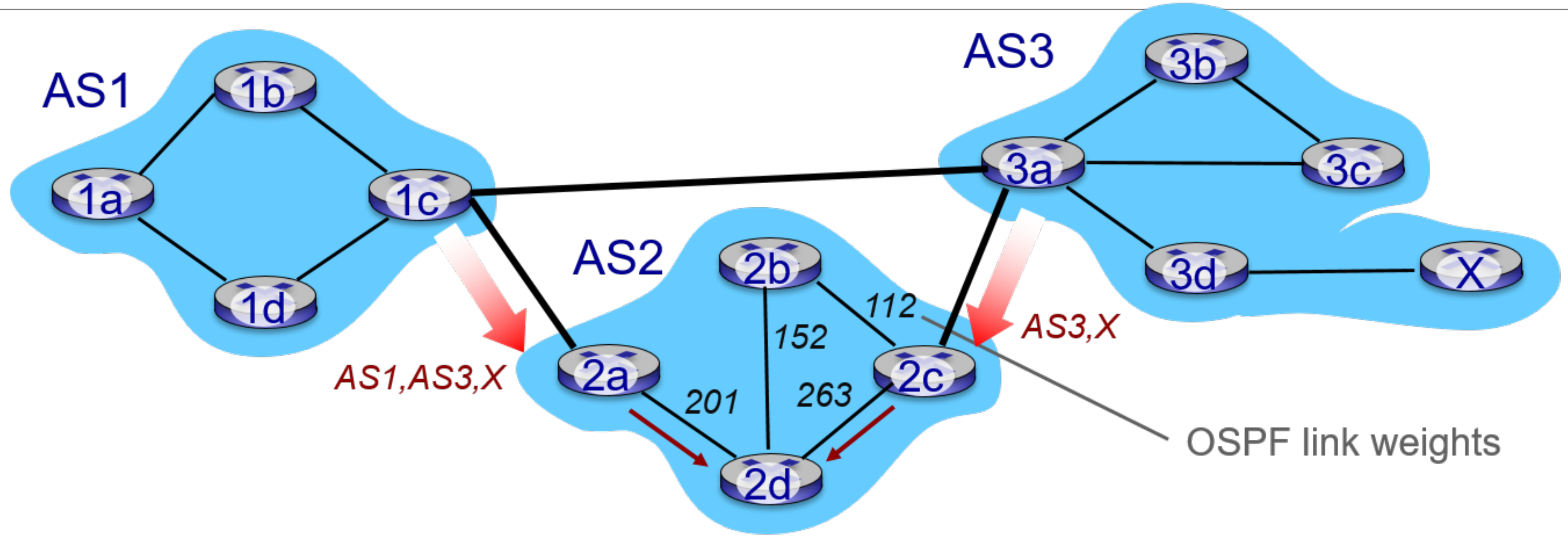
Q: how does router set forwarding table entry to distant prefix

- 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



dest	interface
…	…
X	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!

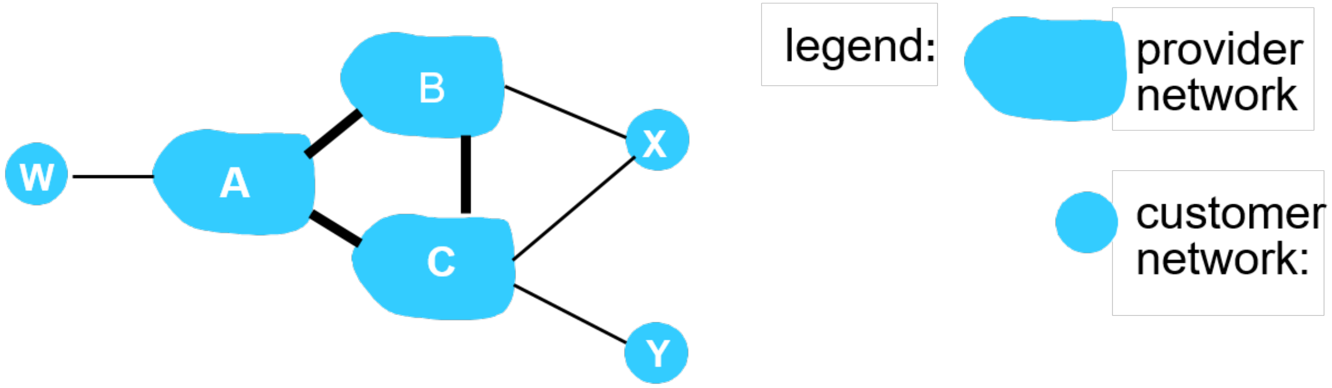
BGP Route Selection Algorithm

- In practice, BGP uses an algorithm that is more complicated than Hot Potato routing!
- A router may learn about **more than one route** to destination AS, BGP sequentially invokes elimination rules until one route remains.
  1. Assign a route with a local preference value (policy decision). Select routes with the highest local preference values.
  2. Among all routes with the same highest local preference value, select routes with shortest AS–PATH (by DV algorithm, using number of AS hops as distance metric)
  3. From the remaining routes, select the closest NEXT-HOP router: hot potato routing!
  4. If more than route still remains, use additional criteria, e.g., BGP identifiers

BGP Routing Policy

- When a router selects a route to a destination, the AS routing policy can trump all other considerations, such as shortest AS path or hot potato routing
- See the first step of BGP route selection algorithm

BGP: Achieving Policy Via Advertisements (2 of 2)

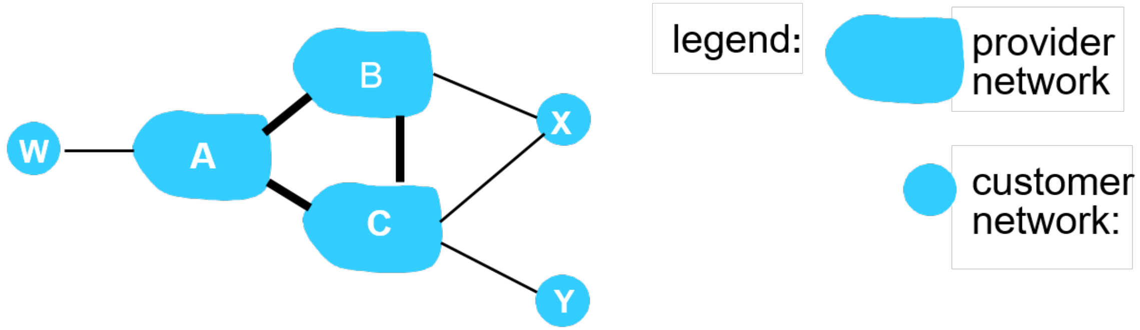


Suppose a customer network (i.e., ISP access network) only wants to route traffic to/from its own customers (does not want to carry transit traffic between other ISPs – no free rides!).

- A,B,C are **provider networks**
- X,W,Y are customer networks (of the attached 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: Achieving Policy Via Advertisements (1 of 2)



Suppose a provider network only wants to route traffic to/from its customer networks (does not want to carry transit traffic between other provider networks – no free rides!)

- Focus on B and its customer X
- A advertises path Aw to B and to C
- B advertises path BAw to its customer X
- 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

Why Different Routing Protocols for Intra-AS and Inter-AS?

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:

- inter – AS: a critical issue as to handle routing to/among large numbers of networks
- intra - AS: scalability is less of a concern: hierarchical routing saves table size, reduced update traffic

performance:

- inter - AS: policy may dominate over performance
- intra - AS: can focus on performance