Lecture 19: Network Layer
Routing in the Internet

COMP 332, Spring 2018

Victoria Manfredi

WESLEYAN
UNIVERSITY

Acknowledgements: materials adapted from Computer Networking: A Top Down Approach 7th edition: ©1996-2016, J.F Kurose and K.W. Ross, All Rights Reserved as well as from slides by Abraham Matta at Boston University, and some material from Computer Networks by Tannenbaum and Wetherall.

Today

1. Announcements
   - homework 7 due today Wed. 11:59p
   - run the traceroute command and look at traffic in wireshark
     - compare with pkts you're generating
2. Internet routing
   - overview
   - intra-AS routing
   - inter-AS routing
3. Internet Control Message Protocol (ICMP)

Internet Routing

OVERVIEW

From graph algorithms to routing protocols

Need to address Internet reality

1. Internet is network of networks
   - hierarchical structure
   - routers not all identical
     - some routers connect different networks together
   - each network admin may want to control routing in its own network

2. Scalability with billions of destinations
   - don't all fit in one routing table
   - can't exchange routing tables this big
     - would use all link capacity

Scalable routing on the Internet

Aggregate routers into regions called Autonomous Systems

Autonomous Systems (AS)

- aka domain
- network under single administrative control
  - company, university, ISP, …
- 30,000+ ASes: AT&T, IBM, Wesleyan …
- each AS has a unique 16-bit AS #
  - Wesleyan: AS167
  - BBN: used to be AS1: was first org to get AS # then L3 later acquired

```
AS160   U-CHICAGO-AS - University of Chicago, US
AS161   TI-AS - Texas Instruments, Inc., US
AS162   DNIC-AS-00162 - Navy Network Information Center (NNIC), US
AS163   IBM-RESEARCH-AS - International Business Machines Corporation
AS164   DNIC-AS-00164 - DoD Network Information Center, US
AS165   DNIC-AS-00165 - DoD Network Information Center, US
AS166   IDA-AS - Institute for Defense Analyses, US
AS167   WESLEYAN-AS - Wesleyan University, US
AS168   UMASS-AMHERST - University of Massachusetts, US
AS169   HANSCOM-NET-AS - Air Force Systems Networking, US
```

Hierarchical routing

Idea

- impose 2nd hierarchy on Internet: limits which routers talk to each other
- 1st hierarchy: address hierarchy governs how packets are forwarded

2-level route propagation hierarchy

- intra AS routing protocol between routers in same AS
  - aka intra domain routing protocol
  - aka interior gateway protocol
  - each AS selects its own
- inter AS routing protocol between gateway routers in different ASes
  - aka inter domain routing protocol
  - aka exterior gateway protocol
  - Internet-wide standard

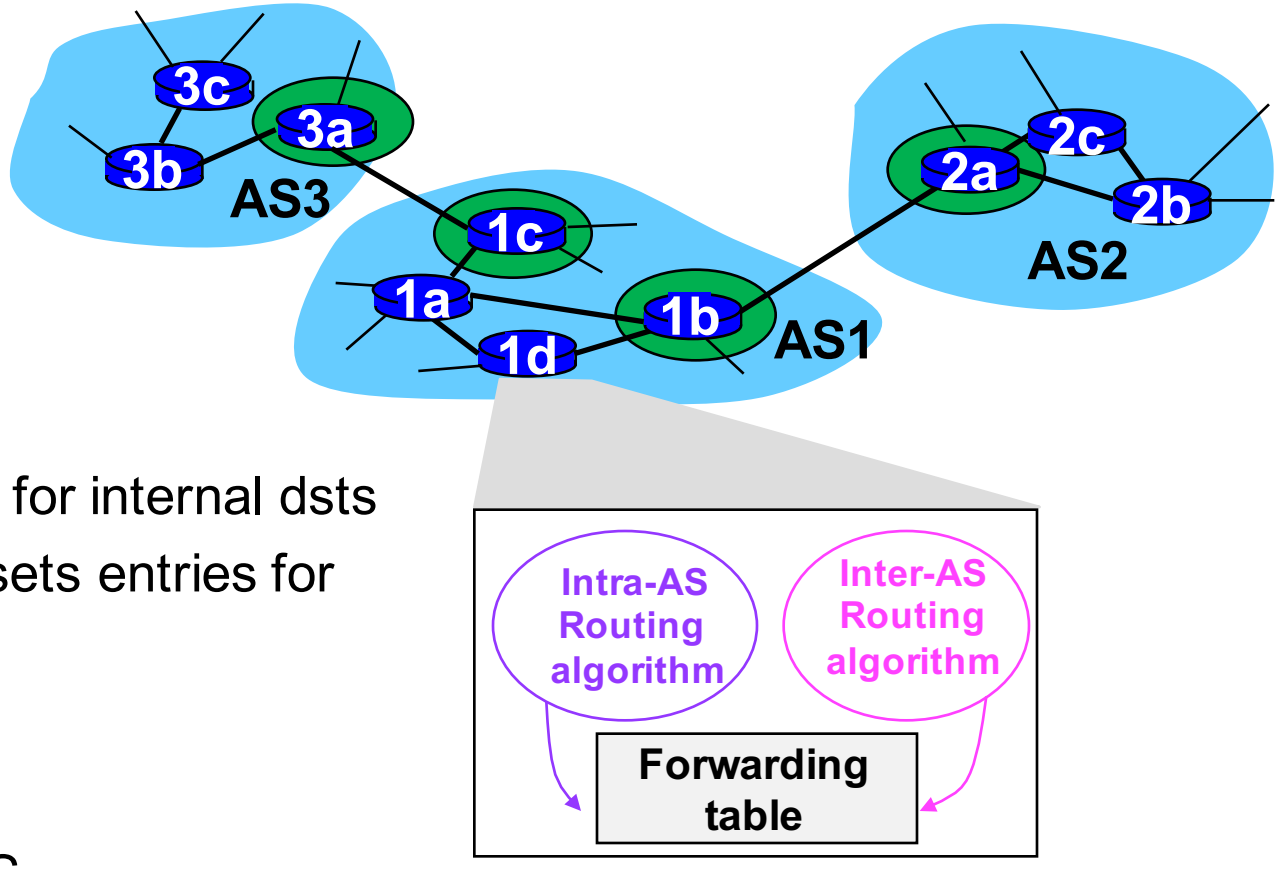
Focus is performance

Policy may dominate performance

Q: Can routers in different ASes run different intra AS routing protocol?

Q: Why are there different intra and inter-AS protocols?

Hierarchical routing



Forwarding table

- intra-AS sets entries for internal dsts
- inter-AS & intra-AS sets entries for external dsts

Gateway router

- at edge of its own AS
- direct link to router in another AS
- perform inter-AS as well as intra-AS routing
- distributes results of inter-AS routing to other routers in AS

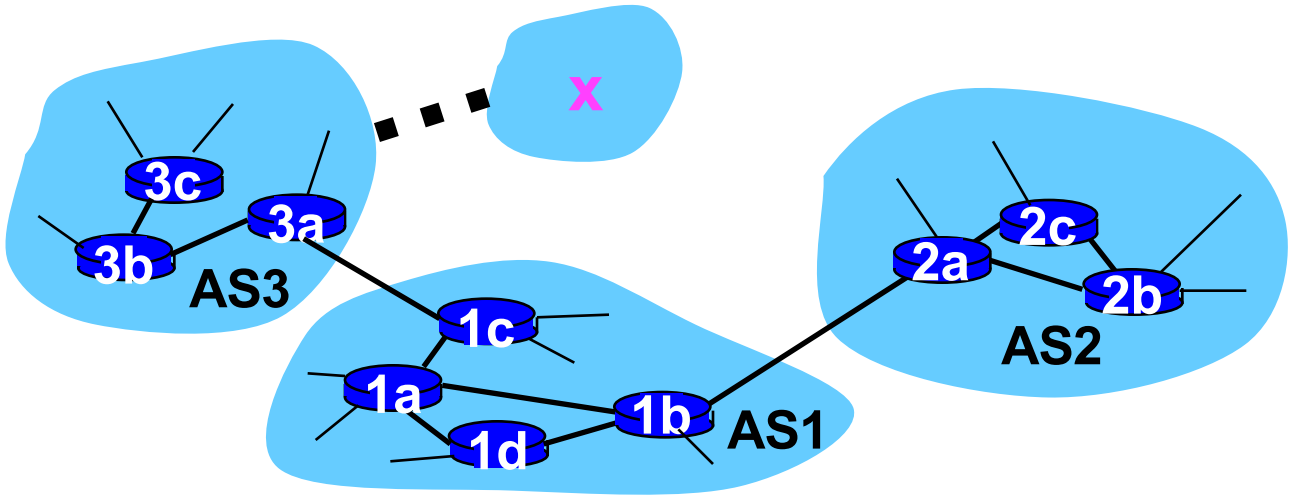
Example: set forwarding table in router 1d

Suppose AS1 learns (via inter-AS protocol)

- subnet **x** is reachable via AS3 (gateway 1c) but not via AS2
- inter-AS protocol propagates reachability info to all internal routers

Router 1d determines from intra-AS routing info

- that its interface **y** is on least cost path to 1c.
- installs forwarding table entry **(x,y)**



Q: What if multiple ASes can be used to reach x?

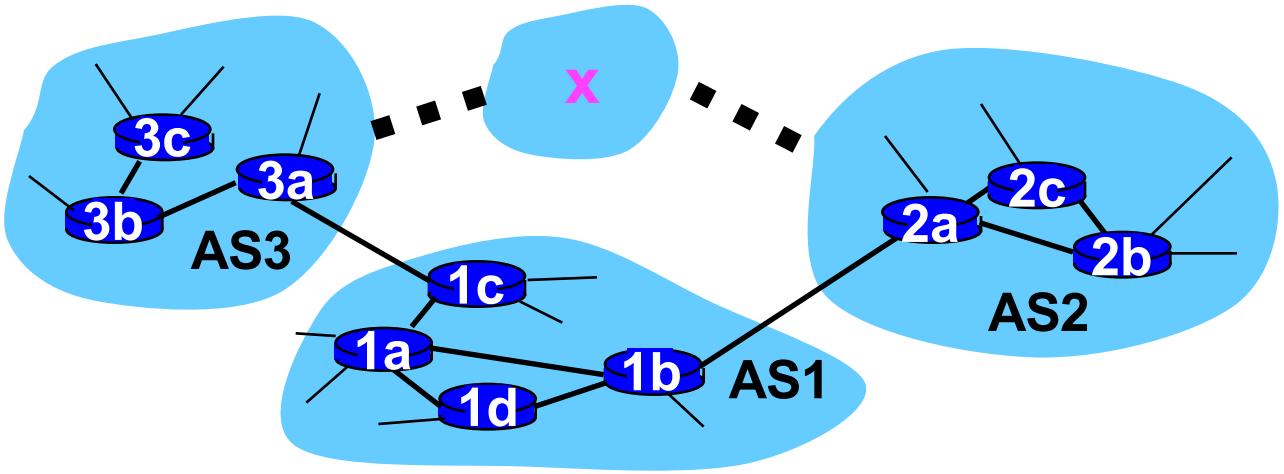
Example: choosing among multiple ASes

Suppose AS1 learns from inter-AS protocol

- subnet x is reachable from AS3 **and** from AS2

To configure forwarding table, router 1d must determine towards which gateway it should forward packets for dst x

- may take policy into account
- this is also job of inter-AS routing protocol!



Internet ROUTING

INTRA-AS ROUTING

Inter-AS tasks

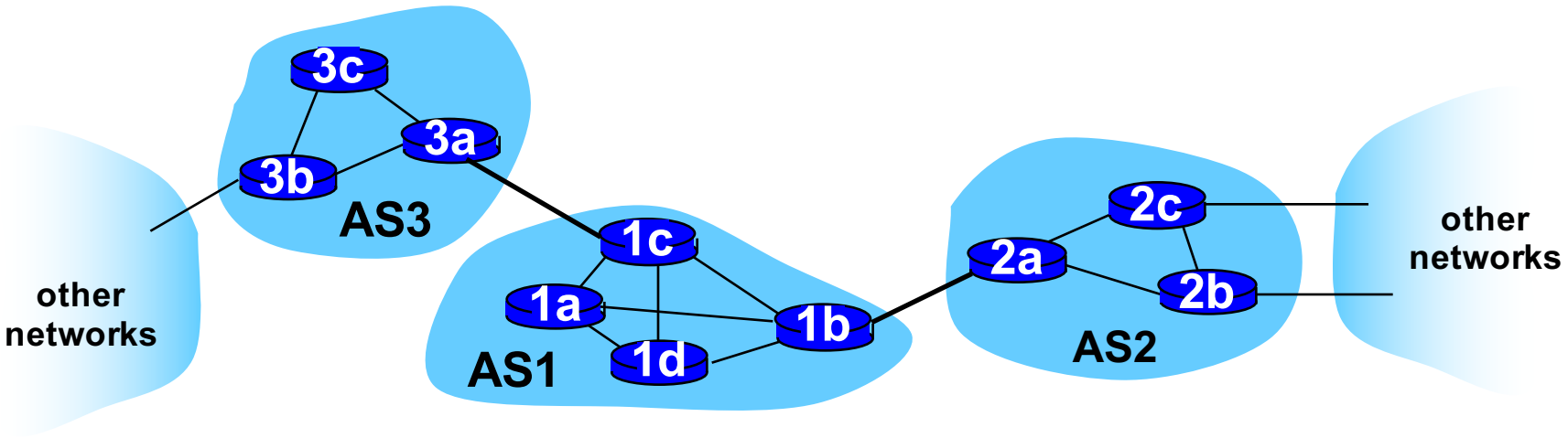
Suppose router in AS1 receives pkt destined outside of AS1

- router should forward packet to gateway router, but which one?

AS1 must

- learn which dsts are reachable through AS2, which through AS3
- propagate this reachability info to all routers in AS1

⇒ job of inter-AS routing!



Most common intra-AS routing protocols

RIP

- Routing Information Protocol
- distance vector protocol

(E)IGRP

- (Enhanced) Interior Gateway Routing Protocol
- Cisco proprietary for decades, until 2016
- distance vector protocol

IS-IS

- Intermediate System to Intermediate System
- link state protocol

OSPF

- Open Shortest Path First
- link state protocol

Open Shortest Path First (OSPF)

Open

- i.e., publicly available

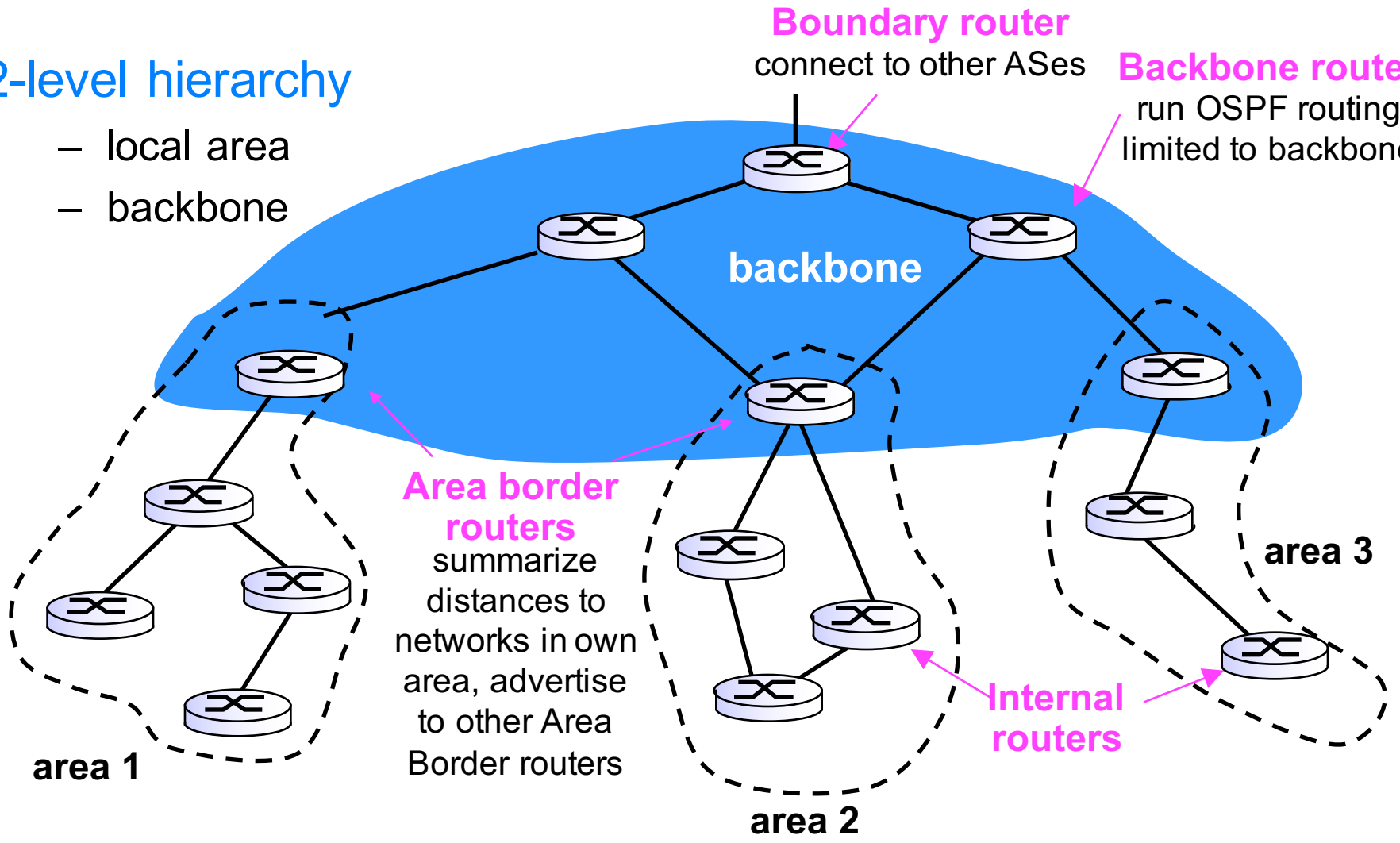
Link-state algorithm

1. each router floods its link state to all other routers in AS
   - messages carried directly over IP
   - message authentication possible
   - supports both unicast (1src –1dst) and multicast (1src - multiple dst)
2. each router builds topology map
3. route computation using Dijkstra's
   - can have multiple paths with same cost
     - traffic can go over different paths
   - can have different costs per link depending on type of service
     - e.g., satellite link cost: low for best effort, high for real time

Hierarchical OSPF in large domains

2-level hierarchy

- local area
- backbone



Link-state advertisements only in area:
internal routers have detailed area topology but only know direction (shortest path) to networks in other areas (like distance vector between areas)

Internet ROUTING

INTER-AS ROUTING

Border Gateway Protocol (BGP)

The de facto inter-domain routing protocol

- "glue that holds the Internet together"
- path vector protocol

BGP provides each AS a means to

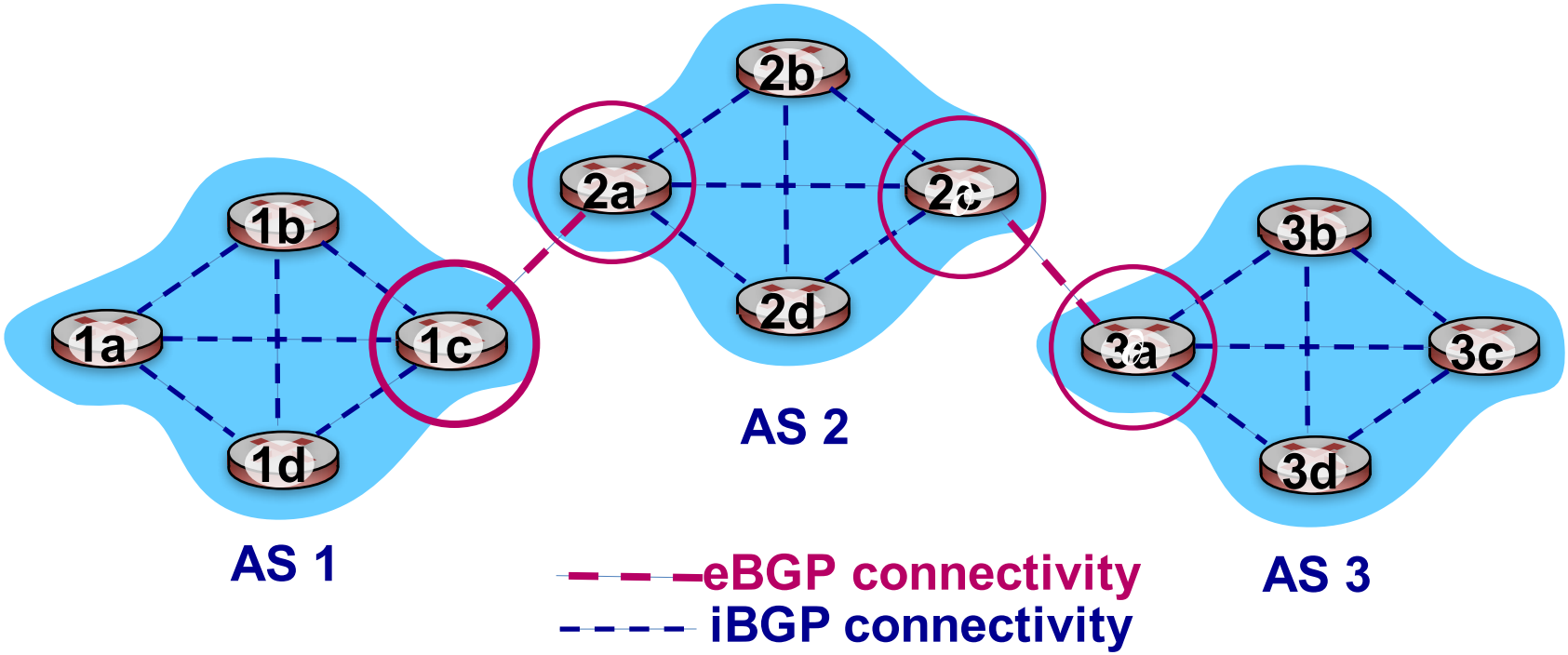
- eBGP: external
  - obtain subnet reachability info (routes) from neighboring ASes
- iBGP: internal
  - propagate externally learned reachability info (routes) to all routers in AS
  - similar to intra-AS routing protocols but more scalable
- determine "good" routes to other networks
  - based on reachability info and policy

Allows subnet to advertise its existence to rest of Internet

- "I am here"

Q: why do all ASes need to use same inter-AS protocol

eBGP vs. iBGP connections

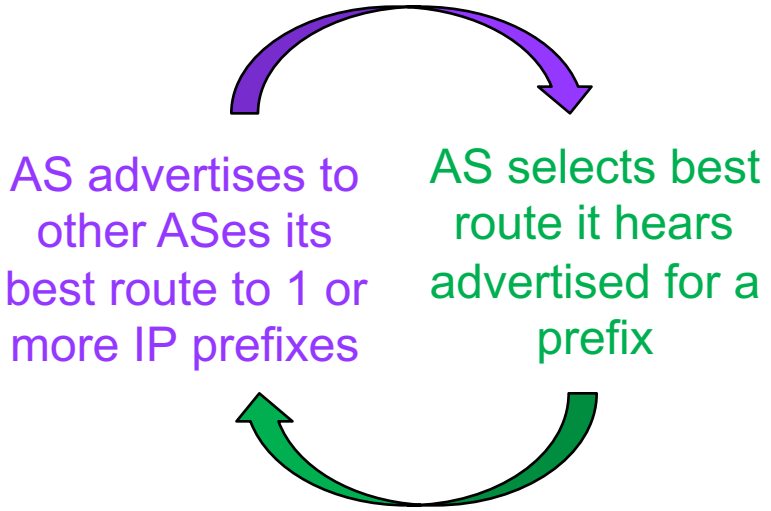


gateway routers run both eBGP and iBGP protocols

How eBGP works

Similarities with distance vector

- per dst route info advertised
- no global sharing of network topology
- iterative distributed convergence



Differences from distance vector

- selects best route based on policy not min cost
- path vector routing
  - advertises entire path for each dst rather than cost
    - allows policies based on full path
    - avoids loop: if your AS is in path then discard
  - selective route advertisements
    - choose not to advertise route to dst for policy reasons
    - aggregate routes for scalability: e.g., a.b.*.* and a.c.*.* become a.*.*.*

Message contents for path advertisement

Advertised prefix includes BGP attributes

- prefix + attributes = BGP "route"

2 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 advertisements
  - uses import policy to accept/decline path
  - e.g., never route through AS Y
- determines whether to advertise path to other neighboring ASes

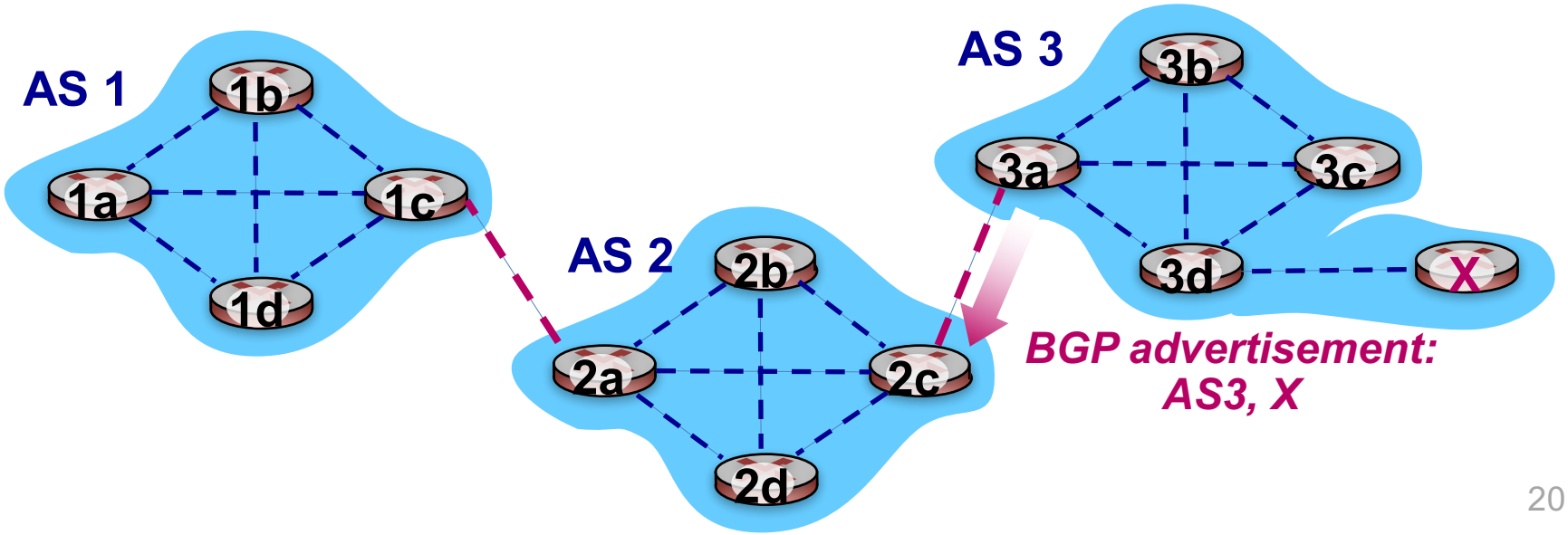
Session

Two BGP routers ("peers") exchange BGP messages

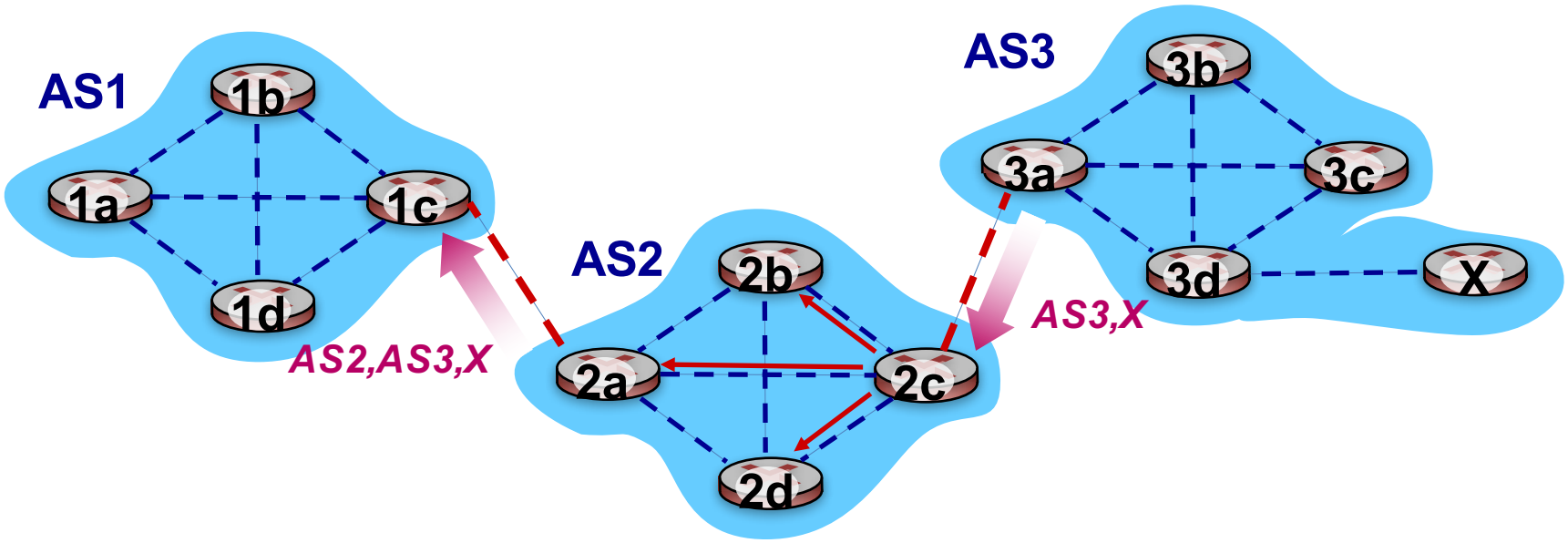
- over semi-permanent TCP connection
- advertise paths to different destination network prefixes

AS3 gateway router 3a

- advertises path ***AS3,X*** to AS2 gateway router 2c
  - i.e., AS3 promises to AS2 it will forward packets towards ***X***



How path advertisement works



AS2 gateway 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
- AS2 router **2a** advertises (via eBGP) path ***AS2,AS3,X*** to AS1 router **1c**

What if there are multiple routes?

Gateway router may learn about multiple routes to dst AS

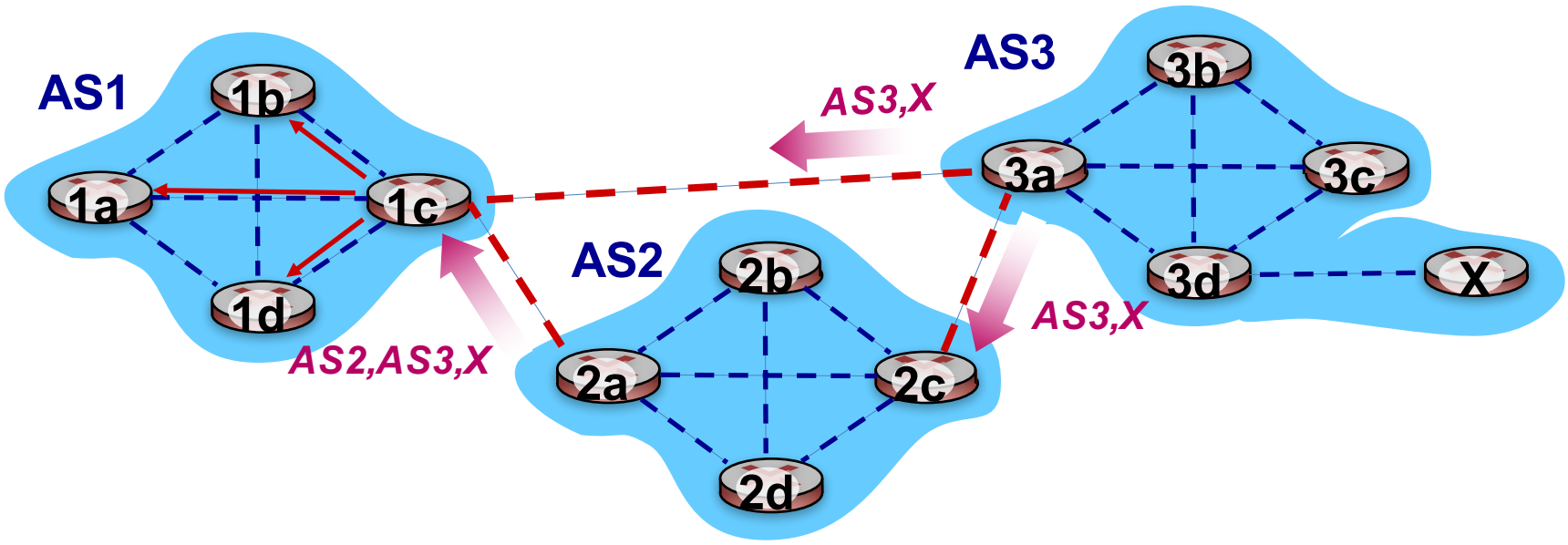
Route to use is up to AS but various strategies

- routes through peer ASes are better (don't pay)
- shorter AS paths are better
- lower cost within AS is better
  - hot potato routing: choose local gateway with lowest intra AS cost
- …

In practice

- BGP uses a more complicated version of hot potato routing

Multiple routes to destination AS

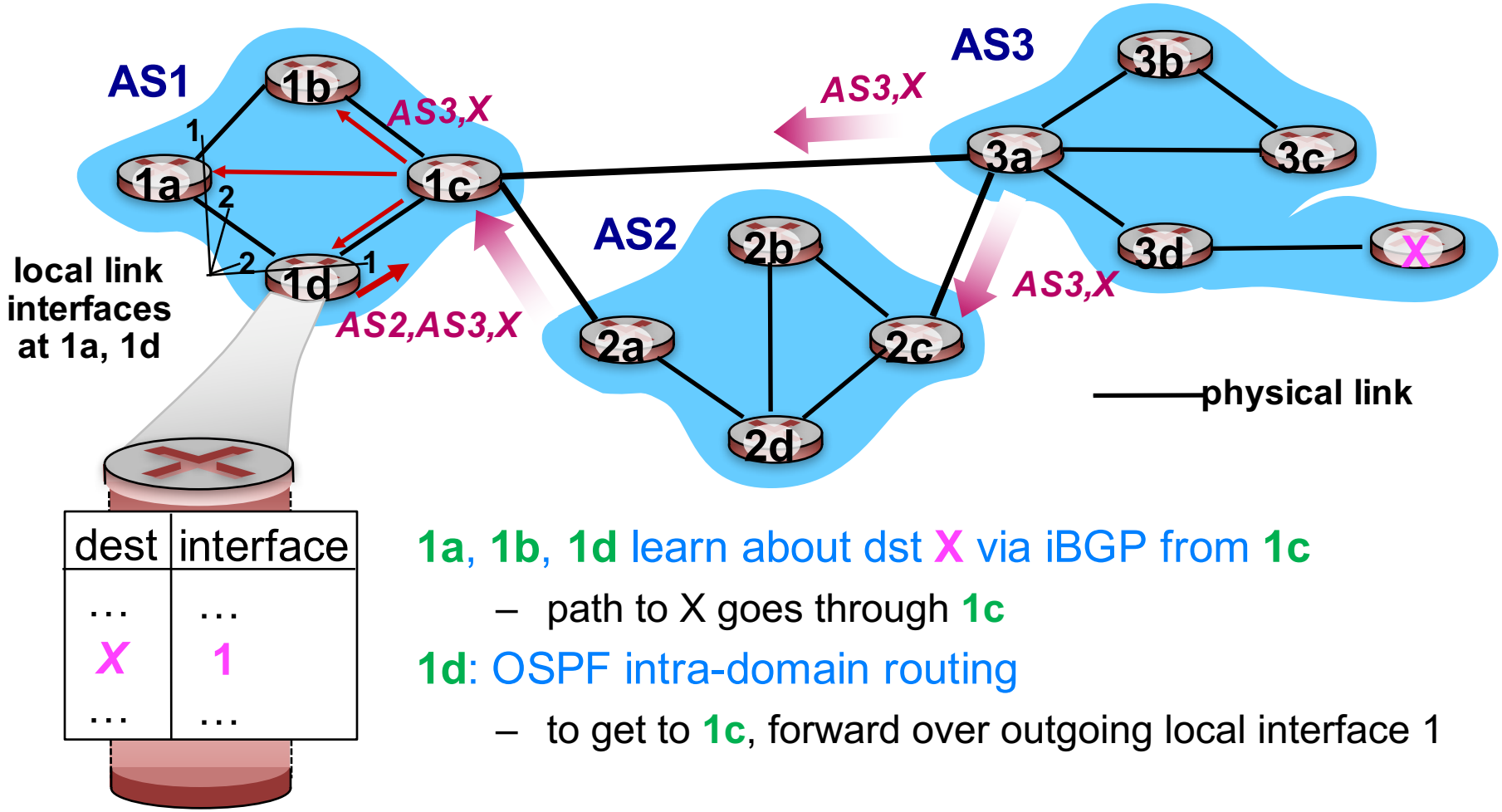


AS1 gateway router **1c**

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

Interactions between BGP and OSPF

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



1a, **1b**, **1d** learn about dst **X** via iBGP from **1c**

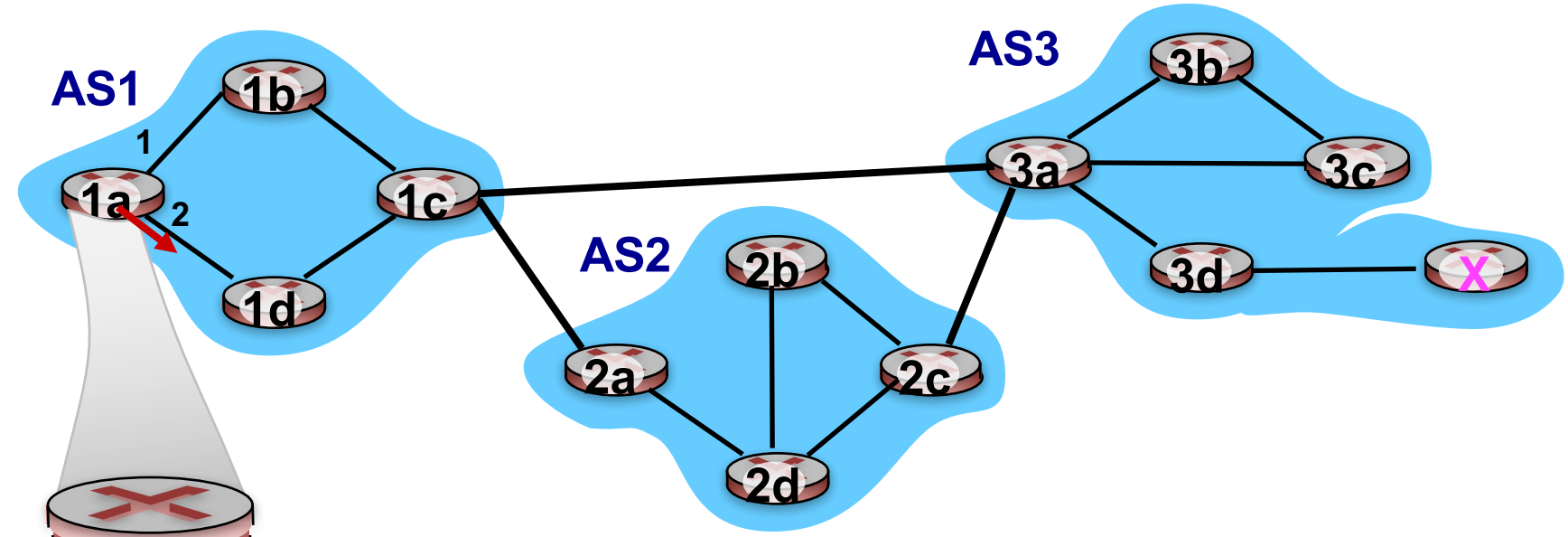
- path to X goes through **1c**

1d: OSPF intra-domain routing

- to get to **1c**, forward over outgoing local interface 1

Interactions between BGP and OSPF

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



dest	interface
…	…
X	2
…	…

1a, **1b**, **1d** learn about dst **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

Policy-shaped route selection

Political, economic, security considerations

Shaped by business relationships between ASes

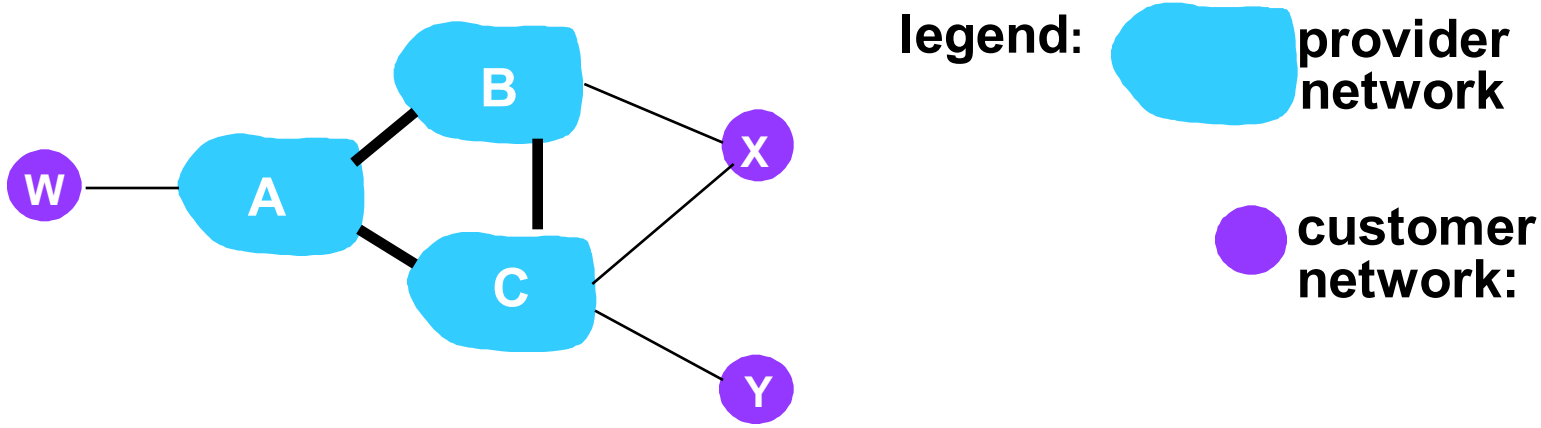
- AS1 is customer of AS2 (AS 1 pays AS2)
- AS1 is provider of AS 2
- AS1 is peer of AS 2 (peers don't pay each other to exchange traffic)

E.g.,

- don't want to carry commercial traffic on university network
- traffic to apple shouldn't transit through google
- pentagon traffic shouldn't transit through Iraq

Why BGP is so complicated!

Achieving policy via advertisements



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

Why different intra- vs. 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

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

INTERNET CONTROL MESSAGE PROTOCOL

OVERVIEW

Internet Control Message Protocol (ICMP)

Used by hosts & routers to communicate network-level information

- error reporting
  - unreachable host, network, port, protocol
- echo request/reply
  - used by ping)
- network-layer above IP
  - ICMP msgs carried in IP pkts

ICMP message

- type, code plus first 8 bytes of IP pkt causing error

Type	Code	Description
0	0	echo reply (ping)
3	0	dest. network unreachable
3	1	dest host unreachable
3	2	dest protocol unreachable
3	3	dest port unreachable
3	6	dest network unknown
3	7	dest host unknown
4	0	source quench (congestion control - not used)
8	0	echo request (ping)
9	0	route advertisement
10	0	router discovery
11	0	TTL expired
12	0	bad IP header

Traceroute and ICMP

Source sends series of segments or packets to destination

- first set has TTL =1
- second set has TTL=2, etc.
- unlikely port number

When *n*th set arrives to nth router

- router discards and sends source ICMP message (type 11, code 0)
- ICMP message includes name of router & IP address

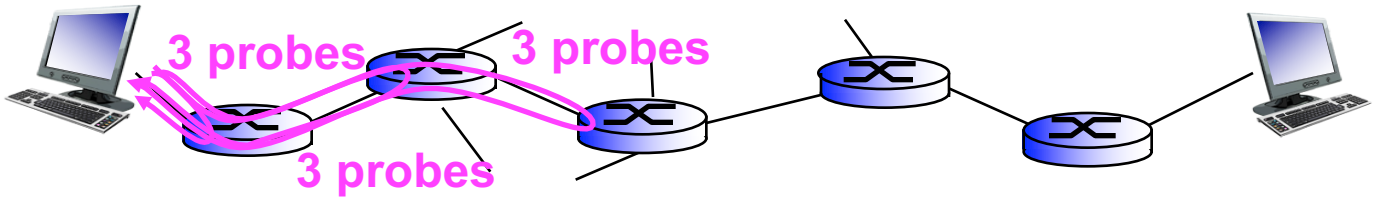
When ICMP msg arrives

- source records RTTs

Stopping criteria

TCP segment or UDP datagram eventually arrives at dst host

- dst returns ICMP "port unreachable" message
- source stops



Q: why can traceroute work with segments, datagrams, or packets?

ICMP traceroute

We're generating an ICMP echo request

Intermediate routers

- respond with ICMP ttl expired

Final destination

- responds with ICMP echo reply