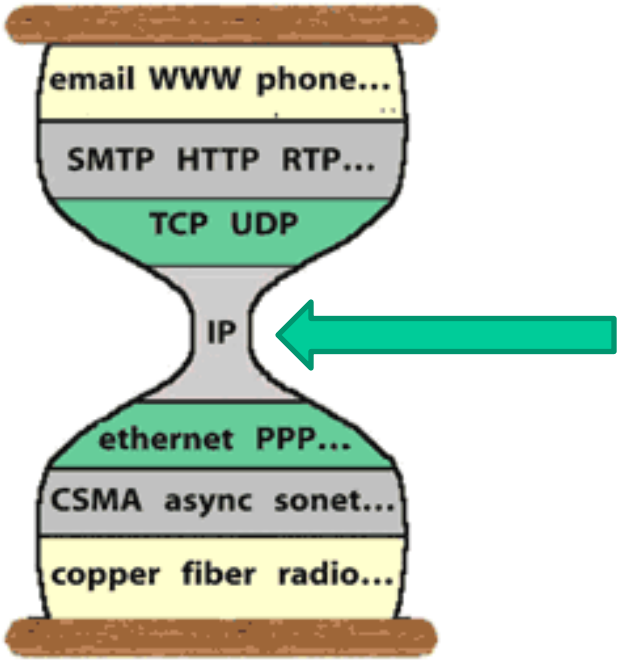
Lecture 14: Routing Protocols



5.1 Introduction

5.2 Routing protocols

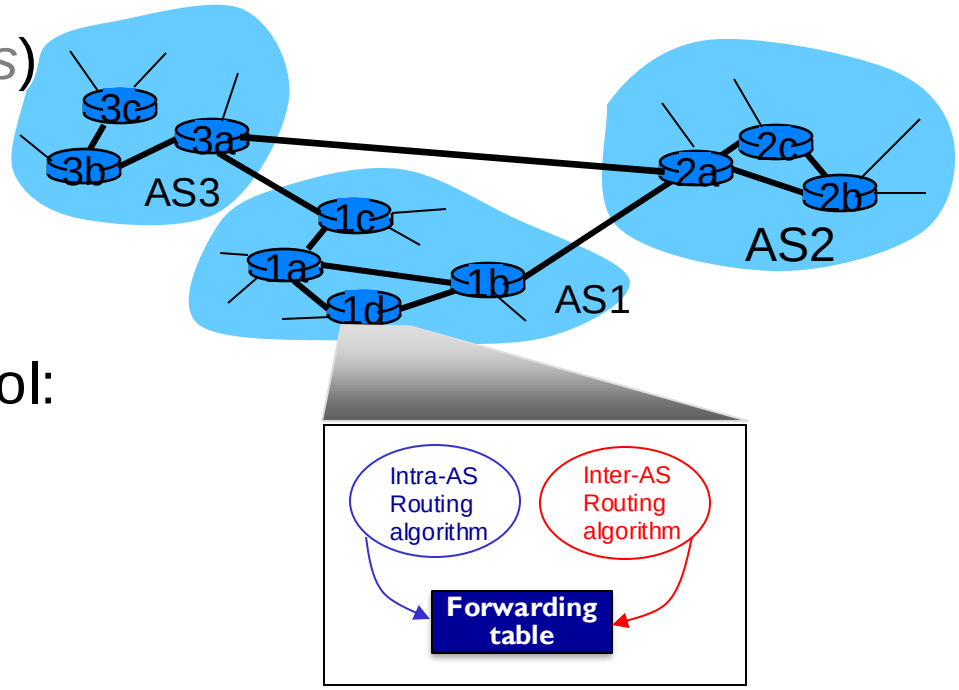
- Link state
- Distance vector

5.3 Intra-AS routing in the Internet: OSPF

5.4 Routing among the ISPs: BGP

Internet routing: 2-level hierarchy

- Intra-AS (within a campus, within an ISP)
  - *Intra*-Domain Routing protocols: RIP, **OSPF** (*and a few others*)
- Inter-AS (between ISPs, between stub and transit ASes)
  - *Inter*-Domain Routing protocol: **BGP** (*the only one*)
- intra- and inter-AS routing protocols jointly fill in each router's forwarding table
  - intra-AS sets entries for internal destinations
  - inter-AS & intra-AS sets entries for external destinations

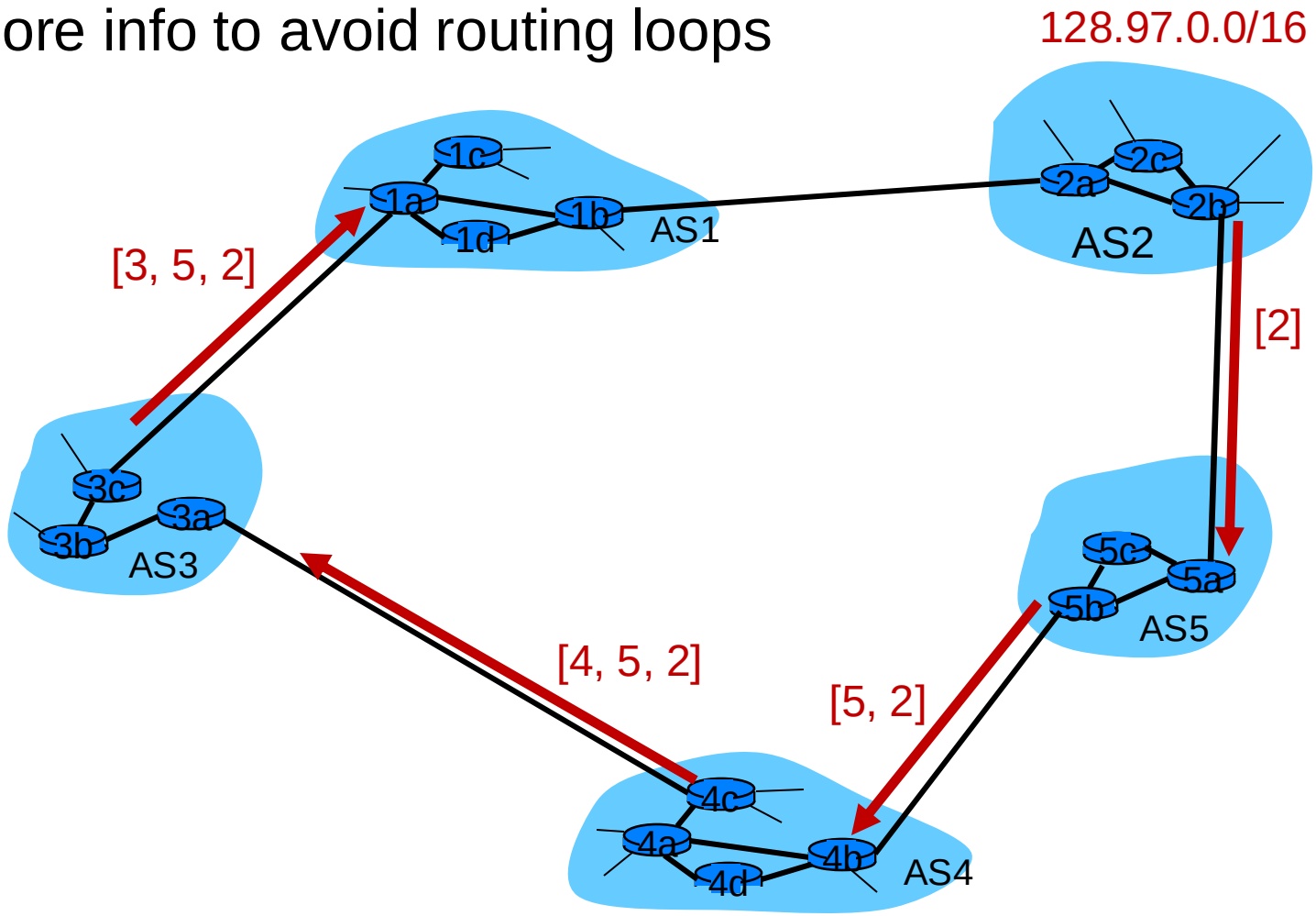


Border Gateway Protocol: BGP

Interior Gateway Protocol: OSPF, etc

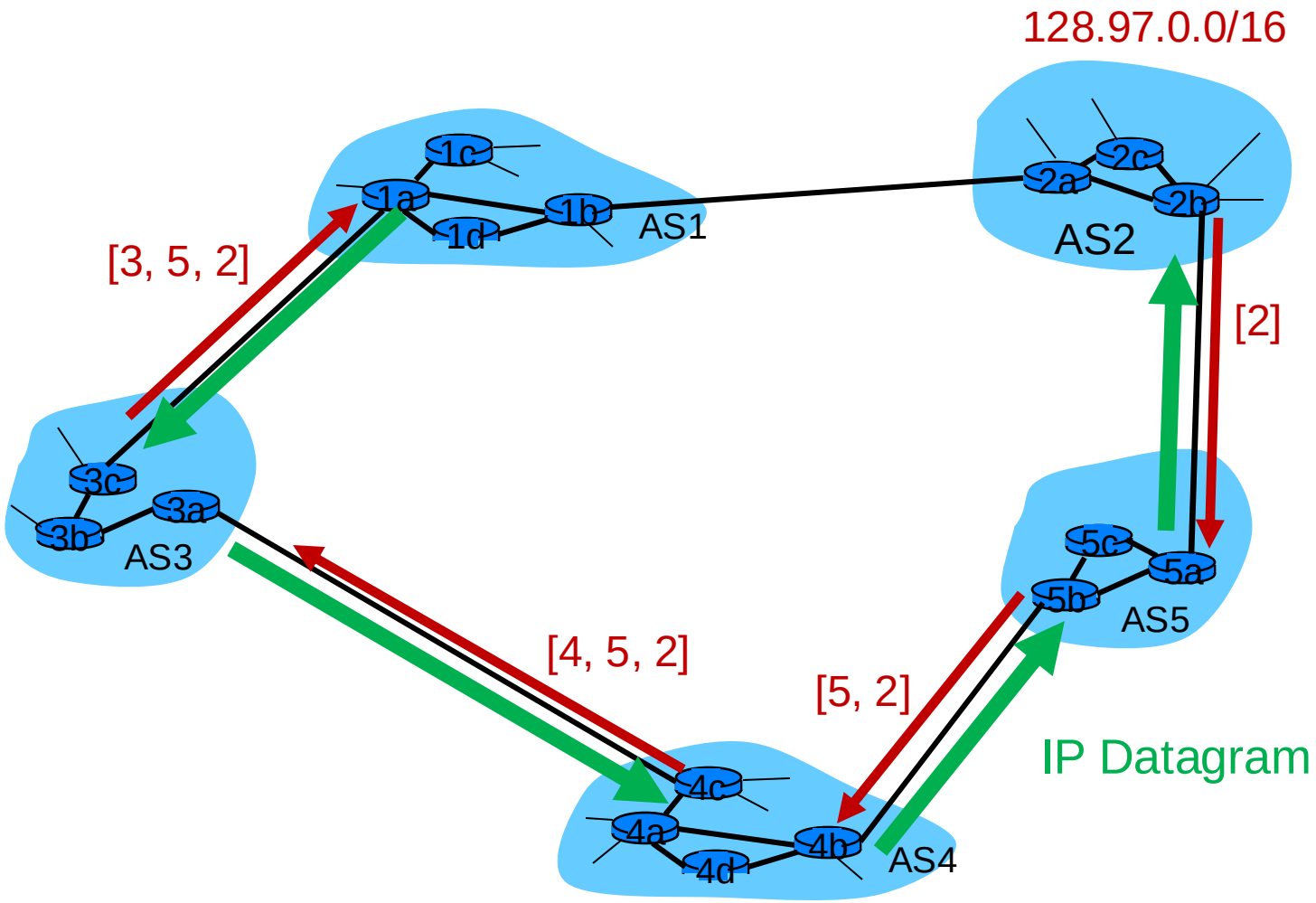
Strawman Solution

- Path-vector, a variant of distance vector
  - Announcing the whole network path
  - More info to avoid routing loops



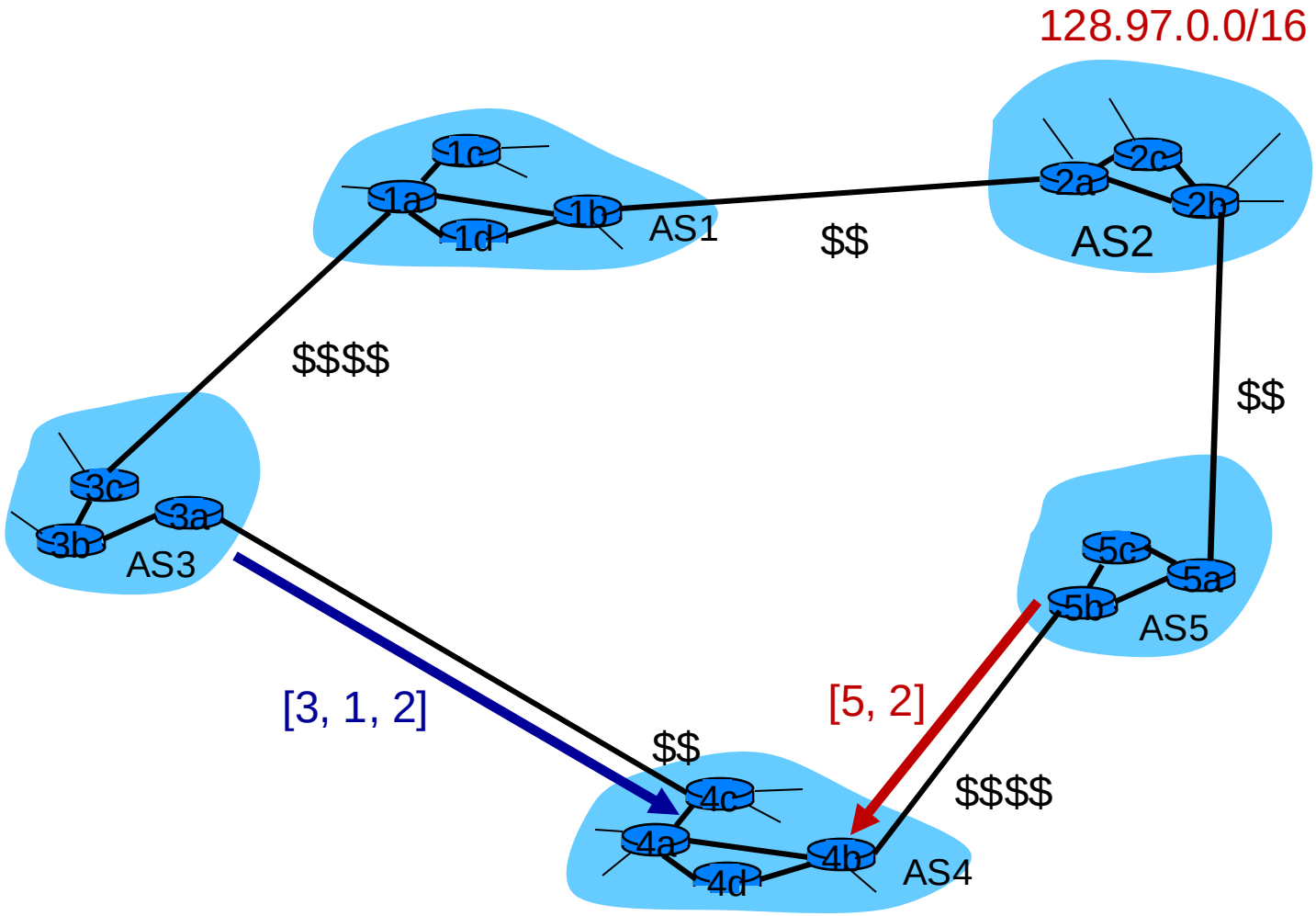
Strawman Solution

- BGP routers injecting prefixes to local OSPF routers
  - IP datagram follows the opposite path



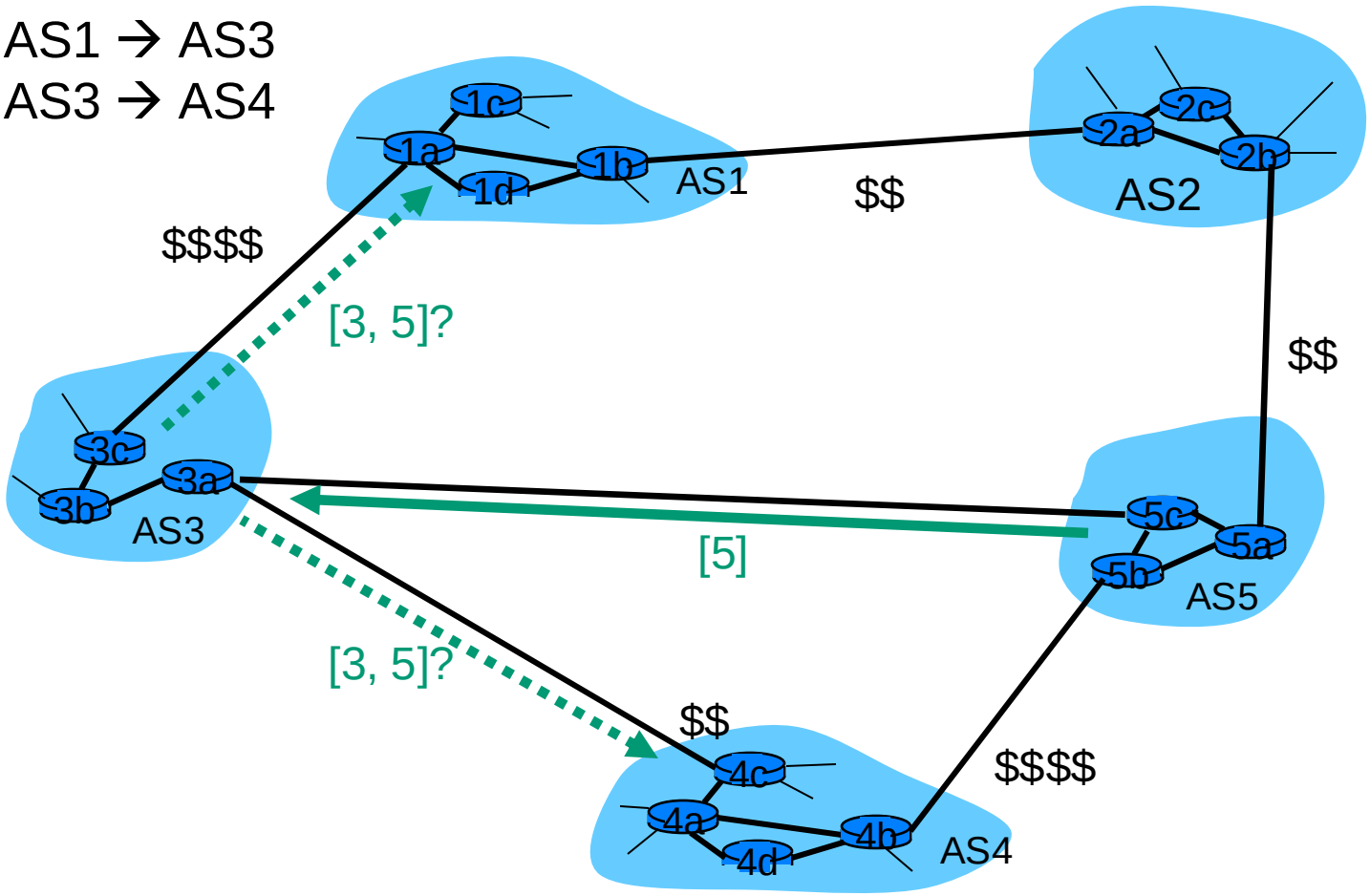
Not an ideal world: case 1

- AS4-AS3 link is cheaper to AS4-AS5 link
  - Which one AS4 prefers? Cheaper or closer?



Not an ideal world: case 2

- AS3 to reach AS5: hey lets directly connect
  - Do you want give AS4 and AS1 a “free” ride [3, 5]?
  - Provider → Customer relation?
    - AS1 → AS3
    - AS3 → AS4

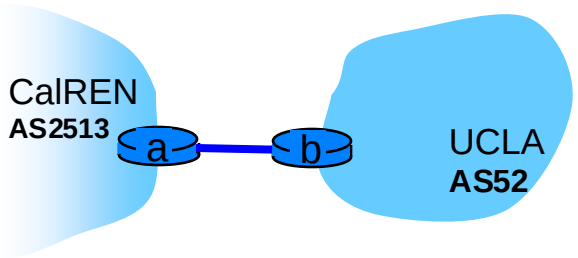


BGP: Border Gateway Protocol

BGP provides each AS a means to:

1. Advertise its own IP address prefixes to the rest of the Internet
2. Obtain IP address prefix reachability info from neighboring ASes
3. Propagate the reachability info to all routers *internal to the AS*.
4. Determine "good" routes to use for learned reachability to destination prefix and policy
   - Also propagate a proper set of the externally learned routers to selected neighbors

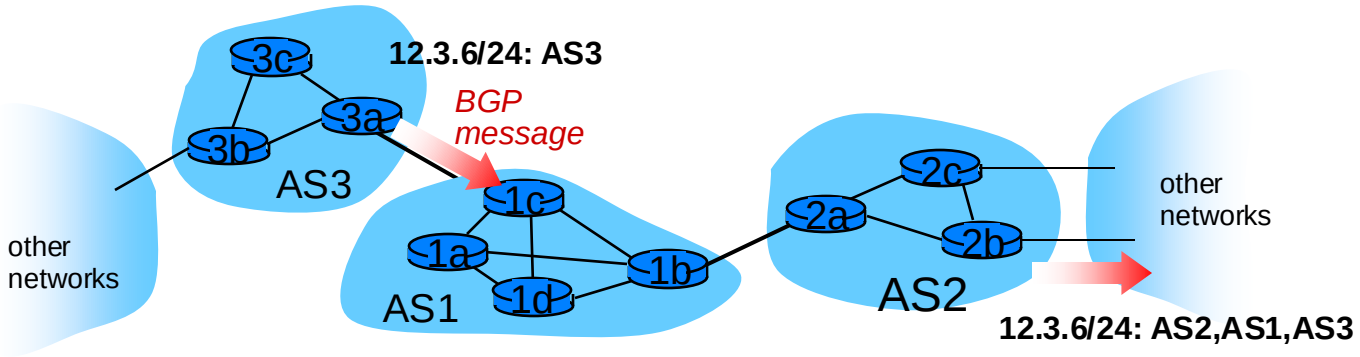
- Performing the above 4 tasks
- propagating (partial) prefix reachability info to (some of) the neighbors



BGP basics: distributing path information

important

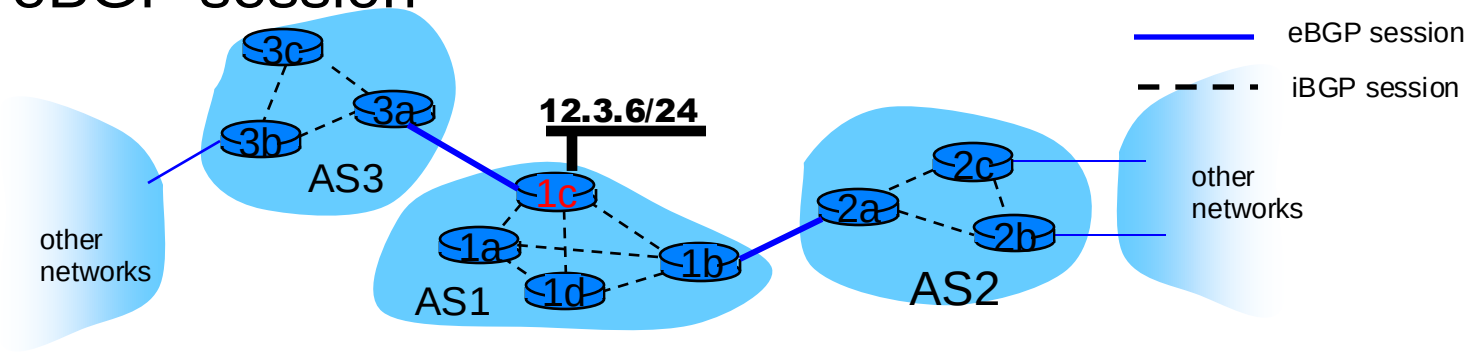
- 2 neighbor BGP routers establish a BGP session over a TCP connection(on port 179)to exchange routing updates
  - advertising routes to destination network prefixes
    - Route = prefix + attributes
- When AS3 router 3a advertises a prefix to AS1:
  - AS3 *promises* it will forward packets towards that prefix
  - AS1 runs local policies to further process



eBGP and iBGP

The problem: a BGP speaker learned reachability to some destination, how to inform other routers inside the same AS? Answer: use iBGP

- iBGP: BGP session between routers *in the same AS*
  - Router **1c** uses iBGP to distribute new prefix info (e.g. 12.3.6.0/24) to all routers in AS1
  - when router learns of new prefix, it creates entry for prefix in its forwarding table.
- eBGP: BGP session *between two different ASes*
  - e.g. the BGP session between 1b and 2a
  - router 1b may advertise 12.3.6.0/24 reachability to AS2 over this eBGP session



BGP messages

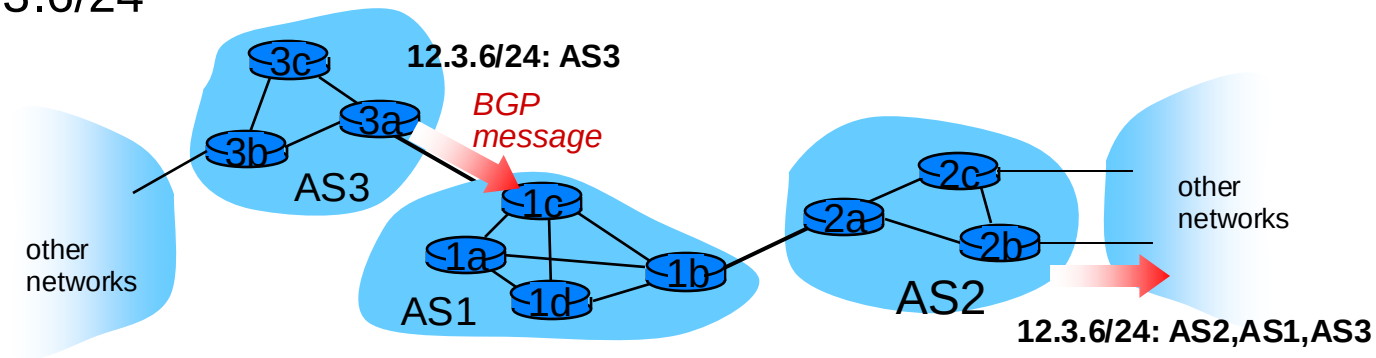
FYI

- Two BGP peers exchange routing messages over a 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 received BGP updates; also used to close connection

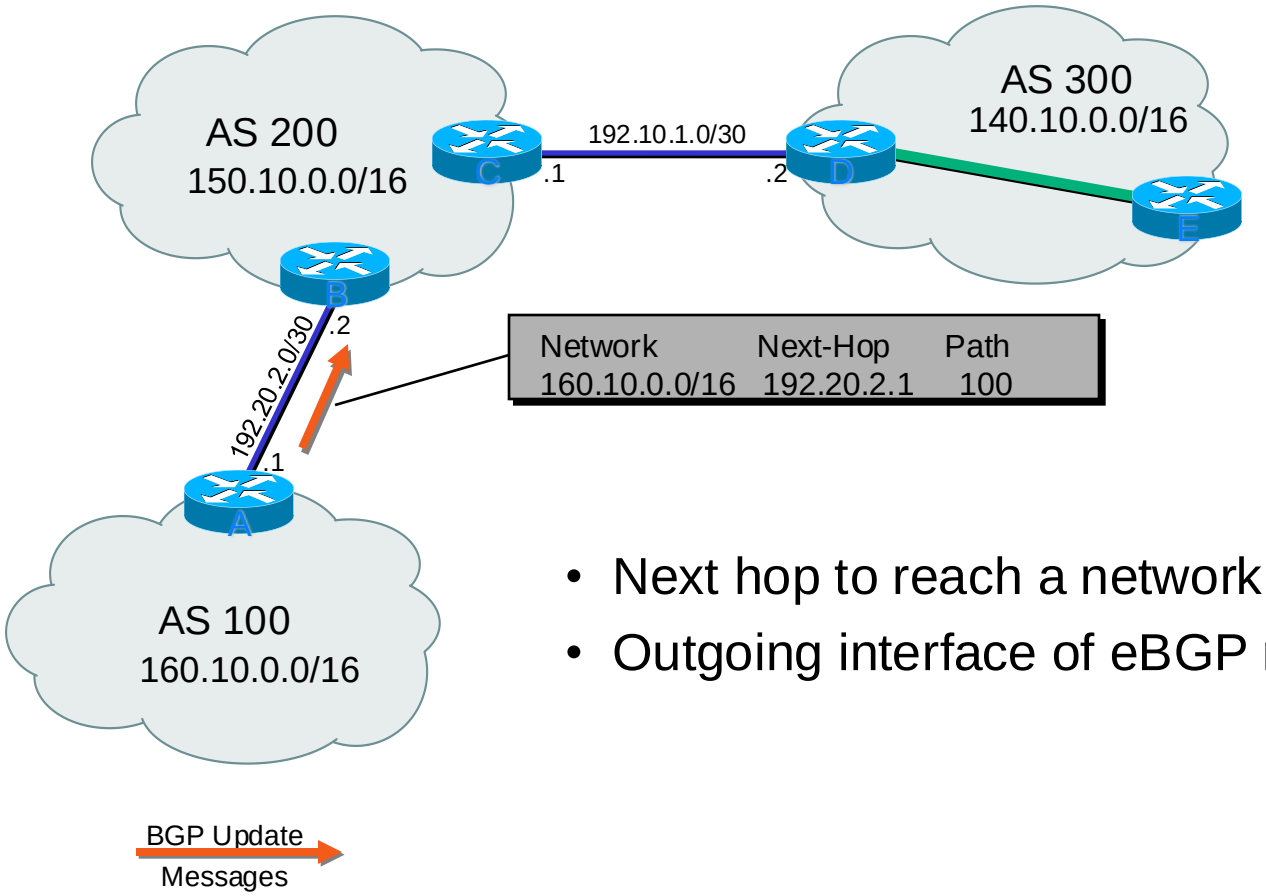
Path attributes and BGP routes

3 most important attributes:

- AS-PATH: a list of ASes, through which prefix advertisement has passed
  - When Router-C receives the announcement for prefix 12.3.6/24: AS-PATH value: [AS4, AS1]
- NEXT-HOP: indicates specific internal-AS router to next-hop AS
  - there can be multiple links from one AS to a neighbor AS
- Local-Preference: policy preference in path selection
  - border routers inject local-preference into received BGP updates
  - internal routers use it in path selection
    - e.g. deciding whether going through AS2 or AS4 to reach destination 12.3.6/24

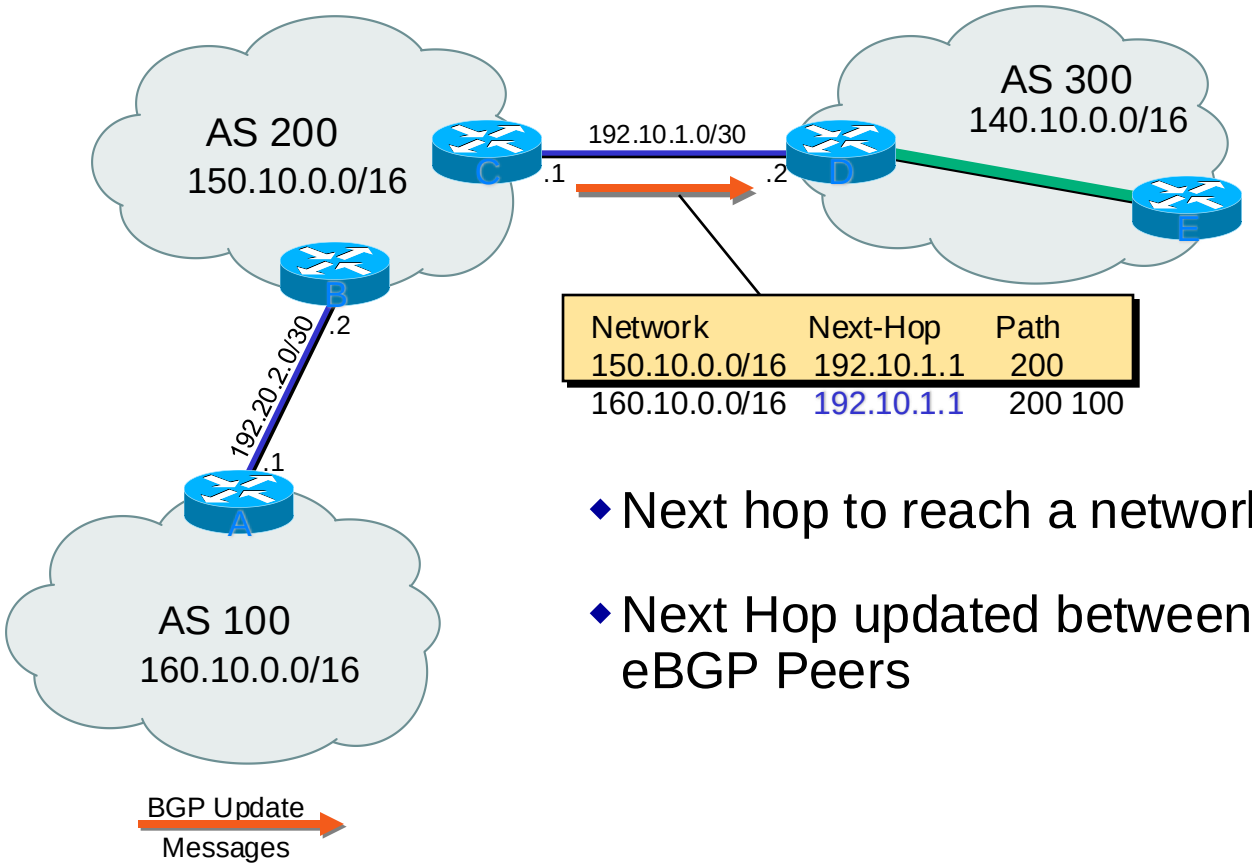


Next Hop Attribute



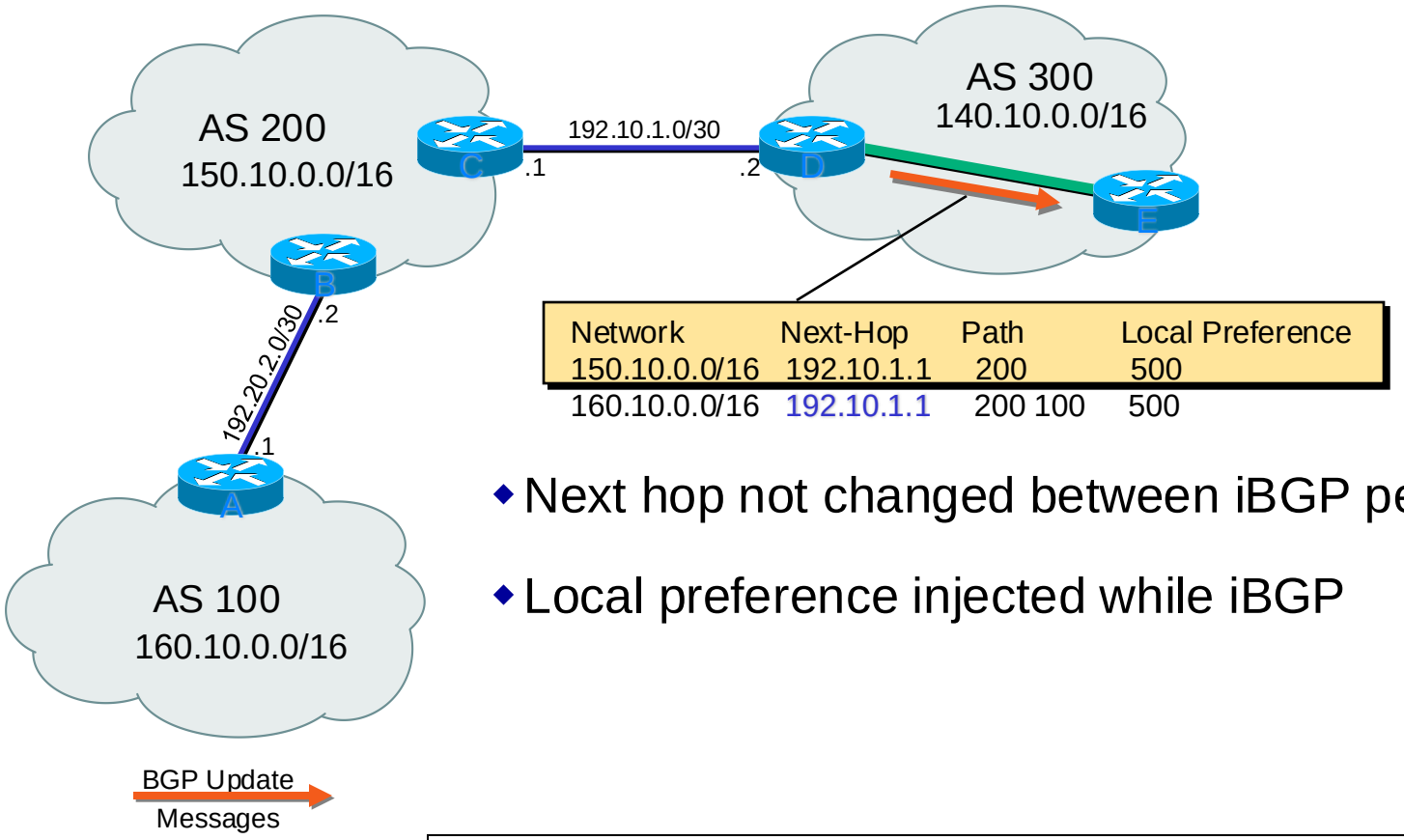
- Next hop to reach a network
- Outgoing interface of eBGP router

Next Hop Attribute



- Next hop to reach a network
- Next Hop updated between eBGP Peers

Next Hop and Local Preference Attribute



Network	Next-Hop	Path	Local Preference
150.10.0.0/16	192.10.1.1	200	500
160.10.0.0/16	192.10.1.1	200 100	500

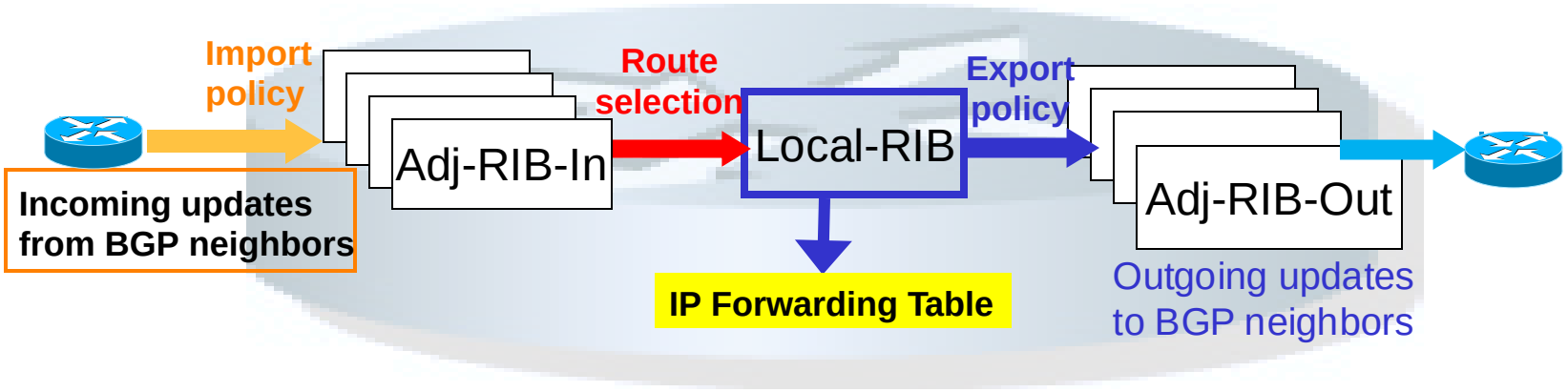
- Next hop not changed between iBGP peers
- Local preference injected while iBGP

Q: how internal OSPF routers reach Next-Hop?

Router reachability vs. prefix reachability

Path attributes & BGP Routing Policies

- Import policy: which paths to keep, which to drop?
  - Filter out unwanted routes from neighbor
- Route selection: among multiple routes to a given destination, pick *one* to use
- Export policy: tell which neighbor about which destination?
  - Filter out the routes you don't want to tell your neighbor



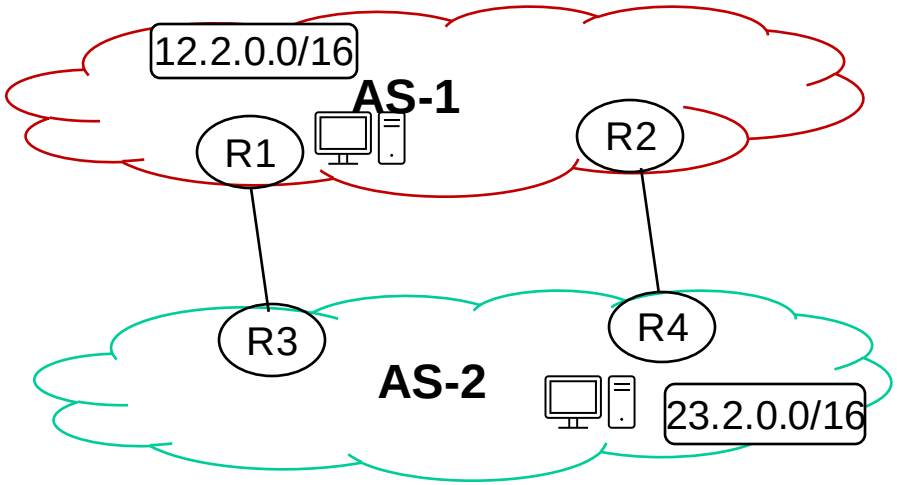
BGP route selection

- When an AS-internal router learns multiple routes to a destination prefix from iBGP: selects the *best Next-Hop* in the following order:
  1. local preference attribute: manually configured value according to AS policies
  2. If same local preference value for multiple routes: choose the one with shortest AS path
  3. Then Lowest IGP cost (hot potato routing)
  4. additional criteria

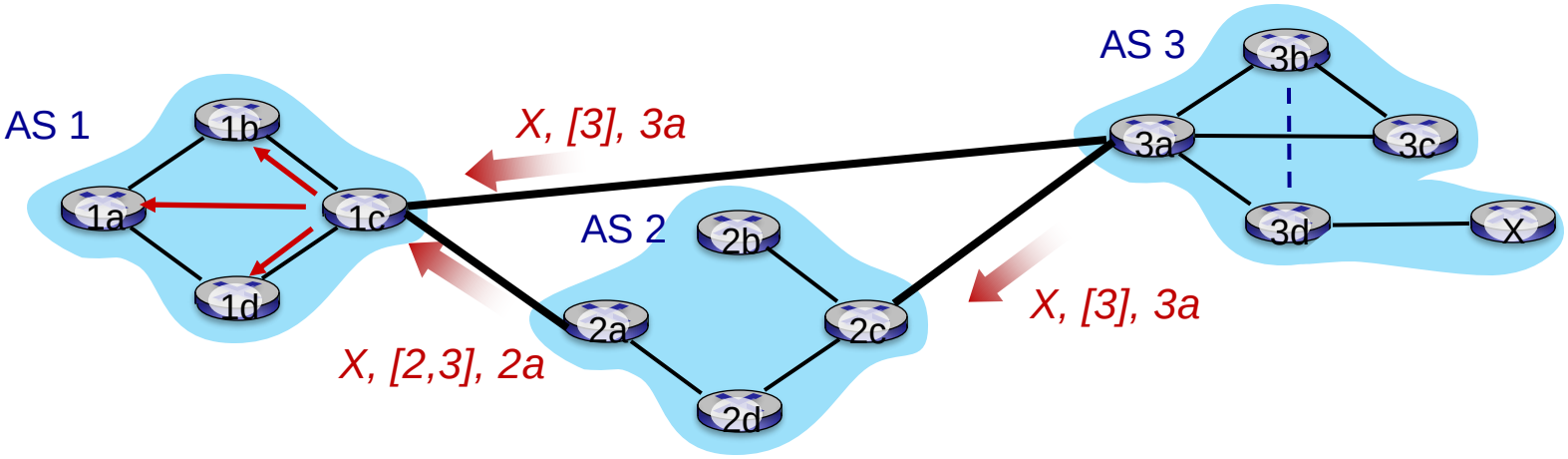
Hot potato routing: an example

- Host A 12.2.2.2 in AS-1 communicates with host B 23.2.1.1
- (If local preferences are the same)
- Packets from A to B path: going through R1-R3
  - AS-1 throws packets over to AS2 as quick as possible
- Packets from B to A path: going through R4-R2
  - AS2 does the same

Between the same pair of hosts, packets are likely forwarded through asymmetric paths



Putting pieces together



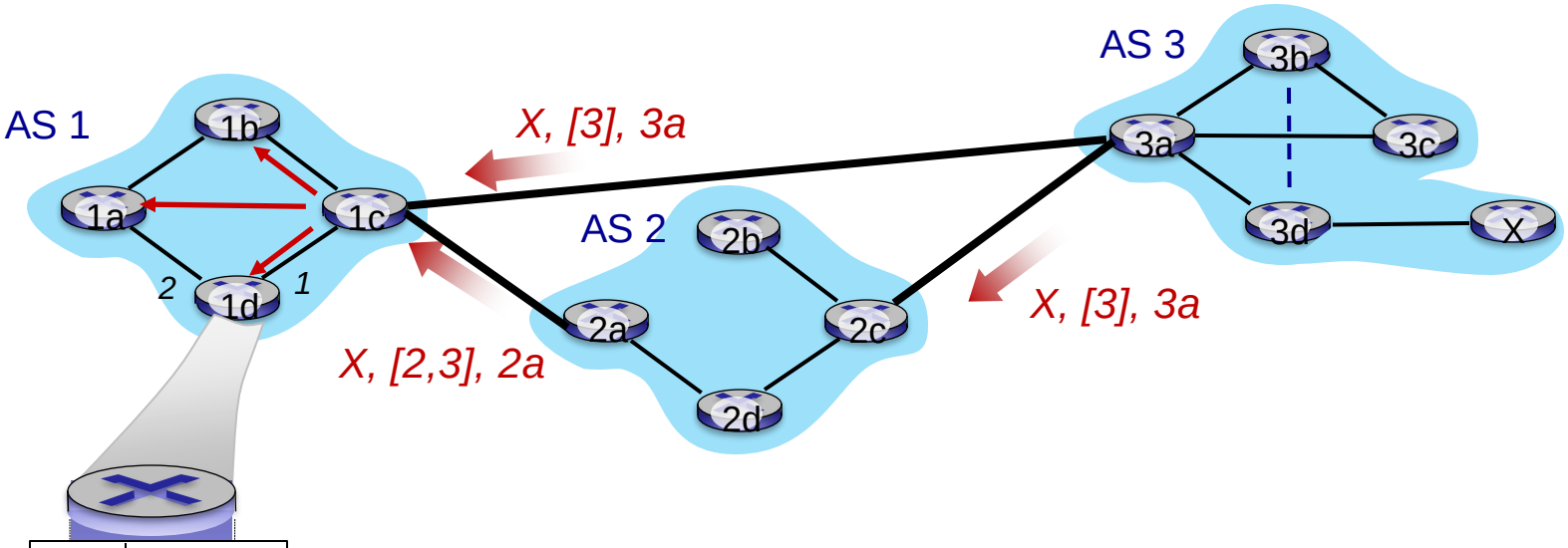
Gateway router may learn about multiple paths to destination:

- AS1 gateway router 1c learns path *AS2, AS3, X* from 2a
- AS1 gateway router 1c learns path *AS3, X* from 3a
- Based on *import policy (e.g., accept both but prefer AS2),* AS1 gateway router 1c advertises routes within AS1 via iBGP
  - ***Prefix: X***, AS-Path: [3], Next-Hop: 3a, Local Preference: 100
  - ***Prefix: X***, AS-Path: [2, 3], Next-Hop: 2a, Local Preference: 200

Putting pieces together

Prefix: X, AS-Path: [3], Next-Hop: 3a, Local Preference: 100

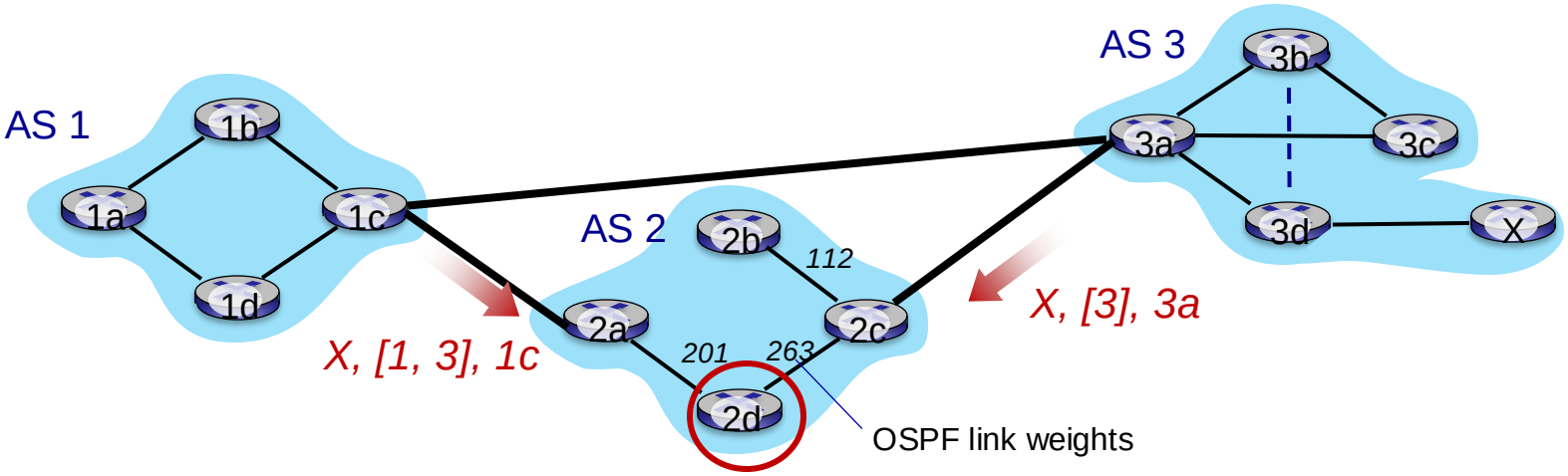
Prefix: X, AS-Path: [2, 3], Next-Hop: 2a, Local Preference: 200



dest	interface
…	…
1c	1
3a	1
2a	1
X	1
…	…

- recall: 1a, 1b, 1d learn via iBGP from 1c: “path to X goes through 3a” “path to X goes through 2a”
- at 1d: OSPF intra-domain routing
  - 1c LSA: I can reach 2a and 3a
  - To get 1c, use interface 1
- at 1d: to get to X, use interface 1

Putting pieces together



2c import policy: accept, set preference 100

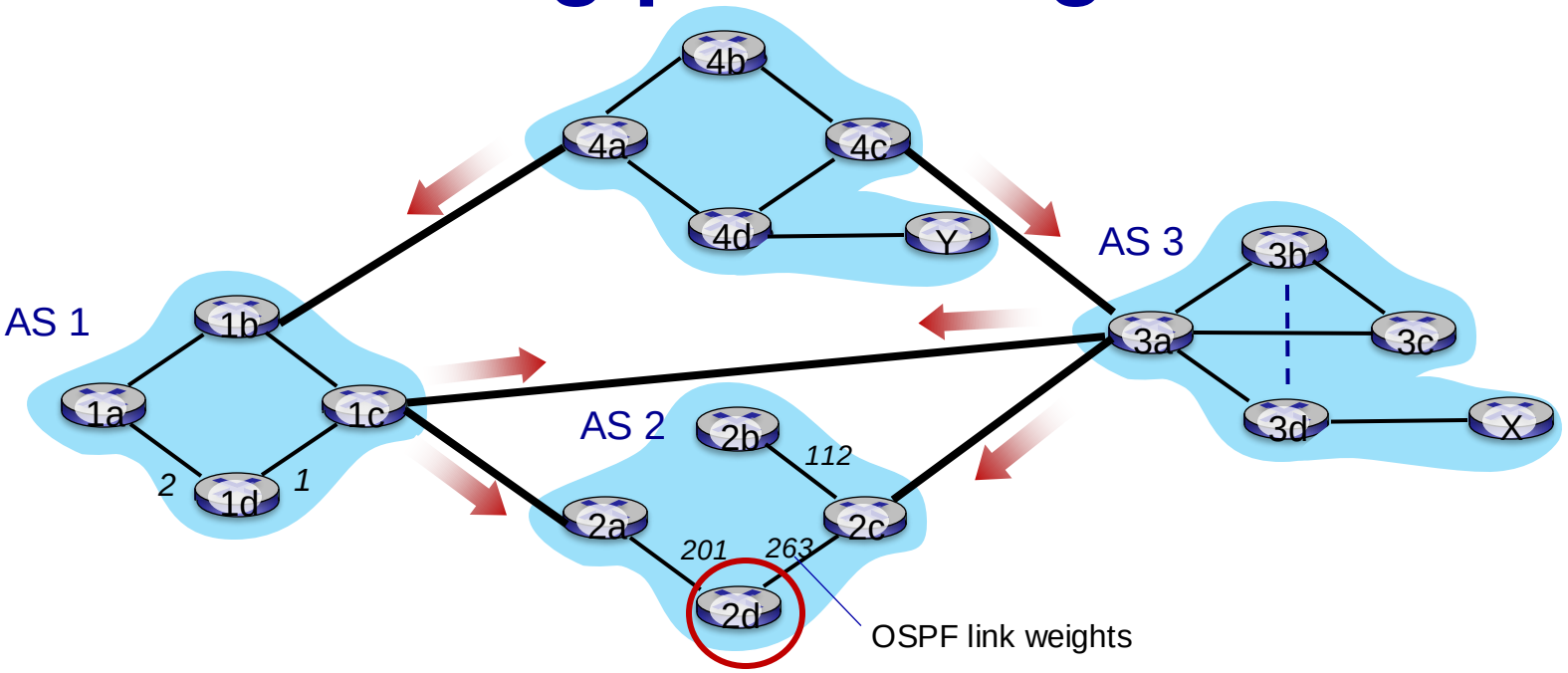
2a import policy: accept, set preference 100

Prefix: X, AS-Path: [3], Next-Hop: 3a, Local Preference: 100

Prefix: X, AS-Path: [1, 3], Next-Hop: 1c, Local Preference: 100

Q: Does hot potato routing happen?

Putting pieces together



2c import policy: accept, set preference 100

2a import policy: accept, set preference 100

Prefix: Y, AS-Path: [3, 4], Next-Hop: 3a, Local Preference: 100

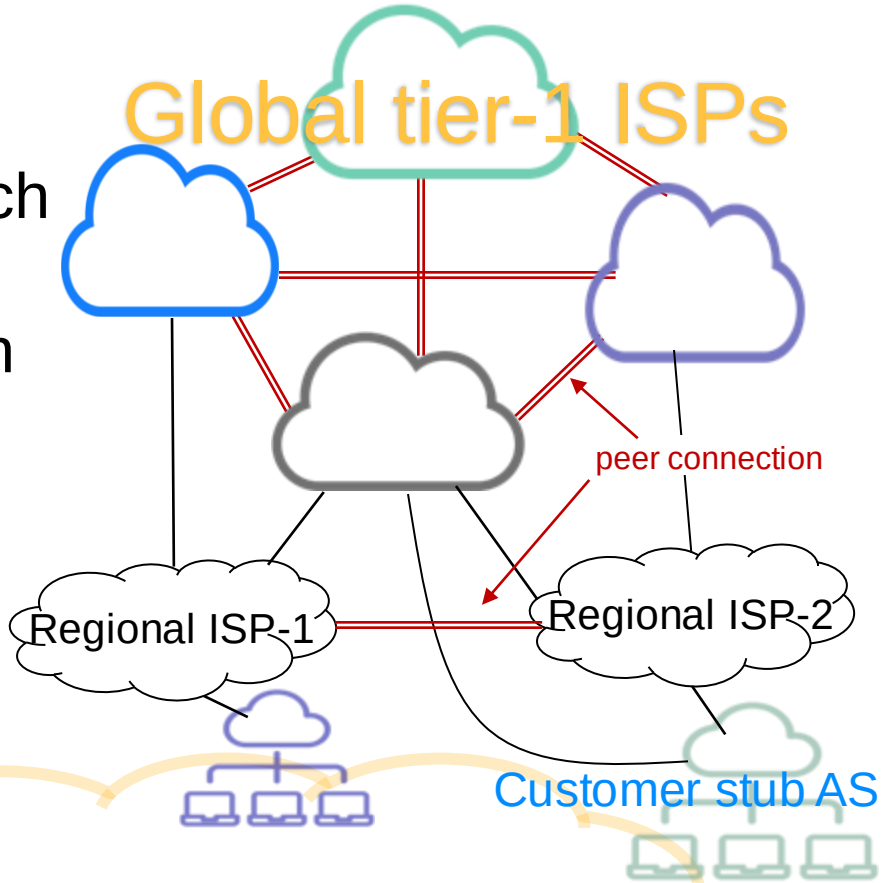
Prefix: Y, AS-Path: [3, 1, 4], Next-Hop: 3a, Local Preference: 100

Prefix: Y, AS-Path: [1, 4], Next-Hop: 1c, Local Preference: 100

Prefix: Y, AS-Path: [1, 3, 4], Next-Hop: 1c, Local Preference: 100

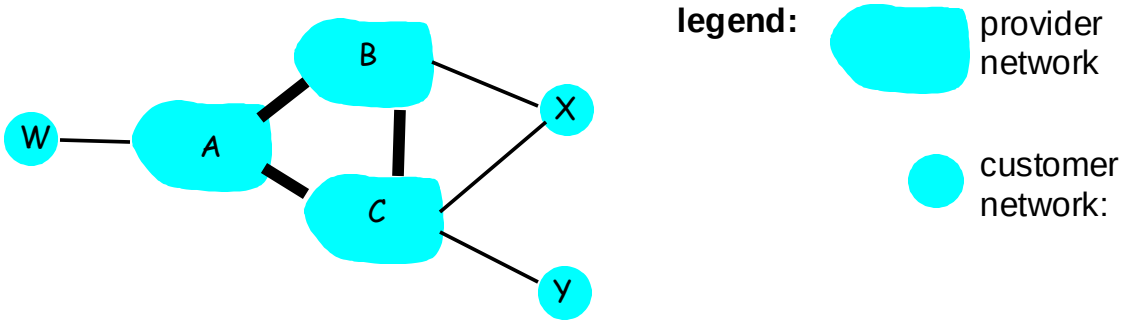
Internet AS interconnects

- Tier-1 Internet service providers
  - Full-mesh connected with each other
  - No has provider; peer-relation with each other
- Regional ISPs
  - Customers of tier-1 ISPs
  - May peer with other regional ISPs
- Customer stub networks
  - Multihomed in general
  - A special type of customer network: Super-giants



BGP: export policy in routing advertisements

important



a provider passes all prefixes to its customer ASes;
a customer does not pass prefixes between providers

“no valley” routing policy

A,B,C are provider network ASes

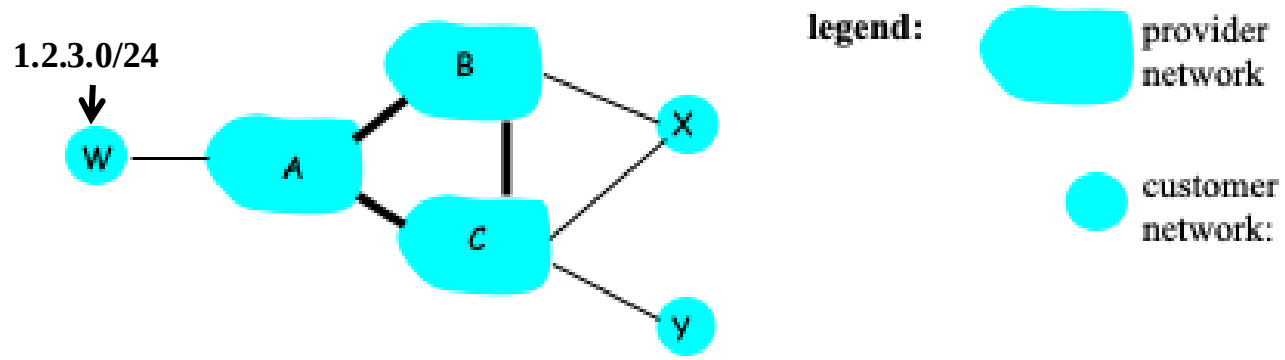
X,W,Y are customer ASes (of provider networks)

X is dual-homed: attached to two provider networks

- X does not want to forward traffic from B to C
- .. so X will not advertise to B any route it learned from C

BGP routing policy: a provider only propagates customers routes to peers

important



- a provider passes all prefixes to its customer ASes;
  - a customer *must not* pass prefixes between providers
- a provider does not pass prefixes that are not its clients' to other provider

A advertises to B the path **[1.2.3.0/24: A-W]**

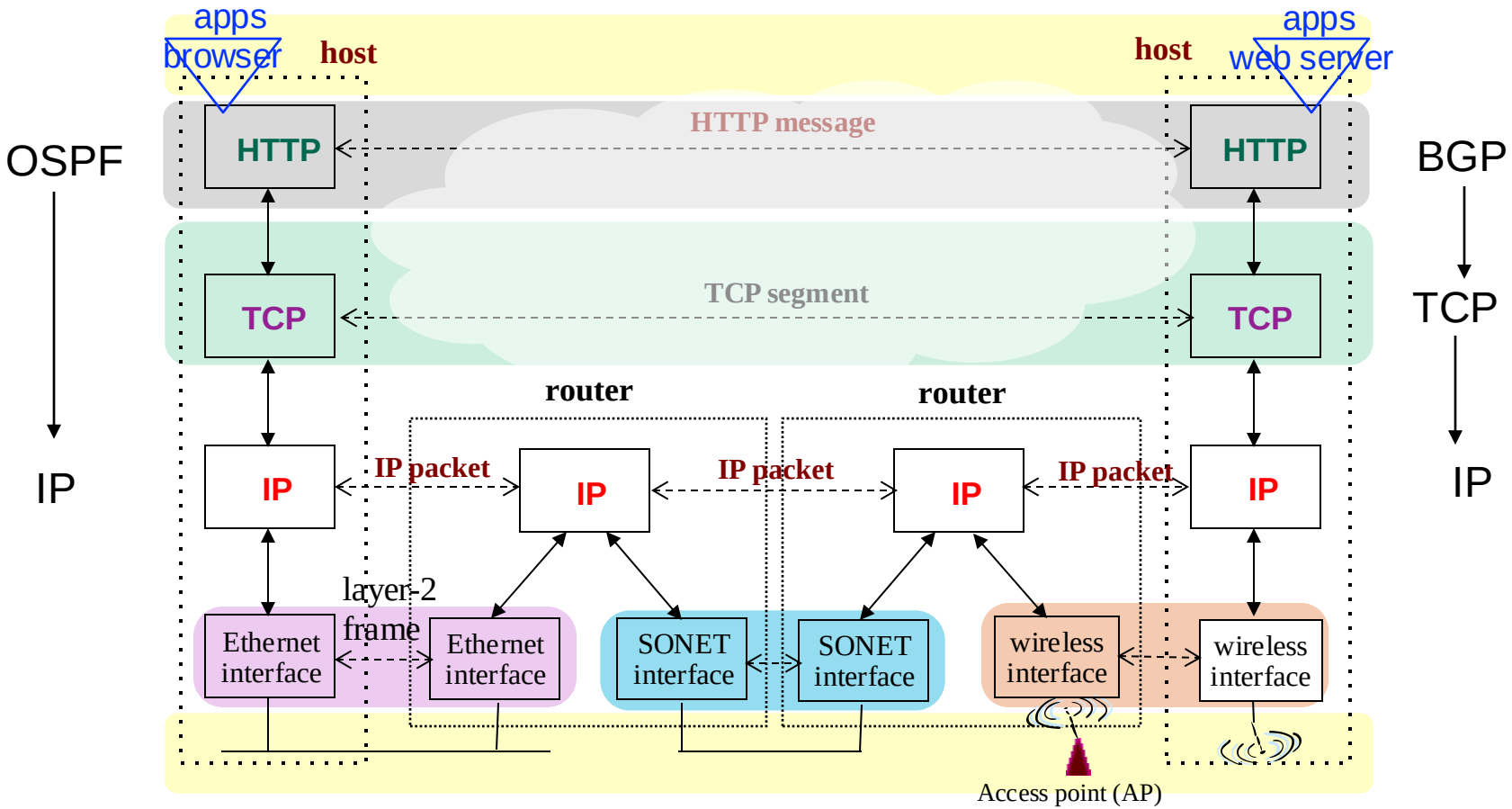
B advertises to X the path **[1.2.3.0/24: B-A-W]**

- B does not advertise to C the path **[1.2.3/24: C-B-A-W]:** neither W nor C is B's customer

Why different Intra- and Inter-AS routing ?

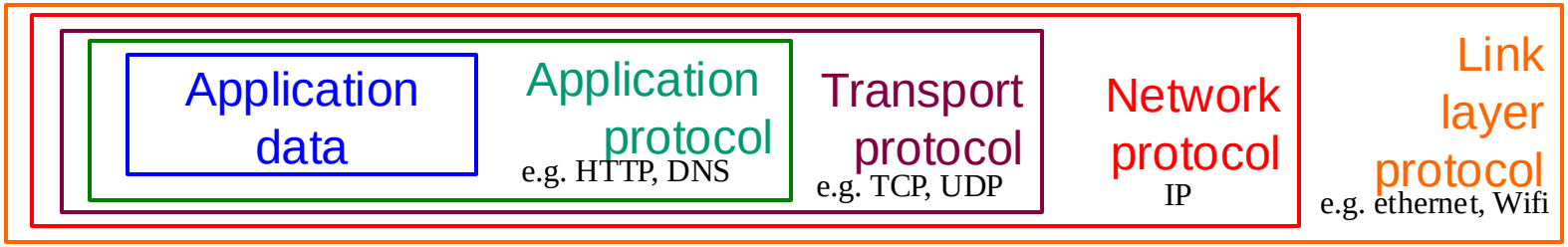
- Policy:
  - Inter-AS: admin wants control over how its traffic routed, who routes through its net.
  - Intra-AS: single admin, so no policy decisions needed
- Scale:
  - hierarchical routing saves table size, reduced update traffic
- Performance:
  - Intra-AS: can focus on performance
  - Inter-AS: policy may dominate over performance

Always keep the big picture in mind



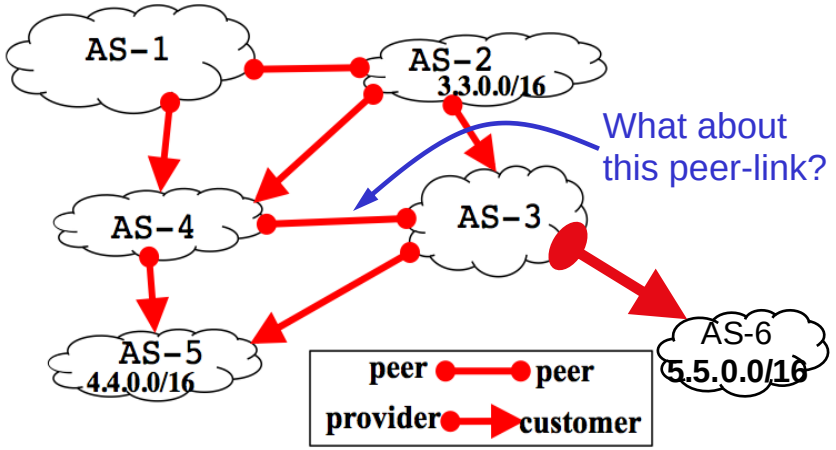
Where are the routing protocols in this layered protocol picture?

- The computation results from routing protocol are installed to the IP layer



Practice Question 1: No-Valley Routing Policy

- to reach destination prefix 3.3.0.0/16 in AS-2
  - List all valid path(s) AS-1 can take
  - List all valid paths that AS-5 can take.
- Considering the reachability to destination prefix 4.4.0.0/16,
  - List *all* the valid path(s) that AS-1 can take.
  - Among these path(s), which valid path does AS-1 prefer the most? Following BGP route selection policy



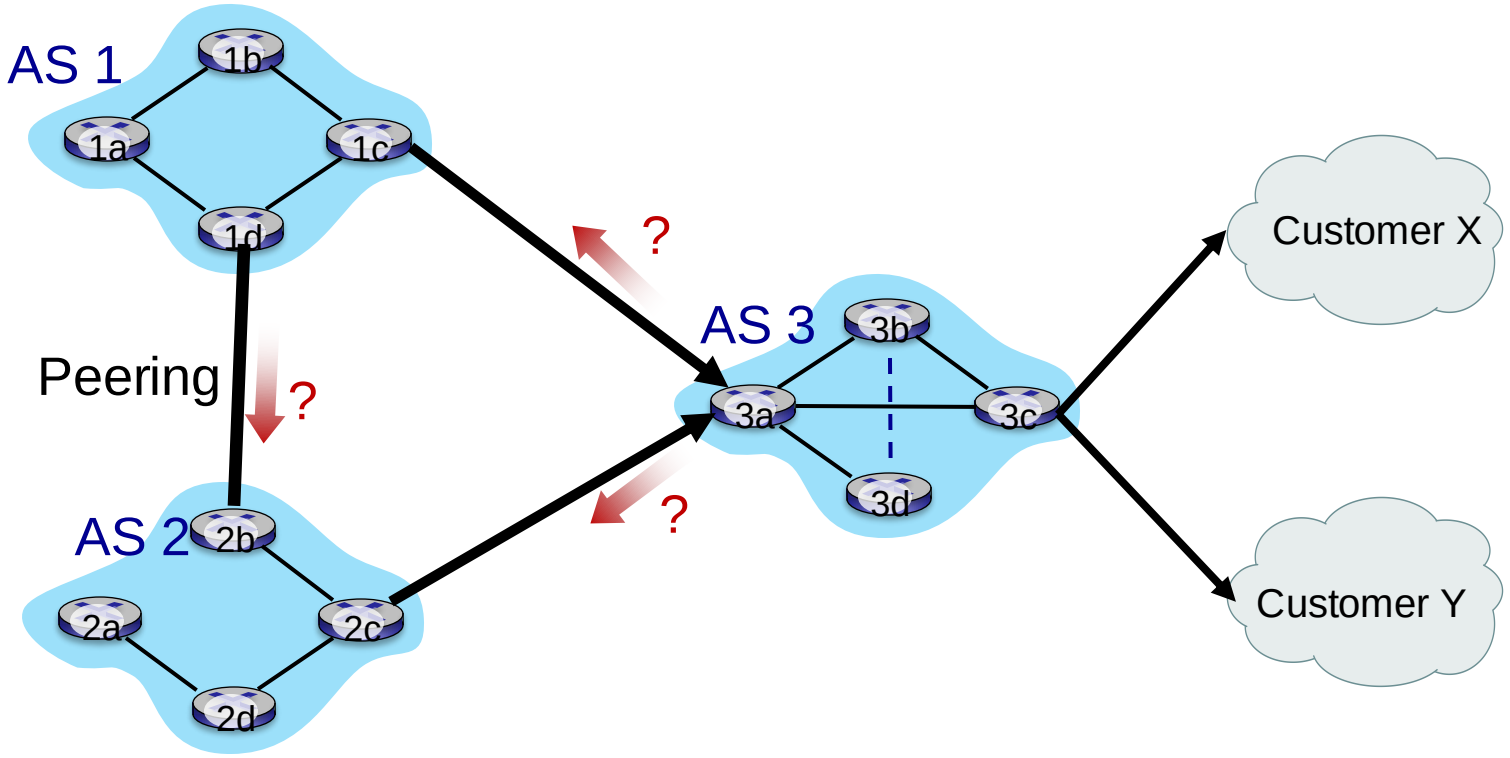
AS-1 to 3.3.0.0/16: all valid paths
- AS-1 → AS-2

AS-5 to 3.3.0.0/16: all valid paths
- **5**-3-2
- 5-4-2
- 5-4-1-2

AS-1 to 4.4.0.0/16: all valid paths
- 1-4-5
- 1-2-4-5
- 1-2-3-5

Practice Question 2: customized policy

- Customer X traffic comes from AS 1 only
- Customer Y traffic comes from AS 2 only



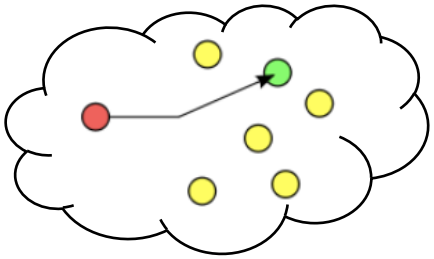
A Quick Summary of Routing Protocols

- OSPF: a link-state routing protocol
  - Each router sends Link-State Packet containing
    - ID of the node that created the LSP + seq# for this LSP
    - a list of direct neighbors, with link cost to each
    - time-to-live (TTL) for information carried in this LSA
  - LSAs are sent periodically, or whenever changes happen
    - flooded everywhere, reliably
  - Neighbor routers use Hello msgs to keep track each other
- BGP: a path-vector routing protocol
  - Running over TCP connection
  - Propagate reachable IP prefixes under the constraints of policies
  - Two types of AS relations: peer-peer, customer-provider
  - BGP routing policy model:
    - A route can have no more than one peer-peer link
    - No valley policy: in an AS route, a provider→customer link can only be the last link, or followed by more provider→customer link

Datagram delivery

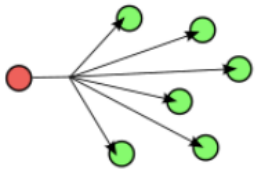
FYI

Unicast Focus



Unicast: A given IP address block ***A*** is announced from a single location

Broadcast



Broadcast: if a packet's destination IP is broadcast, send it everywhere

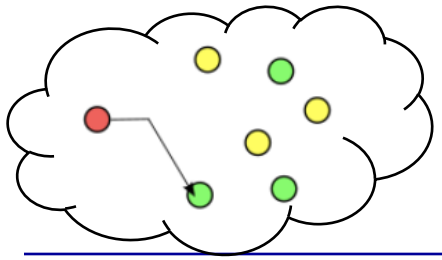
- 255.255.255.255 ==> "this link-local subnet"
- 131.79.196.255 ==> broadcast for 131.79.196.0/24

Multicast

Multicast: an IP multicast address represents a group of recipients

- **224.0.0.5** to represent **link-local** multicast group, Link-local multicast does not go beyond subnet

Anycast

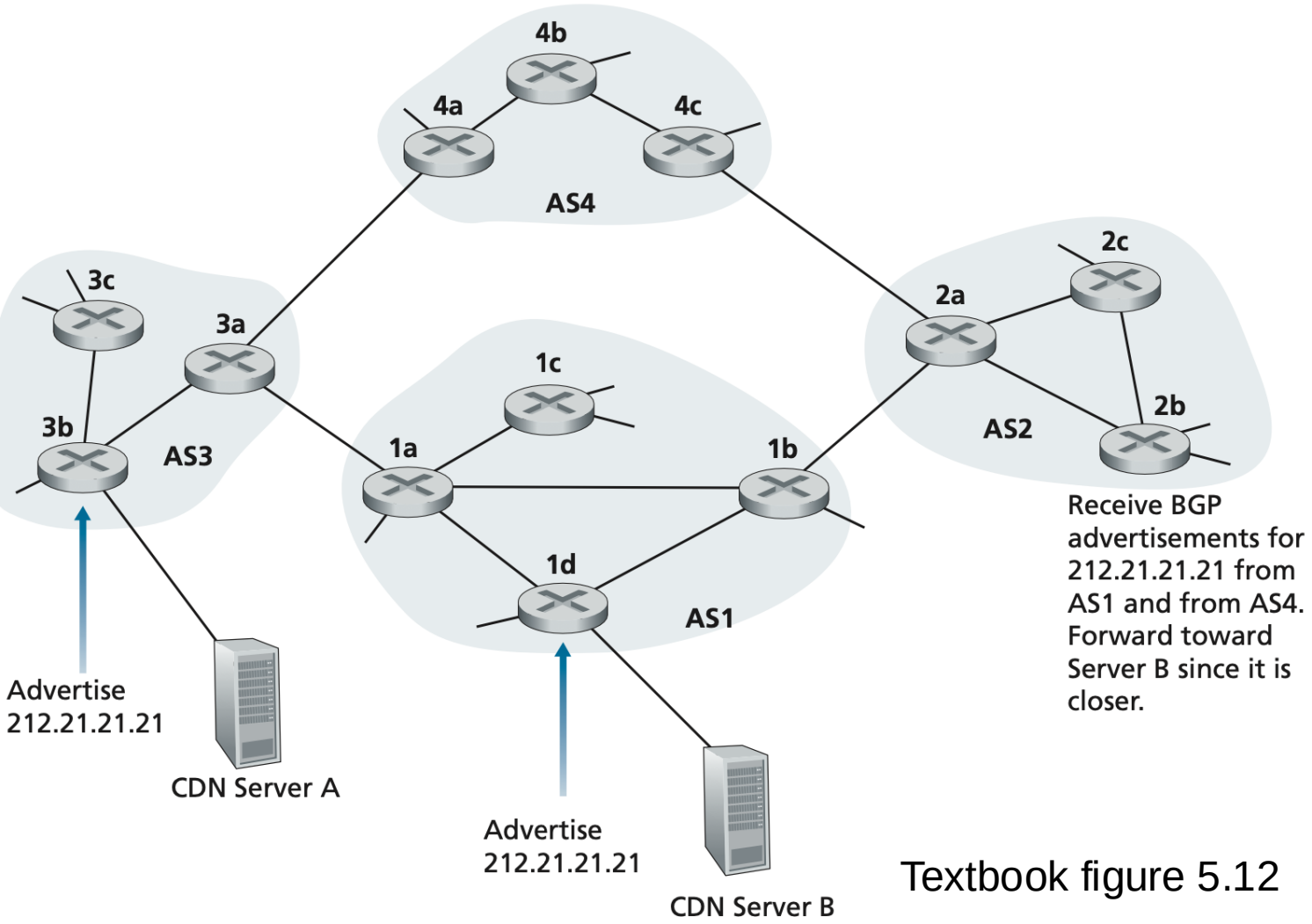


Anycast: A given IP address block ***A*** is announced from multiple locations

- a route receives reachability to ***A*** from multiple neighbors, pick the shortest path to forward packets
- 8.8.8.8

BGP to implement IP anycast

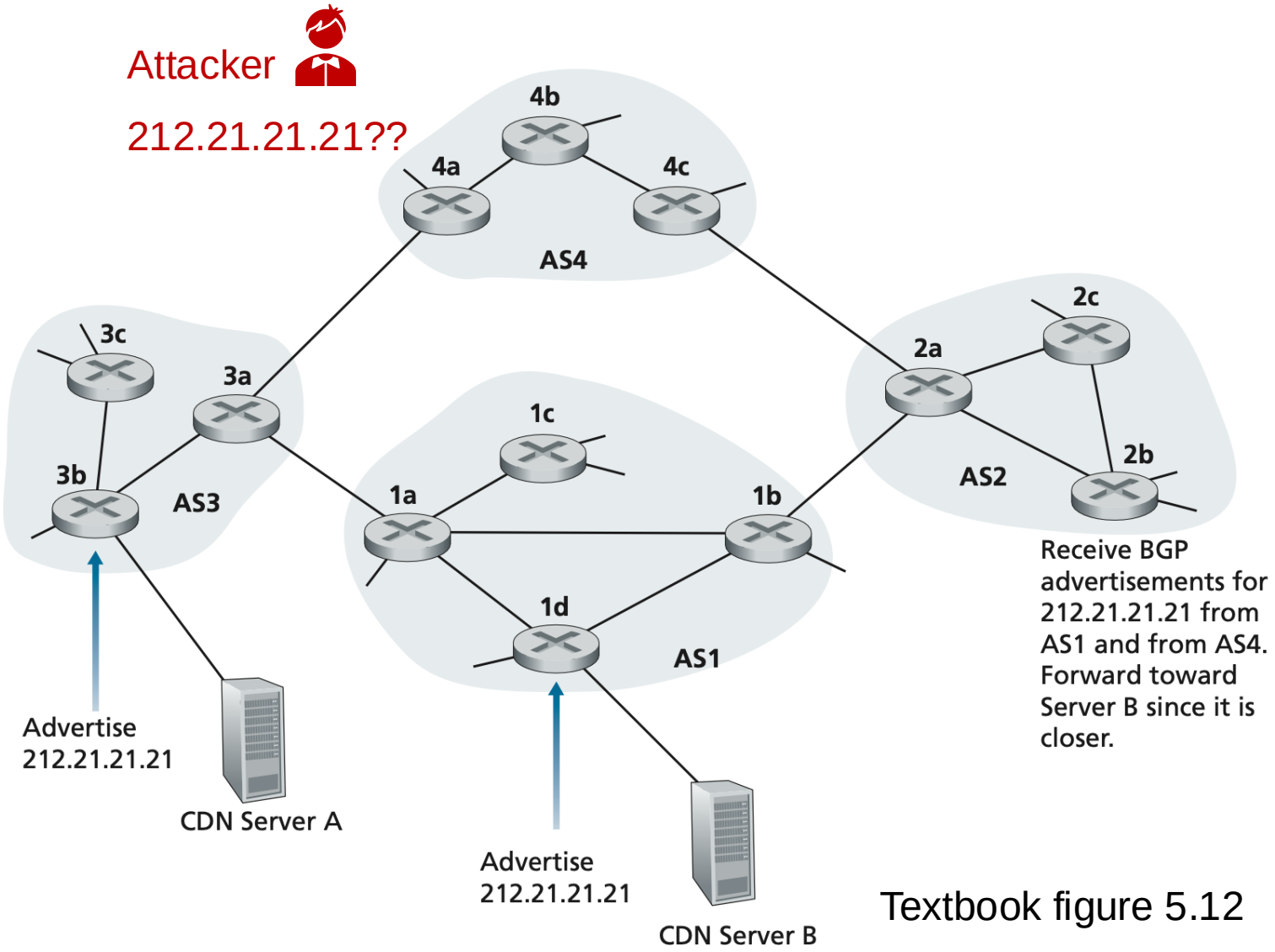
FYI



Textbook figure 5.12

Do you trust the prefix?

FYI



Textbook figure 5.12

It happened...BGP Hijacking

FYI

- (2008) Pakistan Telecom (AS 17557) attempted to block Youtube (AS 36561)
  - Real prefix: 208.65.152.0/22
  - Internally, re-routing 208.65.153.0/24 to its other customers (likely to be blackholes)
  - By accident, the new routes were announced to upstream providers (AS 3491, a tier-1 ISP), and from there broadcast to the whole Internet.
  - BGP prefers more specific prefix
    - 2/3 of Internet was sending youtube traffic to Pakistan Telecom

BGP operates on trust…