Inter-AS Routing in the Internet: BGP

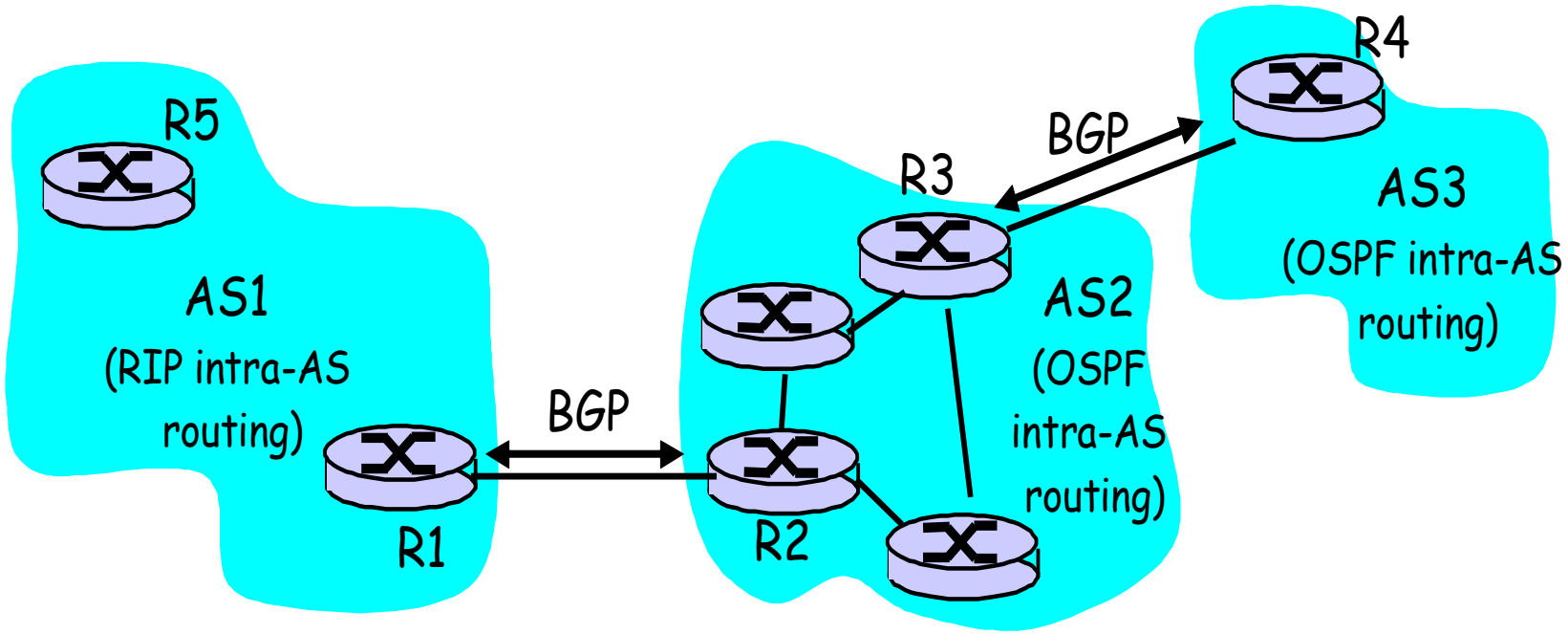


Figure 4.5.2-new2: BGP use for inter-domain routing

BGP (Border Gateway Protocol)

- *The* de facto standard (BGP-4)
- **Path Vector** protocol:
  - similar to Distance Vector protocol
  - each Border Gateway broadcast to neighbors (peers) *entire path* (i.e., sequence of ASes) to destination
  - BGP routes to networks (ASes), not individual hosts
- E.g., Gateway X may announce to its neighbors it "knows" a (AS) path to a *destination network*, Z, via *a series of ASes*:

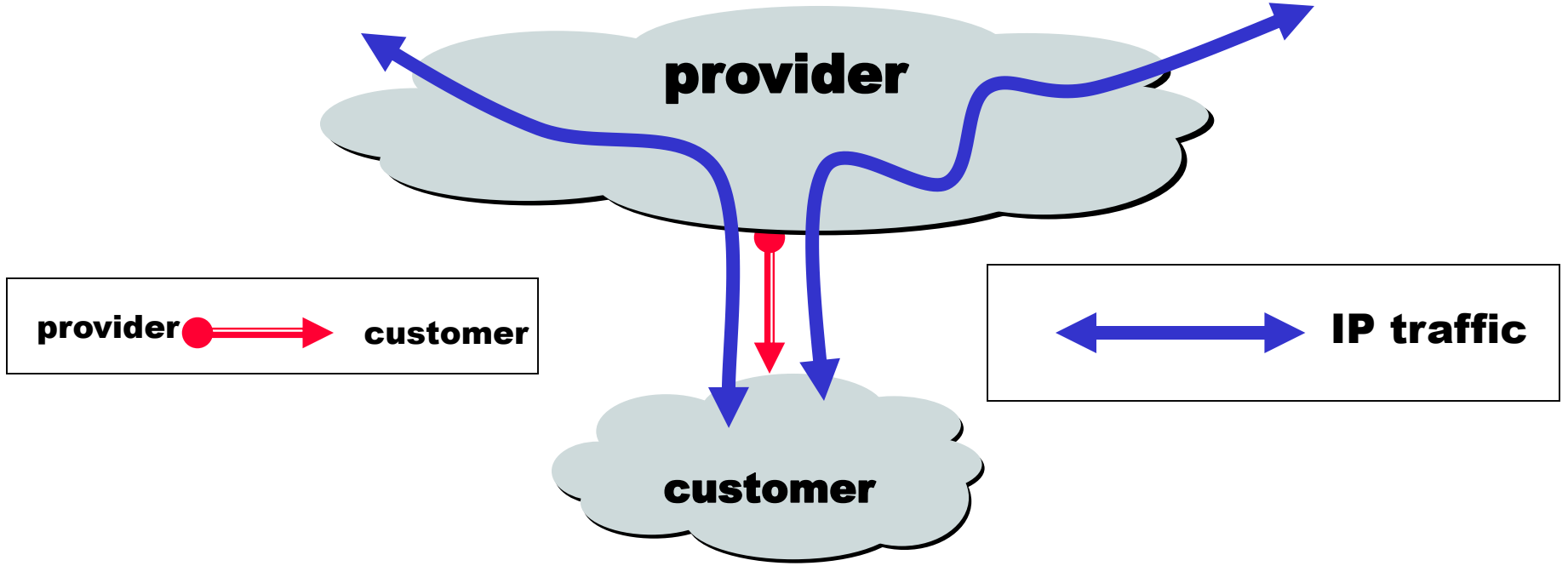
  Path (X,Z) = X,Y1,Y2,Y3,…,Z
- BGP border gateways referred to as BGP *speakers*

BGP Operations: Policy Routing

Q: What does a BGP border gateway do?

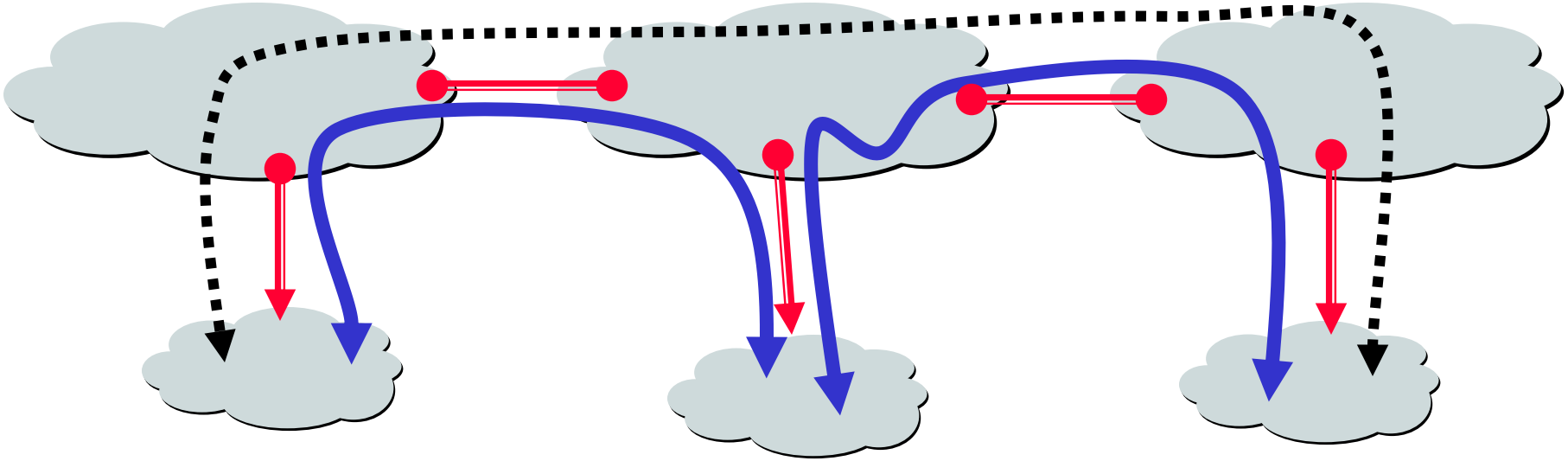
- Receiving and *filtering* route advertisements from directly attached neighbor(s)
  - To accept or not accept route advertisements depends on policies (e.g., whether you "trust" your neighbors)
- Route selection (rank diff. routes to same dest. network).
  - to route to destination X, which path (of several advertised) will be taken?
  - route selection based on policies (e.g., always prefer route advertisement from "good old" neighbor Y)
- Filtering and sending (certain) route advertisements to neighbors
  - what/whether to advertise to your neighbors also depends on policies (e.g., don't tell your neighbor Z that you know a route to destination X)

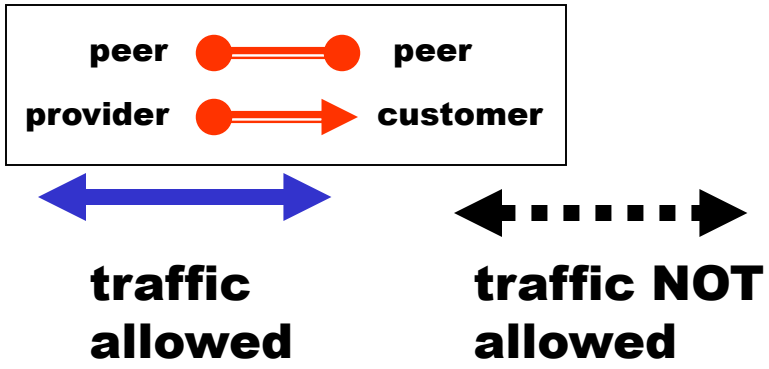
Customers and Providers



Customer pays provider for access to the Internet

The Peering Relationship



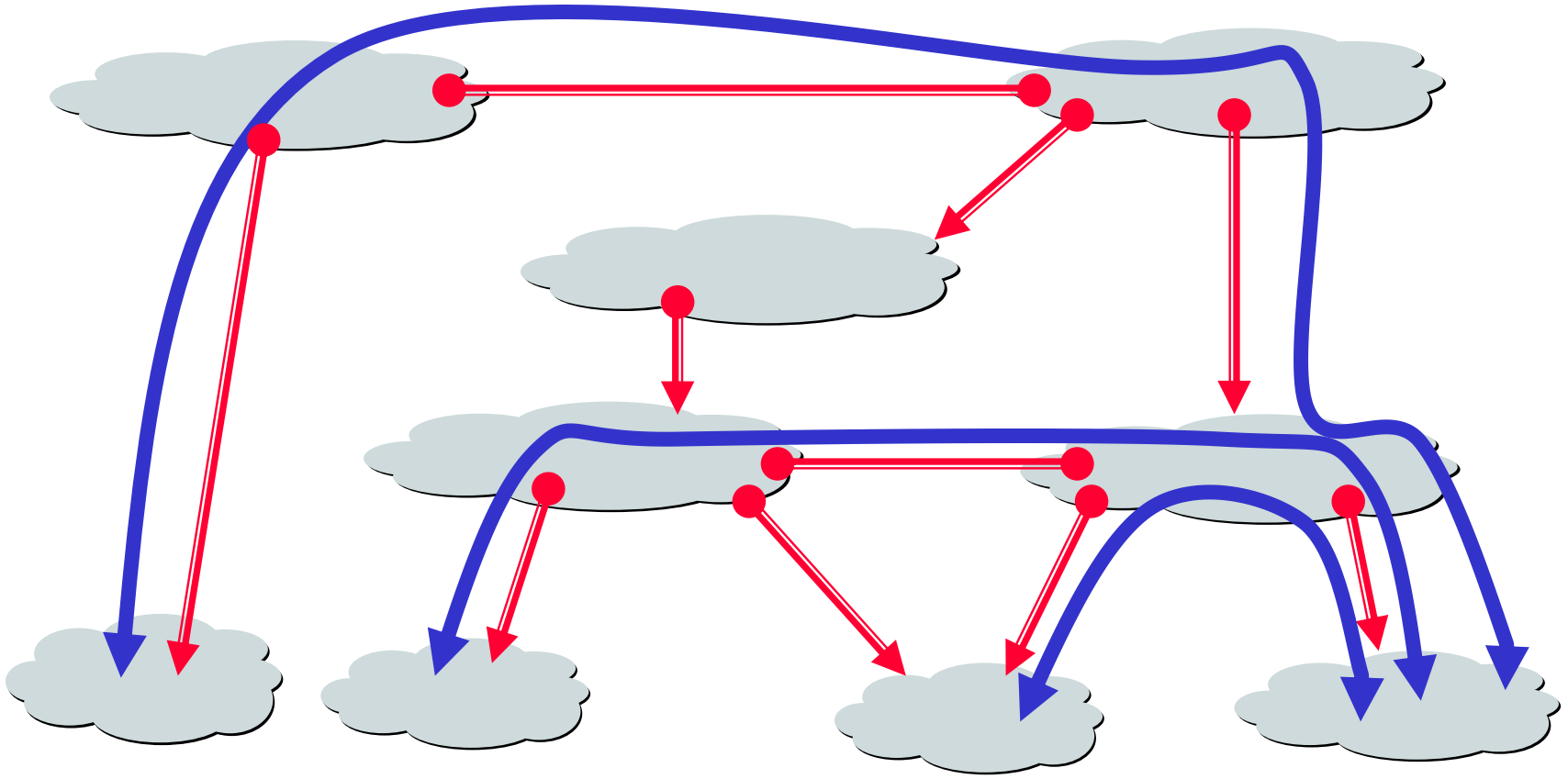


Peers provide transit between their respective customers

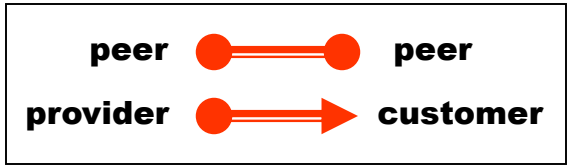
Peers do not provide transit between peers

Peers (often) do not exchange $$$

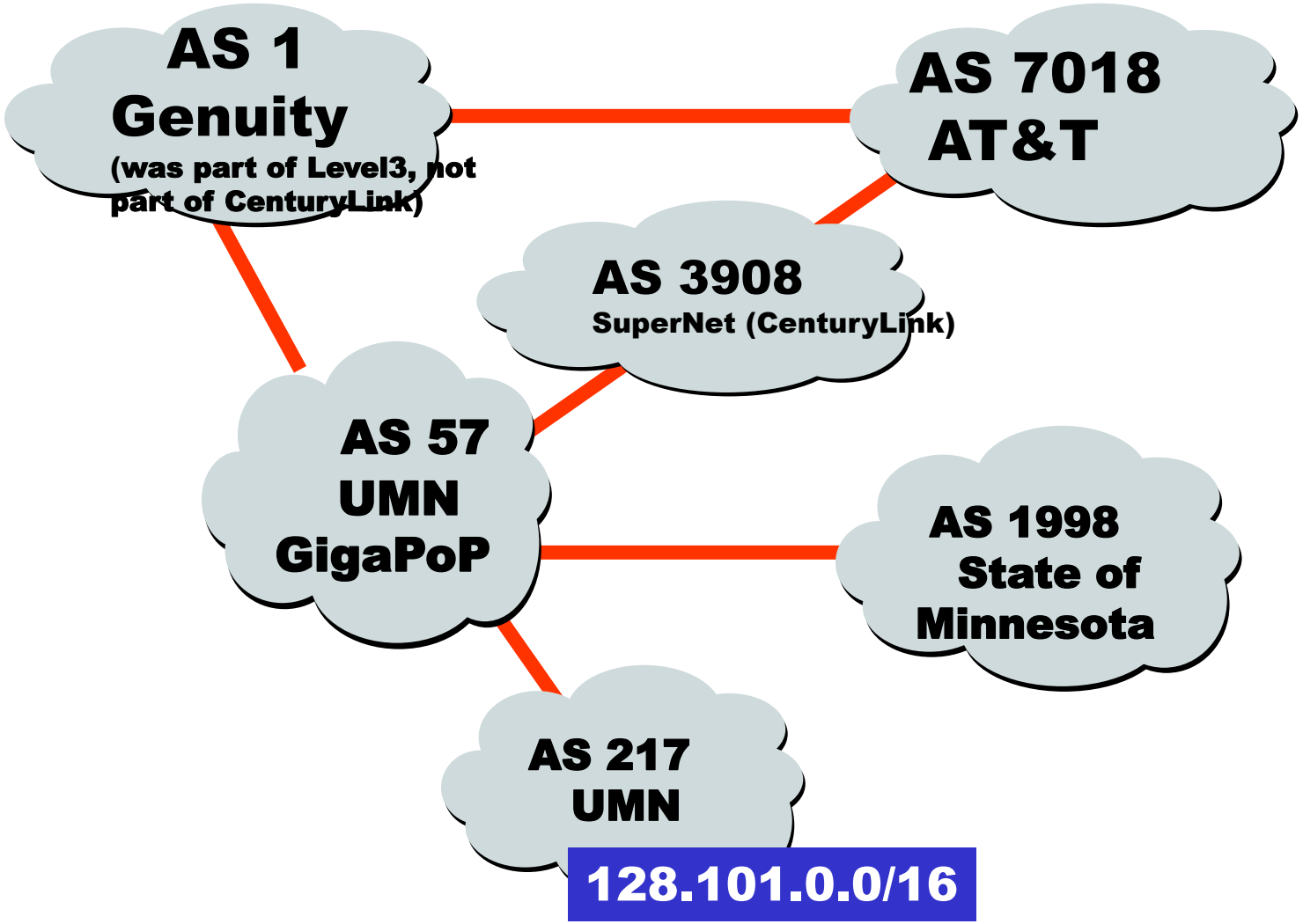
Peering Provides Shortcuts



Peering also allows connectivity between the customers of "Tier 1" providers.



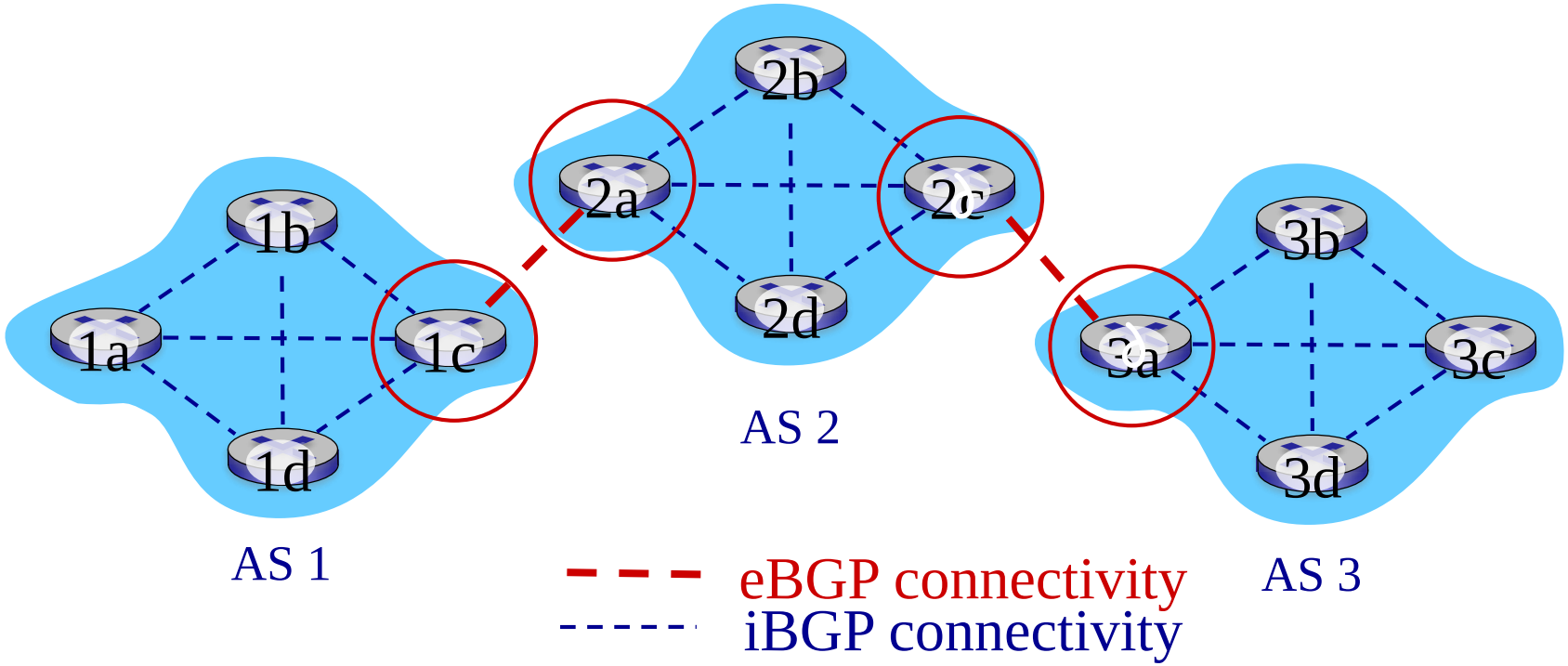
U of Minnesota (*Old AS*) Neighborhood



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 to rest of Internet: *"I am here" (network reachability)*
- BGP provides each AS a means to select a route:
  - eBGP: obtain subnet reachability information and available routes from neighboring ASes
  - iBGP: propagate reachability information and available routes to all AS-internal routers.
  - determine "good" routes to other networks based on reachability information, available routes and *policy*

eBGP, iBGP Connections



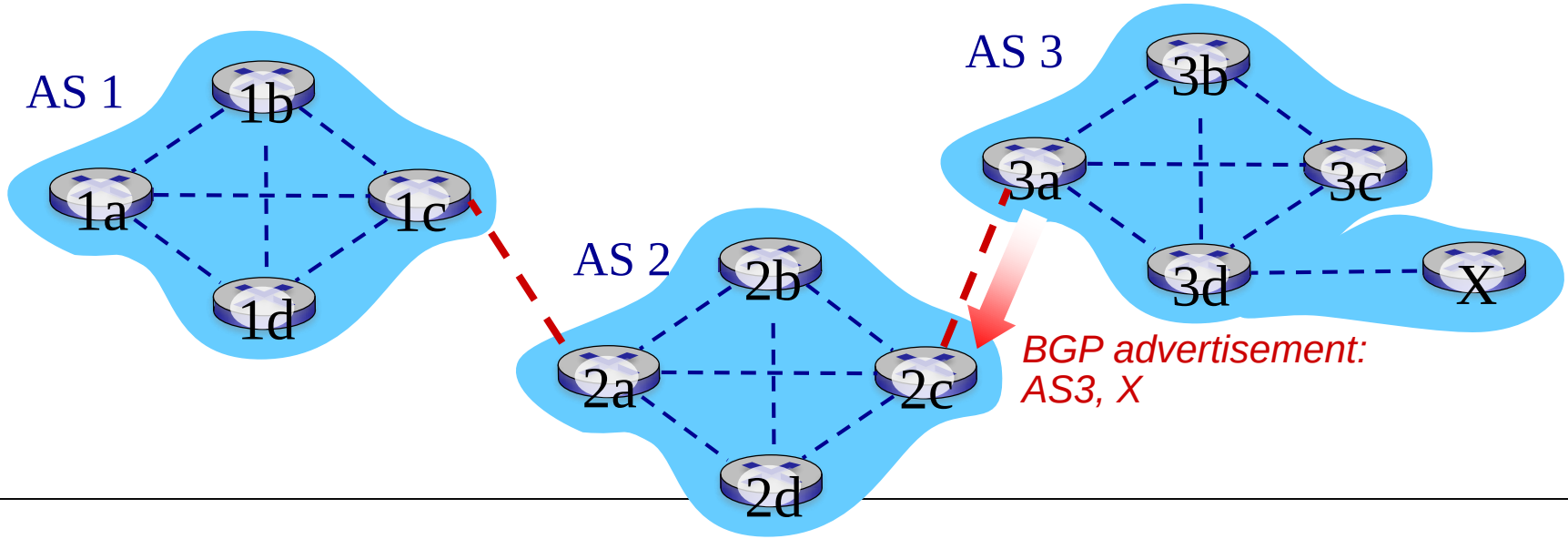
gateway routers run both eBGP and iBGP protocols

BGP Messages

- BGP messages exchanged using TCP.
- BGP messages:
  - OPEN: opens TCP connection to peer and authenticates sender
  - KEEPALIVE keeps connection alive in absence of UPDATES; also ACKs OPEN request
    - OPEN/KEEPALIVE establish & maintain BGP neighbor relation
  - UPDATE: advertises new path (or withdraws old)
  - NOTIFICATION: reports errors in previous msg; also used to close connection

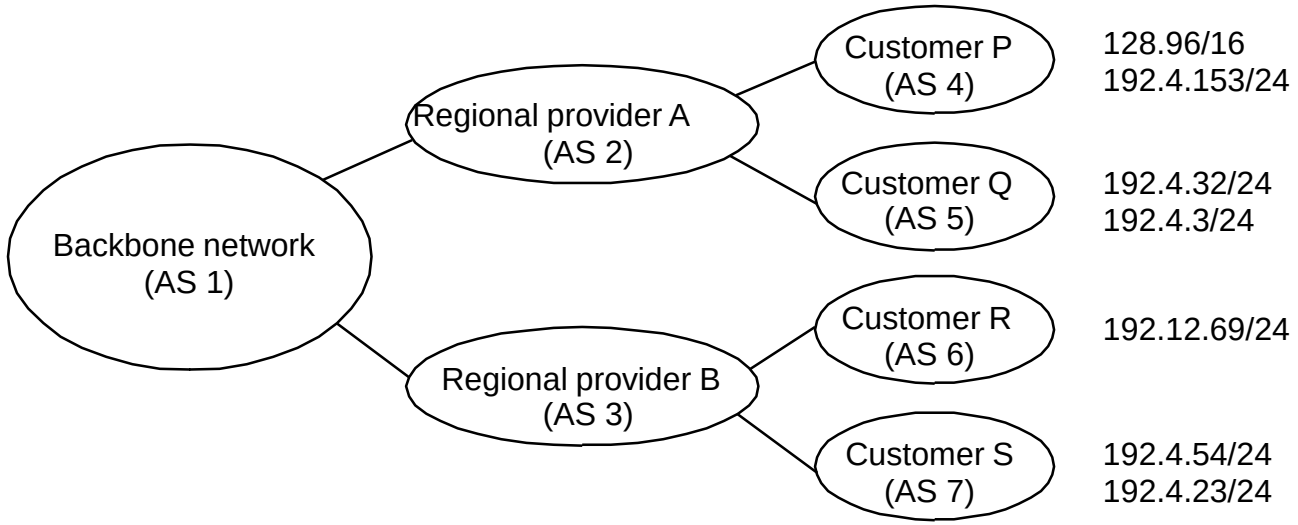
BGP Basics

- 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 Example

- Speaker for AS2 advertises reachability to P and Q
  - **network 128.96/16, 192.4.153/24, 192.4.32/24, and 192.4.3/24, can be reached directly from AS2**



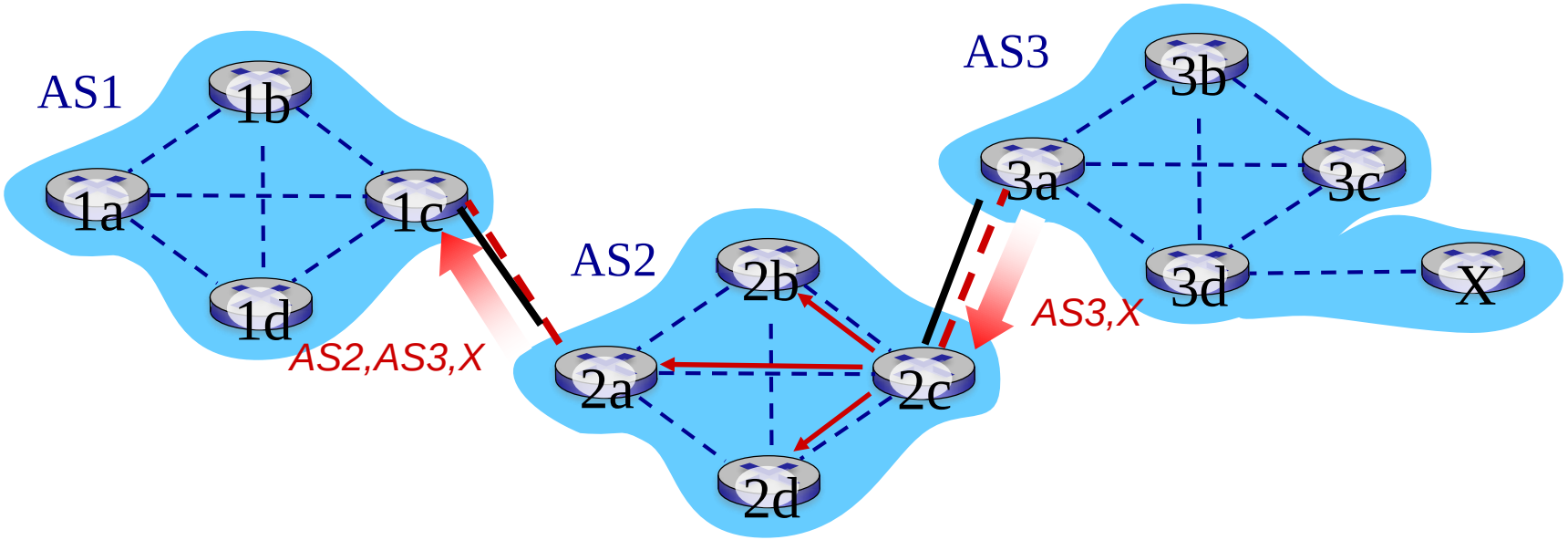
- Speaker for backbone advertises
  - **networks 128.96/16, 192.4.153/24, 192.4.32/24, and 192.4.3/24 can be reached along the path (AS1, AS2).**

- Speaker can cancel previously advertised paths (by sending withdrawal messages)

Path Attributes and BGP Routes

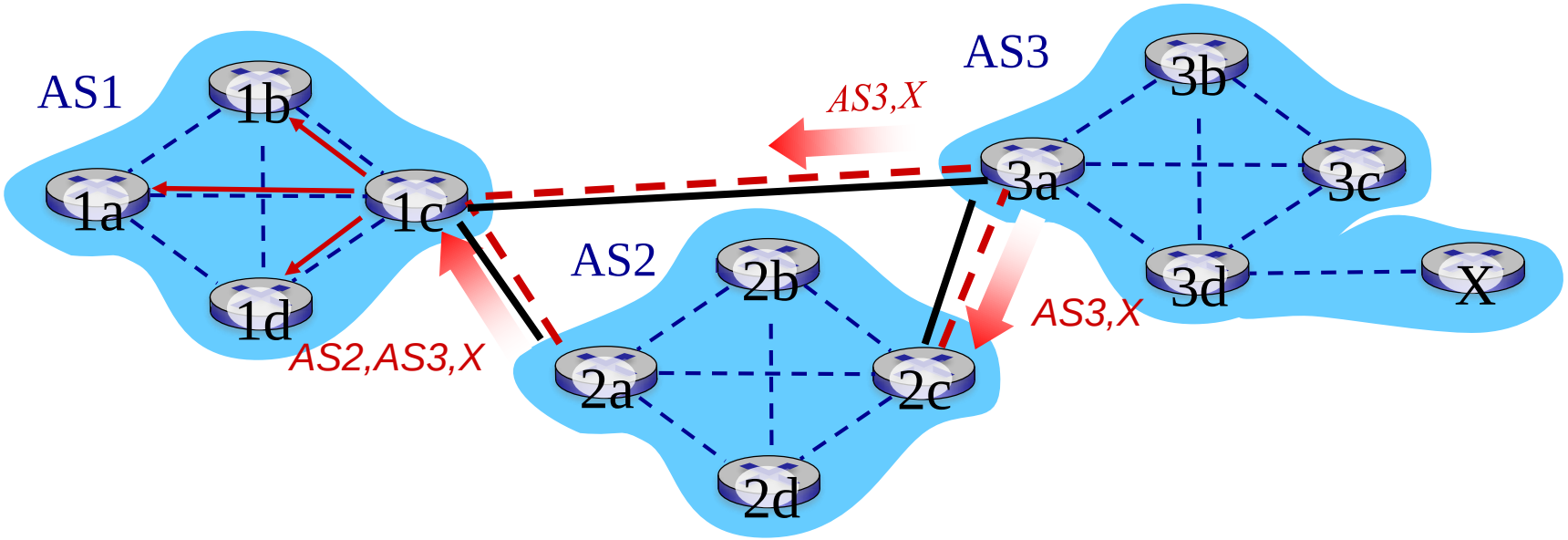
- advertised prefix includes BGP attributes
  - prefix + attributes = "route"
- two important attributes:
  - AS-PATH: list of ASes through which prefix advertisement has passed
  - NEXT-HOP: indicates specific internal-AS router to next-hop AS
- *Policy-based routing:*
  - gateway receiving route advertisement 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 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

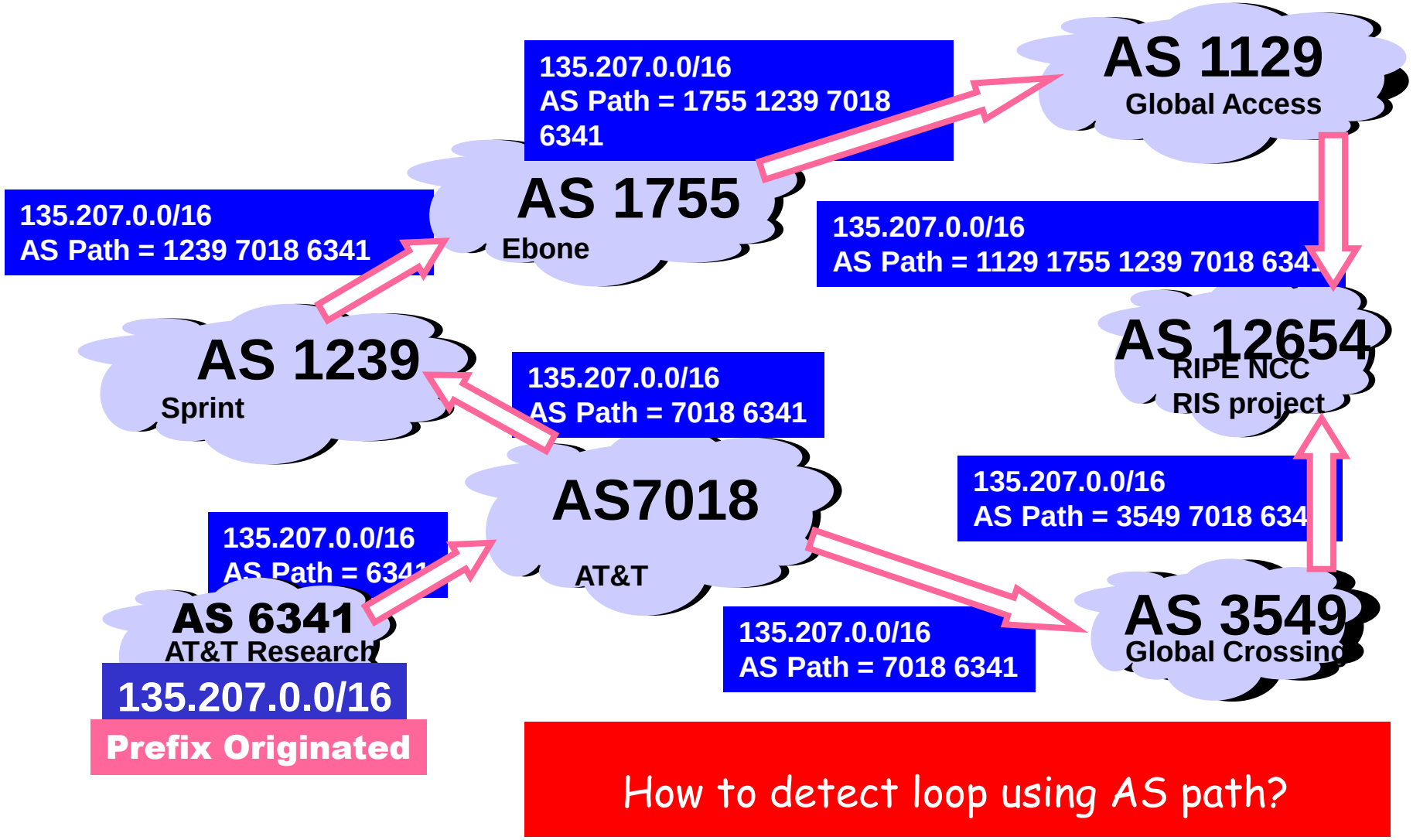
BGP Path Advertisement



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*

AS Path Attribute



BGP: AS Path Advertisement and Policy Routing

Suppose: gateway X send its path to peer gateway W

- W may or may not select path offered by X
  - cost, policy (don't route via competitors AS), loop prevention reasons
  - Policy-based Routing Selection (using BGP attributes)
- If W selects path advertised by X, then:

  Path (W,Z) = W, Path (X,Z)
- Note: X can *control* incoming traffic by controlling its route advertisements to peers:
  - e.g., don't want to route traffic to Z → don't advertise any routes to Z
  - route filtering and export policy (by manipulating attributes)

BGP Attributes

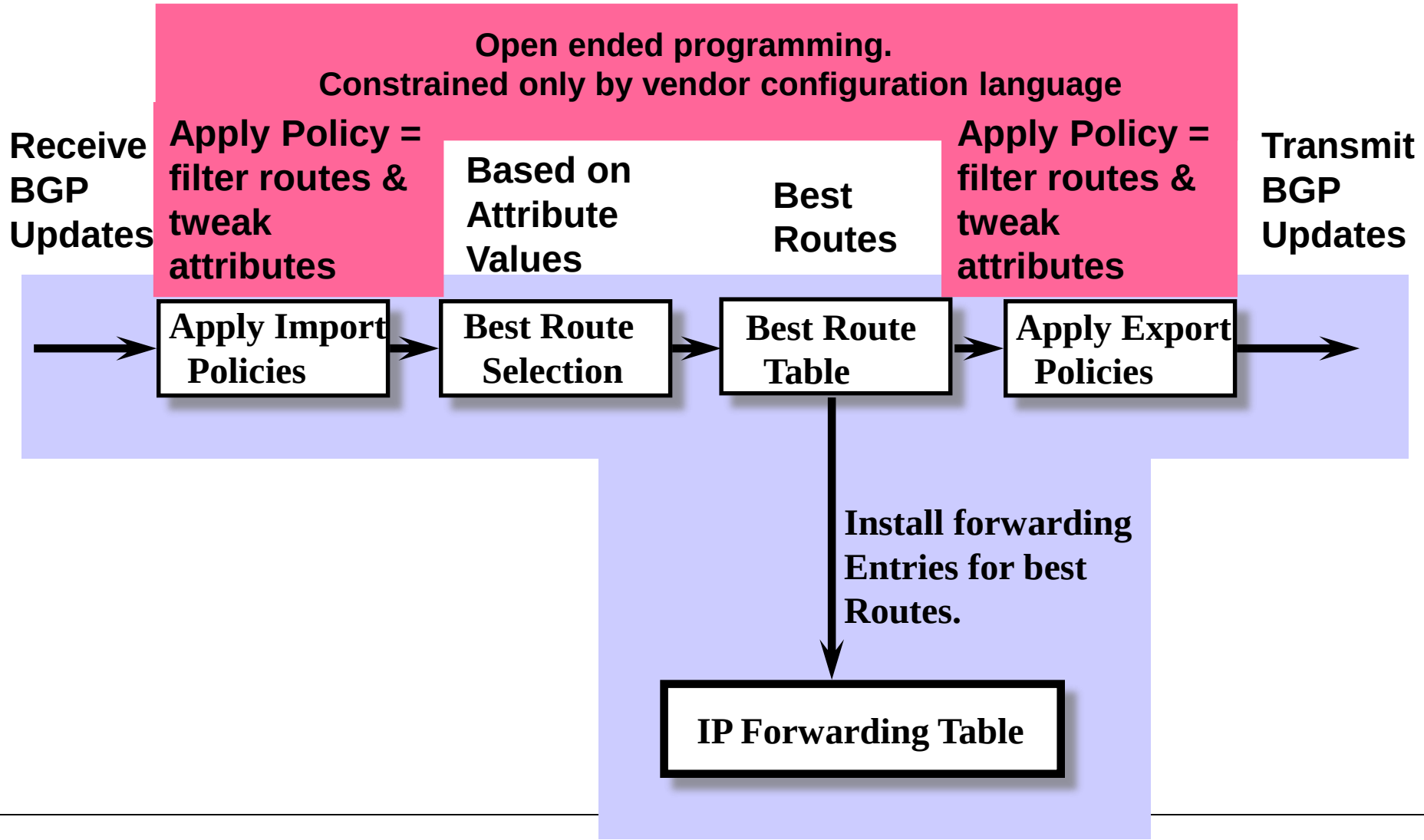
Value	Code	Reference
1	ORIGIN	[RFC1771]
2	AS_PATH	[RFC1771]
3	NEXT_HOP	[RFC1771]
4	MULTI_EXIT_DISC	[RFC1771]
5	LOCAL_PREF	[RFC1771]
6	ATOMIC_AGGREGATE	[RFC1771]
7	AGGREGATOR	[RFC1771]
8	COMMUNITY	[RFC1997]
9	ORIGINATOR_ID	[RFC2796]
10	CLUSTER_LIST	[RFC2796]
11	DPA	[Chen]
12	ADVERTISER	[RFC1863]
13	RCID_PATH / CLUSTER_ID	[RFC1863]
14	MP_REACH_NLRI	[RFC2283]
15	MP_UNREACH_NLRI	[RFC2283]
16	EXTENDED COMMUNITIES	[Rosen]
...		
255	reserved for development	

From IANA: http://www.iana.org/assignments/bgp-parameters

Most important attributes

Not all attributes need to be present in every announcement

BGP Route Processing



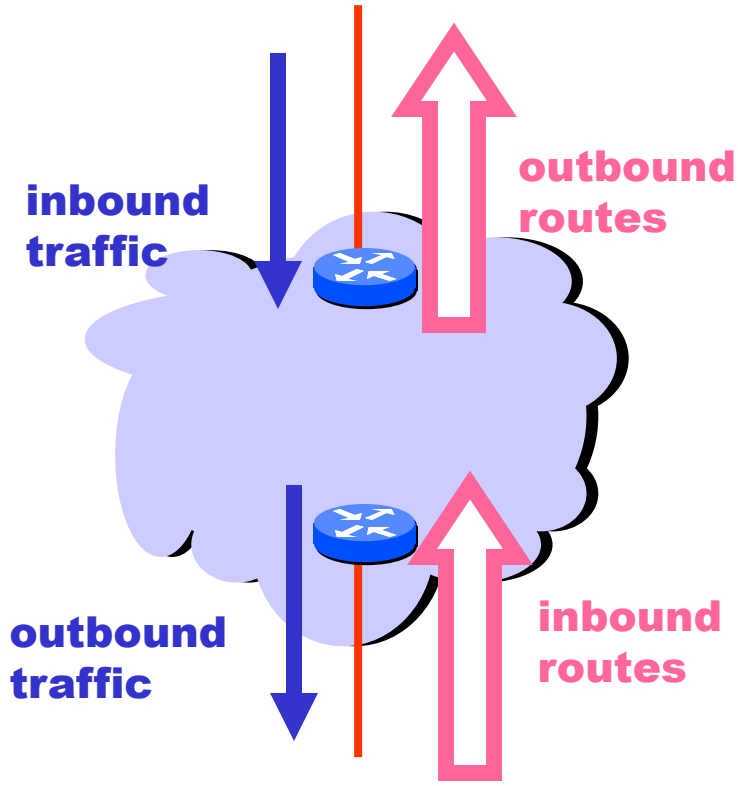
BGP Route Selection

- router may learn about more than one route to destination AS, selects route based on:
  1. local preference value attribute: policy decision
  2. shortest AS-PATH
  3. closest NEXT-HOP router: hot potato routing
  4. additional criteria …

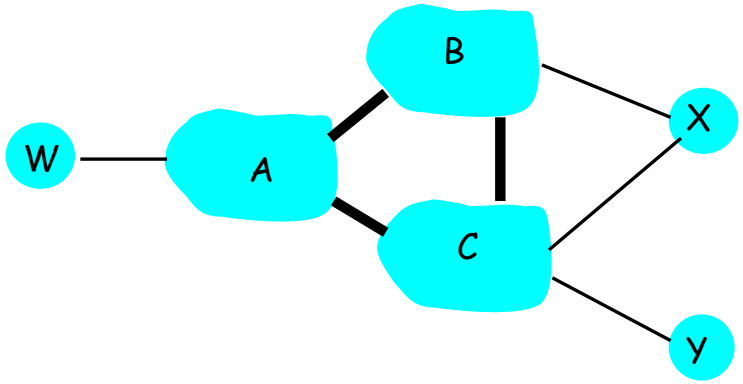
Tweak Tweak Tweak

- For inbound traffic
  - Filter outbound routes
  - Tweak attributes on outbound routes in the hope of influencing your neighbor's best route selection
- For outbound traffic
  - Filter inbound routes
  - Tweak attributes on inbound routes to influence best route selection

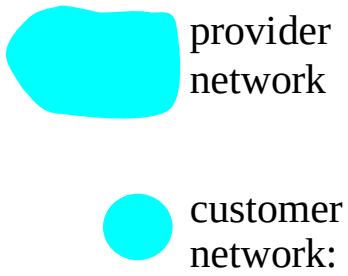
In general, an AS has more control over outbound traffic



BGP: Controlling Who Routes to You



a simple BGP scenario

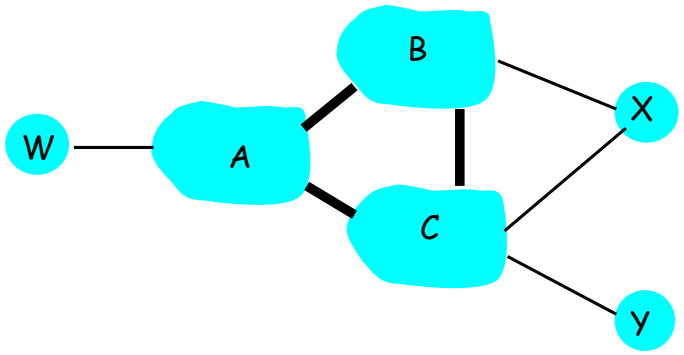


Export Policy for a Provider AS?

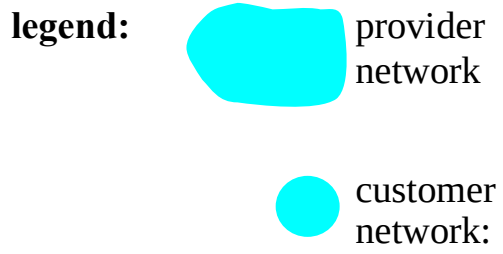
Export Policy for a Customer AS?

- A,B,C are provider networks
- X,W,Y are customer (of provider networks)
- X is dual-homed: attached to two networks
  - C tells X networks belonging to C, i.e., a route to them via C
  - X does not want to carry traffic from B via X to C
  - .. so X will not advertise to B any route to networks in C learned from C

BGP: Controlling Who Routes to You



a simple BGP scenario



Export Policy for a Peer AS?

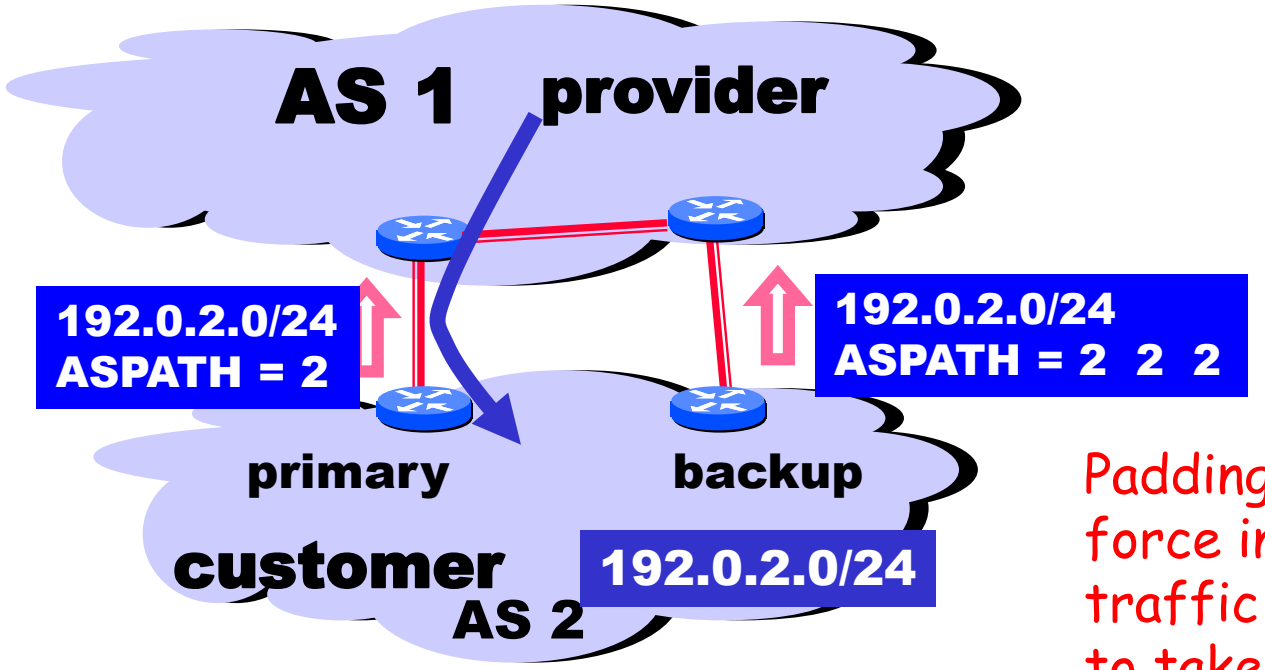
- A advertises to B the path AW
- B advertises to X the path BAW
- Should B advertise to C the path BAW?
  - No way! B gets no "revenue" for routing CBAW since neither W nor C are B's customers
  - B wants to force C to route to W via A
  - B wants to route *only* to/from its customers!

What about route selection?

•which should you prefer? a route learned from a customer AS, a peer AS or a provider AS?

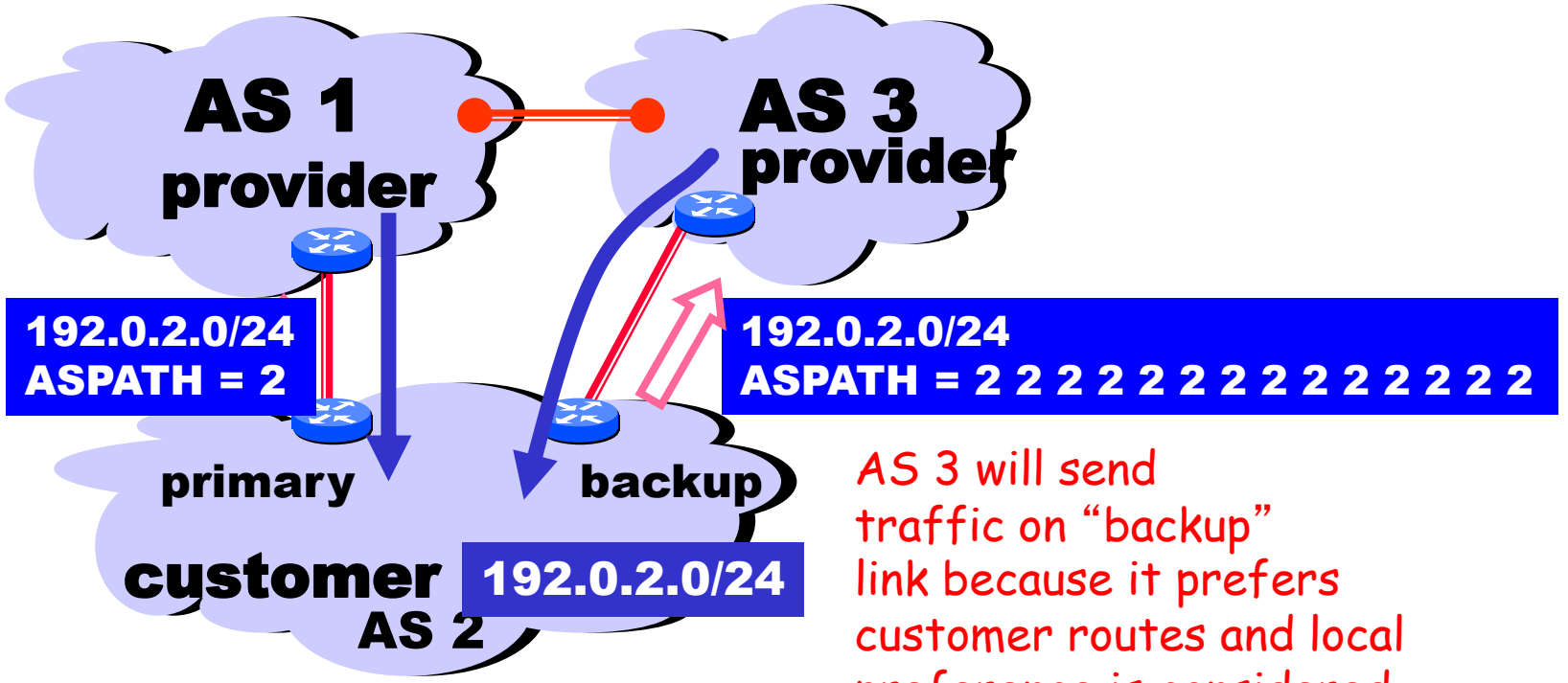
Hint: think how money flows!

Shedding Inbound Traffic with ASPATH Padding Hack



Padding will (usually) force inbound traffic from AS 1 to take primary link

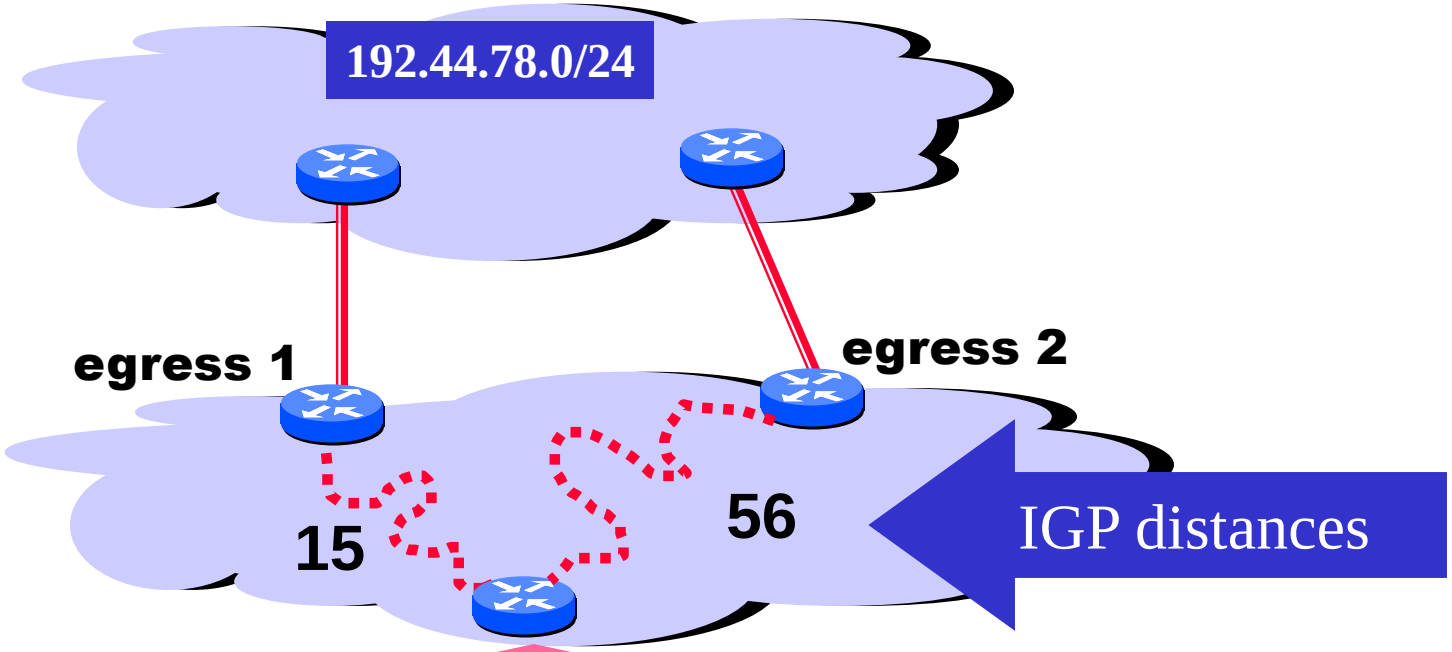
Padding May Not Shut Off All Traffic



AS 3 will send traffic on "backup" link because it prefers customer routes and local preference is considered before ASPATH length!

Padding in this way is often used as a form of load balancing

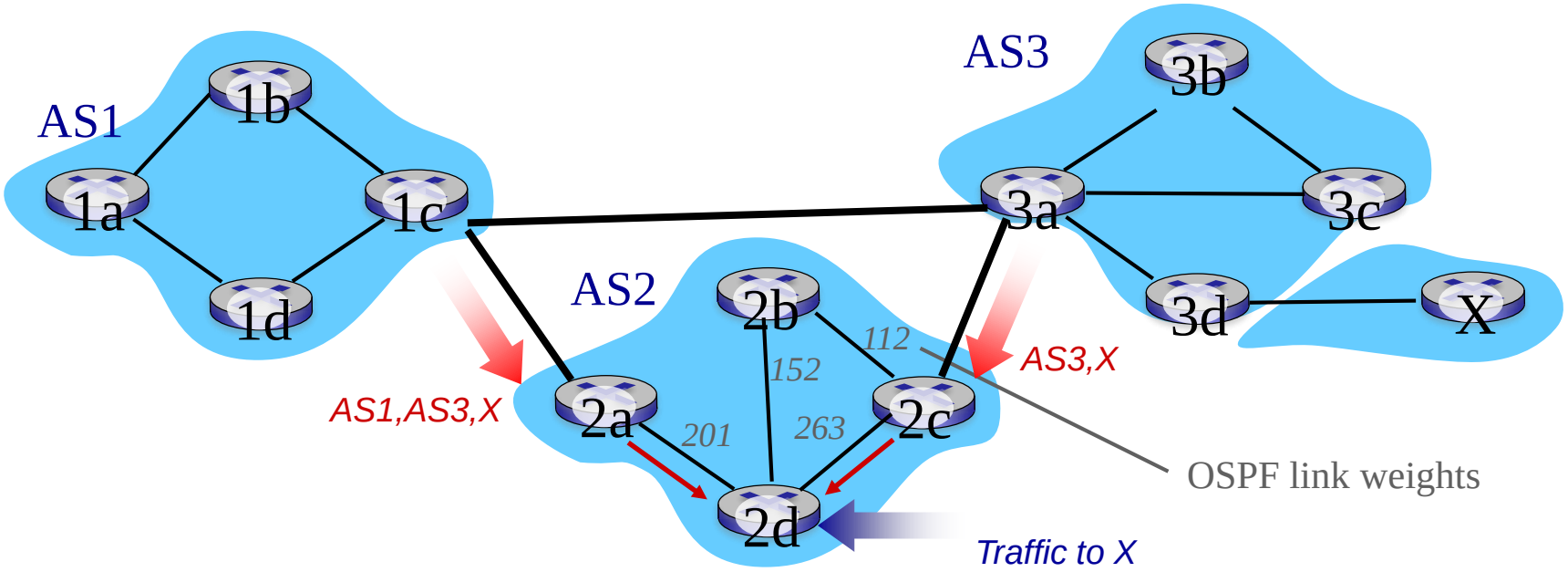
Early Exit or Hot Potato Routing: Go for the Closest Egress Point



This Router has two BGP routes to 192.44.78.0/24.

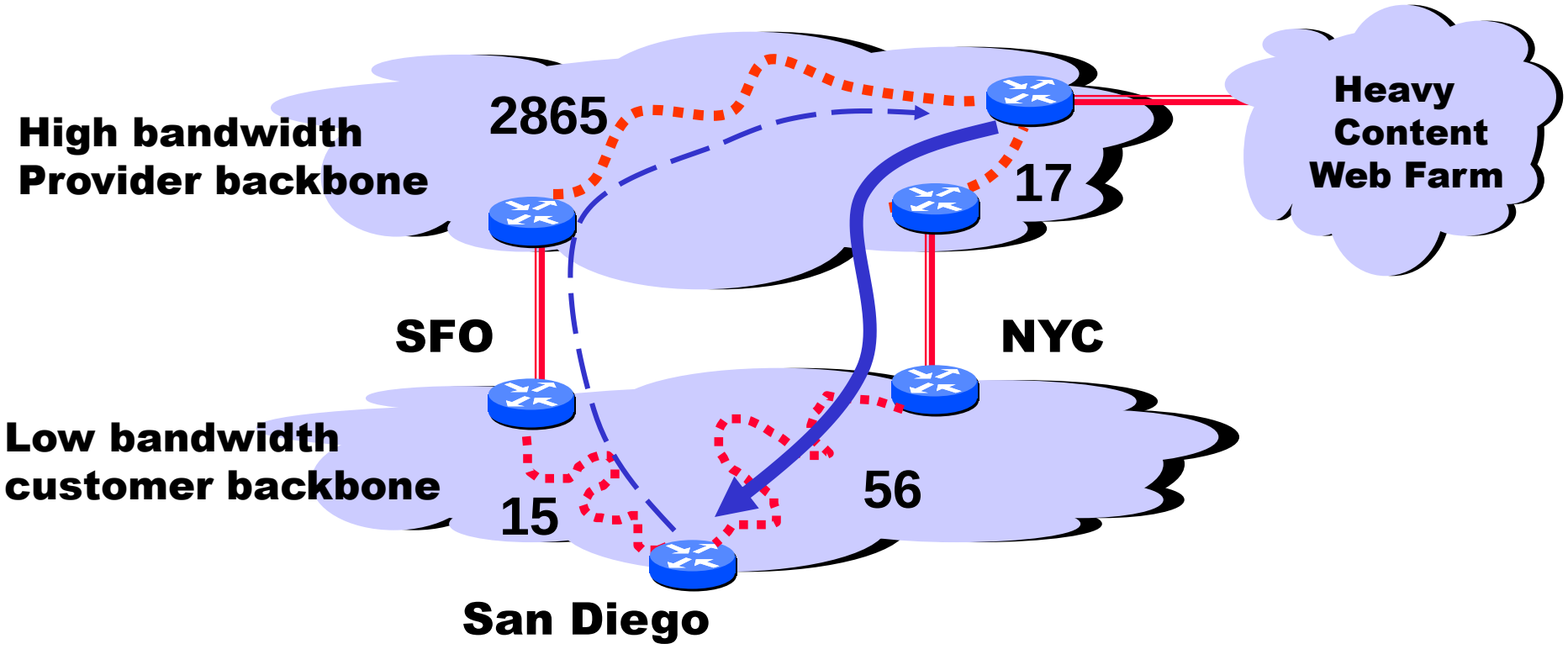
Hot potato: get traffic off of your network as soon as possible. Go for egress 1!

Hot Potato Routing (from authors' lecture notes)



- 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!* (this statement is incorrect! AS-Path attribute precedes the IGP weight criteria in best route selection: hence 2d will select route the from 2c !)
- there is a way to prefer the route from 2a (using local preference attribute) – but this is no longer hot potato routing

Getting Burned by the Hot Potato



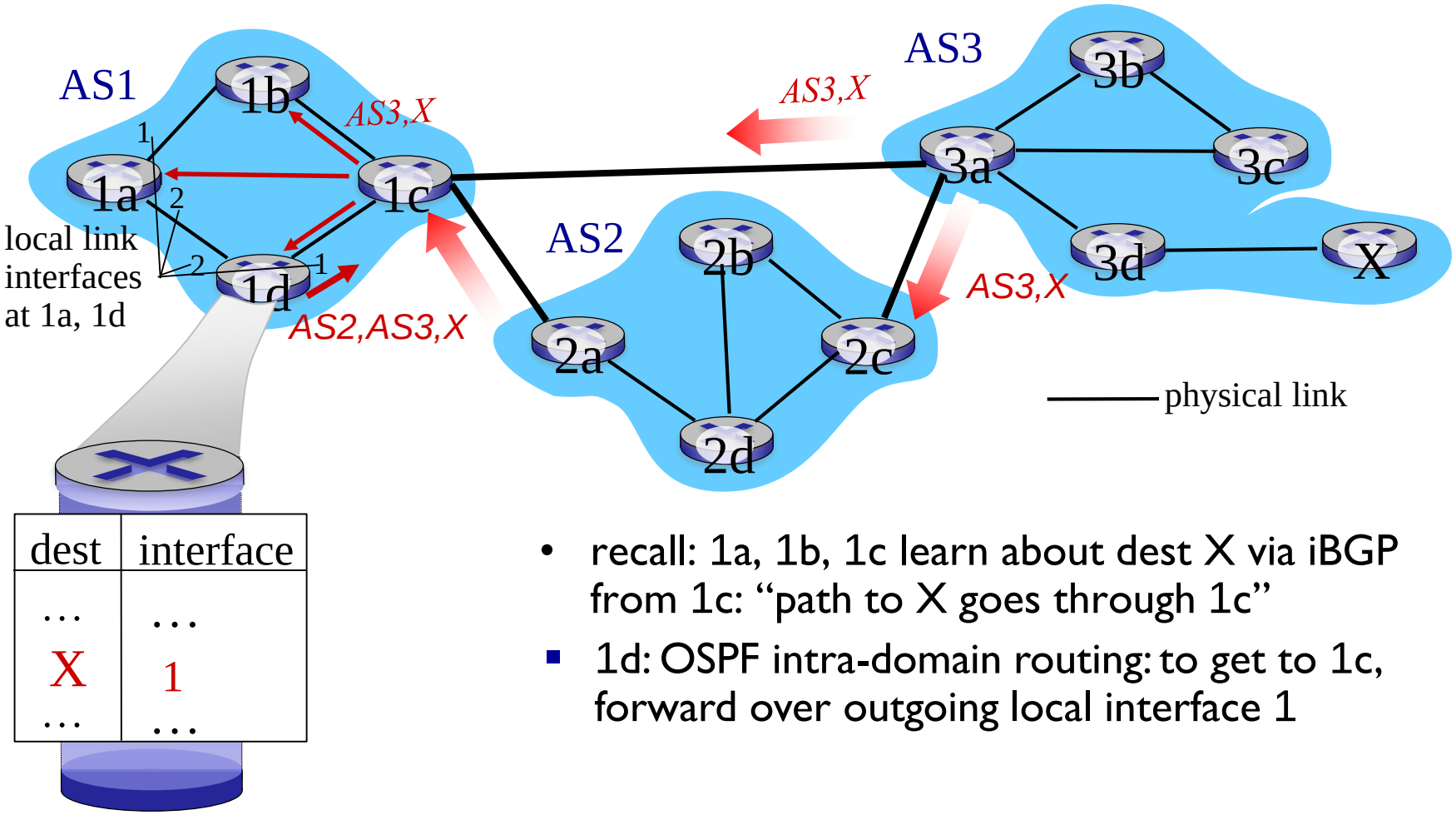
Many customers want their provider to carry the bits!

tiny http request

huge http reply

BGP, OSPF, Forwarding Table Entries

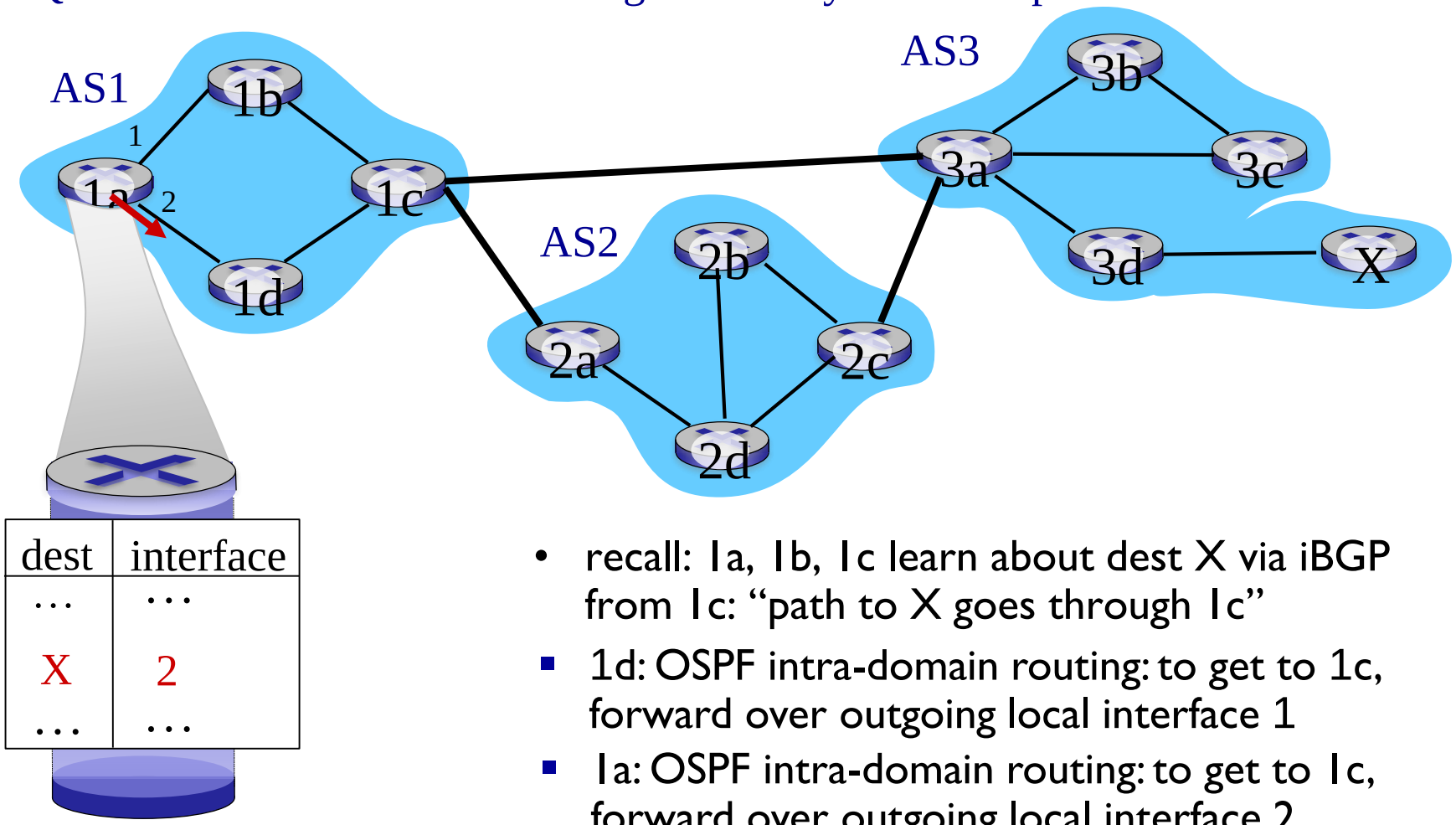
Q: how does router set forwarding table entry to distant prefix?



- 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

BGP, OSPF, Forwarding Table Entries

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

Recap: Why Different Intra-, 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

Network Layer
Control Plane Summary

- Routing is a Key Function in the control plane
  - Basic Issues: topology, path selection, …
- Distributed (intra-AS) routing algorithms: LS vs. DV
  - Link State (LS): How does it work?
  - Distance Vector (DV): How does it work? Issues?
- Centralized control plane and SDN controllers
  - Openflow; POX, OpenDayLight (ODL), ONOS, ..
- Routing in the Internet:
  - Intra-AS vs. Inter-AS routing
  - Distributed intra-AS routing protocols: RIP and OSPF
- Inter-domain (inter-AS) routing: BGP and Policy Routing