CS 352
Routing for the Internet

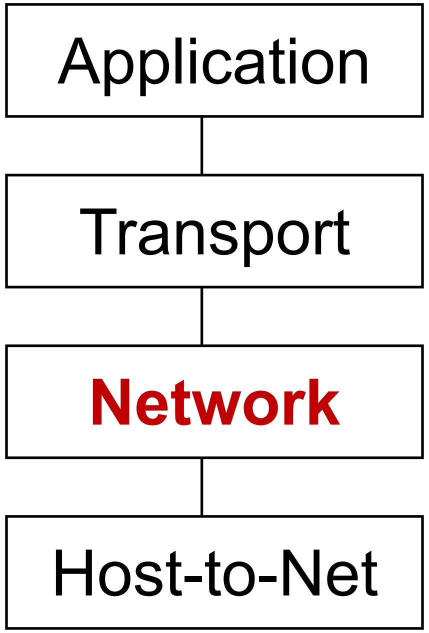
CS 352, Lecture 19.1

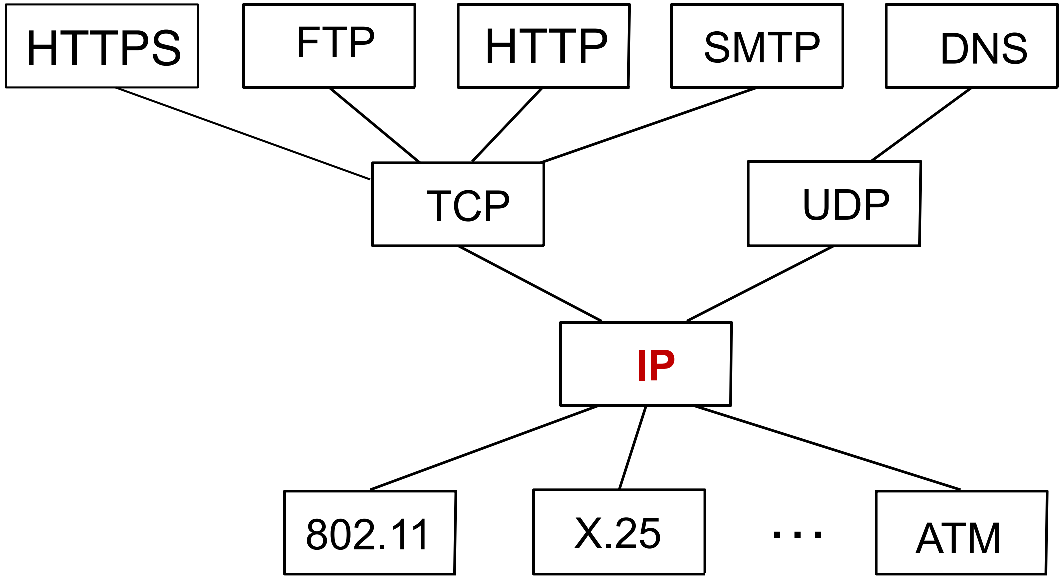
http://www.cs.rutgers.edu/~sn624/352

Srinivas Narayana

RUTGERS
UNIVERSITY | NEW BRUNSWICK

Network



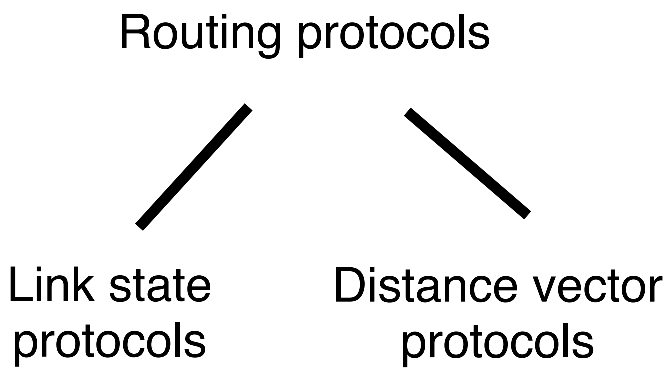


The main function of the network layer is to
move packets from one endpoint to another.

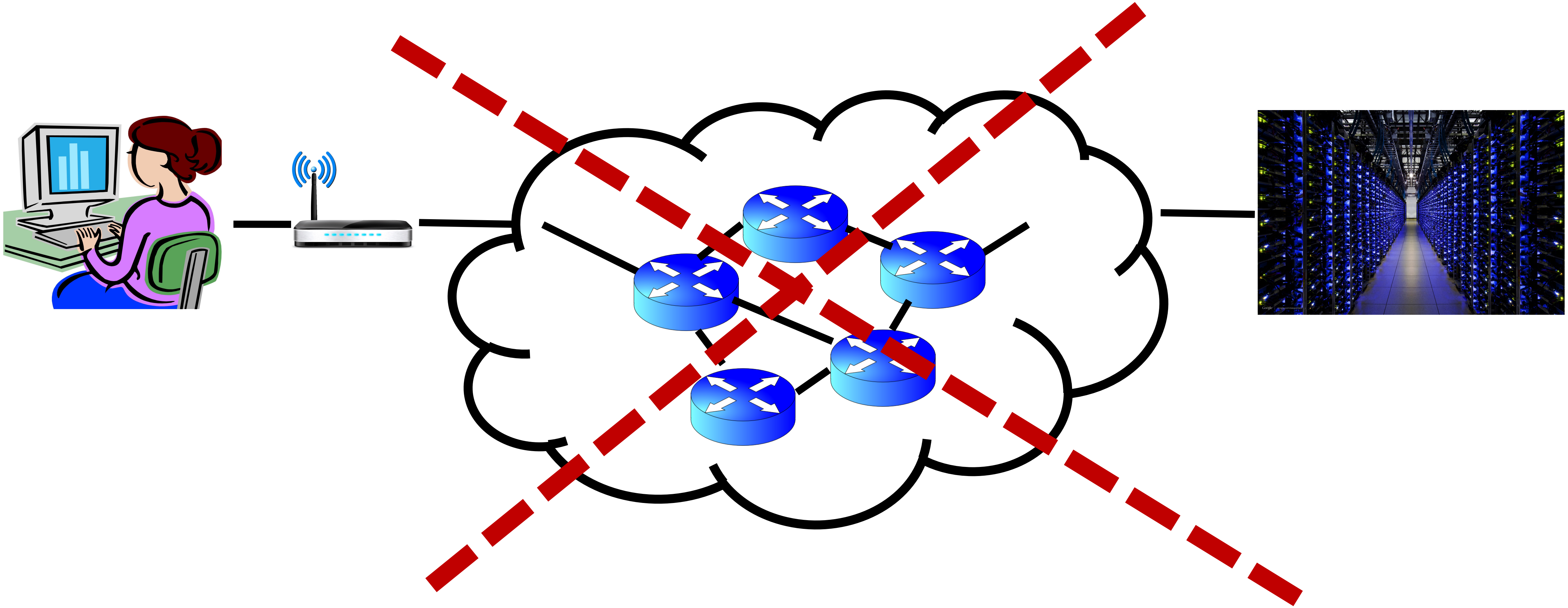
Routing so far

- Routers know the existence of all other routers
  - It's safe to exchange neighborhood information
- All link metrics are known
- It is feasible to exchange info at scale
  - LS: Link state advertisement flooding throughout the network
  - DV: Distance vectors to all other routers is small enough to exchange

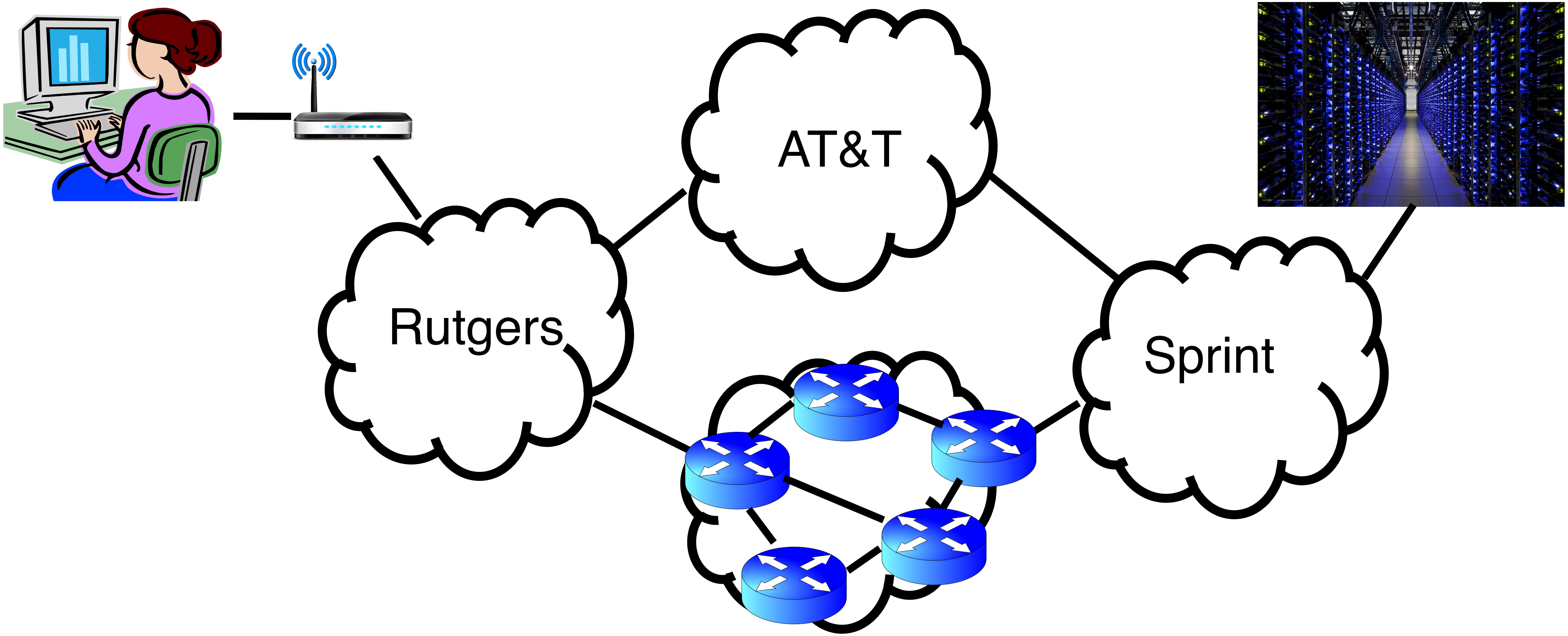
- It is difficult to scale this approach to the Internet



The Internet is not a “flat” network



The Internet is a network of networks



Constraints of the Internet

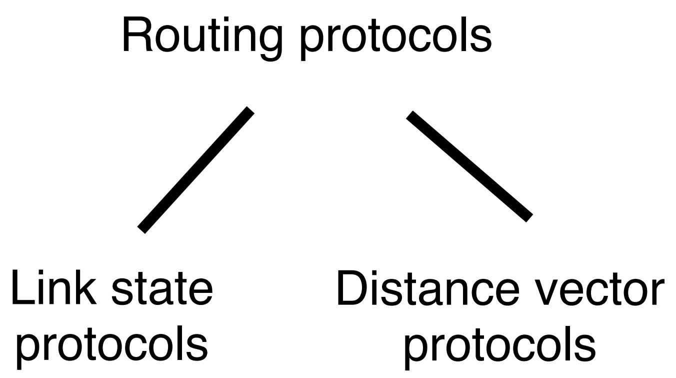
- Administrative autonomy
  - The Internet is not owned by any one organization
  - Rather, it is a network of organizations interconnected with each other
  - The network graph, the link metrics, the IDs and locations of routers are not public information
- Scale
  - It is unscalable to flood LSAs all over the Internet
  - Sending a vector containing distances to all other Internet routers will swamp network links

The Internet's approach

- Split the network into separately administered autonomous systems (AS'es)
  - Rutgers is an autonomous system
  - So are AT&T, Verizon, and Comcast
- Use different approaches for routing within AS'es and routing across AS'es
- Distributing the administration helps scale to larger networks
  - Ex: think about Government: federal → state → city → boro → …

Intra-domain routing: Routing within AS'es

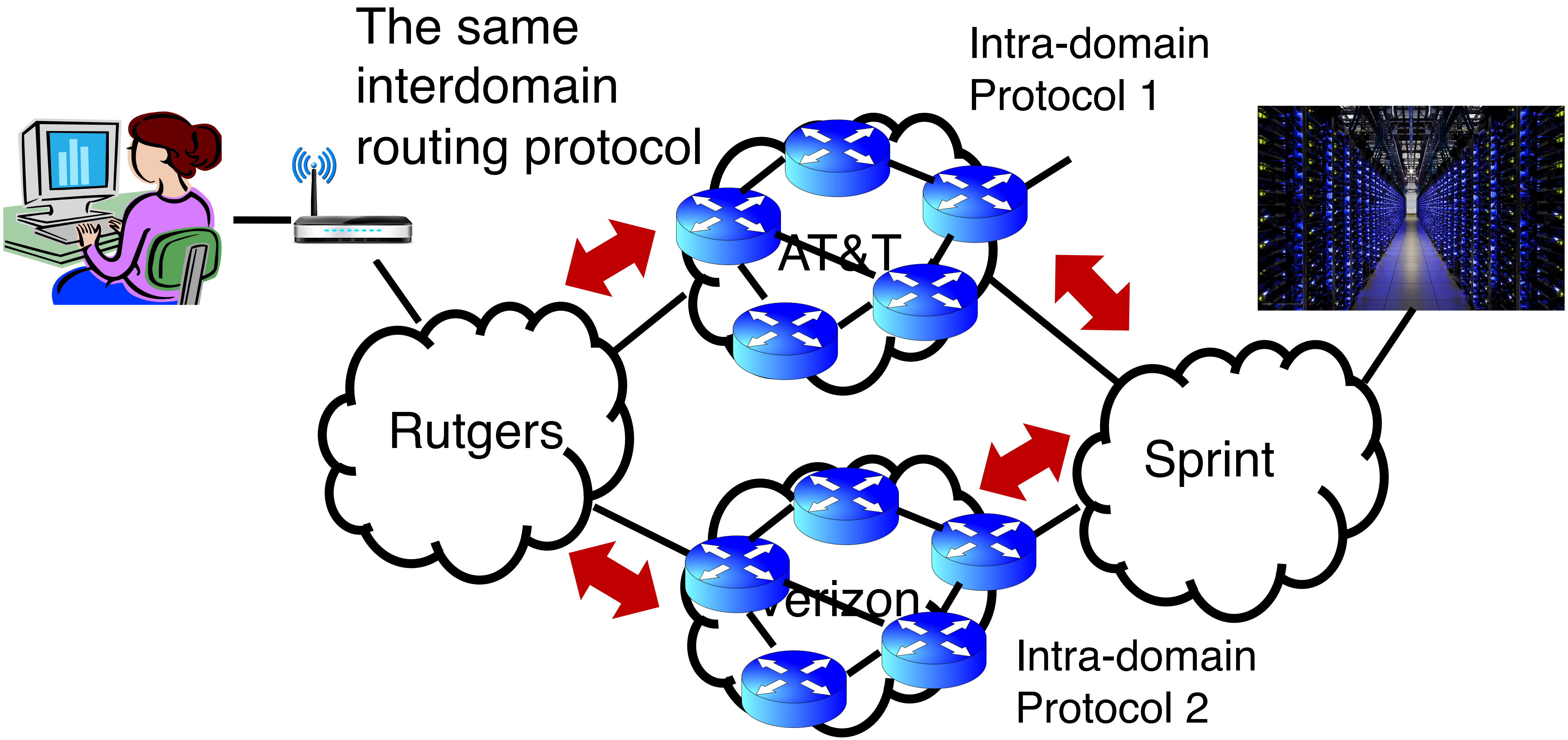
- The approaches we've studied so far are applicable within an AS!
- It is safe for routers within an AS to know the existence of other routers and all link metrics within the same organization
- It is indeed feasible to use link state flooding or exchange distance vectors to all other routers
  - Such approaches won't scale to Internet size
- Different AS'es can use different intra-domain routing protocols: e.g., OSPF (LS), RIP (DV)
- Routers within an AS must use the same protocol



Inter-domain routing: Routing across AS'es

- Routing information is exchanged at a coarser granularity
  - Don't announce per-router info; instead, announce per AS info
  - (The assignment of IP prefixes to AS'es is public information)
  - Path announced per destination, not for all destinations
- Link metrics are not exchanged (not public info)
  - Instead, the entire path available to the destination is exchanged
- Only the routers at the border of two networks need to speak the inter-domain routing protocol: border/gateway routers
- However, all AS'es need to speak the same inter-domain routing protocol
  - Next, we'll study this protocol: Border Gateway Protocol (BGP)

The Internet is a network of networks



Routing protocols

Link state protocols
e.g., OSPF, IS-IS

Distance vector protocols
e.g., RIP, IGRP

Path vector protocols
BGP

Intra-domain protocols

- same protocol within an AS
- different algorithms across ASes
- (semi)global view of the org’s network
- Sometimes called interior gateway protocols (IGP)

Inter-domain protocol

- common across AS’es
- each AS knows little about the others

CS 352
Border Gateway Protocol

CS 352, Lecture 19.2

http://www.cs.rutgers.edu/~sn624/352

Srinivas Narayana

RUTGERS
UNIVERSITY | NEW BRUNSWICK

Routing protocols

Link state protocols
e.g., OSPF, IS-IS

Distance vector protocols
e.g., RIP, IGRP

Path vector protocols
BGP

Intra-domain protocols

Inter-domain protocol

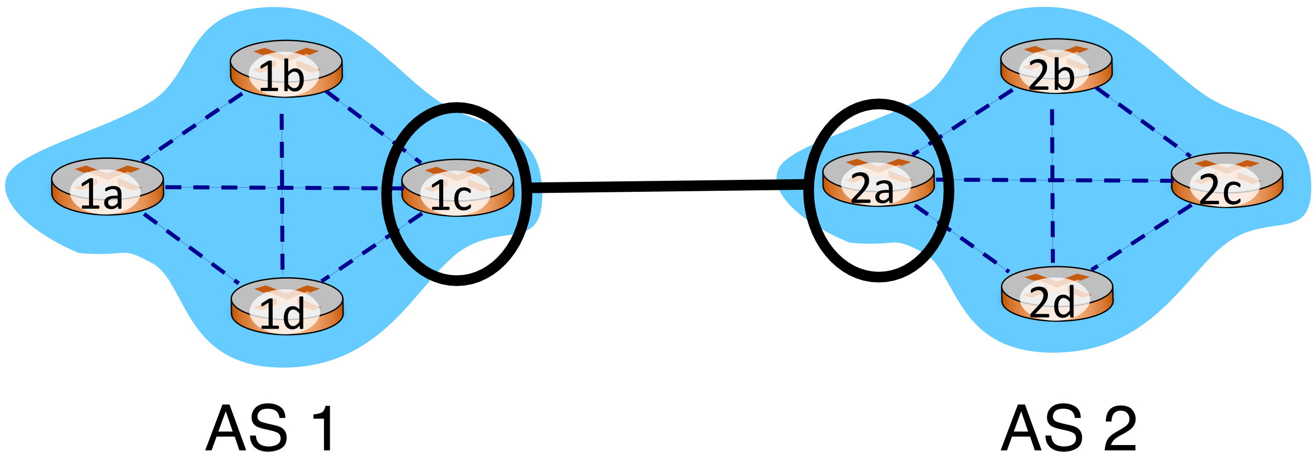
The glue that holds the Internet together.

Border Gateway Protocol

- *The* de facto inter-domain routing protocol
- Two parts to BGP:
  - eBGP: each AS can obtain reachability information from neighboring AS'es
  - iBGP: each AS propagates reachability information about external AS'es to all AS-internal routers.
- Q1: What computation occurs at each router?
- Q2: What information is exchanged?

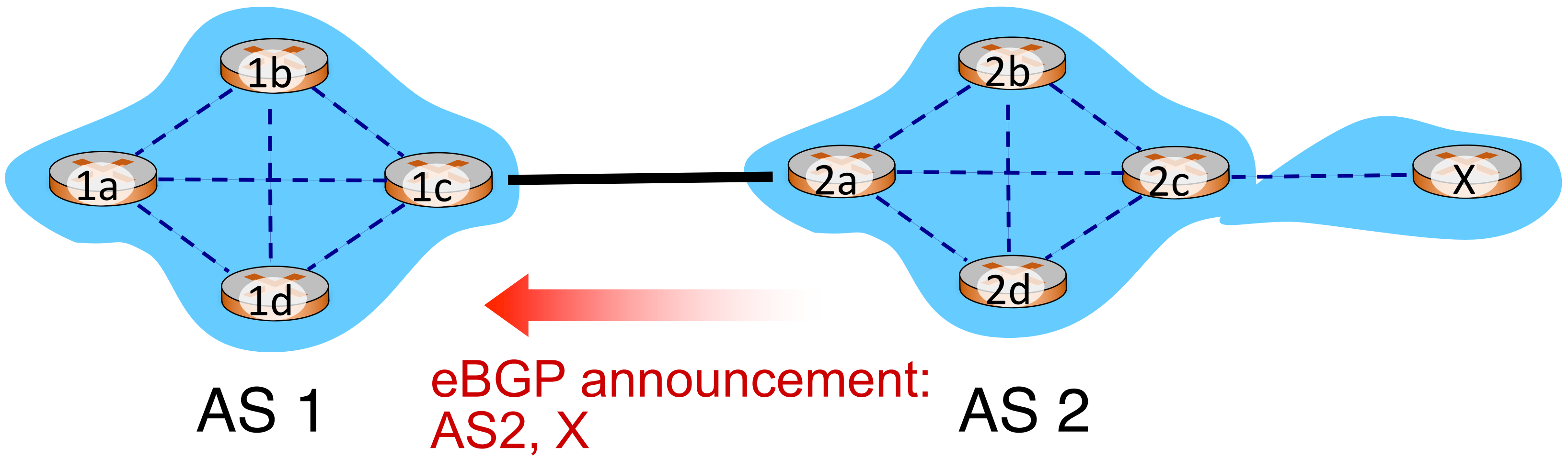
Q2. BGP announcements

- eBGP allows a network to advertise its existence to the rest of the Internet using eBGP announcements
- Announcements occur over a BGP session
  - Semi-permanent TCP connection between gateway routers
- Announcements contain AS-level paths to IP prefixes
  - BGP is a path vector protocol



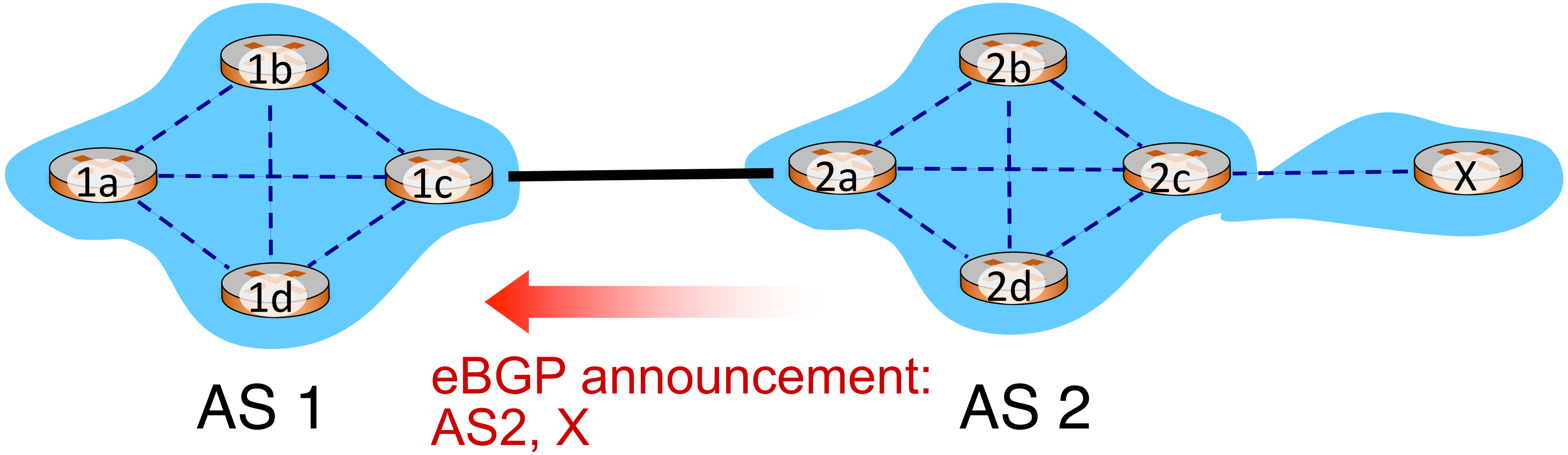
Q2. BGP announcements

- Suppose AS2's gateway router 2a announces path AS2,X to AS1's gateway router 1c
- AS2 promises that it will forward datagrams towards X
- Announcements contain the IP prefix destination as well as attributes
- Two important attributes: AS-path (AS2,X), Next-Hop



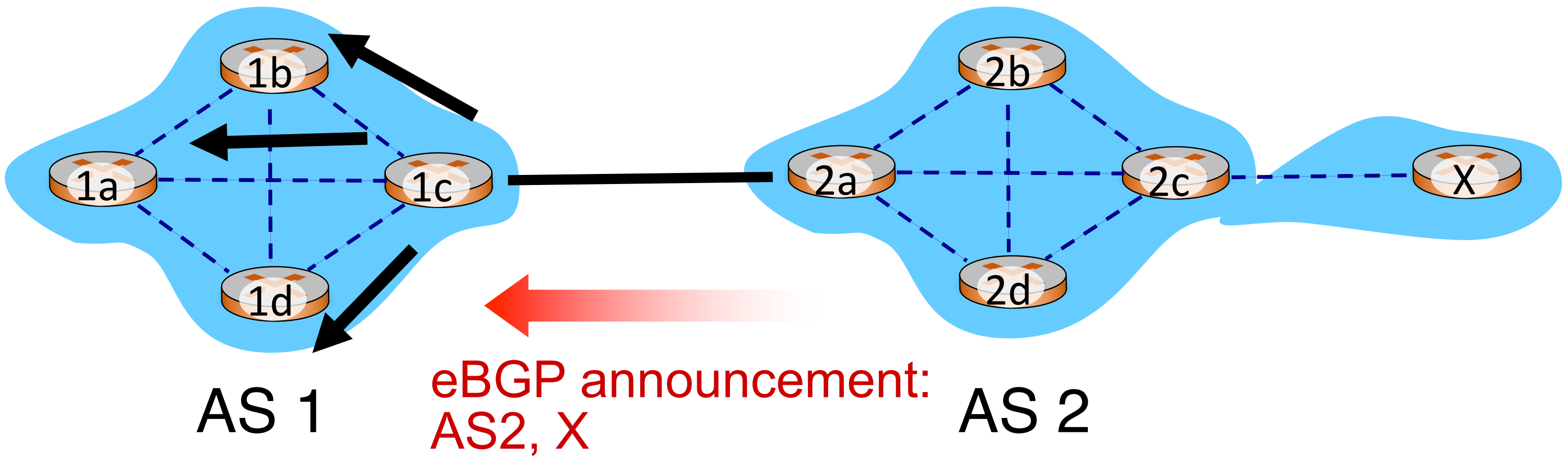
Q2. BGP announcements: Next Hop

- Next hop conceptually denotes the next hop router that must be used to reach a specific destination.
  - However, the meaning of this attribute is context-dependent
- In an eBGP announcement, next hop denotes the router in the next AS which sent the announcement
- Next Hop of the eBGP announcement reaching 1c is 2a



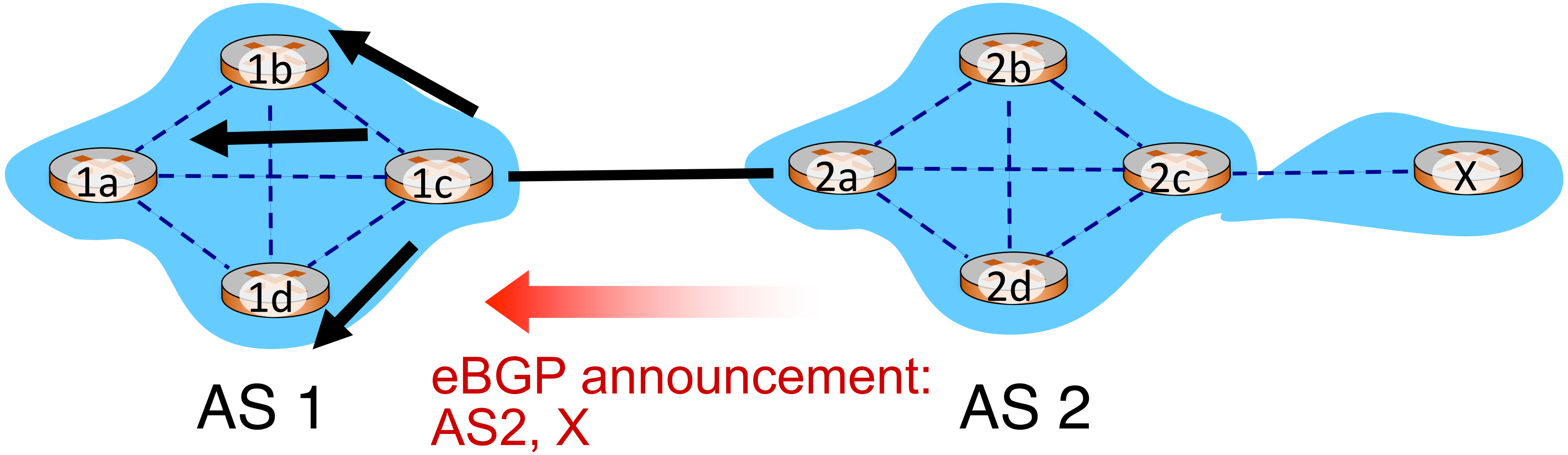
Q2. BGP announcements: Next Hop

- Suppose router 1c accepts the path (more on this soon)
- Router 1c will propagate the announcement inside the AS using iBGP
- The next hop of the iBGP announcement from 1c to 1a is set to router 1c
  - In particular, the next hop is an AS1 internal address



Q1. What is computed?

- Upon receiving an announcement, a BGP router chooses routes to other networks based on policy considerations
- This approach is very different from the link-metrics-based approaches we've seen earlier
- Export policy determines whether a path is announced
- Import policy determines whether a path is accepted

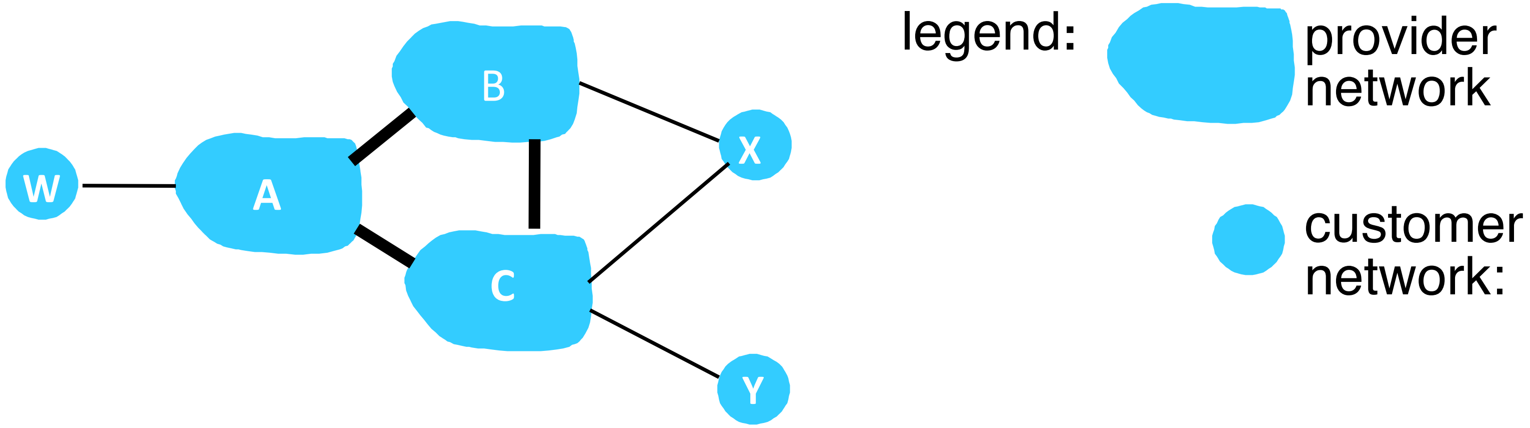


Policies in BGP

Policy arises from business relationships

- Customer-provider relationships:
  - E.g., Rutgers is a customer of AT&T
- Peer-peer relationships:
  - E.g., Verizon is a peer of AT&T
- Business relationships depend on where connectivity occurs
  - "Where", also called a "point of presence" (PoP)
  - e.g., customers at one PoP but peers at another
  - Internet-eXchange Points (IXPs) are large PoPs where ISPs come together to connect with each other (often for free)
- Sometimes, even when there is no direct connectivity
  - "e.g., inteliquent (zoom/webex) traffic should not be charged"

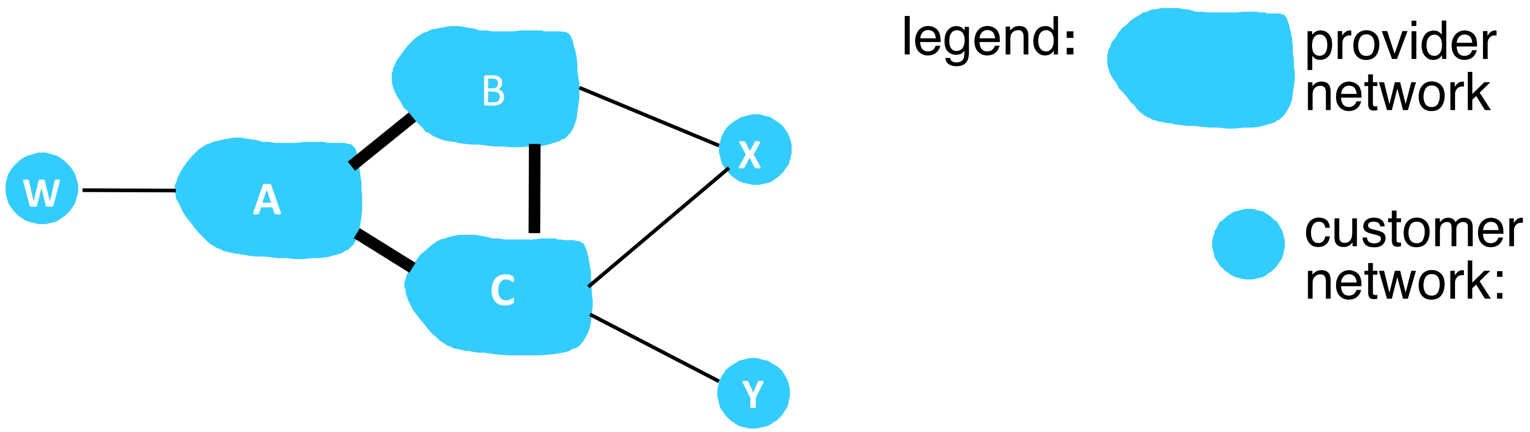
BGP Export Policy



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

- A,B,C are provider networks
- X,W,Y are customers (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 announce to B a route to C

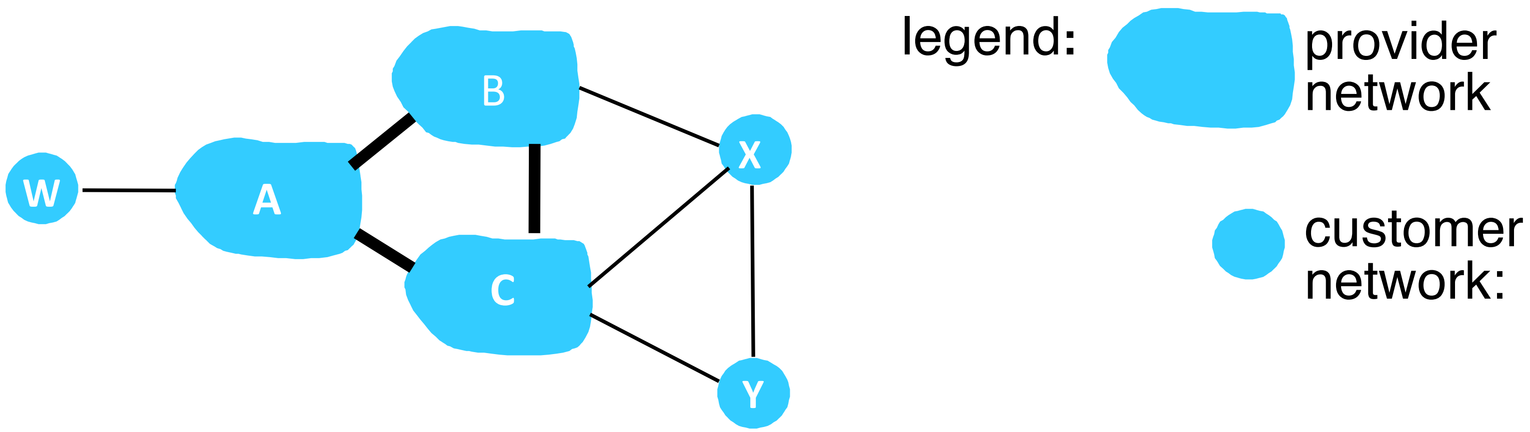
BGP Export Policy



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

- A announces path Aw to B and to C
- B *chooses not to announce* BAw to C:
  - B gets no "revenue" for routing CBAw, since none of C, A, w are B's customers
- C will route CAw (not using B) to get to w

BGP Import Policy



Suppose an ISP wants to minimize costs by avoiding routing through its providers when possible.

- Suppose C announces path Cy to x
- Further, y announces a direct path ("y") to x
- Then x might reject the path Cy towards y in favor of using the direct path ("y") towards y: reduce costs by avoiding provider network

Policies make BGP a complex protocol.

Policy considerations dominate performance considerations (e.g., no “link metrics” for AS paths).

BGP chooses to announce (export) only certain paths.

BGP chooses to accept (import) only certain paths.

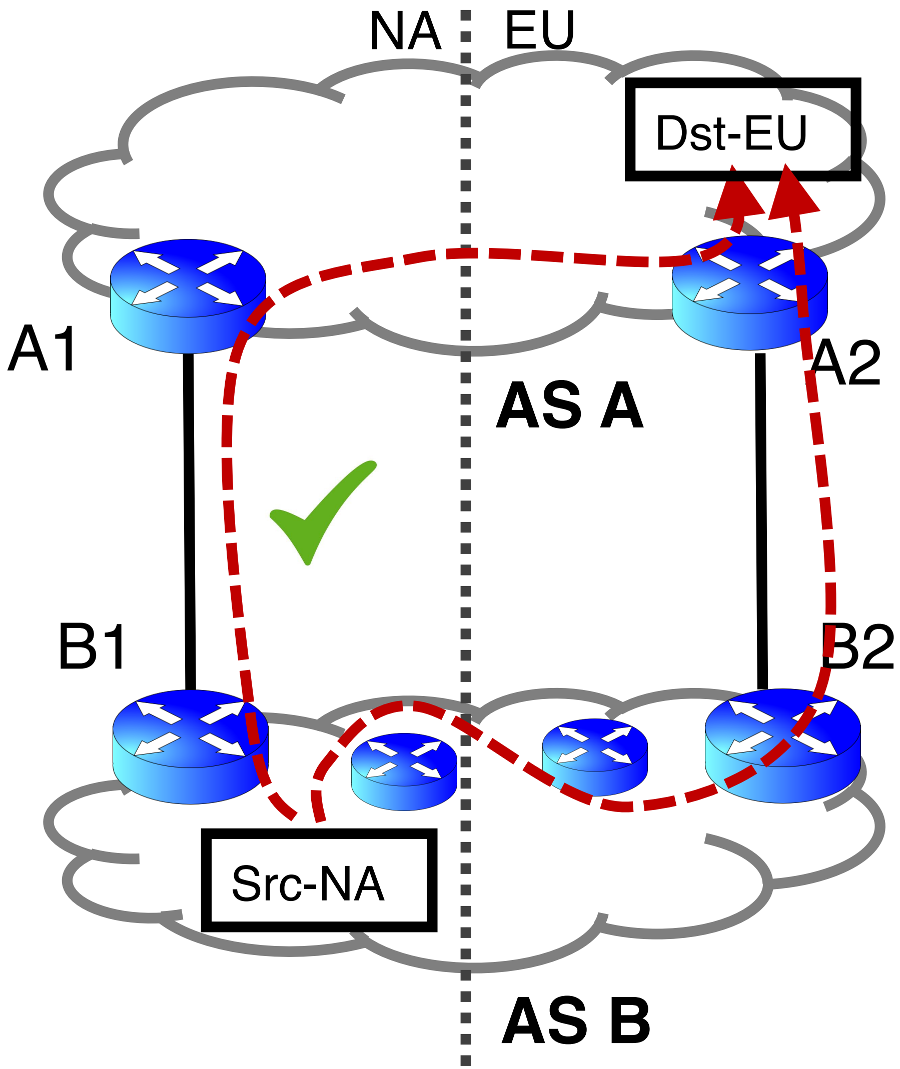
A complex decision process prefers certain paths over others.

Q1. BGP route selection process

- When a router learns more than one acceptable route to a destination AS, it selects route based on:
  1. local preference value attribute (policy decision, set by network admin)
  2. shortest AS-PATH
  3. closest NEXT-HOP router
  4. Several additional criteria: You can read up on the full, complex, list of criteria, e.g., at https://www.cisco.com/c/en/us/support/docs/ip/border-gateway-protocol-bgp/13753-25.html

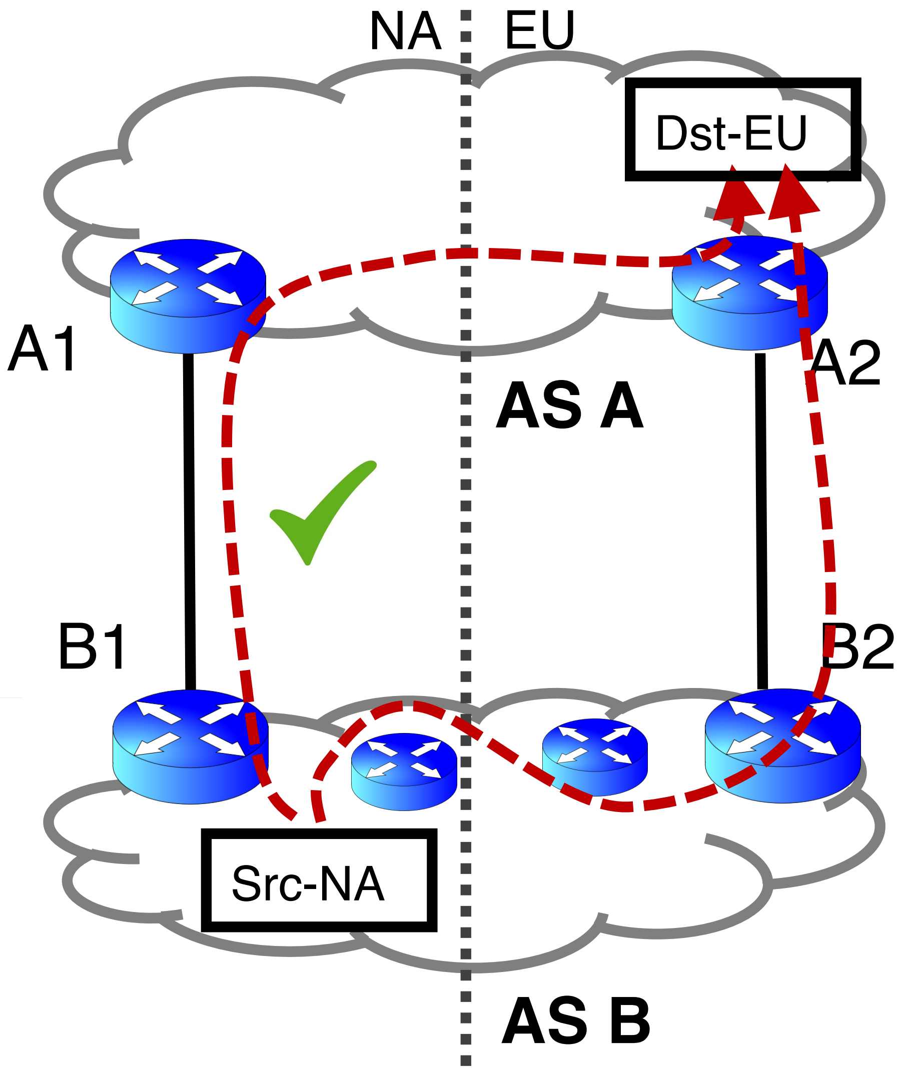
Example of route selection

- Suppose AS A and B are connected to each other both in North America (NA) and in Europe (EU)
- A source in NA wants to reach a destination in EU
- There are two paths available
  - *Assume* same local preference
  - Same AS path length
- Closest next hop-router: choose path via B1 rather than B2



Example of route selection

- Choosing closest next-hop results in early exit routing
  - Try to exit the local AS as early as possible
  - Also called hot potato routing
- Reduces bandwidth use within the local AS
  - … potentially at the expense of another AS



Summary of BGP

- BGP is the protocol that enables communication across multiple autonomous systems in the Internet
- Border routers exchange AS-level paths to prefixes via eBGP, propagate those prefixes to internal routers via iBGP
  - Path vector protocol
- BGP routers use a complex policy-based procedure to choose the final path and next hop for a given IP prefix destination
  - Local pref, AS path length, closest next hop, and other criteria

CS 352
Forwarding to External Destinations

CS 352, Lecture 19.3

http://www.cs.rutgers.edu/~sn624/352

Srinivas Narayana

RUTGERS
UNIVERSITY | NEW BRUNSWICK

Routing protocols

Link state protocols
e.g., OSPF, IS-IS

Distance vector protocols
e.g., RIP, IGRP

Path vector protocols
BGP

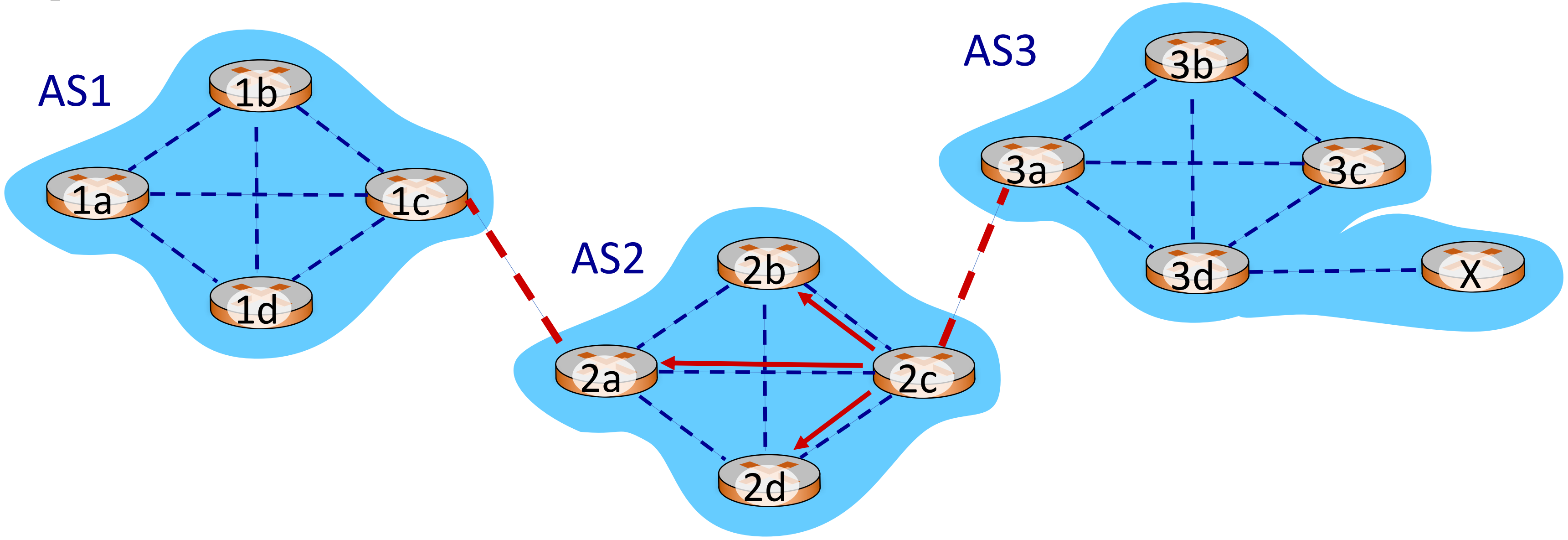
Border Gateway Protocol:
The glue that holds the Internet together.

Review: BGP

- Two parts to BGP:
- eBGP announcements from external AS'es carry information about IP prefixes reachable through an AS
- iBGP propagates announcements received from external AS'es to AS-internal routers
- BGP announcements contain an IP prefix and attributes
- This module: One of the attributes of the BGP announcement, Next Hop, is key to generating forwarding tables for all routers

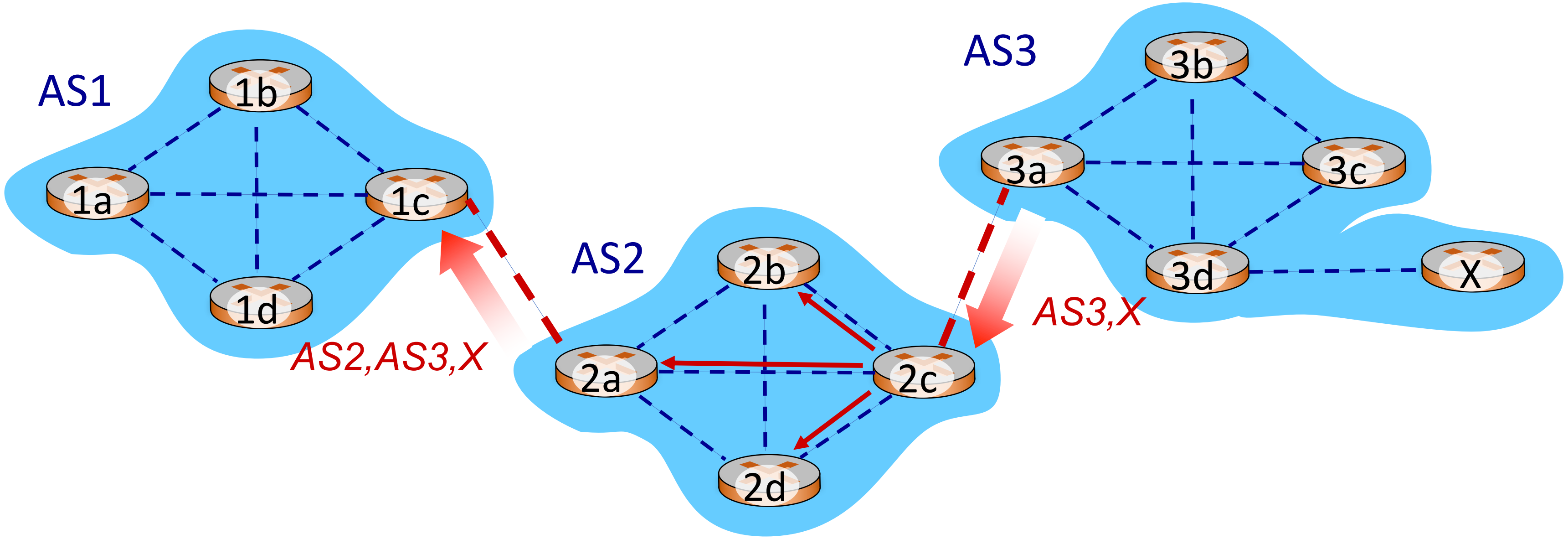
Forwarding to an external prefix

Example scenario



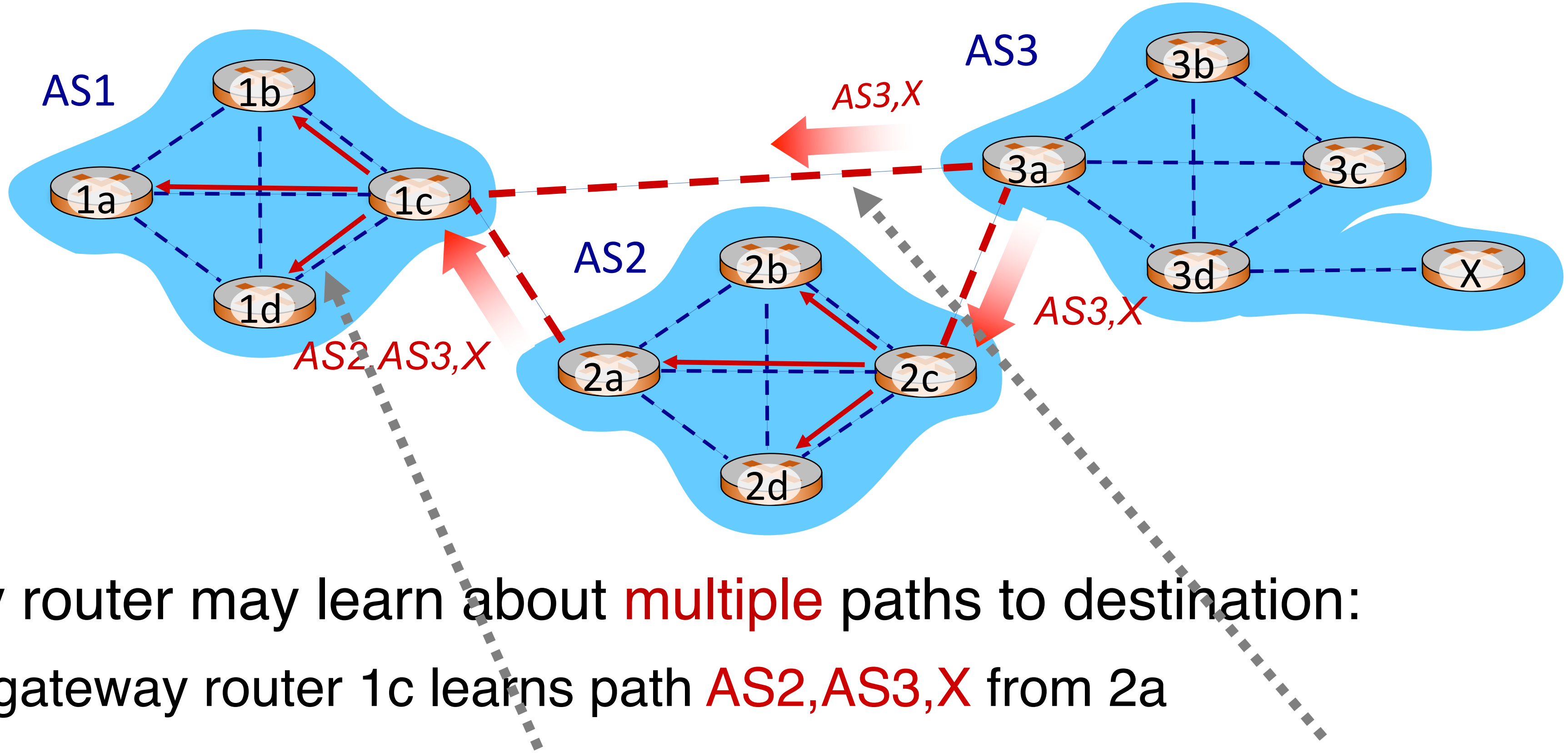
- Suppose a router in AS1 wants to forward a packet destined to external prefix X.
- How is the forwarding table entry for X at 1d computed?
- How is the forwarding table entry for X at 1c computed?

eBGP and iBGP announcements



- AS2 router 2c receives path announcement 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 announces (via eBGP) path AS2, AS3, X to AS1 router 1c

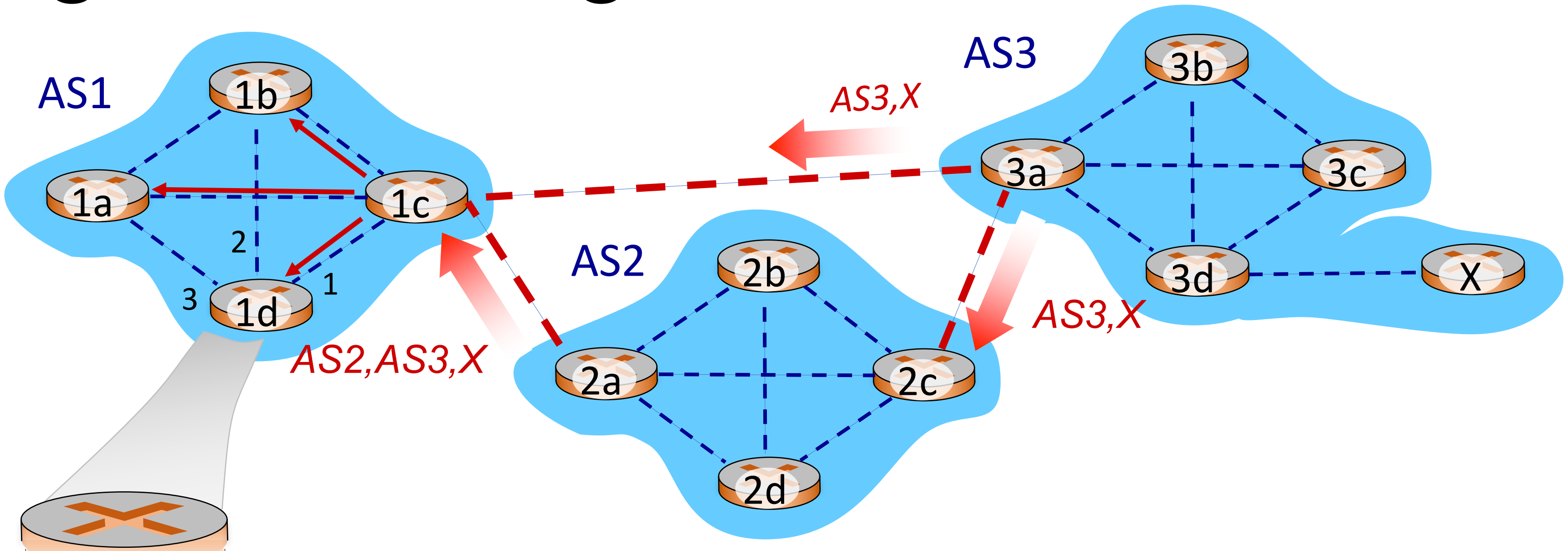
eBGP and iBGP announcements



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 (next hop 3a)
- Based on policy, AS1 gateway router 1c chooses path AS3,X, and announces path within AS1 via iBGP (next hop 1c)

Setting forwarding table entries

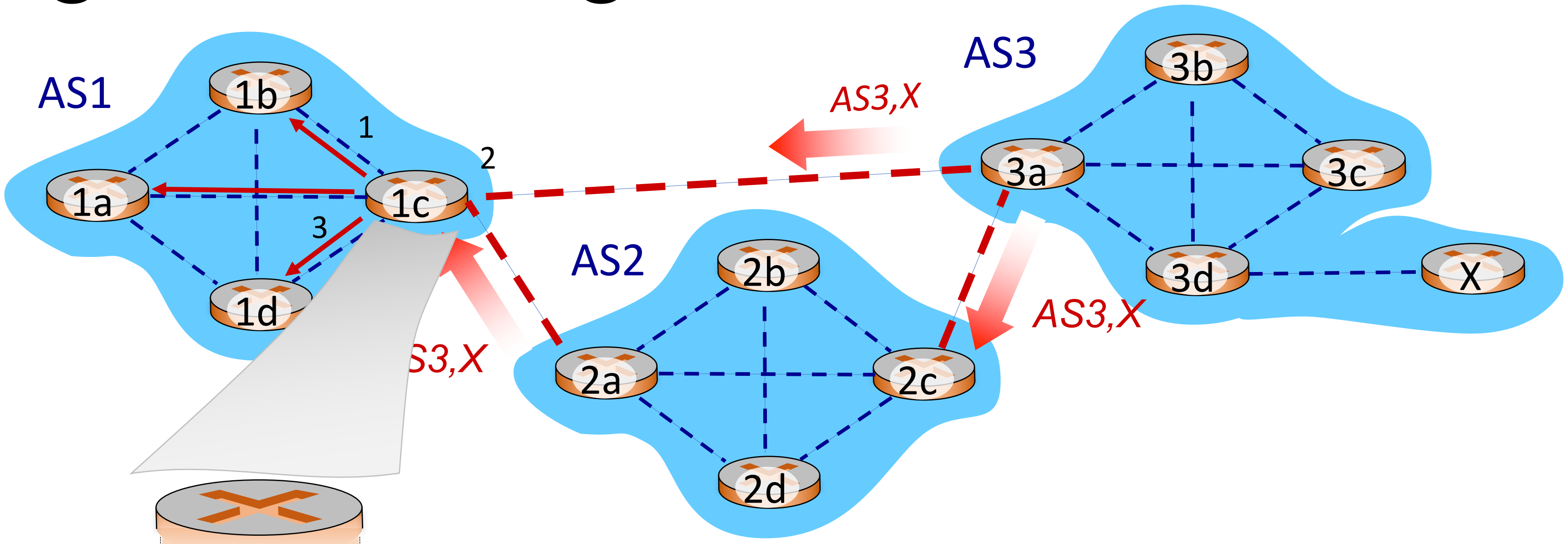


dest	Next-Hop
…	…
X	1c
…	…

NH	interface
…	…
1c	1
…	…

- recall: 1a, 1b, 1d learn about dest X via iBGP from next-hop 1c: “path to X goes through 1c”
- 1d: intra-domain routing: to get to 1c, forward over outgoing local interface 1

Setting forwarding table entries



dest	Next-Hop
…	…
X	3a
…	…

NH	interface
…	…
3a	2
…	…

- recall: 1c learns about dest X via eBGP from next-hop 3a: "path to X goes through 3a"
- 1c: to get to link-local neighbor 3a, forward out interface 2

Summary: Computing forwarding table

- Intra- and inter-domain protocols collaborate to form the final forwarding table at each router

- eBGP next hop is the external router that provided the announcement

- iBGP next hop is the internal router that is used to reach the eBGP next hop