The Network Layer:
Inter-Domain Routing

CS 352, Lecture 17, Spring 2020

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

Srinivas Narayana

RUTGERS
UNIVERSITY | NEW BRUNSWICK

Course announcements

- Mid-term: Sakai scores you received were out of 30
  - We will grade the remaining 15 points manually
- Project 2 due this Friday
- Project 3 will go out this weekend
  - Small, but tricky. Start early.

Routing protocols

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

Distance vector protocols
e.g., RIP, IGRP

Path vector protocols
e.g., BGP

Intra-AS protocols

- same protocol within an AS
- different algorithms across ASes
- Also called interior gateway protocols (IGP)

Inter-AS protocols

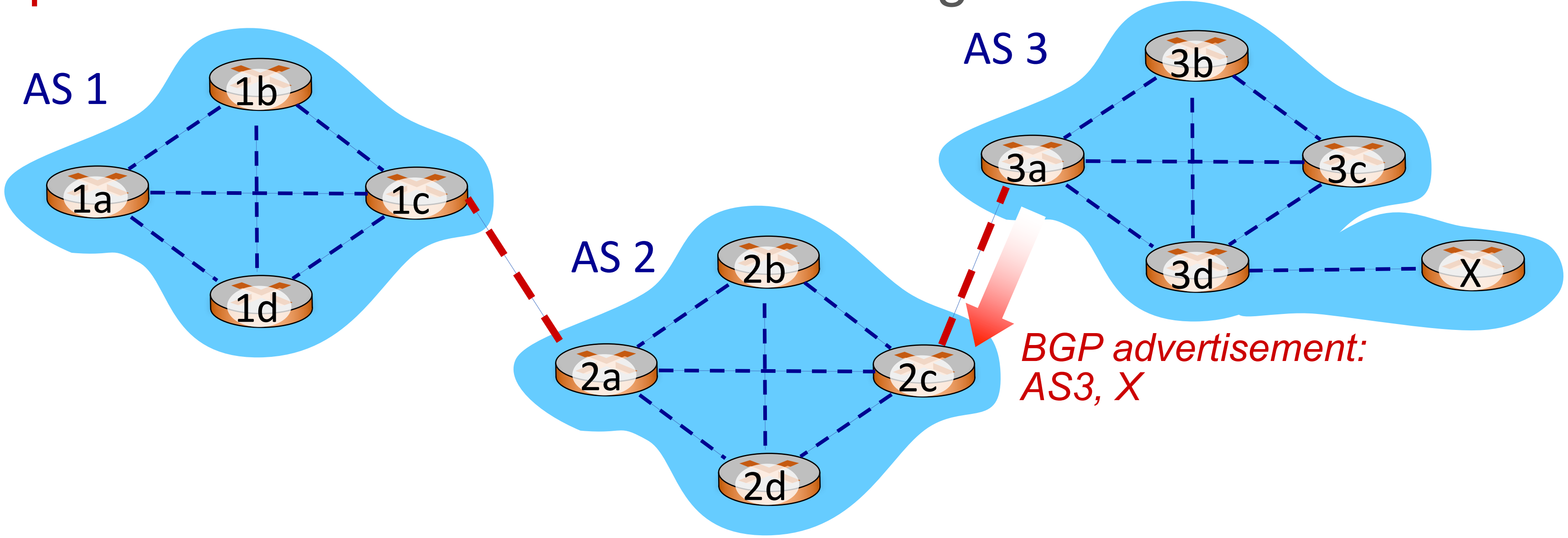
- common across Ases
- each AS knows little about the others
- eBGP, iBGP, gateway routers

Border Gateway Protocol (BGP)

The glue that holds the Internet together

BGP basics

- BGP session: two BGP routers ("peers") exchange BGP messages over semi-permanent TCP connection:
  - advertising paths to different destination network prefixes
  - (compare to distance vectors and link state)
- When AS3 gateway router 3a advertises path AS3,X to AS2 gateway router 2c,
  - AS3 promises to AS2 it will forward datagrams towards X



Path attributes and BGP routes

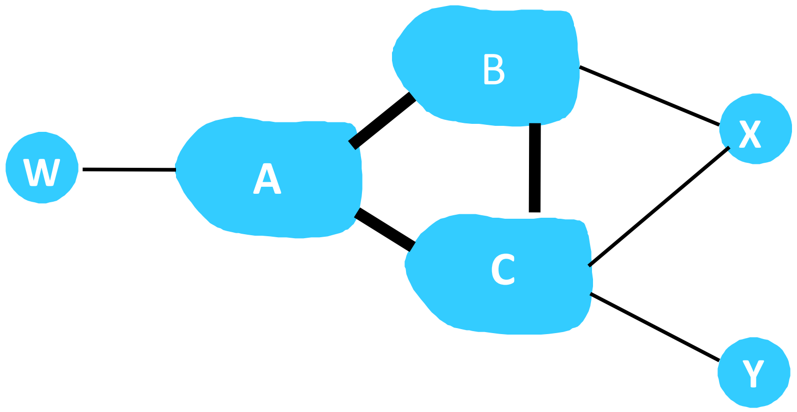
- advertised prefix includes BGP attributes
  - Advertisement of a route = prefix + attributes
- Two important attributes:
  - AS-PATH: list of ASes through which prefix advertisement has passed
  - NEXT-HOP: indicates specific internal-AS router to next-hop AS
- Policy-based routing:
  - gateway receiving route advertisement uses import policy to accept/decline path (e.g., never route through AS Y).
  - AS export policy also determines whether to advertise a path to other other neighboring ASes

Policies in BGP

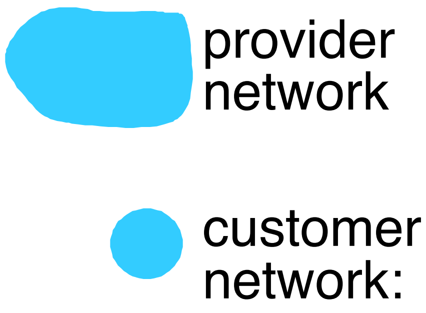
Policy comes from business relationships

- 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
- Sometimes, even when there is no direct connectivity
  - E.g., inteliquent (zoom/webex) traffic not to be charged, acc. to the FCC
- Internet-eXchange Points (IXPs) are large PoPs where ISPs come together to connect with each other (often for free)

BGP Export Policy and Advertisements



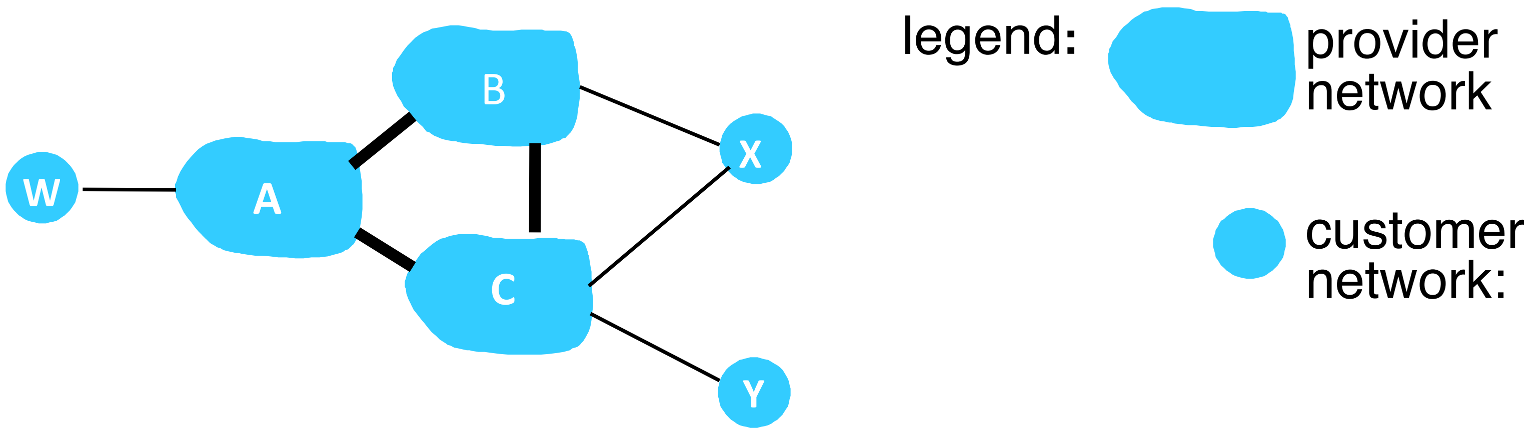
legend:



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 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

BGP Export Policy and Advertisements



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

- A advertises path Aw to B and to C
- 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

Policies make BGP a complex protocol.

Advertise entire paths, not just local info (like link state or distance vectors).

Choose to advertise (export) only certain paths.

Choose to accept (import) only certain paths.

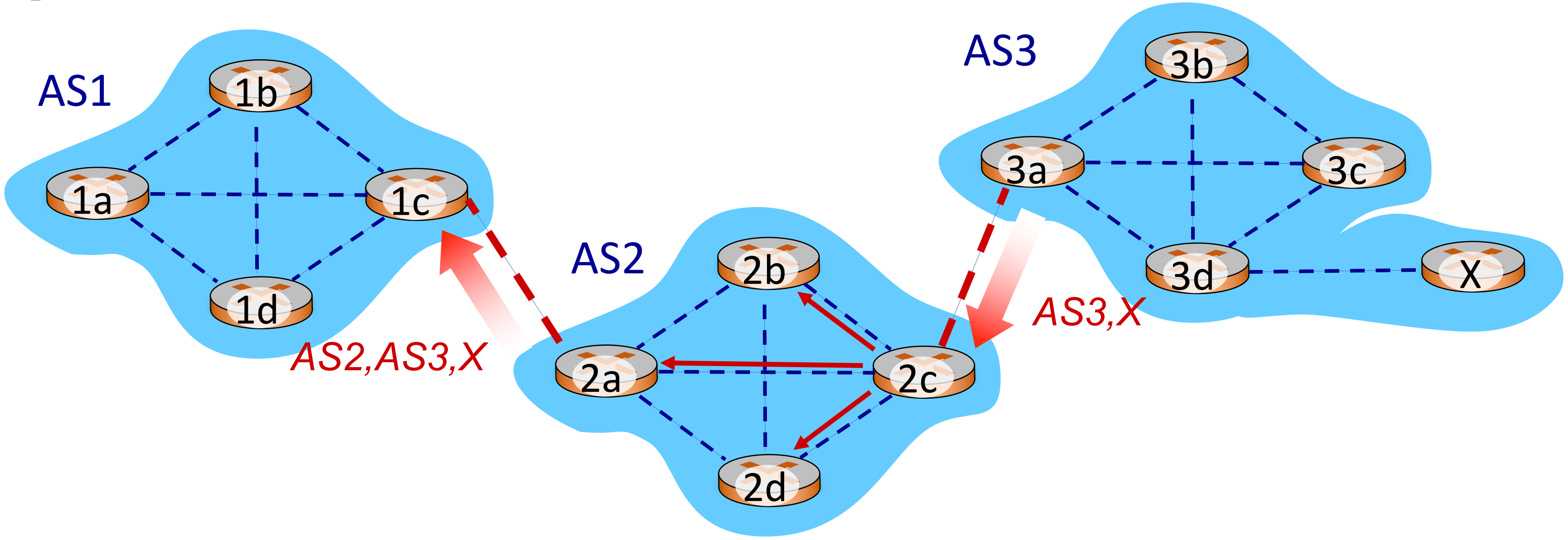
Complex decision process to prefer certain imported paths over others.

Poll #1

- What may be a legitimate business policy used by a BGP-speaking AS?
  - (a) Don’t advertise to one provider paths to another provider
  - (b) Don’t advertise to a peer paths to another peer
  - (c) Do advertise to a customer paths to other customers
  - (d) Any of the above

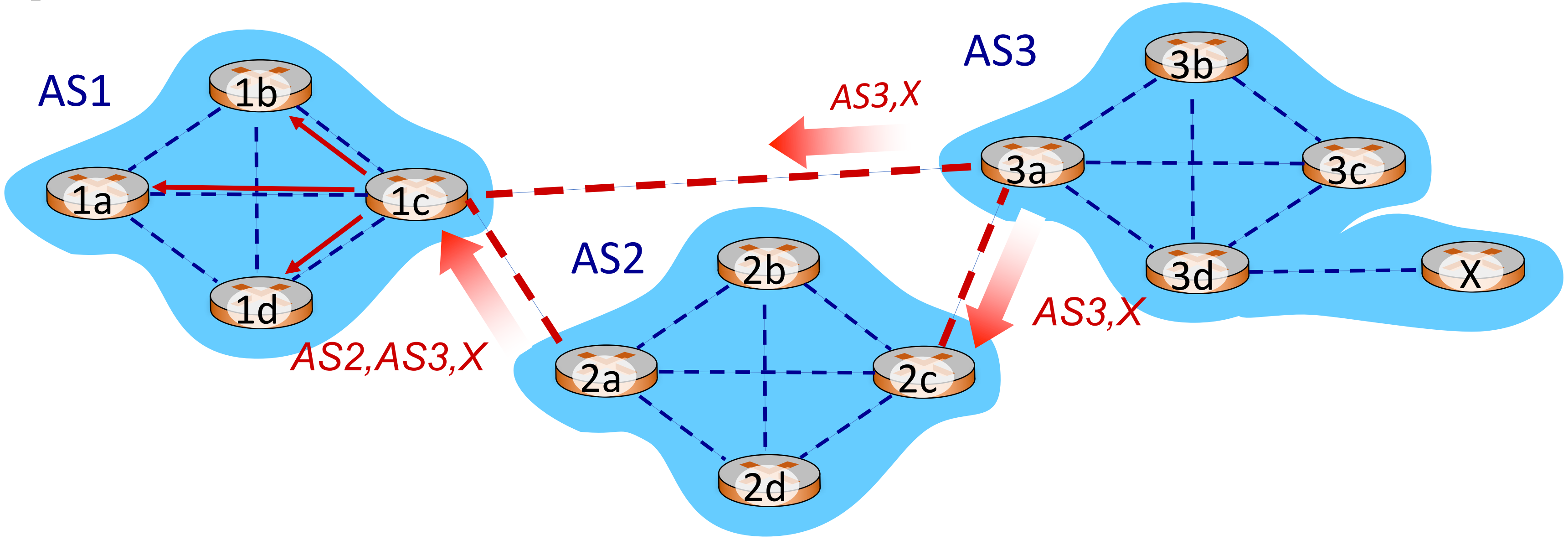
BGP Routing

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:

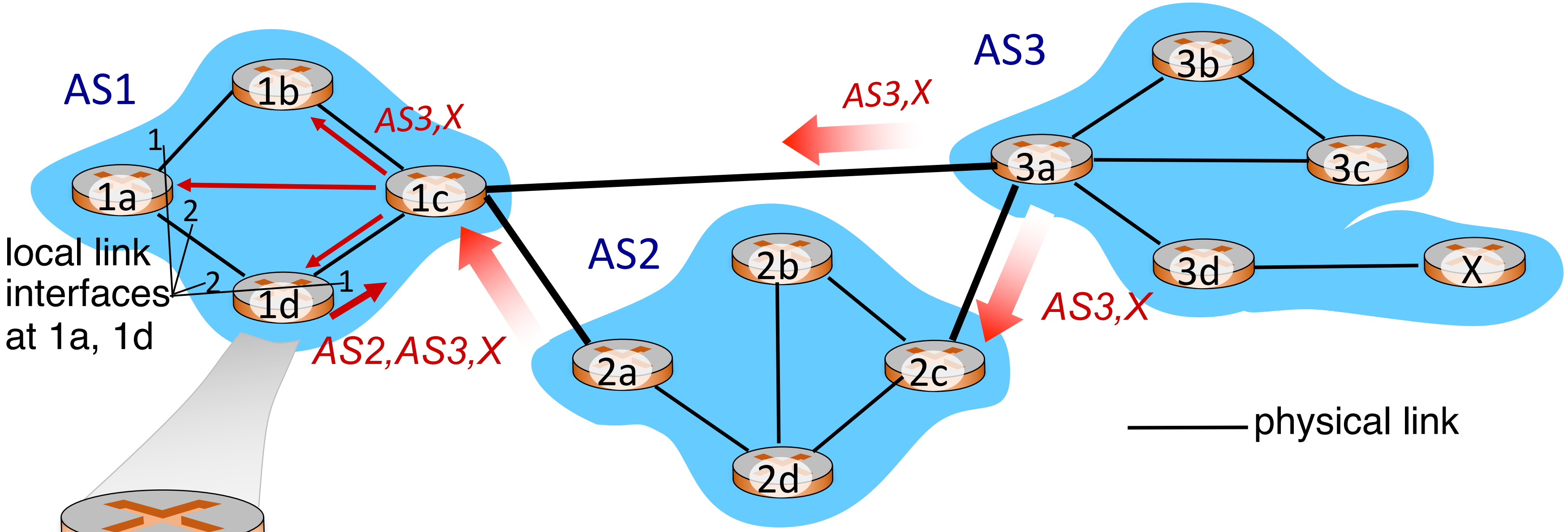
- AS1 gateway router 1c learns path *AS2,AS3,X* from 2a
- AS1 gateway router 1c learns path *AS3,X* from 3a
- Based on policy, AS1 gateway router 1c chooses path *AS3,X, and advertises path within AS1 via iBGP*

BGP messages

- BGP messages exchanged between peers over TCP connection
    - In principle, can establish BGP session with any router
        - Common, but not necessary, that routers are physically adjacent
- 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 previous msg; also used to close connection

BGP, OSPF, forwarding table entries

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



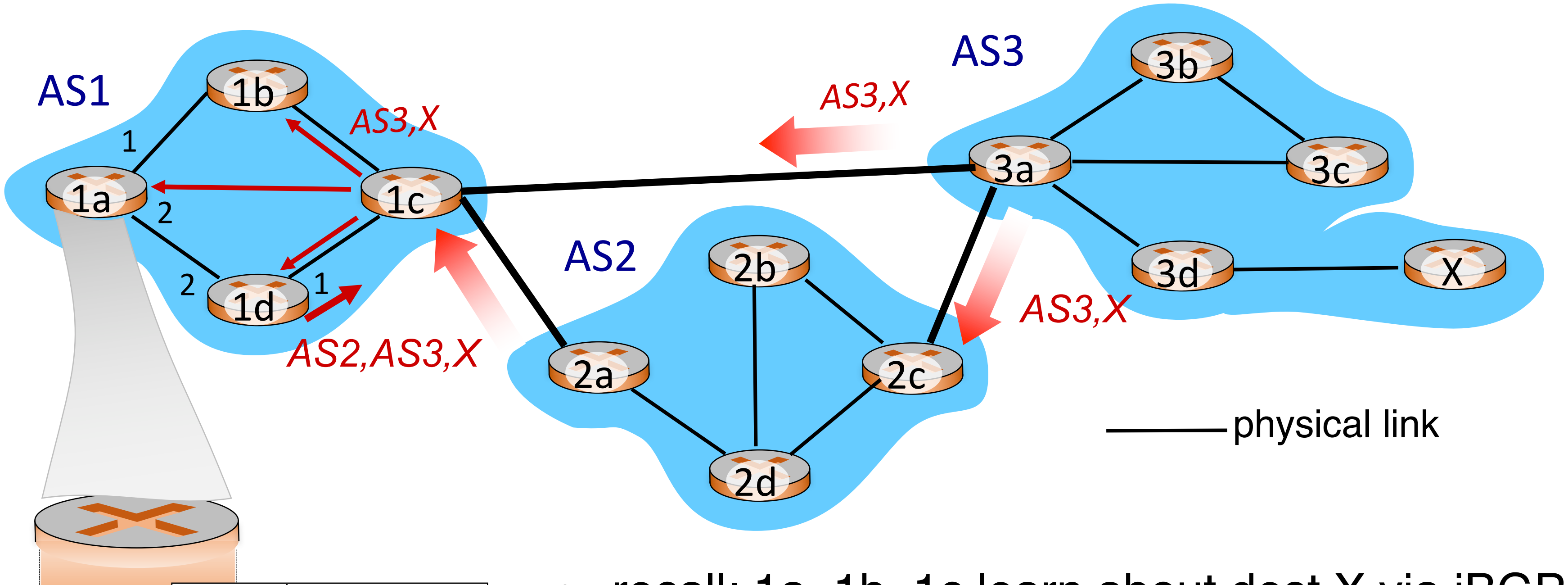
dest	Next-Hop
…	…
X	1c
…	…

NH	interface
…	…
1c	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

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



dest	Next-Hop
…	…
X	1c
…	…

NH	interface
…	…
1c	2
…	…

- 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

Poll #2

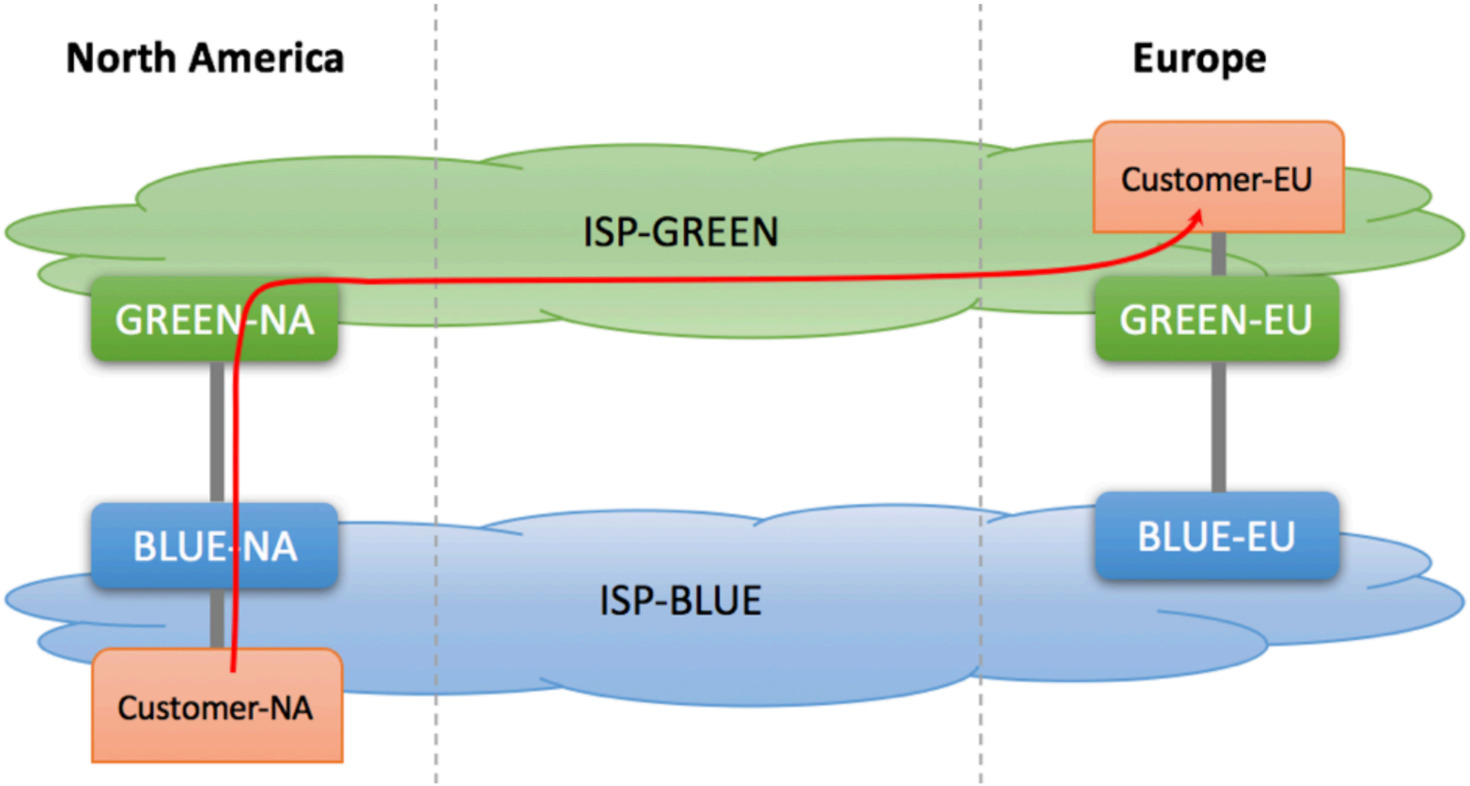
- Suppose an AS uses OSPF as its intra-domain routing protocol.
- Forwarding table entries on AS-internal routers towards destinations outside the AS are computed using information from
  - (a) iBGP
  - (b) OSPF
  - (c) both iBGP and OSPF
  - (d) None of the above

BGP route selection process

- 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

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

Hot-Potato Routing



BGP Hot Potato Routing

Also called early-exit routing

Choose the "next-hop" router that is closest based on intra-AS routing

Reduces utilization on resources inside the AS

Source: http://bgphelp.com/2017/04/25/hot-potato-vs-cold-potato-routing/

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

Quality of Service

How can the network make application performance better?

Network support for applications

- A best effort Internet architecture does not offer any guarantees on delay, bandwidth, and loss
  - Network may drop, reorder, corrupt packets
  - Network may treat transfers randomly regardless of their “importance”
- However, many apps require delay and loss bounds
  - E.g., voice over IP (phone calls) require strict delay guarantees
  - E.g., HD video requires a reasonable minimum bandwidth
  - E.g., remote surgery with 3D-vision requires strict sync & latency
- How to provide quality of service (QoS) for apps?
  - Provision enough resources: make the best of best effort service
  - Mechanisms to handle traffic differently based on importance

How can networks improve the quality of service for applications?

Contention between different applications occurs often in the core of the network, not just at endpoints.

If contention isn’t resolved, performance of some apps may be severely affected.

e.g., zoom session affected by massive concurrent bittorrent downloads

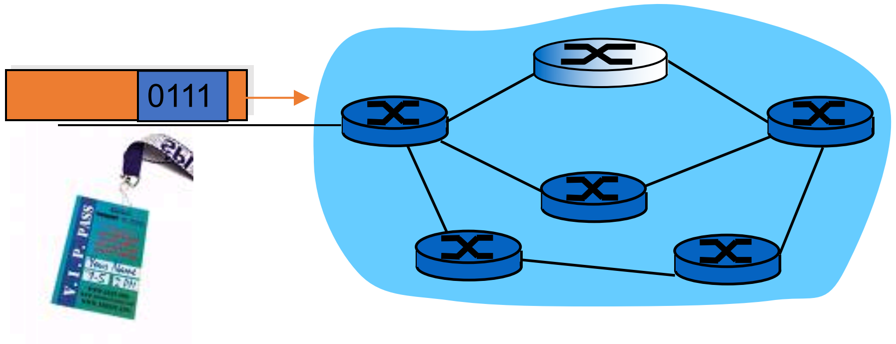
One approach: “dimension” best effort networks well

- deploy enough link capacity so that congestion doesn’t occur, multimedia traffic flows without delay or loss
  - low complexity of network mechanisms (use current “best effort” network)
  - high bandwidth costs
- challenges:
  - *network dimensioning:* how much bandwidth is “enough?”
  - *estimating network traffic demand:* needed to determine how much bandwidth is “enough” (for that much traffic)
- Network operators do this quite well, but there are exceptional circumstances.
  - Superbowl?
  - Pandemics?

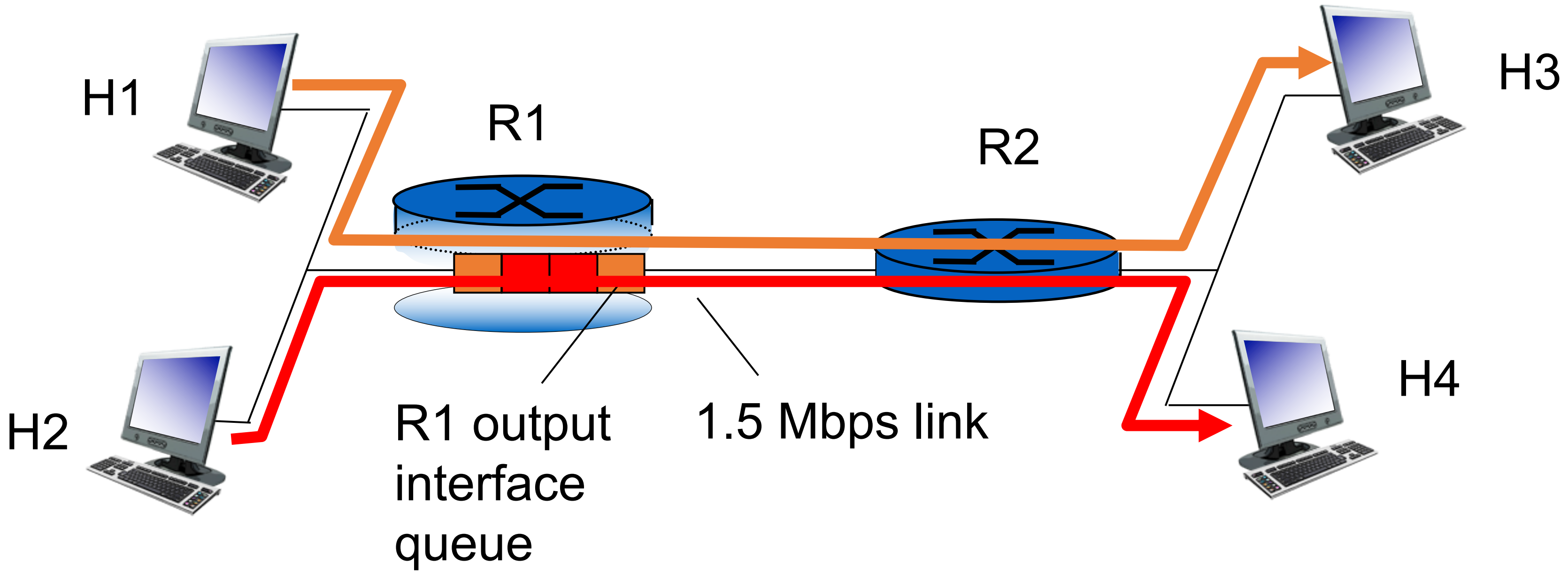
Another approach: Multiple classes of service

- Avoid "one-size fits all" service model
- Use multiple classes of service
  - partition traffic into classes
  - network treats different classes of traffic differently (analogy: premium vs. economy lines at airports)

- granularity: differential service among multiple classes, not among individual connections
- history: ToS bits in IP hdr

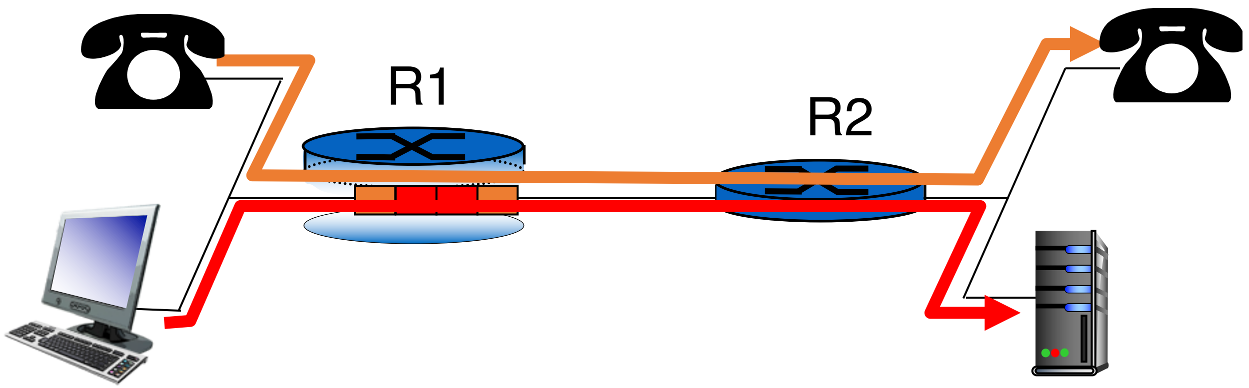


Multiple classes of service: scenario



Scenario 1: mixed HTTP and VoIP

- example: 1Mbps VoIP, HTTP share 1.5 Mbps link.
  - HTTP bursts can congest router, cause audio loss
  - want to give priority to audio over HTTP

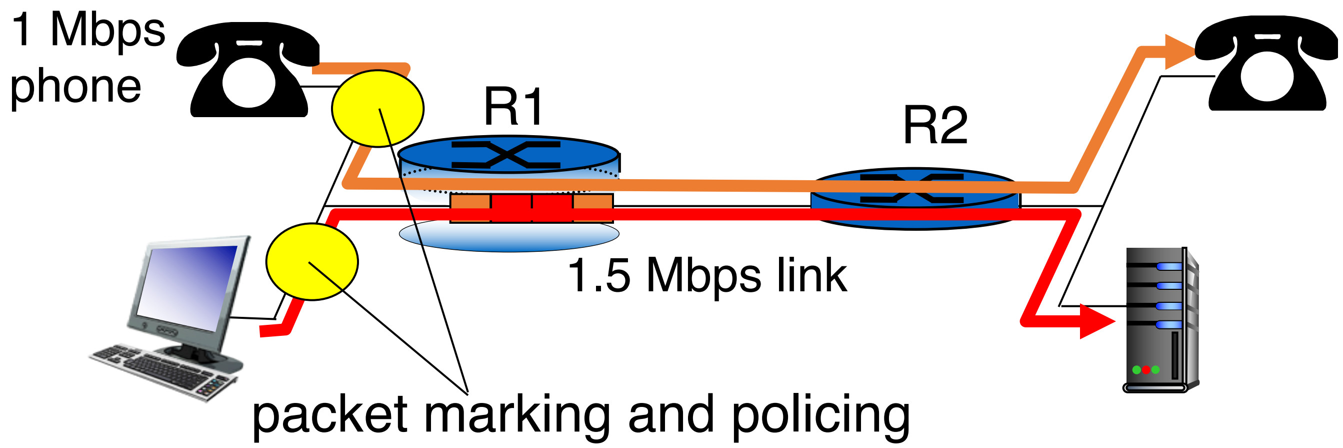


Principle 1

packet marking needed for router to distinguish between different classes; and new router policy to treat packets accordingly

Principles for QOS guarantees (more)

- what if applications misbehave (VoIP sends higher than declared rate)
  - policing: force source adherence to bandwidth allocations
- *marking*, *policing* at network edge

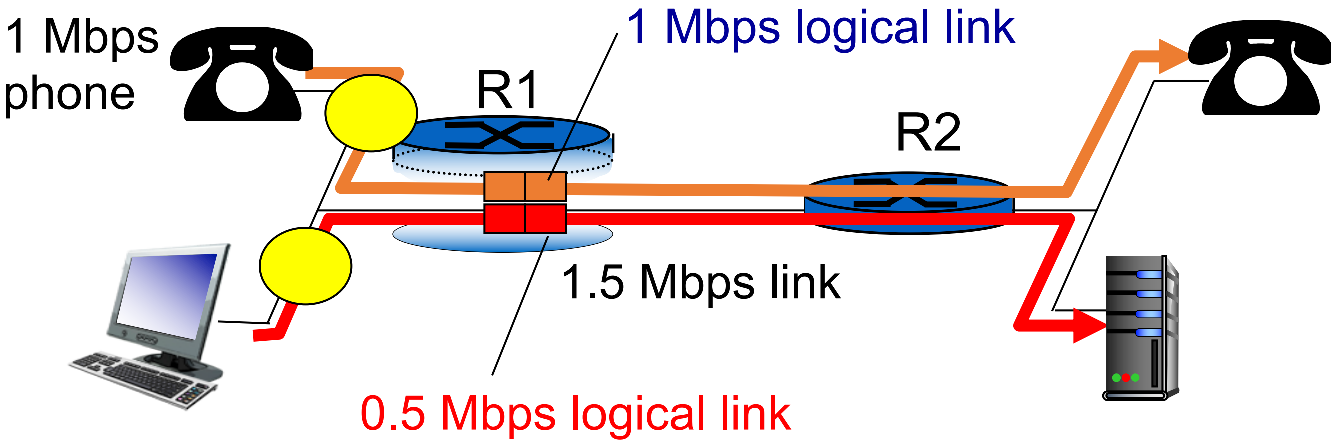


Principle 2

provide protection (isolation) for one class from others

Principles for QoS guarantees (more)

- allocating *fixed* (non-sharable) bandwidth to flow: *inefficient* use of bandwidth if flows doesn't use its allocation



Principle 3

while providing isolation, it is desirable to use resources as efficiently as possible

Work conservation

Poll #3

- Where does contention between different traffic classes (e.g., resulting in long queues) typically occur within routers?
  - (a) switch fabric
  - (b) input line termination
  - (c) output port buffers
  - (d) forwarding table

Poll #4

- What router mechanisms might you use to implement quality of service mechanisms?
  - (a) forwarding
  - (b) scheduling
  - (c) buffer management
  - (d) switching