CS3640

Network Layer (5): Routing Protocols

Prof. Supreeth Shastri
Computer Science
The University of Iowa

Programming Project-2: **Implement Ping**

An important tool for network programmers and users

- implement using ICMP and raw sockets
- No sample code this time; expect to write ~100 LoC

Due: Apr 22

Written Assignment-3: **Numerical Problems**

Serves as a reference for numerical problems to expect in the finals

- there will be one more written assignment (on research topics)
- all written assignments are to be done individually (expect to spend ~2 hours on this)

ICMP: Internet Control Message Protocol

Network Working Group   J. Postel
Request for Comments: 792   ISI
                                                    September 1981

Updates: RFCs 777, 760
Updates: IENs 109, 128

INTERNET CONTROL MESSAGE PROTOCOL

DARPA INTERNET PROGRAM
PROTOCOL SPECIFICATION

Introduction

The Internet Protocol (IP) [1] is used for host-to-host datagram service in a system of interconnected networks called the Catenet [2]. The network connecting devices are called Gateways. These gateways communicate between themselves for control purposes via a Gateway to Gateway Protocol (GGP) [3,4]. Occasionally a gateway or destination host will communicate with a source host, for example, to report an error in datagram processing. For such purposes this protocol, the Internet Control Message Protocol (ICMP), is used. ICMP, uses the basic support of IP as if it were a higher level protocol, however, ICMP is actually an integral part of IP, and must be implemented by every IP module.

ICMP Message Format

<table>
<tr><th></th><th>Bits 0–7</th><th>Bits 8–15</th><th>Bits 16–23</th><th>Bits 24–31</th></tr>
<tr><td rowspan="2">ICMP Header (8 bytes)</td><td>Type of message</td><td>Code</td><td colspan="2">Checksum</td></tr>
<tr><td colspan="4">Header Data</td></tr>
<tr><td>ICMP Payload (optional)</td><td colspan="4">Payload Data</td></tr>
</table>

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
8	0	echo request (ping)
9	0	route advertisement
10	0	router discovery
11	0	TTL expired
12	0	bad IP header

For echo request and reply

- Type == 8 (for request) or 0 (for reply)
- Code == 0 (for both)
- Header Data == 16-bit ID and 16-bit sequence number
- Payload Data == NULL

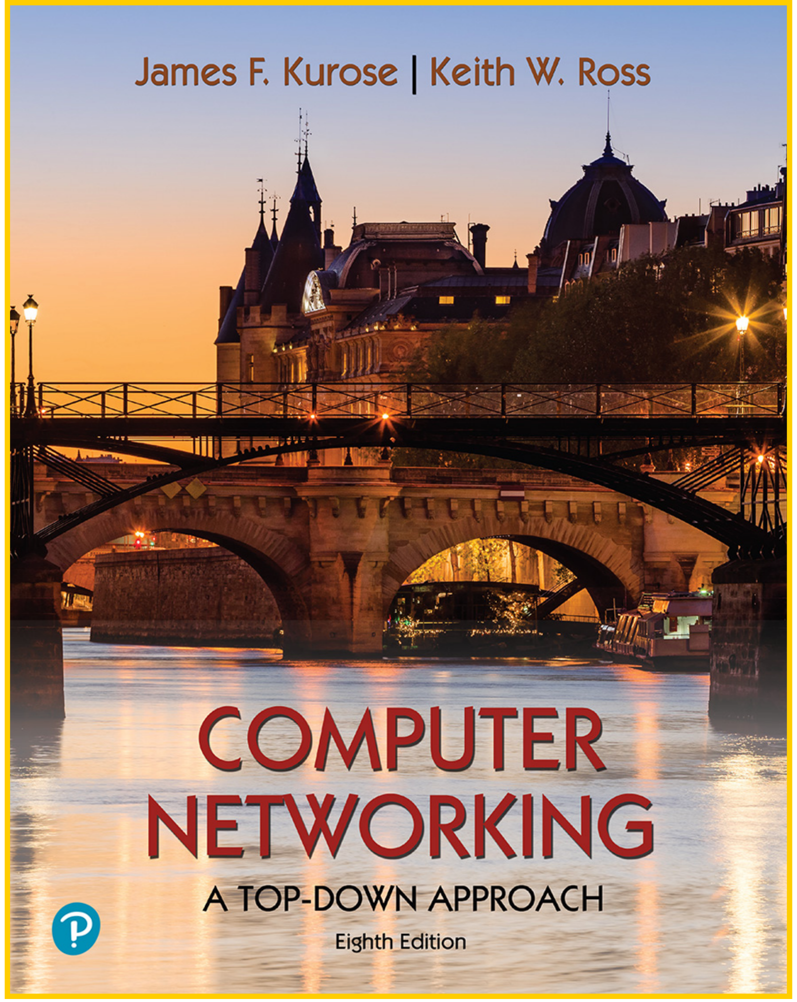
How to Send and Receive ICMP Messages

- Socket API has a special type: SOCK_RAW that allows direct access to IP datagrams

Lecture goals

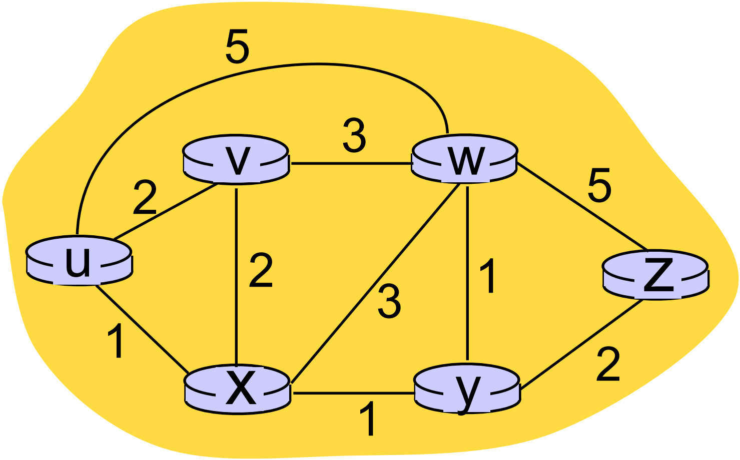
translating routing algorithms to protocols and practice on the Internet

- *OSPF: intra-AS routing*
- *BGP: inter-AS routing*



Chapters 5.3 - 5.4

Our study of routing assumed an idealized view…



- all routers are **identical**
- network has a **flat structure**
- routers work **cooperatively** to route packets efficiently to their destination
- … none of these are true in practice!

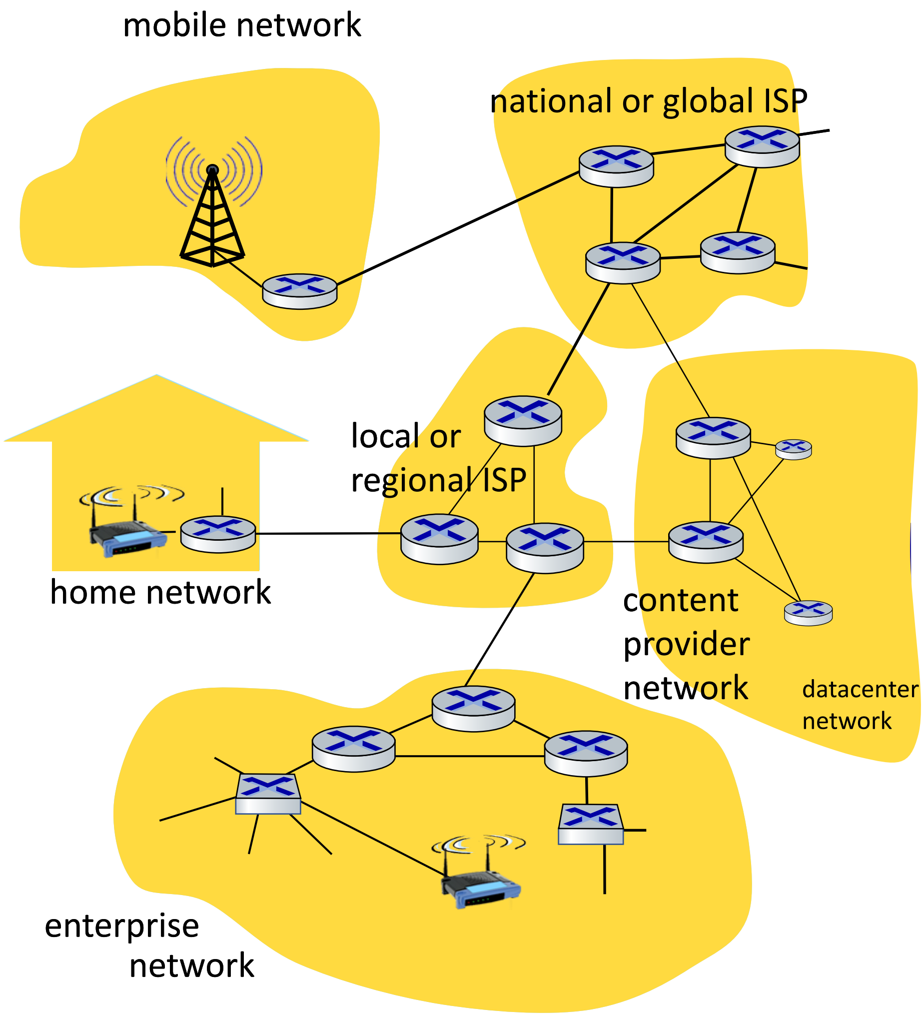
A more realistic picture of the Internet routing

Administrative autonomy

- Internet is a network of networks
- Each network administrator would want to control routing in their own network

Scale: billions of destinations

- can't store all destinations in routing tables
- exchanging link-state or distance-vector information would swamp links



The Internet’s Approach to Routing

aggregate routers into regions known as “domains” or “autonomous systems” (AS)

Intra-AS (or intra-domain) routing

routing within a given AS

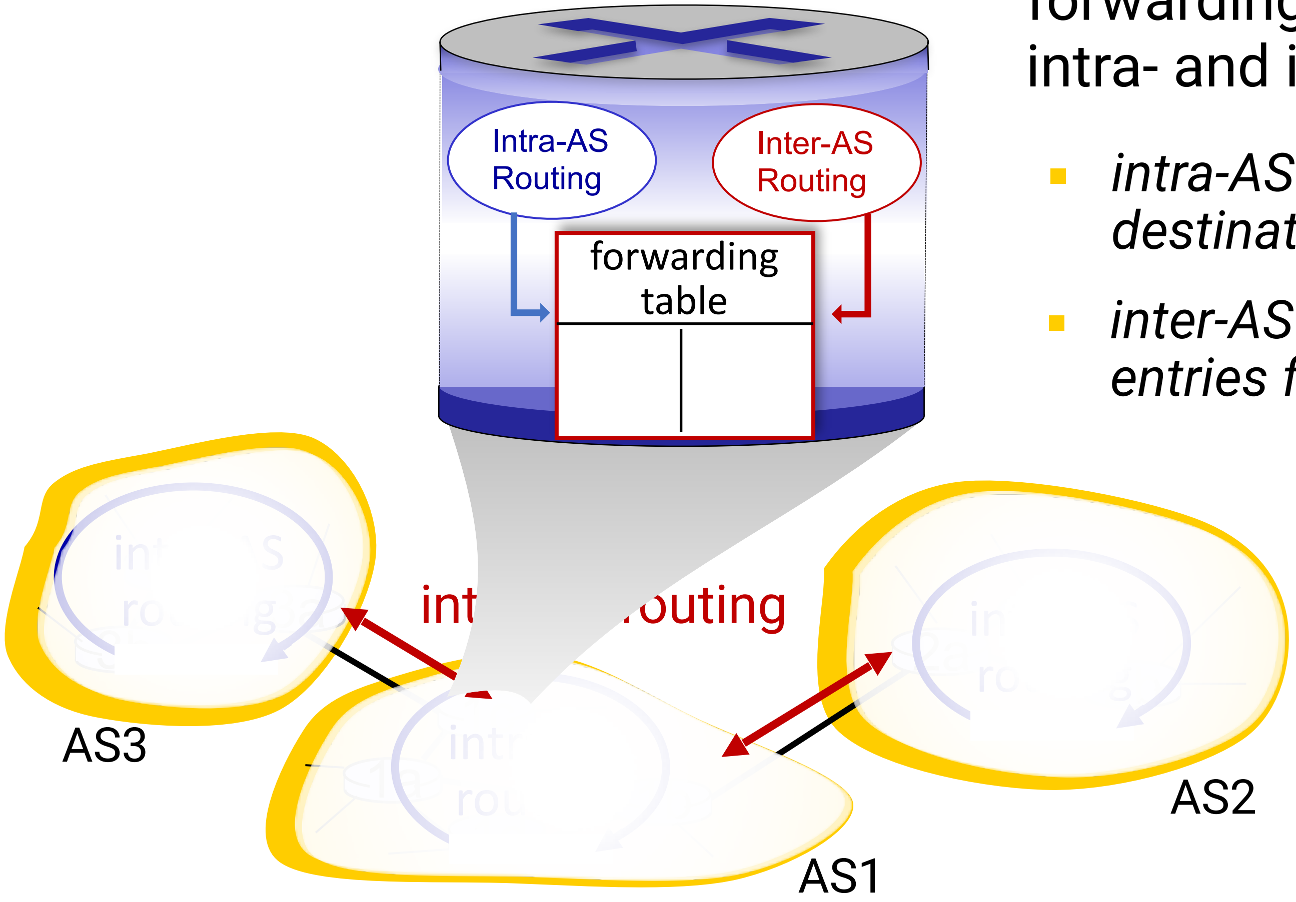
- all routers inside an AS must run the **same** intra-AS routing protocol
- **gateway router**: a router at the edge of an AS that has links to routers in other AS

Inter-AS (or inter-domain) routing

routing across different AS’es

- while AS’es are free to run any intra-AS routing protocols, they must run the **same** inter-AS protocol
- gateway routers perform both inter-domain as well as intra-domain routing

Constructing the Forwarding Table



forwarding table is configured by both intra- and inter-AS routing algorithms

- *intra-AS routing determines the entries for destinations within AS*
- *inter-AS & intra-AS together determine the entries for external destinations*

Open Shortest Path First

a widely used intra-domain routing protocol

Network Working Group
Request for Comments: 2328
STD: 54
Obsoletes: 2178
Category: Standards Track
J. Moy
Ascend Communications, Inc.
April 1998

OSPF Version 2

Status of this Memo

This document specifies an Internet standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "Internet Official Protocol Standards" (STD 1) for the standardization state and status of this protocol. Distribution of this memo is unlimited.

Copyright Notice

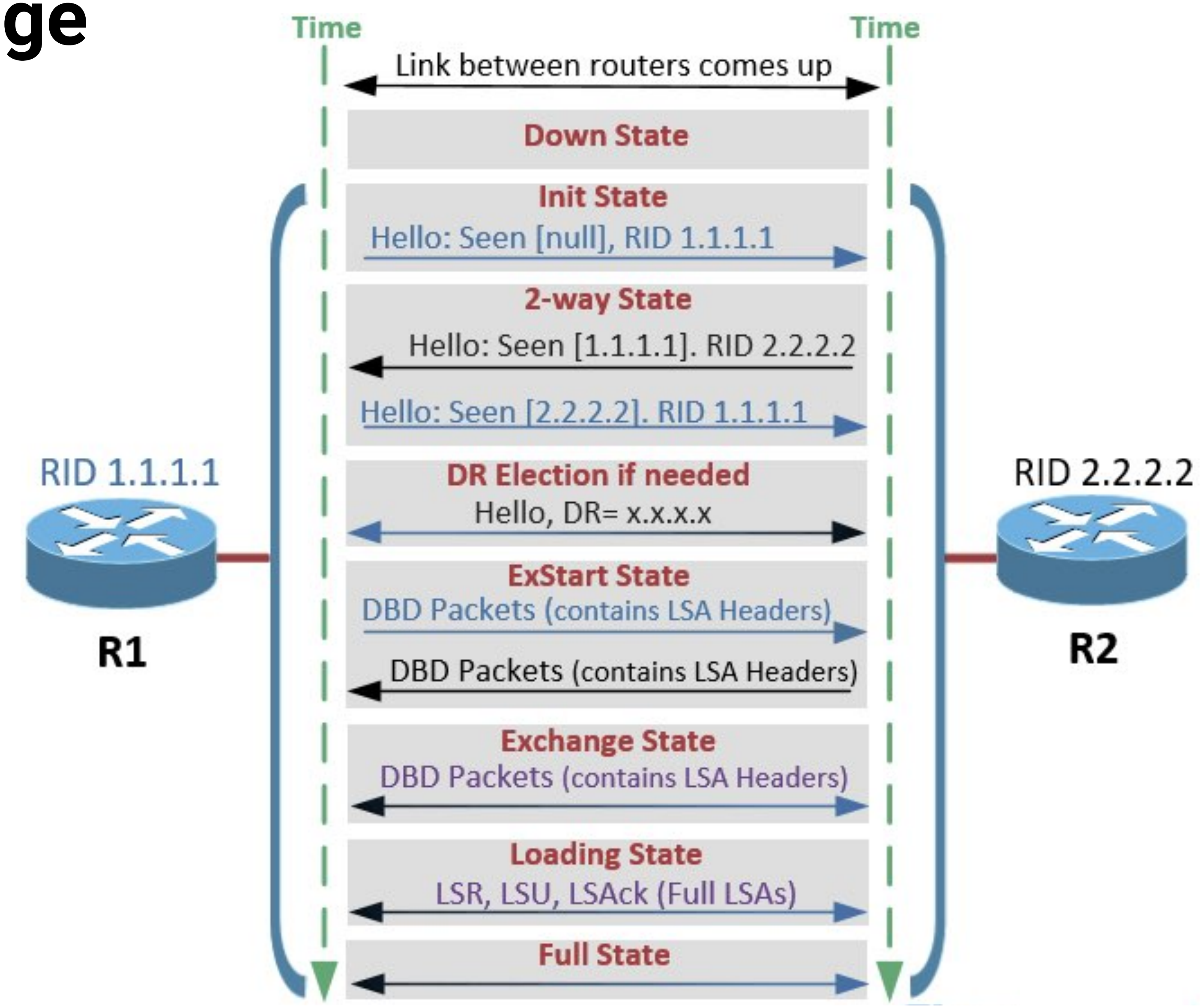
Abstract

This memo documents version 2 of the OSPF protocol. OSPF is a link-state routing protocol. It is designed to be run internal to a single Autonomous System. Each OSPF router maintains an identical database describing the Autonomous System's topology. From this database, a routing table is calculated by constructing a shortest-path tree.

OSPF: message exchange

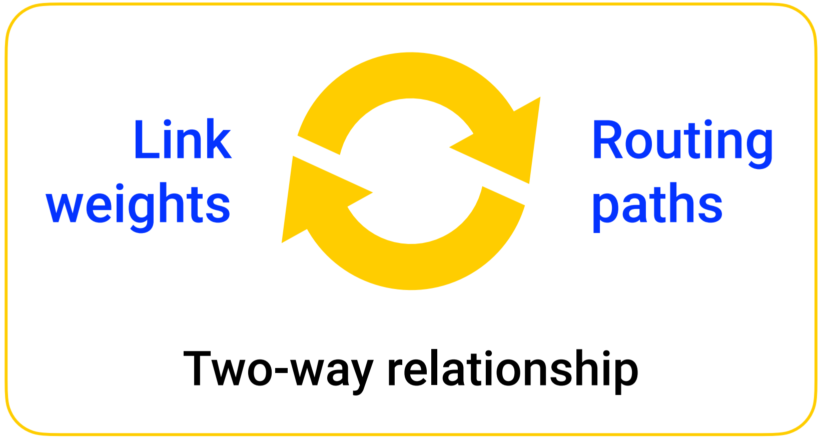
Five Messages

- **Hello**: discover and maintain adjacencies with neighboring routers
- **Database Description:** exchange contents of the link-state database with an adjacent router
- **Link-State Request**: request a specific link-state record
- **Link-State Response**: send the link-state record of a specific link(s)
- **Link-State Ack**: provide reliability to the link-state exchange process



OSPF practices and operation

- **Security**: all OSPF messages are authenticated to prevent malicious intrusions. OSPF supports MD5
- **No transport layer protocol**: all OSPF messages are sent as payload in IP datagrams. Accordingly, OSPF implements its own reliable data transport mechanism.
- **Hierarchical OSPF:** allows separating the network into multiple areas, connected via a backbone network. Each area then runs OSPF by itself and summarize routes across areas using designated ***area border routers***
- **Multiple paths**: when multiple paths to a destination have the same cost, OSPF allows multiple paths to be used (*this allows admins to perform advanced traffic engineering*)



Link weights: setting all weights == 1, results in minimum hop routing; whereas setting weights = 1/link-capacity, discourages traffic to flow through low-capacity links

Intra-AS Routing Protocols

RIP *(RFC 1723)*

Routing Information Protocol
Distance-Vector algorithm
Not widely used

OSPF *(RFC 2328)*

Open Shortest Path First
Link-state algorithm
Most commonly used

EIGRP *(RFC 7868)*

Enhanced Interior Gateway Routing
Distance-Vector algorithm
Proprietary but became open in 2013

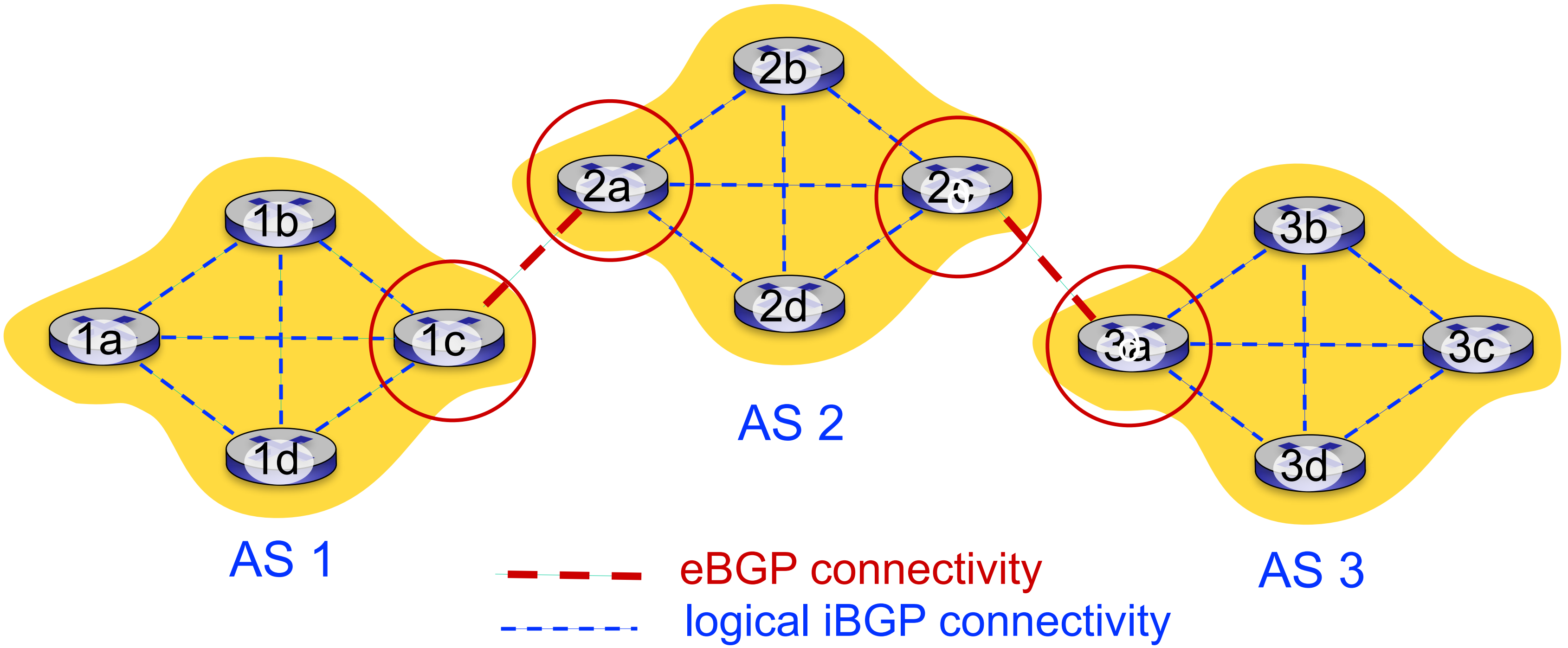
Border Gateway Protocol

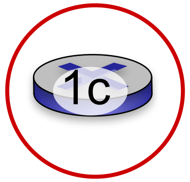
glue that holds the Internet together!

Border Gateway Protocol

- The ***de-facto*** inter-domain routing protocol. The first version was defined in RFC 1105 (in 1989); BGPv4 defined in RFC 4271 (in 2006) is the latest.
- **Goal**: allow autonomous systems to advertise their existence, the destinations they can reach, and thus, enable routing across the wide-area Internet
- More specifically, BGP provides each AS the ability:
  - ➡ to obtain destination network reachability info from neighboring AS'es (**eBGP**)
  - ➡ to compute routes to other networks based on eBGP info and policy
  - ➡ to propagate reachability information to all routers internal to the AS (**iBGP**)
  - ➡ to advertise its own reachability info to its neighboring AS'es (**eBGP**)

BGP basics: eBGP and iBGP





gateway routers run both eBGP and iBGP protocols

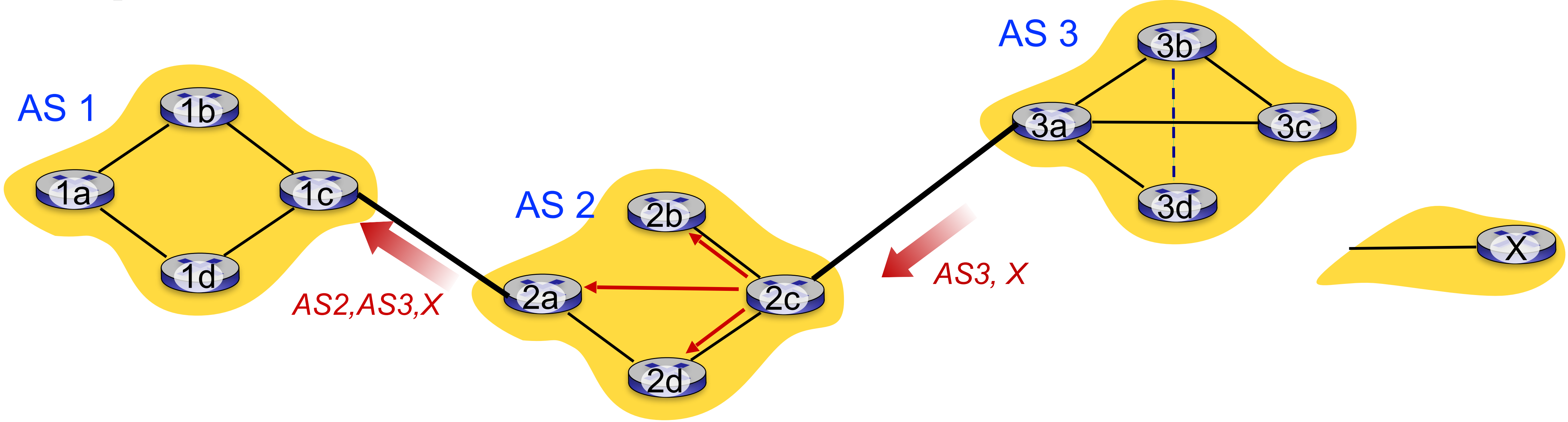
BGP basics: session and advertisements

- **BGP session**: two BGP routers exchange BGP messages over TCP connection
- BGP is a **distance vector** protocol, where advertisements contain the IP prefix of the destination network and a “path vector”
- BGP advertisement is a **promise** that the AS is *willing and able* to forward all datagrams to the advertised prefix

Policy-based routing

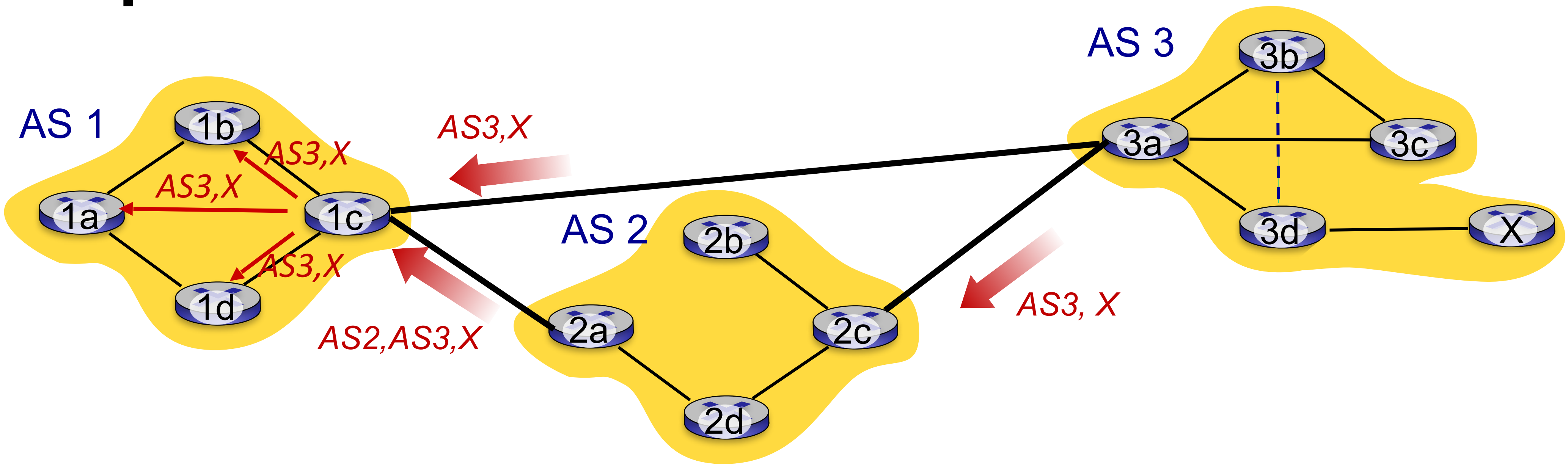
- BGP router, upon receiving route advertisement to destination X, uses policy to accept/reject a path (e.g., never route through AS W, or country Y).
- BGP router uses policy to decide whether to advertise a path to a neighboring AS Z (do I want to route traffic forwarded from Z destined to X?)

BGP path advertisements



- 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

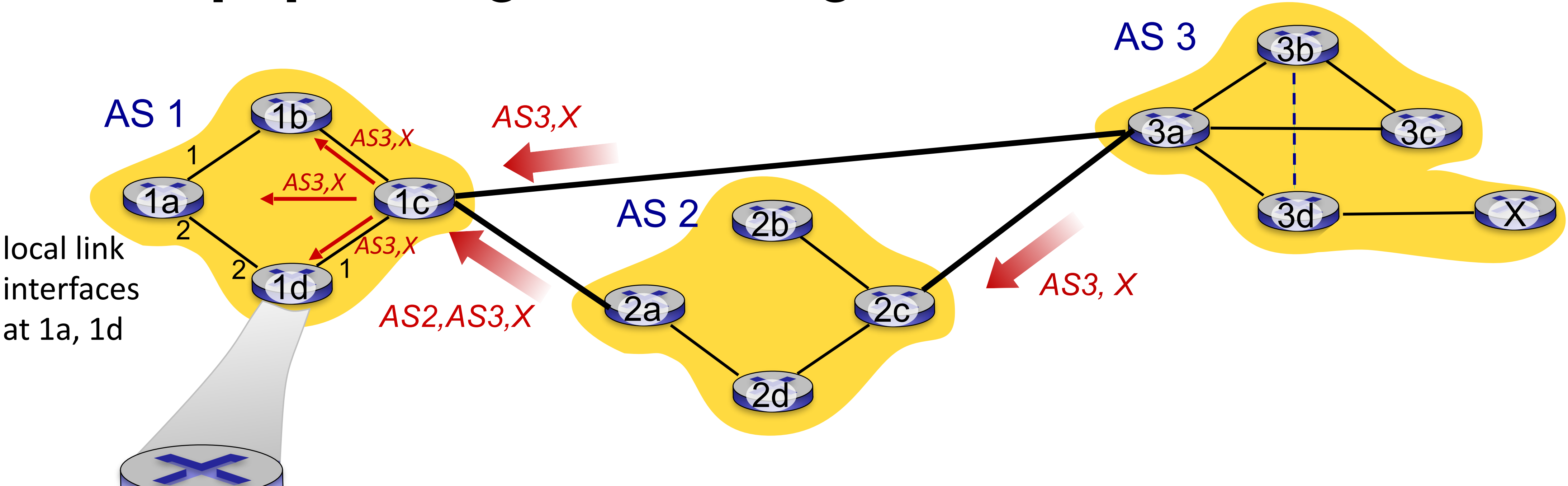
BGP path advertisements



Gateway routers 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

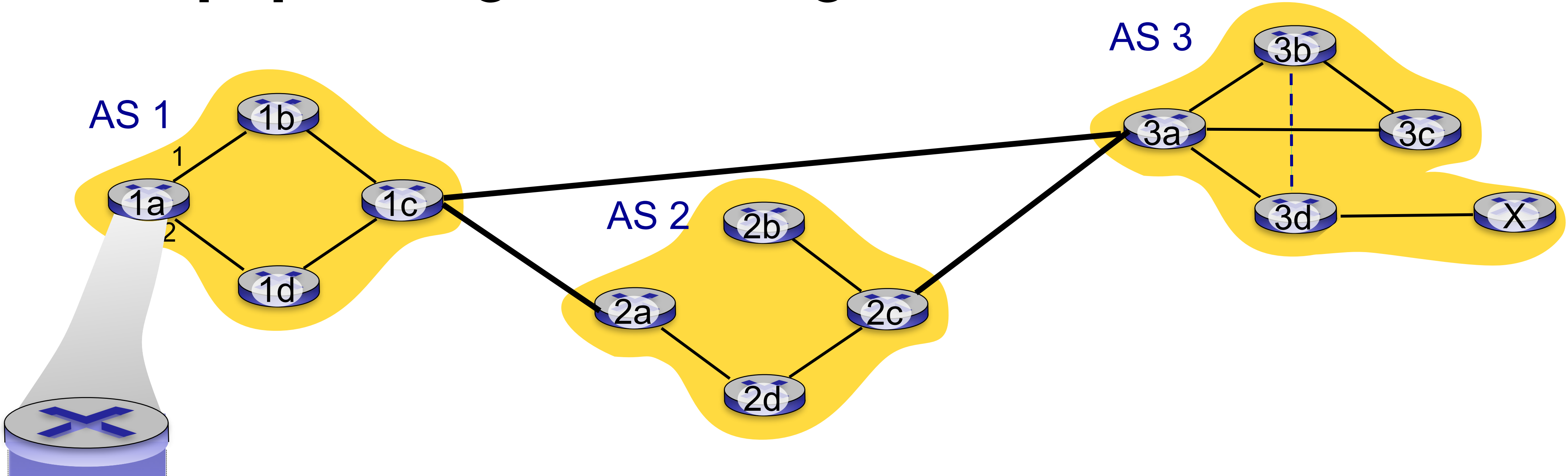
BGP: populating forwarding tables



dest	interface
…	…
1c	1
X	1
…	…

- recall: 1a, 1b, 1d learn via iBGP from 1c: "path to X goes through 1c"
- at 1d: OSPF intra-domain routing: to get to 1c, use interface 1
- at 1d: to get to X, use interface 1

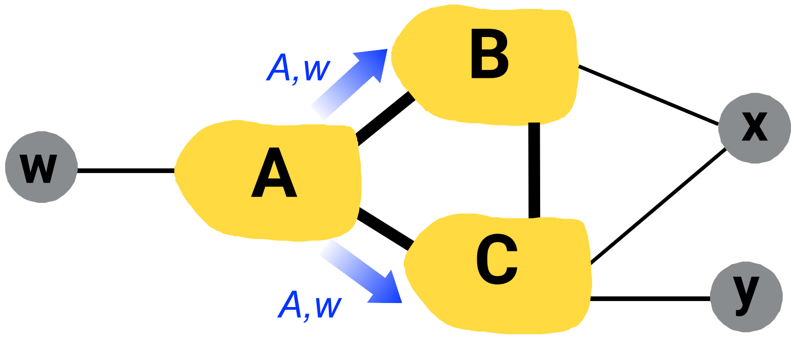
BGP: populating forwarding tables

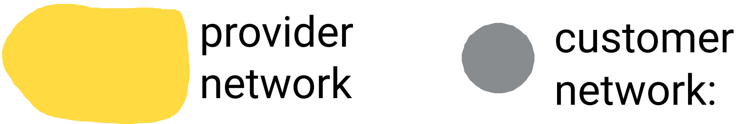


dest	interface
…	…
1c	2
X	2
…	…

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

Achieving policy via BGP advertisements

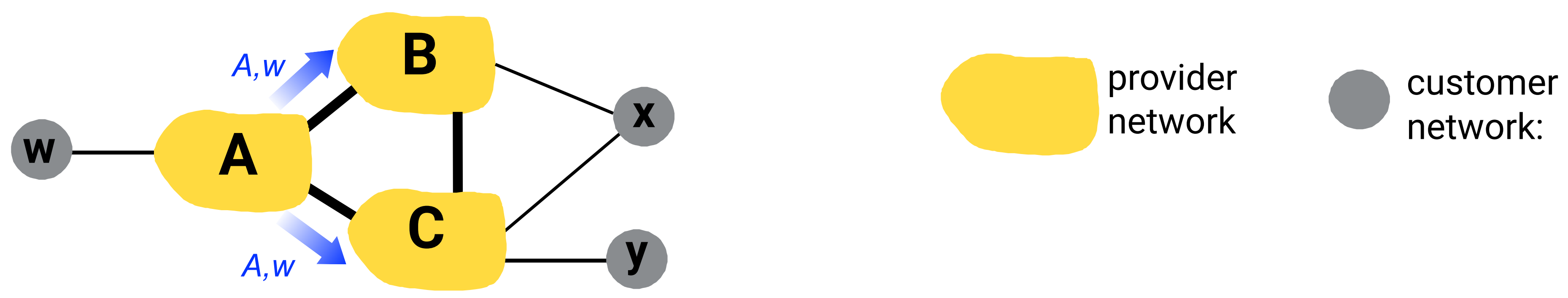




ISP only wants to route traffic to/from its customer networks (does not want to carry transit traffic between other ISPs – a typical "real world" policy)

- A advertises path **A,w** to B and to C
- B chooses not to advertise **BA,w** to C
  - ➡ B gets no "revenue" for routing **CBA,w** since none of C, A, w are B's customers
  - ➡ C does not learn about CBA,w path
- C will route **CA,w** to get to w

Achieving policy via BGP advertisements



ISP only wants to route traffic to/from its customer networks (does not want to carry transit traffic between other ISPs – a typical "real world" policy)

- A, B, C are service provider networks
- w, x, y are customers (of service provider networks)
- x is dual-homed: attached to two networks
- policy decision: x does not want to route traffic from B to C via x
  - ➡ *So, x will not advertise routes to B (or to C)*

Why different protocols for Intra- and Inter-AS routing?

Policy vs Performance

- Inter-AS: administrators want control over how the AS traffic is routed, who routes through its network (balancing economics and geopolitics)
- Intra-AS: single administration; policy not an issue, instead focus on performance

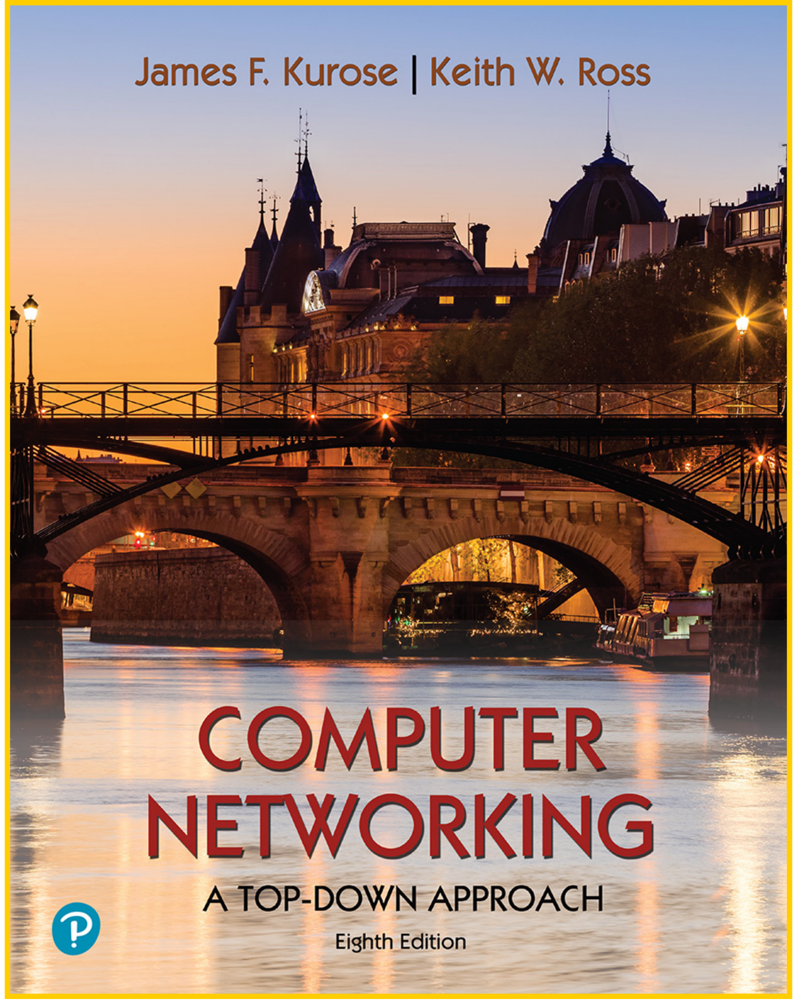
Scale: reducing forwarding table size

- Inter-AS: the routing algorithm, its data structures and protocol operations have to scale to large number of networks
- Intra-AS: scalability it not a concern. Even if an ISP becomes big, they can employ hierarchical OSPF.

Next lecture

a technical overview of the data link layer

- *Link layer services*
- *Network Interface Controller (NIC)*
- *Multiple Access Channels*



Chapters 6.1 - 6.3

Spot Quiz (ICON)