COSC264
Introduction to Computer Networks and the Internet

NAT, IPv6, RIP, OSPF, BGP

Dr. Barry Wu
Wireless Research Centre
University of Canterbury
barry.wu@canterbury.ac.nz

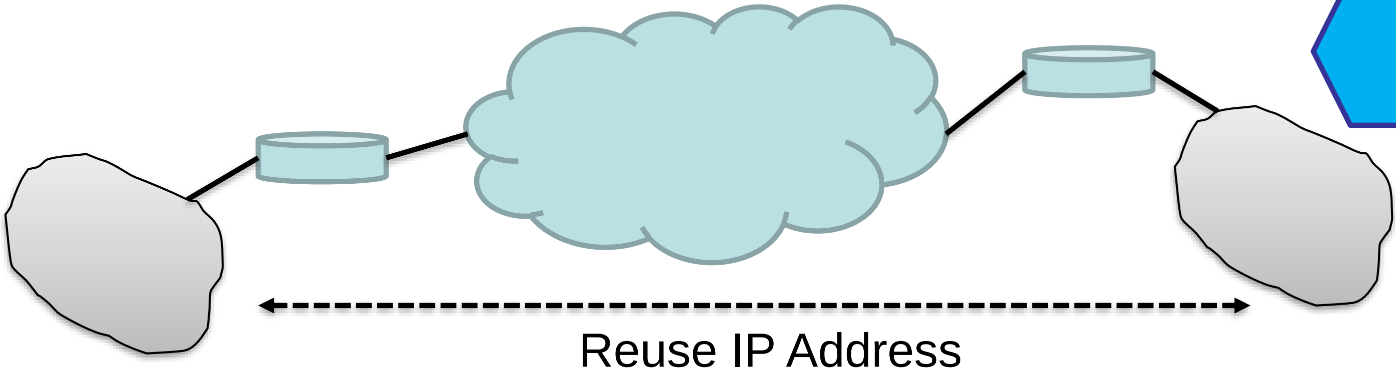
Back to the reality

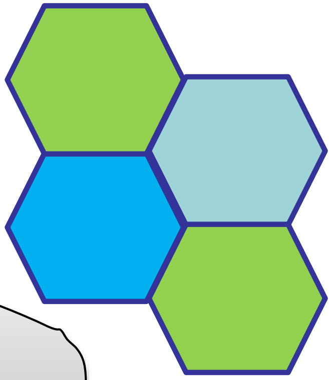
- NAT
- IPv6
- Routing in the Internet
  - Hierarchical routing
  - RIP
  - OSPF
  - BGP

NAT: Network Address Translation

- Every IP-capable device needs an IP address. (IP address depletion problem)
- Short-term solution is CIDR (Classless InterDomain Routing).
- Another solution is to reuse IP address.

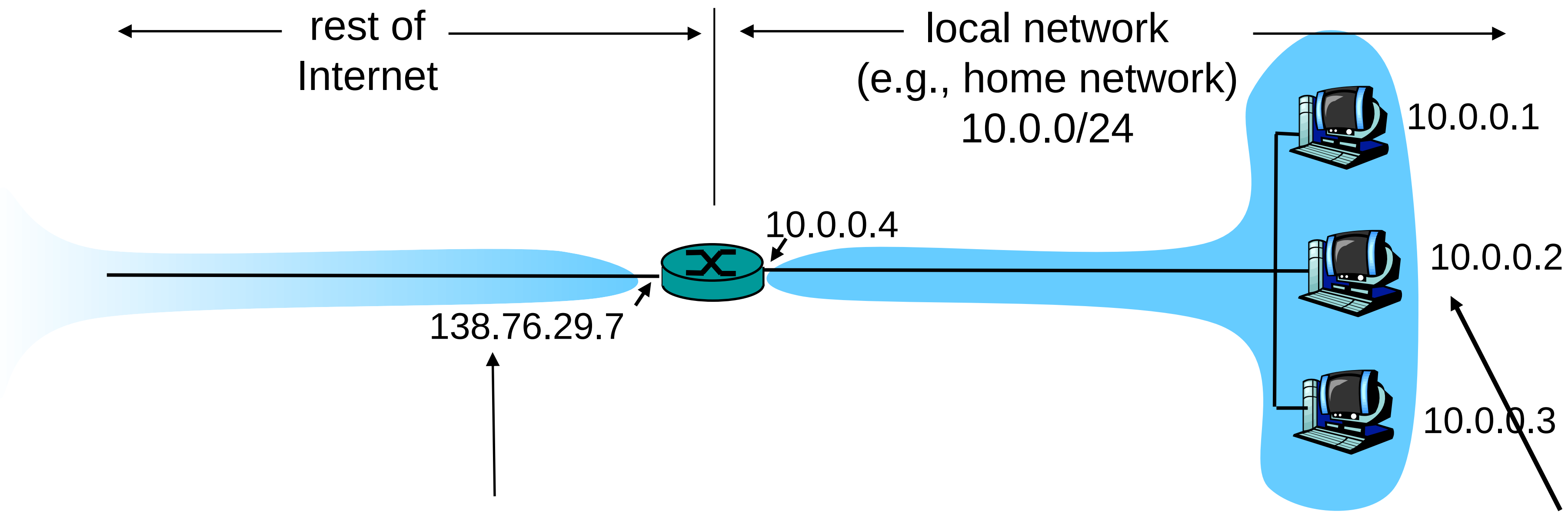
(long-term or short term? IPv6?)





NAT: Network Address Translation

How does the router know which host to forward the traffic?

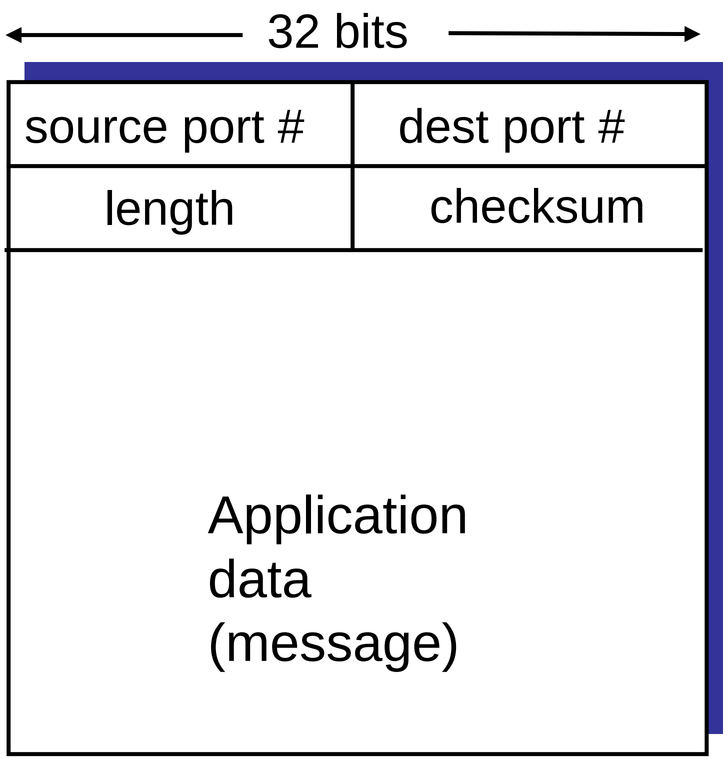


All traffic *leaving* local network have same single source IP address: 138.76.29.7; all traffic entering the home must have a destination address of 138.76.29.7.

Hosts get their IP address through DHCP (Dynamic Host Configuration Protocol).

The trick

WAN side address		LAN side address	
138.76.29.7	xxxx	10.0.0.1	yyyy
138.76.29.7	yyyy	10.0.0.2	zzzz
138.76.29.7	zzzz	10.0.0.3	xxxx



UDP segment format

Port # is used by sockets to identify processes on hosts.
80 is the default port # of HTTP.

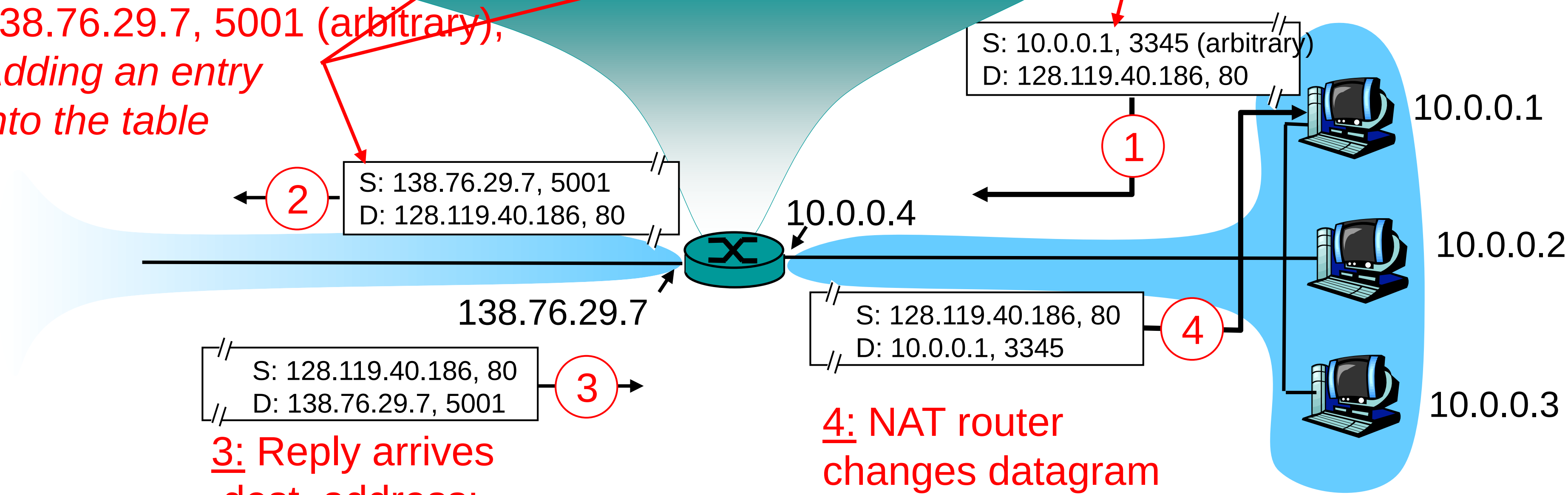
NAT: Network Address Translation

NAT translation table

WAN side addr	LAN side addr
138.76.29.7, 5001	10.0.0.1, 3345
……	……

1: host 10.0.0.1 sends datagram to 128.119.40, 80

2: NAT router changes datagram source addr from 10.0.0.1, 3345 to 138.76.29.7, 5001 (arbitrary), *Adding an entry into the table*



3: Reply arrives dest. address: 138.76.29.7, 5001

4: NAT router changes datagram dest addr from 138.76.29.7, 5001 to 10.0.0.1, 3345

NAT: Network Address Translation

- Motivation: local network uses just one IP address as far as outside world is concerned:
  - no need to be allocated range of addresses from ISP: - just one IP address is used for all devices
  - can change addresses of devices in local network without notifying outside world
  - can change ISP without changing addresses of devices in local network
  - devices inside local net not explicitly addressable, visible by outside world (a security plus).

An example

asn:university of canterbury

asn

```
Total amount of IPs for this ASN: 66,048
```

As Number	9432
As Name	University of Canterbury
CIDR Range	132.181.0.0/16
	Monitor this
As Number	9432
As Name	University of Canterbury
CIDR Range	202.36.178.0/23
	Monitor this

Source: MXtoolbox

```
Wireless LAN adapter Wi-Fi:

   Connection-specific DNS Suffix  . : canterbury.ac.nz
   Link-local IPv6 Address . . . . . : fe80::84d1:5f80:5c58:7b6c%4
   IPv4 Address. . . . . . . . . . . : 10.34.27.79
   Subnet Mask . . . . . . . . . . . : 255.255.0.0
   Default Gateway . . . . . . . . . : 10.34.254.254
```

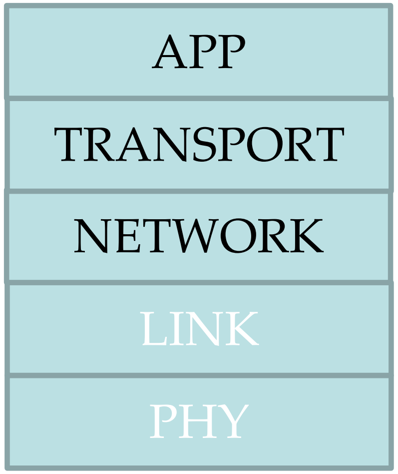
NAT: Network Address Translation

- 16-bit port-number field:
  - 60,000 simultaneous connections with a *single* WAN-side address!

WAN side address		LAN side address	
138.76.29.7	xxxx	10.0.0.1	yyyy
138.76.29.7	yyyy	10.0.0.2	zzzz
138.76.29.7	zzzz	10.0.0.3	xxxx

NAT: Network Address Translation

- NAT is controversial:
  - Port numbers are to address processes – not hosts
  - routers should only process up to layer 3; (Port # is at layer 4).
  - violates end-to-end argument
    - Hosts should talk directly to each other without a middleman.
  - address shortage should instead be solved by IPv6



Back to the reality

- NAT
- IPv6
- Routing in the Internet
  - Hierarchical routing
  - RIP
  - OSPF
  - BGP

Internet Stream Protocol

From Wikipedia, the free encyclopedia

The **Internet Stream Protocol** (**ST**) is a family of experimental protocols first defined in Internet Experiment Note IEN-119 in 1979,[1] and later substantially revised in RFC 1190 (ST-II) and RFC 1819 (ST2+).[2][3][4] The protocol uses the version number 5 in the version field of the Internet Protocol header, but was never known as IPv5.

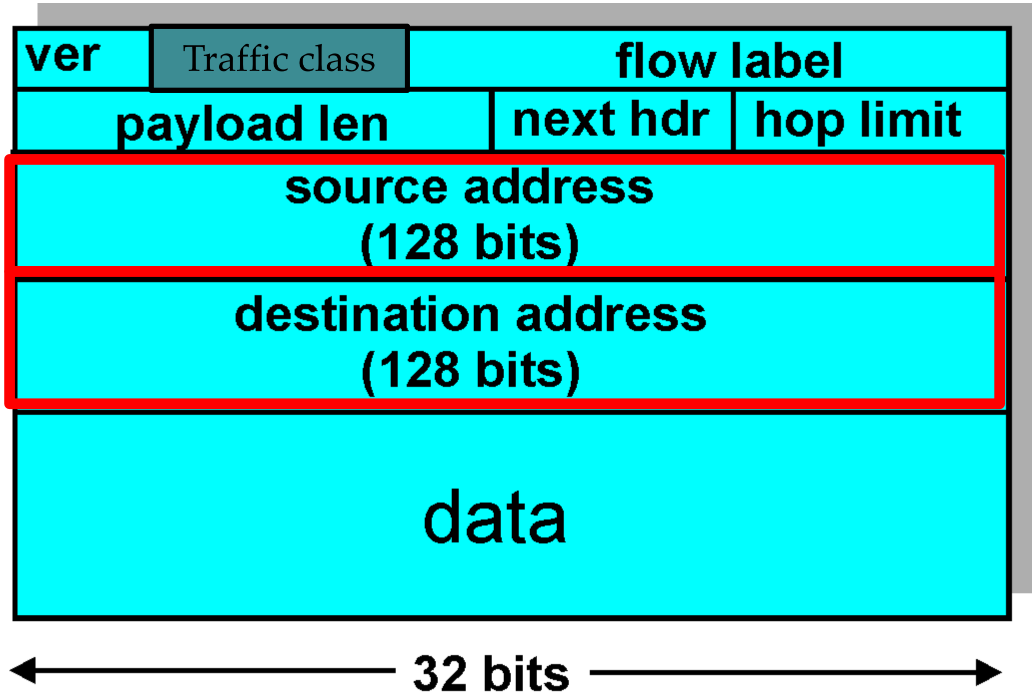
IPv6

- Initial motivation: 32-bit address space soon to be completely allocated.
- Additional motivation:
  - header format helps speed processing/forwarding
  - header changes to facilitate QoS

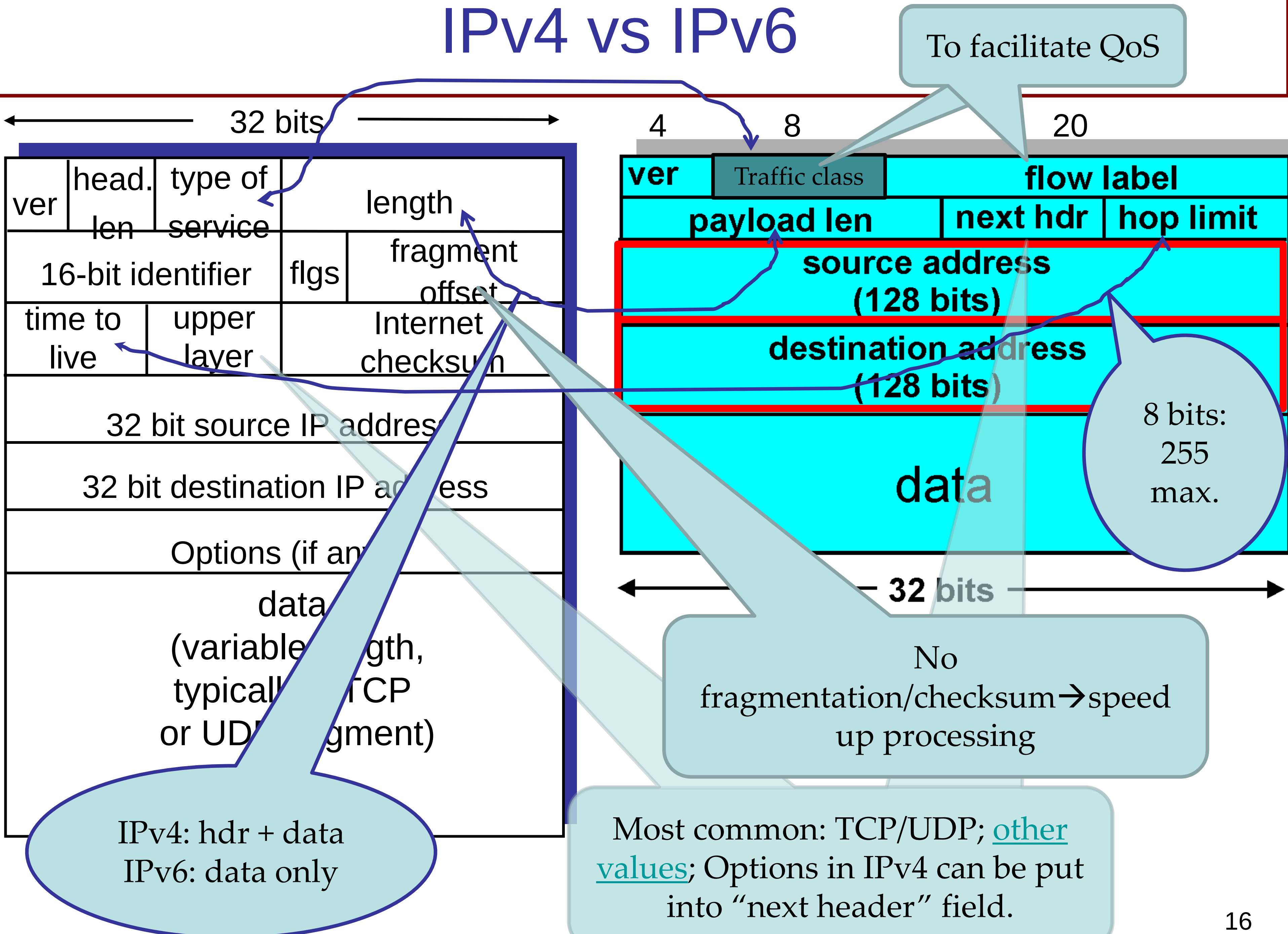
  IPv6 datagram format:
  - fixed-length 40 byte header (20-byte header for IPv4, assuming no options).
  - no fragmentation allowed

IPv6 Header (Cont)

Flow Label: identify datagrams in same "flow."
(concept of "flow" not well defined).
Next header: identify upper layer protocol for data



IPv4 vs IPv6



32 bits
ver
head. len
type of service
length
16-bit identifier
flgs
fragment offset
time to live
upper layer
Internet checksum
32 bit source IP address
32 bit destination IP address
Options (if any)
data (variable length, typically a TCP or UDP segment)
4
8
20
ver
Traffic class
flow label
payload len
next hdr
hop limit
source address (128 bits)
destination address (128 bits)
data
32 bits
To facilitate QoS
8 bits: 255 max.
No fragmentation/checksum→speed up processing
Most common: TCP/UDP; other values; Options in IPv4 can be put into "next header" field.
IPv4: hdr + data
IPv6: data only

Protocol Numbers

Last Updated
2017-10-13
Available Formats
XML HTML Plain text

Registry included below

- Assigned Internet Protocol Numbers

Assigned Internet Protocol Numbers

Registration Procedure(s)
IESG Approval or Standards Action
Reference
[RFC5237][RFC7045]
Note
In the Internet Protocol version 4 (IPv4) [RFC791] there is a field called "Protocol" to identify the next level protocol. This is an 8 bit field. In Internet Protocol version 6 (IPv6) [RFC8200], this field is called the "Next Header" field.
Note
Values that are also IPv6 Extension Header Types should be listed in the IPv6 Extension Header Types registry at [IANA registry ipv6-parameters].

Available Formats
CSV

Decimal	Keyword	Protocol	IPv6 Extension Header	Reference
0	HOPOPT	IPv6 Hop-by-Hop Option	Y	[RFC8200]
1	ICMP	Internet Control Message		[RFC792]
2	IGMP	Internet Group Management		[RFC1112]
3	GGP	Gateway-to-Gateway		[RFC823]
4	IPv4	IPv4 encapsulation		[RFC2003]
5	ST	Stream		[RFC1190][RFC1819]
6	TCP	Transmission Control		[RFC793]
7	CBT	CBT		[Tony_Ballardie]
8	EGP	Exterior Gateway Protocol		[RFC888][David_Mills]
9	IGP	any private interior gateway (used by Cisco for their IGRP)		[Internet_Assigned_Numbers_Authority]
10	BBN-RCC-MON	BBN RCC Monitoring		[Steve_Chipman]
11	NVP-II	Network Voice Protocol		[RFC741][Steve_Casner]
12	PUP	PUP		[Boggs, D., J. Shoch, E. Taft, and R. Metcalfe, "PUP: An In Research Center, CSL-79-10, July 1979; also in IEEE Tran Number 4, April 1980.][XEROX]
13	ARGUS (deprecated)	ARGUS		[Robert_W_Scheifler]
14	EMCON	EMCON		[<mystery contact>]
15	XNET	Cross Net Debugger		[Haverty, J., "XNET Formats for Internet Protocol Version 4", IEN 158, October 1980.][Jack_Haverty]
16	CHAOS	Chaos		[J_Noel_Chiappa]
17	UDP	User Datagram		[RFC768][Jon_Postel]
18	MUX	Multiplexing		[Cohen, D. and J. Postel, "Multiplexing Protocol", IEN 90, USC/Information Sciences Institute, May 1979.][Jon_Postel]
19	DCN-MEAS	DCN Measurement Subsystems		[David_Mills]

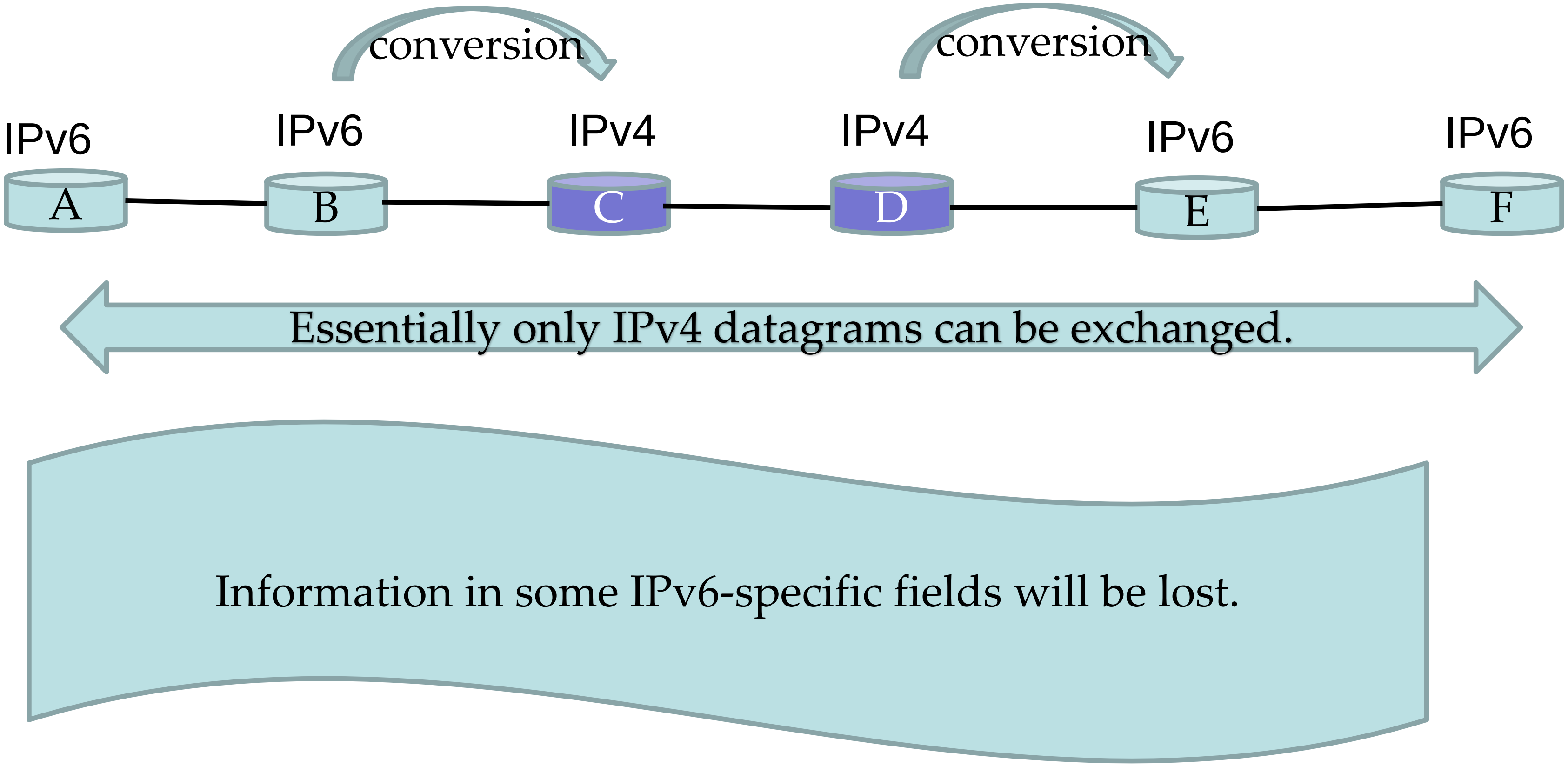
40	IL	IL Transport Protocol		[Dave_Presotto]
41	IPv6	IPv6 encapsulation		[RFC2473]
42	SDRP	Source Demand Routing Protocol		[Deborah_Estrin]
43	IPv6-Route	Routing Header for IPv6	Y	[Steve_Deering]
44	IPv6-Frag	Fragment Header for IPv6	Y	[Steve_Deering]
45	IDRP	Inter-Domain Routing Protocol		[Sue_Hares]
46	RSVP	Reservation Protocol		[RFC2205][RFC3209][Bob_Braden]
47	GRE	Generic Routing Encapsulation		[RFC2784][Tony_Li]

Transition From IPv4 To IPv6

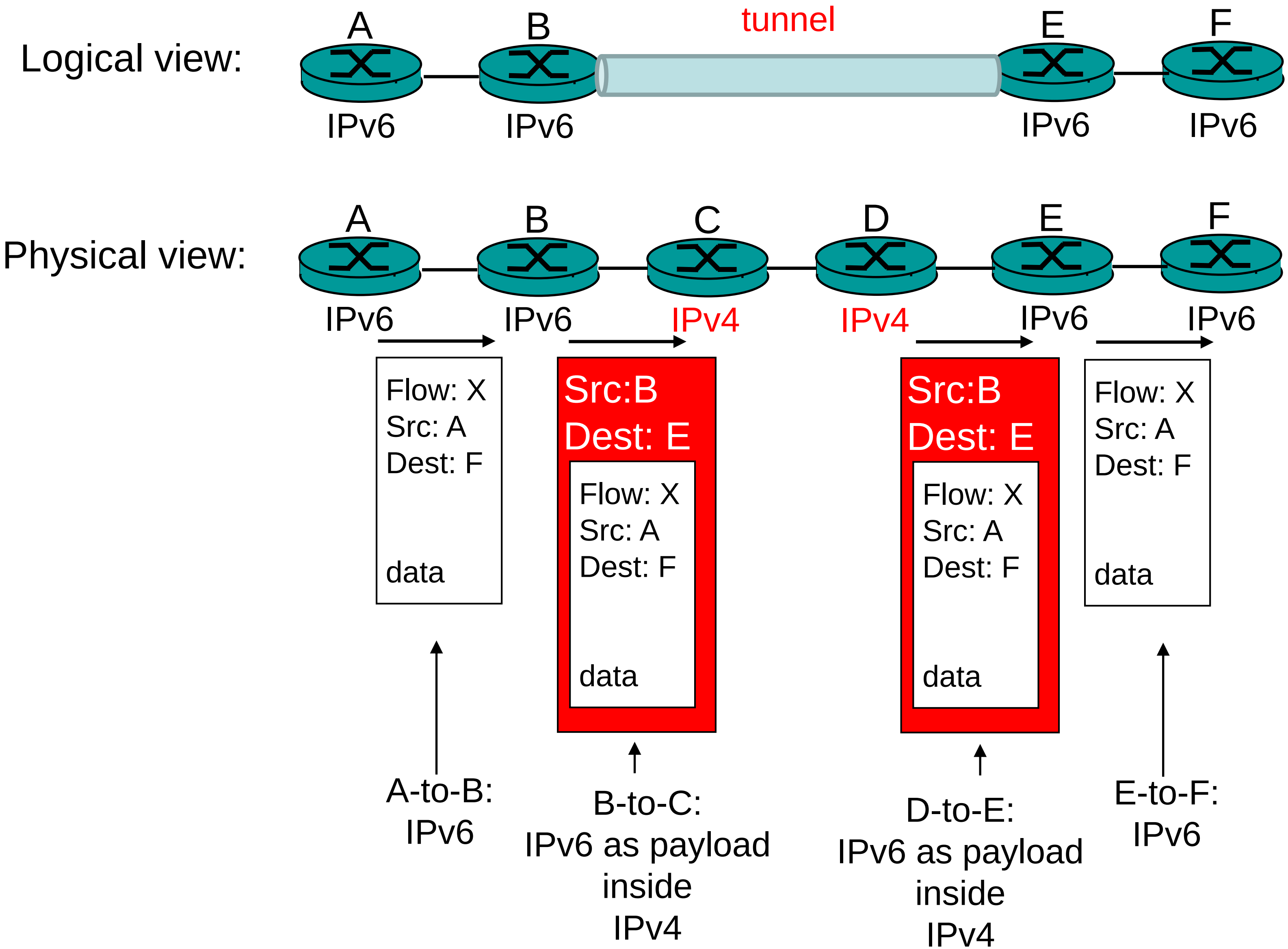
- How will the network operate with mixed IPv4 and IPv6 routers?
  - New devices are IPv4 and IPv6 capable; but old ones only supports IPv4.
  - Not all routers can be upgraded.
- Option one - “flag days” (will not work!)

Dual-stack approach

- All IPv6 nodes also have a complete IPv4 implementation.



Tunneling


Logical view:
A
B
tunnel
E
F
IPv6
IPv6
IPv6
IPv6
Physical view:
A
B
C
D
E
F
IPv6
IPv6
IPv4
IPv4
IPv6
IPv6
Flow: X
Src: A
Dest: F
data
Src:B
Dest: E
Flow: X
Src: A
Dest: F
data
Src:B
Dest: E
Flow: X
Src: A
Dest: F
data
Flow: X
Src: A
Dest: F
data
A-to-B:
IPv6
B-to-C:
IPv6 as payload
inside
IPv4
D-to-E:
IPv6 as payload
inside
IPv4
E-to-F:
IPv6

```
3.5.  IPv4 Header Construction

   When encapsulating an IPv6 packet in an IPv4 datagram, the IPv4
   header fields are set as follows:

      Version:

         4

      IP Header Length in 32-bit words:

         5 (There are no IPv4 options in the encapsulating header.)

      Type of Service:

         0. [Note that work underway in the IETF is redefining the Type
         of Service byte and as a result future RFCs might define a
         different behavior for the ToS byte when tunneling.]

      Total Length:

         Payload length from IPv6 header plus length of IPv6 and IPv4
         headers (i.e. a constant 60 bytes).

      Identification:

         Generated uniquely as for any IPv4 packet transmitted by the
         system.

      Flags:

         Set the Don't Fragment (DF) flag as specified in section 3.2.
         Set the More Fragments (MF) bit as necessary if fragmenting.

      Fragment offset:

         Set as necessary if fragmenting.

      Time to Live:

         Set in implementation-specific manner.

      Protocol:

         41 (Assigned payload type number for IPv6)
```

IPv6 Adoption

- CIDRised address, DHCP (Dynamic Host Configuration Protocol), and NAT has partially solved the IP address shortage problem in the short term.
- Adoption of IPv6 is slow.

Overview

- NAT
- IPv6
- Routing in the Internet
  - Hierarchical routing
  - RIP
  - OSPF
  - BGP

Overview

- NAT
- IPv6
- Routing in the Internet
  - Hierarchical routing
  - RIP
  - OSPF
  - BGP

Hierarchical Routing

Our routing study thus far - idealization

- all routers identical
- A “flat” network

… *not* true in practice

scale: with 200 million destinations:

- can’t store all dest’s in routing tables!
- routing table exchange would swamp links!

administrative autonomy (*not flat*)

- internet = network of networks
- each network admin may want to control routing in its own network

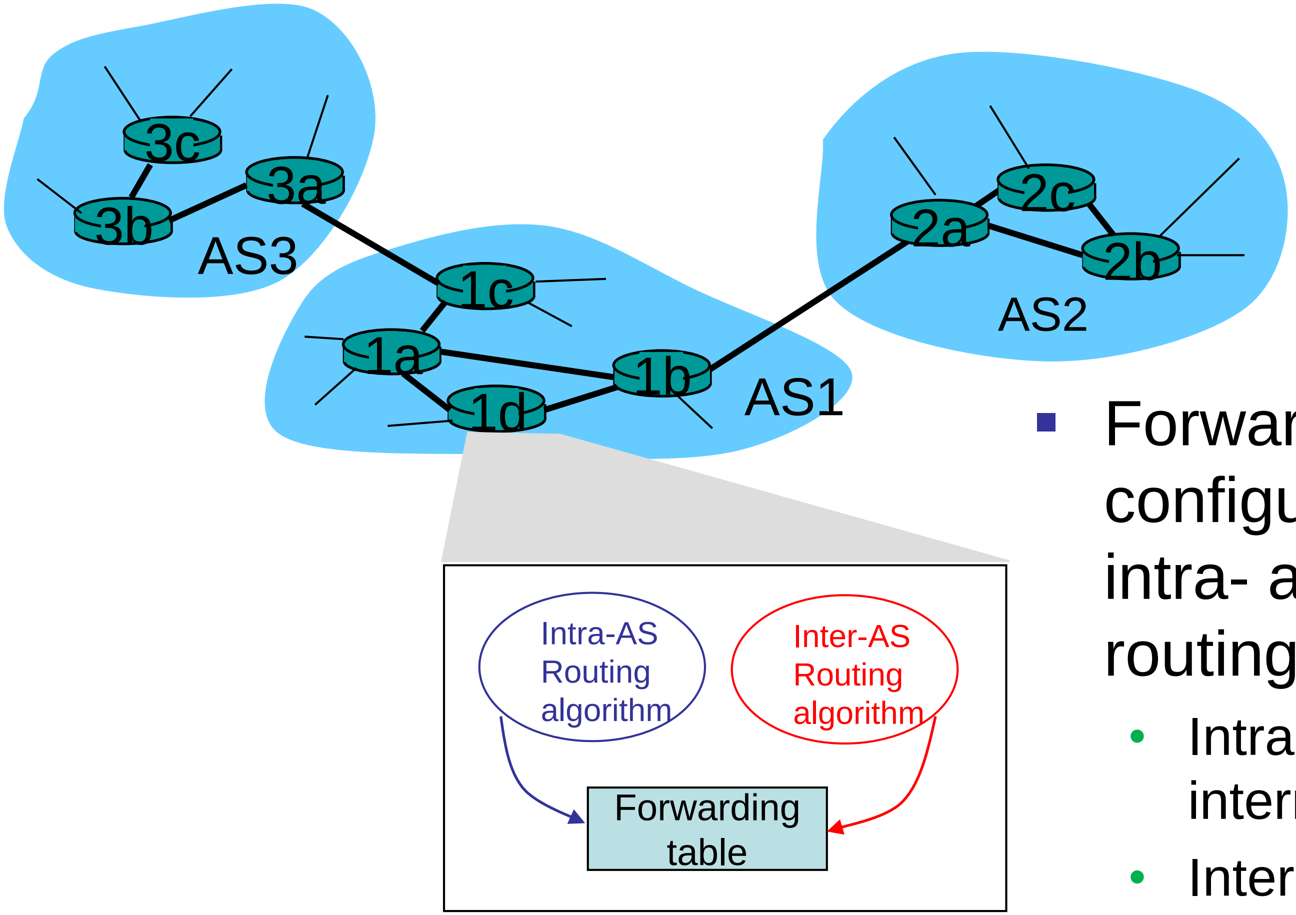
Hierarchical Routing

- aggregate routers into regions, "autonomous systems" (AS)
- routers in same AS run same routing protocol
  - "intra-AS" routing protocol
  - routers in different AS can run different intra-AS routing protocol

Gateway router

- Direct link to router in another AS

Interconnected ASes



- Forwarding table is configured by both intra- and inter-AS routing algorithm
  - Intra-AS sets entries for internal dests
  - Inter-AS & Intra-As sets entries for external dests

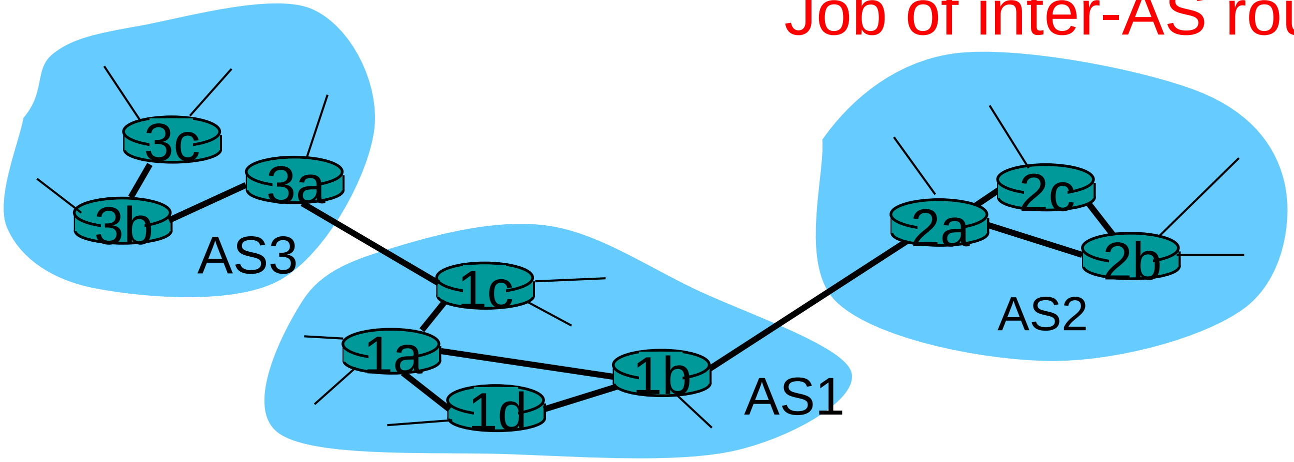
Inter-AS tasks

- Suppose router in AS1 receives datagram for which dest is outside of AS1
  - Router should forward packet towards one of the gateway routers, but which one?

AS1 needs:

1. to learn which dests are reachable through AS2 and which through AS3
2. to propagate this reachability info to all routers in AS1

Job of inter-AS routing!

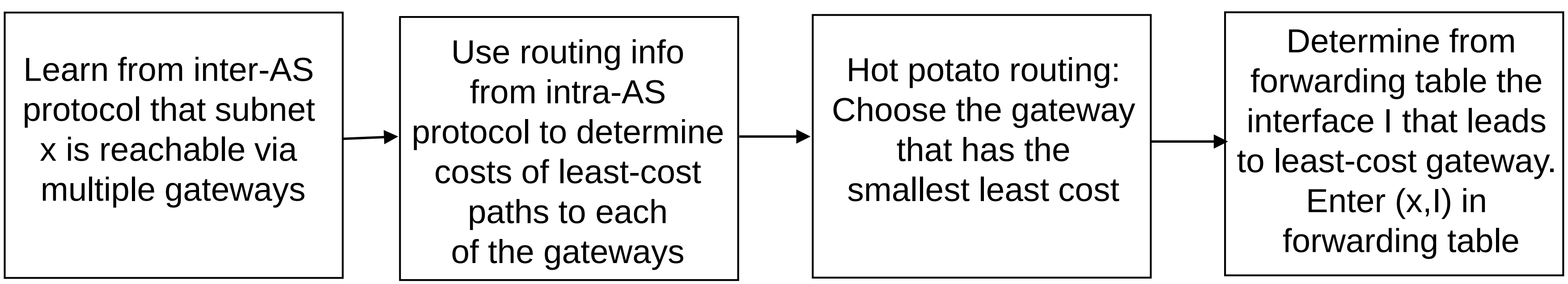


Example: Setting forwarding table in router 1d

- Suppose AS1 learns from the inter-AS protocol that subnet *x* is reachable from AS3 (gateway 1c) but not from AS2.
- Inter-AS protocol propagates reachability info to all internal routers.
- Router 1d determines from intra-AS routing info that its interface *I* is on the least cost path to 1c.
- Puts in forwarding table entry *(x,I)*.

Example: Choosing among multiple ASes

- Now suppose AS1 learns from the inter-AS protocol that subnet *x* is reachable from AS3 *and* from AS2.
- To configure forwarding table, router 1d must determine towards which gateway it should forward packets for dest x.
- This is also the job on inter-AS routing protocol!
- Hot potato routing: send packet towards closest of two routers.



Intra-AS Routing

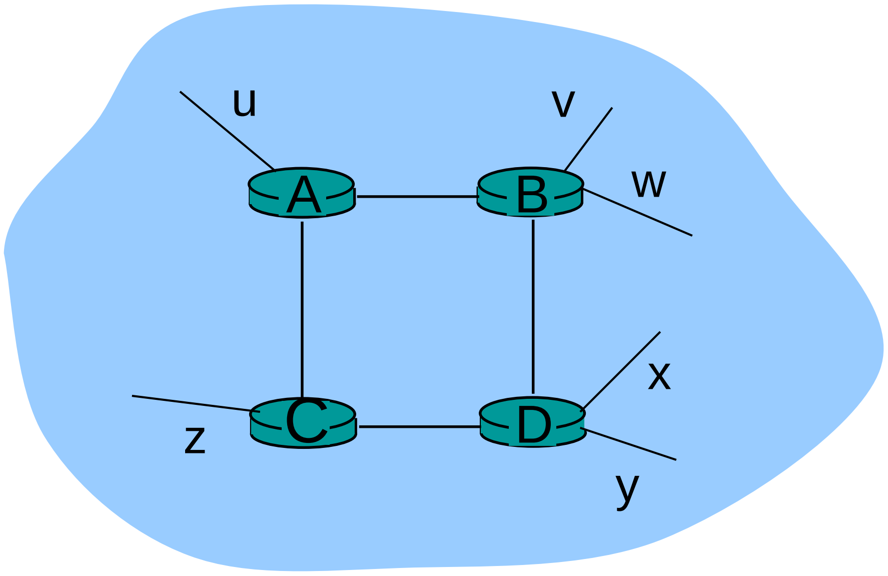
- Also known as Interior Gateway Protocols (IGP)
- Most common Intra-AS routing protocols:
  - RIP: Routing Information Protocol
  - OSPF: Open Shortest Path First
  - IGRP: Interior Gateway Routing Protocol (Cisco proprietary)

Overview

- NAT
- IPv6
- Routing in the Internet
  - Hierarchical routing
  - RIP
  - OSPF
  - BGP

RIP (Routing Information Protocol)

- Distance vector algorithm
- Included in BSD-UNIX Distribution in 1982
- Distance metric: # of hops (max = 15 hops)
  - # of hops: # of subnets traversed along the shortest path from src. router to dst. subnet (e.g., src. = A)

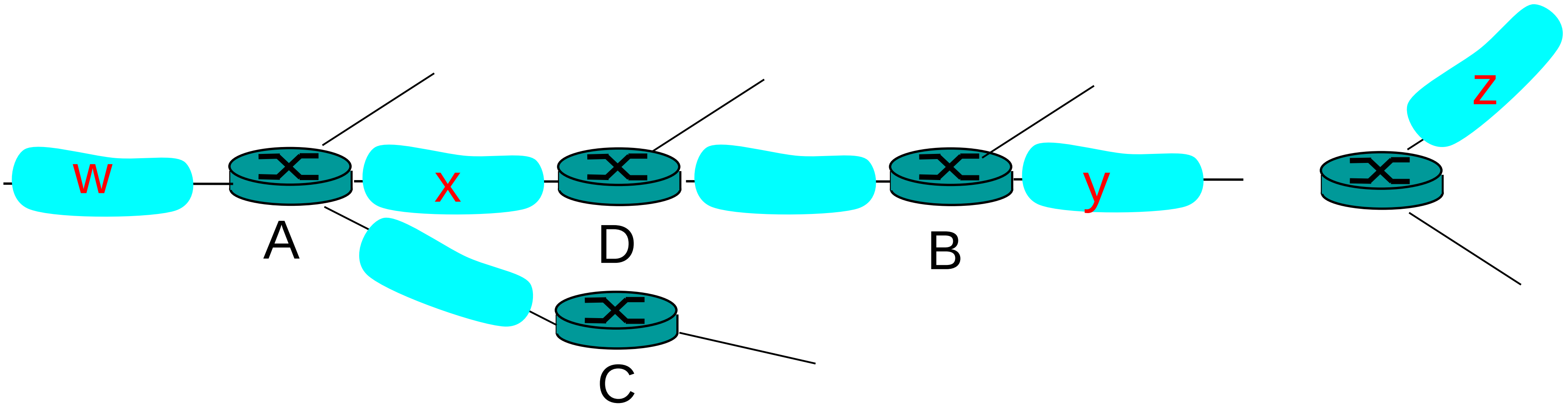


destination	hops
u	1
v	2
w	2
x	3
y	3
z	2

RIP advertisements

- Distance vectors: exchanged among neighbors every 30 sec via Response Message (also called advertisement)
- Each advertisement: list of up to 25 destination nets within AS

RIP: Example



Destination Network	Next Router	Num. of hops to dest.
w	A	2
y	B	2
z	B	7
x	--	1
….	….	….

Routing table in D

Each router that implements RIP is assumed to have a routing table. This table has one entry for every destination that is reachable throughout the system operating RIP. Each entry contains at least the following information:

- The IPv4 address of the destination.

- A metric, which represents the total cost of getting a datagram from the router to that destination. This metric is the sum of the costs associated with the networks that would be traversed to get to the destination.

- The IPv4 address of the next router along the path to the destination (i.e., the next hop). If the destination is on one of the directly-connected networks, this item is not needed.

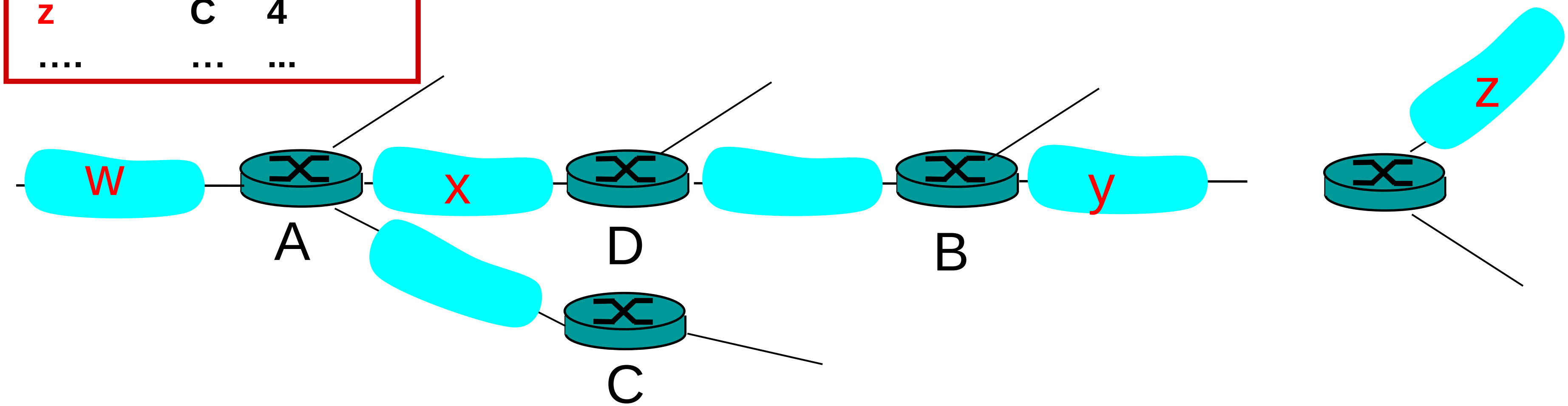
- A flag to indicate that information about the route has changed recently. This will be referred to as the "route change flag."

- Various timers associated with the route. See section 3.6 for more details on timers.

RIP: Example

Dest	Next	hops
w	-	-
x	-	-
z	C	4
….	…	...

Advertisement from A to D



Destination Network	Next Router	Num. of hops to dest.
w	A	2
y	B	2
z	~~B~~ A	~~7~~ 5
x	--	1
….	….	….

Routing table in D

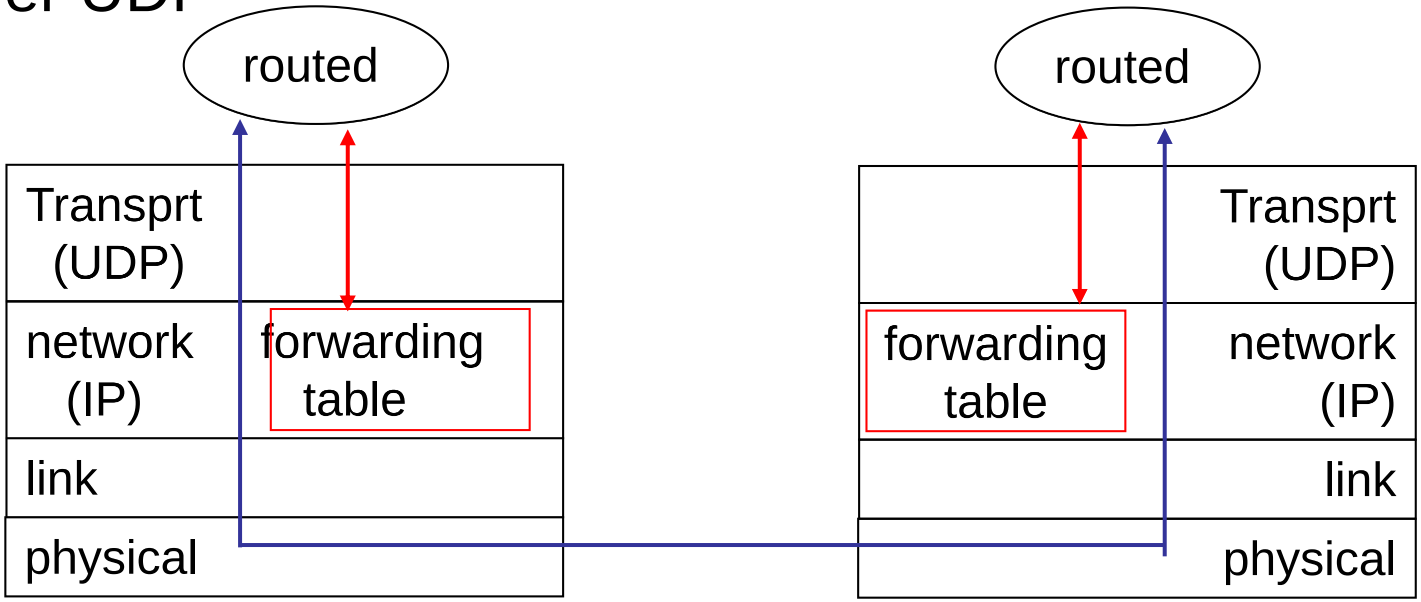
RIP: Link Failure and Recovery

If no advertisement heard after 180 sec --> neighbor/link declared dead

- routes via neighbor invalidated
- new advertisements sent to neighbors
- neighbors in turn send out new advertisements (if tables changed)
- link failure info quickly propagates to entire net
- poison reverse used to prevent ping-pong loops (infinite distance = 16 hops)

RIP Table processing

- RIP routing tables managed by **application-level** process called route-d (daemon)
- advertisements sent in UDP packets, periodically repeated
- RIP implemented as an app-layer protocol running over UDP



Others

- RIPng (RIP for next generation)
  - An extension of RIPv2 for IPv6.
- Zebra project
  - Open source implementation of routing protocols.
  - Quagga is a fork of GNU Zebra; http://www.nongnu.org/quagga/

Overview

- NAT
- IPv6
- Routing in the Internet
  - Hierarchical routing
  - RIP
  - OSPF
  - BGP

OSPF (Open Shortest Path First)

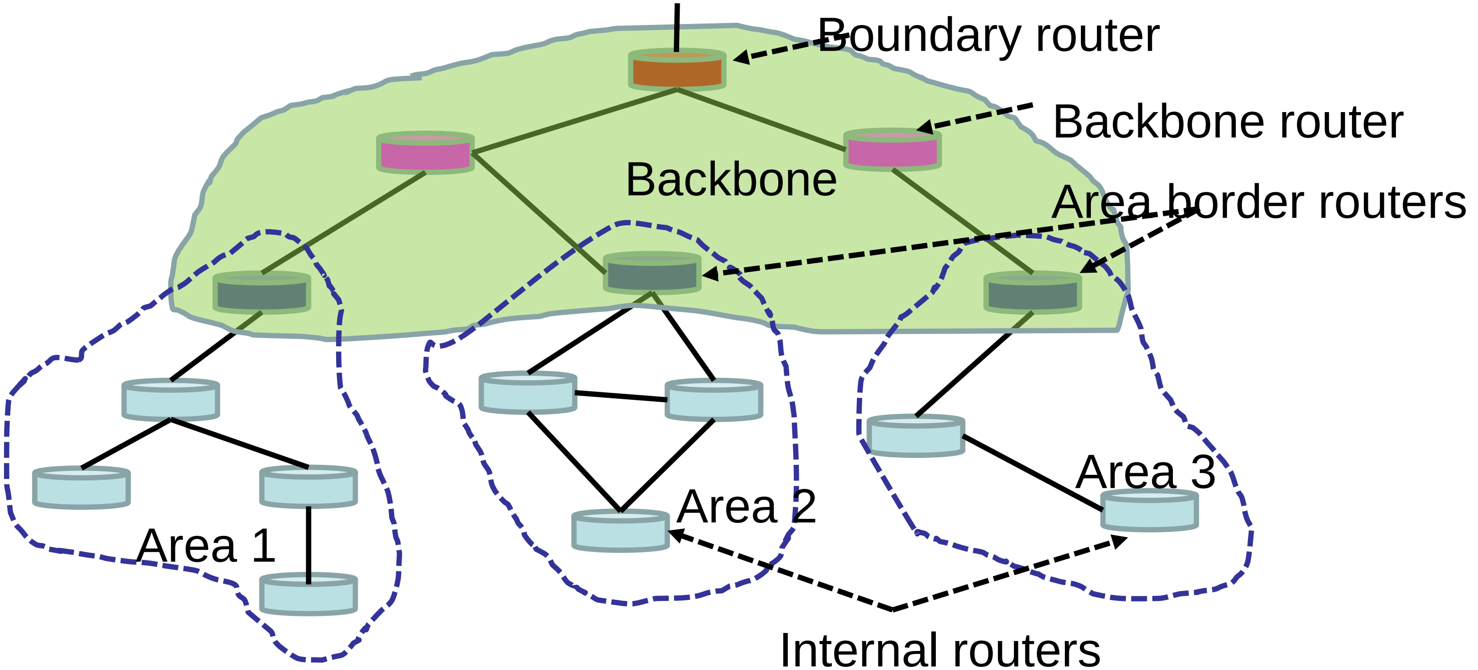
- “open”: publicly available
- Uses Link State algorithm
  - LS packet dissemination
  - Topology map at each node
  - Route computation using Dijkstra’s algorithm
  - Link costs configured by the network administrator
- OSPF advertisement carries one entry per neighbor router
- Advertisements disseminated to entire AS (via flooding)
  - Carried in OSPF messages directly over IP (rather than TCP or UDP
- OSPF and ISIS are typically deployed in upper-tier ISPs, whereas RIP is for lower-tier ISPs.

OSPF “advanced” features (not in RIP)

- Security: all OSPF messages authenticated (to prevent malicious intrusion)
- Multiple same-cost paths allowed (only one path in RIP)
- For each link, multiple cost metrics for different TOS (e.g., satellite link cost set “low” for best effort; high for real time)
- Integrated uni- and multicast support:
  - Multicast OSPF (MOSPF) uses same topology data base as OSPF
- Hierarchical OSPF in large domains.

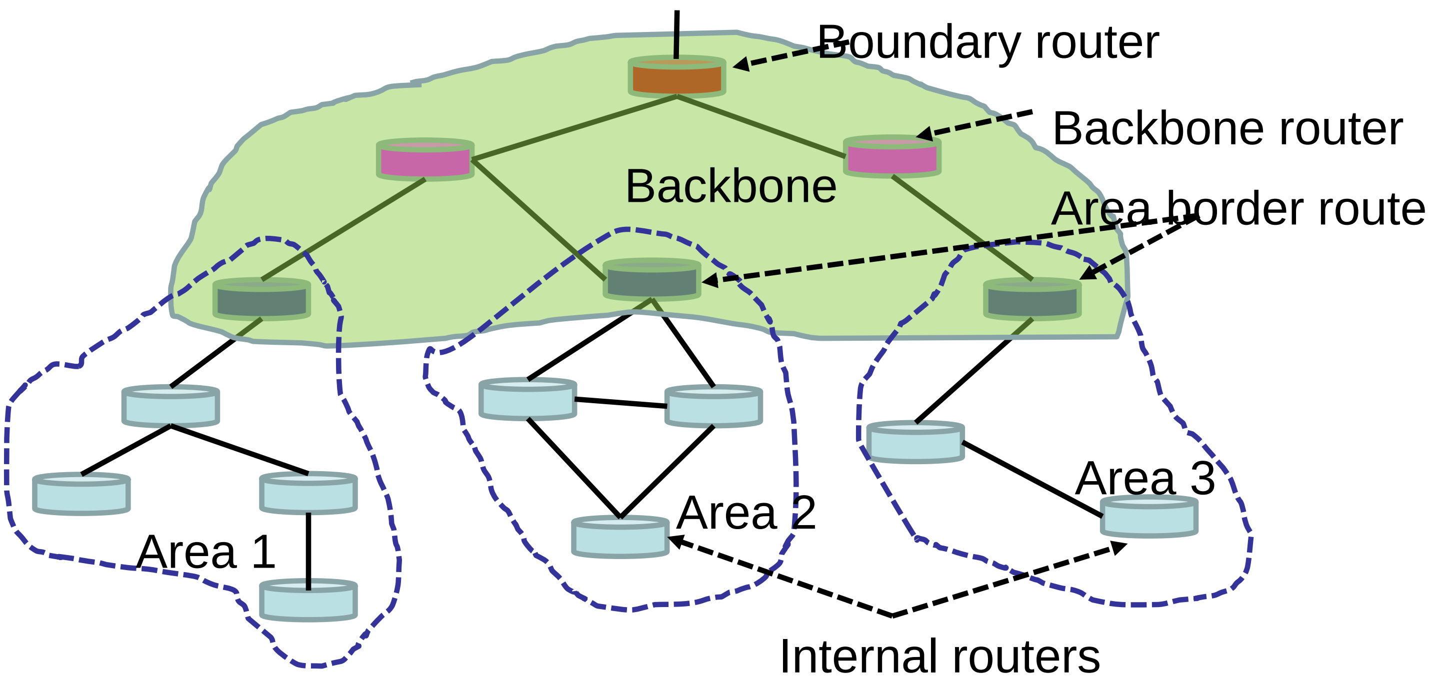
Hierarchical OSPF

- An OSPF autonomous system can be configured into areas;
- Each area runs its own OSPF algorithm;
- Each router broadcasts to all other routers in the same area;
- Only one area is the backbone area;
- Inter-area routing with the AS is done with the help of area border routers.



Hierarchical OSPF

- Two-level hierarchy: local area, backbone.
  - Link-state advertisements only in area
  - each node has detailed area topology;
- Area border routers: "summarize" distances to nets in own area, advertise to other Area Border routers.
- Backbone routers: run OSPF routing limited to backbone.
- Boundary routers: connect to other AS's.



Overview

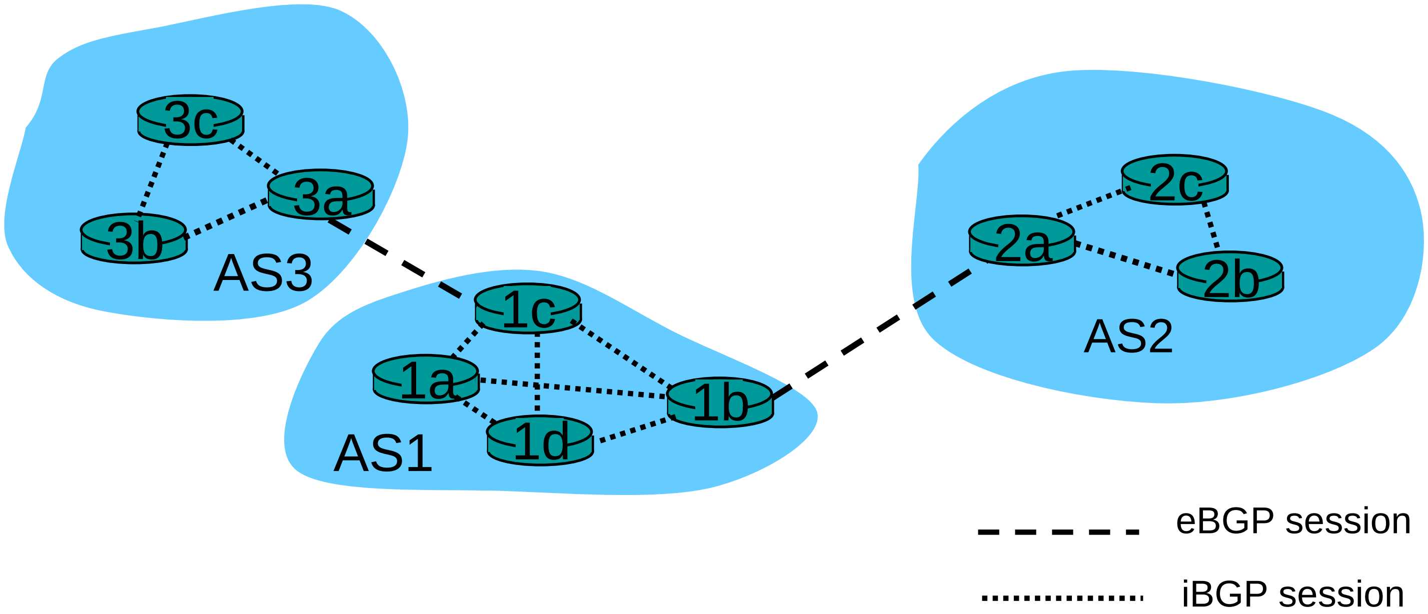
- NAT
- IPv6
- Routing in the Internet
  - Hierarchical routing
  - RIP
  - OSPF
  - BGP

Internet inter-AS routing: BGP

- BGP (Border Gateway Protocol): *the* de facto standard
- BGP provides each AS a means to:
  1. Obtain subnet reachability information from neighboring ASs.
  2. Propagate the reachability information to all routers internal to the AS.
  3. Determine “good” routes to subnets based on reachability information and policy.
- Allows a subnet to advertise its existence to rest of the Internet: *“I am here”*

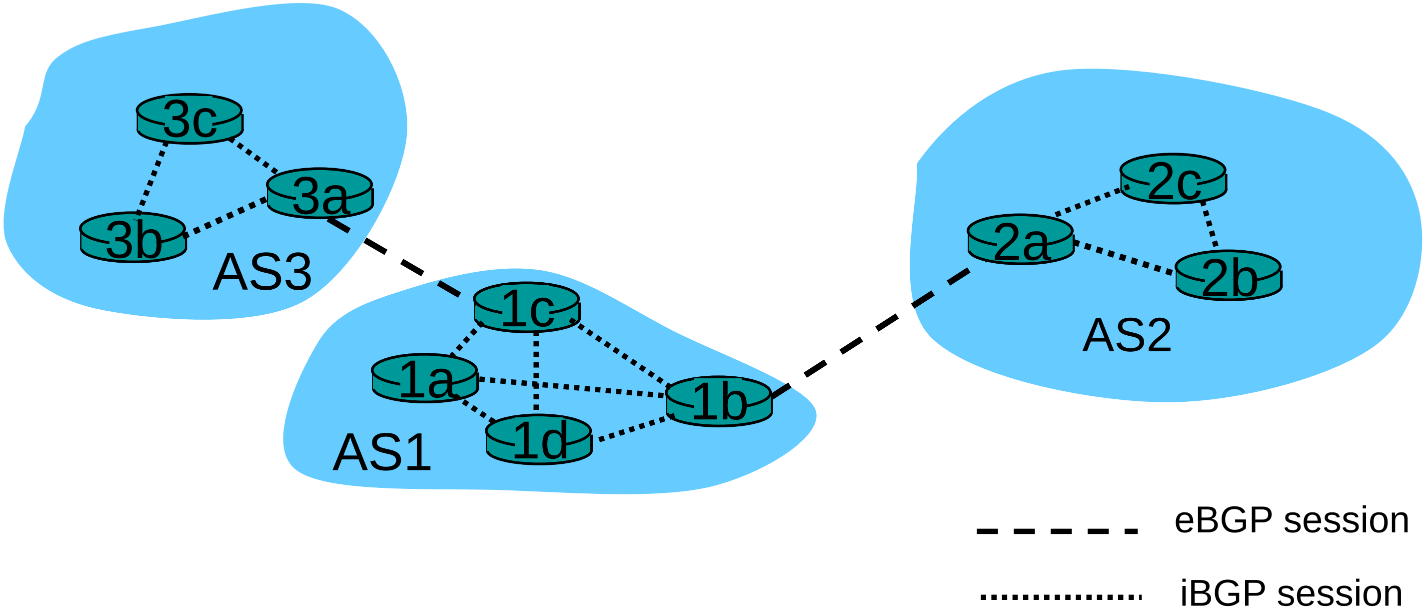
BGP basics

- Pairs of routers (BGP peers) exchange routing info over TCP conections: BGP sessions
- Note that BGP sessions do not correspond to physical links.
- When AS2 advertises a prefix to AS1, AS2 is *promising* it will forward any datagrams destined to that prefix towards the prefix.
    - AS2 can aggregate prefixes in its advertisement



Distributing reachability info

- With eBGP session between 3a and 1c, AS3 sends prefix reachability info to AS1.
- 1c can then use iBGP do distribute this new prefix reach info to all routers in AS1
- 1b can then re-advertise the new reach info to AS2 over the 1b-to-2a eBGP session
- When router learns about a new prefix, it creates an entry for the prefix in its forwarding table.



Path attributes & BGP routes

- When advertising a prefix, advert includes BGP attributes.
  - prefix + attributes = "route"
- Two important attributes:
  - AS-PATH: contains the ASs through which the advert for the prefix passed: AS 67 AS 17
  - NEXT-HOP: Indicates the specific internal-AS router to next-hop AS. (There may be multiple links from current AS to next-hop-AS.)
- When gateway router receives route advert, uses import policy to accept/decline.

An example

```
TIME: 06/01/19 00:00:00
TYPE: TABLE_DUMP_V2/IPV4_UNICAST
PREFIX: 1.44.24.0/21
SEQUENCE: 1236
FROM: 196.60.8.152 AS3491
ORIGINATED: 04/14/19 20:53:33
ORIGIN: IGP
ASPATH: 3491 7473 7474 4804
NEXT_HOP: 196.60.8.152
ATOMIC_AGGREGATE
AGGREGATOR: AS65367 10.194.27.154
COMMUNITY: 3491:2000 3491:2009 7473:10000 7473:12018 7473:12028 7473:12036 7473:
12047 7473:12156 7473:12168 7473:12178 7473:12187 7473:12208 7473:12218 7473:122
26 7473:12237 7473:20000 7473:21079 7473:22010 7473:31119 7473:31149 7473:31209
7473:32090 7473:33909 7473:33919 7473:42105 7474:100 7474:1202 7474:3006 7474:50
03
```

- Oregon Route Views Project
  - http://www.routeviews.org/routeviews/
  - **bgpdump** can be used to converting IPv4 MRT RIB data into text.
  - http://www.ris.ripe.net/source/bgpdump/libbgpdump-1.4.99.11.tar.gz (to download bgpdump)

Overview

- NAT
- IPv6
- Routing in the Internet
  - Hierarchical routing
  - RIP
  - OSPF
  - BGP

References

- [KR3] James F. Kurose, Keith W. Ross, *Computer networking: a top-down approach featuring the Internet*, 3rd edition.
- [PD5] Larry L. Peterson, Bruce S. Davie, *Computer networks: a systems approach*, 5th edition
- [TW5] Andrew S. Tanenbaum, David J. Wetherall, *Computer network*, 5th edition
- [LHBi]Y-D. Lin, R-H. Hwang, F. Baker, *Computer network: an open source approach*, International edition

Acknowledgements

- All slides are developed based on slides from the following two sources:
  - Dr DongSeong Kim's slides for COSC264, University of Canterbury;
  - Prof Aleksandar Kuzmanovic's lecture notes for CS340,Northwestern University, https://users.cs.northwestern.edu/~akuzma/classes/CS340-w05/lecture_notes.htm