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

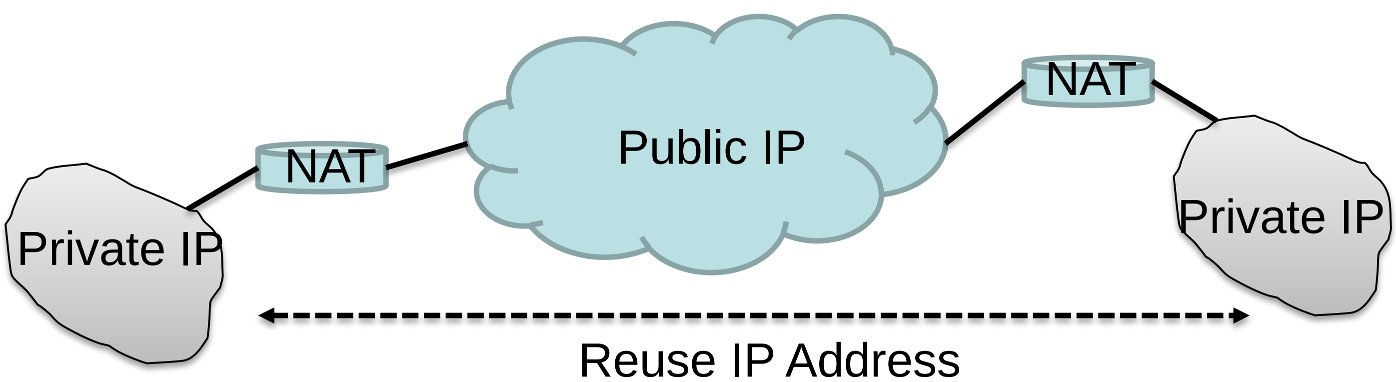
Back to the reality

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

NAT: Network Address Translation

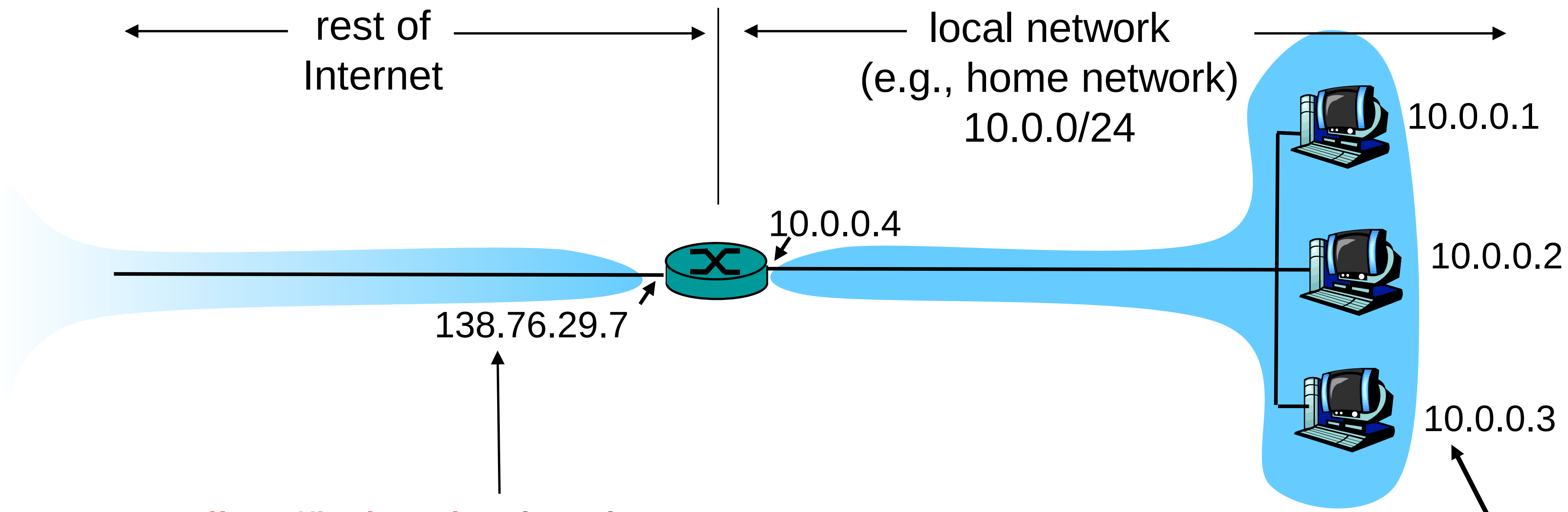
- Every IP-capable device needs an IP address. (IP address depletion problem)
- Short-term solution is CIDR (Classless InterDomain Routing).
- Another solution is to reuse IP address.

(long-term solution: 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.

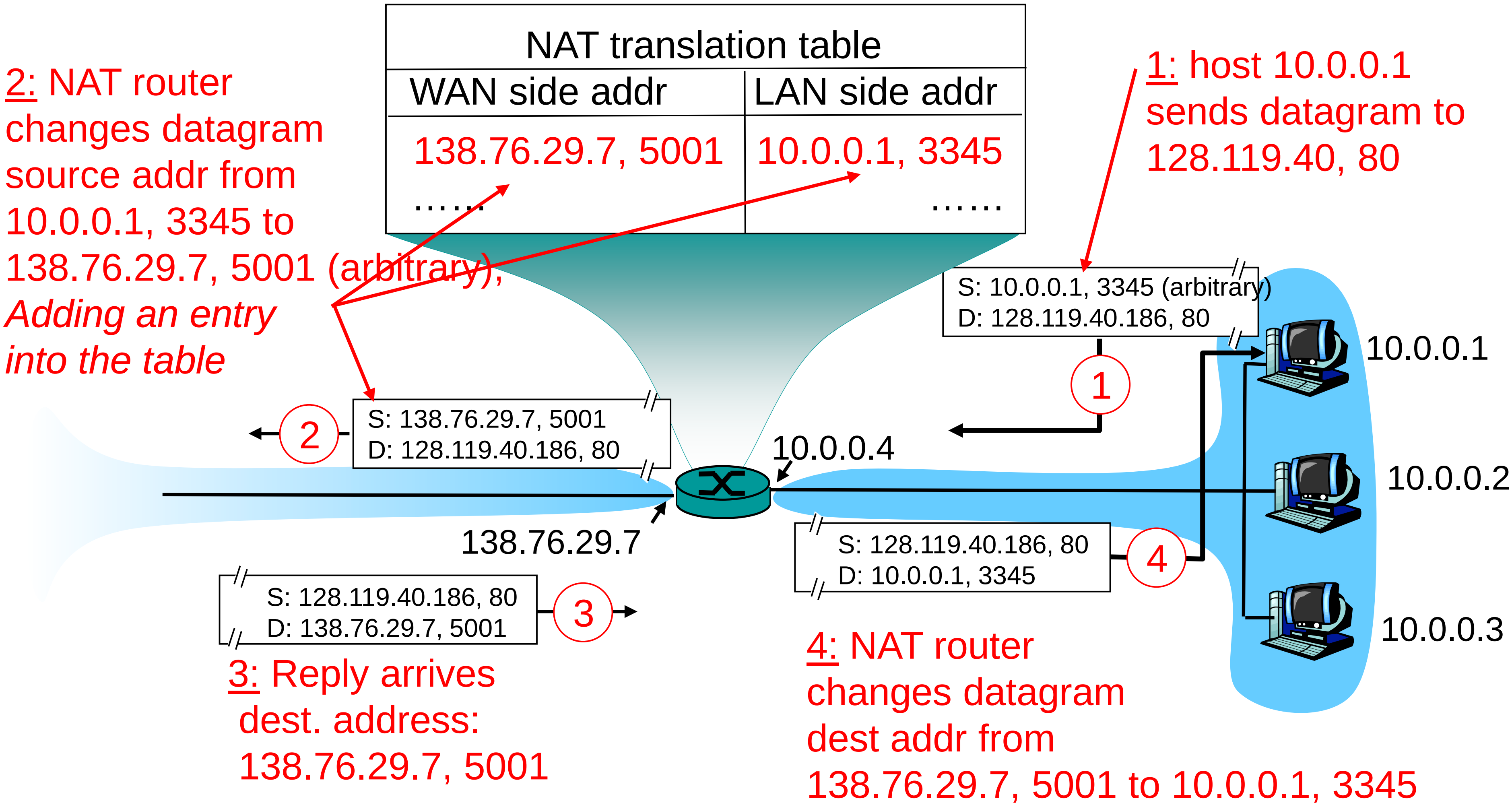
Hosts get their IP address through DHCP (Dynamic Host Configuration Protocol).

The trick

WAN side address		LAN side address	
138.76.29.7	xxxx (port #)	10.0.0.1	ssss
138.76.29.7	yyyy	10.0.0.2	cccc
138.76.29.7	zzzz	10.0.0.3	dddd

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

NAT: Network Address Translation



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
```

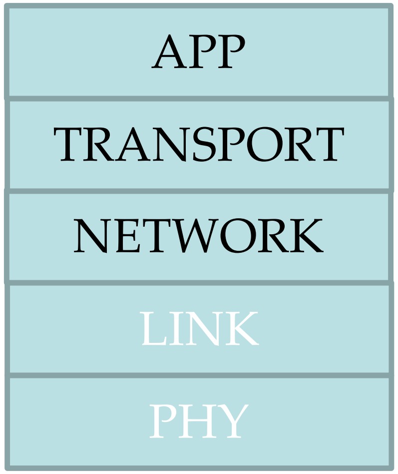
NAT: Network Address Translation

- 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.

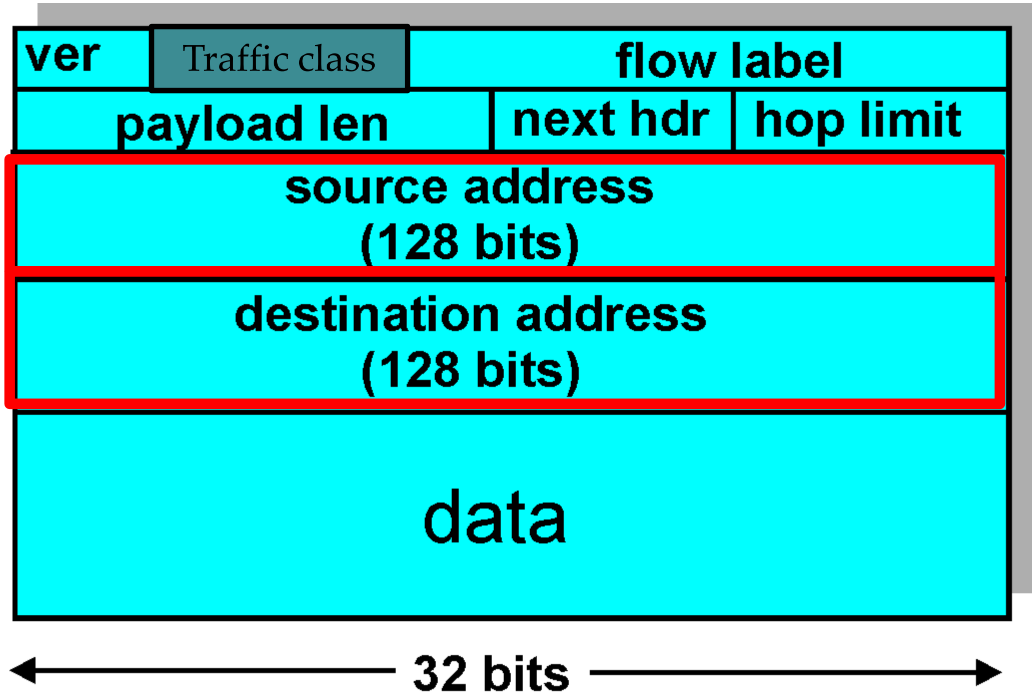
IPv6

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

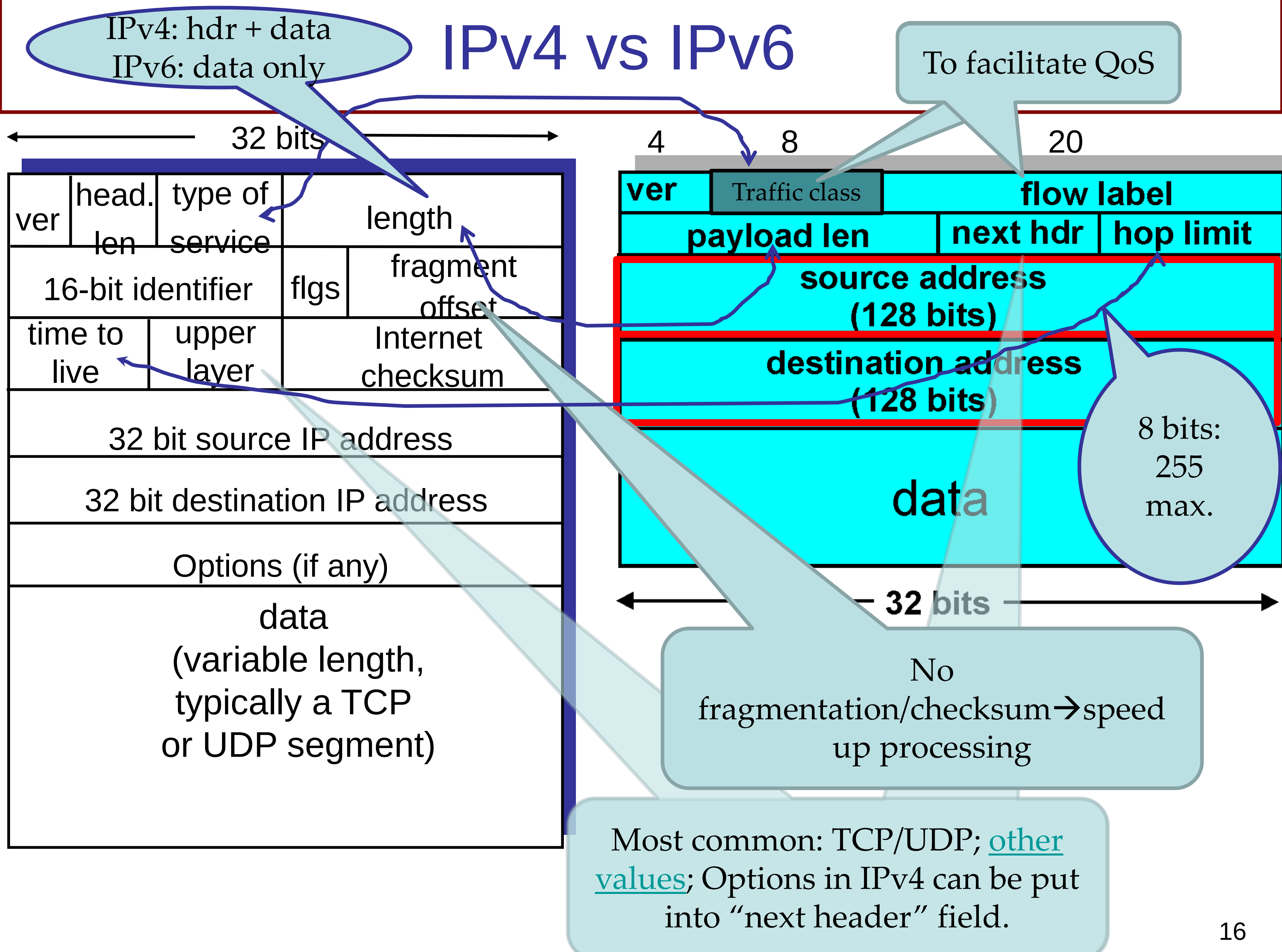
  IPv6 datagram format:
  - fixed-length 40 byte header (20-byte header for IPv4, assuming no options).
  - no fragmentation allowed

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



IPv4: hdr + data
IPv6: data only
To facilitate QoS
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
8 bits:
255
max.
32 bits
No
fragmentation/checksum→speed
up processing
Most common: TCP/UDP; other values; Options in IPv4 can be put into "next header" field.

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]

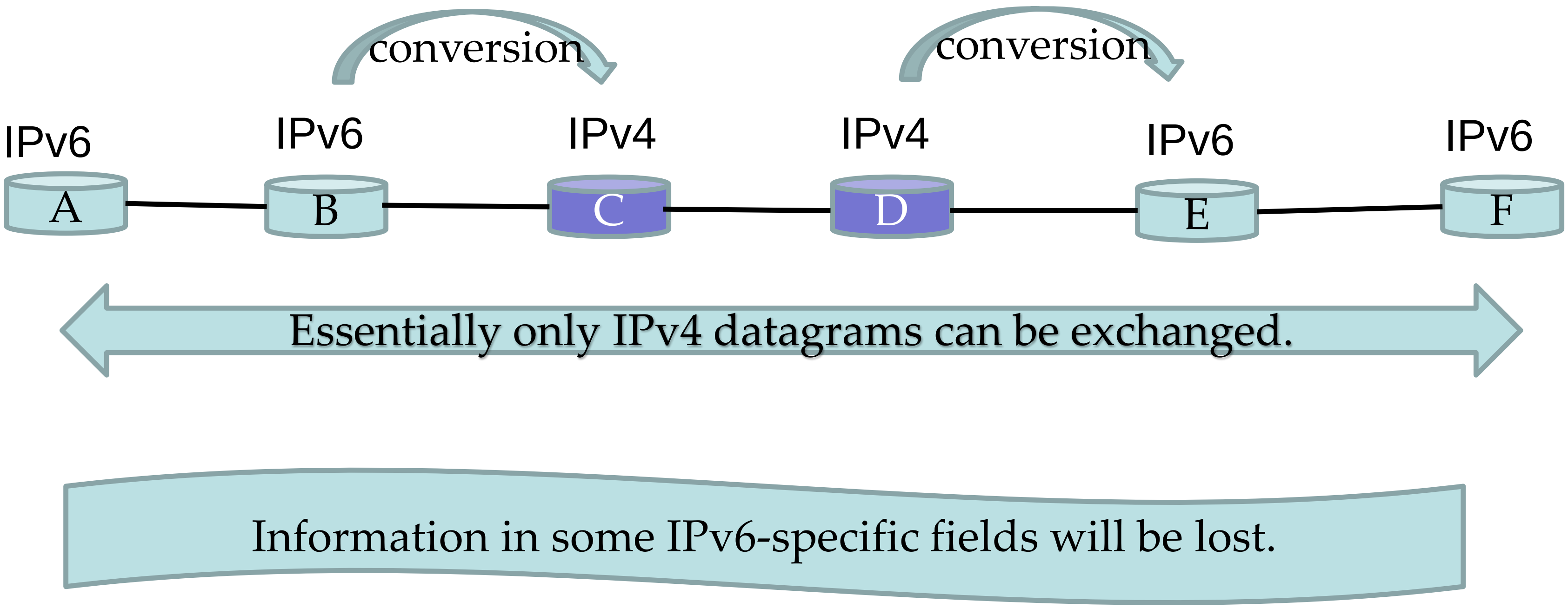
40	IL	IL Transport Protocol		[Dave_Presotto]
41	IPv6	IPv6 encapsulation		[RFC2473]
42	SDRP	Source Demand Routing Protocol		[Deborah_Estrin]
43	IPv6-Route	Routing Header for IPv6	Y	[Steve_Deering]
44	IPv6-Frag	Fragment Header for IPv6	Y	[Steve_Deering]
45	IDRP	Inter-Domain Routing Protocol		[Sue_Hares]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

Transition From IPv4 To IPv6

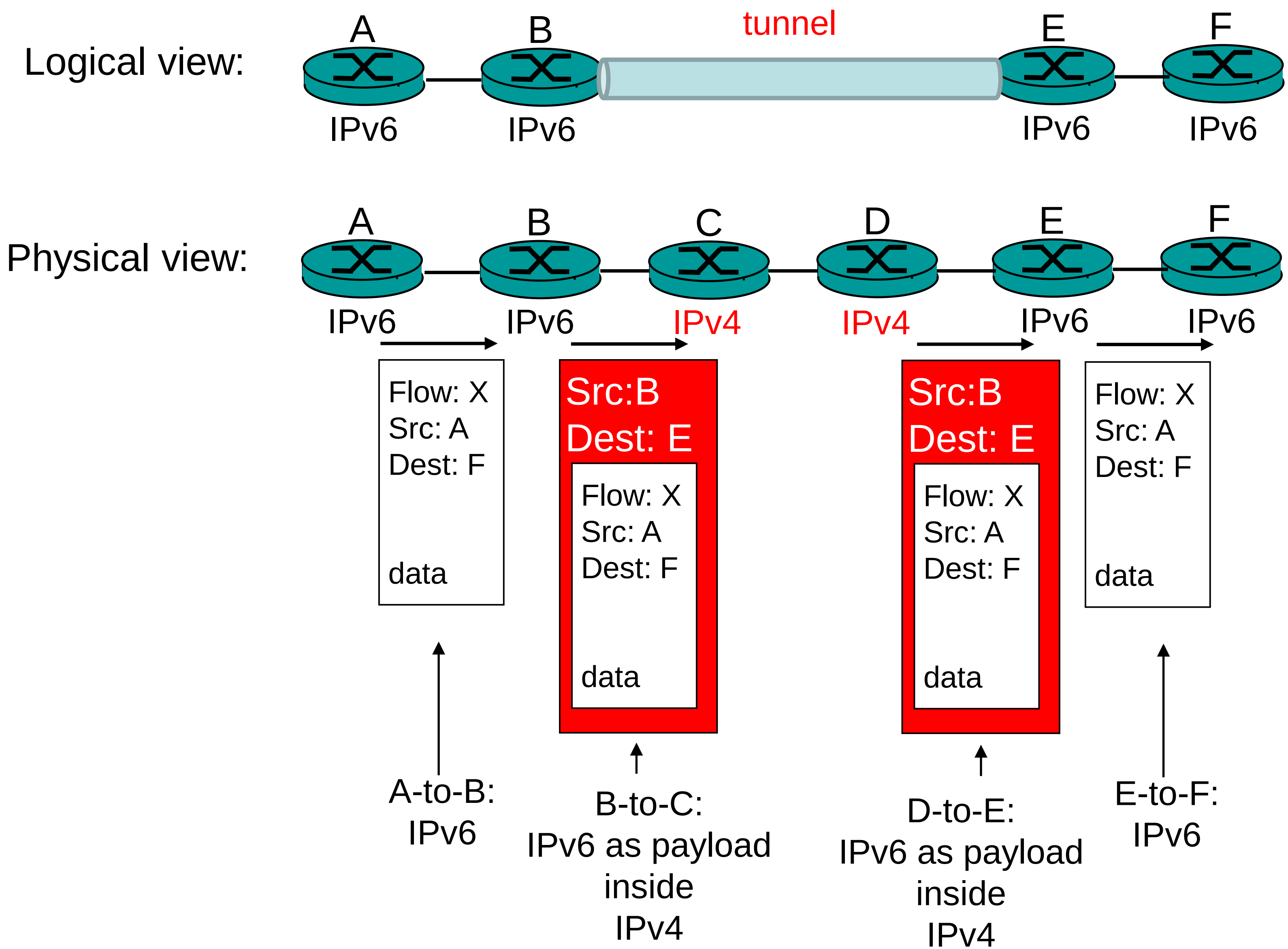
- How will the network operate with mixed IPv4 and IPv6 routers?
  - New devices are IPv4 and IPv6 capable; but old ones only supports IPv4.
  - Not all routers can be upgraded.
- Option one - “flag days” (will not work!)

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 - idealisation

- all routers identical
- A “flat” network

scale: with hundreds of millions of hosts:

- Storing routing information to each of them requires huge memory;
- Broadcasting LS updates; (no bandwidth left for data)
- DV algorithm iterates among such a large # of routers; (never converges)

administrative autonomy (*not flat*)

- An organization should be able to run and administer its own network.

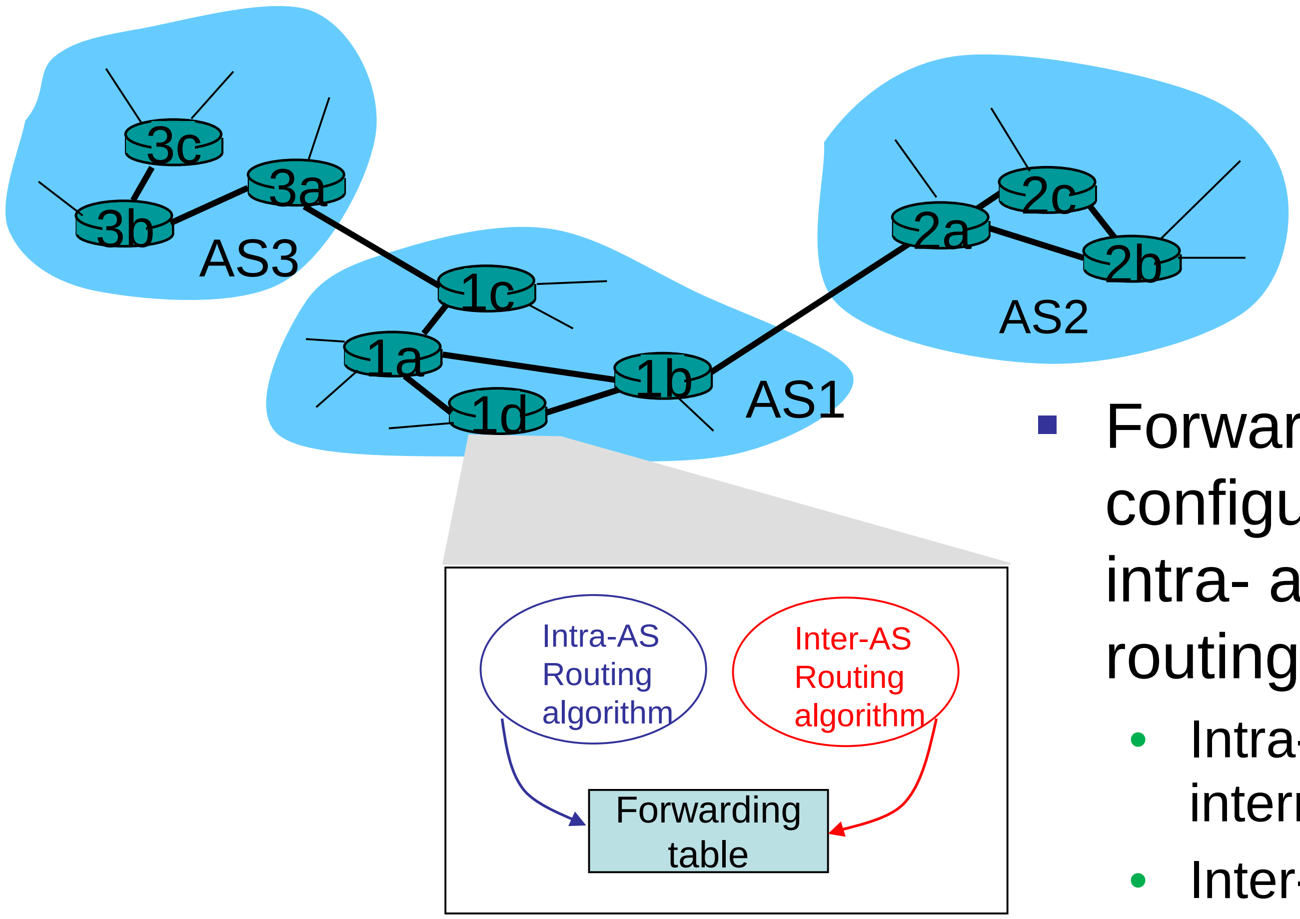
Hierarchical Routing

- aggregate routers into regions, "autonomous systems" (AS)
- routers within the same AS are under the same administrative control;
- routers in the same AS run the same routing protocol
  - "intra-AS" routing protocol
  - routers in different AS can run different intra-AS routing protocols

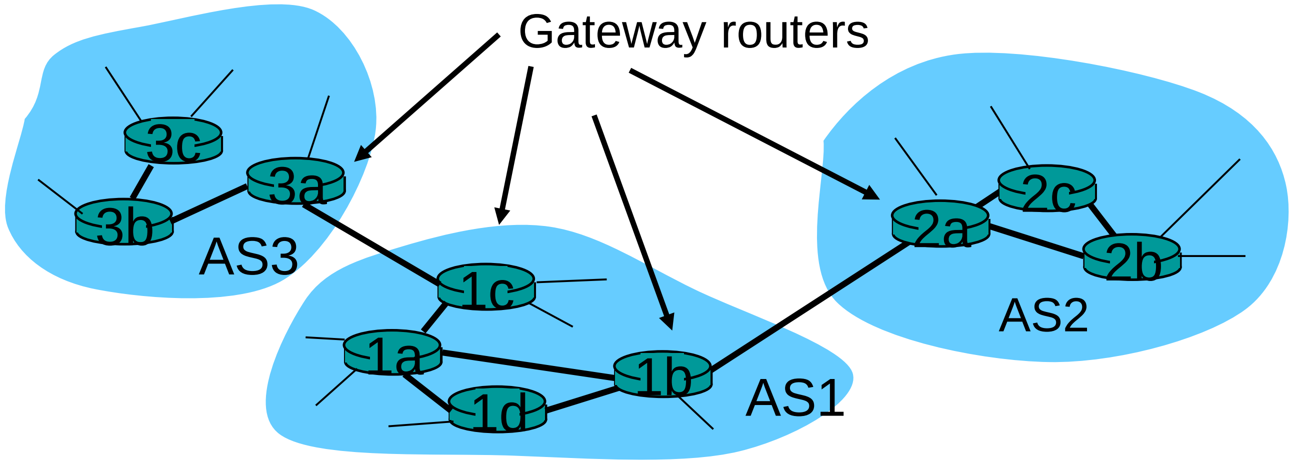
Gateway router

- Direct link to router in another AS

Interconnected ASes

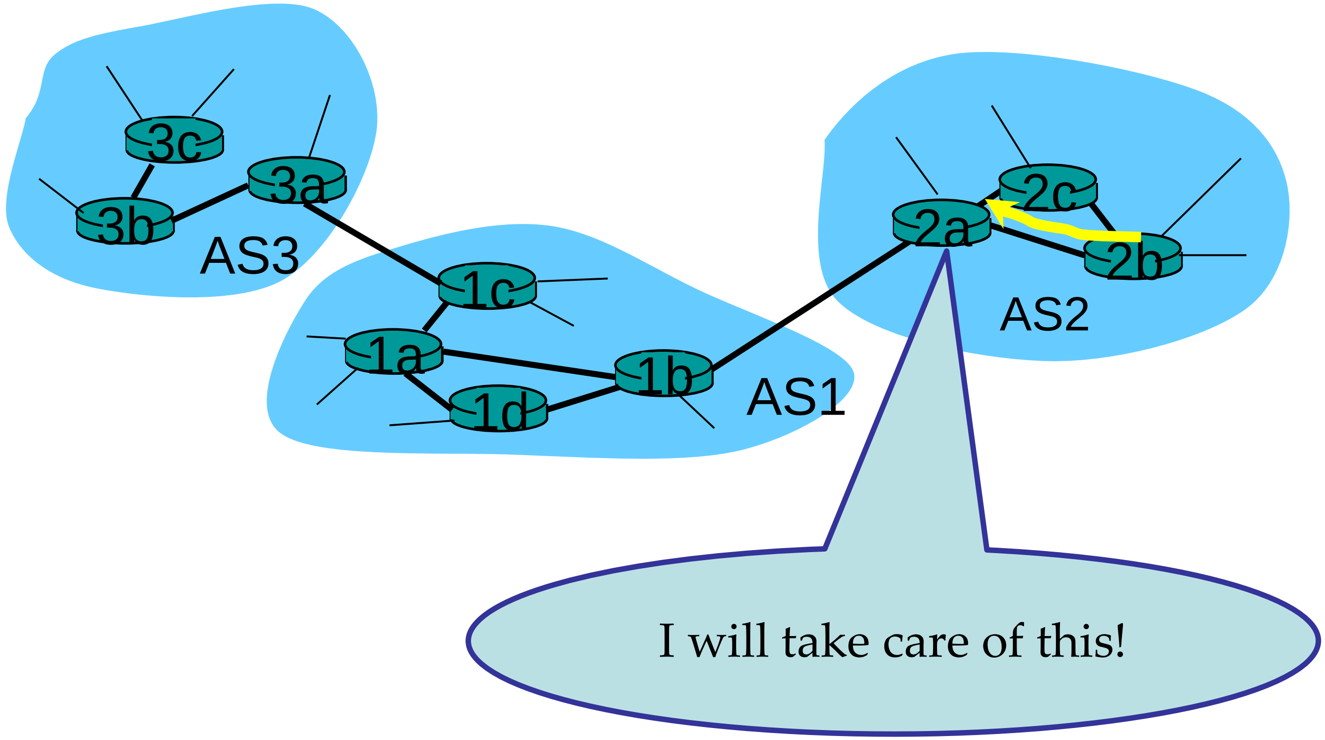


- Forwarding table is configured by both intra- and inter-AS routing algorithm
  - Intra-AS sets entries for internal dests
  - Inter-AS & *Intra-AS* sets entries for external dests



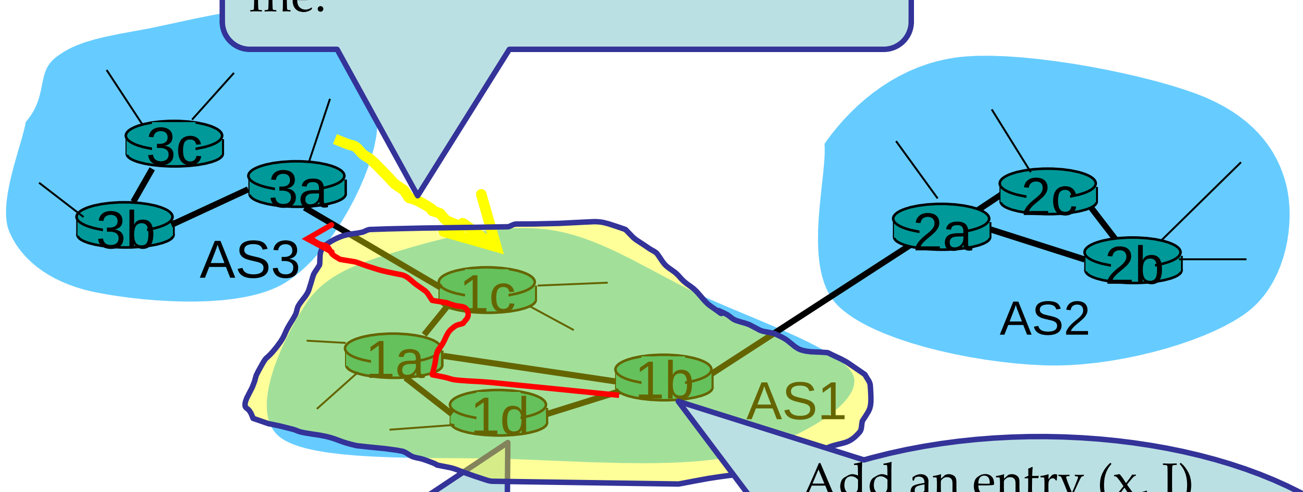
How does a router know how to route a packet to a destination that is outside the AS?

Case 1: Only one gateway router to only one other AS



Case 2: Two or more physical links to the other ASes

Subnet x is reachable from me.

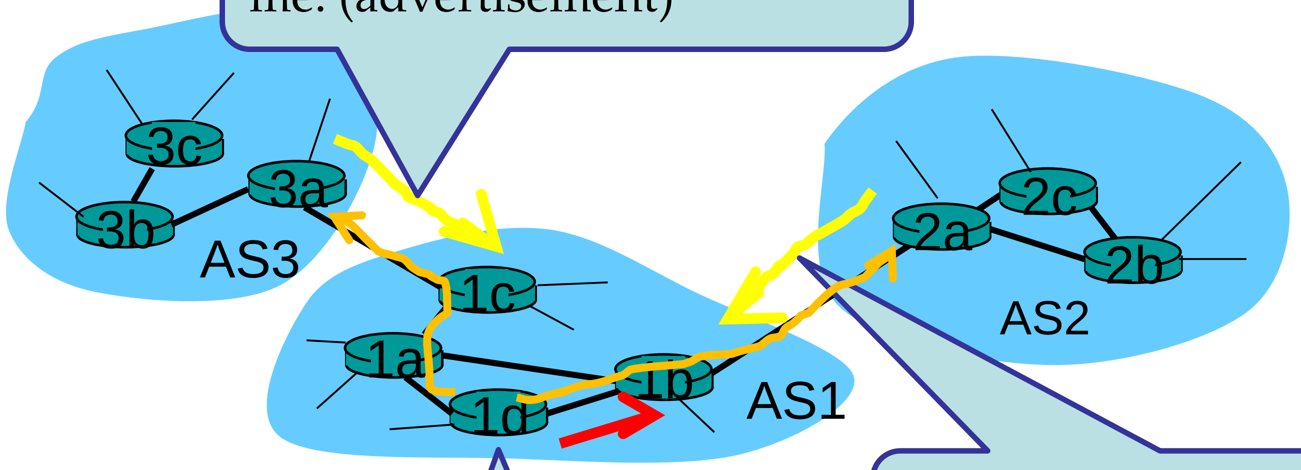


Typically a transit AS:
(1) obtain reachability information from neighbouring AS;
(2) propagating the reachability information to all routers inside the AS;

Add an entry (x, I) into its forwarding table; *intra-AS routing helps to decide which next hop to choose to get to the gateway;*

Case 2: Two or more physical links to the other ASes

Subnet x is reachable from me. (advertisement)



Subnet x is reachable from me.

Which way to go? **Hot-potato routing** – using intra-routing to choose the nearest gateway router

The main steps

Learn from inter-AS protocol that subnet x is reachable via multiple gateways

→

Use routing info from intra-AS protocol to determine costs of least-cost paths to each of the gateways

→

Hot potato routing: Choose the closest gateway

→

Determine from Intra-AS routing the interface I that leads to least-cost gateway. Enter (x,I) in forwarding table

Overview

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

Intra-AS Routing

- Also known as Interior Gateway Protocols (IGP)
- Most common Intra-AS routing protocols:
  - RIP: Routing Information Protocol, RFC 2453 – RIP v2
  - OSPF: *Open* Shortest Path First
  - *IS-IS: OSPF's closely related cousin*
  - IGRP: Interior Gateway Routing Protocol (Cisco proprietary)

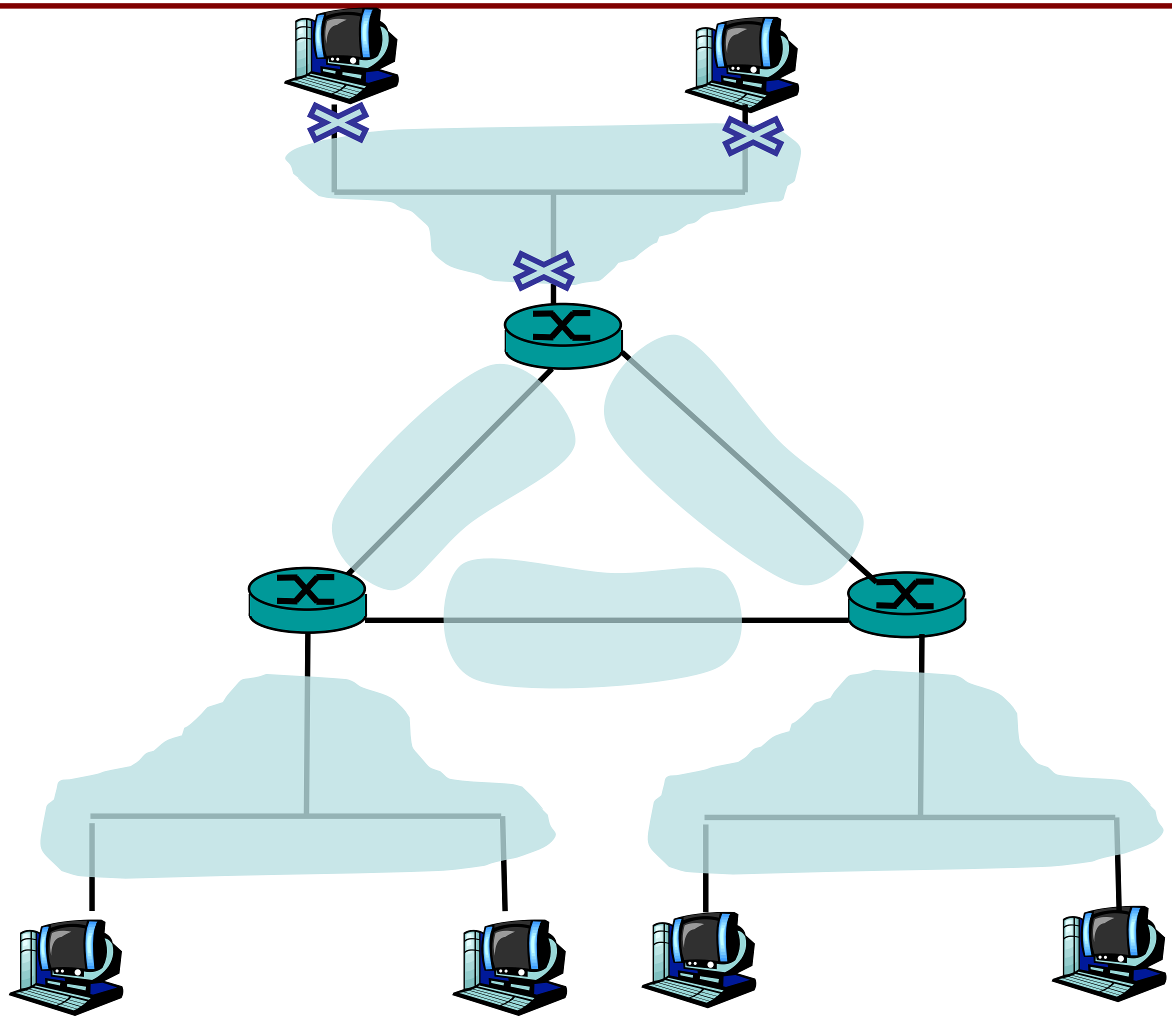
RIP (Routing Information Protocol)

- 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

How to define *subnets*?

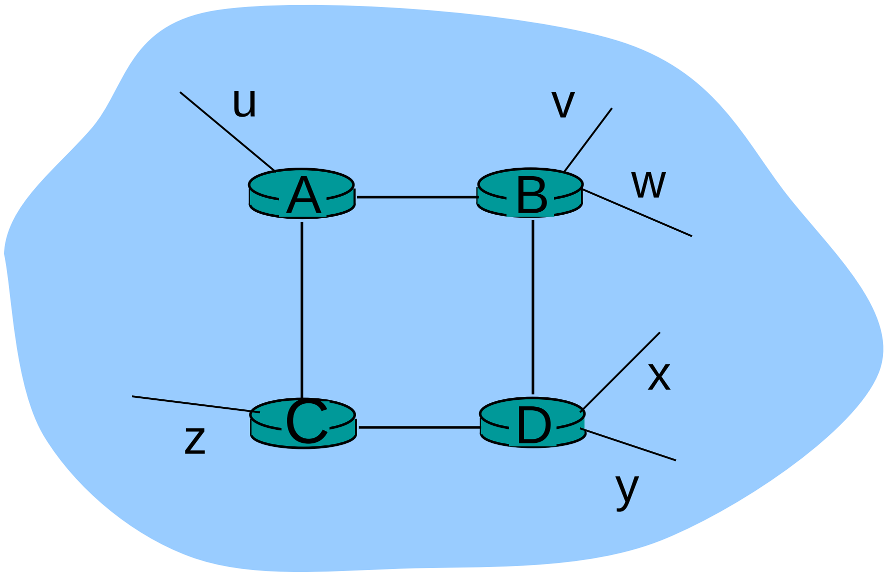
- *To determine subnets, detach each interface from its host or router, creating islands of isolated networks, with interfaces terminating the endpoints of the isolated networks. Each of these isolated network is called a subnet.*

6 subnets!



RIP (Routing Information Protocol)

- Distance metric: # of hops (max = 15 hops)
  - # of hops: # of subnets traversed along the shortest path from src. router to dst. subnet, including the dst. subnet; (e.g., src. = A)

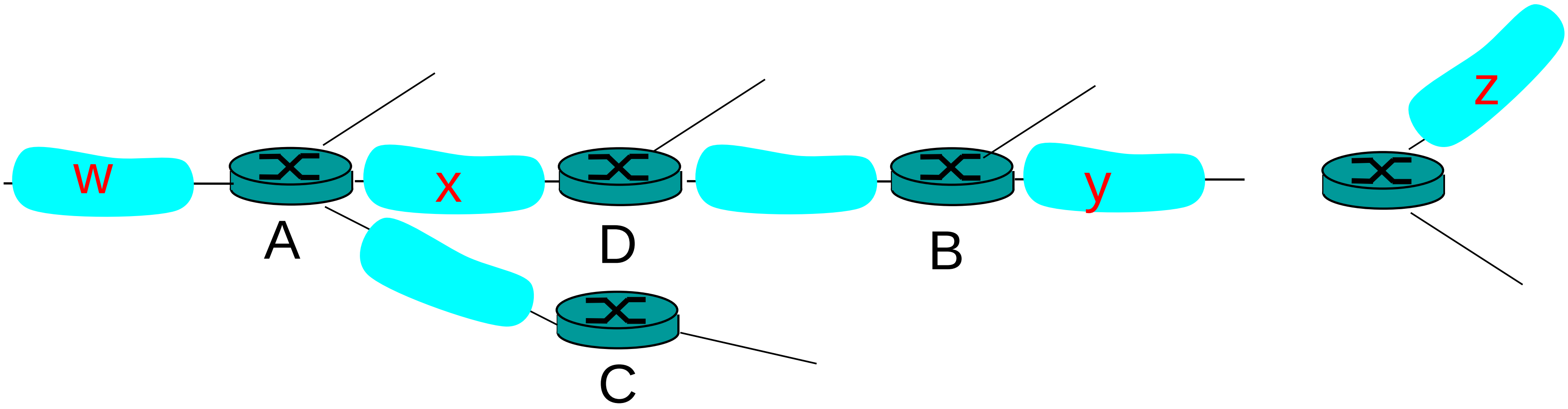


destination	hops
u	1
v	2
w	2
x	3
y	3
z	2

RIP advertisements

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

RIP: Example

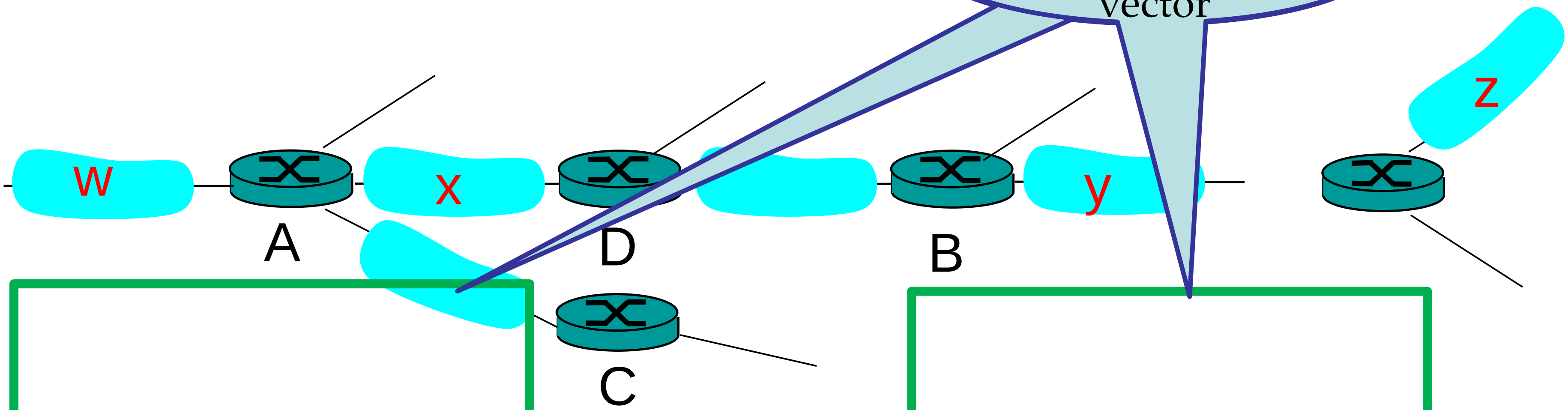


Destination Network	Next Router	Num. of hops to dest.
w	A	2
y	B	2
z	B	7
x	--	1
….	….	….

Routing table in D

RIP: Example

Combined to form a distance vector



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

Routing table in D

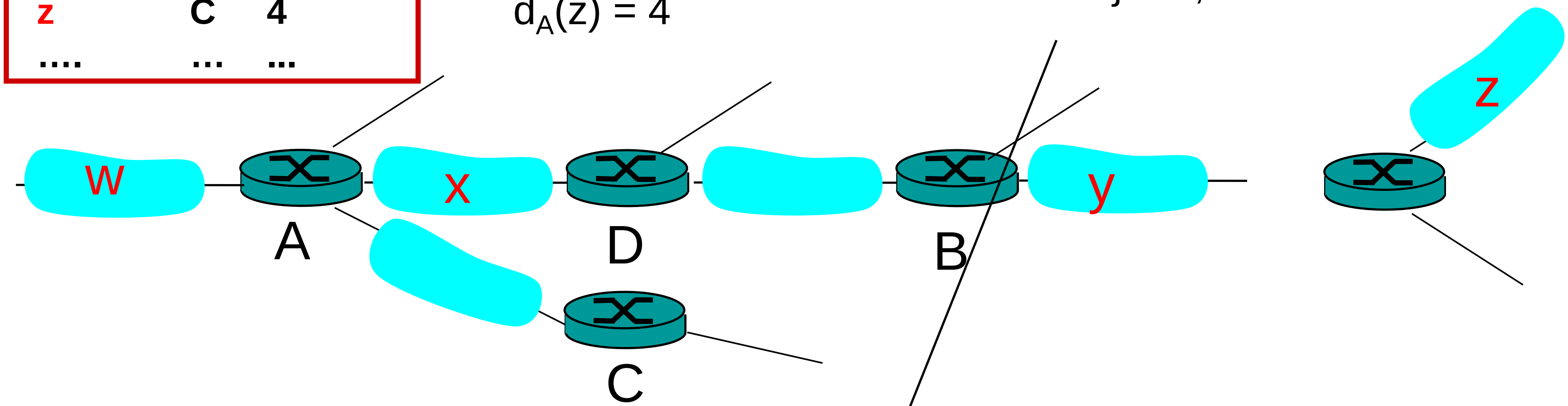
Similar to forwarding table

RIP: Example

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

Advertisement from A to D,
$d_A(z) = 4$

$d_D(z) = \min\{c(D,B)+ d_B(z), C(D,A)+ d_A(z)\} = \min\{7, 1+4\} = 5$; via A.



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

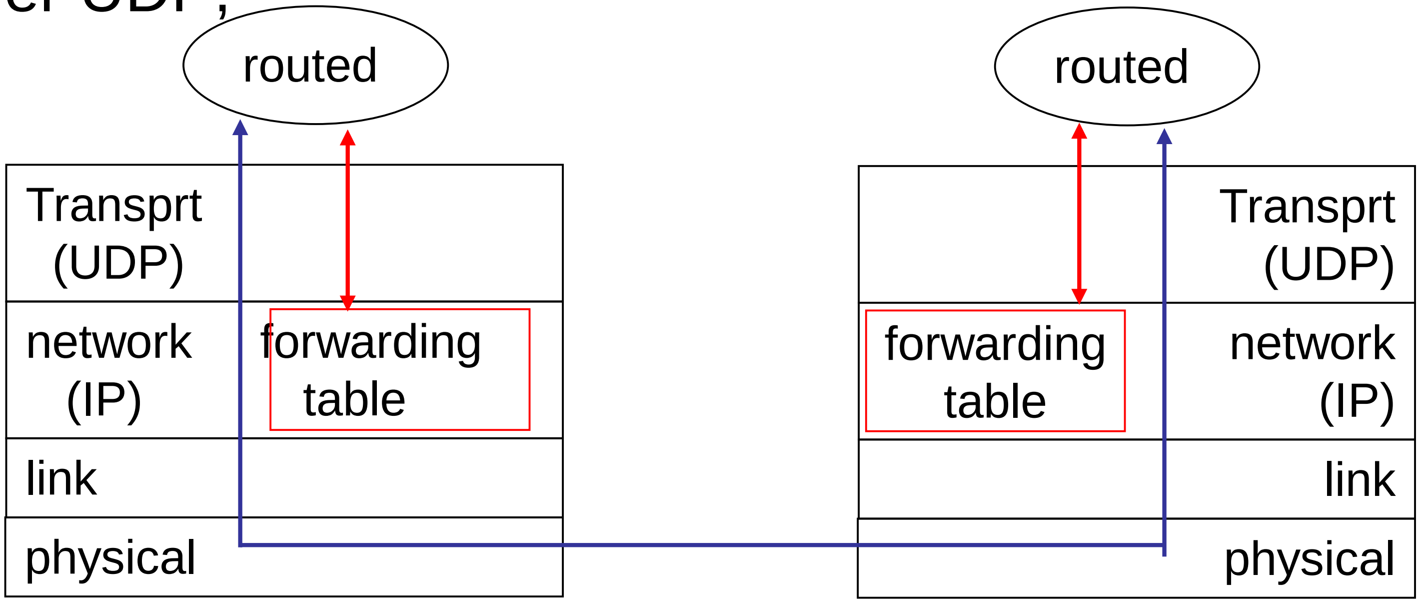
RIP: Link Failure and Recovery

If no advertisement heard after 180 sec (*the use of timer*) --> 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
- Count-to-infinity; *In RIP the maximum cost of a path is limited to 15*.

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;



A comparison of DV algorithm and RIP

DV algorithm	RIP
N (# of nodes)	Dest. Subnets in advertisement <= 25
c(x,y)	Link cost = 1
Node	Subnet
Message - dv (between neighbours)	RIP advertisement – routing table
Check B-F equation explicitly	Check B-F equation implicitly
store own dv and neighbours dvs	Routing table
Update neighbours whenever there is a change in own dv	Routers exchange advertisements about every 30 secs.
Count-to-infinity problem	"Count-to-16"

Overview

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

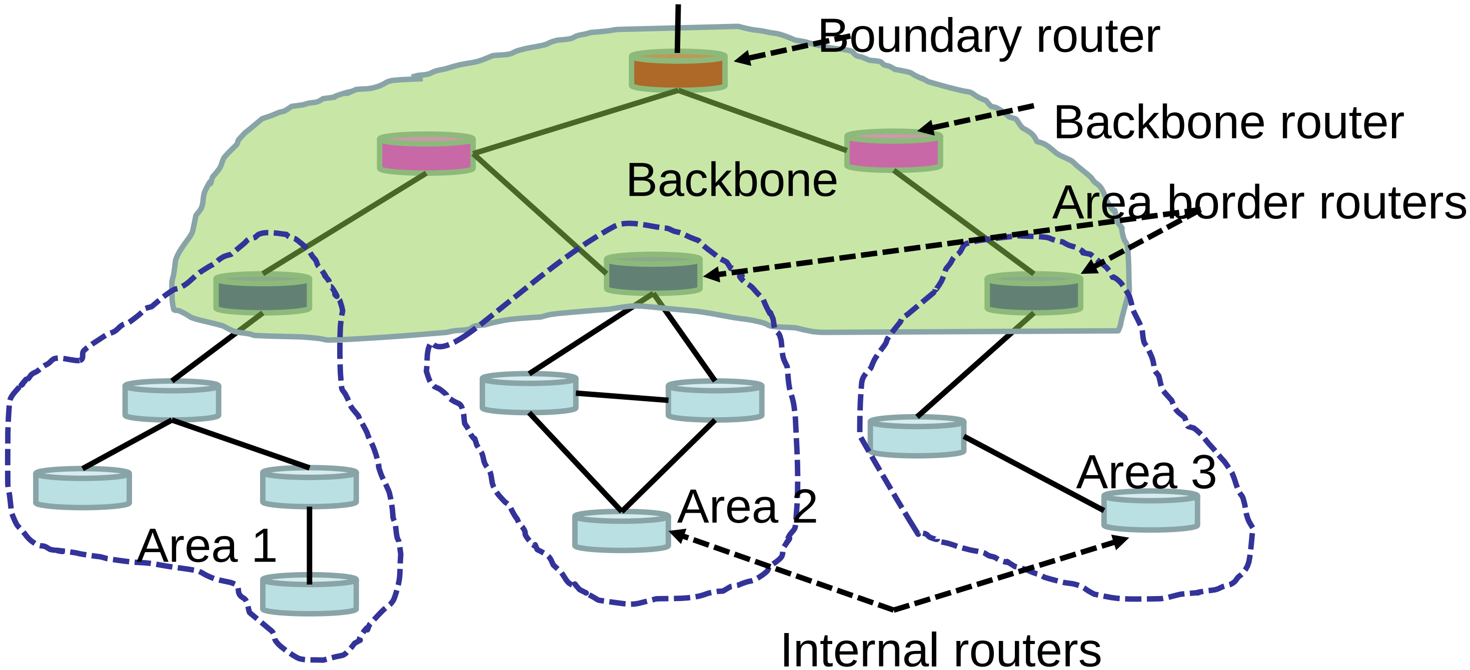
OSPF (Open Shortest Path First)

- "open": publicly available
- Uses Link State algorithm
  - Route computation using Dijkstra's algorithm
- Advertisements disseminated to entire AS (via broadcasting)
  - 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)

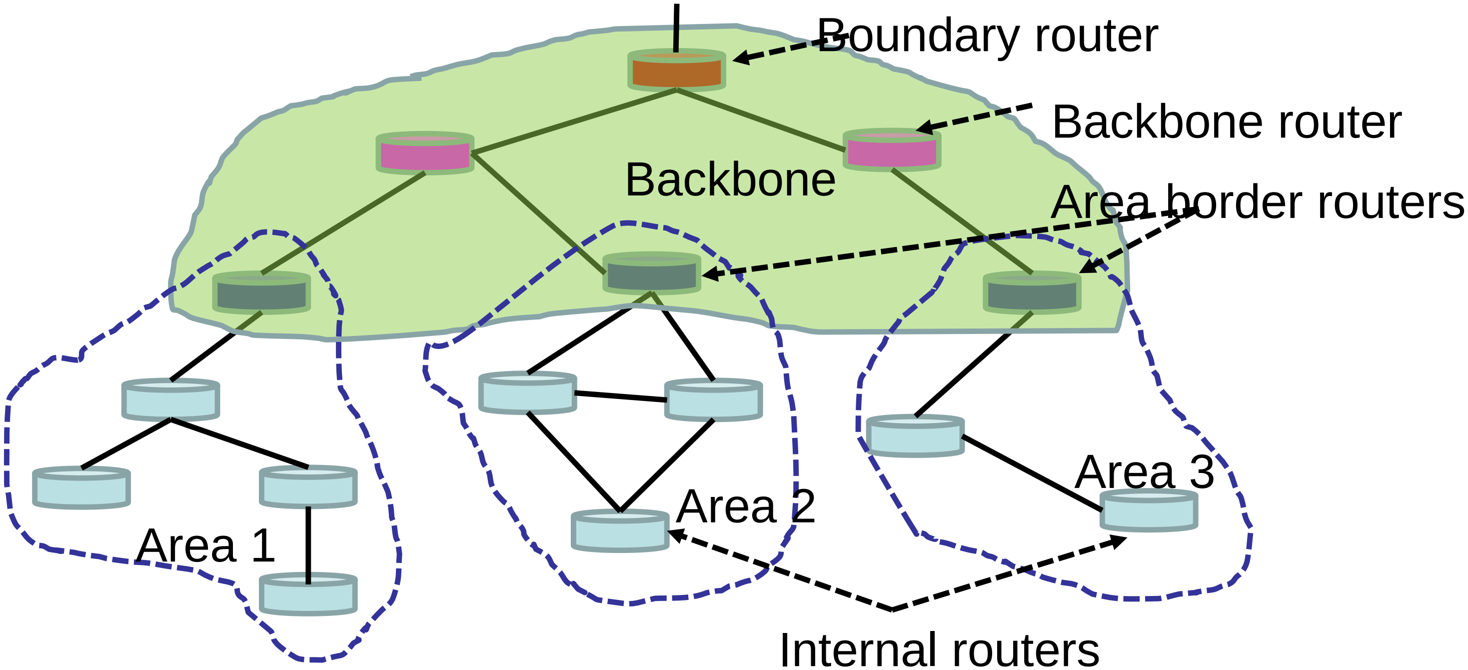
- Security: all OSPF messages authenticated (to prevent malicious intrusion)
- Multiple same-cost paths allowed (only one path in RIP)
- Link cost (weight) can be set by administrators; (traffic engineering)
- Integrated uni- and multicast support; (one-to-one, one-to one group; one-to-all)
- Hierarchical 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 (*dealing with the outside world typically through BGP*).

Overview

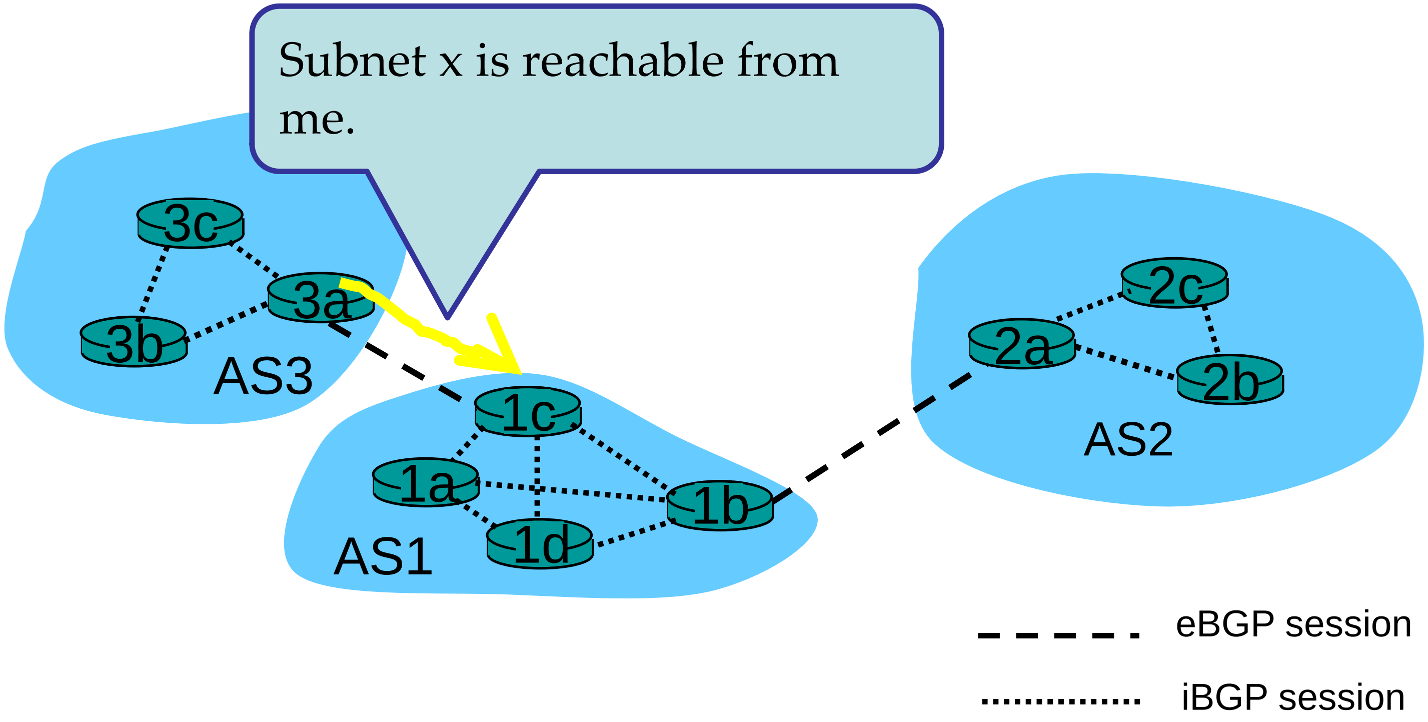
- NAT
- IPv6
- Routing in the Internet
  - Hierarchical routing
  - RIP
  - OSPF
  - BGP

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.
  2. Propagate the reachability information.
  3. Determine “good” routes to subnets.

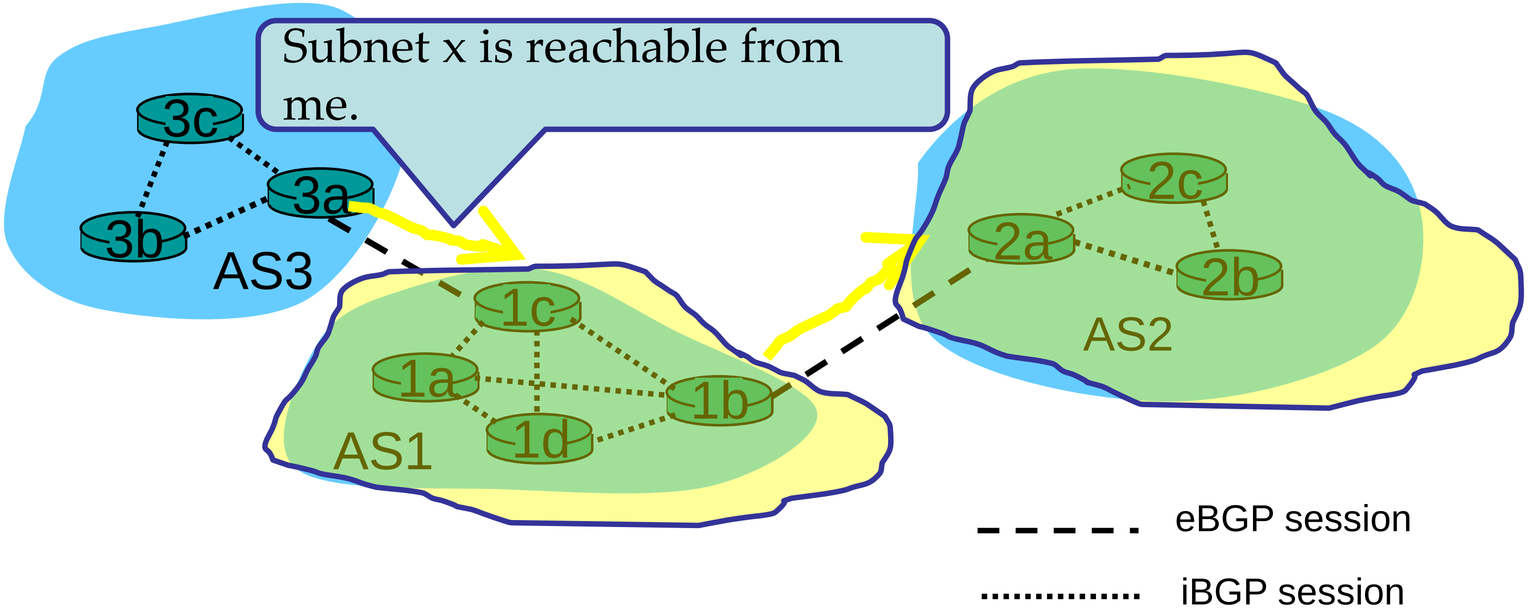
BGP basics

- Pairs of routers (BGP peers) exchange routing info over TCP connections (BGP is implemented as an application): BGP sessions
- When AS3 advertises a prefix to AS1, AS3 is *promising* it will forward any datagrams destined to that prefix.
  - AS3 can aggregate prefixes in its advertisement



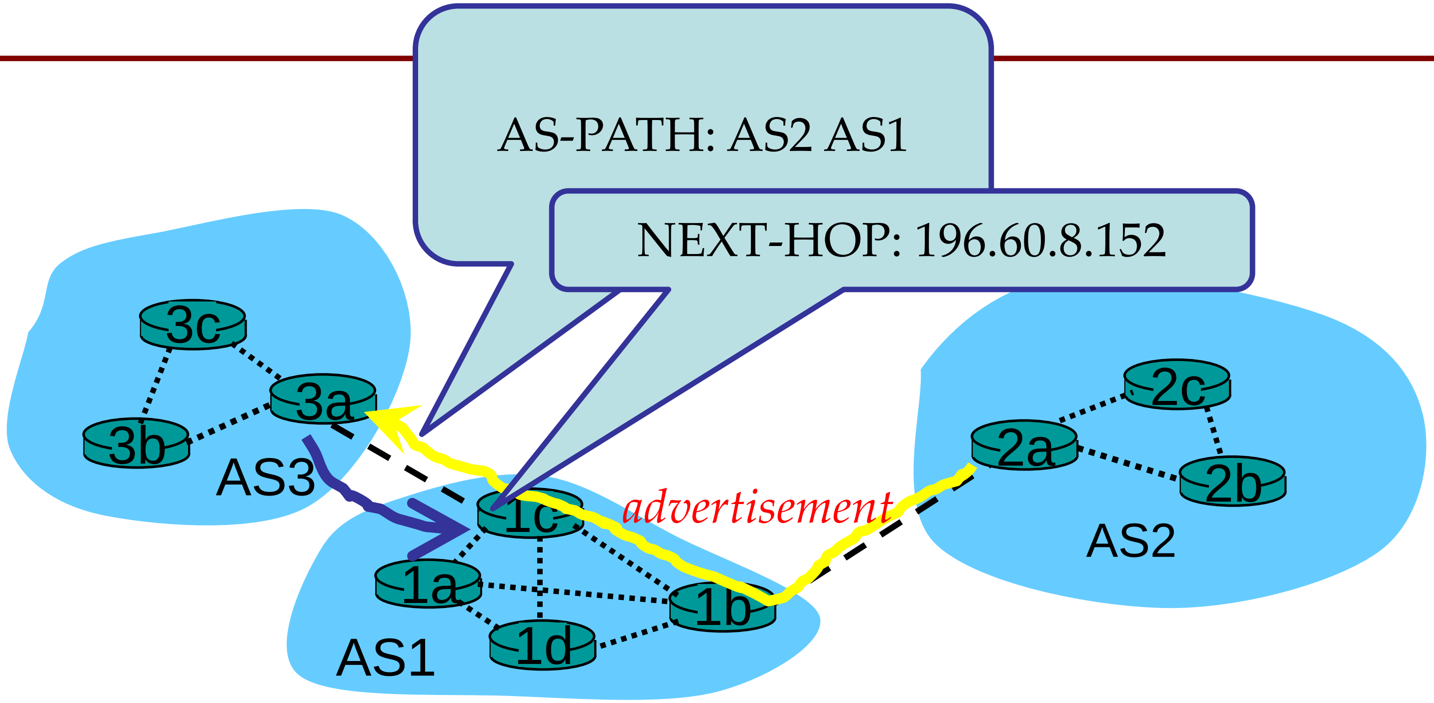
Distributing reachability info

- 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 ; if a router sees its AS is contained in the path list, it rejects the advertisement (to avoid loop).
  - 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.


AS-PATH: AS2 AS1
NEXT-HOP: 196.60.8.152
3c
3a
3b
AS3
advertisement
2c
2a
2b
AS2
1c
1a
1b
1d
AS1

An example

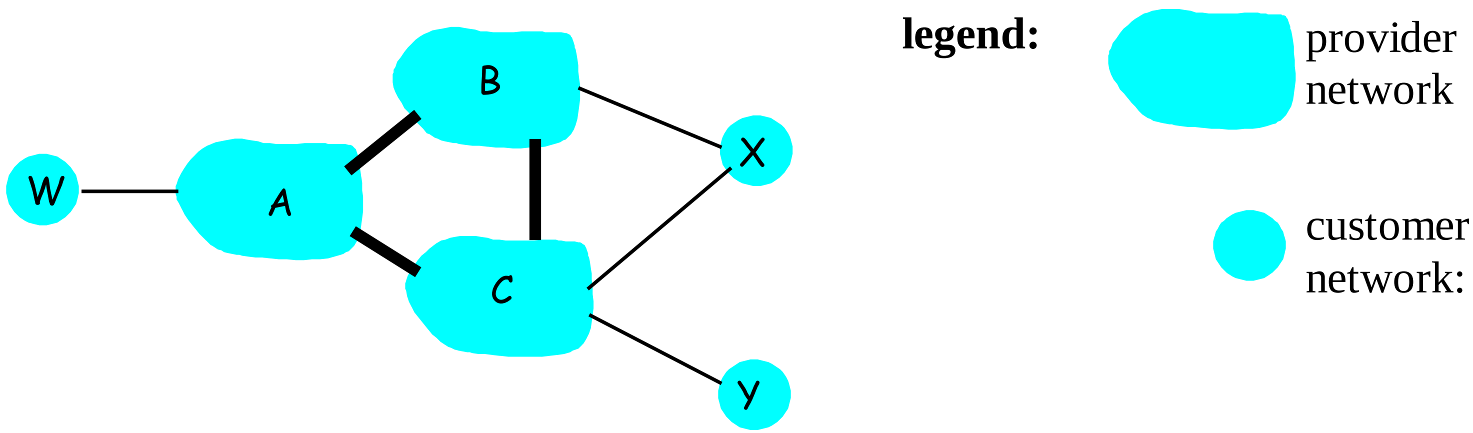
```
TIME: 06/01/19 00:00:00
TYPE: TABLE_DUMP_V2/IPV4_UNICAST
PREFIX: 1.44.24.0/21
SEQUENCE: 1236
FROM: 196.60.8.152 AS3491
ORIGINATED: 04/14/19 20:53:33
ORIGIN: IGP
ASPATH: 3491 7473 7474 4804
NEXT_HOP: 196.60.8.152
ATOMIC_AGGREGATE
AGGREGATOR: AS65367 10.194.27.154
COMMUNITY: 3491:2000 3491:2009 7473:10000 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
```

www.routeviews.org; bgpdump

BGP route selection

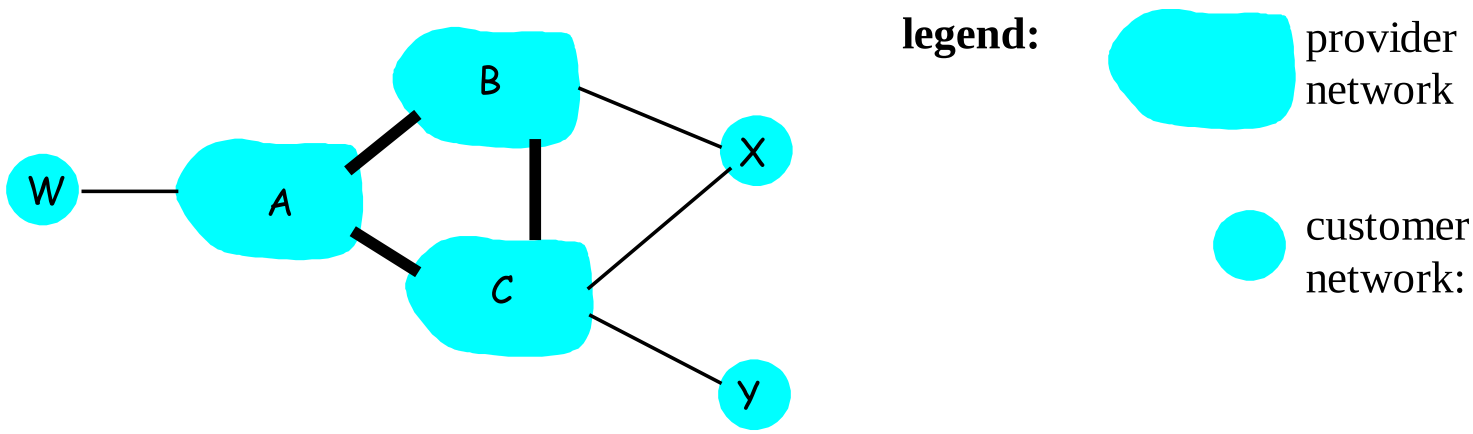
- Router may learn about more than 1 route to some prefix. Router must select route.
- Elimination rules:
  1. Local preference value attribute: policy decision
  2. Shortest AS-PATH (*related to DV algorithm*)
  3. Closest NEXT-HOP router: hot potato routing
  4. Additional criteria

BGP routing policy – *keep silent*



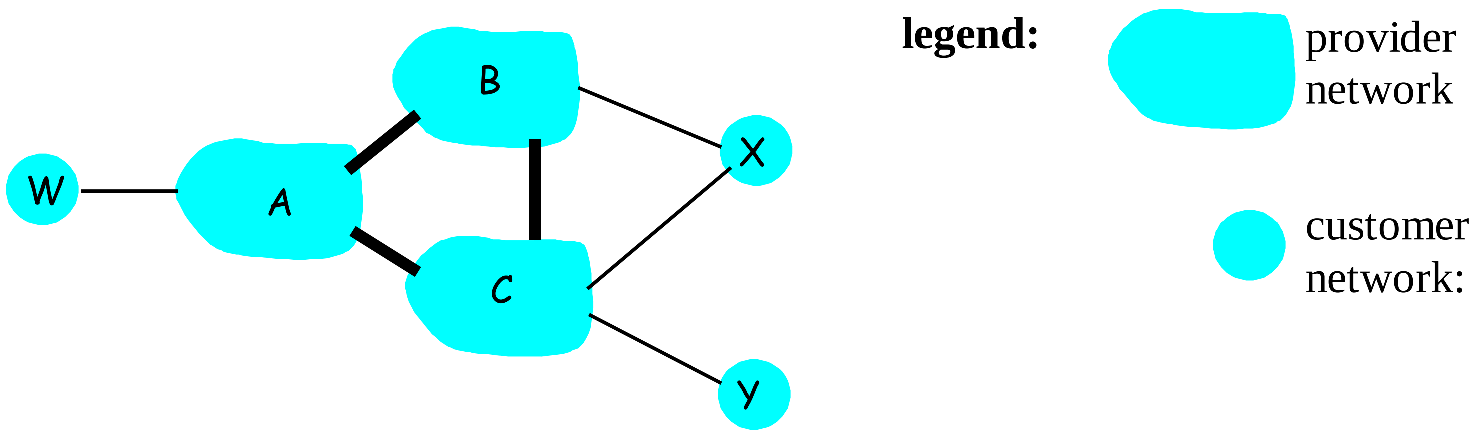
- A,B,C are provider networks
- X,W,Y are customer (of provider networks)
- X is dual-homed (multi-homed): attached to two networks
  - X does not want to route traffic from B via X to C
  - .. so X will not advertise to B a route to C

BGP routing policy (2) – *no free riding*



- A advertises to B the path AW
- B advertises to X the path BAW
- Should B advertise to C the path BAW?

BGP routing policy (2) -*no free riding*



- 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 (no free riding!)
  - B wants to force C to route to W via A
  - B wants to route *only* to/from its customers!

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