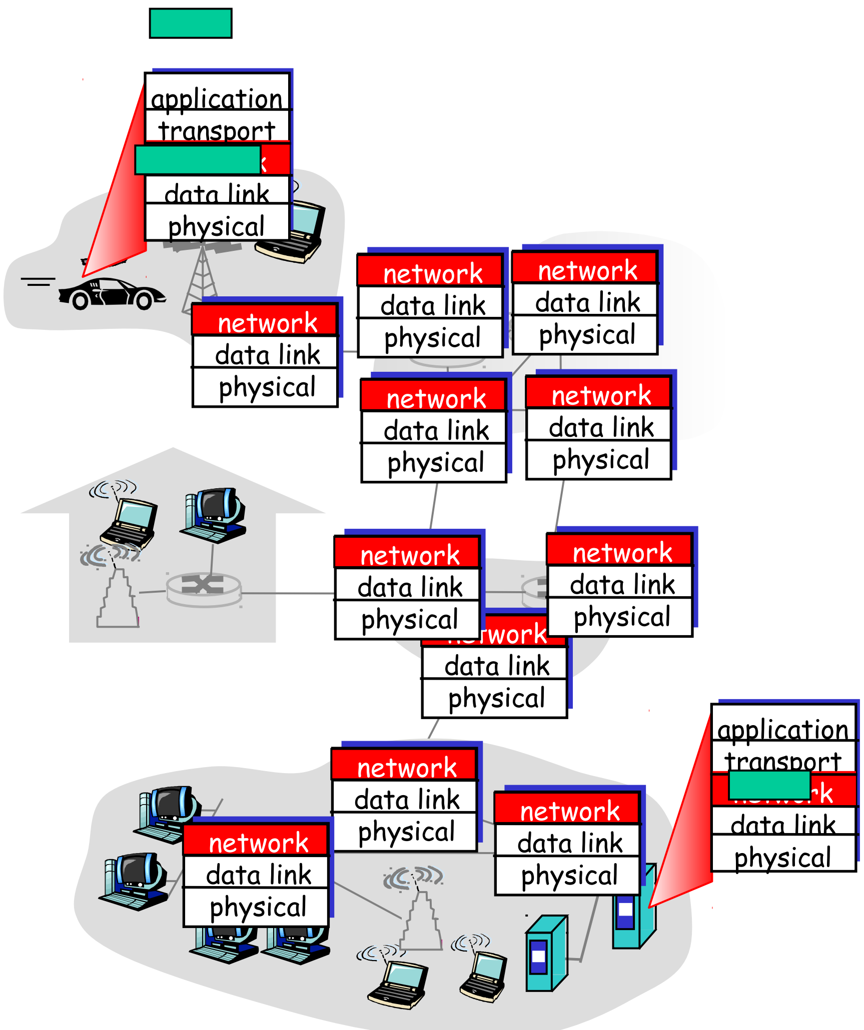
Network layer

Network layer

- transport segment from sending to receiving host
- on sending side puts segments into datagrams
- on rcving side, delivers segments to transport layer
- network layer protocols in *every* host, router

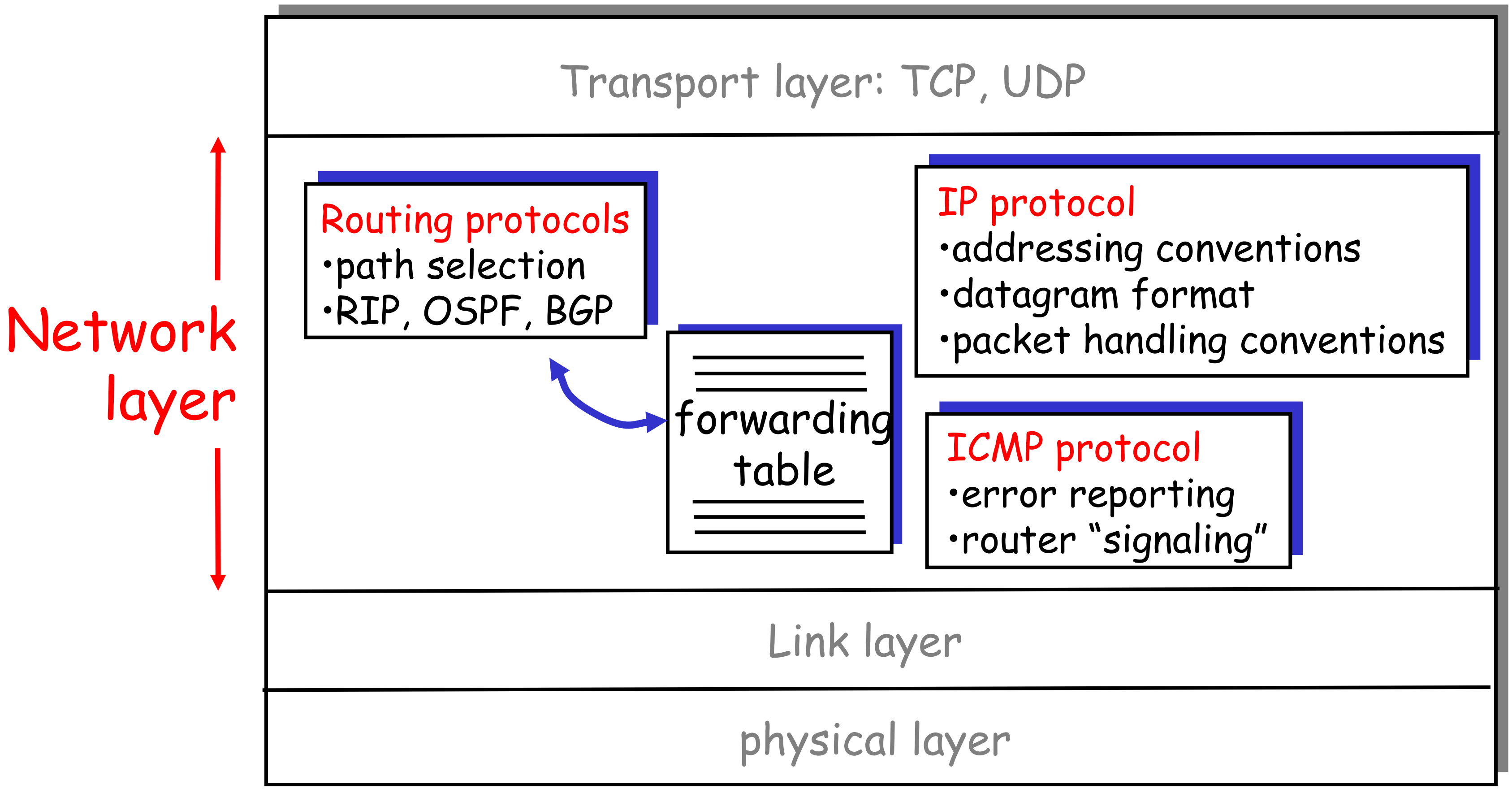


Network layer functions

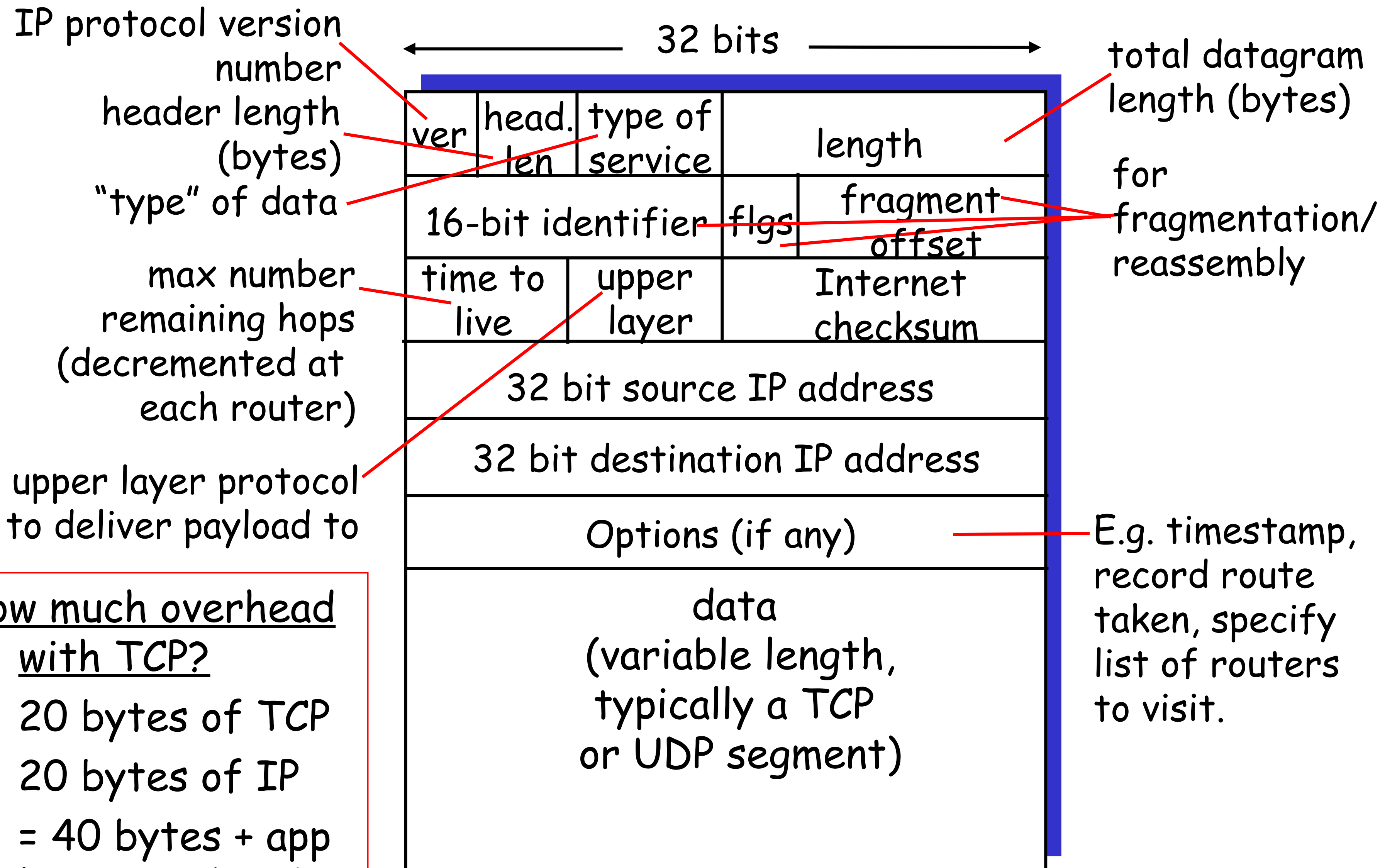
- Connection setup
  - datagram
  - connection-oriented, host-to-host connection
- Delivery semantics:
  - Unicast, broadcast, multicast, anycast
  - In-order, any-order
- Security
  - secrecy, integrity, authenticity
- Demux to upper layer
  - next protocol
  - Can be either transport or network (tunneling)
- Quality-of-service
  - provide predictable performance
- Fragmentation
  - break-up packets based on data-link layer properties
- Routing
  - path selection and packet forwarding
- Addressing
  - flat vs. hierarchical
  - global vs. local
  - variable vs. fixed length

The Internet Network layer

Host, router network layer functions:



IP datagram format



<u>how much overhead with TCP?</u>

- 20 bytes of TCP
- 20 bytes of IP
- = 40 bytes + app layer overhead

Recall network layer functions

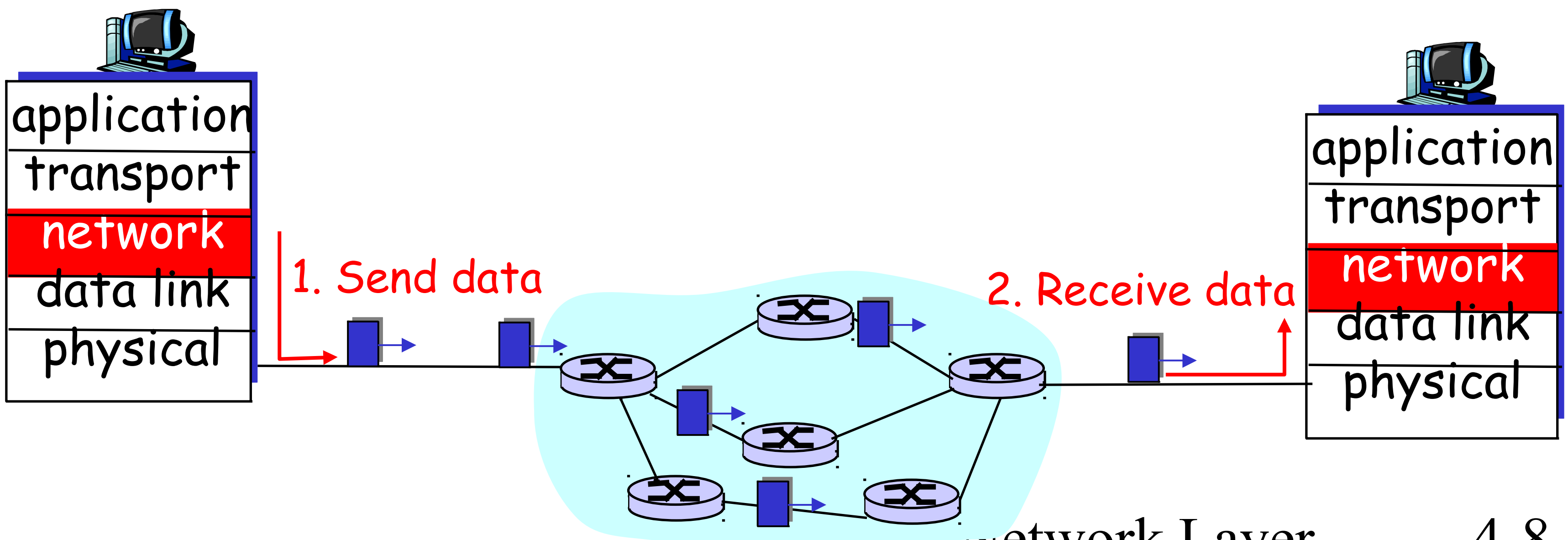
- How does IPv4 support..
  - Connection setup
  - Delivery semantics
  - Security
  - Demux to upper layer
  - Quality-of-service
  - Fragmentation
  - Addressing
  - Routing

IP connection setup

- Hourglass design
- No support for network layer connections
  - Unreliable datagram service
  - Out-of-order delivery possible
  - Connection semantics only at higher layer
  - Compare to ATM and phone network...

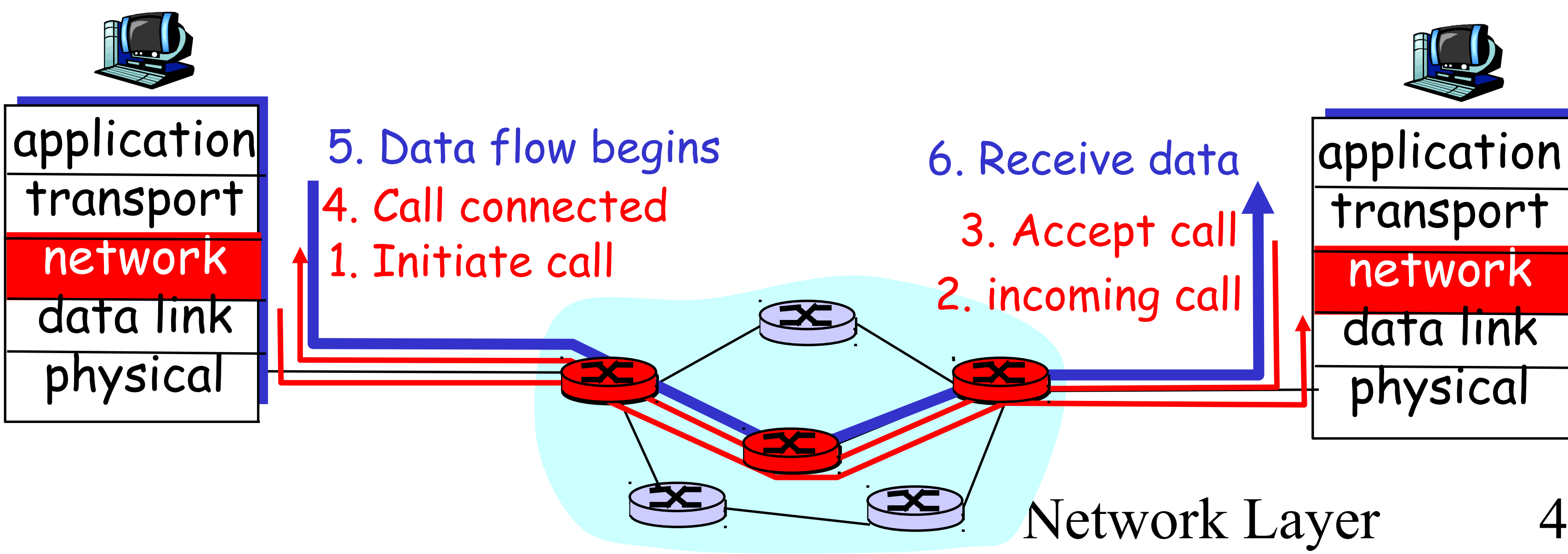
Connectionless network layers

- Postal service abstraction (Internet)
  - Model
    - no call setup or teardown at network layer
    - no service guarantees
  - Network support
    - no state within network on end-to-end connections
    - packets forwarded based on destination host ID



Connection-oriented network layers

- Circuit abstraction
  - Examples: ATM, frame relay, X.25, phone network
  - Model
    - call setup and teardown for each call
    - guaranteed performance during call
  - Network support
    - *every* router maintains "state" for each passing circuit
    - resources allocated per call



IP delivery semantics

- No reliability guarantees
  - Loss
- No ordering guarantees
  - Out-of-order delivery possible
- Unicast mostly
  - IP broadcast (255.255.255.255) not forwarded
  - IP multicast supported, but not widely used
    - 224.0.0.0 to 239.255.255.255

IP security

- Weak support for integrity
  - IP checksum
    - IP has a header checksum, leaves data integrity to TCP/UDP
    - http://www.rfc-editor.org/rfc/rfc1141.txt
  - No support for secrecy, authenticity
- IPsec
  - Retrofit IP network layer with encryption and authentication
  - http://www.rfc-editor.org/rfc/rfc2411.txt

IP demux to upper layer

- http://www.rfc-editor.org/rfc/rfc1700.txt
  - Protocol type field
    - 1 = ICMP
    - 4 = IP in IP
    - 6 = TCP
    - 17 = UDP
    - 88 = EIGRP
    - 89 = OSPF

IP quality of service

- IP originally had "type-of-service" (TOS) field to eventually support quality
  - Not used, ignored by most routers
- Need to provide applications with performance guarantees
  - Mid 90s: Add circuits to the Internet!
    - Integrated services (intserv) and RSVP signalling
    - Per-flow end-to-end QoS support
    - Per-flow signaling and network resource allocation

IP quality of service

- Protocols developed and standardized
  - RSVP signalling protocol
  - Intserv service models
- Failed miserably…Why?
  - Complexity
    - Scheduling
    - Routing (pinning routes)
    - Per-flow signalling overhead
  - Lack of scalability
    - Per-flow state
  - Economics
    - Providers with no incentive to deploy
    - SLA, end-to-end billing issues
  - QoS a weak-link property
    - Requires every device on an end-to-end basis to support flow

IP quality of service

- Now it's diffserv…
  - Use the "type-of-service" bits as a priority marking
  - http://www.rfc-editor.org/rfc/rfc2474.txt
  - http://www.rfc-editor.org/rfc/rfc2475.txt
  - http://www.rfc-editor.org/rfc/rfc2597.txt
  - http://www.rfc-editor.org/rfc/rfc2598.txt

IP Addressing

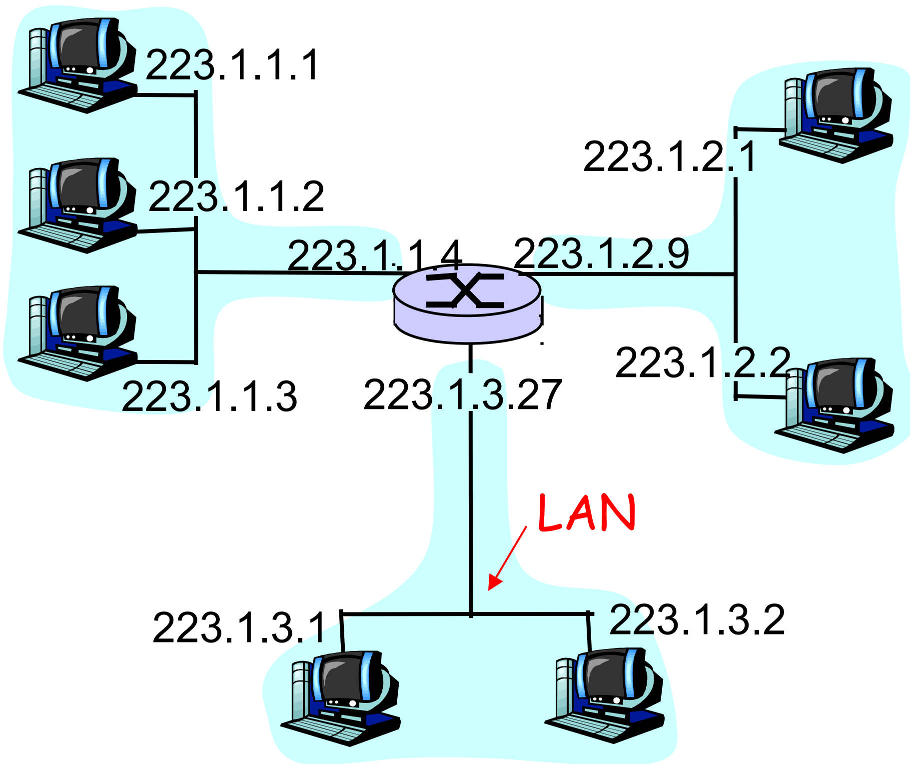
- IP address:
  - 32-bit identifier for host/router *interface*
  - Managed by ICANN: Internet Corporation for Assigned Names and Numbers
    - Allocates addresses, manages DNS, resolves disputes

223.1.1.1 = 11011111 00000001 00000001 00000001

223 1 1 1

IP Addressing

- IP address:
  - Addresses hierarchical (like post office)
  - Network part (high order bits)
  - Host part (low order bits)
- *What's a network ?*
  - all interfaces that can physically reach each other without intervening router
  - each interface shares the same network part of IP address
  - routers typically have multiple interfaces



network consisting of 3 IP networks (for IP addresses starting with 223, first 24 bits are network address)

How did networks get IP addresses?

- Total IP address size: 4 billion
- Initially one large class
  - 256 networks each with 16 million hosts
  - Problem: one size does not fit all
- Then, classful addressing to accommodate smaller networks
  - Class A: 128 networks, 16M hosts
    - 1.0.0.0 to 127.255.255.255
  - Class B: 16K networks, 64K hosts
    - 128.0.0.0 to 191.255.255.255
  - Class C
    - 192.0.0.0 to 223.255.255.255
  - Multicast + reserved
    - 224.0.0.0 to 255.255.255.255

Special IP Addresses

- Private addresses
  - http://www.rfc-editor.org/rfc/rfc1918.txt
  - Class A: 10.0.0.0 - 10.255.255.255 (10.0.0.0/8 prefix)
  - Class B: 172.16.0.0 - 172.31.255.255 (172.16.0.0/12 prefix)
  - Class C: 192.168.0.0 - 192.168.255.255 (192.168.0.0/16 prefix)
- 127.0.0.1: local host (a.k.a. the loopback address)

IP Addressing problems

- Inefficient use of address space
  - Class A (rarely given out, sparse usage)
  - Class B = 64k hosts (sparse usage)
    - Very few LANs have close to 64K hosts
- Address space depletion
  - Classes A and B take huge chunks of space but not used much
  - Not many class C addresses left to give out
- Explosion of routes
  - Increasing use of class C explodes # of routes
  - Total routes potentially > 2,113,664 networks and network routes !

CIDR

CIDR: Classless InterDomain Routing

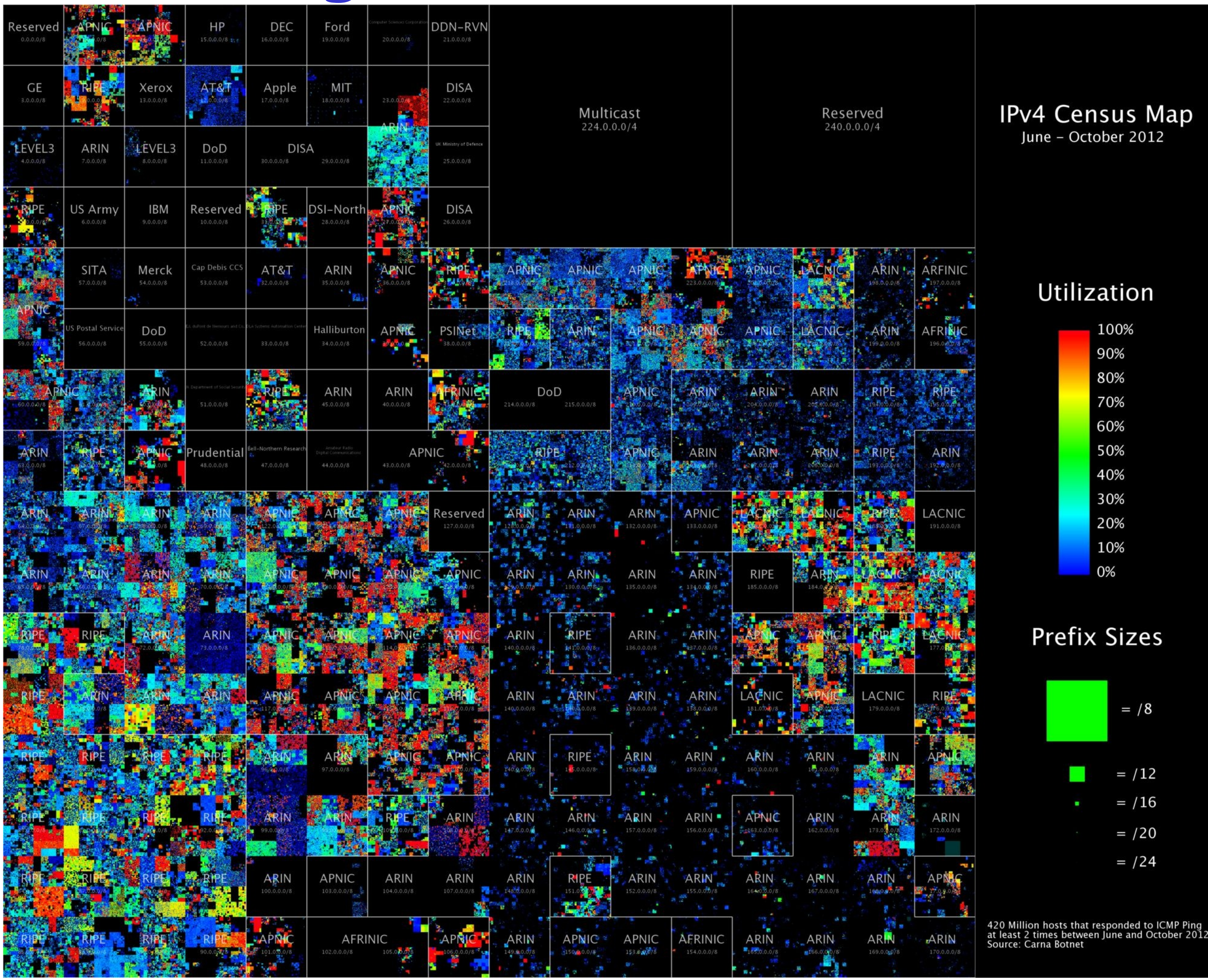
- Arbitrarily aggregate and split up adjacent network address
- Allows one to split large network blocks into multiple smaller ones (increase usage of Class A & B)
- Allows one to combine small network blocks into a single large one (reduce routes from Class C usage)

variable network part | host part

11001000 00010111 00010000 00000000

200.23.16.0/23

IPv4 usage

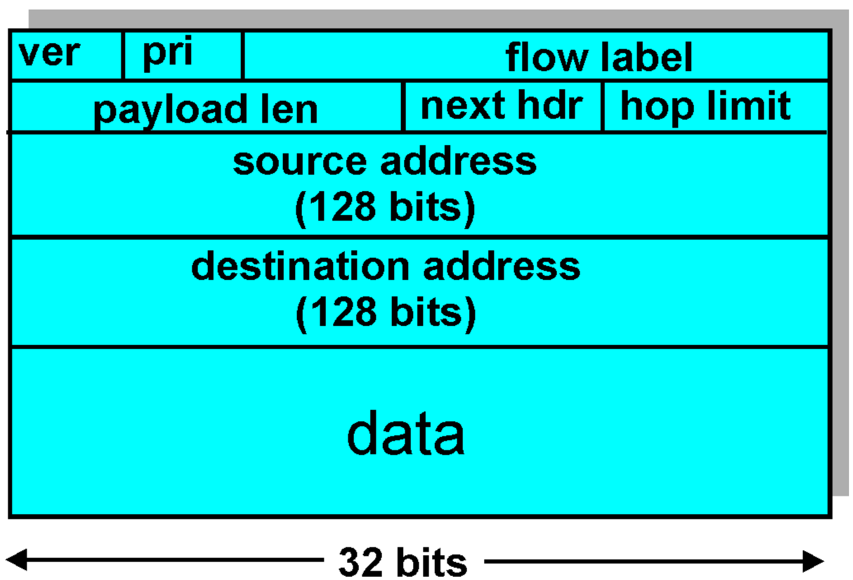

IPv4 Census Map
June – October 2012
Multicast
224.0.0.0/4
Reserved
240.0.0.0/4
Utilization
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Prefix Sizes
= /8
= /12
= /16
= /20
= /24
420 Million hosts that responded to ICMP Ping
at least 2 times between June and October 2012
Source: Carna Botnet

IPv6

- IPv4 running out of addresses
- Need to replace it with a new network protocol
- What changes should be made in….
    - IP addressing
    - IP delivery semantics
    - IP quality of service
    - IP security
    - IP routing
    - IP fragmentation
    - IP error detection

IPv6 Changes

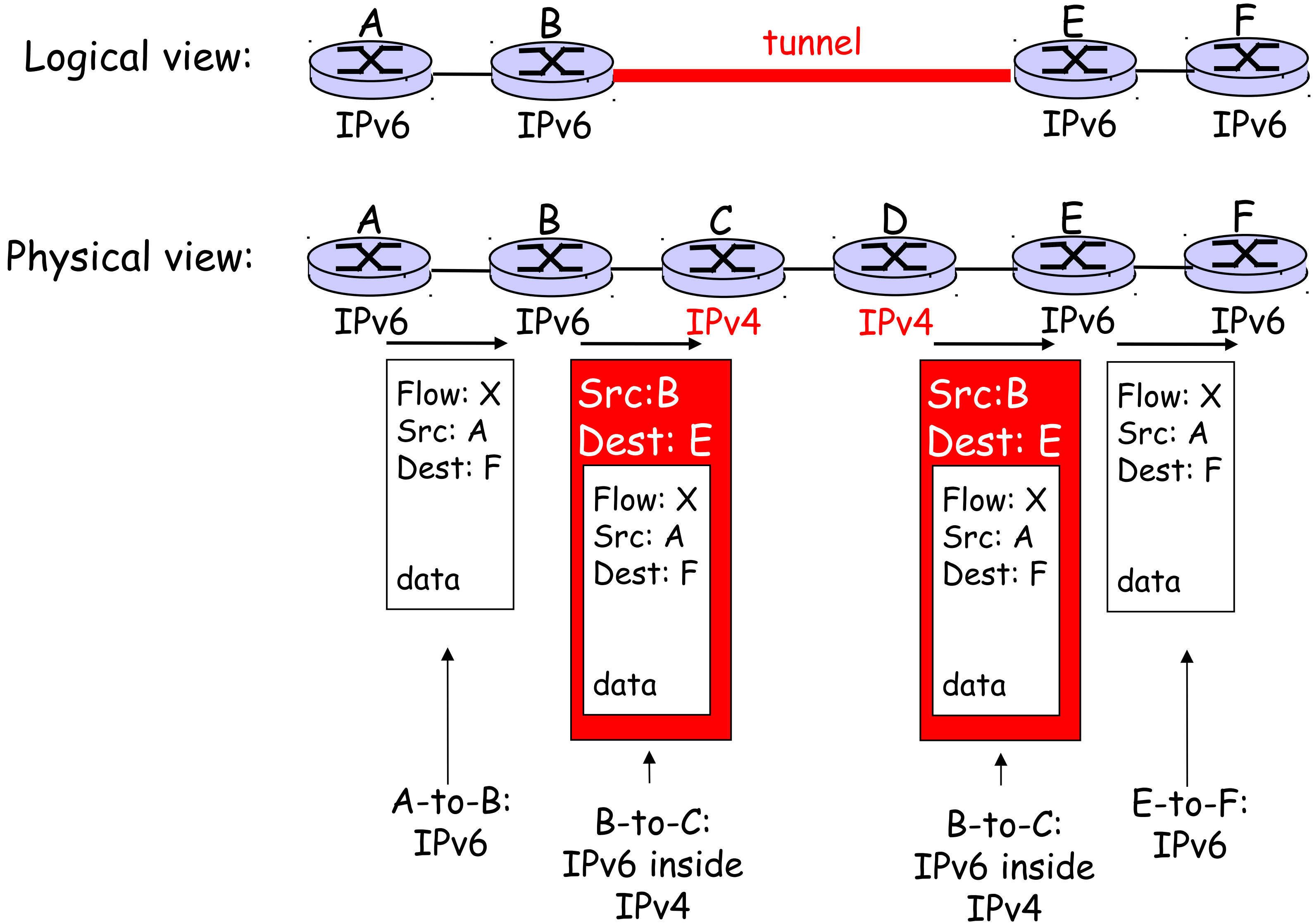
- Addresses are 128bit
- Simplification
  - Removes checksum
  - Eliminates fragmentation



Transition From IPv4 To IPv6

- Not all routers can be upgraded simultaneous
  - How will the network operate with mixed IPv4 and IPv6 routers?
  - *Tunneling:* IPv6 carried as payload in an IPv4 datagram among IPv4 routers

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 inside
IPv4
B-to-C:
IPv6 inside
IPv4
E-to-F:
IPv6

Routing

Internet routing with IP addresses

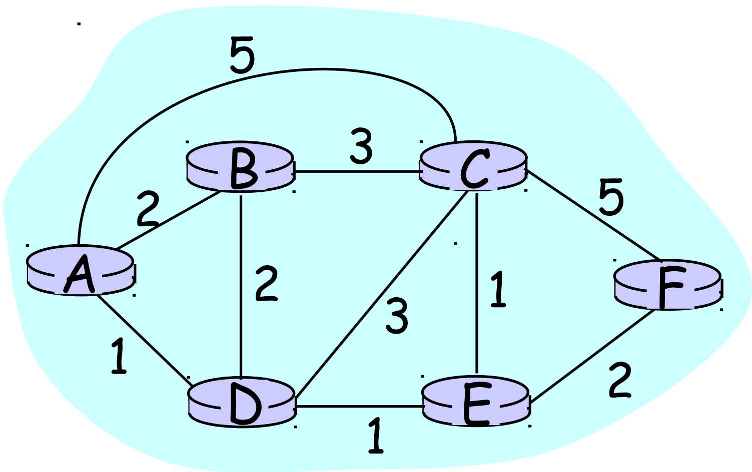
- Internet routing done via hop-by-hop forwarding based on destination IP address
  - Each router has forwarding table of..
    - destination IP → next hop IP address
  - Each router runs a routing protocol to create forwarding table
    - Routing algorithm

Routing protocols and algorithms

Goal: determine "good" path (sequence of routers) thru network from source to dest.

Graph abstraction for routing algorithms:

- Routing algorithms find minimum cost paths through graph



Routing Algorithm classification

Global or decentralized information?

Global:

- all routers have complete topology, link cost info
- "link state" algorithms

Decentralized:

- router knows physically-connected neighbors, link costs to neighbors
- iterative process of computation, exchange of info with neighbors
- "distance vector" algorithms

Hierarchical Routing

scale: with 200 million destinations:

- can't store all dest's in routing tables!
- routing table exchange would swamp links!
- Flat routing does not scale

administrative autonomy

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

Routing Hierarchies

- Key observation
  - Need less information with increasing distance to destination
  - Hierarchical routing
    - saves table size
    - reduces update traffic
    - allows routing to scale

Areas

- Divide network into areas
  - Within area, each node has routes to every other node
  - Outside area
    - Each node has routes for other top-level areas only (not nodes within those areas)
    - Inter-area packets are routed to nearest appropriate border router

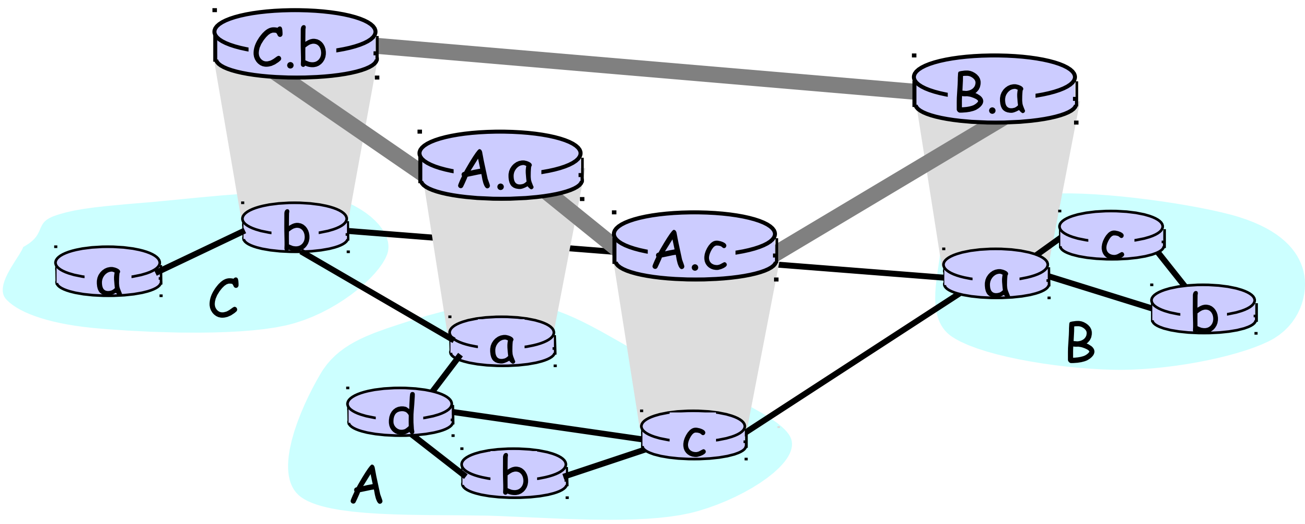
Internet Routing Hierarchy

- Internet areas called "autonomous systems" (AS)
  - administrative autonomy
- routers in same AS run same routing protocol
  - "intra-AS" routing protocol (IGP)

Border routers

- Special routers in AS that directly link to another AS
  - also run inter-AS routing protocol or border gateway protocol (BGP) with other gateway routers in other AS's

Internet Routing Hierarchy



Inter-AS routing

- Done using BGP (Border Gateway Protocol)
  - Uses distance-vector style algorithms
- BGP messages exchanged using TCP.
  - Advantages:
    - Simplifies BGP
    - No need for periodic refresh - routes are valid until withdrawn, or the connection is lost
    - Incremental updates
  - Disadvantages
    - BGP TCP spoofing attack
    - Congestion control on a routing protocol?
    - Poor interaction during high load (Code Red)
    - No authentication of route advertisements
      - Pakistan Youtube incident

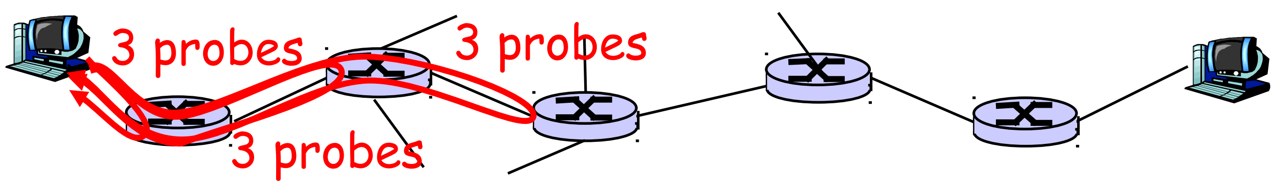
ICMP: Internet Control Message Protocol

- Essentially a network-layer protocol for passing control messages
- used by hosts & routers to communicate network-level information
  - error reporting: unreachable host, network, port, protocol
  - echo request/reply (used by ping)
- ICMP message: type, code plus first 8 bytes of IP datagram causing error
- http://www.rfc-editor.org/rfc/rfc792.txt

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
4	0	source quench (congestion control - not used)
8	0	echo request (ping)
9	0	route advertisement
10	0	router discovery
11	0	TTL expired
12	0	bad IP header

ICMP and traceroute

- What do "real" Internet delay & loss look like?
- **Traceroute program:** provides delay measurement from source to router along end-end Internet path towards destination.



ICMP and traceroute

- Source sends series of UDP segments to dest
  - First has TTL =1
  - Second has TTL=2, etc.
  - Unlikely port number
- When nth datagram arrives to nth router:
  - Router discards datagram
  - And sends to source an ICMP message (type 11, code 0)
  - Message includes name of router& IP address
- When ICMP message arrives, source calculates RTT
- Traceroute does this 3 times

Stopping criterion

- UDP segment eventually arrives at destination host
- Destination returns ICMP "host unreachable" packet (type 3, code 3)
- When source gets this ICMP, stops.

Examples

traceroute: gaia.cs.umass.edu to www.eurecom.fr

Three delay measurements from gaia.cs.umass.edu to cs-gw.cs.umass.edu

1 cs-gw (128.119.240.254) 1 ms 1 ms 2 ms
2 border1-rt-fa5-1-0.gw.umass.edu (128.119.3.145) 1 ms 1 ms 2 ms
3 cht-vbns.gw.umass.edu (128.119.3.130) 6 ms 5 ms 5 ms
4 jn1-at1-0-0-19.wor.vbns.net (204.147.132.129) 16 ms 11 ms 13 ms
5 jn1-so7-0-0-0.wae.vbns.net (204.147.136.136) 21 ms 18 ms 18 ms
6 abilene-vbns.abilene.ucaid.edu (198.32.11.9) 22 ms 18 ms 22 ms
7 nycm-wash.abilene.ucaid.edu (198.32.8.46) 22 ms 22 ms 22 ms
8 62.40.103.253 (62.40.103.253) 104 ms 109 ms 106 ms
9 de2-1.de1.de.geant.net (62.40.96.129) 109 ms 102 ms 104 ms
10 de.fr1.fr.geant.net (62.40.96.50) 113 ms 121 ms 114 ms
11 renater-gw.fr1.fr.geant.net (62.40.103.54) 112 ms 114 ms 112 ms
12 nio-n2.cssi.renater.fr (193.51.206.13) 111 ms 114 ms 116 ms
13 nice.cssi.renater.fr (195.220.98.102) 123 ms 125 ms 124 ms
14 r3t2-nice.cssi.renater.fr (195.220.98.110) 126 ms 126 ms 124 ms
15 eurecom-valbonne.r3t2.ft.net (193.48.50.54) 135 ms 128 ms 133 ms
16 194.214.211.25 (194.214.211.25) 126 ms 128 ms 126 ms
17 * * *
18 * * *
19 fantasia.eurecom.fr (193.55.113.142) 132 ms 128 ms 136 ms

trans-oceanic link

* means no response (probe lost, router not replying)

Try it

- Some routers labeled with airport code of city they are located in
  - traceroute www.yahoo.com
    - Packets go to SEA, back to PDX, SJC
  - traceroute www.oregonlive.com
    - Packets go to SMF, SFO, SJC, NYC, EWR.
  - traceroute www.uoregon.edu
    - Packets go to Pittock block to Eugene
  - traceroute www.lclark.edu
    - Packets go to SEA and back to PDX

Internet overview complete

- Technical background for the rest of the course