

Chapter 4 Network Layer

Reti di Elaboratori

Corso di Laurea in Informatica

Università degli Studi di Roma "La Sapienza"

Canale A-L

Prof.ssa Chiara Petrioli

Parte di queste slide sono state prese dal materiale associato al libro
Computer Networking: A Top Down Approach, 5th edition.

All material copyright 1996-2009

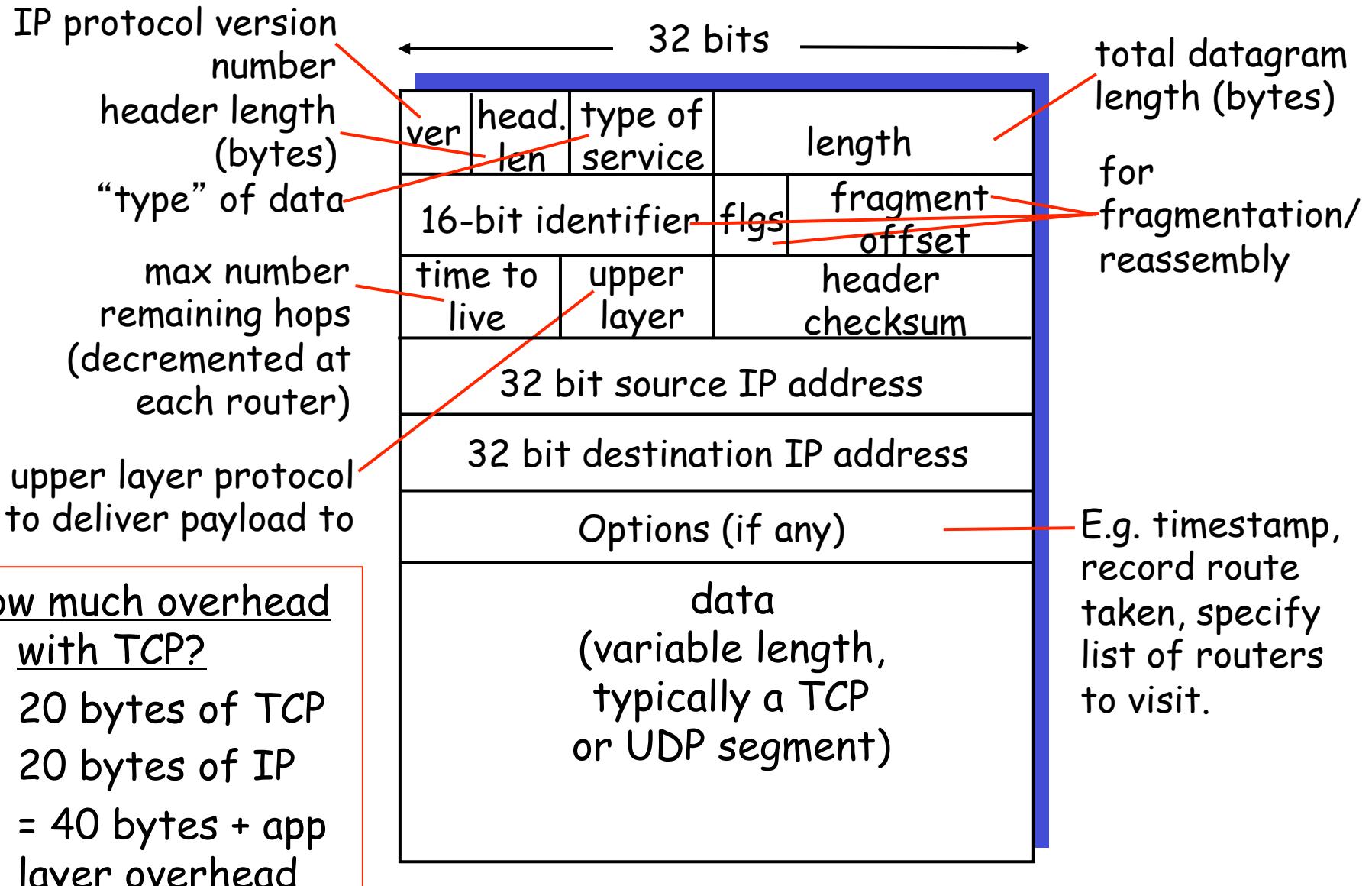
J.F Kurose and K.W. Ross, All Rights Reserved

Thanks also to Antonio Capone, Politecnico di Milano, Giuseppe Bianchi and
Francesco LoPresti, Un. di Roma Tor Vergata

Chapter 4: Network Layer

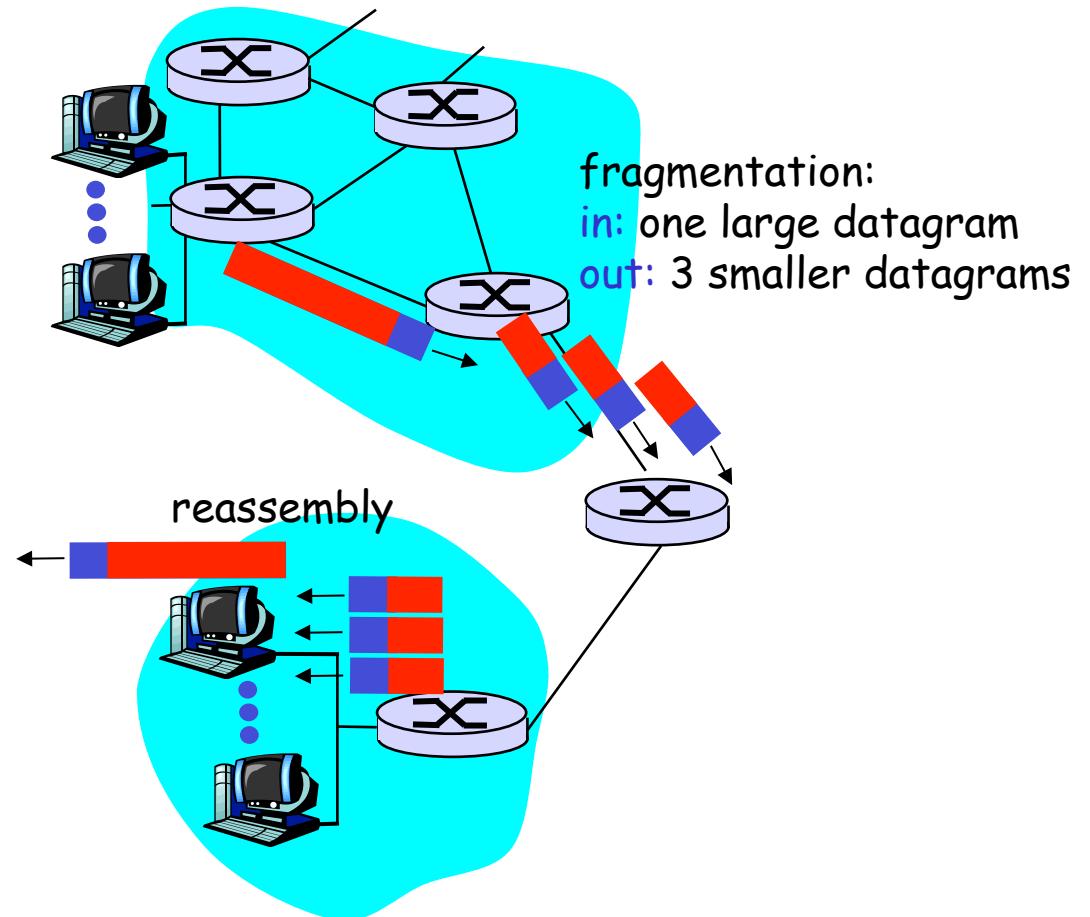
- 4.1 Introduction
- 4.2 Virtual circuit and datagram networks
- 4.3 What's inside a router
- 4.4 IP: Internet Protocol
 - Datagram format
 - IPv4 addressing
 - ICMP
 - IPv6
- 4.5 Routing algorithms
 - Link state
 - Distance Vector
 - Hierarchical routing
- 4.6 Routing in the Internet
 - RIP
 - OSPF
 - BGP
- 4.7 Broadcast and multicast routing

IP datagram format



IP Fragmentation & Reassembly

- network links have MTU (max.transfer size) - largest possible link-level frame.
 - different link types, different MTUs
- large IP datagram divided (“fragmented”) within net
 - one datagram becomes several datagrams
 - “reassembled” only at final destination
 - IP header bits used to identify, order related fragments



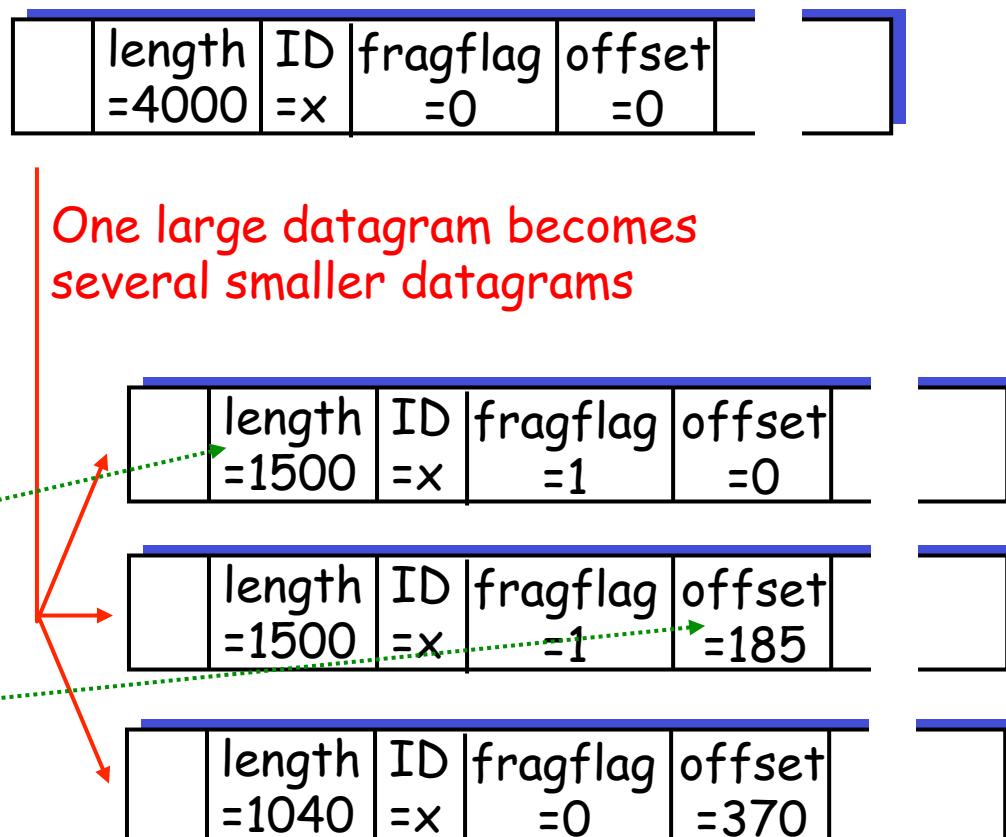
IP Fragmentation and Reassembly

Example

- 4000 byte datagram
- MTU = 1500 bytes

1480 bytes in data field

offset =
 $1480/8$



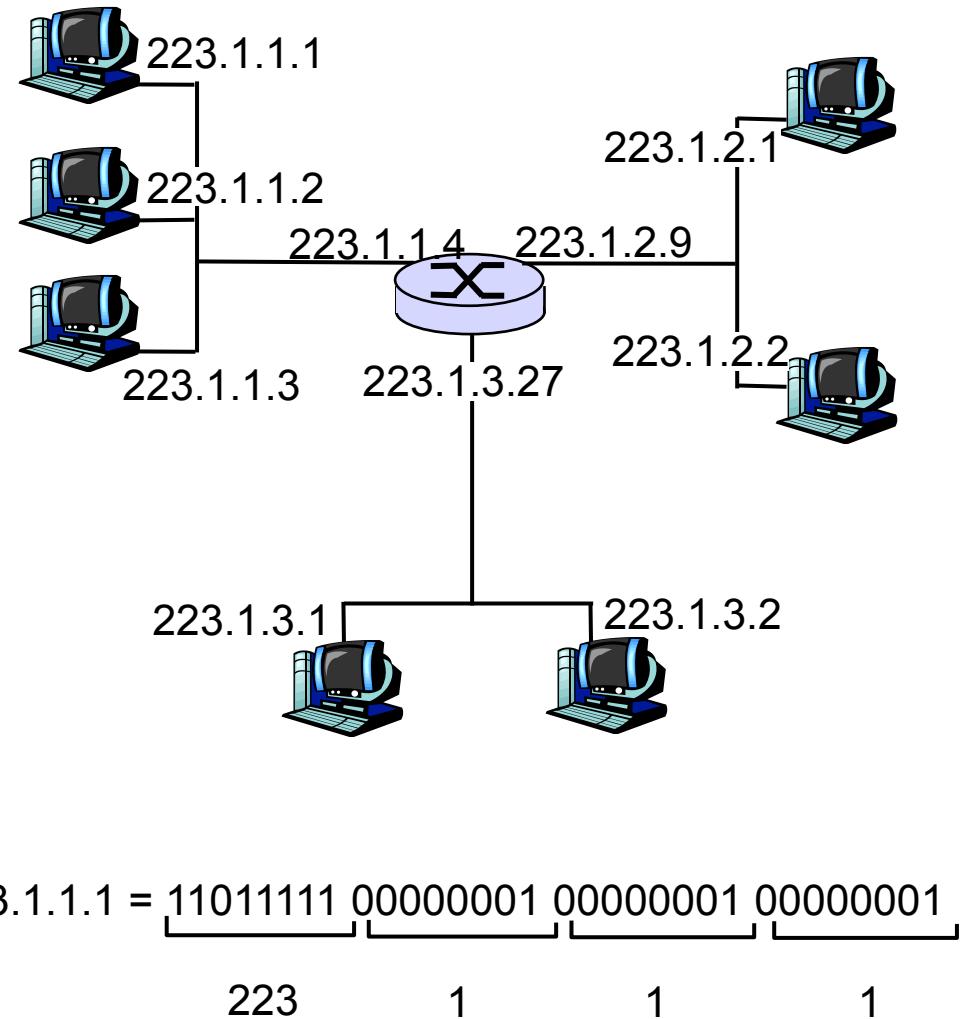
Offset: i dati saranno inseriti a partire
Dal byte YY

Chapter 4: Network Layer

- 4.1 Introduction
- 4.2 Virtual circuit and datagram networks
- 4.3 What's inside a router
- 4.4 IP: Internet Protocol
 - Datagram format
 - IPv4 addressing
 - ICMP
 - IPv6
- 4.5 Routing algorithms
 - Link state
 - Distance Vector
 - Hierarchical routing
- 4.6 Routing in the Internet
 - RIP
 - OSPF
 - BGP
- 4.7 Broadcast and multicast routing

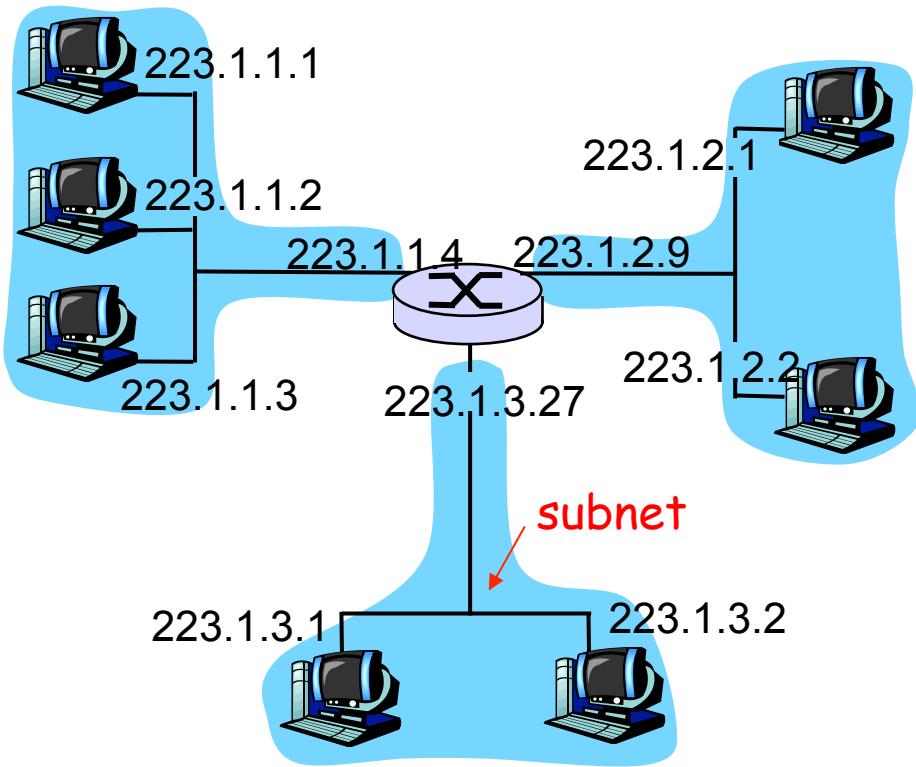
IP Addressing: introduction

- IP address: 32-bit identifier for host, router *interface*
- *interface*: connection between host/router and physical link
 - router's typically have multiple interfaces
 - host typically has one interface
 - IP addresses associated with each interface



Subnets

- IP address:
 - subnet part (high order bits)
 - host part (low order bits)
- *What's a subnet?*
 - device interfaces with same subnet part of IP address
 - can physically reach each other without intervening router

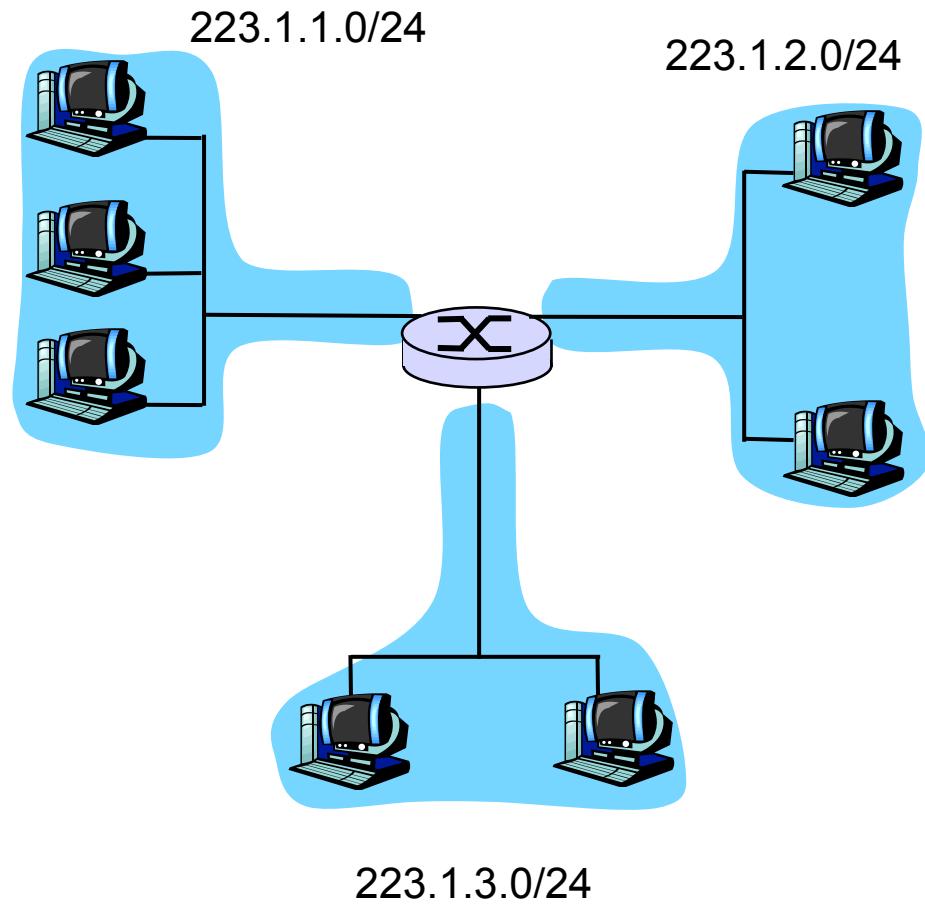


network consisting of 3 subnets

Subnets

Recipe

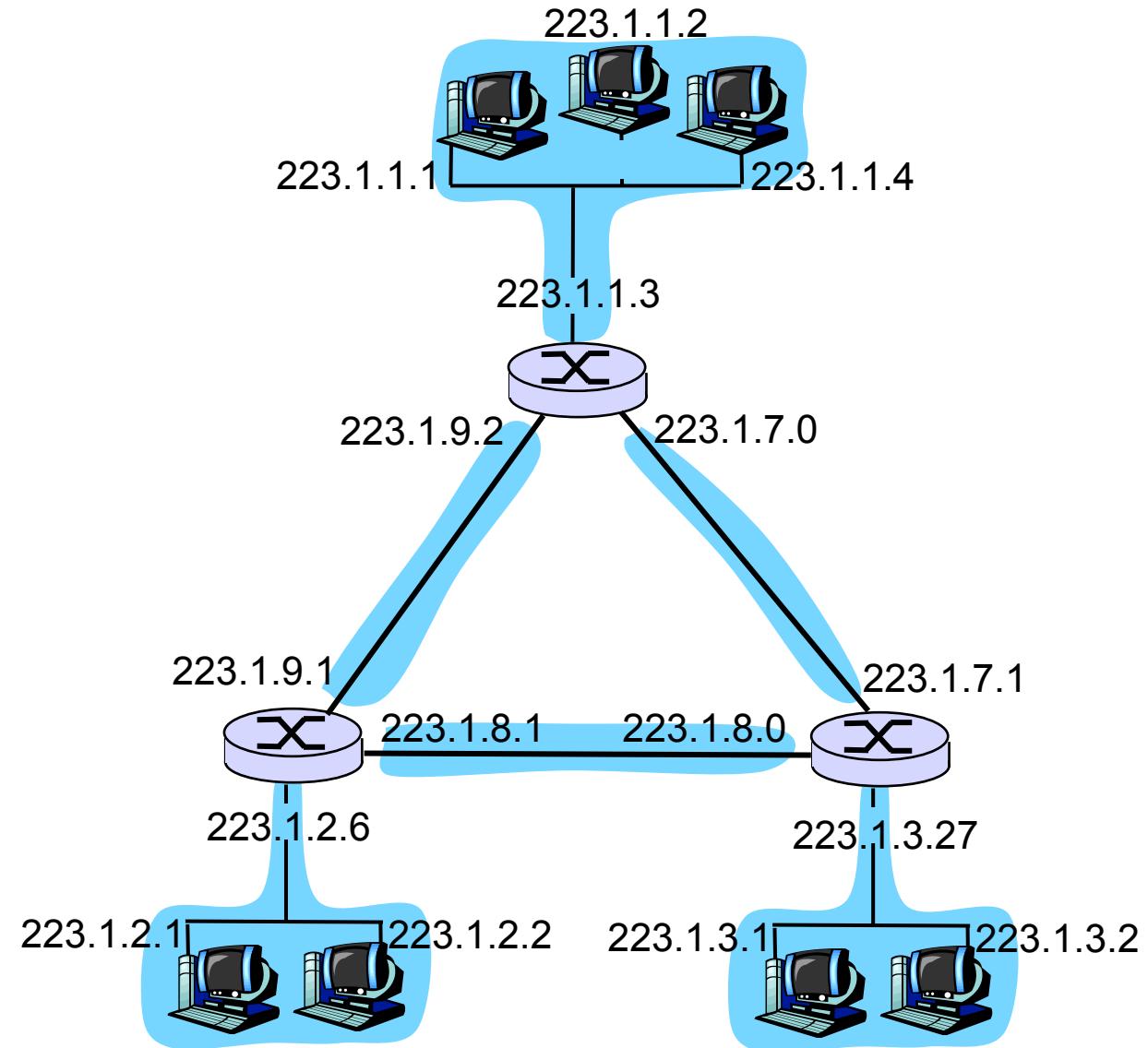
- To determine the subnets, detach each interface from its host or router, creating islands of isolated networks.
Each isolated network is called a **subnet**.



Subnet mask: /24

Subnets

How many?



IP addressing: CIDR

CIDR: Classless InterDomain Routing

- subnet portion of address of arbitrary length
- address format: $a.b.c.d/x$, where x is # bits in subnet portion of address



IP addresses: how to get one?

Q: How does a *host* get IP address?

- hard-coded by system admin in a file
 - Windows: control-panel->network->configuration->tcp/ip->properties
 - UNIX: /etc/rc.config
- **DHCP: Dynamic Host Configuration Protocol:** dynamically get address from server
 - “plug-and-play”

DHCP: Dynamic Host Configuration Protocol

Goal: allow host to *dynamically* obtain its IP address from network server when it joins network

Can renew its lease on address in use

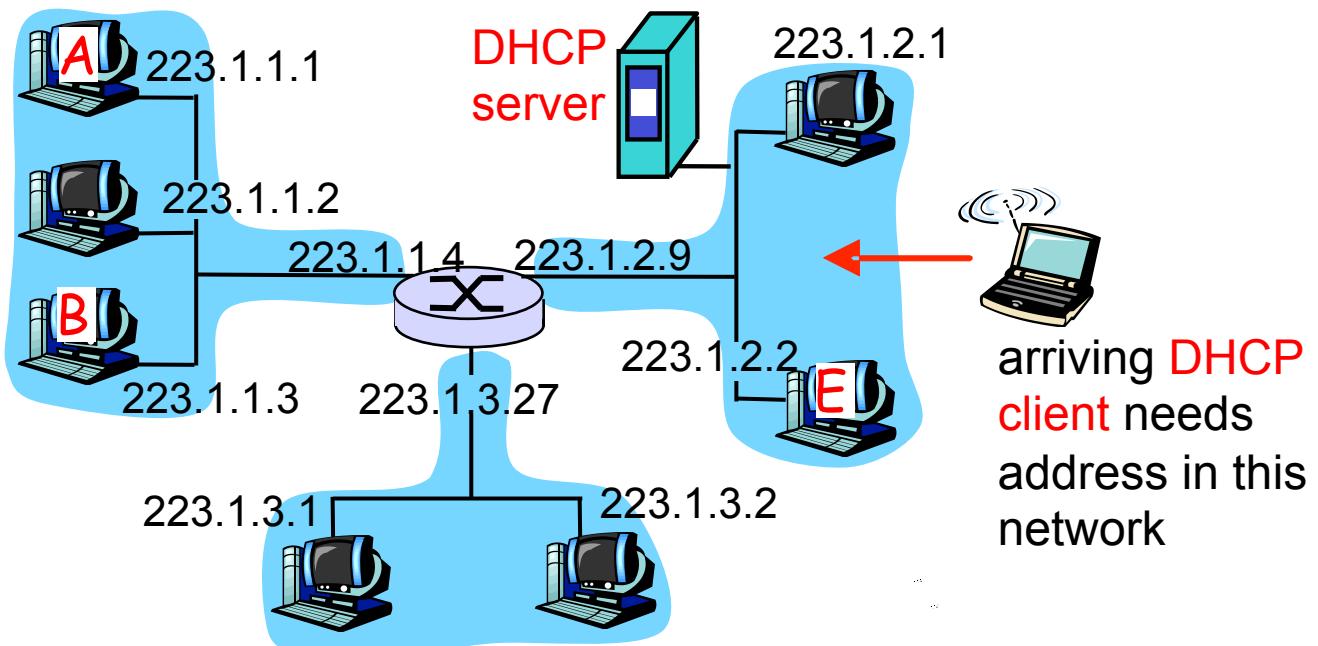
Allows reuse of addresses (only hold address while connected and “on”)

Support for mobile users who want to join network (more shortly)

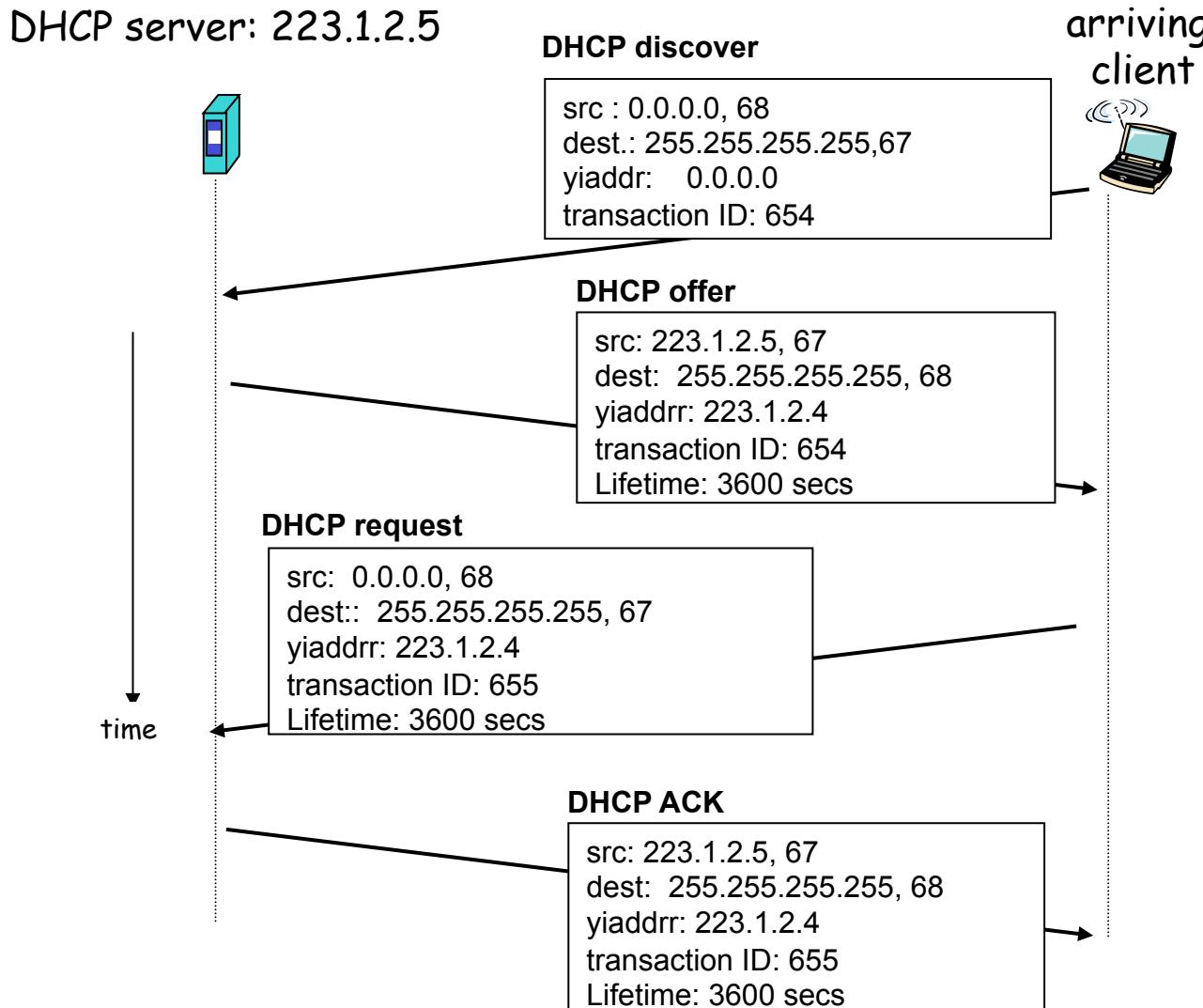
DHCP overview:

- host broadcasts “**DHCP discover**” msg [optional]
- DHCP server responds with “**DHCP offer**” msg [optional]
- host requests IP address: “**DHCP request**” msg
- DHCP server sends address: “**DHCP ack**” msg

DHCP client-server scenario



DHCP client-server scenario



DHCP

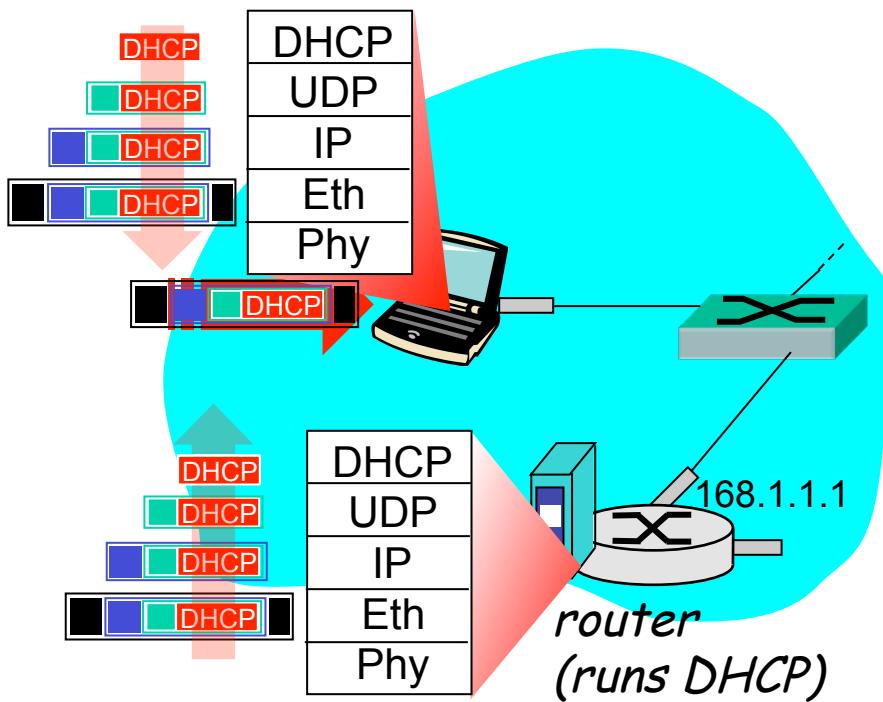
1. The client broadcasts a DHCPDISCOVER message on its local physical subnet. The DHCPDISCOVER message MAY include options that suggest values for the network address and lease duration. BOOTP relay agents may pass the message on to DHCP servers not on the same physical subnet.
2. Each server may respond with a DHCPOFFER message that includes an available network address in the 'yiaddr' field (and other configuration parameters in DHCP options). Servers need not reserve the offered network address, although the protocol will work more efficiently if the server avoids allocating the offered network address to another client. When allocating a new address, servers SHOULD check that the offered network address is not already in use; e.g., the server may probe the offered address with an ICMP Echo Request.

DHCP: more than IP address

DHCP can return more than just allocated IP address on subnet:

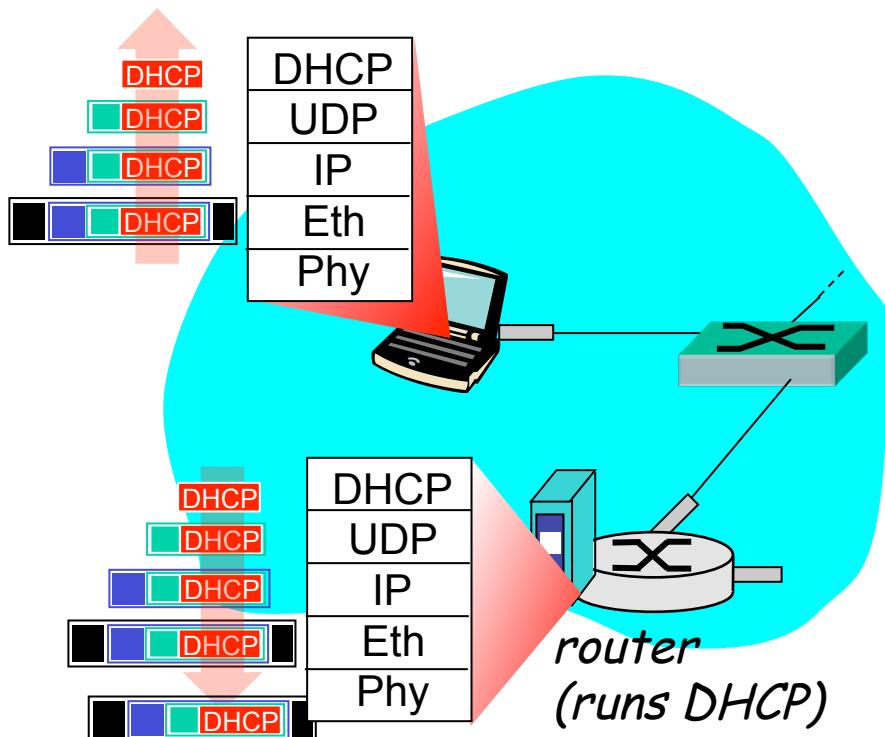
- address of first-hop router for client
- name and IP address of DNS sever
- network mask (indicating network versus host portion of address)

DHCP: example



- connecting laptop needs its IP address, addr of first-hop router, addr of DNS server: use DHCP
- DHCP request encapsulated in UDP, encapsulated in IP, encapsulated in 802.1 Ethernet
- Ethernet frame broadcast (dest: FFFFFFFFFFFF) on LAN, received at router running DHCP server
- Ethernet demux'ed to IP demux'ed, UDP demux'ed to DHCP

DHCP: example



- DCP server formulates DHCP ACK containing client's IP address, IP address of first-hop router for client, name & IP address of DNS server
- encapsulation of DHCP server, frame forwarded to client, demux' ing up to DHCP at client
- client now knows its IP address, name and IP address of DSN server, IP address of its first-hop router

DHCP

DHCPDISCOVER - Client broadcast to locate available servers.

DHCPOFFER - Server to client in response to DHCPDISCOVER with offer of configuration parameters.

DHCPREQUEST - Client message to servers either (a) requesting offered parameters from one server and implicitly declining offers from all others, (b) confirming correctness of previously allocated address after, e.g., system reboot, or (c) extending the lease on a particular network address.

DHCPACK - Server to client with configuration parameters, including committed network address.

DCHPNAK - Server to client indicating client's notion of network address is incorrect (e.g., client has moved to new subnet) or client's lease has expired

DHCPDECLINE - Client to server indicating network address is already in use.

DCHPRELEASE - Client to server relinquishing network address and cancelling remaining lease.

DHCPIINFORM - Client to server, asking only for local configuration parameters; client already has externally configured network address.

DHCP: wireshark output (home LAN)

Message type: **Boot Request (1)**

Hardware type: Ethernet

Hardware address length: 6

Hops: 0

Transaction ID: 0x6b3a11b7

Seconds elapsed: 0

Bootp flags: 0x0000 (Unicast)

Client IP address: 0.0.0.0 (0.0.0.0)

Your (client) IP address: 0.0.0.0 (0.0.0.0)

Next server IP address: 0.0.0.0 (0.0.0.0)

Relay agent IP address: 0.0.0.0 (0.0.0.0)

Client MAC address: Wistron_23:68:8a (00:16:d3:23:68:8a)

Server host name not given

Boot file name not given

Magic cookie: (OK)

Option: (t=53,l=1) **DHCP Message Type = DHCP Request**

Option: (61) Client identifier

Length: 7; Value: 010016D323688A;

Hardware type: Ethernet

Client MAC address: Wistron_23:68:8a (00:16:d3:23:68:8a)

Option: (t=50,l=4) Requested IP Address = 192.168.1.101

Option: (t=12,l=5) Host Name = "nomad"

Option: (55) Parameter Request List

Length: 11; Value: 010F03062C2E2F1F21F92B

1 = Subnet Mask; 15 = Domain Name

3 = Router; 6 = Domain Name Server

44 = NetBIOS over TCP/IP Name Server

.....

request

reply

Message type: **Boot Reply (2)**

Hardware type: Ethernet

Hardware address length: 6

Hops: 0

Transaction ID: 0x6b3a11b7

Seconds elapsed: 0

Bootp flags: 0x0000 (Unicast)

Client IP address: 192.168.1.101 (192.168.1.101)

Your (client) IP address: 0.0.0.0 (0.0.0.0)

Next server IP address: 192.168.1.1 (192.168.1.1)

Relay agent IP address: 0.0.0.0 (0.0.0.0)

Client MAC address: Wistron_23:68:8a (00:16:d3:23:68:8a)

Server host name not given

Boot file name not given

Magic cookie: (OK)

Option: (t=53,l=1) DHCP Message Type = DHCP ACK

Option: (t=54,l=4) Server Identifier = 192.168.1.1

Option: (t=1,l=4) Subnet Mask = 255.255.255.0

Option: (t=3,l=4) Router = 192.168.1.1

Option: (6) Domain Name Server

Length: 12; Value: 445747E2445749F244574092;

IP Address: 68.87.71.226;

IP Address: 68.87.73.242;

IP Address: 68.87.64.146

Option: (t=15,l=20) Domain Name = "hsd1.ma.comcast.net."

IP addresses: how to get one?

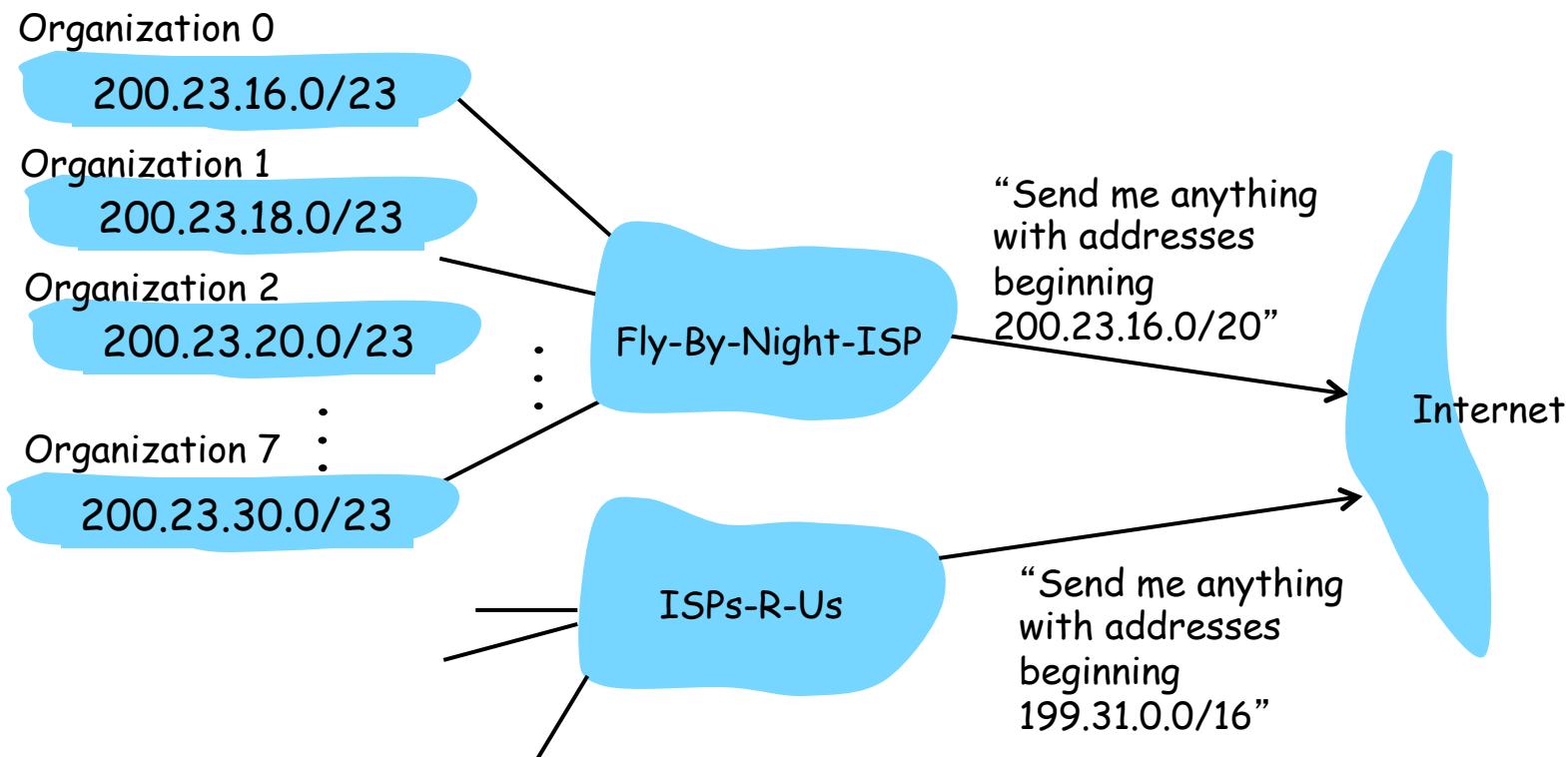
Q: How does *network* get subnet part of IP addr?

A: gets allocated portion of its provider ISP's address space

ISP's block	<u>11001000</u>	<u>00010111</u>	<u>00010000</u>	<u>00000000</u>	200.23.16.0/20
Organization 0	<u>11001000</u>	<u>00010111</u>	<u>00010000</u>	<u>00000000</u>	200.23.16.0/23
Organization 1	<u>11001000</u>	<u>00010111</u>	<u>00010010</u>	<u>00000000</u>	200.23.18.0/23
Organization 2	<u>11001000</u>	<u>00010111</u>	<u>00010100</u>	<u>00000000</u>	200.23.20.0/23
...
Organization 7	<u>11001000</u>	<u>00010111</u>	<u>00011110</u>	<u>00000000</u>	200.23.30.0/23

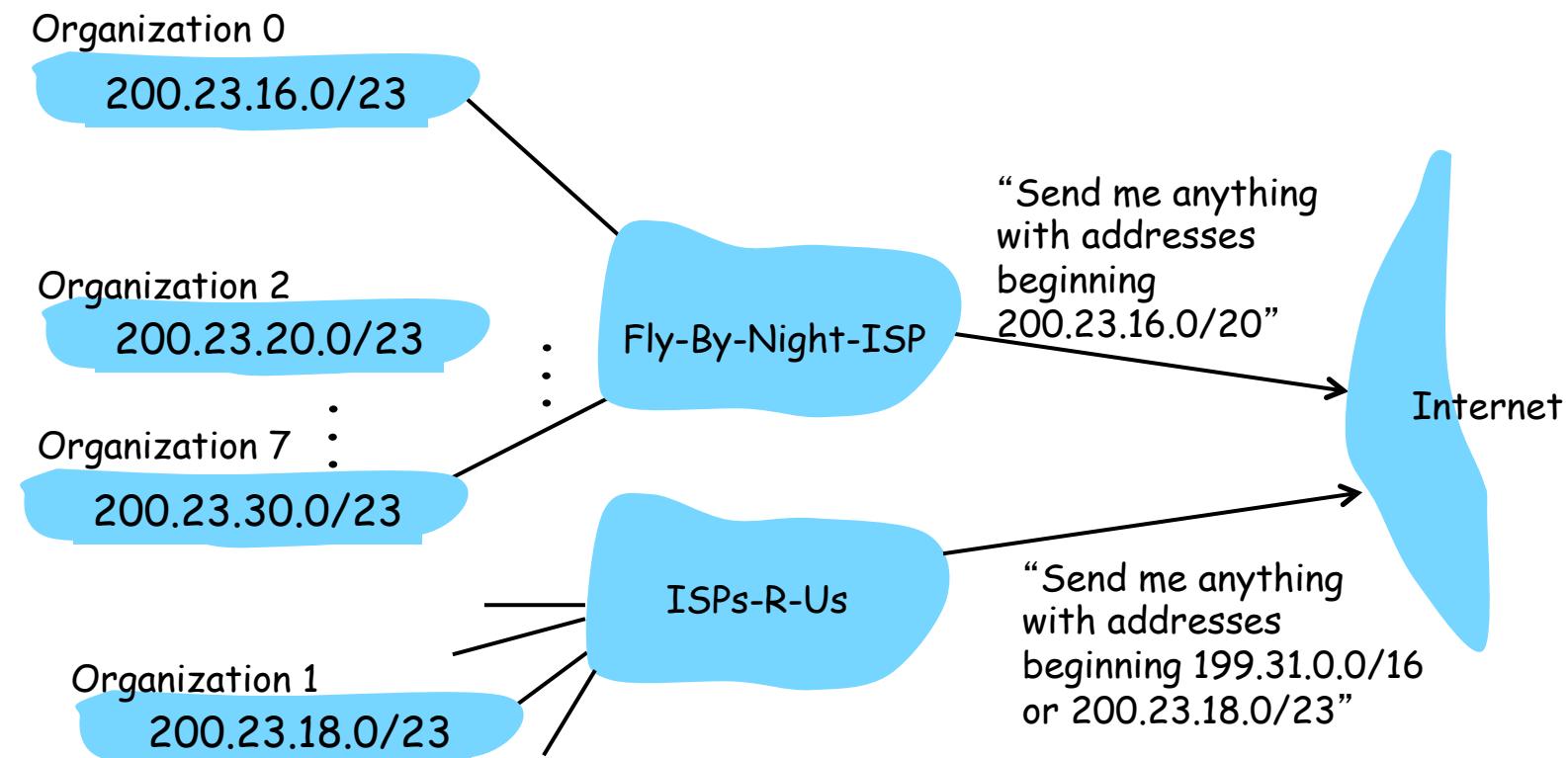
Hierarchical addressing: route aggregation

Hierarchical addressing allows efficient advertisement of routing information:



Hierarchical addressing: more specific routes

ISPs-R-Us has a more specific route to Organization 1



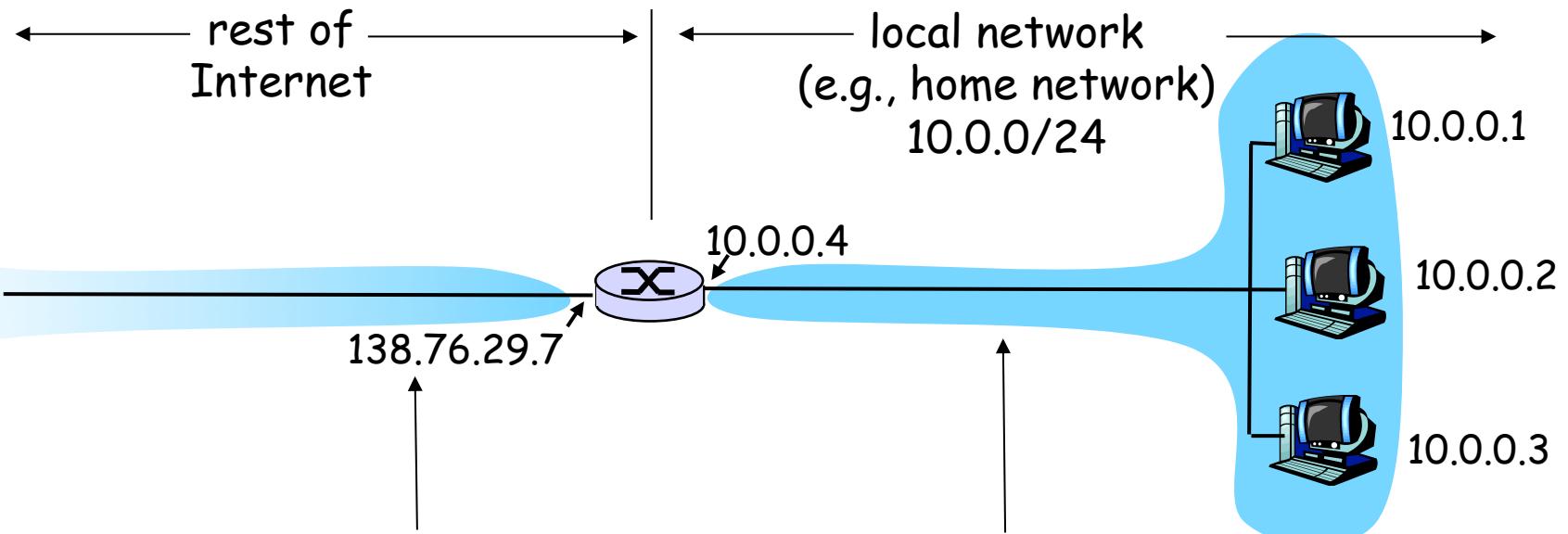
IP addressing: the last word...

Q: How does an ISP get block of addresses?

A: ICANN: Internet Corporation for Assigned
Names and Numbers

- allocates addresses
- manages DNS
- assigns domain names, resolves disputes

NAT: Network Address Translation



All datagrams *leaving* local network have *same* single source NAT IP address: 138.76.29.7, different source port numbers

Datagrams with source or destination in this network have 10.0.0/24 address for source, destination (as usual)

NAT: Network Address Translation

- **Motivation:** local network uses just one IP address as far as outside world is concerned:
 - range of addresses not needed from ISP: just one IP address 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).

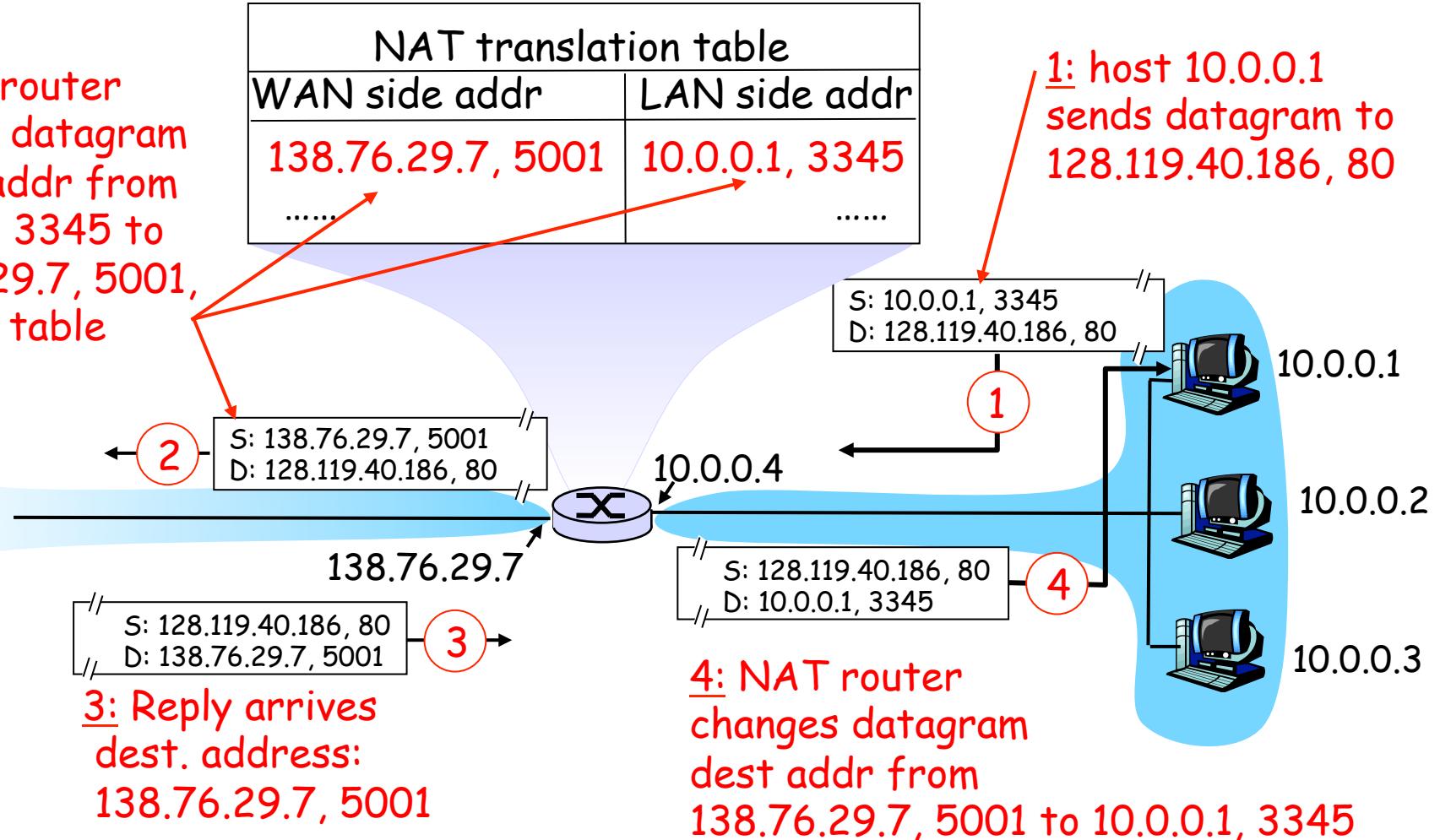
NAT: Network Address Translation

Implementation: NAT router must:

- *outgoing datagrams: replace* (source IP address, port #) of every outgoing datagram to (NAT IP address, new port #)
 - ... remote clients/servers will respond using (NAT IP address, new port #) as destination addr.
- *remember (in NAT translation table)* every (source IP address, port #) to (NAT IP address, new port #) translation pair
- *incoming datagrams: replace* (NAT IP address, new port #) in dest fields of every incoming datagram with corresponding (source IP address, port #) stored in NAT table

NAT: Network Address Translation

2: NAT router changes datagram source addr from 10.0.0.1, 3345 to 138.76.29.7, 5001, updates table

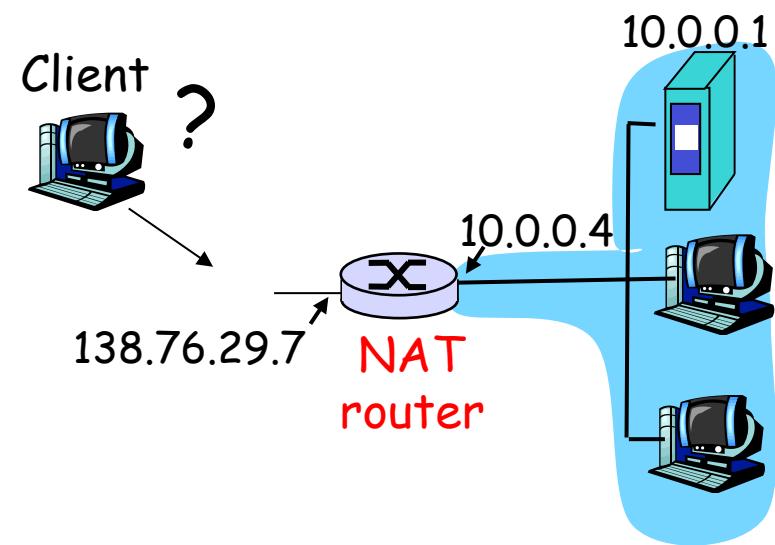


NAT: Network Address Translation

- 16-bit port-number field:
 - 60,000 simultaneous connections with a single LAN-side address!
- NAT is controversial:
 - routers should only process up to layer 3
 - violates end-to-end argument
 - NAT possibility must be taken into account by app designers, eg, P2P applications
 - address shortage should instead be solved by IPv6

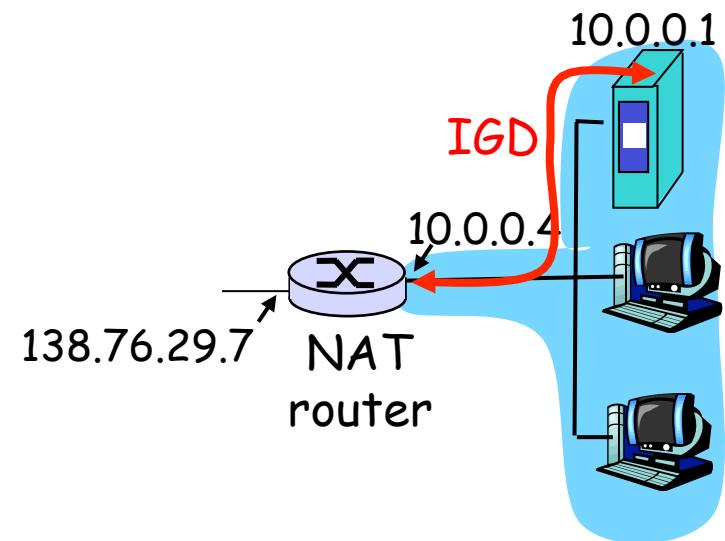
NAT traversal problem

- client wants to connect to server with address 10.0.0.1
 - server address 10.0.0.1 local to LAN (client can't use it as destination addr)
 - only one externally visible NATted address: 138.76.29.7
- solution 1: statically configure NAT to forward incoming connection requests at given port to server
 - e.g., (123.76.29.7, port 2500) always forwarded to 10.0.0.1 port 25000



NAT traversal problem

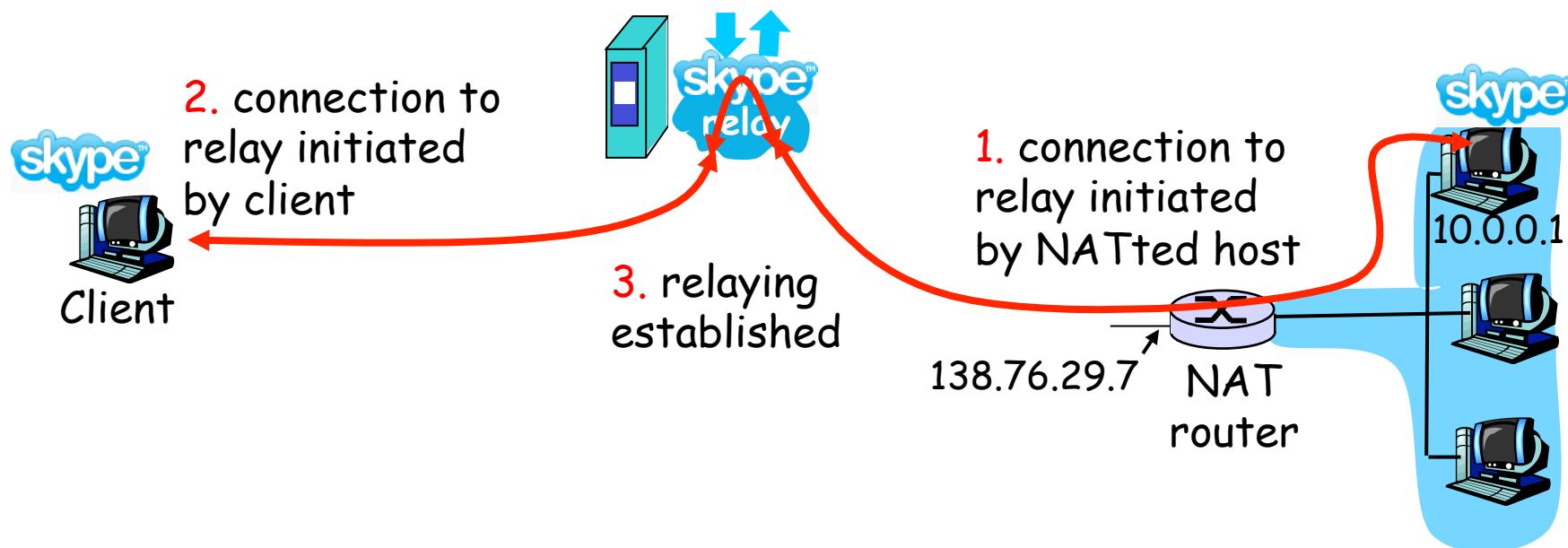
- solution 2: Universal Plug and Play (UPnP) Internet Gateway Device (IGD) Protocol. Allows NATted host to:
 - ❖ learn public IP address (138.76.29.7)
 - ❖ add/remove port mappings (with lease times)



i.e., automate static NAT port map configuration

NAT traversal problem

- solution 3: relaying (used in Skype)
 - NATed client establishes connection to relay
 - External client connects to relay
 - relay bridges packets between two connections



Chapter 4: Network Layer

- 4.1 Introduction
- 4.2 Virtual circuit and datagram networks
- 4.3 What's inside a router
- 4.4 IP: Internet Protocol
 - Datagram format
 - IPv4 addressing
 - ICMP
 - IPv6
- 4.5 Routing algorithms
 - Link state
 - Distance Vector
 - Hierarchical routing
- 4.6 Routing in the Internet
 - RIP
 - OSPF
 - BGP
- 4.7 Broadcast and multicast routing

ICMP: Internet Control Message Protocol

- used by hosts & routers to communicate network-level information
 - error reporting: unreachable host, network, port, protocol
 - echo request/reply (used by ping)
- network-layer “above” IP:
 - ICMP msgs carried in IP datagrams
- **ICMP message:** type, code plus first 8 bytes of IP datagram causing error

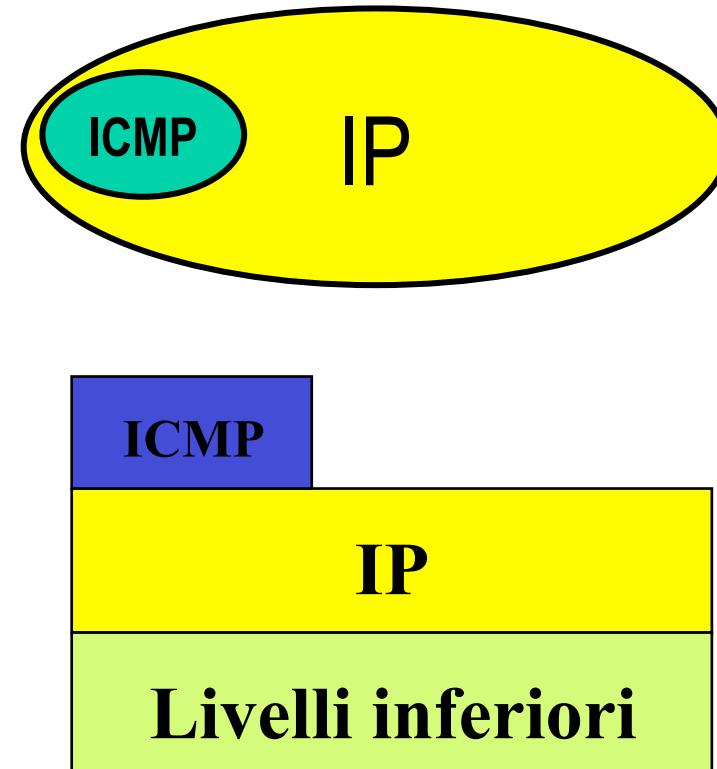
Type	Code	description
0	0	echo reply (ping)
3	0	dest. network unreachable
3	1	dest host unreachable
3	2	dest protocol unreachable
3	3	dest port unreachable
3	6	dest network unknown
3	7	dest host unknown
4	0	source quench (congestion control - not used)
8	0	echo request (ping)
9	0	router advertisement
10	0	router discovery
11	0	TTL expired
12	0	bad IP header

Traceroute and ICMP

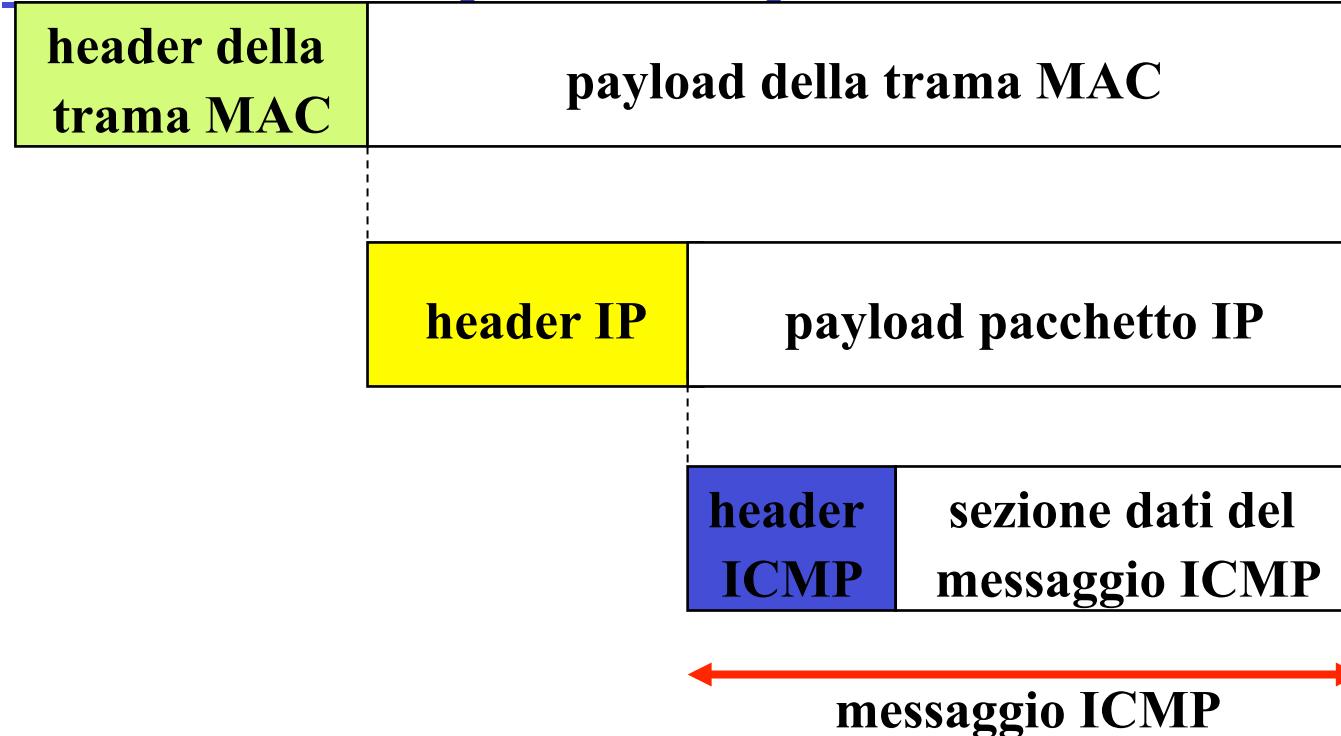
- Source sends series of UDP segments to dest
 - First has TTL =1
 - Second has TTL=2, etc.
 - Unlikely port number
 - When nth datagram arrives to nth router:
 - Router discards datagram
 - And sends to source an ICMP message (type 11, code 0)
 - Message includes name of router& IP address
 - When ICMP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

Internet Control Message Protocol (ICMP)

- E' un protocollo per messaggi di servizio fra host e router per informazioni su errori e fasi di attraversamento della rete
- da questo punto di vista può essere considerato come parte di IP
- i messaggi ICMP sono incapsulati e trasportati da IP, e quindi da questo punto di vista può essere considerato un utente di IP



Internet Control Message Protocol (ICMP)



- Nel pacchetto IP il campo protocol indica il codice dell'ICMP
- il messaggio ICMP viaggia all'interno del pacchetto IP

Messaggi ICMP

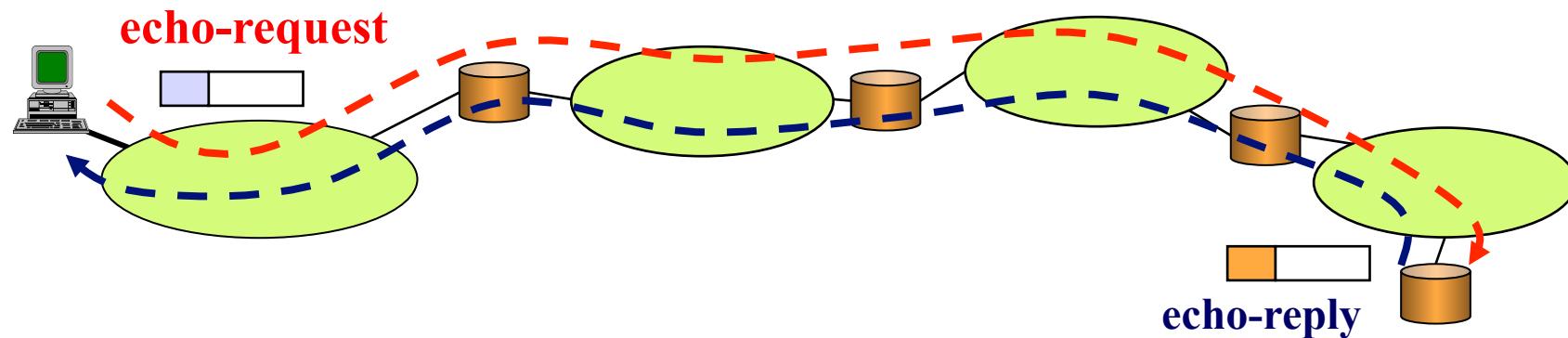
Da informazioni sul codice
Dell'errore (info piu' dettagliate
Sull'errore verificatosi)



Type	Type
0 Echo reply	11 Parameter problem
3 Destination unreachable	13 Timestamp request
4 Source Quench	14 Timestamp reply
5 Redirect (change a route)	17 Address mask request
8 Echo request	18 Address mask reply
11 Time exceeded	

Echo

- I messaggi di Echo-request e Echo-reply sono usati per verificare la raggiungibilità e lo stato di un host o un router
- quando un nodo IP riceve un messaggio di Echo-request risponda immediatamente con un messaggio di Echo reply

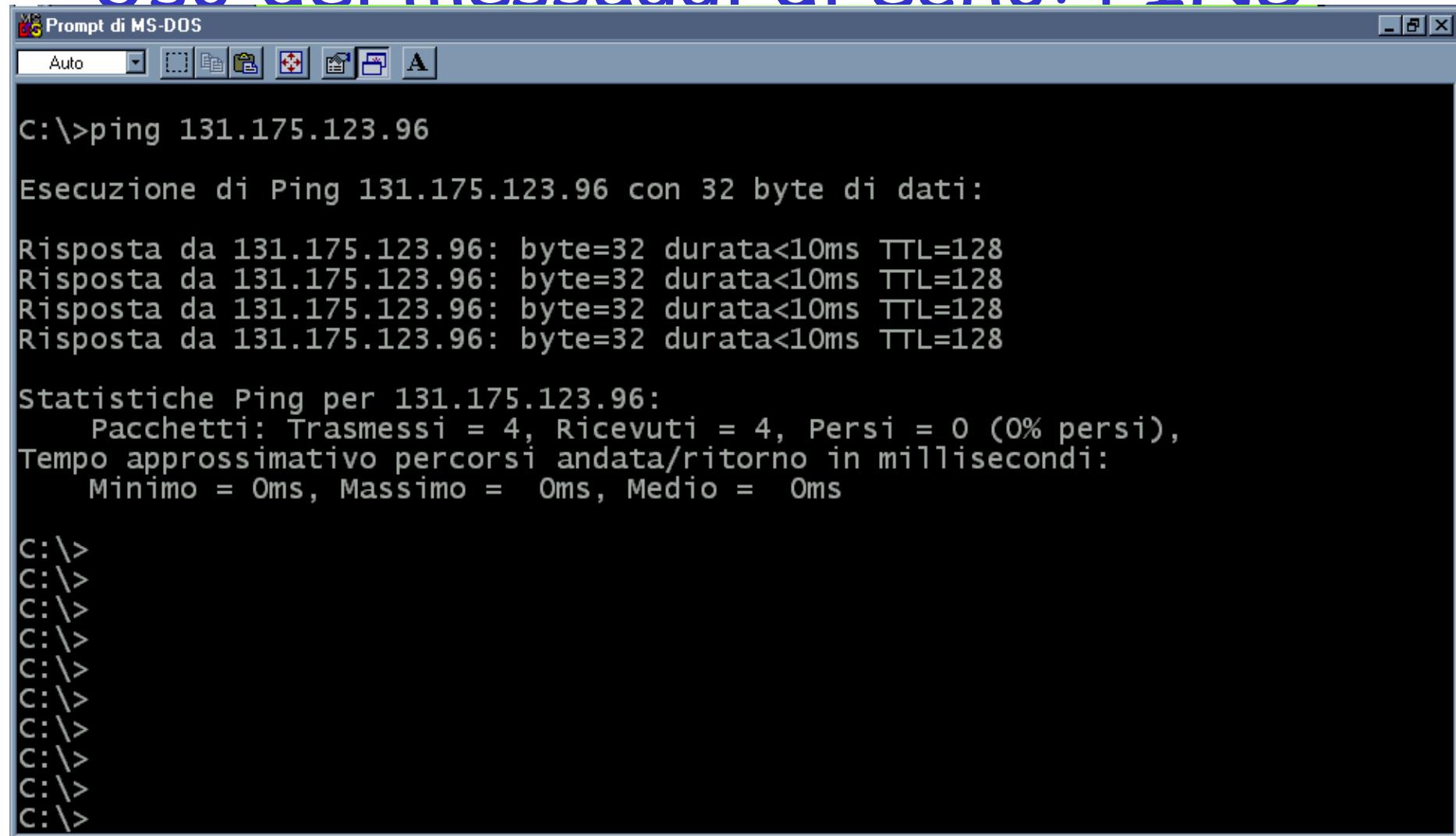


Echo

type (8 request, 0 reply)	code (0)	checksum
identifier	sequence number	
optional data		

- Il campo identifier viene scelto dal mittente della richiesta
- nella risposta viene ripetuto lo stesso identifier della richiesta
- più richieste consecutive possono avere lo stesso identifier e differire per il sequence number
- una sequenza arbitraria può essere aggiunta dal mittente nel campo optional data e deve essere riportata uguale nella risposta

Uso dei messaggi di echo: PING



A screenshot of a Microsoft DOS window titled "Prompt di MS-DOS". The window shows the command "ping 131.175.123.96" being run, followed by the output of four successful ping responses and their statistics.

```
C:\>ping 131.175.123.96

Esecuzione di Ping 131.175.123.96 con 32 byte di dati:

Risposta da 131.175.123.96: byte=32 durata<10ms TTL=128

Statistiche Ping per 131.175.123.96:
  Pacchetti: Trasmessi = 4, Ricevuti = 4, Persi = 0 (0% persi),
  Tempo approssimativo percorsi andata/ritorno in millisecondi:
    Minimo = 0ms, Massimo = 0ms, Medio = 0ms

C:\>
```

Destination unreachable

type (3)	code (0-12)	checksum
non usato (0)		
header + primi 64 bit del pacchetto IP che ha causato il problema		

- Quando un router scarta un pacchetto per qualche motivo normalmente genera un messaggio di errore che invia alla sorgente del pacchetto
- nel campo code è codificato il motivo che ha causato l'errore
- ovviamente la generazione del messaggio avviene solo nei casi in cui il router può accorgersi del problema
- il motivo più comune è il fatto che la destinazione non è presente nella tabella di routing. Un messaggio destination unreachable può anche essere inviato se è necessario frammentare un pacchetto ed il flag DF è a 1.

Destination unreachable

type (3)	code (0-12)	checksum
non usato (0)		
header + primi 64 bit del pacchetto IP che ha causato il problema		

Alcuni Code:

- 0 network unreachable
- 1 host unreachable
- 2 protocol unreachable
- 3 port unreachable
- 4 fragmentation needed and DF set
- 5 source route failed

...

Time exceeded

type (11)	code (0-1)	checksum
non usato (0)		
header + primi 64 bit del pacchetto IP che ha causato il problema		

Code 0

- Il messaggio di time exceeded viene usato quando il router decrementando il TTL lo pone a 0
- il messaggio di time exceeded viene inviato alla sorgente del pacchetto

Code 1

- viene usato dalla destinazione quando non tutti i frammenti di un pacchetto arrivano entro un tempo massimo

Parameter problem

type (12)	code (0-1)	checksum
pointer		non usato (0)
header + primi 64 bit del pacchetto IP che ha causato il problema		

Punta al Byte del datagramma IP originale dove si è verificato il problema

Code 0

- se l'header di un pacchetto IP ha una incongruenza in qualcuno dei suoi campi viene inviato il messaggio di parameter problem; il campo pointer punta al byte del pacchetto che ha causato il problema

Code 1

- viene usato quando un'opzione non è implementata e non può essere soddisfatta

Timestamp request e reply

type (13 request, 14 reply)	code (0)	checksum
identifier		sequence number
originate timestamp		
receive timestamp		
transmit timestamp		

- Questo messaggio viene usato per scambiarsi informazioni sul clock di sorgente e destinazione
- originate timestamp*: viene riempito dalla sorgente
- receive timestamp*: viene riempito dalla destinazione appena ricevuto il pacchetto
- transmit timestamp*: viene riempito dalla destinazione immediatamente prima di inviare il pacchetto di risposta

Address mask request e reply

type (17 request, 18 reply)	code (0)	checksum
identifier	sequence number	
address mask		

- Questo messaggio viene usato per conoscere la netmask di un host/router
- Il campo address mask viene riempito da invia la risposta

Chapter 4: Network Layer

- 4.1 Introduction
- 4.2 Virtual circuit and datagram networks
- 4.3 What's inside a router
- 4.4 IP: Internet Protocol
 - Datagram format
 - IPv4 addressing
 - ICMP
 - IPv6
- 4.5 Routing algorithms
 - Link state
 - Distance Vector
 - Hierarchical routing
- 4.6 Routing in the Internet
 - RIP
 - OSPF
 - BGP
- 4.7 Broadcast and multicast routing

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
 - no fragmentation allowed

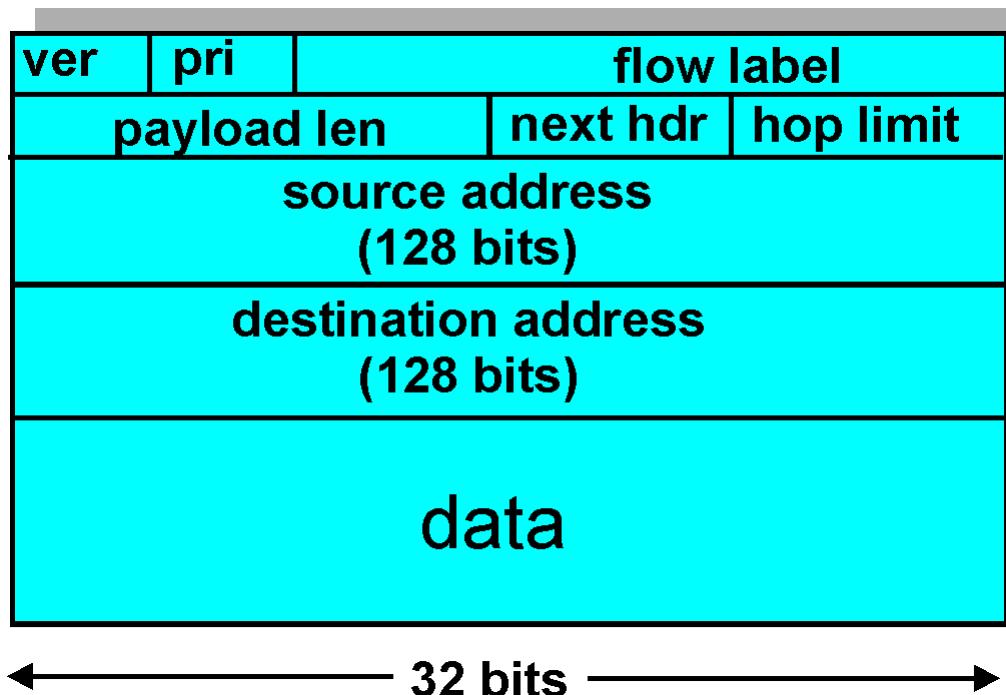
IPv6 Header (Cont)

Priority: identify priority among datagrams in flow

Flow Label: identify datagrams in same “flow.”

(concept of “flow” not well defined).

Next header: identify upper layer protocol for data



Other Changes from IPv4

- **Checksum:** removed entirely to reduce processing time at each hop
- **Options:** allowed, but outside of header, indicated by “Next Header” field
- **ICMPv6:** new version of ICMP
 - additional message types, e.g. “Packet Too Big”
 - multicast group management functions

Transition From IPv4 To IPv6

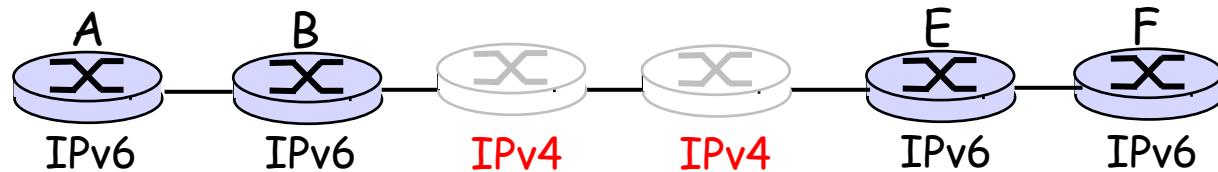
- Not all routers can be upgraded simultaneous
 - no “flag days”
 - How will the network operate with mixed IPv4 and IPv6 routers?
- *Tunneling:* IPv6 carried as payload in IPv4 datagram among IPv4 routers

Tunneling

Logical view:



Physical view:



Tunneling

Logical view:



Physical view:

