Romain Fontugne

2019 Fall Semester

Information Network Systems

The Network Layer (1)

Last lecture:

The link layer

Transfer datagram from one node to physically adjacent node over a link

- Link layer services
- Error detection, correction
- Multiple access protocols
- Ethernet
- Switches
- Data center networking

IP Stack

Application

Transport

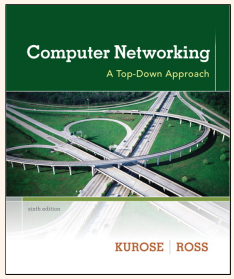
Network

Link

Physical

Today's Lecture: Network layer

1. Introduction, services
2. The Internet network layer
   - IPv4
   - IPv6
   - ICMP

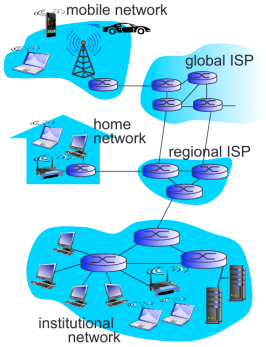


- Chapter 4

Network layer: Introduction

Terminology

- Datagram: layer-3 packet encapsulating segment (layer-4 packet)
- Router: device that forwards datagram between networks



Two key network-layer functions

Forwarding

- Move packets from router's input to appropriate router output (Forwarding table)

Routing

- Determine route taken by packets from source to destination (Routing algorithm)

Analogy:

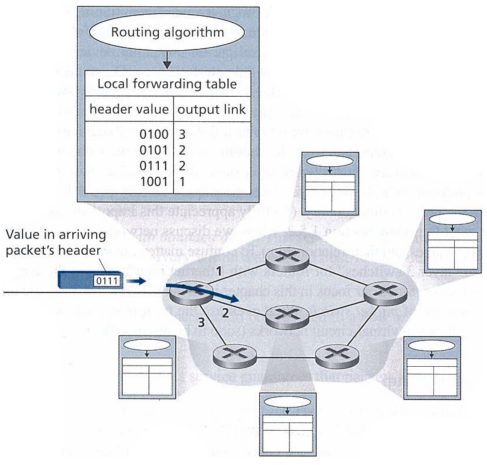
Routing

- Process of planning trip from source to destination

Forwarding

- Process of getting through single interchange

Forwarding and Routing



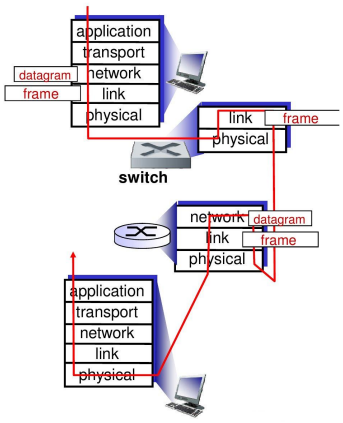
Switches vs. routers

Both are store-and-forward:

- Routers: network-layer devices (examine network-layer headers)
- Switches: link-layer devices (examine link-layer headers)

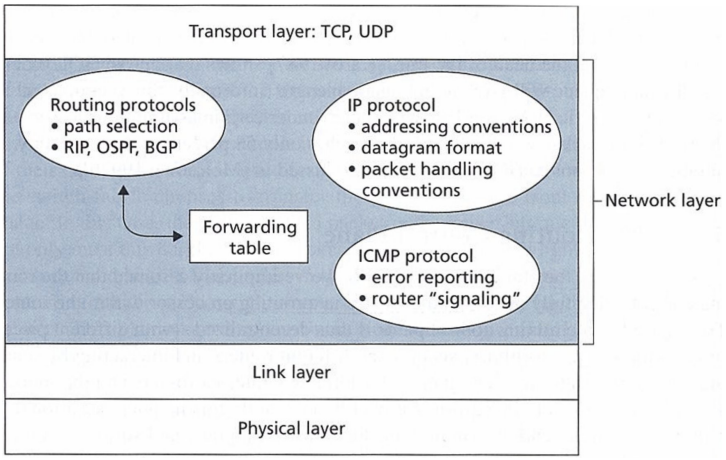
Both have forwarding tables:

- Routers: compute tables using **routing algorithms**, IP addresses
- Switches: learn forwarding table using **flooding**, learning, MAC addresses



The Internet network layer

Host/router network layer functions:



Network layer services

Possible services the network layer could provide:

- **Guaranteed delivery**: Packets will eventually arrive at destination
- **In-order packet delivery**: Packets will arrive in order sent
- **Guaranteed minimal bandwidth**: Data arrives at least at a minimal bandwidth
- **Guaranteed maximum jitter**: The time offset between packets should be similar at the transmitter and receiver
- **Security services**: For instance, encryption of data between source and destination

The Internet's network layer provides only best-effort service

- No guarantees....
- Simple protocols
- Can run on any link layer

The Internet Protocol (IP)

The Internet protocol defines:

- Addressing conventions
- Datagram format
- Packet handling

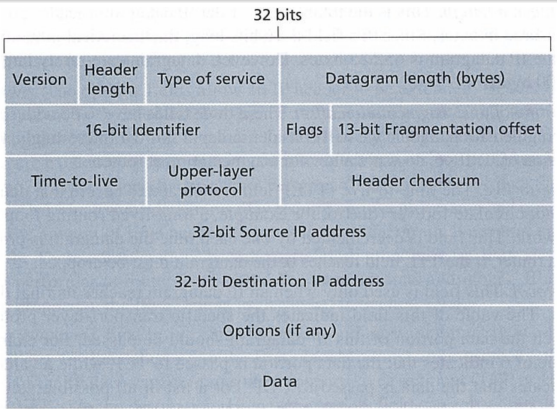
Currently, we are (slowly) migrating from the previous IP protocol (IPv4) to the new standard (IPv6)

- IPv4 carries the vast majority of Internet traffic

The IPv4 protocol

The IPv4 datagram contains 13 key fields

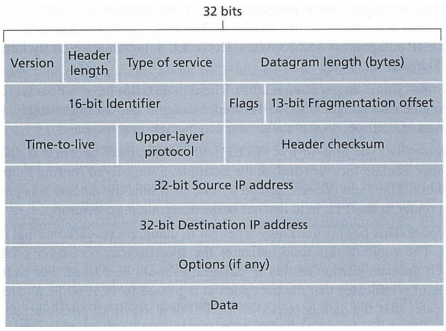
- Overhead: at least 20 bytes



The IPv4 header

Version number (4 bits)

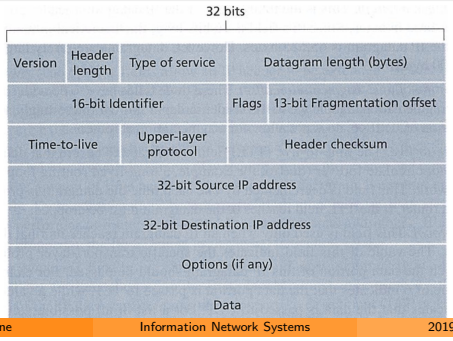
- Specify IP protocol version
- Helps the router to correctly interpret a received datagram



The IPv4 header

Header length (4 bits)

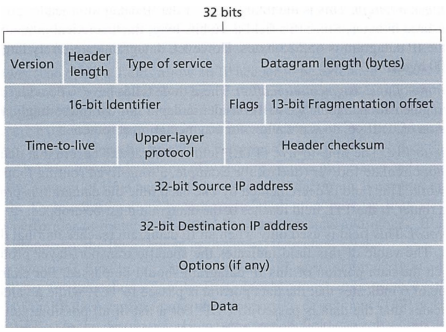
- Since datagram can contain variable number of options, this field specifies when the data begins
- Without options (most common) an IPv4 datagram header is 20 bytes long



The IPv4 header

Type of service (8 bits)

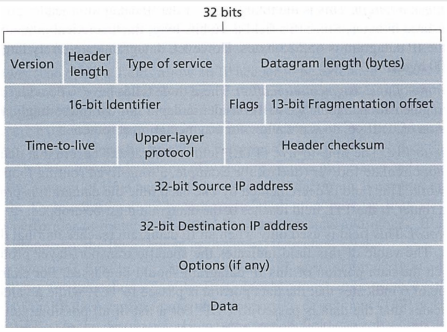
- specifies different types of IP datagrams (low delay, high throughput, reliability)



The IPv4 header

Datagram length (16 bits)

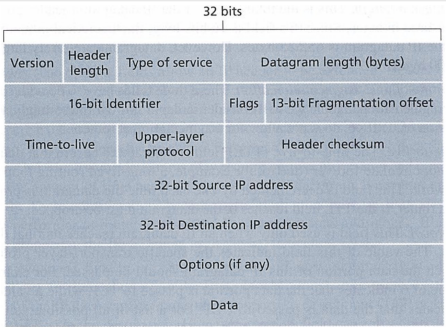
- Total length of the IP datagram (header plus data) measured in bytes
- Theoretical maximum size of IP datagram is 65535 bytes but datagrams are rarely larger than 1500 bytes



The IPv4 header

Identifier, flags, fragmentation offset (16 + 3 + 13 bits)

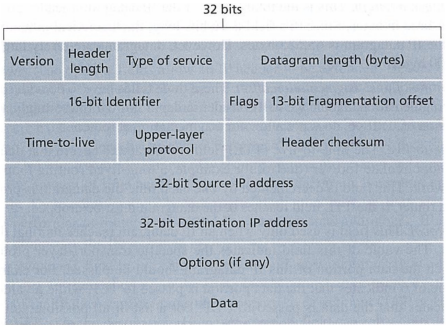
- Information for IP fragmentation
- No longer included in IPv6
- More on fragmentation shortly...



The IPv4 header

Time-to-live (TTL) (8 bits)

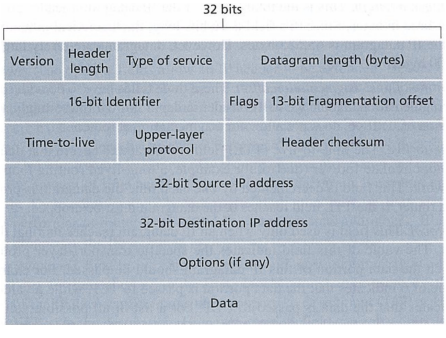
- Ensures that datagrams do not circulate forever
- Decremented by one each time the datagram is processed by a router
- If the TTL field reaches 0, the datagram is dropped



The IPv4 header

Protocol (8 bits)

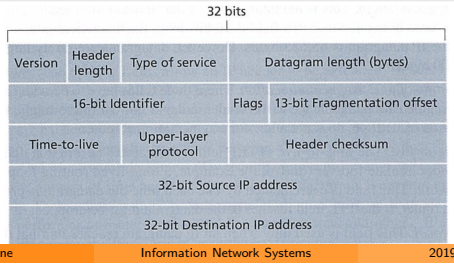
- Used only when datagram reaches its final destination
- Indicates the specific transport-layer protocol to which it should be passed



The IPv4 header

Header checksum (16 bits)

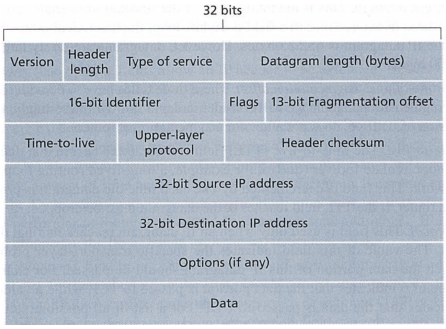
- Detect bit errors in a received datagram **header** (only the header!)
- Computed by treating each 2 bytes in the header as a number and summing these numbers using 1s complement
- If an error is detected, the datagram is discarded
- Routers must adjust the checksum when changing the IP header (e.g. decrementing the TTL)



The IPv4 header

Source and destination IP addresses (32 bits each)

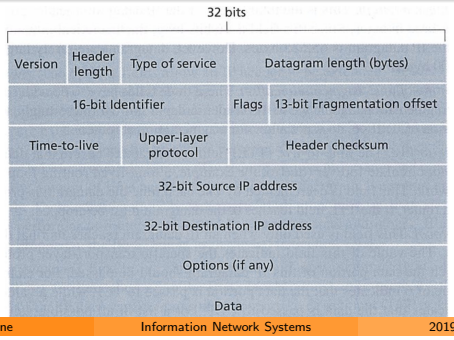
- Addresses of source and destination hosts
- Utilised for forwarding and routing through a network



The IPv4 header

Options (variable length)

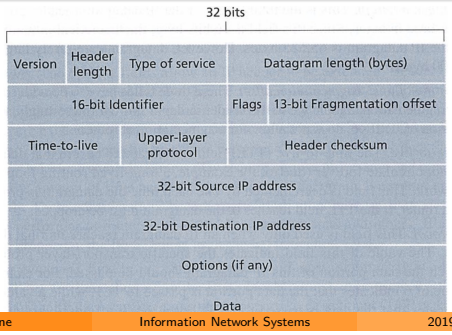
- Enable the specification of specific options
- Rarely used and merely add overhead to the processing of an IP packet
- Not included in IPv6



The IPv4 payload

Data (aka payload) (variable length)

- Contains the data to be transmitted
- Most often this field contains the transport-layer segment (e.g. TCP or UDP)
- Can also carry other kinds of messages such as ICMP



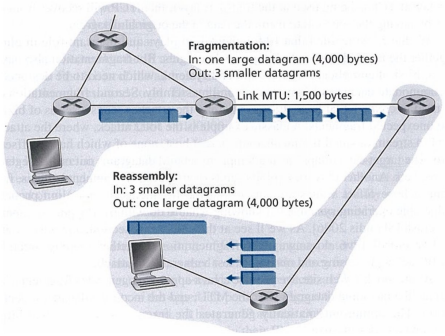
IPv4 fragmentation, reassembly

Link-layer Maximum Transmission Unit (MTU)

- Underlying frames have a maximum size ($\neq$ link types, $\neq$ MTUs)

Large IP datagram are divided ("fragmented") within network

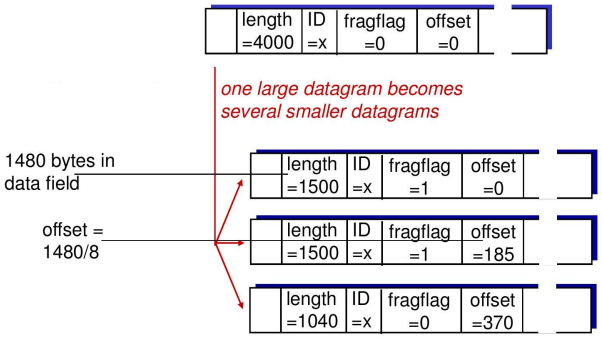
- "reassembled" only at final destination
- IP header bits used to identify, order related fragments



IPv4 fragmentation, reassembly

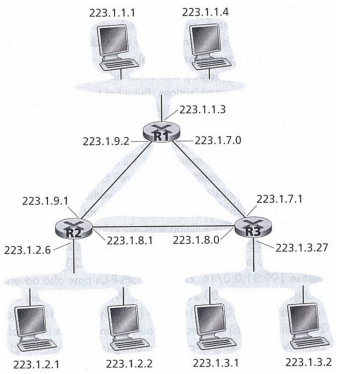
Example:

- Transfer a 4000-byte datagram
- MTU = 1500 bytes



IP addressing: introduction

- IP address: 32-bit identifier for host, router interface
- Interface: connection between host/router and physical link
  - routers typically have multiple interfaces
  - host typically has one or two interfaces (e.g., wired Ethernet, wireless 802.11)
- IP addresses associated with each interface



Notation:

Each byte written in decimal form separated by period: 223.1.1.1
Binary notation: 11011111 00000001 00000001 00000001

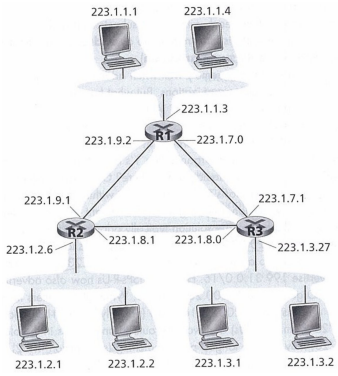
IP addressing: Subnets

IP address: 2 parts

- subnet part - high order bits
- host part - low order bits

Subnet

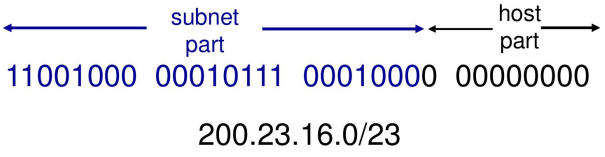
- device interfaces with same subnet part of IP address
- can physically reach each other without intervening router
- namely, if you remove routers the isolated networks are subnets
- how many subnets? how many devices in a subnet??



IP addressing: CIDR

CIDR: Classless InterDomain Routing

- subnet portion of address of arbitrary length
- address format: a.b.c.d/x, where x is the number of bits for the subnet



200.23.16.0/23

IP addressing: CIDR

Notation	Addresses	Subnetmask decimal	Subnetmask binary
/0	4.294.967.296	0.0.0.0	00000000.00000000.00000000.00000000
/1	2.147.483.648	128.0.0.0	10000000.00000000.00000000.00000000
/2	1.073.741.824	192.0.0.0	11000000.00000000.00000000.00000000
/3	536.870.912	224.0.0.0	11100000.00000000.00000000.00000000
/4	268.435.456	240.0.0.0	11110000.00000000.00000000.00000000
/5	134.217.728	248.0.0.0	11111000.00000000.00000000.00000000
/6	67.108.864	252.0.0.0	11111100.00000000.00000000.00000000
/7	33.554.432	254.0.0.0	11111110.00000000.00000000.00000000
/8	16.777.216	255.0.0.0	11111111.00000000.00000000.00000000
/9	8.388.608	255.128.0.0	11111111.10000000.00000000.00000000
/10	4.194.304	255.192.0.0	11111111.11000000.00000000.00000000
/11	2.097.152	255.224.0.0	11111111.11100000.00000000.00000000
/12	1.048.576	255.240.0.0	11111111.11110000.00000000.00000000
/13	524.288	255.248.0.0	11111111.11111000.00000000.00000000
/14	262.144	255.252.0.0	11111111.11111100.00000000.00000000
/15	131.072	255.254.0.0	11111111.11111110.00000000.00000000
/16	65.536	255.255.0.0	11111111.11111111.00000000.00000000
/17	32.768	255.255.128.0	11111111.11111111.10000000.00000000
/18	16.384	255.255.192.0	11111111.11111111.11000000.00000000
/19	8.192	255.255.224.0	11111111.11111111.11100000.00000000
/20	4.096	255.255.240.0	11111111.11111111.11110000.00000000
/21	2.048	255.255.248.0	11111111.11111111.11111000.00000000
/22	1.024	255.255.252.0	11111111.11111111.11111100.00000000
/23	512	255.255.254.0	11111111.11111111.11111110.00000000
/24	256	255.255.255.0	11111111.11111111.11111111.00000000

IP addresses: how to get one?

Question: 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 a server ("plug-and-play")

DHCP: Dynamic Host Configuration Protocol

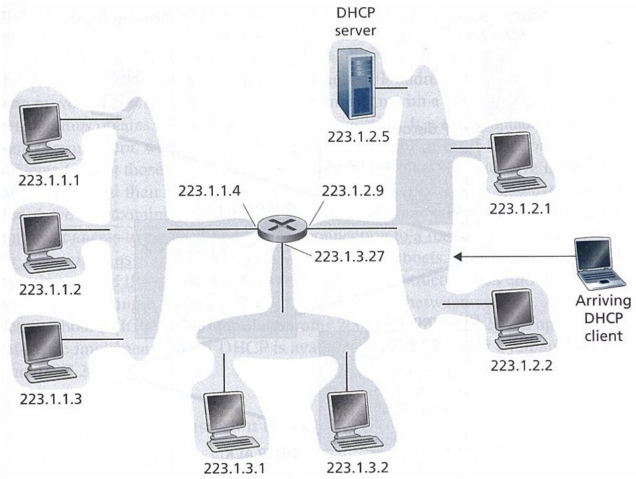
Goal: allow host to dynamically obtain its IP address from network server when it joins network

- can renew its lease on address in use
- allows reuse of addresses (only hold address while connected/"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

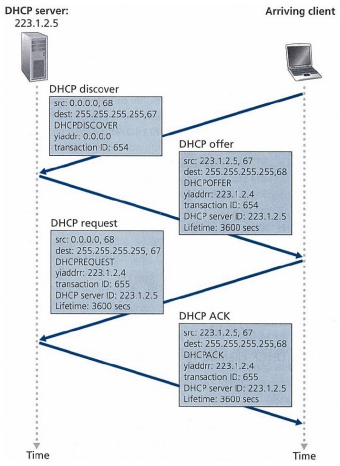
DHCP client-server scenario



DHCP client-server scenario

DHCP server discovery

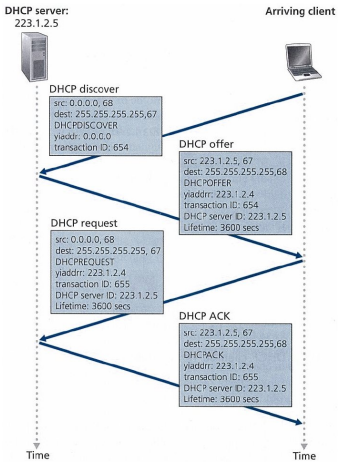
- Encapsulated in IP datagram to the broadcast address 255.255.255.255
- Source IP address set to 0.0.0.0



DHCP client-server scenario

DHCP server offer

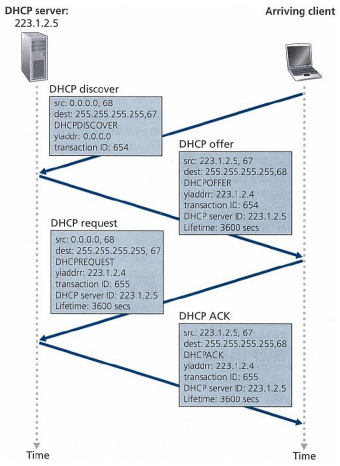
- DHCP server responds with a DHCP offer message
- Broadcast to all nodes (255.255.255.255) in an IP datagram
- Contains transaction ID, received discover message, proposed IP address, network mask and a lease time
- Lease time specifies the amount of time for which the IP address is valid



DHCP client-server scenario

DHCP request

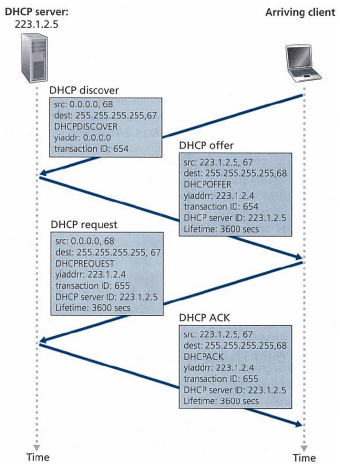
- Client chooses among one or more server offers and responds to the selected offer with a DHCP request message
- Message contains the proposed configuration parameters



DHCP client-server scenario

DHCP ACK

- Server responds and confirms parameter by sending a DHCP ACK message



DHCP: more than IP addresses

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

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

IP addresses: how to get one?

Question: how does network (e.g. DHCP server) get subnet part of IP address?

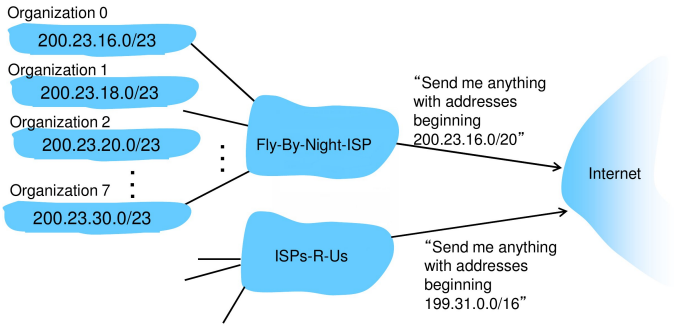
- Answer: Gets allocated portion of its provider ISP's address space

Example

ISP's block	11001000 00010111 00010000 00000000	200.23.16.0/20
Organization 0	11001000 00010111 0001**000**0 00000000	200.23.16.0/23
Organization 1	11001000 00010111 0001**001**0 00000000	200.23.18.0/23
Organization 2	11001000 00010111 0001**010**0 00000000	200.23.20.0/23
...	...	...
Organization 7	11001000 00010111 0001**111**0 00000000	200.23.30.0/23

Hierarchical addressing: route aggregation

Hierarchical addressing allows efficient advertisement of routing information:



IP addressing: the last word...

How does an ISP get block of addresses?

- ICANN: Internet Corporation for Assigned Names and Numbers
- Nonprofit organization
- Allocates IP addresses to Internet Service Providers (ISP)
- Manages DNS
- Assigns domain names, resolves disputes

NAT: Network Address Translation

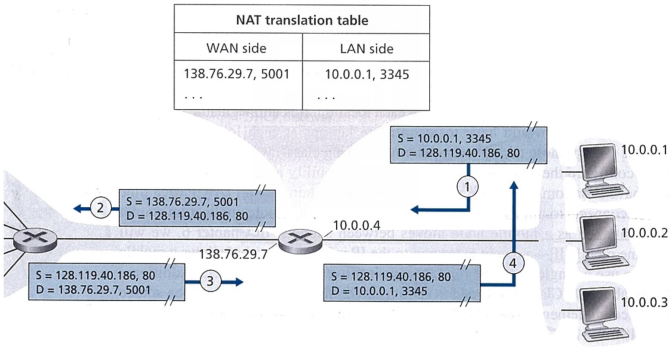
Motivation: A local network behind a router does not necessarily need an IP address for each computer from the ISP or ICANN

- 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 network not explicitly addressable, visible by outside world (a security plus)

NAT: Network Address Translation

NAT router changes datagram addresses between WAN and LAN

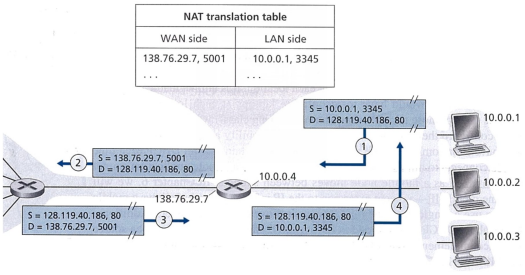
- use 16-bit port-number field (Layer-4):
  - 60,000 simultaneous connections with a single LAN-side address!



NAT: Network Address Translation

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, e.g., P2P applications
- address shortage should instead be solved by IPv6



IPv6: motivation

Initial motivation

- 32-bit address space soon to be completely allocated

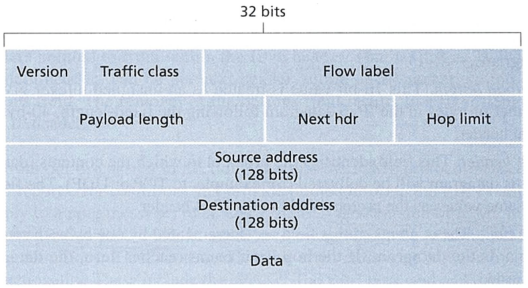
Additional motivation

- header format helps speed processing/forwarding
- header changes to facilitate QoS

IPv6 datagram format

Most important changes

- Expanded addressing capabilities: 128 bit addresses
- Streamlined 40 byte header: some IPv4 fields are removed
- Flow labelling and priority: Sender may request specific handling (QoS) for specific flows



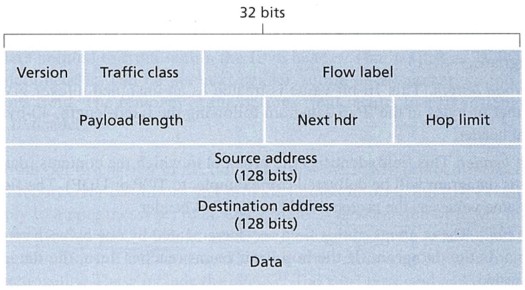
IPv6 datagram format

Version (4 bit)

Describes the IP version

Traffic class (8 bit)

Similar to Type of service field in IPv4



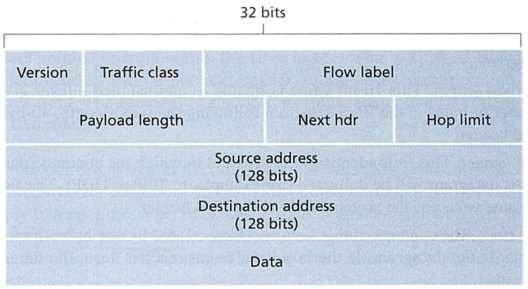
IPv6 datagram format

Flow label (20 bit)

Identify flow of datagrams and their QoS and handling

Payload length (16 bit)

Number of bytes in the datagram following the fixed-length 40 byte header



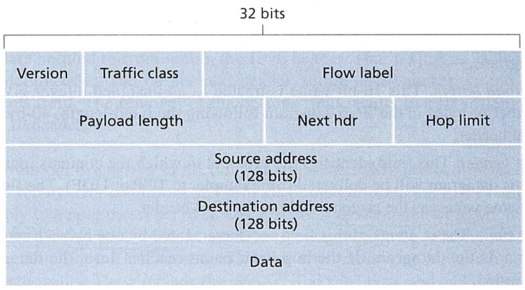
IPv6 datagram format

Next header (8 bit)

Type of the upper layer protocol in data (cf. protocol field in IPv4)

Hop limit (8 bit)

Similar to TTL in IPv4



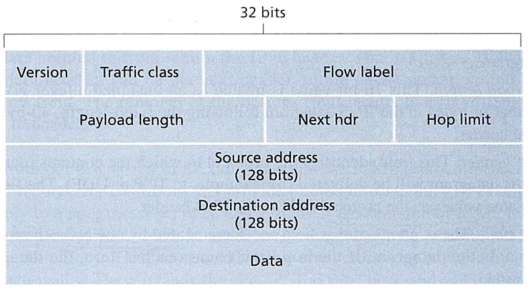
IPv6 datagram format

Source and destination addresses (128 bit each)

128 bit source and destination addresses

data (variable length)

Payload



Other changes from IPv4

Checksum

- Removed entirely to reduce processing time at each hop
- Avoid redundant error check with link layer

Options

- Allowed, but outside of header, indicated by "Next Header" field
- Chain of headers terminated by the header of the upper layer protocol

Fragmentation/Reassembly

- IPv6 does not allow fragmentation at routers
- Fragmentation only at source nodes
- Need to retransmit if packet too big (signaled by ICMP)

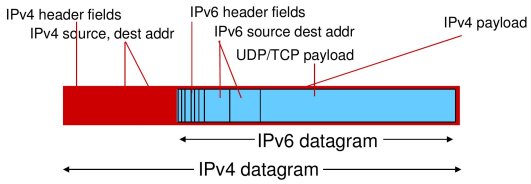
Transition from IPv4 to IPv6

Not all routers can be upgraded simultaneously

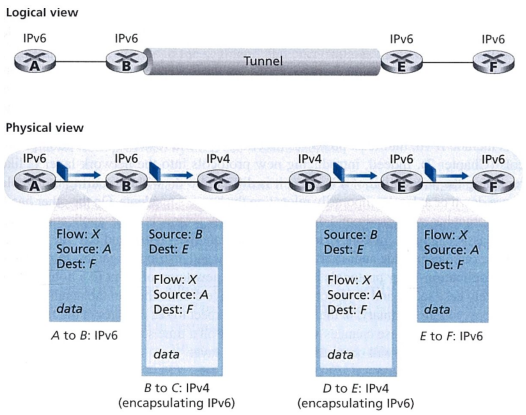
- no "flag days"
- how will network operate with mixed IPv4 and IPv6 routers?

Tunneling

- IPv6 datagram carried as payload in IPv4 datagram among IPv4 routers



Tunneling



Internet Control Message Protocol (ICMP)

ICMP

- Used by host and routers to communicate network-layer information
- Network layer protocol but encapsulated in IP datagrams
- Also utilized by some application programs (ping, traceroute)

ICMP message

- types
- code
- plus the first 8 bytes of IP datagram causing error

Internet Control Message Protocol (ICMP)

ICMP type	Code	Description
0	0	echo reply (to ping)
3	0	destination network unreachable
3	1	destination host unreachable
3	2	destination protocol unreachable
3	3	destination port unreachable
3	4	destination network unknown
4	0	source quench (congestion control)
8	0	echo request
9	0	router advertisement
10	0	router discovery
11	0	TTL expired
12	0	IP header bad

ICMPv6: new version of ICMP

- additional message types, e.g. "Packet Too Big"

Traceroute and ICMP

Traceroute

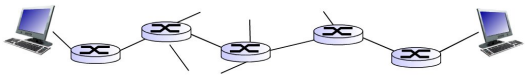
- Example from Tokyo to www.nytimes.com

```
traceroute to www.nytimes.com (170.149.168.130), 30 hops max, 60 byte packets
 1  ve-222.juniper1.hongo.wide.ad.jp (203.178.135.1)  1.309 ms  1.335 ms  1.360 ms
 2  ve-130.foundry4.nezu.wide.ad.jp (203.178.136.185)  0.214 ms  0.240 ms  0.268 ms
 3  ve-42.foundry6.otemachi.wide.ad.jp (203.178.136.65)  0.347 ms  0.369 ms  0.390 ms
 4  ve-51.cisco2.notemachi.wide.ad.jp (203.178.141.142)  0.606 ms  0.663 ms  0.740 ms
 5  ge-8-2.a15.tokyjp01.jp.ra.gin.ntt.net (203.105.72.17)  0.872 ms  0.936 ms  0.971 ms
 6  ae-5.r24.tokyjp05.jp.bb.gin.ntt.net (203.105.72.153)  5.657 ms  5.487 ms ae-5.r25.toky
 7  ae-13.r20.tokyjp01.jp.bb.gin.ntt.net (129.250.6.191)  0.607 ms  0.575 ms ae-12.r20.tok
 8  as-1.r20.sttlwa01.us.bb.gin.ntt.net (129.250.4.189)  98.351 ms  84.700 ms  93.908 ms
 9  ae-1.r05.sttlwa01.us.bb.gin.ntt.net (129.250.5.47)  90.392 ms  105.557 ms  90.339 ms
10  ae-0.level3.sttlwa01.us.bb.gin.ntt.net (129.250.8.74)  84.487 ms  98.076 ms  93.649 ms
11  ae-32-52.ebr2.Seattle1.Level3.net (4.69.147.182)  162.847 ms  171.983 ms  176.430 ms
12  ae-2-2.ebr2.Denver1.Level3.net (4.69.132.54)  198.330 ms  191.047 ms  202.615 ms
13  ae-3-3.ebr1.Chicago2.Level3.net (4.69.132.62)  177.894 ms  168.466 ms  182.681 ms
14  ae-6-6.ebr1.Chicago1.Level3.net (4.69.140.189)  203.643 ms  213.071 ms  200.345 ms
15  ae-2-2.ebr2.NewYork2.Level3.net (4.69.132.66)  166.730 ms  166.527 ms  181.364 ms
16  ae-1-100.ebr1.NewYork2.Level3.net (4.69.135.253)  181.125 ms  176.312 ms  173.774 ms
17  ae-4-4.ebr1.NewYork1.Level3.net (4.69.141.17)  207.655 ms 4.69.201.45 (4.69.201.45)  1
18  ae-2-2.ebr1.Newark1.Level3.net (4.69.132.98)  191.437 ms  179.563 ms  176.368 ms
19  ae-11-51.car1.Newark1.Level3.net (4.69.156.5)  177.557 ms  185.348 ms  189.754 ms
20  NEW-YORK-TI.car1.Newark1.Level3.net (4.30.129.234)  191.326 ms  190.797 ms  191.178 ms
21  170.149.168.130 (170.149.168.130)  169.056 ms  205.439 ms  218.156 ms
```

Traceroute and ICMP

Traceroute algorithm

- Source sends series of IP datagrams to the destination
  - first set has TTL=1, second set has TTL=2, ...
- Each carries a UDP segment with an unlikely UDP port number
- Source starts timer for each datagram
- when n-th set of datagrams arrives to n-th router
  - discard datagram and replies ICMP message "TTL expired"
  - ICMP message include router name and IP
- When ICMP message arrives, source records RTT
- Stopping criteria:
  - UDP segment eventually arrives at destination host
  - destination returns ICMP message "dest. port unreachable"



The Network Layer: Summary

Today's lecture covered the network layer

- Network layer services
- IPv4, IPv6
- IP addressing
- ICMP

In the next lecture

More about the network layer: next time we'll see how routing works!

IP Stack

Application

Transport

Network

Link

Physical

Today's important points

- IP
- IP Addressing
- Traceroute