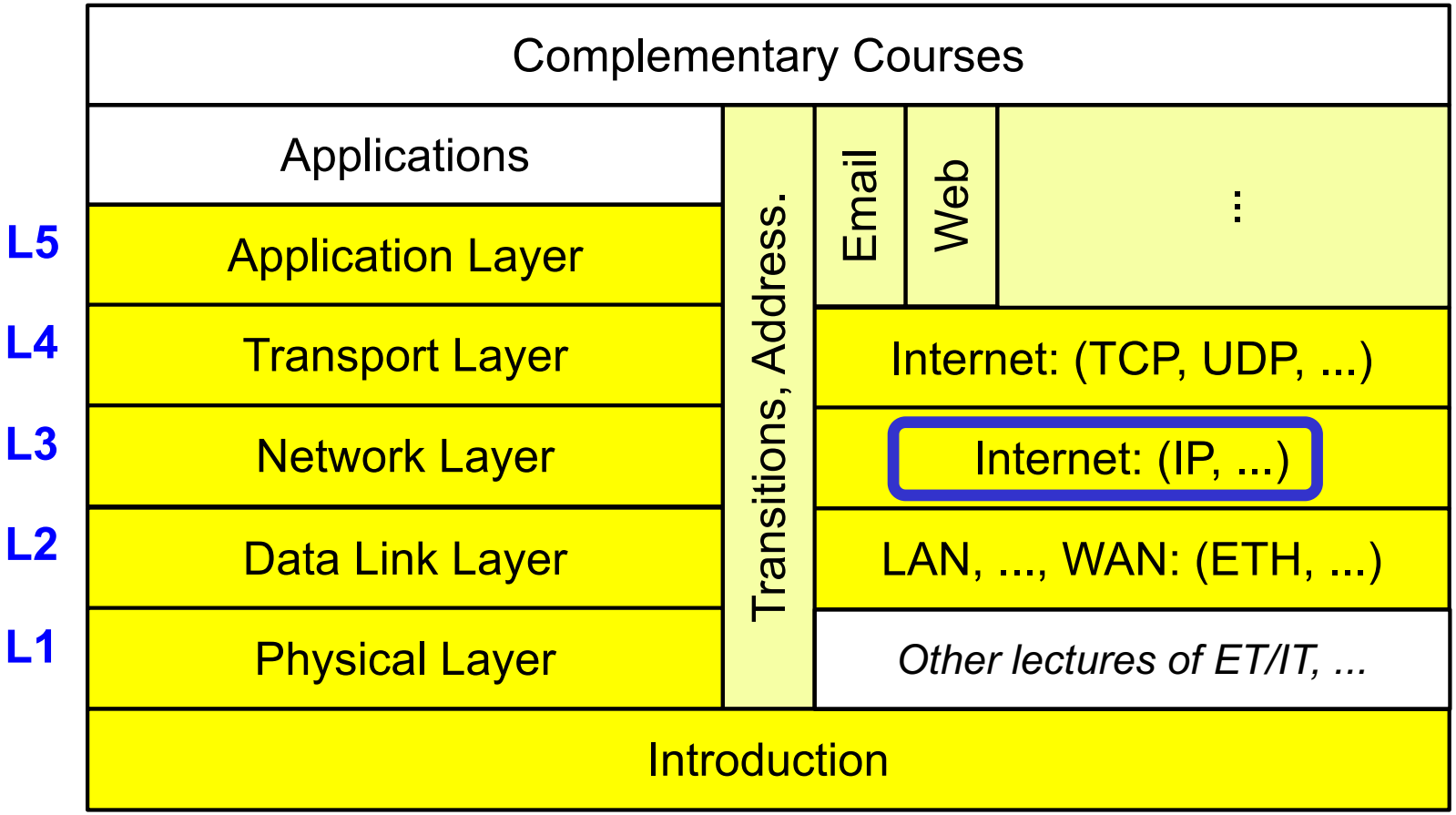
Computer Networks I

Network Layer: Internet Protocols

Prof. Dr.-Ing. **Lars Wolf**

IBR, TU Braunschweig
Mühlenpfordtstr. 23, D-38106 Braunschweig, Germany,
Email: wolf@ibr.cs.tu-bs.de

Scope


Complementary Courses
Applications
Email
Web
...
L5
Application Layer
L4
Transport Layer
Internet: (TCP, UDP, ...)
L3
Network Layer
Internet: (IP, ...)
L2
Data Link Layer
LAN, ..., WAN: (ETH, ...)
L1
Physical Layer
Other lectures of ET/IT, ...
Transitions, Address.
Introduction

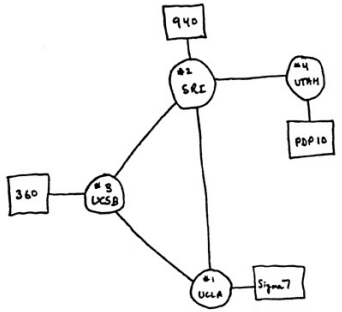
Overview

Overview

1 History and Architecture

ARPANET

- initiated and financed by ARPA
  - Advanced Research Projects Agency, U.S. Department of Defense
- objective:
  - originally: network to survive nuclear war
  - later: network to connect scientific and military institutions
- 1969:
  - experimental network with 4 nodes,
    followed by rapid growth, BBN first contractor

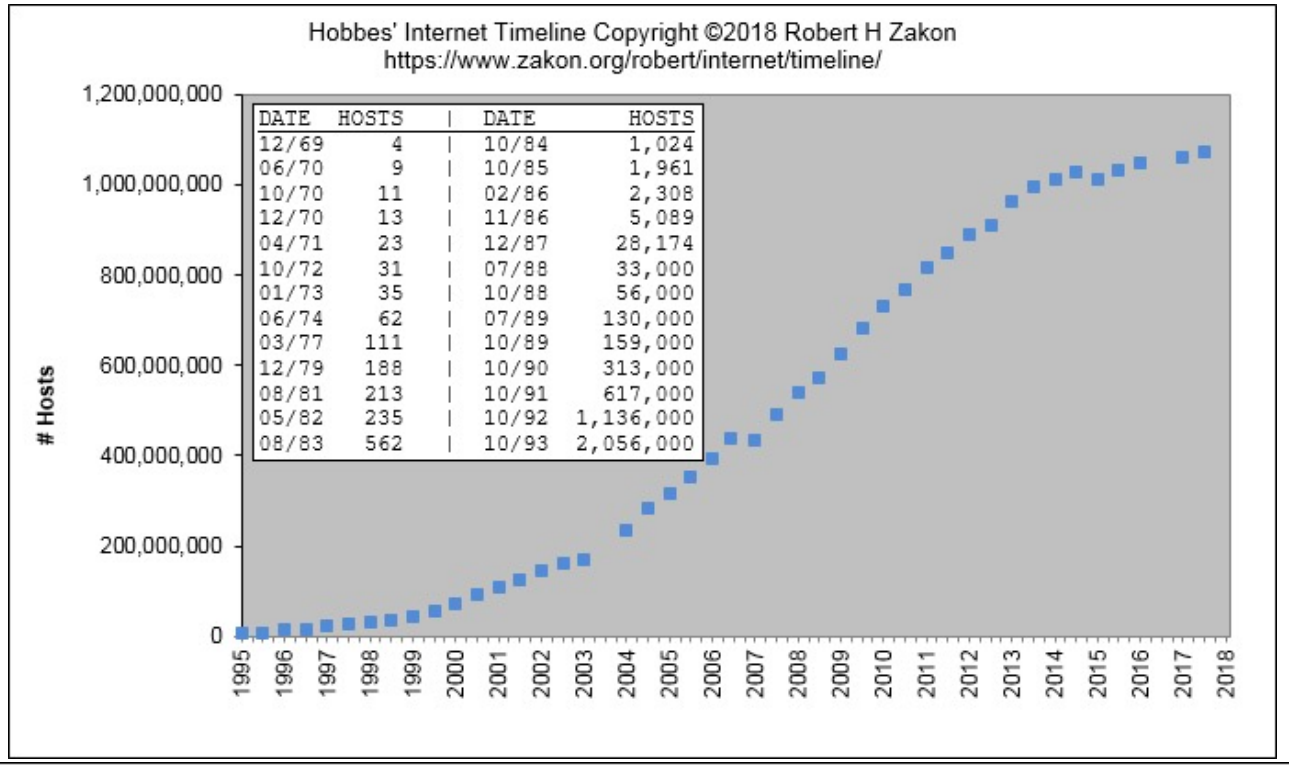


- development of the INTERNET
  - standardized protocols for comm. between networks: TCP/IP (1983)
  - linking military networks (MILNET, MINET)
  - linking satellite networks (SATNET, WIDEBAND)
  - linking the LANs of the universities
- fast spreading of TCP/IP technology as part of UNIX

Some Data about Internet Growth

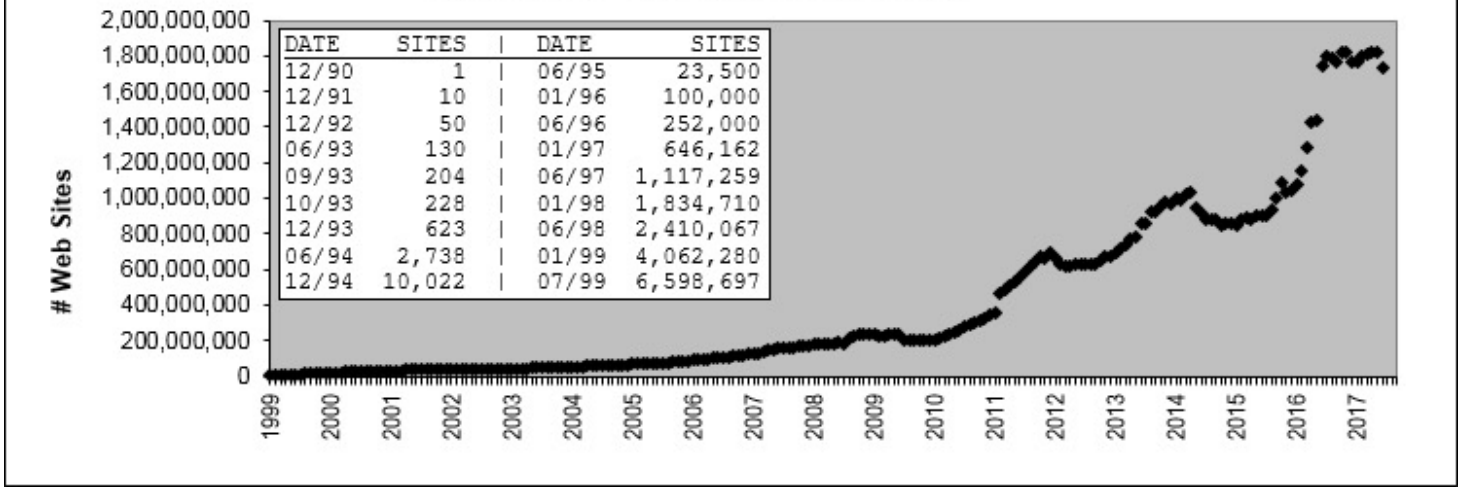
Hobbes' Internet Timeline Copyright ©2018 Robert H Zakon
https://www.zakon.org/robert/internet/timeline/

DATE	HOSTS	DATE	HOSTS
12/69	4	10/84	1,024
06/70	9	10/85	1,961
10/70	11	02/86	2,308
12/70	13	11/86	5,089
04/71	23	12/87	28,174
10/72	31	07/88	33,000
01/73	35	10/88	56,000
06/74	62	07/89	130,000
03/77	111	10/89	159,000
12/79	188	10/90	313,000
08/81	213	10/91	617,000
05/82	235	10/92	1,136,000
08/83	562	10/93	2,056,000



Hobbes' Internet Timeline Copyright ©2018 Robert H Zakon
https://www.zakon.org/robert/internet/timeline/

DATE	SITES	DATE	SITES
12/90	1	06/95	23,500
12/91	10	01/96	100,000
12/92	50	06/96	252,000
06/93	130	01/97	646,162
09/93	204	06/97	1,117,259
10/93	228	01/98	1,834,710
12/93	623	06/98	2,410,067
06/94	2,738	01/99	4,062,280
12/94	10,022	07/99	6,598,697



The Internet: Committees

Internet (Internet Society)

- mid-80s
  - a multiple of networks was designated as the "Internet"
- Jan. 1992:
  - founding of the (actual) Internet Society
  - objective: to spread the use of the Internet (protocols and services)

- IAB: Internet Architecture Board
  - founded in 1983 to involve researchers in the ARPANET
  - today it is the supreme Internet board
- IAB oversees/nominates
  - IETF (INTERNET ENGINEERING TASKFORCE)
    - divided into working groups
    - actual governing board
  - IRTF (Internet Research Taskforce)

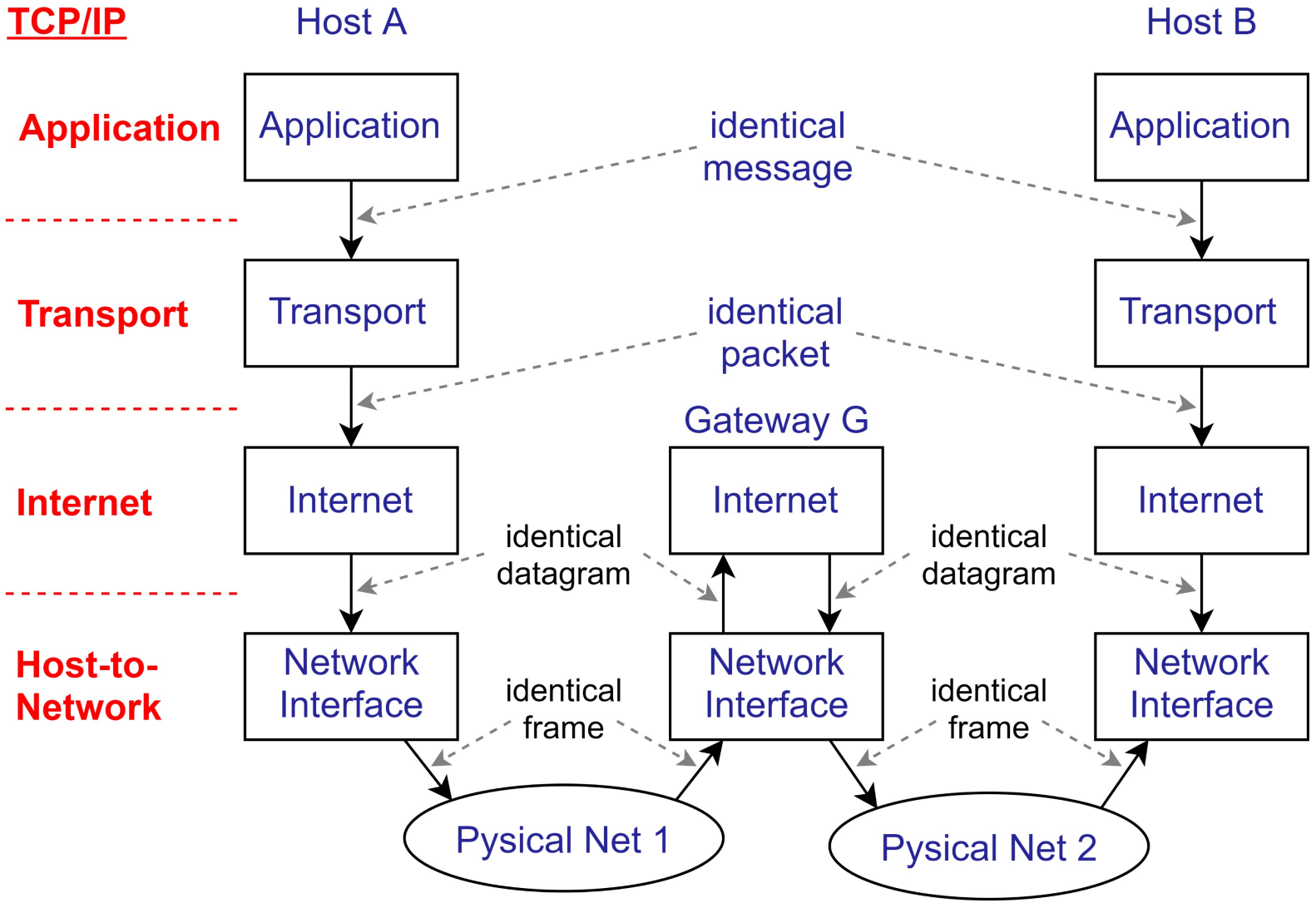
- RFC (REQUEST FOR COMMENTS)
  - recommendations, e.g. June 2021 approx. 9000

The Internet: Tasks

Tasks in the INTERNET

- to connect different networks over gateways
- definition of
  - protocols that work on all subnetworks
  - standardized addressing pattern for a very large network
  - global routing architecture

Internet Architecture



i.e.

- ISO-OSI presentation and session layer not explicitly available
- data link layer and physical layer combined

Internet Architecture

No formal architecture

No unchangeable principles:

The principle of constant change is perhaps the only principle of the Internet that should survive indefinitely.
[RFC 1958, Architectural Principles of the Internet, June 1996]

The Internet approach in very general terms (from RFC 1958):

- the goal is connectivity
- the tool is the Internet Protocol
- the intelligence is end-to-end rather than hidden in the network

Some Internet Guidelines (1)

make sure it works

- do not finalize a design / standard
  before multiple prototypes are interoperable

keep it simple (stupid ... KISS)

- when in doubt, use simplest solution:
  leave unnecessary features out

make clear choices

- if there are alternatives for same thing,
  choose one (avoid too many options)
- re-use good solutions if applicable,
  but avoid duplication of functionality

exploit modularity

- like layers in protocol stacks

…

Some Internet Guidelines (2)

…

expect heterogeneity
- hardware, applications, etc.

avoid static options and parameters
- negotiations among sender and receiver

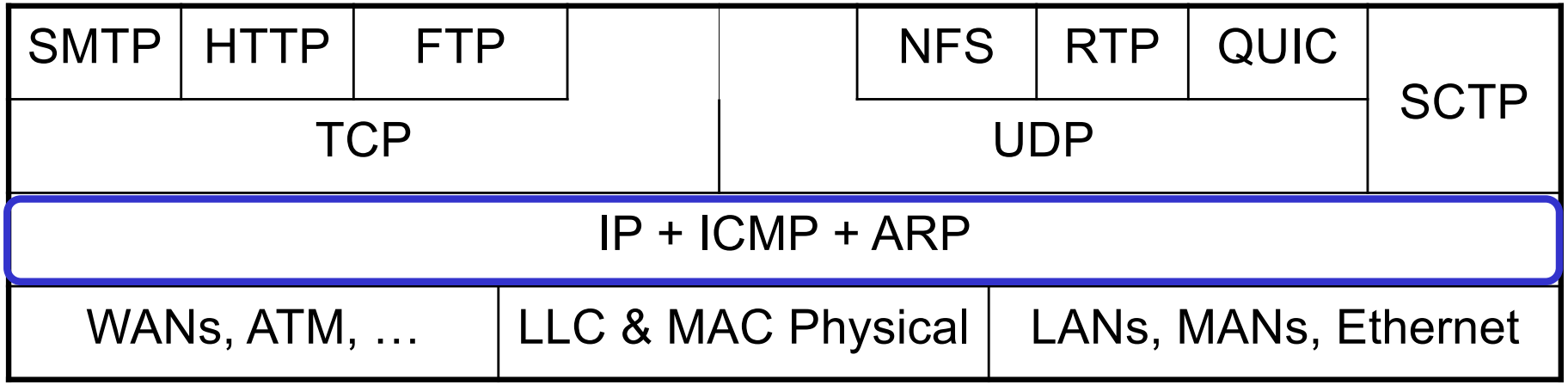
look for good design, but not necessarily perfect
- adopt almost complete solution now,
  don't wait for perfect solution

be strict when sending and tolerant when receiving (*Postel's Law)*
- follow spec.s precisely when sending,
  but tolerate faulty input from net

consider scalability
- many nodes per site and many millions of sites

consider cost and performance

(Some) Well-Known Internet Protocols



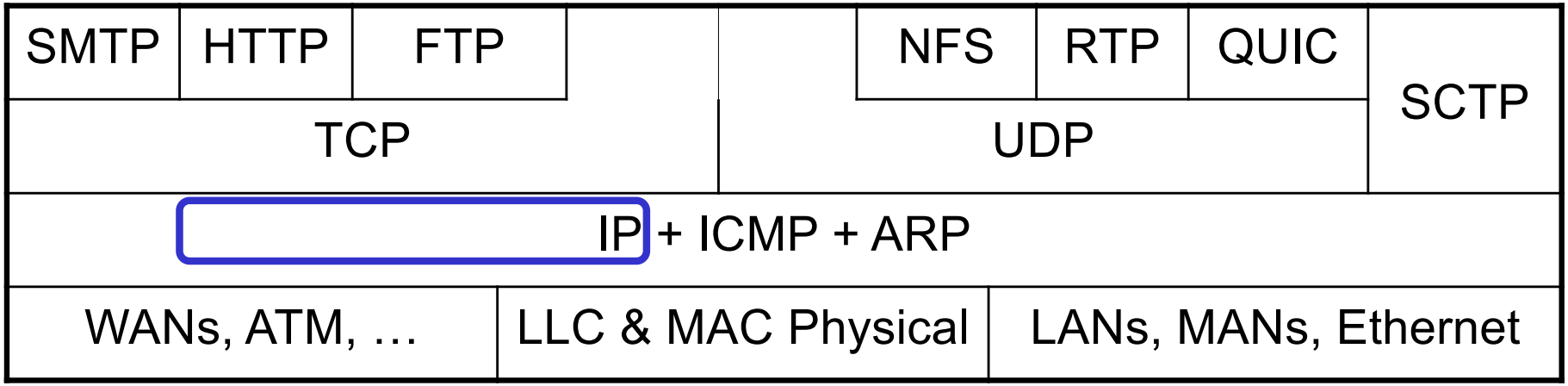
ARP = ADDRESS RESOLUTION PROTOCOL
IP = INTERNET PROTOCOL
ICMP = INTERNET CONTROL MESSAGE PROTOCOL

LLC = Logical Link Control
MAC = Media Access Control

TCP = Transmission Control Protocol
UDP = User Datagram Protocol
SCTP = Stream Control Transmission Protocol

SMTP = Simple Mail Transfer Protocol
HTTP = Hypertext Transfer Protocol
FTP = File Transfer Protocol
NFS = Network File System
RTP = Real-Time Transport Protocol
QUIC (simply the name of the protocol)

2 Internet Protocol (IP)

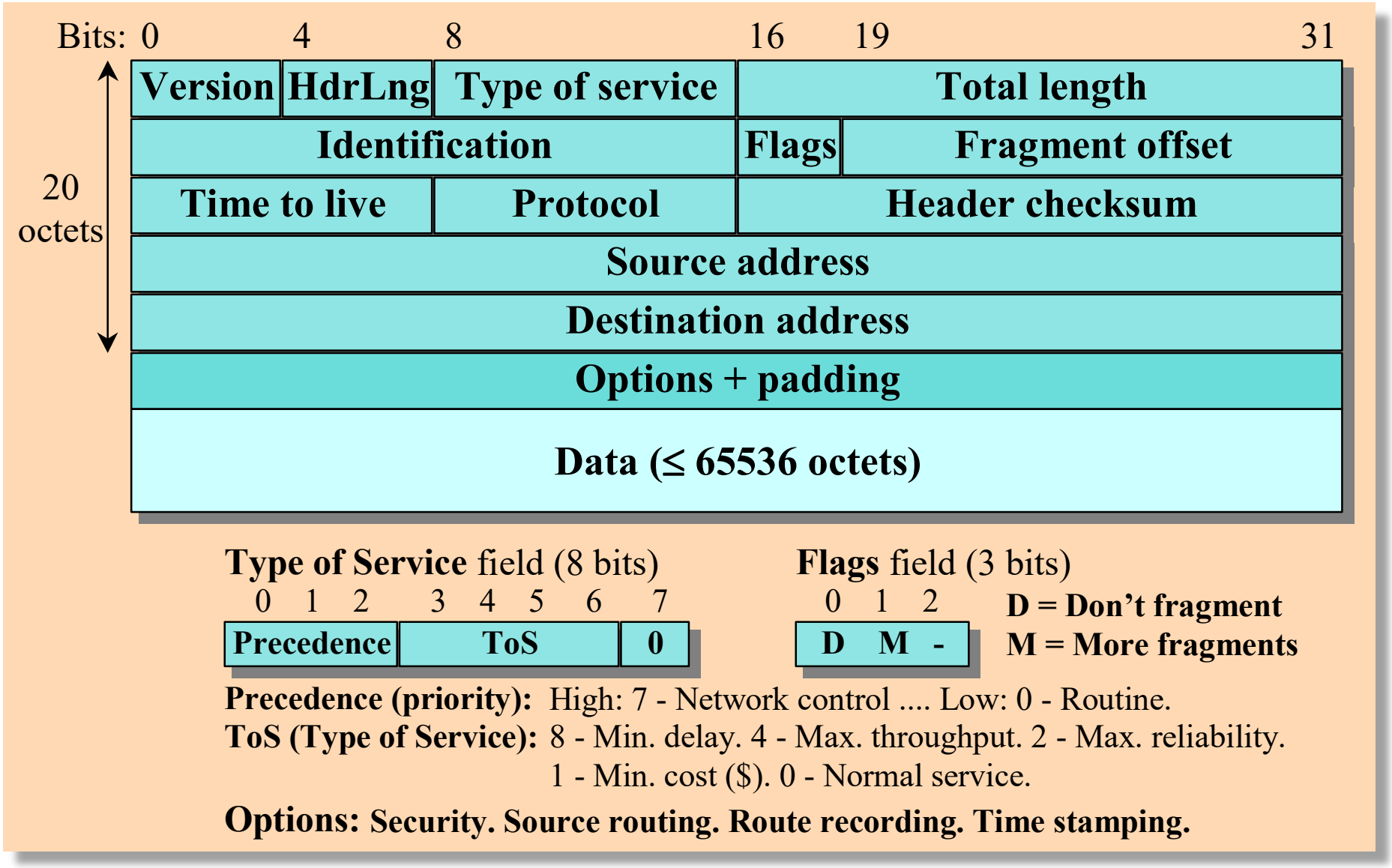


Defined in RFC 791, September 1981 (J.Postel)

connectionless service (datagram)

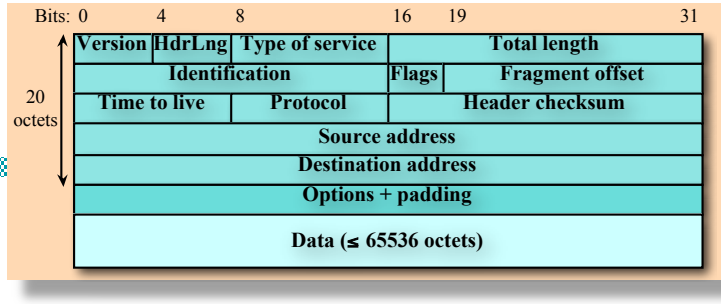
- provide best-effort way to transport datagrams
  - from source to destination
  - without regard whether
    - these machines are on the same network
    - there are other networks in between
- no guarantees

IPv4 Datagram Format



Comment: "Big Endian" transfer order (from left to right, highest version bit first)

IPv4 Datagram Format

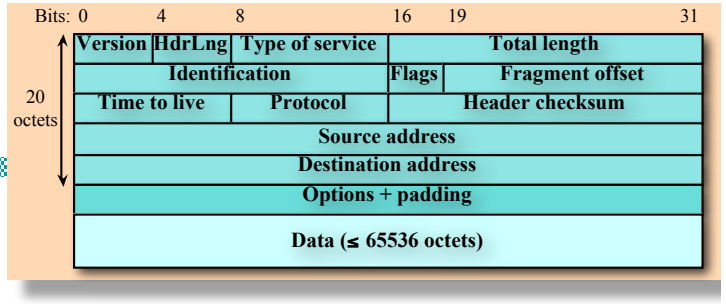


Version
- IP v.4 actual protocol version
- IP v.5 (real time data transfer) ST-2
- IP v.6 successor to IP v.4

Header Length (IHL) (in 32 bit words)
- at least 5 words with 32 bit each = 20 bytes
- at most 15 words with 32 bit each = 60 bytes

Type of Service
- simple QoS: a combination of reliability and delay
  - precedence (3 bit):
    - priority 0 (normal) ...7 (network control)
    - influences the queuing scheme (and not routing)
  - D (1 bit): Delay, e.g. no satellite transmission
  - T (1 bit): Throughput, e.g. no telephone line
  - R (1 bit): Reliability, e.g. no radio channels
  - C (1 bit): low Cost, defined later on
  - 1 bit unused
    comment: C & D activated: e.g. invalid
- practical view: ignored by routers
- redefined for **Differentiated Services** (DiffServ)

IPv4 Datagram Format

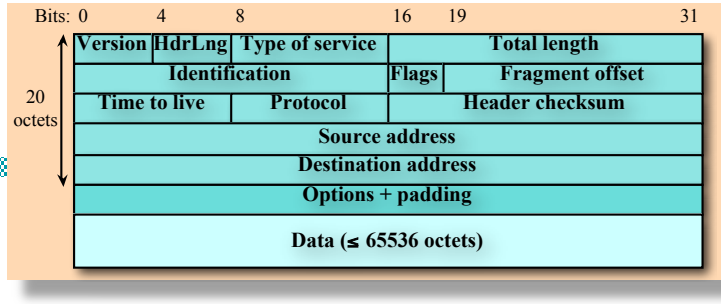


Total length
- full length including the data
- stated in bytes
- all hosts must accept datagrams of up to 576 bytes
- max. 65535 byte, often approximately 1500 byte sent

Identification
- identify group of fragments of single datagram (at destination)
- all fragments of a datagram contain same identification value

Flags
- 1 bit unused
- DF (1 bit): don't fragment
  - packets may have a length of up to 576 byte
- MF (1 bit): more fragments
  - last fragment marked 0

IPv4 Datagram Format



Fragment offset

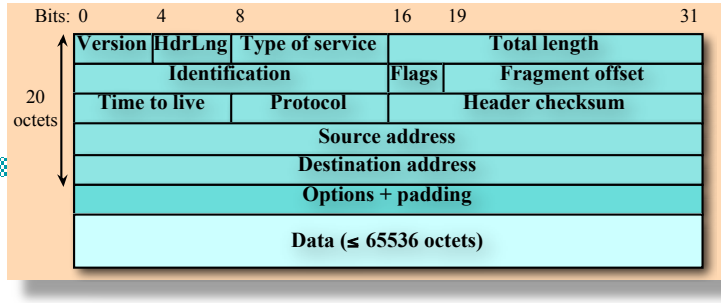
- offset of this fragment,
  i.e. the position within a datagram
- stated in multiples of 8 bytes (elementary fragment unit)
- length of 13 bits ➔ max. 8192 fragments / datagram
  - max. datagram length 65536 bytes

Time To Live (TTL)

- life cycle in seconds, max. 255 sec
- when 0: drop packet, feedback to sender
- must be decremented per hop

- in practical use: counts hops (not sec.)

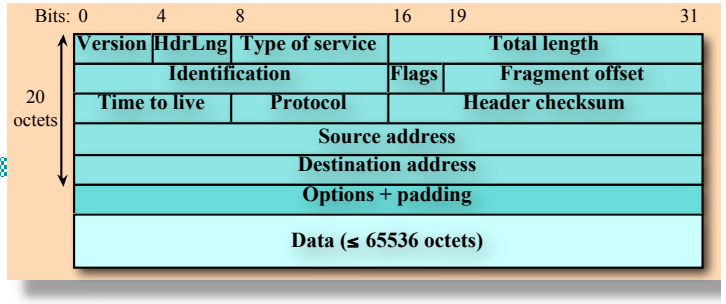
IPv4 Datagram Format



Protocol: higher level protocol in payload of datagram

No.	Abbreviation	Protocol
0		reserved
1	ICMP	Internet Control Message
2	IGMP	Internet Group Management
3	GGP	Gateway to Gateway
4	IP	IP in IP
5	ST	Stream
6	TCP	Transmission Control
…	…	…

IPv4 Datagram Format



Header Checksum

- used to detect errors generated by bad memory words inside an IS
- observed each time when datagram is received
  - both in IS and ES
  - if error is detected, then datagram is dropped
- summation of the header words
  - addition of all 16-bit halfwords in one's complement arithmetic and use one's complement of result (assume this field as zero upon arrival)
- must be recomputed at each hop
  - due to change in Time-to-Live field

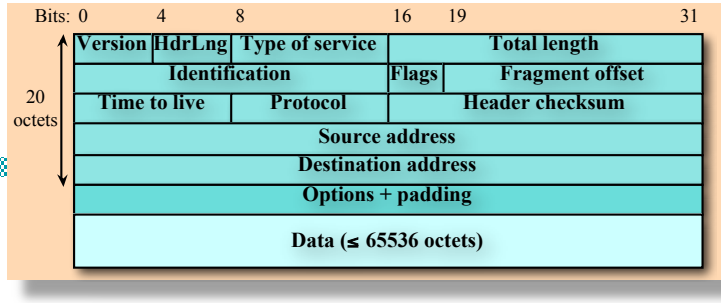
Source Address

- sender's IP address

Destination Address

- receiver's IP address

IPv4 Datagram Format



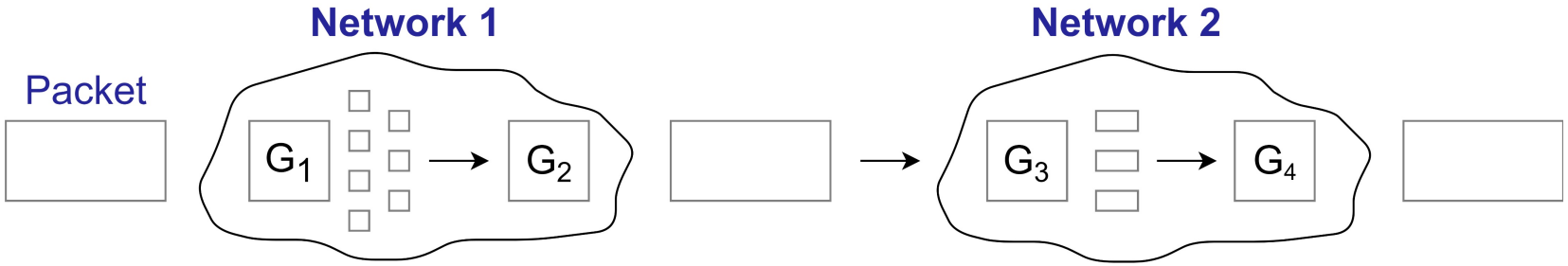
Options
- options for routing, testing and debugging
- conceptual design: as an enhancement for future versions
- variable length: each begins with 1-byte identification code

Padding
- fill up to the word limit

Data
- user data

IP: Segmentation/Reassembling

1. Transparent segmentation



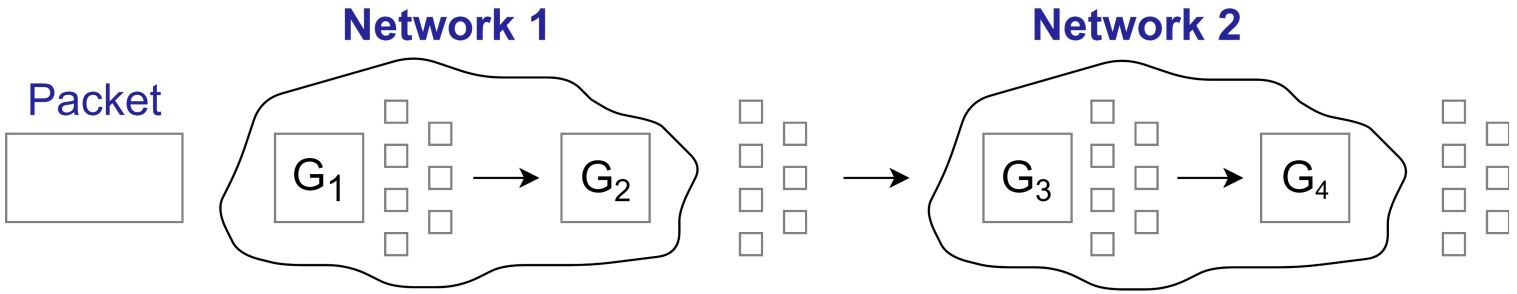
G_1 fragments a large packet

G_2 reassembles the fragments

G_3 fragments again

G_4 reassembles again

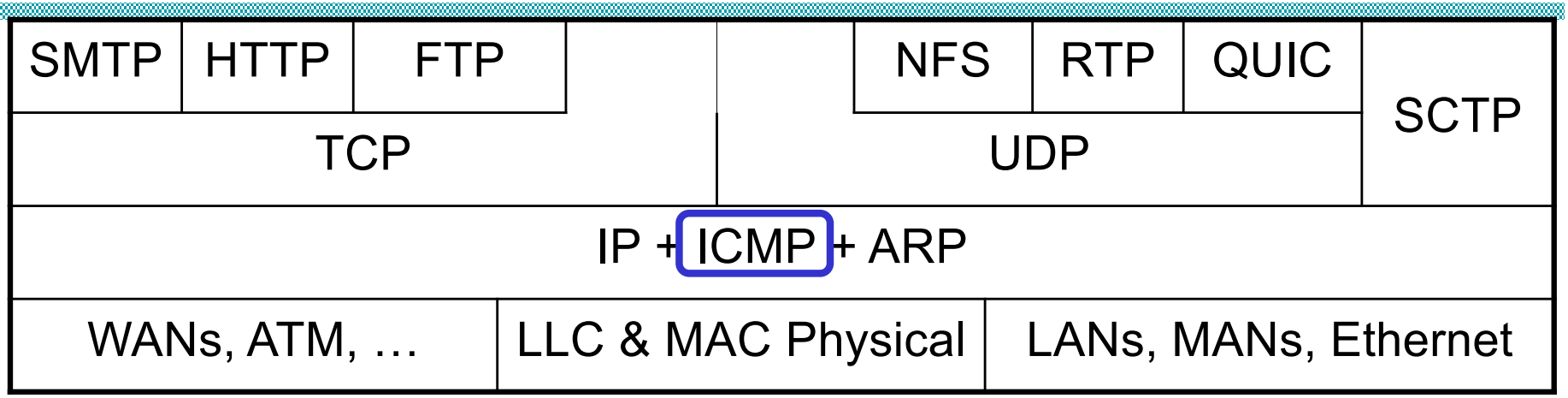
2. Non-transparent segmentation
   - **USED IN IP:**



G_1 fragments a large packet

The fragments are not reassembled until the final destination (a host) is reached

3 Internet Control Message Protocol (ICMP)



History: J. Postel, RFC 792, Sept. 1981

Purpose

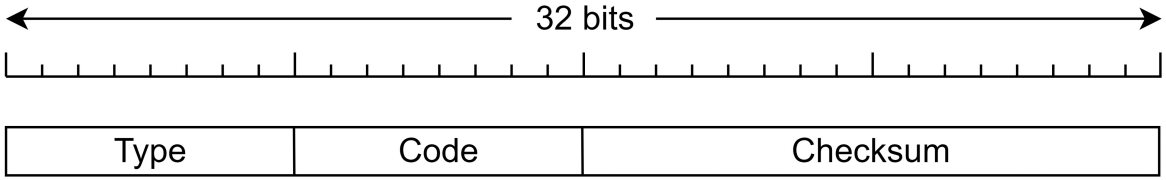
- to communicate network layer information between hosts, routers (and gateways)
  - mostly error reporting
    e.g. in ftp, telnet, http "destination network unreachable"

sent as IP packets

- i. e. the first 32 bits of the IP data field identical as ICMP headers

Internet Control Message Protocol (ICMP)

Header structure



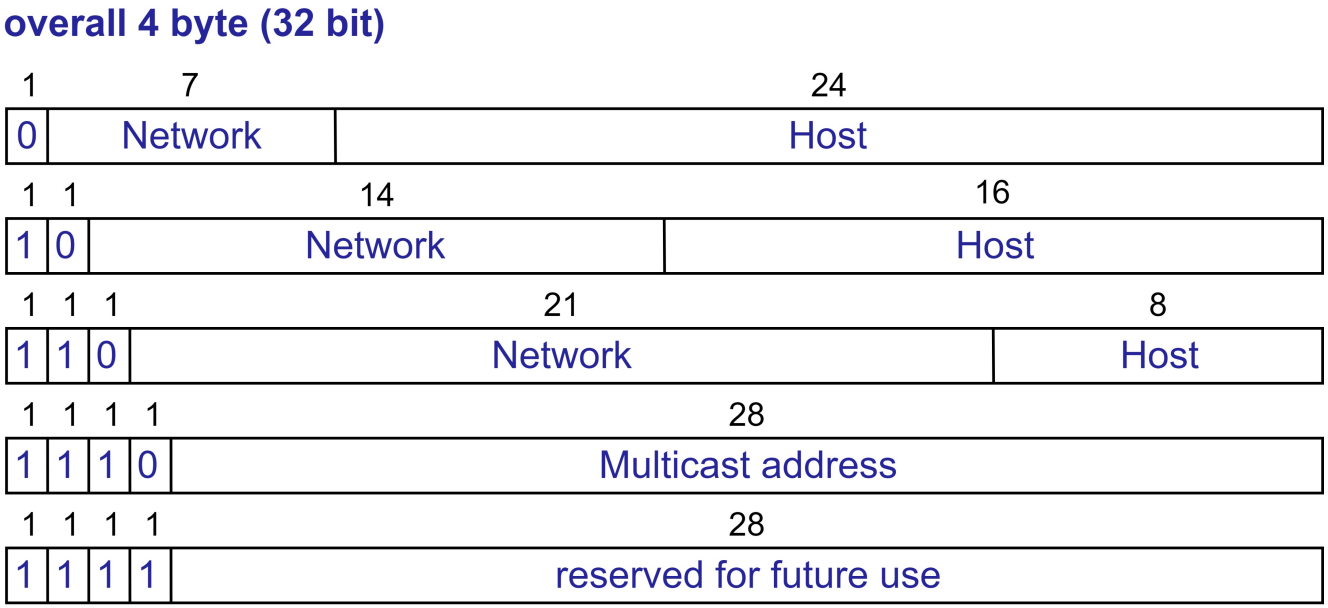
Type

- 16 types, e.g.
  - destination or port or protocol unreachable
  - fragmentation necessary but DF (don’t fragment) DF is set
  - source route failed, redirect (for routing)
  - used by ping program
    - echo request (e.g. for "ping" program)
    - echo reply (response in "ping" program)
  - used by traceroute program
  - source quench (previously used for congestion control: Choke packet)

Code

- states reason if type is "destination unreachable“
  - e.g. net, host, protocol, port unreachable or
  - fragmentation needed, source route failed

4 Internet Addresses and Internet Subnetworks



Global addressing concept for ES (and IS) in the Internet

- unique 32 bit address with net-ID (subnetwork-Id), ES-Id
- i.e., each network interface (not ES) has its own unique address
- originally 5 classes

Network addresses typically written in dotted decimal notation

- A.B.C.D
- e.g., 134.169.34.18

IP Addresses & Forwarding

Forwarding decisions based on IP addresses

- IP addresses of hosts on one network “look similar”
  – have same prefix

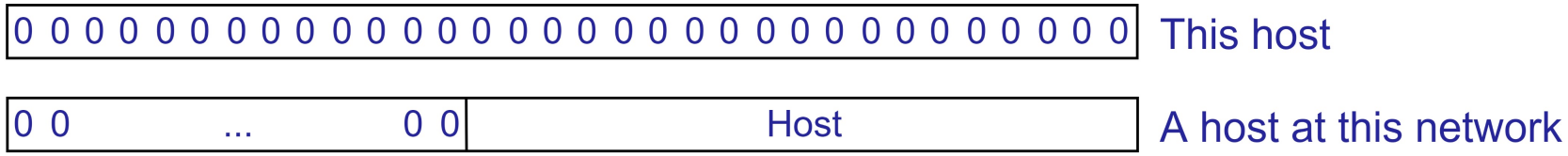
ES checks whether address belongs to same network

- Yes: direct delivery
- No: send to nearest (default) router

4.1 Special Internet Addresses

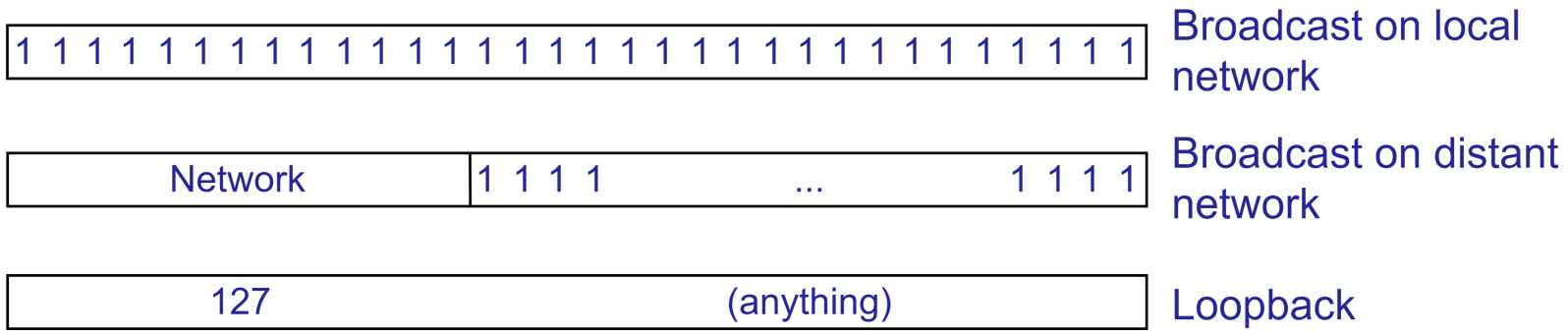
Special IP addresses:

- Source Addresses



Special IP addresses:

- Destination Addresses



(Special) IP Addresses

Public IP addresses, e.g., 134.169.34.1

- valid destination on the global Internet
- must be allocated before use
- nearly no remaining
  → use of Network Address Translation (NAT)
  → IPv6

Private IP addresses

- used within private networks (home, small company)
- 10.0.0.0/8, 172.16.0.0/12, 192.168.0.0/16
- to connect to global Internet, additional mechanisms (middleboxes such as NAT) and public IP address(es) are needed

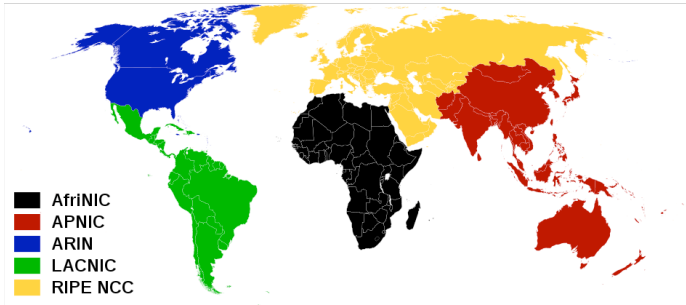
IP Address Assignment

(Public) IP addresses must be unique

- one user only
- must be allocated before use

Internet Assigned Numbers Authority (IANA)

- department of ICANN (Internet Corporation for Assigned Numbers and Names)
- manages network numbers
- delegates parts of the address space to regional authorities: Regional Internet Registry (RIR)
  - NIC Network Information Center
  - E.g. www.denic.de/



4.2 Internet Subnetworks

Structured networks growth

- several networks instead of one preferable
- but getting several address areas is hard
  - since address space is limited
  - e.g.,
    - university may have started with class B address
    - but, doesn't get second one

Problem:

- (original) classes A, B, C refer to
  - one network
  - not collection of LANs

Need

➔ to allow a network to be split into several parts

- for internal use
- still look like single network to outside world

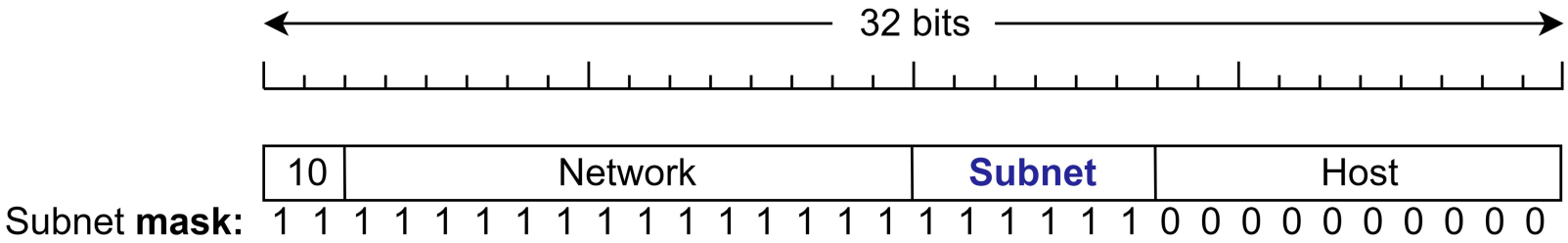
➔ to provide for subnetworks

Internet Subnetworks

Subnets: e.g., Ethernet-based LAN

Idea:

- local decision for subdividing the host part of address into **subnetwork** portion and **end system** portion
- example: class B address: max. 63 subnetworks



Use subnet mask to indicate split between **network + subnet** and **host** part

Forwarding checks using subnet mask

4.3 Classless Routing / IP Prefixes

Problems due to classes

Internet growth led to lack of addresses

- in principle many addresses due to 32-bit address space
  - but nearly all already consumed
- but inefficient allocation due to class-based organization
  - class A network (16 million addresses) too big for most cases
  - class C network (256 addresses) too small
  - → most organizations are interested in class B network,
    - but there are only 16384
    - in reality, class B too large for many organizations

Large number of networks leads to large routing tables

➔ Use prefixes (independent of classes)

- CIDR (Classless InterDomain Routing) (RFC1519)

CIDR: Classless InterDomain Routing

CIDR Principle

- allocate IP ADDRESSES in VARIABLE-SIZED blocks
  - without regard to classes
- e.g., request for 2000 addresses would lead to
  - assignment of 2048 address block starting on 2048 byte boundary

but, dropping classes makes forwarding more complicated

CIDR: Classless InterDomain Routing

Use generalized network prefix

- l-bit prefix: same top l bits for all according addresses
- prefix length of l between 13 and 27 bits

  ➔ blocks of addresses can be assigned to networks
  - as small as 32 hosts up to more than 500000 hosts

Notation: *IP address / length*

- includes
  - the standard 32-bit IP address
  - information on how many bits are used for the network prefix
- e.g. 194.24.8.0 / 24,
  - first 24 bits used to identify unique network
  - remaining bits to identify specific host (256 addresses)

Longer prefix → more specific

- smaller number of IP addresses

Shorter prefix → less specific

- larger number of IP addresses

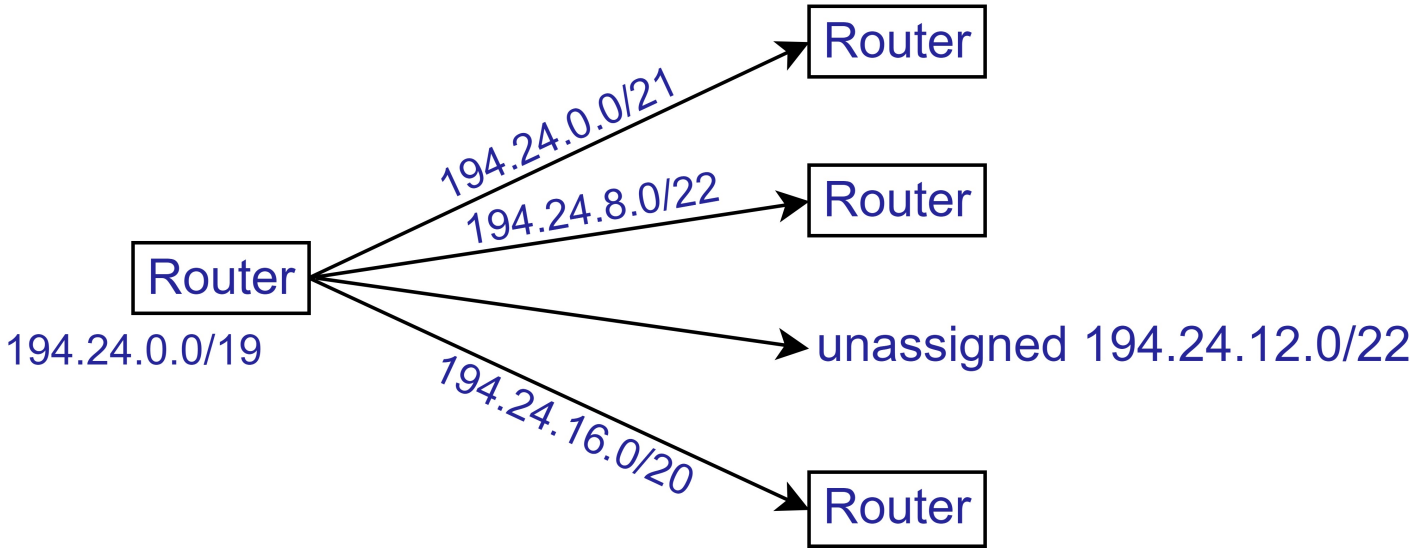
CIDR: Classless InterDomain Routing

Search for **longest matching prefix**

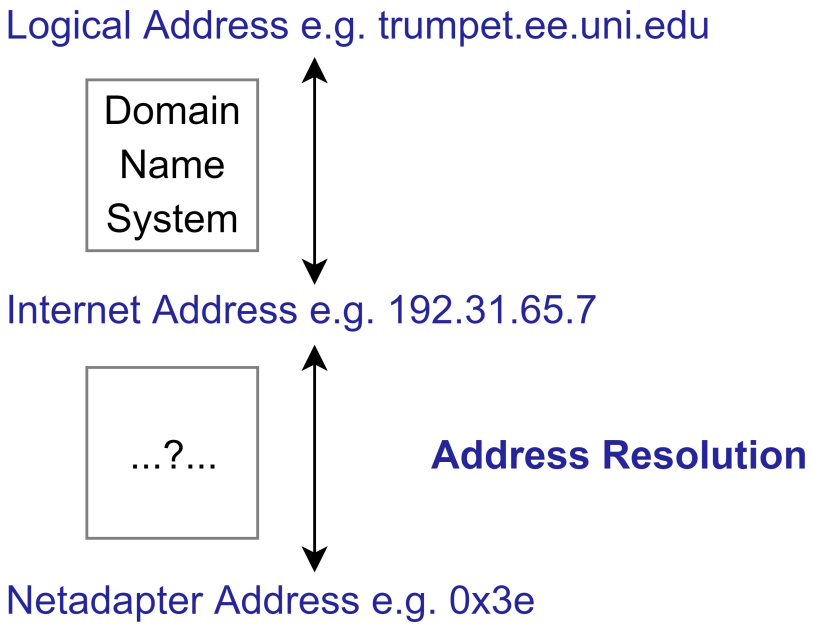
- if several entries with different subnet mask length match
  - then use the one with the longest mask
- i.e., AND operation for address & mask
  - performed for each table entry

E.g., potentially several ’class C’ networks can be characterized by one prefix

Entries may be aggregated to reduce routing tables



5 Address Resolution + NAT



Addressing levels

Host identification and routing specification within a subnetwork
- based on the (local) physical network addresses of ES
  - e.g. station address of the adapter card

Problem:
- INTERNET address (32 bit)
  must be mapped onto the physical network address,
  - usually 48 bit (ADDRESS RESOLUTION)

Address Resolution: Methods

Address resolution in

- source ES, if destination ES is local (direct routing)
- Gateway, if destination ES is not local

Solutions:

1. Direct HOMOGENEOUS ADDRESSING
   - if the physical address can be set by the user, then it could be:
     - physical address = Hostid of the INTERNET address

2. If the physical address is pre-defined or if it has to have a different format, use one of the following:
   - a mapping table from configuration data base (IPaddr → HWaddr),
     - e.g. in the Gateway,
     - may become maintenance nightmare
   - the **Address Resolution Protocol (ARP)**
     - mainly applied in LANs with broadcasting facility

Address Resolution Protocol (ARP)

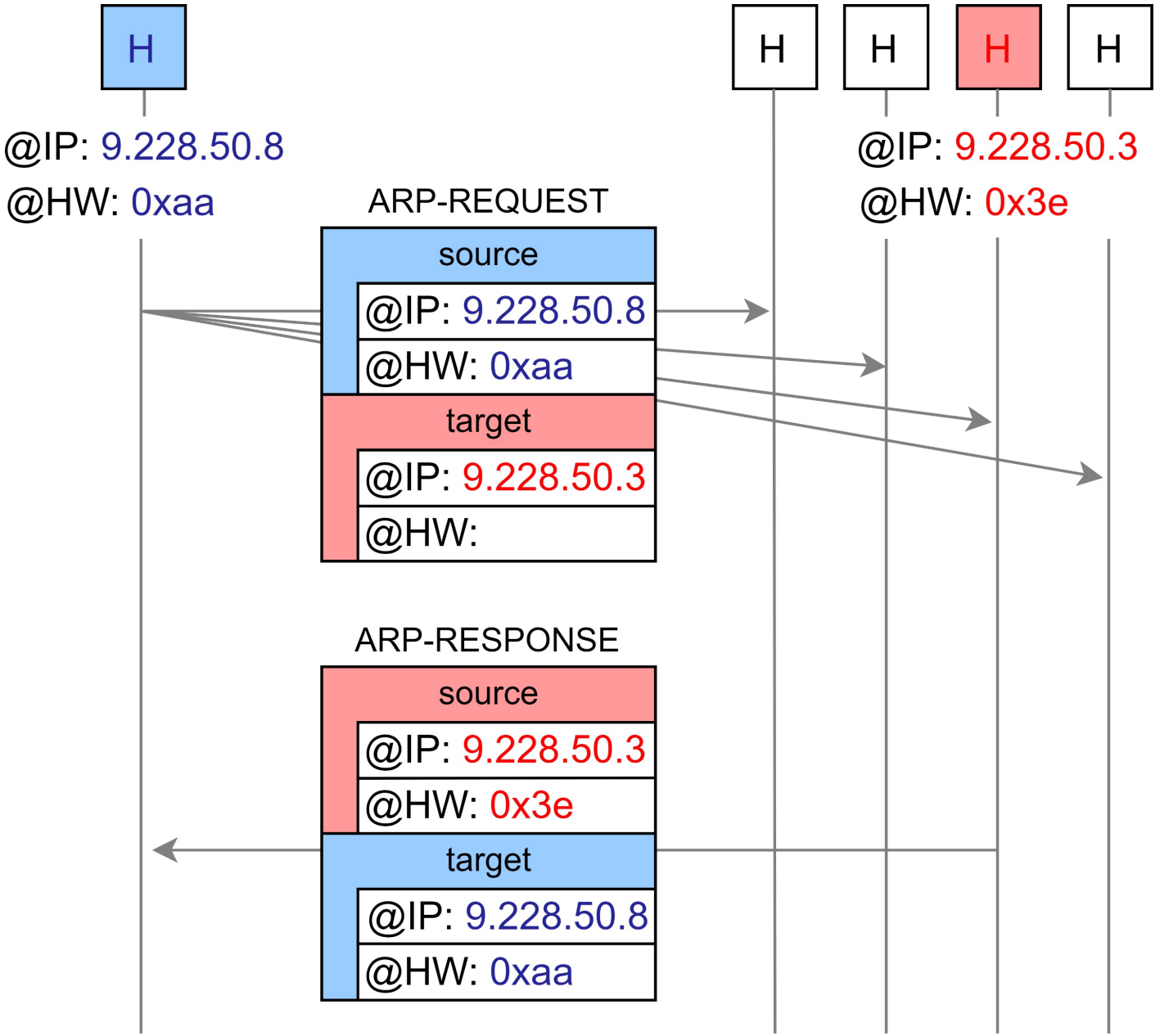
Process:

1. broadcast ARP request datagram on LAN
   - including receiver's INTERNET address (desired value)
   - sender's physical (HW) and INTERNET address (IP)
2. every machine on LAN receives this request and checks address
3. reply by sending ARP response datagram
   - machine which has requested address responses
   - including the phyiscal address
4. store pair (I,P) for future requests

Refinements:

- receiver of ARP request stores sender's (I,P) pair in its cache
- send own table during the boot process
  - (but may be too old)
- entries in ARP cache should time out after some time
  - (few minutes)

Address Resolution Protocol (ARP)



DHCP: Dynamic Host Configuration Protocol

How does a computer know which IP address is assigned to it?

- DHCP helps it

DHCP Objectives

- simplify installation and configuration of end systems
- allow for manual and automatic IP address assignment
- provide additional configuration information
  - (DNS server, netmask, default router, etc.)

DHCP server is used for assignment

- request can be relayed by DHCP relay agent,
  - if server on other LAN

Client broadcasts DHCP DISCOVER packet

- server answers

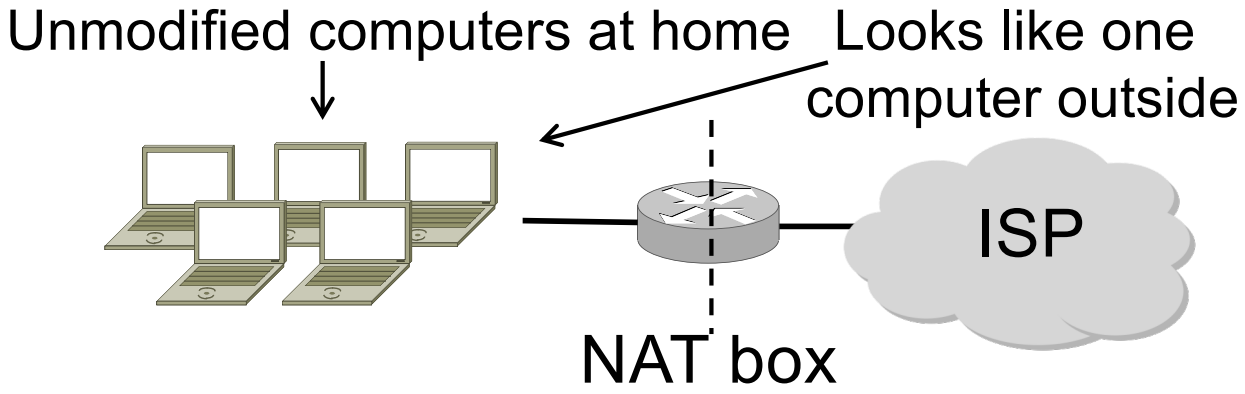
Address is assigned for limited time only

- before the 'lease' expires, client must renew it
- allows to reclaim addresses of disappearing hosts

Network Address Translation (NAT)

Motivation:

- Relax demand on #public IP addresses
- Used at edges of the network, e.g., at homes



From Wetherall & Tanenbaum

"Middlebox" approach

- "gateway" to connect internal network to external network
- peek into packets and "translate addresses"
  - Internal IP address : port → external IP address : port
  - E.g., 192.168,1.12 : 5523 → 44.25.80.3 : 1500
- other "middleboxes" exist as well, e.g., for intrusion detection

Network Address Translation (NAT)

Maintain table: internal / external

- internal IP address : port | external IP address : port
- data from inside: lookup and rewrite source side
- data from outside: lookup and rewrite destination side
- e.g., 192.168.1.12 : 5523 ←→ 44.25.80.3 : 1500

Characteristics:

- violates layering
- breaks connectivity
  - delivery of incoming packets only possible if other packet has been outgoing before
  - operation of servers behind NAT or peer-to-peer applications (e.g., skype) difficult
- other side effects (e.g., FTP)

- But easy to deploy,
- helps with IP address shortage
- provide additional useful functionality such as firewall

8 IP Version 6 (IPv6)

Motivation: Main issues with IPv4

- addressing (presently 32 bit) and
- many other shortcomings in IPv4 (QoS, mobility, ..)

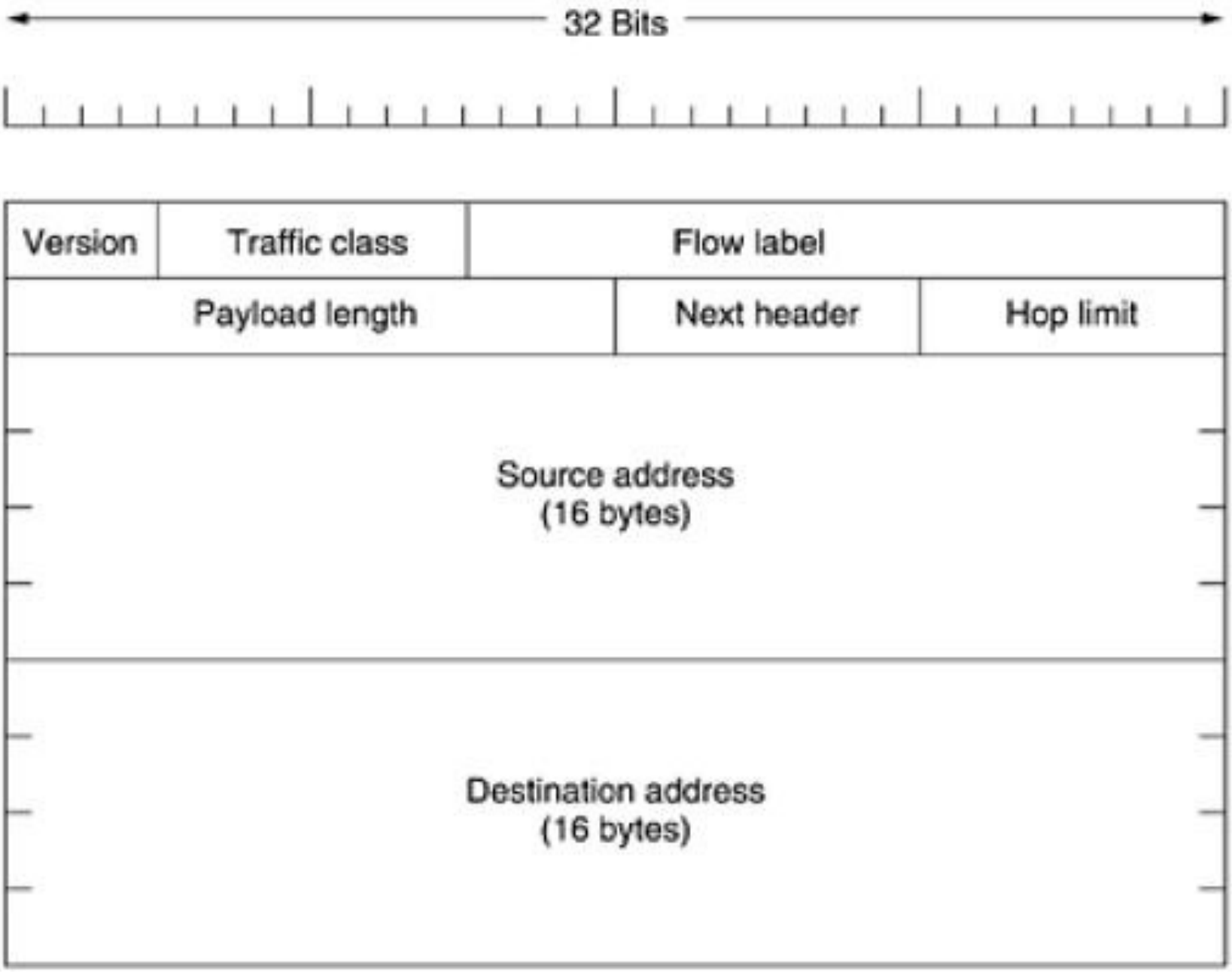
Status

- started early 1990s, since 2011 integrated in all major OS, deployment was much slower than originally expected
- 2016: “IPv6 reaches 10% deployment globally, and becomes the dominant (>50%) Internet protocol for US mobile networks”
  - According to https://www.zakon.org/robert/internet/timeline/; retrieved 15. June 2018
- April 2021: Google's statistics show IPv6 availability of its users at around 30%, in Germany around 50%

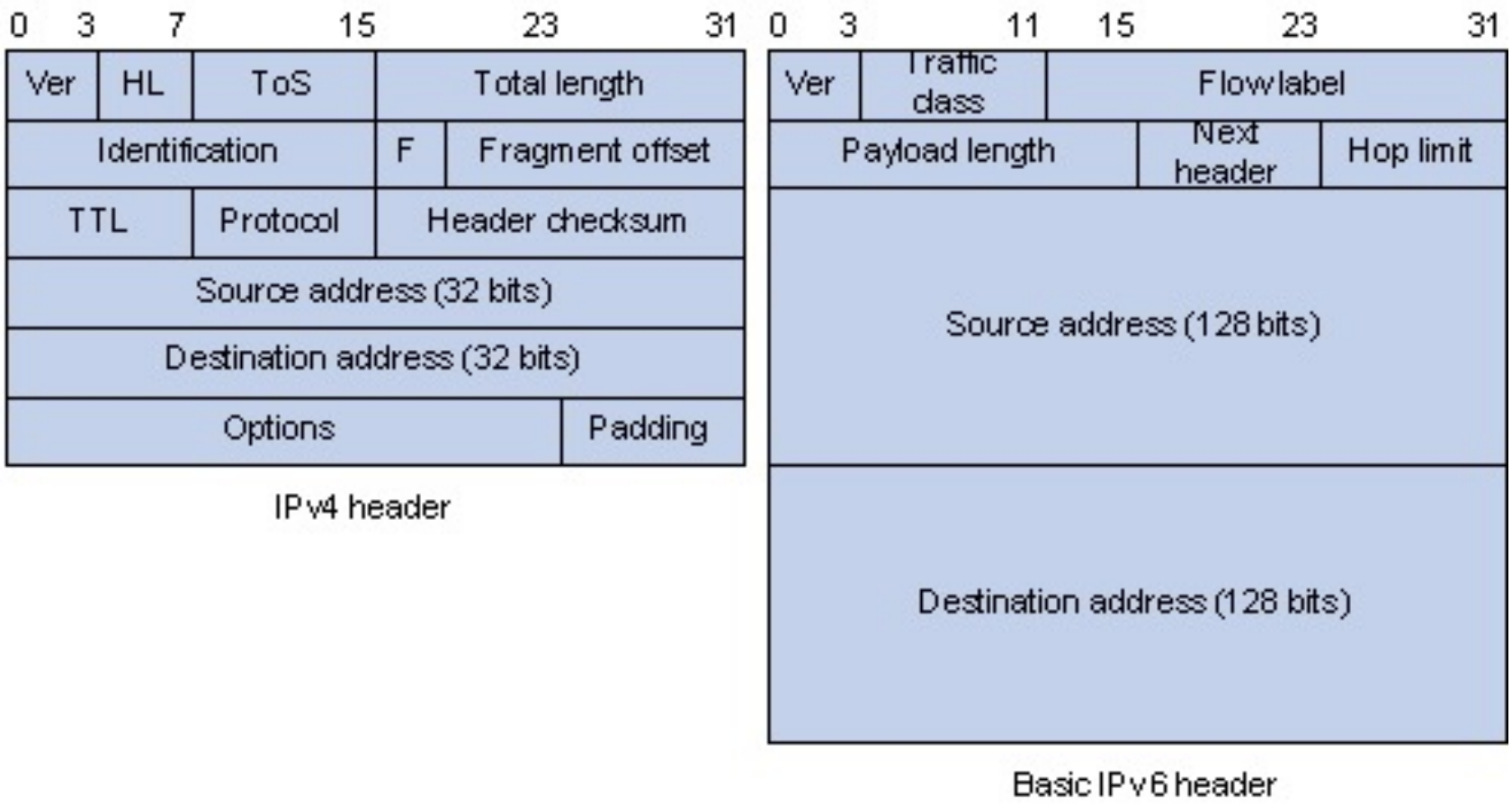
Characteristics

- extended addresses (128-bit) and new addressing schemes
- new flexible and efficient packet formats
- autoconfiguration („plug-and-play“)
- some ‚IPv4 add-ons ‘ integrated (addr. resolution, group mgmt)
- security and mobility mechanisms integrated
- QoS support

IPv6 Header


32 Bits
Version
Traffic class
Flow label
Payload length
Next header
Hop limit
Source address
(16 bytes)
Destination address
(16 bytes)

IPv4 vs IPv6 Header



IPv6 Addresses

Large addresses

- 128 bits

→(theoretically) 2^{128} or approx. 3.4×10^{38} addresses

Notation

- 8 groups separated by colons of 4 hex digits (16 bits)
- omit leading zeros, groups of zeros

Example

- 2001:0db8:0000:0000:0000:8a2e:0370:7334

→ 2001:db8::8a2e:370:7334.

Various types of addresses

- Unicast (global, link-local, site-local, ...)
- Anycast
- Multicast

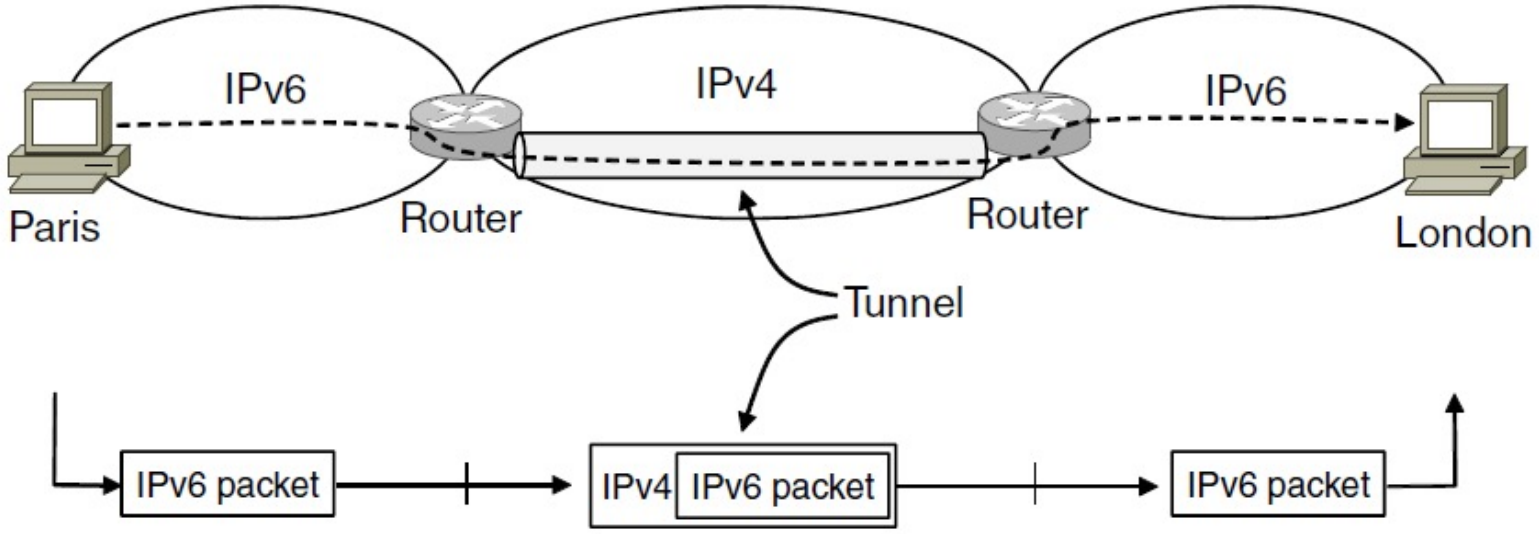
IPv6 Transition

IPv6 is incompatible with IPv4

- How to deploy IPv6?

Several approaches, e.g.,

- Dual stack (IPv4 and IPv6 on a host)
- Translation (convert packets)
- Tunneling (carry IPv6 over IPv4)



From: CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall , 2011