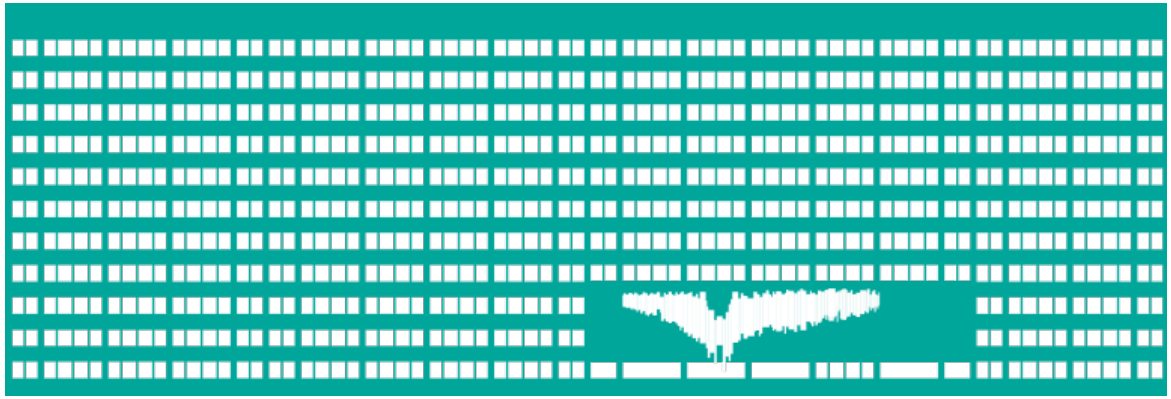
VSB TECHNICAL UNIVERSITY OF OSTRAVA | FACULTY OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE | DEPARTMENT OF COMPUTER SCIENCE

TCP/IP Protocol Suite

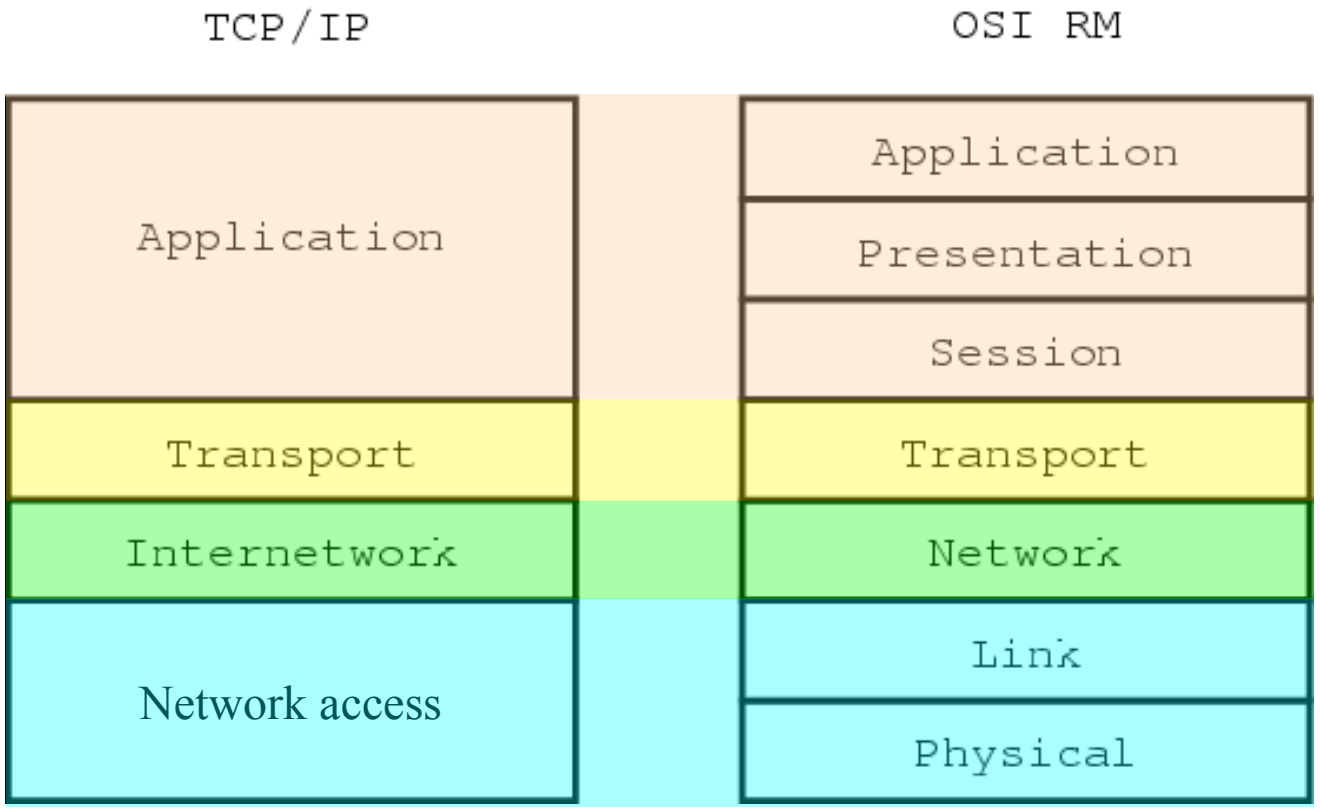


Computer Networks
Lecture 5

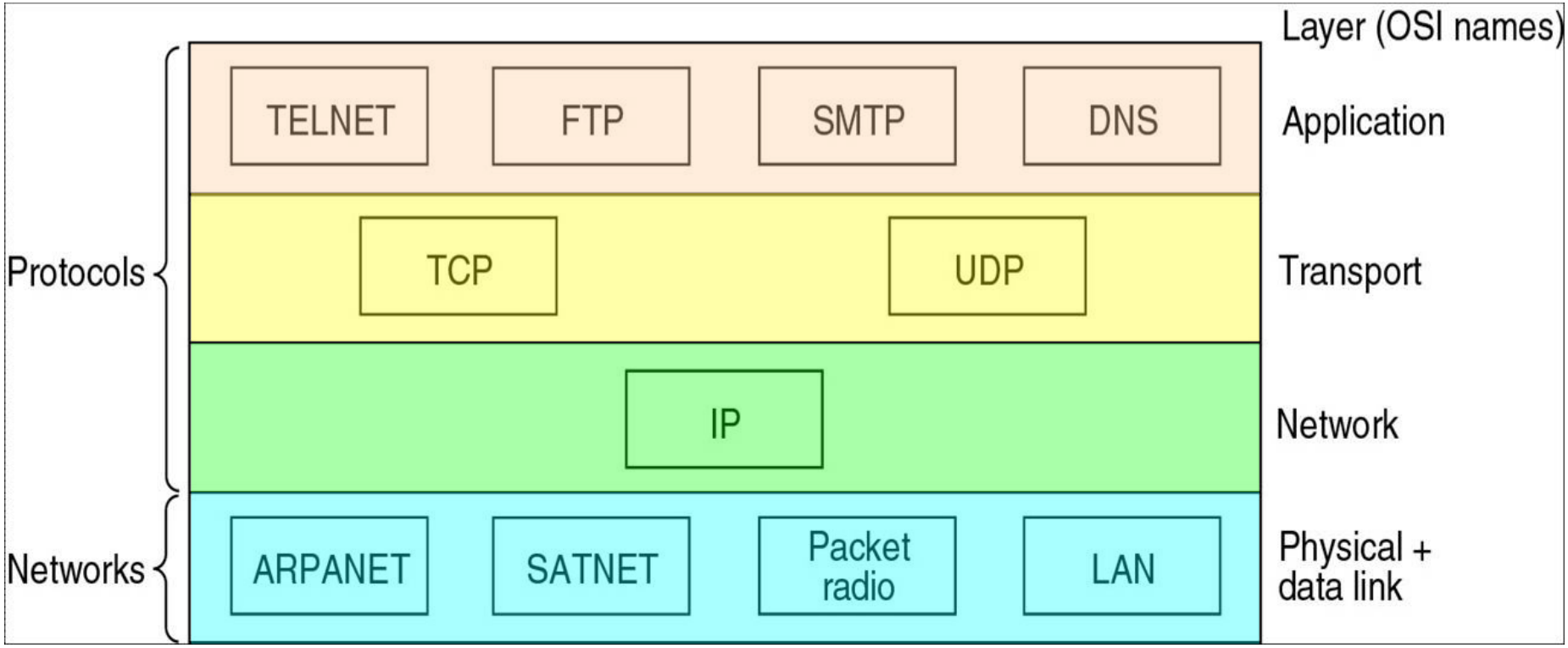
TCP/IP

- Network protocols used in the Internet
  - also used in today's intranets
- TCP – layer 4 protocol
  - Together with UDP
- IP - layer 3 protocol

TCP-IP Layered Model and its Comparison with OSI-RM



Layered TCP/IP Model

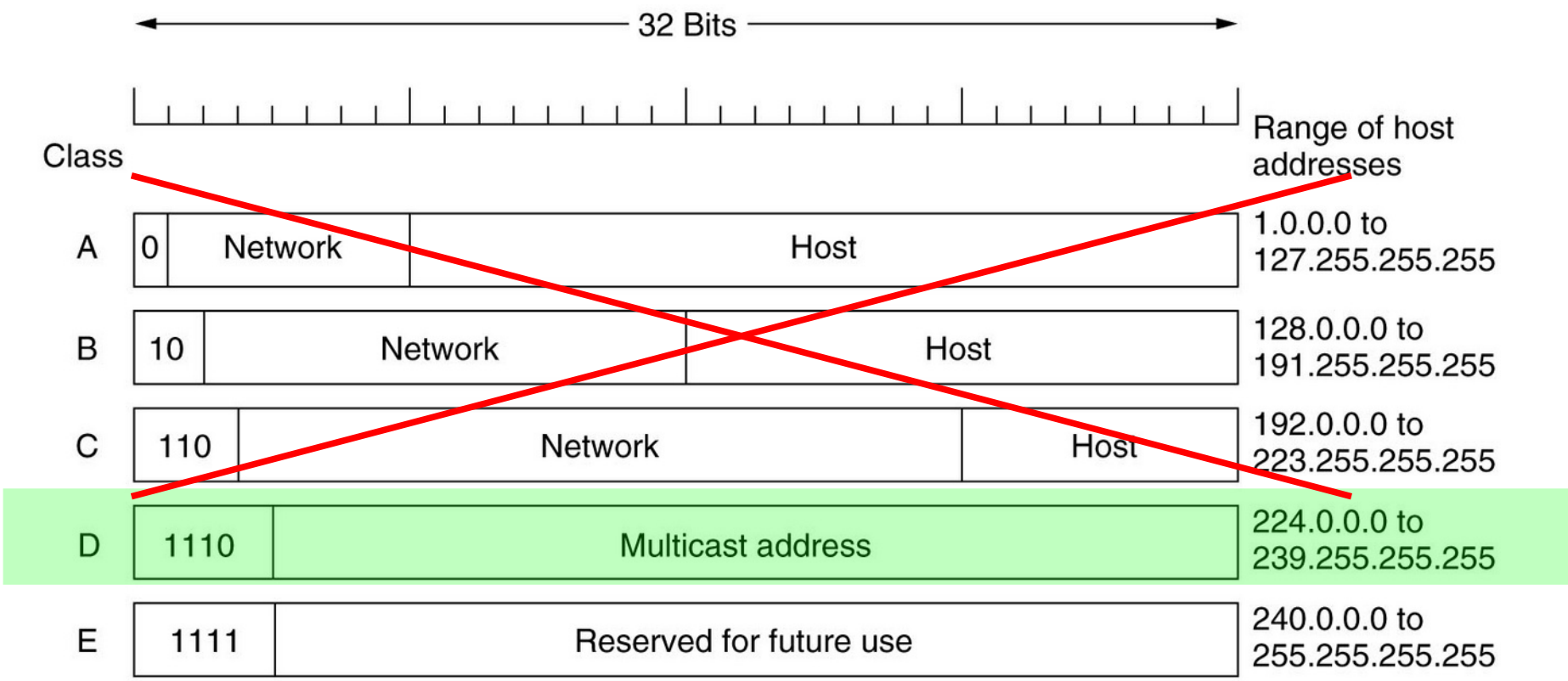


IP Addressing

- 32b addresses (X.X.X.X)
  - Every L3-aware network interface has to have its own IP address
    - e.g. stations and router interfaces
- IP address is divided into network address part and node address part
  - All stations on the same LAN segment (layer 2 broadcast domain) have the same value in the network address part (network prefix)
  - Routers do not have to keep track of all stations' addresses, they just store addresses of individual networks

    → limits the number of records in routing tables

Classes of IP Addresses (used in past)



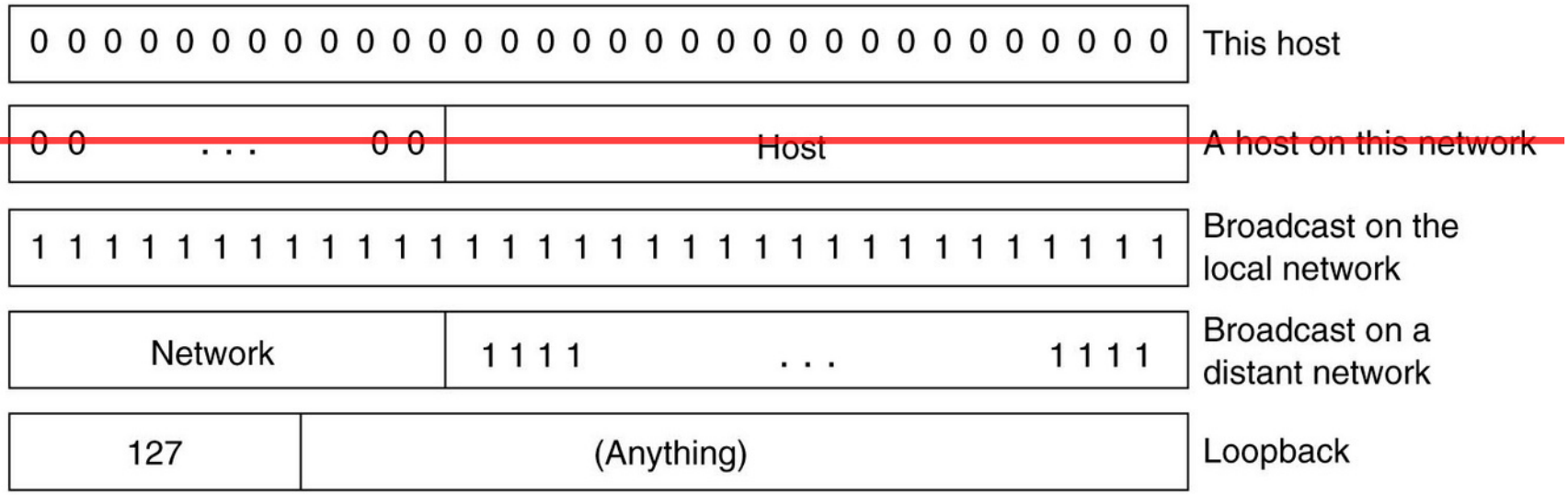
Classless Addressing

- Network prefix of arbitrary length may be allocated
- Classless address has to be accompanied with the subnet mask that specifies the network prefix length
- Classful addresses are no more used at all
  - Classless Inter-Domain Routing (CIDR, RFC 4632)
  - Records of the routing table with the same prefix may be aggregated (supernetting)

IP Addresses Allocation

- Addresses are allocated by the regional Internet Registry (RIPE for Europe)
  - Electronic request form – mediated by the ISP
- Addresses were allocated regardless of the geographical location originally
- Later, the hierarchical addressing was established and allocate network prefixes of the lengths that are really needed
  - Network prefix may be subnetted again
- Private networks may utilize address ranges reserved for private use, but has to avoid leakage of private addresses to the Internet (RFC 1918)
  - 10.0.0.0/8, 172.16.0.0/12 (172.16.*-172.31.*), 192.168.0.0/16, NAT is commonly used to connect such private networks to the Internet
  - Link Local: 169.254.0.0/16 (RFC 3927)

Special IP Addresses



- This host – only as autoconfiguration source address
- Universal broadcast: 255.255.255.255
- Multicast: 224.x.x.x - 239.x.x.x

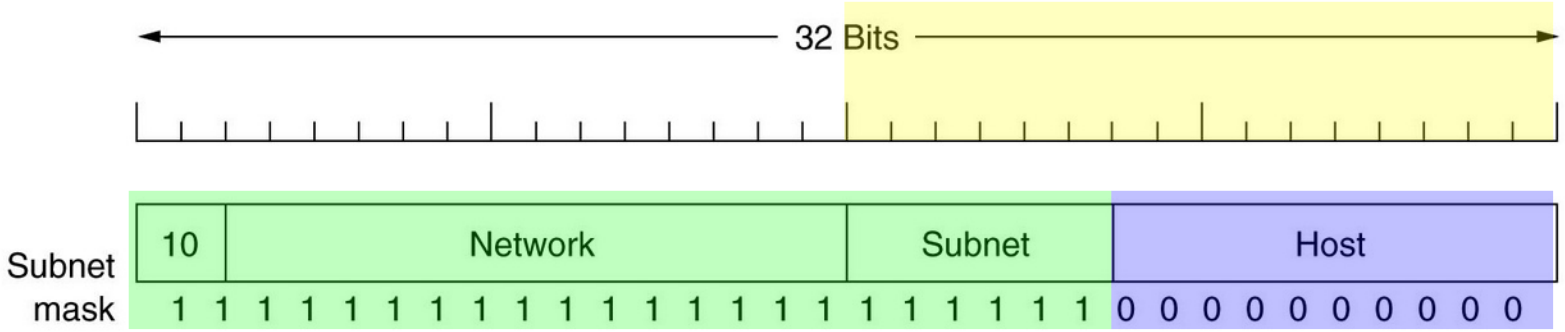
Additional reserved IP addresses

- IETF Protocol Assignments (RFC 5736)
  - 192.0.0.0/24
- Test Networks, documentation (RFC 5737)
  - 192.0.2.0/24 – TEST-NET-1
  - 198.51.100.0/24 – TEST-NET-2
  - 203.0.113.0/24 – TEST-NET-3
- 6to4 Relay Anycast address (RFC 3068)
  - 192.88.99.0/24, e.g. 192.88.99.1 – quite broken, depr.
- Network Interconnect Benchmark (RFC 2544)
  - 198.18.0.0/15
- Shared Address Space (carrier-grade NAT, RFC 6598)
  - 100.64.0.0/10

Subnetting

- Allows to divide network prefix between multiple segments
  - Every segment has to be given an unique subnet address
- The part of the IP address allocated originally for specification of the node is further divided into subnet ID and node ID.
- Address may be split at any bit position according to the required numbers of network stations

Subnet Mask



- Specifies how many bits of the (subnetted) address represent network+subnet
- Binary one at the particular position indicated that the corresponding bit of the IP address belongs to the network+subnet part

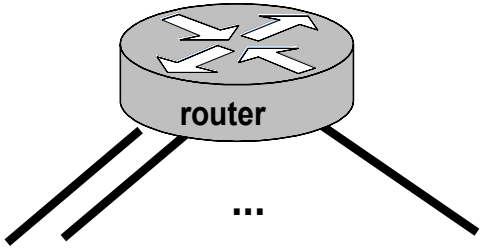
Practical Usage of Subnetting

- Division of the allocated prefix between given number of segments
  (with potentially different number of stations each)
  - Reserved addresses and router interface addresses have to be taken into account
- Specification of the maximum length of the address prefix to ask ISP for needed for addressing of a network with a given number of segments and numbers of stations on individual segments
- WAN addressing plan
  - According to a given network topology number of stations on individual segment

Addressing with Constant Subnet Mask (not used for public IPv4 address ranges now, included for historical reasons)

Division of Address Range (1)

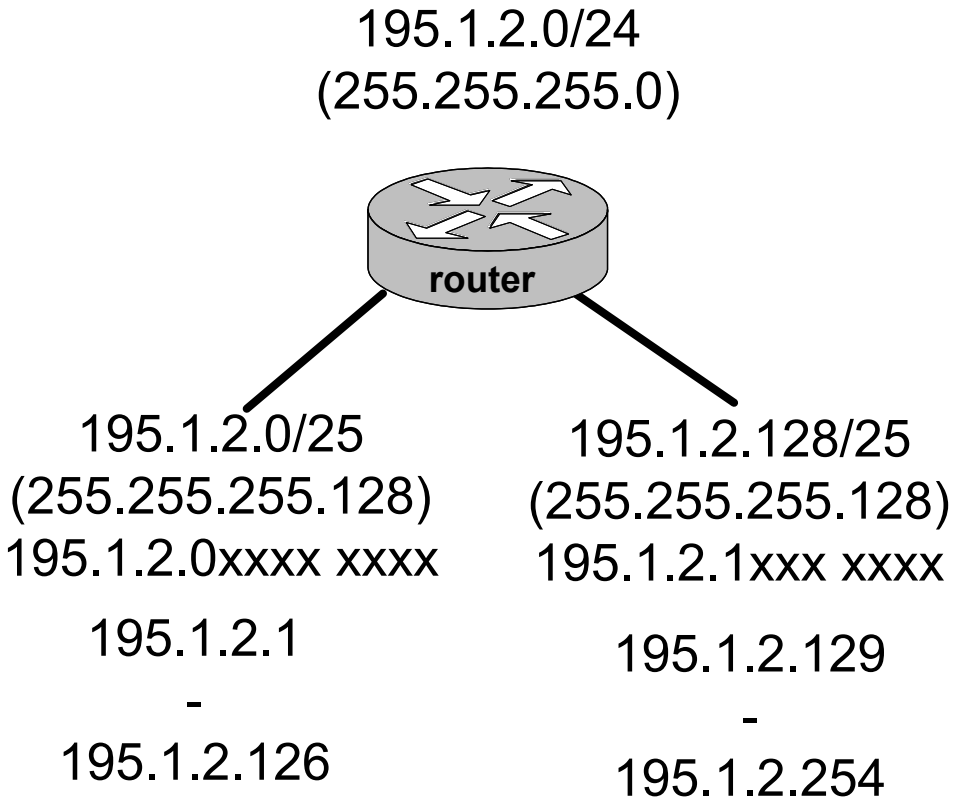
158.196.0.0/16
(255.255.0.0)



...

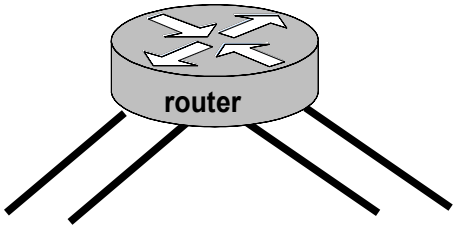
158.196.0.X/24 (255.255.255.0)	158.196.1.X/24 (255.255.255.0)	158.196.255.X/24 (255.255.255.0)
158.196.0.1 - 158.196.0.254	158.196.1.1 - 158.196.1.254	158.196.255.1 - 158.196.255.254

Division of Address Range (2)



Division of Address Range (3)

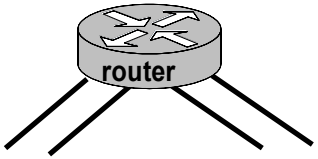
195.1.2.0/24
(255.255.255.0)



195.1.2.0/26	195.1.2.64/26	195.1.2.128/26	195.1.2.192/26
(255.255.255.192)	(255.255.255.192)	(255.255.255.192)	(255.255.255.192)
195.1.2.00xx xxxx	195.1.2.01xx xxxx	195.1.2.10xx xxxx	195.1.2.11xx xxxx
195.1.2.1	195.1.2.65	195.1.2.129	195.1.2.193
-	-	-	-
195.1.2.62	195.1.2.126	195.1.2.190	195.1.2.254

Division of Address Range (4)

195.1.2.160/27
(255.255.255.224)
195.1.2.101x xxxx



195.1.2.160/29
(255.255.255.248)
195.1.2.1010 0xxx
195.1.2.161
-
195.1.2.166

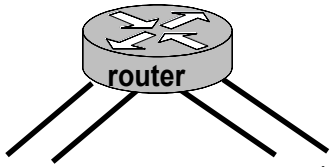
195.1.2.168/29
(255.255.255.248)
195.1.2.1010 1xxx
195.1.2.169
-
195.1.2.174

195.1.2.176/29
(255.255.255.248)
195.1.2.1011 0xxx
195.1.2.177
-
195.1.2.182

195.1.2.184/29
(255.255.255.248)
195.1.2.1011 1xxx
195.1.2.185
-
195.1.2.190

Division of Address Range (5)

195.1.8.0/21
(255.255.248.0)
195.1.0000 1xxx. xxxx xxxx



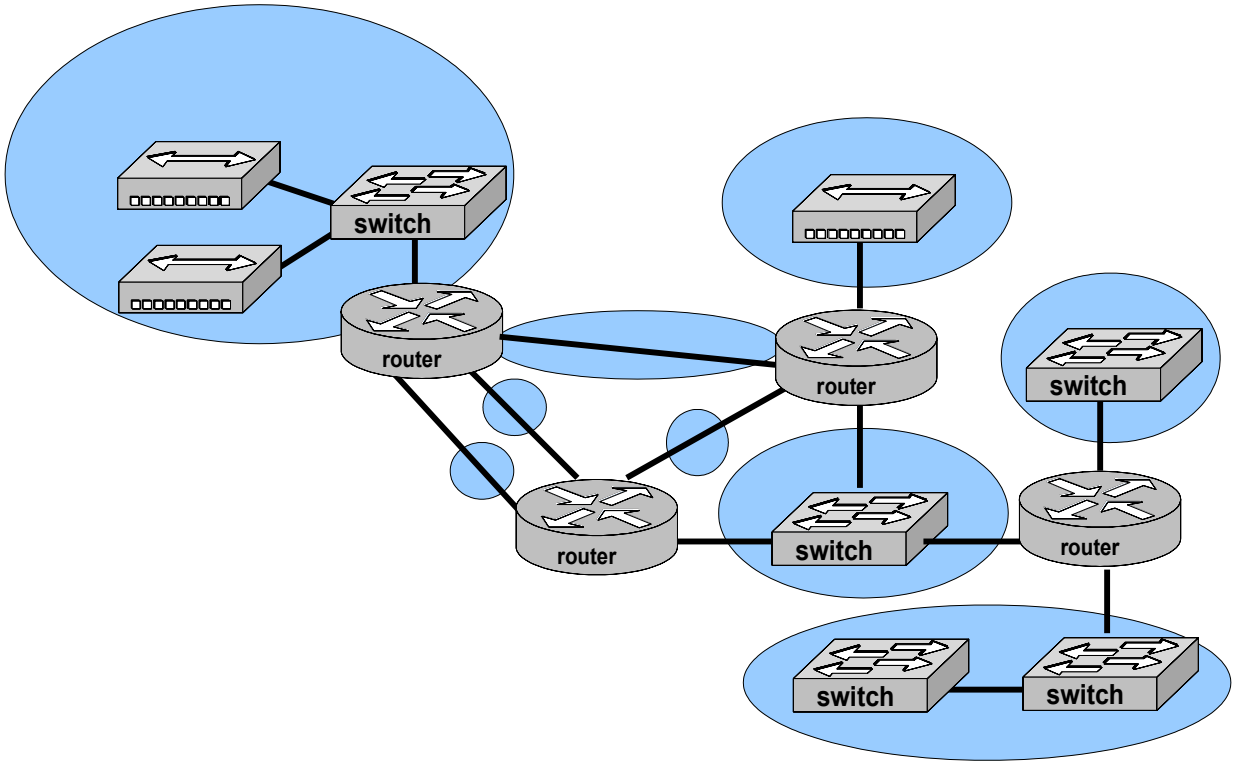
195.1.8.0/23
(255.255.254.0)
195.1.0000 100x. xxxx xxxx
195.1.8.1
-
195.1.9.254

195.1.12.0/23
(255.255.254.0)
195.1.0000 110x. xxxx xxxx
195.1.12.1
-
195.1.13.254

195.1.10.0/23
(255.255.254.0)
195.1.0000 101x. xxxx xxxx
195.1.10.1
-
195.1.11.254

195.1.14.0/23
(255.255.254.0)
195.1.0000 111x. xxxx xxxx
195.1.14.1
-
195.1.15.254

WAN Addressing Plan



- Subnets are separated by Layer 3 devices
  - Routers, stations (not by switches and hubs)

Constraints of the Subnetting

- The minimum number of bits of the node part is 2
  - As we need to represent a subnet (all 0s in host part) and all hosts on the subnet, i.e broadcast (all 1s in host part)
- "Subnet zero" with all 0s in subnet part had been unused in the past but is used normally today
  - Some routers require to explicitly permit usage of subnet zero
- A subnet with all ones in the subnet address part may be also used normally today
  - The usage was not recommended in the past to avoid its address misinterpretation as directed broadcast

Variable-Length Subnet Mask (VLSM) Addressing

See also Czech example at:
http://www.cs.vsb.cz/grygarek/SPS/lect/VLSM/VLSM.html

VLSM – An Example

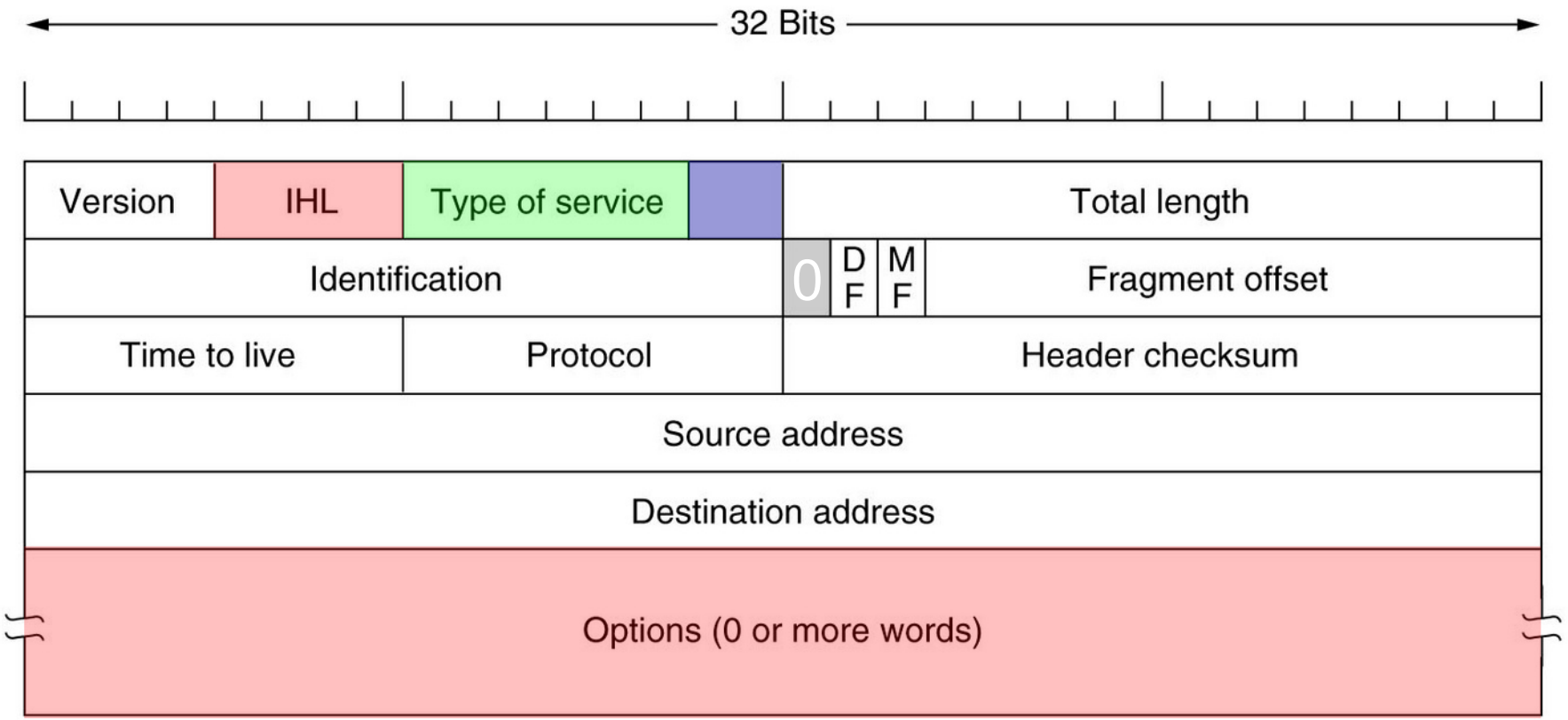
- Let's divide a 1.2.3.0/24 prefix over 4 network segments with 100,50,20 and 10 stations:
  - S1 (100-7b): 1.2.3.1xxxxxxx/25
  - S2 (50-6b): 1.2.3.01xxxxxx/26
  - S3 (20-5b): 1.2.3.001xxxxx/27
  - S4 (10-4b): 1.2.3.0001xxxx/28
- A tree or a rectangle may be used to represent the allocated prefixes

The Internet Protocol

IP – Internet Protocol

- Operates on OSI Layer 3
- Allows to send independent packets between stations of the internetwork
- Unreliable connectionless service
- Defined in RFC 791, 1042 and 894
- Version 4 is still being used today
- A transition to version 6 is ongoing

IPv4 Header



Redefined as Differentiated Services Code Point, DSCP (RFC 2474)

Used for Explicit Congestion Notification (RFC 3168)

Packet Fragmentation

- Applied when the packet has to be routed over a link with insufficient maximum length of data field of the frame (Maximum Transfer Unit, MTU)
  - Either the source station or any router (IPv4) may fragment the packet
  - The packet is reassembled by the destination
    - as fragments may travel along various paths
  - Fragments are grouped together according to the Identification header field
    - The correct ordering is ensured by Fragment Offset field
    - The last fragment does not have More Fragments flag set
- The convention requires all Internet links to support MTUs of at least 576 B

The TCP/IP Supporting Protocols

ARP - Address Resolution Protocol

- Maps destination IP addresses to corresponding MAC addresses
  - ARP Requests with an IP address in question are broadcasted when a corresponding MAC address is needed
    - In addition, a mapping between source MAC and IP address is placed into the request to update ARP caches of all receiving stations and avoid further broadcasting
  - A stations with a required IP address replies with its MAC address
  - The requesting station caches the result
- Works between L2 and L3 (reserved EtherType)

What Destination Addresses will ARP Work for ?

- Used only to resolve MAC addresses for IP addresses on the same segment
  - Stations on the same segment
  - Default gateway address

- ARP is also used to verify whether the configured IP address does not conflict with any other existing address on the segment

ICMP - Internet Control Message Protocol

- Control and information messages
  - Informs about non-standard events during packet transmission
- Carried in data part of IP packets

ICMP Messages (1)

- Echo request , echo reply
- Destination unreachable
  - network, host, port, protocol unreachable, fragmentation prohibited but necessary
  - + administratively prohibited
- Time exceeded
  - TTL=0 or reassembly timeout expired
- Redirect
- Parameter problem

ICMP Messages (2)

Newer messages (not supported by all devices)

- Source quench
  - Asks the source station to lower the speed of packet generation (at the receiver's buffers are becoming full)
- Address mask request
- Address mask reply
- Router solicitation
- Router advertisement

Traceroute: Determination of Packet Route Across the Network

- A command implemented in most OSes
- Allows to detect all routers along the path to the destination network
- Uses value of TTL header field
  - Still increments the TTL value (starting from1)
  - Records addresses from which ICMP Time Exceeded message arrives
    - UDP on non-existent port is used as a probe packet on some Unix implementations

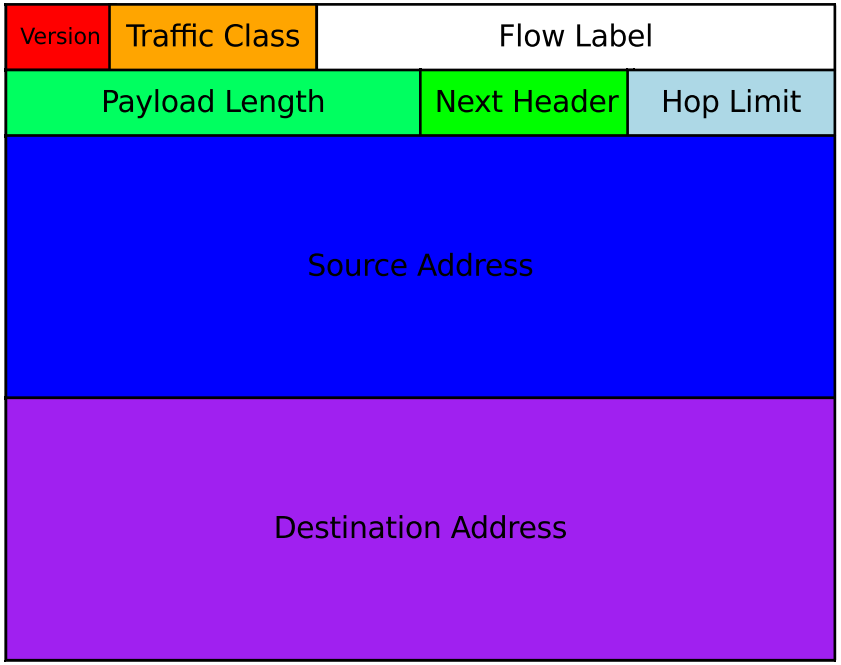
IP version 6

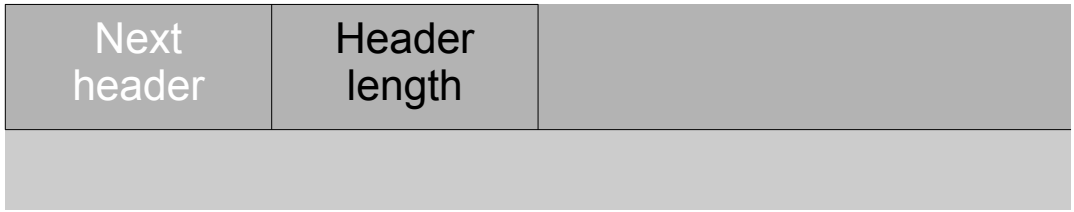
IPv6 Addressing

- 128 bit addresses
  - Written as hexadecimal numbers , e.g. FEDC:0A98:7654:1230:0000:0000:7546:3210
  - Leading zeros in each block may be omitted
    - FEDC:**A98**:7654:1230::7546:3210
  - No broadcasts (only multicasts)
  - Introduction of anycast
  - Hierarchical addressing (aggregable addresses)
- Global and link local addresses
- Stateless address autoconfiguration (SLAAC)
  - Routers advertise the local network prefix, stations append their own MAC address
  - DHCPv6 may be used for extra features
    - e.g. prefix delegation

IPv6 Header

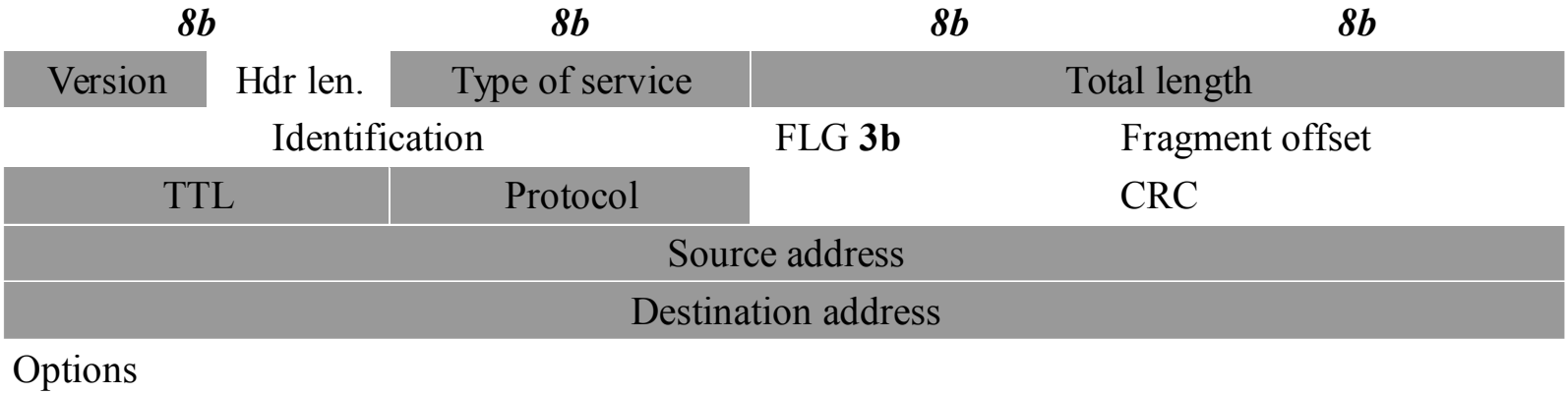
- Simplified in contrast with IPv4
- Header chaining
  - Hop-by-hop options
  - Routing header
  - Fragmentation header
  - Encapsulating Security Payload
  - Authent. header
  - Destination options



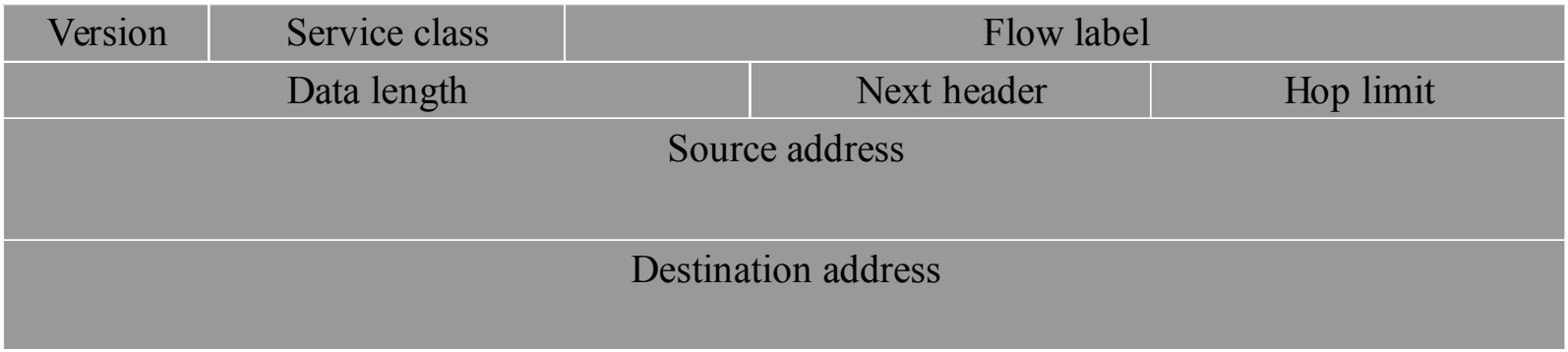


Comparison of IPv6 and IPv4 Header

IPv4



IPv6



- Question: Which one is longer :-) ?

Other Important Differences

- No fragmentation on routers
  - Path MTU discovery procedure have to be applied
  - Only source may fragment packets
    - usage of Fragmentation header
- Optimized IP Option processing
- Support for jumbograms
- No ARP – ICMPv6 used instead
- DNS extensions – AAAA and ipv6.arpa PTR record

IPv4 and IPv6 Coexistence and Interopearation

- Expected to co-exist together for many years
  - Static tunneling
  - Dynamic tunneling (Teredo, etc.)
- Interoperability options
  - Dual-stack hosts
  - Protocol translation
    - includes DNS manipulation
- IPv4 address range is treated as a subset of IPv6 range
  - IPv4-compatible addresses (the latest standard…)

Reserved IPv6 addresses

- Special addresses (RFC 4291)
  - ::1/128 – loopback
  - ::/128 – unspecified
- Discard-Only Address Block (RFC 6666)
  - 100::/64 – blackhole the traffic
- IETF Protocol Assignments (RFC 2928)
  - 2001::/23 – unless specified differently on next slide
- Test Networks, documentation
  - 2001:2::/48 – benchmarking (RFC 5180)
  - 2001:db8::/32 – documentation (RFC 3849)
- Local addresses
  - fe80::/10 – Linked-Scoped Unicast (RFC 4291), i.e. link local addresses
  - fc00::/7 – Unique-Local (RFC 4193)

Multicast IPv6 addresses

- Reserved block – ff00::/8
  - 4th byte in first block indicates scope of the multicast
    - 1 – interface local
    - 2 – link local
    - 3 – IPv4 local
    - 4 – administrative domain
    - 5 – site
    - 8 – organization
    - e – global
- Example multicast groups (link local)
  - ff02::1 – all IPv6 devices on local segment
  - ff02::2 – all IPv6-capable routers
  - ff02::5 – all OSPFv3 routers
  - ff02::9 – all RIPng routers

Additional IPv6 transition addresses

- IPv4 compatibility
  - ~~::0:0/96 IPv4-Compatible IPv6 Address (RFC 2373)~~
  - ::ffff:0:0/96 – IPv4 mapped address (RFC 4291)
    - ::ffff:203.0.133.63 representation is common.
  - 64:ff9b::/96 – IPv4-IPv6 Transl. Address (RFC 6052)
- Transitional technologies
  - 2001::/32 – Teredo (RFC 4380) – may be deprecated
  - 2002::/16 – 6to4
    - IPv4 6to4 anycast 192.88.99.0/24 is deprecated since 2015
- Dual-stack Lite Deployment (RFC 6333) – IPv4
  - 192.0.0.0/29 – IPv4 carried in IPv6 tunnel to carrier-grade (IPv4-IPv4) NAT

Integrated Technologies

- IPSec (encryption, authentication)
- Mobile IP
- Multicasts

- SLAAC – Stateless address autoconfiguration
  - IPv6 hosts can configure themselves automatically when using the Neighbor Discovery Protocol and ICMPv6 router advertisements
  - EUI-64 – for IPv6 address autoconfiguration:
    0A:CD:12:34:56:78 →
    0(A XOR 2)CD:12:FF:FE:34:56:78 →
    FE80::(A XOR 2)CD:12FF:FE34:5678
  - Privacy extension (RFC 4941)

ICMPv6 Messages

- Echo Request & Echo Reply
- Router Solicitation & Router Advertisement
- Neighbor Solicitation & Neighbor Advertisement
- Multicast Router Advertisement, Solicitation & Termination
- Multicast Listener Query, Report, Listener Done
- Destination Unreachable, Time Exceeded
- Packet Too Big
- Parameter Problem, Redirect
- Private Experimentation, …

TCP/IP Transport Layer: UDP and TCP

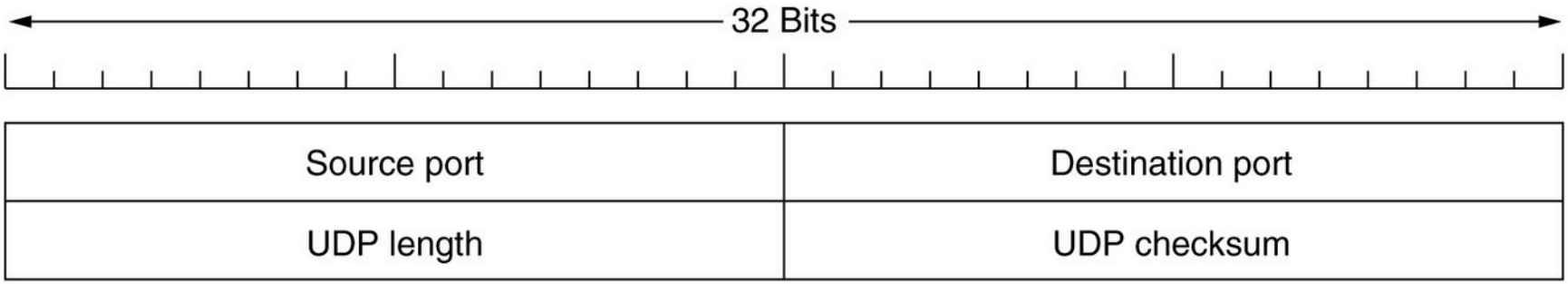
Ports

- The transport-layer entity (i.e. process or service running on a particular machine) is identified by the machine's IP address and port number (which is local to the particular machine)
  - 16bit (0-65535), separately for TCP and UDP
    - 1-1023: well-known services
    - 1024-4096: other registered applications
    - > 49152 (IANA, often depends on used OS) – client (ephemeral) ports
      - Client ports are usually assigned by OS to the applications
- Note that both destination and source port are used to identify the flow

UDP – User Datagram Protocol

- unacknowledged unreliable datagram delivery service
- Support the broadcast and multicast transmission
- The source and destination processes are identified by IP addresses and ports
- User data are protected by (an optional) checksum
  - IP checksum protect only the IP header

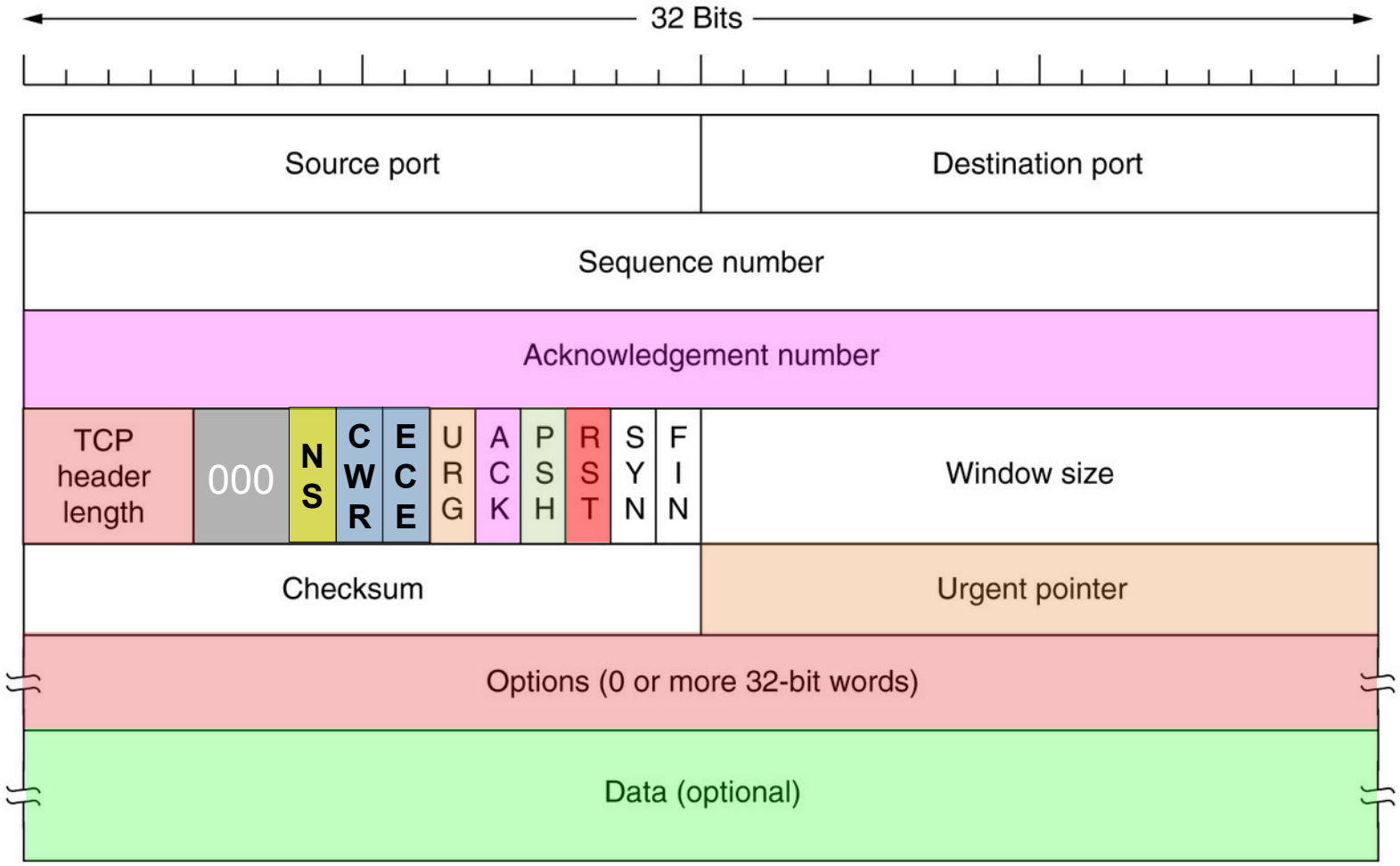
UDP (Pseudo)Header



TCP – Transmission Control Protocol

- Provides a reliable duplex communication
  - Over unreliable IP that may drop and duplicate packets and deliver them out of order
- Segments the data stream into packets
  - Inserts sequence numbers of the first byte into each segment
- Uses Sliding window algorithm (go-back-N)
  - Positive (inclusive) acknowledgments, piggybacking, adaptive retransmission timeout calculation
  - Implements flow control by advertising the current remaining capacity of the receiving buffer, the sending window dynamically adapts to it
  - Robust protocol of connection establishment and termination

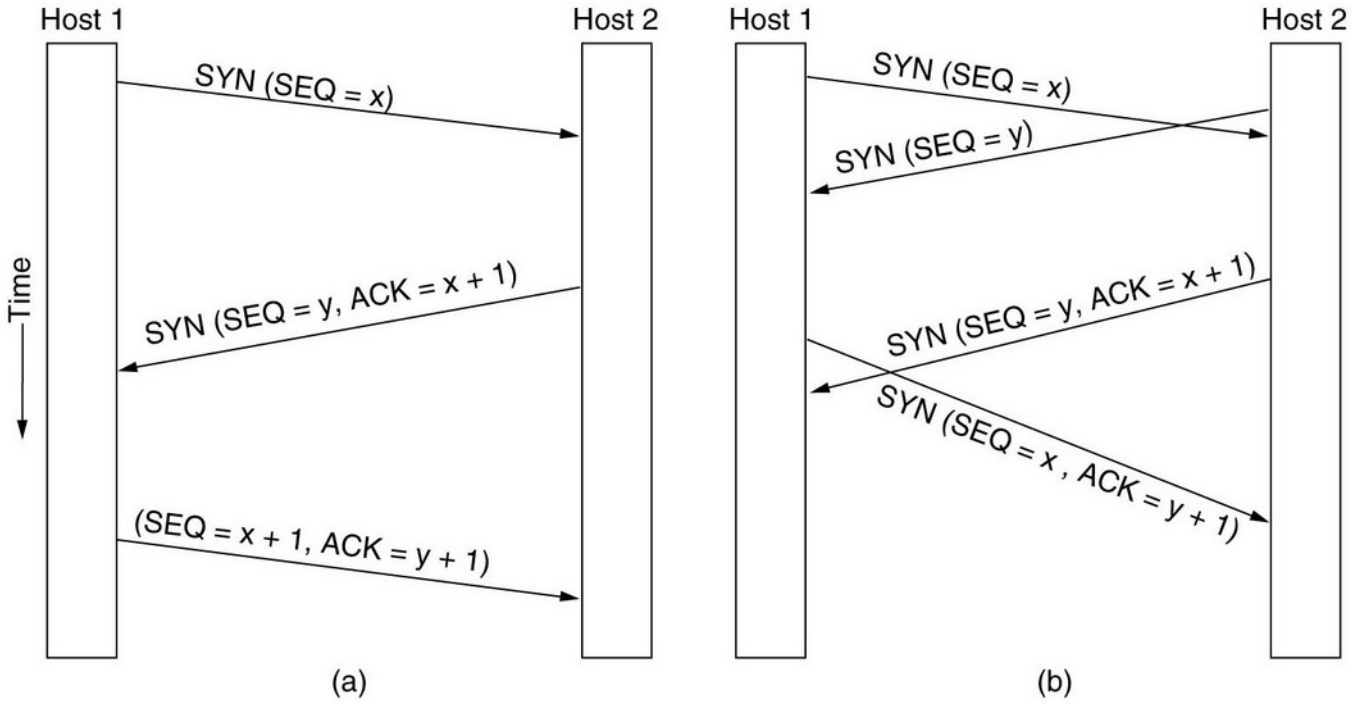
TCP (Pseudo)Header



Used for Explicit Congestion Notification (RFC 3168)

NS – ECN nonce concealed (RFC 3540)

TCP Connection Establishment (1)

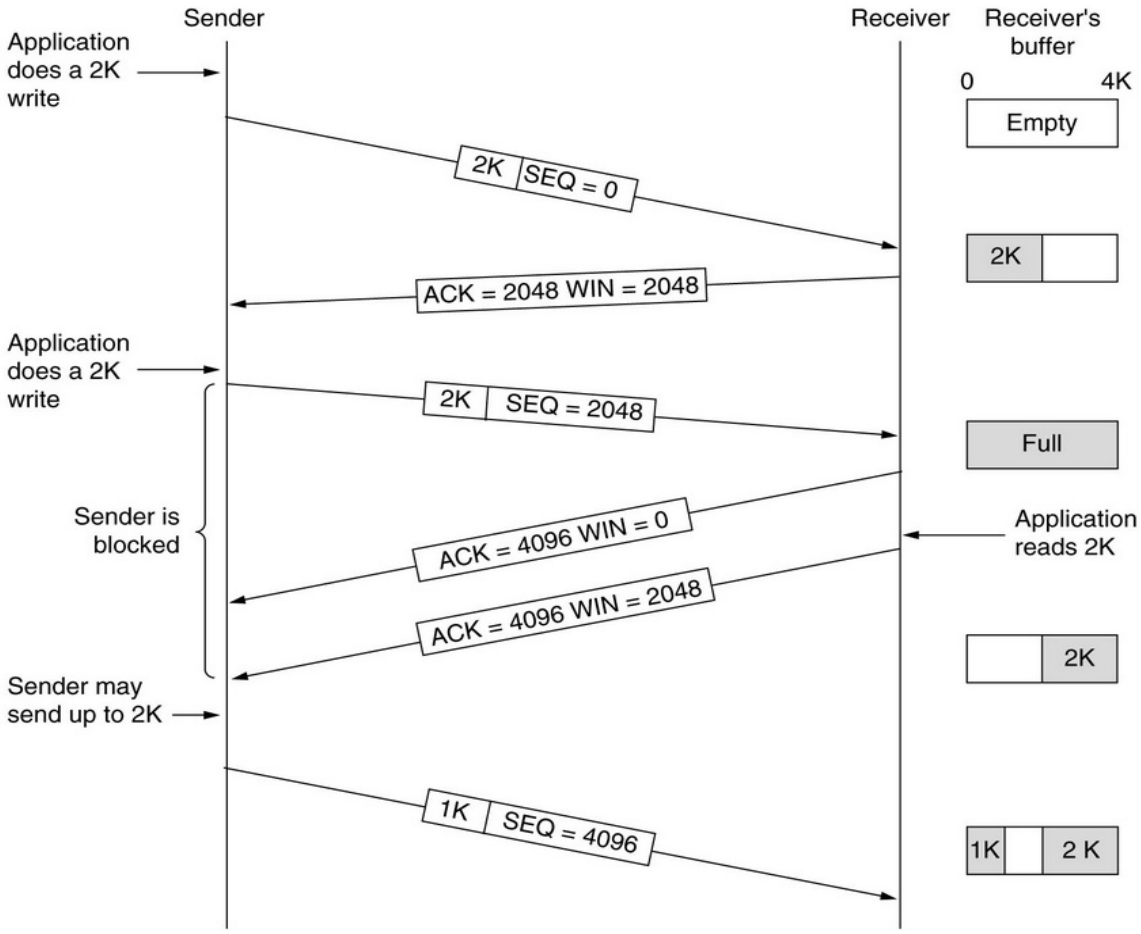


(a) Three-way handshake
(b) Four-way handshake (not seen nowadays)

TCP Connection Establishment (2)

- Three way handshake: SYN, SYN+ACK, ACK
  - Initial sequence number negotiation (independently for both directions)
  - ISNs are „random“ to avoid confusing of the receiving station by delayed packets from previous connection between the same stations
- Opening of a connection by both sides simultaneously results in a single connection

TCP Connection in Action (flow control)



TCP Connection Termination

- Both sides have to close the connection independently
  - Half-closed state (other side can still send data)
  - FIN+ACK (from both sides)
- Any side may close the connection first
- Immediate connection termination (reset) – RST flag
  - No messages need to be exchanged after that