Internet

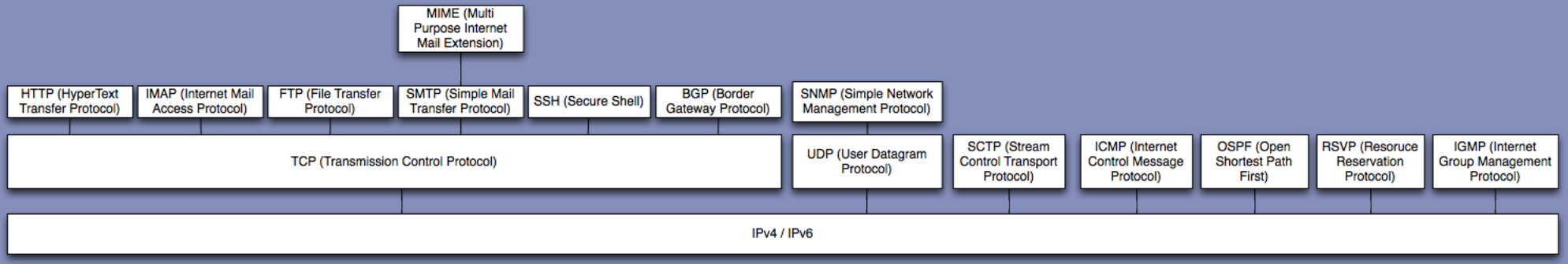
OSI Model

OSI = Open System Interconnection

Level	Layer	Example
7	Application	HTTP, FTP, SMTP
6	Presentation	MIME, SSL
5	Session	TCP (session establishment), RTP
4	Transport	TCP, UDP, SCTP
3	Network	IPv4, IPv6
2	Data Link	PPP, LLC
1	Physical	MLT, QAM

TCP/IP not as layered as the OSI model

Internet Protocol Stack

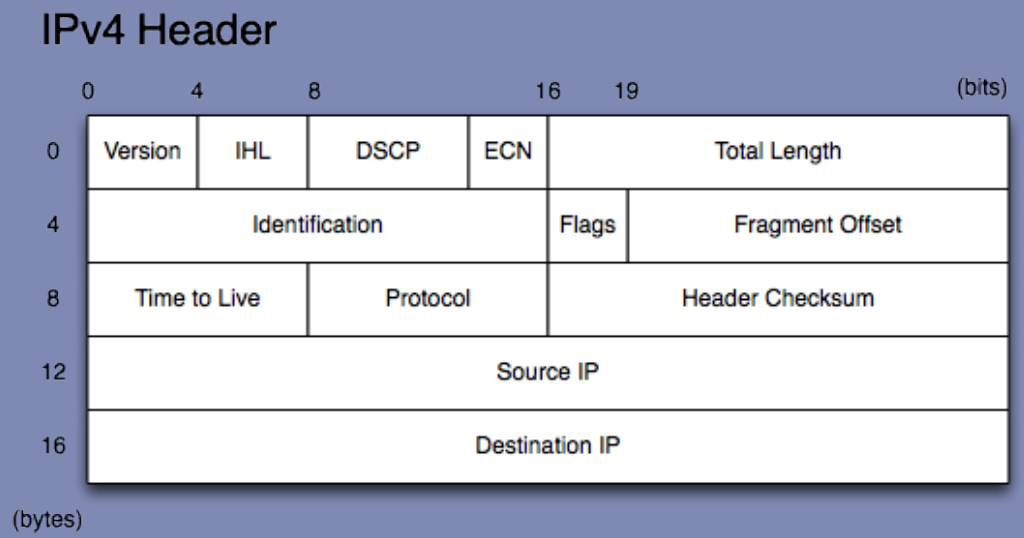


Internet

- IP: Connectionless / datagram service between 2 endpoints
- TCP/IP is roughly layered in 5 levels.
  - Application: communication between processes and applications between hosts.
  - Transport: Data transfer service, hides network details from application layer.
  - Internet: Packet routing
  - Network Access: Connection with subnet (LAN)
  - Physical: Cable standards, modulation.

IPv4

- **32 bit** addresses
- Written using “decimal **octets**” (4 times 8 bits separated by .), i.e. 132.229.16.174
- Total size ≤ 65535 octets (≈ 64KB)



Data...

IPv4

- Version (4 bits): version number
- IHL (4 bits): Internet header length in 32 bit words (min = 5 words / 20 octets).
- DSCP(Differentiated Services Control Point)/ECN (Explicit Congestion Notification). Formally known as TOS(8 bits): Type of service, specifies (but is usually ignored):
  - Priority (8 levels)
  - Reliability: normal or high (avoid dropping packets)
  - Delay: normal or low (minimise delay)
  - Throughput: normal or high (maximise throughput)
- Total length (16 bits): packet length in octets
- Identification (16 bits): sequence number for reassembly of IP datagram from fragments.
- Flags:
  - 1 bit = reserved,
  - 1 bit don't fragment (DF), routers should not split up the packet
  - 1 bit more fragments (MF), multiple fragments exist (last packet MF = 0, fragment offset > 0)
- Fragment Offset (13 bits): In number of 8 octet units, where the fragment should go in the datagram.
- Time to Live (8 bits): max number of router hops until packet self destructs, prevents packets from living forever on the Internet when the network is damaged.
- Protocol (8 bits): 6 = TCP, 17 = UDP
- Checksum (16 bits): one's complement addition of all 16 bit words in the header (except checksum)
- Source and destination address (32 + 32 bits): IP address of sender and destination.

IPv4(Optional)

- Header may contain an *options* fields at the end, this may be followed by padding to make the header size dividable by 32 bit.
- Options include: security, strict source and record route, loose source and record route.
- Often, routers will be configured to drop packets containing some of the options.

IPv4 Addressing (Classes)

Earlier (before 1993) IP addresses where divided in classes.

Class	Start bits	Net	Host	Start IP	End IP	Note
A	0	/8	24	0.0.0.0	127.255.255.255	
B	10	/16	16	128.0.0.0	191.255.255.255	
C	110	/24	8	192.0.0.0	223.255.255.255	
D	1110	N/A	N/A	224.0.0.0	239.255.255.255	Multicast
E	11110	N/A	N/A	240.0.0.0	255.255.255.255	Reserved

IPv4 CIDR

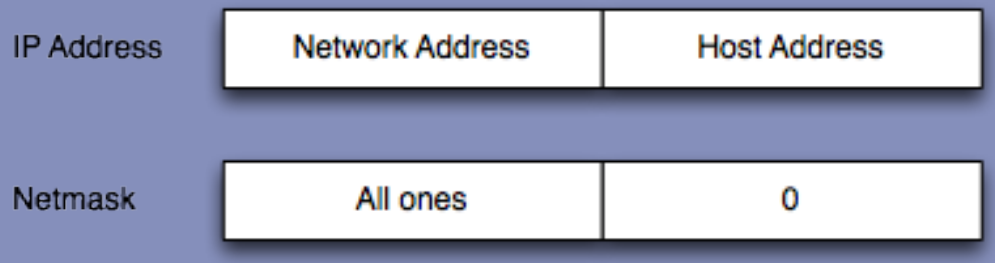
- **Classless Inter-Domain Routing**
- Replaced IP address classes using the CIDR notation.
- IP number registries may assign blocks of different sizes.

IPv4 Addressing (CIDR)

IP addresses are divided in two parts, the network address and the host address.

The network (and host) address can be isolated using a netmask using a bitwise **AND** operation.

Using C/C++ notation:
Network = IP & netmask
Host = IP & ~netmask



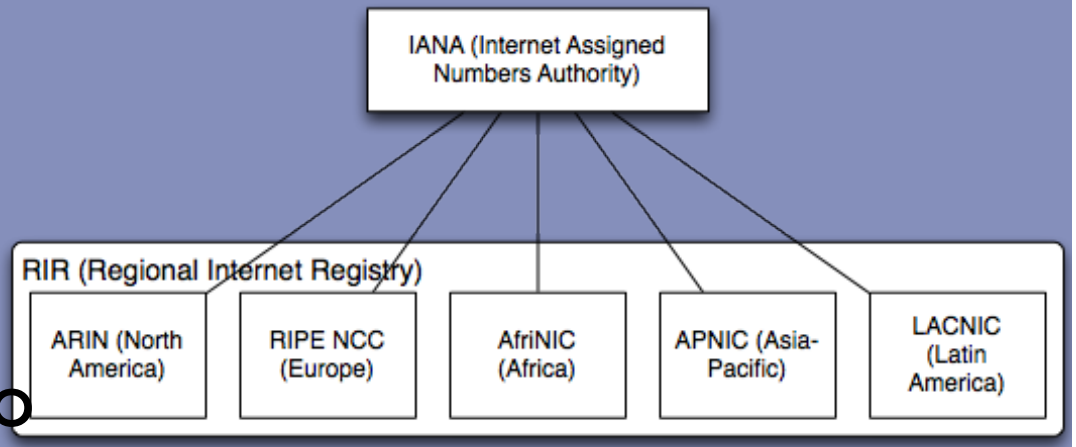
Example:
IP = 192.168.129.3, netmask = 255.255.128.0
Then: network = 192.168.128.0, host part = 0.0.1.3

All practical netmasks start with a sequence of ones followed by zeroes. Therefore, it is common to abbreviate the masks using the CIDR-notation /N, where N is the number of ones in the mask. In the example: 192.168.129.3/17

IPv4 Assignments

IPv4 addresses blocks (mostly /8) handed out by IANA (the central authority) to Regional Internet Registries (RIRs). RIRs allocate address ranges within the geographical regions.

Example: IANA assigns 132.0.0.0/8 to RIR. RIR have the whole block and assigns subnets of this block e.g. 132.229.0.0/16 to Leiden University.



IPv4 Assignments

Some assignments from: http://www.iana.org/assignments/ipv4-address-space/ipv4-address-space.xml

Block	Usage
000/8	Local identification
003/8	General Electric
009/8	IBM
017/8	Apple
051/8	UK Government of Work and Pensions

IPv4 Private Addresses

The Internet Engineering Task Force (IETF) has directed the Internet Assigned Numbers Authority (IANA) to reserve the following IPv4 address ranges for private networks, as published in RFC 1918.

Free for use at home, guaranteed to **not** be routed through the Internet.

Name	Start	End
24 bit block	10.0.0.0	10.255.255.55
20 bit block	172.16.0.0	172.31.255.255
16 bit block	192.168.0.0	192.168.255.255

Often, your ADSL modem will give you an IP like: 192.168.1.2/24, and reserve 192.168.1.1 or 192.168.1.254 as the router's address (192.168.1.255 is reserved for broadcast).

IPv4 Loopback

127.0.0.0/8 is reserved for in host communication and 127.0.0.1 (or any other valid address in the subnet) always represents “your own computer”.

IPv4 Broadcast Address

IPv4 also supports the notion of a broadcast address for the local subnet. This address is your network address plus the host address of all ones.

Example:
IP: 192.168.1.1/24
Then: network = 192.168.1.0/24 and broadcast = 192.168.1.255.

Note that neither the network nor broadcast addresses can represent an individual node.

IPv4 Issues

Vint Cerf (father of IP): “I thought for an experiment 4.3 billion terminations ought to be enough. I didn't know the experiment wasn't going to end.”

IPv4 Issues

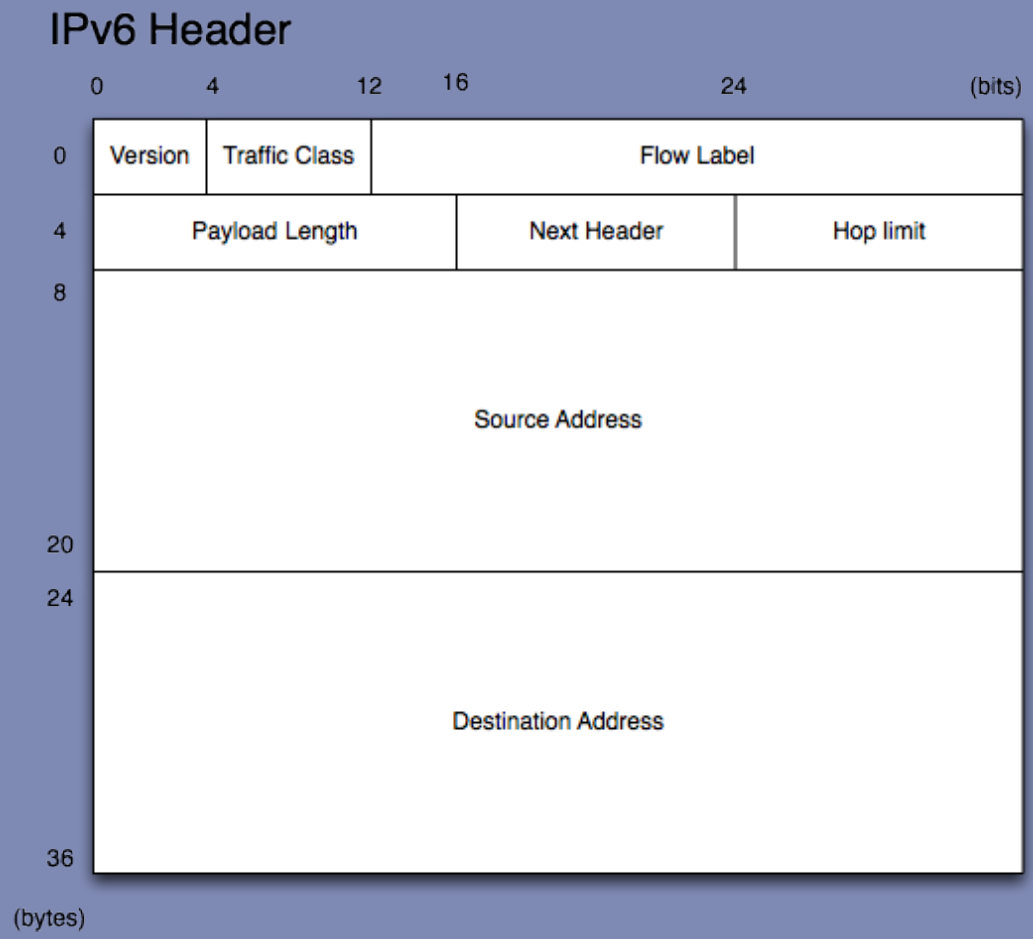
- 32 bit address space is small.
- IANA (allocating blocks to Regional Internet Registries) ran out of /8 address blocks on 2011-01-31.
- APNIC (Asia Pacific RIR) ran out of IPv4 addresses on 2011-04-15
- RIPE NCC (European RIR) ran out of IPv4 addresses on 2012-09-14.
- ARIN (North America) was expected to run out at the end of 2013 / early 2014. Actually they were able to postpone this date till 24 September 2015.

IPv4 Issues

- Lack of addresses leads to workarounds.
- Network Address Translation (NAT)
  - Allows more than one computer to share one public IP address.
  - Major issue for connecting hosts behind two different NATs.
  - Workarounds exists, but are generally complicated.

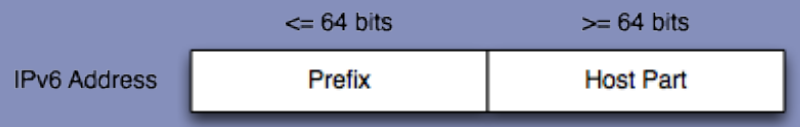
IPv6

- **128 bit** addresses.
- Users don't receive addresses, they receive prefixes.
- NAT not needed at home, every device has a public IP. This also makes IPv6 firewalls very easy to configure for home users.
- Etc.



IPv6

Prefixes of N bits handed out to users. Normally, an ISP will hand out /64 prefixes to home connections. Organisations can expect /56 prefixes (so they can set up their own networks).



IPv6

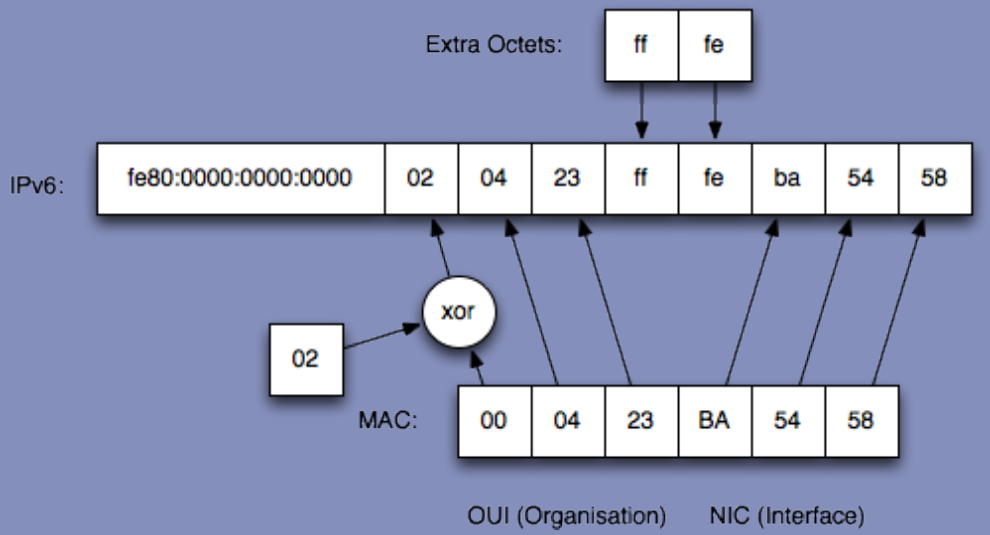
- IPv6 is classless by default.
- Addresses are long and written in hex, may be abbreviated with :: in the case of lots of zeroes.
- fe80::204:23ff:feba:5458 = fe80:0000:0000:0000:0204:23ff:feba:5458
- Auto assign host part of IPv6 from MAC address (00:04:23:BA:54:58)

IPv6 SLAAC

StateLess Address Auto Configuration

fe80::/64 prefix for link local address (in this case, but can be any prefix).

Host part assigned by MAC address with fffe inserted in the middle and the first byte bitwise exclusive or 02



MAC addressing (XOR 2?)

- 6 Bytes addresses grouped into one 3 Bytes Vendor ID or **Organizationally Unique Identifier** (OUI) and 3 Bytes Network Interface Controller (NIC) Address
- Seventh bit is U/L bit indicating whether the MAC address is composed out of OUI and NIC Address (U/L bit is 0 address is universal) or Locally Administered (U/L bit is 1)
- So, if for testing a local MAC address is generated consisting mostly of zero's, then the 7th bit is 1, requiring a manual entering of all bits instead of using ::

IPv6 SLAAC

- Use Neighbor Discovery Protocol (NDP) in order to query for a router and router responds with an IPv6 address prefix.
- Auto create host part of address (see previous slide).
- For privacy reasons, host address can be randomly generated. This prevents the leaking of MAC addresses to the outside world which could be used in order to track a specific machine.

IPv6

- Minimum assignment is a /64 address prefix.
- 128 bits are big numbers...
  - Around 5.1×10^{20} mm^2 surface of the earth.
  - 3.4×10^{38} IPv6 addresses = 6.7×10^{17} addresses per square millimetre.

IPv6 Reserved Addresses

- http://tools.ietf.org/html/rfc5156
- Does not have broadcast functionality
- Loopback = 0:0:0:0:0:0:0:1 = ::1 (::/128)
- IPv4 addresses mapped as v6: "::FFFF:0:0/96"
- LAN: "fe80::/10"
- 6 to 4 tunneling: "2002::/16"
- Multicast: ff00::/8

Mixed Addressing

During the transition of the Internet from IPv4 to IPv6, it is typical to operate in a **mixed addressing environment**.

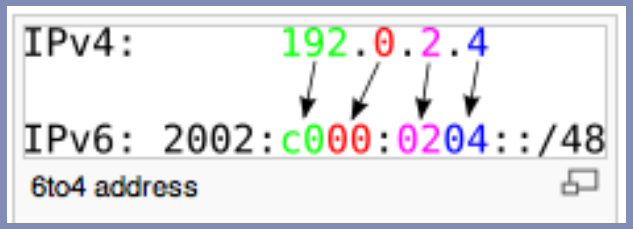
For such use cases, a special notation has been introduced, which expresses IPv4-mapped and IPv4-compatible IPv6 addresses by writing the least-significant 32 bits of an address in the familiar IPv4 dot-decimal notation, whereas all preceding ones are written in IPv6 format. For example, the IPv4-mapped IPv6 address

::ffff:c000:0280 is written as ::ffff:192.0.2.128,

thus expressing clearly the original IPv4 address that was mapped to IPv6.

6to4 tunneling (sending IPv4 packets over an IPv6 network)

- For any 32-bit global IPv4 address that is assigned to a host, a 48-bit 6to4 IPv6 prefix can be constructed for use by that host (and if applicable the network behind it) by appending the IPv4 address to 2002::/16.
- For example the global IPv4 address 192.0.2.4 has the corresponding 6to4 prefix 2002:c000:0204::/48. This gives a prefix length of 48 bits, which leaves room for a 16-bit subnet field and 64 bit host addresses within the subnets.



- Any IPv6 address that begins with the 2002::/16 prefix (in other words, any address with the first two octets of 2002 hexadecimal) is known as a **6to4 address**, as opposed to a *native IPv6 address* which does not use transition technologies.

6to4 Encapsulation (sending IPv6 packets over an IPv4 network)

- 6to4 embeds an IPv6 packet in the payload portion of an IPv4 packet with protocol type 41.
- The IPv4 destination address for the prepended packet header is derived from the IPv6 destination address of the inner packet (which is in the format of a 6to4 address), by extracting the 32 bits immediately following the IPv6 destination address's 2002::/16 prefix.
- The IPv4 source address in the packet header is the IPv4 address of the host or router which is sending the packet over IPv4. The resulting IPv4 packet is then routed to its IPv4 destination address just like any other IPv4 packet

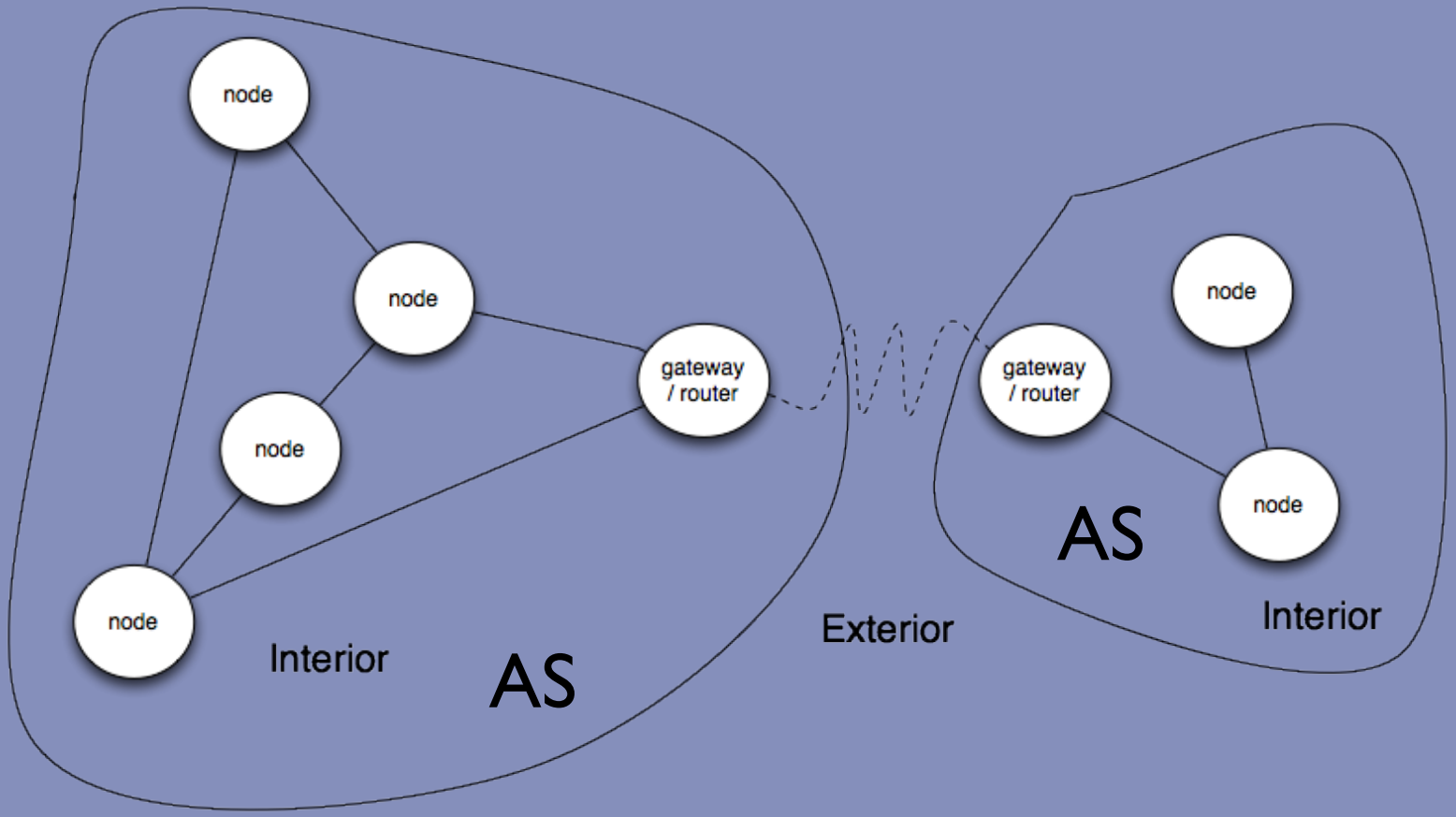
ICMP

- Internet Control Message Protocol
- Two protocols ICMPv4 and ICMPv6

  v4: http://tools.ietf.org/html/rfc792

  v6: http://tools.ietf.org/html/rfc4443
- Used to send network information, e.g. host not reachable.
- End users may use ICMP with the ping/traceroute and ping6/traceroute6 commands for IPv4 and v6 respectively.

Internet Routing



Autonomous System (AS)

A large company which manages its own network and has full control over it. Or an ISP that provides services to local customers.

- **STUB AS.** Only one connection to another AS. Data can be send or received from hosts in the AS to hosts in other AS's. BUT data cannot pass through. So a STUB AS is either source or sink of data transmission. E.g. Small Cooperation.
- **MULTIHOMED AS.** More than one connection to other AS's. No transient data, so again only source or sink of data transmission. E.g. Large Cooperation.
- **TRANSIT AS.** A multihomed AS with transient traffic. E.g. International ISP's. Internet Backbones.

Routing

- Interior routing: Open Shortest Path First (OSPF)
- Exterior routing: Border Gateway Protocol (BGP)

BGP (Border Gateway Protocol)

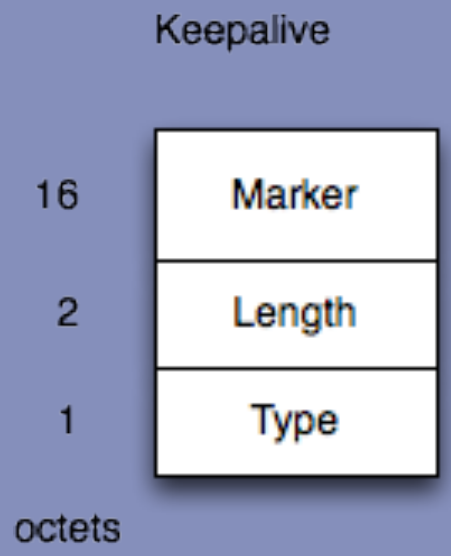
- Preferred for exchange of routing information between gateways.
- Used for:
  - Neighbor acquisition
  - Neighbor reachability
  - Network reachability

BGP

Four message types:

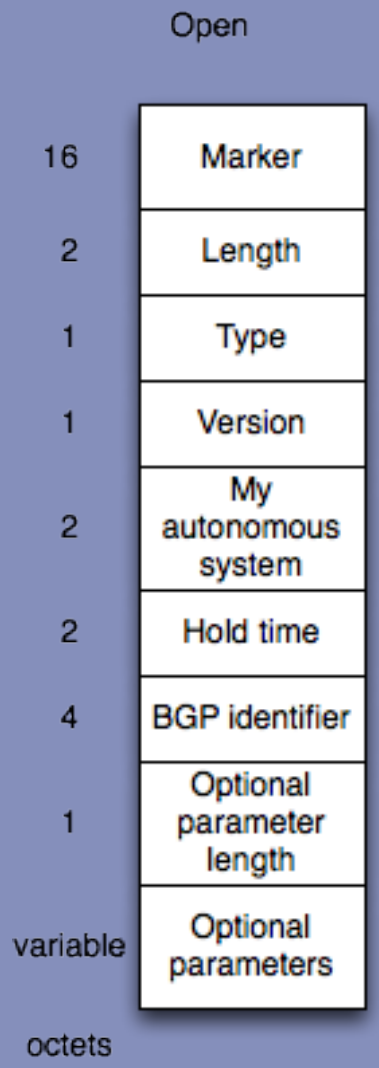
- Open: open neighbor relation
- Update: Transmit info about single route or list with route info
- Keep-alive: Acknowledge open connection or periodic confirmation of neighbor relation.
- Notification: Notifying routers about error conditions.

BGP



- Marker: used for authentication
- Length: in bytes
- Type: Open/update/notification of keepalive.

BGP



- Version: current BGP version (e.g. 4). RFC 4271
- Hold time: max # seconds between keepalive or update messages.

BGP

- BGP exchanges routing information (very complex)
- For example, the update message can include the following info:
  - AS-pad: identity of AS
  - Next hop: IP address of the router in the AS
  - NRLI (BGP Network Layer Reachability Information): a list with the network and IP addresses within the AS.

IP Protocols

- TCP (Transmission Control Protocol): Connection oriented streams of in sequence bytes
- UDP (User Datagram Protocol): connectionless out of order messages
- SCTP (Stream Control Transmission Protocol): in sequence messages, not widely in use (yet)

TCP

- Transmission Control Protocol
- Runs on top of IP
- Standard at: http://tools.ietf.org/html/rfc793
- Two forms of data transmit services:
  - Data stream push: sent data is accumulated in one stream, and sent when enough data has been accumulated.
  - Urgent data signaling: Receiving user (process) is informed that urgent data is available and may take action.

Leiden University. The university to discover

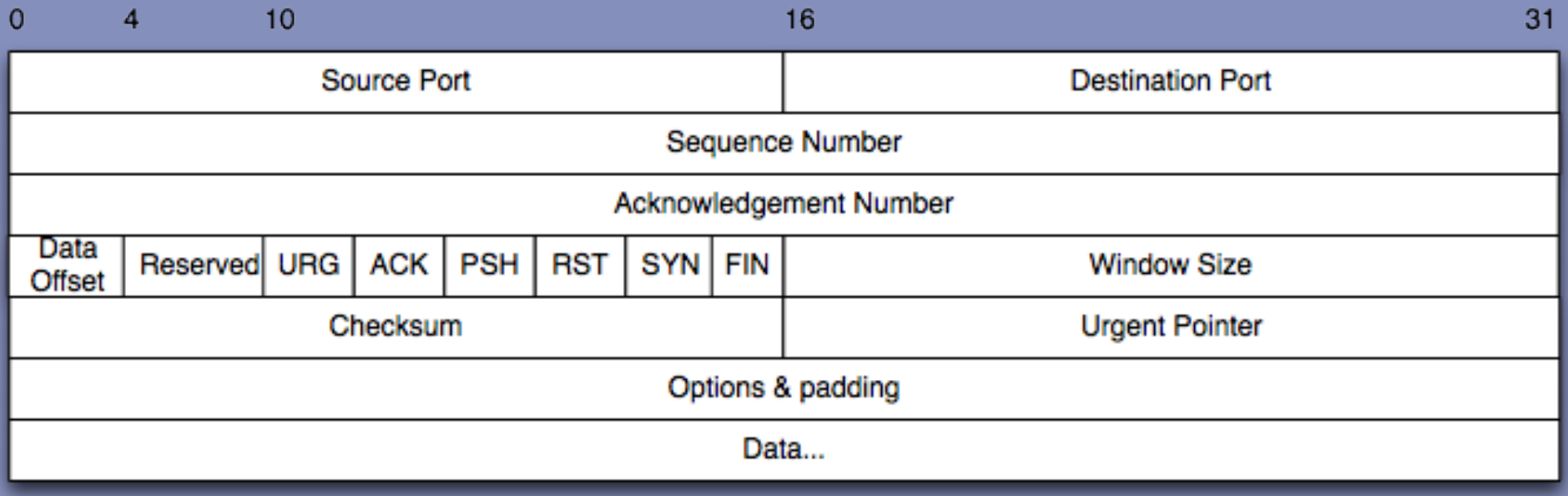
TCP

- TCP provides more functionality than IP
- 17 service primitives (e.g. terminate, status response, send, etc.)
- 21 service parameters (e.g. receive window, timeout, etc.)

Leiden University. The university to discover

TCP Header

TCP packet: header minimum 20 * 8 bits long



TCP Header

- Source Port: port number of sender
- Dest port: port number of destination (e.g. port 80 = http)
- Sequence number: flow control
- Ack number: the next sequence number the receiver expects
- Data offset: number of 32 bit words in the header

Leiden University. The university to discover

TCP header

- Reserved: for future use, some bits used for additional flags not in the initial standard (NS(*Nonce Sum*), CWR(*Congestion Window Reduced*) and ECE(*ECN-Echo*)) to support ECN (Explicit Congestion Notification).
- Flags (1 bit each)
  - URG: urgent
  - ACK: Acknowledgement
  - PSH: push
  - RST: reset connection
  - SYN: synchronise sequence number
  - FIN: no more data from sender

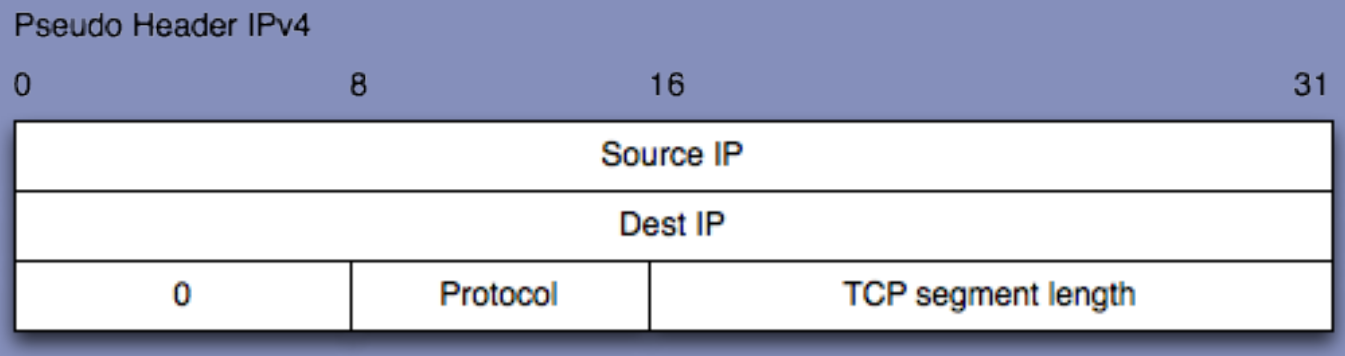
TCP header

- Window: window size (flow control) in octets. Max window = 2^{16} octets ~ 0.5 Mbit
- Checksum: one's complement sum mod 2^{16}-1 of all 16 bit words in the segment and the pseudo header.
- Urgent pointer: points at the last octet in the urgent data (if URG is set)
- Optional: e.g. max acceptable segment size

TCP pseudo header

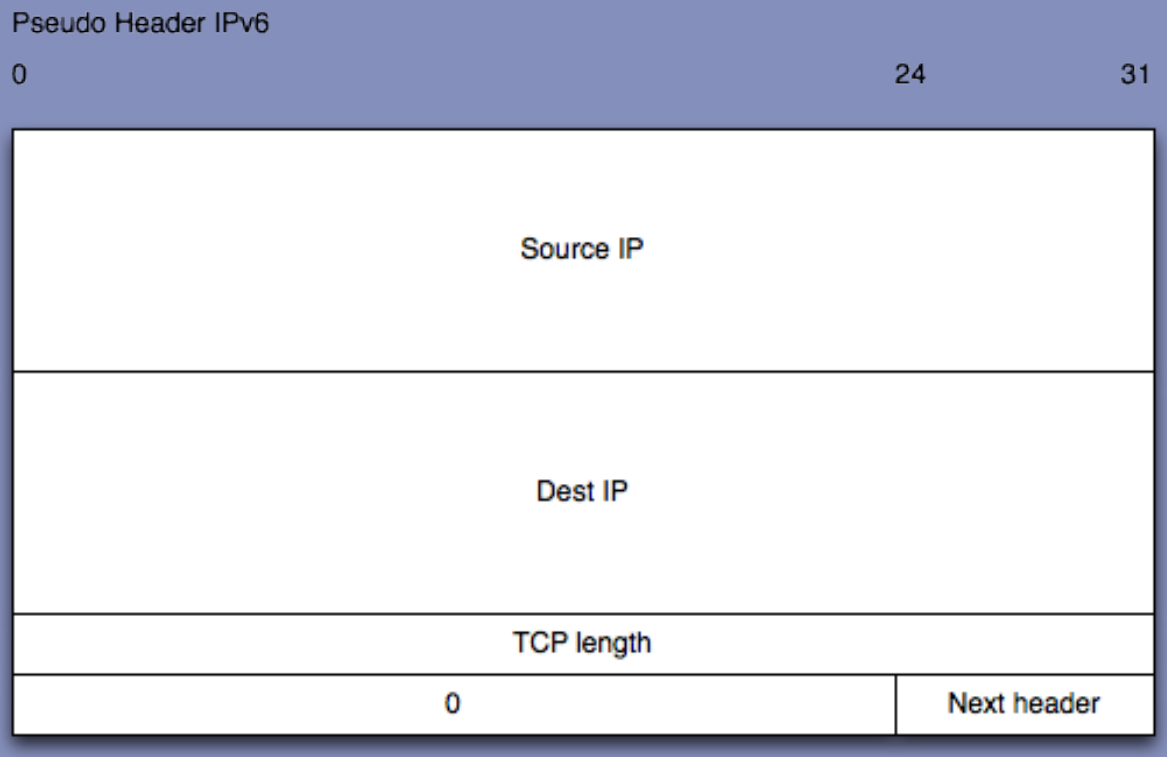
- Crosses OSI layer boundaries.
- For IPv6, IP header has no checksum, so the TCP header checksum takes care of the robustness.
- Checksum is computed over the TCP header and the TCP payload, and IP source and destination address fields, and IP protocol field. Additionally TCP header length and TCP packet length in octets are taken in account.

TCP pseudo header



Protocol = 6 for TCP.

TCP pseudo header



Next header field from IPv6 header