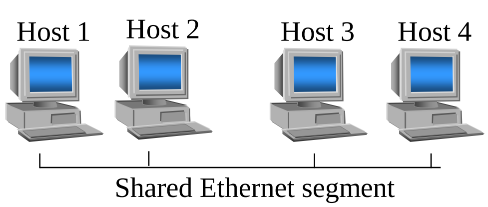
Introduction to routing in the Internet

Ethernet, switching vs. routing
Internet architecture
Addressing, routing principles
Protocols: IPv4, ICMP, ARP

(Chapters 2–3 in Huitema)

Ethernet

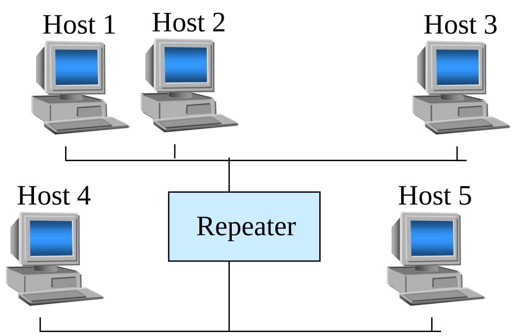
- Most widespread LAN technology
- Shared medium: Carrier sense multiple access with collision detection (CSMA/CD)

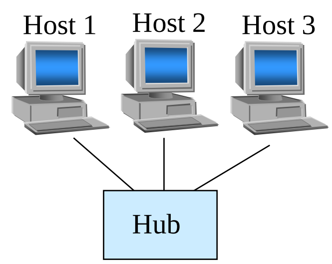


- Everyone receives everyone's traffic!
- Limited length!
- Supports broadcast

Interconnecting Ethernet segments with repeaters

- Repeaters repeats traffic of one segment on the other segment





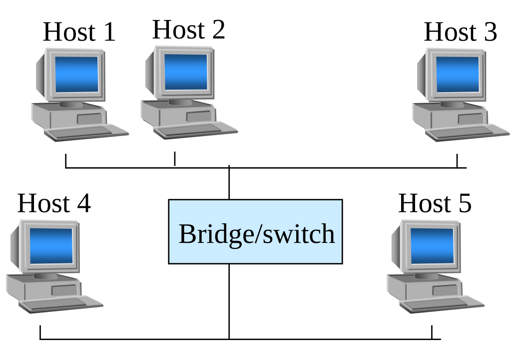
- Hubs are multi-port repeaters
- Still everyone receives everyone's traffic!
- Limited number of repeaters!

S-38.2121 / Fall-07 / RKa, NB

Internet-3

Bridges learn where devices are and forward packets only to the segment where the destination is

- Prior to discovery bridges/switches pass all traffic between segments

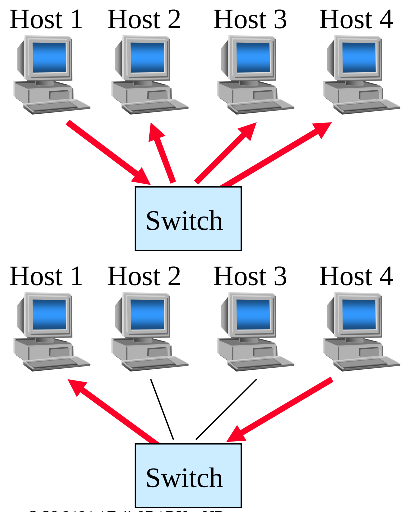


- Broadcast traffic sent to all devices
- Allows different speeds in different segments
- Network size not physically limited

S-38.2121 / Fall-07 / RKa, NB

Internet-4

Example

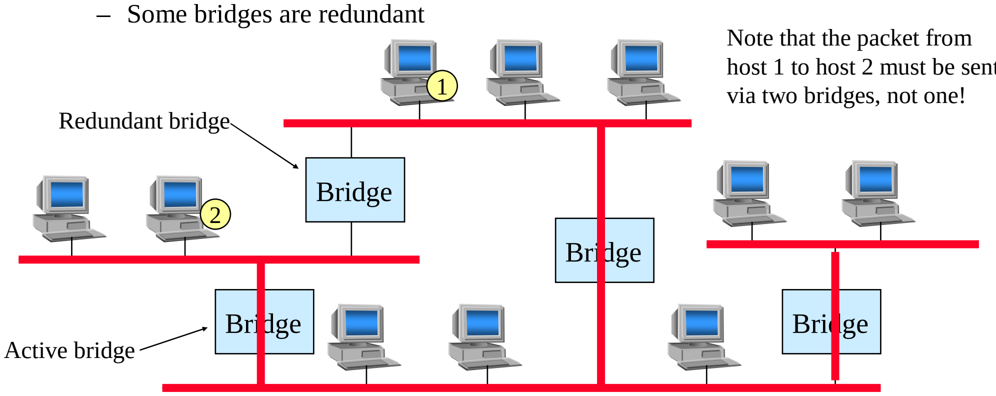


- Host 1 has a packet to send to Host 4
- The switch receives the packet and learns that Host 1 is on interface 1
- The switch does now know Host 4
- The packet is sent on all interfaces

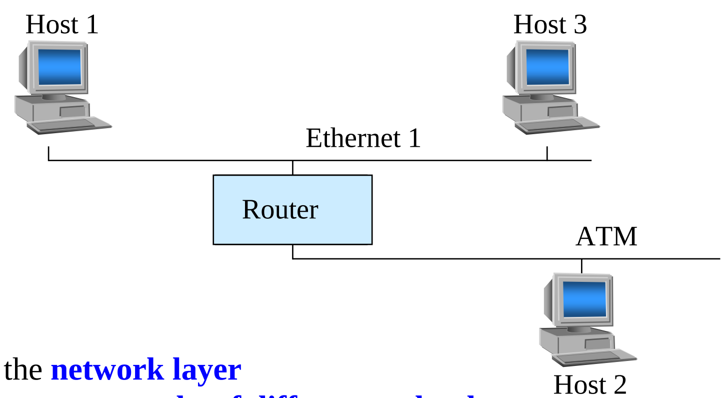
- Host 4 has a packet to send to Host 1
- The switch receives the packet and learns that Host 4 is on interface 4
- The switch now knows that Host 1 is on interface 1
- The packet is sent on interface 1

A spanning tree connects all segments without loops

- Spanning tree protocol
  - Only one possible path between two devices -> loops are impossible
  - Some bridges are redundant



A router allows the shortest paths to be used



A router
- operates on the **network layer**
- can **interconnect networks of different technology** (ATM, Ethernet, Frame relay, FDDI, …)

Routed IP vs. switched Ethernet

Routed IP	Switched Ethernet
• Interworking between network technologies	• Works with Ethernets
• Complex, expensive	• Simple, cheap, less layers

• One route per destination
  • shorter routes
  • load distribution
	• All destinations use the same spanning tree
	  • longer routes
	  • congestion on the tree

• Address allocation	• Fixed addresses
• Hierarchical address space	• Flat address space

Issues to be addressing for use of Ethernet in WANs

- Avoid broadcasting of packets to an unknown destination
- Forwarding table size (flat topology)
- Use the shortest path instead of a spanning tree
- Use all available links for load balancing
- Limit use of broadcasting
- Faster convergence (rapid spanning tree)
- Peering, policies, address hierarchy

Do we re-invent IP when these are implemented in Ethernet?

Internet Architecture Principles

End-to-end principle

by Dave Clark

- Hop-by-hop control
  - In X.25
  - Intelligence in the network
  - Error and flow control on each hop

vs.

- End-to-end control
  - In IP
  - Intelligence in end station
  - Error and flow control in end station
- The network can not be trusted
  - The user must in any case check for errors
    - →Network control is redundant
- Error checking and flow control by TCP in the end stations
- No state information in the network
  - The network is not aware of any connections
  - Packets routed independently
  - If a link fails, another route is used
- Same principle as in distributed systems

Internet Architecture Principles

IP over everything

by Vinston Cerf

- Internet connects different types of networks
  - Each with different framing, addressing, …

Interconnection based on *translation*

- Mapping through a gateway
- Never perfect

Interconnection based on *overlay*

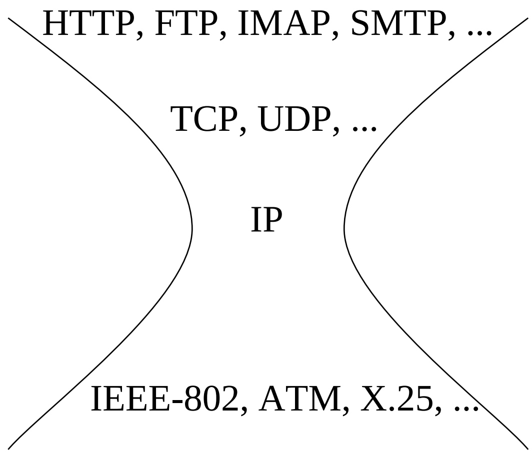
- Approach used by IP
- Single protocol over all underlying networks
- Simple to adapt to new technologies
  - Define framing or encapsulation
  - Define address resolution: IP-address → network address
- Unique IP-address

Translation still needed in many cases

E.g. signaling interworking, IPv4 to IPv6 mapping

Internet Architecture Principles

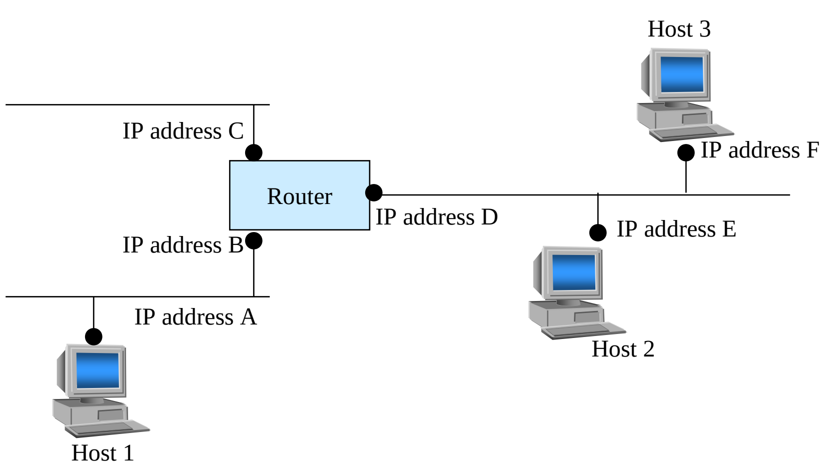
IP over everything



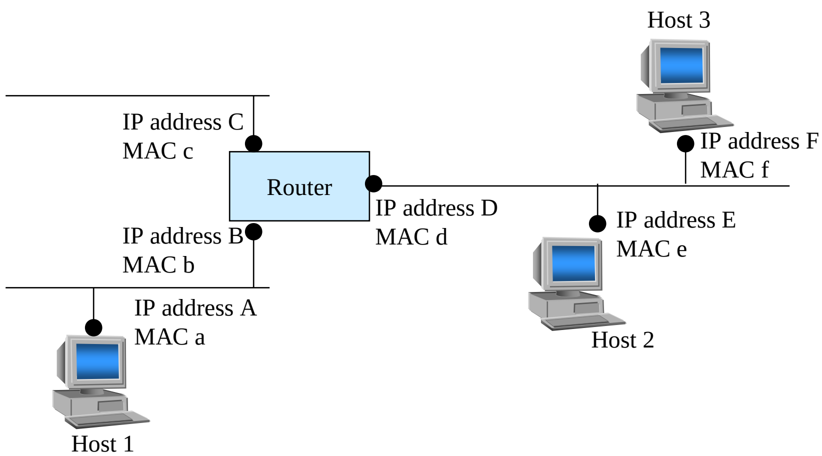
Internet Architecture Principles
Connectivity is its own reward

- The value of a network increases in proportion to the square of the number of nodes on the network (Robert Metcalf's law)
- Be liberal with what you receive, conservative with what you send *by Jon Postel*
  - try to make your best to understand what you receive
  - maximum adherance to standard when sending
- Snowballing effect keeps all interested in connectivity thus keeps adhering to standards

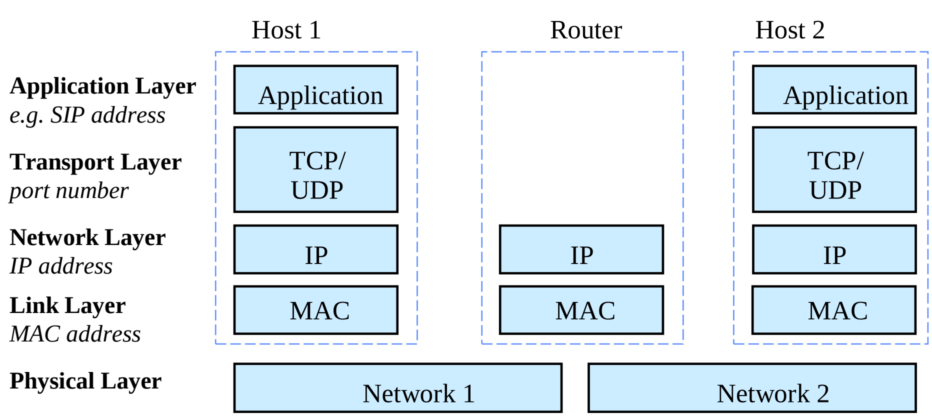
The IP address defines the interface
(not the host)



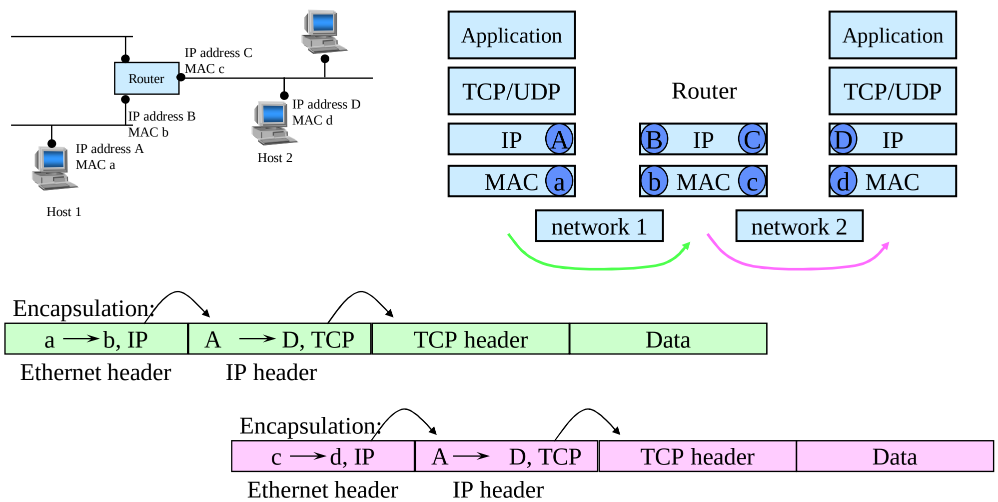
Every interface also has a media specific MAC address



Internet layer model – hosts and routers



Layers and message forwarding



IPv4 address formats

- Originally a two-level (network, host) hierarchy

32 bits

1981

MSB	Network	Host	Class
0	**7 bits**	**24 bits**	**A**
10	**14 bits**	**16 bits**	**B**
110	**21 bits**	**8 bits**	**C**
1110	**28 bits - multicast address**		**D**
1111	**For experimental and future use**		**E**

IPv4 address formats

1984

• A new level for easier network administration

	Network	Subnet	Host	
Example:				
Address: 10.38.154.117	*00001010*	*001001*	*10 10011010 01110101*	
Mask: 255.255.192.0	*11111111*	*111111*	*00 00000000 00000000*	
Network: first bit "0"	*00001010*			*= 10*
Subnet: address* AND mask		*001001*		*= 9 (36)*
Host: address AND NOT mask			*10 10011010 01110101*	*= 2.154.117*

address = address with network part zeroed*

NB: Also written as 10.38.154.117 / 14

IPv4 address formats

• Examples:

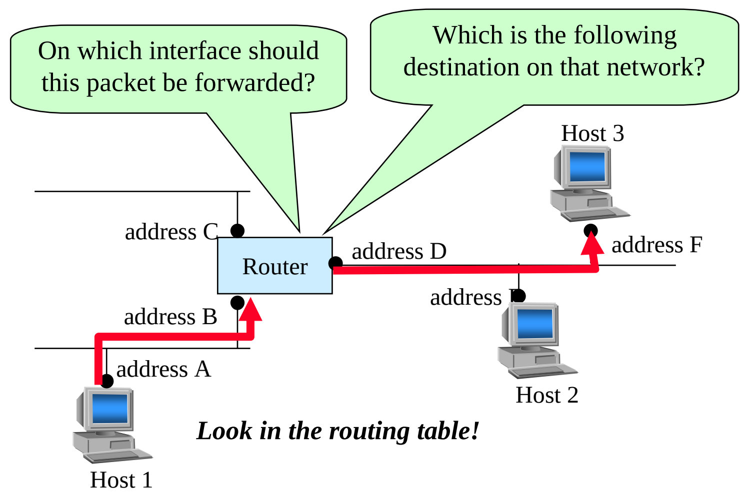
Mask	IP address	Network	Subnet	Host
0xFFFF0000	10.27.32.100	A: 10	27	32.100
0xFFFFFE00	136.27.33.100	B: 136.27	16 (32)	1.100
	136.27.34.141	136.27	17 (34)	0.141
0xFFFFFFC0	193.27.32.197	C: 193.27.32	3 (192)	5

High order bits:
0 ..... 0 - 127. → A-class
10.... 128. - 191. → B-class
110...192. - 223. → C-class

Without right zeroes (and with right zeroes)

Later updated by CIDR (discussed later)

Two functions of a router:
1. Packet forwarding



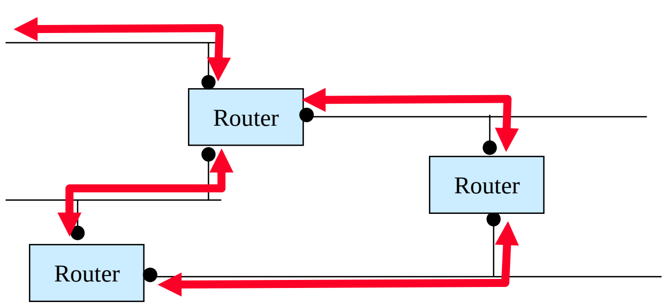
Look in the routing table!

S-38.2121 / Fall-07 / RKa, NB

Internet-21

Two functions of a router:
2. Construction and maintenance of the routing table

- Routers exchange routing information with routing protocols (e.g. RIP, OSPF, BGP)

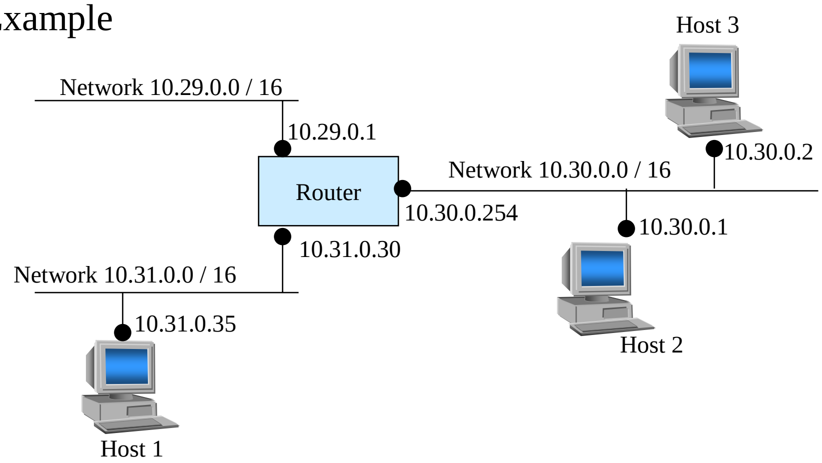


S-38.2121 / Fall-07 / RKa, NB

Internet-22

Routers maintain routes to networks (not to hosts)

- Example

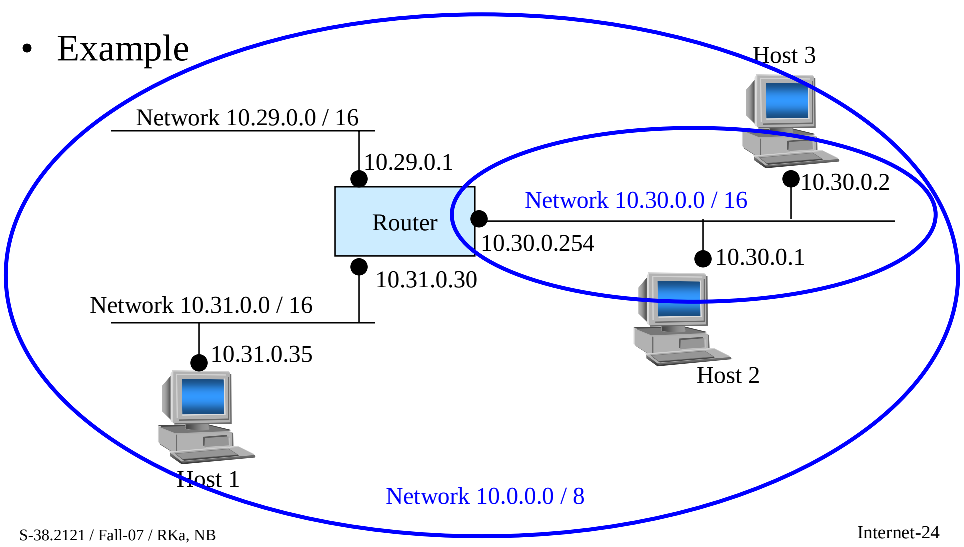


S-38.2121 / Fall-07 / RKa, NB

Internet-23

Aggregation describes several addresses in a single entry to reduce size of routing tables

- Example



S-38.2121 / Fall-07 / RKa, NB

Internet-24

Special purpose addresses

- An unknown network is replaced by 0
  - Only used as source address (e.g. a booting host)
  - 0.0.0.0 = ”this host in this network”
  - 0.X.Y.Z = ”host X.Y.Z in this network”
- Limited broadcast address 255.255.255.255
  - To all hosts in the local network
- Directed broadcast addresses A.255.255.255, B.B.255.255, C.C.C.255
  - To all hosts in a specified network
- Loopback-address 127.X.X.X (usually 127.0.0.1)
  - Internal in one host
- Multicast-addresses (e.g. 224.0.0.2 = all routers on this subnet)

IPv4 packet header

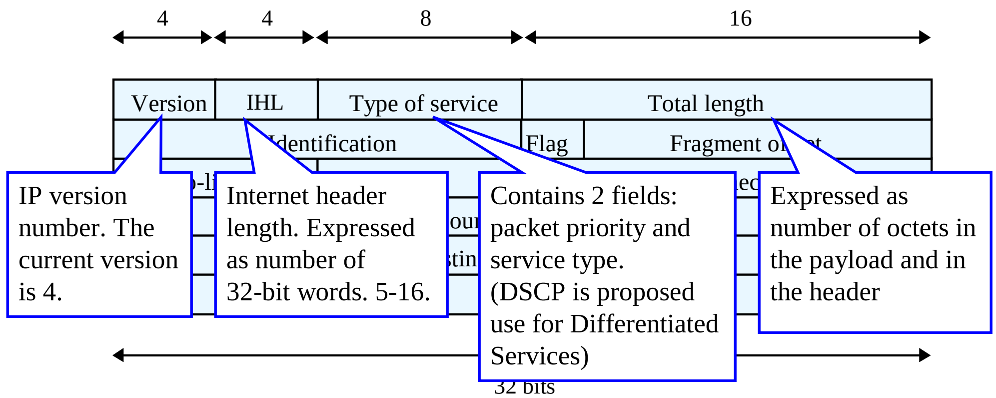
RFC-791

4	4	8	16	
Version	IHL	Type of service	Total length	
Identification			Flag	Fragment offset
Time-to-live (TTL)		Protocol	Header checksum	
Source IP Address				
Destination IP Address				
Optional				Padding

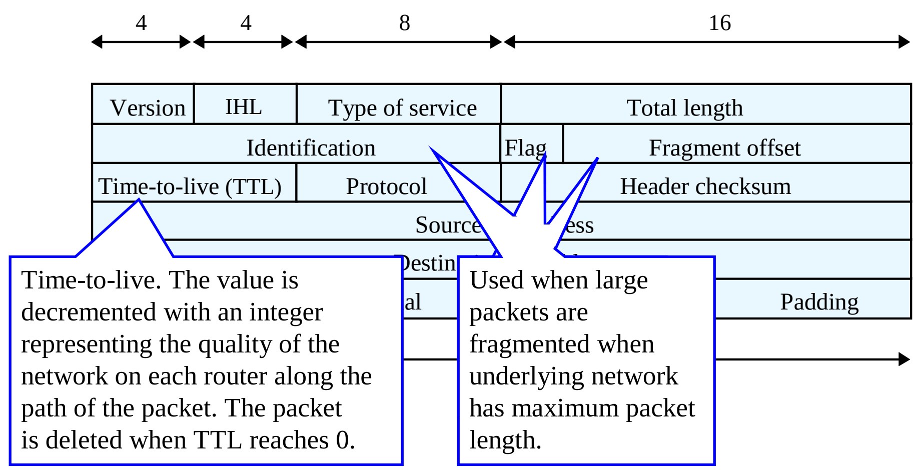
32 bits

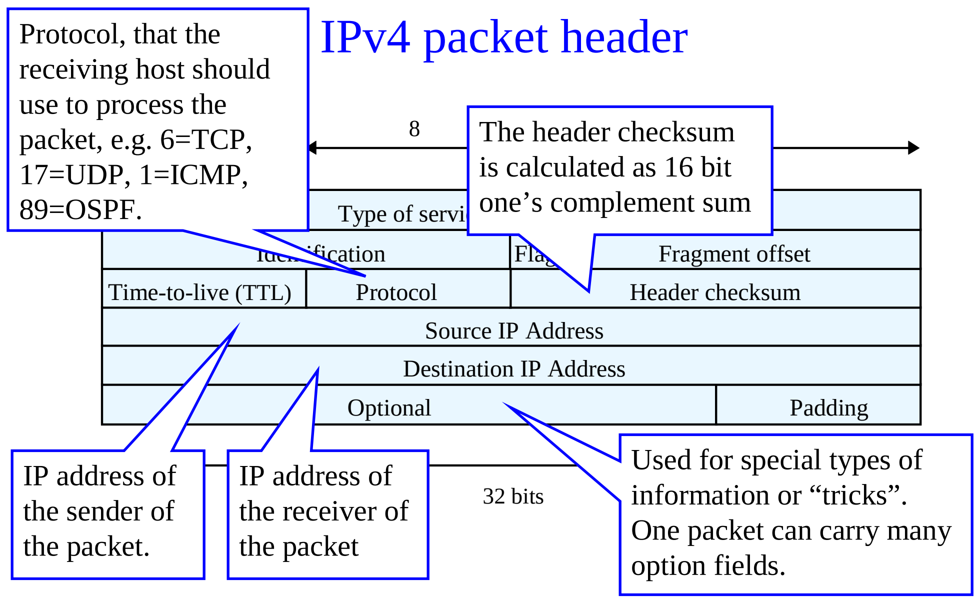
We assume that the sender knows its own IP address.
If not: self configuration protocols such as RARP, BOOTP, DHCP (dynamic host configuration protocol) are used

IPv4 packet header



IPv4 packet header





The most important fields in routing are the destination address and the time-to-live

Version	IHL	Type of service	Total length	
Identification			Flag	Fragment offset
Time-to-live (TTL)		Protocol	Header checksum	
Source IP Address				
Destination IP Address				
Options				Padding

- Every router decrements the TTL → must calculate new checksum
- Options (e.g. source routing, record route, timestamp)
  - rarely/never used in practice.

Type of service

Precedence	D	T	R	C	

- Route selection criteria
  - D – minimization of delay
  - T – maximization of transmission capacity
  - R – maximization of reliability
  - C – minimization of cost
  - Only one can be selected.
- Precedence
  - Packet with the highest precedence is first taken from the queue to be routed.
- In practice, these are not used
- DiffServ uses the field in another way

Source routing

Type	Length	Pointer	Address 1	Address 2	...	Address N

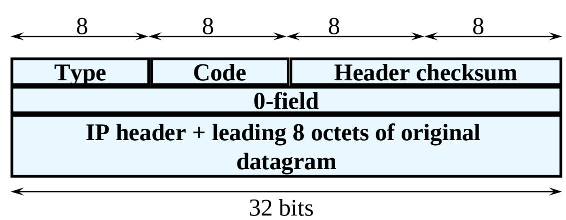
- Implemented with the ”source routing” option
  - Loose source routing (type 131, *10000011*)
    - The packet is sent to the next address in the list using normal routing.
  - Strict source routing (type 137, *10001001*)
    - The packet is sent to the next address in the list. If there is no direct link to the address, the packet is destroyed.
- Slow → Rarely used
  - Can be replaced by encapsulation:

A→C, IP-IP	A→B, TCP	TCP	Data

ICMP – Internet Control Message Protocol

- Gives feedback about the network operation.
- ICMP packet is sent backwards if e.g.
  - The destination is unreachable
  - The router destroys a packet
  - TTL expires
- All hosts and routers must support ICMP.
- ICMP messages are transported in IP packets
- If a ICMP message is dropped, a new one is not generated
  - to avoid the "snowballing effect".

ICMP messages

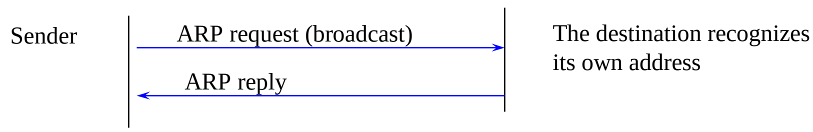


Type
0 - Echo reply (used for "ping")
3 - Destination unreachable
4 - ~~source quench~~ (="slow down")
(dropped from recommendations)
5 - Redirect
8 - Echo (used for "ping")
9 - Router advertisement
10 - Router solicitation
11 - Time exceeded
12 - Parameter problem
13 - Timestamp
14 - Timestamp reply
15 - Information request
16 - Information reply

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

Packet sending – how to determine the next hop

- The sender checks if the destination address is in the same sub-network by comparing the masked values of the source and destination address.
  - If same, the destination is in the same subnet (next hop = destination).
  - Otherwise, the packet must be sent to a router (next hop = router).
- It then obtains the media address (MAC-address) of the next hop using the ARP-protocol.



- The media address is stored in the cache.
  - Note: All hosts in the same subnet stores the address in their cache.

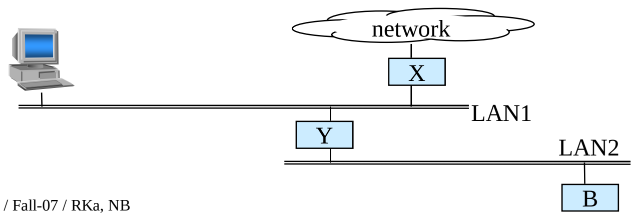
ARP – Address Resolution Protocol

RFC-826

- ARP maps IP to the underlying protocol
- IP-address → MAC-address
- Each network technology requires its own ARP adaptation.
  - Easy if the network supports broadcast or multicast.
    - E.g. Ethernet, Token Ring, FDDI
  - ATM requires a special ARP-server
  - Manually defined address for point-to-point links
    - E.g. X.25, ISDN, Frame-Relay
- Works on top of Ethernet (not on top of IP)

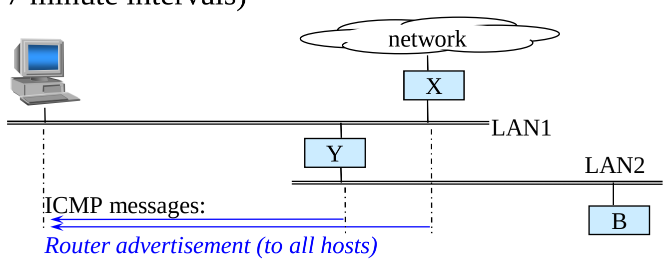
Router discovery

- How to know the address of the router?
  - Configure manually – ”default gateway”
  - Obtain with DHCP
    - Configured by administrator, still needs manual work
  - Listen to routing protocols
    - Uses resources of the host, too many routing protocols → not used today
  - Automatic router discovery with ICMP



ICMP router discovery (1)

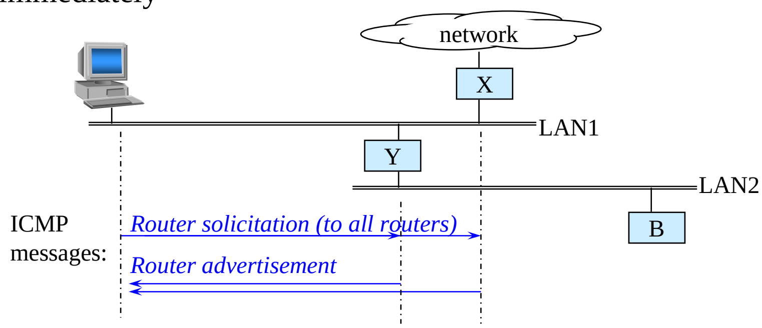
- The routers send *router advertisements* to all hosts periodically (e.g. in 7 minute intervals)



- The advertisement contains
  - a list of the router’s addresses.
  - the preference of the addresses, which are used to identify the normal, reserve, etc. router or router address (the preference of the default router is highest)
  - lifetime of the information (e.g. 30 min)

ICMP router discovery (2)

- The host would have to wait up to 7 minutes before it can send packets outside its sub-network.
- Using a *router solicitation*, the host gets the advertisement immediately



S-38.2121 / Fall-07 / RKa, NB

Internet-39

ICMP router discovery (3)

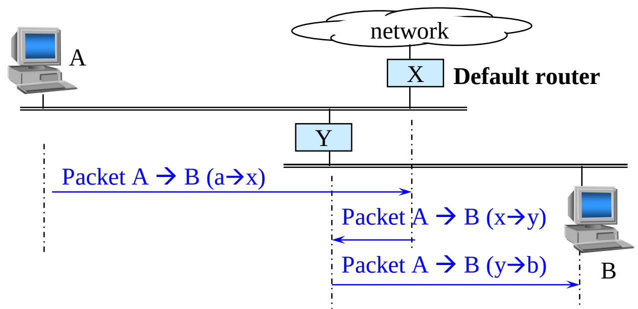
- The host chooses the router with the highest priority as its default router.
- All packets for destinations outside the sub-network are then sent to the default router.
- Any advertisement from a router outside the sub-network is discarded

S-38.2121 / Fall-07 / RKa, NB

Internet-40

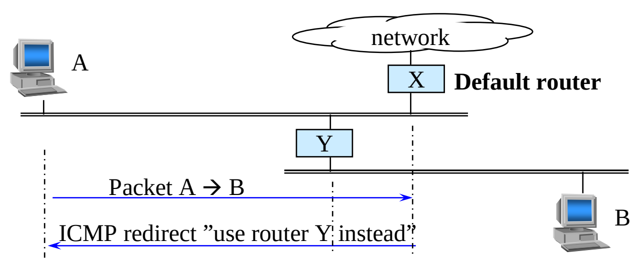
A network may have many routers, the closest to the destination must be found

- A packet sent through the default router reaches the destination, but may waste resources



A network may have many routers, the closest to the destination must be found

- The router can send a redirect to indicate a shorter route to the destination



Type	Code	Header checksum
IP address → router=Y		
IP header + 8 octets of the original datagram		

Type
5 – redirect

Code
0 – redirect for the network (no mask!)
1 – redirect for the host
2 – redir. for type of service and network
4 – redir. for type of service and host

Host must have feedback from the first router to avoid sending to a “black hole”

Feedback may be

- TCP acknowledgements
- Router advertisements
- ARP-replies
- ICMP echo reply (ping)

Between routers, routing protocols provide similar feedback and help in detecting failed router neighbors.

DNS – Domain Name Service

- Host name → IP address
- Why DNS?
  - Easier to remember names than addresses
  - Allows address changes without changing the name
  - Several addresses per host
  - Extensions: service location, ENUM
- DNS does not affect routing, routers only deal with IP addresses

Routing algorithms

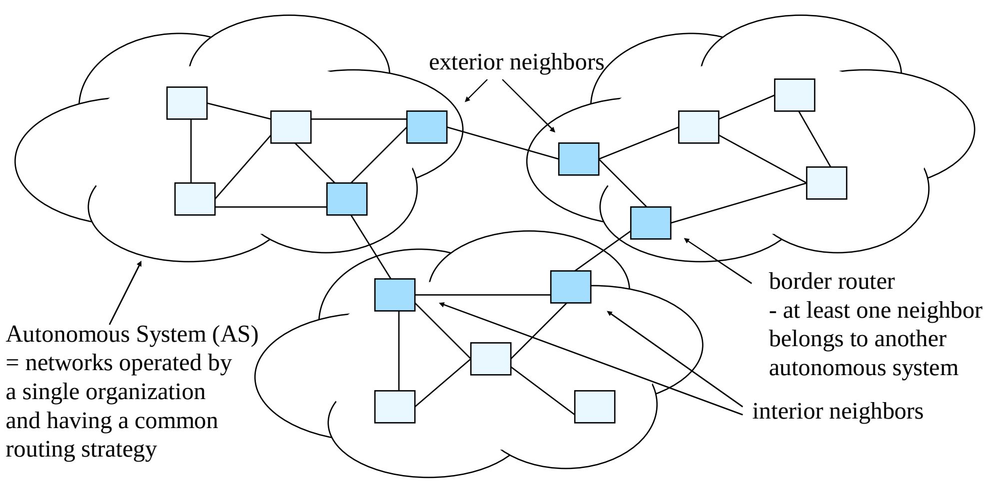
Internet routing is based on routing protocols, which collect information

- Routing is completely automatic
- No offline route planning
- Only dimensioning is made offline
- The routers communicate with a routing protocol
- The routing protocol implements a routing algorithm that finds the shortest (cheapest, fastest, etc.) route to every destination

Routing in the Internet is generally dynamic, but static routing is used in some cases

- Dynamic routing is based on routing protocols which create and maintain the routing tables automatically
  - examples of routing protocols are RIP, OSPF, BGP...
  - E.g. to connect an organization with multiple links to the Internet
- Static routing is based on manually configured routing tables.
  - Static routing is used when e.g. two peer providers do not trust each other
  - To connect an organization to a service provider with a single connection
  - Static routing is difficult to maintain

Routing is divided into interior and exterior



In this course we only deal with interior routing

Routing is divided into interior and exterior

- Interior routing protocols
  - **Routing Information Protocol (RIP), RIP-2**
  - **Open Shortest Path First (OSPF)**
  - Interior Gateway Routing Protocol (IGRP), EIGRP
  - Intermediate System-to-Intermediate System (IS-IS)
- Exterior routing protocols
  - External Gateway Protocol (EGP) *(historical)*
  - **Border Gateway Protocol version 4 (BGP-4)**

Routing algorithms
Proactive vs. reactive

- Proactive
  - The router creates and maintains routes to all destinations
    → The routes are available in advance
  - The routing algorithms in the Internet are proactive
- Reactive
  - Routes are created only when they are needed
  - Used in e.g. ad hoc networks (discussed later in this course)

Routing algorithms
Distance vector vs. link state

- Distance vector
  - Distance vectors are sent, until the state of the network is stable
  - The routers cooperate to generate the routes
  - Example: RIP
- Link state
  - Topology descriptions are sent periodically and nodes generate a map over the network
  - Every router generates the routes independently of the other routers
  - Example: OSPF

Routing algorithms
Distance vector vs. link state

Distance vector	Link state
+ Simple and lightweight	- Complex and heavy
- Slow convergence	+ Fast convergence
- Only one route per destination	+ Several routes per destination
- Only one metric	+ Supports different metrics