Introduction to routing in the Internet

Internet architecture

IPv4, ICMP, ARP

Addressing, routing principles

(Chapters 2–3 in Huitema)

Internet Architecture Principles
End-to-end principle

by Dave Clark

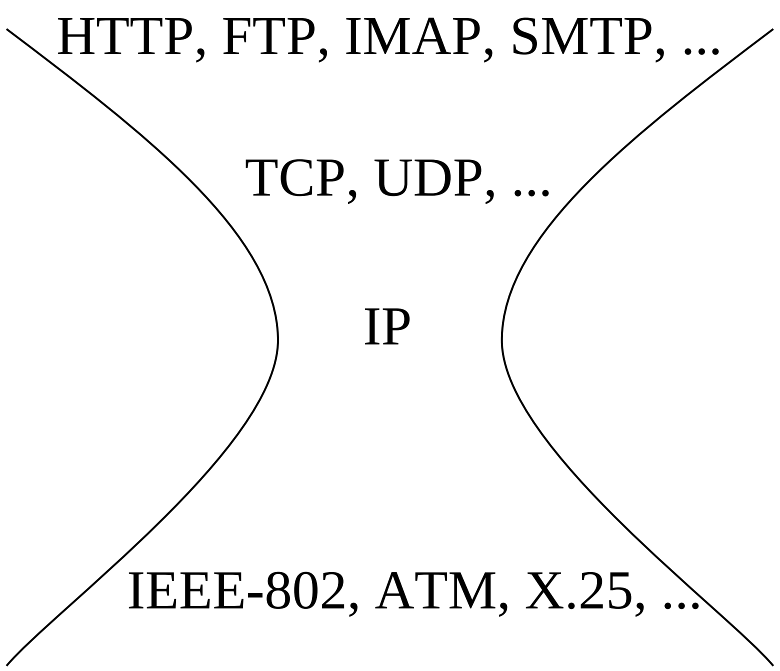
- All control in end stations
  - e.g. error and flow control
- The network can not be trusted
- User must in any case check for errors
  → network control redundant
- Error checking and flow control by TCP
- No state information/connection in the network
  - packets routed independently
  - if a link fails, another route is used
- Same principle as in distributed systems

Internet Architecture Principles
IP over everything

by Vinston Cerf

- Alternative: Interconnection based on *translation*
  - Never perfect
- IP: Interconnection based on *overlay* over all kinds of 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

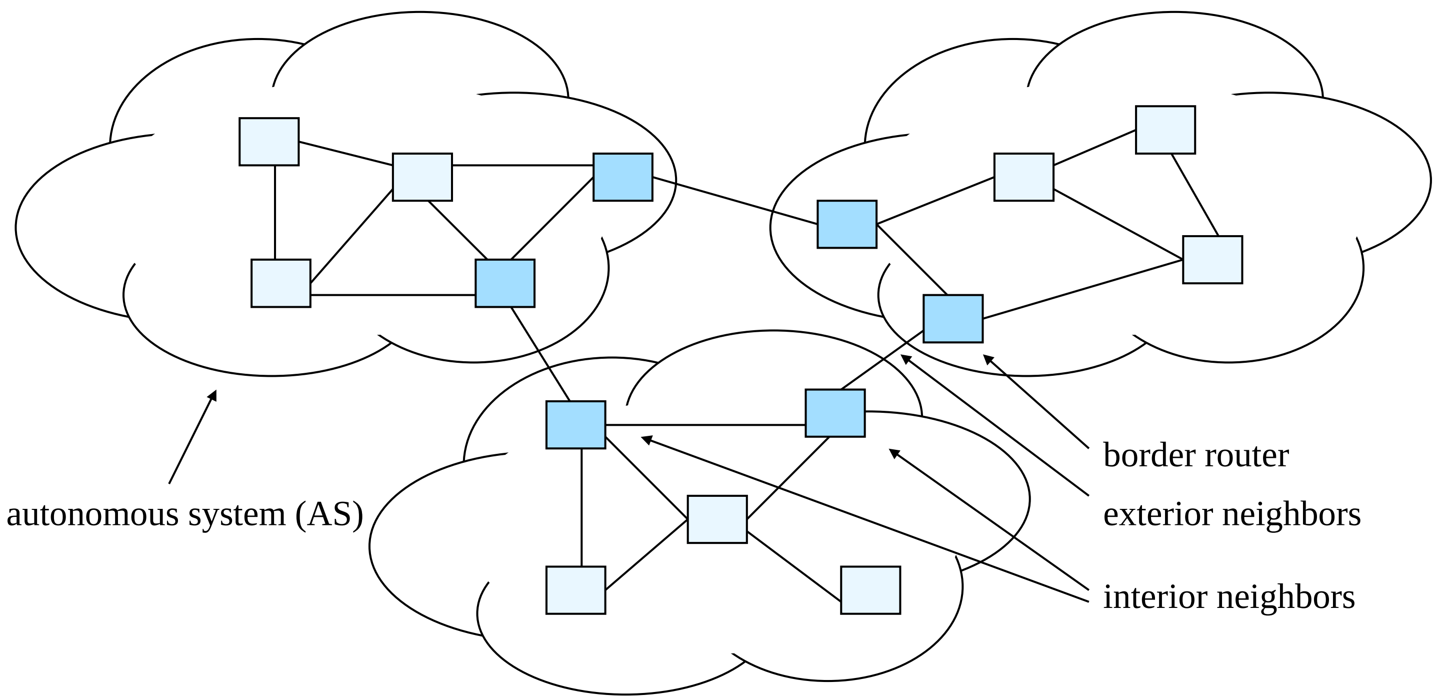


Internet Architecture Principles

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
  - 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

Routing is divided into interior and exterior



In this couse we only deal with interior routing

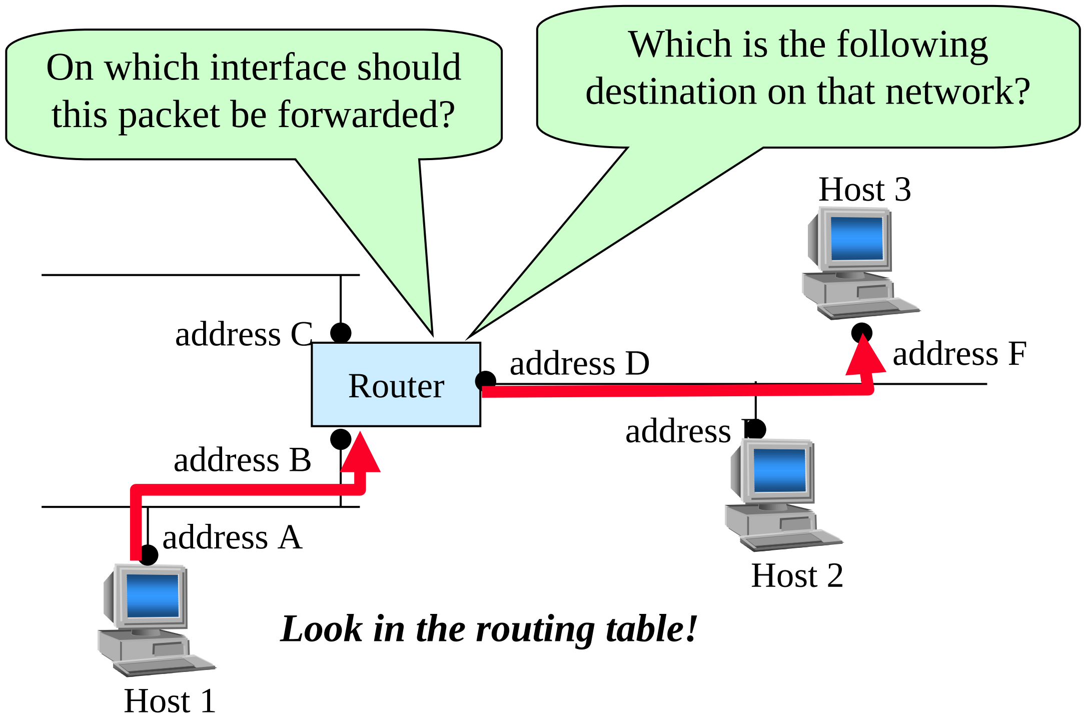
Routing is divided into interior and exterior

- Autonomous system, AS
  - Networks operated by a single organization and having a common routing strategy
- Border router
  - At least one neighbor belongs to another autonomous system

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)
  - **Border Gateway Protocol version 4 (BGP-4)**

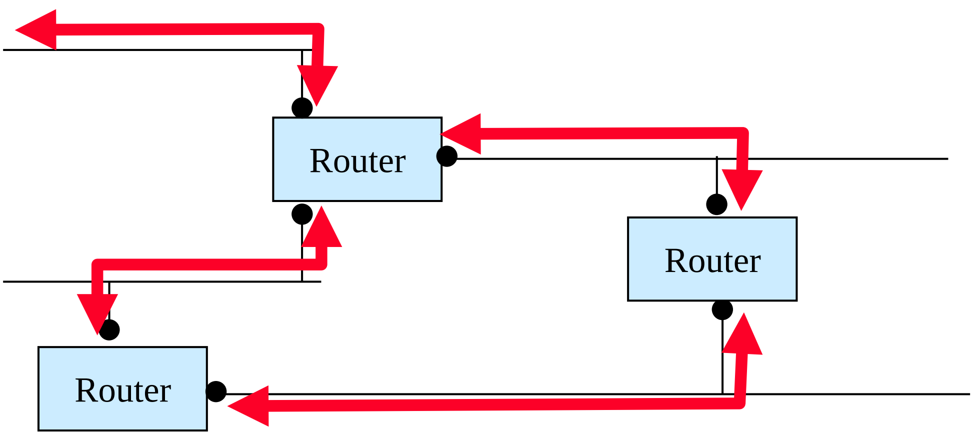
Two functions of a router:
1. Packet forwarding



Look in the routing table!

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)



Internet routing is based on routing protocols, which collect information

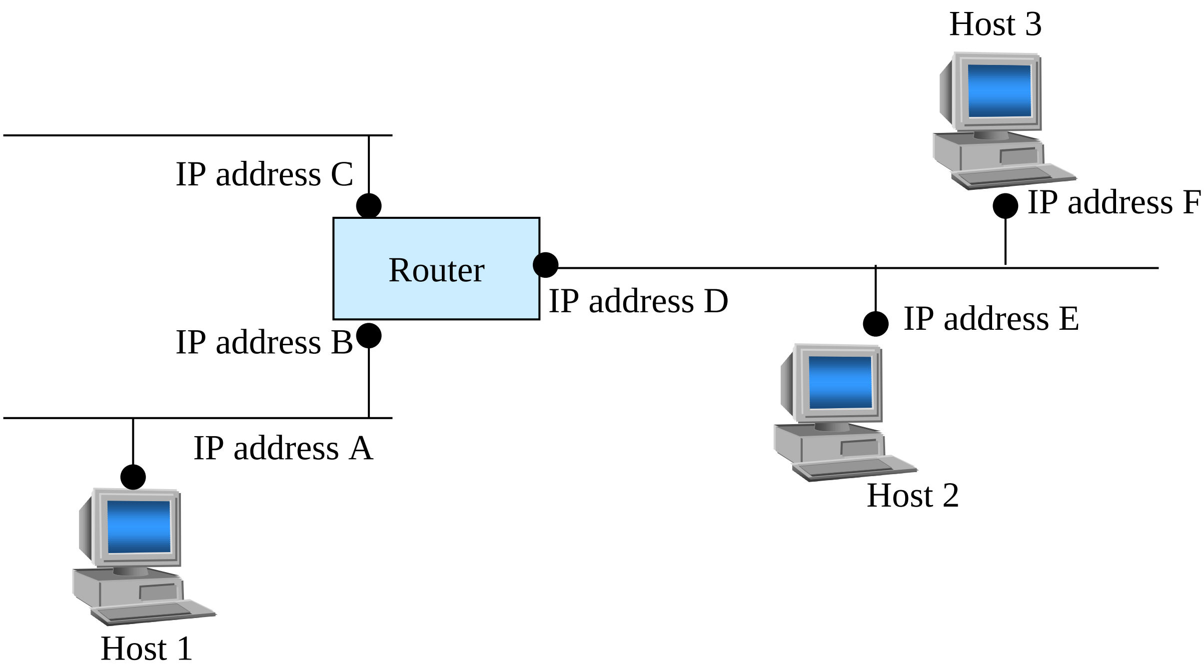
- Routing is completely automatic
- No offline route planning
- Only dimensioning is made offline
- The routers communicate with a routing protocol
- The routing algorithm finds the shortest (cheapest) route to every destination

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

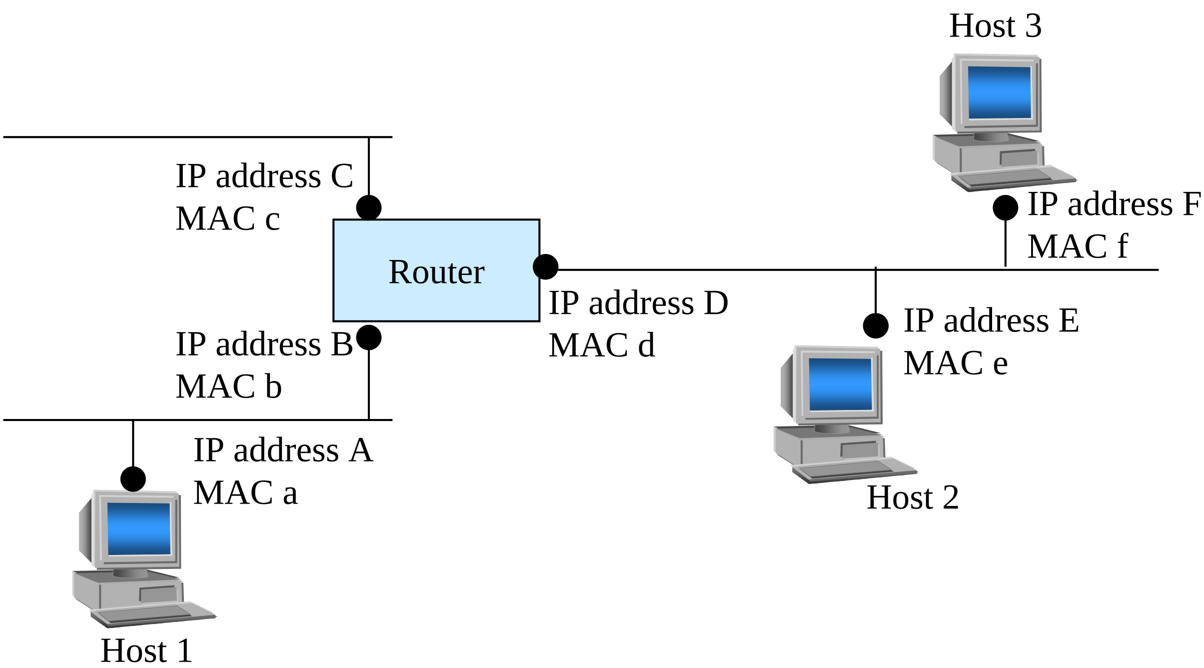
- Dynamic routing is based on routing protocols which create and maintain the routing tables automatically
  - examples of routing protocols are RIP, OSPF, BGP...
  - E.g. to connect an organization with multiple links to the Internet
- Static routing is based on manually configured routing tables.
  - 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

The IP address defines the interface

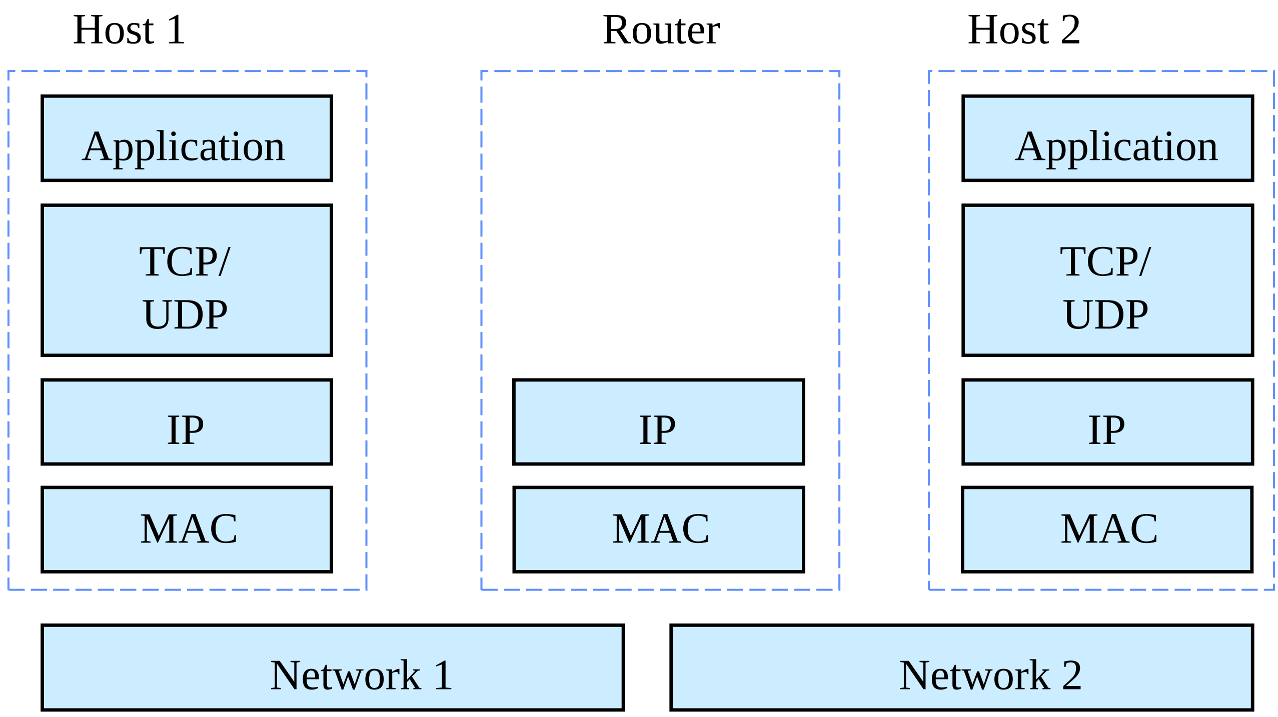
(not the host)



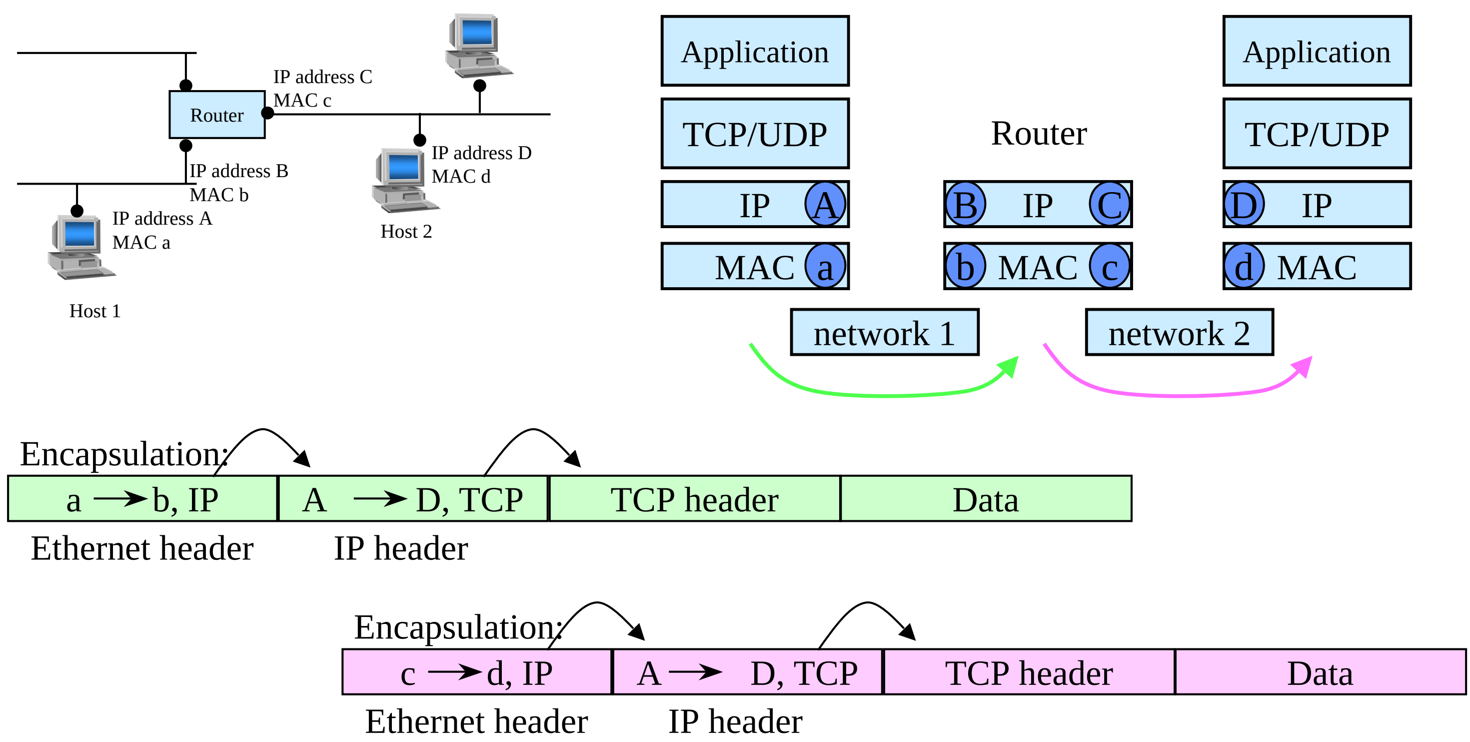
Every interface also has a media specific MAC address



Internet layer model – hosts and routers



Layers and message forwarding



IPv4 address formats

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

1981

32 bits

8 | 8 | 8 | 8

MSB(t)	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

- 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)

IPv4 address formats

Example:

		Network	Subnet	Host	
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*

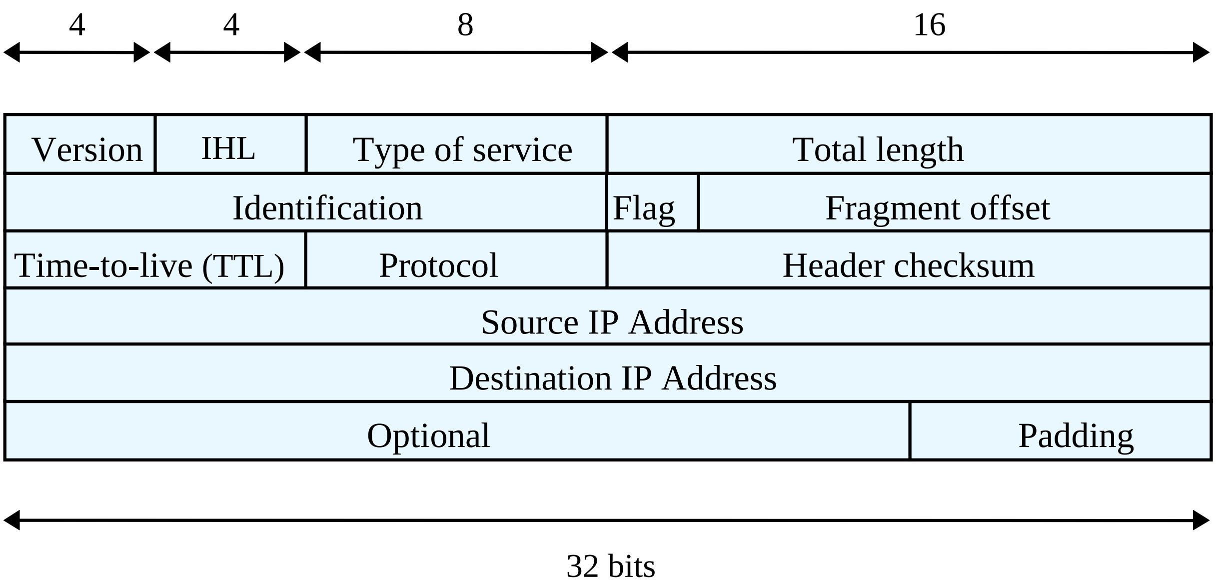
Also written as 10.38.154.117/14

Special 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 host 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)

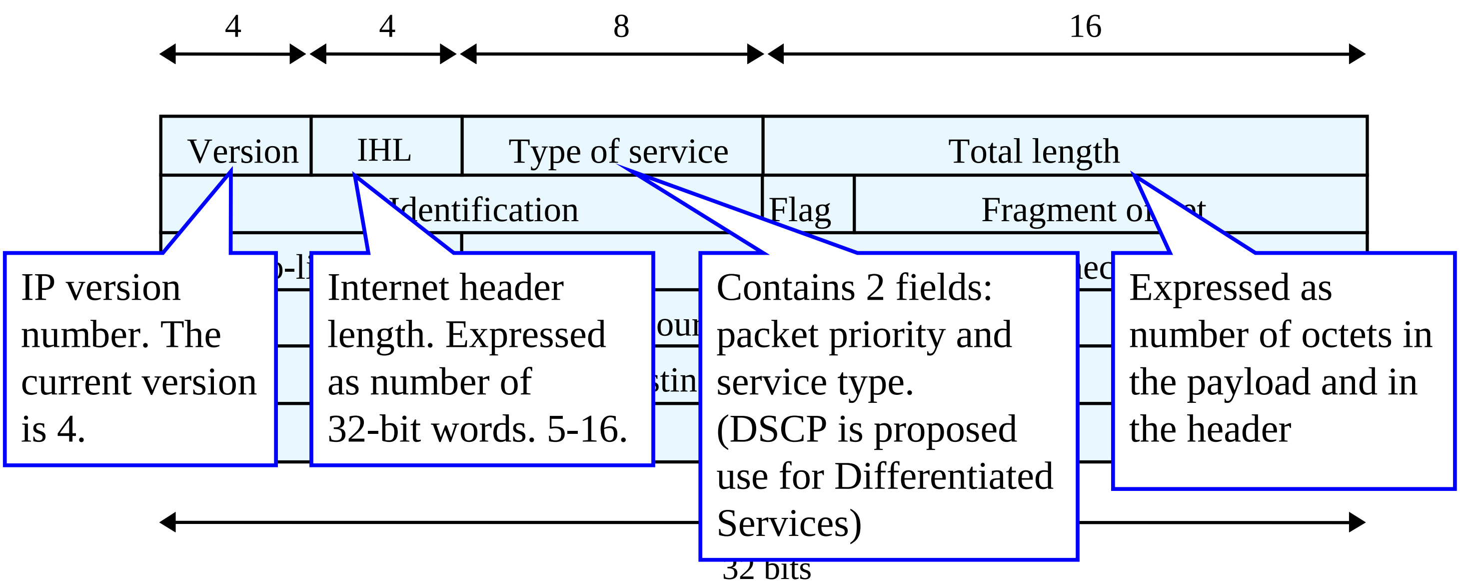
IPv4 packet header

RFC-791

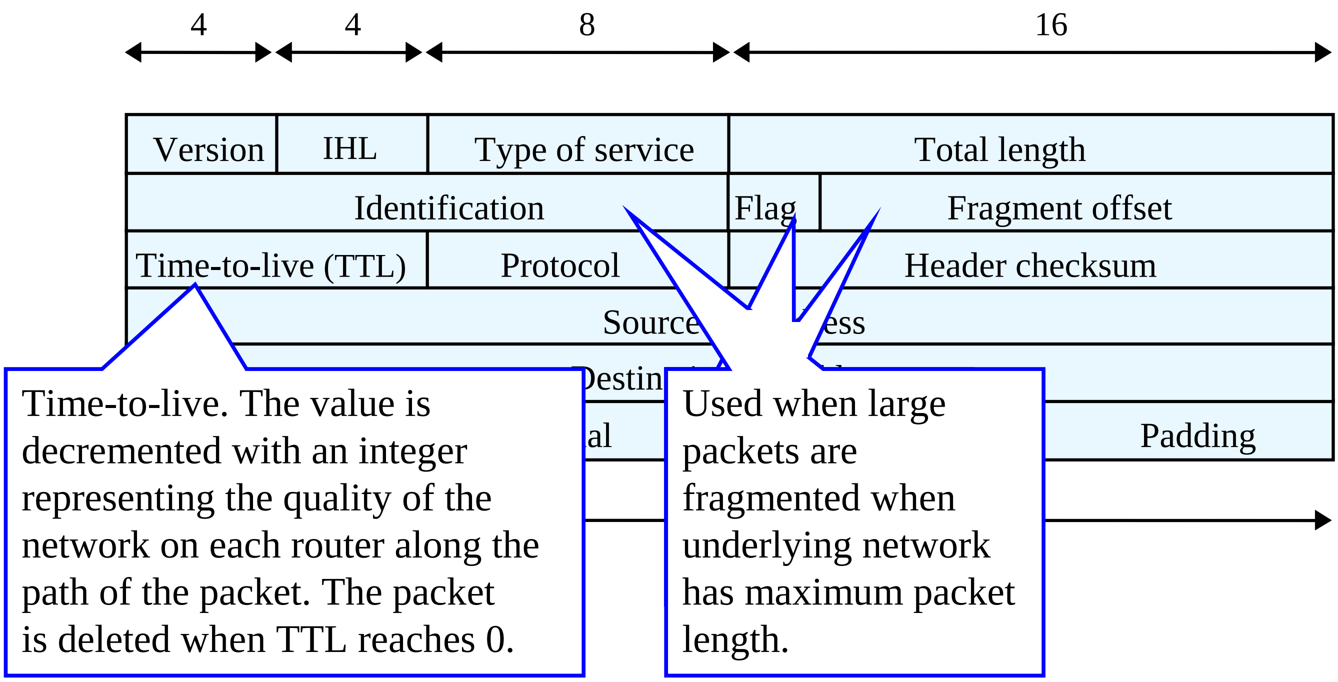


We assume that the sender knows its own IP address.
If not: self configuration protocols such as RARP, BOOTP, DHCP (dynamic host configuration protocol) are used

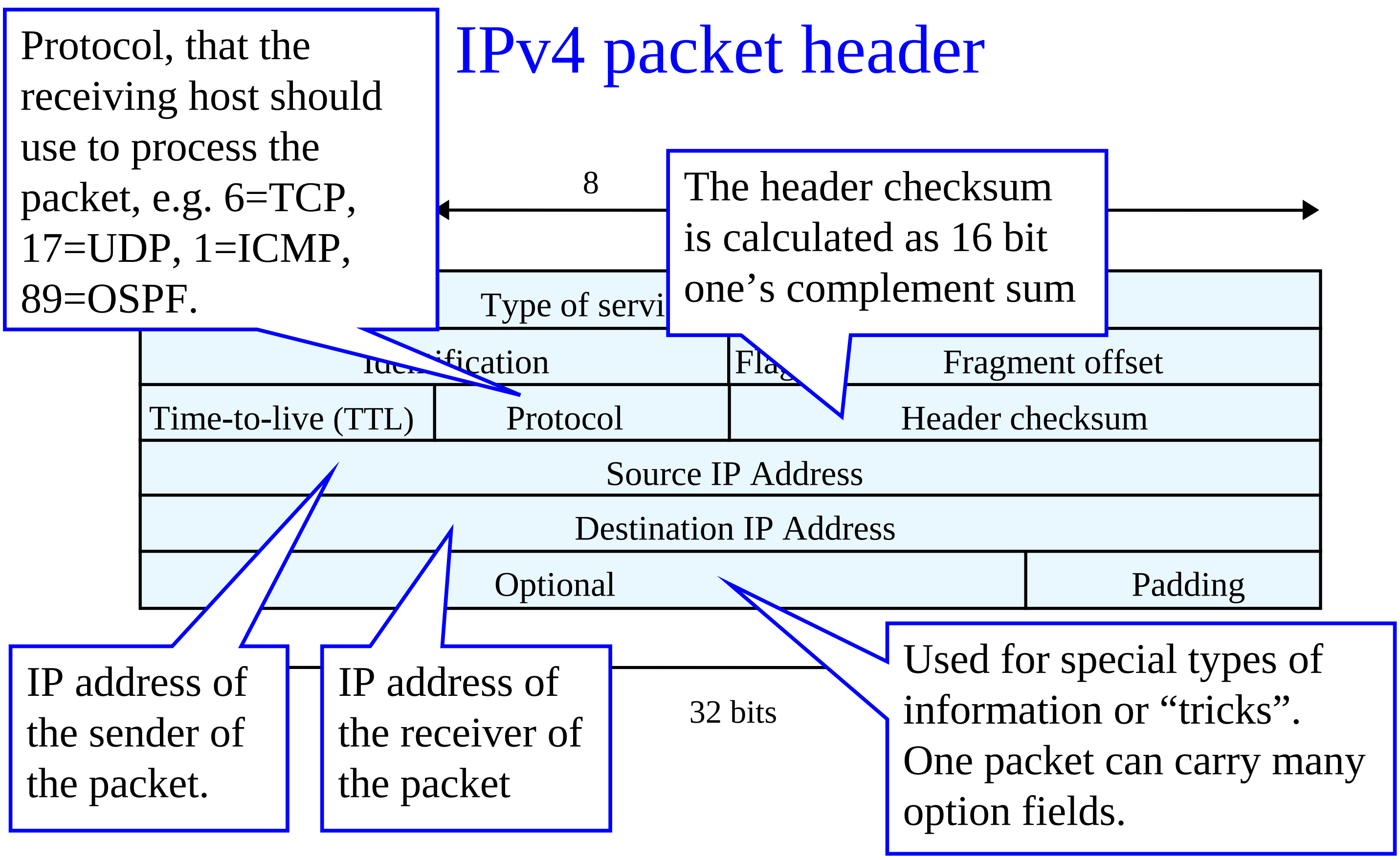
IPv4 packet header



IPv4 packet header



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
  – The largest precedence packet is first taken from the queue to be routed.
- In practise, these are not used
- DiffServ uses the field in another way

Source routing

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

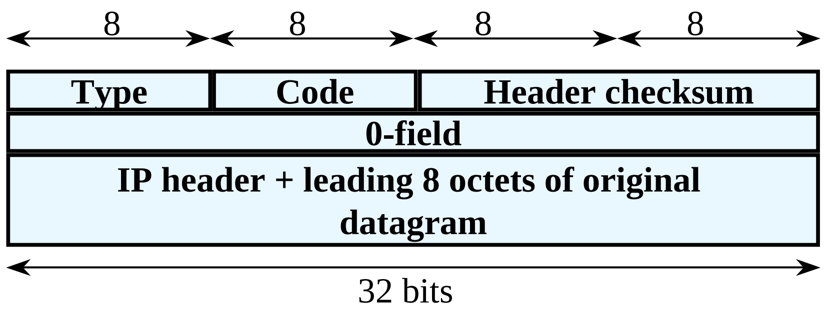
- Implemented with the ”source routing” option
  - Loose source routing (type 131, *10000011*)
    - The packet is sent to the next address in the list using normal routing.
  - Strict source routing (type 137, *10001001*)
    - The packet is sent to the next address in the list. If there is no direct 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 receiver is unreachable
  - The router deletes a packet
  - TTL = 0.
- All hosts and routers must support ICMP.
- ICMP messages are transported in IP packets
- If ICMP message is dropped, a new one is not generated
  - to avoid the “snowballing effect”.

ICMP messages



Type

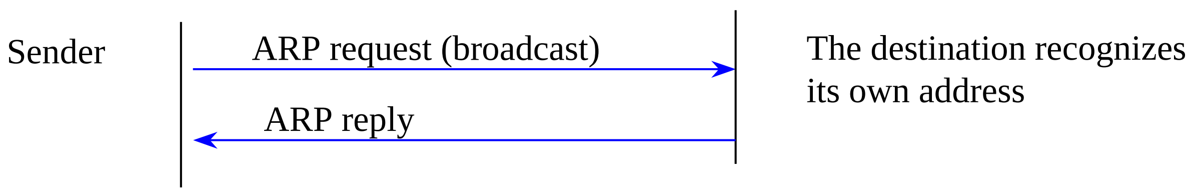
0 - Echo reply (used for “ping”)
3 - Destination unreachable
4 - ~~source quench~~ (=“slow down”)
(dropped from recommendations)
5 - Redirect
8 - Echo (used for “ping”)
9 - Router advertisement
10 - Router solicitation
11 - Time exceeded
12 - Parameter problem
13 - Timestamp
14 - Timestamp reply
15 - 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 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 destination (or router) 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.

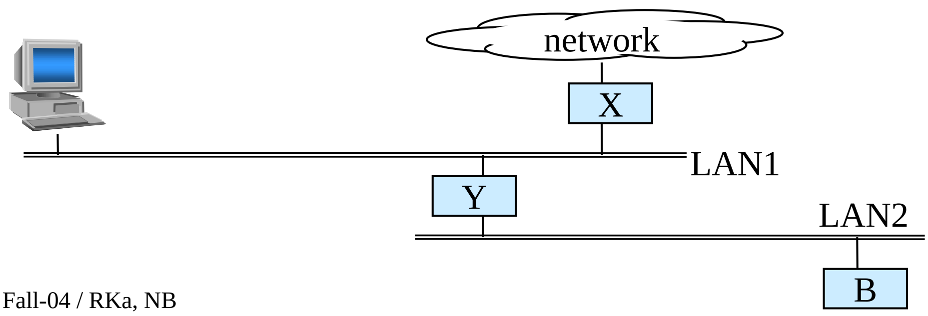
ARP – Address Resolution Protocol

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)

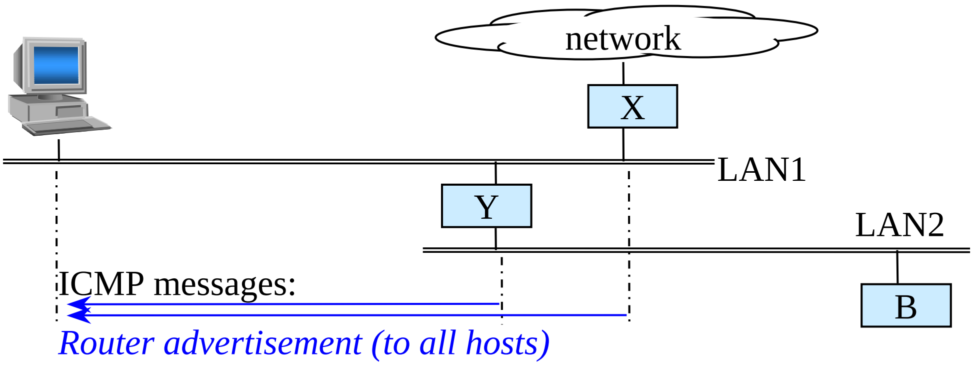
Router discovery

- 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)

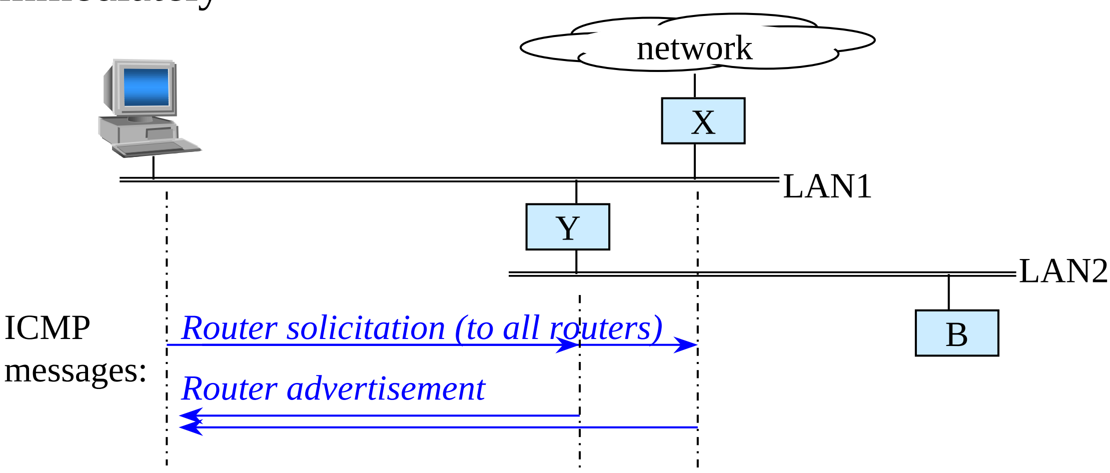
- The routers send *router advertisements* to all hosts periodically (e.g. in 7 minute intervals)



- 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

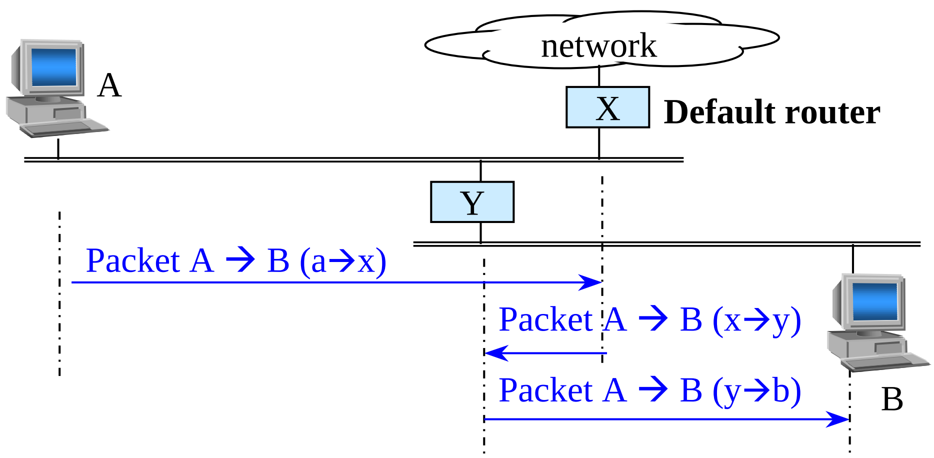


ICMP router discovery (3)

- The host discards advertisements from routers outside its sub-network and chooses the router with the highest priority as its default router.
- All packets for destinations outside the sub-network are sent to the default router.

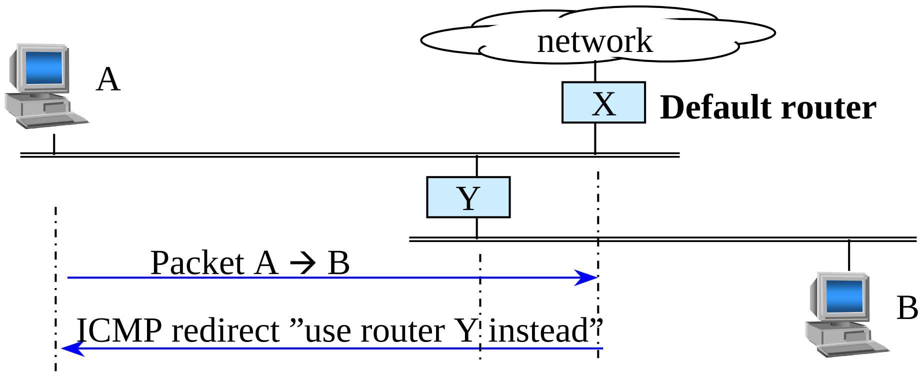
A network may have many routers, the closest to the destination must be found

- 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

Routing algorithms

- 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

Properties of the routing algorithms

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