La Sapienza
Università degli Studi di Roma
Dipartimento di Informatica e Sistemistica

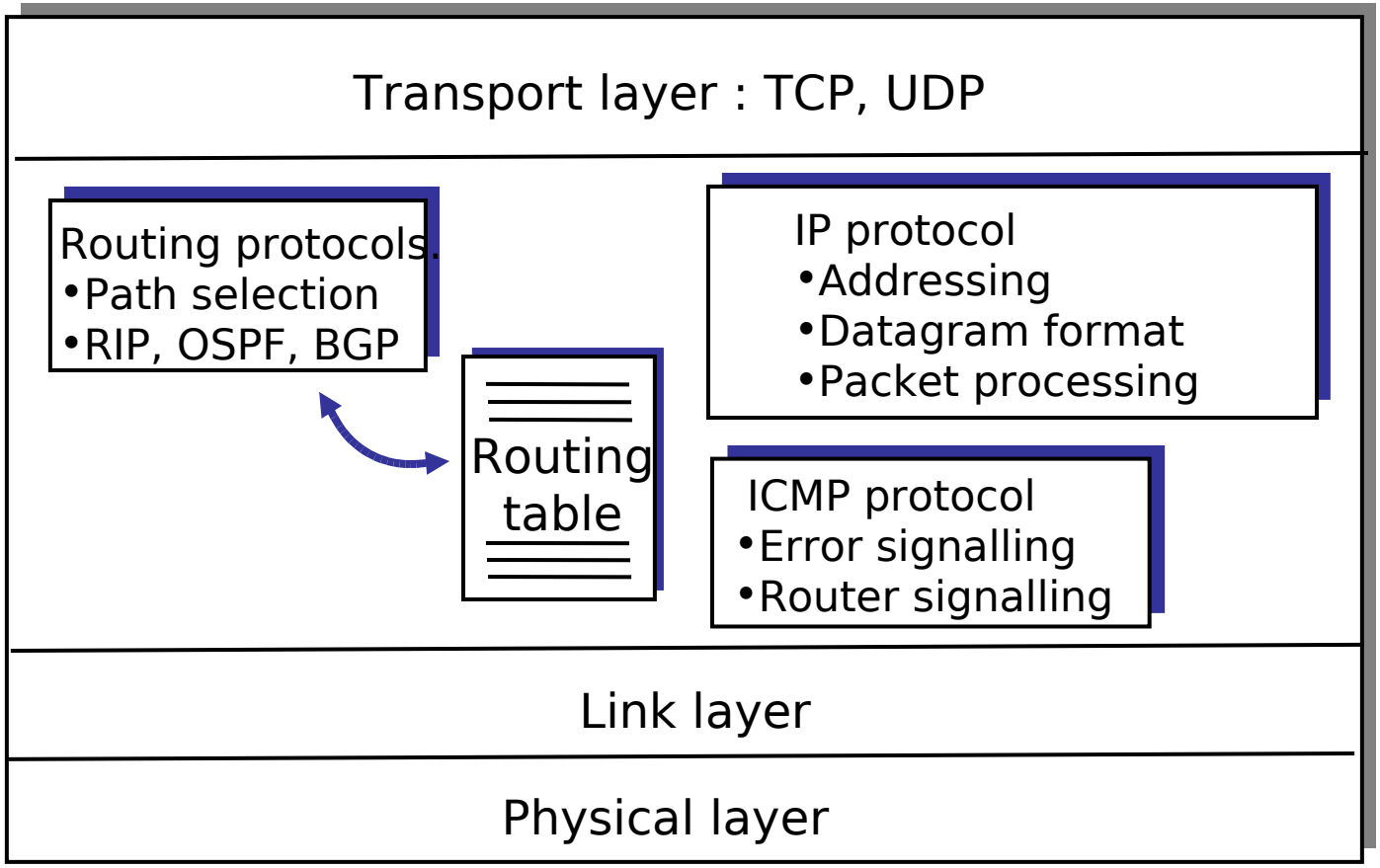
Computer Networks II
IPv4 recap

Luca Becchetti
Luca.Becchetti@dis.uniroma1.it
A.A. 2009/2010

Network layer's functionalities

- *Route discovery:* compute source – destination path followed by packets
  - *Routing algorithms -> Routing tables*
- *Forwarding:* switching of packets from router's input port to correct output port
- *Addressing*
- *Call set-up (if necessary)*
  - Not present in current Internet

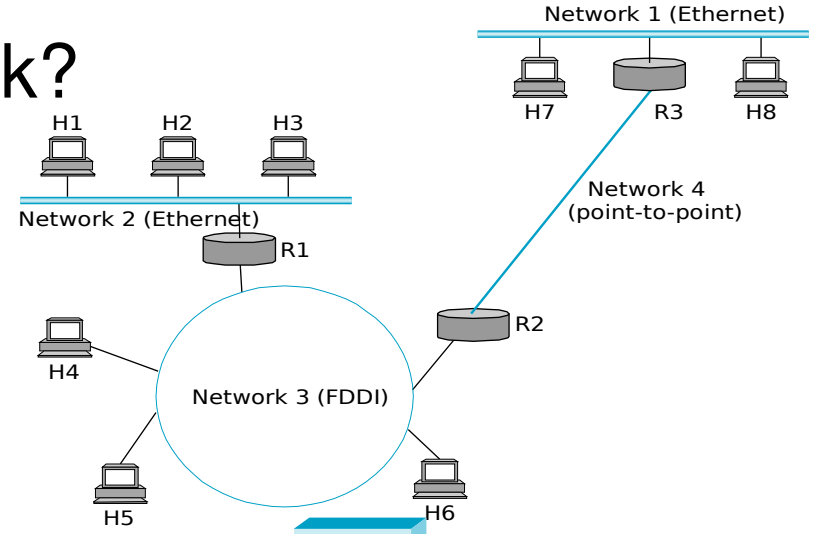
Internet's network layer



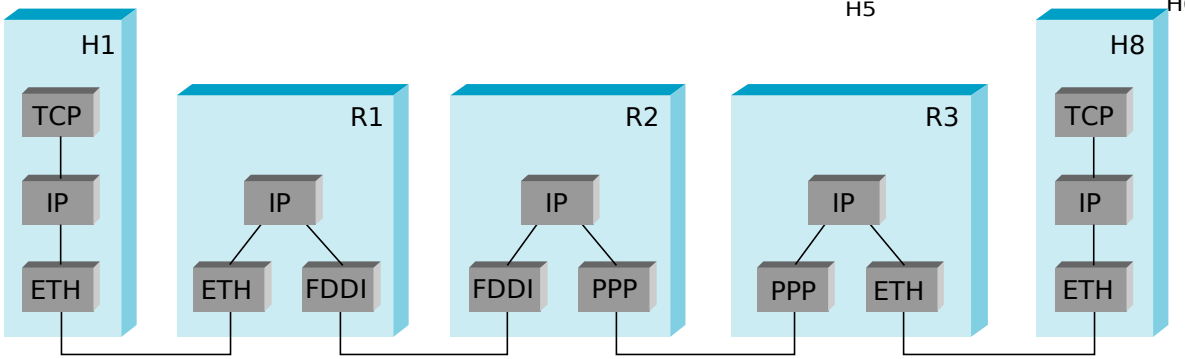
- Set of protocols, not just IP!

IP Internet

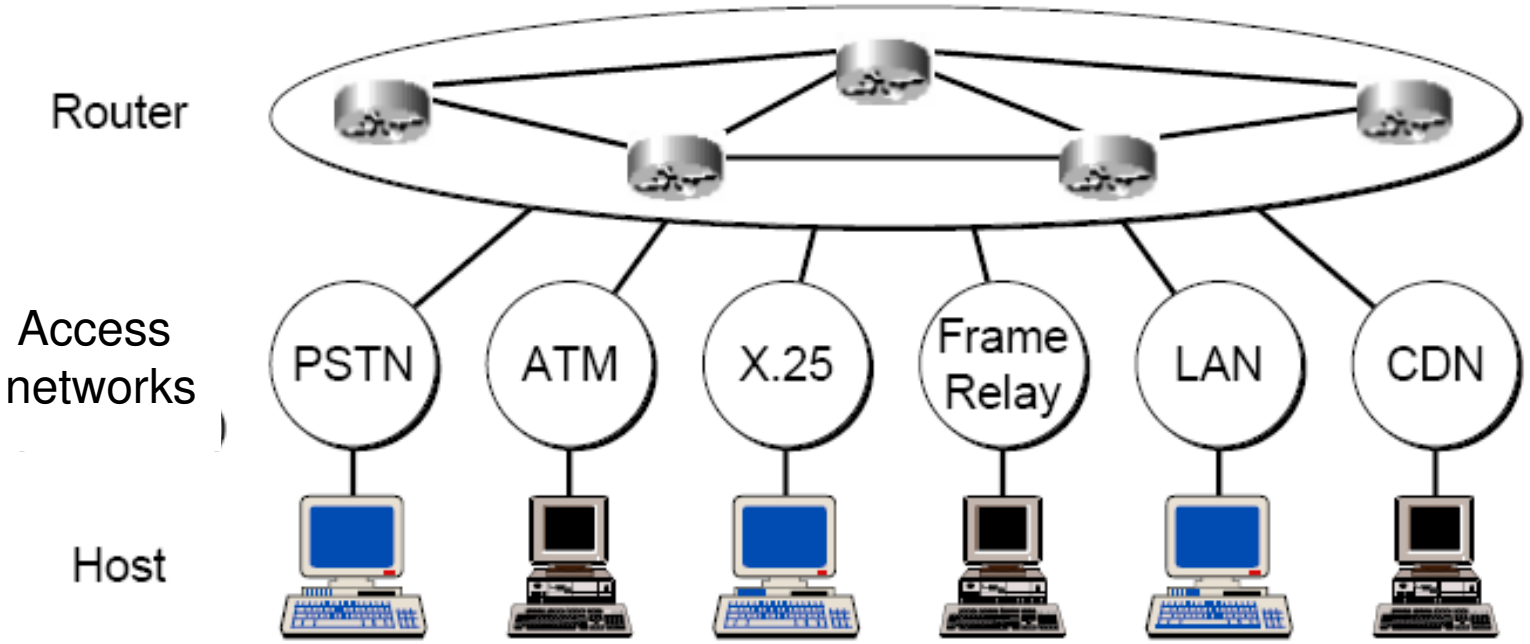
- Set of interconnected networks
  - What is an internetwork?



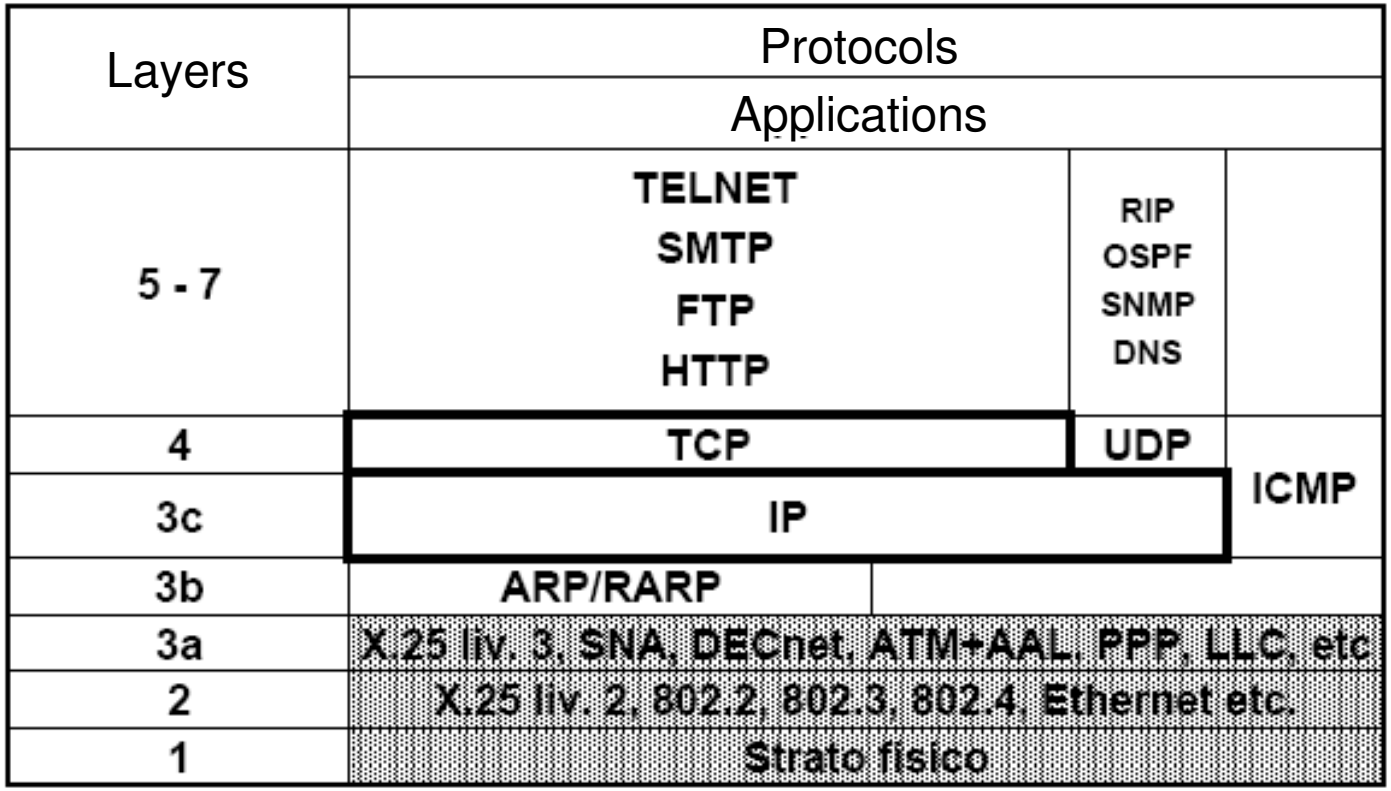
- Protocol Stack



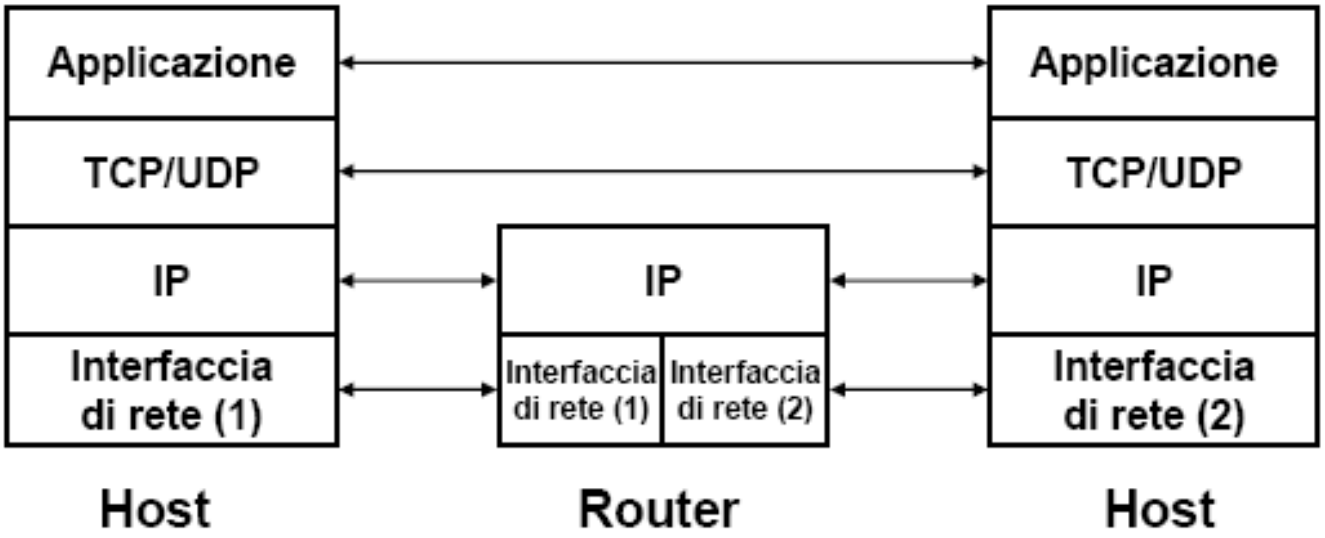
Internet structure



Protocol architecture



Protocol architecture



Routers implement IP, ICMP and routing protocols

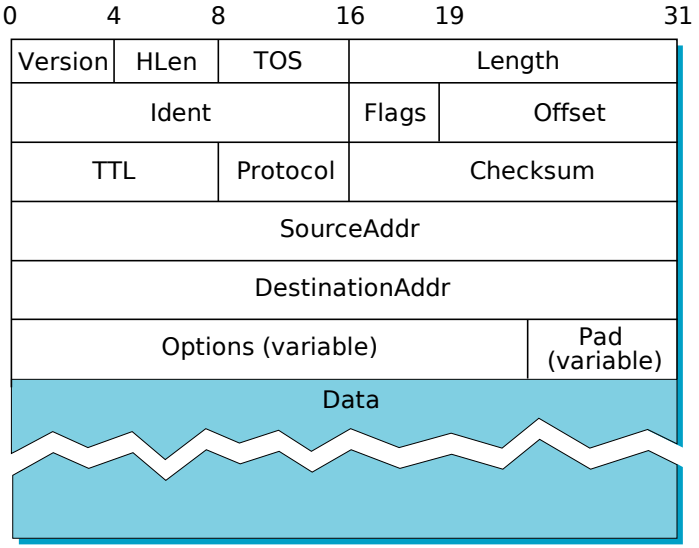
Protocol architecture

IP protocol (RFC 791, 919, 922, 950, 1349):

- Network layer protocol
- Connectionless transfer
- No QoS guarantees (“best effort” service)
- Defines:
  - Protocol Data Unit (PDU: datagram)
  - Max. datagram's size: 65536 bytes
  - Addressing scheme
  - Datagram routing mechanisms
  - Rules for fragmentation/reassembly of datagram units

Service model

- Connectionless (datagram-based)
- Best-effort delivery (unreliable service)
  - packets can be lost
  - packets can be delivered out of order
  - duplicate copies of a packet can be delivered
  - packets can be delayed for a long time
- Datagram format



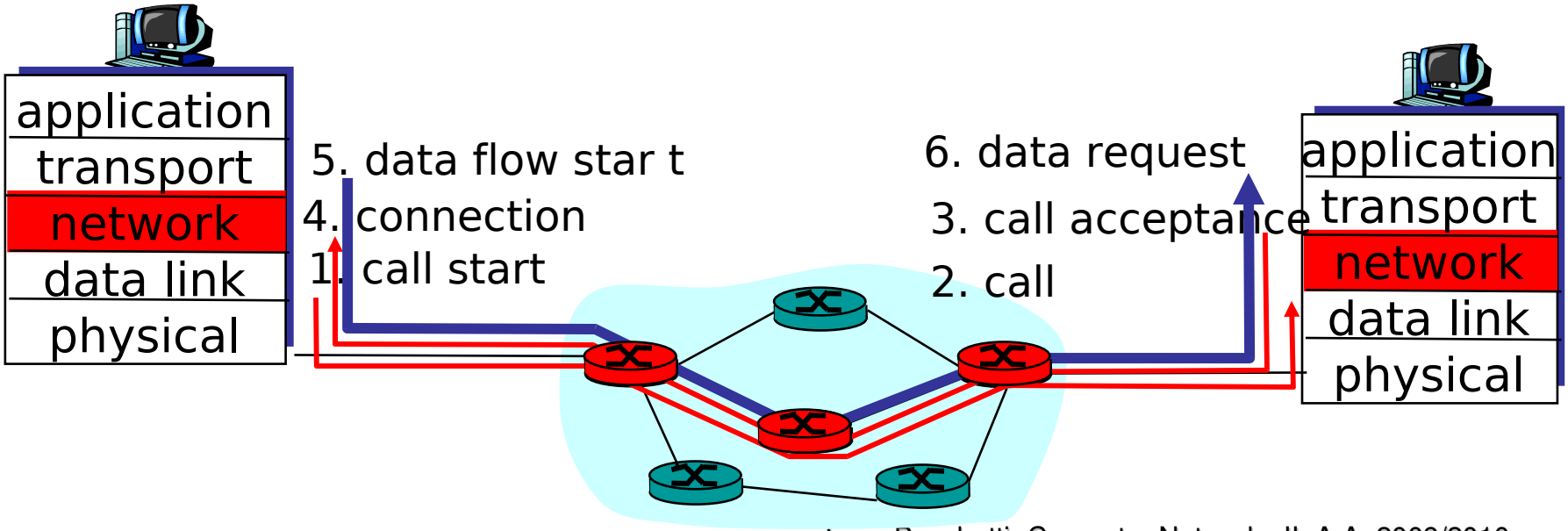
Alternative - virtual circuit

"source-destination path behaves like a telephone circuit
- Tracks performance indices
- Network layer active along source-destination path
- ATM, Frame-Relay, X.25

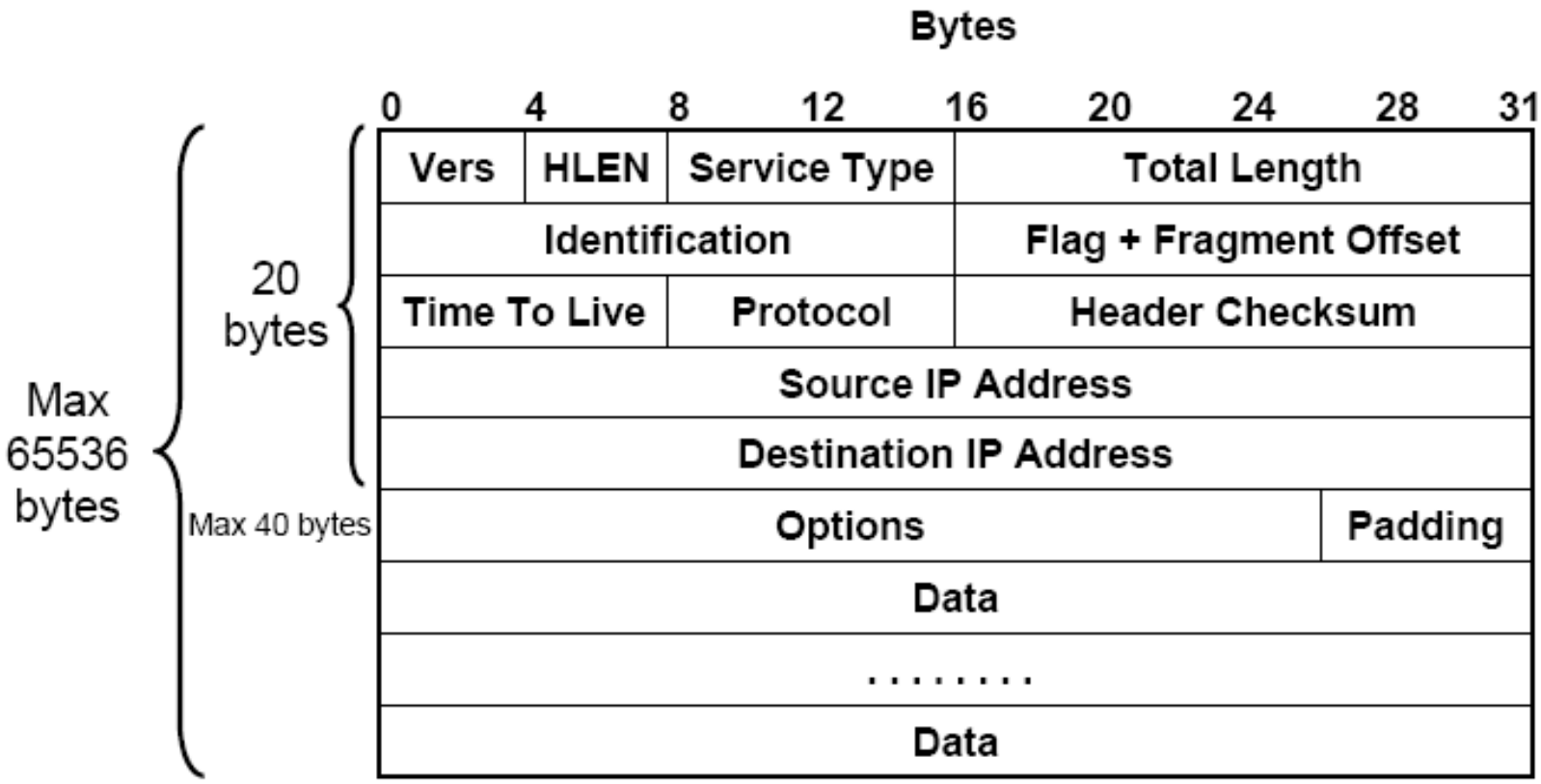
- Call set up and shut down
- Every packet has a VC identifier (not destination address)
- *Every* router along source-destination path maintains state on every traversing connection
- Possible resource allocation to VCs (bandwidth, buffers)

Virtual circuit: signalling

- Used for call set up, maintenance and tear down
- Used in ATM, frame-relay, X.25
- Not used in Internet at present



IPv4 datagram format

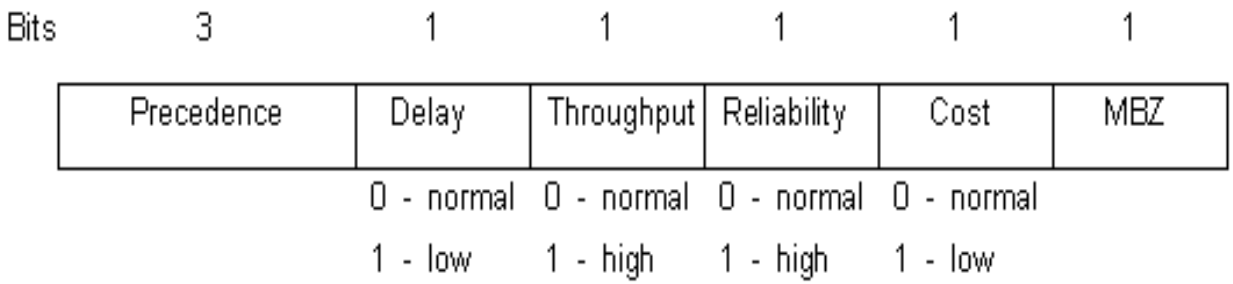


IPv4 datagram format

- **Vers (4 bits)**
  - Protocol version, different version may coexist
- **Header Length (HLEN) (4 bits)**
  - Header length (specified as 32 bit words)
  - Includes fixed part (20 bytes) and optional part
  - Max. size: 60 byte
- **Total length (16 bits)**
  - Overall datagram's length (in bytes)
  - Includes header and payload
  - Max. value: 65536 byte

IPv4 datagram format

- **Service Type (8 bits)**
  - QoS parameters required from sender
  - Precedence (3 bits)
    - Datagram's priority
    - Not used in the past
    - Now implementing DiffServ mechanisms [es. VoIP: 0x0A, 0xE0]
  - Type of Service [ToS – 4 bits]
    - Type of service requested for datagram
    - Default service when all ToS bits are 0



IPv4 datagram format

- **Service Type (8 bits)**
  - Examples

0	0	0	0	1	0	1	0	**0x0A**
1	1	1	0	0	0	0	0	**0xE0**

IPv4 datagram format

- **Time to Live (TTL) (8 bits)**
  - Max. no. of routers a datagram may traverse
  - Initialized by source host and decrements by every traversed router
  - When TTL = 0 → datagram dropped, ICMP notification message to sending host
- **Protocol (8 bits)**
  - Identifies transport layer protocol two which payload has to be delivered-*demultiplexed* (es. TCP=6, UDP=17, ICMP=1)
- **Header Checksum (16 bits)**
  - Protects only datagram's header
  - Drop datagram if error detected

IPv4 datagram format

- **Source Address (32 bits) and Destination Address (32 bits)**
- **Options (variable size in bytes)**
  - Record Route Option (RRO)
    - Empty IP address list, every traversed router inserts own address
  - Timestamp Option
    - Like RRO + time at which every route is traversed
  - Loose Source Routing Option (LSRO)
    - Specifies list of routers the datagram *must* traverse
  - Strict Source Route Option (SSRO)
    - Specifies list of *all* routers the datagram *must* traverse
- **Padding**
  - Used to ensure that overall packet length is multiple of 32 bis [4 byte]
  - **Options at most 40 bytes - Why?**

Fragmentation and Reassembly

- Each network has some MTU
  - Ethernet -> 1500 byte
  - Minimum MTU value is 68 byte
  - Fragmentation and reassembly are needed when the MTU of the physical subnet is lower than IP packet size
- Design decisions
  - fragment when necessary (MTU < Datagram)
  - try to avoid fragmentation at source host
  - re-fragmentation is possible
  - fragments are self-contained datagrams
  - delay reassembly until destination host
  - do not recover from lost fragments

Fragmentation and Reassembly

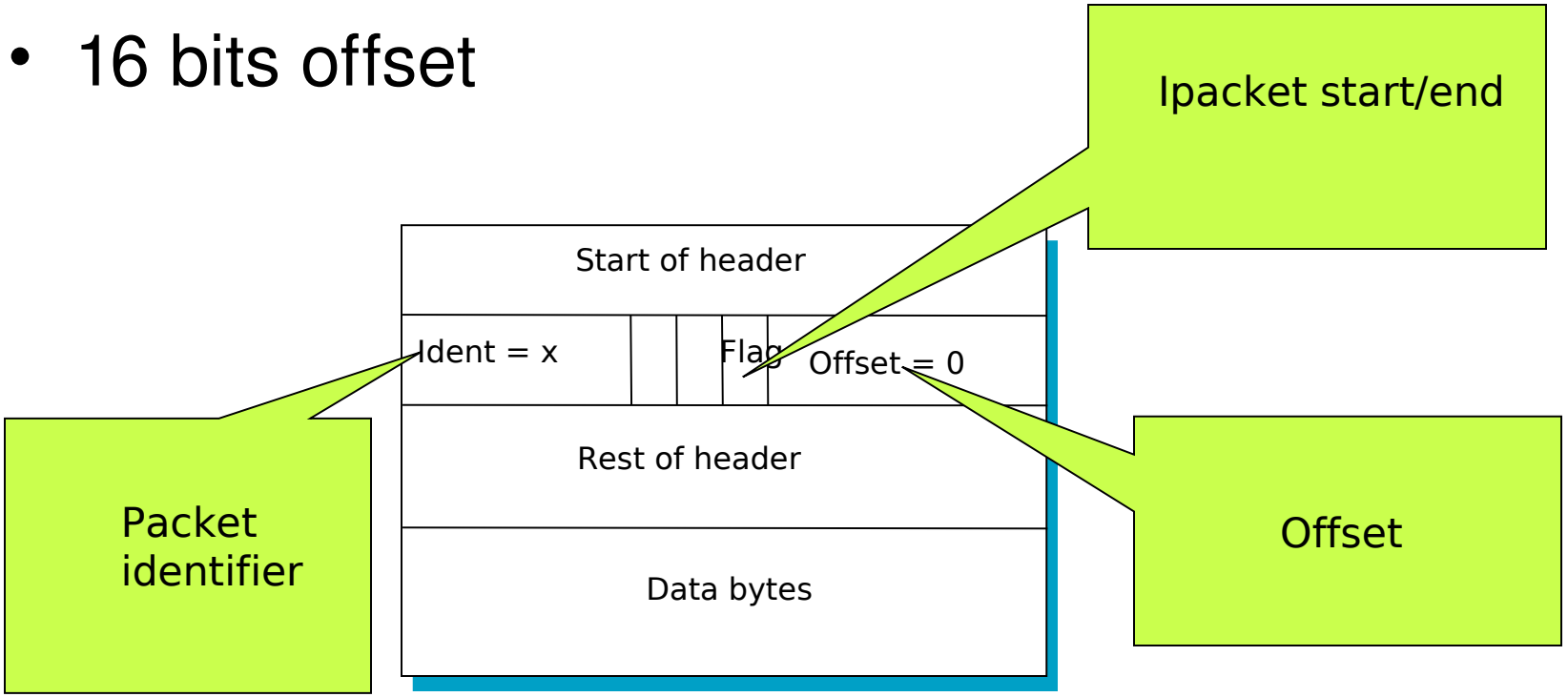
Identification (16 bits)

- Datagram Identifier
- Assigned by source process – same for all fragments

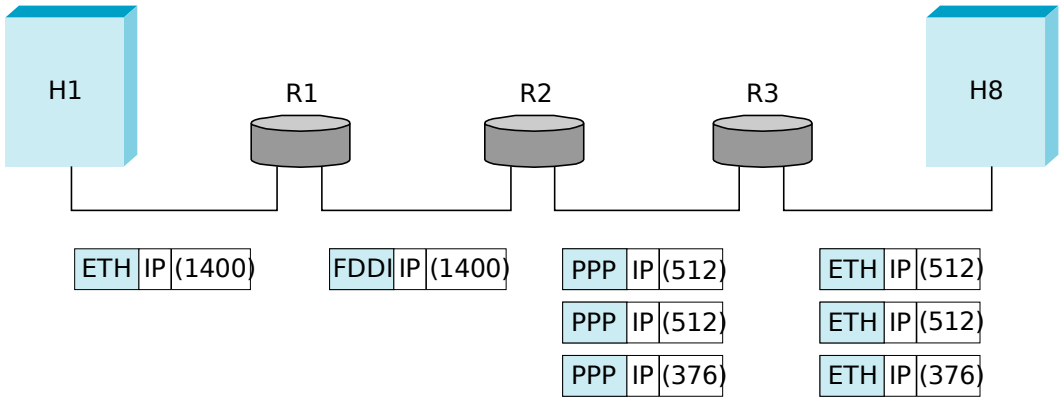
- **Flags (3 bits)**
  - X: not used, must be zero
  - DF: Don't Fragment (0: fragmentation allowed; 1:no fragmentation)
  - MF: More Fragments (0: last fragment; 1: More Fragments to follow)
- **Fragment Offset (13 bits)**
  - Position of fragment within initial datagram's payload (expressed as 8 bytes units) – allows to detect "holes"

Fragmentation/cont.

- 16 bits offset



Example



(a)

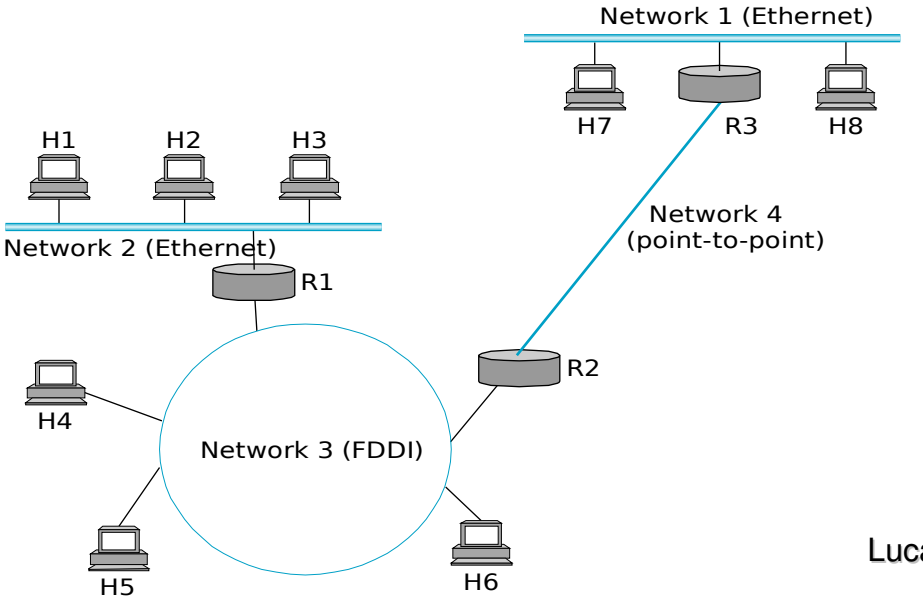
Start of header				
Ident = x			0	Offset = 0
Rest of header				
1400 data bytes				

(b)

Start of header				
Ident = x			1	Offset = 0
Rest of header				
512 data bytes				

Start of header				
Ident = x			1	Offset = 64
Rest of header				
512 data bytes				

Start of header				
Ident = x			0	Offset = 128
Rest of header				
376 data bytes				

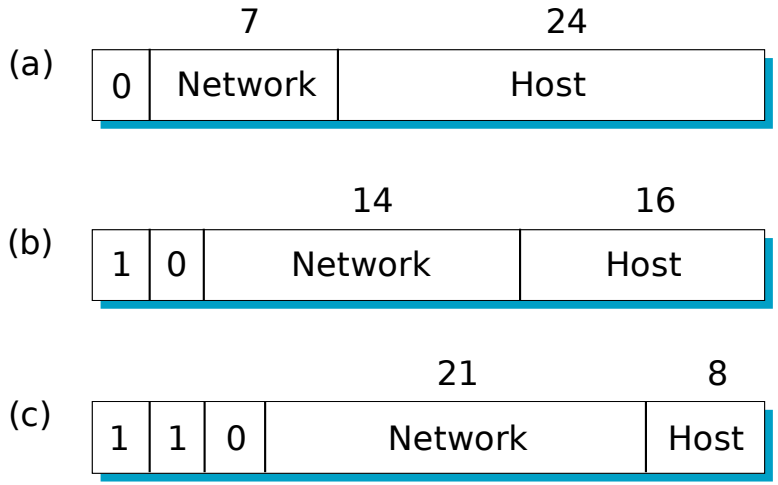


IPv4 addressing scheme

- IP Address identifies host
  - If host connected to more than 1 network (multi-homed) one IP address per network
- IP address unique in network it belongs to
  - Length: 32 bits
- Originally (1981, RFC 1166) consisted of 2 parts
  - Net_Id: subnet identifier
  - Host_Id: host identifier within subnet
  - IP_Address = Net_Id . Host_Id
  - Boundary between Net_Id and Host_Id not fixed [address classes, subnetting]

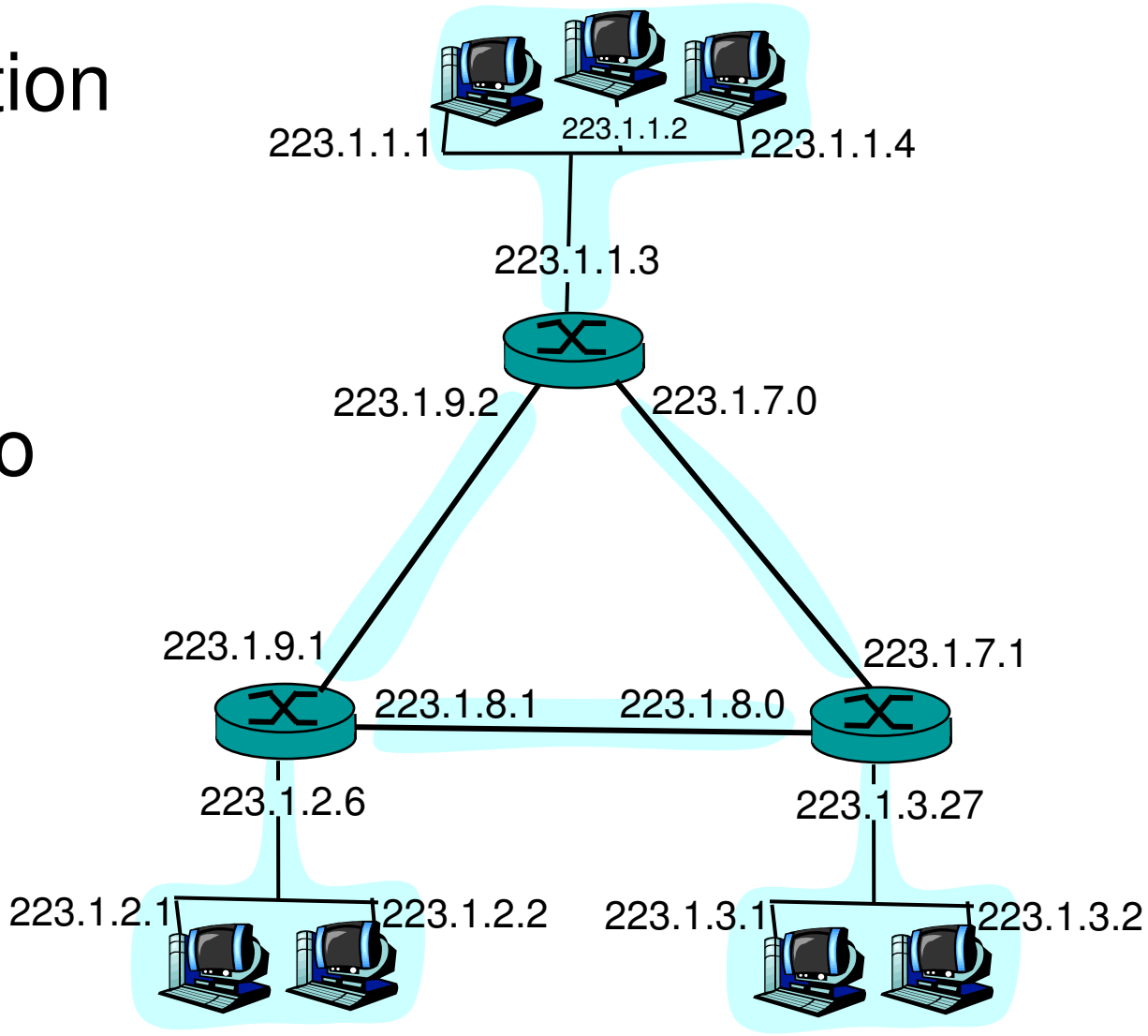
Global Addresses

- Properties
  - globally unique
  - hierarchical: network + host
- Dot Notation
  - 10.3.2.4
  - 128.96.33.81
  - 192.12.69.77



IP subnetworks

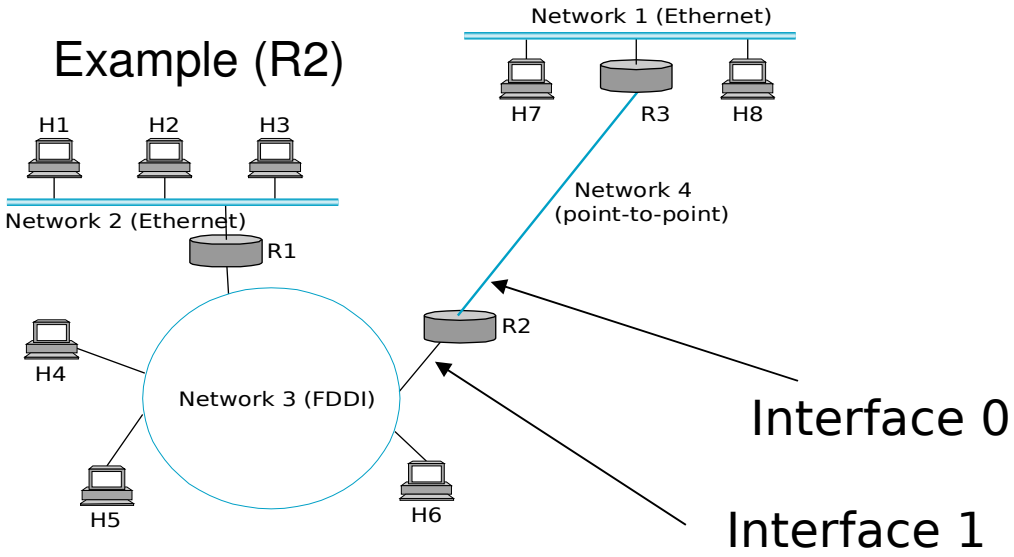
- Interconnection of ? IP networks
- IP address associated to interfaces



Datagram Forwarding

- Strategy
  - every datagram contains destination's address
  - if connected to destination network, then forward to host
  - if not directly connected, then forward to some router
  - forwarding table maps network number into next hop
  - each host has a default router
  - each router maintains a forwarding table

- Example (R2)



Network number	Next hop
1	R3
2	R1
3	interface1
4	interface0

Address Translation within a LAN

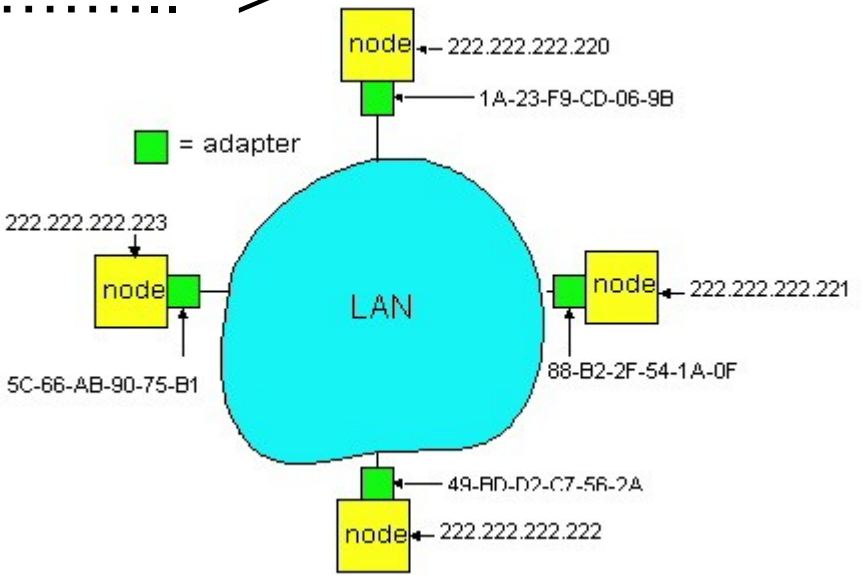
- Map IP addresses into physical addresses
  - destination host
  - next hop router
- Techniques
  - encode physical address in host part of IP address
  - table-based
- ARP
  - table of IP to physical address bindings
  - broadcast request if IP address not in table
  - target machine responds with its physical address
  - table entries are discarded if not refreshed

Address **R**esolution **P**rotocol (RFC 826)

- Every IP node (Host, Router) on a LAN has **ARP** module and **ARP** table
- ARP table: maps IP->MAC for **some** (occasionally all) nodes in LAN

  < IP address; MAC address; TTL>

  < …….. >
- TTL (Time To Live): timer, usually
- 10 - 20 min)



ARP/cont.

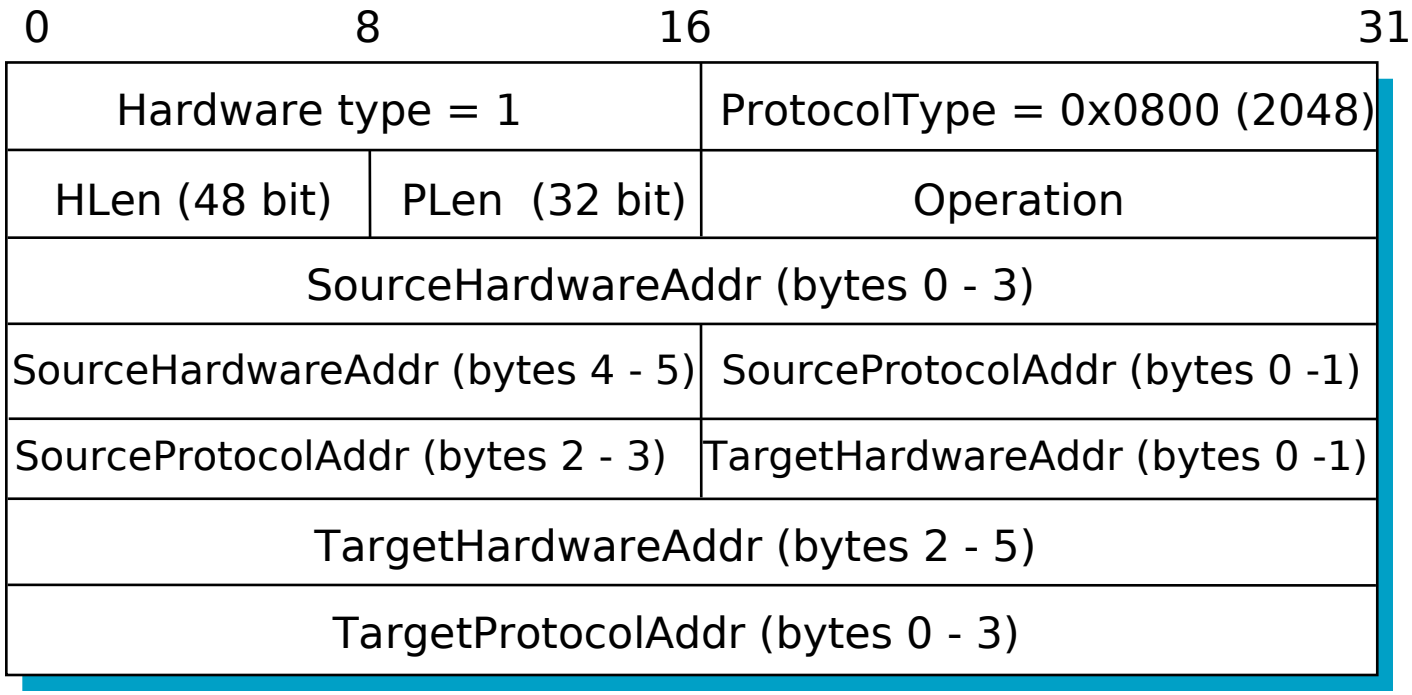
- Node A must send IP packet to IP address XYZ on a given LAN
- A first checks local ARP table if no entry corresponding to XYZ in table, A' s ARP module sends **broadcast** ARP packet:

  < XYZ, MAC (?) >
- All nodes in LAN accept and analyze ARP packet
- Node XYZ replies with ARP **unicast** packet containing own MAC address:

  < XYZ, MAC (XYZ) >
- ARP table acts as cache
- ARP protocol can be used with different technologies

ARP Packet Format

May operate over different layer 2 protocols (not only Ethernet)

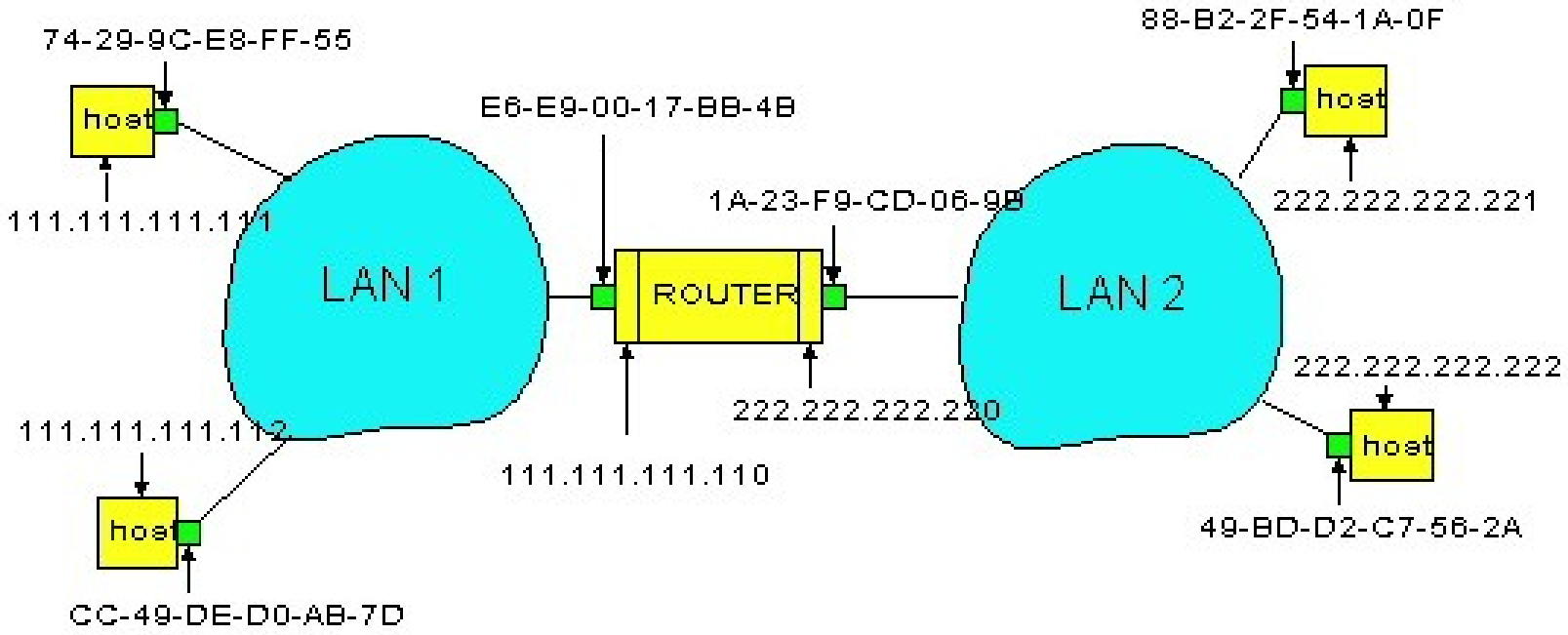


ARP Details

- Request Format
  - HardwareType: type of physical network (e.g., Ethernet)
  - ProtocolType: type of higher layer protocol (e.g., IP)
  - HLEN & PLEN: length of physical and protocol addresses
  - Operation: request or response
  - Source/Target-Physical/Protocol addresses
- Notes
  - table entries timeout in about 10 minutes
  - refresh table if already have an entry
  - otherwise
    - update table with source when you are the target
    - do not update table if not target

Es.: Routing towards different LAN

- Packets from. <111.111.111.111> to <222.222.222.222>

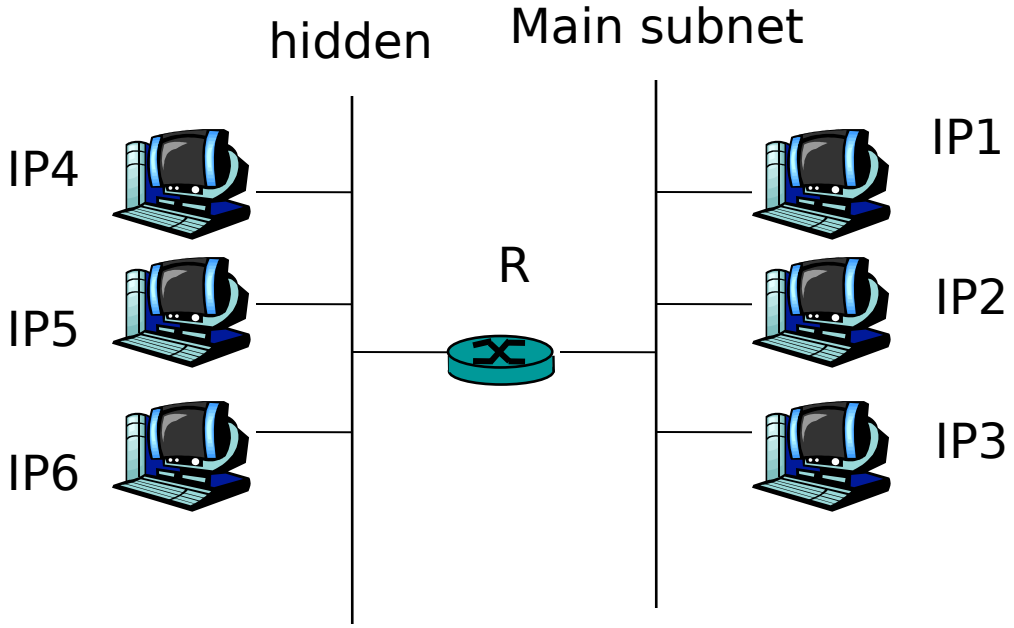


- Find Next hop in Routing table (111.111.111.110)
- Find corresponding MAC address (E6-E9-00-17-BB-4B) in ARP table

ARP proxy

- Allows to partition a network into subnetworks
  - Only larger networks visible outside
  - Successively specified LANs are not visible outside
- Needs special router that:
  - Acts as a switch among different subnets
  - Acts as gateway to/from exterior of the network
- Routers ignore presence of distinct subnets
- Used in the past
- At present used only for specific purposes (e.g., mobile IP)

ARP proxy - cont.



- Dgram IP1-->IP4
  - R captures broadcast ARP request from IP1 and returns own MAC address
  - Datagrams from IP1 to IP4 sent to R and then forwarded to IP4

Pros/cons

- Pros
  - No need to modify RTs of other routers
  - E.g.: external routers need only know that datagrams with destination IP4 must be sent to R
- Cons
  - Routing not entirely automatic
  - Network administrators must manually update routing tables

Internet Control Message Protocol (ICMP)

- Echo (ping)
- Redirect (from router to source host)
- Destination unreachable (protocol, port, or host)
- TTL exceeded (so datagrams don’t cycle forever)
- Checksum failed
- Reassembly failed
- Cannot fragment

References

- Kurose and Ross' textbook
  - Chap.4, in particular 4.1 - 4.5
- Peterson and Davie's textbook
  - Chap. 4, in particular 4.1
- The TCP/IP guide
  - http://www.tcpipguide.com/free/index.htm