Networking v0.9 – Internet Protocol
(3TC / IST / Bachelor / L3)
2012

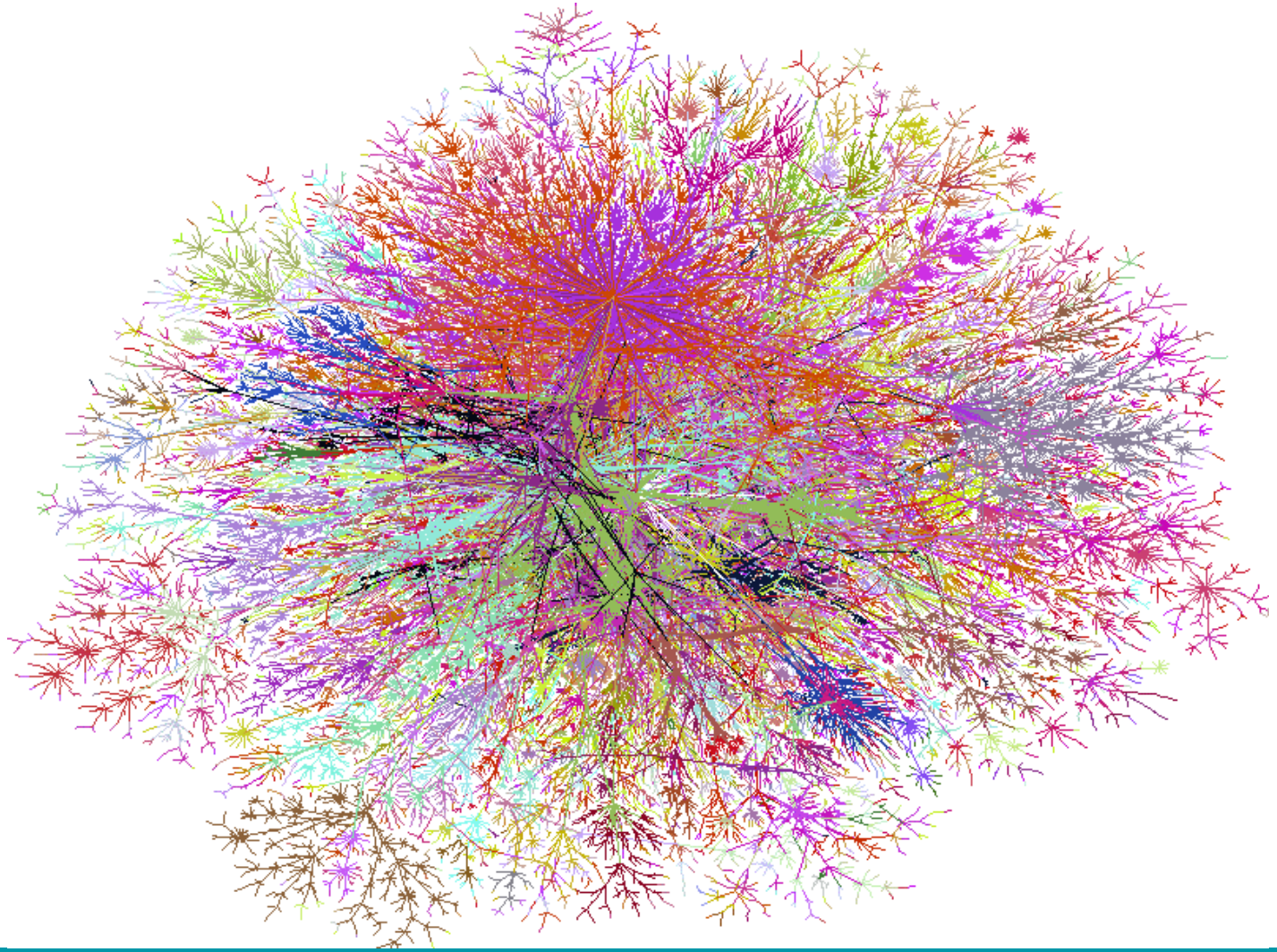
Fabrice Valois, fabrice.valois@insa-lyon.fr

membre de UNIVERSITÉ DE LYON

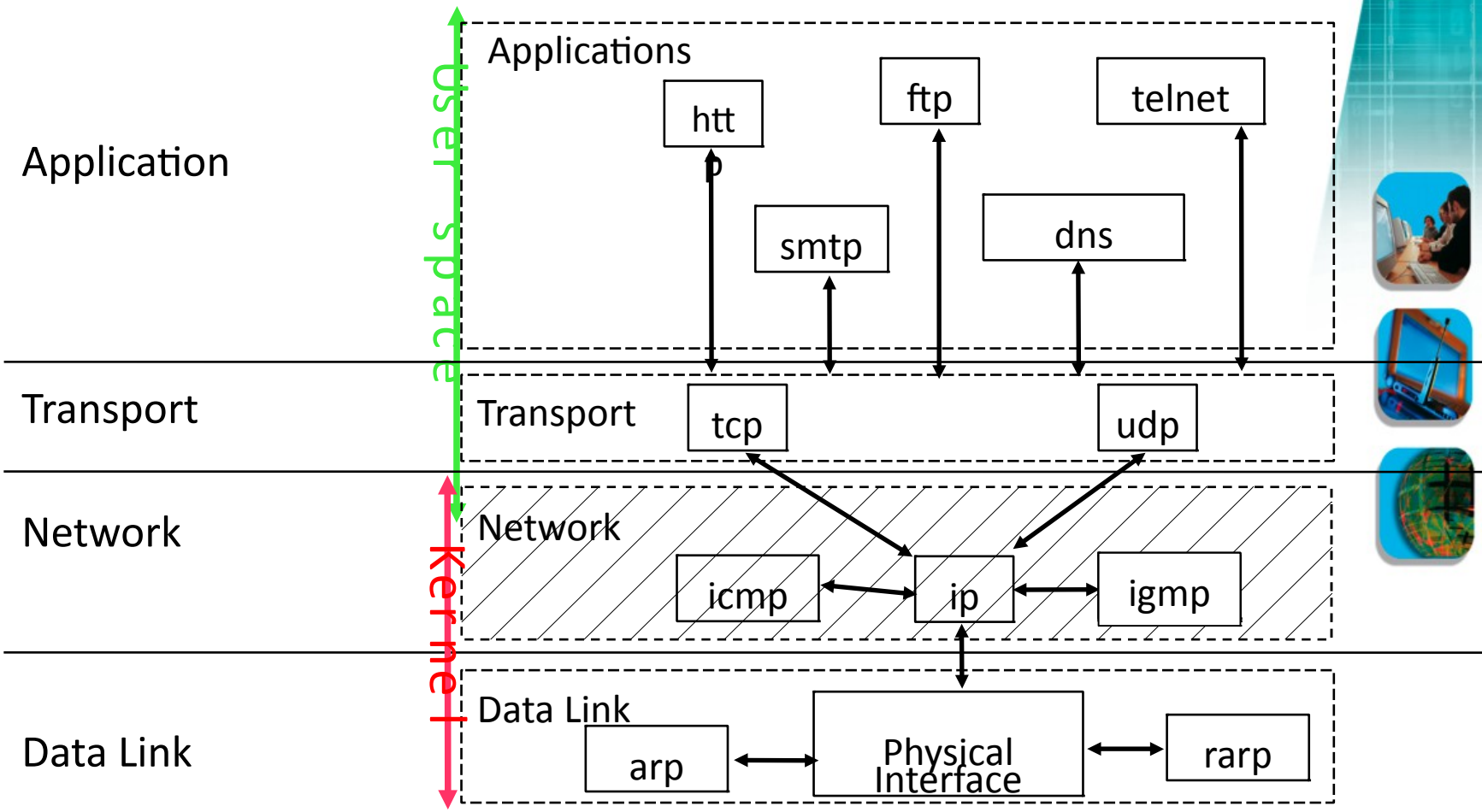
Département Télécommunications
Services & Usages

INSA LYON

General overview of IP



General overview (...)



Remember the warning!!!

* In Networking (and telecommunications), everything has an abbreviation and we always use it....

ARP, VLAN, IP, WI-FI, TCP, DQDB, UDP, P-NAT, TOS, MAC, BAN, FDDI, RSVP, BGP, RTP, VoIP, xTP, SMTP, WAN, ICMP, IGMP, LAN, RARP, OSI, VPN, MPLS, DNS, NAT, QoS, CSMA, HTTP, …

* Be care and be patient...

Chapter 5

IP Protocols : Basic mechanisms, routing protocol and fragmentation

membre de
UNIVERSITÉ DE LYON

Département Télécommunications
Services & Usages

INSA LYON

Agenda

* Back to the fundamentals of IP
* Remember Ethernet Encapsulation and ARP
* Packet format and algorithms
  - Address and routing protocol
  - Fragmentation

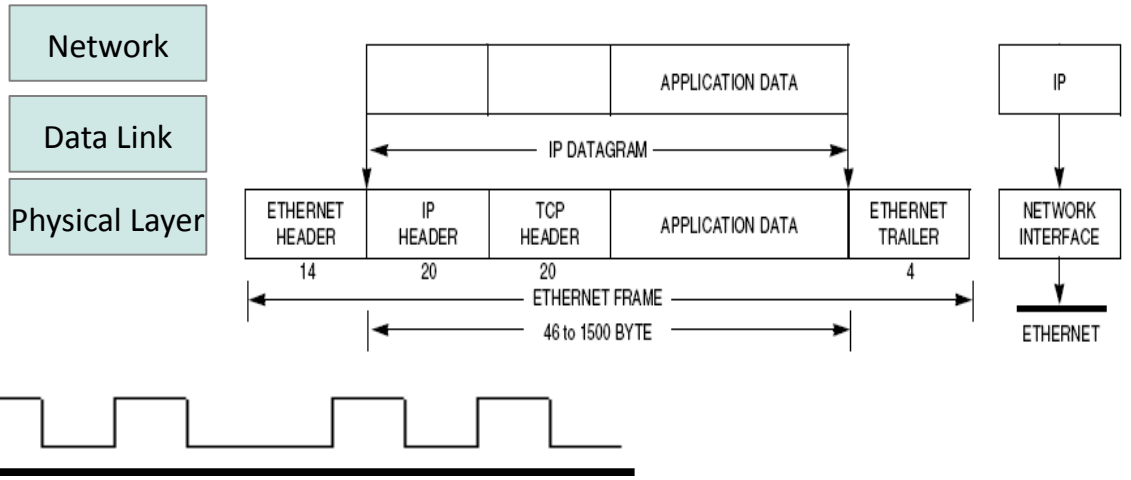


Fundamentals of IP

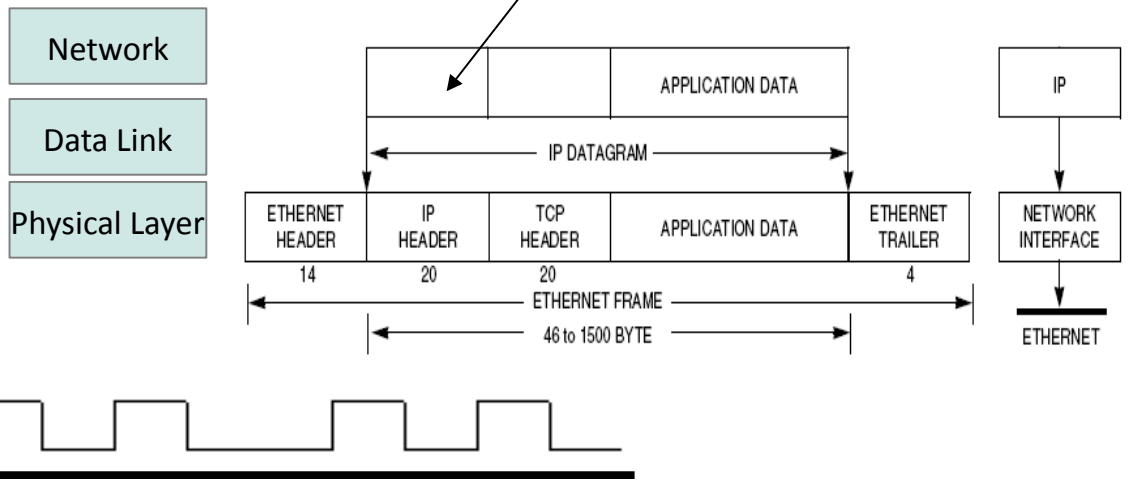
- **Internet Protocol : the heart of Internet**
- **Packet switching, datagram-based protocol**
- **<u>Non-reliable</u> protocol**
  - Robustness (if needed...) is provided by upper layers (transport, eventually application)
  - Simple error management
- **<u>Non-connected</u> protocol**
  - No information about the state of the connection and/or the destination
  - Packets are routed independently
- **Routing protocol, packet fragmentation, basic flow control**
- **rfc 791 [Postel 1981a]**



IP over Ethernet encapsulation



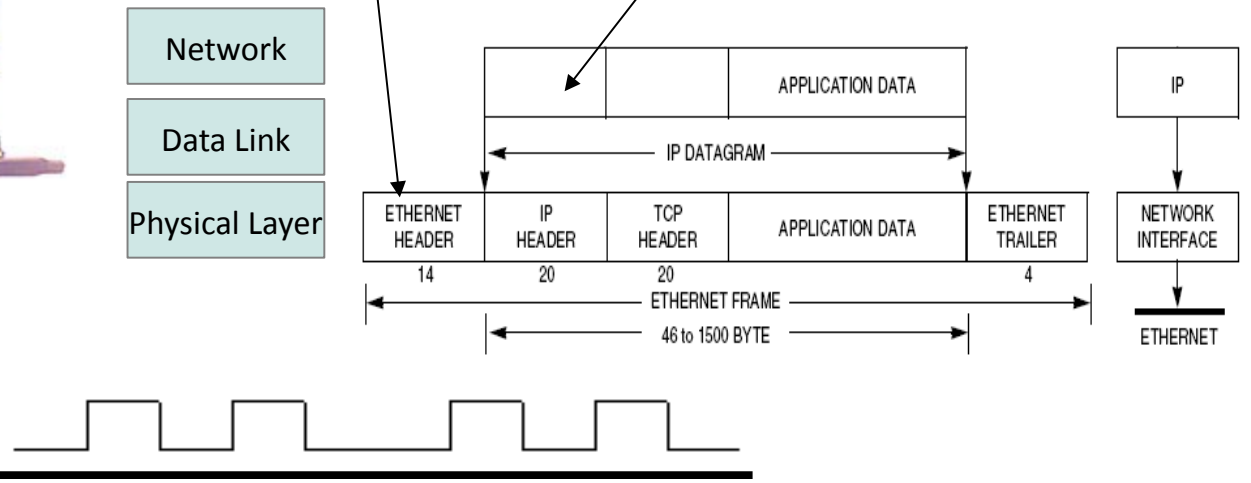
IP over Ethernet encapsulation



IP over Ethernet encapsulation

Physical (hardware) addresses of the source and the destination (at least the next router)

IP addresses of the source and the destination

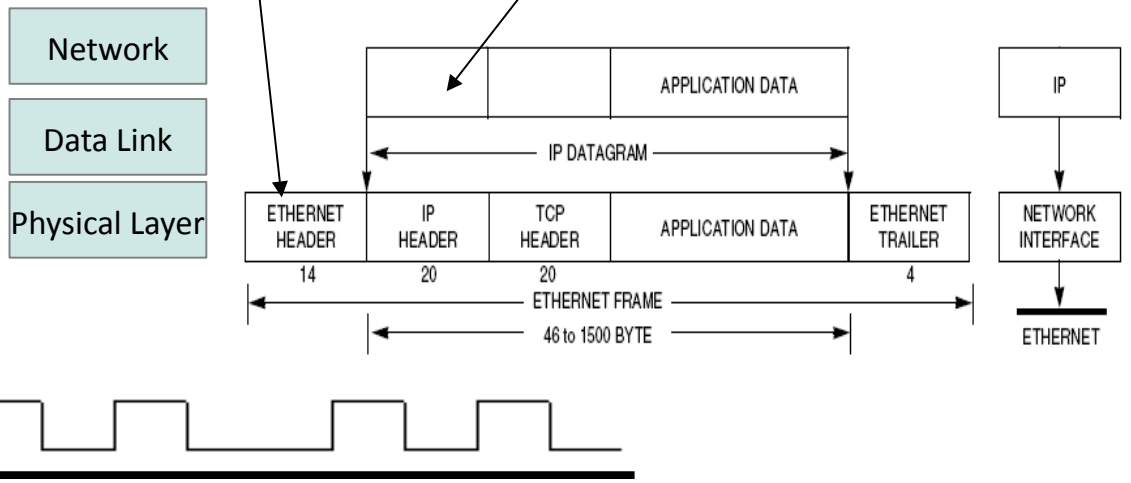


IP over Ethernet encapsulation

How to do the **translation** ?

Physical (hardware) **addresses** of the source and the destination (at least the next router)

IP addresses of the source and the destination



IP over Ethernet encapsulation

* ARP : Address Resolution Protocol
  - On-demand physical to logical address translation for Ethernet purpose (i.e. ARP : @IP → @MAC)
* Reverse ARP (@MAC → @IP)
* Exist also ARP Proxy



12

IP Datagram

Header + Data

Header Information

Version : 4

IHL (*Internet Header Length)*

#words of 32 bits in the header (including options)

ID, Flags, FO

00 00 si pas de fragmentation

Protocol (layer 4)

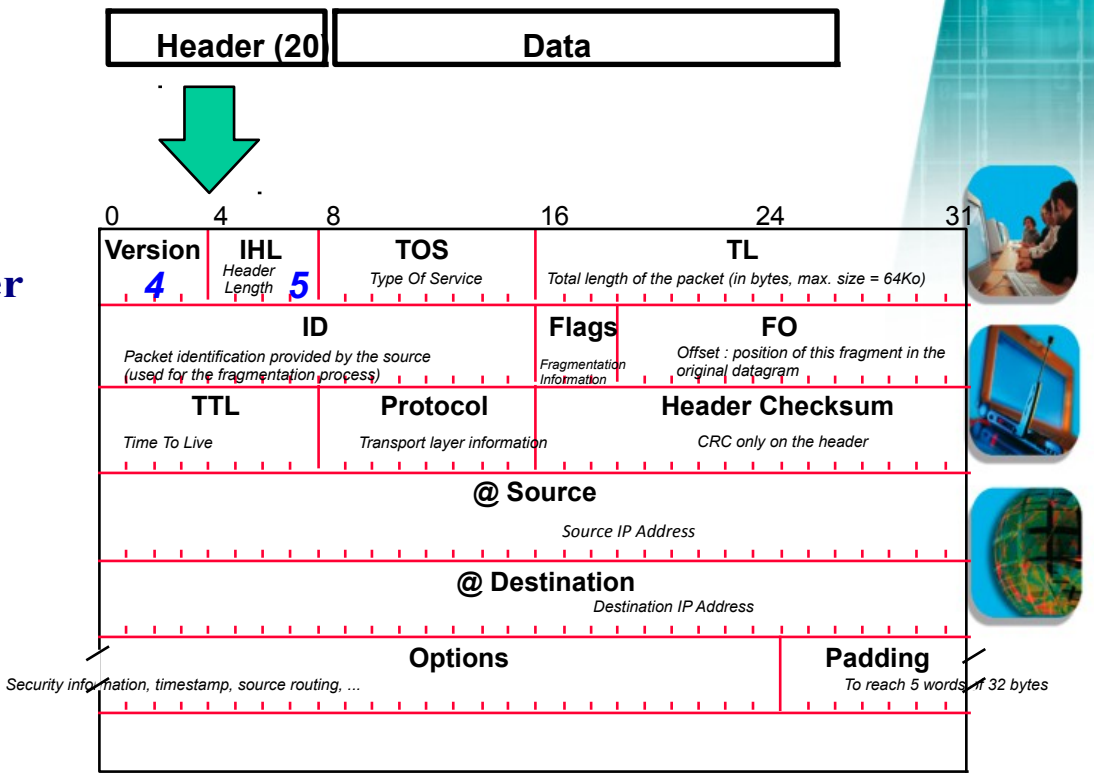
6 for TCP,

17 for UDP, ...

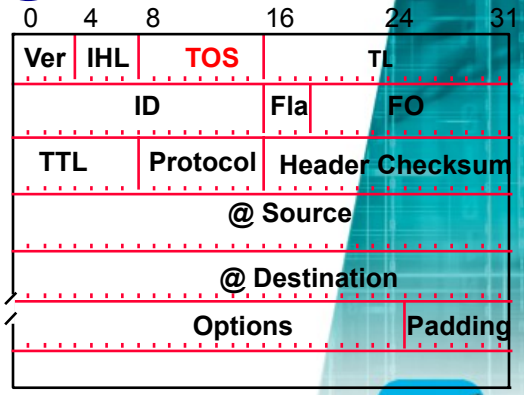
Classicaly :

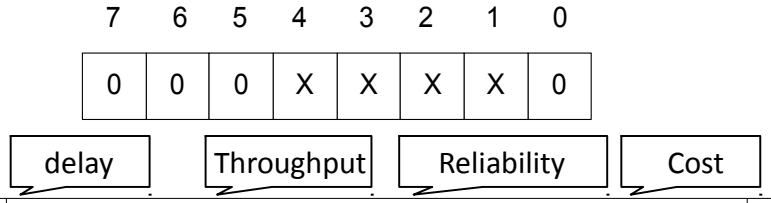
Header starts with : 45 00

@Source / @Destination



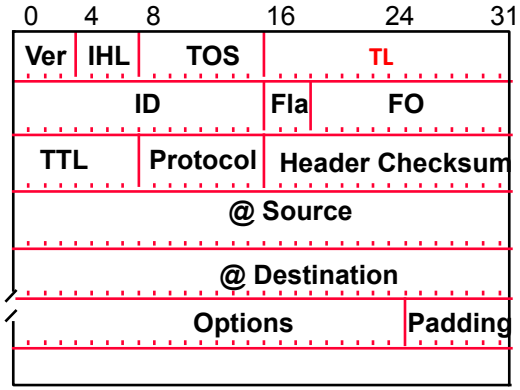
TOS : Type Of Service
(rfc 1340; 1349)





Application	Minimize the delay	Maximize the throughput	Maximize the reliability	Minimize the Cost	Value
Telnet/Rlogin	1	0	0	0	0x10
FTP					
Control	1	0	0	0	0x10
Data	0	1	0	0	0x08
Raw data	0	1	0	0	0x10
TFTP	1	0	0	0	0x10
SMTP					
Control and management	1	0	0	0	0x10
Data and Data management	0	1	0	0	0x08
DNS					
UDP request	1	0	0	0	0x10
TCP request	0	0	0	0	0x00
DNS Server operation	0	1	0	0	0x08
ICMP					
Error	0	0	0	0	0x00
Request	0	0	0	0	0x00
IGP	0	0	1	0	0x04
SNMP	0	0	1	0	0x04
BOOTP	0	0	0	0	0x00

Total Length (16 bits)

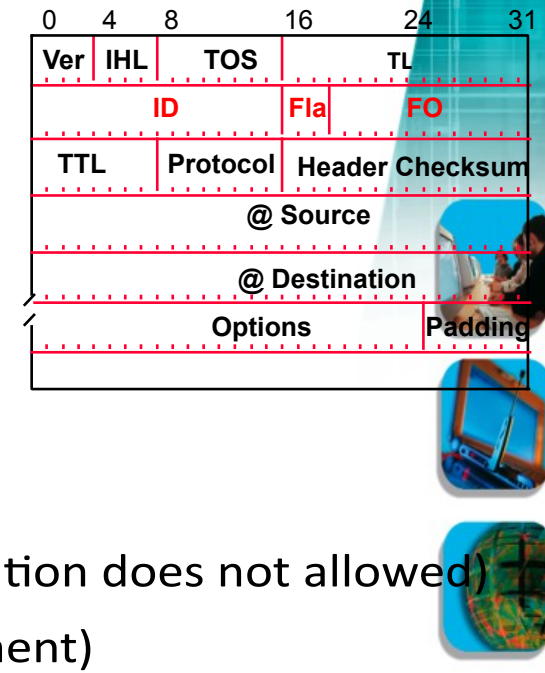


- Max size : 65535 bits
- Mainly 576 bytes
  - Not used by TCP
  - UDP requirement : 512 bytes
- In case of IP over Ethernet encapsulation, this field is used to isolate the padding bits
- This field is modified in case of fragmentation

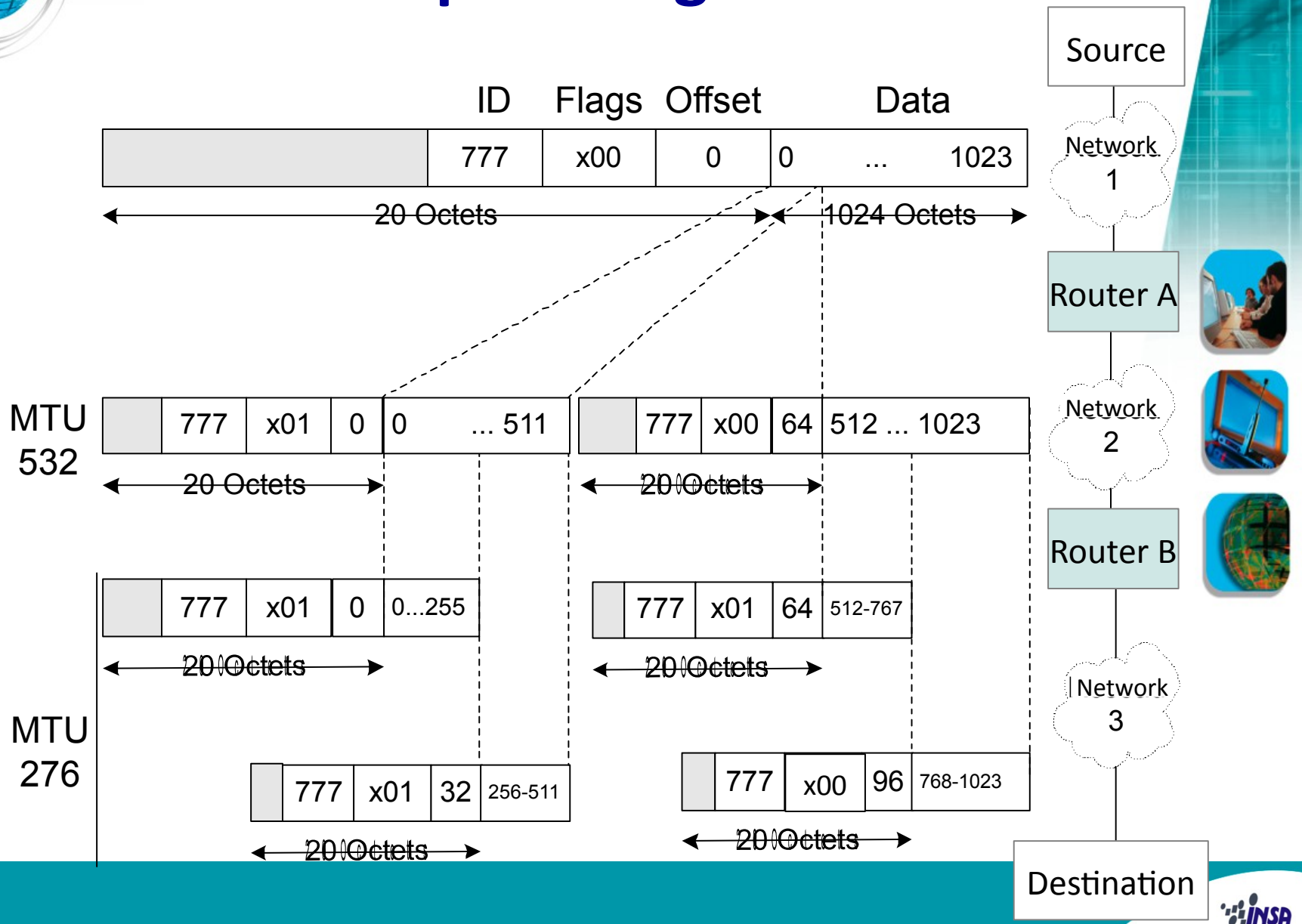


Identification & Fragmentation

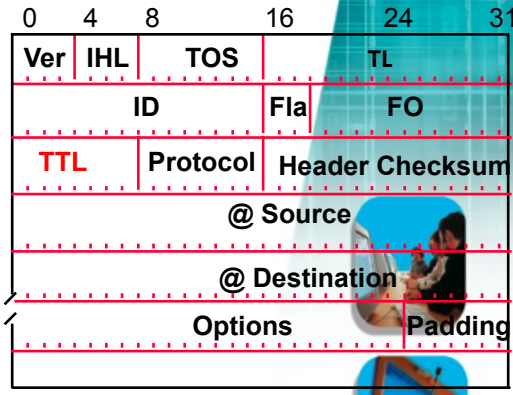
- Fragmentation is due to a lower MTU
- Only routers can do the fragmentation but.. only the destination can re-assembled the fragments
- The Identification fiels is always duplicated in the header of the fragments
- Flag (X/DF/MF):
  - 1 bit always 0
  - 1 bit : Don't Fragment (DF=1 : fragmentation does not allowed)
  - 1 bit : More Fragment (MF=0 : last fragment)
- FO : Fragment OffSet from the beginning
- The size of the fragment is computed (multiple of 8 bytes)
- Retransmission cause a lots of problems



Example : Fragmentation



TTL (8 bits)



Time To Live : counter value = 32 or 64

Limit the lifespan of a datagram in the network

When TTL reaches 0

- Packet is discarded by the router
- ICMP error datagram is sent back to the source

TTL counter is decreased :

- By every router on the route to the destination
- In theory, each second by the destination in case of reassembly (to give an upper bound on the reassembly time)

Transport Protocol field (8 bits, rfc 1700)

For decapsulation process at the destination

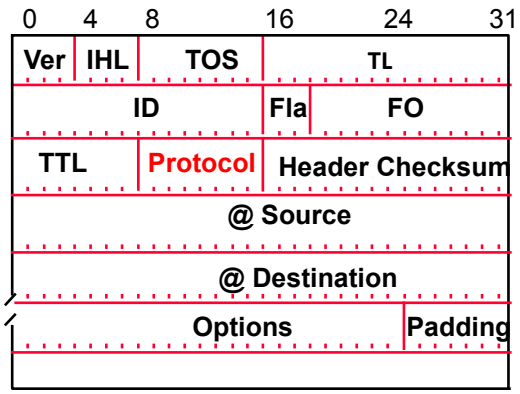
17 → UDP

6 → TCP

1 → ICMP

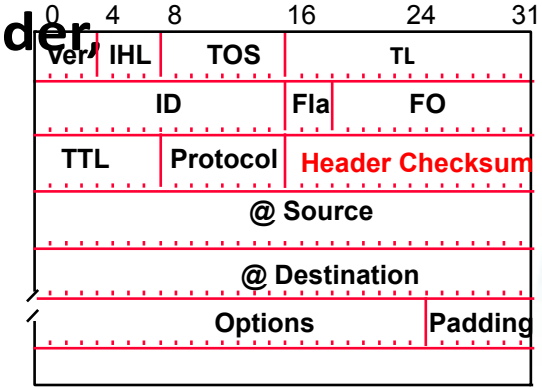
8 → EGP

89 → OSPF



HCS – Header Checksum (rfc 1071, 1141)

Error-checking only on the packet header, not on the data !
→ Motivations?



Main idea of the HCS algorithm :

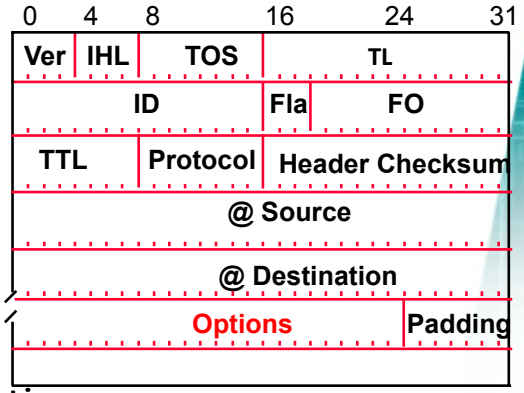
- Initially, HCS = 0000000000000000
- The HCS is the 16-bit one's complement of the one's complement sum of all 16-bit words in the header (mean : 16-bit sum, then 1's complement 16-bit sum, then 1's complement of 1's complement 16-bit sum)

Routers and destination :

- Compute the HC and comparate it to the HCS value in the header



Options (1 byte)



Options can be shared with upper layers

1er bit = 1 :

- In case of fragmentation, the option is duplicated in all the fragments header
- Else, only the first datagram carries the option

The next 2 bits give information about the type of option:

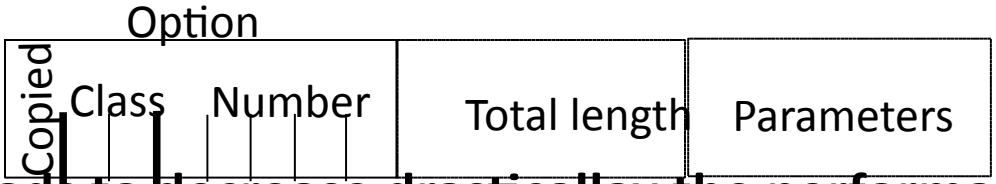
- 00 : control
- 01 : reserved for future use
- 10 : for metrology and bug detection
- 11 : reserved for future use

Options (...)

If the option required parameters, a length field (1 byte) is added :
it gives information about the total length of the option (words of 32 bits)

RFC 1700 give the complete list of options for IPv4

- LSR (*Loose Source Route, 1-00-00011)* : allow source routing instead of dynamic routing
- SEC (*Security,* 1-00-00010) : security information for the datagram



→ The use of options leads to decrease drasticallay the performances (routers congestion ↗ , end-to-end delay ↗ , …)

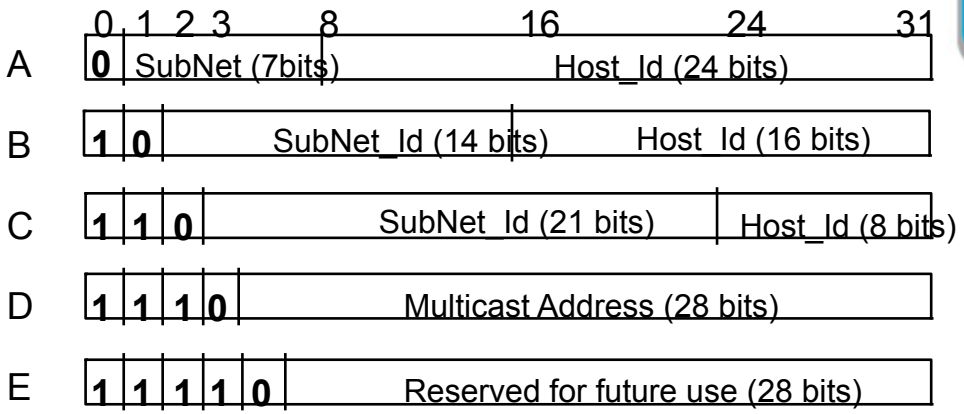
IP address

A global addressing scheme is required

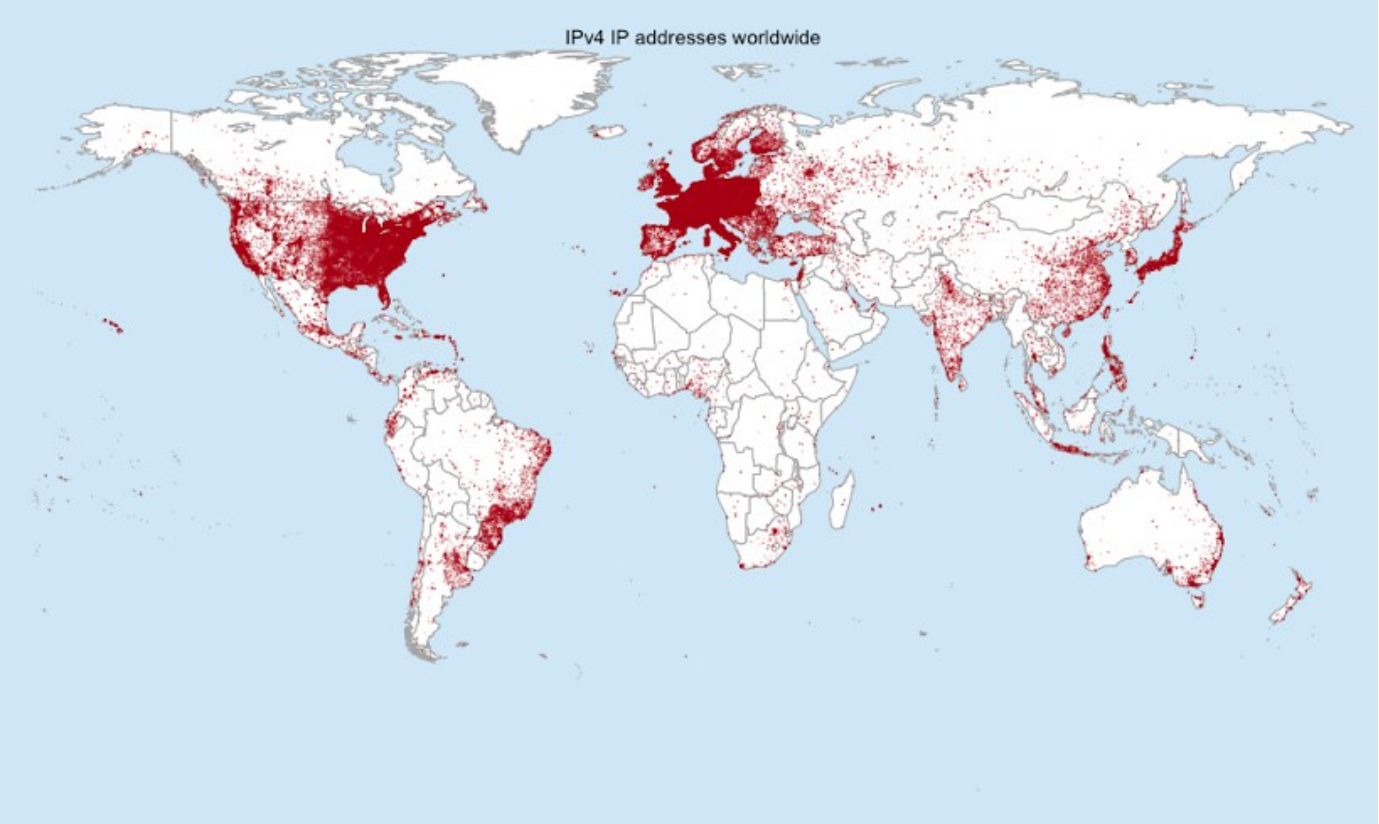
- To give a unique and non replicable identity to the host
- Physical location ↔ global IP address
- Which allow to build a route to the destination
- Distributed management of the adresses space
  (in France : Internic / NIC-France / CISM)

IP address (Internet Protocol) 4 bytes

- *e.g.* 134.214.104.178
- Basically, an IP address uses 3 fields
  Address Class, Network Address, Host Address



IPv4 address distribution



Source: somewhere on Twitter...



IPv4 Subnetting : Class A

7 bits for the subnet id.

1.0.0.0 to 126.0.0.0

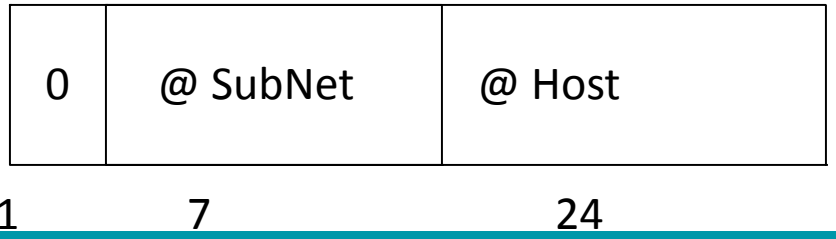
24 bits for the local addressing scheme (subnetting is allowed)

254^3 IP addresses are available (16 277 214)

→ Waste of the address space !

In France, no Class A network

Ex : 16.0.0.0 (DEC) 18.0.0.0 (MIT)



IPv4 Subnetting : Class B

16 bits for the subnet id.

128.1.0.0 à 191.255.0.0

16 bits for the local addressing scheme (subnetting is allowed)

254 x 254 IP addresses available / subnet (65534)

→ Waste of address space !

Example of Class B network (France)

134.214(Rocad, Lyon Tech Campus)

134.157(Jussieu, Paris VI)

No more address available !!

1	0	@ SubNet	@ Host
1	1	14	16

IPv4 Subnetting : Class C

24 bits for the subnet id.

192.0.1.0 à 223.255.255.0

8 bits for the local addressing scheme (subnetting is allowed)

Only 254 IP addresses available / subnet

1	1	0	@ SubNet	@ Host
1	1	1	21	8

IPv4 Subnetting : Class D

Multicast address (RFC 1700)

- Point-to-multipoint applications

Subnetwork from 224.0.0.0 to 231.255.255.255

- Ex : 224.4.4.4

Lack of structure

… because of dedicated use, particular applications, etc.



IPv4 Subnetting : Class E

Network Addresses from 239.a.b.c to 254.a.b.c

Reserved for future use

IP Addressing

Dotted Quad Notation

- 4 numerical values (byte) separated by a dot '.', *e.g.* 134.170.2.48

For each class, adress min, adress max

Reserved addresses

- Bits '0' only : subnet address, *e.g.* 134.214.0.0
- Bits '1' only : local broadcast, all the hosts of the subnet, *e.g.* 134.214.255.255

LoopBack addresses

- 127.0.0.0 : local subnet
- 127.0.0.1 : the host itself, *loopback localhost*

0.0.0.0 :

- Unknown host or, for routing table : default route

Some addresses are non routable on the Internet (rfc 1597)

- 10.*.*.* / 172.16-31.*.* / 192.168.*.* (private networks, restrained to Rocad only)
- 255.255.255.255 indicate all the hosts from all the subnet (*broadcast*)

IP Addressing (...)

Classless InterDomain Routing (CIDR, RFC 1519)

- The use of the 4 previous classes leads to : huge memory consumption in routers, sub-optimal use of the address space (address starvation even if addresses are available, see Class B)

CIDR means:

- Only 1 class to allow addresses aggregation in routing table

- Variable netmask 2 addressLength – maskLength

Example: What means /19 ?

→ 19 bits is dedicated for the subnet address

→ 32-19 bits is dedicated for the host address

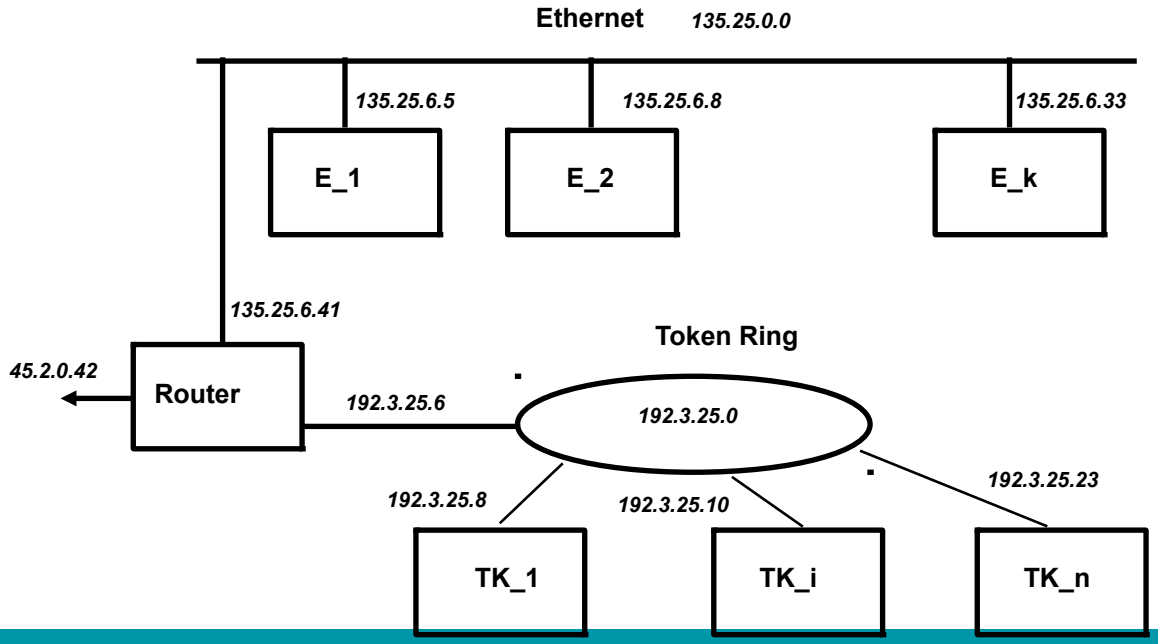
IP Addresses and Hosts

Host : classicaly only 1 interface (1 interface ↔ 1 IP address)

- Hosts on the same subnet have the *subnet mask*

Interconnexion host : router

- Several interfaces → several IP addresses
- Connected to several subnet throught the interfaces



Subnet mask ?

Remember that an IP address is divided into 2 parts :

- Subnetwork Address
- Host Address

The subnet mask is constructed as :

- All the bits of the subnet address are equal to '1'
- All the bits of the host adress are equal to '0'

→ *if* (**(@IP$_{host1}$ && netmask) == (@IP$_{host2}$ && netmask)**) *then* host1 and host2 are on the same subnetwork identified by (@IP$_{host}$ && netmask)



Example :

Subnetwork Address : 134.214.202.0/23
→ Network Mask : 255.255.254.0
Host 1 : 134.214.202.1 is in the subnet

IP Routing Protocol

Hop-by-hop routing (i.e. step by step to the destination)

Transmitting

- IP datagram is send to either the destination or to the next router

Receiving

- From local host or from the network interface

Routing table (neighborhood information)

- $@IP_{destination}$ (@host / @subnet)
- Next hop $@IP_{router}$
- Flags to identifiy : subnet Id. or Host Id.
- Physical interface to use for the datagram



IP Routing : Transmitting

If the destination host is in the same subnet (point-to-point or Ethernet or Wifi …)

- The IP datragram is directly send/deliver to the destination

Else

- The IP datagram is send to the router (according the routing table). This router will process the datagram.



IP Routing : Receiving

If the IP address destination is associated to one of the physical interfaces or if the destination address is a broadcast one or the loopback lo0

- Checksum Control ; fragmentation management if needed
- Decapsulation process to deliver the packet to the transport layer

Else, if the host is a router (or working as a router) :

- Datagram is routed / forwarded to the next router

Else

- Datagram is discarded (ICMP error)



IP Routing : Routing Table

@IP Destination *@machine ou @réseau*	**@IP d'un routeur De saut suivant** *Pour routage de Datagramme*	**Flags** *Précise si @ réseau/machine*	**Interface réseau** *Destination du Datagramme*
127.0.0.0	Néant		Lo0
192.168.0.0	Néant		Eth0
0.0.0.0	192.168.0.1		Eth0

- **There is no information about a complete route associated to the destination.**
- **There is only local information (local hosts, routers connected to the same subnet, interfaces information, subnet mask, …)**
- **The main information is the IP address of the router used to reach the destination**
- **The key idea is to assume that the next router is closer to the destination than the current host**

IP Routing Table: How it works ?

Note : in the routing table, information are classified from the more precise one (host address) to coarse location (subnetwork address), and finally the default route

1. Search an entry associated to the **@IP destination** (NetId/HostId). If found, the datagram is send to the right physical interface to reach either the router or the destination host

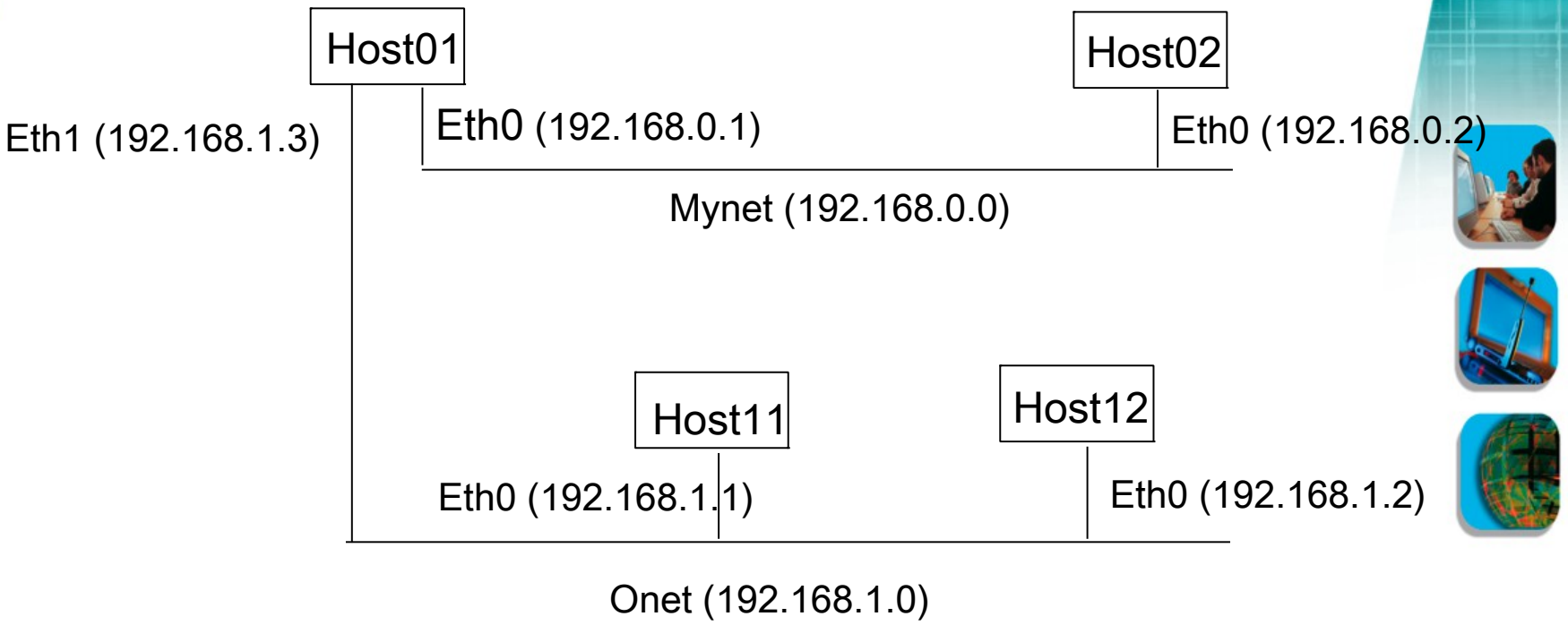
2. Search an entry associated to the **subnetwork address** (require to apply the netmask). If found, the datagram is send to the right physical interface.

3. Use the **default route** and the associated router.

4. If there is no possibility to route the packet according the routing table and the previous rules, the datagram is discarded, and an **ICMP error** packet (*no route to the destination*) is send back to the source.



IP Routing : an illustration



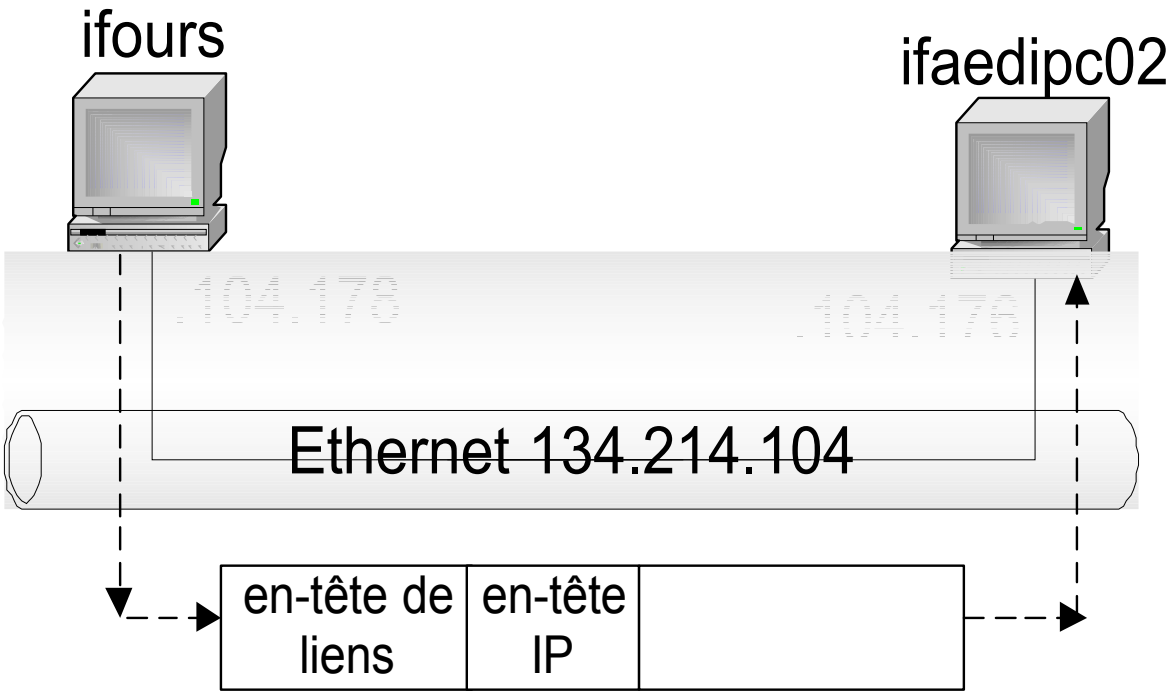
IP Routing : Routing table example

Routing table for the hosts : host01 host02 host11

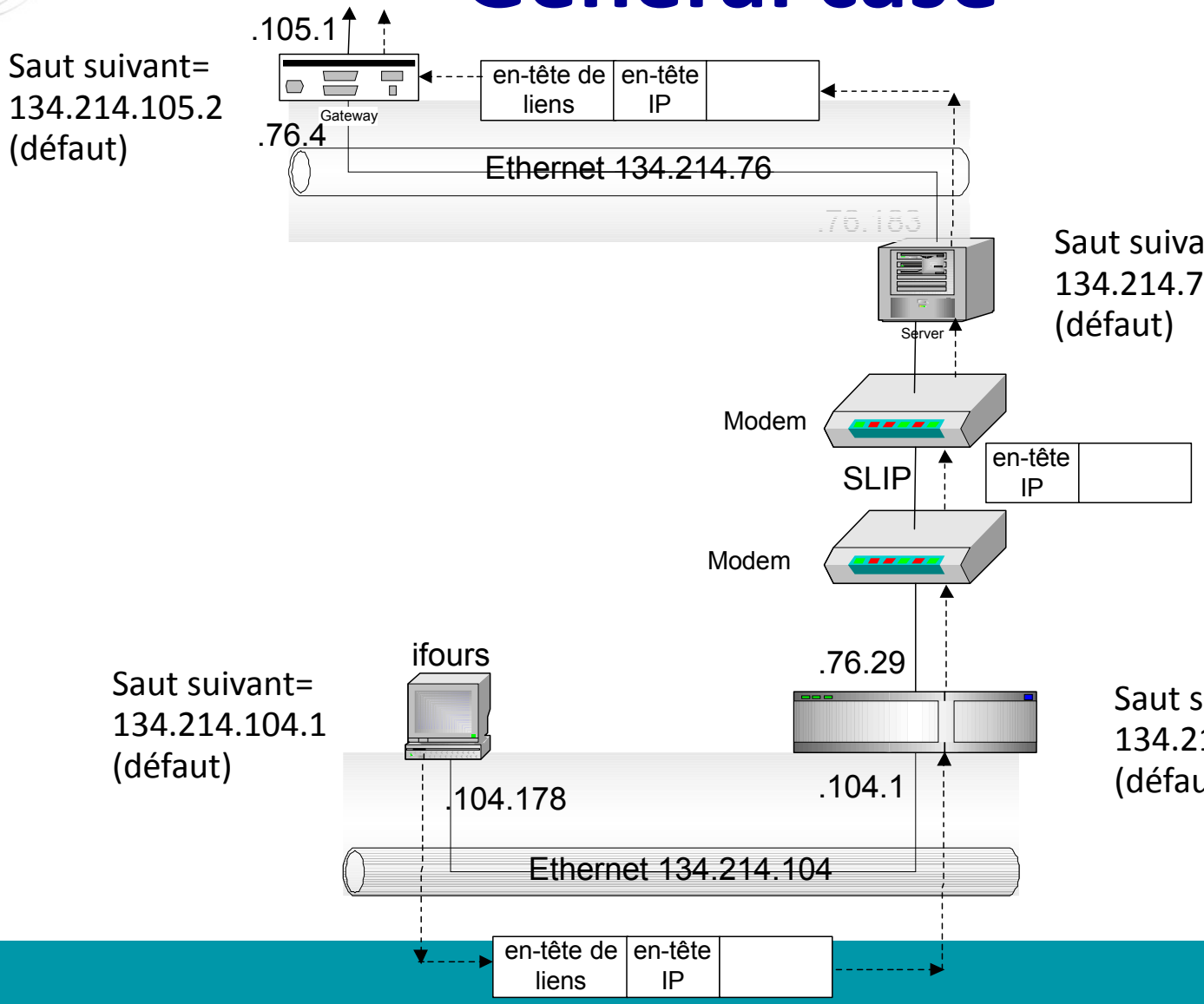
Adress	Netmask	Interface	Gateway
127.0.0.0	255.0.0.0	Lo0	-
192.168.0.0	255.255.255.0	Eth0	-
192.168.1.0	255.255.255.0	Eth1	-
127.0.0.0	255.0.0.0	Lo0	-
192.168.0.0	255.255.255.0	Eth0	-
0.0.0.0	0.0.0.0	Eth0	-
127.0.0.0	255.0.0.0	Lo0	-
192.168.0.0	255.255.255.0	Eth0	-
0.0.0.0	0.0.0.0	Eth0	-



Datagram Transmission : Simple case



Datagram Transmission : General case



IP configuration

Tools : **ipconfig** (NT), **ifconfig** (unix)

```
[root@citi-valois-1 /root]# ifconfig
eth0      Lien encap:Ethernet  HWaddr 00:01:02:20:73:2A
          inet adr:134.214.78.141  Bcast:134.214.79.255  Masque:255.255.252.0
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
          Paquets Reçus:286276 erreurs:0 jetés:0 débordements:0 trames:0
          Paquets transmis:19183 erreurs:0 jetés:0 débordements:0 carrier:0
          collisions:0 lg file transmission:100
          Interruption:10 Adresse de base:0xe400

lo        Lien encap:Boucle locale
          inet adr:127.0.0.1  Masque:255.0.0.0
          UP LOOPBACK RUNNING  MTU:3924  Metric:1
          Paquets Reçus:80 erreurs:0 jetés:0 débordements:0 trames:0
          Paquets transmis:80 erreurs:0 jetés:0 débordements:0 carrier:0
          collisions:0 lg file transmission:0
```



ping & traceroute

ping

Send successive ICMP packet *Echo Request*

Wait for ICMP packet *Echo Reply*

if there is no route to the destination : ICMP error generated by the last router

Use as a connectivity test, performance issues, host reachability, etc.

```
nstouls@balrog:~/$ ping www.google.fr
PING www-cctld.l.google.com (173.194.34.24): 56 data bytes
64 bytes from 173.194.34.24: icmp_seq=0 ttl=52 time=8.693 ms
64 bytes from 173.194.34.24: icmp_seq=1 ttl=52 time=7.456 ms
```

traceroute

Determine the route from the source to the destination, hop-by-hop delay, routers id., ...

Use the `ping` command (increasing the TTL value and DF flag in the IP header)

Traceroute (2010-11-04)

```
nstouls@balrog:~/$ traceroute www.mana.pf
traceroute to www.mana.pf (202.3.227.12), 64 hops max, 52 byte packets
1 psrl152.univ-lyon1.fr (134.214.152.1) 1.093 ms 0.960 ms 0.833 ms
2 cisrezo222-doua2.univ-lyon1.fr (134.214.200.230) 1.401 ms 1.294 ms 2.236 ms
3 rocad-sortie2-int.univ-lyon1.fr (134.214.201.190) 2.338 ms 2.680 ms 2.154 ms
4 193.55.215.6 (193.55.215.6) 2.012 ms 2.336 ms 2.292 ms
5 * * *
6 vl139-te0-3-0-0-lyon1-rtr-001.noc.renater.fr (193.51.189.13) 2.408 ms 2.718 ms 2.615 ms
7 xe-8-0-0.edge5.Paris1.Level3.net (212.73.207.173) 8.495 ms 8.335 ms 8.191 ms
8 ae-34-52.ebr2.Paris1.Level3.net (4.69.139.225) 8.826 ms 9.125 ms 8.814 ms
9 ae-47-47.ebr1.Frankfurt1.Level3.net (4.69.143.141) 18.616 ms 17.820 ms 18.516 ms
10 ae-91-91.csw4.Frankfurt1.Level3.net (4.69.140.14) 18.358 ms 18.303 ms 23.087 ms
11 ae-92-92.ebr2.Frankfurt1.Level3.net (4.69.140.29) 18.548 ms 17.639 ms 18.197 ms
12 ae-43-43.ebr2.Washington1.Level3.net (4.69.137.58) 107.390 ms 106.970 ms 108.994 ms
13 ae-82-82.csw3.Washington1.Level3.net (4.69.134.154) 116.483 ms 107.892 ms 107.703 ms
14 ae-74-74.ebr4.Washington1.Level3.net (4.69.134.181) 112.821 ms 112.719 ms 111.534 ms
15 ae-4-4.ebr3.LosAngeles1.Level3.net (4.69.132.81) 176.045 ms 175.662 ms 175.534 ms
16 ge-4-0-60.ipcolo2.LosAngeles1.Level3.net (4.69.144.47) 172.119 ms 169.805 ms 170.619 ms
17 unknown.Level3.net (63.215.86.130) 170.010 ms 171.588 ms 171.293 ms
18 rvs-rt001-so-0-0-0.globalconnex.net (80.255.35.230) 174.602 ms 175.204 ms 174.152 ms
19 rvs.rt004.vlan-3.globalconnex.net (80.255.37.116) 172.851 ms 174.162 ms 174.779 ms
20 41.194.22.12 (41.194.22.12) 711.552 ms 711.004 ms 711.035 ms
21 202.3.241.11 (202.3.241.11) 704.402 ms 702.555 ms 718.672 ms
22 202.3.227.12 (202.3.227.12) 719.806 ms 716.441 ms 716.664 ms
```



Other useful tools...

arp

Adresses translation (@MAC ↔ @IP)

```
nstouls@balrog:~/Documents/INSA/ISN/Reseaux$ arp 134.214.146.1
psrl146.univ-lyon1.fr (134.214.146.1) at ac:a0:16:a:b6:0 on en0
```

route

Routing table configuration and state

netstat

Current IP connexions

```
nstouls@balrog:~/$ netstat -rn -f inet
Destination        Gateway            Flags        Refs      Use   Netif
default            134.214.146.1      UGSc           11        0     en0
127.0.0.1          127.0.0.1          UH              0     6109     lo0
134.214.146.176    127.0.0.1          UHS             1        1     lo0
169.254.255.255    ac:a0:16:a:b6:0    UHLSW           0        0     en0
```

ICMP

* **Internet Control Message Protocol**
* **Signalling part of IP :**

  Host unreachable

  No route to the destination

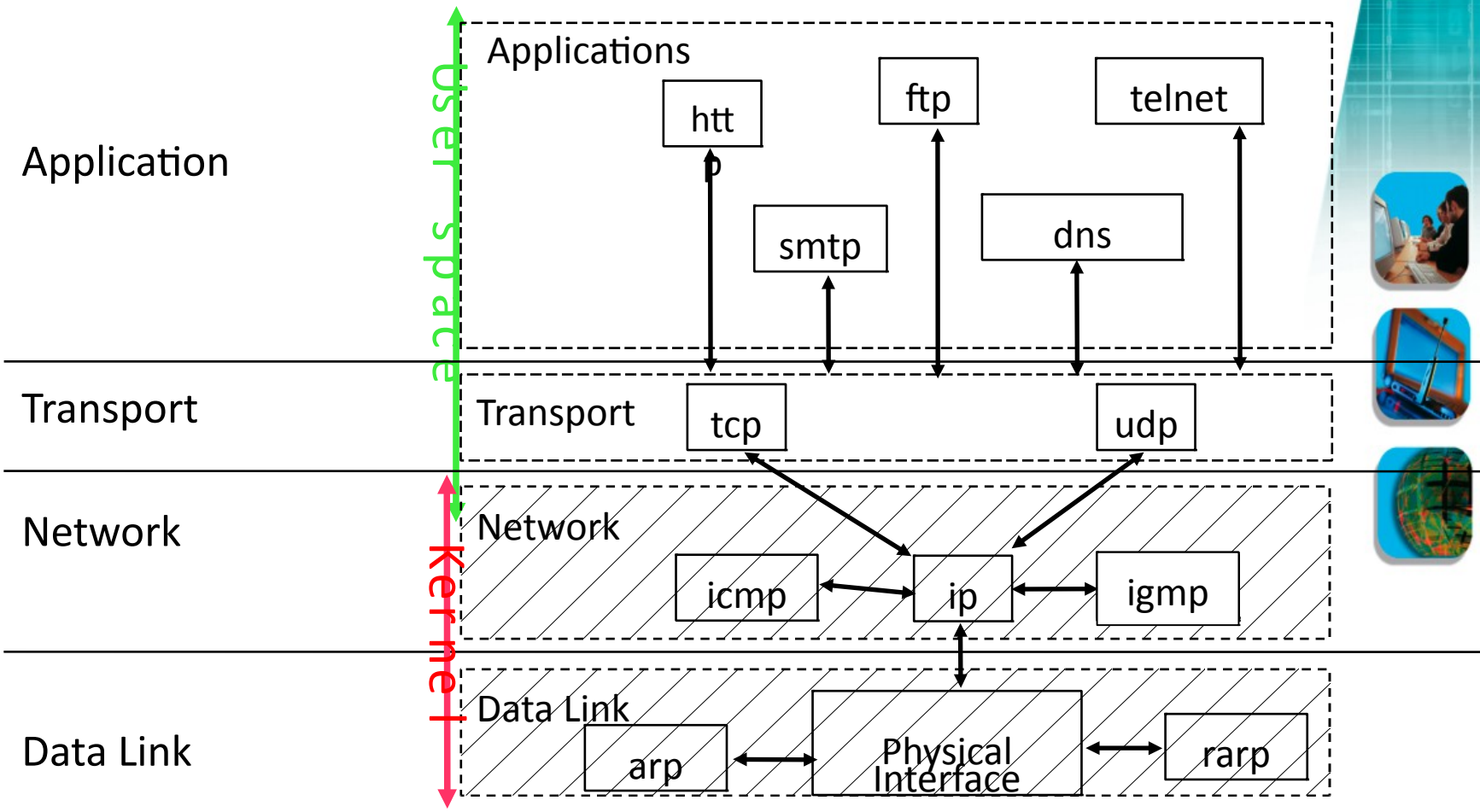
  TTL exceeds the limit value

  …

* **General header including options**
* **ICMP packets are mainly send back to the source**

  When a source receives an ICMP error, it can adapt its parameters

Where are we ?


Application
Transport
Network
Data Link
user space
kernel
Applications
htt
p
ftp
telnet
smtp
dns
Transport
tcp
udp
Network
icmp
ip
igmp
Data Link
arp
Physical
Interface
rarp