CS Communication Systems

CERN IT Department

# IPv6 101

Pre-GDB IPv6 workshop
CERN, 10th of June 2014
edoardo.martelli@cern.ch

CERN IT Department
CH-1211 Genève 23
Switzerland
**www.cern.ch/it**

# IPv6 and IPv4

# Addresses

# IP addresses

## IPv4

32 bits

Written as 4 groups of 8 bits, decimal notation:

**137.138.10.16**

(correspond to: 89.8A.0A.10 Hex)

## IPv6

128 bits

Written as 8 groups of 16 bits, hexadecimal notation:

**2001:0db8:a137:b138:c000:d000:e000:f001**

# Subnets

## IPv4

Netmask (0s in the host part):

**137.138.10.0 255.255.255.0**

Prefix length (number of bits used for the network address):

**137.138.10.0/24**

## IPv6

Only prefix length:

**2001:0db8:a137:b138::/64**

Host part is omitted

Smallest network: /64 (recommendation)

# Network and Host parts

```
1      16         32         48         64         80         92        108        128
1111:2222:3333:4444:5555:6666:7777:8888/64
Site prefix          Subnet                           Host
```

# Number of addresses

## IPv4

32 bits means $2^{32}$ ~= 4 billions

## IPv6

128 bits means $2^{128}$ ~= infinite

A normal allocation for a site/company (/32) gives:

- **$2^{32}$ subnets** (the whole IPv4 space)
- **$2^{64}$ host addresses per subnet** (25000 hosts per square meter on earth, per subnet)

# IPv6 notation

## IPv6

Leading 0s can be omitted:

**2001:<u>0</u>db8:a100:<u>000</u>1:<u>00</u>20:<u>0</u>300:<u>0000</u>:4000**

can also be written:

**2001:db8:a100:1:20:300:0:4000**

Groups of four 0s can be omitted and replaced by **::** (only once):

**2001:0db8:a137:<u>0000</u>:<u>0000</u>:abcd:<u>0000</u>:1234**

can also be written:

**2001:0db8:a137<u>::</u>abcd:<u>0</u>:1234**

# Special addresses

| | IPv4 | IPv6 |
|---|---|---|
| Loopback | 127.0.0.1 | ::1 |
| Unspecified address | | :: |
| Link Local | | FE80::/10 |
| Unique Local | 10.0.0.0/8 (RFC1918) | FC00::/7 |
| Default route | 0.0.0.0/0 | ::/0 |
| Multicast | 224.0.0.0/4 | FF00::/8 |
| Documentation | | 2001:DB8::/32 |

[http://tools.ietf.org/html/rfc4291]

# **Broadcast** **vs** **Multicast**

IPv4 uses broadcast to reach all the nodes on a subnet:

**255.255.255.255**

Broadcast addresses no longer exist in IPv6, but special multicast addresses for groups of hosts. Some examples:

All Nodes Addresses:

**FF02::1** (link-local)

All Routers Addresses:

**FF02::2** (link-local)

**FF05::2** (site-local)

All DHCPv6 servers:

**FF02::1:2** (link-local)

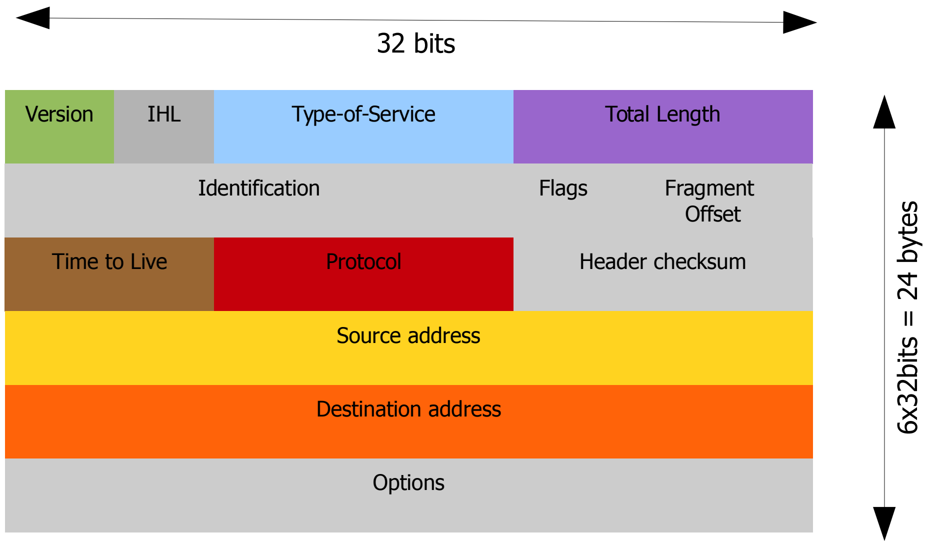
**FF05::1:3** (site-local)

[http://tools.ietf.org/html/rfc2375]

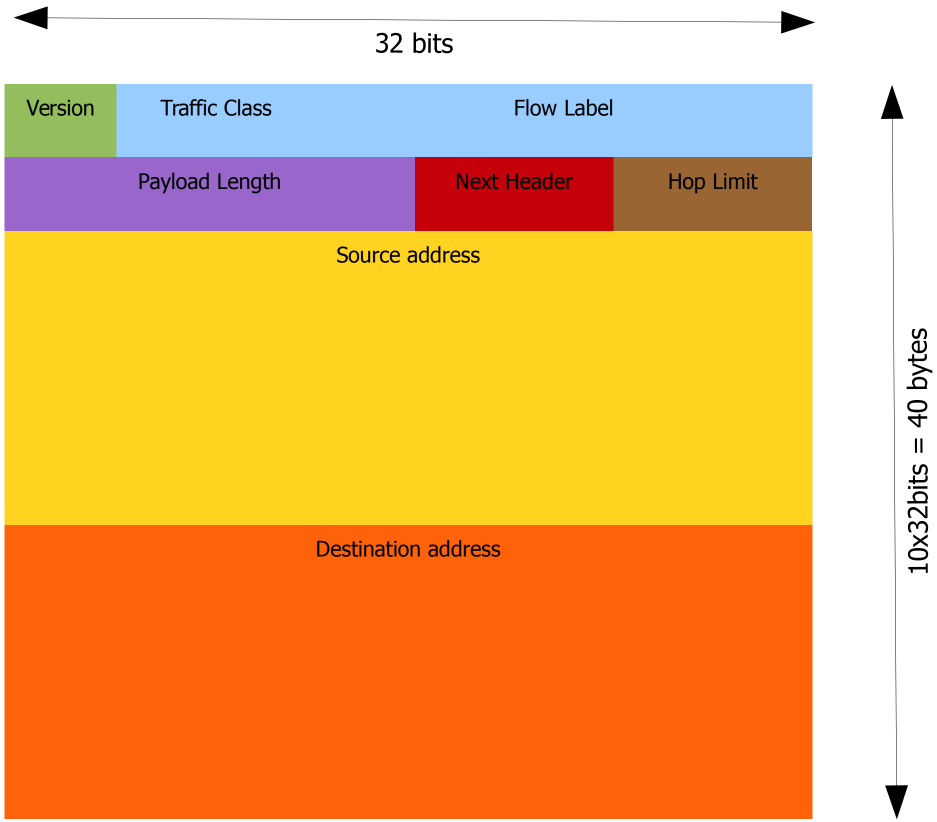
# Packets

# IP headers

## IPv4 header




## IPv6 header




[http://tools.ietf.org/html/rfc2460]

# Fragmentation

**IPv4**: When a packet is too big for the next link over which it is to travel, it can be fragmented by the sender (host or router).

**IPv6**: Fragmentation can only occur at the source node, and reassembly is only done at the destination node.

IPv6 routers never fragment IPv6 packets. Packets exceeding the size of the maximum transmission unit of the destination link are dropped and this condition is signaled by a Packet too Big ICMPv6 type 2 message to the originating node, similarly to the IPv4 method when the Don't Fragment bit set.

End nodes in IPv6 are expected to perform path MTU discovery to determine the maximum size of packets to send, and the upper-layer protocol is expected to limit the payload size. However, if the upper-layer protocol is unable to do so, the sending host may use the Fragment extension header in order to perform end-to-end fragmentation of IPv6 packets.
[http://en.wikipedia.org/wiki/IPv6_packet#Fragmentation]

# MTU

**IPv4**:
Minimum MTU = 576 Bytes
Maximum MTU = 65535 ($2^{16} – 1$) Bytes

**IPv6**:
Minimum MTU = 1280 Bytes
Maximum MTU = 4294967295 ($2^{32} – 1$) Bytes

# Protocols

# Neighbor discovery

IPv4: **ARP** Address Resolution Protocol

IPv6: **NDP** Neighbor Discovery Protocol

# NDP

**NDP** specifies 5 types of ICMP packets:

- **Router Advertisement (RA)**: periodic advertisement of the availability of a router
- **Router Solicitation (RS)**: the host needs RA immediately (at boot time)
- **Neighbor Solicitation (NS)**: to determine the link-layer address of a neighbor (equivalent to ARP request)
- **Neighbor Advertisement (NA)**: answer to a NS packet (equivalent to ARP reply)
- **Redirect**: Used by a router to inform a host of a better route to a given destination

[http://tools.ietf.org/html/rfc4861]

# Host Auto-configuration

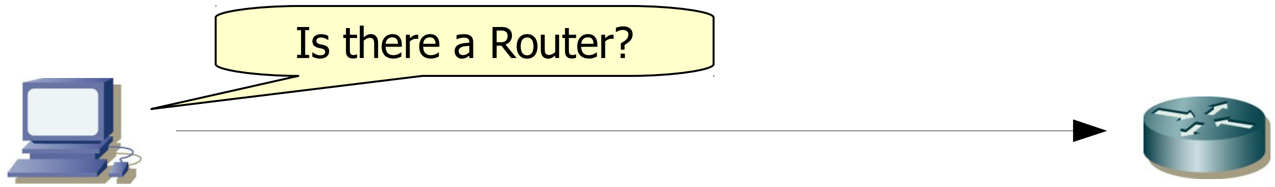
IPv4: **DHCP**

IPv6: **SLAAC** StateLess Address AutoConfiguration

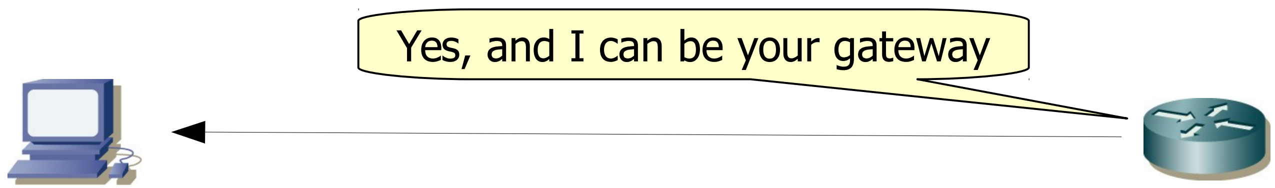
**DHCPv6**

# SLAAC

IPv6 hosts can configure themselves automatically when connected to a routed IPv6 network using ICMPv6 router discovery (RD) messages and EUI-64 for their own unique address.



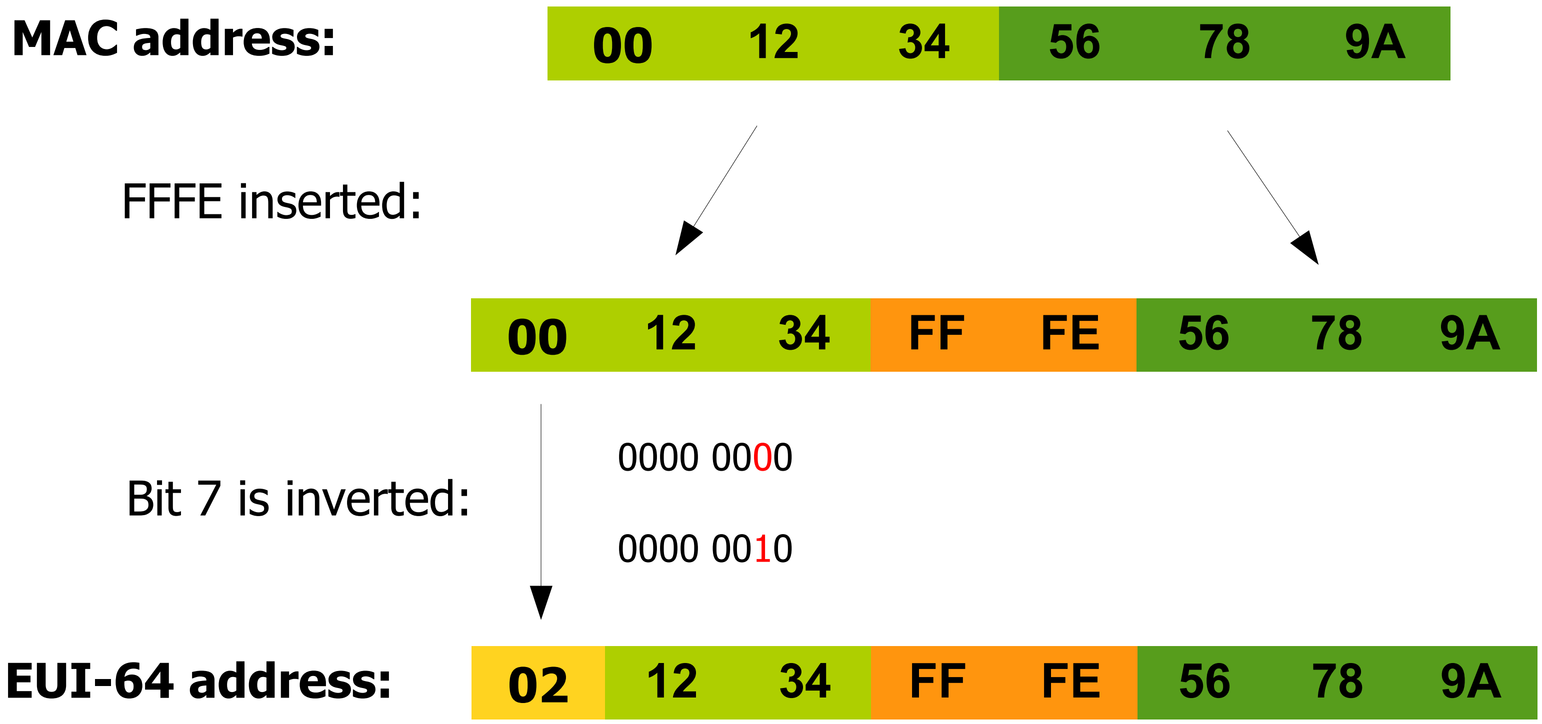

Routers respond to those requests with a router advertisement (RA) packet that contains network configuration parameters (subnet, default gateway).




[http://tools.ietf.org/html/rfc2462]

# EUI-64

EUI-64 is an identifier used to generate a unique host address from the MAC address




[http://tools.ietf.org/html/rfc3513]

# Routing

**RIP**(v2) IPv4 only

**RIPng** IPv6 only

**OSPF**(v2) IPv4 only

**OSPFv3** IPv6 only

**ISIS** IPv4 and IPv6

**Multiprotocol BGP** IPv4 and IPv6

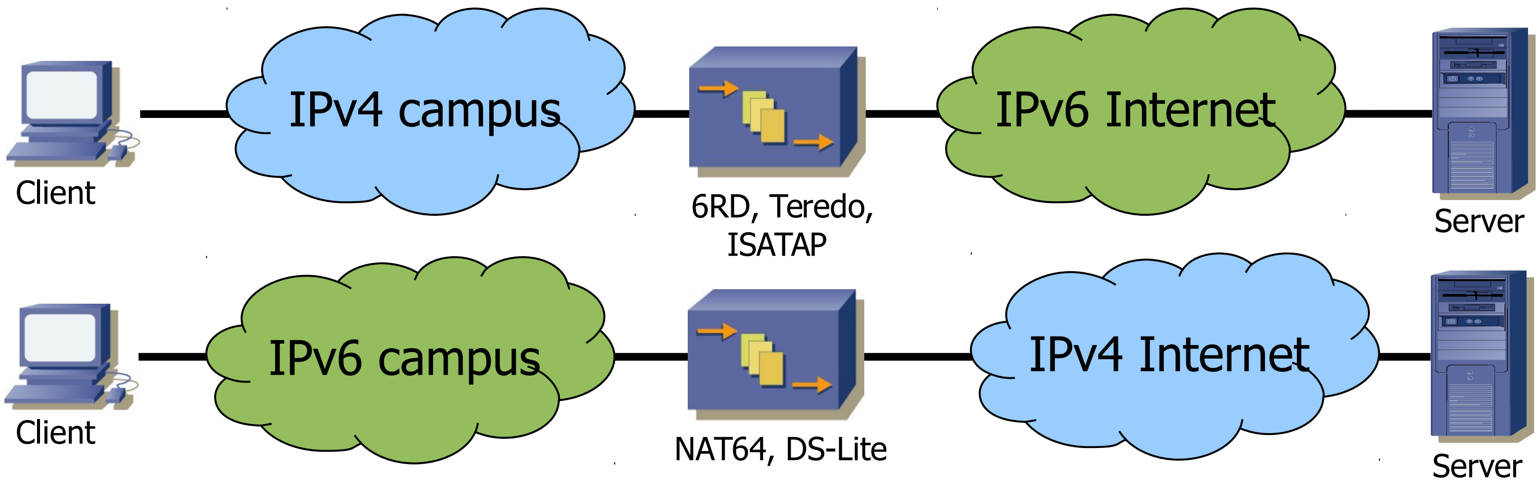
# Deployment

# Change your mindset

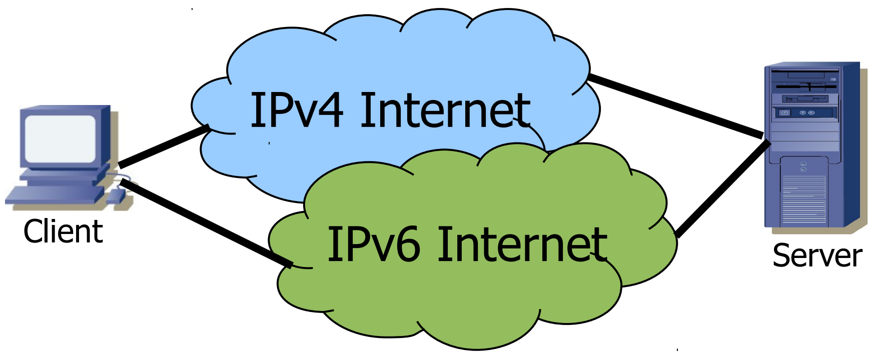
- No fear to waste
- Multiple addresses per interface, even in the same IPv6 subnet
- No NAT (not even designed)

# Transition strategies

**Bridging:**




**Dual Stack:**



# Transition

**Bridging**

- doesn't scale
- no end-to-end connectivity
- all typical issues of NAT
- may be good for an easy start

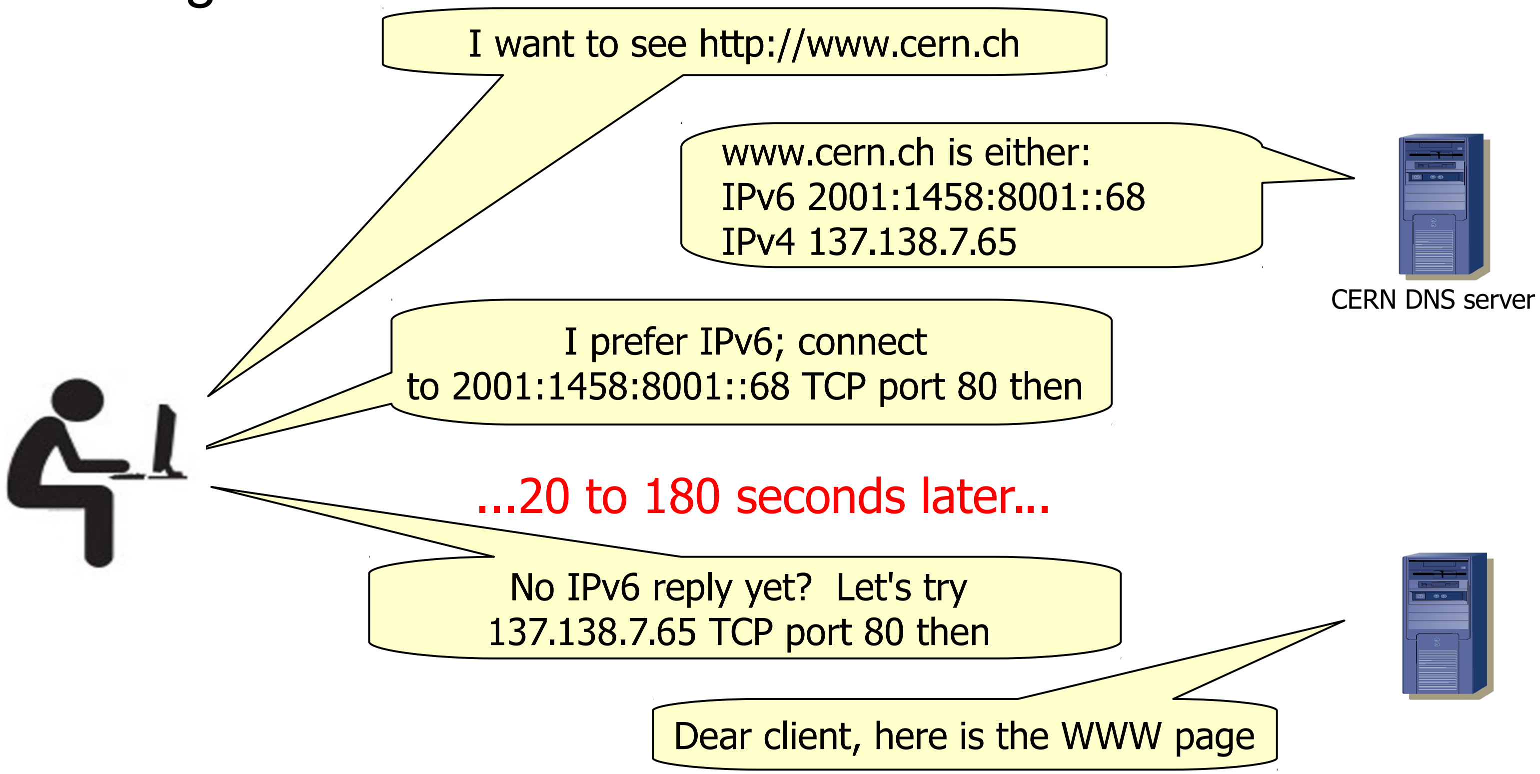
**Dual-Stack**

- two independent networks
- future proof

# Risks

# Multiple services

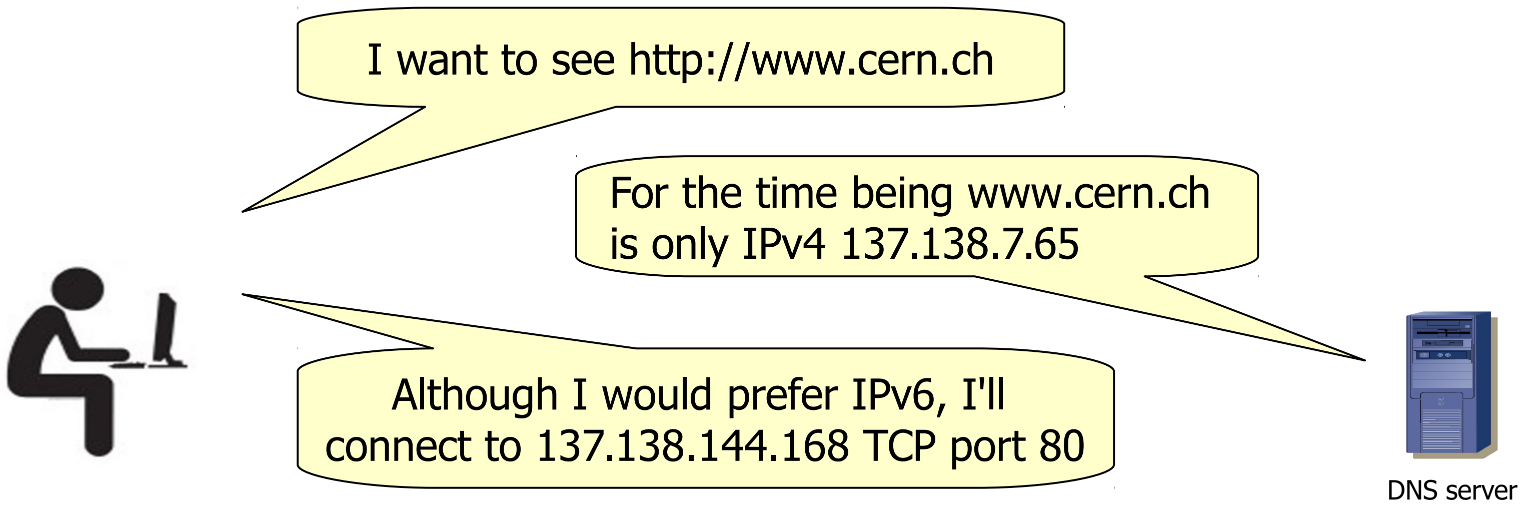
The choice of the IP protocol to be used is up to the client application, based on the DNS reply and its own settings.




**Problem: DNS is not application aware!**

# Multiple services

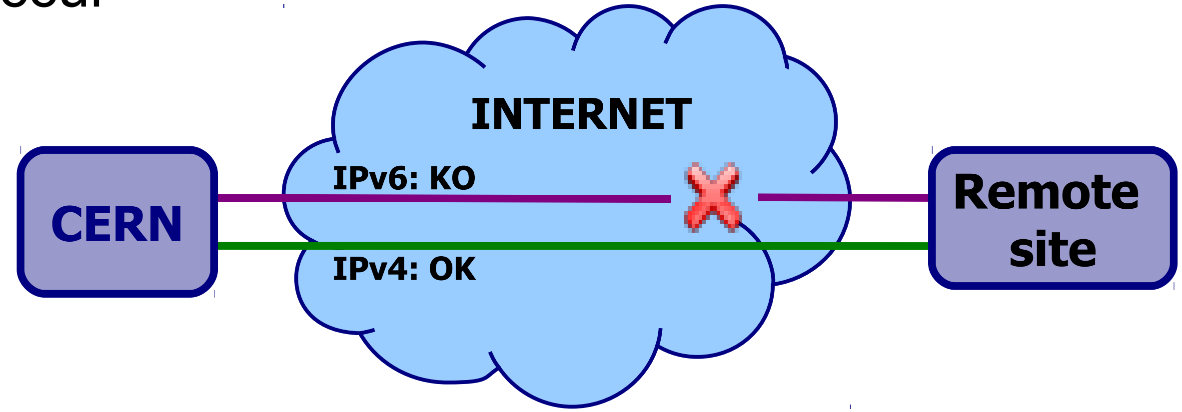
Servers cannot decide which IP protocol the client will use.
**IPv6 can be avoided by the DNS not returning the IPV6 address**



# Broken connectivity

If broken IPv6 connectivity, clients will wait up to 180secs before falling back to IPv4

If degraded IPv6 connectivity, fall back will never occur




**Client's perception: there's a server issue**

# Questions?