Internet Protocol - Addressing and forwarding

Claude Chaudet
Tel: 71 51
Claude.Chaudet@enst.fr

The core of the network layer: IP addressing

- IP addresses are the glue that make the global network stick together

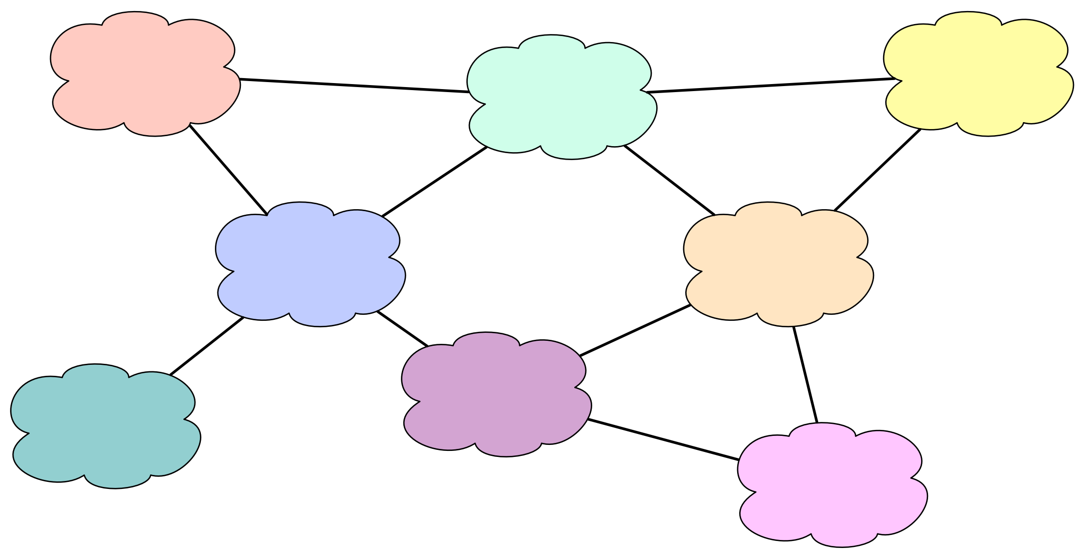
137.194.46.12

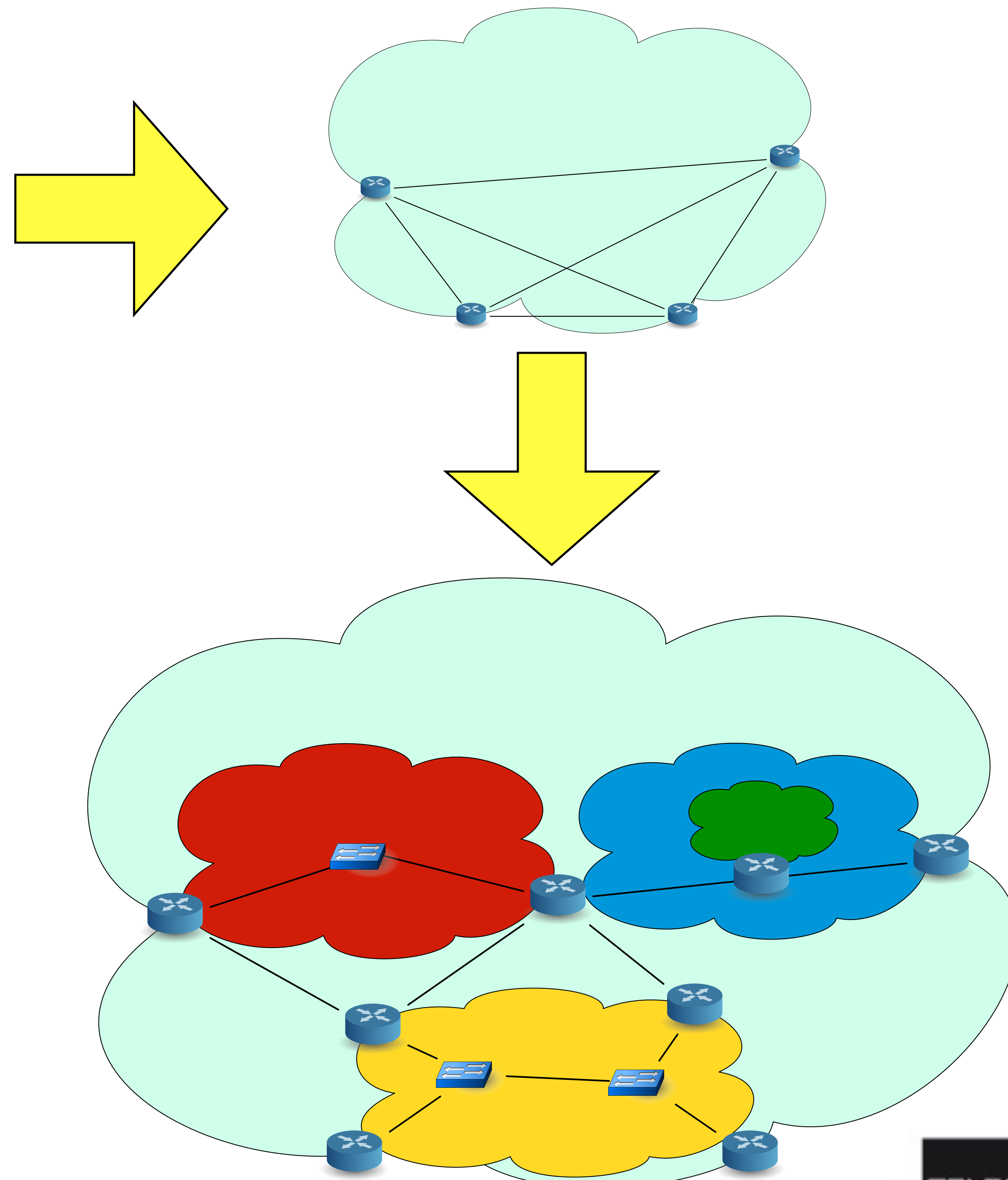
Prefix
Identifies the network

Interface Identifier (IID)
Identifies the device's network card

- IP addresses are the only information that is necessary for finding the path to a machine
  - All other network-level information is related to efficiency, fault reaction, ...

Internet seen as a set of *Autonomous Systems*





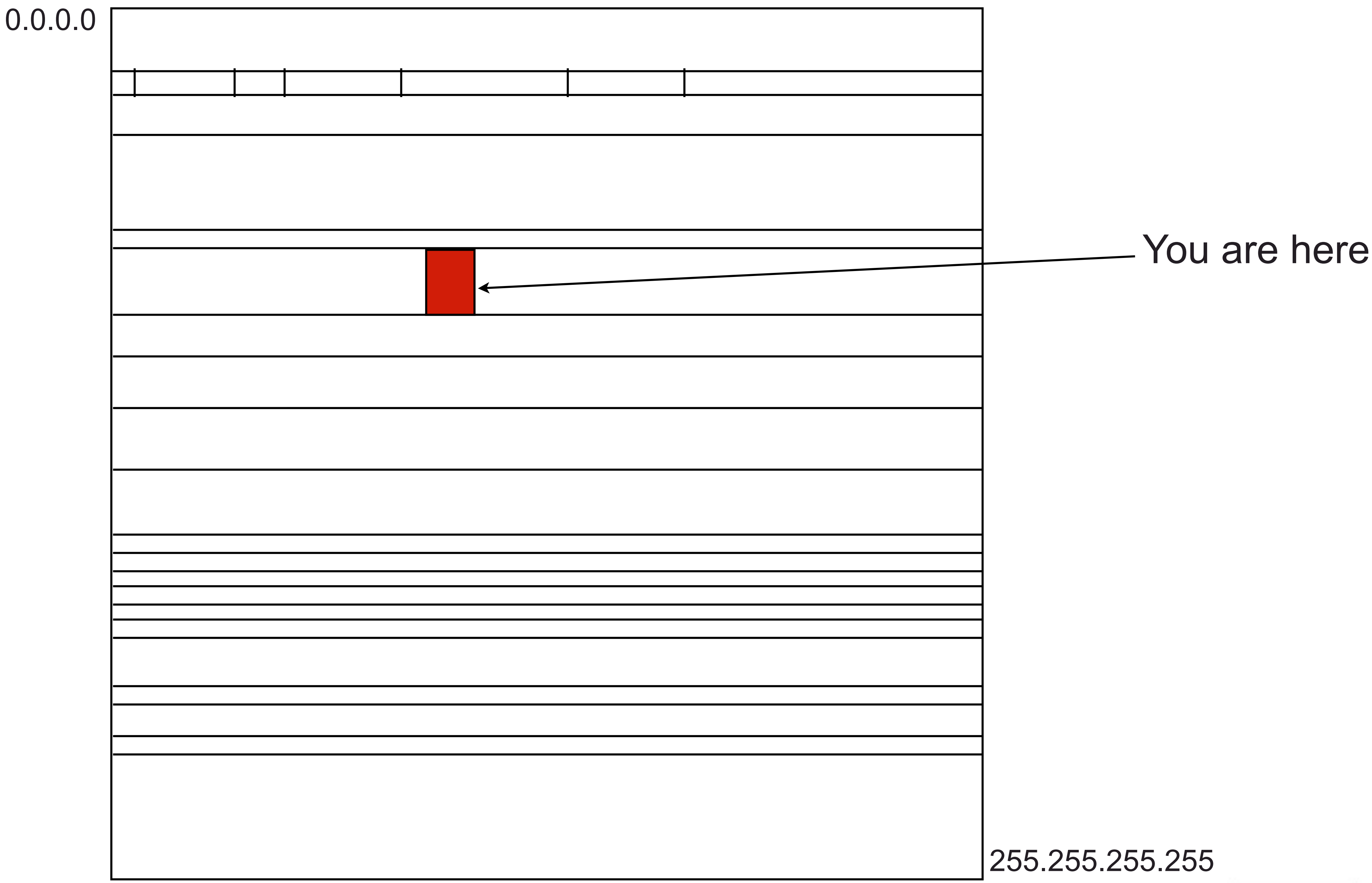
- Each autonomous system can be arbitrarily complex
  - Multiple IP slices

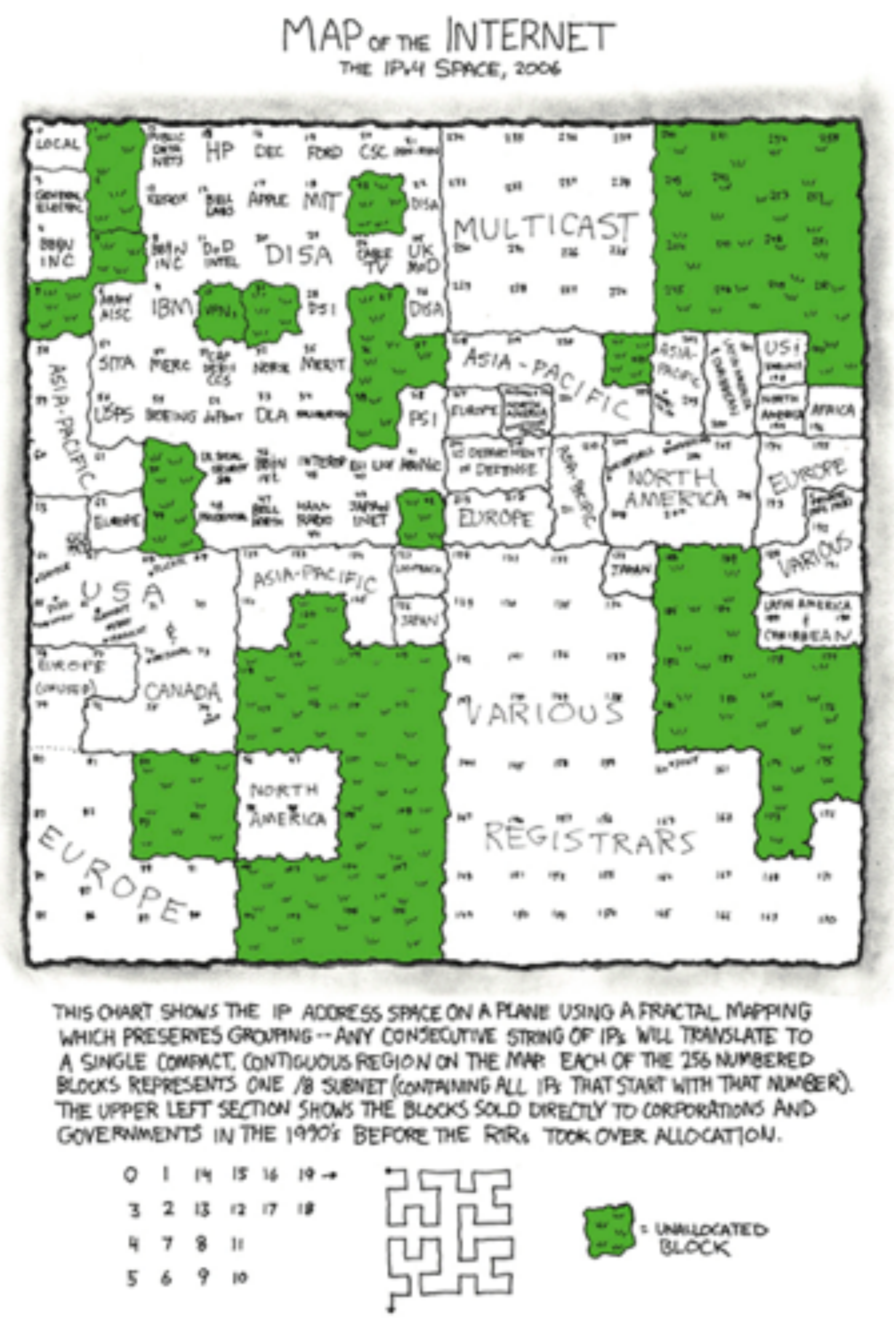
137.194.46.12

sub-network

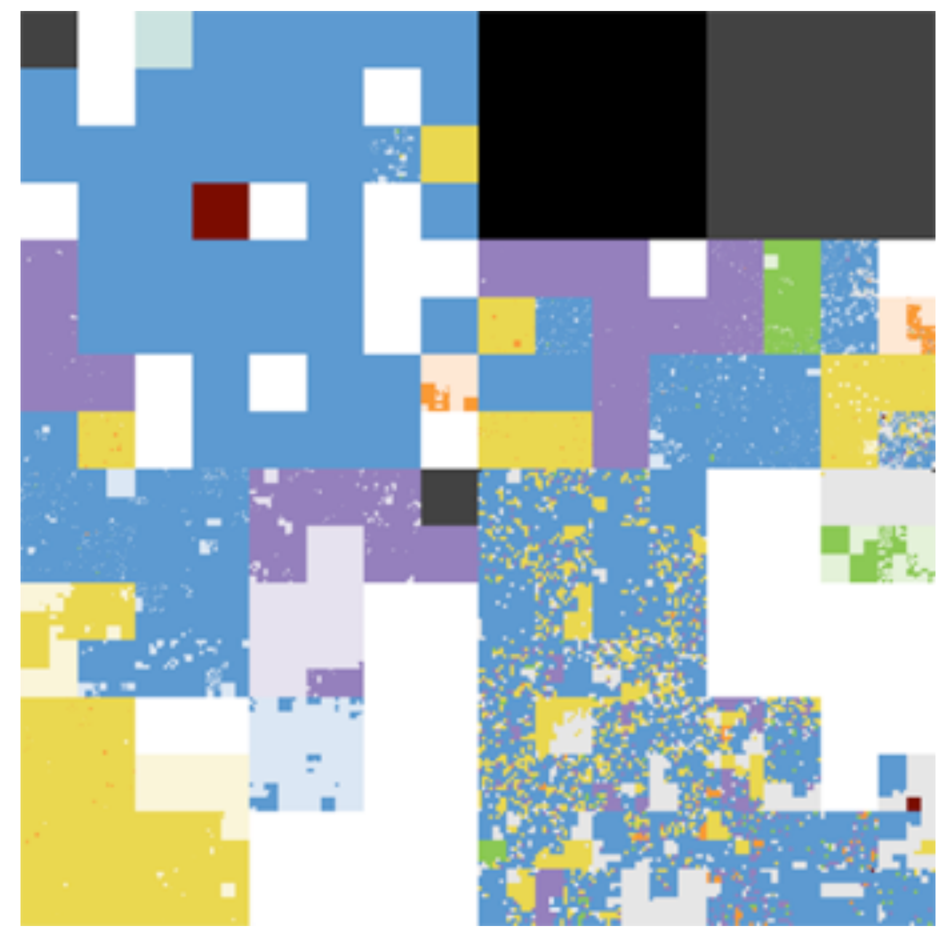
- A sub-network is invisible from the outside

IP addressing space division





Source: http://xkcd.com/195/



Source: ICANN Blog

Who decides of the global IP addressing plan?

- IP prefixes are managed by a central authority
  - IANA — Internet Assigned Numbers Authority
    http://www.iana.org/

- Slices of the global addressing are distributed to delegate regional organizations:
  - One Regional Internet Registry (RIR) per are of the world

- One organization (company, ISP), contacts its RIR to obtain an IP addressing space
  - Often one further delegation step (Local Internet Registry), usually to ISPs

Initial allocation policy: address classes

- Initially (i.e. early 1990's): addresses slices could be of three sizes:
  - Class A: prefix: 1 byte ; IID: 3 bytes (16M devices networks)
    most significant bit = 0
    00000001 → 01111110 ; 1.0.0.0 → 126.0.0.0
  - Class B: prefix: 2 bytes ; IID: 2 bytes (65535 devices networks)
    most significant bits = 10
    10000000.00000000 → 10111111.11111111 ; 128.0.0.0 → 191.255.0.0
  - Class C: prefix: 3 bytes ; IID: 1 bytes (256 devices networks)
    most significant bits = 110
  - 11000000.00000000 → 11011111.11111111 ; 192.0.0.0 → 223.255.255.0
  - Specific slices for reserved uses:
    - Loopback (127.0.0.0/8), private addresses (10.0.0.0/8, 172.16.0.0/12, 192.168.0.0/16), ...
- The problem: class B slices were soon depleted

TELECOM ParisTech

Today's slicing method: CIDR
Classless Interdomain Routing

- Prefix length varies bit per bit
  - More versatile approach, a network prefix is tailored to its actual size
  - CIDR notation:
    137.194.160.24 **/ 22**
    - 22 first bits identify the network
    - Last 10 bits identify the interface

- Alternative representation: network mask
  - 137.194.160.24 / 22
  - ⇔ address 137.194.160.24 ; mask 255.255.252.0

    [binary: 11111111. 11111111. 11111100. 00000000]

  - Network address= prefix shared between all hosts
    network address = any host's address & mask
    On the example: 137.194.160.0

bitwise "AND" operation

Dissecting an IP address

- Most relevant information on the network can be deduced from an host's IP address and the associated mask

- **`ipcalc 137.194.46.124/23`**

```
Address:    137.194.46.124          10001001.11000010.0010111 0.01111100
Netmask:    255.255.254.0 = 23      11111111.11111111.1111111 0.00000000   mask
Wildcard:   0.0.1.255               00000000.00000000.0000000 1.11111111
=>
Network:    137.194.46.0/23         10001001.11000010.0010111 0.00000000   Network prefix
HostMin:    137.194.46.1            10001001.11000010.0010111 0.00000001   Min. device address
HostMax:    137.194.47.254          10001001.11000010.0010111 1.11111110   Max. device address
Broadcast:  137.194.47.255          10001001.11000010.0010111 1.11111111   Broadcast address
Hosts/Net:  510                      Class B
```

Decimal representation

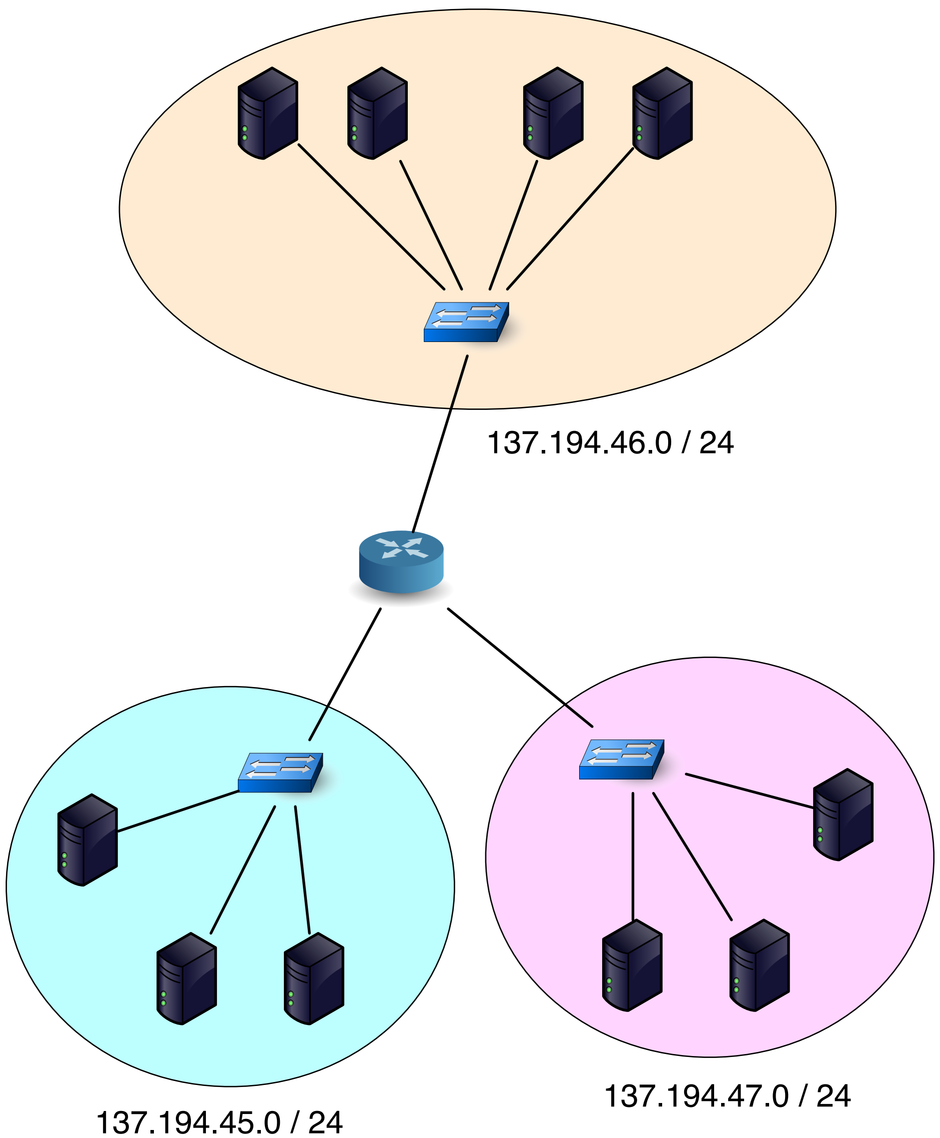
Binary representation

Group Communication - Broadcast

- Broadcast is the action to send a packet from one source to all destinations
  - All = limited to its own network

- IP defines a special address to this extent
  - Routers only forward broadcast frames if the sub-network corresponds to the broadcast address

  - Broadcast address: formed by taking the network address and by setting all IID bits to 1
    - Example: Network address — 137.194.46.0 / 23
    - Broadcast address — 137.194.47.255



Broadcast address	Networks concerned
137.194.46.255	Orange
137.194.47.255	Orange, magenta

Reserved addresses ranges

- Loopback addresses
  - 127.0.0.0 / 8
    - When a machine wants to talk with itself (test network applications)

- Private networks (routers block these addresses)
  - 10.0.0.0 / 8 (10.0.0.1 -> 10.255.255.254)
  - 172.16.0.0 / 12(172.16.0.1 -> 172.31.255.254)
  - 192.168.0.0 / 16 (192.168.0.1 -> 192.168.255.254)
  - 169.254.0.0 / 16 (169.254.0.1 -> 169.254.255.254)
    - Auto-assigned IP addresses (allows communication without DHCP server)

- Multicast addresses (1 to many, group-based, communication)
  - 224.0.0.0 -> 239.255.255.254

- "Test" networks
  - from 240.0.0.0 to 255.255.255.254

Internet Protocol evolution: IPv6

Exhaustion of the IPv4 addressing space

- Today's IPv4 statistics (Oct. 7, 2013)
  - http://www.potaroo.net/tools/ipv4/

- Projected exhaustion per region
  - Asia/Pacific: april 2011 (done)
  - Europe: september 2012 (done)
  - North America: January 2015
  - Latin America: april 2015
  - Africa: august 2022

IPv6: addresses format [RFC 3513]

- Addresses are 16 bytes long (128 bits)
  - 3.10^{38} possible addresses
  - 7.10^{23} per m^2 on earth...
  - More than 1000 unicast addresses per person

- Representation: groups of 2 bytes, written in hexadecimal
  - 8000:0000:0000:0000:0231:3245:AB6F:44FE

- Writing optimizations
  - 8000::0231:3245:AB6F:44FE
  - 8000::231:3245:AB6F:44FE

- Fixed format: prefix = 64 bits, IID = 64 bits
  - 8000:0000:0000:0000:: / 64

Addresses semantics

- Hierarchical notation, “fields” separation:
  2001:0660:330F:00A4:0230:65FF:FEBF:92A2
  - Address type (3 bits) : (001 for unicast)
  - TLA (13 bits) : Top Level Aggregator
  - NLA (32 bits) : Next Level Aggregator
    intermediate authority (ISP)
  - SLA (16 bits) : Site Level Aggregator
    sub-network (local administrator level)
  - IID (64 bits) : can be derived from the MAC address for instance

```
ifconfig en0
en0: flags=8863<UP,BROADCAST,SMART,RUNNING,SIMPLEX,MULTICAST> mtu 1500
        inet6 2001:660:330f:a4:20d:93ff:fe61:dc5e prefixlen 64 scopeid 0x4
        ether 00:0d:93:61:dc:5e
```

Multiple addresses for one interface

- One unicast address per network interface
  - Public address
  - 2000::/3

- One Link-local Unicast address:
  - Mandatory, unique on the link, self-generated, non-routable
  - FE80::/64 (FE80:0:0:0:*:*:*:*)

- One Unique Local Address (ULA) address [RFC 4193]:
  - Private address, unique in the sub-network
  - FC00::/7 (FC*:*:*:*:*:*:*:*)

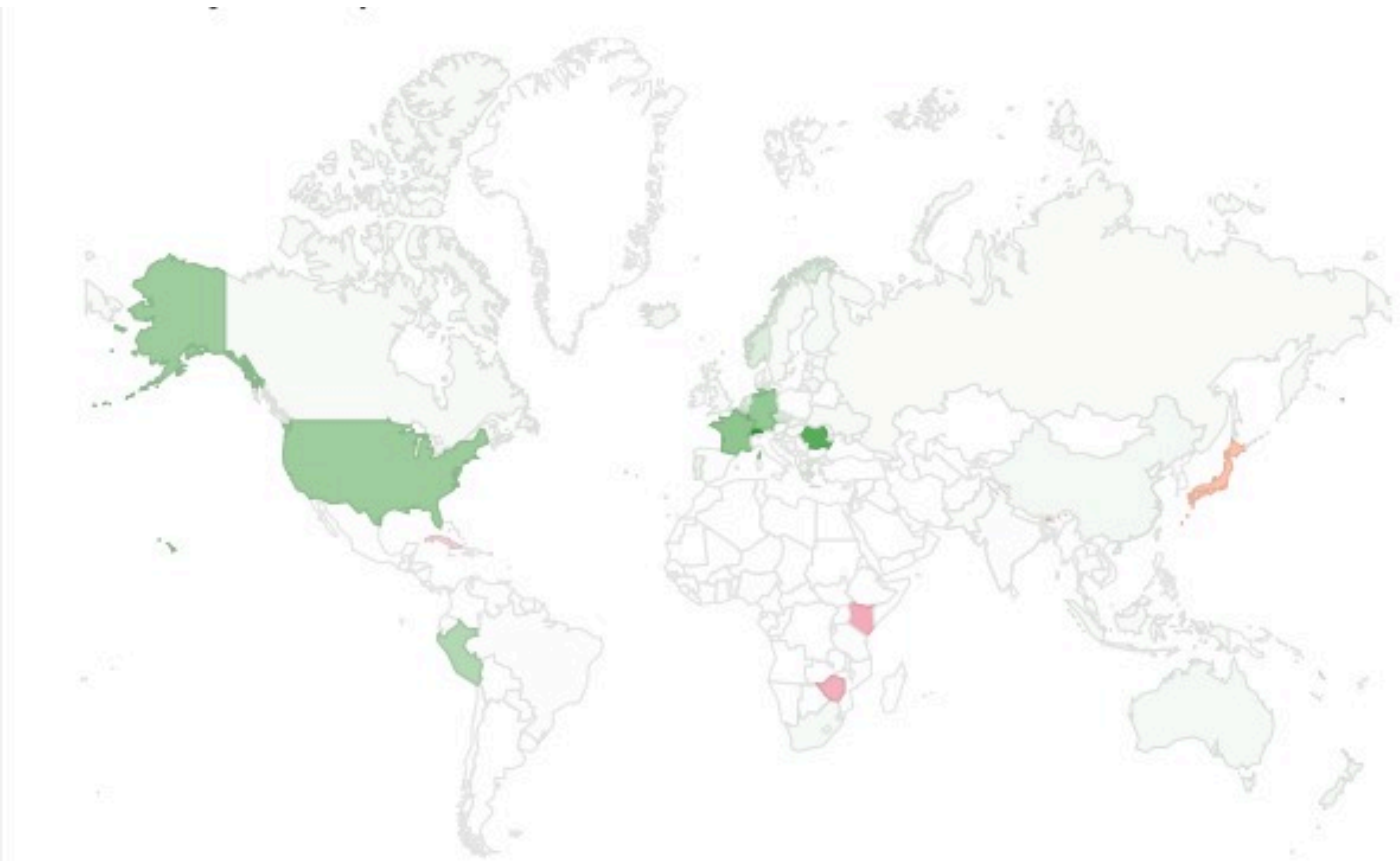
- Loopback (127.0.0.1)
  - ::1

IPv6 - Adoption rate

- Percentage of IPv6 requests received at google
  - http://www.google.com/intl/en/ipv6/statistics.html

IPv6 - per country

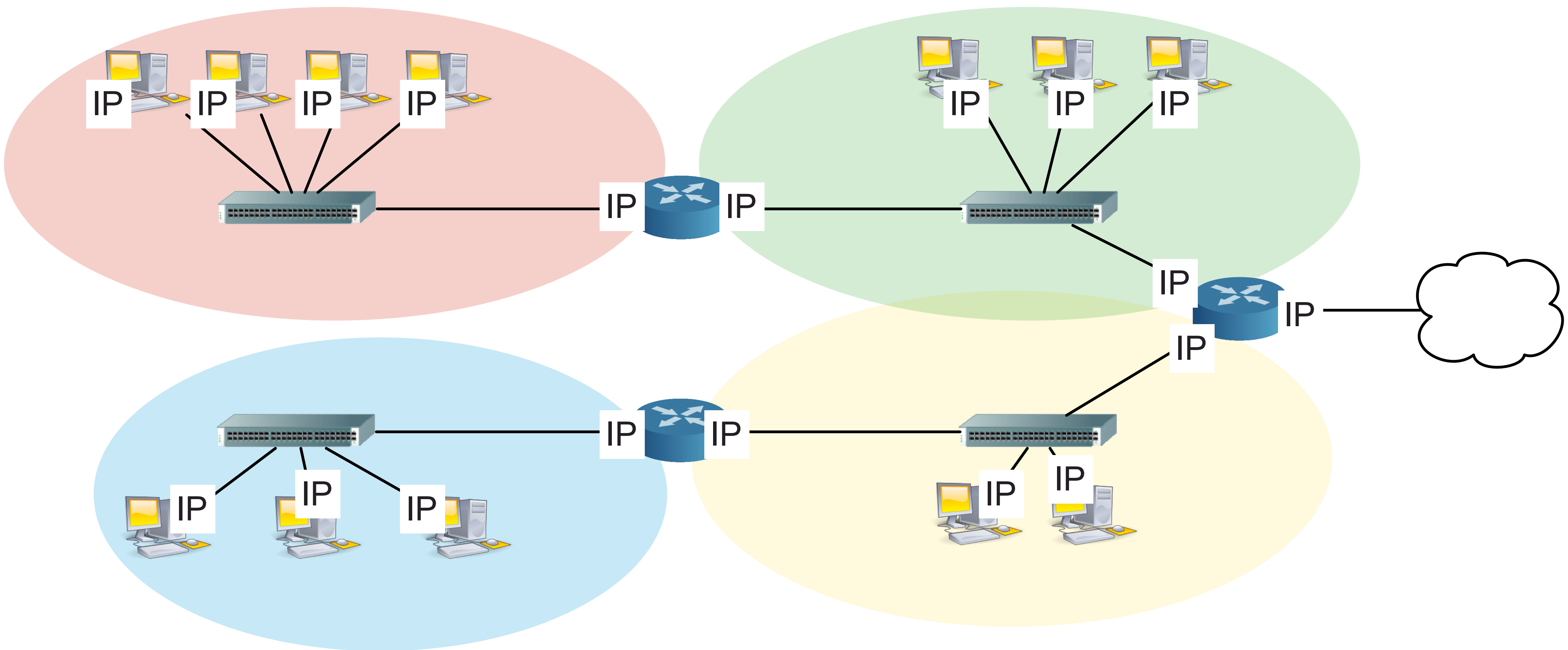
- Source: http://www.google.com/intl/en/ipv6/statistics.html



Networks and sub-networks

Inside a network: IP addresses allocation

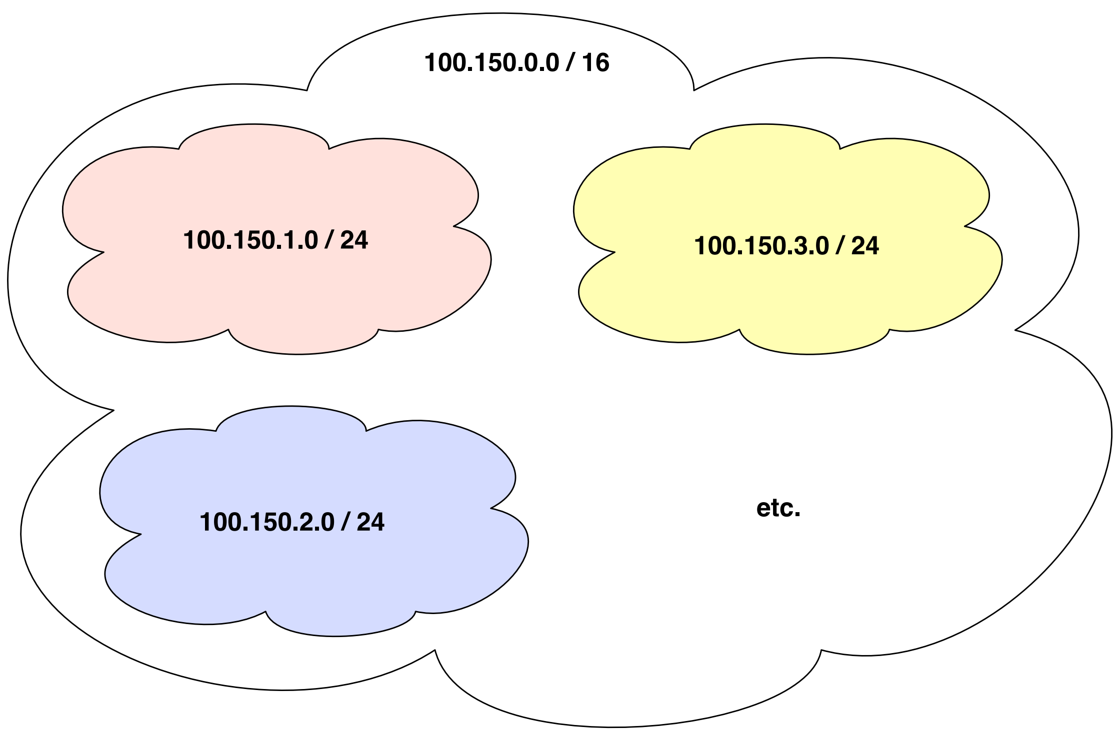
- Not all equipments need to have an IP address
  - Switches work at the MAC level, they do not need an IP address
  - Routers work at the IP level, they have one IP address per interface
- IP addresses of close machines should be coherent
  - Definition of sub-networks within a network

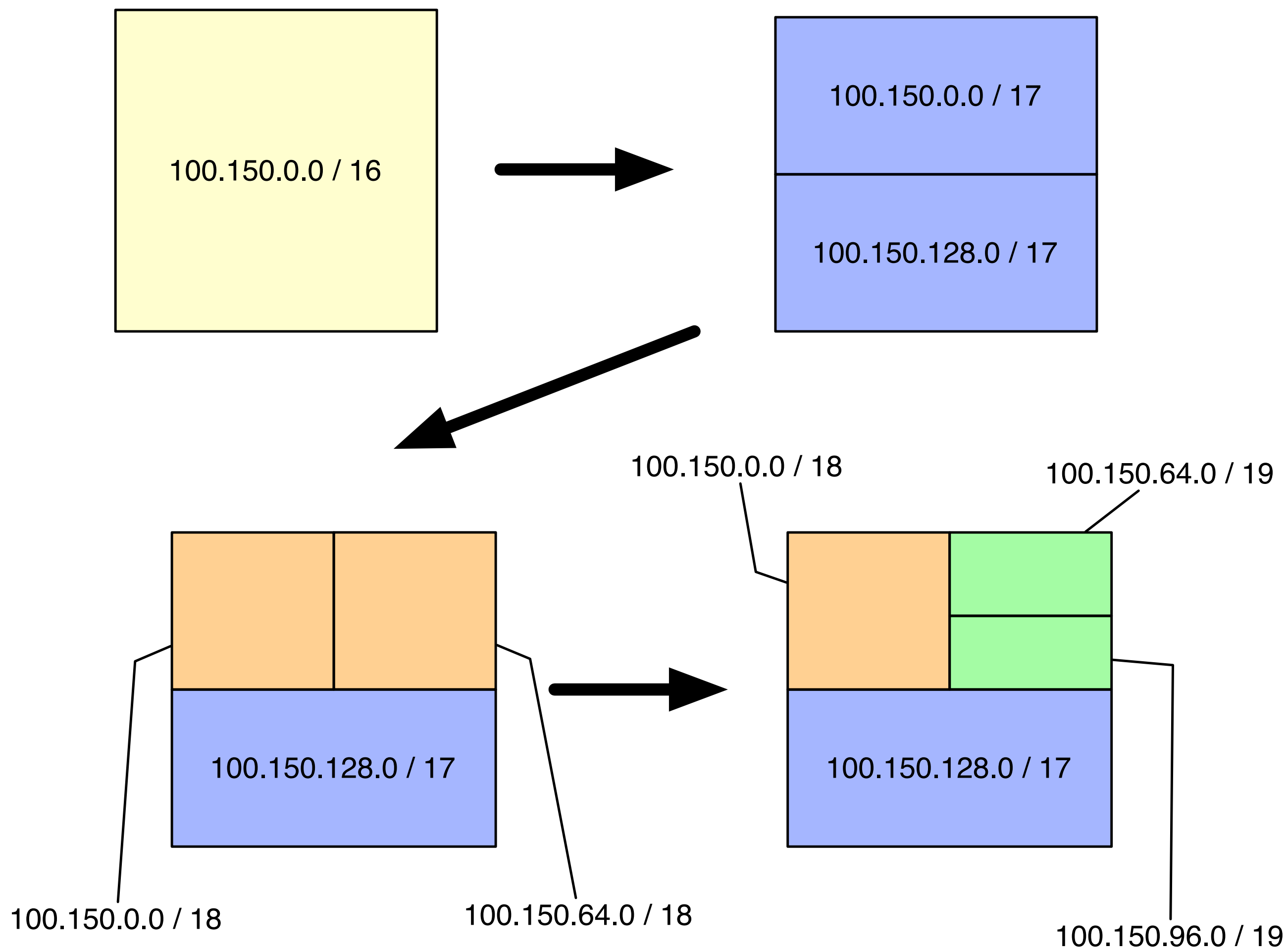


mardi 15 octobre 13

Dividing a network slice

- Once a network slice has been allocated, further division can be applied within the network (usually a good practice for large networks)
  - Separating a network in sub-networks
    - A /22 network can only be cut in two /23, or 4 /24, or one /23 and two /24
    - Etc. recursively

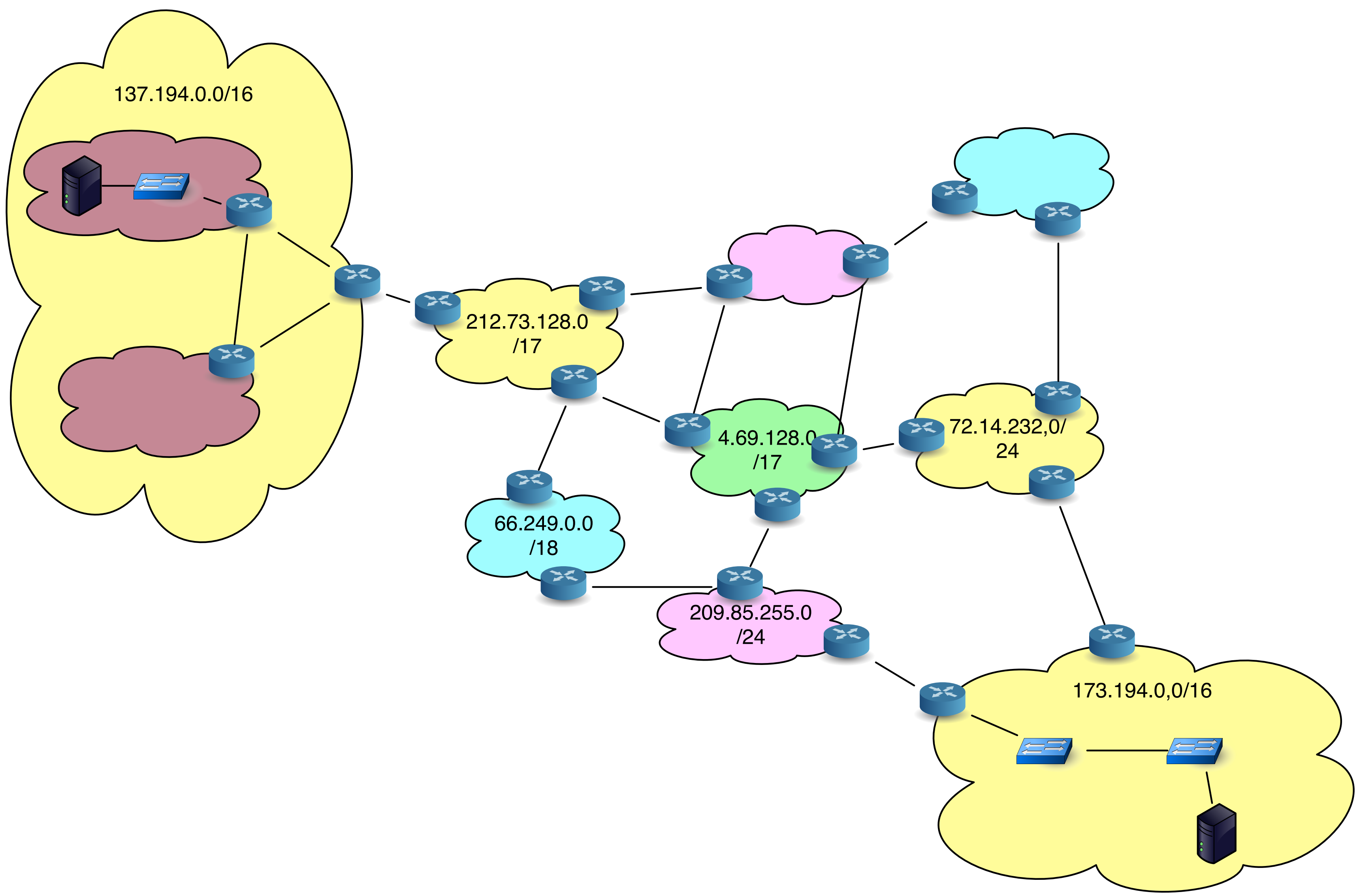




Internet Protocol (IP): forwarding process

IP level forwarding

- IP groups machines according to slices (group of machines that share the same prefix)
- Packets forwarding is realized hop by hop, domain per domain



TELECOM ParisTech

First hop- reaching the first router

- Linux:
  - `ifconfig eth0 192.168.1.2 netmask 255.255.255.0`
  - `route add default gw 192.168.1.1`

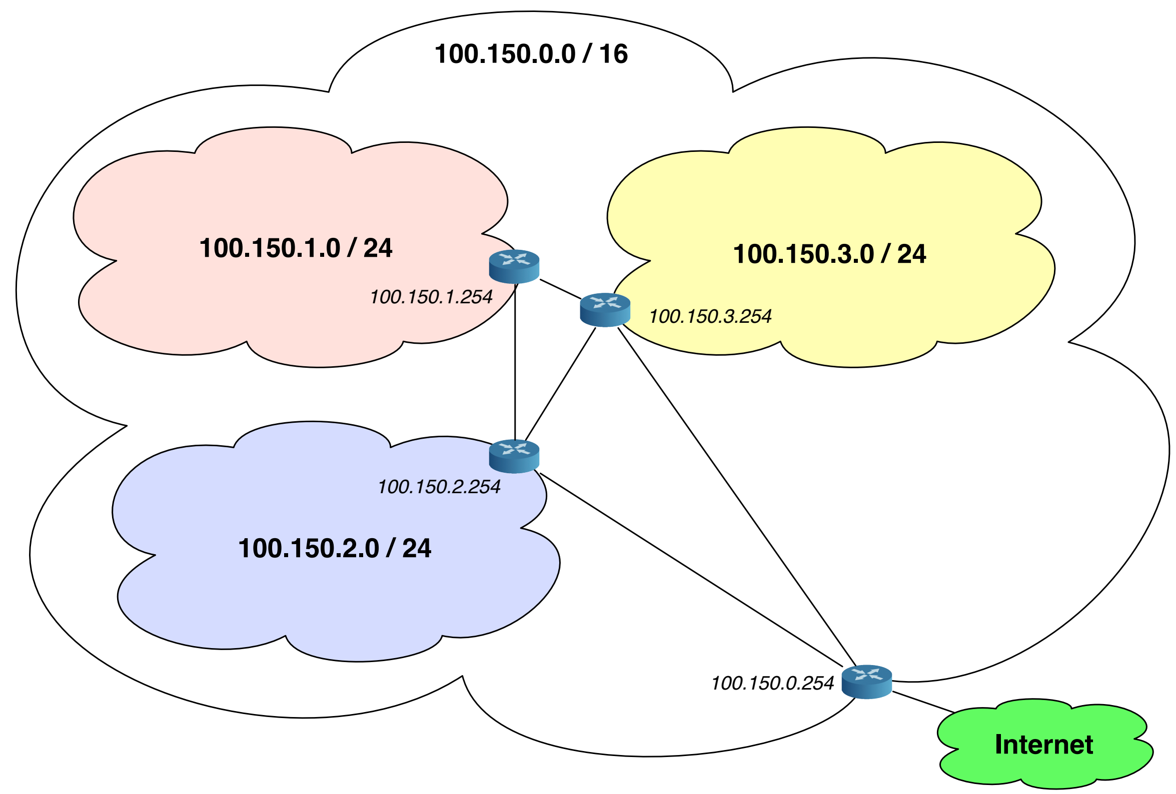
- Windows:

Mac OS X:

- Be careful...
  - The gateway (default router) should be on the same sub-network

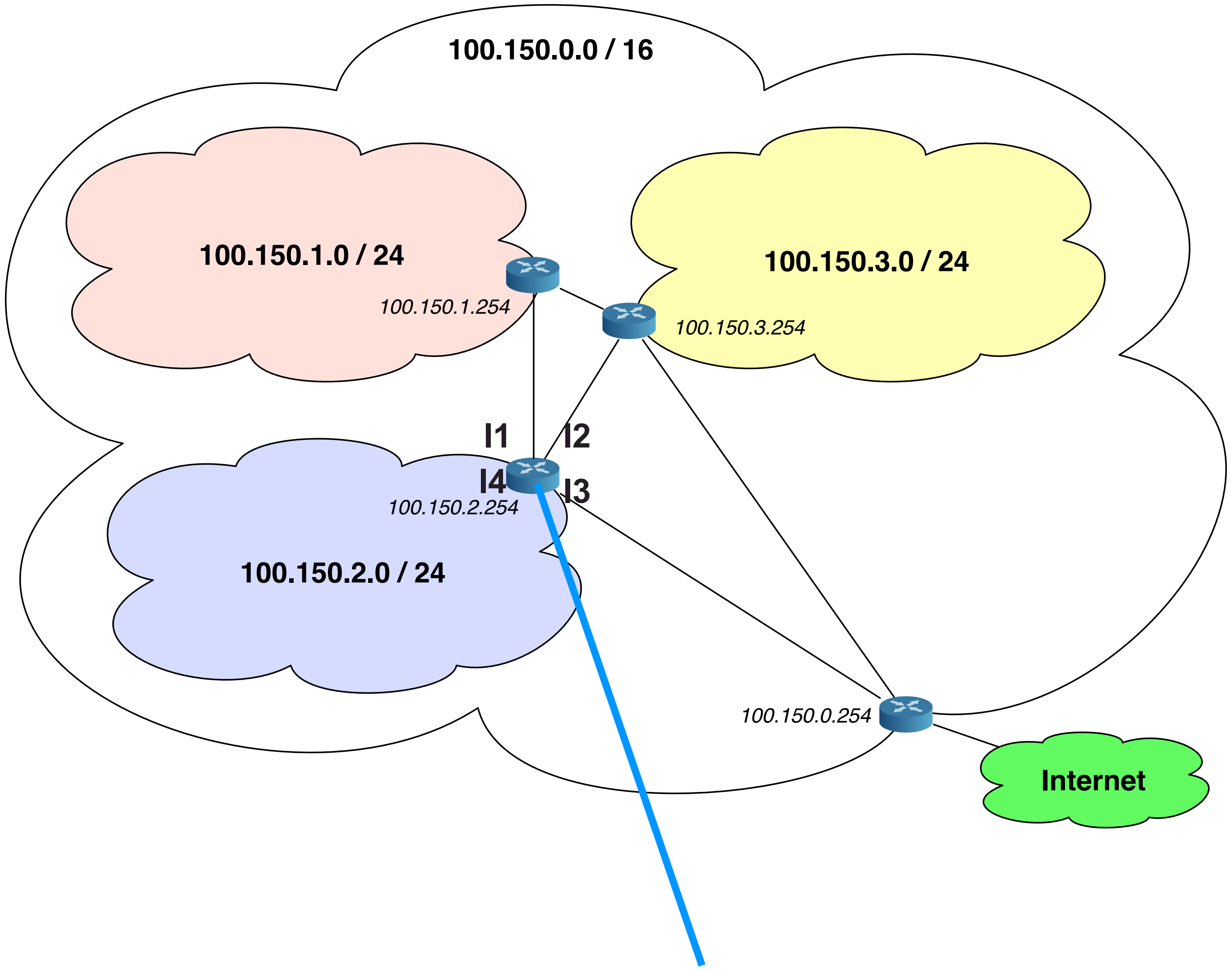
IP forwarding: at the router level

- A router is an interconnection device
  - has multiple network interfaces
  - has one IP address per interface



- Packets forwarding process (base)
  - 1) Look at the destination IP address in the packet header
  - 2) Look in a table (routing table) on which interface the packet should get out and what is the next router's address
  - 3) Transmit the packet to the next router
  - 4) When no match, a default route can be utilized (uncommon in carrier networks)

Routing table example



Destination	Netmask	Gateway	Port
100.150.1.0	255.255.255.0	100.150.1.254	I1
100.150.3.0	255.255.255.0	100.150.3.254	I2
100.150.2.0	255.255.255.0	0.0.0.0	I4
0.0.0.0	0.0.0.0	100.150.0.254	I3

This network is directly connected to the interface

Default route, when no other choice applies

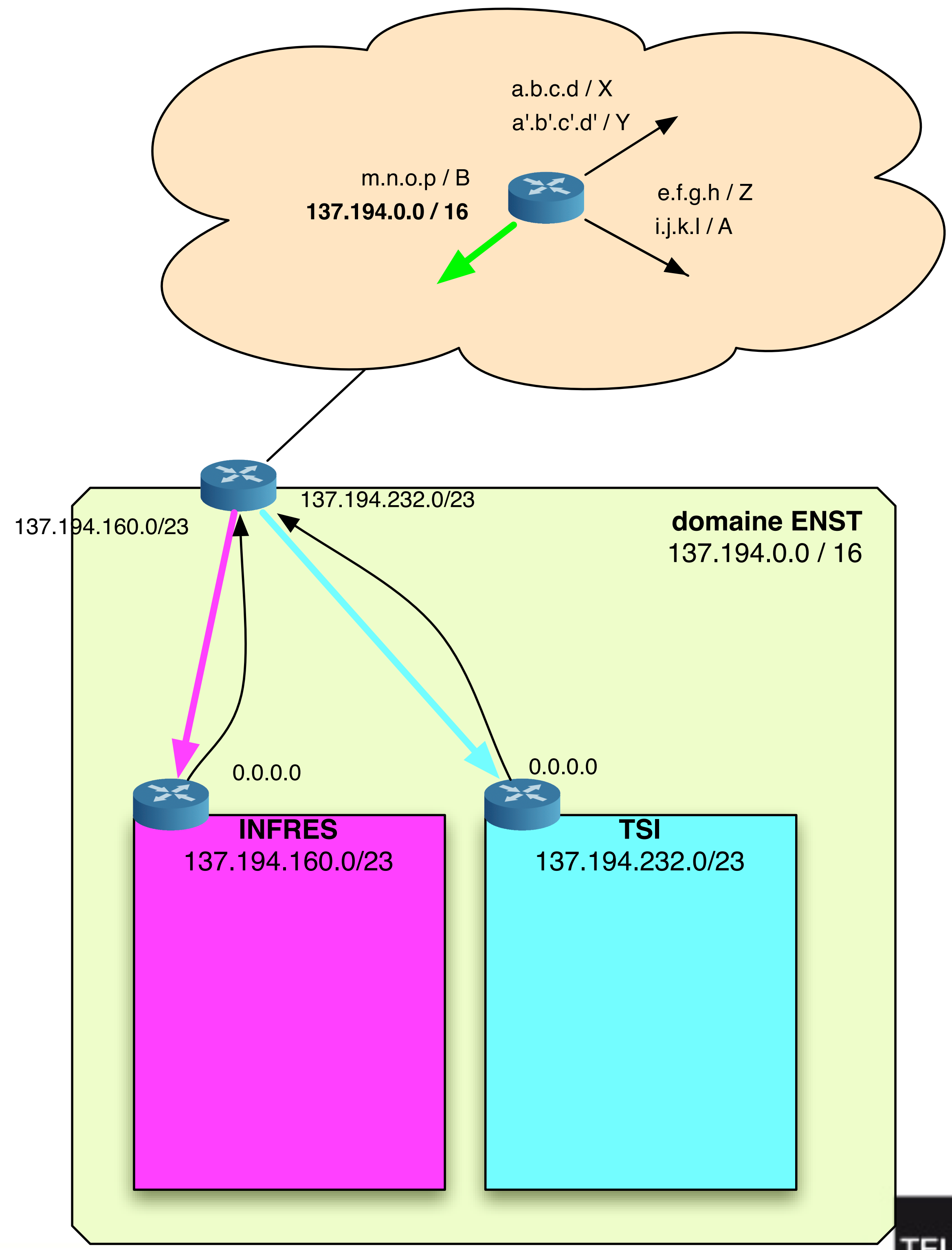
Routing table example (2)

- Partial routing table for one of the school routers
  - Total number of IP routes: 687

Destination	NetMask	Gateway	Port	Cost
137.194.2.0	255.255.254.0	137.194.4.254	v10	2
137.194.4.0	255.255.255.248	137.194.4.253	v10	2
137.194.4.8	255.255.255.248	137.194.4.251	v10	11
137.194.4.192	255.255.255.192	0.0.0.0	v10	1
137.194.6.0	255.255.254.0	137.194.4.254	v10	2
137.194.8.0	255.255.248.0	137.194.4.251	v10	20
137.194.16.0	255.255.255.128	137.194.160.230	v160	11
137.194.16.128	255.255.255.128	137.194.192.102	v192	11
137.194.16.144	255.255.255.240	137.194.192.102	v192	11
137.194.16.176	255.255.255.240	137.194.192.103	v192	20
137.194.17.0	255.255.255.0	137.194.192.103	v192	2
137.194.17.128	255.255.255.240	137.194.192.103	v192	20
137.194.17.144	255.255.255.240	137.194.192.103	v192	20

Forwarding policy: *Longest prefix match*

- Routes are only determined based on the destination address
- Intermediate routers only look at the "network" part of the address
  - As we get closer to the destination, routers have more and more precise information
  - If two routes match a destination address, the router chooses the most precise one (longer prefix)



Routing from one router's point of view

- Examine the packet's destination IP address

- Look into the routing table for the best match
  - Apply the network mask to the destination IP address
  - If there is a match, keep compare it to the current matching
    - If the prefix length is longer, select the new entry, otherwise keep the previous one
  - Forward the frame to the next hop

- Arriving at destination
  - The matching process indicates that the destination network is the one the router is directly connected to (corresp. to one of its own IP addresses / netmask)
  - Forward directly using ARP / L2 technology

Routing tables...

- Everything therefore depends on the routing tables
  - Necessary to keep routing tables as up-to-date as possible

- Regular maintenance
  - Cleanup of entries (timeout)
  - Routing table aggregation

```
137.194.2.0 / 23 → 137.194.1.1 (eth0)
137.194.1.0 / 24 → 137.194.1.1 (eth0)
137.194.0.0 / 24 → 137.194.1.1 (eth0)
```

$\Rightarrow$ `137.194.0.0 / 22 → 137.194.1.1 (eth0)`

- Addition / Refresh of these tables usually by a **routing protocol**
  - Link-state protocols
  - Distance vector protocols
  - ... or manual configuration (static routing)
    - `route add -net 10.1.3.0 netmask 255.255.255.0 gw 10.1.1.1`

Static routing

- All routers are configured by hand
  - Once configured, routes do not change
- Well adapted to a simple network
  - Home network, a single “router”, the set-top-box

- Difficult to manage for a large-scale network
  - Many routers (hundreds)
  - Many routes (hundreds of thousands) — There is usually no default route
  - Failures detection / reconfiguration necessary
  - Load balancing / dynamics difficult to manage by hand

Internet Protocol: IP packet headers

IPv4 header

<table>
<tr><td>V</td><td>L</td><td colspan="2">TOS</td><td colspan="2">Total Length</td></tr>
<tr><td colspan="4">Identification</td><td>F</td><td>Frag</td></tr>
<tr><td colspan="2">TTL</td><td colspan="2">Proto</td><td colspan="2">Checksum</td></tr>
<tr><td colspan="6">Source address</td></tr>
<tr><td colspan="6">Destination address</td></tr>
<tr><td colspan="3">Options</td><td colspan="3">Padding</td></tr>
<tr><td colspan="6">Data
(payload)</td></tr>
</table>

- **V**: Version (IPv4 ; IPv6)
- **L**: Header length (variable because of options)
- **TOS**: *Type of Service* - flow class
- **Total length**: packet length
- **Identification / F / Frag**: fragmentation
- **TTL**: *Time to* Live *;* maximum number of routers traversals authorized
- **Proto**: upper level protocol (TCP or UDP)
- **Checksum**: header integrity check
- **Source address**
- **Destination address**
- **Options** (security, ...)
- **Padding**: to align length to a 32 bits multiple

IPv6: header evolution

<table>
<tr><td>V</td><td>L</td><td colspan="2">TOS</td><td colspan="2">Total Length</td></tr>
<tr><td colspan="4">Identification</td><td>F</td><td>Frag</td></tr>
<tr><td colspan="2">TTL</td><td colspan="2">Proto</td><td colspan="2">Checksum</td></tr>
<tr><td colspan="6">Source address</td></tr>
<tr><td colspan="6">Destination address</td></tr>
<tr><td colspan="3">Options</td><td colspan="3">Padding</td></tr>
</table>

IPv6: header evolution

<table>
<tr><td>V</td><td>L</td><td colspan="2">TOS</td><td colspan="2">Total Length</td></tr>
<tr><td colspan="4">Identification</td><td>F</td><td>Frag</td></tr>
<tr><td colspan="2">TTL</td><td colspan="2">Proto</td><td colspan="2">Checksum</td></tr>
<tr><td colspan="6">Source address</td></tr>
<tr><td colspan="6">Destination address</td></tr>
<tr><td colspan="3">Options</td><td colspan="3">Padding</td></tr>
</table>

<table>
<tr><td>V</td><td>Class</td><td colspan="3">Flow Label</td></tr>
<tr><td colspan="3">Payload length</td><td>Next H</td><td>Hop lim</td></tr>
<tr><td colspan="5">Source address</td></tr>
<tr><td colspan="5">Destination address</td></tr>
</table>

IPv6: header evolution

V	L	TOS	Total Length	
Identification			F	Frag
TTL	Proto	Checksum		
Source address				
Destination address				
Options	Padding			

V	Class	Flow Label	
Payload length		Next H	Hop lim
Source address			
Destination address			

- 7 fields instead of 13
  - Easier routing
  - Fixed size
- Removed fields
  - Checksum ; header length ; fragmentation ; options

IPv6: header

- Version: 6
- Traffic class (QoS)
- Flow Label (QoS)
- Payload length (replaces *packet length*)
- Next header (replaces *protocol*)
- Hop Limit (replaces *TTL*)
- Source address
- Destination address

<table>
<tr><td>V</td><td colspan="2">Class</td><td colspan="2">Flow Label</td></tr>
<tr><td colspan="2">Longueur payload</td><td colspan="2">Next H</td><td>Hop lim</td></tr>
<tr><td colspan="5">Source address</td></tr>
<tr><td colspan="5">Destination address</td></tr>
</table>