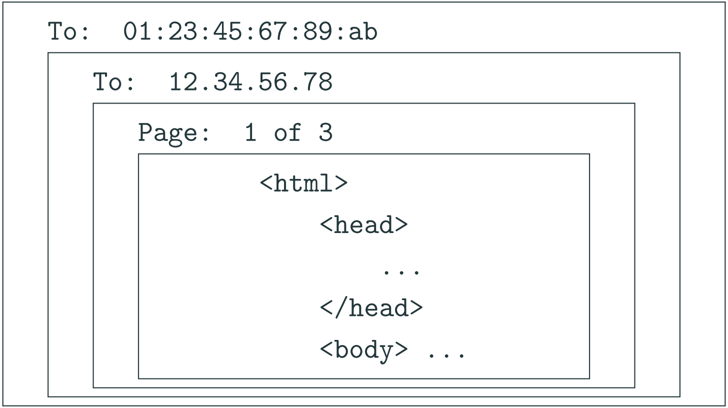
Internet

MPRI 2.26.2: Web Data Management

Antoine Amarilli

General idea

- Several **scales** (local vs global)
- **Stack** of protocols
- Embedded **messages**



To: 01:23:45:67:89:ab

To: 12.34.56.78

Page: 1 of 3

```
<html>
    <head>
        ...
    </head>
    <body> ...
```

OSI model

#	Layer	Examples	Features
7	Application	HTTP, FTP, SMTP	high level task
4	Transport	TCP, UDP, ICMP	sessions, reliable data, fragmentation
3	Network	IPv4, IPv6	routing, addressing
2	Link	Ethernet, 802.11	local addresses
1	Physical	Ethernet, 802.11	physical exchange, unreliable

→ The outermost envelopes are for the lowest layers

Table of Contents

OSI model

Low layers

Higher layers

IP (Internet Protocol), layer 3

- Gives **addresses** to computers
- Routes **packets** between these addresses
- Can get approximate **geographic location** for an IP

	Year	Example	Addresses
IPv4	1981	208.80.152.201	$\leq 2^{32}$
IPv6	1998	2620:0:860:ed1a::1	$\leq 2^{128}$

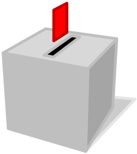
- **Network Address Translation** to get more IPv4 addresses

→ **We can send messages to an address**

POLL: IPv4 vs IPv6

Which proportion of traffic uses IPv6?

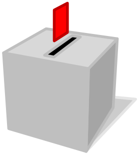
- **A**: less than 25%
- **B**: 25%–50%
- **C**: 50%–75%
- **D**: over 75%



POLL: IPv4 vs IPv6

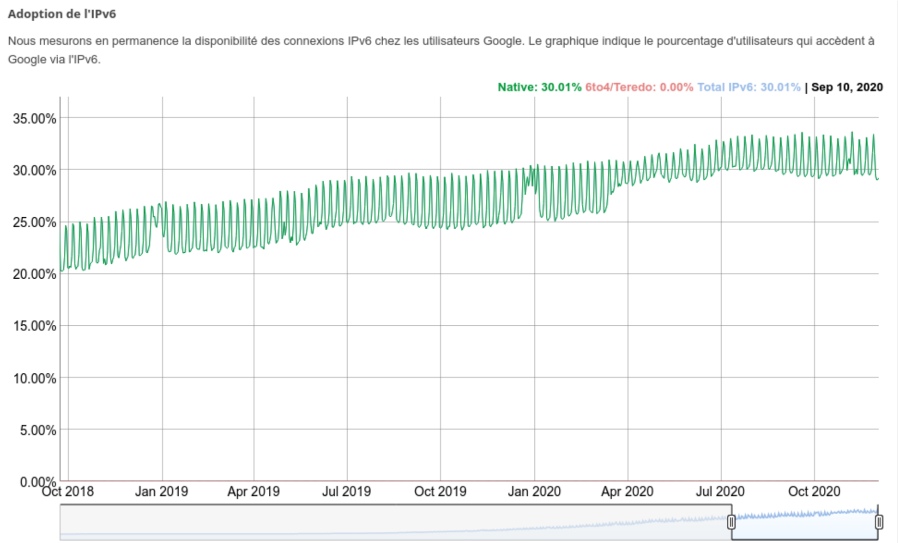
Which proportion of traffic uses IPv6?

- **A**: less than 25%
- **B**: **25%–50%**
- **C**: 50%–75%
- **D**: over 75%



Trafic IPv6 vs IPv4

`https://www.google.com/intl/en/ipv6/statistics.html`



Adoption de l'IPv6

Nous mesurons en permanence la disponibilité des connexions IPv6 chez les utilisateurs Google. Le graphique indique le pourcentage d'utilisateurs qui accèdent à Google via l'IPv6.

Native: 30.01% 6to4/Teredo: 0.00% Total IPv6: 30.01% | **Sep 10, 2020**

DNS (Domain Name System) – side note

- Convert names (www.wikipedia.org) to addresses (208.80.152.201)
- Hierarchy: `org`, `wikipedia.org`, `en.wikipedia.org`, etc.
- gTLDs, registrars, costs, effective TLDs

DNS (Domain Name System) – side note

- Convert names (www.wikipedia.org) to addresses (208.80.152.201)
- Hierarchy: `org`, `wikipedia.org`, `en.wikipedia.org`, etc.
- gTLDs, registrars, costs, effective TLDs
- Caching at several layers, security
- Special characters (IDN, Punycode...) and problems
- Useful indirection layer:
  - Several addresses per domain name (multiple services, load balancing)
  - Multiple domain names per address (virtual host)

DNS (Domain Name System) – side note

- Convert **names** (www.wikipedia.org) to **addresses** (208.80.152.201)
- Hierarchy: `org`, `wikipedia.org`, `en.wikipedia.org`, etc.
- **gTLDs**, registrars, costs, effective TLDs
- **Caching** at several layers, **security**
- **Special characters** (IDN, Punycode...) and problems
- Useful **indirection layer**:
  - Several addresses per domain name (multiple services, load balancing)
  - Multiple domain names per address (virtual host)

→ **Political** implications

→ **Public** DNSes, **alternative** roots, **decentralized alternatives**

→ **We can send messages to a named machine.**

TCP (Transmission Control Protocol), layer 4

- IP is not **reliable**
  - → TCP provides **delivery receipts**
- IP limits the **packet size**
  - → TCP can **fragment** large data
- IP can **mix packets**
  - → TCP ensures **in-order delivery**
- IP is not **multiplexed**
  - → TCP has **sessions** and **ports** (e.g. 80 for the Web)

→ **We can have a two-way communication channel with a machine.**

Table of Contents

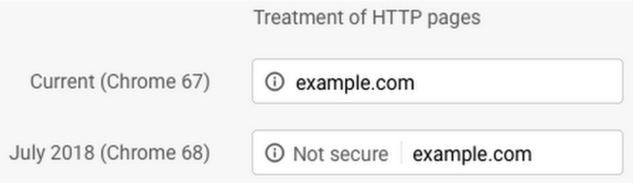
OSI model

Low layers

Higher layers

TLS (Transport Layer Security), layer 5-6

- Communicating in plaintext is risky! (passwords, credit cards...)
- Guarantees: integrity, authenticity, confidentiality
- HTTP + TLS = HTTPS. `https://`.
- Uses asymmetric cryptography
- Does not protect all metadata, possible side channels (size, etc.)
- Ongoing push towards HTTPS (+HSTS), marking HTTP as insecure



Treatment of HTTP pages

Current (Chrome 67) — ⓘ example.com

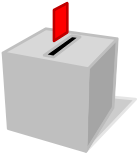
July 2018 (Chrome 68) — ⓘ Not secure | example.com

`https://www.blog.google/products/chrome/milestone-chrome-security-marking-http-not-secure/`

POLL: HTTPS

Which proportion of Web pages loaded by Chrome users is encrypted with HTTPS?[a]

- **A**: less than 25%
- **B**: 25%–50%
- **C**: 50%–75%
- **D**: over 75%

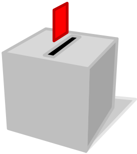


[a] Source: `https://transparencyreport.google.com/https/overview`

POLL: HTTPS

Which proportion of Web pages loaded by Chrome users is encrypted with HTTPS?[a]

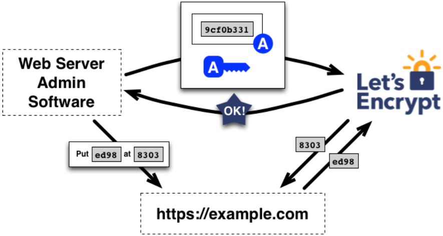
- **A**: less than 25%
- **B**: 25%–50%
- **C**: 50%–75%
- **D**: **over 75%**



[a] Source: `https://transparencyreport.google.com/https/overview`

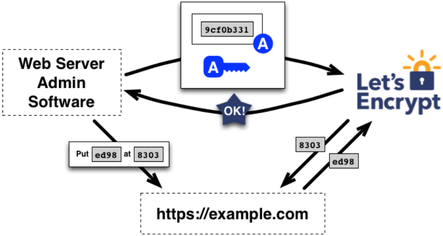
Let’s Encrypt vs extended validation

- Let’s Encrypt: automated check (ACME protocol) and signature of an HTTPS certificate

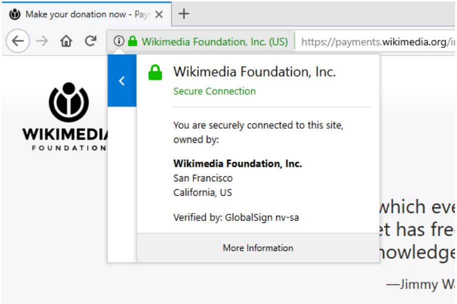


Let’s Encrypt vs extended validation

- **Let’s Encrypt**: automated check (ACME protocol) and signature of an HTTPS certificate



- **Extended Validation** certificates: manual identify check by **trusted parties**


Make your donation now - Pay
Wikimedia Foundation, Inc. (US)
https://payments.wikimedia.org/i
WIKIMEDIA FOUNDATION
which eve
t has fre
owledge
—Jimmy W

Wikimedia Foundation, Inc.
Secure Connection

You are securely connected to this site, owned by:

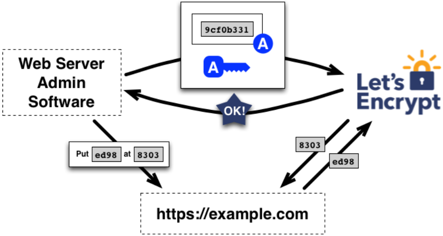
Wikimedia Foundation, Inc.
San Francisco
California, US

Verified by: GlobalSign nv-sa

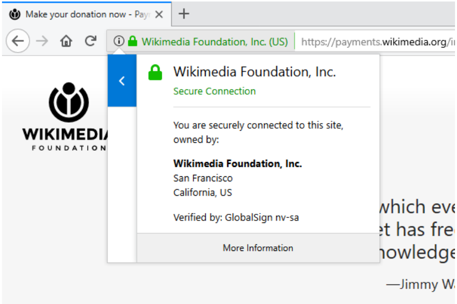
More Information

Let’s Encrypt vs extended validation

- Let’s Encrypt: automated check (ACME protocol) and signature of an HTTPS certificate



- Extended Validation certificates: manual identify check by trusted parties


Make your donation now - Pay
+
Wikimedia Foundation, Inc. (US)
https://payments.wikimedia.org/i
WIKIMEDI
FOUNDATION

Wikimedia Foundation, Inc.
Secure Connection

You are securely connected to this site, owned by:

Wikimedia Foundation, Inc.
San Francisco
California, US

Verified by: GlobalSign nv-sa

More Information

which eve
et has fre
nowledge
—Jimmy W

→ We have an encrypted channel between two machines

`https://letsencrypt.org/how-it-works/`

`Wikimedia_donation_page_with_extended_validation_certificate_in_firefox.png` on Wikimedia commons

Credits

- Matériel de cours inspiré de notes par Pierre Senellart et Georges Gouriten
- Merci à Pierre Senellart pour sa relecture