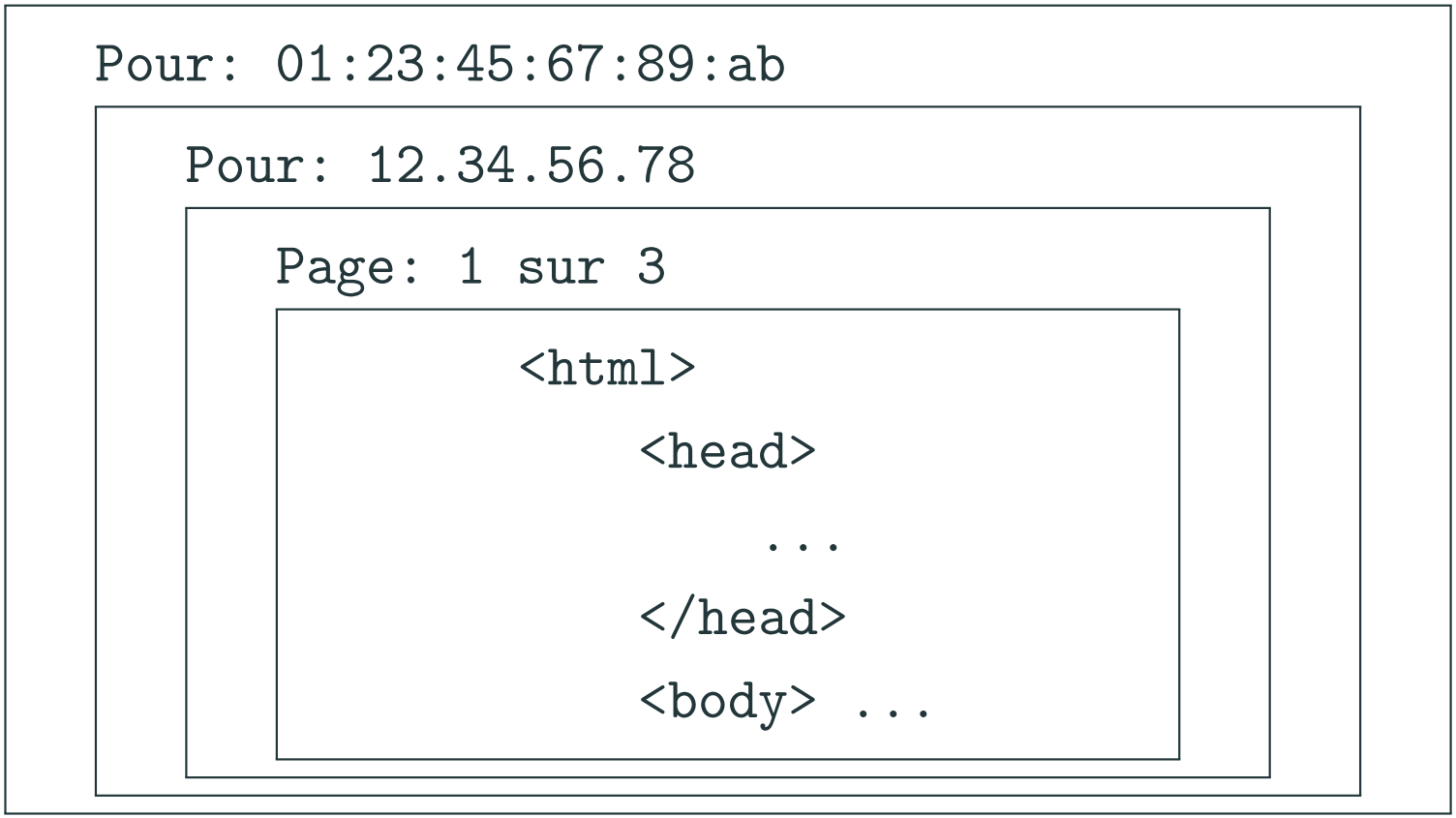
INF344 – Données du Web
Internet

Antoine Amarilli

Vue générale

- Plusieurs **échelles** (locale et globale)
- **Pile** de protocoles
- Messages **imbriqués**



Le modèle OSI

#	Couche	Exemples	Fonctionnalités
7	Application	HTTP, FTP, SMTP	tâche utilisateur de haut niveau
4	Transport	TCP, UDP, ICMP	sessions, fiabilité, fragmentation
3	Réseau	IPv4, IPv6	routage, adressage, non fiable
2	Lien	Ethernet, 802.11	données fiables, adresses locales
1	Physique	Ethernet, 802.11	échange physique, non fiable

→ Plus la couche est basse, plus l'enveloppe est à l'extérieur

Table des matières

IP (Internet Protocol), couche 3

- Donner des **adresses** aux machines
- **Router** des paquets entre ces adresses
- Déterminer la position **géographique** d'une machine

	Année	**Exemple**	**Adresses**
IPv4	1981	**208.80.152.201**	$\leq 2^{32}$
IPv6	1998	**2620:0:860:**ed**1**a**::1**	$\leq 2^{128}$

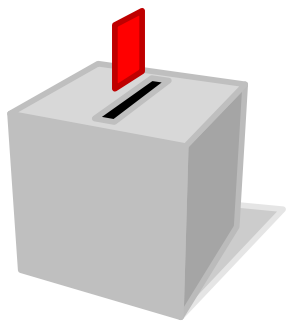
- **Network Address Translation** pour pallier la pénurie d'adresses

→ **On peut envoyer des messages à une adresse.**

SONDAGE : IPv4 vs IPv6

Quelle part du trafic utilise IPv4 plutôt que IPv6 ?

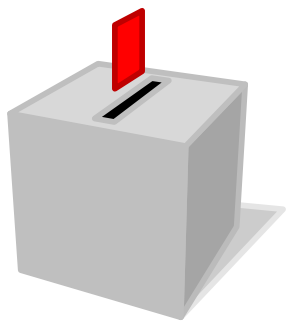
- **A**: moins de 25%
- **B**: 25%–50%
- **C**: 50%–75%
- **D**: plus de 75%



SONDAGE : IPv4 vs IPv6

Quelle part du trafic utilise IPv4 plutôt que IPv6 ?

- **A**: moins de 25%
- **B**: **25%–50%**
- **C**: 50%–75%
- **D**: plus de 75%

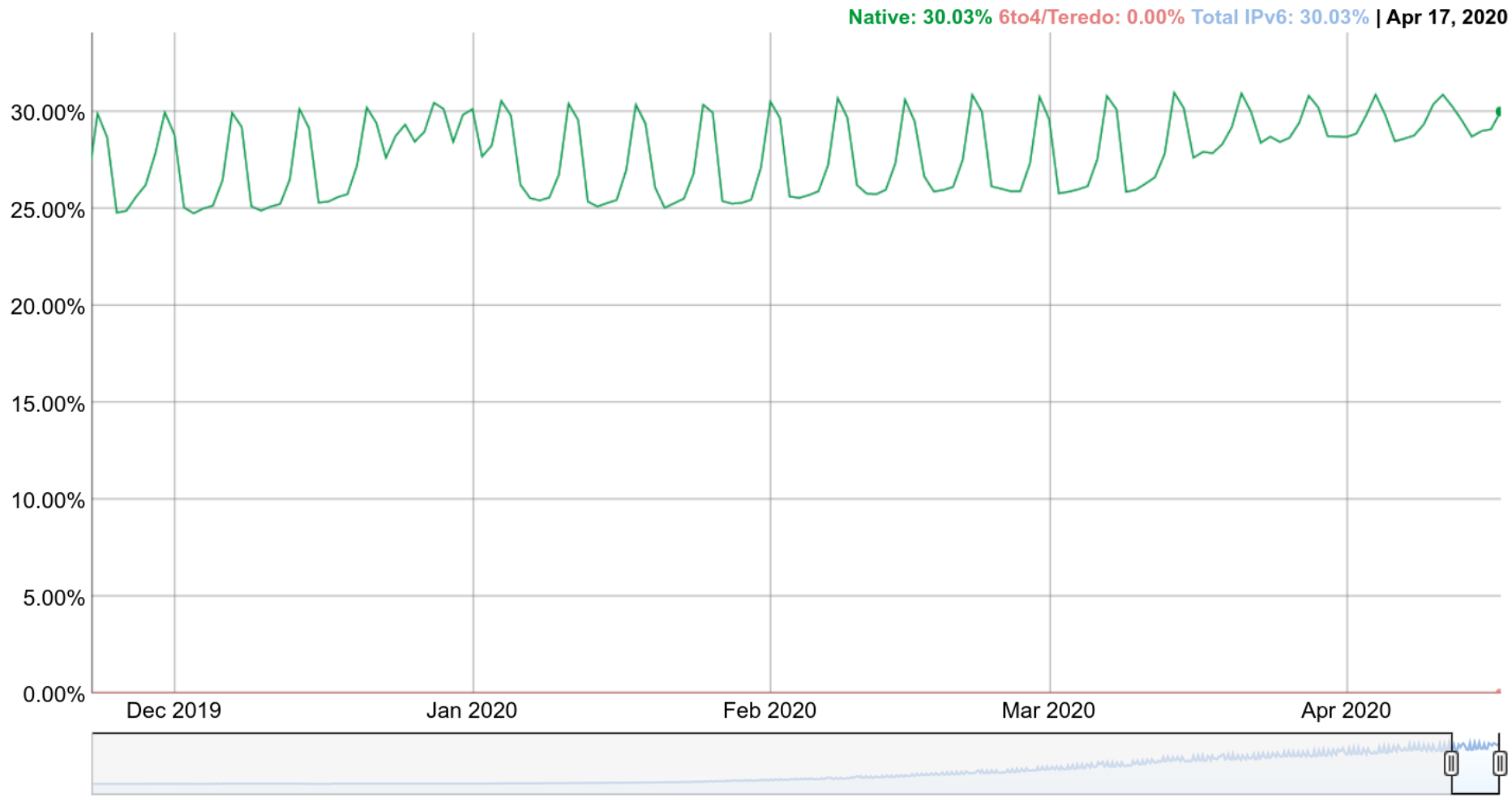


Trafic IPv6 vs IPv4

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

IPv6 Adoption

We are continuously measuring the availability of IPv6 connectivity among Google users. The graph shows the percentage of users that access Google over IPv6.



DNS (Domain Name System), intermède

- Service pour convertir www.wikipedia.org en 208.80.152.201
- Hiérarchie : `org`, `wikipedia.org`, `en.wikipedia.org`, etc.
- Résolution hiérarchique ; gTLDs, registrars, coûts, TLDs effectifs

DNS (Domain Name System), intermède

- Service pour convertir www.wikipedia.org en 208.80.152.201
- Hiérarchie : `org`, `wikipedia.org`, `en.wikipedia.org`, etc.
- Résolution hiérarchique ; gTLDs, registrars, coûts, TLDs effectifs
- Cache à différents niveaux
- Problèmes de sécurisation (authentification, empoisonnement…)
- Caractères spéciaux (IDN, Punycode) et problèmes afférents

DNS (Domain Name System), intermède

- Service pour convertir www.wikipedia.org en 208.80.152.201
- Hiérarchie : `org`, `wikipedia.org`, `en.wikipedia.org`, etc.
- Résolution **hiérarchique** ; **gTLDs**, registrars, coûts, TLDs effectifs
- **Cache** à différents niveaux
- Problèmes de **sécurisation** (authentification, empoisonnement…)
- Caractères **spéciaux** (IDN, Punycode) et problèmes afférents
- **Indirection** :
  - Plusieurs adresses par nom de domaine
    (services multiples, répartition de charge)
  - Plusieurs noms de domaine par adresse (virtual hosts)

→ Qui **gère** le DNS ? Qui a **droit** à un nom de domaine ?

→ DNS alternatifs, autres technologies décentralisées (Namecoin…)

DNS (Domain Name System), intermède

- Service pour convertir www.wikipedia.org en 208.80.152.201
- Hiérarchie : `org`, `wikipedia.org`, `en.wikipedia.org`, etc.
- Résolution **hiérarchique** ; **gTLDs**, registrars, coûts, TLDs effectifs
- **Cache** à différents niveaux
- Problèmes de **sécurisation** (authentification, empoisonnement…)
- Caractères **spéciaux** (IDN, Punycode) et problèmes afférents
- **Indirection** :
  - Plusieurs adresses par nom de domaine
    (services multiples, répartition de charge)
  - Plusieurs noms de domaine par adresse (virtual hosts)

→ Qui **gère** le DNS ? Qui a **droit** à un nom de domaine ?

→ DNS alternatifs, autres technologies décentralisées (Namecoin…)

(Démo : résolution DNS avec `dig`.)

DNS (Domain Name System), intermède

- Service pour convertir www.wikipedia.org en 208.80.152.201
- Hiérarchie : `org`, `wikipedia.org`, `en.wikipedia.org`, etc.
- Résolution **hiérarchique** ; **gTLDs**, registrars, coûts, TLDs effectifs
- **Cache** à différents niveaux
- Problèmes de **sécurisation** (authentification, empoisonnement…)
- Caractères **spéciaux** (IDN, Punycode) et problèmes afférents
- **Indirection** :
  - Plusieurs adresses par nom de domaine
    (services multiples, répartition de charge)
  - Plusieurs noms de domaine par adresse (virtual hosts)

→ Qui **gère** le DNS ? Qui a **droit** à un nom de domaine ?

→ DNS alternatifs, autres technologies décentralisées (Namecoin…)

(Démo : résolution DNS avec `dig`.)

→ **On peut envoyer des messages à une machine nommée.**

TCP (Transmission Control Protocol), couche 4

- IP n'est pas **fiable**
  - → TCP fournit des **accusés de réception**
- IP **limite la taille**
  - → TCP permet de **fragmenter**
- IP peut **mélanger les paquets**
  - → TCP garantit que les paquets arrivent **dans l'ordre**
- IP n'est pas **multiplexé**
  - → TCP introduit des **sessions** et des **ports**. (e.g. 80 pour le Web… mais possible de forcer : `http://localhost:8080/`)

TCP (Transmission Control Protocol), couche 4

- IP n'est pas **fiable**
  → TCP fournit des **accusés de réception**
- IP **limite la taille**
  → TCP permet de **fragmenter**
- IP peut **mélanger les paquets**
  → TCP garantit que les paquets arrivent **dans l'ordre**
- IP n'est pas **multiplexé**
  → TCP introduit des **sessions** et des **ports**. (e.g. 80 pour le Web… mais possible de forcer : `http://localhost:8080/`)

(Démo : communication TCP avec `netcat`.)

TCP (Transmission Control Protocol), couche 4

- IP n'est pas **fiable**
  → TCP fournit des **accusés de réception**
- IP **limite la taille**
  → TCP permet de **fragmenter**
- IP peut **mélanger les paquets**
  → TCP garantit que les paquets arrivent **dans l'ordre**
- IP n'est pas **multiplexé**
  → TCP introduit des **sessions** et des **ports**. (e.g. 80 pour le Web… mais possible de forcer : `http://localhost:8080/`)

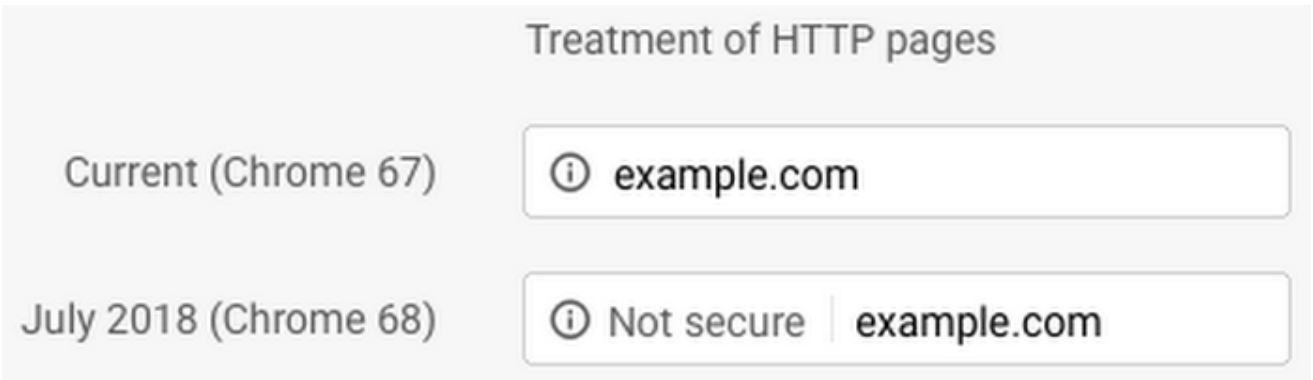
(Démo : communication TCP avec `netcat`.)

→ **On peut avoir un canal de communication avec une machine.**

Table des matières

TLS (Transport Layer Security), layer 5-6

- C'est dangereux de communiquer en clair ! (mots de passe, numéros CB…)
- Garanties : intégrité, authenticité, confidentialité
- HTTP + TLS = HTTPS. `https://`.
- Utilise de la cryptographie asymétrique
- Ne protège pas toutes les métadonnées (taille, etc.)
- Évolution vers HTTPS (+HSTS), HTTP est marqué comme non sécurisé

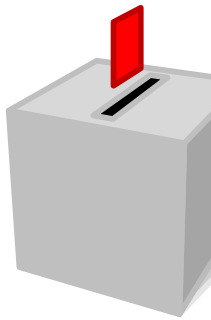


`https://www.blog.google/products/chrome/milestone-chrome-security-marking-http-not-secure/`

SONDAGE : HTTPS

Quelle proportion des pages Web chargées par les utilisateurs de Chrome est chiffrée avec HTTPS ?[a]

- **A**: moins de 25%
- **B**: 25%–50%
- **C**: 50%–75%
- **D**: plus de 75%

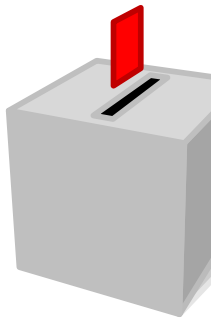


a. Source : `https://transparencyreport.google.com/https/overview`

SONDAGE : HTTPS

Quelle proportion des pages Web chargées par les utilisateurs de Chrome est chiffrée avec HTTPS ?[a]

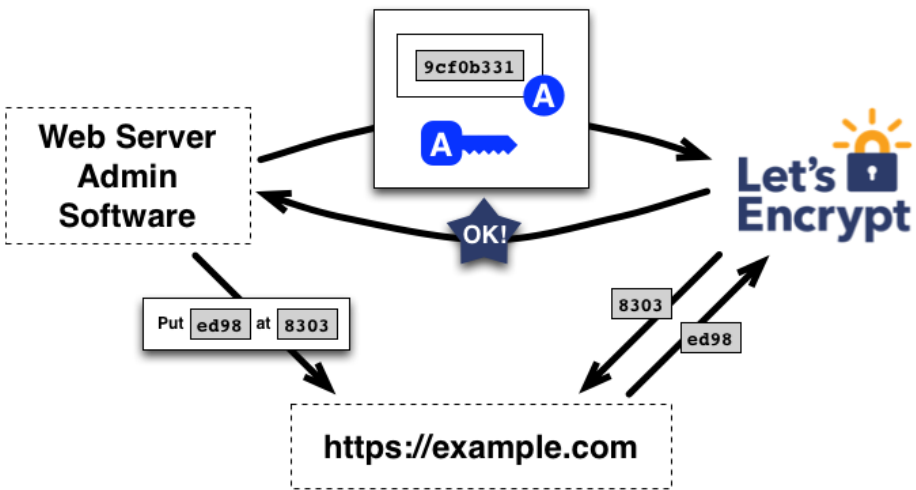
- **A**: moins de 25%
- **B**: 25%–50%
- **C**: 50%–75%
- **D**: **plus de 75%**



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

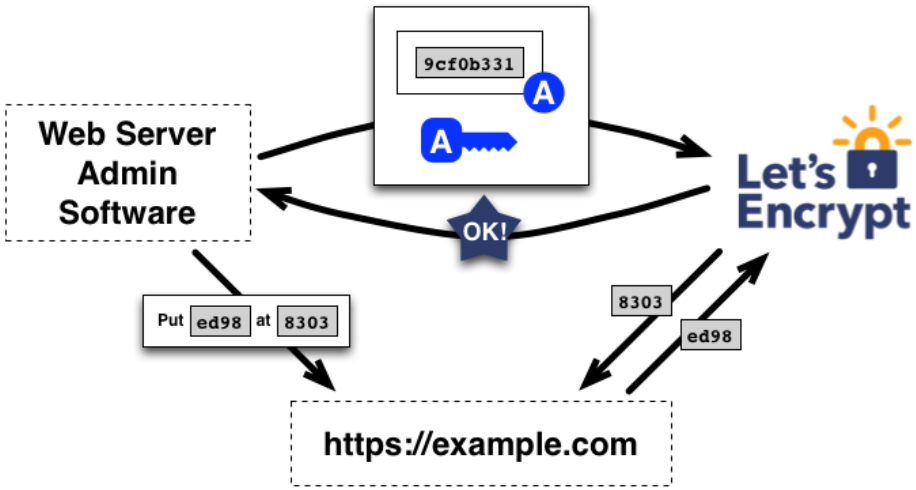
Let’s Encrypt et validation étendue

- Let’s Encrypt : vérification automatique (protocole ACME) et signature d’un certificat HTTPS

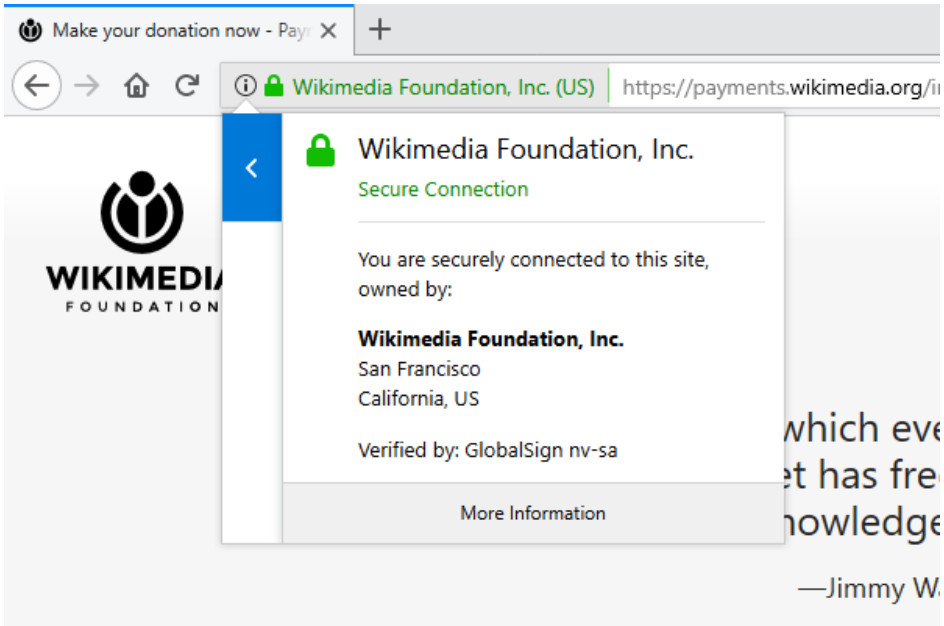


Let’s Encrypt et validation étendue

- Let’s Encrypt : vérification automatique (protocole ACME) et signature d’un certificat HTTPS

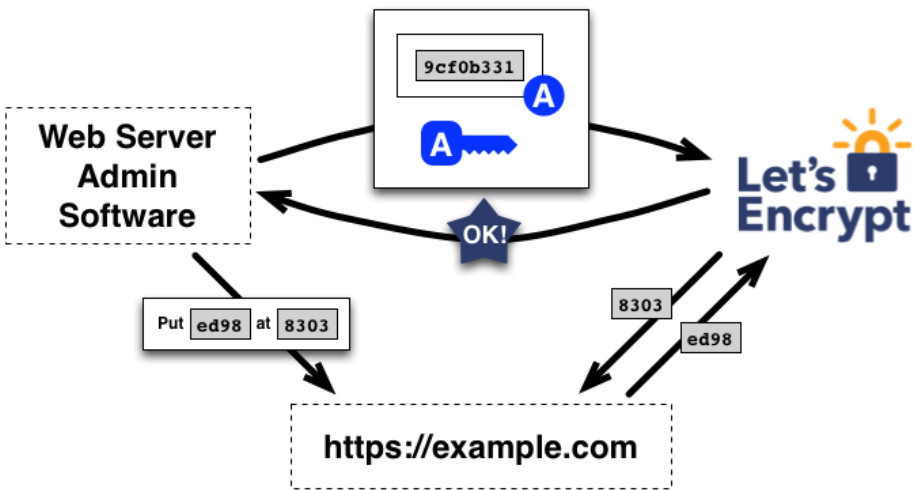


- Certificat Extended Validation : vérification manuelle d’identité par des tiers de confiance

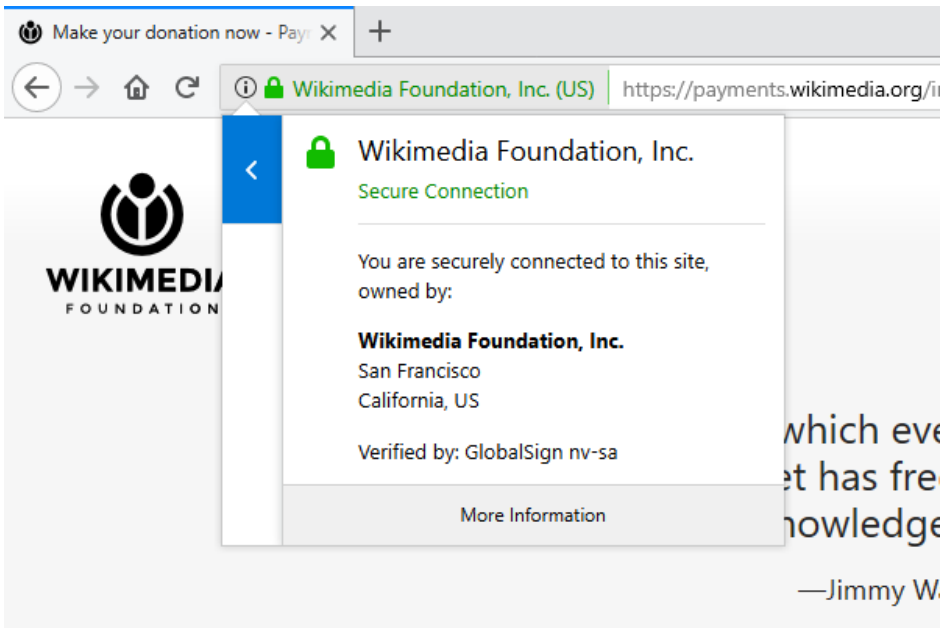


Let’s Encrypt et validation étendue

- Let’s Encrypt : vérification automatique (protocole ACME) et signature d’un certificat HTTPS



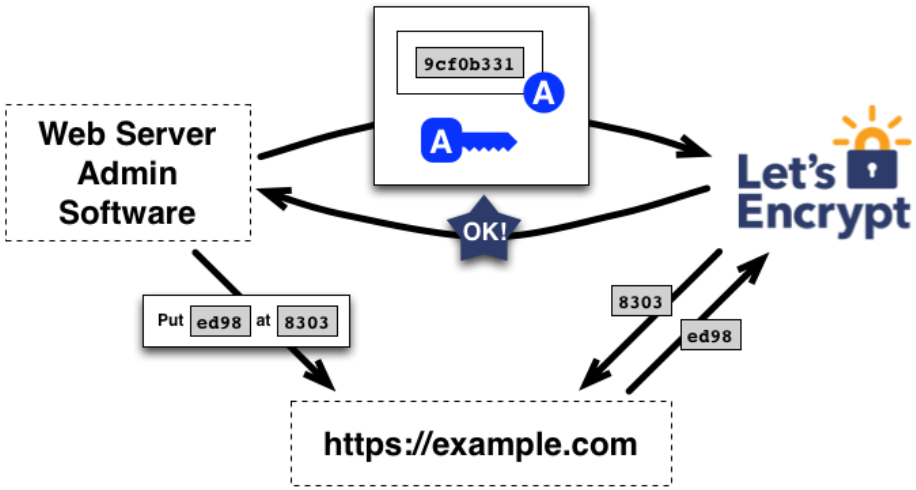
- Certificat Extended Validation : vérification manuelle d’identité par des tiers de confiance



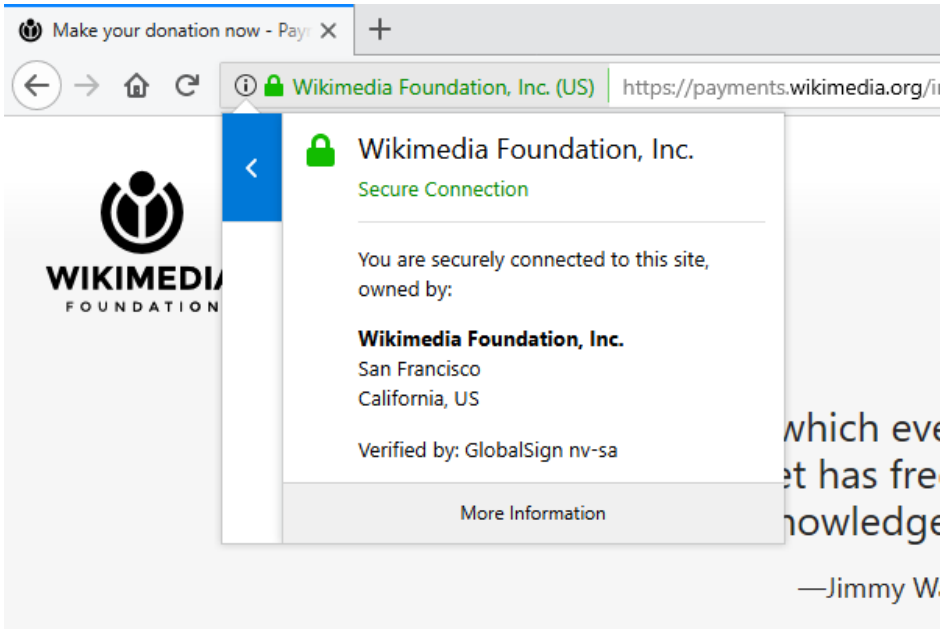
(Démo : communication chiffrée avec `ncat --ssl`, tentatives d’interception infructueuses avec `wireshark`.)

Let’s Encrypt et validation étendue

- Let’s Encrypt : vérification automatique (protocole ACME) et signature d’un certificat HTTPS



- Certificat Extended Validation : vérification manuelle d’identité par des tiers de confiance



(Démo : communication chiffrée avec `ncat --ssl`, tentatives d’interception infructueuses avec `wireshark`.)

→ On a un canal de communication chiffré entre deux machines