Proxy SQUID

Ecrit par **Youenn DUVAL**

Mail: youenn@barbed.fr

Linkedin: Youenn DUVAL

Derniere mise à jour : **25/10/2024**

Table des matières

Un **serveur mandataire**, également appelé **proxy** ou **proxy server**, est un serveur intermédiaire qui agit comme un **intermédiaire** entre un client (par exemple, un navigateur Web) et le serveur de destination (par exemple, un site Web). Le serveur proxy relaie les requêtes entre ces deux entités, ce qui permet de contrôler, filtrer, ou optimiser les communications réseau.

Serveur Mandataire

Fonctionnement d'un serveur mandataire :

- **Client → Proxy → Serveur** : Lorsqu'un client fait une requête pour accéder à une ressource (comme un site Web), la requête est envoyée d'abord au **proxy**. Le proxy évalue la requête, et, selon sa configuration, la transmet au serveur cible. Une fois que le serveur cible répond, le proxy relaye la réponse au client.
- **Cacher l'identité** : Le serveur proxy peut masquer l'adresse IP du client en présentant sa propre adresse IP au serveur de destination.

Types de serveurs mandataires :

- **Proxy HTTP** : Utilisé pour le trafic Web. Il relaie les requêtes HTTP entre les clients (navigateurs) et les serveurs Web.
- **Proxy de cache** : Stocke localement les réponses des serveurs Web pour accélérer les requêtes futures. Si une ressource a déjà été demandée par un client, le proxy de cache peut fournir cette ressource directement sans avoir à contacter le serveur d'origine.
- **Proxy inverse** : Placé devant un ou plusieurs serveurs, il reçoit les requêtes des clients et les distribue à l'un des serveurs en backend. Utilisé pour équilibrer la charge, améliorer la sécurité ou la performance des serveurs backend.
- **Proxy transparent** : Intercepte les communications sans que les clients ne s'en aperçoivent. Il ne modifie pas les requêtes et fonctionne comme un intermédiaire invisible.
- **Proxy anonyme** : Cache l'adresse IP du client. Le serveur de destination ne connaît pas l'adresse réelle du client, ce qui protège son anonymat.

Avantages d'un serveur mandataire :

- **Sécurité** : Le proxy peut filtrer les contenus indésirables, bloquer l'accès à certains sites Web ou surveiller le trafic réseau. Il peut aussi masquer les adresses IP des utilisateurs pour protéger leur anonymat.

- **Contrôle de l'accès** : Dans une entreprise ou un établissement scolaire, le proxy peut être configuré pour limiter l'accès à certains sites Web ou services, ou pour surveiller les activités en ligne des utilisateurs.
- **Cache et performances** : Les **proxies de cache** stockent les copies des pages Web visitées. Si une page est demandée plusieurs fois, le proxy peut fournir la version mise en cache, ce qui réduit la latence et la charge sur le serveur d'origine.
- **Anonymat** : En passant par un proxy, l'adresse IP réelle du client est masquée, ce qui permet de naviguer sur Internet de manière plus anonyme.
- **Compression de données** : Certains proxies compressent les données (comme les images) avant de les transmettre au client, ce qui peut réduire la bande passante utilisée.

Inconvénients d'un serveur mandataire :

- **Latence supplémentaire** : L'ajout d'un intermédiaire peut augmenter le temps de traitement, surtout si le proxy est mal configuré ou distant.
- **Confidentialité** : Certains serveurs proxy peuvent enregistrer les activités des utilisateurs, compromettant leur anonymat et confidentialité.
- **Limitations** : Les proxies peuvent restreindre l'accès à certains sites ou services, ce qui peut être frustrant si la politique de filtrage est trop restrictive.

Cas d'utilisation :

- **Entreprises** : Les entreprises utilisent souvent des serveurs mandataires pour filtrer le trafic Internet et contrôler l'accès à certains sites (réseaux sociaux, sites de streaming, etc.).
- **Sécurité** : Un proxy peut agir comme un **pare-feu** pour bloquer les attaques potentielles contre un réseau en filtrant le trafic indésirable.
- **Caches Web** : Les proxies de cache peuvent accélérer la navigation en stockant localement les pages Web fréquemment visitées, réduisant ainsi les délais d'accès.
- **Accès restreint** : Un proxy peut être utilisé pour accéder à des contenus géo-restreints en modifiant l'adresse IP apparente du client (par exemple, accéder à du contenu bloqué dans une région).

Exemple :

Dans un environnement d'entreprise, lorsque les employés tentent de visiter un site Web, leur navigateur envoie une requête au **serveur mandataire** configuré par l'entreprise. Le proxy vérifie si le site est autorisé et, si c'est le cas, transmet la requête au serveur cible. S'il a une copie du site Web en cache, il peut également la fournir directement, accélérant ainsi la navigation.

Proxy Squid

Exercice démarrer une stack SQUID

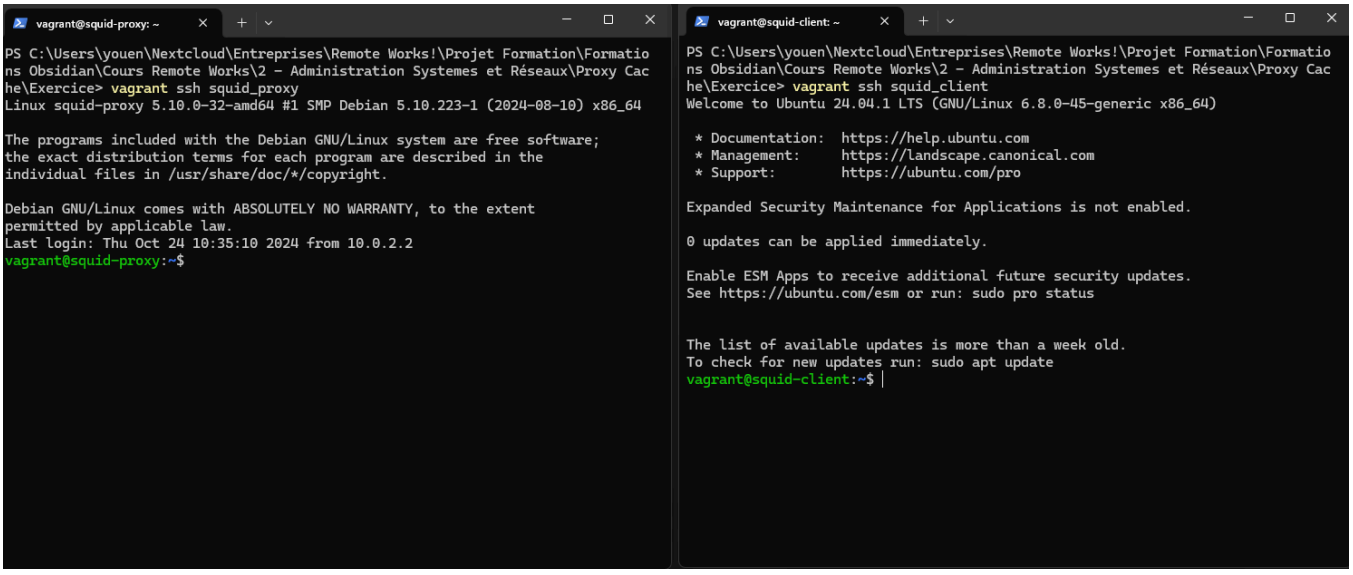
Vous trouverez le Vagrantfile pour démarrer plusieurs un serveur squid ainsi qu'un client pour tester.

Cela va vous démarrer deux machines.

- Squid_Server qui heberge un serveur squid et va s'occuper de gérer du cache utilisateur et préconfiguré pour laisser tout passer.
- Squid_Client qui est configuré pour utilisé un serveur mandataire pour plusieurs protocoles et nous servira pour tester des requetes.

je peux ouvrir deux terminaux (dans le meme dossier que le Vagrantfile). L'objectif est de connecter un terminal dans chaque machine avec

- `vagrant ssh squid_proxy`
- `vagrant ssh squid_client`



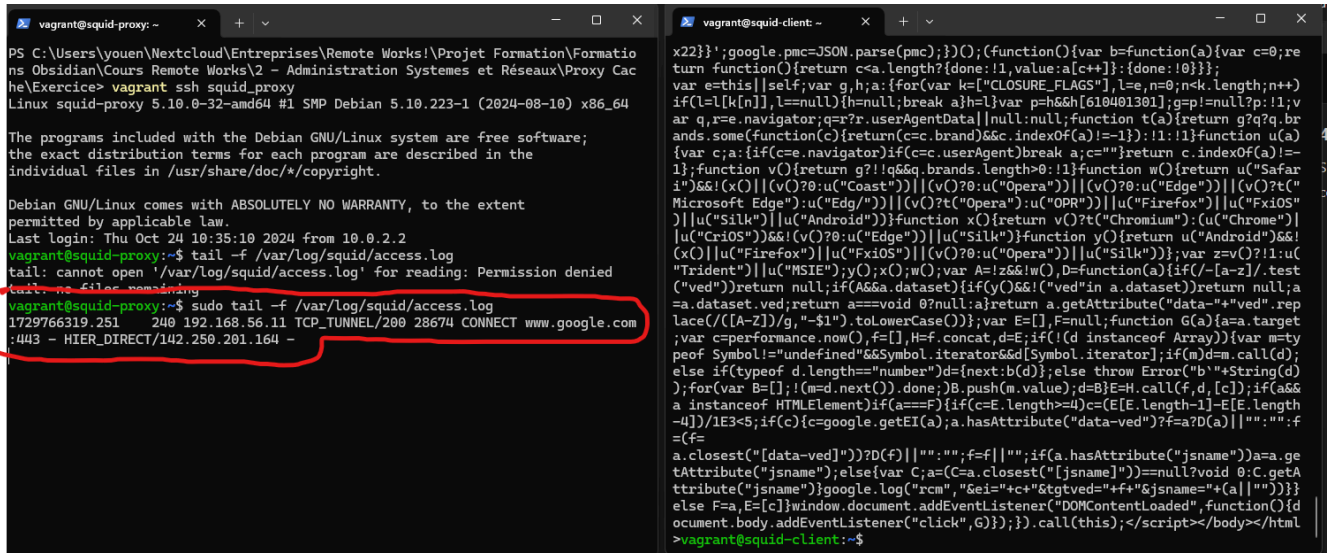
Je vais tester le fonctionnement en ouvrant les logs d'accés sur proxy squid

```
sudo tail -f /var/log/squid/access.log
```

Cela va ouvrir le fichier de log et m'afficher chaque changement.

Pour en voir apparaitre je vais simplement allez sur la machine client et faire une requete vers google.fr pour l'exemple. Pour le faire en ligne de commande je vais utiliser l'utilitaire CURL.

```
curl https://www.google.com
```



A droite je peux voir le retour de google et le contenue de la page. A gauche je peux voir que la requête est correctement passé par le serveur mandataire (proxy).

Il est possible d'obtenir plus d'information avec l'utilitaire squidclient que l'on peut installer avec `sudo apt install squidclient -y`.

Ensuite il suffit d'utiliser la commande: `squidclient -h localhost -p 3128 mgr:info`

```
vagrant@squid-proxy:~$ squidclient -h localhost -p 3128 mgr:info
HTTP/1.1 200 OK
Server: squid/4.13
Mime-Version: 1.0
Date: Thu, 24 Oct 2024 10:41:40 GMT
Content-Type: text/plain;charset=utf-8
Expires: Thu, 24 Oct 2024 10:41:40 GMT
Last-Modified: Thu, 24 Oct 2024 10:41:40 GMT
X-Cache: MISS from squid-proxy
X-Cache-Lookup: MISS from squid-proxy:3128
Via: 1.1 squid-proxy (squid/4.13)
Connection: close

Squid Object Cache: Version 4.13
Build Info: Debian linux
Service Name: squid
Start Time:     Thu, 24 Oct 2024 10:29:14 GMT
Current Time:   Thu, 24 Oct 2024 10:41:40 GMT
Connection information for squid:
        Number of clients accessing cache:      2
        Number of HTTP requests received:       1
        Number of ICP messages received:        0
        Number of ICP messages sent:    0
        Number of queued ICP replies:   0
        Number of HTCP messages received:       0
        Number of HTCP messages sent:   0
        Request failure ratio:   0.00
        Average HTTP requests per minute since start:   0.1
        Average ICP messages per minute since start:    0.0
        Select loop called: 1808 times, 412.730 ms avg
Cache information for squid:
        Hits as % of all requests:      5min: 0.0%, 60min: 0.0%
        Hits as % of bytes sent:        5min: -0.0%, 60min: -0.0%
        Memory hits as % of hit requests:       5min: 0.0%, 60min: 0.0%
        Disk hits as % of hit requests: 5min: 0.0%, 60min: 0.0%
        Storage Swap size:      0 KB
        Storage Swap capacity:   0.0% used,  0.0% free
        Storage Mem size:       216 KB
        Storage Mem capacity:    0.1% used, 99.9% free
        Mean Object Size:       0.00 KB
        Requests given to unlinkd:      0
Median Service Times (seconds)  5 min    60 min:
        HTTP Requests (All):   0.00000  0.00000
        Cache Misses:          0.00000  0.00000
        Cache Hits:            0.00000  0.00000
        Near Hits:             0.00000  0.00000
        Not-Modified Replies:  0.00000  0.00000
        DNS Lookups:           0.01686  0.01686
        ICP Queries:           0.00000  0.00000
Resource usage for squid:
        UP Time:        746.215 seconds
        CPU Time:       0.080 seconds
        CPU Usage:      0.01%
        CPU Usage, 5 minute avg:        0.01%
        CPU Usage, 60 minute avg:       0.01%
        Maximum Resident Size: 94512 KB
        Page faults with physical i/o: 0
Memory accounted for:
        Total accounted:          562 KB
        memPoolAlloc calls:      3719
        memPoolFree calls:       3736
File descriptor usage for squid:
        Maximum number of file descriptors:   1024
        Largest file desc currently in use:     14
        Number of file desc currently in use:    6
        Files queued for open:                   0
        Available number of file descriptors: 1018
        Reserved number of file descriptors:   100
        Store Disk files open:                   0
Internal Data Structures:
           52 StoreEntries
           52 StoreEntries with MemObjects
            0 Hot Object Cache Items
            0 on-disk objects
```

Si vous voulez tester avec plus d'url, voici un script a lancer sur le client:

```
#!/bin/bash
# Liste de 100 sites web
sites=(
  "https://www.google.com"
  "https://www.facebook.com"
  "https://www.twitter.com"
  "https://www.wikipedia.org"
  "https://www.amazon.com"
  "https://www.microsoft.com"
  "https://www.apple.com"
  "https://www.linkedin.com"
  "https://www.instagram.com"
  "https://www.netflix.com"
  "https://www.reddit.com"
  "https://www.ebay.com"
  "https://www.bing.com"
  "https://www.cnn.com"
  "https://www.yahoo.com"
  "https://www.wordpress.com"
  "https://www.nytimes.com"
  "https://www.bbc.com"
  "https://www.imdb.com"
  "https://www.weather.com"
  "https://www.pinterest.com"
  "https://www.whatsapp.com"
  "https://www.airbnb.com"
  "https://www.github.com"
  "https://www.medium.com"
  "https://www.stackoverflow.com"
  "https://www.quora.com"
  "https://www.theguardian.com"
  "https://www.twitch.tv"
  "https://www.dropbox.com"
  "https://www.paypal.com"
  "https://www.salesforce.com"
  "https://www.shopify.com"
  "https://www.wikipedia.com"
  "https://www.adobe.com"
  "https://www.spotify.com"
  "https://www.nasa.gov"
  "https://www.forbes.com"
  "https://www.tesla.com"
  "https://www.msn.com"
  "https://www.aliexpress.com"
  "https://www.hulu.com"
  "https://www.discord.com"
  "https://www.craigslist.org"
  "https://www.booking.com"
  "https://www.squarespace.com"
  "https://www.tumblr.com"
  "https://www.vimeo.com"
  "https://www.soundcloud.com"
  "https://www.kickstarter.com"
  "https://www.ted.com"
  "https://www.wikipedia.fr"
  "https://www.lemonde.fr"
  "https://www.lefigaro.fr"
  "https://www.liberation.fr"
  "https://www.amazon.fr"
  "https://www.laposte.fr"
  "https://www.orange.fr"
  "https://www.ikea.com"
  "https://www.decathlon.fr"
  "https://www.leroymerlin.fr"
  "https://www.cdiscount.com"
  "https://www.fnac.com"
  "https://www.carrefour.fr"
  "https://www.ouest-france.fr"
  "https://www.france.tv"
```

```
  "https://www.tf1.fr"
  "https://www.m6.fr"
  "https://www.gouvernement.fr"
  "https://www.elysee.fr"
  "https://www.boursorama.com"
  "https://www.lci.fr"
  "https://www.public.fr"
  "https://www.voici.fr"
  "https://www.lequipe.fr"
  "https://www.fifa.com"
  "https://www.sport.fr"
  "https://www.footmercato.net"
  "https://www.uefa.com"
  "https://www.espn.com"
  "https://www.nba.com"
  "https://www.moto.com"
  "https://www.formula1.com"
  "https://www.ultimate-guitar.com"
  "https://www.guitarworld.com"
  "https://www.theverge.com"
  "https://www.techcrunch.com"
  "https://www.engadget.com"
  "https://www.wired.com"
  "https://www.xkcd.com"
  "https://www.slashdot.org"
  "https://www.reuters.com"
  "https://www.euronews.com"
  "https://www.aljazeera.com"
  "https://www.scientificamerican.com"
  "https://www.nationalgeographic.com"
  "https://www.disney.com"
)

# Nombre de sites
total_sites=${#sites[@]}

# Boucle pour envoyer 100 requêtes
for ((i=0; i<100; i++)); do
  # Sélectionne un site aléatoire dans la liste
  random_site=${sites[$RANDOM % $total_sites]}

  # Exécute la requête curl via le proxy Squid (modifiez l'adresse et le port du proxy si nécessaire)
  echo "Envoi de la requête vers $random_site via Squid"
  curl -s -o /dev/null -w "%{url_effective} -> %{http_code}\n" "$random_site" &
done

# Attendre que toutes les requêtes soient terminées
wait

echo "Test terminé. 100 requêtes envoyées."
```

SQUID et HTTPS

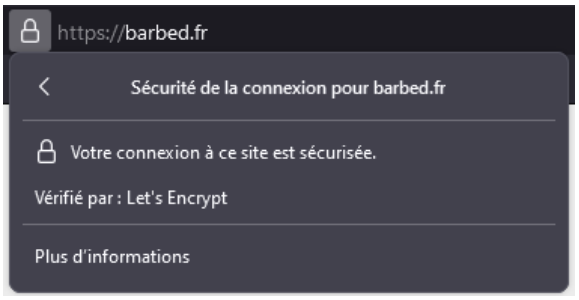
Tout ce qui transite en HTTPS est chiffré. Sans exceptions. Il devient donc compliqué de pouvoir faire du cache avec des fichiers qui transiteront de manière chiffrées et a chaque passage chiffré différemment.
En l'état SQUID va donc mettre en place du cache uniquement sur ce qui transite en HTTP et aujourd'hui, autant dire pas grand chose.

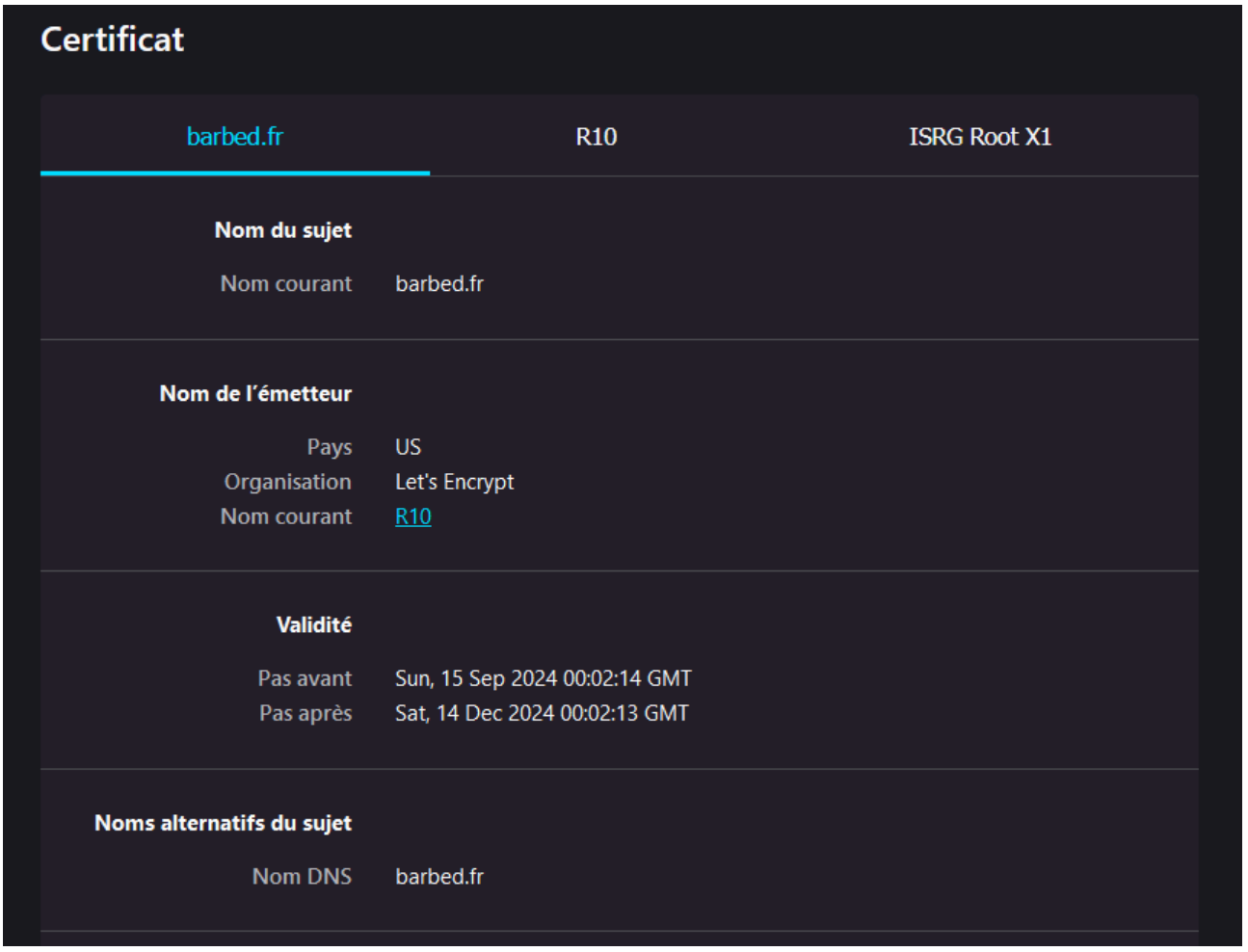
Pour permettre du cache sur du HTTPS il donc falloir être capable de déchiffrer du HTTPS pour voir le contenue des trames.

Faire ceci est un véritable enjeux de sécurité pour vos utilisateurs. Le proxy squid aura alors la possibilité de voir l'entièreté du contenue. Des pages visités, aux média mais aussi les mots de passe.

Il n'est aujourd'hui pas possible de "casser" a la volée le chiffrement de protocole tel que TLS2 ou TLS3 (du moins pas officiellement). Il va donc être nécessaire de mettre en place un subterfuge.

HTTPS gère des certificats et toutes vos machines ont des certificats de préinstallé pour permettre aux autorités certifications tierces d'etre reconnue par votre OS.
Exemple avec le domaine barbed.fr qui est certifié par Let's Encrypt.



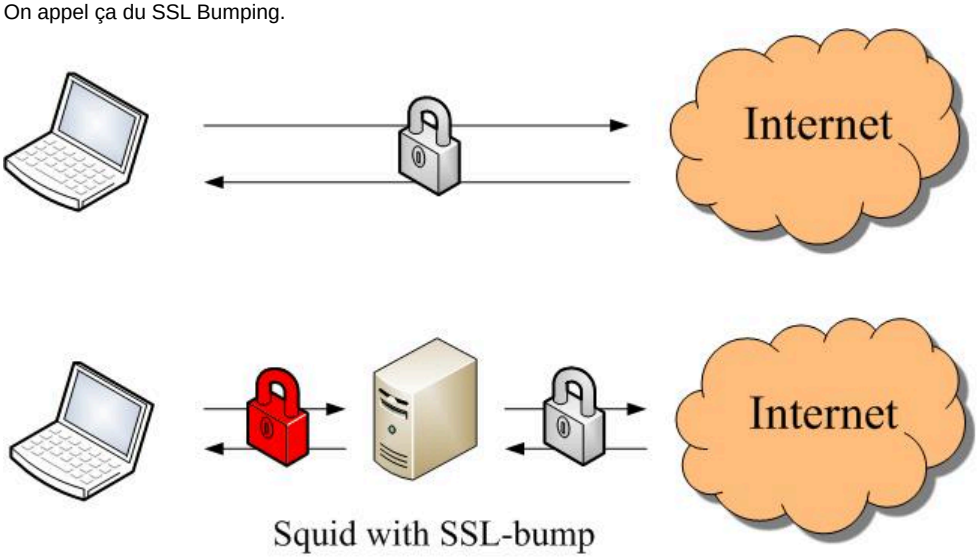


Pour que votre navigateur valide l'authenticité du certificat fournis par le serveur il faut qu'il ait un certificat fournis par l'organisme tier qui permettra de valider cryptographiquement l'authenticité de la chaine de certificat.

SQUID devrait casser cette chaine pour réussir a s'insérer au milieu.

Il y a plus simple. Votre entreprise va déployer sa propre autorité de certification, déployer les certificats dans vos navigateurs puis le proxy fera en sorte de dire a votre navigateur que la totalité des sites que vous visitez est validé par votre entreprise. Cela lui permettra de chiffrer la communication entre votre pc et lui. Analyser le trafic puis retransmettre aupres du vrai site externe en respectant la chaine de certificat d'origine.

On appel ça du SSL Bumping.



Ce n'est généralment pas une bonne pratique de sécurité de le mettre en place sauf si c'est dans un environnement qui le nécéssite pour d'autres raisons de sécurités.

Du coup à quoi ça sert d'avoir un SQUID aujourd'hui ?

Si on voit tout ce qui transite on va pouvoir le filtrer!

Projet SQUIDGuard

Vous pouvez partir du fichier Vagrant.
Vous devez installer un serveur web sur le serveur Squid qui affichera une page ou serons retourné les URL bloqués.

Vous installerez SQUIDGuard.
Vous allez créer une blacklist de domaine interdit:

Vous allez créer une blacklist d'URL spécifiques interdites:

Quand le client essayera d'accéder aux pages blacklistés il sera redirigé vers la page de blocage.

Livrable: Les fichiers de configuration que vous aurait mis en place.
La documentation de votre projet.
Des points bonus si vous fournissez le projet entièrement fonctionnel depuis un projet Vagrant.

Danger

Votre fichier comportera en en-tête votre Classe, le cours, nom et prénom.
Il sera également nommé comme suit: Classe - Cours- Nom Prenom.pdf
Exemple: `SN1-Windows Client-DUVAL Youenn.pdf`
Le non respect de cette consigne me coutant du temps et de l'agacement entrainera un malus de 8 points sur la note sur 80 soit 2 points sur la note sur 20.