Analysis of Dailymotion

Loucheur Benoît | 3613-16-00

I. Introduction

Ce document rapporte l'analyse effectuée sur le site `www.dailymotion.com` dans le cadre du cours LINGI1341 - Réseaux informatiques à l'Université catholique de Louvain. Cette étude est divisée en 4 parties : une analyse DNS, HTTP, TLS et finalement TCP.

Dailymotion est un site internet français proposant un service d'hébergement et de partage de vidéos en ligne.

II. Domain Name System

A. Serveurs de noms

L'analyse DNS du domaine *dailymotion.com* m'a permis de tirer les informations suivantes :

L'adresse IPv4 du serveur est 195.8.215.136, celui-ci est établi en France. L'IPv6 n'est actuellement pas disponible alors que ce protocole tend à remplacer tout doucement son ainé l'IPv4.

Le site dailymotion.com est supporté par 2 serveurs DNS :

- `a.dailymotion.com`
- `b.dailymotion.com`

Ces serveurs ne sont accessible que via l'ipv4. Le *Time to Live* du record A ainsi que des 2 serveurs de noms est de 43200 secondes soit 12 heures. Un TTL aussi long n'est pas vraiment étonnant pour un site de l'envergure de dailymotion. En effet, au plus le TTL est court, au plus la charge de données sera importante sur le DNS.

B. CDN

Content Delivery Network : depuis 2016, Dailymotion utilise un service du fournisseur CDN **Limelight** [1] nommé : Limelight Orchestrate. En effet, ils avaient besoin de ressources supplémentaires pour supporter leur croissance dans des régions comme l'Asie.

D'ailleurs Youtube utilisait ce même CDN avant leur rachat par Google.

C. CNAME

Les 2 serveurs n'utilisent pas de CNAME.

D. MX

Dailymotion possède un serveur de mail : `dailymotion-com.mail.protection.outlook.com` localisé au Royaume-Uni et joignable uniquement via l'IPv4.

E. TXT

Dailymotion possède plusieurs records TXT. Dans l'un des records, on remarque le champs `v=spf1`, celui-ci permet de vérifier le nom de domaine de l'expéditeur lors de l'envoi d'un mail pour éviter le vol d'identité [2]. Il dispose aussi du champs `facebook-domain-verification` et `google-site-verification`. En effet, il est possible de se connecter sur Dailymotion via Facebook ou Google. Ces records attestent que la propriété du domaine est confirmée par Facebook et Google [3].

III. Requêtes HTTP

L'un des composants majeurs du World Wide Web est le protocole HTTP pour HyperText Transfer Protocol. HTTP est un protocole d'échange entre un client et un serveur. Le client envoie par exemple une requête et le serveur renvoie une réponse. Dans cette section, quelques requêtes et réponses seront analysées. En premier lieu, les différents domaines recontrés lors de la navigation sur `dailymotion.com` et les différentes ressources échangées.

A. Domaines

Voici une brève liste des domaines rencontrés lors du chargement de la page principale de `dailymotion.com`

- `connect-facebook.com` | Fichier JavaScript
- `proxy-XXX.ix7.dailymotion.com` | Fichier .mp2t qui est un format spécifique pour le vidéo streaming
- `static1.dmcdn.net` |Fichier JavaScript
- `s1-ssl.dmcdn.net` | Fichier jpeg : bannières des vidéos
- `s2-ssl.dmcdn.net` | Fichier jpeg : bannières des vidéos
- `ajax.googleapis.com` | Fichier JavaScript

B. Port utilisés

Le seul port utilisé est le port **443**, cela est dû au fait que Dailymotion est accessible en HTTPS. Lorsqu'on essaye d'accéder en HTTP, celui-ci renvoie la réponse **307 Temporary Redirect** et force à établir une connection TLS et donc avec le port 443.

C. Pixel tracking

En plus de tous les domaines cités ci-dessus, d'autres ont comme but de traquer les utilisateurs du site [4]. Voici une liste de ceux présent sur la page principale de `dailymotion.com` :

- `google-analytics.com` -> **Google**
- `sb.scorecardresearch.com` -> **ComScore**

- `ads.stickyadstv.com` -> **Comcast**
- `a.tribalfusion.com` -> **Exponential**
- `stats.g.doubleclick.net` -> **Google**
- `sync.tidaltv.com` -> **Groupon, Sales-force, Valhalla**
- `t.teads.tv` -> **Altice**

La plupart de ces entreprises hormis Google, sont toutes spécialisées dans la publicité de masse.

Le point commun entre ces domaines est qu'ils utilisent le *pixel tracking*. Cette technique consiste à faire télécharger à l'utilisateur du site un pixel espion de taille 1x1, ici ils sont sous le format `gif`. Ces domaines utilisent cette stratégie pour récolter de l'information sur l'activité des utilisateurs. Le pixel a en réalité aucune importance graphique, il est cependant utile pour tromper le navigateur, afin de lui faire croire qu'il est une ressource utile.

Voici une analyse de requête envoyée par doubleclick.

TABLE I: Requête

Paramètre	Contenu
Accept	*/*
Accept-Encoding	gzip, deflate, br
Accept-Language	fr,fr-FR;q=0.8,en-US;q=0.5,en;q=0.3
Connection	keep-alive
Cookie	UID=1CB92a12253a1492a0066471544618735; UIDR=1544618735; _gaḠA1.2.1526710939.1544657544; _gidḠA1.2.1370511359.1544657544
Host	sb.scorecardresearch.com
Referer	https://www.dailymotion.com/be
User-Agent	Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:64.0) Gecko/20100101 Firefox/64.0

TABLE II: Réponse

Paramètre	Contenu
Cache-Control	private, no-cache, no-cache=Set-Cookie, no-store, proxy-revalidate
Connection	keep-alive
Content-Length	43
Dates	Thu, 13 Dec 2018 17:49:58 GMT
Expires	Mon, 01 Jan 1990 00:00:00 GMT
Pragma	no-cache

Je vais à présent analyser quelques paramètres.

Cache-Control [5] est un en-tête définissant la politique en matière de cache des contenus autant dans les requêtes que les réponses. Elle peut avoir plusieurs valeurs :

- **no-cache** : indique de renvoyer systématiquement la requête au serveur et ne servir une éventuelle version en cache que dans le cas où le serveur le demande.
- **no-store** : la requête ou les données ne devraient jamais être mises en cache.
- **private** : indique que la réponse ne doit être mise en cache que pour un utilisateur donné et ne doit donc pas être mise en cache par un cache partagé.

Pragma [6] : l'option **no-cache** oblige l'utilisateur à télécharger des fichiers. Vu le site faisant objet comme analyse, il est assez normal de souhaiter des mises à jours régulièrement.

Expire [7] : date et heure à partir de laquelle l'option est considérée comme périmée. En d'autres termes, à partir de cette date, l'utilisateur serait obliger à ne plus utiliser sa mémoire cache pour cet élément particulier. La date ayant été dépassée, la requête ne sera jamais mise en cache. Cette technique n'est d'ailleurs pas anodine et est utilisée dans une majorité des sites internet.

User-Agent [7] : contient des informations à propos du navigateur client. Ce champ contient une liste de navigateur.

TABLE III: Réponse

Navigateur	Contenu
Google Chrome	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/70.0.3538.110 Safari/537.36
Mozilla Firefox	Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:64.0) Gecko/20100101 Firefox/64.0
Internet explorer	Mozilla/5.0 (Windows NT 10.0; WOW64; Trident/7.0; rv:11.0) like Gecko

Pour l'expliquer, il faut se rappeler la "guerre" [8] entre tous les navigateurs. A l'époque, NetScape ("père" de Mozilla et ancien navigateur dominant du marché) était en guerre avec Internet explorer. Ce dernier voulait éviter d'être rejeté par des sites qui rejetaient des navigateurs autre que Netscape. Les navigateurs actuels simulent donc qu'ils sont Mozilla Firefox. Opera, à ses début, a d'ailleurs essayé de ne pas inclure Mozilla comme *User-Agent*, pour des raisons assez pointues, ils ont décidé à partir d'`Opera 9` d'inclure finalement Mozilla comme User-Agent.

D. Réponse de Google

Dailymotion utilise des ressources des serveurs de Google, tel que des fichiers JavaScript et HTML.

Un fichier JavaScript est téléchargé depuis le service `googletagservice.com` et le header de la réponse contient quelques champs intéressants :

- `alt-svc: quic="googleads.g.doubleclick etc`. Quic [9] est un protocol expérimental de Google basé sur UDP qui a pour but de remplacer TCP car plus puissant que ce dernier.
- `x-content-type-options: nosniff` [10] : cette option non standard empêche le navigateur de faire du "MIME-type sniffing" qui lui permet de déduire le format des fichiers en inspectant le flux de bytes. `x-xss-protection:1; mode=block` [11]: cette option non standard est un filtre pour empêcher le navigateur de charger la page s'il détecte un "cross-site scripting".

Ces en-têtes sont assez communes à Google et se retrouvent dans la plupart de leurs requêtes.

E. Type de requêtes

Les requêtes HTTP provenant de Dailymotion sont en HTTP 1.1. Les requêtes HTTP provenant de Google sont en HTTP 2.0 comme l'atteste l'en-tête : `X-firefox-Spdy: h2`.

F. Type de ressources

L'analyse du chargement rapide de la page d'accueil a été effectuée avec et sans AdBlock qui est un bloqueur de publicité. Je ne laissais pas le temps à la vidéo de la page principale de charger, car celle-ci générait beaucoup trop de requêtes rapidement. La différence fondamentale se situe au niveau des images. Avec AdBlock, on en charge très peu. En effet, comme expliqué plus haut, Dailymotion utilise la technique du pixel-tracking. AdBlock permet entre autre de bloquer ces fameux pixels.

TABLE IV: AVEC | SANS -> AdBlock

Requests Type	Requets	Percentage
script	33	54%
other	7	11%
link	6	9.8%
css	5	8.2%
iframe	3	4.9%
fetch	3	4.9%
img	2	3.3%
xmlhttprequest	2	3.3%

Requests Type	Requets	Percentage
script	34	49%
img	9	13%
other	7	10%
link	6	8.7%
css	5	7.2%
iframe	3	4.3%
fetch	3	4.3%
xmlhttprequest	2	2.9%

En fait, Dailymotion ne compte pas réellement beaucoup de pubs intempestives. Les publicités se retrouvent plutôt dans le lecteur vidéo. De ce fait, lors du chargement de la page principale, utiliser AdBlock ne bloquera que le pixel tracking sauf celui provenant de `google-analytics.com`.

G. Cookie

Je vais me concentrer sur les cookies renvoyés par Dailymotion. Pour effectuer cette analyse, j'ai utilisé plusieurs sites internet tel que `www.webcookies.org/`, `www.cookie-checker.com` ainsi qu'un addon firefox : *Cookies Manager*.
Dans le cas où rien n'est précisé, le cookie est de type persistant à fortiori il sera de session. **Persistent cookie** : stocké sur le disque dur et reste actif tant que son temps de vie n'est pas expiré ou que l'utilisateur ne le supprime pas. **Session cookie** : disparaît lorsque l'utilisateur ferme son navigateur.

- **d7s_spc** : il a une durée de 25 minutes avec comme valeur un entier représentant un compteur. A chaque fois qu'on change de page, il s'incrémente de 2.
- **damd** : cookie de session. Il est constitué d'une chaine de caractères liée à la vidéo que l'on regarde.
- **euconsent** : (**A vérifier !**) Il a une durée de 5h30 avec comme valeur une longue chaine de caractères. En me renseignant, j'ai découvert 2 hypothèses sur sa fonctionnalité : Il représente le consentement des fournisseurs qui participent au cadre de consentement de l'OER. L'OER "Open Educational Resources" ou REL "Ressources Educatives Libres" représente donc le contenu mis à disposition librement et gratuitement.
  Ou alors, il serait plutôt lié à la directive européenne qui demande le consentement des utilisateurs pour les cookies.
- **lang** : cookie de session. Ex: "fr_FR" ou "en_US". Il représente la langue d'affichage du site. J'ai réussi à le modifier en changeant manuellement la langue du navigateur.
- **__ga | __gid | __gat** : ils ont une durées d'un peu plus d'un an. Ils proviennent tous de google-analytics.
- **__ gads** : il a une durée de vie d'un an. Il provient de *Doubleclick* qui appartient lui-même à google.
- **client-token** : il a une durée de vie de 8 heures. Il peut être représenté comme un ID qui reste fixe lorsque je navigue sur le site. Cependant, dès que je me connecte soit via Google ou Facebook. Ce cookie disparait pour être replacé par **access-token** mais avec la même clé.
- **lri** : cookie de session. Il n'existe que lorsqu'on se connecte via Google. Il contient notre mail,un id, le pseudo et notre image en URL.

Il en existe encore bien d'autres, cependant ils sont beaucoup plus complexes à analyser.

IV. TLS ET DNSSEC

A. TLS

Le Transport Layer Security (TLS) est un protocole de sécurité sur la couche de transport. Il permet de sécuriser les échanges sur internet. `dailymotion.com` supporte TLS 1.0 1.1 et 1.2 et préfère utiliser ce dernier.

B. Cipher

Le site `dailymotion` utilise le Cipher suivant : `ECDHE-RSA-AES256-GCM-SHA384` [12].

- `ECDHE` : clé de Diffie-Hellman [13] basée sur les courbes elliptique. Cette caractéristique de cipher utilisé par Dailymotion garantit [14] la Perfect Forward Secrecy même si un attaquant découvre la clé privée de notre correspondant, ça ne compromet pas la confidentialité des échanges passés.
- `RSA` : Rivest-Shamir-Adelman. Algorithme de cryptographie asymétrique. Il y a une paire de clé publique pour chiffrer et une clé privée pour déchiffrer.
- `ASE` : Advanced Encryption Standard. Algorithme de chiffrement symétrique.
- `GCM` : Galoi/Counter-Mode est un mode d'opération de chiffrement par bloc en cryptographie symétrique.
- `SHA` : Secure Hash Algorithm, contient plusieurs fonctions de hashages. Ici on dispose de `SHA-384` qui produit un résultat sur 384 bits.

Voici le Cipher entier fourni par `www.ssllabs.com` : `TLS_ECDHE_RSA_WITH_AES_256_GCM_SHA384`. Lors de l'établissement de la connection, on retrouve le fameux handshake. Voici un résumé de l'échange qui se passe sur Dailymotion :

1) Le serveur fournit en premier un certificat contenant une clé publique `RSA` utilisée pour l'identification.
2) L'échange des clés se fera via `ECDHE`.
3) `AES-GCM` est utilisé comme chiffrement symétrique avec une clé de 256 bits.
4) On utilise `SHA-384` pour fournir la fonction de hashage sur 384 bits.

Dailymotion supporte le **Secure Renegotiation** [15] qui permet de refaire un nouveau handshake à l'intérieur d'une session déjà sécurisée.

Dailymotion supporte aussi le **Downgrade Attack Prevention** [16] qui empêche des attaques visant à forcer la session à abandonner le cryptage et donc abaisser la sécurité de la connexion.
Dailymotion supporte aussi le **HSTS : HTTP Strict Transport Security** [17] qui remplace tous les liens HTTP par un lien HTTPS. Plus précisement, le Strict Transport Security force les connexions HTTPS avec chiffrement sans prendre en compte les appels de scripts pour charger des ressources HTTP. Voici donc le fameux liens avec la redirection discutée dans la section B de la partie HTTP.
Dailymotion supporte aussi le **OCSP stapling** [18] qui permet au site de fournir un OCSP : Online Certificate Status Protocol qui soulage le client de devoir demander au Certificate Authorithies. Au départ, cette méthode était très coûteuse pour le serveur, mais depuis l'extension **RFC 6961**, l'ajout de réponses OCSP multiples est possible.
Finalement, au niveau du **Perfect Forward Secrecy** [19], le site nous indique *"with modern browsers"*. Ce qui signifie que le serveurs priorise les échanges avec les navigateurs disposant du Forward Secrecy. Cette algorithme est d'ailleurs très coûteux.

C. DNSSEC

Le **DNSSEC** [20] Domain Name System Security Extensions est un protocole permettant de résoudre des problèmes de sécurité liés au protocole DNS. L'analyse est effectuée via les résultats de `www.dnsviz.net`
Lorsqu'on recherche le domaine `dailymotion.com`, on va tout d'abord interroger le domaine ".com" de manière sécurisée. Par contre, une fois qu'on interrogera le domaine `dailymotion.com`, la communication n'est plus du tout sécurisée.

V. TCP

Dans cette section, je vais décrire les différentes caractéristiques du protocole TCP : Transmission Control Protocol. Cette analyse est uniquement effectuée sur le domaine principal de Dailymotion, je ne prends donc pas en compte les domaines secondaires (publicités, etc).

A. Nombre de connexions simultanée

Le nombre de connexions maximum simultanées que j'ai pu observer est de 2 sur Google Chrome et d'1 sur Firefox.

B. Options TCP

Voici la configuration que j'ai pu récupérer grâce à l'outil Wireshark :

- **Window Scale** :
  - Mon ordinateur propose **8** soit multiplié par **256**
  - Dailymotion répond et propose **11** soit multiplié par **2048**
- **Maximum Segment Size** :
  - Mon ordinateur propose **1460** bytes.
  - Dailymotion répond et propose **1460** bytes.
- L'**aquitement sélectif** est permis permettant au receveur de préciser les paquets qu'il a reçu dans son ACK pour éviter le renvoi de tous les paquets en cas de perte.
- **Timestamp** : cette option ne semble pas être supportée. L'option **TCP Fast Open** [21] est absente. Du fait qu'il y a très peu de connexions simultanées, il n'est pas nécessaire d'accroître la vitesse d'ouverture.

Depuis mon domicile à Arlon, le RTT est en moyenne entre 25 et 30ms, ce résultat correspond avec RIPE Atlas qui m'indique un RTT entre 20 et 30 pour sa probe située en Belgique.

C. Happy Eye balls

Comme expliqué dans le section DNS, Dailymotion ne supporte pas encore l'IPv6, l'utilisation de Happy Eye balls n'a donc pas de sens car Dailymotion n'accepte que l'IPv4.

D. Ouverture de connexion

Etant dans une implémentation TCP, l'ouverture se fait via le three way handshake

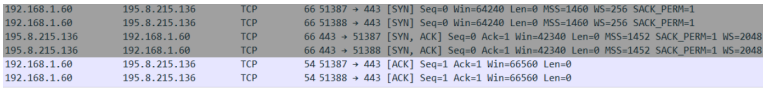

Fig. 1: Ouverture de la connexion

E. Fermeture de connexion

Les connections TCP s'achèvent à l'aide un segment [FIN,ACK] envoyé soit par le client ou le serveur. En cas d'interruption, le segment [RST,ACK] est envoyé par le client pour interrompre les connections ouvertes.

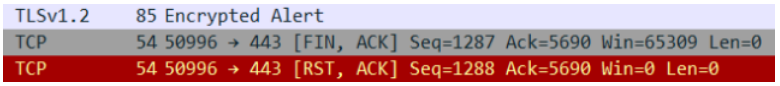

Fig. 2: Fermeture de la connexion

VI. Conclusion

Ainsi se conclut le rapport d'analyse du site `dailymotion.com`. L'analyse TCP ne fut pas très étonnante pour un site de cette envergure. La partie DNS est plus intéressante, car le site ne dispose d'une seule adresse IPv4.
Dailymotion repose sur les services de Google pour des fichiers Javascript et pour une analyse statistique.
L'analyse HTTP a été aussi très interessante au niveau des cookies.
Ce projet m'a aussi permis d'apprendre de nombreuses choses sur le fonctionnement pratique d'un site. On passe tous la plupart de notre temps sur le Web, et on le connaît à peine. Lorsqu'on navigue naïvement sur internet, on se retrouve sur la partie émergée de l'iceberg.

REFERENCES

[1] https://www.cdnplanet.com/cdns/limelight/
[2] Mxtoolbox | https://mxtoolbox.com/problem/spf/txt-record
[3] Google Support | https://support.google.com/a/answer/183895?hl=fr
[4] Quora | https://www.quora.com/How-does-a-tracking-pixel-work
[5] https://developer.mozilla.org/fr/docs/Web/HTTP/Headers/Cache-Control
[6] StackOverFlow | https://stackoverflow.com/questions/10314174/differencebetween-pragma-and-cache-control-headers
[7] Flavioscope | https://flaviocopes.com/cookies/
[8] https://stackoverflow.com/questions/1114254/why-do-all-browsers-useragents- start
[9] Apnic | https://blog.apnic.net/2018/05/15/how-much-of-the-internet-isusing- quic/
[10] StackOverFlow | https://stackoverflow.com/questions/18337630/what-isx- content-type-options-nosniff
[11] Keycdn | https://www.keycdn.com/blog/x-xss-protection
[12] https://security.stackexchange.com/questions/166063/tls-rsa-vs-tlsecdhe-rsa-vs-static-dh
[13] https://security.stackexchange.com/questions/33233/ecdh-and-forwardsecrecy
[14] https://www.namecheap.com/support/knowledgebase/article.aspx/9653/38/howto-check-whether-the-server-supports-forward-secrecy
[15] Rescorla, Ray, Dispensa, Oskov | https://tools.ietf.org/html/rfc5746
[16] Hayden Blauzvern | https://p16.praetorian.com/blog/man-in-the-middletls- ssl-protocol-downgrade-attack
[17] https://fr.wikipedia.org/wiki/HTTP_Strict_Transport_Security - withmozilla
[18] Keycdn | https://www.keycdn.com/support/ocsp-stapling
[19] https://vincent.bernat.ch/fr/blog/2011-ssl-perfect-forward-secrecy
[20] Cloudflare | https://www.cloudflare.com/dns/dnssec/how-dnssec-works/
[21] Keycdn | https://www.keycdn.com/support/tcp-fast-open