Analyse de site web

Gildas Mulders
NOMA : 24601600

***Résumé*—Ce rapport a été réalisé dans le cadre du cours "LINGI1341 : Réseaux Informatiques" et a pour but d'analyser un site web. Le site web qui est analysé dans ce rapport est "dailymotion.com".**

I. Introduction

Ce rapport analyse le site "dailymotion.com" selon plusieurs aspects, à savoir DNS, HTTP, TLS et TCP. Les outils en ligne utilisés et les sites consultés pour ces analyses seront renseignés à la fin de ce document.

II. Analyse DNS

A. Analyse générale

Par la commande dig, on peut voir que le site dailymotion repose sur deux serveurs qui possèdent une adresse IPv4 mais pas d'IPv6 ni de CNAME.

- dailymotion.com 195.8.215.136
- a.dailymotion.com 195.8.214.1
- b.dailymotion.com 195.8.214.2

Ces serveurs et tous leurs records ont un TTL de 43200 secondes, soit 12 heures. Ce TTL ne varie pas lorsqu'on fait une requête dig depuis le réseau de l'UCL ou depuis un resolver DNS public comme celui proposé par google.
L'outil en ligne CDN finder montre que dailymotion utilise Limelight comme CDN. En pratique, Limelight possède plus de 80 points de présence sur la planète. Chaque requête qu'un utilisateur fait au site est donc redirigée vers le point de présence le plus proche. Et chaque point de présence a soit les données demandées en cache, soit il les demande aux serveurs a.dailymotion.com ou b.dailymotion.com. Ce système permet de réduire considérablement les temps de latences qui auraient été très élevés pour des utilisateurs très éloignés du serveur.

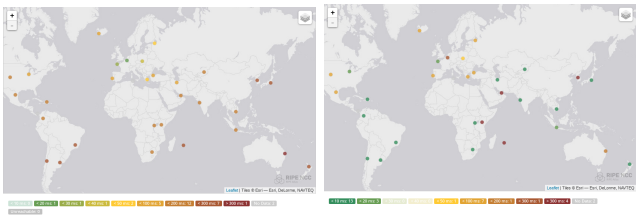

Figure 1. Ping et DNS A de dailymotion à travers le monde

Sur la figure 1 obtenue grâce au site RIPE Atlas, à gauche, on voit la latence due à un ping au serveur de dailymotion tandis qu'à droite, on voit la latence due à une requête au DNS de l'IPv4 du serveur, les points verts représentant les latences les moins importantes et les rouges représentant des latences plus longues. La première image montre que le serveur de dailymotion est basé en France (il s'agit bien d'une entreprise française), mais la deuxième montre que le CDN permet d'avoir une latence très acceptable où qu'on se situe dans le monde.

B. Autres enregistrements DNS

La commande dig peut également montrer d'autres enregistrements intéressants. Ainsi, dailymotion possède entre autre un enregistrement MX correspondant à une messagerie outlook "dailymotion-com.mail.protection.outlook.com". Dailymotion possède également un certain nombre de résultats pour des enregistrements TXT. Parmi ceux-ci, on peut noter la présence des éléments suivant :

- "facebook-domain-verification=12wx..."
- "google-site-verification=jb-qAE0Qy..."

Ceux-ci permettent de valider et vérifier la propriété du site. Il existe également un record de type SOA, dont l'analyse montre la présence d'une adresse email d'administrateur "admins@dailymotion.com". Cette analyse nous apprend également que le serveur maitre de la zone DNS est a.dailymotion.com, on en déduit que b.dailymotion.com fait partie du même cluster que le serveur maitre, mais est un serveur secondaire.

III. Analyse HTTP

A. HTTP en général

Le port TCP utilisé est le 443 pour toutes les requêtes faites sur le site web. Quant à la version d'HTTP utilisée, la plupart des requêtes utilisent HTTP/2.0 à l'exception de la première requête, qui sera analysée plus en détail dans le sous-chapitre "En-têtes" et des trackers qui seront traités dans le sous-chapitre "Domaines".

B. Domaines

Le site webpagetest.org fournit des analyses http tels que les graphes de la figure 2. Ces graphes représentes le nombres de requêtes et le nombre de bytes que dailymotion tire d'autres domaines. Dans le graphe qui montre les différents domaines selon le nombres de bytes échangés fait ressortir :

- "static1.dmcdn.net" en rose(57.6%).
- "s1-ssl.dmcdn.net" en bleu foncé(13.1%).
- "s2-ssl.dmcdn.net" en mauve(4.7%).

En regardant le site dailymotion avec la console développeur présente sur plusieurs navigateurs comme Firefox ou Chrome, on peut voir que "static1.dmcdn.net" est responsable de toute la part constante, ou "statique" du site, comme l'image de fond, les logos,.. D'un autre côté les domaines "..-ssl.dmcdn.net" sont responsables des videos et images qui

défilent sur le fil d'actualité de dailymotion. Tous les web objects de ces domaines ont un header "Expires" fixé en l'an 2028, ce qui correspond au fait que ces objets "n'expirent jamais". En conséquence, tous ces éléments sont mis en cache.

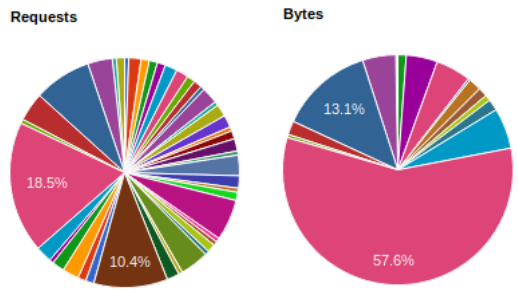

FIGURE 2. Domaines contactés par dailymotion

Dans le graphe des requests, et dans la liste des domaines représentés par ce graphe, on retrouve beaucoup de trackers. Notamment :

- ads.stickyadstv.com (en brun sur le graphe (10.4%))
- public-prod-dspcookiematching.dmxleo.com (mauve (5.8%))
- ib.adnxs.com (4%)
- match.adsrvr.org
- id5-sync.com
- ...

On observe que dailymotion utilise le cookie matching, concept expliqué à la référence [7], qui consiste à regrouper les différents trackers pour mieux cibler les utilisateurs.

C. En-têtes

La première requête faite lorsqu'on navigue vers une page web de dailymotion est toujours de type html et vient du domaine dailymotion.com. En pratique, ce fichier html contient des scripts qui lancent les requêtes suivantes et construisent ainsi la page. Il ne sera jamais mis en cache grâce à la présence du header "cache-control : no-cache". Cette première requête contient en plus des headers standards, une série d'en-têtes personnalisées :

- X-DM-BackNode : web-285.adm.dc3.dailymotion.com :80
- X-DM-BackNode-Response-Time : 34
- X-DM-LB-IP : 195.8.215.136
- X-DM-LB-Name : lb-08
- X-Dm-Page : us.html.feed
- X-Protected-By : Sqreen

On peut supposer que ces headers font référence à un "load balancer" [1]. Cela signifierait que notre requête est envoyée vers l'IPv4 195.8.215.136 [2]. Ce serveur serait donc un load balancer qui redistribue la charge vers des serveurs internes, dans notre cas, le serveur "web-285.adm.dc3.dailymotion.com :80" (référencé dans le header "X-DM-BackNode"). On remarque que lorsqu'on rafraichit la page, notre demande est traitée par un autre "BackNode".

On notera également que dailymotion utilise Sqreen, décrit à la référence [8], pour sécuriser leur application web. On retrouve des en-têtes similaires pour les réponses provenant de "s1-ssl.dmcdn.net" et "s2-ssl.dmcdn.net".

Lorsqu'on fait une recherche puis qu'on lit une video, on constate que dailymotion affiche des publicités en faisant des requêtes au domaine "e.deployads.com". Il y a également beaucoup de requêtes aux serveurs "s1-ssl.dmcdn.net" et "s2-ssl.dmcdn.net" pour afficher toutes les images de preview des videos. Mais surtout, les requêtes qu'on retrouve le plus correspondent à des trackers. Mis à part ces éléments, la circulation sur le site ne fait pas apparaitre de nouvelles requêtes ou de nouveaux headers particuliers.

D. Cookies

Lorsqu'on se connecte avec gmail à dailymotion, une requête est envoyée à un iFrame, ce qui sert à inclure un site dans un autre. Cette requête utilise des cookies pour établir cette connexion au compte. Ceux-ci sont illisibles par un humain, on peut supposer qu'ils sont hashés ou qu'il s'agit de chaine de caractères qui n'ont un sens que pour le serveur. Beaucoup de cookies chiffrés sont également utilisés lors de la première requête au site web, à l'exception d'un champ en clair, le "consent", qui rend compte que l'utilisateur a accepté l'utilisation des cookies. En revanche, on peut voir que les cookies envoyés au site lors du premier accès à celui-ci (le navigateur est dit "vierge" parce qu'on a vidé les cookies) ne comportent pas autant d'informations. En comparant les cookies des deux requêtes, on remarque plusieurs différences, dont les principales :

- Absence de consent lors du premier accès : sur le site, un message demande d'accepter l'utilisation des cookies, en cliquant sur le bouton "j'accepte", dailymotion envoie une requête à l'URL "https://www.dailymotion.com/.well-known/pubvendors.json" avec le consent dans les cookies de la requête. Pubvendors.json est un fichier consulté par un certain nombre de publicitaires et qui spécifie ce que l'utilisateur a accepté qu'on fasse de ses données.
- Client_token à la place de Access_token : cela est relatif au fait qu'on se connecte avec un compte ou pas. Lorsqu'on se connecte avec un compte, une requête est envoyée avec le username et le password en chiffré dans les paramètres, et la réponse contient un Access_token qui vient alors remplacer le Client_token dans les requêtes à venir.

Après la connexion à un compte sur dailymotion, une série de requêtes sont envoyées à "s1-ssl.dmcdn.net" pour afficher des suggestions de videos sans doute plus ciblées pour la personne qui s'est connectée.

IV. ANALYSE TLS

Les requêtes de dailymotion sont sécurisées par un protocole TLSv1.2 à l'exception de celles adressées à facebook et à un certain domaine responsable de cookie

1. On suppose que l'abréviation "LB" signifie Load Balancer
2. Cette IP correspond à dailymotion.com, voir Chapitre II

matching, qui eux, sont sécurisés par un protocole TLSv1.3. Une seule requête se démarque, sur la page de fil d'actualité. Il s'agit d'une requête à une url apparentée au domaine ads.stickyadstv.com, et ses paramètres semblent indiquer qu'elle vise juste à obtenir une image d'un pixel, soit probablement un tracker.
La cipher suite préférée par dailymotion est "TLS_ECDHE_RSA_WITH_AES_256_GCM_SHA384"
ce qui peut être compris comme suit :

- TLS est le protocole utilisé.
- ECDHE_RSA désigne la façon dont la clé de chiffrement sera échangée entre le client et le serveur. ECDHE_RSA garantit la "forward secrecy", ce qui signifie que si une clé est compromise à l'avenir, cela ne compromettra pas la clé actuelle.
- WITH_AES_256 représente l'algorithme de chiffrement des données utilisé. AES est un algorithme développé par deux cryptographes belges. C'est un algorithme dit "block cipher", ce qui signifie que les données sont chiffrées par blocks de bits. AES crypte les données par groupes de 128 bits et le fait que AES soit suivi par 256 signifie que la clé de chiffrement a une taille de 256 bits, ce qui est la taille maximale de clé pour AES. Enfin, AES est un algorithme à clé symétrique, ce qui permet aussi bien de chiffrer et déchiffrer les données avec la même clé.
- GCM est le "block cipher mode of operation", soit l'algorithme qui permet de chiffrer un ensemble bien plus important de données que la limite d'un block du block cipher.
- SHA384 est une fonction de hash utilisée pour vérifier l'intégrité des données.

Dans la console développeur et dans le site SSL Labs, accessible depuis la référence [9], on observe que dailymotion supporte HTTP Strict Transport Security. Cela signifie que l'attaque consistant à enlever le "s" dans le https de l'url du site est bloquée, comme cela est expliqué dans l'example de l'analyse de uclouvain.be présente sur le blog [11].
La commande openssl renseigne que la clé générée par RSA a une longueur de 2048 bits. Elle montre également la certificate chain comme on peut le voir sur la figure 3. Le certificat SSL, preuve de la sécurité du site a été délivré par digicert [3]. On sait cela car digicert est à l'origine du root certificate de la certificate chain. Avant d'arriver au certificat lié au domaine de dailymotion lui-même, un certificat intermédiaire est délivré. On peut constater la présence de la mention "SHA2" dans ce certificat intermédiaire, SHA2 correspond à une série de fonctions de hash contenant entre autre la fonction SHA384 utilisée par dailymotion. Ces résultats sont le fruit d'un secure handshake que la commande openssl effectue avec le serveur de dailymotion. L'option -reconnect de openssl montre que ce certificat ainsi que ce tous les résulats du secure handshake, comme la cipher suite et la plupart des informations traitées dans ce chapitre sont gardés en cache, car en se déconnectant et en se reconnectant au serveur, on obtient les mêmes informations qu'à la première connexion sans avoir à effectuer plus d'une fois le secure handshake.

3. https://www.digicert.com/

```
---
Certificate chain
 0 s:/C=FR/L=Paris/O=Dailymotion SA/CN=*.dailymotion.com
   i:/C=US/O=DigiCert Inc/OU=www.digicert.com/CN=DigiCert SHA2 High Assurance Server CA
 1 s:/C=US/O=DigiCert Inc/OU=www.digicert.com/CN=DigiCert SHA2 High Assurance Server CA
   i:/C=US/O=DigiCert Inc/OU=www.digicert.com/CN=DigiCert High Assurance EV Root CA
---
```

FIGURE 3. Certificate Chain

Le site SSL Labs nous apprend également que dailymotion pratique le OCSP stapling, c'est à dire que lors du premier handshake, il fournit une garantie sur la présence d'un certificat pour le site, signée et datée par le Digicert. Le client qui essaye d'accéder à dailymotion a donc directement cette information sans devoir adresser d'autre requête pour avoir le certificat du site. Enfin, le dernier élément intéressant appris par SSL Labs est que dailymotion sait faire face aux Downgrade attacks. Par défaut, lorsqu'un handshake échoue, les clients vont accéder aux serveurs en utilisant une version antérieure du protocole TLS. Certains profitent de cela en exploitant les failles plus évidentes dans ces versions antérieures. Une downgrade attack consiste à ce qu'une tierce personne utilise cela pour détériorer la version du protocole de sécurité actif entre le client et le serveur. Dailymotion sait prévenir ces attaques grâce à la définition de SCSV, "Signaling Cipher Suite Value", par laquelle le client peut indiquer au serveur que l'échec de la connexion n'est pas du au fait qu'il ne supporte pas la version utilisée de TLS.

V. ANALYSE TCP

A. Options TCP

En analysant les paquets échangés à l'aide de wireshark, on constate qu'il y a deux paquets [SYN] destinés à dailymotion sans paquet de FIN entre les deux, il y a donc deux connexions au site en parallèle. Ces paquets et les [SYN, ACK] qui leur correspondent montrent aussi une série d'option qui sont négociées entre l'ordinateur et le serveur de dailymotion. La figure 4 reprend les options du paquet [SYN] à gauche, envoyé de l'ordinateur à dailymotion, et les options du paquet [SYN, ACK] à droite, qui est la réponse de dailymotion à l'ordinateur.

```
Maximum segment size: 1460 bytes      Maximum segment size: 1460 bytes
SACK permitted                        SACK permitted
Timestamps: TSval 877721903, TSecr 0  Timestamps: TSval 852600589, TSecr 877721903
No-Operation (NOP)                    No-Operation (NOP)
Window scale: 7 (multiply by 128)     Window scale: 11 (multiply by 2048)
```

FIGURE 4. Options TCP

La présence de window scale correspond au fait qu'une version antérieure du protocole TCP était limitée à une window size de 65535 bytes, cette valeur correspondant à la valeur maximum représentée par 16 bits, les window size ont alors été implémentés sur 16 bits. Mais cette valeur maximum a vite été considérée comme limitée, et pour ne pas perdre la compatibilité avec les implémentations précédentes de TCP, plutôt que d'augmenter cette représentation de 16 à 32 bits,

on a eu recours à un arguments supplémentaire dans l'en-tête TCP : le window scale. Le window scale désigne le nombre de fois qu'on doit décaler la window size vers la gauche pour avoir la réelle window size, qui peut alors être bien plus grande, puisque chaque décalage la fait doubler. Après un court calcul, on constate donc que la window size réelle de l'ordinateur est de $29200 * 2^7 = 3737600$ bytes alors que celle de dailymotion est de $43440 * 2^{11} = 88965120$ bytes. La window du serveur de dailymotion est donc considérablement plus grande que celle de l'ordinateur.
La figure 4 montre également que l'ordinateur et dailymotion s'accordent pour une MSS de 1460 bytes et qu'ils pratiquent les Selective acknowledgements, ce qui permet au serveur de ne renvoyer que les paquets qui n'ont pas été reçus par le client au lieu de renvoyer tous les paquets à partir de celui qui a été perdu. Enfin, le Timestamp est composé de deux parties, le TSVal est rempli avec la valeur de clock de celui qui envoie le paquet et le TSecr, présent uniquement dans des paquets de type ACK, reprend la TSVal d'un paquet précédemment reçu. Ici, on constate bien que la réponse de dailymotion a placé la TSVal du paquet envoyé par l'ordinateur dans son champ TSecr.

B. Happy Eyeballs

Avant le premier paquet [SYN] déjà traité, le navigateur effectue deux requêtes DNS, de type A et AAAA pour résoudre l'adresse www.dailymotion.com. Comme l'analyse DNS le préconisait, la requête AAAA ne génère aucune réponse, tandis que celle de type A renvoie bien l'IPv4 de dailymotion.com. Comme le montre la figure 5, il semblerait donc que Mozilla pratique le Happy Eyeballs, c'est à dire qu'il essaie systématiquement de résoudre les DNS en IPv4 et en IPv6, mais dailymotion ne comporant pas d'IPv6, on ne peut pas observer s'il préfère bien l'IPv6 à l'IPv4.

FIGURE 5. Connexion à dailymotion

C. Flags

En se concentrant sur les flags des échanges entre l'ordinateur et dailymotion, on observe que lors de la retransmission d'un paquet, le flag Push est spécifié, ce qui signifie que le paquet ne doit pas être mis en buffer mais directement envoyé. Mis à part cela, les échanges ne montrent pas de flags particuliers et la fin de connexion se fait par l'envoi d'un [FIN, ACK] auquel un [FIN, ACK] est envoyé en réponse, suivi d'un dernier [ACK]. Le RTT varie entre 15 et 75 ms, ces temps assez raisonnables sont permis grâce à l'utilisation de CDN, comme le chapitre sur l'analyse DNS en rendait compte. La présence de CDN n'aurait pas pu être devinée par un simple test depuis un ordinateur, surtout sachant que les serveurs de dailymotion ne sont pas extrêmement loin de l'endroit où le test est effectué , mais les sondes placées un peu partout par RIPE Atlas le suggèrent beaucoup mieux.

Un dernier moyen d'en savoir plus sur la gestion faite de TCP par dailymotion est la commande scapy. L'effet de cette commande capté par wireshark est repris sur la figure 6. Dans l'exemple repris, on envoie un paquet [SYN, ECN, CWR] à dailymotion avec l'option TCP Fast Open. ECN signifie qu'on veut utiliser l'Explicit Congestion Notification et CWR annonce que la fenêtre de congestion est réduite.

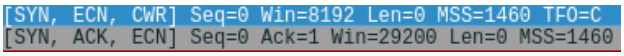

FIGURE 6. Commande Scapy capturée par wireshark

Dailymotion répond à ce paquet avec le flag ECN, ce qui prouve qu'il supporte l'Explicit Congestion Notification. En revanche, l'option TFO est absente et donc pas supportée par dailymotion.

RÉFÉRENCES

[1] Wikipedia_Liste des enregistrements DNS, URL : https://fr.wikipedia.org/wiki/Liste_des_enregistrements_DNS, source : site IANA, autheurs : Roy Arends, Frederico AC Neves, Olafur Gudmundsson, Ray Bellis. Consulté en novembre 2018.

[2] Wikipedia_SOA record, URL : https://en.wikipedia.org/wiki/SOA_record, source : RFC1035, consulté en novembre 2018.

[3] RIPE Atlas, URL : https://atlas.ripe.net/, créé en 2010, consulté en novembre 2018.

[4] WebPagetest, développé par AOL en 2008, URL : https://www.webpagetest.org/, consulté en novembre 2018.

[5] Cache Expiration, rédigé par ATT, URL : https://developer.att.com/video-optimizer/docs/best-practices/cache-expiration, consulté en décembre 2018.

[6] MDN Web Docs, Mozilla, URL : https://developer.mozilla.org/en-US/, consulté en décembre 2018.

[7] "Le Cookie Matching… pour une expérience utilisateur sublimée", article rédigé le 17 février 2016 sur un blog, URL : http ://blog.makazi.com/data-management-platform-le-cookie-matching-pour-une-experience-utilisateur-sublimee/, consulté en décembre 2018.

[8] Web Application and User protection, Sqreen, URL : https://www.sqreen.io/, consulté en décembre 2018.

[9] Qualis. SSL Labs, URL : https://www.ssllabs.com/ssltest/analyze.html?d=dailymotion.com&latest, consulté en décembre 2018.

[10] PacketLife.net, blog, URL : http://packetlife.net/blog, consulté en décembre 2018.

[11] Networking Notes, écrit en novembre 2018 par O. Bonaventure, https://obonaventure.github.io/cnp3blog/, consulté en novembre et décembre 2018.

[12] Dailymotion, URL : https://www.dailymotion.com/be, créé en 2005, consulté en novembre et décembre 2018.