LINGI1341 : Projet 2
Analyse d'un site web
nordpresse.be

de Meeûs d'Argenteuil Corentin
École Polytechnique de Louvain

***Abstract*—Ce travail contient l'analyse web d'un site de presse parodique : nordpresse.be.**

I. Introduction

Nordpresse est un site de presse parodique indépendant, jouant sur la naïveté des lecteurs. Étant un site de presse indépendant, sa survie dépend de la publicité, donc il y en a beaucoup.

Nordpresse.be utilise WordPresse pour la création du site. Sa configuration, du point de vue de la sécurité, est donc assez standard.

II. Analyse DNS

Le site nordpresse.be est hébergé par OVH sur les serveurs de CloudFlare. Il ne possède aucun CNAME, mais deux serveurs différents.

A. Noms des serveurs

CloudFlare est un hébergeur américain. Les noms des deux serveurs utilsés par nordpresse sont maya.ns.cloudflare.com et justin.ns.cloudflare.com.
Ces deux serveurs sont disponibles en IPv4 ainsi qu'en IPv6. Sur les serveurs de CloudFlare, il y a une technologie qui permet de diviser le trafic du site sur plusieurs machines différentes. C'est pour cela que lors d'une analyse traceroute depuis la Belgique, on va passer par le seveur d'Amsterdam, alors que depuis la France, le trafic passera par les serveurs parisiens.

B. IPv4 et IPv6

Nordpresse.be est accessible en IPv4 ainsi qu'en IPv6. Les adresses IPv4 de nordpresse.be sont 104.18.57.34 et 104.18.56.34 alors que les adresses IPv6 sont 2606:4700:30::6812:3922 et 2606:4700:30::6812:3822. Pour ces 4 adresses, le TTL est de 300ms.

La connexion au site est partagée entre les deux serveurs, comme il est observable dans les analyses RIPE Atlas [2].

Il s'agit d'adresses IP assez nouvelles. En effet, celles-ci ont été modifiées le 11 novembre dernier. Grâce au site viexdns.info [3], on conclut que le site nordpresse.be change très souvent d'adresse IP (plus de 13 fois en 2018!).

C. Txt

Le site nordpresse.be possède un record DNS de type TXT. Il s'agit d'une chaîne de caractères commençant par "google-site-verification". Celle-ci permet à Google de vérifier si l'utilisateur courant est le propriétaire du nom de domaine afin que ce dernier puisse avoir accès à son site dans Google analytique et Google Search Console

D. MX

On observe via une commande *dig* que les serveurs mails de nordpresse passent par les serveurs mx1, mx2 et mxb d'ovh.net.

III. Analyse HTTP

Le site nordpresse.be utilise beaucoup de requêtes HTTP. En effet, il y a énormément de pubs sur tout le site. Juste sur la page d'accueil, AdBlock trouve déjà 19 pubs. En allant un peu plus loin, par exemple en cliquant sur un article au hasard, il y a encore plus de pubs ; AdBlock en dénombre 27 sur la page. En descendant dans les articles suggérés, on remarque que la majorité d'entre eux sont des publicités (et ne sont pas bloqués par AdBlock). Et, encore plus étonnant, cette liste d'articles suggérés et de pubs apparaît à l'infini, sans limitation. A force de descendre dans la page, de nouvelles pubs apparaissent.

A. Redirection permanentes

Le site possède certaines redirections permanentes. Par exemple, en se rendant sur www.nordpresse.be, on est immédiatement redirigé vers nordpresse.be. Lors de l'envoi de la méthode get, une erreur 301 (redirection permanente) est lancée.

Il y a le même phénomène lorsqu'on se rend sur le site via http://nordpresse.be au lieu de https://nordpresse.be.

En appliquant les deux en même temps, soit en se rendant sur http://www.nordpresse.be, les deux redirections se font l'une après l'autre. La première est celle pour passer de http à https et la deuxième pour enlever les www.

B. Domaines appelés

Avec une analyse d'accès à nordpresse.be depuis urlscan.io [4], il y a 68 adresses ip contactées, passant par 9 pays et 50

noms de domaines pour un total de 220 requêtes HTTP.

D'une fois à l'autre, il y a beaucoup de différences. Dans l'exemple des 220 requêtes, le tableau à la figure 1, tirée de urlscan.io, représente les 10 domaines les plus appelés par le chargement de la page d'accueil. 61 de ces requêtes sont dirigées vers nordpresse.be, ce qui est normal étant donné que c'est le serveur source.

Regardons maintenant les autres domaines présents dans ce top 10. Le deuxième est adnxs.com. D'après le the-guardian.com [5], *"Adnxs.com is run by AppNexus, a company that provides technology, data and analytics to help companies buy and sell online display advertising"*. Il s'agit donc de requêtes utilisées pour la publicité. Du 3ème au 10ème du classement, il s'agit aussi de services de pub ; des régies publicitaires. Le site nordpresse.be réalise donc beaucoup d'appels publicitaires. En regardant via le *developper tool* de Firefox, le site nordpresse charge des images publicitaires jusqu'à environ 580 requêtes. Après cela, tous les certains temps (environ 1min30), toutes les publicités de la page sont rechargées, donc il y a de nouvelles requêtes HTTP.

Domain / Subdomains		Transfer
61 nordpresse.be		836 KB
20 adnxs.com	5 redirects	27 KB
15 themoneytizer.com		173 KB
14 smartadserver.com	1 redirects	15 KB
11 gstatic.com		117 KB
10 ayads.co	1 redirects	62 KB
9 stickyadstv.com	7 redirects	6 KB
7 adikteev.com		157 KB
7 consensu.org		103 KB
6 sascdn.com		23 KB

Fig. 1. Domaines les plus appelés lors de l'accès à nordpresse.be

C. Headers non standard

En analysant les requêtes HTTP, on remarque que certains headers comportent des informations non standard. En voici une partie :

- P3P : il s'agit d'un protocole obsolète qui permet au site de déclarer les informations qu'il collecte à l'utilisateur, et donc d'informer de sa politique de protection des données personnelles.
- x-amz-[...] : il s'agit de valeurs ajoutées par CloudFront (Amazon). Cela indique qu'une des publicités provenait des serveurs d'Amazon.
- X-Firefox-Spdy: SPDY est un protocole de transport sécurisé qui remplace HTTP/1 dans Firefox (h2). Il n'est donc pas présent avec les autres navigateurs.
- x-goog-[...] : Il s'agit d'extensions customisées d'en-tête HTTP fournies par Google. Dans les requêtes analysées, il y avait par exemple des requêtes `x-goog-generation`. Il s'agit d'une indication sur la version de la donnée à laquelle on essaye d'accéder.
- alt-svc : Cette fonctionnalité permet au header de dire une façon alternative d'accéder au domaine ou un autre protocole. Elle a provoqué des problèmes pour Firefox en 2015 car il y avait une faille de sécurité à cause de celle-ci. En effet, il y avait moyen de contourner la vérification du certificat en l'utilisant. [6]
- traffic-path : il s'agit probablement d'une fonctionnalité qui force le paquet à suivre un chemin particulier.
- Mais aussi par exemple AN-X-Request-Uuid, cf-cache-status, cf-ray, X-Proxy-Origin, X-IPLB-Instance, Fastly-Debug-Digest, ...

D. Technologies utilisées

Dans la figure 2, on peut voir les technologies utilisées lors de l'appel de la page d'accueil de nordpresse (tableau tiré de urlscan.io [4]).

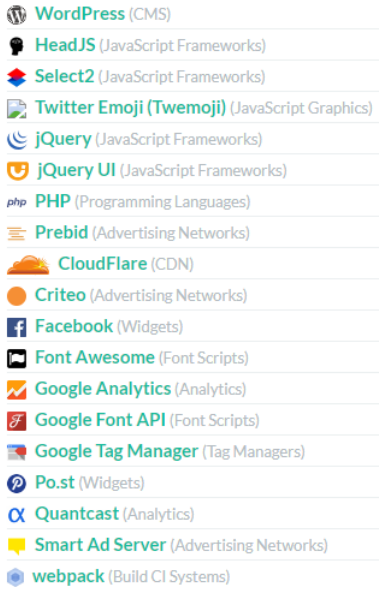

Fig. 2. Technologies utilisées

Nordpresse est un site créé avec l'aide de WordPress. Il s'agit donc naturellement d'une ressource utilisée. Il y a ensuite plusieurs packages de javascript, comme par exemple Select2 qui ajoute des box de sélections personnalisables ou jQuery qui permet de simplifier la navigation et la manipulation des fichiers HTML. Il y a également des appels à Facebook, Po.st et Twitter Emoji pour insérer leurs widgets respectifs. Il y a bien entendu l'utilisation de certaines technologies publicitaires (Prebid, Criteo, Smart Ad Server) et l'utilisation de technologies permettant de réaliser des statistiques comme Google analytics et Quantcast.

E. Cookies

En se basant sur le site cookie-checker.com [7], on peut voir la quantité de cookies présente lors de connexion à la page d'accueil de nordpresse. Cookie-checker signale l'utilisation de 9 "First Party Cookies". Il s'agit de cookies enregistrés par le site en lui-même sur l'ordinateur du client.

Seul le site en question peut les lire. Ils sont utilisés pour des fonctionnalités du site, comme par exemple pour retenir le contenu d'un panier dans un webshop. Dans notre cas, on y voit par exemple un cookie nommé *browsiSessionID*. Il contient les informations de connexions.

Il y a également 48 "Third Party Cookies". Il s'agit des cookies enregistrés par des services tiers, très souvent des entreprises publicitaires. Ce nombre très élevé est prévisible vu la quantité abusive de publicités. Ces cookies peuvent être lus depuis un autre site également ; c'est ce qui permet de faire des publicités personnalisées.

IV. Analyse TLS

Du point de vue de la sécurité, le site nordpresse.be est assez bien sécurisé. En effet, le site ssllabs.com lui donne une note de A. Le graphe à la figure 3 en montre les cotes détaillées. [8]

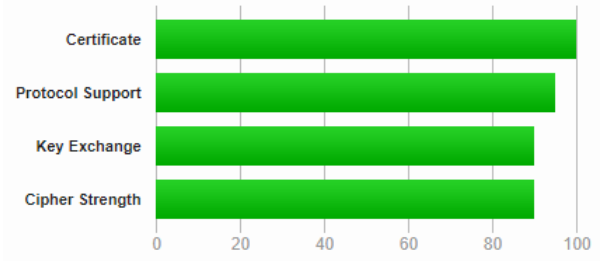

Fig. 3. Notes sur 100 d'après ssllabs.com

D'une part, le graphique montre que le site nordpresse.be utilise un protocole de chiffrement. En effet, il utilise la version 1.3 de TLS pour les dernières versions de navigateur comme Chrome 70 sur Windows 10. Il accepte également les anciennes versions de TLS (1.0, 1.1 et 1.2) pour les autres navigateurs. Par contre, pour les plus vieilles versions comme java6u45 ou android 2.3.7, une erreur fatale est lancée (handshake_failure) car il n'y a pas de certificat disponible. Nordpresse.be a besoin que le navigateur possède un support SNI (Serveur Name Indication) pour fonctionner.

D'autre part, ce graphique nous montre que le certificat de nordpresse.be est bon. Il y en a même trois différents. En voici les caractéristiques :

- Algorithme de signature SHA256withECDSA avec une clé EC de 256 bits. Il est issu de COMODO ECC Domain Validation Secure et expire mi-février.
- Algorithme de signature SHA384withECDSA avec une clé EC de 256 bits. Il est issu de COMODO ECC Certification Authority et expire en 2029.
- Algorithme de signature SHA384withRSA avec une clé EC de 384 bits. Il est issu de AddTrust External CA Root et expire en mai 2020.

La `Secure Renegotiation` est supportée. Elle permet de sécuriser la connexion.

Par contre, comme le montre la figure 4, Firefox signale qu'il peut y avoir un problème de sécurité. En effet, certaines des informations, des images ici, ne sont pas cryptées lorsqu'elles sont envoyées sur le réseau.

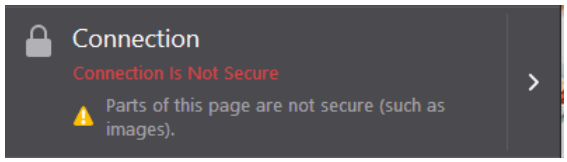

Fig. 4. Message de sécurité de Firefox

V. Analyse TCP

A. Ouverture de la connexion

Avec une analyse faite via Wireshark, on sait analyser les paquets envoyés par le client vers le serveur ou du serveur vers le client. La figure 5 représente les deux premiers paquets TCP envoyés.

```
57684 → 443 [SYN] Seq=0 Win=64260 Len=0 MSS=1428 WS=256 SACK_PERM=1
443 → 57684 [SYN, ACK] Seq=0 Ack=1 Win=27200 Len=0 MSS=1360 SACK_PERM=1 WS=1024
```

Fig. 5. Les deux premiers paquets échangés lors de la connexion

Le premier contient le protocole TCP que le client veut utiliser et le suggère au serveur. On y observe que la taille de segment maximal acceptée par le client est de 1428 bytes. Le client propose un windows scale de 8 (256) et signale que les SACK sont permis.

Le deuxième paquet TCP est la réponse du serveur. Celui-ci précise une windows scale de 10 (1024), dit également que les SACK sont permis et que sa taille maximum de segment est de 1360 bytes.

Une observation est donc que le serveur a une fenêtre 4 fois plus grande que celle du client. L'option timestamps n'est pas négociée dans ces deux premiers paquets. Cela signifie qu'il n'y a pas de vérifications du fait que le paquet reçu ne soit pas un vieux paquet qui est hors séquence. Il n'y a aucune autre option en négociation tels que TCP Fast Open ou TCP Explicit Congestion Notification.

Il est possible de voir l'effet de l'option SACK assez facilement via Wireshark. En effet, pour plusieurs paquets de données envoyés par le serveur, le client n'envoie qu'un seul ACK.

En se basant sur l'analyse Wireshark, on sait également voir le nombre de connexions TCP parallèles qui sont utilisées par le navigateur. Ces analyses ont été effectuées en navigation privée en désactivant le cache et sur windows 10. Voici une liste récapitulative du nombre de connexions TCP en fonction du navigateur.

- Chrome : 2 en même temps

- Firefox : 4 en même temps, 6 différentes au total
- Edge : 4 en même temps

Grâce au site cdnplanet.com [9], on peut voir que la valeur initiale de la fenêtre de congestion du serveur de CloudFlare était de 10 en 2017.

Le TTFB (Time To First Byte) est de 1,002ms. Il s'agit du temps entre l'envoi de la requête Get et la réception du premier byte.

B. RTT

Comme on peut le voir sur le RIPE Atlas de nordpresse.be [2], le round-trip-time mesuré avec ping varie énormément d'un endroit à l'autre du monde. Depuis la Belgique, il est de 14.892ms. Expérimentalement, en navigant sur le site, le RTT est un peu plus élevé. En navigant en Wi-Fi via Chrome, j'ai un RTT d'environ 20ms en moyenne. Via la commande ping, j'ai un RTT moyen de 12ms. Le pire ping d'accès à nordpresse présent dans le RIPE Atlas est en île Maurice, avec un ping à 1341.686ms. Les deux meilleurs pings, eux, sont des tests effectués depuis les US. C'est assez logique étant donné que le fournisseur CloudFlare est américain.

C. IPv4 ou IPv6 ?

Lors de l'accès à nordpresse.be, les trois navigateurs testés (Chrome, Firefox et Edge) préfèrent l'utilisation de l'IPv6, et cela dans plusieurs conditions et lieu d'utilisation (Wi-Fi ou Ethernet, à domicile où via eduroam à l'UCL). Par contre, l'utilisation d'une adresse IPv6 ou de l'autre change en permanence. Même, parfois, entre la première requête (lorsqu'on commence à taper nordpresse dans la barre de recherche et que la suggestion du site apparaît) et l'accès au site en lui-même, l'adresse change.

D. Fermeture de connexions

Les connexions TCP se terminent d'une façon totalement différente en fonction du lieu et de la méthode de test. Par exemple, chez moi, en Wi-Fi, j'ai beaucoup de paquets TCP qui sont finis de façon abrupte avec l'envoi d'un RST par le serveur ou le client. Toujours de chez moi, mais cette fois en étant connecté via Ethernet, il y a moins de RST et plus d'échanges de FIN. Et enfin, via la connexion Wi-Fi en salle Intel, je n'ai presque plus de RST. Cette observation est identique tant avec Chrome qu'avec avec Firefox.

VI. Conclusion

Nordpresse.be est un site avec beaucoup de publicités et donc beaucoup de requêtes HTTP. Le site est accessible via deux adresses IPv4 et deux adresses IPv6. Pour améliorer la vitesse de la connexion, il y a plusieurs connexions TPC en parallèle.
La sécurité y est relativement bonne. Enfin, il y a de nombreux champs de header non-standard dans les requêtes HTTP.

References

[1] Bonaventure Oliver, *Computer Networking : Principles, Protocols and Practice*, http://cnp3book.info.ucl.ac.be/2nd/html/
[2] Analyse Ripe Atlas IPv6 : https://atlas.ripe.net/measurements/17177690/
Analyse Ripe Atlas IPv4 : https://atlas.ripe.net/measurements/17177689/
[3] Évolution de l'adresse IP : https://viewdns.info/iphistory/?domain=nordpresse.be
[4] Analyse HTTP : https://urlscan.io/result/f91dbd41-0155-4ae1-ae3f-58ed9394042a
[5] Geary Joanna et Nicola Hughes, *Battle for the internet Cookies and web tracking*, https://www.theguardian.com/technology/2012/apr/23/adnxs-tracking-trackers-cookies-web-monitoring
[6] Fail de sécurité Alt-Svc : https://www.mozilla.org/en-US/security/advisories/mfsa2015-44/
[7] Analyse des cookies : http://www.cookie-checker.com/check-cookies.php?url=nordpresse.be%2F
[8] Analyse des certificats : https://www.ssllabs.com/ssltest/analyze.html?d=nordpresse.be&
[9] Fenêtre initiale CloudFlare : https://www.cdnplanet.com/blog/initcwnd-settings-major-cdn-providers/