Estelle Monier 3TC2

Synthèse bibliographique

Introduction

Toutes les connections Internet sont précédées d'une recherche DNS sans protection du trafic, ce qui implique la possibilité pour des entités de surveiller et tracker les utilisateurs, voire de les censurer. Pour remédier à cela, de grandes entreprises telles que Google, Cloudflare ou encore Mozilla ont standardisé des protocoles permettant de chiffrer le trafic DNS : DNS-over-TLS (DoT) et DNS-over-HTTPS (DoH). Ces protocoles permettent une protection contre l'analyse du trafic, par bourrage du trafic ou par multiplexage du trafic DNS avec d'autres trafics. Cependant, même si les communications sont chiffrées, certaines caractéristiques du trafic peuvent révéler des informations sur son contenu. Quelle est alors l'efficacité de ces nouveaux mécanismes en terme de protection de la vie privée ?

Etat de l'art

Tout d'abord, nous savons qu'il existe des problèmes de sécurité liés à l'utilisation du trafic DNS sur les pages web [2]. En effet, les requêtes DNS ont de nombreux champs qui peuvent être assez dangereux pour préserver la privacité de l'utilisateur : l'adresse IP source (ainsi que le port utilisé) et le QNAME. Ce dernier peut être particulièrement dangereux lors de visites de sites web pouvant être considérés comme offensants ou peu recommandables. De plus, lors d'un requête DNS, des informations sont envoyées à un résolveur récursif afin de faire le lien entre l'adresse IP et le nom de domaine. En effet, le client envoie une requête DNS à un résolveur récursif (un serveur avec des capacités de résolutions DNS et de cache), et si la résolution de domaine n'est pas dans le cache du résolveur récursif, il contacte un certain nombre de serveurs de noms faisant autorité contenant une base de données du mappage entre noms de domaines et adresses IP. Le résolveur récursif va alors parcourir cette base de données jusqu'à trouver la réponse à la requête DNS qu'il a reçu. Il va alors la transmettre au client et ce dernier pourra utiliser cette adresse IP pour se connecter à l'hôte de destination. Ces résolveurs récursifs ont donc de très nombreux clients avec une relation one-to-many avec les serveurs d'autorité, ce qui permet un risque d'attaque assez bas sur le lien résolveur récursif-serveur [3]. Cependant le trafic DNS entre le client et le résolveur est lié à une adresse IP spécifique et est donc exposée à un grand nombre d'entités. Ainsi, il est possible de traquer un utilisateur en analysant le cache d'un résolveur récursif, ce qui peut encore une fois révéler des données sensibles et cette possibilité fait donc preuve d'un manque de fiabilité sur la protection des données.

Un moyen de protéger ces données est donc de chiffrer le trafic DNS, du moins jusqu'au résolveur récursif. C'est pour cela que deux protocoles ont été mis en place afin de chiffrer le trafic DNS : DoT et DoH. Le protocole DoT, standardisé en 2016, fonctionne de la sorte : Le client établit une session TLS avec un résolveur récursif en utilisant un port TCP dédié pour le trafic DNS (facile à traquer et bloquer), et échange des requêtes et réponses DNS sur une connexion chiffrée, alors que le protocole DoH, standardisé en 2018, se base sur

une connexion HTTPS. En effet, dans ce protocole, le résolveur DNS local établit une connexion HTTPS avec le résolveur récursif et encode les requêtes DNS dans le corps des requêtes HTTP (et donc pousse les réponses DNS qui vont avoir tendance à suivre une recherche DNS, ce qui réduit la latence) [1].

Malgré cela, il est aujourd'hui reconnu que même des caractéristiques du trafic DNS chiffré tel que le volume d'informations ou les synchronisations peuvent révéler des informations quant au contenu du trafic DNS [4]. On fait donc face à des failles de sécurité, et on se pose alors des questions sur l'efficacité de ces protocoles.

Résultats et comparaison d'articles

Dans l'article de référence, les auteurs ont étudié l'efficacité des attaques par analyse du trafic en se concentrant sur le protocole DoH, et ils ont comparé la protection offerte par DoH par rapport à DoT [1].

Pour cela, ils ont considéré un adversaire passif dont le but est d'identifier quelles pages web l'utilisateur visite, ou de pratiquer de la surveillance ou de la censure. Les caractéristiques utilisées pour une attaque sur le trafic web n'étant pas applicables sur le trafic DNS, il ont conçu un nouvel ensemble de caractéristiques permettant des attaques efficaces pour identifier des sites web visités sur un trafic DNS chiffré.

Dans un article sur le chiffrement du trafic DNS [5], nous constatons différents modèles d'attaque entre le client (et donc le résolveur récursif) et le serveur. On voit alors que des attaques peuvent être efficacement menées avec le trafic chiffré car l'adversaire, qui peut être un réseau malveillant ou tout simplement un espion, a accès au temps entre les demandes et les réponses, à la taille des paquets de données échangés et aux dépendances de confiance transitive. Le cache ne pouvant être exploité suite au chiffrement bout-à-bout du trafic, il suffit à l'adversaire d'analyser le motif des requêtes DNS et de leur réponse pour avoir accès à la latence et à la taille de la réponse et donc lui permettre de deviner la requête DNS.

De plus, d'autres études ont montré que la taille des messages DNS était unique pour chaque site, et que souvent, si la taille était plus ou moins similaire entre différents sites web, c'est qu'ils appartenaient à la même organisation. En outre, ces mêmes études ont aussi montré que sur plusieurs jours, la taille de ces données ne changeait pas, ou alors de manière minime [6]. Ainsi toutes ces études sont d'accord pour dire que la taille des données est un facteur très important dans l'analyse du trafic DNS chiffré.

Le trafic DNS étant composé de petits paquets, la technique utilisée dans l'article de référence nécessite 124 fois moins de données que celle utilisée pour analyser le trafic web, et est tout autant efficace, voire plus [1]. Ces caractéristiques permettent également à l'adversaire de connaître l'environnement sur lequel l'attaque sera déployée, contrairement à la plupart des anciens travaux sur le sujet, ce qui permet une attaque de haute performance sur tous les environnements.

Ils ont alors testé leur attaque sur des défenses existantes du trafic, et ont remarqué que la solution de rembourrage (« padding ») EDNS0 (par Google et Cloudflare) ne pouvait pas complètement empêcher leur attaque [1]. Le padding permet de changer la taille de l'information lors d'une requête DNS, et est donc censé être une solution pour garantir une meilleure privacité des données. En effet, comme expliqué précédemment, la taille des données peut être un facteur important pour deviner la requête DNS [6]. Cependant, cette solution de padding ne garantit pas vraiment une sécurisation optimale de données, puisque l'attaque menée par les auteurs de l'article de référence n'a pas été empêchée. De plus, d'autres études ont prouvé que cette solution n'était pas 100% efficace. En effet, il a été montré que même un « padding » complet ne pouvait pas empêcher les analyses du trafic DNS chiffré, et qu'il y avait de graves pertes de confidentialité des informations, même pour des adversaires passifs [7].

En revanche, bien que Tor ne témoigne pas d'une grande protection contre l'empreinte de l'utilisateur sur des sites web sur le trafic web, il est extrêmement efficace contre celle sur le trafic DNS chiffré [1].

Les chercheurs de l'article de référence ont ensuite montré que même en chiffrant le trafic DNS, il était possible pour l'attaquant de trouver le paquet contenant la recherche DNS pour le 1er domaine. Ils ont également quantifié le compromis entre le nombre de données d'un site blacklisté qu'un utilisateur peut télécharger, et le nombre de sites non-blacklistés qui sont censurés par effet secondaire du blocage basé sur l'analyse du trafic. Et enfin, ils ont aussi rassemblé un premier ensemble de données de DNS encrypté collectés dans une grande gamme d'environnements.

Conclusion

Au final, toutes les études sont d'accord sur un point : le chiffrage du trafic DNS permet de mieux sécuriser les données des requêtes DNS, mais ce n'est pas suffisant. Même les solutions de padding se sont montrées inefficaces face aux attaques que nous pouvons réaliser de nos jours. Il est aujourd'hui possible d'avoir accès à nos données lorsque nous effectuons des recherches sur Internet. Un des facteurs les plus inquiétants qu'il faudrait sécuriser est la taille des messages DNS échangés. Ainsi, bien que de nombreuses avancées aient été faites dans ce domaine, il y a encore beaucoup de recherche à faire sur ce sujet afin de permettre une protection 100% efficace des données des utilisateurs.

Références :

Article de référence :

[1] S. Siby, M. Juarez, C. Diaz, et al. Encrypted DNS --> Privacy? A Traffic Analysis Perspective. arXiv:1906.09682 [cs], Oct. 2019 [En ligne]. Disponible sur : http://arxiv.org/abs/1906.09682 (Consulté le 04/04/2021).

Articles d'état de l'art :

[2] S. Bortzmeyer. DNS Privacy Considerations. 2015 [En ligne]. Disponible sur : https://www.hjp.at/doc/rfc/rfc7626.html (Consulté le 04/04/2021).

[3] S. Siby , M. Juarez , N. Vallina-Rodriguez et al. DNS Privacy not so private: the traffic analysis perspective. 2018, 2 p.

[4] A. Panchenko, F. Lanze, A. Zinnen, et al. Website fingerprinting at internet scale. In Network & Distributed System Security Symposium NDSS. IEEE Computer Society, 2016, 15 p.

Articles de comparaison de résultats :

- Présent dans la bibliographie :

[5] H. Shulman. Pretty bad privacy: Pitfalls of DNS encryption. In Proceedings of the 13th Workshop on Privacy in the Electronic Society. ACM, 2014.

- Absents de la bibliographie :

[6] S. Siby , M. Juarez , N. Vallina-Rodriguez et al. DNS Privacy not so private: the traffic analysis perspective. 2018, 2 p.

[7] J. Bushart et C. Rossow. Padding Ain't Enough: Assessing the Privacy Guarantees of Encrypted DNS. arXiv:1907.01317 [cs.CR], Jul. 2019, 2019 [En ligne]. Disponible sur : https://arxiv.org/abs/1907.01317 (Consulté le 04/04/2021).