RFC 9076 : DNS Privacy Considerations

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 24 juillet 2021

Date de publication du RFC : Juillet 2021

https://www.bortzmeyer.org/9076.html

La surveillance généralisée sur l'Internet est un gros problème pour la vie privée. L'IETF s'active donc à améliorer la protection de la vie privée contre cette surveillance (RFC 7258 [1]). Un protocole qui avait souvent été négligé dans ce travail est le DNS. Ce RFC décrit les problèmes de vie privée liés au DNS. Il remplace le RFC 7626, avec deux changements important, l'intégration des techniques de protection développées et déployées depuis l'ancien RFC et, malheureusement, beaucoup de propagande anti-chiffrement imposée par les acteurs traditionnels des résolveurs DNS <https://www.bortzmeyer.org/resolveur-dns.html> qui ne veulent pas se priver de leurs possibilités de contrôle et de surveillance.

Ce RFC est en fait à la croisée de deux activités. L'une d'elles consiste à documenter les problèmes de vie privée, souvent ignorés jusqu'à présent dans les RFC. Cette activité est symbolisée par le RFC 6973, dont la section 8 contient une excellente analyse de ces problèmes, pour un service particulier (la présence en ligne). L'idée est que, même si on ne peut pas résoudre complètement le problème, on peut au moins le documenter, pour que les utilisateurs soient conscients des risques. Et la seconde activité qui a donné naissance à ce RFC est le projet d'améliorer effectivement la protection de la vie privée des utilisateurs du DNS, en marchant sur deux jambes : minimiser les données envoyées et les rendre plus résistantes à l'écoute, via le chiffrement. L'objectif de diminution des données a débouché sur la "*QNAME minimization*" spécifiée dans le RFC 9156 et le chiffrement a donné DoT (RFC 7858) et DoH (RFC 8484).

Donc, pourquoi un RFC sur les questions de vie privée dans le DNS ? Ce dernier est un très ancien protocole, dont l'une des particularités est d'être mal spécifié : aux deux RFC originaux, les RFC 1034 et RFC 1035, il faut ajouter dix ou vingt autres RFC dont la lecture est nécessaire pour tout comprendre du DNS. Et aucun travail de consolidation n'a jamais été fait, contrairement à ce qui a eu lieu pour XMPP,

1. Pour voir le RFC de numéro NNN, https://www.ietf.org/rfc/rfcNNN.txt, par exemple https://www.ietf.org/rfc/rfc7258.txt

HTTP ou SMTP. Or, le DNS est crucial, car quasiment toutes les transactions Internet mettent en jeu au moins une requête DNS (ne me dites pas des bêtises du genre « moi, je télécharge avec BitTorrent, je n'utilise pas le DNS » : comment allez-vous sur `thepiratebay.am`?) Mais, alors que les questions de vie privée liées à HTTP ont fait l'objet d'innombrables articles et études, celles liées au DNS ont été largement ignorées pendant longtemps (voir la bibliographie du RFC pour un état de l'art). Pourtant, on peut découvrir bien des choses sur votre activité Internet uniquement en regardant le trafic DNS.

Une des raisons du manque d'intérêt pour le thème « DNS et vie privée » est le peu de compétences concernant le DNS : le protocole est souvent ignoré ou mal compris. C'est pourquoi le RFC doit commencer par un rappel (section 1) du fonctionnement du DNS.

Je ne vais pas reprendre tout le RFC ici. Juste quelques rappels des points essentiels du DNS : il existe deux sortes de serveurs DNS, qui n'ont pratiquement aucun rapport. Il y a les **serveurs faisant autorité <https://www.bortzmeyer.org/serveur-dns-faisant-autorite.html>** et les **résolveurs <https://www.bortzmeyer.org/resolveur-dns.html>**. Les premiers sont ceux qui connaissent de première main l'information pour une zone DNS donnée (comme `fr` ou `wikipedia.org`). Ils sont gérés par le titulaire de la zone ou bien sous-traités à un hébergeur DNS. Les seconds, les résolveurs, ne connaissent rien (à part l'adresse IP des serveurs de la racine). Ils interrogent donc les serveurs faisant autorité, en partant de la racine. Les résolveurs sont gérés par le FAI ou le service informatique qui s'occupe du réseau local de l'organisation. Ils peuvent aussi être individuels <https://www.bortzmeyer.org/son-propre-resolveur-dns.html>, ou bien au contraire être de gros serveurs publics comme Google Public DNS <https://www.bortzmeyer.org/google-dns.html>, gros fournisseur de la NSA. Pour prendre un exemple concret (et en simplifiant un peu), si M. Michu veut visiter le site Web `http://thepiratebay.am/`, son navigateur va utiliser les services du système d'exploitation sous-jacent pour demander l'adresse IP de `thepiratebay.am`. Le système d'exploitation va envoyer une requête DNS au résolveur (sur Unix, les adresses IP des résolveurs sont dans `/etc/resolv.conf`). Celui-ci va demander aux serveurs de la racine s'ils connaissent `thepiratebay.am`, il se fera rediriger vers les serveurs faisant autorité pour `am`, puis vers ceux faisant autorité pour `thepiratebay.am`. Le résolveur aura alors une réponse qu'il pourra transmettre au navigateur de M. Michu.

Principal point où j'ai simplifié : le DNS s'appuie beaucoup sur la **mise en cache** des données, c'est-à-dire sur leur mémorisation pour gagner du temps la fois suivante. Ainsi, si le même M. Michu, cinq minutes après, veut aller en `http://armenpress.am/`, son résolveur ne demandera rien aux serveurs de la racine : il sait déjà quels sont les serveurs faisant autorité pour `am`.

Le trafic DNS est un trafic IP ordinaire, typiquement porté par UDP. Mais il peut aussi fonctionner sur TCP et bientôt sur QUIC <https://www.bortzmeyer.org/quic.html>. Il peut être écouté, et comme il n'est pas toujours chiffré, un indiscret peut tout suivre. Voici un exemple pris avec tcpdump sur un serveur racine (pas la racine officielle, mais, techniquement, cela ne change rien) :

```
15:29:24.409283 IP6 2001:67c:1348:8002::7:107.10127 > \
    2001:4b98:dc2:45:216:3eff:fe4b:8c5b.53: 32715+ [1au] \
    AAAA? www.armenpress.am. (46)
```

On y voit que le client `2001:67c:1348:8002::7:107` a demandé l'adresse IPv6 de `www.armenpress.am`.

Pour compléter le tableau, on peut aussi noter que les logiciels génèrent un grand nombre de requêtes DNS, bien supérieur à ce que voit l'utilisateur. Ainsi, lors de la visite d'une page Web, le résolveur va envoyer la requête **primaire** (le nom du site visité, comme `thepiratebay.am`), des requêtes **secondaires** dues aux objets contenus dans la page Web (JavaScript, CSS, divers traqueurs et autres outils de cyberflicage ou de cyberpub) et même des requêtes **tertiaires**, lorsque le fonctionnement du DNS lui-même

nécessitera des requêtes. Par exemple, si `abc.xyz` est hébergé sur des serveurs dans `google.com`, une visite de `http://abc.xyz/` nécessitera de résoudre les noms comme `ns1.google.com`, donc de faire des requêtes DNS vers les serveurs de `google.com`.

Bien sûr, pour un espion qui veut analyser tout cela, le trafic DNS représente beaucoup de données, souvent incomplètes en raison de la mise en cache, et dont l'interprétation peut être difficile (comme dans l'exemple ci-dessus). Mais les organisations qui pratiquent l'espionnage massif, comme la NSA, s'y connaissent en matière de "*big data*" et savent trouver les aiguilles dans les bottes de foin.

Les sections 3 à 7 du RFC détaille les risques pour la vie privée dans les différents composants du DNS. Notez que la confidentialité du contenu du DNS n'est pas prise en compte (elle l'est dans les RFC 5936 et RFC 5155). Il est important de noter qu'il y a une énorme différence entre la **confidentialité du contenu** et la **confidentialité des requêtes**. L'adresse IP de `www.charliehebdo.fr` n'est pas un secret : les données DNS sont publiques, dès qu'on connait le nom de domaine, et tout le monde peut faire une requête DNS pour la connaitre. Mais le fait que vous fassiez une requête pour ce nom ne devrait pas être public. Vous n'avez pas forcément envie que tout le monde le sache. (On peut quand même nuancer un peu le côté « public » des données DNS : on peut avoir des noms de domaine purement internes à une organisation. Mais ces noms « fuitent » souvent, par la marche sur la zone décrite dans le RFC 4470, ou via des systèmes de « "*passive DNS*" <`https://www.bortzmeyer.org/dnsdb.html`>).

Pour comprendre les risques, il faut aussi revenir un peu au protocole DNS. Les deux informations les plus sensibles dans une requête DNS sont l'adresse IP source et le nom de domaine demandé ("*qname*", pour "*Query Name*", cf. RFC 1034, section 3.7.1). L'adresse IP source est celle de votre machine, lorsque vous parlez au résolveur, et celle du résolveur lorsqu'il parle aux serveurs faisant autorité. Elle peut indiquer d'où vient la demande. Lorsque on utilise un gros résolveur, celui-ci vous masque vis-à-vis des serveurs faisant autorité (par contre, ce gros résolveur va avoir davantage d'informations). Ceci dit, l'utilisation d'ECS (RFC 7871) peut trahir votre adresse IP, ou au moins votre préfixe. (Cf. cette analyse d'ECS <`https://00f.net/2013/08/07/edns-client-subnet/`>. D'autre part, certains opérateurs se permettent d'insérer des informations comme l'adresse MAC dans des options EDNS de la requête.)

Quant au "*qname*", il peut être très révélateur : il indique les sites Web que vous visitez, voire, dans certains cas, les logiciels utilisés. Au moins un client BitTorrent fait des requêtes DNS pour `_-bittorrent-tracker._tcp.domain.example`, indiquant ainsi à beaucoup de monde que vous utilisez un protocole qui ne plait pas aux ayant-droits. Et si vous utilisez le RFC 4255, pas mal de serveurs verront à quelles machines vous vous connectez en SSH...

Donc où un méchant qui veut écouter votre trafic DNS peut-il se placer ? D'abord, évidemment, il suffit qu'il écoute le trafic réseau. On l'a dit, le trafic DNS aujourd'hui est souvent en clair donc tout "*sniffer*" peut le décoder. Même si vous utilisez HTTPS pour vous connecter à un site Web, le trafic DNS, lui, ne sera pas chiffré. (Les experts pointus de TLS noteront qu'il existe d'autres faiblesses de confidentialité, comme le SNI du RFC 6066, qui n'est pas encore protégé, cf. RFC 8744.) À noter une particularité du DNS : le trafic DNS peut passer par un autre endroit que le trafic applicatif. Alice peut naïvement croire que, lorsqu'elle se connecte au serveur de Bob, seul un attaquant situé physiquement entre sa machine et celle de Bob représente une menace. Alors que son trafic DNS peut être propagé très loin, et accessible à d'autres acteurs. Si vous utilisez, par exemple, le résolveur DNS public de FDN <`http://blog.fdn.fr/?post/2014/12/07/Filtrer-The-Pirate-Bay-Ubu-roi-des-Internets`>, toute la portion de l'Internet entre vous et FDN peut facilement lire votre trafic DNS.

Donc, l'éventuel espion peut être près du câble, à écouter. Mais il peut être aussi dans les serveurs. Bercé par la musique du "*cloud*", on oublie souvent cet aspect de la sécurité : les serveurs DNS voient

passer le trafic et peuvent le copier. Pour reprendre les termes du RFC 6973, ces serveurs sont des **assistants** : ils ne sont pas directement entre Alice et Bob mais ils peuvent néanmoins apprendre des choses à propos de leur conversation. Et l'observation est très instructive. Elle est utilisée à de justes fins dans des systèmes comme DNSDB <https://www.bortzmeyer.org/dnsdb.html> (section 6 du RFC) mais il n'est pas difficile d'imaginer des usages moins sympathiques comme dans le cas du programme NSA MORECOWBELL <https://git.gnunet.org/bibliography.git/plain/docs/mcb-fr.pdf>.

Les résolveurs voient tout le trafic puisqu'il y a peu de mise en cache en amont de ces serveurs. Il faut donc réfléchir à deux fois avant de choisir d'utiliser tel ou tel résolveur ! Il est déplorable qu'à chaque problème DNS (ou supposé tel), des ignorants bondissent sur les réseaux sociaux pour dire « zyva, mets 8.8.8.8 [Google Public DNS] comme serveur DNS et ça ira plus vite » sans penser à toutes les données qu'ils envoient à la NSA ainsi.

Les serveurs faisant autorité voient passer moins de trafic (à cause des caches des résolveurs) mais, contrairement aux résolveurs, ils n'ont pas été choisis délibérement par l'utilisateur. Celui-ci peut ne pas être conscient que ses requêtes DNS seront envoyées à plusieurs acteurs du monde du DNS, à commencer par la racine. Le problème est d'autant plus sérieux que, comme le montre une étude <https://blog.imirhil.fr/vie-privee-et-le-dns-alors.html>, la concentration dans l'hébergement DNS est élevée : dix gros hébergeurs hébergent le tiers des domaines des 100 000 sites Web les plus fréquentés (listés par Alexa).

Au passage, le lecteur attentif aura noté qu'un résolveur personnel (sur sa machine ou dans son réseau local) a l'avantage de ne pas envoyer vos requêtes à un résolveur peut-être indiscret mais l'inconvénient de vous démasquer vis-à-vis des serveurs faisant autorité, puisque ceux-ci voient alors votre adresse IP. Une bonne solution (qui serait également la plus économe des ressources de l'Internet) serait d'avoir son résolveur local **et** de faire suivre les requêtes non résolues au résolveur du FAI. Du point de vue de la vie privée, ce serait sans doute la meilleure solution mais cela ne résout hélas pas un autre problème, celui des DNS menteurs <https://www.bortzmeyer.org/censure-francaise.html>, contre lesquels la seule protection est d'utiliser uniquement un résolveur de confiance. On peut résoudre ce problème en ayant son propre résolveur mais qui fait suivre à un résolveur public de confiance. Notez que la taille compte : plus un résolveur est petit, moins il protège puisque ses requêtes sortantes ne seront dues qu'à un petit nombre d'utilisateurs.

Enfin, il y a aussi les serveurs DNS « pirates » (installés, par exemple, via un serveur DHCP lui-même pirate) qui détournent le trafic DNS, par exemple à des fins de surveillance. Voir par exemple l'exposé de Karrenberg à JCSA 2012 <https://www.afnic.fr/fr/l-afnic-en-bref/actualites/actualites-generales/6171/show/succes-pour-la-journee-du-conseil-scientifique-sous-le-html> disponible en ligne <http://fr.slideshare.net/AFNIC/03-karrenbergsomethoughsvulnresil> (transparents 27 et 28, "*Unknown*" et "*Other*" qui montrent l'existence de clones pirates du serveur racine K.root-servers.net).

Pour mes lecteurs en France férus de droit, une question intéressante : le trafic DNS est-il une « donnée personnelle » au sens de la loi Informatique & Libertés ? Je vous laisse plancher sur la question, qui a été peu étudiée.

Les changements depuis le RFC 7626 sont résumés dans l'annexe A. Outre des retours d'expérience basés sur le déploiement de solutions minimisant et/ou chiffrant les données, ils consistent essentiellement en remarques négatives sur le chiffrement, défendant le mécanisme traditionnel de résolution DNS. Il n'y avait d'ailleurs pas forcément d'urgence à sortir un nouveau RFC. Comme le demandait Alissa Cooper, l'auteure du RFC 6973, « "*Why not wait to see how QUIC, DOH, ADD, ODNS, etc. shake out in the next few years and take this up then ?*" ». C'est en raison de ces changements négatifs que je ne

suis pas cité comme auteur du RFC (contraitement à son prédécesseur). Sara Dickinson, qui avait géré l'essentiel du travail pour les débuts de ce nouveau RFC, s'est également retirée, en raison de la dureté des polémiques qui ont déchiré l'IETF sur ce document.

Depuis la sortie du RFC 7626, il y a eu un certain nombre de déploiement de techniques améliorant la protection de la vie privée dans le cas du DNS, notamment la minimisation des données (RFC 9156) et le chiffrement (DoT, RFC 7858, DoH, RFC 8484, et le futur DoQ <https://datatracker.ietf.org/doc/draft-ietf-dprive-dnsoquic/>). Cela permet des retours d'expérience. Par exemple, sur l'ampleur du déploiement de la minimisation <https://labs.ripe.net/Members/wouter_de_vries/make-dns-a-bit-more-private-with-qname-minimisation>, ou sur les performances du chiffrement (voir « "*An End-to-End, Large-Scale Measurement of DNS-over-Encryption*" <https://www.zhangmingming.org/publication/imc19-doh/> »). Il y a eu également des discussions politiques à propos de ces techniques et de leur déploiement, souvent malhonnêtes ou confuses <https://www.bortzmeyer.org/doh-et-ses-adversaires.html> (et ce RFC en contient plusieurs) comme l'accusation de « centralisation », qui sert surtout à protéger l'état actuel, où l'opérateur de votre réseau d'accès à l'Internet peut à la fois vous surveiller et contrôler ce que vous voyez, via sa gestion de la résolution DNS. Mais il est vrai que rien n'est simple en matière de sécurité et que DoH (mais pas DoT) est trop bavard, puisqu'il hérite des défauts de HTTP, en envoyant trop d'informations au serveur.

Le RFC essaie de décourager les utilisateurs d'utiliser le chiffrement (dans l'esprit du RFC 8404) en notant que ce n'est pas une technologie parfaite (par exemple, l'observation des métadonnées peut donner des indications ou, autre exemple, une mauvaise authentification du résolveur peut vous faire envoyer vos requêtes à un méchant). C'est vrai mais c'est le cas de toutes les solutions de sécurité et on ne trouve pas de tels avertissements dans d'autres RFC qui parlent de chiffrement. L'insistance des opérateurs pour placer ces messages anti-chiffrement dans ce RFC montre bien, justement, l'importance de se protéger contre la curiosité des opérateurs.

Le RFC remarque également qu'il n'existe pas actuellement de moyen de découvrir automatiquement le résolveur DoT ou DoH. Mais c'est normal : un résolveur DoT ou DoH annoncé par le réseau d'accès, par exemple via DHCP, n'aurait guère d'intérêt, il aurait les mêmes défauts (et les mêmes qualités) que le résolveur traditionnel. DoT et DoH n'ont de sens qu'avec un résolveur configuré statiquement. Le RFC cite des déploiements de résolveurs DNS avec chiffrement faits par plusieurs FAI (aucun en France, notons-le). Cela montre une grosse incompréhension du problème : on ne chiffre pas pour le plaisir de chiffrer, mais parce que cela permet d'aller de manière sécurisée vers un autre résolveur. Chiffrer la communication avec le résolveur habituel ne fait pas de mal mais on ne gagne pas grand'chose puisque ce résolveur a les mêmes capacités de surveillance et de modification des réponses qu'avant. Notons que le RFC, inspiré par la propagande des telcos, raconte que le risque de surveillance est minimisé par le chiffrement du lien radio dans les réseaux 4G et après. Cela oublie le fait que la surveillance peut justement venir de l'opérateur.

Plus compliqué est le problème de l'endroit où se fait cette configuration. Traditionnellement, la configuration du résolveur à utiliser était faite globalement par le système d'exploitation. Ce n'est pas du tout une règle du protocole DNS (les RFC normalisant le DNS ne contrôlent que ce qui circule sur le câble, pas les choix locaux de chaque machine) et il serait donc absurde d'invoquer un soi-disant principe comme quoi le DNS serait forcément géré au niveau du système. Certains déploiements de DoH (celui de Firefox, par exemple), mettent le réglage dans l'application (en l'occurrence le navigateur Web). La question est discutable et discutée mais, de toute façon, elle ne relève pas de l'IETF (qui normalise ce qui passe sur le réseau) et il est anormal que le RFC en parle.

Le temps passé depuis le RFC 7626 a également permis la mise en service d'un plus grand nombre de résolveurs publics, comme Quad9 <https://www.bortzmeyer.org/quad9.html> ou comme celui de Cloudflare. Mais il n'y a pas que les grosses boîtes états-uniennes qui gèrent des résolveurs DNS

publics. Ainsi, je gère moi-même un modeste résolveur public <https://www.bortzmeyer.org/doh-bortzmeyer-fr-policy.html>. Le très intéressant RFC 8932 explique le rôle des politiques de vie privée dans ces résolveurs publics et comment en écrire une bonne. Bien sûr, comme toujours en sécurité, la lutte de l'épée contre la cuirasse est éternelle et des nouvelles actions contre la vie privée et la liberté de choix apparaissent, par exemple des réseaux qui bloquent l'accès à DoT (en bloquant son port 853) ou aux résolveurs DoH connus (DoH a été développé en partie pour être justement plus difficile à bloquer).