RFC 7626 : DNS privacy considerations

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 27 août 2015

Date de publication du RFC : Août 2015

https://www.bortzmeyer.org/7626.html

Les révélations d'Edward Snowden en juin 2013 ont fait comprendre à tous, même aux plus obtus qui s'obstinaient à nier l'ampleur et la gravité de la surveillance généralisée, que les problèmes de vie privée ne sont pas un simple détail. Au contraire, il est maintenant établi que l'Internet, et les technologies numériques en général, sont utilisés comme arme de surveillance massive par le gouvernement des États-Unis et, quasi-certainement, par plusieurs autres, dans la mesure de leurs moyens. Cela a mené l'IETF, qui avait parfois tendance à relativiser l'intensité du problème, à se pencher sérieusement sur la question. Plusieurs travaux ont été lancés. L'un d'eux concerne la protection de la vie privée lors de l'utilisation du DNS et ce nouveau RFC, écrit par votre serviteur, est son premier résultat.

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 [1], 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 (c'est le sous-projet "*qname minimization*") et les rendre plus résistantes à l'écoute, via le chiffrement. La diminution des données est étudiée dans le groupe de travail IETF dnsop <https://tools.ietf.org/wg/dnsop> (le principal document est "*DNS query name minimisation to improve privacy*" <https://tools.ietf.org/id/draft-ietf-dnsop-qname-minimisation>) et le chiffrement dans le groupe de travail IETF dprive <https://tools.ietf.org/wg/dprive> (le principal document est "*TLS for DNS : Initiation and Performance Considerations*" <https://tools.ietf.org/id/draft-ietf-dprive-start-tls-for-d qui propose de faire circuler les requêtes DNS sur TLS). Le groupe dprive (le nom, contraction de "*DNS privacy*", suit la tradition des noms de groupes humoristiques, ce qui est fréquent à l'IETF, et est dû à

1. Pour voir le RFC de numéro NNN, https://www.ietf.org/rfc/rfcNNN.txt, par exemple https://www.ietf.org/rfc/rfc6973.txt

Alexander Mayrhofer) est également l'enceinte où a été développé le RFC 7626, objet principal de cet article.

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, 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 étaient largement ignorées (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 TCP/IP ordinaire, typiquement porté par UDP. Il peut être écouté, et comme il n'est aujourd'hui pas 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.

La section 2 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.

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).

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 presque toujours 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.) 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 connecter 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 3 du RFC) mais il n'est pas difficile d'imaginer des usages moins sympathiques comme dans le cas du programme NSA MORECOWBELL <https://gnunet.org/morecowbell>.

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.

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 français 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.

Quelques élements d'histoire de ce RFC pour finir, puisque j'ai vécu tout son cycle de vie (c'est mon premier RFC publié). Sa première version a été entièrement écrite, dans un avion de la KLM entre Amsterdam et Vancouver, où j'allais à une réunion IETF <https://www.bortzmeyer.org/ietf-securite-espionnage-bis.html> en novembre 2013. Long vol sans connexion Internet, donc sans distractions, sièges de la classe affaires et présence de plusieurs collègues qui vous stimulent et avec qui discuter (merci au passage à Olaf Kolkman pour avoir insisté « il faut que tu publies cela »), cela fait d'excellentes conditions de travail.

Le RFC venant d'être publié, on voit qu'il aura fallu moins de deux ans en tout, alors qu'on raconte souvent que l'IETF est lente (sur le fonctionnement de l'IETF, voir mon exposé à JRES <https://www.

bortzmeyer.org/expose-ietf-a-jres.html>). Pour avoir une idée des changements entre le premier « draft KLM » et l'actuel RFC, vous pouvez voir un calcul automatique des différences. <http://tools.ietf.org//rfcdiff?url1=draft-bortzmeyer-perpass-dns-privacy-00&url2=http://www.rfc-editor.org/rfc/rfc7626.txt>

Quelques autres lectures utiles :

— Le début du travail au CENTR <https://centr.org/LR41-Boulvard-Bortzmeyer-DNS_data_sharing> à Amsterdam en juin 2013 (juste avant Snowden...),
— Un exposé à l'OARC <https://indico.dns-oarc.net/event/19/contribution/0/material/slides/0.pdf> à Varsovie en mai 2014 pendant que le travail était en cours,
— Un article de synthèse dans le numéro 79 de la revue MISC <http://boutique.ed-diamond.com/home/859-misc-79.html>, donc plutôt pour techniciens de la sécurité,
— Un article plus grand public sur le blog de l'AFNIC <https://www.afnic.fr/fr/ressources/blog/vers-un-dns-moins-indiscret-3.html>.
— Une discussion sur LinuxFr <https://linuxfr.org/users/bortzmeyer/journaux/creation-du-group