

Votre serveur DNS peut-il faire passer des paquets de toutes les tailles ?

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 13 juillet 2009. Dernière mise à jour le 27 janvier 2010

<https://www.bortzmeyer.org/dns-size.html>

Il y a très longtemps, lorsque les ministres ne parlaient pas de l'Internet à la télévision, la taille des paquets DNS était limitée à 512 octets (RFC 1035¹, section 2.3.4). Cette limite a été supprimée par le RFC 2671 en 1999. Mais dix ans sont une durée très courte pour le conservatisme de certains et beaucoup d'administrateurs réseaux qui mettent à jour leur version de Flash toutes les deux semaines n'ont jamais vérifié la configuration de leur pare-feu. L'OARC <<http://www.dns-oarc.net/>> vient donc de publier un excellent outil <<https://www.dns-oarc.net/oarc/services/replysizetest>> qui permet de tester facilement si votre environnement DNS est correct.

L'outil en question consiste simplement en un serveur DNS spécifique, qu'on peut interroger (le nom est `rs.dns-oarc.net` et le type est TXT) avec n'importe quel client DNS. Voici un exemple avec dig :

```
% dig +short rs.dns-oarc.net txt
rst.x4001.rs.dns-oarc.net.
rst.x3985.x4001.rs.dns-oarc.net.
rst.x4023.x3985.x4001.rs.dns-oarc.net.
"192.168.1.1 sent EDNS buffer size 4096"
"192.168.1.1 DNS reply size limit is at least 4023 bytes"
```

Ici, on voit que les réponses DNS de 4023 octets peuvent arriver.

Avec des résolveurs DNS mal configurés ou mal connectés (ici, ceux d'Alice, via une AliceBox) :

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc1035.txt>

```
% dig +short rs.dns-oarc.net txt
rst.x486.rs.dns-oarc.net.
rst.x454.x486.rs.dns-oarc.net.
rst.x384.x454.x486.rs.dns-oarc.net.
"213.228.63.58 lacks EDNS, defaults to 512"
"213.228.63.58 DNS reply size limit is at least 486 bytes"
```

486 octets arrivent à passer, ce qui est insuffisant dans de nombreux cas.

Malheureusement, les environnements DNS bogués sont fréquents. Ce n'est pas forcément à cause du serveur de noms : cela peut être à cause d'un pare-feu stupidement configuré pour refuser les paquets DNS de plus de 512 octets ou bien par la faute d'une "middlebox" programmée avec les pieds en Chine. D'où l'importance un tel système de test.

Vous pouvez le tester sur votre machine habituelle, au bureau et à la maison. Si vous obtenez moins de 2048 octets, danger : DNSSEC, pour ne citer que lui, va vous causer des problèmes. La plupart des technologies vieilles de moins de dix ans (comme IPv6) auront également des problèmes.

Si les paquets de plus de 512 passent, mais pas ceux de plus de 1500, il s'agit probablement d'un problème de MTU (avec filtrage stupide de l'ICMP) ou bien d'un pare-feu qui ne gère pas les fragments. Par exemple, sur un Cisco, il faut penser <http://www.cisco.com/en/US/docs/ios/12_3t/12_3t8/feature/guide/gt_vfrag.html> à mettre `ip virtual-reassembly` **avant** le passage par un pare-feu (il y en a beaucoup) qui ne gère pas les fragments.

Le même logiciel a aussi été installé par le RIPE-NCC sur ses serveurs donc on peut aussi comparer avec :

```
% dig +short test.rs.ripe.net txt
rst.x3828.rs.ripe.net.
rst.x3833.x3828.rs.ripe.net.
rst.x3839.x3833.x3828.rs.ripe.net.
"192.134.4.162 sent EDNS buffer size 4096"
"192.134.4.162 summary bs=4096,rs=3839,edns=1,do=1"
"192.134.4.162 DNS reply size limit is at least 3839 bytes"
```

Si on est sur une machine qui n'a pas dig mais qui a nslookup (cas du logiciel privateur MS-Windows, on peut probablement tester (je n'ai pas regardé en détail donc je ne garantis rien) :

```
% nslookup -q=txt rs.dns-oarc.net
Server:          ::1
Address:         ::1#53

Non-authoritative answer:
rs.dns-oarc.net canonical name = rst.x3827.rs.dns-oarc.net.
rst.x3827.rs.dns-oarc.net canonical name = rst.x3837.x3827.rs.dns-oarc.net.
rst.x3837.x3827.rs.dns-oarc.net canonical name = rst.x3843.x3837.x3827.rs.dns-oarc.net.
rst.x3843.x3837.x3827.rs.dns-oarc.net text = "192.134.4.69 DNS reply size limit is at least 3843"
rst.x3843.x3837.x3827.rs.dns-oarc.net text = "192.134.4.69 sent EDNS buffer size 4096"
rst.x3843.x3837.x3827.rs.dns-oarc.net text = "Tested at 2010-01-27 10:20:02 UTC"

Authoritative answers can be found from:
x3843.x3837.x3827.rs.dns-oarc.net nameserver = ns00.x3843.x3837.x3827.rs.dns-oarc.net.
ns00.x3843.x3837.x3827.rs.dns-oarc.net internet address = 149.20.58.136
```

D'autres outils pour faire ce genre de test existent mais plus complexes :

- Une application à télécharger en local <<http://labs.ripe.net/content/testing-your-resolver-dns-r>> qui nécessite Sun Java,
- Un autre outil qui dépend d'un simple nom de domaine et est donc utilisable par dig est `pathtest.isc.org`: tapez `dig $RECORDSIZE.pathtest.isc.org TXT +bufsize=$YOURBUFFERSIZE` et le serveur faisant autorité générera pour vous un enregistrement de type TXT et de taille \$RECORDSIZE. Vous verrez bien si vous arrivez à le récupérer...
- L'outil Netalyzr <<http://netalyzr.icsi.berkeley.edu/>>, qui nécessite Sun Java,
- Des instructions détaillées, uniquement avec dig, par Mark Andrews (le mainteneur de BIND), sur la liste des utilisateurs BIND <<https://lists.isc.org/pipermail/bind-users/2010-February/078755.html>> et une variante plus synthétique sur la liste Nanog <<http://mailman.nanog.org/pipermail/nanog/2010-February/018363.html>>.

La racine du DNS devant être complètement signée en mai 2010 <<https://www.bortzmeyer.org/la-racine-commence-signature.html>>, cette question de taille :-> se posera donc souvent, comme dans l'article « *Preparing K-root for a Signed Root Zone* » <<http://labs.ripe.net/content/preparing-k-root-signed-root-zone>> ou dans les excellentes mesures de « *Measuring DNS Transfer Sizes - First Results* » <<http://labs.ripe.net/content/measuring-dns-transfer-sizes-first-resu>>

Si le test montre que les paquets de plus de 1500 octets (voire les paquets de plus de 512 octets) ne peuvent pas passer, faut-il paniquer ? Oui et non. Le fait qu'ils ne peuvent pas passer est inquiétant, dix ans après que la vieille limite de 512 octets aie été supprimée. Mais cela n'entraînera pas forcément une catastrophe. Cela dépend de beaucoup de choses <<https://www.bortzmeyer.org/risques-reels-dns-limite.html>>. Par exemple, si le résolveur ne gère pas du tout EDNS (ici, Google DNS <<https://www.bortzmeyer.org/google-dns.html>>) :

```
% dig @8.8.8.8 +short test.rs.ripe.net txt
rst.x477.rs.ripe.net.
rst.x481.x477.rs.ripe.net.
rst.x486.x481.x477.rs.ripe.net.
"209.85.228.94 DNS reply size limit is at least 486 bytes"
"209.85.228.94 lacks EDNS, defaults to 512"
"209.85.228.94 summary bs=512,rs=486,edns=0,do=0"
```

Certes, c'est une limite anormale en 2010 mais cela ne cassera sans doute pas lors de la signature de la racine : si le résolveur n'émet pas de paquets EDNS, il ne pourra pas mettre le bit DO à 1 (RFC 3225, section 3) et ne recevra donc pas les grosses signatures DNSSEC.

Autre cas intéressant, celui de Free Les clients de ce FAI ont des résultats apparemment incohérents, un coup ça marche, un coup ça ne marche pas. C'est parce que derrière chaque résolveur de Free (par exemple 212.27.40.241), il y a plusieurs machines (c'est banal), qui ont des configurations différentes (c'est assez étonnant, et je ne vois pas de raison valable de faire cela). Notez bien que l'adresse IP dans la réponse du test est différente à chaque fois, montrant bien que plusieurs machines se partagent le travail.