Fonctionnement du DNS : une explication fausse mais courante

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 29 mars 2013

https://www.bortzmeyer.org/fonctionnement-dns.html

Hier, je lisais un article <http://www.theregister.co.uk/2013/03/28/i_accidentally_the_internet/> qui contenait une explication du fonctionnement du DNS qui est **fausse**. Mais comme cette explication est courante, je pense utile de documenter pourquoi elle est fausse et pourquoi c'est important.

L'article en question <http://www.theregister.co.uk/2013/03/28/i_accidentally_the_internet/> est intéressant (un témoignage concret sur une erreur de configuration courante <https://www.bortzmeyer.org/fermer-les-recursifs-ouverts.html>) mais il a le tort de vouloir expliquer le fonctionnement du DNS sans le connaître vraiment. (Le journal qui le publie, The Register, est coutumier de ce genre de bavures. Il vise le sensationnel à forte composante technique mais ne font jamais vérifier leurs articles.) Que dit l'article ? « *"In a recursive configuration the DNS server asks the list of root servers it has preconfigured who owns the ".com" domain. It then asks the .com servers who owns "google.com". It then asks "google.com" who owns "www.google.com" and delivers that address back to you.*" » Cette explication est courante (je tape sur The Register mais ils ne sont pas les seuls, loin de là) mais inexacte.

Non, le serveur récursif (le **résolveur**) ne demande pas à la racine « Qui gère `.com` ? » Il ne demande pas aux serveurs de Verisign (registre de `.com`) « Qui gère `google.com` ? » À chaque étape, le résolveur envoie la demande complète, le FQDN original. La question est toujours « Quelle est l'adresse de `www.google.com` ? » Et la réponse peut être une adresse IP (quand on est arrivé chez Google) ou bien une référence ("*referral*") : « Je ne sais pas mais demande à Verisign ».

Alors, est-ce du pur pinaillage de ma part que de noter cette erreur fréquente ? Non, car cela a une conséquence importante : les serveurs **faisant autorité** (ceux de la racine, des TLD, etc, qui connaissent les réponses et qui sont interrogés par les résolveurs) reçoivent un nom complet. Cela leur permet de faire des études, des analyses, des recherches qui ne seraient pas possible autrement. On voit ainsi passer sur ces serveurs des noms qui sont parfois révélateurs, contenant par exemple le type d'application utilisée (« Quelle est l'adresse IP de `_bittorrent-tracker._tcp.XXXX.abo.wanadoo.fr` ? »).

Pourquoi est-ce que les résolveurs procèdent ainsi ? Ne respecterait pas t-on mieux la vie privée en n'envoyant que le dernier composant du nom, puis les deux derniers, etc ? Mais on ne peut pas : le résolveur ne connait pas le découpage du DNS en zones (ce découpage ne se fait **pas** sur les points entre les composants d'un nom). Il ne sait pas si la racine fait autorité pour `.com` ou non, par exemple (pendant longtemps, c'était le cas). Pour prendre un exemple actuel, si un résolveur cherche `www.redressement-productif.` Demander aux serveurs de l'AFNIC (registre de `.fr`) qui gère `gouv.fr` ne marcherait pas : `gouv.fr` n'est pas une zone séparée, il est dans la zone `.fr`.

Pour ceux et celles qui veulent expérimenter avec dig, voici la vraie requête envoyée par un résolveur à la racine pour `www.redressement-productif.gouv.fr` :

```
% dig @a.root-servers.net A www.redressement-productif.gouv.fr
...
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 12929
;; flags: qr rd; QUERY: 1, ANSWER: 0, AUTHORITY: 7, ADDITIONAL: 11

;; AUTHORITY SECTION:
fr. 172800 IN NS e.ext.nic.fr.
fr. 172800 IN NS d.ext.nic.fr.
fr. 172800 IN NS f.ext.nic.fr.
fr. 172800 IN NS d.nic.fr.
fr. 172800 IN NS g.ext.nic.fr.
fr. 86400 IN DS 1336 8 2 C7AEE4D32904728741DB270E72899673D7DFAF212E517F2400C9424B 231CF56B
fr. 86400 IN RRSIG DS 8 1 86400 20130405000000 20130328230000 40323 . VMBW3MNHCI9LJRU+SptuwnO3pJgvvaMHCE11AF

;; ADDITIONAL SECTION:
d.ext.nic.fr. 172800 IN A 192.5.4.2
d.ext.nic.fr. 172800 IN AAAA 2001:500:2e::2
d.nic.fr. 172800 IN A 194.0.9.1
d.nic.fr. 172800 IN AAAA 2001:678:c::1
e.ext.nic.fr. 172800 IN A 193.176.144.22
e.ext.nic.fr. 172800 IN AAAA 2a00:d78:0:102:193:176:144:22
f.ext.nic.fr. 172800 IN A 194.146.106.46
f.ext.nic.fr. 172800 IN AAAA 2001:67c:1010:11::53
g.ext.nic.fr. 172800 IN A 194.0.36.1
g.ext.nic.fr. 172800 IN AAAA 2001:678:4c::1
```

Le serveur racine ne connaissait pas la réponse (`ANSWER: 0`) mais a redirigé vers les serveurs de `.fr`. On peut alors continuer, et c'est vraiment ce que fait le résolveur. Vous ne me croyez pas ? Regardez les statistiques publiques d'un serveur faisant autorité, par exemple le serveur racine K <`http://k.root-servers.org/statistics/ROOT/qtypes.html`>. On voit que les requêtes de type NS (« qui gère tel domaine ? ») sont en quantité négligeable, le gros du trafic étant fait de requêtes A (adresses IPv4), AAAA (adresses IPv6) et MX (relais de courrier électronique).

Autre solution pour vérifier : utilisez tcpdump sur un résolveur qui vient de démarrer (pour avoir un cache vide) :

```
09:32:33.282971 IP (tos 0x0, ttl 64, id 7716, offset 0, flags [none], proto UDP (17), length 91)
    192.168.2.4.61342 > 193.0.14.129.53: [bad udp cksum 0x9286 -> 0x8ef4!] 10282% [1au] A? www.redressement-
09:32:33.328175 IP (tos 0x0, ttl 54, id 43405, offset 0, flags [none], proto UDP (17), length 606)
    193.0.14.129.53 > 192.168.2.4.61342: [udp sum ok] 10282- q: A? www.redressement-productif.gouv.fr. 0/7/1

09:32:33.335300 IP6 (hlim 64, next-header UDP (17) payload length: 71) 2a01:e35:8bd9:8bb0:ba27:ebff:feba:909
09:32:33.401859 IP6 (hlim 56, next-header UDP (17) payload length: 623) 2a00:d78:0:102:193:176:144:22.53 > 2

09:32:33.695786 IP (tos 0x0, ttl 64, id 28675, offset 0, flags [none], proto UDP (17), length 91)
    192.168.2.4.36882 > 193.108.167.41.53: [bad udp cksum 0x2b9b -> 0x546e!] 10536% [1au] A? www.redressemen
09:32:33.765944 IP (tos 0x0, ttl 48, id 33891, offset 0, flags [DF], proto UDP (17), length 260)
    193.108.167.41.53 > 192.168.2.4.36882: [udp sum ok] 10536*- q: A? www.redressement-productif.gouv.fr. 1/
```

Ici, 193.0.14.129 est k.root-servers.net, un des serveurs de la racine. On voit bien qu'il a reçu la question complète, avec le nom intégral. 2a00:d78:0:102:193:176:144:22 est e.ext.nic.fr, un des serveurs de .fr. Notez que sa réponse n'a pas donné les serveurs de gouv.fr (il n'y en a pas) mais directement ceux de redressement-productif.gouv.fr. Quant à 193.108.167.41, c'est un des serveurs de nom du ministère et il donne enfin autre chose qu'une référence.