À partir d'un nom de domaine, trouver le domaine « responsable »

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 15 mars 2013

https://www.bortzmeyer.org/trouver-domaine-responsable.html

C'est un problème qui revient souvent sur l'Internet. On a un nom de domaine et on voudrait trouver quel est le « nom responsable » ou « nom enregistré ». Par exemple, étant donné `machin.truc.bortzmeyer.org`, le nom responsable est `bortzmeyer.org`. Cela parait trivial, mais cela ne l'est pas, comme cela a encore été illustré cette semaine dans les discussions à la réunion IETF d'Orlando.

D'abord, voyons **pourquoi** on veut faire cela. Il existe de nombreux cas où c'est utile (merci à John Levine pour une compilation) :

— Le cas le plus connu est celui des "*cookies*" du Web (RFC 6265 [1]). Le serveur peut indiquer pour quel domaine un "*cookie*" est valable, par exemple le serveur en `machin.example.org` peut dire que ses "*cookies*" sont valables pour tout `example.org` et que le client devra donc les envoyer à, mettons, le serveur `truc.example.org`. Mais un serveur en `example.com` peut-il dire que ses "*cookies*" valent pour tout `.com` ? Évidemment non, car `.com` n'est pas sous la même administration. Mais comment le navigateur Web pourrait-il le savoir ?
— Certains mécanismes d'authentification partent d'un domaine et cherchent le domaine responsable pour définir une politique pour les sous-domaines. C'est le cas de DMARC. Ou, autre exemple, des autorités de certification X.509 qui, lorsqu'on leur demande un certificat pour `www.example.net` vont interroger le titulaire de `example.net` pour vérifier si cette demande est autorisée. Et, si vous demandez un certificat pour `*.com`, il devra normalement être refusé.
— En cas d'abus, on cherche à prévenir le responsable. Si `machin.truc.example.org` héberge un site de hameçonnage, dois-je prévenir le titulaire de `truc.example.org` ou bien celui de `example.org` ?

1. Pour voir le RFC de numéro NNN, `https://www.ietf.org/rfc/rfcNNN.txt`, par exemple `https://www.ietf.org/rfc/rfc6265.txt`

Certaines personnes, arrivées là, se demandent où est le problème. Il suffit de retirer le **composant** le plus à gauche du nom de domaine et on trouve le domaine responsable. Un problème avec `www.example.com`? Le responsable est `example.com`. Mais cet algorithme trivial ne marche pas dans des cas comme `truc.machin.example.com` qui est parfaitement légal (pour un exemple réel, regardez le site Web de la mairie d'Hiroshima <`http://www.city.hiroshima.lg.jp/`>, en `www.city.hiroshima.l`
Bon, pas grave, se disent les « simplificateurs ». Ne gardons que les deux composants les plus à droite. Comme cela, le responsable de `www.example.com` est `example.com` et celui de `truc.machin.example.com` est aussi `example.com`. Mais cela ne marche pas car tous les TLD ne font pas de délégation au deuxième niveau : `.uk` et `.jp` ne délèguent qu'au troisième niveau (l'Université de Cambridge est en `cam.ac.uk`, pas `cam.uk`). Et il existe des TLD qui délèguent au deuxième niveau **et** au troisième, comme `.fr` ou comme tout TLD qui a un registre parmi ses clients (par exemple `eu.org` : `machin.eu.org` n'est pas sous la même administration que `truc.eu.org`). Donc, un algorithme portant sur le nom n'est **pas** une solution.

D'autres personnes, plus techniques, se disent qu'il suffit de faire une requête DNS et de regarder le domaine de l'enregistrement SOA, qui donnera le nom du domaine responsable (je vous épargne quelques petits détails techniques comme le fait que le SOA n'est renvoyé que si les données demandées n'existent pas) :

```
% dig A foo.bar.doesnotexist.cam.ac.uk
...
;; ->>HEADER<<- opcode: QUERY, status: NXDOMAIN, id: 39210
...
;; AUTHORITY SECTION:
cam.ac.uk. 10800 IN SOA authdns0.csx.cam.ac.uk. hostmaster.ucs.cam.ac.uk. 1363381148 14400 3600 604800 1440
```

On voit que ce nom n'existe pas et que l'autorité est en `cam.ac.uk`. Mais la notion d'« administration » n'est pas technique. Cette méthode DNS ne détecte que les frontières techniques entre les zones DNS (ce qu'on nomme les "*zone cuts*"), pas les politiques appliquées.

```
% dig A www.doesnotexist.asso.fr
...
;; ->>HEADER<<- opcode: QUERY, status: NXDOMAIN, id: 16619
...
;; AUTHORITY SECTION:
fr. 5312 IN SOA nsmaster.nic.fr. hostmaster.nic.fr. 2222270543 3600 1800 3600000 5400
```

Comme `asso.fr` est dans la même zone DNS que `.fr`, on peut avoir l'impression que les délégations se font sous `.fr` (alors qu'il y en a sous `asso.fr`).

Bref, il n'y a pas de solution algorithmique. Il faut récupérer l'information quelque part. Il existe plusieurs listes compilées à la main et qui ont toutes en commun d'être insuffisantes, non officielles et pas à jour. Ce dernier problème est amené à s'aggraver avec le programme de création de nouveaux gTLD. La plus connue de ces listes est gérée par Mozilla et se nomme la "*Public Suffix List*" <`http://publicsuffix.org/`>. Elle est notamment utilisée par les navigateurs Firefox et Chrome. Cela a provoqué des frustrations comme lorsque `.cw` n'était pas reconnu par Chrome <`https://lists.dns-oarc.net/pipermail/dns-operations/2013-January/009637.html`> qui envoyait les visiteurs sur un moteur de recherche !

`https://www.bortzmeyer.org/trouver-domaine-responsable.html`

Alors, quelles solutions ont été discutées à l'IETF ? Elles avaient toutes en commun de permettre aux gérants des zones DNS de publier l'information (contrairement à la "*Public Suffix List*", gérée par des tiers qui ne sont pas forcément au courant). La plus aboutie, due à Andrew Sullivan et décrite dans l'"*Internet Draft*" `draft-sullivan-domain-origin-assert`, consiste à placer des enregistrements d'un nouveau type, SOPA (pour "*Start Of Policy Authority*") qui déclarent « ce nom est sous la même autorité que moi ». Ainsi, si on a cet enregistrement SOPA :

```
bortzmeyer.org. 86400 IN SOPA www.bortzmeyer.org.
```

et qu'un serveur Web accédé sous le nom `www.bortzmeyer.org` va tenter de placer un "*cookie*" pour le nom `bortzmeyer.org`, le navigateur Web pourra vérifier que `bortzmeyer.org` avait donné son autorisation (en se déclarant sous la même administration que `www.bortzmeyer.org`). Cette solution est donc très générale (on peut imaginer des schémas très rigolos comme `example.net` déclarant que `example.com` est sous la même autorité) mais cela peut être un peu complexe parfois.

Une autre proposition plus simple (mais pas encore documentée dans un texte écrit), moins générale mais collant mieux à la nature arborescente du DNS, est de permettre à toute zone DNS de dire « je suis un point de délégation public, c'est-à-dire que les noms en dessous ne sont pas sous mon autorité ». Ainsi, `fr`, `lg.jp` ou `eu.org` pourraient dire (le nom de l'enregistrement, PHB, vient des initiales de l'auteur, Phillip Hallam-Baker, puisque la proposition n'est pas encore écrite) :

```
org.   86400 IN  PHB PUBLIC
```

et tout le monde pourrait alors savoir que `.org` est une zone de délégation publique. Les sous-domaines de `.org` pourraient ne rien mettre, mettre un enregistrement disant explicitement qu'ils ne sont pas une zone de délégation publique, ou bien, comme `eu.org`, avoir un enregistrement disant qu'eux aussi délèguent.

(Une autre solution proposée a été d'utiliser un autre protocole que le DNS, par exemple le successeur de whois sur lequel travaille le groupe WEIRDS.)

A priori, pour l'instant, les participants à l'IETF penchent pour la solution SOPA. (À noter que, trois mois après cet article, une autre proposition technique, plus simple, a été faite, en `draft-levine-orgboundary`.)

`https://www.bortzmeyer.org/trouver-domaine-responsable.html`