Censure administrative du Web en France, un premier regard technique

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 15 mars 2015. Dernière mise à jour le 17 mars 2015

http://www.bortzmeyer.org/censure-francaise.html

Le décret n° 2015-125 du 5 février 2015 « relatif au blocage des sites provoquant à des actes de terrorisme ou en faisant l'apologie et des sites diffusant des images et représentations de mineurs à caractère pornographique » <http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000030195477&dateTexte=&categorieLien=id> est désormais entré en application. Cela fait quoi, en pratique ? Regardons un peu ce qui se passe pour un site Web censuré.

Prenons l'exemple de http://islamic-news.info/ dont, je suppose, d'après son nom de domaine, qu'il s'agit d'un site de propagande intégriste (je n'ai pas pu le consulter, pas à cause de la censure française, facile à contourner, mais parce qu'il semble en panne). Si on essaie de le consulter depuis un FAI français typique, on obtient à la place une page Web avec une main rouge menaçante (une référence historique très malencontreuse, pour une décision étatique, il aurait mieux valu utiliser une main de justice ; la fameuse main rouge a été supprimée vers le 24 mars) et un texte TOUT EN MAJUSCULES « VOUS AVEZ ÉTÉ REDIRIGÉ VERS CE SITE OFFICIEL CAR VOTRE ORDINATEUR ALLAIT SE CONNECTER À UNE PAGE DONT LE CONTENU PROVOQUE À DES ACTES DE TERRORISME OU FAIT PUBLIQUEMENT L'APOLOGIE D'ACTES DE TERRORISME » (ce texte a été légèrement modifié vers le 24 mars). Comment est-ce réalisé ?

Les différents textes officiels sur ces mesures de censure n'imposaient pas aux FAI une technique particulière mais les conditions de mise en œuvre (liste noire de noms de domaine « Les adresses électroniques figurant sur la liste [noire] comportent soit un nom de domaine (DNS), soit un nom d'hôte caractérisé par un nom de domaine précédé d'un nom de serveur », obligation de rediriger vers « CE SITE OFFICIEL » situé au ministère de l'Intérieur), fait que la solution la plus simple est de mettre en place un résolveur DNS menteur <http://www.bortzmeyer.org/dns-menteur.html>. Ce résolveur est la machine qui répond normalement aux requêtes DNS de l'utilisateur, en lui donnant l'adresse IP du serveur à contacter. (Au passage, une explication plus grand public du DNS avait été faite par Andréa Fradin <http://owni.fr/2012/07/05/internet-par-la-racine/>.) Essayons avec dig depuis une machine au Japon, pour voir la vraie adresse IP :

```
% dig +short A islamic-news.info
37.59.14.72
```

Et depuis un FAI français (Free, mais on verra plus loin qu'il n'est pas le seul) :

```
% dig +short A islamic-news.info
90.85.16.52
```

L'adresse obtenue est différente, il y a bien un résolveur DNS menteur. Si on utilise un résolveur DNS public comme Google Public DNS <http://www.bortzmeyer.org/google-dns.html>, on a également la vraie adresse IP :

```
% dig @8.8.4.4 +short A islamic-news.info
37.59.14.72
```

Est-ce bien ce trompeur `90.85.16.52` qui est le site officiel servant la « main rouge » ? Testons-le en HTTP. Il y a plusieurs méthodes pour cela mais j'ai utilisé une des plus simples, mettre dans mon fichier local `/etc/hosts` l'adresse de ce site pour un nom bidon :

```
% cat /etc/hosts
...
90.85.16.52 front-liberation-potamocheres.example
```

Et, en visitant `http://front-liberation-potamocheres.example/`, j'ai bien la page à la main rouge. On peut bien sûr tester directement `http://90.85.16.52/` mais des astuces techniques utilisées par le ministère font que cela ne marche pas toujours si le champ `Host:` indique une adresse IP, et qu'on récupère parfois un 403 "*Forbidden*". Si vous ne voulez pas modifier votre `/etc/hosts` comme moi, vous pouvez vous servir des noms créés par Pierre Beyssac, , ou , ou encore les noms dans le domaine de Laurent Penou, `gsec.ovh`. Commme il y a un joker dans ce nom de domaine, n'importe quel nom convient, comme `cazeneuve-a-raison.gsec.ovh`.

Si vous voulez tester avec d'autres noms de domaine censurés, vous pouvez essayer `jihadmin.com`, `is0lamnation.blogspot.fr`, `mujahida89.wordpress.com`, ou `alhayatmedia.wordpress.com` (sans que j'en recommande leur contenu, œuvre des crapules intégristes !).

Notez que cela veut dire que le ministère de l'Intérieur est au courant de mon intérêt pour cet animal : avec ce système, l'utilisateur est redirigé, **à son insu**, vers un serveur Web du ministère de l'Intérieur, qui aura accès, via le champ HTTP `Host:` (RFC 7230 [1], section 5.4) au nom originellement demandé. Lourde responsabilité pour le FAI qui, en configurant son résolveur DNS pour rediriger, fait cette redirection, il envoie ses clients vers un site qui saura si on aime les potamochères, le djihad ou les photos pédophiles.

Voici, vue avec tshark, la requête HTTP qu'envoie votre machine au serveur géré par le ministère de l'Intérieur, où on voit bien le champ `Host` :

1. Pour voir le RFC de numéro NNN, `https://www.ietf.org/rfc/rfcNNN.txt`, par exemple `https://www.ietf.org/rfc/rfc7230.txt`

```
Internet Protocol Version 4, Src: 192.168.2.1 (192.168.2.1), Dst: 90.85.16.52 (90.85.16.52)
Transmission Control Protocol, Src Port: 56532 (56532), Dst Port: http (80), Seq: 450, Ack: 311, Len: 468
Hypertext Transfer Protocol
    GET / HTTP/1.0\r\n
    Host: front-liberation-potamocheres.example\r\n
    User-Agent: Mozilla/5.0 (X11; Linux i686; rv:31.0) Gecko/20100101 Firefox/31.0 Iceweasel/31.4.0\r\n
    Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8\r\n
    Accept-Language: en-US,en;q=0.5\r\n
    Accept-Encoding: gzip, deflate\r\n
    Pragma: no-cache\r\n
    Connection: keep-alive\r\n
    \r\n
    [Full request URI: http://front-liberation-potamocheres.example/]
```

On voit également que **votre** adresse IP source apparait mais on y voit moins (c'est dans le reste de la connexion) que votre navigateur Web envoie également bien d'autres informations qui peuvent vous identifier, et qui sont peut-être également journalisées. (Voir le Panopticlick <https://panopticlick.eff.org/> pour une impressionnante démonstration.) Donc, ce « SITE OFFICIEL » part d'une bonne intention (rendre la censure explicite, comme à Dubaï <http://www.bortzmeyer.org/censure-a-dubai.html>) mais a des conséquences très dangereuses. Pour s'amuser un peu, dans ce monde de brutes, on peut d'ailleurs utiliser cette propriété pour envoyer un message bien senti à ceux qui lisent ces journaux, avec curl (attention, cela se fera avec **votre** adresse IP : le but est de prendre position, pas de se cacher) :

```
% curl -o /dev/null -H "Host: Halte a la censure administrative du Web" \
      -H "X-Charlie: Je suis Charlie" \
      http://90.85.16.52/\?$RANDOM
```

Une solution technique pour éviter que vos utilisateurs soient redirigés à leur insu vers un serveur HTTP qui va noter leur adresse et peut-être d'autres informations, est de bloquer les accès à ce serveur. Une requête whois montre que ce serveur est dans la plage `90.85.16.32/27` (notez que rien n'indique de lien avec le ministère de l'Intérieur, on a juste un nom, à Metz). Pour bloquer cette plage, si on utilise Shorewall <http://www.bortzmeyer.org/filtrage-avec-shorewall.html>, ce sera dans `/etc/shorewall/blacklist` quelque chose comme :

```
# 2015-03-16: may do logging of visitors (French censorship system)
90.85.16.32/27
```

Si vous utilisez directement Netfilter, ce sera une commande du genre :

```
# iptables --insert OUTPUT --destination 90.85.16.32/27  --jump REJECT
```

Notez que la CNIL avait formulé un avis sur ce point <http://legifrance.gouv.fr/affichTexte.do;jsessionid=?cidTexte=JORFTEXT000030236542&dateTexte=&oldAction=dernierJO&categorieLien=id>, et noté que « [les textes officiels ne] permet ni la collecte ni l'exploitation, par l'OCLCTIC, des données de connexion des internautes qui seraient redirigés vers la page d'information du ministère de l'intérieur. Elle [la CNIL] rappelle que si des traitements de données à caractère personnel spécifiques étaient alimentés par ces données, ils devraient être soumis à l'examen préalable de la commission. » Bref, la CNIL ne se mouille pas. Pas d'avis concret, du genre « l'adresse IP source et le champ `Host:` ne devront **pas** être journalisés", ce qui serait la moindre des choses, et laisserait quand même le ministère de l'Intérieur faire des statistiques sur le nombre de visiteurs des sites censurés. (Note technique au passage, si vous utilisez Apache, la directive pour journaliser l'adresse IP source est `%h` - ou `%a` - et celle pour le champ `Host:` est `%v`. Ce sont ces deux directives qu'il faudrait

retirer du `LogFormat` du ministère pour ne pas fliquer les utilisateurs. Cf. la documentation d'Apache <http://httpd.apache.org/docs/2.2/fr/mod/mod_log_config.html#logformat>.)

Le décret cité au début comprend plusieurs motifs de censure, et on note que l'hébergeur a prévu des sites Web différents, avec des messages différents, selon le motif. Comparez les messages en <http://90.85.16.52/> (terrorisme), <http://90.85.16.51/> (pédophilie) et <http://90.85.16.50/> (le reste, non spécifié).

Continuons avec d'autres détails techniques : le « SITE OFFICIEL » est sans doute en place depuis très peu de temps car DNSDB <http://www.bortzmeyer.org/dnsdb.html> (qui n'est **pas** temps réel, loin de là), n'a pas encore vu de noms se résolvant en `90.85.16.52` (DNSDB permet de chercher par le contenu de la réponse DNS, pas juste par le nom demandé). Il est vrai que DNSDB a probablement peu ou pas de capteurs en France (je n'ai rien vu non plus sur les bases équivalentes comme PassiveDNS.cn <http://www.bortzmeyer.org/passivedns-cn.html>.)

Quant au type de serveur HTTP utilisé, il prétend être un nginx (dans l'en-tête HTTP `Server:`) mais il ressemble plutôt à un Apache : le message d'erreur en cas d'accès refusé (testez `http://90.85.16.51/server-st` et le fait que les "*Etags*" ont le format Apache <http://stackoverflow.com/a/44939/15625> à trois champs, pas le format nginx à deux champs <http://trac.nginx.org/nginx/browser/nginx/src/http/ngx_http_core_module.c?rev=942283a53c289397131c9c2d1e0909af869fd4a4#L1802> (analyse faite par Kim-Minh Kaplan, merci).

Quelle est l'ampleur du déploiement de ces DNS menteurs ? Pour cela, on va utiliser les sondes Atlas <https://atlas.ripe.net/>, petits ordinateurs placés un peu partout, et interrogeables via leur API <http://www.bortzmeyer.org/atlas-udm.html>. Avec le programme resolve-name.py <https://github.com/RIPE-Atlas-Community/ripe-atlas-community-contrib/blob/master/resolve-name.py>, on va interroger les résolveurs DNS des sondes Atlas allemandes, pour commencer :

```
% python resolve-name.py --country=DE islamic-news.info
Measurement #1895738 for islamic-news.info/A uses 499 probes
Probe 4407 failed (trailing junk)
[] : 16 occurrences
[37.59.14.72] : 677 occurrences
Test done at 2015-03-15T16:03:37Z
```

Toutes (sauf quelques-unes qui n'ont pas pu résoudre du tout le nom) trouvent la bonne adresse. (Il y a davantage de résultats que de sondes car chaque sonde fait plusieurs essais.) Essayons en France :

```
% python resolve-name.py --country=FR islamic-news.info
Measurement #1895736 for islamic-news.info/A uses 498 probes
[] : 22 occurrences
[90.85.16.52] : 346 occurrences
[37.59.14.72] : 403 occurrences
Test done at 2015-03-15T15:39:15Z
```

À peu près la moitié des sondes Atlas en France voient la censure. Notez que les sondes Atlas ne sont pas du tout représentatives : installées par des volontaires, sans doute dans des réseaux plus "*geeks*" que la moyenne, elles ont plus de chances d'avoir un résolveur local non menteur (voir plus loin, pour cette solution anti-censure).

Par défaut, ce programme fait des requêtes de type `A` (demande d'une adresse IPv4). On peut lui demander ce qu'il en est pour les `AAAA` (demande d'une adresse IPv6). Normalement, il ne devrait pas y avoir de réponse (le nom n'a pas d'adresse IPv6 associé) mais un FAI a trouvé le moyen de mentir pour IPv6 et quel mensonge :

```
% python resolve-name.py -t AAAA -c FR islamic-news.info
Measurement #1895755 for islamic-news.info/AAAA uses 498 probes
[] : 586 occurrences
[::1] : 191 occurrences
Test done at 2015-03-15T20:25:57Z
```

Il renvoie, non pas vers la page du gouvernement, mais vers la machine locale...

Et si on regarde par AS et plus par pays :

```
% python resolve-name.py --as=3215 islamic-news.info
Measurement #1895737 for islamic-news.info/A uses 133 probes
[90.85.16.52] : 189 occurrences
[37.59.14.72] : 32 occurrences
Test done at 2015-03-15T15:59:52Z

% python resolve-name.py --as=20766 islamic-news.info
Measurement #1895739 for islamic-news.info/A uses 4 probes
[37.59.14.72] : 5 occurrences
Test done at 2015-03-15T16:06:34Z
```

Le premier AS est celui d'Orange, où il semble que la majorité des sondes voient le résolveur légal. Le second est celui de Gitoyen, dont les clients (comme FDN) ont plus de chances d'utiliser un résolveur non-menteur. (Notez toutefois que le faible nombre de sondes doit rendre prudent dans l'analyse.)

Et DNSSEC ? Est-ce qu'il résoudrait ce problème ? Non, car la validation est faite dans le résolveur. S'il est menteur, le fait de valider ne changera rien. Seul avantage, les gens qui valident avec DNSSEC ne verront pas la page d'information du ministère de l[Caractère Unicode non montré [2]]intérieur (et ne seront pas enregistrés par ledit ministère) puisque le mensonge dans le résolveur entraînera une erreur (`SERVFAIL` : `Server Failure`) sur les domaines signés (ce qui n'est de toute façon pas le cas de `islamic-news.info`).

À propos de cryptographie, notez que le site vers lequel on est redirigé n'a **pas** HTTPS. Cela veut dire que le djihadiste ou le pédophile prudent, qui n'utilise que HTTPS, ne verra jamais la jolie main rouge (juste un "*timeout*").

Quelles solutions sont disponibles si on veut quand même voir la propagande djihadiste ? La seule solution propre techniquement est d'avoir son propre résolveur DNS <http://www.bortzmeyer.org/son-propre-resolveur-dns.html>. En attendant, on peut utiliser un résolveur non-menteur (en supposant qu'il ne soit pas détourné <http://www.bortzmeyer.org/turquie-dns-frnog.html> et que le port 53 <http://www.bortzmeyer.org/port53-filtre.html> ne soit pas filtré). Dans tous les cas, il est sûr que la stabilité et la sécurité de l'Internet <http://www.bortzmeyer.org/resolution-de-demain.html> vont en souffrir. Sinon, on peut aussi s'auto-radicaliser un peu plus et franchement passer à Tor pour naviguer sur le Web.

Quelques bonnes lectures :

— Les explications du ministère de l'Intérieur <http://www.interieur.gouv.fr/Actualites/L-actu-du-Ministere/Lutte-contre-le-terrorisme-et-la-pedopornographie>.

2. Car trop difficile à faire afficher par LaTeX

— L'analyse politique de la Quadrature du Net <http://www.laquadrature.net/fr/la-france-persist sur cette censure administrative.
— L'article de Taziden <https://www.libre-parcours.net/post/censure-administrative-sites/ >
— Et celui de Pierre Beyssac <https://signal.eu.org/blog/2015/03/15/censure-sans-juge-dinte >
— L'excellent rapport du Conseil Scientifique de l'AFNIC <http://www.afnic.fr/fr/l-afnic-en-bref/ actualites/actualites-generales/6573/show/le-conseil-scientifique-de-l-afnic-part html> sur la censure via le DNS.
— Un précédent article de réflexion sur le filtrage DNS <http://www.bortzmeyer.org/dns-filtering. html>.
— Un amusant pastiche <http://metamorforme42.neocities.org/Site_masque.html> du site Web du ministère.
— Sur les aspects Communication, une réflexion d'Andréa Fradin <http://rue89.nouvelobs. com/2015/03/16/the-voice-main-rouge-linterieur-prend-les-pieds-com-258225> et une autre réflexion <http://affordance.typepad.com/mon_weblog/2015/03/pouce-bleu-inde html> sur les couleurs de la censure.
— Sur l'aspect politique de la censure, la bonne comparaison de Clochix <http://esquisses. clochix.net/2015/03/16/censure/>.
— Dans les médias : chez Libération <http://ecrans.liberation.fr/ecrans/2015/03/16/ cinq-sites-web-projihad-bloques-de-l-interieur_1222042>, dans Rue89 <http: //rue89.nouvelobs.com/2015/03/16/terrorisme-blocage-sites-internet-a-commence-258 chez Numérama <http://www.numerama.com/magazine/32492-un-site-d-information-islami html>, le Monde <http://www.lemonde.fr/pixels/article/2015/03/16/premier-cas-de-site 4594083_4408996.html> (et des détails ici <http://www.lemonde.fr/pixels/article/ 2015/03/17/les-premiers-blocages-administratifs-de-sites-djihadistes-en-7-questi 4594952_4408996.html>) et chez NextInpact <http://www.nextinpact.com/news/93457-islamic htm>.

Et merci à David Thomson pour avoir apparemment été le premier à repérer cette censure <https: //twitter.com/_DavidThomson/status/577076816868933632>, aujourd'hui, dimanche 15 mars.