Mesurer l'efficacité du pare-feu national chinois

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 10 novembre 2021

https://www.bortzmeyer.org/measuring-gfw.html

L'Internet en Chine est censuré par un dispositif souvent décrit sous le nom de GFW ("*Great Fire-Wall*"). C'est sans doute le dispositif de censure le plus perfectionné au monde. Il a donc été assez étudié mais rarement autant en détail que dans l'excellent article « "*How Great is the Great Firewall ? Measuring China's DNS Censorship*" <https://arxiv.org/abs/2106.02167> », dont les auteurs ont surveillé le GFW sur une longue période.

Le GFW n'a pas une technique unique de censure. Une de ses forces est de combiner plusieurs techniques. L'une d'elles est la génération, par le réseau lui-même (et pas par un résolveur <https://www.bortzmeyer.org/resolveur-dns.html> menteur), de fausses réponses DNS. Vous demandez un nom censuré et paf vous recevez une réponse donnant une adresse IP qui n'a rien à voir avec la question posée. Dans la plupart des pays qui censurent avec le DNS, la réponse mensongère est un NXDOMAIN - code indiquant que le nom n'existe pas - ou bien l'adresse d'un site Web qui affichera un message explicatif. Au contraire, les censeurs chinois sont soucieux de brouiller les pistes et renvoient une adresse réelle, ce qui rendra plus difficile de comprendre ce qui se passe.

Voici un exemple. J'interroge l'adresse IP 113.113.113.113, qui est en Chine, sur le réseau de China Telecom. Autant que j'en sache, aucune machine ne répond à cette adresse (testé avec nmap). Si je l'interroge sur un nom de domaine, je n'ai, logiquement, pas de réponse :

```
% dig @113.113.113.113 A mit.edu
...
;; connection timed out; no servers could be reached
```

Mais si je l'interroge sur un nom censuré, là, le réseau génère une réponse mensongère :

```
%  dig @113.113.113.113 A scratch.mit.edu

...
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 56267
;; flags: qr rd ra; QUERY: 1, ANSWER: 1, AUTHORITY: 0, ADDITIONAL: 1
...
;; ANSWER SECTION:
scratch.mit.edu. 248 IN A 157.240.6.18

;; Query time: 210 msec
;; SERVER: 113.113.113.113#53(113.113.113.113)
;; WHEN: Tue Nov 09 08:35:29 UTC 2021
;; MSG SIZE  rcvd: 60
```

L'adresse IP `157.240.6.18` appartient à Facebook (normalement, `scratch.mit.edu` est chez Fastly), exemple typique des mensonges générés par le GFW.

Pour étudier en détail ce mécanisme, les auteurs de l'article « *"How Great is the Great Firewall?"* <https://arxiv.org/abs/2106.02167> » ont développé le logiciel GFWatch <https://gfwatch.org/> qui permet de faire des études du GFW sur une longue période, entre autres sur les adresses IP renvoyées par le GFWatch.

GFWatch utilise des listes de noms de domaine dans des TLD importants comme `.com`, augmentées de noms dont on a appris qu'ils étaient censurés (`scratch.mit.edu` est censuré - cf. exemple plus haut - mais `mit.edu` ne l'est pas, donc utiliser les listes de noms sous les TLD comme `.edu` ne suffit pas). Il interroge ensuite des adresses IP en Chine, adresses qui ne répondent pas aux questions DNS (rappelez-vous que les réponses mensongères sont fabriquées par des "*middleboxes*" et qu'il n'y a donc même pas besoin que l'adresse IP en question réponde, comme dans le cas de `113.113.113.113`). Toute réponse est donc forcément une action de censure. Le logiciel GFWatch stocke alors les réponses. L'article utilise des données de 2020, collectées pendant neuf mois. (On peut voir les domaines censurés <https://gfwatch.org/censored_domains>.)

Les résultats ? Sur 534 millions de domaines testés, 311 000 ont déclenché une réponse mensongère du GFW. Les censeurs chinois n'ont pas peur des faux positifs et, par exemple, `mentorproject.org` est censuré, probablement uniquement parce qu'il contient la chaine de caractères `torproject.org`, censurée parce que les censeurs n'aiment pas Tor.

GFWatch peut ainsi obtenir une longue liste de domaines censurés, et essayer de les classer, ce qui permet d'obtenir une idée de la politique suivie par les censeurs (inutile de dire que les gérants du GFW ne publient pas de rapports d'activité détaillant ce qu'ils font…). On trouve par exemple des domaines liés à la pandémie de Covid-19 (les autorités chinoises ne veulent pas laisser l'information sur la maladie circuler librement).

Une des particularités du GFW est le renvoi d'adresses IP sans lien avec le nom demandé (comme, plus haut, une adresse de Facebook renvoyée à la place de celle du MIT). Quelles sont ces adresses IP ? Combien sont-elles ? Comment sont-elles choisies ? C'est l'un des gros intérêts d'un système comme GFWatch, de pouvoir répondre à ces questions. L'adresse retournée n'est clairement pas prise au hasard dans tout l'espace d'adressage IPv4. Seules 1 781 adresses IPv4 ont été vues par GFWatch, presque la moitié étant des adresses Facebook. Le GFW renvoie aussi des adresses IPv6 :

```
%  dig @113.113.113.113 AAAA scratch.mit.edu
...
;; ANSWER SECTION:
scratch.mit.edu. 84 IN AAAA 2001::4a75:b6b7
```

Et toutes appartiennent au préfixe réservé pour Teredo (RFC 4380[1]), une technologie désormais abandonnée.

Quant aux adresses IPv4, leur nombre varie dans le temps (de nouvelles adresses apparaissent de temps en temps), et le choix ne semble pas aléatoire, certaines adresses apparaissant davantage que les autres.

Du fait que les réponses mensongères sont générées par le réseau (plus exactement par une "*middlebox*"), et pas par un serveur, le GFW brouille parfois les réponses de serveurs légitimes. Plusieurs cas sont cités par l'article, mais je vais plutôt mentionner un cas très récent, le brouillage des réponses du serveur racine `k.root-servers.net` car le résolveur d'un FAI mexicain a eu le malheur d'interroger l'instance pékinoise de `k.root-servers.net` et le GFW a donc envoyé ses réponses mensongères. Le point a été discuté sur la liste dns-operations de l'OARC <`https://lists.dns-oarc.net/pipermail/dns-operations/2021-November/021437.html`> en novembre 2021 et il semble que l'annonce BGP de l'instance pékinoise ait été transmise bien au delà de sa portée voulue (un problème relativement fréquent avec les serveurs "*anycastés*").

L'article montre d'ailleurs que certains résolveurs DNS publics ont reçu des réponses générées par le GFW et les ont mémorisées. Les réponses de cette mémoire ainsi empoisonnée ont ensuite été servies à d'innocents utiisateurs. Bref, on ne répétera jamais assez qu'il faut utiliser DNSSEC (signer les zones, et vérifier les signatures ; les gros résolveurs publics vérifient tous les signatures mais cela ne marche que si la zone est signée).

Comment lutter contre cette censure ? Déjà, l'article note que le GFW est en général « sur le côté » et pas « sur le chemin ». Il injecte un mensonge mais ne bloque pas la vraie réponse. Parfois, celle-ci arrive même avant le mensonge, si le GFW a été lent à réagir. Une solution possible serait donc d'attendre un peu voir si on ne reçoit pas une autre réponse, plus vraie. Certains motifs dans la réponse mensongère (comme l'utilisation du préfixe `2001::/32`, normalement inutilisé, pour les requêtes AAAA) pourraient permettre d'ignorer les réponses de la censure. (Vous pouvez voir les adresses retournées par les menteurs sur le site de GFWatch <`https://gfwatch.org/forged_ips`>.) Mais, comme dit plus haut, la solution est évidemment DNSSEC, avec un lien sécurisé vers le résolveur validant (par exemple avec DoT ou DoH). Ne vous fiez pas à ce que raconte l'article sur des soi-disant « problèmes de compatibilité », qui ne sont pas détaillés. Mais attention, cela ne résout que la censure du DNS ; le GFW emploie une combinaison de techniques et y échapper n'est pas facile (et peut, si vous êtes en Chine, attirer l'attention de gens assez désagréables et en uniforme).

Vous pouvez aussi regarder la censure chinoise avec un résolveur situé en Chine et accessible via le fédivers, `ResolverCN@mastodon.xyz`, par exemple voici ce qu'il voyait pour Scratch <`https://mastodon.xyz/@ResolverCN/107246521999282095`>. Mais ce n'est même pas nécessaire, comme on l'a vu plus haut, donc vous pouvez aussi vous servir du DNS Looking Glass <`https://www.bortzmeyer.org/dns-lg-usage.html`>, par exemple en . Dans les réponses ci-dessous, `162.125.32.5` appartient à Dropbox, pas Facebook :

1. Pour voir le RFC de numéro NNN, `https://www.ietf.org/rfc/rfcNNN.txt`, par exemple `https://www.ietf.org/rfc/rfc4380.txt`