RFC 7535 : AS112 Redirection using DNAME

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 15 mai 2015

Date de publication du RFC : Mai 2015

https://www.bortzmeyer.org/7535.html

Les serveurs DNS, notamment ceux de la racine, sont en permanence bombardés par un trafic « inutile », des demandes de résolution DNS d'origine locale, et qui n'auraient jamais dû sortir. Par exemple, dans les réseaux qui utilisent les adresses IP privées du RFC 1918[1], les demandes de « résolution inverse » (« quel est le nom correspondant à 192.168.24.66 ? ») devraient toujours être résolues localement (après tout, il s'agit d'adresses privées) mais souvent les gestionnaires de ces réseaux ne configurent pas correctement leurs résolveurs et paf, la requête 66.24.168.192.in-addr.arpa arrive à la racine, qui ne peut pas en faire grand'chose. Idem si un réseau local utilise (bien à tort <https://www.bortzmeyer.org/pourquoi-le-tld-local-n-est-pas-une-bonne-idee.html>) un TLD local comme .home ou .corp. Les requêtes pour des TLD inexistants font plus de 25 % du trafic atteignant la racine. Pour les noms correspondants aux adresses du RFC 1918, le problème a été réglé par la création d'un « puits DNS », l'AS112 (RFC 7534). Les serveurs de l'AS112 sont des machines sacrifiées, à qui on envoie toutes les requêtes pour les noms en in-addr.arpa correspondant au RFC 1918. Cela marche très bien depuis de nombreuses années. Tellement bien qu'on se dit qu'on pourrait jeter dans ce puits bien d'autres choses comme les équivalents IPv6 du RFC 1918 (d.f.ip6.arpa, par exemple, cf. RFC 4193), ou comme les TLD tel que .home. Seulement, voilà, l'AS112 n'a pas été prévu pour cela et, avec leur configuration actuelle, ses serveurs ne répondraient pas correctement à ces requêtes. Il suffit de changer cette configuration ? Plus facile à dire qu'à faire, car l'AS112 est composé de nombreux serveurs plus ou moins coordonnés. Il est probablement impossible de les reconfigurer tous. D'où la nouvelle technique, utiliser (pour la première fois en grand) les enregistrements DNAME.

La nature même de l'AS112, système très décentralisé, sans direction claire, fait qu'il est impossible de changer la configuration de tous les nœuds. Même si l'IETF, l'ARCEP, l'ICANN, la NSA, le CSA et Jean-Kevin <https://twitter.com/jeank3vin> étaient tous d'accord, on ne pourrait pas garantir qu'il ne traine pas des nœuds AS112 ayant l'ancienne liste. Et puis on ne veut pas faire juste un changement une fois. On voudrait davantage de souplesse, permettant d'ajouter (ou de retirer) des noms

1. Pour voir le RFC de numéro NNN, https://www.ietf.org/rfc/rfcNNN.txt, par exemple https://www.ietf.org/rfc/rfc1918.txt

à l'AS112 facilement, sans passer à chaque fois des mois, voire des années, à rappeler aux gérants des serveurs AS112 qu'ils doivent changer leur configuration.

Comme toujours avec l'AS112, l'idéal serait que les administrateurs DNS fassent tous correctement leur travail et suivent le RFC 6303, qui leur impose de servir des zones comme `*.168.192.in-addr.arpa` localement. Mais, dans le monde réel, ça n'arrive pas et l'AS112 (nommé, j'ai oublié de le dire, d'après le numéro de système autonome qui lui a été attribué) reste indispensable.

On a vu plus haut que l'AS112 n'était que faiblement coordonné. Ses serveurs sont anycastés (RFC 4786) et composés de nombreuses machines, administrées par des personnes différentes. Si on décidait de déléguer, mettons, `.home`, aux serveurs de l'AS112, `blackhole-1.iana.org` et `blackhole-2.iana.org`, certaines des instances de ces deux serveurs seront correctement configurées et répondront, à juste titre, `NXDOMAIN` (ce domaine n'existe pas : c'est la réponse unique de l'AS112). Mais d'autres instances, administrées avec retard, ou gérées par des gens surmenés qui ont raté l'annonce faite sur la liste des gérants AS112, ne sauront pas qu'elles doivent répondre aux requêtes pour `*.home`. Elles seront en "*lame delegation*", avec des résultats plus ou moins bons (selon le logiciel utilisé) tels qu'une réponse `REFUSED` (qui poussera les résolveurs à essayer un autre serveur de la zone, alors que l'objectif était au contraire de diminuer le trafic).

L'IETF a donc choisi une voie différente pour l'évolution de l'AS112 : créer un nouveau préfixe, qui recevra les délégations des nouveaux `in-addr.arpa`, `ip6.arpa`, TLD, etc. Les nœuds utilisant ce nouveau préfixe utiliseront une nouvelle technique, à base de DNAME (RFC 6672), pour ne **pas** avoir à être configuré avec la liste des domaines à servir. Cela permettra de faire évoluer cette liste, ce qui est organisationnellement impossible avec les préfixes actuels de l'AS112.

Un enregistrement DNAME dit qu'une zone est un alias d'une autre. Ainsi :

```
example.com.     DNAME    example.net.
```

va faire que tous les noms sous `example.com` (mais pas `example.com` lui-même, attention), sont les alias des noms correspondants en `example.net`. Si on veut trouver l'adresse IPv6 associée à `foobar.example.c` le résolveur DNS, en trouvant le DNAME, fera une requête pour `foobar.example.net` et en transmettra le résultat à l'utilisateur.

La nouvelle zone de l'AS112 se nomme `empty.as112.arpa` et elle est déléguée à un seul serveur (anycasté), `blackhole.as112.arpa` (`192.31.196.1 / 2001:4:112::1`). Ce serveur n'a **pas** les mêmes adresses que les serveurs traditionnels de l'AS112, il y aura deux jeux de serveurs complètement différents, puisqu'ils auront une configuration distincte (une liste fixée de domaines pour les anciens serveurs, cf. RFC 6304, et une seule zone, la cible des DNAME, pour les nouveaux serveurs). Certains serveurs pourront faire partie des deux jeux à la fois (par exemple des anciens serveurs qui seraient modifiés pour gérer les nouvelles adresses en même temps que les anciennes). L'ancien AS112 continue à fonctionner comme avant. Les nouveaux préfixes anycast correspondants sont `192.31.196.0/24` et `2001:4:112::/48`, qui ne servent qu'à l'AS112. (Vous pouvez les chercher dans le "*looking glass*" de votre choix. L'AS d'origine sera le 112. L'AS112 est désormais à l'origine de quatre préfixes <`http://bgp.he.net/AS112`>.)

Ces adresses IP du serveur du nouvel AS112 ont été enregistrées dans le registre des adresses spéciales <`https://www.iana.org/assignments/iana-ipv4-special-registry/iana-ipv4-special-registr` `xml`> créé par le RFC 6890.

Donc, si vous voulez rediriger une zone DNS vers le puits du nouvel AS112, il suffit d'en faire un DNAME de **empty.as112.arpa**. Oui, aucun besoin de prévenir qui que ce soit, ou de demander une autorisation, vous n'avez qu'à ajouter cet enregistrement.

Un petit mot pour les fanas de gouvernance : le RFC 3172 confie la gestion politique de `.arpa` à l'IAB et c'est donc celle-ci qui a eu à approuver, sinon ce RFC, du moins le domaine dont il avait besoin, `as112.arpa`.

À l'heure actuelle, je n'ai pas l'impression qu'il y ait déjà des redirections vers `empty.as112.arpa` (en tout cas, DNSDB <https://www.bortzmeyer.org/dnsdb.html> n'en trouve pas). Il faut dire qu'il y a encore peu de serveurs dans le nouvel AS112. Une fois qu'il sera peuplé, et pourra donc gérer le trafic, on pourra mettre des redirections. Rappelez-vous que l'installation d'une redirection ne nécessite aucun changement dans les serveurs du nouvel AS112. Pour tester, vous pouvez utiliser ma zone `sink.bortzmeyer.fr`. Normalement, `nimportequoi.sink.bortzmeyer.fr` doit vous obtenir un `NXDOMAIN` de la part du serveur AS112 « le plus proche ». Voici, vu par les sondes Atlas <https://atlas.ripe.net/> les différentes réponses possibles pour le nouvel AS112, ce qui semble indiquer qu'il n'a que quatre serveurs :

```
%  python resolve-name.py -t SOA empty.as112.arpa
Measurement #2004698 for empty.as112.arpa/SOA uses 497 probes
[] : 11 occurrences
[as112.ottix.net. dns.ottix.net. 1 604800 60 604800 604800] : 2 occurrences
[blackhole.as112.arpa. dns.ripe.net. 1 604800 60 604800 604800] : 226 occurrences
[as112.hivane.net. info.hivane.net. 1 604800 60 604800 604800] : 184 occurrences
[blackhole.as112.arpa. noc.dns.icann.org. 1 604800 60 604800 604800] : 60 occurrences
Test done at 2015-05-14T10:59:36Z
```

Et l'ancien AS112 ? Il continue à fonctionner comme avant. Si le nouvel AS112 marche bien, s'il est largement peuplé, on pourra changer les actuelles délégations des zones RFC 1918 comme `10.in-addr.arpa` et les remplacer par une redirection DNAME vers `empty.as112.arpa`, le nouvel AS112. Une fois que cela sera fait, l'ancien AS112 et ses serveurs `blackhole-1.iana.org` et `blackhole-2.iana.org` pourront être supprimés. Il n'y a toutefois pas de plan précis pour cela : il faut d'abord que le nouvel AS112 fasse ses preuves.

Quelles zones seront mises dans le nouvel AS112 ? Comme indiqué plus haut, n'importe quelle zone peut être déléguée au nouvel AS112 de manière unilatérale. Mais ce service est surtout utile pour les zones à fort trafic, comme celles listées dans le RFC 6303 (qui ne sont pas toutes dans l'ancien AS112, par exemple aucune zone en `ip6.arpa` n'y est) ou comme certains TLD (les plus importants à l'heure actuelle sont, dans l'ordre, `.local`, `.home`, `.html`, `.localdomain` et `.internal`). Dans le cas des TLD, toutefois, il est probable que l'incroyable bureaucratie ICANN et la mainmise du gouvernement états-unien sur le contenu de la racine ne retardent considérablement le projet.

Le choix d'utiliser DNAME n'a pas été sans mal. En effet, DNAME ne fait pas partie du DNS original, et ce n'est pas un simple type d'enregistrement pour des données passives, il nécessite un traitement spécifique dans les résolveurs (et, dans une certaine mesure, dans les serveurs faisant autorité). On ne peut donc pas compter sur 100 % de serveurs DNS gérant DNAME. Est-ce un problème ? Le RFC 6672 prévoit un mécanisme de secours, où l'enregistrement DNAME est accompagné d'un CNAME (les alias classiques) synthétisé à la demande par le serveur. En théorie, tout doit donc bien se passer, même dans un monde imparfait où tous les serveurs ne gèrent pas DNAME. Et en pratique ? L'annexe A décrit les expérimentations qui avaient été faites avant de choisir la solution DNAME (quelques autres avaient été proposées). Le test avait consisté en quatre images sur une page Web que devait charger les navigateurs. La page était distribuée sous forme d'une publicité payante, envoyée par la régie publicitaire à de nombreux navigateurs. Le premier URL, `http://a.UNIQUE_STRING.dname.example.com/1x1.png?a.UNIQUE_STRING.dname` utilise un domaine qui a un

DNAME vers un autre domaine. Si tout l'Internet peut travailler avec des DNAME sans problème, l'image correspondante sera chargée dans 100 % des cas. (La chaîne de caractères UNIQUE_STRING est la même pour toutes les images mais varie à chaque chargement de la page. C'est donc l'identificateur d'un test donné.) Le second URL, `http://b.dname.example.com/1x1.png?b.UNIQUE_STRING.dname`, qui, contrairement au premier, utilise un nom de domaine qui a un DNAME, et qui n'a pas de partie unique et est donc fréquemment mis en cache. Le troisième URL, `http://c.UNIQUE_STRING.target.example.net/1x1.png?c.UNIQUE_STRING.target`, ressemble beaucoup au premier mais n'utilise pas du tout de DNAME. C'est du DNS on ne peut plus traditionnel, et il sert donc de contrôle : si le score est de moins de 100 % sur cette troisième image, c'est que le navigateur, la machine sur laquelle il tourne, ou bien le réseau qui l'héberge, ont d'autres problèmes que les DNAME. Un script en Flash dans la page mesure le résultat et le temps nécessaire pour l'obtenir. Enfin, une quatrième image, `http://results.recorder.example.net/1x1.png? results.UNIQUE_STRING?za=FIRST_RESULT&zb=SECOND_RESULT&zc=THIRD_RESULT`, sert juste à transmettre les résultats. Le journal du serveur HTTP ressemblera à :

```
GET /1x1.png?results.UNIQUE_STRING?za=1822&zb=1674&zc=1582
```

(Les chiffres sont les temps de chargement en millisecondes.)

Les résultats ? 98,1 % des tests ont permis le chargement de la première ou de la seconde image, celles qui utilisaient les DNAME. Il y a donc 1,9 % de clients qui ne peuvent pas utiliser les DNAME ? Non, car 2,8 % des tests de chargement de la troisième image (qui n'utilise pas de DNAME, que du DNS d'avant-guerre) ont également échoué. Ce qui montre que ce chiffre de 1,9 % est inférieur au pourcentage d'erreur (soubresauts du réseau, etc). Bref, les DNAME marchent très bien.

À noter que les serveurs du nouvel AS112, eux, n'ont aucun besoin de savoir gérer les DNAME. Seuls les résolveurs, et les serveurs faisant autorité pour les zones redirigées doivent le faire.

Le `as112.arpa` a été délégué début février 2015 :

```
% check-soa -i as112.arpa
a.iana-servers.net.
199.43.132.53: OK: 2014122469 (2 ms)
2001:500:8c::53: OK: 2014122469 (14 ms)
b.iana-servers.net.
199.43.133.53: OK: 2014122469 (149 ms)
2001:500:8d::53: OK: 2014122469 (156 ms)
c.iana-servers.net.
2001:500:8e::53: OK: 2014122469 (13 ms)
199.43.134.53: OK: 2014122469 (15 ms)
```

Et son fils :

```
% check-soa -i empty.as112.arpa
blackhole.as112.arpa.
        2001:4:112::1: OK: 1 (31 ms)
        192.31.196.1: OK: 1 (23 ms)
```

Voici une réponse actuelle pour le nouveau domaine :

```
% dig ANY empty.as112.arpa

; <<>> DiG 9.8.4-rpz2+rl005.12-P1 <<>> ANY empty.as112.arpa
;; global options: +cmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 30414
;; flags: qr rd ra; QUERY: 1, ANSWER: 4, AUTHORITY: 1, ADDITIONAL: 5

;; OPT PSEUDOSECTION:
; EDNS: version: 0, flags: do; udp: 4096
;; QUESTION SECTION:
;empty.as112.arpa. IN ANY

;; ANSWER SECTION:
empty.as112.arpa. 604800 IN SOA blackhole.as112.arpa. dns.ripe.net. (
1          ; serial
604800     ; refresh (1 week)
60         ; retry (1 minute)
604800     ; expire (1 week)
604800     ; minimum (1 week)
)
empty.as112.arpa. 3599 IN RRSIG NSEC 8 3 3600 20150515204633 (
20150508164936 52494 as112.arpa.
lZf9d6oYISRAq8i6gLXXCwUoQP+qmOfAEIiY3hNr6PvD
DE1kSeGGR3lo3fK6P/CajYv/8twZm/CNQvpAxlRLUyrY
Hc2WHhgxCtiQx01pLrY9d/QRhiL1WqYZaMSrfNZX0H0S
GUD1BOcVNzT/lYPz0EdZZKzEXle98ZGWwkIvewE= )
empty.as112.arpa. 3599 IN NSEC hostname.as112.arpa. NS RRSIG NSEC
empty.as112.arpa. 3599 IN NS blackhole.as112.arpa.

;; AUTHORITY SECTION:
empty.as112.arpa. 3599 IN NS blackhole.as112.arpa.

;; ADDITIONAL SECTION:
blackhole.as112.arpa. 3599 IN A 192.31.196.1
blackhole.as112.arpa. 3599 IN AAAA 2001:4:112::1
blackhole.as112.arpa. 3599 IN RRSIG A 8 3 3600 20150516053029 (
20150508164936 52494 as112.arpa.
B4eU9u5ZQVGf+Haro2CeCanWwFLeK3hvil8dIlpz1fMm
xR8K1No4rWTV5hWME1GhFatZVgpATfat9A3rghGWB9Xm
hcmsaE5uHTOB+56DNhiokWsVtj+WT828naDMlfvGWiP4
cXIxF/tLcR10XYviczlKYYR/SgAVxgmwjFkBHXg= )
blackhole.as112.arpa. 3599 IN RRSIG AAAA 8 3 3600 20150515212644 (
20150508164936 52494 as112.arpa.
Yk2l+kWkYbYruCNHIKZwGg8GZPDp9y5Qezqk+Ogq5rGF
/3+R/UjPPw240zdnLi4D2DeBFwlvM8rDq0xt3sreEmdk
jMdxGcAc8eEfM6O1heP7lgRJW4eCzwOdNX6f1IXEIerg
XwYWu+3VjI6y4NsrYoczfo+ORAHsdUvz9rdYumk= )

;; Query time: 854 msec
;; SERVER: 217.70.184.225#53(217.70.184.225)
;; WHEN: Fri May  8 20:43:04 2015
;; MSG SIZE  rcvd: 726
```