RFC 7477 : Child To Parent Synchronization in DNS

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 14 mars 2015

Date de publication du RFC : Mars 2015

https://www.bortzmeyer.org/7477.html

Le DNS est un système décentralisé, mais arborescent : toute zone (à part la racine) dépend d'une zone **parente**, qui doit notamment indiquer les serveurs de noms de ses zones **filles**. En théorie, l'information dans la zone parente et celle dans la zone fille doivent parfaitement coïncider mais, en pratique, des différences apparaissent souvent, et elles ont parfois des conséquences ennuyeuses, pouvant aller jusqu'à mettre en danger le bon fonctionnement de la zone. La principale raison de cette « désynchronisation » est qu'il n'existait aucun mécanisme standard par lequel le gestionnaire d'une zone fille pouvait transmettre à la zone parente les données à distribuer. C'est ce manque que comble notre tout nouveau RFC, avec l'introduction des enregistrements DNS de type `CSYNC` ("*Child SYNChronization*").

Les deux types principaux d'enregistrement à copier de la zone fille vers la zone parente sont la liste des serveurs de noms de la zone (enregistrements NS) et les adresses IP de ces serveurs, si et seulement si leur nom est lui-même dans la zone servie, et qu'on ne peut donc pas compter sur la résolution DNS normale pour trouver les adresses. Ces adresses servies par la zone parente sont nommées **colles** ("*glue*") et sont des enregistrements de type A et de type AAAA.

La résolution DNS dépend de cette information. En effet, le résolveur DNS part de la racine, et descend, de zone parente en zone fille, jusqu'à trouver l'information qu'il veut. S'il cherche `www.hacklab-esgi.fr`, il passera par les serveurs de la racine (qui connaissent la liste des serveurs de noms de `.fr`) puis par ceux de `.fr` (qui connaissent la liste des serveurs de `hacklab-esgi.fr`). Voici, reproduite avec dig, la requête qui sera envoyée à un serveur de `.fr` :

```
% dig +norecurse @d.nic.fr A www.hacklab-esgi.fr
...
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 30993
;; flags: qr; QUERY: 1, ANSWER: 0, AUTHORITY: 2, ADDITIONAL: 0

;; QUESTION SECTION:
;www.hacklab-esgi.fr. IN A

;; AUTHORITY SECTION:
hacklab-esgi.fr. 172800 IN NS ns0.online.net.
hacklab-esgi.fr. 172800 IN NS ns1.online.net.
```

On y voit qu'une réponse correcte nécessite que les serveurs de l'AFNIC (ici `d.nic.fr`) connaissent les serveurs de `hacklab-esgi.fr`. Mais la liste de ces serveurs est également dans la zone :

```
% dig +norecurse @ns0.online.net. NS hacklab-esgi.fr
...
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 59586
;; flags: qr aa; QUERY: 1, ANSWER: 2, AUTHORITY: 0, ADDITIONAL: 0

;; QUESTION SECTION:
;hacklab-esgi.fr. IN NS

;; ANSWER SECTION:
hacklab-esgi.fr. 14400 IN NS ns0.online.net.
hacklab-esgi.fr. 14400 IN NS ns1.online.net.
```

Ce sont ces deux listes qui devraient idéalement rester synchrones. (On note au passage que la liste dans les serveurs de `hacklab-esgi.fr` **fait autorité** : on a le bit `aa` - *"Authoritative Answer"* - dans la réponse, alors que la copie dans les serveurs de `.fr` ne fait pas autorité.)

Le processus typique actuel d'**avitaillement**, c'est-à-dire d'enregistrement des informations nécessaires à la **délégation**, se fait de diverses manières. Rappelez-vous qu'il n'y avait pas de standard pour cela. Prenons deux cas dans le TLD imaginaire `.example`, Jeanne Michu qui gère sa zone `jeanne.example` sur ses propres serveurs de noms, et utilise la société Foo comme BE, et Paul Michu, moins *"geek"*, qui utilise la société Bar à la fois comme BE et comme hébergeur DNS, pour sa zone `paul.example`. Jeanne utilise emacs pour modifier le fichier de zone et y mettre :

```
jeanne.example. IN NS ns1.jeanne.example.
         IN NS ns2.hosted-by-a-friend.example.
ns1             IN      AAAA    2001:db8:f54::2:53
```

On note que Jeanne n'indique pas l'adresse IP de `ns2.hosted-by-a-friend.example`, qui est dans une autre zone. Elle recharge ensuite son serveur de noms pour prendre en compte les informations (`nsd-control reconfig` si elle utilise nsd4 <https://www.bortzmeyer.org/nsd4.html>) et elle lance un programme qu'elle a écrit qui, utilisant l'API que le BE Foo met à la disposition de ses utilisateurs, envoie les données au BE qui les transmettra au registre, typiquement en EPP. Paul, lui, ne fait rien, il a enregistré le domaine auprès de la société Bar et celle-ci se charge de mettre l'information dans ses propres serveurs de noms, et de prévenir le registre (puisque cette société est BE). Le résultat sera le même pour les deux zones, une fois les serveurs de la zone parente mis à jour, tout sera synchrone. (Rappelez-vous que ce ne sont que deux exemples, il peut y avoir **beaucoup** d'autres configurations possibles.)

Là où il peut y avoir des problèmes, c'est si la communication se passe mal. Par exemple, si Jeanne se plante dans une manipulation compliquée, ou oublie. Elle modifie l'ensemble NS dans sa zone mais ne pense pas ensuite de signaler le changement au parent. Ou bien elle lance le programme qui va modifier les données chez le parent mais les informations sont rejetées par la zone parente, sauf que ce refus n'est pas transmis à la titulaire du nom de domaine. Il existe des tas de raisons qui peuvent entrainer une désynchronisation de la zone fille et de la parente. C'est ce genre de problème que veut traiter notre nouveau RFC.

La solution proposée est de créer un nouveau type d'enregistrement DNS, nommé `CSYNC`, que le gérant technique de la zone DNS fille mettra dans la zone, et qui indiquera quelles données de la zone

doivent être recopiées dans la zone parente. Qui fera cette copie ? Ce rôle reviendra au "*parental agent*" (terme récent et pas encore traduit), qui peut être un BE (charge à lui de transmettre au registre, sans doute en EPP) ou directement le registre.

À première vue, on pourrait penser que cette technique conviendrait très bien également pour DNSSEC (et cela avait été sérieusement envisagé dans le groupe de travail), afin de permettre la synchronisation des enregistrements DS de la zone parente avec les clés (enregistrements DNSKEY) de la zone fille. Mais le cas de DNSSEC est différent (notamment parce que la zone parente fait autorité pour les `DS`, contrairement aux `NS` et colles). Un mécanisme différent est donc utilisé, normalisé dans le RFC 7344 [1].

Un autre point que ne traite **pas** notre nouveau RFC est celui de la configuration initiale. La zone fille doit évidemment être sécurisée avec DNSSEC pour que la parente puisse être sûre d'avoir récupéré le bon `CSYNC`. La configuration initiale de DNSSEC, ainsi que celle du "*parental agent*" pour le commencement des opérations, ne font pas l'objet d'une standardisation. Les `CSYNC` permettent de maintenir la synchronisation, pas de la créer au début.

La section 2 de notre RFC décrit l'enregistrement `CSYNC`. Il est placé à l'**apex** (au sommet) de la zone fille. Il doit être unique. Son numéro, enregistré à l'IANA `<https://www.iana.org/assignments/dns-parameters/dns-parameters.xhtml#dns-parameters-4>`, est le 62. Ses données comportent trois champs :

— Un numéro de série (identique à celui qu'on trouve dans le SOA de la zone),
— Des options (décrites plus loin),
— Une liste de types d'enregistrement à recopier dans la zone parente (en général, `NS`, `A` et `AAAA`). Elle est encodée selon le mécanisme un peu compliqué du RFC 4034, section 4.1.2.

Cette dernière liste est à respecter totalement ou pas du tout. Si elle compte un type que le "*parental agent*" ne connait pas, il ne doit faire aucun changement. D'une manière générale, les opérations liées à un CSYNC sont atomiques.

Voici un exemple d'enregistrement CSYNC, avec le format texte standard, pour la zone de Jeanne Michu :

```
jeanne.example. 3600 IN CSYNC 2015031401 1 NS AAAA
```

Il se lit ainsi :

— Le numéro de série est 2015031401,
— Les options se limitent à celle du premier bit, `immediate`,
— Il faut recopier les enregistrements `NS` et `AAAA` (il n'y a pas de colle de type `A` dans la zone). Si vous voyez un `TYPEnnn` dans un enregistrement `CSYNC`, c'est que le type était inconnu de votre client DNS (RFC 3597, section 5).

Le numéro de série doit être une copie de celui du `SOA` de la zone fille, ou bien avoir une valeur supérieure (si on ne change pas les informations de délégation, on n'est pas obligé de mettre à jour le numéro de série dans le `CSYNC`). Voir le RFC 1982 pour une définition rigoureuse de « avoir une valeur supérieure ».

L'utilisation de ce numéro de série dépend des options (le deuxième champ). Ces options consistent en seize bits dont seulement deux sont définis aujourd'hui (tout cela est stocké dans un registre IANA

1. Pour voir le RFC de numéro NNN, `https://www.ietf.org/rfc/rfcNNN.txt`, par exemple `https://www.ietf.org/rfc/rfc7344.txt`

<https://www.iana.org/assignments/dns-parameters/dns-parameters.xhtml#csync-flags>, tout ajout nécessitant la rédaction d'une norme). Les deux options définies sont `immediate` et `soaminimum`.

La section 3 du RFC décrit les détails du traitement d'un enregistrement `CSYNC` par le "*parental agent*". Cet agent ne doit agir automatiquement **que** si l'option `immediate` est mise (c'est le cas dans l'enregistrement d'exemple ci-dessus). Autrement, il doit attendre une validation, par exemple via une action dans une interface Web (« cette modification de vos serveurs de noms attend votre confirmation »). Quant à l'autre option normalisée, `soaminimum`, elle indique au "*parental agent*" qu'il ne doit copier les données indiquées que si le numéro de série de la zone (indiqué dans l'enregistrement `SOA`) est **supérieur** au numéro de série de l'enregistrement `CSYNC`. (Dans l'exemple plus haut, l'option `soaminimum` n'était pas activé et le numéro de série dans le `CSYNC` est donc ignoré.)

J'ai dit plus haut que la copie des données de la zone fille vers la parente devait être atomique (tout copier, ou ne rien changer). Cela ne concerne pas les TTL que la zone parente peut modifier lors de la copie.

D'autre part, la zone parente n'est pas **obligée** de copier les données quelles qu'elles soient. En effet, la zone parente a sa propre politique (des trucs du genre « deux serveurs de noms minimum » ou bien « je ne tiens compte de la colle que si elle concerne des serveurs qui sont dans la zone qu'ils servent », cf. section 4.3) et peut donc décider de ne pas utiliser les données fournies, rejetant ainsi la mise à jour demandée.

Voyons maintenant (en section 4) quelques problèmes concrets liés à l'utilisation de cette technique. Premier point à garder en tête, il n'y a aucun mécanisme pour remonter une erreur ou un refus (comme celui cité dans le paragraphe précédent) depuis la zone parente vers le gestionnaire de la zone fille. L'information doit donc passer par une autre voie (un message envoyé automatiquement, par exemple). Le gestionnaire de la zone fille ne doit donc pas tenir pour acquis que les données de la zone parente seront modifiées : il doit vérifier.

Le mécanisme des `CSYNC` nécessite une requête DNS depuis le "*parental agent*". Si celui-ci a peu de zones à surveiller, il peut simplement faire de la scrutation répétée. Mais cela ne passe pas forcément à l'échelle. Les "*parental agents*" avec beaucoup de zones à surveiller préféreront peut-être attendre un signal explicite de la part du gérant de la zone (un bouton sur une interface Web du BE, par exemple et/ou une fonction dans l'API).

Pour faciliter la vie des gérants de zones DNS, il faudrait de toute façon que le "*parental agent*" documente ses capacités : s'il traite les `CSYNC`, quels types d'enregistrements accepte-t-il (à part les évidents `NS`, `A` et `AAAA`), le mécanisme de scrutation (à quel intervalle ?), la notification des erreurs...

Une fois que la zone parente a été modifiée, la zone fille peut-elle retirer le `CSYNC` ? Oui, le changement n'est fait qu'une fois, et la zone parente ne va pas l'annuler même si le `CSYNC` disparait après.

Le "*parental agent*" ne doit agir que si les données ont été validées par DNSSEC (section 5, sur la sécurité du mécanisme). Si votre zone n'est pas encore signée (n'avez-vous pas honte de ce retard ?), vous ne pourrez pas bénéficier des nouveaux services comme CSYNC.

Il n'y a pas encore de mise en œuvre de cette technique (mais Wireshark sait décoder ces paquets <https://code.wireshark.org/review/#/c/7701/>). Je serais curieux de voir une liste des "*parental agents*" potentiels qui la déploient mais, début 2015, c'est encore trop tôt.