RFC 6471 : Overview of Email DNS-Based List (DNSBL) Best Practice

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 19 janvier 2012

Date de publication du RFC : Janvier 2011

https://www.bortzmeyer.org/6471.html

En raison de l'importance du problème du spam (et d'autres comportements tout aussi nuisibles), un grand nombre de sites utilisent des listes noires des gens avec qui on ne veut pas communiquer (DNSBL pour "*DNS-based Black List*"). Ces listes sont souvent distribuées via le DNS, et gérées par des organismes très divers, dont le niveau de sérieux et d'honnêteté est très variable. La question étant très polémique, documenter le comportement attendu de ces organismes n'a pas été une mince affaire. Ce RFC 6471 [1] décrit donc ce qu'on espère d'un gérant de DNSBL.

Rien ne dit qu'il sera suivi, bien sûr. Mais l'espoir est que ce document serve à faire évoluer les choses dans le bon sens. À noter que le groupe en charge de ce RFC, l'"*Anti-Spam Research Group*" <http://irtf.org/asrg> de l'IRTF, n'a pas toujours réussi à se mettre d'accord sur le comportement idéal de la DNSBL. Dans ces cas, ce RFC 6471 se contente d'indiquer des recommandations sur la forme (« le gérant de DNSBL devrait documenter ce qu'il fait » si le RFC ne peut pas dire ce qui devrait être fait). La section 1.4 rappelle d'ailleurs que les règles de l'IRTF n'imposent pas de consensus au sein du groupe avant publication. À noter que ce RFC ne parle pas de protocole, les questions purement techniques étant traitées dans le RFC 5782.

Le sujet est tellement polémique que je ne vais pas forcément me contenter d'un compte-rendu neutre, objectif et ennuyeux de ce RFC et que je risque de glisser quelques opinions personnelles, pour lesquelles je demande l'indulgence de mes lecteurs. C'est que les DNSBL sont un marécage de gens bizarres, de racketteurs (« pour être retiré de notre liste, il faut payer »), d'éradicateurs (« nous avons mis les trois-quarts de l'Afrique en liste noire ») et de quelques personnes sérieuses. Le groupe ASRG représente plutôt le point de vue des éradicateurs (« dans le doute, on filtre »).

1. Pour voir le RFC de numéro NNN, https://www.ietf.org/rfc/rfcNNN.txt, par exemple https://www.ietf.org/rfc/rfc6471.txt

La section 1 du RFC rappelle ce qu'est une DNSBL et comment elle fonctionne. Les détails techniques du fonctionnement de ce service sont exposés dans le RFC 5782. L'idée est de publier dans le DNS les noms de domaine des méchants, ou bien les adresses IP de leurs serveurs de messagerie. Le serveur qui reçoit du courrier peut alors, en consultant le DNS (ce qui est simple et rapide), prendre connaissance de la « note » attribuée au domaine ou à l'adresse IP par le gérant de la DNSBL. Sur la base de cette note, il peut alors décider de filtrer, de contrôler plus étroitement, etc. Rappelons-bien que c'est le destinataire qui décide quoi faire du courrier, pas le gérant de la DNSBL, qui se contente de publier une information. Au bout du compte, c'est le gérant du serveur de messagerie de destination qui est responsable de l'usage qu'il fait de cette information. Beaucoup sont irresponsables, et filtrent aveuglément sur la base de listes noires sur lesquelles ils n'ont même pas pris la peine de se renseigner.

Comme le DNS est une technique éprouvée, très fiable et rapide, cet usage original des DNSBL s'est étendu. On voit aujourd'hui des DNSWL (listes blanches, désignant ceux qui ont droit à un traitement de faveur), listes d'URI dont la présence dans un courrier indique un spam, etc. On voit aussi des changements dans l'usage de ces listes. Très binaire autrefois (l'adresse IP est présente dans la liste noire =¿ on rejette le message), il est devenu plus souple, les résultats d'une recherche dans la liste noire étant souvent un facteur de décision parmi d'autres (SpamAssassin fonctionne ainsi). Poursuivant cette idée d'utiliser le DNS pour récupérer de l'information, on voit aussi des services de géolocalisation utilisant le DNS et d'autres encore plus techniques (je me sers beaucoup du domaine `aspath.routeviews.org` de Route Views <`http://routeviews.org`>, qui permet de récupérer les informations de routage d'une adresse). Enfin, les listes distribuées par le DNS sont désormais utilisées pour bien d'autres choses que le courrier, par exemple pour du contrôle d'accès à des serveurs IRC ou des formulaires Web. Bref, on peut tout faire avec le DNS, qui a cessé il y bien longtemps d'être uniquement un service qui « traduit des noms de domaine en adresses IP ».

Quelles sont les DNSBL aujourd'hui ? Il y a de tout, certaines sont privées, gérées par une organisation pour son besoin propre, d'autres publiques et gratuites, d'autres payantes (beaucoup de DNSBL sont gratuites en dessous d'un certain seuil d'utilisation et payantes ensuite). On estime qu'il existe plus de 700 listes publiques, la plus ancienne ayant été créée en 1997. À cette époque, les spammeurs utilisaient leurs propres machines et c'est en grande partie en raison des DNSBL, qui distribuaient rapidement les adresses de ces machines, que les spammeurs sont passés à d'autres tactiques (comme les relais de courrier ouverts), tactiques à lesquelles les DNSBL ont dû s'adapter.

Il n'y a pas que le statut juridique et administratif qui différencient les DNSBL actuelles. Elles se différencient également par leurs politiques (comment une entrée est-elle ajoutée à la base et, plus important, comment elle est retirée) et, comme le note pudiquement le RFC, les DNSBL se différencient aussi par le niveau d'honnêteté de leurs responsables.

Justement, cette politique (qui va être mis dans la liste noire, sur quels critères ?) est un point de controverse permanent. Des cow-boys qui vous "*blacklistent*" un /20 entier parce qu'**une** adresse a envoyé un spam, aux gens sérieux qui prennent beaucoup de temps et d'effort avant d'ajouter une adresse à la base, il y a un large spectre d'opinions. Ce RFC 6471 n'essaie pas de trancher entre ces opinions. Il ne dit pas quelle est la bonne politique, simplement que le gérant de la DNSBL doit documenter ses critères d'inclusion et d'exclusion, et faire ensuite réellement ce qu'il annonce (ce qui est très loin d'être le cas). C'est ensuite à l'utilisateur de la DNSBL (l'administrateur d'un serveur de messagerie qui décide d'interroger cette base avant d'accepter un message) de s'informer et de s'assurer que la politique de la DNSBL lui convient.

La section 1.2 fournit une bonne "*check-list*" pour ledit utilisateur, sous forme d'une série de questions à se poser avant d'utiliser une DNSBL. Par exemple (entre parenthèses, une étude de l'application de cette "*check-list*" à Spamhaus, pour sa liste SBL) :

`https://www.bortzmeyer.org/6471.html`

— Est-ce que les politiques d'inclusion et d'exclusion suivies par la liste sont nettement marquées sur son site Web ? Sont-elles claires ? (Pour la SBL, la réponse est oui, pour l'inclusion et l'exclusion <http://www.spamhaus.org/sbl/policy.html>.)
— Existe-t-il une évaluation indépendante de cette liste, permettant de comparer la théorie et la pratique ? (Je n'en sais rien, pour la SBL.)
— Qu'en pensent vos pairs, les collègues qui travaillent dans des conditions analogues aux vôtres ? Comme le note justement le RFC, les DNSBL baignent souvent dans un climat de rude controverse. Il faut bien vérifier que les opinions sur une liste sont formulées par des gens qui s'y connaissent, et pas par Jean-Kevin Boulet qui crie bien fort « Ouais, cette liste est super, marche trop bien ». (La SBL est une des listes les plus utilisées et Spamhaus une organisation ancienne et appréciée.)

Il faut en outre réviser ces questions de temps en temps, les listes évoluent et plus d'une a cessé tout fonctionnement.

Le RFC enfonce le clou à plusieurs reprises : **c'est l'utilisateur qui est responsable, au final, pas la DNSBL**. Certes, c'est un plaidoyer pour les gérants de DNSBL (« ce n'est pas nous qui filtrons », remarque courante des DNSBL face aux contestations) mais cela reflète la réalité. Filtrer avec une DNSBL, c'est **sous-traiter** une partie de sa sécurité, c'est confier les décisions à d'autres (la section 4 revient sur ce point et rappelle que des DNSBL ont déjà listé `0/0` c'est-à-dire tout l'Internet). Il est donc crucial d'évaluer les sous-traitants. Le "*postmaster*" responsable comprend ce point : la DNSBL exprime une opinion, mais c'est lui qui décide d'agir sur la base de cette opinion.

Bref, quels sont les conseils effectifs de ce RFC 6471 ? Ils commencent en section 2. D'abord, la transparence de l'offre. La DNSBL doit écrire noir sur blanc quels sont les critères pour être mis dans la liste noire et quels sont ceux pour être enlevés. Par exemple, une liste qui s'appuie sur un pot de miel peut avoir comme critère d'ajout « On ajoute toute adresse IP qui a envoyé au moins trois messages dans une semaine aux adresses du pot de miel » et comme critère de retrait « On retire toute adresse qui n'a rien écrit au pot de miel depuis deux mois ». Cette politique doit ensuite être suivie rigoureusement et honnêtement. Dans la jungle des DNSBL, on a déjà vu des gérants de liste qui ajoutaient à la liste noire les adresses IP des gens qui les critiquaient ("*spite listing*")... Il y a des tas de politiques raisonnables, l'important est de se tenir à celle publiée. Si on prétend gérer une liste de relais de courrier ouverts, on ne doit pas y mettre des adresses IP pour une autre raison, même en rapport avec le spam.

La transparence n'empêche pas des politiques dingues, mais ouvertement assumées par le gérant de la liste. Un bel exemple est chez UCEprotect, qui menace directement les critiques « "*Should you want to contact us, you should keep this in mind and behave rationally and calmly in order not to aggravate your situation. [...] Applying legal action or other pressure against us will result in your IP address and/or your network range being listed in our database.*" <http://www.uceprotect.net/en/index.php?m=8&s=0> ».

À noter que cette règle de transparence n'impose pas de donner tous les détails. Par exemple, l'opérateur de la liste ne va évidemment pas publier les adresses du pot de miel, cela détruirait complètement son efficacité.

Ensuite, la traçabilité. La DNSBL devrait maintenir un historique des ajouts et retraits, avec les raisons, et publier cet historique. Là encore, l'historique publié peut être expurgé mais il doit contenir suffisamment d'informations pour qu'un utilisateur (ou l'administrateur d'une machine listée dans la liste noire) puisse comprendre pourquoi. Si on prend (un peu au hasard), la liste CBL <http://cbl.abuseat.org/>, on trouve juste : « IP Address X.Y.Z.170 is listed in the CBL. It appears to be infected with a spam sending trojan or proxy. It was last detected at 2011-12-17 12 :00 GMT (+/- 30 minutes), approximately 4 hours ago. », ce qui est un peu court.

Beaucoup de gérants de DNSBL sont du genre « éradicateur » et n'hésitent pas devant les dommages collatéraux. Le RFC dit d'ailleurs franchement qu'on ne fait pas d'omelette sans casser d'œufs.

Ainsi, il arrive que des listes répondent pour des adresses qui n'ont pas eu de comportement malveillant mais font partie du même préfixe (mettre dans la liste tout un /28 lorsqu'une seule adresse IPv4 de ce /28 a mal agi, par exemple). Le RFC recommande que cette pratique d'élargissement soit clairement documentée.

L'entrée dans la liste est une chose. Mais, avec la plupart des DNSBL, les problèmes sont encore pires pour sortir de la liste. Même quand l'entrée est correctement faite, et à juste titre, on peut avoir des difficultés incroyables pour se faire rayer de la liste. Par exemple, une machine Windows devient un zombie, crache du spam à tout va, et se retrouve « noirlistée », ce qui est normal. Ensuite, l'administrateur système prend les choses en main, reformate le disque, installe NetBSD à la place, et va tenter de « délister » la machine, désormais propre. Avec la plupart des DNSBL, il aura un mal fou. Être enregistré dans une liste est facile, la quitter est très très dur. C'est le même problème lorsqu'une organisation disparait et que ses adresses IP sont récupérées par une autre, qui va s'apercevoir, mais trop tard, que le RIR lui a passé des adresses plombées par une mauvaise réputation <https://www.bortzmeyer.org/evaluation-adresses-ip.html>.

Le RFC demande donc que les gérants de liste changent de perspective. Au lieu de considérer le retrait comme une opération en soi, toute la liste devrait être considérée comme temporaire et la sortie devrait être automatique au bout d'un moment.

Quelle durée avant cette expiration automatique ? Cela dépend de comment est constituée la liste. Si elle est faite manuellement, sur la base d'informations relativement statiques, comme les allocations de préfixes IP, alors les durées de vie peuvent être longues. Si la détection est automatique (par exemple une liste noire de relais HTTP ouverts), alors une durée de vie très courte est plus raisonnable. Ainsi, la machine dont la configuration a été corrigée disparaîtra vite de la liste et la machine qui rechute sera vite réadmise dans la liste. Dans tous les cas, le RFC demande que la politique d'expiration soit elle aussi publiée. Des informations comme « dernière vérification le tant » sont très précieuses pour évaluer la qualité d'une entrée de la base.

Autre bonne pratique, la fourniture d'un canal privé pour demander le retrait d'une entrée dans la base. Il est nécessaire qu'un tel canal soit disponible. Cela peut être aussi simple qu'une adresse de courrier ou qu'un formulaire Web. Mais, en tout cas, une DNSBL ne devrait pas exiger de discussion publique de la demande de retrait. (Oui, certaines le font, comme forme de punition des administrateurs systèmes qui ont fait preuve de négligence, et laissé leur machine être utilisée pour le spam.)

Et, bien sûr, la DNSBL doit fournir un canal de communication qui ne soit pas lui-même bloqué par la DNSBL... Si l'administrateur de 192.0.2.25 veut demander le retrait de cette adresse de la liste, et que le canal des demandes de retrait est une adresse de courrier, sur un serveur qui utilise la DNSBL, on se retrouverait dans une situation très kafkaïenne. Bien sûr, ces adresses de demande de retrait reçoivent beaucoup de spam mais le filtrage qui les protège devrait être très prudent, pour ne pas rejeter des demandes légitimes.

Les réponses à ces demandes de retrait devraient être raisonnablement rapides. Le RFC suggère deux jours (et sept au grand maximum).

Certaines DNSBL n'acceptent de demandes de retrait de la liste noire que lorsqu'il y a eu une forme d'authentification que le demandeur est bien en charge de l'adresse en question (vérification de l'adresse de courrier dans les bases des RIR, par exemple). Cette pratique est déconseillée car une telle authentification est souvent très difficile à fournir dans l'Internet d'aujourd'hui.

Le RFC recommande même de sérieusement envisager la possibilité de retirer automatiquement les adresses incriminées, sur simple requête (politique dite « pas de question » car on ne demande rien au demandeur). Si l'inscription dans la liste est le résultat d'un test automatique, l'adresse IP « coupable » sera très vite remise, de toute façon. Si on craint une guerre d'ajout/retrait dans la base, il suffit de mettre une limitation au rythme des requêtes (par exemple, deux retraits maximum en vingt-quatre heures) ou de détecter ces guerres et de verrouiller alors l'adresse dans la base. Une telle politique « pas de question » permet de corriger très vite les erreurs dans la base, et ne diminue pas sensiblement l'efficacité globale de la DNSBL (qui ne dépend pas de quelques adresses). Enfin, cette politique permet de désamorcer les (nombreux) conflits entre les gérants de la DNSBL et les responsables des adresses noirlistées.

On l'a vu, ce RFC ne veut pas trancher la question de savoir si une politique d'ajout dans la liste est correcte ou pas. Par contre, la liste des bonnes pratiques demande qu'il y ait symétrie entre la politique d'ajout et celle de retrait. Normalement, si une adresse est ajoutée pour une raison X, et que X cesse d'être vrai, l'adresse devrait être retirée. Par exemple, si une DNSBL stocke les adresses IP des relais de courrier ouverts, elle devrait retirer les adresses qui ne sont plus de tels relais (parce que l'administrateur a réparé la configuration). Cette recommandation du RFC va à l'encontre de la pratique de certaines DNSBL de punir les gérants des machines en question, en demandant une autocritique publique, voire carrément de l'argent, pour être délisté.

Enfin, dernière bonne pratique dans la section 2, cette question du paiement. Le RFC explique qu'il est normal qu'une DNSBL fasse payer ses clients pour y accéder (la définition d'une DNSBL commerciale, après tout). Mais faire payer les administrateurs système, responsables des adresses listées, est, comme le dit le RFC avec euphémisme « proche du racket ». Pour convaincre les administrateurs de listes noires que cette pratique devrait être abandonnée, le RFC note que, même lorsqu'elle est faite avec les meilleures intentions du monde, elle est susceptible de déclencher des réactions négatives, et de mettre en péril le principe même des DNSBL.

Les points traités dans la section 2 étaients de nature plutôt politique. La section 3 touche plutôt aux questions techniques. Quelles sont les bonnes pratiques opérationnelles ? Il y en a plusieurs, par exemple l'importance de disposer d'un ensemble de serveurs de noms suffisant pour faire face à la charge. Il s'agit non seulement de la charge normale, mais également de celle, bien plus élevée, qui surviendra lors d'une dDoS (les spammeurs n'ont pas le sens de l'humour et les attaques par déni de service contre les listes noires ne sont pas rares). Pour être sûr que le nom de domaine « principal » de l'opérateur (mettons `spamhaus.org`) soit à l'écart des problèmes, le RFC conseille que la DNSBL soit dans un sous-domaine délégué, avec ses propres serveurs de noms :

```
% dig NS sbl-xbl.spamhaus.org

; <<>> DiG 9.7.3 <<>> NS sbl-xbl.spamhaus.org
;; global options: +cmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 16914
;; flags: qr rd ra; QUERY: 1, ANSWER: 21, AUTHORITY: 0, ADDITIONAL: 1

;; OPT PSEUDOSECTION:
; EDNS: version: 0, flags:; udp: 4096
;; QUESTION SECTION:
;sbl-xbl.spamhaus.org. IN NS

;; ANSWER SECTION:
sbl-xbl.spamhaus.org. 86400 IN NS b.ns.spamhaus.org.
sbl-xbl.spamhaus.org. 86400 IN NS 5.ns.spamhaus.org.
sbl-xbl.spamhaus.org. 86400 IN NS 0.ns.spamhaus.org.
sbl-xbl.spamhaus.org. 86400 IN NS k.ns.spamhaus.org.
sbl-xbl.spamhaus.org. 86400 IN NS d.ns.spamhaus.org.
```

```
sbl-xbl.spamhaus.org. 86400 IN NS o.ns.spamhaus.org.
sbl-xbl.spamhaus.org. 86400 IN NS x.ns.spamhaus.org.
sbl-xbl.spamhaus.org. 86400 IN NS h.ns.spamhaus.org.
sbl-xbl.spamhaus.org. 86400 IN NS l.ns.spamhaus.org.
sbl-xbl.spamhaus.org. 86400 IN NS 2.ns.spamhaus.org.
sbl-xbl.spamhaus.org. 86400 IN NS 4.ns.spamhaus.org.
sbl-xbl.spamhaus.org. 86400 IN NS i.ns.spamhaus.org.
sbl-xbl.spamhaus.org. 86400 IN NS 3.ns.spamhaus.org.
sbl-xbl.spamhaus.org. 86400 IN NS r.ns.spamhaus.org.
sbl-xbl.spamhaus.org. 86400 IN NS f.ns.spamhaus.org.
sbl-xbl.spamhaus.org. 86400 IN NS g.ns.spamhaus.org.
sbl-xbl.spamhaus.org. 86400 IN NS t.ns.spamhaus.org.
sbl-xbl.spamhaus.org. 86400 IN NS 8.ns.spamhaus.org.
sbl-xbl.spamhaus.org. 86400 IN NS q.ns.spamhaus.org.
sbl-xbl.spamhaus.org. 86400 IN NS 1.ns.spamhaus.org.
sbl-xbl.spamhaus.org. 86400 IN NS c.ns.spamhaus.org.

;; Query time: 341 msec
;; SERVER: 127.0.0.1#53(127.0.0.1)
;; WHEN: Sat Dec 17 17:40:40 2011
;; MSG SIZE  rcvd: 388
```

Un des problèmes récurrents avec les DNSBL est que la plupart des administrateurs système ne testent jamais si leur configuration est toujours bonne, et lorsqu'une DNSBL cesse son activité, elle continue à être interrogée. Pour éviter cela, certaines DNSBL en fin de vie ont choisi la solution radicale de lister la totalité de l'Internet. À chaque adresse, elles répondaient qu'elle était listée. Pour détecter rapidement ce genre de problèmes, le RFC rappelle qu'on peut tester `127.0.0.1`. Cette adresse ne devrait jamais apparaître dans une DNSBL. Si elle le fait, c'est que la liste a un gros problème et ne devrait plus être utilisée. Si la liste comprend des noms de domaines et pas des adresses IP, c'est le domaine `invalid` qui joue ce rôle. (Le RFC 5782 donne des détails sur ces conventions. D'autres domaines réservés par le RFC 2606 peuvent être utiles comme `test`.)

Néanmoins, la première responsabilité est celle du gérant de la liste : s'il arrête le service, il doit le faire proprement (prévenir sur son site Web au moins un mois à l'avance, etc) et surtout pas en se mettant soudain à lister tout l'Internet dans sa base. Il est important de garder le nom de domaine actif : si celui-ci était libéré, un méchant pourrait l'enregistrer et monter une fausse liste noire, qui serait utilisée par tous les clients distraits qui ont oublié de changer leur configuration.

Du point de vue technique, la méthode recommandée est de changer les serveurs de noms de la zone où se trouve la liste pour des adresses IP qui ne peuvent pas exister, par exemple celles réservées pour la documentation (RFC 5735). Ainsi, on est sûr que le domaine ne marchera pas et qu'aucun enregistrement ne sera retourné, montrant bien aux clients que la liste n'est plus en service. Par exemple, avec la syntaxe standard des fichiers de zone DNS :

```
dnsbl.example.com.  604800  IN  NS  u1.example.com.
                    604800  IN  NS  u2.example.com.

u1.example.com.     604800  IN  A   192.0.2.1
u2.example.com.     604800  IN  A   192.0.2.2
```

Le RFC a aussi des recommandations opérationnelles à faire pour le cas spécifiques des DNSBL qui listent les adresses IP de machines présentant une certaine vulnérabilité (relais SMTP ou HTTP ouverts, par exemple), détectée par un programme. D'abord, le programme ne devrait pas tester des machines préventivement (la question est controversée : certains défendent la légitimité de balayages

systématiques de tout l'Internet, d'autres ne sont pas d'accord). Il ne devrait le faire que si quelque chose (un rapport d'envoi de spam, par exemple) attire l'attention sur des machines spécifiques.

Ensuite, une fois une adresse listée, les tests périodiques qui visent à évaluer si la machine est toujours vulnérable devraient être relativement espacés (pas plus d'une fois par jour). Et, bien sûr, le test ne doit pas avoir d'effet négatif (pas d'envoi de grandes quantités de données, par exemple).

Les logiciels qui sont derrière la DNSBL, comme tous les logiciels, ont des bogues. Et les utilisateurs d'une DNSBL peuvent faire des erreurs de configuration. Il est donc important que tout soit vérifié et testé et que des procédures soient en place pour faire face aux problèmes (par exemple, le gérant de la DNSBL doit être prêt à vider manuellement la base ou une partie de celle-ci, si une erreur entraîne le listage erronée d'adresses IP).

À noter que le Wikipédia anglophone a un intéressant tableau de comparaison des DNSBL.

Pour résumer, on a vu que ce RFC résultait d'un compromis entre ceux qui voulaient des listes noires opérant comme elles le voulaient et ceux qui souhaitaient les rendre un peu plus responsables. Le compromis a été de donner peu de recommandations concrètes **excepté** de **documenter** les choix effectués. À l'heure actuelle, le moins qu'on puisse dire est que la plupart des listes noires ne font même pas ce minimum...