RFC 7073 : A Reputation Response Set for Email Identifiers

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 22 novembre 2013

Date de publication du RFC : Novembre 2013

https://www.bortzmeyer.org/7073.html

Le cadre général du système de requête sur la **réputation** a été défini dans le RFC 7070[1]. Il reste à le décliner en diverses applications. L'une des utilisations les plus importantes pour un système de réputation est évidemment la lutte contre le spam et c'est donc à cela qu'est consacré notre RFC : la réputation des identifiants de courrier électronique.

L'application se nomme donc `email-id` (et est enregistrée dans le registre des applications <https://www.iana.org/assignments/reputation-parameters/reputation-parameters.xhtml#applications>). Les assertions possibles sur un acteur du courrier sont :

- "*abusive*" : envoie-t-il des messages abusifs, par exemple de harcèlement ou de menaces ?
- "*fraud*" : envoie-t-il des messages frauduleux, par exemple de hameçonnage ? (Voir à ce sujet le RFC 5901.)
- "*invalid-recipients*" : envoie-t-il des messages à des utilisateurs inexistants (ce qui est courant chez les spammeurs, soit parce que leurs listes sont de mauvaise qualité, soit parce qu'ils testent de nouvelles adresses).
- "*malware*" : envoie-t-il du logiciel malveillant ?
- "*spam*" : envoie-t-il du spam ?

Dans le cadre du RFC 7070, ces assertions ne sont pas binaires : une entité a, pour chaque assertion, un classement qui va de 0 (assertion tout à fait fausse) à 1 (assertion vraiment vraie). Ainsi, `example.net` pourrait avoir un classement de 0,01 à l'assertion "*malware*" (il n'envoie quasiment jamais de logiciels malveillants) mais de 0,8 à l'assertion "*spam*" (il envoie souvent du spam). Les classements sont linéaires donc une entité qui aurait un classement de 0,4 pour le spam pourrait être décrite par « est deux fois moins spammeur que `example.net` ».

1. Pour voir le RFC de numéro NNN, `https://www.ietf.org/rfc/rfcNNN.txt`, par exemple `https://www.ietf.org/rfc/rfc7070.txt`

On note que toutes ces assertions sont « négatives », décrivent un comportement qu'on désapprouve. L'annexe A du RFC rappelle que certains pensent qu'il serait plus intéressant de travailler sur des assertions positives (et donc des bonnes réputations), car on peut échapper à des mauvaises réputations (on achète un nouveau nom de domaine et hop, on repart de zéro). Cela se fera peut-être dans le futur.

En réponse à une requête d'application `email-id` et comportant une ou plusieurs de ces assertions (celles qui intéressent le client du service de réputation), le serveur de réputation renvoie un **reputon**, une information structurée en JSON comportant un certain nombre de membres obligatoires (cf. RFC 7071) et, dans le cadre de cette application `email-id`, deux autres :

— `email-id-identity` qui indique comment a été identifié l'émetteur (la réputation ne vaut évidemment pas grand'chose sans authentification mais, parfois, on n'a pas le choix et on a des identités non authentifiées) : DKIM (l'identité doit alors être la valeur de l'étiquette `d=`), SPF, adresse IP, valeur du champ `HELO` dans la connexion SMTP, valeur du champ `MAIL FROM` dans la communication SMTP (ce qu'on nomme le « RFC 5321 from »), valeur de l'en-tête `From:` du message (ce qu'on nomme le « RFC 5322 from »),

— `sources` qui indique le nombre de sources qui ont contribué à l'établissement de cette réputation (le membre standard `sample-size` est le total de rapports, mais ils peuvent tous provenir d'une seule source).

L'identité (`email-id-identity`) est cruciale car on dispose de plusieurs identités dans un message (le `From:` de l'en-tête n'est pas forcément le même que le `MAIL FROM` de l'enveloppe SMTP), et elles n'ont pas toujours la même force. Par exemple, le `From:` de l'en-tête n'a subi aucune validation et peut valoir n'importe quoi. Au contraire, une signature DKIM permet de rattacher un message à un domaine responsable. Si le client n'est intéressé que par un seul type d'identité (par exemple SPF), il peut le préciser dans la requête.

Un exemple de reputon ? OpenDKIM <http://www.opendkim.org/> a un service de distribution de réputation DKIM (donc, seuls les domaines utilisant DKIM y sont présents) :

```
% wget -O - 'http://repute.opendkim.org/repute.php?subject=amazon.com&assertion=spam&application=email-id&se
...
{
  "application": "email-id",
  "reputons": [
    {
"rater": "repute.opendkim.org",
"assertion": "spam",
"rated": "amazon.com",
"rating": 0.000229625,
"identity": "dkim",
"rate": 1013,
"sample-size": 181,
"generated": 1384331169
    }
  ]
}
```