Interdire l'accès à un site Mason à certains

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 20 mai 2009

https://www.bortzmeyer.org/mason-403.html

J'assure la maintenance d'un service Web dynamique écrit en Mason. Plusieurs machines se sont mises récemment à l'utiliser en boucle, violant ses règles d'usage. Comment leur interdire l'accès ?

Mason est un environnement de développement en Perl pour faire des sites Web dynamiques. Il est utilisé notamment pour amazon.com. Il s'appuie sur l'intégration de Perl dans Apache, mod_perl. Comme il tourne sur Apache, on peut bloquer l'accès à certaines adresses IP via les mécanismes Apache normaux <http://httpd.apache.org/docs/2.2/mod/mod_authz_host.html>. Mais j'ai finalement choisi une solution complètement « Masonienne ». L'un de ses intérêts est qu'on a toute la puissance de Perl à sa disposition pour imaginer des règles d'accès aussi baroques qu'on veut (même si cette possibilité n'est pas utilisée ici).

Le service en question <http://www.generic-nic.net/dyn/whois/> est une interface Web à whois, protocole normalisé dans le RFC 3912 [1] pour l'accès aux base de données des registres. Comme une des grosses faiblesses de whois est le fait qu'il faut désigner explicitement le serveur requis, ce service Web essaie de trouver automatiquement le bon serveur (il existe plusieurs méthodes pour cela, le site documente celles utilisées).

Un tel service est relativement rare sur le Web et il est donc logique que certains « petits malins » se mettent à utiliser des robots pour récolter des données via ce site. C'est gênant car, comme le service effectue les requêtes whois pour le compte de ses clients Web, c'est lui qui risque de se faire mettre en liste noire par les serveurs whois, en cas de requêtes trop fréquentes. Je préviens donc le responsable de l'adresse IP « coupable » puis, en cas de non-réponse, je coupe.

Par exemple, 212.27.63.204, alias pperso.free.fr fait de telles requêtes. La base du RIPE-NCC, interrogée en whois, nous apprend que les protestations doivent être envoyées à abuse@proxad.net qui, comme la très grande majorité des services abuse@, ne répond jamais et ne lit probablement pas son courrier <https://www.bortzmeyer.org/personne-ne-s-est-plaint.html>. Place donc aux mesures fascistes.

Le code Mason a été modifié ainsi :

1. Pour voir le RFC de numéro NNN, https://www.ietf.org/rfc/rfcNNN.txt, par exemple https://www.ietf.org/rfc/rfc3912.txt

```
<%once>
my %blacklist = (
    '212.27.63.204'         => 1, # Proxad pperso.free.fr 2009-05-12
    '69.59.186.12'          => 1, # Unknown robot 2009-05-12
    '2001:660:3003:8::4:69' => 1, # Test only 2009-05-12
);
</%once>
```

La section `<%once>` est exécutée une seule fois au chargement de la page. Ici, elle ne fait qu'initialiser le dictionnaire `%blacklist` qui indique les adresses IP des méchants et la date d'entrée dans la liste noire `<https://www.bortzmeyer.org/mettre-date-fichiers-config.html>`.

Ensuite, il faut, pour chaque requête (donc dans la section `<%init>`), tester l'appartenance du client HTTP à cette liste :

```
<%init>
if (exists $blacklist{$ENV{'REMOTE_ADDR'}} ) {
   $m->clear_and_abort(403);
}
...
```

Mason, comme CGI, rend accessible un certain nombre d'informations sur le client, via des variables d'environnement (dictionnaire Perl `%ENV`). `REMOTE_ADDR` est l'adresse IP du client. Si elle est dans la liste noire, on appelle la méthode `clean_and_abort` (dont le nom indique bien ce qu'elle fait) de l'objet `$m` (qui identifie la requête Mason `<http://www.masonhq.com/docs/manual/Devel.html#mason_objects>`, et qui est une variable « magique », créée automatiquement par Mason).

Le paramètre passé à `clean_and_abort`, 403, est le code de statut HTTP. 403, défini dans le RFC 2616, section 10.4.4, signifie « Accès interdit ».

Et voici le rejet d'une connexion, vue dans le journal d'Apache :

```
212.27.63.204 - - [20/May/2009:09:18:52 +0200] "GET /dyn/whois/ask?query=82.230.221.96 HTTP/1.0" 403 403 "-"
```

Voilà, solution rapide et simple, prochaine étape, mettre les adresses IP dans un fichier de configuration plutôt que dans le code Mason, mais, bon, c'est une solution vite fait.

Merci à Éric Redon pour ses idées, son aide et son débogage.