SPAM ASSASSIN

Mathieu KERN -

Décembre 2014

Table des matières

Première partie
Présentation de SpamAssassin

1 Problématique

Le mail(ou courriel) est aujourd'hui le moyen privilégié de communication à travers le monde. Massivement utilisé, d'une certaine fiabilité et éprouvé par des décennies d'utilisation il reste le moyen le plus répandues pour les communications entres les personnes. Malheureusement, mail est également aujourd'hui synonyme de spam, ces messages indésirables qui s'entassent dans nos boites mails. C'est ici qu'entre en jeu SpamAssassin.

1.1 Le SPAM

Avant de poursuivre sur SpamAssassin, rappellons concrètement ce qu'est le SPAM et ce qu'il implique.

Comment reconnaître un SPAM :

- De par sa nature, un SPAM n'est pas désiré par l'utilisateur qui le reçoit.
- La réception d'un SPAM résulte d'un envoi massif : une machine (souvent un bot) envoi le même message à plusieurs destinataire sans aucun discernement. Cela s'oppose aux messages ciblés par exemple de commerçant, qui n'envoie que à leurs prospects.
- Son contenu n'est pas destiné spécifique à l'utilisateur (chaque personne reçoit le même contenu)
- Une importante liste de destinataires
- Entête des messages souvent corrompues ou ne respectant pas les normes

Statut légal La loi pour la confiance dans l'économie numérique du 21 juin 2004 contient une transposition de la directive européenne du 12 juillet 2002 [1] relative à la protection de la vie privée dans le secteur des communications électroniques :

> Est interdite la prospection directe au moyen d'un automate d'appel, d'un télécopieur ou d'un courrier électronique utilisant, sous quelque forme que ce soit, les coordonnées d'une personne physique qui n'a pas exprimé son consentement préalable à recevoir des prospections directes par ce moyen.

1. Le principe introduit figura également à l'article L.34.5 du code des postes et des communications électroniques

Les SPAM sont donc connus du droit français et encadrés par des textes spécifiques.

2 Le projet

2.1 Informations

Développeur	Apache Software Foundation1
Dernière version	3.4.0 (11 février 2014) [+/-]
Environnements	Multiplate-forme
Type	Anti-spam
Licence	Licence Apache 2.0

Le projet SpamAssassin est actif depuis plus d'une décénnies et est constament en devloppement pour s'adapter aux developpements des méthodes qu'utilisent les spammeurs. C'est en outre le programme anti-spam le plus utilisé à cause de son efficacité.

2.2 Développement

SpamAssasin contient environ 300 000 lignes de codes ce qui

2.3 Q'est ce que SpamAssassin

SpamAssassin est un programme écrit en PERL dont le but est de filtrer activement les Emails en se basant sur des mécanismes internes. SpamAssassin n'effectue aucune action envers les mails, il ajoute seulement des informations personnalisés qui peuvent être utilisée par d'autres programmes pour effectuer des actions sur les mails (les rangers des dossiers distincts, les supprimer, les bloquer, ...)

Il peut être utiliser de plusieurs manière :

— En mode client, lancé à chaque fois que l'on fait appel à lui
— En mode demon grâce à *smapd* , les appels au demon étant fait avec l'utilitaire *spamc*.
— Comme une interface de programation : des programmes qui nécéssitent des fonctionnalitées de filtrage de SPAM peuvent s'interfacer avec SpamAssin pour construire des solutions utilisant ses fonctionnalités

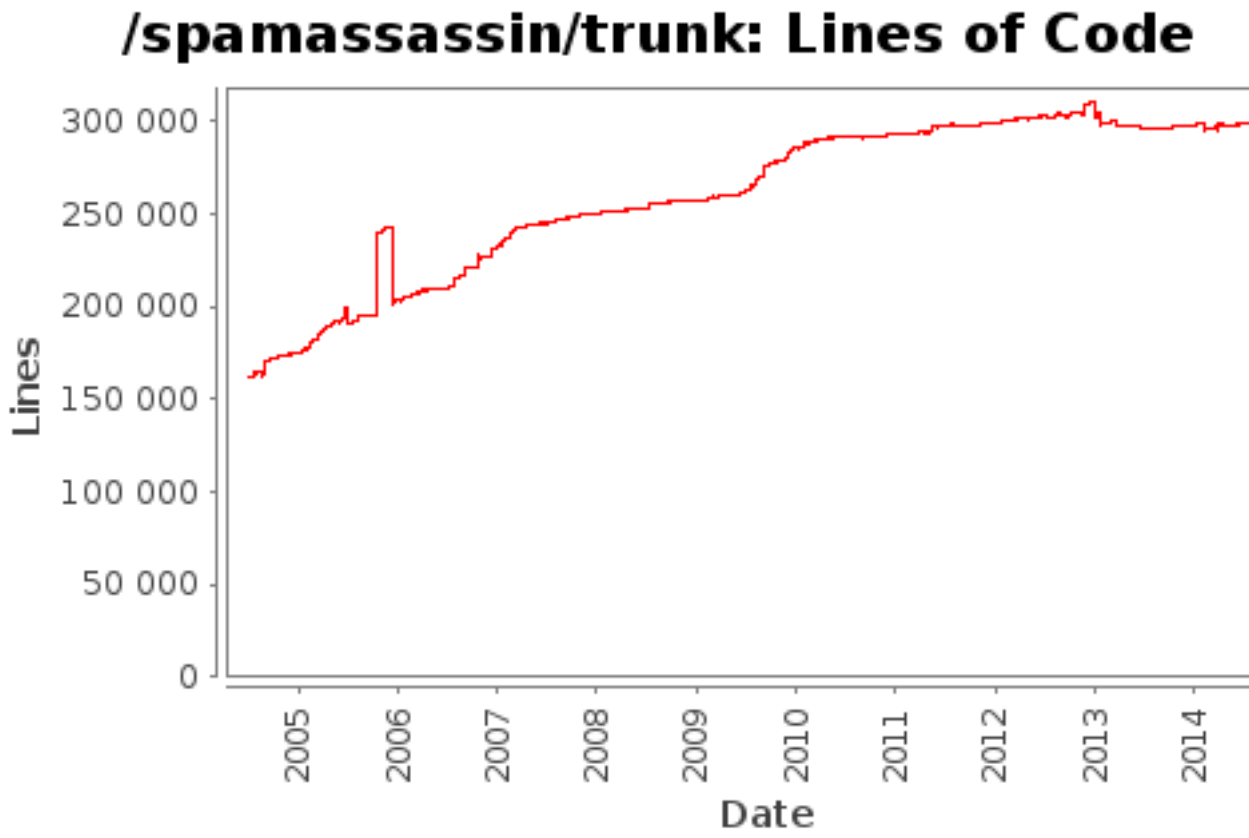

FIGURE 1 – Evolution du nombre de ligne de codes

3 Ses fonctionnalités

3.1 Comment il filtre

Champs d'entête En se basant sur la forme des entêtes et en les comparant avec des schéma connus par SpamAssasin. En effet on peut se base sur la façon dont certains systèmes de SPAMs construisent leurs messages pour les filtrer.

Corps du message Bien sur SpamAssasin permet de filtrer les mails suivant les mots et expressions qu'ils contiennent. "ceci n'est pas un SPAM", "Bonjour jes suis une princesse d'un royaume africain", "Venez chechez votre lot", ...sont des expressions typiques pour des SPAMs.

Filtre bayésien Filtrer les entêtes et le corps d'un message resultera toujours en de multiples faux positifs. C'est ici que le filtarge bayésien se révèle interessant car il va prendre en considération ce que l'on considère comme SPam et non SPAM soit des "bon mails" ("HAM" en anglais). Il va ensuite utiliser les repertoires de SPAMs connu et de "HAM" connus, pour y identifier les mots et phrases (Définits comme "Tokens" en anglais) qui n'apparaissent que dans les SPAMs et que dans les "HAMs". Un token SPAM trouvé resultant d'une hausse du score (voir 3.2) SPAM, un token résultant en une baisse de ce niveau. Ce filtrage permet d'être plus précis et d'éviter les faux positifs, en ne se basant sur un mot ou une phrase mais des enssembles.

List noire/blanche automatique SpamAssassin garde automatiquement une liste blanches des .

Comme précédement si une adresse email envois un SPAM, son score augmente. A l'inverse si elle envoie un "HAM" son score baisse.

Spam Assassin effectue sur chaque mail qui lui est donné à traiter une serie de test, qui vont ensuite donner lieu à un score, qui sera indiqué dans un entête si il est considéré comme SPAM. Ce résultat sera ensuite utilisé par d'autres programmes pour déterminer des actions à entreprendre.

3.2 Le score

C'est la base du fonctionnement du programme. Après avoir effectué une série de test le programme va déterminer une note.