SpamAssassin

Lutter contre le SPAM

Présentation

- SpamAssassin est un logiciel de classification de courriels.
- Il utilise tout un ensemble de règles pour analyser un message électronique et le qualifier de SPAM.
- Il marque ensuite le message.
  - Ajout de HEADER
    - X-Spam-Status
    - X-Spam-Flag
    - X-Spam-Checker-Version
    - X-Spam-Level
  - Il appartient alors à l'utilisateur ou à un autre programme de décider ce qu'il faut faire des messages marqués.

Présentation

- Comment SpamAssassin qualifie t-il un message de SPAM ?
  - Analyse de la conformité des en-têtes avec les RFC
  - Recherche dans les en-têtes et le corps des messages de mots-clefs ou de phrases couramment utilisés dans les messages de type SPAM
    - Viagra, Valium, etc.
    - Recherche multicritères : MUA, html only, avec une ou des images par exemple.
  - Utilisation de listes noires externes
    - Noms de domaines, adresses de MTA, adresses email d'expéditeurs, etc.

Présentation

- SpamAssassin repose sur des modules PERL.
- Un script PERL sert de point d'entrée pour les messages et appelle à son tour d'autres scripts pour chaque tâche.
- Configuration par fichiers
- Il existe une version de SpamAssassin sous forme de démon (spamd) et un client écrit en C pour de meilleures performances.
- SpamAssassin peut être appelé à plusieurs endroits de la chaine SMTP : sur les MTA, depuis les MUA ou bien au moment du dépôt dans les mailbox (MDA)
  - Usage courant : sur les MTA

Tests et règles

- SpamAssassin utilise les tests définis dans son fichier de configuration pour analyser et qualifier les messages.
- Exemples de tests
  - header FROM_STARTS_WITH_NUMS From =~ /^\d\d/
    - header : on recherche dans les en-têtes SMTP
    - FROM_STARTS_WITH_NUMS : nom du test
    - From =~ ... : expression régulière
  - describe FROM_STARTS_WITH_NUMS From: starts with nums
    - Description du test en langage « humain »
  - score FROM_STARTS_WITH_NUMS 0.390 1.574 1.044 0.579
    - Calcul du « scoring » du test

Tests et règles

- Le corps du message contient du javascript :
  - body HTML_OPEN_WIN eval:html_test('window.open')
  - describe HTML_OPEN Contains Javascript Popup
  - score HTML_OPEN 0
    - Note : 0 : test désactivé.
- Scoring
  - Peut être global ou bien défini par utilisateur/domaine
  - Il est également possible d'utiliser une base de données pour stocker les paramètres liés au scoring.

Apprentissage

- SpamAssassin peut être configuré pour affiner son analyse en mode Apprentissage (heuristique).
- auto_whitelisting
  - A partir de l'historique des courriels reçus, SpamAssassin adapte le scoring.
  - Le score attribué à chaque expéditeur va dépendre du nombre de SPAM reçus depuis chaque expéditeur ET du nombre de messages HAM (NON SPAM).

Apprentissage

- Les filtres Bayésiens
- C'est une méthode statistique d'analyse basée sur les probabilités.
- SpamAssassin utilise cette méthode pour décider de manière empirique si un message est plus ou moins probablement un SPAM.
- En principe, plus l'historique est important, plus la détection par cette méthode est pertinente.
- Inconvénient : la configuration est complexe.
- Il est possible (recommandé) de collecter les SPAM qui ont échappé aux filtres pour alimenter l'historique.
- Il est également recommandé de faire « désapprendre » les faux positifs.