LinuxFocus article number 279
http://linuxfocus.org

Lutter contre le "Spam"

par Katja and Guido Socher

<katja/at/linuxfocus.org
guido/at/linuxfocus.org>

L´auteur:

Katja est l’éditrice Allemande de LinuxFocus. Elle aime Tux, le cinéma & la photo et la mer. Sa page personnelle se trouve ici.

Guido est un fan de Linux de longue date et il aime ce système pour les choix et la liberté qu’il offre. Il permet de sélectionner et de développer des solutions adaptées à vos besoins.

Résumé:

Le Spam dans le courrier !? Le Spam dans le courrier électronique augmente à une vitesse alarmante et devient un problème majeur pour presque tout le monde.
Dans cet article nous expliquerons comment réagir contre ce fléau.

Traduit en Français par:
Georges Tarbouriech
<gt/at/linuxfocus.org>

Qu’est-ce que le "Spam" ?

Le Spam possède plusieurs noms. Certains le nomment UCE (Unsolicited commercial email - Courrier commercial non sollicité), d’autres l’appellent Courrier non désiré, mais dans tous les cas ces noms n’expliquent pas de quoi il s’agit vraiment. Si vous ne recevez pas (encore) de spam, consultez cette collection de spam (spam_samples.html). C’est une sélection aléatoire de messages reçus seulement en quelques jours. Lisez-les et vous comprendrez vite que ça n’a rien à voir avec le commerce ou les

affaires. Ces "spammers" sont des criminels. Aucun "commercial" sérieux n'ennuierait ou n'offenserait des millions de personnes afin de trouver quelques "gogos" susceptibles d'acheter ses trucs.

C'est un classique malentendu de la part des personnes ayant peu utilisé Internet de croire que ce type d'annonce peut être comparé aux informations qu'ils reçoivent de temps en temps de leur supermarché. Les produits vendus par ce style de courrier sont souvent illégaux ou bien il ne s'agit même pas de produits. Ce sont des moyens destinés à vous soutirer de l'argent.

Qu'est-ce que ça représente ?

Les "spammers" obtiennent vos adresses de courrier électronique à partir des pages web, des groupes de discussion ou des registres de domaines (si vous possédez votre propre domaine). Certains utilisent des robots pour extraire les adresses, les gravent su CD et les vendent à petit prix à d'autres "spammers". Si votre adresse figure en texte clair sur votre site, de tels programmes sont capables de l'extraire, et dans quelques temps vous aurez un gros problème que vous ne pourrez plus arrêter. Il s'amplifiera tous les jours !

En 1998 le pourcentage de spam envoyé à LinuxFocus représentait moins de 10%. En Novembre 2002, voilà à quoi ressemblent les statistiques :

Notre serveur reçoit environ 4075 messages par semaine. 3273 sont du "spam" !
=> **80% de tout le courrier est du Spam.**

C'est-à-dire que 80% de la capacité du serveur de courrier et 80% de la bande passante sont utilisés pour quelque chose que personne ne veut.

Parmi ces 3273 messages non désirés, environ 40% viennent d'Amérique (essentiellement Canada, Etats-Unis, Mexique) et environ 30% viennent d'Asie (essentiellement Corée, Chine, Taiwan).

Que faire contre le Spam ?

Si vous regardez les "échantillons" vous remarquerez que la plupart offrent la possibilité de se désinscrire de la liste. Ne le faites pas ! Vous avez affaire à des criminels. Aucun des "spammers" ne recevra quoi que ce soit si tant est qu'il maintienne une véritable liste de désinscription. Pourquoi alors continuer à proposer cette possibilité ? La réponse est simple. L'impression sur le lecteur est bien meilleure et c'est de plus un excellent outil statistique. Les "spammers" peuvent ainsi vérifier immédiatement que le courrier a bien été reçu. En d'autres termes **vous accusez réception du courrier !**

L'idée de liste de désinscription pose également un simple problème technique. LinuxFocus n'est pas un très gros site, mais il nous faudrait une personne à plein temps pour désinscrire 3273 messages non désirés par semaine et ce en désinscrivant un courrier par minute. Chaque "spammer" utilise une méthode différente et ce serait un travail stupide et stérile. Les listes de désinscription sont une absurdité et ne servent qu'aux "spammers".

La seule chose à faire consiste à effacer ces messages.

Logiciels pour contrôler le spam

Il existe plusieurs moyens de filtrer ce type de courrier et c'est une très bonne chose puisqu'il devient plus difficile pour les "spammers" de les contourner. C'est malgré tout une course à l'armement. Les outils de filtrage deviennent de plus en plus sophistiqués mais les "spammers" améliorent aussi leurs méthodes.

Il existe deux types de filtres :

1. Les moyens de contrôle disponibles dans le MTA (Message Transfer Agent=Serveur de courrier). Ils permettent habituellement de rejeter le courrier. C'est-à-dire que vous ne le stockez même pas. Vous renvoyez un code d'erreur dès que vous identifiez le courrier comme étant du "spam" pendant la réception. Les outils classiques de ce type sont des listes de blocage basées sur les IP et des vérificateurs d'en-têtes de messages.

2. Le filtrage après réception. Dans ce cas le courrier est effectivement distribué mais il est filtré a posteriori.

Nous allons maintenant étudier les différentes possibilités ainsi que leurs avantages et inconvénients. La meilleure solution pour se débarrasser du spam consiste à utiliser différents outils.

Rejeter le courrier directement depuis le MTA

Si vous rejetez le courrier depuis le serveur pendant la réception, le "spammer" peut recevoir un code d'erreur lui indiquant que l'adresse ne fonctionne pas. S'il fait partie de la catégorie des "graveurs de CD", il est possible qu'il retire l'adresse. Cette méthode permet aussi d'économiser de la bande passante puisque vous ne recevez pas le message complet. Vous pouvez envoyer le code d'erreur dès que vous avez identifié qu'il s'agit de spam.

Pour ce faire, il vous faut un bon MTA le plus souple possible. Malheureusement les deux serveurs les plus répandus, Sendmail et celui de Bill Gates, ne sont pas du tout adaptés à cette tâche. Il existe deux autres très bonnes solutions : Postfix et Exim. Si vous ne pouvez pas changer de serveur vous pouvez ajouter un proxy smtp en amont du serveur (smtp = Simple Mail Transfer Protocol, le protocole de courrier d'Internet).

Nous allons maintenant présenter quelques techniques courantes de filtrage ainsi que leur fonctionnement. Bien sûr, il ne s'agira pas de décrire exactement la configuration de chaque MTA. L'article serait trop long. A la place, je propose de lire la documentation du MTA que vous utilisez. Postfix et Exim sont très bien documentés.

- Listes de blocage en temps réel :
  Ce sont des listes basées sur les DNS. Vous vérifiez l'adresse IP du serveur qui veut vous envoyer du courrier en la comparant à une liste noire des "spammers" connus. Parmi les listes courantes vous avez www.spamhaus.org ou ordb.org. Il existe aussi un outil nommé blq (voir les références) permettant de rechercher manuellement dans ces listes de blocage pour vérifier si telle adresse IP

en fait partie. Ne soyez pas pour autant trop enthousiastes et choisissez soigneusement les listes. Certaines bloquent des étendues entières d'adresses IP seulement parce qu'un "spammer" a utilisé une connexion par modem à partir du prestataire concerné. Personnellement, je recommande au moins ordb.org afin de rejeter le courrier concernant des serveurs mal ou peu administrés. L'expérience montre que ces listes bloquent de 1% à 3% du spam.

- Caractères 8 bits dans la ligne Sujet :
  Environ 30% du spam vient de Chine, de Taiwan ou d'autres pays Asiatiques. Si vous ne lisez pas le Chinois vous pouvez rejeter le courrier comportant des caractères 8 bits (non ASCII) dans le sujet du message. Certains MTA proposent une option de configuration spécifique mais vous pouvez aussi utiliser des expressions régulières correspondant à l'en-tête :

  /^Subject:.*[^ -~][^ -~][^ -~][^ -~]/

  Ceci rejettera le courrier contenant plus de 4 caractères consécutifs non compris dans l'étendue ASCII de l'espace au tilde, dans la ligne "sujet". Si les expressions régulières ne vous sont pas familières, apprenez-les (Voir l'article 53 de LinuxFocus) vous en aurez besoin. Exim et Postfix peuvent être compilés avec le support des expressions régulières de Perl (voir www.pcre.org). Perl possède les plus puissantes des expressions régulières.
  Cette méthode est très bonne et permet de rejeter de 20 à 30% du spam.
- Listes contenant les adresses d'origine ("From") des "spammers" connus :
  Oubliez-les. Elles fonctionnaient en 1997. Aujourd'hui les "spammers" utilisent des adresses truquées ou celles de personnes innocentes.

- Rejet du courrier émanant d'expéditeurs non FQDN (Fully Qualified Domain Name) et de domaines inconnus :
  Quelques "spammers" utilisent des adresses inexistantes dans la ligne "De" (From). Il est impossible de vérifier l'adresse entière mais vous pouvez contrôler la partie hôte/domaine en questionnant le serveur DNS.
  Ceci permet de rejeter de 10 à 15% du courrier non désiré et de toutes façons vous ne pourriez pas répondre à ce type de message même si ce n'était pas du "spam".

- Adresses IP ne possédant pas d'enregistrement PTR dans le DNS:
  Ceci vérifie que l'adresse IP d'où vous recevez le courrier peut être résolue en tant que nom de domaine. C'est une option très puissante capable de rejeter une grande quantité de courrier. Je ne la recommanderais pas ! Elle ne contrôle pas la qualité de l'administrateur du serveur de courrier mais celle du prestataire de "backbone". Les FAI achètent des adresses IP à leurs fournisseurs de "backbone" et de préférence aux plus "gros". Qu'il s'agisse des prestataires de "backbone" ou des fournisseurs d'accès, ils doivent tous configurer leur DNS correctement pour le bon fonctionnement de toute la chaîne. Si au milieu de la chaîne quelqu'un fait une erreur ou se moque de la configuration, rien ne fonctionnera correctement. Rien n'est précisé concernant le serveur de courrier qui se trouve à la fin de la chaîne.

- Commande HELO :
  Lorsque 2 MTAs (serveurs de courrier) discutent entre eux (via smtp) ils disent d'abord qui ils sont (ex. mail.linuxfocus.org). Certains logiciels de spam ne le font pas. Ceci permet de rejeter de 1 à 5% du courrier non souhaité.

- Commande HELO et rejet des serveurs inconnus :
  Vous récupérez le nom obtenu par la commande HELO et vous demandez au DNS s'il s'agit bien d'un serveur enregistré. C'est une méthode efficace parce qu'un "spammer" qui utilise une simple connexion par modem ne prendra sans doute pas le temps de lui configurer un DNS enregistré valide.
  Ceci permet de bloquer de 70 à 80% du spam mais rejette aussi du courrier légitime venant de sites possédant plusieurs serveurs de courrier pour lesquels un administrateur négligent a oublié d'inscrire les hôtes de tous les serveurs dans le DNS.

Certains MTA possèdent encore plus d'options mais celles qui précèdent sont habituellement disponibles dans tout bon logiciel serveur de courrier. L'avantage de ces vérifications vient de ce qu'elles ne sont pas consommatrices de ressources. Le fait d'appliquer ces contrôles ne vous obligera pas à investir dans du nouveau matériel pour votre serveur de courrier.

Filtrage du courrier déjà reçu

Les techniques suivantes sont généralement appliquées à l'ensemble du courrier et le serveur expéditeur ne remarquera même pas que les messages n'ont pas été distribués. Cela signifie aussi qu'un expéditeur légitime ne recevra aucun message de non réception. Le message disparaît tout simplement.
Cela dit, je dois préciser que ce qui précède est incorrect dans la mesure où tout dépend des possibilités de filtrage du serveur de courrier. Exim, par exemple, est très souple et permet d'écrire des filtres personnalisés.

- SpamAssassin (http://spamassassin.org/):
  Il s'agit d'un logiciel de filtrage écrit en perl. Il utilise des règles précises écrites manuellement et attribue des points à certaines chaînes typiques du spam comme "strong by", "you receive this mail because", "Viagra", "limited time offer".... Si les points se situent au-dessus d'un certain niveau, le courrier est considéré comme spam. Le problème de ce type de filtre est d'être très gourmand en mémoire et en temps processeur. Vous devrez sans doute améliorer votre serveur de courrier si celui-ci a déjà 2 ou 3 ans. Personnellement je ne recommande pas d'utiliser ce logiciel directement sur le serveur de courrier. SpamAssassin propose un programme spamd (spamd= le démon spam + spamc= un client pour se connecter au démon) qui contribuera à réduire le temps de démarrage et la consommation de temps machine, mais tout cela reste quand même très exigeant en ressources.

  Pour filtrer le courrier vous devez créer un fichier .procmailrc (et un fichier .forward) semblable à clui-ci :

```
# The condition line ensures that only messages smaller than 50 kB
# (50 * 1024 = 56000 bytes) are processed by SpamAssassin. Most spam
# isn't bigger than a few k and working with big messages can bring
# SpamAssassin to its knees. If you want to run SpamAssassin without
# the spamc/spamd programs then replace spamc by spamassassin.
:0fW:
* < 56000
| /usr/bin/spamc
# All mail tagged as spam (eg. with a score higher than the set threshold)
# is moved to the file "spam-mail" (replace with /dev/null to discard all
# spam mail).
:0:
* ^X-Spam-Status: Yes
spam-mail
```

L’installation est facile et SpamAssassin filtrera plus de 90% du spam.

- procmail (http://www.procmail.org):
  Procmail n’est pas un système de filtrage mais vous pouvez l’utiliser pour en créer un. Procmail est également très léger tant que vous limitez le nombre de règles à un chiffre raisonnable (moins de 10). Pour ce faire, créez un fichier .forward dans votre répertoire personnel et ajoutez-y la ligne suivante :

  ```
  "| exec /usr/bin/procmail"
  ```

  Certains recommandent d’écrire
  "|IFS=’ ’ && exec /usr/bin/procmail"
  mais cela provoque de nouveaux problèmes en créant un processus supplémentaire qui n’est plus sous le contrôle du serveur de courrier. Des serveurs sécurisés comme postfix ou exim n’auront pas ce type de problèmes avec le fichier .forward.

  Procmail est surtout pratique dans un environnement où vous communiquez seulement avec un groupe restreint. Par exemple, pour les employés d’une société dans laquelle la plupart du courrier émane de vos collègues ou de quelques amis. Voici un exemple pour "mycompany.com" :

  ```
  # .procmailrc file.
  # search on header for friends:
  :0 H:
  * ^From.*(joe|paul|dina)
  /var/spool/mail/guido

  # search on header for mails which are not coming from
  # inside mycompany.com and save them to maybespam
  :0 H:
  * !^From.*(@[^\@]*mycompany\.com)
  /home/guido/maybespam

  # explicit default rule
  :0:
  /var/spool/mail/guido
  ```

  Ceci rend l’effacement du spam beaucoup plus facile et vous n’en trouvez plus au milieu de votre courrier normal.

  Procmail est très souple et peut aussi être utilisé à d’autres tâches. Voici un exemple très différent :

  Procmail propose un programme de "réponse à l’expéditeur" nommé formail. Il peut être utilisé par exemple pour renvoyer un message à quelqu’un. Un terrible fléau vient de ces messages contenant des documents Word. Si vous êtes un développeur Linux échangeant des informations par courrier électronique sur vos projets ou sur Linux en général, vous n’avez que faire des gens qui écrivent du texte dans un document Word et le joignent aux messages. Les virus se répandent très facilement de cette manière. Normalement ils n’infectent pas Linux mais ce n’est pas une bonne idée d’utiliser MS-Word pour envoyer du texte puisque cela implique que le destinataire possède la même version pour pouvoir le lire. Il existe des formats ouverts tels que RTF ou HTML qui ne répandent pas de virus, sont lisibles sur toutes les plates-formes et n’ont pas de problèmes

de versions.

```
# Promail script to
# reject word documents. Reject the mail, but do not reply to
# error messages "From MAILER-DAEMON"
# If you use ":0 Bc" instead of ":0 B" then you will still get the mail
:0 H
* !^From.*DAEMON
{
  # The mime messages with word documents look like this in the body
  # of the message:
  #------=_NextPart_000_000C_01C291BE.83569AE0
  #Content-Type: application/msword;
  #        name="some file.doc"
  #Content-Transfer-Encoding: base64
  #Content-Disposition: attachment;
  #        filename="real file.doc"
  :0 B
  * ^Content-Type:.*msword
  | (formail -r ; cat /home/guido/reject-text-msword ) | $SENDMAIL -t
}

# explicit default rule
:0:
/var/spool/mail/guido
```

Le fichier /home/guido/reject-text-msword doit contenir un texte expliquant que les documents MS-Word peuvent répandre des virus et que l'expéditeur doit renvoyer le document au format RTF, par exemple.

Comment utiliser procmail et à quoi correspondent tous ces caractères étranges dans le fichier de configuration est très bien expliqué dans la page de manuel de procmailrc.

- bogofilter (http://www.tuxedo.org/~esr/bogofilter):
  Bogofilter est un système de filtrage Bayesien. Il est écrit en C et il est très rapide (comparé à SpamAssasssin). Un filtre Bayesien est un filtre statistique auquel vous devez d'abord apprendre ce qu'est le spam. Il vous faut environ une centaine de messages pour la période d'apprentissage (triés comme désirés et non désirés) jusqu'à ce que le filtre soit capable d'analyser correctement le nouveau courrier.

  Bogofilter est rapide mais ne fonctionne pas dès le premier jour comme SpamAssassin. Après la phase d'apprentissage, il deviendra aussi performant que SpamAssassin en filtrant plus de 90% du spam.

- razor (http://razor.sf.net/):
  C'est un système de détection distribué et collaboratif. Les checksums des messages reconnus comme spam sont stockés dans une base de données. A la réception d'un nouveau message vous calculez son checksum et le comparez à ceux contenus dans la base de données centrale. Si le checksum trouve une correspondance vous pouvez considérer le message comme spam. Razor fonctionne grâce à des comptes de courrier spéciaux qui ont été disséminés sur Internet dans le but d'entrer dans les listes d'adresses des "spammers". Ces comptes n'attirent que le spam et non le courrier normal. De plus, il est possible d'envoyer des messages à razor afin de les définir en tant que spam. Il est fort probable que ces messages soient déjà reconnus comme spam avant même

d’atteindre votre boîte à lettres. Le système filtre environ 80% du spam. Razor possède une particularité que les autres techniques n’ont pas : razor ne détecte pratiquement aucun faux positifs. C’est-à-dire que le nombre de messages déclarés comme spam mais qui ne le sont pas est très faible avec razor.

De nombreuses autres solutions existent pour combattre le spam. Je pense que celles qui précèdent couvrent l’essentiel.

La meilleure solution consiste à contrôler dans le MTA dans un premier temps et à se débarrasser du spam restant dans un deuxième temps à l’aide d’un filtrage a posteriori.

Courrier HTML

Une forme particulièrement dangereuse de spam est celle du courrier au format HTML.

La plupart des "spammers" proposent la "possibilité de désinscription" afin de voir combien de messages aboutissent. Le courrier au format HTML offre un bien meilleur moyen de contrôle : les images. Vous pouvez comparer ce système à celui des compteurs de visites trouvés sur certaines pages web. Le "spammer" peut connaître exactement le nombre de messages lus et à quel moment ils l’ont été. Si vous étudiez attentivement le spam, vous découvrirez que dans certains cas l’URL des images incluses contient un numéro séquentiel : le "spammer" peut savoir qui lit le courrier et quand. Une incroyable faille de sécurité.

Les programmes modernes de lecture de courrier n’afficheront pas les images téléchargées à partir d’une URL. Toutefois, il n’existe pas franchement de lecteur de courrier HTML moderne et sûr. Kmail et la toute dernière version de Mozilla mail proposent de désactiver les images provenant d’une source externe. La plupart des autres programmes fourniront d’excellentes statistiques au "spammer".

La solution ? N’utilisez pas de programmes aptes à lire le courrier HTML ou bien téléchargez votre courrier et déconnectez-vous avant de lire vos messages.

D’où vient le spam ?

Ne faites jamais confiance à l’adresse de l’expéditeur se trouvant dans le champ "From" du courrier non désiré ! Il s’agit soit d’utilisateurs fantômes ou de victimes innocentes. Il est très rare que ce soit la véritable adresse du "spammer". Si vous voulez savoir d’où vient le courrier vous devez lire l’en-tête complet :

```
...

Received: from msn.com (dsl-200-67-219-28.prodigy.net.mx [200.67.219.28])
        by mailserver.of.your.isp (8.12.1) with SMTP id gB2BYuYs006793;
        Mon, 2 Dec 2002 12:35:06 +0100 (MET)
Received: from unknown (HELO rly-xl05.dohuya.com) (120.210.149.87)
        by symail.kustanai.co.kr with QMQP; Mon, 02 Dec 2002 04:34:43
```

Ici, un hôte inconnu dont l’adresse IP est 120.210.149.87 prétendant être rly-xl05.dohuya.com envoie un courrier à symail.kustanai.co.kr. symail.kustanai.co.kr prend le relais et envoie le message à son destinataire.

Le "spammer" se cache quelque part derrière 120.210.149.87 qui est probablement l'adresse IP dynamique d'une connexion par modem.

En d'autres termes, la police pourrait trouver cette personne en se rendant chez le propriétaire de kustanai.co.kr et en réclamant les logs du serveur et une liste des communications de la compagnie téléphonique locale. Vous avez donc très peu de chances de trouver de qui il s'agit.

Il est également possible que la première partie de l'adresse soit truquée et que le "spammer" se trouve réellement derrière dsl-200-67-219-28.prodigy.net.mx. C'est parfaitement envisageable puisqu'il n'y a aucune raison pour laquelle symail.kustanai.co.kr devrait envoyer le courrier à msn.com par la connexion par modem dsl (dsl-200-67-219-28.prodigy.net.mx). Le serveur.de.votre.fai (nom symbolique) est le serveur de votre prestataire et seule la partie de la ligne "Received:" est fiable.

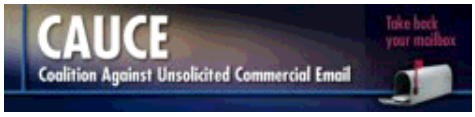
Il est possible de trouver le "spammer" mais il vous faut l'aide des services secrets et de la police pour aller chez prodigy.net.mx.

Conclusion

Si le spam continue à se développer à ce rythme, Internet transportera bientôt plus de spam que de véritable courrier électronique. Le spam circule aux frais du destinataire. De plus en plus de bande passante est nécessaire et souvent les systèmes de courrier doivent être améliorés pour pouvoir gérer le spam.
Les lois de la plupart des pays ne font pas grand chose pour protéger les gens contre ces criminels. En fait, certains pays ont des lois qui servent à pénaliser les honnêtes gens (Gestion des droits "numériques", etc) et à aider les criminels (par exemple à obtenir d'excellentes statistiques sur le spam).

Rejoignez la "Coalition Against UCE" ! (Coalition contre le courrier commercial non désiré).

http://www.euro.cauce.org/en/



http://www.cauce.org/

Les Fournisseurs d'Accès à Internet devraient contrôler leurs systèmes de courrier. Aucun accès non authentifié aux serveurs de courrier ne devrait être accordé et la quantité de message qu'un utilisateur peut envoyer par minute devrait être limitée.

Références

- http://spamassassin.org/: site de spamassassin
- http://www.procmail.org/: site de procmail
- http://www.postfix.org/: site du MTA postfix
- http://www.exim.org/: site du MTA exim
- http://messagewall.org/: site du messagewall smtp proxy
- http://www.unicom.com/sw/blq/: le script perl blq pour questionner les listes de blocage basées sur les DNS

- http://www.ordb.org/: Liste de blocage "open relay" basée sur les DNS
- http://www.spamhaus.org: Liste de blocage basée sur les DNS
- http://www.samspade.org/: D'où vient le spam ?
- http://www.geektools.com/cgi-bin/proxy.cgi: geektools Whois proxy
- http://www.tuxedo.org/~esr/bogofilter/bogofilter, filtrage de courrier
- http://razor.sf.net/: razor
- http://pyzor.sourceforge.net: razor en python
- http://lwn.net/Articles/9460/: Article de Linux weekly news comparant bogofilter et spamassassin.

Site Web maintenu par l´équipe d´édition LinuxFocus	Translation information:
© Katja and Guido Socher	en --> -- : Katja and Guido Socher <katja/at/linuxfocus.org guido/at/linuxfocus.org>
"some rights reserved" see linuxfocus.org/license/	en --> fr: Georges Tarbouriech <gt/at/linuxfocus.org>
http://www.LinuxFocus.org	

2005-01-14, generated by lfparser_pdf version 2.51