SecurityFocus™

Infocus
< http://www.securityfocus.com/infocus/1763 >

Sécurité et Solutions Anti-Spam

par Dr. Neal Krawetz
last updated February 26, 2004
Traduction française personnelle : Jérôme ATHIAS (jerome{@}ATHIAS.fr)
Dernière mise à jour : 11/07/2004

1. Vue d'ensemble

Dans une étude récente, 93% des personnes interrogées ont exprimées leur mécontentement par rapport au large volume de mails (courriers électronique) non sollicités (spam) qu'ils reçoivent. [ref 1] Le problème a pris de l'ampleur au point que pratiquement 50% des mails dans le monde sont du spam [ref 2], alors que seulement quelques centaines de groupes sont responsables. [ref 3] Beaucoup de solutions anti-spam ont été proposées et peu ont été mises en place. Malheureusement, ces solutions n'empêchent pas le spam autant qu'elles interfèrent avec les mails de « tous les jours ».

Les problèmes posés par le spam sont passés du stade de simples contrariétés à d'importants problèmes de sécurité. Le déluge de spam entraîne des pertes estimées à 20 millions de $ par jour en terme de perte de productivité, en accord avec le même document, le spam pour une entreprise peut coûter entre 600 et 1000$ par an, par utilisateur. [ref 4]

1.1 Problèmes de sécurité

En plus du temps de nettoyage passé à voir et supprimer le spam, le spam engendre également des risques de sécurité comme :

- **Vol d'identité.** Le phishing (l'usurpation d'identité) et les arnaques (scams) sont distribués comme du spam, conduisant directement au vol d'identité et à la fraude. Selon le Groupe de Travail Anti-Phishing (Anti-Phishing Working Group), le spam par fishing a augmenté de 52% en Janvier. [ref 5]
- **Virus.** Les nouveaux virus, vers informatiques (worms) et codes malicieux (malwares), comme Melissa, Love Bug, et MyDoom utilisent des techniques de spam pour se propager après avoir été déclenchés par l'utilisateur.
- **Combinaison d'exploits et de spam.** La distinction entre les pirates (hackers) malicieux et les spammeurs (expéditeurs de courrier électronique non sollicité) est devenue moins évidente. Beaucoup de spammeurs ont inclus du code malicieux qui cible les navigateurs, les failles HTML et Javascript. Par exemple, le 31 Décembre 2002, un groupe de pirates au Brésil a envoyé un spam contenant un Javascript hostile à des millions d'utilisateurs. Comme autre exemple ; le récent problème d'affichage d'URL (adresse d'une ressource Internet) dans Internet Explorer, où un « %01 » avant le nom d'hôte (hostname) pouvait être utilisé pour masquer le nom d'hôte réel [ref 6], fut utilisé dans du spam quelques semaines après sa déclaration publique.
- **Combinaison de virus et de spam.** Il est largement suspecté que quelques virus soient conçus pour assister les spammeurs. Par exemple ; le ver SoBig a installé des proxys publics qui furent utilisés pour relayer du spam. Du fait que le spam devient plus répandu, l'utilisation de malwares et de spywares (logiciels espions) pour assister le spam a vraisemblablement tendance à augmenter.

Les solutions anti-spam existantes et proposées tentent d'atténuer le problème du spam et répondent aux besoins de sécurité. En identifiant correctement le spam, l'impact des virus par mails, des exploits, et vol d'identité peut être réduit. Ces solutions mettent en oeuvre différents types de sécurités en vue de déjouer le spam.

Les solutions anti-spam actuelles rentrent dans 4 catégories primaires : filtres, résolution inversée (reverse

lookup), chiffrement avec utilisation de chiffres aléatoires (ou clés) (challenges), et cryptographie. Chacune de ces solutions offrent quelques soulagements face au problème du spam, mais elles ont également des limitations significatives. Le premier volet de ce document en deux parties traite des solutions de filtres et la résolution inverse. La seconde partie se concentre sur les différents types de challenges, comme le challenge-response (chiffrement asynchrone) et les challenges calculés ainsi que les solutions de cryptographie. Comme il y a beaucoup d'aspects différents en regard avec ces solutions, ce document ne présente que les aspects intéressants les plus courants et significatifs – ce document ne se veut pas une liste exhaustive des possibilités de mise en oeuvre, solutions, et problèmes.

1.2 Terminologie usuelle

- **Expéditeur (Sender).** La personne ou le processus qui est responsable de la génération (à la source) du mail.
- **Destinataire (Recipient).** N'importe quel compte email qui reçoit le mail. Cela peut être spécifié dans le mail par un « To : », « CC : », ou « BCC : ».

1.3 Filtres

Les filtres sont utilisés par un système de réception pour identifier et organiser le spam. Il existe différents types de systèmes de filtre incluant :

- **Listes de mots (Word lists).** Listes de mots simples et complexes qui sont connus pour être associés à du spam. Par exemple ; « viagra ».
- **Listes noires (Black lists) et listes blanches (White lists)**. Ces listes contiennent des adresses IP connues d'expéditeurs spammeurs et non spammeurs.
- **Tables de hachage** (**Hash-tables)**. Ces systèmes répertorient des adresses emails sous forme de valeurs quasi-uniques. Les signes répétés de valeurs de hachage sont symptomatiques d'un mailing de masse (bulk mailing).
- **Intelligence Artificielle** et **Systèmes Probabilitaires**. Les systèmes comme les réseaux bayésiens sont employés pour connaître les fréquences d'utilisation de mots et phrases qui sont couramment associés avec parallèlement, les messages de spam et de non-spam.

Les filtres sont classés comparativement en se basant sur leurs résultats faussement-négatifs et faussement-positifs. Un résultat *faussement négatif* indique un message de spam réel qui traverse le filtre. Au contraire, un résultat *faussement positif* indique qu'un mail non-spam est incorrectement classifié comme du spam. Un filtre de spam idéal ne génèrerait aucune analyse faussement-négative et que quelques analyses faussement-positives.

Ces approches anti-spam basées sur le filtrage ont trois limitations significatives :

- **Contournement des filtres**. Les expéditeurs de spam et leurs programmes de mailing en masse ne sont pas statiques – ils s'adaptent rapidement en fonction des filtres. Par exemple, pour contrecarrer les listes de mots, les expéditeurs de spam randomisent l'écriture des mots (« viagra », V1agra, « Viaagra »). Les hash-busters (séquences de caractères au hasard qui changent à chaque email) furent créés pour outrepasser les filtres de hash. Et les filtres bayésiens populaires actuellement sont dépassés par l'inclusion de mots et phrases au hasard. La plupart des filtres anti-spam ne sont efficaces que durant quelques semaines au mieux. Afin de maintenir la viabilité des systèmes anti-spam, les règles de filtrage doivent être mises à jour constamment – en général sur une base journalière ou hebdomadaire.
- **Résultats faussement-négatifs**. Plus un filtre anti-spam est efficace, plus la probabilité de mauvaise classification d'un mail comme du spam diminue. Par exemple, un mail contenant le mot « viagra » (comme le texte de spam « viagra gratuit » ou un mail personnel non-spam « T'as vu la pub marrante pour le viagra pendant la coupe du monde ? ») est presque sûr d'être considéré comme du spam, étant donné le contenu. De même, un mail provenant du subnet Comcast's 24.8.0.0/15 est bloqué par la liste noire SORBS car il est associé avec les adresses DHCP et pas parce que l'expéditeur est associé à du spam. Réciproquement, les filtres anti-spam qui ne génèrent virtuellement aucun résultat faussement-positif génèrent vraisemblablement une grande quantité de résultats faussement-négatifs.
- **Examen du filtre**. Etant donné la possibilité de résultats faussement-positifs, les messages identifiés

comme du spam ne sont pas supprimés immédiatement. Au lieu de cela, ces messages sont placés dans des « boîtes de messages de spam » pour être examinés plus tard. Malheureusement, cela signifie que les utilisateurs doivent toujours voir le spam, ne serait-ce que par le sujet, lorsqu'ils cherchent les mails mal classifiés. Par essence, les filtres ne servent qu'à trier les mails entrants.

Plus important que les limitations des filtres anti-spam, est le mythe courant sur le succès des filtres – il existe une croyance largement répandue comme quoi les filtres arrêtent le spam. **Les filtres anti-spam n'arrêtent pas le spam**. Dans tous les cas, le spam est toujours généré, traverse toujours le réseau, et continue d'être distribué. Et à moins que l'utilisateur ne s'occupe pas de passer occasionnellement à côté de mails souhaitables, le spam est toujours visionné. Tant que les filtres aident à organiser et séparer les mails en groupes « spam » et « non-spam », les filtres n'empêchent pas le spam.

1.4 Résolution inverse

Pratiquement tous les spams utilisent de fausses adresses d'expéditeur ("From:"/ «De :») ; très peu de mails de spam utilisent la véritable adresse de l'expéditeur. Par exemple, en 15 mois, notre base de données de spam a recensée 9300 emails qui prétendent provenir de 2400 domaines uniques. Le domaine « yahoo.com » représentant à peu prêt 20% des adresses d'expéditeur dans le fichier. De même, « aol.com » et « hotmail.com » représentant pour chacun 5%, et « msn.com » représentant 3% du spam, exception faite de tous ces domaines (cumulativement), le spam reçu représente moins de 1%.

Les expéditeurs de spam falsifient les mails pour différentes raisons.

- **Illégales.** Beaucoup de messages de spam sont des arnaques et illégaux dans la plupart des pays. En falsifiant l'adresse de l'expéditeur, l'expéditeur de spam peut rester anonyme et éviter les poursuites.
- **Indésirable.** La plupart des expéditeurs de spam savent que leurs messages sont indésirables. En falsifiant l'adresse de l'expéditeur, ils peuvent atténuer la répercussion de l'envoi de millions de messages à des millions de destinataires exacerbés.
- **Limitations des FAI.** La plupart des Fournisseurs d'Accès Internet ont des clauses de contrat qui interdisent le spam. En falsifiant l'adresse de l'expéditeur, ils diminuent le risque que leur FAI annule leur accès internet.

En résolvant le problème de falsification, les expéditeurs de spam perdraient leur possibilité de rester anonyme. Sans être capable de rester anonyme, les lois comme l'Acte américain CAN-SPAM deviendraient applicables pour les spammeurs opérant depuis les Etats Unis.
Dans le but de limiter la possibilité de falsifier les adresses des expéditeurs, plusieurs systèmes ont été mis au point pour valider l'adresse email de l'expéditeur. Ces systèmes incluent :

- **Echangeur de Mails Inversé (Reverse Mail Exchanger (RMX))**
- <http://www.ietf.org/internet-drafts/draft-danisch-dns-rr-smtp-03.txt>
- **Expéditeur Autorisé Depuis (Sender Permitted From (SPF)**. <http://spf.pobox.com/>
- **Protocole de Mailers Désignés (Designated Mailers Protocol (DMP))** <http://www.pan-am.ca/dmp/>

Ces approches sont sensiblement similaires et sur plusieurs aspects, elles sont identiques. DNS est un service réseau global utilisé pour faire correspondre les adresses IP avec les noms d'hôte et vice versa. En 1986, DNS fut étendu pour associer les enregistrements (« MX ») d'échangeur de mail. [ref 7] Lorsqu'il délivre un mail, un serveur de mail détermine où il doit transmettre le message en se basant sur l'enregistrement MX associé avec le nom de domaine du destinataire.

De manière similaire aux enregistrements MX, les solutions de résolution inverse définissent des enregistrements MX inversés (« RMX » pour RMX, « SPF » pour SPF, et « DMP » pour DMP) afin de déterminer si un mail d'un domaine particulier est autorisé à provenir de n'importe qu'elle adresse IP. L'idée de base est que les adresses email falsifiées ne proviennent pas de la bonne classe d'adresses RMX (ou SPF, ou DMP) et peuvent ainsi être immédiatement identifiées comme falsifiées.

Bien que ces solutions soient viables dans certaines situations, elles partagent quelques limitations significatives.

1.4.1 Domaines sans-hôte et “Vanity domains”

L’approche par résolution inverse requiert que les mails proviennent d’un serveur de mail connu et accepté identifié par une adresse IP connue (l’enregistrement MX-inverse). Malheureusement, la majorité des domaines ne sont pas associés avec des adresses IP statiques. Hormis les cybers squatters, en général ce sont des particuliers et de petites sociétés qui souhaitent utiliser leur propre domaine plutôt que celui de leur FAI, mais ne peuvent pas s’offrir leur propre adresse IP statique et serveur de messagerie. Les services d’enregistrement de DNS, comme GoDaddy, fournissent des services de relais de mail gratuits aux personnes qui enregistrent des domaines sans-hôte ou des vanity domains. Bien que ces services de relais de mails puissent gérer les mails entrants, ils n’offrent pas d’accès pour les mails sortants.

Les solutions de résolution inverse causent quelques problèmes pour ces utilisateurs de domaines sans-hôte et vanity domains :

- **Pas d’enregistrement de MX-inverse**. Les gens qui envoient des mails depuis un domaine sans-hôte ou d’un vanity domain configurent simplement leur application de messagerie pour envoyer des mails depuis leur nom de domaine enregistré. Malheureusement, une résolution de l’adresse IP de l’expéditeur ne trouvera pas le domaine de l’expéditeur. Ce cas est particulièrement courant pour les mobiles, accès commutés (dialup), et autres utilisateurs qui changent fréquemment d’adresses IP.
- **Pas de mails sortants.** Une solution possible nécessite de relayer tous les mails sortants au serveur SMTP du FAI. Cela fournira un enregistrement MX-inverse valide pour les mails envoyés. Malheureusement, beaucoup de FAI ne permettent pas le relais lorsque le domaine de l’expéditeur n’est pas le même que le domaine du FAI.

Dans les deux cas, quelqu’un qui utilise un vanity domain, ou un domaine qui ne possède pas son propre serveur de messagerie, sera bloqué par les systèmes de résolution inverse.

1.4.2 Informatique nomade

L’informatique mobile est une pratique très courante. Les gens prennent leur portable à des conférences, réunions hors-site, et chez eux pour travailler en dehors de la société ou dans un endroit agréable. Hôtels, aéroports, bars, répondent à l’affluence de l’informatique mobile. Malheureusement, la solution de résolution inverse empêche beaucoup d’utilisateurs nomades d’envoyer des mails.

- **Envoi direct.** Il y a deux manières d’envoyer un mail. Un utilisateur peut se connecter à un système de messagerie par le biais d’un compte POP/IMAP/SMTP, web mail ou service similaire, ou bien un utilisateur peut envoyer un mail directement. La plupart des entreprises n’autorisent pas l’accès externe à leurs services de messagerie ; les utilisateurs nomades configurent souvent leurs ordinateurs portables pour envoyer les mails directement. Malheureusement, les problèmes rencontrés en envoyant les mails directement sont exactement les mêmes que les problèmes rencontrés avec les domaines sans-hôte – une résolution inverse du domaine ne donnera pas l’adresse IP de l’expéditeur, et une résolution inverse de l’adresse IP des expéditeurs ne révèlera pas le domaine.
- **Relais de mails**. L’alternative à l’envoi direct nécessite que toutes les sociétés et systèmes de domaines mettent en place des services de messagerie externes pour leurs utilisateurs hors-site et nomades. Dans beaucoup de cas, c’est à la fois non souhaité et non réalisable. Par exemple, d’un point de vue strictement sécurité-réseau, POP3 transmet les noms d’utilisateurs et mots de passe en clair. De cette manière, n’importe quel attaquant reniflant le réseau (sniffing) verra des informations de connection valides. IMAP peut être utilisé avec SSL et supporte l’autentification sécurisée, mais tous les serveurs ne le supportent pas. SMTP supporte également SSL ou TLS, mais là encore, beaucoup de serveurs de sociétés ne supportent pas celà ou utilisent seulement les sertificats coté-serveur. Web mail à travers HTTPS est seulement aussi sécurisé que les certificats coté-client. Du fait que la plupart des sites n’utilisent que les certificats coté-serveur, HTTPS n’offre qu’une très légère protection contre les attaques réseaux de “l’homme du milieu” (man-in-the-middle).

Alors que les solutions de résolution inverse sont viables pour les réseaux internes, elles ne sont globalement pas

réalisables pour une mise en place externe. Les sociétés qui souhaitent supporter les domaines sans-hôte, les vanity domains, et les utilisateurs nomades pourraient vouloir reconsidérer l'implémentation des technologies de résolution inverse anti-spam.

2. Résumé

Le spam prend des allures d'épidémie et les gens cherchent des solutions rapides de toute sorte. Les filtres anti-spam représentent à l'heure actuelle la solution la plus efficace – les filtres tentent d'identifier le spam et limitent l'exposition des destinataires. Mais les filtres n'empêchent pas plus le spam que d'enregistrer une émission de télévision avec un magnétoscope n'empêchent les publicités. Les systèmes de résolution inverse tentent de pallier au problème de falsification. Bien que les résolution inverse soient viables dans un environnement cloisonné, comme un réseau interne d'entreprise, les solutions ne sont pas assez générales pour une admission mondiale.

La seconde partie de cette enquête se concentrera sur les systèmes basés sur le challenge (introduction d'un nombre aléatoire) et les solutions de cryptages disponibles.

A propos de l'auteur

Neal Krawetz possède un Doctorat en Informatique et plus de 15 ans d'expérience dans la sécurité informatique. Le Docteur Krawetz est considéré comme l'un des plus éminant expert dans l'étude du spam et des technologies anti-spam. En plus d'étudier la nature du spam, il dirige l' "Equipe d'Evaluation des Menaces Externes" (ETAT : **External Threat Assessment Team**) de la Secure Science Corporation, une société de services professionnels et logiciels qui développe une technologie avancée pour protéger les actifs en ligne.

Références

[ref 1] "Majority in Favor of Making Mass-Spamming Illegal Rises to 79% of Those Online." The Harris Poll ® #38. July 16, 2003.
[ref 2] "Spam On Course to Be Over Half of All Email This Summer," Brightmail press release. July 1, 2003.
[ref 3] According to SpamHaus, a spam content tracking organization, less than 200 spam groups generate more than 90% of spam messages. SpamHaus ROKSO, September 22, 2003.
[ref 4] Source: "Spam Costs $20 Billion Each Year in Lost Productivity", by Jay Lyman. December 29, 2003.
[ref 5] Source: "Phishing e-mail fraud rises 52% in January, report says", February 18, 2004.
[ref 6] Reference: "Multiple Browser URI Display Obfuscation Weakness"
[ref 7] Source: "Domain System Changes and Observations", RFC973 by Paul Mockapetris. January 1986.