SecurityFocus™

Infocus
< http://www.securityfocus.com/infocus/1766 >

Sécurité et Solutions Anti-Spam, Partie 2

par Dr. Neal Krawetz
last updated February 26, 2004
Traduction française personnelle : Jérôme ATHIAS (jerome{@}ATHIAS.fr)
Dernière mise à jour : 16/07/2004

Note de l'éditeur : la première partie de cette série d'articles est disponible ici.

1. Vue d'ensemble

Le protocole SMTP (Simple Mail Transfer Protocol) n'a jamais été conçu pour la sécurité. SMTP provient d'une extension du protocole FTP de 1973. [ref 1] En 1973, la sécurité informatique n'était pas une préoccupation significative, et les architectes d'Internet n'étaient pas encore certains de leur implémentation du protocole de mails. Par exemple, la RFC 524 décrit les bases du SMTP comme un protocole distinct. L'auteur a inclus cette mise en garde:

> « Bien que quelqu'un puisse (je pense) et pourrait, implémenter un programme sur les bases de ce document, ceci EST VRAIMENT une Requête de Commentaires. Tous commentaires, questions, documents d'avis sont sollicités. Il y a, j'en suis sûr, des bogues dans le protocole spécifié ici, et j'espère que les lecteurs les signaleront via les RFC lorsqu'ils les découvriront. »

Bien que le jeu de commandes ait évolué avec le temps, il apparaît que les gens ont implémenté le SMTP sur les bases de la RFC 524 et il fut supposé que les bogues, comme les problèmes de sécurité, seraient abordés plus tard. Malheureusement, en 2004 les oublis de la RFC 524 sont toujours en train d'être corrigés et le SMTP est trop populaire pour le remplacer du jour au lendemain. Le spam est un exemple d'abus du protocole SMTP – la plupart des outils de spam sont conçus pour falsifiés les entêtes des mails, maquiller les expéditeurs, et cacher le système originaire.

Pour petit rappel de la première partie de cette série d'articles, les solutions anti-spam actuelles rentrent dans quatre catégories primaires: filtres, résolution inverse, systèmes challenges, et cryptographie. Chacune de ces solutions offre quelques soulagements face au problème du spam, mais elles présentent également des limitations significatives. Le premier article traitait des filtres et des solutions de résolution inverse. Cette seconde partie se concentre maintenant sur les différents types de systèmes basés sur le challenge et les solutions de cryptographie. Du fait qu'il existe 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.

1.1 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 : ».

1.2 Challenges

Les expéditeurs de spam utilisent des programmes d'envoi de mails en masse pour générer des millions de mails par jour. Les challenges tentent de gêner les expéditeurs-en-masse en ralentissant le processus d'envoi de mails en masse. Les personnes qui envoient quelques mails à la fois ne devraient pas être impactés significativement. Malheureusement, les challenges ne sont seulement efficaces que lorsque peu de personnes les utilisent. Comme leur popularité augmente, ils tendent plus à interférer avec les mails désirables qu'à dissuader le spam non désiré.

Il existe deux principaux types de challenges : la réponse-challenge et les challenges calculés proposés.

1.2.1 Stimulation-Réponse (Challenge-Response : CR)

Les systèmes de réponse-challenge (RC) maintiennent une liste d'expéditeurs acceptés. Un mail d'un nouvel expéditeur est temporairement stocké sans être distribué. Un mail qui fournit un challenge (habituellement un clic sur un lien ou un mail de réponse) est envoyé au nouvel expéditeur. Après le challenge complété, le nouvel expéditeur est ajouté à la liste des expéditeurs accrédités et le mail original est transmis. La croyance est que les expéditeurs de spam utilisant de fausses adresses email expéditrices ne recevront jamais le challenge, et que les expéditeurs de spam utilisant de véritables adresses email ne pourront pas répondre à tous les challenges. Malheureusement, les systèmes RC présentent un bon nombre de limitations incluant :

- **Impasse RC.** Isabelle demande à Paul d'envoyer un mail à son ami Jérôme. Paul envoie un mail à Jérôme. Le système RC de Jérôme intercepte le mail et envoie un challenge à Paul. Malheureusement, le système RC de Paul intercepte le challenge de Jérôme et émet son propre challenge. Du fait qu'aucun utilisateur ne reçoit réellement le challenge, aucun utilisateur ne recevra le mail. Et du fait que les mails sont non-sollicités et non-attendus, l'utilisateur ne sait jamais qu'il doit examiner le challenge en cours. Par essence, si deux personnes utilisent toute deux des systèmes RC, elles ne pourront pas communiquer ensemble.
- **Systèmes automatisés.** Les systèmes de listes de diffusion (mailing lists) et automatisés comme "Envoyer à un ami" ne peuvent pas répondre aux challenges. (L'argument « vous pouvez toujours les ajouter manuellement » n'est valide que si vous savez que le mail doit arrivé. Je reçois souvent des nouveaux articles que des amis trouvent intéressants et me font suivre. Ces mails ne sont pas attendus et non-sollicités, mais pas indésirables.)
- **Challenges d'interprétation.** Beaucoup de systèmes RC utilisent les challenges d'interprétation. Ces systèmes RC complexes incluent la reconnaissance de caractère ou l'assortiment de motifs qui peuvent facilement être automatisés. Par exemple, le système RC utilisé par Yahoo pour créer de nouveaux comptes email est vulnérable à des systèmes d'IA simples qui réalisent la reconnaissance de caractère. Le système RC de Hushmail requiert de trouver une image sur un fond bleu (parcourir le fond, trouver l'image, et soumettre les coordonnés – pas de problème).

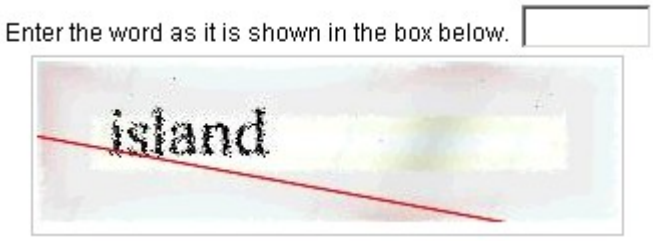

Figure 1. Le challenge de création de compte de Yahoo. Ce système est vulnérable face aux logiciels de reconnaissance de caractères.

Figure 2. Le challenge graphique d'Hushmail. L'utilisateur doit cliquer sur la serrure. Ce système est vulnérable à un traitement simple de l'image.

Le mythe marketing met en relief deux fausses idées: (1) un humain doit réaliser le challenge, et (2) ces problèmes sont trop complexes pour des solutions automatisées. En réalité, la plupart des expéditeurs de spam ignorent ces systèmes RC car ils n'ouvrent pas une grande quantité de boites mails, et ce, non pas parce que le challenge est difficile. Beaucoup d'expéditeurs de spam utilisent des adresses email valides pour leurs arnaques ou pour valider les mailing lists. Lorsque les systèmes RC commenceront à interférer avec les opérations de spam, les spammeurs automatiseront les réponses à ces challenges.

1.2.2 Challenge Calculé

Il existe beaucoup de systèmes de Challenge Calculé (CC) qui tentent d'ajouter un « coût » à l'envoi de mails. La plupart des systèmes CC utilisent des algorithmes complexes qui ont l'intention de prendre du temps. Pour un utilisateur indépendant, le temps n'est vraisemblablement pas ressenti. Mais pour un expéditeur de mails en masse, comme un expéditeur de spam, les petits délais additionnels, font que cela prend trop de temps pour envoyer des millions de mails. Hash Cash [ref 2] et Black Penny de Microsoft sont quelques exemples des systèmes CC proposés. [ref 3] Malheureusement, les systèmes CC possèdent leurs propres ensembles de problèmes d'implémentation qui tendent plus à empêcher leur adoption rapide qu'à empêcher le spam. Quelques exemples de leurs limitations :

- **Pénalisation inégale.** Les challenges calculés, s'ils dépendent du Processeur, de la mémoire, ou du réseau, pénalisent davantage les personnes avec des systèmes plus lents que les personnes avec des systèmes plus rapides. Par exemple, le calcul d'un challenge qui prend 10 secondes sur un ordinateur cadencé à 1Ghz prendra 20 secondes sur un autre à 500MHz. [ref 4]
- **Les listes de diffusion (mailings lists).** Beaucoup de mailing lists comptent des centaines voir des millions de destinataires. Les listes de diffusion populaires, comme le BugTraq, vont être pénalisées autant que les spammeurs. Les challenges calculés rendent la gestion de liste de diffusion impraticable. Et si il existe un moyen pour que les listes de diffusion légitimes puissent passer outre le challenge, alors les spammeurs pourront également passer outre le challenge.
- **Les armées de robots.** Comme nous l'avons constaté avec Sobig et d'autres virus utilisant le spam, beaucoup d'expéditeurs de spam contrôlent des dizaines de milliers de systèmes compromis. Les expéditeurs de spam peuvent aisément distribuer n'importe quels coûts chers à travers leurs systèmes possédés (« 0wned »), annihilant l'impact de n'importe quel coût.
- **Armées légales de robots.** Les expéditeurs de spam génèrent du spam car cela dégage des revenus significatifs. Des groupes de spam importants pourraient offrir d'acheter les services de centaines de systèmes pour distribuer un quelconque coût calculé. Cela pourrait être fait légalement et en ne compromettant aucun système avec un virus.

Les challenges calculés disponibles actuellement ne semblent pas être largement adoptés – ils ne semblent pas réduire le problème du spam et semblent incommoder les mailers légitimes.

1.3 Cryptographie

Seulement quelques solutions utilisant la cryptographie sont disponibles pour vérifier l'expéditeur de spam. Majoritairement, ces systèmes utilisent des certificats pour effectuer l'authentification. Sans un certificat correct, un mail falsifié peut clairement être identifié. Quelques solutions cryptographiques existantes :

- AMTP. http://www.ietf.org/internet-drafts/draft-weinman-amtp-02.txt
- MTP. http://www.ietf.org/internet-drafts/draft-danisch-email-mtp-00.txt
- S/MIME et PGP/MIME. http://www.imc.org/smime-pgpmime.html

Le protocole de mail existant (SMTP) n'intègre pas de support explicite pour l'authentification cryptographique. Certaines de ces solutions disponibles étendent le SMTP (par ex. ; S/MIME, PGP/MIME, et AMTP), alors que d'autres visent à remplacer l'infrastructure mail existante (par ex. ; MTP). De manière intéressante, l'auteur de MTP stipule « SMTP date de plus de 20 ans, alors que des exigences modernes se sont développées ces 5-10 dernières années. Le nombre important d'extensions existantes à la syntaxe et sémantique de SMTP montre, que le SMTP pur ne répond pas à ces exigences et qu'il est trop inflexible pour être étendu sans modification de sa syntaxe. [sic] » [ref 5] Il peut aisément être discuté que le nombre importants d'extensions existantes au SMTP démontre sa flexibilité, et non son inflexibilité, et qu'un protocole de transport de mail complètement nouveau n'est pas nécessaire.

En utilisant des certificats, comme X.509 ou TLS, un certain type d'autorité de certification doit être disponible. Malheureusement, si les certificats sont stockés au niveau du DNS, alors, les clés privées doivent être disponibles pour la validation. (Et si un spammeur a accès aux clés privées, alors il peut générer des clés publiques valides.) Comme alternative, une autorité de certification (CA) centrale reconnue pourrait être utilisée. Malheureusement, le mail est un système distribué et personne ne veut voir une unique CA avoir le contrôle de tous les mails. Beaucoup de solutions autorisent même différents systèmes CA où, par exemple, le certificat X.509 identifie le serveur CA de validation. Cette extension est vulnérable dans le cas où un spammeur fait tourner un serveur CA privé.

Lorsqu'il n'y a pas d'autorité de certification, il faut une méthode de distribution des clés entre l'expéditeur et le destinataire. PGP, par exemple, nécessite des clés publiques pré-partagées. Bien que cette approche soit viable dans des réseaux clos ou pour des groupes d'amis, cela ne s'étend pas très bien au sein de larges groupes d'individus, particulièrement lorsque de nouveaux contacts doivent être établis entre un expéditeur et un destinataire quelconques. Essentiellement, les clés pré-partagées font face aux mêmes problèmes que les listes blanches des filtres : seuls les expéditeurs connus et reconnus peuvent contacter le destinataire.

Malheureusement, ces solutions cryptographiques ne semblent pas arrêter le spam. Par exemple, admettons qu'une de ces solutions (n'importe laquelle) soit acceptée par tout le monde. Ces approches ne valide pas le fait que l'adresse email est réelle – elles ne valident uniquement le fait que l'expéditeur détient les clés correctes pour l'email. Cela créé quelques problèmes :

- **Abus automatisé.** Lorsque implémentée à échelle globale, il sera nécessaire d'avoir un moyen pour générer les certificats ou les clés pour tous les utilisateurs (ou les serveurs de messagerie, ou les clients de messagerie, en fonction de la solution choisie). Le système semble mettre à disposition les clés via une solution automatisée tant qu'il demeure un problème pour une solution manuelle: on estime à 605.60 le nombre de personnes en ligne. [ref 6] Malheureusement, il est réaliste de croire que les spammeurs abuseront de n'importe quel système automatisé après une semaine de déploiement et l'utiliseront pour continuer à envoyer du spam « authentifié ».
- **Problèmes d'utilisabilité.** Il y a également des soucis avec l'utilisabilité. Par exemple, qu'arrive-t-il si le serveur CA est indisponible ? Les mails seront-ils suspendus, retransmis à l'expéditeur, ou considérés comme valides ? Les spammeurs ont récemment conduit une attaque de type déni-de-service longue d'un mois contre environ une demi-douzaine de sites fournissant des listes noires. [ref 7] Une autre croyance largement soutenue, est que les spammeurs ont attaqués les services de listes noires pour empêcher les clients de recevoir les mises à jour. Il n'est pas surréaliste de penser qu'un serveur CA unique (ou même un réseau CA distribué) puisse être la cible d'une attaque similaire.

Résumé des solutions Anti-spam

Le spam prend des allures d'épidémie et les gens cherchent des solutions rapides de toute sorte. Il y a beaucoup de solutions anti-spam existantes et disponibles. Bien que ces options soient viables dans des circonstances limitées, elles semblent toutes avoir des limitations significatives au regard de leur acceptation globale et de leur capacité à empêcher le spam.

Dans la première partie nous avons vu que les filtres anti-spam, alors qu'ils constituent des options viables pour identifier le spam, n'empêchent pas le spam et requièrent une maintenance continue. Les systèmes de résolution inverse (reverse-lookup) tentent d'identifier les expéditeurs falsifiés mais restreignent l'utilisabilité des mails en bloquant les domaines sans hôte et les vanity domains, en restreignant les capacités des utilisateurs mobiles à envoyer des mails de n'importe où n'importe quand. Dans la seconde partie nous avons observé que les systèmes de challenge-response ne sont seulement viables que tant qu'ils maintiennent que peu de caractéristiques, et que les challenges calculés ne semblent pas dissuader les spammeurs. Les solutions cryptographiques, alors qu'elles identifient avec précision les mails falsifiés, ne s'étendent pas simplement à une échelle globale.

Bien que beaucoup de personnes pensent que n'importe qu'elle solution anti-spam vaut mieux que rien, la plupart de ces solutions gênent davantage les utilisateurs classiques plus qu'elles ne freinent les spammeurs. Bien que quelquesunes des options proposées semblent avoir effectivement arrêter le spam lors de tests limités, elles ne prennent pas en compte le fait que les spammeurs adaptent rapidement leur code, dansun délai de quelques jours ou semaines – une bonne solution aujourd'hui ne semble plus être bonne demain.

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**] RFC 458 (Feb. 20, 1973): Mail retrieval via FTP. RFC 510 (May 30, 1973): Network mailbox addresses (user@system). RFC 524 (June 13, 1973): Branching from FTP to a standalone protocol. RFC 561 (Sept. 5, 1973): Standard mail headers.
[**ref 2**] Source: http://www.cypherspace.org/adam/hashcash/.
[**ref 3**] Source: http://groups.yahoo.com/local/service-spamguard.html.
[**ref 4**] The delay is actually more than double due to operating system overhead.
[**ref 5**] Source: http://www.ietf.org/internet-drafts/draft-danisch-email-mtp-00.txt.
[**ref 6**] Source: Nua Internet How Many Online http://www.nua.ie/surveys/how_many_online/index.html, 5-February-2004.
[**ref 7**] Source: "Virus and dDoS Attacks on Spamhaus", http://www.spamhaus.org/cyberattacks/index.html.